Dissertation
submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by

Mgr. Jakub Macošek
born in: Karviná
Oral examination: 22. 11. 2019

Structural insights into regulation of gene expression

Referees: Dr. Christoph Müller
Prof. Irmgard Sinning

Zusammenfassung

Diese Arbeit besteht aus zwei Teilen, in denen strukturelle Aspekte verschiedener Mechanismen der Translationskontrolle untersucht werden.

Das erste Kapitel untersucht den Mechanismus zur Aufhebung der Blockade von Ribosomen (Ribosom stalling). Polyprolin-Abfolgen führen häufig zur Blockade des Ribosoms, die aufgehoben werden muss, um die mRNA translatieren zu können. In Bakterien erfolgt dies durch den Elongationsfaktor P (Elongation factor P (EF-P). EF-P wird durch verschiedene posttranslationale Modifikationen (PTMs) einer positiv geladenen Aminosäure aktiviert. Eine solche PTM ist die Argininglycosylierung, die in etwa 10 % all Bakterienarten, darunter einigen gefährlichen Pathogenen (z.B. *Pseudomonas aeruginosa*) konserviert ist. Arginin wird dabei durch die Glycosyltransferase EarP mit Rhamnose verbunden. Beeinträchtigung der Glycosylierung vermindert die Pathognität der Bakterien. EarP ist jedoch bislang kaum charakterisiert, da es der erste dokumentierte Fall von N-Glycosylierung eines Arginins in Prokaryoten ist. Daher wurde der Mechanismus der EF-P rhamnosylierung durch EarP mit Hilfe von Kernspinresonanzspektroskopie, Röntgenstrukturanalyse und verschiedene biochemische Experimente *in vivo* und *in vitro* untersucht. Die Struktur von EarP mit seinem Substrat dTDP-Rhamnose wurde bestimmt und beleuchtet zusammen mit den in-vivo und in-vitro Experimenten den wahrscheinlichen Mechanismus der EarP Rhamnosylation. Dies legt die Basis für die gezielte Entwicklung neuer Antibiotika.

Das zweite Kapitel behandelt die Hemmung der Translation der hunchback-mRNA (hb). Die hb mRNA bildet während der Embryonalentwicklung der Fruchtfliege Drosophila einen Proteingradienten aus, der zur Ausbildung der Anterior-Posterior-Achse führt. Dieser Gradient entsteht durch die Unterdrückung der mRNA durch einen Komplex aus drei Proteinen – Pumilio (Pum), Nanos and Brain tumor (Brat). Die Expression von Nanos in einem Gradienten mit entgegengesetzter Orientierung zum Hunchback-Gradienten definiert hierbei die räumliche Orientierung. Die Expression der RNA-bindenden Proteine Pum und Brat, die die hb-mRNA spezifisch erkennen, ist hingegen gleichförmig. Sowohl die Struktur des an die hb mRNA gebundenen Brats als auch des Pum/Nanos/hb-mRNA-Komplexes ist bekannt. Es ist jedoch bislang unklar, wie diese drei Proteine zusammen auf der RNA angeordnet sind und ob sie strukturell und funktional direkt verbunden sind. Ein Modell des Komplexes wurde anhand von Daten aus Röntgen- und Neutronenkleinwinkelstreuungsexperimente und Abstandsinformationen aus der Massenspektrometrie des quervernetzten Komplexes (Crosslinking mass spectrometry) erstellt. Diese Daten wurden vervollständigt und bestätigt durch verschiedene *in vitro* assays, wie z.B. Affinitätselektrophorese oder

Isotherme Titrationskalorimetrie. Die Untersuchungen geben erste Einblicke in die Struktur des Komplexes und ebnen den Weg für zukünftige Versuche zur vollständigen Aufklärung der Struktur des Komplexes.

Summary

The thesis comprises two parts investigating structural aspects of various mechanisms of translational control.

The first chapter investigates a mechanism behind ribosome stalling alleviation. Poly-proline stretches often induce ribosome stalling, which needs to be alleviated to translate the mRNA. In bacteria the ribosomes are rescued by Elongation Factor P (EF-P). EF-P is activated by diverse post-translational modifications (PTMs) of a positively charged amino acid. Such PTM is the glycosylation of arginine conserved in approximately 10% of all bacterial species including severe pathogens (e.g. *Pseudomonas aeruginosa*). The arginine is glycosylated by a glycosyltransferase EarP which attaches rhamnose to the arginine. Impairing the glycosylation reduces the pathogenicity of the bacteria. However, EarP is an uncharacterized glycosyltranferase as it is only the first documented case of arginine N-glycosylation in prokaryotes. Hence, the mechanism of EF-P rhamnosylation by EarP was investigated using nuclear magnetic resonance spectroscopy, X-ray crystallography and various *in vivo* and *in vitro* assays. The atomic structure of EarP with its substrate dTDP-rhamnose was solved and the *in vivo* and *in vitro* assays together with subsequent studies elucidate the putative mechanism of EarP rhamnosylation thus providing basis for targeted antibiotic drug design.

The second chapter investigates the translational suppression of the *hunchback (hb)* mRNA. The *hb* mRNA forms during *Drosophila* development a protein gradient governing the anterior-posterior body axis formation. The Hunchback protein gradient results from the suppression of *hb* mRNA at posterior by a complex of three proteins – Pumilio (Pum), Nanos and Brain tumor (Brat). Nanos, expressed in an opposing gradient to Hunchback, provides spatial information. Uniformly expressed Pum and Brat are RNA binding proteins that specifically recognize the *hb* mRNA. The structure of Brat bound to the *hb* mRNA, and the structure the complex of Pum and Nanos bound to the *hb* mRNA have been previously solved. However, it remains unclear how exactly these three proteins assemble on the *hb* mRNA together, and if they are structurally and functional directly linked. The complex was investigated using modelling based on small-angle X-ray and neutron scattering data combined with additional restraints from cross-linking/mass spectrometry. The data were further complemented and validated by various *in vitro* assays such as electrophoretic mobility shift assays and isothermal titration calorimetry. The investigation provides initial insights about the complex and paves the way for future approaches to fully elucidate the structure of the complex.

Acknowledgement

I would like to thank to Dr. Janosch Hennig for his supervision and help with challenging resonance assignments, Dr. Jürgen Lassak by Dr. Ralph Krafczyk as well as other members from the group of Prof. Kirsten Jung and other collaborators for very fruitful collaboration, Dr. Jaelle Foot, Sophie Winter and Alžběta Roeselová for their work on the *hunchback* mRNA suppression project, Dr. Bernd Simon, Dr. Pravin Jagtap and Dr. Pawel Masiewicz for their help with nuclear magnetic resonance spectroscopy, X-ray crystallography and electrophoretic mobility shift assays, respectively and Dr. Kathryn Perez, Dr. Brice Murciano, Dr. Mandy Rettel and Dr. Vladimir Rybin for help in various EMBL core facilities.

I would further like thank Dr. Anne Ephrussi, Prof. Irmgard Sinning and Dr. Christoph Müller for providing scientific advise and mentoring throughout my PhD. I would also like to thank the Hennig group for creating a great working environment, my family and friends for supporting me and Carolina Araújo Sousa for her help with editing and her understanding and support.

Contents

List of Figures

List of Supplementary Figures

List of Tables

List of Supplementary Tables

List of Abbreviations

DMF	Dimethylformamid
DSS	Disuccinimidyl Suberate
DTT	Dithiothreitol
IAA	Iodoacetamide
MD	Molecular Dynamics
PEG	Polyethylene Glycol
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
SDS PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size-Exclusion Chromatography

General Introduction

The most fundamental molecular process to create life is the process of gene expression. The genetic information of an organism during this process gradually manifests into biological function. First, the DNA is transcribed into RNA, which is in turn translated into a protein[8]. It is tightly controlled by several mechanisms which are in sum referred to as the regulation of gene expression. Regulation is possible at any step, so it goes without saying that regulation of gene expression spans a vast plethora of processes. The true importance of those processes regulating gene expression is revealed upon examining genome sizes across species[9]. The increase in functional complexity of organisms does not necessarily stem from an increase in genome size. It must therefore originate in an increase of complexity with which the genomes are expressed. This requires the universal process of gene expression to be plastic. The plasticity of gene expression lies in its regulation. Furthermore, regulation of gene expression allows the organisms to adapt phenotypically to changes in the environment.

Regulation happens both at the level of transcription of a DNA into an RNA and the level of translation of the RNA into a protein. In general, transcription regulation seems to be less conserved and largely associated with phenotypical adaptation and diversity, while translation regulation is more conserved and is postulated to be a buffer[10]. As messenger RNAs (mRNAs) have a generally long half-life, translation regulation is thought to provide fast response to maintain homeostasis. Translation regulation is then crucial for dynamic gene expression in processes common to all organisms such as cell cycle, metabolism, growth and development.

The key mechanisms of translation regulation are the best demonstrated in the context of translation[11]. Translation consists of four phases: initiation, elongation, termination and recycling[12]. Some general mechanisms are shared between eukaryotes and prokaryotes, but there are also substantial differences. The most obvious difference is that in prokaryotes due to the lack of the nuclear membrane, translation is physically coupled to transcription. As a consequence, bacterial mRNAs are often polycistronic (code multiple proteins in a single molecule) and have a rather simple structure. The eukaryotic mRNAs, on the other hand, need to be exported to the cytoplasm, so they are equipped with a m^7G-cap at the 5′ end and a polyA tail at the 3′ end.

The initiation of translation starts with the binding of the small ribosomal subunit (30S or 40S in prokaryotes or eukaryotes, respectively) to the mRNA and of the initiator transfer RNA ($tRNA_i$) to a methionine start codon. The binding

of both the tRNA and the small ribosomal subunit to the mRNA is promoted by numerous initiation factors. In the second step of the initiation the complex of the small ribosomal subunit, the tRNA and the initiation factors on the mRNA recruit the large ribosomal subunit (60S or 70S in prokaryotes or eukaryotes, respectively). As a result, a fully formed ribosome is assembled on the mRNA with the first tRNA ready in the peptidyl site (P site).

This can be already controlled by preventing the assembly of the ribosome on the mRNA in two general manners - either by affecting the initiation factors or ribosomal subunits or by targeting the mRNA itself by the means of RNA-binding proteins or microRNAs[13,14]. The first way provides an opportunity to control translation globally, whereas the second one has consequences only for the synthesis of a specific protein. A prime example of the regulation of translation by RNA-binding proteins is the suppression of *Hunchback* mRNA translation in *Drosophila* development, which is the central topic of chapter 2.

During elongation, the ribosome proceeds to synthesize the protein from the template mRNA. It starts with an amino acid-charged tRNA binding its complementary codon presented in the aminoacyl site (A site) of the ribosome. Then the P site amino acid is transferred to the A site by forming a peptide bond with the previously arrived amino acid. The discharged P site tRNA then translocates to the exit site (E site) of the ribosome and the A site tRNA translocates to the P site. Eventually, the E site tRNA is released from the ribosome. This cycle then repeats itself to synthesize the new protein.

Generally, elongation is thought to occur at its maximum rate and initiation is the rate limiting step in translation. Therefore, initiation is the logical target for efficient translation regulation, especially considering that it is the step when the least has been invested so far. However, regulation of elongation contributes valuably to dynamic regulation of gene expression. The rate of elongation is controlled by the sequence and structure of the mRNA itself and slowing the translating ribosome down provides an opportunity to increase the variety of products. For example, stalled ribosomes can alter the structure of the mRNA downstream, thus allowing synthesis of another protein[15]. It also provides an opportunity to re-program the translation, for example by shifting the ribosome a base upstream or downstream to access an alternative reading frame on the same mRNA[16]. While the first is an example of regulation that plays only a minor role, the second example is an extremely important feature that allows viruses to have particularly compact genomes[17]. A great illustrative example of how simply mRNAs control the rate of the elogantion is proline translation[18]. Prolines are due to their chemical nature incorporated into the nascent peptide chain significantly slower than other amino acids. Consecutive stretches of prolines can then easily slow down and stall the ribosome. The mechanism behind the alleviation of this poly-proline induced stalling in bacteria is the central topic of chapter 1.

When a stop codon reaches the A site of the translating ribosome the termination of translation is triggered. The termination is orchestrated by a set of factors, which first stimulate the cleavage of the peptidyl-tRNA bond causing the release of the peptide and then stimulate the dissociation of the ribosome into subunits.

Eventually, additional factors dissociate the tRNA and together with the ribosomal subunits recycle it for another round of translation. The stop codon itself provides another vital opportunity to increase the variety of translation products. The two main mechanisms to control termination of translation involve either suppressing the stop codon by misreading it by an amino acid-charged tRNA[19] and read-through of the stop codon by erroneously incorporating another amino acid[20]. Both mechanisms result in continued elongation thus synthesizing an alternative gene product.

This illustrates how common and how important regulation of translation is in context of regulation of gene expression. In this work two mechanisms behind regulation of translation were investigated in details.

Chapter 1

Rhamnosylation of Elongation factor P in alleviation of ribosome stalling

1.1 Introduction

1.1.1 Translation proceeds at non-uniform rates

Translation elongation can principally only be used for regulation if substantial control of the rate of elongation is possible. Indeed, the first evidence that translation does not proceed in constant rate has been provided more than 30 years ago[21,22]. Possibly any molecular biologist who ever expressed a protein recombinantly must be aware of this, because the rate of elongation was first linked to the frequency of synonymous codon occurrence - a fact commonly exploited to optimize yields of recombinant proteins[23,24].

However, it is only with the revival of ribosome profiling that the non-uniform rate of translation emerged in its full scale[25]. This prompted the opening of new avenues to study the mechanisms that control the rate of elongation and their roles in translation regulation[26]. Since then evidence emerged that the rate of elongation does not depend only on codon decoding, but also on the amino acid incorporated into the nascent peptide chain[18,27]. Proline was shown to be the amino acid, which is incorporated into the peptide chain particularly slowly. This was suggested to be due to its cyclic chemical nature. In parallel, stretches of consecutive prolines (poly-proline stretches) were found in a large portion of peptides inducing stalling *in vivo*[15,28]. In bacteria, proteins with poly-proline stretches amount to about 10% of the genome[29,30]. To control the poly-proline induced stalling there must also be a mechanism to alleviate it. For that all domains of life use a translation factor called Elogantion Factor P (EF-P) in bacteria, or Initiation Factor 5A (IF-5A) in eukaryotes and archea[31–34].

1.1.2 Consecutive prolines induce stalling

The elongation is fairly universal across organisms[35]. A productive elongation cycle starts with the peptidyl-tRNA bound in the P site of the ribosome. The

unoccupied A site is then bound by a new acetyl-tRNA in a complex with GTP and EF-Tu or EF1A (in prokaryotes or eukaryotes, respectively). The tRNA then base-pairs with the mRNA codon in the A site to decode the codon. The A site tRNA amino acid is then positioned in Peptidyl Transfer Centre (PTC) of the ribosome in vicinity of the nascent peptide chain bound to the tRNA in the P site. A new peptide bond is then formed transferring the nascent peptide chain to the A site tRNA and leaving the P site tRNA deacetylated. The tRNAs (and the mRNA) are then translocated to leave the A site vacant with the use of GTP by EFG or EF2 (in bacteria or eukaryotes, respectively). The deacetylated tRNA is now in the E site and leaves the ribosome. The peptidyl-tRNA is back to the P site with the nascent peptide chain one amino acid longer and ready to initiate another round of elongation.

When poly-proline stretches are translated, the peptide bond formation seems to be the problem causing ribosome stalling. This is thought to be due to the observed tendency of Pro-Pro dipeptides to adopt a conformation which would cause the nascent peptide to clash with the exit tunnel of the ribosome[36]. The nascent poly-proline peptide would be only accommodated by destabilizing the P site peptidyl-tRNA, which in turn would prevent the accommodation of the A site acetyl-tRNA. That causes the ribosome to stall and requires EF-P (IF-5A) to alleviate the stalling.

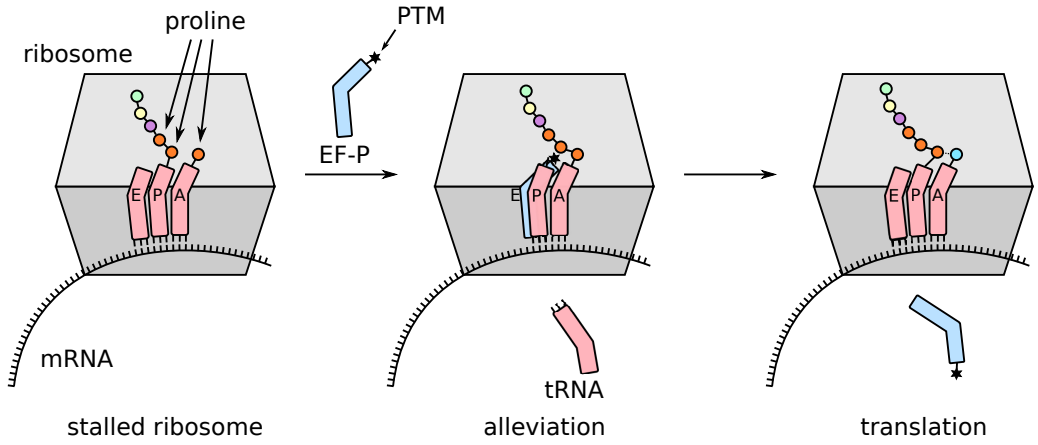


Figure 1.1 Alleviation of ribosome stalling by Elongation Factor P (EF-P). Translation of poly-proline stretches induces ribosome stalling. The stalling is alleviated by EF-P, which is activated by a post-translational modification (PTM) of a positively charged residue in loop at the tip of the N-terminal domain. EF-P binds between the E and P sites of the ribosome, where the modified residue reaches near the Peptidyl Transfer Centre and stimulates peptide bond formation thus alleviating the stalling. The translation resumes saving the nascent peptide chain, the mRNA and the assembled ribosome.

1.1.3 The Elongation Factor P alleviates stalling

Generally, EF-P and IF-5A act in a similar way[33,36–38]. Both factors bind between the P and the E site of the ribosome and extend a loop towards the acceptor stem of the tRNA in the vicinity of PTC. The EF-P (IF-5A) loop contains a conserved positively

charged residue that for full activity of the factor needs to be post-translationally modified. This protruding modified residue interacts with the CCA end of the tRNA forcing a structural rearrangement that alters the conformation of the nascent peptide chain. The conformational change of the nascent peptide chain leads to a favourable substrate geometry and the peptide bond may form. The positively charged residue reaching the acceptor stem was found to be modified by a variety of post-translational modifications (PTMs)[1,31,33,39–41].

1.1.4 Post-translational modifications of the Elongation Factor P

EF-P is an L-shaped molecule consisting of three domains generally resembling the shape of a tRNA[42–44](unpublished, PDBID 1YBY). The first, N-terminal domain folds into a β-barrel comprising six β-strands termed KOW-like domain, which is homologous to SH3-like domain family typical for transcription factors. The second common domain adopts the OB-fold, which is typically a five stranded β-barrel capped by a helix between strand 3 and 4 and is commonly seen binding nucleic acids. The third, C-terminal domain also adopts an OB-fold. The N-terminal KOW-like domain is the domain that harbours the positively charged residue needed for efficient stalling alleviation. The residue is located in a loop between strands β3 and β4. In IF-5A the residues is a conserved lysine. However, the lysine is conserved only in about 26% of bacteria[1]. In about 10% of all bacteria species the conserved lysine is replaced by an arginine. This lysine or arginine is post-translationally modified by diverse pathways.

The lysine of IF-5A was shown to be modified by hypusination[45]. This is a modification of lysine in a two step reaction. First, deoxyhypusine synthase modifies the lysine using spermidine as a substrate. The resulting deoxyhypusine is then hydroxylated by deoxyhypusine hydroxylase to form hypusine. The lysine of EF-P was so far found to undergo two different modifications. The first is a three-step β-lysination and hydroxlation of the lysine. First, the aminomutase EpmB converts an α-lysine into (R)-β-lysine[46]. EmpA then uses (R)-β-lysine to acetylate the EF-P lysine[47,48]. Eventually, the β-lysinated lysine is hydroxylated by EmpC[49]. Alternatively, the lysine of EF-P is modified by 5-aminopentanol[41]. However, the molecular pathway of this modification remains unclear.

The last documented modification system activates the EF-P with a conserved arginine[1]. The arginine is glycosylated by an enzyme termed EarP using dTDP-rhamnose as the substrate. This pathway is conserved in about 10% of all bacteria including some severe pathogens such as *P. aeruginosa*. Moreover, pathogenicity in *P. aeruginosa* was shown to depend on the EF-P rhamnosylation pathway. However, arginine glycosylation was only reported in a very few cases[50,51] and at the start of this work, no molecular details of the catalysis were known. Therefore, we investigated the chemistry of EF-P rhamnosylation by EarP.

1.2 Aims

To shed light into the mechanism of EF-P rhamnosylation by EarP this study aimed to investigate the following aspects of the rhamnosylation:

- The interaction of EF-P with EarP.
- The configuration of the anomeric carbon of the rhamnose throughout the reaction.
- The interaction of EarP and dTDP-rhamnose.
- The structure of EarP.
- The structure of a complex of EarP with dTDP-rhamnose and/or EF-P.

1.3 Results

1.3.1 EarP interacts with the N-terminal domain of Elongation factor P

The investigation of EF-P rhamnosylation was initiated by studying the unmodified and rhamnosylated EF-P by nuclear magnetic resonance (NMR) spectroscopy. First, the resonance frequencies of both versions of EF-P were assigned using the standard triple resonance approach[52] yielding assignment of 98% of the backbone resonances. Then, the interaction of EF-P with EarP was studied by NMR titrations. Both versions of EF-P were expressed ^{15}N labelled, purified and then titrated by unlabelled EarP. The titrations were carried out step-wise until a 1:2 (EF-P:EarP) molar ratio. Unmodified and rhamnosylated EF-P over the course of the titration undergo significant chemical shift perturbations (Figure 1.2). Morever, a number of peaks show substantial line broadening including the rhamnosylation target residue R32 and its surrounding residues, which broaden beyond the limit of detection. This indicates that EarP indeed binds EF-P *in vitro*. The changes

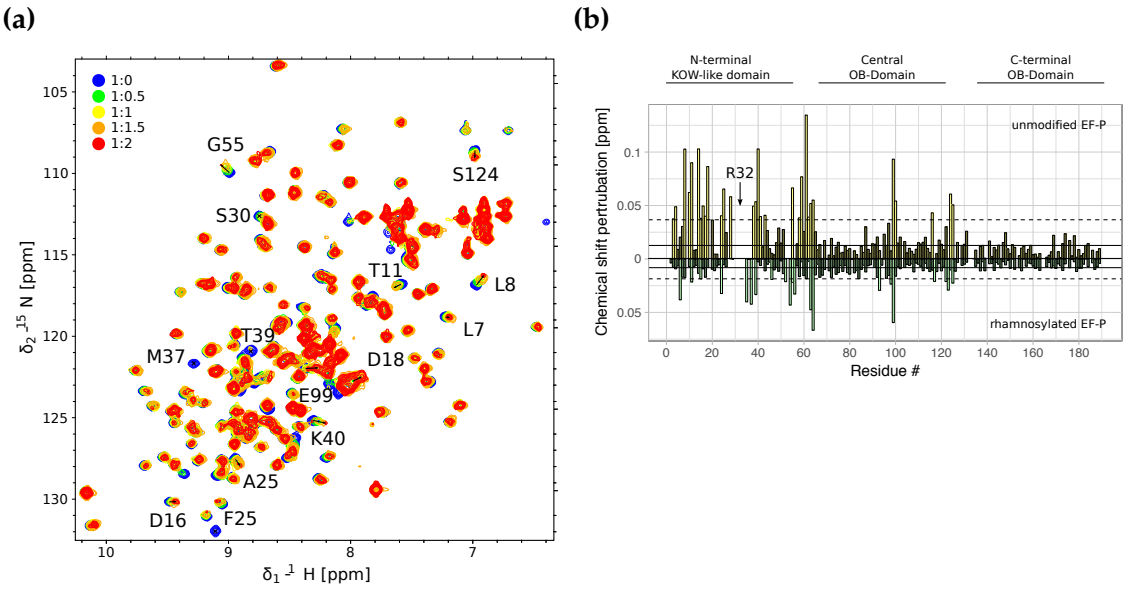


Figure 1.2 The interaction of Elongation Factor P (EF-P) with EarP. (a) Nuclear magnetic resonance spectroscopy titration of EF-P by EarP. Unmodified EF-P was titrated by EarP to 1:2 EF-P:EarP molar ratio and the progress of the titration was monitored by recording a ^{15}N HSQC spectrum at each step of the titration. The recorded spectra are displayed overlaid with colour coding for respective titration steps indicated in the upper left corner. Examples of peaks undergoing high chemical shift perturbations (CSP) or severe line-broadening are shown by labels indicating the assignment of a given peak. **(b)** CSPs of unmodified and rhamnosylated EF-P titrated by EarP. Unmodified and rhamnosylated EF-P were titrated by EarP, the CSPs were calculated according to the Equation 1.1 and plotted against residue number back to back with the CSP unmodified EF-P oriented upwards and the CSP of the rhamnosylated oriented downwards. Full lines indicate mean CSP, dashed lines indicate mean CSP plus standard deviation and residues with CSPs higher than median plus standard deviation are shown in a brighter shade of respective colours. The arrow with the label in the plot indicated the position of the rhamnosylated residue R32. Domain organization is outlined above the plot. The figure is adapted from Krafczyk et al. [2].

of chemical shifts were then for both versions of EF-P assesed quantitatively by calculating the chemical shift perturbations (CSPs, Figure 1.2b). The assessment reveals that the CSPs are overall weaker for the rhamnosylated EF-P than for the unmodified EF-P, which indicate that EF-P interacts stronger with EarP unmodified. This is not surprising considering that modified EF-P is the substrate of EarP and the rhamnosylated EF-P is the product. The highest CSP are located mostly in the N-terminal KOW-like domain as expected considering that it is the site of the rhamnosylated R32. However, almost no above average CSPs were observed in the central and C-terminal EF-P domain. The eukaryotic and archeal homolog of EF-P, IF-5A, interestingly misses the C-terminal domain completely. EF-P was then truncated by removing either the C-terminal domain or both the C-terminal and the central domain and tested for rhamnosylation[2]. The EF-P comprising only the N-terminal domain was still found rhamnosylated.

1.3.2 EarP is an inverting glycosyltransferase

The attached rhamnose moiety was then investigated in detail. First, the sugar ring was assigned using ^{13}C edited NOESY-HSQC spectrum (Figure 1.3a). The main parameter of interest was then the configuration of the anomeric carbon of the sugar ring. The substrate sugar derivative EarP uses is dTDP-β-L-rhamnose. Determining the anomeric configuration of the attached sugar reveals whether it was retained or inverted throughout the reaction. The inversion or retention of the anomeric configuration is a criteria of classification for glycosyltransferases as relates to the chemistry the enzyme uses[53].

Two approaches were taken in order to reveal the anomeric configuration of the bound rhamnose. One bond scalar coupling between the H1′ and C1′ ($^1J_{H1',C1'}$) was reported to be indicative of the anomeric configuration in sugars[54]. For an α-L-rhamnose values around 170 Hz are expected, whereas for a β-L-rhamnose the coupling should only be 160 Hz. This coupling is normally of no importance and would increase the difficulty of the interpretation of a spectrum, so it is commonly suppressed by decoupling. Therefore, ^{13}C-HSQC spectra were measured without decoupling in order to measure the scalar coupling. Three variants of the experiment were measured to ascertain the value - one with the coupling resolved in the direct dimension and two with the coupling resolved in the ^{13}C dimension, one with a large number of points for better resolution and the other with a smaller number of points to not resolve C-C coupling. The $^1J_{H1',C1'}$ coupling was measured to be 167 Hz.

α-L-rhamnose and β-L-rhamnose also differ substantially in the H1′-H5′ distances. In α-L-rhamnose these hydrogens are 3.5 Å apart, whereas in β-L-rhamnose the distance is only 2.4 Å. The distance can be determinend experimentally from the ^{13}C edited NOESY-HSQC spectrum used for the assignment. The NOESY spectrum of rhamnosylated EF-P lacks a cross-peak between the H1′ and the H5′ indicating the longer of the two distances.

Collectively these data show that the attached rhamnose has α configuration on the anomeric carbon. The substrate has β configuration, so this classified EarP

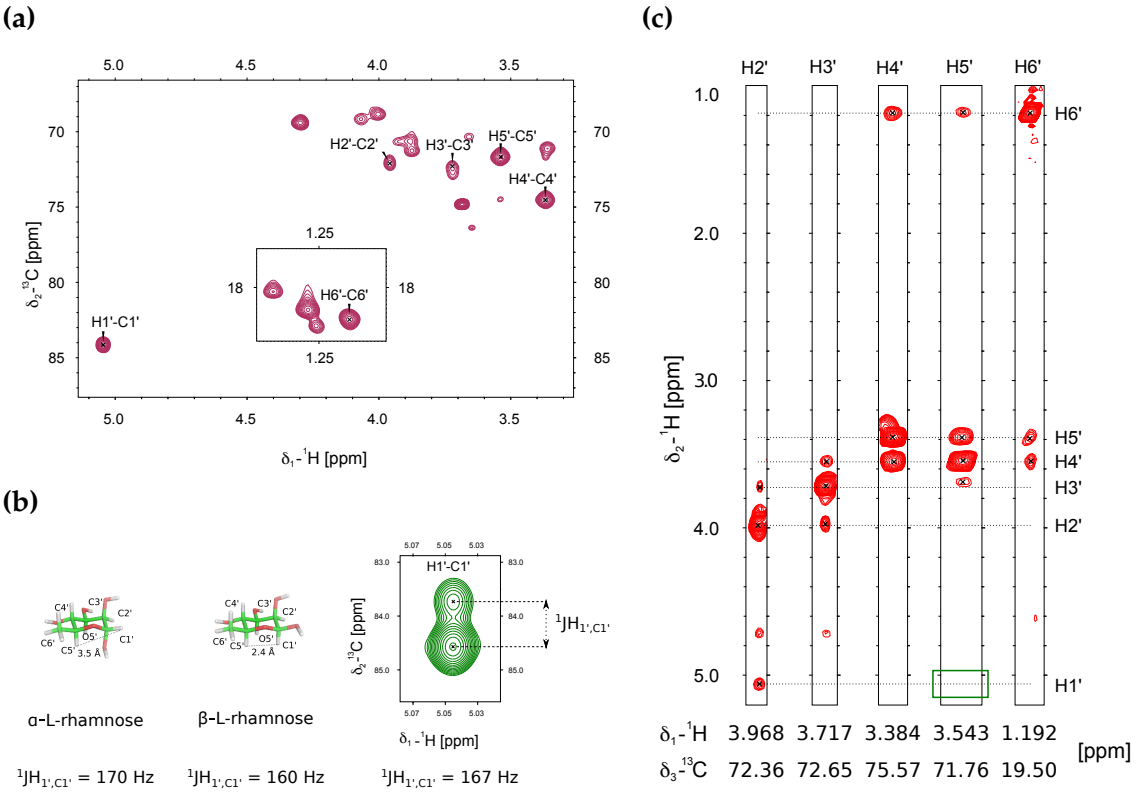


Figure 1.3 The anomeric configuration of the rhamnose attached to Elongation Factor P (EF-P). **(a)** The assignment of the rhamnose attached to EF-P. The sugar resonance region of ^{13}C HSQC of rhamnosylated EF-P is shown with labels indicating the assignment of the sugar ring. The rhamnose moiety was assigned using a ^{13}C edited NOESY-HSQC spectrum. **(b)** Stick representation of α-L and β-L-rhamnose with typical values of $^{1}J_{H1',C1'}$ scalar coupling[54] and H1′ to H5′ distance, and a display of the H1′,C1′ peak of the ^{13}C-HSQC spectrum without decoupling with the measured value of the scalar coupling. **(c)** Strips of the ^{13}C edited NOESY-HSQC spectrum. The green box indicates the lack of the observable NOE between H1′ and H5′. The figure is adapted from Li et al. [55]

as an inverting glycosyltransferase[55,56].

1.3.3 Two opposing Rossman fold domains form EarP

The structure of Earp was then investigated to gain further insights into the rhamnosylation EF-P by EarP. First, EarP was studied by NMR. EarP is in its full-length a 43 kDa protein and is beyond the size limit of structures typically solved by NMR. Unfortunately, soluble truncation of EarP could not be found, so the full length protein was studied by NMR. Deuteration of EarP and the use of transverse-relaxation optimized NMR spectroscopy (TROSY) was required, but eventually led to the assignment of 62% of EarP backbone resonances. ^{2}H,^{15}N-labelled EarP was then titrated by dTDP-rhamnose to 1:5 EarP:dTDP-rhamnose molar ratio. The titration shows that EarP undergoes chemical shift perturbations (Supplementary Figure 1a). While this clearly confirms that EarP binds dTDP-rhamnose *in vitro*, further quantification of the changes by CSP were not particularly informative as

structural information is needed to extract information about the exact interaction surface (Supplementary Figure 1b). The structure of EarP was then investigated by X-ray crystallography.

EarP was co-crystallized with dTDP-rhamnose. The crystallization conditions were optimized to 0.2 M ammonium acetate, 0.1 M bis-Tris, 27% (w/v) PEG 3350, pH 6.0 at 4°C. The crystals provided several full datasets up to 2.3 Å resolution. However, no suitable model could be found for molecular replacement. To experimentally determine the phases EarP was then crystallized again labelled with seleno-methionine (SeMet). The phases were then solved on 3 Å dataset with the anomalous signal extending to 3.4 Å by single-wavelength anomalous dispersion (SAD). The native dataset was then used to extend the resolution to 2.3 Å.

The electron density reveals a well-defined C-terminal domain and partially an N-terminal domain. A weak electron density for the N-terminal domain can be observed, however it is not sufficient for building a model unambiguously. The resulting structure then shows an R-free value of 0.35, which is unusually higher than expected at this resolution . Extensive data quality control was then carried out to find the cause of the bad quality of the electron density for the N-terminal domain. The space group was determined to be I4. To check if it is indeed the correct space group the data were processed again in C2 and P1 space group. This would reveal if non-crystallographic symmetry is mistaken for crystallography symmetry as C2 and P1 are subgroups of I4 with more molecules in the asymmetric unit. However, the electron density did not improve when the data were processed as C2 or P1 space groups. Furthemore the data do not show any signs of anisotropy or twinning. Crystal slippage or radiation damage is not visible during XDS processing. EarP has 6 methionines including the initiatior methionine, which is expected to be disordered as a part of the terminus. Overall 5 SeMet were expected during substructure determination. It is possible to indeed locate 5 selenium atoms, but only the 3 present in the C-terminal domain showed full occupancy. The 2 selenium atoms in the N-terminal domain only have 0.5 occupancy. This might arise from conformational heterogeneity of EarP across different unit cells. Fortunately, the dTDP-rhamnose binding site is located in the part of the structure with the good quality of electron density, so the structure still revealed many important insights about EarP.

EarP comprises two Rossman fold domains (Figure 1.4). The Rossman fold is a super-secondary structure motif characteristic for the alternation of α-helices and β-strands throughout the sequence of the protein. These than arrange so that the β-strands form a central sheet and the α-helices alternate on either side of the sheet. The C-terminal domain of EarP is defined, so the topology of the fold is clearly visible. It features a central β-sheet consisting of four parallel β-strands β8, β9, β11 and β8 and one anti-parallel β-strand β10. The central sheet is then surrounded by α-helices α7-α13 and the C and N-terminal domains are connected by two minor α-helices α5 and α6.

The topology of the N-terminal domain was not completely clear from the structure. Fortunately, secondary structure can be easily determined from NMR chemical shifts, so EarP assignments were then used to calculate the secondary

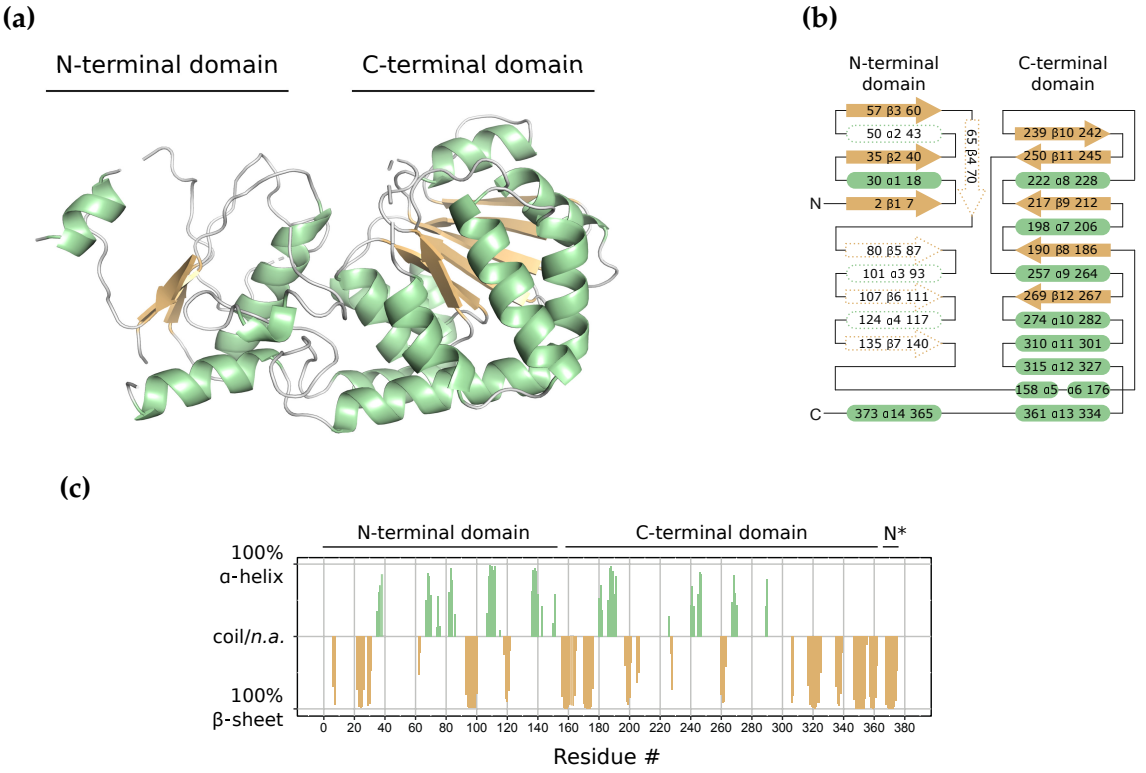
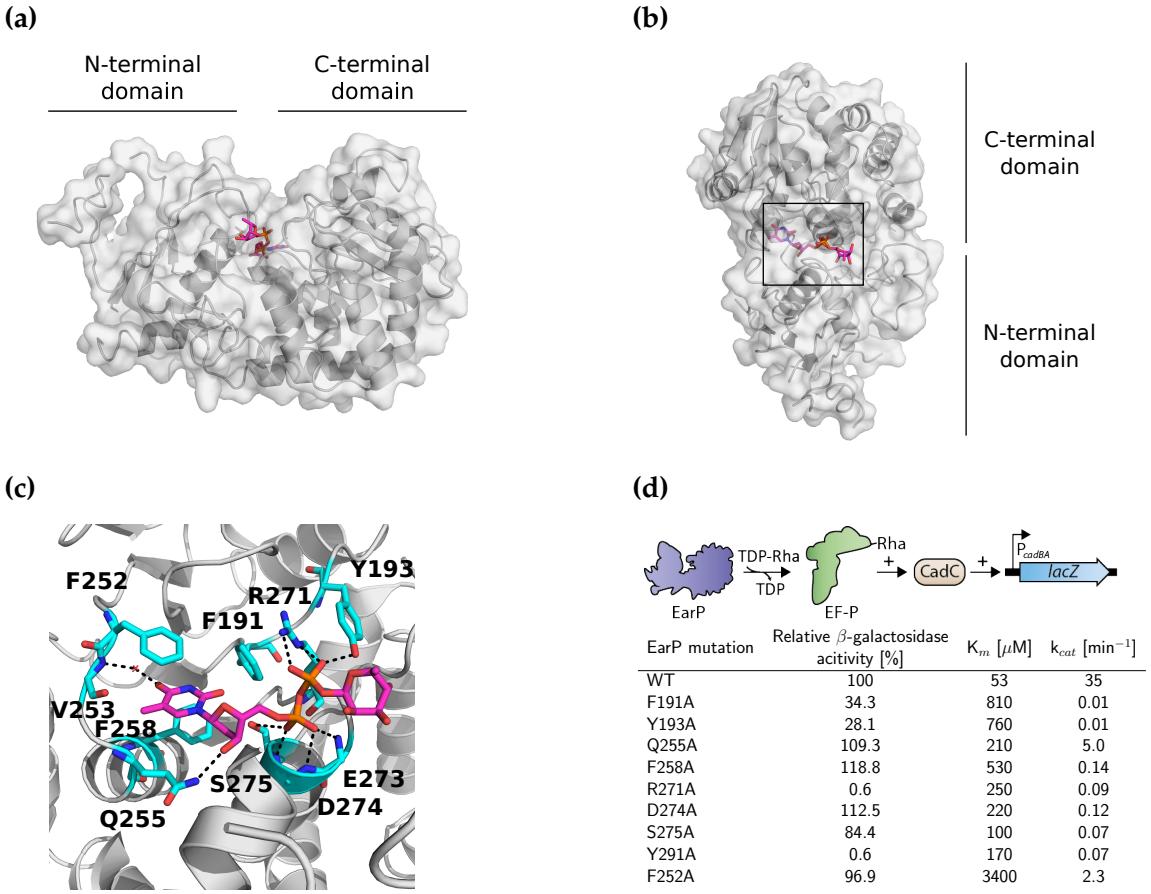


Figure 1.4 The structure of EarP. (a) The structure of EarP determined by X-ray crystallography at 2.3 Å resolution. The structure reveals two Rossman fold domains as indicated above the structure. **(b)** The topology diagram of EarP. The topology of EarP is shown with arrows representing β-sheets and bars representing α-helices. The diagram is divided into the domains which are indicated above. The blank dashed elements indicate secondary structure elements inferred from chemical shifts and homology modeling[2]. **(b)** Secondary structure propensity of EarP determined from NMR chemical shifts. Secondary structure propensity of EarP was calculated from C_α, C_β and backbone N and H secondary chemical shifts using TALOS+[57]. The plot shows the propensity of each residue to form either β-sheet (pointing upwards) or α-helix (pointing downwards). Zero values either indicate the propensity to form a random coil or lack of information. Domain boundaries are indicated above the plot. α-helices are coloured in pale green and β-sheets are colour in beige. The figure is adapted from Krafczyk et al.[2].

structure propensity of EarP (Figure 1.4c). This information was then used together with homology modelling[2] to predict the full topology of the N-terminal domain (Figure 1.4b). The N-terminal domain was predicted to compromise a central β-sheet of β-strands β1 to β7 surrounded by α-helices α1 to α4 and α14, which was visible in the crystal structure to reach back from the C-terminal domain making the N-terminal domain the one to actually harbor both termini.

1.3.4 The inter-domain cleft harbours dTDP-Rhamnose

EarP was crystallized with an excess of dTDP-rhamnose, so the electron density was inspected for a presence of the small molecule. This indeed revealed a molecule of dTDP-rhamnose bound in pocket in the inter-domain cleft (Figure 1.5). Detailed inspection reveals a network of interactions between EarP and



EarP mutation	Relative β-galactosidase acitivity [%]	K_m [μM]	k_{cat} [min^{-1}]
WT	100	53	35
F191A	34.3	810	0.01
Y193A	28.1	760	0.01
Q255A	109.3	210	5.0
F258A	118.8	530	0.14
R271A	0.6	250	0.09
D274A	112.5	220	0.12
S275A	84.4	100	0.07
Y291A	0.6	170	0.07
F252A	96.9	3400	2.3

Figure 1.5 The interaction of rhamnose and EarP. (a) and **(b)** The crystal structure of EarP with bound dTDP-rhamnose. EarP is displayed in grey surface representation, the molecule of dTDP-rhamnose is shown in magenta stick representation. The black square in panel **(b)** outlines the zoomed in area in panel **(c)**. **(c)** Molecular details of dTDP-rhamnose recognition by EarP. The residues of EarP interacting with dTDP-Rhamnose shown in cyan sticks with labels. Dashed lines indicate hydrogen bonds. **(d)** *In vitro* and *in vivo* validation of EarP-dTDP-rhamnose interaction. A list of EarP point mutations was assayed for rhamnosylation activity *in vitro* and *in vivo*. The upper part of the panel shows a diagram of the *in vivo* assay - β-galactoside reporter gene is fused to the P_{cadAB} promoter, which is activated under pH stress by CadC. The CadC translation depends directly on EF-P activity as CadC contains a poly-proline stretch[32]. The results of the *in vivo* assay are given in the second column of the table. The *in vitro* activity was measured by incubating constant amounts of unmodified EF-P and EarP with increasing amounts of dTDP-rhamnose. EF-P rhamnosylation was then quantified by a western blot using antibody against the rhamnosylated arginine[55]. The results of the *in vitro* assays are in the third and the fourth column. K_m is the Michaelis-Menten constant, k_{cat} is the turnover number. The *in vivo* assay diagram was kindly provided by Dr. Lassak. The figure is partially adapted from Krafczyk et al.[2].

dTDP-rhamnose (Figure 1.5c). First, the three aromatic residues F252, F191 and F258 stack against the thymine of the nucleotide moiety. The V253 then via its backbone amide forms a water-mediated hydrogen bond with the O4 of thymine. The side-chain of Q255 forms a hydrogen bond with the O3' in a specific recognition of the ring. Residues E273, D274 and S275 then with their backbone amid groups form several hydrogen bonds with the first phosphate of the diphosphate. The side-chains of R271 and Y193 then form hydrogen bonds with the second phos-

phate. Interestingly, the nucleotide part forms all the contacts with the protein and is buried in the binding site, whereas the rhamnose moiety does not contact the protein at all and is completely solvent exposed. This was further confirmed by saturation transfer difference (STD) NMR experiments, where only signal from the dTDP moiety and not from rhamnose were observed[2].

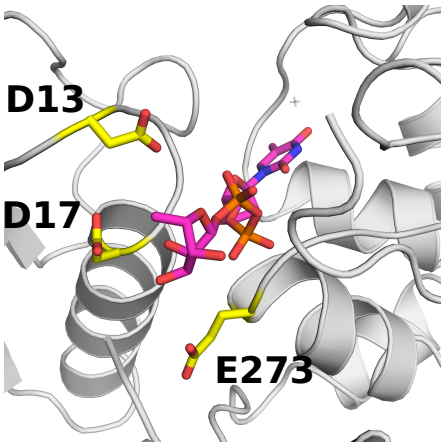
Based on these interactions several of the interacting residues of EarP were mutated to alanines. Then, the activity of the mutated EarP was assayed both *in vivo* and *in vitro*. The *in vivo* assay was a gene-reporter assay based on a previously described EF-P dependent system using CadC (Figure 1.5d)[32]. CadC under pH stress activates the P_{cadAB} promoter. However, CadC contains a poly-proline stretch and directly requires active EF-P for translation. In the assays, endogenous EF-P and EarP are deleted and P_{cadAB} promoter is fused to β-galactosidase reporter gene. β-galactosidase is then used to evaluate the activity of exogenous mutated EarP. The *in vitro* activity of mutated EarP was simply tested by incubating constant amounts of EF-P and EarP with increasing amounts of dTDP-rhamnose , and monitoring the level of rhamnosylation by western blot using antibody against rhamnosylated arginine[55]. F191A, Y193A and R271A mutation showed a strong effect on β-galactosidase activity *in vivo*, while the *in vitro* kinetics were worse for the majority of mutated residues generally validating the structure.

1.3.5 The putative catalytic residues of EarP

The reaction mechanism of inverting glycosyltransferases is S_{N2}-like direct displacement reaction[53]. In such a reaction an acidic residue in the active side of the glycosyltransferase (EarP) serves as a base deprotonating the nucleophile side-chain of the acceptor (EF-P). The acceptor side-chain than displaces the nucleotide resulting in glycosylation. In case of N-glycosylation the deprotonation requires two acidic side-chains[58]. Therefore, sequence alignments of EarP orthologs were combined with details from the structure of EarP to find putative catalytic residues[2]. This approach revealed three residues - D13, D17 and E273 (Figure 1.6a). Each of these is invariant in EarP and in vicinity of dTDP-rhamnose binding site. D13, D17 and E273 were then mutated to alanines and the activity of mutated EarP variants was assessed *in vivo* and *in vitro* as described above. In all three cases the mutation abolishes the activity of EarP *in vivo*. Moreover, no rhamnosylation could be detected *in vitro* whatsoever by EarP containing D13A, D17A or E273A mutations (Figure 1.6b).

The mutations were then tested for dTDP-rhamnose and EF-P binding to see if the loss of activity originates from a loss of binding of either. The rhamnose binding of D13A and D17A variants was tested by performing NMR titrations as for the wild-type(WT) EarP. In both cases similar chemical shift changes were observed as for the WT EarP indicating the interaction is preserved in those mutants[2]. Interestingly, the WT EarP titration shows chemical shift changes in a cluster of residues including G16 and G19. These residues do not directly participate in dTDP-rhamnose binding pocket, but are in the vicinity of D13 and D17 and the binding pocket. That makes these residues valuable indicators of binding of

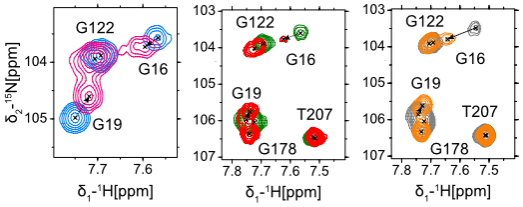
(a)



(b)

EarP mutation	Relative β-galactosidase acitivity [%]	K_m [μM]	k_{cat} [min^{-1}]
WT	100	53	35
D13A	1.9	*n.d.*	*n.d.*
D17A	1.9	*n.d.*	*n.d.*
D273A	2.5	*n.d.*	*n.d.*

(c)



(d)

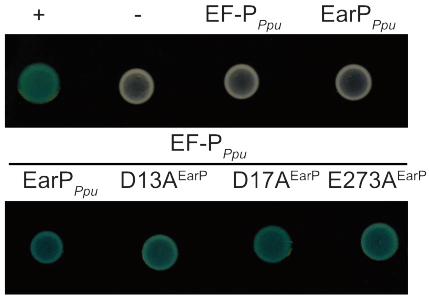


Figure 1.6 The putative catalytic residues of EarP. (a) Molecular details of proposed putative catalytic residues of EarP. The putative catalytic residues are shown in yellow sticks with labels, the dTDP-rhamnose is shown in magenta sticks and EarP is shown in grey cartoon. EarP is an inverting glycosyltransferase which employs S_{N2}-like reaction mechanism, which requires negatively charged catalytic residues to activate the receiving side-chain group[53]. Candidate residues in EarP were identified by searching for conserved negatively charged residues in the vicinity of the bound dTDP-rhamnose[2]. The three putative catalytic residues are invariant in all EarP orthologs. **(b)** *In vitro* and *in vivo* testing of the activity of alanine substitutes of the putative catalytic residues. The results from the *in vivo* assay are in the second column, the results of the *in vitro* assay are in the third and the fourth column. The assay followed the same principles as described in Figure 1.5. The *n.d.* stands for not detectable. **(c)** Test of dTDP-rhamnnose binding by EarP with alanine substitutions of the putative catalytic residues. NMR titrations of EarP variants were done to verify dTDP-rhamnose binding of the EarP with alanine substitutions of the putative catalytic residues. G16 and G19 are in direct vicinity of the D13 and D17 residues and in wild-type EarP their chemical shifts change by the interaction with dTDP-rhamnose. Therefore these residues were observed in the titrations of D13A and D17A as reporters of dTDP-rhamnose binding. The figure shows a section of the ^{15}N HSQC spectra containing the peaks of the G16 and G19 residues, the full spectra are available in Krafczyk et al.[2]. Similar reporter was unavailable for E273 due to missing assignment. Nevertheless dTDP-rhamnose binding was confirmed by STD NMR for E273A EarP[2]. **(c)** Test of EF-P binding to EarP with alanine substitutions of the putative catalytic residues. To test whether the alanine substitutions in EarP impair EF-P binding, bacterial two hybrid assay were performed[59]. The assay tests protein interactions in *E. Coli* by fusing the proteins of interest to the domains of the adenylate cyclese toxin. Protein-protein interactions lead to the reconstitution of the toxin and result in blue phenotype in the commercially supplied kit which was used (Euromedex). The figure is partially adapted from Krafczyk et al.[2].

dTDP-rhamnose in the correct site. This is particularly important as small molecule ligands can often bind multiple sites on a single protein molecule. Observing the G16 and G19 cluster during the NMR titration of D13A or D17A variants in both cases reveals similar chemical shift changes as for the WT EarP (Figure 1.6c).

The D13A, D17A annd E273A EarP variants were then assayed for EF-P bind-

ing using a bacterial two hybrid assay[59]. EarP and EF-P were expressed in *E. Coli* fused to fragments of the adenylate cyclase toxin. Their interaction would lead to the reconstitution of the toxin and result in blue phenotype in the commercially supplied kit which was used. For all three EarP variants in the assay blue phenotype was observed showing that the D13A, D17A and E273A EarP variants still bind EF-P (Figure 1.6d).

In conclusion, the residues D13, D17 and E273 are crucial for EarP glycosyltransferase activity, but are dispensable for EF-P or dTDP-rhamnose binding, which identifies them as promising putative catalytic residues.

1.4 Discussion

This study was quickly followed by more publication by others[3,4]. Together with this structure of EarP from *P. putida* several additional structures were solved. The structure of EarP from *N. meningitidis* was solved in the apo form, in complex with dTDP-rhamnose and in complex with the N-terminal domain of EF-P and dTDP. The structure of EarP from *P. aeruginosa* was solved in the apo form, in complex with dTDP or dTDP-rhamnose and in complex with EF-P and dTDP. These are of tremendous importance as they often fill the gaps in this study. Collectively, these works paint a fairly detailed image of the mechanism of EF-P rhamnosylation by EarP.

The full EarP structure is indeed very similar to the structure proposed here except for some inaccuracies in the prediction of the N-terminal domain, which could not be fully solved due to the poor electron density. The major inaccuracy is that the central β-sheet consists of 8 β-strands instead of the 7 predicted here. However, the C-terminal domain of the *P. putida* EarP presented here aligns very well to the EarP from *P. aeruginosa* (0.6 Å RMSD)[4].

dTDP-rhamnose binds EarP in the conserved binding pocket located in the inter-domain cleft. The binding does not induce any large conformational changes in any of the structures[2–4]. However, the loops in EarP forming the binding pocket of dTDP-rhamnose seem to be disordered in the apo form and well resolved in the bound form suggesting they are stabilized by the donor binding[4]. The interaction with dTDP-rhamnose is very conserved and agrees well with the description here. The base of the nucleotide resides in an aromatic pocket formed by three aromatic residues, the O3' of the deoxyribose is specifically recognized by a glutamine side-chain and the diphosphate is stabilized by a network of backbone and side-chain hydrogen bonds. Additionally, the later studies reveals interactions with the rhamnose moiety[3,4], which are the basis of substrate specificity of EarP, but were not observed here.

The binding of EF-P likely follows the binding of dTDP-rhamnose as saturation of EarP with dTDP-rhamnose was shown to increase affinity for EF-P[4]. Here EF-P was shown to interact with EarP almost exclusively via the rhamnosylated N-terminal KOW-like domain. The domain on its own is sufficient for rhamnosylation *in vitro*[2]. In agreement with this, the structure of the EarP-EF-P complex shows the only the N-terminal KOW-like domain contacting EarP[3,4]. The interaction spans a very large surface covering the N-terminal domain of EF-P which binds to the entrance of the inter-domain cleft of EarP. In *N. meningitidis* EF-P binding induces a domain rotation in EarP, but this is not observed in *P. aeruginosa* EarP[3,4].

The putative catalytic residues were here proposed to be D13, D17 and/or E273. The later studies indeed confirm the importance of these residues. D17 and D13 directly interact with the N_η of R32[3,4]. D17 is then seen in position suitable for deprotonation of the R32 and was therefore proposed to be the base catalyst. The exact role of D13 and E273 is unclear, but the active site residues undergo extensive conformational changes during the reaction, so D13 and E273 may be auxiliary residues facilitating those re-arrangements. Interestingly, in the dTDP-

rhamnose complex D13 forms a hydrogen bond with the side-chain of Y291[4]. Here Y291 was shown to be essential for rhamnosylation both *in vivo* and *in vitro*. These residues together undergo a conformational change which would lead to a clash with the rhamnose moiety in the dTDP-rhamnose complex[4]. Moreover, the rhamnose moiety in the dTDP-rhamnose complex assumes an inactive state with the sugar ring in a conformation incompatible with the S_{N2} reaction[3]. So the conformation changes of residues such as D13, E273 and Y291 was proposed to facilitate conformation change of the sugar ring priming the rhamnose for the nucleophilic attack[3,4].

Furthermore, R271 was here shown to be indispensable for EarP activity *in vivo* and *in vitro*. Curiously, GT-B glycosyltransferases often utilize a positively charged residues during the reaction as well[53]. The positively charged residues would neutralize the emerging negative charge of the diphosphate during the transfer reaction. R271 is in both the *P. putida* EarP structure presented here and in the *P. aeruginosa* EarP[4] structure binding the phosphate of dTDP. This together suggests R271 as a promising candidate for the active site residue facilitating the leaving group departure during the reaction.

In summary, the reaction mechanism of EF-P rhamnosylation emerges as follows. First, the ground state EarP binds dTDP-rhamnose. dTDP-rhamnose stabilizes the active site loops and in turn facilitates EF-P binding. EF-P binding induces further conformational changes in the active site loops. D17 then deprotonates R32, and further residues such as D13, E273 and Y291 force the rhamnose to assume the active conformation. The R32 N_η then performs the nucleophilic attack of the rhamnose further aided by R271, which facilitates the leaving group departure by neutralizing the negative charge of the leaving diphosphate group of the dTDP moiety. Rhamnosylation of R32 then results in steric tensions which lead to the release of rhamnosylated EF-P. dTDP bound EarP is then recycled by the release of the dTDP via unknown mechanisms back to the ground state.

1.5 Materials and Methods

Major part of the experiments described in this chapter were carried our in the laboratory of Prof Dr. Kirsten Jung under the supervision of Dr. Jürgen Lassak and are published[2,55,60]. Therefore, not all experiments are described here and for the details of these experiments please see the published literature.

1.5.1 Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy was used to study the structure and dynamics of EF-P, the rhamnosylation of EF-P, the structure of EarP and the interactions of EF-P and EarP. The details of the experiments are already described[2,55,60], so the following only briefly summarizes the experiments discussed in details in this study.

First, ^{13}C and ^{15}N labelled EF-P was assigned both in its unmodified and rhamnosylated form using the standard triple resonance assignment experiments[52]. The spectra recorded for the assignment were ^{15}N HSQC, HNCA, CBCACONH, and HNCACB (the three later recorded using a constant time during 13C evolution).

Then, a set of experiments was recorded in order to determine the configuration of the anomeric carbon of the rhamnose attached to EF-P. A ^{13}C-edited NOESY-HSQC with simultaneous ^{13}C and ^{15}N evolution during t_2 was recorded to assign the resonances of the attached rhamnose. The $^1J_{CH}$ scalar coupling between the anomeric carbon and directly bound proton (H1'-C1') was then measured by recording three gradient-enhanced ^{13}C-HSQC experiments[61] . The experiments were measured with the ^{1}H decoupling switched off during different parts of the pulse sequence - once during acquisition to resolve the coupling in the ^{1}H direct dimension with 4096 points (160 ms acquisition time) and twice during ^{13}C evolution to resolve the coupling in the ^{13}C dimension with 3584 points (63.6 ms) or with 600 points (10.7 ms) to avoid resolving ^{13}C-^{13}C coupling. The resolved peaks were fitted using the nlinLS built-in function of NMRPipe[62] to determine the scalar coupling.

Due to the large size of EarP (43 kDa) the resonance frequencies were assigned using a ^{2}H, ^{13}C an ^{15}N labelled sample. This allowed to acquire the spectra for the triple resonance backbone assignment in TROSY mode[63,64]. The spectra recorded for EarP assignment were ^{15}N HSQC, HNCA, HNCACB, and CBCACONH all recorded in TROSY mode.

The interactions of EF-P and EarP were studied by NMR titrations. A ^{15}N labelled protein was always titrated by an unlabelled interaction partner and the progress of the titration was monitored by recording a ^{15}N HSQC spectrum after each addition. Unmodified EF-P was titrated by EarP to 1:2 EF-P to EarP molar ratio with subsequent steps at 1:0, 1:0.5, 1:1, 1:1.5 and 1:2 ratios. Rhamnosylated EF-P was titrated directly to 1:2 EF-P:EarP ratio. WT EarP was titrated by dTDP-rhamnose to 1:5 EarP:dTDP-rhamnose molar ratios with steps at 1:0, 1:0.2, 1:1 and 1:3 ratios. The D13A and D17A EarP variants were titrated to 1:10 EarP:dTDP rhamnose ratio without intermediate steps. The NMR titrations were analyzed by

calculating the Chemical Shift Perturbations (CSP) as:

$$CSP_{NH} = \sqrt{(\delta_{\mathrm{Hf}} - \delta_{\mathrm{Hi}})^2 + ((\delta_{\mathrm{Nf}} - \delta_{\mathrm{Ni}})/6.5)^2} \quad (1.1)$$

where $\delta_{\mathrm{Hi/Ni}}$ and $\delta_{\mathrm{Hf/Nf}}$ are initial and final proton/nitrogen chemical shifts, respectively and 6.5 is a scaling factor accounting for nitrogen shifts spanning broader range than proton shifts[65].

The NMR experiments were measured on Bruker Avance III spectrometers with a magnetic field strength corresponding to a proton Larmor frequency of 600 MHz (equipped with a Bruker TXI cryogenic probe head), 700 MHz (equipped with a Bruker room temperature probe head), or 800 MHz (equipped with a Bruker TXI cryogenic probe head). All measurements were done at 25°C . The spectra were all processed in NMRPipe[62] using a gaussian window function and zero filling and the analysis and assignment was done in CARA (http://cara.nmr.ch) and CcpNmr Analysis[66].

The backbone resonance assignments of EarP and EF-P have been deposited to the Biological Magnetic Resonance Bank (BMRB) under the accession numbers 27091 and 27090, respectively.

1.5.2 X-ray crystallography

X-ray crystallography was used to obtain insights into the molecular structure and the chemistry of EarP. The crystallization trials were set up with N-terminally tagged His_6-EarP at 8 mg/ml in 50 mM Tris, 100 mM NaCl, 1 mM DTT, pH 7.6 buffer with 10 mM TDP-Rhamnose. The trials were set up in the EMBL Heidelberg Crystallization Facility using the Mosquito crystallization robot (TTP LabTech) in MRC crystallization plates (2 wells, round bottom). The screens used the vapour diffusion crystallization method in the hanging drop set up with a drop mixed of 100 nl of the protein samples and 100 nl of the mother liqour. The crystallization was tested at 20°C and 4°C. The set of the commercial screens tested is listed in Table 1.1. Eventually, the crystallization conditions were found in the JCSG+ screen (Molecular Dimensions) at 4°C. The initial crystals did not diffract well, so the condition was further refined in a custom design 24-well plate screen to 0.2 M ammonium acetate, 0.1 M bis-Tris, 27% (w/v) PEG 3350, pH 6.0. The details of the optimization screens are in the Supplementary Table 1, Supplementary Table 2 and Supplementary Table 3.

Several complete datasets were then collected at ID23-2 and ID29 at the European Synchrotron Radiation Facility, Grenoble, France. First, a native dataset extending to 2.3 Å was collected at ID23-1, but no suitable model was found to phase the data by molecular replacement (MR). Therefore a seleno-methionine (SeMet) derivative of EarP was crystallized and a full dataset extending to 3 Å was collected at selenium absorption edge (λ = 0.97 Å) at beamline ID29. The space group for both crystals was determined to be I4. The anomalous signal in the SeMet dataset was used to phase the data using single-wavelength anomalous dispersion (SAD). The SAD phasing was done by Crank2 automatic pipeline[67]. The structure factor amplitudes were estimated by Afro (N. S. Pannu, unpublished),

Table 1.1 The crystallization screens tested for EarP.

		Concentrations tested [mg/ml]	
Screen	Supplier	4°C	20°C
Classics	Qiagen	8	8
JCSG+	Molecular Dimensions	8	8
PACT	Molecular Dimensions	8	8
PEGS	Qiagen	8	8
Wizard I+II	Rigaku	8	8

The crystallization screen was done with 8 mg/ml EarP in 50 mM Tris, 100 mM NaCl, 1 mM DTT, 10 mM TDP-Rhamnose, pH 7.6 buffer.

the substructure detection was done by Crunch 2[68] and the density modification was done in Solomon[69]. First, the structure was built using the anomalous dataset by Phenix Autobuild[70] and manual building in Coot[71]. The model was then used for MR of the native high resolution dataset. The published structure was finalized by additional cycles of manual building in Coot and refining in Refmac[72] and is now available with structure factors in the Protein Data Bank under the PDB accession code 5NV8.

Chapter 2

Suppression of *hunchback* mRNA translation

2.1 Introduction

2.1.1 Translation regulation in *Drosophila* development

Development is a prominent example of biological processes governed by the regulation of gene expression[73]. This is probably the best documented on *Drosophila* development[74]. During oogenesis the nurse cells load the developing oocyte with maternally transcribed mRNAs of the maternal effect genes. The mRNAs are produced early but only to be used later during embryogenesis to help the embryo in the synthetically taxing early development.

The maternal effect genes carry the origins of the dorsal-ventral (D-V) and anterior-posterior (A-P) polarity. To define the anterior-posterior axis in the embryo the maternal effect genes will produce protein gradients across the embryo. These protein gradients will then regulate the expression of gap, pair rule and segment polarity genes, which cause the segmentation of the embryo. This illustrates the instrumental need for temporal and spatial regulation of translation in development. However, it does not explain the molecular mechanism behind the temporal and spatial regulation of translation.

2.1.2 The mechanism of the anterior-posterior axis formation

The molecular origins of the A-P polarization can be traced back to maternal activity during oogenesis. However, the A-P axis becomes morphologically apparent only in late embryogenesis. The A-P axis formation starts with the secretion of mRNAs of maternal effect genes into the oocytes by nurse cells. The maternal effect genes induce polarization of the embryo[75,76]. *bicoid* is transcribed by the nurse cells and supplied to the oocyte. The *bicoid* mRNA is then tethered to microtubules at the anterior pole of the oocyte[77–80]. The localization of *bicoid* mRNA is necessary for anterior development[81–83]. The nurse cells also transcribe and transport another maternal effect gene to the oocyte - *nanos*. The *nanos* mRNA

is then concentrated to the posterior pole[84], where it needs to be localized for posterior development[85,86]. Two other maternal effect genes *hunchback* and *caudal* are transcribed by the nurse cells and secreted to the oocyte, but their mRNAs are distributed uniformly throughout the oocyte[87,88]. The mature oocyte then has anterior localized *bicoid* mRNA, posterior localized *nanos* mRNA and uniformly distributed *hunchback* and *caudal* mRNA. These RNAs remain dormant until fertilization.

The oocyte is then fertilized, which triggers synthetic processes. During early embryogenesis the nuclei devide rapidly without cytokinesis forming a syncytial blastoderm. During this phase the maternally supplied mRNAs are translated. The translation of *bicoid* mRNA creates an A-P Bicoid protein gradient[79,89], whereas the translation of *nanos* mRNA results in P-A Nanos protein gradient[84]. However, the gradients are formed by different mechanisms - the Nanos gradient is formed by translation regulation and localization[90–92], whereas the Bicoid gradient forms due to RNA localization[79]. Bicoid then suppresses the translation of maternal *caudal* mRNA at the anterior[93,94] and Nanos suppresses the translation of maternal *hunchback* mRNA at the posterior[95–97]. The embryo then starts to form cell membranes and the syncytial blastoderm develops into cellular blastoderm. When zygotic gene expression starts, Bicoid enhances the zygotic expression of *hunchback*[87,98,99]. As a result, a P-A gradient of Caudal protein[100,101] and an A-P gradient of Hunchback protein forms[98]. The resulting protein gradients then control a group of zygotic genes named gap genes, which divide the embryo into distinct units. The expression of gap genes together with pair rule and segment polarity genes leads eventually to the segmentation of *Drosophila* embryo[102]. During these processes the embryo begins gastrulation and the A-P axis for the first time manifests morphologically.

In summary, it is first necessary to control spatially and temporally the translation of maternal effect genes (e. g. *bicoid*, *nanos*, *caudal* and *hunchback*) in order to transform the maternal molecular polarization of the oocyte into a morphological zygotic A-P axis in the embryo. However, surprisingly little is known about the molecular mechanisms of the spatial and temporal translation regulation of these RNAs[5,92] despite the fact that the resulting protein gradients have been discovered 20-30 years ago. To fully understand the A-P axis formation in *Drosophila* it is therefore necessary to elucidate molecular details of the translational regulation of these genes.

Generally, spatial and temporal translation regulation of the maternal effect genes is mediated by *cis*-acting elements within the regulated mRNA[96,103], which are targeted by *trans*-acting RNA binding proteins (RBPs)[90,92,104–106]. This leads to either localization, translation suppression, translation enhancement or even a combination of those mechanisms might be necessary[92]. For translation suppression the emerging overarching mechanism seems to be a combination of cap-dependent translation inhibition and deadenylation. The cap-dependent translation inhibition is typically driven by protein-protein interaction of the *trans*-acting RBPs that directly or indirectly recruit cap-binding proteins such as eIF4E or a 4EHP[107–109]. This brings together the 5′ and 3′ end of the RNA in a non-productive

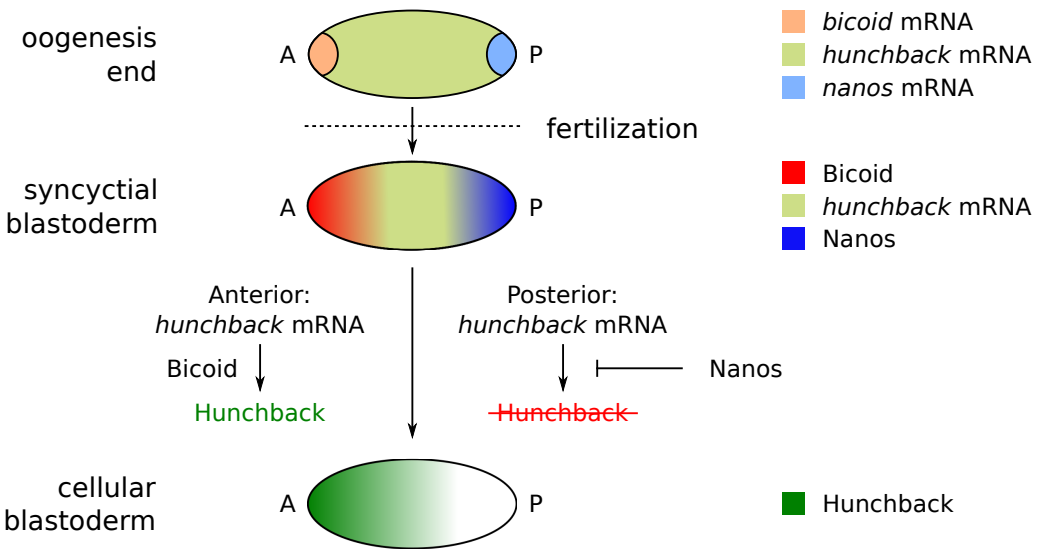


Figure 2.1 The origin of the Hunchback gradient. Mature oocyte harbours several maternally transcribed mRNA which, among others, include *hunchback*, *bicoid* and *nanos* mRNA. The *hunchback* mRNA is distributed uniformly throughout the oocyte and late the embryo, but the *bicoid* mRNA localizes to the anterior and the *nanos* mRNA localizes to the posterior. The *bicoid* and *nanos* mRNAs are after fertilization translated intro steep gradients of Bicoid emanting from anterior and Nanos emanting from posterior. Nanos then suppresses translation of *hunchback* mRNA at posterior, whereas Bicoid stimulates the translation at anterior. As a result the gradient of Hunchback protein forms during embryogenesis.

closed loop conformation preventing the initiation of translation[110,111]. The *trans*-acting RBPs are also often reported to recruit the CCR4 deadenylase[112,113], which removes the polyA tails of the mRNAs priming them for degradation. Specific details of the mechanism of regulation are probably the most well-known for the suppression of *hunchback* mRNA translation.

2.1.3 The suppression of *hunchback* mRNA translation

The Hunchback protein gradient controls the gap genes responsible for abdominal segmentation[114,115]. Therefore, the translation of *hunchback* mRNA is repressed at the posterior by Nanos gradient[95–97]. The responsible *cis*-acting elements were found within the 3′ UTR of the *hunchback* mRNA in two regions of conserved sequence which were termed Nanos Response Elements or NRE[96]. Within each NRE there are two stretches of absolutely conserved sequences which were called BoxA (upstream) and BoxB (downstream). Nanos contains a C-terminal Zinc Finger domain (Nanos ZnF)[116,117]. Zinc Finger domains are known to bind nucleic acids and indeed Nanos ZnF binds RNA with high affinity but low specificity. Additionally, mutations in NREs that disrupt *hunchback* mRNA suppression have *in vitro* no effect on Nanos ZnF binding. Instead other proteins bind the *hunchback* mRNA[118].

One such protein is Pumilio. Pumilio has a C-terimal Homology Domain (Pum HD) which binds the *hunchback* mRNA specifically[6,119–122]. Pum HD is

essential for abdominal segmentation[123], because it recruits Nanos to the *hunchback* mRNA and the recruitment is Pum HD-dependent[105]. The molecular details of the recruitment are explained by the structure of a ternary complex consisting of Pum HD, Nanos ZnF and the mRNA[6]. Pum HD binds its putative binding site overlapping with BoxB, whereas Nanos ZnF is seen interacting with a few bases upstream. However, the structure reveals an additional helix at the very C-terminus of Nanos, which was not previously observed[117]. This additional helix extends out of the ZnF to interact with the C-terminal repeats of Pum which are necessary for the recruitment of Nanos[105]. The remarkable observation regarding the ternary Pum HD-Nanos ZnF-RNA complex is that the interaction of Nanos ZnF is not specific enough to target Nanos to *hunchback* mRNA, while at the same time a stable binary complex Nanos and Pum is not observed. However, all three components form a stable ternary complex, in which not only a functionally essential Nanos-Pum interface is observed, but also sequence requirements are seen for the bases Nanos ZnF is interacts with.

The N-terminal portion of Nanos is divergent and thought to be largely unstructured. However, it was shown to interact with the CCR4-NOT complex[112], which deadenylates the polyA tail of mRNAs priming them for degradation[124]. Pum is able to repress the RNA independently[125], but this ability is irrelevant for *hunchback* mRNA repression since Pum is expressed symmetrically throughout the embryo[126].

A putative model of the suppression of *hunchback* mRNA translation emerges taken all the information together. The *hunchback* mRNA translation is suppressed by the ternary complex of Nanos and Pum bound to the NRE in *hunchback* mRNA 3' UTR. Pum provides specificity to the complex by recognizing the RNA via its HD domain, but carries no information about the position within the embryo. Nanos, on the other hand, shows very little specificity for the RNA, but due to its gradient expression it holds the spatial information. Each of the two proteins can repress an mRNA independently, but they need to come together to create the Hunchback gradient as the repression has to be spatially restricted. This model of *hunchback* mRNA translation suppression proposes translation suppression via the deadenylation pathway by the concerted action of Nanos and Pum. However, it is incomplete.

There is a third protein which participates in the suppression of *hunchback* mRNA translation called Brain tumor (Brat)[106]. Brat has a C-terminal NHL domain via which it interacts with the 3' UTR of the *hunchback* mRNA at a binding site upstream of and partially overlapping with the BoxA of NRE[7,127,128]. Furthermore, Brat interacts with the eIF4E-mimic d4EHP to inhibit translation initiation[108]. However, Brat is, similarly to Pum, uniformly distributed[106]. Thus, it is unlikely that Brat alone creates the gradient of Hunchback. Nevertheless, it remains unclear how exactly the activity of Pumillio, Brat and Nanos is related. Two scenarios are in principle possible: either Brat associates with the Pum-Nanos-RNA ternary complex or Brat forms its own repression complex.

The quaternary complex of Pum-Nanos-Brat and the RNA could benefit from the increased specificity brought by Brat explaining the necessity of Brat for *hunch-*

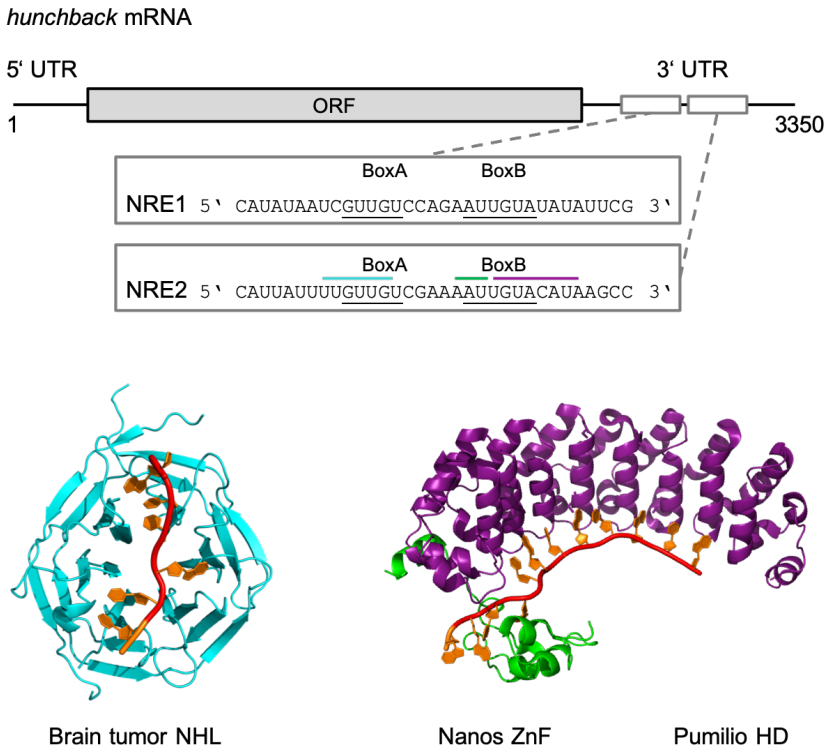


Figure 2.2 *Cis*-acting and *trans*-acting elements in suppression of *hunchback* mRNA translation. The *cis*-acting element in *hunchback* mRNA translation suppression are elements located in the 3′ UTR of the RNA termed Nanos Reponse Element 1 and 2 (NRE1 and NRE2). The full sequence of each NRE is shown the grey box. The underline indicates two conserved sequences in each NRE termed BoxA and BoxB. The *trans*-acting elements include the NHL domain of Brain tumor (in cyan), the HD domain of Pumilio (in purple) and ZnF domain of Nanos (in green). Each domain has its binding site indicated by a line of corresponding colour above the NRE2 sequence.[6,7,95–97,106,119–122,127,128]

back mRNA suppression[106]. Initially, a number of observations were made, suggesting that this could indeed be the case. Yeast four hybrid assays identified several point mutants in Pum HD, Nanos ZnF and Brat NHL that would abrogate formation of the quaternary complex[106]. However, those did not turn out in the light of newly available structural data to be the evidence of the interaction of Brat with the Pum-Nanos-RNA complex[6,7]. The mutations in Pum HD or Nanos ZnF span the Pum-Nanos interface formed by the additional helix of Nanos, whereas the mutation in Brat NHL were later found to be located in the RNA interaction surface. Later, Pum was shown to enhance the affinity of Brat for RNA[128]. This could be interpreted as indirect evidence of the quaternary complex, but the significance of this enhancement remains unclear as the authors themselves observe that Brat and Pum in reporter assays suppress translation independently[128].

Instead Brat could also function to repress *hunchback* mRNA independently. Brat is able to specifically bind *hunchback* mRNA and it has the ability to suppress translation independently[128]. Based on current findings it is impossible to suggest a specific architecture of such a suppression complex, but Brat would likely need to associate with another factor to create the Hunchback gradient, because Brat is

distributed throughout the embryo uniformly[106].

In summary, it is clear that the available data are not sufficient to explain the molecular details of *hunchback* mRNA translation suppression. Further research in two directions seems crucial. First, better understanding of the pathways Brat, Pum and Nanos employ for translational suppression and how these are connected is needed. This is deliberately not discussed in here as it is out of the scope of this work, however others have done interesting research in this direction[129]. Second, further investigation into the events upstream of the repression is crucial. This is signified on the informative power of the structural studies of the Pum HD-Nanos ZnF-RNA and Brat NHL-RNA complexes. The structural information obtained explain previously misunderstood biological data. A similar biophysical approach is now needed to test the existence of the quaternary (or other potential) complexes with Brat and to reveal their architecture and assembly.

2.2 Aims

In order to elucidate how Pum, Brat and Nanos participate in the suppression of *hunchback* mRNA translation a biophysical study aiming to answer the following questions was carried out:

- Do Brat NHL and Pum HD interact *in vitro* directly in the presence or absence of RNA?
- Do Brat NHL and Nanos ZnF interact *in vitro* directly in the presence or absence of RNA?
- Can Brat NHL, Pum HD and Nanos form a quaternary protein-RNA complex?
- What is the architecture of such a quaternary protein-RNA complex and how does it assemble?

2.3 Results

2.3.1 The interplay of Brain Tumor and Pumilio

The Brat NHL domain and the Pum HD domain were both reported to bind *hunchback* mRNA specifically[6,7,119–122,127,128]. However, the interactions were often studied by electrophoretic mobility shift assays (EMSAs), which can be a limited method to report on in solution interactions quantitatively[130]. In case of Pum HD-RNA interactions this in combination with inconsistencies in experimental methods (e. g. using natively purified Pum HD in some studies, but refolded Pum HD in others or using *Homo Sapiens* Pum HD instead of *Drosophila* Pum HD[120,122]) led to inaccuracies in describing the interaction. Therefore, several isothermal titration calorimetry (ITC) experiments were performed in order clarify the understanding of the RNA interactions.

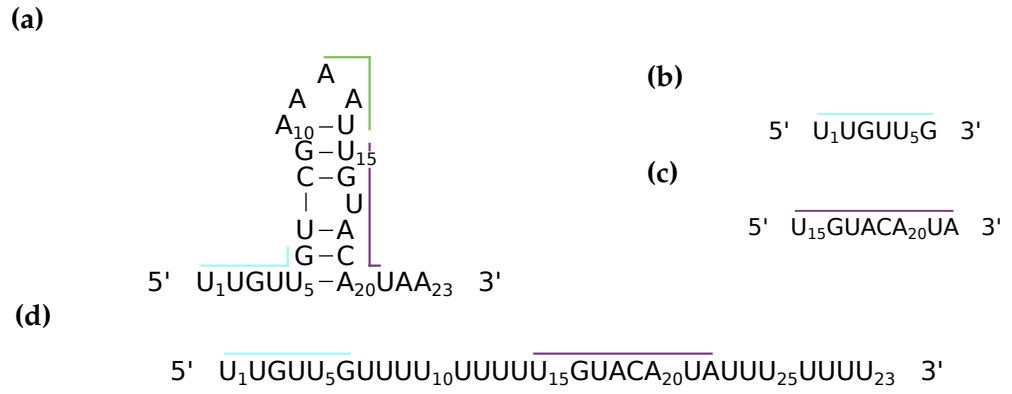
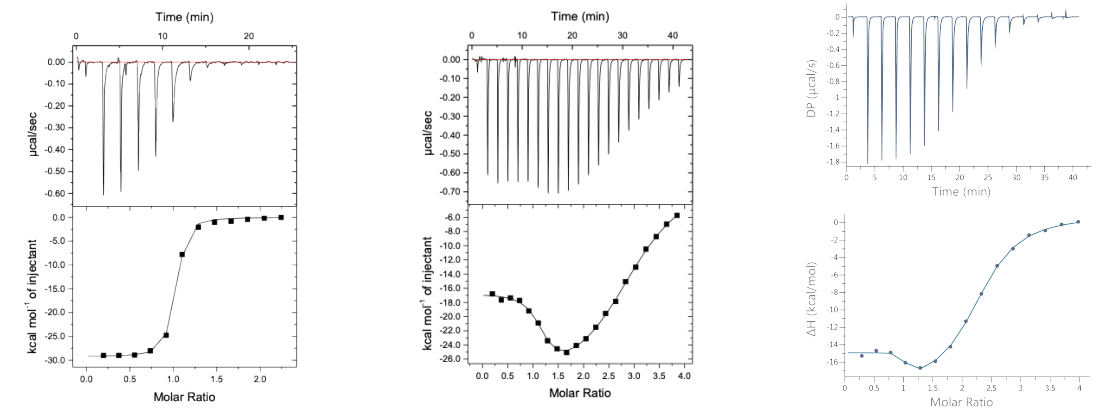


Figure 2.3 Overview of *hunchback* RNAs used. (a) The NRE2 RNA. The secondary structure is indicated based on previous studies[7]. **(b)** The RNA with Brain Tumor (Brat) consensus motif. **(c)** The RNA with Pumilio (Pum) consensus motif. **(d)** The poly-U NRE2 RNA (pNRE2). This RNA is derived from the NRE2 RNA by adding uracils to the 3′ end and replacing the nucleotides connecting Brat and Pum binding sites by uracils to prevent secondary structure formation. Brat, Pum and Nanos binding sites are indicated by cyan, purple and green bars, respectively.

The binding of RNA by Pum HD was measured first (Figure 2.4). The primary site which Pum HD binds within the *hunchback* mRNA is a 8-nucleotide sequence generally termed either Pumilio Response Element (PRE) or Pum consensus (Figure 2.3). The Pum consensus overlaps partially with BoxB, which can be found in both NRE1 and NRE2. The sequence of Pum consensus is UGUAYAUA, where Y is U in NRE1 and C in NRE2. Pum HD binds to Pum consensus with a K_D of 18 $\pm$ 3 nM. The K_D decreases even lower to 2.3 nM when the interaction is measured in the context of the full NRE2 RNA. The NRE2 RNA contains the full sequence of NRE2 including the Brat binding site and connecting nucleotides (Figure 2.3). In the titration with this RNA the high affinity binding is followed by a second lower affinity event with a K_D of roughly 2 μM. To test whether the second binding is to the Brat binding site and whether secondary structure of the RNA influences Pum binding another titration was measured with an RNA termed poly-U NRE2 (pNRE2). The pNRE2 is a derivative of the NRE2 with additional uracils at the 3′

end and with the nucleotides connecting Brat and Pum binding sites replaced by uracils. Since the NRE2 was shown to form a secondary structure[7], the pNRE2 was designed to prevent the RNA from forming any secondary structure. The pNRE2 titration revealed that Pum HD binds the RNA with two binding events as well. The higher affinity event resembles the interaction with Pum consensus with a measured K_D of 26.3 $\pm$ 6.5 nM, whereas the second event resembles the lower affinity binding event in NRE2 with a measured K_D of roughly 2 μM. This result suggests that Pum HD is able to bind any RNA with its consensus with high affinity regardless of its context or the RNA structure and that Pum HD also binds weakly the Brat binding site.

(a) **(b)** **(c)**



N	0.947 $\pm$ 0.0048
K_D	18 $\pm$ 3 nM
ΔH_1	-29.25 $\pm$ 0.245 kcal/mol

N_1	1.10 $\pm$ 0.0188
K_{D1}	26.3 $\pm$ 6.5 nM
ΔH_1	-16.73 $\pm$ 0.197 kcal/mol
N_2	1.880 $\pm$ 0.0348
K_{D2}	1923 $\pm$ 150.9 nM
ΔH_2	-30.82 $\pm$ 0.680 kcal/mol

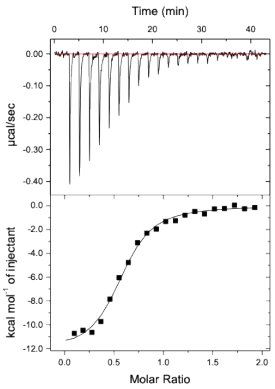
N_1	0.906 $\pm$ 0.0078
K_{D1}	2.3 $\pm$ 0.8 nM
ΔH_1	-15.80 $\pm$ 0.311 kcal/mol
N_2	1.440 $\pm$ 0.0120
K_{D2}	1970 $\pm$ 94.4 nM
ΔH_2	-18.30 $\pm$ 0.720 kcal/mol

Figure 2.4 Isothermal titration calorimentry of Pumilio HD domain (Pum HD). (a) The binding of Pum HD to Pum consensus RNA of the sequence UGUACAUA. **(b)** The binding of Pum HD to poly-U NRE2 RNA (pNRE2). The pNRE2 is derived from the NRE2 RNA by adding uracils to 3′ end and replacing the nucleotides connecting Brat and Pum biding sites by uracils. **(c)** The binding of Pum HD to NRE2 RNA.

Next, the interaction of Brat NHL with RNA was investigated by ITC (Figure 2.5). Brat NHL binds the sequence consisting of UUGUUG present within BoxA of each NRE termed Brat consensus. Isolated Brat consensus RNA interacts with Brat NHL with a K_D of approximately 1.4 μM. The affinity of Brat NHL for the NRE2 RNA is higher with a K_D of approximately 1.2 μM. The μM K_D is not comparable to the affinity of Pum HD. However, when Brat NHL was measured with pNRE2 the affinity increased to 17.5 $\pm$ 3.1 nM. This could mean that Brat NHL is sensitive to RNA secondary structure as it was proposed previously[7]. Interestingly, Brat NHL binds the NRE2 RNA in a single binding event with a 1:1 stoichiometric ratio, but the titration with pNRE2 shows two binding events with an order of magnitude difference in K_D (Figure 2.5b versus Figure 2.5c).

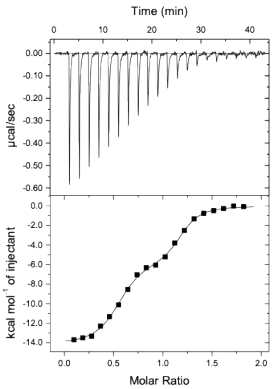
Despite the fact that Pum HD also weakly binds the Brat binding site, it is

(a)



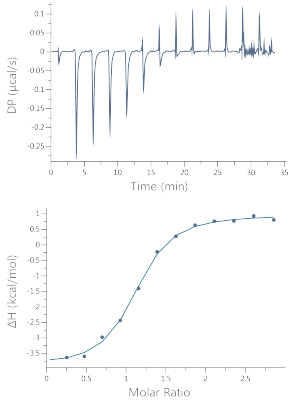
N	0.574 ± 0.0113
K_D	1422 ± 192.6 nM
ΔH_1	-12.24 ± 0.321 kcal/mol

(b)



N_1	0.515 ± 0.0118
K_{D1}	17.5 ± 3.1 nM
ΔH_1	-14.26 ± 0.214 kcal/mol
N_2	0.633 ± 0.0103
K_{D2}	407 ± 83.9 nM
ΔH_2	-5.79 ± 0.379 kcal/mol

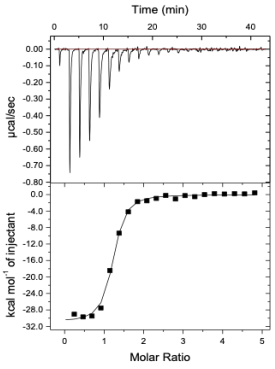
(c)



N	1.06 ± 0.0170
K_D	1190 ± 198 nM
ΔH	-4.90 ± 0.142 kcal/mol

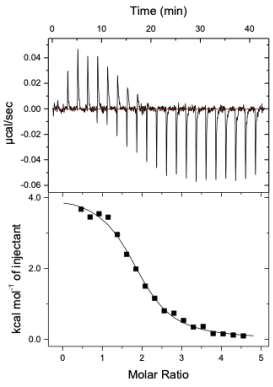
Figure 2.5 Isothermal titration calorimentry of Brain Tumor NHL domain (Brat NHL). (a) The binding of Brat NHL to Brat consensus RNA of the sequece UUGUUG. **(b)** The binding of Brat NHL to poly-U NRE2 RNA (pNRE2). The pNRE2 is derived from the NRE2 RNA by adding uracils to 3′ end and replacing the nucleotides connecting Brat and Pum biding sites by uracils. **(c)** The binding of Brat NHL to NRE2 RNA.

(a)



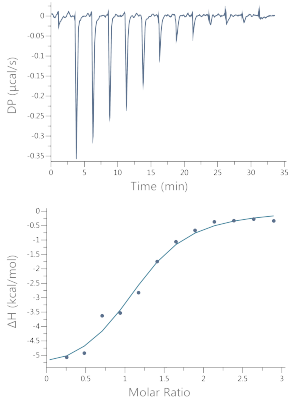
N	1.14 ± 0.0118
K_D	165 ± 24.5 nM
ΔH_1	-30.94 ± 0.451 kcal/mol

(b)



N	1.87 ± 0.0323
K_D	990 ± 131.3 nM
ΔH	-4.10 ± 0.105 kcal/mol

(c)



N	1.12 ± 0.0670
K_D	2910 ± 1370 nM
ΔH	-5.83 ± 0.713 kcal/mol

Figure 2.6 Isothermal titration calorimetry of combined Brain Tumor (Brat) and Pumilio (Pum) binding. (a) The binding of the Pum HD domain to the poly-U NRE2 RNA (pNRE2) pre-incubated with Brat NHL. **(b)** The binding of Brat NHL to pNRE2 pre-incubated with Pum HD. **(c)** The binding of Brat NHL to NRE2 RNA pre-incubated with Pum HD.

unlikely that Pum HD binding would be inhibiting Brat NHL binding as the second binding event of Pum has higher K_D than the first binding event of Brat NHL. In fact Pum HD was shown to mildly increase the affinity of Brat NHL

for RNA in EMSA[7]. To test the cooperativity quantitatively another set of ITC measurements was performed (Figure 2.6). In the first two experiments, Pum HD was pre-incubated with pNRE2 RNA, while Brat NHL was titrated and *vice versa*. In the third experiment the NRE2 RNA was pre-incubated with Pum HD and Brat NHL was titrated. These experiments no longer show high and low affinity binding events. Furthermore, in pre-incubation experiments both Brat NHL and Pum HD show significantly increased K_D in comparison to the analogous experiments with free RNA (e. g. Figure 2.5b versus Figure 2.6b or Figure 2.5c versus Figure 2.6c). Since the mild cooperativity of Brat NHL and Pum HD in RNA binding was previously observed in EMSA[7] analogous pre-incubation experiments were measured orthogonally in EMSA (Supplementary Figure 2). These experiments did not show any clear evidence of cooperativity between Brat NHL and Pum HD binding either. Furthermore, ITC measurements when Pum HD was directly titrated by Brat NHL did not reveal any binding (data not shown).

2.3.2 Reconstitution of the *hunchback* mRNA suppression complex

The lack of cooperativity in RNA binding by Brat NHL and Pum HD does not point to a stable ternary complex of Brat NHL and Pum HD. Such ternary complex of two proteins bound to the *hunchback* mRNA was however reported for Pum HD and Nanos ZnF[6]. Whether the ternary complex recruits Brat NHL remains unclear, as no direct observation of the quaternary complex of Brat NHL, Pum HD and Nanos ZnF bound to the *hunchback* mRNA (Hunchback complex) has been reported yet. To test the existence of Hunchback complex purified Brat NHL, Pum HD and Nanos ZnF were mixed together with a 23-nucleotide NRE2 RNA in an attempt to reconstitute the complex *in vitro*. The mixture then underwent size-exclusion chromatography (SEC) where a broad peak with a maximum at elution volume corresponding to a molecular weight of 77 kDa was observed (Figure 2.7). The expected molecular weight of the complex is approximately 88 kDa, so the identity of the peak was further validated by SDS-PAGE. The gel clearly shows bands corresponding to Pum HD (MW 38.4 kDa), Brat NHL (MW 31.9 kDa) and Nanos ZnF (10.6 kDa). The NRE2 RNA has a molecular weight of only 7 kDa, but the UV trace of the SEC run readily showed strong absorption at 260 nm. Further validation came from a measurement of the UV absorption spectra of the SEC peak which showed a prominent absorption maximum at 260 nm confirming the presence of RNA.

Hunchback complex was rather stable at 4°C, and a repeated SEC run after a week showed a single peak at the same elution volume. Thus, crystallization trails were done in 14 different commercially available crystallization screens (Table 2.2). The trials spanned a concentration range from 1.5 to 10.0 mg/ml, tested every contemporary crystallization chemistry (e. g. PEG precipitants, salts, pH, PEG smears, precipitant mixtures or alternative polymer precipitants) and were done at both 4°C and 20°C. Unfortunately, none of the tested conditions produced crystals, so an alternative crystallization approach was tested using carrier-driven

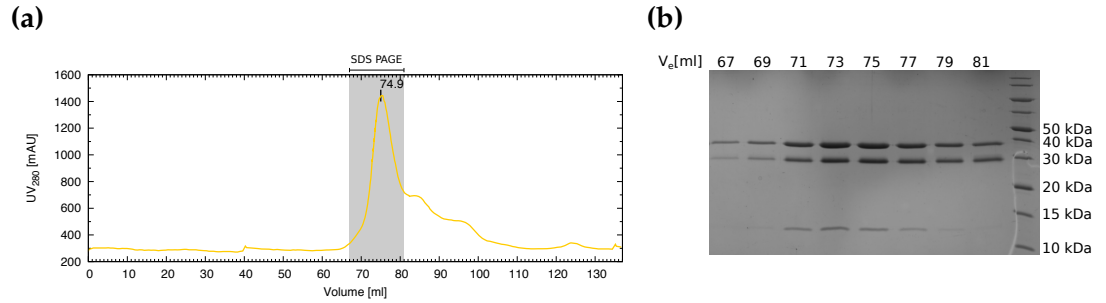


Figure 2.7 The Hunchback complex formation. Hunchback complex was reconstituted *in vitro* by mixing purified Brain Tumor NHL domain, Pumilio HD domain and Nanos Zinc Finger domain with the NRE2 RNA and then performing size-exclusion chromatography. **(a)** A chromatogram of a size-exclusion chromatography run of the reconstituted Hunchback complex on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare). **(b)** SDS-PAGE analysis of the main peak from size-exclusion chromatrography fractions.

crystallization[131]. To this end, Pum HD was fused to MBP using a GSGSGS linker. However, the MBP fusion impairs the formation of Hunchback complex (data not shown).

2.3.3 The interplay of Brain Tumor and Nanos

The attempts to directly crystallize Hunchback complex failed and so alternative strategies to investigate the complex were necessary. As the interaction of Pum and Nanos was described[6] and the ITC experiments do not show any coperativity between Brat NHL and Pum HD (subsection 2.3.1), the next obvious binary relationship to investigate was between Brat NHL and Nanos ZnF. Nanos ZnF was previously cloned, purified and assigned, and due to its rather small size provides a well resolved ^{15}N HSQC spectrum[132]. Such a spectrum is particularly convenient for NMR titrations as any changes in it should be very well visible. Brat NHL is roughly three times the size of Nanos ZnF so a potential interaction with Nanos ZnF should not only manifest in chemical shift perturbations but also in a severe decrease of signal intensity resulting from a change of the tumbling rate of Nanos ZnF upon interaction with Brat NHL.

The samples of natively purified Nanos have been shown to be heterogeneous - running as two bands on SDS-PAGE[105,133]. One of the species is still recruited and crystallized in the ternary complex with Pum HD and RNA, but this heterogeneity is nevertheless visible in NMR as doubling of certain peaks suggesting two alternative conformations in solution[132]. However, upon the addition of zinc the alternative conformation disappears and the prevailing major conformation is likely the active species as it was shown to bind RNA[132]. To study the Brat-Nanos interaction the purified ^{15}N Nanos ZnF was supplemented with zinc and gradually titrated by Brat NHL (Figure 2.8). The ^{15}N HSQC spectrum used to monitor titration confirms the homogeneity of Nanos at the beginning of the titration. Interestingly, the overlay of spectra from the titration reveals chemical shift perturbations of Nanos ZnF upon Brat NHL addition indicating an interaction.

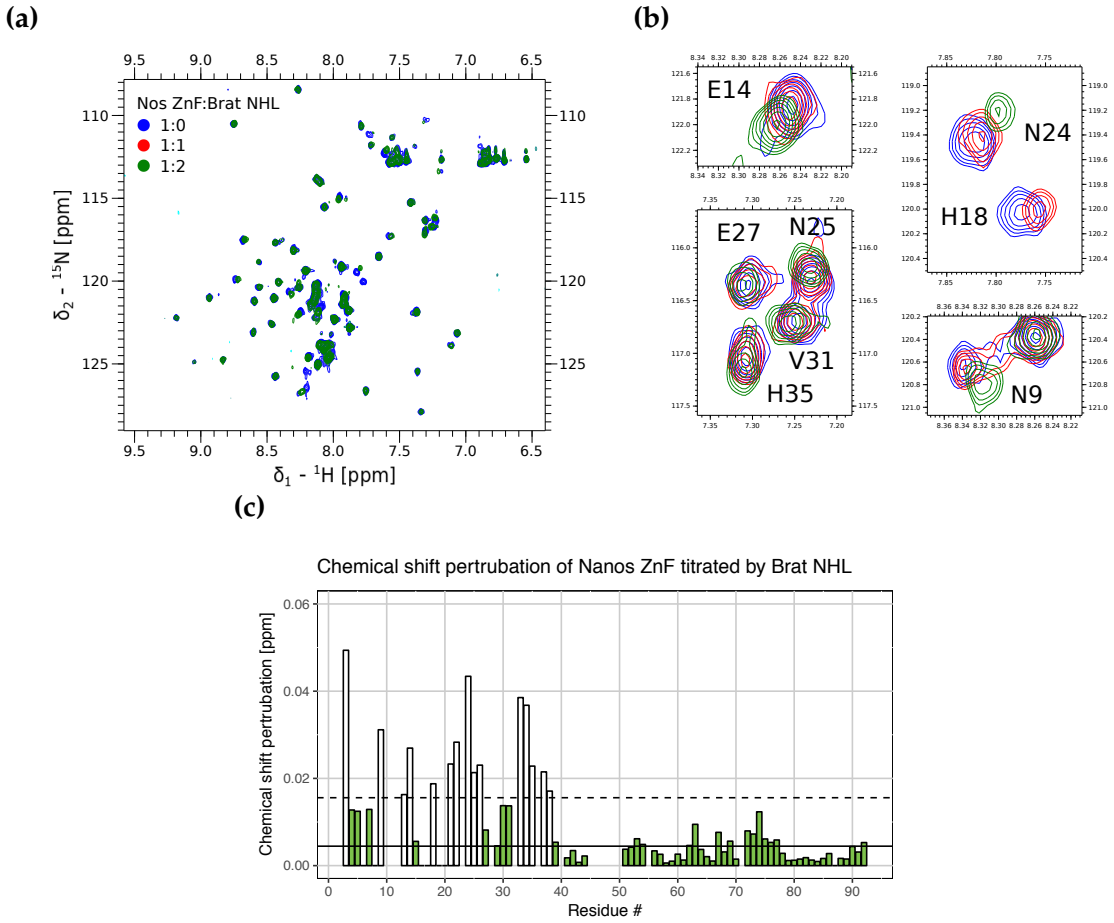


Figure 2.8 The NMR titration of Nanos ZnF by Brat NHL. (a) Overlay of the ^{15}N NMR spectra monitoring the Nanos ZnF-Brat NHL titration. ^{15}N Nanos ZnF was titrated by unlabelled Brat NHL to 1:2 Nanos:Brat ratio. **(b)** Zoom-in of the peaks that shift the most throughout the titration, including peak assignment labels. **(c)** The chemical shift perturbation (CSP) plot derived from the titration experiment. The plot shows the magnitude of changes of Nanos backbone amide chemical shifts for each residue calculated according to Equation 1.1. The full line indicates the mean, the dashed line indicates the mean plus standard deviation. Residues which show higher CSP than the mean plus standard deviation are colored in white.

The chemical shift perturbations map to a continuous surface in the N-terminal half of Nanos ZnF. However, the perturbations are not particularly large and the peaks move in the fast exchange regime, which is typical for weaker interactions.

It would be valuable to describe the interaction more and in an orthogonal assay, such as ITC. However, that requires homogenous Nanos sample. This can be achieved by supplementing zinc in the buffer. However, Brat NHL tolerates zinc only for the short course of the NMR titration, while ITC requires to dialyse both proteins in the same buffer to avoid buffer mismatch and an overnight dialysis to buffer with zinc precipitated Brat NHL. Alternatively, refolding Nanos was tested to obtain a homogeneous active sample of Nanos ZnF which would not require explicit zinc in a buffer. Unfortunately, even a previously published refolding protocol for Nanos ZnF[116] only yielded partially folded Nanos (data not shown).

2.3.4 Characterization of the *hunchback* mRNA suppression complex

The quaternary Hunchback complex comprising Pum HD, Brat NHL, Nanos ZnF and the NRE2 RNA was further studied by small-angle x-ray and neutron scattering (SAXS and SANS). SAXS and SANS both provide low resolution information about the shape of a particle and its size. However, neutron scattering depends mainly on the composition of the nuclei and changes drastically for different isotopes of the same element such as hydrogen and deuterium. This is often exploited to study structures of biomacromolecules using contrast variation. In the solution of a protein complex both the protein and the buffer scatter. Varying the content of D_2O in the buffer and using ^2H-labelled proteins then provides a chance to design sample conditions in which certain parts of the complex are rendered invisible. The scattering measured by contrast variation then provides further insights into the internal structure of the complex of interest.

For Hunchback complex several samples were measured (fully listed in Table 2.3). SAXS was measured on a completely protonated sample at 0% D_2O. SANS was measured on the same sample at 0% and 66% D_2O. Further SANS was measured on samples of the complex with all possible combinations of two ^2H-labelled proteins in complex with ^1H RNA and the remaining ^1H protein. Each of the complexes was generally measured in 0% D_2O buffer, where all components are visible, in approximately 42% D_2O, where the ^1H protein is invisible, and in approximately 65% D_2O where the RNA is invisible and ^1H protein exhibits weak negative contrast.

Thirteen curves in total were obtained and are fully shown in Supplementary Figure 3. First, Guinier analysis was done in order to evaluate the quality of the data (summary in Table 2.1, fits in Supplementary Figure 4). The majority of the curves consistently agrees on the same basic parameter of Hunchback complex - the radius of gyration (R_g). This R_g can be derived from eight of the thirteen curves. The remaining five curves show distinctly lower R_g. These include all samples of the complex with ^2H Brat and ^2H Nanos (labelled 2B1P2N) plus the sample of the complex with ^2H Pum and ^2H Nanos (labelled 1B2P2N) at 68% D_2O. The decrease of R_g in these curves suggests that the integrity of the complex was compromised. Indeed, when the curves of the four different complexes measured at 0% D_2O are compared, the 2B1P2N sample shows clearly different trend in comparison with the other samples (Supplementary Figure 3c). In fact, the curve of the 2B1P2N complex at 0% D_2O does not fit any of the roughly 5000 models of Hunchback complex generated (described later). However, it fits very well a mixture of the Pum HD-Nanos ZnF-RNA and Brat NHL-RNA complexes (Supplementary Figure 3d). The 1B2P2N complex at 0% D_2O fits the models of Hunchback complex rather well and does not fit the mixture (Supplementary Figure 3e). However, the curve of this complex at 68% D_2O consistently across the whole range of Hunchback complex models shows higher intensities at low q values, which is a common sign of aggregation (Supplementary Figure 3f). It is unlikely that these curves provide data about Hunchback complex of high enough quality, so only the eight curves

Table 2.1 The basic parameters derived from the small-angle scattering data.

Sample	Guinier analysis		P(r)			predicted I_0
	R_g [Å]	I_0	R_g [Å]	I_0	D_{max} [Å]	
SAXS	39.6	66.8	39.5	64.8	128	
1B1P1N 0% D_2O	38.9	0.150	42.7	0.089	125	0.239
1B1P1N 66% D_2O	39.6	0.059	41.8	0.035	129	0.056
1B2P2N 0% D_2O	38.2	1.17	36.9	0.928	129	0.938
1B2P2N 41% D_2O	36.1	0.306	36.6	0.276	127.5	0.271
1B2P2N 68% D_2O	33.5	0.098	*n.d.*	*n.d.*	*n.d.*	0.050
2B1P2N 0% D_2O	23.6	0.221	*n.d.*	*n.d.*	*n.d.*	0.808
2B1P2N 44% D_2O	24.2	0.072	*n.d.*	*n.d.*	*n.d.*	0.176
2B1P2N 59% D_2O	18.8	0.045	*n.d.*	*n.d.*	*n.d.*	0.065
2B1P2N 82% D_2O	25.6	0.044	*n.d.*	*n.d.*	*n.d.*	0.000
2B2P1N 0% D_2O	40.2	1.44	39.6	1.33	128	1.406
2B2P1N 33% D_2O	40.0	0.633	39.68	0.557	132	0.671
2B2P1N 62% D_2O	39.8	0.259	41.81	0.208	128	0.245

R_g refers to radius of gyration, I_0 to forward scattering, D_{max} to maximum distance and P(r) to distance distribution function. The parameters grouped under Guinier analysis were calculated by Guinier approximation, whereas the parameters grouped under P(r) were calculated from the distance distribution function. The predicted I_0 was calculated using SASSIE-web (https://sassie-web.chem.utk.edu/sassie2/).

showing consistent data were analyzed further.

SAS reveals the following parameters of Hunchback complex. The R_g of Hunchback complex is around 40 Å. The distance distribution function shows a maximum distance of approximately 125-130 Å with a peak of distances at approximately 30 Å (Supplementary Figure 5). The distance distribution function generally shows a very asymmetric peak gradually tailing of in high distances often with additional local maxima in the tail, which suggests rather an elongated particle than a sphere[134].

The SAS data were then used in conjunction with the available high resolution X-ray structures of Brat NHL-RNA and Pum HD-Nanos ZnF-RNA complexes to model Hunchback complex. The X-ray structures were placed on single NRE2 RNA. The modelling was then done by preserving those parts that contain information from the X-ray structures and sampling the available conformational space. In practice the connecting part of the RNA not bound by any of the proteins was first randomized and then simulated annealing was performed. This generated 5055 models of possible structures of Hunchback complex. Then, scattering curves were back-calculated at corresponding conditions for each of the model. The back calculated curves were then fitted against the experimental curves obtaining a distribution of fits for each experimentally measured curve (Figure 2.9). The quality of the fit was quantified in terms of χ^2. Models that are in the top 0.22 quantile of the χ^2 distribution of each curve were selected to find the overall best fitting

model of Hunchback complex (the points in two shades of blue in Figure 2.9).

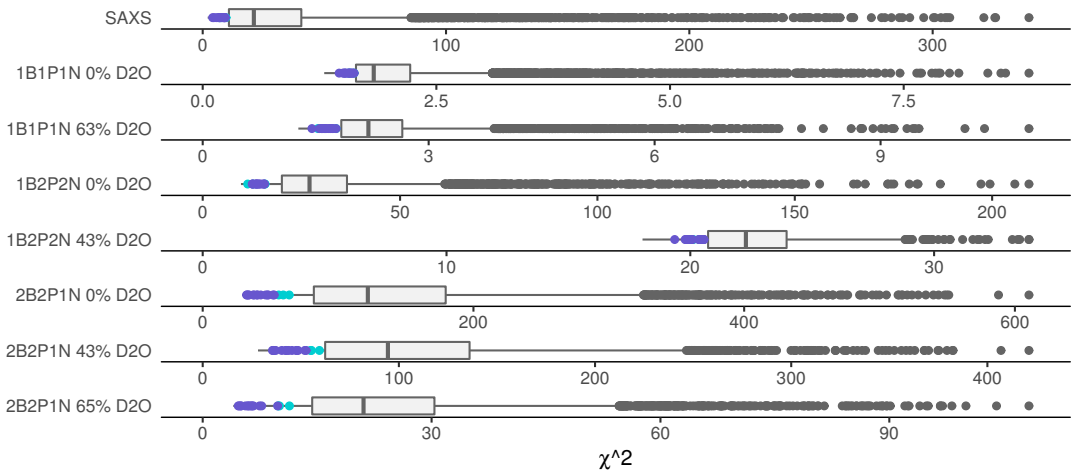


Figure 2.9 The fits of Hunchback complex models to the small angle scattering (SAS) data. A set of 5055 models of Hunchback complex was calculated using the X-ray structures of the components and scattering curves of each model at corresponding condition were back-calculated and fit against the SAS data. Each line shows a distribution of χ^2 values from all models for one experimentally measured curve. The coloured points show the models found in the ensembles of overall best fitting models (models in the top 0.22 quantile for each curve). The purple-blue points indicate models in the major ensemble of 14 models, the cyan points indicate the 3 models in the minor ensemble. The box indicates 0.25 to 0.75 interquantile range, the line in the box indicates the mean and the grey points indicate the outliers.

That yielded 17 models models that reveal two well defined conformations of Hunchback complex (Figure 2.10a). The major conformation ensemble comprises 14 models in which Brat NHL is located in the vicinity of the additional C-terminal helix of Nanos ZnF. The precision of the major ensemble is 8.4 Å RMSD to the mean model of the ensemble. The minor conformation ensemble comprises only 3 models in which Brat NHL localizes to the C-terminus of Pum HD on the opposite side of Pum HD than Nanos ZnF. The precision of the minor ensemble is 5.5 Å RMSD to mean model of the ensemble. The imprecision of either of the ensembles simply reflects the limits of SAS. The difference between the models within the ensembles comes from rotation of Brat NHL about its own axis. Brat NHL is a six-bladed β-propeller that has a six-fold rotational symmetry about its axis. As SAS provides rotationally averaged scattering, the information about the absolute rotation of Brat NHL is lost.

The model closest to the mean coordinates in each ensemble was selected as a representative and its fits to the experimental curves were examined to compare the ensembles (Figure 2.10b and Figure 2.10c). However, no obvious distinction is visible between the major and minor ensemble suggesting that both conformations of Hunchback complex are equally possible. To resolve the ambiguity cross-linking/mass spectrometry (XL/MS) data were collected.

Three sets of XL/MS data were acquired (Supplementary Figure 7). Two sets

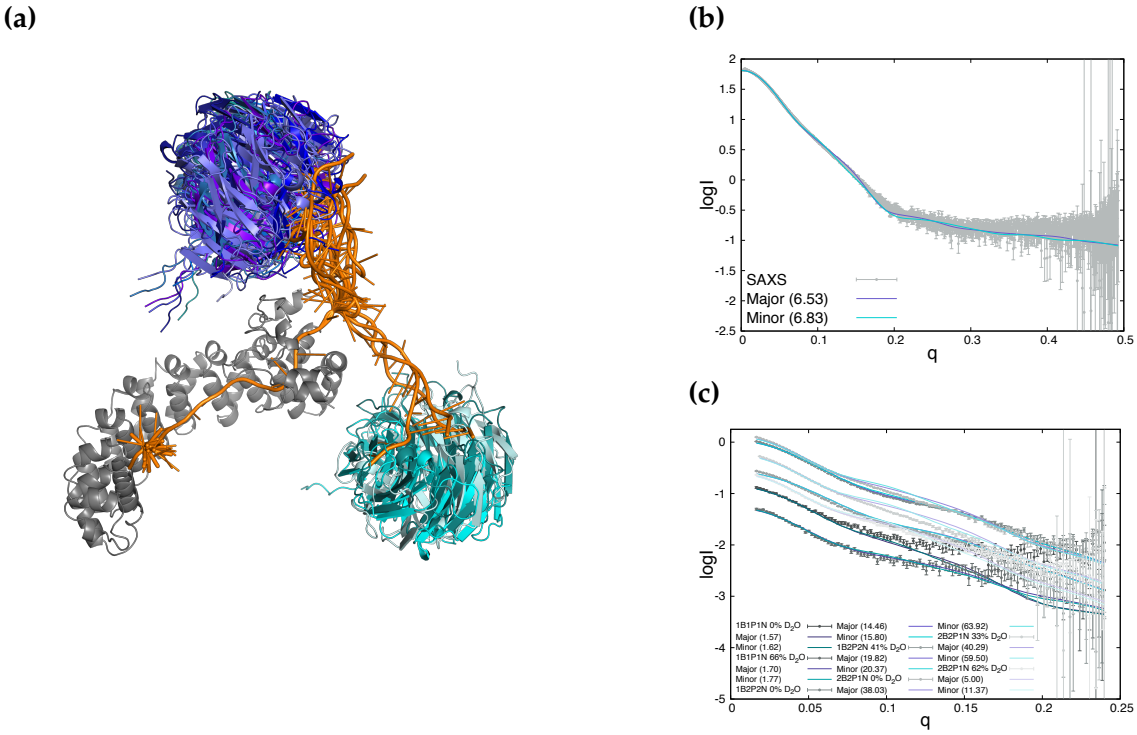


Figure 2.10 The best models of Hunchback complex based on modelling and small angle scattering (SAS) data. **(a)** The top 17 models of Hunchback complex based on their overall fit to the SAS data. The criteria to select the models was a χ^2 of their fit in top 0.22 quantile for each experimental SAS curve. The models fall into two ensembles based on their conformation - the major ensemble comprising 14 models here depicted with Brat NHL domain coloured in shades of purple-blue and the minor ensemble comprising 3 models here depicted with the Brat NHL domain coloured in shades of cyan. The models are always aligned on Pum HD. **(b)** and **(c)** The fits of the representative model of each ensemble to the experimental SAS curves. The representative model was selected as the model closest to the mean structure of the ensemble. The parentheses state the χ^2 of each fit.

come from 1.5 mg/ml Hunchback complex cross-linked with 0.2 mM and 1 mM DSS, respectively. The third was obtained from 0.95 mg/ml Hunchback complex cross-linked by 0.5 mM DSS. The 0.2 mM DSS with 1.5 mg/ml complex yielded 7 unique inter-molecular cross-links. The 1 mM DSS with 1.5 mg/ml complex yielded 74 unique inter-molecular cross-links, while the 0.5 mM DSS with 0.95 mg/ml complex yielded 20 unique inter-molecular cross-links. The 1.5 mg/ml concentration of the complex is likely to be suboptimal for cross-linking. Both datasets acquired at the concentration show significant proportion of cross-links that are obviously not possible within a single monomeric Huchback complex. This includes cross-links when a residue is reported cross-linked to itself (two identical peptides cross-linked), cross-links of residues within a domain that are too distant to be cross-linked in a single molecule or cross-links linking the same site of Brat NHL to the opposing poles of Pum HD. These cross-links clearly do not describe a single monomer of Hunchback complex. At 0.95 mg/ml complex this phenomena is not obvious, so the 0.5 mM DSS dataset was analyzed in more detail (Figure 2.11,subsection 2.5.8).

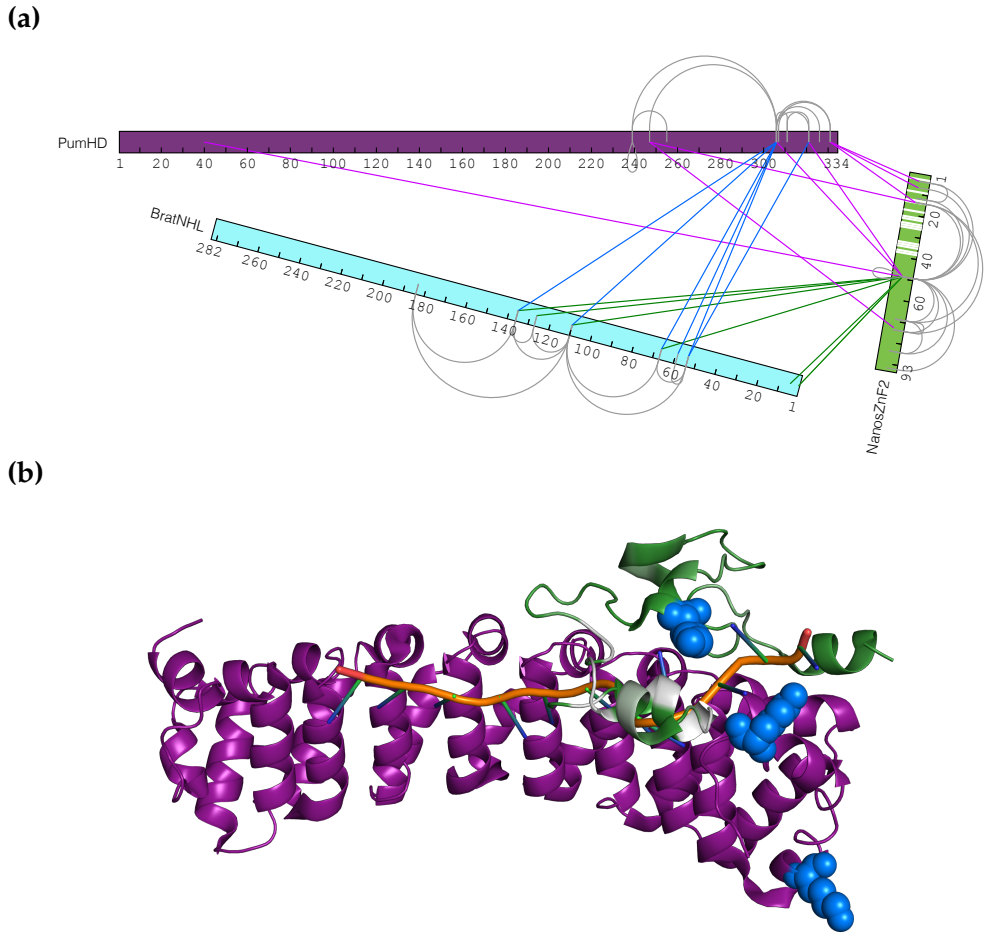


Figure 2.11 The cross-links of Hunchback complex. **(a)** The cross-links detected by cross-linking 0.95 mg/ml Hunchback complex by 0.5 mM DSS. The coloured lines indicate inter-molecular cross-links, while the grey lines indicate intra-molecular cross-links. Only cross-links with xQuest score greater than 25 are shown. The white gaps in Nanos ZnF highlight the residues showing significant chemical shift perturbations (CSPs) in Nanos ZnF-Brat NHL NMR titration shown in Figure 2.8. **(b)** The structure of the Pum HD-Nanos ZnF-RNA complex with highlighted CSPs from Nanos ZnF-Brat NHL NMR titration and the cross-linked residues. The CSP from Figure 2.8 are plotted here in green-white gradient with green representing the least and white representing the most perturbed residues in the titration. The blue spheres represent the Pum HD and Nanos ZnF residues which Brat NHL cross-links to.

The dataset shows Brat NHL to cross-link to both Nanos ZnF and Pum HD. Interestingly, Brat NHL cross-links to the C-terminus of Pum HD, which is also the site of Nanos ZnF binding. Brat NHL is further cross-linked to K49 of Nanos ZnF, which is in vicinity of the Nanos residues showing the largest chemical shift perturbations upon titrating Brat NHL (Figure 2.8). Nanos K49 aligns with the residues of Pum HD, which Brat NHL cross-links to in a single continuous surface (Figure 2.11b). The cross-links of Brat NHL both to Nanos ZnF and Pum HD emanate from the same residues of Brat NHL. Collectively, the XL/MS could suggest that in Hunchback complex Brat NHL, Pum HD and Nanos ZnF form a tripartite interaction surface located at the C-terminus of Pum HD.

The 0.5 mM DSS dataset was then used to repeat the modelling described above. Again, the X-ray structures of the individual components of the complex were used in modelling combining RNA randomization and simulated annealing to generate possible structures of Hunchback complex. However, the XL/MS data were used as additional distance restraints resulting in 4572 models of the complex. Subsequently, theoretical scattering curves were back-calculated for each model at each condition and fitted against the experimental curves (Figure 2.12). The set of models was then reduced to 55 models which are in the top 0.3 quantile of χ^2 distributions for each of the eight experimental SAS curves. From the 55 models top 10% models were selected according to distance restraint energy giving rise to a final ensemble of 6 models (Figure 2.13a).

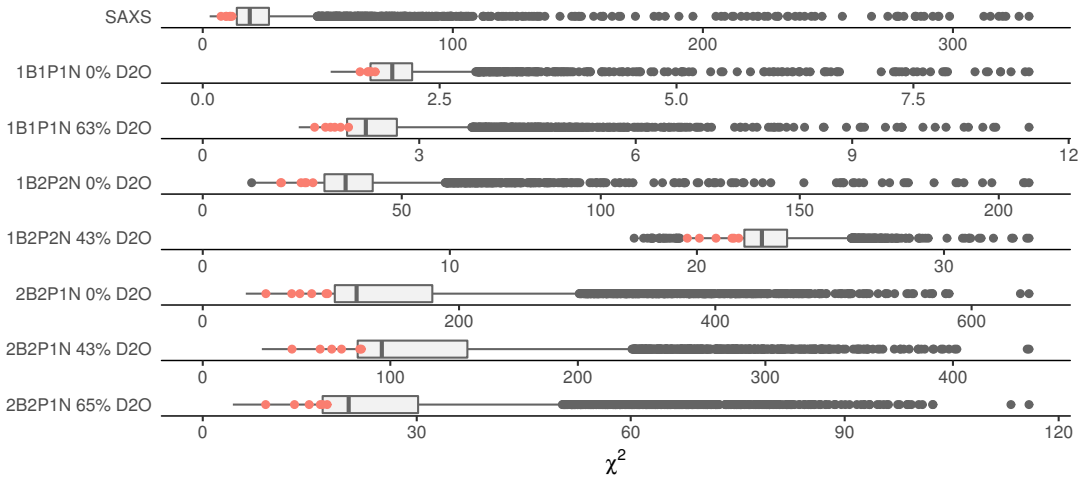


Figure 2.12 The fits of Hunchback complex models generated using the cross-linking/mass-spectrometry (XL/MS) data against the small angle scattering (SAS) curves. A set of 4572 models of Hunchback complex was calculated using the X-ray structures of the components and XL/MS data. Scattering curves of each model at corresponding condition were back-calculated and fit against experimental data. Each line shows a distribution of χ^2 values from all models for one experimentally measured curve. The pink points show the models found in the ensemble of overall best models (the top 10% models with the lowest distance restraint energy of the models in the top 0.3 quantile for each curve). The box indicates 0.25 to 0.75 interquantile range, the line in the box indicates the mean and the grey points indicate the outliers.

The ensemble modelled based on the XL/MS data shows Brat NHL localized to the C-terminus of Pum HD (Figure 2.13a). Brat NHL is positioned similarly to the previously described minor ensemble (Figure 2.13b). However, the NHL domain is oriented differently, in comparison to the minor ensemble it is rotated by roughly 70° about its diameter. The precision of the ensemble is 8.3 Å RMSD to the mean coordinates of the ensemble. The imprecision arises again from uncertain rotational position of Brat NHL, which was not resolved by the addition of the XL/MS data to modelling. The cross-links were then inspected in detail on the structure with the lowest distance restraint energy to see the information they bring to the modelling (Figure 2.14).

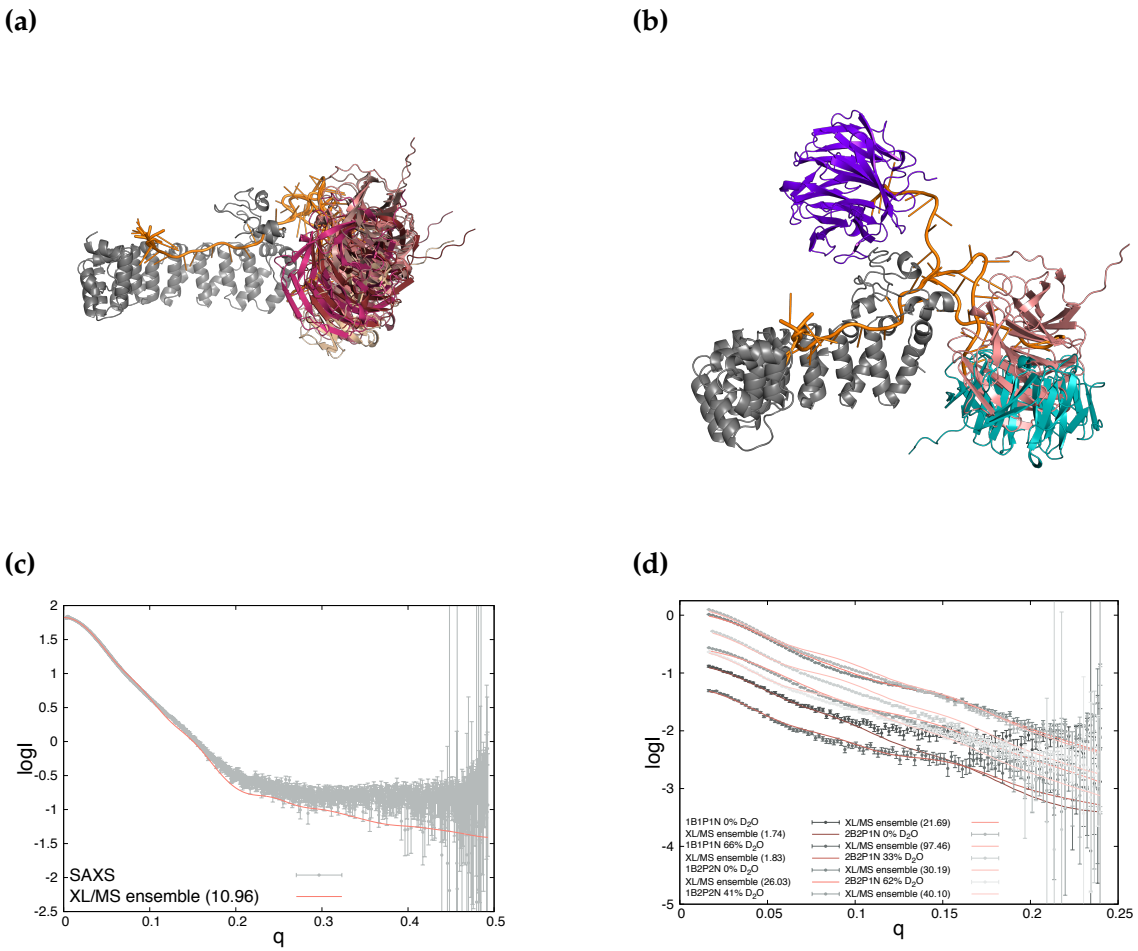


Figure 2.13 The best models of Hunchback complex modelled using the cross-linking/mass-spectrometry and small angle scattering (SAS) data. (a) The top 6 models of Hunchback complex based on their distance restraint energy and overall fit to the SAS data. The ensemble was obtained by finding the models with fits in the top 0.3 quantile for each experimental curve and selecting the models with top 10% distance restraint energy. The models are always aligned on Pum HD. **(b)** The alignment of representative models of ensembles from modelling based only on SAS and on XL/MS and SAS. The representative models are the models from the ensemble closest to the mean structure (modelling based only on SAS) or the model with the lowest distance restraint energy (modelling based on XL/MS and SAS). The colour code is the same as in Figure 2.10. **(c)** and **(d)** The fits of the representative model from modelling based on XL/MS and SAS to the experimental scattering curves. The parentheses state the χ^2 of each fit.

In total 12 crosslinks between Brat NHL and Pum HD or Nanos ZnF were used. However, only 5 of these are satisfied in the representative model of the ensemble (Figure 2.14a). The crosslinks were also inspected in the context of the model with the overall lowest distance restraint energy to ensure it is not the SAS fitting that filters out conformations which satisfy the XL/MS data better. Interestingly, in this model only 6 out of 12 cross-links between Brat NHL and Pum HD or Nanos ZnF are satisfied (Figure 2.14b). A visual inspection of cross-links reveals that the majority of the cross-links form in the lower half of Brat NHL. The upper half of the NHL shows only two cross-links to Nanos ZnF from the very N-terminus of Brat NHL, which is likely to be flexible. The modelling yields a structure that does not

satisfy the cross-links well for two reasons. First, the space Brat NHL can occupy is limited by the RNA - it is sterically not possible for Brat NHL to rotate so that the RNA interaction surface would be completely facing away from the complex, because the NHL domain is approximatelly 30 Å high and the RNA is simply not long enough to span this distance even in its most extended conformation. Secondly, manual manipulation of Brat NHL indicates that cross-links to Pum HD and Nanos ZnF cannot be fully satisfied simultaneously (Figure 2.14c). This suggests that the Brat NHL-Pum HD interface and Brat NHL-Nanos ZnF interface are two separate interfaces and explains why simply adding the XL/MS data does not yield an accurate model of Hunchback complex.

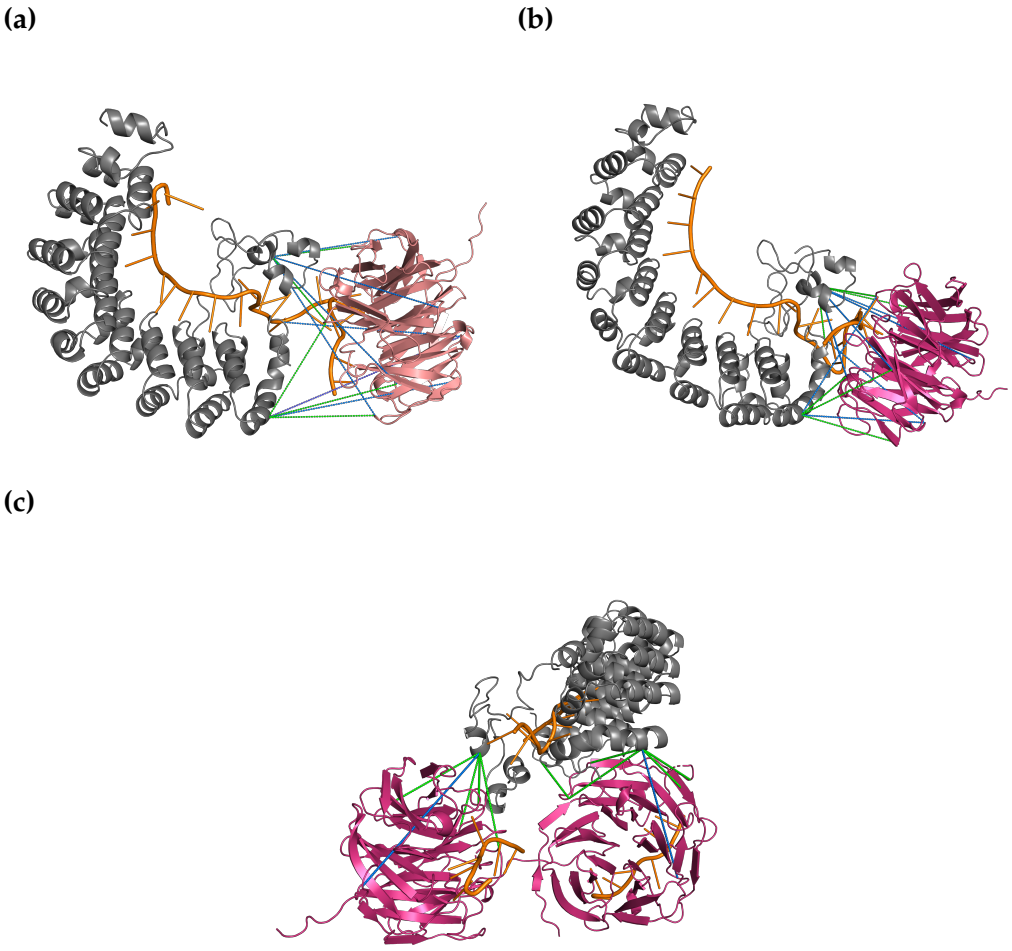


Figure 2.14 The cross-links demonstrated on Hunchback complex models. (a) The representative model of the ensemble of Hunchback complex models modelled using the cross-linking/mass-spectrometry and small angle scattering data with visualized cross-links. **(b)** The overall lowest distance restraint energy model from the aforementioned modelling with visualized cross-links. **(c)** A manually built model with duplicated Brat NHL illustrating the chance of the Brat NHL-Nanos ZnF and Brat NHL-Pum HD cross-links to be satisfied better separately. Satisfied cross-links are coloured green in each panel, unsatisfied cross-links are coloured blue in each panel.

2.4 Discussion

The first informative conclusions can be drawn from the ITC experiments. Previously, it was shown that Pum is capable of binding *hunchback* mRNA at both Pum HD and Brat binding site with comparable affinity[122]. This naturally raises a question about the interplay of Pum and Brat RNA binding. The results of the ITC experiments indeed show that Pum HD is able to bind Pum binding site and the Brat binding site, but the K_D differs in three orders of magnitude (Figure 2.4). This contrasts with the previously published similar affinity of Pum HD for both sites. However, this discrepancy could arise from the different experimental set up. In the published fluorescence anistropy assays the individual affinities for each site were measured in independent experiments with two different RNAs that each contained one of the two bindings sites. The ITC experiments reported here measure the two binding events on the same RNA simultaneously thus providing a more realistic observation that Pum HD has a clear preference for its binding site.

The binding of Pum HD to the Brat binding site is not strong enough to impair Brat binding to the RNA (Figure 2.6c). On the contrary, Brat NHL was shown to bind RNA mildy cooperatively with Pum HD[7]. The mechanism was proposed to be unfolding of RNA structure as a result of Pum HD binding which would increase the accessibility of the RNA for Brat NHL. However, the results here do not indicate any cooperativity. In fact binding of Pum HD slightly increases the K_D of Brat NHL (Figure 2.6). When structured NRE2 RNA and unstructured pNRE2 RNA are compared in terms of Brat NHL binding, Brat shows higher affinity to the pNRE2, but this RNA contains a poly-U stretch and Brat NHL was shown to interact with U-rich sequences[7]. Indeed, the ITC experiments show two binding events when Brat NHL is titrated to pNRE2 and a previous study also clearly showed that the affinity of Brat for RNA increases with the increase of Brat binding sites[128]. This suggests that the increase of affinity observed for Brat NHL with the pNRE2 originates rather in the introduction of another binding site.

To conclude, Brat NHL and Pum HD bind both specifically their respective consensus binding sites. Pum HD is a stronger RNA binder. However, these two interactions are in the context of a single NRE2 fairly independent as they are not observed to significantly reinforce or impair each other. In addition, ITC assays did not show any direct interaction of Brat NHL with Pum HD.

The interplay of Pum HD and Nanos ZnF has been recently elucidated in the structure of Pum HD-Nanos ZnF-RNA complex[6]. The Brat NHL-Nanos ZnF interaction remained unclear. The NMR titrations presented here reveal signs of interaction of Nanos ZnF and Brat NHL in the absence of RNA. The interaction seems rather weak. However, this study provides direct evidence of the existence of the quaternary Hunchback complex of Pum HD, Nanos ZnF, Brat NHL and the NRE2 RNA and within that complex Brat NHL cross-links to Nanos ZnF.

The complete atomic structure of Hunchback complex remains elusive as none of the extensive crystallization trials tested here found a crystallization conditions yielding diffracting crystals. However, this study reveals some initial characteristics of the complex and paves the way to describing the structure of the complex

by integrative approaches.

The small angle scattering and cross-linking/mass-spectrometry data presented here indicate that Hunchback complex is rather extended than compact and reveal that Brat NHL is localized in the vicinity of the N-terminus of Pum HD, where Brat NHL interacts with the N-terminal portion of Nanos ZnF and the Pum HD N-terminus likely in two independent interactions (Figure 2.11b and Figure 2.14c).

The attempts to model a more detailed architecture of Hunchback complex in here did not provide a unique solution. Nevertheless, the modelling provides valuable suggestions about the possible identity of the complex. The modelling approach presented here implicitly assumes that Brat NHL, Pum and Nanos ZnF form a monomeric quaternary complex with the Hunchback mRNA and aims to find a single conformation this complex adapts. The modelling fails when either of the three assumptions are false - either the proteins interact differently in the context of the quaternary complex in comparison to the X-ray structures, the complex adapts multiple conformations or the complex is not monomeric.

There is no evidence suggesting that the interactions would change upon complex formation. The interactions sites within the RNA are well resolved and the experiments in here suggest that at least Brat NHL and Pum HD do not interfere in RNA binding. Hence, drastic changes of the interaction in Hunchback complex formation would be very surprising. However, no experimental data explicitly describing the interaction of the individual domains with the RNA within the Hunchback complex were obtained.

Multiple conformations of the complex in solution would explain why the cross-links cannot be fully satisfied by a single conformation during the modelling. This can be easily tested using the SAS data. In SAS, a mixture of components gives rise to a signal which is an average of the scattering of the individual components. If the SAS data describe a mixture of multiple conformations of the complex, fitting a mixture of back-calculated scattering curves of multiple models should results in a better fit. However, if the fit of the mixture is only as good as the best fitting conformation, the data are most likely not describing a mixture of components. Such a preliminary test was performed by fitting the SAXS curve with a mixture of representative models of the major and minor ensembles obtained from the modelling without XL/MS data. Interestingly, the mixture fits the data better than the individual models (Supplementary Figure 8). However the quality of the fits increases only minimally suggesting that even this mixture does not describe the solution accurately. Therefore, a more sophisticated search for a mixture to fit is required to test this hypothesis properly.

The third explanation why the modelling approach does not seem to converge to a clear solution is that the assumption that Hunchback complex is monomeric is false. The modelling approach presented here does not model anything else than a single monomer of the complex, so if the data describe any other than monomeric state, the modelling will not provide a satisfying solution. In such a scenario the cross-links describing the interaction within a monomer and between multiple monomers would be then interpreted as describing interactions only within a

monomer, which would result in a large proportion of cross-links unsatisfied at any modelled conformation and a fundamentally unrealistic model. It is worth noting that XL/MS experiments on Hunchback complex at 1.5 mg/ml show a significant proportion of cross-links that are physically impossible within a single monomer of the complex. In fact, SAS and XL/MS are both quite sensitive towards minor populations under certain conditions, so it is conceivable that the complex would have a very weak tendency to dimerize. The scattering intensity in SAS scales with the square of the volume of the particle, so one globular particle in a million with an order of magnitude higher radius contributes to the signal equally as the remaining 599,999 smaller particles. This illustrates how even a minor population of particles of larger volume can affect the scattering curve significantly. Furthermore, the SAS data show a disagreement between the R_g and I_0 determined from Guinier analysis and distance distribution function (Table 2.1), which often is a sign of forming particles of higher molecular weight[134]. XL/MS essentially capture an immediate snapshot of the solution state, so any equilibria between multiple states would be visible. In summary, possibly even such a weak dimerization tendency could be observable by SAS and XL/MS data, that would still allow the complex to run as an apparent monomer during size-exclusion chromatography. To test the possibility that Hunchback complex is in such an equilibrium, it is necessary to interpret the data accordingly. To model the hypothetical oligomer and compare how well a mixture of a monomer and the oligomer fits the experimental data. The existence of such an oligomer would then need to be confirmed by orthogonal experiments. Then, its relevance for function would be the question of primary importance. However, the 3′ UTR of *Hunchback* mRNA contains two NREs, so it is conceivable to imagine a monomer of Hunchback complex assembling on each single NRE and such a monomer forming a higher order dimer structure. Moreover, Brat also contains a Coiled-Coil domain and its recently solved structure reveals a dimer[135].

Lastly, a more detailed understanding of the interplay between Brat NHL and Nanos ZnF is essential to properly understand the architecture of Hunchback complex. Therefore, follow-up experiments are absolutely crucial.

The conclusions drawn from this study may seem fairly limited. However, Hunchback complex has been an elusive object of interest for approximately three decades. This study paves an alternative way to elucidate Hunchback complex architecture using integrative structural biology.

2.5 Materials and Methods

2.5.1 Cloning, protein expression and purification

Pumilio Homology Domain

The Pumilio Homology Domain (Pum HD) was prepared as previously described[128,136,137]. The pHUE Pum HD vector was kindly provided by Dr. Inga Loedige. The vector carries a Pum HD construct encompassing Pum residues 1093 to 1426 fused by a short linker containing Usp2cc cleavage site to a His_6-Ubiquitin tag. The vector carries an Ampicillin resistance marker. To express Pum HD the vector was transformed into *E. coli* BL-21(DE3) Rosetta cells. The cells were grown until OD_{600} of 0.6-1.0 at 37°C. The protein expression was induced by the addition of IPTG to 1 mM final concentration and the protein was expressed at 23°C overnight.

The cells were harvested by centrifugation at 5000 rpm at 4°C for 40 minutes. Then, the cells were resuspended in 50 mM Tris, 1 M NaCl, 5% glycerol, 10 mM Imidazol, pH 8.0 with 1 mg/ml lysozime, 1 μg/ml DNase I, 2 μg/ml RNase and protease inhibitors tablets (1 per 50 ml, Roche cOmplete) and incubated for 20 minutes on ice. The cell lysis was completed by sonication. Then, the lysate was spun at 18000 g at 4°C for 60 minutes. The supernatant was loaded on a HisTrap HP 5 ml column (GE Healthcare) and Pum HD was eluted using 10 - 100 mM Imidazole linear gradient followed by 250 mM Imidazole. Fractions containing Pum HD were pooled and cleaved by Usp2cc during overnight dialysis to 50 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.4. The Pum HD was separated from the His_6-Ubiquitin and His_6-Usp2cc by reversing the previous Nickel affinity chromatography. In the final step Pum HD was concentrated and further purified by SEC on the HiLoad 16/600 Superdex 75 column (GE Healthcare) in the same buffer.

Pumilio Homology Domain and Maltose-binding protein fusion

A set of Maltose-binding protein (MBP) fusion constructs of Pum HD was designed for additional crystallization trials. The Pum HD costruct from pHUE Pum HD vector was cloned in parallel into the pMALX vector system[138] (kindly provided by the author) and a modified pETM-41 vector. The pETM-41 plasmid was modified by replacing the linker (including the cleavage site) with a GSGSGS (modified by Kevin Haubrich M.Sc., the vector provided by EMBL Protein Expression and Purification Facility). The pMALX vectors carry an Ampicillin resistance, whereas the pETM-41 carries a Kanamycin resistance. The pETM-41 Pum HD was optimized for expression and purification as follows.

The vector was transformed in *E. coli* BL-21(DE3) Rosetta cells. The cells were grown at 37°C until OD_{600} of 0.6-1.0 and then induced by the addition of IPTG to a final concentration of 1 mM followed by expression at 23°C overnight.

The cells were then harvested as for Pum HD and resuspended in the same

buffer, but without Imidazole. The same procedure as for Pum HD followed, but Pum HD MBP was purified using a MBPTrap HP 5 ml column (GE Healthcare). Pum HD MBP was eluted by a gradient of 0 to 10 mM Maltose in the lysis buffer. The protein was then dialysed into 50 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.4. Lastly, Pum HD MBP was loaded to a HiTrap HP Heparin 5ml column (GE Healthcare) and then eluted by 150 mM to 1.5 M NaCl gradient.

Brain Tumor NHL domain

The residues 756 to 1037 of Brain Tumor (Brat NHL) were cloned by Dr. Jaelle Foot into the pHUE vector[136,137] resulting in a construct similar to the Pum HD construct with His_6-Ubiquitin fusion. The expression and purification of Brat NHL was carried out as for the Pum HD construct with the exception of the last purification step. After overnight dialysis and cleavage Brat-NHL was loaded on in series connected HisTrap HP 5 ml and HiTrap HP Heparin 5ml columns (both GE Healthcare). First, Brat NHL was eluted by a 150 mM to 1.5 M NaCl gradient. Then, the His_6-Ubiquitin tag and His_6-Usp2cc were eluted by 1M Imidazole.

Nanos Zinc Finger domain

The Nanos construct used in this study was cloned and its expression was optimized by Sophie Winter, M.Sc.[132]. The construct comprises the residues 301-392 of Nanos, which cover the Zinc Finger domain of Nanos (Nanos ZnF), in pETM11 SUMO vector (EMBL Protein Expression and Purification Core Facility). The vector contains Nanos ZnF fused by a linker with SenP2 cleavage site to SUMO3 protein with His_6 tag and carries a Kanamycin resistence marker.

To obtain Nanos ZnF the pETM11 SUMO vector was transformed into *E. coli* BL-21(DE3) Rosetta cells. The cells were grown at 37°C until OD_{600} of 0.6-1.0. The expression was induced by the addition IPTG to 0.3 mM final concentration and the protein was expressed at 16°C overnight.

The cells were harvested by centrifugation at 5000 rpm at 4°C for 40 minutes. Then, the cells were resuspended in 50 mM Tris, 500 mM NaCl, 1 mM DTT, 10 mM Imidazole, pH 8.0 with 1 mg/ml lysozime, 1 μg/ml DNase I, 2 μg/ml RNase and protease inhibitors tablets (1 per 50 ml, Roche cOmplete). After 20 minutes incubation, the cell lysis was completed by sonication. Then, the lysate was spun at 18000 g at 4°C for 60 minutes. The supernatant was loaded on HisTrap HP 5 ml column (GE Healthcare) and Nanos ZnF was purified using a 10 - 100 mM linear Imidazole gradient followed by 250 mM Imidazole. Fractions containing Nanos ZnF were pooled and cleaved by Sen2P during overnight dialysis to 50 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.4. Nanos ZnF was then loaded to HisTrap HP 5 ml and HiTrap HP Heparin 5ml columns (both GE Healthcare). First, Nanos ZnF was eluted by a 150 mM to 1.5 M NaCl gradient. Then, the His_6-SUMO3 tag and the His_6-SenP2 were eluted by 1M Imidazole.

On some occasions throughout this study the purified Nanos ZnF was further refolded by the following procedure. First, Nanos ZnF was dialysed to 6M Urea,

50 mM Tris, 150 mM NaCl, 20 mM DTT, 20 mM EDTA, pH 7.4. Then, Nanos ZnF was dialysed twice to 50 mM Tris, 150 mM NaCl, 2 mM $ZnSO_4$, pH 7.4.

Brain Tumor Coiled-Coil domain

To study the Brain Tumor Coiled-Coil domain (Brat CC) a construct was designed spanning residues 375 to 530 of Brat. The Brat CC construct was cloned in pETM11 SUMO vector with a linker containing SenP2 cleavage site and a His_6-SUMO3 tag carrying Kanamycin (EMBL Protein Expression and Purification Core Facility). The expression and purification of Brat CC followed the same protocol as Pum HD except that Brat CC was kept in 50 mM Tris, 1 M NaCl, 5% glycerol, pH 8.0 buffer until the size-exclusion chromatography step.

Expression media and isotope labelling

The choices of media used for the expression of desired proteins followed the same general rules. All unlabelled proteins were expressed in standard Lysogeny broth (LB) medium. M9 minimal medium was used for isotope labeling with ^{13}C-glucose and/or ^{15}N-NH_4Cl as the sole carbon and nitrogen source, respectively. To obtain deuterated proteins the M9 media was prepared in D_2O and, if necessary, ^{2}H-glucose was used as the sole carbon source for perdeuteration.

2.5.2 Hunchback complex formation

The quaternary protein-RNA complex which suppresses *Hunchback* mRNA translation was investigated in this study. It consisted of Pum HD, Brat NHL, Nanos ZnF and the Nanos Response Element 2 of *Hunchback* mRNA (NRE2 RNA). The sequence of NRE2 RNA was 5' UUGUUGUCGAAAAUUGUACAUAA 3'. Hunchback complex was reconstituted as follows. First, Pum HD, Nanos ZnF and NRE2 RNA were diluted to 10 μM and Brat NHL was diluted to 20 μM. Diluted Pum HD was then added to NRE2 RNA, the mixture was shortly incubated on ice, then Nanos ZnF was added, the mixture was again shortly incubated on ice and eventually, Brat NHL was added. The same volume of all proteins was added, so the complex was prepared by incubating the Pum HD, Nanos ZnF, NRE2 RNA and Brat NHL in 1:1:1:2 molar ratio. The mixture was then concentrated by a concentrator with a 3 kDA cutoff to around 1 ml. Then, SEC was used to purify the complex using the HiLoad 16/600 Superdex 200 pg column (GE Healthcare). The SEC run was analyzed by SDS PAGE.

2.5.3 Electrophoretic mobility shift assays

Electrophoretic mobility shift assays were used to asses cooperativity in RNA binding by Brat NHL and Pum HD. All RNA-binding reactions were performed in 50 mM Tris-HCl, 150 mM NaCl, 1 mM DTT, 10 % [v/v] glycerol, pH 7.4 buffer. Reactions were equilibrated for 1 hour at 4°C. Next, the samples were resolved on 6%

native polyacrylamide gel in 0.5x TBE at 4°C for 3-4 hours. Each reaction contained a radiolabeled probe (1 nM NRE2 RNA) which was obtained by 5′ end labeling with T4 Kinase (NEB) and ATP [γ-32P] (Hartmann Analytic). Gels containing radioactive RNAs were dried onto Whatman filter paper. The radioactive gels were then exposed to a storage phosphor screen (GE) for 16 hours. EMSAs were imaged with a Typhoon Trio imager (GE Healthcare) and subsequently quantified using ImageQuant TL software.

2.5.4 Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy in this study was mostly used to investigate the interactions between the components of Hunchback complex. The main interest was the potential interaction of Brat NHL with Nanos ZnF. For that the resonance assignments of Nanos ZnF and Brat NHL were previously done by Sophie Winter, M.Sc.[132] and Dr. Janosch Hennig (unpublished) and therefore shall not be reported here. NMR titrations were performed to test the Nanos ZnF-Brat NHL interaction. A ^{15}N labelled Nanos ZnF (prepared as described in subsection 2.5.1) was titrated by adding unlabelled Brat NHL to 1:0, 1:1 and 1:2 ZnF:Brat NHL ratio and each step of the titration was monitored by recording a gradient-enhanced ^{1}H, ^{15}N-HSQC spectrum. The NMR measurements were done on a Bruker Avance III spectrometer with a magnetic field strength corresponding to an 800 MHz proton Larmour frequency equipped with a Bruker TXI cryo-probe head. The measurements were done at 25°C in 50 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.4 buffer. The spectra were processed using NMRPipe[62]. To analyze the NMR titrations chemical shift perturbations (CSP) were calculated according to Equation 1.1.

2.5.5 X-ray Crystallography

Hunchback complex was prepared as described in subsection 2.5.2 to attempt crystallization. All crystallization trials were set up with the complex in 50 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.4 buffer. The complex was first concentrated in a 3 kDa cutoff concentrator to 10-15 mg/ml. The sample was then spun down to remove potential precipitate in a table top centrifuge at max speed for 10 minutes. Then, the sample was diluted to desired concentrations and the crystallization trials were set up in the EMBL Heidelberg Crystallization Facility using the Mosquito crystallization robot (TTP LabTech) in MRC Crystallization plates (2 wells, round bottom). An extensive set of commercial screens was tested at various protein concentrations. The full list including the sample concentrations used can be found in Table 2.2. The automated commercial screens used a vapour diffusion crystallization method in the hanging drop set up with a drop made of 100 nl of the protein samples and 100 nl of the mother liquor. The crystallization was tested at 20°C and 4°C.

Table 2.2 The crystallization screens tested for Hunchback complex.

		Concentrations tested [mg/ml]	
Screen	Supplier	4°C	20°C
Classics	Qiagen	1.5, 3.0, 5.13, 10.0	1.5, 3.0, 5.13, 10.0
JCSG+	Molecular Dimensions	1.5, 3.0, 5.13, 10.0	1.5, 3.0, 5.13, 10.0
PACT	Molecular Dimensions	1.5, 3.0, 5.13	1.5, 3.0, 5.13
PEGS	Qiagen	1.5, 3.0, 5.13	1.5, 3.0, 5.13
Wizzard I+II	Rigaku	1.5, 3.0, 5.13, 10.0	1.5, 3.0, 5.13, 10.0
Natrix	Hampton Research	1.5, 3.0, 5.13, 10.0	1.5, 3.0, 5.13, 10.0
Nucliex	Qiagen	1.5, 3.0, 5.13, 10.0	1.5, 3.0, 5.13, 10.0
MPD	Qiagen	1.63, 2.13, 2.63, 3.13, 3.63, 4.13, 4.63, 5.13	-
$AmSO_4$	Qiagen	1.5, 3.0, 5.0	3.0, 5.0
MIDAS	Molecular Dimensions	1.5, 3.0, 5.0	1.5, 3.0, 5.0
Morpheus	Molecular Dimensionns	1.5, 3.0, 5.0	1.5, 3.0, 5.0
PEGs II	Qiagen	1.5, 3.0, 5.0	3.0, 5.0
SaltRX	Hampton Research	1.5, 3.0, 5.0	3.0, 5.0
The BCS screen	Molecular Dimensions	3.0	3.0

2.5.6 Small-angle scattering

Small-angle scattering (SAS) was used in this study to obtain low resolution information about the shape and structure of Hunchback complex.

The overall shapes of the complex were investigated by small-angle X-ray scattering (SAXS). The complex was prepared unlabelled as described in subsection 2.5.2. Prior to the measurements the samples were concentrated to 3.7 mg/ml. Hunchback complex was measured in 50 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.4 buffer at 25°C.

The SAXS measurements were done with the assistance of the beamline scientists at the BM29 beamline in the European Synchrotron Radition Source, Grenoble, France. The samples were measured in the constant flow mode to reduce radiation damage, using a 2D Pilatus detector and X-rays of λ = 1.008 Å.

The complexes for SANS measurements were prepared as described in subsection 2.5.2. Multiple samples were prepared with various combinations of perdeuterated subunits. A detailed list of the samples is in the Table 2.3.

The SANS measurements were done with the assistance of Dr. Frank Gabel and Dr. Anne Martel at D22 instrument at the Institute Laue–Langevin, Grenoble, France. In each measurement 200 μl of the sample were measured in Hellma® 100QS quartz cuvettes. The following set of standards was always measured: the empty beam, an empty quartz cell, a boron sample and a calibration curve of a buffer with varying D_2O content. The boron sample was measured for 5 minutes, the empty cell for 10 minutes, the buffers for 20 minutes each and the samples of Hunchback complex were measured for 60 minutes each. Sample transmission was measured prior to each measurement for 1 minute. The wavelenght of the neutrons used was 6 Å. The measurements were done with a laterally shifted detector to obtain the best q range while reducing the time requirements for the measurement. This allowed to collect data faster for the whole q range that contained useful information in a single measurement. Hunchback complex was measured at 4 meters detector and collimator distance. The transmission measurements of the

buffers with different D_2O content were used to calculate a calibration curve for the determination of the actual exact D_2O concentration in Hunchback complex samples.

All experimental curves were buffer subtracted and potential initial points showing signs of aggregation were removed in PRIMUS[139]. The data was further analyzed by various other software of the ATSAS package[140] and in ScaÅtter[141].

Table 2.3 The list of samples for small-angle scattering measurements.

Sample	Perdeuteration scheme				Concentration [mg/ml]	D_2O concentration [%]
	Brat NHL	Pum HD	Nanos ZnF	NRE2 RNA		
SAXS	^{1}H	^{1}H	^{1}H	^{1}H	3.7	0
1B1P1N 0% D_2O	^{1}H	^{1}H	^{1}H		3.7	0
1B1P1N 66% D_2O	^{1}H	^{1}H	^{1}H		3.7	66
1B2P2N 0% D_2O	^{1}H	^{2}H	^{2}H		5.0	0
1B2P2N 41% D_2O	^{1}H	^{2}H	^{2}H		4.3	41
1B2P2N 68% D_2O	^{1}H	^{2}H	^{2}H		4	68
2B1P2N 0% D_2O	^{2}H	^{1}H	^{2}H		4	0
2B1P2N 44% D_2O	^{2}H	^{1}H	^{2}H		4	44
2B1P2N 59% D_2O	^{2}H	^{1}H	^{2}H		4	59
2B1P2N 82% D_2O	^{2}H	^{1}H	^{2}H		4	082
2B2P1N 0% D_2O	^{2}H	^{2}H	^{1}H		4	0
2B2P1N 33% D_2O	^{2}H	^{2}H	^{1}H		4	33
2B2P1N 62% D_2O	^{2}H	^{2}H	^{1}H		4	62

2.5.7 Isothermal titration calorimetry

Isothermal titration calorimetry measurements were done to determine the binding parameters and kinetics of the purified proteins interacting with RNAs. Sets of experiments were measured for both the NRE2 RNA and the pNRE2 RNA (for full sequences see subsection 2.5.2). The pNRE2 RNA measurements were done in 20 mM Tris, 150 mM NaCl, pH 7.4 and the NRE2 RNA measurements were done in 50 mM Tris, 150 mM NaCl, 0.5 mM TCEP, pH 7.4 buffer.

The measurements involving the NRE2 RNA were all done on MicroCal PEAQ-ITC instrument (Malvern) at 20°C. The NRE2 RNA was always first snap-cooled by incubating it shaking at 65°C and then quickly transferring it to ice for an incubation of at least 20 minutes. The snap-cooling was done to ensure the RNA adapts its thermodynamically most favoured confirmation. The pNRE2 measurements were done on the iTC200 (MircoCal) at 20°C. pNRE2 RNA is designed to have no propensity to form secondary structure, so for the measurements including this RNA the samples were just titrated directly. The samples were always degassed prior to the titration. The proteins at high concentrations in the syringe were then titrated to the diluted NRE2 RNA in the cell. The titrations were performed with an initial injection of 0.4 μl followed by either 12, 15 or 20 injections of either 3, 2.5 or 2 μl, respectively. The number of injection was selected according to the enthalpy change for the best compromise between the amount of signal and the maximum number of points. The initial delay and the delay between injections were set to 60 and 150 seconds, respectively. The samples were stirred at 750 rpm and the instrument feedback was set to high. The reference power was set to 10 μcal/s.

Table 2.4 The overview of the isothermal titration calorimetry measurements.

Experiment	Cell	Cell concentration [μM]	Syringe	Syringe concentration [μM]	Injections	Injection size [μl]
Figure 2.4a	Pum consensus RNA	30	Pum HD	300	13	3 (0.4)
Figure 2.4b	pNRE2 RNA	20	Pum HD	400	21	2 (0.4)
Figure 2.4c	NRE2 RNA	20	Pum HD	400	16	2.5 (0.4)
Figure 2.5a	Brat consensus RNA	30	Brat NHL	290	21	2 (0.4)
Figure 2.5b	pNRE2 RNA	30	Brat NHL	390	20	2
Figure 2.5c	NRE2 RNA	25	Brat NHL	375	13	3 (0.4)
Figure 2.6a	pNRE2 RNA + Brat NHL	20	Pum HD	440	21	2 (0.4)
Figure 2.6b	pNRE2 RNA + Pum HD	20	Brat NHL	430	21	2 (0.4)
Figure 2.6c	NRE2 RNA + Pum HD	24	Brat NHL	375	13	3 (0.4)

Part of the measurements were performed by Dr. Vladimir Rybin and Dr. Karthryn Perez, who also provided further support. The instrumentation was provided by the EMBL Protein Expression and Purification Core Facility.

2.5.8 Cross-linking

Cross-linking and mass spectrometry (XL/MS) was used to obtain distance restrains for the modelling of Hunchback complex. All XL/MS experiments were done with the complex reconstituted as described in subsection 2.5.2 and DSS as the cross-linking reagent. The cross-linking reaction was done in 20 mM HEPES, 150 mM NaCl, 1 mM DTT, pH 7.4 buffer and followed a previously published protocol[142].

First, a 1.5 mg/ml Hunchback complex was cross-linked by four different concentrations of DSS - 0.1 mM, 0.5 mM, 1 mM and 2.5 mM. The individual proteins at same molar concentrations were cross-linked as controls by 1 mM DSS. The cross-linked samples together with not cross-linked proteins were then analyzed by SDS PAGE. The cross-linking was then repeated to collect the data with an equimolar mixture of ^{1}H-DSS and ^{2}H-DSS at 0.2 mM, 0.5 mM and 1 mM concentration. The 0.2 mM DSS reaction was done once at 1.5 mg/ml complex concentration, the 0.5 mM DSS reaction was done in two replicates at 0.95 mg/ml complex concentration and the 1 mM reaction was done in three replicates at 1.5 mg/ml complex concentration.

First, 50 mM DSS dissolved in DMF was added up to the desired cross-linking reagent concentration of 1.5 mg/ml complex in 200 μl. The reaction mixture was then incubated shaking at 37°C for 30 minutes. 1 M NH_4HCO_3 was then added to a final concentration of 50 mM to quench the reaction and the mixture was incubated at 37°C for 20 minutes. A few μl of the samples were analyzed by SDS PAGE and the sample was kept at 4°C overnight. The next day, 10 M Urea in 250 mM NH_4HCO_3 was added to 8 M final Urea concentration. Then, RapiGest surfactant (Waters) dissolved in 10 mM NH_4HCO_3 was added to 1 mg/ml final concentration. The mixture was then sonicated on ice by a single 60s pulse at 10% power. 1 M DTT in H_2O was then added to 10 mM final concentration and the mixture was incubated shaking at 37°C for 30 minutes. Then, 100 mM IAA in H_2O was added to 15 mM final concentration and the mixture was incubated shaking at 37°C for 30 minutes in the dark. Next, 0.1 μg/ml LysC (Wako) dissolved in 10 mM NH_4HCO_3 was added in 1:100 protease:protein ratio. The mixture was then

incubated shaking at 37°C for 3.5 hours. HPLC grade H_2O was then added to dilute the sample to 1.5 M Urea concentration. Then, 1 mg of Trypsin (Promega) dissolved in supplied buffer was added in 1:50 protease:protein ratio and the sample was incubated shaking 37°C overnight. On the third day, TFA was added to 1% v/v final concentration. The sample was then incubated shaking 37°C for 30 minutes. At the end, the sample was spun in a table top centrifuge to remove any potential precipitate for 20 minutes at max speed . Finally, the samples were submitted to the EMBL Proteomics Core Facility for the MS analysis.

2.5.9 Modelling

Structural modelling was done to obtain insights into the architecture of the complexes using the Crystallography & NMR System (CNS)[143,144]. The Pum HD-Brat NHL-pNRE2 RNA complex was modelled based on the starting structures of Pum HD in the complex with the RNA (PDB ID 1M8W)[145] and Brat NHL in the complex with RNA (PDB ID 4ZLR)[7]. Hunchback complex was modelled using the same Brat NHL-RNA complex structure and using the Pum HD-Nanos ZnF-NRE2 RNA complex structure(PDB ID 5KL1)[6]. The Brat NHL-RNA complex structure actually contains two NHL domains bound to a single long RNA, so for the modelling only the one bound to the reported Brat binding site on the *Hunchback* mRNA was used. The starting models were built by merging together the individual RNA complexes and connecting and extending the RNAs from the structures by nucleotides according to the RNA sequences (see subsection 2.5.2). These nucleotides were built manually in random extended conformations.

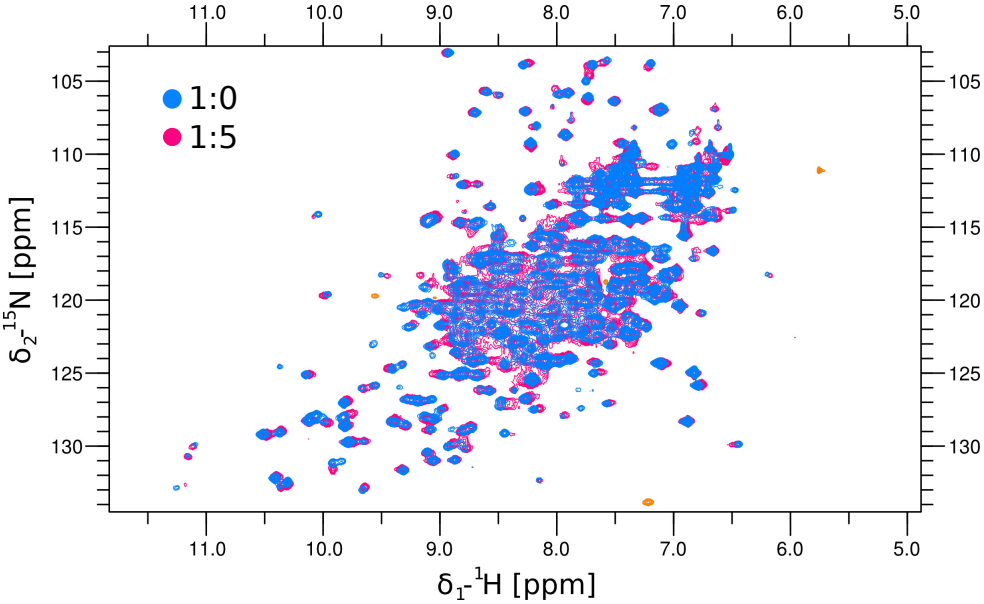
The general modelling done in this study adapted a protocol set up by Dr. Bernd Simon previously described[146]. All the modelling was done by simulated annealing using ARIA[147] in torsion angle space. During the modelling the parts of the model that corresponded to already elucidated structures were kept rigid. The parts of the RNA that were not included in any of the crystal structures were then kept free during the modelling. First, the conformation of the free nucleotides of the RNA was randomized. Then, a standard simulated annealing with the initial temperature set to 10,000 K followed. Hunchback complex modelling without distance restraints generated 5055 models in 40,000 MD steps with each step 6 fs long. Hunchback complex modelling with XL/MS data used as distance restraints was done by generating 4572 models in 120,000 steps with each step 2 fs long. The distance restraints used were set as 30 Å C_α-C_α upper distance limit with log harmonic potential.

The generated models were then used to back-calculate their theoretical scaterring curves in the same specific conditions used for SAS measurements (see subsection 2.5.6) by CRYSOL[148] and CRYSON[149]. These back-calculated curves were than fitted against the experimental curves. The energy, the distance restraints energy and the quality of the fit to the SAS experimental data were then used as a selection criteria to search for the most correct structures.

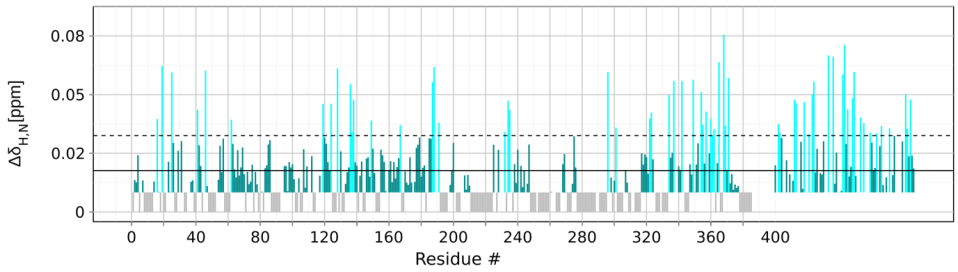
Appendix

Supplementary Material

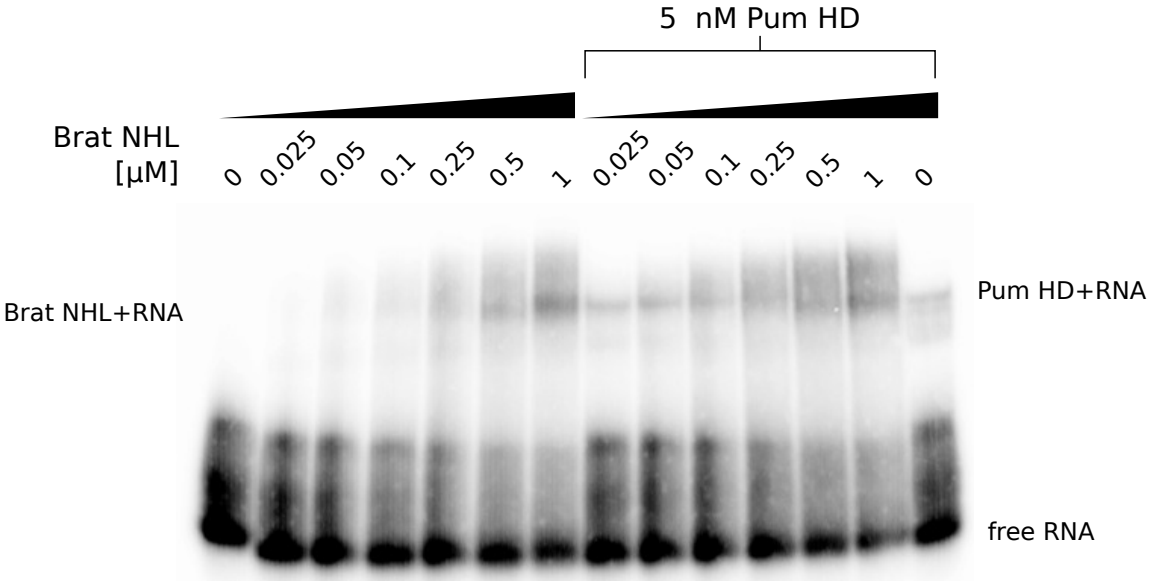
(a)



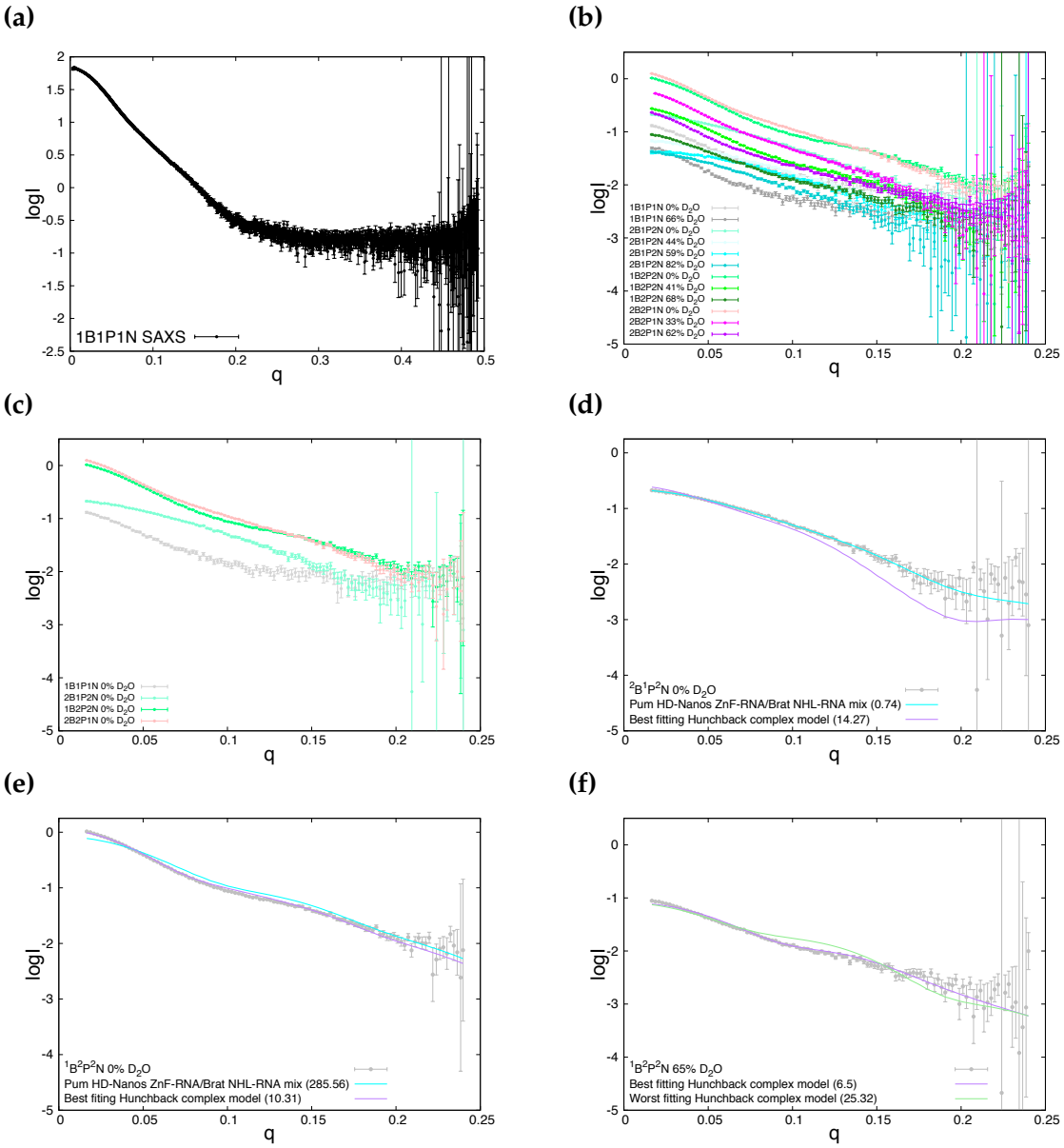
(b)



Supplementary Figure 1 The NMR titration of EarP by dTDP-Rhamnose. (a) Overlay of the ^{15}N HSQC spectra of EarP titrated by dTDP-rhamnose. ^{15}N labelled EarP was as titrated by dTDP-rhamnose to 1:5 EarP:dTDP-rhamnose molar ratios with steps at 1:0, 1:0.2, 1:1 and 1:3 ratios. Only the first and the last steps are shown. **(b)** Chemical Shift Perturbation (CSP) plot of the titration. The CSP was calculated according to Equation 1.1 for each peak and then plotted against residues number. The full line indicates the mean, the dashed line the mean plus standard deviation. The bars in the right portion of the plot without numbering show the unassigned peaks.

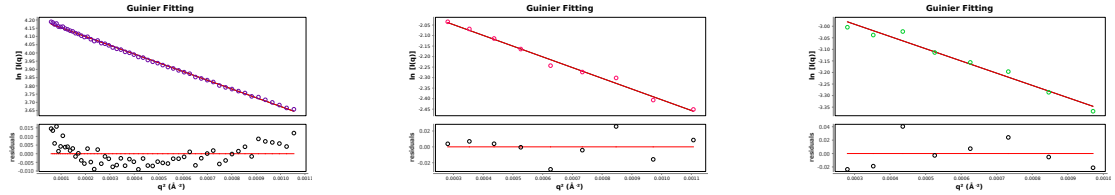


Supplementary Figure 2 Brain tumor-RNA interaction in electrophoretic mobility shift assays (EMSA). Brain tumor NHL domain (Brat NHL) was incubated with 1 nM radiolabeled NRE2 RNA at concentrations indicated above the gel for 1 hours at 4°C. The gel was then resolved on native 6% polyacrylamide gel over 3-4 hours at 4°C and then storage phosphor screen was exposed to the gel for 16 hours. The experiments in the second half of the gel contained additional 5 nM Pum HD in the mixture.

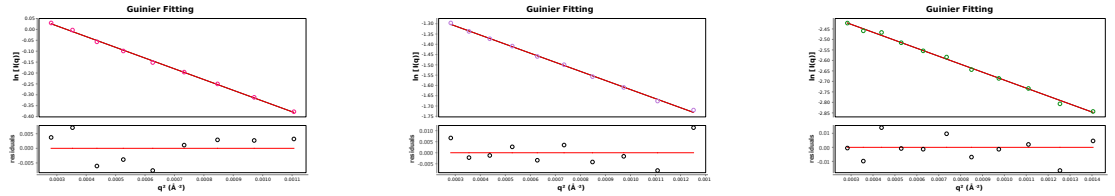


Supplementary Figure 3 The small-angle X-ray and neutron scattering of the Hunchback complex. **(a)** The small-angle X-ray scattering (SAXS) curve of the Hunchback complex. **(b)** The small-angle neutron scattering (SANS) curves of the Hunchback complex. The curves were measured on complex samples with varying deuteration schemes at different concentrations of D_2O in the buffer. This is indicated in labelling of the curves where B stands for Brat NHL, P stands for Pum HD and N stannds for Nanos ZnF. The preceding number than indicates whether the protein is protonated (1) or perduetarated (2) and then concentration of D_2O in the buffer is stated in % [v/v]. The samples arefully described in (Table 2.3. **(c)** The comparison of SANS curves of the four variously deutared Hunchback complexes measured in 0% D_2O buffer. **(d)** The fits to the SANS curve of the Hunchback complex with ^{2}H Brat NHL and Nanos ZnF measured in 0% D_2O buffer. The two fits compared are the fit of the back-calculated curve of the Hunchback complex model which fits this curve the best and the back-calculated curve of the mixture of Pum HD-Nanos ZnF-RNA complex and Brat NHL-RNA complex in 0.56:0.44 volume fractions. The χ^2 values of the fits are given in the parentheses. **(e)** The fits to the SANS curve of the Hunchback complex with ^{2}H Pum HD and Nanos ZnF measured in 0% D_2O buffer. The shows analogous fits to **(d)** except that the Hunchback complex model is the one which fits this curve the best. **(f)** The fits to the SANS curve of the Hunchback complex with ^{2}H Pum HD and Nanos ZnF measured in 65% D_2O buffer. The plot follows the same scheme as in **(d)** and **(e)** except that the Hunchback complex models are the ones which fits this curve the best and the worst.

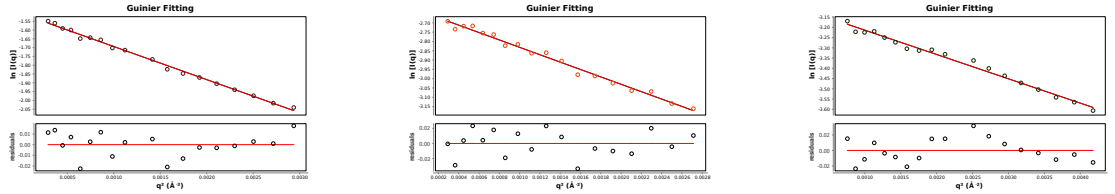
(a) SAXS **(b)** 1B1P1N 0% D_2O **(c)** 1B1P1N 66% D_2O



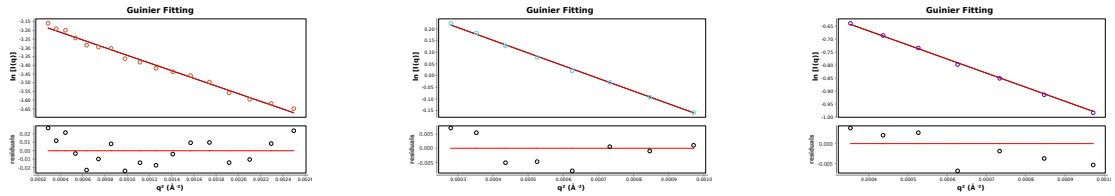
(d) 1B2P2N 0% D_2O **(e)** 1B2P2N 41% D_2O **(f)** 1B2P2N 68% D_2O



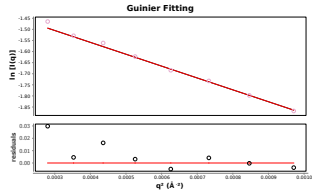
(g) 2B1P2N 0% D_2O **(h)** 2B1P2N 44% D_2O **(i)** 2B1P2N 59% D_2O



(j) 2B1P2N 82% D_2O **(k)** 2B2P1N 0% D_2O **(l)** 2B2P1N 33% D_2O

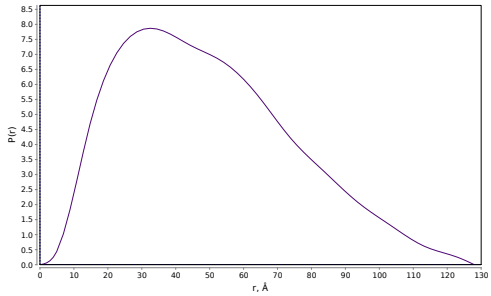


(m) 2B2P1N 62% D_2O

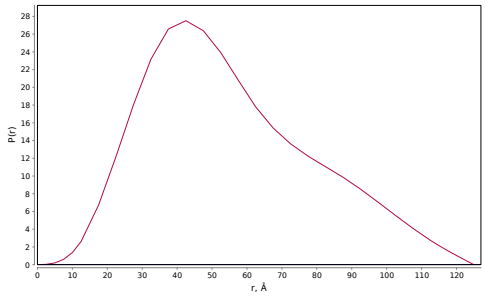


Supplementary Figure 4 Guinier analysis of the small-angle scattering (SAS) curves. Each panel shows the fit from Guinier analysis for one experimentally measured scattering curve in the top panel with the residuals in the bottom panel. All curved were fitted to obey $q \times R_G < 1.3$[150].

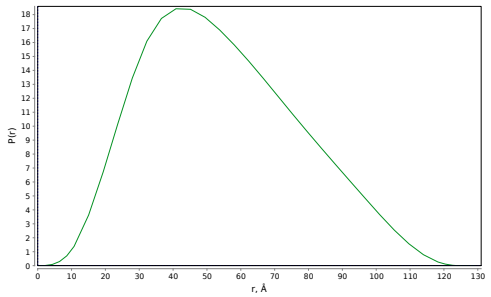
(a) SAXS



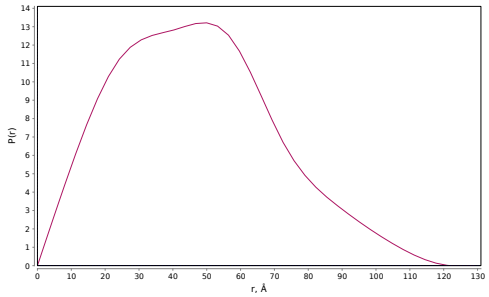
(b) 1B1P1N 0% D_2O



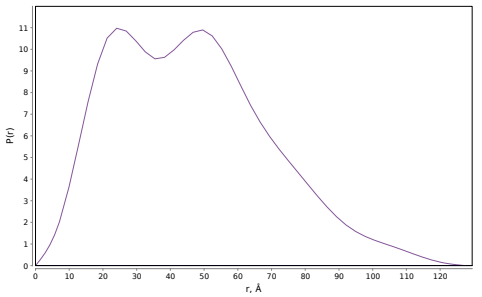
(c) 1B1P1N 66% D_2O



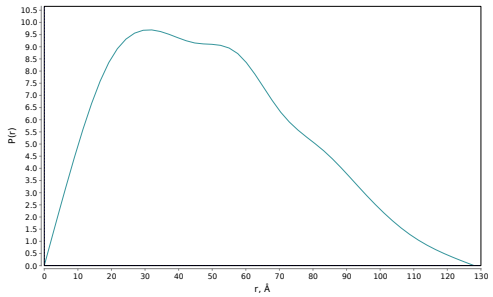
(d) 1B2P2N 0% D_2O



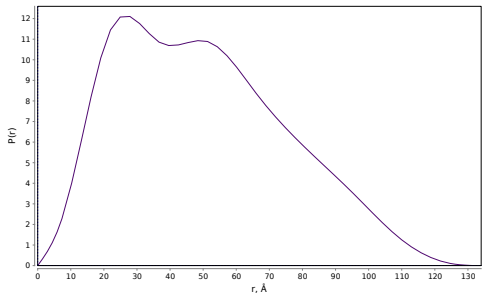
(e) 1B2P2N 41% D_2O



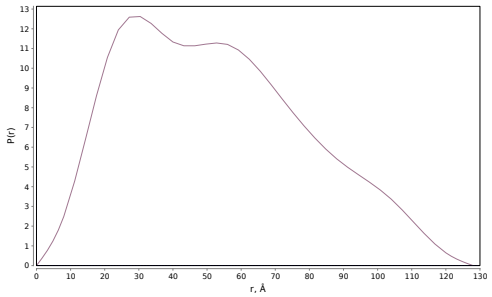
(f) 2B2P1N 0% D_2O



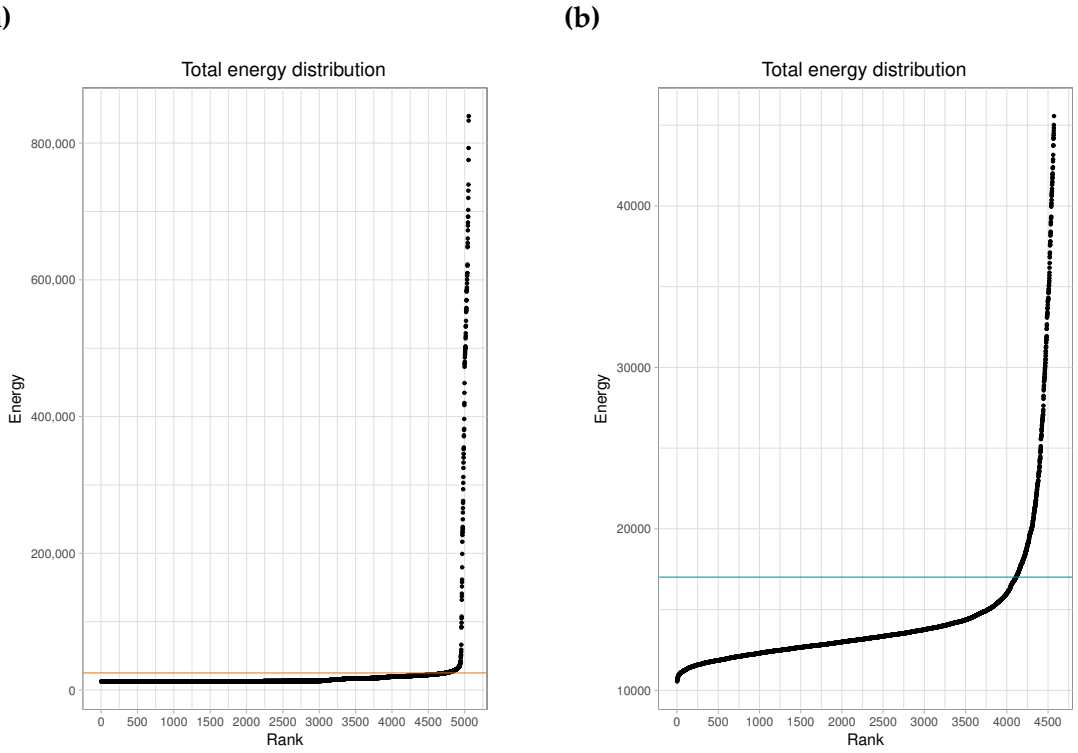
(g) 2B2P1N 33% D_2O



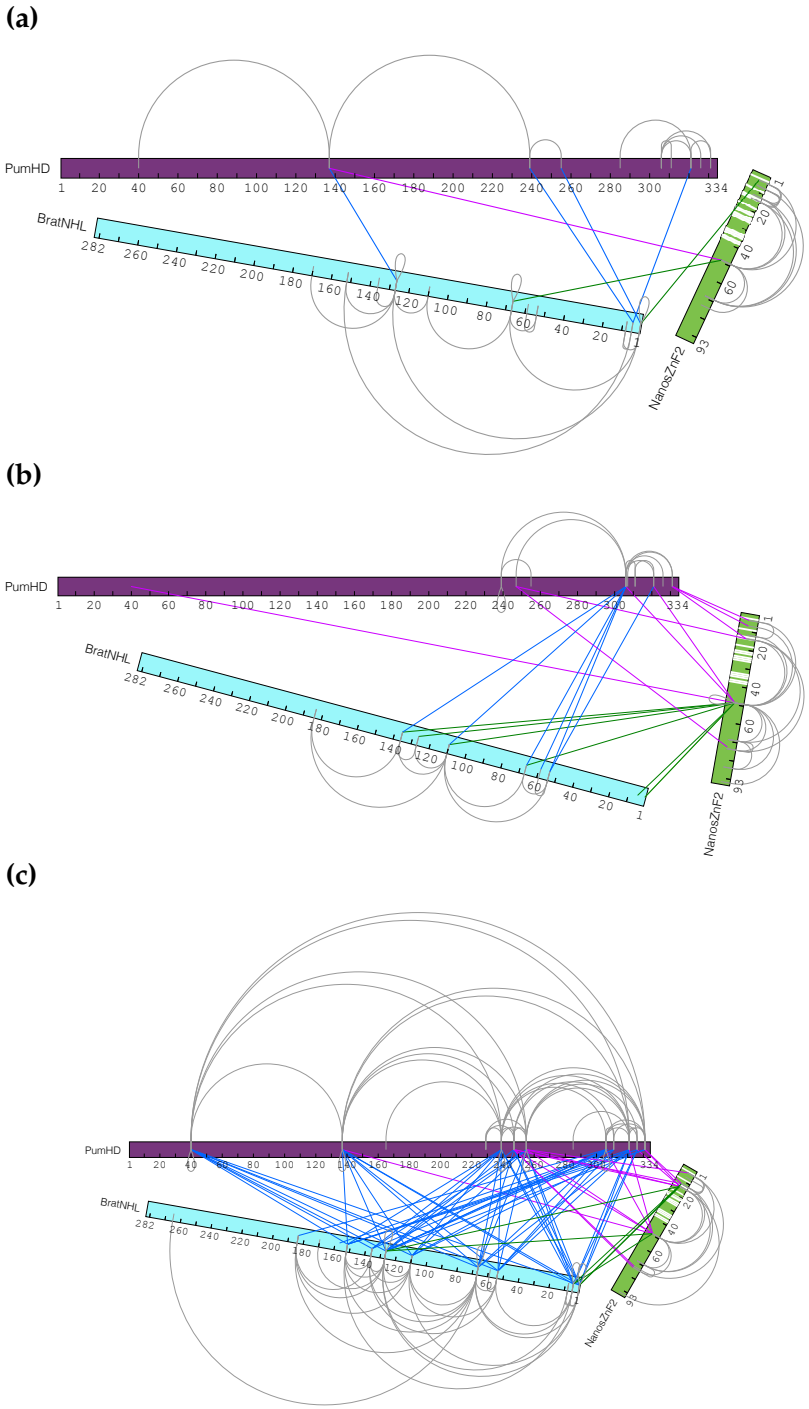
(h) 2B2P1N 62% D_2O



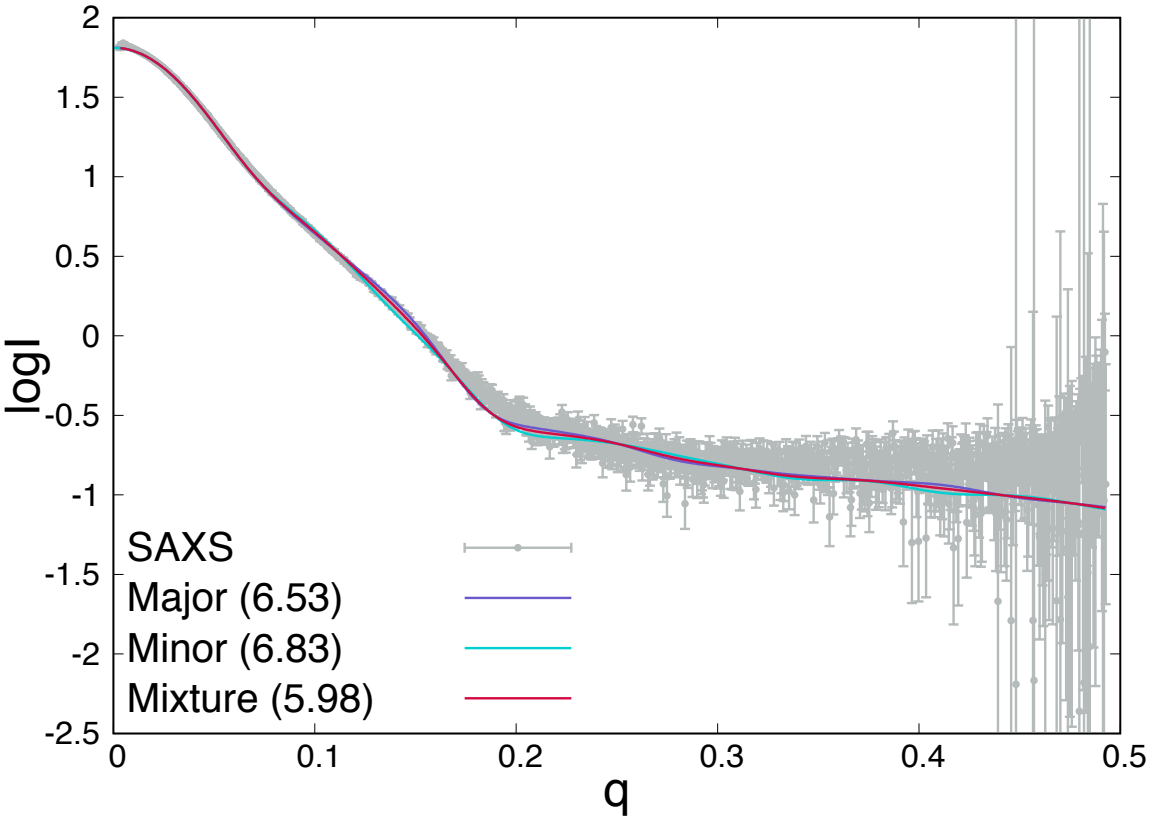
Supplementary Figure 5 Distance distribution function of selected small-angle scattering (SAS) curves. Each plot shows a distance distribution function of one experimentally measured scattering curve.



Supplementary Figure 6 The distribution of the overall energy of all models of the Hunchback complex. (a) The total energy of the 5055 models calculated during the modelling of the Hunchback complex using only small-angle scattering (SAS) data. The red line shows an energy cutoff above which the models were discarded. **(a)** The total energy of the 4572 models calculated during the modelling of the Hunchback complex using cross-linking/mass spectrometry and SAS data. The teal line shows an energy cutoff above which the models were discarded.



Supplementary Figure 7 The cross-linking/mass spectrometry of the Hunchback complex. (a) The cross-links obtained using 1.5 mg/ml Hunchback complex and 0.2 mM DSS. **(b)** The cross-links obtained using 1.5 mg/ml Hunchback complex and 1 mM DSS. **(c)** The cross-links obtained using 0.95 mg/ml Hunchback complex and 0.5 mM DSS. Inter-protein cross-links are in blue, magenta or green, intra-protein cross-links are in grey. Only cross-links with xQuest score above 25 are shown. The white stripes in Nanos ZnF indicate the residues undergoing significant chemical shift perturbation in NMR titration of Nanos ZnF by Brat NHL (Figure 2.8).



Supplementary Figure 8 The fit of a mixture of Hunchback complex models to the small-angle X-ray scattering curve. The fits of the representative models of the major and minor ensembles of models of the Hunchback complex obtained by modelling using the small-angle scattering data are compared the fit of a mixture of the two models. The fit of the mixture represent 0.52:0.48 major:minor ensemble volume fractions. The χ^2 values of the fits are given in the parentheses.

Supplementary Table 1 The pH against PEG 3350 custom crystallization screen for EarP.

	Column / Row	1	2	3	4	5	6
PEG 3350 [% w/v]	A	19	21	23	25	27	29
0.1 M bis-Tris pH		5.5	5.5	5.5	5.5	5.5	5.5
PEG 3350 [% w/v]	B	19	21	23	25	27	29
0.1 M bis-Tris pH		6.0	6.0	6.0	6.0	6.0	6.0
PEG 3350 [% w/v]	C	19	21	23	25	27	29
0.1 M bis-Tris pH		6.5	6.5	6.5	6.5	6.5	6.5
PEG 3350 [% w/v]	D	19	21	23	25	27	29
0.1 M bis-Tris pH		7.0	7.0	7.0	7.0	7.0	7.0

The crystallization screen was done 4°C with 8 mg/ml EarP in 50 mM Tris, 100 mM NaCl, 1 mM DTT, 10 mM TDP-Rhamnose, pH 7.6 buffer.

Supplementary Table 2 The pH against PEG 3350 with Amonium Acetate custom crystallization screen for EarP.

	Column / Row	1	2	3	4	5	6
PEG 3350 [% w/v]	A	19	21	23	25	27	29
0.1 M bis-Tris pH		5.5	5.5	5.5	5.5	5.5	5.5
Amonium Acetate [M]		0.1	0.1	0.1	0.1	0.1	0.1
PEG 3350 [% w/v]	B	19	21	23	25	27	29
0.1 M bis-Tris pH		6.0	6.0	6.0	6.0	6.0	6.0
Amonium Acetate [M]		0.1	0.1	0.1	0.1	0.1	0.1
PEG 3350 [% w/v]	C	19	21	23	25	27	29
0.1 M bis-Tris pH		6.5	6.5	6.5	6.5	6.5	6.5
Amonium Acetate [M]		0.1	0.1	0.1	0.1	0.1	0.1
PEG 3350 [% w/v]	D	19	21	23	25	27	29
0.1 M bis-Tris pH		7.0	7.0	7.0	7.0	7.0	7.0
Amonium Acetate [M]		0.1	0.1	0.1	0.1	0.1	0.1

The crystallization screen was done 4°C with 8 mg/ml EarP in 50 mM Tris, 100 mM NaCl, 1 mM DTT, 10 mM TDP-Rhamnose, pH 7.6 buffer.

Supplementary Table 3 The pH against PEG 10k with Amonium Acetate custom crystallization screen for EarP.

	Column / Row	1	2	3	4	5	6
PEG 10k [% w/v]		11	13	15	17	19	21
0.1 M bis-Tris pH	A	5.5	5.5	5.5	5.5	5.5	5.5
Amonium Acetate [M]		0.1	0.1	0.1	0.1	0.1	0.1
PEG 10k [% w/v]		11	13	15	17	19	21
0.1 M bis-Tris pH	A	6.0	6.0	6.0	6.0	6.0	6.0
Amonium Acetate [M]		0.1	0.1	0.1	0.1	0.1	0.1
PEG 10k [% w/v]		11	13	15	17	19	21
0.1 M bis-Tris pH	A	6.5	6.5	6.5	6.5	6.5	6.5
Amonium Acetate [M]		0.1	0.1	0.1	0.1	0.1	0.1
PEG 10k [% w/v]		11	13	15	17	19	21
0.1 M bis-Tris pH	A	7.0	7.0	7.0	7.0	7.0	7.0
Amonium Acetate [M]		0.1	0.1	0.1	0.1	0.1	0.1

The crystallization screen was done 4°C with 8 mg/ml EarP in 50 mM Tris, 100 mM NaCl, 1 mM DTT, 10 mM TDP-Rhamnose, pH 7.6 buffer.

List of Publications

M. Hassler, I. A. Shaltiel, M. Kschonsak, B. Simon, F. Merkel, L. Tharichen, H. J. Bailey, J. Macošek, S. Bravo, J. Metz, J. Hennig, and C. H. Haering. Structural Basis of an Asymmetric Condensin ATPase Cycle. *Mol. Cell*, 74(6):1175–1188, Jun 2019.

W. Volkwein, R. Krafczyk, P. K. A. Jagtap, M. Parr, E. Mankina, J. Macošek, Z. Guo, M. J. L. J. Furst, M. Pfab, D. Frishman, J. Hennig, K. Jung, and J. Lassak. Switching the Post-translational Modification of Translation Elongation Factor EF-P. *Front Microbiol*, 10:1148, May 2019.

R. Krafczyk*, J. Macošek*, P. K. A. Jagtap, D. Gast, S. Wunder, P. Mitra, A. K. Jha, J. Rohr, A. Hoffmann-Roder, K. Jung, J. Hennig, and J. Lassak. Structural Basis for EarP-Mediated Arginine Glycosylation of Translation Elongation Factor EF-P. *MBio*, 8(5), Sep 2017.

X. Li, R. Krafczyk, J. Macošek, Y. L. Li, Y. Zou, B. Simon, X. Pan, Q. Y. Wu, F. Yan, S. Li, J. Hennig, K. Jung, J. Lassak, and H. G. Hu. Resolving the α-glycosidic linkage of arginine-rhamnosylated translation elongation factor P triggers generation of the first ArgRha specific antibody. *Chem Sci*, 7(12):6995–7001, Dec 2016.

*equal contribution

Contributions

The expression and purification of EarP and EF-P, the *in vitro* and *in vivo* activity assays and the bacteria hybrid assays were done in the laboratory of Prof Dr. Kirsten Jung under the supervision of Dr. Jürgen Lassak by Dr. Ralph Krafczyk and others as described in previously published works[2,55,60]. The resonance assigned of EarP was done with the help of Dr. Janosch Hennig. The resonance assignment of the EF-P-attached rhamnose was done by Dr. Bernd Simon, who also provided help with recording the ^{13}C HSQC spectra without decoupling. The crystals structure of EarP was solved with the help of Dr. Pravin Jagtap.

The cloning, the optimization of expression and purification, and the resonance assignment of Nanos ZnF was done by Sophie Winter, M.Sc. as described in the previously published thesis[132]. The cloning, the optimization of expression and purification of Brat NHL was done by Dr. Jaelle Foot and the resonance assignment of Brat NHL was done Dr. Janosch Hennig. The electrophoretic mobility shift assays were done by Dr. Pawel Masiewicz. Some of the isothermal titration calorimetry measurements were done by Dr. Vladimir Rybin and Dr. Kathryn Perez. The mass spectrometry was done by Dr. Mandy Rettel. The crystallization trails were set up by Dr. Brice Murciano.

Bibliography

[1] J. Lassak, E. C. Keilhauer, M. Furst, K. Wuichet, J. Godeke, A. L. Starosta, J. M. Chen, L. Søgaard-Andersen, J. Rohr, D. N. Wilson, S. Haussler, M. Mann, and K. Jung. Arginine-rhamnosylation as new strategy to activate translation elongation factor P. *Nat. Chem. Biol.*, 11(4):266–270, Apr 2015. [PubMed Central:PMC4451828] [DOI:10.1038/nchembio.1751] [PubMed:23028358].

[2] R. Krafczyk, J. Macošek, P. K. A. Jagtap, D. Gast, S. Wunder, P. Mitra, A. K. Jha, J. Rohr, A. Hoffmann-Roder, K. Jung, J. Hennig, and J. Lassak. Structural Basis for EarP-Mediated Arginine Glycosylation of Translation Elongation Factor EF-P. *MBio*, 8(5), 09 2017. [PubMed Central:PMC5615199] [DOI:10.1128/mBio.01412-17] [PubMed:24025841].

[3] T. Sengoku, T. Suzuki, N. Dohmae, C. Watanabe, T. Honma, Y. Hikida, Y. Yamaguchi, H. Takahashi, S. Yokoyama, and T. Yanagisawa. Structural basis of protein arginine rhamnosylation by glycosyltransferase EarP. *Nat. Chem. Biol.*, 14(4):368–374, 04 2018. [DOI:10.1038/s41589-018-0002-y] [PubMed:29440735].

[4] C. He, N. Liu, F. Li, X. Jia, H. Peng, Y. Liu, and Y. Xiao. Complex Structure of Pseudomonas aeruginosa Arginine Rhamnosyltransferase EarP with Its Acceptor Elongation Factor P. *J. Bacteriol.*, 201(13), Jul 2019. [PubMed Central:PMC6560138] [DOI:10.1128/JB.00128-19] [PubMed:31010899].

[5] R. M. Arvola, C. A. Weidmann, T. M. Tanaka Hall, and A. C. Goldstrohm. Combinatorial control of messenger RNAs by Pumilio, Nanos and Brain Tumor Proteins. *RNA Biol*, 14(11):1445–1456, 11 2017. [PubMed Central:PMC5785226] [DOI:10.1080/15476286.2017.1306168] [PubMed:16537387].

[6] C. A. Weidmann, C. Qiu, R. M. Arvola, T. F. Lou, J. Killingsworth, Z. T. Campbell, T. M. Tanaka Hall, and A. C. Goldstrohm. Drosophila Nanos acts as a molecular clamp that modulates the RNA-binding and repression activities of Pumilio. *Elife*, 5, 08 2016. [PubMed Central:PMC4995099] [DOI:10.7554/eLife.17096] [PubMed:24997599].

[7] I. Loedige, L. Jakob, T. Treiber, D. Ray, M. Stotz, N. Treiber, J. Hennig, K. B. Cook, Q. Morris, T. R. Hughes, J. C. Engelmann, M. P. Krahn, and G. Meister. The Crystal Structure of the NHL Domain in Complex with RNA Reveals

the Molecular Basis of Drosophila Brain-Tumor-Mediated Gene Regulation. *Cell Rep*, 13(6):1206–1220, Nov 2015. [DOI:10.1016/j.celrep.2015.09.068] [PubMed:26527002].

[8] F. Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, Aug 1970. [DOI:10.1038/227561a0] [PubMed:4913914].

[9] T. R. Gregory. Synergy between sequence and size in large-scale genomics. *Nat. Rev. Genet.*, 6(9):699–708, Sep 2005. [DOI:10.1038/nrg1674] [PubMed:16151375].

[10] Z. Wang, X. Sun, Y. Zhao, X. Guo, H. Jiang, H. Li, and Z. Gu. Evolution of gene regulation during transcription and translation. *Genome Biol Evol*, 7(4):1155–1167, Apr 2015. [PubMed Central:PMC4419805] [DOI:10.1093/gbe/evv059] [PubMed:339816].

[11] J. W. Hershey, N. Sonenberg, and M. B. Mathews. Principles of translational control: an overview. *Cold Spring Harb Perspect Biol*, 4(12), Dec 2012. [PubMed Central:PMC3504442] [DOI:10.1101/cshperspect.a011528] [PubMed:22767671].

[12] Geoffrey Cooper. *The Cell: A Molecular Approach*. Sinauer Associates Inc, 2000. ISBN 9780878931064.

[13] R. J. Jackson, C. U. Hellen, and T. V. Pestova. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.*, 11(2):113–127, Feb 2010. [PubMed Central:PMC4461372] [DOI:10.1038/nrm2838] [PubMed:7610048].

[14] N. Sonenberg and A. G. Hinnebusch. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, 136(4):731–745, Feb 2009. [PubMed Central:PMC3610329] [DOI:10.1016/j.cell.2009.01.042] [PubMed:19300492].

[15] K. Ito, S. Chiba, and K. Pogliano. Divergent stalling sequences sense and control cellular physiology. *Biochem. Biophys. Res. Commun.*, 393(1):1–5, Feb 2010. [PubMed Central:PMC2885147] [DOI:10.1016/j.bbrc.2010.01.073] [PubMed:18809677].

[16] J. D. Dinman. Mechanisms and implications of programmed translational frameshifting. *Wiley Interdiscip Rev RNA*, 3(5):661–673, 2012. [PubMed Central:PMC3419312] [DOI:10.1002/wrna.1126] [PubMed:21109664].

[17] M. Gale, S. L. Tan, and M. G. Katze. Translational control of viral gene expression in eukaryotes. *Microbiol. Mol. Biol. Rev.*, 64(2):239–280, Jun 2000. [PubMed Central:PMC98994] [DOI:10.1128/mmbr.64.2.239-280.2000] [PubMed:9211946].

[18] M. Y. Pavlov, R. E. Watts, Z. Tan, V. W. Cornish, M. Ehrenberg, and A. C. Forster. Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proc. Natl. Acad. Sci. U.S.A.*, 106(1):50–54, Jan 2009. [PubMed Central:PMC2629218] [DOI:10.1073/pnas.0809211106] [PubMed:11955004].

[19] H. Beier and M. Grimm. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Res.*, 29(23):4767–4782, Dec 2001. [PubMed Central:PMC96686] [DOI:10.1093/nar/29.23.4767] [PubMed:3894360].

[20] M. Dabrowski, Z. Bukowy-Bieryllo, and E. Zietkiewicz. Translational readthrough potential of natural termination codons in eucaryotes–The impact of RNA sequence. *RNA Biol*, 12(9):950–958, 2015. [PubMed Central:PMC4615788] [DOI:10.1080/15476286.2015.1068497] [PubMed:11237617].

[21] S. Varenne, J. Buc, R. Lloubes, and C. Lazdunski. Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J. Mol. Biol.*, 180(3):549–576, Dec 1984. [DOI:10.1016/0022-2836(84)90027-5] [PubMed:6084718].

[22] T. Ikemura. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. *J. Mol. Biol.*, 151(3):389–409, Sep 1981. [DOI:10.1016/0022-2836(81)90003-6] [PubMed:6175758].

[23] G. F. Chen and M. Inouye. Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the Escherichia coli genes. *Nucleic Acids Res.*, 18(6):1465–1473, Mar 1990. [PubMed Central:PMC330513] [DOI:10.1093/nar/18.6.1465] [PubMed:271968].

[24] E. Angov. Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnol J*, 6(6):650–659, Jun 2011. [PubMed Central:PMC3166658] [DOI:10.1002/biot.201000332] [PubMed:7037777].

[25] N. T. Ingolia, S. Ghaemmaghami, J. R. Newman, and J. S. Weissman. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, 324(5924):218–223, Apr 2009. [PubMed Central:PMC2746483] [DOI:10.1126/science.1168978] [PubMed:10622731].

[26] A. R. Buskirk and R. Green. Ribosome pausing, arrest and rescue in bacteria and eukaryotes. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 372(1716), 03 2017. [PubMed Central:PMC5311927] [DOI:10.1098/rstb.2016.0183] [PubMed:9233821].

[27] F. F. Chevance, S. Le Guyon, and K. T. Hughes. The effects of codon context on in vivo translation speed. *PLoS Genet.*, 10(6):e1004392, Jun 2014. [PubMed Central:PMC4046918] [DOI:10.1371/journal.pgen.1004392] [PubMed:17095544].

[28] D. R. Tanner, D. A. Cariello, C. J. Woolstenhulme, M. A. Broadbent, and A. R. Buskirk. Genetic identification of nascent peptides that induce ribosome stalling. *J. Biol. Chem.*, 284(50):34809–34818, Dec 2009. [PubMed Central:PMC2787343] [DOI:10.1074/jbc.M109.039040] [PubMed:19394297].

[29] A. A. Morgan and E. Rubenstein. Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome. *PLoS ONE*, 8(1):e53785, 2013. [PubMed Central:PMC3556072] [DOI:10.1371/journal.pone.0053785] [PubMed:10771706].

[30] A. L. Starosta, J. Lassak, L. Peil, G. C. Atkinson, C. J. Woolstenhulme, K. Virumae, A. Buskirk, T. Tenson, J. Remme, K. Jung, and D. N. Wilson. A conserved proline triplet in Val-tRNA synthetase and the origin of elongation factor P. *Cell Rep*, 9(2):476–483, Oct 2014. [PubMed Central:PMC4847715] [DOI:10.1016/j.celrep.2014.09.008] [PubMed:24003132].

[31] L. K. Doerfel, I. Wohlgemuth, C. Kothe, F. Peske, H. Urlaub, and M. V. Rodnina. EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. *Science*, 339(6115):85–88, Jan 2013. [DOI:10.1126/science.1229017] [PubMed:23239624].

[32] S. Ude, J. Lassak, A. L. Starosta, T. Kraxenberger, D. N. Wilson, and K. Jung. Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science*, 339(6115):82–85, Jan 2013. [DOI:10.1126/science.1228985] [PubMed:23239623].

[33] E. Gutierrez, B. S. Shin, C. J. Woolstenhulme, J. R. Kim, P. Saini, A. R. Buskirk, and T. E. Dever. eIF5A promotes translation of polyproline motifs. *Mol. Cell*, 51(1):35–45, Jul 2013. [PubMed Central:PMC3744875] [DOI:10.1016/j.molcel.2013.04.021] [PubMed:15213400].

[34] L. Peil, A. L. Starosta, J. Lassak, G. C. Atkinson, K. Virumae, M. Spitzer, T. Tenson, K. Jung, J. Remme, and D. N. Wilson. Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. *Proc. Natl. Acad. Sci. U.S.A.*, 110(38):15265–15270, Sep 2013. [PubMed Central:PMC3780873] [DOI:10.1073/pnas.1310642110] [PubMed:4865540].

[35] T. M. Schmeing and V. Ramakrishnan. What recent ribosome structures have revealed about the mechanism of translation. *Nature*, 461(7268):1234–1242, Oct 2009. [DOI:10.1038/nature08403] [PubMed:15317937].

[36] P. Huter, S. Arenz, L. V. Bock, M. Graf, J. O. Frister, A. Heuer, L. Peil, A. L. Starosta, I. Wohlgemuth, F. Peske, J. Nova?ek, O. Berninghausen, H. Grubmuller, T. Tenson, R. Beckmann, M. V. Rodnina, A. C. Vaiana, and D. N.

Wilson. Structural Basis for Polyproline-Mediated Ribosome Stalling and Rescue by the Translation Elongation Factor EF-P. *Mol. Cell*, 68(3):515–527, Nov 2017. [DOI:10.1016/j.molcel.2017.10.014] [PubMed:29100052].

[37] C. Schmidt, T. Becker, A. Heuer, K. Braunger, V. Shanmuganathan, M. Pech, O. Berninghausen, D. N. Wilson, and R. Beckmann. Structure of the hypusinylated eukaryotic translation factor eIF-5A bound to the ribosome. *Nucleic Acids Res.*, 44(4):1944–1951, Feb 2016. [PubMed Central:PMC4770232] [DOI:10.1093/nar/gkv1517] [PubMed:17245413].

[38] S. Melnikov, J. Mailliot, B. S. Shin, L. Rigger, G. Yusupova, R. Micura, T. E. Dever, and M. Yusupov. Crystal Structure of Hypusine-Containing Translation Factor eIF5A Bound to a Rotated Eukaryotic Ribosome. *J. Mol. Biol.*, 428(18):3570–3576, 09 2016. [PubMed Central:PMC5408928] [DOI:10.1016/j.jmb.2016.05.011] [PubMed:23239624].

[39] M. H. Park. The post-translational synthesis of a polyamine-derived amino acid, hypusine, in the eukaryotic translation initiation factor 5A (eIF5A). *J. Biochem.*, 139(2):161–169, Feb 2006. [PubMed Central:PMC2494880] [DOI:10.1093/jb/mvj034] [PubMed:4334286].

[40] J. H. Park, H. E. Johansson, H. Aoki, B. X. Huang, H. Y. Kim, M. C. Ganoza, and M. H. Park. Post-translational modification by β-lysylation is required for activity of Escherichia coli elongation factor P (EF-P). *J. Biol. Chem.*, 287(4):2579–2590, Jan 2012. [PubMed Central:PMC3268417] [DOI:10.1074/jbc.M111.309633] [PubMed:10987361].

[41] A. Rajkovic, K. R. Hummels, A. Witzky, S. Erickson, P. R. Gafken, J. P. Whitelegge, K. F. Faull, D. B. Kearns, and M. Ibba. Translation Control of Swarming Proficiency in Bacillus subtilis by 5-Amino-pentanolylated Elongation Factor P. *J. Biol. Chem.*, 291(21):10976–10985, May 2016. [PubMed Central:PMC4900249] [DOI:10.1074/jbc.M115.712091] [PubMed:4214946].

[42] S. Choi and J. Choe. Crystal structure of elongation factor P from Pseudomonas aeruginosa at 1.75 Å resolution. *Proteins*, 79(5):1688–1693, May 2011. [DOI:10.1002/prot.22992] [PubMed:21365687].

[43] K. Hanawa-Suetsugu, S. Sekine, H. Sakai, C. Hori-Takemoto, T. Terada, S. Unzai, J. R. Tame, S. Kuramitsu, M. Shirouzu, and S. Yokoyama. Crystal structure of elongation factor P from Thermus thermophilus HB8. *Proc. Natl. Acad. Sci. U.S.A.*, 101(26):9595–9600, Jun 2004. [PubMed Central:PMC470720] [DOI:10.1073/pnas.0308667101] [PubMed:7491491].

[44] M. C. Franklin, J. Cheung, M. J. Rudolph, F. Burshteyn, M. Cassidy, E. Gary, B. Hillerich, Z. K. Yao, P. R. Carlier, M. Totrov, and J. D. Love. Structural genomics for drug design against the pathogen Coxiella burnetii. *Proteins*, 83(12):2124–2136, Dec 2015. [DOI:10.1002/prot.24841] [PubMed:26033498].

[45] M. H. Park and E. C. Wolff. Hypusine, a polyamine-derived amino acid critical for eukaryotic translation. *J. Biol. Chem.*, 293(48):18710–18718, 11 2018. [PubMed Central:PMC6290153] [DOI:10.1074/jbc.TM118.003341] [PubMed:9430712].

[46] E. Behshad, F. J. Ruzicka, S. O. Mansoorabadi, D. Chen, G. H. Reed, and P. A. Frey. Enantiomeric free radicals and enzymatic control of stereochemistry in a radical mechanism: the case of lysine 2,3-aminomutases. *Biochemistry*, 45(42):12639–12646, Oct 2006. [PubMed Central:PMC2553251] [DOI:10.1021/bi061328t] [PubMed:6168630].

[47] T. Yanagisawa, T. Sumida, R. Ishii, C. Takemoto, and S. Yokoyama. A paralog of lysyl-tRNA synthetase aminoacylates a conserved lysine residue in translation elongation factor P. *Nat. Struct. Mol. Biol.*, 17(9):1136–1143, Sep 2010. [DOI:10.1038/nsmb.1889] [PubMed:8654381].

[48] H. Roy, S. B. Zou, T. J. Bullwinkle, B. S. Wolfe, M. S. Gilreath, C. J. Forsyth, W. W. Navarre, and M. Ibba. The tRNA synthetase paralog PoxA modifies elongation factor-P with (R)-β-lysine. *Nat. Chem. Biol.*, 7(10):667–669, Aug 2011. [PubMed Central:PMC3177975] [DOI:10.1038/nchembio.632] [PubMed:10966471].

[49] L. Peil, A. L. Starosta, K. Virumae, G. C. Atkinson, T. Tenson, J. Remme, and D. N. Wilson. Lys34 of translation elongation factor EF-P is hydroxylated by YfcM. *Nat. Chem. Biol.*, 8(8):695–697, Aug 2012. [DOI:10.1038/nchembio.1001] [PubMed:17896105].

[50] D. G. Singh, J. Lomako, W. M. Lomako, W. J. Whelan, H. E. Meyer, M. Serwe, and J. W. Metzger. beta-Glucosylarginine: a new glucose-protein bond in a self-glucosylating protein from sweet corn. *FEBS Lett.*, 376(1-2):61–64, Nov 1995. [DOI:10.1016/0014-5793(95)01247-6] [PubMed:8521968].

[51] J. S. Pearson, C. Giogha, S. Y. Ong, C. L. Kennedy, M. Kelly, K. S. Robinson, T. W. Lung, A. Mansell, P. Riedmaier, C. V. Oates, A. Zaid, S. Muhlen, V. F. Crepin, O. Marches, C. S. Ang, N. A. Williamson, L. A. O'Reilly, A. Bankovacki, U. Nachbur, G. Infusini, A. I. Webb, J. Silke, A. Strasser, G. Frankel, and E. L. Hartland. A type III effector antagonizes death receptor signalling during bacterial gut infection. *Nature*, 501(7466):247–251, Sep 2013. [PubMed Central:PMC3836246] [DOI:10.1038/nature12524] [PubMed:11206551].

[52] Michael Sattler, Jürgen Schleucher, and Christian Griesinger. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 34(2):93–158, March 1999. ISSN 0079-6565. URL `http://www.journals.elsevierhealth.com/periodicals/jpnmrs/article/S0079-6565(98)00025-9/fulltext`.

[53] L. L. Lairson, B. Henrissat, G. J. Davies, and S. G. Withers. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.*, 77:521–555, 2008. [DOI:10.1146/annurev.biochem.76.061005.092322] [PubMed:18518825].

[54] William A. Bubb. NMR spectroscopy in the study of carbohydrates: Characterizing the structural complexity. *Concepts in Magnetic Resonance*, 19A(1): 1–19, 2003. doi: 10.1002/cmr.a.10080. URL https://doi.org/10.1002/cmr.a.10080.

[55] X. Li, R. Krafczyk, J. Macošek, Y. L. Li, Y. Zou, B. Simon, X. Pan, Q. Y. Wu, F. Yan, S. Li, J. Hennig, K. Jung, J. Lassak, and H. G. Hu. Resolving the α-glycosidic linkage of arginine-rhamnosylated translation elongation factor P triggers generation of the first ArgRha specific antibody. *Chem Sci*, 7(12):6995–7001, Dec 2016. [PubMed Central:PMC5363779] [DOI:10.1039/c6sc02889f] [PubMed:24025841].

[56] S. Wang, L. Corcilius, P. P. Sharp, A. Rajkovic, M. Ibba, B. L. Parker, and R. J. Payne. Synthesis of rhamnosylated arginine glycopeptides and determination of the glycosidic linkage in bacterial elongation factor P. *Chem Sci*, 8(3):2296–2302, Mar 2017. [PubMed Central:PMC5363394] [DOI:10.1039/c6sc03847f] [PubMed:24025841].

[57] Y. Shen, F. Delaglio, G. Cornilescu, and A. Bax. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, 44(4):213–223, Aug 2009. [PubMed Central:PMC2726990] [DOI:10.1007/s10858-009-9333-z] [PubMed:1737021].

[58] C. Breton, S. Fournel-Gigleux, and M. M. Palcic. Recent structures, evolution and mechanisms of glycosyltransferases. *Curr. Opin. Struct. Biol.*, 22(5):540–549, Oct 2012. [DOI:10.1016/j.sbi.2012.06.007] [PubMed:22819665].

[59] M. G. Olson, M. Goldammer, E. Gauliard, D. Ladant, and S. P. Ouellette. A Bacterial Adenylate Cyclase-Based Two-Hybrid System Compatible with Gateway® Cloning. *Methods Mol. Biol.*, 1794:75–96, 2018. [DOI:10.1007/978-1-4939-7871-7_6] [PubMed:29855952].

[60] W. Volkwein, R. Krafczyk, P. K. A. Jagtap, M. Parr, E. Mankina, J. Macošek, Z. Guo, M. J. L. J. Furst, M. Pfab, D. Frishman, J. Hennig, K. Jung, and J. Lassak. Switching the Post-translational Modification of Translation Elongation Factor EF-P. *Front Microbiol*, 10:1148, 2019. [PubMed Central:PMC6544042] [DOI:10.3389/fmicb.2019.01148] [PubMed:31178848].

[61] Lewis Kay, Paul Keifer, and Tim Saarinen. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *Journal of the American Chemical Society*, 114(26):10663–10665, 1992. doi: 10.1021/ja00052a088. URL https://doi.org/10.1021/ja00052a088.

[62] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, 6(3):277–293, Nov 1995. [PubMed:8142349].

[63] K. Pervushin, R. Riek, G. Wider, and K. Wuthrich. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U.S.A.*, 94(23):12366–12371, Nov 1997. [PubMed Central:PMC24947] [DOI:10.1073/pnas.94.23.12366] [PubMed:2819934].

[64] M. Salzmann, K. Pervushin, G. Wider, H. Senn, and K. Wuthrich. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 95(23):13585–13590, Nov 1998. [PubMed Central:PMC24862] [DOI:10.1073/pnas.95.23.13585] [PubMed:7077675].

[65] F. A. Mulder, D. Schipper, R. Bott, and R. Boelens. Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins. *J. Mol. Biol.*, 292(1):111–123, Sep 1999. [DOI:10.1006/jmbi.1999.3034] [PubMed:10493861].

[66] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, 59(4): 687–696, Jun 2005. [DOI:10.1002/prot.20449] [PubMed:15815974].

[67] P. Skubak and N. S. Pannu. Automatic protein structure solution from weak X-ray data. *Nat Commun*, 4:2777, 2013. [PubMed Central:PMC3868232] [DOI:10.1038/ncomms3777] [PubMed:21460451].

[68] R. A. de Graaff, M. Hilge, J. L. van der Plas, and J. P. Abrahams. Matrix methods for solving protein substructures of chlorine and sulfur from anomalous data. *Acta Crystallogr. D Biol. Crystallogr.*, 57(Pt 12):1857–1862, Dec 2001. [DOI:10.1107/s0907444901016535] [PubMed:11717499].

[69] J. P. Abrahams and A. G. Leslie. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D Biol. Crystallogr.*, 52(Pt 1):30–42, Jan 1996. [DOI:10.1107/S0907444995008754] [PubMed:15299723].

[70] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, 66(Pt 2):213–221, Feb 2010. [PubMed Central:PMC2815670] [DOI:10.1107/S0907444909052925] [PubMed:10745011].

[71] P. Emsley and K. Cowtan. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, 60(Pt 12 Pt 1):2126–2132, Dec 2004. [DOI:10.1107/S0907444904019158] [PubMed:15572765].

[72] P. Skubak, G. N. Murshudov, and N. S. Pannu. Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr. D Biol. Crystallogr.*, 60(Pt 12 Pt 1):2196–2201, Dec 2004. [DOI:10.1107/S0907444904019079] [PubMed:15572772].

[73] Scott F. Gilbert. *Developmental Biology*. Sinauer Associates, 2000. ISBN 0878932437. URL `https://www.amazon.com/Developmental-Biology-Scott-F-Gilbert/dp/0878932437?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0878932437`.

[74] O. Johnstone and P. Lasko. Translational regulation and RNA localization in Drosophila oocytes and embryos. *Annu. Rev. Genet.*, 35:365–406, 2001. [DOI:10.1146/annurev.genet.35.102401.090756] [PubMed:11700288].

[75] C. Nusslein-Volhard, H. G. Frohnhofer, and R. Lehmann. Determination of anteroposterior polarity in Drosophila. *Science*, 238(4834):1675–1681, Dec 1987. [DOI:10.1126/science.3686007] [PubMed:3686007].

[76] T. Schupbach and E. Wieschaus. Maternal-effect mutations altering the anterior-posterior pattern of the Drosophila embryo. *Rouxs Arch. Dev. Biol.*, 195(5):302–317, Jul 1986. [DOI:10.1007/BF00376063] [PubMed:6766208].

[77] T. Berleth, M. Burri, G. Thoma, D. Bopp, S. Richstein, G. Frigerio, M. Noll, and C. Nusslein-Volhard. The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. *EMBO J.*, 7(6):1749–1756, Jun 1988. [PubMed Central:PMC457163] [PubMed:16453608].

[78] D. St Johnston, W. Driever, T. Berleth, S. Richstein, and C. Nusslein-Volhard. Multiple steps in the localization of bicoid RNA to the anterior pole of the Drosophila oocyte. *Development*, 107 Suppl:13–19, 1989. [PubMed:2483989].

[79] A. Spirov, K. Fahmy, M. Schneider, E. Frei, M. Noll, and S. Baumgartner. Formation of the bicoid morphogen gradient: an mRNA gradient dictates the protein gradient. *Development*, 136(4):605–614, Feb 2009. [PubMed Central:PMC2685955] [DOI:10.1242/dev.031195] [PubMed:17632061].

[80] S. C. Little, G. Tka?ik, T. B. Kneeland, E. F. Wieschaus, and T. Gregor. The formation of the Bicoid morphogen gradient requires protein movement from anteriorly localized mRNA. *PLoS Biol.*, 9(3):e1000596, Mar 2011. [PubMed Central:PMC3046954] [DOI:10.1371/journal.pbio.1000596] [PubMed:17632061].

[81] Hans Georg Frohnhöfer and Christiane Nüsslein-Volhard. Organization of anterior pattern in the drosophila embryo by the maternal gene bicoid. *Nature*, 324(6093):120–125, November 1986. doi: 10.1038/324120a0. URL https://doi.org/10.1038/324120a0.

[82] W. Driever and C. Nusslein-Volhard. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell*, 54(1): 95–104, Jul 1988. [DOI:10.1016/0092-8674(88)90183-3] [PubMed:3383245].

[83] W. Driever, V. Siegel, and C. Nusslein-Volhard. Autonomous determination of anterior structures in the early Drosophila embryo by the bicoid morphogen. *Development*, 109(4):811–820, Aug 1990. [PubMed:2226200].

[84] C. Wang and R. Lehmann. Nanos is the localized posterior determinant in Drosophila. *Cell*, 66(4):637–647, Aug 1991. [DOI:10.1016/0092-8674(91)90110-k] [PubMed:1908748].

[85] R. Lehmann and C. Nusslein-Volhard. The maternal gene nanos has a central role in posterior pattern formation of the Drosophila embryo. *Development*, 112(3):679–691, Jul 1991. [PubMed:1935684].

[86] E. R. Gavis and R. Lehmann. Localization of nanos RNA controls embryonic polarity. *Cell*, 71(2):301–313, Oct 1992. [DOI:10.1016/0092-8674(92)90358-j] [PubMed:1423595].

[87] C. Schroder, D. Tautz, E. Seifert, and H. Jackle. Differential regulation of the two transcripts from the Drosophila gap segmentation gene hunchback. *EMBO J.*, 7(9):2881–2887, Sep 1988. [PubMed Central:PMC457082] [PubMed:6302193].

[88] M. Mlodzik, A. Fjose, and W. J. Gehring. Isolation of caudal, a Drosophila homeo box-containing gene with maternal expression, whose transcripts form a concentration gradient at the pre-blastoderm stage. *EMBO J.*, 4(11): 2961–2969, Nov 1985. [PubMed Central:PMC554605] [PubMed:6330741].

[89] W. Driever and C. Nusslein-Volhard. A gradient of bicoid protein in Drosophila embryos. *Cell*, 54(1):83–93, Jul 1988. [DOI:10.1016/0092-8674(88)90182-1] [PubMed:3383244].

[90] C. A. Smibert, J. E. Wilson, K. Kerr, and P. M. Macdonald. smaug protein represses translation of unlocalized nanos mRNA in the Drosophila embryo. *Genes Dev.*, 10(20):2600–2609, Oct 1996. [DOI:10.1101/gad.10.20.2600] [PubMed:8895661].

[91] A. Dahanukar and R. P. Wharton. The Nanos gradient in Drosophila embryos is generated by translational regulation. *Genes Dev.*, 10(20):2610–2620, Oct 1996. [DOI:10.1101/gad.10.20.2610] [PubMed:8895662].

[92] S. Andrews, D. R. Snowflack, I. E. Clark, and E. R. Gavis. Multiple mechanisms collaborate to repress nanos translation in the Drosophila ovary and embryo. *RNA*, 17(5):967–977, May 2011. [PubMed Central:PMC3078745] [DOI:10.1261/rna.2478611] [PubMed:10882131].

[93] J. Dubnau and G. Struhl. RNA recognition and translational regulation by a homeodomain protein. *Nature*, 379(6567):694–699, Feb 1996. [DOI:10.1038/379694a0] [PubMed:8602214].

[94] R. Rivera-Pomar, D. Niessing, U. Schmidt-Ott, W. J. Gehring, and H. Jackle. RNA binding and translational suppression by bicoid. *Nature*, 379(6567): 746–749, Feb 1996. [DOI:10.1038/379746a0] [PubMed:8602224].

[95] V. Irish, R. Lehmann, and M. Akam. The Drosophila posterior-group gene nanos functions by repressing hunchback activity. *Nature*, 338(6217):646–648, Apr 1989. [DOI:10.1038/338646a0] [PubMed:2704419].

[96] R. P. Wharton and G. Struhl. RNA regulatory elements mediate control of Drosophila body pattern by the posterior morphogen nanos. *Cell*, 67(5): 955–967, Nov 1991. [DOI:10.1016/0092-8674(91)90368-9] [PubMed:1720354].

[97] M. Hulskamp, C. Schroder, C. Pfeifle, H. Jackle, and D. Tautz. Posterior segmentation of the Drosophila embryo in the absence of a maternal posterior organizer gene. *Nature*, 338(6217):629–632, Apr 1989. [DOI:10.1038/338629a0] [PubMed:2704418].

[98] D. Tautz. Regulation of the Drosophila segmentation gene hunchback by two maternal morphogenetic centres. *Nature*, 332(6161):281–284, Mar 1988. [DOI:10.1038/332281a0] [PubMed:2450283].

[99] W. Driever and C. Nusslein-Volhard. The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo. *Nature*, 337 (6203):138–143, Jan 1989. [DOI:10.1038/337138a0] [PubMed:2911348].

[100] P. M. Macdonald and G. Struhl. A molecular gradient in early Drosophila embryos and its role in specifying the body pattern. *Nature*, 324(6097):537–545, 1986. [DOI:10.1038/324537a0] [PubMed:2878369].

[101] M. Mlodzik and W. J. Gehring. Expression of the caudal gene in the germ line of Drosophila: formation of an RNA and protein gradient during early embryogenesis. *Cell*, 48(3):465–478, Feb 1987. [DOI:10.1016/0092-8674(87)90197-8] [PubMed:2433048].

[102] C. Nusslein-Volhard and E. Wieschaus. Mutations affecting segment number and polarity in Drosophila. *Nature*, 287(5785):795–801, Oct 1980. [DOI:10.1038/287795a0] [PubMed:6776413].

[103] E. R. Gavis, L. Lunsford, S. E. Bergsten, and R. Lehmann. A conserved 90 nucleotide element mediates translational repression of nanos RNA. *Development*, 122(9):2791–2800, Sep 1996. [PubMed:8787753].

[104] M. Gotze, J. Dufourt, C. Ihling, C. Rammelt, S. Pierson, N. Sambrani, C. Temme, A. Sinz, M. Simonelig, and E. Wahle. Translational repression of the Drosophila nanos mRNA involves the RNA helicase Belle and RNA coating by Me31B and Trailer hitch. *RNA*, 23(10):1552–1568, 10 2017. [PubMed Central:PMC5602113] [DOI:10.1261/rna.062208.117] [PubMed:21642421].

[105] J. Sonoda and R. P. Wharton. Recruitment of Nanos to hunchback mRNA by Pumilio. *Genes Dev.*, 13(20):2704–2712, Oct 1999. [PubMed Central:PMC317116] [DOI:10.1101/gad.13.20.2704] [PubMed:10488336].

[106] J. Sonoda and R. P. Wharton. Drosophila Brain Tumor is a translational repressor. *Genes Dev.*, 15(6):762–773, Mar 2001. [PubMed Central:PMC312658] [DOI:10.1101/gad.870801] [PubMed:10949924].

[107] M. R. Nelson, A. M. Leidal, and C. A. Smibert. Drosophila Cup is an eIF4E-binding protein that functions in Smaug-mediated translational repression. *EMBO J.*, 23(1):150–159, Jan 2004. [PubMed Central:PMC1271664] [DOI:10.1038/sj.emboj.7600026] [PubMed:7954828].

[108] P. F. Cho, C. Gamberi, Y. A. Cho-Park, I. B. Cho-Park, P. Lasko, and N. Sonenberg. Cap-dependent translational inhibition establishes two opposing morphogen gradients in Drosophila embryos. *Curr. Biol.*, 16(20):2035–2041, Oct 2006. [PubMed Central:PMC2238800] [DOI:10.1016/j.cub.2006.08.093] [PubMed:10731138].

[109] P. F. Cho, F. Poulin, Y. A. Cho-Park, I. B. Cho-Park, J. D. Chicoine, P. Lasko, and N. Sonenberg. A new paradigm for translational control: inhibition via 5'-3' mRNA tethering by Bicoid and the eIF4E cognate 4EHP. *Cell*, 121(3): 411–423, May 2005. [DOI:10.1016/j.cell.2005.02.024] [PubMed:15882623].

[110] I. Abaza and F. Gebauer. Trading translation with RNA-binding proteins. *RNA*, 14(3):404–409, Mar 2008. [PubMed Central:PMC2248257] [DOI:10.1261/rna.848208] [PubMed:18212021].

[111] K. E. Baker and J. Coller. The many routes to regulating mRNA translation. *Genome Biol.*, 7(12):332, 2006. [PubMed Central:PMC1794424] [DOI:10.1186/gb-2006-7-12-332] [PubMed:17176455].

[112] T. Raisch, D. Bhandari, K. Sabath, S. Helms, E. Valkov, O. Weichenrieder, and E. Izaurralde. Distinct modes of recruitment of the CCR4-NOT complex by Drosophila and vertebrate Nanos. *EMBO J.*, 35(9):974–990, 05 2016. [PubMed Central:PMC5207322] [DOI:10.15252/embj.201593634] [PubMed:26527002].

[113] J. L. Semotok, R. L. Cooperstock, B. D. Pinder, H. K. Vari, H. D. Lipshitz, and C. A. Smibert. Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early Drosophila embryo. *Curr. Biol.*, 15(4):284–294, Feb 2005. [DOI:10.1016/j.cub.2005.01.048] [PubMed:15723788].

[114] G. Struhl, P. Johnston, and P. A. Lawrence. Control of Drosophila body pattern by the hunchback morphogen gradient. *Cell*, 69(2):237–249, Apr 1992. [DOI:10.1016/0092-8674(92)90405-2] [PubMed:1568245].

[115] R. Lehmann and C. Nusslein-Volhard. hunchback, a gene required for segmentation of an anterior and posterior region of the Drosophila embryo. *Dev. Biol.*, 119(2):402–417, Feb 1987. [DOI:10.1016/0012-1606(87)90045-5] [PubMed:3803711].

[116] D. Curtis, D. K. Treiber, F. Tao, P. D. Zamore, J. R. Williamson, and R. Lehmann. A CCHC metal-binding domain in Nanos is essential for translational regulation. *EMBO J.*, 16(4):834–843, Feb 1997. [PubMed Central:PMC1169684] [DOI:10.1093/emboj/16.4.834] [PubMed:8475383].

[117] H. Hashimoto, K. Hara, A. Hishiki, S. Kawaguchi, N. Shichijo, K. Nakamura, S. Unzai, Y. Tamaru, T. Shimizu, and M. Sato. Crystal structure of zinc-finger domain of Nanos and its functional implications. *EMBO Rep.*, 11(11):848–853, Nov 2010. [PubMed Central:PMC2966957] [DOI:10.1038/embor.2010.155] [PubMed:8614464].

[118] Y. Murata and R. P. Wharton. Binding of pumilio to maternal hunchback mRNA is required for posterior patterning in Drosophila embryos. *Cell*, 80(5):747–756, Mar 1995. [DOI:10.1016/0092-8674(95)90353-4] [PubMed:7889568].

[119] P. D. Zamore, J. R. Williamson, and R. Lehmann. The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. *RNA*, 3(12):1421–1433, Dec 1997. [PubMed Central:PMC1369583] [PubMed:9404893].

[120] P. D. Zamore, D. P. Bartel, R. Lehmann, and J. R. Williamson. The PUMILIO-RNA interaction: a single RNA-binding domain monomer recognizes a bipartite target sequence. *Biochemistry*, 38(2):596–604, Jan 1999. [DOI:10.1021/bi982264s] [PubMed:9888799].

[121] R. P. Wharton, J. Sonoda, T. Lee, M. Patterson, and Y. Murata. The Pumilio RNA-binding domain is also a translational regulator. *Mol. Cell*, 1(6):863–872, May 1998. [PubMed:9660969].

[122] Y. K. Gupta, T. H. Lee, T. A. Edwards, C. R. Escalante, L. Y. Kadyrova, R. P. Wharton, and A. K. Aggarwal. Co-occupancy of two Pumilio molecules on a single hunchback NRE. *RNA*, 15(6):1029–1035, Jun 2009. [PubMed Central:PMC2685510] [DOI:10.1261/rna.1327609] [PubMed:10653785].

[123] Ruth Lehmann and Christiane Nüsslein-Volhard. Involvement of the pumilio gene in the transport of an abdominal signal in the drosophila embryo. *Nature*, 329(6135):167–170, September 1987. doi: 10.1038/329167a0. URL `https://doi.org/10.1038/329167a0`.

[124] C. Temme, M. Simonelig, and E. Wahle. Deadenylation of mRNA by the CCR4-NOT complex in Drosophila: molecular and developmental aspects. *Front Genet*, 5:143, 2014. [PubMed Central:PMC4033318] [DOI:10.3389/fgene.2014.00143] [PubMed:15687258].

[125] C. A. Weidmann and A. C. Goldstrohm. Drosophila Pumilio protein contains multiple autonomous repression domains that regulate mRNAs independently of Nanos and brain tumor. *Mol. Cell. Biol.*, 32(2):527–540, Jan 2012. [PubMed Central:PMC3255780] [DOI:10.1128/MCB.06052-11] [PubMed:9404893].

[126] P. M. Macdonald. The Drosophila pumilio gene: an unusually long transcription unit and an unusual protein. *Development*, 114(1):221–232, Jan 1992. [PubMed:1576962].

[127] J. D. Laver, X. Li, D. Ray, K. B. Cook, N. A. Hahn, S. Nabeel-Shah, M. Kekis, H. Luo, A. J. Marsolais, K. Y. Fung, T. R. Hughes, J. T. Westwood, S. S. Sidhu, Q. Morris, H. D. Lipshitz, and C. A. Smibert. Brain tumor is a sequence-specific RNA-binding protein that directs maternal mRNA clearance during the Drosophila maternal-to-zygotic transition. *Genome Biol.*, 16:94, May 2015. [PubMed Central:PMC4460960] [DOI:10.1186/s13059-015-0659-4] [PubMed:21238926].

[128] I. Loedige, M. Stotz, S. Qamar, K. Kramer, J. Hennig, T. Schubert, P. Loffler, G. Langst, R. Merkl, H. Urlaub, and G. Meister. The NHL domain of BRAT is an RNA-binding domain that directly contacts the hunchback mRNA for regulation. *Genes Dev.*, 28(7):749–764, Apr 2014. [PubMed Central:PMC4015489] [DOI:10.1101/gad.236513.113] [PubMed:16537387].

[129] A. Bashirullah, S. R. Halsell, R. L. Cooperstock, M. Kloc, A. Karaiskakis, W. W. Fisher, W. Fu, J. K. Hamilton, L. D. Etkin, and H. D. Lipshitz. Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in Drosophila melanogaster. *EMBO J.*, 18(9):2610–2620, May 1999. [PubMed Central:PMC1171340] [DOI:10.1093/emboj/18.9.2610] [PubMed:2632239].

[130] L. M. Hellman and M. G. Fried. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc*, 2(8):1849–1861, 2007. [PubMed Central:PMC2757439] [DOI:10.1038/nprot.2007.249] [PubMed:3678204].

[131] D. S. Waugh. Crystal structures of MBP fusion proteins. *Protein Sci.*, 25(3):559–571, Mar 2016. [PubMed Central:PMC4815407] [DOI:10.1002/pro.2863] [PubMed:21625560].

[132] Sophie Winter. Structural characterization of *Drosophila* nanos in the context of *hunchback* mrna regulation. Master's thesis, Heidelberg University, 2017.

[133] C. Qiu, A. C. Goldstrohm, and T. M. Tanaka Hall. Preparation of cooperative RNA recognition complexes for crystallographic structural studies. *Meth. Enzymol.*, 623:1–22, 2019. [DOI:10.1016/bs.mie.2019.04.001] [PubMed:31239042].

[134] C. D. Putnam, M. Hammel, G. L. Hura, and J. A. Tainer. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.*, 40(3):191–285, Aug 2007. [DOI:10.1017/S0033583507004635] [PubMed:18078545].

[135] C. Liu, Z. Shan, J. Diao, W. Wen, and W. Wang. Crystal structure of the coiled-coil domain of Drosophila TRIM protein Brat. *Proteins*, 87(8):706–710, Aug 2019. [DOI:10.1002/prot.25691] [PubMed:30958583].

[136] A. M. Catanzariti, T. A. Soboleva, D. A. Jans, P. G. Board, and R. T. Baker. An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci.*, 13(5):1331–1339, May 2004. [PubMed Central:PMC2286746] [DOI:10.1110/ps.04618904] [PubMed:3018930].

[137] R. T. Baker, A. M. Catanzariti, Y. Karunasekara, T. A. Soboleva, R. Sharwood, S. Whitney, and P. G. Board. Using deubiquitylating enzymes as research tools. *Meth. Enzymol.*, 398:540–554, 2005. [DOI:10.1016/S0076-6879(05)98044-0] [PubMed:16275357].

[138] A. F. Moon, G. A. Mueller, X. Zhong, and L. C. Pedersen. A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. *Protein Sci.*, 19(5):901–913, May 2010. [PubMed Central:PMC2868234] [DOI:10.1002/pro.368] [PubMed:8729668].

[139] Petr V. Konarev, Vladimir V. Volkov, Anna V. Sokolova, Michel H. J. Koch, and Dmitri I. Svergun. *PRIMUS*: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, 36(5):1277–1282, Oct 2003. doi: 10.1107/S0021889803012779. URL `https://doi.org/10.1107/S0021889803012779`.

[140] D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, and D. I. Svergun. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr*, 50(Pt 4):1212–1225, Aug 2017. [PubMed Central:PMC5541357] [DOI:10.1107/S1600576717007786] [PubMed:27738419].

[141] S. Förster, L. Apostol, and W. Bras. *Scatter*: software for the analysis of nano- and mesoscale small-angle scattering. *Journal of Applied Crystallography*, 43 (3):639–646, Jun 2010. doi: 10.1107/S0021889810008289. URL `https://doi.org/10.1107/S0021889810008289`.

[142] A. Leitner, T. Walzthoeni, and R. Aebersold. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat Protoc*, 9(1): 120–137, Jan 2014. [DOI:10.1038/nprot.2013.168] [PubMed:20736410].

[143] A. T. Brunger. Version 1.2 of the Crystallography and NMR system. *Nat Protoc*, 2(11):2728–2733, 2007. [DOI:10.1038/nprot.2007.406] [PubMed:18007608].

[144] A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, 54(Pt 5):905–921, Sep 1998. [DOI:10.1107/s0907444998003254] [PubMed:9757107].

[145] X. Wang, J. McLachlan, P. D. Zamore, and T. M. Hall. Modular recognition of RNA by a human pumilio-homology domain. *Cell*, 110(4):501–512, Aug 2002. [DOI:10.1016/s0092-8674(02)00873-5] [PubMed:12202039].

[146] A. Lapinaite, B. Simon, L. Skjaerven, M. Rakwalska-Bange, F. Gabel, and T. Carlomagno. The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature*, 502(7472):519–523, Oct 2013. [DOI:10.1038/nature12581] [PubMed:20864039].

[147] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T. E. Malliavin, and M. Nilges. ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, 23(3):381–382, Feb 2007. [DOI:10.1093/bioinformatics/btl589] [PubMed:17121777].

[148] D. Svergun, C. Barberato, and M. H. J. Koch. *CRYSOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography*, 28(6):768–773, Dec 1995. doi: 10.1107/S0021889895007047. URL `https://doi.org/10.1107/S0021889895007047`.

[149] D. I. Svergun, S. Richard, M. H. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. U.S.A.*, 95(5):2267–2272, Mar 1998. [PubMed Central:PMC19315] [DOI:10.1073/pnas.95.5.2267] [PubMed:4856347].

[150] L. A. Feigin and D. I. Svergun. *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*. Plenum Press, 1987. ISBN 978-1-475-76624-0.