

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

presented by
M.Sc. Biophysics and Chemical Biology

Jun Hee Kang
Seoul, South Korea

11.10.2016

Precise mapping of disordered regions of nucleoporins *in situ*

Referees: Dr. Martin Beck

Prof. Dr. Frauke Melchior

Summary

The nuclear pore complex (NPC) is an immense structure which functions as a primary gateway for transport of biomolecules in and out of the nucleus. The NPC allows passive diffusion of small molecules (< 40 kDa, < 4 nm) while blocking large cargo molecules unless they are bound to nuclear transport receptors (NTRs). It is composed of roughly 30 different types of proteins called nucleoporins (Nups). The central transport channel is filled with Nups containing numerous phenylalanine-glycine (FG) repeats which play a major role in forming the permeability barrier.

Recent advancements in structural biology approaches have greatly contributed to deciphering the structure of the ring-like NPC scaffold. However, insights about the fundamental nature of the FG Nups such as their spatial organization and dynamics in the central transport channel of the NPC still remains challenging. This is mainly due to technical limitations when aiming to study the flexible intrinsically disordered regions of FG Nups.

The goal of my thesis was to devise an experimental strategy to study the spatial organization of FG Nups with residue precision in cells using super-resolution microscopy (SRM). One main bottleneck to achieve this goal was to develop a tool to site-specifically label intracellular proteins with synthetic fluorophores (SFs). In this thesis, I developed and established a labeling strategy implementing genetic code expansion (GCE) technology and click-chemistry reaction in order to label FG Nups *in situ* in mammalian cells with residue specificity. Moreover, I designed a dual-color imaging strategy using spectral demixing coupled to a robust automated custom-written analysis to quantitatively assess and compare different labeling positions along the disordered regions of FG Nups. I validated the developed labeling and analysis strategy by determining the localization of Nup153, which is one of the largest and in terms of function, one of the most debated FG Nups in the NPC of mammalian cells.

Ultimately, I fully utilized the new approach to map the spatial organization of distinct Nup98 regions *in situ* in mammalian cells. I will discuss possible interpretations and pitfalls of the acquired results and also the steps that are needed to make the labeling strategy live-cell compatible. Further application of my developed approach to other FG Nups in the central channel of the NPC will contribute in acquiring a deeper understanding of the molecular mechanisms behind nucleocytoplasmic transport. In general, my work presents a complementary approach to study the structure and dynamics of flexible proteins inside cells and paves the way towards precise in-cell structural biology.

This work was carried out in the lab of Edward A. Lemke in EMBL, Heidelberg.

Zusammenfassung

Der Kernporenkomplex (nuclear pore complex, NPC) ist eine ungemein grosse Struktur, die als zentrales Portal für den nuklearen Transport von Biomolekülen in und aus dem Nukleus fungiert. Der NPC erlaubt die passive Diffusion von kleinen Molekülen (> 40 kDa, > 4 nm), während grosse Moleküle blockiert werden, falls diese nicht an Nukleartransportrezeptoren (NTRs) gebunden sind. Der NPC besteht aus 30 unterschiedlichen Arten von Proteinen, den sogenannten Nukleoporinen (Nups). Der zentrale Kanal ist gefüllt mit Nups, bestehend aus vielfachen Phenylalanine-Glycine (FG) Wiederholungen, die eine Hauptrolle für die Funktion der Permeabilitätsbarriere einnehmen.

Kürzliche Fortschritte in strukturbiologischen Ansätzen haben in hohem Masse zum Aufschlüsseln der ringartigen Struktur des NPC beigetragen. Allerdings bleiben Einblicke in die grundlegende Nature der FG Nups, wie die räumliche Organisation sowie die Dynamik im zentralen Transportkanal des NPCs, immer noch eine grosse Herausforderung. Dies ist grösstenteils technischen Einschränkungen zu Buche zu schreiben, wenn man versucht, die flexiblen ungeordneten Regionen der FG Nups zu untersuchen.

Das Ziel meiner Doktorarbeit war es eine experimentelle Strategie zu entwickeln, um die räumliche Organisation der FG Nups mit der Präzision im Bereich von nahezu einem Aminosäurerest, mithilfe von hochauflösender Mikroskopie (super-resolution microscopy, SRM) zu untersuchen. Eine der grössten Herausforderungen beim Erreichen dieses Ziels war es, seiten-spezifisch intrazelluläre Proteine mit synthetischen Farbstoffen (SFs) zu markieren. In dieser Doktorarbeit habe ich eine Markierungsstrategie entwickelt und etabliert, unter Verwendung der genetischen Kode-Expansions-Technologie (genetic code expansion, GCE) kombiniert mit Klick-Chemie, um FG Nups in situ in Säugetierzellen mit Aminosäuregenauigkeit zu markieren. Darüber hinaus habe ich eine zweifarbige, bildgebende Strategie entworfen, unter der Verwendung spektraler Entmischung. An diese ist eine robuste, automatisierte und selbst geschriebene Analyse gekoppelt, die es erlaubt verschiedene Markierungsstellen entlang der ungeordneten Regionen der FG Nups quantitativ abzuschätzen und vergleichen zu können. Ich habe die entwickelten Markierungs- und Analyse-Strategien durch Bestimmung der Lokalisation von Nup153 validiert, welches eines der grössten und in Bezug auf die Funktion, eines der meist diskutierten FG Nups im NPC von Säugetierzellen ist.

Letztendlich habe ich den neuen Lösungsansatz ausgenutzt, um die räumliche Organisation von bestimmten Nup98 Regionen in situ in Säugetierzellen zu entschlüsseln. Ich werde mögliche Interpretationen und Schwierigkeiten der erlangten Ergebnisse und auch die Schritte diskutieren, die nötig sind, um die Markierungsstrategie auch für lebende Zellen kompatibel zu machen. Weitere Anwendungen meiner Entwicklung auf anderen FG Nups im zentralen Kanals des NPCs wird zum Erwerben eines tieferen Verständnisses der molekularen Mechanismen des nukleozytoplasmatischen Transport, beitragen. Generell stellt meine Arbeit ein komplementäres Vorgehen dar, um die Struktur und Dynamik von flexiblen Proteinen in der Zelle zu untersuchen und ebnet den Weg zu einer präzisen in-zell Strukturbiologie.

Diese Arbeit wurde im Labor von Edward A. Lemke am EMBL, Heidelberg.

Publications list

- 1) Kozma, E., Nikic, I., Varga, B. R., Aramburu, I. V., **Kang, J. H.**, Fackler, O. T., Lemke, E. A., Kele, P. (2016). Hydrophilic trans-cyclooctenylated non-canonical amino acids for fast intracellular protein labeling. *ChemBioChem*, *Just accepted*.
- 2) Nikic, I.*, **Kang, J. H.***, Girona, G. E.*, Aramburu, I. V., Lemke, E. A. (2015). Labeling proteins on live mammalian cells using click chemistry. *Nat Protoc*, *10*, 780-91.

* Equal contribution

Table of contents

Summary	I
Zusammenfassung.....	II
Publications list.....	III
Table of contents.....	V
List of figures	VIII
List of Tables.....	VIII
Abbreviation list	IX
1 Introduction.....	1
1.1 Nuclear Pore Complex.....	1
1.1.1 Nucleocytoplasmic transport	2
1.1.2 Models of nucleocytoplasmic transport.....	4
1.2 Methods to study the spatial organization of Nups in the NPC.....	8
1.2.1 Cryo-electron microscopy	9
1.2.2 Atomic force microscopy.....	10
1.2.3 Super-resolution microscopy	10
1.3 Methods to site specifically label proteins in mammalian cells.....	15
1.3.1 Labeling methods after fixing cells.....	15
1.3.2 Labeling methods for live cells	17
1.3.3 Click-chemistry reactions for labeling proteins.....	21
1.4 Objectives.....	25
1.4.1 Site-specific labeling of membrane proteins in mammalian cells.....	25
1.4.2 Site-specific labeling of Nups inside the NPC in mammalian cells	25
1.4.3 Visualize the spatial organization of FG Nups in mammalian cells	26
2 Materials and methods	27
2.1 Mammalian cell culture and transfection	27
2.2 Labeling of proteins in mammalian cells using GCE and click-chemistry.....	28
2.2.1 Labeling of IR at the surface of mammalian cells.....	28
2.2.2 Intracellular labeling of Nups in mammalian cells	28
2.2.3 Labeling of IR and Nups with streptavidin conjugated to AF 647	29
2.2.4 Live-cell intracellular labeling of Nups with SiR.....	30

2.2.5	Synthesis of AF 647 H-tet from NHS.....	30
2.3	Experimental setup	30
2.4	Dual-color single-molecule localization microscopy (SMLM) imaging.....	31
2.4.1	Single-molecule imaging procedure.....	31
2.4.2	Spectral demixing with RapidSTORM.....	32
2.4.3	Custom-written automated evaluation of localization	33
3	Results	35
3.1	Labeling of IR in live mammalian cells using click-chemistry	35
3.2	Intracellular labeling of Nups in mammalian cells	38
3.2.1	Reduction of non-specific background using Strep-AF 647	39
3.2.2	Intracellular labeling of Nups using streptavidin conjugated to synthetic fluorophores.....	41
3.2.3	Use of high salt buffer to wash out non-specifically sticking ncAA and SFs.....	44
3.2.4	Intracellular labeling of live cells with hydrophilic SFs and ncAA.....	47
3.2.5	Assessment of incorporation of Nup153 to the NPC	49
3.3	Dual-color SRM of labeled Nups.....	51
3.3.1	Dual-color SRM of Nup153	51
3.3.2	Automated custom-written analysis to evaluate spatial organization of Nups.....	54
3.3.3	Dual color SRM of disordered regions in Nup98	56
3.3.4	<i>In silico</i> experiments to validate dual-color imaging of Nup98.....	60
4	Discussion	61
4.1	Implementation of click-chemistry reactions for tagging proteins with SFs.....	61
4.2	Intracellular labeling of Nups in mammalian cells	62
4.2.1	Live cell intracellular labeling with hydrophilic ncAA and SiR-based fluorophores ...	63
4.2.2	SR imaging with DNA-PAINT to circumvent photobleaching	64
4.2.3	Other factors to consider when labeling and imaging Nups in mammalian cells.....	65
4.3	Dual color SRM of Nups.....	67
4.3.1	Custom-written analysis for evaluation of labeled Nups	69
4.4	Towards the spatial organization of Nup98 <i>in situ</i>	70
4.4.1	Dual-color SRM results of Nup98 provide clues to its spatial organization	71
4.5	Perspectives.....	74
5	Acknowledgments.....	77

6	Appendices	79
6.1	Appendix I: Spectral demixing using RapidSTORM	79
6.1.1	STEP 1: Localization analysis for alignment.....	79
6.1.2	STEP 2: Alignment of the two channels.....	81
6.1.3	STEP 3: Localization analysis for reconstruction	81
6.1.4	STEP 4: Reconstruction of SR image.....	84
6.2	Appendix II: Manual for analysis of labeled Nups	86
7	References.....	91

List of figures

Figure 1 Schematic representation of the NPC and its sizes.	1
Figure 2 Schematic diagram of the import and Ran cycle	4
Figure 3 Different models proposed to explain the nucleocytoplasmic transport	7
Figure 4 Schematic representation of the NPC architecture	8
Figure 5 Methods to specifically label proteins in fixed cells	16
Figure 6 Methods to specifically label proteins in live mammalian cells.	20
Figure 7 Click-chemistry reactions to specifically label proteins.	24
Figure 8 SPIEDAC labeling of IR with Cy5 H-tet using different ncAAs.	37
Figure 9 Labeling of IR with Cy5 H-tet with and without permeabilization.	39
Figure 10 Schematic scheme of labeling method using Strep-AF 647	40
Figure 11 Labeling of IR with Cy5 H-tet and with streptavidin conjugated to AF 647.	41
Figure 12 Amber mutant expression of Nup98 in HEK293T cells.	42
Figure 13 Amber mutant expression and labeling of Nup98 in U2OSf cells.	43
Figure 14 SR imaging of Nup98 labeled with Strep-AF 647	44
Figure 15 A Schematic labeling method using high salt buffer and SF	45
Figure 16 Labeling of Nup153 in HEK293T cells using high salt buffer.	46
Figure 17 Labeling of Nup153 in HEK293T cells using SiR-dim-tet.	48
Figure 18 Expression of Nup153 in HEK293T cells with hydrophilic ncAA labeled with AF 647 H-tet.	48
Figure 19 SR imaging of Nup153 labeled with nanobodies conjugated to AF 647.	50
Figure 20 Dual-color SR imaging of Nup153 labeled with nanobodies conjugated to AF 647.	52
Figure 21 Dual-color SR imaging of Nup153 labeled with AF 647.	53
Figure 22 Schematic representation showing the general flow of the dual-color SRM and analysis	54
Figure 23 Representative images for each step of custom-written analysis.	55
Figure 24 Schematic representation of Nup98 and its labeled positions	57
Figure 25 Dual-color SR imaging of GFP-Nup98(TAG) labeled with AF 647.	58
Figure 26 Localizations of GFP-Nup98(TAG) labeled with AF 647 H-tet.	59
Figure 27 Synthetic generation of data based on localization precision.	60
Figure 28 Schematic representation of central FG Nups.	73
Figure 29 DNA-PAINT in COS-7 cells via GCE.	76
Figure 30 Recommended input settings of localization analysis for channel alignment	79
Figure 31 Recommended input settings for output options of localization analysis for channel alignment	80
Figure 32 Recommended input settings alignment of the two channels	81
Figure 33 Recommended input settings for localization analysis of two channels.	83
Figure 34 Recommended input settings for reconstruction of SR images	85
Figure 35 A sample bitmap image after setting the parameters for analysis.	87
Figure 36 Output window of the custom-written analysis.	88
Figure 37 Additional outputs of custom-written analysis	89

List of tables

Table 1 Fluorophores and their corresponding dichroic mirror and emission filters used in this study	31
Table 2 Analysis of different mutants of Nup98	57

Abbreviation list

aaRS: amino-acyl tRNA synthetase	racBCN: racemic mixture of bicyclo[6.1.0] nonyne lysine
AFM: atomic force microscope	ROD: reduction of dimensionality
APD: autopreolytic domain	SCO: strained cyclooctyne lysine
BOC: N- ϵ -t-butyloxycarbonyl-L-lysine	SF: synthetic fluorophore
CMV: cytomegalovirus	SIM: structured illumination microscopy
COS: CV-1 in origin with SV40 genes	SPAAC: strain-promoted alkyne-azide cycloaddition
DNA: deoxyribonucleic acid	SPIEDAC: strain-promoted inverse electron demand Diels-Alder cycloaddition
DO-TCO*: dioxo trans-cyclooctene lysine	SPANC: strain-promoted alkyne-nitrone cycloaddition
dSTORM: direct stochastic optical reconstruction microscopy	SR: super-resolution
DHFR: dihydrofolate reductase	SRM: super-resolution microscopy
eRF: eukaryotic release factor	SSIM: saturated structural illumination microscopy
EGFR: epidermal growth factor receptors	STED: stimulated emission depletion
EM: electron microscopy	TCO*: trans-cyclooctene lysine
FAP: fluorogen activating proteins	TMP: trimethoprim
FG: phenylalanine-glycine	TO: thiazole orange
FP: fluorescent protein	WGA: wheat germ agglutinin
FRC: fourier ring correlation	WT: wild-type
FWHM: full-width at half maximum	
GCE: genetic code expansion	
GDP: guanosine diphosphate	
GSDIM: ground state depletion microscopy followed by individual molecular return	
GTP: guanosine triphosphate	
HEK293T: human embryonic kidney 293T	
IBB: importin β binding	
IR: insulin receptor	
kDa: kilodalton	
MDa: megadalton	
MG: malachite green	
ncAA: non-canonical amino acid	
NES: nuclear export signal	
NLS: nuclear localization signal	
NL-SIM: nonlinear-structured illumination microscopy	
NBO: norbornene lysine	
NPC: nuclear pore complex	
NTR: nuclear transport receptor	
Nup: nucleoporins	
PAINT: point accumulation for imaging in nanoscale topography	
PALM: photoactivated localization microscopy	
POI: protein of interest	
PSF: point spread function	
pyIRS: pyrrolysine amino-acyl tRNA synthetase	

1 Introduction

1.1 Nuclear Pore Complex

Nuclear pore complex (NPC) is an immense macromolecular protein complex which functions as a primary gateway for selective transport of biomolecules through the nuclear envelope. It allows passive diffusion of small molecules (< 40 kDa, < 5 nm) while blocking large cargo molecules when not bound to nuclear transport receptors (NTRs). NPCs are composed of approximately 1000 protein molecules called nucleoporins (Nups) which assemble into a ~110 MDa complex.

Nups can be broadly categorized into different types according to their function. Membrane Nups containing transmembrane helices anchors the whole complex to the nuclear membrane [1], while central channel Nups are responsible for interacting with the NTRs to regulate nucleocytoplasmic transport. Scaffold Nups connect the central channel Nups to the membrane Nups [2]. Central channel Nups contain structured domains which enable them to interact with scaffold Nups. Many of these central channel Nups also contain intrinsically disordered regions which are composed of numerous FG repeats, called FG Nups, which interact with NTRs and play a major role in regulation of nucleocytoplasmic transport. Fiber-like Nups extend from the scaffold to the cytoplasm while on the nucleoplasmic side these form a basket-like structure [3].

The overall diameter of the whole complex has been determined to be 100-125 nm while the length along the transport axis including the emanating Nups and the nuclear basket is 150-200 nm. The diameter of the transport channel at its beginning in the nuclear and cytoplasmic side is 60-70 nm which narrows down to 25-45 nm at the center [2, 4, 5]. A simple schematic representation is shown below.

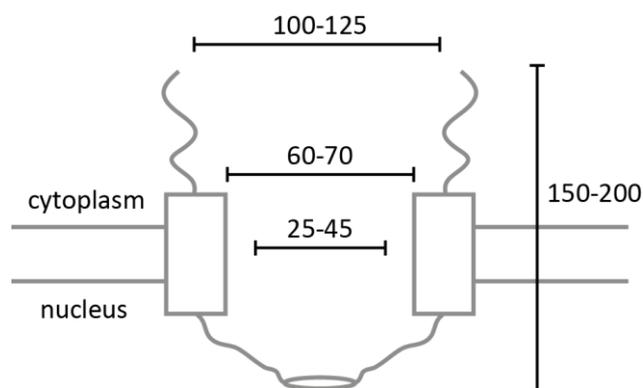


Figure 1 Schematic representation of the NPC and its sizes¹.

¹ The sizes are in nm.

1.1.1 Nucleocytoplasmic transport

Nucleocytoplasmic transport of cargo molecules through NPC can be divided into three steps: 1) formation of cargo complex, 2) transport of cargo complex and 3) dissociation of cargo complex. Additionally, the whole process requires a source of energy to provide directionality of the transport process. In this section, I will describe a classical import pathway involving NTRs to have a generalized overview of nucleocytoplasmic transport.

The first step of the classical import pathway involves the formation of the cargo complex in the cytoplasm. Cargo molecules possess nuclear localization signal (NLS) composed of a stretch of basic amino acids which gets recognized by NTRs. The most important group of NTRs (~100 kDa) belong to the family of β karyopherin proteins which are structurally characterized to contain HEAT repeats which are contained in proteins like huntingtin, elongation factor 3, protein phosphatase 2A, and the yeast PI3-kinase TOR1 [6]. These repeats are characterized by stacking of two anti-parallel α -helices linked by a flexible loop creating a helicoidal structure with two arches [7, 8]. Importin is one of these proteins containing HEAT repeats which are responsible for transporting cargo molecules into the nucleus. The cargo molecule with NLS forms a complex with Importin α and β in two different ways. The NLS of the cargo molecule can first interact with Importin α which then interacts with Importin β . An alternative path is also known which involves formation of complex between Importin α and β before interacting with the cargo [9]. The basic residues in Importin β binding (IBB) domain of Importin α interact with the acidic residues of Importin β . On the other side, the acidic residues of Importin α interact with the basic residues of NLS to form the whole cargo complex [10].

The second step involves actual translocation of the cargo complex through the central channel of NPC filled with FG Nups. Structural studies have shown that HEAT repeats 5 and 7 of Importin β contain a hydrophobic pocket that interacts with phenylalanine residue of FG repeat. Additional binding sites are also reported in HEAT repeats 14 – 16 [11]. *In vitro* experiments with purified FG repeats showed that phenylalanine plays a critical role during interaction with NTR. FG repeats which contained Ser or Ala residues instead of Phe didn't interact with NTRs while substitution to Tyr and Trp residues reduced the binding strength [12]. Single molecule (SM) fluorescence studies have shown that larger cargo molecules require higher number of NTRs compared to small cargo molecules and have observed the time scales of translocation events through NPC in the order of milliseconds [13]. Based on results from an integrative approach using SM FRET, molecular dynamics (MD) simulations, stopped-flow and NMR spectroscopy, the specific but fast interaction between FG Nups and NTRs was described by multivalent and low-affinity binding interactions in which specific conformations are not a prerequisite [14]. A number of models have been proposed in order to describe in details the interaction between FG Nups and cargo molecules. These models will be discussed in the next section.

1 Introduction

For dissociation of the imported cargo complex at the nucleus, RanGTP and CAS are both required to dissociate Imp β , Imp α and cargo molecules [15]. Ran protein drives the overall directionality of nucleocytoplasmic transport which will be discussed in details in the next paragraph. CAS, a protein initially discovered for its role in cellular apoptosis, has been known to be involved in recycling of Imp α back to cytoplasm after nuclear import of cargo complex [16]. When RanGTP binds to Imp β , it is known to cause conformational changes to Imp β to release Imp α /cargo complex [17]. Sun et al. have confirmed with fluorescence tracking experiments that CAS, most likely to be in complex with RanGTP, interacts with the released Imp α /cargo complex. They were able to observe either RanGTP/CAS/Imp α or Imp α /cargo complex after dissociation with the help of Nup50 [15]. Both the complex of Imp β and α as a complex on their own and travels back to the cytoplasm. The release of RanGTP from each Importin complex is mediated by RanGAP and RanBP. Then, the Importins are ready for another cycle of nucleocytoplasmic transport. Figure 2 schematically summarizes the nucleocytoplasmic import of cargo complexes and how each component is recycled.

Ran protein exists in nucleotide bound forms either with guanosine diphosphate or triphosphate (GDP or GTP). The concentration gradient of these two components drives the directionality of nucleocytoplasmic transport. Concentration of RanGTP is higher in nucleus while RanGDP has higher concentration in the cytoplasm [18]. RanGDP gets converted to RanGTP by a protein called RCC1 in the nucleus to retain the concentration gradient, while, RanGAP is involved in the conversion of RanGTP to RanGDP in the cytoplasm [19]. RanGTP is required for transport of cargo complex out of the nucleus. It is also required to dissociate the cargo complex once it arrives in the nucleus. Transport of RanGDP from the cytoplasm to nucleus is accompanied by an NTR called NTF2 [20].

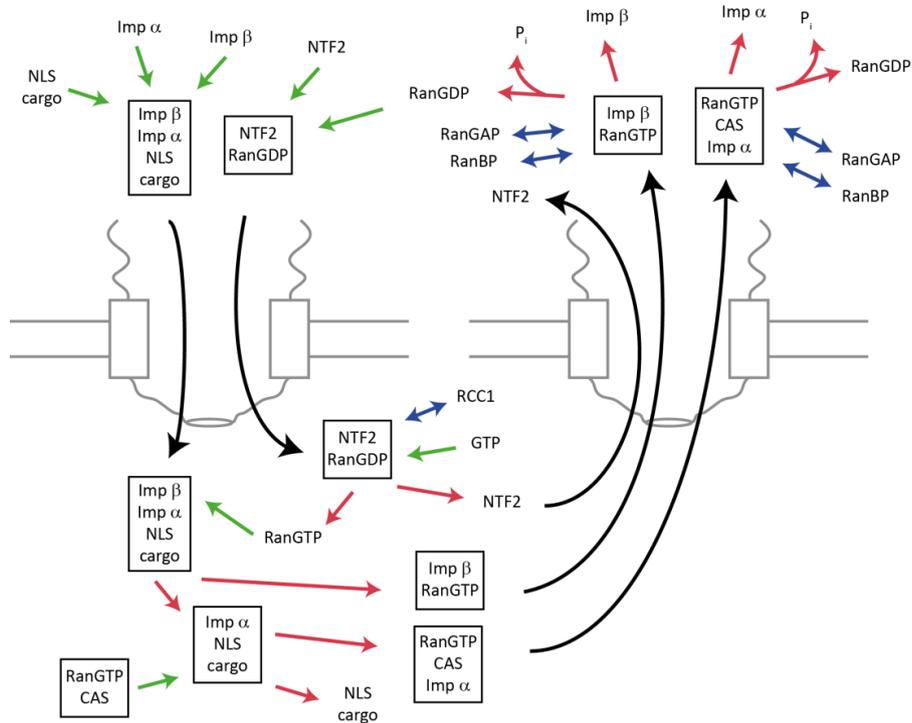


Figure 2 Schematic diagram of the import and Ran cycle

The color of the arrows represent association (green), dissociation (red), interaction (blue) and translocation (black) of proteins and complexes (shown in boxes).

1.1.2 Models of nucleocytoplasmic transport

Ever since the discovery of the NPC, numerous models have been proposed in order to explain nucleocytoplasmic transport. The models heavily debate on the structural arrangement of FG Nups in the central channel due to the critical role in nucleocytoplasmic transport. In the following section, I will point out the similarities and differences of the models.

Selective phase/hydrogel model

The model was initially proposed by Ribbeck et al. in 2001, based on cell-based assays which show that the central channel of the NPC might have a “semi-liquid” phase which regulates nucleocytoplasmic transport [21]. Afterwards, the same group was able to form a hydrogel using FG domains of a single Nup and showed that the gels were able to recapitulate the permeability barrier of the NPC [22]. It was observed that the hydrogel allowed the entry of small molecules while blocking cargo molecules unless bound to NTRs. The hydrogel in the absence of FG repeats

were observed not to function as a permeability barrier. The authors proposed that the hydrophobic FG domains interact with each other forming a sieve-like barrier. The mesh size was imagined to be big enough for small molecules (< 5 nm) to pass through. While, The NTRs with cargo molecules disrupted the FG interactions to overcome the barrier to translocate through the NPC[23].

Polymer brush model

Lim et al. suggested that the selectivity barrier is based on entropic fluctuations of the disordered regions in FG nups [24]. Experiments were performed using atomic force microscopy (AFM) to observe the collective behavior of FG Nups tethered to gold surface. These Nups exhibited long-range repulsive force blocking big cargo molecules not interacting with NTRs. Further experiments showed that FG domains of Nup153 collapsed into compact molecular conformations when exposed to NTRs [25]. Thus, during nucleocytoplasmic transport the reversible collapse of the central channel will provide path for the cargo molecules with NTRs to pass through. Based on recent results from high-speed AFM, the authors propose a transient highly dynamic interaction inside the central channel of the NPC [26].

Virtual gating model

In this model, Rout et al. describes the movements of FG Nups similar to the polymer brush model [27]. The diffusive movements of the disordered regions of FG Nups are suggested to form an entropic barrier, restricting the flow of molecules through the NPC. The transport of NTR/cargo molecules through the narrow NPC channel means its entropy has to be lowered due to restricted movements during the translocation. The model proposes that binding of the NTR with the FG Nups in the central channel will lower the entropic barrier which will favor transport of NTR/cargo molecules.

Reduction of dimensionality (ROD) model

The model is originally based on the shape of the NTRs and the localization of the binding sites which are extended rather than localized. The simplified ROD model takes into account the collapsing property of FG domain when interacting with NTRs in order to explain the high transport speed of cargo molecules [28]. Moreover, based on the observation that under physiological conditions NTRs were bound in NPC in large numbers, the FG Nups are predicted to form a ~10 nm thick FG layer at the wall of the NPC [8]. This wall of FG domains is then used by the NTRs to move through the transport channel by two-dimensional random walk [29]. The model also mentions that FG Nups are not involved in rejecting cargo molecules that are not bound to NTRs.

Forest model

The model proposes two distinct zones of traffic in the transport channel based on physical properties of FG domains of Nups purified from yeast. The model categorizes FG domains in two different types [30]. The first type is proposed to have a collapsed coil configuration with low content of charged amino acids. Based on *in vitro* experiments, this type formed low affinity binding interactions with one another. On the other hand, the second type of FG domain is proposed to contain more charged residues and the configuration was more dynamic (e.g. extended-coil) than the first type. The first type is referred to as the “shrubs” and the second type as “trees” due to their resemblance to the shapes. Smaller cargo molecules bound to NTRs will more likely to interact with the shrubs while bigger cargo molecules like ribosomal nucleoproteins with NTRs will interact with trees at the center of the transport conduit.

In summary, Figure 3 shows a schematic representation of the abovementioned models. The models all agree that binding or interaction of NTRs to FG repeats is a prerequisite for nucleocytoplasmic transport. However, the models disagree on specific details on the molecular mechanism of the interaction. This is due to lack of experimental techniques to address this controversy. Thus, many assumptions of the models have been primarily based on *in vitro* experiments.

It is of main interest in this study to directly visualize the spatial organization and interaction of FG Nups with NTRs *in situ* to shed light to some of the most fundamental mechanisms about nucleocytoplasmic transport. In the next sections, I will discuss how different methods have made an attempt to visualize spatial organization of Nups and how some other techniques can be improved further to complement the existing methods regarding their temporal and spatial constraints.

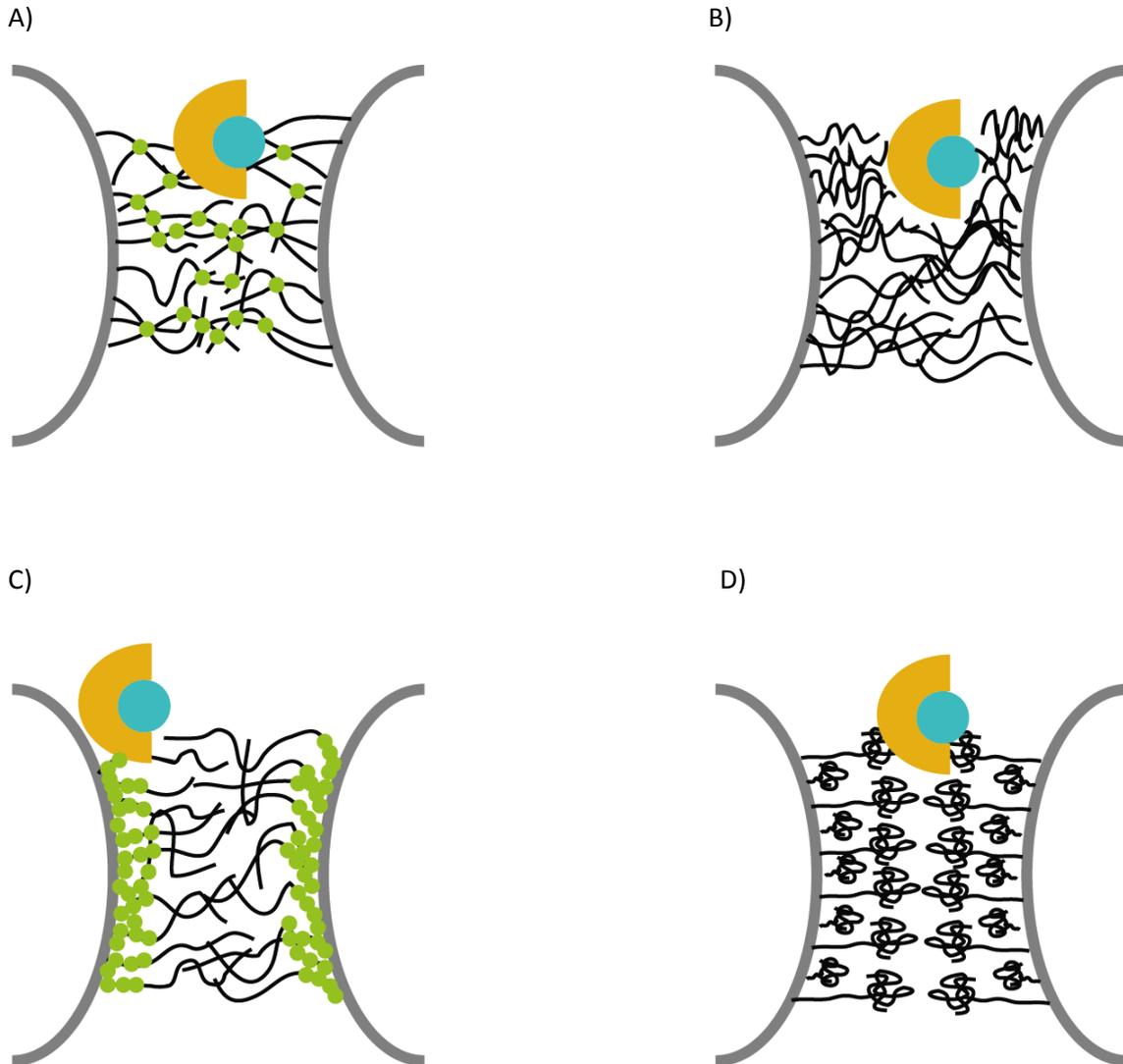


Figure 3 Different models proposed to explain the nucleocytoplasmic transport

A) Selective phase, B) Polymer brush / Virtual gating, C) Reduction of dimensionality and D) Forest model (green circles: interacting FG repeats, blue circles: cargo molecules, yellow boat-like structures: NTRs).

1.2 Methods to study the spatial organization of Nups in the NPC

In here, I will summarize the latest results obtained by different techniques that tried to solve the giant structural puzzle of the NPC. The results mentioned in this section complement each other in their common goal to visualize the whole structure of the NPC. I have grouped the results according to the implemented method to emphasize how these methods complement each other and how their technical limitations prevented further application to the FG Nups in the central channel of the NPC. As of now, structural information regarding the disordered FG Nups filling the central channel of the NPC remains elusive. Figure 4 shows the latest simplified representation of the NPC's molecular architecture based on Kosinski et al.

Among the implemented techniques, SRM stands out among the rest due to its compatibility with live cells and its versatility in application to intrinsically disordered proteins like FG Nups as demonstrated in some *in vitro* and in-cell experiments [31-34]. Thus, more emphasis will be given to explaining the basic principles behind SRM and the specific type of SRM which is suitable for studying the spatial organization of FG Nups in the NPC.

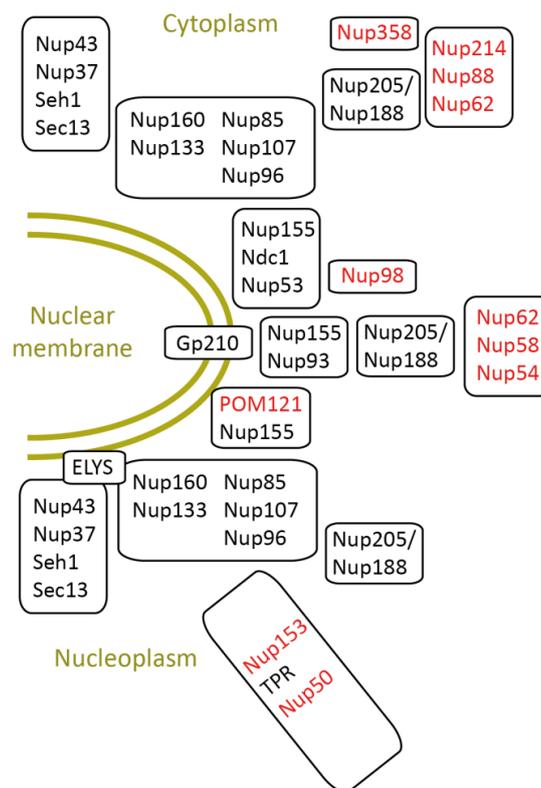


Figure 4 Schematic representation of the NPC architecture

This scheme is based on Kosinski et al. [35]. All the Nups containing at least one FG repeat is in red.

1.2.1 Cryo-electron microscopy

Electron microscopy (EM) is one of the primary tools used by scientists as an alternative to light microscopy. Due to the use of a much shorter wavelength of electron beams compared to visible light, it has been possible to reach a resolution of 0.7 nm with protein samples in as early as 1970s [36]. EM allowed the imaging of the protein complexes but the resolutions of these images were not good enough to reach atomic-resolution of the whole complex. To minimize damaging of molecules by the electron beam, images were acquired with frozen biological samples which reduced the effect of radiation damage [37]. Since then this technique has been combined with tomography to acquire three-dimensional structures of numerous proteins and their complexes [38]. Recent advancements in electron detection and image processing have improved the resolution of the electron density map of the NPC. This has enabled fitting of the isolated Nup subcomplexes into the acquired electron density map to have an overview of the molecular architecture of the NPC. Integration of techniques such as cross-link mass spectrometry and molecular modeling with cryo-EM have enabled fitting of X-ray crystallography structures which gave new insights into the spatial organization of Nups in the NPC. The drastic improvement in the resolution makes it an ideal tool to study protein structure; however, it only provides static images of a protein structure. It usually gives very limited amount of information on flexibility and conformational dynamics of intrinsically disordered proteins. In this section, I will summarize, in brief, the most recent findings about the spatial organization of Nups in the NPC using cryo-EM.

Early discoveries about overall architecture of NPCs gave insights on the overall dimensions of the whole complex as well as the transport channel [39]. It is now widely accepted that the core framework of the NPCs is composed of three stacked rings. The inner ring is located at the merging point of inner and outer nuclear membranes, while the nuclear and cytoplasmic ring sandwich the inner ring from each side. Nups containing transmembrane α -helix domains firmly anchor the stacked rings at the nuclear membrane [1].

Each of the two distal rings is composed of 16 copies of Nup107-160 subcomplex (Y-complex) forming two concentric rings (8 copies per ring) slightly shifted from each other [40]. The cytoplasmic ring interacts with Nup214-Nup88-Nup62 containing Rae1 and Nup98 and Nup358-RanGAP1*SUMO1-Ubc9 subcomplex which functions to assemble and disassemble the NTR-cargo complex [41]. A subcomplex of Nup153-Nup50-TPR binds to the nuclear ring which also interacts with the NTR-cargo complex during nucleocytoplasmic transport. The resolution of a recent tomogram of the NPC by Bui et al., allowed fitting of the available X-ray crystallography structures of some of the Nups with spatial restraints derived from results of crosslinking mass spectrometry [40]. Von Appen et al. were able to identify specific interactions of cytoplasmic filament Nup214 with Nup188 (part of inner ring) together with the role of Nup358 in stabilizing the cytoplasmic ring structure through gene silencing experiments [42]. Nup155 was identified as a potential candidate to connect nuclear and cytoplasmic rings with the help of Nup205/188 [35].

1.2.2 Atomic force microscopy

AFM is a scanning probe technique using a sharp tip to detect interaction forces over the sample surface with piconewton sensitivity [43]. It is possible to acquire topological information of biological molecules or complexes in aqueous buffer eliminating the need of fixing, staining or labeling the biological sample. Moreover, due to recent technical developments, it is now possible to observe dynamic biological molecules with a temporal resolution of ~100 ms [44]. One of the main applications of AFM is quantifying intra- or intermolecular forces to acquire insights on molecular interactions [45]. The main downside of AFM is that the probe has to physically approach the target sample. It is impossible to acquire information on intact cellular compartments or biomolecules which hinders contact with the scanning probe. It would be only possible to acquire data on isolated biological molecules or complexes.

AFM has been continuously implemented to study the NPC *in situ*. Changes in the topology of NPCs isolated from *Xenopus laevis* oocyte have been observed to be affected by NTRs and amphipathic alcohols [46]. The central channel has especially been of interest due to its critical role in nucleocytoplasmic transport. Bestembayeva et al. have reported that the central channel forms a cohesive polymer network in which the cohesiveness varies significantly depending on the radial position in the channel [47]. Their experimental work was also confirmed with an NPC model which contains a central condensate of Nups in the central channel. High-speed AFM has enabled the observation of spatiotemporal dynamics of FG Nups in the NPC of *X. laevis*. The results show that FG Nups are constantly in motion and forms short-lived entanglements (< 500 ms). For its application in studying the NPC, it is most likely limited to topology and indirect quantification of molecular interactions of Nups.

1.2.3 Super-resolution microscopy

The resolution of an optical microscope is limited due to diffraction of light. Ernst Abbe has mathematically defined resolution (R) or the minimum resolvable distance as such:

$$R = \frac{\lambda}{2NA} \quad (1)$$

When imaging a point object with visible light (~550 nm) using a high NA objective (1.4), R would be ~200 nm. It will not be possible to distinguish point objects closer than ~200 nm. Abbe proposed the equation for the case of oblique illumination with a transmission microscope; however, it can also be applied to fluorescence microscopy [48].

Recent technical advancements in the field of SRM have overcome this diffraction barrier. Visualization of components in protein complexes in the regime of less than 100 nm has become possible [34]. This has been achieved by exploiting the on and off switching properties of synthetic fluorophores (SFs) [49]. There are two general ways to achieve this. One way is to excite fluorescent molecules with the use of spatially patterned beams (e.g. SIM, STED). The second way

is to image single molecules sequentially (PALM, STORM). Different techniques of SRM will be discussed in the following section; focusing on how the properties of localization based SRM makes it an ideal tool to implement compared to other methods. I will also summarize on how far these technologies have been applied in studying the spatial organization of Nups in the NPC and on how some technical limitations prevents further application in FG Nups inside the central channel of the NPC.

Structured illumination microscopy (SIM)

SIM is based on the principle of Moirè effect which implies that overlapping two patterns of high spatial frequency can generate a structure with low spatial frequency [48]. The key component of SIM is an optical component to produce patterned excitation with high frequency. The other pattern of high frequency comes from the object details. Thus, when the two high frequency components overlap it leads to a higher resolution image after computational analysis. An epifluorescence microscope together with an optical component containing a fine grating is then used to excite the biological sample decorated with SFs or fluorescent proteins. Around five-fold improvement to the resolution (~50 nm) has been reported by saturating the fluorescence emission with a strong intensity of excitation light [50]. Saturation of the SFs is a source of nonlinearity which is exploited during computational analysis.

Stimulated emission depletion (STED) microscopy

The main principle behind STED to improve resolution is based on switching off the fluorescence of a population of molecules surrounding the target molecule to be observed which are in a distance closer than the diffraction limit. This switching off mechanism is done by having an additional beam, STED beam, with a doughnut shape in which then the center is overlapped with the excitation beam [51]. The depleting / switching off population are nonlinearly dependent on the STED beam. Then together with the excitation beam it is possible to produce a PSF with a full width half maximum (FWHM) of down to typically ~20 nm when working with biological samples [51]. The maximal resolution achievable can be quantitatively described by the following equation:

$$R_{STED} = \frac{\lambda}{2NA \sqrt{1 + I_{max}/I_{sat}}} \quad (2)$$

where I_{max} is the intensity of the STED beam and I_{sat} is saturating intensity for the target fluorophore. It can be seen that infinite increase of I_{max} / I_{sat} will infinitely decrease the resolution. In order to perform STED in biological samples, lower intensity of the STED beam is required to minimize phototoxicity issues which will result in higher resolution. Specialized buffer will improve the behavior of the fluorophores but it is not required. In terms of fluorophores used, some of the fluorescent proteins are compatible but optimal performance is reached with SF.

Localization based SRM

Localization is based on fitting of a mathematical function to a single emitter to determine its position in an image [52]. These positions then can be overlaid after numerous iterations of emission and fitting. Each localized position is then convolved with a Gaussian function to yield a super-resolution (SR) image. The key concept of localization based SRM lies in the ability to separate fluorescence emission from a single emitter with respect to time scale. In an ideal experimental setting, a biological sample would be densely decorated with fluorescence emitters. Localization based SRM is possible only when the fluorescence is detected from just one emitter. This can be done in two ways. One way would be to send all of the emitters to a dark state by a high intensity excitation beam and then only a subset of these emitters will spontaneously come back to the emitting state at a given time to be excited by the excitation beam (e.g. dSTORM, GSDIM). The second way would be done by using another beam to activate only a subset of emitters which can be excited with a second beam to emit fluorescence (e.g. STORM, PALM) [53].

In order for the fluorescence emitters to go to the dark state and blink spontaneously, they have to be exposed to a reducing environment. Direct stochastic optical reconstruction microscopy (d-STORM) and ground state depletion microscopy followed by individual molecular return (GSDIM) both requires imaging of fixed cells with a reducing agent and oxygen scavengers to prevent photobleaching of the fluorescence emitters [54]. Initial attempts of dual-color imaging required two different excitation sources and two different sets of optical filters for both colors [55].

Stochastic optical reconstruction microscopy (STORM) and photo-activated localization microscopy (PALM) both require a shorter wavelength beam to activate the emitter in the dark state. The activated emitter can be excited with a second beam usually longer wavelength than the activation beam for detection of fluorescence. The main difference between STORM and PALM is the fluorescence emitter. STORM uses SFs while PALM is based on genetically encoded fluorescent proteins.

The effective resolution of localization based SRM is mainly affected by two factors, namely, localization precision and labeling density of the fluorescent probe [56]. Localization precision pertains to how well the position of a fluorescence emitter can be estimated. This precision can be determined theoretically and is affected by the background noise, shot noise, pixelation noise. The background noise comes from the experimental conditions and the performance of the detection system including detector noise, readout errors, sample drift and background fluorescence. The pixelation noise stems from the uncertainty of the exact position of photon arrival within a single pixel. Lastly, the shot noise comes from the discrete nature of the detected photons [57, 58].

Taking all these factors into consideration, the localization precision of an ideal detector can be estimated with the following equation,

$$\delta^2 = \frac{s^2 + a^2/12}{N} + \frac{8\sqrt{\pi}s^4b^2}{a^2N^2} \quad (3)$$

where a is the pixelation noise or the pixel size, b is the background noise and s is the standard deviation of the Gaussian function.

The labeling density also affects the effective resolution of the final image. The relationship is described by Nyquist sampling criterion wherein it states that, at least two labeled position within the distance of the resolution must be present. Thus, in order to achieve a resolution of 20 nm a fluorophore must be present at least every 10 nm. Although it is clear that higher labeling density will improve the resolution, recent studies have shown that Nyquist sampling criterion may suggest overestimation of resolution in terms of labeling density [59].

In order to set a standard for determining the resolution of localization based SRM, recent studies report on adopting Fourier ring correlation (FRC) from electron microscopy which allows determination of resolution directly from the final localized positions accounting for localization precision and labeling density [56, 60]. The method can provide a more realistic resolution estimate of the acquired SR image as it takes into account all the experimental factors such as drift and labeling density. Banterle et al. have shown that the proposed method gives resolution values comparable to the ones based on localization precision [56]. Nevertheless, this method is not able to take into account the effect of high blinking rates which will mask the contribution of the low labeling density of the fluorophore.

Recent progress in application of SRM in studying spatial organization of the NPC

SRM techniques have contributed significantly in providing clues to the overall architecture of the NPC. Rego et al. used nonlinear or saturated structural illumination microscopy (NL-SIM or SSIM) to look into Nup98 and POM121 tagged with fluorescent protein at N- and C-terminus, respectively. POM 121 was shown to form a ring-like structure which has a diameter of 40-70 nm while Nup98 was shown to form a dot-like structure [50]. Higher resolution was achieved by using stimulated emission depletion (STED) microscopy and SFs with outstanding photophysical properties and stability [61]. In here, they were able to localize transmembrane Nup Gp210 which showed a ring-like structure with a diameter of ~160 nm and the central channel of the NPC located at the center of the ring-like structure using a different SFs. The localization of Gp210 was confirmed by another study using localization based SRM [62]. In this study, an additional channel was added to look into the central channel of the NPC using wheat germ agglutinin (WGA) which binds to glycosylated Nups. One study looked into the arrangement of the Y-complex in the NPC using localization based SRM. By labeling different components of the Y-complex with their corresponding antibodies, it was possible to conclude that Nup160-Nup37 arm stretches to the periphery of the ring while the Nup85-Seh1 arm protrudes towards the center of the ring-like structure [34]. In a recent study by Platonova et al., they show a proof of concept for simultaneous dual-color imaging using a single

1 Introduction

light source via spectral demixing [63]. They image Nup43 or Nup62 with Nup153 with nanobodies labeled with CF680 and AF 647. Nup62 is localized to the central channel of the pore while Nup43 and Nup153 is similarly localized forming a ring-like structure.

To sum up all the methods mentioned in this section, it is clear that SRM has the greatest potential to be used in studying spatial organization of Nups in the central channel due to its non-invasiveness, live-cell compatibility and high temporal and spatial resolution. Among different SRM methods, STED and localization based SRM seem to have the potential for further applications when considering the spatial resolution that can be achieved. However, for STED the phototoxicity has to be taken into account. On the other hand, for localization-based SRM, it is much easier and convenient to technically implement than STED, especially for dual-color imaging. Moreover, previous studies also proved the compatibility of the method in studying spatial organization of Nups in the NPC. In addition, the method will allow quantitative analysis of the localizations which is not possible with other SRM methods.

Nevertheless, localization-based SRM has been limited to studying only scaffold Nups due to technical limitations. The main advantages of the method are not fully exploited in studying the disordered regions of FG Nups in the central channel of the NPC. To overcome this limitation, it is vital to develop a method to tag SFs to the disordered regions of FG Nups with residue specificity. Moreover, the spatial organization of the disordered regions of FG Nups is unpredictable, thus, a reference point is needed for comparison between different regions and Nups.

1.3 Methods to site specifically label proteins in mammalian cells

The discovery of fluorescent proteins (FPs) has enormously contributed to numerous insights in fundamental processes of life by tracking, measuring and visualizing proteins in living cells [64]. The protein of interest (POI) can be fused to FPs available in different colors. Recent advancements in the field of SRM have generated motivations to develop probes for such applications. Photoswitchable proteins and SFs with outstanding photophysical properties were developed and are still under development for their application in SRM. However, more attention has been given to the application of SFs (~1 kDa) due to their small size compared to FPs (~30 kDa) which can actually perturb the native localization and function of POI. As mentioned in the previous section, the major hurdle for application of SRM in studying the disordered regions of Nups in NPC is the labeling method. In this section, I will discuss different methods of labeling proteins with SFs in mammalian cells. Comparisons will be made in terms of specificity, size, speed and ease of application to biological systems. From these comparisons, it will be clear why genetic code expansion (GCE) in combination with click-chemistry is the only possible way, at this moment, to label disordered regions of Nups in the NPC of mammalian cells with residue specificity.

1.3.1 Labeling methods after fixing cells

The most classical method to label proteins inside cells is via immunolabeling using antibodies that recognize the POI. The method has been a primary choice of labeling especially when the primary antibody is commercially available [65]. The general protocol of the method involves a two-step process. Firstly, the cells are fixed and permeabilized or vice versa, in order for the primary antibodies to penetrate into the interior of the cells to interact with the POI. Then, the secondary antibody usually decorated with SFs is added to tag the primary antibody. Applications in SRM have been possible but localization differences due to the size of the two antibodies (150 kDa, ~10 nm each) must be taken into consideration during analysis [53]. Moreover, due to spatial constraints when working with large antibodies, the labeling density will be limited and thus directly affect the overall resolution in the final image.

In line with the localization artifact from the size of the antibodies, attempts have been successful to provide smaller antibodies. One is by using antibody fragments, which gives comparable efficiency to the full antibody but are about 3-fold smaller (50 kDa, ~9 nm) [66]. Another possible solution is camelid antibodies which are much smaller (13 kDa, ~4 nm) called nanobodies [67]. Different nanobodies for different protein domains can be produced and labeled with SFs.

Another method that has exploited the use of antibodies and oligonucleotides together to overcome photobleaching of SF is the DNA-PAINT method. The acronym stands for point accumulation for imaging in nanoscale topography using DNA strands labeled with SFs [68]. The main principle of the method is based on transient interactions between two complementary strands of DNA. One strand is bound to a target position (docking strand) while the other strand

(imager strand) is freely diffusing with a certain concentration. It is possible to tune the affinity between the two strands by controlling the concentration and the number of nucleotides. When the imager strand is bound to the docking strand the fluorescence is detected and when it is not bound there is no fluorescence detected. These states of on and off of fluorescence due to binding and unbinding interactions make it possible to analyze and acquire SR images. In order to label proteins with this method, it is only possible when the docking strand is bound to the POI. As of now, it has only been shown to work with antibodies which were covalently labeled with the docking strands [68, 69]. However, any other method of tagging the POI with docking strand should in principle enable SRM. Schematic representation of the abovementioned methods are shown in the figure below.

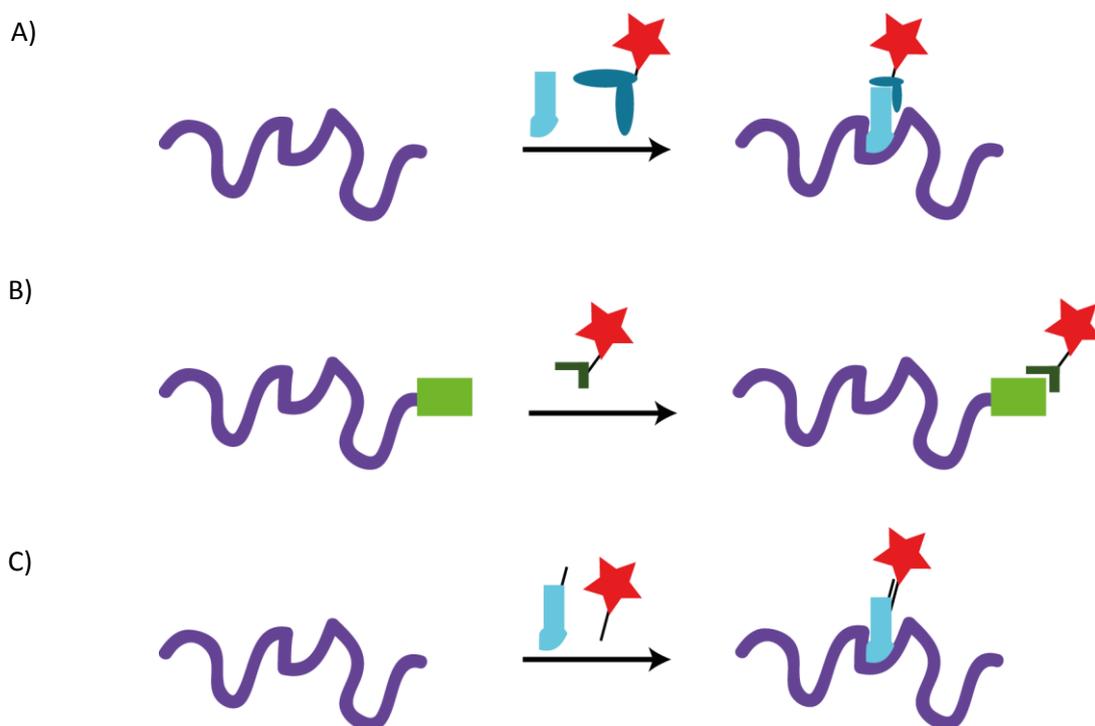


Figure 5 Methods to specifically label proteins in fixed cells

A) Labeling via immunostaining (light blue: primary antibody, dark blue: secondary antibody), B) labeling via nanobody for FP (light green: FP, dark green: nanobody with covalently bound SF), C) DNA-PAINT with immunostaining (light blue: primary antibody with docking strand, SFs is bound to imager strand). Red stars: SFs. This figure is not drawn to scale.

1.3.2 Labeling methods for live cells

1.3.2.1 Self-labeling tags

In this method, a peptide/protein tag is fused to the POI which is recognized by a SF modified with a small molecule. The SFs used in this method are cell permeable which enables live cell labeling. SNAP/CLIP-tag and Halo-Tag are the most popular examples which are also commercially available. SNAP/CLIP tag is derived from the enzyme, O^6 -alkylguanine-DNA-alkyl transferase, which recognizes benzylguanine/cytosine molecules attached to SFs [70]. For Halo-tag, it is derived from Haloalkane dehalogenase and the molecule involved is chloroalkane. Studies have shown their outstanding applications in labeling proteins in living mammalian cells [71, 72]. However, due to their size (~25-30 kDa), it is tagged to either the N- or C-terminus of the POI and thus may also perturb the localization and function of the POI.

Another group of self-labeling tags which is much smaller than the aforementioned protein tags have also been shown to work inside living cells. These tags are based on a short peptide usually 6-8 amino acids containing a tetracysteine core which gets recognized by fluorescein derived SFs containing arsenic (III) molecules [73]. The minimal size of the tag makes it an ideal tool to use to minimize perturbation to the POI. However, the main downside of this method is that it suffers from off-target labeling of thiol containing biomolecules within the living mammalian cell [74].

1.3.2.2 Enzyme mediated tag

The general principle of this method works by introducing an enzyme which recognizes a specific peptide sequence fused to the POI. This enzyme then covalently attaches a substrate molecule to one of the amino acids that is part of the special peptide sequence. Among several of the developed methods, one of them uses *E. coli* lipoic acid ligase. The target POI is fused with a 13 amino acid peptide sequence that is recognized by the enzyme [75]. The SF is ligated to the lysine residue which is part of the fused peptide sequence. The most recent developments in this method have enabled ligation of SFs for SRM. Liu et al. was able to predict mutations to the lipoic acid ligase to be able to accommodate SF, resorufin, for the ligation reaction. They were able to show SR images of not only cellular membrane proteins but also intracellular cytoskeletal proteins [76]. It has to be noted that only SFs that are small enough to be recognized and accommodated by the ligase can be used to tag the POI.

1.3.2.3 Non-covalent tags

A smaller protein derived from *E. coli* was also shown to work in living mammalian cells. An enzyme called dihydrofolate reductase (eDHFR, ~18kDa) is known to non-covalently interact with trimethoprim (TMP). In a recent study, a SF (ATTO655) was modified with TMP and used to image human histone H2B protein in living cells [77].

1 Introduction

Fluorogens are non- or weakly fluorescent molecules which become highly fluorescent when its excited structure is stabilized by conformational or molecular changes [74]. Szent-Gyorgyi et al. have reported that SFs like thiazole orange (TO) and malachite green (MG) increase their fluorescence emission a thousand-fold when they interact with fluorogen activating proteins (FAPs) derived from single-chain antibody fragment (~25 kDa) [78]. Due to the fluorogenicity, background signal is drastically minimized. The availability of cell permeable SFs has also enabled intracellular labeling. However, due to the size of FAP which is not much different compared to other FPs, this method may also perturb the localization of labeled proteins.

Other method utilizing specific interaction between oligonucleotides and proteins has also been demonstrated. These oligonucleotides called aptamers which have similar size compared to nanobodies can be produced for a POI [66]. The aptamers can be conjugated to SFs to be used for labeling of POI. However, the specificity of the interaction between the aptamer and the POI has hindered its widespread use for labeling proteins in mammalian cells. Continuous improvements have been made to enable live cell labeling of transferrin receptor localized at the plasma membrane of mammalian cells [79]. However, if the protein doesn't have any secondary or tertiary structure it is not possible to find a suitable aptamer to label the POI.

It can be seen that the methods are quite similar to each other in terms of overall concept. Factors such as commercial availability and compatibility of the SFs with SRM will be considered before choosing the method for an experiment. However, it has to be noted that all of these aforementioned methods are mostly limited to labeling of proteins only either at the N- or C-terminus which limits its applicability in studies where amino acid specificity is required.

1.3.2.4 *Genetic code expansion technology*

The most versatile and powerful method to label proteins inside living cells with amino acid specificity is by using non-canonical amino acids (ncAAs) via GCE technology [80]. Basically, labeling of proteins in mammalian cells via GCE involves two steps. First step requires incorporation of ncAAs which contain reactive functional groups to the POI. The second step is the actual chemical reaction wherein the reactive functional group of the ncAA is exploited to covalently react with the SF. These site-specific modifications pave way for unprecedented insights to the structure, function and activity of proteins in cells. There are two main ways to apply GCE technology to incorporate ncAA in the POI. One way is to take advantage of the promiscuity of the endogenous amino-acyl tRNA synthetase (aaRS) and the other way is to introduce an orthogonal pair of aaRS/tRNA. The first method is ideally suited for studying the whole proteomics of a cell or an organism while the latter one for protein specific studies [81]. Basic principle of the first method is based on the competition of an ncAA with one of the natural amino acid to be charged by the endogenous aaRS for its cognate tRNA. Recent developments have enabled the system to be applicable in mammalian cells. Mahdavi et al. have shown incorporation of azidonorleucine to replace methionine in hamster, monkey and human cell lines [82].

1 Introduction

The second method of incorporating ncAA specifically to the POI was pioneered by Schultz and coworkers [83]. The overall efficiency of the system depends on the orthogonality of the aminoacyl tRNA synthetase (aaRS) and tRNA. This is usually achieved by introducing the aaRS/tRNA pair from a different organism to the host translational machinery. Due to its orthogonality, the introduced aaRS will not aminoacylate any of the native tRNAs but only its cognate tRNA. This aminoacylated tRNA is designed to recognize a blank codon, usually Amber stop codon, which is introduced to the gene of the POI by site-directed mutagenesis beforehand. Moreover, the orthogonal tRNA should not be charged by any other aaRS except for its cognate orthogonal aaRS. The technology is now shown to be working not only in unicellular organisms and cell culture but also in animals such as nematodes and flies [84].

For application of GCE technology in mammalian cells, initial approaches involved adoption of aaRS/tRNA system discovered in yeast and *Bacillus stearothermophilus* [85, 86]. Further attempts in this direction led to discovery of an orthogonal aaRS/tRNA system from archae *Methanosarcina* which is involved in incorporation of pyrrolysine, a lysine derivative with pyrroline ring, in response to Amber stop codon [87]. For the expression of tRNA, introduction of an extragenic RNA polymerase III promoter such as U6 promoter, allowed its expression in mammalian cells. It was further shown that it is possible to design and evolve the pyrrolysine aaRS (pylRS) from *Methanosarcina mazei* to recognize ncAAs with bulkier functional groups useful for fluorescence labeling [88, 89]. The repertoire of ncAAs was further expanded with additional efforts in synthetic chemistry which enabled incorporation of ncAAs with cyclooctene/cyclooctyne functional groups [88, 90, 91].

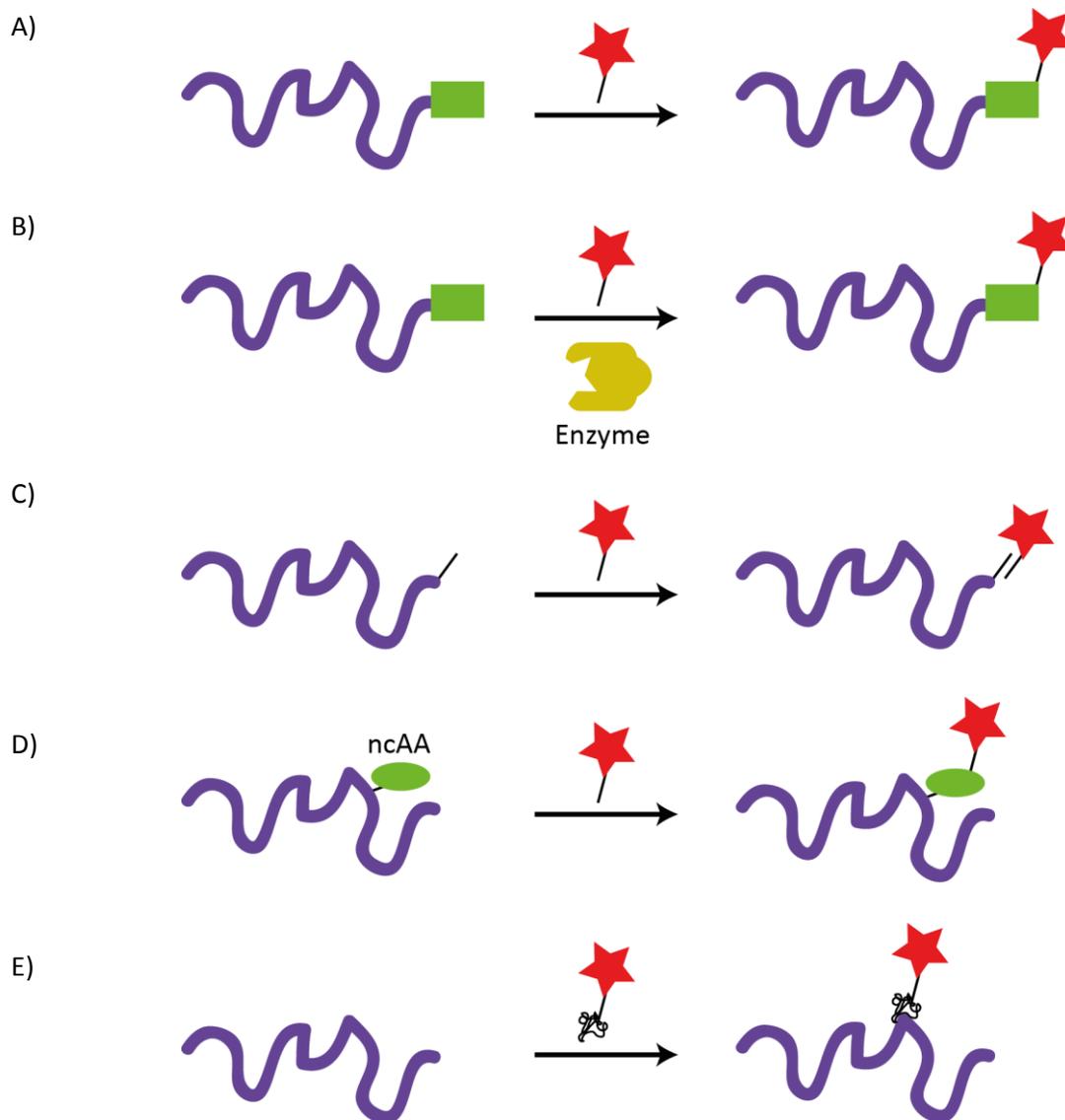


Figure 6 Methods to specifically label proteins in live mammalian cells.

A) Self-labeling tags (green rectangle: a fused enzyme or peptide sequence), B) enzyme mediated tags (green rectangle: peptide sequence recognized by an enzyme), C) non-covalent tags, D) labeling using ncAA through GCE and E) aptamer tags. red star: SF. This figure is not drawn to scale.

In the meantime numerous ncAAs have been developed to be used in different experimental designs. Some recent applications include but are not limited to posttranslational modifications [92], photocrosslinking [93, 94], photocaging [95, 96] and fluorescence spectroscopy and microscopy [97]. Moreover, direct incorporation of fluorescent ncAAs has also been introduced. However, these fluorescent ncAAs are not compatible with SRM, thus limiting their applications in cellular imaging. In order to fully exploit the chemical handles introduced by GCE, it is necessary to find a way to covalently bind a SF compatible with SRM to the incorporated ncAAs.

1.3.3 Click-chemistry reactions for labeling proteins

Introduction of chemical handles to proteins using GCE is only the first step toward residue-specific labeling with SFs. The next step is covalent reaction between the chemical handle of ncAA and the SF. There are a countless number of chemical reactions available for formation of a covalent bond, however, only a handful of these reactions are actually versatile, simple, highly selective, fast and high yielding [98, 99]. Click-chemistry was the term coined by Sharpless [98] to describe these reactions. Moreover, click-chemistry reactions which require and produce non-toxic substances are referred to as biocompatible reactions.

In the following section, I will compare some of the click-chemistry reactions which have been showing promising results in terms of labeling protein *in vitro* and *in situ*. Side by side comparison of these reactions will give a clear picture on their potential use in numerous applications.

1.3.3.1 Staudinger ligation (azide-phosphine reaction)

Azide moiety has been one of the most popular functional groups for click-chemistry reactions especially in biological systems. Its small size and stability and inertness in physiological conditions make them an ideal candidate for labeling proteins in mammalian cells. One of the initial labeling reactions involving azide functional groups uses phosphine modified SFs which also contains an ester group. The nucleophilic phosphorus attacks one of the nitrogen of the azide group which makes the nitrogen nucleophilic. Then the nitrogen will form a bond with the acyl group cleaving the bond with the phosphorus releasing the phosphine oxide (Fig. 5). However, the slow rate of the reaction ($10^{-3} \text{ M}^{-1} \text{ S}^{-1}$), stability of the phosphine group and aqueous solubility have hindered its wide application in labeling cellular proteins in mammalian cells [100].

1.3.3.2 Azide alkyne cycloaddition reaction

A chemical reaction to form a triazole through cycloaddition between azide and alkyne was first mentioned more than 50 years ago [101]. This reaction found numerous applications when a catalyst, Cu (I) was found to speed up the reaction 6 orders of magnitude than the reaction without the catalyst [102]. It has been shown to work in labeling not only for proteins but also other biomolecules such as lipids and nuclei acids [103, 104]. The reduced Cu (I) can be added directly or it can also be generated on-site from Cu (II) by using reducing agents such as sodium ascorbate.

The main advantage of doing so is that Cu (II) is cheaper and that the reduced form will be higher in purity than Cu (I) salts [105]. However, due to intrinsic property of Cu (I) generating reactive oxygen species which are toxic to cells, applications to label proteins in mammalian cells have remained challenging. To overcome this toxicity issue, water-soluble ligands which promote formation of activated copper catalyst even at low concentration of Cu (I) were used to readily catalyze the cycloaddition reaction [106, 107]. These ligands also act as reducing agents to minimize the toxicity effects of ROS. Another alternative way to improve the performance of Cu (I) at low concentration is by increasing the effective concentration at the reaction site by using copper chelating azides. The principle was shown to work not only in labeling surface proteins in living cells [108] but also in cytoplasm of *E.coli* [109].

The concept of adding a ring-strain to the alkyne group to speed up the cycloaddition reaction was already proposed long time ago [110, 111]. The reaction is termed as strain-promoted alkyne-azide cycloaddition (SPAAC). It was revisited by Bertozzi and coworkers in 2004; they synthesized a molecule with a cyclooctyne group and biotin on the other end for confirmation. It was shown to label proteins glycosylated with azido sugars in Jurkat cells analyzed by flow cytometry [100]. From here more compounds with improved stability and reactivity were produced which improved the reaction rate with second-order rate constant of up to $1 \text{ M}^{-1} \text{ s}^{-1}$. These compounds were shown to label azide-bearing glycans in not only mammalian cells but also in *C. elegans* and zebrafish embryos [112, 113].

1.3.3.3 *Cyclooctynes 1,3-Nitrone dipoles cycloaddition*

Other functional groups were also discovered which reacted faster than the azide group with cyclooctyne rings. Ning et al. show that substituted nitrone group gives up to 30 times faster reaction rates compared to azide functional groups [114]. The reaction was termed strain-promoted alkyne-nitrone cycloaddition (SPANAC). Further application of the method showed application in labeling of cellular surface protein epidermal growth factor receptors (EGFRs) in mammalian cells [115]. However, due to the stability of the nitrone group towards hydrolysis, it still needs further improvements before it can be applied to intracellular proteins.

1.3.3.4 *Strained alkene/alkyne tetrazine reaction*

Chemical molecules containing tetrazine functional group has been known to react with a variety of unsaturated compounds with the release of nitrogen in as early as 1950s [116]. Since then it has been determined that the tetrazine acts as the electron poor diene while the alkene/alkyne group acts as the electron-rich dienophile. The reaction between the two groups is termed as inverse-electron-demand Diels-Alder cycloaddition [117]. Sauer et al. reported that strain has a dramatic effect on the reaction rate improving the second-order rate constant up to $12700 \text{ M}^{-1} \text{ S}^{-1}$ [118]. Further experiments in aqueous solution showed similar rates [119]. Stability of the tetrazine functional group has also been improved over the years [120]. Devaraj et al. show application of

1 Introduction

the cycloaddition reaction to commercially available primary antibodies by modifying it with trans-cyclooctyne and modifying a SF with the stable benzylamino tetrazine. Live A549 cells were first exposed to the primary antibody then labeled with the tetrazine modified SF [121]. Due to its selectivity and speed, numerous applications have been reported recently which image small molecules in living cells [122] and labeling of cells with nanoparticles [123, 124].

To sum this up, among all the click-chemistry reactions discussed in this section, one reaction seems to be the most appropriate one for intracellular labeling of proteins in mammalian cells. Inverse electron demand Diels-Alder cycloaddition reaction between strained alkene/alkyne (SPIEDAC) and tetrazine shows the best performance in terms of speed, biocompatibility and bioorthogonality. Previous studies have already noticed the potential capabilities of GCE and click-chemistry and thus have applied it to label membrane and cytoskeletal proteins but not in low density proteins such as Nups in the NPC of mammalian cells [33, 90]. Membrane proteins are straightforward to label due to the exposure of its labeling sites outside of the cell. Thus, the SF does not need to be cell-permeable and will not contribute to any off-target labeling inside the cell. Cytoskeletal proteins have been a classical target for intracellular labeling due to high expression level and distinctive structural features. However, for other intracellular proteins, it has remained challenging mainly due to low abundance and off-target labeling of the SFs.

As mentioned in the previous section, the major hurdle for application of SRM to study spatial organization of FG Nups in NPC of mammalian cells is the lack of labeling method with amino acid specificity. Recent achievements in developing new ncAAs and SFs have paved the way for novel applications of GCE and click-chemistry in numerous biological systems. The main interest of this study is to develop a labeling method by optimizing GCE and click-chemistry to label disordered regions of FG Nups in mammalian cells. This will give access to unbiased visualization of structural arrangement of disordered regions which has never been even tried before due to technical limitations. Major challenges include off-target labeling and lack of quantitative method to assess the spatial organization of disordered regions in the central channel of NPC.

1 Introduction

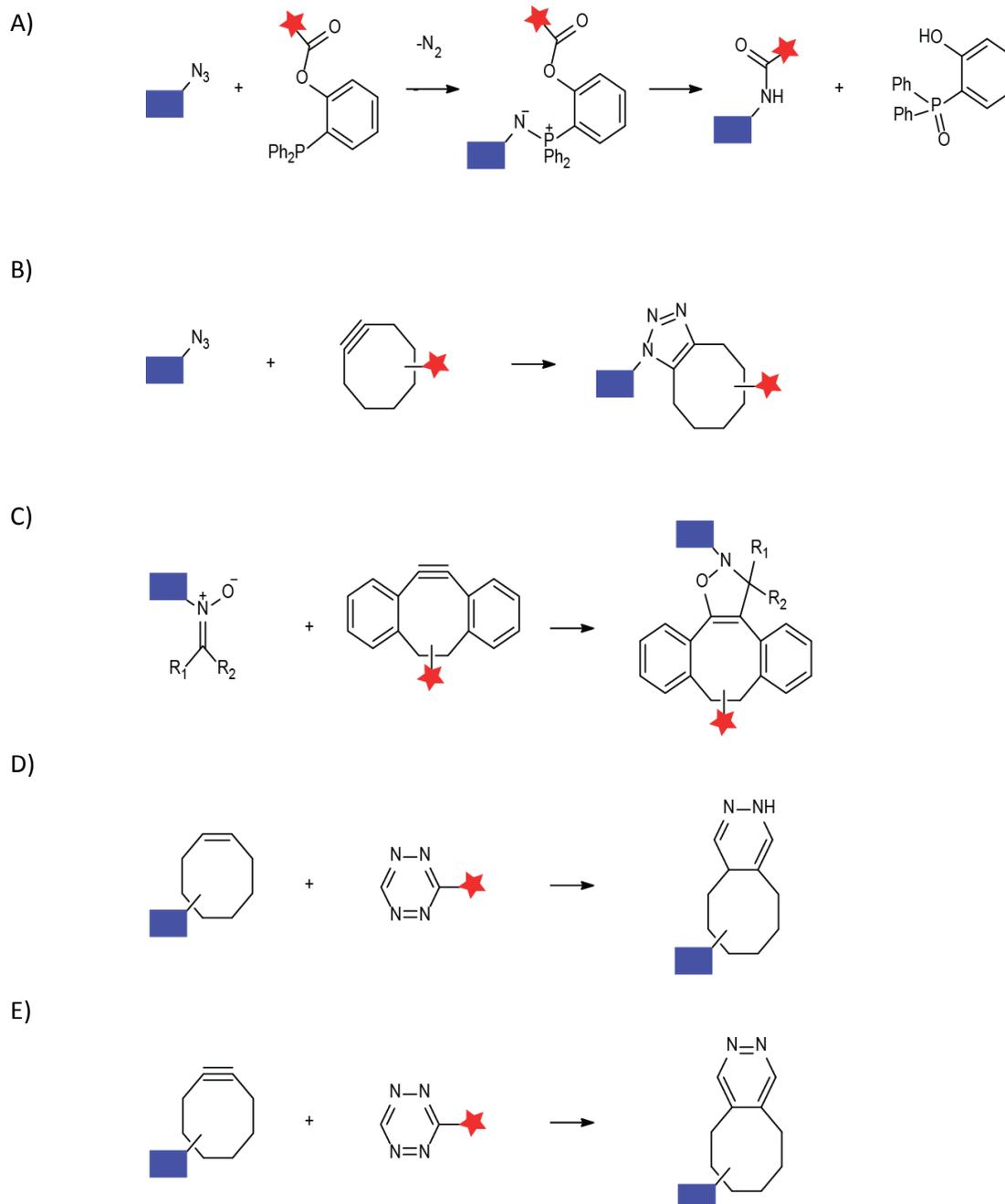


Figure 7 Click-chemistry reactions to specifically label proteins.

A) Staudinger ligation, B) Strain promoted azide-alkyne cycloaddition, C) Cyclooctyne 1,3-Nitrone dipoles cycloaddition, D,E) Strain promoted inverse-electron demand Diels-Alder cycloaddition (blue square: protein, red star: SF).

1.4 Objectives

The main aim of this thesis is to develop a strategy to visualize the spatial organization of FG Nups inside the central channel of the NPC. This will provide novel insights to understand how the transport of biomolecules across the nuclear envelope is regulated. Despite numerous advantages of implementing SRM in studying large protein complexes like the NPC, it has not been possible to study the spatial organization of the flexible Nups in the main transport conduit. One of the major bottlenecks in SRM has been the lack of labeling methods to tag proteins site-specifically with small SFs. Thus, this thesis is inclined towards overcoming of limitations in the currently emerging labeling method in order to label the disordered regions of FG Nups *in situ* in mammalian cells with amino acid specificity. Another challenge in using SRM to study FG Nups in the NPC arises from the need to quantitatively assess and compare labeled positions in order to finally map the spatial organization of the FG Nups. Therefore, I also aim to develop an unbiased multi-color imaging method and a robust analysis pipeline to effectively evaluate the different labeled positions of site-specifically labeled FG Nups.

The main aim of this thesis is split into three specific aims, following a bottom up approach from lower to higher complexity:

1.4.1 Site-specific labeling of membrane proteins in mammalian cells

A currently emerging and versatile method for labeling proteins with amino acid specificity uses GCE combined with click-chemistry. However, the application of this labeling method in mammalian cells is underdeveloped. There are numerous ncAAs and SRM compatible SFs available but their pros and cons for different application has not yet been assessed. I aim to test and evaluate different ncAAs and SFs to give useful insights in choosing the most appropriate combination for labeling proteins in mammalian cells. In line with the bottom up approach, the target POI will be a membrane protein to avoid additional complications that could arise from interactions between SFs and intracellular biomolecules.

1.4.2 Site-specific labeling of Nups inside the NPC in mammalian cells

Once the most appropriate pair of ncAA and SF has been chosen and established for labeling of membrane proteins in mammalian cells, I aim to site-specifically label Nups *in situ* in mammalian cells using GCE and click-chemistry. Much attention will be paid to overcome and circumvent background signal arising from interaction of SFs or ncAAs with intracellular biomolecules other than the target Nups. The labeling method will be cross-validated by determining the localization of a Nup for which its spatial organization inside the NPC is known.

1.4.3 Visualize the spatial organization of FG Nups in mammalian cells

Once the labeling protocol has been validated, I aim to develop a straightforward method to effectively compare the localization of different labeled positions of FG Nups in the NPC. This comparison and evaluation of the labeled positions will be done by introducing a reference point in each experiment. Each labeled position will have a certain distance from the reference point which can be used to compare with other labeled positions. In the end, this will enable the visualization of the spatial organization of FG Nups *in situ* in mammalian cells with residue precision.

2 Materials and methods

In this chapter, I will describe the details on how the results in Chapter 3 were obtained and analyzed.

In section 2.1, I will describe the procedure for handling cell culture and transiently transfecting the cells with plasmids.

In section 2.2, I will give details on the procedures of labeling proteins including how to functionalize the SFs.

In section 2.3, I will describe the custom-built microscope setup that was used for the acquisition of raw data for SRM.

In section 2.4, I will elaborate on the specific details of performing dual-color SR imaging including spectral demixing and custom-written analysis.

2.1 Mammalian cell culture and transfection

Human embryonic kidney 293T (HEK293T, ATCC CRL-3216) cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Life Technologies, cat. no. 41965-039) supplemented with 10% v/v fetal bovine serum (FBS, Sigma, cat. no. F7524), 1% Pen-Strep (Sigma, cat. no. P0781), 1% L-glutamine (Sigma, cat. no. 158127) and sodium pyruvate (Life Technologies, cat. no. 11360). Cells were cultured at 37°C in a 5% CO₂ atmosphere. For imaging, the cells were seeded in a 4/8-well Lab-Tek II chambered coverglass (Thermo Scientific Nunc, cat. no. 155383) pre-coated with poly-L-lysine (PLL, Sigma, cat. no. P5899) at least 4 h before seeding the cells. A stock solution (1 mg/ml) dissolved in ddH₂O was diluted 1:100 and 300 µl/200 µl of this solution was added to an well of 4/8-well Lab-Tek II.

A stable cell line generated by Beck group (HEKFlpIn 153KD) was used to label Nup153, a FG Nup located at the cytoplasmic side of the NPC. This cell line was cultured in the same conditions as HEK293T cells. The cell line expresses enhanced green fluorescent protein (GFP) fused to Nup153 and microRNA which lowers the level of endogenous Nup153. Both expressions were induced with 1 µg/ml of tetracycline at least 48 hrs before transfection. Transfections were performed using JetPrime reagent (Polyplus-transfection, cat. no. 114-15) according to the manufacturer's recommendations.

2.2 Labeling of proteins in mammalian cells using GCE and click-chemistry

2.2.1 Labeling of IR at the surface of mammalian cells

In here, I describe a protocol for labeling IR(676TAG)-GFP at the surface of HEK293T cells. The cells were seeded in a 4-well Labtek II at a density of 110,000 cells/well with a total volume of 500 μ l. The wells have been treated with PLL beforehand as explained in section 2.1. After incubating the cells at 37 °C / 5 % CO₂ for 15-17 hrs, the cells were transfected. For transient transfection, it is also recommended to check the protocol from the manufacturer. In this experiment, 0.5 μ g of the plasmid encoding IR(676TAG)-GFP and 0.5 μ g of plasmid encoding pylRS and tRNA^{pyl} were used. The pylRS contains Y306A and Y384F mutations to allow recognition of ncAAs with cyclooctenes/octynes [89]. If the viability of the cells is greatly reduced after transfection then the amount of plasmid can be lowered to 0.3 μ g per plasmid. The ncAA was prepared and supplied to the cells right after adding the transfection mix. The optimized final concentration of ncAA is 250 μ M. The ncAA was buffered with HEPES (Life Technologies, cat. no. 15630) in a ratio of 1:3 (ncAA : HEPES). The cells were then incubated with the transfection mix and the ncAA for 8 hrs. Afterwards, it was replaced with 500 μ l of medium without ncAA. Then, the cells were incubated for 14-16 hrs in the incubator (37 °C / 5 % CO₂).

For SPIEDAC reaction of sulfonated Cy5 tetrazine (sulfo-Cy5 H-tet, Jena Bioscience, cat. no. CLK-015), a final concentration of 1.5 μ M is recommended. The SF mixture was prepared by diluting 0.9 μ l of 500 μ M stock solution (50 nmol in 100 μ l of DMSO) of sulfo-Cy5 H-tet in 300 μ l of medium. The SF mixture (300 μ l / well) was added to the cells and incubated at 37 °C / 5 % CO₂ for 10 mins. After the labeling reaction, the SF mixture was removed and fresh medium was added. Then, after 2 hrs of incubation at 37 °C / 5 % CO₂, the cells were washed with PBS and fixed for 10 mins with 2 % paraformaldehyde (PFA) in PBS.

All the ncAAs mentioned in this thesis can be purchased from SiChem GmbH (www.sichem.de). SCO (strained cyclooctyne lysine), TCO* (*trans*-cyclooctene lysine) mix (mixture of axial and equatorial isomers), NBO (norbornene lysine), racBCN (racemic mixture of bicyclo [6.1.0]nonyne lysine), endoBCN and exoBCN. Stock solutions of these ncAAs (100 mM) were prepared by dissolving them in 15 % (vol/vol) DMSO/0.2 M NaOH. Detailed characterization of the two isomers of TCO* (axial and equatorial) is further elaborated in Hoffmann et al. [125].

2.2.2 Intracellular labeling of Nups in mammalian cells

HEK293T and HEKFlpIn 153KD cells were seeded at a density of 110,000 and 200,000 cells, respectively, per well in a 24-well plate with a total volume of 500 μ l per well. Then, after 15-17 hrs, transfections were performed with the plasmid of the target Nup and pylRS/tRNA^{pyl}. The ratio of the plasmids was kept at 1:1 with a total amount of DNA at 0.6-1 μ g per well. After 4-6 hours, the transfected cells were supplied with fresh growth media containing 250 μ M of ncAA. The cells

2 Materials and methods

were then incubated with the nAA for 36-48 hrs. Subsequently, the cells were trypsinized and transferred to an 8-well Lab-Tek II chambered coverglass at a density of ~100,000 cells per well with a total volume of 250 μ l per well. Lab-Tek II was treated with PLL before seeding. The cells were left to settle in the incubator for at least 3 hrs before proceeding to labeling and fixation. Before labeling the cells were permeabilized with Digitonin for 2 mins. The working solution of Digitonin is prepared by diluting the stock solution (20 mg/ml in DMSO) 1:500 in high salt buffer. The composition of high salt buffer is as follows: 200 mM Hepes, 1.1 M potassium acetate (KOAc), 50 mM sodium acetate (NaOAc), 20 mM magnesium acetate (MgOAc), 10 mM ethylene glycol-bis(β -aminoethyl ether)-N,N',N'-tetraacetic acid (EGTA), pH 7.3. After permeabilization the cells were rinsed for two times with high salt buffer.

Subsequently, the cells were labeled with 1.5 μ M of Alexa Fluor 647 H-tet (AF647) for 10 mins at room temperature (RT) diluted in high salt buffer. For dual-color imaging, a reference marker for the center of the NPC was used. The cells were then rinsed two times with PBS and then, Wheat germ agglutinin (WGA) labeled with CF 680 (Biotium, cat. no. 29029) at a final concentration of 1 μ g/ml was added to the cells for 10 mins at RT. After a series of washes with PBS for at least 3 times, the cells were fixed with 2% PFA for 10-15 mins.

Labeling of Nups using GFP nanobody was done as a control to confirm integration of Nups expressed via Amber suppression in the NPC. The procedure for labeling with nanobody was also done after treatment with high salt buffer. After permeabilization, instead of adding SF solution, only high salt buffer was added and incubated for 10 mins. Then after fixation, the cells were blocked with a 5 % (w/v) solution of BSA (bovine serum albumin) in PBS for 30 mins. Then nanobody conjugated to SFs diluted in 2.5 % of BSA in PBS was introduced to the cells at a final concentration of 50-200 ng/ml for 90 mins.

2.2.3 Labeling of IR and Nups with streptavidin conjugated to AF 647

For labeling IR with streptavidin conjugated to AF 647 (Strep-AF 647), the protocol becomes different from section 2.2 only before addition of the SF mixture. Instead of adding sulfo-Cy5 H-tet, the linker molecule H-tet biotin (Conju-probe, cat. no. CP-6001) was added at the same final concentration of 1.5 μ M. After fixing the cells, streptavidin (Thermo Fisher, cat. no. S21374) was added at a final concentration of 100-500 nM diluted in 2.5 % of BSA/PBS .

For labeling Nups, the cells were first permeabilized with Digitonin (as in section 2.2.2). After permeabilization, 1.5 μ M of H-tet biotin solution (in 10x diluted high salt buffer) was added to the cells and incubated at 37 $^{\circ}$ C / 5 % CO₂ for 10 mins. Then, after washing the cells with PBS for 3 consecutive times, the cells were then fixed with 2 % PFA for 10-15 mins.

2.2.4 Live-cell intracellular labeling of Nups with SiR

After seeding and transfecting the cells as in section 2.2.2, the cells were supplied with the fresh medium in the presence of 250 μ M of ncAA. After 36-48 hrs of incubation, the cells were washed with fresh medium 2-3 times in 6-9 hrs. The cells were then labeled with 800 nM of SiR-dim-tet dissolved in medium in the presence of 0.1 % (w/v) pluronic F-127 (Sigma, cat. no. P2443) for 20-30 mins. Afterwards, the cells were rinsed 3 times over 2 hrs with fresh medium. The cells were then fixed with 2 % PFA for 10-15 mins and imaged with the microscope. The extrusion of SiR-dim-tet can be minimized by supplying the cells with 10 μ M of verapamil (Sigma, cat. no. V4629) during the labeling reaction.

2.2.5 Synthesis of AF 647 H-tet from NHS

The conversion of AF 647 NHS (N-hydroxysuccinimide) to H-tet can be done in a single step reaction. The synthesis procedure was derived from a patent application publication from Hilderbrand et al. [126]. The main reagents were AF 647 NHS (Thermo fisher, cat. no. A20006) and Amine-tetrazine (Jena Bioscience, cat. no. CLK-001), which are both commercially available. Firstly, a solution of Amine-tetrazine (10 mmol) in anhydrous DMF (0.5 ml) which will give a total concentration of 20 M was prepared. Then 2.5 mmol of AF 647 NHS and 10 mmol of triethylamine (Sigma, cat. no. 471283) were added directly. After shaking overnight in the dark at RT, the final product was isolated by preparative reverse phase HPLC using a gradient of acetonitrile.

2.3 Experimental setup

A custom-built TIRF microscope was used to perform localization-based SRM. The excitation lasers used are Omicron 405 nm 120 mW (Coherent Inc., cat. no. LuxX 405-60), Omicron 488 nm (Coherent Inc., cat. no. LuxX 488-200), Sapphire 568 nm 100 mW (Coherent Inc., cat. no. Sapphire 568 LP) and Omicron 660 nm (Coherent Inc., cat. no. LuxX 660). Beam expanders (Qioptiq, cat. no. 4401) of 7x for the 405 nm laser and 10x for other lasers were used to increase the beam diameter. Each excitation source except 660 nm had its corresponding dichroic mirror to combine the beams in a single path, namely a FF458 (Semrock Inc., cat. no. FF458) for the 405 nm laser and FF495 (Semrock Inc., cat. no. FF495) for the 488 laser and FF605 (Semrock Inc., cat. no. FF605) for the 568 nm laser. Then, the polarizations of the beams were changed to circular using a retarder plate (Qioptiq, cat. no. G362021491).

To be able to illuminate in the TIRF mode, the circularized beam was placed on a linear stage (PI, cat. no. M505) which was used to shift the focus of the beam towards the side of the back-pupil of a UAPON 100X TIRF, NA 1.49 objective (Olympus GmbH). The beam was focused to the objective by a 300 mm focal distance lens (Thorlabs, cat. no. AC508-300-AML). To separate emission from excitation light a quad-edge dichroic beamsplitter (Semrock Inc., cat. no. F-48-890) was used.

2 Materials and methods

For the emission pathway, a notchfilter (Semrock Inc., cat. no.NF01-568/647) was first used to filter out residual excitation light. Then a three-lens system (Qioptiq, cat. no. G322246000 f=180 mm, G322388000, f=120 mm and another G322246000, f=180 mm) was used to focus the emission light to EMCCD (Andor Technology, cat. no. DU-897, iXon X3 EMCCD). Multicolor imaging was performed by using a dichroic mirror and an additional EMCCD (Andor Technology, cat. no. DU-897D, iXon X3 EMCCD) to increase the field of view and minimize aberrations. The following table shows the pairs of fluorophores used in this study and their corresponding dichroic mirror and emission filters

Table 1 Fluorophores and their corresponding dichroic mirror and emission filters used in this study

Pair of Fluorophores	Dichroic mirror	Short wavelength emission filter	Long wavelength emission filter
GFP, AF647	AHF 690	ET 525/50 Chroma	ET 700/75 Chroma
AF647, CF680	AHF 690	ET 700/100 Chroma	ET 700/100 Chroma

A piezo-driven stage was used to move the Lab-Tek II to image different areas of the well while the objective positioner was used to move objective in z-axis to find the focal point. The focus was stabilized by implementing an electronic feedback loop (proportional-integral-derivative controller) to constantly stabilize the position of the objective with respect to the coverslip of the Lab-Tek II. A near-infra red laser (Coherent Inc., cat. no. LuxX 905) was focused at the side of the back-pupil of the objective for total internal reflection at the coverslip which was detected with the photodiode (Thorlabs Inc., cat. no. PDP90A). The change in the focus of the objective will change the detection position of the totally reflected beam which will trigger the feedback loop to stabilize the position of the objective. The different components of the microscope were controlled through custom-written Labview (National Instrument) software.

All the confocal images were acquired with a Leica SP8 STED microscope from the Advanced Light Microscope Facility (ALMF) at EMBL. The microscope was equipped with HCX PL APO 100x/1.4 Oil objective (Leica, Germany). Emission detection range for GFP is 500-550 nm while for Cy5 and AF 647 is 650-700 nm.

2.4 Dual-color single-molecule localization microscopy (SMLM) imaging

2.4.1 Single-molecule imaging procedure

For multicolor imaging, a robust approach was adopted based on spectral demixing. In this approach, a single laser was used to excite CF 680 and AF 647 which are spectrally similar SFs. Both of these SFs provided high performance in terms of brightness and duty cycle in a conventional

2 Materials and methods

blinking buffer. A dichroic beamsplitter as shown in Table 1 was implemented in the emission pathway to separate the two SFs into two different EMCCDs. Cross-talk of emissions led to localization of an emitting SF in both cameras with a ratiometric intensity distribution. This ratio of intensity which is a unique photophysical property of a SF was used to distinguish between CF 680 and AF 647.

The blinking buffer for the SFs was composed of 50 mM Tris (pH 8.0), 10 mM sodium chloride (NaCl), 0.5 mg/ml glucose oxidase (Sigma, cat. no. G0543), 40 μ g/ml catalase (Sigma, cat. no. C3155), 10 % (w/v) glucose (Sigma, cat. no. D9559) and 10 mM β -Mercaptoethylamine (MEA) (Sigma, cat. no. M9768), stored in aliquots of 100 mM in PBS and adjusted to pH 7.4 with hydrochloric acid, HCl). This buffer was exchanged after approximately 60-90 mins of imaging.

By taking into account the brightness of the GFP and the shape of the nuclei, the appropriate cells were selected for imaging. The cells were first illuminated with the maximum power of 660 nm light (~ 1 kW/cm²) to bring both of the SFs into the dark state, typically about 10-20 s. When spontaneous blinking of SFs were observed, then the laser power was reduced to 0.3-0.5 kW/cm² (0.5 A.U. in the LabView program) to acquire a long series of 25,000-30,000 frames with an exposure time of 30 ms. The gradual increase of 405 nm back-pumping laser was done manually to keep the average number of localizations per frame constant (50-100). Initial data analysis for localization and spectral demixing was done with rapidSTORM [127].

2.4.2 Spectral demixing with RapidSTORM

Details on using RapidSTORM for spectral demixing and reconstruction of a SR image are given in section 6. In here, I will discuss about the general outline of the analysis method. The first step of the analysis involves software-based linear alignment of the two acquisition channels. Although the two channels were manually aligned with fluorescent beads before acquisition, adjustments in the range of few nanometers were necessary which can only be done with the software. Both channels were analyzed for localization of SFs then these localizations in each frame were compared to align one channel with respect to the other. The offset values were usually less than 1 μ m.

In the second step, SF intensities of all the localizations were compared for the two channels to determine the ratio. This value depended on the dichroic mirror and the emission filter used in the microscope setup. For the setup used in this thesis, the ratio of CF 680 was ~ 0.2 in the shorter wavelength channel and ~ 0.8 in the longer wavelength channel, while for AF 647 the ration was ~ 0.5 in both channels. After assigning each localization, then two files with positions of each SF were stored as an output.

The last step involved generation of a SR image based on the two localization files. The pixel size for the reconstruction was 10 nm which is based on the resolution or the localization precision of the acquired results (27-34 nm) determined by FRC [56] and Localizer package [128]. However,

RapidSTORM does not provide the option to convolve each localization with a Gaussian filter. Thus, the reconstructed images were convoluted with a Gaussian function (FWHM = 30 nm) using ImageJ .

2.4.3 Custom-written automated evaluation of localization²

The localized positions of the labeled Nups were then analyzed with a custom-written code in Igor Pro. While, the localized positions of WGA were used to determine the center of the NPC. First step towards this is recognizing real fluorescence spots from background which has been of great interest since the advent of fluorescence microscopy.

In this thesis, we implemented “à trous wavelet transform decomposition” of the original image which is based on a pattern recognition algorithm to differentiate fluorescence spots from background noise. Details of the concept are discussed in a study done by Olivo-Marín [129]. The localized spots were analyzed further by fitting a 2D Gaussian function to determine the circularity. A threshold was set to eliminate poorly- or over-labeled NPCs. On the other channel, a clustering algorithm based on density (DBSCAN) [130] was implemented to identify clusters of SFs from labeled Nups. Afterwards, a small window of approximately the size of the NPC was designated around the center of the NPC coming from the WGA channel. Each cluster was analyzed to assess if it fits to the window. This process eliminated analysis of SFs and ncAAs which are non-specifically sticking. Then, all the localizations were combined to give a distribution histogram that can be analyzed to determine the localization of the labeled position with respect to the center of the NPC. A step-by-step guide on how to handle the program can be found in the Appendix section.

² The analysis program was developed together with Daniel Spitz, Lemke group, EMBL.

2 Materials and methods

3 Results

In this chapter, I show the results that I have obtained in order to achieve the three specific aims described in section 1.4. These were highly interdependent and essential to achieve the final goal of this thesis which is to visualize the spatial organization of FG Nups *in situ* in mammalian cells using SRM.

In section 3.1, I evaluated the different ncAAs and commercially available SFs in order to fully develop site-specifically labeling proteins in mammalian cells. For this purpose, I used a membrane protein, the insulin receptor (IR), as a target protein.

In section 3.2, I implemented the most appropriate pair of ncAA and SFs, based on the information gathered from the experiments performed in section 3.1, for labeling Nups *in situ* in mammalian cells. Intracellular labeling of Nups resulted in low signal-to-noise ratio due to background; therefore, I also show the results on identifying and overcoming the sources of the background.

In section 3.3, I developed an imaging strategy which implements dual-color SRM and a custom-written analysis in order to quantitatively map the spatial organization of FG Nups inside the NPC. The imaging strategy and the analysis were first validated with Nup153 and further implemented to study the spatial organization of Nup98 *in situ* in mammalian cells.

3.1 Labeling of IR in live mammalian cells using click-chemistry³

With the aim of labeling Nups inside the NPC, it was first necessary to assess the available click-chemistry reactions in order to choose the most appropriate one. To do this, I adopted and established an optimal protocol to label IR in live mammalian cells. It was of great interest in this experiment to test and evaluate different commercially available ncAAs systematically, which were previously developed in the lab [88, 90], containing cyclooctene and cyclooctyne functional groups by site specifically labeling membrane protein in live mammalian cells.

In here, the experiment was designed to focus on the click-chemistry reaction while excluding other complications especially from the background of the SFs inside the cell. This was possible by having two experimental conditions. One condition was the use of IR, a membrane protein which is known to trigger a variety of biochemical reactions which define biological action of insulin. It was an ideal protein to demonstrate the proof-of-concept for any fluorescent labeling method due to its size, complexity and most of all, the exposure of its structure to the exterior of the cell. Thus, background signal was much easier to avoid as there were much less biomolecules interacting with the SF compared to the interior of the cell. Another condition was by using a SF modified with sulfo

³ The experiments in section 3.1 were done in collaboration with Ivana Nikić and Gemma Estrada Girona, Lemke group, EMBL

3 Results

groups. Due to its negative charge, sulfonated SF was not able to penetrate the cellular membrane. Thus, it was an appropriate SF to focus on the click-chemistry reaction that will occur on the cellular membrane. With these conditions, a straightforward and systematic comparison of different commercially available nCAAs and functionalized SF (sulfo-Cy5) was done.

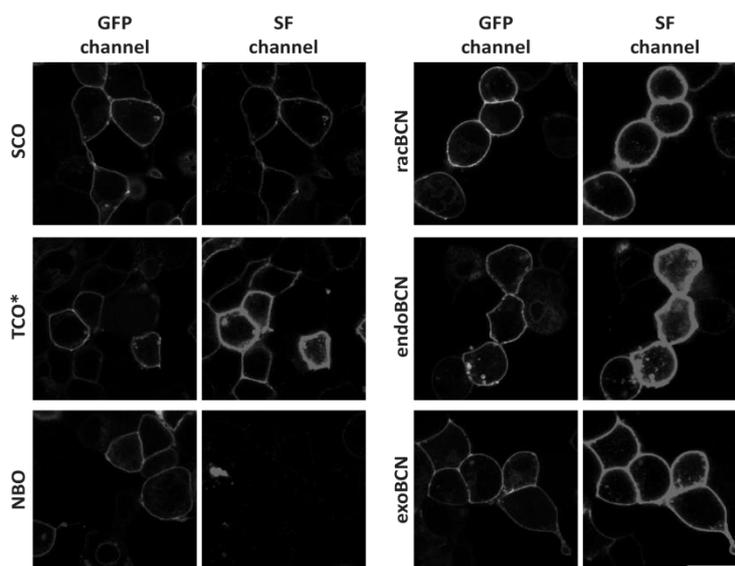
The initial step of the experiment involved preparation of the expression plasmid of IR in mammalian cells. IR gene was cloned into a mammalian expression vector containing cytomegalovirus (CMV) promoter and a GFP (pIR-GFP). GFP was placed at the C-terminus of the IR to use it as a readout for Amber suppression. In this way, GFP signal was observed only when the nCAA was properly incorporated into the IR. The next step was to find a suitable site on the IR gene for Amber mutation (TAG) which allowed Amber suppression and exposed the incorporated nCAA to the exterior of the cellular membrane. Taking these into consideration, a site of IR exposed to the exterior environment (K676) was chosen and mutated to TAG. The construct, pIR(676TAG)-GFP, was used in this experiment.

The experiments were performed with transiently transfected HEK293T cells which gave clearly visible expression patterns of IR(676TAG)-GFP at the cellular membrane as shown in the GFP channel of Figure 8. It is also well known that HEK293T have higher efficiency of transfection compared to other cell lines which can also lead to higher expression of target POI and the Amber suppression system. Cells were transfected with a pIR-GFP and a second plasmid containing pyrrolysyl aminoacyl-tRNA synthetase (pyIRS) and its cognate tRNA (tRNA^{pyl}) derived from *Methanosarcina mazei* [131] which are orthogonal to mammalian cells. GFP signal indicated successful Amber suppression of IR with SCO, TCO*mix (mixture of two isomers, refer to section 2.2 for details), NBO, racBCN, endoBCN and exoBCN, chemical structures are shown in Figure 8. After expression of the protein for at least 8 hrs, cells were incubated with medium without nCAA for 14- 16 hrs. Then, a solution of the SF, sulfonated Cy5 with a 1,2,4,5-tetrazine (H-tet) functional group (sulfo-Cy5 H-tet), was directly added to the culture medium of the cells. The sulfonated group increased the hydrophilicity of the SF making it more soluble in the culture medium. Then after 10 mins, the medium was changed to fresh one and the cells were incubated for 2 hrs before fixation and imaging with a confocal microscope. Cells after labeling are shown in Figure 8, the GFP channel shows the localization of IR to the cellular membrane as expected and the SF channel shows labeling which overlaps nicely with the GFP channel. By comparing the relative intensities of sulfo-Cy5 in the SF channel, the labeling reactivity of each nCAA can be estimated. It can be seen that all BCNs (endo- and exoBCN refer to the two enantiomers which differs in the conformation of the bicyclononyne ring, and racBCN is the mixture of these two enantiomers) and TCO*mix are more reactive than SCO with sulfo-Cy5 H-tet. The reactivity rates have also been analyzed quantitatively using stopped-flow spectroscopy and show the same trend [125]. All of the used nCAAs were previously developed from our lab and are now commercially available from SiChem GmbH. It can be concluded that in terms of reaction rate and incorporation efficiency, all the BCN or TCO*mix seems to be the best choice for optimization of intracellular labeling in mammalian cells. This work was put together as a general protocol to site-specifically label membrane proteins

3 Results

in living mammalian cells and was published in 2015 [132]. In all the following experiments, TCO* axial isomer (TCO*) was used instead of TCO*mix due to its faster reaction rate compared to the equatorial isomer (TCO*e) [125].

A)



B)

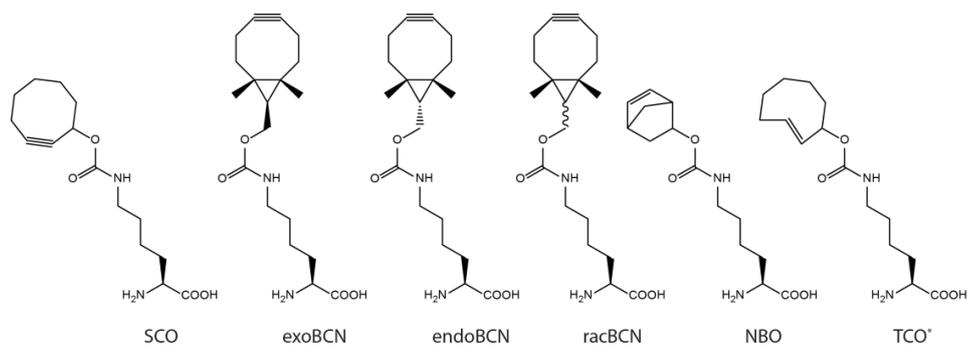


Figure 8 SPIEDAC labeling of IR with Cy5 H-tet using different nAAs.

A) HEK293T cells expressing IR via Amber suppression as shown in the GFP channel. The cells were labeled with Cy5 H-tet as shown in the SF channel. B) The chemical structures of the nAAs used in this experiment. Scale bar: 20 μm . Figure was adapted from Nikic et al. [132].

3.2 Intracellular labeling of Nups in mammalian cells

The next task was to achieve intracellular labeling with SFs after permeabilizing the cells. Before performing the experiment with Nups, IR was used as a target protein due to its localization to the cellular membrane and because it was possible to tag GFP at the C-terminus. Based on the GFP signal which was only observed when Amber codon was suppressed (Figure 8), it was evident that the population of IR localized to the cellular membrane was much higher than the cytoplasmic population. Thus, when IR was labeled the majority of the signal of the SF away from the cellular membrane can be considered as background. In addition to using IR, I determined the source of background by using a non-reactive ncAA, N- ϵ -t-butyloxycarbonyl-L-lysine (BOC). The construct of IR (pIR(676TAG)-GFP) used in section 3.1 was also used in this experiment.

The experiments in this section were first done by labeling IR with sulfo-Cy5 H-tet as shown in section 3.1.1., but only after permeabilizing the cells with Digitonin which is a non-ionic detergent. Digitonin can selectively permeabilize the cellular membrane but not the nuclear membrane [133]. Permeabilization was necessary due to the fact that sulfo-Cy5 H-tet was not a cell-permeable SF. At the time when this experiment was performed, a suitable cell-permeable SF was not available. Therefore, the cells were labeled after permeabilization but before fixation. Obviously, this step can be left out when cell-permeable SFs are used. Development of cell-permeable SFs compatible with SRM came along during the latter course of this study and initial application of the SF seemed to require further optimization. The results are shown in section 3.2.4 and the details are further discussed in section 4.2.1.

When BOC, which does not contain the chemical handle to react with the tetrazine functional group, was used for Amber suppression there was expression of IR, as indicated by the GFP signal but no labeling was observed as expected. (Figure 9, third column from the left, upper and lower panel). The bright dots at the SF channel may be coming from aggregated SFs at the surface of the cells.

From Figure 9, two major issues contributing to the background were observed. One was non-specific sticking of the SF itself inside the cell as evident when non-reactive BOC is used (Figure 9, lower row last panel from the left). The absence of a functional group in BOC reactive with the SF implies that the fluorescence signal comes solely from the non-specific sticking of the SF. The next one was related to the ncAA, which is shown in the SF channel of permeabilized cells when racBCN (BCN) is used (lower row second panel from the left). Fluorescence signal can be observed in the cellular membrane colocalizing with the GFP channel (Figure 9, red arrow) but also inside the cell which is not present in the GFP channel (Figure 9, green arrow). This implies that there was non-specific sticking of not only of the SF itself but also the ncAA which was not incorporated to the POI but rather freely available inside the cell. In the following sections, I will describe how I managed to minimize background coming from both the SF and the ncAA.

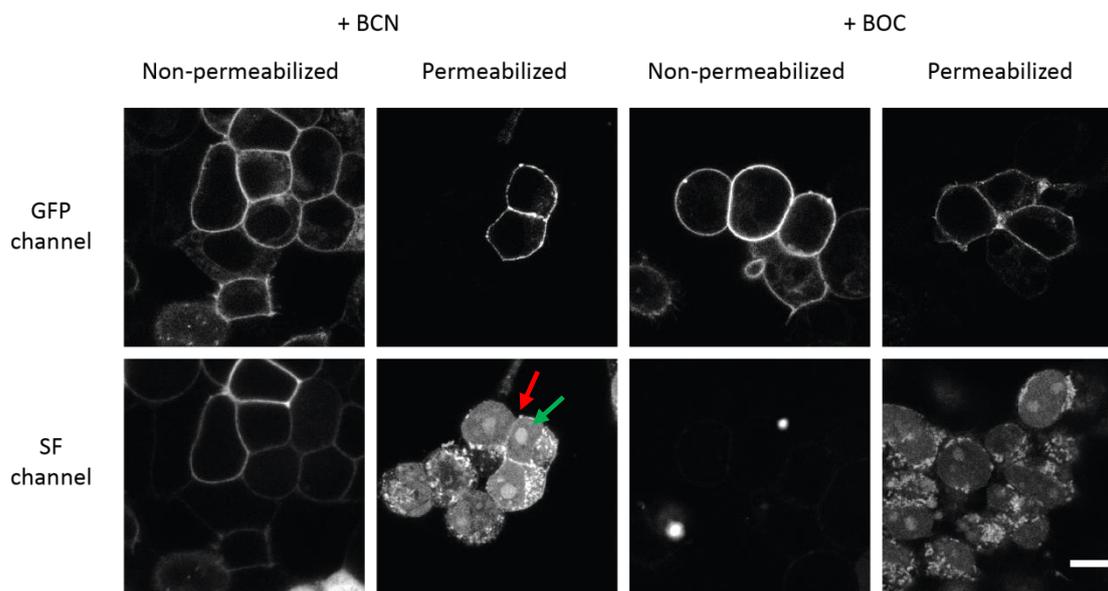


Figure 9 Labeling of IR with Cy5 H-tet with and without permeabilization.

HEK293T cells expressing IR via Amber suppression using BCN and BOC, as shown in the GFP channel. Both permeabilized and non-permeabilized cells were labeled with sulfo-Cy5 H-tet. Non-specific background can be observed in permeabilized cells especially when a reactive ncAA, BCN is used. The red arrow represents labeling of IR with SF while the green arrow represents background. Scale bar: 10 μm .

3.2.1 Reduction of non-specific background using Strep-AF 647

In section 3.2, the main culprits of the background was identified as ncAA which were not incorporated to the target POI and SF which interacted with other biomolecules inside the mammalian cell. In this section, I will describe how I managed to reduce non-specific sticking of the SF. Initial approaches included various washing steps using detergents and organic solvents to wash away sticking SFs. Additional experiments were also done using different permeabilizing reagents. However, none of these attempts showed any significant improvement in reducing non-specific sticking of SFs. Instead, I was able to observe a significant decrease when labeled streptavidin was used instead of directly using the SF to label the protein [134]. I anticipated that the SFs will be washed out more effectively because it's conjugated to streptavidin which is more soluble than the SF itself. A simplified overall scheme of the labeling method using streptavidin conjugated to the SF is shown in Figure 10. To provide the link between the ncAA and the streptavidin, commercially available tetrazine-biotin (H-tet biotin) molecules were used. After expression of IR with ncAA, the cells were permeabilized then H-tet biotin molecules were added. Then, after fixation, the cells were treated with Strep-AF 647, AF 647 is a common SF for localization-based SRM [135]. The results are shown in Figure 11. For both ncAAs used, BCN and

BOC, labeling reaction with SF and Strep-AF 647 are shown. The images using just the SFs were adapted from Figure 9 for direct comparison.

As anticipated, it can be seen in the SF channel of Figure 11 that there was reduction of background when Strep-AF 647 was used instead of sulfo-Cy5 H-tet, as especially evident by inspecting the experiment using Amber suppression with non-reactive BOC. This implies that Strep-AF 647 was much easier to wash away compared to sulfo-Cy5 H-tet. When BCN was used for Amber suppression, colocalization of the GFP signal in the SF channel of both sulfo-Cy5 H-tet and Strep-AF 647 indicated labeling of the IR in the cellular membrane (Figure 11, red arrows). Background was observed inside the cells and most probably it was coming from SF interacting with nCAA not incorporated to IR (Figure 11, green arrows).

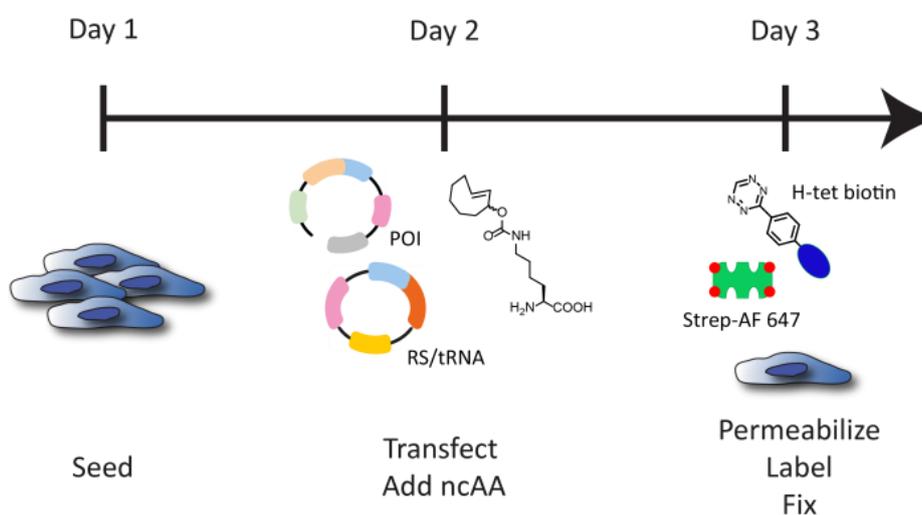


Figure 10 Schematic scheme of labeling method using Strep-AF 647

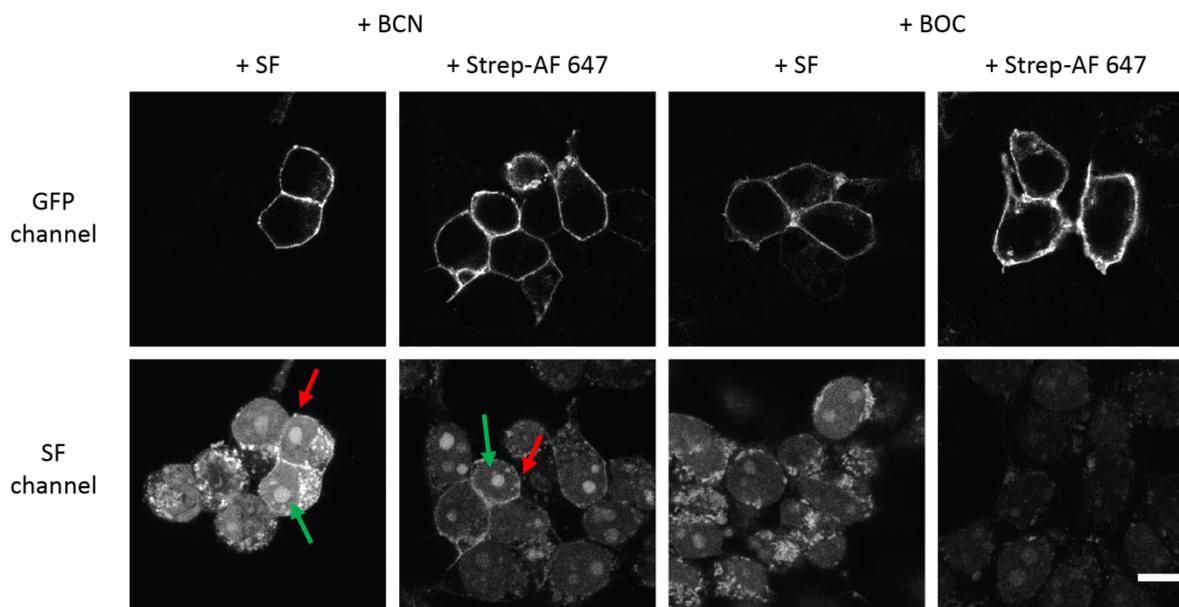


Figure 11 Labeling of IR with Cy5 H-tet and with streptavidin conjugated to AF 647.

HEK293T cells expressing IR via Amber suppression as shown in the GFP channel. Permeabilized cells were labeled with sulfo-Cy5 H-tet and Strep-AF 647. Background was more prevalent when labeling is performed with sulfo-Cy5 H-tet. The red arrows point to labeling of IR with SF while the green arrows point to background. Scale bar: 10 μ m.

3.2.2 Intracellular labeling of Nups using streptavidin conjugated to synthetic fluorophores

Before applying the labeling method using Strep-AF 647 to Nups, it was necessary to first visually confirm the expression of Nups via Amber suppression technology in mammalian cells. Expression plasmids with a CMV promoter containing Nup98 and Nup153 fused to GFP were already available in the lab. As both Nups contain FG repeats and are known to play critical roles in the nucleocytoplasmic transport, these two Nups were used to test and optimize the labeling method. These plasmids were used to transfect HEK293T cells together with the second plasmid coding for $\text{pyIRS/tRNA}^{\text{pyI}}$. Figure 12 shows a comparison of wild-type (WT) Nup98 with Nup98 expressed via Amber suppression. The GFP signal gave a clear nuclear rim staining and formation of punctae due to aggregation [136] in the presence of nCAA while without it only cytoplasmic GFP signal can be observed, which was most probably coming from the GFP with truncated Nup98. It can also be seen that the WT gave higher intensity implying higher expression level. GFP was fused to the N-terminus of Nup98 and the Amber stop codon was placed at K6 position (pGFP-Nup98(6TAG)). It was not possible to fuse the GFP to the C-terminus, as this could inhibit interactions with other Nups to link Nup98 to the NPC.

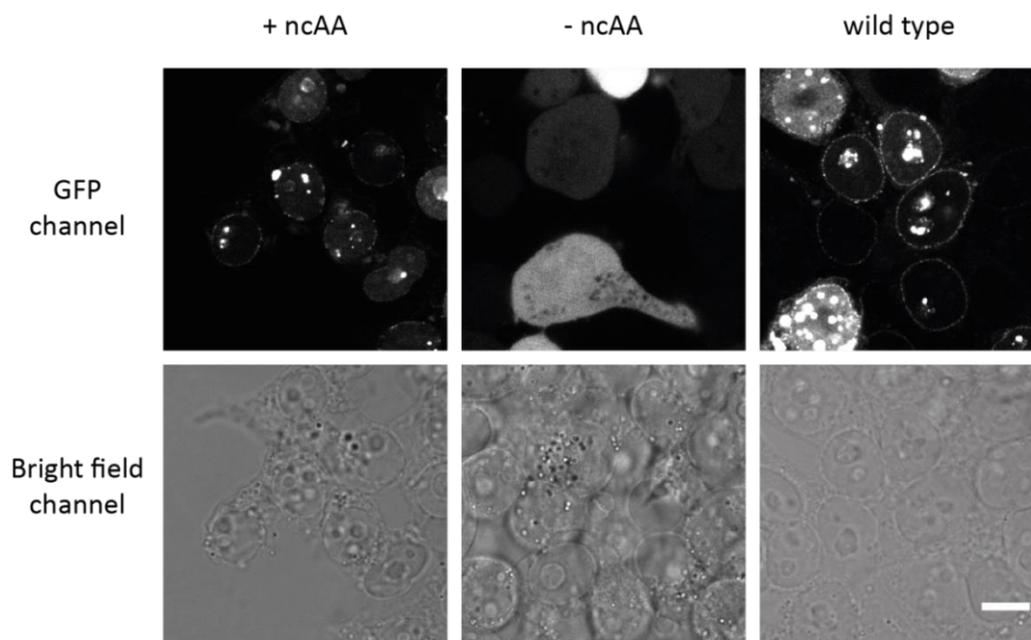


Figure 12 Amber mutant expression of Nup98 in HEK293T cells.

HEK293T cells expressing GFP-Nup98(6TAG) and GFP-Nup98 via Amber suppression as shown in the GFP channel. Nuclear rim staining is not observed in the absence of ncAA. In the presence of ncAA, nuclear rim stain and punctae structure can be observed which is comparable to GFP-Nup98 WT. Scale bar: 10 μ m.

After confirming expression of GFP-Nup98(6TAG) with ncAA, then the labeling method using streptavidin conjugated to SFs was implemented. This time the experiment was done with U2OSf cells. These cells were chosen due to the flat morphology which would enable more convenient acquisition of SR images of the NPCs at the bottom side of the nucleus. Figure 13 shows confocal images of U2OSf cells transfected with GFP-Nup98(6TAG) and $\text{pyIRS/tRNA}^{\text{Dyl}}$. Amber suppression was done in the presence of BCN and BOC. The bottom of the cell was imaged for the GFP channel together with the SF channel which colocalized in the merge channel. With this, it was confirmed that the labeling was feasible in these cells. For the images with BOC, it can be seen that the background coming from non-specific sticking of Strep-AF 647 was minimal as expected.

The cells were subsequently analyzed with GSDIM (ground state depletion followed by individual molecular return), which is a localization based SRM. It can be seen in the TIRF image (Figure 14) that together with the dot-like fluorescence signal in both GFP and SF channel, there were also round micrometer sized nucleolar fluorescent structures. The fluorescence intensity from this structure varies from cell to cell which is why it was not observed in Figure 13. From Figure 14, it was obvious that there were some additional complications that need to be addressed. First of all, the round structures in the nucleus that got labeled need to be minimized especially when imaging

3 Results

the NPCs at the bottom of the nucleus. These structures gave a strong fluorescence signal that masked the fluorescence signal coming from individual NPCs. Secondly, it was impossible to distinguish between specific labeling and non-specific background. This experiment involving GFP-Nup98(K6TAG) was especially challenging because the spatial organization of Nup98 inside the NPC was unknown at the time point when the experiment was performed. Thus, even if the labeling was working properly it was not possible to distinguish it from non-specific sticking of SFs or specific labeling of Strep-AF 647. In the following experiments, I will show how I was able to overcome and circumvent these two complications.

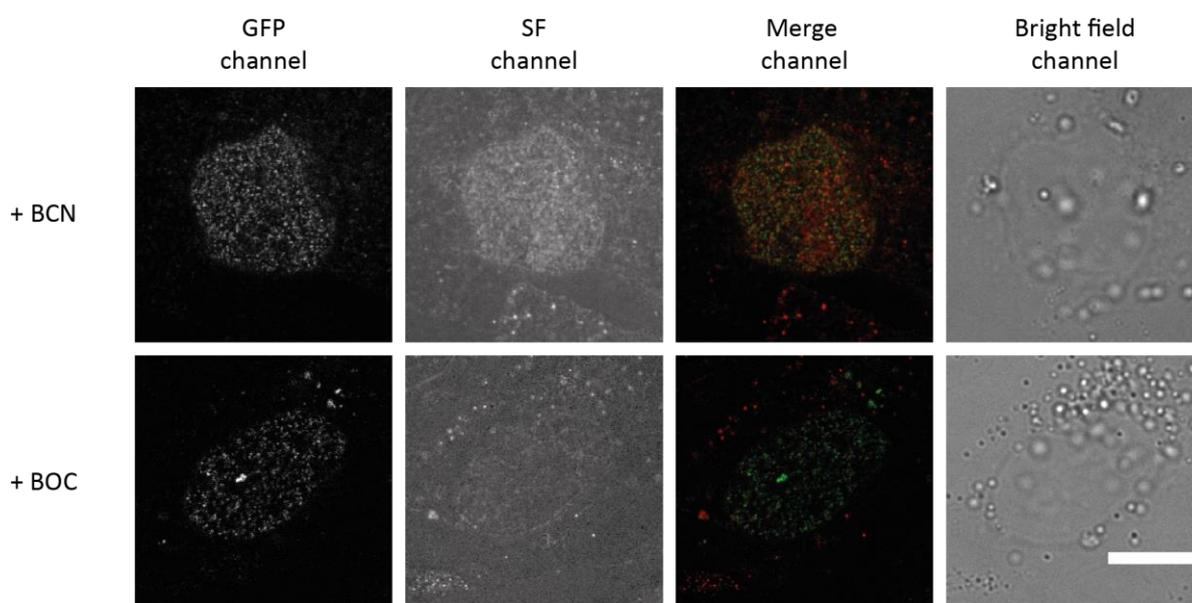


Figure 13 Amber mutant expression and labeling of Nup98 in U2OSf cells.

U2OSf cells expressing GFP-Nup98(K6TAG) via Amber suppression in the presence of BCN or BOC as shown in the GFP channel. These cells were permeabilized and labeled with streptavidin conjugated to ATTO 565 as shown in the SF channel. Merge channel indicates proper labeling in the presence of reactive ncAA, BCN while no labeling is observed with BOC. Scale bar: 10 μ m.

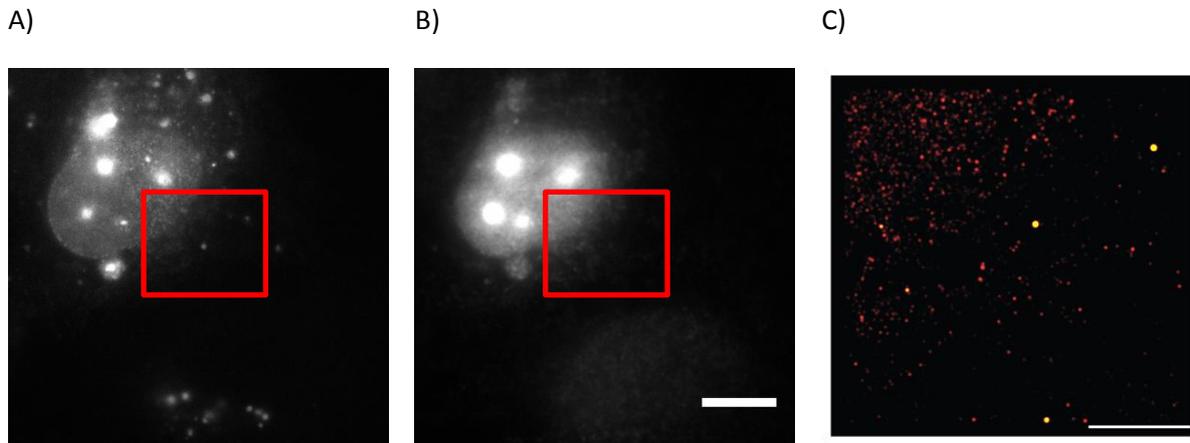


Figure 14 SR imaging of Nup98 labeled with Strep-AF 647

U2OSf cells expressing GFP-Nup98(K6TAG) via Amber suppression with BCN were labeled with streptavidin conjugated to AF 647. A) TIRF image of the bottom of the cell in the GFP channel and B) SF channel. Scale bar: 10 μm . C) Reconstructed SR image of the ROI (red square) in A) and B). Scale bar: 5 μm .

3.2.3 Use of high salt buffer to wash out non-specifically sticking ncAA and SFs

In order to proceed further in visualizing the spatial organization of FG Nups in the NPC, it was necessary to minimize the background signal in the nucleus as shown in Figure 14 which may be also related to non-specific sticking of ncAA, as mentioned in section 3.2. This nucleolar signal was most likely to be coming from the tRNA molecules aminoacylated with ncAA or the RS molecules interacting with ncAA which then reacted with the supplemented SFs [33] and of Nup98 which is known to form aggregates in the nucleus [137]. It was highly undesirable to have this signal from the nucleus especially when imaging Nups in the NPC. Surprisingly, there was an established method to use high salt buffer to extract nucleolar proteins without completely lysing the nuclei [138]. To test if this would give a significant improvement, HEK293T cells were transfected with pGFP(149TAG)-Nup153 and pYRS/tRNA^{pyl}. I made a switch to Nup153 in this experiment, because it was clear that Nup98 was not a proper candidate for unbiased assessment of the labeling method mainly because its spatial organization in the NPC was not clearly known at the time point when the experiment was performed. On the other hand, the N-terminus of Nup153 was better characterized regarding its localization in the NPC [139, 140]. I also switched to a different cell line from previous experiment, due the Amber suppression efficiency which was affected by transfection efficiency. It was much easier to find HEK293T cells with higher Amber suppression efficiency than U2OSf cells. Further details will be discussed in the next chapter.

After expression of GFP(149TAG)-Nup153, the cells were permeabilized and labeled using high salt buffer in which its composition is detailed in section 2.2. As seen in Figure 16A, in the SF channel

3 Results

(second column from the left), the nucleolar signal was minimized when high salt buffer was used during the labeling reaction. The signal at the nuclear membrane indicated proper labeling with the added SF (red arrow). The GFP channel indicated localization of Nups expressed via Amber suppression in the NPC.

In Figure 16C and D, the bottom of the nuclei of HEK293T cells which were transiently transfected with pGFP(149TAG)-Nup153 or pGFP-Nup153 and the plasmid encoding $\text{pylRS/tRNA}^{\text{pyl}}$ in the presence of BCN are shown. High salt buffer was used for the labeling procedure and in addition the cells were immunostained with anti-mouse mAB414 which is a commonly used antibody to localize the NPC. The anti-mouse secondary antibody used was decorated with AF 647. It can be observed that the labeling worked for GFP(149TAG)-Nup153 as indicated in the merge channel of GFP and SF. In addition, it can also be confirmed that the NPC was still intact as indicated by the mAB414 channel. For the cells transfected with pGFP-Nup153 and the plasmid encoding $\text{pylRS/tRNA}^{\text{pyl}}$ which was also supplied with the BCN, the SF signal was minimal as compared to the SF signal from the cells transfected with pGFP(149TAG)-Nup153. With this, it can be confirmed that background due to nCAA has been minimized.

Another important thing to take note is that, the effect of using high salt buffer was significant enough that it was not necessary to continue using Strep-AF 647. Instead, high salt buffer was used as the solvent during the labeling reaction. From this point on, I proceeded with the labeling method using only SFs, unless stated otherwise.

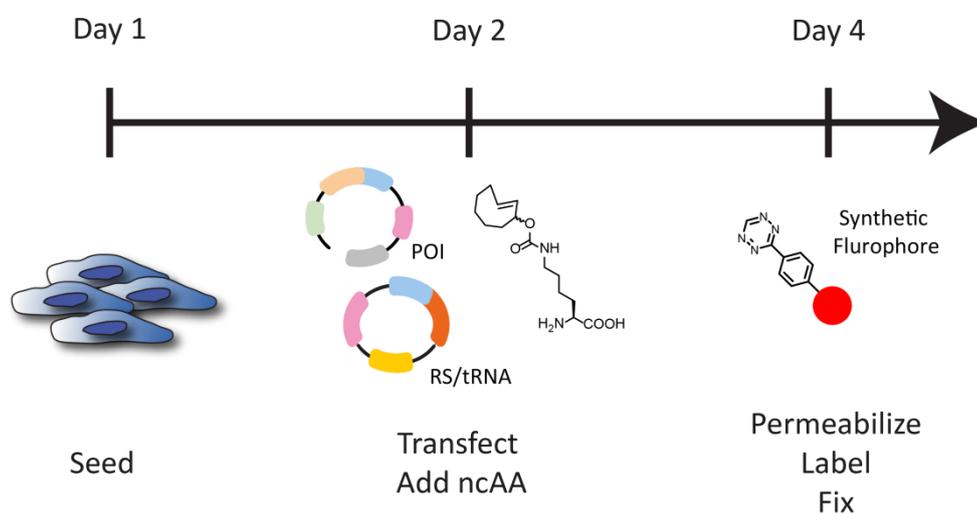


Figure 15 A Schematic labeling method using high salt buffer and SF

3 Results

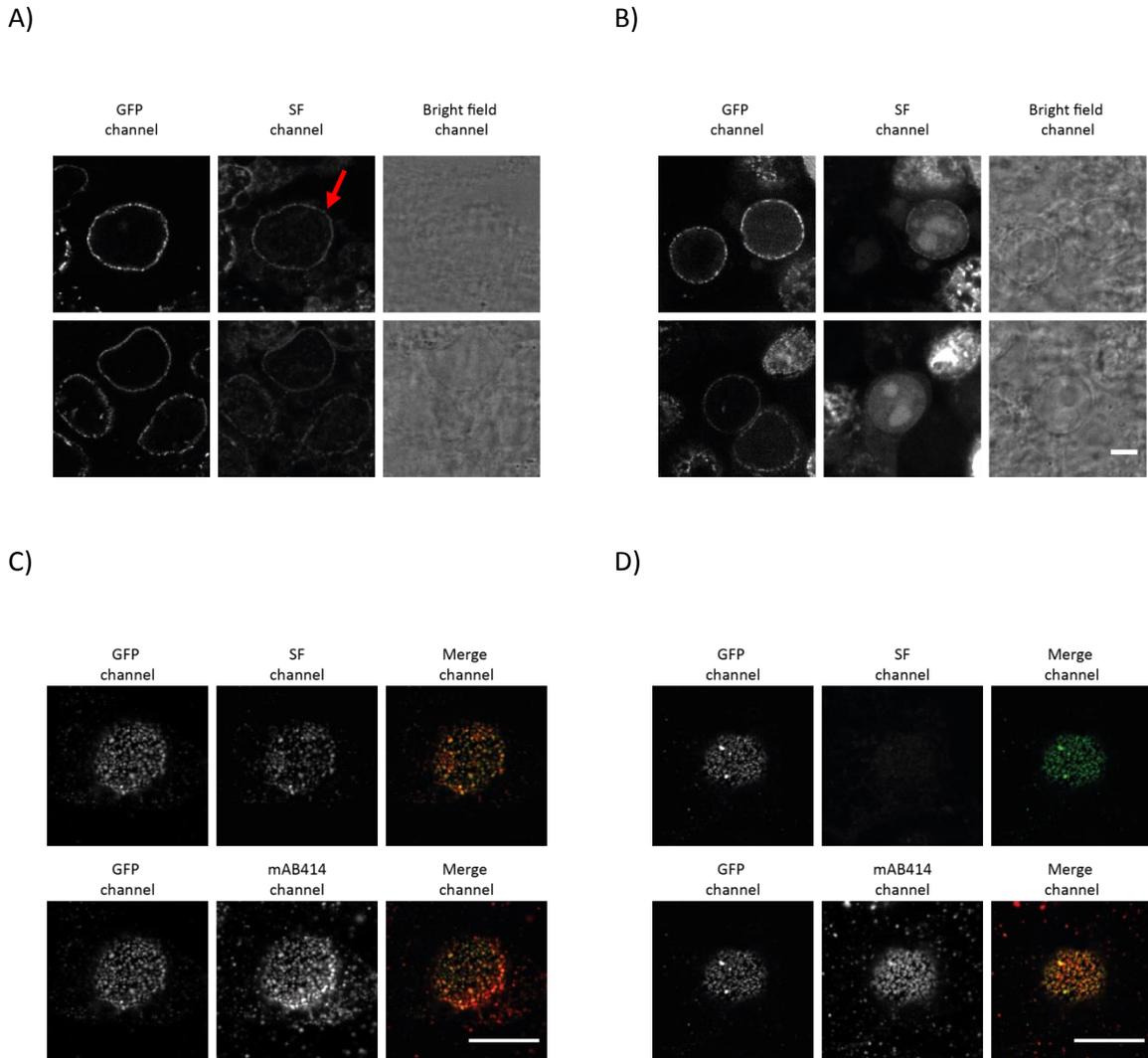


Figure 16 Labeling of Nup153 in HEK293T cells using high salt buffer.

HEK293T cells expressing pGFP(149TAG)-Nup153 via Amber suppression in the presence of TCO. The cells were permeabilized and labeled with sulfo-Cy5 H-tet in A) high salt buffer and B) low salt buffer. The SF channel can be compared wherein with low salt buffer the nucleus gives bright nucleolar signal while high salt buffer gives only nuclear rim staining. The red arrow represents labeling of GFP(149TAG)-Nup153 with SF. Scale bar: 10 μ m.*

The bottom of the cells were imaged C) and compared with D) GFP-Nup153 WT. SF channel indicates proper labeling, while mAB414 channel indicates proper incorporation of labeled Nup153 in the NPC. Scale bar: 10 μ m.

3.2.4 Intracellular labeling of live cells with hydrophilic SFs and ncAA⁴

During the course of my PhD, I participated in testing new advanced SF and ncAA developed in order to minimize background during intracellular labeling of proteins. In this section, I summarize, in brief, the results of implementing these SF and ncAA in labeling Nups in mammalian cells.

Firstly, I had a chance to test a silicon-rhodamine (SiR) based SF with a Me-tet (methyl tetrazine, Figure 17B)) which is less reactive compared to H-tet functional group mentioned in section 3.1. SiR fluorophores are known to penetrate the cellular membrane and at the same time possess favorable photophysical properties which make it an ideal tool for live-cell intracellular labeling [141]. In Figure 17, it can be observed in the SF channel that there was labeling only for the GFP(149TAG)-Nup153 in the presence of both ncAA and *pyIRS/tRNA^{pyl}*. The SF was directly supplied to the cells and incubated in the incubator for the labeling reaction to proceed. However, labeling was only observed in cells which show bright GFP signal indicating higher expression level than other cells. After labeling, the cells were incubated with fresh medium to wash off excess SF. Then, cells were fixed and analyzed with a confocal microscope. While SiR provides potential improvements for live cell labeling, it turned out that my labeling protocol was more appropriate in labeling permeabilized cells with SFs due to labeling efficiency.

For hydrophilic ncAA, namely DO-TCO*, it contains additional oxygen molecules in the cyclooctene ring (Figure 18B)) which increase its solubility in aqueous environment. The increased hydrophilicity enabled much faster and cleaner washout of the ncAA [142] which minimized background. It can be seen in Figure 18, that GFP(149TAG)-Nup153 can once again be labeled with the SF. However, it can also be noticed that only few cells with high Amber suppression efficiency were labeled. When compared to TCO*, which gives higher Amber suppression efficiency, intensity of the SF channel is 2 - 3 fold weaker. It was clear that the incorporation efficiency had to be dramatically increased before further application.

Based on these results, I continued further using TCO* which gave higher incorporation efficiency and was compatible with fast labeling reactions, as well as AF 647 H-tet with my established labeling protocol for labeling Nups in mammalian cells.

⁴ The experiments were done in collaboration with Kele lab, Hungarian Academy of Sciences.

3 Results

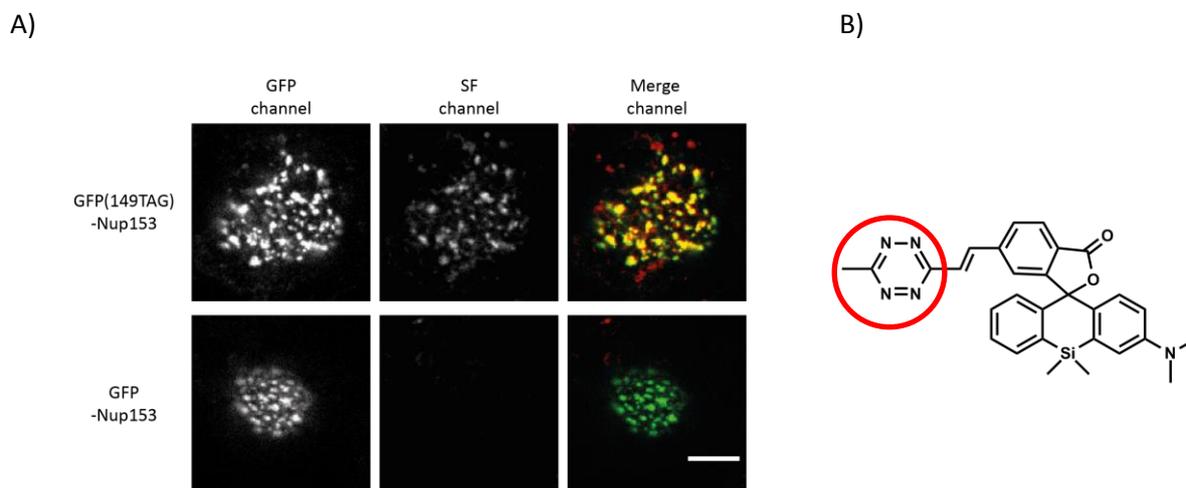


Figure 17 Labeling of Nup153 in HEK293T cells using SiR-dim-tet.

A) HEK293T cells expressed GFP(149TAG)-Nup153 via Amber suppression as shown in the GFP channel. SiR-dim-tet was added to live cells directly to the medium. After labeling, the cells were fixed and imaged. It can be seen in the SF channel that GFP(149TAG)-Nup153 incorporating nAA is labeled while GFP-Nup153 WT is not labeled. Scale bar: 10 μ m. B) chemical structure of SiR-dim-tet. Red circle points to Me-tet.

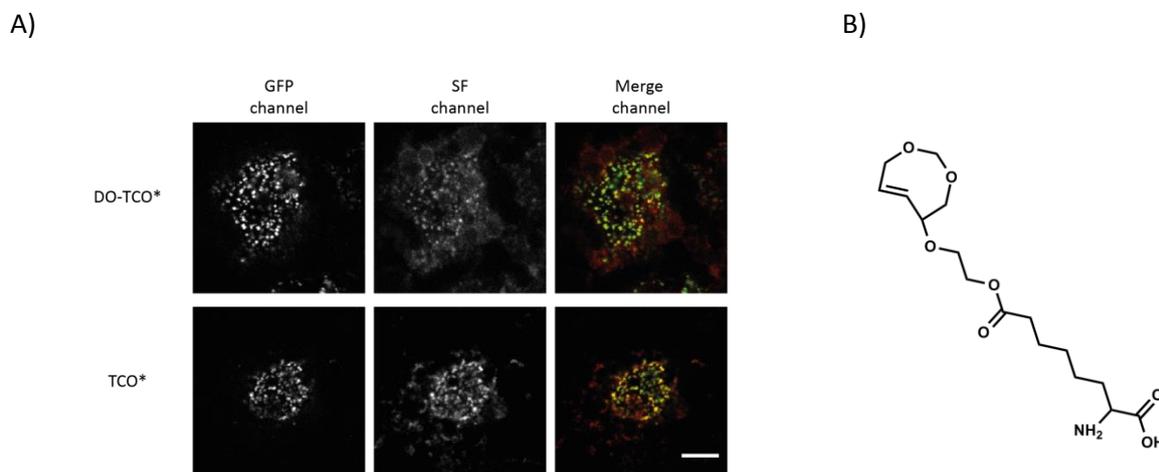


Figure 18 Expression of Nup153 in HEK293T cells with hydrophilic nAA labeled with AF 647 H-tet.

A) HEK293T cells expressed GFP(149TAG)-Nup153 via Amber suppression in the presence of DO-TCO* and TCO* as shown in the GFP channel. AF 647 H-tet was added to the cells after permeabilization with high salt buffer. It can be seen in the SF channel that Nup153 was labeled. The intensity scale in the SF channel of DO-TCO* was increased 5x compared to TCO* to visually confirm the signal. Scale bar: 10 μ m. B) chemical structure of DO-TCO*

3.2.5 Assessment of incorporation of Nup153 to the NPC

In this section, I summarize the results which directly show the effect of heterogeneity of Amber suppression efficiency and the short residence times of Nup153 at the NPC. The cells expressing GFP(149TAG)-Nup153 in the presence of TCO* were labeled with nanobodies conjugated to SFs.

To directly assess if Nup153 expressed via Amber suppression was incorporated to the NPC, GFP(149TAG)-Nup153 was labeled using nanobodies to avoid the effect of the labeling method. Labeling using nanobody conjugated to SFs has been shown to be working with reliable reproducibility [67, 143]. GFP(149TAG)-Nup153 was used in this experiment as the position of the N-terminal region inside the NPC, which can be fused to a fluorescent protein, is known based on fluorescence and EM studies [143]. Moreover, a stable cell (HEKFlpIn 153KD) line from the Beck group, EMBL was readily available. This cell line expresses miRNA to knockdown the expression level of endogenous Nup153, with an efficiency of around 50-60 %.

HEKFlpIn 153KD cells were transfected with GFP(149TAG)-Nup153 and $\text{pyIRS}/\text{tRNA}^{\text{Pyl}}$ and supplied with TCO*. Then, after incubating the cells with TCO* for ~36 hrs, the cells were permeabilized and fixed before incubating with nanobody conjugated to AF647. In Figure 19, ring-like structures can be observed and if these structures were isolated and combined together an average position can be determined. Isolation of single NPCs was done with a custom-written code available in the lab which is based on a density based clustering algorithm [130]. The algorithm was implemented as part of the analysis developed in this thesis and is explained in section 6.2. The determined radius (~47 nm) was comparable to previous studies [139, 144] and thus confirming that the Nups expressed via Amber suppression was integrated into the NPC.

However, it can be seen from Figure 19, that not all of the NPCs showed ring-like structure. Most likely, this was a result of several factors which contributed to the heterogeneity of the system. It has to be noted that it was very tricky to find cells which gave ring-like structures. This observation explicitly implied that the heterogeneity of Amber suppression efficiency and the short residence times of the target Nup in the NPC will affect the final results.

Clearly, there is a need to develop and implement a method to circumvent the issues related to Amber suppression efficiency and short residence times. Improving Amber suppression efficiency has been a topic of interest since its introduction and many improvements have been made in the field which will be discussed further in the next chapter. The short residence time of Nup153 at the NPC is an intrinsic property which should not be altered to preserve its endogenous functionality. Thus, the main solution was to have a reference point at the NPC in order to quantitatively evaluate the positions of the SFs tagged to Nups in the NPC even in the absence of a distinctive structure, such as a ring-like structure. With this reference point, it will be even possible to distinguish specific from non-specific background as mentioned in the previous section (3.1.4.2).

3 Results

Development and implementation of this novel method as well as a general pipeline for analysis will be shown in the next section.

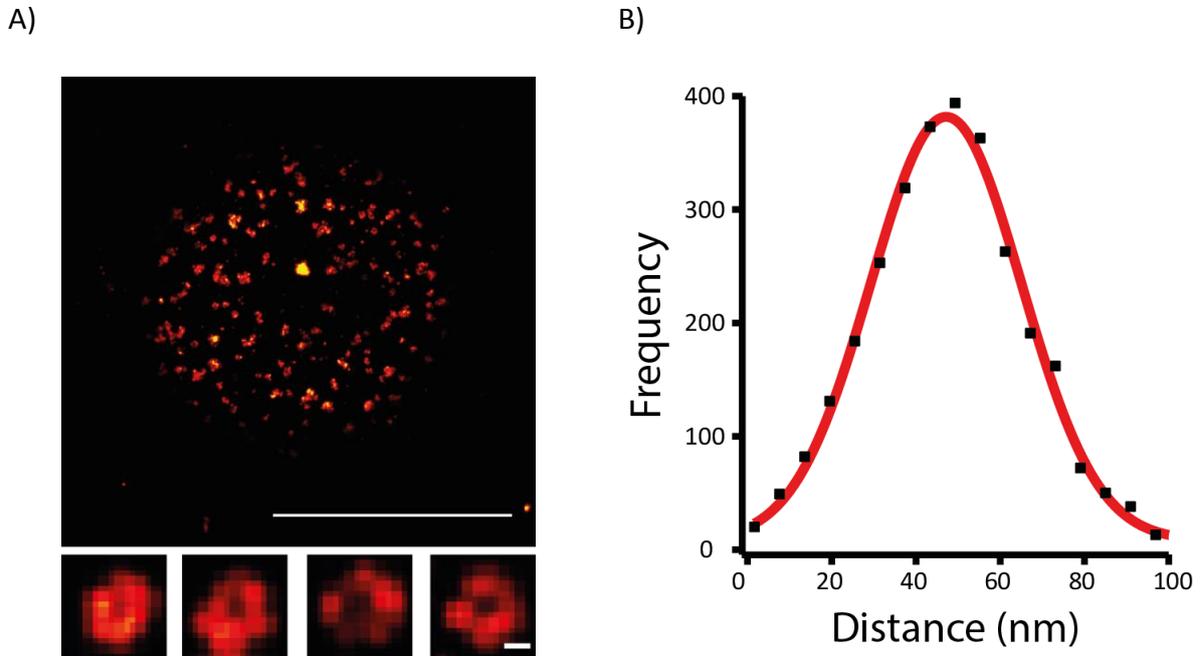


Figure 19 SR imaging of Nup153 labeled with nanobodies conjugated to AF 647.

HEKFlpIn 153KD cells expressing GFP(149TAG)-Nup153 via Amber suppression were labeled with GFP nanobodies conjugated to AF 647 after permeabilization with high salt buffer. A) SR image of a cell which show ring like structure of the pore complex. The lower inset panels show images of selected NPCs. Scale bar: 5 μ m. Inset scale bar: 50 nm. B) Histogram of SF localizations from selected ring-like structures of NPC showing an average radius of ~47 nm.

3.3 Dual-color SRM of labeled Nups

In this section, I will show results on the development and application of dual-color SRM to quantitatively describe the spatial organization of labeled Nups in the NPC. The main reasons for implementing dual-color SRM is to have a common reference point for comparing different labeled positions of a target Nup and to circumvent limitations posed by the heterogeneity in Amber suppression and incorporation to the NPC. By comparing the distances of the labeled positions from the reference point, it will be possible to map how the whole polypeptide chain is spatially organized inside the central channel of the NPC. This concept will be clearly demonstrated in section 3.3.3 when the dual-color SRM results of Nup98 labeled in different positions are analyzed. A major advantage of applying dual-color SRM is that, the method also takes into account the NPCs which are partially or asymmetrically labeled, i.e. even without the presence of a distinctive (ring-like) structure due to suboptimal labeling or incorporation.

For the reference point of the NPC, WGA conjugated to SFs was used. It has already been shown previously that WGA can be used to determine the center of the NPC due to its binding affinity for glycosylated Nups [145]. Thus, commercially available WGA conjugated to SF was used without further purification or modification. WGA was introduced to the cells after the labeling method to minimize interference in labeling the target Nups. Moreover, I adopted spectral demixing to minimize issues related to dual-color SRM such as chromatic aberration and cross-talk. The working principle has been discussed in section 2.4, together with the analysis using RapidSTORM. WGA was decorated with CF 680 and the Nups expressed via Amber suppression were labeled with AF 647.

Moreover, a custom-written analysis was implemented in order to fully utilize the potential of the labeling method. This automated analysis enabled unbiased evaluation of the labeled position with respect to the central reference point from the acquired data using dual-color SRM. I will first describe how dual-color SRM was established and tested with Nup153 then I will move on to explain about the custom-written analysis which automated the analysis of labeled positions of Nups inside the NPC. The last part will be about the application of both the dual-color SRM and the custom-written analysis to look into spatial organization of disordered regions of Nup98 inside the NPC.

3.3.1 Dual-color SRM of Nup153

The concept of dual-color SRM was tested by transfecting HEKFlpIn 153KD cells with pGFP(149TAG)-Nup153 and after expression, the cells were permeabilized, fixed and labeled with nanobodies for GFP conjugated to AF 647. This was done to show the proof of concept of the dual-color SRM while excluding the efficiency of the labeling method using click-chemistry reaction. As shown in Figure 20, not all of the NPCs gave a complete ring-like structure, however, with the use of the reference point it was possible to average the positions of the labeled positions.

3 Results

Subsequently, the labeling was done with SFs. It can be seen in Figure 21 that the labeling GFP(149TAG)-Nup153 with AF647 H-tet works nicely giving high contrast images to resolve the position of Nup153 integrated to the NPC. According to the fit of the Gaussian function to the histogram of all the localized positions of the SF labeled to Nup153, the labeled position is about ~ 42 nm away from the center of the NPC. There were about 1000 localized positions of the SFs coming from one cell in which ~ 20 NPCs were isolated. For the construct used in this experiment, the Amber stop codon is located at the GFP which is tagged at the N-terminus of Nup153. Thus, the localization of the SF should be very close to the N-terminus of Nup153 which is known to interact with the Y-complex located at the NPC scaffold [140]. Hence, if the Nup153 expressed via Amber suppression is labeled properly the average position will be somewhere between 40 and 50 nm. From this result, it was clear that the labeling and the analysis method were working properly.

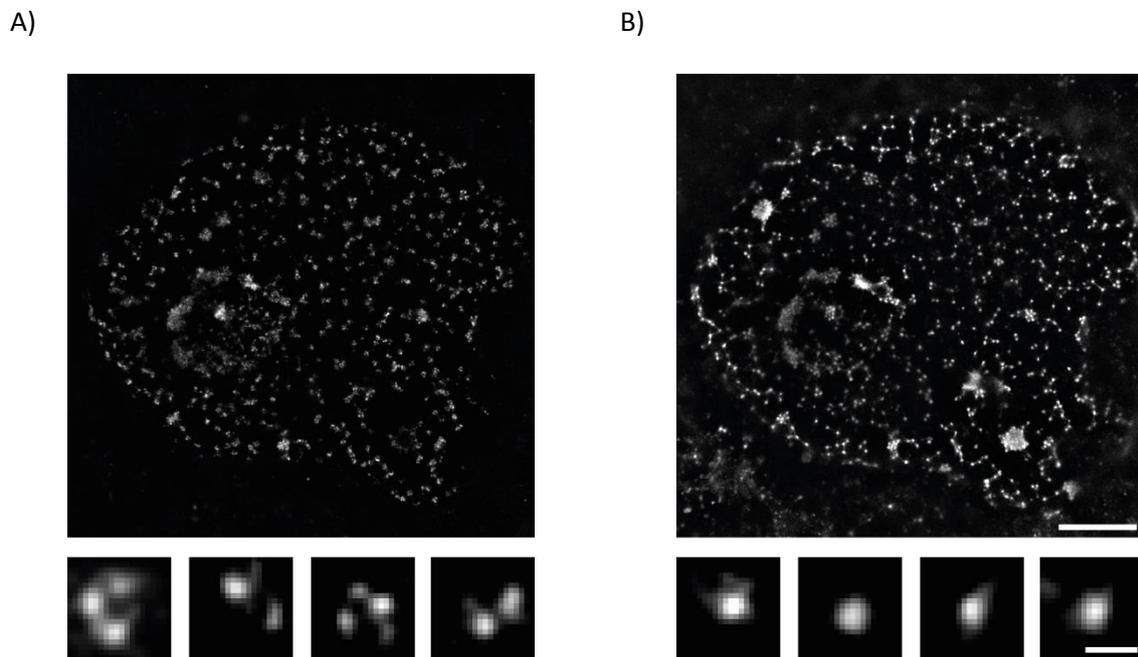


Figure 20 Dual-color SR imaging of Nup153 labeled with nanobodies conjugated to AF 647.

HEKFlpIn 153KD cells expressing GFP(149TAG)-Nup153 via Amber suppression were labeled with nanobodies conjugated to AF 647 after permeabilization and washing with the high salt buffer. Then these cells were incubated with WGA labeled with CF 680. A) SF channel showing labeling of Nup153 forming a ring-like structure. B) WGA channel showing the center of the NPC with a dot-like structure. Upper scale bar: 2 μ m. Inset scale bar: 100 nm.

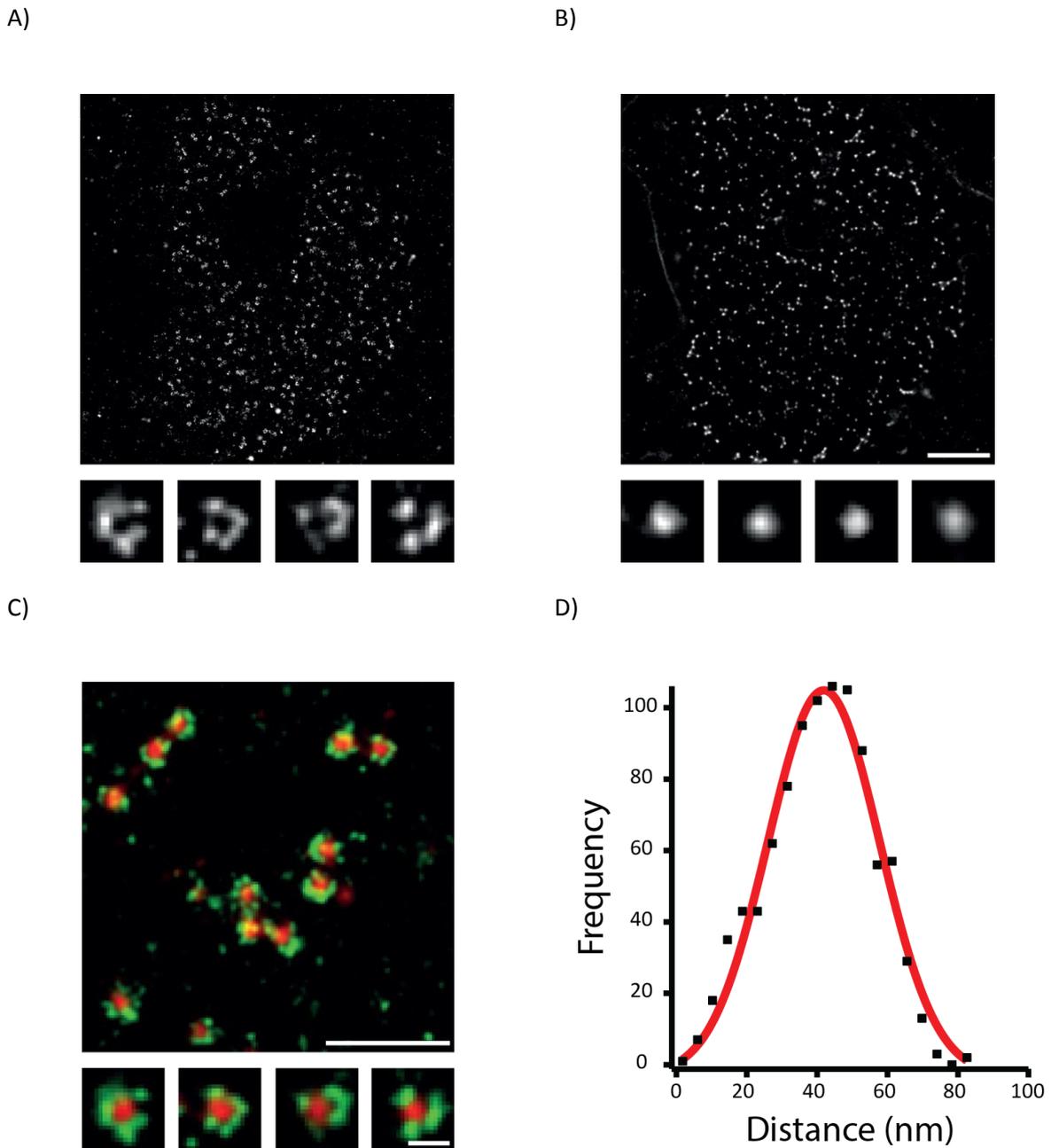


Figure 21 Dual-color SR imaging of Nup153 labeled with AF 647.

HEK293T cells expressing GFP(149TAG-Nup153) via Amber suppression were labeled with AF 647 H-tet after permeabilization and washing with the high salt buffer. Then these cells were incubated with WGA labeled with CF 680. A) SF channel showing proper labeling of Nup153 forming a ring-like structure. B) WGA channel showing the center of the NPC with a dot-like structure. Scale bar: 2 μm . C) The merge channel shows WGA in red and Nup153 in green. Scale bar: 500 nm. Inset scale bar: 100 nm. D) Histogram of AF 647 localization from selected NPCs showing a mean of ~42 nm.

3.3.2 Automated custom-written analysis to evaluate spatial organization of Nups

In order to fully utilize the dual-color SRM scheme in studying the spatial organization of Nups, it was necessary to develop a general analysis tool for more robust and efficient evaluation. A general schematic flow of the analysis is shown in Figure 22. After acquisition, there were two movies from the two EMCCDs which were analyzed using RapidSTORM [127]. Output of analysis from RapidSTORM was two text files which contain localization information for the two SFs separately. This was the only requirement to start with the automated analysis pipeline. The WGA channel was then analyzed to detect bright spots by using a pattern recognition algorithm based on à trous wavelet transform decomposition of the original image [129]. Parameters for circularity and size are used in order to ignore non-specific background WGAs. The centers of these bright spots were then set as the center of NPCs. For the SF channel, all the SF localizations were analyzed for clustering based on a density-based algorithm [146] to reject background signals and group neighboring localizations. The next step was to set a small window of a defined area around the center of the NPC and sort out which clusters were contained inside. Then, all the localizations in the sorted clusters were combined together to create a histogram for quantitative analysis. Based on this histogram, it was possible to compare different labeling positions from different experiments to map the spatial organization of Nups inside the central channel of the NPC with high statistical accuracy.

The working principles and the specifics of the experimental setup and the whole analysis have been discussed in section 2.4. A step-by-step manual of using RapidSTORM for spectral demixing and custom-written analysis for analyzing labeled positions are explained in section 6.1 and 6.2, respectively.

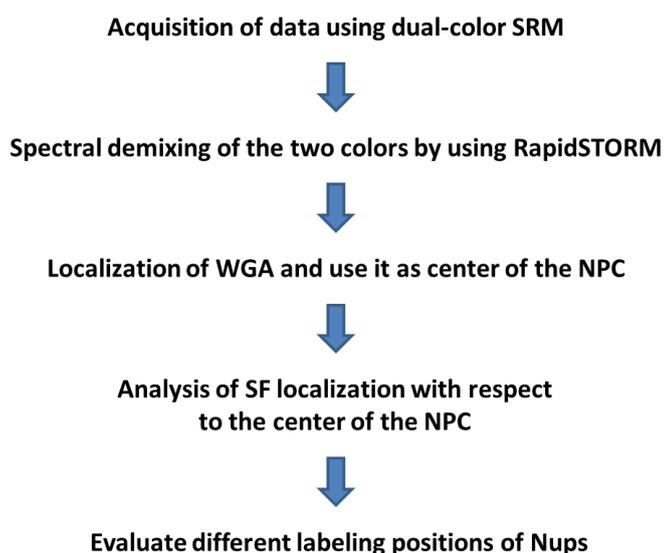
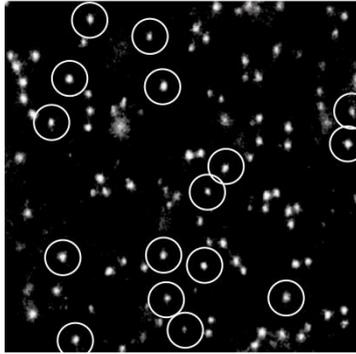


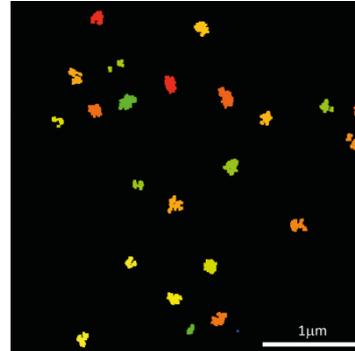
Figure 22 Schematic representation showing the general flow of the dual-color SRM and analysis

3 Results

A)



B)



C)

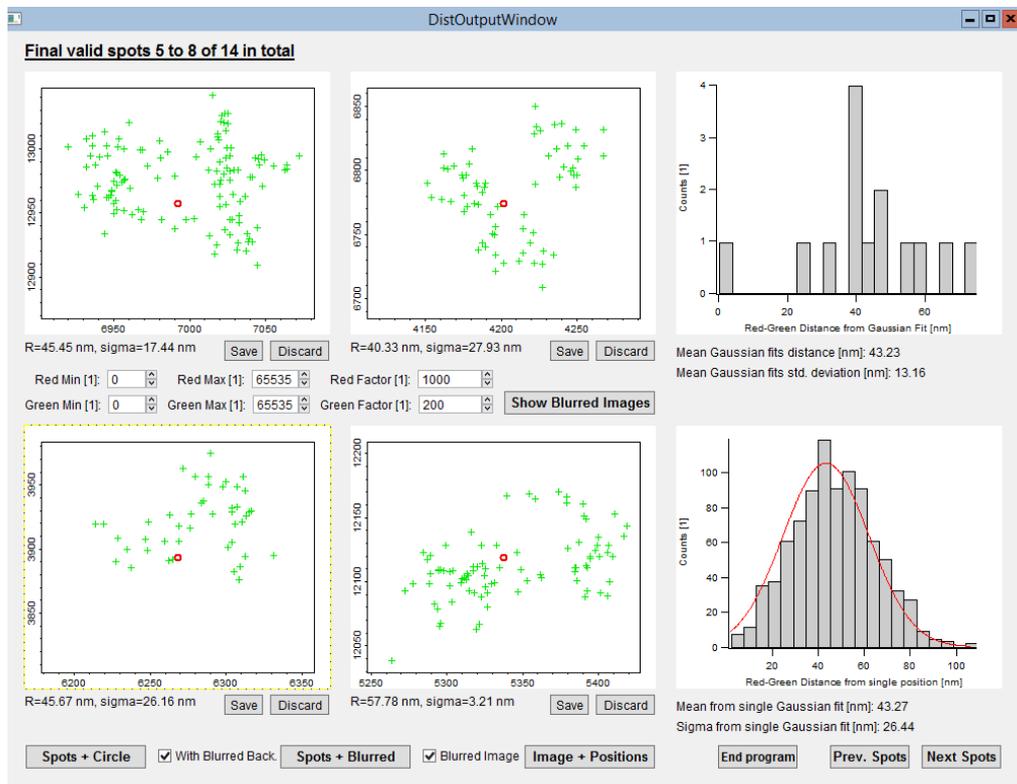


Figure 23 Representative images for each step of custom-written analysis

The custom-written analysis starts with A) localization of bright spots which represents WGA. Then, the next step involves clustering of the SF localization. Afterwards, C) the SF localizations (represented as green crosshairs) are combined and shown as a histogram which is coming from a screenshot of the automated analysis executed in IgorPro. A detailed manual of the analysis is explained in section 6.2.

3.3.3 Dual color SRM of disordered regions in Nup98

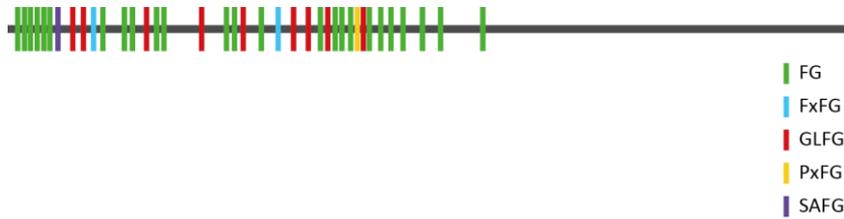
Validation of the labeling method and analysis pipeline was done with pGFP(149TAG)-Nup153 in the previous section. It was confirmed that the labeling method works properly to site-specifically label Nups integrated to NPCs in mammalian cells. The next step was to implement this labeling method in studying the structural organization of disordered regions of FG Nups integrated to the NPC. For this purpose, Nup98 was chosen as the target Nup. Nup98 is known to play a major role in nucleocytoplasmic transport. It has been observed to detach from the NPC while carrying out other biological functions in the cell, such as active participation of regulating transcription not only in unicellular organisms such as yeast[147] but also multicellular organisms such as flies and human cells [148]. As shown in Figure 24A, the N-terminus is mostly composed of FG repeats and the C-terminus is responsible for interacting with other Nups for proper integration into the NPC. Thus, even in the presence of truncated proteins because of the presence of premature Amber stop codon it is not likely to be integrated to the NPC.

Four constructs each containing an Amber stop codon were prepared by site-directed mutagenesis. Three sites were chosen within the disordered region, which are amino acid position 6, 38 and 165. The last site chosen was in position 662 which is very close to the C-terminus. Representative super resolution images are also shown for all four mutants: amino acid position 6, 38, 165 and 662. Analysis of the labeled positions was done the custom-written code in Igor Pro. The output of the analysis was a histogram of all the localizations of a labeled position. It was possible to fit a Gaussian function to this histogram. The center of the fit can be considered as the average distance of the labeled position with respect to the center of the NPC represented by WGA. The Gaussian fit of the histograms of all four positions are shown in Figure 24C. It can be seen that the labeled position with respect to the reference point moves from the center of the central channel towards the periphery as the labeling site changes from amino acid position 6 to 662. The reconstructed SR images and visual representation of all the localizations for each labeled positions are shown in Figure 25 and Figure 26[149]. Details on the effect of the localization error will follow in section 3.3.4.

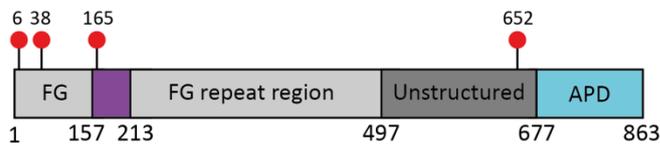
Table 3 shows the analysis of all the labeled positions acquired from the four different mutants of Nup98. It can be observed that the three mutant positions of near the N-terminus are located more close to the center of the NPC channel. On the other hand, position 662 shows higher distance from the center, which is expected due to interaction of the C-terminus region with other scaffold Nups. Further implication of these results in the overall spatial organization of Nup98 inside the NPC will be thoroughly discussed in section 4.3.

3 Results

A)



B)



C)

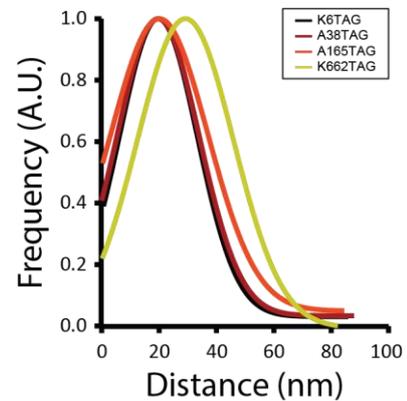


Figure 24 Schematic representation of Nup98 and its labeled positions

(A) Schematic diagram showing the distribution of different types of FG repeats for Nup98. (B) Domain organization of Nup98. FG repeat (grey), GLEBS binding motif (purple), unstructured region (dark grey), autopreolytic domain (APD, aqua blue). The red circles represent labeled positions of Nup98. (C) Gaussian fit of the histograms of the labeled positions from four mutants of Nup98.

Table 2 Analysis of different mutants of Nup98

TAG position in Nup98	Center of Gaussian fit (nm)	Sigma of Gaussian fit (nm)	No. of isolated NPCs	Total no. of localizations
K6	19.76	19.21	46	3110
A38	19.67	20.21	126	7995
K165	20.08	23.89	55	3620
K662	29.20	23.71	86	7080

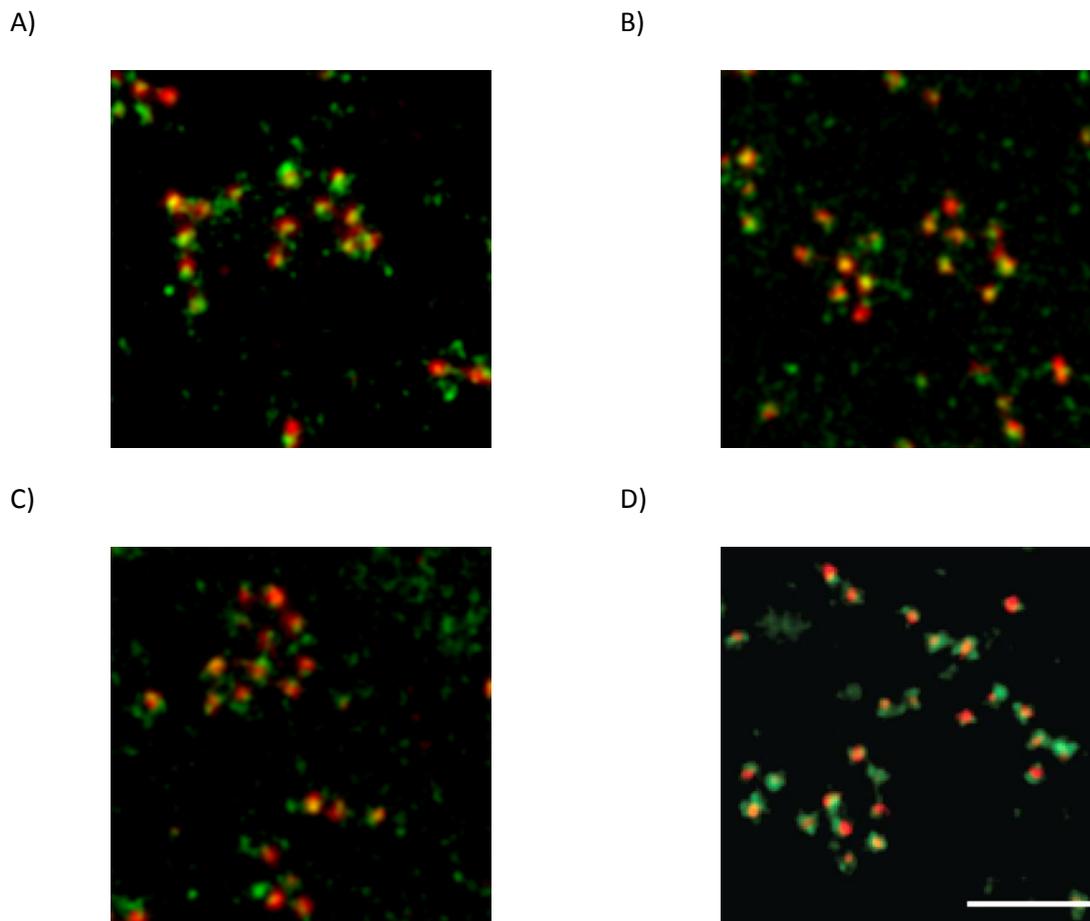


Figure 25 Dual-color SR imaging of GFP-Nup98(TAG) labeled with AF 647.

HEK293T cells expressing GFP-Nup98(TAG) via Amber suppression were labeled with AF 647 H-tet after permeabilization with the high salt buffer. Then these cells were incubated with WGA labeled with CF 680. Merged image showing labeling of Nup98 A) K6TAG, B) K38TAG, C) A165TAG, D) K662TAG. Scale bar: 500 nm. Green channel: AF 647, Red channel: CF 680.

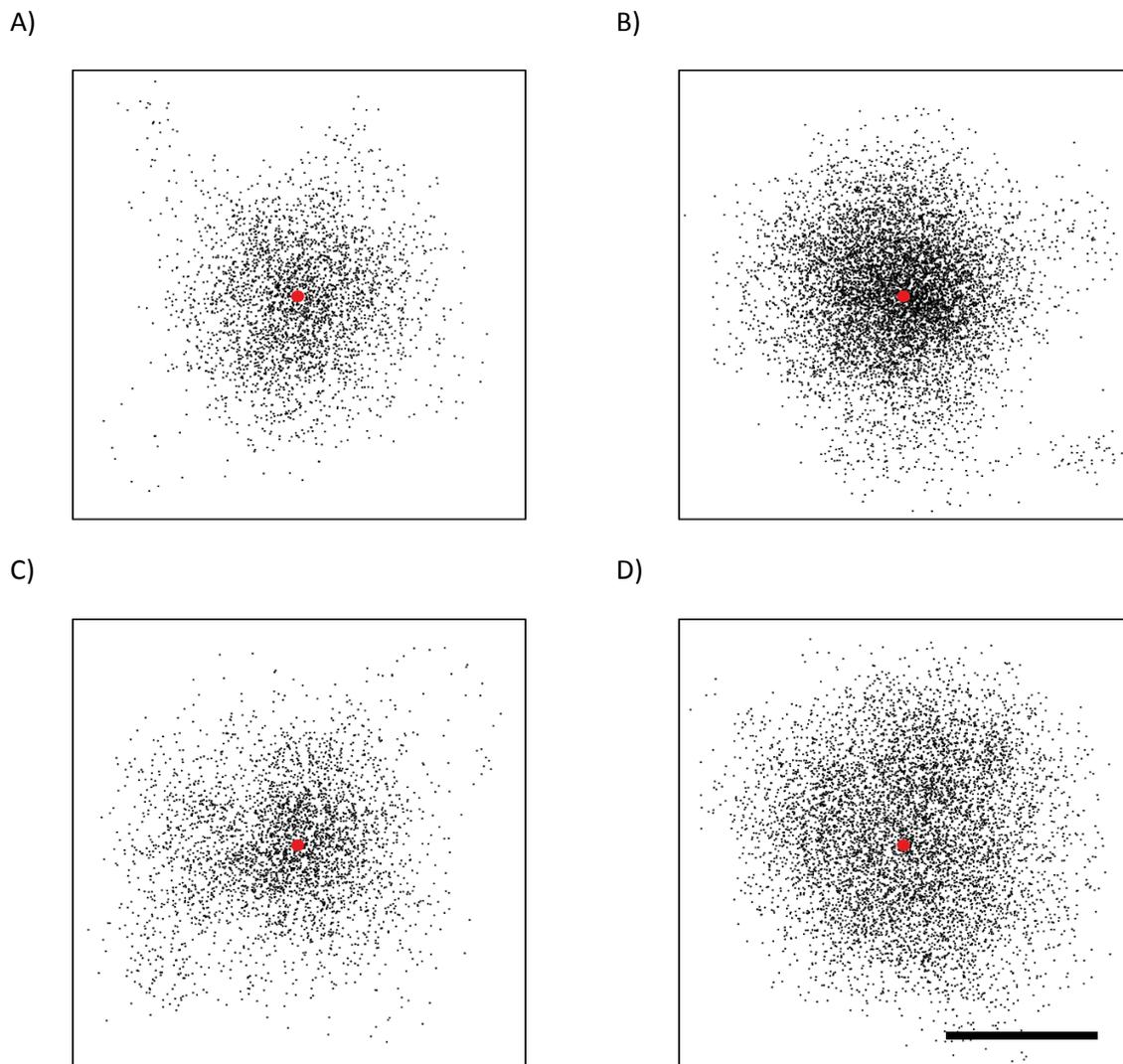


Figure 26 Localizations of GFP-Nup98(TAG) labeled with AF 647 H-tet

Overall localizations of GFP-Nup98(TAG) labeled in positions A) 6, B) 38, C) 165 and D) 662. The red dot represents the center of the NPC. Scale bar: 50 nm.

3.3.4 *In silico* experiments to validate dual-color imaging of Nup98

In this section, I will show results to validate the distance error of the labeled positions from the center of the NPC taking into account the localization precision. I simulated localizations with two input parameters. The first parameter is the distance from the center of the NPC on the assumption that the labeled position will be radially symmetrical with respect to the center of the NPC. This parameter refers to the actual distance of the SF from the center of the NPC. The second one is the localization precision which was simulated by a Gaussian noise with a FWHM of 30 nm which is very close to the experimental value (section 2.4.2). Figure 27A-D shows the resulting 2D plots of simulated localizations for a distance of 5, 20, 30 and 50 nm with a localization precision of 30 nm. The change in the pattern of the localization can be clearly observed. A ring-like structure can be observed by eye when the actual distance from the center is 50 nm. Figure 27E shows the plot of the change in the simulated localizations against the input distance. It can be seen that as the input distance becomes less than 20 nm, the difference increases dramatically until it reaches ~ 12 nm. The implications of these results will be thoroughly discussed in section 4.4.

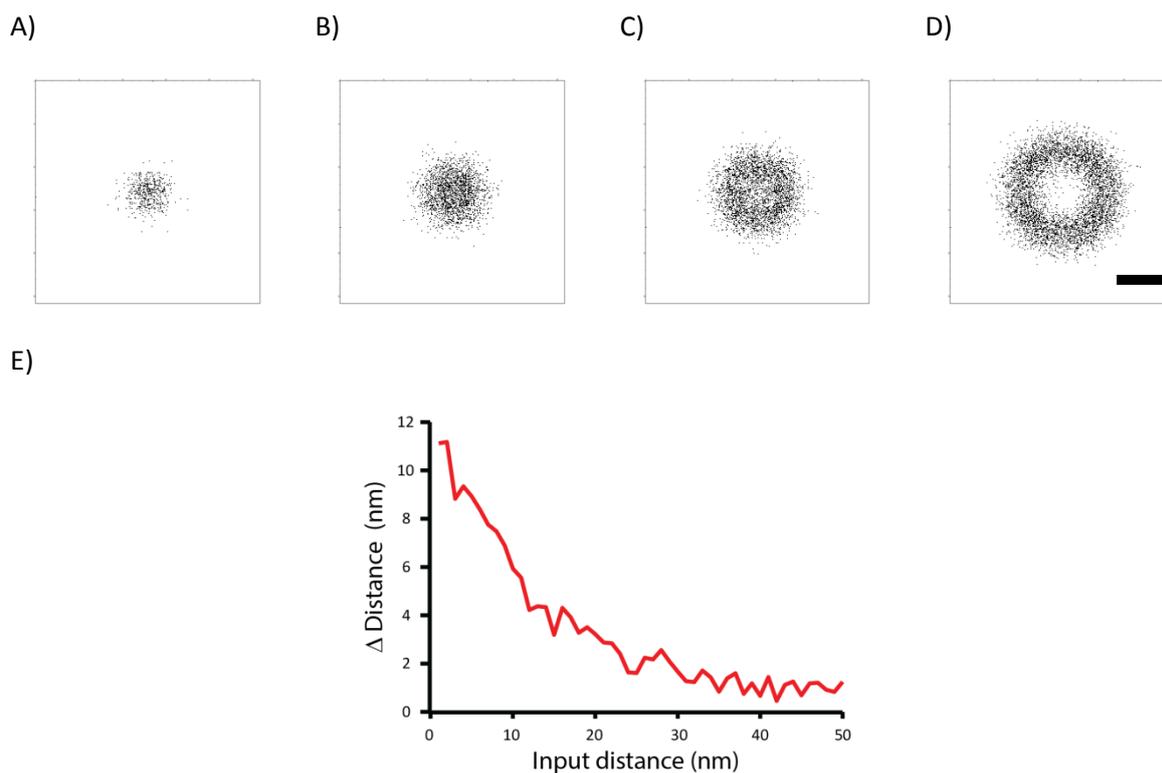


Figure 27 Synthetic generation of data based on localization precision

Localizations were simulated based on the input distance (radius) of A) 5, B) 20, C) 30 and D) 50 nm from the center with a Gaussian noise having a FWHM of 30 nm. Panel E) shows the plot of the offset of the Gaussian center of the generated localizations against the input distance from 2 to 50 nm. (scale bar: 50 nm)

4 Discussion

The main aim of this thesis was to visualize the spatial organization of FG Nups in the NPC using SRM. In order to achieve this, I developed a method to site-specifically label FG Nups with SFs using GCE and click-chemistry reaction. As part of this effort, I also developed a road map to quantitatively assess and compare different labeled positions along the FG Nup based on dual-color imaging and automated custom-written analysis.

To do this, I have grouped this chapter into three main parts which mutually depend on each other to achieve the final aim of visualizing the spatial organization of FG Nups in the NPC of mammalian cells. The first part focuses on the application of click-chemistry reaction for tagging proteins in mammalian cells with SFs which refers to the results in section 3.1. The second part describes the actual experimental establishment of the developed labeling method to label Nups in the NPC of mammalian cells which refers to the results in 3.2. The last part is the development and establishment of dual-color SRM for mapping the spatial organization of Nup98 *in situ* in mammalian cells which refers to the results in section 3.3.

4.1 Implementation of click-chemistry reactions for tagging proteins with SFs

Click-chemistry reactions have the following advantages: fast reaction rate, orthogonality and bio-compatibility. The fast reaction rate allows low concentration (nM to μ M) of the SF which minimizes non-specific background signal. The orthogonality of the chemical reaction increases the specificity of the labeling reaction in order to minimize non-specific labeling of other biomolecules inside the cell. The bio-compatibility allows it to perform these reactions in biological systems such as mammalian cells.

In Section 3.1, I tested different commercially available ncAAs containing strained alkene/alkyne groups in order to evaluate their reaction properties towards tetrazine functionalized SFs. As mentioned in Section 1.3.3, the most suitable click-chemistry reaction when labeling proteins in mammalian cells is the SPIEDAC reaction involving strained alkene/alkyne and tetrazine functional groups. Based on the results with IR, it was concluded that TCO* mix and BCN are the most appropriate ncAAs due to their fast reaction rate compared to other ncAAs that were tested, when labeling Nups in the NPC of mammalian cells in the next step. After this evaluation, I continued further using TCO* mix and SF with H-tet to label Nups.

During the course of my PhD, Li et al. reported about the instability of the product formed between TCO* and tetrazine functional group [150]. However, in all of my experiments using TCO*, no significant effects were observed in terms of labeling and imaging Nups in mammalian cells compared to BCN. It can be speculated that the stability of the reactant before the reaction can also be an issue which has been reported for BCN [151]. Moreover towards the latter part of my

thesis, a newly synthesized TCO* which contains a methylene bridge between the cyclooctene group and the carbamate group was available with the purpose of increasing the stability of the product between TCO* and tetrazine functional group. More systematic analysis of this ncAA is ongoing in the lab to assess the improvement of the stability.

To further decrease the level of non-specific background, it would be even more favorable to have an ncAA with even faster reaction rate. A fast reaction would require lower concentration of both ncAA and SF which will help in minimizing background. Introduction of additional strain to the cyclooctene moiety have demonstrated one of the fastest SPIEDAC reactions in *in vitro* experiments [152]. However, this this strained cyclooctene functional group has not been synthesized as an ncAA thus cannot be tested in mammalian cells.

4.2 Intracellular labeling of Nups in mammalian cells

As introduced in section 1.2, different structural approaches have been implemented to study the spatial organization of Nups *in situ* in mammalian cells. However, none of these approaches have addressed the FG Nups in the central channel of the NPC due to its disordered nature. In my thesis, I have developed and established a labeling method using GCE and click-chemistry to label Nups *in situ* in mammalian cells with amino acid specificity. Uttamapinant et al. implemented GCE and click-chemistry to site-specifically label cytoskeletal proteins with SFs [33]. However, labeling of less abundant proteins such as Nups in the NPC still remains as a challenge.

In this section, I will summarize all of these complications that I encountered when GCE and click-chemistry was implemented to label Nups *in situ* in mammalian cells and how some of these complications were overcome. I will also discuss how to further improve the convenience of the labeling method.

As shown in Figure 9, the major complication that I encountered is non-specific background. With the use of control experiments as shown in Figure 9, I identified the two major sources of background. One major source of is coming from the ubiquitous presence of ncAA inside the cell. It can be imagined that, once the ncAA enters the cell, there will be several populations of ncAA. Some of it would be interacting with the RS, some would be charged to the tRNA, some would be just freely available in the cell waiting to be taken up by the RS and some would be incorporated to the target POI via Amber suppression. In addition, some ncAA populations will be misincorporated or non-specifically sticking. Thus, when SF is added all the populations of ncAAs inside the cells can get labeled. Moreover, background was more prominent in the nucleolus in which we hypothesized that localization of RS/tRNA greatly contributed to it (Figure 9). Another source of background comes from the hydrophobic core structure of many SFs which triggers interaction with numerous biomolecules inside the cell other than the ncAA.

To minimize the non-specific background, I implemented the use of a buffer containing high amounts of salts such as potassium acetate and sodium acetate which extracted nucleolar proteins

and other biomolecules including excess ncAA and SF. This buffer compositions was already previously reported as part of a protocol to extract nucleolar proteins [153]. With this, I was able to reduce non-specific background fluorescence. In Figure 16, it can be seen that the nucleolar signal is no longer visible which leads to higher contrast in SR images.

Even though, I was able to reduce non-specific sticking of the SF by using streptavidin. This labeling method of using Strep-AF 647 was not applied further due to the outstanding washing properties of the high salt buffer. Thus, in the all the labeling reaction after section 3.2.2 was performed in high salt buffer.

4.2.1 Live cell intracellular labeling with hydrophilic ncAA and SiR-based fluorophores

Despite the advantage of using high salt buffer in minimizing non-specific background, it is also necessary to take note on the possible consequences of using high salt buffer. There might be effects on molecular interactions between biomolecules which are highly electrostatic in nature. Although, I have confirmed the presence of FG Nups in the NPC by doing immunostaining with mAB414 after high salt buffer treatment (Figure 16), there is still a possibility that high salt buffer might have affected the organization of FG Nups or the NPC as a whole. In order to avoid using high salt buffer, it would be beneficial to use cell-permeable SFs.

Recent development of cell-permeable SiR (silicon-rhodamine) based SFs has been effectively applied for the visualization of cytoskeletal proteins using SRM. These SiR based SFs show high solubility in aqueous environment which can be exploited to minimize non-specific sticking. Moreover, these SFs also show fluorogenicity 10 to 100 fold which is ideal for intracellular labeling [154]. This increase in fluorescence is coming from structural changes at the core skeleton of the SF when chemically reacting to a specific labeling site. In its unreacted form, these SFs form aggregates which favor the non-fluorescent chemical structure even when non-specifically interacting with hydrophobic structures inside the cell. By using these SFs, it will enable not only live cell labeling but also imaging the cells without any fixation which can also lead to structural artifacts. Standard fixation protocols require the use of strong chemical reagents which cause cell shrinkage and disruption of cellular organelles [155], which may affect localization and dynamics of soluble proteins inside the cell.

Preliminary experiments were performed with the SiR-based SFs provided by our collaborators⁵ and the results are shown in Figure 17. The results show that the labeling of Nup153 was not effective which may be coming from a new tetrazine group conjugated to the SF (Figure 17B). Fluorescence signal from the SF was observed overlapping only with very bright GFP signal indicating high expression of GFP(149TAG)-Nup153. A possible reason for this can be due to the low reactivity of the tetrazine functional group attached to the SiR. This tetrazine functional group

⁵ Peter Kele Lab, Hungarian Academy of Sciences, Hungary

(Figure 17B) is known to be more fluorogenic than the normal H-tet or Me-tet but it seems that the reactivity is too slow for intracellular labeling of Nups. Although the concentration during the labeling reaction can be increased to increase the reactivity, it is not recommended due to non-specific background. It would be more interesting to test SiR SFs with more reactive tetrazine group which are also commercially available.

Coumarin and rhodamine derivatives also show excellent cell-permeability, photophysical properties (i.e. brightness and photostability) and solubility. In addition, recent finding shows that with simple modifications to rhodamine-based SFs, its quantum yield can be greatly improved for applications in SRM [156]. It would be interesting to implement these SFs in studying localization and dynamics of intracellular proteins. Nevertheless, the access to these SFs is limited due to their complicated synthetic procedures and many of them also lack fluorogenicity which contributes to non-specific background signal.

Another interesting SF has also been developed during the course of my PhD. This is a rhodamine-based SF which can blink spontaneously by exploiting intramolecular structural rearrangement (spirocyclization). The difference of this SF from the SiR-based SF of Lukinavicius et al.[141] is that the carboxy group at the core structure which participates in the spirocyclization is replaced with a methoxy group. The main advantage of this SF is that it blinks even without prior intense laser irradiation or introduction of chemical additives such as thiols which enables application in SRM. Uno et al. have shown application of this SF in labeling tubulin in live mammalian cells [157]. It would be interesting to apply this SF in intracellular labeling of Nups to observe their structural dynamics real-time in live mammalian cells.

Another potential candidate to minimize non-specific sticking is the use of hydrophilic ncAAs. It is shown in Figure 18 that it is indeed possible to use hydrophilic ncAAs for expression and labeling Nups in mammalian cells. Recent studies have also shown outstanding washout properties of these newly developed ncAAs proving their potential for application in labeling proteins via GCE and click-chemistry [142]. Nonetheless, there is still room for improvement especially the efficiency of the Amber suppression system using those hydrophilic ncAAs. This can be done by evolving the binding pocket of the currently used pylRS in order to efficiently accommodate the newly synthesized ncAAs. The stability of these ncAAs in reducing environments, such as the interior of a mammalian cell, must also be improved.

4.2.2 SR imaging with DNA-PAINT to circumvent photobleaching

Numerous SFs have been synthesized and evaluated in order to maximize their potential in SRM. However, these SFs have a critical limitation which is photobleaching. The harsh intensity of the excitation beam during acquisition causes permanent loss of fluorescence due to chemical changes to the SF which can occur when the SF non-specifically interacts with reactive oxygen species in the buffer or undergoes structural rearrangement [158]. Additional reagents were added to the

imaging buffer to improve this situation but did not lead to a complete elimination of photobleaching [135].

An exceptional method has been proposed recently in order to permanently subdue this limitation. The method primarily exploits the transient binding interaction of two complementary DNA strands. The specific details of the method have been mentioned in section 1.3.1. Undeniably, the maximum potential of this technology will be unleashed when combined with GCE which will give the freedom to choose site-specific labeling sites for POI. This can be done by introducing a reactive functional group in the docking strand to react with the incorporated ncAA. One disadvantage of the method is that it cannot be applied to live cells. The cells have to be fixed and also permeabilized to allow continuous interaction between the two complementary strands.

4.2.3 Other factors to consider when labeling and imaging Nups in mammalian cells

In order to label the disordered regions of FG Nups, it was necessary to implement a labeling method with residue specificity. This was achieved by implementing GCE in mammalian cells which was underdeveloped during the beginning of this thesis especially for labeling of intracellular proteins with SFs.

In this study, I implemented the $\text{pyIRS/tRNA}^{\text{pyl}}$ system [90] throughout all the experiments involving Amber suppression in mammalian cells. The Amber suppression efficiency was good enough for the experiments; nevertheless, an improvement will increase the overall efficiency of the acquisition process. Increasing the concentration of ncAA can also help but is not the ideal case especially when intracellular labeling is considered due to its contribution to non-specific background. As mentioned in section 4.2, there are different populations of ncAA as soon as it enters the cell. If a higher concentration of ncAA is used then all these different populations of ncAAs will increase as well. On one side this will increase Amber suppression efficiency leading to higher amount of target protein containing the ncAA, however, on the other side it will increase the non-specific background signal coming from the labeling of these other populations of ncAA inside the cell. Thus, a concentration of 250 μM was used throughout the experiments unless stated otherwise.

Application of GCE technology to mammalian cells has made a great progress in the last decade owing to the discovery of pyrrolysine RS and tRNA pair which is orthogonal in both *E. coli* and mammalian cells. Due to this orthogonality, the technology can efficiently be optimized in *E. coli* to accommodate numerous ncAAs and then it has been applied in studying biological processes in mammalian cells. Nevertheless, there are still a number of issues that needs to be addressed to further maximize its application in numerous biological applications.

As mentioned in section 1.3.2.4, Amber stop codon in mammalian cells is used by their translational machinery in order to terminate translation. This process is initiated by the binding of the eukaryotic release factor (eRF) 1 to the Amber stop codon which causes the ribosomal

machinery to dissociate and eventually release the synthesized polypeptide chain with the help of eRF3 and GTP [159, 160]. The working principle of GCE requires introduction of the same Amber stop codon used in translational termination, to a specific position within the gene of the POI which gets recognized by the ribosome to incorporate the growing polypeptide chain with ncAA carried by the orthogonal tRNA. As a consequence of using Amber stop codon for GCE, there is always a high probability that the translation gets terminated instead of incorporating the supplied ncAA to continue the synthesis of the whole POI. Premature termination of translation produces truncated proteins which perturb the function and localization of full length proteins, thus, it must be minimized as much as possible. For example, in the case of Nup153 which contains its pore binding region at the N-terminus, the truncated Nup153 with Amber stop codon near the C-terminus competes with the full length Nup153 containing the ncAA to bind to the NPC. This will lead to improper evaluation of the labeling method and analysis.

Efforts have been made to minimize premature termination of translation by introducing engineered eRF1 in mammalian cells and reducing the level of endogenous eRF1 [161]. Schmied et al. show that there is an increase in Amber suppression efficiency; however, the increase was not significant enough to encourage wider application in different biological systems. The efficiency of Amber suppression has also been addressed by introducing additional copies of tRNAs [161] and creating stable cell lines [162]. However, this stable cell line was not shown for highly expressing proteins and the Amber suppression system was not inducible which may affect the viability of the cell for long term experiments. Efforts are currently going on in our lab to optimize the improvement of Amber suppression efficiency and also to establish an inducible stable cell line. With this, uncertainty from transfection efficiency can be minimized because only the plasmid encoding the POI needs to be transfected. The plasmid for pylRS/tRNA^{pyl} does not need to be transfected as it is already incorporated to the genome of the host cell line. Transfecting with only one plasmid will have a higher efficiency than with two plasmids. The inducibility allows expression of the Amber suppression system machineries only when needed to minimize toxicity issues.

In addition, gene knockdown/silencing systems such as miRNA, siRNA or shRNA can be used to minimize the level of endogenous target POI while increasing the utility of the target POI expressed via Amber suppression inside the cell. These RNAs are usually composed of around 20-30 nucleotides. All of these three systems silence the target gene by interacting with its mRNA. siRNA and shRNA target specific mRNAs while miRNAs can target multiple mRNAs [163]. Thus, miRNA has a higher chance of off-target silencing [164]. The simplest system for effective gene silencing is done by using siRNA which has to be transfected in order to be processed by the target cell but the effect of silencing does not last more than 48 hrs [165]. Genome integration of siRNA can be done by designing shRNA which is its precursor. Intensive optimization and designing are prerequisites for effective application of these systems.

The level of the endogenous Nups which are incorporated to the NPC can dramatically affect the image quality. Gene replacement methods are required to achieve maximal integration of Nups

containing ncAAs to the NPC. In my thesis, HEK293T cells which expressed miRNA to knockdown the level of endogenous Nup153 were readily available. For Nup98, I was able to acquire results even without the means to reduce the level of endogenous protein by implementing dual-color SRM which will be discussed in details in section 4.4.1.

Another suspected major issue of Amber suppression is the nonsense-mediated decay pathway, in which mRNAs containing premature stop codons are degraded before translation occurs. This would drastically decrease the lifetime of mRNAs containing Amber stop codon along the gene of the target POI, which is a prerequisite for GCE. Methods have been proposed to overcome this issue in yeast [167] and *C. elegans* [168] but not yet in mammalian cells. Many of these issues relating to Amber suppression efficiency are currently under investigation by different research groups, it seems like many new discoveries will follow and improve Amber suppression efficiency significantly.

The shape of the nucleus can also affect the quality of the image acquired during SRM. Different types of cell lines tend to have different morphology of the nucleus. It would be much more convenient and efficient if the surface of the nucleus would be flat and not curved. Previous study from Szymborska et al. used U2OS cells which had a nice flat morphology of the nucleus [144]. However, it was not possible for me to directly adopt this cell line primarily due to low transfection and Amber suppression efficiency. Even with the best morphology it would not be straightforward to label Nups containing ncAAs if Amber suppression efficiency is low. This is the reason why in this thesis, HEK293T cells were used due to its high transfection and Amber suppression efficiency. HEK293T cells yielded nice images which could be analyzed. As mentioned in the introduction of this section, Uttamapinant et al. showed the potential of using COS-7 cells in application of GCE and click-chemistry [33]. This cell line behaves like fibroblasts and the morphology of the nucleus is flat which can make them an ideal cell line.

Most of the suggestions and improvements mentioned in this section were require much attention and optimizations and thus were not fully implemented in this thesis. Nevertheless, if the discussed suggestions can be implemented and combined with the dual-color strategy, it will bypass the limitations of Amber suppression and labeling efficiency even more effectively.

4.3 Dual color SRM of Nups

Scaffold Nups which are located at the periphery of the NPC project a ring-like structure when fluorophores are tagged to them [144]. However, for the Nups in the central channel the spatial organization is debated due to its disordered structure. Recent studies integrating different methods such as mass spectrometry, x-ray crystallography and cryo-EM suggest roles and possible locations of folded domains of Nups in the central channel while for the disordered domains their spatial organization is highly debated which ranges from projection to the center to lining to the periphery of the channel as discussed in section 1.1.2.

In order to properly evaluate the spatial organization of the disordered regions of FG Nups in the central channel of the NPC using SRM, it was necessary to have a central reference point at the NPC. With this reference point, it would be possible to analytically evaluate the localization of SFs even without distinctive shape or structure. Thus, even with suboptimal labeling efficiency, incomplete integration or asymmetric organization of the labeled Nups in the NPC, it is still possible to assess and evaluate the localization of each labeled positions. Moreover, different labeling positions can be compared to one another based on their distance from the reference point. This will allow mapping of the spatial organization of Nups, especially the disordered regions of FG Nups, inside the NPC of mammalian cells.

In this study, I used WGA which is known to interact with central channel FG Nups that are glycosylated. It is known that Nup214, Nup62 and Nup98 interact with WGA [170]. It has also been shown that labeled WGA can be used in SRM to determine the center of the NPC [62, 145]. However, it also has to be taken into account that WGA binds to Nup98 which is the target Nup for analysis. Although, the labeling of the disordered region of Nup98 takes place before addition of WGA, there might be some structural changes that might alter the arrangement of FG Nups inside the central channel of the NPC when bound to WGA. In this regard, it would be ideal if an alternative reference point can be implemented to minimize perturbation to the central channel of the NPC. One potential candidate would be a primary antibody binding to Nup62 as shown in a previous study [144].

To image the reference point with the labeled Nup, dual-color SRM method based on spectral demixing was implemented. The main principle of the method is based on the ability to distinguish two spectrally similar SFs based on the ratio of their emission intensities in two separate cameras or channels. The SF pair that I chose was CF 680 and AF 647 for their excellent photophysical properties ideal for SRM. It can be seen from Figure 21 that the SRM method can be used to image Nups in the NPC. Analysis of the acquired movies was done by using RapidSTORM. A step by step guide for using the software is shown in 6.1. The resolution or the localization precision was determined to be in the range of 27 to 34 nm for all the results used for analysis. Details on analysis of the acquired results and the reconstruction of the SR image are outlined in 2.4.

The main advantage of using dual-color SRM via spectral demixing was clearly demonstrated in section 3.3.1. There was no need to implement an additional step to correct for color registration. It was done during the analysis of the intensity ratios of each blinking of the SFs. The acquisition time is also much shorter as only one laser is used to excite both SFs. Chromatic aberration due to the difference of emission range of two SFs was also minimized as the emission range of the SF needs to be similar in range for the method to work properly. As of now, these SFs are one of the best available ones to be implemented in dual-color SRM using spectral demixing.

One possible upgrade to the setup can be done in order to take into account the three-dimensional (3D) structure of the central channel of the NPC. By implementing additional optical components

along with the matching analysis, it would be possible to localize single SFs in 3D [58]. 3D dual-color SRM will allow more accurate mapping of all the disordered regions contained in the central channel of the NPC.

4.3.1 Custom-written analysis for evaluation of labeled Nups

In order to fully exploit the labeling method together with dual-color SRM, a fully custom-written analysis was developed in our lab. The main flow of acquisition and analysis has been shown in Figure 22 and discussed in section 2.4.3. The program was developed in order to automate the analysis of the localizations of the labeled positions. It was necessary to set parameters in order to eliminate the analysis of localizations coming from non-specific sticking of both the SF and WGA.

WGA binding to the sugar molecules at the cellular membrane was eliminated from the analysis by using a pattern recognition algorithm as discussed in section 3.3.2. However, in order to apply a more stringent threshold for labeling of WGA in the center of the NPC, a circularity threshold parameter was added. For the other channel of labeled Nups, a density based clustering algorithm was implemented as mentioned in 2.4.3. The main goal here is to distinguish the fluorescence signal from labeled Nups from the signal coming from non-specific sticking of ncAA and the SF. There are parameters relating to the density such as minimal number of localizations in a designated area and proximity to the center of the NPC. All localizations which were located further than the size of the NPC were discarded. The starting values for these parameters were derived from analyzing cells which gave only non-specific background. These cells are prepared by transfecting them with pGFP-Nup98 (without the Amber stop codon) and pYRS/tRNA^{pyl} in the presence of TCO* and labeling with the same SF.

The main advantage of the analysis is that it is fully automated and it is fully customizable to add in extra features whenever needed. Minor improvements may be implemented to shorten the total calculation time.

Validation of the method and the analysis was done by labeling GFP tagged to the N-terminus of Nup153 with AF 647. The Amber stop codon was placed along the gene of GFP (position 149). Thus, GFP signal will be observed only when Amber suppression occurs. The results show that the SF is localized at around 42 nm from the central reference point determined from WGA signal. It is known that the N-terminal region of Nup153 interacts with the Nup107 complex [140]. According to previous studies, Nup107 (part of the scaffold Nup complex) is known to form a radius of ~44 nm [4] and ~48 nm [144]. The values are comparable wherein the method of labeling has to be taken into account as well. Around ~20 NPCs containing ~1000 localizations were isolated with the custom-written analysis from a single cell which was enough to validate the labeling method and the analysis. These numbers of NPCs and localizations were set as minimum number of data points to have a confidence in the acquired data for the next section when mapping the spatial organization of Nup98. The standard deviation of the Gaussian fit to the data of Nup153 was ~22 nm which is comparable to other studies [4, 144].

4.4 Towards the spatial organization of Nup98 *in situ*

Determining the structure of protein complexes and their components has been an indispensable prerequisite in understanding their function in numerous biological processes. Different structural approaches have been implemented in order to solve the molecular architecture of the NPC, which is one of the biggest protein complexes in mammalian cells. The main function of the NPC is to regulate the transport of biomolecules in and out of the nucleus which is primarily carried out by the central transport channel filled with Nups containing intrinsically disordered regions. Although recent developments in techniques, such as cryo-EM have provided novel insights about the NPC scaffold, it has not been possible to study the spatial organization of the central channel of the NPC due to technical limitations imposed by the disordered regions of its main component which are FG Nups.

Initial experiments with WGA and GFP(149TAG)-Nup153 gave results comparable to previously reported results [143] which gave confidence to proceed further in studying disordered regions of Nup98. Even though Nup153 also contained disordered regions, it was not straightforward to label the disordered regions. This is because the disordered regions of Nup153 are located at the C-terminus while the NPC binding region is located at the N-terminus. Thus, for example, if the Amber stop codon is introduced at position 1330 which is close to the C-terminus. There is a high probability of having truncated GFP-Nup153(1330TAG) which will compete against full length GFP-Nup153(1330TAG) in binding to the NPC because truncated GFP-Nup153(1330TAG) also contained the pore binding region. For Nup98, it was the opposite of Nup153. The N-terminus region of Nup98 is composed of disordered FG repeat regions while the C-terminus contains the pore binding region. Even in the presence of truncated GFP-Nup98, it will not compete against the full length GFP-Nup98 for binding to the NPC.

In my thesis, I developed and integrated the state-of-the-art tools in protein engineering, click-chemistry and SRM in order to visualize the spatial organization of FG Nups in the central channel of the NPC in mammalian cells. Complications such as non-specific sticking of ncAA and SF were minimized and surmounted to reach an optimal experimental condition. I demonstrate the feasibility of the method by labeling four different positions of Nup98 which actively participates in nucleocytoplasmic transport. I was able to label one position at the C-terminus and three positions at the FG domain near the N-terminus with SFs. While, labeled WGA was used as the central reference point of the NPC.

To validate the acquired distance values of the labeled positions of Nup98 in the NPC (

Table 2), *In silico* experiments were done taking into account the localization precision (~30 nm). In Figure 27A-D, simulated localizations are shown with different input distances from the center. It can be observed that a clear ring-like structure is not observed even with a simulated distance of 30 nm from the center due to the limit of localization precision. The ring-like structure is also not clearly observed in the acquired results as shown in the reconstructed SR image in Figure 25D and

the plot of overall localizations in Figure 26D which also gives a Gaussian center distance of ~30 nm. From these simulated localizations, the difference of the Gaussian center of the distances of each localizations to the input distance was plotted against the input position in Figure 27E. The plot implies that, as the labeled position (input distance) approaches the center, the analysis of the localizations will not result to any Gaussian center distance values lower than ~15 nm. Thus, any distance values lower than ~20 nm after analysis should be interpreted more cautiously. With this simulation, it was possible to validate the results that I have acquired with N- and C-terminus of Nup98 which are ~20 and 30 nm, respectively. On the other hand, it also gave an idea of the possible pitfall and limitation of the strategy.

4.4.1 Dual-color SRM results of Nup98 provide clues to its spatial organization

Based on the intensity of the GFP, tagged to the N-terminus of Nup98, cells with high Amber suppression efficiency were imaged with dual-color SRM. Then, by using RapidSTORM, the labeled SFs were localized. As mentioned in the previous section with GFP(149TAG)-Nup153, 1000 was set as the minimum number of localizations. Then these localizations were subject to automated custom-written analysis.

The three labeled positions (6, 38, and 165) near the N-terminus showed a displacement of ~20 nm from the center of the NPC, while position 662 lead to a distance of ~30 nm (

Table 2). For the labeled sites at the N-terminus, this is the first time where localization of disordered region of Nup98 has been mapped inside the NPC of mammalian cells. The closeness of the acquired displacement from the center of positions 6 and 38 provides an internal validation of the labeling and the analysis method. For the C-terminus region of Nup98, recent studies have reported on its localization in the NPC. Pleiner et al. reported SR images of Nup98 labeled with nanobodies that interact with the C-terminus region. They show a ring-like structure of the NPC with a radius of ~30-35 nm [149]. Moreover, the cryo-EM structure of the NPC also shows that the central channel of the NPC has a diameter of ~60 nm [171]. It can be observed that the distance I observed for position 662 (~30 nm) is very close to the radius reported in the literature.

For an extended polypeptide chain with l residues, the contour length can be approximated by bl , where $b=0.38$ nm, which is the average distance between the two $C\alpha$ atoms [172]. Nup98 is composed of ~860 residues in which ~670 residues are considered to be flexible [173]. Thus, an extended form of the flexible region of Nup98 will reach ~250 nm. In order to interpret the distance values of the different labeled positions, a comparison can be made with experimental results using size exclusion chromatography. These results were used to determine the stoke's radius of full length Nup145N, an yeast homolog of human Nup98. Yamada et al. proposed that Nup145N is shaped like a "shrub" as in Figure 28A and that it was divided into two domains based on charge content. The two domains with low and high charges have an estimated diameter of 5 and 6 nm, respectively. With the help of computational tools, they also proposed a model, namely the Forest model which was already introduced in section 1.1.2. Amongst all the models that were

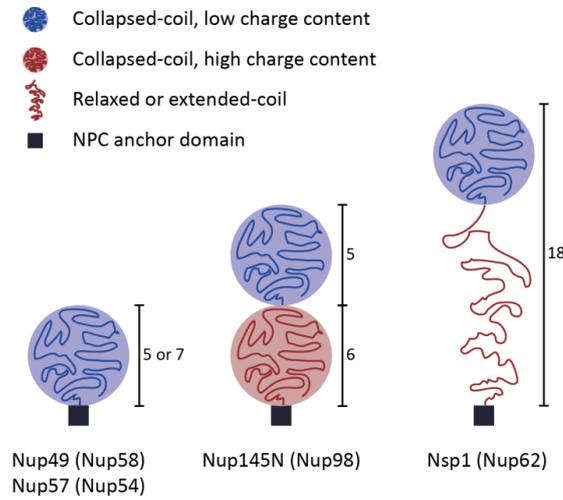
mentioned in section 1.1.2 only Forest model provides actual dimensions of the specific disordered regions of FG Nups in the NPC. Surprisingly, my results of Nup98 agree quite well with this model of the NPC. The proposed overall dimension of ~ 11 nm for purified Nup145N is close to the ~ 10 nm difference that I acquired for the N- and C-terminus positions [30]. Thus, it can be imagined that the huge polypeptide chain somehow fits into a ~ 10 nm space within the NPC channel. The labeled positions are shown in Figure 28B with respect to the NPC. The sizes of the NPC and the Nups in Figure 28B are all in scale based on experimental results. Another feature to take note is the distance of the labeled positions near the N-terminus from the center of the NPC, which is ~ 20 nm. According to the simulated localizations in section 3.3.4, an acquired distance of ~ 20 nm from the center may have a difference of 4-5 nm from its real distance. Nevertheless, this difference can increase by taking into account other factors depending on the experimental conditions. Thus, there is still a chance that the N-terminus is located more towards the center of the NPC.

Although, structural approaches in buffer solution may give useful insights about the size or hydrodynamic behavior of intrinsically disordered proteins such as FG Nups. These approaches do not take into account the local intracellular environment which has critical effects on the structure and function of these proteins. Nevertheless, the results that I have acquired with SRM for the spatial localization of Nup98 *in situ* in mammalian cells nicely complement and support these solution-based approaches.

In section 3.3.3, Table 2 shows a summary of all the labeled positions along Nup98. Each position has different number of total localizations and number of isolated NPCs. For each cell that was imaged, around 10 NPCs can be isolated with around 500 localizations giving a total of 3000 to 7000 localizations depending on the number of cells imaged. I attempted to acquire as much as possible in the given time frame. The number of localizations varies from cell to cell depending on the Amber suppression efficiency. The standard deviation (σ) of the Gaussian function fit to the histogram of the localization data doesn't show a big difference from Nup153 (~ 23 nm). This similarity of standard deviation implies that the disordered regions of Nup98 are organized in a certain structural arrangement and not in a totally random arrangement inside the NPC.

Overall, I have shown the strength of the developed labeling method and analysis in studying the spatial organization of the disordered regions of FG Nups in the central channel of the NPC. The results contribute to the missing puzzle in the molecular architecture of the NPC which has been primarily built up by cryo-EM. My approach clearly demonstrates the possibility of studying the local concentration of FG repeats which is essential in understanding the molecular mechanisms behind the driving force of nucleocytoplasmic transport.

A)



B)

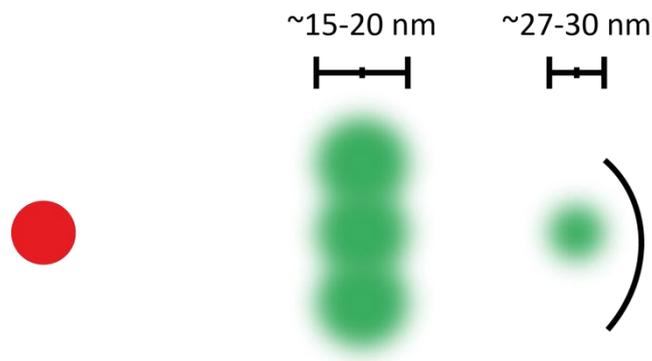


Figure 28 Schematic representation of central FG Nups

A) Representation of three FG Nups based on the Forest model from Yamada et al [30]. The numbers represent hydrodynamic radius in nm. B) In situ localization of Nup98 in the central channel of the NPC based on the SRM results (red and green circles). The center of the NPC is represented by a red circle and the four labeled positions of GFP-Nup98 (TAG) in this thesis are represented in green circles. The numbers represent the distance from the center. The black arc represents the wall of the central channel of the NPC. The radius of NPC central channel in this figure is scaled to ~35 nm based on Pleiner et al [149].

4.5 Perspectives

Based on the localization results of Nup153 and Nup98, it is clear that the developed labeling method together with dual-color SRM and the custom-written analysis has a great strength in mapping the disordered regions of all of the FG Nups in the NPC. Additional positions along the polypeptide chain of Nup98 between amino acid positions ~200 to 600 can be labeled to give more precise mapping of its spatial organization inside the NPC. Moreover, other FG Nups in the central channel such as Nup62, Nup52, Nup58 can also be labeled and analyzed with the same pipeline to complement other structural approaches.

In collaboration with my colleagues, I have some preliminary data showing the possibility of implementing DNA-PAINT in COS-7 cells with Amber suppression⁶ for imaging of FG Nups by SRM. The results are shown in Figure 29. The method was used to image GFP(149TAG)-Nup153 expressed via Amber suppression. As mentioned in section 1.3.1, the biggest advantage of DNA-PAINT is that it overcomes the photobleaching of SFs. If DNA-PAINT is to be applied to the disordered regions of Nups, it is necessary to first establish the dual-color SRM by finding the right pair of SFs. Since the pair of SFs used in spectral demixing (CF 680 and AF 647) has not been tested for its compatibility with DNA-PAINT. For the central reference point, there are other available alternatives to WGA like Nup62 nanobody or primary antibody. The nanobody can be expressed and purified then labeled with the docking strand. For the primary antibody, it is commercially available and thus can be purchased and labeled with the docking strand. In terms of choosing SFs, the same pair of CF 680 and AF 647 as used in this study can be checked if it would be possible to implement it. Otherwise, a different pair of SFs can be used as shown in a recent report by Jungmann et al. They use ATTO 655 and Cy3b to image mitochondria and microtubules, respectively, in HeLa cells with two different excitation sources and detection system [68]. The primary antibodies were modified with DNA strands in order to provide a docking site for the complementary DNA strand with the SF.

Recent outstanding developments in live cell-permeable SFs have created new opportunities to visualize and manipulate proteins to understand fundamental biological processes. These SFs are compatible with SRM which makes them an ideal tool for labeling and imaging of intracellular proteins in living cells while performing their biological functions in real-time. By combining these SFs with GCE and click-chemistry reactions, it would be possible to site-specifically label disordered regions of intracellular proteins in living cells.

As an example of application, the SF can be directly added to live cells to label Nups minimizing perturbations from fixations (section 4.2.1) which can affect the dynamics of the Nups in the NPC. Labeled Nups in live mammalian cells can be used to study their roles in other biological processes. For example, Nup98 is known to play a major role in regulating transcription of developmental genes in eukaryotes [174]. It would be interesting to visualize which region of Nup98 is actually

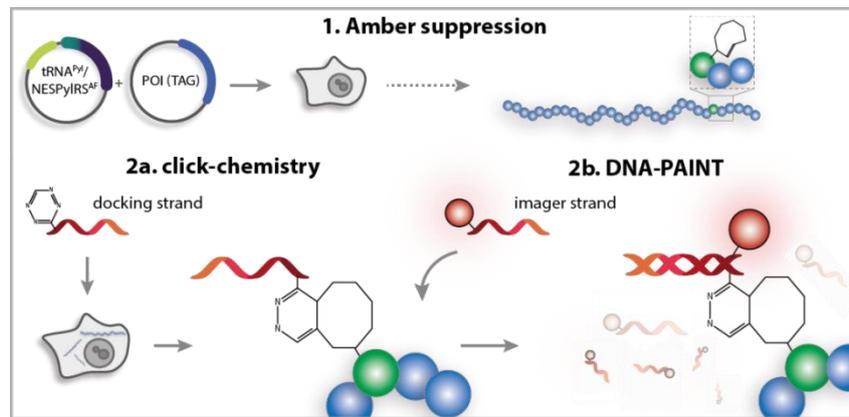
⁶ In collaboration with Ivana Nikić and Gemma Estrada Girona, Lemke group, EMBL.

4 Discussion

involved in interacting with other proteins or the genome in order to carry out its role in regulating transcription. A broader application of the developed method and analysis together with the cell-permeable SFs will be to look into dynamics of FG Nups in the central channel during transport of large cargoes such as virus capsids. It is hypothesized that there will be rearrangement of scaffold and central channel Nups in order to accommodate virus capsids which are as big as the channel itself (~40 nm). It would be interesting to visualize the changes in the spatial organization of the flexible regions of FG Nups in real-time as the virus capsids are transported through the NPC.

The eukaryotic proteome contains 40-50 % of disordered regions, hence termed the “dark proteome” [175]. Thus, the abovementioned applications are only a few of the countless scientific questions which can be addressed. In general, development and implementation of tools such as in cell-permeable SFs, GCE and click-chemistry reactions will bring us one step closer in achieving every scientist’s dream of observing fundamental biological processes with refined spatial and temporal resolution as it occurs in real-time inside the cell.

A)



B)

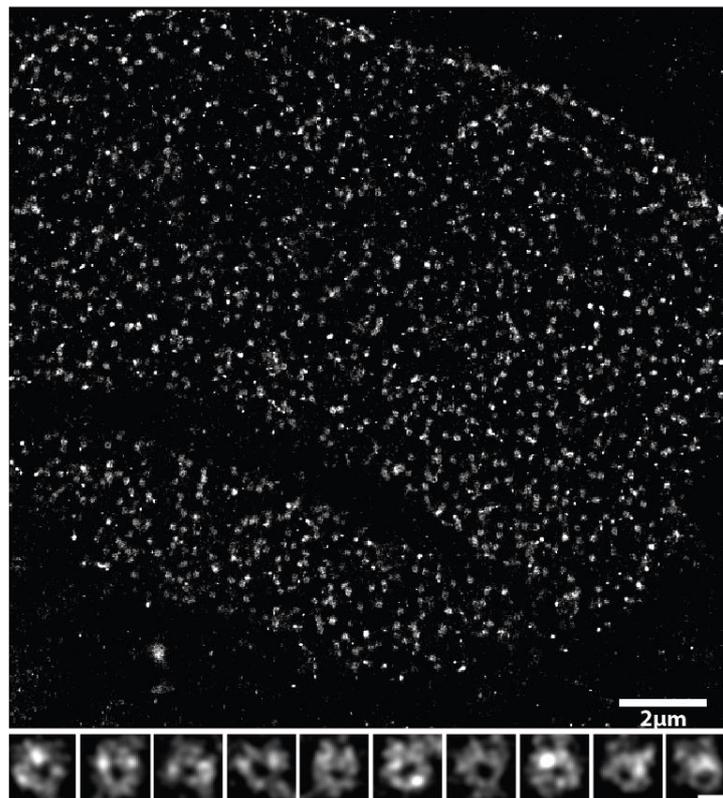


Figure 29 DNA-PAINT in COS-7 cells via GCE

A) Schematic diagram of Amber suppression and DNA-PAINT via click-chemistry. The docking strand contains tetrazine functional group which reacts with the ncAA in the target POI. B) SR image of GFP(149TAG)-Nup153 expressed via Amber suppression with TCO* linker labeled with the docking strand. Scale bar of isolated NPC: 100 nm.

5 Acknowledgments

Even though only my name appears at the cover of my thesis, many people have significantly contributed to its completion both directly and indirectly. I would like to use this opportunity to express my gratitude to all those involved.

First and foremost, I would like to thank Edward for his mentorship, guidance and especially for giving me the opportunity to work in his lab.

I owe my gratitude to all the members of my defense committee for giving me insightful advices and comments.

Hitherto, I am indebted to all the previous and current members of the Lemke group. Sigrid, for showing me the motivation and energy to excel in science. Gustavo, for the quick and effective Spanish lessons and at the same time serious discussions about science. Niccolo, for listening to my silly jokes together with Swati, for giving me enthusiasm about my project in times of failures. He was the main source of my knowledge about the TIRF microscope. Swati, for being my seatmate in the office space for almost three years to answer all my silly and naïve questions about everything in and outside the lab. Thank you again for inviting me to your wedding in India, it was a lifetime experience. Ivana, for all the knowledge about labeling and handling the cells, for all the jokes and funny moments we shared together. I'm very sure that you will be an outstanding group leader. Thank you also for taking your time to read and comment on my thesis. Christine, for all the help especially in the beginning to adjust to the lab, for all the knowledge and insights about Germany and for doing an excellent job in keeping all the people together in the lab, this lab would not have been as it is now without you. Gemma, for all the funny moments and help in the lab, for all the enlightening Spanish tips and for helping me find the emotional side of me. Iker, for all the small and big talks we had during our walking sessions to and from EMBL, for the friendship and support. Giulia, for helping me troubleshoot the problems in the TIRF microscope and for all the fruitful discussions we had. Aritra, for all the informations related to the enjoyment of different cuisines and for keeping me company in the lab during the late hours. Piau Siong, for all the small talks about food, science, Singapore, and everything else under the sun, thank you for taking your time to comment on my thesis during your busiest moments as a father. Sofya, for sharing experiences in the toughest experiment and for having the patience to learn protocols from me. Thank you guys!!!! I will cherish all the moments that we have shared for the rest of my life, :D

I also want to express my sincere gratitude to all the members of the Heidelberg Korean Church for spiritually supporting me outside the lab. God bless you all!!!

Special thanks to Chang Min for bearing with my lousiness and laziness in housekeeping and for providing me with tasty Korean food from time to time. Good luck to your medical licensure examination!!

5 Acknowledgments

Most importantly, I would like to thank my parents for bringing me into this world, without them I would not be who I am today. Thank you very much for continuous support and encouragement in all the choices that I have made through the years. I also thank my brother for cheering me up and boosting my confidence during depressing times.

Last but not the least, I wholeheartedly thank God for guiding me from the very moment I was brought in this world. I believe that he has prepared more exciting plans to make me a better person.

6 Appendices

6.1 Appendix I: Spectral demixing using RapidSTORM

The acquired movies from the two EMCCDs are analyzed using RapidSTORM to separate localizations of CF 680 and AF 647. The whole process of spectral demixing can be divided into four main steps. The first step involves localization for alignment of the two channels. The EMCCD setting must be adjusted in such a way that number of photons is recorded. An online version of the manual from the developers of the software is also available at <https://idefix.biozentrum.uni-wuerzburg.de/doc/rapidstorm/manual.html>. I am showing here optimized analysis for FG Nups together with WGA in the NPC.

6.1.1 STEP 1: Localization analysis for alignment

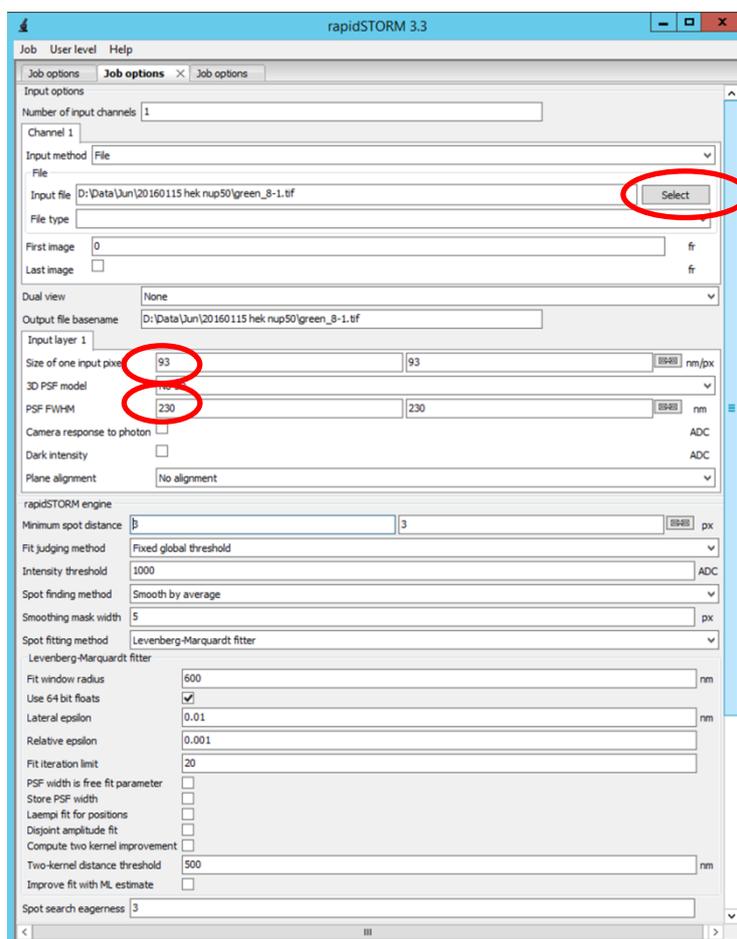


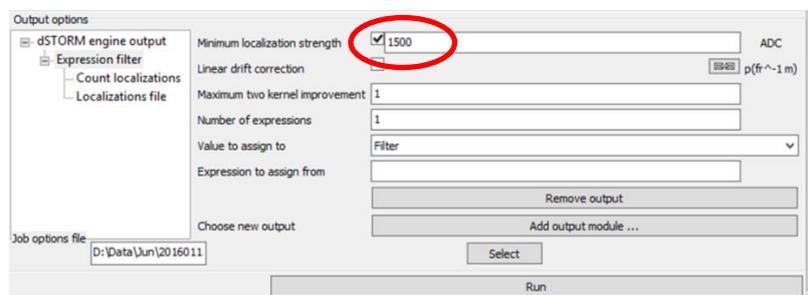
Figure 30 Recommended input settings of localization analysis for channel alignment

The red circles represent settings which need to be manually set.

Firstly, set the “User level” to “intermediate” and select Job > Localization > Minimal. The recommended settings are shown in Figure 30. Select the input movie file acquired by the EMCCD. The localization is done separately for the two movies one after the other. For the “size of the pixel”, it was determined by previous members in the lab as 93 nm for the EMCCDS used in this thesis. The “PSF FWHM” is used when fitting a single SF with a Gaussian function. The value does not need to be precise but the estimate can be determined by emission wavelength divided by twice the NA of the objective in use.

Then, as an output option add “Expression filter” and use “Minimum localization strength” to 1500. In “Expression filter” add output module “Count localizations” and “Localization file”. This will give information on how many localizations have been determined for the analysis. Then in “Localization file”, select the folder where the localization file will be saved. Then click on “Run” to execute the analysis. During the analysis, the number of localization is shown. It is recommended that the total number of localization is similar to the total number of frames of the acquired movie. To increase the number of localizations the “localization strength” can be decreased while vice versa if you want to decrease it. Repeat the whole process by selecting the movie from the other EMCCD. The output of this step will be two text files containing localizations from each EMCCD. Specific details of other parameters will be explained in details in STEP 3.

A)



B)

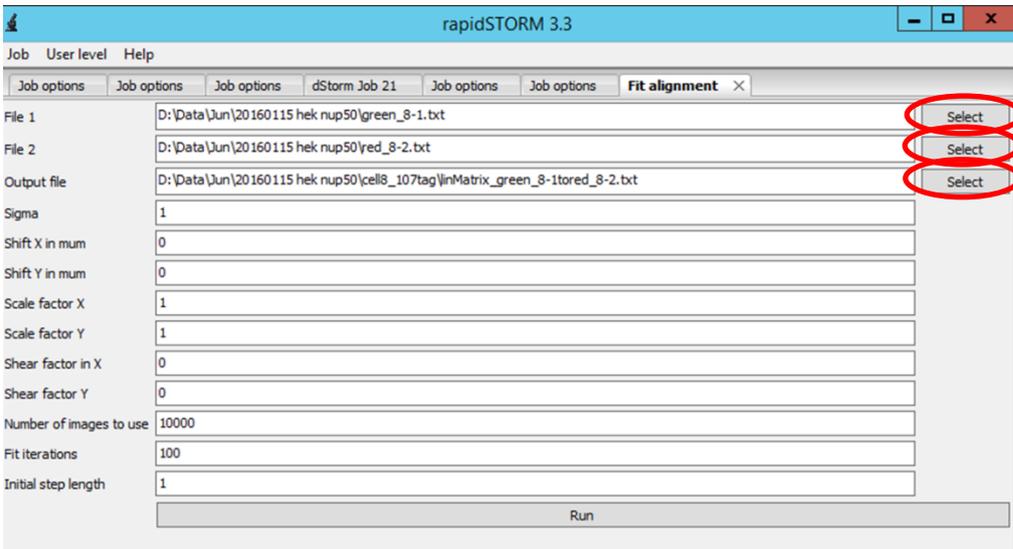


Figure 31 Recommended input settings for output options of localization analysis for channel alignment

Input settings for A) “Expression filter” and B) “Localization file”. The red circles represent settings which need to be manually set.

6.1.2 STEP 2: Alignment of the two channels

The next step is the actual aligning of the two channels based on the localizations from the first step. The recommended settings are shown in Figure 32. The expected value of sigma is ~ 1 and for the shifts it is usually less than $1 \mu\text{m}$. As discussed in section 2.3, the emission dichoric mirror separates the fluorescence into two EMCCDs, shorter wavelength is referred to as “channel 1” while longer wavelength is referred to as “channel 2”. “File 1” is the localization file of channel 1 while “File 2” is channel 2. The localization files are the text files from STEP 1. Please follow the format of the name for the output file as shown in the figure below as recommended by the developers, to minimize bugs.



The screenshot shows the 'rapidSTORM 3.3' software window with the 'Fit alignment' dialog box open. The dialog contains the following fields and values:

Field	Value
File 1	D:\Data\Jun\20160115 hek nup50\green_8-1.txt
File 2	D:\Data\Jun\20160115 hek nup50\red_8-2.txt
Output file	D:\Data\Jun\20160115 hek nup50\cell8_107tag\inMatrix_green_8-1tored_8-2.txt
Sigma	1
Shift X in mum	0
Shift Y in mum	0
Scale factor X	1
Scale factor Y	1
Shear factor in X	0
Shear factor Y	0
Number of images to use	10000
Fit iterations	100
Initial step length	1

Three 'Select' buttons are circled in red, corresponding to the File 1, File 2, and Output file fields.

Figure 32 Recommended input settings alignment of the two channels

The red circles represent settings which need to be manually set.

6.1.3 STEP 3: Localization analysis for reconstruction

After this step is the main localization of the acquired results. Localization in the first step was done for alignment, while in this step localization is done for reconstruction of the SR image. The recommended settings are shown in Figure 33. In brief, the “input files” are the movie files directly from the EMCCD and “number of input channels” is set as 2. For convenience, the shorter wavelength channel is set as “Channel 1”. For the “size of the pixel” and “PSF FWHM”, it is as described in the STEP 1. The “Transmission of fluorophore 0”, AF 647, is 0.5 in both channels. While for CF 680 it is 0.2 and 0.8 for “Channel 1” and “2”, respectively. A “fixed global threshold” for fit judging method is used and set as 50 for the intensity threshold to not continue the fitting if a pixel’s value is less than 50 photons.

In the “output options”, “Expression filter” is selected to only analyze fluorophores which give a certain total number of photons. In my case, I set it to 1000 following a previous study using the same SF [62]. These values relating to the number of photons can be adjusted in order to increase the contrast of the final SR image. By running this analysis, a new window will pop up showing the comparison of isolated PSF in both channels. If only PSFs from a single channel is shown, then there is something wrong in the analysis, either the alignment or the acquisition itself. After this analysis, a single file is saved in the designated folder which contains localization information of PSF from both channels.

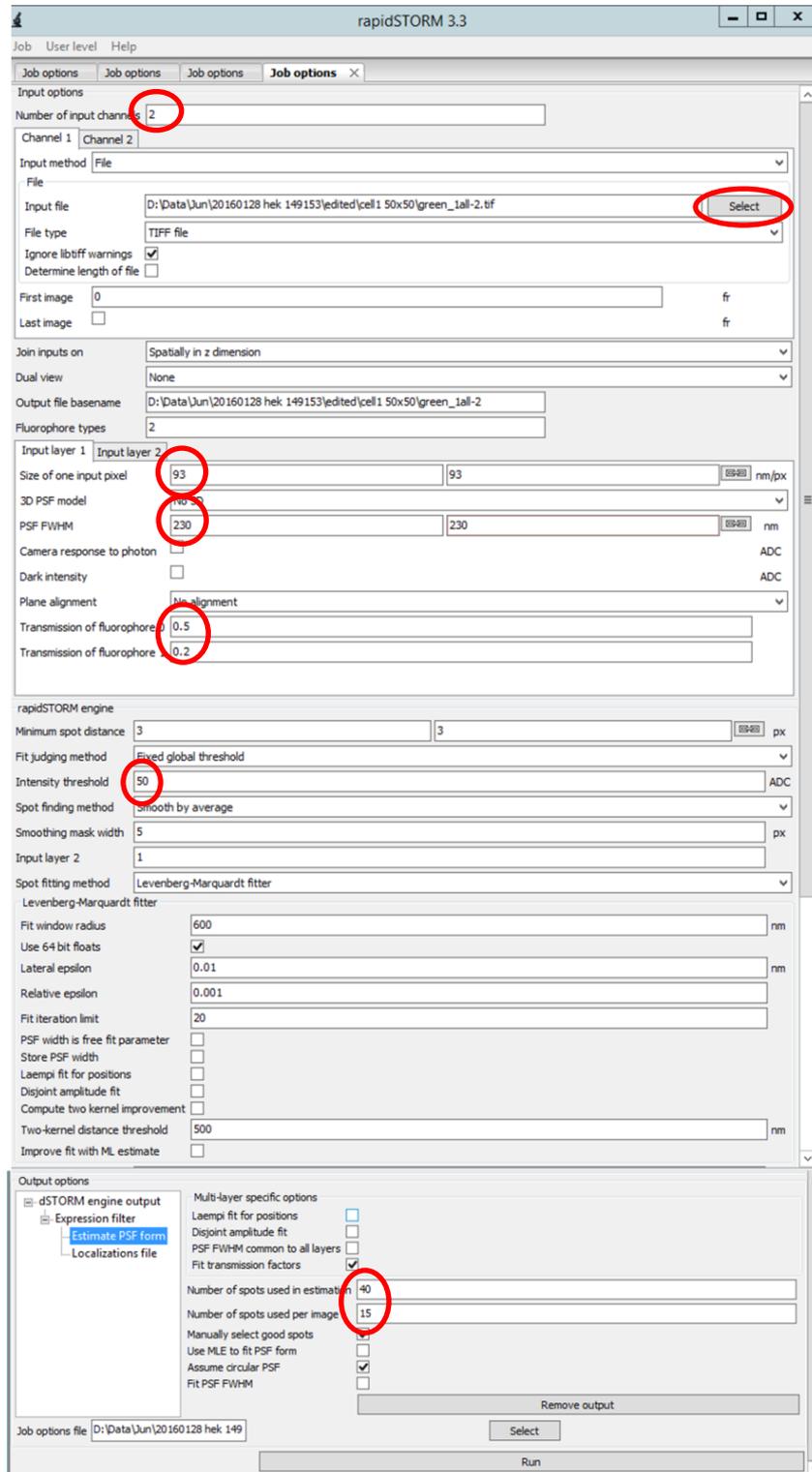


Figure 33 Recommended input settings for localization analysis of two channels

The red circles represent settings which need to be manually set.

6.1.4 STEP 4: Reconstruction of SR image

The last step is using this file containing information about the PSF from both channels to reconstruct the SR image. But before reconstruction, it is recommended to determine the resolution of the localized position using FRC and mentioned in section 1.2.3 details of the procedure is also shown by Banterle et al. [176]. It can also be determined from the localizer package from Dedecker et al. [128] using Igor Pro. This is done by acquiring the histogram of localization error from the output file using “MLEwg” as a fitting function. The center of the histogram is the sigma of the resolution which can be converted to resolution by multiplying 2.355. The pixel size can be approximately half of the determined resolution which can be adjusted comparing the final SR images. For imaging Nups in the NPC, I used a pixel size of 10 nm based on FRC and localizer package resolution. This reconstruction has to be done for each channel, by selecting the appropriate values in the “expression filter” which refers to channel number. Afterwards, the reconstructed images were convolved with a Gaussian function with a standard deviation of 1 pixel using ImageJ. One thing to be noted is that for quantitative analysis of the combined localizations for a single labeled position, it is not at all affected by the SR image. The image is just a representation of the localized positions from the labeled Nups in the NPC.

6 Appendices

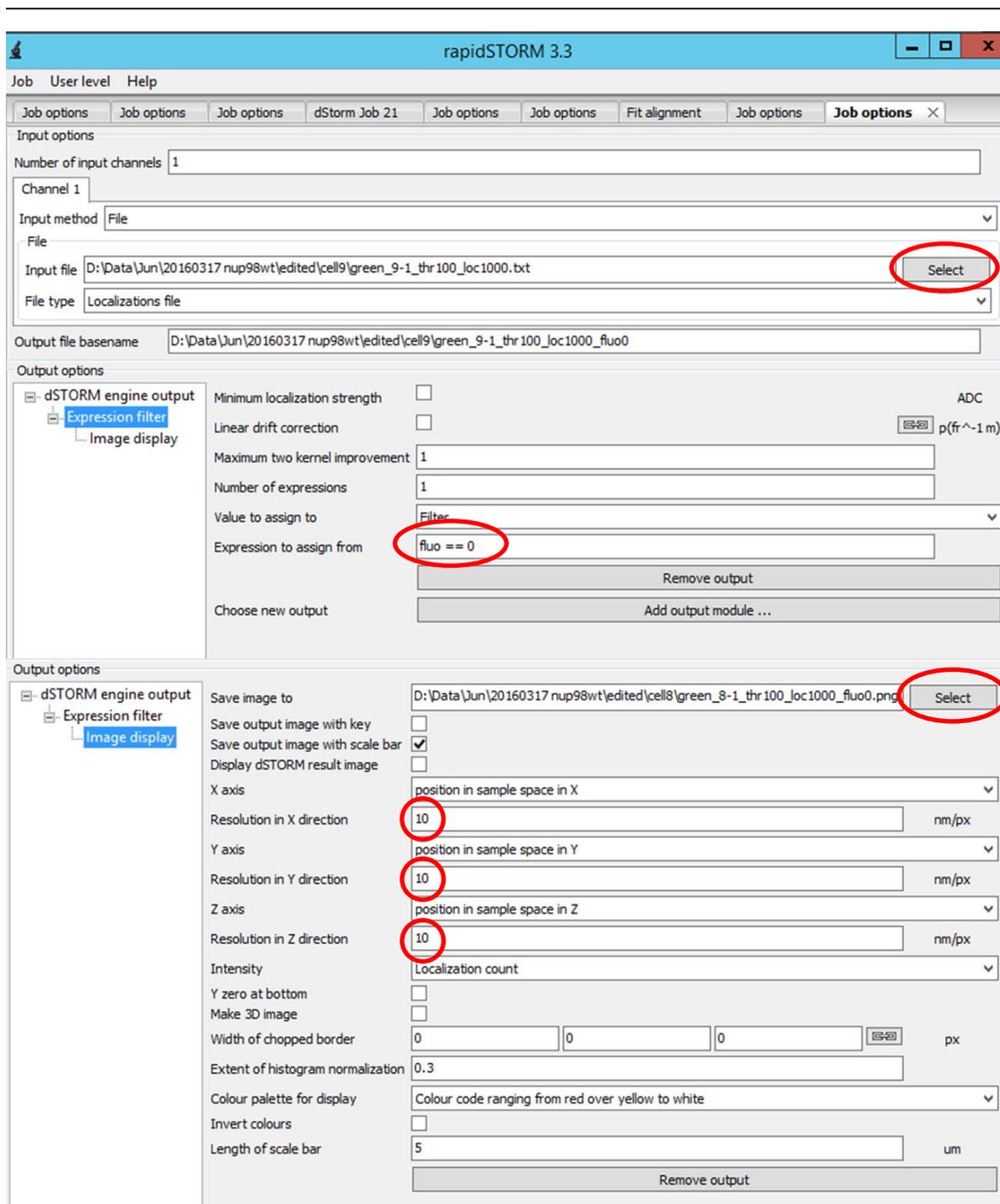


Figure 34 Recommended input settings for reconstruction of SR images

Two images can be reconstructed for each fluorophore used in the experiment.

6.2 Appendix II: Manual for analysis of labeled Nups

First load the ipf file into Igor Pro. Then select Analysis > STORM: Get Fluorophore distances. Then you will be prompted to input values for the following parameters to continue with the analysis. I will discuss in here the purpose of each parameter.

- PSFWidthGlobal is a value used during reconstruction of bitmap image. 6 is recommended.
- Pixel2Distance is the pixel size of the reconstructed SR image.
- IMAGEWIDTH_CONSTANT is the total number of pixels in the final reconstructed SR image. It can be determined by multiplying number of pixels in the acquired movies from the EMCCD to the pixel size. 1280 for the movie with 128x128 pixels and 10 nm pixel size for reconstruction.
- MINDISTANCE_RED is the minimum distance between the isolated WGAs. If two isolated WGA are closer than this value then both are discarded.
- RedSpotSigmaThreshold (pixels) is the threshold sigma value of the Gaussian function that is used to fit the isolated WGA. If the sigma from the fitted Gaussian function is higher than this value then the WGA is discarded.
- MinCircularityThreshold can be used instead of RedSpotSigmaThreshold when the ellipticity of the WGA is used as a criteria to sort out poorly labeled WGAs.
- BlurredSizeFactor is a value used when reconstructing image for the isolated NPCs. A value of 1 can be used for 10 nm pixel size.
- WINDOWSIZE_X and Y is the size of the window designated around the WGA for analysis of the labeled Nups (in nm). 150 nm seems to be the most appropriate value.
- MINPOINTS_CLUSTER is the minimum number of localizations in a certain distance from each other to be considered as a cluster. 25 is recommended which is based on control experiments with the WT Nups.
- MINDISTANCE_GREEN is minimum distance to distinguish one cluster from the other.
- DISTANCEHISTO_BINS is the number of bins in the histogram of localized positions of labeled Nups.
- MinRemovePixels is the minimum number of pixels around the center pixel required after second round of wavelet transformation to be considered as a WGA for analysis. 3 is the recommended value for WGA.
- MinAreaRedParticles is the minimum area (in pixels) of an isolated WGA after wavelet transformation. 1 is the recommended value for WGA.
- WaveletDepth is the number of iteration for wavelet transformation. 3 is the recommended value.
- TooCloseThreshold is the threshold value used. to combine bright spots as contained in a single NPC. 3 is the recommended value.
- GAUSSIANBLURRED_PIXEL_X and Y is the number of pixels of isolated NPC in the final output.

6 Appendices

- `RGBIMAGE_SIGMA`: is the sigma value of the convoluted Gaussian in the final reconstructed SR image.
- `RED_SINGLE_THRESHOLD_DIST` can be ignored because wavelet transformation is used. A value of 1 million will ignore its function.
- `RED_SINGLE_THRESHOLD_N` also can be ignored because wavelet transformation is used. It was used to discard NPCs which were over-labeled with WGAs.
- `APPLY_GREENQUADRANGLE`, if yes is selected then in the next window the coordinates of the window has to be selected. Clustering algorithm will be applied only in this area.

Then after setting the parameters, next step is to load the localization file of two channels from RapidSTORM analysis. After loading the file, a window will be shown as in Figure 35. In here, it is also possible to select a quadrangle in order to analyze only the nucleus and not the background noise outside the nucleus. Then, click on “Proceed”. After this, the clustering algorithm will start which may take as short as 3 hrs to as long as 6 hrs depending on the number of localizations.

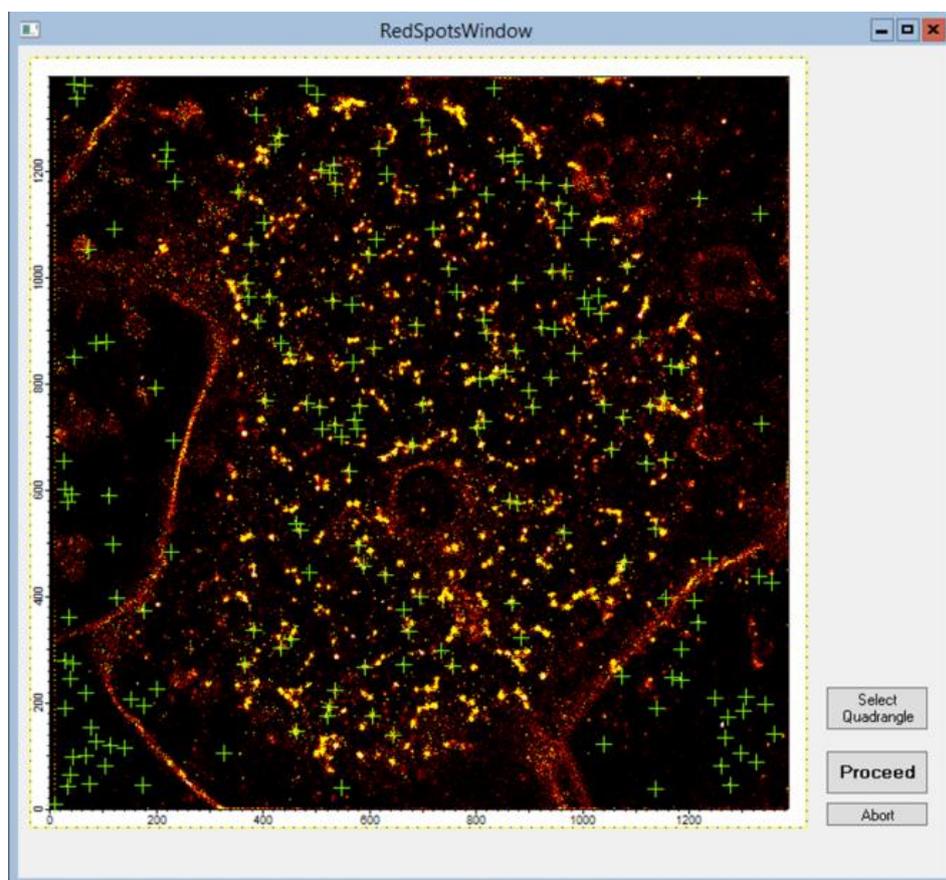


Figure 35 A sample bitmap image after setting the parameters for analysis

In this bitmap image, it is possible to select a quadrangle in order to analyze fluorescence spots only in the selected area.

The output of the analysis is shown in Figure 36. “Save” and “Discard” button is to take out outliers to see how it affects the overall Gaussian fit which can be used for troubleshooting. “Min” and “Max” for “Red” and “Green” channels can be used to change the contrast of the images. R and sigma is shown for each isolated NPC after analyzing the localizations inside the window. “Spots + Circle” will show all the combined localized positions of the isolated NPCs and shown at the background is a circle with the radius from the Gaussian fit of the histogram of the localized positions and its sigma for the thickness of the circle. “Spots + Blurred” will show all the localizations which are convolved with a Gaussian function. The parameters of the Gaussian function are set at the beginning of the analysis. “Image + Positions” will show the reconstructed SR image showing the number of each NPC included in the analysis and the also the total area that was selected for analysis (Green quadrangle). Sample images are shown in **Figure 37**.

It is also possible to combine all the analysis from different sets of analysis. A function called “Analyzefolder(foldername)” will be able to this. In all the single set of analysis, you will be able to find a wave called “Spots” in subfolder “Results”. If the wave is not present, click on “Spots + Gaussian” button to generate the wave. Copy all the “Spots” wave folder in a same subfolder then execute “Analyzefolder(foldername)”. Foldername is the name of the folder containing all the “Spots” waves. Then a histogram together with a Gaussian blurred localization image will be shown.

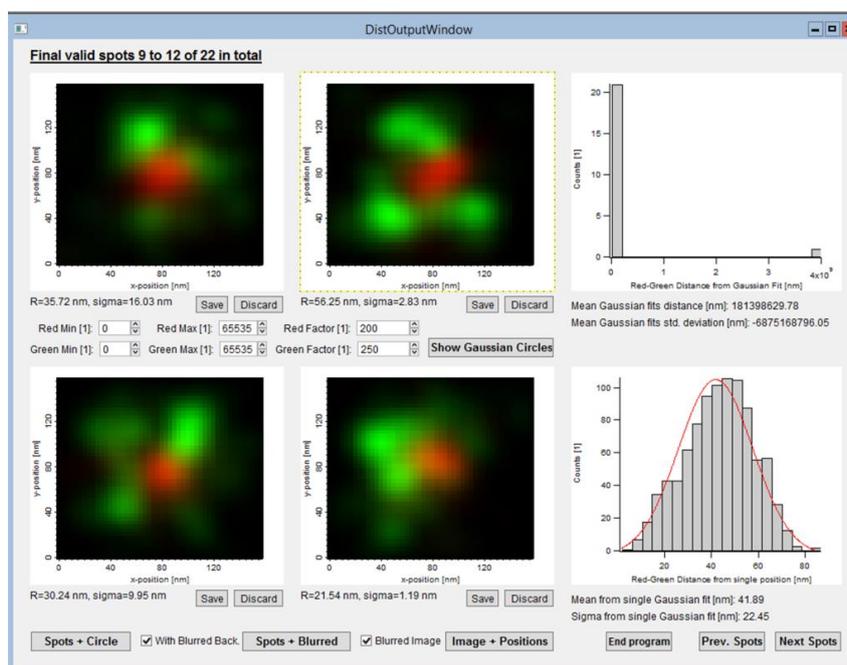


Figure 36 Output window of the custom-written analysis

The output shows the isolated NPCs (left side), a histogram of the average Gaussian fits of each NPCs (right upper) and an overall histogram of all the localizations from isolated NPCs (right lower)

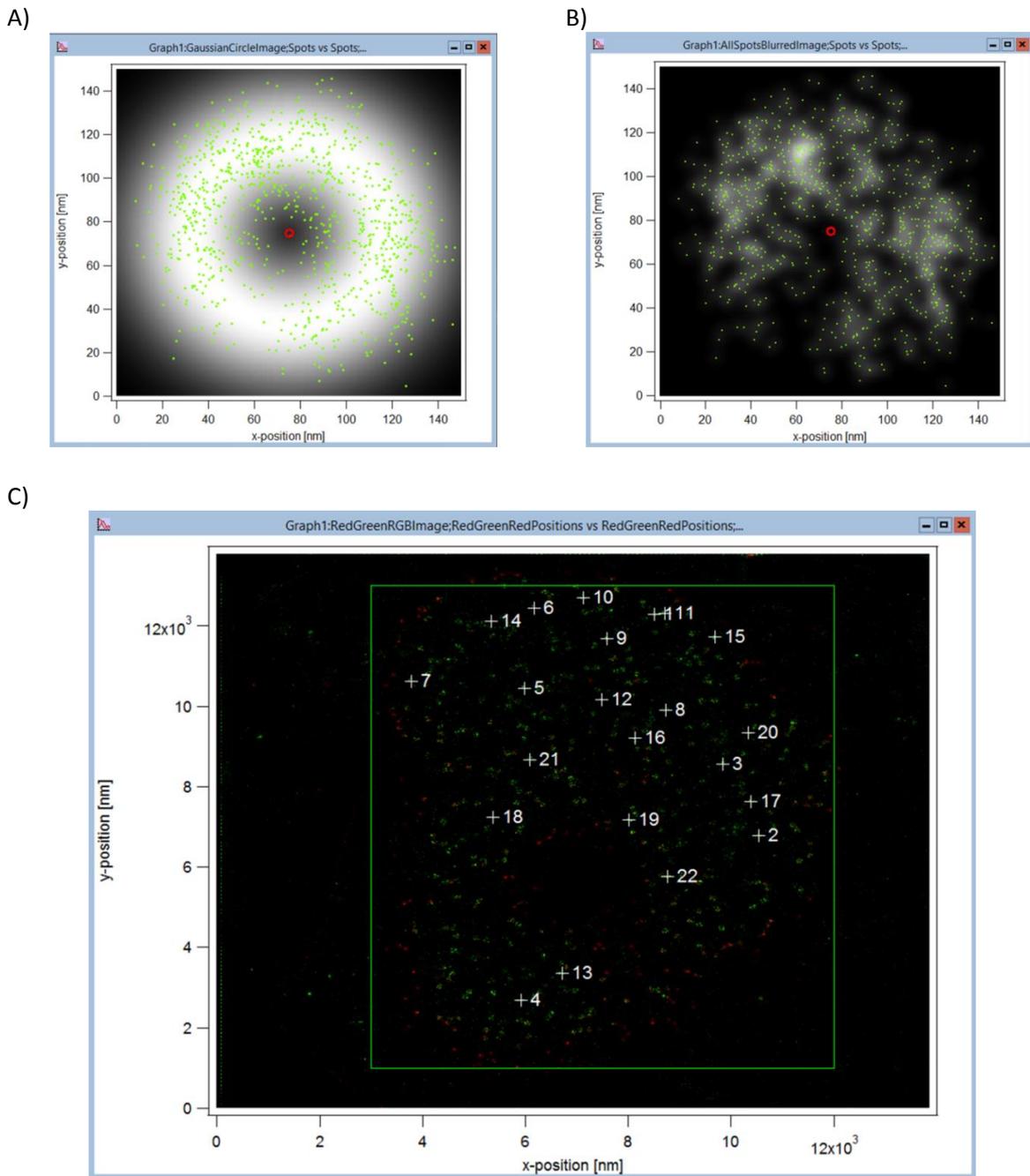


Figure 37 Additional outputs of custom-written analysis

A) shows all the combined localizations as green dots. The radius of the circle is the center of the Gaussian function fitted to the histogram of all the combined localizations. The thickness of the circle is from the sigma of the fitted Gaussian function. B) also shows all the localizations as green dots. The white background is from Gaussian blurring of each of the localizations. C) is the reconstructed SR image without Gaussian convolution of showing which NPC was isolated for analysis

7 References

1. Grossman, E., O. Medalia, and M. Zwerger, *Functional Architecture of the Nuclear Pore Complex*. Annual Review of Biophysics, Vol 41, 2012. **41**: p. 557-584.
2. Beck, M., et al., *Snapshots of nuclear pore complexes in action captured by cryo-electron tomography*. Nature, 2007. **449**(7162): p. 611-5.
3. Frenkiel-Krispin, D., et al., *Structural Analysis of a Metazoan Nuclear Pore Complex Reveals a Fused Concentric Ring Architecture*. Journal of Molecular Biology, 2010. **395**(3): p. 578-586.
4. Ori, A., et al., *Cell type-specific nuclear pores: a case in point for context-dependent stoichiometry of molecular machines*. Mol Syst Biol, 2013. **9**: p. 648.
5. Maimon, T., et al., *The Human Nuclear Pore Complex as Revealed by Cryo-Electron Tomography*. Structure, 2012. **20**(6): p. 998-1006.
6. Andrade, M.A. and P. Bork, *HEAT repeats in the Huntington's disease protein*. Nat Genet, 1995. **11**(2): p. 115-6.
7. Vetter, I.R., et al., *Structural view of the Ran-importin beta interaction at 2.3 angstrom resolution*. Cell, 1999. **97**(5): p. 635-646.
8. Peters, R., *Translocation through the nuclear pore: Kaps pave the way*. Bioessays, 2009. **31**(4): p. 466-477.
9. Goldfarb, D.S., et al., *Importin alpha: a multipurpose nuclear-transport receptor*. Trends in Cell Biology, 2004. **14**(9): p. 505-514.
10. Cingolani, G., et al., *Structure of importin-beta bound to the IBB domain of importin-alpha*. Nature, 1999. **399**(6733): p. 221-229.
11. Bednenko, J., G. Cingolani, and L. Gerace, *Importin beta contains a COOH-terminal nucleoporin binding region important for nuclear transport*. Journal of Cell Biology, 2003. **162**(3): p. 391-401.
12. Patel, S.S. and M.F. Rexach, *Discovering novel interactions at the nuclear pore complex using bead halo*. Molecular & Cellular Proteomics, 2008. **7**(1): p. 121-131.
13. Tu, L.C., et al., *Large cargo transport by nuclear pores: implications for the spatial organization of FG-nucleoporins*. Embo Journal, 2013. **32**(24): p. 3220-3230.
14. Milles, S., et al., *Plasticity of an Ultrafast Interaction between Nucleoporins and Nuclear Transport Receptors*. Cell, 2015. **163**(3): p. 734-745.
15. Sun, C., et al., *Choreography of importin-alpha/CAS complex assembly and disassembly at nuclear pores*. Proc Natl Acad Sci U S A, 2013. **110**(17): p. E1584-93.
16. Kutay, U., et al., *Export of importin alpha from the nucleus is mediated by a specific nuclear transport factor*. Cell, 1997. **90**(6): p. 1061-71.
17. Jamali, T., et al., *Nuclear Pore Complex: Biochemistry and Biophysics of Nucleocytoplasmic Transport in Health and Disease*. International Review of Cell and Molecular Biology, Vol 287, 2011. **287**: p. 233-286.
18. Wenthe, S.R. and M.P. Rout, *The nuclear pore complex and nuclear transport*. Cold Spring Harb Perspect Biol, 2010. **2**(10): p. a000562.
19. Hutten, S., et al., *The Nup358-RanGAP complex is required for efficient importin alpha/beta-dependent nuclear import*. Mol Biol Cell, 2008. **19**(5): p. 2300-10.
20. Macara, I.G., *Transport into and out of the nucleus*. Microbiol Mol Biol Rev, 2001. **65**(4): p. 570-94, table of contents.

7 References

21. Ribbeck, K. and D. Gorlich, *Kinetic analysis of translocation through nuclear pore complexes*. *Embo Journal*, 2001. **20**(6): p. 1320-1330.
22. Frey, S. and D. Gorlich, *A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes*. *Cell*, 2007. **130**(3): p. 512-523.
23. Frey, S. and D. Gorlich, *FG/FxFG as well as GLFG repeats form a selective permeability barrier with self-healing properties*. *Embo Journal*, 2009. **28**(17): p. 2554-2567.
24. Lim, R.Y., et al., *Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport*. *Proc Natl Acad Sci U S A*, 2006. **103**(25): p. 9512-7.
25. Lim, R.Y., et al., *Nanomechanical basis of selective gating by the nuclear pore complex*. *Science (New York, N.Y.)*, 2007. **318**(5850): p. 640-643.
26. Sakiyama, Y., et al., *Spatiotemporal dynamics of the nuclear pore complex transport barrier resolved by high-speed atomic force microscopy*. *Nat Nanotechnol*, 2016.
27. Rout, M.P., et al., *Virtual gating and nuclear transport: the hole picture*. *Trends Cell Biol*, 2003. **13**(12): p. 622-8.
28. Kubitscheck, U., et al., *Nuclear transport of single molecules: dwell times at the nuclear pore complex*. *J Cell Biol*, 2005. **168**(2): p. 233-43.
29. Peters, R., *Translocation through the nuclear pore complex: Selectivity and speed by reduction-of-dimensionality*. *Traffic*, 2005. **6**(5): p. 421-427.
30. Yamada, J., et al., *A Bimodal Distribution of Two Distinct Categories of Intrinsically Disordered Structures with Separate Functions in FG Nucleoporins*. *Molecular & Cellular Proteomics*, 2010. **9**(10): p. 2205-2224.
31. Milles, S. and E.A. Lemke, *Nup153 in Nuclear Transport: Plasticity of Nucleoporin/Transport-Receptor Complexes*. *Biophysical Journal*, 2013. **104**(2): p. 30a-30a.
32. Ferreon, A.C.M., et al., *Modulation of allostery by protein intrinsic disorder*. *Nature*, 2013. **498**(7454): p. 390-+.
33. Uttamapinant, C., et al., *Genetic Code Expansion Enables Live-Cell and Super-Resolution Imaging of Site-Specifically Labeled Cellular Proteins*. *Journal of the American Chemical Society*, 2015. **137**(14): p. 4602-4605.
34. Szymborska, A., et al., *Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging*. *Science*, 2013. **341**(6146): p. 655-8.
35. Kosinski, J., et al., *Molecular architecture of the inner ring scaffold of the human nuclear pore complex*. *Science*, 2016. **352**(6283): p. 363-365.
36. Henderson, R. and P.N.T. Unwin, *3-Dimensional Model of Purple Membrane Obtained by Electron-Microscopy*. *Nature*, 1975. **257**(5521): p. 28-32.
37. Salih, S.M. and V.E. Cosslett, *Radiation-Damage in Electron-Microscopy of Organic Materials - Effect of Low-Temperatures*. *Journal of Microscopy-Oxford*, 1975. **105**(Dec): p. 269-276.
38. Milne, J.L., et al., *Cryo-electron microscopy--a primer for the non-microscopist*. *FEBS J*, 2013. **280**(1): p. 28-45.
39. Rout, M.P., et al., *The yeast nuclear pore complex: Composition, architecture, and transport mechanism*. *Journal of Cell Biology*, 2000. **148**(4): p. 635-651.
40. Bui, K.H., et al., *Integrated Structural Analysis of the Human Nuclear Pore Complex Scaffold*. *Cell*, 2013. **155**(6): p. 1233-1243.
41. Ritterhoff, T., et al., *The RanBP2/RanGAP1*SUMO1/Ubc9 SUMO E3 ligase is a disassembly machine for Crm1-dependent nuclear export complexes*. *Nat Commun*, 2016. **7**: p. 11482.

7 References

42. von Appen, A., et al., *In situ structural analysis of the human nuclear pore complex*. Nature, 2015. **526**(7571): p. 140-3.
43. Muller, D.J. and Y.F. Dufrene, *Atomic force microscopy as a multifunctional molecular toolbox in nanobiotechnology*. Nat Nanotechnol, 2008. **3**(5): p. 261-9.
44. Ando, T., T. Uchihashi, and N. Kodera, *High-speed AFM and applications to biomolecular systems*. Annu Rev Biophys, 2013. **42**: p. 393-414.
45. Moy, V.T., E.L. Florin, and H.E. Gaub, *Intermolecular forces and energies between ligands and receptors*. Science, 1994. **266**(5183): p. 257-9.
46. Jaggi, R.D., et al., *Modulation of nuclear pore topology by transport modifiers*. Biophysical Journal, 2003. **84**(1): p. 665-670.
47. Bestembayeva, A., et al., *Nanoscale stiffness topography reveals structure and mechanics of the transport barrier in intact nuclear pore complexes*. Nature Nanotechnology, 2015. **10**(1): p. 60-64.
48. Jost, A. and R. Heintzmann, *Superresolution Multidimensional Imaging with Structured Illumination Microscopy*. Annual Review of Materials Research, Vol 43, 2013. **43**: p. 261-282.
49. Hell, S.W., *Microscopy and its focal switch*. Nature Methods, 2009. **6**(1): p. 24-32.
50. Rego, E.H., et al., *Nonlinear structured-illumination microscopy with a photoswitchable protein reveals cellular structures at 50-nm resolution*. Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(3): p. E135-E143.
51. Hell, S.W., *Far-field optical nanoscopy*. Science, 2007. **316**(5828): p. 1153-1158.
52. Gould, T.J., S.T. Hess, and J. Bewersdorf, *Optical nanoscopy: from acquisition to analysis*. Annu Rev Biomed Eng, 2012. **14**: p. 231-54.
53. Huang, B., M. Bates, and X.W. Zhuang, *Super-Resolution Fluorescence Microscopy*. Annual Review of Biochemistry, 2009. **78**: p. 993-1016.
54. Heilemann, M., et al., *Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes*. Angew Chem Int Ed Engl, 2008. **47**(33): p. 6172-6.
55. Winterflood, C.M., et al., *Dual-Color 3D Superresolution Microscopy by Combined Spectral-Demixing and Biplane Imaging*. Biophysical Journal, 2015. **109**(1): p. 3-6.
56. Banterle, N., et al., *Fourier ring correlation as a resolution criterion for super-resolution microscopy*. J Struct Biol, 2013. **183**(3): p. 363-7.
57. Thompson, R.E., D.R. Larson, and W.W. Webb, *Precise nanometer localization analysis for individual fluorescent probes*. Biophys J, 2002. **82**(5): p. 2775-83.
58. Deschout, H., et al., *Precisely and accurately localizing single emitters in fluorescence microscopy*. Nat Methods, 2014. **11**(3): p. 253-66.
59. Fitzgerald, J.E., J. Lu, and M.J. Schnitzer, *Estimation theoretic measure of resolution for stochastic localization microscopy*. Phys Rev Lett, 2012. **109**(4): p. 048102.
60. Nieuwenhuizen, R.P.J., et al., *Measuring image resolution in optical nanoscopy*. Nature Methods, 2013. **10**(6): p. 557-+.
61. Gottfert, F., et al., *Coaligned Dual-Channel STED Nanoscopy and Molecular Diffusion Analysis at 20 nm Resolution*. Biophysical Journal, 2013. **105**(1): p. L1-L3.
62. Loschberger, A., et al., *Super-resolution imaging visualizes the eightfold symmetry of gp210 proteins around the nuclear pore complex and resolves the central channel with nanometer resolution*. Journal of Cell Science, 2012. **125**(3): p. 570-575.

7 References

63. Platonova, E., C.M. Winterflood, and H. Ewers, *A Simple Method for GFP- and RFP-based Dual Color Single-Molecule Localization Microscopy*. *Acs Chemical Biology*, 2015. **10**(6): p. 1411-1416.
64. Chudakov, D.M., et al., *Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues*. *Physiological Reviews*, 2010. **90**(3): p. 1103-1163.
65. Griffiths, G. and J.M. Lucocq, *Antibodies for immunolabeling by light and electron microscopy: not for the faint hearted*. *Histochemistry and Cell Biology*, 2014. **142**(4): p. 347-360.
66. Opazo, F., et al., *Aptamers as potential tools for super-resolution microscopy*. *Nature Methods*, 2012. **9**(10): p. 938-939.
67. Ries, J., et al., *A simple, versatile method for GFP-based super-resolution microscopy via nanobodies*. *Nature Methods*, 2012. **9**(6): p. 582-+.
68. Jungmann, R., et al., *Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT*. *Nature Methods*, 2014. **11**(3): p. 313-U292.
69. Jungmann, R., et al., *Quantitative super-resolution imaging with qPAINT*. *Nature Methods*, 2016. **13**(5): p. 439-442.
70. Xia, T., N. Li, and X. Fang, *Single-molecule fluorescence imaging in living cells*. *Annu Rev Phys Chem*, 2013. **64**: p. 459-80.
71. Keppler, A., et al., *A general method for the covalent labeling of fusion proteins with small molecules in vivo*. *Nature Biotechnology*, 2003. **21**(1): p. 86-89.
72. Keppler, A., et al., *Labeling of fusion proteins with synthetic fluorophores in live cells*. *Proceedings of the National Academy of Sciences of the United States of America*, 2004. **101**(27): p. 9955-9959.
73. Hoffmann, C., et al., *Fluorescent labeling of tetracysteine-tagged proteins in intact cells*. *Nature Protocols*, 2010. **5**(10): p. 1666-1677.
74. Yan, Q. and M.P. Bruchez, *Advances in chemical labeling of proteins in living cells*. *Cell and Tissue Research*, 2015. **360**(1): p. 179-194.
75. Uttamapinant, C., et al., *A fluorophore ligase for site-specific protein labeling inside living cells*. *Proceedings of the National Academy of Sciences of the United States of America*, 2010. **107**(24): p. 10914-10919.
76. Liu, D.S., et al., *Computational design of a red fluorophore ligase for site-specific protein labeling in living cells*. *Proceedings of the National Academy of Sciences of the United States of America*, 2014. **111**(43): p. E4551-E4559.
77. Wombacher, R., et al., *Live-cell super-resolution imaging with trimethoprim conjugates*. *Nature Methods*, 2010. **7**(9): p. 717-719.
78. Szent-Gyorgyi, C., et al., *Fluorogen-activating single-chain antibodies for imaging cell surface proteins (vol 26, pg 235, 2008)*. *Nature Biotechnology*, 2008. **26**(4): p. 470-470.
79. de Castro, M.A., B. Rammner, and F. Opazo, *Aptamer Stainings for Super-resolution Microscopy*. *Methods Mol Biol*, 2016. **1380**: p. 197-210.
80. Milles, S., et al., *Click Strategies for Single-Molecule Protein Fluorescence*. *Journal of the American Chemical Society*, 2012. **134**(11): p. 5187-5195.
81. Milles, S. and E.A. Lemke, *What precision-protein-tuning and nano-resolved single molecule sciences can do for each other*. *Bioessays*, 2013. **35**(1): p. 65-74.
82. Mahdavi, A., et al., *Engineered Aminoacyl-tRNA Synthetase for Cell-Selective Analysis of Mammalian Protein Synthesis*. *J Am Chem Soc*, 2016. **138**(13): p. 4278-81.

7 References

83. Noren, C.J., et al., *A general method for site-specific incorporation of unnatural amino acids into proteins*. *Science*, 1989. **244**(4901): p. 182-8.
84. Chin, J.W., *Expanding and reprogramming the genetic code of cells and animals*. *Annu Rev Biochem*, 2014. **83**: p. 379-408.
85. Sakamoto, K., et al., *Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells*. *Nucleic Acids Res*, 2002. **30**(21): p. 4692-9.
86. Liu, W., et al., *Genetic incorporation of unnatural amino acids into proteins in mammalian cells*. *Nat Methods*, 2007. **4**(3): p. 239-44.
87. Srinivasan, G., C.M. James, and J.A. Krzycki, *Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA*. *Science*, 2002. **296**(5572): p. 1459-62.
88. Plass, T., et al., *Genetically Encoded Copper-Free Click Chemistry*. *Angewandte Chemie-International Edition*, 2011. **50**(17): p. 3878-3881.
89. Yanagisawa, T., et al., *Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzoyloxycarbonyl) lysine for site-specific protein modification*. *Chem Biol*, 2008. **15**(11): p. 1187-97.
90. Plass, T., et al., *Amino Acids for Diels-Alder Reactions in Living Cells*. *Angewandte Chemie-International Edition*, 2012. **51**(17): p. 4166-4170.
91. Nikic, I., et al., *Minimal tags for rapid dual-color live-cell labeling and super-resolution microscopy*. *Angew Chem Int Ed Engl*, 2014. **53**(8): p. 2245-9.
92. Neumann, H., S.Y. Peak-Chew, and J.W. Chin, *Genetically encoding N(epsilon)-acetyllysine in recombinant proteins*. *Nat Chem Biol*, 2008. **4**(4): p. 232-4.
93. Chin, J.W. and P.G. Schultz, *In vivo photocrosslinking with unnatural amino acid mutagenesis*. *Chembiochem*, 2002. **3**(11): p. 1135-7.
94. Wang, L. and P.G. Schultz, *Expanding the genetic code*. *Chem Commun (Camb)*, 2002(1): p. 1-11.
95. Gautier, A., et al., *Genetically encoded photocontrol of protein localization in mammalian cells*. *J Am Chem Soc*, 2010. **132**(12): p. 4086-8.
96. Lemke, E.A., et al., *Control of protein phosphorylation with a genetically encoded photocaged amino acid*. *Nat Chem Biol*, 2007. **3**(12): p. 769-72.
97. Brustad, E.M., et al., *A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer*. *J Am Chem Soc*, 2008. **130**(52): p. 17664-5.
98. Kolb, H.C., M.G. Finn, and K.B. Sharpless, *Click Chemistry: Diverse Chemical Function from a Few Good Reactions*. *Angew Chem Int Ed Engl*, 2001. **40**(11): p. 2004-2021.
99. Jewett, J.C. and C.R. Bertozzi, *Cu-free click cycloaddition reactions in chemical biology*. *Chem Soc Rev*, 2010. **39**(4): p. 1272-9.
100. Agard, N.J., J.A. Prescher, and C.R. Bertozzi, *A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems*. *J Am Chem Soc*, 2004. **126**(46): p. 15046-7.
101. Huisgen, R., *Kinetics and Mechanism of 1,3-Dipolar Cycloadditions*. *Angewandte Chemie International Edition in English*, 1963. **2**(11): p. 633-645.
102. Lang, K. and J.W. Chin, *Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins*. *Chemical Reviews*, 2014. **114**(9): p. 4764-4806.
103. Jao, C.Y., et al., *Metabolic labeling and direct imaging of choline phospholipids in vivo*. *Proceedings of the National Academy of Sciences of the United States of America*, 2009. **106**(36): p. 15332-15337.

7 References

104. Shelbourne, M., et al., *Fast copper-free click DNA ligation by the ring-strain promoted alkyne-azide cycloaddition reaction*. Chem Commun (Camb), 2011. **47**(22): p. 6257-9.
105. Rostovtsev, V.V., et al., *A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes*. Angewandte Chemie-International Edition, 2002. **41**(14): p. 2596-+.
106. Soriano Del Amo, D., et al., *Biocompatible copper(I) catalysts for in vivo imaging of glycans*. J Am Chem Soc, 2010. **132**(47): p. 16893-9.
107. Hong, V., et al., *Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation*. Angew Chem Int Ed Engl, 2009. **48**(52): p. 9879-83.
108. Uttamapinant, C., et al., *Fast, cell-compatible click chemistry with copper-chelating azides for biomolecular labeling*. Angew Chem Int Ed Engl, 2012. **51**(24): p. 5852-6.
109. Yang, M., et al., *Biocompatible click chemistry enabled compartment-specific pH measurement inside E. coli*. Nat Commun, 2014. **5**: p. 4981.
110. Alder, K., G. Stein, and H. Finzenhagen, *Über das abgestufte Additionsvermögen von ungesättigten Ringsystemen*. Justus Liebig's Annalen der Chemie, 1931. **485**(1): p. 211-222.
111. Wittig, G. and A. Krebs, *Zur Existenz niedergliedriger Cycloalkine, I*. Chemische Berichte, 1961. **94**(12): p. 3260-3275.
112. Baskin, J.M. and C.R. Bertozzi, *Bioorthogonal click chemistry: Covalent labeling in living systems*. Qsar & Combinatorial Science, 2007. **26**(11-12): p. 1211-1219.
113. Laughlin, S.T., et al., *In vivo imaging of membrane-associated glycans in developing zebrafish*. Science, 2008. **320**(5876): p. 664-667.
114. Ning, X.H., et al., *Protein Modification by Strain-Promoted Alkyne-Nitrone Cycloaddition*. Angewandte Chemie-International Edition, 2010. **49**(17): p. 3065-3068.
115. McKay, C.S., et al., *Strain-promoted cycloadditions of cyclic nitrones with cyclooctynes for labeling human cancer cells*. Chemical Communications, 2011. **47**(36): p. 10040-10042.
116. Carboni, R.A. and R.V. Lindsey, *Reactions of Tetrazines with Unsaturated Compounds. A New Synthesis of Pyridazines*. Journal of the American Chemical Society, 1959. **81**(16): p. 4342-4346.
117. Clavier, G. and P. Audebert, *s-Tetrazines as Building Blocks for New Functional Molecules and Molecular Materials*. Chemical Reviews, 2010. **110**(6): p. 3299-3314.
118. Thalhammer, F., U. Wallfahrer, and J. Sauer, *Reactivity of Simple Open-Chain and Cyclic Dienophiles in Inverse-Type Diels-Alder Reactions*. Tetrahedron Letters, 1990. **31**(47): p. 6851-6854.
119. Wijnen, J.W., et al., *Substituent effects on an inverse electron demand hetero Diels-Alder reaction in aqueous solution and organic solvents: Cycloaddition of substituted styrenes to di(2-pyridyl)-1,2,4,5-tetrazine*. Journal of Organic Chemistry, 1996. **61**(6): p. 2001-2005.
120. Seckute, J. and N.K. Devaraj, *Expanding room for tetrazine ligations in the in vivo chemistry toolbox*. Current Opinion in Chemical Biology, 2013. **17**(5): p. 761-767.
121. Devaraj, N.K., et al., *Fast and Sensitive Pretargeted Labeling of Cancer Cells through a Tetrazine/trans-Cyclooctene Cycloaddition*. Angewandte Chemie-International Edition, 2009. **48**(38): p. 7013-7016.
122. Devaraj, N.K., et al., *Bioorthogonal Turn-On Probes for Imaging Small Molecules inside Living Cells*. Angewandte Chemie-International Edition, 2010. **49**(16): p. 2869-2872.
123. Haun, J.B., et al., *Bioorthogonal chemistry amplifies nanoparticle binding and enhances the sensitivity of cell detection*. Nature Nanotechnology, 2010. **5**(9): p. 660-665.

7 References

124. Han, H.S., et al., *Development of a Bioorthogonal and Highly Efficient Conjugation Method for Quantum Dots Using Tetrazine-Norbornene Cycloaddition*. Journal of the American Chemical Society, 2010. **132**(23): p. 7838-+.
125. Hoffmann, J.E., et al., *Highly Stable trans-Cyclooctene Amino Acids for Live-Cell Labeling*. Chemistry, 2015. **21**(35): p. 12266-70.
126. Hilderbrand, S., Devaraj, N. K., Weissleder, R., *Compositions and methods for delivering a substance to a biological target*, in *United States Patent Application Publication*, U.S.P.A. Publication, Editor. 2013: U.S.A.
127. Wolter, S., et al., *rapidSTORM: accurate, fast open-source software for localization microscopy*. Nature Methods, 2012. **9**(11): p. 1040-1041.
128. Dedecker, P., et al., *Localizer: fast, accurate, open-source, and modular software package for superresolution microscopy*. J Biomed Opt, 2012. **17**(12): p. 126008.
129. Olivo-Marin, J.C., *Extraction of spots in biological images using multiscale products*. Pattern Recognition, 2002. **35**(9): p. 1989-1996.
130. Ester, M., *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. Kdd, 1996. **96**(34): p. 226-231.
131. Plass, T., et al., *Amino acids for Diels-Alder reactions in living cells*. Angew Chem Int Ed Engl, 2012. **51**(17): p. 4166-70.
132. Nikic, I., et al., *Labeling proteins on live mammalian cells using click chemistry*. Nat Protoc, 2015. **10**(5): p. 780-91.
133. Adam, S.A., *Nuclear Protein Transport in Digitonin Permeabilized Cells*. Methods Mol Biol, 2016. **1411**: p. 479-87.
134. Niederwieser, A., et al., *Two-color glycan labeling of live cells by a combination of Diels-Alder and click chemistry*. Angew Chem Int Ed Engl, 2013. **52**(15): p. 4265-8.
135. Dempsey, G.T., et al., *Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging*. Nat Methods, 2011. **8**(12): p. 1027-36.
136. Griffis, E.R., S. Xu, and M.A. Powers, *Nup98 localizes to both nuclear and cytoplasmic sides of the nuclear pore and binds to two distinct nucleoporin subcomplexes*. Mol Biol Cell, 2003. **14**(2): p. 600-10.
137. Griffis, E.R., et al., *Nup98 is a mobile nucleoporin with transcription-dependent dynamics*. Mol Biol Cell, 2002. **13**(4): p. 1282-97.
138. Abmayr, S.M., et al., *Preparation of nuclear and cytoplasmic extracts from mammalian cells*. Curr Protoc Mol Biol, 2006. **Chapter 12**: p. Unit 12 1.
139. Winterflood, C.M. and H. Ewers, *Single-Molecule Localization Microscopy using mCherry*. Chemphyschem, 2014. **15**(16): p. 3447-3451.
140. Ball, J.R. and K.S. Ullman, *Versatility at the nuclear pore complex: lessons learned from the nucleoporin Nup153*. Chromosoma, 2005. **114**(5): p. 319-30.
141. Lukinavicius, G., et al., *A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins*. Nat Chem, 2013. **5**(2): p. 132-9.
142. Kozma, E., et al., *Hydrophilic trans-Cyclooctenylated Non-Canonical Amino Acids for Fast Intracellular Protein Labeling*. Chembiochem, 2016.
143. Platonova, E., et al., *Single-molecule microscopy of molecules tagged with GFP or RFP derivatives in mammalian cells using nanobody binders*. Methods, 2015. **88**: p. 89-97.
144. Szymborska, A., et al., *Nuclear Pore Scaffold Structure Analyzed by Super-Resolution Microscopy and Particle Averaging*. Science, 2013. **341**(6146): p. 655-658.

7 References

145. Loschberger, A., et al., *Correlative super-resolution fluorescence and electron microscopy of the nuclear pore complex with molecular resolution*. Journal of Cell Science, 2014. **127**(20): p. 4351-4355.
146. Sander, J., et al., *Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications*. Data Mining and Knowledge Discovery, 1998. **2**(2): p. 169-194.
147. Schmid, M., et al., *Nup-PI: the nucleopore-promoter interaction of genes in yeast*. Mol Cell, 2006. **21**(3): p. 379-91.
148. Liang, Y., et al., *Dynamic Association of NUP98 with the Human Genome*. Plos Genetics, 2013. **9**(2).
149. Pleiner, T., et al., *Nanobodies: site-specific labeling for super-resolution imaging, rapid epitope-mapping and native protein complex isolation*. Elife, 2015. **4**.
150. Li, J., S. Jia, and P.R. Chen, *Diels-Alder reaction-triggered bioorthogonal protein decaging in living cells*. Nature Chemical Biology, 2014. **10**(12): p. 1003-+.
151. Murrey, H.E., et al., *Systematic Evaluation of Bioorthogonal Reactions in Live Cells with Clickable HaloTag Ligands: Implications for Intracellular Imaging*. J Am Chem Soc, 2015. **137**(35): p. 11461-75.
152. Darko, A., et al., *Conformationally strained trans-cyclooctene with improved stability and excellent reactivity in tetrazine ligation*. Chemical Science, 2014. **5**(10): p. 3770-3776.
153. Abmayr, S.M., et al., *Preparation of nuclear and cytoplasmic extracts from mammalian cells*. Curr Protoc Pharmacol, 2006. **Chapter 12**: p. Unit12 3.
154. Lukinavicius, G., et al., *Fluorogenic probes for live-cell imaging of the cytoskeleton*. Nat Methods, 2014. **11**(7): p. 731-3.
155. Schnell, U., et al., *Immunolabeling artifacts and the need for live-cell imaging*. Nat Methods, 2012. **9**(2): p. 152-8.
156. Grimm, J.B., et al., *A general method to improve fluorophores for live-cell and single-molecule microscopy*. Nat Methods, 2015. **12**(3): p. 244-50, 3 p following 250.
157. Uno, S.N., et al., *A spontaneously blinking fluorophore based on intramolecular spirocyclization for live-cell super-resolution imaging*. Nat Chem, 2014. **6**(8): p. 681-9.
158. Joo, C., et al., *Advances in single-molecule fluorescence methods for molecular biology*. Annu Rev Biochem, 2008. **77**: p. 51-76.
159. Kisselev, L.L. and R.H. Buckingham, *Translational termination comes of age*. Trends Biochem Sci, 2000. **25**(11): p. 561-6.
160. Jackson, R.J., C.U. Hellen, and T.V. Pestova, *Termination and post-termination events in eukaryotic translation*. Adv Protein Chem Struct Biol, 2012. **86**: p. 45-93.
161. Schmied, W.H., et al., *Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1*. J Am Chem Soc, 2014. **136**(44): p. 15577-83.
162. Elsasser, S.J., et al., *Genetic code expansion in stable cell lines enables encoded chromatin modification*. Nat Methods, 2016. **13**(2): p. 158-64.
163. Lam, J.K., et al., *siRNA Versus miRNA as Therapeutics for Gene Silencing*. Mol Ther Nucleic Acids, 2015. **4**: p. e252.
164. Moore, C.B., et al., *Short hairpin RNA (shRNA): design, delivery, and assessment of gene knockdown*. Methods Mol Biol, 2010. **629**: p. 141-58.
165. Rao, D.D., et al., *siRNA vs. shRNA: similarities and differences*. Adv Drug Deliv Rev, 2009. **61**(9): p. 746-59.

7 References

166. Rabut, G., V. Doye, and J. Ellenberg, *Mapping the dynamic organization of the nuclear pore complex inside single living cells*. Nat Cell Biol, 2004. **6**(11): p. 1114-21.
167. Wang, Q. and L. Wang, *New methods enabling efficient incorporation of unnatural amino acids in yeast*. J Am Chem Soc, 2008. **130**(19): p. 6066-7.
168. Parrish, A.R., et al., *Expanding the genetic code of Caenorhabditis elegans using bacterial aminoacyl-tRNA synthetase/tRNA pairs*. ACS Chem Biol, 2012. **7**(7): p. 1292-302.
169. Lin, D.H., et al., *Architecture of the symmetric core of the nuclear pore*. Science, 2016. **352**(6283): p. aaf1015.
170. Miller, B.R., et al., *Identification of a new vertebrate nucleoporin, Nup188, with the use of a novel organelle trap assay*. Mol Biol Cell, 2000. **11**(10): p. 3381-96.
171. Hoelz, A., E.W. Debler, and G. Blobel, *The structure of the nuclear pore complex*. Annu Rev Biochem, 2011. **80**: p. 613-43.
172. Zhou, H.X., *Polymer models of protein stability, folding, and interactions*. Biochemistry, 2004. **43**(8): p. 2141-54.
173. Ren, Y., et al., *Structural and functional analysis of the interaction between the nucleoporin Nup98 and the mRNA export factor Rae1*. Proc Natl Acad Sci U S A, 2010. **107**(23): p. 10406-11.
174. Franks, T.M. and M.W. Hetzer, *The role of Nup98 in transcription regulation in healthy and diseased cells*. Trends in Cell Biology, 2013. **23**(3): p. 112-117.
175. Perdigo, N., et al., *Unexpected features of the dark proteome*. Proc Natl Acad Sci U S A, 2015. **112**(52): p. 15898-903.
176. Banterle, N., et al., *Fourier ring correlation as a resolution criterion for super-resolution microscopy*. Journal of Structural Biology, 2013. **183**(3): p. 363-367.