INTEGRATIVE STRUCTURAL MODELING OF NUCLEAR PORE COMPLEXES

Dissertation

to obtain the degree of

Doctor of Natural Sciences (Dr. rer. nat.)

University of Hamburg (UHH)

Faculty of Mathematics, Informatics and Natural Sciences

at the Department of Chemistry

by

Vasileios Rantos

from Kalabaka, Greece

Hamburg, 2022

Thesis Reviewers and Examination Committee members

This Thesis (cumulative format) titled "Integrative structural modeling of nuclear pore complexes" was reviewed by:

- **Dr. Jan Kosinski**, group leader at the European Molecular Biology Laboratory (EMBL) and the Centre of Structural Systems Biology (CSSB), Hamburg, Germany

- **Prof. Dr. Andrew Torda**, Professor and group leader in Biomolecular Modeling in the Center for Bioinformatics at the Chemistry Dept. in University of Hamburg (UHH), Hamburg, Germany

The Examination Committee consisted of the following members:

- **Prof. Dr. Kay Grünewald (Chair)**, Professor and group leader in Institute of Biochemistry and Molecular Biology at the Chemistry Dept. in University of Hamburg (UHH), Hamburg, Germany

- **Prof. Dr. Andrew Torda (Deputy Chair)**, Professor and group leader in Biomolecular Modeling in the Center for Bioinformatics at the Chemistry Dept. in University of Hamburg (UHH), Hamburg, Germany

- **Dr. Christian Löw (Member)**, group leader at the European Molecular Biology Laboratory (EMBL) and the Centre of Structural Systems Biology (CSSB), Hamburg, Germany

The Doctoral Disputation took place on **25.11.2022**.

Place and duration of the dissertation

This dissertation was conducted in the group of Dr. Jan Kosinski in the facilities of European Molecular Biology Laboratory (EMBL) in Hamburg, Germany. The carrying out period was from 12.10.2017 until 11.12.2021.

I. Table of contents

II. List of publications 6
III. List of Abbreviations 7
IV. List of figures 9
1. Zusammenfassung 11
2. Abstract 13
3. Introduction 15
3.1 The nuclear pore complex (NPC) 15
3.1.1 Nucleoporins: The building blocks of NPCs 16
3.1.2 The architecture of scaffold NPC subcomplexes 17
3.1.3 Peripheral NPC modules 19
3.1.4 Nucleocytoplasmic transport and NPC constriction 22
3.2 Integrative structural modeling of protein assemblies 24
3.2.1 Fundamentals of integrative structural modeling methods 24
3.2.2 Integrative approaches for the modeling of NPCs across species 30
4. Research objectives and motivation 33
5. Synopsis of listed publications 34
5.1 PUBLICATION I: "In-cell architecture of the nuclear pore and snapshots of its turnover" 34
5.1.1 Structural architecture of the *S. cerevisiae* NPC scaffold 35
5.1.2 Revised conformation of the mRNA export platform 40
5.1.3 Integrative density allocation of the linker Nup116 42
5.1.4 Integrative models of NPCs from Nup116Δ *S. cerevisiae* cells 45
5.2 PUBLICATION II: "Nuclear pores dilate and constrict in cellulo" 48
5.2.1 Density allocation of the scaffold SpNPC subcomplexes 48
5.2.2 Structural elucidation of the SpNPC architecture 55
5.3 PUBLICATION III: "Integrative structural modeling of macromolecular complexes using Assembline" 57
5.4 PUBLICATION IV: "Structure of prototypic peptide transporter DtpA from *E. coli* in complex with Valganciclovir provides insights into drug binding of human PepT1" 61
6. Discussion 68
6.1 Challenges in integrative modeling and future Assembline applications 68

6.2 The first in-cell NPC models and their impact on future structural studies 72
7. Bibliography 75
8. Appendix 85
8.1 List of hazardous substances according to GHS (hazard symbols H- and P) 85
9. Acknowledgements 86
10. Declaration of academic honesty (Eidesstattliche Versicherung) 87

II. List of publications

1. Ural-Blimke, Y., Flayhan, A., Strauss, J., **Rantos, V.**, Bartels, K., Nielsen, R., Pardon, E., Steyaert, J., Kosinski, J., Quistgaard, E. M. & Löw, C. Structure of Prototypic Peptide Transporter DtpA from *E. coli* in Complex with Valganciclovir Provides Insights into Drug Binding of Human PepT1. *J. Am. Chem. Soc.* 141, 2404–2412 (2019).

2. Allegretti, M., Zimmerli, C. E., **Rantos, V.**, Wilfling, F., Ronchi, P., Fung, H. K. H., Lee, C. W., Hagen, W., Turoňová, B., Karius, K., Börmel, M., Zhang, X., Müller, C. W., Schwab, Y., Mahamid, J., Pfander, B., Kosinski, J. & Beck, M. In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature* 586, 796–800 (2020).

3. Zimmerli, C. E., Allegretti, M., **Rantos, V.**, Goetz, S. K., Obarska-Kosinska, A., Zagoriy, I., Halavatyi, A., Hummer, G., Mahamid, J., Kosinski, J. & Beck, M. Nuclear pores dilate and constrict in cellulo. *Science* (80-.). (2021).

4. Rantos, V., Karius, K. & Kosinski, J. Integrative structural modeling of macromolecular complexes using Assembline. *Nat. Protoc.* 1–25 (2021).

III. List of Abbreviations

NE	Nuclear envelope
ER	Endoplasmic reticulum
NPC	Nuclear pore complex
INM	Inner nuclear membrane
ONM	Outer nuclear membrane
Nup	Nucleoporin
FG-domain	Phenylalanine-Glycine rich domain
FG-Nup	Phenylalanine-Glycine rich nucleoporin
IRC	Inner ring complex
IR	Inner ring
CR	Cytoplasmic ring
NR	Nuclear ring
NTR	Nuclear transport receptor
DID	Dynein-interacting domain
LR	Luminal ring
POM	Pore membrane protein
Ig-like	Immunoglobulin-like
SLiM	Short linear motif
NLS	Nuclear localization signal
NES	Nuclear export signal
cryo-EM	Cryo-electron microscopy
cryo-ET	Cryo-electron tomography
CrNPC	*C. reinhardtii* NPC
ScNPC	*S. cerevisiae* NPC
ScNup	*S. cerevisiae* Nup
SpNPC	*S. pombe* NPC

MFS	Major facilitator superfamily
POT	Proton-dependent oligopeptide transporter
hPepT1	Human PepT1
hPepT2	Human PepT2
HMM	Hidden Markov Model
ED	Energy depletion
NMR	Nuclear magnetic resonance
SAXS	Small-angle X-ray scattering
IMP	Integrative Modeling Platform
PMI	Python Modeling Interface
GUI	Graphical user interface
AI	Artificial intelligence
GPU	Graphics processing unit

IV. List of figures

Figure 1: Conceptual localization and organization of conserved human and *S. cerevisiae* nucleoporins arranged in the structural framework of the human NPC 16

Figure 2: Overview of the human NR Y-complex architecture 18

Figure 3: Dissection of the human IR 19

Figure 4: Structural architecture of the peripheral P-complex 20

Figure 5: Molecular representation of the Tpr N-terminus 21

Figure 6: Schematic description of a typical integrative modeling workflow 25

Figure 7: Example results from sampling exhaustiveness assessment for PDB: 1AVX 28

Figure 8: Graphical summary of the multilayered organization of IMP 29

Figure 9: Depiction of the composite human NPC model 30

Figure 10: Overview of the integrative ScNPC model 31

Figure 11: Comparative top view of the in vitro and in-cell ScNPC maps 35

Figure 12: Systematic fitting of inner and outer ring components into the in-cell *S. cerevisiae* NPC map 37

Figure 13: Overview of the in-cell ScNPC architecture and comparison with the previous in vitro model 39

Figure 14: Validation of the Nup159 complex positioning 41

Figure 15: Integrative density assignment of Nup116 43

Figure 16: Revised positioning of Nup116 and possible Nup188 SH3-like domain interactions with linker ScNups 44

Figure 17: In-cell-based integrative models of ScNPCs built with Assembline 46

Figure 18: Comparison of cryo-ET maps of SpNPCs under control and energy depletion conditions 49

Figure 19: Systematic fitting of composite SpNPC subcomplex models to the control SpNPC maps 51

Figure 20: Density differences between the control SpNPC map and maps from knock-out *S. pombe* strains 52

Figure 21: Representative results from systematic fitting of composite SpNPC subcomplexes to intermediate and constricted SpNPC maps (from ED *S. pombe* cells) 54

Figure 22: Overview of the SpNPC integrative models built based on datasets from control and energy depleted *S. pombe* cells 56

Figure 23: Schematic representation of the fit library generation step from Assembline 59

Figure 24: Depiction of the complete workflow from Assembline 60

Figure 25: Representation of the DtpA crystal structure without the bound ligand (valganciclovir) 62

Figure 26: Schematic representation of valganciclovir bound to DtpA 63

Figure 27: Curated pairwise sequence alignment of hPepT1 and DtpA 64

Figure 28: Representative hPepT1 homology model with bound valganciclovir 65

Figure 29: Local structural quality of human PepT1 homology model assessed by QMEANBrane 66

Figure 30: Graphical comparison of the ligand binding sites of DtpA and hPepT1 in complex with valganciclovir 67

1. Zusammenfassung

In der Vergangenheit haben Strukturbiologen hauptsächlich zwei experimentelle Methoden zur Strukturbestimmung von Proteinen angewandt: Röntgenkristallographie und Kernspinresonanzspektroskopie (NMR). In jüngster Zeit wurde die Kryo-Elektronenmikroskopie (Kryo-EM) aufgrund bedeutender experimenteller und rechnergestützter Fortschritte ebenfalls in die Sammlung häufig genutzter Strukturbestimmungsmethoden aufgenommen. Diese Methode ermöglicht die Aufklärung großer makormolekularer Strukturen, die zuvor aufgrund Ihrer Komplexität oder subzellulären Lokalisierung nicht untersucht werden konnten. Trotz aller Erfolge der Kryo-EM zeigen die meisten Strukturen großer Proteinkomplexe, die mit Hilfe dieser Methode aufgelöst wurden, hochaufgelöste Merkmale, die sich hauptsächlich auf den Kern des Proteins bezogen. Gleichzeitig wurden die flexibleren peripheren Domänen in der Regel nur unzureichend abgebildet. Solche Fälle traten besonders bei Anwendungen der Kryo-Elektronentomographie (Kryo-ET) auf, wo die gelösten Strukturen typischerweise eine mittlere bis niedrige Auflösung aufweisen, was das Verständnis der Funktion dieser Komplexe auf atomarer Ebene erschwert, insbesondere in ihrem ursprünglichen zellulären Kontext.

Eine Methode, die zunehmend angewandt wird, um Strukturmodelle großer Proteinkomplexe zu erhalten, ist die computergestützte integrative/hybride Strukturmodellierung. Ein typischer integrativer Modellierungs-Workflow kombiniert vorhandene Informationen aus komplementären Methoden wie Rönntegenstrukturanalyse, Kryo-EM, NMR, Crosslinking-Massenspektrometrie, und Homologiemodellierung, um Ensembles von Modellen zu erstellen. Diese Modelle haben einen höherer Genauigkeit und Präzision im Vergleich zu Modellen die nur durch einzelne rechnerische oder experimentelle Ansätze bestimmt wurden. Die derzeit verfügbare integrative Modellierungssoftware und die entsprechenden Protokolle sind jedoch entweder auf Komplexe mit einfacher Architektur beschränkt oder erfordern bei der Anwendung auf komplexere Proteinsysteme umfangreiche projektspezifische Modifikationen. Dies setzt ein großes Fachwissen seitens der Benutzer voraus.

In dieser Arbeit beschreibe ich die Funktionalitäten und algorithmischen Neuerungen von Assembline, einem vielseitigen Modellierungspaket für die effiziente und einfach zugängliche integrative Modellierung von Proteinkomplexen mit hochkomplexer Zusammensetzung. Assembline wurde unter Verwendung der Software-Bibliothek “Integrative Modeling Platform” implementiert. Auf dieser Software-Bibliothek aufbauend wurden benutzerdefinierte Algorithmen zur effizienten suche im Konformationsraum, eine vielseitige Konfigurationssprache, und ein grafisches Interface für die Eingabevorbereitung, Modellierung und Analyse entwickelt. Außerdem werden durch die Implementierung zusätzliche, benutzerdefinierte Randbedingungen ermöglicht, sowie die flexible und symmetrische Modellierung unterstützt. Zusammengenommen machen diese Eigenschaften Assembline zu einer der intuitivsten und am einfachsten zu nutzenden Programme für die integrative Modellierung, die derzeit verfügbar ist.

Ich habe die in Assembline gebündelten integrativen Modellierungsmethoden erfolgreich eingesetzt, um Strukturmodelle des Kernporenkomplexes (NPC) für zwei Hefearten zu erstellen. Im ersten Fall basierten die von mir erstellten NPC-Modelle für *S. cerevisiae* in erster Linie auf

in-cell cryo-ET-Datensätzen und zeigten auffällige strukturelle Unterschiede im Vergleich zu dem veröffentlichten integrativen NPC-Modell, das zuvor auf der Grundlage von in vitro Daten erstellt wurde. Die neuen, in vivo basierten NPC-Modelle erschienen im Vergleich zu den in vitro Modellen in ihrem Durchmesser deutlich erweitert. Gleichzeitig schienen wichtige Strukturelemente, wie die mRNA-Exportplattform, in den in vivo NPC-Modellen völlig anders positioniert zu sein. Für den zweiten Fall habe ich einen deutlich weniger untersuchten NPC aus *S. pombe* modelliert. Diese integrativen NPC-Modelle basierten ebenfalls auf in-cellulo Kryo-ET daten, die unter nativen und energiearmen zellulären Bedingungen aufgenommen wurden, um Schnappschüsse möglicher Konformationsänderungen des NPC zu erfassen. Die NPC-Modelle von *S. pombe* deuten auf die Existenz einer gespaltenen Y-Komplex-Schnittstelle am zytoplasmatischen Ring des NPC hin, die die strukturelle Integrität des äußeren Rings unterbricht und somit das seit langem bestehende Dogma einer aus drei Ringen bestehende NPC-Architektur in Frage stellt. Darüber hinaus zeigten die NPC-Modelle aus den energiearmen Zellen einen deutlich verringerten Durchmesser im Vergleich zu den nativen NPC-Modellen. Dies verdeutlicht die strukturelle Plastizität der NPCs als Reaktion auf physikalisch-chemische Einflüsse.

Beide Modellierungsanwendungen für den NPC heben die Notwendigkeit, die Strukturen solch wichtiger Proteinkomplexe in ihrem nativen zellulären Kontext zu untersuchen, hervor. Sie sind ein eindrucksvolles Beispiel für die Kombination modernster experimenteller Methoden mit den neu entwickelten integrativen Modellierungsmethoden.

2. Abstract

Historically, structural biologists have applied mainly two experimental methods for the structure determination of proteins: X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. Recently, cryo-electron microscopy (cryo-EM) was also added to the regular toolkit of structural biologists due to significant experimental and computational-based advances. These developments allowed the elucidation of complex structural architectures that were previously impossible to study due to the target's complexity or sub-cellular localization. Despite all the recent cryo-EM successes, though, most of the structures of large protein systems that were resolved by such experiments exhibited high-resolution features mostly towards the protein's core. At the same time, more flexible peripheral domains were usually poorly imaged. Such cases were very prominent in cryo-electron tomography (cryo-ET) applications where the solved structures typically range from mid-to-low resolution, thus hindering the process of understanding the function of these complexes at the atomic level, especially in their native cellular context.

A state-of-the-art method that has been increasingly applied to obtain structural models of large protein assemblies is computational integrative/hybrid structural modeling. A typical integrative modeling workflow combines existing information from complementary techniques, such as X-ray crystallography, cryo-EM, NMR, cross-linking mass spectrometry, or homology modeling to produce ensembles of models with higher accuracy and precision compared to models produced by individual computational or experimental approaches. However, the currently available integrative modeling software and relevant protocols are either limited to complexes of simple architectures or require extensive custom modifications, and thus major expertise from the users, when applied to more complex protein systems.

In this Thesis, I describe the functionalities and algorithmic novelties of Assembline, a versatile modeling package for efficient and accessible integrative modeling of protein complexes with very complex architectures. Assembline is implemented using the Integrative Modeling Platform library, on top of which it includes custom algorithms for efficient sampling of the conformational space, versatile protein system configuration language, and graphical interface for input preparation, modeling, and analysis, as well as additional custom restraints. Flexible and symmetrical modeling are also supported, hence establishing Assembline as one of the most straightforward and intuitive software for integrative modeling currently available.

I successfully applied Assembline functionalities to build structural models of the nuclear pore complex (NPC) for two yeast species. In the first case, the *S. cerevisiae* NPC models I built were primarily based on in-cell cryo-ET datasets and revealed striking structural differences compared with the published integrative NPC model built earlier based on in vitro data. The new in-cell-based NPC models appeared significantly dilated compared to their in vitro-based equivalents. At the same time, major structural elements comprising them, such as the mRNA export platform, seemed reoriented entirely in the in-cell NPC models. For the second case, I modeled a significantly less studied NPC from *S. pombe*. These integrative NPC models were also based on in-cell cryo-ET maps acquired under native and energy depletion cellular conditions to capture structural snapshots of possible conformational changes. The *S. pombe* NPC models revealed the

unusual cytoplasmic side of this NPC that does not form a continuous ring, which challenges the long-standing dogma of the conserved three-ringed NPC architecture. Additionally, the NPC models from the energy-depleted cells revealed a significantly constricted architecture compared to the native NPCs and highlighted the structural plasticity of these complexes in response to physicochemical cues.

Both modeling applications on the NPCs highlight the need to study the structures of such important protein complexes in their native cellular context. They are striking examples of the combination of state-of-the-art experimental methods with the newly developed integrative modeling methods.

3. Introduction

3.1 The nuclear pore complex (NPC)

Eukaryotes (the term is derived from Greek, meaning "true kernel") emerged over one and a half billion years ago and the exact evolutionary events that led to their occurrence are not yet fully understood[1]. Unlike prokaryotes, eukaryotic cells developed a complex subcellular compartmentalization architecture by membrane systems, such as the nucleus with surrounding nuclear envelope (NE) and the endoplasmic reticulum (ER)[2]. Although this feature provided eukaryotes with the advantage of conducting several functions in specialized organelles, it increased the cellular "logistical" costs of precisely localizing and transporting molecules across the organelle membranes[3]. An example of such precise localization and transport of molecules is the import into the cell nucleus of chromatin-binding proteins and polymerases, which are responsible for gene transcription.

The NE is the double-membrane defining the boundaries of the nucleus, the largest and most distinct endomembrane compartment of the eukaryotic cell. This membrane bi-layer separates the cellular genome (and nuclear environment in general) from the rest of the cell and serves as the ultimate barrier for nucleocytoplasmic molecular transport. Nuclear pore complexes (NPCs) reside within the NE where they fuse the inner- and outer nuclear membrane (INM, ONM) hence forming a central channel-like structure[4]. They are the main facilitators and regulators of nucleocytoplasmic transport.

NPCs are amongst the largest known proteinaceous assemblies in eukaryotic cells, with an impressive molecular weight of approximately 60 MDa in yeast and up to 120 MDa in humans[5]. These complexes are built from hundreds up to thousands of nucleoporin (Nup) copies, totaling up to 1000 molecules in human NPCs. Despite this high number of Nup copies, only ~30 different Nups comprise each NPC. The Nups assemble into functionally and structurally distinct NPC subcomplexes. The overall structural arrangement of these subcomplexes in every NPC gives rise to the symmetrical architecture of the latter as they form an 8-fold rotational symmetric pore[6,7]. Although the exact Nup subcomplex arrangement and stoichiometries may vary significantly across eukaryotic species, all studied NPCs (prior to the work presented in this Thesis) exhibit a common architectural "blueprint". Essential element of this architecture is the inner ring (IR). This ring is forming a ~50 nm wide aqueous central channel lined with phenylalanine-glycine (FG)-rich repeat domains of the so-called FG-domains, which are establishing the pore's selectively permeable diffusion barrier at the plane of the NE[8,5]. The two outer rings, namely the cytoplasmic and nuclear ring (CR and NR) together with the IR, give rise to the 3-stacked rings conformation of NPCs, with specific peripheral subcomplexes such as the mRNA export platform, the nuclear basket, and cytoplasmic filaments emanating from the outer rings[9,10].

In the following NPC-related sections of the Introduction, I will first describe the distinct NPC subcomplexes. Furthermore, I will elaborate on the modes of nucleocytoplasmic transport and possible implications in NPC conformation. Finally, in the integrative modeling-related sections

of the Introduction (section 3.2), I will provide an overview of the NPC models published before the work that is presented in this Thesis.

3.1.1 Nucleoporins: The building blocks of NPCs

In general, Nups exhibit large molecular weights ranging from 30-358 KDa and sequences of a few hundred up to thousands of amino acids[8]. Furthermore, they are roughly categorized into scaffold Nups, which contain mostly folded protein domains forming the scaffold "protein skeleton" of the NPC, and FG-Nups (Figure 1). The latter, include intrinsically disordered domains rich in FG-repeats that mediate interactions with the soluble nucleocytoplasmic transport machinery while also maintaining the diffusion barrier in the NPC channel[11,12]. However, this functional categorization of Nups might not be as clear-cut as it seems since many FG-Nups are also heavily involved in scaffolding functions by taking advantage of their structured domains and important interaction motifs[13].

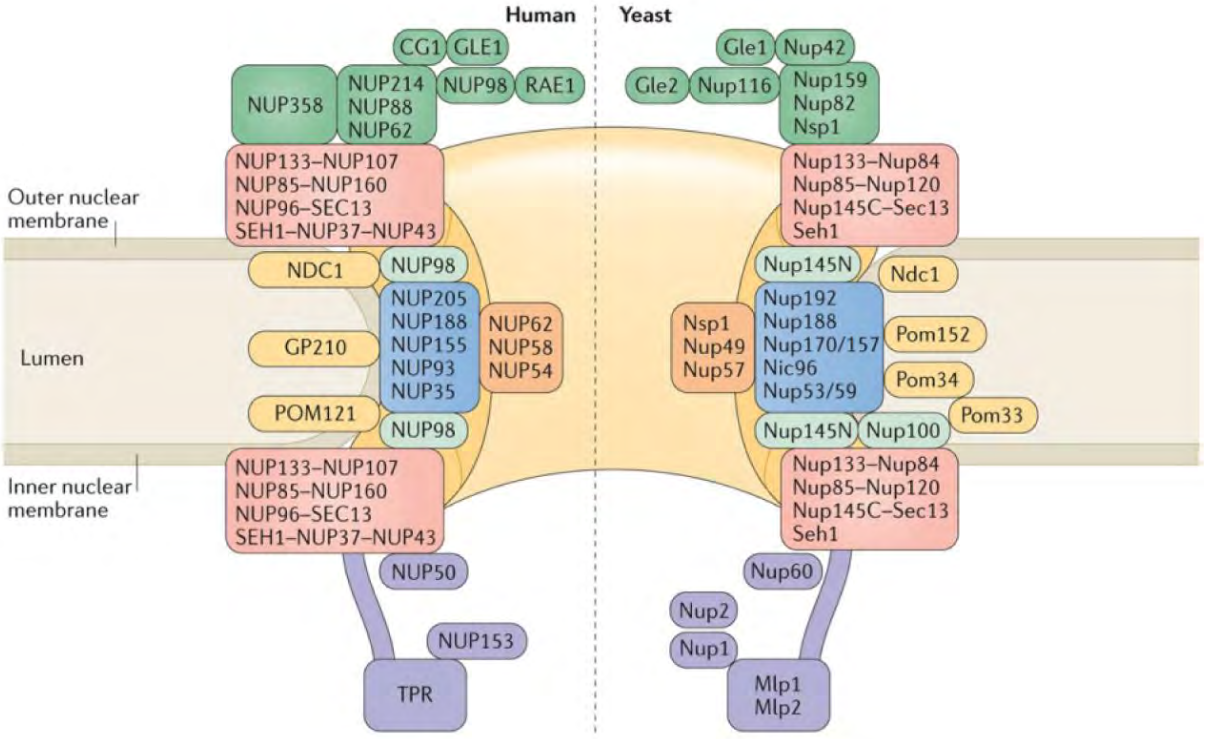


Figure 1: Conceptual localization and organization of conserved human and *S. cerevisiae* nucleoporins arranged in the structural framework of the human NPC. The different Nups are colored according to the NPC subcomplexes that they form: (blue) inner ring subcomplex, (orange) central channel complex, (yellow) transmembrane nucleoporins, (red) Y-complex, (green) cytoplasmic filaments and mRNA export platform, (violet) nuclear basket, (light green) disordered connector Nups. Figure adapted from[5].

Although sequence conservation of scaffold Nups across different species is poor, their structures are highly conserved[5]. Three main domain folds are the most abundant in scaffold Nups, namely: β-propeller domains (6–7 blade-shaped β-sheet subunits arranged around a central axis), elongated α-solenoid domains (antiparallel pairs of α-helices organized to form a coil) and coiled-coil domains (2–3 α-helices coiled and held together by hydrophobic interactions)[14,15,16]. While some Nups are comprised by a single and complete domain, e.g. a single β-propeller[17], others include combinations, e.g. N-terminal β-propeller followed by a carboxy-terminal α-solenoid fold[18]. The β-propeller/α-solenoid organization can be found in many membrane binding complexes such as the coat protein complex I and II (COPI and COPII) and clathrin[19]. Nups that include the aforementioned domain organization usually bind to membranes through α-helical amphipathic motifs in their β-propellers, called ArfGAP1 lipid packing sensor motifs (ALPS)[20,21]. Coiled-coils are probably the least abundant structural domains of scaffold Nups, with such domains residing in the cytoplasmic-facing side of the NPC as parts of the mRNA export apparatus[22], in the central channel as parts of the inner ring complex (IRC), and in the nuclear basket[23]. Furthermore, three transmembrane Nups (four in yeast) anchor the NPC scaffold to the NE[8].

In vertebrates, there are ∼10 different Nups with FG-repeat domains totaling ∼13 MDa of the NPC molecular weight and summing up to ∼5000 FG-repeat motifs[24,25]. Furthermore, the FG-repeats are found in different variations in vertebrate Nups including GLFG, FxFG, or PAFG amino acid signatures[26,27,28,29] which allow them to interact with the incoming/outcoming nuclear transport receptors (NTRs) with usually weak affinity[30]. However, FG-Nups not only differ in FG-motif constitution but also in their localization in the NPC scaffold, with both aspects shaping the functional role of these Nups. One of the most important functional roles of FG-Nups is the formation of a mesh-like entity in the center of the NPC which spans the central channel hence forming a selective diffusion barrier[31,32,33]. This barrier can easily be penetrated by NTRs with their cargo, thus ensuring the regulated nucleocytoplasmic transport. FG-Nups are also involved in mRNA export[22], chromatin remodeling, and cargo transport from the nuclear basket[34].

3.1.2 The architecture of scaffold NPC subcomplexes

The transmembrane “nature” and large size of NPCs posed formidable challenges in elucidating their structures. Despite that fact, it became apparent from early on that the NPC rings and the Nups that constitute them are organized into distinct subcomplexes[35,36,5]. Some of these subcomplexes are solely responsible for NPC scaffold support, while others are also involved in several NPC-specific functions such as nucleocytoplasmic transport, mRNA maturation/export, and chromatin organization.

The most thoroughly studied scaffold subcomplex to date is the so-called Y-complex (Nup107 complex in humans), named after its characteristic Y-shaped architecture. Y-complexes account for up to ~20% of the mass of the NPC[37] and their depletion abolishes the formation of NPCs[38]. These complexes adopt a head-to-tail arrangement and are the main building blocks of the NPC outer rings[8,5]. Across different eukaryotes, the homologous Y-complexes are assembled by varying

numbers of Nups. For example, the human Y-complex consists of ten different Nups whereas, the *S. cerevisiae* Y-complex includes only seven homologs. In both human and *S. cerevisiae* Y-complexes, a conserved vertex is formed by two arms, a long and a short one, followed by a longer structure, referred to as stem[39,40]. The long arm, also called Nup160 arm in humans (Nup120 in yeast), is pointing towards the outer periphery of the NPC while the short arm, also called Nup85-Seh1 complex arm, is pointing towards the central pore channel. The stem of the Y-complex, which consists of Nup133 and Nup107, as well as the long arm, bind the NE membrane through amphipathic helical motifs in β-propeller domains of Nup133 and Nup160 respectively[20,21] (Figure 2). In general, Y-complexes of vertebrates and some fungal species contain additional Nups such as Nup37 and ELYS which are binding the long arm[41,42] and Nup43 which binds to the short arm[43]. Sec13 and Nup96 (Nup145c in yeast) reside in the vertex of the Y-complex where the stem of the complex is bound[44].

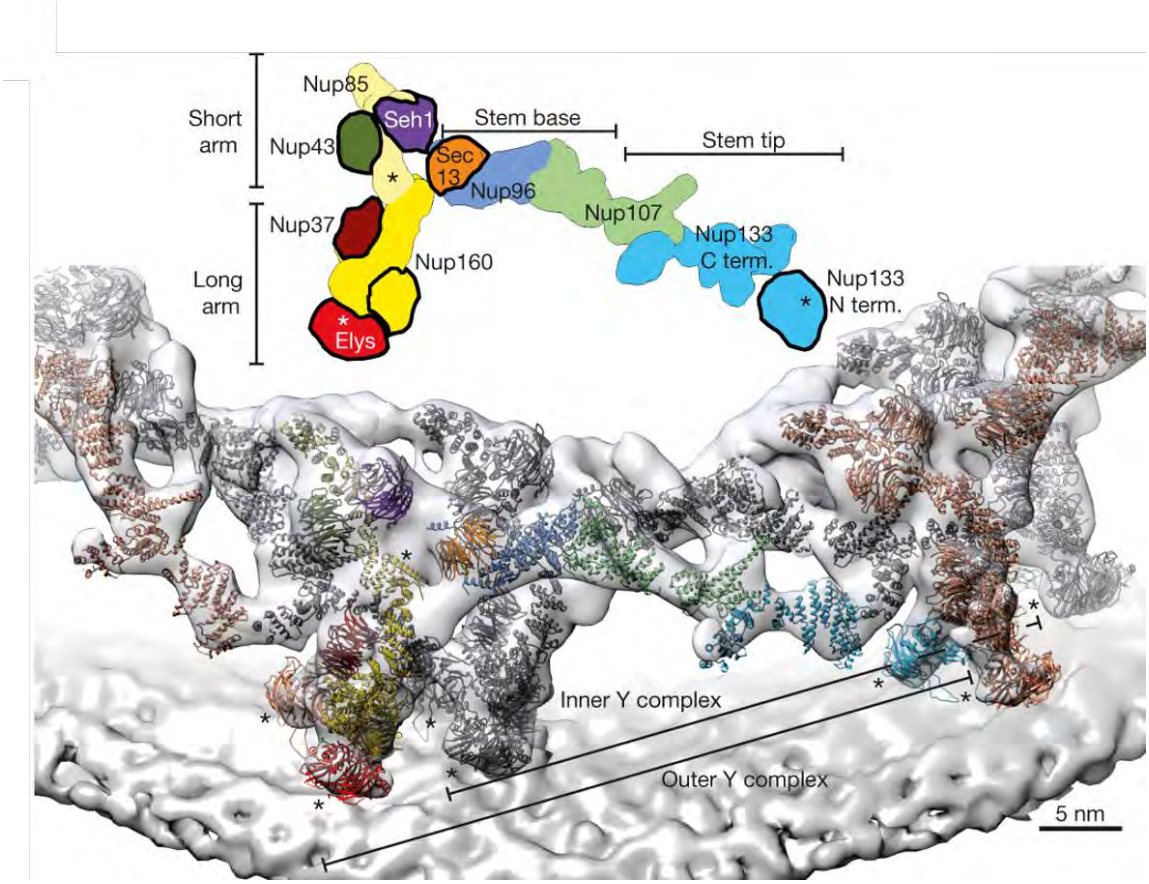


Figure 2: Overview of the human NR Y-complex architecture. (Top) Color-coded schematic representation of the Nup arrangement within the human NR Y-complex. (Bottom) Cartoon representation of double human NR Y-complexes, with an inner and an outer copy per asymmetric unit of the NPC. The structures of the Y-complexes are depicted inside the relevant NPC cryo-electron density[45] (grey density). The coloring of the Nups follows the same scheme as the representation above. The asterisks (*) denote Nup structures unambiguously positioned with respect to the overall Y-complex architecture, though their orientation remains ambiguous. Figure adapted from[45].

Another major scaffold NPC module that consists of two subcomplexes is the IR itself, with recent findings shedding more light on its overall structural architecture[46,47]. The subcomplexes that form a single IR asymmetric unit are the IRC and the central channel Nup62 subcomplex (Nsp1 subcomplex in yeast). There are four copies of each of these subcomplexes per spoke (i.e. per NPC asymmetric unit) resulting in a C2-symmetry as two of the copies of each subcomplex face the cytoplasmic side and two face the nuclear side. In both cytoplasmic and nuclear facing copies, the two-member copies exhibit an outer and inner copy arrangement with respect to the plane of the central pore channel (Figure 3). More specifically, four copies of Nup155 (Nup170 in yeast), of which two copies each are arranged in parallel and diagonally along the nuclear-cytoplasmic axis, tether the IRC to the INM-ONM of the NE. Additionally, four copies of Nup93 (Nic96 in yeast) also adopt a similar parallel diagonal conformation towards the central channel, while four copies of Nup188/Nup205 (Nup188/Nup192 in yeast) exhibit a parallel architecture residing "on top" of the Nup93 copies[8]. The heterotrimeric Nup62 (Nsp1 in yeast) - or central channel complex – is bound to the IRC and forms a triangulated triple coiled-coil present also in four copies. The three Nups constituting the Nup62 complex are Nup62 (Nsp1 in yeast), Nup54 (Nup57 in yeast), Nup58 (Nup49 in yeast). They are all FG-Nups with their FG-rich repeat domains spanning toward the central channel in order to form transient interactions with NTRs and regulate the selective diffusion barrier of the pore[48].

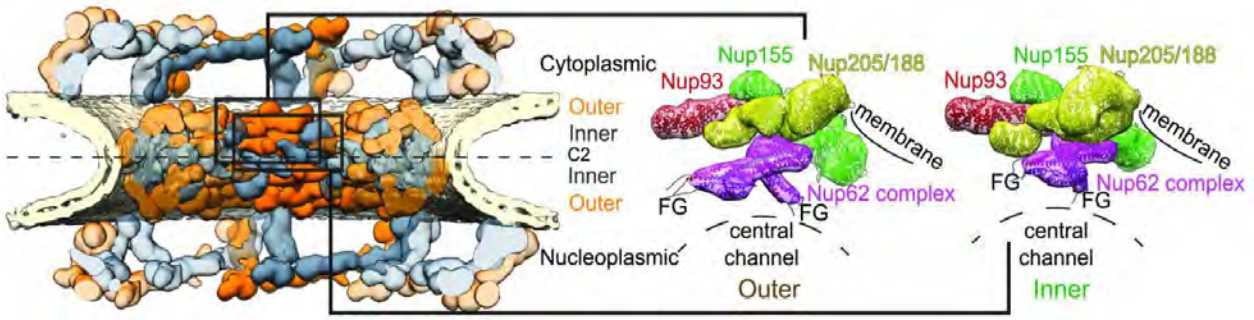


Figure 3: Dissection of the human IR. (Left) Overview of the human scaffold NPC including the densities of outer (orange) and inner (grey) Y-complexes as well as outer (orange) and inner (grey) copies of the IR core module which comprise the IRC. The densities are shown in the context of the relevant cryo-electron density[45]. (Right) Zoomed in view of the human IR core module configuration and comparison between the outer and inner copies. Notation of the Nup-coloring scheme is also provided. Figure adapted from[47].

3.1.3 Peripheral NPC modules

The NPC scaffold subcomplexes also serve as tethering/docking hubs for peripheral modules of the NPC. Such subcomplexes are the mRNA export platform and nuclear basket, which emanate from the CR and NR, respectively.

More specifically, the asymmetrically localized and conserved Nup214 (Nup159 in yeast) subcomplex, also known as P-complex due to its characteristic overall shape[22], binds the small

arm of cytoplasmic Y-complexes (i.e. Nup85–Seh1 Y-complex arm in yeast). This module is considered part of the cytoplasmic filaments[49] and its main constituents are Nup88 (Nup82 in yeast), Nup62 (Nsp1 in yeast) and Nup214 (Nup159 in yeast). These Nups can form many interactions with regulatory Nups and other proteins, making this subcomplex a key "player" in nucleocytoplasmic transport of proteins and mRNPs. To this date, there is no complete structural analysis of this complex due to the flexibility of its subunits and the presence of intrinsically disordered domains[22]. Therefore, only fragments of some relevant domains have been solved using X-ray crystallography[50,51] while the full P-complex was imaged solely by negative stain electron microscopy (EM)[52]. The three P-complex subunits interact with each other using their C-terminal domains (CTDs) forming coiled-coil interactions[53], presumably similar to homologous Nups from the central channel complex[54]. The most prominent Nup of the *S. cerevisiae* P-complex is Nup159 which is responsible for recruiting the DEAD-box RNA helicase Dbp5 (DDX19 in humans)-important P-complex interactor for the termination of messenger RNP export[55,56]. This recruitment occurs through interactions with the Nup159 N-terminus and its conserved β-propeller[57]. Additionally, in *S. cerevisiae*, Nup159 contains dynein-interacting-domains (DID) with which dynein light chain (Dyn2) homodimers interact in order to promote dimerization and stabilization of the P-complex by forming a dynein arm[52,58] (Figure 4).

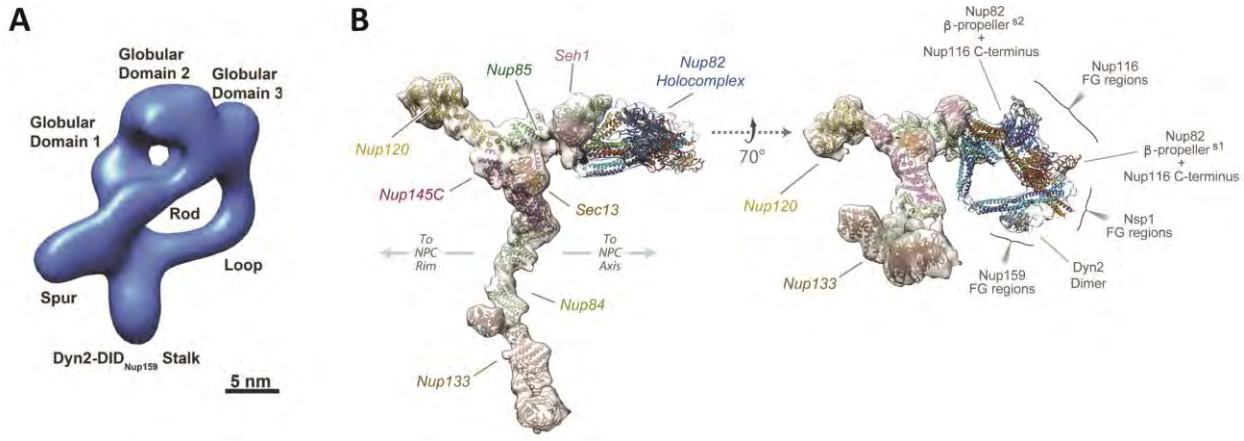


Figure 4: Structural architecture of the peripheral P-complex. (A) Visualization of a negative stain map of the *S. cerevisiae* P-complex (blue density) [52] with notations for the rough localization of domains from P-complex Nups. (B) Molecular representation of the *S. cerevisiae* Y-complex and P-complex (Nup82 holocomplex) models[22]. The structural core of the P-complex is formed by a heterotrimeric coiled-coil arrangement of Nup159 (Nup214 in humans), Nup82 (Nup88 in humans) and Nsp1 (Nup62 in humans) while the subcomplex appears anchored at the short arm of the cytoplasmic Y-complex and pointing towards the center of the pore channel. Each Y-complex and P-complex Nup is denoted. Figure adapted from[22].

The predominant module in the nucleoplasmic-facing side of the NPC is the nuclear basket, which was first observed by EM at the nuclear face of NPCs[59]. This assembly is formed by eight long and flexible filaments tethered to NR scaffold complexes and projects towards the inner nuclear side to form a distal ring-like structure[60]. The nuclear basket serves different roles in the context of the NPC, with the most important being nuclear transport and chromatin structure regulation[2,61].

In humans, the nuclear basket consists of three nucleoporins, Nup50, Nup153, and Tpr, homologs of which can be found across different eukaryotes[8]. *S. cerevisiae* carries two homologs of Tpr, namely Mlp1 and Mlp2, one homolog of Nup50, termed Nup2, and Nup1 and Nup60 which are functionally equivalent to Nup153 in humans[60,8]. Although some subunits of the nuclear basket are FG-Nups (e.g. Nup153), the main structural constituent of the nuclear basket filaments is the non FG-repeat containing Nup Tpr (Mlp1/Mlp2 in yeast)[62,63,64]. This Nup has a molecular weight of 270 KDa and it consists of a ~1600 residue coiled-coil-rich N-terminal domain. This domain is followed by an 800 residue C-terminal stretch which is highly acidic and considered to be mostly disordered[65]. Furthermore, Tpr associates with the NPC by nuclear pore-binding motifs in the N-terminus coiled-coil rich domain[66] and is recruited to the nuclear pore by Nup153. The latter Nup is responsible for the anchoring of other nuclear basket components to the NPC as well[67]. Despite the intensive structural and biochemical studies, the structure of Tpr has not yet been solved, with only the first ~140 residues of its N-terminal domain being experimentally determined by X-ray crystallography[65]. Although this structure revealed an antiparallel tetramer of parallel dimer coiled-coil domains, the exact homodimerization state and mechanism in vivo remain unknown as of today (Figure 5).

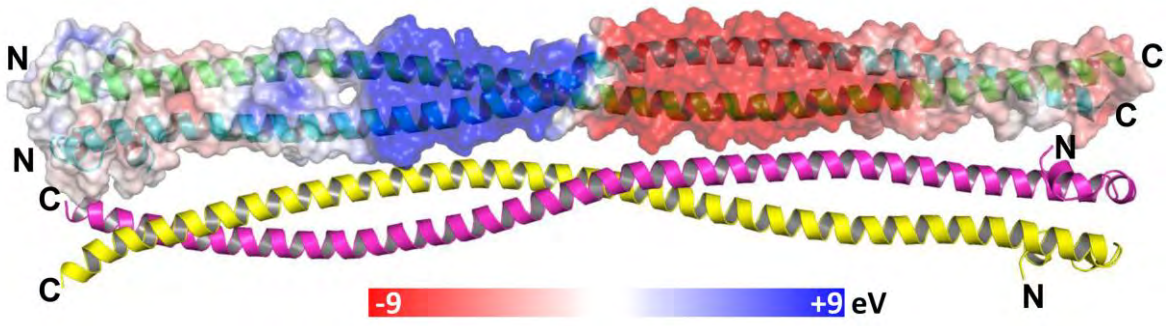


Figure 5: Molecular representation of the Tpr N-terminus. The N-terminus of human Tpr (PDB id: 5TO6) is shown in cartoon representation as an anti-parallel homotetrameric coiled-coil bundle. One of the two homodimers is also overlaid with calculated surface electrostatic potentials (blue for positive charge and red for negative charge) which showed a strong electrostatic dipole in this domain[65]. Figure adapted from[65].

Furthermore, an NPC component that is separated by the nuclear membrane from all other NPC modules is the luminal ring (LR). This assembly is localized in the lumen of the NE surrounding the formed NPC at the plane of the fused NE membranes[68,69]. The transmembrane nucleoporins that belong (or form) to the LR are termed pore membrane proteins (POMs) and are amongst the least conserved and structurally analyzed Nups[8]. The major LR subunit is the glycoprotein POM210 (also known as GP210), which can be found across the whole eukaryotic kingdom (also known as POM152 in yeast)[70]. *S. cerevisiae* POM152 has an N-terminal NPC-associating region proximally to the IR followed by a single transmembrane region while in GP210 these domains are located at the C-terminus (i.e. reversed domain organization)[71,72]. Additionally, POM152 contains a large luminal domain which consists of at least eight repeated immunoglobulin-(Ig) like domains with a head-to-tail arrangement, with the exact number of Ig-like β-sandwich domain

repeats and architecture being unknown to date[73]. Currently, the role of LR and more specifically of GP210 in proper NPC assembly and cell viability in vertebrates[74] is still under investigation, with POM152 knockout experiments in *S. cerevisiae* showing proper formation of NPCs[75].

Finally, recent biochemical studies revealed that many of the NPC subcomplexes that were described earlier are capable of forming inter-subcomplex interactions (both scaffold and peripheral complexes) with the result being several distinct supercomplexes[4,76]. Interactions of this kind are mediated by flexible rope-like linker domains contained in scaffold and FG-Nups[47,46]. These Nup regions include short unstructured amino acid stretches termed short linear motifs (SLiMs) and they are great binding candidates of structured Nups. Some important examples of SLiM-containing Nups are Nic96, which tethers the central channel complex to the IRC via a SLiM in its unstructured domain, and Nup145n (two homologs in *S. cerevisiae*, Nup116 and Nup100), which has been shown to interact with either Nup145c or Nup82 in a mutually exclusive way through its C-terminal autoproteolytic domain[50].

3.1.4 Nucleocytoplasmic transport and NPC constriction

As mentioned earlier, NPCs are the main facilitators of nucleocytoplasmic transport. These transmembrane assemblies allow for both passive diffusion of ions and small molecules and active transport of macromolecules in a bi-directional manner between the nuclear and cytoplasmic compartments[49]. As a general rule, molecules of 30–40 KDa are freely diffused across the NPC central channel with relatively reasonable diffusion times. For larger molecules, the translocation from/to cytoplasm to/from the nucleus is mediated by cargo-carrier proteins called karyopherins or NTRs[77], which interact with the FG-repeats for rapid cargo translocation[5,78]. This procedure is usually, but not always, coupled with energy cues from the nuclear Ran GTPase[79].

This small Ran GTPase is responsible for maintaining a GTP-GDP gradient across the NE environments by actively shuttling through the NPC. Hence at the nuclear side, Ran is loaded with GTP, the hydrolysis of which will happen consequently in the cytoplasmic side upon the translocation of Ran, NTRs and the respective cargo. NTRs, which can be classified as importins and exportins, across both sides of the NE recognize and interact with cargoes based on nuclear localization signals (NLS) and nuclear export signals (NES). The GTP/GDP gradient together with the Ran GTPase interactions with the NTRs and FG-repeats of the NPC are establishing the directional and usually energy-dependent active nuclear transport through the NPCs[77].

On a different note, it has already been suggested that NPCs might be able to undergo conformational changes in order to efficiently accommodate the transport of large cargoes or as a response to mechanical stress[8,34]. These changes usually result in significant increase or reduction of the NPC diameter[80,81]. More specifically, in vitro studies and super-resolution imaging performed on *Xenopus* oocytes revealed the constriction of NPCs during different developmental stages coupled with decreased nuclear transport[82]. Further evidence regarding the unidentified link between active nuclear transport and NPC dilation were provided by studies capturing the structural architecture of the actively transporting NPC within its intact cellular environment[83,43].

From the latter studies, it became evident that the NPC in cellulo exhibits a larger diameter compared to the structures captured of non-transporting NPCs from semi-purified NE. Taken together, the findings from the aforementioned studies suggest high flexibility of NPCs in their native cellular context with the ability to adopt alternative conformations based on environmental cues and active transport state. However, experimental evidence regarding the NPC dilation or constriction directly in response to specific mechanical stresses, as well as the underlying mechanisms of conformational change are still missing.

3.2 Integrative structural modeling of protein assemblies

Most cellular processes are facilitated by proteins which tend to organize into macromolecular assemblies. Therefore, elucidating the structure of such protein complexes is often the first step towards the understanding of their mechanistic function.

Since the determination of the first protein structure back in the 1950s, two experimental methods have "monopolized" the interest of structural biologists for such applications. These methods are X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. Both techniques, helped the advancement of the structural biology field tremendously by helping to resolve structures at atomic and/or near-atomic resolution. As a result, most of the structures in the relevant databases were determined by these methods. Though recent advances in the experimental setup, data acquisition and analysis in cryo-electron microscopy (cryo-EM) and especially in cryo-electron tomography (cryo-ET) have placed this method among the most utilized techniques together with the two aforementioned methods for structure determination nowadays, especially for studying large protein assemblies[84,85,86,87,88].

Over the last few years, a rapidly increasing number of protein complex structures of various sizes and flexibility was determined by integrative/hybrid computational modeling methods. These methods are information-driven and utilize data arising from diverse experimental and/or in silico approaches such as X-ray crystallography, EM, NMR, small-angle X-ray scattering (SAXS), cross-linking mass spectrometry, homology models and more[89,90,91]. By combining datasets of high and low structural resolution, computational integrative modeling pipelines are capable of producing more accurate and precise structural models of large protein complexes[89,92,93,94] even at in-cell conditions[95,96,97] compared to any other structure determination method when used individually. However, the existing integrative modeling software and approaches are either limited to modeling of protein complexes of simple architectures or require custom modifications, and thus necessitate a major expertise when applied to more complex systems.

In the following sections of this Introduction, I will describe the fundamental principles of integrative structural modeling approaches and provide a brief comparison of existing integrative models of NPCs.

3.2.1 Fundamentals of integrative structural modeling methods

Integrative structure determination approaches have been applied for many decades in the field of structural biology. Though, it was only recently that major advances in commonly used experimental methods as well as hardware and software-related innovations gave rise to computational protocols for integrative structural modeling in an automated fashion.

Such modeling approaches aim to utilize every available information regarding an individual protein or a protein assembly in an attempt to increase the final model accuracy and precision[90, 91,92,98,99]. These methods are dependent on complementary datasets arising from various

experimental and computational protocols with the main sources being X-ray crystallography, NMR spectroscopy, cryo-EM, and chemical cross-linking with mass spectrometry[89,90]. The datasets produced by these methods are used as input during the integrative modeling in order for spatial restraints to be derived from them with the aim of producing structural models with properties that agree with the provided information as much as possible[90].

Every integrative structural modeling protocol is an iterative procedure composed of four stages (Fig. 6): 1) gathering of structural information (datasets and bibliographical knowledge) for the protein targets of interest, 2) creating representation of the protein system to be modeled as well as defining a scoring function, 3) sampling of models in the respective conformational space and finally 4) validating the resulting models. An optional but strongly recommended step upon the successful application of an integrative modeling procedure is archiving of resulting model or ensemble of models in PDB-Dev[100,101,102], a standalone archiving, validating and curating system for structures that have been determined or predicted by integrative modeling applications.

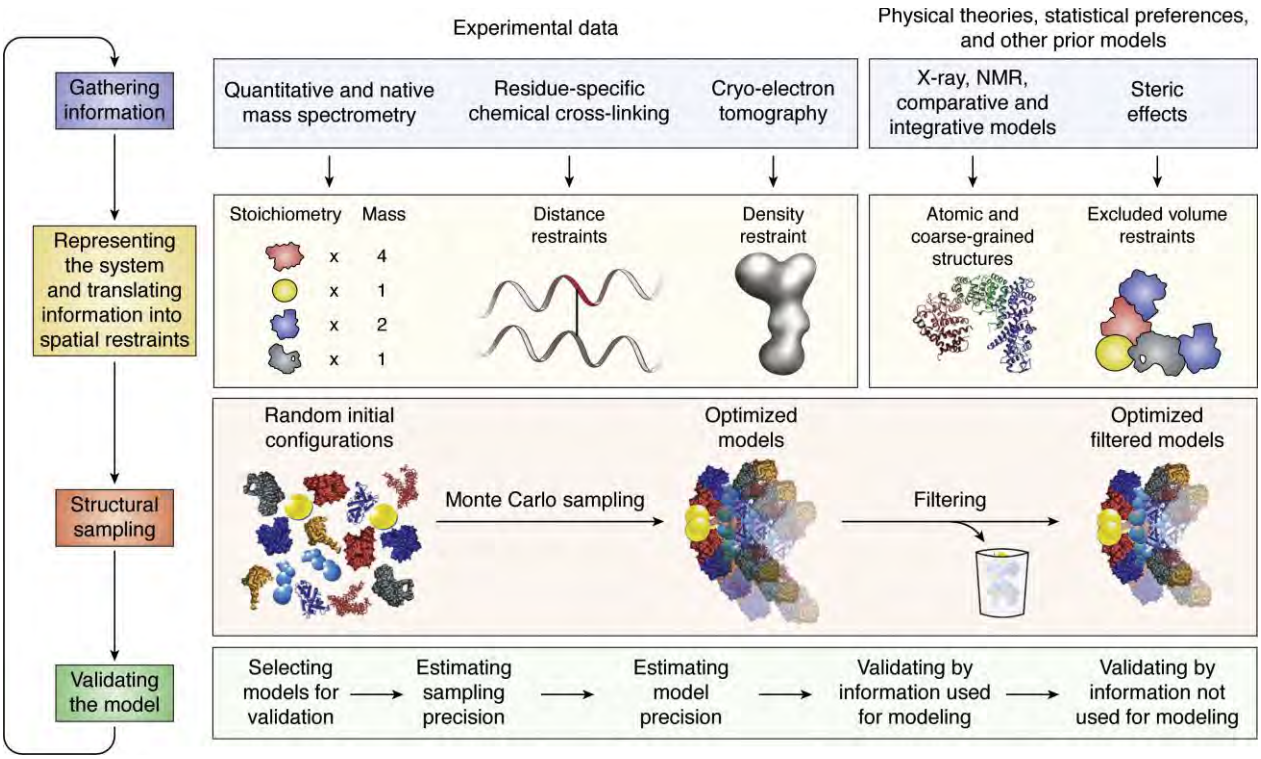


Figure 6: Schematic description of a typical integrative modeling workflow. (Left) Depiction of a typical integrative modeling workflow consisting of the four fundamental stages, i.e. gathering of information, representation of the system, model sampling and model validation. (Middle and right) Examples of different data sources that will be utilized as input during integrative modeling of protein complexes, alternative system representations and translations of the input datasets to distance restraints, sampling and filtering methods of models from final ensembles, model validation steps including the calculation of sampling exhaustiveness and precision. Figure adapted from[90].

During the first stage of an integrative modeling protocol, available information regarding the protein system to be modeled must be gathered, which are usually in the form of datasets. In the case of modeling of protein complexes, if there are no experimental structural models of individual domains or subunits of the complex, then homology models or models predicted by other methods are utilized as input. Additionally, any prior bibliographical information regarding intra-complex interactions, stoichiometry or domain organization that are not included in the gathered datasets can be gathered as well under the condition that they can be converted to some type of distance restraints during the modeling.

The gathered input information as well as their uncertainty will be used to define the representation of the protein system before modeling. Hence, the protein system will be represented in a coarse-grained (e.g. rigid or flexible beads per residue) or even atomistic manner with the copy number of each protein or individual domain dictated by the input information. Other than informed decisions regarding the proper representation of the protein system, the input information will also guide the definition of a scoring function based on which each model will receive a numeric score upon sampling[90]. Very often, the scoring function is a simple weighted linear sum of individual scoring terms which represent the spatial restraints derived from the input datasets. Each spatial restraint violation is basically a quantified deviation of the computed property of a model from its experimental or predicted measurement. Therefore, it is safe to assume that the scoring function is in other words the quantification of the degree of "match" between a tested model and input data.

The next major step for integrative modeling after the gathering of all input data and the selection of system representation is the sampling of conformational space. Due to computational hardware and protocol limitations, sampling of the total 3D-space for an integrative model is impossible. Therefore, the selection of proper sampling approaches/algorithms for efficient "exploration" of the conformational landscape, such as Markov Chain Monte Carlo sampling[103], is of crucial importance in integrative modeling approaches[90]. Therefore, it quickly becomes apparent that the implementation of stochastic algorithms for minimizing the score from the aforementioned scoring function (i.e. minimizing the violation of spatial restraints) is of crucial importance.

As mentioned earlier, such integrative modeling methods are usually applied to model structures of either large individual proteins or protein assemblies, the modeling of which is based on "low-to-mid" resolution (or low-to-mid confidence) input data. Therefore, the final outcome is not a single model rather an ensemble of models satisfying the spatial restraints to varying degrees. A filtering step is often followed prior to the actual model validation which is the last stage during an integrative modeling protocol. The filtering of models from the resulting ensemble can be assisted and tested against a subset of the available input information that were not used during modeling[104]. This filtering approach is particularly useful in modeling cases where the usage of certain datasets is computationally very expensive during the actual modeling.

Upon filtering of the resulting integrative models, the final model validation is proceeding. As in every predictive computational method, this stage will assist in the evaluation of the overall modeling and sampling performance in order to avoid overinterpretation of the output models based on the input data[105]. In order to quantify the modeling performance, the sampling and individual model precision need to be determined in a similar fashion as described by[99] (Fig. 7).

In short, the sampling precision is defined as the clustering threshold at which the sampling of models is exhaustive according to some quantitative criteria (Fig. 8). The model precision is defined by the variability among the final acceptable models given that sufficient sampling was performed. There are various tests that can be applied to the output model ensemble, both score-based and structure-based in order to assess the sampling and model precision[99]. Though, due to the uncertainty and alternative sources of data, there are currently no "golden standards" for objectively quantifying the exhaustiveness of integrative modeling methods, unlike more established experimental methods such as X-ray crystallography. Finally, the described four-stage procedure for integrative modeling can be repeated until a sufficient number of final models satisfying the spatial restraints is retrieved.

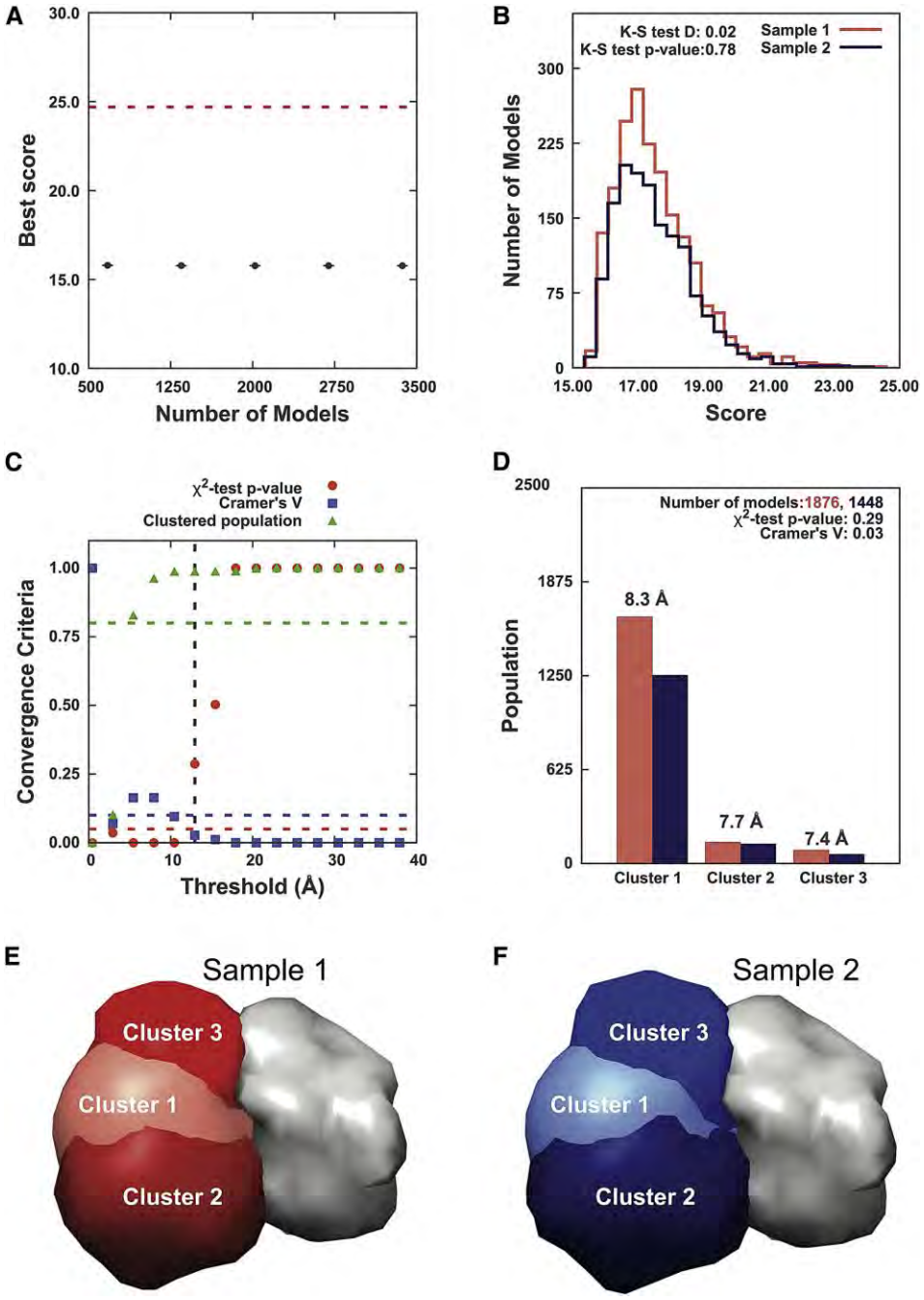


Figure 7: Example results from sampling exhaustiveness assessment for PDB: 1AVX. (A) Example plot for first score-based test assessing the score convergence of models in the final ensemble. (B) Example plot for second score-based test assessing the similarity of score distributions between two randomly split samples of models from final ensemble. (C) Example plot for first structure-based test assessing sampling precision based on three statistical criteria. (D) Example plot depicting the populations of models from the two randomly split samples upon clustering and individual cluster precision. (E & F) Depiction of localization probability densities for models from sample 1 (red) and sample 2 (blue) for visual inspection and comparison as part of the second structure-based test. The grey density (depicted in both E & F) belongs to the bound receptor which was kept rigid during integrative modeling[106]. Figure adapted from[99].

Over the recent years, different modeling software and computational protocols were developed capable of performing integrative structural modeling. Some of the most widely used software for such modeling tasks are UCSF Chimera[107], Situs[108], iMODFIT[109] and MultiFit[110]. Additionally, great modeling successes were achieved by computational protocols that were developed based on the integrative modeling libraries: Integrative Modeling Platform (IMP)[111,112], HADDOCK[113], and Rosetta[114]. Despite their achievements, all these integrative modeling solutions exhibit pros and cons in terms of the possible modeling targets and input information that can be handled.

The state-of-the-art computational package from the aforementioned modeling approaches is considered to be IMP[111,112], which has already been successfully applied for the modeling of large protein assemblies such as the NPC[115], Elongator complex[116], 26S proteasome[117] and expressome[96]. This library is an open-source computational package that allows users to define custom spatial restraints from datasets of many different structural and biochemical methods, or use the predefined restraint types, select or define multiple system representations, define custom scoring functions and sampling methods while also offering modeling output analysis tools[98,99].

IMP is actually at the lowest level of a multilayered platform offering a variety of "components" for the compilation of custom modeling protocols by expert structural bioinformaticians using the C++ and Python programming languages (Fig. 8). Furthermore, a higher-level interface called Python Modeling Interface (PMI)[118] offers significant simplifications for the compilation of modeling protocols using IMP. Technically, PMI is a top-down system with many predefined macros for selecting system representations, defining scoring functions etc., all with using the Python language. The simplified "nature" of PMI comes with the cost of reduced flexibility of IMP for the sake of ease-of-use[118]. Moreover, at even higher abstraction levels than PMI, inexperienced users are able to utilize several IMP and PMI functionalities through user-friendly applications dedicated to specific modeling tasks and aspects[112].

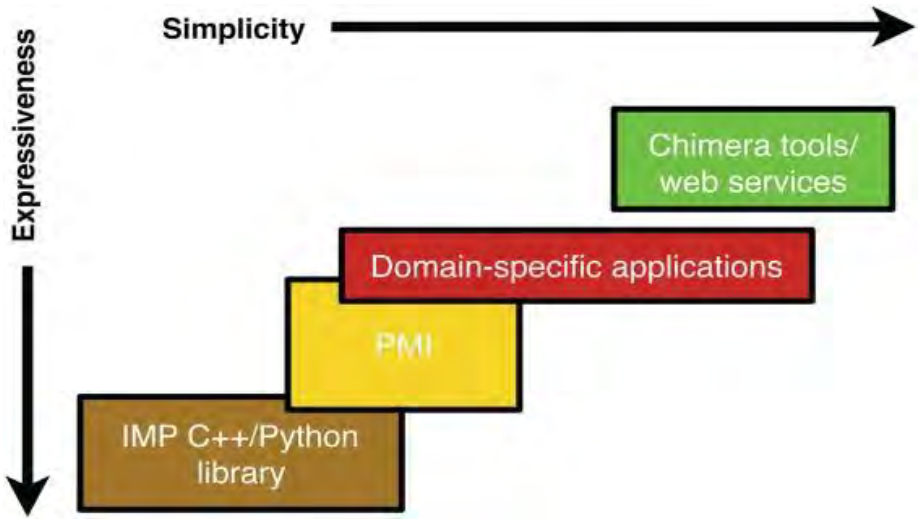


Figure 8: Graphical summary of the multilayered organization of IMP. IMP lies at the lowest abstraction level of the multilayered suite. At the very next abstraction level "comes" PMI with simplified IMP functionalities and access. Finally, at the highest abstraction levels online-services and tools with limited access to IMP functionalities are found (task-specific applications). Figure adapted from[112].

3.2.2 Integrative approaches for the modeling of NPCs across species

Tentative structural and biochemical research that was conducted on NPCs over the last decade revealed that despite the commonalities across various eukaryotes, such as the overall symmetrical architecture and existence of homologous subcomplexes, these protein assemblies exhibit striking species-dependent differences[5]. Some of the differences that have been observed are the alternative stoichiometries of Nups and different subcomplex architectures, especially between evolutionarily distant species. Such differences became quickly apparent from the first integrative structural models of the most studied NPCs from human, *S. cerevisiae* and *C. reinhardtii*.

More specifically, the current integrative structural model of the human NPC[45,47] was built based on high resolution structures from X-ray crystallography, cryo-tomographic maps, cross-links from mass spectrometry and homology models. This model confirmed the highly conserved architecture of the IR with the typical 8-fold rotational symmetry sandwiched between the also 8-fold rotationally symmetric outer rings. Furthermore, both NR and CR appeared connected to the IR by 16 copies of the scaffold nucleoporin Nup155. Both outer rings contain 16 copies each of the Y-complex arranged as two concentric rings with head-to-tail arrangement of 8 copies (Fig. 9).

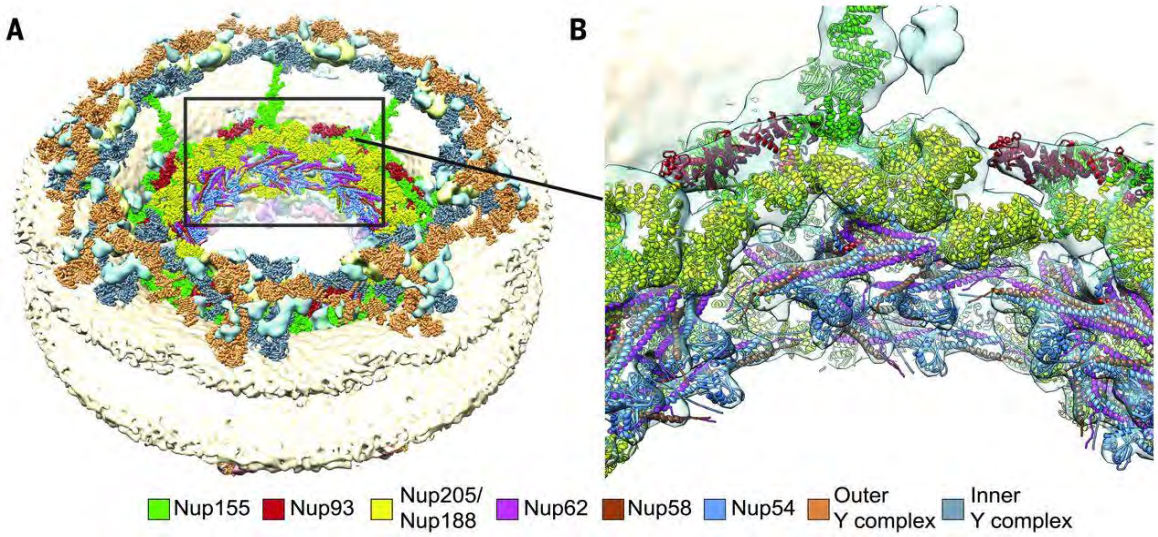


Figure 9: Depiction of the composite human NPC model. (A) Overview of the integrative human NPC model shown together with the unassigned protein densities (cyan). (B) Zoomed-in view of the human NPC IR with the high-resolution structures that were used for the production of the integrative model shown as cartoon ribbons inside their corresponding density (light cyan). The color-coding for specific Nups and subcomplexes is also provided. Figure adapted from[47].

Another NPC architecture that was established with an integrative approach came from the organism *C. reinhardtii*[43]. The production of this low-resolution model was based on cryo-ET maps (~30 Å), multiple sequence alignments for the detection of remote Nup homologs and

homologous structures from human Nup subcomplexes. For this model, the integrative modeling procedure that was applied enabled the density allocation (inside the relevant cryo-ET map) of all major scaffold subcomplexes from the human NPC. The IR architecture as dictated by the *C. reinhardtii* NPC (CrNPC) model appeared similar to the human IR conformation confirming once again the high degree of conservation of this part of the NPC. The NR also revealed a familiar double Y-complex arrangement similar to the human NR with probably additional copies of Nup188/Nup205 (as in human NPC) enclosed between the two concentric Y-complex rings. A striking difference arose in the CR side of the CrNPC which is the existence of only 8 Y-complex copies in this outer ring. Additionally, another observable difference that was highlighted between this algal NPC and its human counterpart was the dilated pore and more flattened conformation of subcomplexes comprising the outer rings.

Until recently, the most complete model of the NPC from any species was the integrative structural model of the *S. cerevisiae* NPC (ScNPC)[46]. The model was built mostly based on cryo-ET maps from detergent-extracted ScNPCs (~28 Å resolution), high resolution structures from X-ray crystallography, prior integrative models of NPC subcomplexes[21,22], homology models of individual Nups, cross-links from mass spectrometry[119] and SAXS profiles. The IR architecture of the ScNPC appeared identical to the human IR, confirming the high degree of conservation of subcomplexes residing at the IR plane. On the contrary, both CR and NR seemed to include only 8 copies of the Y-complex each, which accounts for a striking difference compared to the equivalent dimerized human outer rings subcomplexes (Fig. 10). Furthermore, the large repertoire of diverse datasets that were utilized for the production of this integrative model enabled the placing of models from peripheral ScNPC subcomplexes, such as the P-complex and the luminal ring Pom152, as well as a mesh network of FG-repeats.

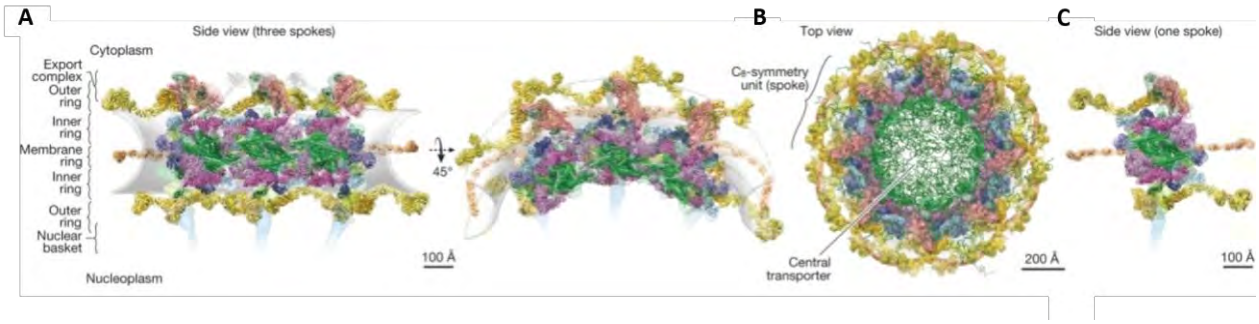


Figure 10: Overview of the integrative ScNPC model. (A) Two cross-section views of three asymmetric ScNPC units depicting the overall organization of Nup subcomplexes which are shown in cartoon representation inside the localization probability densities (colored densities) as determined by the integrative modeling procedure[46]. (B) Cytoplasmic view of the modeled ScNPC showing the same Nup subcomplexes as (A) including modeled FG-repeat Nup regions. (C) Side view of a single ScNPC asymmetric unit. Figure adapted from[46].

Furthermore, the ScNPC model elucidated the significantly smaller overall dimensions of the imaged NPCs compared to the algal NPC. Though, it was not clear whether this difference is a species-specific difference or if the NPC changes its conformation upon purification. In situ or in cellulo-based structural studies regarding the NPC architecture could provide further insights into

the observed differences and possibly validate the physiological relevance of the aforementioned NPC models.

4. Research objectives and motivation

The determination of a protein's structure is an important step for understanding how these macromolecules function either individually or in the context of complexes. Historically, two experimental methods have been widely used for protein structure determination, namely X-ray crystallography and NMR spectroscopy. In the recent past, major advances in instrumentation, data acquisition and data processing in cryo-EM have placed this method as well in the main "arsenal" of structural biologists, especially for the study of large and complex protein assemblies. All these cryo-EM advances coincided with crucial developments in computational hardware and software hence giving rise to integrative computational methods for the elucidation of complex protein architectures. The integrative modeling methods are data-driven and rely heavily on diverse input information with the goal of increasing the accuracy, precision and completeness of the final protein model. There is already a number of integrative structural modeling tools publicly available that were successfully applied to various protein targets. Despite that fact, the vast majority of such tools either lack versatility in terms of input information that can be used for the modeling or demand a high computational expertise level from the users, thus hindering their broad adoption by structural biologists.

Until recently, some of the most prominent applications of integrative modeling were the structural elucidations of the human and *S. cerevisiae* NPC architecture. The NPCs are amongst the largest and most studied protein assemblies in eukaryotes due to their role as the main facilitators of nucleo-cytoplasmic transport. Both these models were able to provide crucial insight into the structural architecture of NPCs, with the *S. cerevisiae* model being considered a complete "depiction" of the ScNPC. Multiple and diverse datasets were utilized for the production of the ScNPC model with cryo-EM maps from detergent-extracted nuclear pores, influencing the most the integrative modeling approach. Therefore, despite the fact that the ScNPC model included all known Nup structures, the physiological relevance of this model remained unclear. Additionally, although a few integrative NPC models have been put forward in recent years, most of the model organisms, such as *Schizosaccharomyces pombe*, are still lacking a structural nuclear pore model.

In my PhD Thesis, I aim to address the following questions:

- How feasible is the development and application of an integrative modeling approach that can be applied to any large protein complex target with minimal user effort?

- How does the in-cell architecture of the *S. cerevisiae* NPC compare to the previous model that was based on detergent-extracted NPCs?

- What is the *S. pombe* NPC scaffold architecture within intact cells and what is the structural response of such complexes to cellular cues like energy depletion?

5. Synopsis of listed publications

As mentioned in Introduction (see section 3.1), despite the commonalities of NPCs in their overall architecture, they exhibit striking differences in terms of Nup subcomplex organization and relevant stoichiometries. Additionaly, despite the extensive research that was conducted on these complexes, the exact mechanism of nucleocytoplasmic transport as well as the detailed steps for the assembly of the NPCs remain unclear.

In the following sections of the Synopsis (section 5), I will describe how my integrative modeling efforts, as part of large collaborative projects, contributed to the structural elucidation of NPC architectures from different yeast species at unprecedented resolution. The integrative NPC models that I produced in this context were based mostly on in-cell datasets and therefore, among other results, provided answers regarding the native NPC architecture and how it can be affected under different cellular conditions. Finally, I will present the integrative modeling tool that I co-developed and applied for the production of the NPC models, called Assembline and briefly mention the different modeling modes included which will enable structural biologists to "tackle" their own complex modeling tasks.

5.1 PUBLICATION I: "In-cell architecture of the nuclear pore and snapshots of its turnover"

Despite the exhaustive structural research that has been conducted regarding the NPC architecture and function across different species, it was only recently that advances in experimental and computational methods gave rise to the first structural models of NPC scaffolds and peripheral subcomplexes. The nuclear pores are large and quite heterogeneous transmembrane assemblies. This fact poses challenges for their structural elucidation and many diverse datasets need to be combined to model NPCs. Among the methods that played a crucial role in modeling NPCs have been cryo-electron tomography (cryo-ET) and integrative structural modeling (see Introduction section 3).

The previous status quo regarding the architecture of *S. cerevisiae* NPC was the integrative model produced by[46]. Although this model contains most of the known *S. cerevisiae* Nups (ScNups), it was substantially influenced by cryo-EM and cross-linking datasets obtained from purified NPCs (see Introduction section), hence raising questions regarding its physiological relevance in the native cellular context.

Matteo Allegretti and Christian E. Zimmerli from Martin Beck's group (EMBL Heidelberg, MPI Frankfurt) obtained cryo-ET maps of ScNPCs, resolved at ~25 Å, which I received as part of our collaboration (in the following sections these maps are referred to as in-cell ScNPC maps). By using our integrative modeling pipeline that was bundled as a Python-based package called Assembline[120] (see Synopsis section 5.3) together with these in-cell maps, already published X-

ray crystallography structures and homology models as structural inputs, I managed to produce the first in-cell model of the ScNPC[95].

In the following sections, I will describe the integrative modeling of the in-cell ScNPCs. Compared with the previous in vitro model by[46], my models revealed striking structural differences in the structure of Y-complexes in the outer rings and in the localization and orientation of the mRNA export platform machinery in the CR.

5.1.1 Structural architecture of the *S. cerevisiae* NPC scaffold

The in-cell cryo-ET maps of wild type ScNPCs at a resolution of ~25 Å that I received from our collaborators revealed that in cells the ScNPCs exhibit ~100 nm diameter between the two INM and ONM membrane fusion points and a central channel diameter of ~63 nm with the height of the entire NPC (i.e., from outermost outer ring densities in the maps) being 56 nm. Compared with the detergent-extracted ScNPCs by[46], the in-cell ScNPCs appear wider by ~20 nm (Fig. 11).

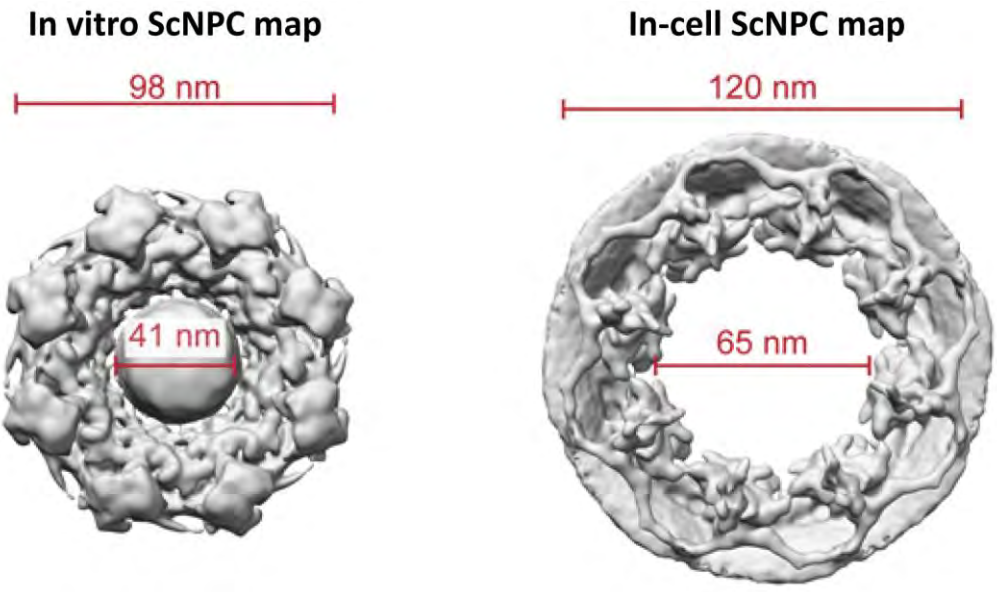


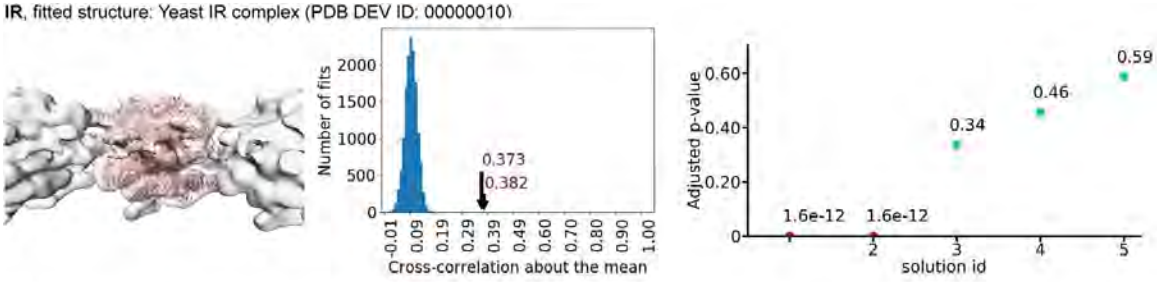
Figure 11: Comparative top view of the in vitro and in-cell ScNPC maps. (Left) Top view of the detergent-extracted ScNPC cryo-ET map at ~28Å[46]. (Right) Top view of the in-cell cryo-ET ScNPC map at ~25Å (from our study[95]). Pore diameters are indicated in red.

To identify locations, orientations, and stoichiometry of the main scaffold subcomplexes of the ScNPC in the received cryo-ET maps, I first used Assembline[120] to systematically fit (see Synopsis section 5.3) the integrative model of the ScNPC IR asymmetric unit from[46] and its sub-complexes to the in-cell maps. The complete model of the IR unit fitted with significant p-values (p-value = $1.6*10^{-12}$) with the top resulting fitted conformations explaining the main density observed

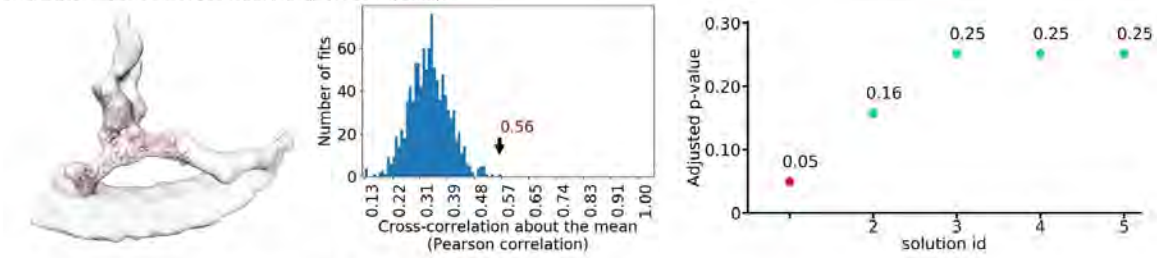
between the outer rings (top panel in Fig. 12). At this point it has to be mentioned that the second best IR unit fit exhibited almost identical fitting scores and p-values with the top best fit as they differ by 180º of rotation (internal symmetries of IR unit as described in Introduction section 3.1). Given the information of the allocated IR unit density in the in-cell ScNPC map, the refinement of all eight copies of the IR unit was performed with Assembline (see Synopsis section 5.3) in order to produce a model that would compensate for the dilated by ~20 nm in-cell NPCs.

On a similar note, I systematically fitted with Assembline the integrative models and the individual rigid bodies (i.e. predefined atoms that are constrained to change coordinates all together as a group) comprising them (e.g. the crystal structure with PDB ID 4XMM by[54], of CR and NR Y-complexes produced by[46]). For both CR and NR Y-complex cases, many of the structures were fitted with significant p-values, a fact that allowed me to unambiguously allocate the densities of these complexes inside our in-cell ScNPC maps (Fig. 12).

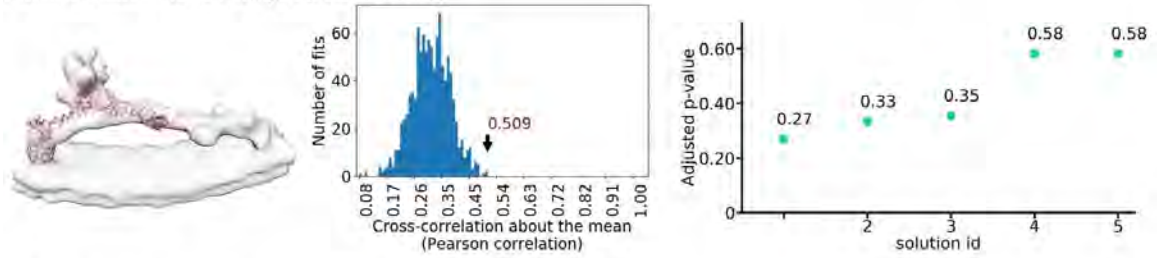
IR, fitted structure: Yeast IR complex (PDB DEV ID: 00000010)



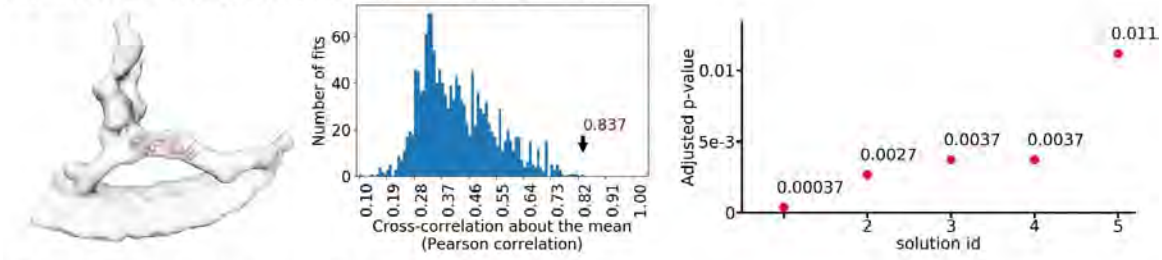
CR, fitted structure: Yeast Y complex (PDB ID: 4XMM)



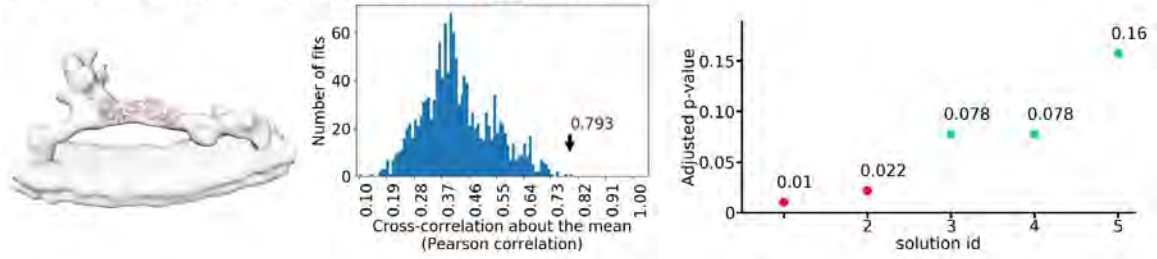
NR, fitted structure: Yeast Y complex (PDB ID: 4XMM)



CR, fitted structure: Yeast Sec13-Nup145C-Nup84 complex (PDB ID: 3IKO)



NR, fitted structure: Yeast Sec13-Nup145C-Nup84 complex (PDB ID: 3IKO)



NR, fitted structure: Yeast Nup84-Nup133 complex (homology model, Fernandez-Martinez, 2016)

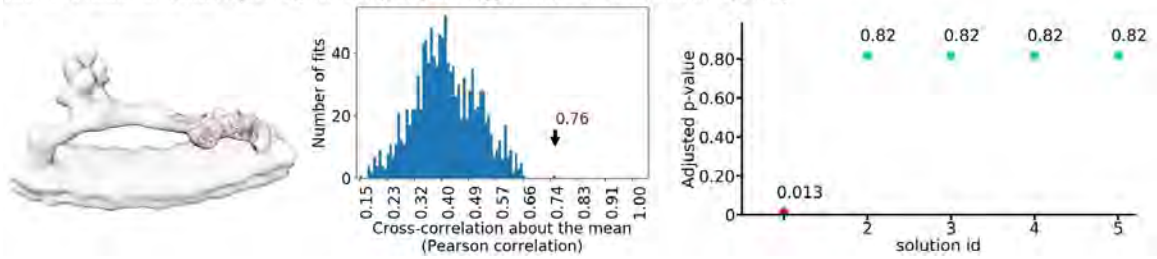


Figure 12: Systematic fitting of inner and outer ring components into the in-cell *S. cerevisiae* NPC map. Each row shows the visualization of the top fits (left), the histogram of raw scores (middle), and a plot of the top five p-values (right). In the p-value plots, the statistically significant fits are colored red (p-value < 0.05). The top fits are indicated in the histograms with an arrow and the score value. The number of sampled fits used to calculate p-values after clustering of similar solutions was 14,348, 1,015, 1,039, 1,479, 1,354, and 1,183 for the rows from top to bottom. For the IR, the integrative model of the single spoke of the IR[46] was used as the fitted structure. For the outer rings (CR and NR), the crystal structure of the yeast Y-complex (PDB id 4XMM) was fitted or its parts corresponding to subcomplexes. Figure adapted from[95].

Based on these systematic fitting results, I confirmed what was also observed in the in vitro model, that eight symmetrical copies of Y-complexes are present in each of the ScNPC outer rings (i.e., 16 copies in total). Despite this agreement between the in vitro and in-cell results, fitting of the complete in vitro integrative CR and NR Y-complex models from[46] revealed that the curvature of the tail and to some degree of the arms was strikingly different in the native cellular context (Fig. 13). Therefore, in order to build the CR and NR Y-complex models that would better account for the in-cell EM densities, I performed integrative modeling with the global optimization mode from Assembline. For this, the individual rigid bodies that were used for systematic fitting together with the in-cell ScNPC maps were used as inputs.

For the CR Y-complex, the global optimization modeling converged to best scoring models with a significantly more extended stem (i.e. Nup133 and Nup84) conformation compared to the in vitro model from[46] (Fig. 13). This architecture of more extended CR Y-complexes explained the allocated density while not violating any prior structural knowledge. As the final ScNPC CR Y-complex model, the best scoring model from the aforementioned global optimization was selected (Fig. 13).

The global optimization for the NR Y-complex converged to best scoring models with a similar extended stem architecture as in the CR Y-complex case. These integrative NR Y-complex models though, were heavily clashing with neighboring Y-complexes. Moreover, the membrane binding domains of Nup133 and Nup120 (carrying ALPS motifs, see Introduction) were not in contact with the nuclear envelope, probably hinting towards differences in stem curvature between the CR and NR Y-complexes (or differently resolved local features at these regions in the in-cell maps). Thus, I selected the best scoring globally optimized NR Y-complex model and refined it with Assembline. This approach led to NR Y-complex models with resolved clashes between neighboring complexes thanks to simultaneous refinement with Assembline of 8 co-axial Y-complex copies. The best scoring model out of the refined models was selected as the final ScNPC NR Y-complex model (Fig. 13). Finally, I built the full in-cell ScNPC scaffold model by merging the selected CR and NR Y-complex models with the refined model of the IR asymmetric unit (Fig. 13).

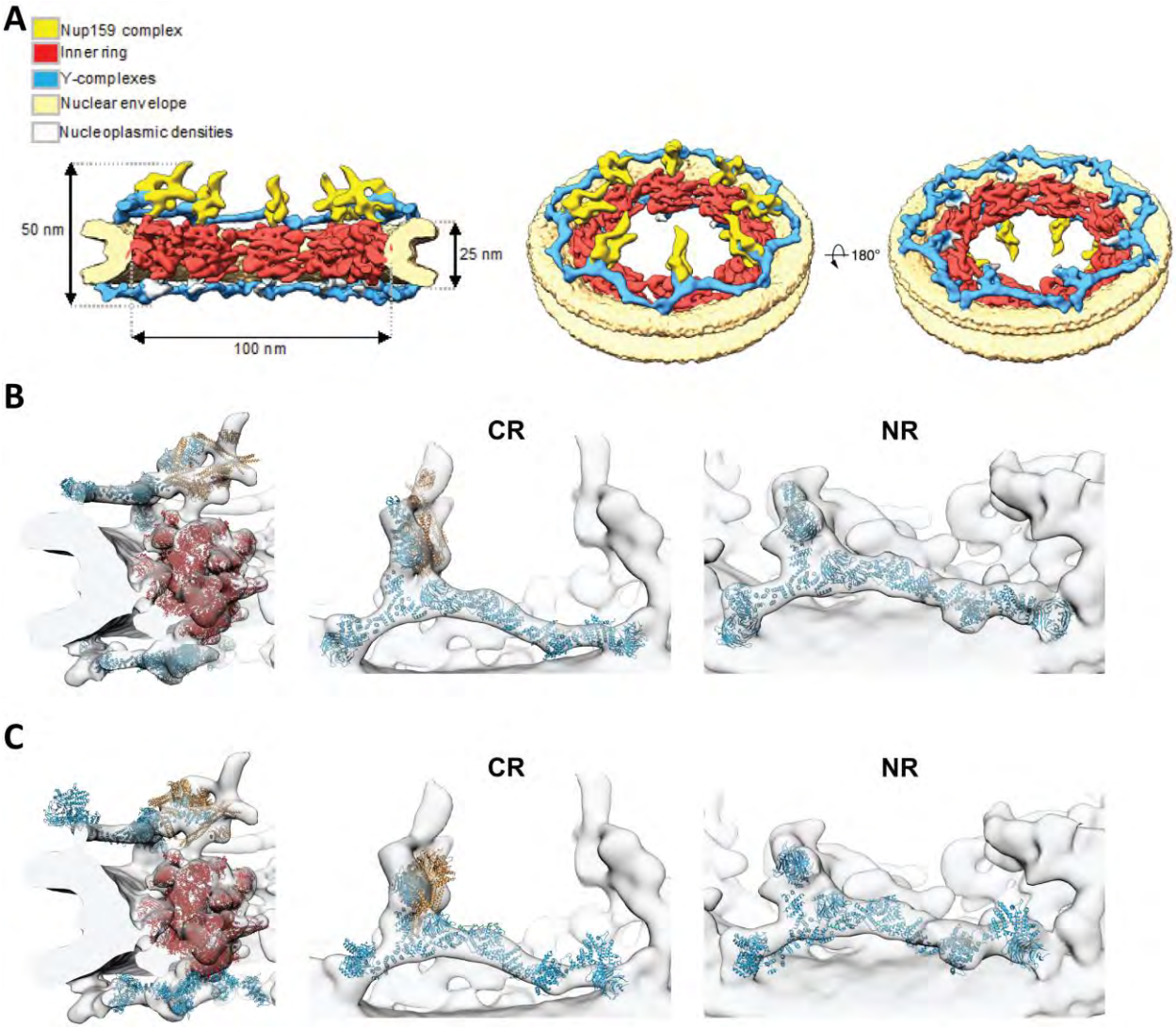


Figure 13: Overview of the in-cell ScNPC architecture and comparison with the previous in vitro model. (A) Segmentation of the entire in-cell ScNPC cryo-ET map shown in cross-section view with respect to the central axis (left) and from the cytoplasmic (middle) and nuclear (right) perspective. The individual subcomplexes such as the mRNA export platform (yellow), the inner ring asymmetric unit (IR asymmetric unit) (red), the outer rings (CR and NR) (blue) and the nuclear membrane (light brown) are highlighted by the indicated colors. (B) Representative in-cell integrative structural models of CR and NR Y-complex (blue ribbons) shown with the integrative model of P-complex (yellow ribbons) and refined IR (red ribbons) from[46] fitted to the allocated density of the in-cell map (grey density). The in-cell Y-complexes appear more extended as compared to their in vitro counterparts. (C) Representative in vitro integrative structural models of CR and NR Y-complex (blue ribbons), P-complex (yellow ribbons) and IR (red ribbons) from[46] fitted to the in-cell cryo-ET map (grey density) with respect to spatial reference frame from[46]. Figure adapted from[95].

5.1.2 Revised conformation of the mRNA export platform

An additional prominent density in the CR seemed to most likely belong to the mRNA export platform machinery as it resembled the characteristic P-shape of the Nup159 complex that was shown by previous studies[52]. This localization of the Nup159 complex in the CR was also included in the in vitro model from[46], which together with the CR Y-complex accounted for most of the observed density in this ring.

In order to allocate this observed CR density in the in-cell EM map, using Assembline, I performed systematic fitting of the previously published negative staining EM map of the P-shaped Nup159 complex[52] to the in-cell ScNPC maps. For negative staining map fitting, I used a different fitting score from UCSF Chimera, more suitable for such maps. The best scoring fit of the P-shaped map was statistically significant (p-value = 0.0027) (Fig. 14), strongly suggesting that indeed this additional CR density belonged to the Nup159 complex. As a negative control, I employed the same systematic fitting procedure for a mirror negative staining map (mirror to the original negative staining map) of the Nup159 complex[52]. The fitting of the mirror map did not yield any significant results, with the best scoring fits occupying partially the already allocated density of CR Y-complex (Fig. 14). In order to validate this density assignment even further, I additionally utilized a third-party fitting tool, called colores, which is part of the Situs software package[108]. The best scoring fits were in complete agreement with the systematic fitting results from Assembline.

Taking into account the characterized long Nup159-Dyn2 arm from the previously published negative staining map[52] and the best scoring systematic fit of this map to the in-cell ScNPC density, I superposed and locally refitted with the fitmap tool from UCSF Chimera a representative integrative model of the Nup159 complex from[46,22]. Compared with the previous in vitro based orientation proposed by[46], this analysis revealed that the P-shaped complex is rotated by 180º around the axis that points towards the central channel. In the context of this revised Nup159 complex orientation, the Nup82 β-propellers are positioned towards the IR, as opposed to the previous suggested orientation with the Nup82 β-propellers pointing towards the cytoplasm (Fig. 14). Additionally, the Nup159 arm, clearly visible in the in-cell ScNPC map, which carries DID domains responsible for binding multiple dimers of Dyn2 (see Introduction section 3.1), projects towards the cytoplasm in a ~45º angle with respect to the nucleocytoplasmic axis. In contrast, in the previously suggested orientation, the arm would face the IR[46]. Also, in the revised P-shaped complex orientation, FG-repeats of the Nsp1 protein are placed towards the central channel and hence being primed to form transient interactions with NTRs (Fig. 14).

Finally, as an additional validation of the revised Nup159 complex orientation, I mapped the in vitro-based inter-crosslinks to the ScNPC model. Indeed, in this revised Nup159 complex orientation all four crosslinks between the P-shaped complex and the Y-complex were satisfied (Fig. 14).

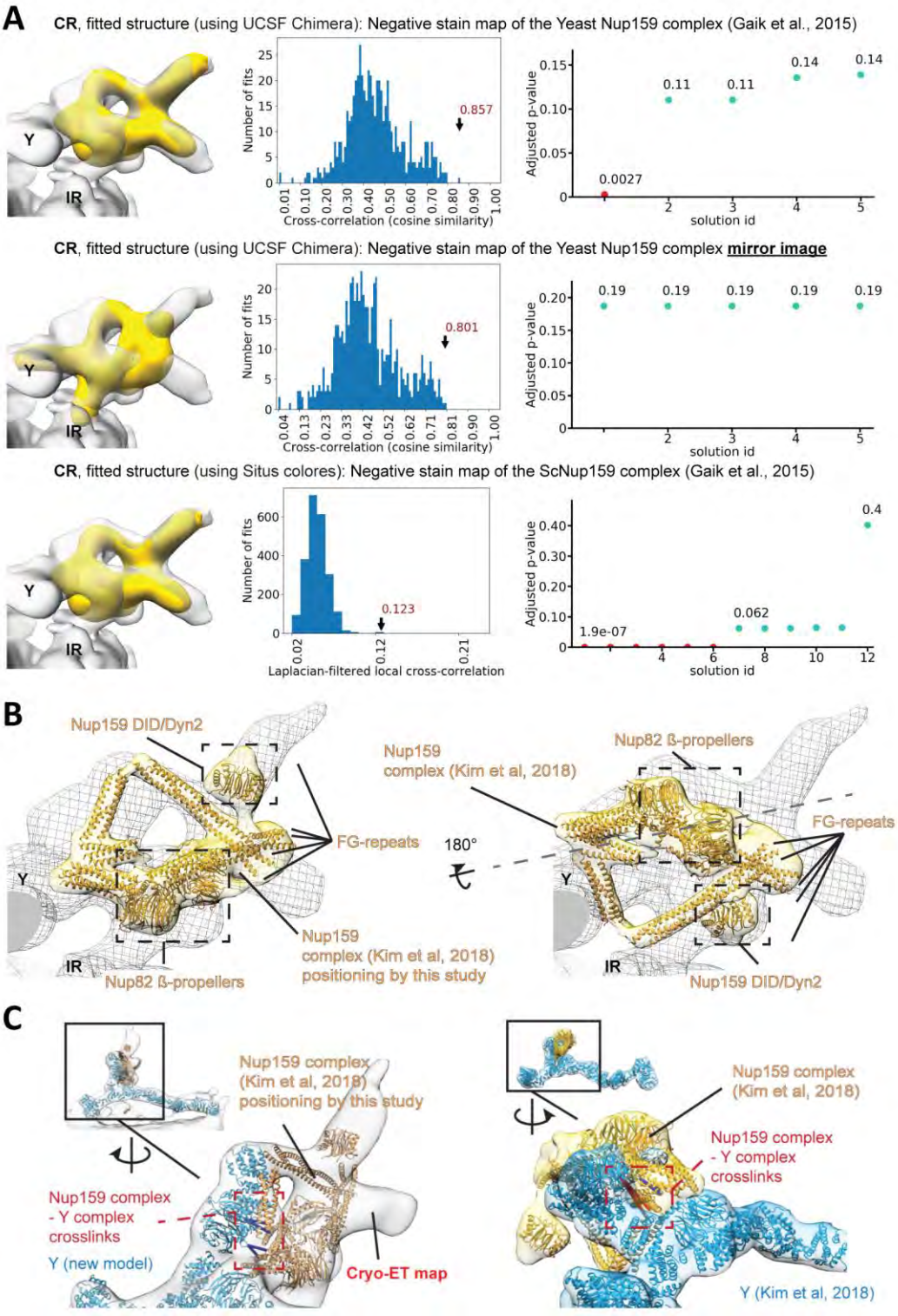


Figure 14: Validation of the Nup159 complex positioning. (A) Systematic fitting results of the negative staining map and the mirror image of the Nup159 complex (yellow) from[52] to the in-cell ScNPC map (grey) using the underlying functionality of UCSF Chimera[107] in Assembline[120] and colores tool from the Situs package[108]. Each row shows the visualization of the best scoring fits (left), the histogram of raw scores (middle), and a plot of the p-values of the five best scoring fits (right). In the p-value plots, the statistically significant fits are colored red (p-value < 0.05).

The top fits are indicated in the histograms with an arrow and the score value. The number of sampled fits used to calculate p-values after clustering of similar solutions was 585, 599, and 2243, for top, middle, and bottom rows respectively. (B) Representative integrative Nup159 complex model from[46] inside the in-cell ScNPC map (grey mesh) in the revised orientation (left) versus the in vitro determined orientation (right). The Nup159 complex model is shown in orange ribbons within yellow localization probability density from[46] locally fitted with the fitmap tool from UCSF Chimera. The in vitro orientation was reproduced by first fitting the entire model from[46] to the in-cell cryo-EM map and then locally fitting the Nup159 complex to the density (which was needed to bring the Nup159 complex into the density and preserve the orientation). The dashed grey line indicates the rotation axis between the two fits (orientations). (C) Crosslinks between the Nup159 complex from[46] and the Y-complex support the new orientation (left) compared to the in vitro orientation (right). Satisfied and violated crosslinks are depicted as blue and red bars respectively, while the in vitro Nup159 and Y-complexes from[46] are depicted within the relevant localization probability densities produced by[46]. Figure adapted from[95].

5.1.3 Integrative density allocation of the linker Nup116

Within this collaboration, I also received in-cell ScNPC maps from Nup116Δ *S. cerevisiae* cells under permissive (25 °C) and non-permissive (37 °C) temperatures (cryo-EM maps produced by Matteo Allegretti and Christian E. Zimmerli). The two maps were resolved at approximately 25 Å and 50 Å resolution, respectively. Since Nup116 is considered a key CR-IR connector (see Introduction section 3.1) by interacting with Nup82 and Nup159[52] and CR Y-complex members, these map reconstructions could confirm the density allocation of this connector Nup, thus validating the revised orientation of the P-shaped complex.

To allocate the Nup116 density, I first superposed the crystal structure of the C-terminal domain of Nup116 in complex with Nup159 and Nup82 (PDB id: 3PBP)[50] with the respective domains of two copies of Nup159 and Nup82 contained in the Nup159 complex in its updated orientation (Fig. 15). Additionally, I used the same crystal structure for systematic fitting with Assembline in the wild type in-cell ScNPC maps (Fig. 15). The results from both superposition and systematic fitting analyses indicated that two copies of the autoproteolytic C-terminal domain of Nup116 are placed into two yet unassigned CR-IR connecting densities that are proximate to each other and project towards the IR. In contrast, the structural analysis that I conducted based on previously published in vitro-based crosslinks[46,22] between Nup116 – Y-complex, Nup116 – Nup159 complex and Nup116 – IRC could not validate the revised orientation of the P-shaped complex since in both in-cell and in vitro-based conformations the relevant crosslinks were satisfied equally (Fig. 16).

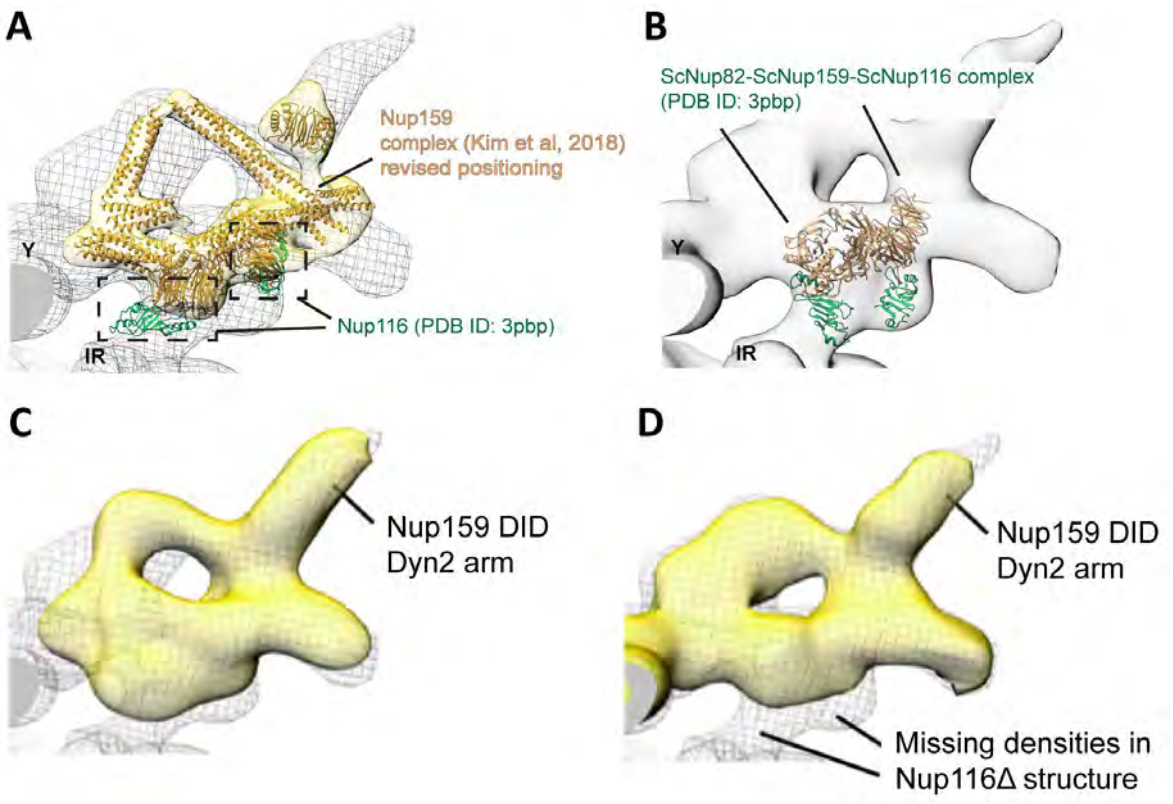


Figure 15: Integrative density assignment of Nup116. (A) Superposition of two copies of the crystal structure with PDB ID 3PBP[50] to the Nup82 β-propellers from the representative integrative Nup159 complex model from[46] in the revised orientation predicts the position of Nup116 (green ribbons). (B) Visualization of the two best-scoring systematic fits of the 3PBP crystal structure[50] to the wild type in-cell cryo-ET map confirms the positioning of Nup116 in the connecting CR-IR density proximally to Nup159-complex. (C) The Nup159 complex region of the in-cell wild type ScNPC map (grey mesh) superposed with the best-scoring systematic fit of the negative stained map of the Nup159 complex[52] (yellow surface). (D) Same as (C) but superposed with the in-cell ScNPC map of a Nup116Δ strain grown at permissive temperature (yellow surface). Figure adapted from[95].

The density allocation of Nup116 was finally validated by superposing the Nup116Δ ScNPC map under permissive temperature (25 °C) (which should have a normal morphological appearance[121] compared to the wild type ScNPC map) with the wild type map. The superposition showed that despite the great overall similarities of the two nuclear pores, there were missing densities in the Nup116Δ ScNPC map at the positions proximally to the Nup159 complex and Nup188, as predicted (Fig. 15). Interestingly, density proximally to the NR-IR connections (as observed in the wild type ScNPC map), was also diminished. Further analysis based on the same in vitro crosslinks[46,22] concerning Nup188 and the proximal linker Nups (such Nup116, Nup100 and Nup145N) hinted towards a biased cytoplasmic localization of Nup116 (Fig. 16).

Taken together, these results allowed me to unambiguously assign the linker density to Nup116 and confirm the in-cell-based orientation of the P-shaped Nup159 complex. The revised Nup159 complex conformation is spatiotemporally more in line with the extensive biochemical data regarding the mRNP export process[32,122].

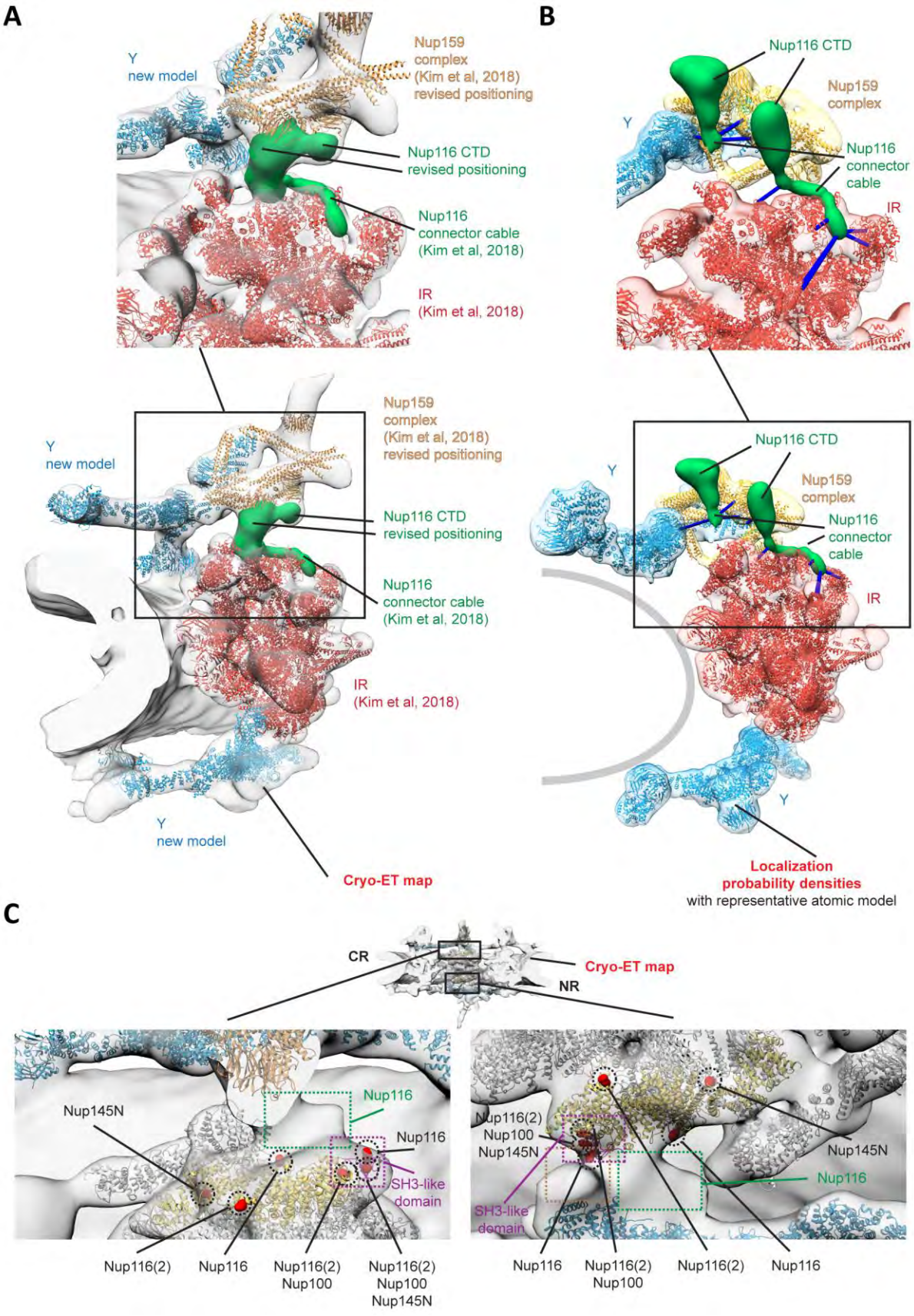

A
Y
new model
Nup159
complex
(Kim et al, 2018)
revised positioning
Nup116 CTD
revised positioning
Nup116
connector cable
(Kim et al, 2018)
IR
(Kim et al, 2018)
Y
new model
Nup159
complex
(Kim et al, 2018)
revised positioning
Nup116 CTD
revised positioning
Nup116
connector cable
(Kim et al, 2018)
IR
(Kim et al, 2018)
Y
new model
Cryo-ET map
B
Nup116 CTD
Nup159
complex
Y
Nup116
connector
cable
IR
Nup116 CTD
Y
Nup159
complex
Nup116
connector
cable
IR
Y
Localization
probability densities
with representative atomic model
C
CR
Cryo-ET map
NR
Nup145N
Nup116
Nup116
SH3-like
domain
Nup116(2)
Nup116
Nup116(2)
Nup100
Nup116(2)
Nup100
Nup145N
Nup116(2)
Nup100
Nup145N
Nup145N
Nup116
SH3-like
domain
Nup116
Nup116(2)
Nup100
Nup116(2)
Nup116

Figure 16: Revised positioning of Nup116 and possible Nup188 SH3-like domain interactions with linker ScNups. (A) Revised positioning of the Nup116 versus (B), previously published integrative model from[46] (right). The two *S. cerevisiae* NPC models were superimposed with respect to their IR so that they are aligned to the same reference frame. The Nup116 position is shown either as the density assigned to Nup116 based on the Nup116 knockout structure in the CR (A) or as localization probability densities retrieved from[46] (B). The major structural elements of the ScNPC are indicated. The Nup116 connector cable in the in-cell model (A) has been taken from[46] based on its position relatively to the IR. Blue bars represent crosslinks from Nup116 to neighboring complexes. For the in-cell model, the cryo-ET map is displayed; for the in vitro model, the localization probability densities (i.e. not an EM map) are shown instead. (C) SH3-like domain (magenta ribbons and dotted frames) of Nup188 (yellow ribbons) bridges the inner-outer rings interactions. The positions of Nup188 crosslinking to Nup116 (the number of times that some Nup188 residues crosslink with more than one Nup116 residues are denoted in the labels in parenthesis), Nup100 and Nup145N are indicated in sphere representation (red spheres) and their location suggests that they link the connecting interfaces between the IR and the outer rings: NR brown dotted frame corresponds to an unassigned density, CR green dotted frame corresponds to Nup116 density, NR green dotted frame corresponds to a second Nup116 density. Figure adapted from[95].

5.1.4 Integrative models of NPCs from Nup116Δ *S. cerevisiae* cells

The in-cell ScNPC cryo-ET maps from Nup116Δ *S. cerevisiae* cells that were produced by our collaborators (Matteo Allegretti and Christian E. Zimmerli, Beck group in EMBL HD and MPI Frankfurt) were utilized not only for validating the revised positioning of the Nup159 complex and density allocation of the linker Nup116, but also for integrative modeling. In order to produce ScNPC models from the mutant *S. cerevisiae* strains that were grown at permissive (25 °C) and non-permissive (37 °C) temperature respectively, I first systematically fitted with Assembline the integrative models of the scaffold subcomplexes that were built based on the wild type in-cell ScNPC map. This step was conducted to unambiguously assign the densities to the main scaffold components of the mutant ScNPCs. The density of all scaffold subcomplexes (i.e. CR and NR Y-complexes, IR asymmetric unit) was allocated in the case of ScNPCs from Nup116Δ cells grown under permissive temperature (25 °C, ~25 Å) with the integrative models of the wild type-based subcomplexes being fitted to the respective map with significant p-values. These results were well in line with previous observations regarding the normal morphological appearance of ScNPCs under such conditions[121]. On a different note, superposition analysis and visual observation of the ScNPCs from Nup116Δ cells grown under non-permissive temperature (37 °C, ~50 Å) revealed a missing CR and roughly half of the IR. These observations were confirmed by the systematic fitting with Assembline which was able to confidently allocate the density of only the outer nuclear copies of the IR asymmetric unit and the NR Y-complex. These NPCs displayed an envelope morphology reminiscent of interphase assembly intermediates in human cells. Therefore, we hypothesized that this in-cell ScNPC map represented a failed "inside-out" NPC assembly[123].

Starting from the best-scoring fits of the scaffold ScNPC subcomplex models to the mutant ScNPC maps (the models built earlier based on the wild type ScNPC map), the integrative modeling of the mutant ScNPCs was completed by using the refinement mode from Assembline. The final refined integrative models of the mutant NPCs explained very well the observed density in both cases. Hence, the Nup116Δ ScNPC model from cells grown under permissive temperature included the same scaffold complexes as the wild type ScNPC while the Nup116Δ ScNPC model

from cells grown under the non-permissive temperature included only the outer nuclear copies of the IR unit and the NR Y-complex (Fig. 17). Finally, to make all the three in-cell-based ScNPC models publicly available, I developed a semi-automated computational protocol that produced ready-to-deposit integrative models for PDB-Dev[100].

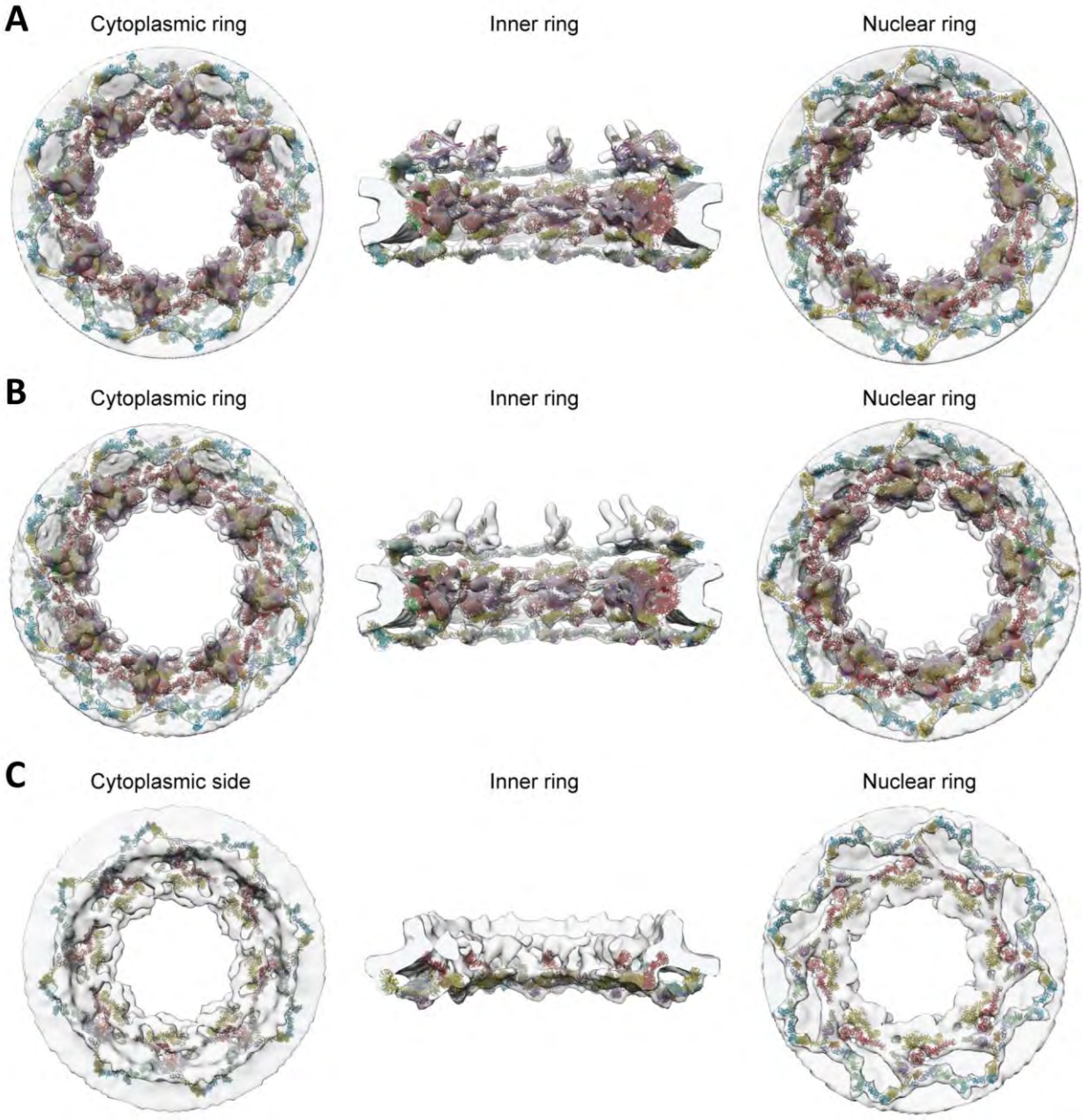


Figure 17: In-cell-based integrative models of ScNPCs built with Assembline. Overview of cytoplasmic ring (left), inner ring (middle) and nuclear ring (right) of (A) the wild type ScNPC, (B) the ScNPC from Nup116Δ cells (at 25 °C) and (C) the ScNPC from Nup116Δ cells (at 37 °C) where the cytoplasmic ring is entirely missing (left). Figure adapted from[120].

All in all, the in-cell cryo-ET maps from wild type and Nup116Δ *S. cerevisiae* cells that I received from our collaborators enabled me to produce the first in-cell models of ScNPCs. By applying our own integrative structural modeling tool called Assembline, which is described in Synopsis section 5.3, I revealed the architecture of wild type ScNPCs inside the cells, which is dilated compared to the previous in vitro studies. The unprecedented resolution of the in-cell ScNPC maps allowed for unambiguous localization and orientation of important subcomplexes and NPC "ring linkers" such as the P-complex and Nup116. Taken together, the integrative models of ScNPCs from either the wild type and knock-out cells together with experiments conducted by our collaborators have shed more light not only on the NPC-related functions such as the mRNA export but also on the assembly and proper formation of these large complexes.

Article

In-cell architecture of the nuclear pore and snapshots of its turnover

https://doi.org/10.1038/s41586-020-2670-5

Received: 7 August 2019

Accepted: 1 June 2020

Published online: 2 September 2020

Check for updates

Matteo Allegretti[1,8], Christian E. Zimmerli[1,2,8], Vasileios Rantos[3], Florian Wilfling[4], Paolo Ronchi[5], Herman K. H. Fung[1], Chia-Wei Lee[4], Wim Hagen[1], Beata Turoňová[1], Kai Karius[3], Mandy Börmel[5], Xiaojie Zhang[1], Christoph W. Müller[1], Yannick Schwab[5,6], Julia Mahamid[1,6], Boris Pfander[4]✉, Jan Kosinski[1,3]✉ & Martin Beck[1,6,7]✉

Nuclear pore complexes (NPCs) fuse the inner and outer membranes of the nuclear envelope. They comprise hundreds of nucleoporins (Nups) that assemble into multiple subcomplexes and form large central channels for nucleocytoplasmic exchange[1,2]. How this architecture facilitates messenger RNA export, NPC biogenesis and turnover remains poorly understood. Here we combine in situ structural biology and integrative modelling with correlative light and electron microscopy and molecular perturbation to structurally analyse NPCs in intact *Saccharomyces cerevisiae* cells within the context of nuclear envelope remodelling. We find an in situ conformation and configuration of the Nup subcomplexes that was unexpected from the results of previous in vitro analyses. The configuration of the Nup159 complex appears critical to spatially accommodate its function as an mRNA export platform, and as a mediator of NPC turnover. The omega-shaped nuclear envelope herniae that accumulate in *nup116Δ* cells[3] conceal partially assembled NPCs lacking multiple subcomplexes, including the Nup159 complex. Under conditions of starvation, herniae of a second type are formed that cytoplasmically expose NPCs. These results point to a model of NPC turnover in which NPC-containing vesicles bud off from the nuclear envelope before degradation by the autophagy machinery. Our study emphasizes the importance of investigating the structure–function relationship of macromolecular complexes in their cellular context.

NPCs are giant macromolecular assemblies with an intricate architecture. About 30 different Nups assemble in multiple copies to form eight-fold rotationally symmetric NPCs, comprising approximately 550 protein building blocks in yeast[1]. The NPC consists of two outer rings, also called nuclear (NR) and cytoplasmic (CR) rings, which are located distally to the inner ring (IR). Scaffold Nups contain folded domains and form a cylindrical central channel. This channel is lined with FG Nups, which harbour intrinsically disordered Phe and Gly (FG)-rich repeats that interact with nuclear transport receptors. The Y-complex forms the outer rings, whereas the inner ring complex builds the IR[2]. The yeast Nup159 complex associates asymmetrically with the Y-complex at the CR and facilitates the terminal steps of mRNA export. Its core consists of two Nup159–Nup82–Nsp1 heterotrimers that dimerize into a characteristic P-shaped structure[4,5].

Genetic perturbation of Nup159 complex-associated Nup116[3] and Gle2[6] or nuclear envelope proteins such as Apq12[7] lead to the formation of clustered nuclear envelope herniae, omega-shaped morphologies that have been linked to neurological diseases such as dystonia[8,9]. Herniae engulf a membrane opening and contain Nups at their neck[3,9]; however, whether they engulf an entire or partial NPC remains unknown[10]. It has been suggested that herniae could be the result of inside-out assembly events of NPCs in which the fusion of the two nuclear membranes failed, or alternatively, the could comprise defective NPCs that were sealed off with membranes by the ESCRT machinery[10]. Recently, an autophagy pathway was described that degrades NPCs by direct interaction of the ubiquitin-like autophagosomal protein Atg8 with the cytosolic filament Nup159, which serves as an intrinsic autophagy receptor[11]. However, it remains unknown whether the formation of herniae is functionally related to the selective autophagy of NPCs.

Understanding how Nup subcomplexes are positioned relative to each other in cells requires the convergence of in vitro and in situ approaches[12]. In situ structural analysis is available for vertebrates and algae, and has revealed that key features of the NPC architecture are not conserved[1,13], such as the stoichiometry of Y-complexes within the outer rings. Notably, the P-shaped outline of the yeast Nup159 complex[5] was not apparent in any of the cryo-electron microscopy (cryo-EM) maps, suggesting either structural diversity, or questioning its physiological relevance.

There remains a lack of in situ structural analysis of the *S. cerevisiae* NPC, which has been extensively studied as a model organism for Nup

[1]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. [2]Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences, Heidelberg, Germany. [3]Centre for Structural Systems Biology (CSSB), DESY and European Molecular Biology Laboratory, Hamburg, Germany. [4]Max Planck Institute of Biochemistry, Martinsried, Germany. [5]Electron Microscopy Core Facility (EMCF), European Molecular Biology Laboratory, Heidelberg, Germany. [6]Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, Germany. [7]Department of Molecular Sociology, Max Planck Institute of Biophysics, Frankfurt, Germany. [8]These authors contributed equally: Matteo Allegretti, Christian E. Zimmerli. ✉e-mail: bpfander@biochem.mpg.de; jan.kosinski@embl.de; martin.beck@biophys.mpg.de

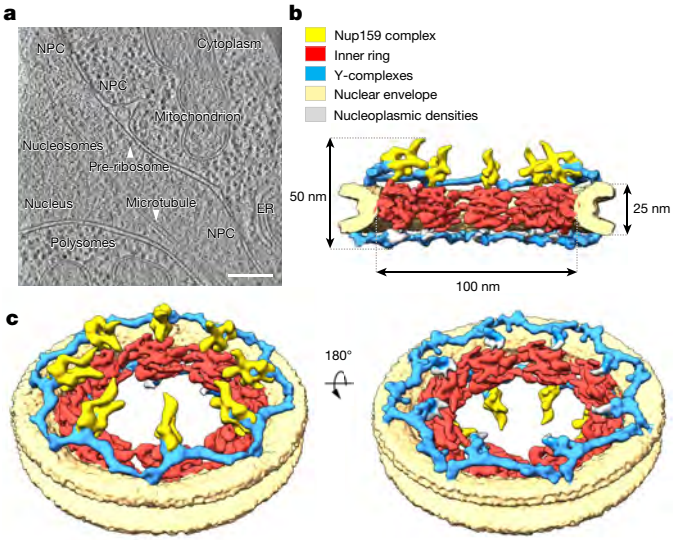


Fig. 1 | In-cell structure of *S. cerevisiae* NPC. a, Cryo-tomographic slice of a *S. cerevisiae* cell during division. $n = 240$ tomograms of *S. cerevisiae* wild-type cells were acquired from grids prepared on four different days from different cell cultures (biological replicates). ER, endoplasmic reticulum. Scale bar, 200 nm. **b**, Segmentation of the cryo-EM map of the *S. cerevisiae* NPC cut in half along the central axis. **c**, Tilted view of the entire NPC showing both the cytoplasmic (left) and the nucleoplasmic (right) faces.

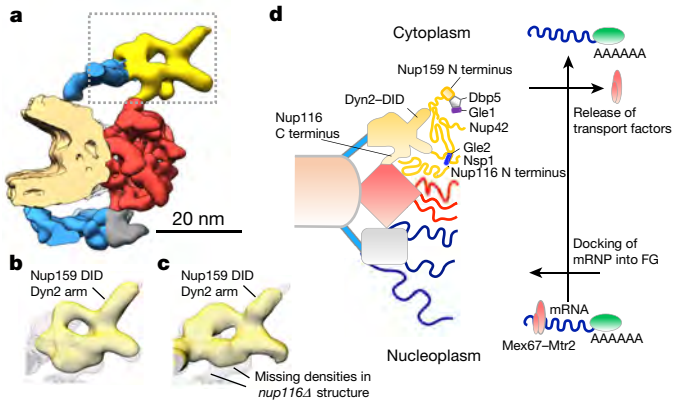


Fig. 2 | Nup159 complex architecture. a, Individual spoke of the NPC (colour-coded as in Fig. 1). The Nup159 complex is highlighted with a dotted frame. **b**, The Nup159 complex region of the in-cell NPC map (grey mesh) superimposed with the top-scoring systematic fit (see Methods) of the negative-stain map of the Nup159 complex[5] (yellow surface). **c**, Same as **b** but superimposed with the in-cell NPC map of a *nup116Δ* strain grown at permissive temperature (yellow surface). **d**, Spatial model of how mRNA export is accommodated by the orientation of the Nup159 complex. A poly-A mRNA with transport factors Mex67–Mtr2 (red) and proteins that protect the poly-A tail from degradation (light green) docks to the unstructured region of the basket Nups[30] (grey box with blue filaments). The Nup159 N-terminal domain that mediates the release of the transport factors from the mRNA[31] is anchored at the cytoplasmic side.

structure, NPC architecture, biogenesis, surveillance and turnover, and the mechanism of mRNA export[1,11,14]. An integrative model of the entire *S. cerevisiae* NPC architecture based on extensive experimental analysis in vitro has been recently put forward[15], but it remains unknown to what extent the native architecture of the NPC has been preserved.

Structure of the NPC in the cellular context

To characterize the architecture of *S. cerevisiae* NPC in situ, we prepared thin cryo-focused ion beam (cryo-FIB) lamellae[16] of cells in the exponential growth phase. We acquired 240 cryo-electron tomograms (Fig. 1a) containing around 500 NPCs and determined the structure by subtomogram averaging. The resulting cryo-EM map at approximately 25 Å resolution (Extended Data Fig. 1a, b, Extended Data Table 1) provides a detailed overview of the native configuration and conformation of subcomplexes of actively transporting NPCs in cells (Fig. 1b, c, Extended Data Fig. 1b, c). A visual inspection of the structure reveals marked disparities, not only compared with the previously analysed structures of the algal[13] and human[17] NPCs, but also compared with previous analysis of isolated *S. cerevisiae* NPC[15] (Extended Data Fig. 1c, d). With minor refinements, the integrative model of one asymmetric unit of the IR[15] fits unambiguously into the observed density ($P = 1.6 \times 10^{-12}$) (Extended Data Fig. 2), but the entire IR has to be dilated by about 20 nm in diameter, thereby spatially separating the eight individual spokes (Extended Data Fig. 3a). This analysis underlines the plasticity of the NPC within cells[13,16], which might be physiologically relevant for the transport of large cargos and inner nuclear membrane proteins.

As expected[15,18], 16 nuclear and cytoplasmic copies of the Y-complex are apparent in our cryo-EM map (Fig. 1c, d, Extended Data Fig. 1c). The crystal structures and homology models of the yeast Y-complex vertex[19] (Sec13–Nup145C–Nup84–Nup120–Seh1–Nup85) or its fragments fit into the observed Y-shaped density (Extended Data Fig. 2). We used an integrative modelling procedure taking into account the more extended in situ conformation and obtained complete structures of Y-complexes (Extended Data Fig. 3b, c, Supplementary Video 1). At the NR, additional densities appear at the connection to the IR and around the Y-complex, in particular after local masking (Figs. 1b, c, 2a, Extended Data Fig. 3d). The densities around the Y-complex form a rod that probably corresponds to fragments of the nuclear basket coiled-coil filaments as in the structure in ref. [15], but anchored to the Y-complex vertex.

A prominent feature of the electron microscopy map is the P-shaped density in the CR that is highly reminiscent of the previous analysis of the isolated, negatively stained Nup159 complex[4,5] (Fig. 2). Systematic fitting of the Nup159 complex negative-stain map[5] into our in situ structure confirmed this assignment (Fig. 2b, $P = 0.0027$; Extended Data Fig. 4a). On the basis of the top resulting fit, we superimposed a representative integrative model of the Nup159 complex[4,15] and locally fitted it into our cryo-EM map (Extended Data Fig. 4b, Supplementary Video 1). The Nup82 β-propellers are positioned towards the inner ring, but nevertheless the previously published crosslinks[15] between the Nup159 and Y-complexes that map to the structure are satisfied (Extended Data Fig. 4e). The arm of the P-shape that consists of Nup159 DID tandem repeats binding multiple Dyn2 dimers[5,20] is clearly apparent in situ. In contrast to the previous model[15], it projects towards the cytoplasm at an angle of around 45° with respect to the nucleocytoplasmic axis (Fig. 2, Extended Data Fig. 4b).

nup116Δ confirms the orientation of Nup159 complex

Nup116 is one of the three yeast homologues of the essential vertebrate Nup98 and a key Nup for the NPC permeability barrier[21,22], mRNA export[21,23], pre-ribosome traslocation[24] and ring connectivity[15,25,26]. Superposition of the crystal structure[27] of Nup116(966–1111)–Nup82(1–452)–Nup159(1425–1458) with the respective parts of two copies of Nup159 and Nup82 contained in the P-complex (Extended Data Fig. 4c, d) suggested that the autoproteolytic domain of Nup116 is located in unassigned densities projecting towards the Nup188 in the IR. To validate this assignment, we structurally analysed the NPC in *nup116Δ* cells at permissive temperature (25 °C, see Methods), at which NPCs have a normal morphological appearance[3]. The *nup116Δ* structure is overall similar to the map of the wild type (Extended Data Fig. 6a, c,

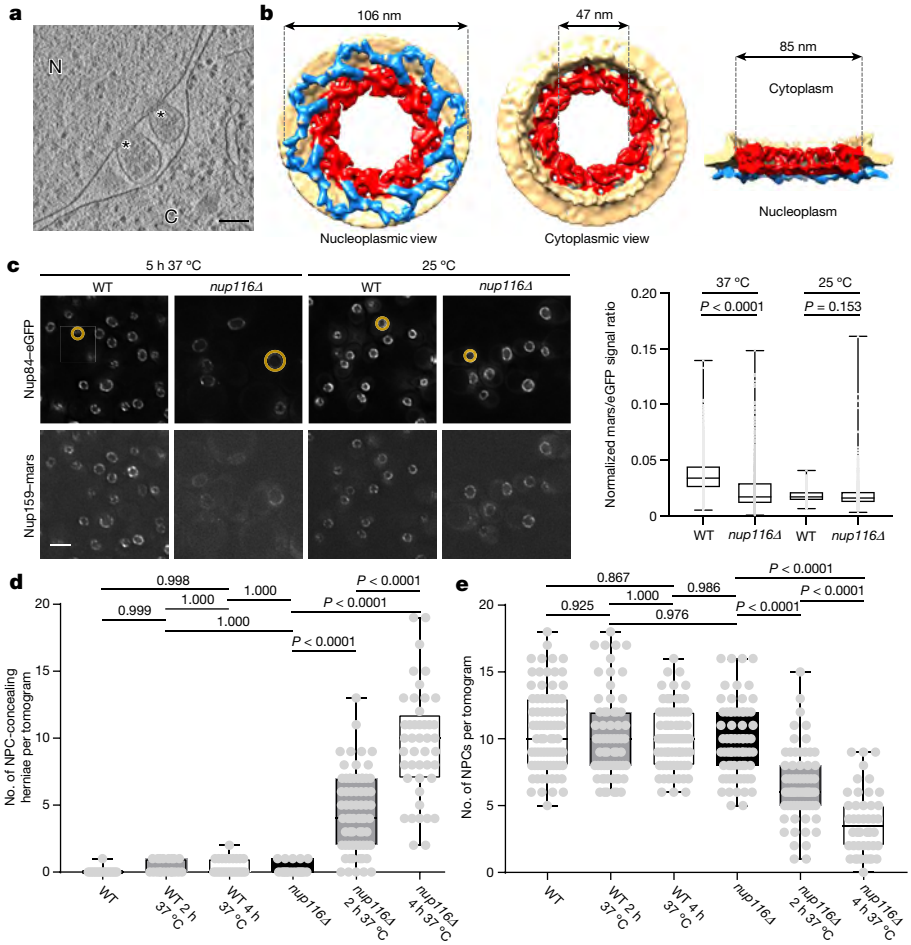


Fig. 3 | NPC-concealing hernia. a, Cryo-tomographic slice of a nuclear envelope with NPC-concealing herniae (*nup116Δ*, 4 h at 37 °C); N, nucleus; C, cytoplasm; asterisks mark densities at the neck of the herniae subjected to subtomogram averaging Scale bar, 100 nm. **b**, Segmentation of the cryo-EM map of the partial NPC found at the basis of NPC-concealing herniae. Colour code as in Fig. 1. **c**, Phase-contrast and maximum-intensity projections of deconvoluted wide-field microscopy stacks showing the nuclear envelope in wild-type (WT) and *nup116Δ* cells at 25 °C or 37 °C with markers Nup84 and Nup159. Yellow circles show the nuclear size (results quantified in Extended Data Fig. 6e). Right, the mars/eGFP ratio shows the significant loss of Nup159 signal in *nup116Δ* at 37 °C (three biologically independent replicates with $n = 300$ nuclei each, evaluated with Kruskal–Wallis test followed by Dunn's multiple comparisons test). **d**, Quantification of number of NPC-concealing herniae from 300 nm tomograms of *S. cerevisiae* cells (plastic sections) **e**, Quantification of number of NPCs not at the herniae from 300 nm tomograms acquired as in **d**. In **d**, **e**, $n = 70$ tomograms on average for each condition, see Supplementary Table 2; one-way ANOVA with Dunn's multiple comparisons test. In box plots (**c**–**e**), box centres are the median, box edges represent first and third quartiles and whiskers show minimum and maximum values.

Supplementary Video 2), but lacks density at the positions proximate to the Nup159 complex and Nup188 as predicted (Fig. 2c, Extended Data Fig. 4c). This finding corroborates the spatial positioning of Nup116 and the orientation of the Nup159 complex (Extended Data Fig. 5a, b). Previous crosslinking analysis[15] and biochemical data suggesting that Nup116 links to Nup188 and Nup192 of the IR[25,26] agree well with this configuration (Extended Data Figs. 4e, 5a, b). The position proximate to the SH3 domain of Nup188[28] in the NR also loses some density (Extended Data Figs. 5c, 6c, Supplementary Video 2), which would be in line with a cytoplasmically biased localization of Nup116[29].

Several studies have highlighted that the accurate positioning of the Nup159 complex with respect to the central channel at the cytoplasmic face of the NPC is critical for the spatial organization and directionality of mRNA export[4,30,31]. Our model of NPC architecture accommodates extensive biochemical analysis of mRNP export[21,30,31]. Indeed, Dbp5 and the N terminus of Nup159—which facilitate the terminal release of Mex67—are positioned towards the cytoplasm[32], whereas the FG repeats of Nsp1 that interact with Mex67 earlier during export, are placed towards the IR. (Fig. 2d). Nup159 itself also contains the Atg8-family interacting motif (AIM), which is located proximate to the DID–Dyn2 arm and facilitates autophagic NPC turnover upon nitrogen starvation[11]. This AIM is exposed towards the cytosolic site, allowing access of Atg8 (Fig. 4a).

Herniae enclose partially assembled NPCs

Although the previous implications of the Nup159 complex in nuclear envelope herniae formation and selective autophagy suggest a functional link to membrane remodelling during NPC assembly and turnover, their exact relationship remains to be determined. We structurally analysed the membrane openings at the bases of the herniae formed in the *nup116Δ* strain at non-permissive temperature[5] (37 °C) (Fig. 3a, Supplementary Video 3) by subtomogram averaging (Fig. 3b, Extended Data Fig. 6a, b). We found that the necks of the herniae enclose partial NPCs that have a smaller diameter compared with the wild type (15 nm difference). Concurrently, the fused nuclear membranes are less curved (about 84° instead of 180°, on average; Fig. 3b, Extended Data Fig. 6a). The cryo-EM map reveals that the entire cytoplasmic ring, including the Nup159 and Y-complexes, and adjacent parts of the inner ring are missing (Fig. 3b, Extended Data Fig. 6a, c). Light microscopy data

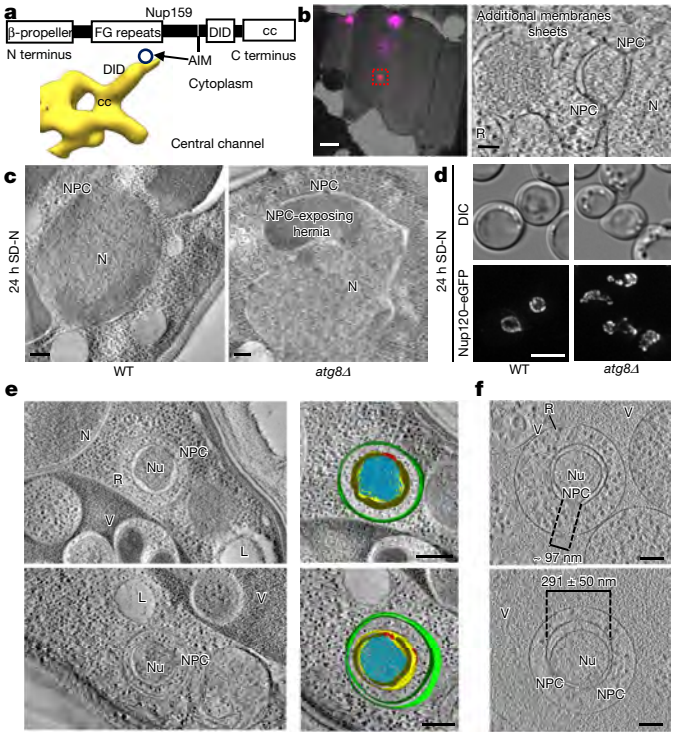


Fig. 4 | NPC-exposing hernia. a, Domain structure of Nup159 with corresponding region of the cryo-EM map showing the position of the Atg8-interacting motif (AIM[11]). cc, coiled coil domain. **b**, Three dimensional cryo-CLEM of Nup159–Atg8–split Venus (red) *nup120Δ* cells after 5.5 h of starvation, shown as cryo-EM overview of the lamella (left; scale bar, 2 µm; fiducials in magenta) and cryo-tomographic slice (right; relevant area framed red; scale bar, 100 nm). $n = 2$ tomograms were acquired in two independent sessions. R, ribosome. **c**, Tomographic slice from 300 nm plastic sections of wild-type and *atg8Δ S. cerevisiae* nuclei after 24 h of starvation (scale bar, 200 nm). $n = 75$ tomograms of independent nuclei (technical replicates) were acquired. SD-N, synthetic minimal medium lacking nitrogen. **d**, Phase-contrast and deconvoluted wide-field maximum-intensity projection of the same strains imaged live before and after 24 h of starvation (three biologically independent replicates and $n = 300$ per replica). **e**, Tomographic slice (plastic section), corresponding segmentation and isosurface rendering of NPC-containing nuclear vesicles in the cytoplasm from *S. cerevisiae atg15Δ* cells starved for approximately 24 h. Vesicles are surrounded by ribosomes and two membranes. Nuclear content (Nu) in cyan, double membrane in green. V, vacuole; L, lipid droplet. $n = 5$ tomograms from different cells (technical replicates) were acquired. Scale bar, 200 nm. **f**, Cryo-tomographic slices of NPC-containing nuclear vesicles in the vacuole from *S. cerevisiae atg15Δ* cells starved for 24 h. The vesicles are surrounded by ribosomes and only one membrane. Scale bar, 100 nm. $n = 35$ vesicles (technical replicates) show an NPC diameter of 97.4 ± 6.5 nm and a nuclear vesicle diameter of 291 ± 50 nm (mean ± s.d.).

confirm that the Nup159 complex fails to integrate into NPCs under these conditions (Fig. 3c). Quantification of the tomographic electron microscopy data revealed that in *nup116Δ* cells at 37 °C, herniae accumulate over time at the nuclear envelope concomitantly with a decrease of non-herniated NPCs (Fig. 3d, e). Even at permissive temperature (25 °C) the nuclear envelope of *nup116Δ* cells in our cryo-tomograms contain a significantly increased number of mushroom-shaped evaginations of the inner nuclear membrane (Extended Data Fig. 6d). This nuclear envelope morphology is reminiscent of interphase assembly intermediates in human cells[33].

Three different lines of evidence, (1) the appearance of evaginations, (2) the fact that partial NPCs are enclosed in herniae, and (3) the interdependence of herniae and NPC number, strongly support models that conceptualize herniae as a consequence of failed, inside-out interphase assembly[34]. They point to a direct or indirect contribution of Nup116 to the fusion of the two nuclear membranes during NPC assembly that if compromised, prevents the association of the cytoplasmic NPC components. Although herniae should not be conceived as 'on-pathway' intermediates of NPC assembly, some Nup interactions relevant for NPC biogenesis might be preserved. We built the structural model refining the WT *S. cerevisiae* NPC model into this map (Supplementary Video 4), which confirms that only the Y-complexes of the NR and partial IR are confidently assigned. The Y-complexes of the NR take up a conformation in which the Nup85 arm contacts the IR in the Nic96 region, forming a continuous coat across NR and IR[35] that was not observed in any of the previously published NPC structures (Extended Data Fig. 6c, Supplementary Video 4).

NPC-exposing herniae are subject to autophagy

The reduction of functionally assembled NPCs under *nup116Δ* conditions is concurrent with an increased nuclear size[3] (Extended Data Fig. 6e, f), emphasizing the association between the accumulation of the herniae, a consequent defective nuclear transport activity and nuclear size control[36]. Since binding of Atg8 to Nup159 is an essential prerequisite for autophagic clearance of NPCs, our structural analysis predicts that herniae of *nup116Δ* cells shifted to non-permissive temperature cannot be cleared by selective autophagy because their NPCs are not accessible to the cytoplasm and they do not include the AIM-containing Nup159 complex (Fig. 3a, b, Extended Data Fig. 6a, Supplementary Videos 3–4). Indeed, the degradation of NPCs upon nitrogen starvation was strongly reduced in *nup116Δ* cells at 30 °C as compared to 25 °C (Extended Data Fig. 7a).

To investigate how the autophagy machinery is recruited to the NPC, we used a split-Venus approach to visualize the Nup159–Atg8 interaction[11] (Extended Data Fig. 8a) using correlative light and electron microscopy[37,38] (CLEM) in an NPC-clustering background (*nup120Δ*[11]) that was shown to induce hernia-related structures in which many NPCs are cytoplasmically exposed[39]. After 5.5 h of nitrogen starvation, we identified the regions of interest (Fig. 4b, Extended Data Fig. 8b) and acquired electron tomograms at positions of NPCs engaged in Atg8 interaction. Although interaction between receptors and Atg8 does not require prior lipidation of Atg8 to the phagophore[40,41], these data revealed agglomerates of herniae similar to those described in ref. [39], surrounded by additional membrane sheets (Fig. 4b, Extended Data Fig. 8b, Supplementary Videos 5, 8–10). These additional membranes were seen in several but not all tomograms, which is expected, given the transient nature of this intermediate of the autophagy pathway. Nevertheless, the structures we observed were often connected to the nuclear envelope and may represent endoplasmic reticulum that contributes to isolation-membrane formation[42–44]. As expected, under these conditions, herniae expose NPCs to the cytosol and thus the AIM motifs are accessible to the selective autophagy machinery. We therefore distinguish between 'NPC-concealing herniae', Ω-shaped nuclear envelope morphologies observed under *nup116Δ* conditions at 37 °C, and 'NPC-exposing herniae', nuclear membrane blebs that cytoplasmically expose NPCs containing the Nup159 complex, as observed under *nup120Δ* conditions.

To clarify whether 'NPC-exposing herniae' are typical for selective autophagy of NPCs during nitrogen starvation or rather a peculiarity of the *nup120Δ*, we analysed *atg8Δ* cells in which the respective autophagy machinery cannot be recruited to the NPC and Nup degradation by autophagy is inhibited. Under these conditions, we observed a prominent accumulation of deformed nuclei (Fig. 4c, d, Extended Data Fig. 7c, d) with larger NPC-exposing herniae compared with *nup120Δ*; NPCs were mildly clustered (Extended Data Fig. 7e) and very dense material accumulated inside them (Supplementary Video 6). A considerably less pronounced phenotype was observed upon interference with the

ESCRT machinery[45] (Extended Data Fig. 7b, f). Together, our results support a model in which NPC-exposing herniae accumulate at the nuclear envelope under conditions in which selective autophagy is triggered (nitrogen starvation), but the autophagy machinery is missing (Fig. 4c, d, Extended Data Fig. 7b–d, f, g). This model predicts that a piece of the nuclear envelope containing clustered NPCs buds off into cytoplasmic intermediates of NPC degradation. To test this prediction, we analysed *atg15Δ* cells in which lipid degradation in the vacuole is perturbed[11]. Indeed, NPC-containing nuclear vesicles surrounded by ribosomes and double membranes typical of autophagosomes were observed in the cytosol, disconnected from the nuclear envelope (Fig. 4e, Supplementary Video 7). They are also found in the vacuole, as shown in ref. [11], surrounded by a single membrane. Cryo-ET analysis revealed that the diameter of NPCs contained in such nuclear vesicles is similar to those in the wild type (Fig. 4f).

Conclusions

Our in situ structural analysis of the *S. cerevisiae* NPC highlights the physiological relevance of the orientation of the Nup159 complex with respect to the NPC scaffold, which accommodates the terminal steps of mRNA export and enables the exposure of the intrinsic AIM. Indeed, our data suggest that the accessibility of the Nup159 complex at the NPC-exposing hernia is critical for nuclear envelope budding and selective autophagy of NPCs, whereas NPC-concealing herniae are probably a result of failed NPC assembly (Extended Data Fig. 9), although some of the individual steps need to be investigated in more detail. Whether the material contained in the NPC-containing vesicles and NPC-exposing herniae can be attributed to specific cargos or is simply unspecific nuclear content remains to be determined. Our structural models of the WT and *nup116Δ S. cerevisiae* NPC reveal conformational changes and interactions previously unobserved in NPC structures, suggesting that NPC assembly might be facilitated by dedicated Nup interactions between the NR and the IR, which require further investigation. Finally, our findings highlight the power of in-cell cryo-EM to provide insights into fundamental processes of eukaryotic cell biology.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2670-5.

1. Lin, D. H. & Hoelz, A. The structure of the nuclear pore complex (an update). *Annu. Rev. Biochem*. **88**, 725–783 (2019).
2. Beck, M. & Hurt, E. The nuclear pore complex: understanding its function through structural insight. *Nat. Rev. Mol. Cell Biol*. **18**, 73–89 (2017).
3. Wente, S. R. & Blobel, G. A temperature-sensitive NUPll6 null mutant forms a nuclear envelope seal over the yeast nuclear pore complex thereby blocking nucleocytoplasmic traffic. *J. Cell Biol*. **123**, 275–284 (1993).
4. Fernandez-Martinez, J. et al. Structure and function of the nuclear pore complex cytoplasmic mRNA export platform. *Cell* **167**, 1215–1228 (2016).
5. Gaik, M. et al. Structural basis for assembly and function of the Nup82 complex in the nuclear pore scaffold. *J. Cell Biol*. **208**, 283–297 (2015).
6. Murphy, R., Watkins, J. L. & Wente, S. R. GLE2, a *Saccharomyces cerevisiae* homologue of the *Schizosaccharomyces pombe* export factor RAE1, is required for nuclear pore complex structure and function. *Mol. Biol. Cell* **7**, 1921–1937 (1996).
7. Scarcelli, J. J., Hodge, C. A. & Cole, C. N. The yeast integral membrane protein Apq12 potentially links membrane dynamics to assembly of nuclear pore complexes. *J. Cell Biol*. **178**, 799–812 (2007).
8. Pappas, S. S., Liang, C. C., Kim, S., Rivera, C. O. & Dauer, W. T. TorsinA dysfunction causes persistent neuronal nuclear pore defects. *Hum. Mol. Genet*. **27**, 407–420 (2018).
9. Laudermilch, E. et al. Dissecting Torsin/cofactor function at the nuclear envelope: a genetic study. *Mol. Biol. Cell* **27**, 3964–3971 (2016).
10. Thaller, D. J. & Patrick Lusk, C. Fantastic nuclear envelope herniations and where to find them. *Biochem. Soc. Trans*. **46**, 877–889 (2018).
11. Lee, C.-W. et al. Selective autophagy degrades nuclear pore complexes. *Nat. Cell Biol*. **22**, 159–166 (2020).
12. Hoelz, A., Glavy, J. S. & Beck, M. Toward the atomic structure of the nuclear pore complex: when top down meets bottom up. *Nat. Struct. Mol. Biol*. **23**, 624–630 (2016).
13. Mosalaganti, S. et al. In situ architecture of the algal nuclear pore complex. *Nat. Commun*. **9**, 2361 (2018).
14. Ungricht, R. & Kutay, U. Mechanisms and functions of nuclear envelope remodelling. *Nat. Rev. Mol. Cell Biol*. **18**, 229–245 (2017).
15. Kim, S. J. et al. Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482 (2018).
16. Mahamid, J. et al. Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**, 969–972 (2016).
17. von Appen, A. et al. In situ structural analysis of the human nuclear pore complex. *Nature* **526**, 140–143 (2015).
18. Rajoo, S., Vallotton, P., Onischenko, E. & Weis, K. Stoichiometry and compositional plasticity of the yeast nuclear pore complex revealed by quantitative fluorescence microscopy. *Proc. Natl Acad. Sci. USA* **115**, E3969–E3977 (2018).
19. Stuwe, T. et al. Nuclear pores. Architecture of the nuclear pore complex coat. *Science* **347**, 1148–1152 (2015).
20. Stelter, P. et al. Molecular basis for the functional interaction of dynein light chain with the nuclear-pore complex. *Nat. Cell Biol*. **9**, 788–796 (2007).
21. Strawn, L. A., Shen, T. & Wente, S. R. The GLFG regions of Nup116p and Nup100p serve as binding sites for both Kap95p and Mex67p at the nuclear pore complex. *J. Biol. Chem*. **276**, 6445–6452 (2001).
22. Schmidt, H. B. & Görlich, D. Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. *eLife* **4**, e04251 (2015).
23. Adams, R. L., Terry, L. J. & Wente, S. R. Nucleoporin FG domains facilitate mRNP remodeling at the cytoplasmic face of the nuclear pore complex. *Genetics* **197**, 1213–1224 (2014).
24. Stage-Zimmermann, T., Schmidt, U. & Silver, P. A. Factors affecting nuclear export of the 60S ribosomal subunit in vivo. *Mol. Biol. Cell* **11**, 3777–3789 (2000).
25. Fischer, J., Teimer, R., Amlacher, S., Kunze, R. & Hurt, E. Linker Nups connect the nuclear pore complex inner ring with the outer ring and transport channel. *Nat. Struct. Mol. Biol*. **22**, 774–781 (2015).
26. Onischenko, E. et al. Natively unfolded FG repeats stabilize the structure of the nuclear pore complex. *Cell* **171**, 904–917.e19 (2017).
27. Yoshida, K., Seo, H.-S., Debler, E. W., Blobel, G. & Hoelz, A. Structural and functional analysis of an essential nucleoporin heterotrimer on the cytoplasmic face of the nuclear pore complex. *Proc. Natl Acad. Sci. USA* **108**, 16571–16576 (2011).
28. Andersen, K. R. et al. Scaffold nucleoporins Nup188 and Nup192 share structural and functional properties with nuclear transport receptors. *eLife* **2**, e00745 (2013).
29. Rout, M. P. et al. The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol*. **148**, 635–652 (2000).
30. Terry, L. J. & Wente, S. R. Nuclear mRNA export requires specific FG nucleoporins for translocation through the nuclear pore complex. *J. Cell Biol*. **178**, 1121–1132 (2007).
31. Weirich, C. S., Erzberger, J. P., Berger, J. M. & Weis, K. The N-terminal domain of Nup159 forms a β-propeller that functions in mRNA export by tethering the helicase Dbp5 to the nuclear pore. *Mol. Cell* **16**, 749–760 (2004).
32. Vallotton, P. et al. Mapping the native organization of the yeast nuclear pore complex using nuclear radial intensity measurements. *Proc. Natl Acad. Sci. USA* **116**, 14606–14613 (2019).
33. Otsuka, S. et al. Nuclear pore assembly proceeds by an inside-out extrusion of the nuclear envelope. *eLife* **5**, e19071 (2016).
34. Zhang, W. et al. Brr6 and Brl1 locate to nuclear pore complex assembly sites to promote their biogenesis. *J. Cell Biol*. **217**, 877–894 (2018).
35. Beck, M., Mosalaganti, S. & Kosinski, J. From the resolution revolution to evolution: structural insights into the evolutionary relationships between vesicle coats and the nuclear pore. *Curr. Opin. Struct. Biol*. **52**, 32–40 (2018).
36. Cantwell, H. & Nurse, P. Unravelling nuclear size control. *Curr. Genet*. **65**, 1281–1285 (2019).
37. Kukulski, W. et al. Correlated fluorescence and 3D electron microscopy with high sensitivity and spatial precision. *J. Cell Biol*. **192**, 111–119 (2011).
38. Arnold, J. et al. Site-specific cryo-focused ion beam sample preparation guided by 3D correlative microscopy. *Biophys. J*. **110**, 860–869 (2016).
39. Aitchison, J. D., Blobel, G. & Rout, M. P. Nup120p: a yeast nucleoporin required for NPC distribution and mRNA transport. *J. Cell Biol*. **131**, 1659–1675 (1995).
40. Shvets, E., Abada, A., Weidberg, H. & Elazar, Z. Dissecting the involvement of LC3B and GATE-16 in p62 recruitment into autophagosomes. *Autophagy* **7**, 683–688 (2011).
41. Sawa-Makarska, J. et al. Cargo binding to Atg19 unmasks additional Atg8 binding sites to mediate membrane-cargo apposition during selective autophagy. *Nat. Cell Biol*. **16**, 425–433 (2014).
42. Osawa, T. et al. Atg2 mediates direct lipid transfer between membranes for autophagosome formation. *Nat. Struct. Mol. Biol*. **26**, 281–288 (2019).
43. Schütter, M., Giavalisco, P., Brodesser, S. & Graef, M. Local fatty acid channeling into phospholipid synthesis drives phagophore expansion during autophagy. *Cell* **180**, 135–149 (2020).
44. Valverde, D. P. et al. ATG2 transports lipids to promote autophagosome biogenesis. *J. Cell Biol*. **218**, 1787–1798 (2019).
45. Thaller, D. J. et al. An ESCRT–LEM protein surveillance system is poised to directly monitor the nuclear envelope and nuclear transport system. *eLife* **8**, e45284 (2019).

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Yeast strains, growth conditions, fluorescence imaging

Strains used in this study are listed in Supplementary Table 1. Genotypes were verified using PCR and by selectable markers. Cells were grown at indicated temperatures (25 °C, 30 °C and 37 °C) to mid-log phase in YPAD (1% yeast extract, 2% peptone, 0.04% adenine, 2% dextrose). Nitrogen starvation was carried out by switching cells grown at OD_{600} ~0.5 with YPAD into synthetic minimal medium lacking nitrogen (SD-N medium) (0.17% yeast nitrogen base, without amino acids and ammonium sulfate, supplemented with 2% glucose) for indicated times.

Fluorescence microscopy

For fluorescence microscopy, yeast cells were grown in low-fluorescence synthetic growth medium (yeast nitrogen base without amino acids and without folic acid and riboflavin (FORMEDIUM)) supplemented with all essential amino acids and 2% glucose. The next day, cells were diluted to OD 0.1 and grown to mid log phase (0.5–0.8 OD) before imaging. Microscopy slides were pretreated with 1 mg ml^{-1} concanavalin A (ConA) solution. The wide-field imaging was performed at the Imaging Facility of the Max Planck Institute of Biochemistry (MPIB-IF) on a GE DeltaVision Elite system based on an OLYMPUS IX-71 inverted microscope, an OLYMPUS (100× 1.40 NA UPLSAPO and 60× 1.42 NA PLAPON) objective and a PCO sCMOS 5.5 camera. Images were deconvolved using the softWoRx Software (default values except: method, additive enhanced, 20 iterations). For Extended Data Fig. 8a, images were acquired using a wide-field fluorescence microscope (Olympus IX81) equipped with an Olympus PlanApo 100× 1.40 NA oil-immersion objective and a CCD camera (Orca-ER; Hamamatsu Photonics). All light microscopy figures were analysed in FIJI[46]. For quantifying nuclear size fold increase in Extended Data Fig. 6e, round nuclei from maximum-intensity projections were circled as shown in Fig. 3c to calculate area and the radius (r) using FIJI. The volume was extrapolated using the formula $(4/3) \times \pi r^3$. For fluorescent intensity quantifications in Extended Data Fig. 8a, cell background was subtracted from each intensity measurement.

Immunoblot techniques

We collected 3×10^7 cells (OD 1.5) at 25 °C and 30 °C (at 37 °C the *nup116Δ* strain is not viable under starvation conditions) for the indicated time points, and total cell protein extracts were obtained using alkaline lysis (2 M NaOH and 7.5% (v/v) 2-mercaptoethanol for 15 min on ice) followed by trichloroacetic acid precipitation (to a final concentration of 22% for 10 min on ice). Protein pellets were collected by centrifugation (14,000 rpm, 20 min), solubilized in HU loading buffer (8 M urea, 5% SDS, 200 mM Tris-HCl pH 6.8, 20 mM dithiothreitol and 1.5 mM bromophenol blue) and disrupted by vortexing with an equal volume of acid-washed glass beads (425–600 µm) for 6 min. The samples were then incubated at 65 °C (1,400 rpm) for 10 min. Proteins were separated using NuPAGE 4–12% gradient gels (Invitrogen), transferred onto polyvinylidene fluoride membranes (Immobilon-P) and then analysed using the specific antibodies. Monoclonal antibody against Dpm1 (1:10,000; clone 5C5A7) was purchased from Invitrogen. The monoclonal GFP antibody (1:500; clone B-2) was purchased from Santa Cruz Biotechnology. To calculate degradation of eGFP-fused protein levels (%), the intensity of the full-length eGFP-fused protein was normalized to the intensity of the respective loading control, and shown as relative to wild-type cells at 0 h. To calculate degradation of eGFP′ (eGFP resistant to vacuolar degradation) levels (fold change), the intensity of the free eGFP′ was divided by the intensity of the Dpm1 loading control; values are shown relative to that in WT cells (always normalized to the longest starvation time point of the WT cells).

Correlative fluorescence and electron tomography in plastic sections

CLEM analysis was conducted as previously described[37,45]. *S. cerevisiae* cells were high-pressure frozen after ~5 h of starvation (HPM010, AbraFluid) and freeze substituted (EM-AFS2, Leica) with 0.1% uranyl acetate in acetone for 55 h at −90 °C. The temperature was then increased to −45 °C at 3.5 °C h^{-1} and samples were further incubated for 5 h. After rinsing in acetone, the samples were infiltrated in Lowicryl HM20 and the resin was polymerized under UV light. Sections 300 nm thick were cut with a microtome (EM UC7, Leica) and placed on carbon coated 200 mesh copper grids (S160, Plano). The fluorescence microscopy imaging of the sections was carried out as previously described[45,47] using a wide-field fluorescence microscope (Olympus IX81) equipped with an Olympus PlanApo 100× 1.40 NA oil-immersion objective and a CCD camera (Orca-ER; Hamamatsu Photonics). After fluorescence imaging, grids were post-stained with uranyl acetate and lead citrate. Protein A-coupled gold beads (15 nm) were also added as fiducial markers used for overlaying low-mag with high-mag tomograms. Tilt series of the cells of interest (60°to 60°) were acquired semi-automatically on a Tecnai F30 (Thermofisher, FEI) at 300 kV with Serial-EM[48] at 20,000× and at 4,700× to facilitate ease of correlation. Tomograms were reconstructed automatically with IMOD batchruntomo[49]. Tilts were aligned by patch tracking. IMOD[49] was used for manual segmentation of the tomograms. Overlays of fluorescence spots and tomograms was performed with the ec-CLEM Plugin[50] in ICY[51] by clicking manually on corresponding pairs of notable features in the two imaging modalities.

2D and 3D electron microscopy

S. cerevisiae cells upon starvation were high-pressure frozen as described above and freeze substituted with 0.2% uranyl acetate 0.1% glutaraldehyde in acetone for 60 h at −90 °C. The temperature was then increased to −45 °C at 3 °C h^{-1} and samples were further incubated for 9 h. After rinsing in acetone, the samples were infiltrated in Lowicryl HM20 resin and the resin was polymerized under UV light. Grids were poststained as described above. Acquisition was performed with a Tecnai F30 (Thermofisher, FEI) at 300 kV. NPC-concealing hernia, NPCs and NPC-exposing hernia counting was performed using 3D tomographic volumes after acquisition of tilt-series from 300-nm sections with Serial EM[48]. See Supplementary Table 2 for number of tomograms for each strain. Box plots were drawn with Graphpad Prism 6.0. For NPC density calculations (NPCs per μm^2), nuclear surface was calculated in IMOD[49] after nuclei segmentation.

Cryo-FIB milling and cryo-CLEM

S. cerevisiae strains were grown in YPAD liquid medium to an OD_{600} of 0.2 at 30 °C, then transferred to SD-N liquid medium or to a 37 °C incubator accordingly. Per grid, 3.5 µl of cell suspension were applied on 200 mesh copper grids coated with R 2/1 holey carbon or SiO_2 films (Quantifoil Micro Tools) and plunge frozen in liquid ethane at ~−186 °C using a Leica EM GP grid plunger. The blotting chamber conditions were adjusted to 30 °C, 95% humidity and 1.5–2 s blotting time. For cryo-CLEM, Crimson FluoSpheres Carboxylate-Modified Microspheres, 1.0 µm, Crimson fluorescent (625/645), were washed in 1× PBS and added to the cell suspension at 1:20 dilution, and 1 s blotting time was used. The frozen grids were fixed in modified autogrids to allow milling at a shallow angle[52] and transferred into an Aquilos (cryo-FIB-SEM dual beam, Thermofisher). For correlative studies, the clipped grids were imaged on a prototype Leica cryo-confocal microscope based on Leica TCS SP8 CFS equipped with a Cryo Stage (similar to the commercially available EM Cryo CLEM Widefield system). Imaging was performed using a 50× objective, NA 0.90, 552-nm laser excitation, and detecting simultaneously at 560–620 nm and 735–740 nm. In the Aquilos, samples were sputter-coated with inorganic platinum and coated with an organometallic protective platinum layer using the Aquilos gas-injection

system[53]. Lamellae were produced using the Gallium ion beam at 30 kV and stage tilt angles of 17°–19° by milling in two parallel rectangular patterns. The lamella preparation was conducted in a stepwise fashion, gradually reducing the current of the ion beam until the final polishing of a thin slab of biological material of around 150–250 nm. For correlative studies, beads (5 to 10) were picked in the squares of interested and overlaid with fluorescence signals coming from the confocal stacks using 3DCT in both the electron and ion beam images[38]. After choosing the signal of interest in the confocal stacks, 3DCT provides the position to place the milling patterns in *x*, *y* and *z* coordinates. Milling is then performed as described above after choosing grid squares with intact film. After polishing, the signal of interest is retained in the lamellae. Autogrids with lamellae were unloaded and placed in storage boxes. In some cases, a further inorganic platinum layer was added to reduce charging during transmission electron microscopy imaging.

Cryo-electron tomography data collection and processing

Grids with lamellae were loaded into the Krios cassette. Cryo-electron tomographic tilt series (TS) were acquired on a Titan Krios (Thermofisher FEI) operating at 300 kV equipped with a Gatan K2 summit direct electron detector and energy filter. The autogrids were carefully loaded with the lamella orientation perpendicular to the tilt-axis of the microscope before TS acquisition. Lamellae were mapped at low mag (2 nm per pixel, 30 kV slit) and for correlative studies the maps were overlaid with the electron beam images from the Aquilos using serial EM registration points. Spots of interested were chosen accordingly. All data was collected using the K2 operating in dose-fractionation mode at 4k × 4k resolution with a nominal pixel size of 0.34 nm. TS collection was automated at using a modified version of the dose-symmetric scheme[54] taking the lamella pre-tilt into account. Defocus TS were acquired over a tilt range of +65 to −45 for a positive pre-tilt with a tilt increment of 2–3°, a total dose of ~140 e^{-} A^{-2} and a targeted defocus of around 2 to 4.5 µm.

All images were pre-processed and dose-filtered as described in[55]. Tilt-series alignment was performed using IMOD software package 4.9.2[49], by patch tracking function on bin4 image stacks. Initial tilt series alignment was manually inspected and improved by removing contours showing large deviations (that is, a large mean residual error) from the alignment model function. The software package gCTF[56] was used for CTF estimation. The 3dCTF correction and tomogram reconstruction was performed with NovaCTF[57].

Subtomogram alignment and averaging

Subtomogram alignment and averaging was performed with slight modifications from a previously described workflow[58] using the Matlab TOM package re-implemented in C++. In brief, particle coordinates and initial orientations were manually picked and assigned. Initial NPC alignment was performed on 8 and 4 times binned subtomograms followed by a manual inspection and curation of the initial alignment for each particle. From the whole aligned NPCs all 8 spokes were assigned according to an eightfold rotational symmetry and the subunits were extracted from the tomograms removing subunits which are located outside of the lamellae. Subunits were further aligned. The alignment was once more manually inspected and all misaligned subunits were removed. Final alignments were done binning two times the subtomograms using focused masks on the cytoplasmic, inner and nucleoplasmic ring of the NPC. In the *S. cerevisiae* wild-type dataset ~500 NPCs (4,000 asymmetric units) were initially picked from ~240 tomograms, ~250 NPCs (2,000 asymmetric units) from 145 tomograms for the *nup116Δ* strain at permissive temperature and ~60 NPCs from ~70 tomograms for the *nup116Δ* strain at not-permissive temperature (37 °C). The particles were split in two half datasets for gold-standard processing. After alignment and manual curation ~2,000 subunits were used in the final average for the wild-type NPC and ~1,300 for the *nup116Δ* at permissive temperature and 450 for the *nup116Δ* strain at 37 °C. For the latter NPCs, SIRT-like filtered volumes generated with IMOD[49] were used for alignment and weighted back projection volumes to get the final structure. Gold standard Fourier shell correlation (FSC) was calculated with the Fourier shell correlation server (www.ebi.ac.uk/pdbe/emdb/validation/fsc/results/) with half-maps as inputs. Local resolution was calculated with Resmap[59], B-factor sharpening was estimated empirically as in ref. [17]. Difference maps were calculated using USCF Chimera[60].

Systematic fitting of NPC components to the cryo-ET map

An unbiased systematic (global) fitting approach was performed using structural models of various *S. cerevisiae* NPC subcomplexes derived from previously published structures[3,15]. All structural models were low-pass filtered to 40 Å prior the fitting. The resulting model maps were then independently fitted into the NPC cryo-EM map from this study using global fitting as implemented in UCSF Chimera[60]. The fitting of the Y-complex structures and Nup159 complex was performed for CR and NR segments of the in-cell cryo-EM NPC map. The nuclear envelope density was erased before the fitting with the tool 'volume eraser' from UCSF Chimera[60]. The regions of the nuclear envelope distant from apparent contact points between membrane and protein densities were erased before the fitting to eliminate fits overlapping with the membrane. All fitting runs were performed using 100,000 random initial placements and the requirement of at least 30% of the model map to be covered by the NPC density envelope from this study defined at low threshold. For each fitted model, this procedure yielded between 500 and 17,000 fits with unique conformation after clustering.

For fitting the filtered atomic models, the cross-correlation about the mean (cam score, equivalent to Pearson correlation) score from UCSF Chimera[60] was used as a fitting metric as in our previous works[13,17,55]:

$$\mathrm{cam} = \frac{\langle \mathbf{u} - \mathbf{u}_{\mathrm{ave}}, \mathbf{v} - \mathbf{v}_{\mathrm{ave}} \rangle}{|\mathbf{u} - \mathbf{u}_{\mathrm{ave}}||\mathbf{v} - \mathbf{v}_{\mathrm{ave}}|}$$

where $\mathbf{u}_{\mathrm{ave}}$ is a vector with all components equal to the average of the components of $\mathbf{u}$ and $\mathbf{v}_{\mathrm{ave}}$ is defined analogously.

The negative-stain map of the Nup159 complex was fitted using UCSF Chimera's cross-correlation about zero (equivalent to cosine similarity) score (cc):

$$\mathrm{cc} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{|\mathbf{u}||\mathbf{v}|}$$

This score was used instead of the cam score since the cam score tests the linear dependence of the two fitted maps, which is not the case when fitting negative-stain maps, which contain the signal of the electron density at the surface of the complex.

The Nup159 complex was also fitted using the Colores program from the Situs package[61] with settings appropriate for low-resolution negative-stain maps (that is, using Laplacian filtering that emphasizes on contour matching over interior volume matching).

For each fitting run, the statistical significance of the fits was assessed as a *P* value calculated from the normalized cross-correlation scores. To calculate the *P* values, the cross-correlation scores were first transformed to *z*-scores (Fisher's *z*-transform) and centred, from which two-sided *P* values were computed using standard deviation derived from an empirical null distribution (derived from all obtained unique fits and fitted using fdrtool[62] R-package). All *P* values were corrected for multiple testing using Benjamini–Hochberg procedure[63]. Figures were made using UCSF Chimera[60] and Xlink Analyzer[64].

Integrative modelling

To build the integrative model of *S. cerevisiae* NPC, the IR model from ref. [15] was first fitted as a rigid body using the procedure described above (systematic fitting). The fit was statistically significant but

was nevertheless further refined to optimize the fitting of individual domains. For the refinement, the structure of the IR was divided into rigid bodies corresponding to the original crystal structures used as modelling templates from ref. [15] and the positions of the rigid bodies were optimized using a custom refinement protocol implemented with Integrative Modelling Platform (IMP)[65] v.2.12.0, similar to the procedure used in ref. [66]. The protocol employs a Monte Carlo simulated annealing optimization of the rigid bodies' rotations and translations and an integrative scoring function, which consisted of the linear combination of the following restraints: (1) cross-correlation to the EM map (FitRestraint of IMP), (2) connectivity distance between domains neighbouring in sequence, (3) clash score (SoftSpherePairScore of IMP), and (4) a soft restraint favouring approximate twofold symmetry. The structures were simultaneously represented at two resolutions: in Cα-only representation and a coarse-grained representation, in which each ten-residue stretch was converted to a bead. The ten-residue bead representation was used for the clash score and the electron microscopy and symmetry restraint to increase computational efficiency, the Cα-only representation was used for domain connectivity restraints. Five-hundred independent Monte Carlo optimizations were run until convergence and the best scoring model was selected as the final model. The models of the Y-complex fitted into CR and NR could not be obtained by rigid body fitting of the published models[15]. Therefore, we have divided the available crystal structures and homology models[15] into smaller (six for the CR and seven for the NR Y-complex) rigid bodies (with cut points corresponding to boundaries of published crystal structures of Y-subcomplexes) and fitted them simultaneously into the NR and CR segments of the cryo-EM *S. cerevisiae* NPC map from this study using the integrative modelling procedure described previously[55,66] using IMP[65] version develop-54df7231ec. At first, each of the rigid bodies has been independently fitted to the NPC cryo-EM map using UCSF Chimera[60] as described above to generate libraries of alternative non-redundant fits for each rigid body. Then, we generated configurations of all the rigid bodies by recombining the aforementioned fits using simulated annealing Monte Carlo optimization. Each configuration was generated by an independent Monte Carlo optimization comprising 60,000 steps resulting in total 20,000 models and scored. The scoring function for the optimization was a linear combination of the EM fit restraint represented as the *P* values of the precalculated domain fits (from systematic fitting as described above), domain connectivity restraint, a term preventing overlap of the Y components with the nuclear envelope and the assigned Nup159 complex density, and clash score (see ref. [55]. for the implementation details). The structures were simultaneously represented at two coarse-grained resolutions 1 and 10 as above for the refinement of the IR. Since the *P* values were derived from the original EM fits generated with UCSF Chimera, the EM restraint can be regarded as an EM restraint derived from the full atom representation. For the CR, the top scoring model was selected for visualization and proposed as final. For the NR, the final model was additionally refined using the above refinement procedure and additional restraints to anchor membrane-binding domains to the nuclear envelope (see below) and the second best scoring refined model was selected for visualization (the top scoring model exhibited a clearly erroneous displacement of one of the rigid bodies, concerning the Nup84–Nup133 interface, but had similar score and conformation of other rigid bodies).

The maps of the NPCs in the mutant strains were built by first rigidly fitting the Y and IR models of the wild-type NPC and then refining the model using the refinement protocol as for the IR in the wild-type NPC. For the *nup116Δ* 25 °C NPC map, both Y-complexes and all four IR subcomplexes (the complexes of Nup(157–170)–Nup(188–192)–Nic96–Nsp1–Nup57–Nup49) could be fitted at the rigid body fitting step and were used for the refinement. In the case of the *nup116Δ* 37 °C NPC map, only the outer nuclear copy out of the four IR subcomplexes fitted to the IR with a poor fit of the Nsp1–Nup57–Nup49 coiled-coil. Therefore, only the single copy of Nup157–Nup188–Nic96 was refined and included in the final model. Additional restraints were used that (1) favoured preservation of the WT inter-subunit interactions using an elastic network of harmonic distance restraints between interface beads of interacting domains, (2) penalized overlap of protein mass with the nuclear envelope. Also, no symmetry restraint was used for the knockout maps. For the *nup116Δ* 37 °C NPC map, an additional restraint favouring interactions of the Nup157 β-propeller with the nuclear envelope was used (derived using MapDistanceTransform of IMP).

Supplementary Videos 1, 2, 4 were created with UCSF Chimera[60]. Supplementary Videos 3, 5, 6–10 were prepared with IMOD[49] and Movavi Video Editor 15.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The three cryo-EM maps associated with this manuscript have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-10198, EMD-10660 and EMD-10661. The integrative models of S. *cerevisiae* NPC are available at Zenodo at https://doi.org/10.5281/zenodo.3820319 and the PDB-Dev database under accession numbers PDBDEV_00000051, PDBDEV_00000052 and PDBDEV_00000053. Unprocessed western blots are available in Supplementary Fig. 1. Because of their size, original imaging data are available from the corresponding author upon request.

Code availability

The input data and the scripts used for modelling are available at Zenodo at https://doi.org/10.5281/zenodo.3820319.

46. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
47. Hampoelz, B. et al. Pre-assembled nuclear pores insert into the nuclear envelope during early development. *Cell* **166**, 664–678 (2016).
48. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
49. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
50. Paul-Gilloteaux, P. et al. eC-CLEM: flexible multidimensional registration software for correlative microscopies. *Nat. Methods* **14**, 102–103 (2017).
51. de Chaumont, F. et al. Icy: an open bioimage informatics platform for extended reproducible research. *Nat. Methods* **9**, 690–696 (2012).
52. Rigort, A. et al. Focused ion beam micromachining of eukaryotic cells for cryoelectron tomography. *Proc. Natl Acad. Sci. USA* **109**, 4449–4454 (2012).
53. Schaffer, M. et al. Optimized cryo-focused ion beam sample preparation aimed at in situ structural studies of membrane proteins. *J. Struct. Biol.* **197**, 73–82 (2017).
54. Hagen, W. J. H., Wan, W. & Briggs, J. A. G. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 (2017).
55. Kosinski, J. et al. Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363–365 (2016).
56. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
57. Turoňová, B., Schur, F. K. M., Wan, W. & Briggs, J. A. G. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4Å. *J. Struct. Biol.* **199**, 187–195 (2017).
58. Beck, M., Lucić, V., Förster, F., Baumeister, W. & Medalia, O. Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. *Nature* **449**, 611–615 (2007).
59. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
60. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
61. Wriggers, W. Conventions and workflows for using Situs. *Acta Crystallogr. D* **68**, 344–351 (2012).
62. Strimmer, K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **24**, 1461–1462 (2008).
63. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
64. Kosinski, J. et al. Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.* **189**, 177–183 (2015).
65. Webb, B. et al. Integrative structure modeling with the Integrative Modeling Platform. *Protein Sci.* **27**, 245–258 (2018).

66. Dauden, M. I. et al. Architecture of the yeast Elongator complex. *EMBO Rep*. **18**, 264–279 (2017).

Acknowledgements We thank J. Baumbach for critical reading of the manuscript; the members of the Beck and Mahamid laboratories and M. Schorb for invaluable input and support; D. Thaller, P. Lusk and E. Hurt for critical discussions and for providing *S. cerevisiae* strains; J. Sun for help in tomographic segmentation; and W. Wriggers for help with the Situs software. We thank Leica for a collaboration to develop the prototype cryo-confocal. We acknowledge support by EMBL's electron microscopy core facility (EMCF) and IT Services. M.A. was funded by an EMBO a long-term fellowship (ALTF-1389–2016); J.M. received funding from the European Research Council (ERC 3DCellPhase 760067); H.K.H.F. is supported by a fellowship from the EMBL Interdisciplinary (EI3POD) programme under Marie Skłodowska-Curie Actions COFUND (664726); B.P. acknowledges funding by Max Planck Society; F.W. was supported by an EMBO Long-Term Fellowship ALTF 764-2014; and M. Beck acknowledges funding by EMBL, the Max Planck Society and the European Research Council (ComplexAssembly 724349).

Author contributions M.A. conceived the project, designed experiments, collected cryo-EM, electron microscopy, CLEM and light microscopy data, analysed data and wrote the manuscript. C.E.Z. designed experiments, collected cryo-EM data, analysed data and wrote the manuscript. V.R. performed integrative modelling, analysed data and wrote the manuscript. F.W. designed experiments, collected light microscopy and functional autophagy data and analysed data. P.R. designed experiments, collected CLEM and electron microscopy data and analysed data. H.K.H.F. collected cryo-confocal data and analysed data. X.Z. collected cryo-confocal data. C.-W.L. collected functional autophagy data and analysed data. M. Börmel collected electron microscopy data. W.H. collected cryo-EM data. B.T. analysed subtomograms. K.K. performed integrative modelling. C.W.M., J.M. and Y.S. supervised the project. B.P., J.K. and M. Beck conceived the project, designed experiments, supervised the project and wrote the manuscript.

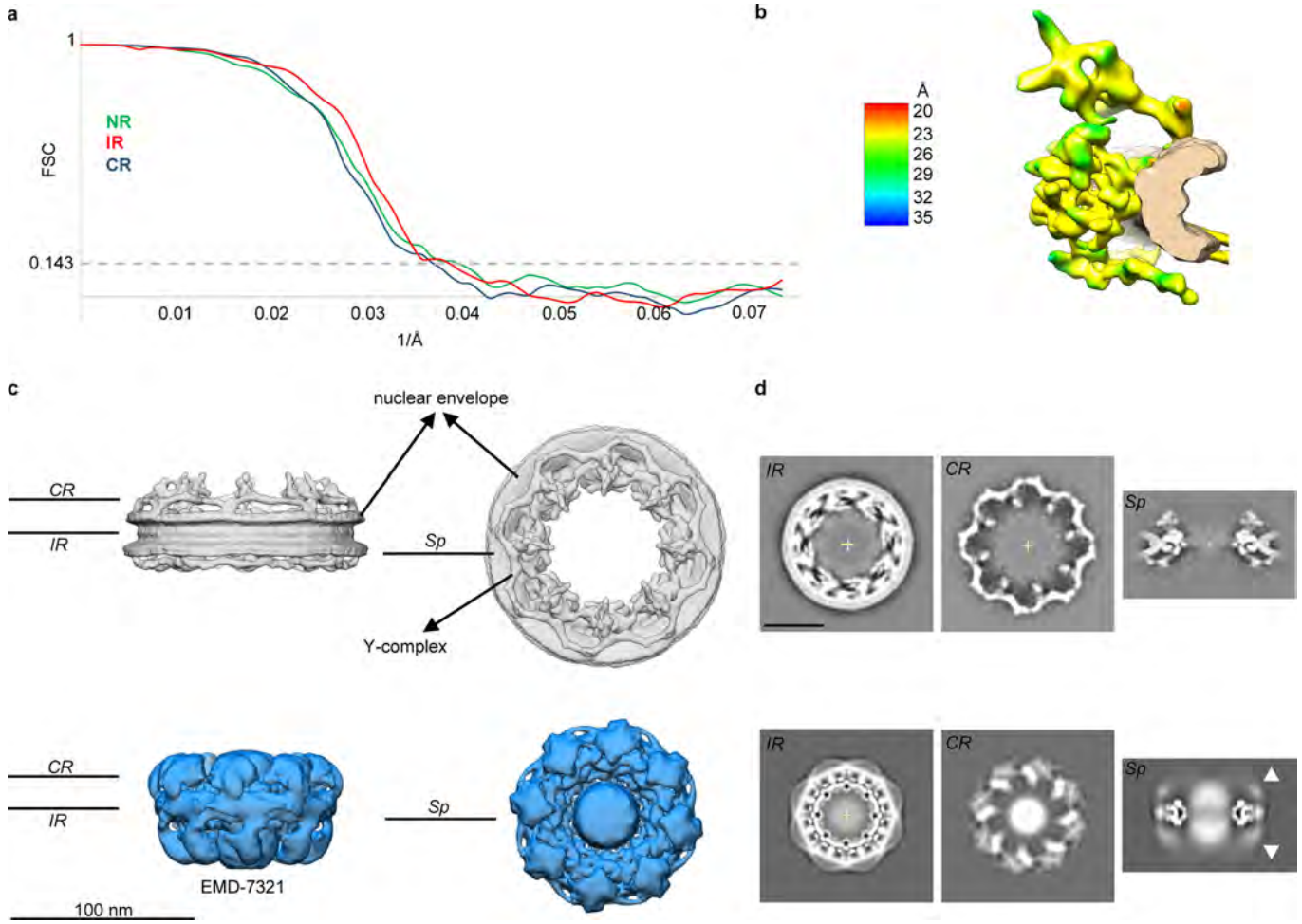
Competing interests The authors declare no competing interests.

Additional information

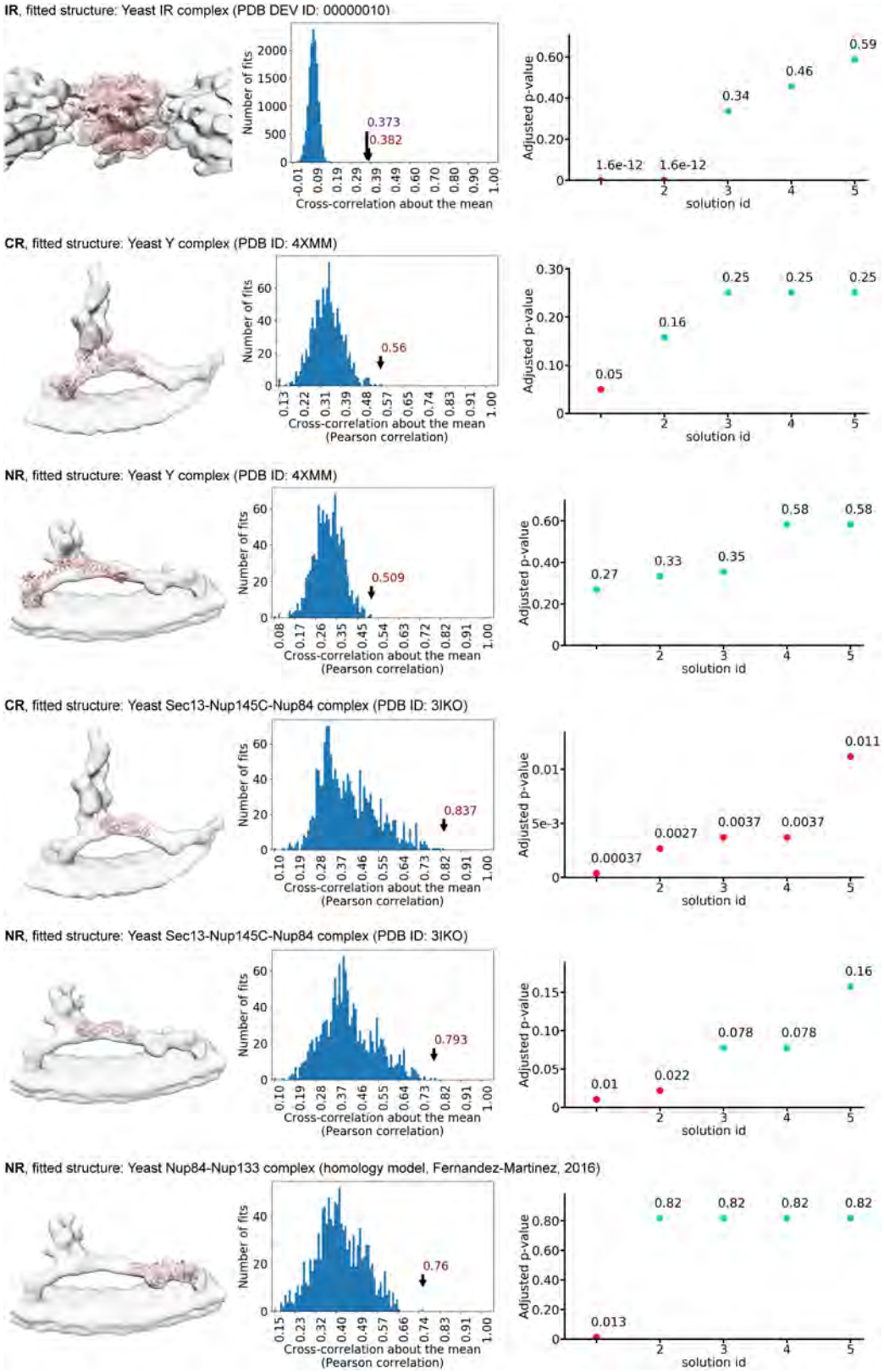
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2670-5.

Correspondence and requests for materials should be addressed to B.P., J.K. or M.B.

Peer review information *Nature* thanks James Hurley and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

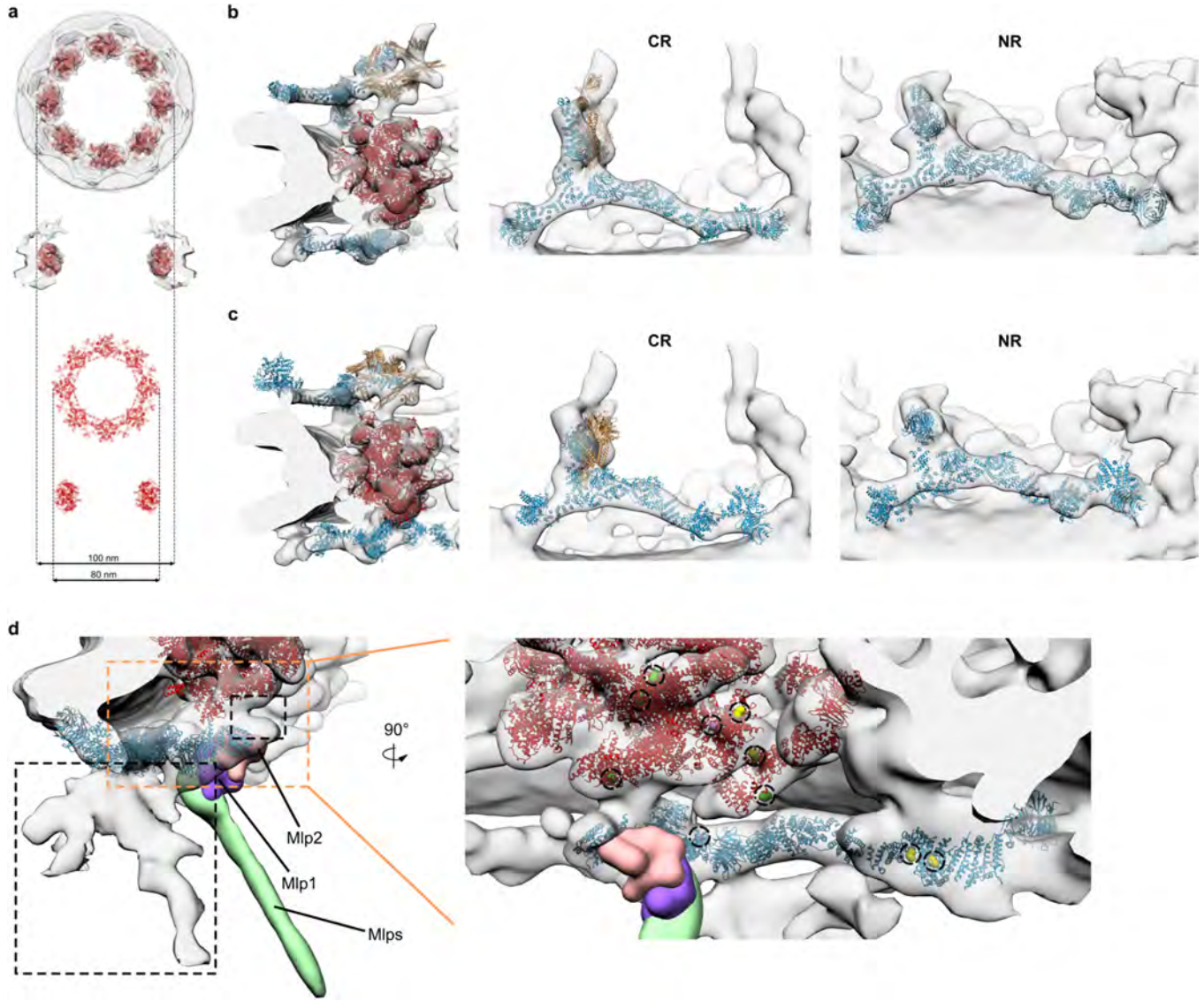


Extended Data Fig. 1 | In-cell structure of the *S. cerevisiae* NPC vs detergent-extracted NPC (EMD-7321). a, Gold standard FSCs of in-cell cryo-EM map of the *S. cerevisiae* NPC. All the curves (nuclear ring NR, cytoplasmic ring CR, inner ring IR) intersect the 0.143 criterium at around 25 Å resolution. **b**, Local resolution analysis[59] with colour-coded bar. **c**, In-cell cryo-EM map (grey, average of $n = 250$ NPCs) in comparison to cryo-EM map of detergent-extracted *S. cerevisiae* NPCs (blue, EMD-7321 at the suggested contour level) show significant differences in diameter and in interpretable features. The nuclear membranes and the Y-complexes are clearly discerned in the in-cell cryo-EM map, in contrast to EMD-7321. **d**, Tomographic slices through the maps shown in **c** at the level of the cytoplasmic (CR) and inner rings (IR) and an individual spoke (Sp). Lines in **c** indicate slicing position shown in **d**. Arrowheads indicate blurred features in the outer rings. Scale bar: 50 nm.

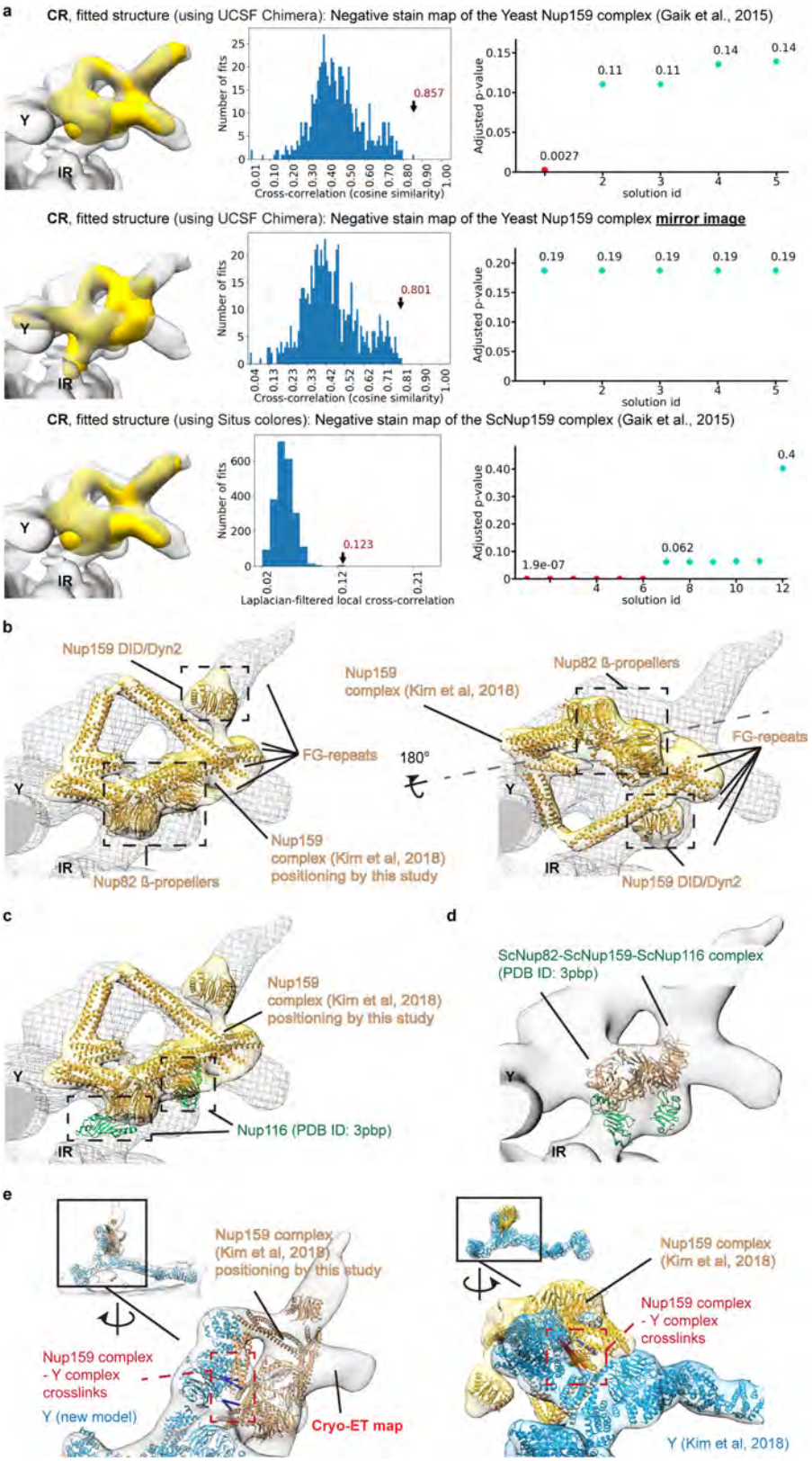


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Systematic fitting of inner and outer ring components into the *S. cerevisiae* NPC map. Each row shows the visualization of the top fits (left), the histogram of raw scores (middle), and a plot of the top five *P* values (right). In the *P* value plots, the statistically significant fits are coloured red ($P < 0.05$; *P* values were calculated using the two-sided test as implemented in fdrtool R-package, see Methods). All *P* values were adjusted for multiple comparisons using Benjamini-Hochberg procedure. The top fits are indicated in the histograms with an arrow and the score value. The number of sampled fits used to calculate *P* values after clustering of similar solutions was 14,348, 1,015, 1,039, 1,479, 1,354, and 1,183 for the rows from top to bottom. For the IR, the integrative model of the single spoke of the IR[15] was used as the fitted structure. For the outer rings (CR and NR), the crystal structure of the yeast Y-complex was fitted or its parts corresponding to subcomplexes. The structures were fitted by an unbiased global search using UCSF Chimera[60] and scored using the cross-correlation score about the mean as explained in the Methods. The IR complex was fitted to the entire spoke map, while the other structures were fitted to individual CR and NR segments.

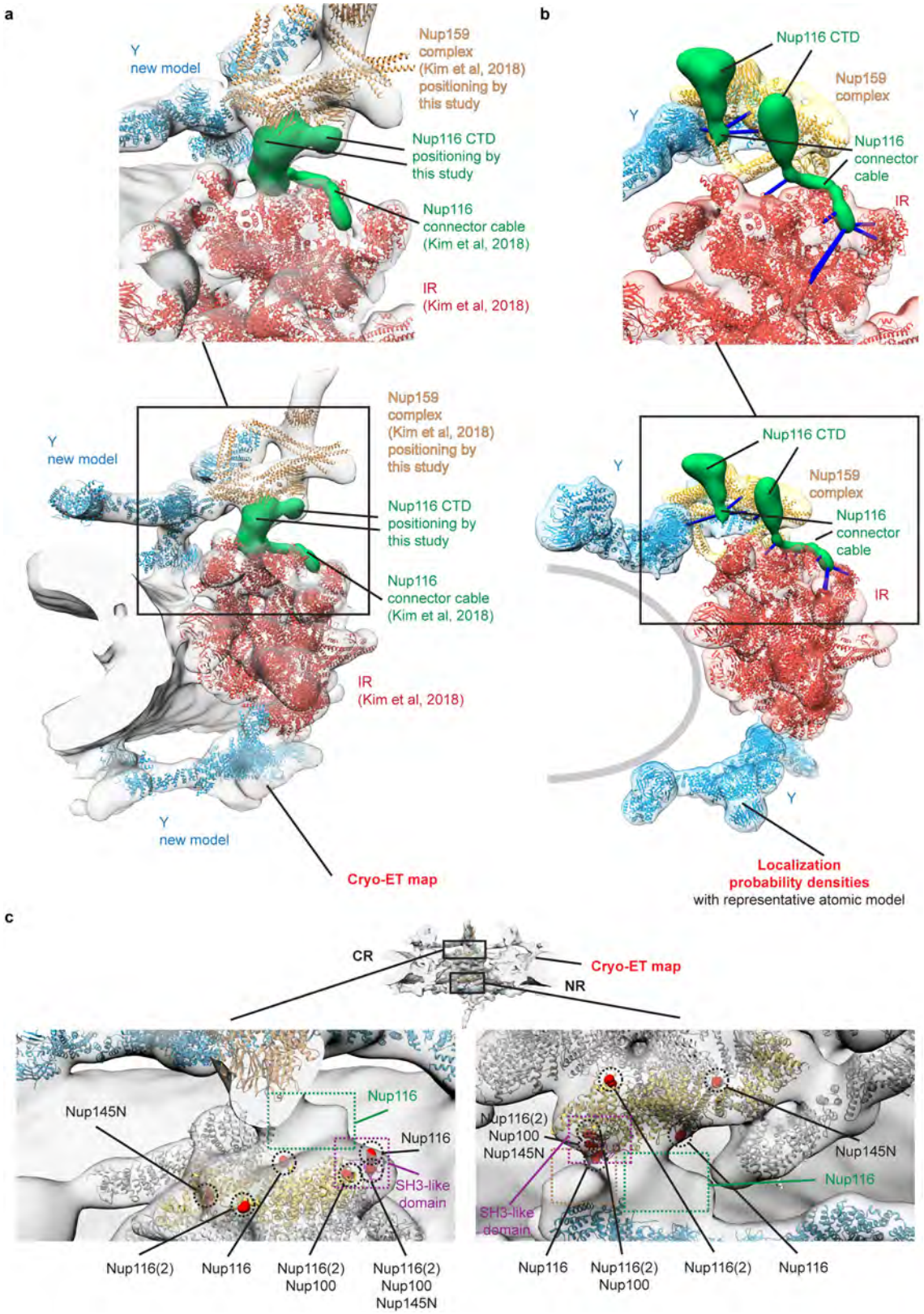


Extended Data Fig. 3 | Architectural model of *S. cerevisiae* NPC. **a**, Comparison of fitted (to the *S. cerevisiae* NPC map from this study, depicted in grey) integrative inner ring complex models (red ribbons) from[15] with 20 nm diameter difference. **b**, Representative integrative structural models of CR and NR Y-complex (blue ribbons) built in this work (see Methods) shown with the integrative model of P-complex (yellow ribbons) and refined IR (red ribbons) from[15] fitted to the allocated density of the in-cell map (grey density) from this study. The Y-complexes are more extended as compared to reference[15] version by ~40 Å. **c**, Representative integrative structural models of CR and NR Y-complex (blue ribbons), P-complex (yellow ribbons) and IR (red ribbons) from[15] (PDB-Dev ID: PDBDEV_00000010) fitted to the *S. cerevisiae* NPC map (grey density) from this study, with respect to spatial reference frame from[15]. **d**, Representative integrative structural model of NR Y-complex (blue ribbons) built in this work shown together with the integrative model of the refined IR (red ribbons) from[15] fitted to the allocated density of the *S. cerevisiae* NPC map (grey density). The localization probability densities of Mlp1, Mlp2 and Mlp1/Mlp2 ensemble from[15] are displayed in orientation relative to the NR Y-complex as in[15]. Local mask refinement of $n = 250$ NPCs recovers extra densities in the NR that locate around the Nup120-Nup145C junction and the Nup107-Nup133 stem (left dark dotted frame). Most likely these densities account for parts of the basket since it resembles a filamentous structure[15]. It anchors to the regions of the Y-complex that crosslink to Mlp1 or Mlp2[15] and connects to the Nup85-Seh1 arm as proposed by the previous model[15], but contrary to[15], the filamentous region centres at the Y-complex vertex rather than at the Nup85 arm. On the right panel residues crosslinking to Mlp1 or Mlp2[15] are shown in sphere representation and coloured according to their confidence score (from[15]) with yellow – $P < 0.01$, green – $P < 0.1$, magenta $P > 0.1$. The shown residues, together with the yet unmodelled crosslink to residue 2 of Nup85[15], emphasize that the highest confidence crosslinks are associated to the Y-complex, while the lowest confidence ones are associated to the IR. A second density (right dark dotted frame) remains unassigned, being proximate to residues crosslinking with Nup116, Nup100, Nup145N (see Extended Data Fig. 5c).



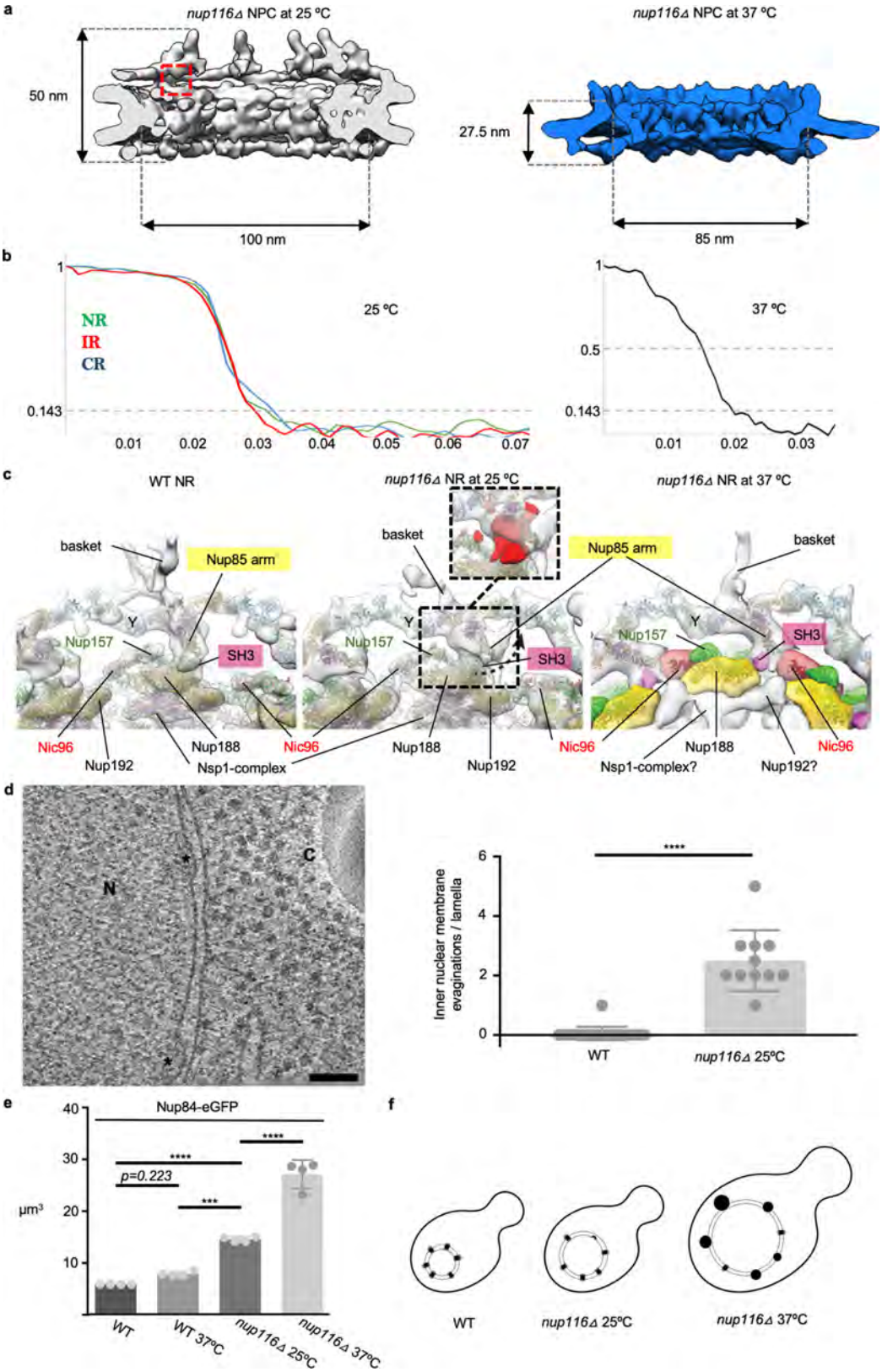
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Validation of the orientation of the Nup159 complex. **a**, Systematic fitting of the negative stain map of the Nup159 complex (yellow) into the *S. cerevisiae* NPC map (grey) using UCSF Chimera[60] and Colores program from the Situs package[61]. As a negative control, also the mirror image of the negative stain map was fitted but did not lead to significant scores, further underlining the unambiguous nature of the fits. Each row shows the visualization of the top fits (left), the histogram of raw scores (middle), and a plot of the top five *P* values (right). In the *P* value plots, the statistically significant fits are coloured red ($P < 0.05$; *P* values were calculated using the two-sided test as implemented in fdrtool R-package, see Methods). All *P* values were adjusted for multiple comparisons using Benjamini-Hochberg procedure. The top fits are indicated in the histograms with an arrow and the score value. The number of sampled fits used to calculate *P* values after clustering of similar solutions was 585, 599, and 2243, for top, middle, and bottom rows respectively. **b**, Representative integrative Nup159 complex model from[15] inside the in-cell *S. cerevisiae* NPC map (grey mesh) in the orientation determined in this work (left) versus the previously published orientation (right). The Nup159 complex model is shown in orange ribbons within yellow localization probability density from[15] locally fitted with UCSF Chimera[60]. The previous orientation was reproduced by first fitting the entire model from[15] to the in-cell cryo-EM map and then locally fitting the Nup159 complex to the density (which was needed to bring the Nup159 complex into the density and preserve the orientation). The dashed grey line indicates the flipping axis between the two fits. **c**, Superimposition of the crystal structure 3PBP[27] onto the Nup82 β-propellers from the representative integrative Nup159 complex model from[15] in the revised orientation predicts the position of Nup116, as confirmed by our knockout study (Fig. 2c). **d**, Visualization of two of the top resulting systematic fits of the 3PBP crystal structure into the cryo-ET map presented in this study confirms our *nup116Δ* structure (Fig. 2c). **e**, Crosslinks between the Nup159 complex and the Y-complex from[15] support the new orientation (left) compared to the published orientation (right). Satisfied and violated crosslinks are depicted as blue and red bars respectively while the Nup159 complex and the Y-complex from[15] are depicted within the relevant localization probability densities from[15].



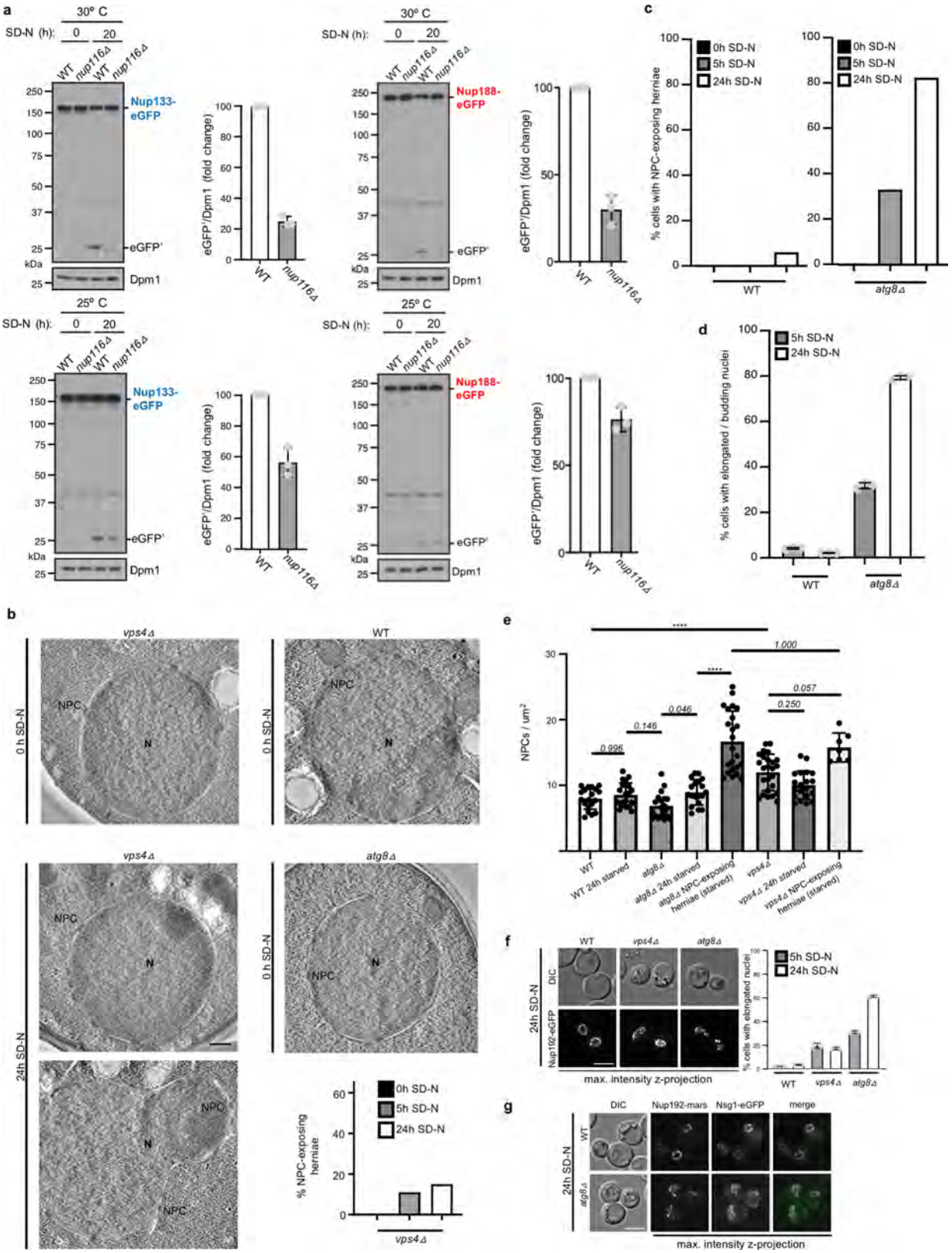
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Nup116 positioning and possible Nup188 SH3-like domain interactions. a, New positioning of the Nup116 versus **b**, previously published integrative model (PDB-Dev ID: PDBDEV_00000010 from[15], right). The two *S. cerevisiae* NPC models were superimposed such that the IRs are aligned to the same reference frame. The Nup116 position is shown either as the density assigned to Nup116 based on the Nup116 knockout structure in the CR (**a**) or as localization probability densities retrieved from reference[15] model (**b**). The major structural elements of the NPC are indicated. The Nup116 connector cable in the in-cell model (**a**) has been taken from reference[15] based on its position relatively to the IR. Blue bars represent crosslinks from Nup116 to other Nups[15]. For the in-cell model, the cryo-ET map is displayed; for the model from[15], the localization probability densities (not an EM map) are shown instead. **c**, SH3-like domain (magenta ribbons and dotted frames) of Nup188[28] (yellow ribbons) bridges the interactions between the inner and the outer rings. The positions of Nup188 crosslinking to Nup116 (the number of times that some Nup188 residues crosslink with more than one Nup116 residues are denoted in the labels in parenthesis), Nup100 and Nup145N are indicated in sphere representation (red spheres) and their location suggests that they link the connecting interfaces between the IR and the outer rings: NR brown dotted frame corresponds to an unassigned density, CR green dotted frame corresponds to Nup116 density, NR green dotted frame corresponds to a second Nup116 density, see Extended Data Fig. 6c.



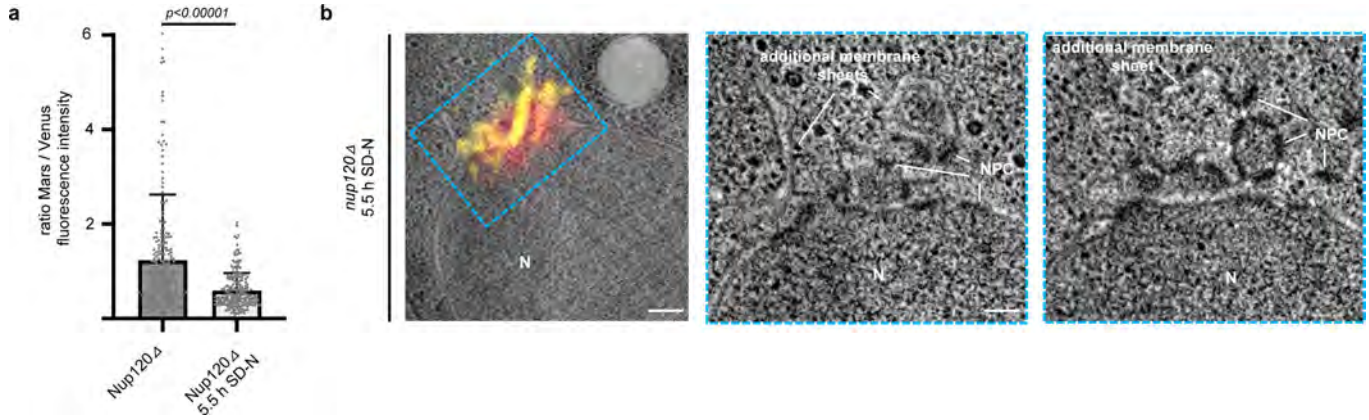
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | *nup116Δ* NPC structure and nuclear envelope remodelling. a, In-cell structure of *nup116Δ* NPC cut in half along the central axis, at 25 °C on the left and at 37 °C on the right. At 25 °C diameter and general dimensions are alike the WT structure of Fig. 1b. The red dashed square indicates the position of the missing densities at neck region corresponding to Nup116 (see also Fig. 2c and Extended Data Fig. 5a). **b**, FSCs of *nup116Δ* NPC at 25 °C after masking the three rings (NR, IR, CR respectively nuclear, inner and cytoplasmic ring). The curves intersect 0.143 at around 35 Å resolution. On the right FSC of *nup116Δ* NPC at 37 °C calculated with a mask enclosing both the IR and the NR. The average was performed with a pixel size of 13.8 Å (original pixel size 4 times binned). The resolution is ~50 Å. **c**, Structural differences in the NR of the *S. cerevisiae* NPC from WT (left), and *nup116Δ* cells at the permissive (25 °C, middle) and non-permissive (37 °C, right) temperatures. The grey envelope represents the overlay of EM maps generated with two different masks – one enclosing the entire NR and IR and the second centred at the region of the extra densities around the Y-complex. The regions of the major differences are indicated with dashed frames. A shift of the Nup85 arm away from the IR is indicated with a black dashed arrow. The difference density around the Nup85 arm between the NPC maps from *nup116Δ* 25 °C and WT cells is shown in the inset in red and indicates a putative nuclear copy of Nup116 and the conformational shift of the Nup85 arm. Due to the low resolution, the IR of the *nup116Δ* 37 °C is shown as segmented densities assigned based on the fitting of the outer nuclear copy of the IR subcomplex. Question marks indicate predicted assignments based on similarity to the WT map but with a poor density insufficient for fitting. The HideDust tool of UCSF Chimera[60] was used for clarity. **d**, Cryo-tomographic slice of nuclear envelope with inner nuclear membrane evaginations marked with asterisks in *nup116Δ* NPC at 25 °C. The histograms show that the number of inner nuclear membrane evaginations is significantly higher in the nuclear envelope of the mutant cells (32 in $n = 145$ cryo-tomograms, 42 lamellae) in comparison to WT envelopes (1 in $n = 240$ cryo tomograms, 100 lamellae); centre values represent the mean and error bars the SD; $P < 0.0001$ (Mann–Whitney test, two-tailed); N marks the nucleus and C marks cytoplasm, scale bar: 100 nm. **e**, Box plot showing the increase in nuclear size in *nup116Δ* cells in comparison to WT cells (~2.5-fold increase in volume) and a 4 to fivefold increase in comparison to WT upon shift to 37 °C. Light microscopy data of Nup84-eGFP, shown in Fig. 3c, were used to quantify the difference in nuclear volume as explained in the methods section. $n = 100$ nuclei were measured for each strain from four independent biological replicates (filled circles represent averages of the independent biological replicates). The statistical significance was evaluated with by one-way ANOVA with Dunn's multiple comparisons test; centre values represent the mean and error bars the SD; $^{***}P < 0.0002$, $^{****}P < 0.0001$). **f**, Cartoon model depicting the summary of the results coming from Fig. 3c, d, e and Extended Data Fig. 6d, e. NPC are represented as black cylinders, NPC-concealing herniae as black circles, a nuclear membrane evagination is shown as semicircle.

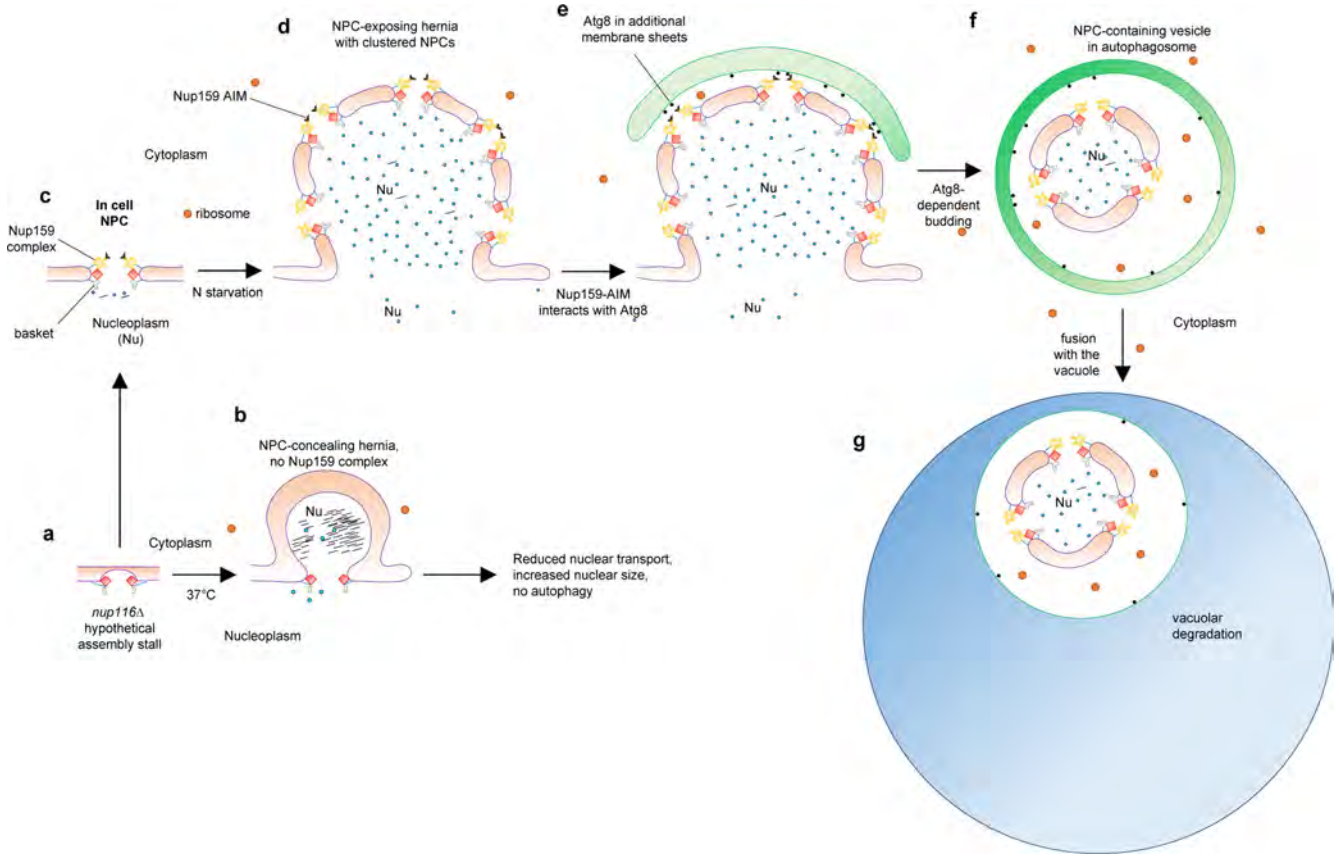


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | NPC degradation and nuclear envelope reshaping upon starvation. a, western blot and quantification of the degradation of Nup133, Nup188 upon Nitrogen starvation at 25 °C and 30 °C in WT and *nup116Δ* strain. eGFP was measured using anti-eGPF immunoblotting. eGFP′ denotes vacuolar eGFP remnant. Dpm1 was used as a loading control. Centre values represent the mean and error bars the s.d. of $n = 3$ biologically independent replicates. At 30 °C, Nup133, Nup188 are degraded fourfold less in comparison to WT. **b**, Typical tomographic plastic section of WT, *atg8Δ S. cerevisiae* nuclei after 0h of starvation (negative control of Fig. 4c) or of *vps4Δ S. cerevisiae* nuclei after 0 or 24 h of starvation. Quantification of number of NPC-exposing herniae at 0, 5 and 24 h of starvation is shown for *vps4Δ* cells. We show an example of a round nucleus and of an NPC-exposing hernia from those cells. $n = 85$ tomograms of nuclei for WT; $n = 84$ for *vps4Δ*; $n = 83$ for *atg8Δ*; $n = 62$ for *vps4Δ* 24h starved; $n = 64$ for *vps4Δ* 5h starved; see Supplementary Table 2. N marks the nucleus. Scale bar, 200 nm. **c**, Quantification of of NPC-exposing herniae from plastic section tomograms as in Fig. 4c at 5h and 24h of starvation. After 24h, 82% of nuclei in the *atg8Δ* cells contain NPC-exposing herniae ($n = 78$ tomograms in average for each condition, see Supplementary Table 2). **d**, Quantification of deconvoluted wide-field maximum intensity projection images as in Fig. 4d from three biologically independent replicates and $n = 300$ per replica. **e**, Quantification of NPC density (NPCs/μm^2) from plastic sections tomograms as in **c** shows clustering of NPCs in NPCs-exposing hernia. The statistical significance was evaluated with one-way ANOVA with Dunn's multiple comparisons test ($n = 20$ tomograms per strain except *vps4Δ* NPC-exposing herniae where $n = 5$ tomograms; the average nuclear envelope surface measured per strain is 20 μm^2, except *atg8Δ* NPC-exposing herniae where is 11 μm^2 and *vps4Δ* NPC-exposing herniae where is 3.5 μm^2; $^{****}P < 0.0001$; centre values represent the mean and error bars the SD). **f**, Phase contrast and deconvoluted wide-field max intensity projection of WT, *vps4Δ* and *atg8Δ* cells imaged live before and after 24h of starvation with eGFP-tagged Nup192 as marker. The quantification is derived from three independent biological replicates where five images with at least $n = 250$ nuclei per replicate; central bars represent the mean and error bars the s.d. The trend of nucleus deformation is the same seen in **b**. **g**, Same as **c** for WT and *atg8Δ* cells, but with the inner nuclear membrane marker Nsg1. Data are derived from three independent biological replicates with at least $n = 250$ nuclei per replicate.



Extended Data Fig. 8 | Interaction Nup159-Atg8 upon starvation. **a**, Fluorescence intensity analysis of *nup120Δ*, Nup170-Mars, Nup159-Atg8-split Venus nuclei spots before and after 5.5 h of starvation. The ratio Mars/Venus shows the significant increase in Venus signal as compared to Mars signal in this starved strain. Three biological replicates and $n = 100$ spots per replicate were measured ($P < 0.00001$, Mann–Whitney, two-sided; box centres represent the mean and error bars the s.d.). **b**, Tomographic slice (plastic section) overlaid with wide-field fluorescent image obtained by on section-CLEM[37,45] of the strain in **a** upon ~6 h of starvation. $n = 36$ correlated tomograms (technical replicates) were acquired and 24 (75%) had similar results. 8 spots come from uranyl autofluorescence in the cytoplasm. Venus signal is shown in yellow. Nup170-Mars is shown in red. The cyan dashed rectangle indicates the area zoomed into the two right panels showing tomographic slices at two different Z-heights (N marks the nucleus, scale bar: 200 nm). See also Supplementary Videos 9, 10.



Extended Data Fig. 9 | Model of membrane remodelling in NPC-concealing and NPC-exposing hernia. a, Plausible inside-out assembly intermediate[33] or inner nuclear membrane evaginations (Extended Data Fig. 6d) progress into **b**, NPC-concealing herniae (Fig. 3a, Supplementary Video 3) which accumulate over time (Fig. 3d), concomitantly with increasing nuclear size (Extended Data Fig. 6e;[3]). Black lines indicate dark material present in cryo tomograms, possibly poly-A mRNA[3]. **a** matures into **c**, fully assembled NPCs as structurally analysed in cells frozen in exponential growth phase (Fig. 1b, c, Extended Fig. 3d). **d**, When autophagy is triggered by nitrogen starvation, NPCs cluster at the NPC-exposing hernia as in *atg8Δ* (Fig. 4c, d, Extended Data Fig. 7c–e, Supplementary Video 6). **e**, NPC clustering allows high avidity between Atg8 and Nup159 (Fig. 4b, Extended Data Figs. 7e, 8a, b, Supplementary Videos 5, 8, 9, 10) causing nuclear envelope budding depicted in **f**. **f**, Autophagosomes harbouring NPC-containing nuclear vesicles (Fig. 4e, Supplementary Video 7) are transported through the cytosol for degradation in the vacuole (**g** as in Fig. 4f). Colour-code as in Fig. 4e except that nuclear envelope membranes are depicted pink. Ribosomes are depicted as red circles; nuclear material as blue circles.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	#1 *Sc*NPC WT (EMDB-10198)	#2 *Sc*NPC nup116Δ 25°C (EMDB-10660)	#3 *Sc*NPC nup116Δ 37°C (EMDB-10661)
Data collection and processing			
Magnification	42.000x	42.000x	42.000x
Voltage (kV)	300	300	300
Electron exposure (e–/Å^2)	~140	~140	~140
Defocus range (μm)	-2 to -4.5	-2 to -4.5	-2 to -4.5
Pixel size (Å)	3.37	3.37	3.37
Symmetry imposed	C8 / C1	C8 / C1	C8 / C1
Initial particle images (no.)	~500 NPCs / ~4000 AU	~250 NPCs/ ~2000 AU	~60 NPCs/ ~480 AU
Final particle images (no.)	~2000 AU	~1300 AU	~450 AU
Map resolution (Å)	25	35	50
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	20-30	-	-

Corresponding author(s): Martin Beck, Jan Kosinski, Boris Pfander

Last updated by author(s): May 11, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Used software are described in the methods. Serial EM is an open source software regularly updated
Data analysis	Used software are described in the methods. Fiji, IMOD (regurarly updated), SoftWoRx (regularly updated) ICY 2.0.0.0, NovaCTF, Gctf 1.06, 3DCT, UCSF Chimera 1.13 are open source. Graph Pad Prism is a commercially available software available at https://www.graphpad.com/scientific-software/prism/. MATLAB is a commercial software available at https://www.mathworks.com/downloads/. Movavi Video Editor 15 is a commercially available software at https://www.movavi.com/mac-video-editor/. The input data and the scripts used for modeling are available at Zenodo (https://zenodo.org/) under the DOI: 10.5281/zenodo.3820319.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The three EM maps associated with this manuscript have been deposited in the Electron Microscopy Data Bank (EMD-10198, EMD-10660, EMD-10661). The integrative models of ScNPC are available at Zenodo (https://zenodo.org/) under the DOI: 10.5281/zenodo.3820319.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size were chosen from to the practical limitation of the methodology used it was sufficient to interpret the data in a robust way according to standard statistical tests like ANOVA or Mann-Whitney and to previous publications
Data exclusions	No data was excluded from the analyses
Replication	A minimum of three biological replicates were performed for the light microscopy and western blot experiments. Electron microscopy data comes from EM grids prepared in different days from different cell cultures (biological replicates). All attempts of replication were successful
Randomization	All light microscopy and electron microscopy data comes from randomly selected cells
Blinding	Blinding is not relevant for this study because cells are randomly distributed

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Monoclonal antibody against Dpm1 (1:10,000; clone 5C5A7) was purchased from Invitrogen (Catalog # A-6429) EGFP (1:500; clone B-2) was purchased from Santa Cruz Biotechnology
Validation	The antibodies used in this study were used and validated in previous studies of S. cerevisiae (DOI:10.1038/s41556-019-0459-2). The antibodies are against GFP (clone B-2, Santa Cruz, https://www.scbt.com/p/gfp-antibody-b-2) and against Dpm1 (clone 5C5A7, Invitrogen, https://www.thermofisher.com/antibody/product/DPM1-Antibody-clone-5C5A7-Monoclonal/A-6429).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Mating_type / Strain_Name / Genotype / Created 1) alpha,WT MATα, his3-Δ200, leu2-3,2-112, lys2-801, trp1-1(am), ura3-52 Fig. 1, 2; Extended Data Fig.1-4 2)a WT MATa his3-Δ200, leu2-3,2-112, lys2-801, trp1-1(am), ura3-52 Fig. 4c, Extended Data Fig.7b 3)alpha Split-Venus-Nup159-Atg8 nup120Δ MATα his3-Δ200, leu2-3,2-112, lys2-801, trp1-1(am), ura3-52 Nup159-VC::HIS3MX6, natNT2::ADH::VN-Atg8, Nup170::mars::kanMX4, nup120Δ::hphNT1 Lee et al, 2020 Fig. 4b, Extended Data Fig. 8b 4) alpha Split-Venus-Nup159-Atg8 MATα his3-Δ200, leu2-3,2-112, lys2-801, trp1-1(am), ura3-52 Nup159-VC::HIS3MX6, natNT2::ADH::VN-Atg8, Nup170::mars::kanMX4 Lee et al, 2020 Revisions 5) a nup116Δ MAT a nup116::HIS3 ade2 his3 leu2 trp1 ura3 kind gift by Hurt lab Fig. 2, 3; Extended Data Fig. 5, 6 6) a atg15Δ Nup192p-yeGFP Nup159p-mars MATa atg15::HIS3MX6 NUP192GFP::kanMX6 NUP159mars::natNT2 or atg15Δ Nup192p-TagGFP Nup159p-mars, his3-Δ200, leu2-3,2-112, lys2-801, trp1-1(am), ura3-52 Lee et al, 2020 Fig. 4e, f

7) alpha nup116Δ Nup188p-yeGFP nup116::natNT2 NUP188GFP::klTRP1 This study Fig. 3f
8) alpha Nup188p-yeGFP NUP188GFP::klTRP1 This study Fig. 3f
9) a nup116Δ Nup133p-yeGFP nup116::natNT2 NUP133GFP::klTRP1 This study Extended Data Fig. 7a
10) alpha Nup133p-yeGFP NUP133GFP::klTRP1 This study Extended Data Fig. 7a
11) alpha Nup120p-yeGFP NUP120GFP::klTRP1 This study Fig. 4d
12) alpha atg8Δ atg8::KAN kind gift by Lusk lab Fig. 4c, Extended Data Fig. 7b
13) alpha atg8Δ Nup120p-yeGFP atg8::kanMX6 NUP120GFP::klTRP1 This study Fig. 4d
14) alpha atg8Δ Nup192p-mars Nsg1p-yeGFP atg8::kanMX6 NUP192Mars::natNT2 NSG1GFP::HIS3MX6 This study Extended Data Fig. 7d
15) alpha Nup192p-mars Nsg1p-yeGFP NUP192Mars::natNT2 NSG1GFP::HIS3MX6 This study Extended Data Fig. 7d
16) alpha vps4Δ vps4::HYG kind gift by Lusk lab Extended Data Fig.7b
17) alpha vps4Δ Nup192p-yeGFP vps4::natNT2 NUP192GFP::klTRP1 This study Extended Data Fig. 7c
18) alpha atg8Δ Nup192p-yeGFP atg8::kanMX6 NUP192GFP::klTRP1 This study Extended Data Fig. 7c
19) alpha Nup192p-yeGFP NUP192GFP::klTRP1 This study Extended Data Fig. 7c
20) alpha Nup84p-yeGFP Nup159p-mars NUP84GFP::klTRP1 NUP159mars::hphNT1 This study Fig. 3c, Extended Data Fig. 6e
21) a nup116Δ Nup84p-yeGFP Nup159p-mars nup116::natNT2 NUP84GFP::klTRP1 NUP159mars::hphNT1 This study Fig. 3c, Extended Data Fig. 6e

Authentication

Authentication of cell lines not created in this study was done on the basis of the expected phenotype of the line (PMID: 32029894)

Mycoplasma contamination

No test was needed since only S. cerevisiae strains were used

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used

5.2 PUBLICATION II: "Nuclear pores dilate and constrict in cellulo"

Apart from *S. cerevisiae*, there are other fungal species that serve as model organisms for biochemical and structural biology research. One of these fungi is *S. pombe*, also known as fission yeast. Although the NPCs of other fungal species have been tentatively studied (see Introduction and Synopsis section 5.1)[95,46], the structural knowledge regarding the *S. pombe* NPC (SpNPC) is very limited. Fluorescence and immunoelectron microscopy studies performed on SpNPC[124] have identified ~30 Nups with the nuclear pore totaling ~46 MDa[125]. Despite the fact that some domains of specific SpNups have been crystallized, there is currently no complete model of the SpNPC or even of its scaffold.

To dissect the architecture of the SpNPC, our collaborators Matteo Allegretti and Christian E. Zimmerli from Martin Beck's group (EMBL Heidelberg, MPI Frankfurt) provided us with in-cell cryo-ET maps of NPCs from exponentially growing *S. pombe* cells. Additional in-cell maps were received based on data collected from *S. pombe* cells subjected to energy depletion (ED). Taken together, these maps could potentially allow us to capture a molecular movie of the relevant NPC conformational changes and potentially prove the high NPC plasticity in cells.

We managed to produce the first in-cell SpNPC models from exponentially growing and energy depleted cells by utilizing our integrative structural modeling software Assembline[120]. The input datasets for modeling were similar to the ScNPC modeling approach (Synopsis section 5.1), which included the aforementioned maps together with the few published X-ray structures and homology models (produced by Agnieszka Obarska-Kosinska, Beck and Kosinski groups).

In the following sections, I will describe how I allocated the densities of the scaffold SpNPC subcomplexes in order to proceed with the integrative modeling of control and energy depleted SpNPCs. The modeling results revealed striking differences compared to the ScNPC architecture in terms of Nup organization in the outer rings. Finally, the integrative SpNPC models from this work suggested that NPCs can undergo significant conformational changes upon different cellular stresses, thus exhibiting high plasticity in living cells.

5.2.1 Density allocation of the scaffold SpNPC subcomplexes

The in-cell SpNPC maps from exponentially growing cells that I received from our collaborators were resolved at ~23 Å and reconstructed from a total of 726 NPC particles from 176 tomograms. In the case of tomograms that were acquired from energy depleted *S. pombe* cells, our collaborators picked and split the particles into two subpopulations based on NPC diameter criteria (i.e. above and below 50 nm). This procedure yielded intermediate and fully constricted SpNPC conformations with the maps being resolved at ~30 Å. Initially, I superposed the three maps from the control (i.e. from exponentially growing cells), intermediate (i.e. from energy depleted cells with > 50 nm diameter) and constricted (i.e. from energy depleted cells with < 50 nm diameter) SpNPCs. The superposition revealed a difference of ~20 nm central channel diameter between the

control and constricted SpNPCs (Fig. 18) as well as a discontinuous CR. Although the dynein-arm characteristics from the *S. cerevisiae* NPC are lacking in *S. pombe*, in the CR side of all three maps there is density that resembles in shape the *S. cerevisiae* counterpart of the mRNA export platform.

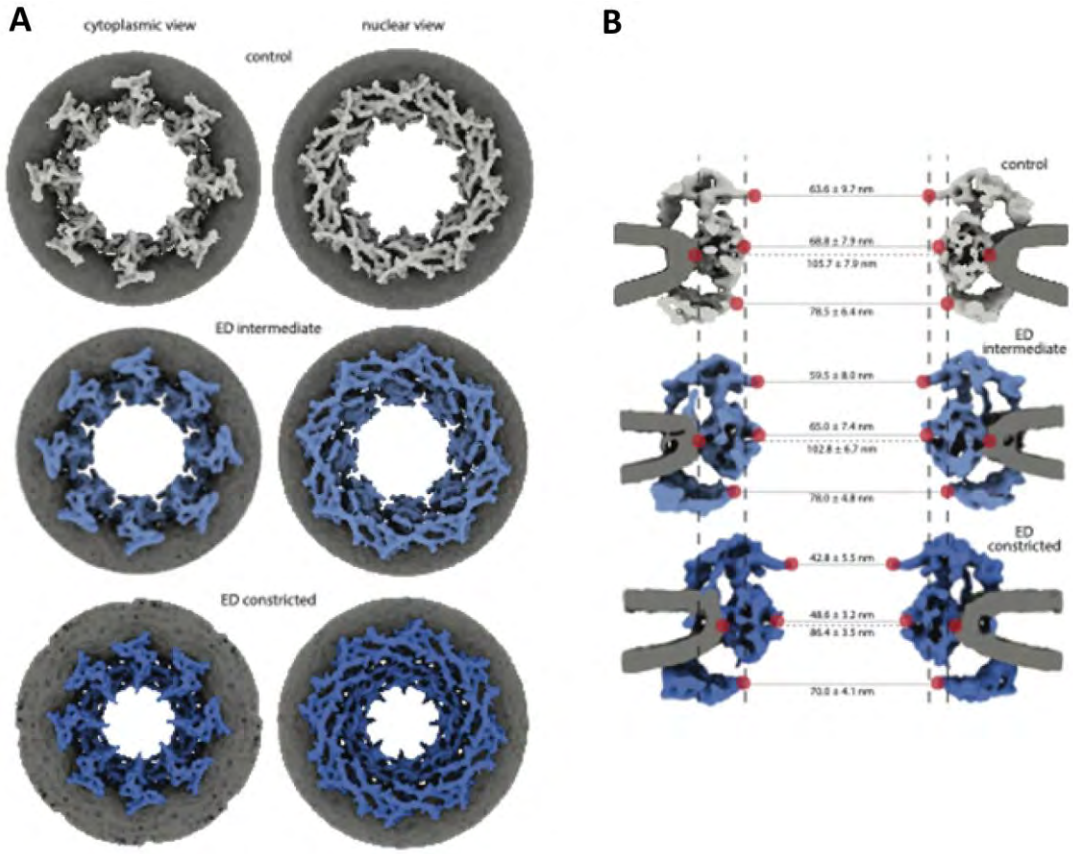


Figure 18: Comparison of cryo-ET maps of SpNPCs under control and energy depletion conditions. (A) Cytoplasmic (left) and nuclear view (right) of SpNPC map reconstructions of control (grey), intermediate- (light blue) and fully constricted ED (dark blue) NPCs (top to bottom). The maps illustrate the overall conformational change leading to a central channel diameter constriction from ~70 nm to <50 nm. (B) Same as (A) but shown as cross-section side view and overlayed with diameter measurements of individually aligned rings. The measurements were performed at the INM-ONM fusion point (dashed horizontal line) and the most centrally exposed scaffold points (solid horizontal lines). Figure adapted from[97].

To allocate the densities and produce models of SpNPCs based on all three in-cell maps, I used the same approach as in the case of ScNPC (see Synopsis section 5.1). I performed systematic fitting of integrative models of corresponding subcomplexes from other eukaryotes and smaller *S. pombe* rigid bodies (from *S. pombe* homology models, produced by Agnieszka Obarska-Kosinska, Beck and Kosinski groups). First, I used the homology models of the SpIR asymmetric unit. I built the model of the entire SpNPC IR unit by superposing the subunits onto the ScNPC counterparts from the wild type ScNPC model that I built previously. The composite SpIR unit model was fitted with Assembline[120] in the control SpNPC map with very high confidence (p-value = $9.3*10^{-13}$)

(Fig. 19), thus allocating the density of the IR unit unambiguously. This result further confirmed the high conservation of the IR unit across many eukaryotes[5]. Subsequently, I produced composite CR and NR Y-complex models in a similar fashion. First, I superposed homology models of the *S. pombe* Y-complex vertex, tail and the crystal structure including Nup120-Nup37[17] (PDB ID: 4GQ2) with previously published human integrative Y-complex models[45] (PDB ID: 5A9Q). The complete composite CR Y-complex was fitted in the control map with high confidence (p-value = 0.00013). The p-value, though was not significant when the same composite model was fitted without the Nup107-Nup131/132 rigid body (i.e. Y-complex stem) (Fig. 19). The fact that the best scoring fit of the full composite CR Y-complex model explained very well the observed CR density hinted towards a disrupted head-to-tail arrangement and incomplete CR in general. This observation is quite a striking result since such an architecture was never shown in NPCs from other eukaryotes. Interestingly enough, the density in the NR side of the control map was more similar to the NR side of the human NPC than *S. cerevisiae*. Therefore, after failing to fit with statistical significance the composite NR Y-complex model to the control map, I produced a dimerized composite NR Y-complex model based on the dimerized conformation of the human equivalent models[45] (PDB ID: 5A9Q). Indeed, when the NR Y-complex was systematically fitted as a dimer, it yielded statistically significant results (p-value = $1.1*10^{-7}$) (Fig. 19). This fact indicated the existence of two NR Y-complexes per SpNPC spoke (totaling 16 copies in the NR side).

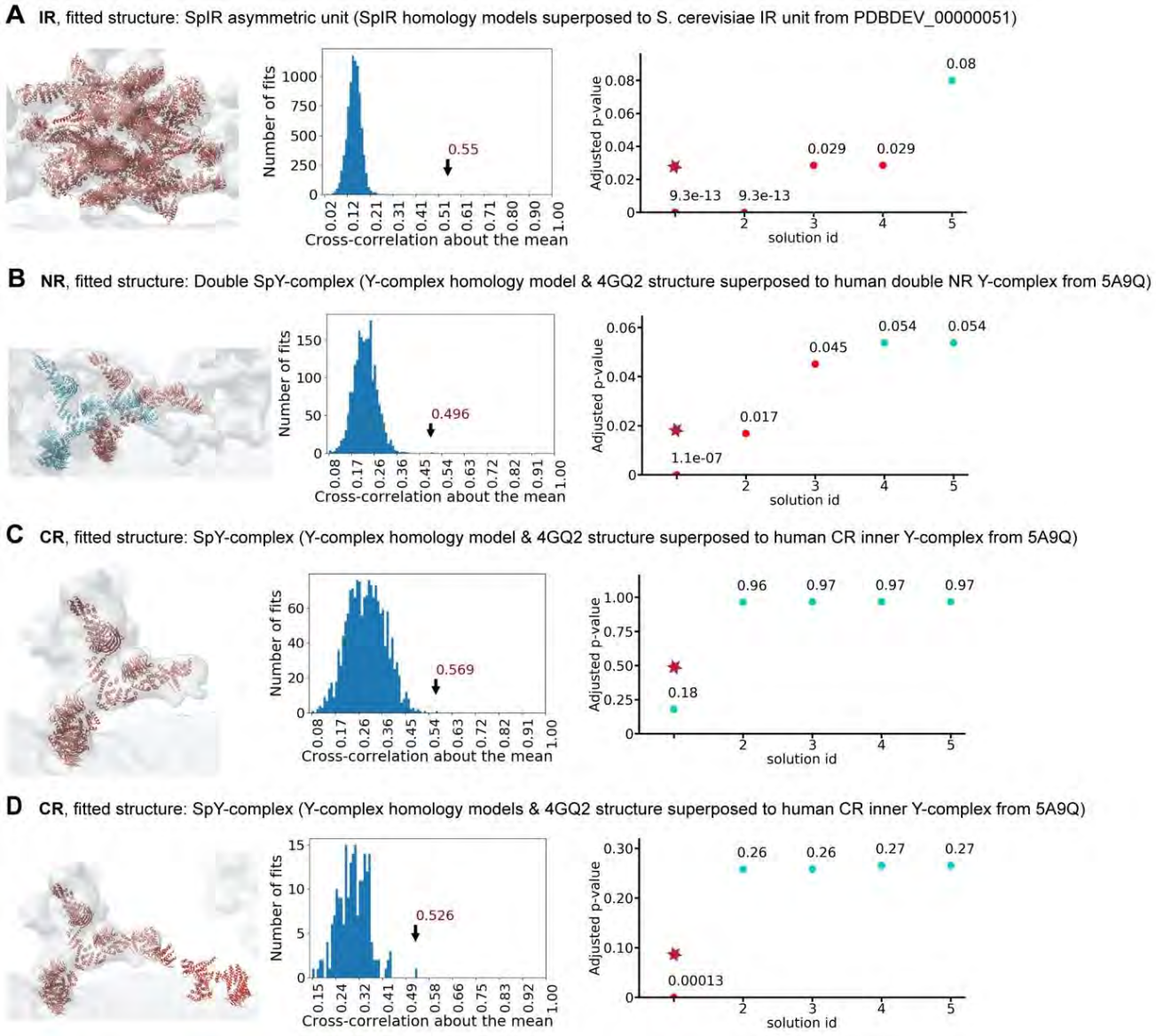


Figure 19: Systematic fitting of composite SpNPC subcomplex models to the control SpNPC maps. Each panel depicts the top fit (left), the histogram of the cross-correlation about the mean (cam) scores (middle), and a plot of the top five p-values (right). In the p-value plots, the statistically significant fits are colored red (p-value < 0.05). The top fits are indicated in the histograms with an arrow and the score value and in the p-value plots the top fits are indicated with a red star. The number of sampled fits used to calculate p-values after clustering of similar solutions was 8371 (A), 2030 (B), 1605 (C) and 210 (D). (A) Fitting of the IR. All IR asymmetric unit homology models were superposed to the integrative model of the single spoke of the IR unit of the wild type *S. cerevisiae* NPC[95] (PDB-DEV: 00000051) and the resulting composite model was used for the fitting. The two best scoring fits represent the orientations of the C2 symmetric rigid body rotated by 180°. (B-D) Fitting of the cytoplasmic NPC subunits and of the NR. The crystal structure of Nup120-Nup37 (PDB: 4GQ2) and the SpY-complex homology models (covering Nup120-Nup189c-Sec13-Seh1-Nup85-Nup107, Nup107-Nup131/Nup132 and Nup131/Nup132 β-propeller) were superposed to the human Y-complex models (PDB: 5A9Q) and systematically fitted. The two composite models of NR Y-complexes shown in (B) were fitted as a single rigid body and colored differently for visualization purposes only (outer Y-complex blue, inner Y-complex red). All components were fitted to the cryo-EM maps of the individual cytoplasmic side, IR and NR segments. Figure adapted from[97].

From both best scoring fits for the composite NR and CR SpY-complexes, a yet-to-be-assigned density was apparent in the long arm of the complex, on top of the Nup120 α-helical domain. The localization and size of this density would match the nucleoporin Ely5 (ELYS in human). In fungi this Nup consists of a ca. 300 amino acid long α-helical domain and is absent in *S. cerevisiae*. To verify this hypothesis, our collaborators obtained in-cell cryo-ET maps of SpNPCs from Nup37Δ and Nup37Δ-Ely5Δ cells. The latter maps were poorly resolved compared to the control map, though the resolution was sufficient for calculating the difference in density between the control SpNPC map and the maps from the knock-out cells. Indeed, the calculated difference in densities revealed that the SpNPC maps from knock-out cells lacked densities proximally to the Nup120 arm (Fig. 20) in both NR and CR. In the knock-out maps, the missing density seemed to belong to Nup37 according to the systematic fitting results, while the rest seemed enough to accommodate Ely5 in both CR and NR. This was a striking difference compared to the exclusive localization of Ely5 homologs in the NR side in higher eukaryotes.

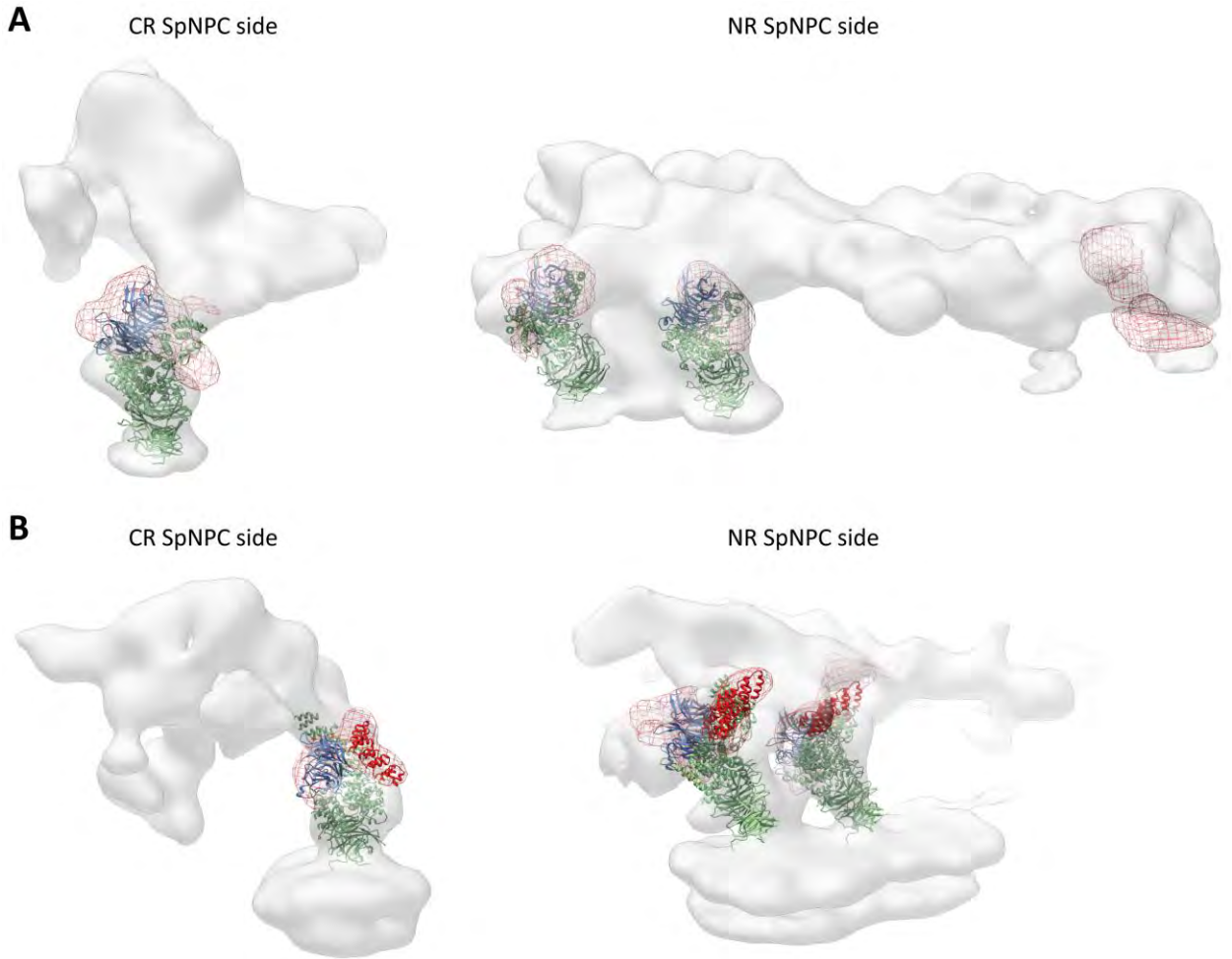


Figure 20: Density differences between the control SpNPC map and maps from knock-out *S. pombe* strains. Representation of locally fitted X-Ray structure of Nup120-Nup37 (with PDB id: 4GQ2)[17] inside the CR (left) and NR (right) side of the control SpNPC map (grey density). The calculated difference in density between the control and SpNPC map from Nup37Δ cells (A) and Nup37Δ-Ely5Δ cells (B) is depicted in red mesh. In both (A) and (B), the previously determined by systematic fitting density of Nup37 appears to be missing in CR and NR sides of the SpNPC scaffold with additional density missing respectively in both sides in (B).

Having allocated the densities of major scaffold SpNPC subcomplexes in the control map I systematically fitted the composite SpNPC scaffold subcomplexes (and the smaller rigid bodies comprising them) into the ED maps. Even from the initial visual inspection after the superposition of the three SpNPC (i.e. from control and ED cells) maps, a significant similarity of density features quickly became apparent. Therefore, it was no surprise that systematic fitting of these subunits to the intermediate and constricted SpNPC maps yielded very similar results compared to the fitting of the same subunits in the control map. The composite IR asymmetric unit was fitted in both maps from the ED cells with significant p-values while the statistically significant fits for the composite Y-complexes hinted towards the existence of two concentric rings and an incomplete cytoplasmic side for NR and CR respectively (Fig. 21).

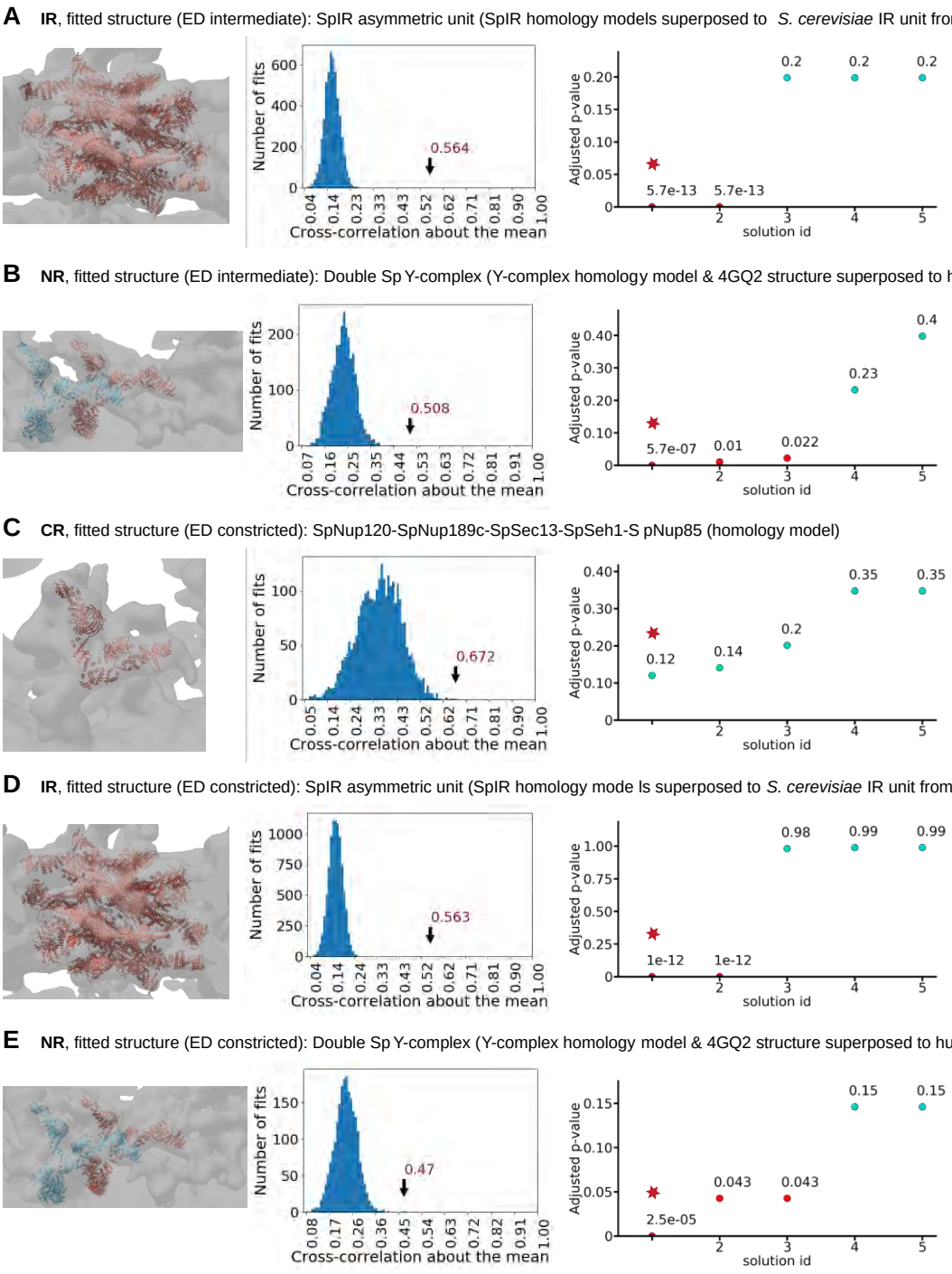


Figure 21: Representative results from systematic fitting of composite SpNPC subcomplexes to intermediate and constricted SpNPC maps (from ED *S. pombe* cells). Each panel shows the visualization of the top fit (left), the histogram of the cross-correlation about the mean (cam) scores (middle), and a plot of the top five p-values (right). In the p-value plots, the statistically significant fits are colored red (p-value < 0.05). The top fits are indicated in the histograms with an arrow and the score value while in the p-value plots the top fits are indicated with a red star. The number of sampled fits used to calculate p-values after clustering of similar solutions was 5134 (A), 2684 (B), 2778 (C), 9189 (D) and 2150 (E). (A) Fitting of the composite IR asymmetric unit to the intermediately constricted SpNPC map. (B) Fitting of NR components to the intermediately constricted SpNPC map. The crystal structure covering Nup120-Nup37[17] (PDB id: 4GQ2) and the Y-complex homology model (covering Nup120-Nup189c-Sec13-Seh1-

Nup85-Nup107) were superposed to human Y-complexes[45] (PDB: 5A9Q) and systematically fitted as a single rigid body. (C) Fitting of the cytoplasmic side NPC subunits to the most constricted SpNPC map. The Y-complex homology model (covering Nup120-Nup189c-Sec13-Seh1-Nup85) was superposed to the human Y-complex models[45] (PDB: 5A9Q) and systematically fitted. (D) Fitting of the IR asymmetric unit to the most constricted SpNPC map. As in (A), all IR homology models were superposed to the integrative model of the single spoke of the IR unit from control *S. cerevisiae* NPC[95] (PDB-DEV: 00000051) and the resulting composite model was used for systematic fitting. (E) Fitting of NR components to the most constricted SpNPC map. As in (B), the crystal structure covering Nup120-Nup37[17] (PDB: 4GQ2) and the Y-complex homology model (covering Nup120-Nup189c-Sec13-Seh1-Nup85-Nup107) were superposed to human Y-complexes[45] (PDB: 5A9Q) and systematically fitted. Figure adapted from[97].

5.2.2 Structural elucidation of the SpNPC architecture

With the density of the scaffold SpNPC subcomplexes allocated in all three SpNPC maps (i.e. maps from control and ED cells) and the fit libraries of *S. pombe* subunits already produced by systematic fitting, Jan Kosinski (EMBL Hamburg) continued with the production of integrative models of SpNPCs. The integrative modeling approach with Assembline that was followed for the SpNPC models was very similar to the one used for the production of the ScNPC models.

The construction of the control integrative SpNPC model started with the local refinement of the best scoring composite IR asymmetric unit model and de novo modeling runs with Assembline (see Synopsis section 5.3) of the CR and NR Y-complexes. For this step, the fit libraries of smaller rigid bodies (the ones used to produce the composite Y-complex models described earlier) were utilized as input together with the control SpNPC map. Ensembles of the produced subcomplex models were then refined altogether in order to create the complete control SpNPC scaffold model (as the single model for the control SpNPC the best scoring refined model was selected).

An ensemble of best scoring control SpNPC scaffold models (refinement run mentioned above) were fitted as rigid bodies (subcomplexes per NPC ring) and further refined with respect to the EM densities from the ED cells.

The three integrative models (Fig. 22) of SpNPCs confirmed that in both the control and ED *S. pombe* cells, the SpNPCs exhibit an unprecedented architecture of incomplete CR and dimerized Y-complexes in the NR. Additionally, judging by the conformational differences between the three models, we hypothesized that in order for SpNPCs to adopt the constricted architecture, both CR and NR subcomplexes need to change their curvature. In order for the central channel constriction to be accommodated, larger lateral displacements of the 8 spokes (at the IR level) that move as independent entities need to occur. All these structural movements suggested by our integrative models hint towards the high in-cell plasticity of SpNPCs upon cellular cues.

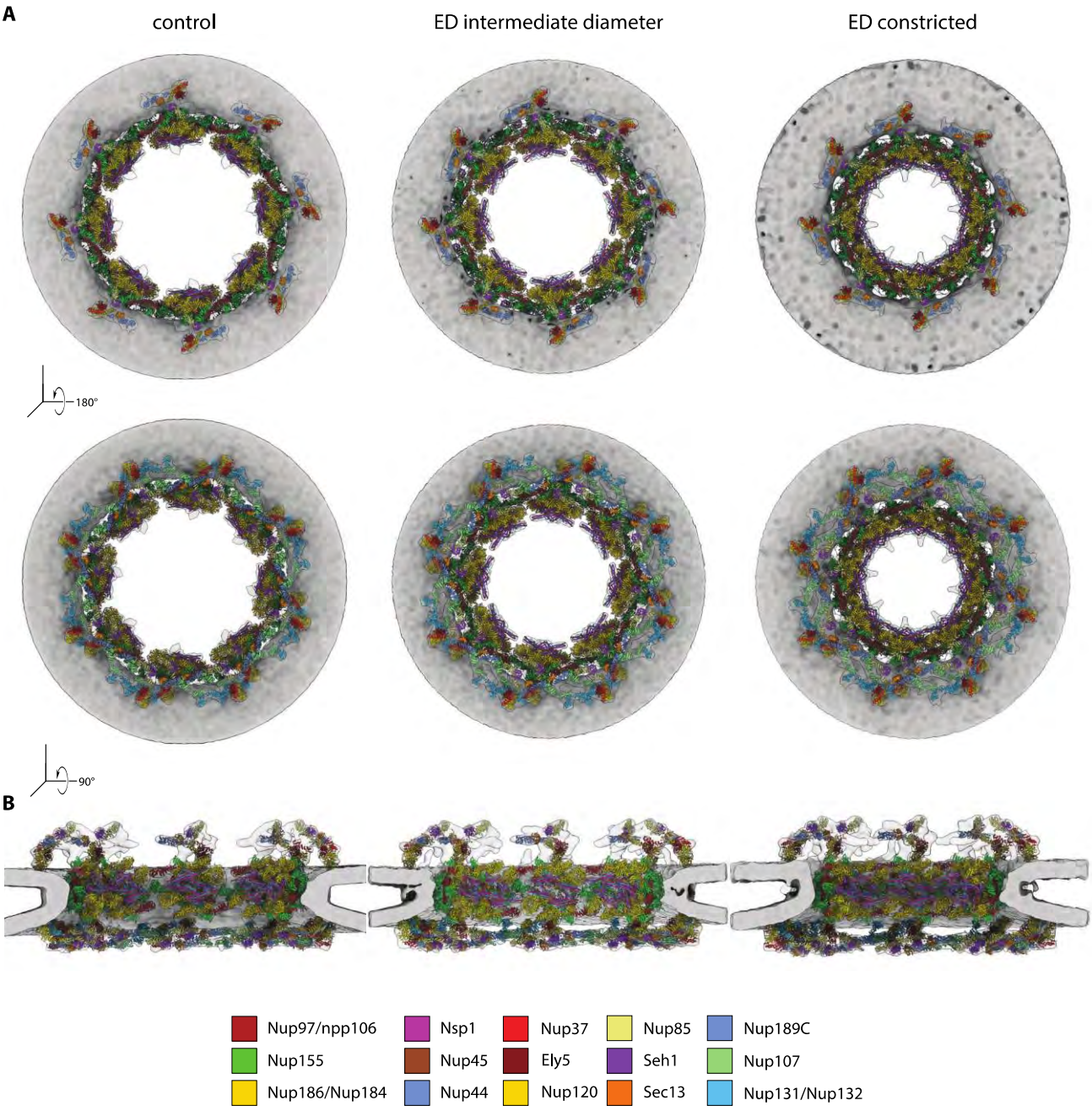


Figure 22: Overview of the SpNPC integrative models built based on datasets from control and energy depleted *S. pombe* cells. (A) Cytoplasmic view (top) and nuclear view (bottom) of cryo-EM maps (grey density) superimposed with the respective integrative models from control, intermediate and fully constricted SpNPCs. The models illustrate the overall conformational changes leading to a central channel diameter constriction from ~70 nm to ~50 nm. (B) Same as (A) but shown as cross-section side view from the IR level. In both (A) and (B), the color-coding scheme of Nups comprising the models is provided. Figure adapted from[97].

RESEARCH ARTICLE SUMMARY

CELL BIOLOGY

Nuclear pores dilate and constrict in cellulo

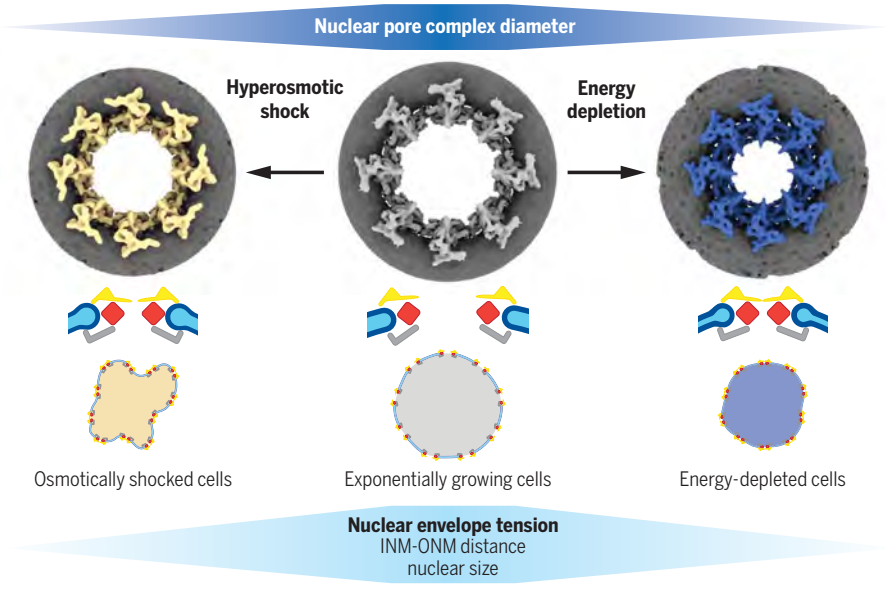
Christian E. Zimmerli†, Matteo Allegretti†, Vasileios Rantos, Sara K. Goetz, Agnieszka Obarska-Kosinska, Ievgeniia Zagoriy, Aliaksandr Halavatyi, Gerhard Hummer, Julia Mahamid, Jan Kosinski*, Martin Beck*

INTRODUCTION: The nucleus harbors and protects the genetic information in eukaryotes. It is surrounded by the nuclear envelope, which separates the nucleoplasm from the cytosol. The double-membrane system of the nuclear envelope is physically connected to the cytoskeleton and chromatin on either side. It senses and signals mechanical stimuli. Nuclear pore complexes (NPCs) mediate nucleocytoplasmic exchange. They fuse the inner and outer nuclear membranes of the nuclear envelope to form an aqueous central channel. Their intricate cylindrical architecture consists of ~30 nucleoporins. The nucleoporin scaffold embraces the membrane fusion topology in a highly dynamic fashion. Variations in NPC diameter have been reported, but the physiological circumstances and the molecular details remain unknown.

RATIONALE: To investigate the dynamics of NPC conformation in vivo, we structurally analyzed NPCs in *Schizosaccharomyces pombe* cells. We used cryo–electron tomography with subsequent subtomogram averaging to quantify the diameter of NPCs in cells exposed to different stress stimuli. We found a dilated NPC scaffold in exponentially growing cells that constricted under conditions of energy depletion and hyperosmotic shock. Structural analysis in combination with integrative structural modeling revealed that the NPC scaffold underwent large-scale movements and bending during constriction. Thereby, the volume of the central channel was lessened to about one-half. Reduced nucleocytoplasmic diffusion was apparent. Scaffold nucleoporins do not have any known motor activity that would allow them to exert forces. Thus, it appears likely that the observed dynamics of the diameter of the NPC were the result of forces that are laterally applied within the nuclear envelope. Under conditions of both energy depletion and hyperosmotic shock, cellular and nuclear volumes were reduced such that nuclear shrinkage led to an excess of nuclear membranes. We therefore hypothesized that the NPC scaffold is mechanosensitive. We surmised that a reduction of nuclear envelope membrane tension resulted in a consequent NPC constriction into a conformational ground state.

RESULTS: We outline several predictions of a conceptual model in which nuclear envelope membrane tension regulates NPC diameter. In such a scenario, nuclear volume and NPC diameter should be dependent on each other. The rigidity of the NPC scaffold should counteract its dilation. On the basis of membrane elastic theory, we predicted that osmotic pressure in the nuclear envelope lumen and membrane tension in the nuclear envelope act equivalently. As a consequence, the distance between the inner and outer nuclear membranes should grow linearly with both. We set out to test those predictions experimentally. We found that nuclear shrinkage correlated with NPC constriction. Recovery experiments established that this behavior was reversible. Constricted NPCs dilated when shifted back to normal medium. Nuclear volumes recovered, and cells remained viable. In exponentially growing cells, genetic perturbation of the NPC scaffold resulted in further dilation. We segmented the inner and outer nuclear membranes in cryo–electron tomograms and quantified their distance. As expected, the diameter of the NPC directly correlated with the distance between the inner and outer nuclear membranes.

CONCLUSION: Our data strongly suggest that the NPC scaffold is mechanosensitive and that membrane tension regulates its diameter. The data link the conformation of the NPC to the mechanical status of the nuclear envelope. This finding has implications for various cellular processes during which the nucleus may be exposed to mechanical forces. This includes cell differentiation and migration, nuclear envelope maintenance in mechanically active tissues, the import of very large cargos such as viral capsids, metastasis, and changing osmotic conditions. Under such conditions, nuclear envelope membrane tension will change globally or locally. The resulting changes in the diameter of the NPC may have functional consequences that remain to be investigated further in the future. ■



NPCs constrict under conditions of cellular energy depletion and hyperosmotic shock. Cryo–electron microscopy maps of different NPC conformations are shown isosurface rendered as seen from the cytosol in exponentially growing (light gray), hyperosmotically shocked (yellow), and energy-depleted cells (blue) (top). Nuclear membranes are shown in dark gray. Cartoons indicate the respective conformational changes. Upon exposure to hyperosmotic shock and energy depletion, NPCs constrict to smaller central channel diameters. The nuclear volume is decreased, and the distance between the inner and outer nuclear membranes is reduced (bottom), which is indicative of a loss of nuclear envelope membrane tension. INM, inner nuclear membrane; ONM, outer nuclear membrane.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: jan.kosinski@embl.de (J.K.); martin.beck@biophys.mpg.de (M.B.)
†These authors contributed equally to this work.
Cite this article as C. E. Zimmerli *et al.*, *Science* **374**, eabd9776 (2021). DOI: 10.1126/science.abd9776

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abd9776

RESEARCH ARTICLE

CELL BIOLOGY

Nuclear pores dilate and constrict in cellulo

Christian E. Zimmerli[1,2,3]†, Matteo Allegretti[1,3]†, Vasileios Rantos[4,5], Sara K. Goetz[1,2], Agnieszka Obarska-Kosinska[3,5], Ievgeniia Zagoriy[1], Aliaksandr Halavatyi[6], Gerhard Hummer[7,8], Julia Mahamid[1], Jan Kosinski[1,4,5]*, Martin Beck[1,3]*

In eukaryotic cells, nuclear pore complexes (NPCs) fuse the inner and outer nuclear membranes and mediate nucleocytoplasmic exchange. They are made of 30 different nucleoporins and form a cylindrical architecture around an aqueous central channel. This architecture is highly dynamic in space and time. Variations in NPC diameter have been reported, but the physiological circumstances and the molecular details remain unknown. Here, we combined cryo–electron tomography with integrative structural modeling to capture a molecular movie of the respective large-scale conformational changes in cellulo. Although NPCs of exponentially growing cells adopted a dilated conformation, they reversibly constricted upon cellular energy depletion or conditions of hypertonic osmotic stress. Our data point to a model where the nuclear envelope membrane tension is linked to the conformation of the NPC.

Nuclear pore complexes (NPCs) bridge the nuclear envelope (NE) and facilitate nucleocytoplasmic transport. Across the eukaryotic kingdom, ~30 different genes encode for NPC components, termed nucleoporins (Nups). Although specialized Nups have been identified in many species, extensive biochemical and structural studies have led to the consensus that the core scaffold inventory is conserved. It consists of several Nup subcomplexes that come together in multiple copies to form an assembly of eight asymmetric units, called spokes, that are arranged in a rotationally symmetric fashion (*1*). The Y-complex (also called the Nup107 complex) is the major component of the outer rings [the nuclear ring (NR) and the cytoplasmic ring (CR)], which are placed distally into the nuclear and cytoplasmic compartments. The inner ring complex scaffolds the inner ring [(IR) also called the spoke ring] that resides at the fusion plane of the nuclear membranes. It consists of the scaffold Nups 35, 155, 188, and 192 as well as the Nsp1 complex. The IR forms a central channel lined with phenylalanine-glycine (FG) repeats containing Nups that interact with cargo complexes. The Nup159 complex (also called the P-complex) asymmetrically associates with the Y-complex of the CR and mediates mRNA export. Despite these common features of quaternary structure, in situ structural biology studies have revealed that the higher-order assembly is variable across the eukaryotic kingdom (*1*, *2*).

NPC architecture is conformationally highly dynamic, and variations in NPC diameter have been observed in various species and using different methods (*3*–*7*). Dilated states have been observed in intact human cells (*3*, *8*, *9*), contrasting with the constricted state in semi-purified NPCs (*10*–*12*). It has been shown that dilation is part of the NPC assembly process (*13*, *14*). However, it remains controversial whether NPC dilation and constriction play a role during active nuclear transport (*15*) and whether the dilation is required to open up peripheral channels for the import of inner nuclear membrane (INM) proteins (*16*–*18*). It has been argued that the constricted state may be a result of purification (*4*, *8*). It is difficult to conceive that such large-scale conformational changes can occur on similar time scales as individual transport events (*19*, *20*), which would be the essence of a physical gate. Nevertheless, several cues that potentially could affect NPC diameter have been suggested, such as exposure to mechanical stress, mutated forms of importin β, varying Ca^{2+} concentrations, or exposure to hexanediol (*7*, *21*–*26*). The biological relevance of these cues remains elusive because the analysis of NPC architecture under physiological conditions is experimentally very challenging. These previous studies did not explore NPC dilation and constriction and its functional cause and consequences within intact cellular environments, nor did they structurally analyze the conformational changes of nuclear pores in high molecular detail.

Here, we demonstrate that *Schizosaccharomyces pombe* NPCs (SpNPCs) constrict in living cells under conditions of energy depletion (ED) or hyperosmotic shock (OS), which is concomitant with a reduction of NE membrane tension. Using in cellulo cryo–electron microscopy (cryo-EM) and integrative structural modeling, we captured a molecular movie of NPC constriction. Our dynamic structural model suggests large-scale conformational changes that occur by movements of the spokes with respect to each other but largely preserve the arrangement of individual subcomplexes. Previous structural models obtained from isolated NEs (*10*–*12*, *27*–*29*) thereby represent the most constricted NPC state.

In cellulo cryo-EM map of the SpNPC

To study NPC architecture and function in cellulo at the best possible resolution and structural preservation, we explored various genetically tractable model organisms for their compatibility with cryo–focused ion beam (cryo-FIB) specimen thinning, cryo–electron tomography (cryo-ET), and subtomogram averaging (STA). *Saccharomyces cerevisiae* cells were compatible with high-throughput generation of cryo-lamellae and acquisition of tomograms. STA of their NPCs resulted in moderately resolved structures (*4*). By contrast, a larger set of cryo-tomograms from *Chaetomium thermophilum* cells did not yield any meaningful averages, possibly because their NPCs displayed a very large structural variability. We therefore chose to work with *S. pombe* cells that are small enough for thorough vitrification; offer a superior geometry for FIB milling compared with *C. thermophilum*, with the advantage of covering multiple cells; and, compared with *S. cerevisiae*, have a higher number of NEs and NPCs per individual cryo-lamella and tomogram, leading to increased data throughput (fig. S1).

To obtain a high-quality cryo-EM map of SpNPCs, we prepared cryo-FIB–milled lamellae of exponentially growing *S. pombe* cells and acquired 178 tomograms from which we extracted 726 NPCs. Subsequent STA resulted in an in cellulo NPC average of very high quality in both visible features and resolution (Fig. 1 and figs. S2 and S3). Systematic fitting of the *S. pombe* IR asymmetric unit model, built based on the *S. cerevisiae* NPC, resulted in precisely one highly significant fit, as expected (figs. S4A and S5A and Materials and methods). The subsequent refinement with integrative modeling led to a structural model that explains most of the observed EM density in the IR (Fig. 1B, fig. S5B, and movie S1). The IR architecture appears reminiscent to NPC structures of other eukaryotes (fig. S6), further corroborating its evolutionary conservation (*2*, *30*) (see table S1 for nomenclature of Nups across different species). Systematic fitting revealed that the NR of the SpNPC is composed of two concentric Y-complex rings (Fig. 1A,

[1]Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany. [2]Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences, 69120 Heidelberg, Germany. [3]Department of Molecular Sociology, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany. [4]Centre for Structural Systems Biology (CSSB), 22607 Hamburg, Germany. [5]EMBL Hamburg, 22607 Hamburg, Germany. [6]Advanced Light Microscopy Facility, EMBL, 69117 Heidelberg, Germany. [7]Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany. [8]Institute of Biophysics, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany.
*Corresponding author. Email: jan.kosinski@embl.de (J.K.); martin.beck@biophys.mpg.de (M.B.)
†These authors contributed equally to this work.

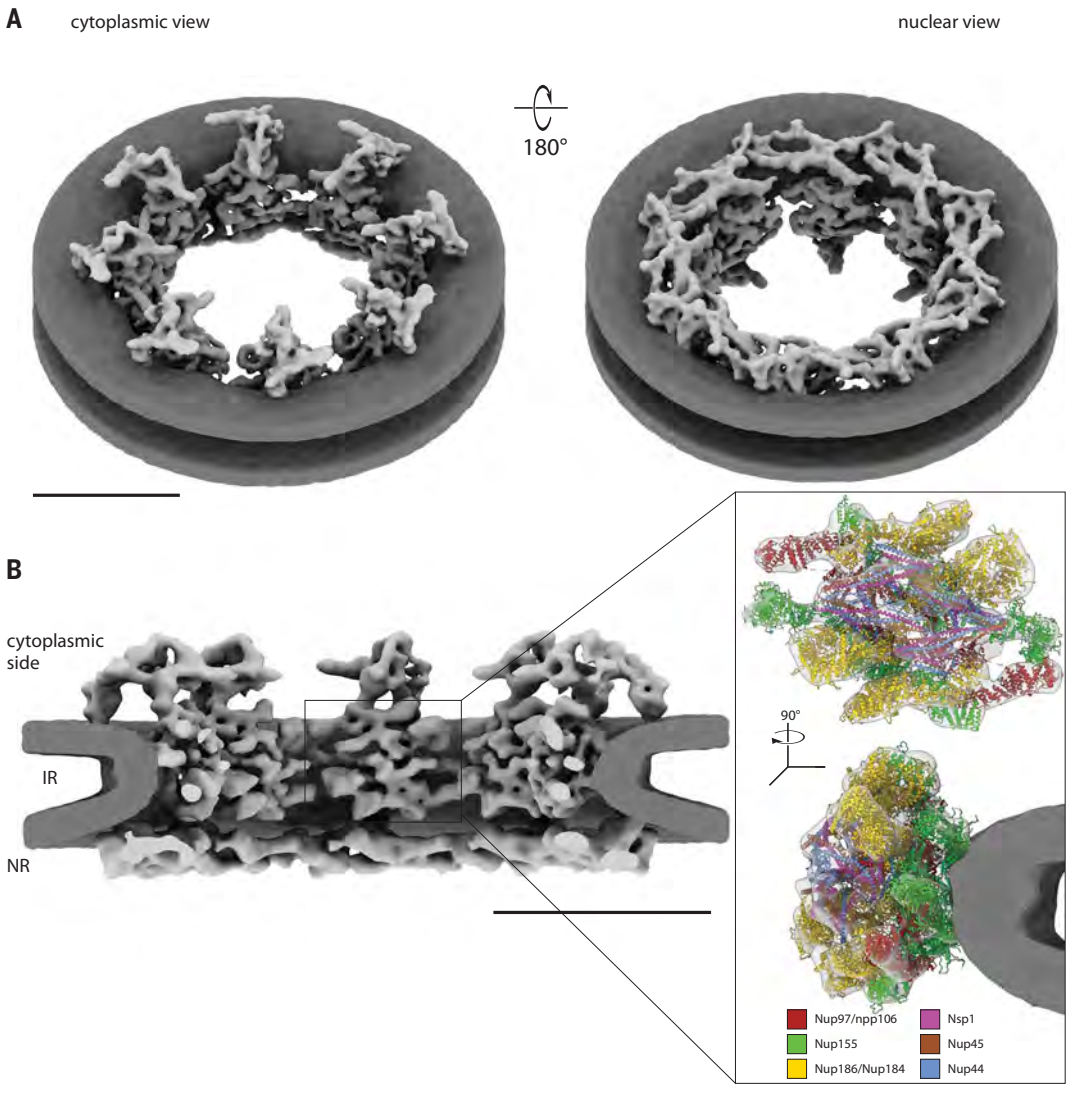


Fig. 1. In cellulo cryo-EM map of the SpNPC. (**A**) Isosurface rendered views of the SpNPC eightfold assembly as seen from the cytoplasm (left) and the nucleoplasm (right) (with membranes in dark gray and protein in light gray). Note, the eightfold symmetric assembly represents an artificial NPC conformation; within living cells, NPCs adopt more-flexible conformations deviating from a perfect circle. Although the cytoplasmic view (left) reveals eight disconnected protein entities instead of a CR, the nuclear view (right) shows two concentric nuclear Y-complex rings. Scale bar, 50 nm. (**B**) Same as (A) but shown as a cutaway view. Although the asymmetric curvature of the nuclear membranes and the arrangement of the cytoplasmic side is unprecedented in other species, the IR architecture is highly conserved, as highlighted in the inset (see also fig. S6). Scale bar, 50 nm. The inset shows one individual asymmetric IR unit map as obtained during STA (see Materials and methods); the density threshold is chosen at a level comparable to the entire eightfold assembly. In the side view (inset, below), the membrane is shown for reference. Fitting of the IR Nup homology models explains most of the observed electron optical density.

fig. S4B, and movie S1) as in vertebrates and algae but as opposed to the single ring observed in *S. cerevisiae* (fig. S3A) (*4*, *11*, *31*). Integrative modeling of the entire Y-complex in the NR revealed a Y-complex architecture with the typical head-to-tail oligomerization (Fig. 2A). It recapitulates the domain orientations and interactions known from other species. However, *S. pombe* Y-complex Nups do localize to the NR, in contrast to previous proposals (*32*).

Closer inspection of the cytoplasmic side of the cryo-EM map revealed an unanticipated architectural outline—it did not form an apparent ring. Instead, we observed eight spatially separated entities (Fig. 1A), which serve as anchor points for the mRNA export platform (*4*, *33*), the Nup159 complex (Fig. 2B). Although the dynein-arm characteristic for the *S. cerevisiae* NPC (*34*) is lacking, the Nup159 complex resembles its *S. cerevisiae* counterpart in shape (fig. S7A). Systematic fitting and subsequent refinement with integrative modeling revealed that only a single Y-complex vertex fits into the density observed at the cytoplasmic side (Fig. 2B, fig. S4C, and movie S1). The density potentially accounting for Nup107 and SpNup131-SpNup132 was missing (figs. S4D and S7B) and could not be recovered by local refinement (fig. S7C). To independently confirm the identity of the observed vertex-like density, we analyzed *nup37*Δ and *nup37*Δ-*ely5*Δ strains. The binding of both Nup37 and Ely5 to Nup120 has been shown previously in vitro (*35*, *36*), and, as expected, density was missing in the respective positions of all Y-complexes (Fig. 2, C and D, and figs. S8 and S9). A density that could accommodate the Ely5 homology model was missing also in the cytoplasmic Y-complex, suggesting that Ely5 is present in *S. pombe* at both sides, unlike its human counterpart Elys (*11*, *37*) (Fig. 2D and movie S1). Otherwise, the NPC architecture remained mostly unchanged, despite some increased flexibility in the Nup120 arm of the outer nuclear Y-complex pointing to a decreased NR stability under knockout conditions. These results independently identify the cytoplasmic structure as bona fide Y-complex vertex but lacking the EM density beyond the edge of SpNup189C. This is consistent with previous work suggesting a split of the SpNup189C-Nup107 interface (*32*, *38*). Alternatively, this observation may be explained by flexibility, although hinges within the Y-complex were identified at different locations (*10*, *39*).

ED is concomitant with a constriction of the central channel

Previous cryo-EM structures of NPCs obtained from isolated NEs (*10*–*12*, *27*–*29*) or by detergent extraction (*40*) had a smaller diameter compared with those obtained from intact cells (*3*, *4*, *31*, *41*). We therefore hypothesized that NPC diameter may depend on the biochemical energy level, which is diminished

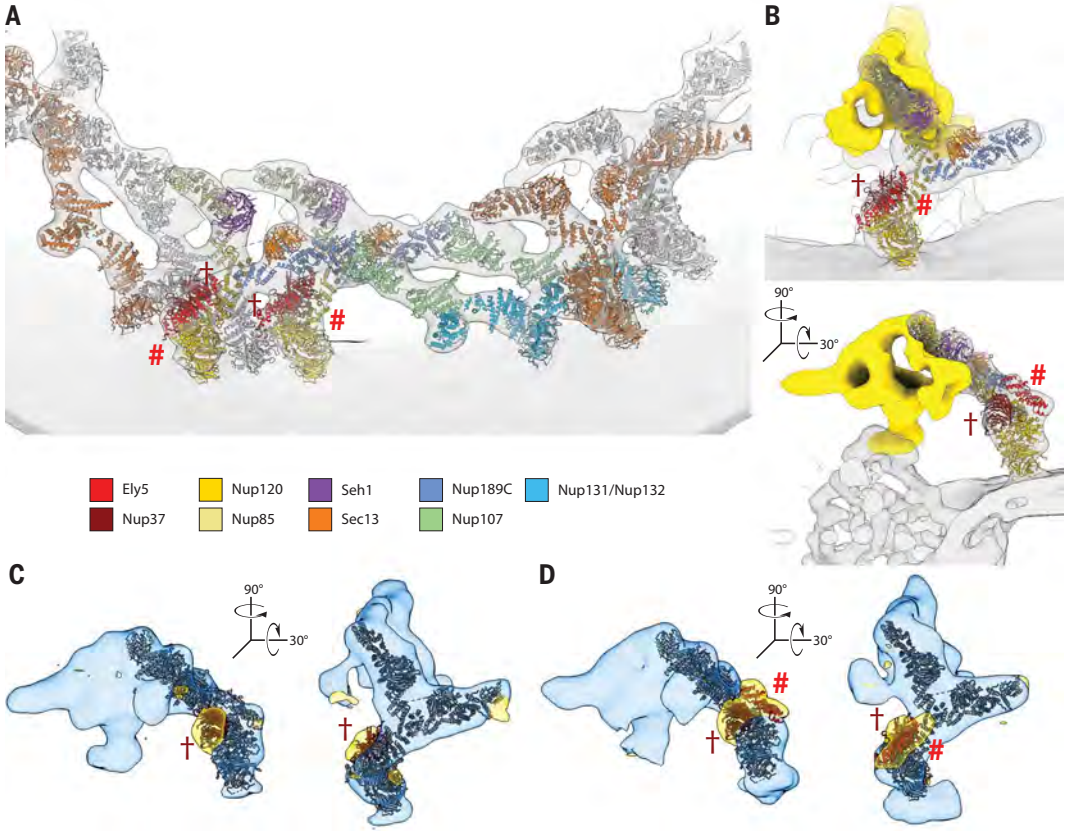


Fig. 2. Architecture of SpNPC outer rings. (**A**) Systematic fitting and integrative modeling of all *S. pombe* Y-complex Nups reveal a head-to-tail arrangement with two concentric Y-complex rings on the nuclear side of the SpNPC, as in the human NPC (*11*). The cryo-EM map of a NR segment is shown rendered as an isosurface in transparent light gray. The adjacent inner Y-complexes are shown in gray, and the outer Y-complexes are shown in orange. The homology models of SpNup131 and SpNup132 fit to the Y-complex tail region equally well, rendering these two proteins indistinguishable by our approach. (**B**) Integrative model of the cytoplasmic protein entities. The fit of the Y-complex vertex explains most of the observed density. The mRNA export platform as identified in (*4*, *33*) is segmented in yellow. (**C**) The *nup37*Δ cryo-EM map is shown in blue (threshold: 0.1) and overlaid with the difference map (yellow; threshold: 0.175) of the WT and *nup37*Δ maps, both filtered to 27 Å. The missing density in the long arm of the Y-vertex coincides with the position of Nup37 (indicated by the dark red dagger symbol) in the Y-complex vertex (dark gray) (see also fig. S8). (**D**) The *nup37*Δ-*ely5*Δ double knockout map (blue; threshold: 0.08) overlaid with the corresponding difference map (yellow; threshold: 0.188) comparing the WT and double knockout map at 34 Å. Differences are apparent at the location of both Nup37 (dark red dagger symbol) and Ely5 (light red number symbol) with respect to the fitted Y-complex model (dark gray) (see also fig. S9).

in preparations of isolated NEs or NPCs but may also be reduced within intact cells—e.g., during stress conditions. We first set out to analyze NPCs in *S. pombe* cells under conditions of ED, namely after 1 h of exposure to nonhydrolyzable 2-deoxy-glucose in combination with the respiratory chain inhibitor antimycin A, as previously established (*42*) (see Materials and methods). Recovery experiments demonstrated the viability of cells exposed to ED, and the subcellular architecture of the respective cells remained intact, as apparent by cryo-ET (fig. S10). Cells of various organisms including *S. pombe* show a rapid shut down of active nuclear transport and mRNA export when depleted of adenosine 5′-triphosphate (ATP) because of a concomitant reduction of guanosine 5′-triphosphate (GTP) levels, which leads to the loss of the nucleocytoplasmic RanGTP-RanGDP (GDP, guanosine diphosphate) gradient (*43*–*46*). To confirm a loss of active nucleocytoplasmic transport under ED conditions, we used live-cell imaging of *S. pombe* cells expressing a green fluorescent protein (GFP) variant tagged with a nuclear localization signal (NLS) or nuclear export signal (NES) on its N and C terminus (NLS-GFP-NLS or NES-GFP-NES), which show a nearly exclusive nuclear or cytoplasmic localization under control conditions (fig. S11, A to D). After 30 min of ED, most of the NLS-GFP-NLS localized to the cytoplasm, whereas NES-GFP-NES equilibrated into the nucleus (fig. S11, A to D). This was reverted in the vast majority of cells when recovered in glucose control medium (fig. S11, E to G), thus underlining their viability.

We analyzed 292 NPCs structurally by STA under these conditions and found a considerable constriction of the central channel diameter. However, the averages appeared blurred, and manual inspection of the data indicated a large variation of diameters. To generate a conformationally more-homogeneous ensemble, we manually assigned the NPCs from the ED dataset into two classes with central channel diameters of <50 nm and >50 nm and refined them separately (corresponding to 533 and 1012 subunits, respectively) (Fig. 3 and fig. S12, A to C). Both conformations of the ED state showed a smaller NPC diameter compared with the conformation observed in NPCs of exponentially growing cells, from here on referred to as control conditions. The intermediate conformation was ~65-nm wide at the IR compared with ~70 nm under control conditions, whereas the most-constricted conformation showed a diameter of ~50 nm, comparable to the diameter observed in isolated NEs (*10*–*12*, *27*–*29*).

To better understand how NPCs accommodate such massive conformational changes on the molecular level, we systematically fitted individual subcomplexes (fig. S13) and built structural models of the three different diameter states on the basis of the cryo-EM maps (fig. S14) using integrative modeling (fig. S5B). Both systematic fitting and integrative modeling showed that the relevant subunits are present in the control state and the ED states. At the cytoplasmic side and the NR, changes in the curvature of the Y-complexes and inward bending of the mRNA export platform toward the center of the pore were apparent (movies S4 to S6). By contrast, the central channel constriction of the IR is more elaborate and mediated by a lateral displacement of the eight spokes that move as independent entities to constrict or dilate the IR (fig. S14 and movies S4 to S6). In the control state, ~3- to 4-nm-wide gaps are formed in-between the neighboring

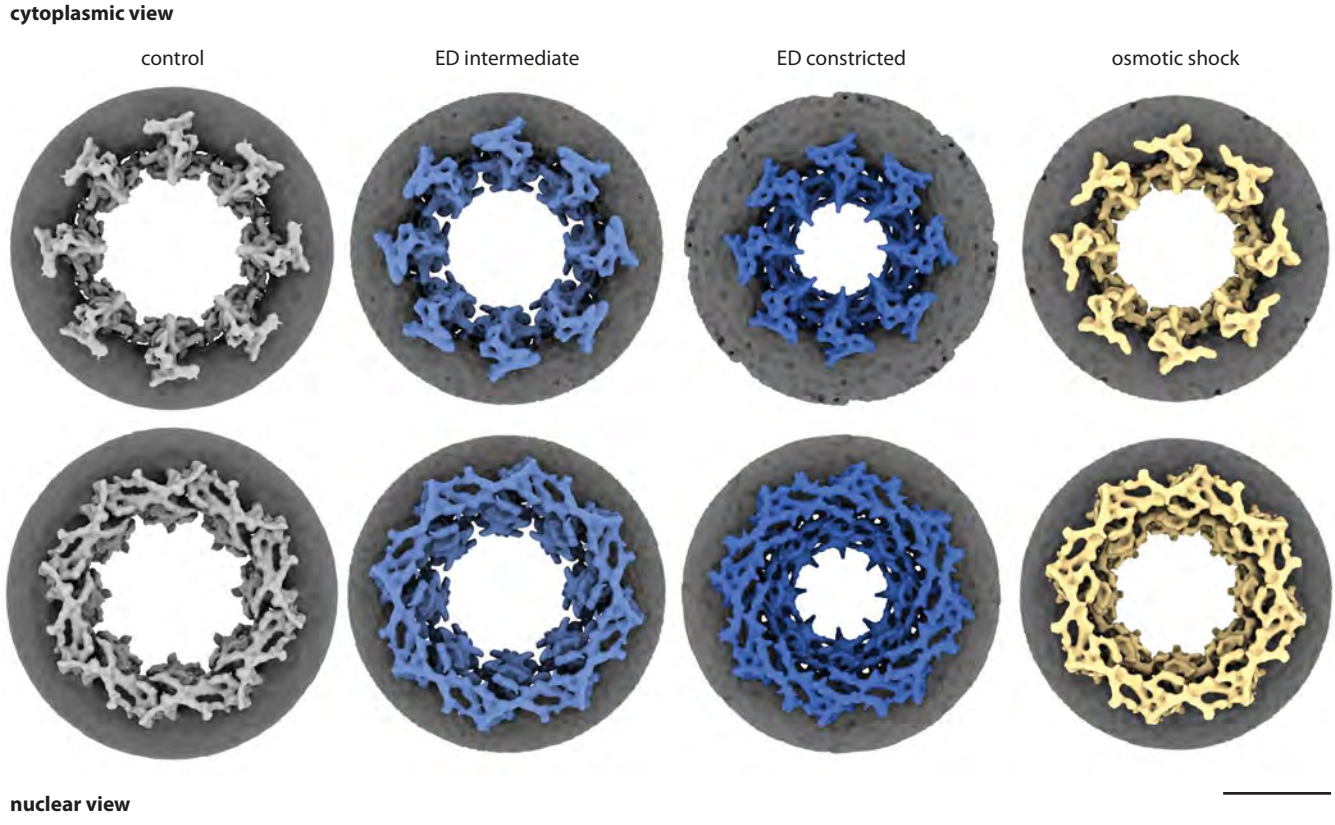


Fig. 3. NPCs constrict during ED and OS. Cytoplasmic and nuclear views of NPC conformations during control, intermediate, and fully constricted ED and OS conditions, illustrating the overall conformational change leading to a central channel diameter constriction from ~70 to <50 nm. The cytoplasmic and IR spokes move as individual entities and contribute the most to the central channel diameter change, whereas the NR constricts to a lesser extent. Scale bar, 50 nm.

spokes. By contrast, in the constricted state, the spokes form extensive contacts (Fig. 4, A and B) equivalent to those in the previously published structures of the human NPC in isolated NEs (*12*). It is plausible that such opening and closing of peripheral gaps may regulate the translocation of INM proteins (*16*–*18*). Notably, under conditions of ED, additional density is arching out into the lumen of the NE (Fig. 4, A, B, and D), contrasting control conditions. It has been previously proposed that such luminal structures are formed by Pom152 (ScPom152 or HsGP210) (*47*, *48*). In terms of their shape, the observed arches are reminiscent of those observed in isolated *Xenopus laevis* NEs (*28*). Our data imply that the luminal ring conformation becomes more prominent upon constriction. Whether this has any mechanical benefits to keep NPCs separated (*28*), or limit the maximal dilation or constriction, remains unclear.

Hyperosmotic shock results in constriction of the central channel

We noticed that ED was concomitant with a substantial reduction of nuclear size and speculated that it could be functionally related to NPC constriction. To test this hypothesis, we triggered nuclear shrinkage independently from ED. We caused a OS by exposing cells to 1.2 M sorbitol. This treatment leads to a pronounced loss of cellular and nuclear volume (*49*). A cryo-EM map from 200 NPCs (Fig. 3; Fig. 4, C and D; and fig. S15A) revealed a strongly constricted conformation. Individual structural features, such as tight lateral packing of IR segments and arching out of the luminal domain, were very similar to those observed with ED, with one notable exception: The N-terminal domain of Nup107 was now clearly resolved at the cytoplasmic side, with the expected angle of engagement of the Nup189C-Nup107 interface (fig. S15B) and consistent with the previously described hinge of the Y-complex (*39*). This observation may be interpreted as either compositional rearrangement or reduced flexibility. Notably, our findings emphasize that nuclear shrinkage correlates with NPC constriction independent of ED.

Passive nucleocytoplasmic transport is reduced under ED and OS conditions

To assess whether the NPC constriction under OS and ED conditions correlates with reduction of passive transport across the NE, we performed fluorescence recovery after photobleaching (FRAP) experiments of nuclei at different time points after ED in cells expressing freely diffusing GFP, as compared with control conditions (Fig. 5, A and B, and fig. S16, A and B) (see Materials and methods). GFP diffusion rates into the nucleus were significantly decreased upon ED (Fig. 5C), contrasting a minor, negligible effect observed within the cytoplasm (fig. S16C). Recovery experiments restored the initially observed diffusion rates (fig. S16D). Under OS conditions, nucleocytoplasmic diffusion decreased to a much lesser extent (Fig. 5, D to F). In contrast to ED, cytoplasmic diffusion also decreased (fig. S16, E to I) in line with previous reports, which demonstrate that OS strongly increases cellular crowding (*42*, *50*). Because ED also leads to a considerable cellular reorganization, including solidification of the cytoplasm, pH-shift, and water loss (*42*, *51*–*53*), it is difficult to disentangle whether the observed reduction in passive exchange rates is a direct consequence of NPC constriction. How exactly NPC constriction affects the passive and active transport of different types of cargos remains to be determined.

NE tension regulates NPC diameter

Scaffold Nups do not have any known motor activity that would allow them to exert forces.

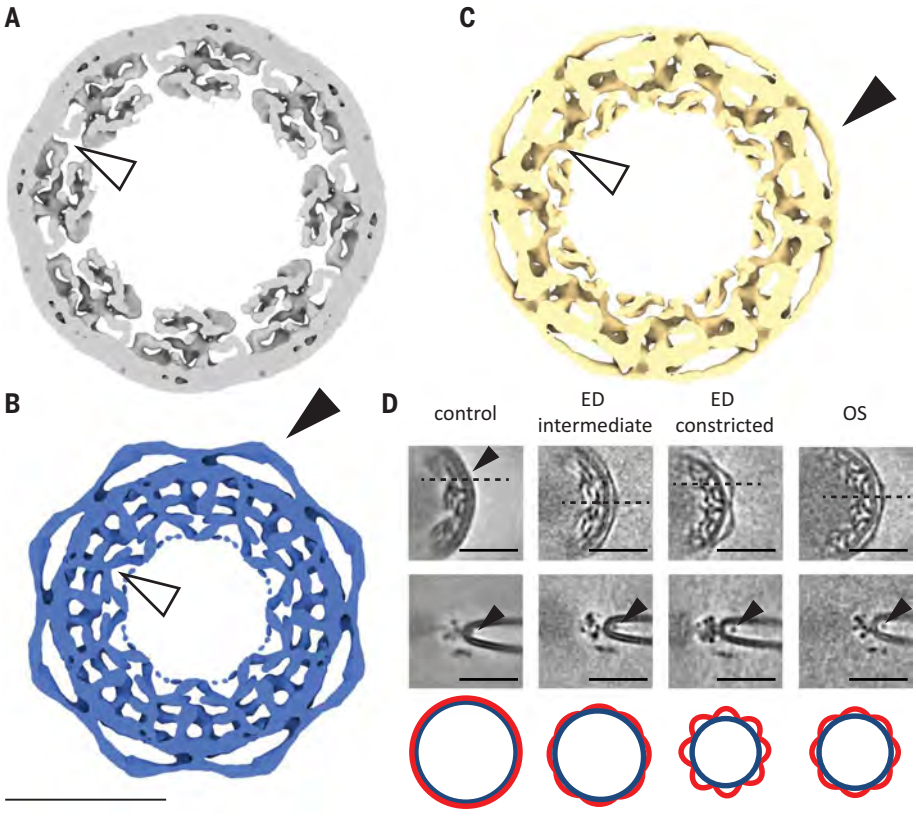


Fig. 4. Conformational changes of the luminal ring and IR during NPC constriction. (**A** to **C**) Central slice through an isosurface representation of NPCs under control (A), most-constricted ED (B), and OS (C) conditions. Under control conditions, IR spokes are clearly separated by 3- to 6-nm gaps and form extensive contacts during NPC constriction (white arrowheads). Luminal densities (black arrowheads) are arching into the NE lumen during NPC constriction. Scale bar, 50 nm. (**D**) Horizontal slices through the IR of cryo-EM maps of control, intermediate, and most-constricted ED and OS conditions (left to right). NPCs reveal the appearance of an arch-like density (black arrowheads) during constriction, likely corresponding to the C-terminal domains of Pom152, whereas the N terminus is attached to the NPC (*48*), which is consistent with previous work on isolated NEs (*28*). Dashed lines indicate the position of the orthogonal slices (below) with the luminal density marked by arrowheads. The cartoons (bottom) indicate how a constriction of the nuclear pore membrane (blue) leads to an arching out of Pom152 C-terminal luminal domains (red), whereas the N termini remain closely attached to the NE. Scale bar, 50 nm; slice thickness, 1.35 to 1.38 nm. In control conditions, the luminal densities are directly proximal to the NPC and thus not easily discernible from the membrane.

Thus, it appears more likely that NPC diameter changes result from mechanical forces that are applied externally. Membrane tension, as the energetic cost to expand the membrane area, results in lateral forces within the NE. We hypothesized that the rigidity of the NPC scaffold counteracts lateral forces imposed by NE tension. In such a scenario, an excess of membrane induced by nuclear shrinkage during OS and ED conditions would reduce tension and result in NPC constriction. This would be consistent with the fact that the most constricted conformation of NPCs is observed in isolated NEs where no mechanical forces act on the scaffold (*10*–*12*, *27*–*29*, *40*).

A model in which membrane tension pulls apart the NPC scaffold to dilate the central channel makes several predictions: (i) Perturbation of the structural integrity of the NPC scaffold should cause NPC dilation; (ii) the more rigid parts of the *S. pombe* scaffold, such as the NR, should be less responsive to lateral forces and constrict less compared with the more fragile part, namely the IR and cytoplasmic side, which in *S. pombe* does not form a rigid ring; (iii) nuclear shrinkage and deformation that are indicative of reduced NE tension (*54*, *55*) should coincide with NPC constriction; (iv) a reduced distance between the INM and outer nuclear membrane (ONM) should directly correlate with NPC constriction; and (v) recovery of nuclear volume, shape, and INM-ONM distance should restore the initial, dilated NPC diameters. Regarding the latter two predictions, it had been proposed that a reduced distance between the INM and ONM is indicative of reduced pressure within the NE lumen (*56*). Using membrane elastic theory, we show that osmotic pressure in the NE lumen and membrane tension in the NE act equivalently (supplementary text) and that the INM-ONM distance grows linearly with both.

To test predictions (i) and (ii), we systematically measured the NPC diameter across all investigated conditions based on centroids of opposite asymmetric subunits as obtained by STA (Fig. 6A and fig. S17, A to F) (see Materials and methods). Knockouts of the scaffold Nups Ely5 and Nup37 caused a dilation of the scaffold (Fig. 6A). We quantified the diameters of each ring under NPC-constricting conditions and found that although all three rings constrict significantly, the conformational changes are considerably smaller at the NR that has a more elaborate scaffold as opposed to the cytoplasmic side that does not form a closed ring (Fig. 6B and movies S4 to S6). These findings are consistent with our model.

To test a relationship of nuclear shrinkage and deformation with NPC diameter [prediction (iii)], we quantified nuclear volume and shape on the basis of three-dimensional (3D) reconstructions of nuclear membranes from confocal fluorescence light microscopy z-stacks and segmentation of nuclear membranes in cryo–electron tomograms for ED- and OS-treated cells. The nuclear volume was reduced in both cases and was reverted when cells were recovered from ED (Fig. 7, A to D, and fig. S18A). Nuclear shrinkage should lead to an excess of nuclear membranes if the total nuclear surface area remains constant, which we verified by quantification of the number of NPCs per NE surface in electron tomograms (Fig. 7E). Under conditions of OS, nuclei were also considerably deformed (Fig. 7, B, F, and G, and fig. S18, B and C). Thus, nuclear shrinkage is concomitant with NPC constriction in both OS and ED conditions (Fig. 7H). Nuclear deformation coincides with NPC constriction during OS but is less apparent during ED (Fig. 7, A and I, and fig. S18, C and D), which we attribute to the missing energy-dependent nuclear deformation forces derived from microtubules (*57*, *58*). Alternatively, the more consistent response of cells to OS compared with ED conditions may explain this effect.

Next, we investigated a relationship between the distance between the INM and ONM and the NPC diameter [prediction (iv); supplementary text]. Under both ED and OS conditions, the average INM-ONM distance was reduced from 22.6 nm in control cells to 15.8 and 19.5 nm, respectively, and reverted to control levels upon recovery from OS, further underlining a loss of NE tension (*56*, *59*) (Fig. 7J) and in agreement with our theoretical model (supplementary text). On the level

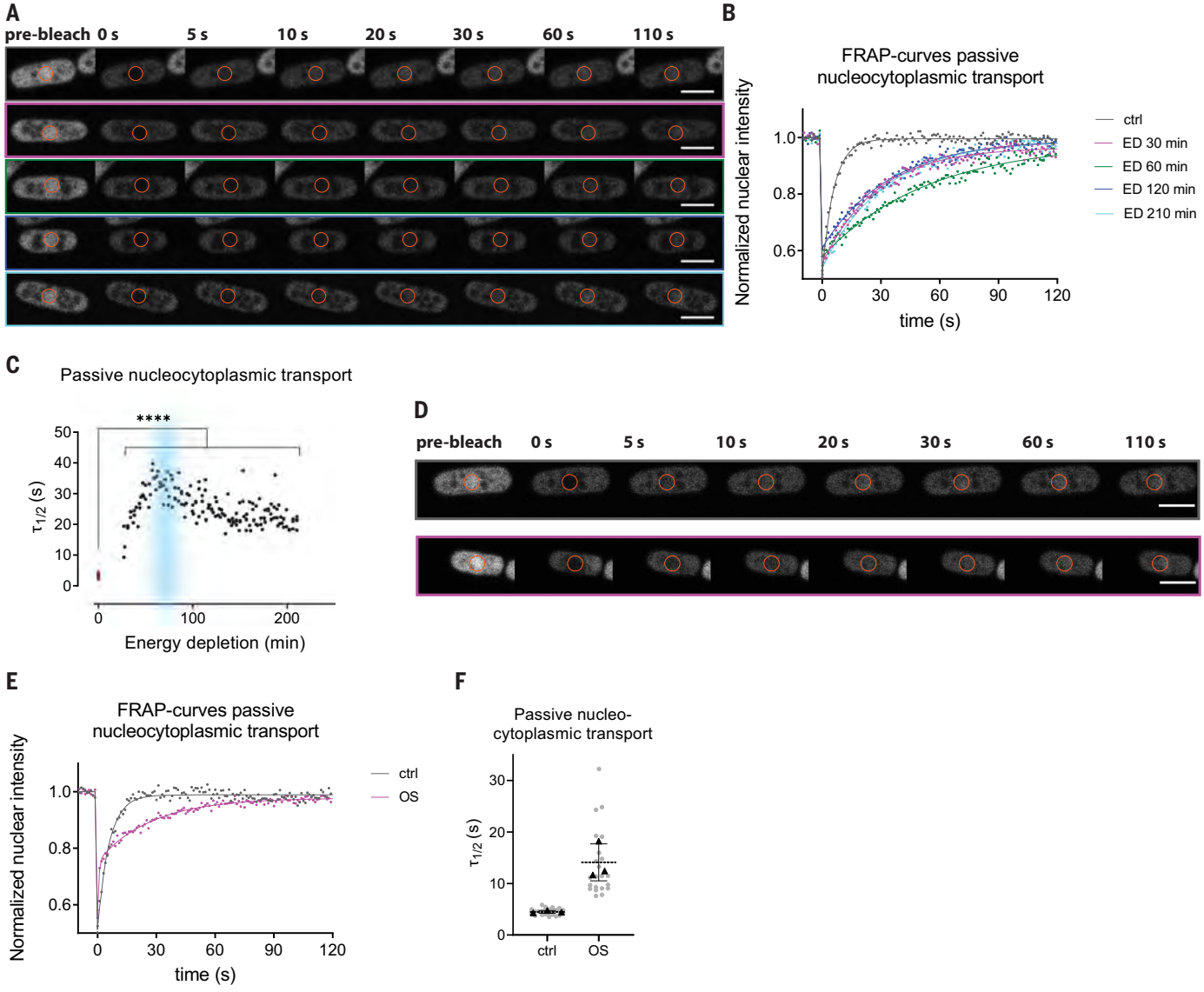


Fig. 5. FRAP experiments and quantification of passive nucleocytoplasmic transport during ED and OS. (**A**) Representative FRAP images from time series of passive nucleocytoplasmic diffusion acquired during control conditions (gray) and at 30 min (purple), 60 min (green), 120 min (dark blue), and 210 min (cyan) after ED (from top). Scale bars, 5 μm. (**B**) FRAP curves corresponding to time series in (A). ctrl, control. (**C**) FRAP recovery half-life times of nuclear signal from freely diffusing GFP at various time points during ED are significantly longer compared with control conditions (red dots). Passive transport of free GFP reaches a minimum after ~1 hour of ED and subsequently recovers slightly. The blue area shows the time point at which cryo-EM grids were prepared for structural analysis of ED NPCs. Data are from three experiments, and triangles indicate means of individual replicates (n = 24 cells in control and n = 154 cells in ED time series). Unpaired t test; ****P < 0.0001. (**D** and **E**) Representative snapshots of passive nucleocytoplasmic transport in control (top) and OS conditions (bottom) (D) and corresponding FRAP curves (E). The orange circles in (D) indicate the bleached area during FRAP experiments, and the time after bleaching is indicated on top. Scale bars, 5 μm. (**F**) Nuclear GFP signal shows a strong increase in recovery half-life time upon OS. Data are from three experiments, and triangles indicate means of individual replicates (n = cells 24 in control and n = 22 cells in OS).

of individual tomograms, the average NPC diameter correlated strongly with the average INM-ONM distance (Spearman correlation coefficient = 0.4803) (Fig. 7K).

Finally, we established that NPC constriction is reversible [prediction (v)]. Because OS conditions lead to a strong and homogeneous constriction of NPC diameters, we performed an OS recovery experiment. We analyzed 278 NPCs structurally by STA from cells that were previously exposed to a OS and shifted back to glucose control medium. Measurements of individual NPC diameters confirmed NPC diameter reversibility (Fig. 6A). As expected, the INM-ONM distance and nuclear volume recovered concomitantly, whereas the nuclear membrane surface–to-NPC ratio remained constant (Fig. 7, E, J, and K, and fig. S18D). These findings further underscore a model in which increasing NE tension causes NPC dilation, whereas a reduction of NE tension allows the NPC scaffold to relax and constrict (Fig. 8).

Conclusions

Our study reveals that NPCs within living cells populate a much larger conformational space than previously anticipated. Massive conformational changes mediate NPC constriction and dilation in response to physiological cues,

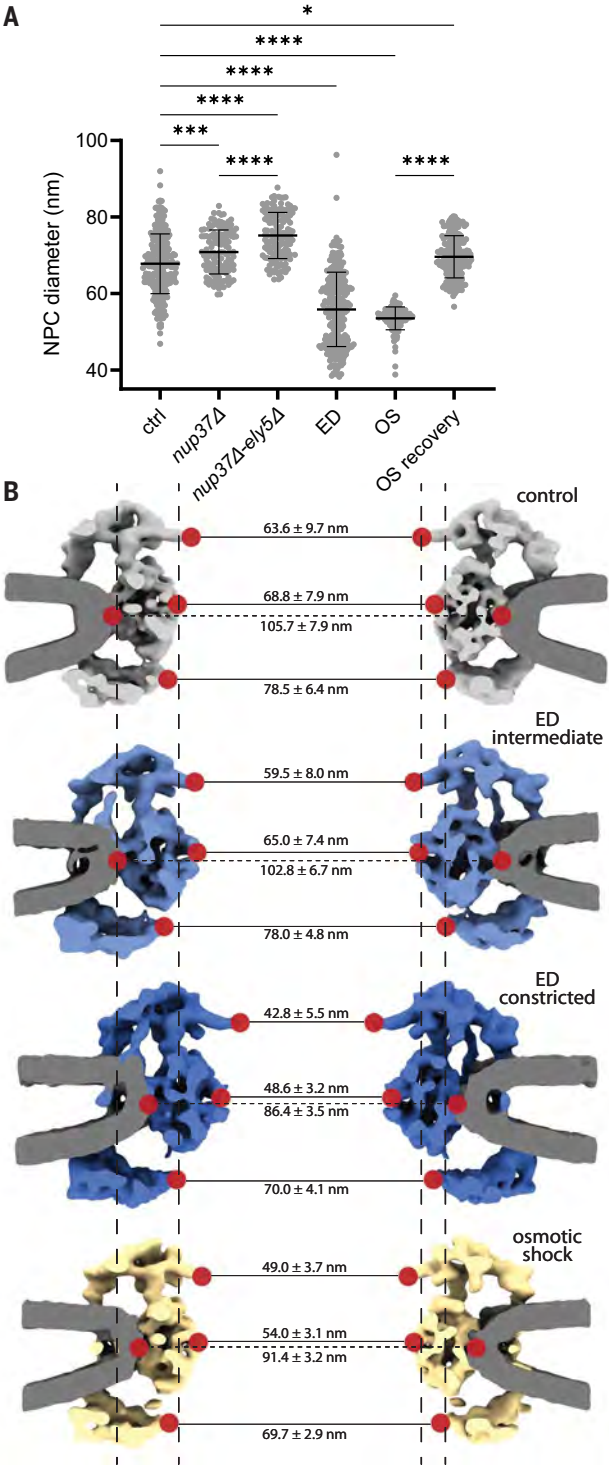


Fig. 6. NPC diameter measurements across different conditions. (**A**) Analysis of NPC central channel diameters measured at the equatorial center of the IR subcomplex based on subunit positions obtained by STA under control, *nup37*Δ, *nup37*Δ-*ely5*Δ, ED, OS, and OS recovery conditions. Data are from one or more experiments [means ± SDs; n = 270 NPCs (control), n = 129 NPCs (*nup37*Δ), n = 141 NPCs (*nup37*Δ-*ely5*Δ), n = 271 NPCs (ED), n = 141 NPCs (OS), and n = 197 NPCs (OS recovery)]. One-way analysis of variance (ANOVA) and Šidák's multiple comparison test; ****P < 0.0001; ***P < 0.001; *P < 0.05 (see also fig. S17). (**B**) Same as Fig. 3 but shown as cutaway side view and overlaid with diameter measurements of individually aligned rings. Measurements are taken at the INM-ONM fusion points (dashed horizontal lines) and the most centrally exposed scaffold points (solid horizontal lines). Positions outlining the central channel were chosen corresponding to the tip of the mRNA export platform in the cytoplasmic side, the equatorial center of the IR subcomplex, and the tip of the inner nuclear Nup85 arm. The vertical lines represent the positions of the respective measurement points in the control conformation. Data are from one or more experiments (means ± SDs; n = 270 NPCs in control, n = 136 NPCs in intermediate ED, n = 68 NPCs in most constricted ED, and n = 141 NPCs in OS).

such as the energy status of the cell or exposure to OS. Notably, ED resulted in a large heterogeneity of NPC constriction levels and two subpopulations with intermediate and strongly constricted NPC diameters. The average diameter observed under OS conditions was more defined at ~55 nm, with a relatively small standard deviation (Fig. 6A). This finding is consistent with the idea that OS affects membrane tension more directly than ED.

The observed conformational changes involve stretching the Y-complexes but are particularly pronounced at the IR, where the spokes move alongside the fused INM and ONM. Notably, the spokes do not move entirely as rigid bodies, but some conformational changes occur within the Nup155 and Nsp1 complex regions (movies S7 and S8). Those are, however, distinct from the previously proposed conformational sliding (*15*) and consistent with an overall preserved intrasubcomplex arrangement (*19*, *20*).

It has been suggested that mechanical forces directly applied to the NE could regulate transport of mechanosensitive cargo and alter NPC diameters (*25*, *60*, *61*). Our data point to a model where NE tension regulates NPC diameter (Fig. 8). During OS and ED, NPC constriction is dependent on nuclear shrinkage and a reduction in INM-ONM distance, which both indicate a reduced NE tension (Fig. 7). Furthermore, the conformational changes of the NPC scaffold appear to be dictated by the movement of the nuclear membranes. During dilation and constriction, the IR spokes move strictly together with the fused INM and ONM. Finally, the knockout of the nonessential Y-complex members Nup37 and Ely5 leads to an increased flexibility and NPC dilation compared with control cells (Fig. 6A), suggesting that the NPC counteracts lateral forces imposed by NE tension, which keep it in an open conformation. Thus, the loss of NE tension leads to a relaxation and constriction in both ED and OS conditions.

The unexpected split Y-complex arrangement of *S. pombe* breaks the long-standing dogma of a three-ringed architecture. This differential cytoplasmic and nuclear architecture, however, helps in assessing how the double–Y-complex arrangement contributes to the mechanical robustness of NPC architecture. The double–Y-complex ring arrangement of the NR constricts much less than the IR and the cytoplasmic structures. Knockouts of Y-complex components within the head-to-tail contact region destabilize the NR and lead to NPC dilation. These findings suggest that the Y-complex ring arrangement provides rigidity to the overall cylindrical architecture. It thus appears plausible that mechanical stress load on the NE—e.g., during cell migration or in osmotically variable environments that may result in unfavorable NE rupture events—imposed an evolutionary selection pressure for higher numbers of Y-complexes and ring formation in mammals or algae. By contrast, organisms less exposed to mechanical NE stresses, such as *S. cerevisiae* or *S. pombe*, reduced their Y-complex copy numbers and head-to-tail contacts during evolution. The NPC scaffold may have a wider range of functions beyond providing grafting sites for FG-Nups. It conformationally responds to mechanical cues and thus may have stabilizing but also mechanically sensory functions (*25*, *60*, *61*).

Materials and methods

S. pombe *culture and cryo-EM grid preparation*

Frozen stocks of *S. pombe* cells were freshly thawed and maintained on YES-agar plates

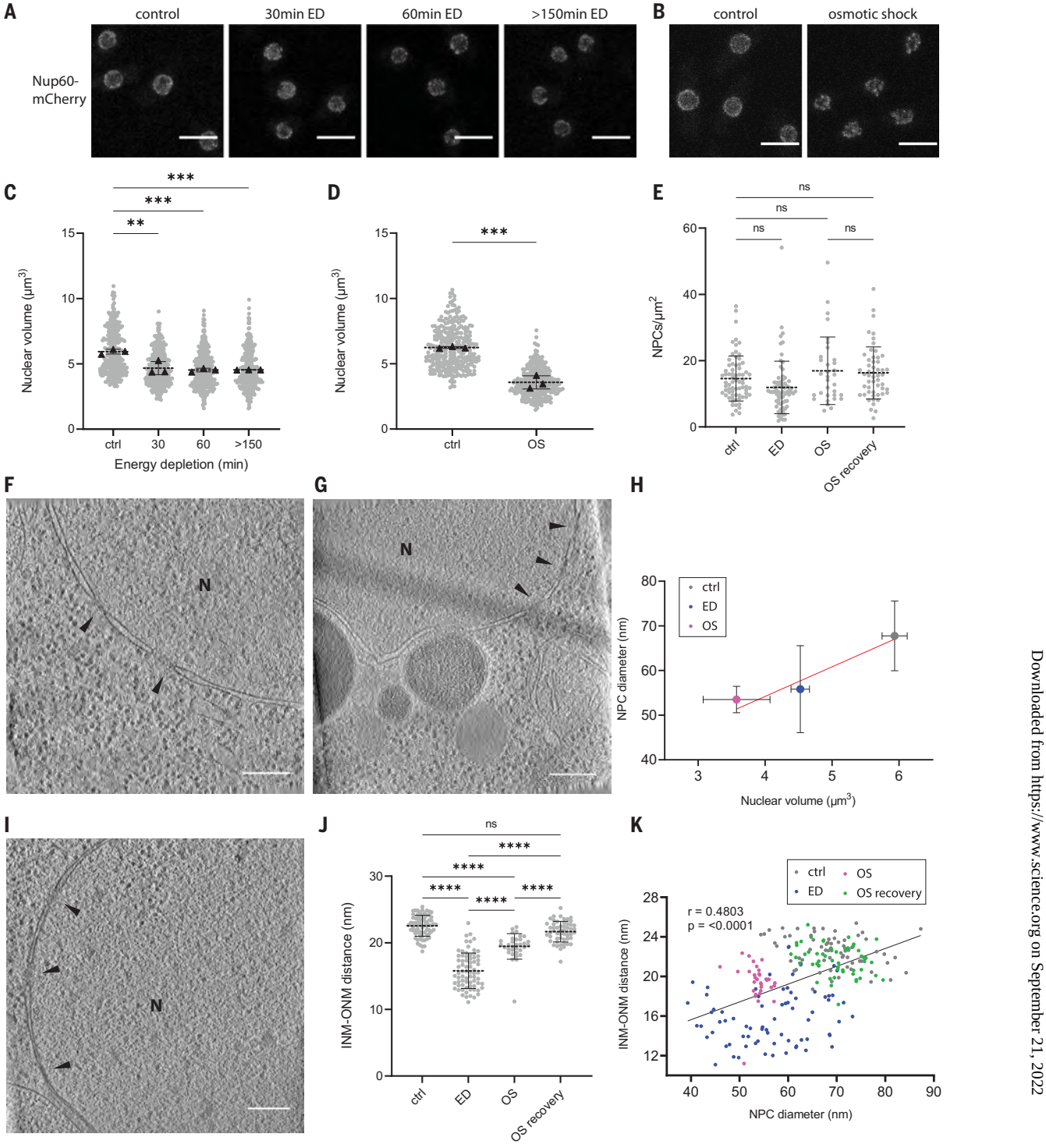


Fig. 7. NE tension is reduced during ED and OS and leads to NPC constriction. (**A**) Maximum projections of Nup60-mCherry channel from live-cell imaging confocal stacks of NLS-GFP-NLS and Nup60-mCherry–expressing cells used for nuclear shape and volume analysis. Scale bar, 5 µm. (**B**) Same as (A) but for cells expressing ubiquitous sfGFP and Nup60-mCherry under OS conditions. (**C** and **D**) Quantification of nuclear volume based on 3D reconstructions of Nup60-mCherry signal corresponding to (A) and (B). Data are from three experiments, and triangles indicate means of individual replicates (means ± SDs; n = 300 nuclei per condition under ED and n = 300 nuclei per condition under OS conditions). One-way ANOVA and Dunnett's multiple comparison test; ***P < 0.001; **P < 0.01. (**E**) Number of NPCs per estimated NE surface area under control, ED, OS, and OS recovery conditions (see Materials and methods). Data are from one or more experiments (means ± SDs; n = 71 tomograms under control, n = 72 tomograms under ED, n = 33 tomograms under OS, and n = 53 tomograms under OS recovery conditions). One-way ANOVA and Šidák's multiple comparison test; ns indicates not significant, P > 0.05. (**F** and **G**) Deconvolved slice through representative tomograms under control (F) and OS conditions (G). N marks the nucleus, and arrowheads mark NPCs. Scale bars, 200 nm. (**H**) Mean NPC diameter (see also Fig. 6A) plotted against mean nuclear volume [same as (C) and (D)] under and control and 60 min after ED and OS conditions. Linear regression

(red line) of mean nuclear volume values against scattered distribution of NPC diameters with means ± SDs and *n* values from (Fig. 6A) indicates a correlation between nuclear volume and NPC diameter. Data are from one or more experiments [coefficient of determination (R^2) = 0.3598]. (**I**) Deconvolved slice through representative tomograms under ED conditions. N marks the nucleus, and arrowheads mark NPCs. Scale bar, 200 nm. (**J**) Measurement of median INM-ONM distance per tomogram under control, ED, and OS conditions shows a clear decrease of INM-ONM distance during ED and a slight decrease during OS. Data are from one or more experiments (means ± SDs; *n* = 71 tomograms under control, *n* = 72 tomograms under ED, *n* = 33 tomograms under OS, and *n* = 53 tomograms under OS recovery conditions). One-way ANOVA with Tukey's multiple comparison test; ****$P < 0.0001$; ns, $P > 0.05$. (**K**) The mean NPC diameter per tomogram plotted against the median INM-ONM distance per tomogram shows a strong positive correlation. Data are from one or more experiments (*n* = 71 tomograms under control, *n* = 72 tomograms under ED, *n* = 33 tomograms under OS, and *n* = 53 tomograms under OS recovery conditions). Spearman correlation coefficient (*r*) and *P* values are plotted, and the black line represents a simple linear regression line.

(YES-broth from Formedium +20 g agarose/L) at 30°C for maximum 3 days and re-streaked on fresh YES-plates before liquid culture inoculation. Five to 10 ml YES-broth were inoculated with individual colonies and incubated over night at 30°C shaking at 200 rpm, where OD_{600} (the optical density of a sample measured at a wavelength of 600 nm) was kept below 1.0. Cryo-EM grid preparation was performed with a Leica EM GP plunger with a set chamber humidity of >90% and temperature of 30°C. A volume of 3.5 to 4 µl cell suspension with previously adjusted OD_{600} to 0.3 were applied to Cu200 or Au200 mesh R2/1 SiO_2 grids (Quantifoil) glow discharged on each side for 45 s before sample application. The grids were blotted with Whatman 597 paper for 1 to 2 s and plunge-frozen in liquid ethane or a mix of ethane-propane (60 and 40%) at −189°C or −194°C, respectively. A list of all yeast strains and their source used in this study can be found in table S2.

C. thermophilum *culture and cryo-EM grid preparation*

C. thermophilum (La Touche) *var. thermophilum* [from DSMZ, Braunschweig, Germany (no. 1495)] mycelium was maintained on CCM-plates, as described in (*62*). For cryo-EM grid preparation, Au200 mesh R2/1 SiO_2 grids (Quantifoil) were glow discharged on both sides for 45 s and placed on the CCM-plates at the edge of the mycelium growth rim with the SiO_2 foil facing up. Plates were incubated at 55°C until the mycelium covered at least half of the grid, typically 2 to ~3.5 hours. Grids were then mounted in the chamber of a Leica EM GP plunger with a set chamber humidity and temperature at >90% and 55°C, respectively. Before plunge freezing in liquid ethane at −186°C, 3.5 µl plunging phosphate buffer (5% trehalose, 0.013 M Na_2HPO_4, 0.045 M KH_2PO_4, pH 6.5) were applied to the sample and blotted for 2 s with Whatman 597 paper.

Generation of knockout and fluorescently tagged S. pombe *strains*

A *nup37*Δ knockout cassette containing a clonNat resistance marker corresponding to pFA6a-natMX6 (*63*), flanked by 70-basepair-long homologous sequences to the nup37 gene was synthesized by Geneart and transformed in a K972 h- wild-type (WT) *S. pombe* strain after a modified protocol from (*64*).

The *nup37*Δ-*ely5*Δ strain was generated by crossing the *nup37*Δ strain with an *ely5*Δ strain (*36*), provided by the Schwartz laboratory at MIT in Cambridge, Massachusetts, USA. Sporulation was triggered on sporulation plates prepared from SpoVB salts containing 8.2 g NaAc, 1.9 g KCl, 2.9 ml $MgSO_4$ 1 M, and 4.1 ml 5 M NaCl and 20 g agarose in 1 L H_2O. Tetrads of sporulated colonies were dissected in 40 µl sorbitol 1 M with 10 µl zymolyase-T100. Selected colonies were replica plated on selection plates containing 100 µg/ml clonNat (Jena Bioscience) or 100 µg/ml G418 (Sigma-Aldrich) and double gene knockout was confirmed by polymerase chain reaction (PCR).

To generate the strains CZ001 [expressing ubiquitous free super-folder GFP (sfGFP) with a Nup60-mCherry marker for FRAP analysis] and CZ007 (expressing NES-sfGFP-NES construct with a Nup60-mCherry marker), a Nup60-mCherry tagging cassette was amplified by PCR from isolated genomic DNA of the GD250 strain (*65*), provided by the Baum laboratory at MRC LMCB London, UK. The product was transformed into the sfGFP or the NES-sfGFP-NES expressing strain (AV0890 and AV1201, respectively) (*66*) obtained from the Yeast Genetic Resource Center (YGRC), as described above.

ED of S. pombe *cells*

Overnight cultures of *S. pombe* cells were grown at 30°C in 5 ml dextrose-free Edinburgh minimal medium (EMM) supplemented with 20 mM glucose (glucose-control medium) while cell density was kept below OD_{600} 0.8. Cells were collected and adjusted to obtain a cell amount of ~1 ml of cells with OD_{600} ~0.3 by centrifugation in Eppendorf tubes at 4000 rpm for 3 min at room temperature. Cells were then washed three times with 1 ml of ED buffer [20 mM 2-deoxy-glucose (Sigma-Aldrich) in dextrose free EMM and 10 µM antimycin A (Sigma-Aldrich)] and resuspended in 1 ml of ED buffer. Control samples were washed three times in glucose control medium. For EM-grid preparation cells were incubated for 1 hour at 30°C on a table-top shaker at 900 rpm. The cell concentration was adjusted to an OD_{600} of 0.2 before EM-grid preparation, as described above.

ED recovery experiments and colony counting viability assays

For recovery experiments, cells were energy depleted as described above and shifted back to glucose control medium ~1 hour after ED. Light microscopy experiments were performed ~20 min after shift back to energy rich medium.

For colony counting, three biological replicates of WT cells were cultivated and energy depleted for 1 hour, as described above. The corresponding control samples were washed in glucose control medium. All samples were then washed and taken up in glucose control medium and 1000 cells were plated on fresh YES plates in duplicates. Colonies were counted after incubation at 30°C for 3 days, and values were reported as mean survivability corrected for plating efficiency in control conditions. For spot assays replicates were prepared as described above and the concentration of each sample was adjusted to OD_{600} = 0.1. A serial dilution with 10x dilution steps was then spotted on a fresh YES plate.

OS of S. pombe *cells*

OS experiments were performed with *S. pombe* cells expressing ubiquitous sfGFP and mCherry-tagged Nup60 (strain CZ001). Cells were grown at 30°C in 5 ml dextrose-free Edinburgh minimal medium (EMM) supplemented with 20 mM glucose (glucose-control medium) while cell density was kept below OD_{600} 0.8. Cells for plunge freezing were grown to OD_{600} of 0.6 and concentrated 2x in 1.5 ml Eppendorf tubes by centrifugation at 4000 rpm for 3 min before OS, whereas cells for light microscopy analysis were grown to OD_{600} 0.4 before OS exposure. The OS was induced by diluting the cell suspension 1:1 with glucose control medium containing 2.4 M sorbitol, resulting in a OS exposure to 1.2 M sorbitol. Before plunge freezing cells were checked by fluorescent light microscopy to confirm the OS reaction and cryo-EM grids were subsequently prepared as described above 1 to 12 min after OS exposure. Light microscopy experiments were performed as described below. In OS recovery experiments cells were treated with 1.2 M sorbitol for ~10 min as described above and subsequently shifted back to glucose control medium where they were allowed to equilibrate for 30 to

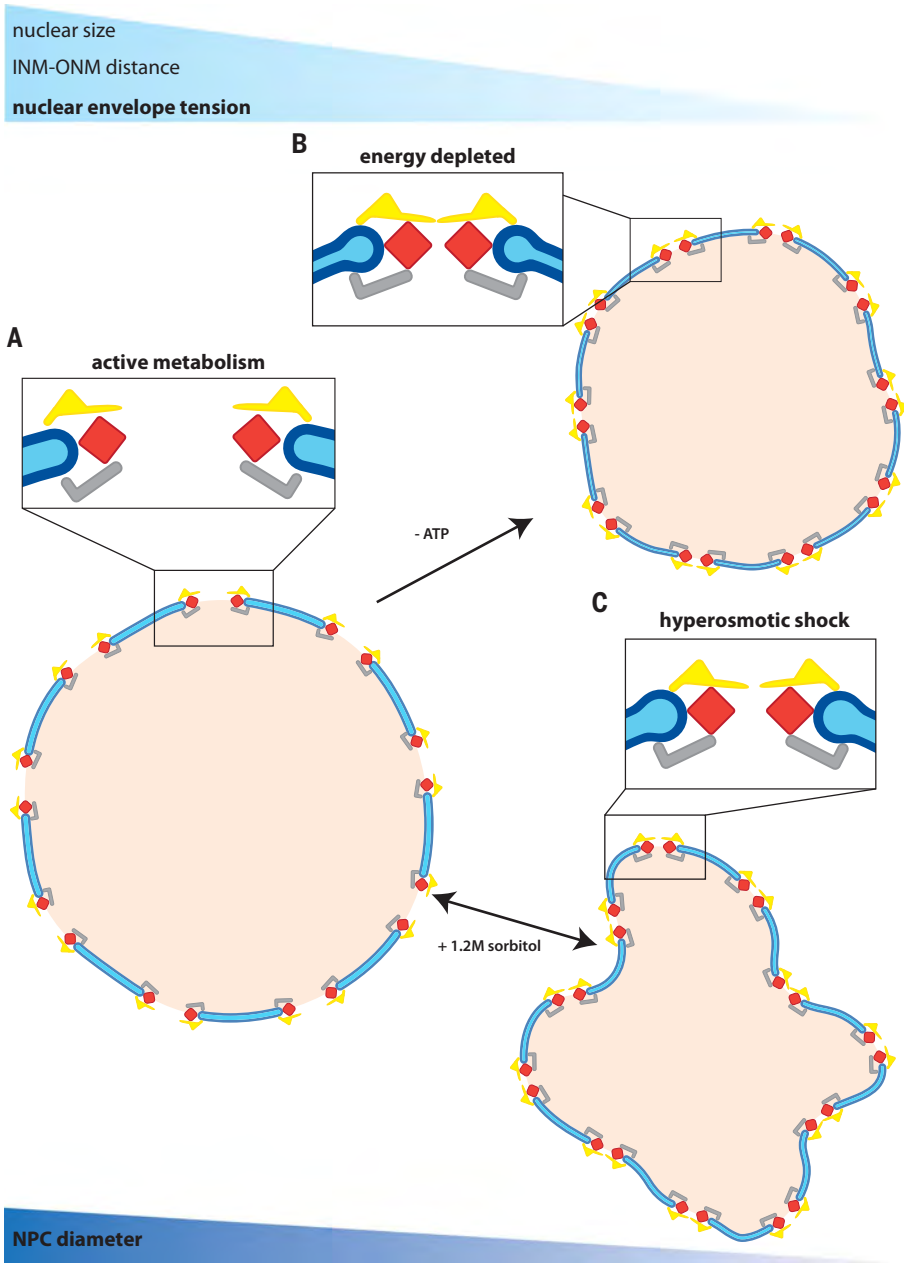


Fig. 8. Conceptual model of how of NE tension regulates NPC diameter. (**A**) NPCs adopt a dilated conformation in metabolically active cells. (**B** and **C**) Both ED (B) and OS (C) lead to a shrinkage of nuclear volume, a reduced INM-ONM distance, and an increased nuclear deformation, all of which point to a reduction in NE tension under both conditions. Such reduced NE tension leads to a constriction of the NPC scaffold architecture, where the central channel volume is reduced to approximately half compared with that in metabolically active cells.

40 min before cryo-EM grid preparation, as described above.

Live-cell fluorescence microscopy and FRAP imaging

Ibidi μ-Slide eight-well Ibidi-treat dishes were coated with 1 mg/ml Concanavalin A (Sigma-Aldrich) in phosphate-buffered saline (PBS) for 10 to 40 min at room temperature and washed three times with ddH_2O. Then, 300 μl of energy-depleted cells were seeded into the Ibidi μ-Slide dish and 150 μl of control cells were seeded and topped up with 150 μl glucose control medium to lower the amount of seeded cells and allow cells to keep dividing. Time of ED was measured from the first wash in ED buffer, typically ~12 min before finalizing the final wash and seeding eight-well dishes. Live-cell imaging and FRAP experiments were all performed on a laser scanning confocal microscope Zeiss LSM-780 equipped with an incubation chamber maintained at 30°C. All images were acquired with a Plan-Apochromat 63x/1.40 Oil differential interference contrast (DIC) objective.

For live-cell imaging of GD250 and CZ007 1532x1532 pixel z-stacks were acquired with a pixel size of 0.0887 μm, an optical section of 0.8 μm, and z-step of 0.38 μm using sequential line scanning and two line averaging rounds with 0.426 μs pixel-dwell time. GFP was excited with a 488 nm line of argon laser and detected with GaAsP spectral detector in the range 500 to 550 nm. mCherry was excited with a DPSS laser at 561 nm and detected with GaAsP spectral detector in the range 568 to 620 nm. Per image stack, 15 to 17 z-slices were acquired to cover the entire *S. pombe* cells. In total, three independent biological replicates were acquired in three independent experiments with each containing about five individual z-stacks. For energy-depleted cells, z-stacks were acquired at time intervals of 30 to 50 min, 55 to 70 min, and 150 to 180 min after ED (after first wash with ED buffer), and control conditions were acquired more than 2 hours after sample preparation. For energy-depleted cells, data acquisition of each time point was started with the first z-stack the given time point after ED. Z-stacks of osmotically shocked CZ001 cells were acquired 1 to 16 min after hypertonic medium exposure.

FRAP measurements of bleached nuclear and cytoplasmic GFP signal in the CZ001 strains were performed to measure passive nuclear transport and cytoplasmic diffusion, respectively.

Passive nucleocytoplasmic transport was measured by individual FRAP experiments performed in the GFP channel consisting of a time series of 140 152x90 pixel (13.6 μm × 8.1 μm) frames acquired over 138.45 s with an interval of 1 s between different frames, a frame time of 0.17 s, and a pixel size of 0.09 μm. The 488 nm line of the argon laser was used both for bleaching and for excitation during time lapse imaging. Bleaching was performed after acquiring 10 images. One bleaching iteration of 1.2 s was performed in a circular area with a diameter of 33 pixel (2.97 μm), previously assigned manually based on the Nup60-mCherry NE outline to cover the nuclear area. Per individual biological replicate independent time series were acquired on individual cells over a period 30 to 210 min after ED, i.e., after the first wash with ED buffer. In total, 48 to 55 FRAP experiments per biological replicate were performed consecutively with an interval of ~3 to 5 min between the individual measurements. Control measurements were performed after completion of

data acquisition on energy-depleted cells i.e., >210 min after sample preparation.

To measure cytoplasmic diffusion, FRAP datasets were acquired in the GFP channel with image dimensions of 80x30 pixel (13.6 μm × 5.1 μm), a pixel size of 0.17 μm, and 140 frames over a period of 2.11 s with a time interval between frames equal to 0.03 s. A rectangular area covering around half of the *S. pombe* cytoplasm and excluding the nucleus was bleached after acquiring 20 initial images. One bleaching iteration was performed in 20 ms with 100% power of 488 laser line and additional 40% of 405 nm laser line. In three independent biological replicates each, eight individual FRAP experiments at time points 30 min (27 to 44 min), 60 min (55 to 68 min), and >150 min (155 to 168 min) after ED were recorded on individual cells. Control measurements were performed after 180 min of sample preparation. FRAP experiments of CZ001 cells exposed to a OS were performed as described above for ED with the following modifications: Each measurement was performed in three biological and eight subsequent technical replicates ranging over a timeframe of 3 to 25 min after OS exposure. To accommodate for the significantly slower cytoplasmic diffusion upon OS, the imaging parameters described above for nuclear FRAP were used where a 33x33 pixel square was bleached within the cytoplasm and recovery was measured over 120 s with a 1-s interval.

Quantification of fluorescence microscopy images and FRAP analysis

To quantify the localization of NLS-GFP in GD250, z-stacks were maximum projected and the ratio of the average pixel intensity of a circular area within the nucleus compared with a similar area in the cytoplasm was measured manually in Fiji (*67*). The nuclear position was determined in the Nup60-mCherry channel. Cells that were not well attached to the eight-well dish and moving during z-stack acquisition or not entirely contained within the z-stack were removed before analysis, and cytoplasmic areas of clearly visible vacuoles were avoided. NES-GFP localization signal was quantified similarly as described above for NLS-GFP. To avoid signal bias from above and below the nucleus, only a single image slice approximately at the central plane of each cell was quantified.

Nuclear volume and sphericity of GD250 and CZ001 z-stacks were quantified based on 3D reconstructions of individual n of the Nup60-mCherry (with excluded areas of cells moving or not entirely contained within the imaged volume) using the 3DMembraneReconstruction (*68*) workflow in Mathwork's MATLAB.

FRAP curves were analyzed with FRAPAnalzer 2.1.0 (*69*). In brief, a double normalization was carried out normalizing against the background and the entire cellular surface as reference. Recovery half-life times were extracted by fitting a one exponential recovery equation to the normalized data. In the case of nuclear FRAP experiments under OS conditions, a two-exponential fitting was performed to account for the first (fast) recovery most likely corresponding to recovery of molecules bleached outside of the nucleus and a second (slower) recovery corresponding to the nuclear fluorescence recovery, the half-life time of the latter was reported as nuclear fluorescence half-life time. Moving cells and cells where the nuclear signal did not recover to a plateau were excluded from analysis.

Cryo-FIB milling

Plunge-frozen sample grids were FIB milled in an Aquilos FIB-SEM (Thermo Fisher) as previously described (*4*). In brief, samples were sputter coated with inorganic platinum at 10 kV voltage and 10 mPa argon gas pressure for ~10 s (Pt-sputtering). Subsequently a protective layer of organometallic platinum was deposited using the gas injection system (GIS-coating) for ~12 s. If needed, a second round of Pt-sputtering was performed to avoid charging during FIB milling. Between 5 and 10 lamellae per grid were milled in a step-wise fashion and polished to a final thickness of <280 nm using decreasing FIB current steps of 1 nA, 0.5 nA, 0.3 nA, and 50 pA. Before unloading the sample, an additional layer of Pt was deposited for 1 s to avoid charging effects during subsequent imaging in the transmission electron microscope (TEM).

Automated tomogram acquisition

The majority of WT *S. pombe* tomograms (136/178) as well as all tomograms from knockout strains were acquired on a Titan Krios G3 (Thermo Fisher), operating at 300 keV and equipped with a Gatan K2 Summit direct electron detector and energy-filter. All tilt-series (TS) were acquired in dose-fractionation mode at 4k × 4k resolution at a nominal pixel size of 3.45 Å. Automated TS acquisition was performed as described previously (*4*) using a dose-symmetric acquisition scheme (*70*) with an effective tilt range of −50° to 50°, considering a tilt-offset to compensate for the lamella angle (typically positive or negative 8° to 13°) (*4*), a 3° tilt increment, a total dose of 120 to 150 e/Å^2, and target defocus of −1.5 to −4.5 μm.

The remaining 42 WT tomograms and 15 of the total 76 tomograms from energy-depleted *S. pombe* cells were acquired on a Titan Krios (Thermo Fisher) equipped with a Gatan K2 Summit direct electron detector and energy-filter and a Volta potential phase plate (VPP) (*71*). Automated dose-symmetric tilt series acquisition was performed in 4k × 4k dose-fractionation mode with a nominal pixel size of 3.37 Å using a total dose of 100 to 125 e/Å^2 distributed over a tilt range of −60° to +60° or −50° to 50° with and a tilt step interval of 3° or 2°, respectively, a defocus range of −2 to −4 μm, and conditioning the VPP by exposure to the electron beam up to 5 min. An additional 11 tomograms of ED *S. pombe* cells were acquired using the same parameters but without usage of the VPP. All 57 tomograms of cells exposed to a OS were acquired with similar parameters including an effective tilt range of −50° to 50°, considering a tilt-offset to compensate for the lamella angle, a 3° tilt increment, a total dose of 120 to 150 e/Å^2, and target defocus of −2 to −5 μm.

The remaining 50 tomograms of ED *S. pombe* cells were acquired on a Titan Krios G3 (Thermo Fisher) operating at 300 keV equipped with a Gatan K3 direct electron detector and energy-filter operated in dose-fractionation mode and 5.7k × 4k resolution with a nominal pixel size of 3.425 Å. Automated dose-symmetric tilt series acquisition was performed with a nominal defocus range of −2 to −3.5 μm, an effective tilt range of −50° to 50° including a tilt offset to compensate for the lamella pretilt, a tilt step increment of 3°, and a total dose of 133 e/Å^2.

All 78 tomograms of cells recovered from a OS were acquired on a Titan Krios G2 (Thermo Fisher) equipped with a Gatan K3 direct electron detector and energy-filter with similar parameters as above including an effective tilt range of −50° to 50°, considering a tilt-offset to compensate for the lamella angle, a 3° tilt increment, a total dose of 142 e/Å^2, a target defocus of −2 to −5 μm, and a nominal pixel size of 3.372 Å. *C. thermophilum* tomograms were acquired as described above for WT *S. pombe* tomograms with a total dose of 130 to 140 e/Å^2 and a target defocus range of −2 to −4 μm.

Image preprocessing and tomogram reconstruction

Images were preprocessed, and contrast transfer function (CTF) was estimated as described previously (*4*). Tilt series were aligned automatically in a tailored workflow using the IMOD package including the patch-tracking functionalities (*72*). In detail, four-times binned (pixel size: 1.348 nm) and dose-filtered TS were low-pass filtered with seven empirically predefined low-pass filter parameters and seven initial patch-tracking attempts using 7x7 patches per tilt were performed in IMOD. The filter-set and patch tracking output yielding the lowest alignment residuals during the initial round of patch-tracking was selected for further iterative refinement. At each iteration, the contour with the largest residual error was removed before the next round of patch-tracking. This process was repeated until the overall residual dropped below 0.5 pixel or until a ratio of known/unknown in the IMOD tilt

align program dropped below a threshold of 10. The final aligned TS were reconstructed using the SIRT-like filtering option in IMOD (*73*) and manually inspected. Only tomograms yielding a high-quality alignment, both, visually and in terms of overall residual error (typically better than 1 pixel) were used for 3D CTF correction using the phase-flipping option in novaCTF (*74*) and subsequent STA workflow.

NE segmentation and tension analysis

Four-times binned 3D CTF-corrected tomograms were deconvolved in MATLAB using tom_deconv function (https://github.com/dtegunov/tom_deconv) and subsequently segmented using the convolutional neural network (CNN)–based tomogram segmentation workflow in EMAN2.2 (*75*) using a custom trained CNN to recognize NEs. The resulting segmentation was improved by applying the tensor voting–based membrane segmentation workflow TomoSegmemTV (*76*). The final binary segmented volume was manually inspected and curated, whereby false positive density such as NE connected ER was removed and tomograms resulting in overall inferior segmentation quality were excluded. Binary segmented NE volumes were subsequently loaded into MATLAB where a sphere was fitted to the coordinates of each individual segmentation. On the fitted sphere surface a sampling grid was defined with a kernel size (distance between sampling points) of 20 pixel (resulting in a subdivision of the spherical surface in ~27 nm × 27 nm patches). At each sampling point several measurements were performed.

First, at each sampling point the mean radius of all segmented voxels within a 20-pixel thick ray extending form the center of the fitted sphere through the sampling point were measured and the difference between the resulting mean NE radius and the fitted sphere radius was reported as deviation of the NE form a perfect sphere at each given sampling point. For each individual tomogram the variance or standard deviation of the deviation form a perfect sphere was then used as a measure of NE wobbliness (how much the NE deviates form a perfect sphere).

Second, the INM-ONM distance at each sampling point was measured by analyzing the distribution of radii of all voxels belonging to the NE segmentation within the 20-pixel-thick ray at each sampling point. Only radius distributions which showed a bimodal distribution were further used to perform a K-means clustering and obtain the mean NE radius of two classes. Sampling points which did not show a bimodal radius distribution were omitted. The shorter radius was then attributed to the INM radius, whereas larger value described the ONM radius at the given sampling point. The difference between the INM and ONM radius was reported as INM-ONM distance at any given sampling point, and the median over the entire tomogram was used as INM-ONM distance measurement for each individual tomogram.

Third, the total segmented NE surface was estimated by projecting the segmented NE voxels on the fitted sphere surface. Subsequently all sampling points covering at least one NE voxel within the 20-pixel radius were counted as 729 nm^2 (27 nm × 27 nm) membrane patch each and summed up to yield the total estimated NE surface. The number of initially picked NPCs (see below) per tomogram was divided by the corresponding estimated NE surface area to estimate the number of NPCs per square micrometer of NE.

Particle identification and STA

NPC coordinates and initial orientations were determined manually in four-times binned SIRT-like filtered (*73*) tomograms as described earlier (*4*). Additionally, a model describing the lamella slab geometry was generated manually by picking the coordinates of all eight corners of the lamella slab for each tomogram. Particle extraction, subtomogram alignment, and averaging (STA) was performed with novaSTA (*77*) from 3D CTF-corrected tomograms, as described previously (*4*). Initial alignment of NPCs was carried out using eight- and four-times binned particles. After obtaining an initial four-times binned average, coordinates of the eight individual NPC spokes were identified for symmetry-independent alignment, as described before (*6*). In brief, the coordinates of an individual subunit within the eightfold average was defined manually, and the remaining asymmetric subunit positions were calculated based on the eightfold symmetry (fig. S2A). The coordinates were then used to extract each individual asymmetric subunit form the original tomograms, where only subunits with coordinates retained in the lamella slab geometry model were retained to avoid the inclusion of subunits lying outside the FIB-milled lamellae. The individually extracted asymmetric subunits were then used for further alignment using a mask covering the entire asymmetric unit (cytoplasmic side, IR, and NR) (fig. S2B). After an initial subunit alignment, each subtomogram and its assigned orientation was inspected manually and misaligned or remaining false positive particles, i.e., subunits extracted outside of the FIB-milled lamellae, were removed. Subtomogram alignment was further continued using four- and two-times binned subtomograms, where the alignment was focused on the CR, IR, or NR using localized masks (fig. S2C). VPP data were excluded from further STA of the WT averages at the transition from bin4 to bin2 processing. If needed, the particle boxes were recentered around the individual rings to achieve better subtomogram alignment. In case of the WT subtomogram average, particles with low final cross correlation scores were removed for the final average. All subtomogram averages of WT, OS, and knockout datasets were b-factor sharpened empirically (*11*) and filtered to the given resolution determined on gold-standard FSC calculations at the 0.143 criterion. For ED averages, instead of b-factor sharpening, the amplitude of the final averages (containing VPP and defocus data) were matched to the amplitudes of a corresponding average generated by averaging only the defocus type subset of data (excluding all VPP data) with the EMAN2 (*78*) e2proc3d.py -matchto function, to overcome a previously described low-pass filter-like artifact occurring during STA of VPP data (*79*).

To generate the eightfold symmetric assemblies, the final individual ring maps were first fit to the initially aligned whole asymmetric unit map to find their correct relative position. The root mean square density value of all individual maps were scaled to similar values to allow appropriate representation of all three rings at a common threshold level. The resulting asymmetric unit was then assembled to an eightfold symmetric assembly based on the coordinates defined during the initial subunit extraction (see above).

Details about tomogram and particle numbers of each average can be found in table S3.

Difference map calculation

To calculate the difference map of the *nup37*Δ and *nup37*Δ-*ely5*Δ knockout to the WT, the maps of the individual knockout rings were filtered according to their reported resolution (*nup37*Δ cytoplasmic side: 27 Å, IR: 27 Å, and NR: 31 Å; cytoplasmic side: 35 Å, IR: 35 Å, and NR: 34 Å of *nup37*Δ-*ely5*Δ, respectively) (table S3 and figs. S8 and S9), and WT maps were filtered to the corresponding knockout resolution. To account for different b-factors and amplitudes, the power spectra of the WT maps were adjusted to the corresponding knockout map using RELION (*80*) relion_image_handler -adjust_power command. To obtain comparable difference maps across the density of the individual knockout rings, the WT cytoplasmic and nuclear map were scaled to match the root mean square of the corresponding WT IR map. The WT and corresponding knockout maps were then brought to the same reference frame by fitting the maps against each other and difference maps of the individual rings were calculated in UCSF Chimera (*81*). The final maps are shown at the same threshold and overlaid with the difference map (WT minus knockout) and the inverse (negative threshold) of the difference map, respectively (Fig. 2 and figs. S8 and S9)

NPC diameter and volume measurements

NPC diameters were measured with an in-house MATLAB script based on the final coordinates and orientations obtained for each individual subunit during STA. First a feature of interest in the subunit average was identified in UCSF Chimera (*81*) by placing a 1-voxel sphere mask on the point of interest. The offset between the center of the average and the mask was then used to calculate the coordinates of the area of interest within each averaged subunit in respect to the original tomograms. To calculate the diameter of each individual NPC, only NPCs with a subunit-occupancy of five or more were considered. For each individual NPC, vectors were derived connecting the opposing subunits. Based on those vectors, the center of each NPC was defined as the center of the intersection area between all vectors corresponding to a given NPC. The average distance between the newly identified center and each respective SU was chosen as a representative NPC radius for the given pore. By this means, an accurate average NPC radius measurement for a specific feature of interest within each individual NPC is obtained. Additionally, the z-translation of each subunit from the NPC center plane and the pairwise distance opposing subunits based on coordinates obtained from whole asymmetric subunit alignment were reported (fig. S17, A to G).

Homology modeling

Modeling templates were detected and selected using HHpred server (*82*). Sequence alignments for modeling were refined in Swiss PDB Viewer (*83*). The models were built based on the templates and alignments using Modeller (*84*). The templates used for modeling are listed in table S4.

Systematic fitting of SpNPC rigid bodies to cryo-ET maps

To assign the number of Y-complexes and IR subcomplexes, and to prepare the input for integrative modeling (fig. S5A), an unbiased global fitting approach was applied using several SpNPC structural models. Two of them were experimental x-ray structures previously published (*35*, *36*), whereas the remaining were homology modeled on the basis of *S. cerevisiae* and *C. thermophilum* NPC components (*20*, *85*–*88*) (table S4). All the aforementioned high-resolution structures were filtered to 40 Å before fitting. The resulting simulated model maps were subsequently fitted into individual segments of the SpNPC cryo-EM maps from this study by global fitting as implemented in UCSF Chimera (*81*). More precisely, structures of the Y-complex were fitted into cytoplasmic and NR segments of the cryo-EM SpNPC maps while the IR model was fitted in the IR. The segmented maps that were used for fitting did not include NE density to eliminate the possibility of fits substantially overlapping with the membrane.

All fitting runs were performed using 100,000 random initial placements with the requirement of at least 30% of the simulated model map to be covered by the SpNPC density envelope defined at a low threshold. For each fitted model, this procedure resulted in ~250 to 17,200 fits with nonredundant conformations upon clustering. The cross-correlation about the mean (cam score, equivalent to Pearson correlation) score from UCSF Chimera (*81*) was used as a fitting metric for each atomic structure, similarly to our previously published works (*4*, *11*, *12*, *31*). The statistical significance of every fitted model was evaluated as a *P* value derived from the cam scores. The calculation of *P* values was performed by first transforming the cross-correlation scores to z-scores (Fisher's z-transform) and centering, from which subsequently two-sided *P* values were computed using standard deviation derived from an empirical null distribution [based on all obtained nonredundant fits and fitted using fdrtool (*89*) R-package]. Finally, the *P* values were corrected for multiple testing with Benjamini-Hochberg (*90*). Figures were produced by UCSF Chimera (*81*) and UCSF ChimeraX (*91*).

Integrative modeling

Structural models of the SpNPCs were built using the integrative procedure (fig. S5) implemented in our Assembline software (*92*). Assembline implements a custom modeling pipeline based on Integrative Modeling Platform (IMP) (*93*) version 2.14 and Python Modeling Interface (PMI) (*94*). The pipeline is similar to the one we used previously for human and *S. cerevisiae* NPCs (*4*, *12*). The modeling procedure consisted of three steps (fig. S5B).

Step 1

First, ensembles of alternative models of the individual rings in the control NPC were constructed. To model the cytoplasmic Y-complexes and NR, the input homology models were divided into smaller rigid bodies (with cut points corresponding to boundaries of published crystal structures) and simultaneously fitted to the cryo-ET map using the global optimization step in Assembline. In this step, first, libraries of alternative nonredundant fits for each rigid body were generated using the systematic fitting procedure described above. Then, the rigid bodies were fitted simultaneously to the maps of individual rings by sampling alternative fits from these libraries using the simulated annealing Monte Carlo method. To cover a large landscape of possible models, the optimization was repeated 4000 and 40,000 times for the cytoplasmic side and NR, respectively, with each repetition leading to a candidate model. To achieve convergence and sampling exhaustiveness [assessed using the procedure by Viswanath *et al.* (*95*)], each repeat run was composed of 130,000 (cytoplasmic side) and 271,000 (NR) Monte Carlo steps at decreasing temperatures. The higher number of steps and models for NR was used because of the higher number of rigid bodies in this ring. The scoring function for the modeling optimization was a linear combination of the EM fit restraint represented as the *P* values of the precalculated rigid-body fits (from systematic fitting as described above), clash score (SoftSpherePairScore of IMP), connectivity distance between domains neighboring in sequence, a term preventing overlap of the protein mass with the NE, a restraint promoting the membrane-binding loops of Nup131, Nup132, and Nup120 to interact with the envelope, implemented using MapDistanceTransform of IMP [predicted by similarity to known or predicted ALPS motifs in human and *S. cerevisiae* homologs (*1*, *96*)], the restraint for Ely5 localization in the region identified based on the difference between the *nup37*Δ and *nup37*Δ-*ely5*Δ knockout NPC maps to the control NPC, and restraint for biochemically characterized interaction restraint between Ely5 and Nup120 (*36*). For the NR, a restraint promoting Nup37 localization from the difference between the *nup37*Δ NPC and the control NPC was also used. During the optimization, the structures were simultaneously represented at two resolutions: in Cα-only representation and in a coarse-grained representation, in which each 10-residue stretch was converted to a bead. The 10-residue bead representation was used for all restraints to increase computational efficiency except for the domain connectivity restraints, for which the Cα-only representation was used. Because the *P* values used for the EM restraint were derived from the original EM fit libraries generated with UCSF Chimera, the EM restraint can be regarded as an EM restraint derived from the full atom representation.

Owing to high structural conservation of the IR as determined by the systematic fitting, this ring was modeled by fitting a homology model of the entire IR built based on *S. cerevisiae* NPC, followed by the refinement with 50,000 Monte Carlo steps at a single temperature. The refinement involved the Monte Carlo simulated annealing optimization of the rigid bodies' rotations and translations, but this time without using the libraries of alternative nonredundant fits, allowing the rigid bodies to move in the EM map with small rotation and translation increments. The scoring function consisted of cross-correlation to the EM map (FitRestraint of IMP), domain connectivity restraint, the clash score, the envelope exclusion restraints as above, and a restraint for the membrane-binding loop of Nup155 (*12*).

Step 2

The above steps led to ensembles of separate models for the cytoplasmic side, NR, and IR. To build the model of the full NPC, the alternative fits of each subunit in the cytoplasmic side and NR models were extracted into a new set of fit libraries. Then, together with the top-scoring model of the IR, the global optimization step was repeated using the new fit libraries and now fitting all subunits from all three rings of the NPC simultaneously. The resulting models were refined as above for the IR, leading to an ensemble of models of the full NPC. An elastic network restraint was used to keep the Ely5-Nup120 orientation similar in all rings.

Step 3

To generate the models of the ED NPCs, the 1000 top-scoring models of the NPC under control conditions were fitted into the ED EM maps and optimized using the refinement procedure. The scoring function comprised the same restraints as for the refinement of the individual rings except for the restraint for Nup37 localization in the difference density. Each of the 1000 starting models was refined with 150,000 steps at five temperature levels leading to an ensemble of 1000 alternative models.

Model analysis

To assess convergence and exhaustiveness of sampling, and to assess the precision of sampling and of the models, we used the procedure by Viswanath *et al.* (*95*). The quantities were assessed at each modeling step of the procedure (fig. S5B). The sampling precision for the control, ED medium, and constricted diameter were 8.4, 12.6, and 12.8 Å, respectively. At this sampling precision, we obtained two to four clusters of models with the model precision between 7.2 and 9.7 Å (fig. S5B). The clusters differed mostly in the orientations of Ely5, Nup120, and the N-terminal propeller of Nup133. The model with the orientation of Ely5 consistent with the model of the human NPC (*11*) was used as a representative model for the figures. For clarity, a single model was shown in the figures and the top 10 models for each ensemble were shown in fig. S5.

REFERENCES AND NOTES

1. D. H. Lin, A. Hoelz, The Structure of the Nuclear Pore Complex (An Update). *Annu. Rev. Biochem.* **88**, 725–783 (2019). doi: 10.1146/annurev-biochem-062917-011901; pmid: 30883195
2. M. Beck, S. Mosalaganti, J. Kosinski, From the resolution revolution to evolution: Structural insights into the evolutionary relationships between vesicle coats and the nuclear pore. *Curr. Opin. Struct. Biol.* **52**, 32–40 (2018). doi: 10.1016/j.sbi.2018.07.012; pmid: 30103204
3. J. Mahamid *et al.*, Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**, 969–972 (2016). doi: 10.1126/science.aad8857; pmid: 26917770
4. M. Allegretti *et al.*, In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature* **586**, 796–800 (2020). doi: 10.1038/s41586-020-2670-5; pmid: 32879490
5. G. J. Stanley, A. Fassati, B. W. Hoogenboom, Atomic force microscopy reveals structural variability amongst nuclear pore complexes. *Life Sci. Alliance* **1**, e201800142 (2018). doi: 10.26508/lsa.201800142; pmid: 30456374
6. M. Beck, V. Lučić, F. Förster, W. Baumeister, O. Medalia, Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. *Nature* **449**, 611–615 (2007). doi: 10.1038/nature06170; pmid: 17851530
7. J. Sellés *et al.*, Nuclear pore complex plasticity during developmental process as revealed by super-resolution microscopy. *Sci. Rep.* **7**, 14732 (2017). doi: 10.1038/s41598-017-15433-2; pmid: 29116248
8. A. P. Schuller *et al.*, The cellular environment shapes the nuclear pore complex architecture. *Nature* **598**, 667–671 (2021). doi: 10.1038/s41586-021-03985-3; pmid: 34646014
9. V. Zila *et al.*, Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell* **184**, 1032–1046.e18 (2021). doi: 10.1016/j.cell.2021.01.025; pmid: 33571428
10. K. H. Bui *et al.*, Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155**, 1233–1243 (2013). doi: 10.1016/j.cell.2013.10.055; pmid: 24315095
11. A. von Appen *et al.*, *In situ* structural analysis of the human nuclear pore complex. *Nature* **526**, 140–143 (2015). doi: 10.1038/nature15381; pmid: 26416747
12. J. Kosinski *et al.*, Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363–365 (2016). doi: 10.1126/science.aaf0643; pmid: 27081072
13. S. Otsuka *et al.*, Postmitotic nuclear pore assembly proceeds by radial dilation of small membrane openings. *Nat. Struct. Mol. Biol.* **25**, 21–28 (2018). doi: 10.1038/s41594-017-0001-9; pmid: 29323269
14. S. Otsuka *et al.*, Nuclear pore assembly proceeds by an inside-out extrusion of the nuclear envelope. *eLife* **5**, e19071 (2016). doi: 10.7554/eLife.19071; pmid: 27630123
15. J. Koh, G. Blobel, Allosteric regulation in gating the central channel of the nuclear pore complex. *Cell* **161**, 1361–1373 (2015). doi: 10.1016/j.cell.2015.05.013; pmid: 26046439
16. A. C. Meinema *et al.*, Long unfolded linkers facilitate membrane protein import through the nuclear pore complex. *Science* **333**, 90–93 (2011). doi: 10.1126/science.1205741; pmid: 21659568
17. R. Ungricht, M. Klann, P. Horvath, U. Kutay, Diffusion and retention are major determinants of protein targeting to the inner nuclear membrane. *J. Cell Biol.* **209**, 687–703 (2015). doi: 10.1083/jcb.201409127; pmid: 26056139
18. A. Boni *et al.*, Live imaging and modeling of inner nuclear membrane targeting reveals its molecular requirements in mammalian cells. *J. Cell Biol.* **209**, 705–720 (2015). doi: 10.1083/jcb.201409133; pmid: 26056140
19. H. Chug, S. Trakhanov, B. B. Hülsmann, T. Pleiner, D. Görlich, Crystal structure of the metazoan Nup62•Nup58•Nup54 nucleoporin complex. *Science* **350**, 106–110 (2015). doi: 10.1126/science.aac7420; pmid: 26292704
20. T. Stuwe *et al.*, Architecture of the fungal nuclear pore inner ring complex. *Science* **350**, 56–64 (2015). doi: 10.1126/science.aac9176; pmid: 26316600
21. V. Shahin, T. Danker, K. Enss, R. Ossig, H. Oberleithner, Evidence for Ca^{2+}- and ATP-sensitive peripheral channels in nuclear pore complexes. *FASEB J.* **15**, 1895–1901 (2001). doi: 10.1096/fj.00-0838com; pmid: 11532969
22. D. Stoffler, K. N. Goldie, B. Feja, U. Aebi, Calcium-mediated structural changes of native nuclear pore complexes monitored by time-lapse atomic force microscopy. *J. Mol. Biol.* **287**, 741–752 (1999). doi: 10.1006/jmbi.1999.2637; pmid: 10191142
23. A. Rakowska, T. Danker, S. W. Schneider, H. Oberleithner, ATP-Induced shape change of nuclear pores visualized with the atomic force microscope. *J. Membr. Biol.* **163**, 129–136 (1998). doi: 10.1007/s002329900377; pmid: 9592077
24. L. Kastrup, H. Oberleithner, Y. Ludwig, C. Schafer, V. Shahin, Nuclear envelope barrier leak induced by dexamethasone. *J. Cell. Physiol.* **206**, 428–434 (2006). doi: 10.1002/jcp.20479; pmid: 16110478
25. I. Liashkovich, A. Meyring, A. Kramer, V. Shahin, Exceptional structural and mechanical flexibility of the nuclear pore complex. *J. Cell. Physiol.* **226**, 675–682 (2011). doi: 10.1002/jcp.22382; pmid: 20717933
26. R. D. Jäggi *et al.*, Modulation of nuclear pore topology by transport modifiers. *Biophys. J.* **84**, 665–670 (2003). doi: 10.1016/S0006-3495(03)74886-3; pmid: 12524319
27. M. Eibauer *et al.*, Structure and gating of the nuclear pore complex. *Nat. Commun.* **6**, 7532 (2015). doi: 10.1038/ncomms8532; pmid: 26112706
28. Y. Zhang *et al.*, Molecular architecture of the luminal ring of the *Xenopus laevis* nuclear pore complex. *Cell Res.* **30**, 532–540 (2020). doi: 10.1038/s41422-020-0320-y; pmid: 32367042
29. G. Huang *et al.*, Structure of the cytoplasmic ring of the *Xenopus laevis* nuclear pore complex by cryo-electron microscopy single particle analysis. *Cell Res.* **30**, 520–531 (2020). doi: 10.1038/s41422-020-0319-4; pmid: 32376910
30. M. C. Field, M. P. Rout, Pore timing: The evolutionary origins of the nucleus and nuclear pore complex. *F1000Research* **8**, 369 (2019). doi: 10.12688/f1000research.16402.1; pmid: 31001417
31. S. Mosalaganti *et al.*, In situ architecture of the algal nuclear pore complex. *Nat. Commun.* **9**, 2361 (2018). doi: 10.1038/s41467-018-04739-y; pmid: 29915221
32. H. Asakawa *et al.*, Asymmetrical localization of Nup107-160 subcomplex components within the nuclear pore complex in fission yeast. *PLOS Genet.* **15**, e1008061 (2019). doi: 10.1371/journal.pgen.1008061; pmid: 31170156
33. J. Fernandez-Martinez *et al.*, Structure and Function of the Nuclear Pore Complex Cytoplasmic mRNA Export Platform. *Cell* **167**, 1215–1228.e25 (2016). doi: 10.1016/j.cell.2016.10.028; pmid: 27839866
34. P. Stelter *et al.*, Molecular basis for the functional interaction of dynein light chain with the nuclear-pore complex. *Nat. Cell Biol.* **9**, 788–796 (2007). doi: 10.1038/ncb1604; pmid: 17546040
35. X. Liu, J. M. Mitchell, R. W. Wozniak, G. Blobel, J. Fan, Structural evolution of the membrane-coating module of the nuclear pore complex. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16498–16503 (2012). doi: 10.1073/pnas.1214557109; pmid: 23019579
36. S. Bilokapic, T. U. Schwartz, Molecular basis for Nup37 and ELY5/ELYS recruitment to the nuclear pore complex. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15241–15246 (2012). doi: 10.1073/pnas.1205151109; pmid: 22955883
37. B. A. Rasala, A. V. Orjalo, Z. Shen, S. Briggs, D. J. Forbes, ELYS is a dual nucleoporin/kinetochore protein required for nuclear pore assembly and proper cell division. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17801–17806 (2006). doi: 10.1073/pnas.0608484103; pmid: 17098863
38. S. W. Baï *et al.*, The fission yeast Nup107-120 complex functionally interacts with the small GTPase Ran/Spi1 and is required for mRNA export, nuclear pore distribution, and proper cell division. *Mol. Cell. Biol.* **24**, 6379–6392 (2004). doi: 10.1128/MCB.24.14.6379-6392.2004; pmid: 15226438
39. K. Kelley, K. E. Knockenhauer, G. Kabachinski, T. U. Schwartz, Atomic structure of the Y complex of the nuclear pore. *Nat. Struct. Mol. Biol.* **22**, 425–431 (2015). doi: 10.1038/nsmb.2998; pmid: 25822992
40. S. J. Kim *et al.*, Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482 (2018). doi: 10.1038/nature26003; pmid: 29539637
41. T. Maimon, N. Elad, I. Dahan, O. Medalia, The human nuclear pore complex as revealed by cryo-electron tomography. *Structure* **20**, 998–1006 (2012). doi: 10.1016/j.str.2012.03.025; pmid: 22632834
42. M. C. Munder *et al.*, A pH-driven transition of the cytoplasm from a fluid- to a solid-like state promotes entry into dormancy. *eLife* **5**, e09347 (2016). doi: 10.7554/eLife.09347; pmid: 27003292
43. W. D. Richardson, A. D. Mills, S. M. Dilworth, R. A. Laskey, C. Dingwall, Nuclear protein migration involves two steps: Rapid binding at the nuclear envelope followed by slower translocation through nuclear pores. *Cell* **52**, 655–664 (1988). doi: 10.1016/0092-8674(88)90403-5; pmid: 3125984
44. N. Shulga *et al.*, In vivo nuclear transport kinetics in *Saccharomyces cerevisiae*: A role for heat shock protein 70 during targeting and translocation. *J. Cell Biol.* **135**, 329–339 (1996). doi: 10.1083/jcb.135.2.329; pmid: 8896592
45. W. A. Whalen, J. H. Yoon, R. Shen, R. Dhar, Regulation of mRNA export by nutritional status in fission yeast. *Genetics* **152**, 827–838 (1999). doi: 10.1093/genetics/152.3.827; pmid: 10388805
46. E. D. Schwoebel, T. H. Ho, M. S. Moore, The mechanism of inhibition of Ran-dependent nuclear transport by cellular ATP depletion. *J. Cell Biol.* **157**, 963–974 (2002). doi: 10.1083/jcb.200111077; pmid: 12058015
47. R. W. Wozniak, G. Blobel, M. P. Rout, POM152 is an integral protein of the pore membrane domain of the yeast nuclear envelope. *J. Cell Biol.* **125**, 31–42 (1994). doi: 10.1083/jcb.125.1.31; pmid: 8138573
48. P. Upla *et al.*, Molecular Architecture of the Major Membrane Ring Component of the Nuclear Pore Complex. *Structure* **25**, 434–445 (2017). doi: 10.1016/j.str.2017.01.006; pmid: 28162953

49. S. Abuhattum *et al*., Intracellular Mass Density Increase Is Accompanying but Not Sufficient for Stiffening and Growth Arrest of Yeast Cells. *Front. Phys.* **6**, 131 (2018). doi: 10.3389/fphy.2018.00131
50. A. Miermont *et al*., Severe osmotic compression triggers a slowdown of intracellular signaling, which can be explained by molecular crowding. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5725–5730 (2013). doi: 10.1073/pnas.1215367110; pmid: 23493557
51. R. P. Joyner *et al*., A glucose-starvation response regulates the diffusion of macromolecules. *eLife* **5**, e09376 (2016). doi: 10.7554/eLife.09376; pmid: 27003290
52. G. Marini, E. Nüske, W. Leng, S. Alberti, G. Pigino, Reorganization of budding yeast cytoplasm upon energy depletion. *Mol. Biol. Cell* **31**, 1232–1245 (2020). doi: 10.1091/mbc.E20-02-0125; pmid: 32293990
53. M. B. Heimlicher *et al*., Reversible solidification of fission yeast cytoplasm after prolonged nutrient starvation. *J. Cell Sci.* **132**, jcs.231688 (2019). doi: 10.1242/jcs.231688; pmid: 31558680
54. A. J. Lomakin *et al*., The nucleus acts as a ruler tailoring cell responses to spatial constraints. *Science* **370**, eaba2894 (2020). doi: 10.1126/science.aba2894; pmid: 33060332
55. V. Venturini *et al*., The nucleus measures shape changes for cellular proprioception to control dynamic cell behavior. *Science* **370**, eaba2644 (2020). doi: 10.1126/science.aba2644; pmid: 33060331
56. A. Agrawal, T. P. Lele, Mechanics of nuclear membranes. *J. Cell Sci.* **132**, jcs229245 (2019). doi: 10.1242/jcs.229245; pmid: 31308244
57. P. T. Tran, L. Marsh, V. Doye, S. Inoué, F. Chang, A mechanism for nuclear positioning in fission yeast based on microtubule pushing. *J. Cell Biol.* **153**, 397–412 (2001). doi: 10.1083/jcb.153.2.397; pmid: 11309419
58. S. M. Schreiner, P. K. Koo, Y. Zhao, S. G. J. Mochrie, M. C. King, The tethering of chromatin to the nuclear envelope supports nuclear mechanics. *Nat. Commun.* **6**, 7159 (2015). doi: 10.1038/ncomms8159; pmid: 26074052
59. A. Agrawal, T. P. Lele, Geometry of the nuclear envelope determines its flexural stiffness. *Mol. Biol. Cell* **31**, 1815–1821 (2020). doi: 10.1091/mbc.E20-02-0163; pmid: 32583742
60. A. Elosegui-Artola *et al*., Force Triggers YAP Nuclear Entry by Regulating Transport across Nuclear Pores. *Cell* **171**, 1397–1410.e14 (2017). doi: 10.1016/j.cell.2017.10.008; pmid: 29107331
61. B. Enyedi, P. Niethammer, Nuclear membrane stretch and its role in mechanotransduction. *Nucleus* **8**, 156–161 (2017). doi: 10.1080/19491034.2016.1263411; pmid: 28112995
62. N. Kellner *et al*., Developing genetic tools to exploit *Chaetomium thermophilum* for biochemical analyses of eukaryotic macromolecular assemblies. *Sci. Rep.* **6**, 20937 (2016). doi: 10.1038/srep20937; pmid: 26864114
63. H. Amelina *et al*., Sequential and counter-selectable cassettes for fission yeast. *BMC Biotechnol.* **16**, 76 (2016). doi: 10.1186/s12896-016-0307-4
64. J. Bähler *et al*., Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14**, 943–951 (1998). doi: 10.1002/(SICI)1097-0061(199807)14:10<943::AID-YEA292>3.0.CO;2-Y; pmid: 9717240
65. G. Dey *et al*., Closed mitosis requires local disassembly of the nuclear envelope. *Nature* **585**, 119–123 (2020). doi: 10.1038/s41586-020-2648-3
66. A. Vještica *et al*., A toolbox of stable integration vectors in the fission yeast *Schizosaccharomyces pombe*. *J. Cell Sci.* **133**, jcs240754 (2020). doi: 10.1242/jcs.240754; pmid: 31801797
67. J. Schindelin *et al*., Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012). doi: 10.1038/nmeth.2019; pmid: 22743772
68. J. F. Williams, S. G. J. Mochrie, M. C. King, A versatile image analysis platform for three-dimensional nuclear reconstruction. *Methods* **157**, 15–27 (2019). doi: 10.1016/j.ymeth.2018.10.009; pmid: 30359725
69. A. Halavatyi, S. Terjung, in *Standard and Super-Resolution Bioimaging Data Analysis: A Primer*, A. Wheeler, R. Henriques, Eds. (Wiley, 2017), pp. 99–141.
70. W. J. H. Hagen, W. Wan, J. A. G. Briggs, Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 (2017). doi: 10.1016/j.jsb.2016.06.007; pmid: 27313000
71. R. Danev, B. Buijsse, M. Khoshouei, J. M. Plitzko, W. Baumeister, Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15635–15640 (2014). doi: 10.1073/pnas.1418377111; pmid: 25331897
72. D. N. Mastronarde, S. R. Held, Automated tilt series alignment and tomographic reconstruction in IMOD. *J. Struct. Biol.* **197**, 102–113 (2017). doi: 10.1016/j.jsb.2016.07.011; pmid: 27444392
73. D. N. Mastronarde, Dual-axis tomography: An approach with alignment methods that preserve resolution. *J. Struct. Biol.* **120**, 343–352 (1997). doi: 10.1006/jsbi.1997.3919; pmid: 9441937
74. B. Turoňová, F. K. M. Schur, W. Wan, J. A. G. Briggs, Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *J. Struct. Biol.* **199**, 187–195 (2017). doi: 10.1016/j.jsb.2017.07.007; pmid: 28743638
75. M. Chen *et al*., Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *Nat. Methods* **14**, 983–985 (2017). doi: 10.1038/nmeth.4405; pmid: 28846087
76. A. Martinez-Sanchez, I. Garcia, S. Asano, V. Lucic, J. J. Fernandez, Robust membrane detection based on tensor voting for electron tomography. *J. Struct. Biol.* **186**, 49–61 (2014). doi: 10.1016/j.jsb.2014.02.015; pmid: 24625523
77. B. Turoňová *et al*., In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **370**, 203–208 (2020). doi: 10.1126/science.abd5223
78. G. Tang *et al*., EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007). doi: 10.1016/j.jsb.2006.05.009; pmid: 16859925
79. B. Turoňová *et al*., Benchmarking tomographic acquisition schemes for high-resolution structural biology. *Nat. Commun.* **11**, 876 (2020). doi: 10.1038/s41467-020-14535-2; pmid: 32054835
80. S. H. W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012). doi: 10.1016/j.jsb.2012.09.006; pmid: 23000701
81. E. F. Pettersen *et al*., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
82. L. Zimmermann *et al*., A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018). doi: 10.1016/j.jmb.2017.12.007; pmid: 29258817
83. N. Guex, M. C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997). doi: 10.1002/elps.1150181505; pmid: 9504803
84. A. Šali, T. L. Blundell, Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993). doi: 10.1006/jmbi.1993.1626; pmid: 8254673
85. T. Stuwe *et al*., Nuclear pores. Architecture of the nuclear pore complex coat. *Science* **347**, 1148–1152 (2015). doi: 10.1126/science.aaa4136; pmid: 25745173
86. D. H. Lin *et al*., Architecture of the symmetric core of the nuclear pore. *Science* **352**, aaf1015 (2016). doi: 10.1126/science.aaf1015; pmid: 27081075
87. C. S. Weirich, J. P. Erzberger, J. M. Berger, K. Weis, The N-terminal domain of Nup159 forms a β-propeller that functions in mRNA export by tethering the helicase Dbp5 to the nuclear pore. *Mol. Cell* **16**, 749–760 (2004). doi: 10.1016/j.molcel.2004.10.032; pmid: 15574330
88. K. Yoshida, H. S. Seo, E. W. Debler, G. Blobel, A. Hoelz, Structural and functional analysis of an essential nucleoporin heterotrimer on the cytoplasmic face of the nuclear pore complex. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16571–16576 (2011). doi: 10.1073/pnas.1112846108; pmid: 21930948
89. K. Strimmer, fdrtool: A versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **24**, 1461–1462 (2008). doi: 10.1093/bioinformatics/btn209; pmid: 18441000
90. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995). doi: 10.1111/j.2517-6161.1995.tb02031.x
91. T. D. Goddard *et al*., UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018). doi: 10.1002/pro.3235; pmid: 28710774
92. V. Rantos, K. Karius, J. Kosinski, Integrative structural modelling of macromolecular complexes using Assembline. *bioRxiv* 2021.04.06.438590 [Preprint] (2021). doi: 10.1101/2021.04.06.438590
93. B. Webb *et al*., Integrative structure modeling with the Integrative Modeling Platform. *Protein Sci.* **27**, 245–258 (2018). doi: 10.1002/pro.3311; pmid: 28960548
94. D. Saltzberg, C. H. Greenberg, S. Viswanath, I. Chemmama, B. Webb, R. Pellarin, I. Echeverria, A. Sali, in *Biomolecular Simulations*, M. Bonomi, C. Camilloni, Eds., *Methods in Molecular Biology, Volume 2022* (Humana, 2019), pp. 353–377.
95. S. Viswanath, I. E. Chemmama, P. Cimermancic, A. Sali, Assessing Exhaustiveness of Stochastic Sampling for Integrative Modeling of Macromolecular Structures. *Biophys. J.* **113**, 2344–2353 (2017). doi: 10.1016/j.bpj.2017.10.005; pmid: 29211988
96. G. Drin *et al*., A general amphipathic α-helical motif for sensing membrane curvature. *Nat. Struct. Mol. Biol.* **14**, 138–146 (2007). doi: 10.1038/nsmb1194; pmid: 17220896
97. J. Kosinski *et al*., Data for integrative modeling of the Nuclear Pore Complex from *Schizosaccharomyces pombe*, *Zenodo* (2021). doi: 10.5281/zenodo.5585949

ACKNOWLEDGMENTS

We thank B. Turoňová, N. Kellner, W. Hagen, F. Weiss, C. Tischer, J. Baumbach, S. Welsch, M. Linder, E. Hurt, E. Lemke, and all the members of the Mahamid, Kosinski, and Beck laboratories for advice and support. We thank E. Hurt, B. Baum, and T. Schwartz for providing yeast strains. We acknowledge support from the Electron Microscopy Core Facility, the Advanced Light Microscopy Facility, and IT services of EMBL Heidelberg and the Microscopy Core Facility of the Max Planck Institute of Biophysics, Frankfurt am Main. **Funding:** M.A. was funded by an EMBO long-term fellowship (ALTF-1389–2016). J.M. received funding from the European Research Council (ERC 3DCellPhase 760067). J.K. was supported by funding from the Federal Ministry of Education and Research of Germany (FKZ 031L0100). M.B. acknowledges funding by EMBL, the Max Planck Society, and the European Research Council (ComplexAssembly 724349). **Author contributions:** C.E.Z. conceived the project, designed and performed experiments, acquired all types of data, designed and established data analysis procedures, analyzed all types of data, and wrote the manuscript; M.A. conceived the project, designed and performed experiments, acquired data, analyzed data, and wrote the manuscript; V.R. analyzed data and wrote the manuscript; S.K.G. designed and performed experiments and acquired data; A.O.-K. analyzed data; I.Z. designed and performed experiments; A.H. designed and performed experiments; J.M. designed experiments and supervised the project; G.H. designed experiments, analyzed data, and derived the theoretical model based on membrane elastic theory; J.K. conceived the project, designed and established data analysis procedures, analyzed data, supervised the project, and wrote the manuscript; and M.B. conceived the project, designed experiments, supervised the project, and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Associated with the manuscript are accession numbers EMD-11373, EMD-11374, EMD-11375, and EMD-13081 (EM Data Bank, www.ebi.ac.uk/emdb/), as well as PDBDEV_00000094, PDBDEV_00000095, and PDBDEV_00000096 (PDB-Dev database, https://pdb-dev.wwpdb.org/). The code for the structural modeling and the input files are available on Zenodo (97).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abd9776
Supplementary Text
Figs. S1 to S18
Tables S1 to S4
References (*98*–*106*)
MDAR Reproducibility Checklist
Movies S1 to S7

View/request a protocol for this paper from *Bio-protocol*.

24 July 2020; resubmitted 14 October 2021
Accepted 27 October 2021
Published online 11 November 2021
10.1126/science.abd9776

Nuclear pores dilate and constrict in cellulo

Christian E. ZimmerliMatteo AllegrettiVasileios RantosSara K. GoetzAgnieszka Obarska-KosinskaIevgeniia ZagoriyAliaksandr HalavatyiGerhard HummerJulia MahamidJan KosinskiMartin Beck

Science, 374 (6573), eabd9776. • DOI: 10.1126/science.abd9776

Mechanosensitive nuclear pores

The nucleus of eukaryotic cells is enclosed by the nuclear envelope, a double membrane punctuated with nuclear pore complexes (NPCs). These giant channels in the nuclear envelope mediate nucleocytoplasmic exchange. Zimmerli *et al*. show that the mechanical status of the nuclear membranes controls their nuclear pore diameter. Pulling forces imposed through nuclear membranes lead to stretching of NPCs and dilation of their diameter, whereas relief of such forces causes NPC constriction. Thus, the control of nuclear size and shape is functionally linked with NPC conformation and nucleocytoplasmic transport activity. —SMH

5.3 PUBLICATION III: "Integrative structural modeling of macromolecular complexes using Assembline"

Most cellular processes are facilitated by proteinaceous assemblies that must adopt a proper structural architecture in order to accomplish such functions in cells. Therefore, it is crucial to study the structures of complexes using experimental and computational methods. To acquire a mechanistic understanding of their cellular functions, it is also important to study protein complexes under native or near-native conditions.

For the past decades, two experimental methods have "monopolized" protein structure determination, namely X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy[89,92]. Although these techniques can solve structures of large complexes, they are mostly applied in the elucidation of structures at atomic or near-atomic resolution of small protein systems, often just single proteins (e.g. DtpA transporter, see Synopsis section 5.4). Recent advances in the experimental setup, data acquisition and analysis in cryo-EM have placed this method as well amongst the most utilized techniques for structure determination, especially for studying large and heterogeneous protein assemblies[84]. Most structures of large protein assemblies solved by cryo-EM exhibit high-resolution features towards the protein's core, while more flexible peripheral domains are usually poorly imaged. Moreover, structures solved using in-cell cryo-ET are typically mid-to-low resolution, thus hindering the process of understanding in detail the function of these protein complexes in their native cellular context.

The low resolution of some cryo-EM maps motivated the development of various methods for integrative structural modeling. These methods allow modeling structures of macromolecular assemblies based on spatial restraints from diverse input data[90,126]. Different integrative modeling approaches have already been applied to many macromolecular assemblies[46,95,96,97].

A typical integrative modeling workflow, described in detail for example by[112], starts with the representation of the target system in a coarse-grained or atomistic manner (see Introduction section 3.2). The selection of finer or cruder system representations should be based on the level of uncertainty of the input information. Then, spatial restraints are derived from the input data, which are used as scoring terms in a scoring function for the optimization of models satisfying the data. Various methods for finding the minimum of the scoring function can be applied. Usually, the output is not a single model but rather an ensemble of models equally satisfying the input information. In any case, the precision and exhaustiveness of sampling need to be assessed[99] and contribute to decisions regarding further modeling iterations.

In this context, I contributed significantly to the development, documentation and application of a ready-to-use package for integrative structural modeling of complex protein assemblies. The package is an open-source Python pipeline called Assembline[120] and is freely available online (https://www.embl-hamburg.de/Assembline/). Assembline was based on older versions of the underlying modeling algorithm and approach that was developed by Jan Kosinski (EMBL Hamburg) and were already applied to diverse modeling targets[47,116,43].

In general, Assembline can be applied to systems for which EM data are not available, though the main input data are usually EM densities and atomic structural models. Despite the fact that there are also other packages that can perform integrative modeling based on mid-to-low resolution cryo-EM maps, they all come with their own disadvantages. Some of such drawbacks are the lack of the ability to define protein systems in a versatile manner (e.g. complexes with symmetries of higher order) or the lack of the option to define additional or non-standard spatial restraints. Moreover, many such tools require comprehensive computational knowledge.

By utilizing the in-house built graphical user interface (GUI) from Xlink Analyzer[127], Assembline is offering a straightforward way to configure input files for modeling. This step provides the definition of the composition of the given assembly, which can be further edited by the user (e.g. to define symmetrical subunit copies). Extensive online documentation is provided along with step-by-step tutorials, including my application of Assembline in the modeling of ScNPCs[95] (see Synopsis section 5.1). Assembline was also applied for the modeling of in situ SpNPCs[97] (see Synopsis section 5.2), proving its suitability for modeling very complex targets. Based on the provided manual and modeling tutorials, there are no coding or heavy modifications needed from the user. In this context, Assembline is enabling structural biologists with less experience in such computational methods to perform modeling of their complexes.

Assembline also offers the option to define multiple system representations, thus enabling multi-scale modeling. The multi-scale representation has been implemented using the programming libraries of IMP[111] and PMI[118]. Additionally, a variety of cryo-EM, cross-linking and protein interaction (and proximity localization) restraints can be used for modeling with Assembline. Our tool also allows multi-state modeling and fitting of flexible (or not solved experimentally) protein domains represented as spherical beads.

Two very time-consuming challenges encountered when modeling large protein complexes based on cryo-EM data are the efficient sampling of the conformational space and the cross-correlation calculation between the experimental and simulated from the atomic models densities. Assembline overcomes both challenges by pre-calculating libraries of alternative fits of each rigid body in the complex. These fits, here referred to as fit libraries, and their respective cross-correlation scores, are then used as possible positions during integrative modeling (Fig. 23). The fit libraries, which may include up to hundreds of thousands of non-redundant fitted conformations, can then be used as input in integrative modeling runs without having to compute cross-correlation scores all over. Additionally, as was illustrated in the case of integrative modeling of ScNPCs[95] and SpNPCs[97] with Assembline (see Synopsis sections 5.1, 5.2), this generation of fit libraries step is also a powerful tool for allocating previously unknown or unobserved densities.

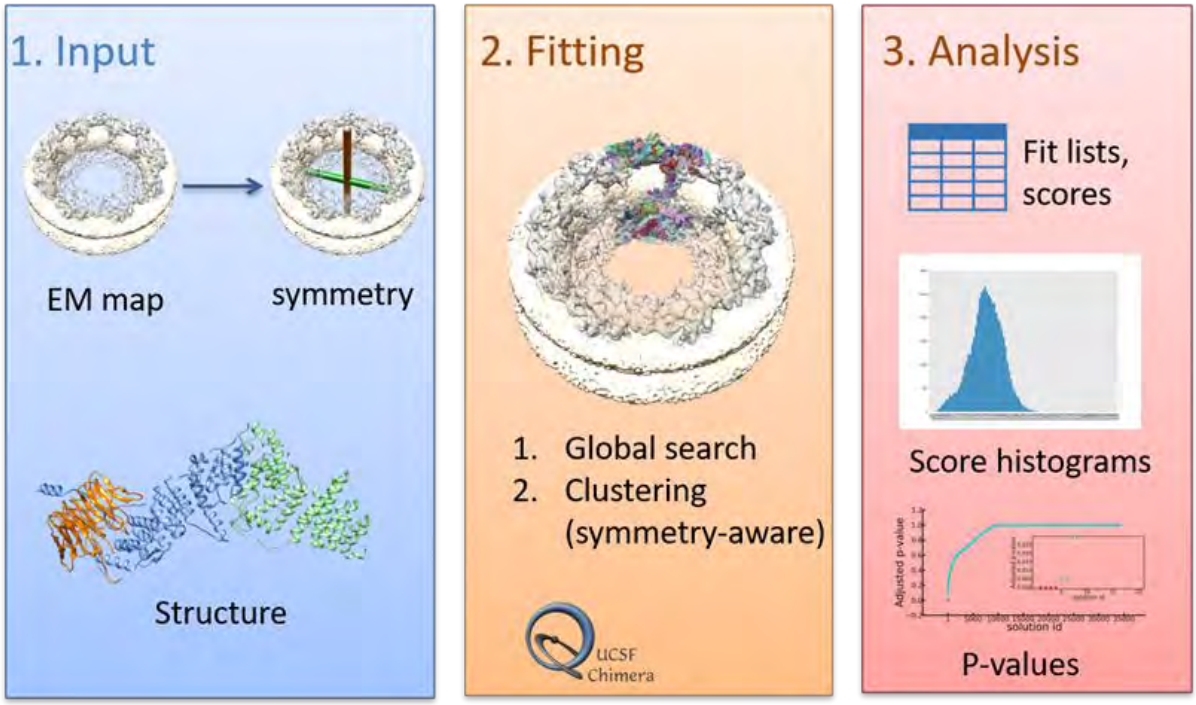


Figure 23: Schematic representation of the fit library generation step from Assembline. The pre-calculation of fit libraries in Assembline generates ensembles of alternative rigid body fits. The pipeline starts by simulating densities of atomic models and proceeds by random fitting attempts with the Global search algorithm from UCSF Chimera[107]. The fitting is performed multiple times (user-defined settings) and the resulting conformations are clustered. Plots of the distribution of scores and corresponding p-values are produced with optional reconstruction of best-fitted rigid bodies.

To build models of the entire target complex, Assembline uses a Monte Carlo Simulated Annealing (MCMC) algorithm for stochastically exploring the possible conformational space. The MCMC algorithm aims at minimizing the scoring function, which as mentioned earlier, includes scoring terms concerning spatial restraints that were derived from the input datasets. Therefore, having calculated the fit libraries a priori, the MCMC algorithm at the global optimization step in Assembline is left with the sole task of producing combinations of rigid body fits which are sampled from the fit libraries. The computed integrative models can then be used as input in the next steps provided by Assembline in order to enrich the conformational sampling (recombination step) and/or perform refinement (refinement step) of the complexes with the local refinement mode (Fig. 24). Noteworthy is also the fact that at each modeling step, the user can analyze the output model ensembles and assess the sampling convergence and exhaustiveness as described by[99].

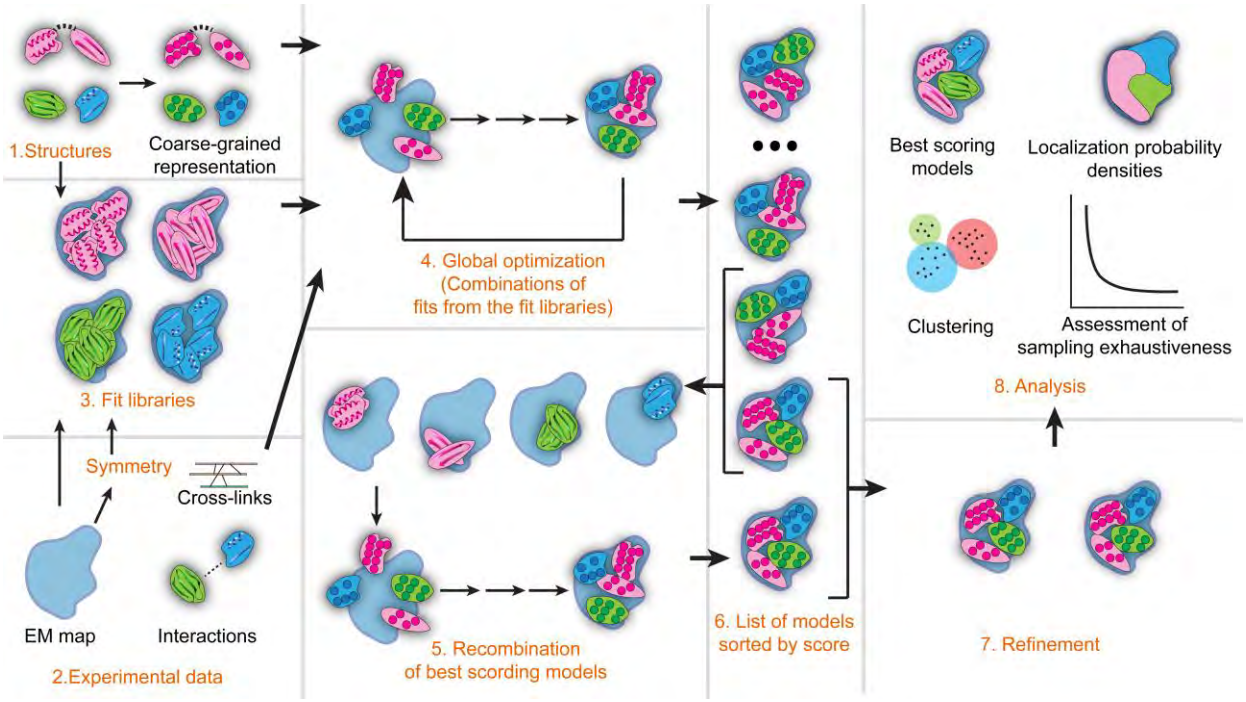


Figure 24: Depiction of the complete workflow from Assembline. The gathered atomic models that will be utilized as input along with the spatial restraints and the system representation are specified first for the configuration prior to modeling (steps 1 and 2). In case of the existence of cryo-EM maps, the systematic fitting step (calculated fit libraries in step 3) is performed prior to modeling with the step of global optimization (step 4) following. The top-scoring integrative models from step 4 can then be locally refined in the refinement step (step 7). Alternatively, before refinement, the models from the global optimization step can be additionally recombined in the recombination step (step 5). Finally, the resulting ensembles of integrative models can be assessed for their precision and quality during the analysis step (step 8). Figure adapted from[120].

In both ScNPC and SpNPC modeling cases that were described earlier (see Synopsis sections 5.1, 5.2), a very similar modeling approach with Assembline was followed (with case-tailored differences). NPCs were not the only modeling target though since an older version of Assembline was also applied for the modeling of yeast Elongator complex[116] and type VII secretion system[128], hence proving that Assembline could already be the "go-to" solution for such modeling targets. This claim is supported by the fact that Assembline offers great versatility due to IMP[111] that lies in its' "algorithmic heart" but also ease of use for structural biologists and other researchers with less computational knowledge. Finally, by circumventing some very time-consuming modeling steps, Assembline offers one of the most computationally efficient integrative modeling solutions to date, especially in cases of large assemblies.

nature protocols

PROTOCOL

https://doi.org/10.1038/s41596-021-00640-z

Check for updates

Integrative structural modeling of macromolecular complexes using Assembline

Vasileios Rantos[1,2], Kai Karius[1,2] and Jan Kosinski[1,2,3]✉

Integrative modeling enables structure determination of macromolecular complexes by combining data from multiple experimental sources such as X-ray crystallography, electron microscopy or cross-linking mass spectrometry. It is particularly useful for complexes not amenable to high-resolution electron microscopy—complexes that are flexible, heterogeneous or imaged in cells with cryo-electron tomography. We have recently developed an integrative modeling protocol that allowed us to model multi-megadalton complexes as large as the nuclear pore complex. Here, we describe the Assembline software package, which combines multiple programs and libraries with our own algorithms in a streamlined modeling pipeline. Assembline builds ensembles of models satisfying data from atomic structures or homology models, electron microscopy maps and other experimental data, and provides tools for their analysis. Compared with other methods, Assembline enables efficient sampling of conformational space through a multistep procedure, provides new modeling restraints and includes a unique configuration system for setting up the modeling project. Our protocol achieves exhaustive sampling in less than 100–1,000 CPU-hours even for complexes in the megadalton range. For larger complexes, resources available in institutional or public computer clusters are needed and sufficient to run the protocol. We also provide step-by-step instructions for preparing the input, running the core modeling steps and assessing modeling performance at any stage.

Introduction

Macromolecular complexes are crucial to many biological processes. The function of complexes depends on their three-dimensional (3D) structure—the relative arrangement of the subunits, which assemble to form structural scaffolds, active and ligand binding sites, and regulatory modules[1]. Determining the structure of complexes is therefore key to understanding how they assemble and function.

One of the most widely used methods for determining structures of large macromolecular complexes is cryo-electron microscopy (cryo-EM). It involves a 3D reconstruction of a so-called electron microscopy (EM) density map, which is used to build a structural model. The resolution of cryo-EM maps is sometimes sufficiently high for building atomic models using the information only from the maps. Frequently, however, the high resolution is limited to more rigid regions of a complex, while flexible peripheral domains, which are less well resolved, remain ambiguous[2,3]. Moreover, some complexes are difficult to resolve at high resolution because of technical challenges specific to the given sample or because of compositional heterogeneity of the sample[4].

Recent work has demonstrated that EM maps of individual complexes can be obtained in their native cellular environment by applying cryo-electron tomography (cryo-ET) to vitrified cells[5–9], followed by sub-tomogram averaging[10]. In-cell cryo-ET foreshadows a new era of native structural biology, but currently the cryo-ET maps rarely reach resolution beyond 1 nm[7,11]. Thus, methods for interpreting low-resolution EM maps will be crucial for realizing the potential of in-cell structural biology.

Integrative structural modeling is an approach that can be used to determine structures on the basis of low-resolution EM maps[12], or even for complexes where an EM map cannot be obtained[13–15]. It leverages information from other structural biology techniques such as X-ray crystallography, homology modeling, NMR, small-angle scattering and cross-linking mass spectrometry (XL-MS)[16,17]. Integrating data from different techniques allows models to be built at a higher precision than using a single technique alone[7,18,19]. Multiple structures of macromolecular complexes have been obtained this way, even under in-cell conditions[6,7,20].

[1]Centre for Structural Systems Biology (CSSB), Hamburg, Germany. [2]European Molecular Biology Laboratory Hamburg, Hamburg, Germany. [3]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. ✉e-mail: jan.kosinski@embl.de

In a typical integrative modeling workflow[18,21], structures of individual subunits or domains of the subunits are first collected using X-ray crystallography, NMR or homology modeling[22]. In some modeling software, the structures are converted to a coarse-grained representation. The following general steps are then performed:

- Experimental data are translated into spatial restraints, which are used to define a scoring function for subsequent optimization
- Various optimization methods are applied to sample alternative arrangements of the input structures and find the arrangement that minimizes the scoring function. The output is not a single model, but rather an ensemble of models equally satisfying the experimental restraints
- The exhaustiveness of the sampling and the precision of both the sampling and the model ensembles are assessed[19]; this is done to interpret the models and quantify their uncertainty (see 'Analysis')

We have developed a versatile integrative modeling protocol that can be applied to complexes as large as 50–100 MDa nuclear pore complexes (NPCs)[6,20,23], using high- and low-resolution EM maps derived from cryo- and negative-stain EM, and cryo-ET maps resolved in cells. The protocol is a computational pipeline integrating our Xlink Analyzer[24] graphical interface for input preparation, UCSF Chimera software[25] for EM fitting and analysis of results, and the programming libraries of Integrative Modeling Platform[21,26] (IMP) and Python Modeling Interface[27] (PMI; an interface to IMP by the same authors) for modeling.

The protocol implements a multistep scoring and sampling procedure that enables efficient exploration of conformational space (see 'Assembline workflow overview'). In addition to the programming interface of IMP and PMI, our approach enables the user to define the target system through the graphical user interface of Xlink Analyzer[24] and text-based configuration files, implement additional restraints and use command-line scripts that can be applied to modeling cases without any modification by the user.

The Procedure provides step-by-step guidelines for applying Assembline to any protein complex.

Development of the protocol

We developed the first version of our protocol to build a model of the scaffold of the human NPC based on cryo-ET and XL-MS data[23]. The first version included all the steps of the procedure described here, except the analysis of sampling exhaustiveness and precision. Subsequently, we used it to model the yeast Elongator complex based on negative-stain EM maps and XL-MS data[28], with the model being later validated by a high-resolution cryo-EM structure[29]. For the Elongator modeling project, we extended the protocol with flexible refinement of loops and implemented a number of improvements in code, speed and user interface. Recently, we have used an updated version of the protocol to build models of yeast NPCs[6,20], and to model peripheral subunits of the mycobacterial Type VII secretion system[30]. The updates included new restraints (such as a log-harmonic cross-link restraint, a restraint that enforces similarity between interfaces in different pairs of the same proteins, or a re-implemented symmetry restraint), implementations of the elastic network, further code and sampling speed improvements, and introduction of the sampling exhaustiveness and precision analysis step. In this work, we are describing the most recent version of the protocol as used for the *Saccharomyces cerevisiae* nuclear pore complex[6] (ScNPC). For this purpose, we organized the code and scripts into a software package called Assembline.

Applications

Assembline can model protein complexes based on EM, XL-MS and protein–protein interaction data, e.g., affinity pull-down experiments indicating interactions between proteins or protein domains. Assembline supports other experimental data types through restraints available in a programming interface from IMP and PMI. When only EM data are available, Assembline can be used as a fitting program for fitting multiple structures simultaneously. Although in our published applications we have primarily used EM data, Assembline can be also applied to cases where EM data are not available, for example to build approximate topological models based on XL-MS data only. Complexes from simple dimers to assemblies of multiple subunits and complex symmetries are also amenable. For highly symmetric complexes, Assembline can model complexes with hundreds of subunits, as demonstrated by our applications to NPCs or bacterial secretion systems.

Exemplary results from our previous applications are described in 'Anticipated results'.

Assembline workflow overview

The core algorithmic feature of our protocol is a unique multistep sampling procedure that efficiently explores structural configurations when an EM map is available (Fig. 1). Modeling protein complexes based on EM data, particularly at low resolution, is computationally demanding, because it is necessary to sample a large conformational space (i.e., within the EM map there are many possible locations and orientations for each of the multiple subunits) and many complex calculations are required to cross-correlate the EM map with the modeled structure.

Assembline overcomes both challenges through a global optimization step that first calculates ensembles of alternatively fitted conformations of individual subunit and domain structures (called 'fit libraries'). Then, it generates good-scoring combinations of those fits according to the scoring function defined by the user for integrative modeling. Since all the cross-correlation scores of fitted structures in the 'fit libraries' have been calculated a priori, there is no need to recalculate the cross-correlation at the stage of generating combinations, drastically speeding up calculations and enabling efficient sampling of the conformational space. Notably, as the original EM scores were calculated using the original atomic structures, the EM scores are derived from the atomic representation even if proteins are coarse-grained for the actual optimization.

The models from the global optimization can be then used as input in the next steps, to either enrich conformational sampling through a recombination between top-scoring models and/or refine the complexes using a conventional sampling, in which rigid bodies are moved in a continuous space through random rotations and translations and cross-correlation is calculated 'on the fly'.

Representation of proteins for modeling

During modeling, input protein structures can be represented in either atomic or, to accelerate computation, coarse-grained representation as beads at a desired 'resolution' (e.g., one bead represents ten amino acid residues). As a trade-off between computing efficiency and accuracy, multiple

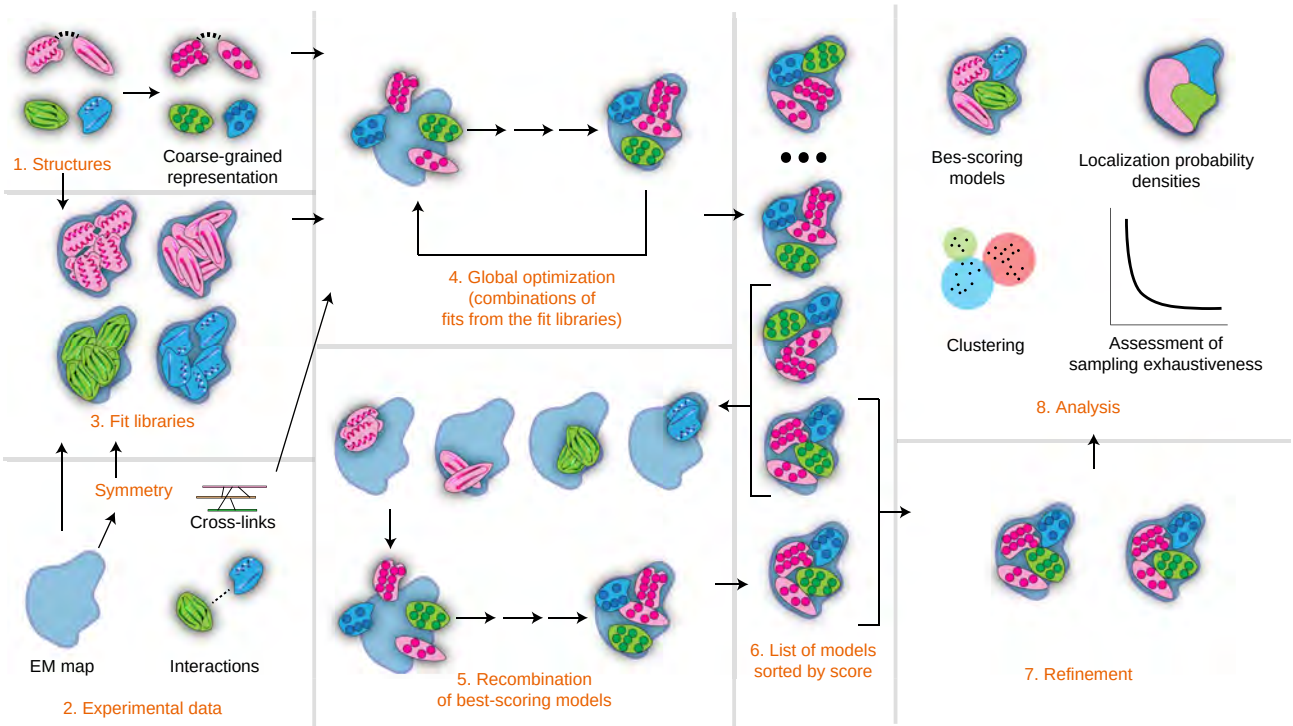


Fig. 1 | Assembline workflow. Available structures and experimental data are specified as input through configuration files. Optionally, input atomic structures are converted to coarse-grained beads (Steps 1 and 2). If EM data are used, libraries of fits for each structure can be calculated (Step 3, drawn as alternative fits in various locations in the EM map). The fits of all structures are sampled simultaneously by combining fits from the libraries and scored based on input restraints (global optimization, Step 4). The best-scoring models from the global optimization can be additionally recombined, generating potentially better scoring models (Step 5). The best-scoring models (Step 6) are selected for local rigid-body or flexible refinement (Step 7). During analysis, sampling convergence and exhaustiveness are assessed, and the output integrative models are clustered into groups of similar models and analyzed with respect to restraint satisfaction. Also, localization probability densities can be calculated for each subunit or rigid body (Step 8).

coarse-grained representations can be used in parallel according to the precision of restraints. For example, a ten-residue per bead representation can be used for the costly excluded volume restraint (steric clash score), while the C_α representation can be used for XL-MS data-based distance restraints. The multiscale representation is implemented using the PMI library[27].

By default, all input structures are treated as rigid bodies (i.e., groups of atoms or beads constrained not to change coordinates relatively to each other). However, users can optionally account for the conformational flexibility of the target subunits and implement it in two ways. First, flexible linkers can be defined explicitly as single-residue beads connected by distance restraints. Second, the input structures can be divided into smaller rigid bodies and restrained using an elastic network to prevent distortions of the input structure. The user can define both ways in configuration files by defining residue ranges that should be treated rigidly or flexibly. The elastic network is calculated automatically based on corresponding directive in the configuration as well.

The user can also include information about homo-oligomeric complexes and cyclic symmetry. Assembline automatically creates the copies of the input structures and imposes symmetry based on user-provided symmetry definitions, and resolves an ambiguity arising when restraints are applied to multiple copies of the same subunit. For example, for cross-link or interaction restraints, it is sufficient that only one of the copies satisfies the restraint.

Scoring function

The scoring function for modeling is a linear combination of restraint scores as implemented in IMP[26]. Several custom restraints have been implemented along with restraints directly sourced from IMP.

Custom restraints in Assembline include:

- EM restraints based on the *P* values of fits in the fit libraries (see 'Calculation of fit libraries')
- Symmetry restraints
- Excluded EM densities restraints, which can be used to penalize penetration of certain EM densities, e.g., a segmented membrane density
- EM density proximity restraints (e.g., to favor interactions of membrane-binding proteins with a membrane)
- Cross-linking restraints including a log-harmonic cross-link restraint[13]
- Elastic network restraints for preservation of protein complex interfaces, and binary interprotein or interdomain interactions.

Standard restraints sourced from IMP include connectivity distance between neighboring domains in sequence, a steric clash score, standard EM fit restraint based on cross-correlation, and EM envelope penetration restraints.

Calculation of fit libraries

The first step of our protocol is the generation of fit libraries, which are ensembles of nonredundant rigid body fits of the input atomic structures within the provided EM maps (before coarse-graining). The fit libraries are used as alternative positions of the rigid bodies during the subsequent global optimization step. Typically, hundreds to thousands of fits are generated per structure. If the initial model of the complex has already been constructed by other means (e.g., with a different integrative modeling protocol), or no EM map is used as a restraint, this step can be omitted and the model can be optimized immediately by the refinement step.

Fit libraries are generated using the Fit In Map tool from UCSF Chimera[25] in 'global search' mode. A command line interface is provided to fit multiple structures into multiple maps (for example, different versions of the EM map or maps of different states of the protein complex). The resulting fits are then clustered using the built-in clustering feature of UCSF Chimera from Fit in Map tool, wherein representative models from each cluster constitute the final ensemble of nonredundant fits. The user can modify several parameters, including the number of random starting positions for the global fitting, clustering threshold settings, resolution of simulated densities and three different cross-correlation score types provided by UCSF Chimera (for details, see the documentation of UCSF Chimera at https://www.cgl.ucsf.edu/chimera/current/docs/UsersGuide/midas/fitmap.html).

Finally, *P* values are calculated from the cross-correlation scores of alternative fits as published before[6,20,23,28,30,31]. The *P* values assess how likely a fit with the same score or better can be found by chance for the given pair of the EM map and the fitted structure. To calculate the *P* values, the cross-correlation scores are first transformed to *z*-scores (Fisher's *z*-transform[32]) and centered, from which two-sided *P* values are computed using standard deviation derived from an empirical null distribution

(derived by fitting the normal distribution to 95% of the fits using fdrtool R-package[33]). All *P* values are then corrected for multiple testing using Benjamini–Hochberg procedure[34]. The corrected *P* values are used as restraints during the global optimization step.

The option of reconstructing the best-scoring fitted models at atomic representation for visual inspection is also available. Even in cases where the best-scoring fits have been mislocalized and exhibited nonsignificant *P* values, the following modeling steps of Assembline can optimize their localization (Supplementary Fig. 1).

In addition, the user can use *P* values to identify unambiguous fits that can be kept fixed in their positions in the subsequent modeling steps to limit the conformational search space.

Global optimization

In the previous step, each rigid body structure was fitted independently to create libraries of alternative fits. The objective of the global optimization step is to generate models of the entire complex based on these fit libraries via simultaneous sampling of alternative fits using Monte Carlo simulated annealing optimization[35] (Fig. 1, Step 4). The optimization algorithm randomly draws upon the precalculated fits from the libraries to generate candidate combinations, scores the combinations according to their fit in the EM map and satisfaction of other restraints, and iteratively repeats the random fit selection and scoring to find better scoring solutions. The EM fit restraint is calculated based on *P* values precalculated during the previous stage, and as such is derived from the atomic representation. Usually, this step consists of hundreds to thousands of independent runs (i.e., optimization trajectories restarted from random initial orientation), each run including thousands of Monte Carlo sampling steps, depending on the size of the system. The scoring function, representation of structures and optimization algorithms are implemented using functionalities from the underlying IMP and PMI programming libraries.

The output of this step is an ensemble of alternative integrative models, each scored based on violation of the restraints comprising the scoring function. At this point, the user can assess the sampling performance and select final models (see 'Analysis') and/or continue to the next integrative modeling stages.

Recombination

The optional recombination stage (Fig. 1, Step 5) allows enriching the sampling of good-scoring models. This is particularly useful for systems with many rigid bodies or thousands of alternative fits in the fit libraries. In this step, the global optimization protocol is run again, but this time using only the precalculated alternative fits of the subunits that led to top models in the global optimization run. The precalculated fits that led to those models are retrieved from the original libraries to create smaller fit libraries, and used as input to the optimization algorithm identical to the global optimization. As the fits are coming from good-scoring models, this leads to preferential sampling of the conformational space in best-scoring areas and often yields models with scores substantially better than in the first global optimization stage.

Refinement

Refinement (Fig. 1, Step 7) optimizes integrative models from the global optimization (including the recombination) step using the underlying IMP and PMI programming libraries. The main difference between the global optimization and refinement is that in the global optimization the EM fit restraint is precalculated from *P* values of fits in the fit libraries, while in the refinement the raw cross-correlation coefficient is used and calculated on the fly. This difference is based on the fact that the component (i.e., rigid body) positions in the refinement stage are not drawn from the ensemble of precalculated fits but translated and rotated in small increments.

In practice, an ensemble or single best-scoring model(s) can be used as input. The refinement calculations can be computationally expensive. Therefore, it is recommended that the user provide input structures that are already approximately fit to the EM map (e.g., output structures from the global optimization). Refinement can also be used independently of any prior integrative modeling runs, e.g., to optimize models obtained using other modeling software.

The refinement, as in the case of global optimization, starts with representing the defined system at desired coarse-grained resolutions and proceeds with stochastic sampling of alternative rigid body conformations by Monte Carlo simulated annealing. The coordinates of flexible beads are optimized using the conjugate gradient algorithm. All output models are sorted by total score, and at this point, the user can reconstruct the top modeling solutions at atomic representation and continue with the analysis.

Analysis

Analysis of the output model ensemble (Fig. 1, Step 8) entails the assessment of the sampling convergence and exhaustiveness, estimation of the sampling and model precision, quantification of restraint violation and selection of representative models or model ensembles. For the assessment of exhaustiveness and precision, Assembline provides a command line interface to generate modeling output compatible with the imp-sampcon[19] toolkit from IMP.

The toolkit assesses the exhaustiveness based on four statistical criteria and estimates sampling precision as the highest precision at which the sampling can be considered exhaustive.

For this, the output ensemble of good-scoring models is split randomly into two samples of approximately equal size. The first two tests assess the convergence of scores and similarities between the distribution of scores in the two model samples. The remaining two tests, performed upon structural clustering of the models from the two samples (i.e., using a root-mean-square deviation between model coordinates as the similarity metric), evaluate the structural similarity of models in the samples by checking whether each cluster includes models from each sample proportionally to its size and if the modeled structures from the two samples are similar in each cluster. The tests are performed automatically by imp-sampcon[19]. This analysis outputs several files describing the contents and including information on whether the sampling converged and at which precision the sampling was exhaustive, as well as precision values for the entire sampling and for each structural cluster. In addition, imp-sampcon provides graphical plots summarizing the results of sampling precision assessment[19]. Finally, for each cluster, imp-sampcon calculates so-called localization probability densities, which are 3D maps expressing the probability of finding each protein at each point in space. The details of the tests are described in the imp-sampcon publication[19], and example plots are shown in 'Anticipated results'.

Comparison with other methods

Other integrative modeling software have been published and applied to a variety of systems. IMP[26] and PMI[27], which we use in our protocol, can on their own be used for integrative modeling. They offer the possibility to construct custom-made integrative modeling protocols based on multiple types of restraints, and defining the initial system architecture and its representation generically. However, a user of PMI or IMP would typically require Python programing expertise to modify and adapt the tutorial scripts to their specific modeling case. In contrast, our protocol brings an advantage of a versatile input configuration system consisting of a graphical interface and text configuration files that can be applied to a wider range of complexes without extra programming. This system enables configuring the input for even the most intricate complexes, for which structures of subunits or domains may be scattered across dozens of Protein Data Bank (PDB) files (parts of the complex might be subject to different restraints), or for which multiple symmetries might be present (as with the NPCs and bacterial secretion systems). Our provided scripts can be applied to new systems without any modification.

Despite the high-level interface, Assembline nevertheless offers the full functionality of IMP and PMI through a Python interface of Assembline. On top of IMP and PMI, our protocol implements its own algorithms for efficient sampling of the conformational space and provides additional modeling restraints. In particular, our fit library approach, which enables fast exploration of the conformational space, and the multistep optimization algorithm are features not available in IMP or PMI. Thus, Assembline benefits from all features of IMP and PMI and extends them with novel and easily accessible functionalities.

Other tools, such as FoXS[36], EMageFIT (from IMP) and MultiFit[37], have been implemented using IMP as an underlying programming library. These tools, however, focus on specific applications and do not offer the full flexibility that Assembline or IMP and PMI provide. Our optimization approach is conceptually similar to MultiFit[37], which also first performs discrete optimization of possible configurations followed by refinement. Some limitations of MultiFit[37] that motivated us to develop a new protocol is the reliance of MultiFit[37] on a priori density segmentation that defines subareas of the map for fitting, but which might exclude good fits before the optimization, less extensibility compared with IMP/PMI and Assembline, and no systematic input management system as included in Assembline. Thus, while MultiFit[37] and Assembline could lead to similar results in some cases, Assembline offers more flexibility, and it is more universal, especially for large complexes.

HADDOCK[38] and M3 (ref. [39]) (which uses HADDOCK internally) offer integrative modeling algorithms specializing in higher-resolution modeling with a physical force-field applied to atomistic representations or a coarse-grained MARTINI[40] representation. HADDOCK can be run either

through a web server or command line interface. Advantages of HADDOCK and M3 include full support of flexibility and the physical scoring function that can complement experimental restraints. Nevertheless, HADDOCK and M3, because of the high-resolution molecular representation, still cannot be applied to complexes as large as the NPC, and do not provide a procedure for efficient sampling of the EM map at a large scale. Thus, HADDOCK and M3 could be used to refine smaller complexes modeled with Assembline or to refine selected interfaces in models of large complexes.

Another versatile package for integrative modeling is PyRy3D (http://genesilico.pl/pyry3d), which also offers accessible user interface and a variety of restraints. In comparison, Assembline enables more efficient sampling algorithms and, by integrating with IMP and PMI, provides more restraints and tools for analysis. In principle, ROSETTA software[41] can also be used for integrative modeling and offers a vast spectrum of modeling algorithms and full support of structural flexibility. However, its versatility for integrative modeling applications depends on the development of additional customized protocols and thus a considerable level of modeling and programming expertise.

When only an EM map is available, Assembline can be used as a tool for simultaneous fitting of multiple components into EM maps. There are several other tools capable of producing models through single or multiple fitting to medium- or low-resolution EM maps by flexible or rigid body fitting, for example: UCSF Chimera[25], Situs[42], Flex-EM[43], MDFF[44], γ-TEMPy[45], CAMPARI[46], iMODFIT[47], MDFIT[48], FOLD-EM[49] or ATTRACT-EM[50]. For many EM-only applications and small complexes, these tools are sufficient and can be used instead of Assembline. Nevertheless, Assembline provides an efficient sampling protocol for large complexes, the advantage of a complete and versatile package that integrates all modeling steps, the possibility of restraints other than, and in addition to, EM, and an integrated pipeline for analysis of the final models.

In summary, Assembline brings broad usability and unique algorithmic advantages compared with other solutions and seamlessly integrates with complementary modeling programs.

Limitations

Out of the box, our protocol is designed for medium-to-low-resolution modeling, approximately worse than 4 Å resolution. Although it can be applied to fit subunit structures to high-resolution EM maps, other methods are more suitable for atomic-level de novo model building or flexible refinement of the resulting models. Currently, it does not support nucleic acids and offers only partial support for flexible deformation of structures during modeling through modeling of loops as flexible beads and elastic-network restraints between rigid bodies. We plan to address these two limitations in future versions of the protocol. As with most optimization algorithms, the protocol also relies on multiple parameters that have to be adjusted by the user, which we facilitate through appropriate guidelines and examples (Supplementary Manual and Supplementary Tutorials 1 and 2).

Expertise needed to implement the protocol

Users of Assembline should be familiar with the basic principles of structural modeling and structural analysis. Familiarity with Unix command line and executing command line programs is necessary, but the provided tutorials (Supplementary Tutorials 1 and 2) assume only basic knowledge, and guide users through all steps with explanations. No programming skills are needed, but users should familiarize themselves with the syntax of the configuration files and graphical molecular visualization software, which is needed to set up the input and analyze the output. Finally, the ability to access and use a computer cluster is beneficial, especially for larger systems.

Materials

Equipment

Hardware

- Personal computer or computer cluster with minimum 50 GB of free disk space and 4 GB of random-access memory
- Internet connection to access the online versions of Assembline usage manual and tutorial material (optional), i.e., for ScNPC and Elongator complex modeling ▲ **CRITICAL** For faster calculations, we advise running Assembline on a workstation with as many processors as possible or on a computer cluster.

Input datasets

- Sequences of protein subunits to be modeled in FASTA format
- Atomic protein structures (in PDB format) that will be used to model the protein complex

- cryo-EM densities of the modeling target complex in MRC (Medical Research Council) format and/or XL-MS data in Xlink Analyzer[24] format ▲ **CRITICAL** Example datasets can be found at https://git.embl.de/rantos/scnpc_tutorial.git (ScNPC modeling material) and https://git.embl.de/kosinski/elongator_tutorial.git (Elongator complex modeling material). Other structural data types can be used through the standard IMP and PMI Python programming interface within a `custom_restraints()` function of the parameter file.

Example input files

- Input sequence FASTA file, PDB structures and cryo-EM maps used for integrative modeling of the cytoplasmic ring (CR) Y-complex from ScNPC[6] (https://git.embl.de/rantos/scnpc_tutorial.git)
- Input sequence FASTA file, PDB structures, XL-MS datasets and negative-stain EM maps used for integrative modeling of the Elongator complex[28] (https://git.embl.de/kosinski/elongator_tutorial.git)

Software prerequisites

▲ **CRITICAL** Detailed instructions are provided for installation of all prerequisite software in the Assembline manual (Supplementary Manual and in the online version: https://assembline.readthedocs.io/en/latest/installation.html#).

- Unix-based operating system (e.g., Linux, Ubuntu, CentOS)
- UCSF Chimera[25] (free molecular visualization and analysis software)
- Xlink Analyzer[24] plugin for UCSF Chimera[25] (graphical interface for Assembline)
- Anaconda software (open-source distribution for Python and R for scientific programming, documentation: https://docs.anaconda.com/)
- Python programming language (version 3) distribution from Anaconda
- Assembline package (Anaconda-bundled package from `kosinskilab` channel in Anaconda)
- IMP[26] (version 2.14 or newer, package from `salilab` channel in Anaconda)
- Scipy[51], numpy[52], scikit-learn[53], matplotlib[54] and pandas (https://doi.org/10.5281/zenodo.3715232) Python packages (packages from generic channel in Anaconda)
- pyRMSD[55] package (package from `salilab` channel in Anaconda)
- hdbscan[56] package (package from `conda-forge` channel in Anaconda)
- R programming language distribution from Anaconda
- fdrtool[33], psych (https://cran.r-project.org/web/packages/psych/index.html), ggplot2 (https://cran.r-project.org/web/packages/ggplot2/index.html), tidyr[57], data.table packages (packages from generic R channel in Anaconda)
- (Optional) Modeller[58] package (comparative modeling software free for academic users)
- gnuplot package (http://www.gnuplot.info/) ▲ **CRITICAL** The software prerequisites along with the Assembline package that we describe in this work have been thoroughly tested in Unix-based systems with `bash` terminal; therefore, the protocol might not work as expected in OS X systems or other shell environments.

Procedure

Setup prior to modeling ● Timing ~2 h

▲ **CRITICAL** To proceed with the following steps, make sure you have first installed all prerequisites listed in 'Software prerequisites'.

1 Activate the Assembline modeling virtual environment, where all software prerequisites have been installed, with the following command:

```
source activate Assembline
```

or, depending on your Anaconda setup:

```
conda activate Assembline
```

2 Create a new directory for the modeling project, which will be referred to as the project directory, with the following example command for Unix-based systems:

```
mkdir Elongator
```

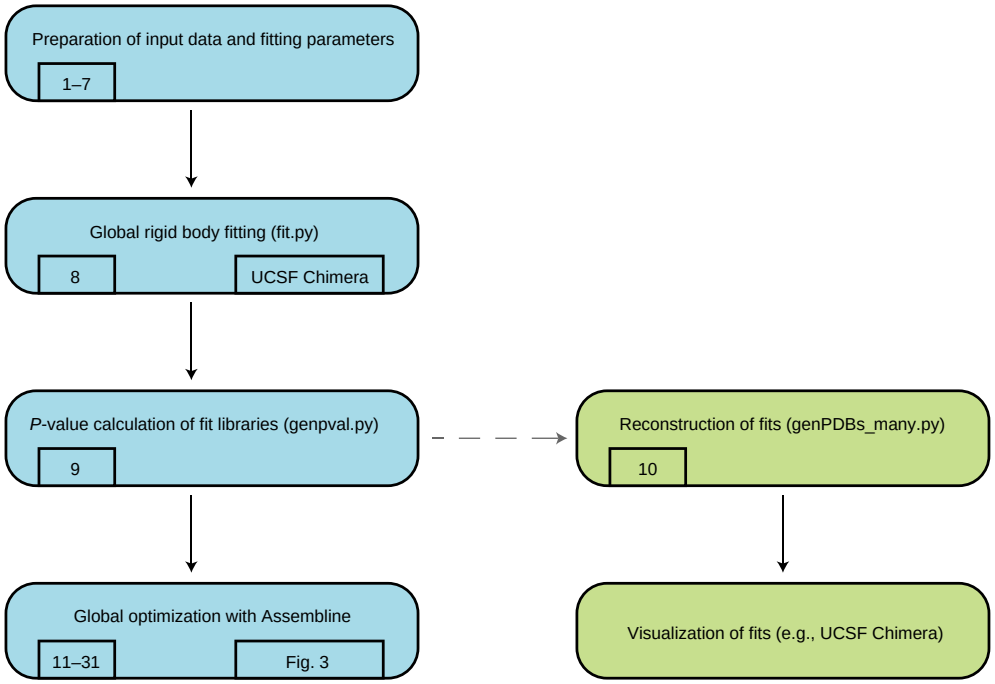


Fig. 2 | Detailed workflow regarding the calculation of fit libraries. The required steps are shown in blue and optional actions are shown in green. The external software and Assembline scripts used at each step, as well as the numbers of the corresponding steps from the Procedure, are indicated.

▲ **CRITICAL STEP** Input and configuration files for Assembline can be located in arbitrary locations on the disk, although it is recommended that they be stored in a single directory for easier management of the modeling project.

3 Create a single file (in FASTA format) with all sequences of the target subunits (i.e., subunit is a protein of a target complex) and store it in the project directory. Example FASTA file with sequences for the Elongator complex is provided at https://git.embl.de/kosinski/elongator_tutorial.git (Elongator complex modeling material).

4 In the project directory, create a new subdirectory and store all the structures of the target complex subunits in PDB format. Note that the subunit chains can be organized in the PDB files in any way, e.g., a PDB file can contain single or multiple subunits, or extra proteins not used in modeling, etc. ▲ **CRITICAL STEP** Make sure that the protein sequence and residue numbering in the PDB files correspond to the sequences in the FASTA file. Additionally, it is recommended that the PDB files be prepared such that each PDB file corresponds to an anticipated rigid body (i.e., there is one-to-one mapping between the PDB files and the anticipated rigid bodies) that will be used to model the target complex. Regions that are to be treated as flexible in modeling can be included in the PDB files. If they are not included but requested by the user, they will be added automatically as chains of flexible beads.

5 In the project directory, create a new subdirectory and store all the EM maps in MRC format. These maps will be used for the definition of the EM restraints (read more about the EM restraints in the Supplementary Manual).

6 Optionally, in the project directory, create new directories and store other available datasets that will be used as input for integrative modeling. For example, create a new directory called `xlinks` and store XL-MS datasets in Xlink Analyzer format.

Calculation of fit libraries ● Timing ~5 h

▲ **CRITICAL** The following steps for the calculation of fit libraries (see also Fig. 2) can be tested by retrieving and utilizing the fitting data and fitting parameters provided in https://git.embl.de/rantos/scnpc_tutorial.git (ScNPC modeling material) and https://git.embl.de/kosinski/elongator_tutorial.git (Elongator complex modeling material). Furthermore, the fit libraries will be used as 'discrete restraints' (read more in the Supplementary Manual) during the global optimization runs with Assembline.

7 In the project directory, create and store a parameter file that includes the paths to input PDB structures and EM maps to be used for fitting, the fitting parameters, and options for execution of fitting on a computer cluster (recommended). An example parameter file is provided in https://git.embl.de/kosinski/elongator_tutorial.git (Elongator complex modeling material), and detailed explanations are provided in the Supplementary Manual.

8 Run the fitting of the specified input PDB structures in the experimental maps with the following command:

```
fit.py efitter_params.py
```

? TROUBLESHOOTING

9 Upon completion of the fitting, start the analysis of the fit libraries by calculating the *P* values of the individual fits with the following command:

```
genpval.py <fitting directory name>
```

? TROUBLESHOOTING

10 (Optional) Enter the output directory (which would have a name that includes the fitting parameters e.g., `results/search100000_metric_cam_inside0.3_radius500`) and generate the models (in PDB format) of best-scoring fits for visualization. Examples of best-scoring fitted models are depicted in Supplementary Fig. 1. For example, to generate top five fits from a fit library for each input structure use:

```
genPDBs_many.py -n5 top5 */*/solutions.csv
```

? TROUBLESHOOTING

Global optimization ● Timing ~1 h to days, depending on the protein complex size and computational resources

▲ CRITICAL STEP Note that the global optimization with Assembline should be run only upon successful generation of the fit libraries in the previous steps. An overview of the following steps is provided in Fig. 3.

11 Open the Xlink Analyzer graphical interface (Fig. 4); this can be accessed as a plugin in UCSF Chimera. Create a project for the target complex; this will be used as the modeling configuration file (in JSON (JavaScript Object Notation) format).

▲ CRITICAL STEP Read more on how to install Xlink Analyzer as a plugin for UCSF Chimera and how to create a project for the target complex from scratch in the tutorial of Xlink Analyzer described by Kosinski et al.[24] (available at https://www.embl-hamburg.de/XlinkAnalyzer/XlinkAnalyzer.html) and the Elongator modeling tutorial (https://elongator-tutorial.readthedocs.io/en/latest/json_setup.html).

12 Add all subunits using the Xlink Analyzer interface.

13 For each subunit, enter the name, the chain ID or comma-separated multiple IDs, define the color, and click the 'Add' button. Set up the chain IDs in the final models according to user's preference; they do not have to correspond to chain IDs in your input PDB files.

14 (Optional) Define domains within subunits, which are user-defined regions of subunits and can be used later for, e.g., defining restraints specific to the domains, by clicking the 'Domains' button and entering the names of the domains and corresponding residue ranges in the pop-up window.

15 Add sequences using the Setup panel in Xlink Analyzer and map the sequences to names of subunits using the 'Map' button.

16 (Optional) Add available XL-MS datasets and map the cross-linked protein names to names of the subunits using the 'Map' button.

17 Make a local copy of the Xlink Analyzer project file, e.g., copy the `XlinkAnalyzer_project.json` as `X_config.json`. The new `X_config.json` will be used as the configuration file for modeling, whereas the original file should be kept for analysis of models.

18 Open a text editor and edit `X_config.json` project file manually to add modeling directives, which include: definition of series (grouping of subunits and domains), symmetry information, definition of input PDB structures and rigid bodies, specification of fit libraries and definition of spatial restraints.

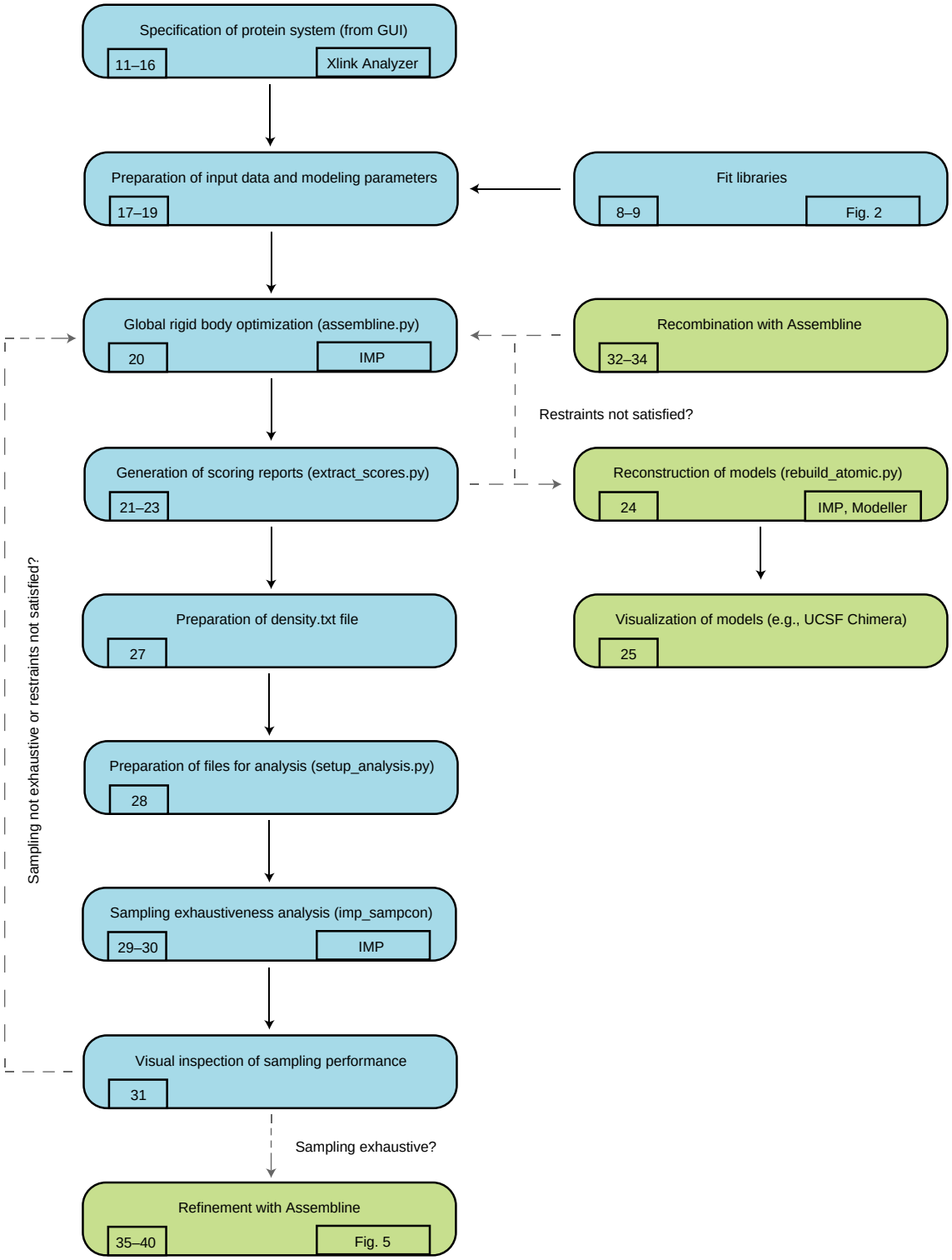


Fig. 3 | Workflow of the global optimization step. The required steps are shown in blue and optional actions are shown in green. The external software and Assembline scripts used at each step, the numbers of the corresponding steps from the Procedure, and the steps at which users have to make informed decisions whether to continue or repeat the modeling steps are also indicated.

▲ **CRITICAL STEP** Due to the different degrees of complexity of every modeling project and the plethora of available options and combinations in setting up the configuration file, a detailed description of how to set up this configuration file along with explanations regarding the definition steps are provided in the Supplementary Manual. We recommend text editors with syntax

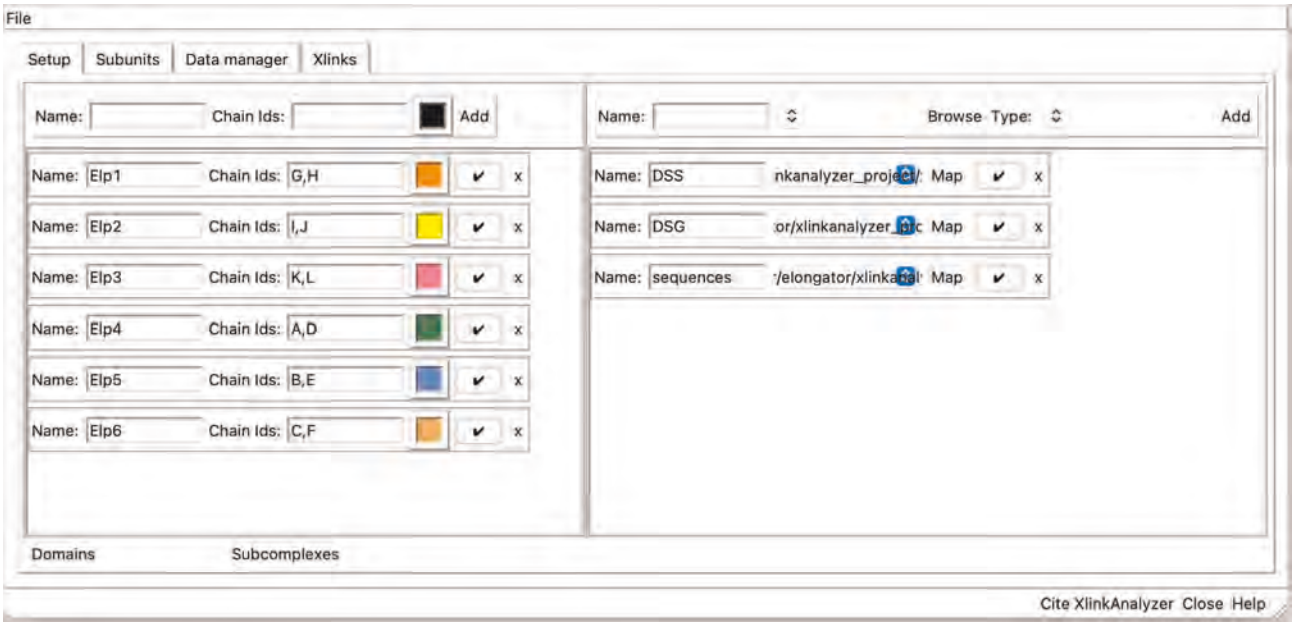


Fig. 4 | An example of Xlink Analyzer interface used for configuring the modeling project. Subunits can be defined in the panel on the left, while data are loaded and mapped to subunits in the panel on the right. Here, six protein subunits are defined (Elp1–6), and each subunit has been assigned a color. Each subunit is present in two copies in the complex; hence, two user-defined chain IDs are defined per subunit. On the right, full protein sequences and two cross-link datasets are defined, obtained with two different cross-linker types (disuccinimidyl suberate (DSS) and disuccinimidyl glutarate (DSG)). The 'Map' button opens a pop-up window that allows mapping of the protein names in the data files to the subunit names. More explanations and step-by-step instructions are included in the Elongator tutorial and the online XlinkAnalyzer documentation (https://www.embl-hamburg.de/XlinkAnalyzer/documentation.html).

highlighting to edit the configuration file (in JSON format), for example, SublimeText (https://www.sublimetext.com/) and Atom (https://atom.io/). Example configuration files can be found in https://git.embl.de/rantos/scnpc_tutorial.git (ScNPC modeling material) and https://git.embl.de/kosinski/elongator_tutorial.git (Elongator complex modeling material).

19 Create a parameter file in Python language format (e.g., `params.py`) that defines the modeling protocol, scoring functions, output parameters and some restraints for the global optimization stage, and save the file in the project directory.

▲ **CRITICAL STEP** Note that, although the file is in the Python language, no programming skills are required. If the modeling is going to be performed on a computer cluster, this parameter file should also include cluster submission settings. A detailed description of how to set up this file is provided in the Supplementary Manual. We recommend text editors with syntax highlighting to edit the configuration file (in JSON format), for example, SublimeText (https://www.sublimetext.com/) and Atom (https://atom.io/). Example configuration files can be found in https://git.embl.de/rantos/scnpc_tutorial.git (ScNPC modeling material) and https://git.embl.de/kosinski/elongator_tutorial.git (Elongator complex modeling material).

20 Using the Unix command line, navigate to the project directory and run the global optimization using the following example command, which will submit all runs to the computer cluster or to a standalone workstation in chunks of N models (i.e., parallel runs) according to the number of processors defined through `ntasks` parameter in the parameter file. The following command will submit 1,000 modeling jobs in the computer cluster queue or run them consecutively in chunks on a standalone workstation with each job leading to one model:

```
assembline.py --traj --models -o out --multi --start_idx 0 --njobs 1000
X_config.json params.py &>log&
```

▲ **CRITICAL STEP** Note that the number of processors to be allocated for the modeling runs or cluster submission commands and templates should be specified a priori in the parameters file. Read more on how to customize the submission of multiple modeling runs according to the computational system architecture in the Supplementary Manual. We also suggest performing a

single test run before deploying multiple jobs to a computer cluster or even a local workstation with a command similar to the following example:

```
assembline.py --traj --models --prefix 0000000 -o out config.json params.py
```

? TROUBLESHOOTING

21 Upon completion of the global optimization runs, enter the output directory, inspect the output log files to validate that the included information (e.g., the list of final parameter values used for modeling, the summary of the molecular system created, the scores of defined spatial restraints) matches the expected output.

▲ CRITICAL STEP If the output modeling logs do not include all the modeling parameters and restraints set prior to modeling, then repeat the previous step after correcting the parameter file accordingly. Read more on how to edit the parameter file and how to evaluate the output logs in the Supplementary Manual.

? TROUBLESHOOTING

22 While in the output directory, generate reports with total scores and individual scoring terms for all modeling runs with the following command:

```
extract_scores.py
```

▲ CRITICAL STEP The `extract_scores.py` script will generate several files with lists of models and their respective scores. Make sure that the `all_scores.csv` file and `all_scores_sorted_uniq.csv` files were generated as they will be used as input for the analysis in the following steps (both for the following optional steps and the steps of sampling exhaustiveness assessment).

? TROUBLESHOOTING

23 (Optional) Plot histograms of all total scores and scoring terms for all modeling runs from global optimization with the following command:

```
plot_scores.R all_scores.csv
```

24 (Optional) While in the output directory, generate the best-scoring models from the global optimization run in macromolecular crystallographic information file (mmCIF) format. For example, to generate the single best-scoring model:

```
rebuild_atomic.py --project_dir <full path to the original project
directory> --top 1 all_scores_sorted_uniq.csv
```

▲ CRITICAL STEP The `--project_dir` option is only necessary if relative paths were used in the JSON configuration file for global optimization. Also, note that by default the `rebuild_atomic.py` script will generate only the parts of the models that correspond to the specified input rigid bodies, meaning that the flexible bead parts of the system (if any) will not be rebuilt. Read more on how to generate flexible beads of subunits, or even how to generate best-scoring models in formats other than mmCIF (e.g., PDB format, although this format is not recommended for large systems) in the Supplementary Manual.

? TROUBLESHOOTING

25 (Optional) Visualize and inspect the modeling trajectory of the best-scoring models (or any other candidate model) by navigating and finding the corresponding trajectory file in `traj/` directory in the output folder from modeling and displaying it with UCSF Chimera. Supplementary Video 1 shows an example trajectory of the best-scoring model from CR Y-complex (from the in-cell ScNPC model[6]).

? TROUBLESHOOTING

26 (Optional) While in the modeling output directory, run a quick sampling convergence test and inspect the convergence plots, which will be stored in a PDF file called convergence.`pdf`. To do this, run the following command (example command to include 20 trajectories from the modeling runs):

```
plot_convergence.R total_score_logs.txt 20
```

▲ CRITICAL STEP If the plots clearly indicate no convergence, go back to Step 19, increase the number of Monte Carlo steps and repeat global optimization.

27 While in the output modeling folder, create a file containing all subunits (and the residue indexes). This file will be used to calculate localization probability densities (e.g., `density.txt`) during the sampling exhaustiveness assessment with `imp-sampcon exhaust` tool[19] from IMP. Create the file with the following command:

```
create_density_file.py --project_dir <path to the original project dir>
config.json --by_rigid_body
```

▲ **CRITICAL STEP** The output file (e.g., `density.txt`) from `create_density_file.py` script is required for the following steps of the sampling exhaustiveness analysis; therefore, the `density.txt` file has to be generated with respect to a very specific format. Read more on how to generate the `density.txt` file (as well as how to compile this file for complexes including symmetrical copies etc.) in the Supplementary Manual. Example `density.txt` files can be found in https://git.embl.de/rantos/scnpc_tutorial.git (ScNPC modeling material) and https://git.embl.de/kosinski/elongator_tutorial.git (Elongator complex modeling material).

If the modeled complex is homo-oligomeric, an extra file defining the symmetry is needed for the next step. This file can be generated using a command like the following example:

```
create_symm_groups_file.py --project_dir <full path to project dir>
config.json params.py
```

28 Run the following command (`setup_analysis.py` script), which will automatically prepare the input files for the sampling exhaustiveness analysis with `imp-sampcon` tool[19] from IMP based on the resulting integrative models:

```
setup_analysis.py -s <abs path to all_scores.csv file produced by
extract_all_scores.py> \
-o <specified output dir> \
-d <density.txt file generated in the previous step> \
-n <number of top scoring models to be analyzed, default is all models> \
-k <restraint score based on which to perform the analysis, default is
total score>
```

▲ **CRITICAL STEP** Read more about example commands and further options that can be used for the `setup_analysis.py` script in the Supplementary Manual.

? TROUBLESHOOTING

29 Go into the output analysis directory created by the previous step (i.e., output directory from `setup_analysis.py`) with the following command:

```
cd <analysis directory>
```

30 Run the sampling exhaustiveness analysis with the `imp-sampcon exhaust` tool[19] from IMP to assess the sampling performance of the modeling:

```
imp_sampcon exhaust -n <prefix for output files> \
--rmfA sample_A/sample_A_models.rmf3 \
--rmfB sample_B/sample_B_models.rmf3 \
--scoreA scoresA.txt --scoreB scoresB.txt \
-d <path to density.txt file>/density.txt \
-m <calculator selection> \
-c <number of processors to use> \
-gp \
-g <float with clustering threshold step> \
--ambiguity <symmetry group file if applicable)
```

▲ **CRITICAL STEP** We highly recommended running the sampling exhaustiveness analysis on multiple-processors workstation or a computer cluster as some of the testing steps during this analysis are computationally demanding. Examples for cluster submission options for `imp-sampcon exhaust` are provided in the Supplementary Manual and Supplementary Tutorials 1 and 2.

31 Inspect the four output plots and text files. If the sampling has not converged and sampling exhaustiveness has not been achieved, perform additional global optimization runs or adjust the modeling parameters, for example the number of simulated annealing steps. Sampling exhaustiveness results from ScNPC and Elongator complex modeling are provided in 'Anticipated results'.

Recombination (optional) ● Timing ~1 h to days, depending on the protein complex size and computational resources

▲ **CRITICAL STEP** Note that the recombination step should only be run if the global optimization step has been completed.

32 While in the output directory of the global optimization run, use the following command to automatically generate the JSON format configuration file for this step:

```
setup_recombination.py \
--json <JSON file used for global optimization> \
--scores all_scores_uniq.csv \
-o <output directory for the new fit libraries> \
--json_outfile <desired name for JSON config forrecombinations> \
--project_dir <original project dir> \
--score_thresh <score threshold for selecting the models> \
--top <number of top scoring models to use for extracting the fit
libraries>
```

33 Navigate back to the main project directory and run a command similar to the following example:

```
assembline.py --traj --models -o out --multi --start_idx 0 --njobs 1000
--prefix recomb config_recomb.json params.py &>log&
```

Using this exact command, Assembline will perform 1,000 modeling runs and store the output in the global optimization output folder. The command will perform the modeling by recombinations with Assembline using the freshly generated JSON formatted configuration file (e.g., `config_recomb.json`) and modeling parameter file generated during global optimization.
▲ **CRITICAL STEP** For recombination, the number of simulated annealing steps in `params.py` can usually be decreased for quicker calculations, because the used fit libraries are smaller. Read more on which settings are available and how to modify the parameter file generation procedure in the Supplementary Manual.
? TROUBLESHOOTING

34 Analyze the output from the modeling by rigid body recombinations with Assembline by following the exact analysis procedure described from Steps 21–31 (i.e., repeat Steps 21–31). It is expected that this step generates additional good-scoring solutions.
? TROUBLESHOOTING

Refinement ● Timing ~3 h to days, depending on the protein complex size and computational resources

▲ **CRITICAL** The refinement can be run without prior global optimization (or recombination) if the input rigid bodies are already approximately fitted in the EM map (e.g., if the model was already produced by a different integrative modeling protocol). In this case, all steps from the calculation of fit libraries, global optimization and recombination (i.e., Steps 7–34) can be omitted. An overview of the following steps of the refinement procedure is provided in Fig. 5.

If the results of the global optimization are satisfactory, you could choose to skip this refinement section. We would, however, recommend running the refinement to further optimize the models. The refinement mode would be also used as the first and the only modeling step if no EM map is available.

Read more regarding the refinement options in the Supplementary Manual or inspect examples of refinement applications in Supplementary Tutorials 1 and 2.

35 While in the main project directory, run a command similar to the following:

```
gen_refinement_template.py --out_json refine_template.json --params
params.py --add_series X_config.json
```

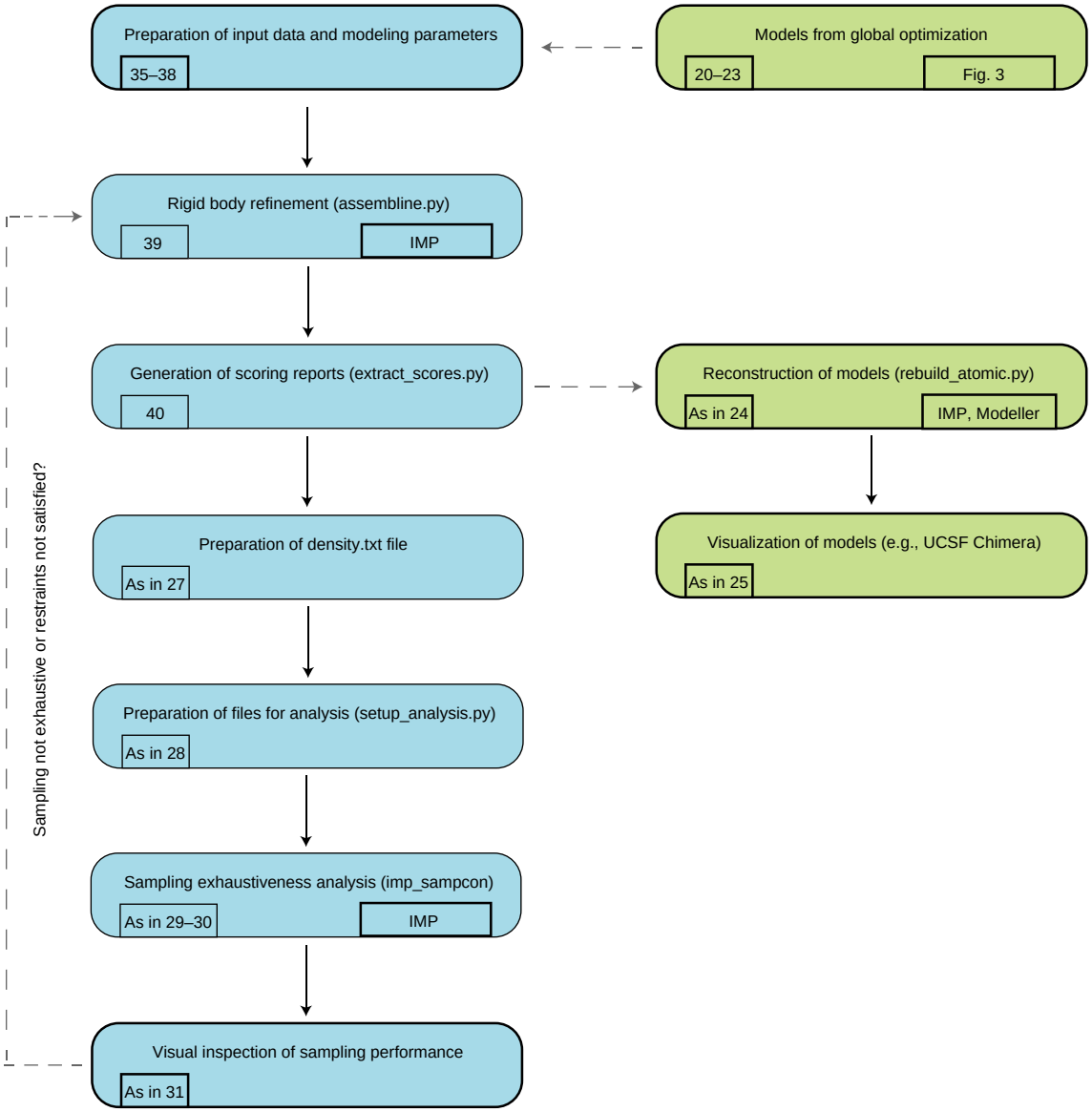


Fig. 5 | Workflow of the refinement step. The required steps are shown in blue and optional actions are shown in green. The external software and Assembline scripts used at each step, the numbers of the corresponding steps from the Procedure, and the steps at which users have to make informed decisions whether to continue or repeat the modeling steps are also indicated.

This example command will automatically generate a generic configuration file (in JSON format) based on the configuration and parameter files from global optimization:

36 Open a text editor and edit the generated (from previous step) JSON configuration file manually to add modeling directives that might include: updated definitions of rigid bodies, restraints specific for the refinement, adjusted restraint weights.

▲ **CRITICAL STEP** A detailed description of how to set up this configuration file along with explanations regarding the definition steps is provided in the Supplementary Manual. Example configuration files can be found at https://git.embl.de/rantos/scnpc_tutorial.git (ScNPC modeling material) and https://git.embl.de/kosinski/elongator_tutorial.git (Elongator complex modeling material).

37 In the main project directory, create and edit a copy of the global optimization parameter file in Python language format (if global optimization was not run beforehand, follow Steps 11–19 to generate it from scratch). The final format of the file should be almost identical with the global optimization parameter file, with the main differences being that the modeling protocol to be applied is `refinement` and the scoring function includes scoring terms derived from restraints specific to the refinement method of Assembline.

▲ **CRITICAL STEP** A detailed description of how to set up this file is provided in the Supplementary Manual.

38 Prepare the top best-scoring models from global optimization for refinement with a command similar to the following:

```
setup_refine.py \
--top 100 \
--scores out/all_scores_uniq.csv \
--previous_json config.json \
--refine_json_template config_refine_template.json \
--refine_json_outname config_refine.json \
--previous_outdir out/\
--refine_outdir out/refinement
```

This example command will create a directory containing 100 folders with input datasets and configuration files for the top 100 models from global optimization.

▲ **CRITICAL STEP** If the global optimization was not run beforehand and the refinement is applied to a pre-existing model, skip this step and specify in the configuration file the paths to the PDB files that will be used for refinement.

39 Run the refinement for all selected top-scoring models from global optimization (or single run in case global optimization was not run beforehand) with a series of commands similar to the following example (a `bash` loop that will apply refinement to the top 100 models from global optimization by running ten refinement runs for each model, i.e., 1,000 runs in total):

```
for model_id in 'ls --color=never out/refinement';
  do
    echo $model_id
    assembline.py \
    --traj --models \
    -o out/refinement/"$model_id"/out \
    --multi --start_idx 0 --njobs 10 \
    --prefix  refine_"$model_id"  \  out/refinement/"$model_id"/config_
    refine.json \ params_refine.py
  done
```

▲ **CRITICAL STEP** Note that the number of processors to be allocated for the modeling runs or cluster submission commands and templates should be specified a priori in the parameters file. Read more on how to customize the submission of multiple modeling runs according to the computational system architecture in the Supplementary Manual. Additionally, it is suggested to perform a single test run before running multiple jobs with commands similar to the following example:

```
model_id='ls --color=never out/refinement | head -n 1'
assembline.py \
--traj --models --prefix 0000000 \
-o out/refinement/"$model_id"/out \ out/refinement/"$model_id"/config_
refine.json \ params_refine.py \

  rm -r out/refinement/"$model_id"/out
```

? TROUBLESHOOTING

40 Analyze the output from the modeling by refinement by following the exact procedure described from Step 21 up to Step 31 (i.e., repeat Steps 21–31), but now generating the scores using the command:

```
extract_scores.py --multi
```

An example trajectory of the best-scoring model (produced with refinement) from nuclear ring (NR) Y-complex (from the in-cell ScNPC model[6]) can be inspected in Supplementary Video 2.

? TROUBLESHOOTING

Troubleshooting

Troubleshooting advice can be found in Table 1.

Table 1 | Troubleshooting table

Step	Problem	Possible reason	Solution
All	Assembline executables not found	Assembline `conda` package was not installed properly or virtual environment was not initiated before use	Make sure to activate the environment before use. Set up a new virtual environment with Anaconda and install Assembline and all remaining dependencies again
8	UCSF Chimera executable not found	UCSF Chimera was not installed properly	Make sure that the most updated version of UCSF Chimera is installed (download from https://www.cgl.ucsf.edu/chimera/download.html) and `chimera` command is available
	PDB structure files and/or EM maps not found	Directory paths to these files were wrongly specified in fitting parameters file or input files are absent	Make sure that these input files exist and specify in the parameters file the absolute paths to them or correct the relative paths used
9	No `solutions.csv` files produced (or only some)	Fitting runs crashed unexpectedly or no fitting positions respecting the fitting parameters could be found	Inspect the log files for error messages. Perform initial test run with less fitting iterations (computational optimization) and then rerun the fitting with `run_crashed_only` option set to `'True'` with less strict fitting parameters
	Very low number of fits found (i.e., rows in the `solutions.csv`)	Big structure fitted to a large map, low-resolution map, high clustering threshold	Use less strict clustering parameters
	Very high number of fits found, which would require computationally extensive global optimization	Small structure fitted to a big and/or low-resolution map	Use stricter clustering parameters or select only top fits for global optimization (using a provided option)
10	`genPDBs_many.py` crashes complaining that ori_pdb.pdb files do not exist or are invalid	`ori_pdb.pdb` symbolic link files that link to the original PDB files are absent or corrupted	Re-create the symbolic links by running `fit.py --update_links efitter_params.py`
20, 33, 39	PDB model files, and/or EM input maps, and/or sequences FASTA file, and/or `solutions_pvalues.csv` not found	Directory paths to these files are wrongly specified in configuration file or input files are absent	Make sure that these input files exist and specify in the configuration file the absolute paths to them or correct the relative paths used. The relative paths must be relative to the JSON file (configuration file) or the running directory
	Assembline crashes with `json.decoder.JSONDecodeError` (or similar JSON format related errors)	JSON syntax mistake in the `config.json` file	Consult the Assembline usage manual in the Supplementary Manual (online version: https://assembline.readthedocs.io/en/latest/json.html) on how to generate the `config.json` file properly, and test the compiled JSON file with any online JSON format validation tool
	Residue position mismatch between PDB and FASTA sequence reported in logs	Wrong residue numbering in the input PDB files or wrong sequence assigned to the subunit	Edit the FASTA sequence and/or relevant PDB structure so the residue numbering is identical
	Optimization takes a very long time to complete	The representation (bead) resolution is too high for the size of the system, or excessive number of Monte Carlo steps was specified	Decrease the representation resolution or Monte Carlo steps in the modeling parameters file
21, 34, 40	Defined spatial restraints not listed in the log files in the scores section	Spatial restraints were not defined properly, were not added to the scoring function, or were assigned to a wrong subunit	Consult the Assembline usage manual in the Supplementary Manual (online version: https://assembline.readthedocs.io/en/latest/json.html), redefine the problematic restraints and perform validation test run
	Some restraints scores dominating the scoring function (outweighing the rest of scores)	Scoring terms were not weighted based on similar scale	Perform test run, inspect the logs and rescale weights of restraints accordingly
22, 34, 40	Fewer output models ($0 \leq$ models $<$ requested) compared with requested and log files not including explicit errors	Some of the modeling runs crashed unexpectedly	Inspect the logs, determine the amount of crashed modeling jobs, and then rerun the modeling for the remaining jobs to total the initially requested models

Table continued

Table 1 (continued)

Step	Problem	Possible reason	Solution
24, 25, 34, 40	Spatial restraints not applied to all series of subunits	Misconfigured grouping of subunits in series	Consult the Assembline usage manual in the Supplementary Manual (online version: https://assembline.readthedocs.io/en/latest/json.html), redefine the problematic series and perform validation test run
	The models contain steric clashes	The excluded volume restraint was too weak or optimization has not converged	Evaluate the convergence and score files, and increase either the weight for excluded volume restraints or number of optimization steps
	Defined modeling target symmetry not respected	Wrongly defined symmetry axis	Consult the Assembline usage manual in the Supplementary Manual (online version: https://assembline.readthedocs.io/en/latest/json.html), redefine target's symmetry and perform validation test run
24, 34, 40	Specified project directory not found	Wrong path provided for the main project directory	Provide the full or correct relative path to the original project directory and run again the script `rebuild_atomic.py`
28, 34, 40	`all_scores.csv` file not found	File was not produced by `extract_scores.py` analysis script, wrong path to file provided, or no models in the output	Run `extract_scores.py` script before the sampling analysis, and define the correct path to `all_scores.csv`. Check if any models were produced in `models_txt` and `scores` directories and inspect logs
34	A rigid body does not obtain a defined position in the models and the localization probability density is very broad	Insufficient sampling, poor quality of the data, or the input structure of the rigid body or the EM map does not have defined density for this rigid body (e.g., due to flexibility)	Calculate more models or remove this rigid body from the system (assuming that the data do not allow for fitting this rigid body or the EM map does not have resolved density for this rigid body)

Timing

Steps 1–6, installation, setup of virtual environment for modeling and project directory architecture: ~2 h per modeling project, assuming the input datasets are already prepared

Steps 7–10, calculation of fit libraries: ~1–5 h, depending on number of input PDB structures to be fitted, number of EM maps to be used for fitting and allocated computational resources. The calculation of fit libraries for ScNPC takes ~3–5 h per input structure and for Elongator complex takes ~1–3 h per input structure

Steps 11–31, modeling with global optimization and analysis pipeline: ~1 h to days, depending on modeling parameters, system configuration settings, size of the complex, and allocated computational resources. A single global optimization run (using a single CPU core) for ScNPC takes ~1 h (~20,000 runs to be performed) and for Elongator complex 6 min (~1,000 runs to be performed)

Steps 32–34, modeling with the recombination step and analysis pipeline: ~1 h to days, depending on modeling parameters, system configuration settings, size of the complex, the number of selected models to be used for recombination, and allocated computational resources. A single recombination run (using a single CPU core) for Elongator complex takes ~4 min (~1,000 runs to be performed)

Steps 35–40, modeling with refinement and analysis pipeline: ~3 h to days, depending on modeling parameters, system configuration settings, size of the complex, the number of selected models to be used for refinement, and allocated computational resources. A single refinement run (using a single CPU core) for ScNPC takes ~3–4 h (~20,000 runs to be performed) and for Elongator complex takes ~20 min (~1,000 runs to be performed)

Anticipated results

The post-processed output of Assembline is an ensemble of models, along with an analysis of model uncertainty. The models can be exported in either coarse-grained or atomic resolution as PDB or CIF files.

Overview of integrative modeling output

The raw output of the integrative modeling stages in Assembline is models stored in bead representation (see Supplementary Videos 1 and 2 with example trajectories in bead representation) and simple text files that list the rigid body transformations and their respective scores. Assembline provides a structural analysis toolkit with which any number of best-scoring models can be converted to atomic representation and stored in the most common formats (e.g., PDB, CIF).

By default, when reconstructing models in atomic representation, only the rigid bodies will be included in the models. Flexible loops can be optionally rebuilt in a full atom representation using a provided option (which uses Modeller[58] to build loops based on the starting conformations of the loops derived from the bead representation).

As examples, in the following sections we describe how our protocol was applied in our recent study to model the NPC from *S. cerevisiae*[6] (Supplementary Tutorial 1). We also demonstrate the results of modeling of the Elongator complex from *S. cerevisiae*[28]. In the original publication, this complex was modeled using an older version of Assembline. For this protocol, we reanalyzed these datasets using the current version of Assembline using a simplified setup suitable for an introductory tutorial (Supplementary Tutorial 2).

Integrative modeling of ScNPC

NPCs are large macromolecular assemblies that fuse the nuclear envelope and facilitate nucleocytoplasmic transport[59]. They are constructed from ~30 different nucleoporins present in multiple copies, which are organized in a triple-stacked ring conformation that forms a central transport channel[23,31,59,60]. In a recent study[6], we built models of ScNPCs based on in-cell cryo-ET data collected from wild-type and knockout cells.

The model of the wild-type ScNPC (Fig. 6a) was constructed by integrating the data from cryo-ET maps of individual rings at ~25 Å resolution and biochemical data on protein–protein interactions and membrane-binding motifs[61]. Two of the three rings—the CR and NR—were modeled using the multistep procedure of global optimization followed by refinement as described above (Figs. 3 and 5). As an example of the analysis step, the results for the global optimization step of the CR Y-complex are depicted in Supplementary Fig. 2. The analysis showed that the sampling converged and led to a top-scoring model clearly standing out from the remaining models. The inner ring (IR) model was constructed by immediately applying the refinement step to a previously published model of in-vitro-purified ScNPC[62].

Two other cryo-ET datasets were collected under conditions where one of the subunits, Nup116, linking the outer and the inner rings, was genetically knocked out and cells were grown either at permissive (25 °C) temperature, at which the nuclear envelope and NPCs have normal morphology, or nonpermissive (37 °C) temperature, at which nuclear envelope and NPCs are distorted[6]. The two conditions led to EM maps with approximately 25 Å and 50 Å resolution, respectively (the lower resolution at the permissive temperature stemming from lower number of particle images available for cryo-ET reconstruction). Modeling was performed by first rigid-body fitting of the wild-type scaffold NPC rings to the knockout EM maps and then optimizing the models using the refinement step directly, based on the EM maps and the interaction restraints. The refinement of the ScNPC model at the nonpermissive temperature was challenging due to the low resolution of the EM map—the subunits were diverging from the initial structure, and many different models were equally satisfying the restraints. As a solution, we applied an elastic network restraint available in Assembline, which enables the preservation of the interfaces between rigid bodies in the starting structure as long as the other restraints are not in conflict. The resulting *nup116* knockout NPC model from cells grown under permissive temperature included the same scaffold complexes as the wild-type ScNPC and exhibited a very similar architecture, except for the missing density for Nup116 (Fig. 6b). In the case of the *nup116* knockout NPC model from cells grown under the nonpermissive temperature, only the outer nuclear copies of the IR unit and the NR could be confidently included in the final model. Therefore, we hypothesized that our model represented a failed 'inside-out' NPC assembly (Fig. 6c).

Detailed instructions for reproducing the wild-type and knockout ScNPC modeling are provided in Supplementary Tutorial 1 (and online at https://scnpc-tutorial.readthedocs.io/en/latest/).

Integrative modeling of the Elongator complex

Elongator is a complex involved in tRNA modification[63]. In yeast, it contains six subunits, two copies each. In 2017, we published the integrative model of the yeast Elongator based on negative-stain EM

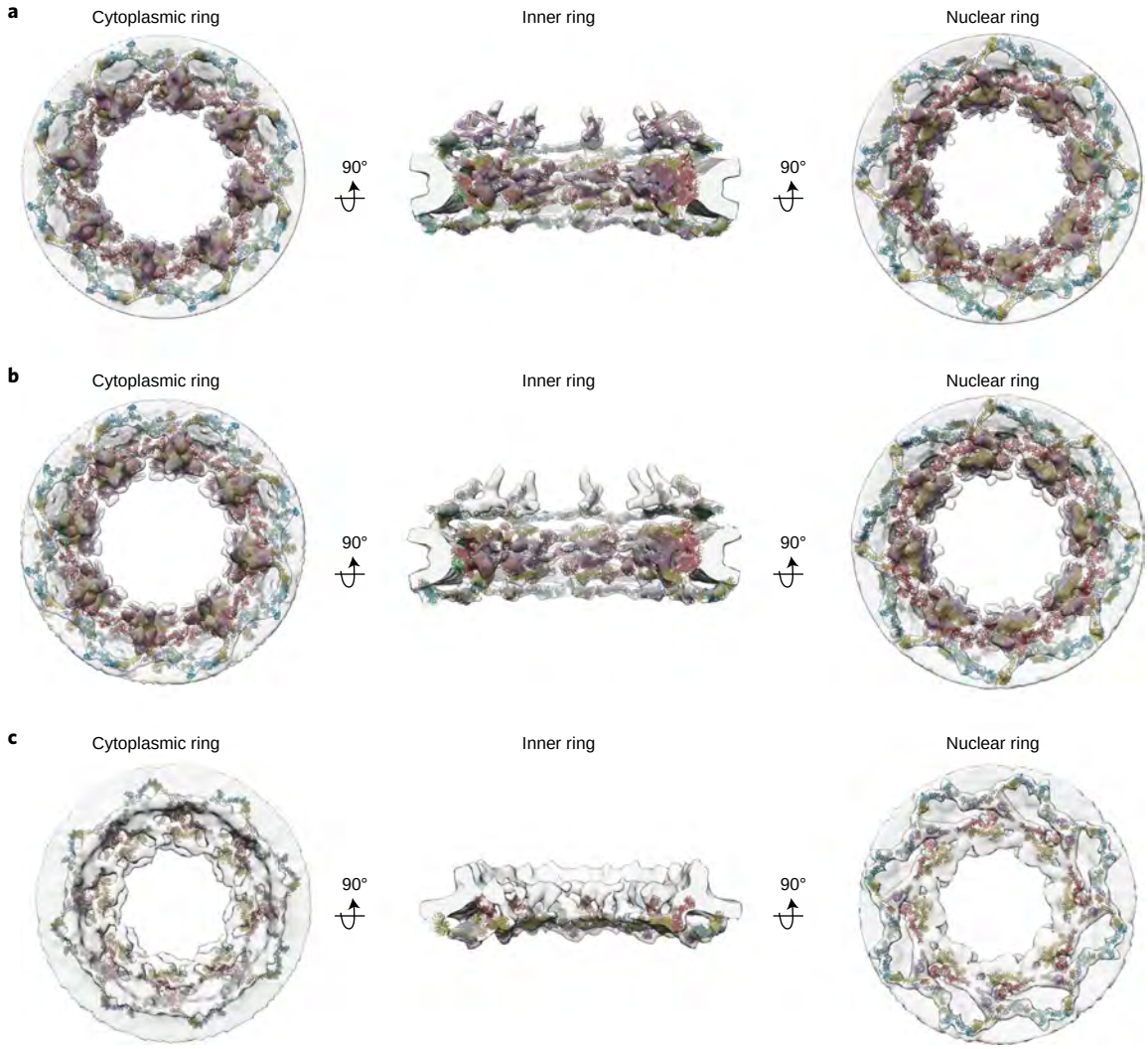


Fig. 6 | Integrative models of ScNPC. a,**b**,**c**, Overview of CR (left), IR (middle) and NR (right) of the wild-type ScNPC (**a**), the ScNPC from cells with *nup116* gene knockout (at 25 °C) (**b**) and the ScNPC from cells with *nup116* gene knockout (at 37 °C) where the CR is entirely missing (left) (**c**).

map at resolution of 27 Å and XL-MS data[28]. For the purpose of this protocol article, we repeated the modeling of the Elp123 subcomplex of Elongator by applying the updated procedure presented in this work.

The Elongator modeling case contained three subunits, each in two copies. The crystal structures and homology models of entire proteins or individual domains were used as input, grouped in nine rigid bodies (four rigid bodies for each of the two asymmetric units and one rigid body encompassing two copies of the same subunit) (Fig. 7a and Supplementary Tutorial 2). The twofold symmetry was applied as a constraint. From the calculation of fit libraries, between 200 and 500 fits were obtained for each rigid body. In the global optimization step, 1,000 models were generated. At this stage, the models already converged to a specific architecture (Fig. 7b) leading to nine clusters at a sampling precision of ~5 Å. The largest cluster contained almost identical models, while the remaining clusters exhibited cluster (or model) precision between 6 Å and 12 Å (Supplementary Fig. 3a-d). The top-scoring models represented very good fits to the EM map and only slightly violated the cross-link distance of 30 Å (Fig. 7b). Since the sampling was exhaustive and at high-sampling precision, the recombination stage did not need to be performed. All 1,000 models were refined, performing three refinement runs for each of the starting models, yielding 3,000 refined models. Two loop regions harboring cross-links were treated flexibly during refinement as in the C_α representation. The sampling precision obtained was ~6 Å, and ten clusters were obtained with precision of individual clusters between 8 Å and 16 Å (Supplementary Fig. 3e-h). The top-scoring model from refinement

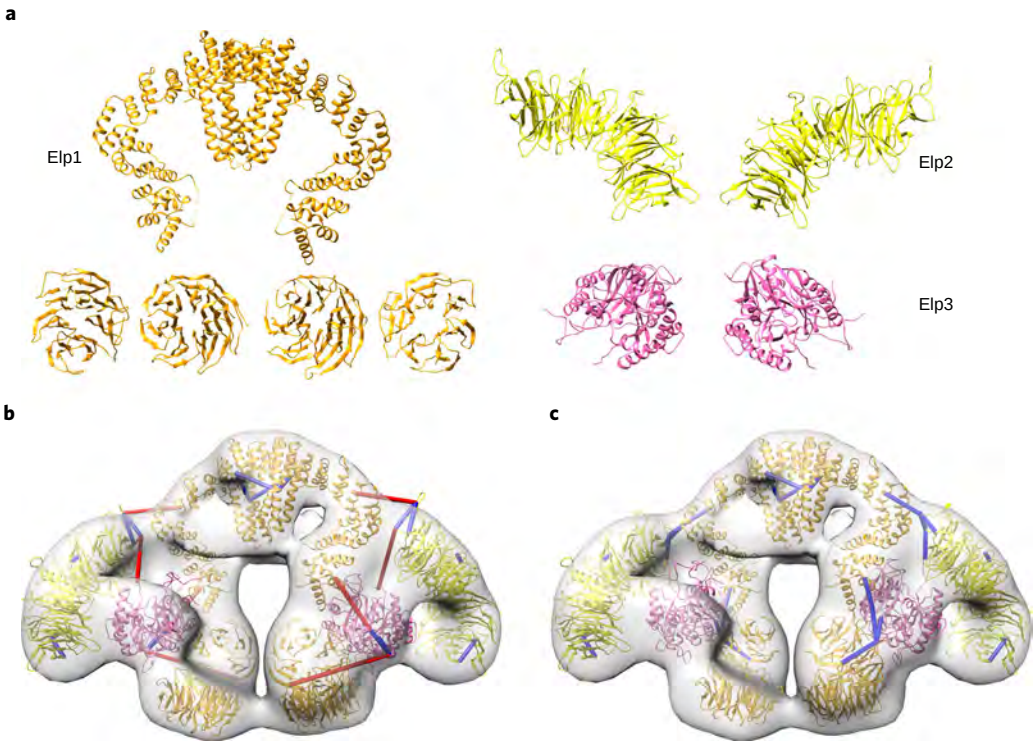


Fig. 7 | Integrative modeling of Elongator complex. a, Input structures (rigid bodies) used for modeling. **b**, The top-scoring model after global optimization within the EM map (gray transparent surface); cross-links indicated as blue and red bars (with red color indicating that the cross-links exceed the expected distance of 30 Å, and blue indicating that they satisfy this distance). **c**, The top-scoring model after the refinement satisfies all cross-links.

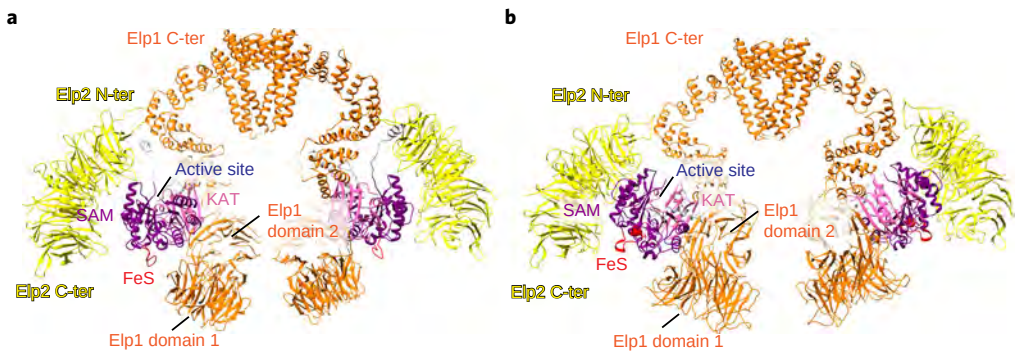


Fig. 8 | Validation of the integrative model of the Elongator complex by a high-resolution cryo-EM structure. a,**b**, Comparison of the integrative model (**a**) and the high-resolution cryo-EM structure (**b**) of the Elp123 subcomplex of Elongator. The subunits, domains and functional sites are indicated for comparison. C-ter, C terminus; FeS, iron-sulfur cluster binding site; KAT, lysine acetyl transferase binding site; N-ter, N terminus; SAM, S-adenosyl-L-methionine binding site.

belonged to the largest cluster and satisfied all cross-links (Fig. 7c, Supplementary Fig. 3i). Alternative models in the largest cluster were very similar to each other (Supplementary Fig. 3j).

One challenge that we encountered in this case, which is common for negative-stain EM maps, is that the calculated fitting libraries contained a very low number of alternative fits when generated with default options. This is likely due to high density in few regions of the map, which are not reflecting the internal structure of the complex, but are an artifact of the negative stain method. The fit libraries are generated with UCSF Chimera by placing input structures in random positions in the map and then optimizing their local fit according to the EM cross-correlation score. If high-density regions are present, the optimization would shift the fitted structures to those regions, as they would give higher scores, resulting in most of the fits falling in the same locations. Thus, in the subsequent

global optimization the models contained many clashes and did not lead to plausible models. To overcome this, we generated fit libraries with decreased clustering thresholds and fewer optimization steps, which led to a wider spectrum of fits and good-scoring models from the global optimization.

The 3.7 Å cryo-EM structure of Elongator[29], published in 2019, confirmed the model (Fig. 8). The entire architecture was predicted correctly, including not only the localization of the domains but also their orientations. Although some features could be predicted only approximately, such as the exact orientation of the Elp3 subunit and the N-terminal domains of Elp1, functional sites were correctly localized relative to other subunits. This demonstrates that integrative modeling can generate models that enable functional interpretation of the structure.

Details and instructions in order to reproduce the Elongator modeling are provided in Supplementary Tutorial 2 (and online at https://elongator-tutorial.readthedocs.io/en/latest/).

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All datasets needed for the step-by-step tutorials (i.e., for yeast NPC and Elongator complex modeling) are provided at https://git.embl.de/rantos/scnpc_tutorial.git (for the yeast NPC) and https://git.embl.de/kosinski/elongator_tutorial.git (for the Elongator complex).

Code availability

The Assembline software is freely available as an open-source Python package (https://www.embl-hamburg.de/Assembline/), which can be installed from source code (https://git.embl.de/kosinski/Assembline) or from the Anaconda repository (https://anaconda.org/kosinskilab/assembline). Documentation is provided in the Supplementary Manual and can also be found online (https://assembline.readthedocs.io/en/latest/), and in the form of the step-by-step tutorials for modeling the yeast NPC (Supplementary Tutorial 1, online version: https://scnpc-tutorial.readthedocs.io/en/latest/) and Elongator complex (Supplementary Tutorial 2, online version: https://elongator-tutorial.readthedocs.io/en/latest/).

References

1. Marsh, J. A. & Teichmann, S. A. Structure, dynamics, assembly, and evolution of protein complexes. *Ann. Rev. Biochem.* **84**, 551–575 (2015).
2. Patel, A. B. et al. Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* **362**, eaau8872 (2018).
3. Vos, S. M., Farnung, L., Urlaub, H. & Cramer, P. Structure of paused transcription complex Pol II–DSIF–NELF. *Nature* **560**, 601–606 (2018).
4. Lyumkis, D. Challenges and opportunities in cryo-EM single-particle analysis. *J. Biol. Chem.* **294**, 5181–5197 (2019).
5. Beck, M. & Baumeister, W. Cryo-electron tomography: can it reveal the molecular sociology of cells in atomic detail? *Trends Cell Biol.* **26**, 825–837 (2016).
6. Allegretti, M. et al. In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature* **586**, 796–800 (2020).
7. O'Reilly, F. J. et al. In-cell architecture of an actively transcribing-translating expressome. *Science* **369**, 554–557 (2020).
8. Turk, M. & Baumeister, W. The promise and the challenges of cryo-electron tomography. *FEBS Lett.* **594**, 3243–3261 (2020).
9. Wagner, F. R. et al. Preparing samples from whole cells using focused-ion-beam milling for cryo-electron tomography. *Nat. Protoc.* **15**, 2041–2070 (2020).
10. Bharat, T. A. M. & Scheres, S. H. W. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nat. Protoc.* **11**, 2054–2065 (2016).
11. Fäßler, F., Dimchev, G., Hodirnau, V. V., Wan, W. & Schur, F. K. M. Cryo-electron tomography structure of Arp2/3 complex in cells reveals new insights into the branch junction. *Nat. Commun.* **11**, 6437 (2020).
12. Malhotra, S., Träger, S., Dal Peraro, M. & Topf, M. Modelling structures in cryo-EM maps. *Curr. Opin. Struct. Biol.* **58**, 105–114 (2019).
13. Ferber, M. et al. Automated structure modeling of large protein assemblies using crosslinks as distance restraints. *Nat. Methods* **13**, 515–520 (2016).
14. Orbán-Németh, Z. et al. Structural prediction of protein models using distance restraints derived from cross-linking mass spectrometry data. *Nat. Protoc.* **13**, 478–494 (2018).

15. Gräwert, T. W. & Svergun, D. I. Structural modeling using solution small-angle X-ray scattering (SAXS). *J. Mol. Biol.* **432**, 3078–3092 (2020).
16. Koukos, P. I. & Bonvin, A. M. J. J. Integrative modelling of biomolecular complexes. *J. Mol. Biol.* **432**, 2861–2881 (2020).
17. Braitbard, M., Schneidman-Duhovny, D. & Kalisman, N. Integrative structure modeling: overview and assessment. *Annu. Rev. Biochem.* **88**, 113–135 (2019).
18. Rout, M. P. & Sali, A. Principles for integrative structural biology studies. *Cell* **177**, 1384–1403 (2019).
19. Viswanath, S., Chemmama, I. E., Cimermancic, P. & Sali, A. Assessing exhaustiveness of stochastic sampling for integrative modeling of macromolecular structures. *Biophys. J.* **113**, 2344–2353 (2017).
20. Zimmerli, C. E. et al. Nuclear pores constrict upon energy depletion. Preprint at *bioRxiv* https://doi.org/10.1101/2020.07.30.228585 (2020).
21. Webb, B. et al. Integrative structure modeling with the Integrative Modeling Platform. *Protein Sci.* **27**, 245–258 (2018).
22. Bordoli, L. et al. Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **4**, 1–13 (2009).
23. Kosinski, J. et al. Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363–365 (2016).
24. Kosinski, J. et al. Xlink analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.* **189**, 177–183 (2015).
25. Pettersen, E. F. et al. UCSF Chimera-A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
26. Russel, D. et al. Putting the pieces together: Integrative Modeling Platform software for structure determination of macromolecular assemblies. *PLoS Biol.* **10**, e1001244 (2012).
27. Saltzberg, D. et al. Modeling biological complexes using Integrative Modeling Platform. *Methods Mol. Biol.* **2022**, 353–377 (2019).
28. Dauden, M. I. et al. Architecture of the yeast Elongator complex. *EMBO Rep.* **18**, 264–279 (2017).
29. Dauden, M. I. et al. Molecular basis of tRNA recognition by the Elongator complex. *Sci. Adv.* **5**, eaaw2326 (2019).
30. Beckham, K. S. H. et al. Structure of the mycobacterial ESX-5 type VII secretion system pore complex. *Sci. Adv.* **7**, eabg9923 (2021).
31. Bui, K. H. et al. Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155**, 1233–1243 (2013).
32. Fisher, R. A. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* **10**, 507 (1915).
33. Strimmer, K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **24**, 1461–1462 (2008).
34. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
35. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science* **220**, 671–680 (1983).
36. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–W544 (2010).
37. Lasker, K., Topf, M., Sali, A. & Wolfson, H. J. Inferential optimization for simultaneous fitting of multiple components into a CryoEM map of their assembly. *J. Mol. Biol.* **388**, 180–194 (2009).
38. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).
39. Karaca, E., Rodrigues, J. P. G. L. M., Graziadei, A., Bonvin, A. M. J. J. & Carlomagno, T. M3: an integrative framework for structure determination of molecular machines. *Nat. Methods* **14**, 897–902 (2017).
40. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & De Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).
41. DiMaio, F., Tyka, M. D., Baker, M. L., Chiu, W. & Baker, D. Refinement of protein structures into low-resolution density maps using Rosetta. *J. Mol. Biol.* **392**, 181–190 (2009).
42. Wriggers, W. Conventions and workflows for using Situs. *Acta Crystallogr. D. Biol. Crystallogr.* **68**, 344–351 (2012).
43. Topf, M. et al. Protein structure fitting and refinement guided by cryo-EM density. *Structure* **16**, 295–307 (2008).
44. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683 (2008).
45. Pandurangan, A. P., Vasishtan, D., Alber, F. & Topf, M. γ-TEMPy: simultaneous fitting of components in 3D-EM maps of their assembly using a genetic algorithm. *Structure* **23**, 2365–2376 (2015).
46. Vitalis, A. & Caflisch, A. Equilibrium sampling approach to the interpretation of electron density maps. *Structure* **22**, 156–167 (2014).
47. Lopéz-Blanco, J. R. & Chacón, P. IMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J. Struct. Biol.* **184**, 261–270 (2013).
48. Ratje, A. H. et al. Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature* **468**, 713–716 (2010).

49. Saha, M. & Morais, M. C. FOLD-EM: automated fold recognition in medium-and low-resolution (4–15 Å) electron density maps. *Bioinformatics* **28**, 3265–3273 (2012).
50. de Vries, S. J. & Zacharias, M. ATTRACT-EM: a new method for the computational assembly of large molecular machines using cryo-EM maps. *PLoS One* **7**, e49733 (2012).
51. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
52. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
53. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
54. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
55. Gil, V. A. & Guallar, V. pyRMSD: a Python package for efficient pairwise RMSD matrix calculation and handling. *Bioinformatics* **29**, 2363–2364 (2013).
56. McInnes, L., Healy, J. & Astels, S. hdbscan: hierarchical density based clustering. *J. Open Source Softw.* **2**, 205 (2017).
57. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
58. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
59. Beck, M. & Hurt, E. The nuclear pore complex: understanding its function through structural insight. *Nat. Rev. Mol. Cell Biol.* **18**, 73–89 (2017).
60. Lin, D. H. et al. Architecture of the symmetric core of the nuclear pore. *Science* **352**, aaf1015 (2016).
61. Drin, G. et al. A general amphipathic α-helical motif for sensing membrane curvature. *Nat. Struct. Mol. Biol.* **14**, 138–146 (2007).
62. Kim, S. J. et al. Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482 (2018).
63. Dauden, M. I., Jaciuk, M., Müller, C. W. & Glatt, S. Structural asymmetry in the eukaryotic Elongator complex. *FEBS Lett.* **592**, 502–515 (2018).

Acknowledgements

We thank coauthors of the publications in which the Assembline protocol has been applied. We are grateful to D. Ziemianowicz and K. Kaszuba for comments on the manuscript and the protocol, and A. Obarska for feedback on the protocol. The work has been supported by the Federal Ministry of Education and Research of Germany (FKZ 031L0100)

Author contributions

V.R., K.K. and J.K. developed the protocol. V.R and J.K wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41596-021-00640-z.

Correspondence and requests for materials should be addressed to Jan Kosinski.

Peer review information *Nature Protocols* thanks André Hoelz, Peter J. Peters and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 20 December 2020; Accepted: 30 September 2021;
Published online: 29 November 2021

Related links

Key references using this protocol

Beckham, K. S. H. et al. *Sci. Adv.* **7**, eabg9923 (2021): https://doi.org/10.1126/sciadv.abg9923
Allegretti, M. et al. *Nature* **586**, 796–800 (2020): https://doi.org/10.1038/s41586-020-2670-5
Zimmerli, C. E. et al. Preprint at *bioRxiv* (2020): https://doi.org/10.1101/2020.07.30.228585
Dauden, M. I. et al. *EMBO Rep.* **18**, 264–279 (2017): https://doi.org/10.15252/embr.201643353
Kosinski, J. et al. *Science* **352**, 363–365 (2016): https://doi.org/10.1126/science.aaf0643

Essential data used in this protocol

Allegretti, M. et al. *Nature* **586**, 796–800 (2020): https://doi.org/10.1038/s41586-020-2670-5
Dauden, M. I. et al. *EMBO Rep.* **18**, 264–279 (2017): https://doi.org/10.15252/embr.201643353

nature research

Corresponding author(s): Jan Kosinski

Last updated by author(s): Sep 2, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Not applicable
Data analysis	The Assembline software is freely available as an open-source Python package (website: https://www.embl-hamburg.de/Assembline/), which can be installed from source code (https://git.embl.de/kosinski/Assembline) or from the Anaconda repository (https://anaconda.org/kosinskilab/assembline).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All datasets needed for the step-by-step tutorials (i.e., for yeast NPC and Elongator complex modelling) are provided in https://git.embl.de/rantos/scnpc_tutorial.git (for the yeast NPC) and https://git.embl.de/kosinski/elongator_tutorial.git (for the Elongator complex).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample-size calculation is not relevant to our study
Data exclusions	No data has been excluded
Replication	The reproducibility has been addressed through independent Monte Carlo runs and exhaustiveness analysis.
Randomization	Not applicable
Blinding	Investigators were not blinded to identity of analyzed data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

5.4 PUBLICATION IV: "Structure of prototypic peptide transporter DtpA from *E. coli* in complex with Valganciclovir provides insights into drug binding of human PepT1"

In addition to my work regarding the integrative modeling of yeast NPCs (see Synopsis sections 5.1, 5.2), during my PhD I was involved in another collaborative project regarding the structural elucidation of a human peptide transporter. The results of this project are thus presented below.

The cytoplasmic membrane serves as a semi-permeable barrier to molecules "trying" to enter or exit the cells due to the existence of many anchored or embedded membrane proteins[129]. In general, there are two types of membrane proteins, namely the integral and peripheral, with the first category including proteins which are fully embedded in the membrane. All the transporter protein families belong to the integral membrane protein category.

One of the most studied superfamilies of transporter proteins is the major facilitator superfamily (MFS) of transporters. The members of this superfamily are capable of transporting solutes, including small peptides and drugs, through the membrane. Different protein families constitute the MFS superfamily, with the proton-dependent oligopeptide transporters (POTs) exhibiting probably the highest pharmacological relevance due to their ability to actively transport short drug-like peptides to the cytoplasm as well as the existence of homologs in both prokaryotes and eukaryotes. The conserved MFS structural architecture consists of 12 transmembrane α-helices forming the transporter core organized in two bundles, the N- and C-terminal bundle helices. Prokaryotic homologs, however, have two additional transmembrane helices, namely HA and HB, which seem to play a functional role in active solute transport[130]. Eukaryotic homologs on the other hand, have an extra soluble extracellular domain. There are currently more than eight published atomic structures of POTs that were acquired with X-ray crystallography (some of them were co-crystallized with short peptides bound) and from these structural studies together with further biochemical characterizations, it was found that different POTs transport different types of peptides and solutes.

Two mammalian POT homologs, human PepT1 and PepT2 (hPepT1, hPepT2), transport di/tripeptides and peptide-like compounds with cotransport of protons. These POTs can transport also various drugs and prodrugs such as valganciclovir and valacyclovir (L-valine derivatives of the guanosine analogs ganciclovir and acyclovir), which are used for the treatment of cytomegalovirus and herpes simplex virus infections[131]. Although the ligand binding activity of both human transporters is well characterized, prior to our study, there were still no available structures that could provide a detailed overview of the transport mechanism and drug uptake. In this context, Yonca Ural-Blimke from Christian Loew's group (EMBL Hamburg) successfully produced well-diffracting crystals that led to the structural elucidation at near-atomic resolution of a bacterial hPepT1 homolog called DtpA with and without the prodrug valganciclovir. DtpA exhibits high conservation with hPepT1 in the binding site domains as well as in the transportable solute repertoire.

The crystal structure of DtpA was resolved at 3.30 Å resolution in the ligand-free form and at 2.65 Å resolution in complex with the antiviral prodrug valganciclovir which was proven earlier that

can be transported by this bacterial transporter and its human homolog (i.e. hPepT1). Additionally, nanobodies were developed and used for the crystallization of DtpA which are known to improve the protein thermal stability. All the aforementioned crystal structures of DtpA exhibited the typical domain arrangement of POTs with two distinct six α-helical bundles separated by the additional helices HA and HB (Fig. 25). From the three possible and distinct conformations that POTs can adopt during active transport of solutes, the DtpA structures were "captured" in the inward open orientation.

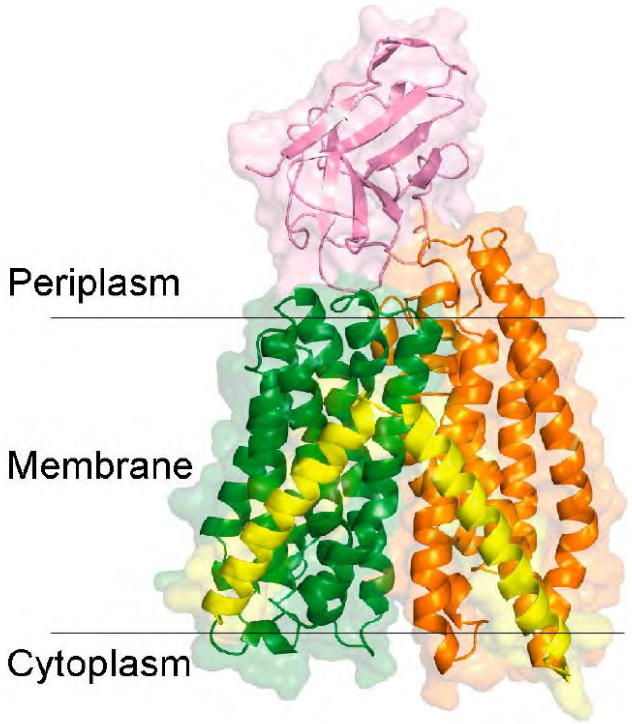


Figure 25: Representation of the DtpA crystal structure without the bound ligand (valganciclovir). DtpA appeared to be structurally organized as a typical POT family member, with two six helical bundles (green and orange) being separated by the additional α-helices HA and HB (yellow). The stabilizing nanobody that was used for the crystallization is also depicted in the periplasmic-facing side of the transporter (pink). The relevant cellular environments are also noted. Figure adapted from[132].

The DtpA structure, which was in complex with the bound valganciclovir, revealed an unexpected binding conformation of this prodrug. Valganciclovir has a similar length to natural tripeptides which have a characterized binding orientation based on previously published structures of POTs[133]. Given that fact, the N-terminal valine of valganciclovir was expected to mimic the N-terminal residue of a typically bound tripeptide. However, the structure revealed that it is the guanine moiety and not the valine residue that occupies the position corresponding to the N-terminal residue of a bound tripeptide (Fig. 26).

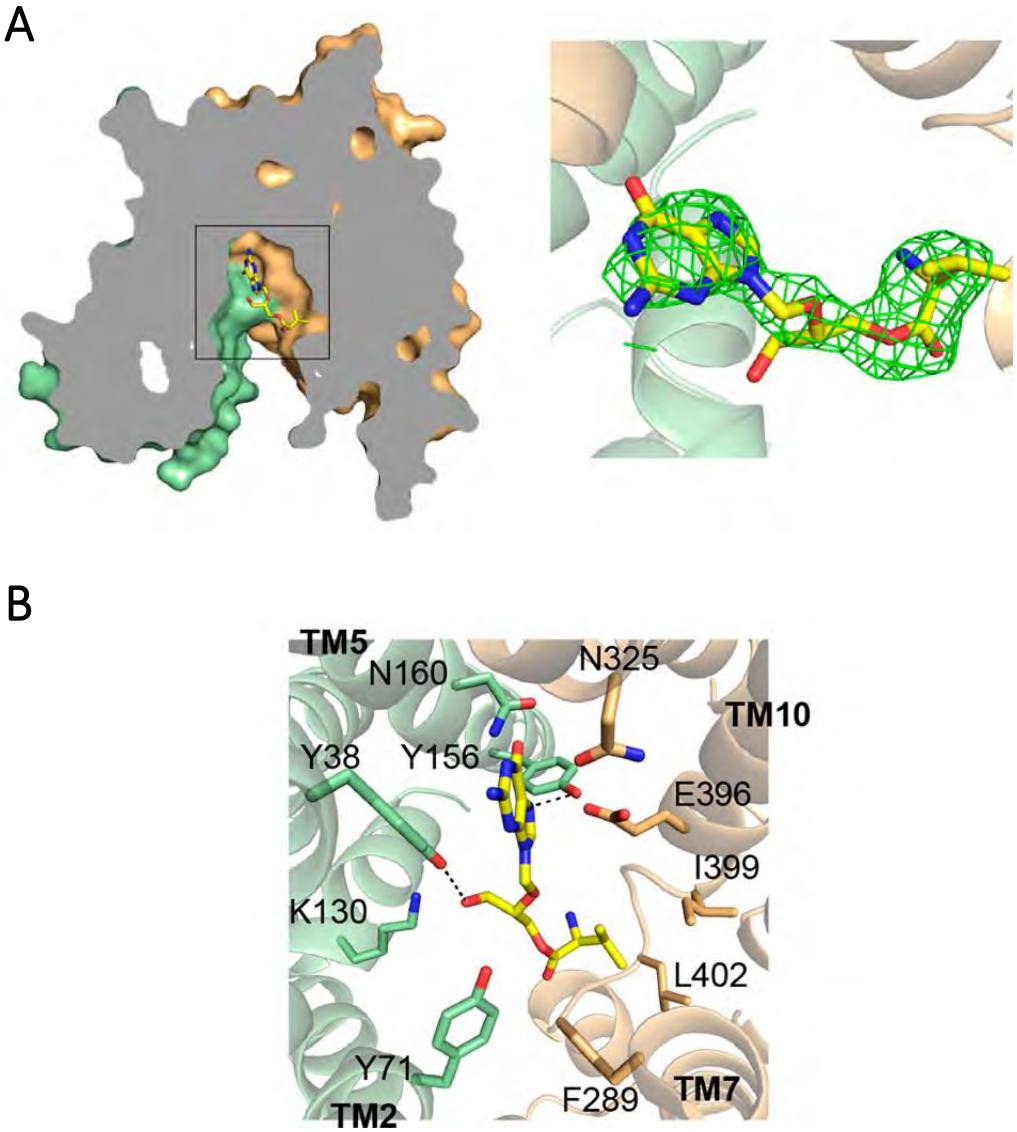


Figure 26: Schematic representation of valganciclovir bound to DtpA. (A) (Left) Ball and sticks representation of the prodrug valganciclovir inside the DtpA (shown in surface representation) binding cavity. (Right) Illustration of the omit map for the bound prodrug in the DtpA active site. (B) Depiction of the possible interaction patterns formed between the bound prodrug (ball and stick representation) and DtpA (ribbon representation) residues from the binding site. All the possible interacting residue positions and types are denoted. Figure adapted from[132].

Further mutational studies with respect to the binding site of DtpA were conducted by Yonca Ural-Blimke (Loew group, EMBL Hamburg) in order to characterize the possible binding modes of di-/tri-peptides. Indeed, the latter experiments confirmed that the observed binding affinities of the constructed mutants for selected di/tripeptides and two other drugs were strongly reduced relatively to wild type DtpA.

Considering that DtpA shares a relatively high sequence conservation with its human homolog hPepT1, I used the structure of DtpA with the bound prodrug to build a homology model of hPepT1. I first validated in a sequence-based manner whether DtpA could actually be used for building an accurate hPepT1 model. Indeed, when the hPepT1 sequence was aligned against the

databases (including both Multiple Sequence Alignments and Hidden Markov Model profiles) of HH-suite package[134], the sequence of DtpA was always detected as a close homolog ranking amongst the top aligned hit sequences from these databases. Next, I subjected the hPepT1 sequence to a series of sequence-based and structure-based predictions, including secondary structure predictions with tools such as GeneSilico MetaServer[135]. Upon retrieving the pairwise sequence alignment of hPepT1 with DtpA, after aligning them individually with a Hidden Markov Model (HMM) profile that I constructed for this family based on relevant subset sequences from the HH-suite databases, I performed manual adjustments of the alignment using Swiss-Pdb Viewer[136] guided by the secondary structure predictions (Fig. 27).

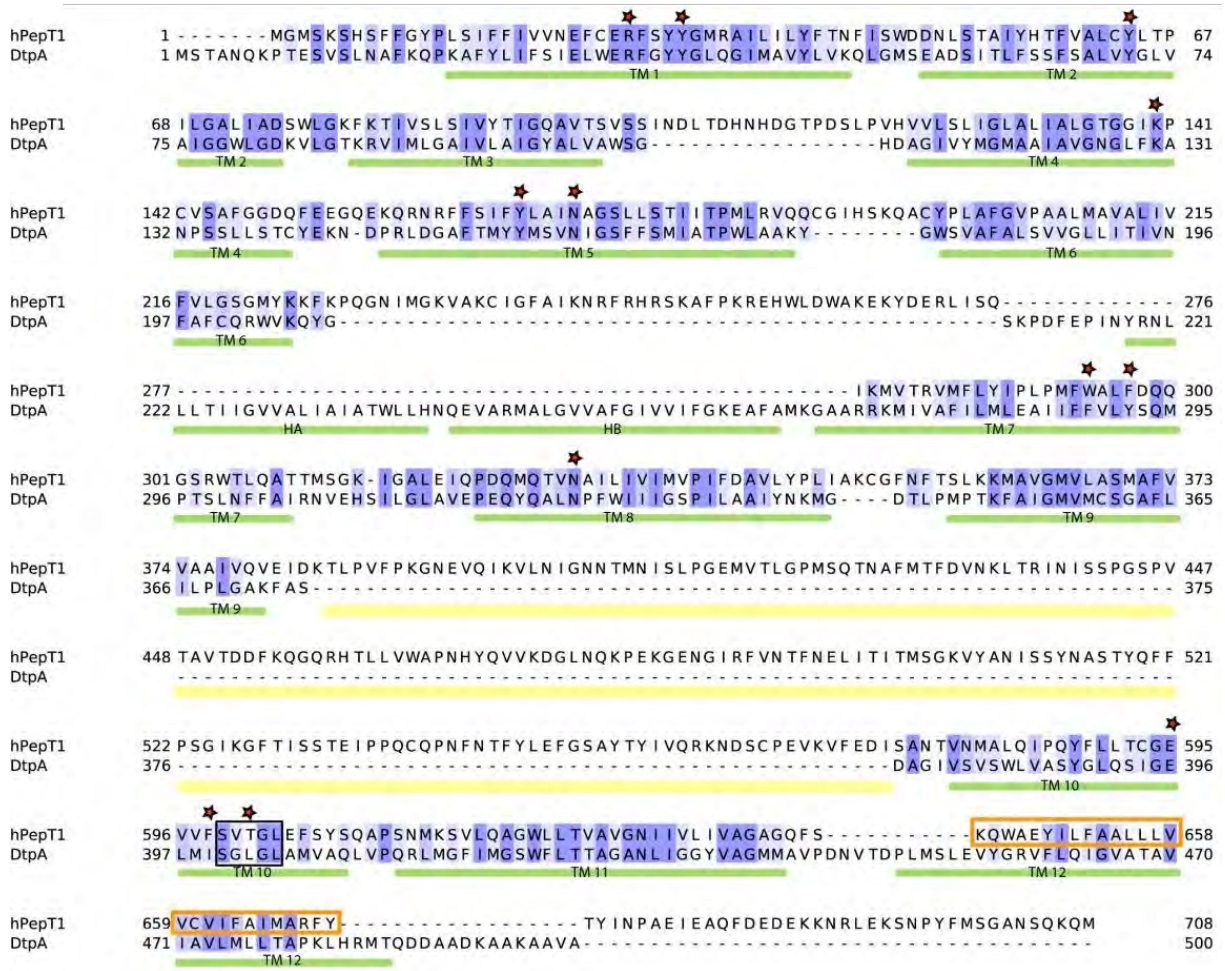


Figure 27: Curated pairwise sequence alignment of hPepT1 and DtpA. The two homologs exhibit high sequence conservation in the helical bundle domains and their active site, while they seem to follow precisely the documented overall fold of MSF POT members. The secondary structure elements (green bars), the eukaryote-exclusive ECD (yellow), the ligand binding residues (red stars) as well as the pairwise sequence similarity (darker to lighter blue) are noted. Figure adapted from[132].

Having retrieved the final curated pairwise alignment of sequences for these transporters, I produced homology models of hPepT1 with bound valganciclovir after utilizing the solved DtpA model as a structural template. The hPepT1 models were produced by a custom homology modeling pipeline that I established based on the modeling suite MODELLER[137]. Their structural architecture followed precisely the overall fold of the DtpA structure that was used as a modeling template (Fig. 28).

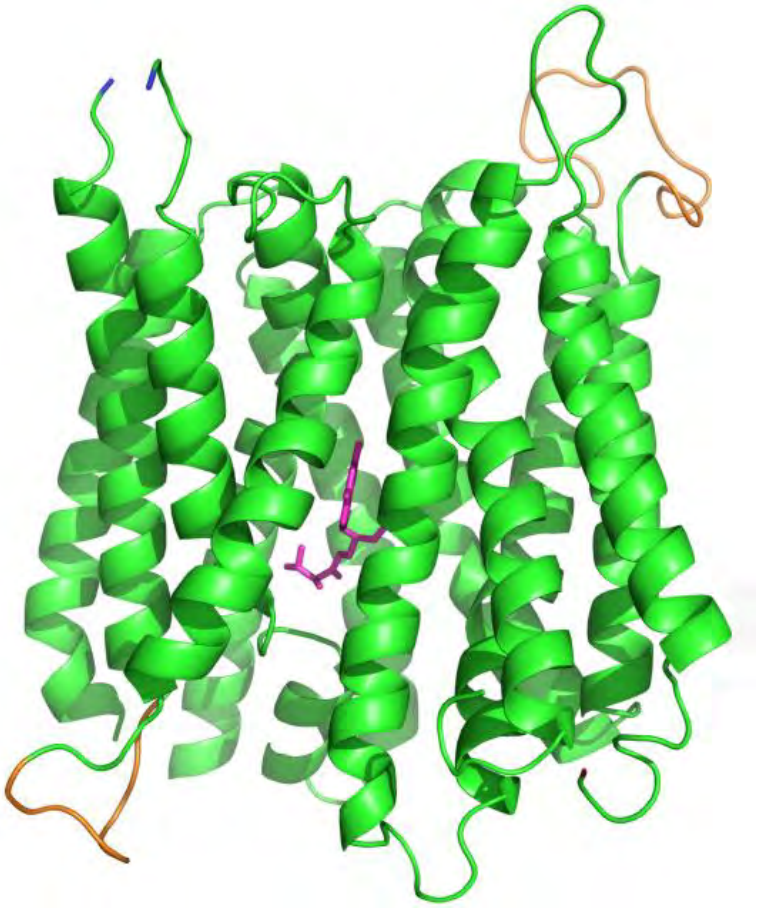


Figure 28: Representative hPepT1 homology model with bound valganciclovir. Representative homology model of hPepT1 (green ribbons) in complex with the prodrug valganciclovir (pink ball and stick) from an ensemble of calculated models that were built based on the DtpA structure from this study. The parts of the structure that correspond to the non-modeled long linker (red) and ECD (blue) as well as the variable loops (orange) are denoted.

The eukaryote-specific ECD as well as a long connecting loop between the two helical bundles were omitted from modeling. The resulting hPepT1 homology model was deposited in Model Archive (relevant structural database for such models)[138]. External tools, such as QMEANBrane[139], were also utilized to assess the accuracy of the produced hPepT1 homology model (Fig. 29).

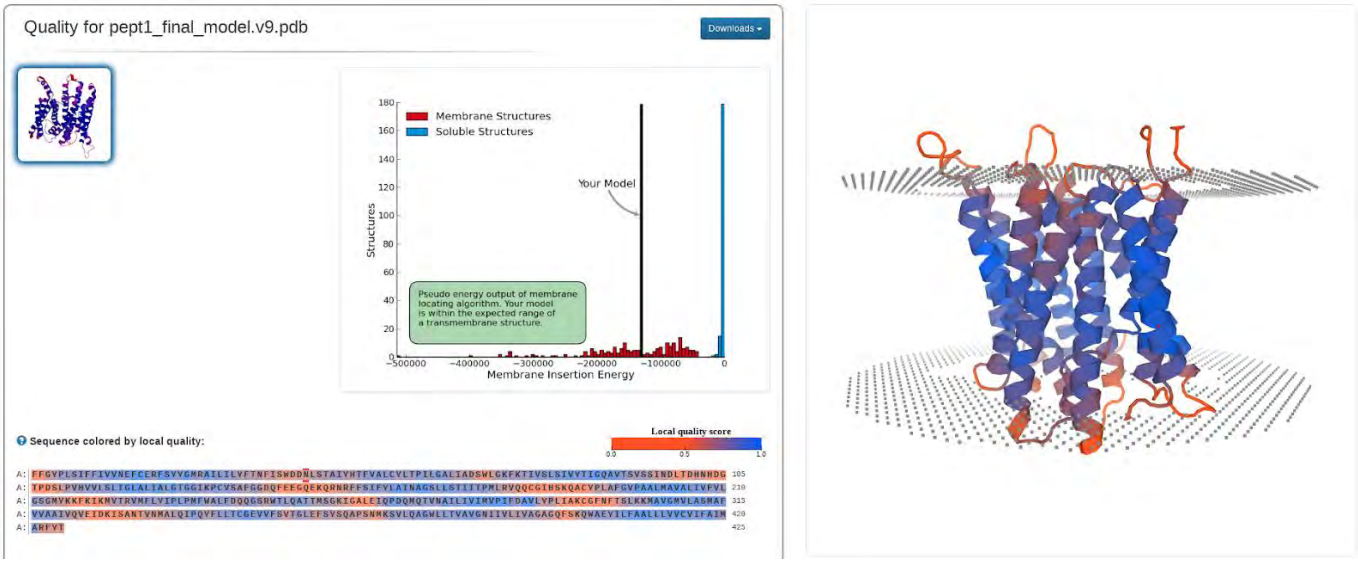


Figure 29: Local structural quality of human PepT1 homology model assessed by QMEANBrane. (Left) The membrane insertion energy assessment from the tool QMEANBrane[139] confirmed the predicted transmembrane nature of the model and revealed the high overall structural confidence of predicted domains. The color-coding for quality scores is noted. (Right) Structural representation of the human PepT1 homology model[132] colored according to the per-residue structural quality scores scheme provided (as left). The helical bundles of the PepT1 model exhibit the highest structural confidence with the loop regions receiving lower quality scores.

By close inspection and superposition analysis between the two homologous structures, it was revealed that the residues coordinating the guanine base of valganciclovir that was mentioned earlier, are fully conserved between DtpA and hPepT1. In addition, residues interacting with the N-terminal valine residue from the prodrug form similar hydrophobic pockets in both proteins (Fig. 30).

Taken together, the resolved structure of DtpA in the ligand-bound state from this study revealed the binding mode of the prodrug valganciclovir. The DtpA structure in complex with the prodrug was used to produce the structure of the human homolog hPepT1. The produced homology models together with further computational analysis, revealed the structure of hPepT1 in complex with valganciclovir and suggested that the binding mode of the prodrug is the same in the case of the human homolog. These novel structural insights could help in the design of improved prodrugs targeting this human transporter.

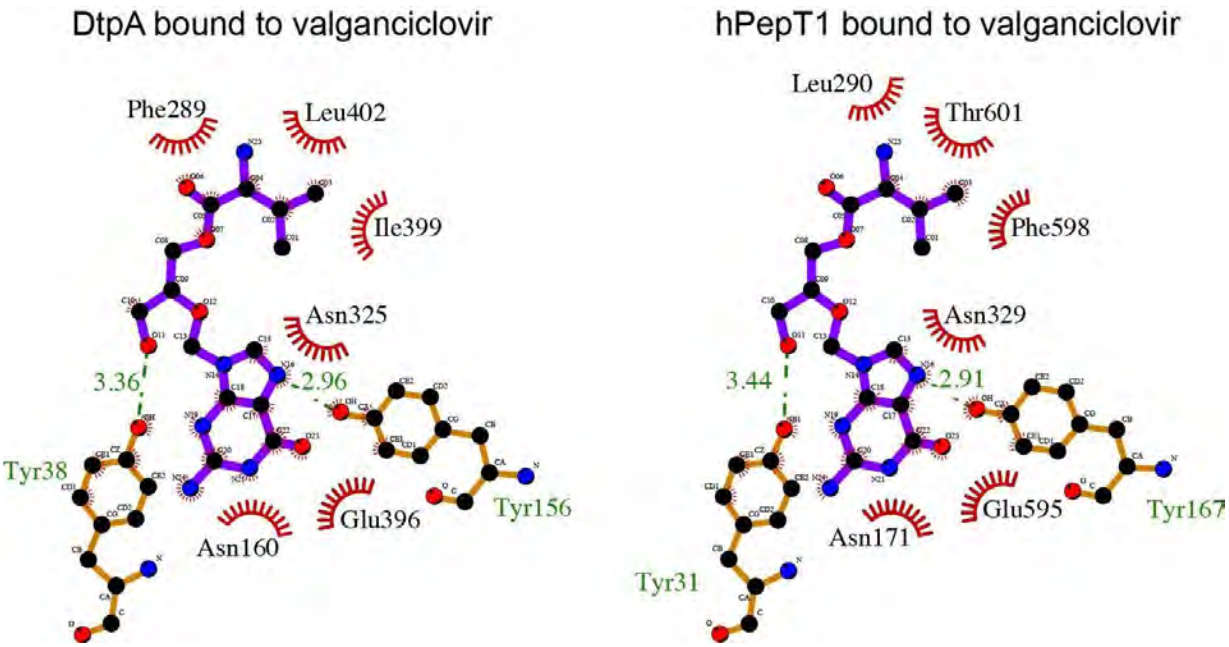


Figure 30: Graphical comparison of the ligand binding sites of DtpA and hPepT1 in complex with valganciclovir. The interaction patterns formed in the binding site of DtpA (left) and hPepT1 (right) with the prodrug valganciclovir highlight the significant similarities regarding interacting residue positions and types. The 2D diagrams were produced with LigPlot+[140], with potential hydrogen bonds shown as green dashed lines and hydrophobic interactions shown as red spoked arcs. Figure adapted from[132].

JACS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

Cite This: J. Am. Chem. Soc. 2019, 141, 2404−2412

pubs.acs.org/JACS

Structure of Prototypic Peptide Transporter DtpA from *E. coli* in Complex with Valganciclovir Provides Insights into Drug Binding of Human PepT1

Yonca Ural-Blimke,[†] Ali Flayhan,[†] Jan Strauss,[†] Vasileios Rantos,[†] Kim Bartels,[†] Rolf Nielsen,[†] Els Pardon,[‡,§] Jan Steyaert,[‡,§] Jan Kosinski,[†,∥] Esben M. Quistgaard,[†,⊥,#] and Christian Löw*[,†,⊥]

[†]Centre for Structural Systems Biology (CSSB), DESY and European Molecular Biology Laboratory Hamburg, Notkestrasse 85, D-22607 Hamburg, Germany

[‡]Structural Biology Brussels, Vrije Universiteit Brussel (VUB), Brussels 1050, Belgium

[§]VIB-VUB Center for Structural Biology, VIB, Brussels 1050, Belgium

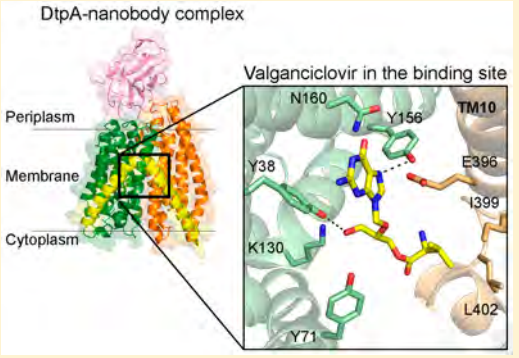
[∥]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

[⊥]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Scheeles väg 2, SE-17177 Stockholm, Sweden

[#]Department of Molecular Biology and Genetics − DANDRITE, Gustav Wieds Vej 10, Aarhus University, DK-8000 Aarhus C, Denmark

Supporting Information

ABSTRACT: Members of the solute carrier 15 family (SLC15) transport di- and tripeptides as well as peptidomimetic drugs across the cell membrane. Structures of bacterial homologues have provided valuable information on the binding and transport of their natural substrates, but many do not transport medically relevant drugs. In contrast, a homologue from *Escherichia coli*, DtpA (dipeptide and tripeptide permease), shows a high similarity to human PepT1 (SLC15A1) in terms of ligand selectivity and transports a similar set of drugs. Here, we present the crystal structure of DtpA in ligand-free form (at 3.30 Å resolution) and in complex with the antiviral prodrug valganciclovir (at 2.65 Å resolution) supported by biochemical data. We show that valganciclovir unexpectedly binds with the ganciclovir moiety mimicking the N-terminal residue of a canonical peptide substrate. On the basis of a homology model we argue that this binding mode also applies to the human PepT1 transporter. Our results provide new insights into the binding mode of prodrugs and will assist the rational design of drugs with improved absorption rates.



INTRODUCTION

The proton-dependent oligopeptide transporters (POTs) constitute a subfamily of the major facilitator superfamily (MFS), transporting di/tripeptides and peptide-like compounds coupled to cotransport of protons. Mammalian POTs are also referred to as solute carrier 15 transporters (SLC15s). Human PepT1 (SLC15A1) and PepT2 (SLC15A2) are of great pharmacological interest due to their roles in intestinal uptake and renal reabsorption of not only dietary peptides but also drugs such as β-lactam antibiotics (ceftibuten), angiotensin-converting enzyme inhibitors (enalapril), and antiviral prodrugs (valganciclovir and valacyclovir).[1] The prodrugs valganciclovir and valacyclovir are L-valine derivatives of the guanosine analogs ganciclovir and acyclovir, used for treating cytomegalovirus and herpes simplex virus infections, respectively. Conjugation to valine makes these compounds more soluble and improves their uptake rates in humans by turning them into substrates for PepT1 and PepT2.[2,3] PepT1 and PepT2 have been well characterized in terms of ligand binding, but their structures remain unknown. The bacterial homologue of PepT1, DtpA (dipeptide and tripeptide permease), provides an excellent prototype to understand the molecular mechanism of peptide and drug transport due to the high conservation of the binding site as well as the highly similar substrate specificity profile compared to PepT1.[4,5] However, like for their human homologues, structural insights into relevant drug binding are missing for bacterial peptide transporters. Therefore, we combined biochemical, biophysical, and structural approaches to obtain molecular insights into binding and transport of peptides and drugs of the prototypic DtpA transporter, which provided the basis for structural modeling of the human PepT1 transporter.

Received: October 21, 2018
Published: January 15, 2019

Table 1. Crystallographic Data Collection and Refinement Statistics for the Ligand-Free and Valganciclovir-Bound DtpA-N00 Structures[a]

	DtpA-N00 in glycine buffer, pdb id 6GS7	DtpA-N00 in MES buffer, pdb id 6GS1	DtpA-N00-Valganciclovir, pdb id 6GS4
data collection			
space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
cell dimensions			
a, *b*, *c* (Å)	55.14, 120.53, 163.43	55.46, 120.72, 163.33	54.94, 120.19, 163.67
α, β, γ (deg)	90, 90, 90	90, 90, 90	90, 90, 90
resolution (Å)	19.94−3.30 (3.42−3.30)	48.16−3.29 (3.41−3.29)	19.68−2.65 (2.74−2.65)
R_{merge}	21.74 (217.7)	26.73 (186.4)	10.74 (207.7)
R_{pim}	7.97 (74.81)	11.14 (78.46)	4.65 (86.77)
I/σ	7.04 (0.83)	6.37 (1.06)	10.74 (1.06)
CC 1/2	99.2 (49.4)	99.4 (56.0)	99.8 (46.3)
completeness (%)	99.02 (99.70)	99.34 (99.47)	98.98 (97.29)
redundancy	8.5 (8.8)	6.7 (6.6)	6.4 (6.6)
refinement			
resolution (Å)	19.94−3.30 (3.42−3.30)	48.16−3.29 (3.41−3.29)	19.68−2.65 (2.74−2.65)
no. of reflns	16919 (1671)	17219 (1679)	32178 (3122)
R_{work}/R_{free}	21.03/24.75	25.11/26.60	21.57/23.96
no. of atoms			
protein	4568	4576	4542
ligand/ion			60
water		1	7
clashscore	4.46	6.41	3.57
B-factors			
protein	108.69	83.22	89.74
ligand/ion			133.49
water		59.18	74.70
rms deviations			
bond lengths (Å)	0.004	0.004	0.004
bond angles (deg)	0.92	0.91	0.64

[a]Values in parentheses are for the highest-resolution shell.

RESULTS AND DISCUSSION

To produce well-diffracting crystals of DtpA, we used nanobodies as crystallization chaperones, which can improve crystallization by increasing the thermal stability of the target protein, by locking it in a specific conformation, and by mediating crystal contacts. Four nanobodies were generated, which were confirmed to bind a conformational epitope of DtpA and increased its thermal stability by more than 10 °C (Supplementary Figure 1). Nanobody N00 was essential to obtain well-diffracting crystals. Three structures of the DtpA-N00 complex were determined (Table 1, Supplementary Figure 2), including two 3.30-Å ligand-free DtpA structures from different crystallization conditions (MES or glycine buffer) (Figure 1a) and a 2.65-Å valganciclovir-bound structure (glycine buffer). The higher resolution of the drug-bound structure in comparison to the ligand-free structures may relate to the fact that valganciclovir stabilizes the DtpA-nanobody complex further and reduces its conformational flexibility (Figure 1b).

The transporter adopts the same basic fold as previously published bacterial POTs:[6−13] It is built from two six-helical MFS domains separated by two additional transmembrane helices (TMs) named HA and HB (Figure 1a). MFS transporters can adopt inward facing, occluded, or outward facing conformational states.[14] Here, the transporter is in the inward open form, which is stabilized by the conformation-selective binding of N00 across the interface of the two MFS domains on the periplasmic side (Supplementary Figure 3).

The valganciclovir-bound DtpA-N00 structure is highly similar to the ligand-free structures, with an rmsd of 0.33 Å over 3809 atoms (glycine buffer condition) and 0.42 Å over 3869 atoms (MES buffer condition). Clear electron density could be observed for the bound valganciclovir molecule (Figure 2a). Valganciclovir has two chiral centers: one at the Cα of L-valine and the other at the C2 position (the latter is highlighted with a red star in Figure 1b). It is synthesized as a mix of two diastereomers that only differ at the C2 position. Both diastereomers can be transported by hPepT1 (human peptide transporter 1).[15] It was not possible to distinguish between the two diastereomers in the electron density map of DtpA (Figure 2b).

The valganciclovir-bound structure reveals an unexpected binding mode for the prodrug. The ligand-binding site of POTs has been previously characterized from structures of different transporters bound to di/tripeptides[9,10,16,17] or the phosphonodipeptide alafosfalin.[7,18] Since valganciclovir has a length similar to tripeptides, it has been suggested that the N-terminal valine residue of the prodrug would mimic the N-terminal residue of a di/tripeptide when bound to the transporter.[19] However, in our DtpA-N00-valganciclovir structure, it is the guanine moiety and not the valine residue that occupies the position corresponding to the N-terminal residue of a peptide substrate. The guanine base is coordinated by Y38, Y156, N160, N325, and E396, which are equivalent to the residues that have previously been observed to bind the N-terminal residues of peptide substrates[16,18] (Figure 2c, Supplementary Figure 4).

DOI: 10.1021/jacs.8b11343
J. Am. Chem. Soc. 2019, 141, 2404−2412

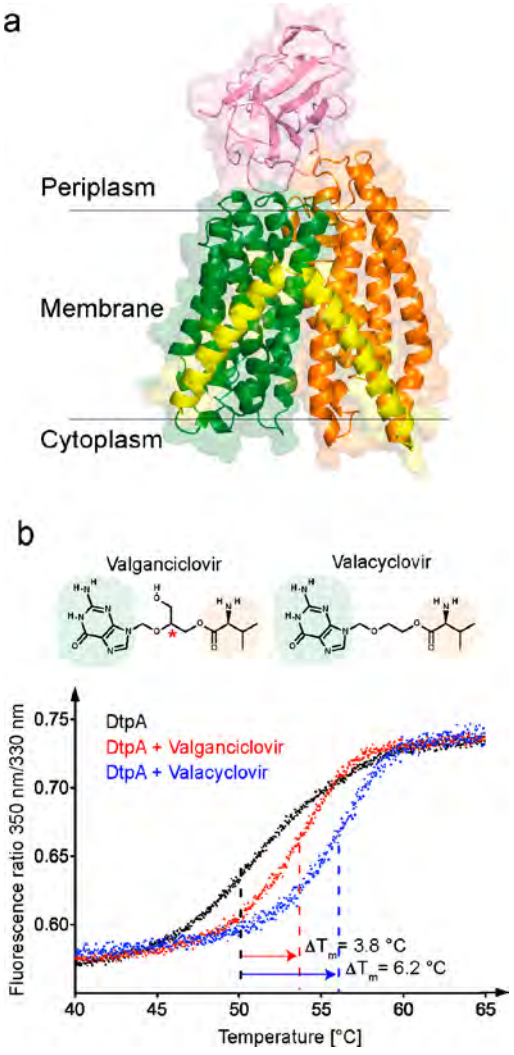


Figure 1. Structure of the peptide transporter-nanobody complex DtpA-N00, and thermal stability with the prodrugs valganciclovir and valacyclovir. (a) Structure of the DtpA-N00 complex (in MES buffer). N-terminal MFS domain is colored in green, C-terminal MFS domain in orange, HA-HB domain in yellow, and N00 in pink. (b) Thermal stabilization effect of valganciclovir and valacyclovir on DtpA, and chemical structure of the prodrugs. Guanine base is highlighted in pale green and the N-terminal valine residue in light orange. Valganciclovir is a mix of two diastereomers with alternative conformations at the chiral center highlighted with a red star.

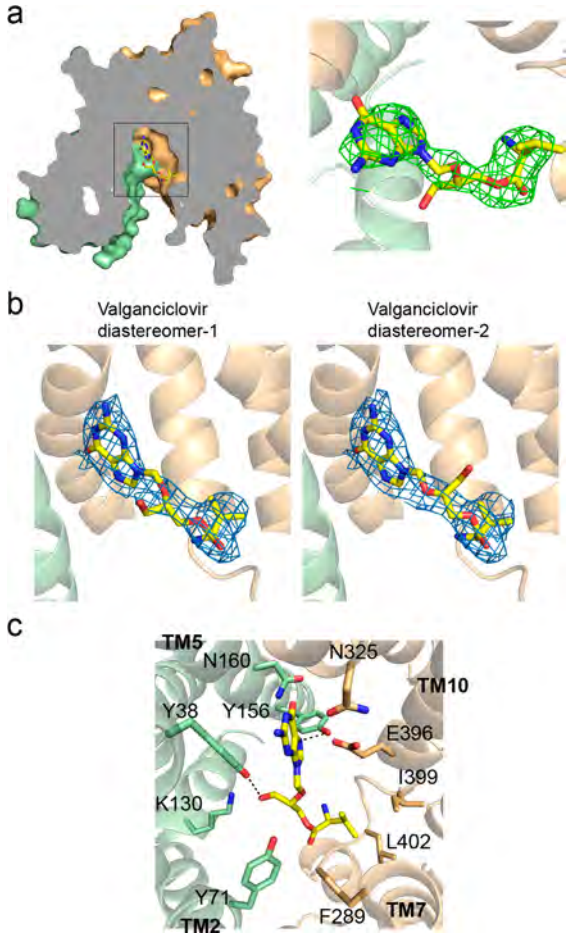


Figure 2. Binding mode of valganciclovir in the crystal structure of the DtpA-N00-valganciclovir complex. (a) (Left) Valganciclovir is shown in sticks in the ligand-binding site of DtpA, which is illustrated as a surface model (MFS domains are colored as in Figure 1). Nanobody is omitted for clarity. (Right) Omit map for valganciclovir contoured at the 3 − σ level. (b) $2F_o - F_c$ electron-density map contoured at 1 − σ of the refined DtpA-N00 structure bound to two diastereomers of valganciclovir. All further figures are prepared with diastereomer-1 (left). (c) Interactions of valganciclovir in the binding site. Same coloring as in other panels, but the cartoon representation is at 50% transparency. Potential hydrogen bonds ≤ 3.4 Å are shown as dashes.

The N-terminal valine residue of valganciclovir extends instead into a pocket formed by TM7 and an intrahelical loop in TM10 (I399-SGLG-L404). Here it forms van der Waals interactions with F289 from TM7 as well as with I399 and L402 from the TM10 loop (Figure 2c). In MFS transporters, it is not uncommon to find intrahelical loops in cavity-lining helices.[14] However, among POTs, a loop within TM10 has previously only been observed in $PepT_{So2}$ (peptide transporter from bacterial species *Shewanella oneidensis*).[18] Here TM7 and the loop within TM10 line the pocket that accommodates the side chains in position two of di- and tripeptides, i.e., "pocket 2" (P2).[10] In $PepT_{St}$ (peptide transporter from bacterial species *Streptococcus thermophilus*), which has an uninterrupted TM10, P2 is instead formed by TM2, TM7, and TM11[9] and is located in a partially overlapping position that is closer to the center of the cavity.

Compared to POT structures with an uninterrupted TM10, the periplasmic half of the helix superimposes well whereas the cytoplasmic half is shifted considerably, which in turn causes TM11 to be pushed away from the binding cavity (Figure 3). For example, in an overlay of $PepT_{St}$ from *S. thermophilus* on DtpA, the distance between L404 Cα in the cytoplasmic half of TM10 in DtpA and the equivalent L408 Cα atom in $PepT_{St}$ is 6.1 Å and the distance between F424 Cα in TM11 of DtpA and the equivalent F428 Cα atom in $PepT_{St}$ is 6.8 Å (Figure 3a). As a result, the binding cavity is significantly enlarged in POTs with a split TM10 (Figure 3b). Indeed, in comparison to other POT transporters with known structures, the binding site of DtpA is the largest (Figure 3b).

We conclude that DtpA binds valganciclovir in a particularly large binding cavity by accommodating its guanine base in a site that in other POTs generally engages the N-terminus of peptides and inserting its valine residue into a pocket, which is largely equivalent to the P2 peptide side chain binding pocket of $PepT_{So2}$.

DOI: 10.1021/jacs.8b11343
J. Am. Chem. Soc. 2019, 141, 2404−2412

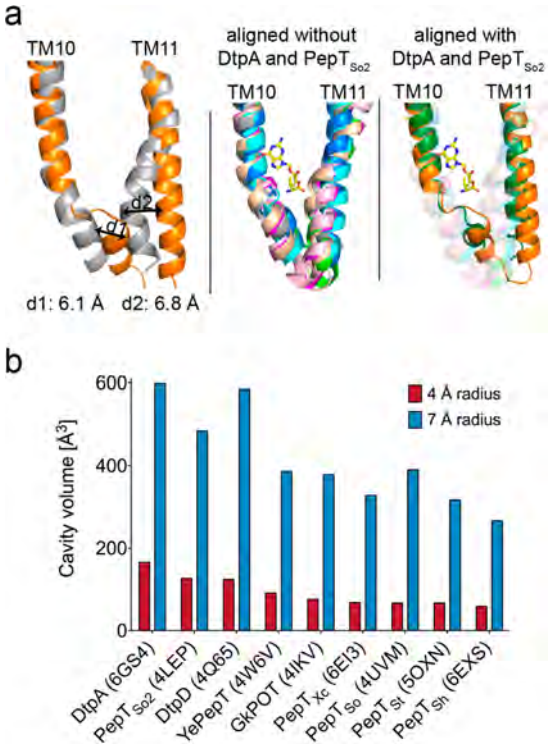


Figure 3. Structural comparison of TM10 and TM11 in DtpA and other known POTs. (a) Structural comparison of the TM10-TM11 region of DtpA (orange) and PepT$_{St}$ (pdb id 5oxn in gray). Loop within TM10 of DtpA (d1) shifts the C-terminal part of TM10 and displaces the N-terminal part of TM11 by 6.8 Å (d2) relative to the position in PepT$_{St}$. Equivalent displacement is not found in PepT$_{So}$ (pdb id 4uvm in magenta), GkPOT (pdb id 4ikv in light green), DtpD (pdb id 4q65 in cyan), YePepT (pdb id 4w6v in blue), PepT$_{Xc}$ (pdb id 6ei3 in light pink), or PepT$_{Sh}$ (pdb id 6exs in wheat) but only in DtpA and PepT$_{So2}$ (pdb id 4tph in green). Valganciclovir is shown in yellow. (b) Cavity volume around the Cα atom of the valine moiety in valganciclovir, which is intruding into the space created by the TM10 intrahelical loop, was calculated with a radius of 4 and 7 Å using POVME 2.0[20,21] for all published POT structures.

To improve our understanding of DtpA and verify its substrate preferences toward peptides and drugs of different size, we screened a ligand library including 18 dipeptides, 10 tripeptides, and 7 drugs. The substrate specificity of DtpA has so far only been characterized in the whole cell context, mostly using a competition assay with the fluorophore-conjugated dipeptide AK-AMCA (β-Ala-Lys coupled to the fluorescent (7-amino-4-methylcoumarin-3-yl)acetic acid), where the reduction of the uptake of the fluorescent molecule into the cell is measured in the presence of increasing amounts of di/tripeptides or drugs.[4] Here, in addition to this *in vivo* AK-AMCA competition assay, we performed an *in vitro* thermal shift assay using differential scanning fluorimetry (DSF) (Figure 4a, Supplementary Table 1). The results from both assays are in good agreement. Ligands that are more stabilizing, as determined by DSF, also show stronger inhibition of AK-AMCA uptake (Figure 4b, Supplementary Figure 5, and Supplementary Table 1).

With regard to binding of peptides, a U-test performed on the DSF results indicates that binding of tripeptides increases the thermal stability of DtpA (median = 7.1 °C) significantly more than binding of dipeptides (median = 2.4 °C) (W = 46, n_1 = 10, n_2 = 18, P = 0.035, two tailed; Figure 4a, right). In line with this, the increase of median values for the AK-AMCA competition is significantly higher for tripeptides (median = 68%) compared to dipeptides (median = 53%) (W = 37, n_1 = 10, n_2 = 18, P = 0.005, two tailed; Figure 4b, right). The significant difference of binding and transport between di- and tripeptides was further substantiated by a principal component analysis of the physicochemical properties of the tested peptide library. This analysis showed that the molar volume of a peptide ligand is indeed the most strongly contributing variable (Supplementary Figure 6). Finally, quantitative binding affinity measurements were carried out using microscale thermophoresis (MST) for a subset of these ligands, all showing binding affinities in the low to medium micromolar range (Supplementary Figure 7 and Supplementary Table 1). Notably, the highest binding affinity measured by MST was obtained for the tripeptide LLA (K_d = 58 ± 10 μM; Figure 4c), which also induced the highest thermal stabilization, as determined by DSF, and the strongest inhibitory effect on AK-AMCA uptake of all tested ligands (Figure 4a and 4b and Supplementary Table 1). Taken together, our results indicate that DtpA preferably binds and likely transports tripeptides over dipeptides. This is in contrast to most other POTs with known structures, which show preferences toward dipeptides.[9] However, PepT$_{So2}$,[10] which shares the presence of an intrahelical loop in TM10 with DtpA, is an exception to this rule (Figure 3a), supporting the notion that a split TM10 and the associated enlargement of the binding cavity may be an adaptation for improving or enabling binding and transport of larger substrates.

From all drugs tested, only valacyclovir and valganciclovir showed a thermal stabilization effect on DtpA and prominent AK-AMCA competition (Figure 4a and 4b). Remarkably, these two drugs bind as tightly to the protein as the best binding tripeptides. The affinities measured using MST were 60 ± 13 μM for valacyclovir and 76 ± 16 μM for valganciclovir (Supplementary Table 1), similar to the K_d of 49 ± 0.3 μM recently measured for the binding of valacyclovir to human PepT1.[22]

Next, we measured the effect of N00 on peptide uptake and affinity *in vivo* and *in vitro*. AK-AMCA uptake in cells overexpressing DtpA together with N00 was strongly reduced (Figure 4d). In addition, DtpA modified with a site-specific unnatural amino acid could be cross-linked to N00 upon UV irradiation (Supplementary Figure 8) in lysates, suggesting that N00 stabilizes the protein in the inward open conformation not only in the crystal and in detergent-solubilized form but also in the lipid membrane. Differences in binding affinity upon addition of N00 were typically small (about 2-fold or less), but an increase in affinity of 4−5-fold could be observed for two of the peptides, LA and AL (Supplementary Figure 7 and Supplementary Table 1), indicating that substrate binding affinities may in some cases vary between the inward and the outward open states.

To confirm the location of the ligand-binding site of DtpA by functional assays, five residues (Y38, Y71, K130, N160, and I399) were mutated to alanine. The correct folding of the mutants was confirmed by binding of the conformational nanobody N00 (Supplementary Figure 9a). Binding affinities of these mutants for selected di/tripeptides and the two drugs were strongly reduced relative to wild type (Supplementary Figure 9b and Supplementary Table 2), and AK-AMCA uptake was abolished (Figure 4d). Two additional binding site mutants (Y156A and F289L) that were characterized

DOI: 10.1021/jacs.8b11343
J. Am. Chem. Soc. 2019, 141, 2404−2412

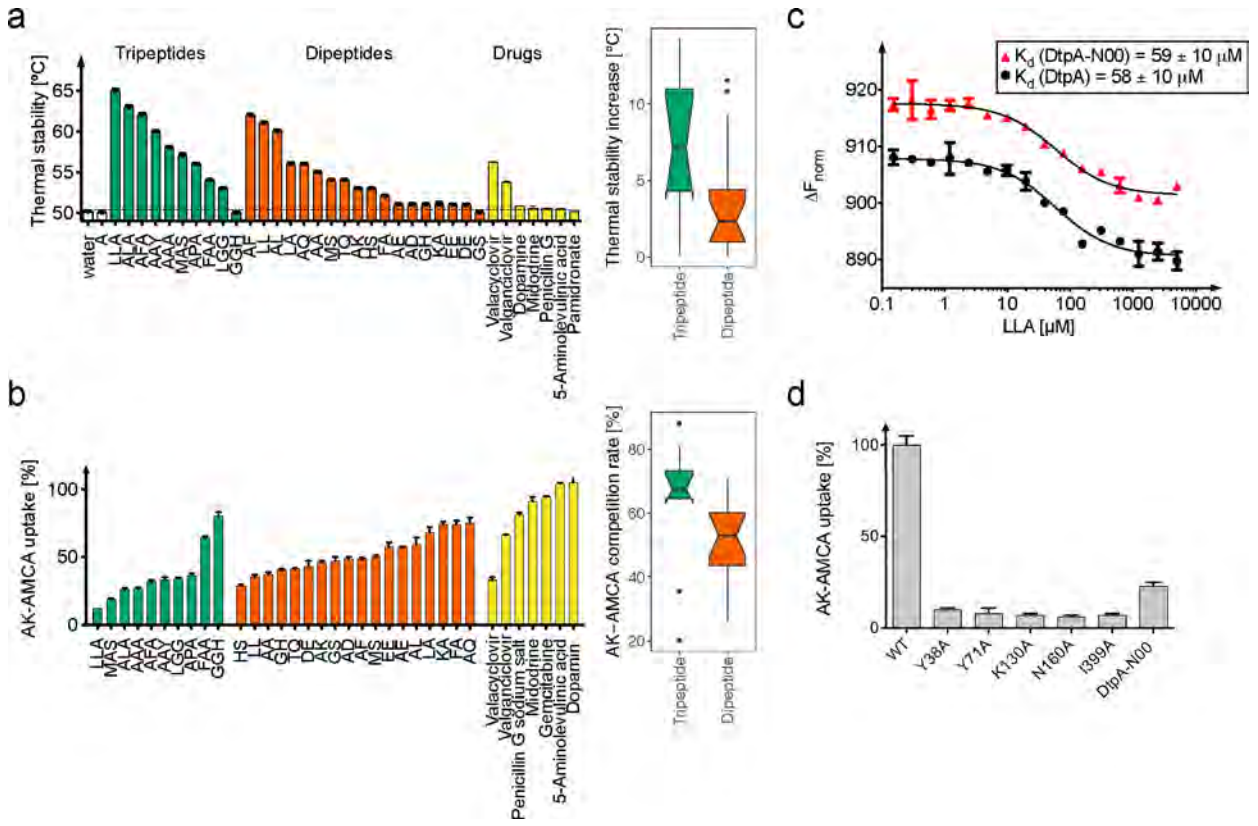


Figure 4. Functional characterization of DtpA. (a) (Left) Thermal stability of DtpA screened with the ligand library at 5 mM concentration. (Right) Notched box plot to compare the thermal stability of DtpA in the presence of di- and tripeptides ($W = 46$, $n_1 = 10$, $n_2 = 18$, $P = 0.035$, two tailed). (b) (Left) AK-AMCA uptake with the same ligand library at 0.5 mM. (Right) Notched box plot to compare AK-AMCA competition uptake for di- and tripeptides. Results here were subtracted from 100% to ease comparison with the thermal stability results ($W = 37$, $n_1 = 10$, $n_2 = 18$, $P = 0.005$, two tailed). (c) Binding affinity curve of DtpA with the tripeptide LLA in black and DtpA-N00 with LLA in red. (d) AK-AMCA uptake of DtpA mutants and cells coexpressing DtpA and N00.

previously were also found to abrogate AK-AMCA uptake.[23] We therefore conclude that the mutational and structural data are in good agreement.

Taking advantage of the high sequence conservation in the binding site (Supplementary Figure 10), a homology model of human PepT1 (hPepT1) in complex with valganciclovir was built based on the DtpA structure, covering the whole sequence except for the extracellular domain (ECD) and the linker between the two MFS domains (Figure 5). All residues involved in valganciclovir binding in DtpA are observed in corresponding positions in hPepT1. Remarkably, the residues coordinating the guanine base of valganciclovir are fully conserved among DtpA (Y38, Y156, N160, E396) and hPepT1 (Y31, Y167, N171, E595) (Supplementary Figure 11). The residues interacting with the N-terminal valine residue form similar hydrophobic pockets in both proteins.

TM5, TM7, and TM10 of hPepT1 have previously been systematically studied by mutations. Here, Y167, N171, and S174 on TM5,[24] F293 and F297 on TM7,[25] and E595 and Y598 on TM10[26] were suggested to play a role in ligand coordination (Supplementary Figure 12). Indeed, all of these residues form part of the ligand-binding site in our model, and most are in close proximity to valganciclovir. Furthermore, high solvent accessibility was observed for residues T601-E604,[26] in line with these residues forming an intrahelical loop in TM10, as suggested by the hPepT1 model. Thus, this model explains substrate-binding similarities between hPepT1 and DtpA and, in the absence of a crystal structure of hPepT1, can serve as a model for analyzing di/tripeptide and drug interactions.

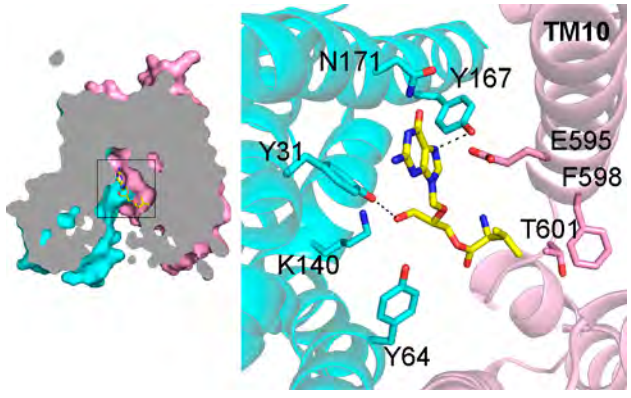


Figure 5. Ligand-binding site of the human PepT1 model with valganciclovir. hPepT1 model is colored in cyan (N-terminal MFS domain) and in pink (C-terminal domain). Binding site residues are shown in sticks and labeled, and potential hydrogen bonds are shown as dashes (distances ≤ 3.4 Å).

CONCLUDING REMARKS

We determined the ligand-free and valganciclovir-bound structures of DtpA, a bacterial SLC15 homologue with high similarity to hPepT1, not only in terms of sequence but also regarding substrate specificity and functional properties

DOI: 10.1021/jacs.8b11343
J. Am. Chem. Soc. 2019, 141, 2404−2412

relating to drug transport. We find that DtpA, in contrast to most characterized bacterial POTs to date, prefers to bind and transport tripeptides over dipeptides, which appears to relate to the presence of a characteristic intrahelical loop in TM10. The structure of DtpA in complex with valganciclovir shows that the prodrug binds in a similar position to di/tripeptides but with the guanine base instead of the valine residue mimicking the N-terminal residue of a di/tripeptide in the ligand-binding site. On the basis of homology modeling and a survey of published mutational data, we find that this binding mode most likely also applies to hPepT1. These novel insights into the structure and substrate binding of DtpA may facilitate future development of prodrugs with improved absorption rates for hPepT1, thereby lowering the pharmacologically effective dose and reducing side effects.

METHODS

Chemicals. Terrific broth (TB) was from Melford, isopropyl-β-D-thiogalactopyranoside (IPTG) and 1,4-dithiothreitol (DTT) were from Roth, DNase I was from AppliChem, complete EDTA-free protease inhibitor cocktail was from Roche, *n*-dodecyl-β-D-maltopyranoside (DDM) was from Anatrace, and crystallization reagents were from Molecular Dimensions. Amino acids, di/tripeptides, and drugs were obtained from Sigma-Aldrich, Bachem, and Fluka. The unnatural amino acid H-p-Bz-Phe-OH (pBpF) was purchased from Bachem.

Gene Construction, Protein Expression, and Purification of DtpA. The full-length *dtpA* (Uniprot ID: P77304) gene was amplified from the *Escherichia coli* genome and cloned into the ptH27 vector.[27] The construct has an N- and C-terminal His_6-tag and a TEV cleavage site after the N-terminal tag. Point mutations were generated by the blunt-end PCR method. DtpA, wild type (WT) and mutants, were expressed and purified as previously described.[28] Briefly, DtpA was overexpressed in *E. coli* C41(DE3) cells in TB medium. The cell pellet was resuspended in lysis buffer (20 mM NaP, pH 7.5, 300 mM NaCl, 5% glycerol, 15 mM imidazole) with 5 units/ml DNase I, 1 tablet of protease inhibitor/100 mL buffer, 1 mg/mL lysozyme, and 0.5 mM TCEP followed by cell lysis with an emulsifier (EmulsiFlex-C3, Avestin, three passages). The lysate was centrifuged 12 min at 10 000*g* and the supernatant centrifuged for 50 min at 95 000*g* (Optima XE-90, Beckman Coulter). The pellet containing the membrane fraction was solubilized in 1% DDM (Anatrace, sol-grade) for 1 h. The sample was centrifuged for 50 min at 70 000*g* and the supernatant applied to Ni-IMAC beads (ThermoFischer). After 45 min incubation on a rotating wheel, the suspension was transferred to a gravity column. Following two wash steps with lysis buffer supplied with 0.5 mM TCEP, 0.03% DDM, and 40 mM imidazole, DtpA was eluted with lysis buffer containing 0.5 mM TCEP, 0.03% DDM, and 300 mM imidazole. TEV cleavage was performed overnight during dialysis in gel filtration buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM DTT, 0.5% glycerol, and 0.03% DDM), and the protein was further purified on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare). All steps following cell lysis were performed at 4 °C. The protein was concentrated to 11 mg/mL using a 50 kDa cutoff concentrator (Corning Spin-X UF concentrators), flash frozen, and stored at −80 °C until further use.

Selection, Expression, and Purification of Nanobodies Against DtpA. For the generation of DtpA-specific nanobodies, two noninbred llamas were injected six times at weekly intervals with a mixture of 94 different proteins (50 μg of each antigen weekly). After 6 weeks of immunization, two separate phage display libraries were constructed, one from each animal, in the pMESy2 vector, which is a derivative of pMESy4 that contain a C-terminal EPEA-tag for affinity purification. After pooling both libraries, nanobodies were selected against individual antigens in two rounds of parallel panning in 96-well plates containing one immobilized antigen in each well. After two selection rounds on DtpA, 70 clones were picked for sequence analysis, 22 clones encoded antigen-specific nanobodies as tested in ELISA, grouping them in 16 different sequence families. A nanobody family is defined as a group of nanobodies with a high similarity in their CDR3 sequence (identical length and >80% sequence identity). Nanobodies from the same family derive from the same B-cell lineage and likely bind to the same epitope on the target. Immunizations, library construction, selection by panning, and nanobody characterization were performed according to standard procedures.[29] Four nanobodies were further characterized.

The nanobodies were expressed in *E. coli* WK6 cells and purified following standard procedures. Specifically, the cell pellet was resuspended in TES buffer (0.2 M TRIS, pH 8, 0.5 mM EDTA, 0.5 M sucrose) supplemented with one protease inhibitor tablet (Roche). Osmotic shock was performed by the addition of diluted TES buffer to release the periplasmic proteins. The solution was first centrifuged for 20 min at 10 000*g* and additionally for 30 min at 142 000*g*. The supernatant was applied to CaptureSelect beads (Thermo Fisher Scientific), which were equilibrated with wash buffer (20 mM NaP, pH 7.5, 20 mM NaCl). After three column volumes of washing, the nanobody was eluted with 20 mM HEPES, pH 7.5, 1.5 M $MgCl_2$. The nanobodies were further purified on a HiLoad 16/600 Superdex 75 pg column in 20 mM HEPES, pH 7.5, 150 mM NaCl, 5% glycerol, concentrated with a 5 kDa cutoff concentrator to 13 mg/mL, flash frozen, and stored at −80 °C until further use.

Analytical Gel Filtration Assay. The DtpA-nanobody complex formation was evaluated with an analytical gel filtration chromatography setup using a Superdex 200 5/150 GL home-packed column (GE Healthcare) attached to a 1260 Infinity liquid chromatography system (Agilent technologies). The size exclusion profiles of the samples containing both DtpA and nanobodies were compared to the separate profiles for DtpA and the nanobody (at the same concentrations). The samples were run in 20 mM NaP, pH 7.5, 150 mM NaCl, 5% glycerol, 0.5 mM TCEP, and 0.03% DDM. The concentration of DtpA was 0.2 mg/mL. All samples were run in duplicates at 10 °C following the UV absorption and the fluorescence (excitation at 280 nm and emission at 350 nm). The data analysis was performed in GraphPad Prism 5.0.

Evaluation of Conformational or Linear Epitope Binding. To determine whether the selected nanobodies are conformational or linear epitope binders, Western blot analysis was used (the principle of this approach is illustrated in Supplementary Figure 1a). SDS-PAGE using a 4−12% Bis-TRIS gel (ThermoFisher) under reducing conditions was performed with DtpA using a total amount of 0.3 and 0.03 μg per lane. On the same gel a positive control (N00, amount 15 μg) and a negative control (DtpC, another *E. coli* POT member with 0.3 and 0.03 μg) were included. The protein was transferred to a PVDF membrane (Biorad). One percent bovine serum albumin in TBS-T (TRIS buffer with Tween20, from Sigma) was used for blocking; TBS-T was also used as washing buffer. The membrane was incubated for 1 h with one of the nanobodies N87, N89, N93, or N00 containing an EPEA tag (1 μg/mL nanobody diluted in TBS-T buffer). Anti-EPEA antibody coupled to biotin (ThermoFisher) was used as primary and streptavidin coupled to horseradish peroxidase (ThermoFisher) as secondary antibody. The blot was developed using Super Signal West Pico Substrate (ThermoFisher) and Super Signal West Femto Substrate (ThermoFisher) in a 1:10 ratio. The image was acquired with a ChemiDoc MP device (BioRad).

Cross-Linking of DtpA and N00 in the Lipid Bilayer. In a cross-linking experiment we investigated if N00 could bind to DtpA in the native membrane. For this experiment two mutants were generated with an UV-inducible cross-linker in different positions. One mutant (G53TAG) was at the proximity of the N00 binding site on the periplasmic side of the transporter, and the second mutant (F265TAG) was on the cytoplasmic side of the transporter as a control. The amber suppression plasmid pEvol-pBpF and the DtpA expression plasmid carrying the amber mutation were cotransformed and overexpressed in the *E. coli* stain C41(DE3). The cross-linking protocol from Farrell et al. was followed.[30] Shortly, DtpA mutants were overexpressed in medium containing 1 mM pBpF and the lysed cells were incubated with N00. Cross-linking was performed with UV light (365 nm, 60 min), and the DtpA mutant was purified as described above. Sodium dodecyl sulfate polyacrylamide gel electro-

DOI: 10.1021/jacs.8b11343
J. Am. Chem. Soc. 2019, 141, 2404−2412

phoresis (SDS-PAGE) using a 4−12% Bis-TRIS gel (ThermoFisher) was performed and transferred on a PVDF membrane (Biorad). Two percent bovine serum albumin in TBS-T (TRIS buffer with Tween20, from Sigma) was used for blocking, TBS-T was used as washing buffer. The membrane was incubated with HisProbe-HRP conjugate antibody (ThermoFisher) for 1 h. The blot was developed using Super Signal West Pico Substrate (ThermoFisher) and Super Signal West Femto Substrate (ThermoFisher) in a 1:10 ratio. The image was acquired with a ChemiDoc MP device (BioRad).

Crystallization and Structure Determination. DtpA and N00 were mixed in 1:1.2 molar ratio 1 h prior to crystallization and incubated at 4 °C. The final concentration of DtpA was 8.5 mg/mL. Crystallization plates were prepared with an automated liquid handler (mosquito, TTP Labtech) using the sitting drop vapor diffusion technique with a final drop volume of 300 nL (at 1:1, 1:2 and 2:1 (v/v) ratios of protein to precipitant). The initial crystals of the ligand-free structure were obtained from the MemGold2 crystallization screen (Molecular Dimensions) at room temperature. After several rounds of optimization, the best diffracting crystals grew in 0.02 M magnesium chloride, 0.1 M MES at pH 6.5, and 32% PEG 600. Diffraction data were collected at the MASSIF-1 (ID30A-1) beamline at ESRF (Grenoble, France)[31] using the fixed 0.969 Å wavelength at 100 K temperature. The data were processed using the XDS package.[32] The space group was identified as $P2_12_12_1$, and the resolution limit was at 3.29 Å. Initial phases were obtained by molecular replacement using DtpD (PDB ID 4Q65) and a nanobody structure as search models in Phaser[33] as part of PHENIX.[34] The model was further built manually in Coot[35] and refined in PHENIX in iterative cycles. The Ramachandran statistics for the final model are 97.3% for favored regions, 2.7% for allowed regions, and 0.0% for outliers. In the case of the DtpA-N00 structure bound to valganciclovir, the drug was added to a DtpA solution in powder form (estimated concentration of 20 mM) and N00 was added as described above. The workflow for structure determination of the drug-bound complex was the same as for the ligand-free form, but the best diffracting crystals grew in the following condition: 0.1 M glycine, pH 9.0, 35% PEG 400, 0.15 M $CaCl_2$, and 0.02% Anapoe-$C_{12}E_{10}$. The data set was collected at beamline P13 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany)[36] at 0.966 Å wavelength and 100 K temperature. The ligand-free structure was used as a search model for molecular replacement. The Ramachandran statistics were the following: 98.1% favored, 1.9% allowed, and 0.0% for outliers. A second ligand-free structure was determined from crystals derived in a similar condition as the complex structure to validate that the additional density originates from the added compound and not from a buffer or precipitant molecule. Crystals for the second ligand-free form grew in in 0.1 M glycine, pH 9.0, 0.15 M $CaCl_2$, and 35% PEG 400, and the data were collected at beamline P14 at PETRA III at 0.976 Å wavelength and 100 K temperature and processed as described for the valganciclovir-bound structure. The Ramachandran statistics for the final model are 96.6% for favored regions, 3.4% for allowed regions, and 0.0% for outliers. Structure figures were generated using PyMOL.[37]

Analysis of the Binding Site Volume. The cavity introduced by the TM10 intrahelical loop and the shift of TM11 in DtpA was investigated by comparing the cavity volumes of all available POT structures. The volume calculation was performed with POVME 2.0[21] using the default parameters. All structures were overlaid, and the coordinates of the Cα atom of the valine moiety in valganciclovir were selected as the center. The volume of the binding site based on a 4 and 7 Å radius sphere was calculated for the following POT structures: DtpA (6GS4), $PepT_{So2}$ (4LEP), DtpD (4Q65), YePepT (4W6V), GkPOT (4IKV), $PepT_{Xc}$ (6EI3), $PepT_{So}$ (4UVM), $PepT_{St}$ (5OXN), and $PepT_{Sh}$ (6EXS). The results were analyzed in GraphPad Prism 5.0 (GraphPad Software, CA, USA).

Thermal Stability Assay. The differential scanning fluorimetry (DSF) method was used to follow the thermal unfolding event of DtpA with a Prometheus NT.48 device (NanoTemper Technologies, Munich, Germany). Here, the fluorescence at 330 and 350 nm is recorded over a temperature gradient scan. The temperature at the transition point of the fluorescence ratio from 330 to 350 nm corresponds to the melting temperature (T_m). The shift of T_m in the presence of a nanobody or a ligand is interpreted as potential binding. DtpA was diluted to 0.5 mg/mL with assay buffer (100 mM HEPES, pH 7.5, 150 mM NaCl, 0.03% DDM). All ligand lyophilizates were dissolved in water to a final concentration of 50 mM. An 18 μL amount of DtpA at 0.5 mg/mL was mixed with 2 μL of 50 mM ligand/substrate (5 mM final concentration). In the case of the nanobodies, the same concentration of DtpA was mixed with a 1:1.2 molar ratio of the tested nanobody. Samples were incubated for 10 min at room temperature before loading them with standard capillaries into the Prometheus device. The excitation power was set between 15% and 25%, and the tested temperature range was from 20 to 85 °C. All measurements were done in triplicate, and the results were analyzed with Excel (Microsoft) and GraphPad Prism 5.0 (GraphPad Software, CA, USA).

Binding Affinity Assay. The binding affinity of DtpA and the di/tripeptides were measured with a label-free microscale thermophoresis device Monolith (NanoTemper Technologies). However, due to the intrinsic signal from valganciclovir and valacyclovir another approach was taken for these drugs. DtpA was labeled on the His_6-tag with RED-tris-NTA dye (NanoTemper), and the binding affinity was measured with microscale thermophoresis following the RED-dye signal. To compare the results from unlabeled and labeled DtpA, the binding affinity of the tripeptide AFA was measured in both systems.

For the measurements on unlabeled DtpA, the protein was diluted to 0.5 μM with assay buffer (100 mM HEPES, pH 7.5, 150 mM NaCl, 0.03% DDM). A dilution series of the substrate was also prepared with the same buffer. DtpA was then mixed with the substrate dilution series and incubated for 10 min at room temperature before loading the samples in standard capillaries to the MST LabelFree device (NanoTemper Technologies, Munich, Germany). For the measurements with the DtpA-N00 complex, the concentration of DtpA was kept constant and a 1:1.2 molar ratio of N00 was mixed 1 h prior to the measurements. As a negative control, the binding affinity of N00 alone, at 2 μM concentration, to selected peptides was measured under the same conditions. The settings were 20% LED and 20% MST power. For the measurements with valganciclovir and valacyclovir, DtpA was diluted to 1.6 μM and labeled with RED-tris-NTA dye according to the supplier's protocol but using the assay buffer in all steps. The MST NT.115 device (NanoTemper Technologies, Munich, Germany) was used with the standard capillaries and 20% LED and 20% MST power settings. All measurements were performed in duplicate, and the results were analyzed with GraphPad Prism 5.0.

In Vivo Uptake and Competition Assay. The *in vivo* uptake of the β-Ala-Lys peptide coupled to the fluorescent AMCA (7-amino-4-methylcoumarin-3-yl)acetic acid) group (AK-AMCA) was performed as previously described with minor modifications.[4] *E. coli* C41(DE3) cells containing the DtpA plasmid were grown in 5 mL of LB medium supplemented with 100 μg/mL ampicillin to an $OD_{600\ nm}$ of 0.7. DtpA expression was induced by the addition of 0.2 mM IPTG and followed by an incubation period of 3 h at 37 °C. The cell culture was then centrifuged for 3 min at 3000 rpm and resuspended to a final $OD_{600\ nm}$ of 10 in the assay buffer (100 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM glucose). 40 μL of buffer, 40 μL of cells, 10 μL of AK-AMCA (final concentration 50 μM), and 10 μL of the competing ligand (final concentration 5 or 0.5 mM) were mixed in a 96-well plate and incubated for 20 min at 37 °C. The reaction was stopped by adding 200 μL of ice-cold assay buffer, and the cells were then washed twice with the same buffer. Finally, the cells were resuspended in 200 μL of assay buffer, and the fluorescence was measured in a M1000 microplate reader (TECAN) with excitation at 350 nm and emission at 450 nm. All experiments were performed in triplicate. The results were normalized by the fluorescence value of the control (cells overexpressing DtpA incubated with AK-AMCA) and plotted as AK-AMCA uptake rate percentage. The analysis was done in Excel and GraphPad Prism 5.0.

Principal Component Analysis. The physicochemical properties of di- and tripeptides investigated in this study were obtained from the

DOI: 10.1021/jacs.8b11343
J. Am. Chem. Soc. 2019, 141, 2404−2412

ChemSpider database[38] using experimental or predicted physicochemical properties (PhysChem Module of the ACD/Laboratories Percepta Platform; Advanced Chemistry Development Inc., Toronto, Canada). A statistical analysis of the data set was performed in the R Statistical Software Environment.[39] An exploratory statistical analysis by principal component analysis (PCA) was carried out using the R package FactoMineR[40] to investigate potential relationships between physicochemical properties of di- and tripeptide ligands and the two experimentally determined biophysical parameters AK-AMCA competition and thermal stability. Results were visualized with factoextra[41] and ggplot2.[42] Prior to the PCA, the data were filtered for zero variance physicochemical properties. A selection of the 10 most highly contributing properties used for PCA were plotted on a biplot together with the supporting experimentally determined parameters AK-AMCA competition rate and thermal stability that were not used for PCA. Additionally, nonparametric statistical testing was performed to compare the experimentally determined biophysical parameters AK-AMCA competition rate and thermal stability for the groups of di- and tripeptides.

Model Building of Human PepT1 (hPepT1, SLC15A1) with Valganciclovir. The homology model of hPepT1 (Uniprot ID P46059) transmembrane domain was constructed using MODELLER[43] based on the DtpA structure as the modeling template. The initial query-template alignment for the homology model was created with the HH-suite.[44] The alignment was subsequently adjusted manually using the Swiss-PdbViewer[45] according to predictions of secondary structure and membrane helices from GeneSilico MetaServer[46] and to minimize distortions around loop regions. The regions including the ECD domain[47] as well as the eukaryotic linker between TM6 and TM7 were excluded from modeling.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b11343.

Nanobody screening for DtpA binding; electron density maps of DtpA; DtpA-N00 binding site; stereoimage of the ligand-binding site with valganciclovir; AK-AMCA uptake and concentration-dependent competition assay with DtpA; principal component analysis of physicochemical data for di- and tripeptide ligands; binding curves derived from microscale thermophoresis experiment for DtpA and the DtpA-N00 complex; cross-linking of DtpA and N00 in a lipid environment; thermal stability and binding affinity analysis of DtpA mutants; sequence alignment of DtpA and hPepT1; ligand-binding site of DtpA and hPepT1 with valganciclovir; key residues for function in TM5, TM7, and TM10 highlighted in the hPepT1 homology model; thermal stability (T_m), binding affinity (K_d), and AK-AMCA uptake results for DtpA and DtpA-N00; binding affinity results for DtpA mutants; references (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*christian.loew@embl-hamburg.de

ORCID

Jan Strauss: 0000-0002-6208-791X

Christian Löw: 0000-0003-0764-7483

Notes

The authors declare no competing financial interest.

Atomic coordinates and structure factors were deposited in the protein data bank with the following accession numbers: ligand-free DtpA-N00 structure in MES buffer, 6GS1; DtpA-N00 structure in glycine buffer, 6GS7; and DtpA-N00-valganciclovir structure, 6GS4. The homology model of hPepT1 was deposited in the Model Archive (https://www.modelarchive.org/) and the relevant record can be found at https://dx.doi.org/10.5452/ma-ajiw9.

■ ACKNOWLEDGMENTS

We thank the Sample Preparation and Characterization facility of EMBL Hamburg for support with nanoDSF measurements, MST, and protein crystallization. We would like to thank the group of Thomas R. Schneider at EMBL Hamburg for access to the EMBL beamlines P13 and P14 as well as the staff at ID30A-1/MASSIF-1 at the European Synchrotron Radiation Facility. We acknowledge the support and the use of resources of Instruct-ERIC, part of the European Strategy Forum on Research Infrastructures (ESFRI), and the Research Foundation-Flanders (FWO) for their support to the nanobody discovery. We thank Saif Saifuzzaman and Alison Lundqvist for technical assistance during nanobody discovery. We want to acknowledge Pär Nordlund from the Karolinska Institutet for his support in the beginning of the project. This work was funded by the Horizon2020 program of the European Union (iNEXT grant, project no. 653706), the Swedish Research Council (grant no. 621-2013-5905), and a grant from the German-Israeli Foundation (GIF grant no: G-1288-207.9/2015). J.S. was additionally supported by a fellowship from the EMBL Interdisciplinary Postdoc (EIPOD) programme under Marie Sklodowska-Curie Actions COFUND programme (grant no. 664726).

■ REFERENCES

(1) Brandsch, M.; Knütter, I.; Bosse-Doenecke, E. Pharmaceutical and Pharmacological Importance of Peptide Transporters. *J. Pharm. Pharmacol.* **2008**, *60*, 543.

(2) Jung, D.; Dorr, A. Single-Dose Pharmacokinetics of Valganciclovir in HIV- and CMV-Seropositive Subjects. *J. Clin. Pharmacol.* **1999**, *39*, 800.

(3) Beauchamp, L. M.; Orr, G. F.; de Miranda, P.; Bumette, T.; Krenitsky, T. A. Amino Acid Ester Prodrugs of Acyclovir. *Antivir. Chem. Chemother.* **1992**, *3*, 157.

(4) Weitz, D.; Harder, D.; Casagrande, F.; Fotiadis, D.; Obrdlik, P.; Kelety, B.; Daniel, H. Functional and Structural Characterization of a Prokaryotic Peptide Transporter with Features Similar to Mammalian PEPT1. *J. Biol. Chem.* **2007**, *282*, 2832.

(5) Prabhala, B. K.; Aduri, N. G.; Iqbal, M.; Rahman, M.; Gajhede, M.; Hansen, P. R.; Mirza, O. Several hPepT1-Transported Drugs are Substrates of the Escherichia coli Proton-Coupled Oligopeptide Transporter YdgR. *Res. Microbiol.* **2017**, *168*, 443.

(6) Solcan, N.; Kwok, J.; Fowler, P. W.; Cameron, A. D.; Drew, D.; Iwata, S.; Newstead, S. Alternating access mechanism in the POT family of oligopeptide transporters. *EMBO J.* **2012**, *31*, 3411.

(7) Doki, S.; Kato, H. E.; Solcan, N.; Iwaki, M.; Koyama, M.; Hattori, M.; Iwase, N.; Tsukazaki, T.; Sugita, Y.; Kandori, H.; Newstead, S.; Ishitani, R.; Nureki, O. Structural Basis for Dynamic Mechanism of Proton-Coupled Symport by the Peptide Transporter POT. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 11343.

(8) Boggavarapu, R.; Jeckelmann, J.-M.; Harder, D.; Ucurum, Z.; Fotiadis, D. Role of Electrostatic Interactions for Ligand Recognition and Specificity of Peptide Transporters. *BMC Biol.* **2015**, *13*, 58.

(9) Martinez Molledo, M.; Quistgaard, E. M.; Flayhan, A.; Pieprzyk, J.; Löw, C. Multispecific Substrate Recognition in a Proton-Dependent Oligopeptide Transporter. *Structure* **2018**, *26*, 467.

(10) Guettou, F.; Quistgaard, E. M.; Raba, M.; Moberg, P.; Löw, C.; Nordlund, P. Selectivity mechanism of a bacterial homolog of the human drug-peptide transporters PepT1 and PepT2. *Nat. Struct. Mol. Biol.* **2014**, *21*, 728.

(11) Quistgaard, E. M.; Martinez Molledo, M.; Löw, C. Structure Determination of a Major Facilitator Peptide Transporter: Inward Facing PepTSt from Streptococcus thermophilus Crystallized in Space Group P3121. *PLoS One* **2017**, *12*, No. e0173126.
(12) Minhas, G. S.; Bawdon, D.; Herman, R.; Rudden, M.; Stone, A. P.; James, A. G.; Thomas, G. H.; Newstead, S. Structural Basis of Malodour Precursor Transport in the Human Axilla. *eLife* **2018**, *7*, e34995.
(13) Parker, J. L.; Li, C.; Brinth, A.; Wang, Z.; Vogeley, L.; Solcan, N.; Ledderboge-Vucinic, G.; Swanson, J. M. J.; Caffrey, M.; Voth, G. A.; Newstead, S. Proton Movement and Coupling in the POT Family of Peptide Transporters. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 13182.
(14) Yan, N. Structural Biology of the Major Facilitator Superfamily (MFS) Transporters. *Annu. Rev. Biophys.* **2015**, *44*, 257.
(15) Sugawara, M.; Huang, W.; Fei, Y. J.; Leibach, F. H.; Ganapathy, V.; Ganapathy, M. E. Transport of Valganciclovir, a Ganciclovir Prodrug, via Peptide Transporters PEPT1 and PEPT2. *J. Pharm. Sci.* **2000**, *89*, 781.
(16) Lyons, J. A.; Parker, J. L.; Solcan, N.; Brinth, A.; Li, D.; Shah, S. T.; Caffrey, M.; Newstead, S. Structural Basis for Polyspecificity in the POT Family of Proton-Coupled Oligopeptide Transporters. *EMBO Rep.* **2014**, *15*, 886.
(17) Martinez Molledo, M.; Quistgaard, E. M.; Löw, C. Tripeptide Binding in a Proton-Dependent Oligopeptide Transporter. *FEBS Lett.* **2018**, *592*, 3239.
(18) Guettou, F.; Quistgaard, E. M.; Trésaugues, L.; Moberg, P.; Jegerschöld, C.; Zhu, L.; Jong, A. J. O.; Nordlund, P.; Löw, C. Structural Insights into Substrate Recognition in Proton-Dependent OligoPeptide Transporters. *EMBO Rep.* **2013**, *14*, 804.
(19) Samsudin, F.; Parker, J. L.; Sansom, M. S. P.; Newstead, S.; Fowler, P. W. Accurate Prediction of Ligand Affinities for a Proton-Dependent Oligopeptide Transporter. *Cell Chem. Biol.* **2016**, *23*, 299.
(20) Durrant, J. D.; de Oliveira, C. A. F.; McCammon, J. A. POVME: an Algorithm for Measuring Binding-Pocket Volumes. *J. Mol. Graphics Modell.* **2011**, *29*, 773.
(21) Durrant, J. D.; Votapka, L.; Sørensen, J.; Amaro, R. E. POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J. Chem. Theory Comput.* **2014**, *10*, 5047.
(22) Clémençon, B.; Lüscher, B. P.; Hediger, M. A. Establishment of a Novel Microscale Thermophoresis Ligand-Binding Assay for Characterization of SLC Solute Carriers Using Oligopeptide Transporter PepT1 (SLC15 Family) as a Model System. *J. Pharmacol. Toxicol. Methods* **2018**, *92*, 67.
(23) Malle, E.; Zhou, H.; Neuhold, J.; Spitzenberger, B.; Klepsch, F.; Pollak, T.; Bergner, O.; Ecker, G. F.; Stolt-Bergner, P. C. Random Mutagenesis of the Prokaryotic Peptide Transporter YdgR Identifies Potential Periplasmic Gating Residues. *J. Biol. Chem.* **2011**, *286*, 23121.
(24) Kulkarni, A. A.; Haworth, I. S.; Lee, V. H. L. Transmembrane Segment 5 of the Dipeptide Transporter hPepT1 Forms a Part of the Substrate Translocation Pathway. *Biochem. Biophys. Res. Commun.* **2003**, *306*, 177.
(25) Kulkarni, A. A.; Haworth, I. S.; Uchiyama, T.; Lee, V. H. L. Analysis of Transmembrane Segment 7 of the Dipeptide Transporter hPepT1 by Cysteine-scanning Mutagenesis. *J. Biol. Chem.* **2003**, *278*, 51833.
(26) Xu, L.; Haworth, I. S.; Kulkarni, A. A.; Bolger, M. B.; Davies, D. L. Mutagenesis and Cysteine Scanning of Transmembrane Domain 10 of the Human Dipeptide Transporter. *Pharm. Res.* **2009**, *26*, 2358.
(27) Woestenenk, E. A.; Hammarström, M.; van den Berg, S.; Härd, T.; Berglund, H. His tag Effect on Solubility of Human Proteins Produced in Escherichia coli: a Comparison between Four Expression Vectors. *J. Struct. Funct. Genomics* **2004**, *5*, 217.
(28) Flayhan, A.; Mertens, H. D. T.; Ural-Blimke, Y.; Martinez Molledo, M.; Svergun, D. I.; Löw, C. Saposin Lipid Nanoparticles: A Highly Versatile and Modular Tool for Membrane Protein Research. *Structure* **2018**, *26*, 345.
(29) Pardon, E.; Laeremans, T.; Triest, S.; Rasmussen, S.; Wohlkönig, A.; Ruf, A.; Muyldermans, S.; Hol, W. G. J.; Kobilka, B. K.; Steyaert, J. A General Protocol for the Generation of Nanobodies for Structural Biology. *Nat. Protoc.* **2014**, *9*, 674.
(30) Farrell, I. S.; Toroney, R.; Hazen, J. L.; Mehl, R. A.; Chin, J. W. Photo-Cross-Linking Interacting Proteins with a Genetically Encoded Benzophenone. *Nat. Methods* **2005**, *2*, 377.
(31) Bowler, M. W.; Nurizzo, D.; Barrett, R.; Beteva, A.; Bodin, M.; Caserotto, H.; Delagenière, S.; Dobias, F.; Flot, D.; Giraud, T.; Guichard, N.; Guijarro, M.; Lentini, M.; Leonard, G. A.; McSweeney, S.; Oskarsson, M.; Schmidt, W.; Snigirev, A.; von Stetten, D.; Surr, J.; Svensson, O.; Theveneau, P.; Mueller-Dieckmann, C. MASSIF-1: a Beamline Dedicated to the Fully Automatic Characterization and Data Collection from Crystals of Biological Macromolecules. *J. Synchrotron Radiat.* **2015**, *22*, 1540.
(32) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125.
(33) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658.
(34) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213.
(35) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126.
(36) Cianci, M.; Bourenkov, G.; Pompidor, G.; Karpics, I.; Kallio, J.; Bento, I.; Roessle, M.; Cipriani, F.; Fiedler, S.; Schneider, T. R. P13, the EMBL Macromolecular Crystallography Beamline at the Low-Emittance PETRA III Ring for High- and Low-Energy Phasing with Variable Beam Focusing. *J. Synchrotron Radiat.* **2017**, *24*, 323.
(37) Delano, W. L. *PyMOL Molecular Graphics System*; DeLano Scientific: Palo Alto, CA, USA, 2002; http://www.pymol.org.
(38) Pence, H. E.; Williams, A. ChemSpider: An Online Chemical Information Resource. *J. Chem. Educ.* **2010**, *87*, 1123.
(39) *R: A language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria. 2018.
(40) Lê, S.; Josse, J.; Husson, F. FactoMineR: an R Package for Multivariate Analysis *J. Stat. Soft.* **2008**, *25*. DOI: 10.18637/jss.v025.i01
(41) Kassambara, A.; Mundt, F. Factoextra: Extract and visualize the results of multivariate data analyses. *R package*, version 1.0.3; 2015.
(42) Wickham, H. *ggplot2*; Springer, 2009.
(43) Sali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779.
(44) Zimmermann, L.; Stephens, A.; Nam, S.-Z.; Rau, D.; Kübler, J.; Lozajic, M.; Gabler, F.; Söding, J.; Lupas, A. N.; Alva, V. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **2018**, *430*, 2237.
(45) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an Environment for Comparative Protein Modeling. *Electrophoresis* **1997**, *18*, 2714.
(46) Kurowski, M. A.; Bujnicki, J. M. GeneSilico Protein Structure Prediction Meta-Server. *Nucleic Acids Res.* **2003**, *31*, 3305.
(47) Beale, J. H.; Parker, J. L.; Samsudin, F.; Barrett, A. L.; Senan, A.; Bird, L. E.; Scott, D.; Owens, R. J.; Sansom, M. S. P.; Tucker, S. J.; Meredith, D.; Fowler, P. W.; Newstead, S. Crystal Structures of the Extracellular Domain from PepT1 and PepT2 Provide Novel Insights into Mammalian Peptide Transport. *Structure* **2015**, *23*, 1889.

DOI: 10.1021/jacs.8b11343
J. Am. Chem. Soc. 2019, 141, 2404−2412

6. Discussion

In this section of the Thesis, I will put my work in the context of the recent advances in molecular modeling. In the first part (see Discussion section 6.1), I will discuss the major limitations of Assembline as an integrative structural modeling method and propose future improvements that could enable wider adoption of such modeling methods by researchers in the field. The potential of integrative modeling for the structural elucidation of large protein assemblies will also be briefly discussed with respect to novel AI-assisted modeling methods.

In the second part (see Discussion section 6.2), my findings from the production of the integrative NPC models will be discussed with respect to our mechanistic understanding of NPCs. Since the models were built based on low resolution datasets, their biological relevance and future research scopes are also highlighted.

6.1 Challenges in integrative modeling and future Assembline applications

Recent hardware and software-related advances in some of the most used experimental methods for protein structure determination pushed the boundaries for the development of novel modeling protocols. Integrative structural modeling is such an approach and is considered a state-of-the-art methodology to obtain structures of very complex protein assemblies by combining information from complementary techniques. The increased application of integrative modeling workflows for various protein targets has led so far to a dramatic increase of integrative models deposited in publicly available structural databases.

Despite the fact that integrative modeling is the "go-to" solution, especially for complex protein architectures, such modeling workflows are usually multi-step processes that demand a number of informed decisions by the user (see Introduction section 3.2). The conglomerate of these user-based decisions before and during modeling can often lead to significantly different modeling outputs. Currently, neither Assembline nor any other existing integrative modeling solution offer "golden" standards for the aforementioned decisions therefore, the users are still often reaching the desired modeling results by "trial and error".

The first and most important modeling decision that can influence dramatically the modeling performance of integrative methods like Assembline is the definition of the system architecture and selection of a proper representation. Each protein system in this context can be represented in an atomistic (or near-atomistic) or coarse-grained manner, depending on the types and quality of the input datasets as well as on available computational resources for the calculation of the derived spatial restraints (see Introduction section 3.2). As it becomes apparent, selection of system representations at "high resolution", especially in cases of large protein complexes, poses the usual modeling "dangers" of e.g. data overinterpretation and costly computations. So far, the system representation resolution is selected based on prior experience and subjective user judgment as there are currently no objective methods for assessing the suitability of a specific representation.

Assembline[120] offers a variety of options for defining the system representation together with the GUI of Xlink Analyzer[127]. Although these options are certainly helping less experienced researchers to prime their modeling projects, the main issue regarding the subjectivity of system representation remains. Different approaches to circumvent the subjectivity of this step have already been proposed by researchers in the field. One such approach is the selection of optimal representation with respect to a set of predefined representations based on available system information and data uncertainty as described by[141]. This approach is already implemented in IMP[111] (remember that the algorithmic "core" of Assembline is IMP) and it could be one of the main approaches for selection of system representation offered in future Assembline releases. Additionally, I believe that the continuous development of software and hardware will lead to such an abundance of computational resources that even for very large protein architectures only the atomic representation would be utilized for modeling.

The data availability regarding a modeling target is a crucial factor for selecting the appropriate system representation, and any "inherent" uncertainties should be taken into consideration. However, even in cases when there are multiple orthogonal datasets available regarding a protein complex, most of the current integrative modeling methods cannot utilize them all due to the limited variety of modeling restraints they offer. Thanks to the functionalities of IMP, Assembline is offering a wider selection of predefined spatial restraints which can easily be extended by more advanced users. Even in Assembline though, not every data type can be utilized for modeling, with such an example being SAXS datasets which cannot be represented as distance restraints. Nevertheless, restraints for datasets from methods like SAXS are available in IMP and incorporation of such restraints is already under consideration for future Assembline releases.

In the case of Assembline, the availability of certain types of data can also influence the selection of modeling steps (different modeling modes offered by Assembline, see Synopsis section 5.3) and consequently the overall performance of the modeling. For example, the generation of fit libraries step, which will provide the main input for the global optimization step, can only be performed when cryo-EM density maps are available for the modeling target. In such cases, as described earlier, the modeling process is significantly faster compared to standard IMP modeling protocols due to the precalculated fits of the complex subunits. This implementation of precalculated rigid body fits could be optimized even further in the context of efficient sampling of fit space and increased sensitivity for "good" fits. One such optimization approach that takes advantage of graphics processing units (GPUs) together with entropy-based scores has been developed by Kai Karius (Kosinski group, EMBL Hamburg) and is going to be integrated into the next version of Assembline. Additionally, the main integrative modeling steps in Assembline, i.e. global optimization, recombination and refinement, are all based on a Monte Carlo Simulated Annealing algorithm for sampling the solution space. As it becomes apparent, these steps could also be optimized further with the implementation of more advanced conformational sampling algorithms or modern variations of MCMC, such as a Replica Exchange Monte Carlo[142]. Taken together, these improvements would significantly reduce the computation time of the complete integrative modeling process with Assembline and increase the quality of the output model ensemble even for very complex protein targets.

As described earlier (see Introduction section 3.2), most integrative modeling protocols, including Assembline, produce ensembles of output models instead of a single solution. In this context, each modeling solution is scored during modeling based on a predefined scoring function. This scoring function is usually a linear combination of restraint scoring terms that are weighted in order to adjust their modeling influence. Assembline is also based on such scoring functions that can and should be edited (i.e. redefined) in a case-specific manner since we are not able to provide a universal and objective method for selecting the most suitable scoring function for every modeling case. In order to tackle this issue, several approaches have been proposed by researchers in the field, with one of the most prevalent being the implementation of a Bayesian scoring function as described by[143], which allows for a more objective optimization of weights. This scoring function could also be available among other options in the next version of Assembline as this solution is already included in the most recent versions of IMP.

The final stage of every integrative modeling protocol is as mentioned earlier, the assessment of sampling performance (see Introduction section 3.2). Based on the output from this assessment, better-informed decisions can be made regarding e.g. the reiteration of modeling for further model refinement. There is currently no standard method for quantifying the quality of structural sampling, though the tool imp-sampcon[99] provided by IMP is capable of assessing the sampling exhaustiveness and convergence of the modeling runs. The modeling performance of Assembline can also be assessed with imp-sampcon[99] thanks to the algorithmic interface we are providing between the two packages (i.e. Assembline and imp-sampcon from IMP). Assembline though includes even further modeling analysis steps for each key stage of the multi-step procedure. Despite the fact that imp-sampcon is the main performance assessment tool for Assembline, we are aware of the limitations it poses when applied to models of more complex protein systems. Such a limitation is the danger of precision over-estimation caused by exhaustive sampling at a local minimum of the scoring "landscape" due to the MCMC algorithm. Nevertheless, imp-sampcon is still under active development by the authors and updating Assembline to take into account the improved imp-sampcon versions that address these limitations will be a straightforward process.

Around the time that the development and release of Assembline were completed, two new structural modeling methods became publicly available. The new tools, called AlphaFold[144] and RoseTTAfold[145] respectively, are both machine learning-based approaches (a specific type of Artificial Intelligence, i.e. AI) that include sophisticated neural network architectures. An essential element of every typical machine learning framework is the training stage which should be performed with extreme caution in order to avoid model overfitting for example. Hence, the network architectures of both tools were rigorously trained based on the available protein structures from PDB and are able to produce models based on multiple sequence alignments and structural templates. So far, AlphaFold has proven to be the most successful method judging by the groundbreaking structural predictions in the context of the 14th Critical Assessment of protein Structure Prediction (CASP14)[146] and the wider adoption by researchers. AlphaFold and RoseTTAfold are able to produce protein models starting solely from the input amino acid sequence in relatively short running times, depending on the available computational resources (a couple of hours on average). Furthermore, this tool has already been used to produce models of

entire proteomes, including the human[147]. Despite the outstanding performance of AlphaFold in structural modeling of individual proteins or small complexes, the method is still far from considered complete for the modeling of large and heterogeneous assemblies. By the time this Thesis was being compiled, continuous improvements on AlphaFold by the developers and other expert research groups[148] were being implemented with regards to protein complex modeling. These developments together with the need for accurate predictions of inter-protein interactions gave rise to AlphaFold-Multimer, a version of AlphaFold specialized in structure predictions of relatively simple protein complexes[149]. At the current stage though, AlphaFold is unable to produce accurate models of large complexes due to reasons including the dependence on available GPU memory resources and inability of AlphaFold to optimize proteins or complexes with more than around 3,000-4,000 residues. However, it is apparent that in case AlphaFold in the near future achieves comparable modeling accuracy for monomers and complex multi-mers, all structural studies will be accelerated significantly.

Assembline was recently successfully utilized together with AlphaFold, by researchers in Kosinski group (EMBL Hamburg), in order to produce the most complete and accurate integrative structural models of the human NPC[150]. In this context, AlphaFold served as the main modeling engine for individual Nups and smaller Nup subcomplexes which were then subjected all together to integrative modeling with Assembline. These studies highlighted that the application of AlphaFold, as a starting point for predicting accurately the structure and interactions of rigid bodies, together with Assembline could be the preferred solution by researchers in the field for producing models of very complex structural architectures.

All in all, despite the subjective decisions that need to be made by users of integrative modeling methods, such approaches are still the "go-to" solution for producing models of complex protein systems. From the existing integrative modeling tools, Assembline has proven to be one of the most efficient and versatile, especially in modeling cases of large complexes such as the NPCs. The novel AI-based methods for structure predictions such as AlphaFold and RoseTTAfold that emerged recently, appear to have already increased the impact and significance of Assembline in the field of integrative structural modeling, meaning that less experienced (algorithmically) researchers would now be enabled to study even the most intriguing of protein structure architectures.

6.2 The first in-cell NPC models and their impact on future structural studies

The NPCs are probably the largest proteinaceous assemblies in eukaryotes playing a crucial role in several cellular processes. Determining the structure of NPCs at atomic resolution will pave the way for better understanding of functions such as nucleocytoplasmic cargo trafficking and viral invasion through the nuclear pores. The task of structural elucidation of the NPC architecture proved extremely challenging over the years due to the dimensions, transmembrane nature and heterogeneity of such complexes.

During my PhD research, I produced integrative structural models of NPCs from different yeast species. Additionally, my modeling efforts included the production of NPC models based on datasets from different cellular growth conditions (see Synopsis sections 5.1 and 5.2). All the aforementioned models were considered to be the first in-cell NPC models since the integrative modeling approach that was applied was mainly based on in-cell acquired cryo-EM datasets.

In the case of *S. cerevisiae*, our collaborative experimental and modeling efforts revised the previous "status quo" NPC models which were produced by combining in vitro datasets such as cryo-ET maps from detergent-extracted ScNPCs, atomic models, crosslinks and more[46]. Our in-cell ScNPC models resolved at ~25 Å, which were built with Assembline[120] based mostly on cryo-ET datasets, appeared dilated by ~20 nm compared to the previous in vitro models, a fact that led us to question the biological relevance of the latter. Accordingly, major NPC scaffold subcomplexes such as the Y-complexes exhibited a more "relaxed" conformation with less tail and stems curvature. Despite the fact that the architecture of our in-cell Y-complexes explained very well the in-cell-based observed density, with a single CR and NR Y-complex per spoke, new structural research concerning the structural variability of ScNPCs suggested the existence of dimerized NR Y-complexes in some of the observed nuclear pores[151]. Therefore, further microscopy and/or mass spectrometry-based studies should be conducted in the future in order to provide more detailed answers about scaffold Nup stoichiometries under different cellular conditions.

Moreover, my ScNPC modeling work highlighted the proper placement and orientation of the mRNA export platform machinery, i.e. P-complex, in wild type NPCs. The in-cell ScNPC models revealed that the in-cell orientation of the P-complex appeared rotated by 180° compared to the previous in vitro model, which was solely based on cross-linking datasets (as the features of this complex were completely absent from the in vitro NPC density). Our proposed orientation was much more in line with what was already suggested about the spatiotemporal sequence of molecular events with regards to remodeling of mRNPs during export. This striking P-complex orientation was revealed by utilizing Assembline together with the corresponding integrative model from the in vitro NPC model[46]. Therefore, AI-based modeling methods such as AlphaFold could be utilized in order to produce a more accurate and extended (near-complete) architecture of the mRNA export platform which would definitely provide a more detailed mechanistic understanding of the cytoplasmic transport steps.

As mentioned earlier, I first produced integrative models of NPCs from wild type cells growing under native conditions followed by the production of models from knock-out cells growing under different temperatures (see Synopsis section 5.1). The wild type in-cell ScNPC model was utilized as a starting point for building ScNPC models from the knock-out cells based on respective in-cell cryo-ET maps. The Nup that was knocked-out from these NPCs was Nup116, a critical linker Nup for the CR and IR, and the cells from which the NPCs were imaged were grown under permissive and non-permissive temperatures. The ScNPC model from the permissive temperature exhibited an almost identical overall architecture compared to the wild type as opposed to the model from the non-permissive temperature, where approximately half of the NPC subunits were missing. The architecture of the latter model resembled a structural phenotype observed during failed inside-out NPC assembly. Therefore, we hypothesized that this model represented such a failed-to-assemble ScNPC. Despite this hypothesis, further experiments should be conducted in order to confirm the biological relevance of this model as well as to shed more light on the time-resolved events of NPC assembly and biogenesis.

After the ScNPC modeling success, I utilized Assembline[120] once again for the production of integrative models from another yeast species, namely the fission yeast. The NPC of *S. pombe* was significantly less studied compared to other eukaryotes (e.g. human and *S. cerevisiae*). In this case, our SpNPC models, which were based on in-cell cryo-ET datasets with resolution ~23 Å, offered the first glimpse into the overall architecture of this complex in *S. pombe*. The modeling logic that was followed for the SpNPC was very similar to the case of ScNPCs (see Synopsis section 5.2).

The aforementioned SpNPC models highlighted a striking scaffold NPC organization that was never observed previously. This unprecedented architecture consisted of a non-continuous CR, with the densities belonging to the "tail" Y-complex Nups being non-existent, and dimerized NR. The fact that the Y-complexes in the cytoplasmic scaffold side of the NPC did not appear as complete "head-to-tail" protein complex entities directly challenged a long-standing dogma regarding the integrity of NPC rings for the proper assembly and function of such complexes. In the NR side, the SpNPC exhibited a very similar conformation with the human and algal NPCs where the Y-complexes appeared complete and dimerized, hence totaling 16 copies in the nuclear-facing side of the pore. The datasets that were utilized for SpNPC modeling with Assembline did not provide any further "clues" though regarding the responsible Nup or Nups for the Y-complex dimerization, which could be the target of future structural NPC studies. Also, further work regarding these radical stoichiometric differences in essential NPC subcomplexes between the yeast species would possibly shed more light in the molecular evolution of such complexes[151]. Additionally, such future research attempts could possibly reveal "missing" connections between the observed stoichiometrical variability and the physiological cell state during the conduction of experiments.

On a different note, it was already shown that various physico-chemical cellular cues could have a major impact in diameter changes of the pores. In this context, we utilized the SpNPC model that I built under native conditions together with in-cell cryo-ET maps (lower resolution compared to native SpNPC maps) of SpNPCs from energy depleted cells in order to produce respective integrative models. The latter NPC models revealed the significantly constricted conformation of

pores (see Synopsis section 5.2), which was indeed coupled to dramatically decreased (almost shut-down) active transport, as was shown by further biochemical experiments in this collaborative study. Additionally, apart from the constricted architecture of the energy depleted SpNPCs, their overall structural organization was almost identical to the NPCs from cells growing under native conditions. The aforementioned constricted conformation appeared to be achieved by independent lateral movements of many subunits belonging to scaffold NPC subcomplexes. Despite the capturing of a molecular movie regarding the NPC constriction and dilation based on these integrative SpNPC models, the exact mechanism for the initiation and proper completion of such large conformational changes remains unknown. Some of the "molecular suspects" of the NPC diameter regulation could possibly be the luminal ring constituents, e.g. SpPom152 (HsGP210), and nuclear envelope tension. It would definitely be interesting to investigate if and how the NPC diameter is affected by changes of the SpPOM152 length (e.g. less or more Ig-like domains) which in our cryo-ET datasets under energy depletion conditions exhibited clear arches not present under native conditions. Furthermore, judging by the increased pore diameters of our in-cell ScNPC and native SpNPC models, it would be interesting to investigate in the future whether the publicly available NPC models that were built based on in vitro or detergent-extracted NPC datasets were equivalent to the constricted versions (energy depletion conditions) of the SpNPC. Such future experiments would elucidate the "key" factors behind the NPC constriction and rule out the structural effects different physico-chemical cellular stresses could cause.

In general, the ScNPC and SpNPC models that I have built during my PhD research augmented our structural and functional understanding of these large complexes while revealing several aspects of their structural "behavior" under near-native conditions. Nevertheless, all these models were built based on mid-to-low resolution datasets with "inherent" uncertainties hence leaving much room for resolution- and accuracy-wise improvements from studies in the near future. For example, peripheral elements like the nuclear basket and cytoplasmic filaments are completely absent from our models due to inexistent or very low resolution datasets. Despite the fact that we were able to identify possible nuclear basket anchor points in our cryo-ET datasets from ScNPCs we were not able to allocate the observed densities unambiguously to Nups like Mlp1/Mlp2. Another example is the absence of the mesh-like network of FG-repeats from our models, which although included "solid" anchor points along the modeled NPC scaffolds for projections of FG-repeats we did not have the computational means to model them efficiently. We hope that in the near future, modeling methods like AlphaFold together with Assembline and other modeling simulation tools will be able to complete the structural puzzle of NPCs. Such modeling attempts could start from our in-cell models and integrate datasets of higher resolution while taking advantage of the ever-increasing amount of available computational resources. All in all, our NPC findings clearly highlight the need for in-cell structural studies in order to reveal the architecture of protein complexes in their native cellular context and showcase the power of hybrid modeling methods in producing reliable structural models even from mid-to-low resolution data.

7. Bibliography

1. Dacks, J. B. *et al.* The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together. *Journal of Cell Science* vol. 129 3695–3703 (2016).
2. Strambio-De-Castillia, C., Niepel, M. & Rout, M. P. The nuclear pore complex: Bridging nuclear transport and gene regulation. *Nature Reviews Molecular Cell Biology* vol. 11 490–501 (2010).
3. Hampoelz, B., Andres-Pons, A., Kastritis, P. & Beck, M. Structure and Assembly of the Nuclear Pore Complex. *Annual Review of Biophysics* vol. 48 515–536 (2019).
4. Lin, D. H. *et al.* Architecture of the symmetric core of the nuclear pore. *Science* 352, aaf1015 (2016).
5. Beck, M. & Hurt, E. The nuclear pore complex: understanding its function through structural insight. *Nat. Rev. Mol. Cell Biol.* 18, 73–89 (2017).
6. Gall, J. G. OCTAGONAL NUCLEAR PORES. *J. Cell Biol.* 32, 391–399 (1967).
7. Donnaloja, F., Jacchetti, E., Soncini, M. & Raimondi, M. T. Mechanosensing at the nuclear envelope by nuclear pore complex stretch activation and its effect in physiology and pathology. *Frontiers in Physiology* vol. 10 896 (2019).
8. Lin, D. H. & Hoelz, A. The structure of the nuclear pore complex (An Update). *Annu. Rev. Biochem.* 88, 725–783 (2019).
9. Hinshaw, J. E., Carragher, B. O. & Milligan, R. A. Architecture and design of the nuclear pore complex. *Cell* 69, 1133–1141 (1992).
10. Knockenhauer, K. E. & Schwartz, T. U. The Nuclear Pore Complex as a Flexible and Dynamic Gate. *Cell* vol. 164 1162–1171 (2016).
11. Lim, R. Y. H. *et al.* Nanomechanical basis of selective gating by the nuclear pore complex. *Science (80-.).* 318, 640–643 (2007).
12. Terry, L. J. & Wente, S. R. Flexible gates: Dynamic topologies and functions for FG nucleoporins in nucleocytoplasmic transport. *Eukaryotic Cell* vol. 8 1814–1827 (2009).
13. Bailer, S. M. *et al.* Nup116p and Nup100p are interchangeable through a conserved motif which constitutes a docking site for the mRNA transport factor Gle2p. *EMBO J.* 17, 1107–1119 (1998).
14. Hartmann, M. D. *et al.* α/β coiled coils. *Elife* 5, (2016).
15. Devos, D. *et al.* Simple fold composition and modular architecture of the nuclear pore complex. *Proc. Natl. Acad. Sci. U. S. A.* 103, 2172–2177 (2006).
16. Floch, A. G. & Palancade, B. Fifty Years of Nuclear Pores and Nucleocytoplasmic Transport Studies: Multiple Tools Revealing Complex Rules. *Methods Cell Biol.* 122, 1–40 (2014).

17. Liu, X., Mitchell, J. M., Wozniak, R. W., Blobel, G. & Fan, J. Structural evolution of the membrane-coating module of the nuclear pore complex. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16498–16503 (2012).

18. Brohawn, S. G., Partridge, J. R., Whittle, J. R. R. & Schwartz, T. U. The Nuclear Pore Complex Has Entered the Atomic Age. *Structure* vol. 17 1156–1168 (2009).

19. Mettlen, M., Chen, P. H., Srinivasan, S., Danuser, G. & Schmid, S. L. Regulation of Clathrin-Mediated Endocytosis. *Annual Review of Biochemistry* vol. 87 871–896 (2018).

20. Drin, G. *et al.* A general amphipathic α-helical motif for sensing membrane curvature. *Nat. Struct. Mol. Biol.* 14, 138–146 (2007).

21. Kim, S. J. *et al.* Integrative structure-function mapping of the nucleoporin Nup133 suggests a conserved mechanism for membrane anchoring of the nuclear pore complex. *Mol. Cell. Proteomics* 13, 2911–2926 (2014).

22. Fernandez-Martinez, J. *et al.* Structure and Function of the Nuclear Pore Complex Cytoplasmic mRNA Export Platform. *Cell* 167, 1215-1228.e25 (2016).

23. Frosst, P., Guan, T., Subauste, C., Hahn, K. & Gerace, L. Tpr is localized within the nuclear basket of the pore complex and has a role in nuclear protein export. *J. Cell Biol.* 156, 617–630 (2002).

24. Schmidt, H. B. & Görlich, D. Transport Selectivity of Nuclear Pores, Phase Separation, and Membraneless Organelles. *Trends in Biochemical Sciences* vol. 41 46–61 (2016).

25. Ori, A. *et al.* Cell type-specific nuclear pores: A case in point for context-dependent stoichiometry of molecular machines. *Mol. Syst. Biol.* 9, 648 (2013).

26. Hurt, E. C. A novel nucleoskeletal-like protein located at the nuclear periphery is required for the life cycle of Saccharomyces cerevisiae. *EMBO J.* 7, 4323–4334 (1988).

27. Wente, S. R., Rout, M. P. & Blobel, G. A new family of yeast nuclear pore complex proteins. *J. Cell Biol.* 119, 705–723 (1992).

28. Schmidt, H. B. & Görlich, D. Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity. *Elife* 2015, (2015).

29. Denning, D. P. & Rexach, M. F. Rapid evolution exposes the boundaries of domain structure and function in natively unfolded FG nucleoporins. *Mol. Cell. Proteomics* 6, 272–282 (2007).

30. Bayliss, R. *et al.* Interaction between NTFP and xFxFG-containing nucleoporins is required to mediate nuclear import of RanGDP. *J. Mol. Biol.* 293, 579–593 (1999).

31. Patel, S. S., Belmont, B. J., Sante, J. M. & Rexach, M. F. Natively Unfolded Nucleoporins Gate Protein Diffusion across the Nuclear Pore Complex. *Cell* 129, 83–96 (2007).

32. Strawn, L. A., Shen, T., Shulga, N., Goldfarb, D. S. & Wente, S. R. Minimal nuclear pore complexes define FG repeat domains essential for transport. *Nat. Cell Biol.* 6, 197–206 (2004).

33. Frey, S., Richter, R. P. & Görlich, D. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science (80-.).* 314, 815–817 (2006).

34. Hoelz, A., Debler, E. W. & Blobel, G. The Structure of the nuclear pore complex. *Annu. Rev. Biochem.* 80, 613–643 (2011).

35. Allen, N. P. C. *et al.* Deciphering Networks of Protein Interactions at the Nuclear Pore Complex *. *Mol. Cell. Proteomics* 1, 930–946 (2002).

36. Schwartz, T. U. Modularity within the architecture of the nuclear pore complex. *Curr. Opin. Struct. Biol.* 15, 221–226 (2005).

37. Nordeen, S. A., Turman, D. L. & Schwartz, T. U. Yeast Nup84-Nup133 complex structure details flexibility and reveals conservation of the membrane anchoring ALPS motif. *Nat. Commun.* 11, 1–12 (2020).

38. Harel, A. *et al.* Removal of a single pore subcomplex results in vertebrate nuclei devoid of nuclear pores. *Mol. Cell* 11, 853–864 (2003).

39. Nordeen, S. A. *et al.* A nanobody suite for yeast scaffold nucleoporins provides details of the nuclear pore complex structure. *Nat. Commun.* 11, 1–13 (2020).

40. Kelley, K., Knockenhauer, K. E., Kabachinski, G. & Schwartz, T. U. Atomic structure of the Y complex of the nuclear pore. *Nat. Struct. Mol. Biol.* 22, 425–431 (2015).

41. Bilokapic, S. & Schwartz, T. U. Structural and functional studies of the 252 kDa nucleoporin ELYS reveal distinct roles for its three tethered domains. *Structure* 21, 572–580 (2013).

42. Thierbach, K. *et al.* Protein interfaces of the conserved Nup84 complex from chaetomium thermophilum shown by crosslinking mass spectrometry and electron microscopy. *Structure* 21, 1672–1682 (2013).

43. Mosalaganti, S. *et al.* In situ architecture of the algal nuclear pore complex. *Nat. Commun.* 9, 2361 (2018).

44. Nagy, V. *et al.* Structure of a trimeric nucleoporin complex reveals alternate oligomerization states. *Proc. Natl. Acad. Sci. U. S. A.* 106, 17693–17698 (2009).

45. von Appen, A. *et al.* In situ structural analysis of the human nuclear pore complex. *Nature* 526, 140–143 (2015).

46. Kim, S. J. *et al.* Integrative structure and functional anatomy of a nuclear pore complex. *Nature* 555, 475–482 (2018).

47. Kosinski, J. *et al.* Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* 352, 363–5 (2016).

48. Fischer, J., Teimer, R., Amlacher, S., Kunze, R. & Hurt, E. Linker Nups connect the nuclear pore complex inner ring with the outer ring and transport channel. *Nat. Struct. Mol. Biol.* 22, 774–781 (2015).

49. Fahrenkrog, B. & Aebi, U. The nuclear pore complex: nucleocytoplasmic transport and

beyond. *Nat. Rev. Mol. Cell Biol.* 4, 757–766 (2003).

50. Yoshida, K., Seo, H.-S., Debler, E. W., Blobel, G. & Hoelz, A. Structural and functional analysis of an essential nucleoporin heterotrimer on the cytoplasmic face of the nuclear pore complex. *Proc. Natl. Acad. Sci. U. S. A.* 108, 16571–6 (2011).

51. Stuwe, T. *et al.* Architecture of the fungal nuclear pore inner ring complex. *Science (80-.).* 350, 56–64 (2015).

52. Gaik, M. *et al.* Structural basis for assembly and function of the Nup82 complex in the nuclear pore scaffold. *J. Cell Biol.* 208, 283–97 (2015).

53. Belgareh, N. *et al.* Functional characterization of a Nup159p-containing nuclear pore subcomplex. *Mol. Biol. Cell* 9, 3475–3492 (1998).

54. Stuwe, T. *et al.* Architecture of the nuclear pore complex coat. *Science (80-.).* 347, 1148–1152 (2015).

55. Tran, E. J., Zhou, Y., Corbett, A. H. & Wente, S. R. The DEAD-Box Protein Dbp5 Controls mRNA Export by Triggering Specific RNA:Protein Remodeling Events. *Mol. Cell* 28, 850–859 (2007).

56. Tieg, B. & Krebber, H. Dbp5 - From nuclear export to translation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* vol. 1829 791–798 (2013).

57. Hodge, C. A., Colot, H. V., Stafford, P. & Cole, C. N. Rat8p/Dbp5p is a shuttling transport factor that interacts with Rat7p/Nup159p and Gle1p and suppresses the mRNA export defect of xpo1-1 cells. *EMBO J.* 18, 5778–5788 (1999).

58. Stelter, P. *et al.* Molecular basis for the functional interaction of dynein light chain with the nuclear-pore complex. *Nat. Cell Biol.* 9, 788–796 (2007).

59. Jarnik, M. & Aebi, U. Toward a more complete 3-D structure of the nuclear pore complex. *J. Struct. Biol.* 107, 291–308 (1991).

60. Mészáros, N. *et al.* Nuclear Pore Basket Proteins Are Tethered to the Nuclear Envelope and Can Regulate Membrane Curvature. *Dev. Cell* 33, 285–298 (2015).

61. Soop, T. *et al.* Nup153 affects entry of messenger and ribosomal ribonucleoproteins into the nuclear basket during export. *Mol. Biol. Cell* 16, 5610–5620 (2005).

62. Niepel, M. *et al.* The nuclear basket proteins Mlp1p and Mlp2p are part of a dynamic interactome including Esc1p and the proteasome. *Mol. Biol. Cell* 24, 3920–3938 (2013).

63. Cordes, V. C., Reidenbach, S., Rackwitz, H. R. & Franke, W. W. Identification of protein p270/Tpr as a constitutive component of the nuclear pore complex-attached intranuclear filaments. *J. Cell Biol.* 136, 515–529 (1997).

64. Krull, S., Thyberg, J., Björkroth, B., Rackwitz, H. R. & Cordes, V. C. Nucleoporins as components of the nuclear pore complex core structure and Tpr as the architectural element of the nuclear basket. *Mol. Biol. Cell* 15, 4261–4277 (2004).

65. Pal, K. *et al.* Structural Basis of TPR-Mediated Oligomerization and Activation of

Oncogenic Fusion Kinases. *Structure* 25, 867-877.e3 (2017).

66. Hase, M. E., Kuznetsov, N. V. & Cordes, V. C. Amino acid substitutions of coiled-coil protein Tpr abrogate anchorage to the nuclear pore complex but not parallel, in-register homodimerization. *Mol. Biol. Cell* 12, 2433–2452 (2001).

67. Walther, T. C. *et al.* The nucleoporin Nup153 is required for nuclear pore basket formation, nuclear pore complex anchoring and import of a subset of nuclear proteins. *EMBO J.* 20, 5703–5714 (2001).

68. Zhang, Y. *et al.* Molecular architecture of the luminal ring of the Xenopus laevis nuclear pore complex. *Cell Res.* 30, 532–540 (2020).

69. Stoffler, D. *et al.* Cryo-electron tomography provides novel insights into nuclear pore architecture: Implications for nucleocytoplasmic transport. *J. Mol. Biol.* 328, 119–130 (2003).

70. Neumann, N., Lundin, D. & Poole, A. M. Comparative genomic evidence for a complete nuclear pore complex in the last eukaryotic common ancestor. *PLoS One* 5, e13241 (2010).

71. Greber, U. F., Senior, A. & Gerace, L. A major glycoprotein of the nuclear pore complex is a membrane-spanning polypeptide with a large lumenal domain and a small cytoplasmic tail. *EMBO J.* 9, 1495–1502 (1990).

72. Upla, P. *et al.* Molecular Architecture of the Major Membrane Ring Component of the Nuclear Pore Complex. *Structure* 25, 434–445 (2017).

73. Hao, Q., Zhang, B., Yuan, K., Shi, H. & Blobel, G. Electron microscopy of Chaetomium pom152 shows the assembly of ten-bead string. *Cell Discovery* vol. 4 56 (2018).

74. Cohen, M., Feinstein, N., Wilson, K. L. & Gruenbaum, Y. Nuclear pore protein gp210 is essential for viability in HeLa cells and Caenorhabditis elegans. *Mol. Biol. Cell* 14, 4230–4237 (2003).

75. Madrid, A. S., Mancuso, J., Cande, W. Z. & Weis, K. The role of the integral membrane nucleoporins Ndc1p and Pom152p in nuclear pore complex assembly and function. *J. Cell Biol.* 173, 361–371 (2006).

76. Meinema, A. C. *et al.* Long unfolded linkers facilitate membrane protein import through the nuclear pore complex. *Science (80-.).* 333, 90–93 (2011).

77. Dickmanns, A., Kehlenbach, R. H. & Fahrenkrog, B. Nuclear Pore Complexes and Nucleocytoplasmic Transport: From Structure to Function to Disease. *Int. Rev. Cell Mol. Biol.* 320, 171–233 (2015).

78. Khan, A. U., Qu, R., Ouyang, J. & Dai, J. Role of Nucleoporins and Transport Receptors in Cell Differentiation. *Front. Physiol.* 0, 239 (2020).

79. Timney, B. L. *et al.* Simple rules for passive diffusion through the nuclear pore complex. *J. Cell Biol.* 215, 57–76 (2016).

80. Solmaz, S. R., Blobel, G. & Melčák, I. Ring cycle for dilating and constricting the nuclear pore. *Proc. Natl. Acad. Sci.* 110, 5858–5863 (2013).

81. Pulupa, J., Prior, H., Johnson, D. S. & Simon, S. M. Conformation of the nuclear pore in living cells is modulated by transport state. *Elife* 9, 1–24 (2020).

82. Sellés, J. *et al.* Nuclear pore complex plasticity during developmental process as revealed by super-resolution microscopy. *Sci. Reports 2017 71* 7, 1–8 (2017).

83. Mahamid, J. *et al.* Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science (80-.).* 351, 969–972 (2016).

84. Srivastava, A., Tiwari, S. P., Miyashita, O. & Tama, F. Integrative/Hybrid Modeling Approaches for Studying Biomolecules. *Journal of Molecular Biology* vol. 432 2846–2860 (2020).

85. Glaeser, R. M. How Good Can Single-Particle Cryo-EM Become? What Remains Before It Approaches Its Physical Limits? *https://doi.org/10.1146/annurev-biophys-070317-032828* 48, 45–61 (2019).

86. Dunstone, M. A. & de Marco, A. Cryo-electron tomography: an ideal method to study membrane-associated proteins. *Philos. Trans. R. Soc. B Biol. Sci.* 372, (2017).

87. Turk, M. & Baumeister, W. The promise and the challenges of cryo-electron tomography. *FEBS Lett.* 594, 3243–3261 (2020).

88. Murata, K. & Wolf, M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochim. Biophys. Acta - Gen. Subj.* 1862, 324–334 (2018).

89. Koukos, P. I. & Bonvin, A. M. J. J. Integrative Modelling of Biomolecular Complexes. *Journal of Molecular Biology* vol. 432 2861–2881 (2020).

90. Sali, A. From integrative structural biology to cell biology. *J. Biol. Chem.* 296, 100743 (2021).

91. Braitbard, M., Schneidman-Duhovny, D. & Kalisman, N. Integrative structure modeling: Overview and assessment. *Annual Review of Biochemistry* vol. 88 113–135 (2019).

92. Rout, M. P. & Sali, A. Principles for Integrative Structural Biology Studies. *Cell* vol. 177 1384–1403 (2019).

93. Soni, N. & Madhusudhan, M. S. Computational modeling of protein assemblies. *Curr. Opin. Struct. Biol.* 44, 179–189 (2017).

94. Vakser, I. A. & Deeds, E. J. Computational approaches to macromolecular interactions in the cell. *Curr. Opin. Struct. Biol.* 55, 59–65 (2019).

95. Allegretti, M. *et al.* In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature* 586, 796–800 (2020).

96. O'Reilly, F. J. *et al.* In-cell architecture of an actively transcribing-translating expressome. *Science (80-.).* 369, 554–557 (2020).

97. Zimmerli, C. E. *et al.* Nuclear pores dilate and constrict in cellulo. *Science (80-.).* eabd9776 (2021).

98. Saltzberg, D. J. *et al.* Using Integrative Modeling Platform to compute, validate, and archive

a model of a protein complex structure. *Protein Sci.* 30, 250–261 (2021).

99. Viswanath, S., Chemmama, I. E., Cimermancic, P. & Sali, A. Assessing Exhaustiveness of Stochastic Sampling for Integrative Modeling of Macromolecular Structures. *Biophys. J.* 113, 2344–2353 (2017).

100. Vallat, B., Webb, B., Westbrook, J. D., Sali, A. & Berman, H. M. Development of a Prototype System for Archiving Integrative/Hybrid Structure Models of Biological Macromolecules. *Structure* 26, 894-904.e2 (2018).

101. Burley, S. K. *et al.* PDB-Dev: a Prototype System for Depositing Integrative/Hybrid Structural Models. *Structure* 25, 1317–1318 (2017).

102. Doerr, A. A home for integrative structural models. *Nat. Methods 2018 156* 15, 409–409 (2018).

103. Heilmann, N. *et al.* Sampling of the conformational landscape of small proteins with Monte Carlo methods. *Sci. Reports 2020 101* 10, 1–13 (2020).

104. Velázquez-Muriel, J. *et al.* Assembly of macromolecular complexes by satisfaction of spatial restraints from electron microscopy images. *Proc. Natl. Acad. Sci.* 109, 18821–18826 (2012).

105. Schneidman-Duhovny, D., Pellarin, R. & Sali, A. Uncertainty in integrative structural modeling. *Curr. Opin. Struct. Biol.* 28, 96–104 (2014).

106. Song, H. K. & Suh, S. W. Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from Erythrina caffra and tissue-type plasminogen activator. *J. Mol. Biol.* 275, 347–363 (1998).

107. Pettersen, E. F. *et al.* UCSF Chimera-A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612 (2004).

108. Wriggers, W. Conventions and workflows for using Situs. *Acta Crystallogr. D. Biol. Crystallogr.* 68, 344–51 (2012).

109. Lopéz-Blanco, J. R. & Chacón, P. IMODFIT: Efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J. Struct. Biol.* 184, 261–270 (2013).

110. Lasker, K., Topf, M., Sali, A. & Wolfson, H. J. Inferential Optimization for Simultaneous Fitting of Multiple Components into a CryoEM Map of Their Assembly. *J. Mol. Biol.* 388, 180–194 (2009).

111. Russel, D. *et al.* Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies. *PLoS Biol.* 10, e1001244 (2012).

112. Webb, B. *et al.* Integrative structure modeling with the Integrative Modeling Platform. *Protein Sci.* 27, 245–258 (2018).

113. Van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* 428, 720–725 (2016).

114. DiMaio, F., Tyka, M. D., Baker, M. L., Chiu, W. & Baker, D. Refinement of Protein Structures into Low-Resolution Density Maps Using Rosetta. *J. Mol. Biol.* 392, 181–190 (2009).

115. Joong Kim, S. *et al.* Structure and Functional Anatomy of the Nuclear Pore Complex. *Biophys. J.* 114, 372a (2018).

116. Dauden, M. I. *et al.* Architecture of the yeast Elongator complex. *EMBO Rep.* 18, 264–279 (2017).

117. Lasker, K. *et al.* Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl. Acad. Sci.* 109, 1380–1387 (2012).

118. Saltzberg, D. *et al.* Modeling Biological Complexes Using Integrative Modeling Platform. in *Methods in Molecular Biology* vol. 2022 353–377 (Humana Press Inc., 2019).

119. Shi, Y. *et al.* Structural Characterization by Cross-linking Reveals the Detailed Architecture of a Coatomer-related Heptameric Module from the Nuclear Pore Complex *. *Mol. Cell. Proteomics* 13, 2927–2943 (2014).

120. Rantos, V., Karius, K. & Kosinski, J. Integrative structural modeling of macromolecular complexes using Assembline. *Nat. Protoc. 2021* 1–25 (2021).

121. Wente, S. R. & Blobel, G. A temperature-sensitive NUP116 null mutant forms a nuclear envelope seal over the yeast nuclear pore complex thereby blocking nucleocytoplasmic traffic. *J. Cell Biol.* 123, 275–284 (1993).

122. Adams, R. L., Terry, L. J. & Wente, S. R. Nucleoporin FG domains facilitate mRNP remodeling at the cytoplasmic face of the nuclear pore complex. *Genetics* 197, 1213–1224 (2014).

123. Otsuka, S. *et al.* Nuclear pore assembly proceeds by an inside-out extrusion of the nuclear envelope. *Elife* 5, (2016).

124. Asakawa, H. *et al.* Asymmetrical localization of Nup107-160 subcomplex components within the nuclear pore complex in fission yeast. *PLOS Genet.* 15, e1008061 (2019).

125. Asakawa, H. *et al.* Characterization of nuclear pore complex components in fission yeast Schizosaccharomyces pombe. *Nucleus* 5, 149–62 (2014).

126. Ziegler, S. J., Mallinson, S. J. B., St. John, P. C. & Bomble, Y. J. Advances in integrative structural biology: Towards understanding protein complexes in their cellular context. *Comput. Struct. Biotechnol. J.* 19, 214–225 (2021).

127. Kosinski, J. *et al.* Xlink analyzer: Software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.* 189, 177–183 (2015).

128. Beckham, K. S. H. *et al.* Structure of the mycobacterial ESX-5 type VII secretion system pore complex. *Sci. Adv.* 7, eabg9923 (2021).

129. Engel, A. & Gaub, H. E. Structure and Mechanics of Membrane Proteins. *https://doi.org/10.1146/annurev.biochem.77.062706.154450* 77, 127–148 (2008).

130. Zhao, Y. *et al.* Crystal Structure of the E. coli Peptide Transporter YbgH. *Structure* 22, 1152–1160 (2014).

131. Jung, D. & Dorr, A. Single-Dose Pharmacokinetics of Valganciclovir in HIV- and CMV-Seropositive Subjects. *J. Clin. Pharmacol.* 39, 800–804 (1999).

132. Ural-Blimke, Y. *et al.* Structure of Prototypic Peptide Transporter DtpA from E. coli in Complex with Valganciclovir Provides Insights into Drug Binding of Human PepT1. *J. Am. Chem. Soc.* 141, 2404–2412 (2019).

133. Martinez Molledo, M., Quistgaard, E. M., Flayhan, A., Pieprzyk, J. & Löw, C. Multispecific Substrate Recognition in a Proton-Dependent Oligopeptide Transporter. *Structure* 26, 467-476.e4 (2018).

134. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* 430, 2237–2243 (2018).

135. Kurowski, M. A. & Bujnicki, J. M. GeneSilico protein structure prediction meta-server. *Nucleic Acids Res.* 31, 3305–3307 (2003).

136. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723 (1997).

137. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815 (1993).

138. Schwede, T. *et al.* Outcome of a workshop on applications of protein models in biomedical research. *Structure* 17, 151–159 (2009).

139. Studer, G., Biasini, M. & Schwede, T. Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). *Bioinformatics* 30, i505–i511 (2014).

140. Laskowski, R. A. & Swindells, M. B. LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model.* 51, 2778–2786 (2011).

141. Viswanath, S. & Sali, A. Optimizing model representation for integrative structure determination of macromolecular assemblies. *Proc. Natl. Acad. Sci. U. S. A.* 116, 540–545 (2019).

142. Kannan, S. & Zacharias, M. Simulated annealing coupled replica exchange molecular dynamics—An efficient conformational sampling method. *J. Struct. Biol.* 166, 288–294 (2009).

143. Bonomi, M. *et al.* Bayesian Weighing of Electron Cryo-Microscopy Data for Integrative Structural Modeling. 27, 175-188.e6 (2019).

144. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nat. 2021 5967873* 596, 583–589 (2021).

145. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science (80-.).* 373, 871–876 (2021).

146. Pereira, J. *et al.* High-accuracy protein structure prediction in CASP14. *Proteins Struct. Funct. Bioinforma.* (2021) doi:10.1002/PROT.26171.

147. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nat. 2021 5967873* 596, 590–596 (2021).

148. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods 2022 196* 19, 679–682 (2022).

149. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034 (2022) doi:10.1101/2021.10.04.463034.

150. Mosalaganti, S. *et al.* AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science (80-.).* 376, (2022).

151. Akey, C. W. *et al.* Comprehensive structure and functional adaptations of the yeast nuclear pore complex. *Cell* 185, 361-378.e25 (2022).

8. Appendix

8.1 List of hazardous substances according to GHS (hazard symbols H- and P)

No hazardous substances were used for the work presented in this Thesis. My doctoral dissertation was conducted exclusively with computational resources.

9. Acknowledgements

First, I want to thank my EMBL PhD supervisor Jan Kosinski for granting me the opportunity to work on such novel and challenging modeling projects. Thank you very much for your continuous and fruitful feedback during the process of my PhD research and you have been an unlimited source of inspiration.

I would also like to thank the members of my Thesis Advisory Committee Martin Beck, Kay Gruenewald and Christian Loew for their insightful suggestions and advice during each evaluation stage.

A big thank you as well to all Kosinski group members, the current ones as well as the former ones. Especially during the last years, you became more of a family to me from which I was receiving all the needed scientific and mental support for a successful PhD. It was a pleasure and honor for me that I worked by your side and it will forever be a memorable journey.

Furthermore, I would like to thank Matteo Allegretti and Christian E. Zimmerli from Beck group (EMBL Heidelberg, MPI Frankfurt) for their continuous and in-depth feedback during our collaborative efforts over these years.

At this point, I would like to thank the rest of EMBL Hamburg scientific personnel, the administration team and the IT group for providing a great and supportive environment to conduct cutting-edge research. Additionally, a great thank you goes to EMBL, which through the EIPP enabled me to grow both scientifically and personality-wise surrounded by a group of bright and like-minded people.

Finally, I want to say a big thank you to my family and friends who have always tirelessly been there for me and supported me so far in every important step of my life.

Σας ευχαριστώ από καρδιάς!

10. Declaration of academic honesty (Eidesstattliche Versicherung)

I hereby declare under oath that this doctoral dissertation is my own work and that I have not used any sources other than those listed.

23.09.2022, Vasileios Rantos

Hiermit versichere ich an Eides statt, die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt zu haben. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Ich versichere, dass diese Dissertation nicht in einem früheren Promotionsverfahren eingereicht wurde.

23.09.2022, Vasileios Rantos