# Cryo-EM analysis of three macromolecular machines: RNA polymerase, ribosome, and E3 ubiquitin ligase

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Paul Ferdinand Krupp

2020

Berlin

This thesis is the result of my work from June 2016 until December 2019 at the Institute of Medical Physics and Biophysics at the Charité - Universitätsmedizin Berlin, under the direction of Prof. Dr. Christian M. T. Spahn.

1. Reviewer: Prof. Dr. Christian M. T. Spahn
2. Reviewer: Prof. Dr. Markus Wahl
Date of Defense: 12.2.2021

**Selbstständigkeitserklärung**

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbstständig und unter Zuhilfenahme der angegebenen Literatur erstellt habe.

Berlin, den _____

(Unterschrift)

**Acknowledgements**

I would like to thank my supervisor Prof. Dr. Christian Spahn for his support, encouragement and advice throughout my studies. I am grateful to him for providing me with the opportunity and the environment to perform excellent scientific work and engage in numerous informative discussions on image processing, macromolecular machines and research politics.

Additionally, I would like to express my gratitude to Prof. Dr. Markus Wahl for his continuous support as a collaborator and as the second reviewer of this thesis.

Furthermore, I am in debt to all members of the Institute of Medical Physics and Biophysics at the Charité - Universitätsmedizin Berlin between 2015 and 2020. A special thanks goes to my generation of graduate students within the vibrant co-working space: Annika Balke, Dennis Kwiatkowski, Nicolas Heyder, and David Speck.

Besides, I would like to specifically acknowledge Matthew Kraushar, David Schwefel, Rainer Nikolay, Magdalena Schacherl, Karin Chiluvane, and Justus Loerke for continuous collaboration and support.

I had the pleasure to collaborator on all matters of sample preparation, data collection and analysis with Jörg Bürger and Thorsten Mielke located at the MPI for Molecular Genetics.

Additionally, I would like to express my gratitude for proofreading my thesis to Magdalena Schacherl, Nelly Said, David Schwefel, Matt Kraushar, Sasa Petrovic, and Dominique Sydow.

I also owe my deepest gratitude to my ever-supporting family and friends. Moreover, I would like to specifically thank Dominique and our 4-legged companion Janosch for their in-house support.

“… aber mit Promotion!”
Gerda Krupp

# Table of Content

# Abstract

Cryo-electron microscopy (cryo-EM) together with single particle analysis (SPA) has emerged as a powerful tool to elucidate the 3D structure of macromolecular machines (MMMs) at atomic resolution. In this thesis, three projects are presented where cryo-EM was successfully employed as part of an integrative structural biology approach. The results of these projects provided insights into the molecular interactions and dynamics of the MMMs under investigation. All structural findings are discussed together with the corroborating results from collaboration partners.

(i) The λN-mediated transcription antitermination complex (λN-TAC) represents a well-studied example of a viral protein that hijacks the bacterial transcription apparatus. Here, λN induces transcriptional anti-termination in the context of specific stop signals within the RNA, leading to the transcription of viral genes downstream of the stop signal. The presented results shed light onto the mechanism by which the intrinsically disordered protein λN influences a large elongating complex consisting of RNA polymerase, several regulatory Nus protein factors, DNA, and RNA. A multi-pronged strategy is revealed by which λN traverses through the λN-TAC, establishing several favorable interactions to promote anti-termination.

(ii) Neuronal development is a tightly controlled process during which translational regulation at high spatiotemporal precision is required. *Ex vivo* derived ribosomal fractions from the mouse neocortex were purified at different time points during development. Within them, Ebp1 was found as a highly enriched ribosomal regulatory factor. Structural elucidation located Ebp1 at the nascent chain peptide tunnel exit, where it is involved in the translation of several neuron-specific proteins. Thereby, Ebp1 was discovered to be an important regulator of gene expression during neuronal development.

(iii) Mammalian cells counteract viral infection by the activity of numerous antiviral factors. One such antiviral factor, SAMHD1, decreases the NTP pool available for viral replication. Lentiviruses like HIV or SIV have evolved the accessory proteins Vpx/Vpr to counteract such antiviral factors. Vpx/Vpr both recruit SAMHD1 as a *neo*-substrate to the E3 ubiquitin ligase $CRL4^{DCAF1}$, leading to its ubiquitination and subsequent degradation. In this study, the Vpr-SAMHD1 binding interface was defined, allowing the formulation of a diverging mechanism for the Vpx/Vpr family. Furthermore, insights into the dynamics of the $CRL4^{DCAF1}$ are provided.

Moreover, preliminary work regarding the structural and biochemical characterization of Cdt2, a master regulator of genome stability is presented.

These studies combined have shown the power of an integrative structural biology approach that includes cryo-EM/SPA analyses to gain structural insights on a molecular level.

# Zusammenfassung

Die Kryo-Elektronenmikroskopie (Kryo-EM) in Verbindung mit der Einzelpartikelanalyse (EPA) ist eine etablierte mächtige Methode zur Strukturaufklärung von Makromolekularen Maschinen (MMM). In der vorliegenden Doktorarbeit werden drei verschiedene Projekte beschrieben, bei denen Kryo-EM als Teil eines integrativen Strukturbiologie-Ansatzes erfolgreich angewendet wurde. Die Ergebnisse dieser Projekte geben Aufschluss über molekulare Interaktionen und die Dynamik der verschiedenen MMMs. Die strukturellen Ergebnisse werden zusammen mit unterstützenden Resultaten von Kooperationspartner diskutiert.

(i) Der λN-gesteuerte transkriptionelle Anti-Terminationskomplex (TAK) ist ein gut erforschtes Beispiel für die virale Übernahme des bakteriellen Transkriptionsapparates. Dabei induziert λN transkriptionelle Anti-Termination im Kontext von bestimmten Stop-Signalen innerhalb der RNA. Das führt zur Transkription von viralen Genen strangabwärts des Stop-Signals. Die hier präsentierten Ergebnisse deuten auf den Mechanismus hin, mit dem das intrinsisch ungeordnete Protein λN den großen Elongationskomplex bestehend aus RNA-Polymerase, verschiedenen Nus-Proteinen, DNA und RNA beeinflusst. In einer mehrgleisigen Strategie schlängelt sich λN durch den λN-TAK und geht dabei mehrere vorteilhafte Interaktionen ein, um die Anti-Termination einzuleiten.

(ii) Die neuronale Entwicklung ist ein engmaschig kontrollierter Prozess, der eine hohe raumzeitliche Präzision der Translationskontrolle benötigt. Zu verschiedenen Zeitpunkten während der Entwicklung des Neocortex wurden ribosomale Fraktionen *ex vivo* aufgereinigt. Deren Analyse zeigte einer großen Menge von Ebp1, einem regulatorischen Faktor des Ribosoms. Dank der Strukturaufklärung wurde Ebp1 an dem ribosomalen Ausgangstunnel lokalisiert, wo es in die Proteinbiosynthese von verschiedenen neuron-spezifischen Proteinen involviert ist. Durch diese Ergebnisse wurde Ebp1 als ein wichtiger Regulator der Genexpression während der Neuronalentwicklung identifiziert.

(iii) Säugetierzellen erwehren sich gegen virale Angriffe durch eine Vielzahl antiviraler Faktoren. Einer dieser antiviralen Faktoren, SAMHD1, dezimiert den für die Virusreplikation nötigen NTP-Pool. Lentiviren wie HIV oder SIV haben die akzessorischen Proteine Vpr/Vpx entwickelt, um diesen antiviralen Faktoren entgegenzuwirken. Vpx/Vpr rekrutieren unter anderem SAMHD1 als *neo*-Substrat zur E3 Ubiquitin Ligase $CRL4^{DCAF1}$, was zur Ubiquitinierung und dem anschließenden Abbau von SAMHD1 führt. In dieser Studie, wurde die SAMHD1-Vpr Interaktion definiert, was Einblicke in den divergenten Mechanismus der Vpx/Vpr-Familie erlaubt. Außerdem wurden neue Erkenntnisse in die Dynamik des $CRL4^{DCAF1}$-Komplexes gewonnen.

Des Weiteren werden vorläufige Ergebnisse einer strukturellen und biochemischen Charakterisierung von Cdt2, einem Masterregulator für Genomstabilität, gezeigt.

Zusammengenommen stellen diese Studien die Leistungsfähigkeit eines integrativen Strukturbiologie-Ansatzes dar, bei dem Kryo-EM / EPA Einblicke auf molekularer Ebene gewährt.

# List of relevant publications

Said N.*, Krupp F., Anedchenko E., Santos K. F., Dybkov O., Huang Y.-H., Lee C.-T., Loll B., Behrmann E., Bürger J., Mielke T., Loerke J., Urlaub H., Spahn C. M. T., Weber G., Wahl, M. C. (2017) Structural basis for λN-dependent processive transcription antitermination, Nature Microbiology 2, 17062 (2017)

*FK screened samples for cryo-EM imaging conditions, collected and processed medium- and high-resolution electron microscopic data, analyzed results. All of my results were submitted in 2016 to the Faculty of Life Science at the Humboldt Universität zu Berlin as part of my master thesis entitled:* "Structural Investigation of two DNA-binding Complexes using Cryo-electron Microscopy." *Furthermore, FK prepared figures, parts of the manuscript and revision as part of his graduate studies.*

Krupp F.*, Said N.*, Huang Y.-H.*, Loll B., Bürger J., Mielke T., Spahn C. M. T., Wahl M. C. (2019) Structural Basis for the action of an all-purpose transcription anti-termination factor. Molecular Cell Volume 74, Issue 1, 4 April 2019, Pages 143-157.e5

*FK collected and processed medium- and high-resolution electron microscopic data, analyzed results, wrote part of the manuscript, and prepared figures and movies.*

Kraushar M. L.*, Krupp F., Harnett D., Turko P., Ambrozkiewicz M. C., Sprink T., Imami K., Günnigmann M., Zinnall U., Vieira-Vieira C.-H., Schaub T., Münster-Wandowski A., Bürger J., Borisova E., Yamamoto H., Rasin M.-R., Beule D., Landthaler M., Mielke T., Tarabykin V., Landthaler M., Kramer G., Vida I., Selbach M., and Spahn C. M. T.. The architecture of protein synthesis in the developing neocortex at near-atomic resolution reveals Ebp1-mediated neuronal proteostasis at the 60S tunnel exit. *Under review*

*FK assisted in sample screening, analyzed high-resolution electron microscopic data, interpreted results, prepared figures, and wrote parts of the manuscript.*

Banchenko S.*, Krupp F.*, Gotthold C., Bürger J., Graziadei A., O'Reilly F., Sinn L., Ruda O., Rappsilber J., Spahn C. M. T., Mielke T., Taylor I. and Schwefel D. Structural insights into Cullin4-RING ubiquitin ligase remodeling by Vpr from simian immunodeficiency viruses. *Under review*

*FK collected and processed medium- and high-resolution electron microscopic data, analyzed results, interpreted results, prepared figures and wrote parts of the manuscript.*

* denotes first author

# Comments on notation

Several different proteins, their domains and their interactions between them, and with nucleic acid elements will be described in the following. Therefore a consistent nomenclature is ued:

- Superscript symbolizes specific domains or residues of individual proteins, e.g. $NusA^{S1}$ or $NusA^{1-20}$
- *Italic font* denotes specific sites within RNA or DNA, e.g. *rut* binding site
- Δ represents deletion of domains from proteins, e.g. $NusA^{\Delta AR2}$
- Furthermore, superscript can also denote a specific E3 ubiquitin ligases belonging to the cullin family, e.g. $CRL4^{DCAF1}$ abbreviates the complex Cul4:Roc1:DDB1:DCAF1
- Hyphens designate interactions between elements, e.g. the RNAP-NusA interaction.
- Depending on the context, hyphens can also represent fusion proteins, e.g. T4L-SAMHD1
- Colons indicate formed complexes, e.g. the RNAP:NusA complex
- Additionally, Latin words are also marked in *italic font*, e.g. *in vitro* or *in vivo*

# 1 Introduction

## 1.1 Single particle cryo-electron microscopy

Cryo-electron microscopy (cryo-EM) in combination with single particle analysis (SPA) is a powerful method to determine the structure of macromolecular machines (MMMs). One experiment in cryo-EM consists of three sub-steps: sample preparation, data collection and image processing, and is often part of an iterative cycle (Figure 1.1) (1, 2). Sample preparation aims in preserving the structure of the MMM on an EM grid for its analysis in the vacuum of the electron microscope (EM). Data collection in the EM is performed at low dose conditions to avoid radiation damage of the sample in an automatic fashion, acquiring small (several hundred) to large (several thousand images) data sets. Each image displays 10s to 100s 2D projections of the three-dimensional MMM in random orientations. Image processing is applied to these images to obtain statistical information in 2D and 3D of the sample (3).

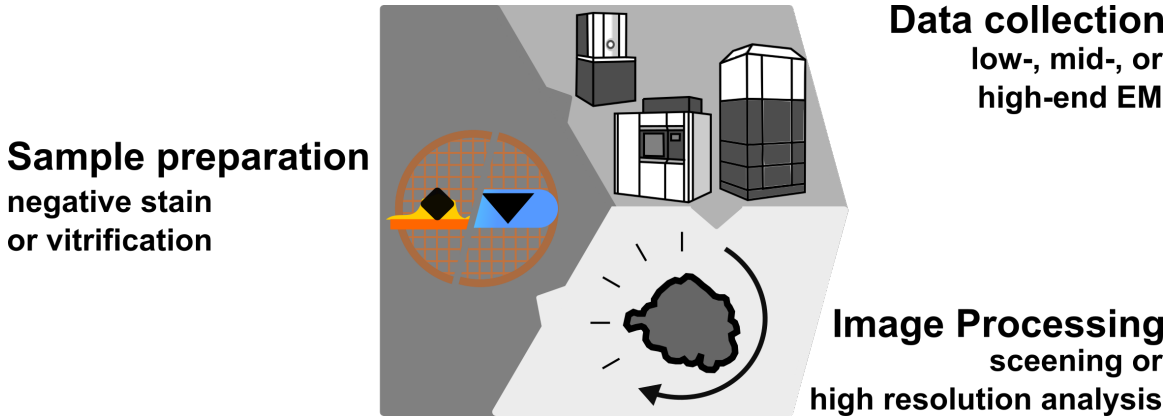



**Figure 1.1: Schematic depiction of one EM experiment.**
*The three sub-steps of an EM experiment and different options therein are depicted: sample preparation, data collection, and image processing. The cyclic nature is implied.*

In order to resolve the structure of an MMM at high resolution several such experiments with different conditions are performed in a five-step workflow (Figure 1.2): Biochemical sample purification, screening under negative staining and vitrification conditions, respectively, high resolution analysis, and interpretation. Once a sample has been proven to be of sufficient quality at one step, the next step is initiated. Thereby, an efficient use of the available resources is ensured. Depending of the biological sample and its behavior, such a linear workflow is rarely sufficient to answer the biological question. Therefore, several iterative cycles are often required in order to optimize the steps of biochemistry, sample preparation and data collection (arrows in Figure 1.2).

In the following, the fundamental concepts for each step of this workflow will be described. In the end, a table will summarize similarities and differences between the three steps of negative stain screening, vitrification screening and high resolution analysis.

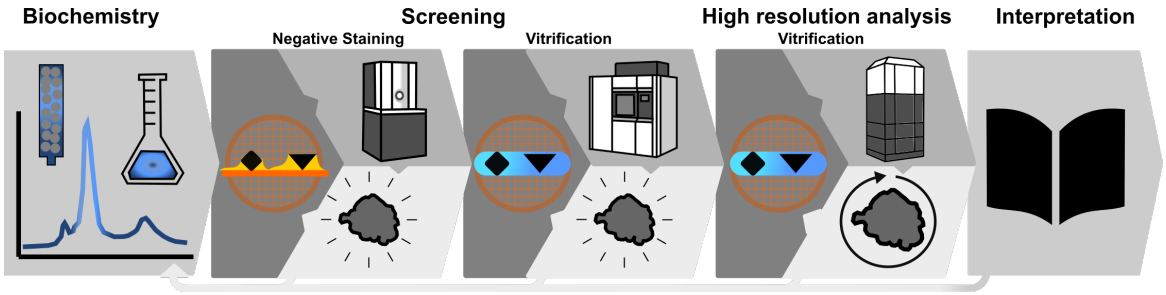



**Figure 1.2: Workflow of a cryo-EM SPA project.**
*Five-step workflow of a complete cryo-EM SPA project consisting of biochemistry, screening under negative stain and vitrification conditions, high resolution analysis, and interpretation. The iterative optimization starting from the step of Biochemistry is implied by arrows below.*

### 1.1.1 Biochemistry

The first and most important step provides the biological sample for all steps to come. In order to obtain sufficiently pure MMMs at a necessary concentration several approaches are possible. For example, a recombinant expression of the MMM from “gene to protein” can be performed in different organisms or systems (4). The MMM can be obtained by its assembly from individually purified components *in vitro* or via co-expression *in vivo*. Furthermore, the MMM can be purified *ex vivo* (5). Often times a disparity exist between an optimal sample in a biochemical and a cryo-EM sense. High homogeneity, e.g. a single peak in a gel filtration analysis, is no guarantee for well behavior during all subsequent steps. Due to these and other reasons, the step of “Biochemistry” is currently recognized as a major bottleneck in structural determination (6, 7).

### 1.1.2 Sample preparation

Sample preparation aims in fixating the biological sample for its insertion into high vacuum of the EM column and ensures its stability in the high radiation environment of the electron beam. Two preparation techniques are used for imaging of single particle MMMs via EM: negative staining and vitrification.

In both cases, the sample is applied to an EM grid, a disc of 3 mm in diameter, that consists of a fine mesh of copper, rhodium, gold, or a mixture of them (Figure 1.3 a). The mesh consists of several hundred squares covered with a thick carbon layer (dark grey in Figure 1.3 b). Usually, the thick carbon layer contains several tens to hundreds of regularly arranged holes (Figure 1.3 b and c). Optionally, the thick carbon layer can be covered with a thin (2-5 nm) carbon support film, which binds the particles (grey, horizontal line in bottom Figure 1.3 e).

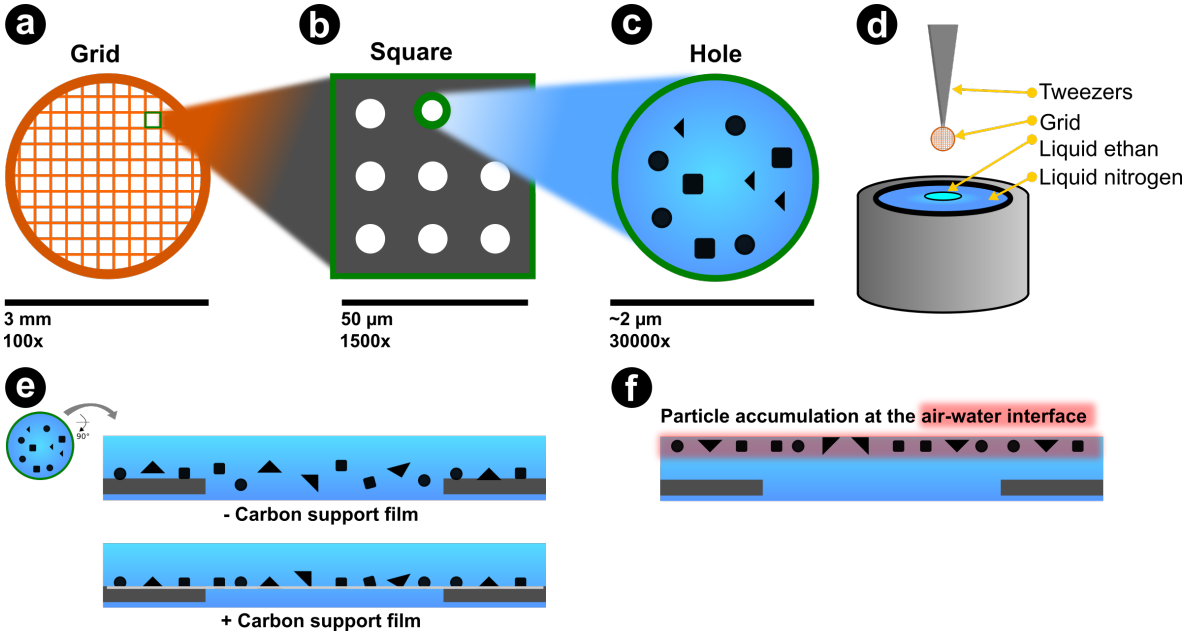



**Figure 1.3: Schematic depiction of the grid architecture and vitrification.**
***(a - c)*** *Architecture of a typical EM-grid. The typical size and used magnification are indicated. A grid* ***(a)*** *consists of several hundred squares* ***(b)****, covered with a thick carbon film (dark grey). Within this film regularly arranged, circular holes* ***(c)*** *are located, which are used for the data acquisition.* ***(d)*** *Depiction of the vitrification process: The grid, held by a tweezer, is plunged into liquid ethane cooled by a surrounding volume of liquid nitrogen.* ***(e)*** *Side view of c), without (-) and with (+) a thin carbon support film (light grey). In both cases, particles are located in the middle of the ice film away from the air-water interface.* ***(f)*** *Accumulation of particles at the air-water interface often has a negative impact on the particle quality due to partial denaturation.*

**Negative staining** is used for a first assessment and subsequent optimization of the sample quality. For negative staining of an MMM, heavy metal salts such as uranyl acetate and EM grids are used in a simple and fast preparation process: A droplet of the sample is spread onto the EM grid, excess fluid is removed, the negative staining reagent is added, excess fluid in removed once more and the grid dried. Afterwards, individual particles of the MMM are outlined by a dense coat of heavy metal stain, allowing its imaging in the electron microscope(1, 8).

**Vitrification** of the sample aims in its fixation in a native, hydrated state. To vitrify a biochemical sample for cryo-EM, the sample is applied to an EM grid (Figure 1.3 a and d), the excess of fluid is removed via blotting, resulting in a thin layer of MMM molecules emerged in buffer. The sample is rapidly vitrified in liquid ethane at -196 °C (Figure 1.3 d), where water forms amorphous rather than crystalline ice (9). Vitrification renders the sample transmittable for electrons, preserves a native, hydrated state, and allows the insertion of the sample into the vacuum of an EM (9, 10). The sample is also more resistant to radiation damage caused by the electrons (11, 12). Ideally, the formed ice layer is slightly larger than the largest particle diameter and allows for a random orientation of the particles therein (Figure 1.3 e).

### 1.1.3 Data Collection

Data collection entails the automatic acquisition of images using the prepared EM grids. Automated data collection for a SPA projects usually comprises the selection of suitable grids, squares, and holes by the user or the software package based on criteria like quality of the negative stain or ice, particles distribution, or number of particles per hole. Either commercial package like EPU (13) or open source software packages such as SerialEM and Leginon (14, 15) are used. These software packages control the EM and adjust the magnification automatically. Typically, magnifications of around 30000x are used for the acquisition of images (micrographs), leading to a pixel size of around 1 Å/px. Subsequently, automatic data collection is carried out ranging from a few hours to several days, collecting hundreds to thousands of images under low-dose conditions to reduce radiation damage of the sensitive sample.

The acquired experimental images are convoluted by the contrast transfer function (CTF) of the microscope, which manipulates the spatial frequencies of the resulting image (1, 16). The CTF depends on the phase shift, the spherical aberration Cs, the wavelength λ of the electron beam, and the defocus Δ, quantifying the distance of the image plane from the true focal plane. While defocus and astigmatism represent variables, the other contributing factors are constants for a given EM. The contrast of biological objects can be improved by effects from the spherical aberration and image defocus. For each recorded image the CTF is estimated and its influence restored during the step of image processing. The exact parameters for the data acquisition vary, depending on the used EM hardware and acquisition scheme. While the steps of “screening” are usually performed in low- or medium-end EM, high resolution data are collected on high-end EMs.

Such low-end or medium-EMs are usually equipped with an electron source, consisting of a tungsten filament or a $LaB_6$ cathode operated at an acceleration voltage of 100-200 kV, and with a charge-coupled device (CCD) detector. Here, one image is collected per field of view and exposure time. Such a set-up would typically allow to resolve details in the range of a few nanometers before and around 1 nm (10 Å) after image processing. The usage of low acceleration voltage enhances the contrast on the image, which is helpful during initial characterization of the sample.

High-end EMs are equipped with a field emission gun (FEG) as electron source, operated at 200-300 kV, and a direct electron detector (DED) (17, 18). The development and commercially availability of DEDs has fueled the “resolution revolution” in recent years (18, 19). For each field of view, several images or frames (e.g. 20-40) are collected over the exposure time (e.g. 5-10 s) and arranged as a movie stack. Through-out the exposure time the sample exhibits drift by microscopic instabilities, charging and beam induced motion,

leading to a decrease of information during the exposure. However, this information is mostly recovered during motion correction as part of the image processing (20–22).

For **Negative staining screening** usually a few images or small data sets (10s-100s images) are collected on low-end EMs, allowing first insights into the quality of the sample. Thereby information on the size, shape, symmetry of individual particles and insights into the homogeneity of the sample in general can be received. However, the harsh conditions of the negative staining reagents (pH 5) and the drying can lead to the collapse of fragile MMMs. Furthermore, artifacts of the staining reagents like crystals, thick or thin layers of stain reagent require substantial experience during the interpretation. Nevertheless, negative staining presents a fast and simple method to assess the quality of the sample.

For **Vitrification screening** usually a small data set (100s-1000s images) is collected on a low- or medium-end EM allowing the judgement of particle distribution and behavior under vitrification conditions. Interactions of the particles with the air-water interface (AWI), the grid material, or other particles can lead to unwanted behavior such as aggregation, denaturation, or preferential orientation (Figure 1.3 f). The latter can lead to artifacts during the image processing due to uneven sampling of the viewing directions. Both, preferential orientation and sample denaturation, can be dealt with by finding optimal vitrification parameters, adjusting the ice thickness, using different buffers or grids, adding surfactants or stabilizing molecules, or adapting the data acquisition strategy by additionally tilting the grid (6, 10, 23, 24).

For **high resolution analysis** usually a large data set (1000s-10000s movie stacks) of an optimized vitrified sample is collected on a high-end EM. Due to the usage of DEDs, specific demands are placed on the acquisition, transfer, storage and processing of these large data sets, usually in the range of 1 to 10 Tb. Furthremore, the stability of the cryo-EM needs to be ensured by reducing external influences such as mechanical vibrations or fluctuations of the temperature or the magnetic field.

### 1.1.4 Image processing

Image processing in SPA uses several hundreds or thousands of noisy particle images, representing 2D projections of the MMM's coulomb potential to reconstruct one or several 3D volumes, also called (coulomb potential density) map of an MMM at high resolution (1, 3). This is achieved by selecting particle images, finding the orientation parameters, x, y, $\phi$, $\theta$, and $\psi$, for each particle image, and removing "bad" ones. The five orientation parameters are (i) the rotational angle $\psi$ and the translation shift x and y, denoting in-plane operations, and (ii) two Euler angles $\varphi$ and $\theta$ defining the projection direction (Figure 1.5 a).

Image processing represents the third part of an SPA experiment. Negative stain screening aims in an initial characterization of a sample at low resolution (20-15 Å) as well as the reconstruction and verification of an *ab initio* volume. Vitrification screening intents to characterize and optimize the sample under vitrification conditions at low to medium resolution

(10 Å) as well as independently reproduce of the previous findings from negative stain screening. High resolution analysis reaches after the ultimate goal of a SPA project, the reconstruction of one or several 3D volumes at high resolution (below 3 Å).

Depending on the step within the workflow, screening or high resolution analysis, image processing consist of different sub-steps. For the two steps of screening, only a limited set of sub-steps is usually performed to reconstruct and verify an *ab initio* model (Figure 1.4 a). The verified *ab initio* model can then be used as a 3D reference during the high resolution analysis (Figure 1.4 b). In the following all sub-steps will be described for the high resolution analysis and some comments on image processing during screening will be given.

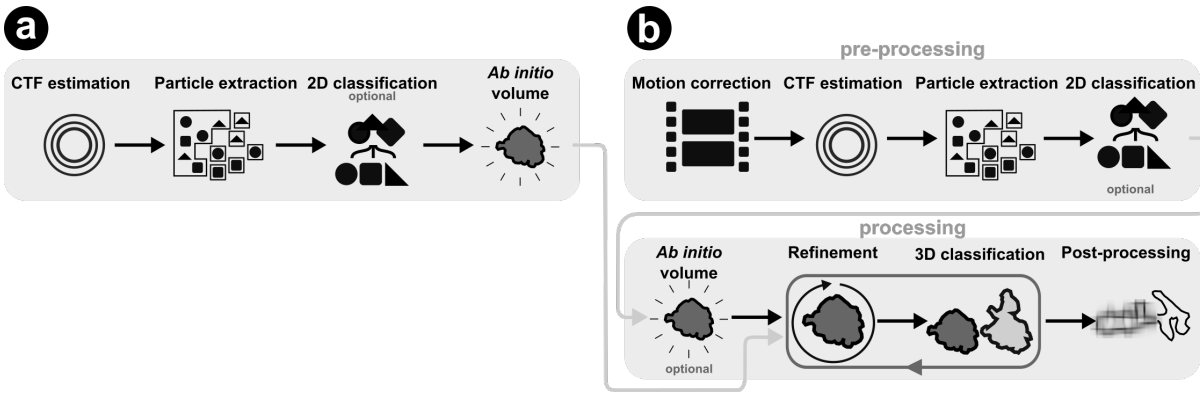



**Figure 1.4: Image Processing for screening and high resolution analysis.**
***(a)*** *The sub-steps of image processing during the screening and* ***(b)*** *high resolution analysis are depicted. For the latter, a division into pre-processing and processing was performed. The iterative combination of the steps refinement and 3D classification are depicted. The usage of an* ab initio *volume derived from the screening step is marked.*

### 1.1.4.1 Pre-processing

Here, pre-processing will be defined as a four-step process, which can principally be performed without user intervention. Pre-processing is ideally performed automatically and "on-the-fly" during the data collection. That way sample quality and the microscope performance can be monitored and adjusted, if necessary. The complete order of events is shown in the upper panel of Figure 1.4 b. Several software packages do exist that perform one specific (e.g. CTFFind4 (25), gCTF (26) or MotionCor2 (21)) or all of the necessary tasks (e.g. SPHIRE (27), Warp (28), Relion (29), cryoSPARC (30) or cisTEM (31) in a user-defined or automated way. In order to perform these steps in a sequential manner, scripting languages such as Bash or Python, or specifically designed meta-software packages like Appion (32) or Scipion (33) can be employed to distribute the task to either CPUs or GPUs, depending on the used program. The four sub-steps of pre-processing are described in the following:

**Motion correction.** The individually acquired frames in a movie stack are aligned to recover beam-induced motion that blurs the particles in the field of view. The alignment can be performed by using the entire frame or by dividing each frame into several patches (20, 21,

28, 34). This sub-step often includes the dose weighting procedure, allowing to recover the degradation of the signal during the exposure time due to radiation damage by the electrons (20). After alignment and dose weighting a summation is performed yielding one image per movie stack. This sub-step is usually not required for the analysis of screening data sets, since no movie stacks are collected on low- or medium-end EMs.

**CTF estimation.** The parameters of the contrast transfer function (CTF) are estimated, by measuring the defocus and the astigmatism within each image (25, 26). This estimation can be performed for the entire image or based on either patches or particle positions, allowing to recover non-uniform CTF-values across the image (28, 29). The found CTF-values are used during image restoration later.

**Particle picking and extraction.** Each image is used to select particle positions. Several software packages (e.g. cisTEM (31), EMAN2 (35), Relion (29)) are available to perform this task either manually, semi-automatically or automatically. Template-based particle picking can be performed using either a Gaussian blob or a known 2D template. Recently, the help of neural networks was successfully employed for the task of selecting suitable classes (28, 36–38). After determining the particle positions, particle images are extracted from the image (dimensions of around 4000 x 4000 Å) by cropping each particle into smaller boxes (e.g. 200 x 200 Å). Each particle image contains a centered particle and surrounding background. These particle images are arranged as particle stacks together with their associated CTF-values and are used for subsequent processing.

**2D classification.** Particle images are aligned and grouped into homogeneous 2D classes (3) using k-means algorithms or a maximum likelihood approach (39). By aligning and averaging all particle images within one class, a higher signal-to-noise ratio (SNR) can be achieved, allowing to visually differentiate “good” from “bad” classes. Classes considered as “good” represent different possible views of an MMM and exhibit low in-class variance. Classes considered as “bad” do not represent meaningful views or contain artifacts like aggregates, ice-fragments or edges. However, the classification is far from perfect due to the low contrast in each individual particle image or the existence of rare “good” particles. Therefore, the possibility of false positive and false negative classifications needs to be considered. Particle images in “good” classes are forwarded to the following steps of image processing. Usually, this selection is highly subjective and depends on the user’s experience. However, recent usage of neural networks allows for an automatic selection of “good” classes (40). The sub-step of 2D classification can be circumvented altogether in order to remove “bad” particle images later, during 3D classification.

### 1.1.4.2 Processing

Processing describes the main part of the image processing. As depicted in Figure 1.4 b, it consists of the four sub-steps *ab initio* volume (optionally), refinement, 3D classification, and post-processing. Before described each sub-steps individually, the fundamental process of reconstructing a 3D volume from 2D particle images will be introduced.

**3D reconstruction.** In SPA the reconstruction of a 3D volume is based on several 2D images, all possessing the previously mentioned five orientation parameters x, y, ψ, ϕ, and θ (2, 41) (Figure 1.5 a). These orientation parameters are found during the process of refinement as described in below.

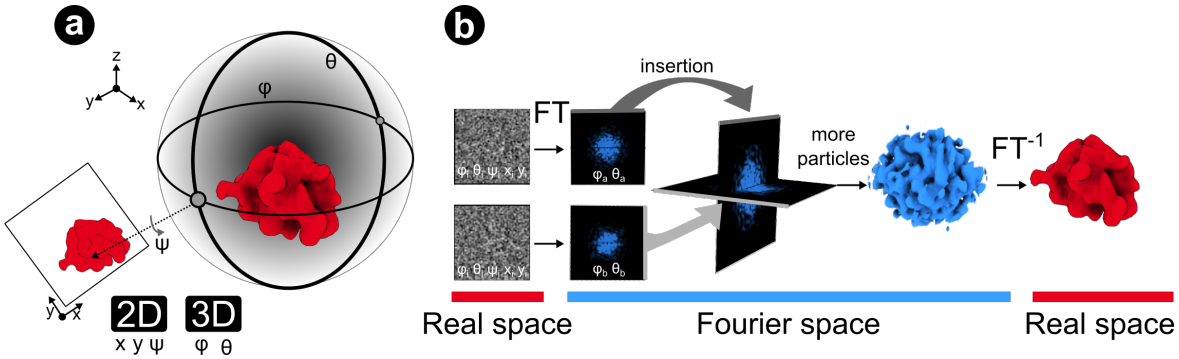



**Figure 1.5: Orientation parameters and 3D reconstruction.**
***(a)*** *Relationship between the 2D projection and the 3D volume. In order to generate a 2D projection from a 3D volume, a point on the Euler sphere is defined by ϕ and θ, representing the direction of a view. The three parameters x, y and ψ associated with the 2D projection are in-plane parameters for shift in x and y and rotation, respectively.* ***(b)*** *The process of 3D reconstruction using the method of direct Fourier inversion is illustrated. (i) Two 2D real space images are (ii) Fourier transformed and (iii) inserted into the 3D volume according to their Euler angles ϕ and θ. (iv) Inserting more images yields a robust 3D Fourier volume. (v) By inversion of the Fourier transform, a 3D real space volume is generated, based on the input data.*

In general, reconstruction algorithms used for SPA need to meet the following requirements (2, 42): They need to yield meaningful results in the presence of high degrees of noise, handle very large data sets with over $10^5$ particle images, work with box sizes of more than $256^2$ pixels, account for an uneven distribution of projection directions or spatial frequencies, be computationally cheap, minimize artifacts due to gaps within the Fourier space coverage, allow for the correction of CTF effects, and result in a volume that is linear and shift-invariant to ensure meaningful resolution measurements.

Two main groups of reconstruction methods exist matching the stated requirements, i.d. algebraic and transform methods (41). The transform methods, which are based on the central section theorem, are often used in SPA. Within the transform methods, the subgroup of direct Fourier inversion algorithms is computationally very efficient for this task. Here, the 2D particle images are Fourier transformed and then inserted as slices into the 3D volume at their respective Euler angles (Figure 1.5 b). Thereby, an interpolated 3D Fourier volume is

assembled. Upon calculation of the invertion of the Fourier transform, a 3D volume in real space is obtained.

***Ab initio* volume.** When investigating an MMM, its 3D structure is often unknown. Here, the steps of screening aim in providing a first approximation of the 3D structure in form of a 3D volume at low resolution. Such a 3D volume is needed for all subsequent steps as a 3D reference. PDB structures, rough geometric shapes or results from previous projects can be used as a 3D reference in the beginning. However, the independent generation of an *ab initio* volume ensures the consistency within the sample and can help to avoid biases. Several computational and experimental methods exist to create an *ab initio* volume:

- Computational methods to generate a 3D model based on the noisy 2D particle images or 2D class averages can be grouped into three approaches: common line (43–45), random angular assignment (31, 46), and stochastic gradient descent (29, 30). All three methods are capable of reconstructing an *ab initio* volume in an iterative way but cannot determine the handedness of the *ab initio* volume.
- Experimental methods aim at imaging the same area of the grid at several defined rotation angles. These known angles are subsequently used for the reconstruction of an *ab initio* volume. Three procedures exist: tomography (1, 47, 48), orthogonal tilt (49) and random-conical tilt (RCT) (50).

Once an *ab initio* volume has been created, computational validation should be performed to ensure its consistency of the data set with the calculated results. Such comparison can be (i) the comparison of 2D class averages with their in-class particle images, or (ii) the comparison of re-projections of a *ab initio* volume with 2D class averages or raw particle images from the same data set (51).

A validated initial 3D volume is used in the next step, refinement, as the 3D reference to calculate orientation parameters for each particle image.

**Refinement**

The orientation parameters (x, y, ψ, ϕ, and θ) for each particle image are unknown in the beginning and are iteratively optimized by aligning each particle image (2D) against an existing 3D reference. In the very first iteration, this 3D volume equals the aforementioned initial 3D volume, e.g. the *ab initio* volume. After each iteration, the newly found parameters are used to reconstruct a new, better resolved 3D volume, which is then used in the next iteration as the new 3D reference. This interplay between alignment and reconstruction, called refinement, is performed until convergence is reached. The concept of 3D reconstruction was described before, in the following the concept of alignment explained in more detail.

**Alignment**. Conceptually, alignment can be defined as the computational process of synchronizing two signals by finding optimal orientation parameters. When aligning two 2D images, one represents the query and one the reference image. The translation, x and y, as

well as the in-plane rotation ψ, are optimized such that the query overlays the reference image. These orientation parameters are optimized by maximizing an objection function such as maximum-likelihood method (52) or a cross-correlation function (41). The objection function is a similarity measure, often in the range [0, 1], where 0 and 1 represent no similarity and perfect similarity, respectively.

**Projection matching.** In order to assign Euler angles and in-plane parameters to each particle image, a comparison with a 3D volume (3D reference) is required. Therefore, the projection matching approach is applied (53), consisting of four steps (Figure 1.6) and employing the cross-correlation coefficient (CCC) as a measure of similarity:

1. Create evenly spaced projections of the 3D volume at multiple sets of Euler angles ($\phi$ and $\theta$) defined by the step size $\delta$, resulting in several 2D projection images associated with specific $\phi$ and $\theta$.
2. For each 2D particle image:
   a. For each 2D projection image (associated with $\phi$ and $\theta$):
      - For each set of in-plane parameters ($\psi$, x, and y):
        - Align both images and calculate CCC value.
   b. Select the best alignment based on the highest CCC.
   c. Retrieve the associated orientation parameters (x, y, $\psi$, $\phi$, $\theta$).
3. Reconstruct a new 3D volume using all 2D particle images and their optimized orientation parameters.
4. go back to 1 as long as convergence is not reached.

In order to find the optimal orientation parameters for each particle image, the number of projection images used in 1. is crucial. In the following, global and local alignment as well as convergence of the orientation parameters will be discussed in more detail.

**Global alignment.** In the first few iterations of a refinement, approximate orientation parameters for each particle image are found by global alignment (left side of Figure 1.6). Therefore, a large step size δ is used to generate a few projection images, e.g. δ=15° resulting in 81 projection images. Once, approximate orientation parameters are found, the step size δ is decreased, leading to more projection images, e.g. δ=3.75° resulting in 3025 projection images.

However, using a small step size establishes a computational burden for two reasons: (i) the large number of projection images needs to be kept in memory, and (ii) comparing particle images to more projection images increases the computational cost. The latter reason is circumvented by performing local alignment.

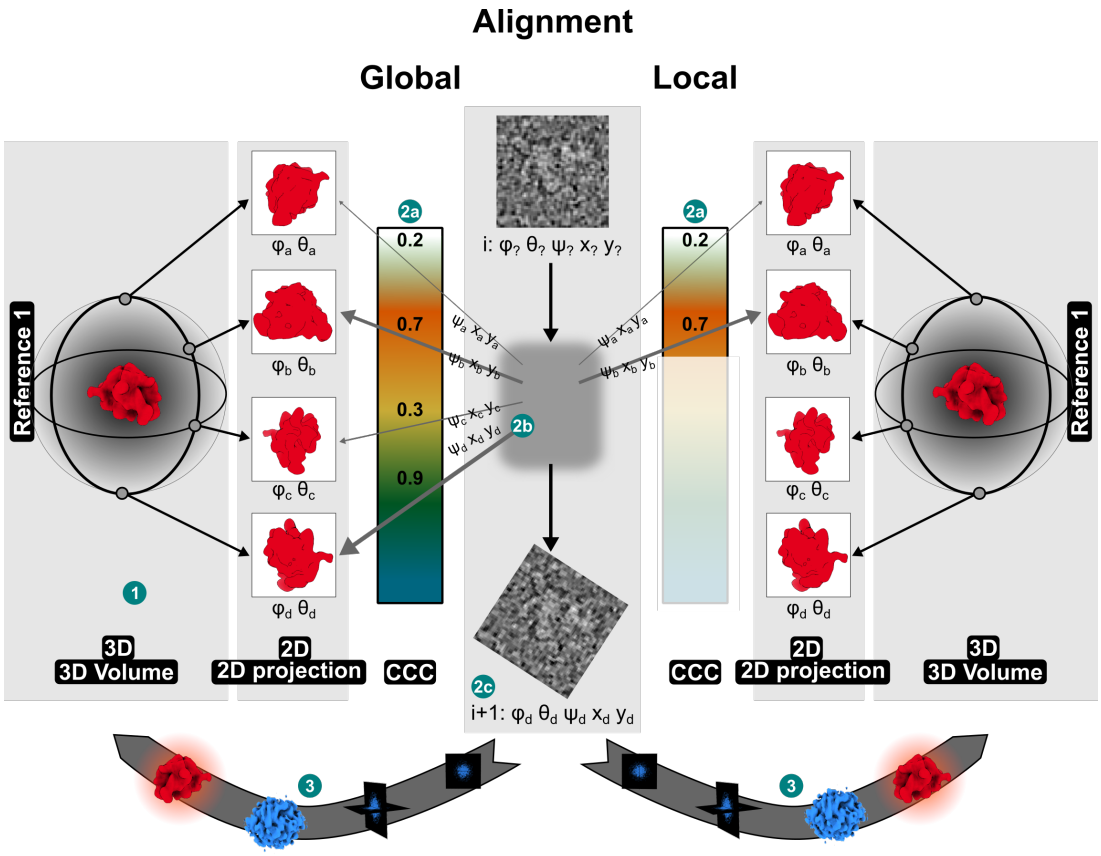



**Figure 1.6: Global and local alignment.**
*Global and local alignment are depicted, left and right from the central axis, respectively. The particle image does not possess a set of orientation parameters at iteration i, denoted by question marks. (1) Using a 3D volume, 2D projections are generated (black arrows). For each particle image the following steps are applied: (2a) The template-matching operations is performed, comparing the particle image with a set of 2D projections (grey arrow). The thickness of the arrows represents the obtained CCC values. (2b) The differences in search space for global and local alignment are depicted by two color gradients. The highest CCC value is selected and (2c) the corresponding parameters are associated with the particle image at the next iteration i+1. Once all particle images have been aligned, (3) a new 3D volume is reconstructed and (1) used as the reference in the next alignment iteration.*

**Local alignment.** Local alignment restricts the search space for the projection matching a each particle image to the neighboring projections on the Euler sphere (right side of Figure 1.6). Usually, this selection is located within a cone of a few degree around the current Euler angles ϕ and θ. Thereby, local alignment allows the finding of more accurate orientation parameters, reduces the computational burden and speeds up the alignment process.

**Convergence.** Most modern software packages such as SPHIRE (27), cryoSPARC (30), Relion (29), or cisTEM (31) possess some form of automatic refinement routine. Usually the step size δ of 8-15 ° is used in the beginning during global alignment while being automatically and incrementally reduced to a few degrees during local alignment. The refinement is automatically stopped, once convergence is reached. Convergence can be defined as a stagnation of the changes in orientation parameters or in the resolution improvement.

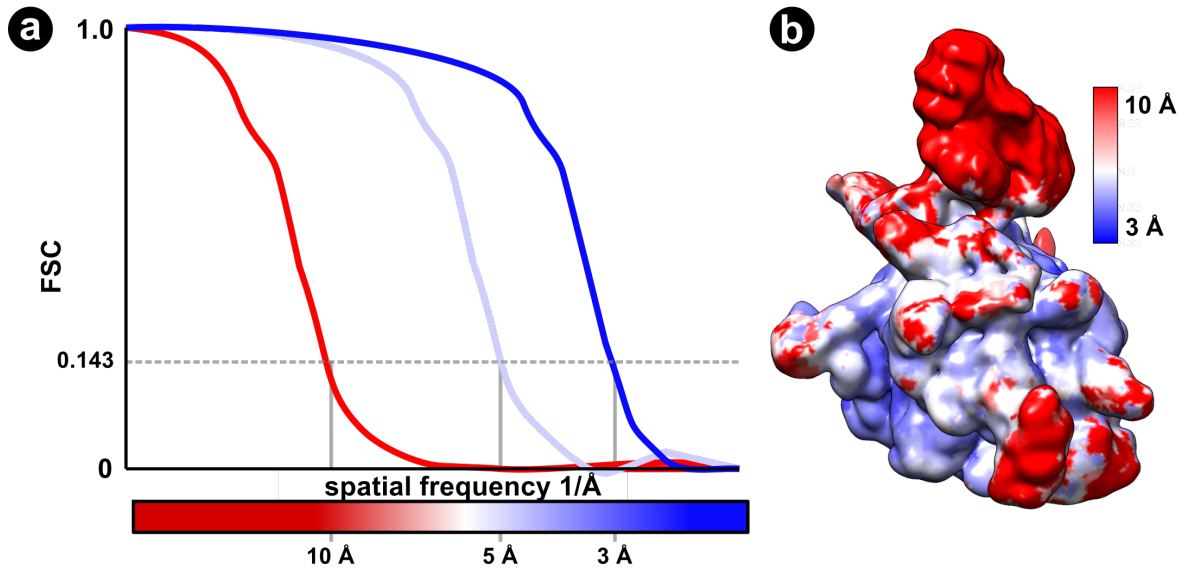



**Figure 1.7: Global and local resolution assessment.**
***(a)*** *Global resolution assessment using a Fourier shell correlation (FSC) curve. The x-axis and y-axis denote the resolution in spatial frequency in 1/Å and the FSC-value, respectively. The crossing point of the plot with the FSC threshold value of 0.143 marks the resolution on the x-axis. Three plots in red, light blue, and blue are shown with resolutions of 10, 5 and 3 Å, respectively.* ***(b)*** *Local resolution mapped onto the surface of a 3D volume using the same color code as in (a). While the upper part and some extensions exhibit a lower resolution of around 10 Å, the lower part exhibits resolution of 5 Å resolution and better.*

**Resolution assessment.** Determining the resolution of a cryo-EM map is an important and very controversial matter in the field. For cryo-EM reconstructions, resolution is defined as a single number for the entire map, stating the maximum spatial frequency at which information can be considered reliable. However, due to the highly subjective term “considered reliable”, numerous papers on the theoretical and practical implementations have been published (54–57).

Currently, the gold-standard Fourier shell correlation (FSC) is the most widely adopted method to determine the resolution of a 3D volume in cryo-EM. While the coordinates of a volume in real space are the three axes x, y and z, in Fourier space the three axes denote spatial frequencies 1/Å (shells) into the different directions. Spatial frequencies range from low resolution (e.g. 1/200 Å) to high resolution (1/2 Å). The curve typically shows as fall-off from 1.0 at low resolution to 0.0 at high resolution. The gold-standard FSC measures the self-consistency within the data set in Fourier space in the following way:

1. Divide the data set randomly into two half-sets and refine them independently.
2. Reconstruct both half-sets as volumes “half1” and “half2”.
3. Fourier transform these two volumes and calculated the CCCs at each Fourier shell.

4. Plot the CCC at each Fourier shell, denoted as spatial frequency 1/Å, and thereby creating the FSC curve.
5. Use a CCC threshold of 0.143 to find the intersection with the FSC curve. The x-value at the intersection is converted to a resolution measure in Å and stated as the global resolution.

During refinement, the resolution of the 3D volume is determined after each round via the gold-standard FSC measurement. The obtained FSC curve or the resolution value are used for filtration in order to prepare the reference volume for the next iteration. This aims in minimizing the bias of the used reference by restricting the influence of high frequency noise.

Since this global resolution represents an average value over the entire volume, additional resolution measurements have been developed to quantify the resolution with high spatial accuracy (58–60). Such information can be used to locally filter the volumes, often improving the interpretation of volumes exhibiting a wide range of resolution.

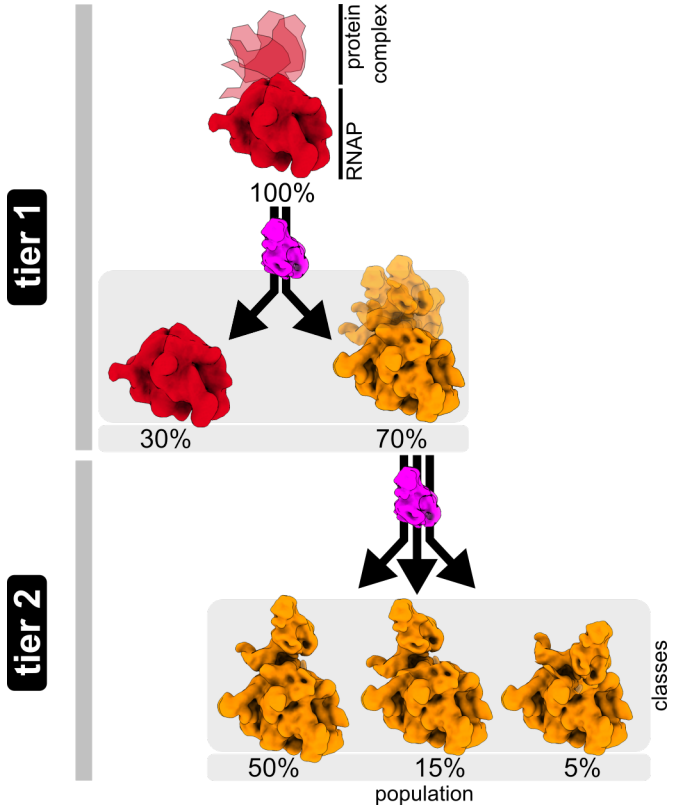



**Figure 1.8: Hierarchical sorting of a heterogeneous MMM.**
***(tier 1)*** *Depiction of hierarchical sorting in two tiers. The full data set (100 %) possesses a high level of heterogeneity depicted by the transparent area in the upper part. Using a focus mask (magenta), the data set is classified into two classes representing RNAP alone (30 %) and a RNAP:protein complex (70 %).* ***(tier 2)*** *Since the latter still depicts some levels of heterogeneity, it is further classified into three classes with populations of 50%, 15% and 5%, respectively. These three classes all represent the RNAP:protein complex exhibiting different conformational states of the bound protein. These states can potentially represent distinct structural states of RNAP during regulation by the protein complex.*

### 3D classification

Once the refinement has converged, the reconstructed volume usually depicts different levels of heterogeneity. While central regions are often stably aligned, flexible domains at the periphery or sub-stoichiometrically bound ligands represent regions of high heterogeneity. Heterogeneity within the data set can be described as compositional and conformational (2, 61): (i) compositional heterogeneity: sub-stoichiometric presence of subunits or ligands within a complex. (ii) Conformational heterogeneity: a mixture of different conformation states of one or several subunits.

Heterogeneous region within a 3D volume often appear as fragmented densities. 3D classification is performed to separate the heterogenous data set into discrete homogenous classes. The particle images in a class are termed the population, quantified by an absolute number and a percentage. 3D classification in several tiers (levels) is called hierarchical sorting (5, 62). Such a hierarchical sorting tree is illustrated in Figure 1.8. In the first tier, 3D classification is often performed on large differences, here the presence or absence of a protein complex bound on top of the RNA polymerase (RNAP). In the second tier, smaller differences in the conformation of the protein complex are used for their classification.

In order to reduce the influence of noise on the outcome, a mask is used to restrict the classification to the region of interest (ROI). This approach has been termed focused classification (63, 64) and requires three steps: (i) detection of heterogeneous regions, (ii) design of a focus mask and (iii) performance of the 3D classification. In the following, all three steps will be explained using a RNAP:protein complex as an example.

**(i) Detection of heterogeneity.** Regions of high heterogeneity can be indirectly detected when visualizing the map in Chimera (65). As an example, Figure 1.9 a illustrates the same 3D volume at different density thresholds. While the lower part is present at both thresholds, the upper part is not. Filtration of the 3D volume to a resolution of 10 Å resolution can visualize the heterogeneous regions by suppressing signal and noise beyond that resolution (Figure 1.9 b).

In order to directly visualize regions of heterogeneity, several approaches have been developed. Often they rely on calculating the 3D variance and/or co-variance maps (63, 64, 66–68). However, these methods are often computationally too expensive, not yet publicly available or are not universally applicable to all samples.

The tool “3D variability” was developed (5, 27, 69), where particle images from the same point on the Euler sphere are grouped and 2D averages as well as 2D variances are calculated. Afterwards, 2D averages and 2D variances are used for the reconstruction of two 3D volumes, i.e. the “3D average” volume and the “3D variability” volume. While the “3D average” volume merely serves as a control, the “3D variability” volume shows a high signal at regions of high 2D variance and can therefore be used to directly visualize heterogeneity within the data set (Figure 1.9 c). This spatial information is then used to design a focus mask for 3D classification of the data set.

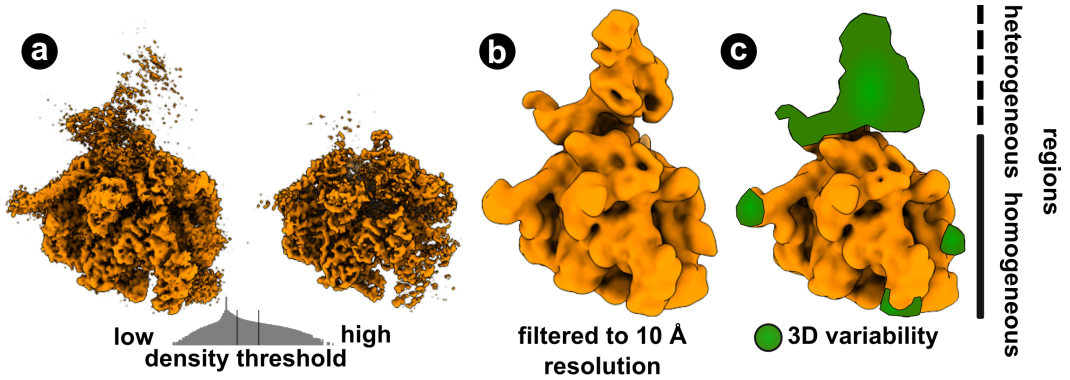



**Figure 1.9: Detection of heterogeneity in a 3D volume.**
***(a)*** *Depiction of a high-resolution map at a low- and high-density threshold. The upper, heterogenous region is only visible as fragmented density, when a low threshold is applied.* ***(b)*** *The same map as in (a) filtered to 10 Å resolution is shown. Here, the upper region can be clearly visualized.* ***(c)*** *Result of 3D variability analysis depicting regions with high heterogeneity (green).*

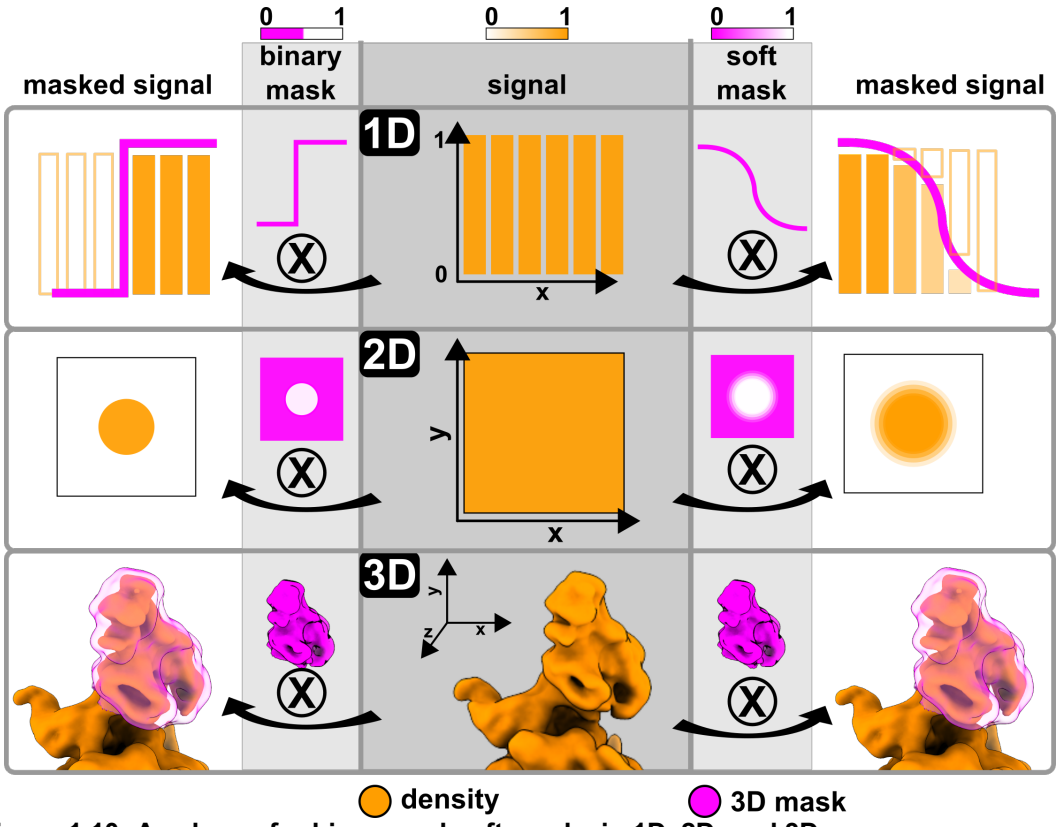



**Figure 1.10: A scheme for binary and soft masks in 1D, 2D, and 3D.**
*Masks are depicted in magenta. The multiplication of the mask with the signal is depicted with an arrow resulting in the masked signal on the far left and right, respectively. The column in the middle (grey) represents the signal in different dimension. The major difference between binary and soft mask is best visualized in 1D, where the binary mask corresponds to a step function while the soft mask possesses a smooth fall off, e.g. a Gaussian fall-off. 3D masks are shown transparently to visualize the enclosed area.*

**(ii) Design of focus mask.** In general, masks are used to restrict the computational analysis to the region of interest (ROI) and can thereby reduce the influence of noise or surrounding regions to the outcome. Masks are one, two or three dimensional (1D, 2D, or 3D, respectively) objects, that contain the value 1 inside and 0 outside the ROI. Masks are multiplied with the signal resulting in a masked signal where only the ROI is present (Figure 1.10). Masks have the same dimensions as the signal. In cryo-EM, two types of masks can be used for image processing, binary and soft mask (Figure 1.10):

- **Binary masks** possess a step function at the edge between the ROI and the outside. They are used for focused classification, to only consider the signal within the ROI.
- **Soft masks** exhibit a smooth Gaussian or cosine fall-off at the edge ranging over several pixels. Thereby, the signal is smoothly suppressed to 0 which reduces artifacts at the edges of the masked signal. Soft masks are often used to diminish the influence of the background or amorphous densities such as detergent micelles.

Both types of masks are used at different steps during image processing. Masks are either designed by the user or automatically.

**(iii) perform 3D focused classification.** In general, 3D classification of particle images aims at separating heterogenous data sets into discrete classes. 3D classification can be performed in two different ways - simultaneously with alignment, and sequentially, after the alignment:

- During simultaneous alignment and 3D classification, each particle image is compared to projections from two or more reference volumes. Thereby, orientation parameters and class assignments are found at the same time.
- For sequentially 3D classification, the orientation parameters found during the alignment are used and only the assignment of particle images into different classes is changed. This way will be discussed in the following.

When restricting the 3D classification to a specific ROI, it is called 3D focused classification. For that purpose, a binary or focused mask is used, which encompasses the ROI (63). The procedure of focused classification follows these steps (Figure 1.11):

1. For each particle image the following steps are applied:
    a. Retrieve current orientation parameters x, y, ψ, ϕ, θ.
    b. Use them to create one projection from each 3D volume.
    c. Use them to create one projection from the focus mask and binarize it.
    d. Multiply projections of volume with projection of focus mask.
    e. Multiply particle image with projection of focus mask.
    f. Compare masked particle image with masked projection of volume, resulting in a CCC value each.
    g. Select best matching class based on the CCC-values.
    h. Assign particle image to that class
2. Reconstruct new 3D volumes using all assigned particle images.

This procedure is performed iteratively aiming for stable class assignments after a few iterations. In order to further separate the data set, more 3D volumes can be added as references. Thus, particle images with marked difference within the ROI are separated into different classes. Upon iteration of this procedure, a clear separation of the heterogenous data sets into discrete classes should be achieved.

However, since MMMs often exist not in discrete conformations but rather exhibit continuous motions, problems can arise. Specific methods have been designed for that purpose: 3DVAR as part of cryoSPARC (30), manifold embedding (70), hybrid EM normal mode analysis (71), WarpCraft (72), and multi-body refinement as part of Relion (73).

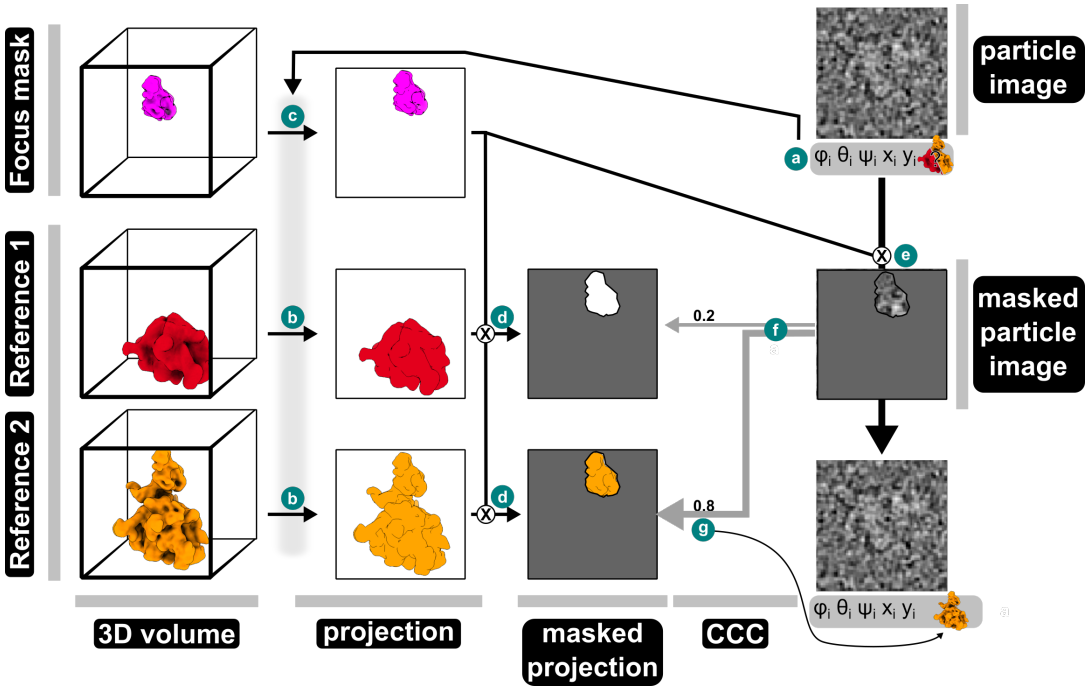



**Figure 1.11: 3D focused classification.**
*Schematic process of 3D focused classification: One 3D focus mask (pink) and two 3D references (red and orange, respectively) are depicted. From the current particle image, the orientation parameters (a) are used for the projection of both 3D references (b) and projection and binarization of the 3D focus mask (c). The 2D mask is multiplied with the reference projections (d). The masked particle image (e) is compared to both projections (f) and the particle image is assigned to the class with the highest CCC-value (g).*

### 1.1.4.3 Post-processing

Once 3D classification yielded one or several final 3D volumes, post-processing is performed comprising several steps, with a few important being listed here:

**Assessment of resolution.** The global resolution of the map in Å is quantified using the previously described gold-standard FSC procedure. Since the gold-standard FSC measurement results in one average number for the entire map, local variations are not detected. Local variations in resolution can arise from conformational or compositional heterogeneity. A quantitative assessment of the local resolution and its mapping onto the 3D volume can be done to visualize heterogeneous regions and support the model building. Several software packages exist to evaluate the local resolution, e.g. LocalRes, where resolution is evaluate against the background noise (58), or local FSC as implemented in SPHIRE (27) or Relion (29), where, small corresponding regions in both half maps are used for FSC calculations (59). This information can be used to filter the 3D volume locally, according to its resolution.

**Sharpening.** A reduction of the map quality is caused by several factors during data acquisition and image processing: specimen movement, sample charging, radiation damage, inelastic electron-scattering events, partial beam coherence, alignment, and classification

errors as well as compositional and conformational heterogeneity. A combination of all these effects results in the down-weighting of high-frequency information resulting in an overall reduction of the resolution (74). To restore these high-resolution information, a filter function is applied to the reconstructed 3D volume. This process is also called “sharpening”, since it enhances the level of details seen in the 3D volume and in the surrounding noise. Different approaches exist to create a suitable filter function: (i) Global methods apply the derived filter function to the entire volume, e.g. PDB enhancement (75–77) and B-factor sharpening (74). (ii) Local methods apply specifically adjusted filter functions to local regions individually and thereby preserve their local character, examples being LocScale (78) and tools within Phenix (79, 80).

### 1.1.5 Interpretation

The final, sharpened 3D volume is used for interpretation using an atomic model. In general, a 3D volume is interpreted using previously available models, e.g. derived from other cryo-EM, MX or NMR studies, homology models or using a *de novo* built model. Since the focus of this work lies on image processing, the reader is referred to excellent reviews regarding model building in cryo-EM (81, 82). In order to generate an atomic model, a large pool of software packages is available, e.g. PyMol (83), Coot (84, 85), Phenix (86, 87), and RosettaES (88), Chimera (65), ChimeraX (89).

The validation of the generated model is performed to ensure the quality of the model itself and the model-to-volume (model-to-map) fit (81). (i) Model quality assessment is widely performed using the package MolProbity (90). A good model has for example a low clash score as well as a low number of Ramachandran outlier and bad side chain rotamers, respectively. (ii) The global model-to-map fit quality has evolved as a standard to judge high-resolution maps using the CCC as a metric (91). Additionally, cross-validation is performed by detecting overfitting using a model-map FSC (92).

In order to not only validate the 3D volume but the entire research hypothesis, an integrative structural biological approach should be employed, combining complementary techniques.

### 1.1.6 Summary of cryo-EM workflow

The depicted five-step workflow of a cryo-EM project (Figure 1.2) and typical values for its steps and sub-steps are summarized in Table 1.1. The exact structure of the workflow can, among other factors, vary on the sample, the available equipment, and the scientific question. Therefore, the presented workflow should be taken as an optimistic example of how structural determination via cryo-EM could be performed.

**Table 1.1: Summary of cryo-EM workflow.**

| | step in workflow | Screening | Screening | Analysis |
|---|---|---|---|---|
| | **aims of experiment°** | - initial characterization (e.g. distribution, size, shape, homogeneity)<br>- 2D classification, *ab initio* 3D reconstruction<br>- Comparison results to to the literature | - optimization of vitrification conditions (e.g. ice thickness, particle distribution and orientation, contrast)<br>- independent reconstruction of initial model<br>- detection of subunits or proteins in 3D volume | - collection and processing of high resolution data<br>- reconstruction at high resolution and 3D classification<br>- atomic model building |
| **Sample prep.** | **$t_{sample\ prep.}$*** | 10 - 30 min | 30 – 60 min | 30 – 60 min |
| | **v[µl] @ c[µM]*** | 4 @ 0.1 – 0.01 | 4 @ 1 – 0.01 | 4 @ 1 – 0.01 |
| | **stored @** | room temperature | -196°C (liquid nitrogen) | -196°C (liquid nitrogen) |
| **Data collectio** | **typical EM** | low-end microscope 120 kV + CCD | mid-end microscope 120-200 kV + CCD | high-end microscope 200-300 kV + DED |
| | **$t_{data\ collection}$*** | several hours | several hours | hours to days |
| | **#imgs** | 10s-100s | 100s-1000s | 1000s-10000s |
| **Image processing** | **disc-space** | several Gb | several Gb | several Tb |
| | **#ptcl imgs** | 100s – 10000s | 1000s – 100000s | 10000s – Mio.s |
| | **motion cor** | | (x) | x |
| | **CTF** | x | x | x |
| | **ptcl X** | x | x | x |
| | **2D class** | x | x | x |
| | ***ab Initio*** | x | x | x |
| | **Refinement** | | (x) | x |
| | **3D class** | | (x) | x |
| | **post Proc** | | | x |
| | **max reso*** | 20 - 15 Å | 20 – 10 Å | <4 Å |

(°) Aims of an individual experiment can vary depending on the question, sample quality and available equipment. (*) The stated numbers are examples and can vary depending on several factors like used EM, size of data set, heterogeneity of sample and future development in hard- and software. Abbreviations used: sample prep. – sample preparation, stored @ – grids stored at temperature , v @ c – needed volume (µl) per grid at given concentration (µM), Gb and Tb – gigabyte and terabyte, $t_{sample\ prep}$ / $t_{data\ collection}$– time required for sample preparation / data collection, #imgs – number of images, #ptcl imgs – number of particle images, motion cor – Motion correction, CTF – CTF estimation, ptcl X – particle extraction, 2D and 3D class – 2D and 3D classification, Post Proc – Post Processing, max reso – maximum resolution

## 1.2 Integrative structural biology

The structure of an MMM alone does not provide enough information to understand its function and mechanism of action comprehensively. Therefore, it is necessary to combine data from multiple complementary techniques such as macromolecular crystallography (MX), nuclear magnetic resonance (NMR), cryo-EM, small-angle x-ray scattering (SAXS), functional assays, cross-linking and mass spectrometry (XLMS), co-purification, bioinformatics, and biophysical characterization. By combining several techniques in an integrative way, questions like the following can be tackled:

- Where does a protein bind to the MMM?
- Which interactions between amino acids, small molecules, metabolites, and nucleic acids are important for the MMM's function?
- What are the structural states of an MMM and in what sequence do they occur?
- What context, e.g. other proteins, localization, nucleic acids, does the MMM require to perform its function?
- What composition does the MMM adopt in a different context (cellular localization or bound DNA/RNA sequence)?
- Which interactions to other cellular components arise during a MMM's lifetime?
- How are the transcription, translation, activation and degradation of the MMM regulated?

In the following chapters, four studies are presented as my doctoral thesis, employing an integrative structural biology approach to resolve the structure and provide insights into the mechanism of the bacterial transcriptional anti-termination complex by λN, Ebp1:ribosome complex, and $CRL4^{DCAF1}$ Vpr-mediated SAMHD1 recruitment.

## 1.3 Bacterial RNA polymerase

### 1.3.1 Structure of the RNA polymerase

DNA-dependent RNA polymerases (RNAPs) synthesize RNA in all organisms. In bacteria, their catalytic core consists of five subunits (α, α, β, β' and ω), creating an overall "crab-claw" shape (93, 94) (Figure 1.12). The two subunits β and β' form the active site channel, buried within in a 27 Å deep cleft. The interior of this cleft exhibits predominately positively charged residues. The dimeric N-terminal domain of the α subunits ($\alpha^{NTD}$) are located opposite from the deep cleft and support the correct complex assembly (93, 95). The ω subunit wraps around the C-terminal domain of β' ($\beta'^{CTD}$), contacting several conserved regions therein, supporting the correct assembly, and regulating its conformations (96). Additionally, ω plays a functional role in binding guanosin-3',5'-bispyrophosphat and is involved in the stringent response (97). Disordered linkers connect the $\alpha^{CTDs}$ to the core of the RNAP. They have been structurally elucidated to bind DNA and regulatory proteins (98, 99). Two $Zn^{2+}$-binding elements were found exclusively in prokaryotic RNAPs: One located within a discrete zinc binding domain (ZBD) located in proximity to the flap domain close to the RNA exit channel, the other within the β', potentially facilitating the folding of this subunit (100, 101). The inside of the RNAP was divided into different channels, which fulfil specific functions (Figure 1.13 a).

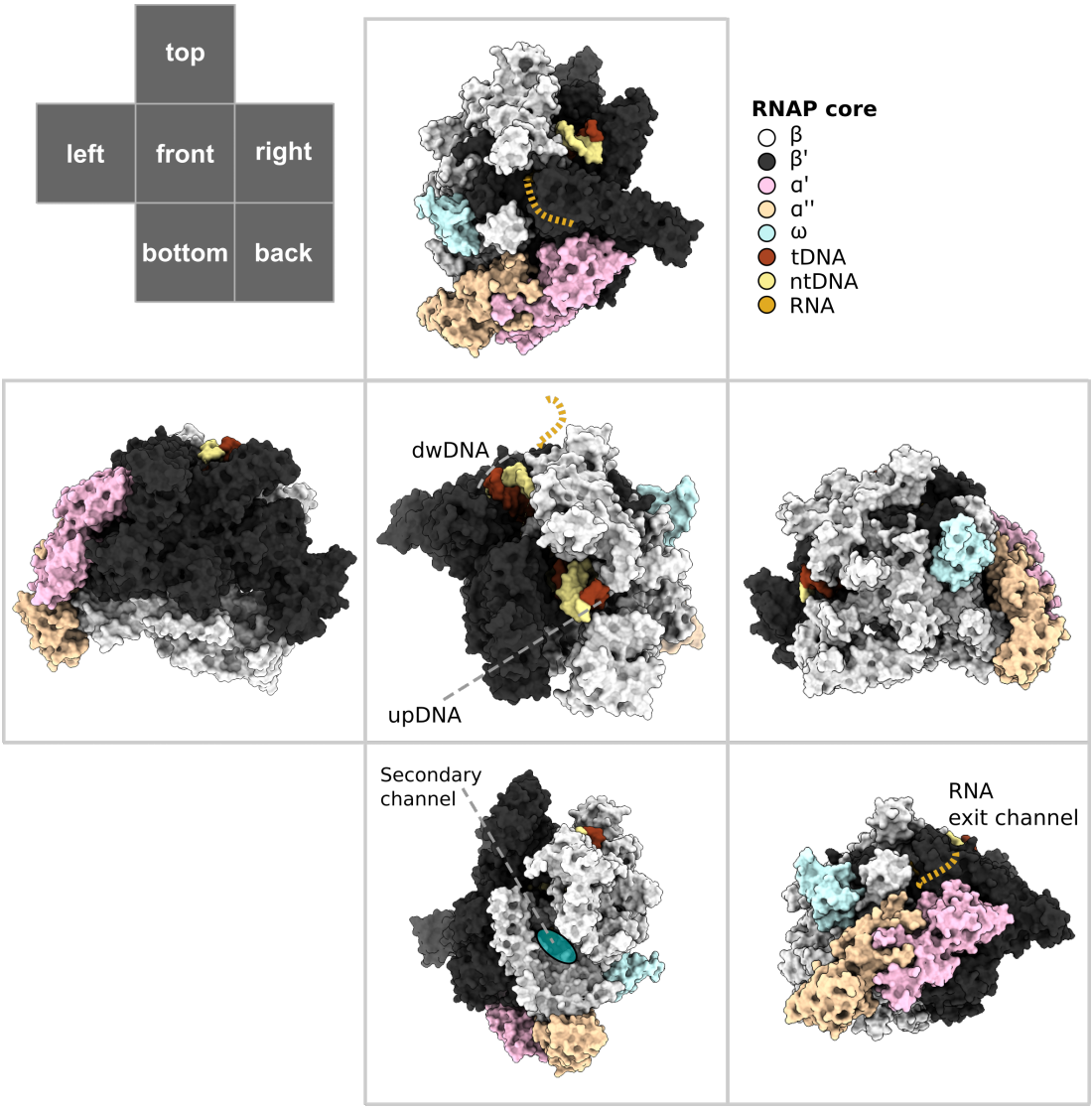



**Figure 1.12: The structure of the RNAP.**
*Structure of the RNAP (PDB ID: 6ALH) in surface representation from different viewing directions* (102). *Important structural elements are highlighted. The two large subunits β and β' (white and black, respectively) are forming the cleft, which binds the DNA and possesses the active site. The two α subunits are depicted as a pink and peach heterodimer and the ω subunit in light blue.*

**The active site channel.** The active site channel of the RNAP (Figure 1.12 and Figure 1.13) is highly conserved across bacterial RNAPs and lined with important structural elements for the catalysis and maintenance of the nucleic acid scaffold (Figure 1.13 b). The active site consists of three universally conserved aspartate residues, embedded within the NADFDGD motif of β', chelating an $Mg^{2+}$ ion (103). Three sites were defined to explain the mechanism of nucleic addition: *E*, *i+1* and *i* site. The newly incoming nucleotide binds at the E site (entrance site). While the *i*-site describes the last position of the DNA:RNA hybrid, the next nucleotide is added to the waiting DNA template at the *i+1* site (Figure 1.13 b). Downstream of the *i+1* site lies the bridge helix (BH) and the trigger loop (TL), two important structural features of the RNAP active site.

Both structures depict dynamic behavior during all RNAP activities and have been shown to adopt different conformations throughout its mechanism. The BH can adopt a straight and a bent state (100, 104) and requires two hinges as supported by mutational, computational, and structural studies (105, 106). The tip of the TL adopts different conformations in the presence and absence of nucleoside triphosphate (NTP) (107, 108), respectively. During the absence of NTP, the tip is disordered. While, in the presence of NTP, the tip folds into the trigger helix (TH) and gains access to the active site. Together with the BH, the TH forms a three-helix bundle, closing the active site and positioning the NTP substrate (see below) (109). Furthermore, the β' rudder, located further upstream, is important for the separation of template and non-template strand upstream of the hybrid. The β subunit's flap domain (βflap) as well as the β' lid and zipper domains mark the edge of the active site channel (100) (Figure 1.13 b).

**The secondary channel.** The BH separates the main channel into the downstream DNA entry channel and the 10-12 Å wide secondary channel. It is too limited to pass double-stranded nucleic acid through, but allows NTPs to access the active center (93, 100). During backtracking, the 3' end of the RNA transcript invades through the secondary channel (110).

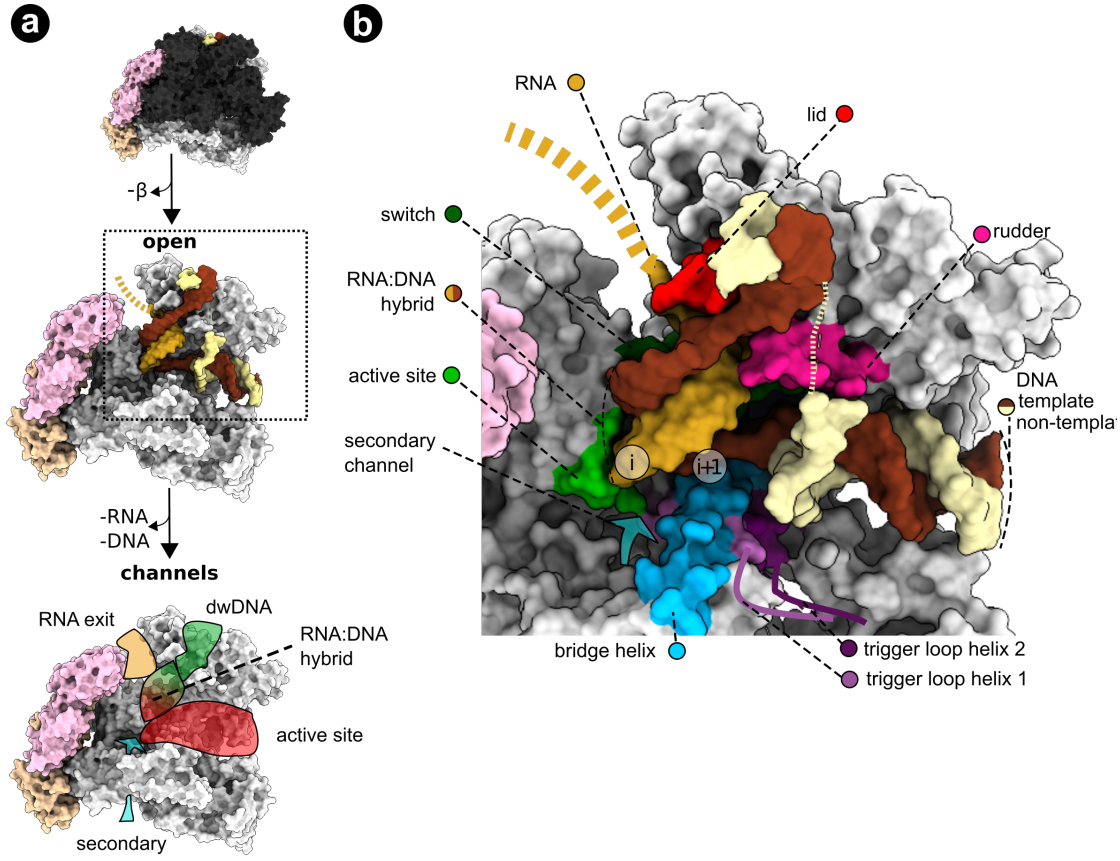



**Figure 1.13: The structure of the RNAP active site.**
***(a)*** *Left side view of the RNAP depicting the interior and the channels. (**b**) Detailed view of the structural elements within the interior (PDB ID: 6ALH)* (102).

**RNA exit channel.** The nascent transcript leaves the RNAP through the RNA exit channel (Figure 1.12 and Figure 1.13 b). Here, the RNA passes by the β flap and the β' lid, which can detect the formation of pause/termination hairpins within the nascent RNA. Such signals can lead to allosteric changes within the active site and thereby alter the elongation behavior of the enzyme (111, 112).

**Downstream DNA channel.** The *downstream DNA channel* is formed by the β lobe and the β' jaw, protecting 15-20 base pairs (bp) within the RNAP from nuclease cleavage (113). In order to form a stable transcription elongation complex (TEC), at least 9 bp of dsDNA downstream of the active site are required (114). Additionally, the sequence of this downstream DNA can modulate the TEC's rate of elongation (115), its response to pause (116), and termination signals (117, 118), suggesting an intricate interplay between the downstream DNA and RNAP.

**Mobile modules.** For the RNAP to act as an MMM, extensive conformation movements within the core subunits are required. Based on several RNAP crystal and cryo-EM structures, several mobile domains were identified (Figure 1.13 b). The mobile modules move relative to the immobile core, which is includes $\alpha^{NTD}$, ω and parts of the β and β' near the active site (103). The mobile modules are listed in the following sorted by their size:

- The β' pincer is formed by the clamp domain and closes down around the active site channel, fixating the DNA and the DNA:RNA-hybrid within the active site.
- The β pincer is formed by the lobe and protrusion, and is able to open and close the active site channel.
- The BH controls translocation and obstructs entry to the active site by adopting a kinked position. The TL transitions from unfolded to folded, where it forms a three-helix bundle together with the BH (109, 119).
- The β flap covers the RNA exit channel and recognizes exiting RNA (103, 112). The β flap tip helix (βFTH) aids in the formation of weak RNA hair pins (120).
- The switches 1-5 are highly conserved polymorphous elements at the base of the clamp. They are involved in the opening of the clamp and thereby widening of the main cleft. Compaction of the switches stabilizes contacts between the closed clamp and the nucleic acids in the elongating RNAP (121–123).

### 1.3.2 Transcription cycle

Transcription represents the first step of gene expression in all organisms. Thereby, the cells adapt to new environments, interact symbiotically or pathogenically with hosts, respond to stress, or perform basic house-keeping functions. Transcription is divided into three steps: initiation, elongation, and termination (103, 112). RNAPs transcribing along a DNA can be visualized as “beads-on-a-string” with the nascent RNA pointing away from the DNA (Figure 1.14 a). At each of these three steps, additional protein factors, small molecules, RNAs or signals within the DNA or RNA regulate the RNAP (103). In the following, an introduction into all three steps and their sub-steps is given (Figure 1.14).

**Initiation.** Transcriptional initiation comprises the loading of the RNAP to the DNA at the promoter site. Initiation itself can be separated into three sub-steps: promoter recognition, promoter opening, and promoter escape (orange area in Figure 1.14 b). First the σ factor binds to the RNAP forming the so-called holoenzyme. Subsequently, this complex binds to and moves along the double stranded DNA, until a promotor region is recognized (124). Afterwards, the open initiation complex is formed, including bending and unwinding of the DNA to form a transcription bubble. In the last step of initiation, promoter escape is achieved in short rounds for abortive transcription by the initial transcribing complex (ITC), resulting in short RNA segments.

**Elongation and the nucleotide addition cycle.** After leaving the promoter, the RNAP is part of the transcriptional elongation complex (TEC). During elongation, the affinity of the σ factor to the RNAP decreases gradually, allowing its dissociation while several other transcription regulator factors are recruited to the RNAP, ensuring a stable and processive elongation through the gene (green area in Figure 1.14 b). During elongation, nucleotides are added to the nascent RNA, depicted in the nucleotide addition cycle (NAC) in Figure 1.15. Throughout the NAC, RNAP can adopt two states, defined by the position of the catalytic site relative to the 3’ end of the RNA:DNA hybrid: pre- and post-translocated state.

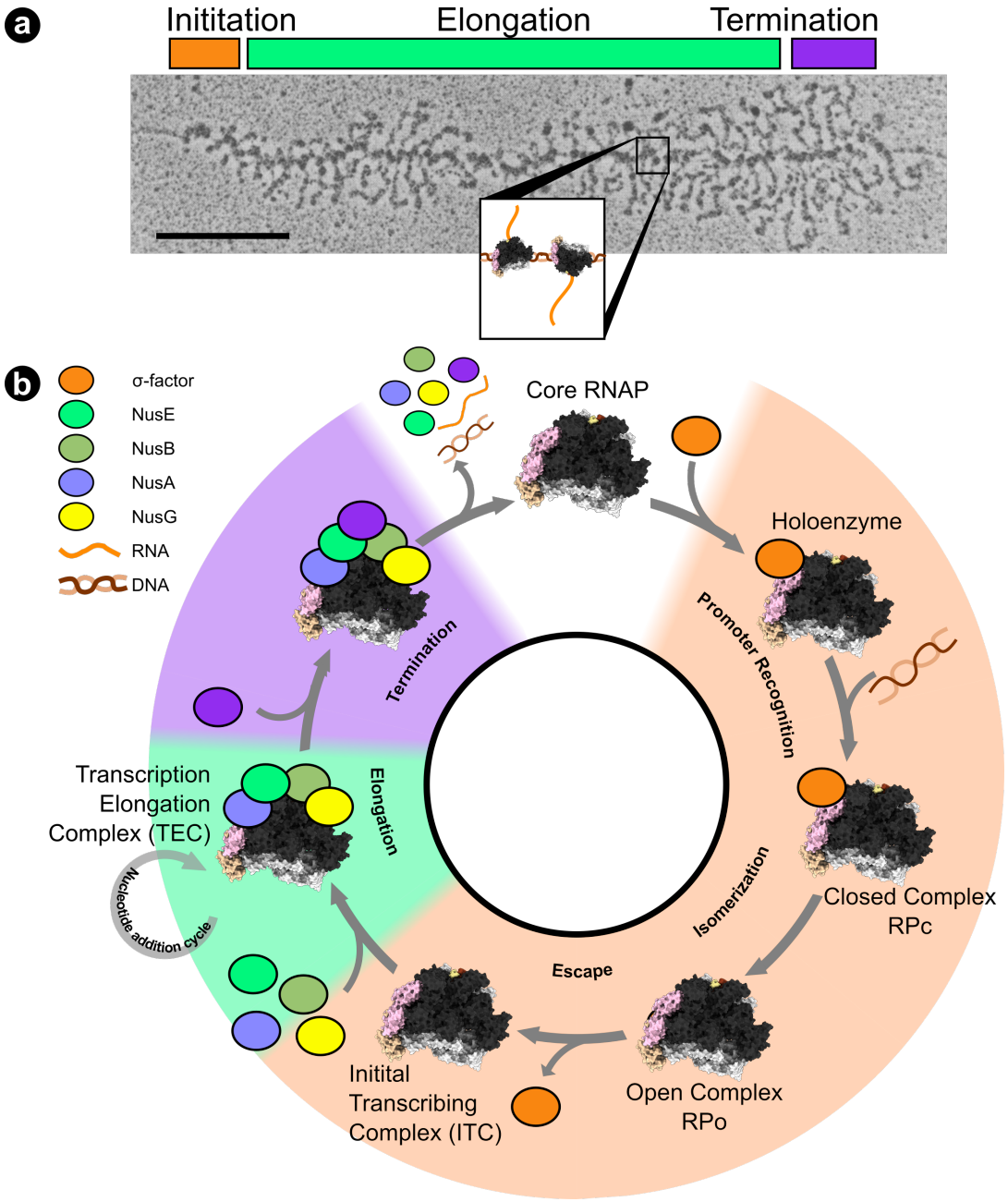



**Figure 1.14: Transcription cycle.**
***(a)*** *The three steps of transcription, i.e. initiation, elongation, and termination, are depicted using an electron microscopic image of the Miller Tree adapted from* (125)*. The DNA strand runs from left to right. Individual RNAP molecules appear as "beads-on-a-string". Nascent RNA molecules at different positions of the transcribed gene are spread out to the upper and lower part. An inlet depicts a scheme of two RNAP molecules.* ***(b)*** *Schematic overview of the transcription cycle depicting all steps for a whole round of transcription (PDB ID: 6ALH)* (102)*.Further details are described in the text. The three steps are highlighted with the same colors as in (a). Binding and dissociation events of the different regulatory factors are depicted with arrows and colored ellipsoids, respectively. The nucleotide addition cycle (NAC) is indicated at the elongation step.*

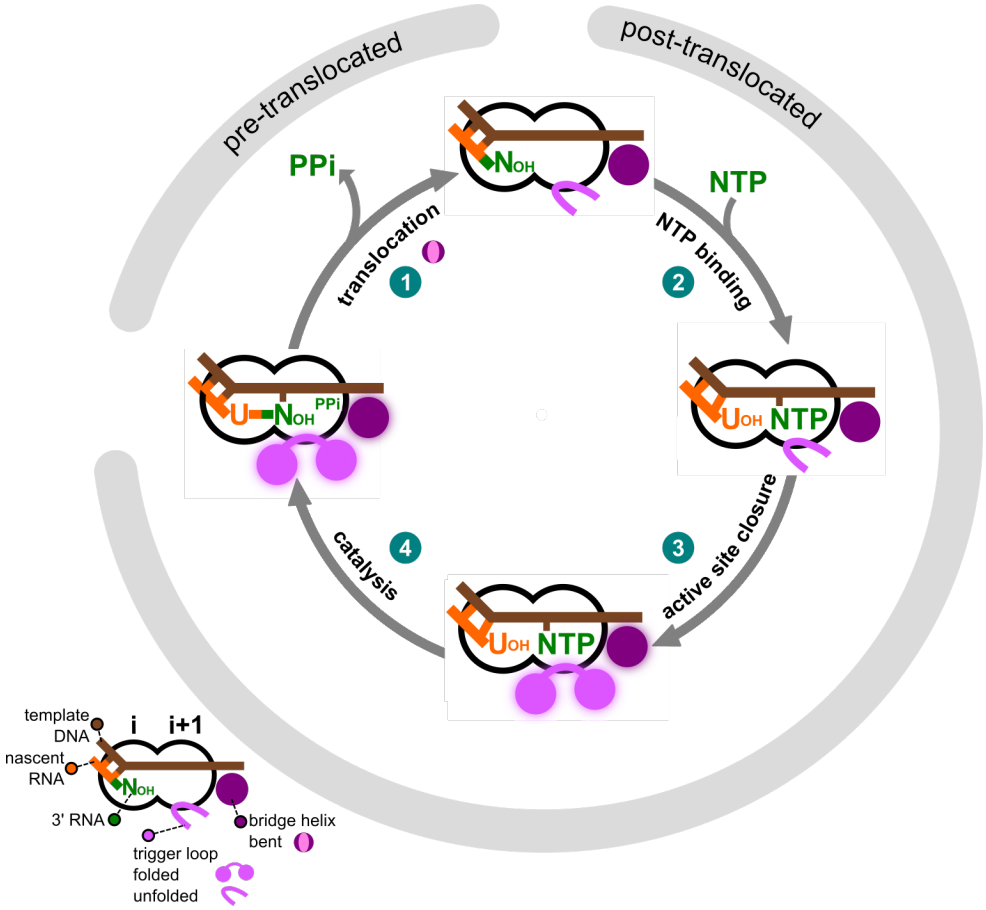



**Figure 1.15: Nucleotide addition cycle.**
*Steps of the nucleotide addition cycle including (1) translocation, (2) NTP binding, (3) active site closure and (4) catalysis. A legend of the active site is provided in the lower left corner. The sites i and i+1 are depicted as two intersecting spheres. The template DNA and the nascent RNA are depicted in brown and orange, respectively. NTP and PPi are colored in green. The added nucleotide monophosphate is abbreviated as $N_{OH}$. The RNA's OH ends are depicted. The grey outer shell illustrates the pre- and post-translocated state, respectively.*

In the pre-translocated state, the last added nucleotide is positioned inside the *i+1* site. In a process termed translocation, the last nucleotide moves from the *i+1* site to the *i* site ((1) in Figure 1.15). This movement is accompanied by bending and unbending of the BH, the TL unfolding, and the release of pyrophosphate (PPi) (126).

Once the post-translocated register is reached, a new incoming NTP can enter the vacant *i+1* site ((2) in Figure 1.15). Afterwards, the TL folds into the two THs. The THs together with the BH form the three-helix bundle and thereby closes down the active site ((3) in Figure 1.15). In the catalytical step, the nascent RNA is extended by one nucleotide monophosphate and one PPi are generated ((4) in Figure 1.15). This step requires two $Mg^{2+}$ ions (I and II) within the active site (Figure 1.16).

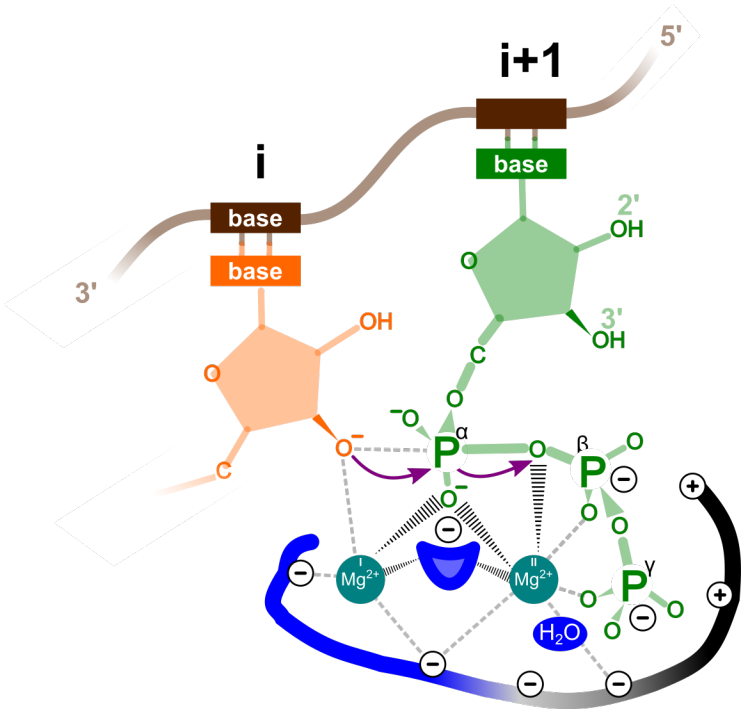



**Figure 1.16: The $S_N2$ mechanism of nucleotide addition.**
*In the post-translocated TEC, the 3' RNA end (orange) is located at the i site and the incoming substrate NTP (green) at the i+1 site. The substrate NTP has successfully base paired with the acceptor template base and the triphosphate reaches into the substrate pocket. The two $Mg^{2+}$ ions are labeled as I and II and are represented as teal spheres. The high-affinity $Mg^{2+}$ ion I is coordinated by the catalytic triad residues (blue). The low affinity $Mg^{2+}$ ion II is bound to β and γ phosphates of the NTP and coordinated by additional residues of the pocket (grey and black). The transfer of electrons during the $S_N2$ nucleophilic substitution are indicated by purple arrows. The different regions within the substrate binding pocket are colored in blue, grey, and black, respectively.*

$Mg^{2+}$ ion I lowers the affinity of the 3' OH for the hydrogen, thereby facilitating the $S_N2$ nucleophilic attack of the 3' $O^-$ onto the α phosphate. $Mg^{2+}$ ion II assists in stabilizing and leaving of the pyrophosphate. Both ions stabilize the structure and charge of the trigonal-bipyramidal transition state of the α phosphate. This state is universally conserved for all polynucleotide polymerases (127, 128). By adding one nucleotide, the RNA:DNA hybrid is extended by one base, resulting in a pre-translocated register. Subsequent relaxation into the post-translocated state involves the release of PPi, opening of the active site by unfolding of the TL, and translocating the RNAP upstream along the template DNA (4 in Figure 1.15) (119, 129–131).

**Termination.** Transcriptional termination fulfills four functions (132) as part of the overall gene regulation:

- Prevent inappropriate expression of downstream genes and interference of antisense transcripts.
- Define 3' ends for precise RNA structures formation and regulation.
- Recycle RNAP for efficient gene expression.

- Minimize collision between the TEC and the replication complex, causing damaging double-strand breaks.

Unsurprisingly, termination events are heavily controlled, since pre-mature termination has severe consequences for the cell. Programmed transcription termination in bacteria can be classified into two pathways: intrinsic termination and factor-dependent termination (112, 132).

For **intrinsic termination** to occur, DNA and RNA elements at the termination site direct a sequence of events to cause termination of the TEC within a discrete location of 2-3 nt. This releases the RNAP and the transcript from the chromosome. The canonical intrinsic terminator sequence can be defined by two elements: (i) a guanosine and cytosine (GC)-rich dyad that forms the termination hairpin, located 7-8 nt from the 3' end of the transcript and (ii) a partially conserved 7-8 nt U-rich tract (132, 133). The process of intrinsic termination was separated into four subsequent steps (132): (i) Pause at the 3' end of the U-rich tract. (ii) Nucleate the termination hairpin. (iii) Complete the hairpin and inactivate the TEC. (iv) Dissociate the RNAP from both the DNA template and the RNA transcript.

However, the conformational rearrangements within the RNAP during termination remain poorly understood. Already the order of RNA and DNA release remains unknown.

The process of **factor-dependent termination** is best described for termination factor ρ. The ρ-dependent termination relies on the recognition of specific signals within the nascent RNA by ρ. In general, such signals consist of a pyrimidine-rich RNA element called the ρ utilization (*rut*) site, stretching over a length of 30 nt. (112, 134).

Afterwards, RNA is threaded through the central pore of the ρ factor to trigger its translocase activity. Thereby, all upstream RNA is translocated leading to its dissociation from the RNAP. This process occurs in four steps (132):

(i) Recognize and bind the ρ-utilization elements within the RNA transcript emerging from the exit channel. The *rut* sites are constituted of C-rich and G-poor sequences lacking extensive secondary structures and are located upstream of the termination site with a length of 60-100 nt (135–139).

(ii) Close the ρ-ring and activate the catalysis by binding of *rut* and ATP (140). Afterwards, ρ translocates along the nascent RNA under NTP hydrolysis (141), while the *rut* sites remain stably bound to ρ. Thereby, a RNA loop with growing size is produced. This model is known as the tethered-tracking model (142–144).

(iii) ρ and the TEC are kinetically coupled to each other, describing the competition between translocation of both (145). TECs are being most susceptible to ρ at specific long-lived pause sites.

(iv) Release RNA and dissociate the TEC once ρ contacts the exit channel, likely inducing conformational changes within the nucleic acid scaffold and the RNAP itself. Two models for the reorganization of nucleic scaffold have been proposed: hypertranslocation (146) and hybrid-shearing (147). Both models are based on ρ's motor ability (148).

**Anti-termination.** The termination of a transcriptional event can be modulated by a variety of factors in form of proteins, small molecules, or RNA structures. Such regulation can either enhance termination or induce anti-termination. Anti-termination results in the overwriting of a termination signal, leading to the transcription of genes positioned downstream of the termination signal (112). A wide variety of mechanism exists for viral proteins inducing anti-termination of the bacterial RNAP (149).

Anti-termination can be divided into *cis*-acting and processive. Examples for *cis*-acting anti-terminators are riboswitches, which can adopt different, mutually exclusive conformations, leading to either termination or anti-termination of the RNAP. Here, conformational changes of the riboswitch can be induced by proteins, small molecules or growth conditions (150).

Processive anti-termination, on the other hand, involves the direct regulation of RNAP by protein factors, enabling the RNAP to read through termination signals while transcribing along the DNA (112). In the following, examples for processive anti-termination will be given. The two best-studied cases of anti-termination are the λN-mediated and λQ-mediated ones, both of which have been derived from the bacteriophage λ.

**Processive anti-termination in bacteriophage λ.** Bacteriophage λ, is a temperate phage that infects *E. coli* host cells. It was first discovered and isolated in 1951 (151). A bacteriophage particle possesses a characteristic morphology consisting of an icosahedral head and a long, flexible tail (Figure 1.17 a). The head contains the linear dsDNA genome (48,502 bp) and codes for around 50 proteins (152). Gene regulation within and by bacteriophage λ was studied in great detail and it became a model system for bacteriophage-host studies (153).

Upon infection of its host, bacteriophage λ adopts two alternative pathways of development: lytic or lysogenic. During the lytic pathway, the host cell replicates bacteriophage DNA, bacteriophage proteins are translated, new bacteriophage particles

are assembled, and released by lysis of the host cell. During the lysogenic pathway, on the other hand, the bacteriophage DNA is integrated into the genome of the host cell. In this prophage state, the DNA of the bacteriophage is replicated within the host cell until transition to the lytic pathway occurs. Such a transition is induced by less-favorable environmental conditions. The precise sequential execution of this event is of utmost importance to avoid premature cell lysis (154). Therefore, a system of successive expression of genes evolved based on tight transcriptional control. Three stages of gene expression are performed, which involve immediate-early, delayed-early, and late genes.

The immediate-early genes (inner orange circle in Figure 1.17 b) encode various functions important for the activation of either the lytic or the lysogenic pathway. Once the lytic pathway is activated, bacterial RNAP initiates transcription from the early promoters $P_L$ and $P_R$, located within the control region of the λ genome (brown circle in Figure 1.17 b). During this first round of transcription the gene *N* is expressed before transcription terminates at the termination sites $tL_1$ and $tR_1$.

During the expression of delayed-early genes, λN-modified RNAPs (red arrow) initiate transcription at the same promoters $P_L$ and $P_R$ but continue further downstream, reading through $tL_1$, $tR_1$, other intrinsic, and ρ-dependent terminators (not shown). This leads to the expression of delayed-early genes located downstream of $tR_1$ and $tL_1$ termination sites. In the right direction, transcription processes into the *Q* gene and is terminated at the tR' termination site.

The expression of late genes is ensured by the λQ-mediated anti-termination (blue arrow). Therefore, λQ recognizes the DNA signal QBE (Q-binding element) during initiation at the $P_{R'}$ promoter, leading to a read-through at the terminator tR'. This last stage allows the expression of genes required for lysis and assembly of head and tail, thereby finalizing the lysogenic pathway.

In the following section, several bacterial and viral protein factors, influencing the RNAP during the transcription cycle will be introduced.

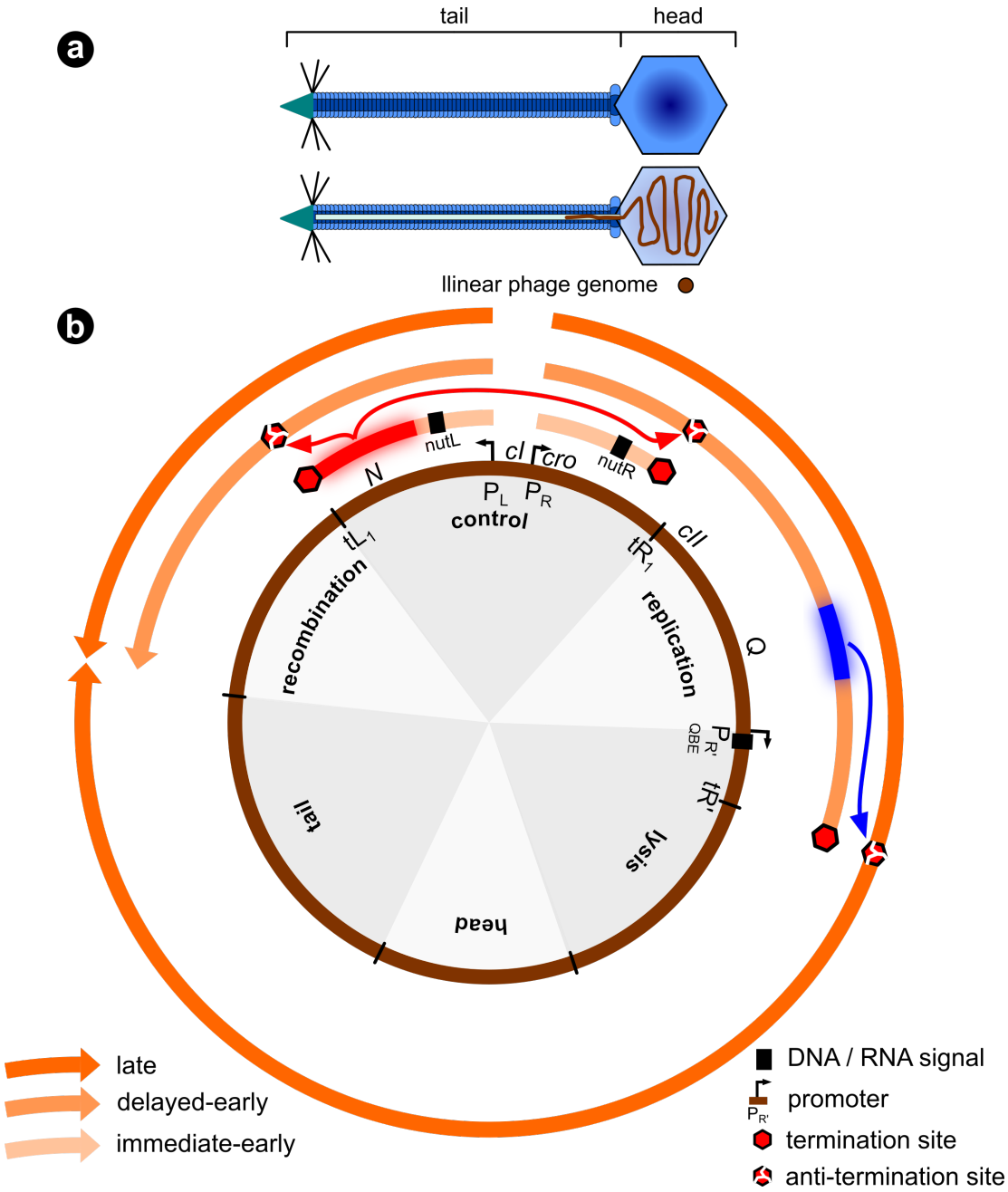



**Figure 1.17: Morphology and gene expression of bacteriophage λ.**
***(a)*** *Schematic overview of the bacteriophage λ in an opened and closed view. The head contains the genome, which is transferred into the host cell via the tail.* ***(b)*** *Simplified chromosome organization and temporal regulation of gene expression by bacteriophage λ. The inner circle (brown) represents the circularized DNA with labeled gene locations, showing important regulatory genes (N, cI, cro, cII and Q), early promoters ($P_L$ and $P_R$), a late promoter ($P_{R'}$), corresponding terminator sites ($tL_1$, $tR_1$ and tR'), and regulatory sequences (*nutL *and* nutR *within the RNA and QBE within the DNA as black boxes). Three orange circles represent the transcribed mRNA at the different stages immediate-early, delayed-early, and late. The relative positions of the genes N and Q as well as their anti-terminator effect are depicted.*

### 1.3.3 Bacterial regulatory factors

Due to its wide importance for survival and proliferation, transcription is a process tightly regulated by proteins (130), RNAs (155), and small molecules (156). Here, some examples of the broad pool of regulatory protein factors will be introduced, which are important to understand the transcriptional cycle.

**σ factors.** The σ factors (σ) represent key players in the process of transcriptional initiation. They associate with the RNAP forming the holoenzyme and perform the steps of initiation. Here, the domains of the $\sigma^{70}$ factor will be introduced, since it represents the best studied one. The protein $\sigma^{70}$ is required for initiation but is unable to recognize DNA alone. $\sigma^{70}$ can be separated into four domains connected by flexible linkers: $\sigma^{1.1}$, $\sigma^{2}$, $\sigma^{3}$, and $\sigma^{4}$ (Figure 1.18 a).

**NusG.** The protein NusG is essential for cell viability and represents the only universally conserved transcription factor (157). It consists of two domains, $NusG^{NTD}$ and $NusG^{CTD}$, connected by a flexible and extended linker (Figure 1.18 b) (158). The $NusG^{NTD}$ consists of four anti-parallel β-strands enclosed by three α-helices comprising the NusG amino-terminal (NGN) motif. The $NusG^{CTD}$ comprises a β-barrel consisting of five anti-parallel β-strands (159) containing the Kyprides-Onzonis-Woese (KOW) motif, related to ribosomal proteins (160).

In *E. coli*, NusG displays several different functions: NusG increases elongation rates by reducing pausing and inhibiting backtracking (112, 161, 162). It supports anti-termination in the presence of λN (163) and within ribosomal *rrn* operons (112, 161). It enhances termination by the transcription termination factor ρ (164, 165). $NusG^{NTD}$ binds to the RNAP, contacting the β' coiled helix (β'CH) and the β gate loop (βGL) (162, 166–168). The NTD thereby closes down the clamp, locking the RNAP in a closed conformation, which helps to stabilize the elongations state of RNAP (107). Additionally, the NTD separates the upstream and downstream portion of the template strand (166, 169). The $NusG^{CTD}$ was shown to recruit additional regulatory factors. For instance, interaction of CTD with the termination factor ρ takes place during ρ-dependent termination (170). CTD interacts also with NusE, coupling transcription and translation (171), or within the λN-TAC (172).

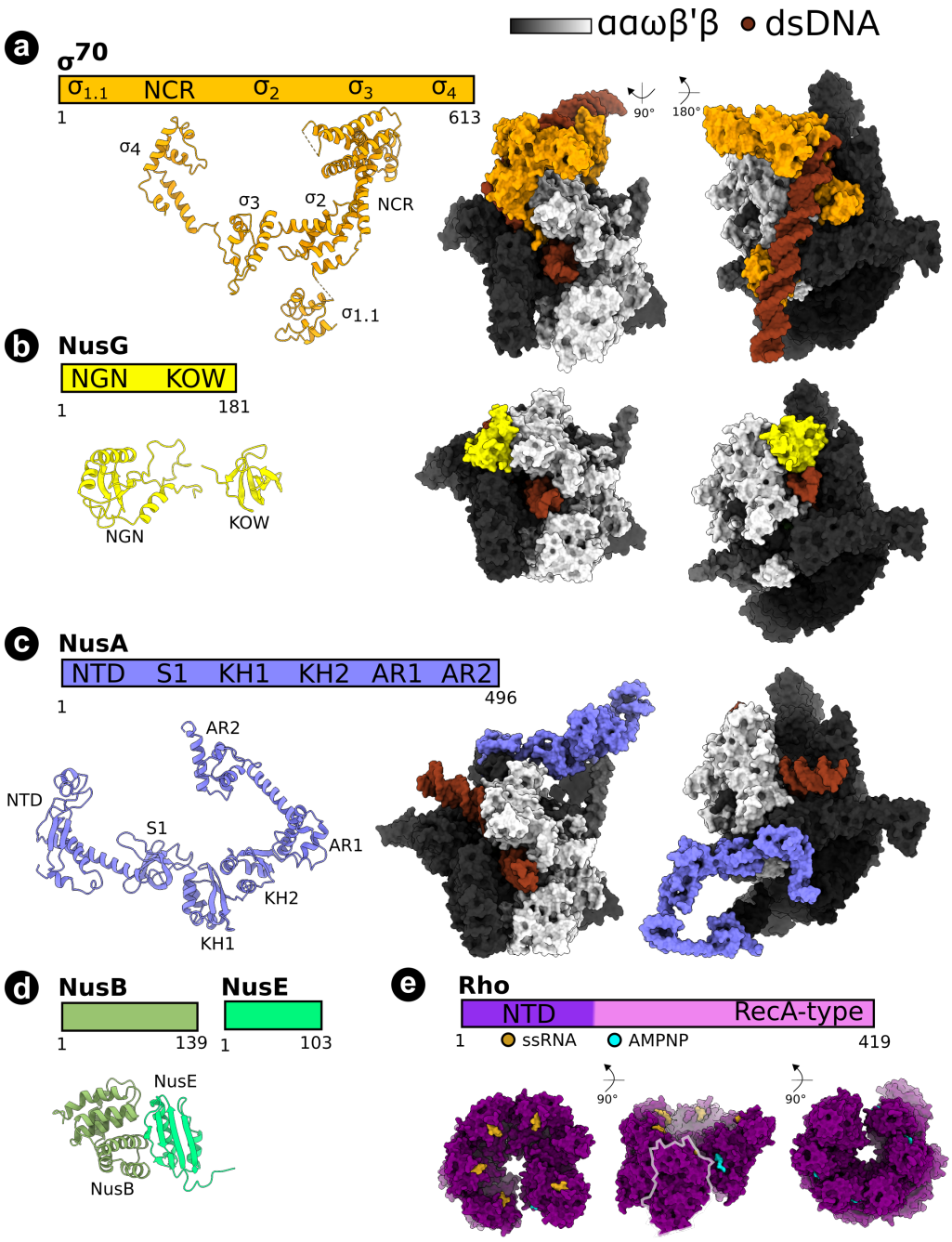



**Figure 1.18: Overview of bacterial regulatory factors at the RNAP**
*Depicted are the 1D structure including the domain organization, the 3D structure, and the complex of the bacterial regulatory factors bound to the RNAP if available:* ***(a)*** *$\sigma^{70}$ (PDB ID: 4JK1)* (173), ***(b)*** *NusA (6FLQ)* (174), ***(c)*** *NusG (6C6U)* (162), ***(d)*** *the heterodimeric complex NusB:NusE (3D3B)* (172), *and* ***(e)*** *the hexameric protein ρ (1PVO)* (175).

**NusA.** NusA is the largest and, besides NusG, the most conserved elongation factor. It binds shortly after initiation (176) and competes with the $\sigma^{70}$ factor for the binding interface (95, 177). It is involved in a large variety of regulatory functions, such as (i) the formation of stable TECs together with other Nus-factors, (ii) the support of read-through at termination sites (178), and (iii) the recognition of pausing and termination signals within the RNA (174, 179, 180). When bound to the RNAP, NusA interacts with the βFTH and is in close proximity to the RNA exit channel (Figure 1.18 c) (174). In *E. coli*, NusA comprises 495 amino acids (aa) (55 kDa) and can be separated into 6 domains: NTD, S1, KH1, KH2 and Acidic Repeat 1 and 2 (AR1 and AR2, respectively).

- The $NusA^{NTD}$ interacts with the βFTH of the RNAP in proximity to the RNA exit channel (168, 179, 181, 182) and the $\alpha^{CTD}$ (174). It was suggested the NTD and the ßFTH rearrange the catalytic site in an allosteric fashion to induce pausing (183–185). These findings are further supported by the observation, that deletion of the ßFTH abolishes pausing effects of NusA (111). Furthermore, the $NusA^{NTD}$ is involved in termination and pausing events by interacting with pause or terminator stem loops emerging from the RNA exit channel (174, 179).
- The three domains $NusA^{S1}$, $NusA^{KH1}$ and $NusA^{KH2}$ are often named the $NusA^{SKK}$ super domain (186, 187) and are connected via a highly conserved linker helix. All three are homologs to known RNA-binding proteins (187–189) and hold ssRNA binding activity alone and in complex with the RNAP (179, 190). They are all needed for efficient intrinsic termination (191). S1 is located close to the RNA exit channel, hovering above the exiting RNA likely to support RNA folding (174, 178). The domains KH1 and KH2 constitute one of the most abundant RNA-binding motifs (192) and are involved in transcriptional and translational regulation (193). In NusA, the two conserved GxxG loops are responsible for ssRNA binding (190).
- The domains $NusA^{AR1}$ and $NusA^{AR2}$ each exhibit two helix-hairpin-helix motifs (194–196). In general, such motifs have been found to perform DNA binding (197) and protein-protein interactions (198). Both AR domains were so far only detected in γ-proteobacteria and a few other species and do not display high sequence conservation (199). AR1 and AR2 have different functionalities:
    - $NusA^{AR1}$ was shown to bind λN *in vitro* (191, 195).
    - $NusA^{AR2}$ was proposed to possess two related functions: (i) It inhibits RNA binding by masking the SKK domain in an auto-inhibitory fashion (199, 200). (ii) The release of that auto-inhibitory state was shown to involve the binding of $NusA^{AR2}$ to the $\alpha^{CTD}$ (200). Alternatively, the λN protein may induce its release during anti-termination (194, 199).

In conclusion, NusA is involved in a wide range of different regulatory processes, making it a versatile tool in the arsenal of transcriptional control.

**NusE.** NusE was first described as ribosomal protein S10 as part of the 30S subunit where it is incorporated during late steps of the ribosome biogenesis (201). Additionally, NusE was also described in the context of transcriptional anti-termination (202). NusE (11.8 kDa) adopts a four-stranded anti-parallel β-sheet fold, stabilized by two α-helices on each side (light green in Figure 1.18 d) (172). NusE is unstable in isolation (203, 204), but forms a stable heterodimer with NusB in the context of the λN-TAC (171, 172). It was suggested that transcription-translation coupling is achieved via an interaction between NusG and NusE (171, 205).

**NusB.** The protein NusB is involved in transcription and translation (201). NusB possesses an all-α-helical fold with two perpendicular three-helix bundles, which have no close structural homologs (olive green in Figure 1.18 d) (206, 207). The monomer has medium affinity to specific ssRNA sequences (208, 209), while the NusE:NusB heterodimer exhibits high affinity (209, 210). This increase in affinity was explained with the formation of a mosaic RNA-binding interface upon dimer-formation (172). While NusB is dispensable for the formation of λN-TAC (172, 211, 212), it is required during anti-termination within the context of bacterial ribosomal *rrn* operons (213).

**ρ factor.** The RNA/DNA helicase ρ is required for factor-dependent termination. ρ is a homohexameric RNA translocase of the RecA-family (Figure 1.18 e). Different structures of ρ were resolved, among them one bound to ssDNA and another to ssRNA and a non-hydrolysable analog of ATP (175).

The protomer is peanut shaped and can be divided into an $\rho^{NTD}$ and a $\rho^{CTD}$ connected by an extended 30 aa linker. The NTD consists of two subdomains, a three-helix bundle and a five-stranded β barrel adopting an oligonucleotide binding (OB)-fold often occurring in single-stranded nucleic acid binding proteins (192).

ρ possesses two RNA binding sites: a primary RNA binding site, which recognizes the *rut* sequence on the nascent transcript and a secondary RNA binding site required for RNA translocation. The primary RNA binding site is formed by an intended interface on one side of this subdomain. The $\rho^{CTD}$ consists of a RecA-type ATPase binding fold and a small helical subdomain. Three loops combined are forming the secondary RNA binding site and make up the central pore within the hexamer.

### 1.3.4 Viral regulatory factors

Several bacteriophages evolved small proteins, which hijack the host transcriptional apparatus in order to control different steps of the transcription process (149): (i) Hijacked initiation is triggered by the proteins gp36, gp76, or gp2, which all diminish the initiation on host cell promoters and stimulate transcription of bacteriophage genes (149, 214, 215). (ii) Hijacked termination is performed by Nun to compete with other bacteriophage proteins. (iii) Hijacked anti-termination leads to the transcription of bacteriophage genes, located further downstream of the usual stop signal. In recent years, several of such interactions have been elucidated using structural data and a few examples will be illustrated here.

**Nun.** The 13 kDa protein Nun from bacteriophage HGK022 excludes superinfection of the viral bacteriophage λ (216) by arresting the TEC specifically on λDNA (217, 218). Therefore, Nun competes and antagonizes the function of λN by provoking premature downstream termination (216). Nun performs its function on the RNAP alone or in the presence of the factors NusA, NusB, and NusE (216, 218, 219). It is a 112 aa long intrinsically disordered protein (IDP) and can be separated into two domains (Figure 1.19 a): (i) The NTD contains an arginine rich motif (ARM) which binds specifically the *boxB* RNA hairpin within the emerging *nut* transcript (220–222). (ii) The CTD (aa 87-112) interacts with the TEC and the DNA within, to block translocation at intrinsic pause sites without changing the RNAP's catalytic activity and dissociating the complex (217, 218). Recently, the structure of the $Nun^{87\text{-}112}$:TEC complex was elucidated, explaining how Nun interferes with translocation and how primarily paused TECs are affected (102). However, the structure of Nun with the RNAP and other Nus factors has not been elucidated yet.

**λQ.** At the late stage during gene expression of bacteriophage λ, the protein λQ acts as a transcriptional anti-terminator allowing the expression of the late operon and thereby completion of the lytic cycle (223). Protein λQ is a globular, helical DNA-binding protein (224–226), recognizing the Q binding element (QBE) in the late operon promotor, and subsequently stays bound to the TEC (Figure 1.19 b). The structure of the protein λQ, in complex with the QBE and the RNAP elucidated the following mechanism: protein λQ forms a torus, termed a "nozzle", extending and narrowing the RNA exit channel. The nascent RNA is threaded through the constricted channel, precluding the formation of pause and terminator RNA hairpins (224).

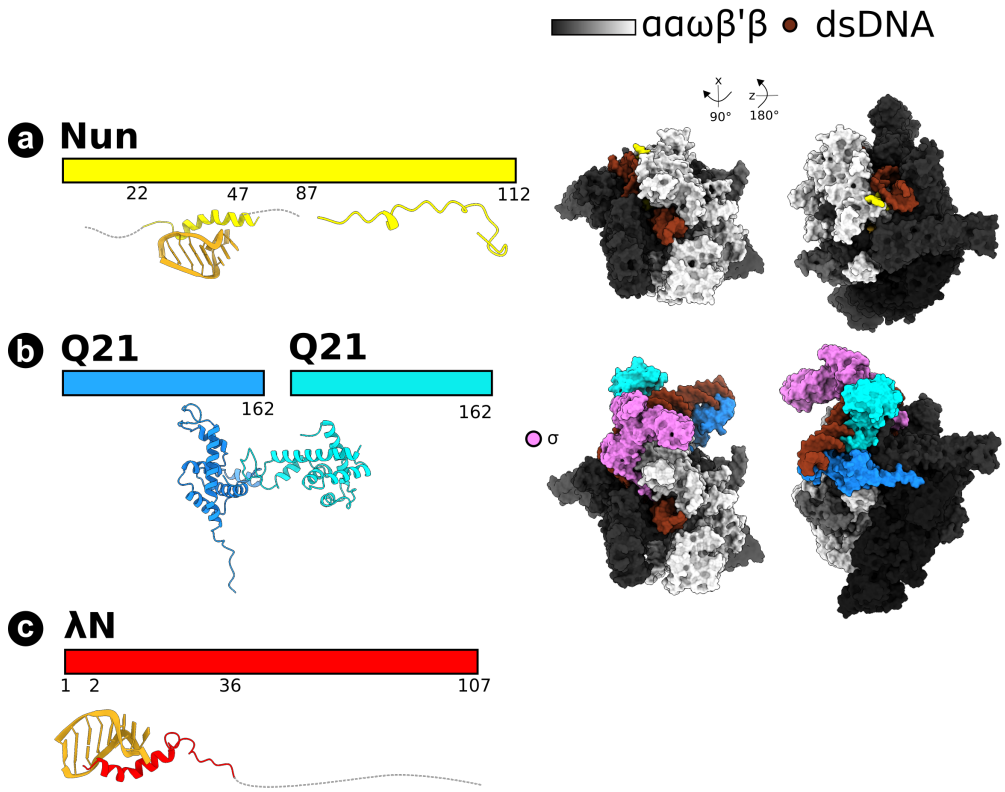



**Figure 1.19: Overview of viral regulatory factors.**
*Depicted are the 1D structure, the 3D structure and the complex of the viral factor bound to the RNAP.* ***(a)*** *Nun is colored in yellow and makes contacts to the upstream and downstream DNA (PDB ID: 6ALG)* (102). ***(b)*** *λQ forms as dimer and makes contacts to the RNA exit channel, the DNA, and the σ factor (pink) (6P18)* (224). ***(c)*** *λN binds RNA at its N-terminal end (1QFQ)* (222).

**λN.** The bacteriophage λ protein N (λN) controls transcription during lytic growth by inducing an anti-termination event depending on a specific nascent RNA site called *nut* (N utilization) and the host elongation factors NusA, NusB, NusE, and NusG (227). When assembled, this complex is named the λN-mediated transcriptional anti-termination complex (λN-TAC). Three elements make up the *nut* RNA: a 5' *boxA* element, a spacer sequence and a 3' stem-loop called *boxB* (228). Since its discovery, the mechanism of λN-mediated anti-termination has been studied extensively (152). λN is a 107 aa long, highly basic IDP protein (Figure 1.19 c) (229, 230). Based on biochemical evidence, λN can be separated into three functionally distinct domains:

- Residues 1-22 make up an arginine-rich motif (ARM), which binds *nut boxB* with high affinity (231–233). This binding event leads to the formation of a stabile α-helical fold as shown by NMR studies (222, 234).
- Residues 34-47 bind $NusA^{AR1}$ with only medium affinity (195, 235).
- Residues 73-107 were found to form the RNAP binding region (227, 236).

The entire λN-TAC is assembled in an cooperative manner via a multitude of weak binary interactions between the components and the surface of the RNAP (227). Since its

identification in 1964, λN has presented a remarkable example on how a small protein hijacks the RNAP in a manner that allows to overwrite intrinsic and ρ-dependent termination signals (237, 238). Despite decade-long efforts, the structure of the λN-TAC has not been elucidated yet.

## 1.4 Neuronal development, the ribosome and Ebp1

### 1.4.1 Neuronal development

The physical basis for consciousness, complex cognition, language, and coordination of voluntary motor activity in mammals is located within the adult neocortex (239, 240). Neocortical functions arise from the activity of two main cell types: glia and neurons. Neurons within the neocortex are divided into two main classes: interneurons and primary projection neurons.

Interneurons are inhibitory GABAergic cells that migrate tangentially into the developing neocortex, exhibiting a distinguished cellular morphology and the expression of specific markers (241, 242). Primary projection neurons, on the other hand, are excitatory glutamatergic cells which perform the majority of signaling within the neocortex and expand processes over long distances. The neocortex consists to 75-85 % of excitatory projection neurons and was divided into six distinct layers (I-VI), based on cell morphological features and axonal targets that define distinct functional circuits (Figure 1.20).

The development of the neocortex encompasses the transition of a pool of neural stem cells into intricately branched neurons forming the layered tissue structure. In mammalian species, the neurogenic phase of stem cell maturation in neocortical development follows a conserved course (242–244) (Figure 1.20). The lateral cortical ventricular zone is lined with neural stem cells (NSC) (3), which undergo symmetric divisions. The newly formed pool of cells represents the cellular basis of the cortical plate. Afterwards, NSC division continues asymmetrically and newly born neurons migrate superficially. These cells ultimately form a layered cortical plate composed of structurally and functionally distinct neurons. The majority of subcortically projecting lower layer neurons are born first, while intracortical projecting upper layer neurons are born second. In mice, lower layer (IV-VI) neocortical neurons are born around embryonic day 12.5 (E12.5), while upper layer neurons (II-III) are born around E15.5.

Neurogenesis is largely completed at postnatal day 0 (P0), the day of birth, and the ongoing stem cell divisions result in glial lineage cells (Figure 1.20).

The projection architecture of each layer defines its distinct functional connectivity, where dendritic inputs and axonal outputs configure the neocortical circuits (245). The establishment and maintenance of these circuits requires fine-tuned proteostasis in the neurite outgrowth and the synapse. Proteostasis, protein homeostasis, at the cellular level is necessary for a precise regulation of cell adhesion molecules to establish neuronal connectivity (246).

Previous studies focused largely on the transcriptional regulation of the developing neocortex. Subgroups of neurons have been defined based on the expression of transcription factors (TFs) (243). Differentially expressed TFs represent one major driver of neuronal fate during differentiation.

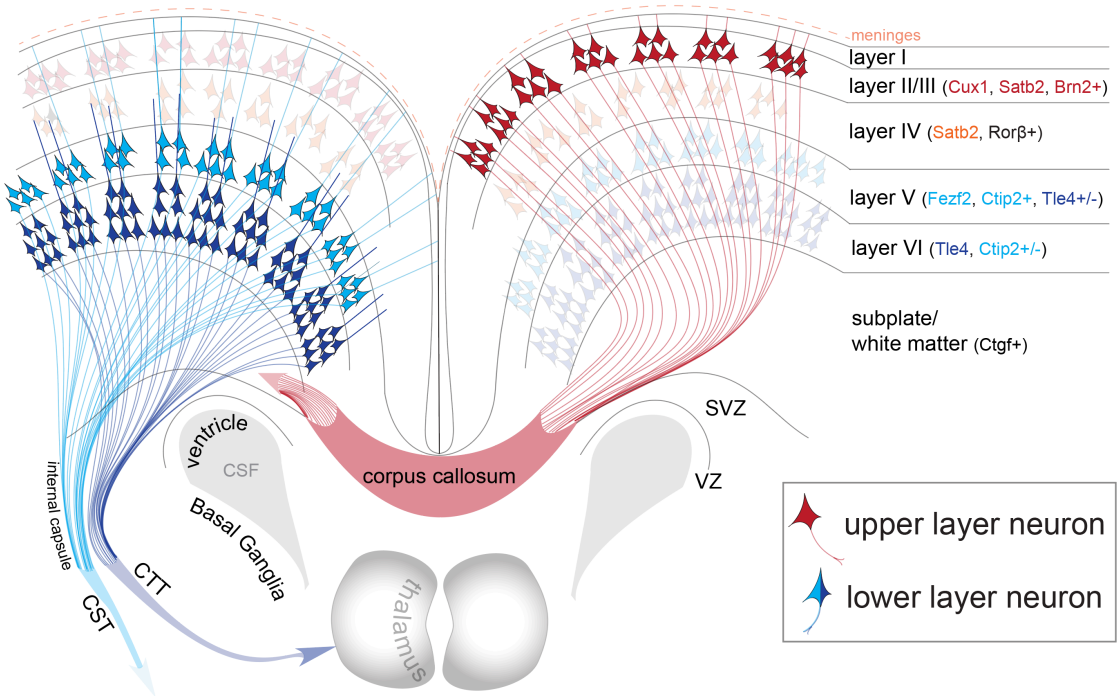



**Figure 1.20: A schematic overview of the neocortex development.**
*The neocortex is a highly organized layered structure. Six layers are defined by neuronal subpopulations of glutamatergic projection neurons, which represent approximately 85 % of all neocortical neurons. These subpopulations of projection neurons are characterized by specific molecular identities, dendritic morphologies, and terminal targets corresponding to each layer. Projection neurons that are born during early stages of prenatal neurogenesis will be predominantly placed in lower layers (blue), while those born later will migrate into upper layers (red). Image adapted from* (242).

However, in recent years a more complex picture for neurodevelopment arose, with post-transcriptional regulation increasingly appreciated as a major driver of neuronal development (242). For example, RNA-binding protein (RBPs) are involved as important regulators of all post-transcriptional steps (splicing, localization, stability/decay, and translation) (242, 247). As the final step of gene expression, translation is executed by the ribosome. Recent findings place the ribosome in a dynamic and executive role at the junction of cellular proliferation, differentiation, and disease (248–250).

### 1.4.2 Ribosome and proteostasis

The mammalian ribosome is a large macromolecular machine consisting of 4 rRNA species and at least 79 ribosomal protein (RPs). The eukaryotic 80S ribosome can be found in the cytoplasm and is composed of two subunits: the 60S large subunit and the 40S small subunit (251). The ribosome translates mRNAs into polypeptide chains, where each triplet of nucleotides within the mRNA is interpreted by one tRNA, charged with a specific amino acid.

The process of translation is divided into four steps: initiation, elongation, termination, and recycling (251, 252). The 40S subunit initiates translation of the mRNA via a 5’ cap-dependent

or cap-independent mechanism. When the 40S and the 60S subunits join, the 80S ribosome (monosome) is formed. Several monosomes engage on the same mRNA molecule, forming a polysome. Polysomes represent the predominate species performing active elongation within the cell. Active translation ends with termination and subunit recycling. The ribosome is regulated throughout translation by a multitude of eukaryotic initiation (eIF) and elongation (eEF) factors. Additionally, the ribosome traverses through several structural intermediates defined by an energy landscape (5, 253). Ribosomal co-factors are themselves extensively regulated by post-translational modifications (PTMs) like phosphorylation (252, 254).

Proteostasis ensures the biological function of proteins by regulating their synthesis, folding, prevention of aggregation, and degradation. Ailments in proteostasis are involved in many metabolic, oncological, neurodegenerative, and cardiovascular disorders (255, 256). The process of translation represents the first regulatory point of proteostasis and therefore an important process in the development of the nervous system in general and the neocortex in particular (257). In case of abnormal proteostasis, neurodevelopment and neurodegenerative diseases often follow (258, 259). Therefore, a crucial understanding of proteostasis is important to understand neurogenesis in the neocortex. However, the architecture of ribosomal complexes specifically involved in neocortical development remain unknown.

### 1.4.3 Ebp1

Ebp1 (*E*rbB3 *b*inding *p*rotein *1*) belongs to the proliferation-associated 2G4 (PA2G4) family and has been linked to cellular development and cancer regulation (260). Ebp1 possesses a methionine aminopeptidase (Metap) “pita bread” structural fold (261–263) (Figure 1.21). This fold is characterized by a deep pocket on the concave surface (Figure 1.21 b) housing the active site. Metap proteins are essential metalloproteases in all kingdoms of life, cleaving off the first methionine once the nascent polypeptide chain emerges from ribosomal tunnel at a length of ~40 aa (264). Metap proteins were classified into two types, Metap1 and Metap2. Metap2 possess a helical subdomain of ~60 aa (orange in Figure 1.21). Ebp1 is Metap2-like; however, it is missing the essential residues for catalytic cleavage activity. Further distinctions from Metap2 are a shortened N-terminus, and an extended C-terminus containing six consecutive lysine residues.

Ebp1 is a highly conserved, broadly expressed and multifunctional protein in metazoans. Due to its ambiguous role in cancer progression, the question of “friend or foe” is still open (260). This ambiguity is founded on its two splice variants p42 and p48. While p42 (55-394 aa) functions as a tumor suppressor, p48 promotes cell proliferation (260). Ebp1 was originally linked to the ErbB3 receptor pathway (hence its name) (265). Interestingly, it has additionally been identified as an IRES *trans*-acting factor (ITA45) involved in the translation initiation on viral mRNA for the FMDV IRES (266). Moreover, Ebp1 binds to the ribosome, and inhibits the

phosphorylation of initiation factor eIF2α (267). Furthermore, Ebp1 is involved in transcriptional repression of some E2F-regulated promoters via binding to HDAC (268, 269).

So far, the protein Arx1, the 60S pre-ribosomal nuclear export factor from *Saccharomyces cerevisiae*, has been assigned as an Ebp1-related protein. Despite their sequence conservation both proteins exhibit different structures and putative functions. While Arx1 binds exclusively the 60S ribosomal subunit during maturation, Ebp1 stably binds 80S ribosomes *in vivo* as shown previously (270). The binding site and mechanism of Ebp1 on the ribosome as well as its role in neuronal development remains to be determined.

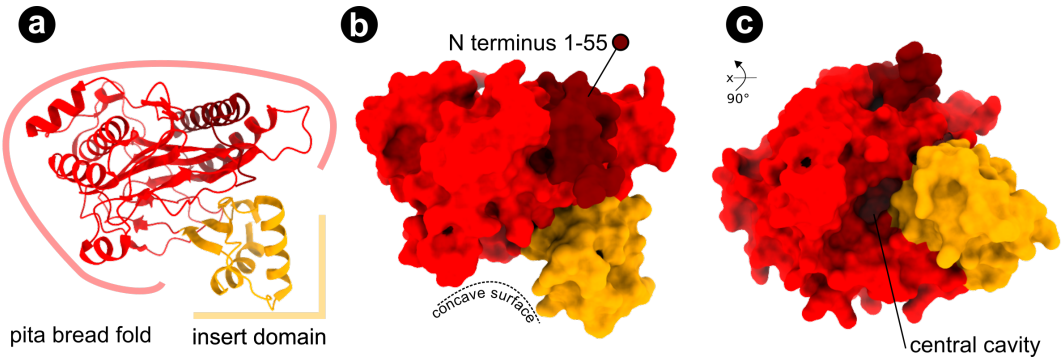


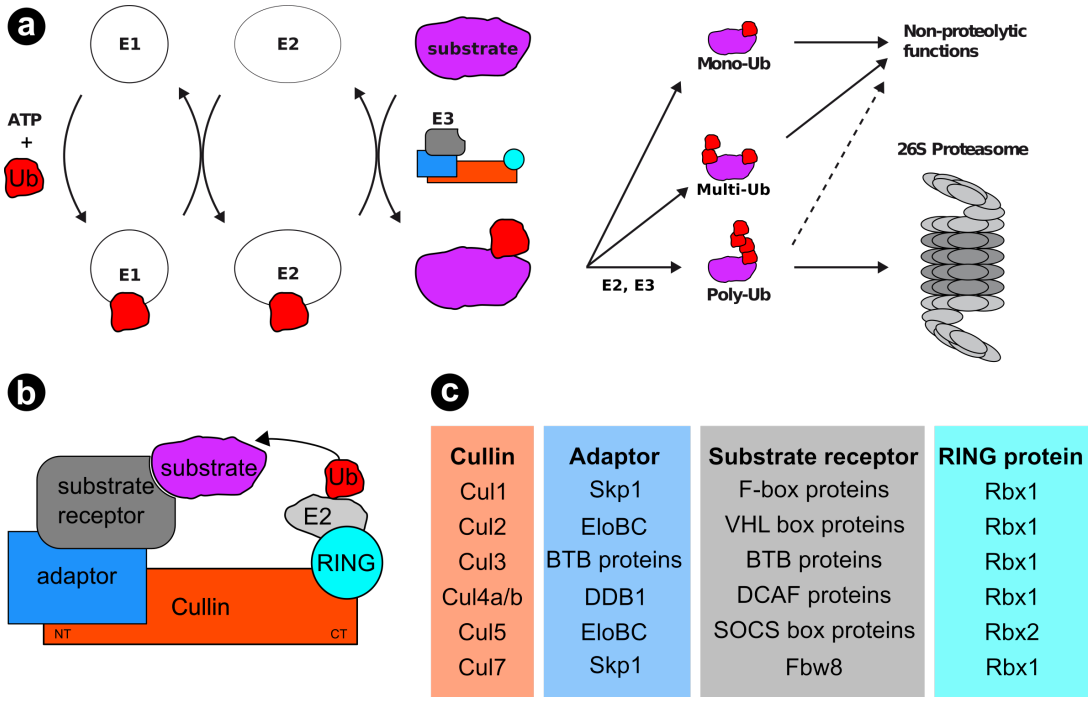

**Figure 1.21: Structure of Ebp1.**
*Side view of Ebp1 (PDB ID: 2Q8K)* (261) *in* ***(a)*** *ribbon and* ***(b)*** *surface representation, respectively. The N terminus (1-55 aa), the pita bread fold and the insert domain are colored in dark red, red and orange, respectively.* ***(c)*** *Bottom view of Ebp1 illustrates the central cavity.*

## 1.5 E3 ubiquitin ligase and viral hijacking

### 1.5.1 Ubiquitination and proteasomal degradation

Ubiquitin (Ub), a small 8.5 kDa protein conserved across all eukaryotes, regulates important cellular processes (271). It is named after its ubiquitous high concentration *in vivo*, e.g. 85 µM in human embryonic kidney cells (272). The covalent attachment of Ub and of Ub-chains of varying topology to a target protein is termed (poly-)ubiquitylation or (poly-)ubiquitination and has different downstream effects on the substrate protein: degradation by the 26S proteasome, mediation of protein-protein interactions, change of subcellular localization or regulation of enzymatic activity (273).

The process of ubiquitination requires as cascade of three enzymes: ubiquitin-activating enzymes (E1), ubiquitin conjugating enzymes (E2), and ubiquitin ligases (E3). The E1 transfers Ub to the E2 in an ATP-dependent manner, forming an E2-ubiquitin conjugate (E2~Ub). The E3 binds E2~Ub and at the same time specifically recognizes the substrate protein, to facilitate the transfer of Ub from the E2 to a lysine side chain of the substrate (Figure 1.22 a) (273).




| Cullin | Adaptor | Substrate receptor | RING protein |
|---|---|---|---|
| Cul1 | Skp1 | F-box proteins | Rbx1 |
| Cul2 | EloBC | VHL box proteins | Rbx1 |
| Cul3 | BTB proteins | BTB proteins | Rbx1 |
| Cul4a/b | DDB1 | DCAF proteins | Rbx1 |
| Cul5 | EloBC | SOCS box proteins | Rbx2 |
| Cul7 | Skp1 | Fbw8 | Rbx1 |

**Figure 1.22: Ubiquitination-proteasome system and CRL4.**
***(a)*** *Ubiquitin (Ub) is activated by E1 in an ATP-dependent reaction. Ub is transferred to an E2 ubiquitin conjugating enzyme (E2~Ub). An E3 ubiquitin ligase binds both, E2~Ub and substrate, in order to facilitate ubiquitin transfer resulting in substrate mono-ubiquitination (priming). Subsequent steps of ubiquitination can result in multi- or polyubiquitination of the substrate. Depending on the "Ub-code" on the substrate, different downstream events such as degradation by the 26s proteasome or various non-proteolytic activities can occur. Figure adapted from* (273) ***(b)*** *Schematic architecture of a CRL. The four main components of a CRL are denoted, as well as the substrate and the E2~Ub are depicted.* ***(c)*** *Table of the four protein components for each CRL family are listed.*

Substrates are typically modified with a single ubiquitin (mono-Ub), several ubiquitin-monomers (multi-mono-Ub), or several Ubs forming a chain (poly-Ub) (Figure 1.22 a). Ubiquitin chains are formed by the linking of several Ubs via their lysine residues (Lys11, 48 or 63) or their N terminus. Such chains are homogenous if the same residue is used throughout chain elongation, e.g. Met1-, Lys11-, Lys48-, or Lys63-linked chains. Furthermore, mixed topologies can arise from alternating linking patterns. Additionally, branched ubiquitin chains with unknown functions can also be generated. Based on the presence of mono-Ub, multi-Ub, or poly-Ub, different outcomes are triggered for the ubiquitinated protein in the cell. Therefore, ubiquitination was described as a code to store and transmit information (274).

The interplay of ubiquitination, de-ubiquitination, and ubiquitin-binding proteins influences the fate of the marked protein. Proteasomal degradation of the ubiquitinated substrate represents the most prominent pathway. Here, the most abundant Lys48-linked poly-Ub chains play an essential role. The ubiquitinated substrate is recognized by additional factors or directly by the proteasome, triggering conformational changes within the proteasome. Subsequently, substrate unfolding, de-ubiquitination, and translocation of the substrate towards the proteolytic site is performed with the help of several cofactors (275). Furthermore, Lys11-linked chains also bind proteasomal receptors, triggering degradation of cell cycle regulators during mitosis (274). The degradation of proteins on the plasma membrane occurs via lysosomes, involving mono-Ub or Lys6-linked chains.

Beside its destructive outcome, ubiquitination can also promote protein functions. For example, the attachment of mono-Ub to PCNA can recruit binding partners to chromatin in response to DNA damage (276). Activation of the transcription factor NF-κB is promoted by the proteosomal cleavage of its inhibitory domain (277). Furthermore, ubiquitination can also regulate protein localization, e.g. mono-Ub has been observed to lead to internalization of proteins from plasma membranes in yeast (278).

### 1.5.2 Cullin-RING ubiquitin ligase

In humans, there are two E1s, UBA1 and UBA6, ~30 different E2, and over 600 different E3s (279). E3s are divided into three main families: HECT (*h*omologous to *E7*AP *c*arboxyl *t*erminus), RING (*r*eally *i*nteresting *n*ew *g*ene), and RBR (*R*ING-*b*etween-*R*ING). The largest among them, the RING family, can be furthermore divided into different superfamilies. Among them, the cullin-RING E3 ligases (CRLs) are one of the largest with over 200 members and accounting for ~20 % of cellular ubiquitination events (273, 280, 281).

The core of all CRLs consists of four components (Figure 1.22 b) (273): (i) a cullin protein serving as scaffold, (ii) a RING protein-domain that recruits the E2 enzyme, (iii) an adaptor protein bridging the substrate receptor to cullin, and (iv) a substrate receptor recognizing the substrate. In mammals, seven canonical cullin proteins are expressed (Cul1, Cul2, Cul3, Cul4A, Cul4B, Cul5, and Cul7) each forming modular, multi-subunit CRLs named CRL1-7

(Figure 1.22 c). Cullins were named after “to cull” originally meaning the selective slaughtering of animals. All CRLs share an overall architecture, despite different structural subunit compositions (Figure 1.22 b). All cullin proteins within the various CRLs act as the central scaffold. The C-terminal domain of cullin (cullin$^{CTD}$) binds the RING protein, which recruits E2~Ub. The N-terminal domain of cullin (cullin$^{NTD}$) interacts with the adaptor protein (Figure 1.22 b).

The regulation of CRLs is achieved by the continuous CAND1-NEDD8 cycle (281–283). The covalent attachment of NEDD8 enhances the activity of CRL1 by providing an interaction platform. NEDD8 together with Cul1 WHB domain, forms the “activation module” and positions the “catalytic module” (E2~Ub and RBX1 RING) relative to the “substrate-scaffolding module” (IκBα:CRL4$^{\beta\text{-}TRCP}$) to activate ubiquitin transfer (284). Deneddylation of CRL1 is performed by the COP9 signalosome and the subsequent replacement of the adaptor and substrate receptor by CAND1. Afterwards, the CAND1:Cul1 complex is biased to bind DCN1, initiating the neddylation of Cul1 again. By an unknown mechanism this leads to the replacement of CAND1 by a new adaptor and substrate receptor, allowing the ubiquitination of a substrate.

### 1.5.3 Cullin4-RING ubiquitin ligase

CRL4s complexes contain the adaptor protein DDB1 (*d*amaged *D*NA *b*inding protein *1*), consisting of three WD40/β-propeller domains (BPA, BPB, and BPC) as well as a helical C-terminal domain (285). The BPB domain interacts with Cul4-NTD and rotates relative to BPA and BPC with a range of 150° (286, 287) (Figure 1.23 b). This flexibility is believed to allow binding of variously sized or shaped substrates, establishing a “ubiquitination zone” where Ub-transfer is possible (288, 289). A specific CRL4 complex together with its substrate receptor DCAF (*D*DB1 and *C*ul4-*a*ssociated *f*actor) is denoted as CRL4$^{DCAF}$.

Substrate receptors are responsible for the recognition of specific motif within substrate protein. If ubiquitination leads to proteasomal degradation these motives have been termed “degrons”, which can be a linear peptide stretches, specific tertiary and quaternary epitopes, or contain specific PTMs, e.g. a phosphodegron (290).

### 1.5.4 DCAF family

The DCAF family of proteins are substrate receptors comprising over 60 members, as identified by affinity purification or standard immuno purification techniques (285, 286, 291–293). Of these, 52 possess WD40 repeats, and either have an H-box or a WDxR motif, respectively. These two motifs have been structural located at the bottom of the WD40 domain, facilitating the connection of DCAF to DDB1 (285).

The family of DCAF substrate receptors represents a large variety of sequences and functional relations, with several members being biochemically and structurally understood in in recent years: DDB2 recognizes photolesions within DNA (288), Cockayne syndrome A

(CSA) is involved in transcription coupled repair (285, 294), and DCAF15, the target of the anti-cancer agent Indisulam acting as a molecular glue degrader in recruiting RBM39 (295, 296). Due to the diverse involvement of CRL4s in various biological processes, their association with a multitude of diseases is not surprising. For instance, $CRL4^{DCAF8}$ has been associated with the promotion of muscle atrophy (297). Furthermore, DCAF2/Cdt2, a master regulator of genome stability, has been implicated with various types of cancer (298, 299). In the following, DCAF1 will be presented as an example of a CRL4-substrate receptors.

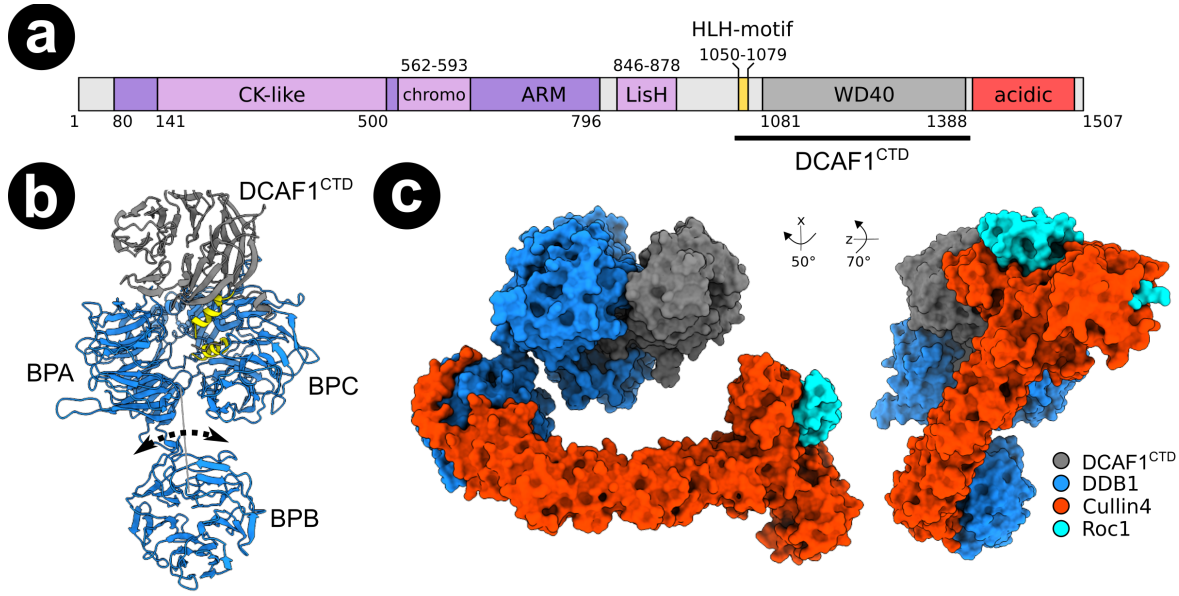



**Figure 1.23: Structural organization of DCAF1.**
***(a)*** *Domain organization of DCAF1. Structural features for the 1507 aa human DCAF1 isoform are shown. The truncated form of* $DCAF1^{CTD}$ *used in the following study is indicated.* ***(b)*** $DDB1{:}DCAF1^{CTD}$ *complex depicted as ribbons (PDB ID: 2HYE)* (286). *The three WD40 domains of DDB1 are labeled (BPA, BPB, and BPC). The axis of rotation between BPB and BPA:BPC is shown. The HLH-motif is colored in yellow.* ***(c)*** *Structural model of a full* $CRL4^{DCAF1(CTD)}$ *is in two views (side and bottom) depicted as surface representation (2HYE)* (286). *The same color code as in Figure 1.22 c is used.*

### 1.5.5 DCAF1, a versatile substrate receptor

DCAF1 contains 1507 aa residues and seven domains (Figure 1.23 a): (i) A large Armadillo (ARM)-like fold (80-796 aa), known to enable protein-protein interactions (300), (ii) a casein kinase (CK)-like domain (141-500 aa) (301), and (iii) a chromo-like domain (562-593 aa) within the ARM-like fold. (iv) A Lis1-homology (LisH; 846-878 aa) has been associated with DCAF1 dimerization (302). (v) A helix-loop-helix motif (HLH-motif or H-box; 1050-1079 aa) that facilitates binding to DDB1 and the (vi) WD40 domain (1081-1388 aa). (vii) An acidic domain (1397-1508 aa) implicated in interactions with additional proteins like p53 or Merlin (303, 304).

So far, the most detailed information does exist on the C-terminus (1046-1396) with several structures revealing its WD40 domain fold and its interactions with DDB1 and various

proteins like simian immunodeficiency virus (SIV) infecting mandrils type 2 (SIVmnd-2) Vpx, SIV infecting sooty mangabey (SIVsm) Vpx as well as human immunodeficiency virus type 1 (HIV-1) Vpr (289, 305, 306).

DCAF1 stands out among DCAFs due to its ability to serve two distinct E3 ubiquitin ligases: CRL4 and the HECT-type E3 ligase EDD/UBR5 (307). DCAF1 is broadly expressed in many tissues and cell types (308) and has been functionally connected to act as a p53-dependent and p53-independet regulator of cell cycle, cell growth, cell division, and survival (308, 309). A gene regulatory effect of DCAF1 through the control of histone modifications has also been reported (301). Despite its widespread role in the cell, DCAF1 was originally discovered as the cellular target of the viral protein r (Vpr), a HIV-1 accessory protein, and hence was named Vpr binding protein (VprBP) (310).

### 1.5.6 Viral hijacking of CRL4s

In recent years, the ability of viruses to counteract host factors during viral infection via the ubiquitin proteasome system (UPS) has emerged (311). Such host-virus interactions disable host defense mechanisms by poly-ubiquitination and degradation of antiviral host proteins. Members of the CRL family are often hijacked by virus proteins, leading to the recruitment of host proteins as *neo*-substrates. Several well studied examples of virus proteins influencing CRLs in different ways are known:

Simian Virus 5 (SV5) possesses the V protein that recruits STAT2/STAT1 to CRL4 to ubiquitinate and degrade STAT1, in order to shut down the host interferon response against viral infection (312, 313). Human Immunodeficiency Virus (HIV) represent a particular interesting example since three, Vif, Vpu and Vpr, out of its 18 proteins interact with five different CRLs (Cul1, 2, 4a/b, and 5) (311). The majority of these interactions lead to poly-ubiquitination of anti-viral host factors. For example, the viral infectivity factor (Vif) binds $CRL5^{ELOBC\text{-}CBF\beta}$ and poly-ubiquitinates the host proteins APOBEC3F and APOBEC3G, which otherwise both negatively affect many steps of the HIV replication cycle (314). The transmembrane protein Vpu hijacks $CRL1^{\beta\text{-}TRCP}$ to mark the CD4 receptor for degradation, thereby preventing superinfection and ensuring efficient release of new viruses (315). The protein Vpr and its relative Vpx will be introduced in the following section.

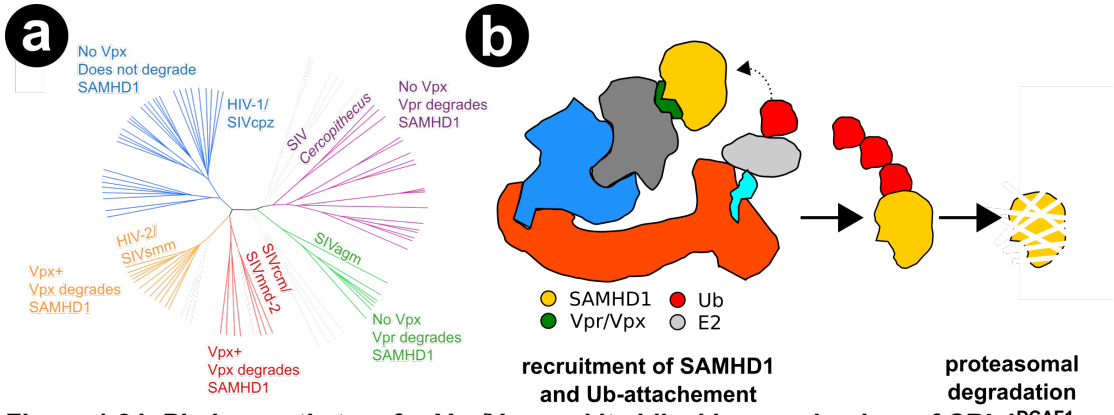



**Figure 1.24: Phylogenetic tree for Vpr/Vpx and its hijacking mechanism of CRL4$^{DCAF1}$.** ***(a)*** *Phylogenetic tree of primate lentiviruses based on an alignment of Vpr/Vpx nucleic acid sequences, with the 5 major lineages of primate lentiviruses labeled. Information regarding activity is based on selected tests in each lineage. Figure was adapted from* (316) ***(b)*** *Mechanism of Vpr/Vpx-hijacking of the CRL4$^{DCAF1}$ to ubiquitinate SAMHD1 leading to proteasomal degradation is shown. Vpr (green) forms a platform to bind DCAF1$^{CTD}$ on the one side and SAMHD1 on the other. Ub-transfer is depicted by a dotted arrow. The same color code as in previous figures is used.*

### 1.5.7 Vpr and Vpx - two accessory proteins

The accessory proteins viral protein r or x (Vpr or Vpx) are encoded by all contemporary HIV or SIV strains (Figure 1.24 a). Both proteins function by modulating the specificity of CRL4$^{DCAF1}$ towards downregulating host proteins. HIV-1 Vpr has been implicated in virus replication in vivo and in macrophage infection models (317). In case of the Vpr, a recent proteomic analysis revealed hundreds of host proteins (316) primarily by a DCAF1- and proteasome-dependet pathway including previously studied targets like UNG (318), HLTF (319), MUS81 (320, 321), MCM10 (322) and TET2 (323). However, exact understanding how this large-scale degradation of host factors promotes virus replication is still missing.

Vpx, on the other hand, is only encoded by two lineages, HIV-2 and SIVmnd (324) (Figure 1.24 a). Vpx recruits a much smaller number of *neo*-substrates, the most important being SAMHD1. SAMHD1 restricts retroviral replication by decreasing the pool of deoxynucleotide triphosphate (dNTP) and thereby inhibiting the function of the viral reverse transcriptase (325–332). In order to promote viral replication, Vpx recruits SAMHD1 to CRL4$^{DCAF1}$ and initiates its proteolytic degradation (333, 334) (Figure 1.24 b), thus leading to promoted virus replication in differentiated myeloid lineage cells, resting T cells, and memory cells (334–336). This relationship between the degradation of SAMHD1 and its antiviral activity induces a constant evolutionary “arms race” between host and virus, with its exact mechanism being highly virus species- and strain-specific: While HIV-2-type Vpx recruits the SAMHD1$^{CT}$, SIVmnd-type Vpx recruits the SAMHD1$^{NT}$ in a different way (289, 305, 324, 337, 338).

The molecular mechanisms of *neo*-substrate recruitment by Vpr and Vpx can be derived from their recently determined structures. SIVmnd Vpx (305), SIVsm Vpx (289) and HIV-1 Vpr (306) all fold into three α helices and make contacts with the same structural elements of

DCAF1$^{CTD}$ by employing typ-conserved residues. However, the interactions between the viral protein and the *neo*-substrate differ substantially: Vpr recruits UNG2 via a cleft between Vpr helices α1 and α2 opposite to the DCAF1-binding side. Both Vpx proteins engage SAMHD1 with their N-terminal tail and an insert loop between helices α2 and α3, with SIVmnd Vpx additionally using helix α2. Therefore, distinct structural regions within Vpr and Vpx bind and recruit the respective *neo*-substrate.

The adaptation to the primate host and the evasion of SAMHD1's antiviral activity places selective pressure on the virus evolution. Therefore, two SIVs (SIVagm and SIVdeb/SIVmus/SIVsyk), that do not possess Vpx, *neo*-functionalized Vpr instead to induce binding and subsequent degradation of SAMHD1 (324, 337, 339). Beside the recruitment of SAMHD1, these "hybrid" Vpr proteins still target some the host factors recruited by HIV-1 Vpr.

# Chapter 2

# 2 Aims of this study

Several applications of cryo-EM as a tool to elucidate the structure of MMMs will be presented in the following. In Chapter 3 and 4 present two studies, in which the λN-TAC was investigated using an integrative structural biology approach. Chapter 5 presents the structural and functional investigation of Ebp1 and its role at the ribosome during neuronal development. Chapter 6 gives insights into the hijacking of a Cullin4-RING ubiquitin ligase by the viral protein Vpr.

**Chapter 3:** In *Said et al. 2017* (340), data from biochemical studies, XLMS, cryo-EM, and MX allowed to elucidate the structure of a small ribonucleoprotein complex, termed the "modifying RNP", consisting of the $\lambda N^{CTD}$, NusA, NusE, NusB, and the *nut* site. This structure alongside extensive biochemical analysis provided a detailed understanding of how $\lambda N^{CTD}$ interacts with the RNA and Nus factors. Additionally, the overall architecture of the λN-TAC was resolved using cryo-EM at medium resolution. This allowed to position the RNP and the RNAP relative to each other, suggesting an invasion of $\lambda N^{NTD}$ into the active site of the RNAP.

**Chapter 4:** In *Krupp, Said, Huang et al. 2019* (341), the full λN-TAC was resolved using cryo-EM at near-atomic resolution, allowing to model the $\lambda N^{NTD}$ *de novo*. Together with corroborating mutational studies and functional assays, a multi-pronged strategy was proposed by which λN influences the interactions of the RNAP with nucleic acids, NusA, NusE, and NusG.

**Chapter 5:** In *Kraushar et al., under review,* the structure of the ribosome:Ebp1 complex was resolved at near-atomic resolution. This structural results allowed the description of the mechanism by which Ebp1 binds the ribosome at the tunnel exit. In combination with additional functional results, Ebp1's role in the control of gene expression during neuronal development was defined.

**Chapter 6:** In *Banchenko, Krupp et al., under review,* the structure of DCAF1:Vpr was resolved by MX and corroborated by biochemical findings. Together with data from XLMS and cryo-EM the binding interface of SAMHD1 was defined. These results allow the formulation of a diverging molecular mechanism for the Vpx/Vpr family and provide insights into the structural flexibility and dynamics of a CRL4-catalysed E3 ubiquitin ligases.


STOP

# Chapter 3

# 3 Structural basis for λN-dependent processive transcription antitermination.

## 3.1 Results of *Said et al. 2017*

λN represents a paradigmatic transcriptional regulator, leading to processive anti-termination. λN assembles on a transcriptional elongation complex (TEC) containing the Nus factors A, G, B, E, and the *nut* RNA, and thereby forms the transcriptional anti-termination complex (λN-TAC).

### 3.1.1 Crystal structure interaction studies of the "modifying RNP"

In this study, the structure of the λN:NusA:NusB:NusE:*nut* RNA complex was resolved using MX. This ribonucleoprotein complex (RNP) will be abbreviated in the following as "modifying RNP". All components were recombinantly expressed and assembled in a stable complex as validated by size exclusion chromatography (SEC). A crystal structure was obtained using full length λN. After truncation of unresolved residues, the structure of the $\lambda N^{1-84}$:$NusA^{\Delta AR2}$:NusB:NusE:*nut*-RNA complex was determined at 3.35 Å resolution using MX. This „modifying RNP" adopts a triskelion shape with the three arms being the NusB:NusE heterodimer, the $NusA^{NTD}$, and the $NusA^{AR1}$ (Figure 3.1 a). The λN can be modeled in an elongated conformation making contact with the NusB:NusE heterodimer, and the $NusA^{KH1}$ and $NusA^{KH2}$.

The structural arrangement of λN are depicted in Figure 3.1 a: Residues $\lambda N^{1-20}$ (region I) form a kinked helix binding the major groove of *boxB* on one side and $NusA^{KH2}$ on the other. $\lambda N^{21-46}$ contact the surface of $NusA^{KH1}$ and $NusA^{KH2}$ on the opposite side of the *nut*-RNA binding site (region II). The residues $\lambda N^{47-52}$ (region IV) contact $NusA^{NTD}$ and $NusA^{S1}$ as well as NusE. $\lambda N^{57-79}$ form an amphipathic helix (region V) arranged in a three-helix bundle together with the N-terminal helix of NusA and the NTD-S1 connector helix (Figure 3.1 c). Independent support for the structural arrangements was obtained by crosslinking/mass spectrometry (XLMS) experiments of the "modifying RNP" (black lines in Figure 3.1 a).

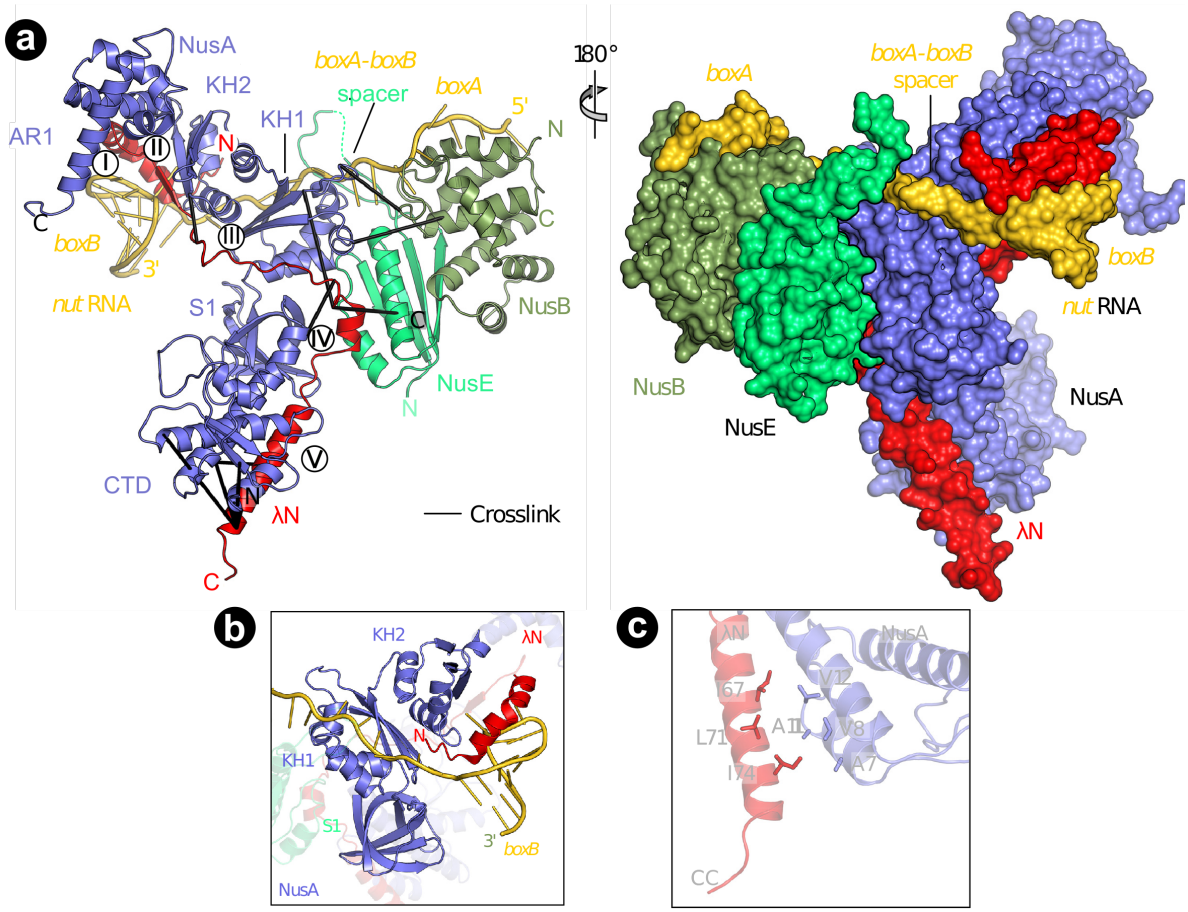



**Figure 3.1: Structure of the "modifying RNP".**
***(a)*** *Front and back view of the "modifying RNP", i.e. λN$^{1-84}$:NusA$^{\Delta AR2}$:NusB:NusE:*nut-*RNA RNP in ribbon and surface representations, respectively. N - N termini; C - C termini; black lines in ribbon view represent intermolecular crosslinks observed for the "modifying RNP" in solution; roman numerals I-V - indicate λN:NusA interaction regions.* ***(b)*** *Close-up view of the interaction sites I and II depicting residues λN$^{1-20}$, the NusA S1, KH1 and KH2, and the boxB of the "modifying RNP".* ***(c)*** *Close-up view of the interaction site V, in which residues λN$^{58-72}$ are packed in a three-helix bundle with two helices from the NusA NTD. Involved residues are shown as sticks.*

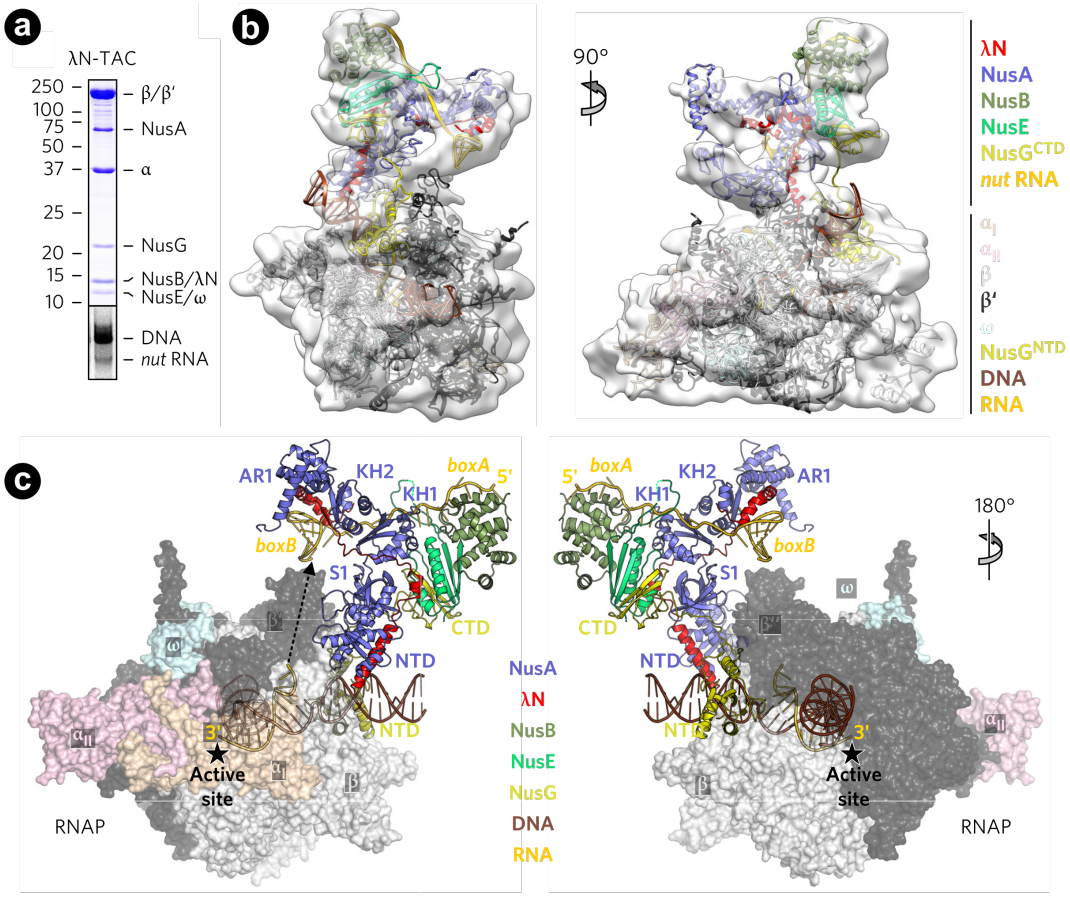


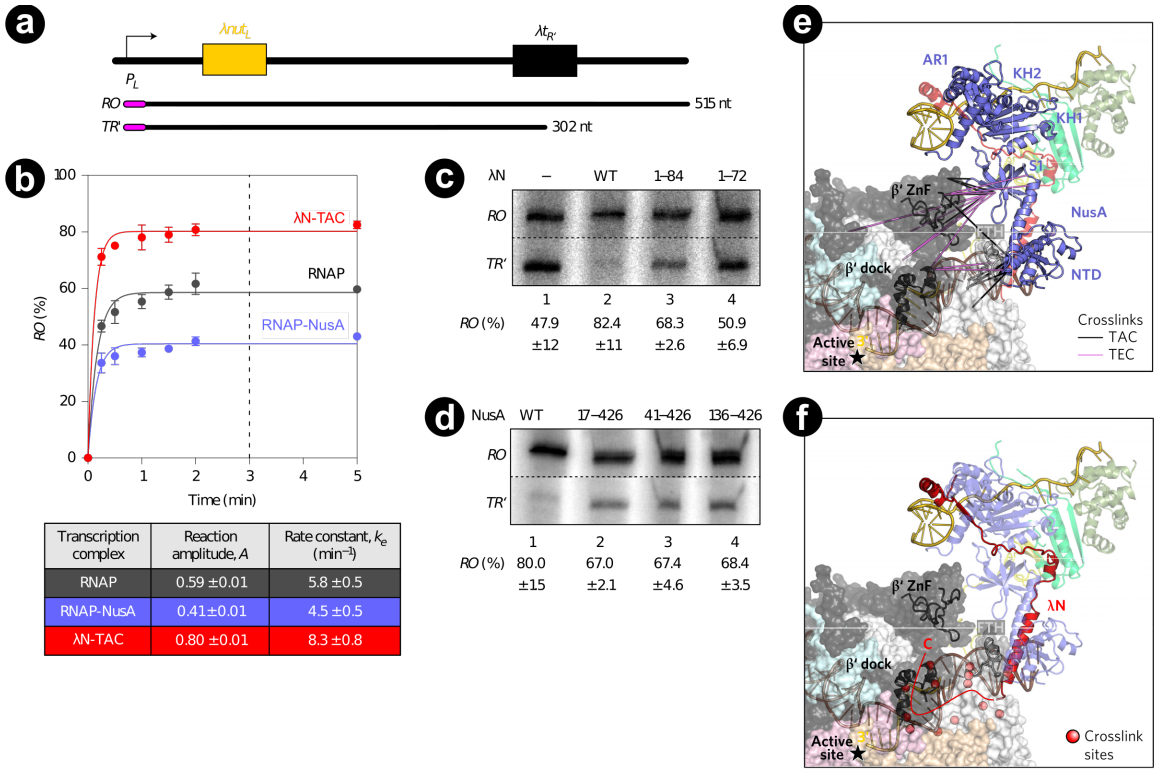

**Figure 3.2: Assembly and structure of the λN-TAC**
***(a)*** *Analysis of the assembled λN-TAC using Coomassie-stained SDS-PAGE (top) and ethidium bromide-stained urea PAGE (bottom). Molecular weights (kDa) are shown on the left.* ***(b)*** *Fitting of the reconstructed cryo-EM density map with the λN-TAC components (ribbon representation) in front and side view. Components belonging to the modifying RNP (top part) and the RNAP (bottom part) are labeled on the right side. Due to their low resolution, the flexible domains NusA$^{AR2}$ and RNAP$^{\alpha CTD}$ where not modeled.* ***(c)*** *Pseudo-atomic model of the λN-TAC in a tilted back and front view, respectively. Asterisk - active site; Dashed arrow - connection between RNA emerging from the exit channel and* boxB*.*

### 3.1.2 Assembly and structure of the λN-TAC

In order to elucidate the complete structure of the λN-TAC, the complex was assembled using the RNAP, an artificial transcription bubble consisting of dsDNA and *nut*-RNA, all Nus factors, and λN (Figure 3.2 a). Several samples were screened using negative stain and cryo-EM conditions. *Ab initio* volumes were reconstructed and validated using computational approaches. In the end, a 9.8 Å resolution map was obtained using cryo-EM and SPA, allowing to localize all components of the λN-TAC (Figure 3.2 b). In the overall structure the „modifying RNP“ binds at the β flap domain close to the RNA exit channel, the β'-zinc finger domain, β'-dock domain, and the CTD of β subunit (Figure 3.2 b and c). NusG$^{NTD}$ can be located at the upstream face of the transcription bubble closing down the RNAP cleft. NusG$^{CTD}$ can be detected as additional density next to NusE. Both locations are in agreement with previous studies (162, 171, 342).

### 3.1.3 Structure-guided functional analysis of the λN-TAC

In order to provide insights into the molecular mechanism of λN's action, structure-guided functional analysis was carried out using *in vitro* transcription assays. The *in vitro* transcription assay measures anti-termination activity as the ratio of runoff product (RO) and the terminated product (TR′) (Figure 3.3 a). Successful termination produces predominantly short TR′ products with a length of 302 nt. In the event of anti-termination the transcript RO (515 nt) dominates the sample and leads to an anti-termination efficiency (RO ratio) of >50% (Figure 3.3 c). Measurements over a time course of 5 min allowed the fitting of the data points using a first-order reaction (Figure 3.3 b).




| Transcription complex | Reaction amplitude, *A* | Rate constant, $k_e$ ($min^{-1}$) |
|---|---|---|
| RNAP | 0.59 ±0.01 | 5.8 ±0.5 |
| RNAP-NusA | 0.41±0.01 | 4.5 ±0.5 |
| λN-TAC | 0.80 ±0.01 | 8.3 ±0.8 |

**Figure 3.3: Analysis of the λN-TAC via functional assays and XLMS.**
***(a)*** *Scheme of the DNA template used for the* in vitro *anti-termination assays and the observed products RO and TR′. RO - runoff product; TR′ - product terminated at $\lambda t_{R'}$; magenta regions - initially transcribed region.* ***(b)*** *Top: Time course performed by the complexes λN-TAC, RNAP, and RNAP:NusA depicting no significant changes after 3 mins. Bottom: Table of the amplitude A and rate constant $k_e$ after fitting the data to a first-order reaction.* ***(c)*** *and* ***(d)*** *Transcription assays monitoring anti-termination efficiency at 3 min time points with different constructs for (c) λN and (d) NusA.* ***(e)*** *Crosslinks between NusA and RNAP subunits obtained from λN-TAC (black lines) or TEC (violet lines), respectively. Elements of the RNA exit channel are shown as ribbons.* ***(f)*** *Crosslinks (red line) between λN C-terminal residues absent in the crystal structure and RNAP, indicating the path of the remaining C terminus into the RNA exit channel. Red spheres - crosslink s on RNAP.*

The calculated anti-termination efficiency identifies the λN-TAC as the complex with the highest anti-termination activity compared to the TEC alone (82 to 48 %, lanes 1 and 2 in Figure 3.3 c). A C-terminal truncated variant of λN, $\lambda N^{1-84}$, shows a reduced anti-termination efficiency of 68% (lane 3), whereas the even further truncated protein $\lambda N^{1-72}$ did not induce anti-termination at all (51%, lane 4). These data highlight the importance of the residues $\lambda N^{72-84}$ for λN-mediated antitermination.

Consistently, several truncations of the NusA N-terminus also reduce anti-termination activity from 80 % (wild type in lane 1) up to ~67 % for various mutants (Figure 3.3 d). These results highlight the importance of the $NusA^{NTD}$ for efficient anti-termination by λN.

XLMS performed on the λN-TAC supports the position of the modifying RNP relative to RNAP (Figure 3.3 e). The unresolved $\lambda N^{CTD}$ likely reaches inside the active site cavity as verified by various crosslinks in that region (Figure 3.3 f).

## 3.2 Discussion of *Said et al. 2017*

In this work, the structures of the λN$^{1–84}$:NusA$^{\Delta AR2}$:NusB:NusE:*nut* RNP ("modifying RNP") as well as the overall architecture of the λN-TAC were elucidated employing an integrative structural biology approach. MX, cryo-EM, and XLMS were employed to obtain spatial information of the components within the λN-TAC. Additionally, structure-guided mutations combined with interaction studies and *in vitro* transcription assays supported a suggested mechanism of λN-mediated processive anti-termination. The protein λN adopts an extended conformation and thereby contacts several widespread areas inside the λN-TAC.

### 3.2.1 λN induces anti-termination by interacting with Nus factors and *nut* RNA

The N-terminus of λN fulfills several tasks. The λN$^{1-20}$ are sandwiched between the NusA KH2 and *boxB* RNA, thereby assisting the assembly of the λN-TAC (178). The following residues of λN occupy the RNA-binding platform of NusA$^{KH2}$. NusA$^{KH1}$, on the other hand, is bound to *nut* RNA. Thereby, a continuous RNA binding interface along the KH1-KH2 surface is sequestered by *nut* and λN, preventing interactions with nascent RNA transcripts. λN$^{47-52}$ contact NusA$^{NTD}$ and NusA$^{S1}$ as well as NusE. The amphipathic helix λN$^{57-79}$ participates in a three-helix bundle restricting the βFTH in its function.

The unresolved C-terminus λN$^{82-107}$ likely binds at the RNA exit channel, probably stabilizing a processive state. Additionally, the C-terminus may counteract DNA:RNA hybrid melting due to terminator hairpin stabilization or may neutralize the action of termination factor ρ. The results shed light on how the intrinsically disordered protein λN cooperates with the Nus factors and the *nut* RNA to achieve anti-termination. The relatively small protein λN can interact with several, compactly arrayed epitopes within the TEC to induce anti-termination.

### 3.2.2 Reprogramming of NusA into an antitermination factor

NusA is the largest and most versatile of the Nus factors, being involved in opposing processes like elongation, pausing, termination, and anti-termination (178, 180). Previous structural and biochemical studies on NusA (111, 174, 179, 182) placed it next to the RNA exit channel with the its, S1, and KH1 hovering above the exit to allow contact with an emerging RNA hairpin (343). While the obtained cryo-EM structure and comparative XLMS results suggest a repositioning of NusA by λN, a high-resolution study together with structure-guided analysis is needed to test this hypothesis.

### 3.2.3 Possible mechanisms for inhibition of ρ-dependent termination

These results also suggest insights into how λN, the Nus factors and *nut*-RNA oppose the termination factor ρ in different ways. The bulky "modifying RNP" may sterically hinder the binding of ρ to the RNAP, since it occupies important landmarks, namely the NusG$^{CTD}$ and the RNA exit. While factor ρ translocates along the RNA until it reaches the RNAP, the "modifying

RNP" could present a stable roadblock, hindering ρ in approaching the RNAP. Furthermore, the nascent RNA downstream of the *nut* site is expected to loop out to the back (black dashed arrow in Figure 3.2 c). A *rut* site within that loop, would point away from the front side, where ρ is likely bound to $NusG^{CTD}$ (344). In order to better understand the mechanism of ρ-dependent termination, high-resolution structures of the TEC and the λN-TAC in the presence of ρ are required in the future.

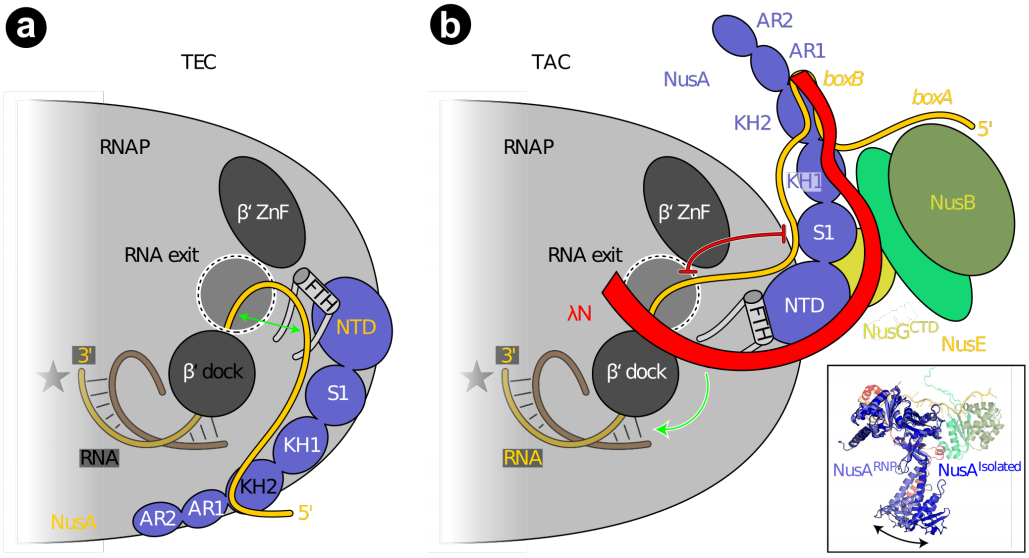



**Figure 3.4: Mechanism of λN-mediated processive anti-termination.**
***a)*** *Model for the activity of NusA in the TEC. Nascent RNA is threaded between the RNAP β' dock domain and NusA. The* $NusA^{NTD}$ *can facilitate the formation of a terminator hairpin by stabilizing the upstream branch and delivering it to the emerging downstream branch (green double arrow). Thereby NusA supports intrinsic termination.* ***b)*** *Model of the λN-TAC. λN repositions NusA, the associated factors and the βFTH, thus opening a gate for the nascent RNA to bind* $NusA^{S1}$*. The latter is able to sequester the upstream branch of a terminator hairpin during the emergence of the downstream branch. Moreover, the λN C-terminus contacts RNAP elements at the RNA exit channel, thus indirectly stabilizing the hybrid (green arrow) and preventing hybrid melting. Inset: Overlay based on the S1-KH1-KH2 region of NusA within the "modifying RNP" and isolated* $NusA^{100\text{-}426}$ *with a modeled* $NusA^{NTD}$*. A double arrow indicates the flexibility of the NTD relative to the overlaid regions of NusA.*

# Chapter 4

# 4 Structural basis for the action of an all-purpose transcription anti-termination factor

## 4.1 Results of *Krupp, Said, Huang et al. 2019*

A recombinant λN-TAC was assembled from purified parts as described previously. Small optimizations of the previous assembly protocol (340) and the collection of a larger data set on a more modern cryo-EM yielded a 3.7 Å resolution reconstruction of the full λN-TAC (Figure 4.1 a). The local resolution ranges from 3.4 Å in the RNAP up to 12 Å at the "modifying RNP". All components were unambiguously docked into the cryo-EM density using our previous model (340) and the $NusG^{CTD}$:NusE heterodimer structure resolved by NMR (171). All elements of the "modifying RNP" directly contacting the RNAP were clearly defined and were modeled *de novo* (Figure 4.1 a).

### 4.1.1 λN reinforces the elongation-competent conformation of RNAP

The overall conformation of the RNAP is similar to the one of an unmodified TEC (102). The λN-TAC comprises an elongation-competent 9 bp hybrid in the post-translocated state with an unpaired +1 position in the DNA template strand (Figure 4.1 c). This is surprising since the nucleic acid scaffold could form a pause-inducing 10 bp hybrid (Figure 4.1 b). In order to confirm that, *in vitro* transcription assays were performed, showing a higher yield and an increased rate of λN-TAC compared to TEC alone (Figure 4.3 b lane 3 and 4). These results suggest a stabilization of the RNAP by λN in an elongation-competent conformation.

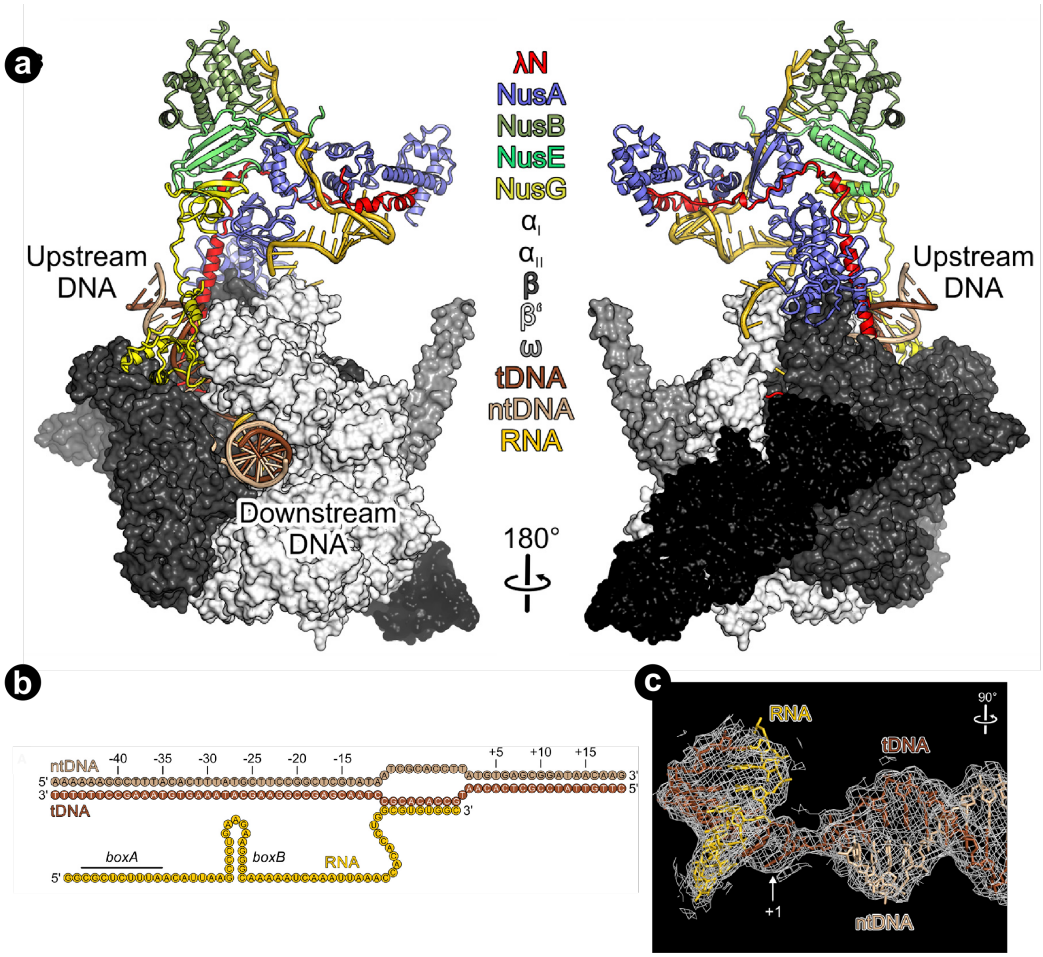



**Figure 4.1: Structural overview of the assembled λN-TAC.**
***(a)*** *λN-TAC with the "modifying RNP" and the nucleic acids depicted in ribbon and the RNAP in surface representation, respectively. The color code is used in all following figures. tDNA - template DNA strand; ntDNA - non-template DNA strand.* ***(b)*** *Nucleic acid scaffold employed in the structural analysis comprising a transcription bubble, a DNA:RNA hybrid and sequences for* boxA *and* boxB. ***(c)*** *Cryo-EM density at the downstream DNA and the DNA:RNA hybrid. The λN-TAC adopts a post-translocated state and the unpaired +1 position is indicated. Rotation symbols here and in all following figures are relative to the view in (a) left side.*

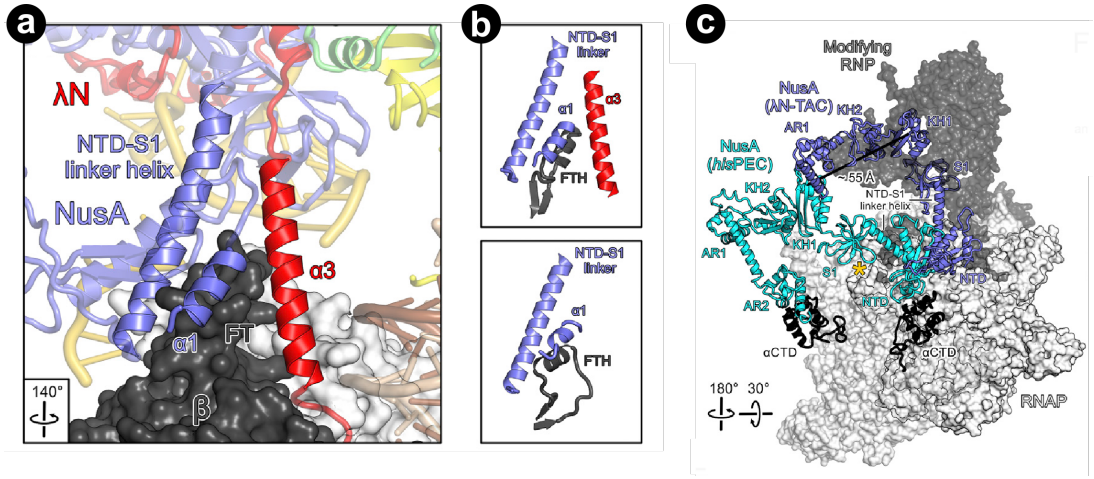



**Figure 4.2: Remodeling of the βFT and repositioning of NusA.**
***a)*** *Interaction of the linker helix and α1 helix of NusA (blue) and the central a3 helix of λN (red) with the RNAP βFT (dark gray surface).* ***b)*** *Comparison of the NusA-λN-FT interaction in the λN-TAC (top) and the NusA-FT interaction in a NusA-hisPEC (PDB ID: 6FLQ) after superposition according to the NusA NTD-S1 linker helix.* ***c)*** *Global repositioning of NusA (blue) by λN and other portions of the "modifying RNP" (dark gray surface), illustrated by the position of NusA in the NusA-hisPEC structure (PDB: 6FLQ; cyan). For the overlay, the RNAPs were aligned (light grey surface). The $\alpha^{CTDs}$ (black) as seen in the NusA-hisPEC are also shown. Black line - distance between equivalent points in the NusA KH1 domains. Golden asterisk - position of the tip of the pause hairpin in the NusA-hisPEC.*

### 4.1.2 The "modifying RNP" is flexibly anchored to RNAP

3D classification of the cryo-EM data focused on the "modifying RNP" led to three sub-states at nominal resolution between 4.2 and 4.8 Å. Comparison of these three states suggested a bi-model movement of the "modifying RNP" as a rigid body. By a "shake" of ~20° and a "nod" of ~10°, the "modifying RNP" moves the *nut boxB* back and forth from the RNA exit tunnel. Throughout all states, all four components of the "modifying RNP" maintain their interactions with the RNAP:

(i) $\lambda N^{58-84}$ forms a long helix α3 contacting the upstream DNA duplex and the RNAP βFT (red in Figure 4.2 and Figure 4.3 c).

(ii) $NusA^{NTD}$ binds on the opposite side of the βFT (purple in Figure 4.2 a and b).

(iii) Nascent RNA reaches from the RNA exit tunnel to the *nut boxB* of the "modifying" RNP (golden in Figure 4.5 a).

(iv) $NusG^{CTD}$ (yellow) interactions with the NusE (light green) and NusA (purple) (Figure 4.6 a).

Taken together, these interactions may be crucial to stabilize the λN-TAC in its active conformation and counteract different modes of pausing and termination. In the following, these interactions will be discussed in more detail.

### 4.1.3 λN remodels the RNAP β flap tip

As mentioned in the previous section, three elements form a tripod, engaging with the βFT (residues 887-915) like a push button (Figure 4.2 a and b): the α-helix $\lambda N^{58-84}$ (red), the $NusA^{NTD}$ helices α1 (2-10), and its linker helix (122-132). Thus, the βFT within the λN-TAC is stabilized by hydrophobic interactions, in contrast to the unmodified TEC, where the βFT is disordered (102). Compared to the hairpin-paused elongation complexes with and without NusA, $\lambda N^{58-84}$ also remodels the interaction of the βFT to the Zinc binding domain (ZBD) and NusA NTD (174, 345). Thereby, λN induces remodeling of the βFT-ZBD interaction, prevents RNA hairpin accommodation, hampers the $NusA^{NTD:S1}$ domains in approaching the hairpin tip, and redirects residues of the ZBD from sensing signals for pausing and intrinsic termination. Taken together, λN interacts with the βFT and thereby influences several mechanisms at once.

### 4.1.4 λN-mediated remodeling alters $\alpha^{CTD}$-NusA interactions

The influence of λN on the elements surrounding NusA (Figure 4.2 a) was investigated by testing the following systematic truncations of λN in an *in vitro* anti-termination assay (Figure 4.3 a and b):

- $\lambda N^{38-107}$ reduces the anti-termination efficiency to ~75% (lane 5) compared to full length λN, by losing interactions with *nut boxB* and both $NusA^{KH}$ domains.
- $\lambda N^{58-107}$ completely loses anti-termination efficiency (lane 6) upon additionally removal of the interactions to NusE.
- $\lambda N^{1-84}$ diminishes the anti-termination activity to 60% (lane 8), missing the 22 residues reaching into the RNAP.
- $\lambda N^{1-62}$ also abolishes anti-termination completely by removing the α3-helix (lane 9).

Compared to the NusA-*his* paused state (NusA-*his*PEC), λN rotates NusA by 55 Å, breaking its interactions to the $\alpha^{CTD}$, which normally support intrinsic termination (174) (Figure 4.2 c). Taken together, these results demonstrate the essential role of λN helix α3 and its neighboring regions in anchoring it to and modifying the components of the λN-TAC.

### 4.1.5 λN stabilizes upDNA, the hybrid and the active state of the RNAP

Due to its location next to the upstream DNA (upDNA), λN helix α3 reinforces base pairing upstream of the transcription bubble, and thereby counteracts RNAP backtracking (Figure 4.3 c). $\lambda N^{85-107}$ enters the RNAP hybrid cavity and does not introduce major distortions within the RNAP or the hybrid. Only a small upward movement of the β flap and β' clamp elements contacting the hybrid can be detected. On its way, λN stitches some of the mobile elements together (Figure 4.3 a), thereby reducing pause-associated conformational changes. λN potentially counteracts hybrid melting by decreasing the inner diameter of the hybrid cavity and reducing the repulsion of charges between the RNA and DNA. The C-terminal end of λN resides between the clamp and the β'-dock, lining the exit tunnel opposite the ZBD, and

thereby counteracting a dilation of the exit tunnel and obstructing the position of the 5'-branch of a hairpin (345) (orange dashed line in Figure 4.3 d).

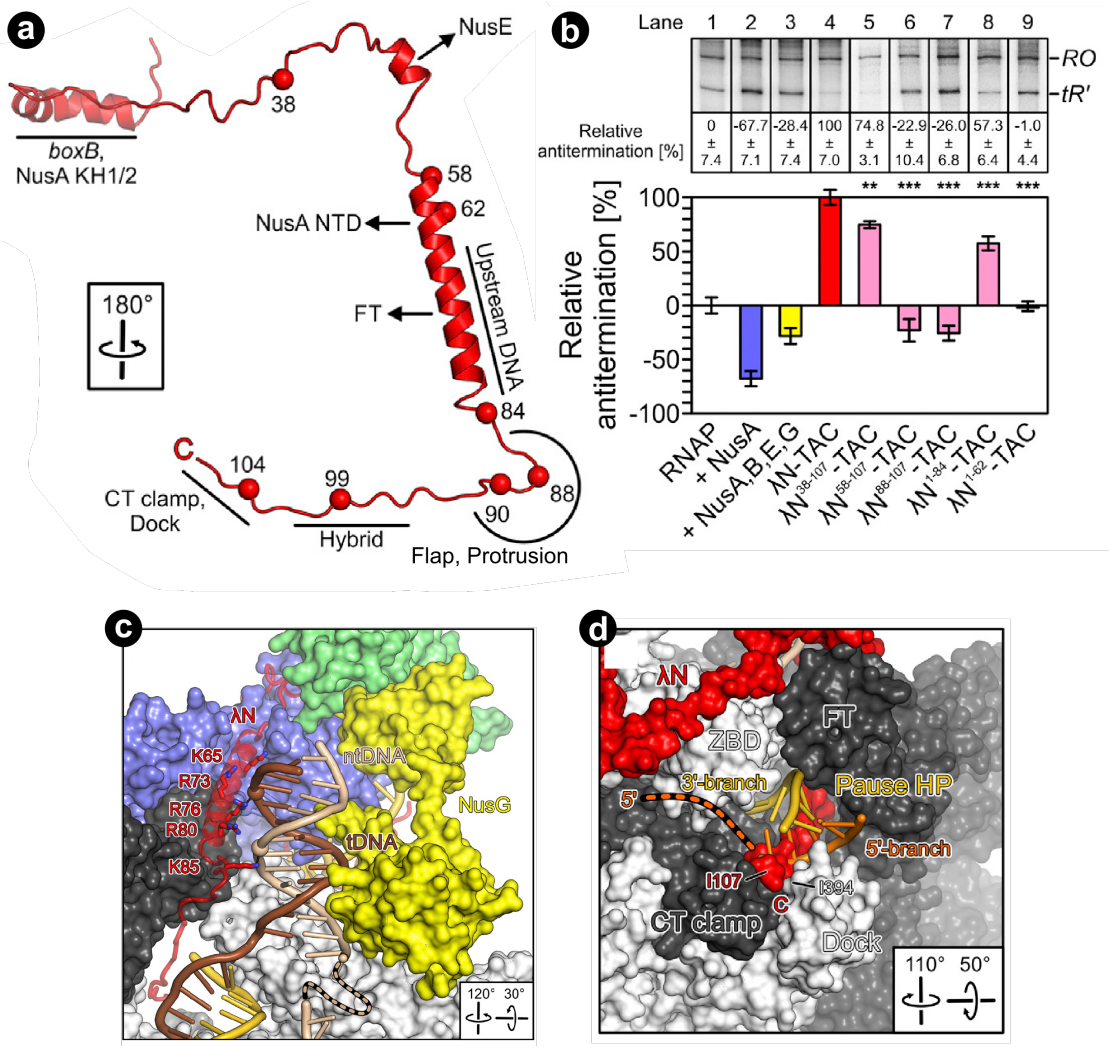



**Figure 4.3: Interaction of λN with various factors and the RNAP.**
***(a)*** *Ribbon representation of the entire λN within the λN-TAC highlighting truncation positions and interaction elements as red spheres.* ***(b)*** *Transcription assay monitoring anti-termination efficiency at 3 min time points. Complexes and truncated forms of λN are indicated at the bottom. Relative anti-termination efficiency defining RNAP alone as 0% and λN-TAC as 100% and scaling quantified data of other complexes accordingly.* ***(c)*** *λN α3-helix running along the upstream DNA with positively charged side chains extending toward the DNA as sticks.* ***(d)*** *λN C terminus within the RNA exit tunnel connecting the CT clamp and the dock domain. The 5'-branch of a pause hairpin (orange), modeled according to NusA-hisPEC (PDB ID: 6FLQ), clashes with the C-terminus of λN. Orange dashed line indicates the presumed path of the 5' hairpin.*

### 4.1.6 λN enhances anti-pausing and anti-termination activities of NusG

NusG$^{NTD}$ binds at the cleft, bridging the subunits β and β‘ and thereby promoting transcriptional processivity (157, 162). A long loop of NusG lines the major groove of the upstream DNA (158) (Figure 4.3 c), rationalizing how NusG counteracts RNAP backtracking (346) and supports re-annealing (347). λN runs along the opposite side of the upstream DNA and could thereby hold the DNA close to NusG.

### 4.1.7 λN might communicate with the RNAP active site

Previous results (111) suggested an allosteric effect of the pause hairpin on the RNAP active site via the connector, a two-stranded β-sheet linking the β flap with the active site (pink in Figure 4.4 a). λN might influence such an interaction by intervening between the connector and the RNA in the exit channel as shown by point-mutants of λN (Figure 4.4 b).

### 4.1.8 Guidance of nascent RNA might oppose hairpin formation

Density for the RNA at a lower contour level suggests that nascent RNA contacts a positively charged surface made up by the βFT, ZBD, and the NusA$^{S1}$ domain (Figure 4.5 a). This guidance of exiting RNA could counteract or delay the formation of a hairpin by occupying its upper branch. This hypothesis is supported by several NusA$^{S1}$ mutants within the proposed RNA path, reducing the anti-termination efficiency (Figure 4.5 b).

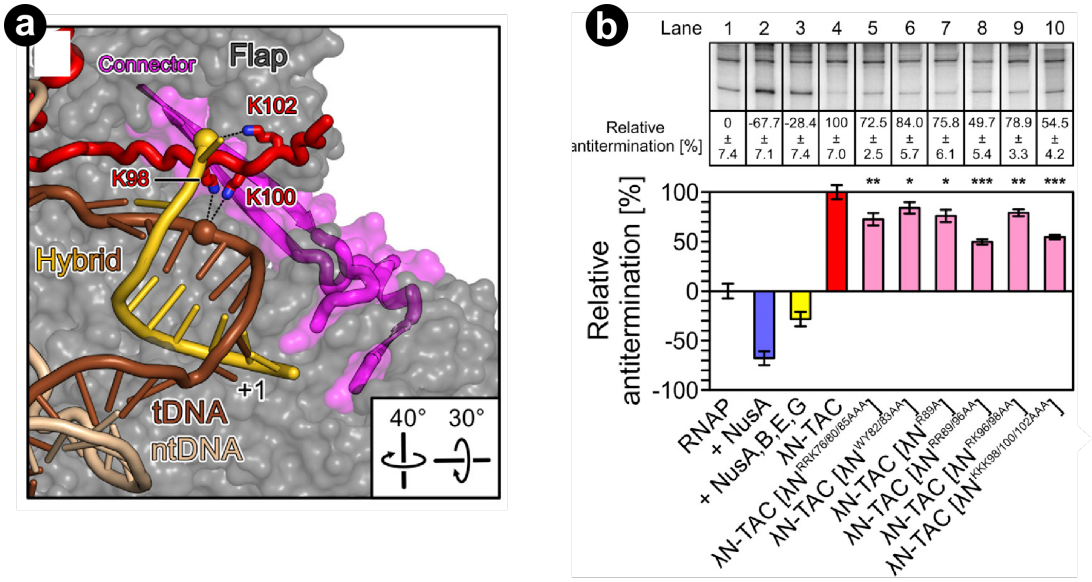



**Figure 4.4: λN contacts several internal RNAP elements.**
***(a)*** *λN contacting the connector (pink), the exiting RNA (golden), and the DNA:RNA hybrid (brown and golden). Positively charged side chains are shown as sticks.* ***(b)*** *Transcription assay monitoring anti-termination efficiency at 3 min time points by complexes indicated in the presence of mutated λN.*

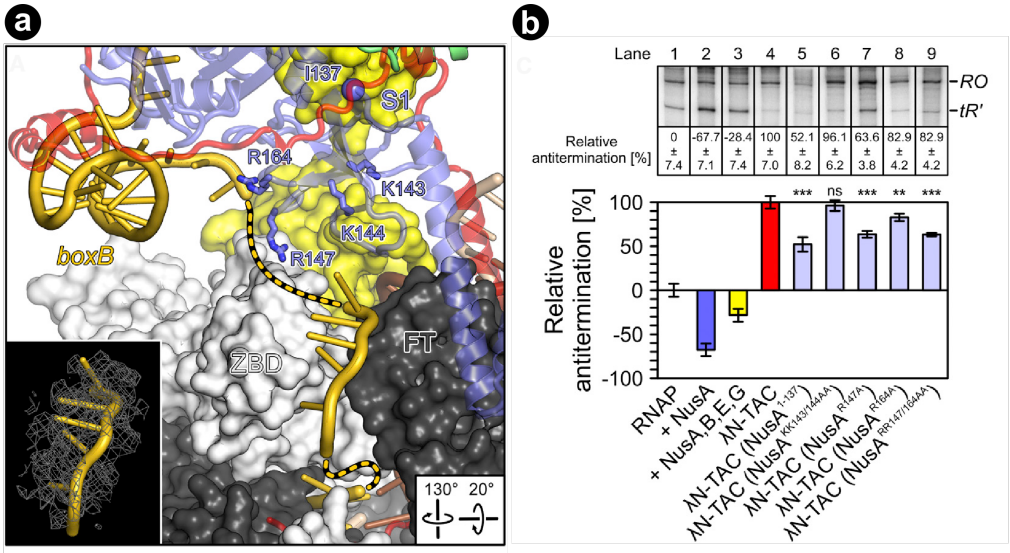



**Figure 4.5: Chaperoning RNA in an extended conformation.**
***(a)*** *Path of the emerging transcript from the RNA exit tunnel to the boxB element across the ZBD, the βFT, and along the NusA$^{S1}$. Golden dashed lines represent regions not defined in the cryo-EM map. Inset: cryo-EM map at the 4σ level around the RNA portion between ZBD and βFT. Side chains shown as sticks around the concave surface of the NusA$^{S1}$ domain, which were mutated for functional tests.* ***(b)*** In vitro *transcription assays monitoring anti-termination activity at 3 min time point by different complexes. Significance relative to λN-TAC are depicted by asterisks.*

### 4.1.9 A NusA-NusG$^{CTD}$-NusE interaction inhibits ρ-dependent termination

NusG$^{CTD}$ makes interactions with NusE, as described previously (171), and with NusA$^{S1}$ (Figure 4.6 a). Additionally, NusG$^{CTD}$ also binds termination factor ρ. Interactions of NusG$^{CTD}$ to either ρ or NusE are mutually exclusive (348). A ρ-dependent termination assay was performed to test whether the interaction between NusA$^{S1}$, NusG$^{CTD}$, and NusE counteract ρ-dependent termination (Figure 4.6 b). As expected λN promotes anti-ρ activity. When modeling a NusA:NusE:NusB:NusG complex using the NusA-*his*PEC structure (174), NusA and NusE are too far away to reach the NusG$^{CTD}$, which is restricted by the NusG$^{NTD}$ binding position (Figure 4.6 c). Therefore, λN repositions NusA$^{S1}$ and NusE and thereby generates an efficient mechanism to sequester NusG$^{CTD}$. Additionally, the “modifying RNP” could sterically hinder the binding of ρ to the RNAP, although its exact bindings site is currently unknown.

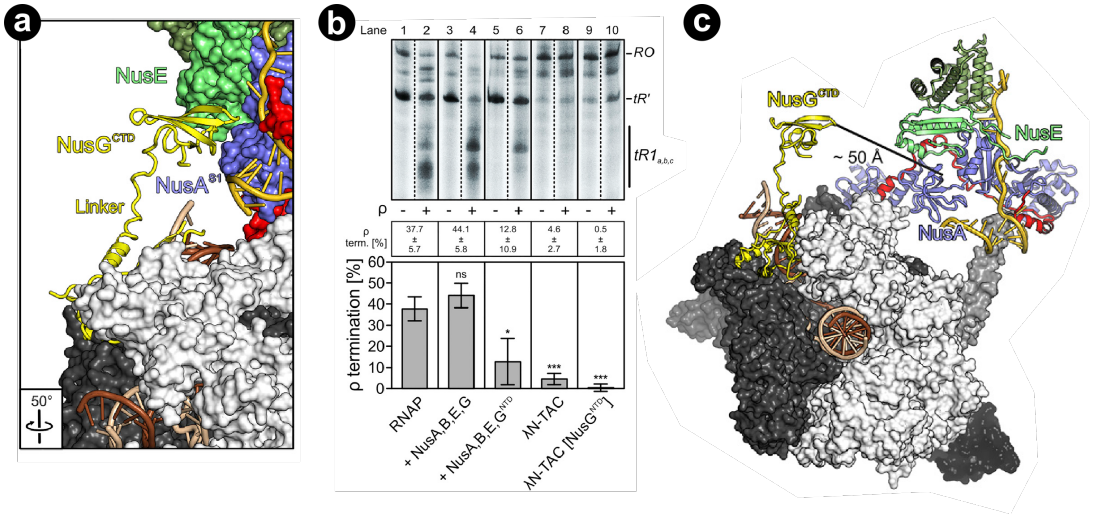



**Figure 4.6: NusG$^{CTD}$ Interactions and Impact on ρ-dependent termination.**
***a)*** *Binding of the NusG$^{CTD}$ at the NusA$^{S1}$:NusE interface.* ***b)*** In vitro *transcription assay monitoring termination of the template λtR1$_{a,b,c}$ at a 3 min time point for different compositions. The DNA template encodes a* rut *site, a ρ-dependent λtR1 and an intrinsic λtR' terminator. Significance is relative to RNAP alone.* ***c)*** *The λN:NusA:NusE:NusE:nut-*RNA *complex was modeled by superposition of the NusA within the "modifying RNP" and the NusA-hisPEC (PDB ID: 6FLQ).*

## 4.2 Discussion of *Krupp, Said, Huang et al. 2019*

In this study, a detailed structural and functional analysis of the λN-TAC was performed. Structural information was obtained by performing cryo-EM SPA with the entire λN-TAC. Additionally, the use of truncated or mutated proteins in *in vitro* transcription assays provided functional insights into the mechanism of λN anti-termination.

### 4.2.1 The λN-TAC structure rationalizes numerous previous observations

The presented results can be used to rationalize results from over 50 years of research on the mechanism of anti-termination by λN. A few examples are listed below:

- λN induces alterations of the interactions between the nascent RNA with the βFT, ZBD, and NusA (178, 349).
- Flexibilities within the βFT (111, 120) and the ZBD (350) are important for intrinsic termination.
- λN destabilizes hairpins during pausing or termination (351).
- λN prevents backtracking (349).
- λN enters the RNAP through the hybrid cavity explaining previous crosslinks (340) and hydroxyl-radical footprints (349).
- An increase in elongation rate in the presence of λN corroborates the observation of an increase in catalytical competence of the RNAP (349).

Taken together, the near atomic cryo-EM structure together with its structure-guided analysis can be used to explain several previous findings, obtained over five decades of research.

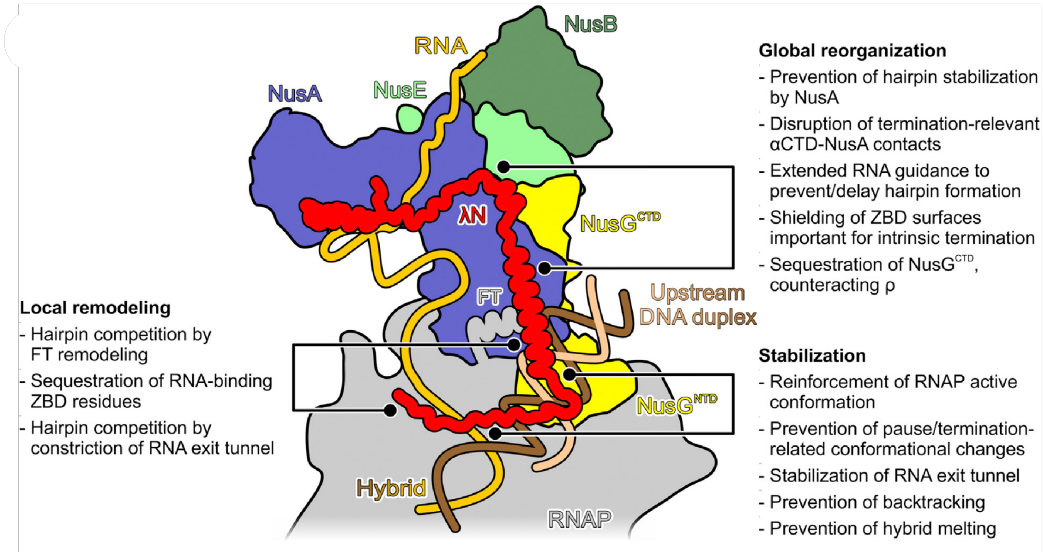



**Figure 4.7: Model illustrating λN action.**
*Summary of suggested mechanisms employed by λN to suppress transcriptional pausing and termination.*

### 4.2.2 Intrinsic disorder allows λN to implement a multi-pronged anti-pausing and anti-termination strategy

From the presented results three main mechanisms can be concluded on how λN ensures resistance to pausing and termination of the λN-TAC (Figure 4.7):

(i) λN globally repositions NusA, thereby

- displacing regions that would otherwise stabilize RNA hairpins in the exit channel,
- breaking NusA-$\alpha^{CTD}$ interactions, that support hairpin-stimulated pausing and intrinsic termination,
- redirecting nascent RNA and thereby counteracting or delaying hairpin formation,
- probably occluding the binding site of ρ, and
- arranging $NusA^{S1}$, thereby breaking interactions with the ZBD which are important for intrinsic termination.

(ii) λN locally remodels RNAP elements such as the

- βFT to hamper binding of the RNA hairpin in the RNA exit tunnel,
- βFT to sequester RNA-binding residues of the ZBD involved in hairpin accommodation, and
- RNA exit tunnel by constricting its diameter.

(iii) λN promotes processive elongation by

- stabilizing the upstream DNA together with $NusG^{NTD}$, likely favoring DNA re-annealing, thereby preventing RNAP backtracking and restricting RNAP conformations, and
- traversing into the hybrid cavity and thereby averting pause-related conformational changes, stabilizing the hybrid, functionally insulating the exit tunnel from the active site, and probably enhancing the catalytic activity of RNAP.

Despite its small size of 107 aa, λN can comprehensively hijack RNAP and its elongation factors. Due to its intrinsic disorder, λN adopts an extended conformation bridging a large distance and interacts with the nascent RNA, upstream DNA, the hybrid, proteins of the regulatory RNP (NusA, NusE), and various elements within the RNAP.

Taken together, λN can be termed a “Swiss army knife” evolved to counteract several modes of pausing and termination. Therefore, it will be interesting to see if other proteins inducing anti-termination operate via an analogous mechanism.

### 4.2.3 Strategies to maintain the elongating conformation

Pausing is a universally conserved state of all multi-subunit RNAPs, exhibiting a tilted 10 bp hybrid, with RNA in the post- and DNA in the pre-translocated state, respectively. The tilted 10 bp hybrid is accommodated by small conformational changes in the β-flap, β‘-lid, clamp and dock elements (Figure 4.8 a). λN may favor an elongation competent 9 bp hybrid and counteract transcriptional slippage by

- contacting all pause related RNAP elements and restricting their movements,
- capturing RNA upstream of the hybrid, and
- binding the dock domain and thereby indirectly promoting hybrid movement during translocation.

DNA rewinding and stabilization of an active conformation is achieved by similar strategies in the prokaryotes (NusG) and eukaryotes (NusG-homolog Spt5) (157).

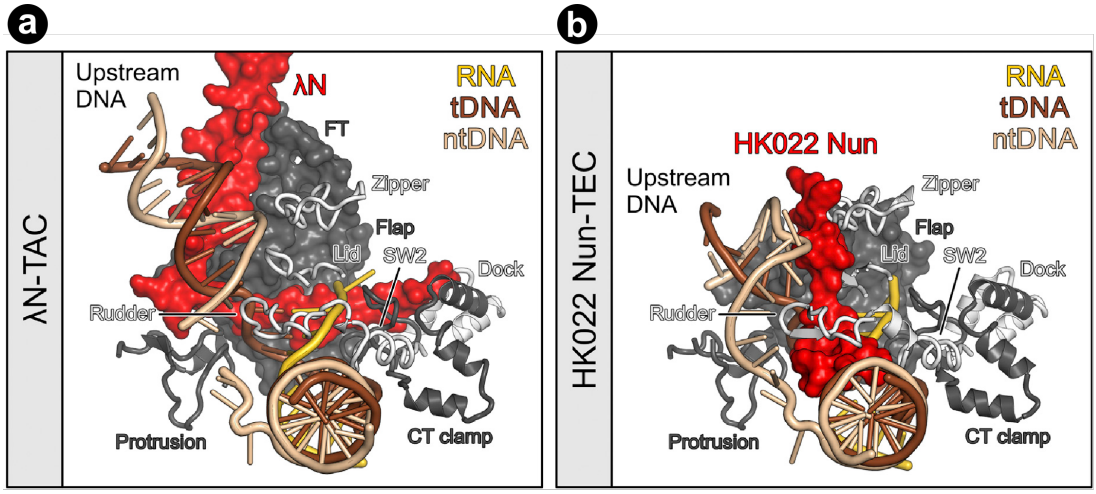



**Figure 4.8: Comparison of λN and HK022 Nun entering the active site cleft.**
***a)*** *Path of λN inside the active site of the λN-TAC.* ***b)*** *Path of the HK022 Nun inside the TEC (PDB: 6ALG).*

### 4.2.4 Regulation via the RNAP inner tunnel system

The C-terminal end of λN embodies a remarkable example how existing channels, tunnels, and cavities can be used for regulation. λN gains access to the active site by introducing minimal adjustments to the upstream DNA. Afterwards, a free passage is available across the hybrid to the RNA exit tunnel.

The recent structure of the HK022 Nun-stalled TEC allows a comparison of their mechanism. λN and Nun access the RNAP through adjacent openings but achieve an opposite effect (Figure 4.8). While λN traverses alongside the nucleic acids, Nun wedges between them, inhibiting their motion and thus arresting RNAP.

# Chapter 5

# 5 The architecture of protein synthesis in the developing neocortex at near-atomic resolution reveals Ebp1-mediated neuronal proteostasis at the 60S tunnel exit

## 5.1 Results of *Kraushar et al.*, under review

Protein synthesis is a crucial driver of neuronal development as it represents the last step of gene expression. Nevertheless, the analysis of ribosomal complexes within the developing mammalian brain is at an early stage of understanding. In the following, we pursued analysis of *ex vivo* derived mammalian neocortex ribosomes during developmental neurogenesis. We identified Ebp1 as a 60S peptide tunnel exit cofactor at near-atomic resolution with cryo-EM. Together with additional analysis Ebp1's function in proteostasis was shown. Thus, Ebp1 was established as a central component at the 60S peptide tunnel exit site for the control of neuronal morphology during protein synthesis.

### 5.1.1 Ebp1 is an abundant, high occupancy translation cofactor during development.

In order to analyze the architecture of neocortical ribosomal complexes across different time points of development, translating ribosomes were extracted *ex vivo* (Figure 5.1 a). Samples where derived from the early NSC predominate embryonic stage (E12.5) up to the post-mitotic neuronal postnatal stage (P0). Neocortex lysates were fractionated by sucrose gradient ultracentrifugation into 80S monosomes and actively translating polysomes and samples were analyzed by MS at all five time points (Figure 5.1 b).

As expected, these data quantified core ribosomal proteins (RPs) as the most enriched proteins in polysomes. Moreover, polysomes were enriched for various translation-associated proteins through development. Unexpectedly, Ebp1 was observed at rates similar to RPs themselves within fractions containing polysomes and monosomes. Additionally, Ebp1 was found to be among the most abundant proteins in the neocortex across development (Figure 5.1 b). Additional analysis of the MS-data measured Ebp1 at approximately stoichiometric levels to the ribosome throughout development, similar to the core RPs. These data suggested a key role of Ebp1 during neocortical development.

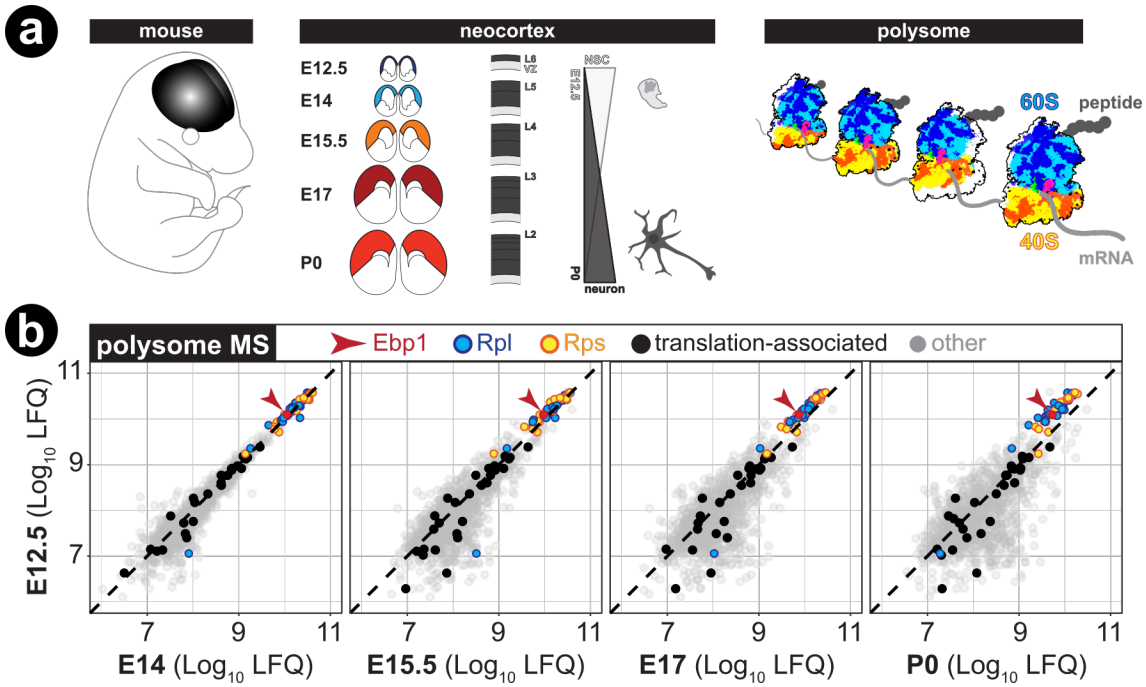



**Figure 5.1: Ebp1 is nearly stoichiometric to RPs in polysomes.**
***(a)*** *Schematic of the experimental system to measure the architecture of active protein synthesis (polysomal ribosomes) from the* ex vivo *neocortex across embryonic (E12.5, E14, E15.5, E17) and early postnatal (P0) neurogenesis.* ***(B)*** *MS analysis of neocortical polysomal complexes across development. Scatter plots compare early neurogenesis E12.5 vs. each subsequent time points, demonstrating the enrichment of Ebp1 (red arrow) among ribosomal proteins (RPs) of the large (Rpl, blue) and small (RPs, yellow) subunits, in contrast to other translation-associated proteins (black).*

### 5.1.2 The structure of Ebp1-ribosome complexes *ex vivo* at near-atomic resolution

In order to analyze the architecture of neocortical ribosome complexes at near-atomic resolution, 80S and polysomes were purified by sucrose density gradient fractionation from P0 neocortex lysates, pooled, and vitrified for cryo-EM. Obtained micrographs verified the mixture of 80S and polysomes within the sample (Figure 5.2 a). High-resolution cryo-EM data collection and SPA without 3D classification resulted in a complete 80S map with additional density adjacent to the 60S peptide tunnel exit. Using the crystal structure of mouse Ebp1 (PDB 2V6C) unambiguously identified Ebp1 in complex with 60S (Figure 5.2 b and c). Continuous density was present for nearly the entire N-terminus, ensuring the presence of full length (p48) Ebp1. This direct visualization of an *ex vivo* derived Ebp1-60S complex strongly supports its the physiological nature in the neocortex.

Hierarchical sorting was performed to entangle the different conformational and compositional states of the *ex vivo* derived ribosomes (5, 62) (Figure 5.3). In the first tier, the data set was divided into rotated and classical state as defined by the rotation of the small subunit (SU) (Figure 5.3 tier 1) and a “junk” population.

These two populations were separated and further classified individually. In tier 2, the rotated state was separated into ribosomal classes with (1) eEF2 and with (2) eEF2+P/E-tRNA. The classical state was divided into a ribosomal class with (3) A/A+P/P tRNA, (4) E/E-tRNA, and an (5) empty one. At tier 3, the separation focused on the presence and absence of Ebp1. Therefore, a modified strategy of 3D focused classification called “nue” mode, was applied.

The “nue” mode, named after the hybrid beast in Japanese folklore, creates a hybrid map for each class. Before the classification step at each iteration a new “nue” volume is created following these 5 steps:

1. Calculate “outside” volume as a weighted average using the class distribution and 3D classes of the last iteration.
2. Define a soft-focus mask based on the focus mask by adding a smooth fall-off.
3. For each 3D class:
    a. Define the “inside” volume by extracting the ROI of the 3D class volume using the soft-focus mask.
    b. Create “nue” volume by combining the general “outside” and the class-specific “inside” volume.
    c. Filter and normalize “nue” volume.
4. Use “nue” volumes as the reference for 3D focus classification.

The “nue” volumes of an iteration only differ within the ROI, thereby influences of heterogenous regions outside the focus mask are diminished, improving the sensitivity of the classification. A similar procedure has been implemented in cisTEM, where the area outside the ROI can be filtered or down-weighted to reduce its influence (31, 46, 352). The application of “nue” allowed a more precise separation of the particle images into 10 classes, 5 with and 5 without Ebp1.

Statistics on the distribution of particle images across these 10 classes revealed an occupancy of 48 % for Ebp1:ribosome complexes (Figure 5.3). High-resolution refinement of Ebp1-bound and unbound populations in the rotated state were performed, respectively. This resulted in four cryo-EM density volumes at 3.1 Å (for the rotated state) and 3.3 Å (for the classical state) global resolution. These four populations illustrate the binding of Ebp1 to both actively translating (classic state with A/A+P/P-tRNA) and non-translating ribosomes (rotated state with eEF2). These findings support a high occupancy of ~50 % of Ebp1 at the 60S subunit, similar to the results obtained by MS (data in manuscript).

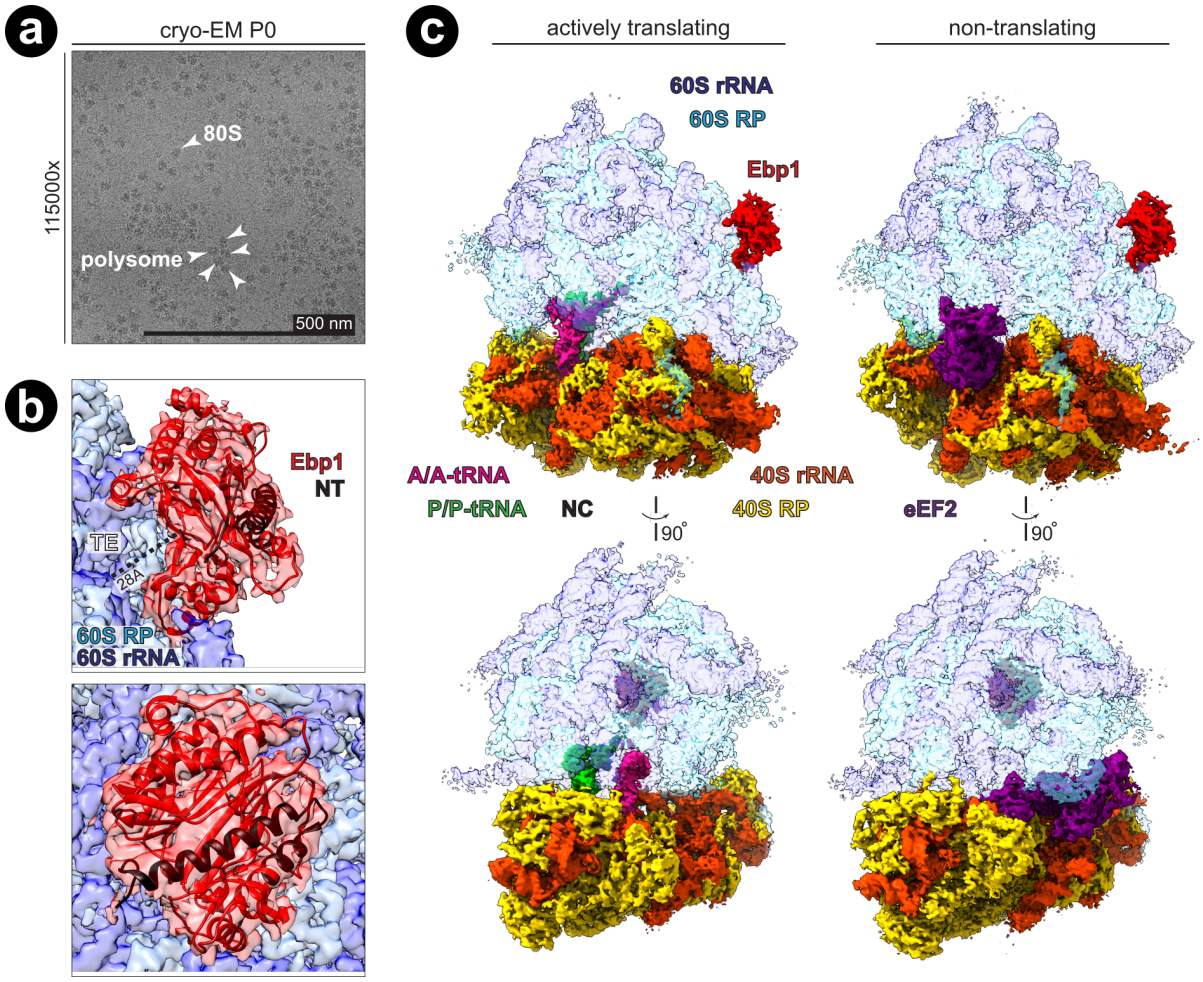



**Figure 5.2: Structure of the Ebp1:60S complexes at near-atomic resolution.**
***(a)*** *Cryo-EM micrograph of pooled 80S and polysome complexes (arrows) derived* ex vivo *from P0 mouse neocortical lysate.* ***(b)*** *Cryo-EM density map with an additionally density (red) corresponding to the structure of full length mouse Ebp1 (PDB 2V6C* (262) *including the N-terminal residues (NT, black ribbon). rRNA (dark blue) and RPs (light blue) of the 60S are depicted.* ***(c)*** *Hierarchical 3D classification resulted in ribosomal complexes with Ebp1 in both actively translating (left, classic) and non-translating states (right, rotated). Color code: Ebp1 (red), A/A tRNA (pink), P/P tRNA (green), eEF2 (purple), 60S rRNA (dark blue), 60S RPs (light blue), 50S rRNA (orange), 40S RPs (yellow), fragmented nascent chain (black).*

The near-atomic resolution allowed the modeling of the complete Ebp1:60S complex. Ebp1 contacts the immediate peptide tunnel exit (TE) surface by interacting with three rRNA helices (H24, H53, and H59) and four RPs (eL19, uL23, uL24, uL29) (Figure 5.4 a and b). The deep cavity of Ebp1 is positioned above the TE vestibule, stabilized by electrostatic interactions (yellow in Figure 5.4 b-e). However, this interaction does not seal the TE, but leaves large gaps, potentially allowing an emerging peptide chain to exit. Taken together, Ebp1 binds at the 60S TE, positioning its cavity above the TE vestibule and thereby forming a porous interface.

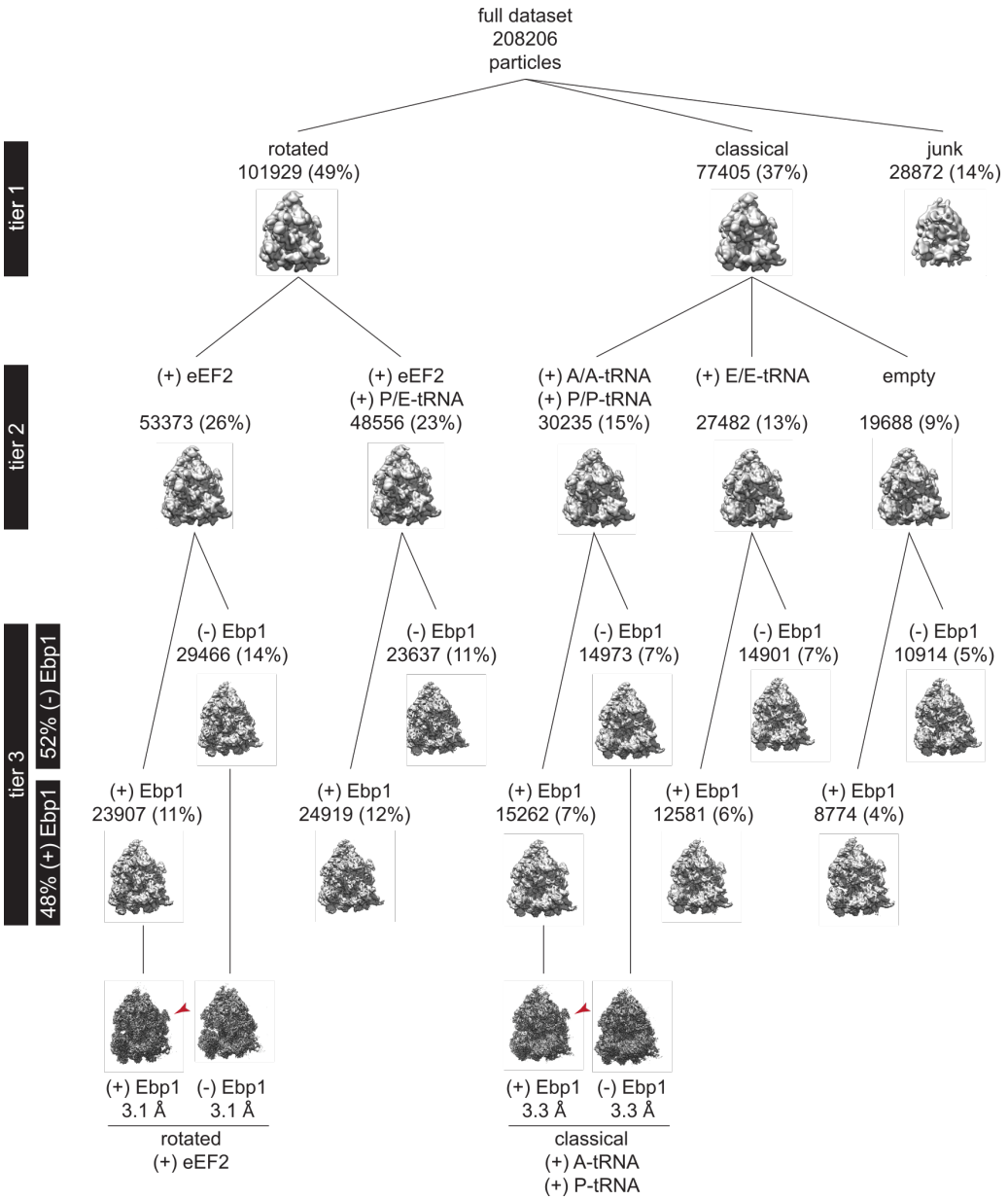



**Figure 5.3: Hierarchical sorting of the *ex vivo* derived polysomes at P0.**
*Cryo-EM imaging yielded 208,206 particle images, which were classified in 3 tiers (tiers 1-3).* ***(tier 1)****, 3D classification based on large scale heterogeneities was achieved, resulting in three classes corresponding to a rotated (49 %) and classical (37 %) state, as well as a "junk" class (14 %).* ***(tier 2)****, 3D focused classification was performed for the branches "rotated" and "classic" individually. For the "rotated" branch the two classes (1) eEF2, and (2) eEF2(+)P/E-tRNA were obtained. For the "classic" branch, three classes, (3) A/A-tRNA(+)P/P-tRNA, (4) E/E-tRNA, and (5) empty ribosome, were obtained.* ***(tier 3)*** *all 5 classes were separated into +Ebp1 and -Ebp1, resulting in 10 classes in total. The four highest populated classes were each refined to near-atomic resolution. Highest resolved classes correspond to eEF2(+)Ebp1, eEF2, A/A-tRNA(+)P/P-tRNA(+)Ebp1, and A/A-tRNA(+)P/P-tRNA.*

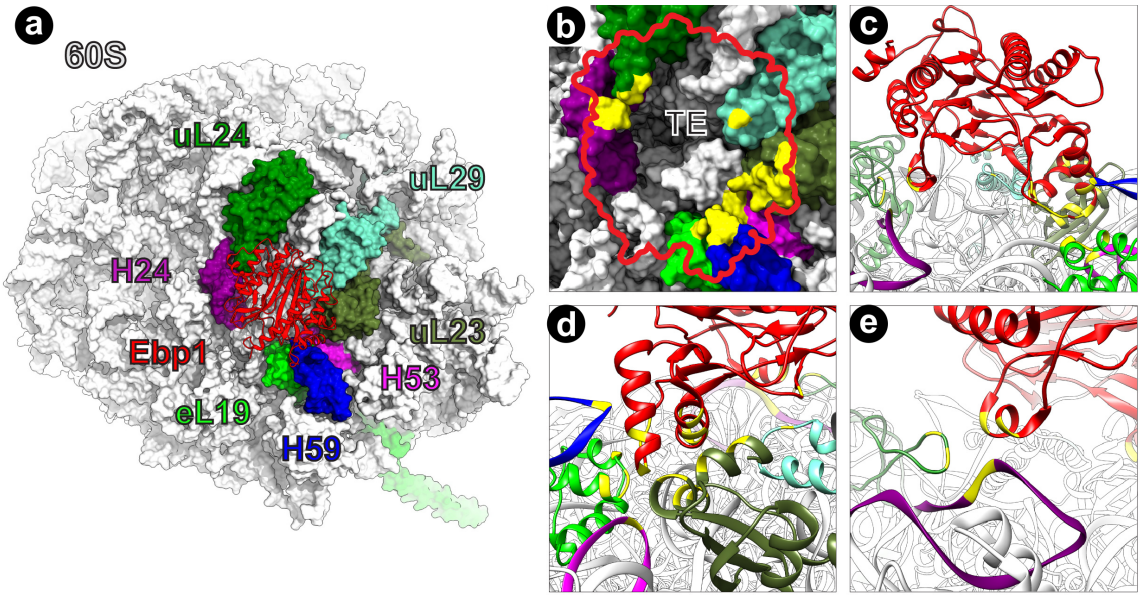



**Figure 5.4: Ebp1 binds above the 60S peptide tunnel exit.**
***(a)*** *Model of Ebp1 (red ribbon) binding at the 60S peptide tunnel exit. The interacting rRNAs and proteins are depicted. 60S rRNAs: H24 (purple), H53 (magenta), and H59 (blue). 60S RPs: eL19 (lime green), uL23 (olive), uL24 (forest green), and uL29 (aquamarine).* ***(b)*** *Arial view of Ebp1's footprint (red outline) at the tunnel exit (TE) with rRNAs and proteins colored as in (a). Electrostatic interaction sites are colored in yellow.* ***(c-e)*** *Detailed view of the Ebp1:60S model with the binding residues/nucleosides highlighted in yellow.*

### 5.1.3 Ebp1 binding requires a swinging latch mechanism

The extensive hierarchical sorting of the cryo-EM data set into Ebp1-bound and unbound classed allowed us to describe Ebp1-binding with an internal negative control. Comparing Ebp1-bound with the unbound structure allowed the identification of G-2690 at rRNA helix H59 to undergo a 235° flip (Figure 5.5 a) to form intra-helical base stacking interactions when Ebp1 is bound. In the absence of Ebp1, H59 G-2690 contacts H53 C-2501, G-2502 and C-2513. In the presence of Ebp1 when G-2690 breaks this interaction and flips, a rearrangement occurs in the tip of H59. This rearrangement in H59 critically includes U-2687 flipping outwards to form interactions with the insert domain of Ebp1. These structural rearrangements were termed "swinging latch" mechanism and have been previously observed for the binding of yeast peptide tunnel quality control factor Arx1 (353, 354). Taken together, Ebp1-binding to the 60S requires a swinging latch mechanism of the H59 similar to the Ebp1-related yeast protein Arx1.

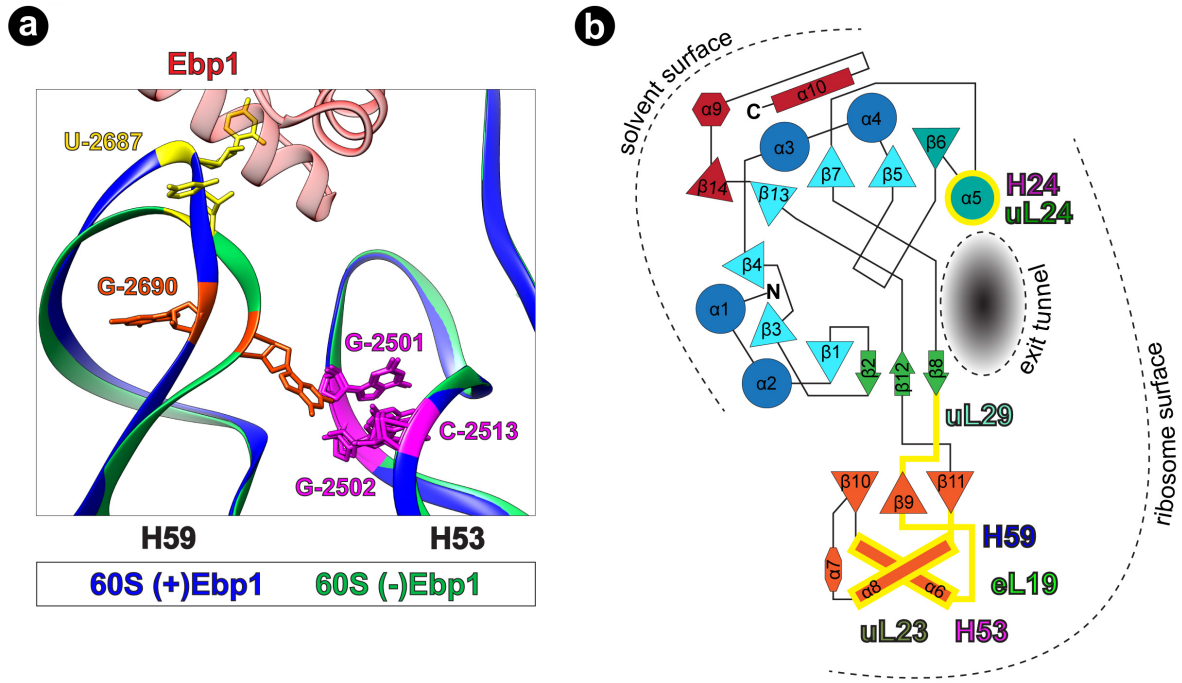



**Figure 5.5: Ebp1 utilizes a swinging latch mechanism for 60S binding.**
***(a)*** *60S rRNA helices H59 and H53 depicted in the presence (blue) and absence (green) of Ebp1, adjacent to the Ebp1 insert domain (red).* ***(b)*** *Diagram of Ebp1's secondary structural elements, their orientation to and interactions with (highlighted in yellow) the ribosome surface. Figure adapted from* (261).

### 5.1.4 Ebp1 binding is mutually exclusive to other peptide tunnel exit cofactors

The structure of the Ebp1:60S complex assigned new functions to structural domains of Ebp1 (Figure 5.5 b, adapted from (261)). Ebp1 contacts the ribosome with its insert domain on one side and the α5 helix on the other, positioning β sheets 1, 3, 4, 5, 7, and 13 directly above the TE.

Ebp1, Metap2, and Arx1 all share similar structural features (Figure 5.6 a): (i) A common insert domain, facilitating 60S binding, (ii) the pita bread fold positioned over the peptide TE, and (iii) the α4 helix available for molecular interactions at the solvent side. All three factors provide different electrochemical environments for the nascent peptide (Figure 5.6 b). Ebp1 and Metap2 both provide a deep, strongly electronegative pocket. In contrast to Metap2, Ebp1 is catalytically inactive due to missing key resides within the pita bread fold (261, 262, 355). However, since the structure of the Metap2:60S complex remains unsolved, structural adjustments of Metap2 during binding may exist.

Structural elucidation of the Arx1:Rei1:ribosome complex (353) revealed the existence of Rei in the peptide tunnel, probing the 60S as a quality-control stop of preempting active translation. Therefore, Ebp1, Metap2, and Arx1, despite their conserved structure and binding mode, interact with functionally distinct states of 60S-nascent chain complexes and provide distinct environments for the nascent chain (Figure 5.6 b).

By comparing the footprints of different eukaryotic TE cofactors, a mutual exclusivity of their binding can be visualized. In Figure 5.6 c the footprints of the following cofactors are

superimposed: Ebp1, Metap2 (355), Arx1 (280), ER translocation channel Sec61 (281), Ltn1-NEMF ubiquitin ligase complex (356), N-terminal acetyltransferase NatA (357), the ribosome-associated complex (RAC) (358), and the nascent polypeptide-associated complex (NAC) (359, 360). The high abundance of Ebp1 compared to these cofactors and its dynamic turnover, might allow binding of other cofactors once a signal within the nascent chain is emerging from the TE. The dynamic enrichment of Ebp1 may represent a specific stage during neurogenesis. Due to its high occupancy, Ebp1 depletion during neurogenesis may disrupt the balance of proteostasis during that crucial time.

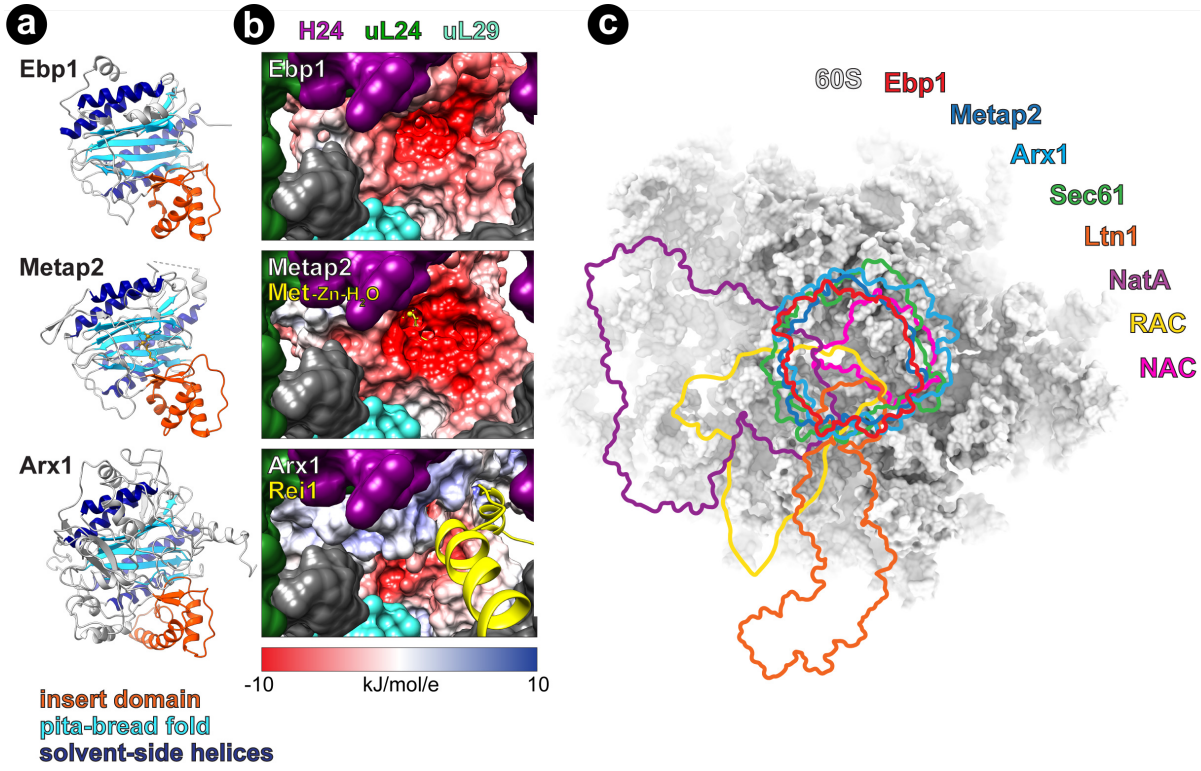



**Figure 5.6: Ebp1 compared in isolation and to other TE cofactors.**
***(a)*** *Comparison of Ebp1, Metap2, and Arx1. All three share the insert domain (orange), the pita bread fold (light blue) and the α4 helix, on the solvent side (dark blue).* ***(b)*** *Electrostatic potential maps for Ebp1, Metap2, and Arx1 viewed from the perspective of the emerging nascent chain.* ***(c)*** *Schematic footprints of known TE cofactors on the surface of 60S. Protein (PDB ID) Metap2 (1KQ9), Arx1 (5APN), Sec61 (3J7R), Ltn1 (3J92), NatA (6HD7). Protein (EMDB ID): RAC (6105), NAC (4938).*

## 5.2 Discussion of *Kraushar et al.*, under review

The obtained structural results of this study elucidated the architecture of the 80S:Ebp1 complex extracted from the developing neocortex (Figure 5.1). The obtained cryo-EM reconstruction allowed the unambigious identification of Ebp1 bound to the TE (Figure 5.2), making contacts so several surrounding rRNAs and proteins (Figure 5.4). Here, Ebp1 forms a porous interface at the TE, allowing emerging peptides chains to exit. Additionally, hierarchical sorting using the “nue” focused classification mode entangled different conformational and compositional states.

Comparison of the Ebp1-bound and -unbound classes, allowed the postulation of a swinging latch mechanism involving Helix H59 and H53 as previously observed for the related protein Arx1 (353). Comparison of Ebp1 to its relatives Metap2 and Arx1 reveales their distinct enivroment for the nascent chain (Figure 5.6 b). By comparison of Ebp1 with different eukaryotic TE cofactors, a mututal exclusivity of their binding site was visualized (Figure 5.6 c).

Statistics on the Ebp1-bound and -unbound classes revealed a high occupany of Ebp1 (48 %) across translating and non-translating ribosomes. Additional data presented in the full manuscript illustrate Ebp1 is a participant in ribosomal regulation with high occupancy and strong affinity to the 60S TE on both translating and non-translating ribosomes.

During the analysis of gene expression throughout neocortical neurogenesis, the primary focus was until recently placed on transcriptional control (361, 362). Recent work improved the analytical depth to the single-cell levels and assigned transcriptional signatures to cell subtypes. Additionally, suggestions about successive layers of regulation (363) like post-transcriptional mechanism and environmental signals arose.

Precisely regulated spatiotemporal gene expression (242–244) during neocortical neurogenesis depends on the ribosome as its final step (364–366). In this work, the neocortical protein synthesis apparatus and more specifically the 60S TE were found as a locus of control in neurogenic gene expression. The 60S TE represents an interaction platform for cofactors, which exhibit a dynamic equilibrium in their competition to sculpt protein synthesis (360, 367).

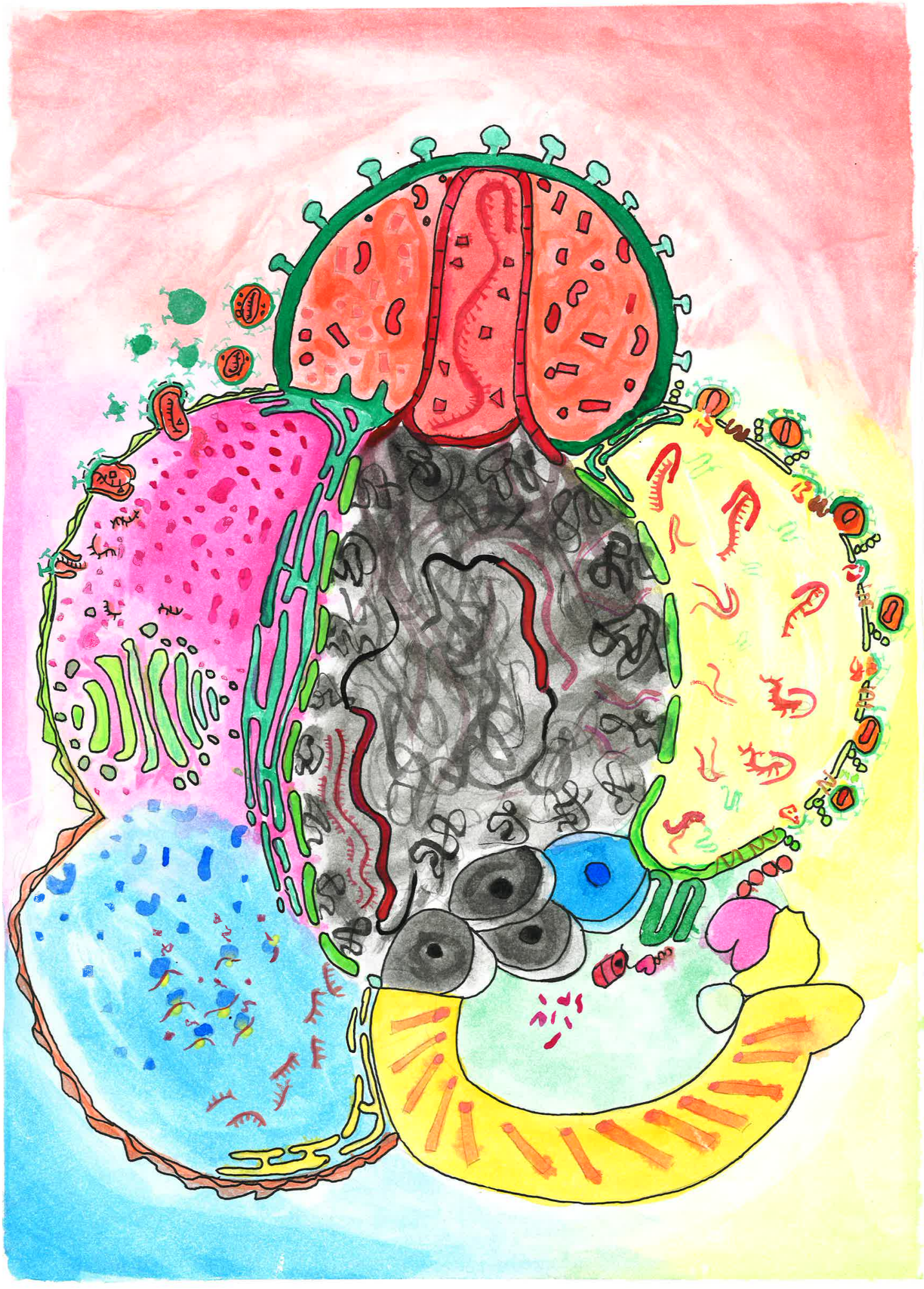
Overall, these findings contradict previous definitions of Ebp1 as a niche regulator of gene expression (266, 267). The expression of Ebp1 is cell type and temporally specific and aligns well with a transient abundance of ribosomal complexes at these developmental stages (data in manuscript). Thereby, Ebp1 contributes to ribosomal complex heterogeneity (250, 368) and is located on the ribosomal surface to shift the balance of proteostasis.

The concentration of Ebp1 with actively translating ribosomes at the synapse hints to a local proteostasis control, opening up future questions of subcellular action of Ebp1:ribosome complexes.

The establishment of Ebp1 into its ribosomal and extra-ribosomal mechanisms in the nervous system provides several unanswered questions. Ebp1 deletion has been implicated

with growth restriction in mice and *Arabidopsis* (369–371). Ebp1 plays a role in gene expression in stem cells of the neuroectoderm lineage (372), the specification of the neural border zone, neural crest, and cranial placode domains in *Xenopus* (373).

How the development of the nervous system, its dysfunctions, functional gene expression, and the role of Ebp1 at the 60S TE intertwine, remains an interesting future direction of research.



# Chapter 6

# 6 Structural insights into Cullin4-RING ubiquitin ligase remodeling by Vpr from simian immunodeficiency viruses

## 6.1 Results of *Banchenko, Krupp et al.*, under review

In the following, results from an integrative structural and biochemical analysis of the Vpr protein derived from SIV infecting moustached monkey (SIVmus) will be presented. The results allow to pinpoint the divergent molecular mechanism of SIVmus Vpr-dependent SAMHD1 recruitment to $DCAF1^{CTD}$. Additionally, cryo-EM was employed to gain insights into the structure and dynamics of a full SAMHD1:Vpr:$CRL4^{DCAF1}$ complex, with implications for the ubiquitin transfer mechanism in general.

### 6.1.1 Molecular mechanism of SAMHD1-targeting

In order to gain mechanistic insights into the Vpr-mediated recruitment of $SAMHD1^{CTD}$ to DCAF1, cryo-EM in combination with SPA was performed. In order to generate an active $CRL4^{DCAF1}$, NEDD8 was enzymatically attached to Cul4. Afterwards, the full complex was reconstituted *in vitro* and purified via gel filtration (GF), consisting of Vpr, SAMHD1, DDB1:$DCAF^{CTD}$ and NEDD8:Cul4:Roc1.

A data set of 3644 micrographs was collected and extensively classified in 2D and 3D. The general architecture of the full $CRL4^{DCAF1}$-complex can be divided into the „stalk“ (subcomplex of NEDD8:Cul4:Roc1) and the „core“ (SAMHD1:Vpr:DDB1:$DCAF1^{CTD}$). The data set exhibited a large degree of heterogeneity: (i) compositional heterogeneity due to the presence or absence of the „stalk“ and (ii) conformational heterogeneity due to the movement

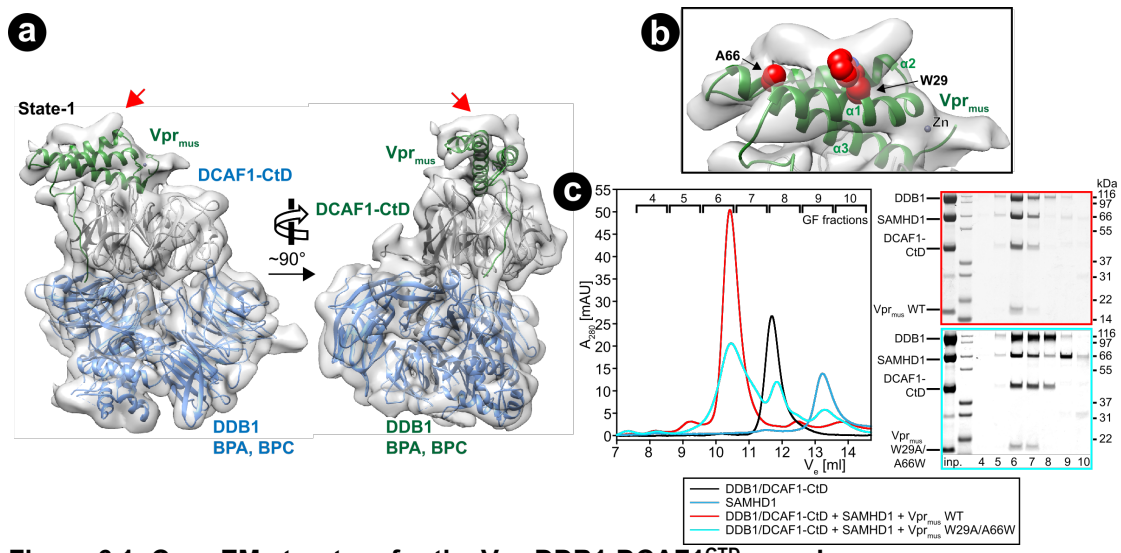



**Figure 6.1: Cryo-EM structure for the Vpr:DDB1:DCAF1$^{CTD}$ complex.**
*Two views of the cryo-EM reconstruction state 1. The determined MX structure was fitted as rigid body into the cryo-EM density. DDB1 model and cryo-EM density were removed for clarity. The red arrows mark an additional density on top of the Vpr helix bundle.* ***(b)*** *Detailed view of the extra electron density in the same orientation as in (a), left panel. Selected Vpr residues W29 and A66, which are in close contact to the additional density, are shown as red spheres.* ***(c)*** *Analysis of the double mutant Vpr$^{W29A/A66W}$ on complex formation via* in vitro *reconstitution followed by analytical GF. The inset shows SDS-PAGE analysis of corresponding GF fractions.*

of the „stalk“ relative to the core. The first analysis of the data set yielded one 3D class, “core”, without a „stalk“ at a resolution of 7.3 Å. Molecular models of DDB1$^{BPB/BPC}$ and of the Vpr:DCAF1$^{CTD}$ crystal structure were placed in the density via rigid body docking (Figure 6.1 a).

Upon closer inspection of the outer Vpr interface, an additional, slightly curved tubular density was identified (red arrows in Figure 6.1 a). While one end of this density feature contacts helix 1, the other is in proximity to helix 2 and 3 of Vpr (Figure 6.1 b). Due to the low local resolution at this position (7.5-8 Å), no atomic model was generated. But in the light of biochemical results (data in manuscript), the additional tubular density was hypothesized to correspond to a sequence stretch in SAMHD1$^{CTD}$ physically interacting with Vpr. Taken the dimensions of the tubular density into account, either ~10 aa in an extended conformation or up to 23 aa in a bent helical arrangement might be accomodated. According to all available crystallographic structures and secondary structure predictions, a disordered state of the SAMHD1$^{CTD}$ downstream of the HD domain (599-626 aa) can be assumed (332). This disordered CTD seems to be bound to Vpr, thereby leaving the globular domains of SAMHD1 attached in a flexible way, adopting a multitude of positions relative to the CRL4 core. Such a high degree of conformational heterogeneity would be averaged out during the cryo-EM reconstruction.

In order to independently assess the binding mode of Vpr to SAMHD1$^{CTD}$, mutations within Vpr were introduced at their putative binding interface. The Vpr double mutant Vpr$^{W29A/A66W}$ was purified, where W29 was mutated to alanine to block a hydrophobic contact (W29A) and

an A66 was replaced by a bulky tryptophane to introduce a steric clash (A66W). The $Vpr^{W29A/A66W}$ was assessed in its ability of complex formation by analytical GF (cyan trace in Figure 6.1 c). Ternary complex formation was reduced by this double mutant (fraction 6), concomitant with (i) the increase of species representing a Vpr:DDB1:$DCAF1^{CTD}$ complex containing sub-stoichiometrically amounts of SAMHD1 (fractions 7-8), (ii) an excess of DDB1:$DCAF1^{CTD}$ and (iii) monomeric SAMHD1 (fractions 9-10). In combination, these biochemical and the cryo-EM results at intermediate resolution locate the $SAMHD1^{CTD}$ binding side on top of the Vpr helix bundle.

In order to assess the topology of SAMHD1:Vpr:$CRL4^{DCAF1}$ and the binding regions of $SAMHD1^{CTD}$, cross-linking mass spectrometry (XLMS) using photo-reactive cross-linking reagent sulfo-SDA was performed (374). A high number of cross-links between SAMHD1 and the C-terminal half of Cul4, $DCAF1^{CTD}$ and DDB1 residues 800-1000 were found, supporting a high variability of the SAMHD1 position relative to the CRL4 (Figure 6.2 a). Several cross-links between the $SAMHD1^{CTD}$, the C-terminal half of Vpr helix α1, the disordered part around $Vpr^{90-100}$, $Cul4^{CTD}$, and the “acidic loop” of DCAF1 were identified. Distance restraints from these cross-links were used for calculation of the interaction space accessible to the center of mass of $SAMHD1^{CTD}$ (Figure 6.2 b) (375). The obtained results are in agreement with the position of the $SAMHD1^{CTD}$ determined by cryo-EM (Figure 6.1 a and b). Interestingly, while residues 622 and 626 within the $SAMHD1^{CTD}$ cross-linked primarily to Vpr, the residues around 595, 596 602-606 of $SAMHD1^{CTD}$ crosslinked to Cul4 and DCAF1. These results suggest the immobilization of the very C-terminal part of SAMHD1 (aa >622) by Vpr while residues further upstream (aa <606) are exposed to the catalytic machinery in close vicinity to the $Cul4^{CTD}$.

These data allow for a structural comparison of *neo*-substrate binding modes with Vpx and Vpr proteins from different retrovirus species (Figure 6.2 c-f): (i) SIVsm/HIV-2 type Vpx recruits $SAMHD1^{CTD}$ to the side of the DCAF BP domain via Vpx helices 1 and 3 (Figure 6.2 d) (289), (ii) SIVmnd2-type Vpx provides a bipartite interface for $SAMHD1^{NTD}$ at the side of DCAF1 BP and on top of the Vpx helix bundle (Figure 6.2 e) (305), and (iii) HIV-1 Vpr engages UNG2 on top of the Vpr helix bundle (Figure 6.2 f) (306). However, all of these “top side” interfaces only partially overlap with the one of SIVmus Vpr:$SAMHD1^{CTD}$, and furthermore involve fundamentally different sets of residues for their interactions. Therefore, the interaction interfaces of Vpr and Vpx for *neo*-substrate recognition are not conserved among related Vpx/Vpr accessory proteins, even when targeting identical $SAMHD1^{CTD}$ regions.

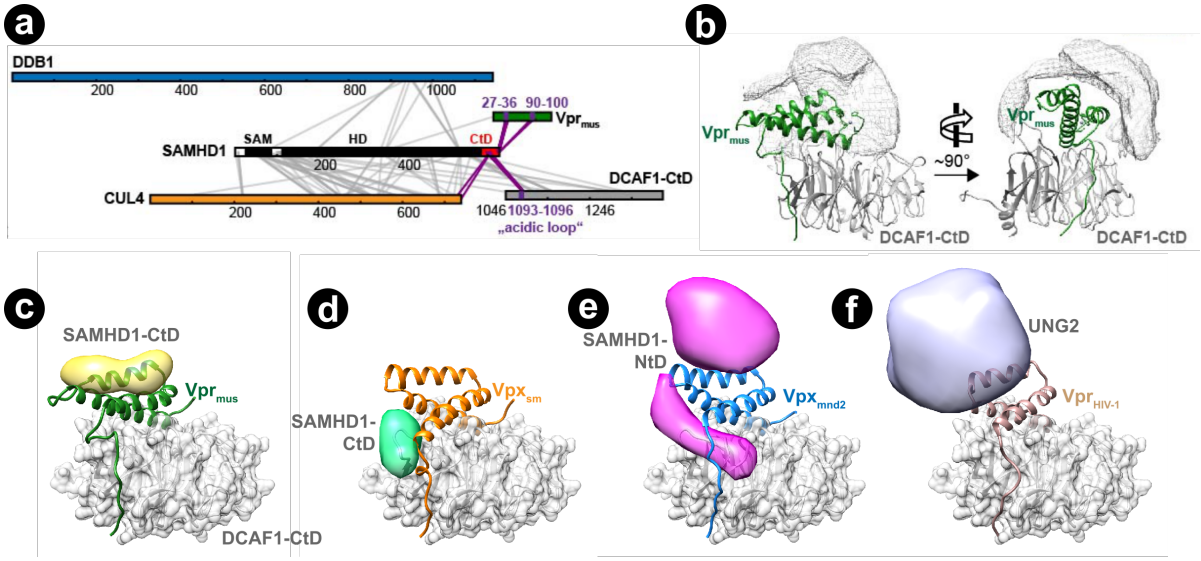



**Figure 6.2: Comparison of substrate recognition modes of different Vpr and Vpx proteins.** ***(a)*** *Schematic representation of cross-links (gray lines) within the SAMHD1:Vpr:CRL4$^{DCAF1}$ complex and cross-links to SAMHD1$^{CTD}$ (purple lines).* ***(b)*** *Accessible interaction space of SAMHD1$^{CTD}$ depicted as gray mesh as calculated by the DisVis server* (375) *using 14 out of 26 observed cross-links.* ***(b-f)*** *Comparison of* neo*-substrate recognition modes of SIVmus Vpr* ***(c)****, SIVsmm Vpx* ***(d)****, SIVmnd Vpx* ***(e)****, and HIV-1 Vpr* ***(f)*** *proteins. For clarity, molecular models of the recruited substrates are shown as strongly filtered, semi-transparent calculated densities.*

### 6.1.2 Cryo-EM analysis of the Vpr-modified CRL4$^{DCAF1}$ reveals different conformational states

In order to assess the conformational heterogeneity of the „stalk“-movement, a second analysis of the cryo-EM data set was performed. Therefore, a more strict selection of high-quality 2D class averages depicting secondary elements was executed, followed by extensive hierarchical 3D classification. By employing an unsupervised 3D classification strategy three different states of the “stalk” relative to the “core” were identified: state 1, 2 , and 3, respectively (Figure 6.3 a). Their nominal resolution ranged from 8-10 Å. The quality of the cryo-EM density maps was sufficient to discriminate α helices as tubular structures and allowed an interpretation using MX models (Figure 6.3 b and e). First, a model for the full complex CRL4$^{DCAF1}$ was assembled by aligning the DDB1:DCAF1$^{CTD}$ crystal structure to the “stalk” in PDB ID 2HYE via superposition of the DDB1 BPB domains (286). Then, this model was separated at the interface between DDB1$^{BPB}$ and DDB1$^{BPA/BPC}$, yielding two individual entities, Cul4:Roc1:DDB1$^{BPB}$ and DDB1$^{BPA/BPC}$:DCAF1$^{CTD}$, which were used for the interpretation of “stalk” and the “core” cryo-EM densities, respectively.

Only fragmented density was found for the RING-protein Roc1. Additionally, the location of NEDD8-attachment was only poorly resolved and did not allow for a comparison with available structures of Cul5 in complex with NEDD8 (376). Structral alignment of the individual “core” cryo-EM volumes of states-1, -2 and -3 indicate a good superposition and no major conformational fluctuations. Accordingly, the “core” can be seen as a rigid platform for the

binding of substrates and the attachment of the “stalk”. The three states-1, -2, and -3 represent different rotational states of DDB1 BPB with the rotational axis located at the intra-DDB1 interface, between BPB and BPA/BPC (Figure 1.23 b). In such a way, the CRL4 catalytic machinery, sited at the distal end of the “stalk”, is able to approach Vpr-tethered SAMHD1 from a variety of angles to catalyse Ub-transfer. These states and their rotation of the “stalk” around the “core” by ~150° are in line with previous comparative crystal structure analysis (286, 288, 294, 377, 378). Additional support for these three states comes from XLMS results, where some of the obtained distance restraints are specific for each state.

Inspection of states 2 and 4 revealed additional molecular interactions between DDB1 and Cul4 in addition to their main interface between $DDB1^{BPB}$ and $Cul4^{NT}$: in state-1, the loop Cul4 (residues 161-169) is in close proximity to a loop in DDB1 BPC (residues 795-801) (Figure 6.3 c). In state-3, the loop Cul4 (275-282) touches a region within the C-terminal helical domain of DDB1 (residues 1110-1127) (Figure 6.3 d). These interactions may provide guidance for the two outmost stalk positions and thereby confine the rotational range of the Cul4-“stalk”.

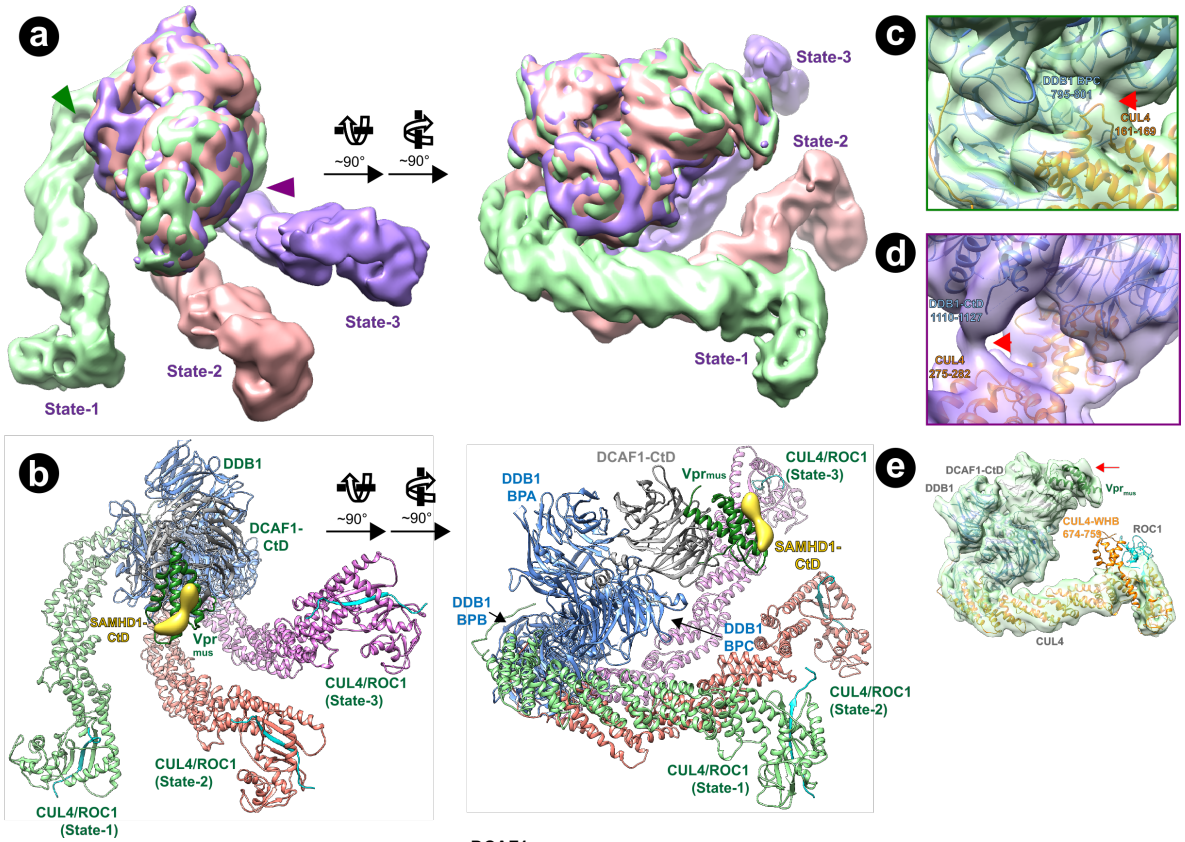



**Figure 6.3: Cryo-EM analysis of $CRL4^{DCAF1}$ conformational state.**
***(a)*** *Two views of cryo-EM reconstructions of SAMHD1:Vpr:$CRL4^{DCAF1}$ conformational states 1-3, light green, salmon, and purple, respectively. All three states have been aligned on the “core” consisting of the DDB1:$DCAF1^{CTD}$. View points for c) and d) are indicated by a green and purple arrow, respectively.* ***(b)*** *Superposition of the PDB models for the “core” and “stalk” as fitted into the states 1-3, as shown in (a). All models are depicted in ribbon representation. Tubular density (yellow) from state-1 corresponding to $SAMHD1^{CTD}$ is shown to illustrate the location of the substrate in the context of the complete $CRL4^{DCAF1}$.* ***(c and d)*** *Detailed view of state-1 (c) and state-3 (d) densities at their additional Cul4-DDB1 interface. Red arrows indicate contact points. (e)*

## 6.2 Discussion of *Banchenko, Krupp et al.*, under review

The present study provides insights into the mechanism of a class of "hybrid" SIV Vpr proteins by employing an integrative structural biological approach. The results allowed the formulation of a divergent molecular mechanism of SIVmus Vpr-dependent SAMHD1 recruitment towards $DCAF1^{CTD}$.

### 6.2.1 SIVmus Vpr utilizes distinct structural elements for SAMHD1 recruitment

SIV Vpr proteins are present in SIVagm and SIVdeb/SIVmus/SIVsyk lineages of lentiviruses and combine characteristics of HIV-1-type Vpr and SIV Vpx accessory proteins. Comparison of the SIVmus Vpr structure reported here to published structures of Vpx (49-51) and HIV-1 Vpr (54) reveals a conserved three helix bundle fold at similar position on the $DCAF1^{CTD}$ surface (Figure 6.2 c-f). Additionally, the majority of DCAF1-interacting side chains are type-conserved within Vpx and Vpr proteins (Figure 6.2 c-f). These observations hint at a common molecular mechanism of Vpr/Vpx-hijacking of $CRL4^{DCAF1}$.

"Hybrid" Vpr accessory proteins, similar to Vpx, promote viral replication by down-regulating the host restriction factor SAMHD1. To achieve that goal, Vpr/Vpx hijack $CRL4^{DCAF1}$ to recruit SAMHD1 as a *neo*-substrate for poly-ubiquitination and subsequent proteasomal degradation. Viral down-regulation of SAMHD1 evokes an evolutionary "arms race" between virus and host, leading to changes within the amino acid sequence of both host SAMHD1 and virus Vpr/Vpx, respectively. This "arms race" formed two clades of Vpx proteins within lentiviruses, possessing divergent amino acid sequences upstream of helix 1 termed variable region 1 (VR1), which in combination with polymorphisms in $SAMHD1^{NTD}$ determine if the SAMHD1 CTD or NTD are recruited to the Vpx helix bundle (289, 305, 338, 379). SIVmus Vpr evolved a different mechanism for specificity towards $SAMHD1^{CTD}$. Instead of utilizing VR1 like Vpx proteins, Vpr employs an interface comprising helices 1 and 3 to immobilize $SAMHD1^{CTD}$ on top of the three-helix bundle. Here, the *neo*-substrate $SAMHD1^{CTD}$ is positioned in a way that precludes direct interactions between SAMHD1 and DCAF1, as it is the case for Vpx (Figure 6.2 d and e). Nevertheless, robust ubiquitination of SAMHD1 was measured *in vitro* (data in manuscript) and its degradation *in vivo* (324).

Other "hybrid" Vpr orthologues exhibit a wide sequence divergence. SIVdeb Vpr, the closest relative to SIVmus Vpr, shares only 50 % of its residues at the putative $SAMHD1^{CTD}$ binding site. Tests *in vitro* and *in vivo* did not yield a clear preference for $SAMHD1^{NTD}$ or $SAMHD1^{CTD}$ recruitment (324, 337). SIVsyk Vpr specifically recruits $SAMHD1^{CTD}$ (337), but does not share a high conversation of putative SAMHD1-binding residues with SIVmus Vpr. SIVagm Vpr differs significantly in possible SAMHD1-contacting residues, containing additional stretches upstream of helix 1 and 3 (337, 339), respectively. Taken together, the evolutionary "arms race" of lentivirus-host adaptation resulted in high species-specificity displaying diverse molecular mechanisms for Vpr-mediated SAMHD1 recruitment to

$CRL4^{DCAF1}$. Further structural analysis of other Vpr/Vpx lineages will be necessary to illustrate the variety of virus-host co-evolution on the molecular level.

HIV-1 Vpr engages UNG2 by mimicking the DNA phosphate backbone, providing a hydrophobic cleft for specific DNA-binding residues of UNG2 (306). Such a mimicry mechanism might explain HIV-1 Vpr's binding promiscuity towards DNA- and RNA-binding host factors (316). Resulting promiscous degradation of DNA- and RNA-binding proteins was suggested to induce cell cycle arrest at the G2/M phase border, the best described phenotype for Vpr proteins so far (316, 317, 380). However, the structural basis for such a DNA-mimicry mechanism does not exist in SIVmus Vpr, due to the absence of such a hydrophobic cleft. Consequently SIVmus Vpr does not down-regulate UNG2 in a human T cell line (316). However, SIVmus, SIVsyk, and SIVagm Vpr also cause G2/M cell cycle arrest, hinting at the existence of further structural elements responsible for *neo*-substrate recognition in these and HIV-1-type Vpr proteins, besides the DNA-mimicking cleft. Therefore, additional structural characterizations are needed to better understand the Vpx/Vpr helical scaffold and its mechanism to promiscuously bind *neo*-substrates. Such information could be applied in a more general scope to rationally design PROTACs or "molecular glue"-like compounds (381, 382).

### 6.2.2 Visualization of conformational states for the $CRL4^{DCAF1}$

Cryo-EM SPA of a complete Vpr-mediated $CRL4^{DCAF1}$ complex with the *neo*-substrate SAMHD1 elucidated three conformational states. Rotation around the $DDB1^{BPB}$-$DDB1^{BPA/BPC}$-interface caused differing positioning of the Cul4-"stalk" relative to the DDB1:$DCAF1^{CTD}$ "core" (Figure 6.3 a and b). These results are in agreement with previous ones obtained by MX (286, 288, 294, 377, 378). Additionally, interactions between Cul4 and DDB1 (Figure 6.3 c-d) might confine the ubiquitination zone of CRL4, i.e. the space accessible for the Ub-transfer from E2 to the substrate (Figure 6.4 a). Also, the lack of stable density for the neddylated WHB domain and the catalytic Roc1 RING-domain indicates a high mobility of these distal stalk elements (Figure 6.4 b).

Two structural studies shed light on the mechanism of NEDD8-mediated stimulation of ubiquitination activity: In Duda et al. (376), NEDD8-attachment leads to re-orientation of the Cul1 WHB domain and the release of the Roc1 RING-domain from the Cul1-scaffold. In Baek et al. (284), neddylated $CRL1^{\beta\text{-}TRCP}$ with its substrate IκBα demonstrated substantial flexibility of NEDD8, Cul1 WHB, and Roc1 RING-domain in the pre-catalytic state. This conformational flexibility is necessary to activate multiple CRL1-dependent processes, such as poly-ubiquitination, substrate priming via RBR-type E3 ligases, or de-neddylation (284, 383–385).

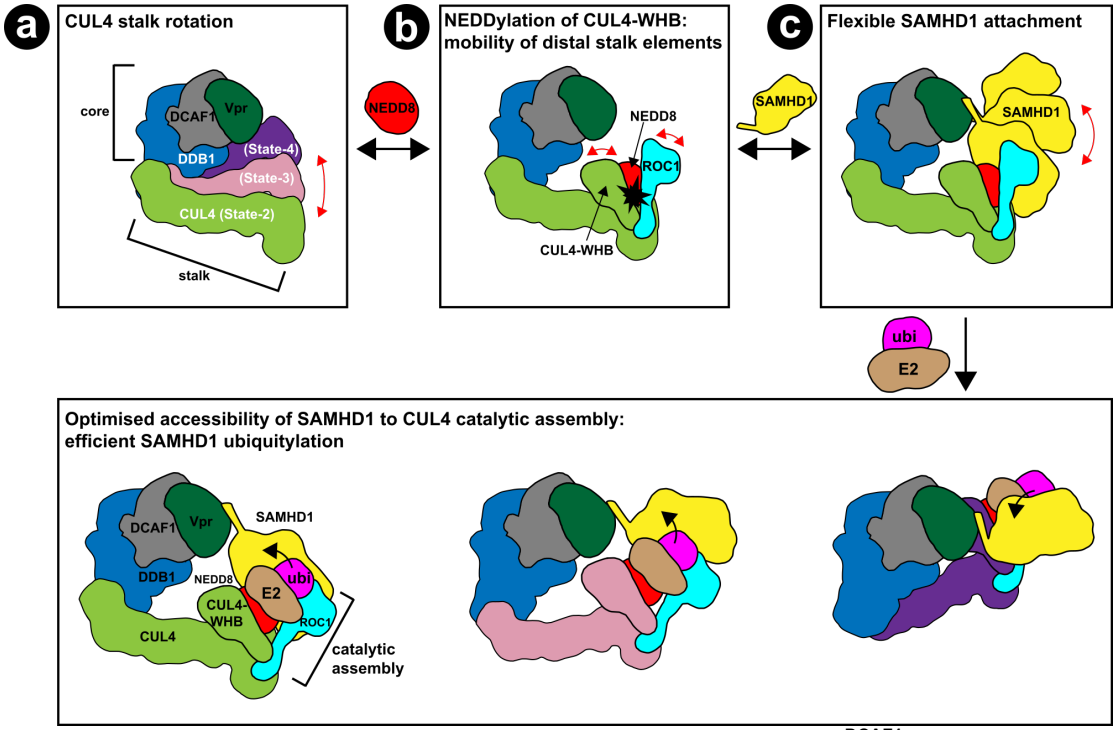



**Figure 6.4: Illustration of molecular motion of a Vpr-hijacked CRL4$^{DCAF1}$.**
*Schematic illustration of molecular motion and flexibility in Vpr-modified CRL4 (upper panel), and how these movements maximize accessibility of surface-exposed lysine side chains in SAMHD1 to the CRL4 catalytic assembly at the distal stalk end (lower panel).*

The cryo-EM structure of a CRL4-complex presented here might indicate that similar principles apply for CRL4. However, in order to elucidate and compare the CRL4 catalytic architecture, sophisticated crosslinking protocols as in Baek et al. (284) will be required.

The conformational states adopted by the CRL4 “stalk” in combination with the different DCAF substrate receptors likely creates an extended “ubiquitination zone” assisting in the ubiquitination of differently sized and shaped substrates. Substantial selective pressure to neutralize the host’s SAMHD1 resulted in the evolution of Vpr/Vpx accessory proteins in HIV-2 and certain SIVs, to hijack the dynamic CRL4$^{DCAF1}$ assembly. By flexibly recruiting either SAMHD1$^{NTD}$ or SAMHD1$^{CTD}$ to DCAF1, the accessibility of surface-exposed SAMHD1 lysine residues is maximized to allow for efficient Ub-transfer and poly-ubiquitination, followed by proteasomal degradation.

# Chapter 7

# 7 Conclusion and Outlook

Detailed functional and structural insights into the mechanism of several different MMMs have been obtained by cryo-EM in combination with SPA. Three different MMMs were studied, exhibiting a large range in molecular weight and topology: the λN-TAC containing the bacterial RNAP, the Ebp1:80S some derived *ex vivo*, and the SAMHD1:Vpr:CRL4$^{DCAF1}$. In all cases, cryo-EM elucidated the structure of the MMM at medium or high resolution, allowing their interpretation using atomic models. Additionally, several distinct 3D classes were obtained representing different conformational or compositional states of the MMM. Delivering such results, cryo-EM represents a valuable tool to study the structure and dynamics of an MMM assembled *in vitro* or derived *ex vivo*. Further optimization of sample preparation, data collection, and image processing will allow the visualization of more heterogeneous samples at higher resolution in the future. Specifically, the area of *in situ* structural biology will gain momentum in the upcoming years with the continuous improvement of cryo-tomography (cryo-ET), focused ion beam milling (FIB milling), and cryo-correlative light and electron microscopy (cryo-CLEM). Such development will further improve the power and necessity for the application of an integrative structural biology approach.

In the first study, *Said et al. 2017*, the structure of the “modifying RNP” was elucidated using MX and its position within the λN-TAC defined by cryo-EM and XLMS. The C-terminus of λN binds at the RNA exit channel and most likely stabilizes the processive state of the RNAP. Additionally, the C-terminus points towards the DNA:RNA hybrid, suggesting a path for the unresolved residues.

In the second study, *Krupp, Said, Huang et al. 2019*, a detailed structural and functional analysis of the full λN-TAC was performed. Structural information on the entire λN-TAC was obtained by cryo-EM and additional functional insights were gained using an *in vitro* transcription assay. The λN-TAC structure rationalizes numerous previous observations and mutants. λN implements a multi-pronged anti-pausing and anti-termination strategy by locally remodeling RNAP elements, globally reorganizing NusA, and promoting elongation. This is achieved over a distance of 60 Å by making extended contacts to several sites within the RNAP and the Nus factors. Taking together from both studies, λN is a remarkable example of an intrinsically disordered protein (IDP) to hijack a large cellular machinery.

By combining the results from both studies, a structural framework for the design of future experiments is provided. Questions addressing the regulation of the RNAP by other factors such as the ρ or NusA in the context of different RNAs could follow. Additionally, the mode of

λN-binding to the RNAP could be exploited to design new antibiotic peptides. Furthermore, the assembly of several elongation complexes along a longer template DNA could be performed to obtained *in vitro* Miller trees.

In the third study, *Kraushar et al.* provided a detailed analysis ranging from a near-atomic resolution structure of the Ebp1:ribosome complex to the impact of Ebp1 on neuronal development. Taken together, Ebp1 was revealed as a common and crucial central component of protein synthesis located at the 60S peptide tunnel exit, controlling gene expression during neuronal development. In order to further delineate the function of Ebp1 in recognizing nascent proteins, nascent peptides emerging from the tunnel exit could be used. Moreover, the analysis of translation regulation as an important layer during neuronal development should be assessed at the cellular level. Therefore, cryo-ET in combination with FIB milling represent a valuable tool to analyze ribosomes *in situ.*

In the fourth study, *Banchenko, Krupp et al.*, results for the structural mechanism of Vpr-hijacking of $CRL4^{DCAF1}$ to induce ubiquitination of the *neo*-substrate SAMHD1 were revealed by a combination of MX, XLMS, cryo-EM, and biochemical characterization. These results allowed the formulation of a diverging molecular mechanism for the Vpx/Vpr family and provided insights into the structural flexibility and dynamics of a CRL4-catalysed E3 ubiquitin ligases.

In the future, the mechanistic analysis of more cellular targets of Vpr/Vpx should provide general insights into their mechanism to hijack the $CRL4^{DCAF1}$. Further studies using SAMHD1 peptides should allow the detection of the precise sequence motif required for Vpr-interaction. Furthermore, the structural elucidation of the Ub-transfer onto the *neo*-substrate could be elucidated using a full SAMHD1:Vpr:$CRL4^{DCAF1}$ complex trapped using a ubiquitination mimicking crosslink.

# 8 Lebenslauf

Aus Gründen des Datenschutzes ist der Lebenslauf in der Online-Version nicht enthalten.

# 9 List of Publications

Krupp F.*, Said N.*, Huang H.-Y.*, Loll B., Bürger J., Mielke T., Spahn C. M. T., Wahl M. C. (2019) Structural Basis for the Action of an All-Purpose Transcription Anti-termination Factor. *Molecular Cell,* Volume 74, Issue 1, 4 April 2019, Pages 143-157.e5 (2019)

Said N.*, Krupp F., Anedchenko E., Santos K. F., Dybkov O., Huang Y.-H., Lee C.-T., Loll B., Behrmann E., Bürger J., Mielke T., Loerke J., Urlaub H., Spahn C. M. T., Weber G., Wahl M. C. (2017) Structural basis for λN-dependent processive transcription antitermination, *Nature Microbiology* **2**, 17062 (2017)

Pilsl M.*, Crucifix C., Papai G., Krupp F., Steinbauer R., Griesenbeck J., Milkereit P., Tschochner H., Schultz P. Structure of the initiation-competent RNA polymerase I and its implication for transcription (2016). *Nature Communication*, **7**, 12126.

# 10 Table of Figures

# 11 Table of Tables

# 12 List of Abbreviations

| | |
|---|---|
| °C | degree Celsius |
| A | Adenine |
| Å | angstroms |
| aa | amino acid |
| AR | acidic-rich |
| ARM | armadillo fold |
| ATP | adenosine triphosphate |
| BH | bridge helix |
| bp | base pair |
| BPA, BPA, BPC | β-propeller domains A, B , and C of DDB1 |
| C | Cytosine |
| CAMs | cell adhesion molecules |
| CCC | cross correlation coefficient |
| Cdt1 | cell division cycle protein 1 |
| Cdt2 | cell division cycle protein 2 |
| CK | casein kinase |
| CPU | central processing unit |
| CRL | cullin-RING E3 ligases |
| CRL4 | cullin4-RING E3 ligases |
| $CRL4^{Cdt2}$ | cullin4-RING E3 ligases with Cdt2 as substrate receptor |
| $CRL4^{DCAF1}$ | cullin4-RING E3 ligases with DCAF1 as substrate receptor |
| cryo-EM | cryo-electron microscopy |
| cryo-ET | cryo-electron tomography |
| cryo-CLEM | cryo-correlative light and electron microscopy |
| CT | C terminus |
| CTD | C terminal domain |
| CTF | contrast transfer function |
| DBD | DNA-binding domain |
| DCAF | DDB1 and Cul4-associated factor |
| DDB1 | damaged DNA binding protein 1 |
| DED | direct electron detector |
| DNA | deoxyribonucleic acid |
| ds | double stranded |
| dwDNA | downstream DNA |
| *e.g.* | *exempli gratia* |
| E1 | E1 ubiquitin activating enzyme |
| E12.5 | embyonic stage day 12.5 |
| E2 | E2 ubiquitin conjugating enzyme |

| | |
|---|---|
| E3 | E3 ubiquitin ligases |
| Ebp1 | ErbB3 binding protein 1 |
| eEF | eukaryotic elongation factor |
| eIF | eukaryotic initiation factor |
| EM | electron microscope |
| FIB milling | focused ion beam milling |
| FSC | Fourier shell correlation |
| FT | Fourier transform |
| G | Guanine |
| GF | gel filtration |
| GO | gene ontology |
| GPU | graph processing unit |
| H-box | helix-loop-helix motif |
| H59 | rRNA helix 59 |
| HECT | homologous to E7AP carboxyl terminus |
| HIV | human immunodeficiency virus |
| HLH | helix-loop-helix motif |
| *i.e.* | *id est* |
| immuno-EM | immuno-electron microscopy |
| IRES | internal ribosomal entry sites (IRESs), |
| kDa | kilo Dalton |
| KH1 | K homology domain |
| LisH | Lis1-homology |
| *Luc* | Luciferase mRNA |
| MDa | mega Dalton |
| Metap2 | methionine aminopeptidase |
| min | minute |
| MMM | macromolecular machine |
| MS | mass spectrometry |
| MX | macromolecular crystallography |
| NAC | nascent polypeptide-associated complex |
| nm | nanometer |
| NMR | nuclear magnetic resonance |
| NSCs | neuronal stem cells |
| NT | N terminus |
| nt | nuelotide |
| NTD | N terminal domain |
| NTP | nucleoside triphosphate |
| Nus | N-utilization substance |
| *nut* | N-utilisation site |
| OB-fold | oligonucleotide binding -fold |

| | |
|---|---|
| P | phosphate |
| P0 | post-mitotic stage |
| p21 | protein 21 kDa |
| PCNA | *p*roliferating *c*ell *n*uclear *a*ntigen |
| $PCNA^{DNA}$ | DNA-loaded PCNA |
| PPi | pyrophosphate |
| PTM | post-translation modification |
| RAC | ribosome associated complex |
| RBR | RING-between-RING |
| RFC | replication factor C |
| RING | really interesting new gene |
| RNA | ribonucleic acid |
| RNAP | RNA polymerase |
| RNP | ribonucleoprotein complex |
| ROI | region of interst |
| RPs | ribosomal proteins |
| RRL | rabbit reticulocyte lysate |
| *rut* | rho-utilisation site |
| s | second |
| SAMHD1 | SAM domain and HD domain-containing protein 1 |
| SAXS | small-angle X-ray scattering |
| SDS-PAGE | sodium dodecyl sulfate-polyacrylamide gel electrophoresis |
| SEC | size exclusion chromatography |
| SIV | simian immunodeficiency virus |
| SIVmnd | SIV infecting mandrills |
| $S_n2$ | nucleophilic substitution type 2 |
| SPA | single particle analysis |
| ss | single stranded |
| SUMO | small ubiquitin-related modifier |
| T | Thymine |
| T4L | T4 lysozyme |
| TAC | transcription anti-termination complex |
| TE | 60S peptide tunnel exit |
| TEC | transcription elongation complex |
| TEM | transmission electron microscope |
| THB | three helix bundle |
| TL | trigger loop |
| U | Uracil |
| Ub | ubiqutin |
| upDNA | upstream DNA |
| Vpr | viral protein r |

| | |
|---|---|
| VprBP | Vpr-binding protein |
| Vpx | viral protein x |
| WD40 domain | domain consisting of seven WD40 repeats |
| XLMS | crosslinking mass spectrometry |
| ZBD | zinc binding domain |
| βFT | β flap tip |
| βFTH | β flap tip helix |
| λN-TAC | λN-mediated transcription anti-termination complex |

# 13 Appendix

## Cdt2 is a master regulator of genomic stability – Starting a structural characterization via cryo-EM.


### Abstract

Cdt2 has emerged as a master regulator of genome stability by ubiquitinating key cell cycle proteins. Cdt2 is part of a Cullin4-RING E3 ligase complex ($CRL4^{Cdt2}$) and possesses a unique mode of substrate recognition. Both Cdt2 and its substrate interact with proliferation cellular nuclear antigen (PCNA) via their PIP box, a linear sequence motif. In this way, the substrate is brought in close proximity to Cdt2, resulting in its ubiquitination and subsequent proteasomal degradation. PCNA and several Cdt2-substrates are functionally involved in DNA replication, DNA damage repair, and cell cycle control, emphasizing Cdt2 as an important regulator of these processes. Deregulation of Cdt2 activity has been associated with various forms of cancer, underlining its importance for genomic stability.

In this study, preliminary results of the biochemical and structural characterization of Cdt2 are presented. Recombinant expression and purification of several $CRL4^{Cdt2}$ components and their *in vitro* ubiquitination activity was achieved. A complex consisting of Cdt2:DDB1:PCNA was successfully assembled and a preliminary structural characterization via cryo-EM was obtained. Taken together, these initial results provide a promising basis for a continued characterization of Cdt2's mechanism by applying an integrative structural biology approach.

## 13.1 Introduction

Cullin4-RING E3 ubiquitin ligases (CRL4) possess a modular architecture consisting of four proteins: (i) Cullin4 (Cul4) scaffold, (ii) Roc1, a RING protein, (iii) an adaptor protein DDB1 (DNA-damage binding protein 1) bridging substrate receptor and Cul4, and (iv) a substrate receptor belonging to the DCAF (*D*DB1 and *C*ul4-*a*ssociated *f*actors) family (273, 298). A CRL4 complex with a specific substrate receptor is denoted as $CRL4^{DCAF}$. The family of DCAF substrate receptors consist of over 60 members, representing a large variety of sequences and functional relations (285, 293, 313, 386, 387).

Among them, the substrate receptor Cdt2 (synonyms: DCAF2, DTL) has emerged as a master regulator of genome stability (298). The $CRL4^{Cdt2}$ (Cdt2:DDB1:Cul4:Roc1) targets key cell cycle proteins for degradation during S phase and after DNA damage via ubiquitination (298, 388). $CRL4^{Cdt2}$ has a unique mode of action: it recognizes its substrate only when presented by the proliferation cellular nuclear antigen loaded onto DNA ($PCNA^{DNA}$) (Figure A13.1 a). PCNA is a homotrimeric protein that encloses the DNA and is involved in DNA synthesis and the repair of DNA damage. Cdt2 as well as several of its substrates possess a PCNA interacting peptide motif (PIP box) that represents a loosely conserved linear 8 aa sequence motif (Figure A13.1 b and c) (389, 390). Through co-recruitment of Cdt2 and its substrates on PCNA, substrate degradation is confined to certain locations on chromatin. In this context, $PCNA^{DNA}$ functions as a landing pad for independent recruitment of $CRL4^{Cdt2}$ and its substrates. It was suggested that the substrate's PIP box binds to $PCNA^{DNA}$. This complex is recognized by Cdt2, initiating substrate ubiquitination(298). The full complex required for substrate recognition, substrate:$PCNA^{DNA}$:Cdt2:DDB1:Cul4:Roc1, and can be denoted as substrate:$PCNA^{DNA}$:$CRL4^{Cdt2}$.

Human Cdt2, encoded by the *DTL* gene, is a 730 aa protein and can be divided into an N- and a C-terminal part (Figure A13.1 d): (i) The N-terminal part (1-398 aa) bears seven WD40 repeats as well as a two WDxR motif. While the latter is responsible for its interactions with DDB1, the former likely recognizes PCNA and the bound degron (298, 388). (ii) The C-terminal half (399-703 aa) has been predicted as an intrinsically disordered region (IDR), interspersed with a DNA-binding domain (DBD), recognizing ssDNA as well as dsDNA. The CTD exhibits a high isoelectric point (pI) of 9.56, and contains several phosphorylation sites and another PIP box at its distal end (298, 388, 391). So far, a detailed structural and biochemical characterization of Cdt2 alone and in context of the whole $CRL4^{Cdt2}$ assembly has not been carried out.

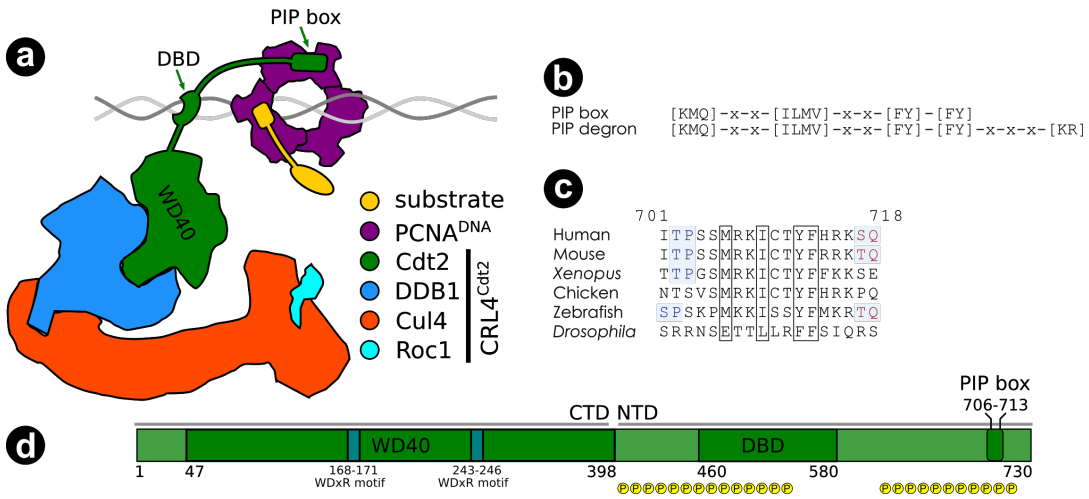



**Figure A13.1: Cdt2 domain organization, architecture and PIP box.**
***(a)*** *Schematic structure of the CRL4$^{Cdt2}$ in complex with PCNA$^{DNA}$ and a substrate. The interaction of PCNA$^{DNA}$ with the substrates' and Cdt2's PIP boxes are depicted as yellow and green boxes, respectively. The DBD domain of Cdt2 makes contact to the DNA.* ***(b)*** *The consensus sequence of the PIP box and the PIP degron are shown.* ***(c)*** *A sequence alignment of Cdt2$^{PIP}$ from different organism and putative phosphorylation sites are depicted (CDK, blue, and ATR, red). (c) and (d) were adapted from* (298). ***(d)*** *Domain organization of the human Cdt2 divided into N- and CTD. At the NTD, the WD40 domain consists of seven WD40 repeats. The CTD contains a DNA-binding region (DBD) and the PIP box. Putative phosphorylation sites within the CTD are marked and regions predicted as disordered are show in transparent color.*

The process of substrate ubiquitination by CRL4$^{Cdt2}$ is envisioned as a five-step process (Figure A13.2) (298): (i) PCNA is loaded onto the DNA during the S phase of the cell cycle or at sites of DNA damage. (ii) Either CRL4$^{Cdt2}$ or its substrate are recruited first to the PCNA$^{DNA}$ via interactions with their PIP boxes. (iii) The next component, either substrate or CRL4$^{Cdt2}$, binds to the PCNA$^{DNA}$. (iv) For degron recognition, the substrate's PIP box and surrounding residues are recognized at the surface of PCNA$^{DNA}$ by Cdt2's WD40 domain. (v) The substrate is ubiquitinated and subsequently degraded by the proteasome (392–395). The PIP degron seems to consist of the substrate's PIP box, a basic residue located 4 aa downstream (+4) of the PIP box, and an acidic residue of PCNA (Figure A13.1 b). The ternary interface consisting of PCNA$^{DNA}$, the substrate PIP box, and Cdt2, ensures the ubiquitination of only PCNA-bound PIP degron-containing substrates (393).

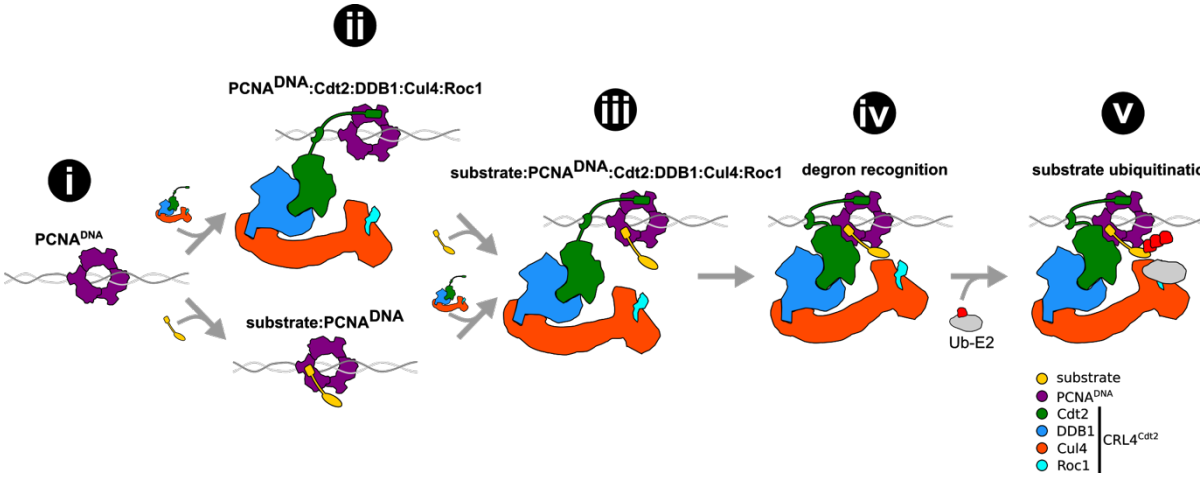



**Figure A13.2: Mechanism of ubiquitination and the central role in genome integrity.**
*During S phase or after DNA damage,* ***(i)*** *PCNA is loaded onto the DNA.* ***(ii)*** *$CRL4^{Cdt2}$ and the substrates bind independently to different subunits of PCNA via their PIP boxes. Cdt2's DNA-binding domain further stabilizes the complex at the DNA.* ***(iii)*** *Once the substrate and the Cdt2 are bound,* ***(iv)*** *Cdt2 recognizes the PIP degron via its WD40 domain and* ***(v)*** *initiates ubiquitination of the substrate.*

Several important regulators of DNA replication have been identified as substrates of Cdt2 in recent years (Table A13.1). The best studied ones are the replication licensing factor Cdt1, the CDK inhibitor p21, and the histone methyltransferase Set8 (388, 396). Additionally, other substrates of $CRL4^{Cdt2}$ are involved in key roles during the cell cycle, DNA replication and DNA repair. Via its regulation of a multitude of factors, Cdt2 has been broadly associated with the regulation of genome stability. Linked to that, Cdt2 was shown to be mis-regulated in multiple tumors (299, 397), identified as a target of oncogenic viruses (398), and linked to Alzheimer's disease (399). Taken together, this wide role of ever-increasing substrates position $CRL4^{Cdt2}$ as an important regulator of endogenous and exogenous threats.

Numerous Cdt2 phosphorylation sites have been implicated in the regulation of Cdt2 recruitment to and activity at $PCNA^{DNA}$. Human Cdt2 possesses 22 putative cyclin-dependent kinase (CDK) phosphorylation sites, 18 of them being located in the C-terminal part (Figure A13.1 d) (298, 391). Binding of Cdt2 to $PCNA^{DNA}$ is regulated by the periodic oscillations of CDK activity and their downstream targets during the cell cycle (298). Cdt2 phosphorylation levels mirror mitotic CDK activity and impact binding to DNA and $PCNA^{DNA}$ (391, 400). However, an exact relationship between individual sites and regulation of Cdt2 activity on the molecular level is still missing.

**Table A13.1: List of Cdt2 substrates and their functional association.**

| substrate | functional involvement | reference |
|---|---|---|
| Cdc6 | licensing factor, interacting with Cdt1 | (401) |
| Cdt1 | licensing factor for replication | (293, 388) |
| CHK1 | effector of ATR kinase during cell cycle and replication | (402) |
| CRY1 | circadian clock | (403) |
| E2F1 | transcription factor, involved in cell cycle regulation | (404) |
| FBH1 | helicase with anti-recombinogenic activity | (405) |
| HBV viral DNA polymerase | viral DNA polymerase | (406) |
| MDM2 | negative regulator of p53 | (407) |
| MGMT | DNA methyltransferase | (408) |
| p12/POLD4 | DNA polymerase subunit | (409, 410) |
| p21/CDKN1A | central regulator of cell cycle progression | (411)(412) |
| p53 | tumor suppressor | (407) |
| PDCD4 | tumor suppressor | (413) |
| SDE2 | involved in DNA replication and cell cycle control inhibiting translesion DNA synthesis | (414) |
| Set8 | histone methyltransferase important for replication licensing and double strand breaks | (415, 416) |
| TDG | participating in base-excision repair | (417, 418) |
| TLS DNA polymerase η | translesion DNA synthesis during DNA repair | (419) |
| XPG | 3‘ DNA endonuclease and recruited gap-filling DNA synthesis | (420) |

## 13.2 Aims of this study

This study aims at recombinant expression and purification of the human proteins PCNA, Cdt2:DDB1, Cdt1, and p21. The enzymatic activity of Cdt2 towards its substrate Cdt1 will be assessed using an *in vitro* ubiquitination assay. Insights into the phosphorylation status of Cdt2 will be gained. Several protein complexes will be assembled, and their stability will be assessed via analytical gel filtration chromatography (GF) analysis. In the end, a structural investigation of the stably assembled complexes will be launched using negative stain electron microscopy (EM) as well as cryo-EM combined with single particle analysis (SPA).

## 13.3 Methods

### 13.3.1 Molecular cloning

In order to recombinantly express all of the individual components, the following human cDNA clones were ordered: *PCNA* (GenBank ID: BC000491.2), *Cdt1* (BC009410.1), *Cdt2* (BC033297.1). Forward and backward primers were designed and ordered (Invitrogen, USA) (Table A13.2). Additionally, the gene for *p21* (541 bp) including forward and reverse primers as well as three oligonucleotides to create a blunt (30/30 nt) and sticky-end (30/20 nt) dsDNA were synthesized (Invitrogen, USA).

**Table A13.2: Overview of primers, vectors, restriction enzymes, and oligonucleotides.**

| gen name, direction | vector | primer sequence | restriction enzyme |
|---|---|---|---|
| PCNA, forward | pHisSUMO | ggc ccc ggg atg ttc gag gcg cgc ctg gtc | XmaI |
| PCNA, reverse | pHisSUMO | ggc gcg gcc gcc tac taa gat cct tct tca tcc tcg atc | NotI |
| p21, forward | pRSF Duet (insert 1) | ggc gga tcc gtt aga agt tct gtt tca agg t | BamHI |
| p21, reverse | pRSF Duet (insert 1) | gcc ctc gag tta tta cgg ttt acg ttt gct ga | XhoI |
| Cdt2, forward | pTriEx6 | ggc ccc ggg atg ctc ttc aat tcg gtg ctc cg | XmaI |
| Cdt2, reverse | pTriEx6 | ggc gcg gcc gct aat tct gtt gag tgt tca gga cc | NotI |
| Cdt1, forward | pTriEx6 | ggc ggt acc aga tgg agc agc gcc gcg tca cc | KpnI |
| Cdt1, reverse | pTriEx6 | ggc gcg gcc gcc agc ccc tcc tca gca cgt g | NotI |
| 30 nt, forward | | ata cga tgg gat acg atg gga tac gat ggg | |
| 30 nt, reverse | | ccc atc gta tcc cat cgt atc cca tcg tat | |
| 20 nt, reverse | | ccc atc gta tcc cat cgt at | |

Several constructs were created by standard molecular cloning techniques including PCR amplification, restriction digest, gel purification, ligation, transformation into DH5α *E. coli* cells, extraction of plasmids (QIAGEN, Germany), preparation of cryo-stocks for long term storage, and test digestion. For the recombinant expression in *E. coli* Rosetta cells, the pHis-SUMO and the pRSFDuet vectors were used containing *hs*PCNA and *hs*p21 respectively. Both proteins possessed a N-terminal His-tag followed by a 3C protease cleavage site. For the production of baculovirus and protein expression in Sf9 insect cells, the vector pTriEx6 was used to assemble pTriEx6-Cdt1 and pTriEx6-Cdt2, both constructs containing 3C-cleavage site followed by a C-terminal His-tag. All constructs were verified by test digestion, agarose gel visualization and sequencing (LGC genomics, Berlin).

### 13.3.2 Buffer and solutions

The following buffers were prepared fresh and stored at 4°C (Table A13.3):

**Table A13.3: List of used buffers.**

| lysis buffer | washing buffer |
|---|---|
| 50 mM HEPES | 50 mM HEPES |
| 500 mM NaCl | 500 mM NaCl |
| 4 mM MgCl2 | 4 mM MgCl2 |
| 0.5 mM TCEP | 0.5 mM TCEP |
| 9 protease inhibitor tablet / 50 ml | 30 mM imidazole HCl pH 7.8 |
| 30 mM imidazole HCl pH 7.8 | filtered with 0.22 µm cut-off membrane |

| elution buffer | GF buffer |
|---|---|
| 50 mM HEPES | 10 mM HEPES |
| 500 mM NaCl | 500 mM NaCl |
| 4 mM MgCl2 | 4 mM MgCl2 |
| 0.5 mM TCEP | 0.5 mM TCEP |
| 300 mM imidazole HCl pH 7.8 | filtered with 0.22 µm cut-off membrane |
| filtered with 0.22 µm cut-off membrane | |

The following purchased media were used to grow E. coli and Sf9 insect cells, respectively: lysogeny broth (LB) medium (10 g/l tryptone, 5 g/l yeast extract, and 5 g/l NaCl) and Insect Xpress serum free (Lonza, Switzerland).

### 13.3.3 Protein expression and purification

**Protein expression in *E. coli* Rosetta cells** was perfromed in the following way: (i) Pre-culture of 25 ml LB medium including antibiotics was inoculated with transformed *E. colis* Rosetta 2(DE3) from cryo-stocks and grown over-night at 37°C. (ii) On the next day, 2 l LB medium containing appropriate antibiotics was inoculated with 20 ml pre-culture. (iii) Cells were grown at 37°C and 150 rpm. (iv) Upon reaching an $E_{600}$ of ~0.7 OD, induction with 200 µM IPTG was performed and (iv) the temperature was reduced to 18°C for 20 hrs protein production. (v) On the next morning, cells were harvested at 3,300 rpm for 15 min using a JLA9.1000 centrifuge rotor (Beckman) and (vi) the equivalent of 1 l culture was resuspended in 35 ml lysis buffer for storage at -20°C.

**Baculovirus production in Sf9 insect cells** consisted of three steps described previously (421): (i) P0 virus fraction was produced by co-transfecting linearized bacmid (baculovirus genome) and the transfer plasmid pTriEx6 containing the protein of interest (pTriEx6-POI) into 2.5 ml Sf9 cells at a density of $0.8e^6$ cells/ml. Sf9 cells were contained in 6-well plates at 28°C for 96 hours. Upon successful infection, the supernatant was removed and stored as recombinant P0 baculovirus (*Autographa californica nucleopolyhedrovirus* clone C6) at 4°C. (ii) Amplification of P0 to P1 (P0>P1) was achieved by infection of 50 ml Sf9 cells at $2.0e^6$ cells/ml with 1 ml P0 baculovirus. Cells were shaking for 72 to 96 hours at 28°C and, once the infection was verified via growth arrest, the supernatant was harvested and stored as

P1 at 4 °C. (iii) Amplification P1>P2 was achieved analogous to the previous step, despite using only 10 µl P1 for the infection of 50 ml Sf9 cells at 2.0e$^6$ cells/ml. Harvested P2 baculovirus fraction was stored in the dark at 4°C and subsequently used for protein production. Sf9 cells were cultured using Insect-EXPRESS medium (Lonza) in an Innova 42R incubator shaker (New Brunswick) at 200 rpm.

**Protein expression in Sf9 insect cells** was achieved using the P2 for small- or large-scale expression in 100 ml or 1000 ml at 3.0e$^6$ cells/ml. (i) A ratio of 1:250 P2:volume was used. (ii) Next Sf9 cells are shaking for 72 hours at 28°C and 200 rpm, and (iii) harvested at 1,000 g for 30 min. (v) The equivalent of 1000 ml medium was resuspended with at least 100 ml lysis buffer including 10 % glycerol and stored at -20°C.

**Protein purification with Ni-NTA and SEC** was achieved in several subsequent steps executed on ice or at 4°C: (i) Cell lysis, (ii) debris removal, (iii) Ni-NTA affinity chromatography, (iv) protease cleavage, (v) concentration, (vi) size exclusion chromatography, (vii) analysis via SDS-PAGE, and (viii) storage at -80°C.

In the following, all seven steps will be described in more detail: (i) Cell lysis was achieved via mechanical disruption using a M-110L micro-fluidizer (Microfluidics, United Kingdom). (ii) Cellular debris was pelleted down using a JA 25.50 rotor (Beckman Coulter, USA) at 48,000 g and 4°C for 45 min. (iii) Supernatant was carefully extracted and (iv) furtheron purified via automatic Ni-NTA affinity chromatography as part of a ÄKTA pure system (cytiva, United Kingdom). At this step, the sample was loaded to an equilibrated 5 ml HisTrap column (cytiva, United Kingdom), washed with a total of 250 ml washing buffer containing 20 mM imidazole, from which 50 ml included 5 mM ATP to remove chaperones. Elution into 10 fractions was achieved by 300 mM imidazole. The protein content in each fraction was assessed by SDS-PAGE and the highest quality fractions were pooled and (v) concentrated to at least 5 ml using a centrifugal filter device (Vivaspin). (vi) Protein sample was loaded to a size exclusion chromatography (SEC) Superdex 200 16/600 column (cytiva, United Kingdom) as part of a ÄKTA prime system (cytiva, United Kingdom). The column was pre-equilibrated with GF buffer. Protein sample was eluted from the column using around 150 ml GF buffer into 51 fractions at a flow rate of 1 ml/min. Chromatogram of the $UV_{280}$ signal was analyzed via Unicorn 7.0 software (cytiva, United Kingdom) and (vii) promising fractions were analyzed via SDS-PAGE and NanoDrop spectrophotometer (ND 1000, Peqlab). Protein concentrations were quantified using theoretical absorption coefficients based on the amino acid sequence calculated by ProtParam on the ExPASy webserver (422). Finally, (viii) appropriate fractions were pooled, concentrated, measured, snap-frozen in liquid nitrogen, and stored at -80°C. If necessary, the cleavage of affinity tags was achieved by overnight-incubation with 100 µg GST-3C protase per mg total protein prior to step (vi).

### 13.3.4 Annealing of dsDNA

Equimolar ratios of two complementary 30 nt strands were mixed, heated to 95°C for 2 min and annealed during the cool-down over 2 hours.

### 13.3.5 Analytical gel filtration

For the formation of protein complexes, individual components were mixed in specific µM ratios and incubated together for at least 15 minutes on ice. For analytical GF, 500 µl with a total concentration of 5 µM protein complex were loaded to the analytical Superdex 200 10/300 GL as part of the ÄKTA pure system. 1 ml fractions were collected at a flow rate of 0.5 ml/min and analyzed via SDS-PAGE. Promising fractions were pooled, concentrated, measured, snap-frozen in liquid nitrogen, and stored at -80°C. All work was performed at 4°C.

### 13.3.6 *In vitro* ubiquitination assay

For the *in vitro* ubiquitination assay (ivUb) the following reagents (concentration) were used: ubiquitin (15 µM), $MgCl_2$ (2.5 mM), ATP (2.5 mM), UBCH5C E2 (0.25 µM), Cul4A:NEDD8:Roc1 (0.125 µM), Cdt2:DDB1 (0.125 µM), Cdt1 (0.5 µM), PCNA (0.125 µM), and 30 nt dsDNA (0.125 µM). All components were mixed forming a 160 µl reaction using the ivUb buffer (20µmM HEPES at pH 7.0, 150 mM NaCl, 3 mM MgCl2). The first time point (0 min) was taken and E1 (0.05 µM) was added to start the ubiquitination time course reaction. Afterwards, the reaction mixture was placed immediately at 37°C. At each time point, 30 µl sample were taken, the reaction was stopped immediately by the addition of 10 µl 4xSDS buffer, and incubation at 95°C for 5 min. Samples of each time point were loaded to an SDS-PAGE for visualization.

### 13.3.7 Negative stain EM sample preparation, image acquisition

Grids covered with a home-made carbon support film were glow-discharged for 30 s and 3.5 µl 0.04 µM protein complex was applied. Upon incubation for 30 s, the solution was blotted away, and 3.5 µl 2 % uranyl acetate negative staining solution was applied for 45 s. Afterwards the staining solution was blotted away, and the grid was dried at room temperature. Image acquisition was performed at 100 kV using a CM100 electron microscope (Philips, Netherlands) equipped with a 1kx1k Fastscan CCD camera (TVIPS, Germany).

### 13.3.8 Cryo-EM sample preparation, image acquisition, and image processing

The protein complex was vitrified using a VitroBot MKII operated at 80 % humidity and 4°C. 3.5 µl protein complex at 0.4 µM were applied to a R1.2/1.3 grid (Quantifoil, Germany), incubated for 30 s, plotted, and vitrified in liquid ethane. 4 grids were prepared and stored in liquid nitrogen until usage. Data acquisition was performed on a 120 kV Tecnai Spirit (FEI, USA) equipped with a 4kx4k F416 CMOS detector (TVIPS) and a Gatan 626 cryo holder. Automatic data acquisition (14) resulted in a data set of 490 micrographs collected at a pixel

size of 2.65 Å/px. All micrographs were inspected visually, and suboptimal ones were discarded, leaving 244 micrographs. Contrast transfer function parameters were estimated using CTFFind4 (25) within cisTEM (31). Particle positions were determined via Gaussian picking, extracting 186,074 particle images at a pixel size of 2.65 Å/px. In a first tier of 2D classification, particle images were classified into 75 classes. Classes not representing suitable 2D projections of a protein complex or depicting signs of edges or artifacts were discarded. The remaining 145,964 particle images were classified in a second tier of 2D classification into 300 classes to explore the large heterogeneity of the data set.

Data acquisition on a high-end cryo-EM was performed using a Polara F30 cryo-EM operated at 300 kV and equipped with a K2 DED and a FEG as electron source.

## 13.4 Results

### 13.4.1 Expression and purification of the individual components

Ubiquitination of the Cdt2 substrates requires the formation of a $CRL4^{Cdt2}$ complex comprising Cdt2:DDB1, $PCNA^{DNA}$ and a substrate like p21 or Cdt1. In order to reconstitute the complex, the individual components PCNA, Cdt2:DDB1, Cdt1, and p21 were recombinantly expressed and purified.

**PCNA**

Human PCNA (*hs*PCNA) is a homotrimer (~99 kDa) that encircles the DNA (Figure A13.3 a). pHis-Sumo-PCNA was cloned, expressed in *E. coli* Rosetta cells, purified via Ni-NTA and GF chromatography, treated with 3C protease (3C) to remove the His-SUMO affinity tag, and further purified via GF to separate PCNA from the affinity tag (grey andf purple trace in Figure A13.3 b). Upon treatment with 3C protease, a size shift from ~40 kDa to 33 kDa (compare "-3C" with "+3C" inFigure A13.3 e) and HisSUMO at ~15kDa appeared (fractions #33-35 Figure A13.3 b and lane 34 in c), indicating the successful cleavage of pHis-SUMO-PCNA. The corresponding GF-fractions containing PCNA were pooled and stored at -80°C. Taken together, large quantities of homogeneous PCNA were produced at sufficient yield, suitable for subsequent biochemical and structural analysis.

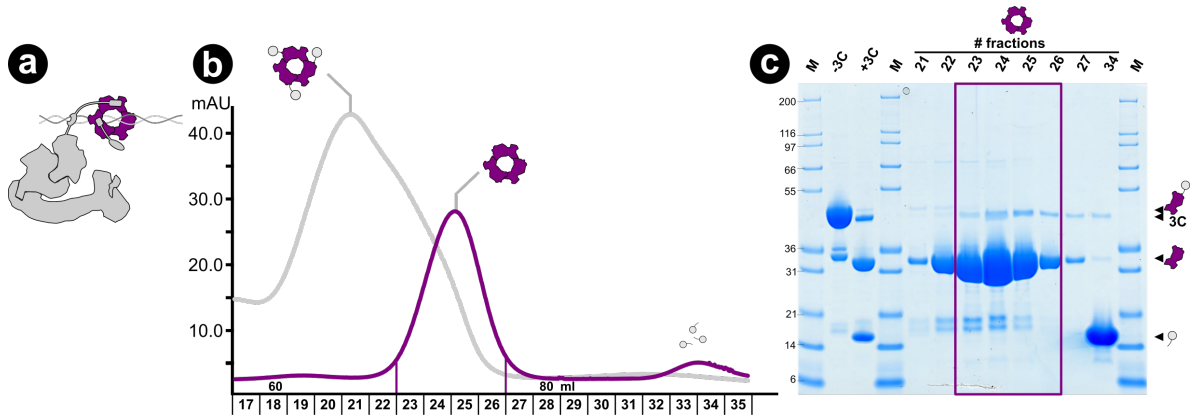



**Figure A13.3: Purification of PCNA via Ni-NTA and GF.**
***(a)*** *Schematic of PCNA and its role in the* $CRL4^{Cdt2}$ *complex.* ***(b)*** *Overlay of chromatograms for GF before (grey) and after (purple) 3C-treatment.* ***(c)*** *SDS-PAGE of the relevant fractions of GF run after overnight 3C-treatment.*

**Cdt2:DDB1**

The substrate receptor Cdt2 was expressed and purified in a heterodimeric complex with the adaptor protein DDB1 (Figure A13.4 a). To this end, Sf9 insect cells were co-infected with recombinant DDB1 and Cdt2 baculoviruses in a ratio of 1:500 and 1:166 per 1000 ml cells, respectively. Purification was achieved via Ni-NTA chromatography (Figure A13.4 b) and subsequent GF over a Superdex 200 (Figure A13.4 c and d). A comparison of the elution profiles of Cdt2:DDB1 and DDB1 alone exhibits a peak shift from fraction #17 to fraction #16

for the Cdt2:DDB1 sample (Figure A13.4 c). SDS-PAGE analysis confirmed the presence of Cdt2 (79 kDa) in that peak with a band running between the 97 and 66 kDa marker lanes. The weaker intensity of the Cdt2-band could stem from its inadequate staining or substoichiometric ratios compared to DDB1. Nevertheless, fractions #13-18 were pooled, concentrated, snap frozen in liquid nitrogen, and stored at -80°C. Taken together, Cdt2:DDB1 could be prepared in sufficient purity for subsequent use.

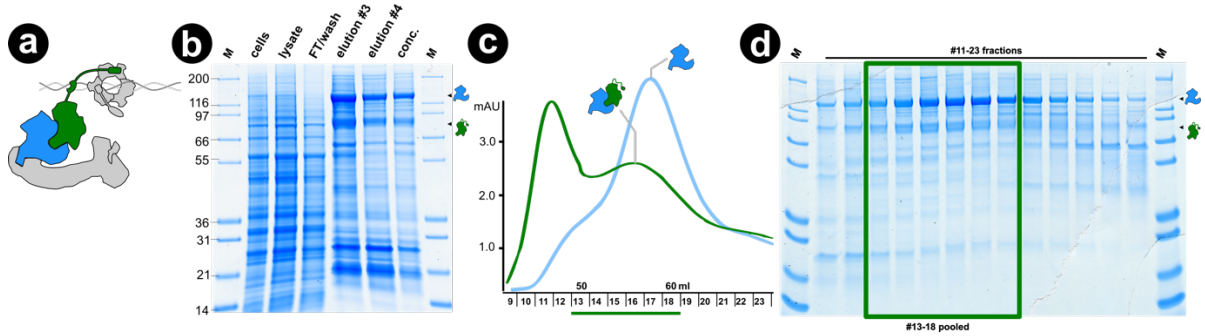



**Figure A13.4: Cdt2:DDB1 purification via Ni-NTA and GF.**
***(a)*** *Schematic of Cdt2:DDB1 and its role in the CRL4$^{Cdt2}$ complex.* ***(b)*** *SDS-PAGE of the steps from lysis to Ni-NTA chromatography.* ***(b)*** *Overlay of GF chromatograms of Cdt2:DDB1 (green) and DDB1 alone (blue).* ***(d)*** *SDS-PAGE of the GF fractions. Pooled fractions are indicated by a green frame and the expected positions of the bands are marked.*

### Cdt1

The substrate Cdt1 (Figure A13.5 a) was expressed in Sf9 insect cells using a baculovirus infection ratio of 1:250 in 1000 ml cells. Purification was achieved via Ni-NTA chromatography (Figure A13.5 c) and subsequent GF over a Superdex 200 (Figure A13.5 b and c). A large peak in the column void volume was detected (fractions #9-12). The fractions #17-20 contained Cdt1 (~60 kDa) as confirmed by SDS-PAGE analysis. Subsequently, fractions #17-20 were pooled, concentrated, snap frozen in liquid nitrogen, and stored at -80°C. The two bands (arrows in Figure A13.5 b) were identified as HSP70 (~70 kDa) and Cdt1 (60 kDa) via mass spectrometry. Taken together, a small amount of impure Cdt1 was purified from Sf9 insect cells.

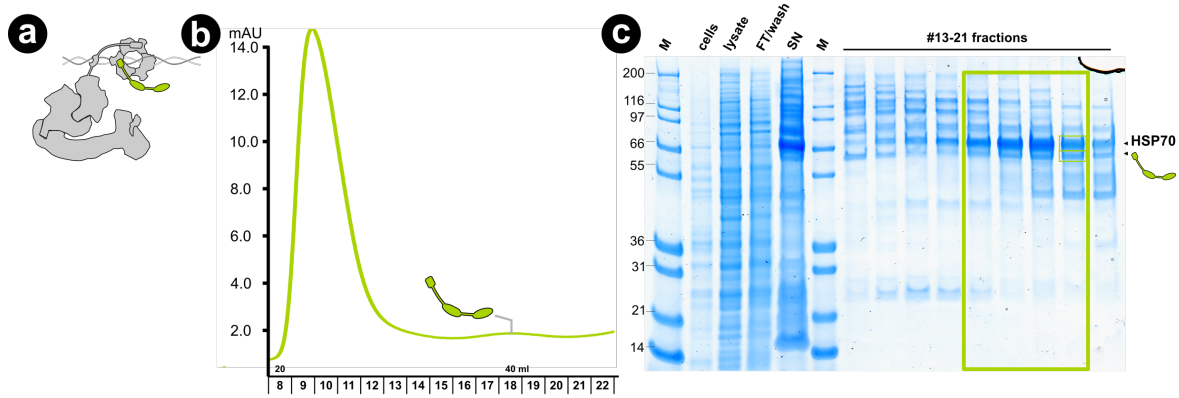



**Figure A13.5: Cdt1 purification via Ni-NTA and GF.**
***(a)*** *Schematic of the substrate Cdt1 and its role in the CRL4$^{Cdt2}$ complex.* ***(b)*** *GF chromatogram for Cdt1.* ***(c)*** *SDS-PAGE of the steps during Ni-NTA chromatography and the relevant fractions after GF. Pooled fractions and contaminations identified by mass spectrometry are indicated.*

**p21**

Human kinase inhibitor p21, another substrate of the CRL4$^{Cdt2}$ (Figure A13.6 a), was expressed in *E. coli* Rosetta cells, purified via Ni-NTA chromatography (Figure A13.6 b), treated with 3C protease and cleaned via GF. A large peak in fractions #25-31 was detected and confirmed to contain p21 together with a contamination at ~25 kDa (Figure A13.6 b and d). MS identified both bands as the GTP cyclohydrolase from *E. coli* and human p21, respectively. The corresponding fractions were pooled and stored at -80°C. Taken together, large quantities of contaminanted p21 were produced for downstream use in biochemical and structural analysis.

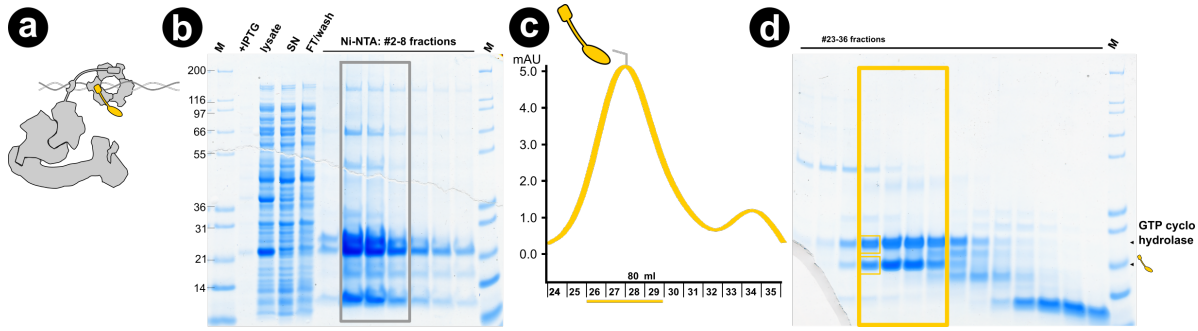



**Figure A13.6: p21 purification via Ni-NTA and GF.**
***(a)*** *Schematic of the substrate p21 and its role in the CRL4$^{Cdt2}$ complex.* ***(b)*** *SDS-PAGE of the steps during Ni-NTA chromatography.* ***(c)*** *GF chromatogram for p21.* ***(d)*** *SDS-PAGE analysis of the fractions after GF. Pooled fractions and contaminations identified by mass spectrometry are indicated.*

### 13.4.2 Functional analysis via *in vitro* ubiquitination assay

In order to prove ubiquitination activity of the purified Cdt2:DDB1 heterodimer, an *in vitro* ubiquitination assay (ivUb) was performed using Cdt1 as the substrate. A loading control for all individual components was performed, exhibiting contaminations in Cdt1, Cdt2:DDB1, and E1 preparations (asterisk in Figure A13.7 a). In the Cdt1-sample, a contamination by HSP70 was identified previously at ~70 kDa via mass spectrometry. Cdt2:DDB1 additionally contained unidentified bands at ~40 kDa and ~58 kDa. Despite these impurities, the preparations were sufficiently pure for their downstream use in ivUb.

For ivUB, all protein components (Ub, E1, E2, Cul4-N8, PCNA, Cdt2:DDB1, and Cdt1) and an aligned 30 nt dsDNA were mixed together and placed at 37°C. With the limited amount of purified components, only two conditions were tested “all” and “-PCNA” in order to investigate the influence of PCNA on the ubiquitination activity of Cdt2.

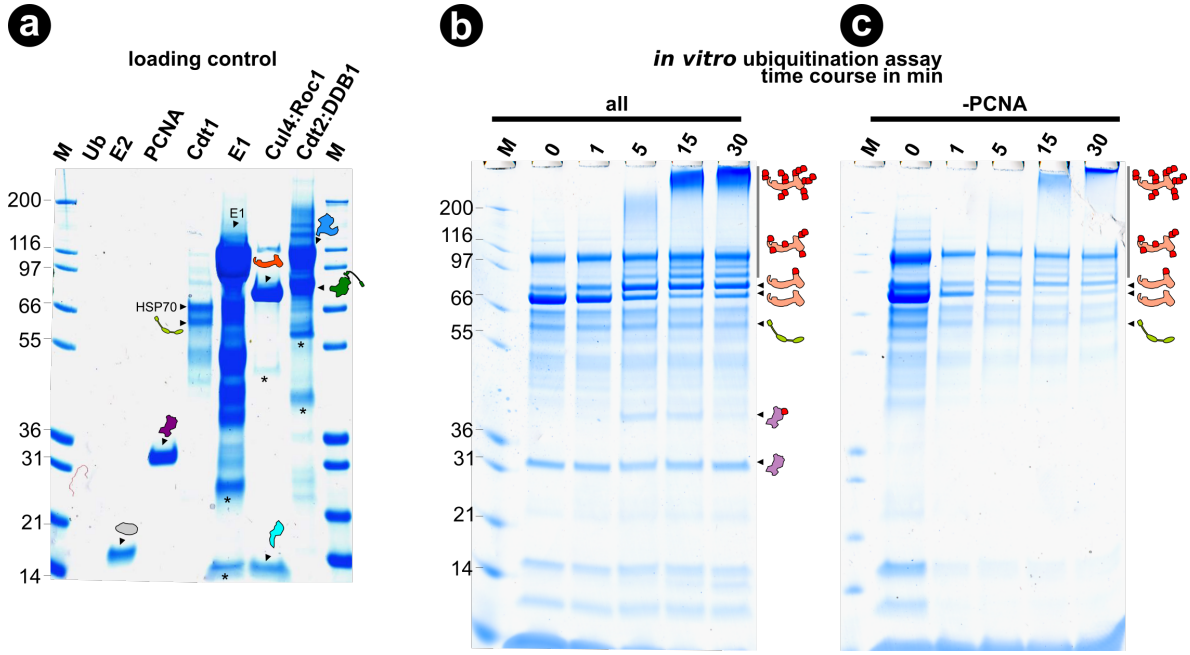



**Figure A13.7: *In vitro* ubiquitination assay.**
***(a)*** *Loading control of all components of the in vitro ubiquitination assay (ivUb). Time course of the ivUb for all components* ***(b)*** *and without PCNA* ***(c)****.*

Ubiquitinated proteins are discernable as characteristic ladders with an increasing number of Ub’s attached. For the “all” condition, a strong ubiquitination activity was detected by the formation of large ubiquitinated species after 5 min (Figure A13.7 b). However, no substantial decrease in the intensity of the Cdt1 band (60 kDa) was noticed, suggesting the ubiquitination of another protein component within the reaction mixture. The large ubiquitinated species might be attributed to auto-ubiquitination of CUL4 and/or DDB1. This auto-ubiquitination was also detected in the absence of PCNA (Figure A13.7 c), suggesting it to be a PCNA-independent mechanism. Additionally, a species of mono-ubiquitinated PCNA (~38 kDa) was

also detected (Figure A13.7 b) supporting previous results of Cdt2 activity towards its binding partner (423).

### 13.4.3 Phosphorylation of Cdt1 and Cdt2

Phosphorylation has previously been shown to regulate Cdt1 and Cdt2 activity (391, 400, 424). Both proteins possess a long, intrinsically disordered region (IDR) with an isoelectric point around 10: Cdt1 at its NTD and Cdt2 at its CTD (Figure A13.8 a and b).

In a first test, the phosphorylation status of Cdt2 was assessed via treatment with λPP, a phosphatase with activity towards phosphorylated serine, threonine, and tyrosine residues. Treatment with λPP (+) did induce a slight but significant shift of the Cdt2 band relative to the untreated condition (-) (Figure A13.8 c). Therefore, the phosphorylation of Cdt2 during recombinant expression inf Sf9 cells can be concluded. However, a more thorough analysis will be necessary in the future, using for example western blotting, autoradiograms, and/or mass spectrometry.

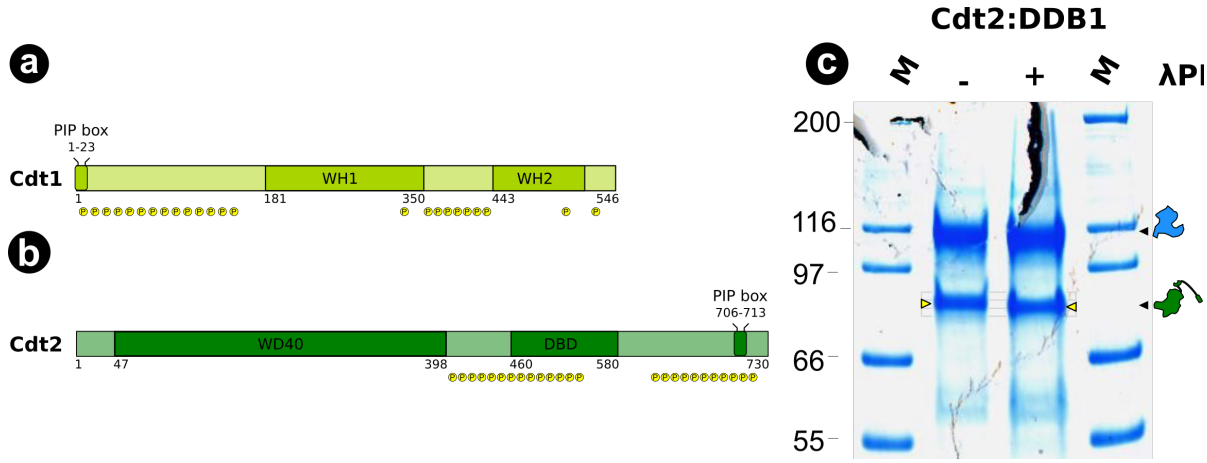



**Figure A13.8: Treatment of Cdt2:DDB1 with λPP.**
*Architecture of Cdt1* ***(a)*** *and Cdt2* ***(b)*** *with domains and phospho-sites labeled. Light colored areas denote predicted disordered regions.* ***(c)*** *SDS-PAGE analysis of λPP untreated (-) and treated (+)Cdt2:DDB1. Yellow triangles and lines are provide as guidance for the eye. Bands for DDB1 and Cdt2 are indicated on the right, while the protein marker bands are labeled on the left.*

### 13.4.4 Formation of $PCNA^{DNA}$:Cdt2:DDB1 complexes

Having all the individual components available allowed the assembly of a Cdt1:$PCNA^{DNA}$:Cdt2:DDB1 complex and its analysis via analytical GF. For the first assembly, equimolar ratios of all four components, PCNA, Cdt1, Cdt2:DDB1, and an 5-fold excess of 30 nt dsDNA were used. After incubation on ice for 30 min, the assembly reaction was separated using a Superdex 200 analytical GF column as part of the ÄKTA pure system. Furthermore, control runs for PCNA, Cdt1, and Cdt2:DDB1 individually were performed (Figure A13.9 a). A shift of the elution volumes between the Cdt2:DDB1-peak at fraction #12 (11 ml) and the assembly-peak at #11(10.5 ml), suggested the formation of a complex. SDS-PAGE of the corresponding fractions supported this assumption, and clearly showed enrichment of

PCNA in the assembly-peak (Figure A13.9 b). Furthermore, two additional peaks were present in the elution profile of the assembly reaction, at #15 (12.5 ml) and #17 (13.5 ml), respectively. These two peaks matched the elution profile of the individual Cdt1-sample in size and shape. The results of the SDS-PAGE analysis for #15-17 identified their content as HSP70 (70 kDa; fraction #15) and Cdt1 (60 kDa, fraction #17) (Figure A13.9 b and e). The PCNA homotrimer alone eluted at #16 and #17 (~13 ml), with an approximate size of 99 kDa (Figure A13.9 d). An additional co-eluting contamination ((1) at ~65 kDa in Figure A13.9 b and c) most probably originated from the Cdt2:DDB1 preparation.

Taken together, these findings elucidate the following: The assembly of a PCNA:Cdt2:DDB1 complex was successful as indicated by the size shift in the elution profile and appearance of a PCNA protein band in fractions #10-13. However, Ctd1 was not incorporated in the complex, indicated by the additional two peaks in the elution profile matching the profile of Cdt1 alone, and by the absence of a SDS-PAGE gel band corresponding to Cdt1 in the assembly-peak.

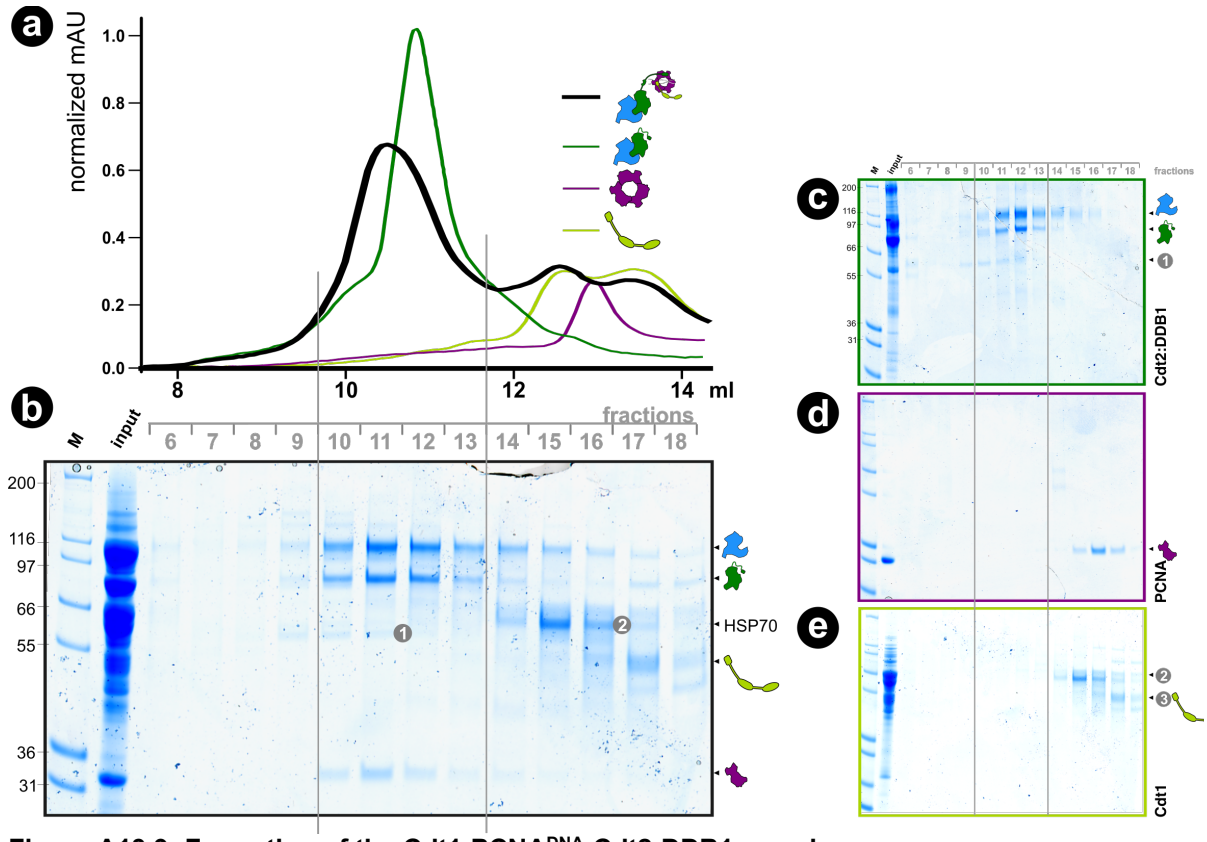



**Figure A13.9: Formation of the Cdt1:PCNA$^{DNA}$:Cdt2:DDB1 complex.**
***(a)*** *GF elution profile for the assembly of Cdt1:Cdt2:DDB1:PCNA$^{DNA}$. SDS-PAGE analysis of the fractions #6-18 for the assembly reaction* ***(b)*** *and the individual components Cdt2* ***(c)****, PCNA* ***(d)****, and Cdt1* ***(e)****. A contamination co-purified with Cdt2:DDB1 is labeled as (1) in (b) and (c). The band for HSP70 is labeled as (2) in (b) and (e).*

In a second attempt to assemble a Cdt2:DDB1-containing complex, a different Cdt2:DDB1 batch was used. Here, two complexes, Cdt1:PCNA$^{DNA}$:Cdt2:DDB1 (2-PIP, red trace) and PCNA$^{DNA}$:Cdt2:DDB1 (1-PIP, orange trace) were assembled using higher protein concentrations (Figure A13.10 a). Consistently, in both cases, the assembly yielded a complex at earlier elution volume than Cdt2:DDB1 alone. For both assemblies, the peak fractions #10 and #11 were pooled, concentrated, and analyzed via SDS-PAGE (Figure A13.10 c), indicating the presence of DDB1, Cdt2, PCNA, and an insignificant contamination (1) as seen before. In the 2-PIP assembly peak, no Cdt1 was detected, suggesting no incorporation into the complex. Indeed, the GF elution profile of the 2-PIP assembly exhibited a peak around fraction #16. Based on SDS-PAGE analysis, its identity was confirmed as Cdt1 (Figure A13.10 d). Taken together, also the second attempt to reconstitute a 2-PIP, including Cdt2 and the substrate Cdt1, was unsuccessful. However, the binding of Cdt2:DDB1 to PCNA (1-PIP assembly) was confirmed. Infact, both assemblies resulted in the formation if 1-PIP complexes.

In summary, the analytical GF experiments confirmed the formation of a PCNA:Cdt2:DDB1 complex from recombinantly expressed components. The formation was identified by a shift in the elution profile relative to Cdt2:DDB1 and the existence of the corresponding protein bands in the SDS-PAGE. Since the purity and concentration was sufficiently high for all three assembled complexes, their structural analysis via cryo-EM was be initiated.

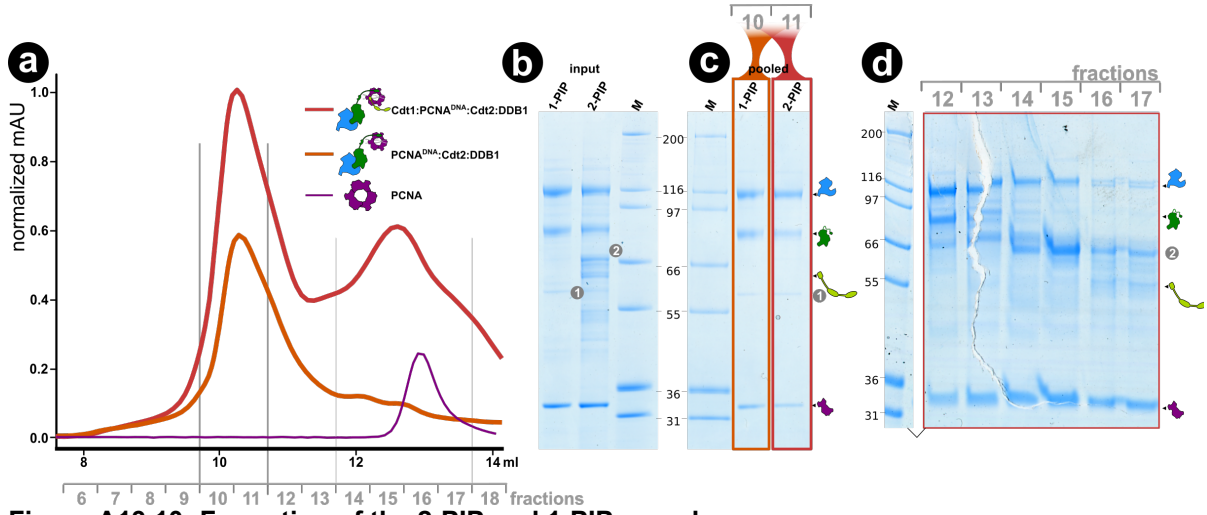



**Figure A13.10: Formation of the 2-PIP and 1-PIP complexes.**
***(a)*** *Elution profile for the assemblies Cdt1:PCNA$^{DNA}$:Cdt2:DDB1 (2-PIP) and PCNA$^{DNA}$:Cdt2:DDB1 (1-PIP) and PCNA are depicted. SDS-PAGE analysis of the* ***(b)*** *input and* ***(c)*** *peak fractions #10 and #11 for both reactions. Peak fractions were pooled, respectively. The bands for DDB1, Cdt2, Cdt1, PCNA, and two contaminations ((1) and (2)) are indicated.* ***(d)*** *SDS-PAGE analysis of the fractions #14-17 for the 2-PIP complex. Marker lane was part of the same gel and the intermediate lanes were removed for clarity. The HSP70 contamination (2) is marked as in (b). A contamination co-eluted with Cdt2:DDB1 is labeled as (1) in (b) and (c).*

### 13.4.5 Structural investigation via negative stain

Structural analysis of a complex via negative stain EM allows for a fast assessment of the sample quality. Here, the PCNA:Cdt2:DDB1 complex preparation with the highest yield (red trace in Figure A13.10 a) was used to prepare grids using negative staining methodology and imaged using a low-end screening EM. A homogeneous sample in size and distribution was obtained (Figure A13.11 b). However, based on the model where Cdt2:DDB1 interacts with PCNA via its PIP box located within the disordered CTD, a flexible arrangement as a 2-body particle is possible (Figure A13.11 a). In a visual analysis, three species can be differentiated (red, green, and purple arrows inFigure A13.11 b and c): (i) A round, ring-shaped particle without a hole, probably corresponding to PCNA (red). (ii) An elongated particle with several domains discernable, potentially representing Cdt2:DDB1 consisting of four WD40 domains (green). (iii) A large, round particle possibly presenting a p97 or proteasome particle possessing similar dimensions (purple). However, the precise interpretation of such individual particles embedded in negative staining solution are in general not reliable and can only be used as a general method to judge the homogeneity of a sample in terms of particle size and concentration.

Taken together, this negative stain EM analysis characterized the PCNA:Cdt2:DDB1 complex preparation as homogeneous enough to continue with cryo-EM analysis.

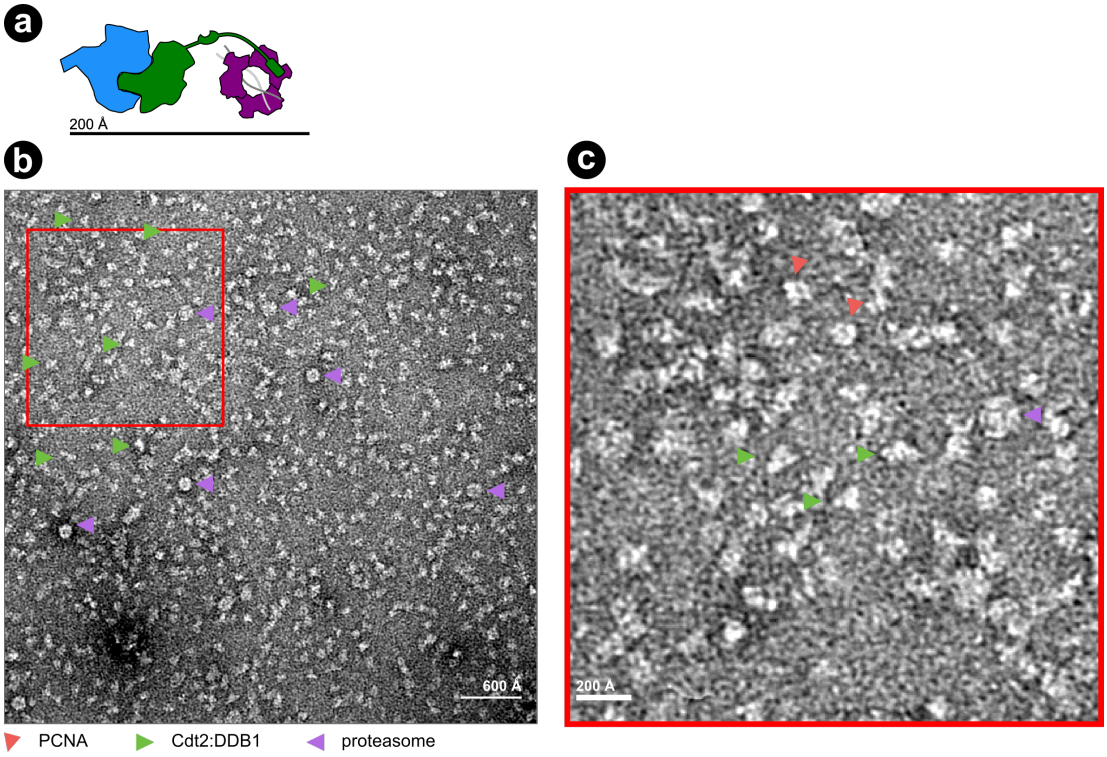



**Figure A13.11: Negative staining imaging of the PCNA$^{DNA}$:Cdt2:DDB1 complex.**
***(a)*** *Schematic representation of a PCNA$^{DNA}$:Cdt2:DDB1 complex with Cdt2:DDB1 and PCNA resembling a 2-body arrangement, connected by the Cdt2$^{CTD}$.* ***(b)*** *A representative micrograph of the PCNA$^{DNA}$:Cdt2:DDB1 complex embedded in negative stain and imaged with a low-end screening EM. Three species are marked with red, green, and purple arrows, respectively. A scale bar is indicated.* ***(c)*** *Detailed view of the boxed area in (b), allowing for a closer inspection of the three particle species.*

### 13.4.6 Structural investigation via cryo-EM

Cryo-EM analysis was initiated by vitrifying 0.4 µM $PCNA^{DNA}$:Cdt2:DDB1 complex on holey carbon grids without carbon support film to avoid preferential orientation. A data set of 490 micrographs was collected automatically (14) on a 120 kV Tecnai Spirit screening microscope. After visual inspection, only 244 micrographs were marked as of sufficient quality and used subsequently for SPA.

A homogeneous particle distribution was observed, consistent with the results obtained from negative staining. Due to the low contrast of the vitrified sample, no direct visual inspection of individual particles was performed.

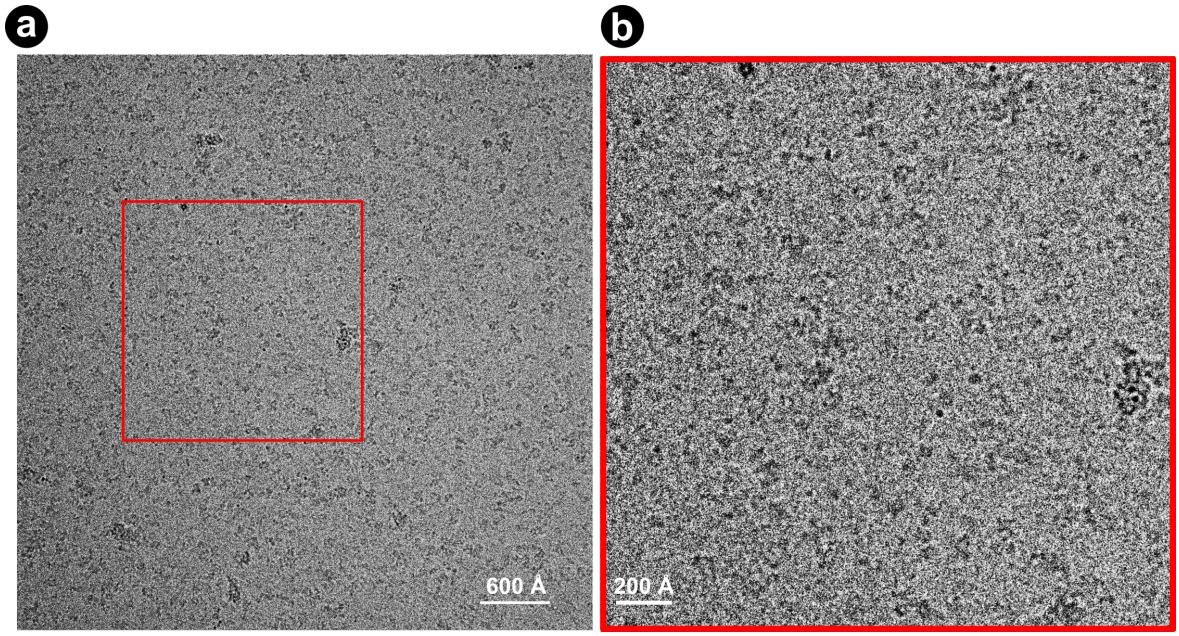



**Figure A13.12: Cryo-EM imaging of the 1-PIP complex.**
***(a)*** *A representative micrograph of the $PCNA^{DNA}$:Cdt2:DDB1 complex embedded in vitrified ice and imaged at 120 kV. A scale bar is indicated.* ***(b)*** *Detailed view of the boxed area in (a), allowing for a closer inspection.*

Instead, a statistical analysis by SPA was carried out using cisTEM (31). CTF parameters were estimated using CTFFind4 and 186074 particle positions were found using Gaussian blob picking. A first tier of 2D classification resulted in the selection of class averages representing suitable projections of protein complexes, containing 145964 particle images. Using this selection, a second tier of 2D classification was launched, further separating the data set into 300 classes to explore its heterogeneity (Figure A13.12 a).

A wide variety of different protein complexes, views and interactions between them were visualized. Upon visual inspection, 7 general groups were defined (Figure A13.13 b): (i) PCNA with stick, (ii) PCNA with loop (iii), Cdt2:DDB1 side view, (iv) and (v) DDB1, and (vi) unknown particle with 7-fold symmetry.

- The groups (i) and (ii) depicted a clear PCNA-presence with an additional density bound at the PCNA-periphery. Several class averages depicted either PCNA alone, PCNA with a stick or with a loop, respectively.

- The group (iii) potentially represent Cdt2:DDB1 in a side, with up to four WD40 domains observable. The ability to discern individual WD40 domains using the given imaging conditions has been shown for the DCAF1$^{CTD}$:DDB1 complex before (Banchenko, Krupp *et al.* under review).
- Groups (iv) and (v) might depict a single DDB1 particle, adopting a Y-shape with its three WD40 domains.
- Group (vi) could be interpreted as the full two-body particle, with PCNA and Cdt2:DDB1 as the two bodies forming the (1-PIP) complex. Due to the long and flexible connection between the two bodies, no optimal alignment could be obtained leading to spurious class averages.
- Group (vii) depicts a large, round 7-fold symmetric particle. It could correspond to the similar shaped particle in the negative stain images (purple arrow inFigure A13.11 b and c). Its 7-fold symmetry would rather hint at a co-purified proteasome (425) than a p97 particle (6-fold symmetry) (426).

In summary, the PCNA$^{DNA}$:Cdt2:DDB1 complex was successfully vitrified and exhibited consistently good behavior, i.e. a homogeneous particle size and distribution. A small data set was collected using a screening microscope and the images subsequently analyzed. 2D class averages suggested a high degree of conformational and compositional heterogeneity. This is expected from such a 2-body complex connected via the disordered Cdt2$^{CTD}$. The obtained class averages were separated into 7 groups by visual inspection, sampling several different views of the PCNA$^{DNA}$:Cdt2:DDB1 complex: PCNA alone (i) and with a bound protein at its periphery (ii) was clearly identified by its circular shape. Side (iii) and top views (iv) of Cdt2:DDB1 might show up to 4 domains, potentially representing the 4 expected WD40 domains. Additionally, a V-shaped particle (v) could be DDB1 alone while group (vi) may represent the full 2-body particle. Furthermore, a large disc-shaped particle with 7-fold symmetry (vii) could correspond to a co-purified proteasome. The high quality of the obtained class averages suggests the readiness of the PCNA$^{DNA}$:Cdt2:DDB1 complex sample for high resolution data collection using a direct electron detector.

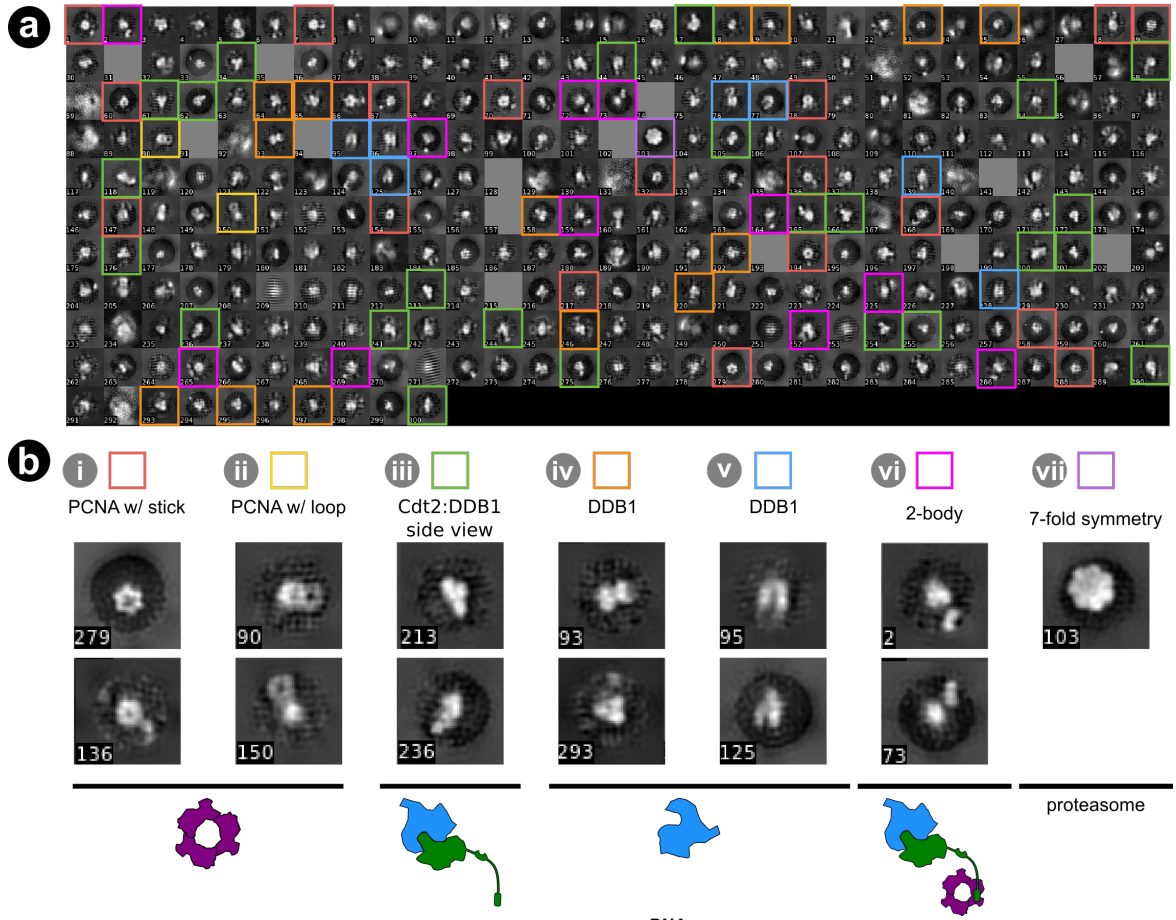



**Figure A13.13: Single particle analysis of the PCNA$^{DNA}$:Cdt2:DDB1 complex.**
***(a)*** *2D class averages after the 2$^{nd}$ tier of 2D classification into 300 classes. Class averages sorted into 7 different groups are labeled by colored boxes.* ***(b)*** *Depiction of the 7 groups with names and class averages as examples. Numbers in the lower left corner of each class average represent their original class number.*

Using the remaining grids from the vitrification batch, several attempts were made to find suitable regions for high-resolution data collection. Therefore, grids were imaged at 300 kV Polara F30 cryo-EM equipped with a Gatan K2 DED. However, thick and crystalline ice as well as dissociated and aggregated protein complexes within the holes hampered the collection of a data set (Figure A13.14). Such problems of irreproducible vitrification have been reported before and remain one of the main problems in cryo-EM (6, 10).

Taken together, despite its promising quality as proven by negative staining EM and a small cryo-EM data set, no high-resolution cryo-EM data set could be collected due to the surprisingly heterogeneous quality of the remaining grids. Therefore, the vitrification of a new batch of PCNA:Cdt2:DDB1 complexes has to be initiated in the future.

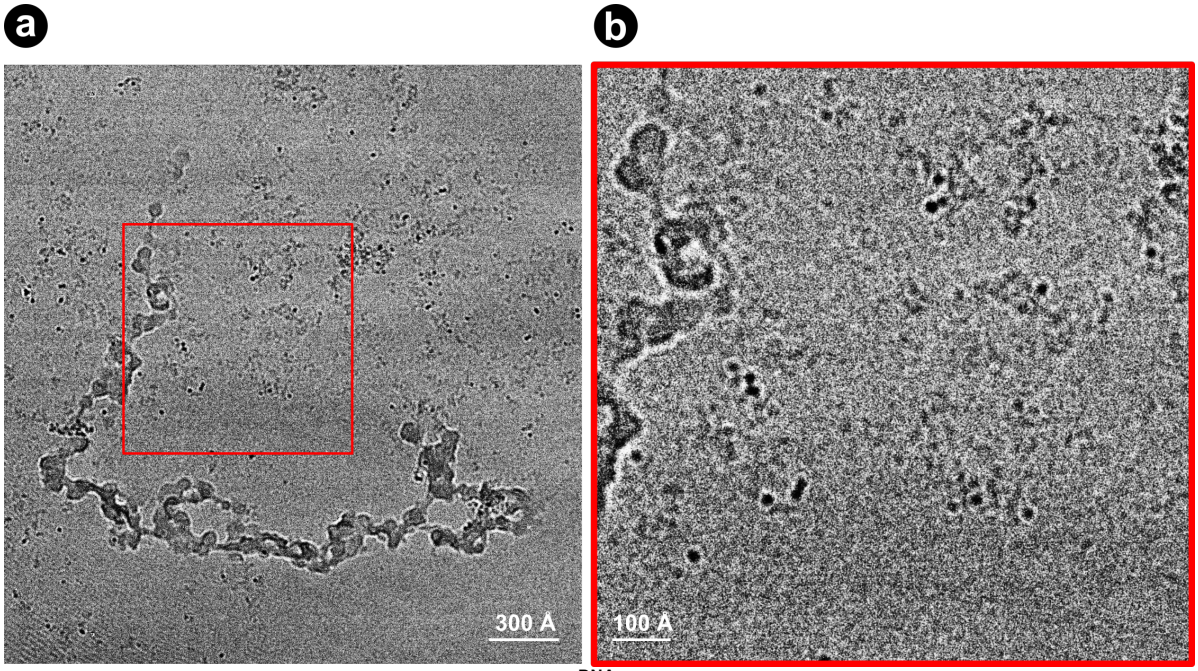



**Figure A13.14: Cryo-EM imaging of the $PCNA^{DNA}$:Cdt2:DDB1 complex using a high-resolution microscope.**
***(a)*** *A representative micrograph of the $PCNA^{DNA}$:Cdt2:DDB1 complex embedded in vitrified ice and imaged at 300 kV. Furthermore, a scale bar is indicated.* ***(b)*** *Detailed view of the boxed area in (a), allowing for a closer inspection.*

## 13.5 Discussion

Cdt2 represents a master regulator of genome-associated processes via its ability to bind $PCNA^{DNA}$ and ubiquitinate PIP-degron-containing proteins. The presented results depict the early phase of a project to uncover Cdt2's mechansims of ubiquitination.

### 13.5.1 Protein expression and purification

At first, plasmids for the expression for several proteins in either *E. coli* or Sf9 insect cells were designed and prepared employing standard molecular cloning techniques. Then, large-scale expression and purification of the human proteins PCNA and p21 were achieved. For both proteins, mg-quantities were purified, following previous successful expression protocols of PCNA (427) and p21 (428). For the latter, no solubilization of p21 from inclusion bodies was required. However, p21 eluted together with GTP cyclohydrolase in a stochiometric ratio. Accordingly, PCNA and p21 are available for biochemical and structural work in high quantities. Their expression in *E. coli* will allow a simple manipulation via site-directed mutagenesis to probe certain residues.

Sf9 insect cells and baculovirus were successfully used for the large-scale expression of the heterodimer Cdt2:DDB1. Cdt2:DDB1 was expressed by co-infection of Sf9 insect cells with baculoviruses for DDB1 and Cdt2, respectively. The purification via Ni-NTA and GF resulted in a small but homogeneous amount of protein, before the Cdt2-baculovirus lost its infectivity. Subsequent expressions using new baculovirus and optimization of the virus titer did not result in success. With this strategy, a well-established protocol for the co-expression of DCAF1:DDB1 in Sf9 insect cells was followed (289, 305). Other groups also reported success with a similar co-expression strategy (429). In the future, a co-expression as Cdt2:DDB1 with freshly made baculovirus for Cdt2 should be investigated. An additional strategy would be the co-expression as a Cdt2:DDA1:DDB1 heterotrimeric complex, since DDA1 improved the yield, stability, and enzymatic activity of a DCAF15:DDB1:DDA1-complex (295). For Cdt1 expression, Sf9 insect cells were used, yielding low levels of Cdt1 contaminated with HSP70. Others used different constructs, insect cell lines, growth conditions, or purification protocols (424, 430). The problem of aggregation could be tackled by using a Cdt1-His6-MBP fusion protein analogous to the methodology used by Parker *et al.* (424), where MBP helps in solubilizing the largely disordered Cdt1.

### 13.5.2 *In vitro* ubiquitination activity

*In vitro* ubiquitination activity (ivUb) towards Cdt1 using purified components did not result in the ubiquitination of Cdt1. The missing of Cdt1-ubiquitination could be associated with numerous reasons falling into two categories: biological or technical reasons. Biological reasons can be one or a combination of the following: the deactivation of Cdt2 by endogenous phosphorylation, shielding of Cdt1 by HSP70 impurities, the use of unspecific E2 proteins

towards Cdt1 (431), or the improper loading of PCNA onto the DNA. Technical reasons could be a low overall concentration of Cdt1, higher auto-ubiquitination activity under the given experimental conditions, or the slower reaction kinetics of Cdt1 ubiquitination versus Cul4/PCNA. Further ivUb assays using newly purified proteins, a different substrate, or other E2s will be necessary in the future to evaluate these results. Additionally, the amount of proteins loaded at the time points >1min for "-PCNA" were reduced compared to the "all" conditions. This further hampers the detection of the bands in the interpretation of the results. To summarize, in this first test, Cdt1-ubiquitination by Cdt2 was not detected due to several possible biological and/or technological reasons. However, Cdt2:DDB1 seems to exhibit auto-ubiquitination activity, providing evidence for its *in vitro* activity. However, a replication of these experiments with additional controls is required to support such claims.

However, ubiquitinated species of PCNA and Cul4 and/or DDB1 were found. These findings verify auto-ubiquitination activity of the purified Cdt2:DDB1. Based on the existing literature, this likely represents the first ivUb using recombinant Cdt2. This method will provide the basis for subsequent biochemical analysis of the Cdt2-ubiqutination reaction in the future. So far published assays relied on the detection of proteins (429) or tags (293, 430) via antibodies. However, due to the lack of recombinant Cdt2:DDB1, no further experiments were possible in this study, leaving the requirements for other protein components, DNA, or phosphorylation untested.

### 13.5.3 Protein phosphorylation

Phosphorylation has been implicated in the binding of Cdt2 to DNA and PCNA prior to its ubiquitination activity (391, 400). Treatment of Cdt2:DDB1 with λPP hints at phosphorylation of the purified Cdt2:DDB1 complex. However, a more thorough analysis is needed to unambiguously assess the phosphorylation status of this protein. To achieve this, purified CDK2/CycA and λPP proteins could be used to add or remove phosphorylation, respectively, allowing to assess changes during complex formation and in ivUB. Additionally, pI and MS measurements could provide insights into the status and location of phosphorylation sites. De-phosphorylation experiments of Cdt2 were successfully reported before (391, 430), reporting that phosphorylation of Cdt2 negatively regulates its binding to and enzymatic activity at $PCNA^{DNA}$.

Furthermore, phase separation of DNA-Cdt1-mixtures have been found to be negatively regulated by CDK-dependent phosphorylation of Cdt1's IDR. (424). Since the $Cdt2^{CTD}$ exhibits similar properties as the $Cdt1^{NTD}$, a similar mechanism for localization and activation might be at work. Similarly, phosphorylation has also been implicated in p21-binding towards PCNA (432).

### 13.5.4 Assembly of protein complexes

The assembly of several complexes and their subsequent investigation via analytical GF resulted in the formation of a PCNA:Cdt2:DDB1 complex (1-PIP). The formation was identified by a shift in the elution profile relative to Cdt2:DDB1 and the existence of the corresponding protein bands in the SDS-PAGE. Missing Cdt1-incorporation might be attributed to its contamination by HSP70. Additional work needs to be done to confirm the existence of the annealed oligonucleotide in the complex. Furthermore, stable incorporation of Cdt1 needs to be achieved in order to obtain a Cdt1:$PCNA^{DNA}$:Cdt2:DDB1 complex (2-PIP). In order to ensure the proper formation of $PCNA^{DNA}$, replication factor C (RFC) could be used to load PCNA onto a specifically designed DNA construct (430, 433). Additionally, purified p21, should be used for *in vitro* reconstitution, due to its higher affinity towards PCNA (430).

So far, the formation of Cdt2-complex has mainly be studied *in vivo* using western blots, fluorescence microscopy, pull-down experiments, and XLMS (293, 313, 387, 389, 413). The present study demonstrates the reconstitution of large quantities of a DDB1:Cdt2-complex from purified components *in vitro* in agreement with previous findings (430). The previously published results reporting a direct Cdt2-PCNA interaction in the absence of a substrate were verified by the formation of a stable Cdt2:DDB1:PCNA complex reported here (430, 434). This homogeneous complex was obtained at high concentration, suitable for analysis via EM. Similar results were previously obtained using recombinant Flag-Cdt1 and co-expressed $CRL4^{Cdt2}$ purified through a glycerol gradient (430).

### 13.5.5 Negative staining and cryo-EM

Negative stain EM for this PCNA:Cdt2:DDB1 complex was performed, underlining the sample homogeneity. Subsequent cryo-EM analysis resulted in a small data set collected at a screening cryo-EM. In summary, the $PCNA^{DNA}$:Cdt2:DDB1 complex was successfully vitrified and exhibited consistently good behavior, i.e. a homogeneous particle size and distribution. A small data set was collected using a screening microscope and the images subsequently analyzed.

2D class averages suggested a high degree of conformational and compositional heterogeneity. This is expected from such a 2-body complex connected via the disordered $Cdt2^{CTD}$. The obtained class averages were separated into 7 groups by visual inspection, sampling several different views of the $PCNA^{DNA}$:Cdt2:DDB1 complex: PCNA alone (i) and with a bound protein at its periphery (ii) was clearly identified by its circular shape. Side (iii) and top views (iv) of Cdt2:DDB1 might show up to 4 domains, potentially representing the 4 expected WD40 domains. Additionally, a V-shaped particle (v) could be DDB1 alone while group (vi) may represent the full 2-body particle. Furthermore, a large disc-shaped particle with 7-fold symmetry (vii) could correspond to a co-purified proteasome. The high quality of the obtained class averages suggests the readiness of the $PCNA^{DNA}$:Cdt2:DDB1 complex sample for high

resolution data collection using a direct electron detector. Despite these promising results, a subsequent analysis of the remaining grids did not allow the collection of high-resolution data due to suboptimal particle density and distribution in the holes.

Structural analysis of Cdt2:DDB1 alone or as part of a larger CRL4 complex has not been carried out so far. Up to this point, only a $Cdt2^{PIP}$ peptide was crystallized in complex with PCNA, elucidating their interactions at atomic detail (430). Predictions of the Cdt2 structure in 2D (435, 436) and 3D (413, 437, 438) were performed, suggesting a β-propeller domain in the N-terminus (1-412 aa), while the C-terminus (413-730 aa) adopts a disordered conformation interspersed with a DNA-binding domain (460-580 aa) (439) and PIP box (706-713 aa) (430, 434). Taken together, the presented results provide a promising basis for an ongoing investigation of the Cdt2-PCNA mechanism.

## 13.6 Conclusion and Outlook

The results presented in this study provide a starting point for further exploration of Cdt2 mechanism of action by following an integrative biological approach in a tripartite way: (i) biochemical, (ii) structural and (iii) cellular.

(i) Biochemical investigation would aim in a characterization of Cdt2's regulation, interaction, and activity. Expression and purification of the clamp loader RFC should be performed to ensure the formation of a $PCNA^{DNA}$ species using specifically designed DNA-constructs (430, 440, 441). In such a way, homogeneous $PCNA^{DNA}$ for all subsequent steps would be available. In order to investigate DNA-binding of Cdt2 and PCNA, a DNA-binding assay would be required, allowing for a quantitative assessment. An analysis of Cdt2 using ivUb would shed light on the enzymatic activity of Cdt2 and its requirements for other proteins, nucleic acids (ssDNA and dsDNA), and PTMs (phosphorylation). In order to decipher the influence of phosphorylation of the $Cdt2^{CTD}$ on its binding and activity, CDK2/CycA kinase (442), λPP phosphatase, and/or Cdt2-mutants (391) should be used. Along that line, an influence of the phosphorylation status on liquid-liquid phase separation of Cdt2 in combination with DNA could be investigated following a recently published methodology (424). In vitro reconstitution of several complexes coupled with GF would provide insights into their formation, stability, and dependency. For the investivation of the entire $CRL4^{Cdt2}$, a crosslinking strategy to mimic ubiquitination transfer intermediates should be employed to reduce the inherent heterogeneity of CRL4s (284).

(ii) Structural investigation should be carried out on all complex along the reaction coordinate, namely Cdt2:DDB1 alone, proceeding with $PCNA^{DNA}$:Cdt2:DDB1 and Cdt1:$PCNA^{DNA}$:Cdt2:DDB1 complexes up to the full $CRL4^{Cdt2}$. Their investigation via cryo-EM and XLMS would probably uncover several structural states.

(iii) A cell-wide identification of Cdt2 substrates and other interacting proteins could be obtained via XLMS of a tagged Cdt2 variant, or a by using a Cdt2-BioID fusion protein (443, 444). In this way, proteins in close proximity to Cdt2 within the cell could be extracted. Additionally, tagged PCNA could be investigated in the same way to experimentally define two sets of proteins, Cdt2- and PCNA-associated. By combining these two sets an overlap for known and unknown substrates and interacting partners of the Cdt2-PCNA pathway might be found.

Taken together, the presented results represent an initial basis for the characterization of Cdt2 in terms of regulation, interaction and activity.

# 14 References


1. Orlova, E. V. et al. (2011) Structural analysis of macromolecular assemblies by electron microscopy. *Chem. Rev.*
2. Frank, J. (2006) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies - Oxford Scholarship* (Oxford University Press; 2 edition (February 2, 2006))
3. Cheng, Y. et al. (2015) A Primer to Single-Particle Cryo-Electron Microscopy. *Cell*
4. Bieniossek, C. et al. (2008) MultiBac: Multigene baculovirus-based eukaryotic protein complex production. *Curr. Protoc. Protein Sci.*
5. Behrmann, E. et al. (2015) Structural Snapshots of Actively Translating Human Ribosomes. *Cell*
6. Carragher, B. et al. (2019) Current outcomes when optimizing ‘standard’ sample preparation for single-particle cryo-EM. *J. Microsc.*
7. Errey, J. C. et al. (2020) Production of membrane proteins in industry: The example of GPCRs. *Protein Expr. Purif.*
8. Peisley, A. et al. (2015) 2D Projection Analysis of GPCR Complexes by Negative Stain Electron Microscopy. *Methods Mol. Biol.*
9. Dubochet, J. et al. (1988) Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys.*
10. Glaeser, R. M. (2018) Proteins, interfaces, and cryo-EM grids. *Curr. Opin. Colloid Interface Sci.*
11. Taylor, K. A. et al. (1976) Electron microscopy of frozen hydrated biological specimens. *J. Ultrasructure Res.*
12. Hayward, S. B. et al. (1979) Radiation damage of purple membrane at low temperature. *Ultramicroscopy*
13. EPU. Accessed 15.10.2020. https://www.thermofisher.com/de/de/home/industrial/electron-microscopy/electron-microscopy-instruments-workflow-solutions/electron-microscopy-software/software-updates.html
14. Carragher, B. et al. (2000) Leginon: an automated system for acquisition of images from vitreous ice specimens. *J. Struct. Biol.*
15. Schorb, M. et al. (2019) Software tools for automated transmission electron microscopy. *Nat. Methods*
16. Kohl, H. et al. (2007) *Transmission Electron Microscopy*
17. Wu, M. et al. (2020) Sub-2 Angstrom resolution structure determination using single-particle cryo-EM at 200 keV. *J. Struct. Biol. X*
18. Glaeser, R. M. (2019) How good can single-particle cryo-EM become? What remains before It approaches its physical limits? *Annu. Rev. Biophys.*
19. Kühlbrandt, W. (2014) The resolution revolution. *Science .*
20. Grant, T. et al. (2015) Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife*
21. Zheng, S. Q. et al. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*
22. Brilot, A. F. et al. (2012) Beam-induced motion of vitrified specimen on holey carbon film. *J. Struct. Biol.*
23. Russo, C. J. et al. (2014) Controlling protein adsorption on graphene for cryo-EM using low-energy hydrogen plasmas. *Nat. Methods*
24. Zi Tan, Y. et al. (2017) Addressing preferred specimen orientation in single-particle cryo-

EMthrough tilting. *Nat. Methods*

25. Rohou, A. et al. (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.*
26. Zhang, K. (2016) Gctf: Real-time CTF determination and correction. *J. Struct. Biol.*
27. Moriya, T. et al. (2017) High-resolution single particle analysis from electron cryo-microscopy images using SPHIRE. *J. Vis. Exp.*
28. Tegunov, D. et al. (2019) Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods*
29. Zivanov, J. et al. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife*
30. Punjani, A. et al. (2017) CryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods*
31. Grant, T. et al. (2018) CisTEM, user-friendly software for single-particle image processing. *Elife*
32. Lander, G. C. et al. (2009) Appion: An integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.*
33. de la Rosa-Trevín, J. M. et al. (2016) Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.*
34. Bartesaghi, A. et al. (2018) Atomic Resolution Cryo-EM Structure of β-Galactosidase. *Structure*
35. Tang, G. et al. (2007) EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.*
36. Wagner, T. et al. (2019) SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.*
37. Bepler, T. et al. (2018) in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*
38. Zhu, Y. et al. (2017) A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. *BMC Bioinformatics*
39. Scheres, S. H. W. et al. (2005) Maximum-likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.*
40. Wagner, T. Cinderella: Automatic 2D class selection. Accessed 14.10.2020. doi:10.5281/zenodo.3672421.
41. Penczek, P. A. (2010) Fundamentals of Three-Dimensional reconstruction from projections. *Methods Enzymol.*
42. Penczek, P. A. (2010) Image restoration in cryo-electron microscopy. *Methods Enzymol.*
43. Elmlund, D. et al. (2012) SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles. *J. Struct. Biol.*
44. Penczek, P. et al. (1996) A common-lines based method for determining orientations for N > 3 particle projections simultaneously. *Ultramicroscopy*
45. Penczek, P. A. et al. (2014) in *2014 IEEE International Conference on Image Processing, ICIP 2014*
46. Grigorieff, N. (2016) Frealign: An Exploratory Tool for Single-Particle Cryo-EM. *Methods Enzymol.*
47. Bartesaghi, A. et al. (2012) Protein secondary structure determination by constrained single-particle cryo-electron tomography. *Structure*
48. Galaz-Montoya, J. G. et al. (2017) The advent of structural biology in situ by single particle cryo-electron tomography. *Biophys. Reports*
49. Leschziner, A. E. et al. (2006) The orthogonal tilt reconstruction method: An approach to

generating single-class volumes with no missing cone for ab initio reconstruction of asymmetric particles. *J. Struct. Biol.*

50. Radermacher, M. (1988) Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *J. Electron Microsc. Tech.*

51. Rosenthal, P. B. et al. (2015) Validating maps from single particle electron cryomicroscopy. *Curr. Opin. Struct. Biol.*

52. Sigworth, F. J. et al. (2010) An introduction to maximum-likelihood methods in cryo-EM. *Methods Enzymol.*

53. Penczek, P. et al. (1992) Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy*

54. Penczek, P. A. (2010) *Resolution measures in molecular electron microscopy*

55. van Heel, M. et al. (2017) Reassessing the Revolution's Resolutions. *bioRxiv*

56. Henderson, R. et al. (2012) in *Structure*

57. Scheres, S. H. W. et al. (2012) Prevention of overfitting in cryo-EM structure determination. *Nat. Methods*

58. Kucukelbir, A. et al. (2014) Quantifying the local resolution of cryo-EM density maps. *Nat. Methods*

59. Cardone, G. et al. (2013) One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.*

60. Vilas, J. L. et al. (2018) MonoRes: Automatic and Accurate Estimation of Local Resolution for Electron Microscopy Maps. *Structure*

61. Alberts, B. (1998) The cell as a collection of protein machines: Preparing the next generation of molecular biologists. *Cell*

62. Loerke, J. et al. (2010) *Multiparticle Cryo-EM of Ribosomes* (Elsevier Inc.). 1st Ed.

63. Penczek, P. A. et al. (2006) A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. *J. Struct. Biol.*

64. Penczek, P. A. et al. (2006) Estimation of variance in single-particle reconstruction using the bootstrap technique. *J. Struct. Biol.*

65. Pettersen, E. F. et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.*

66. Penczek, P. A. et al. (2011) Identifying conformational states of macromolecules by eigen-analysis of resampled cryo-EM images. *J. Invest. Dermatol.*

67. Katsevich, E. et al. (2015) Covariance matrix estimation for the cryo-em heterogeneity problem. *SIAM J. Imaging Sci.*

68. Liao, H. Y. et al. (2015) Efficient Estimation of Three-Dimensional Covariance and its Application in the Analysis of Heterogeneous Samples in Cryo-Electron Microscopy. *Structure*

69. Loerke, J. et al. Sx3dvariability.

70. Dashti, A. et al. (2014) Trajectories of the ribosome as a Brownian nanomachine. *Proc. Natl. Acad. Sci.*

71. Jin, Q. et al. (2014) Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes. *Structure*

72. Plaschka, C. et al. (2015) Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature*

73. Nakane, T. et al. (2018) Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife*

74. Rosenthal, P. B. et al. (2003) Optimal determination of particle orientation, absolute hand, and

contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.*

75. Ratje, A. H. et al. (2010) Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature*

76. Heymann, J. B. (2001) in *Journal of Structural Biology*

77. Heymann, J. B. et al. (2007) Bsoft: Image processing and molecular modeling for electron microscopy. *J. Struct. Biol.*

78. Jakobi, A. J. et al. (2017) Model-based local density sharpening of cryo-EM maps. *Elife*

79. Terwilliger, T. C. et al. (2018) Automated map sharpening by maximization of detail and connectivity. *Acta Crystallogr. Sect. D Struct. Biol.*

80. Adams, P. D. et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.*

81. Malhotra, S. et al. (2019) Modelling structures in cryo-EM maps. *Curr. Opin. Struct. Biol.*

82. Alnabati, E. et al. (2020) Advances in structure modeling methods for cryo-electron microscopy maps. *Molecules*

83. Schrödinger, L. (2015) The PyMol Molecular Graphics System, Versión 1.8. *Thomas Hold.*

84. Emsley, P. et al. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.*

85. Casañal, A. et al. (2019) Current Developments in Coot for Macromolecular Model Building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci.*

86. Liebschner, D. et al. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix . *Acta Crystallogr. Sect. D Struct. Biol.*

87. Afonine, P. V. et al. (2018) New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. Sect. D Struct. Biol.*

88. Frenz, B. et al. (2017) RosettaES: A sampling strategy enabling automated interpretation of difficult cryo-EMmaps. *Nat. Methods*

89. Goddard, T. D. et al. (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.*

90. Chen, V. B. et al. (2010) MolProbity : all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.*

91. Joseph, A. P. et al. (2017) Improved metrics for comparing structures of macromolecular assemblies determined by 3D electron-microscopy. *J. Struct. Biol.*

92. Murshudov, G. N. (2016) in *Methods in Enzymology*

93. Darst, S. A. (2001) Bacterial RNA polymerase. *Curr. Opin. Struct. Biol.*

94. Murakami, K. S. (2015) Structural Biology of Bacterial RNA Polymerase. *Biomolecules*

95. Murakami, K. S. (2013) X-ray crystal structure of Escherichia coli RNA polymerase σ70 holoenzyme. *J. Biol. Chem.*

96. Minakhin, L. et al. (2001) Bacterial RNA polymerase subunit and eukaryotic RNA polymerase subunit RPB6 are sequence , structural , and functional homologs and promote RNA polymerase assembly.

97. Igarashi, K. et al. (1989) Promoter selectivity of Eschenichia coli RNA polymerase: omega factor is responsible for the ppGpp sensitivity.

98. Benoff, B. et al. (2002) Structural basis of transcription activation: The CAP-??CTD-DNA complex. *Science .*

99. Jeon, Y. H. et al. (1995) Solution structure of the activator contact domain of the RNA polymerase α subunit. *Science .*

100. Zhang, G. et al. (1999) Crystal structure of thermus aquaticus core RNA polymerase at 3.3 å

resolution. *Cell*

101. Markov, D. et al. (1999) A zinc-binding site in the largest subunit of DNA-dependent RNA polymerase involved in enzyme assembly. *Genes Dev.*

102. Kang, J. Y. et al. (2017) Structural basis of transcription arrest by coliphage HK022 Nun in an Escherichia coli RNA polymerase elongation complex.

103. Gezvain, K. . et al. (2004) The structure of bacterial RNA polymerase. *Bact. Chromosom.*

104. Cramer, P. et al. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*

105. Jovanovic, M. et al. (2011) Activity map of the Escherichia coli RNA polymerase bridge helix. *J. Biol. Chem.*

106. Seibold, S. A. et al. (2010) Conformational coupling, bridge helix dynamics and active site dehydration in catalysis by RNA polymerase. *Biochim. Biophys. Acta - Gene Regul. Mech.*

107. Vassylyev, D. G. et al. (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*

108. Vassylyev, D. G. et al. (2007) Structural basis for substrate loading in bacterial RNA polymerase. *Nature*

109. Belogurov, G. A. et al. (2019) The Mechanisms of Substrate Selection, Catalysis, and Translocation by the Elongating RNA Polymerase. *J. Mol. Biol.*

110. Nudler, E. (2012) RNA polymerase backtracking in gene regulation and genome instability. *Cell*

111. Toulokhonov, I. et al. (2001) Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science .*

112. Santangelo, T. J. et al. (2011) Termination and antitermination: RNA polymerase runs a stop sign. *Nat. Rev. Microbiol.*

113. Wang, D. et al. (1997) Nuclease cleavage of the upstream half of the nontemplate strand DNA in an Escherichia coli transcription elongation complex causes upstream translocation and transcriptional arrest. *J. Biol. Chem.*

114. Nudler, E. et al. (1996) Transcription processivity: Protein-DNA interactions holding together the elongation complex. *Science .*

115. Holmes, S. F. et al. (2003) Downstream DNA Sequence Effects on Transcription Elongation. *J. Biol. Chem.*

116. Lee, D. N. et al. (1990) Transcription pausing by Escherichia coli RNA polymerase is modulated by downstream DNA sequences. *J. Biol. Chem.*

117. Reynolds, R. et al. (1992) Parameters affecting transcription termination by Escherichia coli RNA. II. Construction and analysis of hybrid terminators. *J. Mol. Biol.*

118. Telesnitsky, A. et al. (1989) Terminator-Distal Sequences Determine the in Vitro Efficiency of the Early Terminators of Bacteriophages T3 and T7. *Biochemistry*

119. Zhang, J. et al. Substrate Loading, Nucleotide Addition, and Translocation by RNA Polymerase.

120. Toulokhonov, I. et al. (2003) The flap domain is required for pause RNA hairpin inhibition of catalysis by RNA polymerase and can modulate intrinsic termination. *Mol. Cell*

121. Gnatt, A. L. et al. (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science*

122. Cramer, P. et al. (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science .*

123. Ederth, J. et al. (2002) The downstream DNA jaw of bacterial RNA polymerase facilitates both transcriptional initiation and pausing. *J. Biol. Chem.*

124. Murakami, K. S. et al. (2003) Bacterial RNA polymerases: the wholo story. *Curr. Opin. Struct.*

*Biol.*

125. French, S. L. et al. (1989) Transcription mapping of the Escherichia coli chromosome by electron microscopy. *J. Bacteriol.*

126. Brueckner, F. et al. (2009) A movie of the RNA polymerase nucleotide addition cycle. *Curr. Opin. Struct. Biol.*

127. Steitz, P. (1998) A mechanism for all polymerases Thomas A. *Nature*

128. Sosunov, V. et al. (2003) Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J.*

129. Kireeva, M. L. et al. (2008) Transient Reversal of RNA Polymerase II Active Site Closing Controls Fidelity of Transcription Elongation. *Mol. Cell*

130. Belogurov, G. A. et al. (2015) Regulation of Transcript Elongation. *Annu. Rev. Microbiol.*

131. Cheung, A. C. M. et al. (2012) A movie of RNA polymerase II transcription. *Cell*

132. Ray-Soni, A. et al. (2016) Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu. Rev. Biochem.*

133. Carafa, Y. d. A. et al. (1990) Prediction of rho-independent Escherichia coli transcription terminators. A statistical analysis of their RNA stem-loop structures. *J. Mol. Biol.*

134. Richardson, L. V. et al. (1996) Rho-dependent termination of transcription is governed primarily by the upstream Rho utilization (rut) sequences of a terminator. *J. Biol. Chem.*

135. Chen, C. Y. et al. (1987) Sequence elements essential for rho-dependent transcription termination at lambda tR1. *J. Biol. Chem.*

136. Zalatan, F. et al. (1992) Effects of decreased cytosine content on rho interaction with the rho-dependent terminator trp t' in Escherichia coli. *J. Biol. Chem.*

137. McSwiggen, J. A. et al. (1988) Interactions of Escherichia coli transcription termination factor rho with RNA. I. Binding stoichiometries and free energies. *J. Mol. Biol.*

138. Lowery Goldhammer, C. et al. (1974) An RNA dependent nucleoside triphosphate phosphohydrolase (ATPase) associated with rho termination factor. *Proc. Natl. Acad. Sci. U. S. A.*

139. Rutherford, S. T. et al. (2009) Allosteric control of escherichia coli rRNA promoter complexes by DksA. *Genes Dev.*

140. Thomsen, N. D. et al. (2009) Running in Reverse: The Structural Basis for Translocation Polarity in Hexameric Helicases. *Cell*

141. Browne, R. J. et al. (2005) Catalytic cooperativity among subunits of Escherichia coli transcription termination factor Rho: Kinetics and substrate structural requirements. *J. Biol. Chem.*

142. Koslover, D. J. et al. (2012) Binding and translocation of termination factor Rho studied at the single-molecule level. *J. Mol. Biol.*

143. Soares, E. et al. (2014) The RNA-mediated, asymmetric ring regulatory mechanism of the transcription termination Rho helicase decrypted by time-resolved Nucleotide Analog Interference Probing (trNAIP). *Nucleic Acids Res.*

144. Steinmetz, E. J. et al. (1990) A short intervening structure can block rho factor helicase action at a distance. *J. Biol. Chem.*

145. Ding Jun Jin et al. (1992) Termination efficiency at rho-dependent terminators depends on kinetic coupling between RNA polymerase and rho. *Proc. Natl. Acad. Sci. U. S. A.*

146. Park, J. S. et al. (2006) Role of DNA bubble rewinding in enzymatic transcription termination. *Proc. Natl. Acad. Sci. U. S. A.*

147. Richardson, J. P. (2002) Rho-dependent termination and ATPases in transcript termination. *Biochim. Biophys. Acta - Gene Struct. Expr.*

148. Schwartz, A. et al. (2007) Transcription termination factor Rho can displace streptavidin from biotinylated RNA. *J. Biol. Chem.*

149. Tabib-Salazar, A. et al. (2019) Xenogeneic Regulation of the Bacterial Transcription Machinery. *J. Mol. Biol.*

150. Henkin, T. M. (2008) Riboswitch RNAs: Using RNA to sense cellular metabolism. *Genes Dev.*

151. Lederberg, E. M. (1951) Lysogenicity of E. coli Strain K-12.

152. Schweimer, K. et al. (2011) in *Flexible Viruses: Structural Disorder in Viral Proteins*, pp 425–444.

153. Casjens, S. R. et al. (2016) Bacteriophage lambda : early pioneer and still relevant.

154. Greenblatt, J. et al. (1993) Transcriptional antitermination. *Nature*

155. Chen, J. et al. (2017) 6S RNA Mimics B-Form DNA to Regulate Escherichia coli RNA Polymerase. *Mol. Cell*

156. Zuo, Y. et al. (2013) The Mechanism of E. coli RNA Polymerase Regulation by ppGpp is suggested by the structure of their complex. *Mol. Cell*

157. Werner, F. (2012) A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *J. Mol. Biol.*

158. Mooney, R. A. et al. (2009) Two structurally independent domains of E. coli NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. *J. Mol. Biol.*

159. Steiner, T. (2002) Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. *EMBO J.*

160. Kyrpides, N. C. et al. (1996) KOW: A novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem. Sci.*

161. NandyMazumdar, M. et al. (2015) Ubiquitous transcription factors display structural plasticity and diverse functions. *BioEssays*

162. Kang, J. Y. et al. (2018) Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. *Cell*

163. Li, J. et al. (1992) NusG, a new Escherichia coli elongation factor involved in transcriptional antitermination by the N protein of phage λ. *J. Biol. Chem.*

164. Sullivan, S. L. et al. (1992) Requirement for E. coli NusG protein in factor-dependent transcription termination. *Cell*

165. Cardinale, C. J. et al. (2008) Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. *Science*

166. Martinez-Rucobo, F. W. et al. (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.*

167. Sevostyanova, A. et al. (2011) The β Subunit Gate Loop Is Required for RNA Polymerase Modification by RfaH and NusG. *Mol. Cell*

168. Drögemüller, J. et al. (2015) Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Sci. Rep.*

169. Klein, B. J. et al. (2011) RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc. Natl. Acad. Sci. U. S. A.*

170. Li, J. et al. (1993) Elongation factor NusG interacts with termination factor rho to. *Genes Dev*

171. Burmann, B. M. et al. (2010) A NusE:NusG Complex Links Transcription and Translation. *Science .*

172. Luo, X. et al. (2008) Structural and Functional Analysis of the E. coli NusB-S10 Transcription Antitermination Complex. *Mol. Cell*

173. Mechold, U. et al. (2013) Differential regulation by ppGpp versus pppGpp in Escherichia coli.

*Nucleic Acids Res.*

174. Guo, X. et al. (2018) Structural Basis for NusA Stabilized Transcriptional Pausing. *Mol. Cell*

175. Skordalakes, E. et al. (2003) Structure of the Rho transcription terminator: Mechanism of mRNA recognition and helicase loading. *Cell*

176. Gill, S. C. et al. (1991) Escherichia coli σ70 and NusA proteins. I. Binding interactions with core RNA polymerase in solution and within the transcription complex. *J. Mol. Biol.*

177. Mooney, R. A. et al. (2009) Regulator trafficking on bacterial transcription units in vivo. *Mol. Cell*

178. Gusarov, I. et al. (2001) Control of Intrinsic Transcription Termination by N and NusA. *Cell*

179. Ha, K. S. et al. (2010) The NusA N-Terminal Domain Is Necessary and Sufficient for Enhancement of Transcriptional Pausing via Interaction with the RNA Exit Channel of RNA Polymerase. *J. Mol. Biol.*

180. Gusarov, I. et al. (1999) The mechanism of intrinsic transcription termination. *Mol. Cell*

181. Liao, M. et al. (2014) Single particle electron cryo-microscopy of a mammalian ion channel. *Curr. Opin. Struct. Biol.*

182. Yang, X. et al. (2010) The interaction between RNA polymerase and the elongation factor NusA. *RNA Biol.*

183. Weixlbaumer, A. et al. (2013) Structural basis of transcriptional pausing in bacteria. *Cell*

184. Hein, P. P. et al. (2014) RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat. Struct. Mol. Biol.*

185. Ma, C. et al. (2015) RNA polymerase-induced remodelling of NusA produces a pause enhancement complex. *Nucleic Acids Res.*

186. Prasch, S. et al. (2009) RNA-binding specificity of E. coli NusA. *Nucleic Acids Res.*

187. Worbs, M. et al. (2001) An Extended RNA Binding Surface through Arrayed S1 and KH Domains in Transcription Factor NusA. *Mol. Cell*

188. Subramanian, A. R. (1983) Structure and Functions of Ribosomal Protein S1. *Prog. Nucleic Acid Res. Mol. Biol.*

189. Siomi, H. et al. (1993) The pre-mRNA binding K protein contains a novel evolutionary conserved motif. *Nucleic Acids Res.*

190. Beuth, B. et al. (2005) Structure of a Mycobacterium tuberculosis NusA-RNA complex. *EMBO J.*

191. Mah, T.-F. et al. (1999) Functional importance of regions in Escherichia coli elongation factor NusA that interact with RNA polymerase, the bacteriophage lambda N protein and RNA. *Mol. Microbiol.*

192. Messias, A. C. et al. (2004) Structural basis of single-stranded RNA recognition. *Acc. Chem. Res.*

193. Valverde, R. et al. (2008) Structure and function of KH domains. *FEBS J.*

194. Eisenmann, A. et al. (2005) The E. coli NusA carboxy-terminal domains are structurally similar and show specific RNAP- and lambdaN interaction. *Protein Sci.*

195. Bonin, I. et al. (2004) Structural basis for the interaction of Escherichia coli NusA with protein N of phage λ. *Proc. Natl. Acad. Sci. U. S. A.*

196. Doherty, A. J. et al. (1996) The helix-hairpin-helix DNA-binding motif: A structural basis for non-sequence-specific recognition of DNA. *Nucleic Acids Res.*

197. Shao, X. (2000) Common fold in helix-hairpin-helix proteins. *Nucleic Acids Res.*

198. Thanos, C. D. et al. (1999) Oligomeric structure of the human EphB2 receptor SAM domain. *Science .*

199. Mah, T. F. et al. (2000) The alpha subunit of E. coli RNA polymerase activates RNA binding by NusA. *Genes Dev.*

200. Schweimer, K. et al. (2011) NusA interaction with the α subunit of E. coli RNA polymerase is via the UP element site and releases autoinhibition. *Structure*

201. Squires, C. L. et al. (2000) Proteins Shared by the Transcription and Translation Machines. *Annu. Rev. Microbiol.*

202. Friedman, D. I. et al. (1981) Evidence that ribosomal protein S10 participates in control of transcription termination. *Proc. Natl. Acad. Sci. U. S. A.*

203. Gopal, B. et al. (2001) Spectroscopic and thermodynamic characterization of the transcription antitermination factor NusE and its interaction with NusB from Mycobacterium tuberculosis. *Biochemistry*

204. Das, R. et al. (2008) Structural Biophysics of the NusB:NusE Antitermination Complex. *J. Mol. Biol.*

205. Burmann, B. M. et al. (2012) An α helix to β barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell*

206. Altieri, A. S. et al. (2000) The structure of the transcriptional antiterminator NusB from Escherichia coli. *Nat. Struct. Biol.*

207. Gopal, B. et al. (2000) The crystal structure of NusB from Mycobacterium tuberculosis. *Nat. Struct. Biol.*

208. Mason, S. W. et al. (1992) Direct interaction between two Escherichia coli transcription antitermination factors, NusB and ribosomal protein S10. *J. Mol. Biol.*

209. Burmann, B. M. et al. (2010) Fine tuning of the E. coli NusB:NusE complex affinity to BoxA RNA is required for processive antitermination. *Nucleic Acids Res.*

210. Greive, S. J. et al. (2005) Assembly of an RNA-protein complex. Binding of NusB and NusE (S10) proteins to boxA RNA nucleates the formation of the antitermination complex involved in controlling rRNA transcription in Escherichia coli. *J. Biol. Chem.*

211. Zuber, M. et al. (1987) Analysis of nutR, a site required for transcription antitermination in phage lambda. *Proc. Natl. Acad. Sci. U. S. A.*

212. DeVito, J. et al. (1994) Control of transcription processivity in phage lambda: Nus factors strengthen the termination-resistant state of RNA polymerase induced by N antiterminator. *Proc. Natl. Acad. Sci. U. S. A.*

213. Torres, M. et al. (2004) In Vivo Effect of NusB and NusG on rRNA Transcription Antitermination. *J. Bacteriol.*

214. Sheppard, C. et al. (2013) A non-bacterial transcription factor inhibits bacterial transcription by a multipronged mechanism. *RNA Biol.*

215. Ooi, W. Y. et al. (2018) A Thermus phage protein inhibits host RNA polymerase by preventing template DNA strand loading during open promoter complex formation. *Nucleic Acids Res.*

216. Robert, J. et al. (1987) The remarkable specificity of a new transcription termination factor suggests that the mechanisms of termination and antitermination are similar. *Cell*

217. Hung, S. C. et al. (1997) The Nun protein of bacteriophage HK022 inhibits translocation of Escherichia coli RNA polymerase without abolishing its catalytic activities. *Genes Dev.*

218. Vitiello, C. L. et al. (2014) Coliphage HK022 Nun protein inhibits RNA polymerase translocation. *Proc. Natl. Acad. Sci.*

219. Robledo, R. et al. (1991) Escherichia coli mutations that block transcription termination by phage {HK}022 {N}un protein. *J. Mol. Biol.*

220. Chattopadhyay, S. et al. (1995) Interaction between the phage HK022 Nun protein and the nut RNA of phage λ. *Proc. Natl. Acad. Sci. U. S. A.*

221. Faber, C. et al. (2001) The structure of the coliphage HK022 Nun protein-λ-phage boxB RNA

complex. Implications for the mechanism of transcription termination. *J. Biol. Chem.*

222. Schärpf, M. et al. (2000) Antitermination in bacteriophage lambda. The structure of the N36 peptide-boxB RNA complex. *Eur. J. Biochem.*

223. Roberts, J. W. et al. (1998) Antitermination by bacteriophage λ Q protein. *Cold Spring Harb. Symp. Quant. Biol.*

224. Yin, Z. et al. (2019) Structural basis of Q-dependent antitermination.

225. Vorobiev, S. M. et al. (2014) Short Article Structure of the DNA-Binding and RNA-Polymerase-Binding Region of Transcription Antitermination Factor I Q. *Struct. Des.*

226. Shi, J. et al. Structural basis of Q-dependent transcription antitermination. *Nat. Commun.*

227. Mogridge, J. et al. (1995) A protein - RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase of a stable antitermination complex containing the λ N protein. *Genes Dev.*

228. Brigati, C. et al. (2003) HIV Tat, its TARgets and the control of viral gene expression. *FEMS Microbiol. Lett.*

229. Mogridge, J. et al. (1998) Independent Ligand-Induced Folding of the RNA-Binding Domain and Two Functionally Distinct Antitermination Regions in the Phage λ N Protein. *Mol. Cell*

230. Conant, C. R. et al. (2005) A quantitative description of the binding states and in vitro function of antitermination protein N of bacteriophage lambda. *J. Mol. Biol.*

231. Tan, R. et al. (1995) Structural variety of arginine-rich RNA-binding peptides. *Proc. Natl. Acad. Sci. U. S. A.*

232. Cilley, C. D. et al. (1997) Analysis of bacteriophage N protein and peptide binding to boxB RNA using polyacrylamide gel coelectrophoresis (PACE). *RNA*

233. Van Gilst, M. R. et al. (1997) Assembly of the N-dependent antitermination complex of phage lambda: NusA and RNA bind independently to different unfolded domains of the N protein. *J. Mol. Biol.*

234. Legault, P. et al. (1998) NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell*

235. Prasch, S. et al. (2006) Interaction of the intrinsically unstructured phage λ N protein with Escherichia coli NusA. *Biochemistry*

236. Greenblatt, J. et al. (1998) Structure and mechanism in transcriptional antitermination by the bacteriophage lambda N protein. *Cold Spring Harb. Symp. Quant. Biol.*

237. Radding, C. M. (1964) Nuclease Activity in defective Lysogens of Phage λ.

238. Echols, H. (1971) Lysogeny: viral repression and site-specific recombination. *Annu. Rev. Biochem.*

239. Weiler, N. et al. (2008) Top-down laminar organization of the excitatory network in motor cortex. *Nat. Neurosci.*

240. Lui, J. H. et al. (2011) Development and evolution of the human neocortex. *Cell*

241. Corbin, J. G. et al. (2001) Telencephalic cells take a tangent: Non-radial migration in the mammalian forebrain. *Nat. Neurosci.*

242. DeBoer, E. M. et al. (2013) Post-transcriptional regulatory elements and spatiotemporal specification of neocortical stem cells and projection neurons. *Neuroscience*

243. Molyneaux, B. J. et al. (2007) Neuronal subtype specification in the cerebral cortex. *Nat. Rev. Neurosci.*

244. Kwan, K. Y. et al. (2012) Transcriptional co-regulation of neuronal migration and laminar identity in the neocortex. *Development*

245. Harris, K. D. et al. (2015) The neocortical circuit: Themes and variations. *Nat. Neurosci.*

246. De Wit, J. et al. (2016) Specification of synaptic connectivity by cell surface interactions. *Nat. Rev. Neurosci.*

247. Pilaz, L. J. et al. (2015) Post-transcriptional regulation in corticogenesis: How RNA-binding proteins help build the brain. *Wiley Interdiscip. Rev. RNA*

248. Kraushar, M. L. et al. (2016) The frontier of RNA metamorphosis and ribosome signature in neocortical development. *Int. J. Dev. Neurosci.*

249. Mills, E. W. et al. (2017) Ribosomopathies: There's strength in numbers. *Science .*

250. Shi, Z. et al. (2015) Translating the Genome in Time and Space: Specialized Ribosomes, RNA Regulons, and RNA-Binding Proteins. *Annu. Rev. Cell Dev. Biol.*

251. Melnikov, S. et al. (2012) One core, two shells: Bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.*

252. Jackson, R. J. et al. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.*

253. Munro, J. B. et al. (2009) Navigating the ribosome's metastable energy landscape. *Trends Biochem. Sci.*

254. Dever, T. E. et al. (2012) Phases of Translation in Eukaryotes. *Cold Spring Harb. Perspect. Biol. Perspect Biol.*

255. Hipp, M. S. et al. (2019) The proteostasis network and its decline in ageing. *Nat. Rev. Mol. Cell Biol.*

256. Balch, W. E. et al. (2008) Adapting proteostasis for disease intervention. *Science .*

257. Rakic, P. (2009) Evolution of the neocortex: A perspective from developmental biology. *Nat. Rev. Neurosci.*

258. Bosco, D. A. et al. (2011) Proteostasis and movement disorders: Parkinson's disease and amyotrophic lateral sclerosis. *Cold Spring Harb. Perspect. Biol.*

259. Kapur, M. et al. (2018) MRNA Translation Gone Awry: Translation Fidelity and Neurological Disease. *Trends Genet.*

260. Nguyen, D. Q. et al. (2018) The role of ErbB3 binding protein 1 in cancer: Friend or foe? *J. Cell. Physiol.*

261. Kowalinski, E. et al. (2007) The crystal structure of Ebp1 reveals a methionine aminopeptidase fold as binding platform for multiple interactions. *FEBS Lett.*

262. Monie, T. P. et al. (2007) Structural insights into the transcriptional and translational roles of Ebp1. *EMBO J.*

263. Lowther, W. T. et al. (2000) Structure and function of the methionine aminopeptidases. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.*

264. Ball, L. A. et al. (1973) Cleavage of the N-terminal formymlethionine residue from a bacteriophage coat protein in vitro. *J. Mol. Biol.*

265. Yoo, J. Y. et al. (2000) Interaction of the PA2G4 (EBP1) protein with ErbB-3 and regulation of this binding by heregulin. *Br. J. Cancer*

266. Pilipenko, E. V. et al. (2000) A cell cycle-dependent protein serves as a template-specific translation initiation factor. *Genes Dev.*

267. Squatrito, M. et al. (2006) Ebp1 is a dsRNA-binding protein associated with ribosomes that modulates eIF2α phosphorylation. *Biochem. Biophys. Res. Commun.*

268. Zhang, Y. et al. (2003) Repression of E2F1-mediated transcription by the ErbB3 binding protein Ebp1 involves histone deacetylases. *Nucleic Acids Res.*

269. Zhang, Y. et al. (2005) The ErbB3 binding protein Ebp1 interacts with Sin3A to repress E2F1 and AR-mediated transcription. *Nucleic Acids Res.*

270. Wild, K. et al. (2020) MetAP-like Ebp1 occupies the human ribosomal tunnel exit and recruits

flexible rRNA expansion segments. *Nat. Commun.*

271. Hershko, A. et al. (1998) THE UBIQUITIN SYSTEM. *Annu. Rev. Biochem.*

272. Kaiser, S. E. et al. (2011) Protein standard absolute quantification (PSAQ) method for the measurement of cellular ubiquitin pools. *Nat. Methods*

273. Nguyen, H. C. et al. (2017) Chapter 12 Cullin-RING E3 Ubiquitin Ligases: Bridges to Destruction Henry. *Subcell. Biochem.*

274. Komander, D. et al. (2012) The Ubiquitin Code. *Annu. Rev. Biochem.*

275. Saeki, Y. (2017) JB special review - Recent topics in ubiquitin-proteasome system and autophagy: Ubiquitin recognition by the proteasome. *J. Biochem.*

276. Hoege, C. et al. (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature*

277. Palombella, V. J. et al. (1994) The ubiquitinproteasome pathway is required for processing the NF-κB1 precursor protein and the activation of NF-κB. *Cell*

278. Terrell, J. et al. (1998) A function for monoubiquitination in the internalization of a G protein-coupled receptor. *Mol. Cell*

279. Wijk, S. J. L. et al. (2010) The family of ubiquitin-conjugating enzymes (E2s): deciding between life and death of proteins. *FASEB J.*

280. Soucy, T. A. et al. (2009) Targeting NEDD8-activated cullin-RING ligases for the treatment of cancer. *Clin. Cancer Res.*

281. Petroski, M. D. et al. (2005) Function and regulation of cullin–RING ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.*

282. Scott, D. C. et al. (2018) SCF E3 Ligase Substrates Switch from CAN-D to Can-ubiquitylate. *Mol. Cell*

283. Wang, K. et al. (2020) in *Advances in Experimental Medicine and Biology*

284. Baek, K. et al. (2020) NEDD8 nucleates a multivalent cullin–RING–UBE2D ubiquitin ligation assembly. *Nature*

285. Lee, J. et al. (2007) DCAFs, the Missing Link of the CUL4-DDB1 Ubiquitin Ligase. *Mol. Cell*

286. Angers, S. et al. (2006) Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature*

287. Zimmerman, E. S. et al. (2010) Structural assembly of cullin-RING ubiquitin ligase complexes. *Curr. Opin. Struct. Biol.*

288. Scrima, A. et al. (2008) Structural Basis of UV DNA-Damage Recognition by the DDB1-DDB2 Complex. *Cell*

289. Schwefel, D. et al. (2014) Structural basis of lentiviral subversion of a cellular protein degradation pathway. *Nature*

290. Lucas, X. et al. (2017) Recognition of substrate degrons by E3 ubiquitin ligases and modulation by small-molecule mimicry strategies. *Curr. Opin. Struct. Biol.*

291. He, Y. J. et al. (2006) DDB1 functions as a linker to recruit receptor WD40 proteins to CUL4 – ROC1 ubiquitin ligases.

292. Higa, L. A. et al. (2006) CUL4-DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation. *Nat. Cell Biol.*

293. Jin, J. et al. (2006) A Family of Diverse Cul4-Ddb1-Interacting Proteins Includes Cdt2, which Is Required for S Phase Destruction of the Replication Factor Cdt1. *Mol. Cell*

294. Fischer, E. S. et al. (2011) The molecular basis of CRL4DDB2/CSA ubiquitin ligase architecture, targeting, and activation. *Cell*

295. Faust, T. B. et al. (2020) Structural complementarity facilitates E7820-mediated degradation of

RBM39 by DCAF15. *Nat. Chem. Biol.*

296. Bussiere, D. E. et al. (2020) Structural basis of indisulam-mediated RBM39 recruitment to DCAF15 E3 ligase complex. *Nat. Chem. Biol.*

297. Nowak, M. et al. (2019) DCAF8, a novel MuRF1 interaction partner, promotes muscle atrophy. *J. Cell Sci.*

298. Panagopoulos, A. et al. (2020) $CRL4^{Cdt2}$: Coupling Genome Stability to Ubiquitination. *Trends Cell Biol.*

299. Pan, W. W. et al. (2013) Ubiquitin E3 ligase CRL4CDT2/DCAF2 as a potential chemotherapeutic target for ovarian surface epithelial cancer. *J. Biol. Chem.*

300. Peifer, M. et al. (1994) A repeating amino acid motif shared by proteins with diverse cellular roles. *Cell*

301. Kim, K. et al. (2013) VprBP Has Intrinsic Kinase Activity Targeting Histone H2A and Represses Gene Transcription. *Mol. Cell*

302. Ahn, J. et al. (2011) The cullin-RING E3 ubiquitin ligase CRL4-DCAF1 complex dimerizes via a short helical region in DCAF1. *Biochemistry*

303. Wang, D. et al. (2016) Acetylation-regulated interaction between p53 and SET reveals a widespread regulatory mode. *Nature*

304. Huang, J. et al. (2008) VprBP targets Merlin to the Roc1-Cul4A-DDB1 E3 ligase complex for degradation. *Oncogene*

305. Schwefel, D. et al. (2015) Molecular determinants for recognition of divergent SAMHD1 proteins by the lentiviral accessory protein Vpx. *Cell Host Microbe*

306. Wu, Y. et al. (2016) The DDB1 – DCAF1 – Vpr – UNG2 crystal structure reveals how HIV-1 Vpr steers human UNG2 toward destruction. *Nat. Publ. Gr.*

307. Nakagawa, T. et al. (2013) VprBP (DCAF1): a promiscuous substrate recognition subunit that incorporates into both RING-family CRL4 and HECT-family EDD/UBR5 E3 ubiquitin ligases. *BMC Mol. Biol.*

308. Schabla, N. M. et al. (2019) DCAF1 (VprBP): Emerging physiological roles for a unique dual-service E3 ubiquitin ligase substrate receptor. *J. Mol. Cell Biol.*

309. Hrecka, K. et al. (2007) Lentiviral Vpr usurps Cul4-DDB1[VprBP] E3 ubiquitin ligase to modulate cell cycle. *Proc. Natl. Acad. Sci. U. S. A.*

310. Zhang, S. et al. (2001) Cytoplasmic retention of HIV-1 regulatory protein Vpr by protein-protein interaction with a novel human cytoplasmic protein VprBP. *Gene*

311. Mahon, C. et al. (2014) Cullin E3 Ligases and their rewiring by viral factors. *Biomolecules*

312. Precious, B. et al. (2005) In vitro and in vivo specificity of ubiquitination and degradation of STA1 and STAT2 by the V protiens of the paramyxoviruses simian virus 5 and human parainfluenza virus type 2. *J. Gen. Virol.*

313. Angers, S. et al. (2006) Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature*

314. Albin, J. S. et al. (2013) Dispersed sites of HIV Vif-dependent polyubiquitination in the DNA deaminase APOBEC3F. *J. Mol. Biol.*

315. Pe, F. J. et al. (2010) Multilayered Mechanism of CD4 Downregulation by HIV-1 Vpu Involving Distinct ER Retention and ERAD Targeting Steps.

316. Greenwood, E. J. D. et al. (2019) Promiscuous Targeting of Cellular Proteins by Vpr Drives Systems-Level Proteomic Remodeling in HIV-1 Infection. *Cell Rep.*

317. Fabryova, H. et al. (2019) Vpr and Its Cellular Interaction Partners: R We There Yet? *Cells*

318. Schröfelbauer, B. et al. (2005) Human Immunodeficiency Virus Type 1 Vpr Induces the Degradation of the UNG and SMUG Uracil-DNA Glycosylases. *J. Virol.*

319. Lahouassa, H. et al. (2016) HIV-1 Vpr degrades the HLTF DNA translocase in T cells and macrophages. *Proc. Natl. Acad. Sci. U. S. A.*

320. Laguette, N. et al. (2014) Premature activation of the slx4 complex by vpr promotes g2/m arrest and escape from innate immune sensing. *Cell*

321. Zhou, X. et al. (2016) SLX4-SLX1 protein-independent down-regulation of MUS81-EME1 protein by HIV-1 viral protein R (Vpr). *J. Biol. Chem.*

322. Romani, B. et al. (2015) HIV-1 Vpr protein enhances proteasomal degradation of MCM10 DNA replication factor through the Cul4-DDB1[VprBP] E3 ubiquitin ligase to induce G2/M cell cycle arrest. *J. Biol. Chem.*

323. Lv, L. et al. (2018) Vpr Targets TET2 for Degradation by CRL4 VprBP E3 Ligase to Sustain IL-6 Expression and Enhance HIV-1 Replication. *Mol. Cell*

324. Lim, E. S. et al. (2012) The Ability of Primate Lentiviruses to Degrade the Monocyte Restriction Factor SAMHD1 Preceded the Birth of the Viral Accessory Protein Vpx. *Cell Host Microbe*

325. Powell, R. D. et al. (2011) Aicardi-Goutières syndrome gene and HIV-1 restriction factor SAMHD1 is a dGTP-regulated deoxynucleotide triphosphohydrolase. *J. Biol. Chem.*

326. Goldstone, D. C. et al. (2011) HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature*

327. Zhu, C. et al. (2013) Structural insight into dGTP-dependent activation of tetrameric SAMHD1 deoxynucleoside triphosphate triphosphohydrolase. *Nat. Commun.*

328. Kim, B. et al. (2012) Tight interplay among SAMHD1 protein level, cellular dNTP levels, and HIV-1 proviral DNA synthesis kinetics in human primary monocyte-derived macrophages. *J. Biol. Chem.*

329. Lahouassa, H. et al. (2012) SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat. Immunol.*

330. St Gelais, C. et al. (2012) SAMHD1 restricts HIV-1 infection in dendritic cells (DCs) by dNTP depletion, but its expression in DCs and primary CD4+ T-lymphocytes cannot be upregulated by interferons. *Retrovirology*

331. Rehwinkel, J. et al. (2013) SAMHD1-dependent retroviral control and escape in mice. *EMBO J.*

332. Morris, E. R. et al. (2019) The missing link: Allostery and catalysis in the anti-viral protein SAMHD1. *Biochem. Soc. Trans.*

333. Hrecka, K. et al. (2011) Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature*

334. Laguette, N. et al. (2011) SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature*

335. Shingai, M. et al. (2015) The Expression of Functional Vpx during Pathogenic SIVmac Infections of Rhesus Macaques Suppresses SAMHD1 in CD4+ Memory T Cells. *PLoS Pathog.*

336. Baldauf, H. M. et al. (2012) SAMHD1 restricts HIV-1 infection in resting CD4 + T cells. *Nat. Med.*

337. Fregoso, O. I. et al. (2013) Evolutionary Toggling of Vpx/Vpr Specificity Results in Divergent Recognition of the Restriction Factor SAMHD1. *PLoS Pathog.*

338. Wu, Y. et al. (2015) Structural basis of clade-specific engagement of SAMHD1 (sterile α motif and histidine/aspartate-containing protein 1) restriction factors by lentiviral viral protein X (Vpx) virulence factors. *J. Biol. Chem.*

339. Spragg, C. J. et al. (2013) Antagonism of SAMHD1 is actively maintained in natural infections of simian immunodeficiency virus. *Proc. Natl. Acad. Sci.*

340. Said, N. et al. (2017) Structural basis for λN-dependent processive transcription antitermination. *Nat. Microbiol.*

341. Krupp, F. et al. (2019) Structural Basis for the Action of an All-Purpose Transcription Anti-

termination Factor. *Mol. Cell*

342. Liu, B. et al. (2017) Structural insights into NusG regulating transcription elongation. *Nucleic Acids Res.*

343. Kolb, K. E. et al. (2014) Antisense oligonucleotide-stimulated transcriptional pausing reveals RNA exit channel specificity of RNA polymerase and mechanistic contributions of NusA and RfaH. *J. Biol. Chem.*

344. Tomar, S. K. et al. (2013) NusG-Spt5 Proteins—Universal Tools for Transcription Modification and Communication. *Chem. Rev.*

345. Kang, J. Y. et al. (2018) RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Mol. Cell*

346. Herbert, K. M. et al. (2010) E. coli NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *J. Mol. Biol.*

347. Turtola, M. et al. (2016) NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *Elife*

348. Lawson, M. R. et al. (2018) Mechanism for the Regulated Control of Bacterial Transcription Termination by a Universal Adaptor Protein. *Mol. Cell*

349. Cheeran, A. et al. (2007) The site of action of the antiterminator protein N from the lambdoid phage H-19B. *J. Biol. Chem.*

350. Epshtein, V. et al. (2010) An allosteric mechanism of Rho-dependent transcription termination. *Nature*

351. Cheeran, A. et al. (2005) Escherichia coli RNA polymerase mutations located near the upstream edge of an RNA:DNA hybrid and the beginning of the RNA-exit channel are defective for transcription antitermination by the N protein from lambdoid phage H-19B. *J. Mol. Biol.*

352. Zhang, C. et al. (2019) Analysis of discrete local variability and structural covariance in macromolecular assemblies using Cryo-EM and focused classification. *Ultramicroscopy*

353. Greber, B. J. et al. (2016) Insertion of the Biogenesis Factor Rei1 Probes the Ribosomal Tunnel during 60S Maturation. *Cell*

354. Bradatsch, B. et al. (2007) Arx1 Functions as an Unorthodox Nuclear Export Receptor for the 60S Preribosomal Subunit. *Mol. Cell*

355. Nonato, M. C. et al. (2006) Human methionine aminopeptidase type 2 in complex with l- and d-methionine. *Bioorganic Med. Chem. Lett.*

356. Shao, S. et al. (2015) Structure and assembly pathway of the ribosome quality control complex. *Mol. Cell*

357. Knorr, A. G. et al. (2019) Ribosome–NatA architecture reveals that rRNA expansion segments coordinate N-terminal acetylation. *Nat. Struct. Mol. Biol.*

358. Zhang, Y. et al. (2014) Structural basis for interaction of a cotranslational chaperone with the eukaryotic ribosome. *Nat. Struct. Mol. Biol.*

359. Gamerdinger, M. et al. (2015) The principle of antagonism ensures protein targeting specificity at the endoplasmic reticulum. *Science .*

360. Shen, K. et al. (2019) Dual Role of Ribosome-Binding Domain of NAC as a Potent Suppressor of Protein Aggregation and Aging-Related Proteinopathies. *Mol. Cell*

361. Lein, E. S. et al. (2017) Transcriptomic Perspectives on Neocortical Structure, Development, Evolution, and Disease. *Annu. Rev. Neurosci.*

362. Silbereis, J. C. et al. (2016) The Cellular and Molecular Landscapes of the Developing Human Central Nervous System. *Neuron*

363. Cadwell, C. R. et al. (2019) Development and Arealization of the Cerebral Cortex. *Neuron*

364. Hanus, C. et al. (2013) Proteostasis in complex dendrites. *Nat. Rev. Neurosci.*

365. Holt, C. E. et al. (2013) The central dogma decentralized: New perspectives on RNA function and local translation in neurons. *Neuron*

366. Jung, H. et al. (2014) Remote control of gene function by local translation. *Cell*

367. Balchin, D. et al. (2016) In vivo aspects of protein folding and quality control. *Science .*

368. Genuth, N. R. et al. (2018) The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life. *Mol. Cell*

369. Zhang, Y. et al. (2008) Alterations in cell growth and signaling in ErbB3 binding protein-1 (Ebp1) deficient mice. *BMC Cell Biol.*

370. Horváth, B. M. et al. (2006) EBP1 regulates organ size through cell growth and proliferation in plants. *EMBO J.*

371. Li, C. et al. (2018) EBP1 nuclear accumulation negatively feeds back on FERONIA-mediated RALF1 signaling. *PLoS Biol.*

372. Somanath, P. et al. (2018) ERBB3-Binding Protein 1 (EBP1) Is a Novel Developmental Pluripotency-Associated-4 (DPPA4) Cofactor in Human Pluripotent Cells. *Stem Cells*

373. Neilson, K. M. et al. (2017) Pa2G4 is a novel Six1 co-factor that is required for neural crest and otic development. *Dev. Biol.*

374. Schneider, M. et al. (2018) Protein Tertiary Structure by Crosslinking/Mass Spectrometry. *Trends Biochem. Sci.*

375. van Zundert, G. C. P. et al. (2017) The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.*

376. Duda, D. M. et al. (2008) Structural Insights into NEDD8 Activation of Cullin-RING Ligases: Conformational Control of Conjugation. *Cell*

377. Li, T. et al. (2006) Structure of DDB1 in complex with a paramyxovirus V protein: Viral Hijack of a propeller cluster in ubiquitin ligase. *Cell*

378. Fischer, E. S. et al. (2014) Structure of the DDB1–CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature*

379. DeLucia, M. et al. (2013) HIV-2 and SIVmac accessory virulence factor Vpx down-regulates SAMHD1 enzyme catalysis prior to proteasome-dependent degradation. *J. Biol. Chem.*

380. Guenzel, C. A. et al. (2014) HIV-1 Vpr-a still “enigmatic multitasker.” *Front. Microbiol.*

381. Hanzl, A. et al. (2020) Targeted protein degradation: current and future challenges. *Curr. Opin. Chem. Biol.*

382. Schapira, M. et al. (2019) Targeted protein degradation: expanding the toolbox. *Nat. Rev. Drug Discov.*

383. Scott, D. C. et al. (2016) Two Distinct Types of E3 Ligases Work in Unison to Regulate Substrate Ubiquitylation. *Cell*

384. Kelsall, I. R. et al. (2013) TRIAD1 and HHARI bind to and are activated by distinct neddylated Cullin-RING ligase complexes. *EMBO J.*

385. Cavadini, S. et al. (2016) Cullin-RING ubiquitin E3 ligase regulation by the COP9 signalosome. *Nature*

386. He, Y. J. et al. (2006) DDB1 functions as a linker to recruit receptor WD40 proteins to CUL4-ROC1 ubiquitin ligases. *Genes Dev.*

387. Higa, L. A. et al. (2006) CUL4-DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation. *Nat. Cell Biol.*

388. Havens, C. G. et al. (2011) Mechanism of CRL4 cdt2, a PCNA-dependent E3 ubiquitin ligase. *Genes Dev.*

389. Havens, C. G. et al. (2009) Docking of a Specialized PIP Box onto Chromatin-Bound PCNA Creates a Degron for the Ubiquitin Ligase CRL4Cdt2. *Mol. Cell*

390. Prestel, A. et al. (2019) The PCNA interaction motifs revisited: thinking outside the PIP-box. *Cell. Mol. Life Sci.*

391. Nukina, K. et al. (2018) Mutations at multiple CDK phosphorylation consensus sites on Cdt2 increase the affinity of CRL4 Cdt2 for PCNA and its ubiquitination activity in S phase Genes to Cells. *Genes to Cells*

392. Havens, C. G. et al. (2009) Docking of a Specialized PIP Box onto Chromatin-Bound PCNA Creates a Degron for the Ubiquitin Ligase CRL4Cdt2. *Mol. Cell*

393. Havens, C. G. et al. (2012) Direct role for proliferating cell nuclear antigen in substrate recognition by the E3 ubiquitin ligase CRL4Cdt2. *J. Biol. Chem.*

394. Abbas, T. et al. (2010) CRL4Cdt2 regulates cell proliferation and histone gene expression by targeting PR-Set7/Set8 for degradation. *Mol. Cell*

395. Tsanov, N. et al. (2014) PIP degron proteins, substrates of CRL4Cdt2, and not PIP boxes, interfere with DNA polymerase g and i focus formation on UV damage. *Nucleic Acids Res.*

396. Petropoulos, M. et al. (2019) Replication Licensing Aberrations, Replication Stress, and Genomic Instability. *Trends Biochem. Sci.*

397. Vanderdys, V. et al. (2018) The Neddylation Inhibitor Pevonedistat (MLN4924) Suppresses and Radiosensitizes Head and Neck Squamous Carcinoma Cells and Tumors. *Mol. Cancer Ther.*

398. Kiran, S. et al. (2018) The Deubiquitinase USP46 Is Essential for Proliferation and Tumor Growth of HPV-Transformed Cancers. *Mol. Cell*

399. Huang, F. et al. (2018) CDT2-controlled cell cycle reentry regulates the pathogenesis of Alzheimer's disease. *Alzheimer's Dement.*

400. Sakaguchi, H. et al. (2012) Checkpoint Kinase ATR Phosphorylates Cdt2, a Substrate Receptor of CRL4 Ubiquitin Ligase, and Promotes the Degradation of Cdt1 following UV Irradiation. *PLoS One*

401. Clijsters, L. et al. (2014) PIP-box-mediated degradation prohibits re-accumulation of Cdc6 during S phase. *J. Cell Sci.*

402. Huh, J. et al. (2013) CRL4CDT2 Targets CHK1 for PCNA-Independent Destruction. *Mol. Cell. Biol.*

403. Tong, X. et al. (2015) CUL4-DDB1-CDT2 E3 Ligase Regulates the molecular clock activity by promoting Ubiquitination-dependent degradation of the mammalian CRY1. *PLoS One*

404. Shibutani, S. T. et al. (2008) Intrinsic Negative Cell Cycle Regulation Provided by PIP Box- and Cul4Cdt2-Mediated Destruction of E2f1 during S Phase. *Dev. Cell*

405. Bacquin, A. et al. (2013) The helicase FBH1 is tightly regulated by PCNA via CRL4(Cdt2)-mediated proteolysis in human cells. *Nucleic Acids Res.*

406. Hou, L. et al. (2019) Restriction of hepatitis B virus replication by c-Abl–induced proteasomal degradation of the viral polymerase. *Sci. Adv.*

407. Banks, D. et al. (2006) L2DTL/CDT2 and PCNA interact with p53 and regulate p53 polyubiquitination and protein stability through MDM2 and CUL4A/DDB1 complexes. *Cell Cycle*

408. Mostofa, A. G. M. et al. (2018) S-phase Specific Downregulation of Human O 6 -Methylguanine DNA Methyltransferase ( MGMT ) and its Serendipitous Interactions with PCNA and p21 cip1 Proteins in. *Neoplasia*

409. Terai, K. et al. (2013) Degradation of p12 subunit by CRL4Cdt2 E3 ligase inhibits fork progression after DNA damage. *J. Biol. Chem.*

410. Lee, M. Y. W. T. et al. (2019) Two forms of human DNA polymerase δ: Who does what and why? *DNA Repair (Amst).*

411. Galanos, P. et al. (2016) Chronic p53-independent p21 expression causes genomic instability by deregulating replication licensing. *Nat. Cell Biol.*

412. Galanos, P. et al. (2018) Mutational signatures reveal the role of RAD52 in p53-independent p21-driven genomic instability. *Genome Biol.*

413. Cui, H. et al. (2019) DTL promotes cancer progression by PDCD4 ubiquitin-dependent degradation.

414. Jo, U. et al. (2016) PCNA-Dependent Cleavage and Degradation of SDE2 Regulates Response to Replication Stress. *PLoS Genet.*

415. Tardat, M. et al. (2010) The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. *Nat. Cell Biol.*

416. Shoaib, M. et al. (2018) Histone H4K20 methylation mediated chromatin compaction threshold ensures genome integrity by limiting DNA replication licensing. *Nat. Commun.*

417. Shibata, E. et al. (2014) CRL4Cdt2 E3 ubiquitin ligase and Proliferating Cell Nuclear Antigen (PCNA) cooperate to degrade thymine DNA glycosylase in S phase. *J. Biol. Chem.*

418. Slenn, T. J. et al. (2014) Thymine DNA Glycosylase Is a CRL4 Cdt2 Substrate *.

419. Ma, X. et al. (2017) Polη O-GlcNAcylation governs genome integrity during translesion DNA synthesis. *Nat. Commun.*

420. Han, C. et al. (2015) Cdt2-mediated xpg degradation promotes gap-filling DNA synthesis in nucleotide excision repair. *Cell Cycle*

421. Zhao, Y. et al. (2003) Improving baculovirus recombination. *Nucleic Acids Res.*

422. Wilkins, M. R. et al. (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.*

423. Terai, K. et al. (2010) CRL4Cdt2 E3 Ubiquitin Ligase Monoubiquitinates PCNA to Promote Translesion DNA Synthesis. *Mol. Cell*

424. Parker, M. W. et al. (2019) A new class of disordered elements controls DNA replication through initiator self-assembly. *Elife*

425. Huber, E. M. et al. (2012) Immuno- and constitutive proteasome crystal structures reveal differences in substrate and inhibitor specificity. *Cell*

426. Tang, W. K. et al. (2010) A novel ATP-dependent conformation in p97 N-D1 fragment revealed by crystal structures of disease-related mutants. *EMBO J.*

427. De Biasio, A. et al. (2015) Structure of p15PAF-PCNA complex and implications for clamp sliding during DNA replication and repair. *Nat. Commun.*

428. Mayrose, D. R. et al. (1996) Purification and crystallization of cyclin-dependent kinase inhibitor p21. *Protein Sci.*

429. Hayashi, A. et al. (2014) PCNA-Dependent Ubiquitination of Cdt1 and p21 in Mammalian Cells.

430. Hayashi, A. et al. (2018) Direct binding of Cdt2 to PCNA is important for targeting the CRL4 Cdt2 E3 ligase activity to Cdt1.

431. Shibata, E. et al. (2011) Selective Ubiquitylation of p21 and Cdt1 by UBCH8 and UBE2G Ubiquitin-Conjugating Enzymes via the CRL4Cdt2 Ubiquitin Ligase Complex. *Mol. Cell. Biol.*

432. Scott, M. T. et al. (2000) Reversible phosphorylation at the C-terminal regulatory domain of p21(Waf1/Cip1) modulates proliferating cell nuclear antigen binding. *J. Biol. Chem.*

433. Ellison, V. et al. (1998) Reconstitution of recombinant human replication factor C (RFC) and identification of an RFC subcomplex possessing DNA-dependent ATPase activity. *J. Biol. Chem.*

434. Leng, F. et al. (2018) Proliferating cell nuclear antigen interacts with the CRL4 ubiquitin ligase subunit CDT2 in DNA synthesis–induced degradation of CDT1. *J. Biol. Chem.*

435. McGuffin, L. J. et al. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*

436. Jones, D. T. et al. (2003) in *Proteins: Structure, Function and Genetics*

437. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*

438. Biasini, M. et al. (2014) SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*

439. Mazian, M. A. et al. (2019) A DNA-binding domain in the C-terminal region of ubiquitin ligase activity for Cdt1.

440. Fujisawa, R. et al. (2017) Human CTF18-RFC clamp-loader complexed with non-synthesising DNA polymerase ϵ efficiently loads the PCNA sliding clamp. *Nucleic Acids Res.*

441. Lancey, C. et al. (2020) Structure of the processive human Pol δ holoenzyme. *Nat. Commun.*

442. Brown, N. R. et al. (1999) The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases.

443. O'Reilly, F. J. et al. (2020) In-cell architecture of an actively transcribing-translating expressome. *bioRxiv*

444. Hesketh, G. G. et al. (2017) in *Methods in Molecular Biology*

# 15 Publications

## 15.1 Publication I

Said N*, Krupp F, Anedchenko E, Santos KF, Dybkov O, Huang Y-H, Lee C-T, Loll B, Behrmann E, Bürger J, Mielke T, Loerke J, Urlaub H, Spahn CMT, Weber G, Wahl, MC (2017) Structural basis for λN-dependent processive transcription antitermination, *Nature Microbiology* **2**, 17062 (2017)

https://doi.org/10.1038/nmicrobiol.2017.62

* denotes first author

# Structural basis for λN-dependent processive transcription antitermination

Nelly Said[1], Ferdinand Krupp[2,†], Ekaterina Anedchenko[1,†,‡], Karine F. Santos[1,#], Olexandr Dybkov[3], Yong-Heng Huang[1], Chung-Tien Lee[4,6], Bernhard Loll[1], Elmar Behrmann[2,‖], Jörg Bürger[2,5], Thorsten Mielke[5], Justus Loerke[2], Henning Urlaub[4,6], Christian M. T. Spahn[2], Gert Weber[1,§], Markus C. Wahl[1,7,*]

[1] Freie Universität Berlin, Laboratory of Structural Biochemistry, Takustraße 6, D-14195 Berlin, Germany

[2] Medizinische Physik und Biophysik, Charité – Universitätsmedizin Berlin, Charitéplatz 1, D-10117 Berlin, Germany

[3] Max Planck Institut für biophysikalische Chemie, Department of Cellular Biochemistry, Am Fassberg 11, D-37077 Göttingen, Germany

[4] Max Planck Institut für biophysikalische Chemie, Bioanalytical Mass Spectrometry, Am Fassberg 11, D-37077 Göttingen, Germany

[5] Max-Planck-Institut für Molekulare Genetik, Microscopy and Cryo-Electron Microscopy Group, Ihnestraße 63-73, D14195 Berlin, Germany

[6] Universitätsmedizin Göttingen, Institut für Klinische Chemie, Bioanalytik, Robert-Koch-Straße 40, D-35075 Göttingen, Germany

[7] Helmholtz-Zentrum Berlin für Materialien und Energie, Macromolecular Crystallography, Albert-Einstein-Straße 15, D-12489 Berlin, Germany

[†] These authors contributed equally to this work.

[‡] Present address: Humboldt-Universität zu Berlin, Molecular Cell Biology, Philippstr. 13, D-10099 Berlin, Germany

[#] Present address: moloX GmbH, Takustraße 6, D-14195 Berlin, Germany

[‖] Present address: Research Group Structural Dynamics of Proteins, Center of Advanced European Studies and Research (Caesar), Ludwig-Erhard-Allee 2, D-53175 Bonn, Germany

[§] Present address: Ernst-Moritz-Arndt-Universität Greifswald, Molekulare Strukturbiologie, Felix-Hausdorff-Str. 4, D-17487 Greifswald, Germany

[*] Correspondence to: mwahl@zedat.fu-berlin.de

## Abstract

**λN-mediated processive antitermination constitutes a paradigmatic transcription regulatory event, during which phage protein λN, host factors NusA, NusB, NusE and NusG and an RNA *nut* site render elongating RNA polymerase termination-resistant. The structural basis of the process has so far remained elusive. Here, we describe a crystal structure of a λN-NusA-NusB-NusE-*nut* site complex and an electron cryo-microscopic structure of a complete transcription antitermination complex, comprising RNA polymerase, DNA, *nut* site RNA, all Nus factors and λN, validated by cross-linking/mass spectrometry. Due to intrinsic disorder, λN can act as a multi-protein/RNA interaction hub, which together with *nut* site RNA arranges NusA, NusB and NusE into a triangular complex. This complex docks *via* the NusA N-terminal domain and the λN C-terminus next to the RNA exit channel on RNA polymerase. Based on the structures, comparative cross-linking analyses and structure-guided mutagenesis, we hypothesize that λN mounts a multi-pronged strategy to reprogram the transcriptional machinery, which may include (i) the λN C-terminus clamping the RNA exit channel, thus stabilizing the DNA:RNA hybrid; (ii) repositioning of NusA and RNAP elements, thus redirecting nascent RNA and sequestering the upstream branch of a terminator hairpin; and (iii) hindering RNA engagement of termination factor ρ and/or obstructing ρ translocation on the transcript.**


In all free-living organisms, transcription is carried out by multi-subunit RNA polymerases (RNAPs). Bacteria harbor a single RNAP that receives regulatory inputs from the substrate DNA, nascent RNA and protein transcription factors[1]. The *Escherichia coli* RNAP core enzyme has a $\alpha_2\beta\beta'\omega$ subunit composition and associates with a σ factor to form a holoenzyme that can initiate transcription at promoters. After promoter escape, σ is replaced by elongation factors NusA and NusG[2], which together with core RNAP form a stable transcription elongation complex (TEC). During elongation, *E. coli* NusA enhances RNAP

pausing at specific sites[3], while NusG increases RNA chain elongation rate[4]. Transcription can be terminated *via* an intrinsic mechanism, elicited by a stable stem-loop structure followed by a stretch of uridine residues in the nascent RNA, which leads to conformational changes in RNAP, destabilization of the DNA:RNA hybrid and release of the transcript[5]. Alternatively, the ρ factor can terminate transcription by engaging nascent RNA *via* ρ-utilization (*rut*) sequences, translocating on the transcript in 5'-to-3' direction and, upon encounter of RNAP, extracting the transcript[4,6]. NusA facilitates intrinsic termination[5] and can support or counteract ρ-dependent termination depending on the context[6]. NusG can directly contact ρ *via* its C-terminal domain (CTD) and supports ρ-dependent termination[4].

During lytic growth, lambdoid phages switch from immediate-early to delayed-early gene expression by processive transcription antitermination[5]. Upon transcription of N-utilization (*nut*) sequences on the λ genome, phage protein λN together with host N-utilization substances (Nus) A, B, E (equivalent to ribosomal [r] protein S10) and G assembles an RNA-protein complex (RNP) on RNAP, which enables the enzyme to read through intrinsic and ρ-dependent terminators[5]. A similar mechanism, with several r-proteins replacing λN, is at work during transcription of ribosomal RNA (*rrn*) genes[7]. The overall organization of the λN-based transcription antitermination complex (TAC) is unknown. Here, we used X-ray crystallography, single-particle electron cryo-microscopy (cryo-EM), chemical cross-linking/mass spectrometry (CX-MS) and structure-guided mutational analyses to elucidate the functional organization of an intact λN-TAC. Our results suggest that λN binds and reorganizes RNAP elements and repositions NusA on RNAP, thereby redirecting nascent RNA. *Via* these activities, it apparently mounts a multi-pronged approach to prevent termination, which might involve clamping of the RNAP RNA exit channel, sequestering the upstream branch of a terminator hairpin and sterically interfering with ρ action.

## Results

**Crystal structure of a λN-NusA-NusB-NusE-*nut* RNP.** We recombinantly produced λN and all *Escherichia coli* Nus factors and generated a 36-nucleotide (nt) *nut* RNA, containing a linear *boxA* element and a *boxB* hairpin, by *in vitro* transcription (Fig. 1a). While a complex containing λN, NusA or NusA$^{\Delta AR2}$ (a variant lacking 70 C-terminal residues that comprise the second of two C-terminal acidic repeats [AR]), NusB, NusE and *nut* RNA could be reconstituted, NusG failed to stably integrate (Fig. 2a,b). We determined the crystal structure of the λN-NusA$^{\Delta AR2}$-NusB-NusE-*nut* RNP at 4.0 Å resolution. Residues lacking electron density (85-107 of λN and the first five residues of *nut* RNA) guided the design of a slightly further truncated complex, which yielded a crystal structure at 3.35 Å resolution (Fig. 1b; Supplementary Table 1).

The λN$^{1-84}$-NusA$^{\Delta AR2}$-NusB-NusE-*nut* RNP resembles a triskelion (Fig. 1b). NusA$^{\Delta AR2}$ comprises an N-terminal domain (NTD) followed by S1, hnRNP K-homology (KH) 1, KH2 and AR1 domains, and forms an arch-like structure along two arms of the triskelion, with KH1 positioned at the center. NusA NTD and S1 as well as KH2 and AR1 are linked *via* long helices, while S1, KH1 and KH2 are connected by short peptides and large interfaces, as in the isolated protein[8]. The third arm of the triskelion is dominated by the NusB-NusE heterodimer, which resembles the structure of the isolated sub-complex[9,10]. *Nut* RNA adopts an elongated conformation with *boxA* extending across NusB-NusE (Fig. 1c), the *boxA*-*boxB* spacer running along NusA KH1 and the *boxB* hairpin neighboring NusA KH2 (Fig. 1d), in agreement with the previous implication of the KH domains in RNA binding[11]. Although NusA S1 is a *bona fide* RNA-binding domain, it does not contact *nut* RNA.

λN is intrinsically disordered in isolation[12]. In the complex, its N-terminal 20 residues form a kinked helix, one face of which binds along the major groove and the loop of *boxB*, similar to the isolated sub-complex[13,14], while the other flank is facing NusA KH2 (Fig. 1b, region I). The five-nt *boxB* loop adopts a GNRA tetraloop-like conformation with Gua25 extruded. Gua25 is stacked between R8 of λN and I318 of NusA, presenting it for hydrogen bonding to

the backbone of the neighboring GXXG motif of NusA KH2 (Fig. 1e) and thus explaining why mutations in Gua25 abolish NusA binding to a λN-*nut* complex[15]. The following five residues of λN further encircle NusA KH2, expanding the domain's β-sheet (Fig. 1b, region II). The irregularly structured $\lambda N^{21-46}$ element extends across the NusA KH1 and KH2 surfaces that face away from *nut* RNA (Fig. 1b, region III). $\lambda N^{47-52}$ folds as a short helix that rests in a cavity formed by the NusA NTD-S1 connector helix, NusA S1, KH1 and the globular portion of NusE ("NusA-NusE cavity"; Fig. 1b, region IV). The following irregular linker leads to a long, amphipathic helix (residues 57-79), that forms a three-helix bundle with the N-terminal helix of NusA and the NTD-S1 connector helix (Fig. 1b, region V). $\lambda N^{34-47}$ has previously been shown to bind NusA AR1[12,16,17] but is remote from AR1 in the present complex, consistent with the $\lambda N^{34-47}$-AR1 interaction being dispensable for antitermination[16,18] and possibly only formed transiently during complex assembly (Supplementary Discussion). CX-MS of a λN-$NusA^{\Delta AR2}$-NusB-NusE-*nut* RNP indicated a very similar overall structure of the complex in solution as in the crystal (Fig. 1b, left; Supplementary Table 2). Cross-links at variance with the structure are explained by the presence of a minor complex based on the alternative $\lambda N^{34-47}$-NusA AR1 interaction in solution[12,16,17].

**Structure-guided interaction studies.** Based on the structure, we probed the stability of individual interfaces and the importance of contact-mediating residues for RNP formation by analytical size-exclusion chromatography (Fig. 2). In the absence of *nut* RNA, only binary λN-NusA and NusB-NusE interactions prevailed (Fig. 2c), showing that the RNA is required to inter-connect the two protein complexes (Fig. 1b, right). *Nut* RNA bound stably to λN, NusB-NusE or $NusA^{\Delta AR2}$ but not to full-length NusA (Fig. 2d-g), consistent with AR2 auto-inhibiting NusA-RNA interactions by folding back onto the S1-KH1-KH2 region[19]. NusA joined a λN-*nut* RNP (Fig. 2h), at least in part *via* NusA-λN contacts, as NusA did not bind a NusB-NusE-*nut* RNP (Fig. 2i). NusB-NusE did not stably associate with $NusA^{\Delta AR2}$-*nut* RNP (Fig. 2f) or λN-*nut* RNP (Fig. 2j). Besides the *boxA*-binding site, a NusB-NusE-RNA structure revealed an unexpected binding site for *boxB*-like elements[10] (Fig. 1f), which is not maintained in the

142 presence of NusA$^{\Delta AR2}$ and λN (Fig. 1c). Thus, NusA$^{\Delta AR2}$ or λN sequester *boxB*, weakening
143 the NusB-NusE-*nut* RNP and rendering NusB-NusE entry dependent on both λN and NusA.
144 NusE alone assembled with NusA and λN on *nut* RNA (Fig. 2k), while NusB in the absence
145 of NusE failed to do so (Fig. 2l), indicating that NusE is required for stable integration of
146 NusB. The S1-KH1-KH2 region of NusA was not sufficient for formation of an intact RNP
147 (Fig. 2m), presumably due to the lack of the NusA NTD-S1 connector helix and consequent
148 destabilization of the NusA-NusE cavity. Residues 73-107 of λN, which contact NusA NTD,
149 were dispensable for RNP formation (Fig. 2n). In contrast, while still binding NusA and *nut*
150 RNA, a λN fragment truncated after residue 47, which lacks the region binding to the NusA-
151 NusE cavity, was defective in higher-order RNP formation (Fig. 2o). Alanine exchanges of
152 NusA E212, E218 and R255, which mediate binary NusA-NusE contacts, did not destabilize
153 the complex. In contrast, an alanine exchange of R258, which is involved in a NusA-NusE-
154 *boxA*-*boxB* spacer triple interaction (Fig. 1h), abrogated complex formation (Fig. 2p). These
155 results expand previous interaction studies[15,16,20], showing that while complex stability is
156 considerably buffered with respect to the loss of individual binary contacts, it critically
157 depends on residues (such as NusA R258) and regions (such as λN$^{47-52}$ or the NusA-NusE
158 cavity) that mediate interactions between multiple other subunits.
159
160 **Structural mechanisms underlying the effects of subunit variants.** Our structure
161 rationalizes λN-related defects associated with numerous previously studied variants of
162 complex components (Fig. 3). A7D, V8A, V8E, A11D, V12D and L31E exchanges in NusA
163 show defects in the propagation of lamboid phages and/or in λN-mediated antitermination[18].
164 Residues A7, V8, A11 and V12 of NusA form the surface of the N-terminal helix of NusA that
165 is involved in binding the C-terminal helix of λN (Fig. 3b), while L31 is part of the hydrophobic
166 core and thus crucial for the fold stability of the NusA NTD. NusA$^{R199A}$ abrogates λN-
167 mediated antitermination and reduces NusA interaction with a λN-*nut* RNP[19,21], in agreement
168 with R199 stabilizing the S1-KH1 arrangement that bridges between *nut* RNA and λN regions
192
169 (Fig. 3c). Thus, in contrast to previous expectation[21], NusA$^{R199A}$ does not abrogate direct *nut*

RNA contacts. NusA$^{L183R}$ (encoded by the *nusA1* allele) blocks[22,23] while NusA$^{R104H/E212K}$ (*nusA10*) reduces λN-dependent antitermination based on the E212K exchange[24]. L183 in S1 interacts with the upper part of the NTD-S1 connector helix (Fig. 3d) and E212 in NusA KH1 forms part of the NusA-NusE interface (Fig. 3e). Thus, both residues help to shape the NusA-NusE cavity that accommodates λN$^{47-52}$. In NusA944, four residues (153-156) in a loop of the S1 domain are replaced by nine residues found at this position in *Salmonella typhimurium* NusA[25]. In line with NusA944 and *S. typhimurium* NusA still binding a λN-*nut* RNP but failing to support antitermination[21,25], our structure reveals that the insertion will alter NusA-λN contacts (Fig. 3f), but again does not directly affect NusA-*nut* RNA interactions. NusA$^{G253D}$ fails to bind a λN-*nut* RNP or to support λN-dependent antitermination[11], as explained by G253 being part of the KH1 GXXG motif that binds the *boxA*-*boxB* spacer, and R258 following this motif engaging in interactions with NusE (Fig. 1h). Consistent with NusA$^{G319D}$ likewise abrogating NusA interaction with a λN-*nut* RNP and exhibiting a mild defect in supporting λN action[11], G319 belongs to the KH2 GXXG motif involved in binding the *boxB* loop (Fig. 1e). NusE$^{A86D}$ (*nusE71*) blocks λN-dependent antitermination[26], reconciled by A86 being located in the NusE-λN interface, where an aspartate would be forced into unfavorable interactions with the neighboring L52 of λN (Fig. 3g). NusB$^{D118N}$ (*nusB101*) exhibits enhanced affinity to *boxA*[9] and suppresses *nusA1* and *nusE71* defects[27]. These findings can be rationalized by D118 neighboring the negatively charged sugar-phosphate backbone (Fig. 3h), which will decrease the NusB-*boxA* affinity compared to an asparagine at this position. K45R, S50R and I55M in λN (*pun134*, *pun150* and *pun165*, respectively) suppress *nusA1* defects[28,29]. λN$^{K45R}$ likely enhances contacts to NusA E132, λN$^{S50R}$ presumably introduces additional salt bridges to NusA E132 and/or E136 and λN$^{I55M}$ might sandwich more stably between the NTD-S1 connector helix and the S1 domain (Fig. 3i). Finally, consistent with the identity of the nt at *boxA* position 9 modulating λN-mediated antitermination[30,31], Ade9 in our structures mediates the only nucleobase contact of *nut* RNA to NusE (Fig. 1h). Taken together, the effects of these variants in light of the present structure emphasize the importance of a precise relative orientation of the NusA S1-KH1-

KH2 domains, around which *nut* RNA, λN and NusE are organized, and point to the NusA-NusE-λN interaction region as a crucial node in the complex.

**Structure of a λN-based TAC.** To further investigate the structural basis of λN-mediated antitermination, we assembled a complete TAC comprising RNAP, double-stranded DNA bearing a transcription bubble, *nut* RNA that can hybridize with the template strand in the transcription bubble, all Nus factors and λN (Fig. 4a,b). We determined a cryo-EM map of the λN-TAC at 9.8 Å resolution, in which all subunits could be located (Fig. 4c,d; Supplementary Fig. 1). CX-MS analysis of the TAC supported the structural assignment (Fig. 4e; Supplementary Table 3). In the TAC, the λN-NusA-NusB-NusE-*nut* RNP towers above the RNA exit channel formed by the β flap domain (including the flexible ß flap tip helix; FTH), C-terminal residues of β, the β' dock domain and the β' Zinc finger (ZnF; Fig. 4d-f). The NusG NTD is located next to the RNAP primary channel, contacting the β' clamp helices and spanning one wall of the upstream DNA-binding site, in agreement with the location of the NusG NTD in a recent RNAP-NusG co-crystal structure[32] and of the homologous Spt5 NTD on eukaryotic RNA polymerase II[33]; however, NusG and Spt5 NTDs have been modeled in different orientations in these previous studies, which cannot be distinguished at the present resolution of the λN-TAC. The NusG C-terminal domain (CTD) is detached from the site seen in an RNAP-NusG crystal structure[32]. Instead, it is bound to NusE as previously characterized in isolation[34] and also contacts NusA S1-KH1, stabilizing the λN-binding NusA-NusE cavity (Fig. 4d). In this arrangement, additional contacts of $NusG^{S163F}$ (*nusG4*) to NusA could ensue and counteract *nusA1*- and *nusE71*-mediated destabilization of the NusA-NusE-λN interaction network[35].

The above analyses suggest strategies, by which λN modulates termination. First, as indicated by λN cross-links to β flap residues 844, 937 and 1041, to β flap tip residues 890 and 914, to β residues 496 and 503 (preceding the β pincer), to β residue 1242 (neighboring the β' rudder) and to β' dock residues 371, 378, 395 and 399 (Supplementary Table 3), the λN C-terminus meanders around the rim of the RNA exit channel (Fig. 4f) and clamps RNAP

elements that bind the DNA:RNA hybrid and are crucial for the stability of elongating complexes[36]. RNAP clamping by the λN C-terminus would reinforce hybrid stability and hinder hairpin-induced conformational changes in RNAP, explaining how the protein could interfere with termination-related pausing, RNA slippage and hybrid melting[37,38]. In the same manner, λN could counteract ρ-mediated destabilization of the hybrid. In agreement with the path of the λN C-terminus, mutations in the β flap, β pincer, β' lid and β' rudder affect λN-dependent antitermination[39-42], and hydroxyl radicals generated from a modified C-terminus of the phage H-19B N protein target the RNAP active site cavity[43]. Furthermore, the same contact areas are also used by σ70 region 3.2 to access the interior of RNAP[44], explaining binding competition between σ70 and λN[12].

Second, in a transcription elongation complex (TEC) lacking λN, NusA NTD and S1 domains are thought to reside next to and guide nascent RNA along the β' dock domain[45,46] (Fig. 5a). Conversely, in the TAC, the NusA NTD bound to the C-terminal λN helix is moved towards the center of the β flap domain, placing the NusA S1 domain next to the β' ZnF (Fig. 4e and 5b). The β flap tip maintains interactions with the NusA NTD and is thus displaced from the β' ZnF and β' dock domain, consistent with the suggestion that λN leads to altered NusA NTD-β flap interactions and a rearrangement of the β flap and β' dock domains[39]. This rearrangement could allow nascent RNA 15-33 nts upstream of the 3'-end to run between the λN C-terminus, the β flap tip and the β' ZnF towards the NusA S1 domain (Fig. 4a and 5b). In striking agreement with this organization, cross-linking analyses have shown that λN directly contacts this RNA region and reinforces its interactions with the RNAP upstream RNA-binding site (UBS, composed of the β' ZnF, C-terminal residues of β and αCTD) and NusA S1[37]. The upstream branch of a terminator hairpin at the time of termination could lie in the RNA region guided along NusA S1 and thus remote from the complementary downstream branch just emerging from the exit channel. Thus, our analysis provides a structural explanation for the suggested sequestration of the upstream terminator branch by λN, RNAP UBS and NusA S1 to prevent termination[37] (Fig. 5b).

**Structure-guided functional analyses.** To test the suggested λN-mediated repositioning of NusA, we analyzed a TEC lacking λN by CX-MS (Supplementary Table 4). In agreement with our model, cross-links of the NusA NTD and S1 domain to the β flap, β' ZnF, β' dock and UBS regions were strongly altered in the absence of λN. Only in the absence of λN, the NusA S1 domain cross-linked to the β' dock domain and the C-terminus of β (Fig. 4e), fully consistent with its presumed positioning next to these elements in the unmodified TEC. We also determined a crystal structure of a large NusA fragment ($NusA^{100-426}$) in isolation (Supplementary Table 1), revealing a flexible relative orientation between the NTD and the remainder of NusA (Fig. 5b, inset). Thus, λN repositions NusA on RNAP, possibly capitalizing on an intrinsic flexibility of NusA.

To further test the suggested antitermination mechanisms, we conducted *in vitro* transcription assays with rationally engineered components (Fig. 6a,b). According to our model, $λN^{1-84}$ is expected to still maintain contacts to the core of the β flap domain, reposition NusA NTD and thus support sequestration of the upstream terminator branch, but lack the ability to suppress pausing and hybrid melting by RNAP clamping. The fragment retained 83 % of the antitermination efficiency of full-length λN (Fig. 6c, lanes 2 and 3), in good agreement with the estimated about 10 % of the antitermination activity of λN in the presence of NusA relying on pause suppression[37]. $λN^{1-72}$ lacks all RNAP-contacting residues of λN. While the fragment still assembled with Nus factors and *nut* RNA on RNAP, the resulting complex lacked antitermination activity (Fig. 6c, lanes 2 and 4). Removal of the N-terminal helix of NusA ($NusA^{17-426}$) or further deletions up to the NTD-S1 connector helix ($NusA^{41-426}$ and $NusA^{136-426}$) should leave RNAP clamping by the λN C-terminus unaffected but partly interfere with sequestration of the upstream terminator branch by hampering rearrangements of the RNAP β flap tip and proper positioning of NusA S1. Consistently, these variants exhibited ca. 85 % of the antitermination efficiency of full-length NusA (Fig. 6d).

## Discussion

Here, we elucidated the structure of a λN$^{1-84}$-NusA$^{\Delta AR2}$-NusB-NusE-*nut* RNP and the global architecture of a complete λN-TAC. Interaction studies, mapping of subunit variants, CX-MS and cryo-EM analyses are in full agreement with the crystal structure of the isolated RNP, indicating that the latter was not influenced by truncations of λN and NusA. The structures together with comparative CX-MS analyses of a λN-TAC and a TEC lacking λN suggest molecular mechanisms underlying λN-mediated antitermination, which are consistent with our structured-guided mutational analyses. We propose that λN in conjunction with Nus factors and *nut* RNA launches a multi-tiered strategy to suppress termination.

**Basal antitermination activity of λN and support by Nus factors and *nut* RNA.** As shown by our crosslinking data, the λN C-terminus binds elements of the RNAP RNA exit channel, likely stabilizing their relative organization and counteracting hybrid melting due to terminator hairpin formation or due to the action of termination factor ρ. Additionally, the λN C-terminus may reposition RNAP elements and thereby redirect nascent RNA, possibly creating a steric barrier for terminator hairpin formation. Based on our structural results, λN could exert these functions also in the absence of Nus factors and *nut* RNA, in agreement with λN alone exhibiting basal antitermination activity[37,47]. Our results also suggest how λN-based antitermination is rendered processive by the Nus factors and *nut* RNA[5]. Nus factors and *nut* RNA almost certainly stabilize and possibly orient λN on RNAP, as indicated by additional NusA-RNAP interactions and by the numerous mutual contacts within the λN-NusA-NusB-NusE-*nut* RNP.

RNAP binding by λN and its cooperation with Nus factors and *nut* RNA hinges on λN being intrinsically unstructured, as a consequence of which essentially all of its side chains are surface-exposed and the protein is conformationally malleable. Despite its relatively small size, λN can thus engage *via* numerous, densely arrayed, short epitopes in interactions with *boxB*, multiple domains of NusA, NusE, various parts of RNAP and nascent RNA at the

same time. Additionally, intrinsic disorder may allow overlapping regions of λN to adapt to different interaction platforms, such as residues 34-52 to NusA AR1 or the NusA-NusE cavity, during TAC assembly (Supplementary Discussion).

**Reprogramming of NusA into an antitermination factor.** A low-resolution EM analysis of a RNAP-NusA complex[3], mapping of the NusA NTD on RNAP by NMR[48] and biochemical data[45,46] indicate that in a TEC lacking λN, the NusA NTD resides next to the β flap tip and β' ZnF, with the S1 and KH domains extending towards and beyond the β' dock domain. In this situation, nascent RNA emerging from the RNA exit channel is suggested to run between the RNAP β' dock domain on one side and the NusA NTD, S1 and KH domains on the other[45,46]. NusA may thus display the upstream branch of a terminator hairpin in close vicinity to the downstream branch emerging from the exit channel (Fig. 5a), consistent with the suggested direct stabilization by NusA of RNA duplexes formed in the exit channel[49]. The NusA NTD contributes most of this termination-promoting activity of NusA, while the remaining domains seem to predominantly enhance NusA association with the TEC[46]. Clearly, full understanding of how NusA functions in a TEC will require elucidation of a high-resolution TEC structure and structure-guided mutational analyses.

Our cryo-EM structure of a λN-TAC and our comparative CX-MS analyses indicate that λN repositions NusA and the β flap tip on RNAP, possibly allowing nascent RNA to thread along the NusA NTD and S1 domain towards the *boxB* element of the *nut* site, the latter being suspended across the NusA KH domains and the NusB-NusE heterodimer remote from the RNA exit (Fig. 5b). The upstream branch of a terminator hairpin could then be bound on a surface of the λN-NusA-NusB-NusE-*nut* RNP, remote from the downstream branch when it just emerges from the exit channel (Fig. 5b). Our structural results suggest that the NusA S1 domain may sequester the upper terminator branch, but it could be supported by other elements such as the NusA NTD. Additionally or alternatively, repositioned RNAP and/or NusA elements around the redirected nascent RNA may sterically prevent terminator hairpin formation. In the future, these hypotheses have to be further scrutinized, for example by

devising a higher-resolution structure of a λN-TAC that would allow precisely targeted mutational studies.

While *nut* RNA has previously been proposed to merely facilitate recruitment of λN[37], our findings suggest that it also aids λN in redirecting NusA functions. In the λN-TAC, the N-terminal 20 residues of λN are sandwiched between NusA KH2 and bound *boxB*, and the following λN residues reorganize the NusA KH2 GXXG motif and occupy an RNA-binding platform neighboring the KH2 β-sheet (Fig. 1d,e), thereby preventing continuous RNA binding along an extended KH1-KH2 surface as seen in an isolated NusA-RNA complex[50] (Fig. 1g). Thus, occupation of RNA binding sites on NusA KH1 and KH2 by λN and *nut* RNA, further supported by NusB-NusE binding *boxA*, will prevent interaction of the NusA domains with alternative nascent RNA sequences and might modulate NusA function in certain scenarios. For example, NusA alone decreases termination by a terminator hairpin bearing an elongated loop, which was previously explained by postulating two NusA molecules per RNAP[37]. However, our analyses suggest that in the absence of λN, NusA KH1 and KH2 would not be occupied by *nut* RNA and be available for sequestering the upstream branch of a long-loop terminator distal to the exit channel.

**Possible mechanisms for inhibition of ρ-dependent termination.** Our results also suggest how the Nus factors and *nut* RNP may support λN in opposing termination factor ρ. ρ action *via* a *rut* site upstream of a *nut* site could be prevented if the λN-NusA-NusB-NusE-*nut* RNP posed a stable roadblock for ρ translocating towards RNAP. A *rut* site either up- or downstream but in close proximity to a *nut* site could also be sequestered *via* surfaces of the λN-NusA-NusB-NusE-*nut* RNP. In addition, the bulky λN-NusA-NusB-NusE-*nut* RNP may sterically hinder ρ engagement of a *rut* site close to a *nut* site. Furthermore, consistent with the observation that ρ associates with RNAP early during transcription[2], ρ might be recruited directly by the NusG CTD to RNAP[4] before being handed off to an emerging *rut* site. RNA downstream of the *nut* site is expected to loop out on the concave side of NusA in the λN-TAC (Fig. 2d). A downstream *rut* site would thus be guided away from NusG-bound ρ by the

intervening λN-NusA-NusB-NusE-*nut* RNP (Fig. 2d), disfavoring ρ engagement of the transcript. Structures of TECs and TACs stalled in the presence of ρ will be required to eventually elucidate the mode of ρ action and how it is prevented during processive antitermination.

## Methods

**Protein production.** DNA fragments encoding NusA and NusG were generated by PCR from *E. coli* chromosomal DNA and cloned into the pETM-11 vectors (European Molecular Biology Laboratories) under the control of a T7 promotor *via Bsa*I and *Xho*I restriction sites to produce proteins with a TEV-cleavable N-terminal $His_6$-tag. A DNA fragment encoding λN was produced by assembly-PCR and cloned into the pGEX-6P-1 vector using *Bam*HI and *Xho*I restriction sites to produce a protein with a PreScission-cleavable N-terminal GST-tag. NusB-NusE[9] and RNAP[51] were produced based on previously described plasmids. Mutations were introduced by site-directed mutagenesis using the QuikChange protocol (Stratagene). All constructs were verified by DNA sequencing (Seqlab).

Plasmids containing the genes of interest were transformed into *E. coli* Rosetta2 (DE3) cells. The cells were grown in auto-inducing medium[52] to an $OD_{600}$ of 0.3 at 37 °C and incubated for an additional 16 hours at 18 °C. Cells were harvested at 4 °C by centrifugation, resuspended, lysed by sonication using a Sonoplus ultrasonic homogenizer HD3100 (Bandelin) and centrifuged at 4 °C and 21,500 rpm for 1 h. NusA, NusB, NusE, NusG, RNAP and σ70 proteins and variants were purified as described previously[9,34,53,54]. GST-λN in 50 mM Tris-HCl, pH 7.6, 500 mM NaCl, 10 mM EDTA, 10 % (v/v) glycerol, 10 mM DTT was captured on Gluthathione Sepharose beads (GE Healthcare), washed, eluted with 20 mM reduced glutahione (Sigma-Aldrich) supplemented with PreScission protease and dialyzed against 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 0.1 mM EDTA, 1 mM DTT. The protein solution was diluted to 50 mM NaCl and applied to a MonoS cation exchange column (GE Healthcare) in 20 mM Tris-HCl, pH 7.6, 50 mM NaCl, 0.1 mM EDTA, 5 % (v/v) glycerol, 1 mM

DTT and eluted with a gradient to 1 M NaCl. Pooled fractions were concentrated and further separated *via* a Superdex 75 size-exclusion column (GE Healthcare) in 10 mM Tris-HCl, pH 7.6, 200 mM NaCl, 1 mM DTT. Peak fractions were pooled, concentrated and stored at -80°C.

**RNA production.** The shortened *nut* site RNA used to obtain optimized crystals was chemically synthesized (Dharmacon). All other RNAs were produced by T7 RNAP-based *in vitro* transcription and purified *via* a non-denaturing method using strong anion-exchange chromatography and size-exclusion chromatography[55,56]. To this end, DNA templates containing a T7 promoter and downstream region of interest were produced by assembly-PCR and cloned into the pUC18 vector (Thermo Scientific) *via Xba*I and *Kpn*I restriction sites. For *in vitro* transcription, a PCR product was generated using reverse oligos that harbored two terminal 2'-O-methylated nucleotides to reduce 3'-end heterogeneity[57,58]. Transcription reactions were performed in 120 mM HEPES-NaOH, pH 7.5, 3 mM rNTPs (ATP, GTP, UTP, CTP), 40 mM DTT, 4.8 mM spermidine, 0.125 mg/ml acetylated BSA, 16 mM $MgCl_2$, 0.04µl/µl reaction PPase, 10 ng/µl PCR product, 0.15 µg/µl T7 RNAP at 37 °C overnight. The reactions were treated with DNaseI, loaded on a MonoQ anion-exchange column (GE Healthcare) in 50 mM HEPES-NaOH, pH 6.9, 0.2 mM EDTA and eluted in a gradient to 2 M NaCl. Pooled fractions were further purified *via* Superdex 75 size-exclusion chromatography in 10 mM HEPES-NaOH, pH 7.5, 50 mM NaCl. Peak fractions were pooled, concentrated and stored at -80°C.

**Interaction assays.** Interactions were studied using analytical size-exclusion chromatography. Stock solutions of proteins and/or nucleic acids were combined to generate equimolar mixtures (20 µM final concentration for each component) in 10 mM HEPES-NaOH, pH 7.5, 50 mM NaCl, 1 mM DTT and incubated for 15 min at room temperature. 50 µl of the mixtures were loaded on a Superdex 200 Increase 3.2/300 column (GE Healthcare) and chromatographed at a flow rate of 50 µl/min in 10 mM HEPES-NaOH, pH 7.5, 50 mM NaCl, 1

mM DTT at 4 °C. 50 µl fractions were collected and analyzed by SDS PAGE (17.5 % separating gel) and urea PAGE (15 %, 8 M urea) to reveal protein and nucleic acid contents, respectively.

**Transcription assays.** A linear DNA fragment, containing the $\lambda P_L$ promoter, $\lambda nut_L$, $\lambda t_{R'}$ intrinsic terminator and additional vector-derived sequences to distinguish between products of termination and read-through, was generated by PCR based on pKC30[59]. In control reactions we employed a DNA fragment, in which the $\lambda nut_L$ element was deleted. *In vitro* transcription was assayed in single-round format as described[60]. Briefly, 100 nM *E. coli* core RNAP, 100 nM σ70 factor and 10 nM template DNA were mixed in a 12.5 µl reaction in transcription buffer (20 mM Tris-acetate, pH 7.9, 100 mM potassium acetate, 5 mM magnesium acetate, 5 % (v/v) glycerol, 1mM DTT) with 100 µM ApU, 2 µM ATP, 2 µM GTP, 2 µM CTP, 2 µCi [α-$^{32}$P] ATP and incubated for 10 min at 32 °C. The initial incubation allowed the formation of a halted ternary RNAP-σ70-DNA complex containing a 15 nucleotide initial transcript, as UTP was omitted from the reaction. RNA chain elongation was started by addition of 12.5 µl of a pre-heated (32 °C) mixture containing 200 nM NusA, 1 µM NusG, 1 µM NusB-NusE, 500 nM λN, 4 mM ATP, 100 µM GTP, 100 µM CTP and 100 µM UTP in transcription buffer and incubated for defined times. Transcription was stopped by the addition of 5 µl proteinase K (5 mg/ml) and incubated for 5 min at 50 °C followed by PCI extraction and ethanol precipitation (3 vol 1:30 3 M sodium acetate:ethanol, 0.5 µl glycoblue). The RNA was washed three times with 70 % ethanol and dried. RNA was dissolved in 1 x TBE additionally containing 0.5 mg/ml urea, 2 mg/ml xylene cyanol, 2 mg/ml bromophenol blue, 200 µl/ml 0.5 M EDTA, pH 8.0, incubated for 2 min at 96 °C and loaded on a pre-heated 6 M urea, 4 % polyacylamide gel. RNA bands were visualized using a Storm phospohorimager and quantified using Image-Quant software (GE Healthcare). The antitermination efficiency was determined as a ratio of the read-through to the sum of the terminated and read-through products.

**Crystallographic procedures.** NusA$^{\Delta AR2}$, NusB, NusE, λN and *in vitro* transcribed *nut* RNA (λN complex) or NusA$^{\Delta AR2}$, NusB, NusE, λN$^{1-84}$ and synthetic *nut* RNA (λN$^{1-84}$ complex) were mixed in equimolar ratios in 10 mM HEPES-NaOH, pH 7.5, 50 mM NaCl, 1mM DTT and passed over a Superdex 200 (GE Healthcare) size-exclusion column. Fractions containing the target complex were pooled and concentrated to 25 mg/ml. Crystallization was performed using the sitting-drop vapor diffusion technique at 22 °C in 48-well plates. Crystals of the λN complex were obtained by mixing 1 µl complex solution with 0.2 µl of 0.1 M potassium tetracyanoplatinate and 0.8 µl of reservoir solution (0.1 M MES-NaOH, pH 6.3, 14 % (w/v) PEG 12000). Crystals of the λN$^{1-84}$ complex were obtained by mixing equal volumes of complex solution and reservoir solution (0.1 M HEPES-NaOH, pH 7.5, 40 % (v/v) ethylene glycol, 5 % (w/v) PEG 3000). Crystals in their mother liquor (λN$^{1-84}$ complex) or after transfer into mother liquor containing 20 % (v/v) ethylene glycol (λN complex) were flash cooled in liquid nitrogen.

Diffraction data were collected at 100 K on beamline P14 of the Petra III synchrotron (Hamburg, Germany) and on beamline 14.1 of the BESSY II storage ring (Berlin, Germany). The data were processed with XDS[61] and the structure of the λN complex was solved by molecular replacement using the structure coordinates of NusA domains (*Thermotoga maritima* S1-KH1-KH2; PDB ID 1L2F[62]; *E*. coli NTD, PDB ID 2KWP; *E*. coli AR1, PDB ID 1U9L[17]), *E*. coli NusB-NusE (PDB ID 3B3D[9]) and λN$^{1-22}$-*boxB* (PDB ID 1QFQ[14]) as search models with Phaser[63] and completed by model building in Coot[64]. The structure of the λN$^{1-84}$ complex was solved by molecular replacement using the structure coordinates of the λN complex and by alternating rounds of model building in Coot and automated refinement using phenix.refine[65] and CNS[66].

**TAC assembly.** For assembly of a TAC, we followed a strategy as used previously for the structural investigation of bacterial RNAP-based[67,68] and eukaryotic RNA polymerase II-based TECs[69-71]. We generated a 66 nt RNA by *in vitro* transcription, whose 3'-terminal 10 nts could pair to the template DNA strand within a 14 nt non-complementary region of a 65 nt

DNA duplex. The 5'-terminal 33 nts of this RNA contained the same *nut* site as used in crystallization. TAC was assembled by pre-incubating RNAP with the DNA duplex and mixing the RNAP-DNA complex with equimolar amounts of NusA, NusG, NusB, NusE, λN and RNA. The mixture was incubated for 15 min at room temperature and separated by Superdex 200 size-exclusion chromatography in 10 mM HEPES-NaOH, pH 7.5, 50 mM NaCl, 1 mM DTT. Fractions containing the target complex were pooled and used without further concentration.

**EM procedures.** For the determination of an initial map, the sample was applied to a freshly glow-discharged holey carbon grid (Quantifoil) covered with an additional thin carbon support film and negatively stained using 3.5 µl of 2 % uranyl acetate solution. Grids were imaged on a Tecnai Spirit electron microscope operated at 120 kV equipped with a 2kx2k Eagle CCD camera (FEI). Micrographs with a pixel size of 5.19 Å/px were acquired fully automatically using Leginon[72]. CTF parameters were estimated using CTFFind 3[73]. 25,543 particle images were selected manually from 128 micrographs using EMAN2[74], centered and grouped into homogenous classes using the ISAC algorithm[75]. 240 stable classes were found accounting for 17,825 particle images. Suitable class averages exhibiting distinctive structural features and representing possible projections of the complex were selected and used for the reconstruction of an initial map *via* the VIPER algorithm[76,77]. The map was used to determine initial parameters by template matching for all particle images using SPARX[78].

For cryo-EM data acquisition, the sample was applied to freshly glow-discharged R3/3 holey carbon grids (Quantifoil) covered with a 2 nm carbon support film and plunge-frozen into liquid ethane using a Vitrobot plunger (FEI) with the blotting chamber set to 4 °C and 100 % humidity. Grids were frozen using a blotting time of 2 to 4 s and stored in liquid nitrogen until use. For the validation of the initial map, a small cryo-EM dataset was acquired on a Tecnai Spirit electron microscope operated at 120 kV equipped with a 2kx2k Eagle CCD camera (FEI). Micrographs were recorded with a pixel size of 5.19 Å/px using Leginon and processed as described before. 46,327 particle images were selected with EMAN2 and used for template matching using SPARX against the previously determined initial map.

For the final cryo-EM reconstruction, micrographs were acquired on a Tecnai G2 Polara electron microscope (FEI) operated at 300 kV, equipped with a K2 direct electron detector (Gatan) operated in super-resolution mode and using the Leginon system. 1,721 micrographs were acquired within a defocus range of 1 to 8 µm under low-dose conditions with a pixel size of 0.64 Å/px. Defocus estimation was performed using the CTFFIND4 package[79]. Dose-fractionated image stacks were aligned using MotionCorr[80]. 102,374 particle images were picked manually using EMAN2. All subsequent image processing steps were performed with Relion 1.4[81,82] unless otherwise noted. 4-fold binned particle images were subjected to 2D classification and further selected by image quality (using a figure-of-merit criterion of better than 8 Å in CTFFIND4), resulting in a total of 79,946 suitable particle images. To find the initial parameters, 3D refinement was performed with all 2-fold binned particles against the previously determined initial map. 3D classification without alignment was then performed to sort the particles into 4 classes, with the classes containing 18, 28, 24 and 30 % of the particles images, respectively. 3D refinement was performed for each class independently, resulting in three low-quality maps and one medium-quality class with an overall gold-standard resolution of 9.8 Å according to the $FSC_{0.143}$ criterion (Supplementary Fig. 1e,g). Power spectrum readjustment was applied to the final map in SPIDER[83] after calculating a spectral weighting function from the raw map and a map simulated from the atomic model and map was low pass filtered to the global FSC estimate. The local resolution of the raw map was calculated using ResMap[84] (Supplementary Fig. 1h,i).

The low overall resolution can be explained with the strong orientational preference of the particle images (Supplementary Fig. 1j). Although results from ResMap indicate a large portion of voxels with a resolution better than 9 Å (Supplementary Fig. 1h), the map itself does not show details expected at this resolution. Therefore, we have chosen the $FSC_{0.143}$ criterion as a more reliable representation of the map resolution. The strong orientational preference might have an impact on the ResMap calculations.

**Modeling the TAC.** A crystal structure of *E. coli* RNAP (PDB ID: 4YG2[44]) was manually docked into the cryo-EM map and the fit was optimized by rigid body refinement in Coot. The RNAP model was locally adjusted by transferring the FTH coordinates from the structure of *E. coli* RNAP in complex with RapA (PDB ID: 4S20[85]) after superpositioning of the RNAP cores. Nucleic acids were transferred from structures of *Thermus thermophilus* RNAP (PDB ID: 2O5I[67]; DNA:RNA hybrid and RNA in the RNA exit channel) and bovine Pol II (PDB ID: 5FLM[86]; upstream dsDNA) by superpositioning of RNAP core subunits. The λN$^{1-84}$-NusA$^{\Delta AR2}$-NusB-NusE-*nut* RNP was docked into a distinctly shaped, unoccupied region of the cryo-EM map and adjusted by rigid body refinement. The structures of NusG NTD and NusG CTD (PDB ID: 2K06 and 2JVV[87]) were added according to the structure of archael Spt4/5 bound to the ß' clamp helices of Pol II (PDB ID: 3QQC[33]) and an E. *coli* NusE-NusG-CTD structure (PDB ID: 2KVQ[34]), respectively.

**Chemical cross-linking and cross-link identification by mass spectrometry.** For CX-MS, we employed the homobifunctional, primary amino group-reactive BS3 (bis(sulfosuccinimidyl)suberate) cross-linker, which can bridge N termini or lysine side chains, when their Cα atoms are < 35 Å apart, or the hetero-bifunctional "zero-length" EDC (N-(3-dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride) and DMTMM (4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride) cross-linkers, which mediate the direct formation of amide bonds between carboxylate (Asp or Glu side chains) and primary amine (Lys side chains) groups.

25 pmol of the λN-NusA$^{\Delta AR2}$-NusB-NusE-*nut* RNP were cross-linked with 150 µM BS3 or 30 mM DMTMM (4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride) in 10 mM HEPES, pH 7.5, 50 mM NaCl and 1 mM DTT and incubated for 30 min at room temperature. The reaction was stopped by 25 mM Tris-HCl. Cross-linking efficiency was analysed by SDS PAGE. The cross-linked λN-NusA$^{\Delta AR2}$-NusB-NusE-*nut* RNP was excised from the gel and digested with trypsin{Schmidt, 2009 #10739}. The digested proteins were dissolved in 20 µl of 5 % (v/v) acetonitrile and 0.1 % (v/v) formic acid and subjected to liquid

chromatography-tandem mass spectrometry (LC-MS/MS) analysis using a Q Exactive (Thermo Scientific) mass spectrometer.

For TAC and TEC, 100 pmol of purified recombinant complex were cross-linked with 150 µM BS3 for 30 min in 10 mM HEPES-NaOH, pH 7.5, 50 mM NaCl or with 4 mM EDC, 1 mM N-hydroxysuccinimide for 60 min in 10 mM MES-NaOH, pH 6.5, 50 mM NaCl at 25 °C. The samples were acetone-precipitated and analyzed essentially as described before[89], with the following modifications: Precipitated material was dissolved in 4 M urea/50 mM ammonium bicarbonate, reduced with DTT, alkylated with iodoacetamide, diluted to 1 M urea and digested with trypsin (1:20 [w:w]). Peptides were reversed-phase extracted using Sep-Pak Vac tC18 1cc cartridges (Waters) and fractionated by size-exclusion on a Superdex Peptide PC3.2/30 column (GE HealthCare). 50 µl fractions corresponding to elution volumes of 1.2-1.8 ml were analyzed on a Q Exactive HF (Thermo Scientific) mass spectrometer. Protein-protein cross-links were identified *via* the pLink1.22 search engine (http://pfind.ict.ac.cn/software/pLink) and filtered at FDR 1 % according to the recommendations of the developers[90]. A cross-link score was evaluated as the negative value of the common logarithm of the original pLink Score (score = $-\log_{10}$[pLink score]). The cross-links observed with > 4 spectra for BS3 and > 7 spectra for EDC and DMTMM are listed in Supplementary Tables 2-4.

**Data availability.** Coordinates and structure factors have been deposited in the RCSB Protein Data Bank (www.pdb.org) with accession codes 5LM7 ($NusA^{\Delta AR2}$-NusB-NusE-$\lambda N^{1-84}$-*nut* RNA complex) and 5LM9 ($NusA^{100-426}$). Cryo-EM data have been deposited in the RCSB Protein Data Bank with accession code 5MS0 and in the Electron Microscopy Data Bank (www.emdatabank.org) with accession code EMD-3561.

## References

1. Zhang, J. & Landick, R. A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. *Trends Biochem. Sci.*, (2016).

2. Mooney, R. A. *et al.* Regulator Trafficking on Bacterial Transcription Units In Vivo. *Mol. Cell* **33**, 97-108, (2009).

3. Yang, X. & Lewis, P. J. The interaction between RNA polymerase and the elongation factor NusA. *RNA Biol.* **7**, 272-275, (2010).

4. Tomar, S. K. & Artsimovitch, I. NusG-Spt5 proteins-Universal tools for transcription modification and communication. *Chem. Rev.* **113**, 8604-8619, (2013).

5. Nudler, E. & Gottesman, M. E. Transcription termination and anti-termination in E. coli. *Genes Cells* **7**, 755-768, (2002).

6. Ciampi, M. S. Rho-dependent terminators and transcription termination. *Microbiol.* **152**, 2515-2528, (2006).

7. Torres, M., Condon, C., Balada, J. M., Squires, C. & Squires, C. L. Ribosomal protein S4 is a transcription factor with properties remarkably similar to NusA, a protein involved in both non-ribosomal and ribosomal RNA antitermination. *EMBO J.* **20**, 3811-3820, (2001).

8. Worbs, M., Bourenkov, G. P., Bartunik, H. D., Huber, R. & Wahl, M. C. An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. *Mol. Cell* **7**, 1177-1189, (2001).

9. Luo, X. *et al.* Structural and functional analysis of the E. coli NusB-S10 transcription antitermination complex. *Mol. Cell* **32**, 791-802, (2008).

10. Stagno, J. R. *et al.* Structural basis for RNA recognition by NusB and NusE in the initiation of transcription antitermination. *Nucleic Acids Res.* **39**, 7803-7815, (2011).

11. Zhou, Y., Mah, T. F., Greenblatt, J. & Friedman, D. I. Evidence that the KH RNA-binding domains influence the action of the E. coli NusA protein. *J. Mol. Biol.* **318**, 1175-1188, (2002).

12. Mogridge, J. *et al.* Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. *Mol. Cell* **1**, 265-275, (1998).

13. Legault, P., Li, J., Mogridge, J., Kay, L. E. & Greenblatt, J. NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell* **93**, 289-299, (1998).

14. Scharpf, M. *et al.* Antitermination in bacteriophage lambda. The structure of the N36 peptide-boxB RNA complex. *Eur. J. Biochem.* **267**, 2397-2408, (2000).

15. Mogridge, J., Mah, T. F. & Greenblatt, J. Involvement of boxA nucleotides in the formation of a stable ribonucleoprotein complex containing the bacteriophage lambda N protein. *J. Biol. Chem.* **273**, 4143-4148, (1998).

16. Mah, T. F., Li, J., Davidson, A. R. & Greenblatt, J. Functional importance of regions in Escherichia coli elongation factor NusA that interact with RNA polymerase, the bacteriophage lambda N protein and RNA. *Mol. Microbiol.* **34**, 523-537, (1999).

17. Bonin, I. *et al.* Structural basis for the interaction of Escherichia coli NusA with protein N of phage lambda. *Proc. Natl. Acad. Sci. USA* **101**, 13762-13767, (2004).

18. Mishra, S., Mohan, S., Godavarthi, S. & Sen, R. The interaction surface of a bacterial transcription elongation factor required for complex formation with an antiterminator during transcription antitermination. *J. Biol. Chem.* **288**, 28089-28103, (2013).

19. Mah, T. F., Kuznedelov, K., Mushegian, A., Severinov, K. & Greenblatt, J. The alpha subunit of E. coli RNA polymerase activates RNA binding by NusA. *Genes Dev.* **14**, 2664-2675, (2000).

20. Mogridge, J., Mah, T. F. & Greenblatt, J. A protein-RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase of a stable antitermination complex containing the lambda N protein. *Genes Dev.* **9**, 2831-2845, (1995).

21. Zhou, Y. *et al.* Interactions of an Arg-rich region of transcription elongation protein NusA with NUT RNA: Implications for the order of assembly of the lambda N antitermination complex in vivo. *J. Mol. Biol.* **310**, 33-49, (2001).

22. Friedman, D. I. & Baron, L. S. Genetic characterization of a bacterial locus involved in the activity of the N function of phage lambda. *Virology* **58**, 141-148, (1974).

23. Ward, D. F. & Gottesman, M. E. The nus mutations affect transcription termination in Escherichia coli. *Nature* **292**, 212-215, (1981).

24. Craven, M. G. & Friedman, D. I. Analysis of the Escherichia coli nusA10(Cs) allele: relating nucleotide changes to phenotypes. *J. Bacteriol.* **173**, 1485-1491, (1991).

25. Craven, M. G. *et al.* Escherichia coli-Salmonella typhimurium hybrid nusA genes: identification of a short motif required for action of the lambda N transcription antitermination protein. *J. Bacteriol.* **176**, 1394-1404, (1994).

26. Friedman, D. I. *et al.* Interactions of bacteriophage and host macromolecules in the growth of bacteriophage lambda. *Microbiol. Rev.* **48**, 299-325, (1984).

27. Ward, D. F., DeLong, A. & Gottesman, M. E. Escherichia coli nusB mutations that suppress nusA1 exhibit lambda N specificity. *J. Mol. Biol.* **168**, 73-85, (1983).

28. Friedman, D. I., Schauer, A. T., Baumann, M. R., Baron, L. S. & Adhya, S. L. Evidence that ribosomal protein S10 participates in control of transcription termination. *Proc. Natl. Acad. Sci. USA* **78**, 1115-1118, (1981).

29. Sparkowski, J. & Das, A. Simultaneous gain and loss of functions caused by a single amino acid substitution in the beta subunit of Escherichia coli RNA polymerase: suppression of nusA and rho mutations and conditional lethality. *Genetics.* **130**, 411-428, (1992).

30. Friedman, D. I. & Olson, E. R. Evidence that a nucleotide sequence, "boxA," is involved in the action of the NusA protein. *Cell* **34**, 143-149, (1983).

31. Friedman, D. I., Olson, E. R., Johnson, L. L., Alessi, D. & Craven, M. G. Transcription-dependent competition for a host factor: the function and optimal sequence of the phage lambda boxA transcription antitermination signal. *Genes Dev.* **4**, 2210-2222, (1990).

32. Liu, B. & Steitz, T. A. Structural insights into NusG regulating transcription elongation. *Nucleic Acids Res.* **45**, 968-974, (2017).

33. Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. M. & Cramer, P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* **30**, 1302-1310, (2011).

34. Burmann, B. M. *et al.* A NusE:NusG complex links transcription and translation. *Science* **328**, 501-504, (2010).

35. Sullivan, S. L., Ward, D. F. & Gottesman, M. E. Effect of Escherichia coli nusG function on lambda N-mediated transcription antitermination. *J. Bacteriol.* **174**, 1339-1344, (1992).

36. Kuznedelov, K., Korzheva, N., Mustaev, A. & Severinov, K. Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA-DNA hybrid length. *EMBO J.* **21**, 1369-1378, (2002).

37. Gusarov, I. & Nudler, E. Control of intrinsic transcription termination by N and NusA: the basic mechanisms. *Cell* **107**, 437-449, (2001).

38. Parks, A. R. *et al.* Bacteriophage lambda N protein inhibits transcription slippage by Escherichia coli RNA polymerase. *Nucleic Acids Res.* **42**, 5823-5829, (2014).

39. Mishra, S. & Sen, R. N protein from lambdoid phages transforms NusA into an antiterminator by modulating NusA-RNA polymerase flap domain interactions. *Nucleic Acids Res.* **43**, 5744-5758, (2015).

40. Cheeran, A. *et al.* Escherichia coli RNA polymerase mutations located near the upstream edge of an RNA:DNA hybrid and the beginning of the RNA-exit channel are defective for transcription antitermination by the N protein from lambdoid phage H-19B. *J. Mol. Biol.* **352**, 28-43, (2005).

41. Georgopoulos, C. P. Bacterial Mutants in Which Gene N Function of Bacteriophage Lambda Is Blocked Have an Altered Rna Polymerase. *Proc. Natl. Acad. Sci. USA* **68**, 2977-2981, (1971).

42. Jin, D. J. *et al.* Effects of rifampicin resistant rpoB mutations on antitermination and interaction with nusA in Escherichia coli. *J. Mol. Biol.* **204**, 247-261, (1988).

43. Cheeran, A., Kolli, N. R. & Sen, R. The site of action of the antiterminator protein N from the lambdoid phage H-19B. *J. Biol. Chem.* **282**, 30997-31007, (2007).

44. Murakami, K. S. X-ray crystal structure of Escherichia coli RNA polymerase sigma70 holoenzyme. *J. Biol. Chem.* **288**, 9126-9134, (2013).

45. Toulokhonov, I., Artsimovitch, I. & Landick, R. Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science* **292**, 730-733, (2001).

46. Ha, K. S., Toulokhonov, I., Vassylyev, D. G. & Landick, R. The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase. *J. Mol. Biol.* **401**, 708-725, (2010).

47. Rees, W. A., Weitzel, S. E., Yager, T. D., Das, A. & von Hippel, P. H. Bacteriophage lambda N protein alone can induce transcription antitermination in vitro. *Proc. Natl. Acad. Sci. USA* **93**, 342-346, (1996).

48. Drogemuller, J. *et al.* Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Sci. Rep.* **5**, 16428, (2015).

49. Kolb, K. E., Hein, P. P. & Landick, R. Antisense oligonucleotide-stimulated transcriptional pausing reveals RNA exit channel specificity of RNA polymerase and mechanistic contributions of NusA and RfaH. *J. Biol. Chem.* **289**, 1151-1163, (2014).

50. Beuth, B., Pennell, S., Arnvig, K. B., Martin, S. R. & Taylor, I. A. Structure of a Mycobacterium tuberculosis NusA-RNA complex. *EMBO J.* **24**, 3576-3587, (2005).

51. Artsimovitch, I., Svetlov, V., Murakami, K. S. & Landick, R. Co-overexpression of Escherichia coli RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. *J. Biol. Chem.* **278**, 12344-12355, (2003).

52. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Express. Purif.* **41**, 207-234, (2005).

53. Kyzer, S., Ha, K. S., Landick, R. & Palangat, M. Direct versus limited-step reconstitution reveals key features of an RNA hairpin-stabilized paused transcription complex. *J. Biol. Chem.* **282**, 19020-19028, (2007).

54. Belogurov, G. A. *et al.* Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol. Cell* **26**, 117-129, (2007).

55. Easton, L. E., Shibata, Y. & Lukavsky, P. J. Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA* **16**, 647-653, (2010).

56. McKenna, S. A. *et al.* Purification and characterization of transcribed RNAs using gel filtration chromatography. *Nat. Protoc.* **2**, 3270-3277, (2007).

57. Kao, C., Zheng, M. & Rudisser, S. A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5**, 1268-1272, (1999).

58. Sherlin, L. D. *et al.* Chemical and enzymatic synthesis of tRNAs for high-throughput crystallization. *RNA* **7**, 1671-1678, (2001).

59. Rao, R. N. Construction and properties of plasmid pKC30, a pBR322 derivative containing the pL-N region of phage lambda. *Gene* **31**, 247-250, (1984).

60. Artsimovitch, I. & Henkin, T. M. In vitro approaches to analysis of transcription termination. *Methods* **47**, 37-43, (2009).

61. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125-132, (2010).

62. Shin, D. H. *et al.* Crystal structure of NusA from Thermotoga maritima and functional implication of the N-terminal domain. *Biochemistry* **42**, 13429-13437, (2003).

63. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63**, 32-41, (2007).

64. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486-501, (2010).

65. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352-367, (2012).

66. Schroder, G. F., Levitt, M. & Brunger, A. T. Deformable elastic network refinement for low-resolution macromolecular crystallography. *Acta Crystallogr. D* **70**, 2241-2255, (2014).

67. Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**, 157-162, (2007).

68. Vassylyev, D. G. *et al.* Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448**, 163-168, (2007).

69. Westover, K. D., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* **303**, 1014-1016, (2004).

70. Westover, K. D., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119**, 481-489, (2004).

71. Kettenberger, H., Armache, K. J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* **16**, 955-965, (2004).

72. Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41-60, (2005).

73. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334-347, (2003).

74. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38-46, (2007).

75. Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J. & Penczek, P. A. Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* **20**, 237-247, (2012).

76. Penczek, P. A. & Asturias, F. J. Ab initio cryo-EM structure determination as a validation problem. *IEEE International Conference on Image Processing (ICIP)* **2014**, 2090-2094, (2014).

77. Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. A primer to single-particle cryo-electron microscopy. *Cell* **161**, 438-449, (2015).

78. Hohn, M. *et al.* SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157**, 47-55, (2007).

79. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216-221, (2015).

80. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584-590, (2013).

81. Scheres, S. H. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114-122, (2015).

82. Bharat, T. A., Russo, C. J., Lowe, J., Passmore, L. A. & Scheres, S. H. Advances in Single-Particle Electron Cryomicroscopy Structure Determination applied to Sub-tomogram Averaging. *Structure* **23**, 1743-1753, (2015).

83. Frank, J., Shimkin, B. & Dowse, H. Spider - a Modular Software System for Electron Image-Processing. *Ultramicroscopy* **6**, 343-357, (1981).

84. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63-65, (2014).

85. Liu, B., Zuo, Y. & Steitz, T. A. Structural basis for transcription reactivation by RapA. *Proc. Natl. Acad. Sci. USA* **112**, 2006-2010, (2015).

86. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* **529**, 551-554, (2016).

87. Mooney, R. A., Schweimer, K., Rosch, P., Gottesman, M. & Landick, R. Two structurally independent domains of E. coli NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. *J. Mol. Biol.* **391**, 341-358, (2009).

88. Schmidt, C. & Urlaub, H. iTRAQ-labeling of in-gel digested proteins for relative quantification. *Methods Mol. Biol.* **564**, 207-226, (2009).

89. Leitner, A., Walzthoeni, T. & Aebersold, R. 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**, 120-137, (2014).

90. Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904-906, (2012).


## Acknowledgements

We are grateful to Claudia Alings, Freie Universität Berlin, for help with crystallization. We thank Irina Artsimovitch, Ohio State University, for plasmids pVS-10 (T7P-α-β-β'$His_6$-ω) and pIA1127 (T7P-$His_6$-TEV-σ70), used for RNAP production, and Max Gottesman, New York

University, for plasmid pKC30, used in transcription assays. We acknowledge access to beamlines of the BESSY II storage ring (Berlin, Germany) *via* the Joint Berlin MX-Laboratory sponsored by the Helmholtz Zentrum Berlin für Materialien und Energie, the Freie Universität Berlin, the Humboldt-Universität zu Berlin, the Max-Delbrück-Centrum and the Leibniz-Institut für Molekulare Pharmakologie, and to beamline 14-1 at the Petra III storage ring (EMBL, Hamburg, Germany). This work was supported by the Deutsche Forschungsgemeinschaft (SFB 740 to T.M., C.M.S. and M.C.W. and grant WA 1126/5-1 to M.C.W.). K.F.S. was supported by a Dahlem International Network PostDoc grant from Freie Universität Berlin. E.B. holds a Freigeist-Fellowship from the Volkswagen Foundation and acknowledges continuous support from the Caesar Foundation.

## Author contributions

N.S., F.K., E.A., K.F.S., O.D., Y.H.H., C.T.L., B.L., E.B., J.B., T.M., J.L. and G.W. performed the experiments. All authors contributed to the analysis of the data and the interpretation of the results. N.S. and M.C.W. wrote the manuscript.

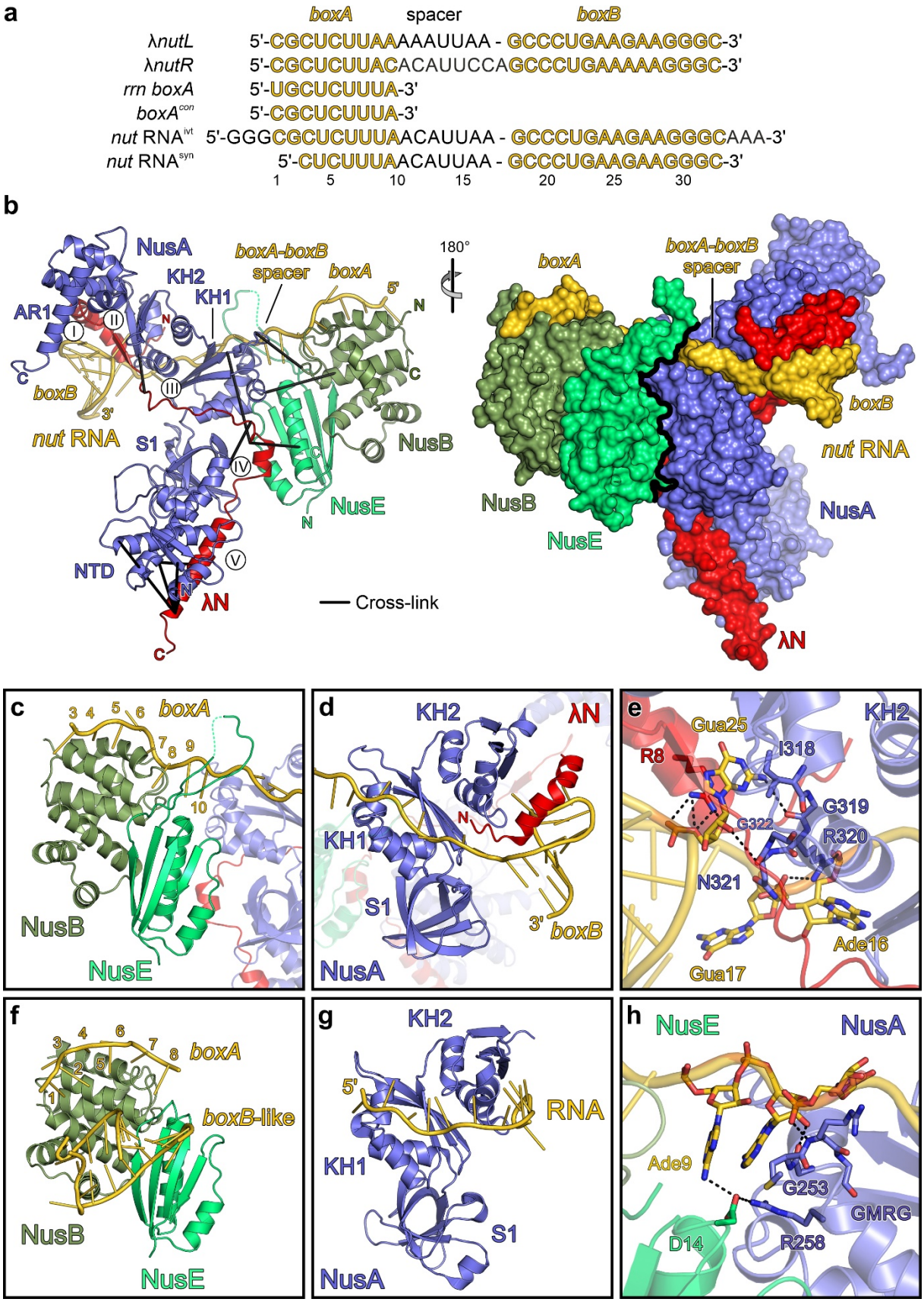
## Additional information

Supplementary information is available for this paper. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.C.W.

## Competing interests

The authors declare no competing interests.

833 # Figure and figure legends

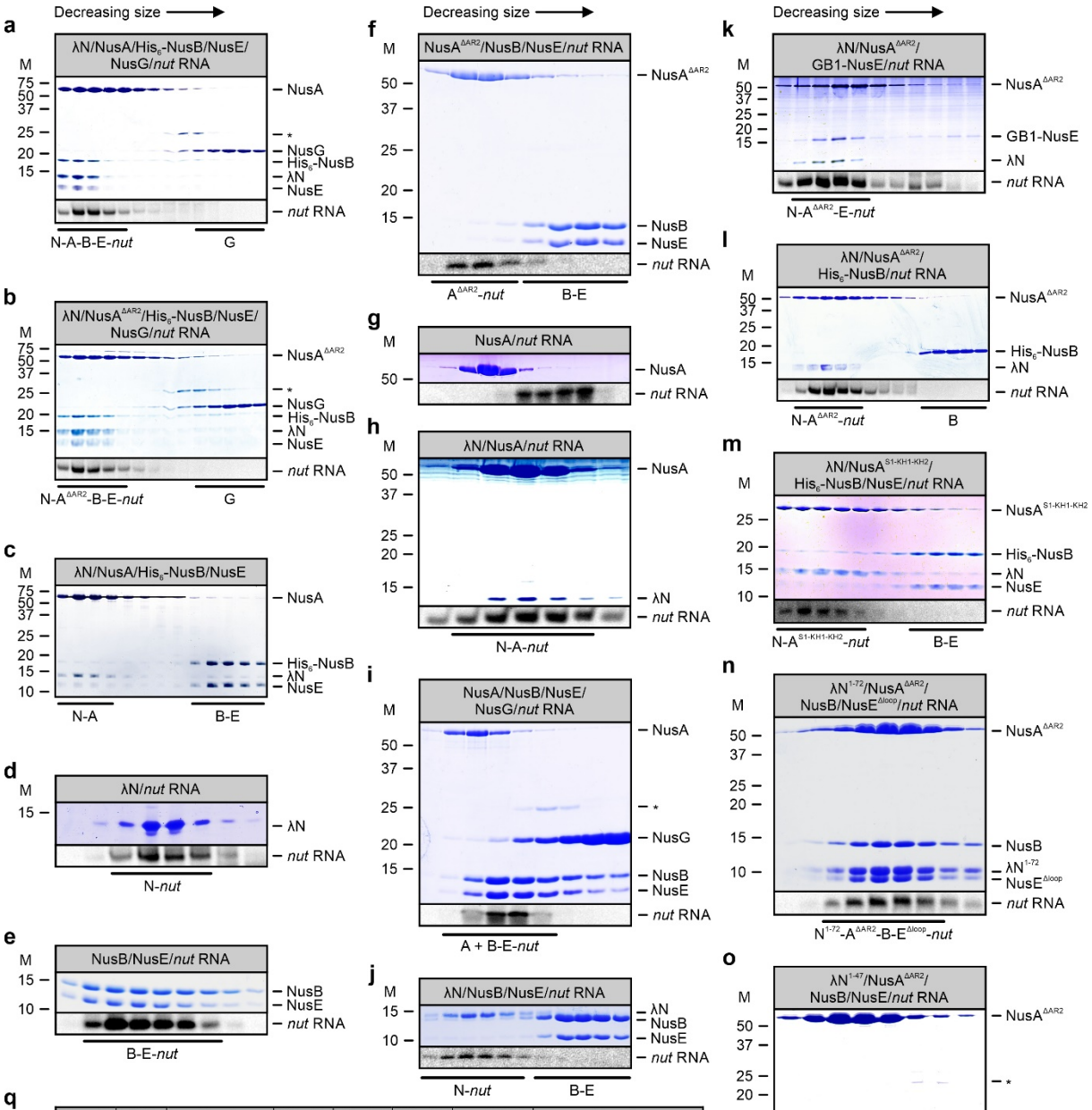



834

214

835

**Figure 1. Structure of a λN$^{1-84}$-NusA$^{\Delta AR2}$-NusB-NusE-*nut* RNP. a**, λ, ribosomal (r) RNA (*rrn*) and consensus (con) *nut* RNA sequences and RNAs used for crystallization (ivt, *in vitro* transcribed; syn, chemically synthesized). **b**, Diametric cartoon and surface representations of the λN$^{1-84}$-NusA$^{\Delta AR2}$-NusB-NusE-*nut* RNP. N – N-termini; C – C-termini. Black lines in cartoon view, inter-molecular cross-links observed in the λN-NusA$^{\Delta AR2}$-NusB-NusE-*nut* RNP in solution. Roman numerals – λN-NusA interaction regions. Black line in surface view, border between the NusB-NusE-*boxA* and λN$^{1-84}$-NusA$^{\Delta AR2}$-*boxB* half-RNPs. **c**, NusB-NusE-*boxA* unit of the λN$^{1-84}$-NusA$^{\Delta AR2}$-NusB-NusE-*nut* RNP. Nucleotide numbering according to (**a**). Orientation as in (**b**), right panel. **d**, λN$^{1-20}$-NusA$^{S1-KH1-KH2}$-*boxB*-spacer unit of the λN$^{1-84}$-NusA$^{\Delta AR2}$-NusB-NusE-*nut* RNP. Rotated 30° about the horizontal axis (top to front) compared to (**b**), right panel. **e**, Interaction of λN and NusA with the extruded nucleobase Gua25, sandwiched between λN R8, NusA I318 and the following GXXG of NusA KH2. R320 of the GXXG motif additionally binds the backbone of Ade16-Gua17. Orientation as in (**b**), left panel. **f**, Comparison of the structure of an isolated NusB-NusE-RNA complex (PDB ID: 3R2C). **g**, Comparison of the structure of an isolated NusA$^{S1-KH1-KH2}$-RNA complex (PDB ID: 2ATW). **h**, Close-up view of the triple interaction between NusA KH1, NusE and the *boxA*-*boxB* spacer. Rotated 10° clockwise about the vertical axis compared to (**b**), right panel.

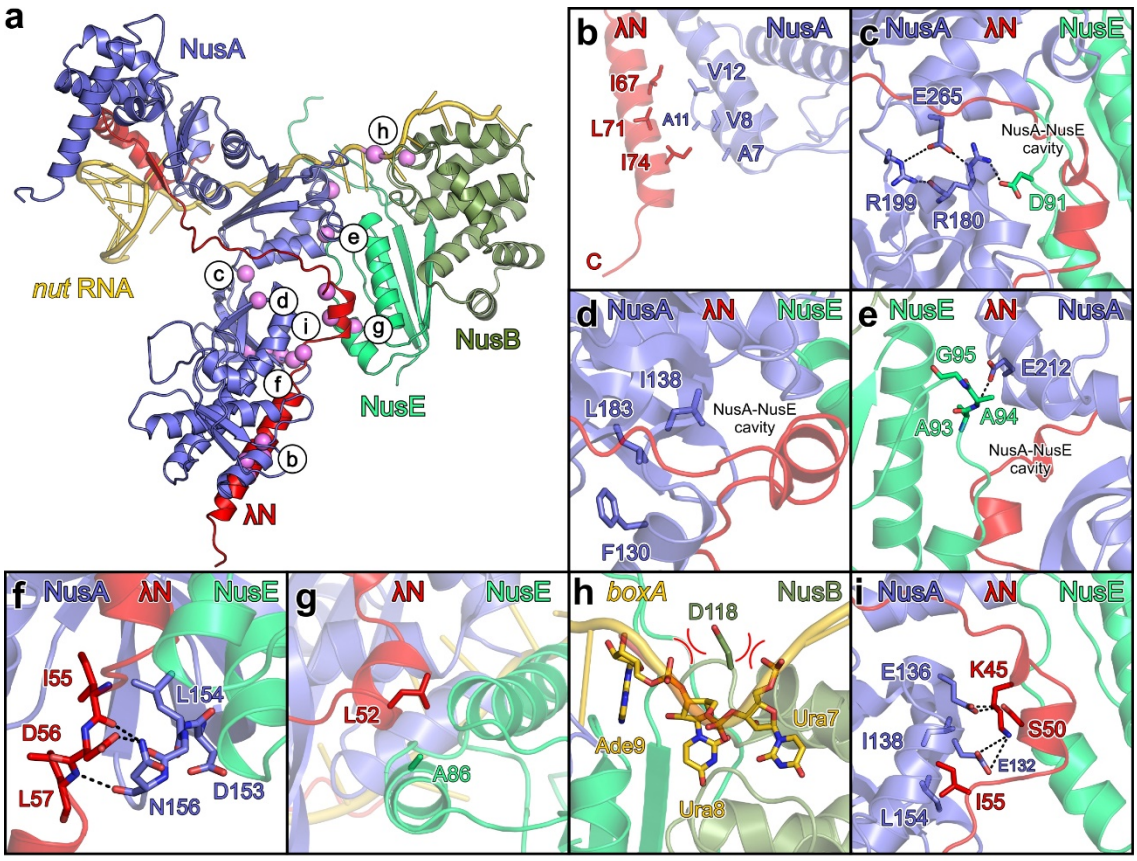



q

| Panel | λN | NusA | NusB | NusE | NusG | *nut* RNA | Result |
|---|---|---|---|---|---|---|---|
| **a** | + | + | $His_6$ | + | + | + | N-A-B-E-*nut* + G |
| **b** | + | ΔAR2 | $His_6$ | + | + | + | N-$A^{\Delta AR2}$-B-E-*nut* + G |
| **c** | + | + | $His_6$ | + | - | - | N-A + B-E |
| **d** | + | - | - | - | - | + | N-*nut* |
| **e** | - | - | + | + | - | + | B-E-*nut* |
| **f** | - | ΔAR2 | + | + | - | + | $A^{\Delta AR2}$-*nut* + B-E |
| **g** | - | + | - | - | - | + | A + *nut* |
| **h** | + | + | - | - | - | + | N-A-*nut* |
| **i** | - | + | + | + | + | + | A + B-E-*nut* + G |
| **j** | + | - | + | + | - | + | N-*nut* + B-E |
| **k** | + | ΔAR2 | - | GB1 | - | + | N-$A^{\Delta AR2}$-E-*nut* |
| **l** | + | ΔAR2 | $His_6$ | - | - | + | N-$A^{\Delta AR2}$-*nut* + B |
| **m** | + | S1-KH1-KH2 | $His_6$ | + | - | + | N-$A^{S1\text{-}KH1\text{-}KH2}$-*nut* + B-E |
| **n** | 1-72 | ΔAR2 | + | Δloop | - | + | $N^{1\text{-}72}A^{\Delta AR2}$-B-$E^{\Delta loop}$-*nut* |
| **o** | 1-47 | ΔAR2 | + | + | - | + | $N^{1\text{-}47}$-$A^{\Delta AR2}$-*nut* + B-E |
| **p** | + | R258A | $His_6$ | + | - | + | N-$A^{R258A}$-*nut* + B-E |

**Figure 2. Interaction studies. a-p**, Coomassie-stained SDS PAGE (proteins; top) and ethidium bromide-stained urea PAGE (RNA; bottom) analyses of analytical size-exclusion

chromatography runs, monitoring interactions among λN, Nus factors and *nut* RNA. Mixtures that were loaded for each run are indicated in the boxes above the gels. Molecular weight sizes are shown on the left (M). Bands are identified on the right, complexes formed are indicated below the gels. N, λN; A, NusA; B, NusB; E, NusE; G, NusG; *nut*, *nut* RNA; $NusE^{\Delta loop}/E^{\Delta loop}$ NusE variant with the long protruding loop deleted, which has no known functional consequences for λN-dependent antitermination[9]. *, unidentified contaminants. **q**, Overview of the components mixed in the experiments shown in panels **a-p** and the results obtained. Gels shown are representative examples of at least two repetitions.




j

| Factor | Mutation | Allele | Effect | Figure | References |
|---|---|---|---|---|---|
| NusA | A7D | - | Defective phage propagation/antitermination | 3b | 18 |
| | V8A | - | Defective phage propagation/antitermination | 3b | |
| | V8E | - | Defective phage propagation/antitermination | 3b | |
| | A11D | - | Defective phage propagation/antitermination | 3b | |
| | V12D | - | Defective phage propagation/antitermination | 3b | |
| | L31E | - | Defective phage propagation/antitermination | - | |
| | R199A | - | Abrogation of λN antitermination, reduced binding to λN-*nut* RNP | 3c | 19,21 |
| | L183R | *nusA1* | Block of λN antitermination | 3d | 22,23 |
| | R104H/E212K | *nusA10* | Reduction of λN antitermination | 3e | 24 |
| | 944 | *nusA944* | Binding to λN-*nut* RNP, failure to support antitermination | 3f | 21,25 |
| | G253D | - | No binding to λN-*nut* RNP, failure to support antitermination | 1h | 11 |
| | G319D | - | No binding to λN-*nut* RNP, mild defect in supporting λN action | 1e | |
| NusE | A86D | *nusE71* | Block of λN but not of *rrn* antitermination | 3g | 26 |
| NusB | D118N | *nusB101* | Enhanced affinity to *boxA*, Suppression of *nusA1* and *nusE71* defects | 3h | 9,27 |
| λN | K45R | *pun134* | Suppression of *nusA1* defects | 3i | 28,29 |
| | S50R | *pun150* | Suppression of *nusA1* defects | 3i | |
| | I55M | *pun165* | Suppression of *nusA1* defects | 3i | |
| ***nut*** | boxA Ade9Uri | - | Abrogation of λN antitermination | 1h | 30,31 |

**Figure 3. Mapping of subunit variants. a**, Overview of previously investigated variants of subunits, with affected residues mapped to the λN$^{1-84}$-NusA$^{\Delta AR2}$-NusB-NusE-*nut* RNP structure. Magenta spheres, Cα or phosphorus atoms of affected protein or RNA residues,

872 respectively. Letters refer to close-up views in the following panels. **b**, Close-up view of the
873 region affected by the A7D, V8A, V8E, A11D and V12D exchanges in NusA. Rotated 90°
874 counter-clockwise about the vertical axis relative to (**a**). **c**, Close-up view of the region
875 affected by the NusA$^{R199A}$ variant. Rotated 20° counter-clockwise about the vertical axis
876 relative to (**a**). **d**, Close-up view of the region affected by the *nusA1* (L183R) allele. Rotated
877 60° about the horizontal axis (top to front) relative to (**a**). **e**, Close-up view of the region
878 affected by the *nusA10* (E212K and R104H) allele. Rotated 150° counter-clockwise about the
879 vertical axis relative to (**a**). **f**, Close-up view of the region affected by an insertion *via* the
880 *nusA944* allele (NusA residues 153-156). Rotated 60° clockwise about the vertical axis
881 relative to (**a**). **g**, Close-up view of region affected by the *nusE71* (A86D) allele. NusB cut
882 away. Rotated 30° clockwise about the vertical axis and 60° about the horizontal axis (top to
883 back) relative to (**a**). **h**, Close-up view of the region affected by the *nusB101* (D118N) allele.
884 Red double arcs, repulsive electrostatic effects. Rotated 20° clockwise about the viewing axis
885 and 40° about the horizontal axis (top to front) relative to (**a**). **i**, Close-up view of the region
886 affected by the *pun* alleles, encoding K45R, S50R, I55M exchanges in λN. Same orientation
887 as in (**a**). **j**, List of variants discussed in the text and their effects.
888

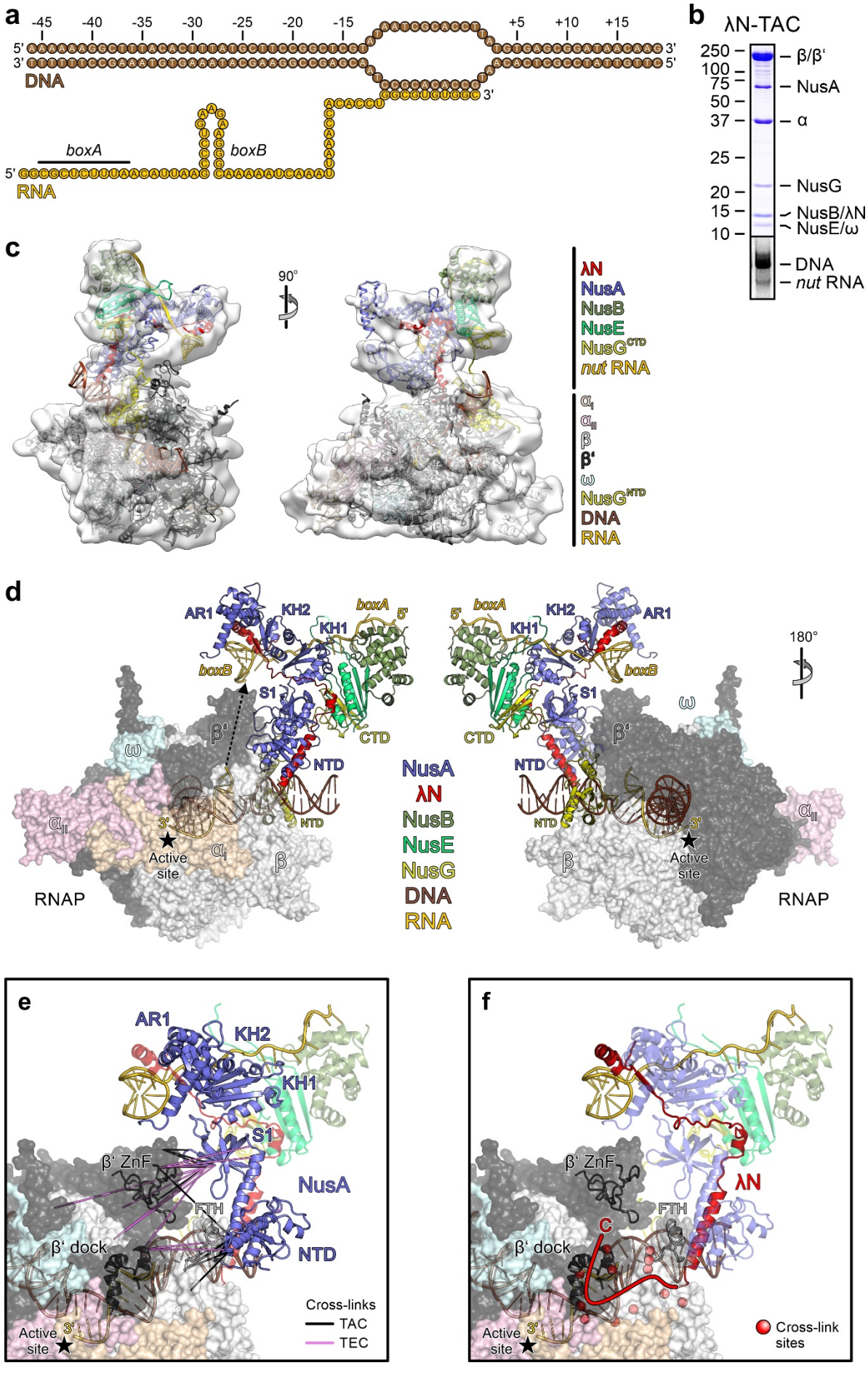

a
DNA
RNA
boxA
boxB
b
λN-TAC
β/β‘
NusA
α
NusG
NusB/λN
NusE/ω
DNA
nut RNA
c
90°
λN
NusA
NusB
NusE
NusG^CTD
nut RNA
α_I
α_II
β
β‘
ω
NusG^NTD
DNA
RNA
d
AR1
KH2
KH1
boxA
boxB
S1
CTD
NTD
ω
β‘
α_II
α_I
β
Active site
RNAP
180°
NusA
λN
NusB
NusE
NusG
DNA
RNA
e
AR1
KH2
KH1
S1
β‘ ZnF
NusA
FTH
β‘ dock
NTD
Cross-links
TAC
TEC
Active site
f
β‘ ZnF
λN
FTH
C
β‘ dock
Cross-link sites
Active site

**Figure 4. Structure of a λN-based transcription antitermination complex. a**, Scheme of nucleic acids used for transcription antitermination complex (TAC) assembly. **b**, Analysis of the assembled TAC. Top, Coomassie-stained SDS PAGE (proteins); bottom, ethidium bromide-stained urea PAGE (nucleic acids). Molecular weight sizes are shown on the left (M). Gels shown are representative examples of three repetitions. **c**, Fitting of the TAC components to the cryo-EM map. At the present resolution, the positions of the NusA AR2 domain and of the α-CTD's could not be modeled reliably and have thus been omitted from the model. **d**, Structural model of the λN-based TAC. Asterisk, active site. Dashed arrow, connection between RNA emerging from the exit channel and *boxB*. Orientations of the λN-NusA-NusB-NusE-*nut* RNP as in Fig. 1b, left panel. **e**, Cross-links between NusA and RNAP subunits in the TAC (black lines) and a TEC lacking λN (violet lines). Elements of the RNA exit channel are shown as cartoons. Rotated 30° counter-clockwise about the vertical axis and 30 ° about the horizontal axis (top to front) relative to (**d**), left panel. **f**, Cross-links between λN C-terminal residues absent in the crystal structure and RNAP, indicating the path of the λN C-terminus around the RNA exit channel (red line). Red spheres, cross-link sites on RNAP. Orientation as in (**e**).

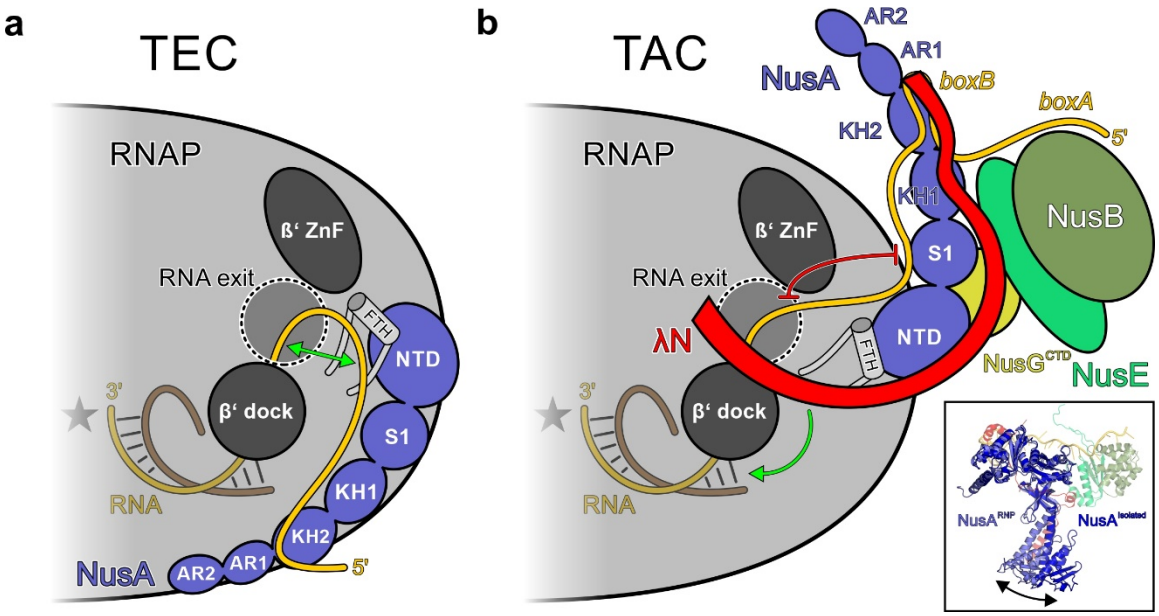
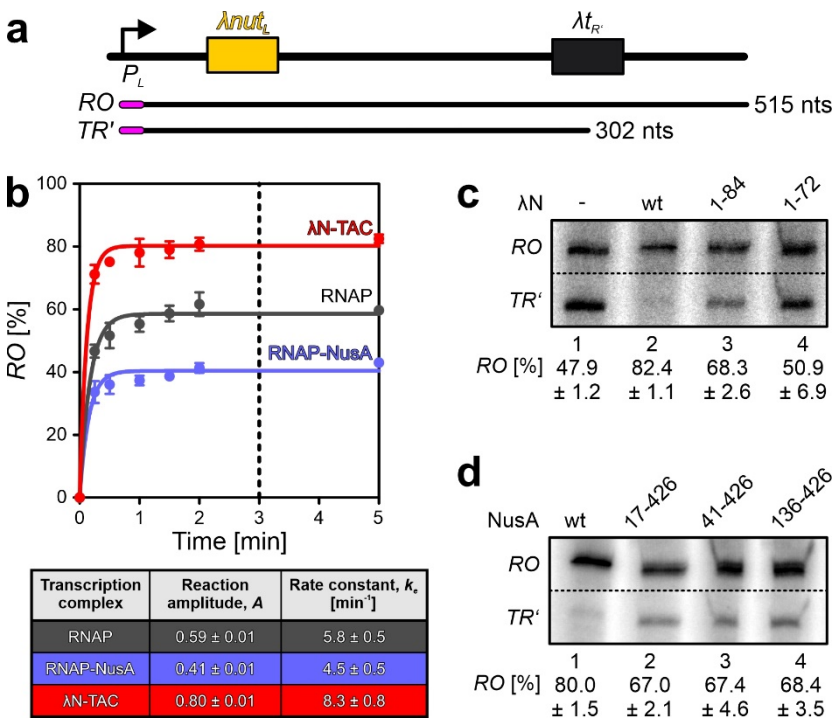



**Figure 5. Mechanism of λN-mediated processive antitermination. a**, Model for NusA activity in a transcription elongation complex (TEC) lacking λN. Nascent RNA is threaded between the RNAP β' dock domain and NusA. The NusA NTD can present the upstream branch of a terminator hairpin for base-pairing with the downstream branch emerging from the RNA exit channel (green double arrow), thus supporting intrinsic termination. **b**, Organization of the λN-based transcription antitermination complex (TAC). Repositioning of NusA and associated factors by λN relocates the NusA S1 domain and displaces the β flap tip, thus opening a gate for the nascent RNA to approach NusA S1. The latter is not blocked by *nut* RNA and can thus sequester the upstream branch of a terminator hairpin during emergence of the downstream branch from the RNA exit channel (red symbol). Additionally, the λN C-terminus clamps RNAP elements around the RNA exit channel, thus indirectly stabilizing the DNA:RNA hybrid (green arrow) and preventing RNA slippage/hybrid melting. Inset, structural comparison of NusA in the $\lambda N^{1-84}$-$NusA^{\Delta AR2}$-NusB-NusE-*nut* RNP and isolated $NusA^{100-426}$ with an NTD modeled according to the orientation of the NTD-S1 connector helix (dark blue) after superposition according to the S1-KH1-KH2 regions. Double arrow in inset, flexible positioning of the NTD with respect to the remainder of NusA. Rotated 45° about the horizontal axis (top to front) relative to Fig. 1b, left panel.




| Transcription complex | Reaction amplitude, *A* | Rate constant, $k_e$ [min$^{-1}$] |
|---|---|---|
| RNAP | 0.59 ± 0.01 | 5.8 ± 0.5 |
| RNAP-NusA | 0.41 ± 0.01 | 4.5 ± 0.5 |
| λN-TAC | 0.80 ± 0.01 | 8.3 ± 0.8 |

**Figure 6. *In vitro* functional analysis of antitermination activity. a**, Scheme of the DNA template used for transcription assays and the observed products. *RO*, run-off product; *TR'*, product terminated at $\lambda t_{R'}$; magenta regions, radioactively labeled initial transcripts. **b**, Top, time courses of transcription by the indicated machineries, showing that product levels do not significantly change after three minutes (dashed line). Data represent means ± SEM of three independent experiments. Bottom, data were fit to a first-order reaction (fraction *RO* = *A*{1-exp(-$k_e$ t)}; *A,* amplitude of the reaction; $k_e$, apparent first-order rate constant of transcription elongation; t, time). **c**,**d**, Transcription assays monitoring antitermination efficiency at three-minute time points by TACs bearing the indicated λN (**c**) or NusA (**d**) variants. Quantified data represent means ± SEM of three independent experiments. Lanes in (**c**) or (**d**) are from the same gel, but duplicate lanes and sections between *RO and TR'* have been removed (dashed lines, cut/paste sites). Gels shown are representative examples of three repetitions.

## 15.2 Publication II

Krupp F*, Said N*, Y-H Huang*, Loll B, Bürger J, Mielke T, Spahn CMT, Wahl MC (2019) Structural Basis for the action of an all-purpose transcription anti-termination factor. *Molecular Cell,* Volume 74, Issue 1, 4 April 2019, Pages 143-157.e5 (2019)

https://doi.org/10.1016/j.molcel.2019.01.016

* denotes first authors

## 15.3 Publication III

Kraushar M. L.*, Krupp F., Harnett D., Turko P., Ambrozkiewicz M. C., Sprink T., Imami K., Günnigmann M., Zinnall U., Vieira-Vieira C.-H., Schaub T., Münster-Wandowski A., Bürger J., Borisova E., Yamamoto H., Rasin M.-R., Beule D., Landthaler M., Mielke T., Tarabykin V., Landthaler M., Kramer G., Vida I., Selbach M., and Spahn C. M. T.. The architecture of protein synthesis in the developing neocortex at near-atomic resolution reveals Ebp1-mediated neuronal proteostasis at the 60S tunnel exit. Molecular Cell, Volume 81, Issue 2, 21 January 2021, Pages 304-322.e16 (2020)

https://doi.org/10.1016/j.molcel.2020.11.037

* denotes first author

# The architecture of protein synthesis in the developing neocortex at near-atomic resolution reveals Ebp1-mediated neuronal proteostasis at the 60S tunnel exit

Matthew L. Kraushar,[1,12,*] Ferdinand Krupp,[1] Dermot Harnett,[2] Paul Turko,[3] Mateusz C. Ambrozkiewicz,[4] Thiemo Sprink,[1] Koshi Imami,[5] Manuel Günnigmann,[6] Ulrike Zinnall,[5] Carlos H. Vieira-Vieira,[5,7] Theres Schaub,[4] Agnieszka Münster-Wandowski,[3] Jörg Bürger,[1,8] Ekaterina Borisova,[4] Hiroshi Yamamoto,[1] Mladen-Roko Rasin,[9] Dieter Beule,[2,10] Thorsten Mielke,[8] Victor Tarabykin,[4] Markus Landthaler,[5] Günter Kramer,[6] Imre Vida,[3] Matthias Selbach,[5] Christian M.T. Spahn[1,11,*]

[1]Institute of Medical Physics and Biophysics, Charité-Universitätsmedizin Berlin, 10117 Berlin, Germany

[2]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, 10115 Berlin, Germany

[3]Institute for Integrative Neuroanatomy, Charité-Universitätsmedizin Berlin, 10117 Berlin, Germany

[4]Institute of Cell Biology and Neurobiology, Charité-Universitätsmedizin Berlin, 10117 Berlin, Germany

[5]Max Delbrück Center for Molecular Medicine, 13092 Berlin, Germany

[6]Center for Molecular Biology of Heidelberg University (ZMBH) and German Cancer Research Center (DKFZ), DKFZ-ZMBH Alliance, D-69120 Heidelberg, Germany

[7]Humboldt Universität zu Berlin, Faculty of Life Sciences, 10117 Berlin, Germany

[8]Max Planck Institute for Molecular Genetics, Microscopy and Cryo-Electron Microscopy Service Group, 14195 Berlin, Germany

[9]Department of Neuroscience and Cell Biology, Rutgers University-Robert Wood Johnson Medical School, Piscataway, NJ, 08854 USA

[10]Core Unit Bioinformatics, Berlin Institute of Health, 10117 Berlin, Germany

[11]Lead Contact

[12]Current address: Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany

*Correspondence: matthew.kraushar@molgen.mpg.de (M.L.K.), christian.spahn@charite.de (C.M.T.S.)

**SUMMARY**

Protein synthesis must be finely tuned in the nervous system, and represents an essential feature of neurodevelopmental gene expression. However, the architecture of ribosomal complexes in the developing mammalian brain has not been analyzed at high resolution. This study investigates the architecture of ribosomes *ex vivo* from the embryonic and perinatal mouse neocortex, revealing Ebp1 as a high-occupancy, cytoplasmic, 60S peptide tunnel exit binding factor during active protein synthesis at near-atomic resolution by multiparticle cryo-electron microscopy. Ribosome profiling analysis with Ebp1-selective and Ebp1-knockdown conditions in neuronal cells demonstrated that Ebp1's 60S occupancy is the highest at the start codon, regulating start codon initiation, with persistent high occupancy during elongation of N-terminal peptides proteome-wide. Ebp1 binding to the 60S tunnel exit then reaches two distinct steady-states throughout elongation until the stop codon, contingent on translocon signal peptide that recruits SRP/Sec61, which putatively displaces Ebp1. In neurons, Ebp1 knockdown particularly impacts the synthesis of membrane-targeted cell adhesion molecules, measured by pSILAC/BONCAT mass spectrometry. Concordantly, *in vivo* embryonic Ebp1 knockdown results in dysregulation of neocortical neuron morpholgy. Our findings reveal Ebp1 as a central component of neocortical protein synthesis, and the 60S peptide tunnel exit as a focal point of gene expression control in the molecular specification of neuronal morphology during development.

**KEYWORDS:** Ebp1, ribosome, proteostasis, neocortex, neurons, development, cryo-EM, pSILAC, click chemistry, mass spectrometry, ribosome profiling, selective ribosome profiling

## INTRODUCTION

Proteostasis, the fine-tuned balance of protein homeostasis, is fundamental in establishing the molecular landscape of the nervous system. The demand for spatially targeted and precisely timed protein synthesis is exceptionally high in mammalian nervous system development, where amorphous neural stem cells generate intricate neuronal morphology through targeted gene expression (Holt et al., 2019; Jayaraj et al., 2019; Jung et al., 2014) . This is particularly true in the evolutionarily advanced mammalian neocortex, the central neuronal circuit of complex cognition in the brain (Silbereis et al., 2016). Concordantly, the nervous system is uniquely susceptible to abnormal proteostasis, a major driver of neurodevelopmental and neurodegenerative disease (Bosco et al., 2011; Kapur et al., 2017; Sossin and Costa-Mattioli, 2018). How proteostasis is achieved, therefore, stands as a crucial question towards understanding neurogenesis in the neocortex.

The neurogenic phase of stem cell maturation in neocortical development follows a trajectory largely conserved across mammalian species (DeBoer et al., 2013; Molyneaux et al., 2007) (**Figure 1A**). Neural stem cells (NSCs) lining the lateral cortical ventricular zone initially divide symmetrically to expand the cellular pool. NSC divisions then transition to yield newly born neurons, which progressively and sequentially undergo superficial migration, ultimately forming a layered cortical plate composed of structurally and functionally distinct neurons. In mice, lower layer neocortical neurons appear at approximately embryonic day 12.5 (E12.5), with the switch to upper layer neurogenesis at E15.5. By postnatal day 0 (P0), neurogenesis is largely complete, with ongoing ventricular stem cell divisions yielding glial cells. The elaboration of intricate neuronal morphology during this developmental window requires tight regulation of the neurite outgrowth and synaptic proteome (Holt et al., 2019; Jung et al., 2014), a fine-tuned balance of membrane proteins like cell adhesion molecules that establish neuronal connectivity (de Wit and Ghosh, 2016).

Analysis of the molecular landscape in the developing neocortex has largely focused on transcriptional regulation (Lein et al., 2017; Silbereis et al., 2016), with the neocortical transcriptome coming into focus recently at the single-cell level (scRNAseq) (Nowakowski et al., 2017; Telley et al., 2019; Yuzwa et al., 2017). However, the ultimate output of gene expression is protein, and bridging the neocortical transcriptome to proteome is the current challenge. The ribosome is the gatekeeper of the proteome, poised at the final essential step of gene expression as the macromolecular hub of protein synthesis, at the crossroads of gene expression in cellular proliferation, differentiation, and disease (Kraushar et al., 2016; Mills and

Green, 2017; Shi and Barna, 2015). However, the architecture of ribosomal complexes and proteostasis control in neocortical development remain unknown.

In this study, we analyze the molecular architecture of native ribosome complexes from the mammalian neocortex during developmental neurogenesis at near-atomic resolution. We find that the ErbB3-binding protein 1 (Ebp1) participates in high occupancy binding to the 60S subunit of both actively translating and non-translating ribosomes through high-affinity interactions with the peptide tunnel exit surface in the embryonic and perinatal neocortex. Ebp1's function in protein synthesis during nervous system development is unknown. Ebp1 enrichment across developmental stages scales directly with dynamic ribosome levels and is cell type-specific: dominantly expressed in early-born NSCs, compared to later-born NSCs and post-mitotic neurons – in contrast to other exit tunnel cofactors. Ebp1•ribosome interaction occurs in the cytoplasm of NSCs in the neocortical ventricular zone at early embryonic stages when ribosomal complex levels are highest, and persists in post-mitotic neurons of the expanding cortical plate as steady state ribosome levels decline. With Ebp1-selective ribosome profiling, we show that Ebp1's highest occupancy on actively translating ribosomes is during start codon initiation, with knockdown resulting in ribosome accumulation at the AUG. Ebp1 binding is maintained during elongation, especially during synthesis of N-terminal peptides throughout the proteome, until translocon signal sequences for membrane targeting emerge and putatively engage competition for a common binding surface with SRP/Sec61. Ebp1 maintains neuronal proteostasis, especially impacting the synthesis of cell adhesion molecules as measured by pSILAC/BONCAT mass spectrometry. Concordantly, *in vivo* embryonic Ebp1 knockdown selectively in early-born neocortical NSCs results in dysregulated morphology during neuronal maturation. This study is the first near-atomic resolution analysis of protein synthesis in the nervous system, positioning Ebp1 and the 60S peptide tunnel exit as a focal point of gene expression control during neocortical neurogenesis.

## RESULTS

### Ebp1 is a high occupancy translation cofactor proportional to dynamic ribosome levels during neocortex development

To analyze the architecture of neocortical ribosome complexes across development, we first optimized a protocol to purify actively translating ribosomes *ex vivo* rapidly and stably without the use of chemical inhibitors that bias its conformational state, capturing the full repertoire of integral translation cofactors. Initial analytical sucrose density gradients revealed that global ribosome levels are dynamic across neocortex development (**Figure 1B**). High levels of 80S

ribosomes (monosomes) and chains of multiple 80S actively translating mRNA (polysomes) predominate at E12.5-E14, transitioning to a lower steady state from E15.5-P0 (**Figure 1C**). This decrease is not wholly accounted for by the availability of individual subunits in the cytoplasm, as 40S-60S levels decrease marginally. Thus, ribosomal complexes exist at elevated levels during early neocortical neurogenesis, and transition to a lower steady state at later stages.

We next performed mass spectrometry (MS) analysis of 80S and polysomes, in addition to corresponding input total lysates, across neocortex developmental stages. Sample reproducibility was observed in hierarchical clustering of the MS data (**Figure S1**). Results from the neocortical polysome MS are shown in **Figure 1D**, comparing protein levels at E12.5 with each subsequent developmental stage. As expected, core ribosomal proteins (RPs) are the most enriched proteins in polysomes, including RPs of the large 60S (Rpl) and small 40S (Rps) subunits. Translation-associated proteins (GO:0006417) associate to varying degrees with polysomes throughout development. Unexpectedly, we observed Ebp1 co-purifying at levels approaching the RPs themselves in polysomes, higher than any other translation-associated protein. Ebp1 is metazoan-specific, and was largely studied in the context of cancer (Nguyen et al., 2018). Ebp1 was observed to play only a niche role in protein synthesis: promoting internal ribosome entry site (IRES) dependent translation of a specific viral mRNA (Pilipenko et al., 2000), and suppressing eIF2a phosphorylation in conditions of cellular stress (Squatrito et al., 2006), by unknown mechanisms. Thus, we were intrigued by Ebp1's exceptionally high polysome enrichment, and observed a similarly robust association with 80S complexes (**Figure S2A**). Furthermore, Ebp1 is among the most abundant proteins measured in total neocortical lysates across development (**Figure S2A**).

To examine the global trajectory of neocortical *Ebp1* and core ribosomal protein gene expression across development, we next analyzed total lysates by RNAseq (**Figure 1E**). *Ebp1* mRNA steadily decreases after E12.5, while *Rpl* and *Rps* mRNA decreases lag behind at E17. However, corresponding MS measurements revealed total Ebp1 protein levels decline abruptly at E15.5 along with total Rpl and Rps levels in the neocortex, suggesting their protein levels are regulated in concert, with protein changes anticipating mRNA changes for the RPs. MS findings were confirmed by Western blot analysis of total neocortex lysates (**Figures 1F and S3A-B**), showing that levels of Ebp1 and the RP uL30 are highest in the early prenatal neurogenic period, and decrease at E15.5, with the lowest Ebp1 levels in the postnatal period. The timing of Ebp1, Rpl, and Rps total protein decreases coincides with the timed decrease of global ribosome levels measured by density gradient fractionation (**Figures 1B-C**).

Ebp1 has been previously reported as a full-length 48kDa protein ("p48"), and a 42kDa isoform ("p42") generated by *Ebp1* mRNA splicing (Liu et al., 2006). Western blot findings with a C-terminal targeting antibody ($Ebp1^{CT}$) recognizing both long and short isoforms (**Figures 1F and S3A**), and with a N-terminal specific antibody ($Ebp1^{NT}$) recognizing only full-length Ebp1 (**Figures 1F and S3B**), compared to polyhistidine-tagged full-length recombinant Ebp1 (Ebp1-His), showed that the dominant isoform of Ebp1 in neocortical development is full-length.

The core of the eukaryotic 80S ribosome is a macromolecular machine consisting of ~79 RPs on a scaffold of 4 rRNAs, with translation-associated proteins transiently binding to catalyze and modulate ribosomal functions. We next calculated the stoichiometry between translation-associated cofactors and core RPs in neocortical ribosomes across development, in addition to their balance in total steady-state (**Figures 1G and S2B**). In contrast to the majority of core RPs, translation-associated proteins are maintained at a wide range of total steady-state levels, and their association with purified ribosomal complexes tend to be substantially sub-stoichiometric. At all stages, Ebp1's total steady-state level is similar to RPs (0.7-1.5 total stoichiometry), and is the 80S and polysome cofactor with the highest association; in the range of 0.4-0.6 80S stoichiometry, and 0.2-0.3 polysome stoichiometry.

Given that Ebp1 is sub-stoichiometric in 80S and polysome complexes while near-stoichiometric in total, a substantial proportion of its total levels are likely extra-ribosomal. To test this, we next measured the balance of ribosome-associated Ebp1 compared to "free" extra-ribosomal Ebp1 in neocortical development (**Figures 1H and S3C-D**). Results showed that, indeed, at each developmental stage the majority of Ebp1 is extra-ribosomal in contrast to the RP uL30 (**Figure S3E**) – consistent with Ebp1 being a ribosome cofactor rather than a core protein. Furthermore, comparing E12.5 with subsequent stages showed a decrease of Ebp1 in ribosomal fractions beginning at E15.5 to P0 that mirrors changes in the RP uL30 (**Figure 1H**). In contrast, free extra-ribosomal Ebp1 is maintained over time.

Taken together, these data suggest that Ebp1 associates with both neocortical 80S and polysomes, maintaining a high and consistent stoichiometry, in concert with decreasing ribosome levels across development. The unusual abundance of Ebp1-ribosome association suggests that Ebp1 may play a more central role in neocortical translation, rather than niche for a small subset of transcripts or during transient conditions as previously reported (Pilipenko et al., 2000; Squatrito et al., 2006).

**Ebp1 is enriched in early-born NSCs and localizes throughout the cytoplasm**

As the above observations were in bulk neocortex tissue, we next analyzed the cellular expression of Ebp1 in neocortex development with scRNAseq data (Telley et al., 2019) measuring the transcriptome of early and late-born NSCs maturing into lower and upper layer neurons, respectively (**Figure 2A**). Strikingly, *Ebp1* mRNA is particularly enriched in early-born NSCs, with levels decreasing abruptly during both neuronal differentiation, and in the later-born NSC pool. Likewise, *Rpl* and *Rps* mRNA levels decline with differentiation; however, the *Rpl* and *Rps* expression patterns are more generic in NSCs, regardless of birthdate.

Consistent with the scRNAseq data, immunohistochemistry analysis across developmental stages (**Figures 2B and S4A**) demonstrated particularly high cytoplasmic Ebp1 protein enrichment in the ventricular zone (VZ) and nascent cortical plate (CP) at E12.5-E14. Ebp1 is persistent in maturing neurons laminating the CP at later stages, albeit at lower levels. Interestingly, Ebp1 enrichment in the P0 VZ that contains early gliogenic progenitor cells (DeBoer et al., 2013; Molyneaux et al., 2007) is substantially lower than in the neurogenic E12.5 VZ. This enrichment in neural progenitors may relate to an observation that Ebp1 is particularly enriched in neurons compared to astroglia in the postnatal period (Ko et al., 2017). Thus, Ebp1 enrichment is specific to cell type, differentiation status, and NSC birthdate in the neocortex.

To assess subcellular Ebp1 localization at higher resolution, we next analyzed the neocortex at E12.5, E15.5, and P0 by immuno-electron microscopy (immuno-EM), probing for Ebp1 with both $Ebp1^{NT}$ (**Figure 2C**) and $Ebp1^{CT}$ (**Figure S4B**) antibodies. Quantification of Ebp1 immunogold labeling demonstrated almost exclusively cytoplasmic signal (**Figures 2D and S4C**), occurring in clusters throughout the cytoplasm most abundantly in early-born NSCs and recently born neurons in the VZ, with lower cytoplasmic levels in CP neurons and VZ glial progenitors at later stages. Ebp1 was largely absent from nuclei, including the nucleolus and nuclear membrane, nor observed in mitochondria, strict proximity to the endoplasmic reticulum, or plasma membrane. Thus, Ebp1 may bind the ribosome in the cytoplasm, rather than as a subunit assembly or export factor in the nucleus, consistent with prior observations (Bradatsch et al., 2007).

Ebp1 was also observed in dendrites of maturing neurons at P0 (**Figure 2C**), suggesting Ebp1 localizes throughout cytoplasmic compartments as neocortical NSCs mature into neurons. We next sought to visualize Ebp1 localization during the progressive differentiation of early-born neocortical NSCs into post-mitotic neurons undergoing neurite outgrowth. Primary cultures were prepared from the E12.5 neocortex of *Nex:Cre;Ai9* mice (Turko et al., 2018), which label post-mitotic pyramidal neurons, followed by immunohistochemical analysis of Ebp1 expression (**Figure 2E**). Ebp1 is enriched in cytoplasmic foci colocalizing with Nestin labeling in NSCs at

div 0, and persists in differentiating Nex-positive neurons at div 2-4. Ebp1 puncta are visualized in robust neuronal protrusions by div 5, and particularly apparent with further magnification of neurites and growth cones, including the most distal aspects of extending processes, consistent with prior observations in hippocampal neurons (Ko et al., 2017; Kwon and Ahn, 2011).

**Ebp1 binds the 60S peptide tunnel exit in actively translating and inactive ribosomes**

To analyze the architecture of neocortical ribosome complexes and visualize the physiologic binding mode of Ebp1 at near-atomic resolution, 80S and polysomes were purified by sucrose density gradient fractionation from P0 neocortex lysates, pooled together, and frozen on grids for cryo-electron microscopy (cryo-EM). Micrographs confirmed the presence of both 80S and polysome complexes in the sample (**Figure 3A**). High-resolution cryo-EM data collection (**Figure S5**) and initial single-particle reconstruction yielded a map of the complete 80S, along with extra-ribosomal density (red) adjacent to the 60S peptide tunnel exit (**Figure 3B**). Fitting the crystal structure of Ebp1 (Kowalinski et al., 2007; Monie et al., 2007) to the extra-ribosomal density unequivocally identified Ebp1 in complex with the neocortical 60S. Robust density was present for nearly the entire N-terminus, identifying the full-length isoform of Ebp1 is bound. Ebp1 forms a concavity above the TE vestibule with a porous interface, including gaps (~28 Å at the widest point) that may permit peptide chain exit. The mouse neocortex 80S core structure was found to be otherwise highly conserved with previously solved human (Behrmann et al., 2015) and rabbit (Flis et al., 2018) structures, with Ebp1 density the greatest distinction.

To disentangle the ribosome conformational states bound by Ebp1, we proceeded with hierarchical multiparticle sorting and 3D classification of both large and small scale heterogeneity intrinsic to the data (Behrmann et al., 2015; Loerke et al., 2010) (**Figure S6**). Ribosome complexes in both the rotated and classical conformations were first sorted, including populations with (1) eEF2 and (2) eEF2+P/E tRNA in the rotated state, and populations with (3) A/A+P/P tRNAs, (4) E/E tRNA, and (5) without tRNAs in the classical state. Within each of these five states, a strategy of modified focused classification was utilized to separate sub-states with and without Ebp1, yielding ten total classes. Across all states, Ebp1 was bound to 48% of ribosomes, with ~50% binding to each of the five sub-states.

We proceeded with high-resolution refinement of Ebp1-bound and unbound populations in the inactive rotated state with eEF2 (3.1 Å global resolutions), and the active classical state with A/A+P/P tRNAs (3.3 Å global resolutions) (**Figure 3C**). When Ebp1 was bound, the structural conformations of both Ebp1 and the 60S binding surface were identical between active and inactive ribosomes. These data indicate Ebp1 binds to both actively translating and

non-translating neocortical ribosome states with approximately equal probability, high occupancy, and identical conformations.

The near-atomic resolution of our data (**Figures S5 and S7**) permitted modeling of the entire neocortical Ebp1•60S complex. **Figure 3D** visualizes the peptide tunnel exit (TE) surface in electrostatic proximity to Ebp1, including four RPs (eL19, uL23, uL24, uL29) and three rRNA helices (H24, H53, H59). An aerial view of the Ebp1 footprint over the TE surface highlights the 60S RP residues and rRNA nucleosides making electrostatic interactions with Ebp1 (**Figure 3E**), demonstrating that Ebp1 contacts the immediate TE surface. The neocortical Ebp1•60S complex establishes previously unassigned functions to Ebp1 structural domains (**Figure 3F**; adapted from (Kowalinski et al., 2007)), where binding by Ebp1's insert domain and α5 helix positions β-sheets 1, 3, 4, 5, 7, and 13 directly over the TE.

**Ebp1 binding requires a conserved 60S helix H59-H53 swinging latch mechanism**

Multiparticle sorting of our data into Ebp1-bound and unbound states enabled identification of 60S structural changes facilitating Ebp1 interactions with an internal negative control (**Figure S6**). Dynamic interactions occur with helix H59 of 28S rRNA (**Figures 4A and S8A**). In the Ebp1-bound state, the tip of H59 undergoes a backbone rearrangement enabled by a 235° flip of H59 G-2690, releasing contact with H53 G-2501, G-2502, and C-2513 as seen in the canonical unbound state – resulting in H59 G-2690 transitioning to intra-helical base stacking interactions. This "swinging latch" mechanism further includes a 73° flip of H59 U-2687, with the base reaching into a pocket of Ebp1's insert domain (**Figure S8B**), locking Ebp1 into position. This particular movement of H59 U-2687 was previously observed for binding of the yeast nuclear export (Bradatsch et al., 2007) and peptide tunnel quality control (Greber et al., 2016) factor Arx1 to the 60S – thus representing a conserved binding mechanism. However, unlike Arx1, Ebp1 binding does not appear to require stabilization by rRNA expansion segment ES27 on the solvent side (Greber et al., 2016) – which we confirmed by 3D-variability calculation (Penczek et al., 2006) and independent map reconstructions with alternative methods (Punjani et al., 2017; Scheres, 2012) (data not shown) – thus representing a distinction in its binding mode.

Protrusion of H59 U-2687 into the insert domain of Ebp1 is stabilized by hydrogen bonds with the backbones of Y-255 and G-256, in addition to S-267, and $\pi$-stacking interactions with F-266 (**Figures 4B and S8C**). Reorientation of H59 brings the adjacent U-2688 in proximity to R-271 yielding further hydrogen bond stabilization. The RP eL19 confers stability to both the

flexible loop and α6 helix in Ebp1's insert domain (**Figures 4B-C and S8C-D**). Hydrogen bonding occurs between eL19 N-34 and N-36 to the backbone and R-group of Ebp1 Y-255, respectively, in addition to interaction between eL19 Q-40 and Ebp1 R-263.

The Ebp1 α6 and α8 binding interfaces at the TE rim further include 28S rRNA H53 and uL23 (**Figures 4C and S8D**). The intramolecular interaction between Ebp1 K-258 and Y-255 is reinforced by H53 C-2505 and uL23 E-84. Ebp1 α6/8 interactions with uL23 further include hydrogen bonds between Ebp1 methionines 259 and 291 with uL23 K-88 and E-91, respectively, while Ebp1 K-287 is coordinated by uL23 L-147, D-148, and N-151, with N-151 also contacting Ebp1 R-290. Contact between Ebp1 and uL29, in contrast, is less robust and mediated by van der Waals interactions.

Finally, 5.8S rRNA H24 and uL24 tether the Ebp1 α5 domain on the opposite side (**Figures 4D and S8E**). Ebp1 K-211 is in proximity to H24 A-383 and C-384, while Ebp1 D-207 contacts the backbone of uL24 G-92 and flanking N-91 and T-93.

**Ebp1•60S binding is incompatible with simultaneous binding of other eukaryotic peptide tunnel exit cofactors**

Ebp1, Metap2, and Arx1 share a common β-α---α-β insert domain that facilitates 60S binding, a "pita-bread" β6 fold motif positioned over the peptide TE, and a solvent-side α4 motif (**Figures 4E-G**). In the event of Ebp1 or Metap2 binding, emerging peptide chain would encounter a deep, strongly electronegative pocket; however, Metap2 β6 fold residues catalyzing aminopeptidase activity (Nonato et al., 2006) are absent in Ebp1 (Kowalinski et al., 2007; Monie et al., 2007), rendering Ebp1 catalytically inactive. Furthermore, the Ebp1 α5 domain facilitating electrostatic contacts with H24 and uL24 is absent in Metap2 (**Figure S8F**); however, a Metap2•60S structure has not yet been solved. In contrast, the yeast Arx1 pita-bread fold binds FG repeat domains of nuclear membrane nucleoporins (Bradatsch et al., 2007) and threads Rei1 into the peptide tunnel to probe the 60S as a quality-control step (Greber et al., 2016) – with contrasts to the metazoan Ebp1 described previously (Bradatsch et al., 2007). Thus, the binding of these distinct TE factors creates unique structural and electrochemical environments for emerging peptide chains.

The binding of Ebp1 would be sterically incompatible with the simultaneous docking of other 60S TE cofactors, competing for limited real estate surrounding an emerging nascent peptide chain (**Figure 4H**). The footprint of Ebp1 is shown superimposed on the footprints of Metap2 (Nonato et al., 2006) and Arx1 (Greber et al., 2016), in addition to: the ER targeting machinery SRP (Kobayashi et al., 2018) and Sec61 (Voorhees et al., 2014); the Ltn1-NEMF

ubiquitin ligase complex (Shao et al., 2015); the N-terminal acetyltransferase NatA (Knorr et al., 2019); the ribosome-associated complex (RAC) coupling nascent chain elongation and folding (Zhang et al., 2014); the nascent polypeptide-associated complex (NAC) preventing ER mistargeting and protein aggregation (Gamerdinger et al., 2015; Shen et al., 2019); and Ttc5, a negative-feedback sensor of tubulin synthesis (Lin et al., 2020). Ebp1 is among the most abundant of these TE factors in the neocortex (**Figure 4I**), comparable to NAC, and the Hspa8 subunit of RAC. Early in development, however, Ebp1 is the most highly associated with 80S and polysomes; later in development, as Ebp1 and ribosome levels decline, only Hspa8 supersedes Ebp1 in 80S and polysomes. Furthermore, the neocortical cell-type and temporal specificity of Ebp1 enrichment is in contrast to some TE cofactors, such as Ttc5, while similar to others, such as RAC (**Figure S9**). Dynamic enrichment of Ebp1 vs. other TE cofactors may represent the differential regulation of protein synthesis in response to the unique demands of particular stages in neurogenesis.

**Ebp1 binds the 60S with high affinity and active turnover**

Given the competition for a common TE surface, we next sought to measure the affinity and dynamics of Ebp1•60S binding. We first confirmed the specificity of Ebp1 for the 60S subunit of both neocortical and rabbit reticulocyte (RRL) ribosomes, in the absence of mRNA (**Figures S10A-B**). We then determined the relative affinity range of Ebp1•60S binding (**Figures 4J and S10C**). The curve best fit to data ($r^2$=0.99) indicates Ebp1 reaches a $K_{d(app)}$ at ~124 nM, with saturated Ebp1•60S binding at ~200 nM, relative to 100 nM 60S. These data indicate Ebp1 binds the 60S with high affinity, reaching saturation at ~2-fold excess Ebp1 over the 60S. The 0.7-1.5 stoichiometry of total steady-state Ebp1 compared to the core ribosome measured by MS (**Figures 1G and S2B**) along with a high relative affinity may account for the high degree of Ebp1•ribosome association measured in the neocortex across development.

We next texted whether Ebp1 binding undergoes dynamic turnover by reconstitution of the following binding conditions in parallel: (1) saturating levels of recombinant Ebp1-His in the presence of rabbit 60S, (2) RRL containing native Ebp1, and (3) saturating Ebp1-His added to RRL (**Figure 4K**). Native Ebp1 in RRL co-pelleted with the ribosome as did Ebp1-His to the 60S, undergoing dynamic binding demonstrated by the nearly complete turnover of native Ebp1 with saturating Ebp1-His. The dynamic turnover of the Ebp1•60S binding mode may permit emerging peptide chain motifs to recruit other TE cofactors and displace Ebp1.

**Start codon initiation and N-terminal peptides are regulated proteome-wide by Ebp1**

356

357 To interrogate the specific translation activity of Ebp1-bound ribosomes at high-resolution
358 proteome-wide, we performed Ebp1-selective ribosome profiling (SeRP) (Schibich et al., 2016)
359 – deep sequencing of ribosome-protected mRNA fragments – in mouse neuronal cultures
360 (Neuro2a). We first confirmed that, like the neocortex, Neuro2a cells dominantly express full-
361 length Ebp1 (**Figure S11A**), which associates with 60S, 80S, and polysomes (**Figures S11B-C**).
362 Comparison of overall mRNA enrichment in the Ebp1•ribosome immuno-precipitation (IP)
363 interactome with the total translatome demonstrated a high degree of correlation (**Figure 5A**),
364 consistent with high-occupancy ribosome binding by Ebp1 during the synthesis of a large, highly
365 generalized cohort of transcripts. Highlighting the subset of mRNAs with ≥ 1.5-fold enrichment in
366 Ebp1•ribosome IP vs. the ribosome total demonstrated a cellular compartment-specific
367 translation preference (**Figure 5B**). Ebp1•ribosome complexes are more likely to be engaged in
368 the translation of proteins localizing to the nucleus and other intracellular compartments, while
369 disfavoring membrane-targeted proteins of the endoplasmic reticulum (ER).

370 Proteome-wide metagene analysis (**Figure 5C**) shows that, on average, Ebp1•ribosome
371 interaction is highest during initiation at the start codon. Proteins ultimately targeted to distinct
372 cellular sub-compartments during translation, including the cytoplasm and ER, share this
373 characteristic enrichment (**Figure 5D**) – consistent with highly comprehensive Ebp1•ribosome
374 transcript interaction at the AUG. During elongation in the coding sequence across protein
375 subsets, Ebp1 occupancy progressively declines until ~120 codons, when steady-state binding
376 is established and maintained until the stop codon.

377 However, for ER-targeted proteins, a distinct, lower steady-state of Ebp1•ribosome
378 interaction is established from ~70-120 codons into the coding sequence (**Figures 5D-E**).
379 Translation is targeted to the ER by 60S TE binding proteins SRP and Sec61, which facilitate
380 recruitment and membrane insertion, respectively, of ribosome-nascent chain complexes by
381 interaction with emerging translocon signal sequences and transmembrane domains (TMD)
382 (Jan et al., 2014; Schibich et al., 2016). Such membrane-targeting domains commonly emerge
383 within this N-terminal window, after traversing the mammalian 60S tunnel that is ~40 amino
384 acids in length (Jan et al., 2014). Alignment to the C-terminal codon of signal peptides
385 demonstrated that decreased Ebp1 interaction occurs ~40 amino acids downstream (**Figure**
386 **5E**). Such signal sequences upstream of a TMD lead to Ebp1 displacement before the first TMD
387 is translated, whereas the first TMD without an upstream signal sequence likewise results in
388 Ebp1 displacement after ~40 codons (**Figure 5F**) The competition between Ebp1 and



262

SRP/Sec61 for a common binding surface at the 60S TE (**Figure 4H**) suggests that Ebp1 is displaced by SRP/Sec61 when recruited by membrane-targeting nascent chain motifs.

To further interrogate the translation-specific function of Ebp1, we next performed ribosome profiling of Neuro2a cells in Ebp1 knockdown and control conditions (**Figure S11D**). Metagene plots demonstrated a redistribution of ribosomes towards the start codon with Ebp1 knockdown, in particular for ER-targeted, transmembrane, and signal peptide-containing proteins (**Figure 5G**). P-site alignment of ribosome-protected fragments and normalization to mapped reads per gene more precisely demonstrated the impact of Ebp1 knockdown on ribosome occupancy at the start codon, where ribosomes accumulate for both cytoplasmic and signal peptide-containing proteins in Ebp1 knockdown conditions (**Figures 5H-I**). Notably, evidence of ribosome accumulation was also observed for the elongation of early N-terminal amino acids in signal peptide-containing proteins. Given high-occupancy Ebp1•ribosome binding at the start codon, and differential occupancy during N-terminal peptide discrimination, the impact of Ebp1 depletion further demonstrates a role for Ebp1 during active protein synthesis at the initiation and elongation phases.

**Ebp1 regulates the morphology and proteome of the neuronal membrane**

Since we observed particularly high Ebp1 enrichment in early-born NSCs of the developing neocortex (**Figures 2A-D and S4A**), we next sought to study the cellular effect of Ebp1 knockdown in early-born NSCs during their maturation into neocortical neurons *in vivo*. *In utero* electroporation (IUE) of a *shEbp1* knockdown or control plasmid along with a CAG-GFP transfection reporter at E12 was followed by analysis at E16 during initial neurite outgrowth (**Figure 6A**). Analysis at E16 demonstrated increased branching of neuronal processes in *shEbp1* conditions compared to control, as normal pyramidal neuron projections include a single unbranched axon extending towards basal white matter tracts, along with an apical dendrite oriented towards the pial surface. Tracing the morphology of transfected neurons (**Figure 6B**) highlighted the impact of Ebp1 depletion on neurite outgrowth at various neurite lengths, with Sholl analysis (**Figure 6C**) demonstrating a significantly increased branch number in *shEbp1* conditions – an approximately two-fold increase for proximal segments. Importantly, this increased branching phenotype was rescued by co-electroporation of an Ebp1 overexpression plasmid (*oeEbp1*) along with *shEbp1*, with neuronal morphology tracing and branching analysis quantified as indistinguishable from control conditions.

To interrogate the potential function of Ebp1 in maintaining proteostasis during neuronal differentiation and neurite outgrowth, we again performed Ebp1 siRNA knockdown in Neuro2a

cells, and grew cultures in low serum conditions to induce maturation of neuronal morphology (Evangelopoulos et al., 2005). The impact of Ebp1 depletion on acute protein synthesis and chronic proteostasis in Neuro2a was measured by MS (Eichelbaum et al., 2012; Howden et al., 2013) with a combination of pulsed stable isotope labeling by amino acids in cell culture (pSILAC) (Schwanhäusser et al., 2009) and bioorthogonal noncanonical amino acid tagging (BONACT) (Dieterich et al., 2006). SILAC isotopes labeled all newly made proteins throughout Ebp1 knockdown for longitudinal proteome changes, while pulse labeling with a methionine analog (AHA) captured a snapshot of newly synthesized proteins at the nadir of Ebp1 levels (**Figure 6D**). Importantly, Ebp1 levels were below the quantification threshold in *siEbp1* conditions, confirming robust knockdown.

Results for both the pSILAC and pSILAC-AHA MS (**Figure 6E**) showed that in Ebp1 knockdown conditions, proportionately more proteins decrease compared to those with increased protein levels relative to control, suggesting Ebp1 generally enhances protein expression. Cell adhesion molecules (CAMs) were highly represented among Ebp1 regulated proteins (**Figure 6F**), such as L1cam, Mcam, Cadm1, and Cdh15, which play a critical role in neurite outgrowth and synaptogenesis (de Wit and Ghosh, 2016). Such membrane-targeted proteins may be particularly susceptible to Ebp1 depletion, given the role of Ebp1 during initiation and elongation at the N-terminus.

Proteins like L1cam were found to change in common between the pSILAC and pSILAC-AHA datasets, reinforcing that their regulation by Ebp1 is direct and protein synthesis-specific. Analysis of *L1cam* translation by Ebp1•ribosome complexes underscores the dynamic interactions of Ebp1 at the 60S TE during initiation and elongation of L1cam peptides (**Figure 6G**), with generally high-occupancy at the start and N-terminus that include fluctuations in binding, transitioning to generally low-occupancy downstream. Ebp1 knockdown redistributes ribosome occupancy in the *L1cam* coding sequence with a trend approximately opposite to Ebp1 binding, with the notable exception at the start, consistent with relatively elevated start codon occupancy with Ebp1 depletion. Given the sensitivity of N-terminal dynamics for proper membrane-targeted translation, the efficient translation of proteins like L1cam may be compromised in Ebp1 deficient conditions.

## DISCUSSION

Taken together, this study analyzes the architecture of protein synthesis in the developing neocortex at high-resolution, positioning Ebp1 among 60S TE cofactors to regulate neuronal proteostasis in the molecular specification of morphology during neurogenesis. With a

multidisciplinary approach, we demonstrate that Ebp1 is a chief component – rather than a niche regulator – of the protein synthesis machinery (**Figure 7**). Ebp1 participates in the initiation and elongation phases of translation in the neuronal cytoplasm, with high-occupancy 60S TE binding during start codon initiation and N-terminal peptide synthesis proteome-wide, reaching a binding equilibrium during elongation unless displaced by translocon targeting motifs. Ebp1 expression is cell-type and temporally specific, enriched in the early-born neural stem cell pool, in direct proportion to the transient abundance of ribosomal complexes at this early developmental stage. During NSC differentiation, Ebp1 particularly impacts the synthesis of CAMs that are essential components of the neuronal membrane and morphology.

The interaction of 60S TE cofactors exists in a dynamic equilibrium, competing for a common binding surface to sculpt proteins synthesized by a modular macromolecular machine (Deuerling et al., 2019; Kramer et al., 2019). While the regime of *Rpl* and *Rps* mRNA expression appears to follow generally elevated levels in all neocortical NSCs compared to their daughter neurons, there is a great diversity of TE cofactor expression patterns in the developing neocortex (**Figure S9**). Ebp1 is particularly enriched in early-born NSCs, similar to RAC subdomains, but in contrast to Metap2, Ltn1, or NAC. Modulating the balance of TE cofactors may be a key determinant of cell type-specific proteostasis, gatekeepers at the very moment a nascent protein emerges from the tunnel.

In our simple model (**Figure 7**), Ebp1 binding to the 60S in the cytoplasm may be a "default state" during translation initiation and early elongation proteome-wide before nascent chain exits the TE. Ebp1 binding with active turnover establishes a dynamic equilibrium throughout elongation, depending on competition from other TE cofactors if recruited by their associated nascent chain moieties. Future work will determine in more detail how Ebp1 acts in coordination with SRP (Chartron et al., 2016; Jan et al., 2014; Schibich et al., 2016) and NAC (Gamerdinger et al., 2015) to organize subcellular targeting and N-terminal processing. Ebp1 may have an initial competitive advantage, given its abundance in the neocortical cytoplasm relative to other TE factors, high affinity interaction with the 60S, and permissive binding requirements – including both translating and non-translating ribosomes.

Whether Ebp1's role in active and inactive complexes is linked or distinct remains unclear; for example, Ebp1 may play a role in 60S recycling for re-initiation, or protect a reserve of inactive, dormant ribosomes available to participate in translation. Exploring such possibilities may help clarify the phenotype of Ebp1 knockdown, which yields increased ribosome occupancy during both start codon initiation generally, and in the transition to elongation for signal peptide-containing proteins.

While this paper was in revision, a structure of Ebp1 bound to inactive 80S from HeLa cells was published (Wild et al., 2020), in agreement with our data for interactions surrounding the TE rim. However, the study shows rRNA expansion segment ES27L interacting with Ebp1 on the solvent-side, which we do not visualize as a binding requirement – a discrepancy of unknown significance, but a possible distinction in mouse brain ribosome expansion segment structure. Furthermore, Ebp1's potential direct interaction with nascent peptide chain and/or recruitment of other ribosome cofactors remains to be established. Since our cryo-EM analysis of native Ebp1•ribosome complexes includes ribosomes engaging with the entire translated proteome, nascent chain density is lacking entirely at the TE vestibule outside the tunnel, secondary to heterogeneity intrinsic in the data, and is a limitation of our study. Future studies in a more homogenous system will be required to interpret potential Ebp1-nascent chain interactions at high resolution.

Many of the proteins impacted by Ebp1 are membrane-associated, particularly cell-cell adhesion pathways regulating neuronal protrusions, with neocortical Ebp1 knockdown resulting in increased neurite branching. How Ebp1 depletion ultimately impacts protein output and membrane architecture is unclear, but may be multifaceted and interconnected if the 60S TE is compromised. Depletion of a global ribosome cofactor like Ebp1 could, for example, impact mRNA degradation coupled to translation (Pelechano et al., 2015), the efficiency of translation initiation and elongation, and/or protein degradation that may occur as a result of misfolding, mistargeting, and ER stress (Martínez et al., 2018).

Ribosomes locally translate mRNAs in neuronal protrusions (Hafner et al., 2019; Zappulo et al., 2017) where Ebp1 is also present (**Figures 2C and 2E**), and several Ebp1-regulated proteins are predominantly translated in neuronal protrusions (**Figure S12**). In conjunction with Ebp1 enrichment, regulation may further derive from shifts in global ribosome levels that transition during neocortical neurogenesis (**Figure 1**), in line with previous observations of RP downregulation in the mouse forebrain during neural tube closure between E8.5-E10.5 (Chau et al., 2018), and further between E13-P0 (Kraushar et al., 2015). Global shifts in steady-state ribosome levels may reflect the dynamics of cellular homeostasis (Delarue et al., 2018; Mills and Green, 2017; Sinturel et al., 2017) in neocortex development. The subcellular and coordinated actions of Ebp1•ribosome complexes are interesting directions for future study.

Transcriptional control has been the principal focus in gene expression analysis during neocortex development (Silbereis et al., 2016), and recent excellent work has advanced this analysis to the single-cell level (Nowakowski et al., 2017; Telley et al., 2019; Yuzwa et al., 2017).

However, while these studies assign transcriptional signatures to cell subtypes, they also strongly suggest that highly generic gene expression programs are refined by successive layers of regulation (Cadwell et al., 2019), such as post-transcriptional mechanisms and environmental signals (Kraushar et al., 2016). Our present work indicates binding events at the 60S TE during translation constitute a locus of control during neurogenic gene expression.

**ACKNOWLEDGMENTS**

We apologize to the authors of key papers who we could not cite due to space limitations. We are particularly grateful to Rainer Nikolay, Anett Unbehaun, Tatyana Budkevich, Justus Loerke, and Dennis Kwiatkowski for fruitful scientific discussions and technical support. We thank Ludovic Telley (Denis Jabaudon lab) for technical support with the scRNAseq data. M.L.K. would like to thank James Millonig and Daniel Mehan of the Rutgers-RWJMS-Princeton Universities MD/PhD Physician-Scientist Program. M.L.K. was supported by an EMBO Long-Term Postdoctoral Fellowship (190-2016), Alexander von Humboldt Foundation Postdoctoral Fellowship, and a NIH NRSA F30 MD/PhD Fellowship (1F30MH106220). The study was further supported by funding from the Deutsche Forschungsgemeinschaft (DFG; SFB-740) to C.M.T.S. and T.M. High resolution cryo-electron microscopy data collection was performed with support from an iNEXT Cryo-electron Microscopy Instrumentation Grant (PID:2227; EMBL Heidelberg, special thanks to Wim Hagen and Felix Weis), and an Instruct Structural Biology Pilot R&D Grant (APPID: 2016-232; Diamond Light Source Oxfordshire, special thanks to Jason van Rooyen) awarded to M.L.K. Funding from a NeuroCure/Charité Cluster of Excellence Innovation Project Grant awarded to M.L.K. and C.M.T.S. further supported this work. This work was also supported by funding from the German Research Foundation (Grant #EXC 257 to I.V.) and the European Research Council (ERC Advanced Grant #743118 to Bernd Bukau, ZMBH-Heidelberg).

**AUTHOR CONTRIBUTIONS**

M.L.K. designed and conducted the study. C.M.T.S. supervised the study, with contributions from M.S., I.V., G.K., M.L., V.T., T.M., and D.B. Mass spectrometry was processed and analyzed by K.I. and C.H.V.-V. U.Z. processed the RNA sequencing and Ebp1-knockdown ribosome profiling. M.G. performed and analyzed Ebp1-selective ribosome profiling experiments. D.H. further analyzed mass spectrometry, RNA sequencing, and ribosome profiling data. P.T. and T.Sc. prepared pSILAC and BONCAT samples. M.L.K. prepared, and J.B. froze, cryo-electron microscopy samples, followed by initial cryo-electron microscopy data collection by J.B.

559 and T.M. F.K. and M.L.K. performed multiparticle cryo-electron microscopy sorting and
560 refinement, T.Sp. modeled the data, and C.M.T.S. and M.L.K. interpreted the data. A.M.W.
561 performed immuno-electron microscopy. M.C.A. and E.B. performed *in utero* electroporation
562 experiments. Primary cell culture and immunocytochemistry was performed by P.T. Tissue
563 preparation and immunohistochemistry was performed by M.L.K. H.Y. cloned and purified
564 recombinant Ebp1-His. M.-R.R. provided mice for a pilot study and expertise. M.L.K. wrote the
565 manuscript and prepared the figures, with valuable input from all authors.
566

567 **DECLARATION OF INTERESTS**

568 The authors declare no competing interests.
569

570 **FIGURE LEGENDS**

571 **Figure 1. Ebp1 is a highly associated cofactor of the neocortex ribosome across**
572 **development**

573 **(A)** Schematic of the experimental system to measure the architecture of active protein
574 synthesis (polysomal ribosomes) from the neocortex across embryonic and early postnatal
575 neurogenesis. **(B)** Analytic density gradient fractionation of A260-normalized neocortex lysates,
576 measuring the relative abundance of ribosomal subunits, 80S ribosomes, and polysomes. A260
577 curves plotted as mean ± s.d. across replicate fractionations ($n$ = 2-3) for each stage, baseline
578 (1.0) centered at onset of 40S peak. **(C)** Statistical comparison of $\Sigma$ A260 within grey marked
579 regions in (B), mean ± s.d. with significance testing by one-way ANOVA and Dunnett's *post hoc*
580 test vs. E12.5. *$p$ < 0.05; **$p$ < 0.01; ***$p$ < 0.001; ****$p$ < 0.0001. **(D)** MS analysis of neocortex
581 polysomal complexes across development, scatter plots comparing E12.5 vs. each subsequent
582 stage for enrichment of Ebp1, ribosomal proteins of the large (Rpl) and small (Rps) subunits,
583 and translation-associated proteins (GO:0006417). See also **Figures S1 and S2A**. **(E)**
584 Neocortex expression of *Ebp1*, *Rpl*, *Rps*, and translation-associated genes measured in total
585 steady-state levels by RNAseq (left) and MS (right) across developmental stages, median
586 expression plotted ± s.d., one-way ANOVA and Bonferroni corrected *post hoc* test vs. E12.5,
587 $p$ < 0.05. **(F)** Western blot probing for Ebp1 ($Ebp1^{CT}$, $Ebp1^{NT}$) in total neocortex lysates,
588 compared to full-length recombinant Ebp1-His, along with the RP uL30 and Gapdh, full blots
589 shown in **Figures S3A-B**. **(G)** Jitter plots comparing the median stoichiometry of Rpl and Rps
590 (centered at 0) with Ebp1 and translation-associated proteins in total, 80S, and polysomes at
591 E12.5. Other stages shown in **Figure S2B**. **(H)** Western blot analysis (top) of Ebp1 enrichment



268

592 in free, 80S, and polysome fractions across development, compared to Gapdh and uL30.
593 Quantification (bottom, *n* = 2 blots) of Ebp1 and uL30 levels vs. E12.5. See also **Figures S3C-E**.
594
595 **Figure 2. Ebp1 is enriched in early-born NSCs and localizes throughout the neuronal**
596 **cytoplasm**
597 **(A)** Expression heat maps of *Ebp1* compared to averaged *Rpl* and *Rps* family mRNA
598 enrichment in scRNAseq analysis of the developing mouse neocortex, derived from (Telley et
599 al., 2019). Relative expression in apical progenitor (AP) NSCs during differentiation into mature
600 neurons (N4d) on the y-axis, corresponding to NSC birthdates E12, E13, E14, E15 on the x-axis.
601 **(B)** Immunohistochemistry analysis of Ebp1 the developing neocortex ventricular zone (VZ) and
602 cortical plate (CP). Early-born NSCs in the VZ generate lower layer (LL) neurons, while later-
603 born NSCs in the VZ generate upper layer (UL) neurons. Axonal white matter (WM); DAPI
604 staining (grey). Zoomed images (inset, left) correspond to the VZ and leading-edge of the CP at
605 each stage, quantified across *n* = 5-7 images (inset, right heat map). See also **Figure S4A**. **(C)**
606 Immuno-electron microscopy with anti-Ebp1$^{NT}$ immunogold labeling (black dots) in the
607 neocortex at E12.5, E15.5, and P0. Neural stem cells (NSC, blue nuclei) and neurons (N, red
608 nuclei). Nucleoli (n), mitochondria (m, green), endoplasmic reticulum (er), dendrite (D), plasma
609 membrane (arrows). **(D)** Quantification of (C), comparing the cytoplasmic vs. nuclear distribution
610 of Ebp1 in VZ stem cells and CP neurons, with 1° antibody leave-out control. See also **Figures**
611 **S4B-C**. **(E)** Primary neuronal cultures from the E12.5 neocortex, immunocytochemistry at div 0,
612 2, 4, and 5 for Nestin, *Nex:Cre;tdTomato*, and Ebp1. Growing neurites and distal growth cones
613 are indicated (arrows).
614
615 **Figure 3. Ebp1 binds the 60S tunnel exit in actively translating and inactive 80S**
616 **complexes**
617 **(A)** Cryo-electron micrograph of pooled monosome and polysome complexes from P0 mouse
618 neocortical lysates *ex vivo*. **(B)** Cryo-EM maps of (A) with extra-ribosomal density conforming to
619 mouse Ebp1 (PDB 2V6C) over the 60S TE (side view, top image; aerial view, bottom image). N-
620 terminal Ebp1 residues (NT, black ribbon) corresponding to full-length "p48" Ebp1. See also
621 **Figures S5-7**. **(C)** Actively translating (left, classical state with A/A and P/P tRNAs) and non-
622 translating (right, rotated state with eEF2) 80S•Ebp1 complexes. **(D)** Model of the Ebp1 binding
623 surface at the 60S peptide tunnel exit, including 60S rRNA helices H24, H53, H59, and 60S RPs
624 eL19, uL23, uL24, uL29. **(E)** Aerial view of the Ebp1 footprint (red outline) over the 60S peptide
625 tunnel exit (TE), with rRNA helices and RP model surfaces colored as in (D),

626 residues/nucleosides making electrostatic interactions with Ebp1 are highlighted (yellow). **(F)** 2D
627 structure diagram of Ebp1 domains adapted from (Kowalinski et al., 2007), orienting Ebp1 on
628 the ribosome surface, with binding domains highlighted (yellow).
629
630 **Figure 4. Ebp1•60S binding utilizes a conserved H59 latch mechanism, and is**
631 **incompatible with simultaneous binding of other TE cofactors**
632 **(A)** 60S rRNA H59 and H53 models in conformations with and without Ebp1, adjacent to the
633 Ebp1 insert domain. See also **Figures S8A-B**. **(B-D)**. Ebp1•60S binding interface in detail, with
634 interacting residues highlighted for Ebp1 (grey) and the 60S (yellow). See also **Figures S8C-E**.
635 **(E-G)** Global alignment of Ebp1, Metap2, and Arx1 (top, ribbon), likewise when viewed from
636 within the 60S tunnel (bottom, electrostatic potential map) from the perspective of emerging
637 nascent chain. See also **Figure S8F**. **(H)** Aerial view with overlapping footprints of eukaryotic TE
638 binding factors superimposed on the neocortex 60S. PDB IDs: Metap2, 1KQ9; Arx1, 5APN;
639 Sec61, 3J7R; SRP, 6FRK; Ltn1, 3J92; NatA, 6HD7; Ttc5, 6T59. EMDB IDs: RAC, 6105; NAC,
640 4938. **(I)** Jitter plots comparing the median stoichiometry of Rpl and Rps (centered at 0) with
641 Ebp1 and other TE cofactors in total steady-state, 80S, and polysomes at E12.5 and P0. See
642 also **Figure S9**. **(J)** Ebp1•60S binding affinity assay (**Figures S10A-C**), with independent
643 replicate experiments (white and grey circles) and curve best fit to the data. **(K)** Ebp1•60S
644 binding dynamics assessed by pelleting assay and Western blot. Binding pellet signal for (1)
645 super-saturating Ebp1-His, (2) native Ebp1 in RRL, and (3) competition between added Ebp1-
646 His and native Ebp1 in RRL. Native Ebp1 (arrow) and Ebp1-His signal (star).
647
648 **Figure 5. Ebp1•60S complexes engage in translation initiation and elongation, with high**
649 **occupancy prior to N-terminal membrane targeting**
650 **(A)** Correlation between the neuronal Ebp1•ribosome interactome and total translatome
651 measured by selective ribosome profiling (*n* = 3), mRNAs with RPKM enrichment ≥ 1.5-fold are
652 highlighted. See also **Figures S11#**. **(B)** Gene ontology analysis of mRNAs enriched in the
653 Ebp1•ribosome interactome vs. total translatome from (A). **(C)** Proteome-wide metagene read
654 density distribution of the Ebp1•ribosome interactome vs. total translatome over the coding
655 sequence, aligned to the start (left) or stop (right) codon. Arbitrary units (AU). **(D)** Metagene
656 plots as in (C), separated by sub-cellular protein localization. **(E)** Cytoplasmic and signal
657 peptide-containing protein metagene plots aligned to the start or stop codon, and overlaid (left
658 figure) with the relative enrichment at 70 codons annotated (grey dashed line). Metagene plot
659 for signal peptide-containing proteins aligned to the C-terminal codon of the signal sequence


270

(right figure), with 60S tunnel transit region 40 codons downstream annotated (grey box). **(F)** Metagene plot for transmembrane domain-containing (TMD) proteins with (left figure) and without (right figure) an upstream signal peptide, aligned to the C-terminal codon of the first TMD. **(G)** Metagene read density distribution comparing Ebp1 knockdown vs. control neuronal ribosome profiling ($n$ = 3), separated by sub-cellular protein localization, and aligned to the start or stop codon. **(H)** Ebp1 knockdown and control P-site count metagene plots, normalized by total read density per gene, and aligned to the stop or stop codon. Inset right, scaled to highlight relative differences at the start codon. **(I)** Fold change P-site counts at the start, stop, and adjacent codons in Ebp1 knockdown vs. control.

**Figure 6. Ebp1 regulates neocortical neuronal morphology during development and the synthesis of membrane-targeted cell adhesion molecules**

**(A)** E12 *in utero* electroporation (IUE) of NSCs followed by neuronal analysis at E16, comparing *shEbp1* and scrambled shRNA control, and rescue by co-electroporation with Ebp1 overexpression (*oeEbp1*). Co-electroporation with CAG-GFP visualizes transfected cells, shown magnified (bottom), including basally projecting axons (arrows) forming white matter (WM) tracts. **(B)** Morphology tracing GFP labeled neurons in control, *shEbp1*, and rescue *sh+oeEbp1* conditions from (A). **(C)** Sholl analysis of (B), comparing branching per unit distance from the soma (top figure) and sum total (bottom figure) ($n$ = 15 cells), mean ± s.d, one-way ANOVA with Bonferroni corrected *post hoc* test vs. control ($^*p < 0.01$). **(D)** Schematic of the strategy to measure both chronic proteostasis and acute protein synthesis responses to Ebp1 knockdown in Neuro2a cells with pSILAC and BONCAT mass spectrometry. **(E)** pSILAC and pSILAC-AHA labeled protein levels in *siEbp1* relative to *siControl* in biological replicates with label swab. Proteins with significantly lower (purple) or higher (orange) levels in *siEbp1* conditions are highlighted, unchanged proteins in grey. Ebp1 levels were below the MS quantification threshold in *siEbp1* conditions, and thus not plotted. Threshold of >2-fold change from control in both replicates (dotted lines) considered significant. **(F)** Gene ontology (GO) pathway analysis of significantly changing proteins in (E). **(G)** Metagene enrichment plots of the Ebp1-interactome (Ebp1-IP) and Ebp1 knockdown ribosome distribution for *L1cam*, aligned to the start and stop codons.

**Figure 7. Model of Ebp1 function in neurodevelopment**

For details, see text. Endoplasmic reticulum (ER), cell adhesion molecules (CAMs).

**REFERENCES: MAIN**

Behrmann, E., Loerke, J., Budkevich, T. V., Yamamoto, K., Schmidt, A., Penczek, P.A., Vos, M.R., Bürger, J., Mielke, T., Scheerer, P., et al. (2015). Structural snapshots of actively translating human ribosomes. Cell *161*, 845–857.

Bosco, D.A., Lavoie, M.J., Petsko, G.A., and Ringe, D. (2011). Proteostasis and Movement Disorders: Parkinson's Disease and Amyotrophic Lateral Sclerosis. Cold Spring Harb. Perspect. Biol. *3*, a007500.

Bradatsch, B., Katahira, J., Kowalinski, E., Bange, G., Yao, W., Sekimoto, T., Baumgärtel, V., Boese, G., Bassler, J., Wild, K., et al. (2007). Arx1 Functions as an Unorthodox Nuclear Export Receptor for the 60S Preribosomal Subunit. Mol. Cell *27*, 767–779.

Cadwell, C.R., Bhaduri, A., Mostajo-Radji, M.A., Keefe, M.G., and Nowakowski, T.J. (2019). Development and Arealization of the Cerebral Cortex. Neuron *103*, 980–1004.

Chartron, J.W., Hunt, K.C.L., and Frydman, J. (2016). Cotranslational signal-independent SRP preloading during membrane targeting. Nature *536*, 224–228.

Chau, K.F., Shannon, M.L., Fame, R.M., Fonseca, E., Mullan, H., Johnson, M.B., Sendamarai, A.K., Springel, M.W., Laurent, B., and Lehtinen, M.K. (2018). Downregulation of ribosome biogenesis during early forebrain development. Elife *7*, e36998.

DeBoer, E.M., Kraushar, M.L., Hart, R.P., and Rasin, M.-R. (2013). Post-transcriptional regulatory elements and spatiotemporal specification of neocortical stem cells and projection neurons. Neuroscience *248*, 499–528.

Delarue, M., Brittingham, G.P., Pfeffer, S., Surovtsev, I. V., Pinglay, S., Kennedy, K.J., Schaffer, M., Gutierrez, J.I., Sang, D., Poterewicz, G., et al. (2018). mTORC1 Controls Phase Separation and the Biophysical Properties of the Cytoplasm by Tuning Crowding. Cell *174*, 338–349.

Deuerling, E., Gamerdinger, M., and Kreft, S.G. (2019). Chaperone Interactions at the Ribosome. Cold Spring Harb. Perspect. Biol. a033977.

Dieterich, D.C., Link, A.J., Graumann, J., Tirrell, D.A., and Schuman, E.M. (2006). Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). Proc. Natl. Acad. Sci. *103*, 9482–9487.

Eichelbaum, K., Winter, M., Diaz, M.B., Herzig, S., and Krijgsveld, J. (2012). Selective enrichment of newly synthesized proteins for quantitative secretome analysis. Nat. Biotechnol. *30*, 984–990.

Evangelopoulos, M.E., Weis, J., and Krüttgen, A. (2005). Signalling pathways leading to neuroblastoma differentiation after serum withdrawal: HDL blocks neuroblastoma differentiation by inhibition of EGFR. Oncogene *24*, 3309–3318.

Flis, J., Holm, M., Rundlet, E.J., Loerke, J., Hilal, T., Dabrowski, M., Bürger, J., Mielke, T., Blanchard, S.C., Spahn, C.M.T., et al. (2018). tRNA Translocation by the Eukaryotic 80S Ribosome and the Impact of GTP Hydrolysis. Cell Rep. *25*, 2676–2688.

Gamerdinger, M., Hanebuth, M.A., Frickey, T., and Deuerling, E. (2015). The principle of antagonism ensures protein targeting specificity at the endoplasmic reticulum. Science. *348*, 201–207.

742 Greber, B.J., Gerhardy, S., Leitner, A., Aebersold, R., Panse, V.G., Ban, N., Boehringer, D.,
743 Leulliot, N., Aebersold, R., Panse, V.G., et al. (2016). Insertion of the Biogenesis Factor Rei1
744 Probes the Ribosomal Tunnel during 60S Maturation. Cell *164*, 91–102.

745 Hafner, A.-S., Donlin-Asp, P.G., Leitch, B., Herzog, E., and Schuman, E.M. (2019). Local
746 protein synthesis is a ubiquitous feature of neuronal pre- and postsynaptic compartments.
747 Science. *364*, 363184.

748 Holt, C.E., Martin, K.C., and Schuman, E.M. (2019). Local translation in neurons: visualization
749 and function. Nat. Struct. Mol. Biol. *26*, 557–566.

750 Howden, A.J.M., Geoghegan, V., Katsch, K., Efstathiou, G., Bhushan, B., Boutureira, O.,
751 Thomas, B., Trudgian, D.C., Kessler, B.M., Dieterich, D.C., et al. (2013). QuaNCAT:
752 Quantitating proteome dynamics in primary cells. Nat. Methods *10*, 343–346.

753 Jan, C.H., Williams, C.C., and Weissman, J.S. (2014). Principles of ER cotranslational
754 translocation revealed by proximity-specific ribosome profiling. Science. *346*, 748–751.

755 Jayaraj, G.G., Hipp, M.S., and Hartl, F.U. (2019). Functional Modules of the Proteostasis
756 Network. Cold Spring Harb. Perspect. Biol. a033951.

757 Jung, H., Gkogkas, C.G., Sonenberg, N., and Holt, C.E. (2014). Remote control of gene function
758 by local translation. Cell *157*, 26–40.

759 Kapur, M., Monaghan, C.E., and Ackerman, S.L. (2017). Regulation of mRNA Translation in
760 Neurons — A Matter of Life and Death. Neuron *96*, 616–637.

761 Knorr, A.G., Schmidt, C., Tesina, P., Berninghausen, O., Becker, T., Beatrix, B., and Beckmann,
762 R. (2019). Ribosome–NatA architecture reveals that rRNA expansion segments coordinate N-
763 terminal acetylation. Nat. Struct. Mol. Biol. *26*, 35–39.

764 Ko, H.R., Hwang, I., Ahn, S.Y., Chang, Y.S., and Park, W.S. (2017). Neuron-specific expression
765 of p48 Ebp1 during murine brain development and its contribution to CNS axon regeneration.
766 BMB Rep. *50*, 126–131.

767 Kobayashi, K., Jomaa, A., Lee, J.H., Chandrasekar, S., Boehringer, D., Shan, S., and Ban, N.
768 (2018). Structure of a prehandover mammalian ribosomal SRP·SRP receptor targeting complex.
769 Science. *360*, 323–327.

770 Kowalinski, E., Bange, G., Bradatsch, B., Hurt, E., Wild, K., and Sinning, I. (2007). The crystal
771 structure of Ebp1 reveals a methionine aminopeptidase fold as binding platform for multiple
772 interactions. FEBS Lett. *581*, 4450–4454.

773 Kramer, G., Shiber, A., and Bukau, B. (2019). Mechanisms of Cotranslational Maturation of
774 Newly Synthesized Proteins. Annu. Rev. Biochem. *88*, 337–364.

775 Kraushar, M.L., Viljetic, B., Wijeratne, H.R.S., Thompson, K., Jiao, X., Pike, J.W., Medvedeva,
776 V., Groszer, M., Kiledjian, M., Hart, R.P., et al. (2015). Thalamic WNT3 Secretion
777 Spatiotemporally Regulates the Neocortical Ribosome Signature and mRNA Translation to
778 Specify Neocortical Cell Subtypes. J. Neurosci. *35*, 10911–10926.

779 Kraushar, M.L., Popovitchenko, T., Volk, N.L., and Rasin, M.-R. (2016). The frontier of RNA
780 metamorphosis and ribosome signature in neocortical development. Int. J. Dev. Neurosci. *55*,
781 131–139.

Kwon, I.S., and Ahn, J.Y. (2011). P48 Ebp1 acts as a downstream mediator of Trk signaling in neurons, contributing neuronal differentiation. Neurochem. Int. *58*, 215–223.

Lein, E.S., Belgard, T.G., Hawrylycz, M., and Molnár, Z. (2017). Transcriptomic Perspectives on Neocortical Structure, Development, Evolution, and Disease. Annu. Rev. Neurosci. *40*, 629–652.

Lin, Z., Gasic, I., Chandrasekaran, V., Peters, N., Shao, S., Mitchison, T.J., and Hegde, R.S. (2020). TTC5 mediates autoregulation of tubulin via mRNA degradation. Science. *367*, 100–104.

Liu, Z., Ahn, J.-Y., Liu, X., and Ye, K. (2006). Ebp1 isoforms distinctively regulate cell survival and differentiation. Proc. Natl. Acad. Sci. U. S. A. *103*, 10917–10922.

Loerke, J., Giesebrecht, J., and Spahn, C.M.T. (2010). Multiparticle Cryo-EM of Ribosomes. Methods Enzymol. *483*, 161–177.

Martínez, G., Khatiwada, S., Costa-Mattioli, M., and Hetz, C. (2018). ER Proteostasis Control of Neuronal Physiology and Synaptic Function. Trends Neurosci. *41*, 610–624.

Mills, E.W., and Green, R. (2017). Ribosomopathies: There's strength in numbers. Science. *358*, eaan2755.

Molyneaux, B.J., Arlotta, P., Menezes, J.R.L., and Macklis, J.D. (2007). Neuronal subtype specification in the cerebral cortex. Nat. Rev. Neurosci. *8*, 427–437.

Monie, T.P., Perrin, A.J., Birtley, J.R., Sweeney, T.R., Karakasiliotis, I., Chaudhry, Y., Roberts, L.O., Matthews, S., Goodfellow, I.G., and Curry, S. (2007). Structural insights into the transcriptional and translational roles of Ebp1. EMBO J. *26*, 3936–3944.

Nguyen, D.Q., Hoang, D.H., Nguyen Vo, T.T., Huynh, V., Ghoda, L., Marcucci, G., and Nguyen, L.X.T. (2018). The role of ErbB3 binding protein 1 in cancer: Friend or foe? J. Cell. Physiol. 1–11.

Nonato, M.C., Widom, J., and Clardy, J. (2006). Human methionine aminopeptidase type 2 in complex with l- and d-methionine. Bioorganic Med. Chem. Lett. *16*, 2580–2583.

Nowakowski, T.J., Bhaduri, A., Pollen, A.A., Alvarado, B., Mostajo-Radji, M.A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S.J., Velmeshev, D., et al. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science. *358*, 1318–1323.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread Co-translational RNA Decay Reveals Ribosome Dynamics. Cell *161*, 1400–1412.

Penczek, P.A., Frank, J., and Spahn, C.M.T. (2006). A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. J. Struct. Biol. *154*, 184–194.

Pilipenko, E. V., Pestova, T. V., Kolupaeva, V.G., Khitrina, E. V., Poperechnaya, A.N., Agol, V.I., and Hellen, C.U.T. (2000). A cell cycle-dependent protein serves as a template-specific translation initiation factor. Genes Dev. *14*, 2028–2045.

Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296.

820 Scheres, S.H.W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure
821 determination. J. Struct. Biol. *180*, 519–530.

822 Schibich, D., Gloge, F., Pöhner, I., Björkholm, P., Wade, R.C., von Heijne, G., Bukau, B., and
823 Kramer, G. (2016). Global profiling of SRP interaction with nascent polypeptides. Nature *536*,
824 219–223.

825 Schwanhäusser, B., Gossen, M., Dittmar, G., and Selbach, M. (2009). Global analysis of cellular
826 protein translation by pulsed SILAC. Proteomics *9*, 205–209.

827 Shao, S., Brown, A., Santhanam, B., and Hegde, R.S. (2015). Structure and assembly pathway
828 of the ribosome quality control complex. Mol. Cell *57*, 433–445.

829 Shen, K., Gamerdinger, M., Chan, R., Gense, K., Martin, E.M., Sachs, N., Knight, P.D.,
830 Schlömer, R., Calabrese, A.N., Stewart, K.L., et al. (2019). Dual Role of Ribosome-Binding
831 Domain of NAC as a Potent Suppressor of Protein Aggregation and Aging-Related
832 Proteinopathies. Mol. Cell *74*, 729–741.

833 Shi, Z., and Barna, M. (2015). Translating the Genome in Time and Space: Specialized
834 Ribosomes, RNA Regulons, and RNA-Binding Proteins. Annu. Rev. Cell Dev. Biol. *31*, 31–54.

835 Silbereis, J.C., Pochareddy, S., Zhu, Y., Li, M., and Sestan, N. (2016). The Cellular and
836 Molecular Landscapes of the Developing Human Central Nervous System. Neuron *89*, 248–268.

837 Sinturel, F., Gerber, A., Mauvoisin, D., Wang, J., Gatfield, D., Stubblefield, J.J., Green, C.B.,
838 Gachon, F., and Schibler, U. (2017). Diurnal Oscillations in Liver Mass and Cell Size
839 Accompany Ribosome Assembly Cycles. Cell *169*, 651–663.

840 Sossin, W.S., and Costa-Mattioli, M. (2018). Translational Control in the Brain in Health and
841 Disease. Cold Spring Harb. Perspect. Biol. a032912.

842 Squatrito, M., Mancino, M., Sala, L., and Draetta, G.F. (2006). Ebp1 is a dsRNA-binding protein
843 associated with ribosomes that modulates eIF2alpha phosphorylation. Biochem. Biophys. Res.
844 Commun. *344*, 859–868.

845 Telley, L., Agirman, G., Prados, J., Amberg, N., Fièvre, S., Oberst, P., Bartolini, G., Vitali, I.,
846 Cadilhac, C., Hippenmeyer, S., et al. (2019). Temporal patterning of apical progenitors and their
847 daughter neurons in the developing neocortex. Science. *364*, eaav2522.

848 Turko, P., Groberman, K., Browa, F., Cobb, S., and Vida, I. (2018). Differential Dependence of
849 GABAergic and Glutamatergic Neurons on Glia for the Establishment of Synaptic Transmission.
850 Cereb. Cortex *29*, 1230–1243.

851 Voorhees, R.M., Fernandez, I.S., Scheres, S.H.W., and Hegde, R.S. (2014). Structure of the
852 mammalian ribosome-Sec61 complex to 3.4A resolution. Cell *157*, 1632–1643.

853 Wild, K., Aleksić, M., Lapouge, K., Juaire, K.D., Flemming, D., Pfeffer, S., and Sinning, I. (2020).
854 MetAP-like Ebp1 occupies the human ribosomal tunnel exit and recruits flexible rRNA
855 expansion segments. Nat. Commun. *11*, 1–10.

856 de Wit, J., and Ghosh, A. (2016). Specification of synaptic connectivity by cell surface
857 interactions. Nat. Rev. Neurosci. *17*, 22–35.

858 Yuzwa, S.A., Borrett, M.J., Innes, B.T., Voronova, A., Ketela, T., Kaplan, D.R., Bader, G.D., and

859 Miller, F.D. (2017). Developmental Emergence of Adult Neural Stem Cells as Revealed by
860 Single-Cell Transcriptional Profiling. Cell Rep. *21*, 3970–3986.

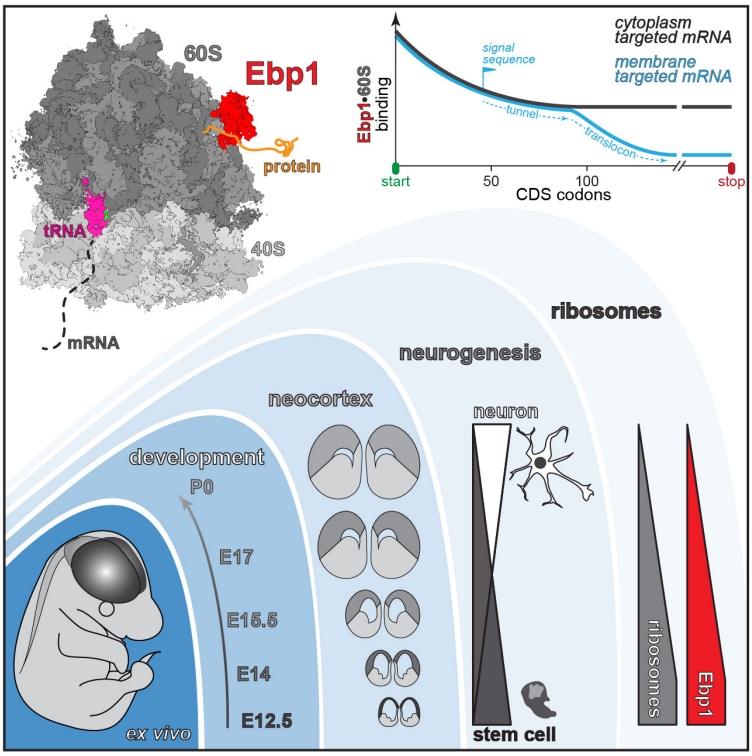
861 Zappulo, A., Van Den Bruck, D., Ciolli Mattioli, C., Franke, V., Imami, K., McShane, E., Moreno-
862 Estelles, M., Calviello, L., Filipchyk, A., Peguero-Sanchez, E., et al. (2017). RNA localization is a
863 key determinant of neurite-enriched proteome. Nat. Commun. *8*, 1–12.

864 Zhang, Y., Ma, C., Yuan, Y., Zhu, J., Li, N., Chen, C., Wu, S., Yu, L., Lei, J., and Gao, N. (2014).
865 Structural basis for interaction of a cotranslational chaperone with the eukaryotic ribosome. Nat.
866 Struct. Mol. Biol. *21*, 1042–1046.

867

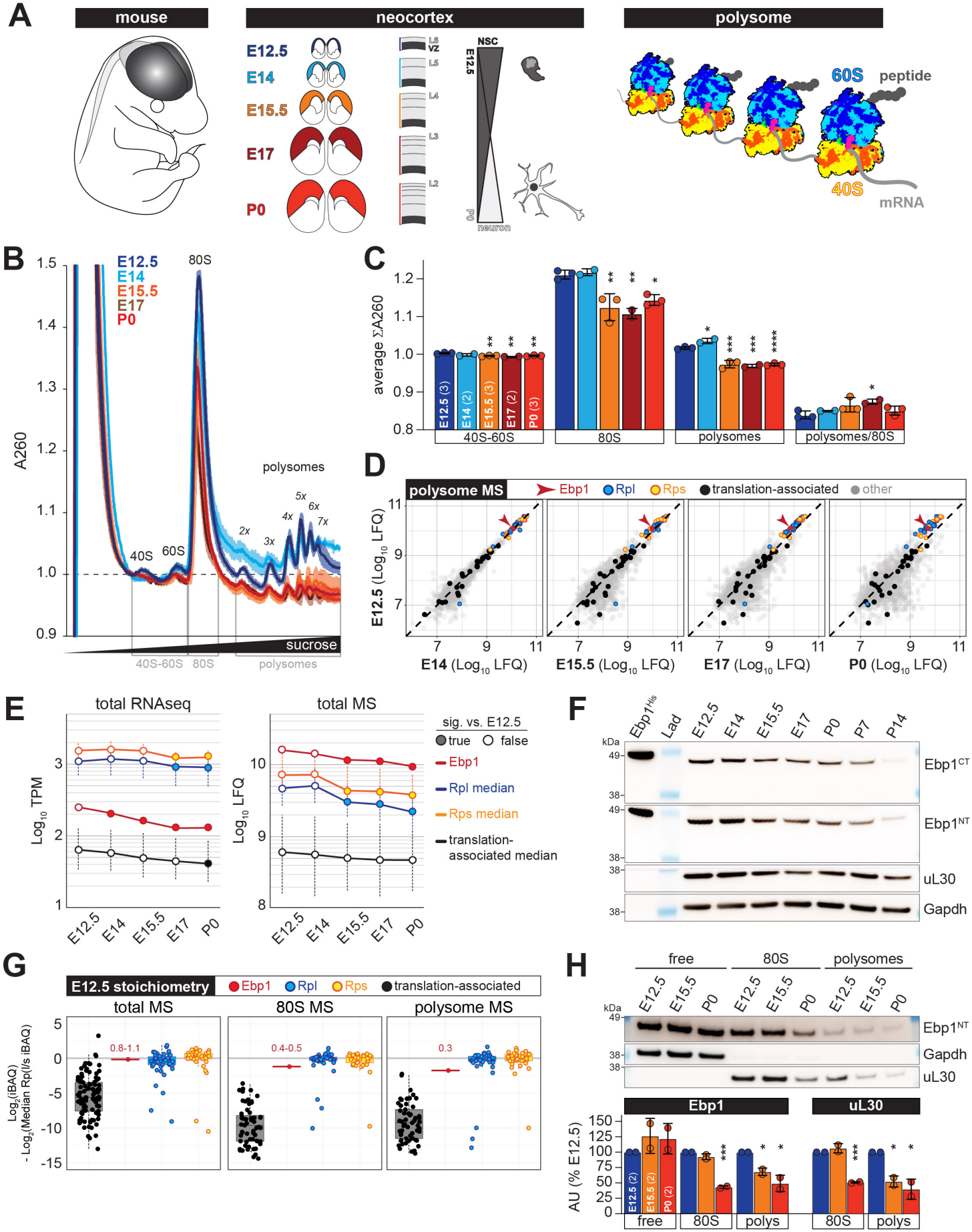
# Graphical abstract

Kraushar ML, *et al*. Revision.



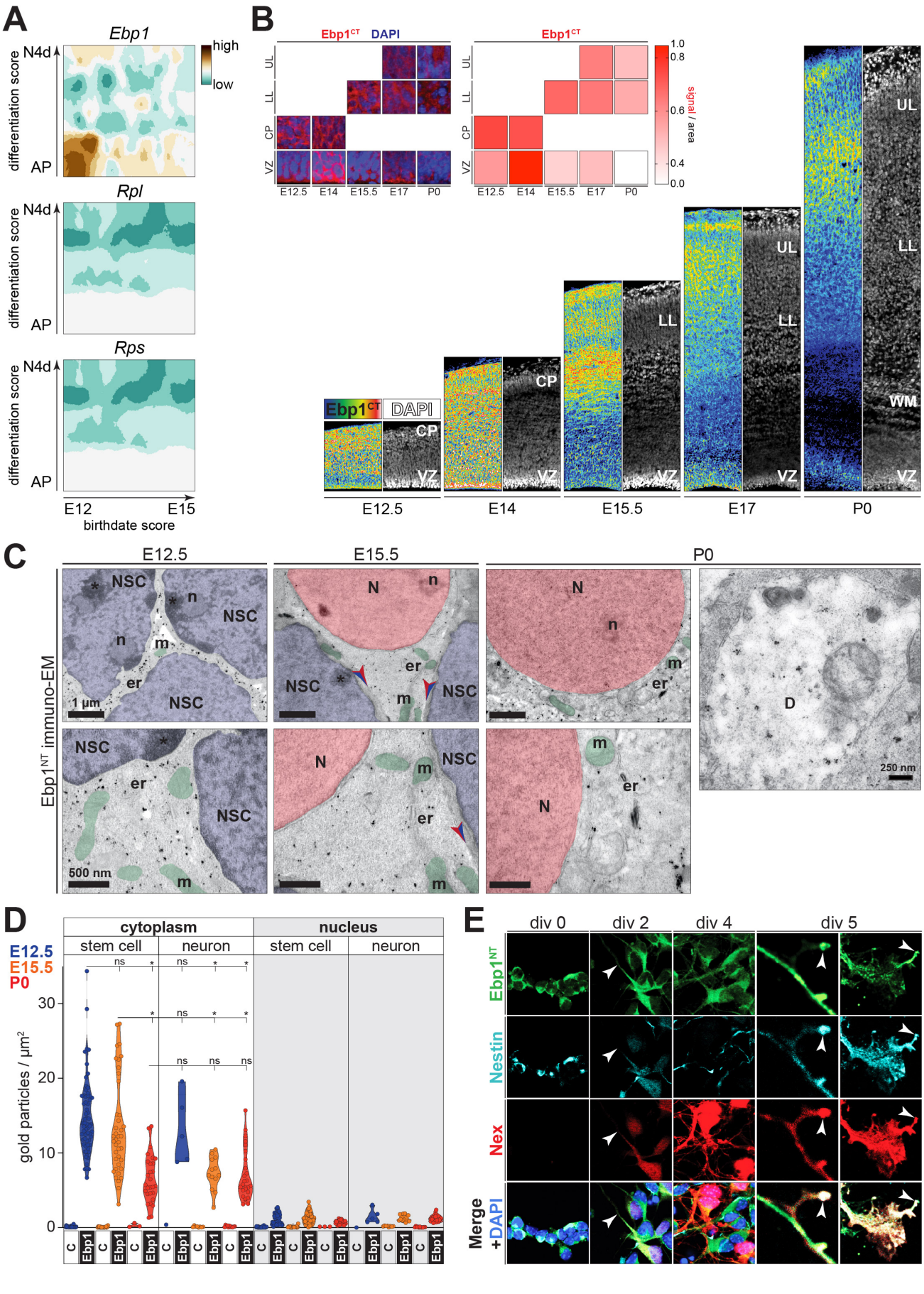
# Figure 1

## Kraushar ML, *et al*. Revision.



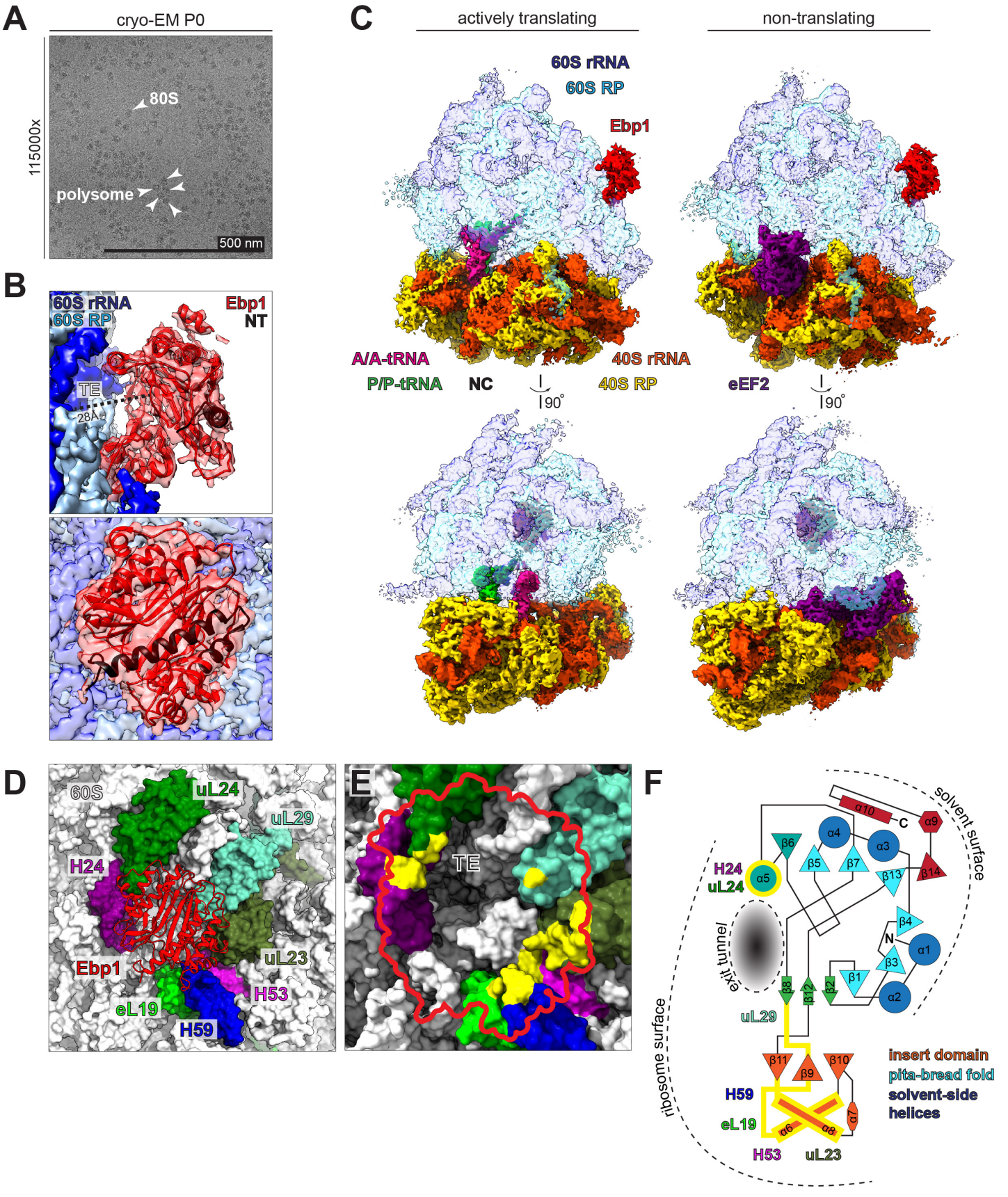
# Figure 2

# Kraushar ML, *et al*. Revision.



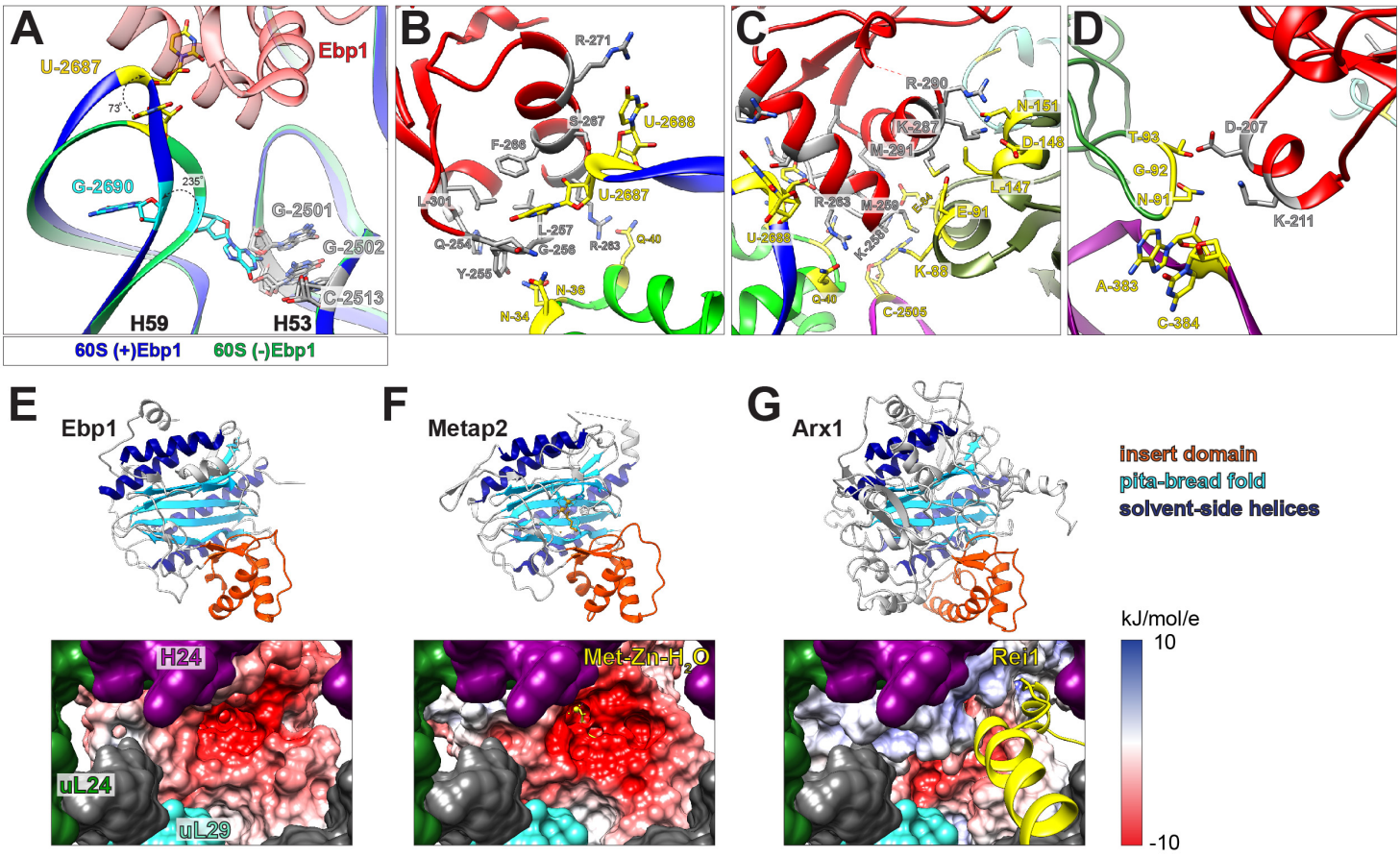
# Figure 3

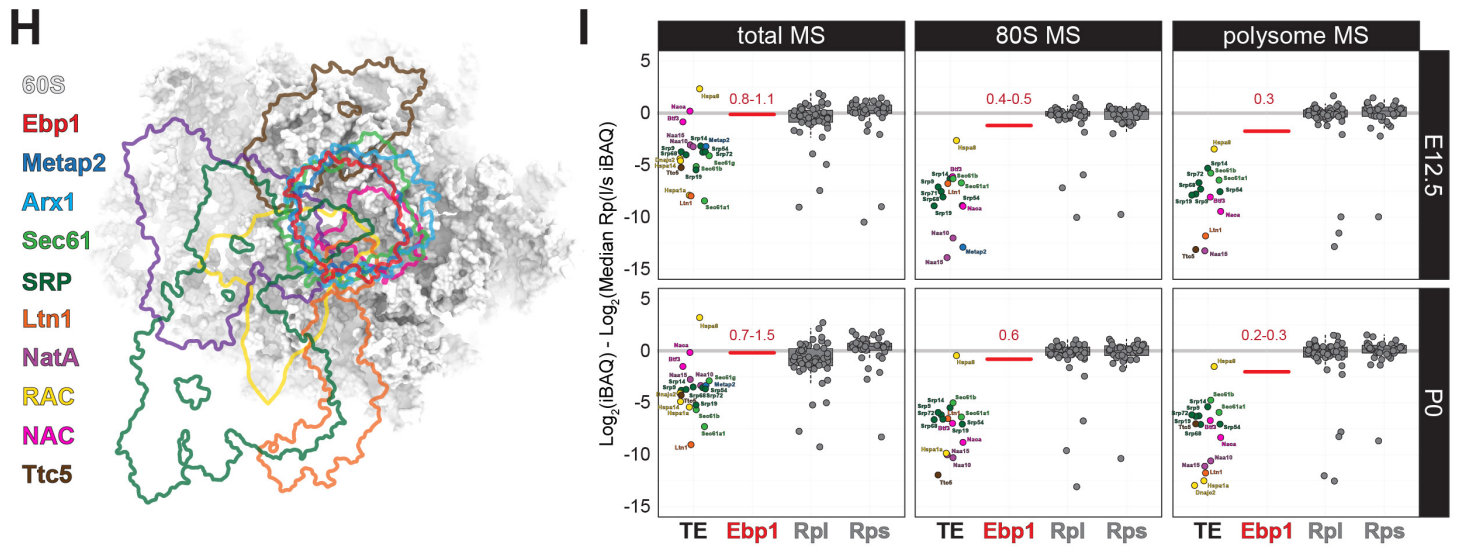
## Kraushar ML, *et al*. Revision.

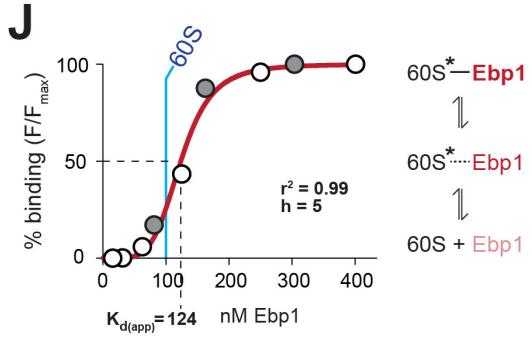


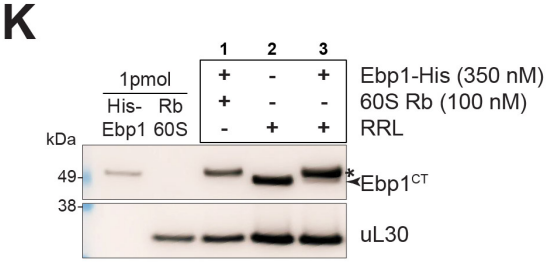
# Figure 4

## Kraushar ML, *et al*. Revision.



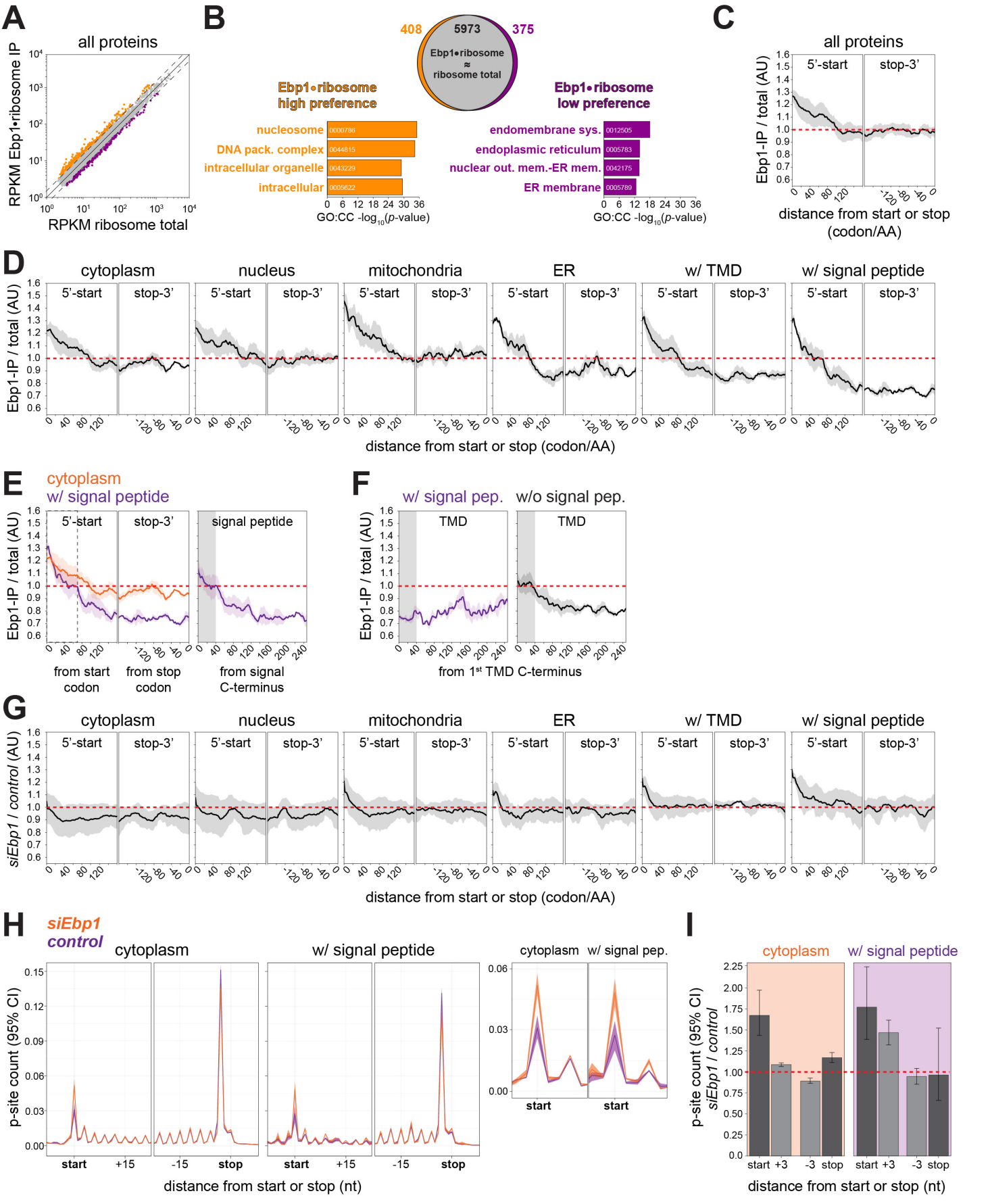



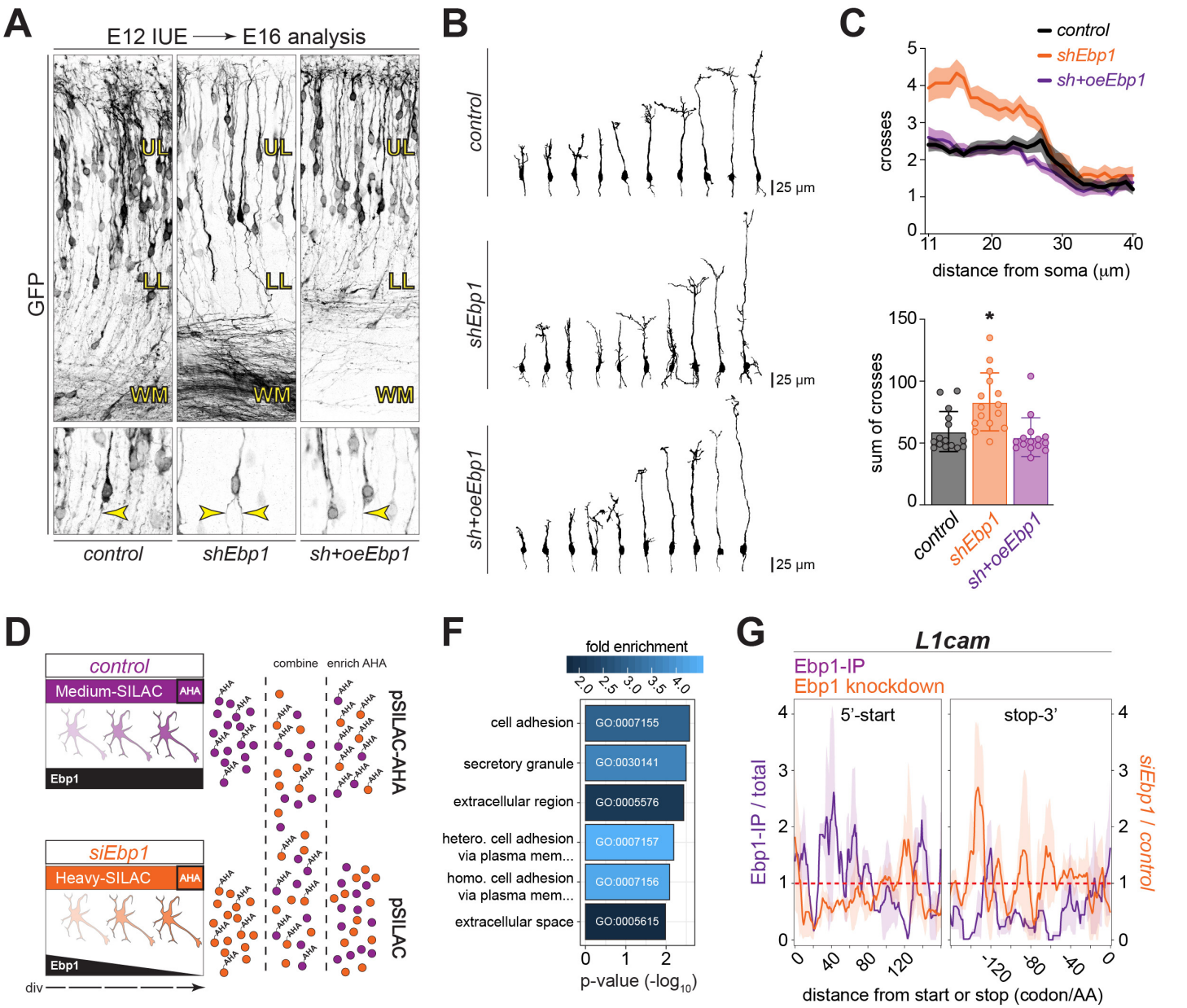
# Figure 5




A
all proteins
RPKM Ebp1•ribosome IP
RPKM ribosome total
B
408
5973
375
Ebp1•ribosome ≈ ribosome total
Ebp1•ribosome high preference
nucleosome
DNA pack. complex
intracellular organelle
intracellular
Ebp1•ribosome low preference
endomembrane sys.
endoplasmic reticulum
nuclear out. mem.-ER mem.
ER membrane
GO:CC -log10(p-value)
C
all proteins
5'-start
stop-3'
Ebp1-IP / total (AU)
distance from start or stop (codon/AA)
D
cytoplasm
nucleus
mitochondria
ER
w/ TMD
w/ signal peptide
E
cytoplasm
w/ signal peptide
signal peptide
from start codon
from stop codon
from signal C-terminus
F
w/ signal pep.
w/o signal pep.
TMD
from 1st TMD C-terminus
G
siEbp1 / control (AU)
H
siEbp1
control
p-site count (95% CI)
start
+15
-15
stop
distance from start or stop (nt)
I
p-site count (95% CI) siEbp1 / control
start +3
-3 stop

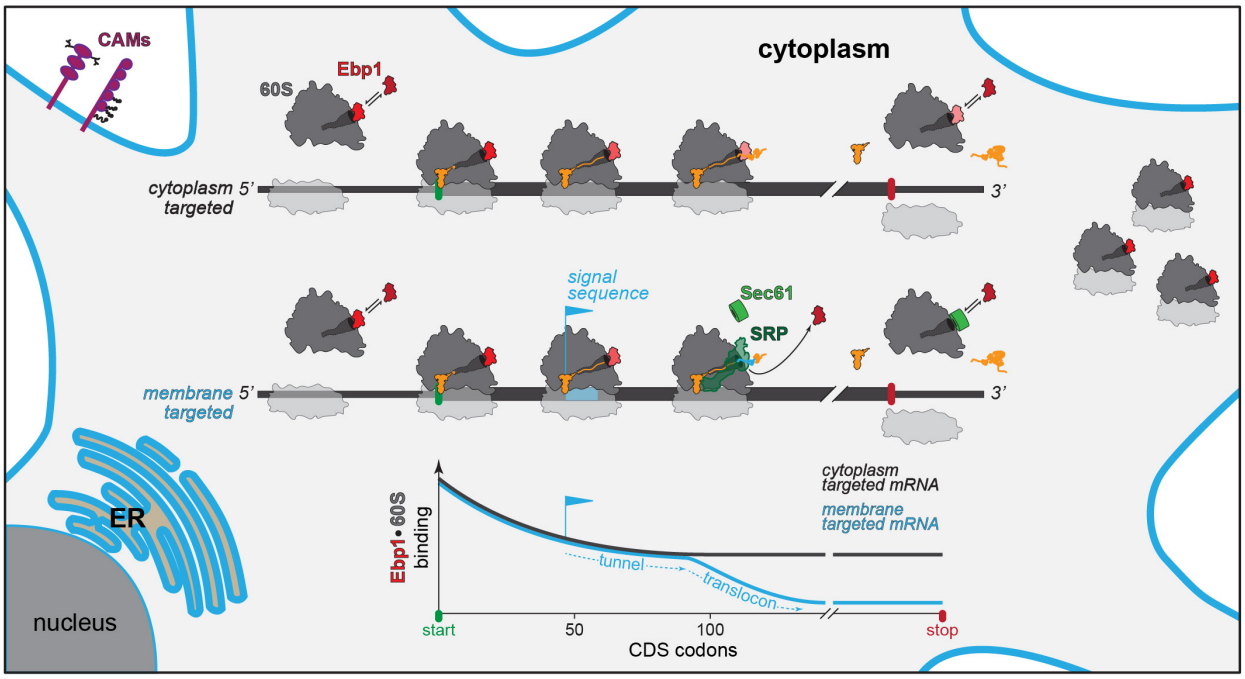
# Figure 6

**Kraushar ML, *et al*. Revision.**



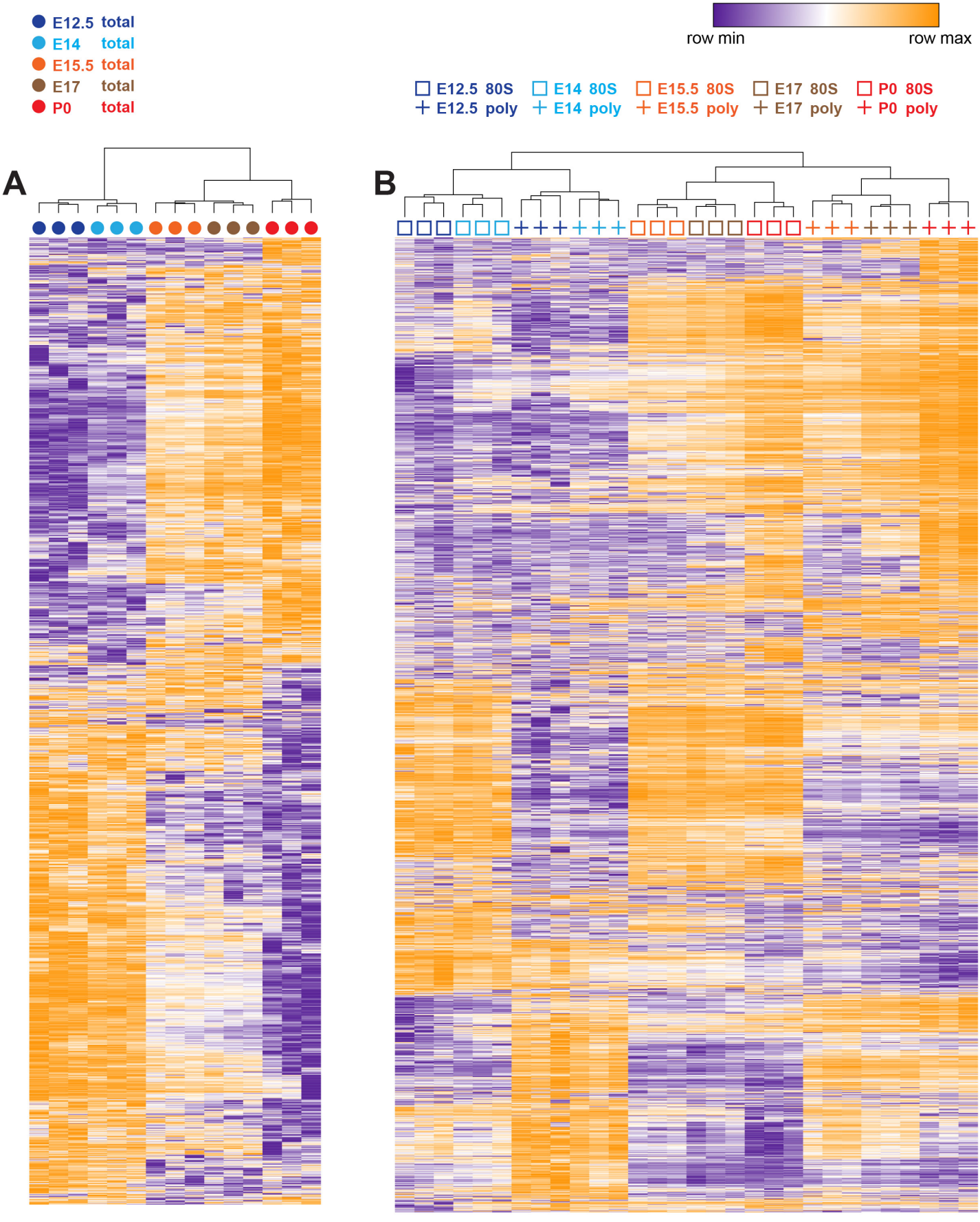
# Figure 7

# Kraushar ML, *et al*. Revision.



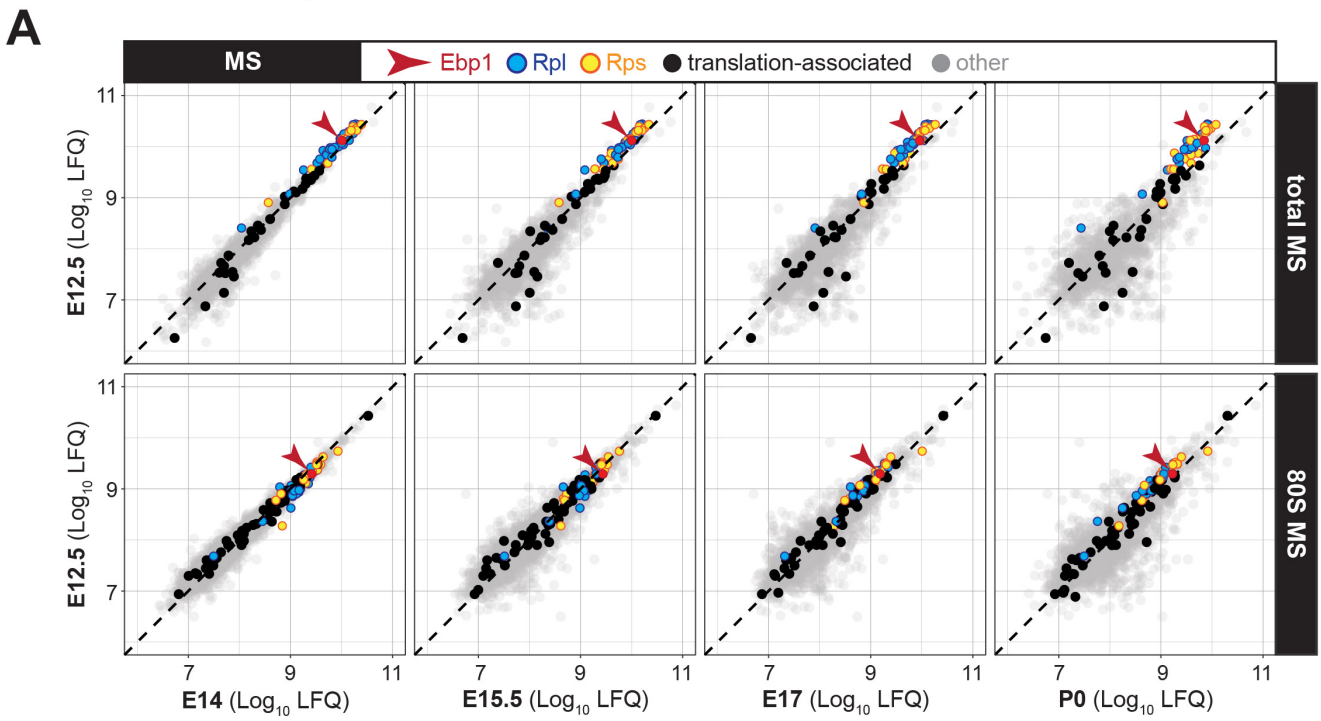
# Supplementary Figure 1

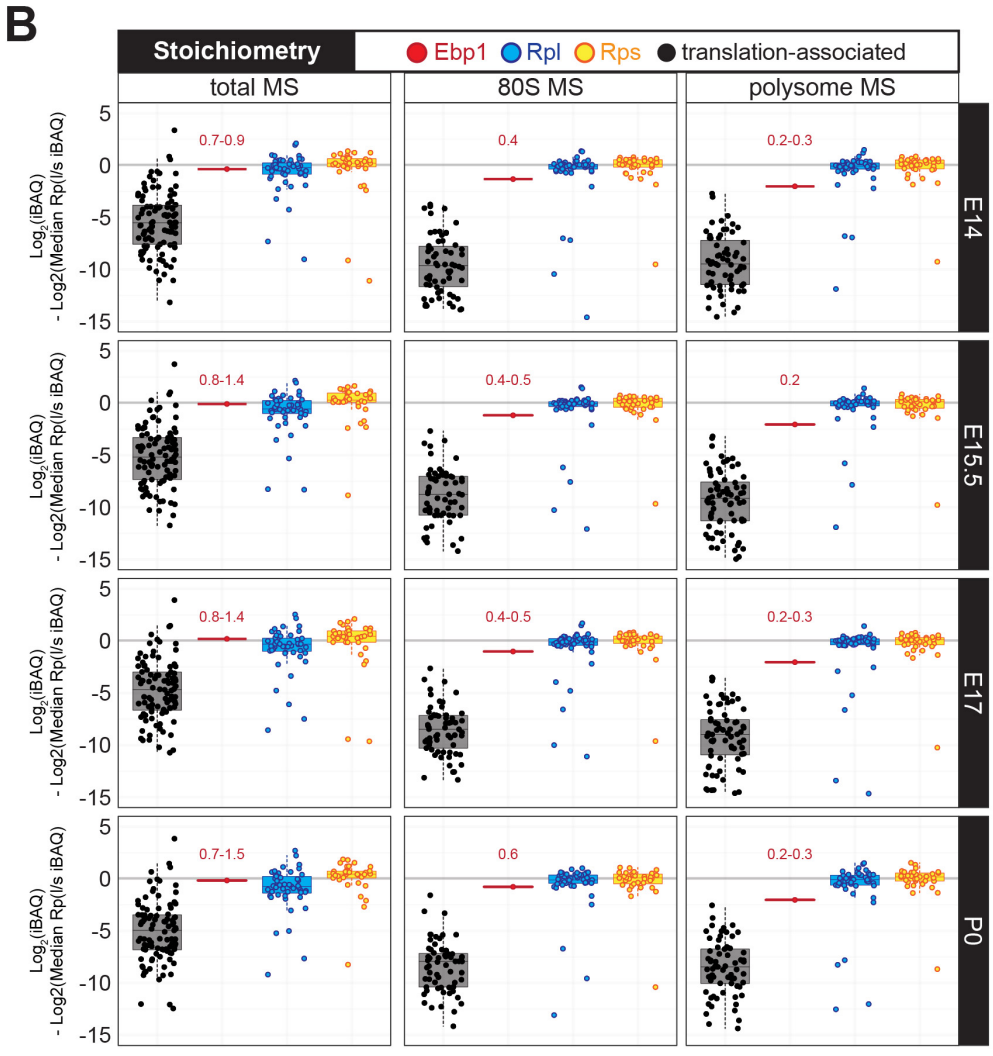
## Kraushar ML, *et al*. Revision.



# Supplementary Figure 2

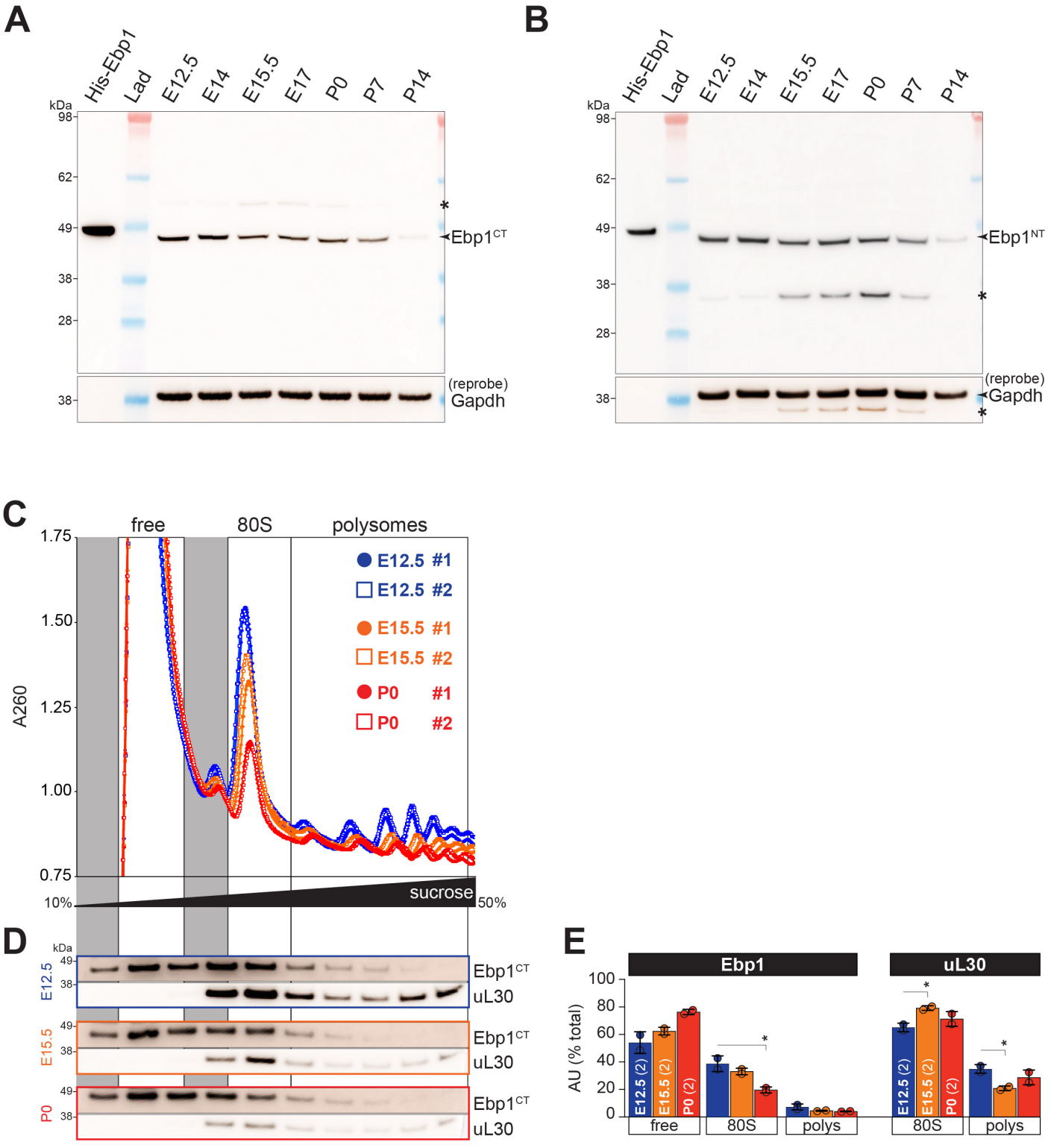
**Kraushar ML, *et al*. Revision.**



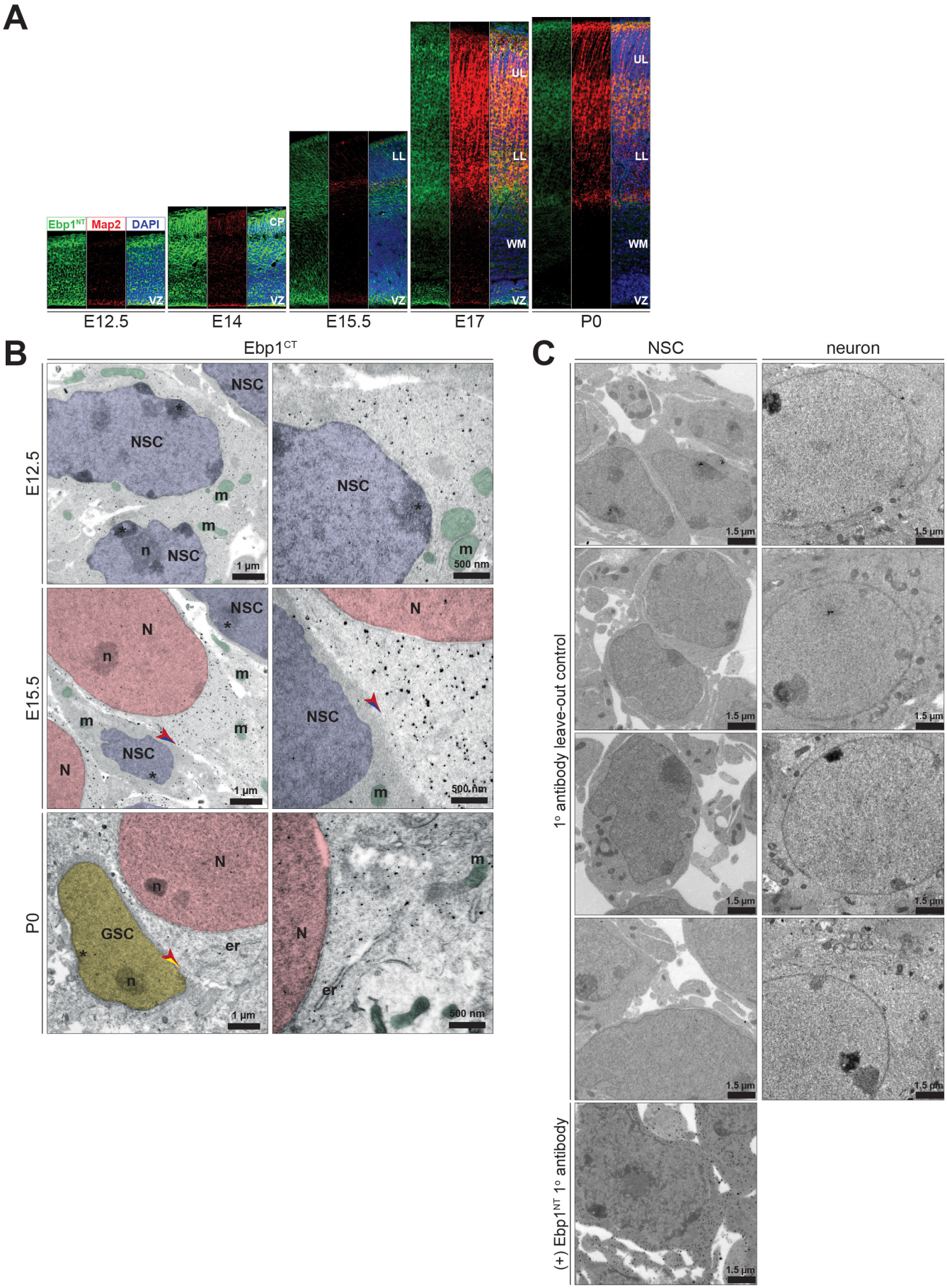
# Supplementary Figure 3

Kraushar ML, *et al*. Revision.



# Supplementary Figure 4

# Kraushar ML, *et al*. Revision.



# Supplementary Figure 5

# Kraushar ML, *et al*. Revision.

**A**

| Data collection | |
|---|---|
| Voltage (keV) | 300 |
| Defocus range (μm) | 0.5-2.5 |
| Pixel size (Å) on object scale | 0.66 |
| Electron dose ($e^-$/Å$^2$) | 31.78 |

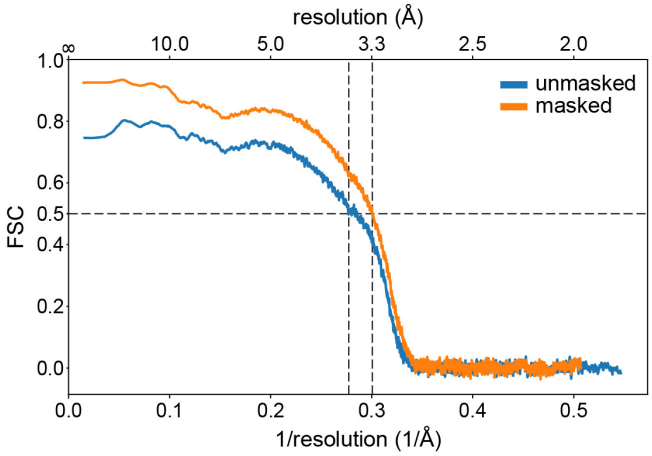
| Map refinement | *Rotated (+)eEF2* | | *Classical (+)A/A, P/P tRNAs* | |
|---|---|---|---|---|
| | *(+)Ebp1* | *(-)Ebp1* | *(+)Ebp1* | *(-)Ebp1* |
| Particles | 23907 | 29466 | 15262 | 14973 |
| Resolution (Å) | 3.1 | 3.1 | 3.3 | 3.3 |

| Model | |
|---|---|
| Composition: | |
| Chains | 58 |
| Atoms | 135402 (Hydrogens: 0) |
| Residues | |
| Protein | 6385 |
| Nucleotide | 3887 |
| Water | 0 |
| Ligands | |
| Zn | 3 |
| Mg | 257 |
| Bonds (RMSD): | |
| Length Å (# > 4σ) | 0.010 (7) |
| Angles (°) (# > 4σ) | 0.702 (45) |
| MolProbity score | 1.47 |
| Clash score | 4.06 |
| Ramachandran plot (%) | |
| Outliers | 0.02 |
| Allowed | 4.1 |
| Favored | 95.89 |
| Rotamer outliers (%) | 0.16 |
| Cβ outliers (%) | 0 |
| Peptide plane (%) | |
| Cis proline/general | 0.4/0.0 |
| Twisted proline/general | 0.4/0.1 |
| CaBLAM outliers (%) | 2.74 |
| ADP (B-factors) | |
| Iso/Aniso (#) | 135402/0 |
| min/max/mean | |
| Protein | 43.66/282.25/79.40 |
| Nucleotide | 44.27/574.49/99.91 |
| Ligand | 29.57/202.48/57.93 |
| Water | --- |
| Occupancy | |
| Mean | 1 |
| occ = 1 (%) | 100 |
| 0 < occ < 1 (%) | 0 |
| occ > 1 (%) | 0 |

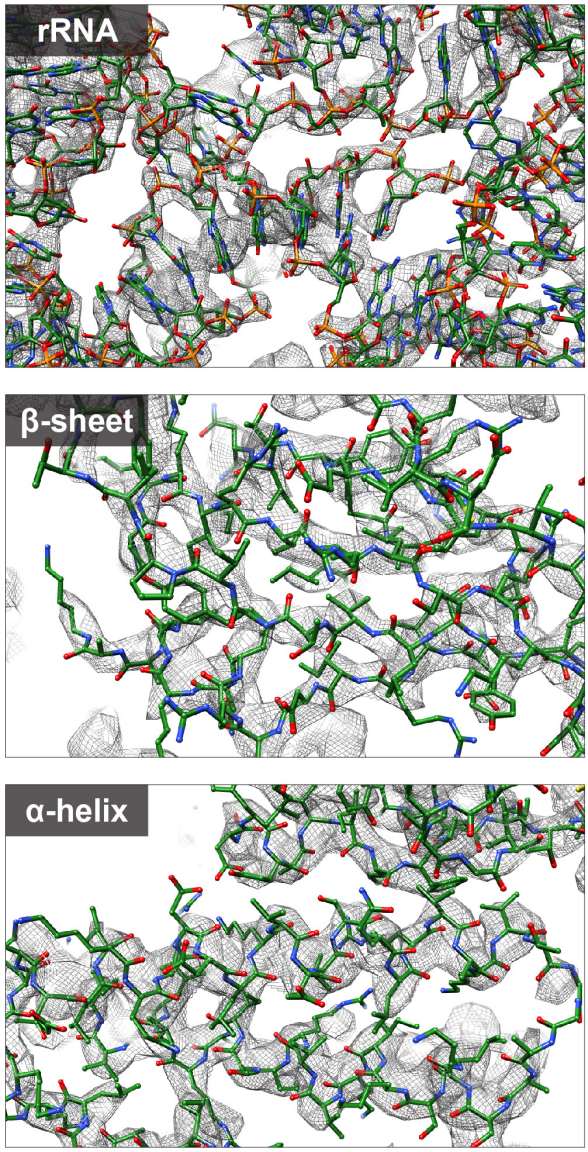
| Data | | |
|---|---|---|
| Box | | |
| Lengths (Å) | 238.81, 218.91, 270.65 | |
| Angles (°) | 90.00, 90.00, 90.00 | |
| Supplied Resolution (Å) | 3.3 | |
| Resolution Estimates (Å) | *Masked* | *Unmasked* |
| d FSC (half maps; 0.143) | --- | --- |
| d 99 (full/half1/half2) | 3.5/---/--- | 3.5/---/--- |
| d model | 3.4 | 3.4 |
| d FSC model (0/0.143/0.5) | 2.9/3.1/3.3 | 2.9/3.1/3.6 |
| Map min/max/mean | -10.76/22.03/0.00 | |

| Model vs. Data | |
|---|---|
| CC (mask) | 0.84 |
| CC (box) | 0.69 |
| CC (peaks) | 0.55 |
| CC (volume) | 0.82 |
| Mean CC for ligands | 0.75 |

**B**

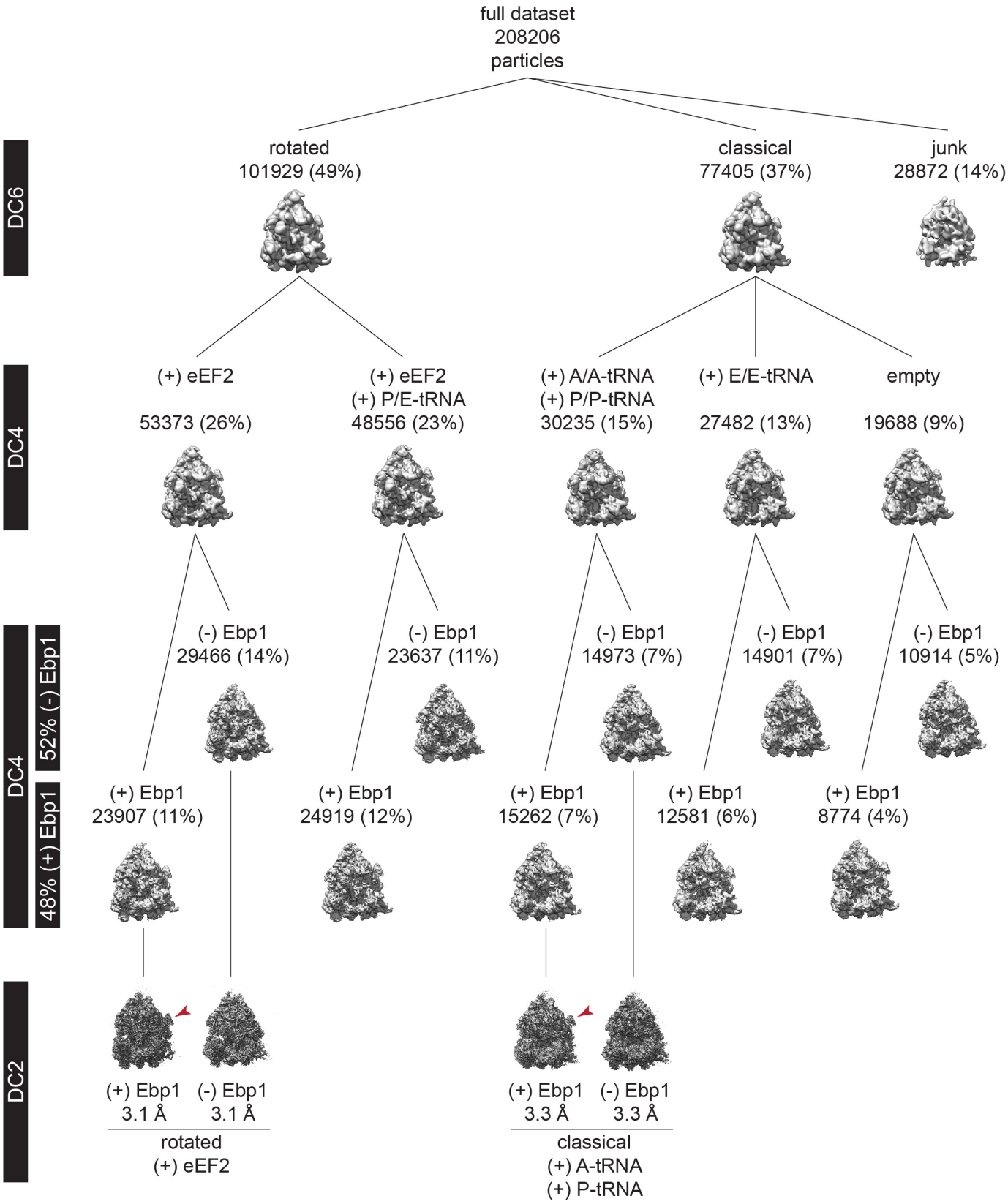



**C**



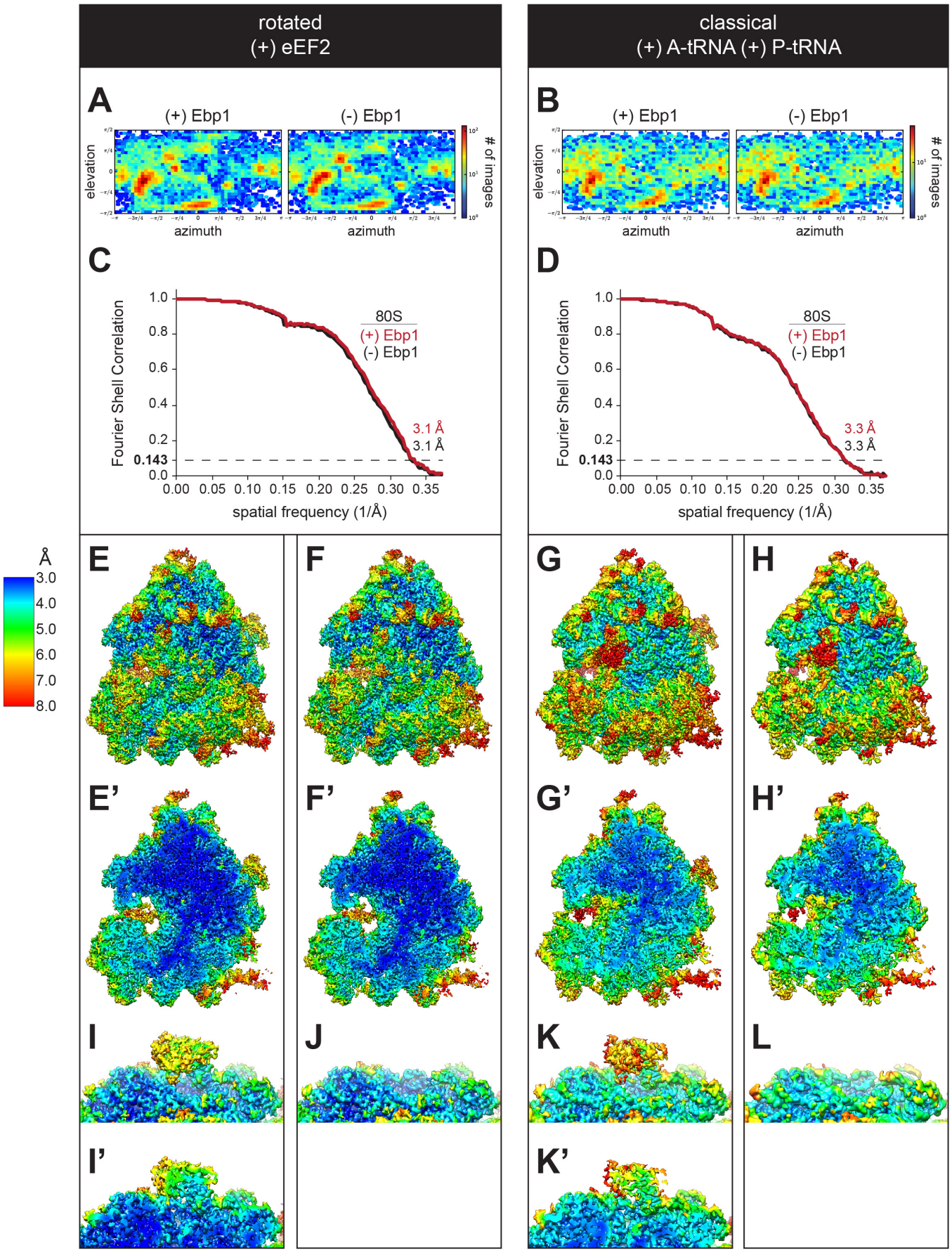
# Supplementary Figure 6

**Kraushar ML, *et al*. Revision.**



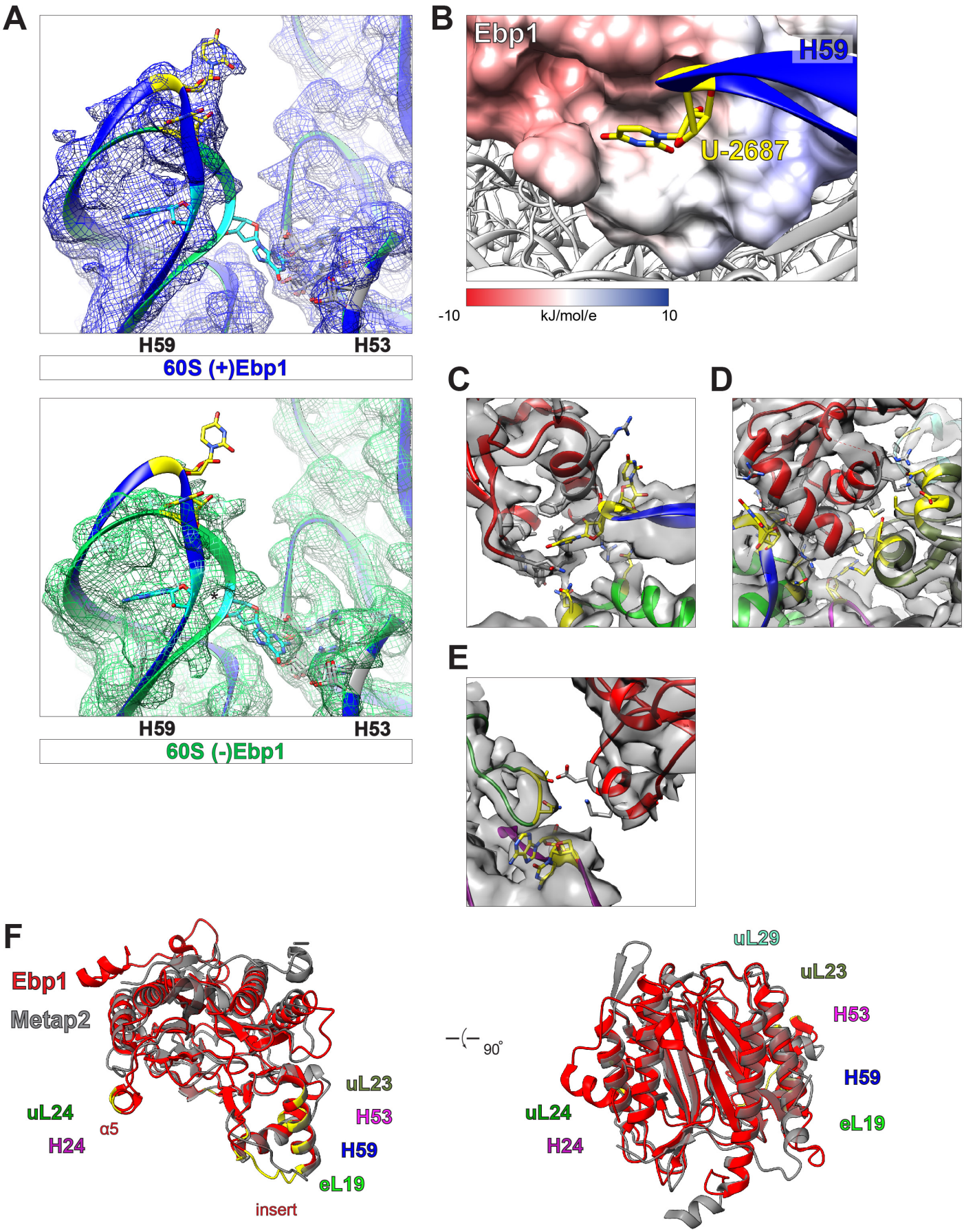
# Supplementary Figure 7

# Kraushar ML, *et al*. Revision.



# Supplementary Figure 8

Kraushar ML, *et al*. Revision.

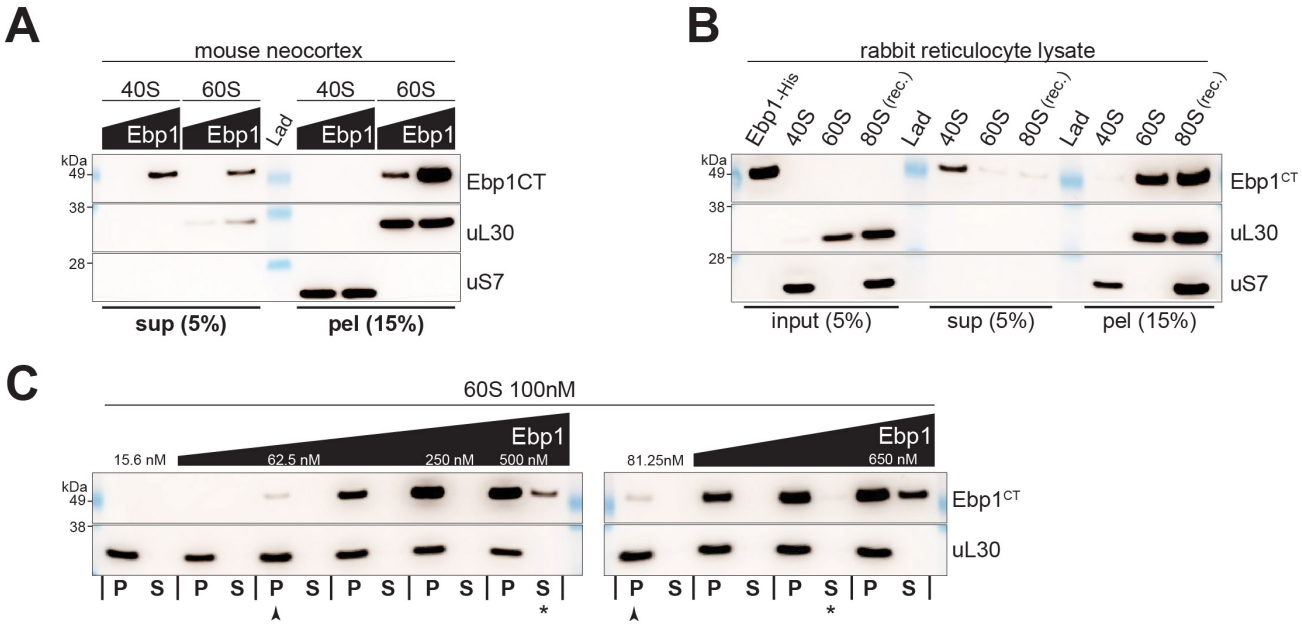


# Supplementary Figure 9

**Kraushar ML, *et al*. Revision.**

# Supplementary Figure 10

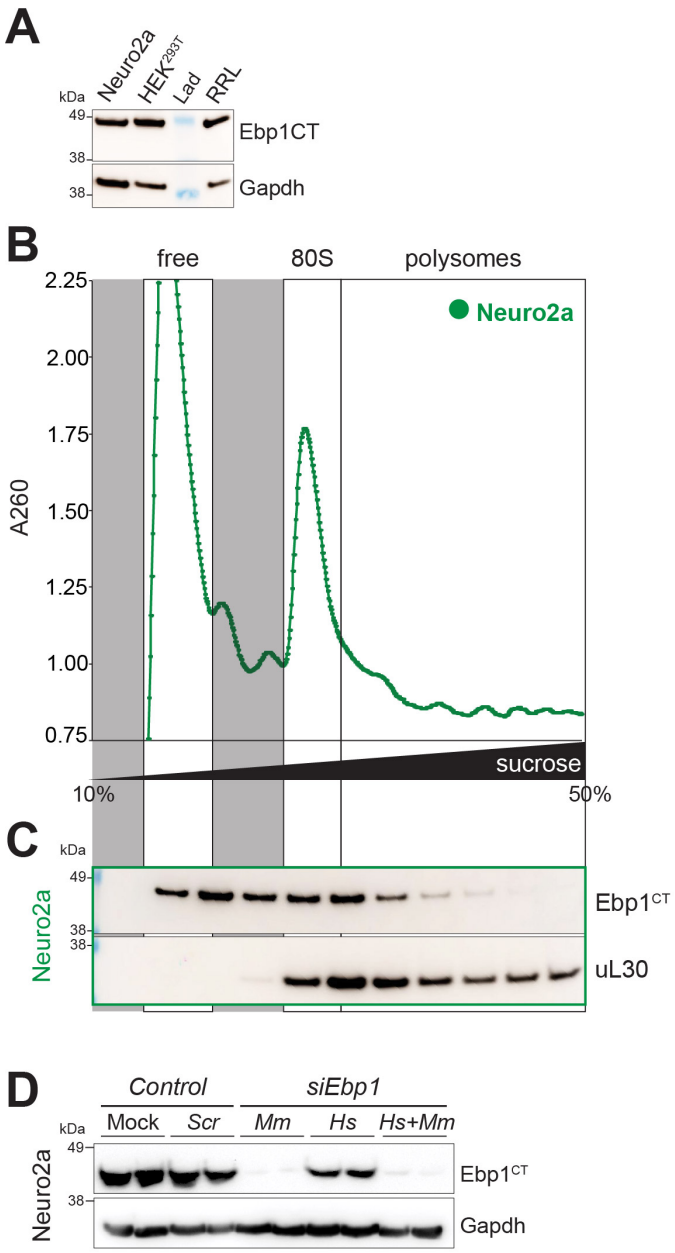
# Kraushar ML, *et al*. Revision.



# Supplementary Figure 11

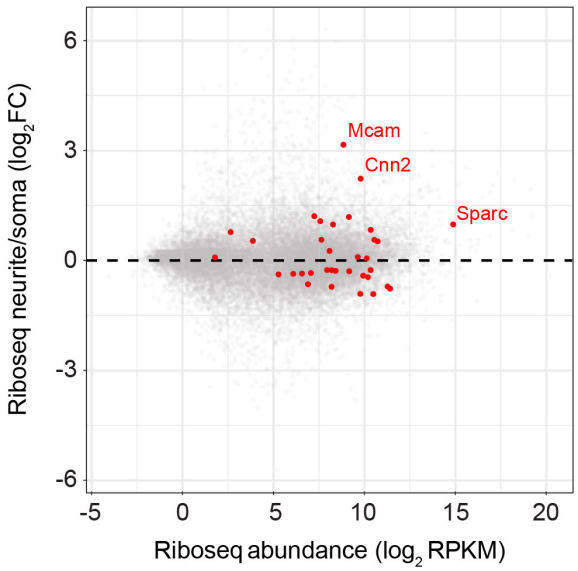
# Kraushar ML, *et al*. Revision.

#



# Supplementary Figure 12

# Kraushar ML, *et al*. Revision.



## 15.4 Publication IV

Banchenko S.*, Krupp F.*, Gotthold C., Bürger J., Graziadei A., O'Reilly F., Sinn L., Ruda O., Rappsilber J., Spahn C. M. T., Mielke T., Taylor I. and Schwefel D. Structural insights into Cullin4-RING ubiquitin ligase remodeling by Vpr from simian immunodeficiency viruses. *Under review PLOS Pathogens*

* denotes first authors

# Structural insights into Cullin4-RING ubiquitin ligase remodelling by Vpr from simian immunodeficiency viruses

Sofia Banchenko[1†], Ferdinand Krupp[1†], Christine Gotthold[1], Jörg Bürger[1,2], Andrea Graziadei[3], Francis O'Reilly[3], Ludwig Sinn[3], Olga Ruda[1], Juri Rappsilber[3,4], Christian M. T. Spahn[1], Thorsten Mielke[2], Ian A. Taylor[5], David Schwefel[1*]

[1] *Institute of Medical Physics and Biophysics, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany*

[2] *Microscopy and Cryo-Electron Microscopy Service Group, Max-Planck-Institute for Molecular Genetics, Berlin, Germany*

[3] *Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany*

[4] *Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh, UK*

[5] *Macromolecular Structure Laboratory, The Francis Crick Institute, London, UK.*

[†]These authors contributed equally to this work

*****Corresponding author

Email: david.schwefel@charite.de

**Abstract**

Viruses have evolved means to manipulate the host's ubiquitin-proteasome system, in order to down-regulate antiviral host factors. The Vpx/Vpr family of lentiviral accessory proteins usurp the substrate receptor DCAF1 of host Cullin4-RING ligases (CRL4), a family of modular ubiquitin ligases involved in DNA replication, DNA repair and cell cycle regulation. $CRL4^{DCAF1}$ specificity modulation by Vpx and Vpr from HIV-2 and simian immunodeficiency viruses (SIV) leads to recruitment, poly-ubiquitylation and subsequent proteasomal degradation of the antiviral host factor SAMHD1, resulting in enhanced virus replication. Here, we present integrative biochemical and structural analyses of the Vpr protein from SIVs infecting *Cercopithecus cephus* ($SIV_{mus}$). X-ray crystallography reveals commonalities in the Vpx/Vpr family with regard to the interaction with DCAF1, while cryo-electron microscopy and cross-linking mass spectrometry highlight a divergent molecular mechanism of SAMHD1 recruitment in $SIV_{mus}$ Vpr. In addition, structural analysis of an entire $SIV_{mus}$ Vpr-bound multi-subunit $CRL4^{DCAF1}$ assembly provides important insights into its dynamic architecture, with general implications for CRL4-catalysed ubiquitylation processes.

## Introduction

A large proportion of viruses have evolved means to co-opt their host's ubiquitylation machinery, in order to improve replication conditions, either by introducing viral ubiquitin ligases and deubiquitinases, or by modification of host proteins involved in ubiquitylation[1-3]. In particular, host ubiquitin ligases are a prominent target for viral usurpation, to redirect specificity towards antiviral host restriction factors. This results in recruitment of restriction factors as non-endogenous *neo*-substrates, inducing their poly-ubiquitylation and subsequent proteasomal degradation[4-8]. This counteraction of the host's antiviral repertoire is essential for virus infectivity and spread[9-11], and mechanistic insights into these specificity changes extend our understanding of viral pathogenesis and might pave the way for novel treatments.

Frequently, virally encoded modifying proteins associate with, and adapt the Cullin4-RING ubiquitin ligases (CRL4)[5]. CRL4 consists of a Cullin4 (CUL4) scaffold that bridges the catalytic RING-domain subunit ROC1 to the adaptor protein DDB1, which in turn binds to exchangeable substrate receptors (DCAFs, DDB1- and CUL4-associated factors)[12-16]. In some instances, the DDB1 adaptor serves as an anchor for virus proteins, which then act as "viral DCAFs" to recruit the antiviral substrate. Examples are the simian virus 5 V protein and mouse cytomegalovirus M27, which bind to DDB1 and recruit STAT1/2 proteins for ubiquitylation, in order to interfere with the host's interferon response[17-19]. Similarly, CUL4-dependent downregulation of STAT signalling is important for West Nile Virus replication[20]. In addition, the hepatitis B virus X protein hijacks DDB1 to induce proteasomal destruction of the structural maintenance of chromosome (SMC) complex to promote virus replication[21,22].

Viral factors also bind to and modify DCAF receptors in order to redirect them to antiviral substrates. Prime examples are the lentiviral accessory proteins Vpr and Vpx. All contemporary human and simian immunodeficiency viruses (HIV/SIV) encode Vpr, while only two lineages, represented by HIV-2 and SIV infecting mandrills, carry Vpx[23]. Vpr and Vpx proteins are packaged into progeny virions and released into the host cell upon infection, where they bind to DCAF1 in the nucleus[24]. In this work, corresponding simian immunodeficiency virus Vpx/Vpr proteins will be indicated with their host species as subscript, with the following abbreviations used: mus – moustached monkey (*Cercopithecus cephus*), mnd – mandrill (*Mandrillus sphinx*), rcm – red-capped mangabey (*Cercocebus torquatus*), sm

– sooty mangabey (*Cercocebus atys*), deb – De Brazza's monkey (*Cercopithecus neglectus*), syk – Syke's monkey (*Cercopithecus albogularis*), agm – african green monkey (*Chlorocebus* spec).

$Vpr_{HIV-1}$ is important for virus replication *in vivo* and in macrophage infection models[25]. Recent proteomic analyses revealed that DCAF1 specificity modulation by $Vpr_{HIV-1}$ proteins results in down-regulation of hundreds of host proteins in a DCAF1- and proteasome-dependent manner[26], including the previously reported $Vpr_{HIV-1}$ degradation targets UNG2[27], HLTF[28], MUS81[29,30], MCM10[31] and TET2[32]. This surprising promiscuity in degradation targets is also partially conserved in more distant clades exemplified by $Vpr_{agm}$ and $Vpr_{mus}$[26]. However, Vpr pleiotropy, and the lack of easily accessible experimental models, have prevented a characterisation of how these degradation events precisely promote replication[25].

By contrast, Vpx, exhibits a much narrower substrate range. It has recently been reported to target stimulator of interferon genes (STING) and components of the human silencing hub (HUSH) complex for degradation, leading to inhibition of antiviral cGAS-STING-mediated signalling and reactivation of latent proviruses, respectively[33-35]. Importantly, Vpx also recruits the SAMHD1 restriction factor to DCAF1, in order to mark it for proteasomal destruction[36,37]. SAMHD1 is a deoxynucleotide triphosphate (dNTP) triphosphohydrolase that restricts retroviral replication in non-dividing cells by lowering the dNTP pool to levels that cannot sustain viral reverse transcription[38-45]. Retroviruses that express Vpx are able to alleviate SAMHD1 restriction and allow replication in differentiated myeloid lineage cells, resting T cells and memory T cells[37,46,47]. As a result of the constant evolutionary "arms race" between the host's SAMHD1 restriction and its viral antagonist Vpx, the mechanism of Vpx-mediated SAMHD1 recruitment is highly virus species- and strain-specific: The Vpx clade represented by $Vpx_{HIV-2}$ recognises the SAMHD1 C-terminal domain (CtD), while $Vpx_{mnd2/rcm}$ binds the SAMHD1 N-terminal domain (NtD) in a fundamentally different way[23,48-51].

In the course of evolutionary adaptation to their primate hosts, and due to selective pressure to evade SAMHD1 restriction, two groups of SIVs that do not have Vpx, $SIV_{agm}$, and $SIV_{deb/mus/syk}$, neo-functionalised their Vpr to bind SAMHD1 and induce its degradation[23,48,52]. Consequently, these species evolved "hybrid" Vpr proteins that retain targeting of some host factors depleted by HIV-1-type Vpr[26], and additionally induce SAMHD1 degradation.

Here we present integrative biochemical and structural analyses of a "hybrid" Vpr isolated from SIV infecting *Cercopithecus cephus*, $Vpr_{mus}$. These studies reveal similarities and differences to Vpx and Vpr proteins from other lentivirus species and pinpoint the divergent molecular mechanism of $Vpr_{mus}$-dependent SAMHD1 recruitment to $CRL4^{DCAF1}$. Furthermore, cryo-electron microscopic (EM) reconstructions of a $Vpr_{mus}$-modified $CRL4^{DCAF1}$ protein complex allow for insights into the structural plasticity of the entire CRL4 ubiquitin ligase assembly with implications for the ubiquitin transfer mechanism.

**Results**

*SAMHD1-CtD is necessary and sufficient for $Vpr_{mus}$-binding and ubiquitylation in vitro*

To investigate the molecular interactions between $Vpr_{mus}$, the *neo*-substrate SAMHD1 from rhesus macaque and CRL4 subunits DDB1/DCAF1 C-terminal domain (DCAF1-CtD), protein complexes were reconstituted *in vitro* from purified components and analysed by gel filtration (GF) chromatography. The different protein constructs that were employed are shown schematically in Fig. S1A. $Vpr_{mus}$ is insoluble after removal of the GST affinity purification tag (Fig. S1B) and accordingly could not be applied to the GF column. No interaction of SAMHD1 with DDB1/DCAF1-CtD could be detected in the absence of $Vpr_{mus}$ (Fig. S1C). Analysis of binary protein combinations ($Vpr_{mus}$ and DDB1/DCAF1-CtD; $Vpr_{mus}$ and SAMHD1) shows that $Vpr_{mus}$ elutes in a single peak together with DDB1/DCAF1-CtD (Fig. S1D) or with SAMHD1 (Fig. S1E). Incubation of $Vpr_{mus}$ with DDB1/DCAF1B and SAMHD1 followed by GF resulted in elution of all three components in a single peak (Fig. 1A, B, red trace). Together, these results show that $Vpr_{mus}$ forms stable binary and ternary protein complexes with DDB1/DCAF1-CtD and/or SAMHD1 *in vitro*. Furthermore, incubation with any of these interaction partners apparently stabilises $Vpr_{mus}$ by alleviating its tendency for aggregation/insolubility.

Previous cell-based assays indicated that residues 583-626 of rhesus macaque SAMHD1 (SAMHD1-CtD) are necessary for $Vpr_{mus}$-induced proteasomal degradation[48]. To test this finding in our *in vitro* system, constructs containing SAMHD1-CtD fused to T4 lysozyme (T4L-SAMHD1-CtD), or lacking SAMHD1-CtD (SAMHD1-ΔCtD, Fig. 1A), were incubated with $Vpr_{mus}$ and DDB1/DCAF1-CtD, and complex formation was assessed by GF chromatography. Analysis of the resulting chromatograms by

SDS-PAGE shows that SAMHD1-ΔCtD did not co-elute with DDB1/DCAF1-CtD/$Vpr_{mus}$ (Fig. 1A, B, green trace). By contrast, T4L-SAMHD1-CtD accumulated in a single peak, which also contained DDB1/DCAF1-CtD and $Vpr_{mus}$ (Fig. 1A, B, cyan trace). These results confirm that SAMHD1-CtD is necessary for stable association with DDB1/DCAF1-CtD/$Vpr_{mus}$ *in vitro*, and demonstrate that SAMHD1-CtD is sufficient for $Vpr_{mus}$-mediated recruitment of the T4L-SAMHD1-CtD fusion construct to DDB1/DCAF1-CtD.

To correlate these data with enzymatic activity, *in vitro* ubiquitylation assays were conducted by incubating SAMHD1, SAMHD1-ΔCtD or T4L-SAMHD1-CtD with purified $CRL4^{DCAF1\text{-}CtD}$, E1 (UBA1), E2 (UBCH5C), ubiquitin and ATP. Input proteins are shown in Fig. S2A, and control reactions in Fig. S2B, C. In the absence of $Vpr_{mus}$, no SAMHD1 ubiquitylation was observed (Fig. 1C, S2D), while addition of $Vpr_{mus}$ resulted in robust SAMHD1 ubiquitylation (Fig. 1D, S2E). In agreement with the analytical GF data, SAMHD1-ΔCtD was not ubiquitylated in the presence of $Vpr_{mus}$ (Fig. 1E, S2F), while T4L-SAMHD1-CtD, was ubiquitylated with similar kinetics as the full-length protein (Fig. 1F, S2F). Again, these data substantiate the functional importance of SAMHD1-CtD for $Vpr_{mus}$-mediated recruitment to the $CRL4^{DCAF1}$ ubiquitin ligase.

*Crystal Structure analysis of apo- and $Vpr_{mus}$-bound DDB1/DCAF1-CtD protein complexes*

To obtain structural information regarding $Vpr_{mus}$ and its mode of binding to the CRL4 substrate receptor DCAF1, the X-ray crystal structures of a DDB1/DCAF1-CtD complex, and DDB1/DCAF1-CtD/T4L-$Vpr_{mus}$ (residues 1-92) fusion protein ternary complex were determined. The structures were solved using molecular replacement and refined to resolutions of 3.1 Å and 2.5 Å respectively (Table S1). $Vpr_{mus}$ adopts a three-helix bundle fold, stabilised by coordination of a zinc ion by His and Cys residues on Helix-1 and at the C-terminus (Fig. 2A). Superposition of $Vpr_{mus}$ with previously determined $Vpx_{sm}$[49], $Vpx_{mnd2}$[50,51], and $Vpr_{HIV\text{-}1}$[53] structures reveals a conserved three-helix bundle fold, and similar position of the helix bundles on DCAF1-CtD (Fig. S3A). In addition, the majority of side chains involved in DCAF1-interaction are type-conserved in all Vpx and Vpr proteins (Fig. S3B-G, Fig. S6A), strongly suggesting a common molecular mechanism of host CRL4-DCAF1 hijacking by the Vpx/Vpr family of accessory proteins. However, there are also significant differences in helix length and register as well

as conformational variation in the loop region N-terminal of Helix-1, at the start of Helix-1 and in the loop between Helices-2 and -3 (Fig. S3A).

$Vpr_{mus}$ binds to the side and on top of the disk-shaped 7-bladed β-propeller (BP) DCAF1-CtD domain with a total contact surface area of ~1600 $Å^2$ comprising three major regions of interaction. The extended $Vpr_{mus}$ N-terminus attaches to the cleft between DCAF1 BP blades 1 and 2 through several hydrogen bonds, electrostatic and hydrophobic interactions (Fig. S3B-D). A second, smaller contact area is formed by hydrophobic interaction between $Vpr_{mus}$ residues L31 and E34 from Helix-1, and DCAF1 W1156, located in a loop on top of BP blade 2 (Fig. S3E). The third interaction surface comprises the C-terminal half of $Vpr_{mus}$ Helix-3, which inserts into a ridge on top of DCAF1 (Fig. S3F, G).

Superposition of the apo-DDB1/DCAF1-CtD and $Vpr_{mus}$-bound crystal structures reveals conformational changes in DCAF1 upon $Vpr_{mus}$ association. Binding of the N-terminal arm of $Vpr_{mus}$ induces only a minor rearrangement of a loop in BP blade 2 (Fig. S3C). By contrast, significant structural changes occur on the upper surface of the BP domain: polar and hydrophobic interactions of DCAF1 residues P1329, F1330, F1355, N1371, L1378, M1380 and T1382 with $Vpr_{mus}$ side chains of T79, R83, R86 and E87 in Helix-3 result in the stabilisation of the sequence stretch that connect BP blades 6 and 7 (“C-terminal loop”, Fig. 2B, S3F). Moreover, side chain electrostatic interactions of $Vpr_{mus}$ residues R15, R75 and R76 with DCAF1 E1088, E1091 and E1093 lock the conformation of an “acidic loop” upstream of BP blade 1, which is also unstructured and flexible in the absence of $Vpr_{mus}$ (Fig. 2B, C, S3D, F).

Notably, in previously determined structures of Vpx/DCAF1/SAMHD1 complexes the “acidic loop” is a central point of ternary contact, providing a binding platform for positively charged amino acid side chains in either the SAMHD1 N- or C-terminus[49-51]. For example, $Vpx_{sm}$ positions SAMHD1-CtD in such a way, that SAMHD1 K622 engages in electrostatic interaction with the DCAF1 “acidic loop” residue D1092 (Fig. 2C, left panel). However, in the $Vpr_{mus}$ crystal structure the bound $Vpr_{mus}$ now blocks access to the corresponding SAMHD1-CtD binding pocket, in particular by the positioning of an extended N-terminal loop that precedes Helix-1. Additionally, $Vpr_{mus}$ side chains R15, R75 and R76 neutralise the DCAF1 “acidic loop”, precluding the formation of further salt bridges to basic residues in SAMHD1-CtD (Fig. 2C, right panel).

To validate the importance of $Vpr_{mus}$ residues R15 and R75 for DCAF1-CtD- and SAMHD1-binding, charge reversal mutations to glutamates were generated by site-directed mutagenesis. The effect of the $Vpr_{mus}$ R15E R75E double mutant on complex assembly was then analysed by GF chromatography. SDS-PAGE analysis of the resulting chromatographic profile shows an almost complete loss of the DDB1/DCAF1-CtD/$Vpr_{mus}$/SAMHD1 complex peak (Fig. 2D, fraction 6), when compared to the wild type, concomitant with enrichment of (i) $Vpr_{mus}$ R15E R75E-bound DDB1/DCAF1-CtD (Fig. 2D, fractions 7-8), and of (ii) $Vpr_{mus}$ R15E R75E/SAMHD1 binary complex (Fig. 2D, fraction 8-9). This suggests that charge reversal of $Vpr_{mus}$ side chains R15 and R75 weakens the strong association with DCAF1 observed in wild type $Vpr_{mus}$, due to loss of electrostatic interaction with the "acidic loop", in accordance with the crystal structure. Consequently, some proportion of Vpr-bound SAMHD1 dissociates, further indicating that $Vpr_{mus}$ side chains R15 and R75 are not central to SAMHD1 interaction.

*Molecular mechanism of SAMHD1-targeting*

To obtain mechanistic insight into $Vpr_{mus}$-recruitment of SAMHD1-CtD, we initiated cryo-EM analyses of the $CRL4^{DCAF1\text{-}CtD}$/$Vpr_{mus}$/SAMHD1 assembly. In these studies, the small ubiquitin-like protein NEDD8 was enzymatically attached to the CUL4 subunit, in order to obtain its active form (Fig. S4A)[54]. A CRL4-$NEDD8^{DCAF1\text{-}CtD}$/$Vpr_{mus}$/SAMHD1 complex was reconstituted *in vitro* and purified by GF chromatography (Fig. S4B). Extensive 2D and 3D classification of the resulting particle images revealed considerable conformational heterogeneity, especially regarding the position of the CUL4-NEDD8/ROC1 subcomplex (stalk) relative to DDB1/DCAF1/$Vpr_{mus}$ (core), (Fig. S4).

Nevertheless, a homogeneous particle population could be separated, which yielded a 3D reconstruction at a nominal resolution of 7.3 Å that contained electron density corresponding to the core (Fig. S4C-F). Molecular models of DDB1 BP domains A and C (BPA, BPC), DCAF1-CtD and $Vpr_{mus}$, derived from our crystal structure (Fig. 2), could be fitted as rigid bodies into this cryo-EM volume (Fig. 3A). No obvious electron density was visible for the bulk of SAMHD1. However, close inspection revealed an additional tubular, slightly arcing density feature, approx. 35 Å in length, located on the upper surface of the $Vpr_{mus}$ helix bundle, approximately 17 Å away from and opposite of the $Vpr_{mus}$/DCAF1-CtD

binding interface (Fig. 3A, red arrows). One end of the tubular volume contacts the middle of $Vpr_{mus}$ Helix-1, and the other end forms additional contacts to the C-terminus of Helix-2 and the N-terminus of Helix-3. A local resolution of 7.5-8 Å precluded the fitting of an atomic model. Considering the biochemical data, showing that SAMHD1-CtD is sufficient for recruitment to DDB1/DCAF1/$Vpr_{mus}$, we hypothesise that this observed electron density feature corresponds to the region of SAMHD1-CtD which physically interacts with $Vpr_{mus}$. Given its dimensions, the putative SAMHD1-CtD density could accommodate approx. 10 amino acid residues in a fully extended conformation or up to 23 residues in a kinked helical arrangement. All previous crystal structure analyses[45], as well as secondary structure predictions indicate that SAMHD1 residues C-terminal to the catalytic HD domain and C-terminal lobe (amino acids 599-626) are disordered in the absence of additional binding partners. Accordingly, the globular domains of the SAMHD1 molecule might be flexibly linked to the C-terminal tether identified here. In that case, the bulk of SAMHD1 samples a multitude of positions relative to the DDB1/DCAF1-CtD/$Vpr_{mus}$ core, and consequently is averaged out in the process of cryo-EM reconstruction.

The topology of $CRL4^{DCAF1\text{-}CtD}$/$Vpr_{mus}$/SAMHD1 and the binding region of SAMHD1-CtD were further assessed by cross-linking mass spectrometry (CLMS) using the photo-reactive cross-linker sulfo-SDA[55]. A large number of cross-links between SAMHD1 and the C-terminal half of CUL4, the side and top of DCAF1-CtD, and BP blades 6-7 of DDB1 were found, consistent with highly variable positioning of the SAM and HD domains of SAMHD1 relative to the CRL4 core (Fig. 3B). Moreover, multiple cross-links between SAMHD1-CtD and $Vpr_{mus}$ were observed, more specifically locating to a sequence stretch comprising the C-terminal half of $Vpr_{mus}$ Helix-1 (residues A27-E36), and to a portion of the disordered $Vpr_{mus}$ C-terminus (residues Y90, Y100). These data are in accordance with the presence of SAMHD1-CtD in the unassigned cryo-EM density and its role as $Vpr_{mus}$ tether. The remaining SAMHD1-CtD cross-links were with the C-terminus of CUL4 and the “acidic loop” of DCAF1 (Fig. 3B). Distance restraints from these SAMHD1-CtD cross-links, together with our structural models of $CRL4^{DCAF1\text{-}CtD}$/$Vpr_{mus}$ (see below, Fig. 4), were employed to visualise the interaction space accessible to the centre of mass of SAMHD1-CtD. This analysis is compatible with recruitment of SAMHD1-CtD on top of the $Vpr_{mus}$ helix bundle as indicated by cryo-EM (Fig. 3C). Interestingly, cross-links to $Vpr_{mus}$ were restricted to the C-terminal end of SAMHD1-CtD (residues K622, K626), while cross-links to CUL4

and DCAF1 were found in the N-terminal portion (residues K595, K596, T602-S606). These observations are consistent with a model where the very C-terminus of SAMHD1 is immobilised on $Vpr_{mus}$, and SAMHD1-CtD residues further upstream are exposed to the catalytic machinery surrounding the CUL4 C-terminal domain.

To further probe the interaction, $Vpr_{mus}$ amino acid residues in close proximity to the putative SAMHD1-CtD density were substituted by site-directed mutagenesis. Specifically, $Vpr_{mus}$ W29 was changed to alanine to block a hydrophobic contact with SAMHD1-CtD involving the aromatic side chain, and $Vpr_{mus}$ A66 was changed to a bulky tryptophan, in order to introduce a steric clash with SAMHD1-CtD (Fig. 3D). This $Vpr_{mus}$ W29A A66W double mutant was then assessed for complex formation with DDB1/DCAF1-CtD and SAMHD1 by analytical GF. In comparison to wild type $Vpr_{mus}$, the W29A A66W mutant showed a reduction of DDB1/DCAF1-CtD/$Vpr_{mus}$/SAMHD1 complex peak intensity (Fig. 3E, fraction 6), concomitant with (i) enrichment of DDB1/DCAF1-CtD/$Vpr_{mus}$ ternary complex, sub-stoichiometrically bound to SAMHD1 (Fig. 3E, fraction 7), (ii) excess DDB1/DCAF1-CtD binary complex (Fig. 3E, fraction 8), and (iii) monomeric SAMHD1 species (Fig. 3E, fractions 9-10). In conclusion, this biochemical analysis, together with cryo-EM reconstruction at intermediate resolution and CLMS analysis, locate the SAMHD1-CtD binding site on the upper surface of the $Vpr_{mus}$ helix bundle.

These data allow for structural comparison with *neo*-substrate binding modes of Vpx and Vpr proteins from different retrovirus lineages (Fig. 4A-D). $Vpx_{HIV-2}$ and $Vpx_{sm}$ position SAMHD1-CtD at the side of the DCAF1 BP domain through interactions with the N-termini of Vpx Helices-1 and -3 (Fig. 4B)[49]. $Vpx_{mnd2}$ and $Vpx_{rcm}$ bind SAMHD1-NtD using a bipartite interface comprising the side of the DCAF1 BP and the upper surface the Vpx helix bundle (Fig. 4C)[50,51]. $Vpr_{HIV-1}$ engages its ubiquitylation substrate UNG2 using both the top and the upper edge of the $Vpr_{HIV-1}$ helix bundle (Fig. 4D)[53]. Of note, these upper-surface interaction interfaces only partially overlap with the $Vpr_{mus}$/SAMHD1-CtD binding interface identified here and employ fundamentally different sets of interacting amino acid residues. Thus, it appears that the molecular interaction interfaces driving Vpx/Vpr-mediated *neo*-substrate recognition and degradation are not conserved between related SIV and HIV Vpx/Vpr accessory proteins, even in cases where identical SAMHD1-CtD regions are targeted for recruitment.

*Cryo-EM analysis of $Vpr_{mus}$-modified CRL4-NEDD8$^{DCAF1\text{-}CtD}$ conformational states and dynamics*

A reanalysis of the cryo-EM data using strict selection of high-quality 2D classes, followed by focussed 3D classification yielded three additional particle populations, resulting in 3D reconstructions at 8-10 Å resolution, which contained both the $Vpr_{mus}$-bound CRL4 core and the stalk (conformational states-1, -2 and -3, Fig. 5A, S4G-J). The quality of the 3D volumes was sufficient to fit crystallographic models of core (Fig. 2) and the stalk (PDB 2hye)[14] as rigid bodies (Fig. 5B, S5A). For the catalytic RING-domain subunit ROC1, only fragmented electron density was present near the position it occupies in the crystallographic model (Fig. S5A). In all three states, electron density was selectively absent for the C-terminal CUL4 winged helix B (WHB) domain (residues 674-759), which contains the NEDD8 modification site (K705), and for the preceding α-helix, which connects the CUL4 N-terminal domain to the WHB domain (Fig. S5A). In accordance with this observation, the positions of CRL5-attached NEDD8 and of the CRL4 ROC1 RING domain are sterically incompatible upon superposition of their respective crystal structures (Fig. S5B).

Alignment of 3D volumes from states-1, -2 and -3 shows that core densities representing DDB1 BPA, BPC, DCAF1-CtD and $Vpr_{mus}$ superimpose well, indicating that these components do not undergo major conformational fluctuations and thus form a rigid platform for substrate binding and attachment of the CRL4 stalk (Fig. 5). However, rotation of DDB1 BPB around a hinge connecting it to BPC results in three different orientations of state-1, -2 and -3 stalk regions relative to the core. BPB rotation angles were measured as 69° between state-1 and -2, and 50° between state-2 and -3. Furthermore, the crosslinks between DDB1 and CUL4 identified by CLMS are satisfied by the state-1 model, but increasingly violated in states-2 and -3, validating in solution the conformational variability observed by cryo-EM. (Fig. S5C). Taken together, this places the CRL4 catalytic machinery, sited at the distal end of the stalk, appropriately to approach the $Vpr_{mus}$-tethered bulk of SAMHD1 for ubiquitylation at a wide range of angles (Fig. 5B).

These data are in line with previous prediction based on extensive comparative crystal structure analyses, which postulated an approx. 150° rotation of the CRL4 stalk around the core[12,14,15,18,56]. However, the left- and rightmost CUL4 orientations observed here, states-1 and -3 from our cryo-EM

analysis, indicate a slightly narrower stalk rotation range (119°), when compared to the outermost stalk conformations modelled from previously determined crystal structures (143°) (Fig. S5D). An explanation for this discrepancy comes from inspection of the cryo-EM densities and fitted models, revealing that along with the main interaction interface on DDB1 BPB there are additional molecular contacts between CUL4 and DDB1. Specifically, in state-1, there is a contact between the loop connecting helices D and E of CUL4 cullin repeat (CR)1 (residues 161-169) and a loop protruding from BP blade 3 of DDB1 BPC (residues 795-801, Fig. S5E). In state-3, the loop between CUL4 CR2 helices D and E (residues 275-282) abuts a region in the C-terminal helical domain of DDB1 (residues 1110-1127, Fig. S5F). These auxiliary interactions might be required to lock the outermost stalk positions observed here in order to confine the rotation range of CUL4.

**Discussion**

Our X-ray crystallographic studies of the DDB1/DCAF1-CtD/$Vpr_{mus}$ assembly provide the first structural insight into a class of "hybrid" SIV Vpr proteins. These are present in the $SIV_{agm}$ and $SIV_{mus/deb/syk}$ lineages of lentiviruses and combine characteristics of related $Vpr_{HIV-1}$ and SIV Vpx accessory proteins.

Like SIV Vpx, "hybrid" Vpr proteins down-regulate the host restriction factor SAMHD1 by recruiting it to $CRL4^{DCAF1}$ for ubiquitylation and subsequent proteasomal degradation. However, using a combination of X-ray, cryo-EM and CLMS analyses, we show that the molecular strategy, which $Vpr_{mus}$ evolved to target SAMHD1, is strikingly different from Vpx-containing SIV strains. In the two clades of Vpx proteins, divergent amino acid sequence stretches just upstream of Helix-1 (variable region (VR)1, Fig. S6A), together with polymorphisms in the SAMHD1-N-terminus of the respective host species, determine if HIV-2-type or $SIV_{mnd}$-type Vpx recognise SAMHD1-CtD or SAMHD1-NtD, respectively. These recognition mechanisms result in positioning of SAMHD1-CtD or -NtD on the side of the DCAF1 BP domain in a way that allows for additional contacts between SAMHD1 and DCAF1, thus forming ternary Vpx/SAMHD1/DCAF1 assemblies with very low dissociation rates[49-51,57]. In $Vpr_{mus}$, different principles determine the specificity for SAMHD1-CtD. Here, VR1 is not involved in SAMHD1-CtD-binding at all, but forms additional interactions with DCAF1, which are not observed in

Vpx/DCAF1 protein complexes (Fig. S6A). Molecular contacts between $Vpr_{mus}$ and SAMHD1 are dispersed on Helices-1 and -3, facing away from the DCAF1 interaction site and immobilising SAMHD1-CtD on the top side of the $Vpr_{mus}$ helix bundle (Fig. S6A). Placement of SAMHD1-CtD in such a position precludes stabilising ternary interaction with DCAF1-CtD, but still results in robust SAMHD1 ubiquitylation *in vitro* and SAMHD1 degradation in cell-based assays[23].

Predictions regarding the molecular mechanism of SAMHD1-binding by other "hybrid" Vpr orthologues are difficult due to sequence divergence. Even in $Vpr_{deb}$, the closest relative to $Vpr_{mus}$, only approximately 50% of amino acid side chains lining the putative SAMHD1-CtD binding pocket are conserved (Fig. S6A). Previous *in vitro* ubiquitylation and cell-based degradation experiments did not show a clear preference of $Vpr_{deb}$ for recruitment of either SAMHD1-NtD or -CtD[23,48]. Furthermore, it is disputed if $Vpr_{deb}$ actually binds DCAF1[58], which might possibly be explained by amino acid variations in the very N-terminus and/or in Helix-3 (Fig. S6A). $Vpr_{syk}$ is specific for SAMHD1-CtD[48], but the majority of residues forming the binding platform for SAMHD1-CtD observed in the present study are not conserved. The $SIV_{agm}$ lineage of Vpr proteins is even more divergent, with significant differences not only in possible SAMHD1-contacting residues, but also in the sequence stretches preceding Helix-1, and connecting Helices-2 and -3, as well as in the N-terminal half of Helix-3 (Fig. S6A). Furthermore, there are indications that recruitment of SAMHD1 by the $Vpr_{agm.GRI}$ sub-type involves molecular recognition of both SAMHD1-NtD and –CtD[48,52]. In conclusion, recurring rounds of evolutionary lentiviral adaptation to the host SAMHD1 restriction factor, followed by host re-adaptation, resulted in highly species-specific, diverse molecular modes of Vpr-SAMHD1 interaction. In addition to the example presented here, further structural characterisation of SAMHD1-Vpr complexes will be necessary to illustrate the manifold outcomes of this particular virus-host molecular "arms race".

Previous structural investigation of DDB1/DCAF1/$Vpr_{HIV-1}$ in complex with the neo-substrate UNG2 demonstrated that $Vpr_{HIV-1}$ engages UNG2 by mimicking the DNA phosphate backbone. More precisely, UNG2 residues, which project into the major groove of its endogenous DNA substrate, insert into a hydrophobic cleft formed by $Vpr_{HIV-1}$ Helices-1, -2 and the N-terminal half of Helix-3[53]. This mechanism might rationalise $Vpr_{HIV-1}$'s extraordinary binding promiscuity, since the list of potential $Vpr_{HIV-1}$

degradation substrates is significantly enriched in DNA- and RNA-binding proteins[26]. Moreover, promiscuous $Vpr_{HIV-1}$-induced degradation of host factors with DNA- or RNA-binding activity has been proposed to induce cell cycle arrest at the G2/M phase border, which is the most thoroughly described phenotype of Vpr proteins so far[25,26,59]. In $Vpr_{mus}$, the N-terminal half of Helix-1 as well as the bulky amino acid residue W48, which is also conserved in $Vpr_{agm}$ and Vpx, constrict the hydrophobic cleft (Fig. S6A, B). Furthermore, the extended N-terminus of $Vpr_{mus}$ Helix-3 is not compatible with UNG2-binding due to steric exclusion (Fig. S6C). In accordance with these observations, $Vpr_{mus}$ does not down-regulate UNG2 in a human T cell line[26]. However, $Vpr_{mus}$, $Vpr_{syk}$ and $Vpr_{agm}$ also cause G2/M cell cycle arrest in their respective host cells[58,60,61]. This strongly hints at the existence of further structural determinants in $Vpr_{mus}$, $Vpr_{syk}$, $Vpr_{agm}$ and potentially $Vpr_{HIV-1}$, which regulate recruitment and ubiquitylation of DNA/RNA-binding host factors, in addition to the hydrophobic, DNA-mimicking cleft on top of the three-helix bundle. Future efforts to structurally characterise these determinants will further extend our understanding of how the Vpx/Vpr helical scaffold binds, and in this way adapts to a multitude of *neo*-substrate epitopes. In addition, such efforts might inform approaches to design novel $CRL4^{DCAF1}$-based synthetic degraders, in the form of proteolysis-targeting chimera-(PROTAC-) type compounds[62,63].

Our cryo-EM reconstructions of $CRL4^{DCAF1\text{-}CtD}$/$Vpr_{mus}$/SAMHD1, complemented by CLMS, also provide insights into the structural dynamics of CRL4 assemblies prior to ubiquitin transfer. The data confirm previously described rotational movement of the CRL4 stalk, in the absence of constraints imposed by a crystal lattice, creating a ubiquitylation zone around the $Vpr_{mus}$-modified substrate receptor (Fig. 5, Fig. 6A)[12,14,15,18,56]. Missing density for the neddylated CUL4 WHB domain and for the catalytic ROC1 RING domain indicates that these distal stalk elements are highly mobile and likely sample a multitude of orientations relative to the CUL4 scaffold (Fig. 6B). These observations are in line with structure analyses of CRL1 and CRL5, where CUL1/5 neddylation leads to re-orientation of the cullin WHB domain, and to release of the ROC1 RING domain from the cullin scaffold, concomitant with stimulation of ubiquitylation activity[64]. Moreover, recent cryo-EM structure analysis of $CRL1^{\beta\text{-}TRCP}$/IκBα demonstrated substantial mobility of pre-catalytic NEDD8-CUL1 WHB and ROC1 RING domains[65]. Such flexibility seems necessary to structurally organise multiple CRL1-dependent

processes, in particular the nucleation of a catalytic assembly, involving intricate protein-protein interactions between NEDD8, CUL1, ubiquitin-charged E2 and substrate receptor. This synergistic assembly then steers the ubiquitin C-terminus towards a substrate lysine for priming with ubiquitin[65]. Accordingly, our cryo-EM studies might indicate that similar principles apply for CRL4-catalysed ubiquitylation. However, to unravel the catalytic architecture of CRL4, sophisticated cross-linking procedures as in reference (65) will have to be pursued.

Intrinsic mobility of CRL4 stalk elements might assist the accommodation of a variety of sizes and shapes of substrates in the CRL4 ubiquitylation zone and might rationalise the wide substrate range accessible to CRL4 ubiquitylation through multiple DCAF receptors. Owing to selective pressure to counteract the host's SAMHD1 restriction, HIV-2 and certain SIVs, amongst other viruses, have taken advantage of this dynamic CRL4 architecture by modification of the DCAF1 substrate receptor with Vpx/Vpr-family accessory proteins. By tethering either SAMHD1-CtD or -NtD to DCAF1, and in this way flexibly recruiting the bulk of SAMHD1, the accessibility of lysine side chains both tether-proximal and on the SAMHD1 globular domains to the CRL4 catalytic assembly might be further improved (Fig. 6C, D). This ensures efficient Vpx/Vpr-mediated SAMHD1 priming, poly-ubiquitylation and proteasomal degradation to stimulate virus replication.

## Materials and methods

*Protein expression and purification*

Constructs were PCR-amplified from cDNA templates and inserted into the indicated expression plasmids using standard restriction enzyme methods (Supplementary Table 2). pAcGHLT-B-DDB1 (plasmid #48638) and pET28-UBA1 (plasmid #32534) were obtained from Addgene. The pOPC-UBA3-GST-APPBP1 co-expression plasmid, and the pGex6P2-UBC12 plasmid were obtained from MRC-PPU Reagents and Services (clones 32498, 3879). Bovine erythrocyte ubiquitin and recombinant hsNEDD8 were purchased from Sigma-Aldrich (U6253) and BostonBiochem (UL-812) respectively. Point mutations were introduced by site-directed mutagenesis using KOD polymerase (Novagen). All constructs and variants are summarised Supplementary Table 3.

Proteins expressed from vectors pAcGHLT-B, pGex6P1/2, pOPC and pET49b contained an N-terminal GST-His-tag; pHisSUMO – N-terminal His-SUMO-tag; pET28, pRSF-Duet-1 – N-terminal His-tag; pTri-Ex-6 – C-terminal His-tag. Constructs in vectors pAcGHLT-B and pTri-Ex-6 were expressed in Sf9 cells, and constructs in vectors pET28, pET49b, pGex6P1/2, pRSF-Duet-1, and pHisSUMO in *E. coli* Rosetta 2(DE3).

Recombinant baculoviruses (*Autographa californica nucleopolyhedrovirus* clone C6) were generated as described previously[66]. Sf9 cells were cultured in Insect-XPRESS medium (Lonza) at 28°C in an Innova 42R incubator shaker (New Brunswick) at a shaking speed of 180 rpm. In a typical preparation, 1 L of Sf9 cells at $3\times10^6$ cells/mL were co-infected with 4 mL of high titre DDB1 virus and 4 mL of high titre DCAF1-CtD virus for 72 h.

For a typical *E. coli* Rosetta 2(DE3) expression, 2 L of LB medium was inoculated with 20 mL of an overnight culture and grown in a Multitron HT incubator shaker (Infors) at 37°C, 150 rpm until $OD_{600}$ reached 0.7. At that point, temperature was reduced to 18°C, protein expression was induced by addition of 0.2 mM IPTG, and cultures were grown for further 20 h. During co-expression of CUL4 and ROC1 from pRSF-Duet, 50 μM zinc sulfate was added to the growth medium before induction.

Sf9 cells were pelleted by centrifugation at 1000 rpm, 4°C for 30 min using a JLA 9.1000 centrifuge rotor (Beckman). *E. coli* cells were pelleted by centrifugation at 4000 rpm, 4°C for 15 min using the same rotor. Cell pellets were resuspended in buffer containing 50 mM Tris, pH 7.8, 500 mM NaCl, 4 mM MgCl2, 0.5 mM tris-(2-carboxyethyl)-phosphine (TCEP), mini-complete protease inhibitors (1 tablet per 50 mL) and 20 mM imidazole (for His-tagged proteins only). 100 mL of lysis buffer was used for resuspension of a pellet from 1 L Sf9 culture, and 35 mL lysis buffer per pellet from 1 L *E. coli* culture. Before resuspension of CUL4/ROC1 co-expression pellets, the buffer pH was adjusted to 8.5. 5 μL Benzonase (Merck) was added and the cells lysed by passing the suspension at least twice through a Microfluidiser (Microfluidics). Lysates were clarified by centrifugation at 48000xg for 45 min at 4°C. Protein purification was performed at 4°C on an Äkta pure FPLC (GE) using XK 16/20 chromatography columns (GE) containing 10 mL of the appropriate affinity resin. GST-tagged proteins were captured on glutathione-Sepharose (GSH-Sepharose FF, GE), washed with 250 mL of wash buffer (50 mM Tris-HCl pH 7.8, 500 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP), and eluted with the same buffer

supplemented with 20 mM reduced glutathione. His-tagged proteins were immobilised on Ni-Sepharose HP (GE), washed with 250 mL of wash buffer supplemented with 20 mM imidazole, and eluted with wash buffer containing 0.3 M imidazole. Eluent fractions were analysed by SDS-PAGE, and appropriate fractions were pooled and reduced to 5 mL using centrifugal filter devices (Vivaspin). If applicable, 100 µg GST-3C protease, or 50 µg thrombin, per mg total protein, was added and the sample was incubated for 12 h on ice to cleave off affinity tags. As second purification step, gel filtration chromatography (GF) was performed on an Äkta prime plus FPLC (GE), with Superdex 200 16/600 columns (GE), equilibrated in 10 mM Tris-HCl pH 7.8, 150 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP buffer, at a flow rate of 1 mL/min. For purification of the CUL4/ROC1 complex, the pH of all purification buffers was adjusted to 8.5. Peak fractions were analysed by SDS-PAGE, appropriate fractions were pooled and concentrated to approx. 20 mg/mL, flash-frozen in liquid nitrogen in small aliquots and stored at -80°C. Protein concentrations were determined with a NanoDrop spectrophotometer (ND 1000, Peqlab), using theoretical absorption coefficients calculated based upon the amino acid sequence by ProtParam on the ExPASy webserver[67].

*Analytical gel filtration analysis*

Prior to gel filtration analysis affinity tags were removed by incubation of 30 µg GST-3C protease with 6 µM of each protein component in a volume of 120 µL wash buffer, followed by incubation on ice for 12 h. In order to remove the cleaved GST-tag and GST-3C protease, 20 µL GSH-Sepharose FF beads (GE) were added and the sample was rotated at 4 °C for one hour. GSH-Sepharose beads were removed by centrifugation at 4°C, 3500 rpm for 5 min, and 120 µL of the supernatant was loaded on an analytical GF column (Superdex 200 10/300 GL, GE), equilibrated in 10 mM Tris-HCl pH 7.8, 150 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP, at a flow rate of 0.5 mL/min. 1 mL fractions were collected and analysed by SDS-PAGE.

*In vitro ubiquitylation assays*

160 µL reactions were prepared, containing 0.5 µM substrate (indicated SAMHD1 constructs, Fig. S2), 0.125 µM DDB1/DCAF1-CtD, 0.125 µM CUL4/ROC1, 0.125 µM HisSUMO-T4L-$Vpr_{mus}$ (residues 1-

92), 0.25 µM UBCH5C, 15 µM ubiquitin in 20 mM Tris-HCl pH 7.8, 150 mM NaCl, 2.5 mM $MgCl_2$, 2.5 mM ATP. In control reactions, certain components were left out as indicated in Fig. S2. A 30 µl sample for SDS-PAGE analysis was taken (t=0). Reactions were initiated by addition of 0.05 µM UBA1, incubated at 37°C, and 30 µl SDS-PAGE samples were taken after 1 min, 2 min, 5 min and 15 min, immediately mixed with 10 µl 4x SDS sample buffer and boiled at 95°C for 5 min. Samples were analysed by SDS-PAGE.

*In vitro neddylation of CUL4/ROC1*

For initial neddylation tests, a 200 µL reaction was prepared, containing 8 µM CUL4/ROC1, 1.8 µM UBC12, 30 µM NEDD8 in 50 mM Tris-HCl pH 7.8, 150 mM NaCl, 2.5 mM $MgCl_2$, 2.5 mM ATP. 2x 30 µL samples were taken for SDS-PAGE, one was immediately mixed with 10 µL 4x SDS sample buffer, the other one incubated for 60 min at 25°C. The reaction was initiated by addition of 0.7 µM APPBP1/UBA3, incubated at 25°C, and 30 µL SDS-PAGE samples were taken after 1 min, 5 min, 10 min, 30 min and 60 min, immediately mixed with 10 µL 4x SDS sample buffer and boiled at 95°C for 5 min. Samples were analysed by SDS-PAGE. Based on this test, the reaction was scaled up to 1 mL and incubated for 5 min at 25°C. Reaction was quenched by addition of 5 mM TCEP and immediately loaded onto a Superdex 200 16/600 GF column (GE), equilibrated in 10 mM Tris-HCl pH 7.8, 150 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP at a flow rate of 1 mL/min. Peak fractions were analysed by SDS-PAGE, appropriate fractions were pooled and concentrated to ~20 mg/mL, flash-frozen in liquid nitrogen in small aliquots and stored at -80°C.

*X-ray crystallography sample preparation, crystallisation, data collection and structure solution*

*DDB1/DCAF1-CtD complex* – DDB1/DCAF1-CtD crystals were grown by the hanging drop vapour diffusion method, by mixing equal volumes (1 µL) of DDB1/DCAF1-CtD solution at 10 mg/mL with reservoir solution containing 100 mM Tri-Na citrate pH 5.5, 18% PEG 1000 and suspending over a 500 µl reservoir. Crystals grew over night at 18°C. Crystals were cryo-protected in reservoir solution supplemented with 20% glycerol and cryo-cooled in liquid nitrogen. A data set from a single crystal was collected at Diamond Light Source (Didcot, UK) at a wavelength of 0.92819 Å. Data were processed

using XDS[68] (Table S1), and the structure was solved using molecular replacement with the program MOLREP[69] and available structures of DDB1 (PDB 3e0c) and DCAF1-CtD (PDB 4cc9)[49] as search models. Iterative cycles of model adjustment with the program Coot[70], followed by refinement using the program PHENIX[71] yielded final $R/R_{free}$ factors of 22.0%/27.9% (Table S1). In the model, 94.5 % of residues have backbone dihedral angles in the favoured region of the Ramachandran plot, the remainder fall in the allowed regions, and none are outliers. Details of data collection and refinement statistics are presented in Table S1.

*DDB1/DCAF1-CtD/T4L-Vpr$_{mus}$ (1-92) complex* – The DDB1/DCAF1-CtD/$Vpr_{mus}$ complex was assembled by incubation of purified DDB1/DCAF1-CtD and HisSUMO-T4L-$Vpr_{mus}$ (residues 1-92), at a 1:1 molar ratio, in a buffer containing 50 mM Bis-tris propane pH 8.5, 0.5 M NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP, containing 1 mg of HRV-3C protease for HisSUMO-tag removal. After incubation on ice for 12 h, the sample was loaded onto a Superdex 200 16/600 GF column (GE), with a 1 mL GSH-Sepharose FF column (GE) connected in line. The column was equilibrated with 10 mM Bis-tris propane pH 8.5, 150 mM NaCl, 4 mM MgCl2, and 0.5 mM TCEP. The column flow rate was 1 mL/min. GF fractions were analysed by SDS-PAGE, appropriate fractions were pooled and concentrated to 4.5 mg/mL.

Crystals were prepared by the sitting drop vapour diffusion method, by mixing equal volumes (200 nL) of the protein complex at 4.5 mg/mL and reservoir solution containing 8-10% PEG 4000 (w/v), 200 mM $MgCl_2$, 100 mM HEPES-NaOH, pH 7.0-8.2. The reservoir volume was 75 µL. Crystals grew after at least 4 weeks of incubation at 4°C. Crystals were cryo-protected in reservoir solution supplemented with 20% glycerol and cryo-cooled in liquid nitrogen. Data sets from two single crystals were collected, initially at BESSY II (Helmholtz-Zentrum Berlin, HZB) at a wavelength of 0.91841 Å, and later at ESRF (Grenoble) at a wavelength of 1 Å. Data sets were processed separately using XDS[68] and XDSAPP[72]. The structure was solved by molecular replacement, using the initial BESSY data set, with the program PHASER[73], and the following structures as search models: DDB1/DCAF1-CtD (this work) and T4L variant E11H (PDB 1qt6)[74]. After optimisation of the initial model and refinement against the higher-resolution ESRF data set, $Vpr_{mus}$ was placed manually into the density, using an NMR model of $Vpr_{HIV-1}$ (PDB 1m8l)[75] as guidance. Iterative cycles of model adjustment with the program Coot[70],

followed by refinement using the program PHENIX[71] yielded final $R/R_{free}$ factors of 21.61%/26.05%. In the model, 95.1 % of residues have backbone dihedral angles in the favoured region of the Ramachandran plot, the remainder fall in the allowed regions, and none are outliers. Details of data collection and refinement statistics are presented in Table S1.

*Cryo-EM sample preparation and data collection*

*Complex assembly* – Purified CUL4-NEDD8/ROC1, DDB1/DCAF1-CtD, GST-$Vpr_{mus}$ and rhesus macaque SAMHD1, 1 µM each, were incubated in a final volume of 1 mL of 10 mM Tris-HCl pH 7.8, 150 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP, supplemented with 1 mg of GST-3C protease. After incubation on ice for 12 h, the sample was loaded onto a Superdex 200 16/600 GF column (GE), equilibrated with the same buffer at 1 mL/min, with a 1 mL GSH-Sepharose FF column (GE) connected in line. GF fractions were analysed by SDS-PAGE, appropriate fractions were pooled and concentrated to 2.8 mg/mL.

*Grid preparation* – 3.5 µl protein solution containing 0.05 µM CUL4-NEDD8/ROC1/DDB1/DCAF1-CtD/$Vpr_{mus}$/SAMHD1 complex and 0.25 µM UBCH5C-ubiquitin conjugate (Fig. S4 A, B) were applied to a 300 mesh Quantifoil R2/4 Cu/Rh holey carbon grid (Quantifoil Micro Tools GmbH) coated with an additional thin carbon film as sample support and stained with 2% uranyl acetate for initial characterisation. For cryo-EM, a fresh 400 mesh Quantifoil R1.2/1.3 Cu holey carbon grid (Quantifoil Micro Tools GmbH) was glow-discharged for 30 s using a Harrick plasma cleaner with technical air at 0.3 mbar and 7 W. 3.5 µl protein solution containing 0.4 µM CUL4-NEDD8/ROC1/DDB1/DCAF1-CtD/$Vpr_{mus}$/SAMHD1 complex and 2 µM UBCH5C-ubiquitin conjugate were applied to the grid, incubated for 45 s, blotted with a Vitrobot Mark II device (FEI, Thermo Fisher Scientific) for 1-2 s at 8°C and 80% humidity, and plunged in liquid ethane. Grids were stored in liquid nitrogen until imaging.

*Cryo-EM data collection* – Initial negative stain and cryo-EM datasets were collected automatically for sample quality control and low-resolution reconstructions on a 120 kV Tecnai Spirit cryo-EM (FEI, Thermo Fisher Scientific) equipped with a F416 CMOS camera (TVIPS) using Leginon[76,77]. Particle images were then analysed by 2D classification and initial model reconstruction using SPHIRE[78], cisTEM[79] and Relion 3.07[80]. These data revealed the presence of the complexes containing both

DDB1/DCAF1-CtD/$Vpr_{mus}$ (core) and CUL4/ROC1 (stalk). High-resolution data was collected on a 300 kV Tecnai Polara cryo-EM (FEI, Thermo Fisher Scientific) equipped with a K2summit direct electron detector (Gatan) at a nominal magnification of 31000x, with a pixel size of 0.625 Å/px on the object scale. In total, 3644 movie stacks were collected in super-resolution mode using Leginon[76,77] with the following parameters: defocus range of 0.5-3.0 µm, 40 frames per movie, 10 s exposure time, electron dose of 1.25 e/Å$^2$/s and a cumulative dose of 50 e/Å$^2$ per movie.

*Cryo-EM computational analysis*

Movies were aligned and dose-weighted using MotionCor2[81] and initial estimation of the contrast transfer function (CTF) was performed with the CTFFind4 package[82]. Resulting micrographs were manually inspected to exclude images with substantial contaminants (typically large protein aggregates or ice contaminations) or grid artefacts. Power spectra were manually inspected to exclude images with astigmatic, weak, or poorly defined spectra. After these quality control steps the dataset included 2322 micrographs (63% of total). At this stage, the data set was picked twice and processed separately, to yield reconstructions of the core (analysis 1) and states-1, -2 and -3 (analysis 2).

For **analysis 1**, particle positions were determined using template matching with a filtered map comprising core and stalk using the software Gautomatch (https://www2.mrc-lmb.cam.ac.uk/research/locally-developed-software/zhang-software/). 712,485 particle images were found, extracted with Relion 3.07 and subsequently 2D-classified using cryoSPARC[83], resulting in 505,342 particle images after selection (Fig. S4C, D). These particle images were separated into two equally sized subsets and Tier 1 3D-classification was performed using Relion 3.07 on both of them to reduce computational burden (Fig. S4D). The following parameters were used: initial model="core", number of classes K=4, T=10, global step search=7.5°, number of iterations=25, pixel size 3.75 Å/px. From these, the ones possessing both core and stalk were selected. Classes depicting a similar stalk orientation relative to the core were pooled and directed into Tier 2 as three different subpopulations containing 143,172, 193,059 and 167,666 particle images, respectively (Fig. S4D).

For Tier 2, each subpopulation was classified separately into 4 classes each. From these 12 classes, all particle images exhibiting well-defined densities for core and stalk were pooled and labelled

"core+stalk", resulting in 310,801 particle images in total. 193,096 particle images representing classes containing only the core were pooled and labelled "core" (Fig. S4D)

For Tier 3, the "core" particle subset was separated into 4 classes which yielded uninterpretable reconstructions lacking medium- or high-resolution features. The "core+stalk" subset was separated into 6 classes, with 5 classes containing both stalk and core (Fig. S4D) and one class consisting only of the core with $Vpr_{mus}$ bound. The 5 classes with stalk showed similar stalk orientations as the ones obtained from analysis 2 (see below, Fig. S5), but refined individually to lower resolution as in analysis 2 and were discarded. However, individual refinement of the core-only tier 3 class yielded a 7.3 Å reconstruction (Fig. S4E, F).

For **analysis 2**, particle positions were determined using cisTEMs Gaussian picking routine, yielding 959,155 particle images in total. After two rounds of 2D-classification, 227,529 particle images were selected for further processing (Fig. S4G, H). Using this data, an initial model was created using Relion 3.07. The resulting map yielded strong signal for the core but only fragmented stalk density, indicating a large heterogeneity in the stalk-region within the data set. This large degree of compositional (+/- stalk) and conformational heterogeneity (movement of the stalk relative to the core) made the classification challenging. Accordingly, alignment and classification were carried out simultaneously. The first objective was to separate the data set into three categories: "junk", "core" and "core+stalk". Therefore, the stalk was deleted from the initial model using the "Eraser"-tool in Chimera[84]. This core-map was used as an initial model for the Tier 1 3D-classification with Relion 3.07 at a decimated pixel size of 2.5 Å/px. The following parameters were used: number of classes K=6, T=10, global step search=7.5°, number of iterations= 25. The classification yielded two classes containing the stalk (classes 3 and 5 containing 23% and 22% of the particle images, respectively) (Fig. S4H). These particles were pooled and directed into Tier 2 3D-classification using the following parameters: number of classes K=6, T=10, global step search=7.5°, number of iterations=25. Three of these classes yielded medium-resolution maps with interpretable features (states-1, -2 and -3, Fig. S4H). These three classes were refined individually using 3D Relion 3.07, resulting in maps with resolution ranging from 7.8 Å – 8.9 Å (Fig. S4H-J).

*Molecular visualisation, rigid body fitting, 3D structural alignments, rotation and interface analysis*

Density maps and atomic models were visualised using Coot[70], PyMOL (Schrödinger) and UCSF Chimera[84]. Rigid body fits and structural alignments were performed using the program UCSF Chimera[84]. Rotation angles between extreme DDB1 BPB domain positions were measured using the DynDom server[85] (http://dyndom.cmp.uea.ac.uk/dyndom/runDyndom.jsp). Molecular interfaces were analysed using the EBI PDBePISA server[86] (https://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver).

*Multiple sequence alignment*

A multiple sequence alignment was calculated using the EBI ClustalOmega server[87] (https://www.ebi.ac.uk/Tools/msa/clustalo/), and adjusted manually using the program GeneDoc[88].

*Cross-linking mass spectrometry (CLMS)*

*Complex assembly* – Purified CUL4/ROC1, DDB1/DCAF1-CtD, GST-$Vpr_{mus}$ and rhesus macaque SAMHD1, 1 µM each, were incubated in a volume of 3 mL buffer containing 10 mM HEPES pH 7.8, 150 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP, supplemented with 1 mg GST-3C protease. After incubation on ice for 12 h, the sample was loaded onto a Superdex 200 16/600 GF column (GE), equilibrated with the same buffer, at a flow rate of 1 mL/min with a 1 mL GSH-Sepharose FF column (GE) connected in line. GF fractions were analysed by SDS-PAGE, appropriate fractions were pooled and concentrated to 6 mg/mL.

*Photo-Crosslinking* – The cross-linker sulfo-SDA (sulfosuccinimidyl 4,4′-azipentanoate) (Thermo Scientific) was dissolved in cross-linking buffer (10 mM HEPES pH 7.8, 150 mM NaCl, 4 mM $MgCl_2$, 0.5 mM TCEP) to 100 mM before use. The labelling step was performed on 18 µg aliquots of the complex at 1 mg/mL with 2, 1, 0.5, 0.25, 0.125 mM sulfo-SDA, added, respectively. The samples were irradiated with UV light at 365 nm, to form cross-links, for 20 min and quenched with 50 mM $NH_4HCO_3$ for 20 min. All steps were performed on ice. Reaction products were separated on a Novex Bis-Tris 4–12% SDS−PAGE gel (Life Technologies). The gel band corresponding to the cross-linked complex was excised and digested with trypsin (Thermo Scientific Pierce)[89] and the resulting tryptic peptides were extracted and desalted using C18 StageTips[90]. Eluted peptides

were fractionated on a Superdex Peptide 3.2/300 increase column (GE Healthcare) at a flow rate of 10 µL/min using 30% (v/v) acetonitrile and 0.1 % (v/v) trifluoroacetic acid as mobile phase. 50 µL fractions were collected and lyophilised to dryness.

*CLMS acquisition* – Samples for analysis were resuspended in 0.1% (v/v) formic acid, 3.2% (v/v) acetonitrile. LC-MS/MS analysis was performed on an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher) coupled on-line with an Ultimate 3000 RSLCnano HPLC system (Dionex, Thermo Fisher). Samples were separated on a 50 cm EASY-Spray column (Thermo Fisher). Mobile phase A consisted of 0.1% (v/v) formic acid and mobile phase B of 80% (v/v) acetonitrile with 0.1% (v/v) formic acid. Flow rates were 0.3 µL/min using gradients optimized for each chromatographic fraction from offline fractionation, ranging from 2% mobile phase B to 55% mobile phase B over 90 min. MS data were acquired in data-dependent mode using the top-speed setting with a 3 s cycle time. For every cycle, the full scan mass spectrum was recorded using the Orbitrap at a resolution of 120,000 in the range of 400 to 1,500 m/z. Ions with a precursor charge state between 3+ and 7+ were isolated and fragmented. Fragmentation by Higher-Energy Collisional Dissociation (HCD) and fragmentation spectra were then recorded in the Orbitrap with a resolution of 50,000. Dynamic exclusion was enabled with single repeat count and 60 s exclusion duration.

*CLMS processing* – A recalibration of the precursor m/z was conducted based on high-confidence (<1% false discovery rate (FDR)) linear peptide identifications. The re-calibrated peak lists were searched against the sequences and the reversed sequences (as decoys) of cross-linked peptides using the Xi software suite (v.1.7.5.1) for identification[91]. Final crosslink lists were compiled using the identified candidates filtered to <1% FDR on link level with xiFDR v.2.0[92] imposing a minimum of 20% sequence coverage and 4 observed fragments per peptide.

*CLMS analysis* – In order to sample the accessible interaction volume of the SAMHD1-CtD consistent with CLMS data, a model for SAMHD1 was generated using I-TASSER[93]. The SAMHD1-CtD, which adopted a random coil configuration, was extracted from the model. In order to map all crosslinks, missing loops in the complex structure were generated using MODELLER[94]. An interaction volume search was then submitted to the DisVis webserver[95] with an allowed distance between 1.5 Å and 22 Å for each restraint using the "complete scanning" option. The rotational sampling interval was set to 9.72°

and the grid voxel spacing to 1Å. The accessible interaction volume was visualised using UCSF Chimera[84].

**Data availability**

The coordinates and structure factors for the crystal structures have been deposited at the Protein Data Bank (PDB) with the accession codes 6ZUE (DDB1/DCAF1-CtD) and 6ZX9 (DDB1/DCAF1-CtD/T4L-$Vpr_{mus}$ 1-92). Cryo-EM reconstructions have been deposited at the Electron Microscopy Data Bank (EMDB) with the accession codes EMD-10611 (core), EMD-10612 (conformational state-1), EMD-10613 (state-2) and EMD-10614 (state-3). CLMS data have been deposited at the PRIDE database[96] with the accession code PXD020453.

**Acknowledgments**

This work was funded by the DFG Emmy Noether Programme (D.S., SCHW1851/1-1), by an EMBO Advanced laboratory start-up grant (D.S., aALTF-1650) and by the Wellcome trust (I.A.T., 108014/Z/15/Z). We thank the MPI-MG for granting access to the TEM instruments of the microscopy and cryo-EM service group. We thank Manfred Weiss and the scientific staff of the BESSY-MX (Macromolecular X-ray Crystallography)/Helmholtz Zentrum Berlin für Materialien und Energie at beamlines BL14.1, BL14.2, and BL14.3 operated by the Joint Berlin MX-Laboratory at the BESSY II electron storage ring (Berlin-Adlershof, Germany) as well as the scientific staff of the ESRF (Grenoble, France, grant no. MX-651) at beamlines ID30A-3, ID30B, ID23-1, ID23-2, and ID29 for continuous support. We acknowledge Diamond Light Source, Didcot, UK, for access and support of the synchrotron beamline I04 (grant no. MX13775) and cryo-EM facilities at the UK's national Electron Bio-imaging Centre (eBIC) (EM19715-1 and EM19715-2), funded by the Wellcome Trust, MRC and BBRSC. Furthermore, the authors acknowledge the North-German Supercomputing Alliance (HLRN) and the HPC for Research cluster of the Berlin Institute of Health for providing HPC resources. This work has additionally been supported by iNEXT (proposal no. 3825), project number 653706, funded by the Horizon 2020 programme of the European Union. The pHisSUMO plasmid was a generous gift from Dr. Evangelos Christodoulou (The Francis Crick Institute, UK). The rhesus macaque SAMHD1 cDNA template was a generous gift from Prof. Michael Emerman (Fred Hutchinson Cancer Research Center, Seattle, USA). Recombinant BAC10:1629KO bacmid was a generous gift from Prof. Ian Jones (University of Reading, UK). pAcGHLT-B-DDB1 was a gift from Ning Zheng (Addgene plasmid

48638). pET28-mE1 was a gift from Jorge Eduardo Azevedo (Addgene plasmid 32534). The FP7 WeNMR (project 261572), H2020 West-Life (project# 675858), BioExcel (project 675728) and the EOSC-hub (project 777536) European e-Infrastructure projects are acknowledged for the use of their web portals, which make use of the EGI infrastructure with the dedicated support of CESNET-MCC, INFN-PADOVA, NCG-INGRID-PT, TW-NCHC, SURFsara and NIKHEF, and the additional support of the national GRID Initiatives of Belgium, France, Italy, Germany, the Netherlands, Poland, Portugal, Spain, UK, Taiwan and the US Open Science Grid.

**Author contributions**

S.B., F.K., C.G., J.B., A.G., F.O'R., L.S., J.R., C.M.T.S., T.M., I.A.T. and D.S. planned experiments; S.B., F.K., C.G., J.B., A.G., F.O'R., L.S., O.R. and D.S. performed experiments; S.B., F.K., A.G., F.O'R., L.S., J.R. and D.S. analysed data; D.S. wrote the manuscript, which was reviewed by all co-authors.

**Conflict of interest**

The other authors declare no competing interests.

## References


1. Randow, F. & Lehner, P.J. Viral avoidance and exploitation of the ubiquitin system. *Nat Cell Biol* **11**, 527-34 (2009).
2. Isaacson, M.K. & Ploegh, H.L. Ubiquitination, ubiquitin-like modifiers, and deubiquitination in viral infection. *Cell Host Microbe* **5**, 559-70 (2009).
3. Gustin, J.K., Moses, A.V., Fruh, K. & Douglas, J.L. Viral takeover of the host ubiquitin system. *Front Microbiol* **2**, 161 (2011).
4. Barry, M. & Fruh, K. Viral modulators of cullin RING ubiquitin ligases: culling the host defense. *Sci STKE* **2006**, pe21 (2006).
5. Mahon, C., Krogan, N.J., Craik, C.S. & Pick, E. Cullin E3 ligases and their rewiring by viral factors. *Biomolecules* **4**, 897-930 (2014).
6. Becker, T., Le-Trilling, V.T.K. & Trilling, M. Cellular Cullin RING Ubiquitin Ligases: Druggable Host Dependency Factors of Cytomegaloviruses. *Int J Mol Sci* **20**(2019).
7. Seissler, T., Marquet, R. & Paillart, J.C. Hijacking of the Ubiquitin/Proteasome Pathway by the HIV Auxiliary Proteins. *Viruses* **9**(2017).
8. Zheng, N. & Shabek, N. Ubiquitin Ligases: Structure, Function, and Regulation. *Annu. Rev. Biochem.* **86**, 14.1-29 (2017).
9. Sauter, D. & Kirchhoff, F. Key Viral Adaptations Preceding the AIDS Pandemic. *Cell Host Microbe* **25**, 27-38 (2019).
10. Sharp, P.M. & Hahn, B.H. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med* **1**, a006841 (2011).
11. Hatziioannou, T. et al. HIV-1-induced AIDS in monkeys. *Science* **344**, 1401-5 (2014).
12. Fischer, E.S. et al. The molecular basis of CRL4DDB2/CSA ubiquitin ligase architecture, targeting, and activation. *Cell* **147**, 1024-39 (2011).
13. Lee, J. & Zhou, P. DCAFs, the missing link of the CUL4-DDB1 ubiquitin ligase. *Mol Cell* **26**, 775-80 (2007).
14. Angers, S. et al. Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature* **443**, 590-3 (2006).
15. Scrima, A. et al. Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell* **135**, 1213-23 (2008).
16. Zimmerman, E.S., Schulman, B.A. & Zheng, N. Structural assembly of cullin-RING ubiquitin ligase complexes. *Curr Opin Struct Biol* **20**, 714-21 (2010).
17. Andrejeva, J., Young, D.F., Goodbourn, S. & Randall, R.E. Degradation of STAT1 and STAT2 by the V proteins of simian virus 5 and human parainfluenza virus type 2, respectively: consequences for virus replication in the presence of alpha/beta and gamma interferons. *J Virol* **76**, 2159-67 (2002).
18. Li, T., Chen, X., Garbutt, K.C., Zhou, P. & Zheng, N. Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase. *Cell* **124**, 105-17 (2006).
19. Trilling, M. et al. Identification of DNA-damage DNA-binding protein 1 as a conditional essential factor for cytomegalovirus replication in interferon-gamma-stimulated cells. *PLoS Pathog* **7**, e1002069 (2011).
20. Paradkar, P.N., Duchemin, J.B., Rodriguez-Andres, J., Trinidad, L. & Walker, P.J. Cullin4 Is Pro-Viral during West Nile Virus Infection of Culex Mosquitoes. *PLoS Pathog* **11**, e1005143 (2015).
21. Decorsiere, A. et al. Hepatitis B virus X protein identifies the Smc5/6 complex as a host restriction factor. *Nature* **531**, 386-9 (2016).
22. Murphy, C.M. et al. Hepatitis B Virus X Protein Promotes Degradation of SMC5/6 to Enhance HBV Replication. *Cell Rep* **16**, 2846-2854 (2016).
23. Lim, E.S. et al. The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. *Cell Host Microbe* **11**, 194-204 (2012).

24. Romani, B. & Cohen, E.A. Lentivirus Vpr and Vpx accessory proteins usurp the cullin4-DDB1 (DCAF1) E3 ubiquitin ligase. *Curr Opin Virol* **2**, 755-63 (2012).
25. Fabryova, H. & Strebel, K. Vpr and Its Cellular Interaction Partners: R We There Yet? *Cells* **8**(2019).
26. Greenwood, E.J.D. et al. Promiscuous Targeting of Cellular Proteins by Vpr Drives Systems-Level Proteomic Remodeling in HIV-1 Infection. *Cell Rep* **27**, 1579-1596 e7 (2019).
27. Schrofelbauer, B., Yu, Q., Zeitlin, S.G. & Landau, N.R. Human immunodeficiency virus type 1 Vpr induces the degradation of the UNG and SMUG uracil-DNA glycosylases. *J Virol* **79**, 10978-87 (2005).
28. Lahouassa, H. et al. HIV-1 Vpr degrades the HLTF DNA translocase in T cells and macrophages. *Proc Natl Acad Sci U S A* **113**, 5311-6 (2016).
29. Laguette, N. et al. Premature activation of the SLX4 complex by Vpr promotes G2/M arrest and escape from innate immune sensing. *Cell* **156**, 134-45 (2014).
30. Zhou, X., DeLucia, M. & Ahn, J. SLX4-SLX1 Protein-independent Down-regulation of MUS81-EME1 Protein by HIV-1 Viral Protein R (Vpr). *J Biol Chem* **291**, 16936-47 (2016).
31. Romani, B., Shaykh Baygloo, N., Aghasadeghi, M.R. & Allahbakhshi, E. HIV-1 Vpr Protein Enhances Proteasomal Degradation of MCM10 DNA Replication Factor through the Cul4-DDB1[VprBP] E3 Ubiquitin Ligase to Induce G2/M Cell Cycle Arrest. *J Biol Chem* **290**, 17380-9 (2015).
32. Lv, L. et al. Vpr Targets TET2 for Degradation by CRL4(VprBP) E3 Ligase to Sustain IL-6 Expression and Enhance HIV-1 Replication. *Mol Cell* **70**, 961-970 e5 (2018).
33. Su, J. et al. HIV-2/SIV Vpx targets a novel functional domain of STING to selectively inhibit cGAS-STING-mediated NF-kappaB signalling. *Nat Microbiol* **4**, 2552-2564 (2019).
34. Chougui, G. et al. HIV-2/SIV viral protein X counteracts HUSH repressor complex. *Nat Microbiol* **3**, 891-897 (2018).
35. Yurkovetskiy, L. et al. Primate immunodeficiency virus proteins Vpx and Vpr counteract transcriptional repression of proviruses by the HUSH complex. *Nat Microbiol* **3**, 1354-1361 (2018).
36. Hrecka, K. et al. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* **474**, 658-61 (2011).
37. Laguette, N. et al. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* **474**, 654-7 (2011).
38. Powell, R.D., Holland, P.J., Hollis, T. & Perrino, F.W. Aicardi-Goutieres syndrome gene and HIV-1 restriction factor SAMHD1 is a dGTP-regulated deoxynucleotide triphosphohydrolase. *J Biol Chem* **286**, 43596-600 (2011).
39. Goldstone, D.C. et al. HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature* **480**, 379-82 (2011).
40. Zhu, C. et al. Structural insight into dGTP-dependent activation of tetrameric SAMHD1 deoxynucleoside triphosphate triphosphohydrolase. *Nat Commun* **4**, 2722 (2013).
41. Kim, B., Nguyen, L.A., Daddacha, W. & Hollenbaugh, J.A. Tight interplay among SAMHD1 protein level, cellular dNTP levels, and HIV-1 proviral DNA synthesis kinetics in human primary monocyte-derived macrophages. *J Biol Chem* **287**, 21570-4 (2012).
42. Lahouassa, H. et al. SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat Immunol* **13**, 223-8 (2012).
43. St Gelais, C. et al. SAMHD1 restricts HIV-1 infection in dendritic cells (DCs) by dNTP depletion, but its expression in DCs and primary CD4+ T-lymphocytes cannot be upregulated by interferons. *Retrovirology* **9**, 105 (2012).
44. Rehwinkel, J. et al. SAMHD1-dependent retroviral control and escape in mice. *EMBO J* **32**, 2454-62 (2013).
45. Morris, E.R. & Taylor, I.A. The missing link: allostery and catalysis in the anti-viral protein SAMHD1. *Biochem Soc Trans* **47**, 1013-1027 (2019).

46. Baldauf, H.M. et al. SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. *Nat Med* **18**, 1682-7 (2012).
47. Shingai, M. et al. The Expression of Functional Vpx during Pathogenic SIVmac Infections of Rhesus Macaques Suppresses SAMHD1 in CD4+ Memory T Cells. *PLoS Pathog* **11**, e1004928 (2015).
48. Fregoso, O.I. et al. Evolutionary toggling of Vpx/Vpr specificity results in divergent recognition of the restriction factor SAMHD1. *PLoS Pathog* **9**, e1003496 (2013).
49. Schwefel, D. et al. Structural basis of lentiviral subversion of a cellular protein degradation pathway. *Nature* **505**, 234-8 (2014).
50. Schwefel, D. et al. Molecular Determinants for Recognition of Divergent SAMHD1 Proteins by the Lentiviral Accessory Protein Vpx. *Cell Host Microbe* **17**, 489-99 (2015).
51. Wu, Y. et al. Structural Basis of Clade-specific Engagement of SAMHD1 (Sterile alpha Motif and Histidine/Aspartate-containing Protein 1) Restriction Factors by Lentiviral Viral Protein X (Vpx) Virulence Factors. *J Biol Chem* **290**, 17935-45 (2015).
52. Spragg, C.J. & Emerman, M. Antagonism of SAMHD1 is actively maintained in natural infections of simian immunodeficiency virus. *Proc Natl Acad Sci U S A* **110**, 21136-41 (2013).
53. Wu, Y. et al. The DDB1-DCAF1-Vpr-UNG2 crystal structure reveals how HIV-1 Vpr steers human UNG2 toward destruction. *Nat Struct Mol Biol* **23**, 933-940 (2016).
54. Enchev, R.I., Schulman, B.A. & Peter, M. Protein neddylation: beyond cullin-RING ligases. *Nat Rev Mol Cell Biol* **16**, 30-44 (2015).
55. Schneider, M., Belsom, A. & Rappsilber, J. Protein Tertiary Structure by Crosslinking/Mass Spectrometry. *Trends Biochem Sci* **43**, 157-169 (2018).
56. Fischer, E.S. et al. Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* **512**, 49-53 (2014).
57. DeLucia, M., Mehrens, J., Wu, Y. & Ahn, J. HIV-2 and SIVmac accessory virulence factor Vpx down-regulates SAMHD1 enzyme catalysis prior to proteasome-dependent degradation. *J Biol Chem* **288**, 19116-26 (2013).
58. Berger, G., Lawrence, M., Hue, S. & Neil, S.J. G2/M cell cycle arrest correlates with primate lentiviral Vpr interaction with the SLX4 complex. *J Virol* (2014).
59. Guenzel, C.A., Herate, C. & Benichou, S. HIV-1 Vpr-a still "enigmatic multitasker". *Front Microbiol* **5**, 127 (2014).
60. Stivahtis, G.L., Soares, M.A., Vodicka, M.A., Hahn, B.H. & Emerman, M. Conservation and host specificity of Vpr-mediated cell cycle arrest suggest a fundamental role in primate lentivirus evolution and biology. *J Virol* **71**, 4331-8 (1997).
61. Planelles, V. et al. Vpr-induced cell cycle arrest is conserved among primate lentiviruses. *J Virol* **70**, 2516-24 (1996).
62. Schapira, M., Calabrese, M.F., Bullock, A.N. & Crews, C.M. Targeted protein degradation: expanding the toolbox. *Nat Rev Drug Discov* **18**, 949-963 (2019).
63. Hanzl, A. & Winter, G.E. Targeted protein degradation: current and future challenges. *Curr Opin Chem Biol* **56**, 35-41 (2020).
64. Duda, D.M. et al. Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. *Cell* **134**, 995-1006 (2008).
65. Baek, K. et al. NEDD8 nucleates a multivalent cullin-RING-UBE2D ubiquitin ligation assembly. *Nature* **578**, 461-466 (2020).
66. Zhao, Y., Chapman, D.A. & Jones, I.M. Improving baculovirus recombination. *Nucleic Acids Res* **31**, E6-6 (2003).
67. Wilkins, M.R. et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **112**, 531-52 (1999).
68. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32 (2010).
69. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66**, 22-5 (2010).
70. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).

71. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).
72. Sparta, K.M., Krug, M., Heinemann, U., Mueller, U. & Weiss, M.S. XDSAPP2.0. *Journal of Applied Crystallography* **49**, 1085-1092 (2016).
73. McCoy, A.J. et al. Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674 (2007).
74. Kuroki, R., Weaver, L.H. & Matthews, B.W. Structural basis of the conversion of T4 lysozyme into a transglycosidase by reengineering the active site. *Proc Natl Acad Sci U S A* **96**, 8949-54 (1999).
75. Morellet, N., Bouaziz, S., Petitjean, P. & Roques, B.P. NMR structure of the HIV-1 regulatory protein VPR. *J Mol Biol* **327**, 215-27 (2003).
76. Carragher, B. et al. Leginon: an automated system for acquisition of images from vitreous ice specimens. *J Struct Biol* **132**, 33-45 (2000).
77. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J Struct Biol* **151**, 41-60 (2005).
78. Moriya, T. et al. High-resolution Single Particle Analysis from Electron Cryo-microscopy Images Using SPHIRE. *J Vis Exp* (2017).
79. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *Elife* **7**(2018).
80. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**(2018).
81. Zheng, S.Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).
82. Mindell, J.A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* **142**, 334-47 (2003).
83. Punjani, A., Rubinstein, J.L., Fleet, D.J. & Brubaker, M.A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017).
84. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).
85. Hayward, S. & Lee, R.A. Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J Mol Graph Model* **21**, 181-3 (2002).
86. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-97 (2007).
87. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**, W636-W641 (2019).
88. Nicholas, K.B., Nicholas Jr., H. B., Deerfield II., D. W. GeneDoc: Analysis and Visualization of Genetic Variation. *embnet.news* **4**, 1-4 (1997).
89. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**, 2856-60 (2006).
90. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**, 663-70 (2003).
91. Mendes, M.L. et al. An integrated workflow for crosslinking mass spectrometry. *Mol Syst Biol* **15**, e8994 (2019).
92. Fischer, L. & Rappsilber, J. Quirks of Error Estimation in Cross-Linking/Mass Spectrometry. *Anal Chem* **89**, 3829-3833 (2017).
93. Yang, J. & Zhang, Y. Protein Structure and Function Prediction Using I-TASSER. *Curr Protoc Bioinformatics* **52**, 5 8 1-5 8 15 (2015).
94. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* **54**, 5 6 1-5 6 37 (2016).
95. van Zundert, G.C. et al. The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **429**, 399-407 (2017).

96. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442-D450 (2019).

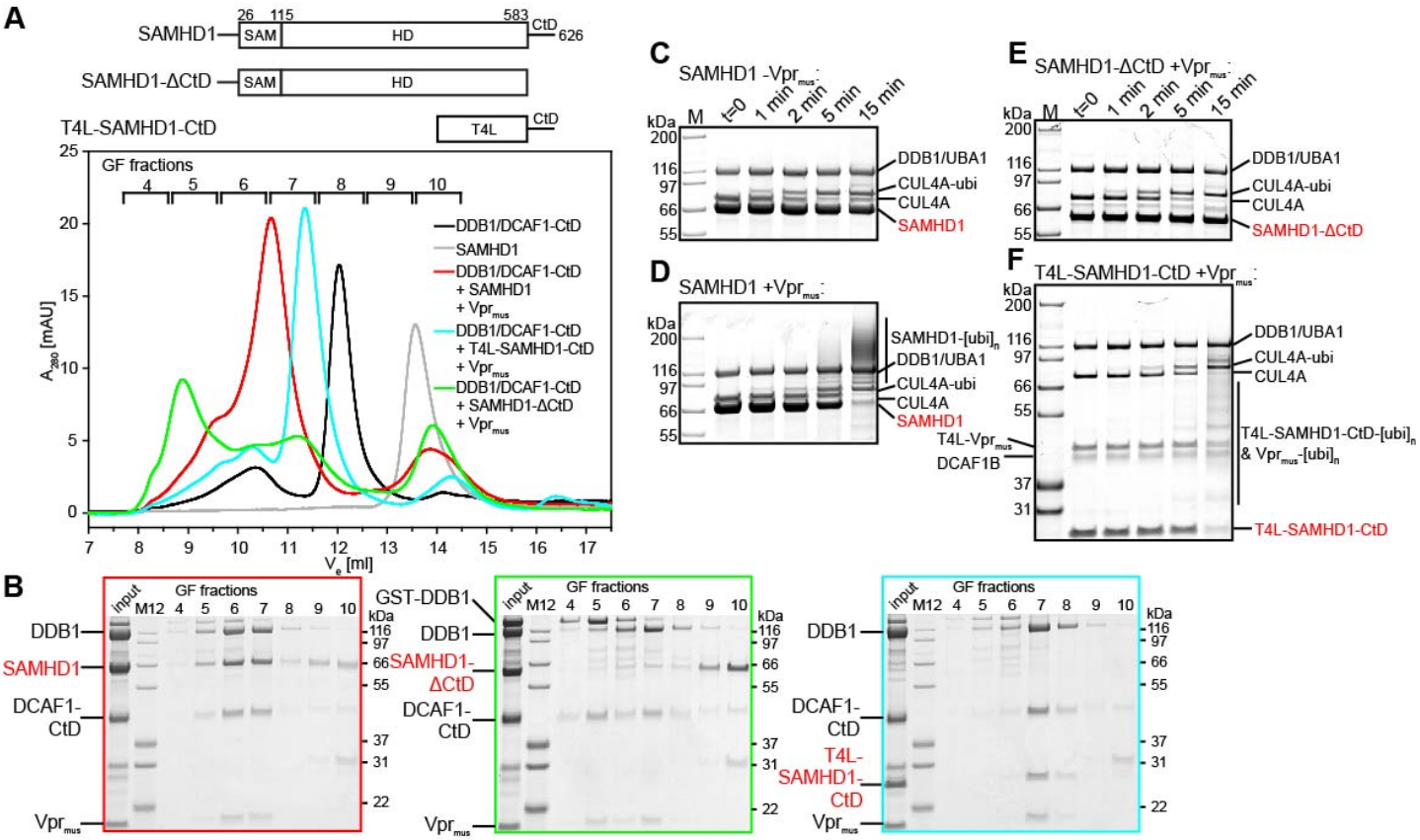



**Fig. 1. Biochemical analysis of $Vpr_{mus}$-induced $CRL4^{DCAF1}$ specificity redirection.** (**A**) GF analysis of *in vitro* reconstitution of protein complexes containing DDB1/DCAF1-CtD, $Vpr_{mus}$ and SAMHD1 constructs. A schematic of the SAMHD1 constructs is shown above the chromatograms. SAM – sterile α-motif domain, HD – histidine-aspartate domain, T4L – T4 Lysozyme. (**B**) SDS-PAGE analysis of fractions collected during GF runs in **A**, boxes are colour-coded with respect to the chromatograms. Note that during preparation of the GF run containing SAMHD1-ΔCtD (green trace), the GST-affinity tag, which forms dimers in solution, was not removed completely from DDB1. Accordingly, the GF trace contains an additional dimeric GST-DDB1/DCAF1-CtD/$Vpr_{mus}$ component in fractions 4-5. (**C-F**) *In vitro* ubiquitylation reactions with purified protein components in the absence (**C**) or presence (**D-F**) of $Vpr_{mus}$, with the indicated SAMHD1 constructs as substrate. Reactions were stopped after the indicated times, separated on SDS-PAGE and visualised by staining.

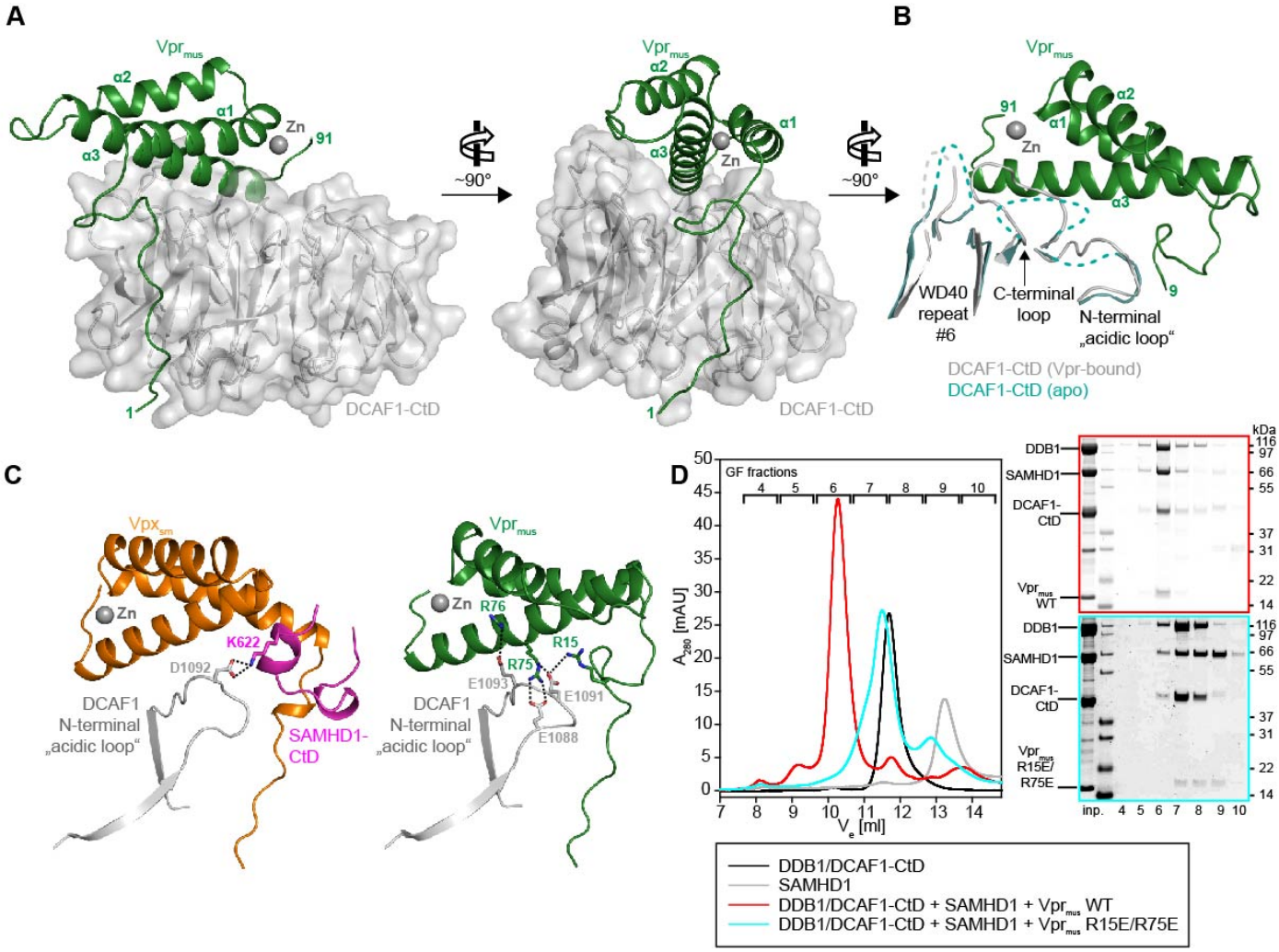



**Fig. 2. Crystal structure of the DDB1/DCAF1-CtD/Vpr$_{mus}$ complex.** (**A**) Overall structure of the complex in two views. DCAF1-CtD is shown as grey cartoon and semi-transparent surface. Vpr$_{mus}$ is shown as a dark green cartoon with the co-ordinated zinc ion shown as grey sphere. T4L and DDB1 have been omitted for clarity. (**B**) Superposition of apo-DCAF1-CtD (light blue cartoon) with Vpr$_{mus}$-bound DCAF1-CtD (grey/green cartoon). Only DCAF1-CtD regions with significant structural differences between apo- and Vpr$_{mus}$-bound forms are shown. Disordered loops are indicated as dashed lines. (**C**) Comparison of the binary Vpr$_{mus}$/DCAF1-CtD and ternary Vpx$_{sm}$/DCAF1-CtD/SAMHD1-CtD complexes. For DCAF1-CtD, only the N-terminal "acidic loop" region is shown. Vpr$_{mus}$, DCAF1-CtD and bound zinc are coloured as in **A**; Vpx$_{sm}$ is represented as orange cartoon and SAMHD1-CtD as pink cartoon. Selected Vpr/Vpx/DCAF1-CtD side chains are shown as sticks, and electrostatic interactions between these side chains are indicated as dotted lines. (**D**) *In vitro* reconstitution of protein complexes containing DDB1/DCAF1-CtD/Vpr$_{mus}$ or the Vpr$_{mus}$ R15E/R75E mutant, and SAMHD1,

analysed by analytical GF. SDS-PAGE analysis of corresponding GF fractions is shown next to the chromatogram.

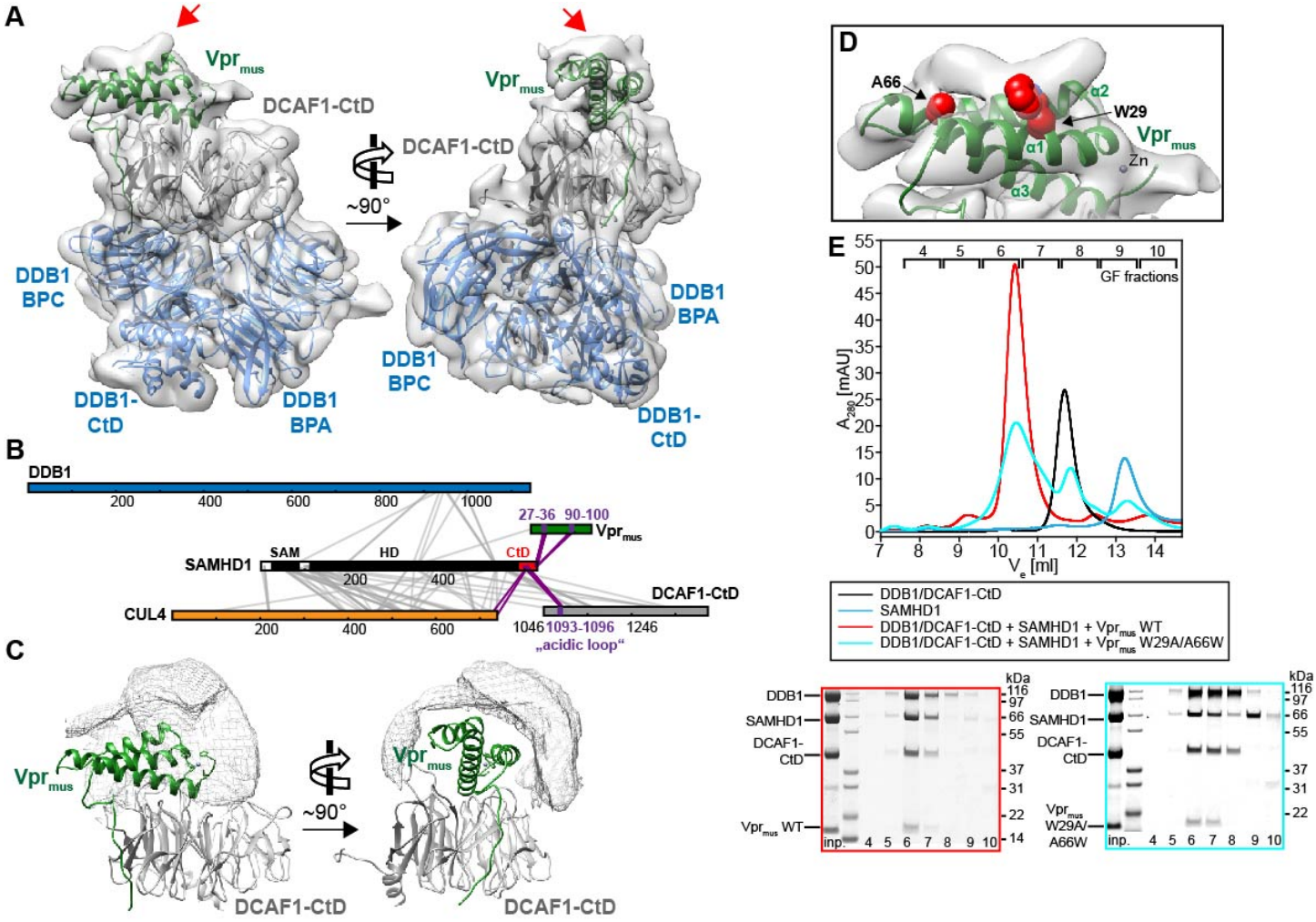



**Fig. 3. Mechanism of SAMHD1-CtD recruitment by Vpr$_{mus}$.** (**A**) Two views of the cryo-EM reconstruction of the CRL4-NEDD8$^{DCAF1\text{-}CtD}$/Vpr$_{mus}$/SAMHD1 core. The crystal structure of the DDB1/DCAF1-CtD/Vpr$_{mus}$ complex was fitted as a rigid body into the cryo-EM density and is shown in the same colours as in Fig. 2A. The DDB1 BPB model and density was removed for clarity. The red arrows mark additional density on the upper surface of the Vpr$_{mus}$ helix bundle. (**B**) Schematic representation of Sulfo-SDA cross-links (grey lines) between CRL4$^{DCAF1}$/Vpr$_{mus}$ and SAMHD1, identified by CLMS. Proteins are colour-coded as in **A**, CUL4 is coloured orange, SAMHD1 black/white. SAMHD1-CtD is highlighted in red, and cross-links to SAMHD1-CtD are highlighted in violet. (**C**) The accessible interaction space of SAMHD1-CtD, calculated by the DisVis server[95], consistent with at least 14 of 26 observed cross-links, is visualised as grey mesh. DCAF1-CtD and Vpr$_{mus}$ are oriented and coloured as in **A**. (**D**) Detailed view of the SAMHD1-CtD electron density. The model is in the same orientation as in **A**, left panel. Selected Vpr$_{mus}$ residues W29 and A66, which are in close contact to the additional density, are shown as red space-fill representation. (**E**) *In vitro* reconstitution

of protein complexes containing DDB1/DCAF1-CtD, $Vpr_{mus}$ or the $Vpr_{mus}$ W29A/A66W mutant, and SAMHD1, assessed by analytical GF. SDS-PAGE analysis of corresponding GF fractions is shown below the chromatogram.

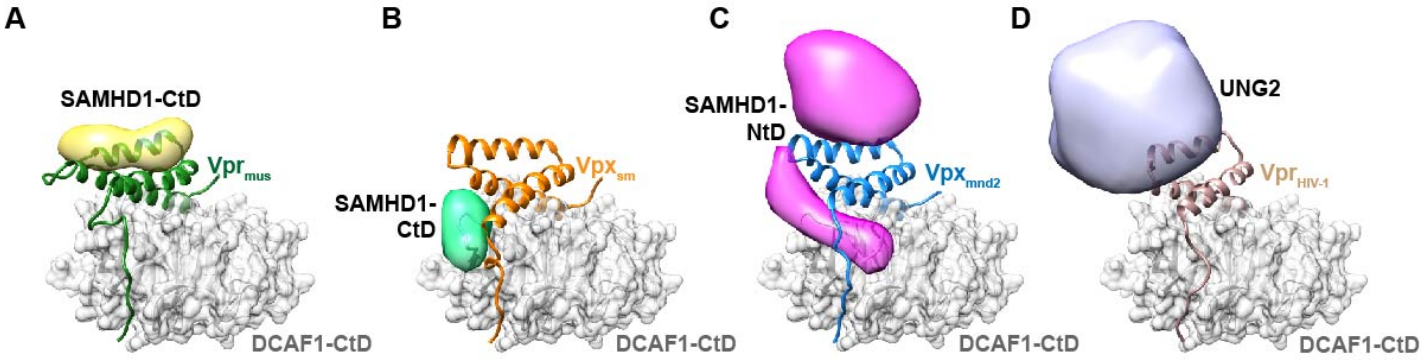



**Fig. 4. Variability of *neo*-substrate recognition in Vpx/Vpr proteins.** Comparison of *neo*-substrate recognition modes of $Vpr_{mus}$ (**A**), $Vpx_{sm}$ (**B**), $Vpx_{mnd2}$ (**C**) and $Vpr_{HIV-1}$ (**D**) proteins. DCAF1-CtD is shown as grey cartoon and semi-transparent surface, $Vpr_{mus}$ – green, $Vpx_{sm}$ – orange, $Vpx_{mnd2}$ – blue and $Vpr_{HIV-1}$– light brown are shown as cartoon. Models of the recruited ubiquitylation substrates are shown as strongly filtered, semi-transparent calculated electron density maps with the following colouring scheme: SAMHD1-CtD bound to $Vpr_{mus}$ – yellow, SAMHD1-CtD (bound to $Vpx_{sm}$, PDB 4cc9)[49] – mint green, SAMHD1-NtD ($Vpx_{mnd2}$, PDB 5aja)[50] – magenta, UNG2 ($Vpr_{HIV-1}$, PDB 5jk7)[53] – light violet.

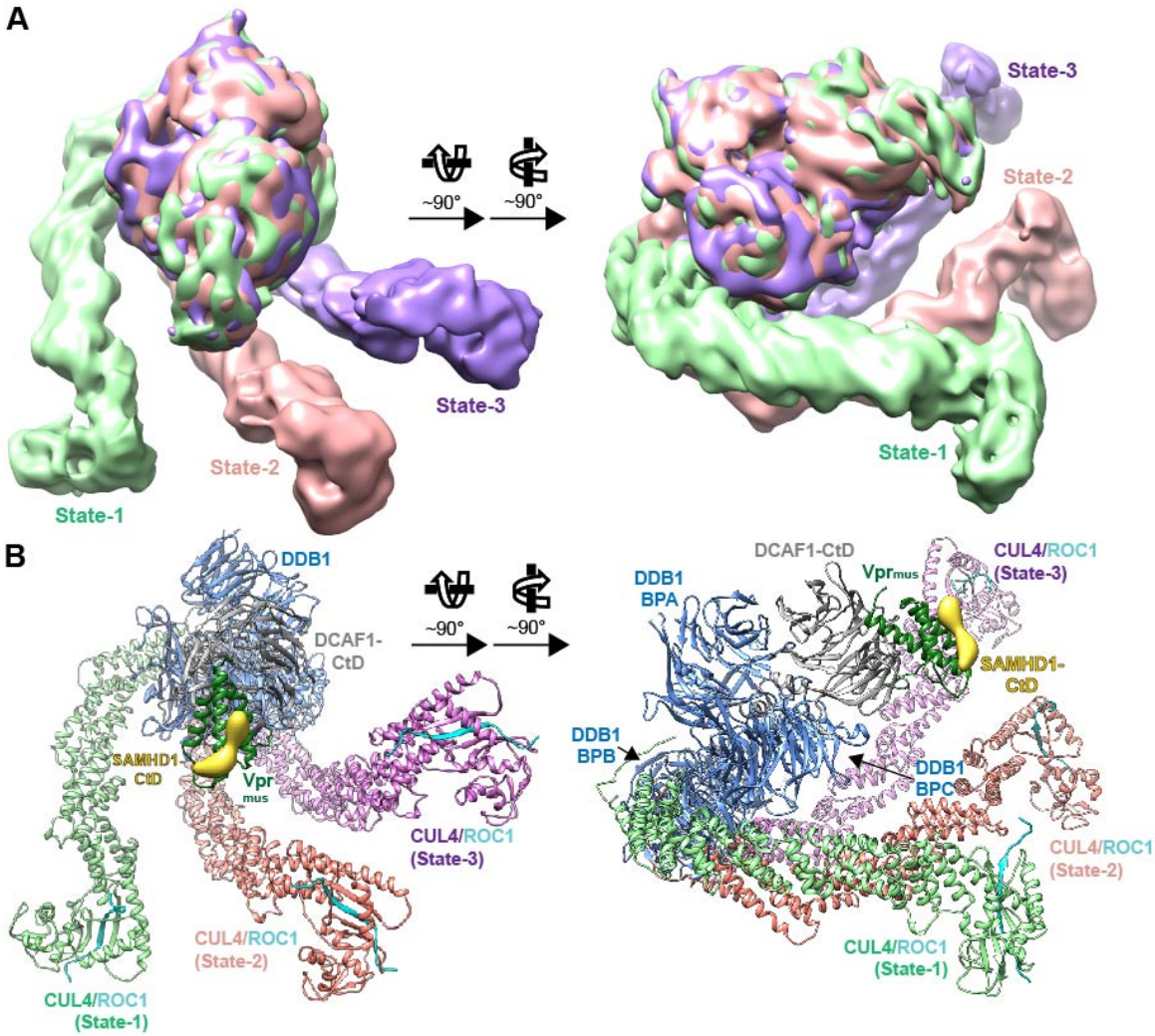



**Fig. 5. Cryo-EM analysis of CRL4-NEDD8$^{DCAF1\text{-}CtD}$ conformational states.** (**A**) Two views of an overlay of CRL4-NEDD8$^{DCAF1\text{-}CtD}$/Vpr$_{mus}$/SAMHD1 cryo-EM reconstructions (conformational state-1 – light green, state-2 – salmon, state-3 – purple). The portions of the densities corresponding to DDB1 BPA/BPC, DCAF1-CtD and Vpr$_{mus}$ have been superimposed. (**B**) Two views of a superposition of DDB1/DCAF1-CtD/Vpr$_{mus}$ and CUL4/ROC1 (PDB 2hye)[14] molecular models, which have been fitted as rigid bodies to the corresponding cryo-EM densities; the models are oriented as in **A**. DDB1/DCAF1-CtD/Vpr$_{mus}$ is shown as in Fig. 2A, CUL4 is shown as cartoon, coloured as in **A** and ROC1 is shown as cyan cartoon. Cryo-EM density corresponding to SAMHD1-CtD is shown in yellow, to illustrate the SAMHD1-CtD binding site in the context of the whole CRL4 assembly.

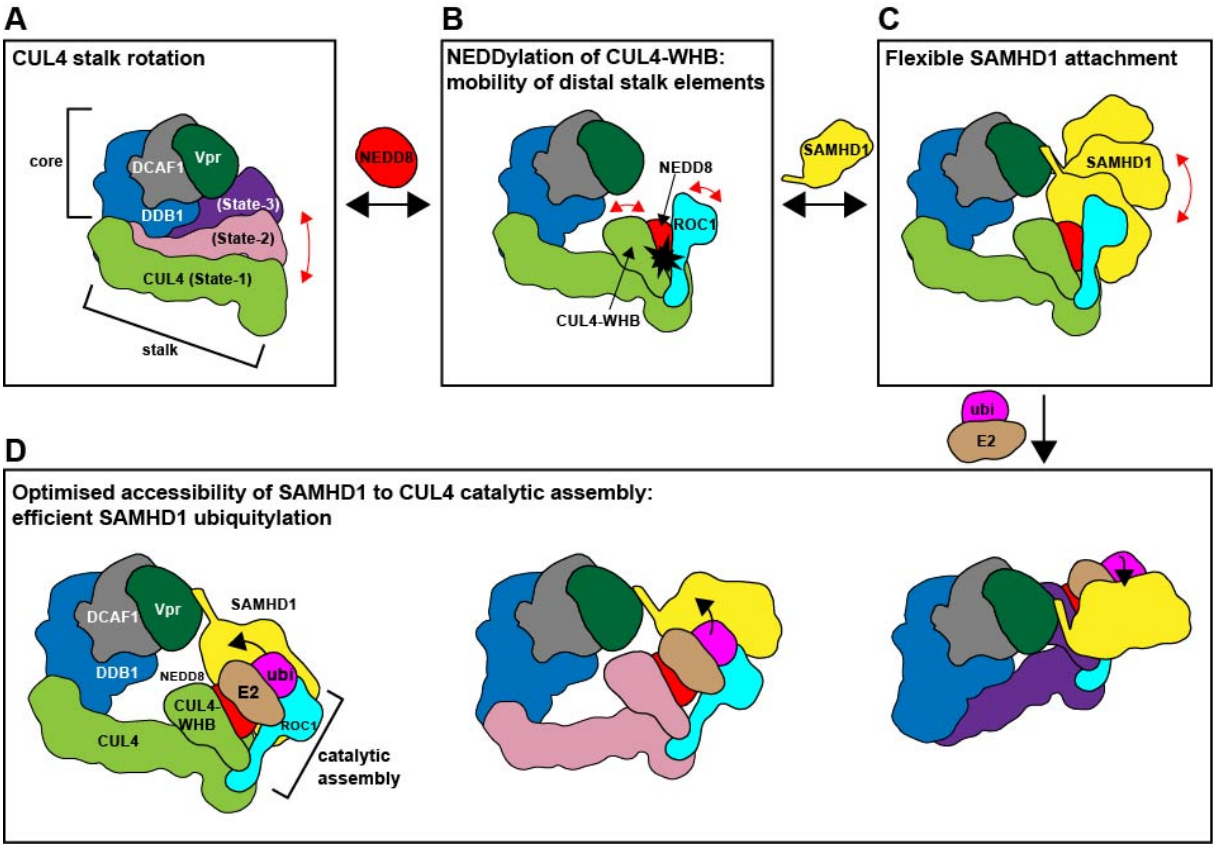



**Fig. 6. Schematic illustration of structural plasticity in $Vpr_{mus}$-modified $CRL4^{DCAF1\text{-}CtD}$, and implications for ubiquitin transfer**. (**A**) Rotation of the CRL4 stalk increases the space accessible to catalytic elements at the distal tip of the stalk, forming a ubiquitylation zone around the core. (**B**) Modification of CUL4-WHB with NEDD8 leads to increased mobility of these distal stalk elements (CUL4-WHB, ROC1 RING domain), further extending the ubiquitylation zone and activating the formation of a catalytic assembly for ubiquitin transfer (see also **D**). (**C**) Flexible tethering of SAMHD1 to the core by $Vpr_{mus}$ places the bulk of SAMHD1 in the ubiquitylation zone and optimises surface accessibility. (**D**) Dynamic processes **A**-**C** together create numerous possibilities for assembly of the catalytic machinery (NEDD8-CUL4-WHB, ROC1, ubiquitin-(ubi-)charged E2) on surface-exposed SAMHD1 lysine side chains. Here, three of these possibilities are exemplified schematically. In this way, ubiquitin coverage on SAMHD1 is maximised.

**Structural insights into Cullin4-RING ubiquitin ligase remodelling by Vpr from simian immunodeficiency viruses**

Sofia Banchenko, Ferdinand Krupp, Christine Gotthold, Jörg Bürger, Andrea Graziadei, Francis O'Reilly, Ludwig Sinn, Olga Ruda, Juri Rappsilber, Christian M. T. Spahn, Thorsten Mielke, Ian A. Taylor, David Schwefel

# Supplementary Information

Supplementary Tables:



Supplementary Figures:



Supplementary References

**Table S1: X-ray data collection and refinement statistics.**

| Sample | DDB1/DCAF1-CtD | DDB1/DCAF1-CtD/T4L-$Vpr_{mus}$ 1-92 | |
|---|---|---|---|
| PDB code | 6zue | - | 6zx9 |
| | | | |
| *Data collection* | | | |
| Space group | I222 | $P2_12_12$ | $P2_12_12$ |
| Cell dimensions | | | |
| a, b, c (Å) | 117.38, 153.63, 223.16 | 266.91, 95.94, 99.35 | 265.90, 95.54, 98.35 |
| α, β, γ (°) | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 |
| Resolution range (Å) | 50.00 (3.28)* – 3.09 | 50.00 (3.83) – 3.61 | 79.07 (2.56) – 2.52 |
| $R_{merge}$ (%) | 8.8 (120.0) | 32.9 (136.1) | 9.9 (162.6) |
| $CC_{1/2}$** | 99.9 (85.3) | 100.0 (35.4) | 99.6 (48.0) |
| I/σ(I) | 13.0 (1.4) | 5.9 (1.3) | 9.8 (1.1) |
| Completeness (%) | 96.8 (97.1) | 99.3 (96.2) | 99.9 (99.8) |
| Redundancy | 4.8 (4.9) | 6.4 (6.1) | 6.6 (6.7) |
| | | | |
| *Refinement* | | | |
| Resolution range (Å) | 48.60 (3.18) – 3.09 | - | 79.07 (2.55) – 2.52 |
| No. reflections | 35922 (2815) | - | 84808 (2880) |
| $R_{work}/R_{free}$ (%) | 22.0/27.9 (35.9/43.5) | - | 21.6/26.1 (41.1/44.6) |
| No. atoms | | | |
| Protein | 11279 | - | 13556 |
| Ligand/ion | - | - | 49 |
| Water | 10 | - | 253 |
| B-factors | | | |
| Protein | 114.6 | - | 75.5 |
| Ligand/ion | - | - | 77.2 |
| Water | 71.7 | - | 54.9 |
| R.m.s. deviations | | | |
| Bond lengths (Å) | 0.004 | - | 0.004 |
| Bond angles (°) | 0.759 | - | 0.794 |

*Numbers in parentheses account for the high-resolution shell
**defined in [1]

**Table S2: Oligonucleotide primer sequences.**

| Restriction enzyme, insert | Sequence | Destination plasmid (restriction enzymes) |
|---|---|---|
| NcoI hsDCAF1 1046 fw. | ggcCCATGGCGccaataaactttacgtcaaggc | pTriEx-6 (NcoI/SacI) |
| SacI hsDCAF1 1396 rev. | ggcGAGCTCctctgccagacgctgcctgcc | |
| XmaI T4L (E11H) 2 fw. | ggcCCCGGGaacatttttgaaatgctgcgtattgatg | pHisSUMO (XmaI/NotI) |
| NotI T4L (E11H) 164 rev. | ggcGCGGCCGCcaggtttttataggcatcccatgtg | |
| AgeI rhSAMHD1 1 fw. | atattACCGGTatgcagcaagccgactcc | pHisSUMO (XmaI/NotII) |
| NotI rhSAMHD1 583 rev. | taattGCGGCCGCTTAatcctgaggcttggtgaaatttc | |
| NotI rhSAMHD1 626 rev. | taattGCGGCCGCTTActttgggtcatctttaaaaagc | |
| NotI rhSAMHD1 582 fw. | ggcGCGGCCGCAcaggatggtgatgttattgcacc | pHisSUMO-T4L (E11H) (NotI/SacI) |
| SacI rhSAMHD1 626 rev. | ggcGAGCTCTTATtatttcggatcatctttaaacagctg | |
| XmaI $Vpr_{mus}$ 1 fw. | ggcCCCGGGatggaacgtgttccgcctagcc | pET49b (XmaI/XhoI) |
| XhoI $Vpr_{mus}$ 135 rev. | ggcCTCGAGTTAttattcatccatacgataacggctc | |
| NotI $Vpr_{mus}$ 1 fw. | ggcGCGGCCGCAatggaacgtgttccgcctagcc | pHisSUMO-T4L (E11H) (NotI/SacI) |
| SacI $Vpr_{mus}$ 92 rev. | ggcGAGCTCTTATTAgcggtgataacaaccttcacgataatg | |
| $Vpr_{mus}$ R15E fw. | GGCATAGCGAAGTTGTTCCGACCACC | pET49b-$Vpr_{mus}$ |
| $Vpr_{mus}$ R15E rev. | GGTCGGAACAACTTGGCTATGCCAAGG | |
| $Vpr_{mus}$ R75E fw. | GATTATATTGAACGTACCCAGACCCTGCTG | |
| $Vpr_{mus}$ R75E rev. | GTCTGGGTACGTTCAATATAATCAATGGCAC | |
| $Vpr_{mus}$ W29A fw. | GCACAGCAGGCCATGGCGGATCTGAATGAAGAAGCA | |
| $Vpr_{mus}$ W29A rev. | TTCTTCATTCAGATCCGCCATGGCCTGCTGTGCCTG | |
| $Vpr_{mus}$ A66W fw. | GGACCGTTGATCAGGCATGGATTGCATGTGCCATTGATTATATTC | |
| $Vpr_{mus}$ A66W rev. | CAATGGCACATGCAATCCATGCCTGATCAACGGTCCAATTC | |
| NdeI ROC1 1 fw. | ggcCATATGgcggcagcgatggatgtgg | pRSF-Duet-1 (NdeI/XhoI) |
| XhoI ROC1 108 rev. | ggcCTCGAGCTActagtgcccatacttttggaattc | |
| BamHI hsCUL4A 2 fw. | ggcGGATCCGgcggacgaggccccgcgg | pRSF-Duet-1-ROC1 (12-108) (BamHI/NotI) |
| NotI hsCUL4A 759 rev. | ggcGCGGCCGCTCAtcaggccacgtagtggtactgattc | |
| BamHI UBCH5C 1 fw. | ggcGGATCCatggcgctgaaacggattaataag | pGex6P1 (BamHI/NotI) |
| NotI UBCH5C 147 rev. | ggcGCGGCCGCTCAcatggcatacttctgagtcc | |

**Table S3: Expression constructs.**

| Protein | UniProt ID | Vector | Expression system |
|---|---|---|---|
| Homo sapiens (hs) DDB1<br>(full length (fl)) | Q16531 | pAcGHLT-B | Sf9 |
| hsDCAF1-CtD<br>(residues 1046-1396) | Q9Y4B6 | pTri-Ex-6 | Sf9 |
| Macaca mulatta<br>(Rhesus macaque, rh) SAMHD1<br>(fl)<br>(ΔCtD, residues 1-583) | G7N4W9 | pHisSUMO | E. Coli Rosetta 2 (DE3) |
| T4L(variant E11H)-rhSAMHD1-CtD<br>(residues 582-626) | G7N4W9 | pHisSUMO | E. Coli Rosetta 2 (DE3) |
| SIVmus Vpr (WT and variants R15E/R75E; W29A/A66W)<br>(fl) | A4UDG5 | pET49b | E. Coli Rosetta 2 (DE3) |
| T4L(variant E11H)-SIVmus Vpr (residues 1-92) | A4UDG5 | pHisSUMO | E. Coli Rosetta 2 (DE3) |
| hsCullin4A (CUL4A) (residues 38-759)<br>hsROC1 (residues 12-108 ) | CUL4A: Q13619<br>ROC1: P62877 | pRSF-Duet-1 (Co-expression) | E. Coli Rosetta 2 (DE3) |
| Mus musculus (mm)<br>mmUBA1<br>(fl) | Q02053 | pET28 | E. Coli Rosetta 2 (DE3) |
| hsUBCH5C<br>(fl) | P61077 | pGex6P1 | E. Coli Rosetta 2 (DE3) |
| hsUBA3<br>(fl)<br>hsAPPBP1<br>(fl) | UBA3: Q8TBC4-2<br>APPBP1: Q13564 | pOPC (Co-expression) | E. Coli Rosetta 2 (DE3) |
| hsUBC12<br>(fl) | P61081 | pGex6P2 | E. Coli Rosetta 2 (DE3) |

342

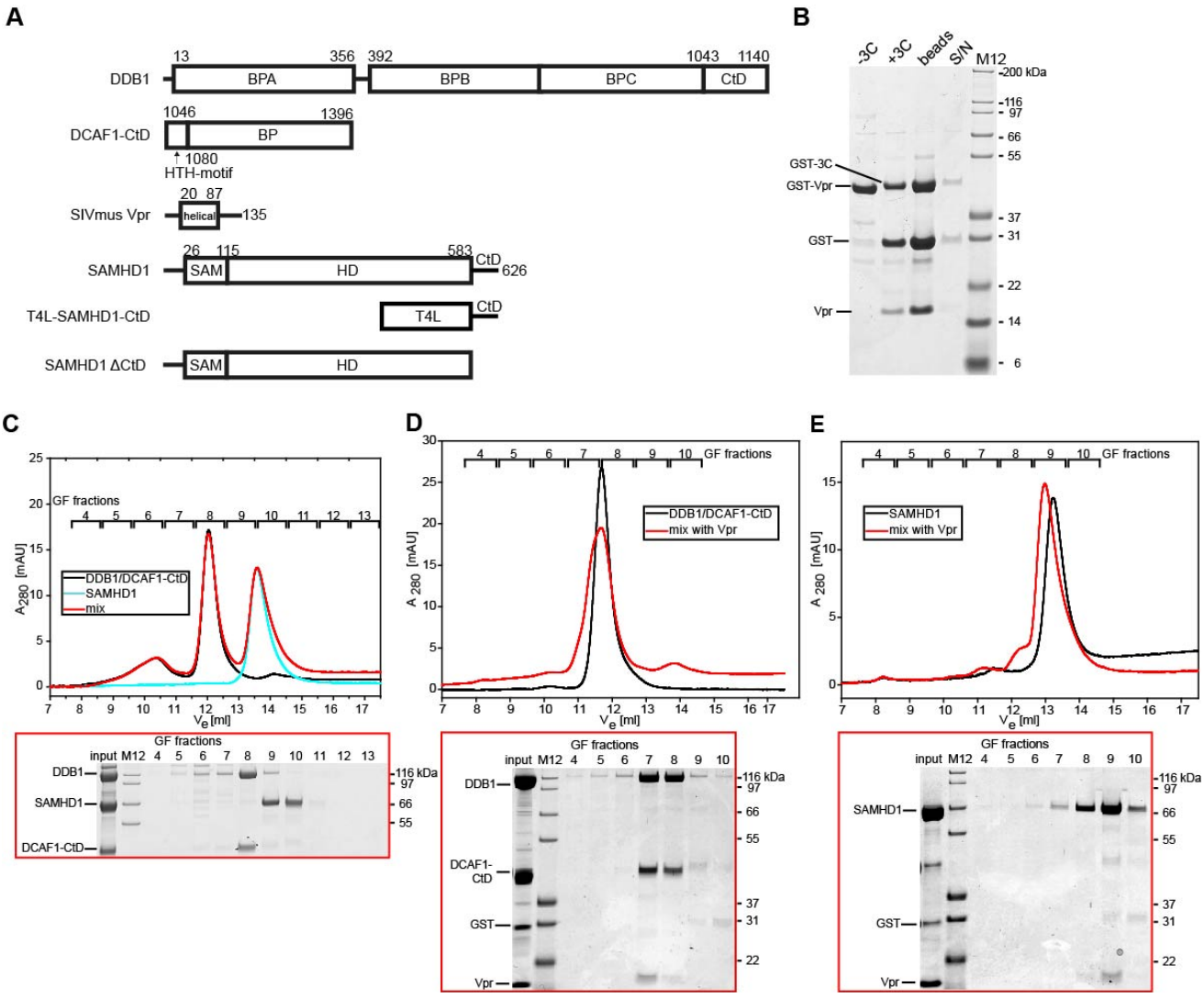



**Fig. S1. Additional biochemical analysis of Vpr$_{mus}$-induced CRL4$^{DCAF1\text{-}CtD}$ specificity redirection towards SAMHD1.** (**A**) Schematic view of the protein constructs used in biochemical analyses. BP – β-propeller domain, HD – histidine-aspartate domain, HTH – helix-turn-helix motif, SAM – sterile alpha motif. (**B**) SDS-PAGE analysis of GST-Vpr$_{mus}$. After treatment with 3C protease to remove the GST-tag (+3C) and GSH-Sepharose pull down to remove protease and tag, no Vpr$_{mus}$ is present in the eluted fraction (S/N) indicating that it interacts non-specifically with the GSH-Sepharose beads and/or becomes insoluble after tag removal. (**C-E**) Analytical GF analysis of DDB1/DCAF1-CtD incubated with SAMHD1 (**C**), DDB1/DCAF1-CtD incubated with Vpr$_{mus}$ (**D**) and SAMHD1 incubated with Vpr$_{mus}$ (**E**). SDS-PAGE of the corresponding GF fractions is shown below each chromatogram.

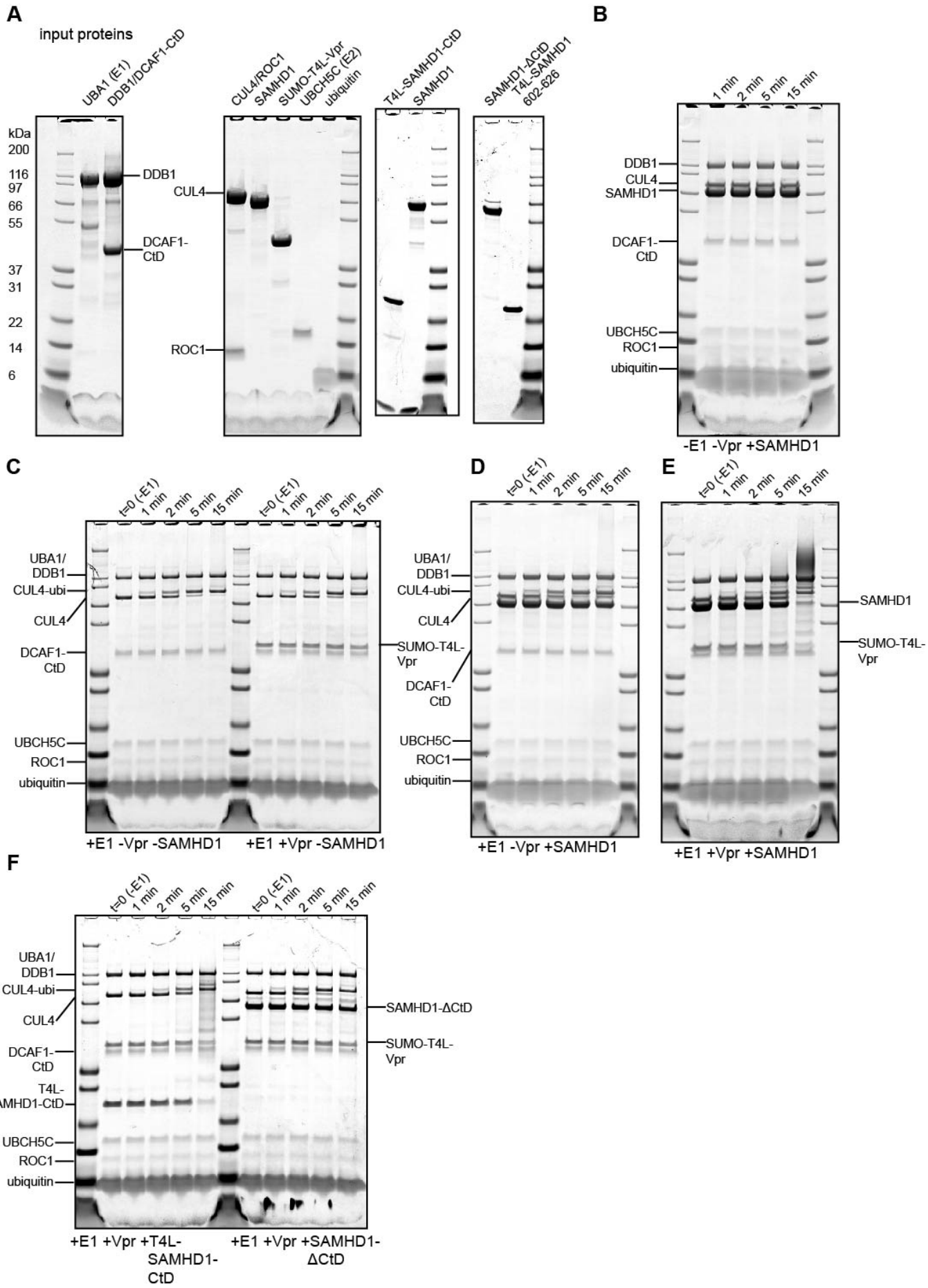

A
input proteins
UBA1 (E1)
DDB1/DCAF1-CtD
CUL4/ROC1
SAMHD1
SUMO-T4L-Vpr
UBCH5C (E2)
ubiquitin
T4L-SAMHD1-CtD
SAMHD1
SAMHD1-ΔCtD
T4L-SAMHD1
602-626
kDa
200
116
97
66
55
37
31
22
14
6
DDB1
CUL4
DCAF1-CtD
ROC1
B
1 min
2 min
5 min
15 min
DDB1
CUL4
SAMHD1
DCAF1-CtD
UBCH5C
ROC1
ubiquitin
-E1 -Vpr +SAMHD1
C
t=0 (-E1)
1 min
2 min
5 min
15 min
UBA1/DDB1
CUL4-ubi
CUL4
DCAF1-CtD
UBCH5C
ROC1
ubiquitin
SUMO-T4L-Vpr
+E1 -Vpr -SAMHD1
+E1 +Vpr -SAMHD1
D
UBA1/DDB1
CUL4-ubi
CUL4
SUMO-T4L-Vpr
DCAF1-CtD
UBCH5C
ROC1
ubiquitin
+E1 -Vpr +SAMHD1
E
SAMHD1
SUMO-T4L-Vpr
+E1 +Vpr +SAMHD1
F
UBA1/DDB1
CUL4-ubi
CUL4
DCAF1-CtD
T4L-SAMHD1-CtD
UBCH5C
ROC1
ubiquitin
SAMHD1-ΔCtD
SUMO-T4L-Vpr
+E1 +Vpr +T4L-SAMHD1-CtD
+E1 +Vpr +SAMHD1-ΔCtD

**Fig. S2. Components, controls and uncropped SDS-PAGE images of *in vitro* ubiquitylation reactions.** (**A**) SDS-PAGE of individually purified protein components used in the *in vitro* ubiquitylation reactions. (**B, C**) Control reactions in the absence of indicated components. (**D-F**) Uncropped gels of reactions shown in Fig. 1C-F. All reactions were incubated at 37°C for the indicated times, stopped by addition of SDS sample buffer and separated on SDS-PAGE.

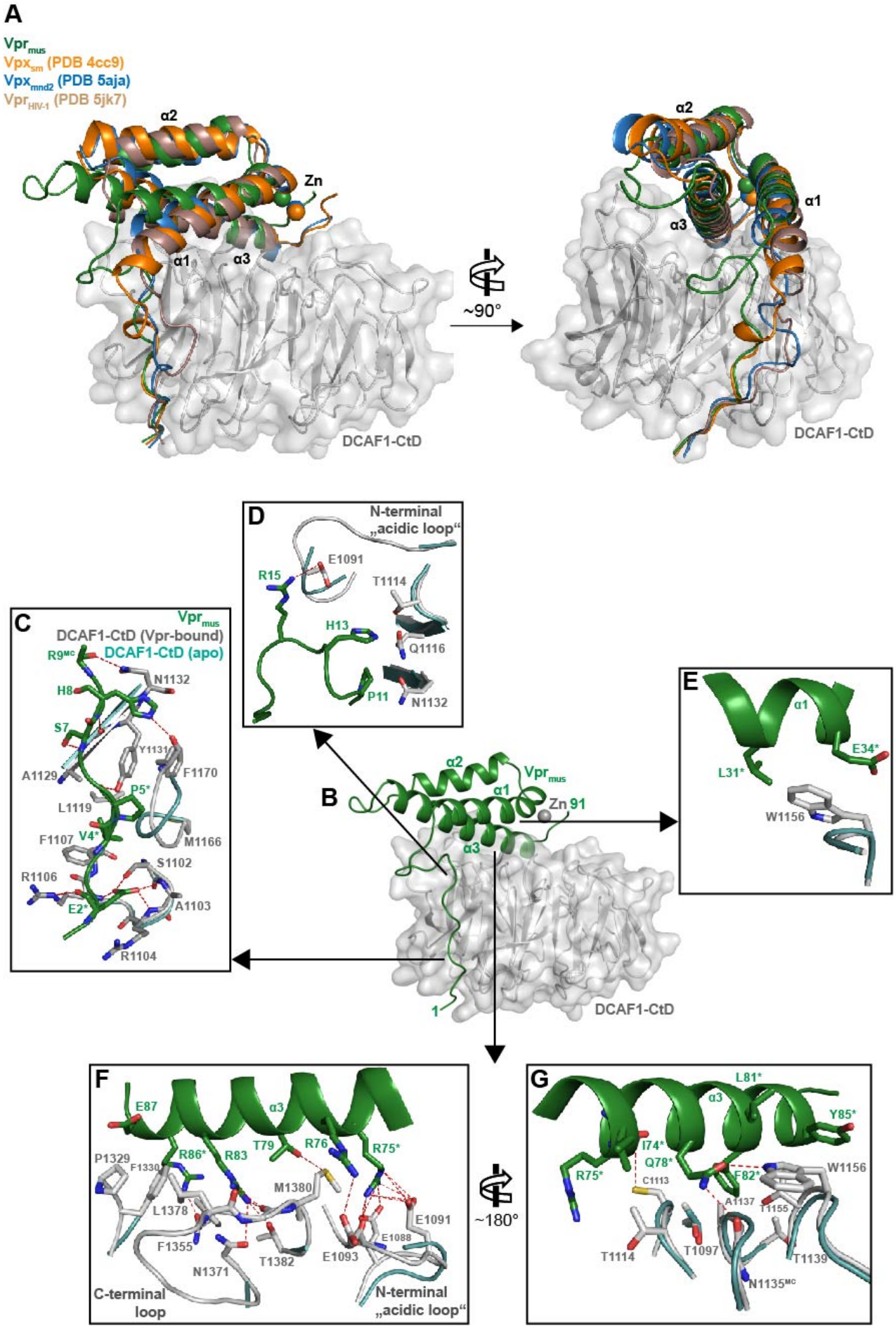

A
Vpr mus
Vpx sm (PDB 4cc9)
Vpx mnd2 (PDB 5aja)
Vpr HIV-1 (PDB 5jk7)
α2
Zn
α1
α3
DCAF1-CtD
~90°
α2
α1
α3
DCAF1-CtD
D
N-terminal „acidic loop“
E1091
R15
T1114
H13
Q1116
P11
N1132
C
Vpr mus
DCAF1-CtD (Vpr-bound)
DCAF1-CtD (apo)
R9MC
N1132
H8
S7
Y1131
A1129
F1170
P5*
L1119
F1107
V4*
M1166
S1102
R1106
E2*
A1103
R1104
B
α2
Vpr mus
α1
Zn
91
α3
1
DCAF1-CtD
E
α1
E34*
L31*
W1156
F
E87
α3
T79
R76
R75*
P1329
F1330
R86*
R83
M1380
L1378
E1091
E1088
F1355
E1093
T1382
N1371
C-terminal loop
N-terminal „acidic loop“
~180°
G
L81*
α3
Y85*
I74*
Q78*
R75*
C1113
F82*
W1156
A1137
T1155
T1114
T1097
T1139
N1135MC

**Fig. S3. Detailed crystal structure analysis of the DDB1/DCAF1-CtD/Vpr$_{mus}$ complex.** (**A**) Superposition of the Vpr$_{mus}$ (green cartoon)/DCAF1-CtD complex with Vpx$_{sm}$ (orange cartoon, PDB: 4cc9)[2], Vpx$_{mnd}$ (blue cartoon, PDB: 5aja)[3] and Vpr$_{HIV-1}$ (light brown cartoon, PDB: 5jk7)[4]. Structures have been aligned with respect to their DCAF1 BP domains but only the DCAF1-CtD from the Vpr$_{mus}$ complex is shown for clarity (grey cartoon and semi-transparent surface). (**B-G**) Details of the DCAF1-CtD/Vpr$_{mus}$ interaction. (**B**) The structure of the complex is shown in the same orientation as Fig. 2A, left panel. The insets (**C**-**G**) show individual interaction areas in more detail, Vpr$_{mus}$ (green), Vpr$_{mus}$-bound DCAF1-CtD (grey) and apo-DCAF1-CtD (light blue). Selected amino acid residues, that make intermolecular interactions, are shown as sticks, and hydrogen bonds/electrostatic interactions as dashed red lines. Vpr$_{mus}$ residues with asterisks are type-conserved within all Vpr/Vpx proteins.

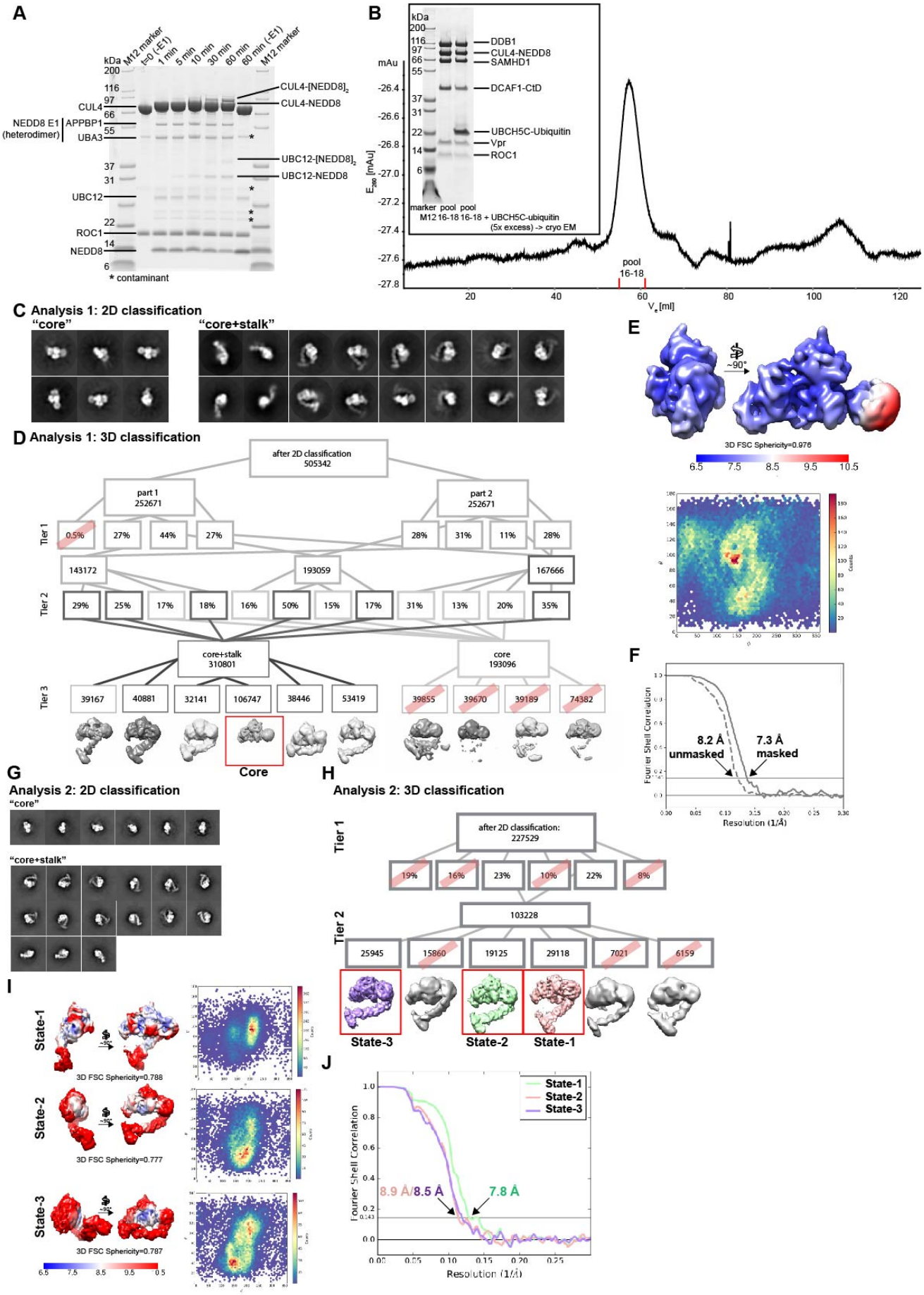

A
kDa
M12 marker
t=0 (-E1)
1 min
5 min
10 min
30 min
60 min
60 min (-E1)
M12 marker
200
116
97
66
55
37
31
22
14
6
CUL4
NEDD8 E1 (heterodimer)
APPBP1
UBA3
UBC12
ROC1
NEDD8
CUL4-[NEDD8]2
CUL4-NEDD8
UBC12-[NEDD8]2
UBC12-NEDD8
* contaminant
B
kDa
DDB1
CUL4-NEDD8
SAMHD1
DCAF1-CtD
UBCH5C-Ubiquitin
Vpr
ROC1
marker pool pool
M12 16-18 16-18 + UBCH5C-ubiquitin (5x excess) -> cryo EM
mAu
E280 [mAu]
pool 16-18
Ve [ml]
C Analysis 1: 2D classification
"core"
"core+stalk"
D Analysis 1: 3D classification
after 2D classification 505342
part 1 252671
part 2 252671
Tier 1
Tier 2
Tier 3
core+stalk 310801
core 193096
Core
E
~90°
3D FSC Sphericity=0.976
F
Fourier Shell Correlation
8.2 Å unmasked
7.3 Å masked
Resolution (1/Å)
G
Analysis 2: 2D classification
"core"
"core+stalk"
H
Analysis 2: 3D classification
after 2D classification: 227529
103228
State-3
State-2
State-1
I
State-1
State-2
State-3
3D FSC Sphericity=0.788
3D FSC Sphericity=0.777
3D FSC Sphericity=0.787
J
Fourier Shell Correlation
8.9 Å/8.5 Å
7.8 Å
Resolution (1/Å)

**Fig. S4. Cryo-EM analysis of the CRL4$^{DCAF1\text{-}CtD}$/Vpr$_{mus}$/SAMHD1 complex.** (**A**) In vitro neddylation of CUL4/ROC1. Protein was mixed with purified neddylation-E1 (APPBP1/UBA3 heterodimer), E2 (UBC12) and NEDD8. The reaction was incubated at 25°C, samples were taken at indicated times, stopped by addition of SDS sample buffer and separated on SDS-PAGE. (**B**) GF analysis of the CUL4-NEDD8/ROC1/DDB1/DCAF1-CtD/Vpr$_{mus}$/SAMHD1 complex with pooled fractions indicated. A 5x molar excess of UBCH5C-ubiquitin was added before plunge-freezing for cryo-EM experiments, in an attempt to stabilise the assembly. However, no density in any of the reconstructions could be assigned to UBCH5C-ubiquitin, indicating low binding affinity and/or heterogeneity in its mode of binding. (**C**) 2D class averages depicting either “core” or “core+stalk” classes of analysis 1. (**D**) Sorting tree after 2D classification. In Tier 3, the core reconstruction was identified, containing 106,747 particle images (red box). (**E**) Local resolution of the core reconstruction after refinement, indicating a resolution range from 6.5 Å in the hydrophobic interior of DDB1 to 10.5Å in the DDB1 BPB domain. On the right, the Euler distribution is shown. (**F**) FSC curve of the core reconstruction after refinement. (**G**) 2D class averages depicting either “core” or “core+stalk” classes of analysis 2. (**H**) 3D sorting tree after 2D classification. Conformational states-1, -2 and -3 are indicated. (**I**) Local resolution and Euler distribution of states-1, -2 and -3. (**J**) FSC curves for state-1, -2 and -3 reconstructions.

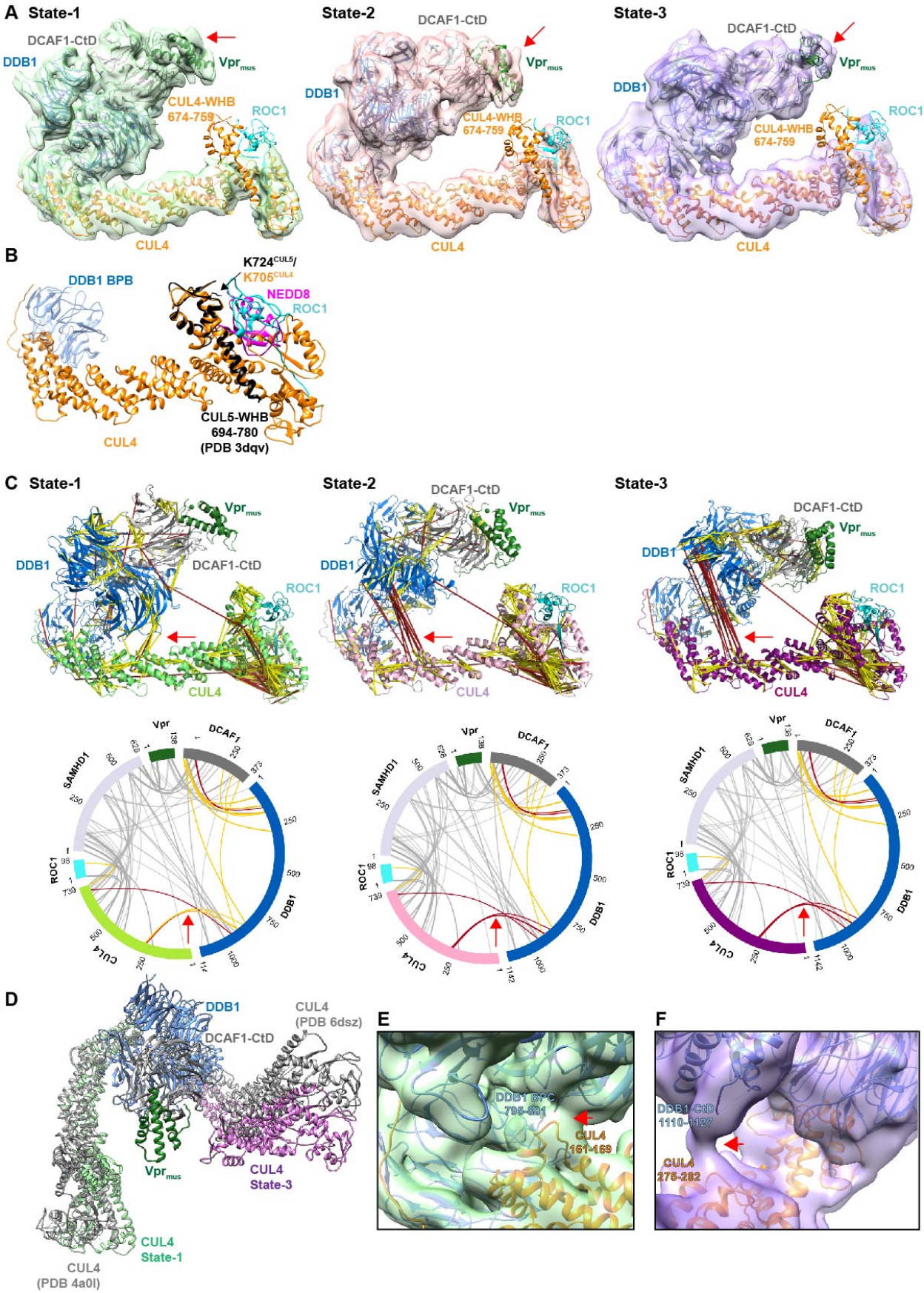

A
State-1
DCAF1-CtD
DDB1
Vpr mus
CUL4-WHB
674-759
ROC1
CUL4
State-2
DCAF1-CtD
DDB1
Vpr mus
CUL4-WHB
674-759
ROC1
CUL4
State-3
DCAF1-CtD
DDB1
Vpr mus
CUL4-WHB
674-759
ROC1
CUL4
B
DDB1 BPB
K724CUL5/
K705CUL4
NEDD8
ROC1
CUL5-WHB
694-780
(PDB 3dqv)
CUL4
C
State-1
State-2
State-3
Vpr mus
DCAF1-CtD
DDB1
ROC1
CUL4
Vpr
DCAF1
SAMHD1
ROC1
DDB1
CUL4
D
DDB1
CUL4
(PDB 6dsz)
DCAF1-CtD
Vpr mus
CUL4
State-3
CUL4
State-1
CUL4
(PDB 4a0l)
E
DDB1 BPC
795-801
CUL4
161-169
F
DDB1-CtD
1110-1127
CUL4
275-282

**Fig. S5. Detailed structural and CLMS analysis of the CRL4$^{DCAF1\text{-}CtD}$/ Vpr$_{mus}$ /SAMHD1 complex.** (**A**) Side-by-side comparison of state-1, 2 and 3 reconstructions, coloured as in Fig. 4. Molecular models of the DDB1/DCAF1-CtD/Vpr$_{mus}$ crystal structure and CUL4/ROC1 (PDB 2hye)[5] have been fitted as rigid bodies into the volumes and are shown as cartoons. DDB1/DCAF1-CtD/Vpr$_{mus}$ is coloured as in Fig. 3, CUL4 is coloured yellow and ROC1 cyan. All states show additional density corresponding to SAMHD1-CtD, indicated by the red arrows. (**B**) Superposition of the neddylated CUL5 C-terminal WHB domain (black cartoon, PDB 3dqv)[6] on the CUL4 WHB (PDB 2hye), coloured as in **A**. Respective lysine residues, which are covalently modified with NEDD8, are indicated. (**C**) Upper panel: cross-links, identified by CLMS, mapped on molecular models representing state-1, -2 and -3. Satisfied crosslinks (<25 Å) are coloured yellow, violated crosslinks red. Red arrows indicate a subset of cross-links between DDB1 and CUL4, whose distance restraints are satisfied in state-1, and increasingly violated in states-2 and -3. Lower panel: circle plot of CLMS data for states-1, -2 and -3, using the same colour scheme as in the upper panel. Grey lines represent crosslinks between residues that are not present in the molecular models. Only crosslinks between subunits are displayed. (**D**) Comparison of outermost CUL4 stalk orientations observed in the cryo-EM analysis presented here (states-1 and -3, coloured as in Fig. 4, show 119.5° rotation of DDB1 BPB) to the two most extreme stalk positions present in previous crystal structures (PDB 4a0l[7], PDB 6dsz[8], coloured grey, show 143.4° DDB1 BPB rotation). (**E**, **F**) Detailed view of state-1 (E) and state-3 (F) cryo-EM density. Red arrows indicate contacts between CUL4A (orange cartoon) and DDB1 BPA/BPC/CtD (blue cartoon).

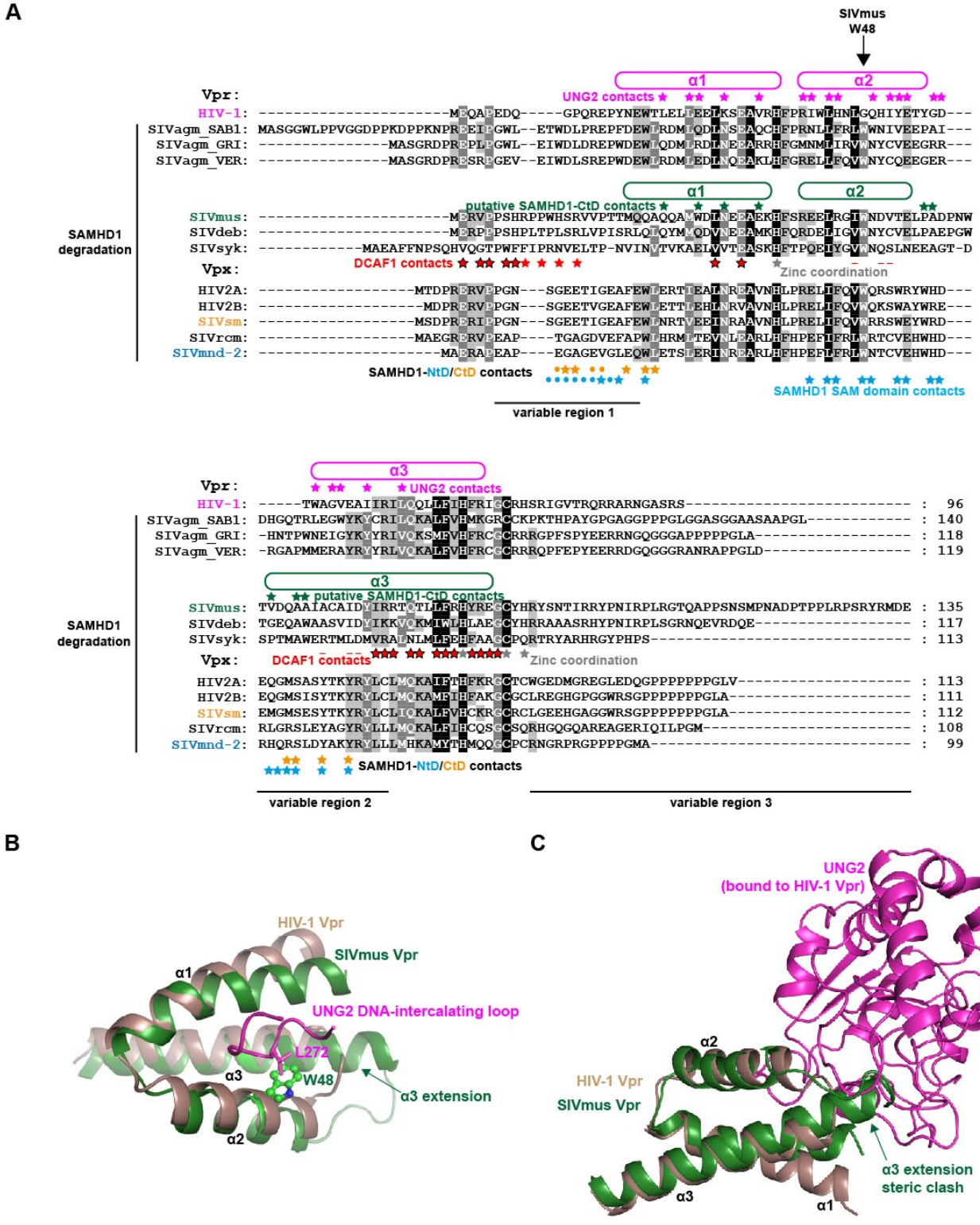



**Fig. S6. Multiple sequence alignment of Vpr/Vpx proteins, detailed structural comparison between $Vpr_{mus}$ and $Vpr_{HIV-1}$.** (**A**) Sequence alignment of indicated Vpr and Vpx proteins. Helices are indicated by the boxes above the amino acid sequences for $Vpr_{HIV-1}$ (pink) and $Vpr_{mus}$ (green). $Vpr_{HIV-1}$ side chains involved in UNG2-binding are indicated with pink asterisks. $Vpr_{mus}$ side chains putatively involved in SAMHD1-CtD-binding are indicated by green asterisks. $Vpx_{sm}$ side chains targeting

SAMHD1-CtD are indicated with orange asterisks, and $Vpx_{mnd2}$ side chains contacting N-terminal SAMHD1 domains are highlighted with blue asterisks. Red symbols mark $Vpr_{mus}$ side chains involved in DCAF1-CtD-binding. The non-outlined symbols indicate DCAF1-CtD-contacting side chains unique to $Vpr_{mus}$, and dashes show DCAF1-binding side chains, which are in contact with DCAF1-CtD in other Vpr/Vpx structures, but not in $Vpr_{mus}$. Grey asterisks mark Vpr/Vpx side chains involved in zinc coordination. (**B**) Structural alignment of $Vpr_{HIV-1}$ (PDB 5jk7[4], light brown) in complex with UNG2 and $Vpr_{mus}$ (green). Protein backbone is shown in cartoon representation. For clarity, only the DNA-intercalating loop of UNG2 is shown (pink), which inserts into a hydrophobic pocket created by the $Vpr_{HIV-1}$ helix bundle. Note the steric clash between UNG2 side chain L272 and $Vpr_{mus}$ residue W48 in the structural superposition. (**C**) Alternative view of the structural alignment of $Vpr_{HIV-1}$ (light brown) in complex with UNG2 (pink) and $Vpr_{mus}$ (green). Note the steric clash between UNG2 and the extended Helix-3 of $Vpr_{mus}$.

353

## Supplementary References


1. Karplus, P.A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030-3 (2012).
2. Schwefel, D. et al. Structural basis of lentiviral subversion of a cellular protein degradation pathway. *Nature* **505**, 234-8 (2014).
3. Schwefel, D. et al. Molecular Determinants for Recognition of Divergent SAMHD1 Proteins by the Lentiviral Accessory Protein Vpx. *Cell Host Microbe* **17**, 489-99 (2015).
4. Wu, Y. et al. The DDB1-DCAF1-Vpr-UNG2 crystal structure reveals how HIV-1 Vpr steers human UNG2 toward destruction. *Nat Struct Mol Biol* **23**, 933-940 (2016).
5. Angers, S. et al. Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. *Nature* **443**, 590-3 (2006).
6. Duda, D.M. et al. Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. *Cell* **134**, 995-1006 (2008).
7. Fischer, E.S. et al. The molecular basis of CRL4DDB2/CSA ubiquitin ligase architecture, targeting, and activation. *Cell* **147**, 1024-39 (2011).
8. Shabek, N. et al. Structural insights into DDA1 function as a core component of the CRL4-DDB1 ubiquitin ligase. *Cell Discov* **4**, 67 (2018).


354