TU Delft

Delft University of Technology

Single-Chromosome Biophysics

Holub, M.

DOI
10.4233/uuid:08912541-02f4-44af-b8a6-c3b7a3647f17

Publication date
2024

Document Version
Final published version

Citation (APA)
Holub, M. (2024). *Single-Chromosome Biophysics*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:08912541-02f4-44af-b8a6-c3b7a3647f17

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

SINGLE-CHROMOSOME BIOPHYSICS

SINGLE-CHROMOSOME BIOPHYSICS

Dissertation

for the purpose of obtaining the degree of doctor

at Delft University of Technology,

by the authority of the Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen,

chair of the Board for Doctorates

to be defended publicly on

Wednesday 8 January 2025 at 17:00 o'clock

by

Martin HOLUB

MASTER OF SCIENCE IN MECHANICAL ENGINEERING
EIDGENÖSSISCHE TECHNISCHE HOCHSCHULE ZÜRICH, SWITZERLAND,

BORN IN ČESKÉ BUDĚJOVICE, CZECH REPUBLIC

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. C. Dekker	Delft University of Technology, promotor
Prof. dr. A. M. Dogterom	Delft University of Technology, promotor
Independent members:	
Dr. C. Broedersz	Vrije Universiteit Amsterdam
Prof. dr. R. T. Dame	Leiden University
Prof. dr. ir. P. A. S. Daran-Lapujade	Delft University of Technology
Dr. S. A. Hoffmann	Wageningen University & Research
Prof. dr. W. T. S. Huck	Radboud University
Prof. dr. ir. S. J. J. Brouns	Delft University of Technology, reserve

Keywords: single-molecule biophysics, chromosome organization, microfluidics, bottom-up biology, SMC proteins, biosecurity, synthetic genomics

Printed by: Proefschriftspecialist.nl

Front & Back: Martin Holub

Copyright © 2024 by Martin Holub

Casimir PhD Series, Delft-Leiden 2023-27

ISBN/EAN: 978-94-93391-90-1

An electronic version of this dissertation is available at
`http://repository.tudelft.nl/`.

CONTENTS

SUMMARY

This doctoral thesis stands on three pillars that emerged from following my scientific interests: i) biophysics, ii) synthetic biology, and iii) biosecurity. The former, biophysics, reflects my desire for deep understanding of biological systems and my affinity for experimental work. The latter, synthetic biology and biosecurity, were born from the conviction that the understanding obtained in pursuing science offers the most fulfillment when applied to the benefits of society.

The main part of this thesis explores methodologies for the extraction, and characterization of large-scale DNA, with a particular focus on the megabase-pair length DNA from bacterial sources. The research aims to bridge the gap between in vivo chromosome studies and in vitro single-molecule techniques by developing approaches that enable the investigation of chromosome structure and dynamics at a more relevant genomic scale. The following chapters detail the experimental approaches, results, and conclusions drawn from this work.

After having introduced the major concepts in this thesis in **Chapter 1**, we start **Chapter 2** by addressing the limitations of current single-molecule DNA studies, which rely on DNA substrates significantly shorter than those found in living cells. This chapter presents a method for obtaining deproteinated DNA of megabase-pair length for in vitro experiments. The procedure involves isolating chromosomes from bacterial cells and enzymatically removing native proteins. The degree of removal is confirmed by mass spectrometry. The resulting DNA polymers are analyzed using fluorescence microscopy, revealing an increase in their radius of gyration while maintaining their megabase-pair length. The practical applications of this method are demonstrated through proof-of-concept experiments, including experiments with Fis and PEG, as well as tracking the motion of fluorescently labeled DNA loci.

Chapter 3 introduces a microfluidic platform designed to overcome the challenges associated with handling fragile megabase-scale DNA molecules. The microfluidic device enables the sequential extraction, purification, and analysis of bacterial nucleoids directly within individual quasi-2D microchambers. This avoids the fragmentation of these large DNA molecules, while allowing for their extended observation in highly controlled conditions. Using the platform, we successfully extract the chromosomal DNA from *E. coli* and *B. subtilis* cells. Additionally, we demonstrate the capability of the platform to introduce proteins to the trapped DNA purely through diffusion. This integrated microfluidic approach represents an important step towards bottom-up assembly of complex biomolecular systems, such as artificial chromosomes.

The **Chapter 4** opens with a broad introduction to the polymer physics of DNA, and

overview of previous research on isolating and characterizing individual chromosomes, with focus on microfluidic approaches. In the experimental part, we build on the methodology established in Chapter 3 by applying the microfluidic platform for detailed biophysical studies of chromosomes isolated in microfluidic traps. We pursue both the structural and dynamic characterization of DNA at the megabase scale, and develop extensive set of image analysis tools to do so. The results demonstrate the ability to trap individual chromosomes for periods exceeding 60 minutes, while following their structure and dynamics at range of spatial and temporal scales. This work represents a detailed characterization of unperturbed chromosomes in microfluidic traps, which paves way for wide range of studies employing the genome-in-a-box methodology.

Chapter 5 shifts the focus to the role of Structural Maintenance of Chromosomes (SMC) proteins, particularly condensin, in DNA compaction and organization at the megabase scale. We conduct controlled microfluidic experiments with yeast condensin and observe that the protein facilitates an ATP-dependent compaction of megabase scale DNA. This is corroborated by simultaneous observation of condensin recruitment to the DNA, and local DNA intensity increase. However, and unfortunately, challenges such as protein-mediated surface interactions, which caused local DNA sticking to the walls of microfluidic device, which severely complicated the analysis. To address these challenges, novel image analysis metrics are developed to quantify the compaction despite the aggregation. This chapter establishes an analytical workflow including comparison with polymer modelling and identifies relevant experimental conditions for future studies.

Chapter 6, is the outcome of 3-month EMBO Scientific Exchange Grant stay at Manchester Institute of Biotechnology, and borne out of the interest to bridge our whole-chromosome experiments to applications, as well as to enter the field of synthetic genomics. The chapter opens up with a general introduction into the field, and a discussion of approaches for isolating synthetic chromosomes, both for use in different organisms, and in-vitro. Next, I propose and prototype an immunoprecipitation-based strategy for isolating synthetic yeast chromosomes. The method is demonstrated on a 190 kbp synthetic chromosome featuring a tandem repeat array of *tetO* binding sites that recruit Tet repressor proteins (TetR). The results, assessed using pulsed-field gel electrophoresis (PFGE), show an enrichment factor of 6- to 15-fold for these synthetic chromosomes. This method shows promise for further development, particularly by testing efficacy with larger chromosomes and varied protein recruitment site arrangements.

Chapters 7 and **8** present results of independent research in the field of biosecurity. The work follows my interest in synthetic biology, and in the advantages and risks it brings to the society as it rapidly develops and becomes more powerful as well as approachable for broad set of actors.

Chapter 7 explores the potential of biofoundries—highly automated facilities designed for processing biological samples—to accelerate innovation in disease surveillance and environmental monitoring. Biofoundries typically support engineering biology by implementing design, build, test, and learn (DBTL) cycles, and by fostering collaboration

between public and private stakeholders. This chapter argues for expanding the scope of biofoundries to include roles in biosurveillance and biosecurity. Through a literature review, we identified ways biofoundries could contribute to these areas, such as by developing measurement standards and protocols, engaging citizens in data collection, collaborating closely with biorefineries, and processing samples for biosecurity purposes. The chapter concludes with a discussion of potential roadblocks to these applications and offers recommendations for policymakers and stakeholders interested in enhancing biosecurity programs through the integration of biofoundries.

Finally, **Chapter 8** explores how a recent technological revolution, large language models (LLMs), can be leveraged to enhance biosecurity efforts. Based on interviews with biosecurity experts, we examine how LLMs could support various biosecurity-related tasks, such as information gathering, report generation, and communication. The findings suggest that approximately 50% of these tasks have a high potential for automation through LLMs. We conclude by arguing for the development of LLM-based tools tailored to the specific needs of biosecurity professionals, and suggest field-specific datasets to help to do so.

SAMENVATTING

Dit proefschrift staat op drie pijlers die voortkomen uit mijn wetenschappelijke interesses: i) biofysica, ii) synthetische biologie en iii) bioveiligheid. De eerste, biofysica, weerspiegelt mijn zoektocht naar inzicht in de werking van biologische systemen. De laatste twee, synthetische biologie en bioveiligheid, zijn geboren uit de overtuiging dat wetenschappelijke kennis de meeste voldoening biedt wanneer het wordt toegepast ten voordele van de maatschappij.

Het grootste deel van dit proefschrift onderzoekt methodes voor de extractie en karakterisering van lange DNA-moleculen, met een specifieke focus op het chromosomale DNA uit bacteriën. Het onderzoek is gericht op het overbruggen van de kloof tussen in vivo chromosoomstudies en in vitro enkele-molecuultechnieken. Dit doen we door een in vitro experimentele methode te ontwikkelen die het mogelijk maakt om de structuur en dynamica van DNA op de relevantere schaal van een compleet genoom (in plaats van het gebruikelijke korte DNA) te onderzoeken. De volgende hoofdstukken beschrijven de experimentele benaderingen, resultaten en conclusies die uit dit werk zijn getrokken.

Nadat Hoofdstuk 1 de belangrijkste concepten in dit proefschrift heeft geïntroduceerd, bespreekt **Hoofdstuk 2** de beperkingen van huidige enkele-molecuul-DNA-studies, die DNA-substraten gebruiken die aanzienlijk korter zijn dan die in levende cellen. Dit hoofdstuk presenteert een methode voor het verkrijgen van DNA van megabasepaarlengte, ontdaan van alle DNA-bindende eiwitten. De procedure omvat het isoleren van chromosomen uit bacteriële cellen en het enzymatisch verwijderen van de DNA-bindende eiwitten. De mate van eiwitverwijdering wordt bevestigd door massaspectrometrie. De resulterende DNA-moleculen worden geanalyseerd met behulp van fluorescentiemicroscopie, waarbij een toename in hun ruimtelijke grootte wordt getoond terwijl hun genomische lengte (in aantal DNA baseparen) onveranderd blijft. De toepassingen van deze methode worden gedemonstreerd door middel van enkele proeven, waaronder experimenten met Fis en PEG, evenals het volgen van de beweging van fluorescent gelabelde DNA-locaties.

Hoofdstuk 3 introduceert een microfluïdisch platform dat is ontworpen om de problemen oplossen die gepaard gaan met het hanteren van fragiele DNA-moleculen van enkele miljoenen baseparen. Het microfluïdische apparaat maakt de afzonderlijke stappen van extractie, zuivering en analyse van bacteriële chromosomen mogelijk, direct in individuele quasi-2-dimensionale microscopische kamertjes. Dit voorkomt de fragmentatie van de grote DNA-moleculen, terwijl ze uitgebreid kunnen worden geobserveerd in zeer gecontroleerde omstandigheden. Met behulp van het platform extraheren we met succes het chromosomale DNA uit *E. coli* en *B. subtilis* cellen. Daarnaast tonen we aan dat het platform in staat is om gepurificeerde eiwitten aan het geïsoleerde DNA toe te

voegen, en de daaropvolgende DNA-compactie te observeren. Deze geïntegreerde microfluïdische benadering vormt een belangrijke stap in de richting van bottom-up assemblage van complexe biomoleculaire systemen, zoals kunstmatige chromosomen.

Hoofdstuk 4 begint met een bredere introductie tot de polymeerfysica van DNA en een overzicht van eerder onderzoek naar het isoleren en karakteriseren van individuele chromosomen, met de focus op microfluïdische benaderingen. In het experimentele deel bouwen we voort op de methodologie uit hoofdstuk 3. We passen het microfluïdische platform toe voor gedetailleerde biofysische studies van chromosomen die zijn geïsoleerd in de microfluïdische kamertjes. We karakteriseren zowel de structuur als de dynamica van het DNA op megabasepaar-schaal en ontwikkelen een uitgebreide waaier aan beeldanalysetools om dit te bereiken. De resultaten tonen dat we in staat zijn om individuele chromosomen te volgen voor langer dan 60 minuten, terwijl hun structuur en dynamica in beeld brengen op verschillende schalen in zowel ruimte als tijd. Dit werk omvat een gedetailleerde karakterisering van intacte chromosomen in microfluïdische kamertjes, en maakt de weg vrij voor een breed scala aan studies met behulp van de Genome-in-a-Box-methodologie.

Hoofdstuk 5 verschuift de focus naar de rol van Structural Maintenance of Chromosomes (SMC)-eiwitten, met name condensin, in DNA-compactie en -organisatie op megabasepaar-schaal. We voeren gecontroleerde microfluïdische experimenten uit met gistcondensin en observeren dat het eiwit een ATP-afhankelijke compactie van DNA op megabase-schaal faciliteert. Dit wordt bevestigd door gelijktijdige observatie van toenemende co-lokalisatie van condensin met het DNA en een lokale DNA-intensiteitstoename. Helaas zijn er nog grote experimentele uitdagingen voor dit project. Zo plakte het DNA via de DNA-bindende eiwitten op verschillende plekken aan de wanden van de microfluïdische kamertjes, wat de analyse van het experiment compliceerde. Om deze uitdagingen aan te pakken, worden nieuwe beeldanalysemethodes ontwikkeld om de DNA-compactie te kwantificeren ondanks de problemen met plakkend DNA. Dit hoofdstuk stelt een analytische workflow vast, inclusief vergelijking met polymeermodellering, en identificeert relevante experimentele omstandigheden voor toekomstige studies.

Hoofdstuk 6 is het resultaat van een drie maanden durend verblijf met een EMBO Scientific Exchange Grant aan het Manchester Institute of Biotechnology. De motivatie voor deze onderzoeksstage was om experimenten aan hele chromosomen te verbinden met praktische toepassingen, en om het veld van synthetische genomica te betreden. Het hoofdstuk begint met een algemene introductie in het veld en een bespreking van benaderingen om synthetische chromosomen te isoleren, zowel voor gebruik in verschillende organismen als in vitro. Vervolgens stellen we een nieuwe, op immunoprecipitatie gebaseerde strategie voor om synthetische gistchromosomen te isoleren. De methode wordt gedemonstreerd op een synthetisch chromosoom van 190 kbp waarop een cluster van 224 tetO-bindingsplaatsen is aangebracht. Hieraan binden Tet-repressorproteïnen (TetR) en op deze manier wordt het synthetische chromosoom geïsoleerd uit het monster. De resultaten, beoordeeld met behulp van gepulst-veld-gel-elektroforese (PFGE), laten een verrijkingsfactor van 6 tot 15 keer zien voor deze synthetische chromosomen.

Deze methode lijkt veelbelovend voor verdere ontwikkeling, met name door de efficiëntie te testen op grotere chromosomen en andere locaties te kiezen voor de cluster van tetO-bindingsplaatsen.

Hoofdstukken 7 en **8** presenteren resultaten van een onafhankelijk onderzoek op het gebied van bioveiligheid. Het werk komt voort uit mijn persoonlijke interesse in de synthetische biologie. Hierbij is het mijns inziens vooral belangrijk om de voordelen en risico's te analyseren die de synthetische biologie met zich meebrengt voor de maatschappij. Het veld ontwikkelt zich namelijk snel en wordt breder toepasbaar en toegankelijker voor een grotere groep actoren.

Hoofdstuk 7 onderzoekt het potentieel van biofoundries (geautomatiseerde faciliteiten die zijn ontworpen voor het verwerken van biologische monsters) om innovatie in ziekte- en milieu-monitoring te versnellen. Biofoundries ondersteunen doorgaans de technische biologie door de design-, build-, test- end learn- (DBTL) cyclus te implementeren en door samenwerking tussen publieke en private belanghebbenden te bevorderen. Dit hoofdstuk pleit om het aandachtsgebied van biofoundries uit te breiden, zodat ze een rol kunnen gaan spelen in de biosurveillance en bioveiligheid. Door middel van een literatuuronderzoek en interviews hebben we manieren geïdentificeerd waarop biofoundries hieraan kunnen bijdragen. Dit zou kunnen door meetnormen en -protocollen te ontwikkelen, burgers te betrekken bij het verzamelen van gegevens, nauw samen te werken met bioraffinaderijen en monsters te verwerken voor bioveiligheidsdoeleinden. Het hoofdstuk wordt afgesloten met een bespreking van mogelijke obstakels voor deze toepassingen en biedt aanbevelingen voor beleidsmakers en belanghebbenden die geïnteresseerd zijn in het verbeteren van bioveiligheidsprogramma's door de integratie van biofoundries.

Tot slot onderzoekt **Hoofdstuk 8** hoe een recente technologische revolutie in de kunstmatige intelligentie, namelijk Large Language Models (LLM's, ofwel grote taalmodellen), kan worden benut om bioveiligheid te verbeteren. Op basis van interviews met bioveiligheidsexperts onderzoeken we hoe LLM's verschillende bioveiligheid-gerelateerde taken kunnen ondersteunen, zoals informatieverzameling, rapportgeneratie en communicatie. De bevindingen suggereren dat ongeveer 50% van deze taken een hoog potentieel heeft voor automatisering via LLM's. We sluiten af door te pleiten voor de ontwikkeling van LLM-gebaseerde toepassingen die zijn afgestemd op de specifieke behoeften van bioveiligheidsprofessionals, en stellen enkele veldspecifieke datasets voor om hierbij te helpen.

1

CHROMOSOME ORGANIZATION: A PUZZLE WITH MANY PIECES

Information stored in DNA is critical for cellular fate and function, and nature has evolved many intricate ways how to organize it in space and time. The physical forces of confinement and crowding act globally on chromosomes, and contribute to their compaction and positioning inside the cell and nucleus. Structural maintenance of chromosome proteins interact with other DNA-bound proteins to confer chromosomes with dynamic structure across scales. Transcription itself is an important contributor to chromosome organization and compaction, injecting torsional strain into the DNA molecule, which leads to looping-out of plectonemes and compaction. Phase separation is a general phenomenon that refers de-mixing of DNA based on its physicochemical properties, which are themselves influenced by proteins that associate with it. Finally, on the finest scale, it is diverse range of proteins that associate with DNA to bend, bridge and coat, locally modifying its properties and contributing the final major piece to the chromosome organization puzzle. All together, these effectors not only compact the chromosomes, but also structure them to domains across range of spatial and temporal scales. These domains provide a scaffold on which the regulation of gene expression and chromosome function across different environments, cell states, and types, takes place. In this chapter, we review the current state-of-the-art in understanding the major organizing principles of genomes and highlight open questions that motivated this thesis.

1

1.1. THE DNA AS A POLYMER AND THE ROLE OF CONFINEMENT IN CHROMOSOME ARCHITECTURE

The number of genes any given cell expresses varies by about three orders of magnitudes across domains of life but for most model organisms, it is of the order of 10'000. The size of genomes, i.e. the total number of DNA bases that the cells store their genetic information in, is more variable. In fact, organisms with comparable amount of encoded genes can have genome sizes differing by three orders of magnitude.[1] Despite this variety, all cells face the same fundamental challenge: to compact their genomes into the confined volume of a cell. This is a daunting task, imposing a requirement for human cells to compact about 1 m of DNA to 10 μm sized nucleus (factor of ~100'000), for yeast cells to compact about 3.6 mm of DNA to a nucleus that is 1.8 μm across (factor of ~2'000) and bacteria to compact about 1.4 mm of DNA to their bodies which are just about 1 μm in diameter (factor of ~1'000). This requirement becomes even more strict at the time of DNA replication and cell division when the genome size is doubled or when the genome does not occupy whole cell (or nucleus) volume. Over the last decade, thanks to development of enabling techniques for chromosome scale studies of DNA conformation, polymer physics computational models and single molecule studies, our understanding of genome architecture in time and space has progressed rapidly. Yet, the descriptions are often only phenomenological and the understanding of diverse phenomena that have on global genome architecture is incomplete.

DNA molecules in cells consist of two antiparallel polymer chains that wrap around each other, stabilized by hydrogen bonds formed between their individual bases. The "double helix" has one helix repeat per 10.5 pairs, with a distance of about 0.34 nm between base pairs.[4] At physiological conditions, the DNA has linear charge of density of about $2e^-$ per base pair and the electrostatic repulsion between DNA strands makes it adopt a tube-like structure with diameter of about 2.3 nm. Bending DNA double helix requires energy as a stiff polymer with a persistence length (L_p, about 50 nm or 150 bp) bends against the thermal motion ($k_B T$).[5] The DNA therefore behaves as a flexible polymer at length scales higher than hundreds of nanometers where it is subject to thermal fluctuations. The bases along DNA are frequently stacked in such a way that they generate curvature, which creates recognition sites for proteins, e.g. histones in eukaryotes[6] or nucleoid-associated-proteins (NAPs) in bacteria,[7] further structuring the DNA molecule.

The contour length of a typical bacterial genome is much higher than the DNA persistence length (by about a factor of 30'000) and the chromosome has an intrinsic tendency to adopt a conformation that maximizes available degrees of freedom for its individual segments. This drives spontaneous condensation into a globule of size $2R \sim L_p(N)^{1/2}$ which for N=30'000 is about 10 μm [3,8] and yields ~100-fold compaction (Fig. 1.1B). Thus, the arguments of polymer physics alone can already explain a significant fraction of the compaction of DNA in bacteria.

Inside cells, DNA is in a poor solvent and prefers to interact with itself. Additional level of compaction is therefore conferred through non-specific volume-exclusion effect that

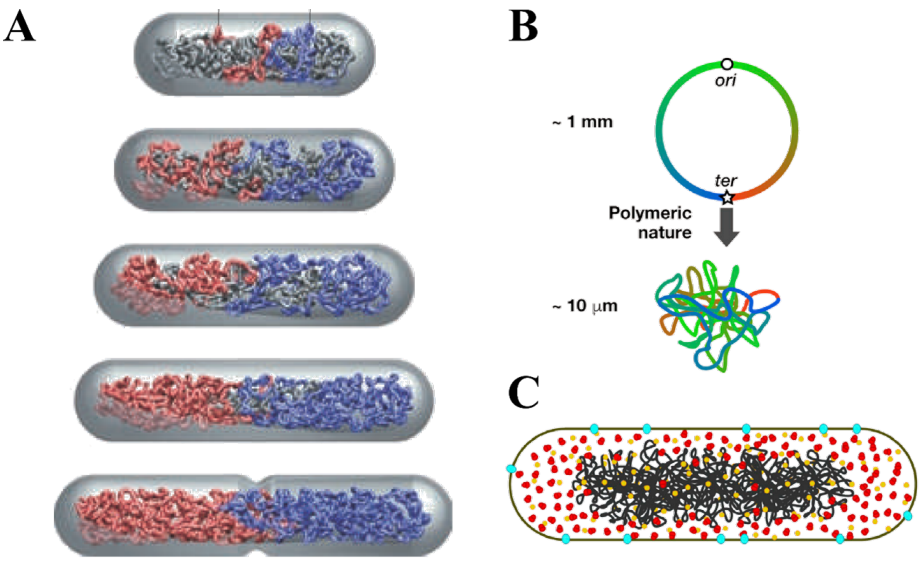

Figure 1.1: Role of polymer physics in organization of bacterial genome. A) Segregation of sister chromosomes may be driven by entropic effects due to demixing of two ring-shaped compacted polymers in a confinement. B) Chromosome-scale DNA behaves as a charged flexible polymer and maximizes its available degrees of freedom, yielding ~100-fold compaction. C) The macromolecular crowding, driven mainly by active (red) and inactive (yellow) ribosomes, leads to excluded volume force that compacts the nucleoid so as to occupy only fraction of cell volume. Most ribosomes are outside of nucleoid and mRNA is degraded at the circumference (blue) which may further promoter localization of transcriptional active sites at the outsides of the nucleoid. However, not all bacteria localize mRNA degradation to periphery. A) Adapted from.[2] B,C) Adapted from.[3]

arises as a direct consequence of macromolecular crowding in confined volume of a cell (Fig. 1.1C).[9,10] Although this effect leads to collapse of large polymer loops and decreases the conformational freedom of the polymer, this is thermodynamically compensated by the gain of accessible volume of the cytoplasmic crowding molecules.

The major crowders are ribosomes, large in size and abundant in negative charge,[11] they are electrostatically repelled from the DNA of the nucleoid. There are around 45'000 ribosomes in a cell, and vast majority of them, 85 – 90% locate outside of the nucleoid volume (which for *E. coli* takes up about 25% the total cell volume),[12] creating an excluded volume force that compacts the chromosome.

1.2. BACTERIAL CHROMOSOME ORGANIZATION BY DNA-ASSOCIATED PROTEINS

Proteins that interact with DNA are major contributors to DNA compaction. These are either histones in eukaryotes, or nucleoid associated proteins (NAPs) in bacteria. Interestingly, archaea evolved to include members of both classes.[13] Histones are the strongest compactors, and it is likely that bacteria did not need to develop them, as their

chromosomes require a ~100-fold lower compaction. Additionally, the bacterial genome is dense in coding sequences which leads to requirement for its high DNA sequence accessibility.

Bacterial chromosomes are folded to a range of conformations by DNA-binding proteins (Fig. 1.2A). The regulation of effect of individual NAPs arises mainly by changes of their intracellular concentration and additional level of control is conferred through post-translational modifications (PTMs) and conformational changes in response to ligands. In this section, we focus our description of the major nucleoid associated proteins, which were the most relevant in our work with bacterial genomes. In addition to Dps mentioned earlier, here we briefly describe H-NS, HU, Fis and IHF.

The histone-like nucleoid structuring protein (H-NS) is a small polypeptide (137 AA in *E. coli*) that binds preferentially to AT rich sequences. It has the ability to chain and forms nucleoprotein filaments that have the ability to bridge two DNA helices, forming a loop-like structure that contributes to DNA compaction and transcriptional regulation. The formation of such bridges by H-NS prevents transcription from coated genes. On the other hand RNAP can displace a DNA-bound H-NS filament. The detailed understanding of details of interaction of RNAP with H-NS filaments continues to be a matter of open research.

HU protein locally induces DNA bending by small out-of-plane angles and preferentially binds at kinked regions and regions with curvature but otherwise shows no sequence specificity. HU is known to promote negative supercoiling and its DNA association patterns drive the uneven supercoiling density that is observed to increase from *ori* to *ter* in starved cells[14] which correlates with the observation that HU expression levels vary for different growth phases.[15] DNA was observed to wrap around HU proteins, making them take up a similar function as histones in eukaryotes. However, unlike histones, the interactions of HU with DNA are weak, transient and electrostatically driven (HU is positively charged). Rather than statically conferring the chromosome with certain architecture, they are likely to facilitate the dynamic structure of bacterial nucleoid that is needed for rapid reorganization during growth and segregation (Fig. 1.2B).[16] Moreover, despite their diffusive nature, HU proteins have been shown to localize preferentially with nucleoid which is suggestive of movement in liquid phase separated macro-compartment of the nucleoid.

Recent studies showed that deletion of small bacterial noncoding RNA (ncRNA4) that mediates HU-DNA interactions restructured the nucleoid and altered CID patterns.[17] Interestingly, multiple chromatin binding proteins including StpA, H-NS and Hfq also possess the ability to bind RNA suggesting an important, yet to date neglected, role of RNA-protein and RNA-DNA interactions in chromatin architecture.[12]

The integration host factor (IHF) shares 40% sequence identity with HU, but unlike HU it bends DNA severely, by about 160º, and has been identified only in Gram-negative bacteria (i.e. including *E. coli*). Although IHF binds specifically, its large intracellular concentration allows it to bind nonspecifically across the genome as well. It associates primarily with promoters and can both activate or repress transcription, possibly by sharp DNA bending.

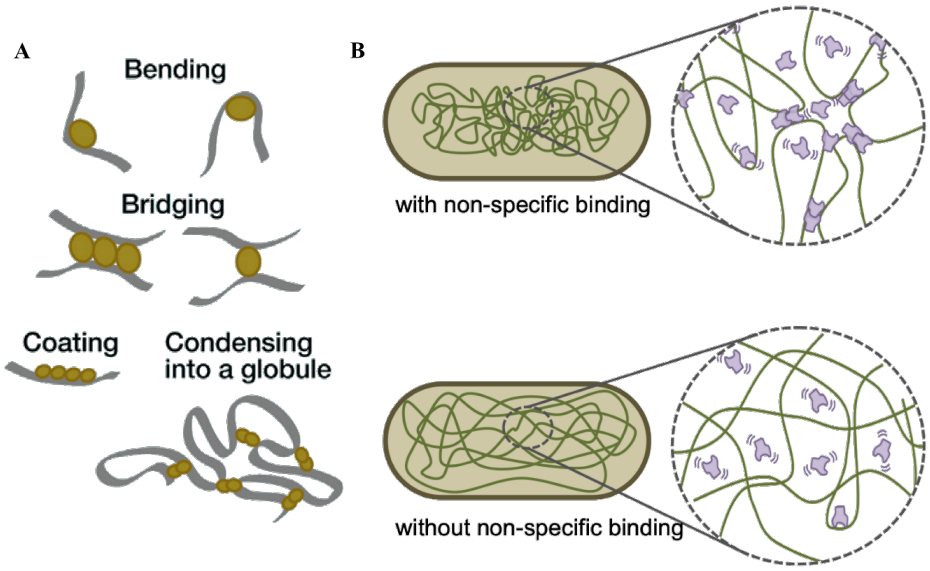

Figure 1.2: Various architectural roles of DNA-binding proteins. A) DNA-binding proteins bend (e.g. HU, IHF, FIS), bridge (e.g. H-NS, StpA, SMCs), coat (e.g. H-NS, HU), or condense (e.g. Dps) chromatin. B) Transient weak interactions of HU with nucleoid were suggested to confer it with increased flexibility while facilitating compaction and distant DNA-DNA contacts (top). A) Adapted from[3], B) from[16].

The Fis (factor for inversion stimulation) protein dimerizes and bends the DNA by 50 – 90°. It associates with sequence-formed groove on DNA and is often found at locations of crossover of DNA helices, likely stabilizing supercoiled state and plectonemes. Fis is the most abundant protein in growth phase in *E. coli* (whereas it is entirely absent in *B. subtilis*) and activates expression of genes that encode products necessary for rapid cell division by binding to RNAP while repressing genes for utilization of alternative carbon sources. It also represses gyrase and therefore indirectly decreases negative supercoiling[12] which in turn promotes DNA melting and facilitates transcription, which is itself an evidence of the ability of supercoiling to serve as gene expression regulator.

1.3. INTERPLAY OF TRANSCRIPTION AND GENOME ORGANIZATION

An important part of chromosome compaction arises as a consequence of supercoiled nature of DNA. This mechanism is thought to be the most important in prokaryotes, who mostly lack histones.[3] DNA is a helix with pitch of about 10 bp per twist. Addition or removal of twists imposes a torsional strain on the molecule. In majority of cases this occurs between two 'fixed' points (e.g. two boundaries of supercoiled domain (SD), with average size of ~15 kb, that are stabilized by DNA-binding proteins) and the strain is partially released by folding of DNA into superhelices (plectonemes) (Fig. 1.3A), which

leads to DNA compaction. Transcription-induced supercoiling domains have recently been described as the major organizing elements of bacterial chromosome.[18]

Cells need to control DNA accessibility and thus chromosomal compaction, and various proteins can increase or decrease the torsional strain. Increase arises e.g. from RNA polymerases unwinding the DNA double helix or gyrase, while the torsional strain can be relaxed by topoisomerases that cut and reseal the double helix allowing it to freely twist.

Transcription is likely to play an important role in segregation and positioning of the bacterial chromosome as well. Due to the absence of physical boundary separating transcribed DNA from protein synthesis, both can happen simultaneously in prokaryotic cells in a process of coupled transcription-translation. Such coupling drags DNA sequences coding for cell-wall associated proteins towards the cell periphery. If the proteins are at the same time inserted into the membrane (coupled transcription-transertion), this together with the longitudinal expansion of cell wall could lead to force pulling on the nucleoid and promoting separation of daughter chromosomes. The effect could be further enhanced by competition for free transertion sites that become occupied directionally, from mid-cell to poles.[19] Nevertheless, more recent results have shown persistent chromosome segregation also in cells where transcription was arrested, questioning the transertion hypothesis.[20]

In prokaryotes, chromosomal interaction domains are demarked by locations of highly expressed genes (HEGs) and were visualized with Hi-C in *C. crescentus*[21] and *B. subtilis* (Fig. 1.3A).[22] Their boundaries could be repositioned by repositioning the HEGs[21] and longer transcribed regions with higher occupancy of elongation complexes (EC) seem to form stronger boundaries.[23] Association of RNAP with DNA and transcription change supercoiled state of the DNA in the vicinity and this likely serves as local gene expression regulator. It was shown that gene expression correlates within domain size of ~10 kb and this agrees well with the size of SDs, boundaries of which constrain supercoil diffusion (Fig. 1.3A, B).[24]

The level of organization based on transcriptional state on fine grain level (compartmental domains) was proposed to be general for all species.[25] In eukaryotes, fewer than 1/3 of compartmental domains are flanked by CTCF sites and the major organizing principle is likely associated with transcription. For example, placing an enhancer and promoter element at distal locations in *Drosophila,* a species that undergoes embryogenesis without CTCF, was enough to promote looping and although not necessary, activation of transcription further enhanced the stability of the loops.[26]

1.4. GENOME ORGANIZATION BY STRUCTURAL MAINTENANCE OF CHROMOSOME PROTEINS

Structural maintenance of chromosome (SMC) protein complexes are molecular motors that actively bring distant regions of chromosomes into closer spatial proximity. An

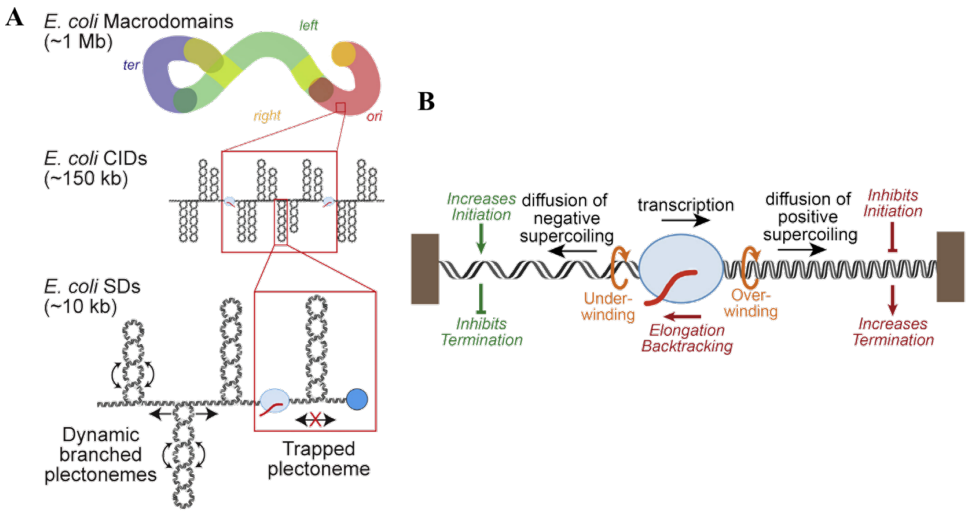

Figure 1.3: Bacterial chromosome organization and the role of transcription. A) Bacterial chromosome is helically twisted to fit into the volume of the cell and organized in 4 macrodomains (6 if unstructured parts of left and right are considered separately) on the scale of ~1Mb. The chromosome arms do not interact in *E.coli*, whereas interaction is common in other species. Macrodomains are further divided into CIDs, and SDs. CIDs are often stabilized by HEGs and SDs form as a consequence of transcription and/or DNA-protein binding (blue). B) Detail of torsional stress induced by transcribing RNAP. Positive supercoiling favors dissociation of elongation complex (EC), whereas DNA unwinding due to negative supercoiling favors DNA melting and transcription initiation, indicating influence of supercoiling state on gene regulation A, B).[12]

SMC complex consists of two SMC coiled-coil proteins that are connected at the hinge. In prokaryotes, the complex is a homodimer, whereas eukaryotic complexes are heterodimeric. In both cases the sites of individual SMCs opposite to the hinge terminate with ATPase heads. To complete the ring-structure of SMC complex, the ATPase heads are connected by a kleisin subunit. SMC protein complexes contribute to chromosome architecture by range of functions, including extruding loops of DNA, sister-chromatid cohesion in eukaryotes[27,28] (Fig 1.5E), and zipping chromosome arms in some prokaryotes (Fig 1.4A,B).[29,30]

1.4.1. SMCS IN PROKARYOTES

Three classes of SMCs have been so far identified in bacterial cells. SMC-ScpAB from *B. subtilis* and *C. crescentus* is loaded on the chromosome at *parS* sequences, recruited by the ParB protein that itself binds to *parS* site[31,32] and is possibly aided by R-loop that forms upstream of transcribing RNAP.[33] The SMC-ScpAB complex progresses from the *parS* sites near the origin of replication towards *ter* region and progressively zips together the chromosomal arms in cell cycle dependent manner (Fig. 1.4A, B).[30,34] The progression rate of SMC-ScpAB is decreased by transcription and oppositely oriented highly expressed genes (HEGs) are often found at the stem of the loop.[29,30]

Further, two SMC-like proteins were identified: MukBEF in γ-proteobacteria (e.g. *E. coli*) and δ-proteobacteria (Fig. 1.4C) and MksBEF that is present in wide range of bacterial species. The SMC-like complexes, which form dimers, are thought to play a role in chromosome decatenation and segregation, possibly by correctly positioning replication origins and interaction with topoisomerase IV.[35,36] No loading factors are known to date. While there is no single-molecule level evidence of ability of SMC-like complexes to extrude genomic loops, the chromosome organization patterns suggest this phenomenon plays a role. For example, the MukBEF SMC has been implied in maintaining an axial core organization of *E. coli* chromosome, which supports loops of several tens of kbps.[37] Loop formation is consistent with the organization of *B. subtilis* chromosome as well.[30]

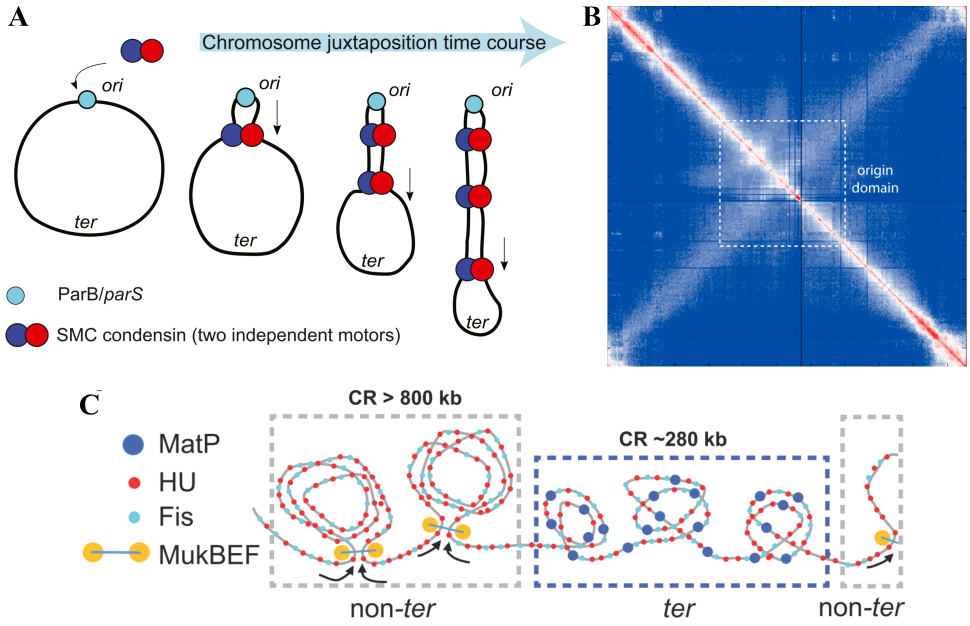

Figure 1.4: SMC-ScpAB is a loop extruding factor that aligns chromosome arms. A) SMC-ScpAB loads on ParB-bound *parS* sites end extrudes loops of DNA. Cartoon visualization of progressive chromosome arm zipping by multiple bound SMCs in *B.subtilis*. Adapted from [38]. B) A chromatin conformation capture map of contact frequencies along B/ suntilis chromosome. The diagonal running from bottom-left to top-right is created by the juxtaposition of chromosome arms by the SMC. Adapted from [22]. C) MukBEF is a dimeric proten that strucures the E. coli genome. It binds uniformly, with exception of *ter* region, where it is excluded by MatP. Outside of *ter* MukBEF promotes long range interactions (CR, contact range). Adapted from[39].

1.4.2. ROLE OF LOOP EXTRUSION IN GENOME ORGANIZATION

Continuous advancements in genomic methods for mapping spatial proximity of DNA in ensembles of cells (Hi-C) revealed that chromosomes are hierarchically structured.[40] That is, chromosomes contain nested regions where sequences preferentially interact with other sequences in that region and remain insulated from sequences outside of the region (Fig. 1.5A, B). Loop extrusion by SMCs is a thought to be the major contributor

to such hierarchical organization, and this organization has been observed both in eukaryotes, where the regions were termed topologically associated domains (TADs), as well as prokaryotes, where they were termed chromosome interaction domains (CIDs). The size of TADs is in the 0.2 - 1.0 Mb range,[41] whereas that of CIDs is about an order of magnitude smaller.[21,22] Improvements in resolution of Hi-C, bringing it down to 1-5 kb,[42] enabled experiments that suggest that chromosomes are organized even at finer scales.

Sites of active promoters have been associated with boundaries between these domains in both prokaryotes and eukaryotes[43] (Fig. 1.5D). In eukaryotes, where loop extrusion seems to play more important role, the borders between neighboring TADs are additionally commonly demarked by either convergent CTCF-binding sites or chromatin modifications (Fig. 1.5C). Although CTCF loops are dubbed after the CTCF transcriptional repressor that recognizes CCCTC motif,[44] which when encountered in convergent pair present the strongest and most common boundary to cohesin-loop enlargement, they seem to be transiently stabilized also by large protein complexes bound to transcribed regions of DNA (e.g. Mediator complex that binds to promoters[45]) or histone modifications. CTCF-loop-like chromosome structure is observed also in eukaryotic species that do not have CTCF orthologue (e.g. fission yeast *S. pombe* [46,47]) and where Hi-C patterns were predicted from chromatin transcriptional state alone.

Definition of TADs is based on Hi-C data and as such is subject to change of its interpretation. The current level of understanding suggests that TADs are much more dynamic, constituted by transient, low frequency, and weak interactions that are enhanced only about 2-3 fold in comparison to interactions outside of the domain.[50] Their structure, i.e. size and location of boundaries, likely arises as combination of processes including interactions of regulatory and genetic elements, contacts mediated by SMCs and active processes (e.g. transcription, replication and repair). Interestingly, recent research has shown a remarkable variability in TAD organization at the single-cell level.[51,52] As an aside, we note that eukaryotic SMCs have been studied at single molecule level, revealing that their extrusion speed is within the range of few hundreds to one thousand bp/s and is hardly influenced by interactions with other DNA-binding proteins.[38,53]

1.4.3. INTERACTION OF TRANSCRIPTION AND LOOP EXTRUSION

An interaction between loop extrusion and transcription was suggested in prokaryotes, where removing all endogenous *parS* sites and inserting one in an asymmetric location (between *ori* and *ter*), resulted in uneven speed of alignment of the individual chromosome arms in *B. subtilis*.[38] This was explained by mode of action of SMC-ScpAB that permits to align the chromosomal arms in an independent fashion. If one arm experiences different “resistance”, this can explain the observed asymmetry in extrusion speeds. The asymmetry of experienced “resistance” was attributed to the fact that bacterial genes have preferential orientation (75% genes in *B. subtilis* are oriented in *ori* ->*ter* direction) and SMC translocating towards the *ori* region will therefore experience more frequent head-to-head collisions with RNAP.

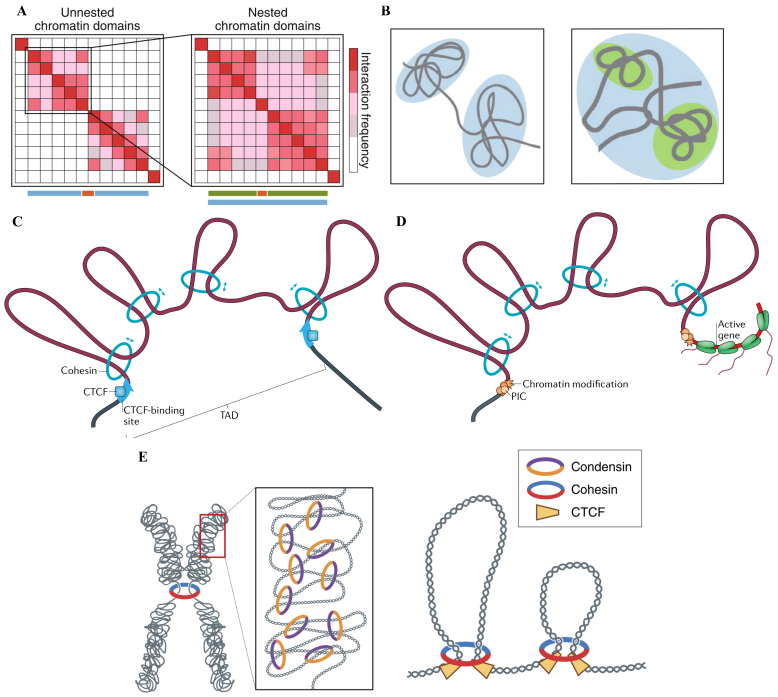

Figure 1.5: Chromosome compartmentalization and the role of loop extrusion. A) Illustrative contact maps, and B) globular interactions of hierarchical domain chromosomal organization. Colors and strips denote TADs (blue), boundaries (red) and subTADs (green). C) Convergent CTCF sites are often found at domain boundaries and represent pause site for cohesin. D) In other cases, boundaries are demarked by sites if active transcription, location of promoters (PIC – transcription pre-imitation complex) and chromatin modifications. E) Loop extruding factors contribute to domain organization. Cohesin extrudes loops of eukaryotic DNA and holds together sister chromatids in mitosis. Condensin further compacts mitotic chromosome. A, B) Adapted from[48], C, D) from [43], E) from[49].

Similarly, transcription has been shown to interact with loop extrusion in eukaryotes. CTCF and WAPL deletion experiments shown that cohesin translocated to 3'end, possibly due to being biased by contacts with progressing RNA polymerase (RNAP) II or indirectly via transcription induced supercoiling.[54] It was further shown that rRNA genes, that exhibit high density of RNAP (~10 RNAP/kb; about 100-fold that of protein-coding gene) and are highly transcribed, serve as stronger barriers to SMC progression. Interestingly single RNAP that transcribes protein coding gene (non-rRNA gene) is on its own stronger steric barrier (about 5-fold) to SMC progression, likely due to the fact that it binds ribosome, which makes it take up larger volume, compared to RNAP transcribing rRNA gene (that is not ribosome-bound). In neither case is the barrier impermeable, and the data suggested that condensins can bypass it in just about 1-5 ATP-hydrolysis steps. More generally, SMCs have been shown to bypass barriers much larger than their lumen size.[55] Further, ChIP-Seq data indicated that SMCs often clustered at loci where RNAP did, and did so with twice the intensity at operons oriented at direction opposing condensin movement.[38] Interestingly, SMCs clustered also at locations that did not overlap

with RNAP enrichment, suggesting role of other DNA-associated proteins in hindering progression of SMCs along the DNA.[56,57]

In conclusion, transcription plays an important role in organizing the chromosome on a fine scale, as well as contributes to higher level organization of active and inactive chromatin. On fine scale, the mechanism that promotes self-interaction of active and inactive DNA is not fully known, but it is likely that multivalent interactions of transcription associated proteins drive de-mixing of the two chromatin states. On larger scale, transcription contributes to formation of domain boundaries and compartmentalization, both through physical mechanism of supercoiling and plectoneme formation, as well as its interaction with loop extrusion.

1.5. CHROMATIN ORGANIZATION BY PHASE SEPARATION

Compartmentalization is a general tool for creating structure. Biology creates functionally and biochemically distinct compartments to spatiotemporally regulate intracellular processes. Recent decade was marked by establishment of our appreciation for how cells can partition their contents without need the for membranes. This is accomplished by process called liquid-liquid phase separation (LLPS) that is driven by multivalent transient weak interactions among subset of intracellular species that entropically favor partitioning into a condensed phase (a droplet; Fig 1.6A). Such condensate has propensity to interact with some molecular species, and exclude others, which confers it with unique functionality and biochemical properties.

The presence of LLPS in eukaryotic cells is established and many bodies are known to form by the process of LLPS: stress granules, Cajal bodies, nucleoli, nuclear speckles and paraspeckles to name just few. More recently, phase separation (PS) has been observed to form transcriptional condensates on the level of single transcriptional unit (Fig. 1.6C).[58] The above results and the nature of LLPS, including the relatively relaxed requirements for its establishment, suggest that the phenomenon is likely universal across domains of life. Indeed, some authors argue that phase separation may be so common that cells may actually have to actively exert energy to prevent most of its contents from forming condensates all the time.[59]

A major driver of phase separation in cells are the intrinsically disordered regions (IDRs), or intrinsically disordered proteins (IDPs). These have gained much attention in the last decade as they defy the previously held dogma of tertiary structure being indispensable for function. Rather, these IDRs and IDPs fulfill specific tasks and functions in a cell precisely thanks to their disorder. Many IDPs have been identified in eukaryotic cells. In relation to chromosomal organization, a major example thereof is HP1α that forms biomolecular condensates upon phosphorylation of its N-terminal domain (CTD). Heterochromatin preferentially partitions into these droplets and is further compacted therein.[60]

1.5.1. TRANSCRIPTION AND PHASE SEPARATION

Components of transcription machinery have been shown to self-associate and concentrate in liquid condensates[61] at sites of active transcription (transcriptional condensates) (Fig. 1.6C). The association is mediated by IDRs of RNAP, Mediator complex, transcription factors (TFs) and transient promoter-enhancer (P-E) interactions. The concentration of active transcription site to a condensate could facilitate the P-E search in 3D space. Further, LLPS could drive fusion of transcriptional condensates and colocalization of multiple transcriptionally active sites, which would e.g. explain the ability of enhancers to influence multiple promoters (and vice versa) and the fact that these interactions are sometimes inter-chromosomal[62], both of which would further confer chromatin with structure (Fig. 1.6A). In eukaryotes, this would form the basis for self-association of euchromatin and on yet larger scale the partitioning to A/B compartments (Fig. 1.6B), which describe a feature of Hi-C data where preferentially interacting sequences were shown to well overlap with their transcriptional state (A – broadly active, B – broadly inactive).

Recent research points to importance of LLPS for structuring genomes (Fig. 1.6D). Multibromodomain protein BRD4 binds DNA and self-associates in regions of low chromatin density and further excludes chromatin from the growing droplet, while simultaneously bringing together distant loci by droplet adhesion and coalescence to minimize surface tension.[65] The preference for condensation in regions of low chromatin density arises from the necessity of newly formed condensation seed (which can theoretically occur homogeneously along the DNA) to grow beyond critical nucleation diameter and displace surrounding DNA in doing so. The surrounding chromatin can be interpreted as viscoelastic matrix. The lower the chromatin density, the lower its stiffness and the easier it is for droplet to grow. The size of the droplet is then set by elasticity of the matrix.[67]

Overall, the results described in the above paragraph evidence of the ability of LLPS not only to compartmentalize, but also to give rise to forces that restructure genome, while also providing specificity in doing so. Further, they show the ability of the genome to govern LLPS by serving not only as location of seeds but also by mechanically constraining condensate growth. In opposite direction, restructuring the genome (by LLPS) has likely consequence for gene regulation. In addition to the above transcriptional regulators, DDX4 protein was shown to exclude chromatin, but not ssRNA which could form the basis for transcriptional regulation by RNA that is pulled into the condensate from adjacent regions.

Although much attention has been given to LLPS, it is likely that adjacent phenomena are equally important in genome architecture. These are the polymer-polymer phase separation (PPPS) (Fig. 1.6D) or the bridging-induced phase separation (BIPS).[68] Indeed, several phenomena that have been previously attributed to LLPS could be equally interpreted as PPPS or BIPS.[66] PPPS and BIPS arise from DNA-binding proteins with multiple chromatin recognition sites that bridge DNA. Such a mechanism could become important in presence of positive feedback-loop mechanism that promotes formation of bridges in already bridged regions and, in some cases would even not require interac-

tions among bridging partners.

1.5.2. PHASE SEPARATION IN PROKARYOTES

To date, phase separation remains less studied in bacteria and prokaryotes in general. This owes to the fact that bacteria are smaller than eukaryotic cells and therefore more difficult to study with optical imaging techniques. Second contributor is the sheer diversity of prokaryotic domain, with many species that are barely studied and even more that are yet to be discovered. However, evidence suggests that phase separation is an ancient process that is employed for organization by all forms of life. For example, over 100 different proteins have been reported to form "patchy-fluorescent-loci" in *C. crescentus* making them possible candidates for LLPS condensates.[69]

Of possible roles of phase separation in prokaryotes, self-association of active chromatin was hypothesized to translocate to nucleoid periphery where it would be more available for ribosomes (Fig. 1.7A). Interestingly, this does not seem to be required in all cell stages, as has been show in experiments with the bacterial starvation protein Dps[71], where DNA formed densely compacted crystalline-like structure that remained largely permissible for access of transcription machinery and translation. Among other example of PS in bacteria belong (Fig. 1.7): FtsZ SlmA and DNA complex formed during cell division[72], the α-proteobacterial (*C. crescentus*) RNA degradosome[73], McdAB protein-cargo positioning and segregation system in β-cyanobacteria[74,75], MDP-1 histone-like protein that contributes to chromatin structure in mycobateria[76], PopZ that is essential for equipartitioning intracellular cargo during cell division[70] and most recently RNA polymerase in *E. coli*[77] and the ParABS chromosome and DNA-cargo segregation system.[59,78] Condensate formation has been shown also in case of H-NS, but due presence of confounding factors the results were not conclusive.[79]

1.5.3. EXPERIMENTAL TECHNIQUES TO STUDY PHASE SEPARATION

The formation of phase separated condensates is governed by a phase diagram that is defined by a set of parameters such as temperature, concentration and interaction strength. At fixed interaction strength and temperature increasing molecular concentration above the saturation results in phase separation, an abrupt effect associated with a step-like change of properties. The concentration of various components can be controlled by the strength of their expression and inducible degradation. At fixed concentration, phase separation can be driven by changes in interaction strength, which can be influenced by factors like protein and nucleic acid sequence, inducible modifications, and ionic strength. The interaction strength can be changed during an experiment e.g. by light or chemically inducible oligomerization. A common approach to check whether a compartment exhibits liquid-like properties is to FRAP it and check whether the fluorescence intensity will be recovered.

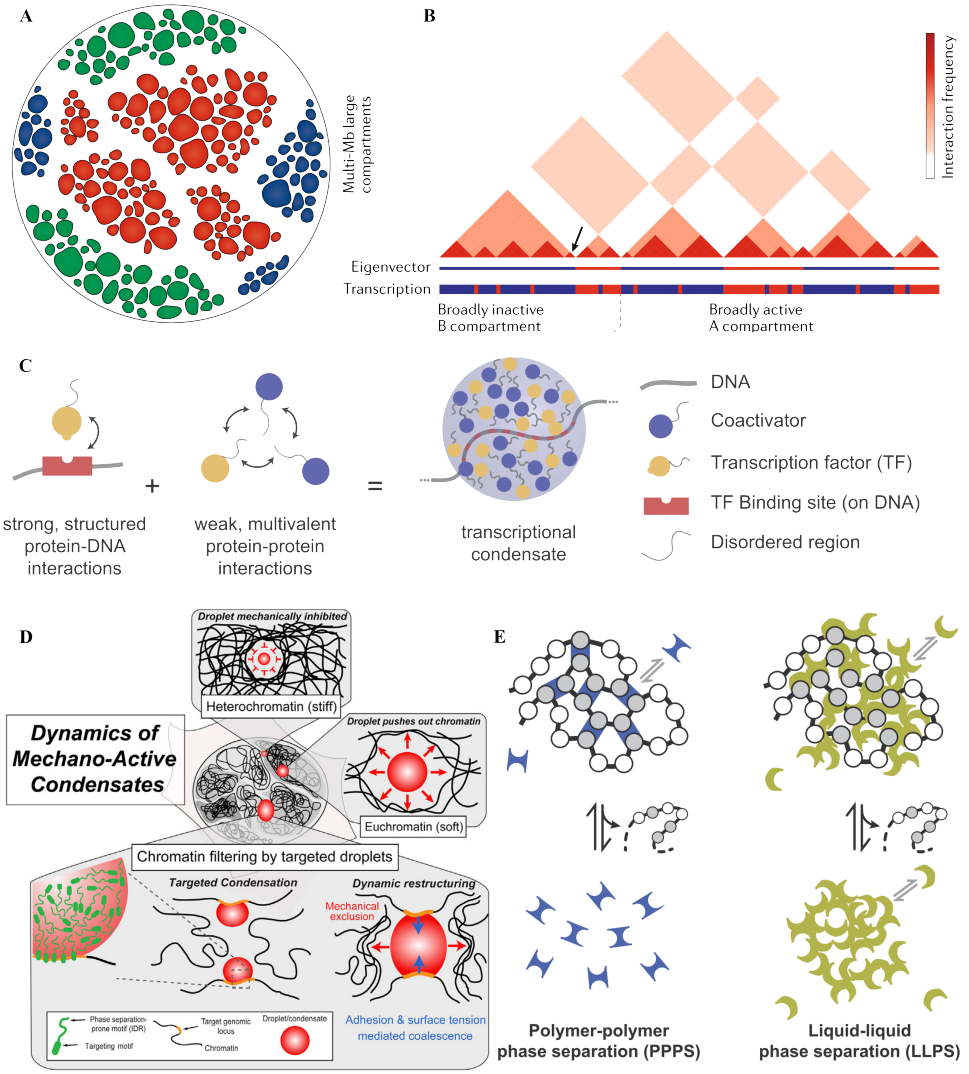

Figure 1.6: Phase separation structures and organizes genome across scales. A) On mesoscopic level, chromatin partitions to different locations within the nucleus based on its properties (red - transcriptionally active euchromatin, blue and green – different phases of heterochromatin). B) Hi-C maps show tendency for self-association of transcribed and silent sequences. The regions are termed A (broadly active) and B (broadly inactive) compartments. D) Transcription machinery at an active site has been shown to associate to a condensate that may have implications for P-E search, gene regulation and higher order chromatin structure. A, B) Adapted from[63], C) from.[64] D) Liquid condensates can promote domain association of active chromatin, exclude silent regions and exert forces capable of restructuring genome while simultaneously serving as a platform for gene expression regulation. E) Despite high popularity of LLPS, PPPS could yield same experimental results while allowing for different interpretation of underlying mechanism. Blue – bridging factor, Lime – self-associating weak chromatin binder D) Adapted from[65], E) from.[66]

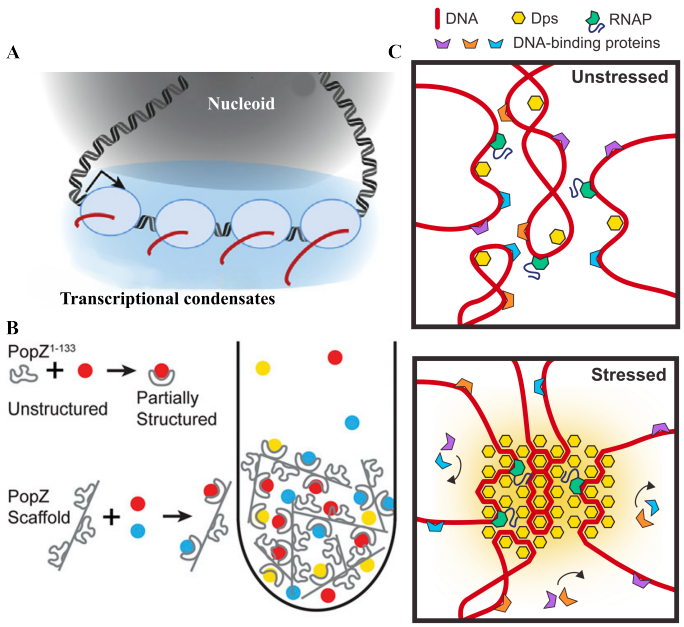

Figure 1.7: Examples of phase separation in prokaryotes. A) Transcriptional condensates are likely present also in prokaryotes where they could promote localization of active chromatin to nucleoid periphery. B) PopZ plays a role in eqi-partitioning cellular content during cell division. Its ability to form multivalent weak interactions with itself and binding partners (colored circles) that interact with its disordered region lead to formation of meshwork that locally changes diffusion dynamics and traps cellular content at poles. C) Dps is the most abundant protein in starvation phase, associates with chromosome and promotes crystalline-like dense structure. Despite strong condensation, RNAP can still access and transcribe the genome and this suggests that Dps may form a condensate that exclude/allows molecules based on their biochemical properties. A) Adapted from[12], B) from[70] and C) from.[71]

1.5.4. OPEN QUESTIONS IN CHROMOSOME ORGANIZATION

While chromosome organization is a well-established topic, a large number of questions remain subject to active research. Below I list a number of selected questions that potentially can be well addressed with the experimental methods developed in this thesis:

- What is the relative importance of crowding and chromatin-associated proteins on chromosome compaction?
- What is the structure and dynamics of isolated chromosomes?
- How stable are domains of chromosome organization in the absence of confinement?
- What enzymatic and nonenzymatic approaches are most suitable for deproteination of megabase-sized DNA scaffold?
- How can a large DNA scaffold be transferred between cells and in vitro without causing excessive damage?
- What are suitable approaches to quantify the integrity and function of isolated DNA?
- What is the effect of ATP-driven and ATP-independent action of SMCs on megabase-scale DNA scaffold?
- What is the degree of compaction that can be achieved on bare DNA by SMCs alone? And what is the average loop size extruded by SMCs in this context?
- How can polymer dynamics simulations synergize with single-molecule biophysics techniques to describe the DNA-protein interactions on a megabase-scale isolated chromosome?
- What is the interplay between transcription and 3D-organized chromatin?
- What is the role of RNA for regulating chromatin architecture independently of transcription?
- What is the mechanistic basis of translocation of actively transcribed DNA?
- What is the mechanism of long-range promoter-enhancer contacts? Are they enriched upon loop extrusion?
- Can we dynamically tune accessibility to a synthetic/isolated chromosome by controlling its phase-separated state?
- How do individual NAPs interact with transcription machinery to structure the chromosome?
- Can we study transcriptional output from subset of genes along a chromosome to characterize how compact and accessible the chromosome is?
- How do we characterize isolated chromosomes from fluorescence images? What metrics are suitable to describe their size, structure, and dynamics?
- Would chromatin capture technology be effective to study the structure of isolated megabase-scale scaffold?

While we were not able to address all of these open questions, we did tackle some, and in doing so, we hope to have laid a solid technical groundwork to future research. It is with interest that I look to future developments in this area and future research leveraging the methods established here.

1.6. OTHER SUBJECTS ADDRESSED IN THIS THESIS

Besides the biophysics of chromosome organization, my graduate work included engagement with iGEM, EMBO fellowship in synthetic genomics, and work on biosecurity. This also gave rise to several, more general, questions:

- Looking ahead to the future where biology is increasingly easy to engineer (and we can build, e.g., whole chromosomes), how do we make sure that biological engineering techniques are available in an appropriate ethical context and with the right set of restraints to prevent misuse?
- What modern tools can assist researchers to reduce the likelihood of such misuse?
- Given the recent pandemic, how do we decrease chance of its recurrence?
- Given the importance of human-animal interfaces for the spread of zoonotic diseases and the ongoing global infringement on biodiversity, how do we monitor this interface adequately?
- Given the rapid advancements and importance of biotechnology to a sustainable future,
- how do we train a large enough workforce to address the needs of bioeconomy?

These questions are addressed in later parts of my thesis.

1

1.7. REFERENCES

1. Milo, R. *Cell Biology by the Numbers.* (Garland Science, Taylor & Francis Group, LLC, New York NY USA, 2016).

2. Jun, S. & Wright, A. Entropy as the driver of chromosome segregation. *Nat Rev Microbiol* **8**, 600–607 (2010).

3. Surovtsev, I. V. & Jacobs-Wagner, C. Subcellular Organization: A Critical Feature of Bacterial Cell Replication. *Cell* **172**, 1271–1293 (2018).

4. Marko, J. F. *Biomechanics in Oncology.* (Springer Science+Business Media, New York, NY, 2018).

5. Hagerman, P. J. Flexibility of DNA. 24.

6. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci USA* **115**, E6697–E6706 (2018).

7. Dame, R. T., Rashid, F.-Z. M. & Grainger, D. C. Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nat Rev Genet* (2019) doi:10.1038/s41576-019-0185-4.

8. Marko, J. F. Physics and Biology (of Chromosomes). *Journal of Molecular Biology* S002228361930703X (2019) doi:10.1016/j.jmb.2019.11.022.

9. Cunha, S., Woldringh, C. L. & Odijk, T. Polymer-Mediated Compaction and Internal Dynamics of Isolated Escherichia coli Nucleoids. *Journal of Structural Biology* **136**, 53–66 (2001).

10. Odijk, T. Osmotic compaction of supercoiled DNA into a bacterial nucleoid. *Biophysical Chemistry* **73**, 23–29 (1998).

11. Schavemaker, P. E., Śmigiel, W. M. & Poolman, B. Ribosome surface properties may impose limits on the nature of the cytoplasmic proteome. *eLife* **6**, e30084 (2017).

12. Shen, B. A. & Landick, R. Transcription of Bacterial Chromatin. *Journal of Molecular Biology* **431**, 4040–4066 (2019).

13. Peeters, E., Driessen, R. P. C., Werner, F. & Dame, R. T. The interplay between nucleoid organization and transcription in archaeal genomes. *Nat Rev Microbiol* **13**, 333–341 (2015).

14. Lal, A. *et al.* Genome scale patterns of supercoiling in a bacterial chromosome. *Nat Commun* **7**, 11055 (2016).

15. Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S. & Ishihama, A. Growth Phase-Dependent Variation in Protein Composition of the Escherichia coli Nucleoid. *Journal of Bacteriology* **181**, 6361–6370 (1999).

16. Bettridge, K., Verma, S., Weng, X., Adhya, S. & Xiao, J. *Single Molecule Tracking Reveals the Role of Transitory Dynamics of Nucleoid-Associated Protein HU in Organizing the Bacterial Chromosome.* http://biorxiv.org/lookup/doi/10.1101/2019.12.31.725226 (2019) doi:10.1101/2019.12.31.725226.

17. Qian, Z. *et al.* A New Noncoding RNA Arranges Bacterial Chromosome Organization. *mBio* **6**, e00998-15 (2015).

18. Bignaud, A. *et al.* Transcription-induced domains form the elementary constraining building blocks of bacterial chromosomes. *Nat Struct Mol Biol* **31**, 489–497 (2024).

19. Woldringh, C. L. The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. *Mol Microbiol* **45**, 17–29 (2002).

20. Wang, X. & Sherratt, D. J. Independent Segregation of the Two Arms of the Escherichia coli ori Region Requires neither RNA Synthesis nor MreB Dynamics. *Journal of Bacteriology* **192**, 6143–6153 (2010).

21. Le, T. B. K., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science* **342**, 731–734 (2013).

22. Marbouty, M. *et al.* Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. *Molecular Cell* **59**, 588–602 (2015).

23. Le, T. B. & Laub, M. T. Transcription rate and transcript length drive formation of chromosomal interaction domain boundaries. *EMBO J* **35**, 1582–1595 (2016).

24. Postow, L. Topological domain structure of the Escherichia coli chromosome. *Genes & Development* **18**, 1766–1779 (2004).

25. Rowley, M. J. *et al.* Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Molecular Cell* **67**, 837-852.e7 (2017).

26. Chen, H. *et al.* Dynamic interplay between enhancer–promoter topology and gene activity. *Nat Genet* **50**, 1296–1303 (2018).

27. Ganji, M. *et al.* Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102–105 (2018).

28. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* eaaz3418 (2019) doi:10.1126/science.aaz3418.

29. Tran, N. T., Laub, M. T. & Le, T. B. K. SMC Progressively Aligns Chromosomal Arms in Caulobacter crescentus but Is Antagonized by Convergent Transcription. *Cell Reports* **20**, 2057–2071 (2017).

30. Wang, X., Brandão, H. B., Le, T. B. K., Laub, M. T. & Rudner, D. Z. *Bacillus subtilis* SMC complexes juxtapose chromosome arms as they travel from origin to terminus. *Science* **355**, 524–527 (2017).

31. Sullivan, N. L., Marquis, K. A. & Rudner, D. Z. Recruitment of SMC by ParB-parS Organizes the Origin Region and Promotes Efficient Chromosome Segregation. *Cell* **137**, 697–707 (2009).

32. Gruber, S. & Errington, J. Recruitment of Condensin to Replication Origin Regions by ParB/SpoOJ Promotes Chromosome Segregation in B. subtilis. *Cell* **137**, 685–696 (2009).

33. Yano, K. & Niki, H. Multiple cis -Acting rDNAs Contribute to Nucleoid Separation and Recruit the Bacterial Condensin Smc-ScpAB. *Cell Reports* **21**, 1347–1360 (2017).

34. Wang, X. *et al.* Condensin promotes the juxtaposition of DNA flanking its loading site in *Bacillus subtilis*. *Genes Dev.* **29**, 1661–1675 (2015).

35. Rajasekar, K. V. *et al.* Dynamic architecture of the Escherichia coli structural maintenance of chromosomes (SMC) complex, MukBEF. *Nucleic Acids Research* **47**, 9696–9707 (2019).

36. Nolivos, S. *et al.* MatP regulates the coordinated action of topoisomerase IV and MukBEF in chromosome segregation. *Nat Commun* **7**, 10466 (2016).

37. Mäkelä, J. & Sherratt, D. J. Organization of the Escherichia coli Chromosome by a MukBEF Axial Core. *Molecular Cell* **78**, 250-260.e5 (2020).

38. Brandão, H. B. *et al.* RNA polymerases as moving barriers to condensin loop extrusion. *Proc Natl Acad Sci USA* **116**, 20489–20499 (2019).

39. Lioy, V. S. *et al.* Multiscale Structuring of the E. coli Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell* **172**, 771-783.e18 (2018).

40. Naumova, N. *et al.* Organization of the Mitotic Chromosome. *Science* **342**, 948–953 (2013).

41. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

42. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).

43. van Steensel, B. & Furlong, E. E. M. The role of transcription in shaping the spatial organization of the genome. *Nat Rev Mol Cell Biol* **20**, 327–337 (2019).

44. Eelena M. Klenova, H. F. P., Robert H. Nicolas & Lobanenkov, V. V. CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-zn-finger protein differentially expressed in multiple forms. *Molecular and Cellular Biology* **13**, 7612–7624 (1993).

45. Phillips-Cremins, J. E. *et al.* Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* **153**, 1281–1295 (2013).

46. Mizuguchi, T. *et al.* Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. *Nature* **516**, 432–435 (2014).

47. Benedetti, F., Racko, D., Dorier, J., Burnier, Y. & Stasiak, A. Transcription-induced supercoiling explains formation of self-interacting chromatin domains in S. pombe. *Nucleic Acids Research* **45**, 9850–9859 (2017).

48. Beagan, J. A. & Phillips-Cremins, J. E. On the existence and functionality of topologically associating domains. *Nat Genet* **52**, 8–16 (2020).

49. Eeftens, J. & Dekker, C. Catching DNA with hoops—biophysical approaches to clarify the mechanism of SMC proteins. *Nature Structural & Molecular Biology* **24**, 1012–1020 (2017).

50. Cardozo Gizzi, A. M., Cattoni, D. I. & Nollmann, M. TADs or no TADS: Lessons From Single-cell Imaging of Chromosome Architecture. *Journal of Molecular Biology* S0022283619307442 (2020)

doi:10.1016/j.jmb.2019.12.034.

51. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).

52. Szabo, Q. *et al.* Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat Genet* **52**, 1151–1157 (2020).

53. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).

54. Racko, D., Benedetti, F., Dorier, J. & Stasiak, A. Transcription-induced supercoiling as the driving force of chromatin loop extrusion during formation of TADs in interphase chromosomes. *Nucleic Acids Research* **46**, 1648–1660 (2018).

55. Pradhan, B. *et al.* SMC complexes can traverse physical roadblocks bigger than their ring size. *Cell Reports* **41**, 111491 (2022).

56. Davidson, I. F. *et al.* CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature* **616**, 822–827 (2023).

57. Analikwu, B. T. *et al.* Telomere protein arrays stall DNA loop extrusion by condensin. Preprint at https://doi.org/10.1101/2023.10.29.564563 (2023).

58. Plys, A. J. & Kingston, R. E. Dynamic condensates activate transcription. *Science* **361**, 329–330 (2018).

59. Guilhas, B. *et al.* *ATP-Driven Separation of Liquid Phase Condensates in Bacteria.* http://biorxiv.org/lookup/doi/10.1101/791368 (2019) doi:10.1101/791368.

60. Larson, A. G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).

61. Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).

62. Lim, B., Heist, T., Levine, M. & Fukaya, T. Visualization of Transvection in Living Drosophila Embryos. *Molecular Cell* **70**, 287-296.e6 (2018).

63. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat Rev Genet* **19**, 789–800 (2018).

64. Shrinivas, K. *et al.* Enhancer Features that Drive Formation of Transcriptional Condensates. *Molecular Cell* **75**, 549-561.e7 (2019).

65. Shin, Y. *et al.* Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell* **175**, 1481-1491.e13 (2018).

66. Erdel, F. & Rippe, K. Formation of Chromatin Subcompartments by Phase Separation. *Biophysical Journal* **114**, 2262–2270 (2018).

67. Style, R. W. *et al.* Liquid-Liquid Phase Separation in an Elastic Network. *Phys. Rev. X* **8**, 011028 (2018).

68. Ryu, J.-K. *et al.* Bridging-induced phase separation induced by cohesin SMC protein complexes. *Science Advances* **7**, eabe5905.

69. Werner, J. N. *et al.* Quantitative genome-scale analysis of protein localization in an asymmetric bacterium. *Proceedings of the National Academy of Sciences* **106**, 7858–7863 (2009).

70. Holmes, J. A. *et al. Caulobacter* PopZ forms an intrinsically disordered hub in organizing bacterial cell poles. *Proc Natl Acad Sci USA* **113**, 12490–12495 (2016).

71. Janissen, R. *et al.* Global DNA Compaction in Stationary-Phase Bacteria Does Not Affect Transcription. *Cell* **174**, 1188-1199.e14 (2018).

72. Monterroso, B. *et al.* Bacterial FtsZ protein forms phase-separated condensates with its nucleoid-associated inhibitor SlmA. *EMBO Rep* **20**, (2019).

73. Al-Husini, N., Tomares, D. T., Bitar, O., Childers, W. S. & Schrader, J. M. α-Proteobacterial RNA Degradosomes Assemble Liquid-Liquid Phase-Separated RNP Bodies. *Molecular Cell* **71**, 1027-1039.e14 (2018).

74. Schumacher, M. A., Henderson, M. & Zhang, H. Structures of maintenance of carboxysome distribution Walker-box McdA and McdB adaptor homologs. *Nucleic Acids Research* **47**, 5950–5962 (2019).

75. Wang, H. *et al.* Rubisco condensate formation by CcmM in β-carboxysome biogenesis. *Nature* **566**, 131–135 (2019).

76. Savitskaya, A. *et al.* C-terminal intrinsically disordered region-dependent organization of the mycobacterial genome by a histone-like protein. *Sci Rep* **8**, 8197 (2018).

77. Ladouceur, A.-M. *et al.* Clusters of bacterial RNA polymerase are biomolecular condensates that assemble through liquid–liquid phase separation. *Proc Natl Acad Sci USA* 202005019 (2020) doi:10.1073/pnas.2005019117.

78. Babl, L. *et al.* CTP-controlled liquid–liquid phase separation of ParB. *Journal of Molecular Biology* **434**, 167401 (2022).

79. Wang, S., Moffitt, J. R., Dempsey, G. T., Xie, X. S. & Zhuang, X. Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proceedings of the National Academy of Sciences* **111**, 8452–8457 (2014).

2

EXTRACTING AND CHARACTERIZING PROTEIN-FREE MEGABASEPAIR DNA FOR *in vitro* EXPERIMENTS

Chromosome structure and function is studied using various cell-based methods as well as with a range of *in vitro* single-molecule techniques on short DNA substrates. Here we present a method to obtain megabasepair length deproteinated DNA for *in vitro* studies. We isolated chromosomes from bacterial cells and enzymatically digested the native proteins. Mass spectrometry indicated that 97-100% of DNA-binding proteins are removed from the sample. Fluorescence-microscopy analysis showed an increase in the radius of gyration of the DNA polymers, while the DNA length remained megabasepair sized. In proof-of-concept experiments using these deproteinated long DNA molecules, we observed DNA compaction upon adding the DNA-binding protein Fis or PEG crowding agents and showed that it is possible to track the motion of a fluorescently labelled DNA locus. These results indicate the practical feasibility of a 'genome-in-a-box' approach to study chromosome organization from the bottom up.

This chapter has been published: Martin Holub*, Anthony Birnie*, Aleksandre Japaridze, Jaco van der Torre, Maxime den Ridder, Carol de Ram, Martin Pabst, Cees Dekker, *Extracting and characterizing protein-free megabasepair DNA for in vitro experiments*, Cell Reports Methods 2, 100366 (2022). *Equal contribution

2.1. INTRODUCTION

Over the past decade, bottom-up synthetic cell research or 'bottom-up biology' has gained traction as a method to study components of living systems. The ultimate aim of such efforts is to build a synthetic cell by assembling biological functionalities from the bottom up. This involves the reconstitution of the various parts of living cells from a set of well-characterized but lifeless molecules such as DNA and proteins.[1] While the end goal of building a functional synthetic cell is yet far off, the bottom-up approach has already successfully been applied to constitute and study minimal cellular systems, for example, intracellular pattern formation,[2] cell division,[3] the cytoskeleton,[4] and cellular communication.[5]

For studying chromosome organization in the eukaryotic nucleus or in bacterial cells, numerous studies have been made on live or fixed cells through imaging,[6,7] chromosome conformation capture techniques,[8,9] *etc.*, while *in vitro* protein-DNA interactions are often characterized at the single-molecule level using techniques such as Atomic Force Microscopy,[10–12] magnetic[13,14] and optical tweezers,[15,16] and DNA visualization assays.[17–21] While these complementary approaches have yielded great insights, they leave a significant gap since typical single-molecule methods address the kilobasepair (kbp) scale while actual genomes consist of $10^5 - 10^{11}$ bp long DNA. It would therefore be useful to study DNA in the megabasepair size range with bottom-up *in vitro* methods, including the emergent collective behavior associated with this length scale. We propose that such experiments, which we coin a 'genome-in-a-box' (GenBox) approach,[22] may provide valuable insights into chromosome organization, somewhat analogous to the 'particle-in-a-box' experiments in physics which proved a useful abstraction to understand basic phenomena in quantum mechanics. However, such a GenBox method has so far been lacking. Expanding from the kbp to the Mbp scale poses technical challenges, both in the handling of long DNA that is prone to shearing,[23–25] and in the availability of long DNA, as common *in vitro* experiments[26–28] are done on viral DNA (such as the 48.5 kbp lambda-phage DNA) which however is limited in length. Several previous studies have proposed methods to extract chromosomes from cells, and some have even used protein-removal steps to obtain deproteinated DNA.[29–34] However, most of these studies lacked an imaging-based characterization of the resulting DNA objects, regarding their size, level of deproteination, and suitability for *in vitro* imaging-based experiments.

Here, we present a methodology for the extraction of chromosomal DNA from *E. coli* bacteria and the subsequent removal of native proteins, resulting in deproteinated DNA of megabasepair size which can be used for *in vitro* bottom-up experiments to study chromosome organization (Figure 2.1). We describe the extraction and purification protocol, characterize the DNA objects obtained, and present some first proof-of-principle experiments.

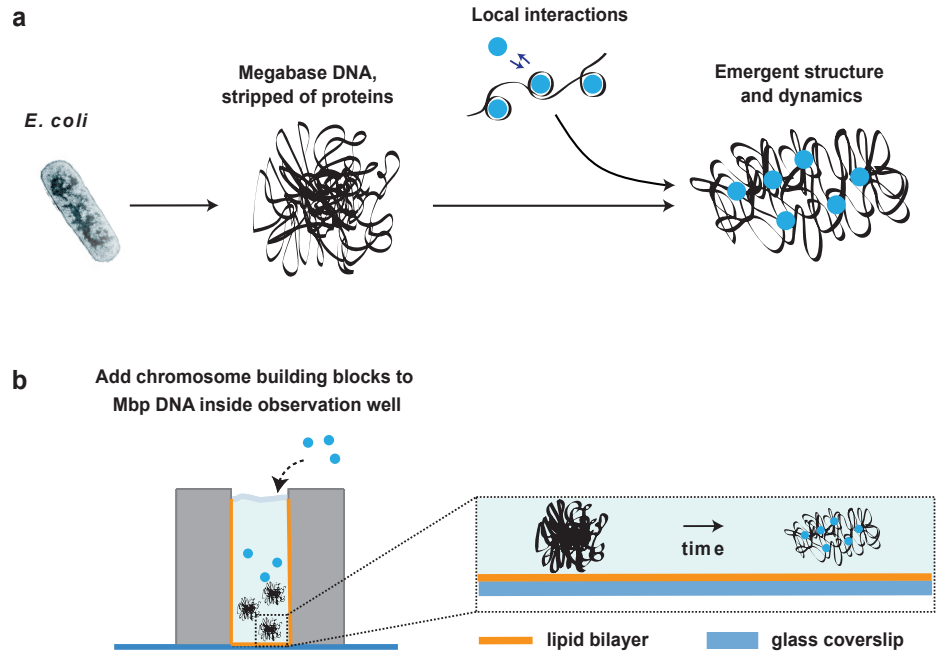

Figure 2.1: Methodology of extracting, purifying, and studying a bacterial chromosome. **(a)** In a Genome-in-a-box (GenBox) approach, one isolates chromosomes from bacterial cells, removes the natively bound proteins, to subsequently add DNA-structuring elements and thus study the resulting emergent DNA structure. **(b)** Typical setup where a deproteinated megabasepair-long DNA is suspended in solution in an observation well attached to a glass coverslip. The surface of the observation well is coated with a lipid bilayer to prevent DNA adhesion to the surface. DNA-binding elements are added and the resulting DNA structure is observed using fluorescence microscopy.

2.2. RESULTS

The workflow to obtain and characterize deproteinated megabasepair DNA consisted of several experimental steps, which are discussed in the following sections. First, we ensured and verified that the *E. coli* bacteria contained a single 4.6 Mbp chromosome by cell-cycle arrest. Then chromosomes were extracted from the cells in one of two routes, either directly in solution or *via* embedding them in an agarose gel plug. Lastly, the isolated chromosomes were deproteinated using a protease enzyme. Mass spectrometry was used to confirm the level of deproteination, followed by microscopy imaging and quantitative analysis of the total fluorescence intensity per object and the radius of gyration (R_g). This was done in order to verify if the chromosomes remained intact throughout the protocol, as well as to assess the effect of deproteination of the size of the DNA objects. Finally, as a proof of concept, three examples of possible experiments are shown.

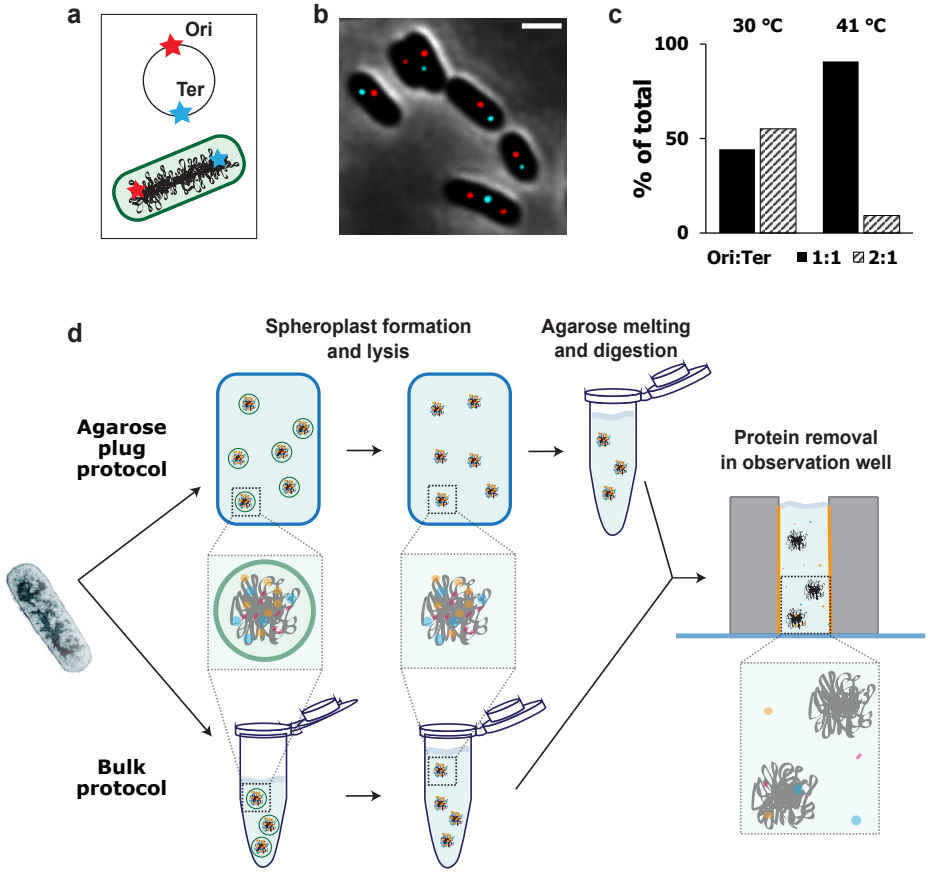

Figure 2.2: Workflow of the protocol. **(a)** The *E. coli* chromosome is circular and contains FROS arrays near the Origin of replication (Ori) and Terminus of replication (Ter). **(b)** Deconvolved image of *E. coli* cells with the Ori and Ter location labeled in red and cyan, respectively. Scale bar 2 μm. **(c)** Origin to Terminus of replication (Ori:Ter) ratio in control and temperature-treated *E. coli* cells (N = 185 and 178, respectively). **(d)** Agarose plug and bulk protocol to prepare deproteinated megabasepair DNA. Starting from *E. coli* cells, the cell wall is digested and the resulting spheroplasts are either embedded and lysed inside an agarose plug or directly lysed in a solution. After lysis in the agarose plugs, the agarose matrix is digested. At this stage, the chromosomes in both protocols are suspended in a solution and transferred to an observation well for protein removal and study of the deproteinated chromosomes.

2.2.1. EXTRACTING A SINGLE CHROMOSOME FROM *E. coli*

We prepared *E. coli* cells that contain only a single chromosome. In the exponential growth phase of bacteria, chromosomes are permanently replicating and typically exhibiting multiple replication forks on the DNA. For the purpose of controlled *in vitro* experiments this is undesirable for two reasons: first, halfway replicated DNA and multiple replication forks make the exact amount of DNA per cell unknown, and second,

DNA near replication forks is prone to damage and breaking.[35] As our aim is to extract DNA of a well-defined size, it is needed to obtain conditions that yield a known number of chromosomes per cell, ideally only a single chromosome per cell.

For this purpose, we used minimal media to avoid the occurrence of nested replication forks[36] as well as a temperature-sensitive *E. coli* strain where replication initiation was arrested by culturing the cells at an elevated temperature.[37,38] We grew cells for 2 hours (*i.e.*, for a time period longer than the doubling time in minimal media) at 41 °C and subsequently determined the number of chromosomes per cell by fluorescence imaging. The *E. coli* cells were engineered to contain Fluorescent Repressor Activator System (FROS) arrays near the Origin (Ori) and Terminus (Ter) locations (Figure 2.2a). At the start of the DNA replication process, the Ori is duplicated upon which the remainder of the chromosome follows, while the Ter is only duplicated at the end. This means that cells with a partly replicated chromosome will contain two Ori spots and a single Ter spot, whereas cells containing a single chromosome will only show one Ori and Ter. By counting the Ori and Ter fluorescence spots per cell, we confirmed that 90% of cells contained a single chromosome (Figure 2.2b and 2.2c, Ori:Ter ratio 1:1), while 10% of cells were still in the process of DNA replication (Ori:Ter ratio 2:1). If one were to extract the DNA from these cells, one would therefore expect a size distribution in which 90% of the objects are 4.6 Mbp, whereas the remaining 10% would contain DNA at an amount of between 4.6 and 9.2 Mbp, depending on how far genome replication in the cell had proceeded at the time of DNA extraction. In control experiment with growth at 30 °C instead of replication arrest, 55% of cells were in a state of active DNA replication whereas 45% contained a single chromosome (Figure 2.2c).

In order to extract the chromosomes from *E. coli* cells, the peptidoglycan cell wall was degraded using lysozyme enzyme, resulting in spheroplasts which are wall-less rounded *E. coli* cells that merely are contained in their plasma membranes. To release the cellular contents including the DNA, the spheroplasts were submerged in a low-osmolarity buffer, which forces water to enter the spheroplasts, thereby rupturing them. This so-called lysis by osmotic shock was achieved on spheroplasts that were prepared with one of two methods (Figure 2.2d): *i)* direct lysis of the cytosolic content of the spheroplasts into solution, based on a protocol developed in the Woldringh lab[30,39] (hereafter called 'bulk protocol'), or *ii)* embedding of spheroplasts inside agarose gel plugs where they were subsequently lysed, following a protocol from the Glass lab[32] (hereafter called 'agarose plug protocol'). Embedding of the spheroplasts inside the agarose plug resulted in intact spheroplasts that did not get lysed prematurely (figure S2.1). Bulk isolation yielded DNA that could be used on the same day, while the agarose-plug protocol produced samples that could be stored for a period of up to weeks after isolation. Depending on the application, the agarose plug protocol may also present advantages regarding the handing of the DNA material, such as a reduced shearing in transferring between experimental steps.

2.2.2. VIRTUALLY ALL PROTEINS CAN BE REMOVED FROM EXTRACTED CHROMOSOMES

DNA in cells is compacted by confinement, crowding, and binding of DNA-associated proteins. After cell lysis, the boundary conditions of confinement and crowding no longer apply, but DNA-binding proteins can in principle remain attached to the DNA. To digest such DNA-binding proteins in the sample, we incubated the bulk and plug protocol samples with a thermolabile Proteinase K enzyme, which is a broad range serine protease that cleaves peptide bonds at the carboxylic sides at a variety of positions (*i.e.*, after aliphatic, aromatic, and hydrophobic amino acids). We observed increased DNA fragmentation after digesting and melting agarose plugs that had undergone proteinase treatment. Contrary to previous work,[32] we therefore opted for treating the agarose sample in liquid, instead of in the gel state. While the bulk protocol sample already was liquid, agarose plugs had to be first digested using beta-agarase enzyme that breaks down the polymers forming the agarose gel. After the 15 min deproteination treatment and subsequent enzyme heat-inactivation (to prevent protein digestion in downstream experiments), we quantified the resulting degree of protein removal by mass spectrometry (MS).

Two categories of proteins were distinguished in the MS experiments, namely DNA-binding proteins and non-DNA-binding proteins. Obviously, the removal of the DNA-binding proteins is most critical for obtaining deproteinated DNA for GenBox experiments. To aid the quantification, we compiled a list of the 38 most abundant DNA-binding proteins as well as DNA-binding protein sub-units (Table S2.1), based on the protein's description in the UniProt database as DNA-binding or DNA processing. For the bulk protocol (Table 2.1 and 2.2), we found that all DNA-binding proteins were removed (100%, at the MS resolution). For the agarose plug protocol (Table 2.1 and 2.3), the vast majority of the DNA-binding proteins, 97%, was removed. These percentages refer to protein abundances relative to control samples that underwent exactly the same treatment steps, but to which no Proteinase K was added. For the agarose plug protocol (Table 2.3), the major remaining DNA-binding proteins were IHF-A (14.8% remaining) and various RNA polymerase sub-units (rpoA/B/C, up to 4.5% remaining). The non-DNA-binding proteins were removed to the degree of 98.1% and 93.0% for the bulk and agarose plug protocol, respectively. More specifically, several ribosomal proteins were still present at large percentages (>40%) in the agarose plug sample.

	Bulk Protocol	**Agarose Protocol**
DNA-binding proteins (%)	0	3.9 ± 1.4
Non-DNA-binding proteins (%)	1.9 ± 0.3	7.0 ± 2.5

Table 2.1: Overall protein removal efficiency as measured by mass spectrometry. Overall percentage of proteins remaining after the protein removal treatment for bulk protocol and agarose protocols.

Protein	Function	Percentage (%) remaining
Non-DNA-binding:		
thrS	Threonine–tRNA ligase	56 ± 22
trxA	Thioredoxin 1	7.0 ± 2.5

Table 2.2: Protein removal efficiency in bulk protocol as measured by mass spectrometry. Individual remaining proteins in the bulk protocol. Only those non-DNA-binding proteins with more than 40% remaining are included in the table. Errors are standard deviation from the mean obtained from three independent experiments per condition ('before' and 'after'). See also Table S2.1.

Protein	Function	Percentage (%) remaining
DNA-binding:		
ihfA	Integration host factor subunit alpha	15 ± 11
rpoC	RNA polymerase subunit beta'	4.5 ± 1.5
rpoA	RNA polymerase subunit alpha	4.2 ± 3.2
rpoB	RNA polymerase subunit beta	0.9 ± 0.4
Non-DNA-binding:		
dppB	Dipeptide transport system permease protein	>100
rpmG	50S ribosomal protein L33	>100
lhgD	L-2-hydroxyglutarate dehy-drogenase	>100
frsA	Esterase FrsA	>100
rpmB	50S ribosomal protein L28	80 ± 61
cydA	Cytochrome bd-I ubiquinol oxidase subunit 1	60 ± 15
uraA	Uracil permease	50 ± 50
mlaB	Intermembrane phospholipid transport system binding protein	50 ± 46
rplU	50S ribosomal protein L21	50 ± 27
rplJ	50S ribosomal protein L10	45 ± 8
yraR	Putative NAD(P)-binding protein	43 ± 42
cyoB	Cytochrome bo(3) ubiquinol oxidase subunit 1	43 ± 15

Table 2.3: Protein removal efficiency in agarose plug protocol as measured by mass spectrometry. Individual remaining proteins in the agarose plug protocol. All remaining DNA-binding protein are included. while for non-DNA-binding proteins only those with more than 40% remaining are included in the table. The agarose plug protocol contained a few lower abundant proteins (dppB, rpmG, IhgD, frsA) for which higher relative abundancies were estimated (denoted with >100%) due to low level of protein removal. Errors are standard deviation from the mean obtained from three independent experiments per condition ('before' and 'after'). See also Table S2.1.

2.2.3. EXTRACTED CHROMOSOMES REMAIN OF MEGABASEPAIR LENGTH AND EXPAND IN SIZE AFTER PROTEIN REMOVAL

We imaged DNA resulting from the bulk and agarose plug protocols before and after protein removal by fluorescence imaging on a spinning disc confocal microscope using the DNA-intercalating dye Sytox-Orange (Figure 2.3c/d and Figure S2.2). From a first visual inspection we observed that, before protein removal, the DNA objects contain a dense/bright core with a lower density 'cloud' surrounding it (Figure 2.3c-purple, Figure 2.3d-orange/purple, and Figure S2.2a/c/d). After protein removal, the objects seemed to be larger and more spread out (Figure 2.3c/d-green, and Figure S2.2b/e). In order to make more quantitative statements, we developed a semi-automated analysis script in Python (see STAR Methods for a detailed description), with which we identified individual DNA objects in the images, segmented them from the background, and quantified their radius of gyration R_g (a measure of the spatial extent of a polymer) as well as the sum of the fluorescence intensity.

In our image analysis, the positions of DNA-objects were automatically determined from three-dimensional z-stacks followed by a manual curation step (Figure 2.3a-object detection). Objects were then segmented in cube-shaped crops centered at each object's center of mass. The DNA objects were further segmented from background within these cubes based on a globally (within the cube) determined threshold,[40] yielding a 3-dimensional foreground mask containing only the DNA object, and a minimal amount of background (Figure 2.3a-segmentation and Figure S2.3b). Masks determined on the individual crops were registered within the full field-of-view volume resulting in a labeled image. Individual masks were additionally checked in a curation step and manually adjusted if upon visual inspection they did not contain single objects or did not mask objects in their entirety. Sum intensity was calculated as the total sum of all pixel intensities within a foreground mask and the radius of gyration was calculated by squaring the sum of all foreground pixels' intensity-weighted distances from the object's center of mass (Figure 2.3b).[41]

In order to monitor the integrity of the extracted chromosomes at various steps of the protocol, we measured the total per-object fluorescence intensity, *i.e.*, the sum of the intensities across all layers of the z-stack. While the sum intensity of a DNA object is expected to be set by the number of DNA basepairs, the measured distributions appeared to be fairly broad. In order to best compare the distributions before and after protein removal, we scaled the sum intensity values of each distribution with the mean value. We assume that the points in the 'before'-distributions (before protein removal) in Figure 2.3e and 2.3g represented those of intact chromosomes. This appears to be is a reasonable assumption since we observed similarly broad distributions of the sum intensity for lambda (λ)-DNA molecules (Figure S2.5).

To estimate the fraction of chromosomes that got fragmented in the process, we counted

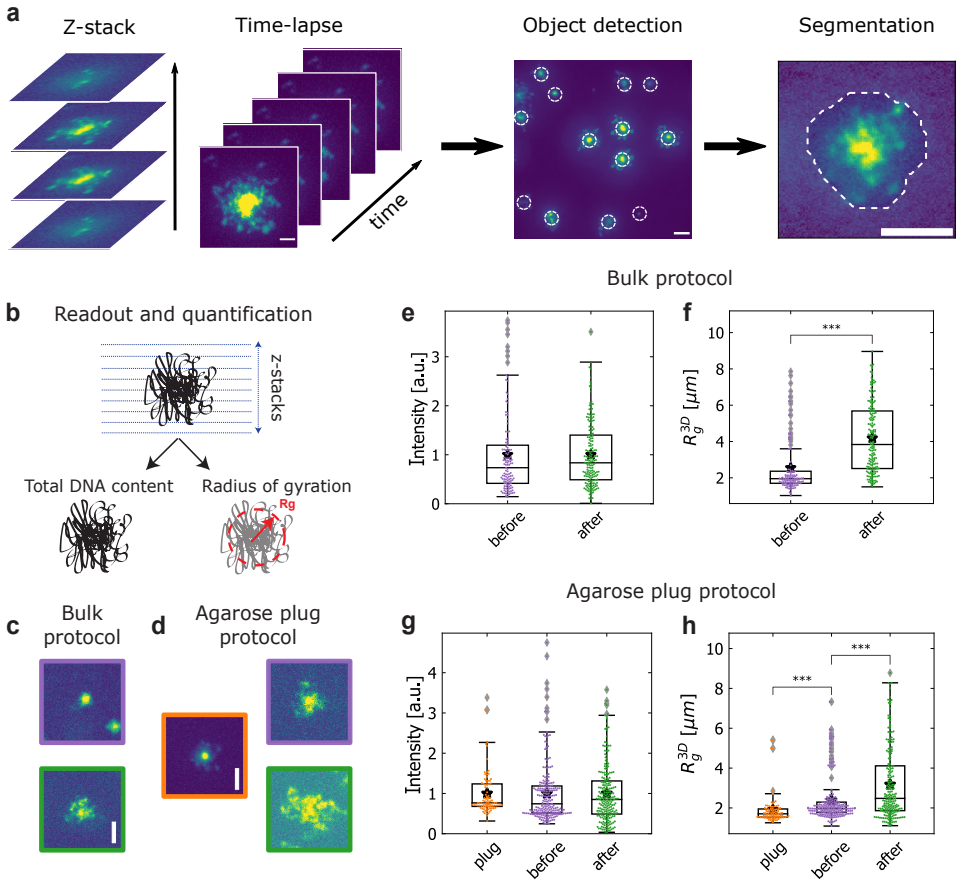

Figure 2.3: Characterization of isolated chromosomes before and after protein removal. **(a)** Image analysis workflow for a GenBox experiment. In each image, objects are detected and segmented from the background. **(b)** Within the segmentation boundary of each DNA object, the R_g and the total fluorescence intensity are calculated. **(c)** Images of typical DNA objects before (violet) and after (green) protein removal. **(d)** Images of typical DNA objects in each condition of the agarose protocol: in plug (orange), before (violet) and after (green) protein removal. **(e)** Total fluorescence intensity per DNA object before and after protein removal for the bulk protocol. **(f)** R_g distribution before and after protein removal for the bulk protocol ($p = 2.5e^{-15}$). **(g)** Total fluorescence intensity per DNA object before (in the plug and after plug melting) and after protein removal for the agarose plug protocol. **(h)** R_g distribution in the plug (orange), before protein removal but plug melting (purple) and after protein removal (green) for the agarose plug protocol ($p = 2.6e^{-5}$, $p = 5.8e^{-8}$ with independent two-sample t-test). Boxplots show the median and 25^{th}-75^{th} percentiles, thick star denotes mean. Scale bars are 5 μm. Intensity values in each distribution in panel e and g are scaled to the mean of the respective sum intensity distribution. Sample sizes in panels e and f are N=125 and 181 for before and after. Sample sizes in panels g and h are N=90, 223, 222 for plug, before, and after, respectively.

the objects in the distributions after protein removal that had a lower sum intensity value than a threshold of 1.5 times below the 25^{th} percentile of the data. For the bulk protocol, this fraction was 4 of 181 objects, while for the agarose plug protocol it was 24 of

222 objects. In other words, only a low percentage of fragmented objects of 2% and 11% was estimated for bulk and agarose plug protocol, respectively. Another indication that our observed DNA objects remain well contained in the megabasepair size range comes from comparing their sum intensities with those of λ-DNA molecules (Table S2.2). We found that the mean of the 'after' sum intensity distribution is a factor 50 (bulk protocol) or 64 (agarose plug protocol) larger than the mean of the sum intensity distribution of the 48.5 kbp long lambda-DNA molecules. Assuming that the sum intensity scales linearly with the number of basepairs, which was demonstrated previously for the dye used here in flow cytometry experiments,[42] this indicates that the DNA objects after protein removal have an average length of 2.4 Mbp (bulk protocol) and 3.1 Mbp (agarose plug protocol). However, these numbers are lower limits and the molecules are likely larger, because, following the same calculation, even the in-plug 4.6 Mbp chromosomes, which clearly are not fragmented, would be estimated to be 3.5 Mbp long.

The effect of deproteination of the extracted chromosomes is also evident from an expansion in the size of the DNA objects, which can be characterized by measuring its radius of gyration. The mean R_g in the bulk protocol increased from 2.55 ± 0.14 μm to 4.24 ± 0.14 μm (mean ± S.E.M) before and after protein removal respectively (Fig. 2.3f, Fig. S2.4a), and from 2.38 ± 0.08 μm to 3.18 ± 0.12 μm for the agarose plug protocol (Figure 2.3h and S2.4b). These results indicate that the removal of the proteins had a clear effect on the mean R_g, namely a 35% to 65% increase of the size for the agarose plug and bulk protocols (p-values $5.8e^{-8}$ and $2.5e^{-15}$), respectively. The measured radii of gyration exhibited a rather broad distribution (Figure 2.3f/h). Notably, the measured R_g values are extracted from momentarily measured snapshot images of the DNA objects, which yielded a broader distribution than the single value for the theoretical radius of gyration of a polymer which is a steady-state property.[43]

2.2.4. FIRST PROOF-OF-PRINCIPLE GENBOX EXPERIMENTS

In order to demonstrate the potential of the GenBox approach, some first example experiments were performed. First, purified protein LacI was added to chromosomes that were deproteinated with the agarose plug protocol. These fluorescently labelled proteins bind sequence-specifically to FROS arrays that were inserted near the Ori position of the chromosomes. This yielded a well-visible fluorescent spot on the isolated chromosome (Figure 2.4a-*ii*). Using a custom tracking script, the spot's locations were tracked and the mean square displacement (MSD) was computed (Figure 2.4a-*iii*). In line with the literature of local motion of chromosomal loci,[45,46] the data for this example indicate that the DNA locus moved in a sub-diffusive manner, as the MSD curve tended to plateau towards longer lag times.

For a second example, the DNA-binding protein Fis was added to deproteinated chromosomes. Figure 2.4b-*ii* shows an example of a typical DNA object before and after addition of 550 nM Fis. A significant compaction of the DNA upon Fis addition is clear. The distributions of R_g can be used to quantify the level of DNA compaction at increasing levels of added Fis (Figure 2.4b-*iii*). As the Fis levels increased from 0 nM to 550

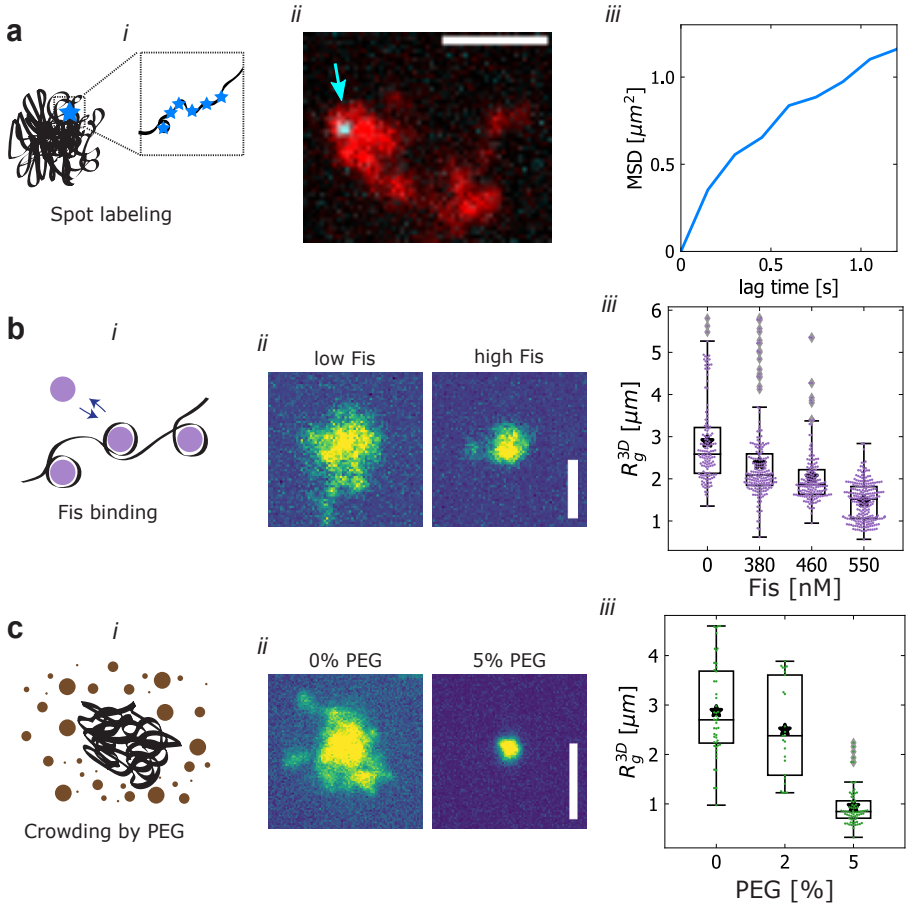

Figure 2.4: Proof-of-concept GenBox experiments. **(a)** Example of a fluorescent spot located near the Ori (cyan). Location on the isolated chromosome (red) is tracked, yielding the MSD *vs.* time (right).[44] **(b)** Fis protein is added at increasing concentrations of 380 nM and 550 nM, and the resulting compaction is observed in the shifting and narrowing distribution of R_g (right). **(c)** PEG crowding agent is added at increasing concentrations of 2% and 5% and the resulting compaction is observed from the shifting and narrowing distribution of R_g. Boxplots show the median and 25$^{\text{th}}$-75$^{\text{th}}$ percentiles, star denotes mean. Sample sizes are N=141, 201 and 242 in panel b and N=48, 25, 74 in panel c. All scale bars are 5 μm.

nM, the average R_g decreased gradually from 2.89 ± 0.08 μm to 1.47 ± 0.03 μm (mean ± S.E.M.), while the standard deviation of the distribution also decreased significantly from 1.00 μm to 0.45 μm. From recent single-molecule Atomic Force Microscopy (AFM) experiments[47] it was observed that Fis induces a strong global compaction of ~30% and also reduced the persistence length by ~20% (at a 1:40 protein:bp ratio). This compacting action was achieved by stabilization of loops and DNA crossovers. Our observation of a strong global DNA compaction of megabasepair DNA at a comparable protein:bp

ratio (1:10) are consistent with these AFM experiments.

For a final example, the crowding agent PEG was added at increasing concentrations to deproteinated chromosomes. A pronounced compaction was observed, when adding 5% PEG (Figure 2.4c-*ii*), consistent with previous reports.[48,49] The increase of PEG from 0% to 2% resulted in the mean R_g decreasing slightly from 2.87 ± 0.14 μm to 2.5 ± 0.2 μm, while the standard deviation remained steady at around 0.95 μm. However, at 5% PEG the mean and standard deviation of the R_g distribution dropped to 0.95 ± 0.05 μm and 0.39 μm, respectively (Figure 2.4c-*iii*).

2.3. DISCUSSION

In this paper, we present a methodology to prepare megabasepair deproteinated DNA, characterized the resulting DNA objects, and we provide first proof-of principle experiments to illustrate the utility of the method. The work expands on previous *in vitro* studies of large DNA molecules.[29,32–34,50,51] For example, Wegner *et al.*[30,49] and Cunha *et al.*[39,52] studied bacterial chromosomes directly after isolation from cells in an aqueous solution, while Pelletier *et al.*[48] used microfluidic devices to perform cell lysis on-chip in cell-sized channels for studying the compaction of DNA with crowding agents. A limitation of these interesting first studies was that the megabasepair DNA substrates still contained an unknown number of natively bound proteins. Our GenBox protocol builds upon these previous experiments by explicitly removing the proteins and characterizing the remaining protein content with mass spectrometry and quantitative fluorescence imaging.

We presented two variants to prepare the deproteinated DNA sample, namely the bulk protocol and the agarose plug protocol. From a practical point of view, the agarose plug protocol has some advantages compared to the bulk protocol. First, samples can be made in advance and stored until needed for further processing. Secondly, unlike the bulk protocol sample, the agarose plugs are compatible with protocols that necessitate washing steps. One the other hand, the main advantage of the bulk protocol is the lower number of experimental steps. Our mass spectrometry data (Table 2.1) showed that the deproteinated chromosomes of the bulk protocol contained fewer remaining DNA-binding proteins than those resulting from the agarose plug sample (0% vs 3%). Additionally, the bulk protocol results in a lower amount of fragmentation compared to the agarose plug protocol (as 98% vs. 89% of DNA objects classified as intact after protein removal). Since long DNA is easily sheared, it is important to limit the number of pipetting steps of DNA in solution. For both the bulk and agarose plug protocol, there is one major pipetting step involving the long DNA, namely the transfer to the observation well before the protein removal treatment. Conducting the chromosome extraction and protein removal inside a microfluidic chip could possibly eliminate this single pipetting step to further increase the number of intact DNA objects.

Modelling would be welcome to describe the observed radius of gyration of the deproteinated chromosomes. Polymer models connect the DNA contour length to a radius

of gyration R_g of the polymer blob that it forms in solution, but a broad spectrum of model variants that have been reported in literature yielded widely ranging values for R_g. Indeed, how the theoretical R_g scales with polymer length depends on multiple external parameters.[43] These include, but are not limited to experimental parameters such as the fluorescent dyes,[53] buffer salts[54,55] and divalent cations,[56–59] which set the solvent conditions and the resulting self-avoidance/attraction of the polymer, as well branches in the form of supercoils, the DNA topology of linear *vs.* circular polymers, *etc.* Variation of these factors can yield very different predicted values for R_g ranging from 1 to 6 μm for 4.6 Mbp DNA, as illustrated in Table S2.3. The values of R_g that we observed in our experiments fall within this range. Notably, bacterial chromosomes may be natively supercoiled.[60] While the removal of supercoil-stabilizing proteins as well as potential local nicks in the DNA will likely reduce the level of supercoiling significantly, some degree of supercoiling may remain in the DNA objects that result from the protocol.

We hope that the results presented in this paper open a way to start GenBox experiments that may subsequently provide a valuable bottom-up approach to the field of chromosome organization. Promising avenues may include encapsulation of megabasepair DNA inside droplets or liposomes to study the effects of spatial confinement, addition of loop extruding proteins such as cohesin or condensin to elucidate the effect of loop formation on the structure of large DNA substrates, and experiments with phase-separating DNA-binding proteins to observe the effects of polymer-mediated phase separation at long length scales.

2.4. LIMITATIONS OF STUDY

While we established and characterized two related strategies to isolate megabasepair deproteinated DNA, the approach inevitably also has limitations. First, while we reduced the number of pipetting steps in the protocols to a single one, this final slow pipetting step may still lead to unwanted DNA damage due to mechanical shearing. Indeed, the isolated megabasepair DNA blobs may contain single- and double- stranded DNA breaks, which also may result in an unknown residual level of supercoiling. Second, due to liquid motion, it proved challenging to track the objects through time in the 3D time-resolved imaging of isolated DNA objects in bulk volume. We were therefore unable to link initial state to a state at some later time during the experiment on object-per-object basis. This disadvantage may be solved by using microfabricated chambers.

Regarding the presence of residual ribosomal subunits after deproteination (Table 2.3), we can make the following comment. Although previous studies with chromosomes isolated by osmotic shock (in absence of protein removal) did not observe any difference in chromosome conformations in the presence or absence of RNase,[30] we opted to perform the RNAse treatment, for which we doubled the supplier's treatment time and used a 100-fold higher amount than the lowest recommended concentration. We suspect that any remaining ribosomal proteins may aggregate and become non-specifically trapped in the agarose matrix, later eluting with fragments of digested agarose.

One might consider the addition of DNase in the protocol for MS sample preparation, in order to ensure that tightly bound proteins would also reach the mass spectrometer. We did not adopt this approach for multiple reasons. Firstly, every enzymatic step reduces the sensitivity of the mass spectrometry quantification by the introduction of additional protein species. Secondly, DNase I treatment has been reported to introduce bias in protein-abundance patterns, and is therefore advised against.[61] Finally, under the used conditions (buffers, incubation time, dilution of crowding) it is unlikely that a protein species would remain bound to DNA so strongly that virtually none of the molecules would dissociate into solution.

2.5. MATERIALS AND METHODS

2.5.1. RESOURCE AVAILABILITY

DATA AND CODE AVAILABILITY

Data reported in this paper will be shared by the lead contact upon reasonable request. The Python code used throughout the analysis has been deposited on Zenodo (DOI: 10.5281/zenodo.6677094). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

2.5.2. METHODS DETAILS

PREPARATION OF SPHEROPLASTS AND IMAGING OF CELLS AND ORI/TER RATIO

E. coli bacterial cells (HupA-mYPet frt, Ori1::lacOx240 frt, ter3::tetOx240 gmR, ΔgalK::tetR-mCerulean frt, ΔleuB::lacI-mCherry frt, DnaC::mdoB::kanR frt)[62] were incubated from glycerol stock in M9 minimal media (1x M9 minimal salts, 0.01% v/v protein hydrolysate amicase, 0.8% glycerol, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$) supplemented with 50 μg/mL Kanamycin antibiotic (K1876, Sigma-Aldrich) in a shaking incubator at 30 °C and 300 rpm and allowed to reach OD_{600} of 0.1 to 0.15. The cells were then grown for 2 to 2.5 hours at 41 °C shaking at 900 rpm in order to arrest replication initiation.

In order to determine the Ori/Ter ratio, 1.25 μL cells were deposited on a cover slip (15707592, Thermo Fischer) and covered with an agarose pad. The cells were imaged with a Nikon Ti2-E microscope with a 100X CFI Plan Apo Lambda Oil objective with an NA of 1.45 and SpectraX LED (Lumencor) illumination system using phase contrast, cyan (CFP filter cube $\lambda_{ex}/\lambda_{bs}/\lambda_{em}$ = 426–446/455/460–500 nm), yellow (triple bandpass filter λ_{em} = 465/25–545/30–630/60 nm) and red (the same triple bandpass filter). Spots corresponding to Ori and Ter were identified on the red and cyan channels and counted either manually or with an automated routine, producing the same results.

Next, appropriate volume of cell culture was spun down at 10000 g for 2.5 min, in order to obtain a pellet at OD_{eq} = 1 (approx. 8 x 10^8 cells). The pellet was resuspended in

475 μL cold (4 °C) sucrose buffer (0.58 M sucrose, 10 mM Sodium Phosphate pH 7.2, 10 mM NaCl, 100 mM NaCl). 25 μL lysozyme (L6876 Sigma-Aldrich, 1 mg/mL in ultrapure water) was immediately added and gently mixed into the cell/sucrose buffer suspension, followed by either *i)* 15 min incubation at room temperature (bulk protocol) or *ii)* a 10 min incubation at room temperature and a 5 min incubation at 42 °C in a heat block (agarose plug protocol). The lysozyme digests the cell wall, resulting in spheroplasts.

PREPARATION OF ISOLATED CHROMOSOMES (BULK PROTOCOL)

Spheroplasts were prepared as described above. Cell lysis and nucleoid release was achieved by pipetting 10 μL of spheroplasts into 1 mL of lysis buffer (20 mM Tris-HCl pH 8) with a cut pipette tip, after which the tube was once gently inverted for mixing. Immediately thereafter, buffer composition was adjusted to match the one of agarose plug protocol (50 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM EDTA pH 8.0 and 5% glycerol). After this stage, we continued to the preparation of the observation well.

PREPARATION OF ISOLATED CHROMOSOMES (AGAROSE PLUG PROTOCOL)

500 μL warmed (42 °C) spheroplast/sucrose buffer suspension was added to 500 μL warm (42 °C) agarose solution (low melting point agarose, V2831 Promega, 2% w/v in sucrose buffer) using a cut pipette tip. In the following steps, the Eppendorf tubes were kept at 42 °C to prevent gelation of the agarose solution. The spheroplast/agarose mixture was gently mixed using a cut pipette tip, and casted in volumes of 100 μL into a plug mold (Bio-Rad laboratories, Veenendaal, The Netherlands). In order to produce a larger number of agarose plugs, it proved most optimal to perform the protocol with multiple Eppendorf tubes in parallel, rather than increasing the number of cells and volumes of sucrose buffer and agarose solution used per Eppendorf tube. To solidify the agarose plugs, the plug mold was stored at 4 °C for 1 h.

The solidified agarose plugs containing spheroplasts were removed from the plug mold and added to 25 mL per plug lysis buffer (10 mM Sodium Phosphate pH 7.2, 10 mM EDTA pH 8.0, 100 μg/mL RNase-A), thereby lysing the cells and thus merely trapping the nucleoids from the spheroplasts in the agarose gel matrix. The plugs were incubated tumbling in the lysis buffer for 1 h. Subsequently, the plugs were removed from the lysis buffer and each plug was stored in 2 mL TE wash buffer (20 mM Tris-HCl pH 8, 50 mM EDTA pH 8.0) at 4 °C until further use.

In order to transfer agarose plugs from one container to another, a sheet of aluminum foil was put over the top of a glass beaker. Using a 200 μL pipette tip holes were punched into the aluminum foil and the foil was gently pressed down into a concave shape to prevent liquid spilling over the edge. The container containing the plugs was emptied through the strainer into the beaker, leaving the agarose plugs behind on the strainer. Using flat-headed tweezers the agarose plugs were transferred to the new container. To prevent cross-contamination, the tweezers were washed after each handing step with 70% ethanol and dried using a pressurized air gun.

For releasing the purified chromosomes from the agarose plugs for experiments, agarose plugs were incubated for 1 hour in buffer A (50 mM Tris-HC pH 8, 50 mM NaCl, 1 mM EDTA pH 8.0, 5% glycerol) and then transferred to 150 μL of buffer A preheated to 71 °C. The plug was then melted at 71 °C for 15 minutes before equilibrating at 42 °C. The agarose was digested by 1 hour incubation at 42 °C with 2 units of beta-agarase (M0392, New England Biolabs). After this stage, we continued to the preparation of the observation well.

IMAGING OF SPHEROPLASTS AND CHROMOSOMES INSIDE THE AGAROSE PLUG

A plug containing spheroplasts was deposited on a KOH-cleaned cover slip. Spheroplasts were imaged with a Nikon Ti2-E microscope with a 100X CFI Plan Apo Lambda Oil objective with an NA of 1.45 and SpectraX LED (Lumencor) illumination system using the channels phase contrast, cyan (CFP filter cube $\lambda_{ex}/\lambda_{bs}/\lambda_{em}$ = 426–446/455/460–500 nm), yellow (triple bandpass filter λ_{em} = 465/25–545/30–630/60 nm) and red (the same triple bandpass filter). The imaging protocol was composed of a single time-point, using a 2 μm z-stack with 200 nm *z*-slices.

For imaging chromosomes after lysing the spheroplasts, a nucleoid-containing plug was incubated in 2 mL buffer A (50 mM Tris-HC pH 8, 50 mM NaCl, 1 mM EDTA pH 8.0, 5% glycerol) at 4 °C for 1 h. The plug was transferred to 2 mL imaging buffer (50 mM Tris-HC pH 8, 50 mM NaCl, 1 mM EDTA pH 8.0, 5% glycerol, 3.5 mM $MgCl_2$, 1 mM DTT, 500 nM Sytox Orange) and incubated for 15 min. Then the plug was deposited on a KOH-cleaned cover slip and 30 μL imaging buffer was added onto the plug to prevent drying. The plug was imaged using an Andor Spinning Disk Confocal microscope with a 100X oil immersion objective, 20% 561 laser, filters, 250x gain, and 10 ms exposure. The imaging protocol resulted in 30 μm *z*-stacks with 250 nm *z*-slices and was repeated at 15 distinct *xy* positions.

TREATMENT WITH PROTEINASE K FOR PROTEIN REMOVAL

Thermolabile Proteinase K (P8111S, New England Biolabs) was added to isolated chromosomes (0.01 unit per 1 μL of nucleoid suspension) in buffer containing 2.5 mM $MgCl_2$ and 50 mM NaCl. The samples were then incubated for 15 minutes at 37 °C for treatment and for 10 minutes at 56 °C for Proteinase K inactivation. The samples were equilibrated to RT for at least 30 minutes before imaging or and further experiments.

MASS SPECTROMETRY

Bulk and agarose plug samples were treated with Proteinase K as described above. Each sample contained nucleoids from an amount of cells corresponding to OD 5.0 (ca. $5x10^9$ cells in 100 μL). With two different DNA isolation approaches (bulk and agarose plug) and two conditions (control and Proteinase K), four triplicate samples were analyzed (twelve samples in total) by mass spectrometry. The control sample underwent exactly the same steps as the treated sample, but equal volume of 50 % glycerol (corresponding

to Proteinase K storage buffer concentration) was used instead of Proteinase K enzyme. 200 mM ammonium bicarbonate buffer (ABC) was prepared by dissolving ammonium bicarbonate powder (A6141, Sigma-Aldrich) in LC-MS grade quality water. 10 mM DTT (43815, Sigma-Aldrich) and iodoacetamide (IAA) (I1149, Sigma-Aldrich) solutions were made fresh by dissolving stock powders in 200 mM ABC. Next, 25 μL of 200 mM ABC buffer was added to each sample to adjust pH, immediately followed by addition of 30 μL of 10 mM DTT and 1 hour incubation at 37 °C and 300 rpm. Next, 30 µL of 20 mM IAA was added and samples were incubated in dark at room temperature for 30 min. Finally, 10 μL of 0.1 mg/mL trypsin (V5111, Promega) was added and samples were incubated overnight at 37 °C and 300 rpm.

On the following day, samples were purified by solid phase extraction (SPE). SPE cartridges (Oasis HLB 96-well μElution plate, Waters, Milford, USA) were washed with 700 μL of 100% methanol and equilibrated with 2x500 μL LC-MS grade water. Next, 200 μL of each sample was loaded to separate SPE cartridge wells and wells were washed sequentially with 700 μL 0.1% formic acid, 500 μL of 200 mM ABC buffer and 700 μL of 5% methanol. Samples were then eluted with 200 μL 2% formic acid in 80% methanol and 200 μL 80% 10 mM ABC in methanol. Finally, each sample was collected to separate low-binding 1.5 μL tubes and speedvac dried for 1-2 hours at 55 °C. Samples were stored frozen at -20 °C until further analysis. Desalted peptides were reconstituted in 15 μL of 3% acetonitrile/0.01% trifluoroacetic acid prior to MS-analysis.

Per sample, 3 μL of protein digest was analysed using a one-dimensional shotgun proteomics approach.[63,64] Briefly, samples were analysed using a nano-liquid-chromatography system consisting of an EASY nano LC 1200, equipped with an Acclaim PepMap RSLC RP C18 separation column (50 μm x 150 mm, 2 μm, Cat. No. 164568), and a QE plus Orbitrap mass spectrometer (Thermo Fisher Scientific, Germany). The flow rate was maintained at 350 nL·min^{-1} over a linear gradient from 5% to 25% solvent B over 90 min, then from 25% to 55% over 60 min, followed by back equilibration to starting conditions. Data were acquired from 5 to 175 min. Solvent A was water containing 0.1% FA, and solvent B consisted of 80% ACN in water and 0.1% FA. The Orbitrap was operated in data-dependent acquisition (DDA) mode acquiring peptide signals from 385–1250 m/z at 70,000 resolution in full MS mode with a maximum ion injection time (IT) of 75 ms and an automatic gain control (AGC) target of 3E6. The top 10 precursors were selected for MS/MS analysis and subjected to fragmentation using higher-energy collisional dissociation (HCD). MS/MS scans were acquired at 17,500 resolution with AGC target of 2E5 and IT of 100 ms, 1.0 m/z isolation width and normalized collision energy (NCE) of 28.

PREPARATION OF OBSERVATION WELLS

Cover slips (15707592, Thermo Fischer) were loaded onto a teflon slide holder. The coverslips were sonicated in a bath sonicator in a beaker containing ultrapure water for 5 min, followed by sonication in acetone for 20 min, a rinse with ultrapure water, sonication in KOH (1 M) for 15 min, a rinse with ultrapure water, and finally sonication in methanol for 15 min. Cleaned cover slips were stored in methanol at 4 °C.

To assemble the observation well, a PDMS block with a 4 mm punched (504651 World Precision Instruments) through hole was bonded on a cleaned coverslip. PDMS block was obtained from PDMS slab of ± 5 mm thickness which was casted from mixture of 10:1 = PDMS:curing agent (Sylgard 184 Dow Corning GmbH) and allowed to cure for 4 hours at 80 °C. The bonding was done immediately after exposing both surfaces, glass and PDMS, to oxygen plasma (2 minutes at 20 W) and the bond was allowed to cure for 10 minutes at 80 °C.

Immediately after the bonding, the inner surface of the observation well was treated to create a lipid bilayer to prevent sticking of DNA and proteins. To do so, DOPC liposomes were used. DOPC and PE-CF lipids from chloroform stocks (both Avanti Polar Lipids, Inc.) were combined in 999:1 mol-ratio DOPC:PE-CF in a glass vial for final lipid concentration of 4 mg/mL. Chloroform was evaporated by slowly turning the vial in a gentle nitrogen steam for 15 minutes or until dry. The vial was then placed in a desiccator for 1 hour to further dry its contents. The lipids were then resuspended in SUV buffer (25 mM Tris-HCl pH 7.5, 150 mM KCl, 5 mM $MgCl_2$) and vortexed until solution appears opaque and homogeneous to the eye. Any large lipid aggregates were broken up by 7 to 10 freeze-thaw cycles of repeated immersion into liquid nitrogen and water at 70-90 °C. The lipid suspension was loaded in a glass syringe (250 μL, Hamilton) and extruded through 30 nm polycarbonate membrane (610002, Avanti Polar Lipids, Inc.) fixed in mini-extruder (610020, Avanti Polar Lipids, Inc.) at 40 °C. Lipids were stored at -20 °C for up to several months. SUV suspension (99.9 mol% DOPC, 0.1 mol% PE:CF - both Avanti Polar Lipids, Inc.) was sonicated for 10 minutes at RT and pipetted into the well to cover the area to be treated. After 1 minute of incubation, the solution was diluted by adding 3x fold excess off SUV buffer (25 mM Tris-HCl pH 7.5, 150 mM KCl, 5 mM $MgCl_2$). Subsequently, the solution in the well was exchange at least 5-times, without de-wetting the surface of the glass, for imaging buffer (50 mM Tris-HC pH 8, 50 mM NaCl, 1 mM EDTA pH 8.0, 5% glycerol, 3.75 mM $MgCl_2$, 1.5 mM DTT, 750 nM Sytox Orange). As final step, a sample with nucleoids from either the bulk or plug protocol was added to the imaging buffer in ratio 1:2 (nucleoids to imaging buffer), after which the well was ready for imaging.

EXPERIMENTS WITH SPOT LABELING, FIS, AND PEG

For the experiments of figure 2.4, the protocol for imaging digested plugs was followed, but with some modifications for the imaging. Plugs with ProtK protein removal treatment were used. The imaging protocol was as follows: *i)* a 30 μm *z*-stack was taken with 250 nm *z*-slices, and this was repeated at 5 *xy* positions; *ii)* a 30 μm *z*-stack was taken with 1 μm *z*-slices at 5 *xy* positions, repeated 10 times; *iii)* the protein of interest was added to the observation well at a final concentration of 1.25 nM (LacI), 380 or 550 nM (Fis), 2 or 5% (PEG-8000, Sigma Aldrich); *iv)* a 30 μm *z*-stack was taken with 1 μm *z*-slices at 5 *xy* positions, repeated 50 times. Once the compaction process reached a steady state, the imaging step *i)* was repeated.

Fis protein was a kind gift of William Nasser, and was purified as described previously.[47] 8xHis-tagged LacI-SNAP fusions in pBAD plasmids were ordered from GenScript. BL21(DE3)-competent *E.coli* cells (New England Biolabs) were transformed with the

plasmids and plated with Ampicillin (Amp). Overnight colonies were inoculated in LB with Amp and incubated overnight at 37 °C and 150 rpm. Cells were diluted 1:100 into fresh media with Amp and grown at 37 °C at 150 rpm until OD_{600} of 0.5 - 0.6 after which 2 g/L arabinose was added to induce expression for 3-4 hours. Next, cells were harvested by centrifugation and resuspended in buffer A (50 mM Tris-HCl pH 7.5, 200 mM NaCl, 5% w/v glycerol). Lysis was performed with French Press and supernatant was recovered after centrifugation. His-tagged proteins were bound to beads in talon resin and column was then in turns washed with 50 mL of buffer A1 (buffer A + 10 mM imidazole), buffer A2 (buffer A + 0.01% Tween-20), and buffer A3 (buffer A + 0.5 M NaCl). Next, the sample was eluted with 15 mL buffer B (buffer A + 3C protease + 1 mM β-Mercaptoethanol) and diluted 10x in buffer C (50 mM Tris-HCl pH 8.0). Anion exchange chromatography was done with Mono Q-ion exchange column (Cytiva) equilibrated with buffer C and sample was eluted to buffer D (50 mM Tris-HCl pH 8.0 with 1 M NaCl). Next, size exclusion chromatography was done on Superdex S200 (Cytiva) column equilibrated with buffer A, collected and fractions were run on gel to check for purity. Finally, purified proteins were labelled with SNAP-Surface Alexa Fluor 647 tag (New England Biolabs) following manufacturer's instructions.

2.5.3. QUANTIFICATION AND STATISTICAL ANALYSIS

IMAGE PROCESSING AND ANALYSIS

We developed a custom analysis pipeline for quantifying DNA objects in fluorescent images obtained from GenBox experiments, written entirely in Python. The analysis proceeds in three main steps: *i)* identification of individual DNA objects, *ii)* segmentation of these objects from background, *iii)* quantification of relevant observables (*e.g.*, a calculation of the radius of gyration).

Positions of individual objects were determined automatically from three-dimensional stacks using *skimage* function *peak_local_max*.[65] Maxima were required to be at least twice as bright as globally determined threshold[40] (see next paragraph for description). If objects' maxima were closer than 30 pixels from each other, or from any image boundary, the objects were discarded from further analysis. Next, all locations were visually inspected with *napari*'s viewer[66] using Image and Points layers. Typically, none or few changes had to be made (*e.g.*, if one object was identified as two or vice-versa).

Objects were segmented from the background in crops corresponding to 25x25x25 μm^3 centered at each object's center of mass. First, the raw data in any crop was binarized based on a globally determined threshold.[40] Pixels' intensity values were sorted increasingly, and two lines were fitted to such curve *a)* a line fitted to the first half of the pixels in the image (estimate of background), and *b)* a line fitted to all pixels brighter than half of the maximum intensity (estimate of foreground). The intensity threshold value was then determined from the point on the sorted intensity curve which was closest to intersection of the two lines (figure S2.3a). Images before and after background subtraction were inspected and confirmed that the approach was able to discriminate background

and foreground well (figure S2.3b). The crops were then traversed plane-by-plane in z-direction, discarding small regions, dilating remaining region(s) and filling holes. The mask contours were smoothed in each plane with a Savitzky-Golay filter with a window size of quarter the contour length of the mask. Finally, only the most central 3D contiguous binary object was retained as foreground mask for each object.

Masks determined on individual crops were subsequently registered within full FOV volume (typically about 100x100x100 μm^3) producing a labeled image. If shared pixels resulted at masks overlap, these pixels were assigned to the mask which center of mass was the closest. Subsequently, the masks were inspected with *napari*'s viewer using Image and Label layers and manually adjusted if upon visual inspection they did not contain single objects or did not mask those in their entirety.

The quantification of the objects' properties was done within the volume of the foreground mask applied onto the raw data after subtracting globally determined threshold (as described earlier) from each crop. Sum intensity was calculated as the total sum of all pixel intensities within a foreground mask and the radius of gyration was calculated by squaring the sum of all foreground pixels' intensity-weighted distances from the object's center of mass. The resulting measurements were saved as structured JSON files, one per each FOV, and aggregated based on condition to produce R_g and intensity plots. The MSD in spot-labeling experiment was calculated using the xy-coordinates of fluorescent spots obtained with the ImageJ TrackMate plugin.[67,68]

MASS SPECTROMETRY ANALYSIS

Mass spectrometry data were analysed against the proteome database from *Escherichia coli* (UniProt, strain K12, Tax ID: 83333, November 2021, https://www.uniprot.org/), including Proteinase K from *Parengyodontium album* (UniProt ID: P06873) and Beta-agarase I from *Pseudoalteromonas atlantica* (UniProt ID: Q59078),[69] using PEAKS Studio X+ (Bioinformatics Solutions Inc., Waterloo, Canada),[70] allowing for 20 ppm parent ion and 0.02 m/z fragment ion mass error, 3 missed cleavages, carbamidomethylation as fixed and methionine oxidation and N/Q deamidation as variable modifications. Peptide spectrum matches were filtered for 1% false discovery rates (FDR) and identifications with ≥ 1 unique peptide matches. For the case that a protein in the sample was identified by only a single peptide in only one out of three runs, the protein identification was only considered if the same peptide sequence was also identified in unpurified control (within a retention time window of ± 2 min).

For determination of relative amounts of protein remaining after Proteinase K treatment, protein abundances were expressed as 'spectral counts' normalized by their molecular weight (*i.e.*, $\frac{\text{spectral counts}}{\text{molecular weight}}$x1000). Using the normalized spectral counts per protein in the three replicate experiments per condition ('before' and 'after'), the mean was calculated for each protein individually and for the aggregated DNA-binding and non-DNA-binding categories. Uncertainties were expressed as standard deviations from the means due to inter-sample variation. Relative amounts (for individual proteins and the aggregated categories) were defined as the ratio of the 'after' over the 'before' means, with uncertainties calculated by propagating the errors through this ratio.

2.6. SUPPLEMENTARY INFORMATION

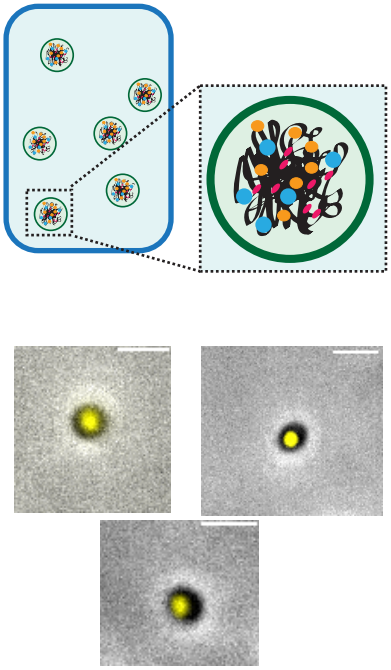

Figure S2.1: Spheroplasts in plug. Related to figure 2.2. Schematic (top) and microscopy images (bottom) of spheroplasts embedded inside an agarose plug. The yellow signal comes from fluorescently labeled HU-protein and thus serves as a DNA marker. The greyscale signal is phase contrast. Scale bars are 2 μm.

Protein	Function
rpoC	DNA-directed RNA polymerase subunit beta'
rpoB	DNA-directed RNA polymerase subunit beta
rpoA	DNA-directed RNA polymerase subunit alpha
gyrA	DNA gyrase subunit A
topA	DNA topoisomerase 1
gyrB	DNA gyrase subunit B
stpA	DNA-binding protein StpA
hupA	DNA-binding protein HU-alpha
dps	DNA protection during starvation protein
ybiB	Uncharacterized protein
Fis	DNA-binding protein Fis
cbpA	Curved DNA-binding protein

Continued on next page

Table S2.1 – *Continued from previous page*

Protein	**Function**
rpoZ	DNA-directed RNA polymerase subunit omega
polA	DNA polymerase I
hupB	DNA-binding protein HU-beta
ihfA	Integration host factor subunit alpha
ihfB	Integration host factor subunit beta
helD	DNA helicase IV
kdgR	Transcriptional regulator
uvrD	DNA helicase II
oxyR	Hydrogen peroxide-inducible genes activator
parE	DNA topoisomerase 4 subunit B
rpoS	RNA polymerase sigma factor
rpoD	RNA polymerase sigma factor
crl	Sigma factor-binding protein
yejK	Nucleoid-associated protein
ybaB	Nucleoid-associated protein
dnaE	DNA polymerase III subunit alpha
dnaA	Chromosomal replication initiator protein
ebfC	Nucleoid-associated protein
slmA	Nucleoid occlusion factor
crfC	Clamp-binding protein CrfC
mukB	Chromosome partition protein
mukF	Chromosome partition protein
matP	Macrodomain Ter protein
topo3	DNA topoisomerase
parC	DNA topoisomerase 4 subunit A
mukE	Chromosome partition protein

Table S2.1: List of DNA-binding proteins used for mass spectrometry analysis. Related to Table 2.2 and Table 2.3. Proteins' description is taken UniProt (UniProt, strain K12, Tax ID: 83333, November 2021) database. Shortlist contains proteins identified as DNA-binding or DNA processing.

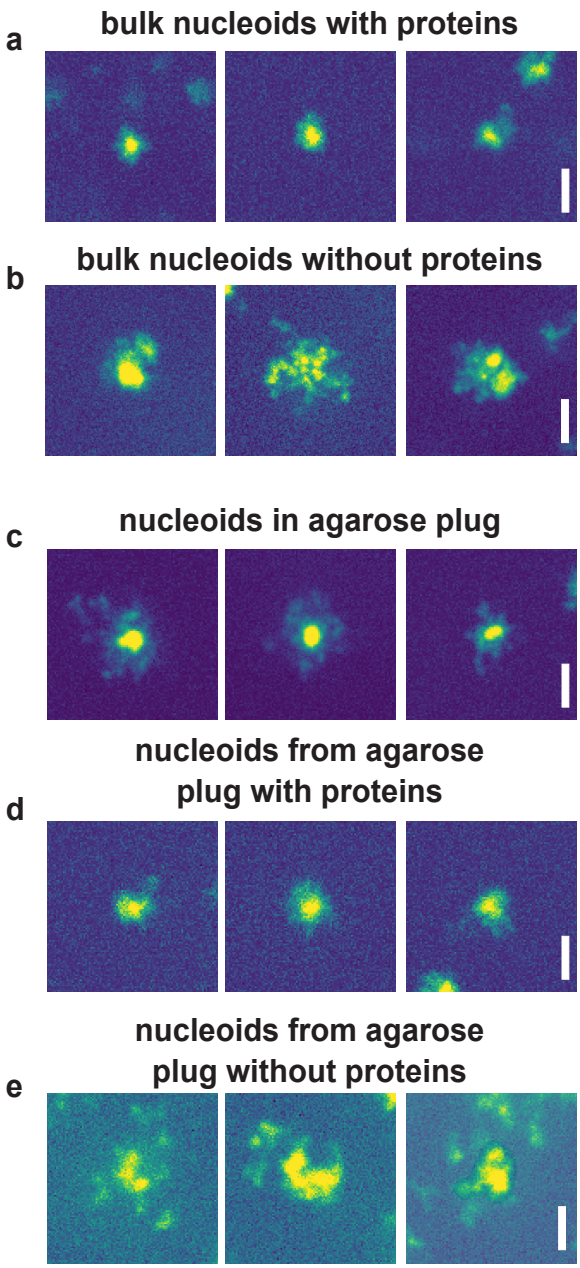

Figure S2.2: Examples of DNA-objects. Related to figure 2.2. Fluorescence images of DNA objects in various conditions: **(a)** Bulk protocol chromosomes before protein removal. **(b)** Bulk protocol chromosomes after protein removal. **(c)** Agarose plug protocol chromosomes inside the agarose plug before protein removal. **(d)** Agarose plug protocol chromosome in solution before protein removal. **(e)** Agarose plug protocol chromosomes in solution after protein removal. Scale bars are 5 μm.

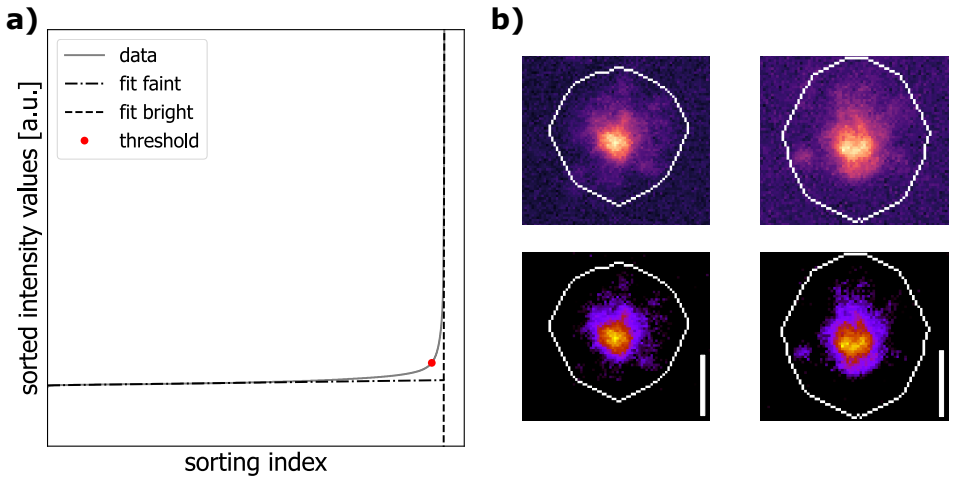

Figure S2.3: Visualization of thresholding procedure. Related to figure 2.3. **(a)** Pixel intensity values were sorted by increasing intensity, and two lines were fitted to this curve: a line fitted to the first half of the pixels in the image (which is the estimate of background, dash-dot), and a line fitted to all pixels brighter than half of the maximum intensity (estimate of foreground, dash). The intensity threshold value was then determined from the point on the sorted intensity curve (red dot) which was closest to intersection of the two lines. **(b)** Images before (top) and after (bottom) background subtraction. Inspection confirmed that the approach was able to discriminate background and foreground well. White line is contour of the mask. Scale bars are 5 μm.

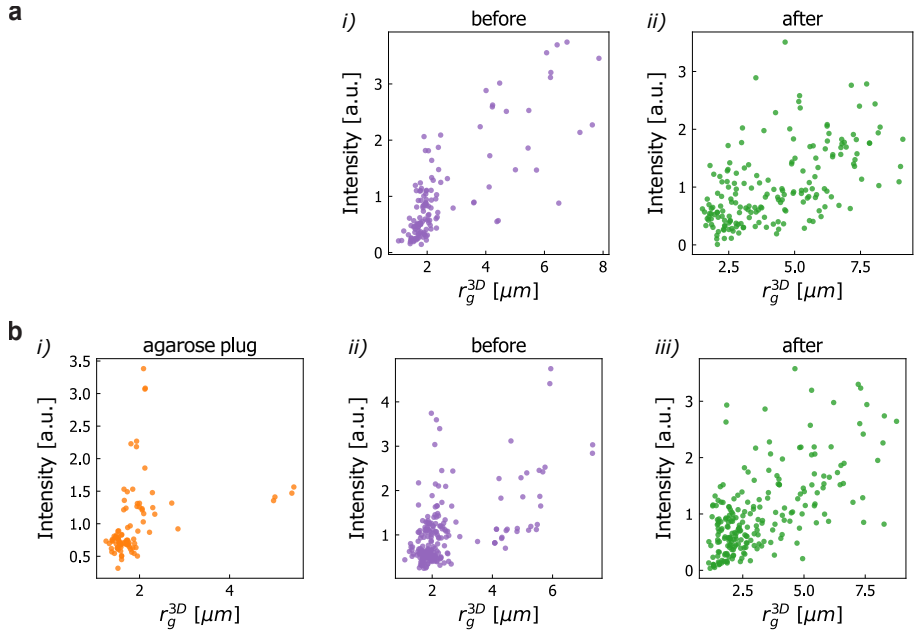

Figure S2.4: Radius of gyration versus sum intensity distributions. Related to figure 2.3. Scatter plots of the radius of gyration and sum intensity of observed DNA objects in various conditions: **(a)** *i)* Bulk protocol chromosomes before protein removal. *ii)* Bulk protocol chromosomes after protein removal. **(b)** *i)* Agarose plug protocol chromosomes inside the agarose plug before protein removal. *ii)* Plug protocol chromosome in solution before protein removal. *iii)* Agarose plug protocol chromosomes in solution after protein removal. Intensity values in each scatter plot are scaled to the mean of the applicable sum intensity distribution. Sample sizes are N=125 and 181 in panel a; and N=90, 223, 222 in panel b.

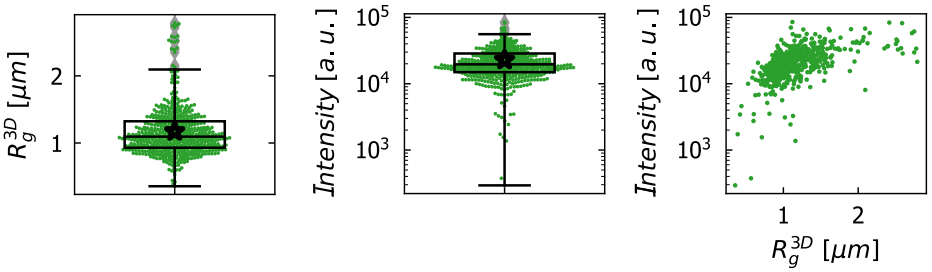

Figure S2.5: Characterization of lambda-DNA molecules. Related to figure 2.3. (left) R_g distribution for lambda DNA molecules. (center) Total fluorescence intensity per identified λ-DNA molecule. (right) R_g *vs.* total fluorescence intensity per DNA object distribution. Boxplots show the median and 25$^{\text{th}}$-75$^{\text{th}}$ percentiles, star denotes mean, N=534.

Sample	**Condition**	**Intensity (a.u.)**	**Relative intensity**	**(theoretical) number of kbp**
bulk	after	1133034	49.5	2403
plug	after	1466184	64.1	3109
plug	in plug	1658675	72.5	3518
lambda		22870	1	48.5

Table S2.2: Total and relative total intensities of DNA molecules. Related to Figure 2.3. Mean sum intensity per molecule is reported. Bulk and plug condition values of intensity are compared relative to lambda-DNA molecules, and from this an expected number of base pairs is calculated.

	Topology	**Solvent**	N = 1 Mbp, L_p = 25 nm	N = 1 Mbp, L_p = 50 nm	N = 4.6 Mbp, L_p = 25 nm	N = 4.6 Mbp, L_p = 50 nm
Ideal chain[71]	L	n.a.	1.7	2.4	3.6	5.1
	R	n.a.	1.2	1.7	2.6	3.6
Worm-like chain[71]	L	n.a.	1.7	2.4	3.6	5.1
	R	n.a.	0.8	1.2	1.8	2.6
Self-avoiding polymer with solvent interaction (Flory theory)[71]	R	good	2.6	3.4	6.3	8.4
	R	ideal	1.2	1.7	2.6	3.6
	R	poor	0.35	0.54	0.6	0.9
Self-avoiding polymer with solvent interaction (Flory theory)[71]	L	good	3.7	4.9	9.0	12.0
	L	ideal	1.7	2.4	3.6	5.1
	L	poor	0.5	0.76	0.8	1.3

Continued on next page

Table S2.3 – *Continued from previous page*

Non-crosslinked supercoiled polymer[39,72]	L/C	n.a.	1.35	0.83	1.5	2.5

Table S2.3: Gyration radii (μm) for various length DNA and various persistence length values. Related to Figure 2.3. The persistence length of bare DNA is commonly 50 nm. However buffer conditions (*e.g.*, high concentrations of mono- and di-valent ions, as well as varying concentrations of intercalating dyes) can substantially decrease it. At conditions used in this study we do not expect persistence lengths lower than 25 nm.[73–75] Topology: L - linear, R - ring. Solvent: good - $\nu = 0.588$, ideal - $\nu = 0.5$, poor - $\nu = 0.36$.

REFERENCES

(1) Schwille, P. Jump-starting life? Fundamental aspects of synthetic biology. *The Journal of Cell Biology* **2015**, *210*, 687–690.

(2) Litschel, T.; Ramm, B.; Maas, R.; Heymann, M.; Schwille, P. Beating Vesicles: Encapsulated Protein Oscillations Cause Dynamic Membrane Deformations. *Angewandte Chemie International Edition* **2018**, *57*, 16286–16290.

(3) Ganzinger, K. A.; Merino-Salomón, A.; García-Soriano, D. A.; Butterfield, A. N.; Litschel, T.; Siedler, F.; Schwille, P. FtsZ Reorganization Facilitates Deformation of Giant Vesicles in Microfluidic Traps. *Angewandte Chemie International Edition* **2020**, *59*, 21372–21376.

(4) Litschel, T.; Kelley, C. F.; Holz, D.; Adeli Koudehi, M.; Vogel, S. K.; Burbaum, L.; Mizuno, N.; Vavylonis, D.; Schwille, P. Reconstitution of contractile actomyosin rings in vesicles. *Nature Communications* **2021**, *12*, 2254.

(5) Joesaar, A.; Yang, S.; Bögels, B.; van der Linden, A.; Pieters, P.; Kumar, B. V. V. S. P.; Dalchau, N.; Phillips, A.; Mann, S.; de Greef, T. F. A. DNA-based communication in populations of synthetic protocells. *Nature Nanotechnology* **2019**, *14*, 369–378.

(6) Bintu, B.; Mateo, L. J.; Su, J.-H.; Sinnott-Armstrong, N. A.; Parker, M.; Kinrot, S.; Yamaya, K.; Boettiger, A. N.; Zhuang, X. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science (New York, N.Y.)* **2018**, *362*, eaau1783.

(7) Ricci, M. A.; Manzo, C.; García-Parajo, M. F.; Lakadamyali, M.; Cosma, M. P. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **2015**, *160*, 1145–58.

(8) Falk, M.; Feodorova, Y.; Naumova, N.; Imakaev, M.; Lajoie, B. R.; Leonhardt, H.; Joffe, B.; Dekker, J.; Fudenberg, G.; Solovei, I.; Mirny, L. A. Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **2019**, *570*, 395–399.

(9) Brandão, H. B.; Ren, Z.; Karaboja, X.; Mirny, L. A.; Wang, X. DNA-loop-extruding SMC complexes can traverse one another in vivo. *Nature Structural & Molecular Biology 2021 28:8* **2021**, *28*, 642–651.

(10) Liang, Y.; van der Valk, R. A.; Dame, R. T.; Roos, W. H.; Wuite, G. J. L. Probing the mechanical stability of bridged DNA-H-NS protein complexes by single-molecule AFM pulling. *Scientific Reports* **2017**, *7*, 15275.

(11) Dame, R. T.; Wyman, C.; Goosen, N. H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Research* **2000**, *28*, 3504–3510.

(12) Japaridze, A.; Muskhelishvili, G.; Benedetti, F.; Gavriilidou, A. F. M.; Zenobi, R.; De Los Rios, P.; Longo, G.; Dietler, G. Hyperplectonemes: A Higher Order Compact and Dynamic DNA Self-Organization. *Nano Letters* **2017**, *17*, 1938–1948.

(13) Kaczmarczyk, A.; Meng, H.; Ordu, O.; Noort, J. v.; Dekker, N. H. Chromatin fibers stabilize nucleosomes under torsional stress. *Nature Communications* **2020**, *11*, 126.

(14) Sun, M.; Nishino, T.; Marko, J. F. The SMC1-SMC3 cohesin heterodimer structures DNA through supercoiling-dependent loop formation. *Nucleic Acids Research* **2013**, *41*, 6149–6160.

(15) Renger, R.; Morin, J. A.; Lemaitre, R.; Ruer-Gruss, M.; Jülicher, F.; Hermann, A.; Grill, S. W. Co-condensation of proteins with single- and double-stranded DNA. *Proceedings of the National Academy of Sciences of the United States of America* **2022**, *119*, DOI: 10.1073/pnas.2107871119.

(16) Lin, S. N.; Dame, R. T.; Wuite, G. J. Direct visualization of the effect of DNA structure and ionic conditions on HU–DNA interactions. *Scientific Reports 2021 11:1* **2021**, *11*, 1–10.

(17) Davidson, I. F.; Bauer, B.; Goetz, D.; Tang, W.; Wutz, G.; Peters, J.-M. M. DNA loop extrusion by human cohesin. *Science (New York, N.Y.)* **2019**, *366*, 1338–1345.

(18) Golfier, S.; Quail, T.; Kimura, H.; Brugués, J. Cohesin and condensin extrude DNA loops in a cell-cycle dependent manner. *eLife* **2020**, *9*, e53885.

(19) Ganji, M.; Shaltiel, I. A.; Bisht, S.; Kim, E.; Kalichava, A.; Haering, C. H.; Dekker, C. Real-time imaging of DNA loop extrusion by condensin. *Science (New York, N.Y.)* **2018**, *360*, 102–105.

(20) Kim, Y.; Shi, Z.; Zhang, H.; Finkelstein, I. J.; Yu, H. Human cohesin compacts DNA by loop extrusion. *Science (New York, N.Y.)* **2019**, *366*, 1345–1349.

(21) Greene, E. C.; Wind, S.; Fazio, T.; Gorman, J.; Visnapuu, M.-L., DNA Curtains for High-Throughput Single-Molecule Optical Imaging In *Methods in Enzymology*; Elsevier Inc.: 2010; Chapter 14, pp 293–315.

(22) Birnie, A.; Dekker, C. Genome-in-a-Box: Building a Chromosome from the Bottom Up. *ACS Nano* **2021**, *15*, 111–124.

(23) Yoo, H.-B.; Lim, H.-M.; Yang, I.; Kim, S.-K.; Park, S.-R. Flow cytometric investigation on degradation of macro-DNA by common laboratory manipulations. *Journal of Biophysical Chemistry* **2011**, *02*, 102–111.

(24) Adam, R. E.; Zimm, B. H. Shear degradation of DNA. *Nucleic Acids Research* **1977**, *4*, 1513–1538.

(25) Vanapalli, S. A.; Ceccio, S. L.; Solomon, M. J. Universal scaling for polymer chain scission in turbulence. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 16660–16665.

(26) Kaur, G.; Lewis, J.; van Oijen, A. Shining a Spotlight on DNA: Single-Molecule Methods to Visualise DNA. *Molecules* **2019**, *24*, 491.

(27) Dufrêne, Y. F.; Ando, T.; Garcia, R.; Alsteens, D.; Martinez-Martin, D.; Engel, A.; Gerber, C.; Müller, D. J. Imaging modes of atomic force microscopy for application in molecular and cell biology. *Nature Nanotechnology* **2017**, *12*, 295–307.

(28) Kriegel, F.; Ermann, N.; Lipfert, J. Probing the mechanical properties, conformational changes, and interactions of nucleic acids with magnetic tweezers. *Journal of Structural Biology* **2017**, *197*, 26–36.

(29) Shintomi, K.; Takahashi, T. S.; Hirano, T. Reconstitution of mitotic chromatids with a minimum set of purified factors. *Nature Cell Biology* **2015**, *17*, 1014–1023.

(30) Wegner, A. S.; Alexeeva, S.; Odijk, T.; Woldringh, C. L. Characterization of Escherichia coli nucleoids released by osmotic shock. *Journal of Structural Biology* **2012**, *178*, 260–269.

(31) Pelletier, J.; Jun, S. Isolation and Characterization of Bacterial Nucleoids in Microfluidic Devices. *Methods in Molecular Biology* **2017**, *1624*, 311–322.

(32) Lartigue, C.; Glass, J. I.; Alperovich, N.; Pieper, R.; Parmar, P. P.; Hutchison, C. A.; Smith, H. O.; Venter, J. C. Genome transplantation in bacteria: Changing one species to another. *Science (New York, N.Y.)* **2007**, *317*, 632–638.

(33) Zhang, M.; Zhang, Y.; Scheuring, C. F.; Wu, C. C.; Dong, J. J.; Zhang, H. B. Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. *Nature Protocols 2012 7:3* **2012**, *7*, 467–478.

(34) Łopacińska-Jørgensen, J. M.; Pedersen, J. N.; Bak, M.; Mehrjouy, M. M.; Sørensen, K. T.; Østergaard, P. F.; Bilenberg, B.; Kristensen, A.; Taboryski, R. J.; Flyvbjerg, H.; Marie, R.; Tommerup, N.; Silahtaroglu, A. Enrichment of megabase-sized DNA molecules for single-molecule optical mapping and next-generation sequencing. *Scientific Reports* **2017**, *7*, 1–10.

(35) Merrikh, H.; Zhang, Y.; Grossman, A. D.; Wang, J. D. Replication–transcription conflicts in bacteria. *Nature Reviews Microbiology 2012 10:7* **2012**, *10*, 449–458.

(36) Bird, R. E.; Louarn, J.; Martuscelli, J.; Caro, L. Origin and sequence of chromosome replication in Escherichia coli. *Journal of Molecular Biology* **1972**, *70*, 549–566.

(37) Saifi, B.; Ferat, J. L. Replication Fork Reactivation in a dnaC2 Mutant at Non-Permissive Temperature in Escherichia coli. *PLOS ONE* **2012**, *7*, e33613.

(38) Japaridze, A.; Gogou, C.; Kerssemakers, J. W.; Nguyen, H. M.; Dekker, C. Direct observation of independently moving replisomes in Escherichia coli. *Nature Communications 2020 11:1* **2020**, *11*, 1–10.

(39) Cunha, S.; Woldringh, C. L.; Odijk, T. Polymer-Mediated Compaction and Internal Dynamics of Isolated Escherichia Coli Nucleoids. *Journal of Structural Biology* **2001**, *136*, 53–66.

(40) Vtyurina, N. What makes long DNA short? Modulation of DNA structure by Dps protein: cooperating & reorganizing, Ph.D. Thesis, Delft University of Technology, 2016, p 184.

(41) Strychalski, E. A.; Geist, J.; Gaitan, M.; Locascio, L. E.; Stavis, S. M. Quantitative measurements of the size scaling of linear and circular DNA in nanofluidic slitlike confinement. *Macromolecules* **2012**, *45*, 1602–1611.

(42) Yan, X.; Habbersett, R. C.; Yoshida, T. M.; Nolan, J. P.; Jett, J. H.; Marrone, B. L. Probing the kinetics of SYTOX Orange stain binding to double-stranded DNA with implications for DNA analysis. *Analytical Chemistry* **2005**, *77*, 3554–3562.

(43) De Gennes, P. G., *Scaling Concepts in Polymer Physics*; Cornell University Press: 1979.

(44) Vink, J. N.; Brouns, S. J.; Hohlbein, J. Extracting Transition Rates in Particle Tracking Using Analytical Diffusion Distribution Analysis. *Biophysical Journal* **2020**, *119*, 1970–1983.

(45) Weber, S. C.; Spakowitz, A. J.; Theriot, J. A. Bacterial Chromosomal Loci Move Subdiffusively through a Viscoelastic Cytoplasm. *Physical Review Letters* **2010**, *104*, 238102.

(46) Javer, A.; Kuwada, N. J.; Long, Z.; Benza, V. G.; Dorfman, K. D.; Wiggins, P. A.; Cicuta, P.; Lagomarsino, M. C. Persistent super-diffusive motion of Escherichia coli chromosomal loci. *Nature Communications* **2014**, *5*, 3854.

(47) Japaridze, A.; Yang, W.; Dekker, C.; Nasser, W.; Muskhelishvili, G. DNA sequence-directed cooperation between nucleoid-associated proteins. *iScience* **2021**, *24*, 102408.

(48) Pelletier, J.; Halvorsen, K.; Ha, B.-Y.; Paparcone, R.; Sandler, S. J.; Woldringh, C. L.; Wong, W. P.; Jun, S. Physical manipulation of the Escherichia coli chromosome reveals its soft nature. *Proceedings of the National Academy of Sciences* **2012**, *109*, E2649–E2656.

(49) Wegner, A. S.; Wintraecken, K.; Spurio, R.; Woldringh, C. L.; de Vries, R.; Odijk, T. Compaction of isolated Escherichia coli nucleoids: Polymer and H-NS protein synergetics. *Journal of Structural Biology* **2016**, *194*, 129–137.

(50) Shintomi, K.; Inoue, F.; Watanabe, H.; Ohsumi, K.; Ohsugi, M.; Hirano, T. Mitotic chromosome assembly despite nucleosome depletion in Xenopus egg extracts. *Science* **2017**, *356*, 1284–1287.

(51) Shintomi, K. Making Mitotic Chromosomes in a Test Tube. *Epigenomes* **2022**, *6*, 20.

(52) Cunha, S.; Woldringh, C. L.; Odijk, T. Restricted diffusion of DNA segments within the isolated Escherichia coli nucleoid. *Journal of structural biology* **2005**, *150*, 226–32.

(53) Japaridze, A.; Benke, A.; Renevey, S.; Benadiba, C.; Dietler, G. Influence of DNA binding dyes on bare DNA structure studied with atomic force microscopy. *Macromolecules* **2015**, *48*, 1860–1865.

(54) Cruz-León, S.; Vanderlinden, W.; Müller, P.; Forster, T.; Staudt, G.; Lin, Y.-Y.; Lipfert, J.; Schwierz, N. Twisting DNA by salt. *Nucleic Acids Research* **2022**, *50*, 5726–5738.

(55) Schlick, T.; Li, B.; Olson, W. K. The influence of salt on the structure and energetics of supercoiled DNA. *Biophysical Journal* **1994**, *67*, 2146–2166.

(56) Thomas, G. J.; Benevides, J. M.; Duguid, J.; Bloomfield, V. A. Roles of Cations in the Structure, Stability and Condensation of DNA. *Fifth International Conference on the Spectroscopy of Biological Molecules* **1993**, 39–45.

(57) Srivastava, A.; Timsina, R.; Heo, S.; Dewage, S. W.; Kirmizialtin, S.; Qiu, X. Structure-guided DNA–DNA attraction mediated by divalent cations. *Nucleic Acids Research* **2020**, *48*, 7018–7026.

(58) Bloomfield, V. A. DNA Condensation by Multivalent Cations. *Biopolymers* **1998**, *44*, 269–282.

(59) Hagerman, P. J. FLEXIBILITY OF DNA. *Ann. Rev. Biophys. Biophys. Chem* **1988**, *17*, 265–86.

(60) Kavenoff, R.; Bowen, B. C. Electron microscopy of membrane-free folded chromosomes from Escherichia coli. *Chromosoma* **1976**, *59*, 89–101.

(61) Acosta-Martin, A. E.; Chwastyniak, M.; Beseme, O.; Drobecq, H.; Amouyel, P.; Pinet, F. Impact of incomplete DNase I treatment on human macrophage proteome analysis. *Proteomics - Clinical Applications* **2009**, *3*, 1236–1246.

(62) Wu, F.; Japaridze, A.; Zheng, X.; Wiktor, J.; Kerssemakers, J. W. J.; Dekker, C. Direct imaging of the circular chromosome in a live bacterium. *Nature Communications* **2019**, *10*, 2194.

(63) Köcher, T.; Pichler, P.; Swart, R.; Mechtler, K. Analysis of protein mixtures from whole-cell extracts by single-run nanoLC-MS/MS using ultralong gradients. *Nature Protocols 2012 7:5* **2012**, *7*, 882–890.

(64) Den Ridder, M.; Knibbe, E.; van den Brandeler, W.; Daran-Lapujade, P.; Pabst, M. A systematic evaluation of yeast sample preparation protocols for spectral identifications, proteome coverage and post-isolation modifications. *Journal of Proteomics* **2022**, *261*, 104576.

(65) Van Der Walt, S.; Schönberger, J. L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J. D.; Yager, N.; Gouillart, E.; Yu, T. Scikit-image: Image processing in python. *PeerJ* **2014**, *2014*, e453.

(66) Sofroniew, N.; Lambert, T.; Evans, K.; Nunez-Iglesias, J.; Bokota, G.; Winston, P.; Peña-Castellanos, G.; Yamauchi, K.; Bussonnier, M.; Doncila Pop, D.; Can Solak, A.; Liu, Z.; Wadhwa, P.; Burt, A.; Buckley, G., et al. napari: a multi-dimensional image viewer for Python. **2022**, DOI: 10.5281/ZENODO.6598542.

(67) Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; Tinevez, J. Y.; White, D. J.; Hartenstein, V.; Eliceiri, K.; Tomancak, P., et al. Fiji: an open-source platform for biological-image analysis. *Nature Methods 2012 9:7* **2012**, *9*, 676–682.

(68) Tinevez, J. Y.; Perry, N.; Schindelin, J.; Hoopes, G. M.; Reynolds, G. D.; Laplantine, E.; Bednarek, S. Y.; Shorte, S. L.; Eliceiri, K. W. TrackMate: An open and extensible platform for single-particle tracking. *Methods* **2017**, *115*, 80–90.

(69) Bateman, A.; Martin, M. J.; O'Donovan, C.; Magrane, M.; Alpi, E.; Antunes, R.; Bely, B.; Bingley, M.; Bonilla, C.; Britto, R.; Bursteinas, B.; Bye-AJee, H.; Cowley, A.; Da Silva, A.; De Giorgi, M., et al. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **2017**, *45*, D158–D169.

(70) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* **2003**, *17*, 2337–2342.

(71) Michael Rubinstein; Ralph H. Colby, *Polymer physics*; Oxford University Press: 2016.

(72) Zimm, B. H.; Stockmayer, W. H. The dimensions of chain molecules containing branches and rings. *The Journal of Chemical Physics* **1949**, *17*, 1301–1314.

(73) Davidson, I. F.; Barth, R.; Zaczek, M.; van der Torre, J.; Tang, W.; Nagasaka, K.; Janissen, R.; Kerssemakers, J.; Wutz, G.; Dekker, C.; Peters, J.-M. CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature 2023 616:7958* **2023**, *616*, 822–827.

(74) Brunet, A.; Tardin, C.; Salomé, L.; Rousseau, P.; Destainville, N.; Manghi, M. Dependence of DNA Persistence Length on Ionic Strength of Solutions with Monovalent and Divalent Salts: A Joint Theory-Experiment Study. *Macromolecules* **2015**, *48*, 3641–3652.

(75) Newton, M. D.; Fairbanks, S. D.; Thomas, J. A.; Rueda, D. S. A Minimal Load-and-Lock RuII Luminescent DNA Probe. *Angewandte Chemie International Edition* **2021**, *60*, 20952–20959.

3

A Microfluidic Platform for Extraction and Analysis of Bacterial Genomic DNA

Bacterial cells organize their genomes into a compact hierarchical structure called the nucleoid. Studying the nucleoid in cells faces challenges because of the cellular complexity while *in vitro* assays have difficulty in handling the fragile megabase-scale DNA biopolymers that make up bacterial genomes. Here, we introduce a method that overcomes these limitations as we develop and use a microfluidic device for the sequential extraction, purification, and analysis of bacterial nucleoids in individual microchambers. Our approach avoids any transfer or pipetting of the fragile megabase-size genomes and thereby prevents their fragmentation. We show how the microfluidic system can be used to extract and analyze single chromosomes from *B. subtilis* cells. Upon on-chip lysis, the bacterial genome expands in size and DNA-binding proteins are flushed away. Subsequently, exogeneous proteins can be added to the trapped DNA via diffusion. We envision that integrated microfluidic platforms will become an essential tool for the bottom-up assembly of complex biomolecular systems such as artificial chromosomes.

.

This chapter has been published as a pre-print: *Joesaar, A.; *Holub, M.; Lutze, L.; Emanuele, M.; Kerssemakers, J.; Pabst, M.; Dekker, C. A Microfluidic Platform for Extraction and Analysis of Bacterial Genomic DNA *bioRxiv* 2024. https://doi.org/10.1101/2024.10.17.618837. *Equal contribution

3.1. INTRODUCTION

The 3D spatial structure of genomes is important for gene expression and other cellular functions.[1] Whereas eukaryotes organize their genomic DNA in a cell nucleus where individual chromosomes occupy territories[2], bacteria organize their DNA into a compact structure called the nucleoid,[3,4,5] which is not enclosed by a nuclear membrane. Despite much research, we still have an incomplete understanding of the 3D organization of the bacterial genome and its effects on various biological processes.[6] There are many fruitful techniques for studying genome organization in cells such as chromosome conformation capture (3C/HiC),[7,8] high-resolution fluorescence microscopy,[9,10] and fluorescence-based localization techniques like FISH.[11] Yet, many questions remain due to the inherent complexity of the cellular environment. *In vitro* single-molecule techniques are powerful since they can study DNA-proteins at the single molecule level in controlled environments, but they typically use short DNA molecules that are orders of magnitude smaller than bacterial genomes.[12,13,14] Recently, we have proposed a novel *in vitro* method ("genome-in-a-box") to study chromosome organization from the bottom up using purified bacterial chromosomes[15], i.e. using DNA molecules of similar size to the genomes of living cells. Extraction of nucleoids from bacteria is nontrivial, although first examples of nucleoid isolation from bacteria date from the 1970's.[16] While we recently presented a method to obtain deproteinated DNA of megabasepair length from *E. coli,*[17] it remains challenging to avoid unwanted DNA damage that occurs due to mechanical shearing during pipetting. A microfluidic system could provide solutions to these limitations, as a precise and well-defined control of fluid flow minimizes the shear forces on the megabase-scale DNA. Furthermore, confining the DNA in microscale compartments allows for continuous monitoring of individual DNA objects. Microfluidic devices have been extensively used for trapping live cells[18] and cell-like synthetic compartments[19,20] and so-called 'mother-machine'[21] devices were developed for studying the growth and controlled cell lysis[22] of bacterial cells.

In this paper, we introduce a microfluidic platform that enables all the individual steps needed for lysis of individual bacterial cells, extraction of the bacterial nucleoid, deproteination of the nucleoid, imaging analysis of the extracted nucleoids, and introduction of DNA-structuring elements to the genomic DNA (Fig. 3.1). Notably, this approach allows for continuous tracking of the individual nucleoids in discrete microchambers that are hydrodynamically isolated from a buffer channel, which eliminates shear forces on the fragile genomic DNA molecules while allowing for addition and exchange of DNA-binding proteins. Flow control is provided by pneumatically actuated on-chip valves.[23]

We validate our microfluidic platform with the extraction and analysis of bacterial chromosomes of *B. subtilis* cells. Using confocal fluorescence microscopy, we can track individual cells from the moment they are inserted into the chambers, whereupon we observe their lysis, followed by deproteination, expansion, and relaxation of their chromosomal DNA. As proof-of-principle experiments of first steps towards the bottom-up assembly of an artificial chromosome, we show the effect of DNA-binding protein Fis and PEG on the 3D structure of isolated megabasepair-long DNA.

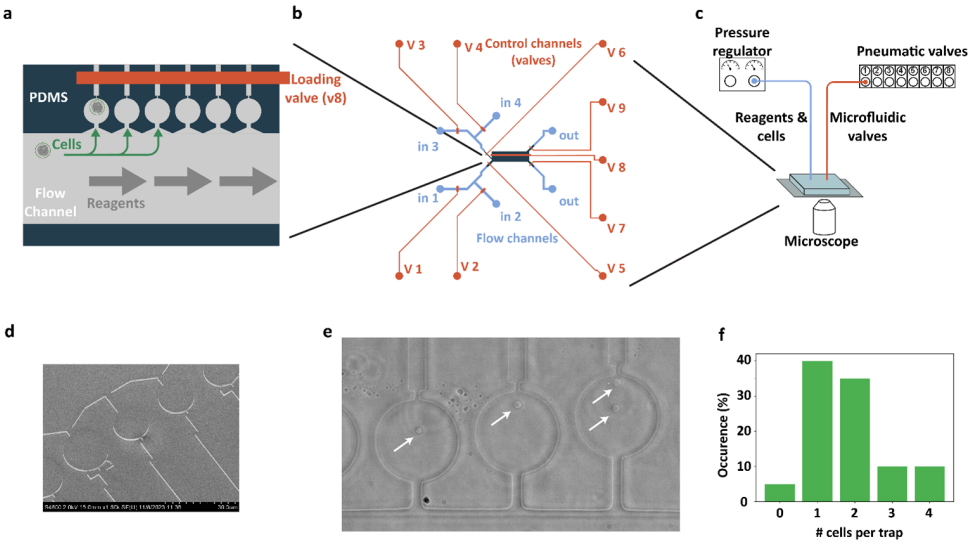

Figure 3.1: A microfluidic platform for extraction and purification of bacterial nucleoids. **a**. A liner array of microfluidic trapping chambers. Cells are inserted into the chambers by directing fluid flow from the large filling channel through the trapping chambers and out of the respective exhaust channels. The exhaust channels are too narrow for cells to pass through, allowing them to stay trapped in the chambers. **b**. Overview of the design of the microfluidic chip. Pneumatically actuated Quake valves are used in a push-down configuration to direct the flow of cells and reagents. **c**. Overview of the setup for 2-layer microfluidic platform with on-chip flow control. **d**. SEM micrograph of the PDMS trap array. **e**. Widefield micrograph of *B. subtilis* spheroplasts (arrows) in the trapping array. Scale bar is 10 μm. **f**. Average number of spheroplasts per trap. In a typical experiment, almost 40% of the traps (n=20) contain a single spheroplast.

3.2. RESULTS

3.2.1. DESIGN OF A MICROFLUIDIC PLATFORM FOR BACTERIAL DNA EXTRACTION

The main objective of our microfluidic platform is to perform bacterial nucleoid extraction and long-term analysis in individual micro-chambers with minimal perturbation of the chromosomal DNA. The microfluidic device is required to switch between a number of different input solutions/fluids to perform the individual steps of bacterial DNA extraction, analysis and reagent addition, while keeping the megabase-scale DNA fixed in the trapping chambers. Our initial tests revealed that megabase-scale DNA molecules are highly sensitive to the shear forces caused by flow rate fluctuations in microfluidic channels. Typically, the volume of fluid within the feeding tubes connected to a microfluidic chip is orders of magnitude larger than the volume of the microfluidic chip itself. Therefore, fluctuations in the flexible tubing led to very substantial fluctuations in the flow rate within the microfluidic chip. With these considerations in mind, we reasoned that to be able to reliably switch between different input reagents without perturbing the megabase-scale DNA molecules, all flow control would have to be incorporated

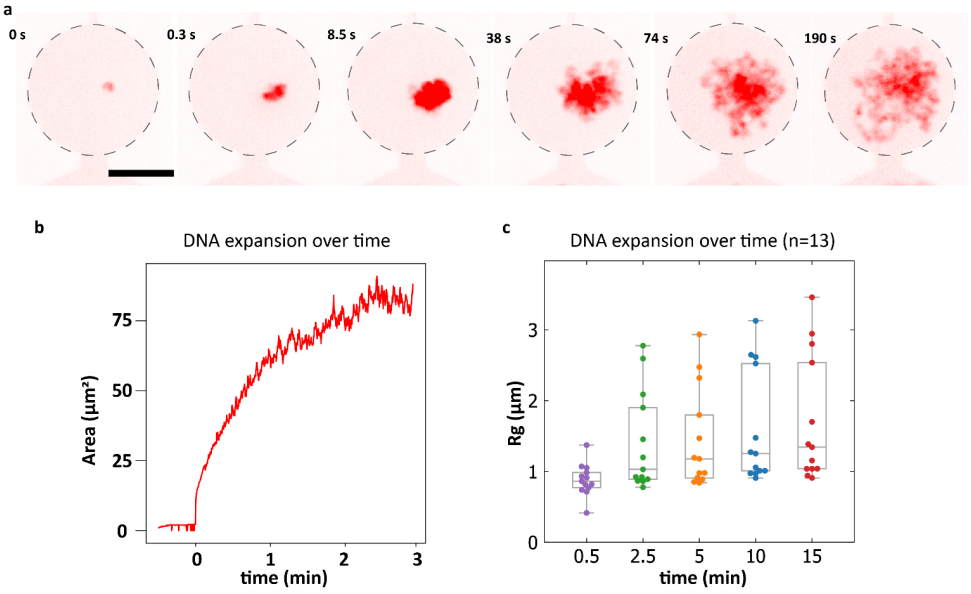

Figure 3.2: Trapping and lysis of quasi-2D confined *B. subtilis* spheroplasts. **a**. Sequence of images showing the lysis event of a single *B. subtilis* cell and the gradual expansion of the Sytox Orange labeled genomic DNA in the quasi-2D confined environment. Scale bar is 10 μm. **b**. Detected area of the chromosomal DNA from **a** over time. **c**. Calculated radius of gyration of n=13 *B. subtilis* nucleoids over time.

into the microfluidic chip. Therefore, we chose to use PDMS/glass for the material of the microfluidic chip as it enables straightforward implementation of on-chip flow control using pneumatically actuated microvalves (Supplementary Fig. 3.2).[23] This approach eliminates the dead volume effects of the connectors and tubing because the fluid flow is manipulated via integrated valves instead of external valves or syringe pumps.

Initially, we designed more conventional microfluidic trapping devices with a 2D grid arrangement of microfluidic traps (Supplementary Fig. 3.1). This configuration worked well for cell trapping and for their lysis, but keeping the extracted nucleoids localized in the traps proved to be impossible during reagent addition, since the flexible DNA polymer would inevitably exit the traps due to the applied flow. Therefore, we switched to a linear array of micro-chambers with individual input and output channels (Fig. 3.1a). The input channels of all these chambers are connected to a single 'filling channel' that runs parallel to the trapping array, while the output channels are actuated with a single pneumatic on-chip valve. The advantage of this 'side chamber' configuration is that it allows for reagents to be added to the chambers using two methods, either via direct flow or via diffusion from the filling channel. While the latter, importantly, avoided any shear forces on the fragile genomic DNA molecules while allowing for addition and exchange of DNA-binding proteins, the former, flow-based filling, was mainly used to insert the bacterial cells into the chambers. The dimensions of the input and output channels were selected such that cells could freely flow through the input channels with a width of 2 μm

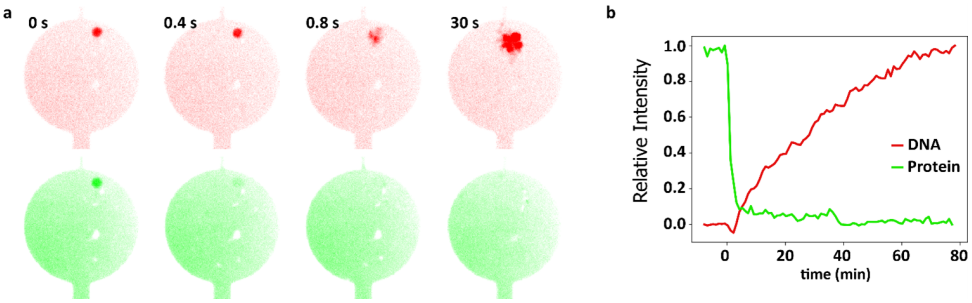

Figure 3.3: Analysis of the extracted bacterial genomic DNA in microfluidic chambers. **a**. Sequence of confocal fluorescence micrographs showing the lysis event of a single *B. subtilis* spheroplast. DNA (red channel) was labeled using Sytox Orange, while an amine reactive fluorescent label Alexa647-NHS (green channel) was used to track the movement of intracellular proteins during the lysis event. Scale bar is 10 μm. **b**. Normalized fluorescence intensities of the of the *B. subtilis* spheroplast from **a**. Green and red traces correspond to the Alexa647-NHS and Sytox Orange intensities respectively.

but would not pass through the output channel with a width of only 0.7 μm, resulting in their entrapment in the cylindrical chambers.

The input channels that connect the filling channel to the trapping microchambers were chosen as wide and short as possible to allow efficient diffusion while still providing enough insulation in order to prevent fluid flow from reaching the chamber and perturbing the DNA. The trapping chambers were 1.6 μm in height and 16 to 20 μm in diameter. The input and output channels that run through the pneumatically actuated valves had a rounded profile and a height of 10 μm (Fig. 3.1b, c). A detailed description of the valve design is given in Supplementary Fig. 3.2.

3.2.2. MICROFLUIDIC SIDE CHAMBERS ENABLE THE ISOLATION AND STUDY OF MEGABASEPAIR DNA WITHOUT SHEAR FLOW

The process of nucleoid extraction and analysis on our microfluidic platform consists of the following steps: 1) preparation of bacterial spheroplasts; 2) injection and trapping the spheroplasts in microchambers; 3) lysis of the spheroplasts which yielded to extraction of the DNA and the disassembly of DNA-binding proteins; and possibly 4) the addition of reagents of interest for follow-up biophysical studies.

Spheroplasts are spherical-shaped bacteria of which the outer cell wall has been removed. Preparation of the spheroplasts was performed in a cell-culture flask using lysozyme to digest the bacterial cell wall. The main reason for preparing the spheroplasts outside the microfluidic device is that the spherical shape and lack of motility makes the spheroplasts much easier to trap compared to the intact cells, which can swim out of the traps. Furthermore, this approach avoids contaminating the trapping chambers with lysozyme and cell-wall degradation products. Spheroplasts were injected into the filling channel of the microfluidic device (Materials and Methods, Fig. 3.1a). The exhaust

channels were then opened, directing the flow through the microfluidic side chambers such that spheroplasts were trapped in them (Fig. 3.1e).

We characterized the trapping efficiency of the system using *B. subtilis* spheroplasts. In a typical experiment, approximately 40% of the traps contained a single spheroplast and were therefore suitable for further analysis (Fig. 3.1e, f, Supplementary Fig. 3.3). In the current configuration, often more than one cell was observed to enter a chamber. The efficiency can potentially be improved by optimization of the geometry of the narrow output channels such that a single cell would block the flow and thus prevent successive cells from entering the same chamber. When the desired amount of spheroplasts was inserted into the trapping chambers, the flow through the traps was stopped and the cells were ready for lysis.

We explored two methods for cell lysis, (i) based on surfactants and (ii) based on osmotic shock. For (i), we used a lysis buffer solution (Materials and Methods) containing 5% surfactant (IGEPAL) and 500 nM of the intercalating fluorescent dye (Sytox Orange) that stains DNA, to detect the chromosomal DNA. When lysis buffer was flowed into the filling channel of the microfluidic device, the trapped *B. subtilis* spheroplasts abruptly ruptured within a minute, which was followed by a rapid expansion of their chromosomal DNA (Fig. 3.2a). Within minutes the DNA expansion reached a stable size (Fig. 3.2b, c), occupying a typical area of order 50 μm^2 (or a 3D volume of approximately 80 μm^3). Lysis method (ii) was performed by flowing a buffer with a low osmolarity (relative to the cell growth medium) through the filling channel of the microfluidic device with trapped spheroplasts. This resulted in a more irregular lysis of the spheroplasts, with some cells lysing but their chromosomal DNA only minimally expanding while others not lysing at all (Supplementary Fig. 3.4). Therefore, in all the following experiments, we used the surfactant-based lysis. However, as residual IGEPAL can potentially interfere with downstream protein-binding experiments, we explored what minimal concentration could be used to still yield robust lysis. We were able to lyse cells with only 0.2% IGEPAL and adopted that as a working concentration.

3.2.3. UPON LYSIS, PROTEINS DISSOCIATE FROM THE MEGABASEPAIR DNA

As the extracted bacterial genomic DNA is intended to be the starting material for studying the binding of chromosome-organizing proteins to bare DNA, we aimed to remove the original cellular proteins from the nucleoids. Upon lysis, most of these in fact spontaneously unbound from the nucleoid and diffused away. To measure how many proteins remained bound to DNA, we first used an amine-reactive fluorescent dye (Alexa647-NHS) to nonspecifically label the cellular proteins in *B. subtilis*. The succinimidyl ester group on this molecule reacts with primary amines (N-terminus and lysine residues), making all proteins viable targets for labeling. Although we expected this dye to react with cellular proteins only after the cells had been lysed, we did, interestingly, find that the dye was able to already permeate the membrane of the spheroplasts and thus enter the cytoplasm of the spheroplasts and label proteins therein (Fig. 3.3A). As the cells were lysed, the Alexa647 signal faded away from the DNA within seconds, indicating that the

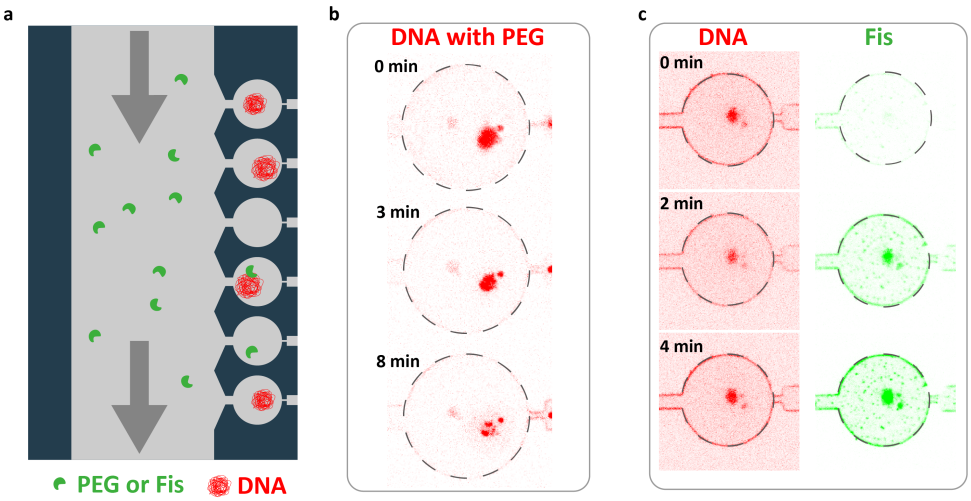

Figure 3.4: Manipulating the extracted chromosome using DNA-organizing elements. **a**. Experimental setup. DNA-interacting elements are flowed through the filling channel from where they can diffuse into the trapping chambers and interact with the trapped DNA molecules. **b**. A sequence of confocal fluorescence micrographs of a *B. subtilis* nucleoid being exposed to a 10% PEG solution. Scale bar is 10 μm. **c**. A sequence of confocal fluorescence micrographs of a *B. subtilis* nucleoid being exposed to a 3 μM Fis solution. Scale bar is 10 μm.

bulk of the *B. subtilis* proteins dissociated from the DNA very rapidly. Since the relatively low signal-to-background ratio of around 5 (Supplementary Fig. 3.5) limited the sensitivity of this assay, we decided to further investigate the degree of protein removal with mass spectrometry (Supplementary Fig. 3.7). Mass spectrometry samples were prepared in similar conditions in dialysis plugs (Materials and Methods) to mimic what happens in the microfluidic device. The mass spectrometry data indicated that a majority of total protein dissociated from both *E. coli* and *B. subtilis* genomic DNA. In particular, the amount of DNA-binding protein was reduced by at least 10-fold upon treatment (Supplementary Table 1) for *B. subtilis*.

3.2.4. DNA-BINDING PROTEINS AND CROWDERS CAN CONDENSE DNA

To demonstrate the capability of our microfluidic platform to introduce DNA-organizing elements to the trapped megabasepair-long and deproteinated DNA, we probed for the effects of a generic molecular crowding agent (PEG) and of a DNA-binding protein Fis to visualize DNA condensation in real time (Fig. 3.4a). Introduction of 10% PEG solution into the filling channel of the microfluidic device resulted in the significant compaction of the chromosomal DNA within the trapping chambers (Fig. 3.4b). We observed that most densely compacted DNA was more prone to adsorption to the walls of the microfluidic chambers (Supplementary Fig. 3.6). When instead 3 μM of fluorescently labeled Fis protein was flowed into the microfluidic device, we observed its binding to the trapped chromosomal DNA (Fig. 3.4c). However, in this case, no significant change to the shape

or size of the DNA was observed. These results are proof-of-principle illustrations of how our microfluidic platform allows for a diffusion-based addition of DNA-organizing elements to bacterial chromosomal DNA without perturbing the fragile megabase-scale DNA in the process.

3.3. CONCLUSIONS

We presented a microfluidic platform for the *in vitro* study of genome-sized DNA where DNA-organizing elements can be added without perturbing the trapped genomes. Studying such genome-sized DNA molecules with a "genome-in-a-box" approach[15] aims to fill the gap between live-cell and single-molecule experiments. A two-layer PDMS chip with integrated valves and cell-trapping chambers was used to trap and subsequently lyse *B. subtilis* spheroplasts, whereupon most of the DNA-binding proteins detached from the nucleoid. The approach allows for the extracted chromosomal DNA to be continuously observed from the moment of cell lysis.

Our work builds on previous studies of isolated *E. coli* nucleoids in bulk solution[24,25] and cell-sized microchannels.[22] A key limitation of the bulk methods is that it is very difficult to continuously track the behavior of individual DNA molecules, especially when new reagents are being introduced to the solution which exposes the DNA to mechanical disruption and concentration gradients. The main advantage of our approach compared to microfluidic devices with cell-sized microchannels is that the precise flow control provided by the integrated valves and the ability to direct the fluid flow through the trapping chambers allows for seamless cell loading and introduction of reagents to the trapping chambers. The use of the quasi-2D geometry in 1.6 micrometer high chambers makes it possible to image the isolated chromosomal DNA in a single plane, and resolve its finer structure and dynamics. Most importantly, the approach allows to locally trap a megabasepair-long DNA molecule and subsequently administer new components by diffusion, i.e. not by a flow which disrupts the DNA. Next to the great potential of the methodology, it also has some limitations, for example, some residual undesired surface interactions of the chromosomal DNA at very high densities, and the fact that custom-made microfluidic devices are single-use which leads to a relatively low overall experimental throughput.

Summing up, we developed a cell lysis method using a small amount of surfactant. This led to a rapid expansion of the chromosomal DNA and dissociation of cellular proteins from the DNA. We used mass spectrometry to verify that the DNA is mostly protein-free after this treatment. Proof-of-principle experiments using a crowding agent and DNA-binding protein Fis demonstrated the feasibility of the microfluidic "genome-in-a-box" approach. We envision to use the new microfluidic platform for further bottom-up studies of genome organization. Examples will include the effects of loop-extruding proteins on a genome-sized DNA, behavior of the nucleoid under spatial confinement, and *in-vitro* transcription-translation from genomic DNA.

3.4. MATERIALS AND METHODS

3.4.1. MICROFLUIDIC DEVICE FABRICATION

The PDMS/glass microfluidic devices were fabricated using 2-layer soft-lithography techniques.[23] The bottom (flow) layer master mold was fabricated using a combination of electron-beam (e-beam), photo-lithography, and DRIE etching. Etching mask for features with 1.6 um height was generated by spin-coating NEB-22 e-beam resist at 1000 rpm for 60 s on a 4" silicon wafer, followed by a 120 s bake at 110 °C. The patterns were then exposed using EBPG-5200 (Raith Nanofabrication), followed by a 120 s bake at 105 °C and developed for 60 s in MF322. The patterns were then DRIE etched into the silicon wafer on Oxford Estrellas using Bosch process at 5 °C in 14 steps. Next the 10 um features with rounded profiles were fabricated by spin-coating AZ10XT positive photoresist at 2000 rpm for 60 s, followed by a 180 s bake at 110 °C. The patterns were then exposed using a Heidelberg uMLA direct writer with a dose of 500 mJ/cm2 and developed for 6 min in AZ400K (diluted 1:3 in demi-water). Rounded profile was then obtained by placing the wafer on a 25 °C hotplate and then ramping the temperature to 120 °C in approximately 5 min, after which the wafer was allowed to cool down by switching off the hotplate. This allows the resist to reflow without introducing cracks or bubbles which often appear when placing a wafer with solidified AZ10XT directly on a 120 °C hotplate.

The top (control) layer master mold was fabricated using photo-lithography and DRIE etching. ARN4400.05 photoresist was spin-coated at 4000 rpm for 60 s on a 4" silicon wafer, followed by a 120 s bake at 90 °C. The patterns were then exposed using a Heidelberg uMLA direct writer with a dose of 60 mJ/cm2, followed by a 5 min bake at 100 °C and developed for 75 s in MF321. The patterns were then DRIE etched 20 μm into the silicon wafer on Oxford Estrellas using Bosch process at 5 °C in 150 steps.

The final devices consisted of bottom (flow) and top (control) layers that were bonded to a glass coverslip. We fabricated the layers with 2-layer soft-lithography techniques using ratios 18:1 and 6:1 of PDMS base to curing agent (Sylgard 184 Silicone Elastomer Kit, Dow Corning GmbH) for the bottom and top layers respectively. PDMS was desiccated before casting over the molds, and the desiccation was repeated for the top layer after casting. The bottom layer was spin-coated at [3'000 rpm for 60 s]. The two layers were baked at 90 °C for around 10 min until the top layer PDMS had hardened while the thin bottom layer PDMS was still slightly sticky to the touch. PDMS slabs were then cut out from the top layer castings and manually aligned and placed on top of the bottom layer. The two PDMS layers were gently pushed together but no weights were used as this often resulted in collapsing the 1.6 μm flow layer channels. Next, the two layers were thermally bonded by baking at 90 °C for 2 to 3 hours. The bonded PDMS devices were gently peeled off the bottom layer wafer, and inlet and outlet holes were manually punched with 0.5 mm diameter biopsy punch. Finally, the PDMS blocks were bonded onto the glass coverslips (#631-0147, 24x50 mm No.1.5, VWR (Avantor) International BV) using oxygen plasma (#119221 Atto, Diener electronic GmbH + Co. KG) at 40 W for 20 s.

3.4.2. BACTERIAL CELL CULTURE

E. coli bacterial cells (BN2179, HupA-mYPet frt, Ori1::lacOx240 frt, ter3::tetOx240 gmR, ΔgalK::tetR-mCerulean frt, ΔleuB::lacI-mCherry frt, DnaC::mdoB::kanR frt)[26] were incubated from glycerol stock in LB media supplemented with 50 μg/mL Kanamycin antibiotic (K1876, Sigma-Aldrich) in a shaking incubator at 30 °C and 300 rpm overnight. The cells were then resuspended in the morning to OD=0.05 and allowed to grow for until reaching OD of 0.1 (approx. 1 hour). The cells were then grown for another hour at 41 °C shaking at 900 rpm in order to arrest replication initiation. Next, appropriate volume of cell culture was spun down at 10000 g for 2.5 min, in order to obtain a pellet at OD_{eq} = 1 (approx. 8 x 10^8 cells). The pellet was resuspended in 475 μL cold (4 °C) sucrose buffer (0.58 M sucrose, 10 mM Sodium Phosphate pH 7.2, 10 mM NaCl, 100 mM NaCl). 25 μL lysozyme (L6876 Sigma-Aldrich, 1 mg/mL in ultrapure water) was immediately added and gently mixed into the cell/sucrose buffer suspension, followed by 30+ min incubation at room temperature to create spheroplasts.

B. subtilis bacterial cells (BSG4623, smc::-mGFP1mut1 ftsY::ermB, hbsU-mTorquais::CAT, ParB-mScarlet::kan, amyE::Phyperspank-opt.rbs-sirA (spec), trpC2)[27] were incubated from glycerol stock in SMM+MSM medium (300 mM Na2-Succinate, supplemented with 0.1% Glutamic acid and 2ug-mL Tryptophan) in a shaking incubator at 30 °C and 300 rpm overnight. The cells were resuspended in a fresh media in the morning (12.5x dilution of the overnight culture) and allowed to grow for 3 hours. Subsequently, 2 mM IPTG was added to the culture to arrest replication, while continuing shaking at 30 °C and 300 rpm. Finally, to create spheroplasts, lysozyme was added to the culture to final concentration of 500 ug/mL for at least 40 minutes. Spheroplasts created in either of two ways were then directly used for on-chip experiments.

3.4.3. EXPRESSION, PURIFICATION, AND LABELLING OF FIS

Full length *Escherichia coli* Fis with an N-terminal His_8 tag followed by a HRV-3C protease site, and appended with a C-terminal cysteine residue, was expressed from (pET28a-derived) plasmid pED72 in *Escherichia coli* ER2566 cells (New England Biolabs, *fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr73::miniTn10--Tet^S)2 [dcm] R(zgb-210::Tn10--Tet^S) endA1* Δ*(mcrCmrr)114::IS10)*. Cells were grown at 37 °C in baffled flasks on LB supplemented with 50 μg/ml kanamycin, expression was induced at an OD_{600} of 0.6 with 0.2 mM IPTG, and cells were harvested after overnight expression at 18 °C (8 min 4500 rpm, JLA8.1000 rotor). After washing the cells in PBS they were resuspended in buffer A (50 mM TrisHCl pH 7.5 (@RT), 750 mM NaCl, 1 mM EDTA, 0.05 mM TCEP, 10% (w/v) glycerol) and lysed using a French Press (Constant Systems) at 20 kpsi, 4°C. Following the addition of 0.35% polyethyleneimine, unbroken cells, DNA and protein aggregates were pelleted in a Ti45 rotor (30 min, 40.000 rpm, 4 °C), and Fis was precipitated from the supernatant by the addition of 476 g/l ammonium sulfate. Following centrifugation (JA-17 rotor, 10 minutes, 8500 rpm, 4 °C) and resuspension in buffer A, the sample was applied to 2 ml Talon Superflow resin (Clontech) pre-equilibrated with buffer A, and incubated for one hour while rotating at 4 °C. Subsequently, the resin was

washed with buffer A supplemented with 20 mM imidazole and finally Fis was eluted in 15 ml of buffer A supplemented with 1 mM β-mercaptoethanol and homemade 3C protease. Proteins were concentrated using a Vivaspin centrifugal concentrator (10 kDa cut-off) and further purified by size exclusion chromatography (SEC) on a Superdex 200 Increase 10/300 column pre-equilibrated with buffer A, eluting at ~16.5 ml. For preparation of fluorescently labelled Fis, 0.5 ml of concentrated protein was incubated 0.1 mM Alexa Fluor™ 647 C2 Maleimide (Invitrogen) for 30 minutes at room temperature, prior to size exclusion chromatography. Purified protein was snap-frozen and stored at -80 °C until use.

3.4.4. OPERATION OF THE MICROFLUIDIC NUCLEOID TRAPPING AND ANALYSIS DEVICE

The microfluidic device was mounted on the stage of a spinning disk confocal microscope (Andor CSU-X Yokogawa Spinning Disk Confocal). The operation of the devices requires precisely controlling pressure on the input lines, as well as supplying steady pressure on the valve lines. The control/valve channels of the device were filled with MilliQ water and actuated using a pneumatic valve array (FESTO), which was in turn actuated using an array of manual switches connected to a benchtop power-supply. The input pressure to the pneumatic valve array was 2 bar. The pressure to the reagent input channels of the microfluidic device was controlled using an adjustable pressure regulator (Fluigent). In a typical experiment, buffer solution (20 mM Tris-HCl ph 7.5, 50 mM NaCl, 1 mg/ml BSA) was connected to inlet 1 of the microfluidic device with a pressure of 300 mbar in order to wet all the flow channels and remove any air bubbles. Next the trapping chamber area of the device was filled with a buffer containing DNA intercalating dye (Sytox Orange, 400 nM) and incubated for 15 minutes.

As a next step, the bacterial spheroplasts should be trapped in the microfluidic chambers. To do so, they were injected into the device from inlet port 1 or 2, typically an input pressure of 1-5 mbar was used. Initially the spheroplast were added to the large filling channel by opening valves 1 (or 2), 5 and 7. After a sufficient number of spheroplasts were present in the filling channel, valve 7 was closed and valves 8 and 9 were opened to enable flow through the exhaust channels and thereby allow the spheroplasts to enter the trapping chambers. When a desired amount of spheroplasts had entered the chambers, valves 8 and 9 were closed and at this point the cells were ready for lysis. To lyse the spheroplasts, lysis buffer (Tris-HCl pH 7.5 40 mM, Potassium Glutamate 50 mM, BSA, 0.2 mg/mL, MgCl2 2.5 mM, Glucose 5%, Sytox Orange 500 nM, with addition of IGEPAL-CA-630 0.2% to aid lysis) was connected to inlet port 3 and was injected into the filling chamber by opening valves 3, 6 and 7 and using an input pressure of 1-2 mbar. Lysis of the individual spheroplasts could then be observed, this proceeded in a sequential manner starting from the upper trapping chambers. Stopping the flow of the lysis buffer would also stop the lysis events from happening in the downstream chambers and this allowed us to analyze the expansion of several nucleoids sequentially with a high frame rate within the same experiment. After all the spheroplasts had been lysed, valves 6 and 7 were closed and a desired reagent (PEG or Fis solution in this case) was connected to

inlet 4. Valves 6 and 7 were then reopened and the reagent solution was allowed to flow into the filling channel and to diffuse into the trapping chambers and interact with the trapped DNA.

3.4.5. IMAGE ACQUISITION AND ANALYSIS

To image isolated nucleoids in microfluidic traps, we used an Andor Spinning Disk Confocal microscope equipped with 100x magnification oil immersion objective. Isolated DNA was labelled by the intercalating dye Sytox Orange (S11368, Thermo Fischer Scientific, MA, USA) at concentration of 500 nM. At this concentration, the dye is known to reduce the persistence length of DNA to 37 nm. The dye was excited with 561 nm laser line (20% power, 250x gain, 10 ms exposure) with 617/73 nm filter on the emission. The acquisition computer was running Andor iQ 3.6 software. Multiple z-planes per each object, with separation of 1 μm between subsequent planes were acquired. For extended observations, we defined xy-positions and imaged them repeatedly over time, usually once every 30 or 60 seconds.

The analysis of nucleoid images within microfluidic traps was conducted using a custom Python code pipeline. We began by selecting circular regions of interest from in-focus plane images, encompassing the area inside the traps. These image sections were then thresholded to eliminate background noise and isolate the pixels containing fluorescent signal associated with nucleoids. The resulting set of pixels, each characterized by [position, intensity] values, was used to compute the radius of gyration for each nucleoid. This same pixel set also provided a measure of the total thresholded area occupied by the nucleoid.

3.4.6. SAMPLE PREPARATION FOR MASS SPECTROMETRY

Dialysis plug were chosen for sample preparation as they allowed to continuously exchange solutions in which nucleoid were suspended, similar to what happens in the microfluid device. This approach also allowed for removal of IGEPAL, which is otherwise incompatible with LC/MS, even at small concentrations [ref]. Spheroplasts were prepared from overnight cultures as described in the section 'Bacterial cell culture'. Lysis buffer contained final concentration of 0.2% IGEPAL and 50 mM Tris-HCl (pH 8). All spheroplast samples were lysed by adding 100 uL of spheroplast suspension to 900 uL of lysis buffer in dialysis plugs. Control and treatment samples were prepared in 3.5-5 kDa (G235029, Repligen Corporation, CA USA), and 300 kDa cut-off plugs (G235036, Repligen Corporation, CA USA) respectively following manufacturer's protocol. Each sample condition was prepared and measured in triplicates.

100 mM ammonium bicarbonate buffer (ABC) was prepared by dissolving ammonium bicarbonate powder (A6141, Sigma-Aldrich) in LC-MS grade quality water. 10 mM DTT (43815, Sigma-Aldrich) and iodoacetamide (IAA) (I1149, Sigma-Aldrich) solutions were made fresh by dissolving stock powders in 100 mM ABC. Next, 50 μL of 100 mM ABC

buffer was added to 200 uL of each sample to adjust pH, immediately followed by addition of 60 μL of 10 mM DTT and 1 hour incubation at 37 °C and 300 rpm in dark. Next, 60 μL of 20 mM IAA was added and samples were incubated in dark at room temperature for 30 min. Finally, 20 μL of 0.1 mg/mL trypsin (V5111, Promega) was added and samples were incubated for 16-20 hours at 37 °C and 300 rpm. On the following day, samples were purified by solid phase extraction (SPE). SPE cartridges (Oasis HLB 96-well μElution plate, Waters, Milford, USA) were washed with 750 μL of 100% methanol and equilibrated with 2x500 μL LC-MS grade H_2O. Next, 200 μL of each sample was loaded to separate SPE cartridge wells and wells were washed sequentially with 700 μL 0.1% formic acid, 500 μL of 200 mM ABC buffer and 700 μL of 5% methanol. Samples were then eluted with 200 μL 2% formic acid in 80% methanol and 200 μL 80% 10 mM ABC in methanol. Finally, each sample was collected to separate low-binding 1.5 μL tubes and speedvac dried for 2-3 hours at 45 °C. Samples were stored frozen at -20 °C until further analysis. Desalted peptides were reconstituted in 15 μL of 3% acetonitrile/0.01% formic acid prior to mass spectrometric analysis. Per sample, 2 μL of protein digest was analyzed using a one-dimensional shotgun proteomics approach[28,29]. Briefly, samples were analyzed using a nano-liquid-chromatography system consisting of an EASY nano LC 1200, equipped with an Acclaim PepMap RSLC RP C18 separation column (50 μm x 150 mm, 2 μm, Cat. No. 164568), and a QE plus Orbitrap mass spectrometer (Thermo Fisher Scientific, Germany). The flow rate was maintained at 350 nL/min over a linear gradient from 5% to 35% solvent B over 90 min, then from 35% to 65% over 30 min, followed by back equilibration to starting conditions. Data were acquired from 0 to 130 min. Solvent A was H2O containing 0.1% FA and 3% ACN, and solvent B consisted of 80% ACN in H2O and 0.1% FA. The Orbitrap was operated in data-dependent acquisition (DDA) mode acquiring peptide signals from 385–1250 m/z at 70,000 resolution in full MS mode with a maximum ion injection time (IT) of 75 ms and an automatic gain control (AGC) target of 3E6. The top 10 precursors were selected for MS/MS analysis and subjected to fragmentation using higher-energy collisional dissociation (HCD). MS/MS scans were acquired at 17,500 resolution with AGC target of 2E5 and IT of 100 ms, 2.5 m/z isolation width and normalized collision energy (NCE) of 28.

Mass spectrometric raw data were analyzed against the proteome database from *Escherichia coli* K12 (UP000000625, Tax ID: 83333, April 2024) or *Bacillus subtilis* strain 168 (UP000001570, Tax ID: 224308, April 2024, downloaded from `https://www.uniprot.org/`)[30] using PEAKS Studio X (Bioinformatics Solutions Inc., Waterloo, Canada)[31] allowing for 20 ppm parent ion and 0.02 m/z fragment ion mass error, 3 missed cleavages, carbamidomethylation as fixed and methionine oxidation, N/Q deamidation and N-terminal Acetylation as variable modifications. Peptide spectrum matches were filtered for 1% false discovery rates (FDR) and identifications with ≥ 1 unique peptide matches. The protein area was determined from the averaged top-3 peptide areas. Protein areas between conditions were compared by label free quantification using PEAKSQ, allowing a retention time shift tolerance of 5.0 minutes, a mass error tolerance of 10.0 ppm, and considering protein identifications filtered for 1% FDR. Peptide ID counts and min confident samples was set to 0 and significance method was set to ANOVA. Otherwise software default parameters were used. Data inspection revealed that one *B. subtilis*

treatment sample was indistinguishable from the control, and highly dissimilar to other two treatment samples. This pointed to an experimental error and this sample was left out from further analysis.

Relative protein abundancies were defined as the ratio of the 'treatment' over the 'control' conditions for the top-3 peptide areas, where the areas were weighted by each protein's molecular mass. For purposes of plotting, where no protein was identified on treatment condition, the fold change was set to 10^{-3}, and where no protein was measured on control condition, the fold change was set to the highest one in the dataset. Similarly fold change was limited between 2^{7} and 2^{-7} and plotted as log2(FC) (e.g. log2(2^{7}) = 7), and the maximum significance was capped at 10^{-20} (i.e. -log10(10^{-20}) = 20) for visualization purposes. To calculate the ratio between conditions presented in Supplementary Table 3.2, the top-3 peptide areas were summed up per each sample, and the values aggregated per each condition. Standard error of the mean from each condition was propagated to the error on the ratio by propagation of uncertainty.

In this study, we conducted a label-free quantification to compare the control with the purified sample. It is important to note that in such an experiment the remaining proteins in the purified sample are expected to appear more abundant than when they are part of a complex mixture. As a result, the apparent abundance of these proteins may seem higher in the purified sample compared to the control. The relative abundance of proteins after purification, as reported in Supplementary Table 3.2, should therefore be considered an upper bound estimate, and the actual quantities are likely significantly lower.

3.5. SUPPLEMENTARY INFORMATION

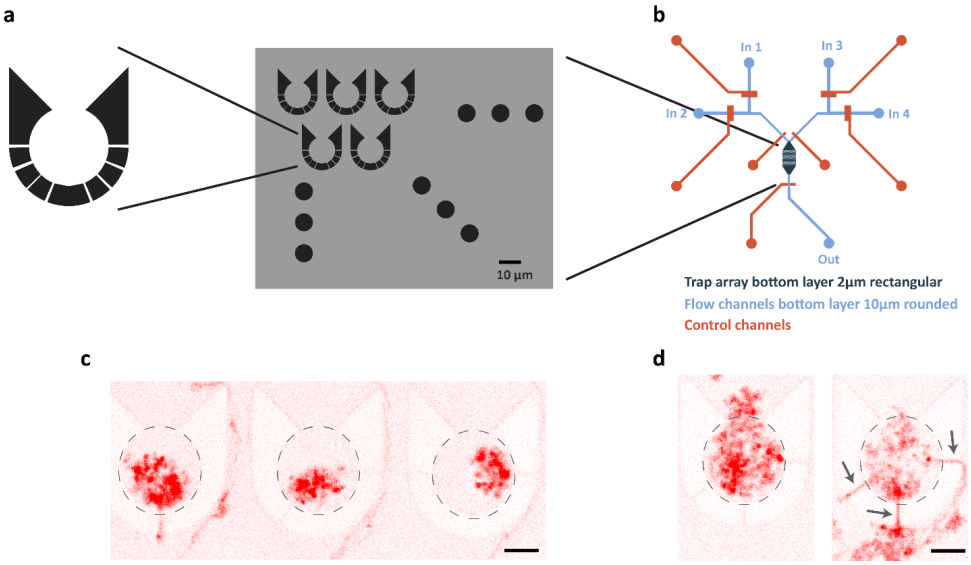

Figure S3.1: Initial design of the microfluidic trapping device for bacterial chromosome extraction. **a**. Configuration of a flow cell with a 2D grid of PDMS traps. **b**. Schematic of the complete microfluidic chip. The left inputs (in1 and in2) were used for loading the cells while the right inputs (in3 and in4) were used for loading the lysis buffer. **c**. Confocal fluorescence micrograph of extracted genomic DNA from 3 *B. subtilis* cells labeled with Sytox Orange. Scale bar is 5 μm. **d**. An example of DNA leakage through the narrow slits of the PDMS trap. Scale bar is 5 μm.

Figure S3.2: Design and fabrication of pneumatically actuated microfluidic valves. Push-down configuration was used for all the valves. The rounded profile in the mold for valves 1, 2, 3, and 4 was realized by reflowing AZ10XT photoresist at 120 °C.

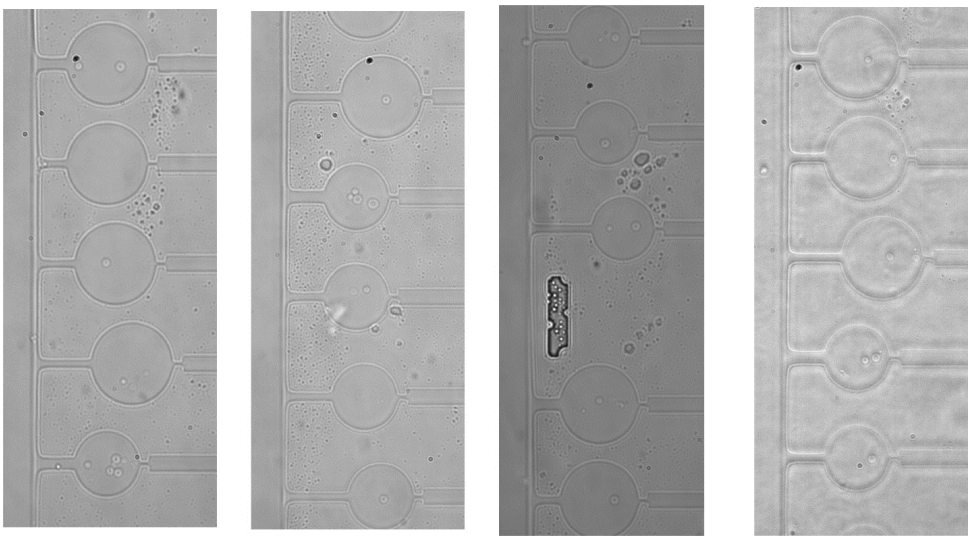

Figure S3.3: Widefield micrographs of *B. subtilis* spheroplasts in the microfluidic trapping chambers. Approximately 40% of the traps contain a single spheroplast.

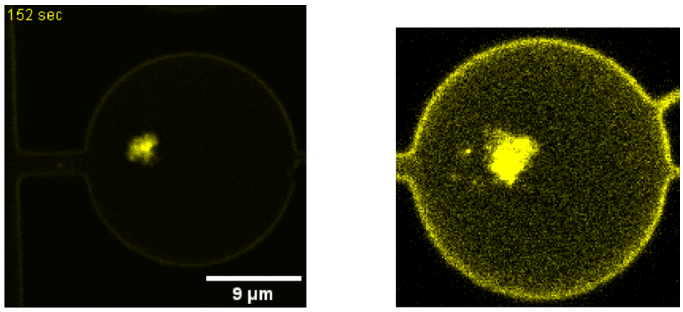

Figure S3.4: Lysis results using osmotic shock. Two examples of Sytox-Orange-labeled genomic DNA that was extracted using an osmotic shock, resulting in a more compacted structure.

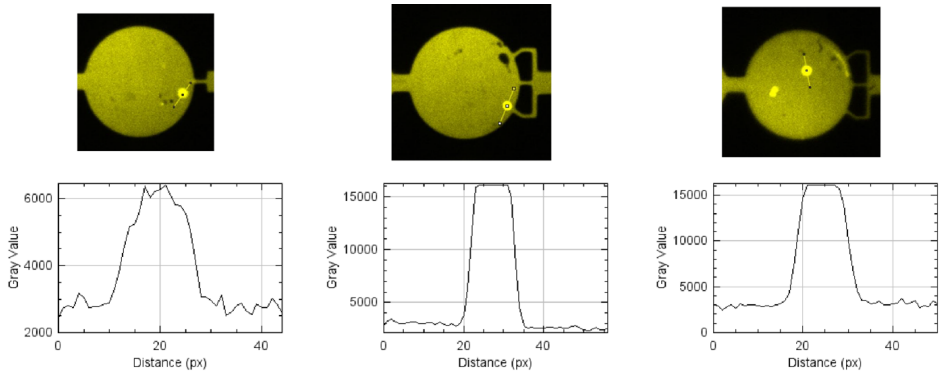

Figure S3.5: Signal-to-background ratio of Alexa647-NHS protein quantification measurements. Three examples of Alexa647 fluorescence intensity profiles of spheroplasts before lysis.

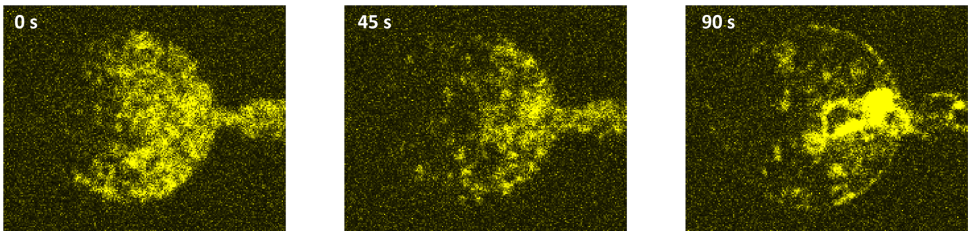

Figure S3.6: Adsorption of condensed DNA to chamber walls. An example of DNA condensation and absorption to the bottom of the trapping chamber upon addition of 20% PEG solution.

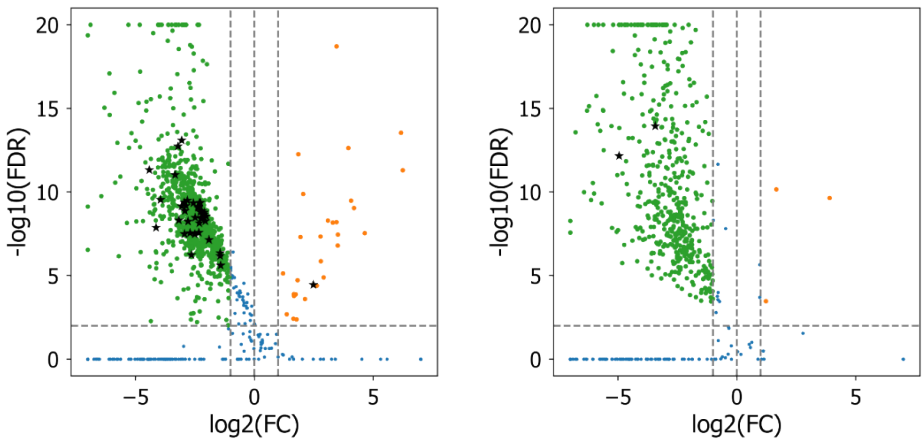

Figure S3.7: Protein abundance after removal for **a**. *E. coli* **b**. *B. subtilis*. Vertical lines indicate 2-fold removal and enrichment respectively, horizontal line corresponds to significance threshold of 1%. Points highlighted with stars represent DNA-binding proteins (cf. Table S2 and S3).

	E. coli	*B. subtilis*
Relative abundance [%]		
DNA-binding	19.2 ± 6.6	9.9 ± 2.6
All	24.6 ± 8.5	17.7 ± 4.4
Number of proteins reduced fewer than two-fold		
DNA-binding	1‡ (of 39)	0 (of 2)
All	69 (of 1246)	13 (of 490)

Table S3.1: Relative protein abundance after treatment for *E. coli* and *B. subtilis*, and the number of proteins that were not reduced more than 2-fold. Proteins were selected on significance threshold of 1% on the fold-change. To obtain relative values, proteins were weighted by their mass. ‡ RpoZ

Protein Name	**Description**
hbs	DNA-binding protein HU 1
rpoY	DNA-directed RNA polymerase subunit epsilon

Table S3.2: Proteins labeled as DNA-binding or DNA-processing in the *B. subtilis* sample (filtered on FDR 1%).

Protein Name	Description
dnaE	DNA polymerase III subunit alpha
topA	DNA topoisomerase 1
cbpA	Curved DNA-binding protein
dps	DNA protection during starvation protein
matP	Macrodomain Ter protein
gyrA	DNA gyrase subunit A
mukE	Chromosome partition protein MukE
rpoB	DNA-directed RNA polymerase subunit beta
ybaB	Nucleoid-associated protein YbaB
hupA	DNA-binding protein HU-alpha
parC	DNA topoisomerase 4 subunit A
mukF	Chromosome partition protein MukF
ybiB	Uncharacterized protein YbiB
hupB	DNA-binding protein HU-beta
mukB	Chromosome partition protein MukB
crp	DNA-binding transcriptional dual regulator CRP
rpoC	DNA-directed RNA polymerase subunit beta'
ompR	DNA-binding dual transcriptional regulator OmpR
ihfB	Integration host factor subunit beta
rpoA	DNA-directed RNA polymerase subunit alpha
rpoD	RNA polymerase sigma factor RpoD
crl	Sigma factor-binding protein Crl
dnaN	Beta sliding clamp
ihfA	Integration host factor subunit alpha
polA	DNA polymerase I
fis	DNA-binding protein Fis
uvrA	UvrABC system protein A
gyrB	DNA gyrase subunit B
nadR	Trifunctional NAD biosynthesis/regulator protein NadR
rpoS	RNA polymerase sigma factor RpoS
yaaA	DNA-binding and peroxide stress resistance protein YaaA
uvrD	DNA helicase II
yejK	Nucleoid-associated protein YejK
oxyR	DNA-binding transcriptional dual regulator OxyR
stpA	DNA-binding protein StpA
parE	DNA topoisomerase 4 subunit B
hns	DNA-binding protein H-NS
kdgR	HTH-type transcriptional regulator KdgR
rpoZ	DNA-directed RNA polymerase subunit omega

Table S3.3: Proteins labeled as DNA-binding or DNA-processing in the *E. coli* sample (filtered on FDR 1%).

3.6. REFERENCES

1. Oudelaar, A. M. & Higgs, D. R. The relationship between genome structure and function. *Nature Reviews Genetics* **22**, 154–168 (2021).

2. Hildebrand, E. M. & Dekker, J. Mechanisms and Functions of Chromosome Compartmentalization. *Trends in Biochemical Sciences* **45**, 385–396 (2020).

3. Toro, E. & Shapiro, L. Bacterial Chromosome Organization and Segregation. *Cold Spring Harb Perspect Biol* **2**, a000349 (2010).

4. Verma, S. C., Qian, Z. & Adhya, S. L. Architecture of the Escherichia coli nucleoid. *PLOS Genetics* **15**, e1008456 (2019).

5. Dame, R. T., Rashid, F.-Z. M. & Grainger, D. C. Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nature Reviews Genetics* **21**, 227–242 (2020).

6. Lioy, V. S., Junier, I. & Boccard, F. Multiscale Dynamic Structuring of Bacterial Chromosomes. *Annu. Rev. Microbiol.* **75**, 541–561 (2021).

7. Brandão, H. B., Ren, Z., Karaboja, X., Mirny, L. A. & Wang, X. DNA-loop-extruding SMC complexes can traverse one another in vivo. *Nat Struct Mol Biol* **28**, 642–651 (2021).

8. Falk, M. *et al.* Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395–399 (2019).

9. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, (2018).

10. Ricci, M. A., Manzo, C., García-Parajo, M. F., Lakadamyali, M. & Cosma, M. P. Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. *Cell* **160**, 1145–1158 (2015).

11. Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).

12. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338–1345 (2019).

13. Greene, E. C., Wind, S., Fazio, T., Gorman, J. & Visnapuu, M.-L. Chapter 14 - DNA Curtains for High-Throughput Single-Molecule Optical Imaging. in *Methods in Enzymology* (ed. Walter, N. G.) vol. 472 293–315 (Academic Press, 2010).

14. Ganji, M. *et al.* Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102–105 (2018).

15. Birnie, A. & Dekker, C. Genome-in-a-Box: Building a Chromosome from the Bottom Up. *ACS Nano* acsnano. °C07397 (2020) doi:10.1021/acsnano. °C07397.

16. Stonington, O. G. & Pettijohn, D. E. The Folded Genome of Escherichia coli Isolated in a Protein-DNA-RNA Complex. *Proceedings of the National Academy of Sciences* **68**, 6–9 (1971).

17. Holub, M. *et al.* Extracting and characterizing protein-free megabase-pair DNA for in vitro experiments. *Cell Reports Methods* 100366 (2022) doi:10.1016/j.crmeth.2022.100366.

18. Deng, Y., Guo, Y. & Xu, B. Recent Development of Microfluidic Technology for Cell Trapping in Single Cell Analysis: A Review. *Processes* **8**, 1253 (2020).

19. Nahas, K. A. *et al.* A microfluidic platform for the characterisation of membrane active antimicrobials. *Lab Chip* **19**, 837–844 (2019).

20. Joesaar, A. *et al.* DNA-based communication in populations of synthetic protocells. *Nature Nanotechnology* **14**, 369 (2019).

21. Wang, P. *et al.* Robust Growth of Escherichia coli. *Current Biology* **20**, 1099–1103 (2010).

22. Pelletier, J. *et al.* Physical manipulation of the Escherichia coli chromosome reveals its soft nature. *PNAS* **109**, E2649–E2656 (2012).

23. Unger, M. A., Chou, H.-P., Thorsen, T., Scherer, A. & Quake, S. R. Monolithic Microfabricated Valves and Pumps by Multilayer Soft Lithography. *Science* **288**, 113–116 (2000).

24. Wegner, A. S., Alexeeva, S., Odijk, T. & Woldringh, C. L. Characterization of Escherichia coli nucleoids released by osmotic shock. *Journal of Structural Biology* **178**, 260–269 (2012).

25. Cunha, S., Woldringh, C. L. & Odijk, T. Polymer-Mediated Compaction and Internal Dynamics of Isolated Escherichia coli Nucleoids. *Journal of Structural Biology* **136**, 53–66 (2001).

26. Wu, F. *et al.* Direct imaging of the circular chromosome in a live bacterium. *Nat Commun* **10**, 1–9 (2019).

27. Tišma, M. *et al.* Direct observation of a crescent-shape chromosome in expanded Bacillus subtilis cells. *Nat Commun* **15**, 2737 (2024).

28. Köcher, T., Pichler, P., Swart, R. & Mechtler, K. Analysis of protein mixtures from whole-cell extracts by single-run nanoLC-MS/MS using ultralong gradients. *Nat Protoc* **7**, 882–890 (2012).

29. den Ridder, M., Knibbe, E., van den Brandeler, W., Daran-Lapujade, P. & Pabst, M. A systematic evaluation of yeast sample preparation protocols for spectral identifications, proteome coverage and post-isolation modifications. *Journal of Proteomics* **261**, 104576 (2022).

30. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158–D169 (2017).

31. Ma, B. *et al.* PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* **17**, 2337–2342 (2003).

4

STRUCTURAL AND DYNAMIC CHARACTERIZATION OF CHROMOSOMES ISOLATED IN MICROFLUIDIC TRAPS

DNA biophysics is investigated both in vivo and in vitro as well as by means of polymer simulations. However, tools for biophysical studies of DNA at the megabase scale are lacking. Here we applied a new single-molecule methodology, genome-in-a-box, to investigate the behavior of isolated bacterial chromosomes in quasi-2D microfluidic traps. Using fluorescent microscopy, we observed a rapid chromosome expansion upon cell lysis in real time. Full subsequent relaxation occurred over 30 minutes, which we attribute to protein dissociation and DNA entanglement resolution. Quantitative image analysis of the deproteinated chromosomes revealed fast (seconds) internal dynamics and slower (minutes) rearrangements at larger scales. Our findings demonstrate the potential of microfluidic trapping for studying chromosome behavior at the megabase scale.

Next to myself, various other people contributed to the content of the work described in this chapter, namely Alex Joesaar, Leander Lutze, Jacob Kerssemakers, Janni Harju, Cees Dekker.

4.1. INTRODUCTION

4.1.1. DNA AS A POLYMER

DNA is the carrier of our genetic code but it also is a physical object, namely a biopolymer. Its bending stiffness is commonly quantified in terms of persistence length, *Lp,* which is the length over which thermal fluctuations will cause a polymer to bend, on average, by 1 radian [1]. Quantitatively, the persistence length of a DNA molecule has been measured to be about 50 nm, but the value depends on range of inherent and extrinsic factors. Ions interacting with the DNA as well as the DNA sequence itself can decrease the value of *Lp* by as much as 20% [2].

The structure and dynamics of DNA on scales beyond the persistence length are governed by polymer physics. A range of theories of varying complexity have been developed to describe its behavior, all introducing some assumptions on how individual DNA segments connect and interact with each other. These segments, or "monomers", are abstractions most commonly representing some multiple of the persistence length.

The *ideal chain model* assumes that there are no interactions between monomers [3]. The simplest example of the ideal chain model is the *freely jointed chain model*, where bonds of length l are connected by fully flexible joints. Similarly simplistic, the *freely rotating chain model* fixes the value of the angle θ at which two bonds are joined, while it allows for full rotational flexibility of the joint. With an effective monomer length $b = l cos\theta$, this yields a mean-squared *end-to-end distance:*

$$\left\langle R_e^2 \right\rangle = Nb^2 \, .$$

The *worm-like chain* is a special case of freely rotating chain. It allows only for small values of the bond angle, $\theta \ll 1$, where θ is the angle in radians, which allows to write $cos\theta \approx 1 - \frac{\theta^2}{2}$ and results in a characteristic monomer size

$$b = 2L_p = 2\left(l\frac{2}{\theta^2}\right) ,$$

where L_p is the persistence length and b is the Kuhn length.

Many relevant biological molecules, including plasmids and bacterial chromosomes, are circular. However the mean-squared end-to-end distance is defined only for linear chains. For circular molecules, instead the *radius of gyration,* the average distance of monomers to polymer's center of mass, is commonly used as a size descriptor,

$$R_g^2 = \frac{1}{N}\sum_i^N \left(\vec{r}_i - \vec{r}_{com}\right)^2 \, .$$

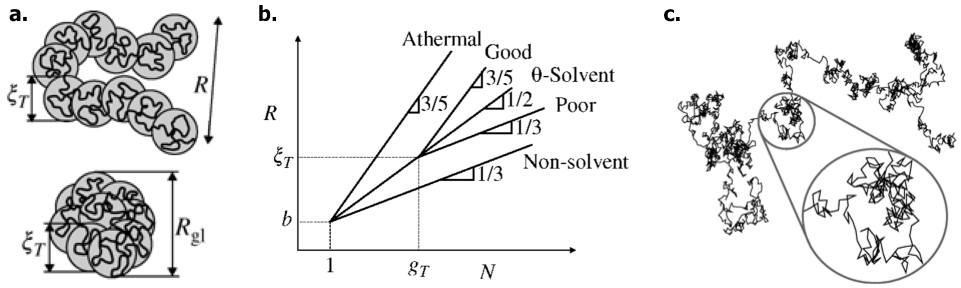

Figure 4.1: a) Polymer conformations in good (top) and poor (bottom) solvents. b) Scaling of polymer size R with number of monomers N for different solvent conditions. c) Illustration of fractal (self-similar) nature of a polymer chain. All figures are from Rubinsten & Colby [3].

In fluorescence imaging applications, such as here, the masses of individual monomers represented by the aggregate fluorescence intensity at each pixel must be factored into this equation as relative weights. Similarly, the monomer position of real polymers is not fixed in time, and their fluctuations should be faithfully accounted for. It is therefore preferable to average the metric over ensemble of allowed conformations, and to report the *mean-squared radius of gyration*, $\left\langle R_g^2 \right\rangle$. For a linear ideal chain, the radius of gyration reduces to

$$\left\langle R_g^2 \right\rangle = \frac{Nb^2}{6} = \frac{\left\langle R_e^2 \right\rangle}{6},$$

and for a ring polymer, the value is twice as small,

$$\left\langle R_g^2 \right\rangle = \left\langle R_e^2 \right\rangle / 12.$$

4.1.2. POLYMERS IN REAL SOLVENTS

The central assumption of the ideal chain model is the absence of any interactions between the monomers. In other words, the monomers are infinitesimally thin, and their movements can be described by a random walk. This translates to properties of ideal chains being Gaussian, following a scaling $R_e \propto N^{0.5}$. However, realistic polymers are made of monomers of finite thickness, that exhibit some mutual interaction. This leads to scaling $R_e \propto N^{\nu}$ where $\nu \neq 0.5$ (Fig. 4.1b). The effective interaction is a balance between attractive forces between monomers, and the effect of the volume excluded by them. It can be overall repulsive, with $\nu > 0.5$, or attractive, with $\nu < 0.5$. These conditions are referred to as good and poor solvent respectively.

The so-called *real chain model*' can be used to describe polymers in these conditions. In comparison to ideal chain, the real chain model accounts not only for the physics

and self-interaction of a polymer, but also for its interactions with the solvent. Here, we describe the basic features of the model. First, we describe the polymer at some level of abstraction that is larger than its individual monomers, an approach named coarse-graining. To do so, we define a characteristic unit of organization of the polymer in its natural shape in thermal equilibrium, a so-called *thermal blob*, with a length scale ξ_T (Fig. 4.1a). This length scale is set by the balance of the excluded volume interactions of a given number of monomers g_T and the thermal energy $k_B T$. Here, g_T is a number of monomers within a thermal blob, a value ususally much smaller than the total number of monomers N. This results in

$$k_B T |V| \frac{g_T^2}{\xi_T^3} \approx k_B T,$$

where V is the excluded volume. Therefore

$$\frac{|V|\, g_T^2}{\xi_T^3} \approx 1,$$

Now, recall that spatial extent of an ideal polymer chain can be modelled as $\langle R^2 \rangle = (N^\nu b)^2$, with $\nu = 0.5$ for the case of no interactions. Assuming that this exponent also holds *inside* the thermal blob, we can write the same relationship as we did for the while chain

$$\xi_T \approx b\, g_T^{0.5}\,.$$

Combining the two preceding equations, we get:

$$\xi_T \approx \frac{b^4}{|V|}\,,$$

$$g_T \approx \frac{b^6}{V^2}\,,$$

for the thermal blob size, and number of monomers therein, respectively. When $V > 0$, the monomers repel each other, or, in other words, they exhibit preferable interaction with their environment. We refer to this situation as good solvent condition.

The end-to-end distance of a polymer R_e can be written in terms of the thermal blob size and number of monomers therein as:

$$R_e \approx \xi_T \left(\frac{N}{g_T}\right)^\nu,$$

which again is a function of the quality of the solvent ν. The exponent ν can be approximated with the Flory theory [4]. Briefly, the theory expresses the free energy of a chain in a solvent as a sum of spring-like connectivity and two-body collisions between monomers. Minimization of the free energy yields $R_e \propto N^{0.6}$. More detailed treatments yield a similar value, $\nu = 0.588$ [3]. For poor solvent conditions, the polymer minimizes its interaction with the environment, and there is effectively an attraction between the monomers, i.e. $V < 0$. The volume of a polymer globule is then just the volume of densely packed thermal blobs, with its extent proportional to the cube root of its volume (Fig. 4.1a),

$$R_{gl} \approx \left(\xi_T^3 \frac{N}{g_T}\right)^{1/3}.$$

In other words, the scaling exponent in poor solvent conditions is $\nu = 1/3$.

Next, we consider the spatial distribution of a polymer. Recall that above, for a thermal blob of size ξ_T, number of monomers g_T and monomer size b, we have written:

$$\frac{\xi_T}{b} = g_T^{\nu} \rightarrow g_T = \left(\frac{\xi_T}{b}\right)^{1/\nu}.$$

This relationship can be written for a polymer on scales independent of the thermal blob size. Setting N as the number of monomers, r as a distance, and $\nu = 0.5$, we can write

$$N \approx \left(\frac{r}{b}\right)^2.$$

Let's denote the probability of finding two monomers separated by a distance r within a unit volume as $g(r)$. This can be approximated as an average number density N/r^3 in a volume r^3. Therefore

$$g(r) \approx \frac{N}{r^3} \approx \frac{1}{rb^2},$$

i.e. the probability scales inversely with increasing distance. Remarkably, this relationship holds for any pair of monomers along the chain. The chain is can be referred to as *self-similar*, or *fractal* (Fig. 4.1c). The *fractal dimension* D turns out to be the inverse of the scaling exponent, $D = 1/\nu$. For an ideal chain, $\nu = 0.5$ and $D = 2$, as we have seen above. For poor solvent, the polymer behaves as a globule and $D = 3$. For good solvent, the chain assumes a random coil configuration, with fractal dimension between 1 and 2, and more specifically, for a good solvent with $\nu = 0.588$, $D \cong 1.7$.

4.1.3. POLYMER DYNAMICS

Polymers in solutions are in constant movement and static descriptors are not sufficient to fully characterize their behavior. Many theoretical descriptions have been developed to describe polymer dynamics, with two approaches in particular, the *Rouse* and the *Zimm models* [3]. Here we briefly describe both, focusing on the timescale of motion of individual polymer segments that they predict.

The *Rouse model* represents a polymer chain as N beads connected by springs of mean size b. There is no coupling between the movement of the beads and the movement of the solvent. In this model the characteristic timescale for motion of individual beads can be written as

$$\tau_0 \approx \frac{\zeta b^2}{k_B T} ,$$

where ζ is friction coefficient of the beads due to their spring-like connectivity. This time is also called the Kuhn monomer relaxation time. The Rouse relaxation time τ_R of a polymer of N monomers scales with exponent $2\nu+1$,

$$\tau_R \approx \tau_0 N^{2\nu+1} .$$

The Rouse model offers a good description of polymer dynamics in dense solutions, but falls short for polymers in dilute solutions.

The *Zimm model* considers the hydrodynamic interactions of monomers with the solvent, and is more applicable in dilute conditions. Consider the Einstein relation for diffusivity in medium with friction coefficient ζ:

$$D = \frac{k_B T}{\zeta} .$$

The friction coefficient ζ is proportional to chain size R_{ch} and the medium viscosity η, hence

$$D = \frac{k_B T}{6\pi\,\eta R_{ch}} = \frac{k_B T}{6\pi\,\eta(bN^{\nu})} .$$

This allows to write a characteristic time for a chain to diffuse away for a distance on the order of its own size. This time is the Zimm relaxation time

$$\tau_Z \approx \frac{R_{ch}^2}{D} \approx \frac{\eta}{kT} R_{ch}^3 \approx \frac{\eta b^3}{kT} N^{3\nu} \approx \tau_0 N^{3\nu} ,$$

Notably, the Zimm time τ_Z has a somewhat weaker dependence on chain length then the Rouse time τ_R in most solutions,

$$3\nu < 2\nu + 1, \ \ (\nu < 1) \, ,$$

and long chains in dilute solutions thus tend to move faster than predicted by the Rouse model.

Finally, experimental techniques, such as single particle tracking, frequently seek to characterize the diffusivity of DNA (segments) by measuring the distance it travels over a time interval Δt. This is the mean-squared displacement *MSD*:

$$MSD(\Delta t) = \left\langle (x(t+\Delta t) - x(t))^2 \right\rangle \approx D_{app}(\Delta t)^{\alpha} \, ,$$

where D_{app} is the apparent diffusion coefficient and α the scaling exponent. To connect with the classical Brownian diffusion model, one can write $\alpha = 1$ and $D = 2dD_{app}$, where d is dimensionality of the system. Biological experiments across prokaryotes and eukaryotes have yielded α between 0.6 and 0.9 for whole-coil diffusion and 0.4 – 0.7 for the diffusion of individual genomic loci [5]. As $\alpha < 1$, this behavior is subdiffusive.

4.1.4. CHROMOSOME ISOLATION EXPERIMENTS

Scientist pursue many experimental approaches for studying the biological physics of chromosomes. Most commonly, chromosomes are studied inside living cells. While this environment is by far the most relevant, detailed studies are hindered by the environment's complexity and limited access for DNA manipulation. To address these shortcomings, the community of single-molecule biophysicists developed experimental approaches for studying DNA molecules of varying length in vitro. Recently, efforts have been made to carry out experiments with DNA molecules of size on the scale of whole chromosomes.

The study of individual chromosomes that were isolated from cells dates back at least 5 decades. Researchers in the group of David Pettijohn at the University of Colorado Medical Center conducted pioneering experiments with isolated phage and *E. coli* genomes. In 1970, they isolated *E. coli* nucleoids, and, using sedimentation studies and gel electrophoresis, observed that these were highly compacted. The authors deduced that the compaction was mainly due to bound molecules of RNA polymerase as these genomes could be unfolded with RNase and heat [6]. In 1974, they used fluorescence microscopy to image isolated *E. coli* nucleoids, observing that they remained similarly compacted as nucleoids in cells, but gradually expanded and became more diffuse [7]. In the same year, they observed that isolated nucleoids can be prevented from decompaction by treatment with the polyamine spermidine [8]. Similar studies were conducted with phage genomes by this and other groups in following years [9]. In 1975, the same group

investigated in vitro transcription from nucleoids isolated from *E. coli*. They observed that the transcription rate from compacted nucleoids was higher than that from expanded ones, and that the compaction state did not change during transcription [10]. In 1978 the authors studied the effect of irradiation-induced breaks on the nucleoid structure. This allowed them to deduce that *E. coli* chromosomes contain, on average, about 100 ± 30 domains of supercoiling [11].

A group of Dutch scientists revived these early studies of isolated *E. coli* genomes at the break of the 21st century. In 1998, the theorist Odijk reflected on in-vivo studies of groups of Woldringh and Westerhoff and presented theory describing the phase-separation-driven compaction of supercoiled DNA into a bacterial cell [12]. This study spurred interest in physical phenomena involved in DNA compaction and gave rise to new *E. coli* chromosome isolation studies. In 2001, researchers in the group of Woldringh isolated *E. coli* chromosomes from cells lysed by osmotic shock, and imaged them by fluorescence microscopy [13]. They observed strong nucleoid compaction upon addition of polyethylene glycol (PEG) [14]. In 2005, the same researchers conducted spot-labeling experiment in isolated nucleoids. Using the LacO/I system they could track labeled spots over 12 s and observed that their motion was restricted to subregions within the nucleoid [15]. In 2006, Zimmerman realized that nucleoids isolated so far were likely to retain a large number of proteins, and experimented with using urea and trypsin for protein removal, observing a partial expansion of the observed nucleoids [16]. He noted that the largest nucleoid dispersion was achieved by treatment with RNaseH, and, similar to previous studies, observed nucleoid compaction with the addition of PEG. In 2012, the groups of Woldringh and Odijk repeated their earlier experiments with lysing *E. coli* nucleoids by osmotic shock, but this time followed up by treatment aimed at removing DNA-bound proteins. Using fluorescence microscopy, they observed a 2-fold volume increase after treating nucleoids with surfactants and Proteinase K [17]. Based on these observations the authors developed a model of nucleoid as a branched DNA supercoil, maintained by proteins. Such model can be viewed as an extension to the original theory of Odijk [12].

In 2011, a group of Japanese scientists studied DNA compaction using atomic force microscopy (AFM) at two different DNA lengths, 166 kb (T4 DNA) and 880 kb [18]. They observed that the larger molecules were more sensitive to compaction by spermidine. It also exhibited an intermediate compaction state, with smaller blobs emerging within the molecule, whereas shorter molecule compacted in one step, without an intermediate. They dubbed the intermediate condition “intrachain segregation”. A similar transition was observed by group of researchers in Singapore in 2018 [19]. They conducted experiments with microscopy and dynamic light scattering (DLS) using chromatinized (i.e. histone-coated) repeats of T4 DNA (166 - 664 kb) and saw intrachain-like segregation behavior upon addition of spermine. Addition of divalent cation Mg2+ also led to compaction, but without the intermediate. Unsurprisingly, the higher the degree of chromatization, the more compact the molecules were. The same authors followed up on this work in 2020, when they conducted experiments with the same chromatinized repeats of T4 genome, but this time investigating the role of BSA as a compacting agent [20].

The experiments with isolated *E. coli* chromosome in groups of Odijk and Woldringh continued to fuel the interest in the physical principles of genome organization. Jun and Mulder presented a theory of entropy as main driver of chromosome segregation in 2006 [21,22]. In 2012, Pelletier and Halvorsen, working in the groups of Jun and Wong, leveraged the mother machine device developed earlier in the group of Jun to carry out biophysical studies of single isolated *E. coli* chromosomes in microfluidic confinement [23]. They observed that, following cell lysis, chromosomes expanded rapidly in matter of seconds, after which a slower equilibration phase continued. It took about 30 minutes for the isolated chromosomes to reach their equilibrium size. Chromosomes isolated from exponentially growing cells expanded to several-fold larger size than chromosomes from cells in stationary phase. The authors went on to apply an optical-trap based micro-piston device that allowed them to measure forces required for DNA compression, revealing that DNA could be compacted with forces as low as 100 pN. Additionally, introduction of PEG resulted in reversible DNA compaction, happening first at individual regions, akin to intrachain segregation observed earlier. In another notable example of applying microfluidics to studying megabasescale DNA, Freitag et al developed a meandering nanofluidic device that allowed them to stretch out the DNA of single yeast chromosome (nearly 6 Mbp) into full length and image it in a single frame with fluorescent microscopy [24].

Most recently, studies of isolated *E. coli* chromosomes were revisited by us [25]. Reviewing earlier experiments, we noted that previous studies worked with chromosomes at arbitrary degrees of replication, and that there was lack of characterization of isolated DNA in terms of residual protein. We also wanted to extend the degree of quantification in studies with recombinant proteins using an isolated genomic scaffold. We were able to isolate megabase-scale DNA from *E. coli* cells, and showed that it was largely deproteinated. We went on to show that the DNA could be compacted by PEG, and observed a concentration-dependent compaction effect of the *E. coli* nucleoid associated protein Fis. However, we also encountered several challenges, including inability to wash away unbound protein or reverse protein binding by introduction of a different buffer. Additionally, due to the large experimental volumes, some experiments required prohibitively large amounts of recombinant protein. Finally, transient but notable temperature and concentration gradients led to flows within the experimental sample, which made it challenging to track isolated nucleoids on timescales longer than just a few minutes.

To address these limitations, we developed the microfluidic trapping device that we presented in Chapter 3 of this thesis. Below, we present, analyze, and discuss data obtained from imaging isolated chromosomes in such microfluidic devices.

4.2. RESULTS AND DISCUSSION

We used the previously established microfluidic trapping device for controlled capture and lysis of *B. subtilis* spheroplasts (Chapter 3). The spheroplasts were prepared with a single chromosome, as discussed in Chapter 3 and flown into quasi-2D traps (height 1.6

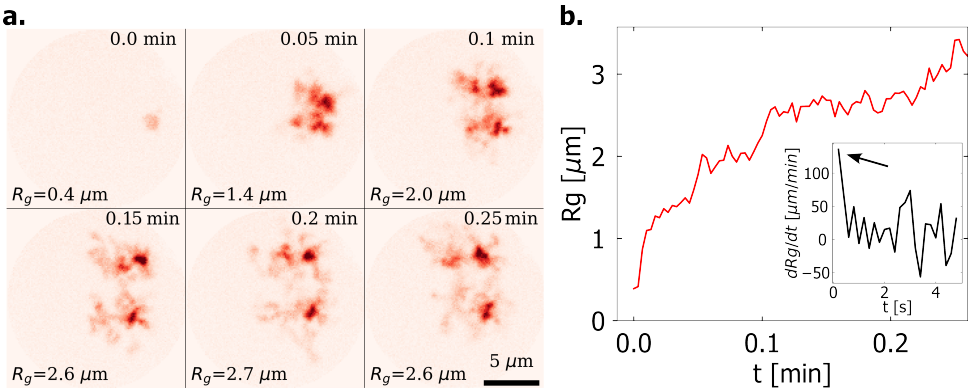

Figure 4.2: Isolated chromosomes expand rapidly within tens of seconds after spheroplast lysis. a) An example of a rapidly expanding nucleoid. Here, two spheroplasts right next to each other lyse simultaneously to give rise to two expanding nucleoids. b) Radius of gyration of an example nucleoid imaged from the onset of lysis. R_g increases from 0.5 μm (DNA in spheroplasts) 6-fold in first 15 seconds after lysis. The speed of expansion reaches 100 μm/ min (inset, black arrow).

μm, diameter 16 μm) in an osmoprotective medium. Next, the spheroplasts were lysed by the introduction of 0.2% solution of IGEPAL that slowly diffused into the traps – while being imaged on a fluorescence microscope (viz. Materials and Methods, Chapter 3). As cells fill traps semi-deterministically, a number of traps contained only a single spheroplast, and this allowed us to image the chromosomes isolated from these cells in detail. To obtain quantitative insights from the imaging data, we developed a python pipeline for semi-supervised object detection and segmentation, as well as characterization of chromosomes' properties (Materials and Methods).

Visually, we noticed that chromosomes expand very rapidly after a lysis event. We therefore first imaged the initial phase of spheroplast lysis and chromosome expansion at high time resolution (Fig. 4.2a). We observed that the size of chromosomes measured by their radius of gyration R_g increased from 0.5 μm (corresponding to the DNA packed within the spheroplast) by factor of ~6-fold within first 15 seconds immediately after lysis. The expansion rate (i.e. the derivative of the radius of gyration over time, dR_g/dt) showed that it can exceed values 100 $\mu m/min$ in the first seconds immediately after expansion onset (Fig. 4.2b, inset).

We observed that trapped chromosomes continue to slowly expand even further after the initial rapid size increase. We were interested in understanding if and when the chromosomes reach a steady state. As fluorescent laser light damages DNA, where the extent of this damage is proportional to the laser power and exposure time, we could not image at high frame rates for periods longer than tens of seconds. We therefore opted for a 1-minute imaging interval, which proved sufficient to capture the long-time chromosome size.

Image analysis corroborated our visual observation and revealed gradual expansion of isolated chromosomes. As observed in Fig. 4.3, the prolonged expansion eventually

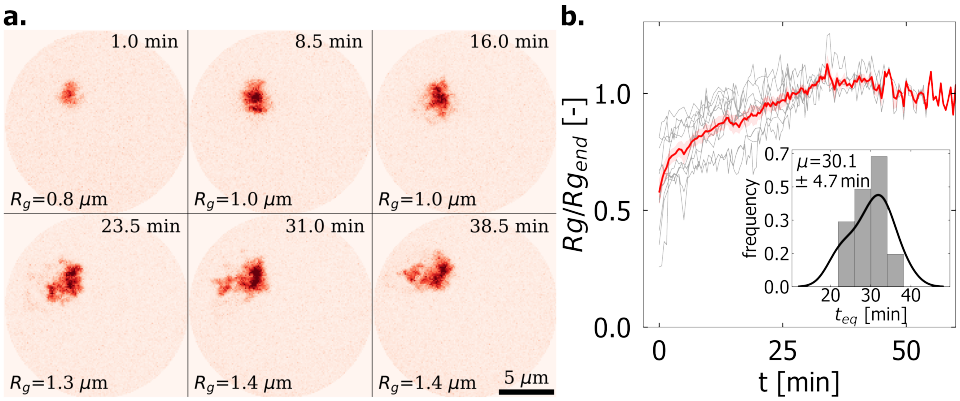

Figure 4.3: Isolated chromosomes continue expanding for 30 minutes after lysis. a) An example of slowly expanding nucleoid. b) Radius of gyration traces for nucleoids expanding over 60 minutes (n=10). Chromosomes expand about two-fold over initial 30 minutes (inset, mean $t_{eq} = 30.1 \pm 4.57$ min, s.d., n=10), after which their size plateaued or slightly decreased. The R_g value was scaled to a value at the end of expansion $R_{g,end}$ to highlight the common trend.

plateaued where the radius of gyration equilibrated. Deducing the time to reach the equilibrium, t_{eq}, as the intersection of piecewise linear fits to the expanding and steady state portions of R_g, we found $\tau_{eq} = 30.1 \pm 4.7$ min (s.d.) (Fig. 4.3b, inset).

Combining these results, we conclude that chromosome expansion in the microfluidic traps consists of two phases: i) a rapid phase associated with about 6-fold increase in size that happens largely in the first 15 seconds after lysis, followed by ii) a slower relaxation associated with about 2-fold increase in size over the next 30 minutes. The rapid phase (i) can be associated with chromosome release from the confinement of spheroplast membrane and the removal of crowding as the cytoplasmic content rapidly diffuses to fill the volume of the trap. The gradual phase (ii) is likely associated with dissociation of some residual fraction of DNA-bound proteins and the gradual resolution of DNA entanglement due to supercoil relaxation.

The observation of a rapid onset of expansion followed by extended slower phase matches well with that of Pelletier and Halvorsen [23] who imaged *E. coli* chromosomes in traps of comparable dimensions, and observed rapid expansion on the scale of tens of seconds, followed up by slower relaxation of 20-30 minutes. In contrast to the work of Pelletier and Halvorsen, however, our experimental platform allowed for the imaging of the bulk of each isolated chromosome in a quasi-2D confinement, offering additional insight into their structure.

We observed that expanded nucleoids exhibited internal inhomogeneities in the intensity, cf. e.g. Fig. 4.4a. Mass spectrometry investigation indicated that the isolated chromosomes can retain trace amounts of residual protein (Chapter 3), and we therefore decided to compare the fluorescence images with molecular dynamics simulation data that included different number of remaining random crosslinks. Here, crosslinks are

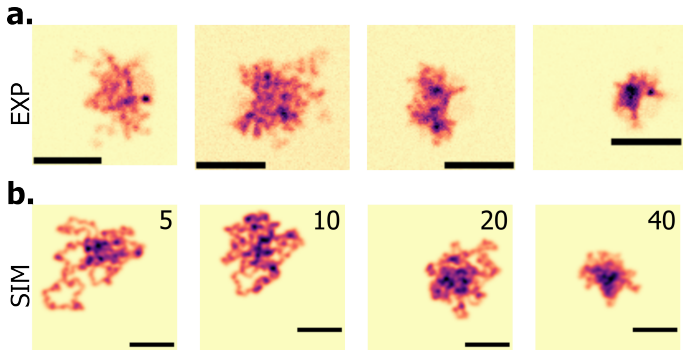

Figure 4.4: Residual crosslinking can explain local DNA density variations. a) Single plane snapshots of experimental data. b) Snapshots from molecular dynamics simulations with n=5, 10, 20 and 40 crosslinkers respectively. Scale bars are 5 μm.

molecules that can bridge DNA in trans, and serve as models for entanglement or residual protein binding (Fig. 4.4). Interestingly, including between only 5 and 20 crosslinkers faithfully recapitulated the internal structural variation that we observed in experimental data. Additionally, including more crosslinkers (as many as 40) allowed to capture the early stages of expansion onset, where chromosomes were transiently more compact.

We sought additional ways how to describe the shape of isolated chromosomes. We noted that their perimeter is highly corrugated, and this led to use the perimeter excess ratio that we defined in Chapter 5 (Chapter 5 – Materials and Methods). Briefly, the perimeter excess ratio, or perimeter ratio, is the ratio of the length of perimeter of the nucleoid to the length of a perimeter of a disk with equivalent surface area (which is defined as πR_g^2 where R_g is the radius of gyration of the nucleoid). Perimeter ratio values larger than 1 characterizes how corrugated the nucleoid shapes are. Calculating the perimeter ratio for both expanding and steady state isolated chromosomes highlighted its ability to follow the changing shape (Fig. 4.5).

For steady state nucleoids, we found the perimeter ratio to have a value of about 3. Specifically it was 3.2 ± 0.4 (s.d., n = 6) for nucleoids imaged for over 30 minutes at 1 minute resolution, and 2.7 ± 0.5 (s.d., n = 19) for nucleoids sampled rapidly at 5 Hz. In both cases the mean coefficient of variation was about 13%. Practically, this means that that the perimeter of expanded nucleoids was about three times as corrugated as would have been the case for a uniform disk with the same radius of gyration.

The sequential imaging of expanding chromosomes highlighted that the chromosomes remained remarkably stable in size for durations exceeding 30 minutes. We were interested in further validating the size of trapped chromosomes. To do so, we imaged already lysed and expanded chromosomes for durations of 60 minutes at 1 minute time resolution. We confirmed that the chromosomes did not visually appear to disintegrate (Fig. 4.6a) and maintain their size (Fig. 4.6b). This is an important feature of the microfluidic experimental platform which can be leveraged to study DNA conformations and effects

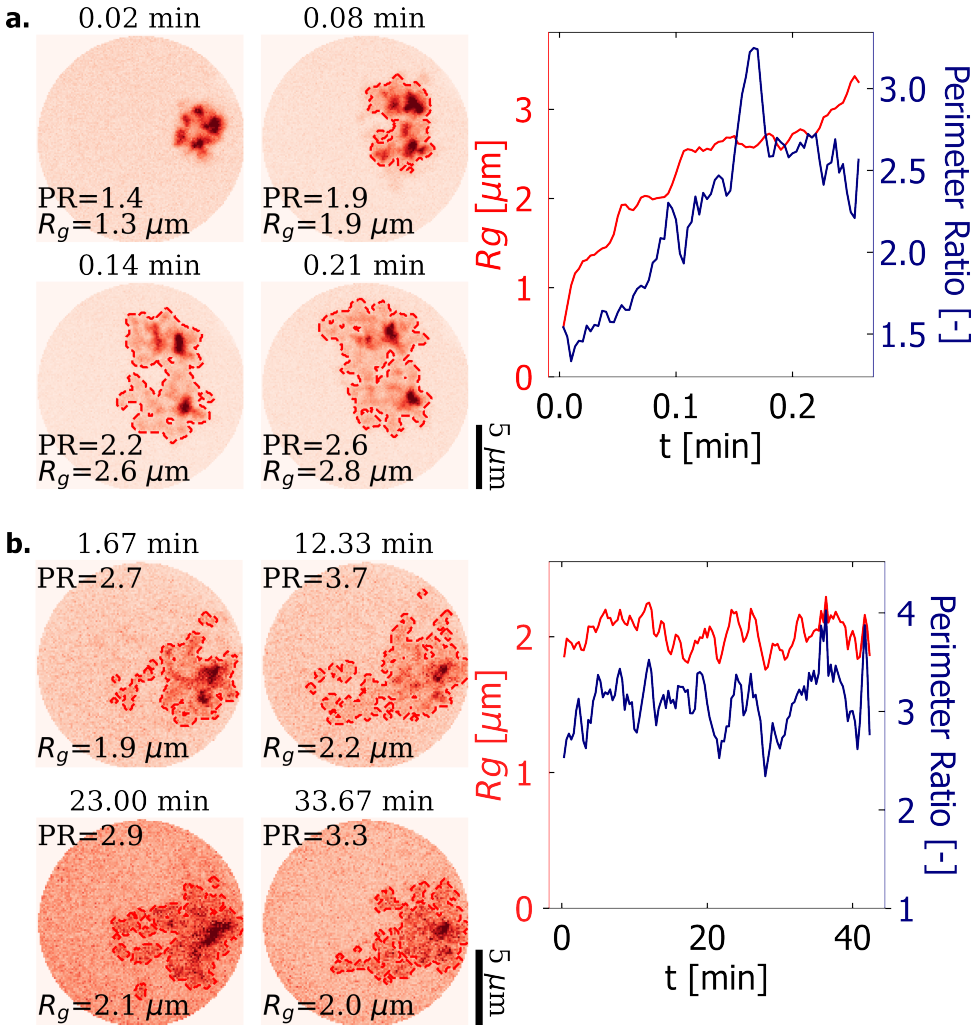

Figure 4.5: Perimeter excess ratio (PR) quantifies the corrugation of chromosome structure. a) Example of a rapidly expanding nucleoid. Visually, its perimeter is becoming more corrugated. The perimeter excess ratio tracks this alongside an increasing radius of gyration (panel on the right). b) Example of a steady-state isolated chromosome. Both the radius of gyration and perimeter excess ratio fluctuate about a mean value; the perimeter excess ratio is more sensitive to shape changes.

of isolated protein factors on megabase DNA organization for extended periods of time.

While we did not visually observe chromosome disintegration at the used sampling rates, we were interested in understanding the possible contribution of single- and double-stranded DNA damage to the gradual expansion. We summed the fluorescence intensity over whole chromosomes after segmentation and background subtraction (Materials and Methods) and observed that it remained largely unchanged for ob-

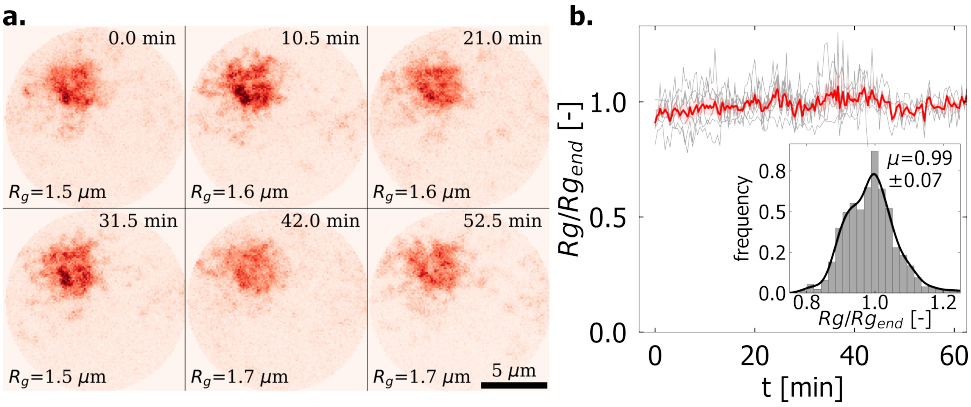

Figure 4.6: Isolated chromosomes are stably trapped and can be imaged for extended period of time. a) Example of expanded nucleoid that maintains its size over 60 minutes. b) Traces of the radius of gyration versus time for n=6 nucleoids. The radius of gyration fluctuates, but on average does not change. Inset shows the time-averaged distribution of scaled radius of gyration $\left\langle R_g / R_{g,\,end} \right\rangle = 0.99 \pm 0.07$ (s.e., n=6).

servation times up to 60 minutes (Fig. 4.7). This indicated that bleaching was not prominent and that the chromosomes were not disintegrating, which would manifest as gradual decrease of the total DNA, and therefore the total sum intensity, due to washout and diffusion of DNA segments and the gradual disappearance of DNA into background. While we cannot exclude that some local DNA nicking takes place, the stability of the sum intensity corroborated visual observation and indicated that the isolated molecules maintain their DNA content.

Next, we quantified the dynamics of isolated chromosomes. Visually, we observed that isolated chromosomes were highly dynamic objects that rapidly explored a range of spatial conformations across time (Fig. 4.8a). We sought to characterize the dynamics of the expanded state. Calculation of the radius of gyration from images of nucleoids acquired at high frame-rate (5 Hz), revealed that it fluctuated around a mean value (Fig. 4.8b). With exception of few outlier datapoints (that are likely caused by switching of our imaging plane in the Z-stack data acquisition) the fluctuations were modest. To quantify this further, we examined the traces on individual level, and found that the coefficient of variation was 0.07 on average.

We characterized the dynamics further by calculating pixel-wise temporal correlations, which revealed that that the individual pixel values were largely uncorrelated, even at sub-second timescales (Fig. 4.9a). Indeed, an exponential fit revealed decay time of 0.38 ± 0.15 s (s.e., n=20). While the DNA object as a whole thus remained relative stable in size as characterized by radius of gyration, local fast fluctuations in DNA density were prominent, at frequencies comparable to or below our sampling rate of 5 Hz.

Finally, we were interested in understanding the timescales involved in larger scale rearrangements of the isolated chromosomes. To gain insight on this, we calculated the temporal correlation of radius of gyration for chromosomes imaged over tens of min-

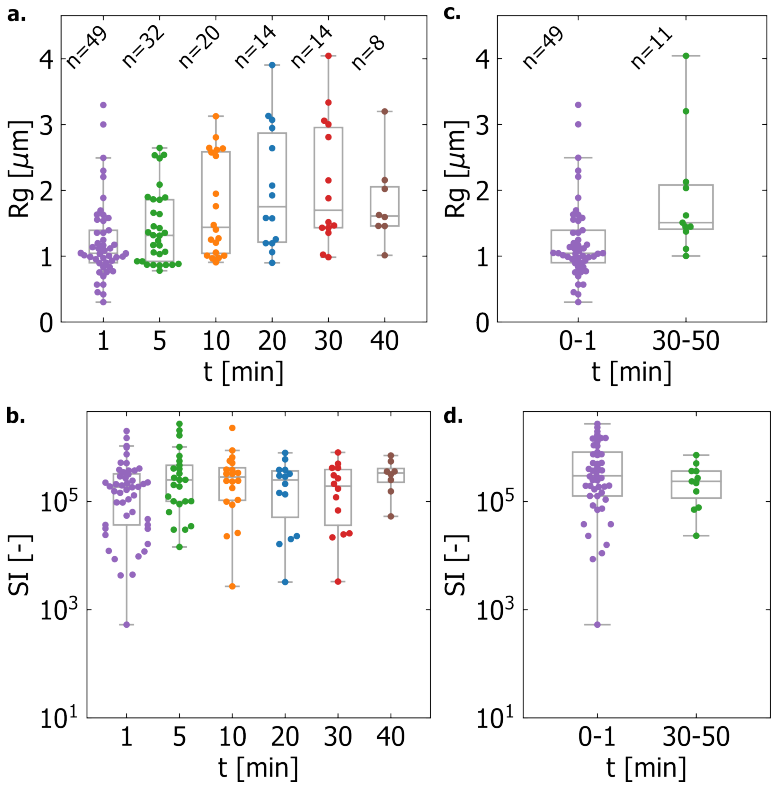

Figure 4.7: Isolated chromosomes expand while maintaining their integrity. a) Population-level radius of gyration for expanding isolated chromosomes versus time. Each point represents a time average of a trace for 2 minute window, except for t=1 min, where the window size was 1 min. b) Same for the sum intensity (SI), which remained largely unchanged. c) Detail of the population level radius of gyration during the first minute of expansion (purple) and during period when full expansion was reached (t > 30 min). Each point represents a trace average over a correspondingly sized window. d) Same for sum intensity. Sample sizes for each distribution are shown in the figures.

utes. We observed that the temporal correlations of radius of gyration decayed on the scale of a few minutes (Fig. 4.9b). An exponential fit revealed the decay time of 5.0 ± 1.8 min (s.e., n=5). This highlights that the overall arrangement of the DNA mass with respect to its center is significantly more stable that fine-scale dynamics, and changes only on the scale of ~5 minutes.

4.3. CONCLUSION AND OUTLOOK

In this study, we investigated the details of the expansion and equilibration process for *B. subtilis* chromosomes isolated in quasi-2D-shaped (pancake-like) microfluidic traps. We observed that the genomic DNA expanded rapidly immediately after lysis, with a rate of increase of radius of gyration reaching 100 $\mu m/min$. This allowed the expanding nu-

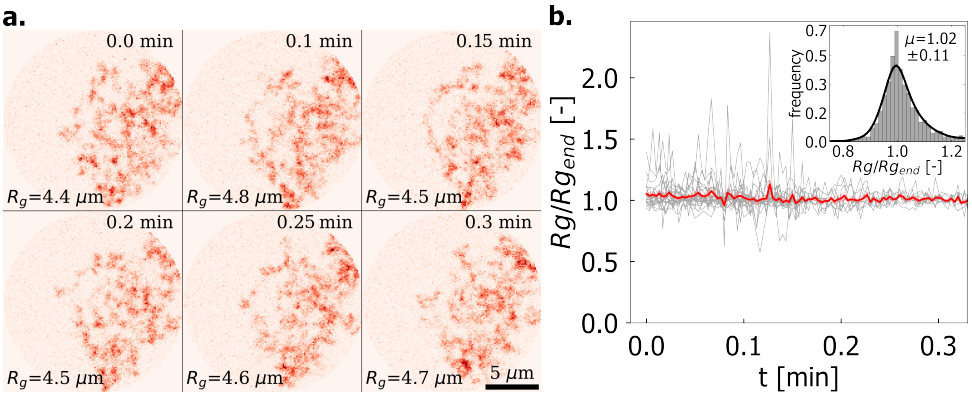

Figure 4.8: Isolated chromosomes show fast scale dynamics. a) An example of isolated chromosome imaged at high frame rate (5 Hz). While its radius of gyration does not change appreciably, the internal structure is found to dynamically reorganize. b) Radius of gyration for n=20 isolated chromosomes captured at a high frame rate (5 Hz). While the mean value of R_g does not change, this does not capture the diffusive behavior and internal conformational changes of the isolated DNA molecules. Inset shows the time-averaged distribution of scaled radius of gyration $\left\langle R_g / R_{g,\,end} \right\rangle = 1.02 \pm 0.11$ (s.e., n=20).

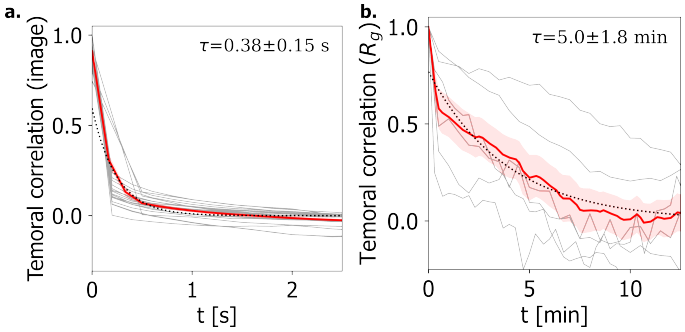

Figure 4.9: Different scales of temporal correlations of isolated chromosomes. a) Pixel-wise correlation of frames acquired at high framerate decays over 380 ± 150 ms, highlighting rapid internal dynamics (n=20). b) Temporal correlation of the radius of gyration of steady state isolated chromosomes that decayed over minutes (exponential decay time $\tau = 5.0 \pm 1.8$ min, s.e., n=6), highlighting that larger rearrangements of the chromosomal mass occur at the minute scale.

cleoid to reach a 6-fold expansion in tens of seconds. Such behavior can be attributed to the rapid disappearance of the mechanical confinement and crowding, and emblematic of DNA behaving as an entropic loaded spring [22]. Continued observation of the expansion process allowed us to uncover a slow relaxation that was accompanied with a further 2-fold size increase that occurred for about 30 minutes after the lysis. This is in agreement with observations in earlier study of isolated *E. coli* chromosomes [23]. We suggest that the gradual expansion phase is likely associated with dissociation of a residual fraction of DNA-bound proteins and the slow resolution of DNA entanglement due to supercoil relaxation.

Next, we demonstrated the utility of this microfluidic trapping assay for time-resolved experiments. Whereas earlier demonstrations of microfluidic trapping succeeded in containing chromosomes, they did so in vertically aligned traps (traps where one of the two longest dimensions is parallel to direction of observation). In such traps, the DNA is constantly moving from and to the plane of observation, complicating studies of its dynamics. Here, we showed a horizontally aligned quasi-2D observation, i.e., traps where both of the two longest dimensions are in the plane of observation, which allowed us to conveniently image most of the megabase-scale DNA in a single frame, even at high magnification (100x). This revealed that fast DNA dynamics occur in the millisecond range. An even lower thickness of the traps (now 1.6 μm) might confine the DNA even better to the focal plane but could also complicate experiments (e.g. in the insertion of spheroplasts). Of course, one can also image at lower magnification (i.e. with longer focal depth) to fully capture all of the DNA within the focal plane.

Finally, whereas earlier experiments with isolated DNA suffered from confounding factors such as temperature and concentration gradients that led to macroscale flow and disturbance to the DNA [25], we here were able to observe individual isolated chromosomes for extended periods (exceeding one hour) without any apparent loss to their integrity. We leveraged this feature to measure the degree to which the chromosome size and shape appear correlated, revealing for example minute-scale decay of the temporal correlations in the radius of gyration. This is indicative of slower internal rearrangements of the DNA mass with respect to its center. We further characterized the shapes by developing quantitative metric of their complexity, the perimeter excess ratio, and demonstrated its utility for expanding and steady state chromosomes. Future studies could seek to describe the shape and internal organization with additional metrics inspired by polymer physics such as the fractal dimension and globule size. Similarly, the process of expansion could be followed with metrics borrowed from in vivo studies, such as characterizing the gradual dissolution of macrodomains and supercoiling domains.

Taken together, this study shows the utility of combining controlled experimental conditions with detailed image analysis for advancing our understanding of properties of isolated megabasescale DNA. Despite clear advantages, our study also has some limitations. Above we already mentioned movement of DNA between imaging planes, and finite sampling rate that was unable to capture the fastest DNA dynamics. Additionally, due to the early stage of this project, where assay development, and not data collection, was the main focus, the sample sizes in this study are rather small. Increasing the sample sizes in future studies, e.g. by increasing number of trapping chambers and placing multiple devices on a single chip, will allow to further support the insights presented in this study. Another approach that we demonstrated here but and will become more powerful as more data becomes available is the corroboration of experiments with polymer dynamics simulations.

4.4. MATERIALS AND METHODS

4.4.1. FLUORESCENCE MICROSCOPY

To image isolated nucleoids in microfluidic traps, we used an Andor Spinning Disk Confocal microscope equipped with 100x magnification oil immersion objective. Isolated DNA was labelled by the intercalating dye Sytox Orange (S11368, Thermo Fischer Scientific, MA, USA) at concentration of 500 nM. At this concentration, the dye is known to reduce the persistence length of DNA to 37 nm [2]. The dye was excited with 561 nm laser line (20% power, 250x gain, 10 ms exposure) with 617/73 filter on the emission. The acquisition computer was running Andor iQ 3.6 software. Multiple z-planes were collected for each object, with a separation of 1 μm between subsequent planes. For extended observations, we defined xy-positions and imaged them repeatedly over time, usually once every 60 seconds. The speed of acquisition at high frame rates was limited by the sum of exposure and readout times, with typically 0.2 seconds for single-plane imaging.

4.4.2. CHROMOSOME IDENTIFICATION AND SEGMENTATION FROM FLUORESCENCE IMAGES

To quantify the properties of DNA molecules in fluorescent images, we wrote a custom Python analysis pipeline. The analysis proceeds in five main steps (Fig 4.10): *i)* identification of trap locations, *ii)* identification of chromosomes inside the traps, *iii)* fitting a foreground mask to the chromosomes, *iv)* inspection and potential adjustments of the masks, and *v)* quantification of relevant observables from inside of the trap (*e.g.,* a calculation of the radius of gyration). Identification of trap locations and fitting a foreground mask happen in semi-automated fashion. First, coarse cropping of traps containing objects of interest is done manually. Second, these image stacks are loaded into an analysis pipeline and user is prompted to click two to three times inside each trap in stack. A circular foreground mask is then fitted to each trap. This mask is then shrunk with a factor 0.95. With our magnification conditions, this shrinkage effectively eliminates signal spillover from traps' edges.

Next, DNA molecules inside the traps were segmented from the background. The procedure is essentially identical to what we described elsewhere [25].First, the raw data in any circular trap was binarized based on a globally determined threshold [26]. Pixels' intensity values were sorted increasingly, and two lines were fitted to such curve as follows: a) a line was fitted to the first half of the pixels in the image (estimate of background), and b) a line was fitted to all pixels brighter than half of the maximum intensity (estimate of foreground). The intensity threshold value was then determined from the point on the sorted intensity curve which was closest to intersection of the two lines. Images before and after background subtraction were inspected and confirmed that the approach was able to discriminate background and foreground well. The crops were then traversed plane-by-plane in z-direction, discarding small regions, dilating remaining region(s) and filling holes. The mask contours were smoothed in each plane with a Savitzky-Golay filter with a window size of quarter the contour length of the mask. Finally, only the most

central 3D contiguous binary object was retained as foreground mask for each object.

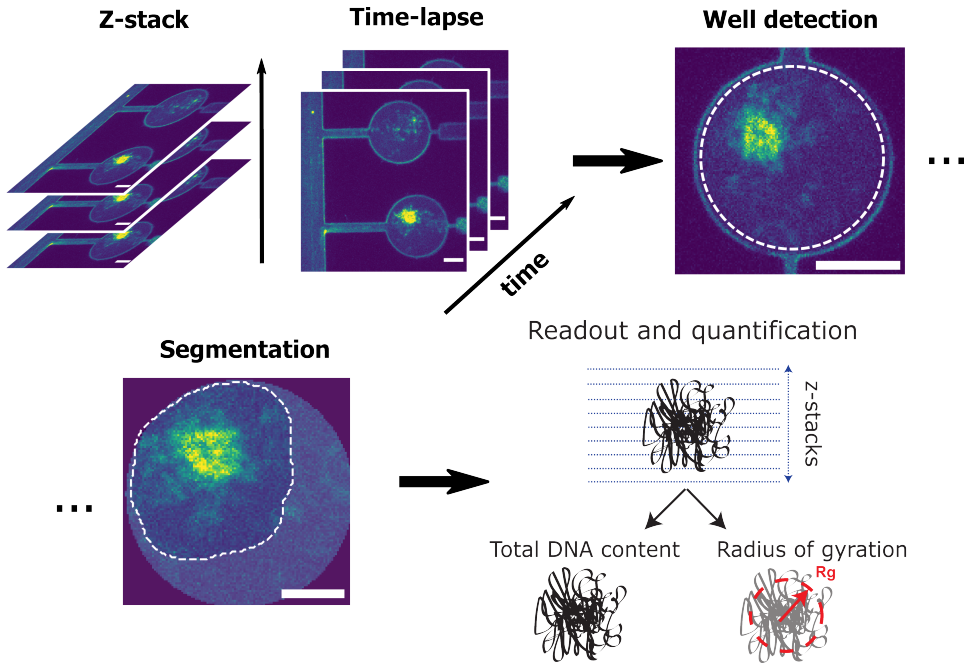

Figure 4.10: Overview of image analysis. Images are acquired as time resolved series of z-stacks. Locations of microwells are fitted and extracted. Isolated chromosomes are identified inside the traps and segmented from the background. The user is prompted to confirm and potentially adjust the masks (not shown). Finally, quantification of relevant observables (e.g. the radius of gyration, ...) is done inside these masks. Scalebars are 5 μm.

4.4.3. SAMPLE SELECTION AND DATA ANALYSIS

Radius of gyration and sum intensity were selected as the main readouts. Sum intensity was calculated as the total sum of all pixel intensities within a foreground mask after subtracting the background. The radius of gyration was calculated by squaring the sum of all foreground pixels' intensity-weighted distances from the object's center of mass, after subtracting the background. The resulting measurements were saved as structured JSON files, one per each field of view (i.e. trap). Sample data were grouped based on frame rate and on the observed behavior (expansion or steady state). To be able to calculate trends from traces with heterogenous sampling frequences, the measurement data (R_g and sum intensity) was interpolated on regular time-sampling interval. For the expansion datasets, the time of the expansion onset ($t = t_0$) was defined in supervised fashion by visual inspection. For reporting a relative radius of gyration, its value was scaled by terminal 10% of its time trace ($R_{g,end}$). To create swarmplots (Fig. 4.6), each measurement trace was aggregated by time-averaging over a window of 2 minutes (unless otherwise specified). The obtained datapoints were used to characterize population level-behaviour, described by the boxplots.

4.4.4. DYNAMICS ANALYSIS

Pixel-wise temporal correlations were calculated on masked and background subtracted data and averaged over whole frame. Radius-of-gyration correlations were calculated from values obtained on masked and background subtracted data.

We obtained the expansion equilibration time as the x-axis value of piecewise linear fit to the normalized expansion time traces. Each of the two parts of the fit was described with relationship $f\left(x^i\right) - f\left(x_0^i\right) = k^i(x^i - x_0^i)$, where x_0^i is the time to reach the equilibrium for the trace i, t_{eq}^i. We used numpy's function 'piecewise' to construct the bilinear, and scipy's 'curve_fit' to fit it to the data.

4.4.5. MOLECULAR DYNAMICS SIMULATIONS

To test whether our images would be consistent with a simple polymer model, we compared the images to molecular dynamics simulations of a circular polymer using the polychrom wrapper for OpenMM. Simulations were carried out by Janni Harju at VU Amsterdam. In these simulations, the chromosome is represented as a bead-spring polymer, with finite excluded volume interactions that allow occasional strand passage (with an overlap penalty of 5 $k_B T$). This allowed to sample steady state configurations more efficiently. Since the bacterial chromosome is circular, we initially simulated a ring polymer in a cylindrical confinement, with the aspect-ratio of the confinement similar to experiments (height 1.2 μm, diameter 16 μm), where the lower height has been chosen to account for negative wall potential. We chose a coarse-graining scale of 1 kb, and set the monomer length to 90 nm to match the experiments' steady-state radius of gyration of 4 μm. We let simulations converge to a steady state, and evaluated the end result by plotting the mean radius of gyration as a function of simulation time. Once simulations converged, we constructed images of the simulated polymer configurations by taking a 3D convolution of the configuration on the middle z-plane of the confinement, based on the resolution of our experimental system. We note that only images with crosslinkers that connect random points along the chromosome recapitulated the experimental data. When we tested crosslinkers that specifically linked chromosome arms (mimicking the action of bacterial condensin), these produced effectively linear polymer configuration that we did not observe experimentally. The absence of crosslinkers led to chromosome configurations of an open ring, which we also did not observe experimentally.

4.4.6. CODE AND DATA AVAILABILITY

The python analysis code used for processing images, analyzing data, and creating figures in this manuscript is available at Zenodo (ID: 13369941). Data is available from the corresponding author upon reasonable request.

4.5. REFERENCES

[1] Marko, J. F. *Biomechanics in Oncology.* (Springer Science+Business Media, New York, NY, 2018).

[2] Davidson, I. F. *et al.* CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature* **616**, 822–827 (2023).

[3] Rubinstein, M. & Colby, R. H. Polymer Physics.

[4] Flory, P. J. Thermodynamics of High Polymer Solutions. *The Journal of Chemical Physics* **10**, 51–61 (1942).

[5] Simon Grosse-Holz, Antoine Coulon, & Leonid Mirny. Scale-free models of chromosome structure, dynamics, and mechanics. *bioRxiv* 2023.04.14.536939 (2023).

[6] Stonington, O. G. & Pettijohn, D. E. The Folded Genome of *Escherichia coli* Isolated in a Protein-DNA-RNA Complex. *Proc. Natl. Acad. Sci. U.S.A.* **68**, 6–9 (1971).

[7] Hecht, R. M., Taggart, R. T. & Pettijohn, D. E. Size and DNA content of purified E. coli nucleoids observed by fluorescence microscopy. *Nature* **253**, 60–62 (1975).

[8] Flink, I. & Pettijohn, D. E. Polyamines stabilise DNA folds. *Nature* **253**, 62–63 (1975).

[9] Gosule, L. C. & Schellman, J. A. Compact form of DNA induced by spermidine. *Nature* **259**, 333–335 (1976).

[10] Giorno, R., Stamato, T., Lydersen, B. & Pettijohn, D. Transcription in vitro of DNA in isolated bacterial nucleoids. *Journal of Molecular Biology* **96**, 217–237 (1975).

[11] Pettijohn, D. E. Interactions Stabilizing DNA Tertiary Structure in the Escherichia coil Chromosome Investigated with Ionizing Radiation. (1978).

[12] Odijk, T. Osmotic compaction of supercoiled DNA into a bacterial nucleoid. *Biophysical Chemistry* **73**, 23–29 (1998).

[13] Cunha, S., Odijk, T., Süleymanoglu, E. & Woldringh, C. L. Isolation of the Escherichia coli nucleoid. *Biochimie* **83**, 149–154 (2001).

[14] Cunha, S., Woldringh, C. L. & Odijk, T. Polymer-Mediated Compaction and Internal Dynamics of Isolated Escherichia coli Nucleoids. *Journal of Structural Biology* **136**, 53–66 (2001).

[15] Cunha, S., Woldringh, C. L. & Odijk, T. Restricted diffusion of DNA segments within the isolated Escherichia coli nucleoid. *Journal of Structural Biology* **150**, 226–232 (2005).

[16] Zimmerman, S. B. Cooperative transitions of isolated Escherichia coli nucleoids: Implications for the nucleoid as a cellular phase. *Journal of Structural Biology* **153**, 160–175 (2006).

[17] Wegner, A. S., Alexeeva, S., Odijk, T. & Woldringh, C. L. Characterization of Escherichia coli nucleoids released by osmotic shock. *Journal of Structural Biology* **178**, 260–269 (2012).

[18] Yoshikawa, Y. *et al.* Critical behavior of megabase-size DNA toward the transition into a compact state. *J. Chem. Phys.* 8 (2011).

[19] Zinchenko, A. *et al.* Single-molecule compaction of megabase-long chromatin molecules by multivalent cations. *Nucleic Acids Research* **46**, 635–649 (2018).

[20] Zinchenko, A. Compaction and self-association of megabase-sized chromatin are induced by anionic protein crowding. *Soft Matter* 8 (2020).

[21] Jun, S. & Mulder, B. Entropy-driven spatial organization of highly confined polymers: Lessons for the bacterial chromosome. *Proceedings of the National Academy of Sciences* **103**, 12388–12393 (2006).

[22] Jun, S. & Wright, A. Entropy as the driver of chromosome segregation. *Nat Rev Microbiol* **8**, 600–607 (2010).

[23] Pelletier, J. *et al.* Physical manipulation of the Escherichia coli chromosome reveals its soft nature. *Proceedings of the National Academy of Sciences* **109**, E2649–E2656 (2012).

[24] Freitag, C. *et al.* Visualizing the entire DNA from a chromosome in a single frame. *Biomicrofluidics* **9**, 044114 (2015).

[25] Holub, M. *et al.* Extracting and characterizing protein-free megabase-pair DNA for in vitro experiments. *Cell Reports Methods* 100366 (2022).

[26] Vtyurina, N. N. *et al.* Hysteresis in DNA compaction by Dps is described by an Ising model. *Proc Natl Acad Sci USA* **113**, 4982–4987 (2016).

5

CONDENSIN-MEDIATED DNA COMPACTION ON THE MEGABASE SCALE

SMC proteins such as condensin and cohesin are essential for chromosome organization. While extensive in vitro studies have been done on their role in structuring DNA at the kilobase scale, such in vitro studies have been lacking at larger scales. Here, we employ a microfluidic assay to investigate condensin's interaction with megabase-sized DNA scaffolds. We observe an ATP-dependent DNA compaction which occurs with a rate that increases with condensin concentration. Unfortunately, protein-mediated surface interactions induce local spots of DNA aggregation at the surface which hinders a detailed analysis. We develop novel metrics to quantify compaction despite this challenge. Our findings provide a starting point for future studies aimed at understanding condensin's role in chromosome architecture and exploring applications in bottom up biology and genome engineering.

Next to myself, various other people contributed to the content of the work described in this chapter, namely Alex Joesaar, Leander Lutze, Jacob Kerssemakers, Janni Harju, Cees Dekker.

5.1. INTRODUCTION

DNA represents a molecular program for every cell. Understanding how this program is read off and interpreted requires studying the DNA-binding proteins that interact with the DNA. Among these, one class that our lab worked on extensively over the recent years are the structural maintenance of chromosome (SMC) proteins. These SMC proteins such as cohesin and condensin are loop-extruding factors that are key to the spatiotemporal organization of chromosomes as they are involved in number of fundamental processes such as sister chromatid cohesion and interphase genome organization (cohesin), chromosome compaction and segregation (condensin) [1,2], neuronal progenitor cell differentiation [3], and antibody diversification [4]. Members of this class of proteins are present both in prokaryotes [5,6] and eukaryotes, but play more important and diverse set of roles in the higher domain of life.

SMC proteins have been the topic of large body of research over the past decade, yielding advanced understanding of their function [2]. Experiments in living cells, e.g. using conformation capture methods and super-resolution imaging, revealed many molecular interactions and chromosome organization patterns [7,8]. In-vitro studies, including single-molecule experiments, allowed to measure the molecular motor properties of these protein complexes, such as their DNA-loop-extrusion speed [9], ability to bypass barriers [10], and interaction with regulatory proteins and RNAs at the mechanistic level [11,12].

However, there are only very limited number of in vitro studies of how SMC proteins act mechanistically on scales beyond a few tens of kbs [13]. This is striking, given their purported importance in structuring chromosomes at the 0.1-10 Mbp scale [14], but can be understood in view of the lack of tools to do so. Cellular differentiation, for example, is a nearly universal process that relies on long-distance gene expression regulation, a process where SMC proteins are thought to play an important role [15]. Clearly, more studies are needed to improve our understanding of phenomena at these scales.

We previously established a 'genome-in-a-box' approach (Chapter 3), a microfluidic-based assay for whole-chromosome in-vitro reconstitution, and demonstrated its applicability for studying the conformational dynamics of individual bacterial chromosomes (Chapter 4). Here, we apply the assay to investigate the DNA-binding and loop-extrusion properties of yeast (Saccharomyces cerevisiae) condensin (further denoted as 'condensin') at the scale of millions of basepairs. Motivated by the promise of this assay to tackle open questions about the role of SMCs at these scales, we can ask multiple questions: What is the structure of a megabase-scale protein-free DNA scaffold after loop extrusion? Is there a common size and structure that the DNA molecules converge to? What is the average size of extruded loops. How variable is this size? And what is the speed of compaction and how does it depend on condensin concentration?

To our knowledge, this represents the first-of-its-kind investigation of condensin interactions with a megabasepair DNA scaffold that was conducted at single-molecule level and in well-defined in vitro conditions. While the lack of time and experimental chal-

lenges hindered our ability to fully answer all the questions we originally posed, we nevertheless anticipate that the results will be valuable in designing future experiments that will not only shed light on the protein's role in living cells but will also be valuable for whole-genome engineering approaches and bottom-up biology.

5.2. RESULTS AND DISCUSSION

For these experiments, we leveraged our previously established microfluidic trapping platform (Chapter 3), and built upon the characterization of bare bacterial chromosomes isolated in these devices (Chapter 4), to investigate the effect of the loop-extruding factor condensin on megabase-scale isolated deproteinated DNA (derived from individual chromosomes from *B. subtilis* cells).

As a first step in this study, we sought to characterize the effect of low concentration of condensin (40 nM) in the absence of ATP. We microfluidically contained bacterial cells in quasi-2D traps (height 1.6 µm, diameter 16 µm) and lysed them with low osmolarity buffer with 0.2% IGEPAL to release individual chromosomes. Next we added condensin by flushing it through the device and allowing it to diffuse to the traps. We observed that the isolated deproteinated chromosomal DNA did not significantly recruit condensin and that it did not exhibit any compaction (Fig. 5.1). The radius of gyration of the isolated chromosomes also remained constant within errors (Fig. 5.1a - top, 5.1b). We did not detect any appreciable recruitment of condensin to the DNA (Fig 5.1a - bottom, 5.2a). Therefore, while reports have shown condensin' s ability to bind DNA in the absence of ATP [16], the effect remained below our limit of detection at these experimental conditions.

Next, we investigated how the presence of ATP modulated condensin's effect at this concentration. To do so, we included 10 mM ATP in the experimental buffer with condensin and imaged isolated chromosomes for over 20 minutes. In contrast to the experiment where no ATP was present, we now observed that the DNA fluorescence intensity slightly increased (Fig. 5.3a,c) for a few minutes, while the radius of gyration did not show any appreciable change over time (Fig. 5.3b), Quantification of the condensin imaging channel revealed an intensity that was increasing over time, which pointed to recruitment of condensin to the DNA (Fig. 5.2b, 5.3b). While the effects were small, and at our limit of quantification, they nevertheless pointed to the ATP-dependent action of condensin in our assay, which we decided to investigate further.

As the ATP concentration is in large surplus at 10 mM and not limiting to the action of condensin, we increased the protein concentration while keeping the concentration of ATP constant to further understand the protein's effect. At 200 nM we observed the DNA becoming progressively more compact over the course of 15 minutes, as well as recruitment of condensin molecules to it (Fig. 5.4a), which was corroborated by an increasing intensity in the condensin imaging channel (Fig. 5.2c). Measuring the radius of gyration revealed that it decreased by ~40% with respective to the value at the start of the experiment (mean relative radius of gyration 0.60 ± 0.10, s.d. n=4; Fig. 5.4b).

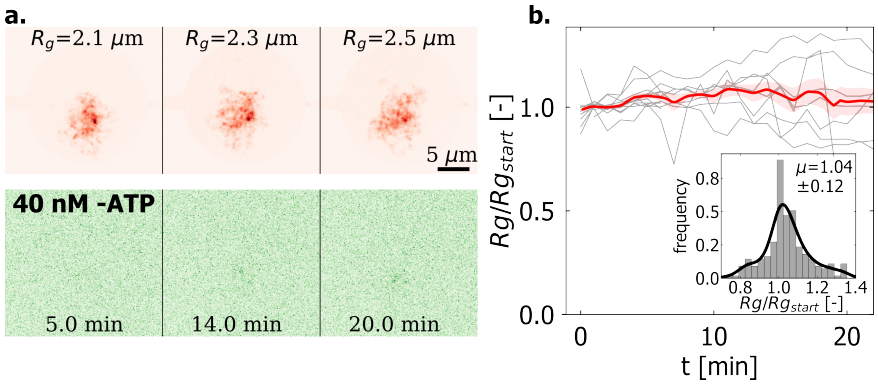

Figure 5.1: Condensin does neither bind nor compact genomic DNA in the absence of ATP. a) Example time-lapse of DNA (top, red) and condensin (bottom, green) channels. No significant condensin binding is seen. b) Time traces of radius of gyration for n=12 nucleoids. Rg does not change significantly from its value at the start of imaging. Inset shows the time-averaged distribution of the scaled radius of gyration as histogram and kernel density estimate, $\langle R_g/R_{g,\,start}\rangle = 1.04 \pm 0.12$ (s.d., n=12).

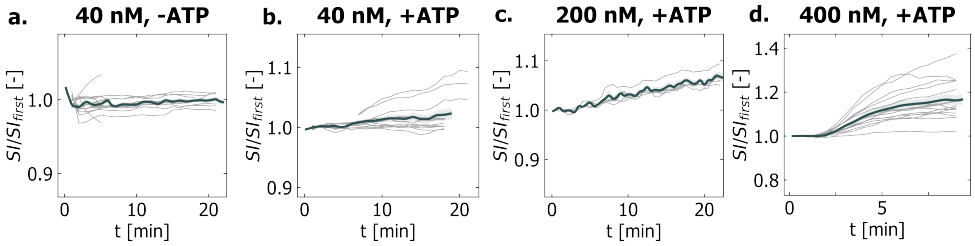

Figure 5.2: Condensin fluorescence intensity for experimental conditions tested in this study. Condensin shows recruitment to DNA over time at all concentrations where ATP was present. Before taking the measurement of intensity, all images were segmented with the same mask as the DNA channel, and background was subtracted.

At the highest tested concentration of 400 nM, the DNA compacted rapidly, within a few minutes. Simultaneously, its maximum fluorescence intensity increased suggesting locally increased DNA concentration (i.e. compaction). The imaging at the condensin channel showed its concurrent recruitment to the DNA (Fig. 5.5a). Averaging over n=9 chromosomes revealed an almost step-wise change of radius of gyration that was happening within two minutes (Fig. 5.5b). Surprisingly, the final degree of compaction was relatively modest, with the mean relative radius of gyration at the end of expansion of 0.82 ± 0.09 (s.d., n=9). This corresponded to less than 20% compaction, which is a more modest reduction than what observed a lower concentration. This did not agree with our expectation the DNA should become more compact as more molecules of loop-extruding condensin are allowed to process it. We turned to clarifying this apparent contradiction.

Detailed inspection of the fluorescence images revealed that that nucleoids become entirely immobile shortly after their compaction onset. While initially allowed to compact

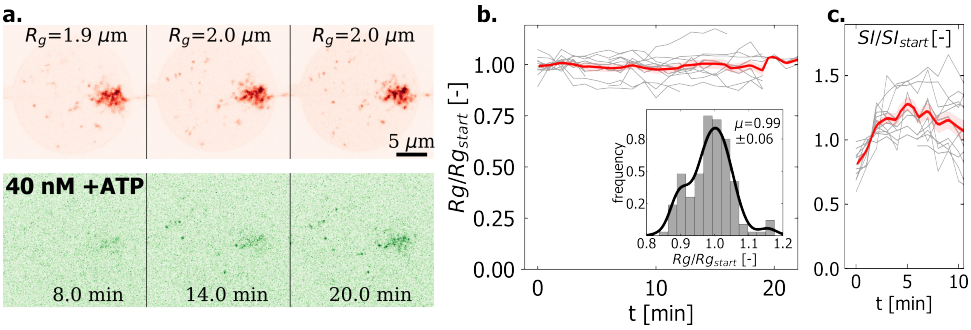

Figure 5.3: Genomic DNA versus time in the presence of ATP and 40 nM condensin. Over time, condensin accumulates on the DNA. a) Example timelapse of DNA (top, red) and condensin (bottom, green) channels. DNA maintains its size. Some degree of condensin binding is evident. b) Time traces of the relative radius of gyration for n=11 nucleoids. Rg does not change significantly from its value at the start of imaging. Inset shows time-averaged distribution of scaled radius of gyration $\langle R_g/R_{g,\,start}\rangle = 0.99 \pm 0.06$ (s.d., n=11) as histogram and kernel density estimate. c) Sum of DNA fluorescence intensity. The intensity is seen to increase over initial 5 minutes suggesting DNA is becoming more concentrated (n=12).

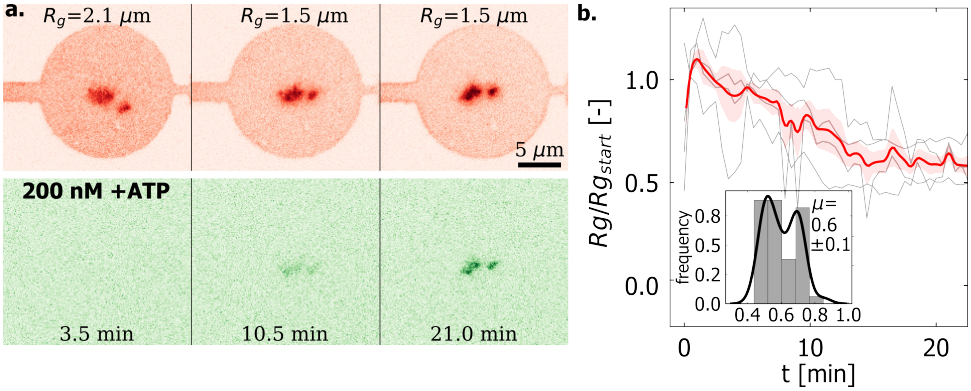

Figure 5.4: Genomic DNA versus time in the presence of ATP and 40 nM condensin. Over time, condensin accumulates on the DNA. a) Example timelapse of DNA (top, red) and condensin (bottom, green) channels. DNA maintains its size. Some degree of condensin binding is evident. b) Time traces of the relative radius of gyration for n=11 nucleoids. Rg does not change significantly from its value at the start of imaging. Inset shows time-averaged distribution of scaled radius of gyration $\langle R_g/R_{g,\,start}\rangle = 0.99 \pm 0.06$ (s.d., n=11) as histogram and kernel density estimate. c) Sum of DNA fluorescence intensity. The intensity is seen to increase over initial 5 minutes suggesting DNA is becoming more concentrated (n=12).

locally, any further compaction appeared to be hindered by DNA's inability to be pulled together Experiments with washing the immobile molecules with a range of buffers (including high salt and hexandiol), where the molecules remain intact, demonstrated that the protein-DNA complex formed an aggregate that became irreversibly attached to the device's surface. The aggregate could be removed only by incubation with DNAse. Clearly, while the current surface treatment (lipid bilayer) was adequate for experiments in the absence, or at a low concentration of, exogenous protein, it necessitates further optimization for experiments at high protein concentrations.

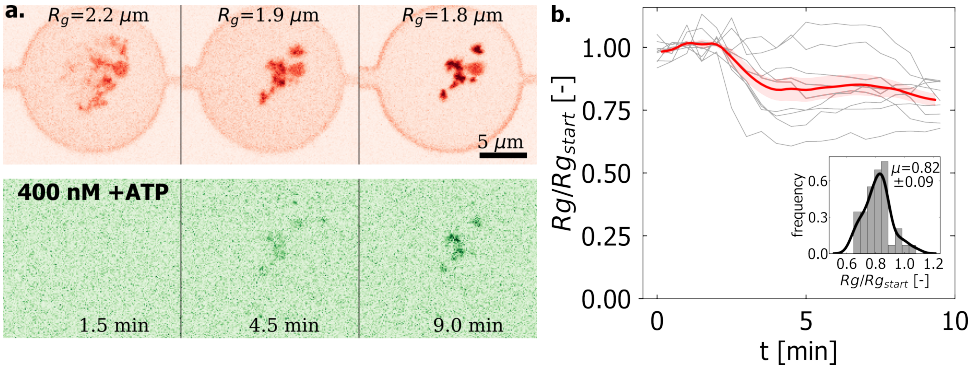

Figure 5.5: Condensin at high concentration (400 mM) compacts genomic DNA within a few minutes in the presence of ATP. a) Example timelapse of DNA (top, red) and condensin (bottom, green) channels. DNA visibly compacts, and is seen to bind condensin. b) Time traces of radius of gyration for n=9 nucleoids. Rg drops by about 20% within about two minutes from the compaction onset. Inset shows the distribution of scaled radius of gyration averaged over last 2.5 minutes, $\langle R_g/R_{g,\,start}\rangle = 0.82 \pm 0.09$ (s.d., n=9), as histogram and kernel density estimate.

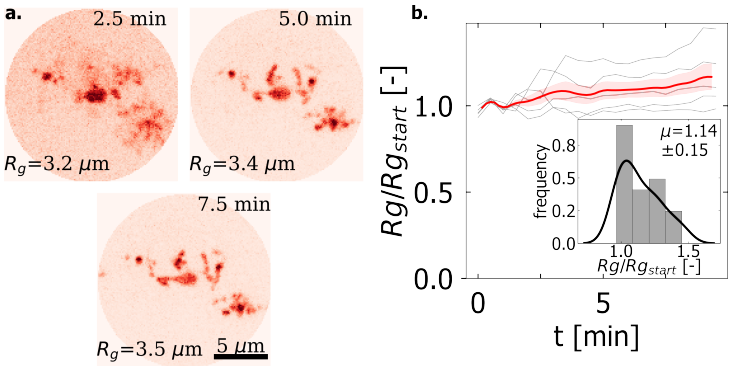

Figure 5.6: Radius of gyration fails to capture the local nature of condensin-driven DNA compaction. a) Example timelapse of condensin driven compaction of isolated nucleoid in a microfluidic trap. The DNA clearly becomes locally more compact. Radius of gyration is in fact seen to *increase* versus time. b) Time traces of scaled radius of gyration for n=5 nucleoids, with the inset showing distribution averaged over minutes 7.5 to 10 as histogram and kernel density estimate. Radius of gyration increases to $\langle R_g/R_{g,\,start}\rangle = 1.14 \pm 0.15$ (s.d., n=5).

We note that at the experimental conditions tested in the study, each microfluidic well ($V_{well} = \pi R^2 h \approx 320\ \mu m^3$) can contain up to the order of ~5'000 and 50'000 condensin molecules, at 40 nM and 400 nM, respectively. In practice this number is likely (very much) lower, due to the large surface to volume ratio of the microfluidic device, and consequent abundant areas for protein to stick to. Furthermore, the onset of condensation happens before the equilibrium concentration establishes in the microfluidic well, and therefore the number of proteins acting on the DNA will be lower than the number predicted by these upper bounds.

We further analysed the imaging data which showed an increase of radius of gyration, despite apparent condensation at the conditions where we saw surface sticking (Fig. 5.6). Clearly, this indicated that radius of gyration becomes a poor metric for characterizing compaction that is driven locally. Especially, if, as it was case here, the DNA mass is prevented to be driven to the chromosome's centre. Importantly, such behaviour could occur as well in living cells, e.g. due to compartmentalization and phase separation. We therefore developed new metrics to characterize DNA behaviour in these conditions.

We sought metrics for characterizing the structure of megabase-scale isolated DNA under the effect of exogenous proteins, that would capture condensation features that the radius of gyration could not. We based these metrics on two features that we observed in the fluorescence images. Namely, a) as the DNA changed shape, the complexity of its perimeter (i.e. how corrugated it is) tended to change; and b) the number of pixels with a high intensity in the DNA images tended to change too. Hence we developed two metrics: i) the 'perimeter excess ratio' (or perimeter ratio), and ii) the 'share of bright pixels' (or bright ratio).

The perimeter excess ratio (cf. Materials and Methods) is the defined as the ratio of the length of a perimeter of the nucleoid, to the length of a perimeter of a hypothetical disk, where the disk dimension is such as to have the same radius of gyration as the nucleoid. A disk with given radius of gyration has well defined area ($A = \pi R_{disk}^2 = 2\pi R_g^2$). Choosing a threshold on the nucleoid fluorescence image that creates a binary image that satisfies this area defines a perimeter. The length of the perimeter can be calculated from the discretized image by computing the nearest-neighbor-distances of its edge pixels. This can, in turn, be compared to the one obtained on the hypothetical disk. The more a shape deviates from a disk, the longer its perimeter will be, and hence the perimeter excess ratio increases. This metric is useful to quantitatively compare conditions where one observes changes to chromosomes shapes. Indeed, we applied the metric also to nucleoids that are expanding after lysis of cell wall, and observed the perimeter ratio captured the expansion process, as the nucleoids became less sphere-like and more corrugated (cf. Chapter 4).

Note that while we have tied the definition of perimeter excess ratio to the radius of gyration, that need not to be the case. Similar metrics that relate object's area to its perimeter, from the fields of economics [17] and digital image processing [18], could be applicable as well. Comparing the perimeter ratio, or equivalent metrics, across magnification or coarse graining scales should allow to study whether the isolated nucleoids exhibit fractal nature.

The second metric, the share of bright pixels is the fraction of pixels that have an intensity above a set threshold value (Materials and Methods). The metric tracks the increasing local density of DNA upon compaction, which manifests as locally increasing fluorescence counts. Setting the threshold constant (here, the background value plus two s.d.) across the entire time lapse allows to follow changes that happen over time.

Figure 5.7 presents the perimeter excess ratio and the share of bright pixels metrics for nucleoids that showed an increase in the radius of gyration (Fig 5.6). Both metrics sig-

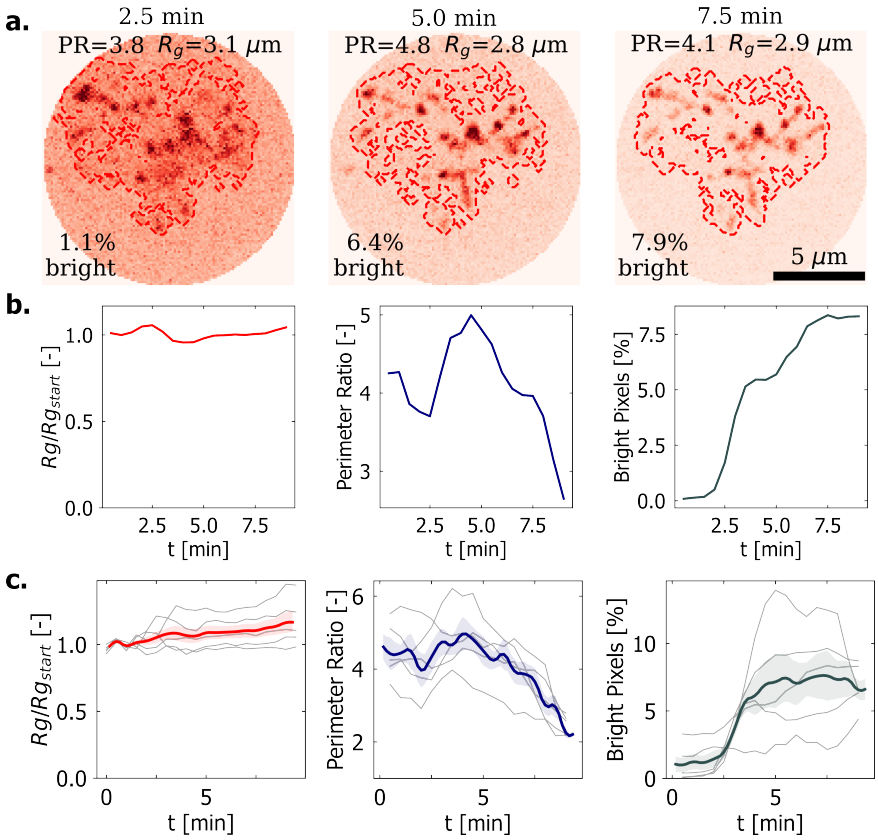

Figure 5.7: Perimeter excess ratio and share of bright pixels accurately capture the local DNA compaction driven by condensin. a) Example snapshots of a nucleoid being compacted by 400 nM condensin in the presence of ATP. PR – perimeter ratio, bright – share of bright pixels. The background becomes relatively less bright as the intensity has been adjusted per frame for visualization. b) Radius of gyration, perimeter excess ratio, and share of bright pixels for the nucleoid in a). While radius of gyration remains largely constant, both the perimeter excess ratio and the bright ratio change, as a consequence of *local* compaction. c) Time traces of the radius of gyration, perimeter excess ratio, and bright ratio for the nucleoids where Rg increased from Fig. 5.6b (n=5).

naled the compaction (unlike the radius of gyration). The perimeter excess ratio decreased, which described the gradual 'packing in' of expanded DNA. The share of bright pixels, on the other hand, increased, which tracked what we observed visually, i.e. a locally increasing DNA density (Fig 5.7b,c). Interestingly, while the share of bright pixels tracked the speed of compaction, the perimeter excess ratio evolved more slowly for some nucleoids. This is likely due to finite background. As faint nucleoid parts become more bright and defined above the level of intensity background due to ongoing compaction, the perimeter becomes more corrugated and therefore longer,.

Having demonstrated that the perimeter excess ratio and the share of bright pixels can handle challenging examples of nucleoid compaction, we sought to apply these metrics

to the full set of experimental conditions that we have studied. As expected, both metrics remained mostly constant at conditions with low condensin concentration (Fig. 5.8a,b). On the other hand, they robustly tracked the changes associate with compaction. The perimeter ratio decreased, highlighting the extended portions of DNA molecules being tied in. Concomitantly, the bright ratio increased, highlighting local DNA density increase manifested by higher pixel intensity. Importantly, these metric performed well not only in the condition with 200 nM condensin (Fig. 5.8c), where also radius of gyration behaved predictably, but also in the condition with 400 nM condensin concentration, where strong interaction with the surface in some samples made the radius of gyration readout uninformative (Fig. 5.7d). Taken together these results demonstrate the utility of these two metrics to characterize the structure of megabase-scale isolated DNA under the effect of exogenous proteins.

While the metrics identified until now are clearly useful to describe the compaction we observed, they are not able to capture emerging internal structure. While this was not fully applicable to our studies due to surface mediated sticking, we wanted to develop a metric for future experiments. To do so, we furthermore plotted the fluorescence intensities as function of distance from nucleoid's center of mass. Plotting these curves over time for the case of rapid compaction (400 nM condensin + ATP) (Fig 5.9), revealed that while the radius of gyration remains constant, the profiles became increasingly more pronounced and thus captured the changes that we observed visually, i.e. the nucleoid becoming more defined and brighter. Additionally, the fact that the intensity peak after the compaction is located off-center (i.e. not at center of mass) indicates emergence of structure other than DNA simply being pulled together to one blob. We applied this metric across the range of studied conditions (Fig 5.10). We found that the profiles did not change over time at low condensin concentration (Fig 5.10a,b), where no compaction happened. Conversely, the profiles became significantly more pronounced, at higher concentration both in the presence and absence of the case of surface-mediated sticking (Fig 5.10c,d).

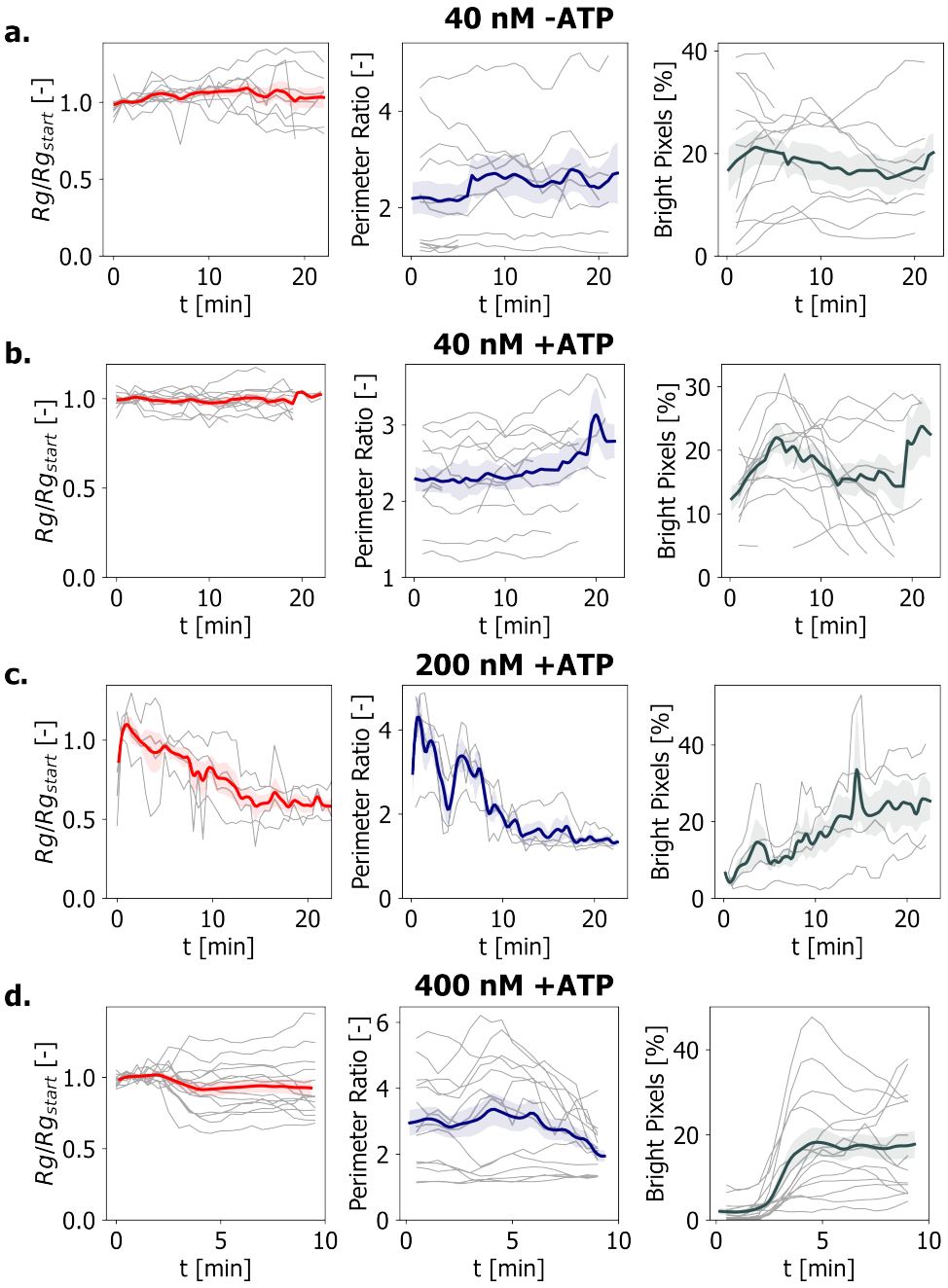

Figure 5.8: Radius of gyration, perimeter excess ratio and share of bright pixels discriminate between conditions with and without condensin-driven DNA compaction. All radius of gyration, perimeter excess ratio and share of bright pixels remain, on average, constant in the absence of compaction (a, b). For condensin-driven DNA compaction, however, the radius of gyration and perimeter excess ratio are seen to decrease (c, d), whereas the share of high intensity pixels increases. Concentrations for these panels were as follows: a) 40 nM condensin -ATP, n=12 (cf. Fig 5.1), b) 40 nM condensin +ATP, n=11 (cf. Fig. 5.2), c) 200 nM condensin +ATP, n=4 (cf. Fig 5.3), d) 400 nM condensin +ATP, n=14 (cf. Fig 5.4-5).

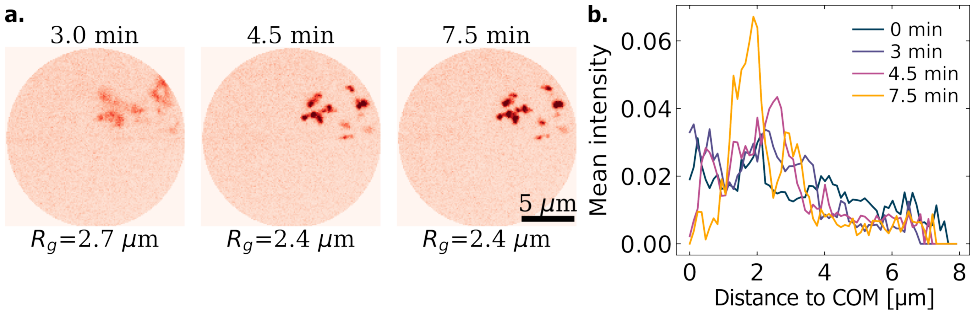

Figure 5.9: Temporal evolution of radial intensity profile captures gradual compaction. a) Timelapse of a compacting nucleoid (400 nM condensin + ATP). While the DNA is seen to become more defined and brighter, the radius of gyration (numbers denoted as insets) captures this only modestly (10% change). b) Temporal evolution of the radial intensity distribution centered at the nucleoids center of mass (COM) reflects the visual observation, with more defined and pronounced intensity peak emerging over time.

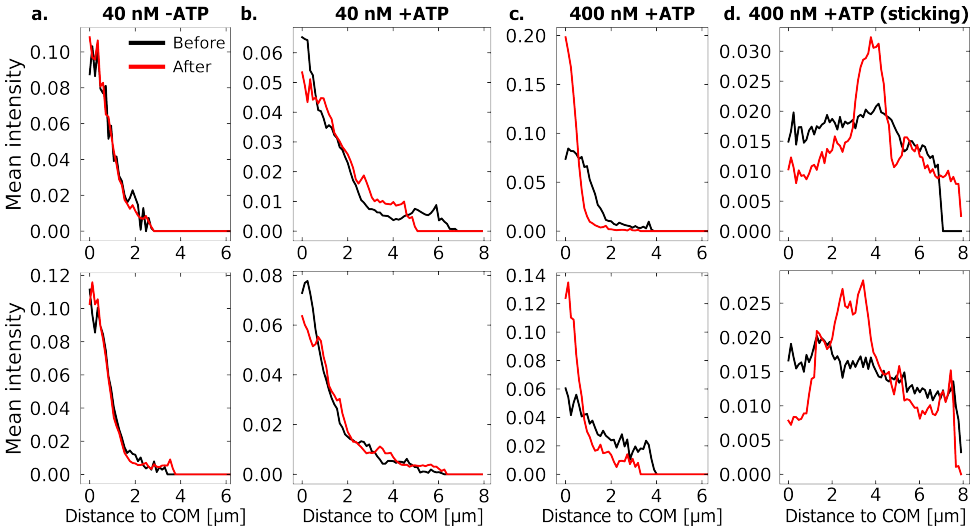

Figure 5.10: Intensity versus distance to center of mass (COM). Examples for two representative nucleoids (top and bottom) for each condition. While the spatial intensity distribution with respect to the center of mass of a nucleoid does not change substantially in condition with low condensin concentration (a,b), it becomes narrower at high concentration (c), showing the DNA mass moves closer to its center. d) In presence of surface mediated sticking during compaction, the distribution narrows as well, but DNA is prevented from being pulled to the nucleoid's center, and as a result peaks appear away from COM. Profiles are taken at the beginning (black) and end (red) of compaction.

5.3. CONCLUSION AND OUTLOOK

In this study, we performed microfluidic experiments investigating the role of SMC protein condensin on megabase DNA scaffold at a single molecule level. We observed concentration-dependent DNA compaction in the presence of ATP. The period over

which the compaction proceeded varied significantly from about 15 to about 2 minutes at 200 and 400 nM respectively. We did not observe any notable compaction at concentrations below about 50 nM. Compaction was also not observed in the absence of ATP, although we noted some DNA binding, which became stronger in the presence of ATP. At the highest protein concentrations, the protein-DNA complex formed an insoluble aggregate at the surface. While our results point to the loop-extrusion activity as mediator of the compaction process, more experiments are needed to characterize the effect of ATP-independent behaviour in our assay.

While the highest measured reduction in radius of gyration was up to 40%, which appears modest, this corresponds to a nearly 80% 3D volume reduction (i.e. if radius reduces by 40%, e.g. going from 1 to 0.6 in relative terms, the corresponding 3D volume reduces by 80%, $V_2/V_1=0.6^3/1^3=0.22$), which is appreciable. Here, we used 3D volume, as highly compacted nucleoids have dimensions smaller than the trap height. Interestingly, recent reports have suggested that eukaryotic SMCs, including condensin, extrude loops asymmetrically, i.e. in one-sided fashion [19]. In this case, the theory predicts only moderate compaction, with radius of gyration for the chromosome backbone decreasing at most by a factor of 3 [20]. Notably, this factor is the upper bound for compaction, for as the loop extrusion proceeds, large portion of DNA is reeled into the loops, and out of the backbone - which has a compensatory effect that increases the effective size. Therefore, the compaction observed here may already be close to this bound. Future experiments with larger sample sizes, and a combination with polymer dynamic simulations, should aim to establish whether this bound holds, and characterize the final state including average loop size.

During our experiments, we experienced challenges with surface interaction of the protein-DNA complex, that became apparent at the highest protein concentration. In these conditions, the radius of gyration, a metric that describes the spatial arrangement of DNA mass around objects centre, became a poor descriptor of the compaction effects that we observed visually. We developed three different metrics for these cases, the perimeter excess ratio, the share of bright pixels, and the radial intensity distribution with respect to centre of mass, each with distinct advantages. The perimeter excess allows to describe change in how corrugated the shape of a large DNA molecule is. The share of bright pixels is a simple and robust metric that most faithfully captures what is seen by the eye, namely some *local* compacted objects becoming more defined and bright. Finally, the radial intensity distribution with respect to centre of mass can be useful in describing any (deviations of a) radially symmetric structure of the isolated chromosomes, such as toroid that has been hypothesized to emerge under the effect of loop extrusion on a topologically closed scaffold. We observed that deviations from a simple symmetric distribution signaled local accumulation of DNA.

Taken together, the results we presented here are an interesting proof of concept and suggest a range of follow-up experiments. First, studies at intermediate condensin concentrations (50 – 200 nM) should be carried out. This will allow to expand sample sizes for conditions initially studied here, and carrying out statistical comparisons among them. It will also shed more insight on degree of compaction and DNA binding that hap-

pens at the lower end of this concentration range, and indicate to what extent these experiments may suffer less from the protein-induced DNA sticking to surface we observed at high concentrations (400 nM). Second, experiments in the absence of ATP should be carried out in the range of intermediate to high condensin concentrations (50 – 400 nM). These experiments will help to elucidate what are the individual contributions of ATP-independent intermolecular interactions between condensin proteins, as well as their ATP-dependent effect. Third and most important, the surface treatment protocol has to be optimized to rule out any significant interaction of the DNA with the surface for the duration of the experiment. In fact, the presence of surface sticking was the single most important confounding factor in this study, and the reason that prevented us from the original plan of comparing experimental results with polymer model simulations. Indeed, carrying out polymer model simulations will be a fruitful complement to results obtained from this microfluidic assay. Notably, they should allow to make progress toward the questions originally posed at the design of this study, such the structure of the megabase-scale DNA after loop extrusion, the size of the extruded loops and its variability, and the proteins' processivity and residence time. While such polymer simulations are greatly valuable, experimental validation is obviously important – for which the current approach provides an interesting avenue.

In parallel to the type of studies presented here, future experiments on DNA-protein interactions employing a megabase scaffold can also employ modalities that do not require microfluidic trapping. Examples include variants of the conformation capture (such as 3-C and Hi-C), and functional studies of transcriptional readout (such as rt-PCR and RNASeq). The advantage of these complementary approaches, in comparison to fluorescence imaging employed here, is that they are not limited by optical resolution. Finally, it is important that the continued development of the microfluidic approach, makes the tradeoffs between device complexity, ease of operation, and throughput explicit. Device designs that emphasize the latter are more likely to be successful, not only at the stage of data collection but also during the assay's development. Furthermore, such approaches have a higher likelihood to be adopted by the broader scientific community.

5.4. MATERIALS AND METHODS

5.4.1. MICROFLUIDIC DEVICE FABRICATION AND OPERATION

The microfluidic device fabrication and operation is described in Chapter 3 of this thesis.

5.4.2. BACTERIAL CELL CULTURE

The conditions for bacterial cell growth, synchronization and spheroplasts preparation are described in Chapter 3 of this thesis. The characterization of DNA and the remaining proteins associated with it was done with mass spectrometry, and is likewise reported in Chapter 3.

5.4.3. EXPRESSION, PURIFICATION, AND LABELING OF YEAST CONDENSIN

Pentameric *S. cerevisiae* condensin complexes were purified as reported previously [21]. Briefly, *S. cerevisiae* cells were transformed with a pair of 2µ-based high copy plasmids containing *pGAL10-YCS4 pGAL1-YCG1 TRP1* and either *pGAL7-SMC4-StrepII$_3$ pGAL10-SMC2 pGAL1-BRN1-His$_{12}$-HA$_3$ URA3* (wild-type, strain C4491), *pGAL7-smc4(Q302L)-StrepII$_3$ pGAL10-smc2(Q147L) pGAL1-BRN1-His$_{12}$-HA$_3$ URA3* (Q-loop ATPase mutant, strain C4724), *or pGAL7-SMC4-StrepII$_3$ pGAL10-SMC2 pGAL1-brn1(M391D, F394D, W402D, W408D)-His$_{12}$-HA$_3$ URA3* (safety belt mutant, strain C5037). Overexpression was induced by addition of galactose to 2% in –Trp-Ura media. Cell lysates were prepared in buffer A (50 mM TRIS-HCl pH 7.5, 200 mM NaCl, 5% (v/v) glycerol, 5 mM β-mercaptoethanol, 20 mM imidazole) supplemented with 1× cOmplete EDTA-free protease inhibitor mix (11873580001, Roche) in a FreezerMill (Spex), cleared by centrifugation, loaded onto a 5-mL HisTrap column (GE Healthcare) and eluted with 220 mM imidazole in buffer A. Eluate fractions were supplemented with 1 mM EDTA, 0.2 mM PMSF and 0.01% Tween-20, incubated overnight with Strep-Tactin Superflow high capacity resin (2-1208-010, IBA), and eluted with buffer B (50 mM TRIS-HCl pH 7.5, 200 mM NaCl, 5% (v/v) glycerol, 1 mM DTT) containing 10 mM desthiobiotin. After concentrating the eluate by ultrafiltration, final purification proceeded by size-exclusion chromatography with a Superose 6 column (GE Healthcare) pre-equilibrated in buffer B containing 1 mM $MgCl_2$. For preparation of fluorescently labelled condensin, 0.5 ml of concentrated protein was incubated with 0.1 mM Janelia Fluor®646 (#6148, Tocris, UK; JF-646) for 30 minutes at room temperature, prior to size exclusion chromatography. Purified protein was snap-frozen and stored at -80 °C until use.

5.4.4. IMAGE ACQUISITION AND ANALYSIS

The protocols for image acquisition and analysis, including link to the Python code, are described in Chapter 4. For the imaging of condensin, we used 640 wavelength 100 mW laser (typically at 10-20% power output) for excitation, and 617/73 filter on emission.

5.4.5. SAMPLE SELECTION AND DATA ANALYSIS

Radius of gyration was selected as a main readout, with perimeter excess ratio, the share of bright pixels, and the radial intensity distribution as secondary readouts. The radius of gyration was calculated by squaring the sum of all foreground pixels' intensity-weighted distances from the object's center of mass, after subtracting the background. The resulting measurements were saved as structured JSON files, one per each field of view (i.e. trap). Sample data were grouped based on experimental condition (40, 200 or 400 nM condensin, with or without ATP). To be able to calculate trends from traces with heterogenous sampling frequences, the measurement data (e.g. R_g, perimeter excess ratio, ...) was interpolated on regular time-sampling interval. The time of start of imaging was simply the time of the start of experiment, without any alignment to the time of (perceived) compaction onset. For reporting a relative radius of gyration, its value was scaled by the average of the three initial measurements in its time trace ($R_{g,start}$).

5.4.6. PERIMETER EXCESS RATIO, SHARE OF HIGH INTENSITY PIXELS, AND RADIAL DISTRIBUTION OF INTENSITY

Perimeter excess ratio, share of high intensity pixels, and the radial intensity distribution centered at the center of mass (COM) were the secondary readouts. All values were calculated from the raw data after applying a binary mask, with the exception of the radial intensity profiles, where background was additionally subtracted.

The perimeter excess ratio is the ratio of the length of a perimeter of a maximum projected binarized nucleoid image, to the length of a perimeter of a hypothetical disk having the same surface area, and the same radius of gyration (Fig. 5.11). Notably a disk with radius of gyration R_g will have an area $A_{disk} = 2\pi R_g^2$. The equivalence of areas is satisfied by sweeping across a range of thresholds on the nucleoid image and selecting the one that gives the lowest least-squares error from the disk's area. The perimeter is obtained from the binarized image by tracing the length of a line that results from taking a difference between the image and the eroded version of the same image (which is effectively about 1 pixel smaller at all points along the edge). The perimeter length is then calculated as the sum of their nearest neighbor distances of these pixels. The same procedure can be repeated for a disk, which allows evaluation of a ratio. This metric is useful to quantitatively compare conditions where one observes changes to chromosomes shapes. The more a shape deviates from a disk, the longer its perimeter will be, and hence the perimeter excess ratio increases.

The share of high intensity pixels was calculated at each time point from a maximum projected image by counting the number of pixels that are above a threshold that was fixed at a constant value across the whole timelapse (Fig. 5.12). The threshold was originally found per each frame (viz. Image Acquisition and Analysis), and here was averaged over all frames, and additionally increased by 2 standard deviations of all non-zero values in the masked image, so as to retain only the brightest pixels.

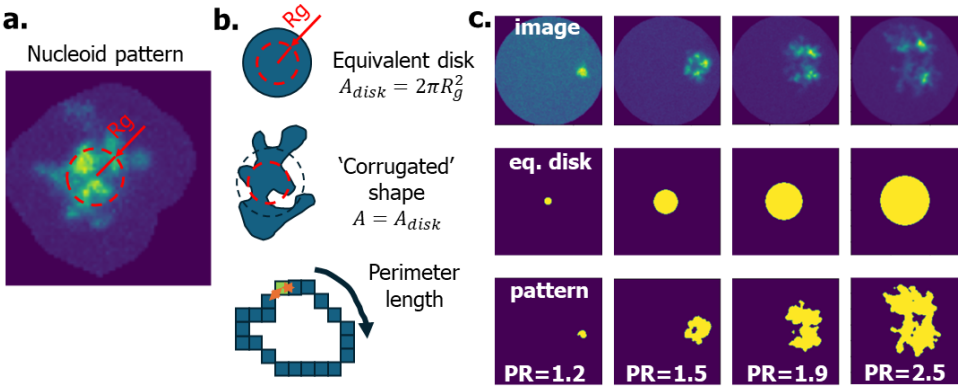

Figure 5.11: Calculation of perimeter excess ratio. a) The input is a nucleoid fluorescence image and its radius of gyration, Rg, that was calculated in data analysis. b) An equivalent disk (eq. disk) is a solid disk with the same radius of gyration as a nucleoid. The area of that disk is A_{disk}, is $\pi R_{disk}^2 = 2\pi R_g^2$.The nucleoid image is thresholded such as to yield the same binary area. Next, the length of the perimeter of this area can be calculated and compared to the one of the disk, yielding perimeter excess ratio. c) Example of nucleoid where the perimeter excess ration (PR) is seen to change over time, reliably following the change of nucleoid's shape.

Finally, the radial intensity distribution around object's center of mass (COM) was simply the radially averaged profile obtained on masked and background subtracted data. One curve was generated per each timepoint and each object. As the center of mass is positioned at an arbitrary position inside the trap, this results to a sharp cutoff of the profile at walls, which happens at different distance from COM for different samples.

5.4.7. BUFFER COMPOSITION AND EXPERIMENTAL CONDITIONS

Bacterial spheroplasts were prepared as described in the section 'Bacterial Cell Culture', introduced on the microfluidic device and guided into the traps by fluid flow, after which they were lysed with low osmolarity buffer (50 mM Tris HCl pH8, 500 nM Sytox Orange). Next, condensin labeled with JF-646 was introduced onto the device at concentration of either 40, 200 or 400 nM in buffer similar to the one used in in vitro experiments previously [22] (2 mM Trolox (an antioxidant preventing formation of reactive oxygen species), 40 mM Tris HCl pH7.5, 50 mM potassium glutamate, 0.2 mg/mL BSA, 2.5 mM $MgCl_2$, 5% glucose, 500 nM Sytox Orange and 10 mM ATP). Notable differences from previous studies include the high concentration of Sytox Orange (which was required because a large portion of the dye was sequestered by device walls) and omission of enzymatic scavenging system such as glucose oxidase (which was deemed superfluous due to lack of any visible DNA damage in its absence). Experiments without ATP were done in the same buffer, where ATP was left out. Condensin aliquots were retrieved from -80 °C freezer and thawed on ice right before use.

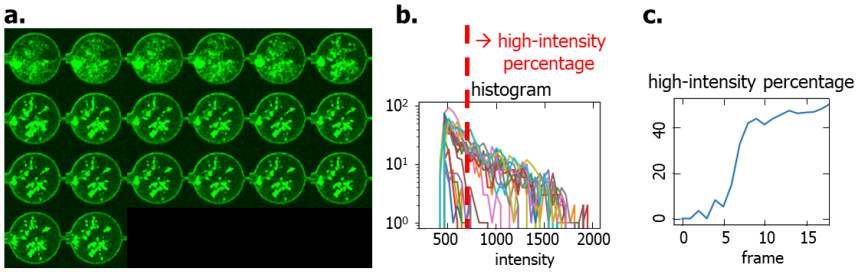

Figure 5.12: Calculation of share of bright pixels. a) The input is a nucleoid fluorescence image. b) Histogram is obtained for each frame (here color-coded), and the share or bright pixels is determined using a global threshold. c) Plotting the share of bright pixels over time highlights if the nucleoid is getting more compact, which manifests as higher DNA density and therefore larger share of bright pixels.

5.5. REFERENCES

[1] Goloborodko, A., Imakaev, M. V., Marko, J. F. & Mirny, L. Compaction and segregation of sister chromatids via active loop extrusion. *eLife* **5**, e14864 (2016).

[2] Kim, E., Barth, R. & Dekker, C. Looping the Genome with SMC Complexes. *Annu. Rev. Biochem.* **92**, 15–41 (2023).

[3] Kiefer, L. *et al.* Tuning cohesin trajectories enables differential readout of the Pcdhα cluster across neurons. *Science* **385**, eadm9802 (2024).

[4] Zhang, X. *et al.* Fundamental roles of chromatin loop extrusion in antibody class switching. *Nature* (2019).

[5] Pradhan, B. *et al.* Loop extrusion-mediated plasmid DNA cleavage by the bacterial SMC Wadjet complex. Preprint at (2024).

[6] Tišma, M. *et al.* Dynamic ParB–DNA interactions initiate and maintain a partition condensate for bacterial chromosome segregation. *Nucleic Acids Research* **51**, 11856–11875 (2023).

[7] Aljahani, A. *et al.* Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. *Nat Commun* **13**, 2139 (2022).

[8] Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).

[9] Ganji, M. *et al.* Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102–105 (2018).

[10] Pradhan, B. *et al.* SMC complexes can traverse physical roadblocks bigger than their ring size. *Cell Reports* **41**, 111491 (2022).

[11] Pan, H. *et al.* Cohesin SA1 and SA2 are RNA binding proteins that localize to RNA containing regions on DNA. *Nucleic Acids Research* **48**, 5639–5655 (2020).

[12] Davidson, I. F. *et al.* CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature* **616**, 822–827 (2023).

[13] Sabaté, T. *et al.* Universal dynamics of cohesin-mediated loop extrusion. Preprint at (2024).

[14] Davidson, I. F. & Peters, J.-M. Genome folding through loop extrusion by SMC complexes. *Nat Rev Mol Cell Biol* **22**, 445–464 (2021).

[15] Vos, E. S. M. *et al.* Interplay between CTCF boundaries and a super enhancer controls cohesin extrusion trajectories and gene expression. *Molecular Cell* S109727652100455X (2021).

[16] Ryu, J.-K. *et al.* Bridging-induced phase separation induced by cohesin SMC protein complexes. *Science Advances* **7**, eabe5905.

[17] Polsby–Popper test. *Wikipedia* (2024).

[18] Roundness. *Wikipedia* (2024).

[19] Barth, R. *et al.* SMC motor proteins extrude DNA asymmetrically and contain a direction switch. *bioRxiv* 2023.12.21.572892 (2023).

[20] Banigan, E. J. & Mirny, L. A. Limits of Chromosome Compaction by Loop-Extruding Motors. *Phys. Rev. X* **9**, 031007 (2019).

[21] Terakawa, T. *et al.* The condensin complex is a mechanochemical motor that translocates along DNA. (2017).

[22] Analikwu, B. T. *et al.* Telomere protein arrays stall DNA loop extrusion by condensin. Preprint at (2023).

6

Developing an Immunoprecipitation Strategy for Isolating Yeast Synthetic Chromosomes

Synthetic genomics, an interdisciplinary field involving the design and construction of new genomes and the modification of existing ones, has heavily relied on *Saccharomyces cerevisiae* as a host. This yeast has enabled significant advances in the field, particularly in the assembly of synthetic chromosomes, some exceeding 10 million base pairs (Mbp). However, transferring these chromosomes to other organisms remains challenging. Spheroplast-fusion and agarose-plug methods have been developed to this end, but are limited to about 1 Mbp. Furthermore, no existing methods have been shown to effectively isolate chromosomes for cell-free systems, with current approaches limited to small plasmids and hampered by background DNA contamination.

In response to these challenges, we developed an immunoprecipitation-based method for isolating synthetic chromosomes. This method aims to preserve chromosome integrity and minimize DNA background, potentially providing an inexpensive and facile solution for the isolation of synthetic chromosomes for in vitro applications. We demonstrate this approach on a 190 kbp synthetic chromosome carrying a tandem repeat array of tetO binding sites that recruit a Tet repressor protein (TetR), and asses the pulldown quality using pulsed-field gel electrophoresis (PFGE). Our results demonstrate the ability to significantly enrich for chromosomes of this size, with an enrichment factor ranging between 6- to 15-fold. Future work should focus on testing isolated chromosomes in vitro, and on assessing effectiveness of pulling down larger chromosomes with varying arrangements of protein recruitment sites.

This project was pursued in the laboratory of Patrick Yizhi Cai at the Manchester Institute of Biotechnology between June and August of 2023. It was made possible by the EMBO Scientific Exchange Grant.

6.1. INTRODUCTION

As this project was carried out as a branching out of the main part of my work, I will first provide an extended introduction with the background for this project, before continuing with description of my actual research efforts later in this chapter.

Biological research has a common denominator – studying living systems that are complex. In face of this complexity, scientists generally settle on an intuitive approach - to make alterations to the systems at hand, and observe their consequences. This approach dominates virtually all of biological research. It is thanks to it that we discovered insulin, understood the role of antibodies in immune response, and determined the genotypic profile of number of hereditary diseases, among others. Such alterations can be made at levels across all of biological hierarchy, from ecology to molecular biology. All are useful, but it is only at the lowest level of organization that we can start to decipher mechanistic relationships.

Making these changes is a critical step in much of biological research. So how does one go about it? The most common way is to manipulate the DNA sequence. DNA is the programming language of life, and it is responsible for majority of what happens in cells. Knocking out a gene, for example, can help to reveal the role of its corresponding protein in health and disease. Changing individual bases in DNA can, in turn, point to importance of specific amino acids in protein's function. Introducing a new gene can help to study its role in an insulated context, or substitute for a malfunctioning protein in a disease model.

Despite the various approaches to study the relationship between genetic sequence and cell's phenotype that have been demonstrated to date, there is room for improvement. Most importantly, it is currently difficult to introduce a large number of modifications at once. The exact number varies based on the type of modification, delivery method, and cell type, but practically ranges from a few to about twenty [1]. Secondly, it is not possible to deliver sequences for genomic integration without some degree of off-target effects [2]. Similarly, especially in context of engineering organisms for novel functions, it is challenging to insulate the newly introduced pathways from the existing genomic background [3]. Clearly, new tools are needed to tackle questions about the relative importance of genomic context for function, DNA sequence redundancy, and genome architecture. The field of synthetic genomics emerged in response to these questions.

6.1.1. BRIEF HISTORY OF SYNTHETIC GENOMICS

Synthetic genomics is an interdisciplinary field that involves the design and construction of new genomes, as well as large-scale modifications of existing ones. The origins of synthetic genomics date back to 1970, when a group of scientists in the laboratory of Har Gobind Khorana reported the first total synthesis of a gene [4]. Building on earlier work of Khorana, who shared the 1968 Nobel Prize in Physiology and Medicine for the "interpretation of the genetic code and its function in protein synthesis", the group syn-

thesized the gene for the yeast alanine transfer RNA ($tRNA^{Ala}$). Another major milestone was achieved in 1995, when a group led by Herbert L. Heyneker reported total synthesis of a 2.7 kbp plasmid from 134 oligos in a single reaction [5]. The work built on the DNA shuffling technology, which was developed one year earlier by Willem P.C. Stemmer [6].

In passing we can add that Heyneker and Stemmer are, respectively were, both Dutch, and that they entered history as major scientific contributors to the modern biopharma industry. Stemmer and Frances H. Arnold shared the 2011 Charles Stark Draper Prize from the National Academy of Engineering for contributions to directed evolution (other technologies awarded this prize include the invention of CCD in 2006, the development of World Wide Web in 2007, and the creation of the C++ programming language in 2018). His patents have been recognized as among the most influential in the biotech industry. Heyneker, in turn, was the first employee of Genentech. Founded in 1976, Genentech is historically regarded as the world's first biotechnology company. Heyenker's scientific contributions were key to the company's early success that included the cloning and production of somatostatin [7], human insulin [8], and human growth hormone [9] in *E. coli*. These achievements impacted the lives of hundreds of millions if not billions of people, and catapulted Genentech to become one of the most valuable pharma/biotech companies worldwide.

The year 2002 marked the first chemical synthesis of a full genome of an organism, when researchers at the Stony Brook University in New York assembled the 7.5 kbp poliovirus DNA [10]. Controversial already at the time due to risk of virus' escape from laboratory and dual-use potential, their research nevertheless represented an important milestone in genome assembly. From there on, the field of synthetic genomics advanced rapidly. In part, this was driven by the Human Genome project in the US, which was completed in 2006, and brought large improvements in DNA synthesis, assembly, and screening. However, the major enabler was the use of *S. cerevisiae* for the assembly of large DNA constructs. The use of yeast allowed a group of scientists at the J. Craig Venture Institute (JCVI) to demonstrate the synthesis and combined in vitro and in vivo assembly of the 583 kbp *Mycoplasma* genome in 2008 [11]. This success was followed up by the redesign and engineering of the 4 Mbp *E. coli* genome in the labs of George Church and Jason Chin in 2016 and 2019, respectively [12,13].

Encouraged by the results in assembly of genomes in prokaryotes, researchers endeavored on genome engineering and assembly projects in eukaryotes. Given the size and complexity of eukaryotic genomes this was facilitated by large international collaborations, first for a synthetic yeast genome with the Sc2.0 project [14], and more recently for genomes of higher eukaryotes, including human and plant cells, with the Genome Project Write [15]. The goal of Sc2.0 is to build a complete synthetic genome of a model eukaryote, creating a platform for systematic studies. The design rationale behind the Sc2.0 genome balances maintenance of the wild-type phenotype with maximizing genetic flexibility and reducing genome instability (cf. also Fig. 6.2). The main design features of the Sc2.0 synthetic chromosomes are i) the replacement of all TAG stop codons with TAA; ii) the inclusion of Cre-recombinase cutting sites for inducible evolution; iii) the removal of repeat elements and number of introns; iv) the relocation of all tRNA

genes on a new separate chromosome; v) the addition of PCRTag sequences for rapid genotyping; and vi) the addition or removal of enzyme recognition sites that facilitate the multistep sequence assembly.

With the total of 17 total chromosomes built by joint effort of 11 different groups, the Sc2.0 project is, as of the moment of writing of this thesis, nearly complete [16]. The project laid a technological foundation for large-scale synthetic genomics efforts, including the development of computational design tools [14,17] and troubleshooting protocols [18,19]. While still in the very early stages of its application, the synthetic yeast built in Sc2.0 project already contributed to advancing understanding of genome function. Some examples include a study leveraging the strain's genetic flexibility to investigate the role of genomic context on transcription [20], a study relating chromosome topology to contact and recombination frequencies [21], or a study that identified new essential genes and allowed for nearly 40% chromosome size reduction [22]. Clearly, synthetic genomics is well posed to contribute to answering questions about fundamental principles of genome organization and function, as well as to find uses in range of biotechnological applications.

6.1.2. SYNTHETIC GENOMICS WORKFLOW

A typical synthetic genomics workflow consists of several standard steps (Fig. 6.1). First, the sequence of to-be-built chromosome is designed on a computer. Here, a number of criteria is taken into account, including the desired functionality, compatibility with the host, feasibility of the synthesis, and the assembly strategy. Second, the sequence design is split into short fragments which are ordered from a DNA synthesis vendor (price generally ranges between 0.1 and 0.3 EUR per base). Alternatively, these fragments can be synthesized in a lab with benchtop synthesizer, an approach that is becoming increasingly affordable and popular, in laboratories where the demand justifies larger start-up costs. Third, these oligos are assembled in vitro with standard DNA ligation techniques relying on sequence overhang homology (e.g. Golden Gate, HiFi, Gibson, ...).

The molecules hitherto obtained are generally in the size range between 2 and 10 kilobasepairs, though larger constructs can be made with some difficulty. While linear, these constructs are almost exclusively integrated to a circular vector for amplification and storage in *E. coli.* Fourth, these sequences can be retrieved by PCR amplification or enzymatic digestion from isolated vectors. Fifth, they are delivered to a host, most commonly *S. cerevisiae,* which uses its recombination machinery to assemble them into a single DNA molecule. Molecules ranging from tens of kilobasepairs [3] up to tens of megabasepairs [22,24] have been built using this technology. Except for very simple assemblies, this step will generally involve extensive screening and troubleshooting. After this process, and if the assembly host is also the target one, the chromosome is ready for use.

It has become of increasing interest to deploy synthetic chromosomes in different organism [25,26], or even in vitro [27]. In these cases the chromosome needs to be transferred

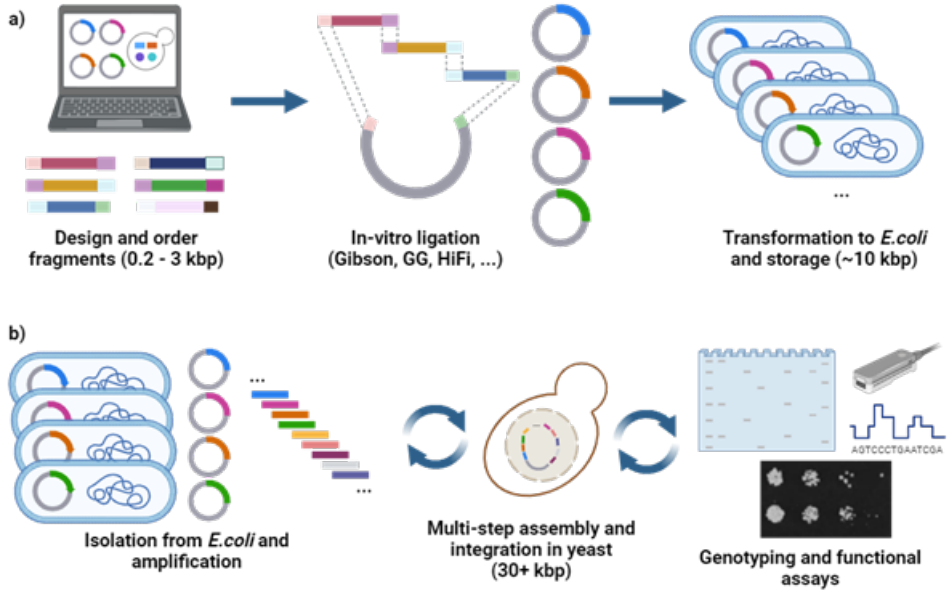

Figure 6.1: A typical synthetic genomic workflow. a) The assembly is design with the help of biological design tools (BDT). Individual fragments are ordered from synthesis vendor and assembled in-vitro. The assembled plasmids are transformed to *E. coli* for long-term storage and amplification. b) Plasmids are retrieved from *E. coli*, and the fragments can be retrieved by PCR. Fragments are transformed to *S. cerevisiae* together with suitable backbone, or integrated to natural and synthetic genomes. Depending on the size and complexity of assembly, this may require multiple steps. Finally, the assembly is verified by genotyping and with functional assays. GG – Golden Gate. Spot assay example reprinted from [23].

to the target organism, or isolated, both of which are discussed later. Before doing so, we now turn to discuss two critical steps of a synthetic genomic workflow, their design and host-based assembly, in more detail.

6.1.3. DESIGN OF SYNTHETIC GENOMES

Synthetic chromosomes are designed to closely emulating natural chromosomal behavior, while incorporating designer features and not carrying substantial metabolic burden (Fig. 6.2). Synthetic chromosomes have to contain host- and target-species specific centromeres (CEN) [26], to assure faithful segregation of chromosomes during cell division. Linear synthetic chromosomes must carry telomeres to prevent undesirable end-to-end chromosome fusions and maintain genomic stability. These can be substantially simplified, as demonstrated by the Sc2.0 design that replaces telomeric and subtelomeric regions with universal telomere caps (UTCs) ~300 bp [14,28]. Nevertheless, circular chromosomes are also common. In order to assure DNA replication, a synthetic chromosome must include a sufficient number of autonomously replicating sequences (ARS) distributed along its length. The overall topology of the chromosome, including the spacing and orientation of these replication origins, is an element of the design, so as to assure reliable replication.

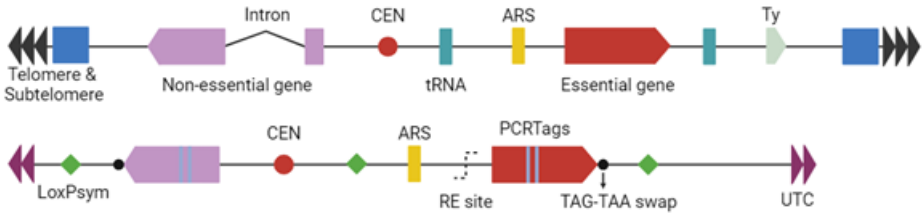

Figure 6.2: A layout comparison example between native (top) and synthetic (bottom) chromosome. CEN – centromere, ARS – autonomously replicating sequence , tRNA – tRNA gene, Ty – yeast transposable element, LoxPsym – LoxPsym site, RE site – restriction enzyme site, UTC – universal telomere cap. Illustration inspired by synIV [30].

A unique feature of yeast synthetic chromosomes is the incorporation of recombinase recognition sites, such as *loxP* and *loxPsym*. Inducing expression of the Cre recombinase in a host with such a synthetic chromosome will result in a wide variety of genomic rearrangements, which is the principle of SCRaMbLE (Synthetic Chromosome Rearrangement and Modification by *LoxP*-mediated Evolution) system [29]. This system allows for inducible, targeted recombination events, including deletions, duplications, inversions, as well as translocations. SCRaMbLE is a powerful tool for studying genome organization and evolution as well as for engineering yeast strains with novel properties. In the context of engineering of industrially relevant strains, this directly allows to probe what chromosome rearrangements lead to optimized production [18]. Notably, Cre is only one of many recombinases. Using other enzymes in parallel, and incorporating corresponding recombination sites to subset of synthetic chromosomes, would allow for chromosome-specific evolution, even in a synthetic strain background.

Additional design elements often include selectable markers and reporter genes, which assist in identifying and isolating cells that contain the synthetic chromosome. Insulator sequences may also be incorporated to prevent unintended interactions between adjacent genetic elements, ensuring that each functional unit operates independently.

6.1.4. HOST-BASED ASSEMBLY OF SYNTHETIC GENOMES

After the initial in-vitro assembly of smaller DNA fragments, these fragments are ready to be introduced into the yeast cells for further integration. *S. cerevisiae* has become the host of choice for this task, owing to its genetic tractability, well-established laboratory protocols, and its ability to recombine DNA efficiently through homologous recombination (HR). Single step assemblies of as many 40 DNA fragments [3] have been reported in this host. Final sizes of chromosomes built in this way are frequently in the megabase-pair range [24,26,31,32].

Despite the advantages of using *S. cerevisiae* for synthetic chromosome assembly, there are common challenges and limitations. Assemblies of individual chromosomes usually must proceed in multiple steps and must be accompanied by careful design, assuring

autotrophy selection of successful clones at each step. Consequently, the whole process is often technologically demanding and time intensive. Another limitation is ensuring the stability and proper function of synthetic chromosomes within the host cell, as large synthetic constructs may impose a metabolic burden or interfere with the host's cellular machinery. Finally, while *S. cerevisiae* is a powerful model organism, findings in yeast may not always be directly translatable to other eukaryotic systems.

Taken together, *S. cerevisiae* has proven as a remarkable host for synthetic genomics. It is not an overstatement to say that the field would likely not have emerged if it was not for the existence of this host. Nevertheless, researchers are increasingly interested studying synthetic chromosomes that were built in yeast in different contexts, i.e. in different organisms or in cell-free systems. In order to be able to do that, the synthetic chromosome must be first transferred to the target organism, or isolated.

6.1.5. STRATEGIES FOR TRANSFERRING SYNTHETIC GENOMES

S. cerevisiae has been the host-of-choice for assembly of synthetic genomes for the reasons described above. It is, however, only one of its many targets. Different microbial strains are common candidates in context of synthetic biology, whereas pharmaceutical production would benefit from transfer to a mammalian cell target [37]. Recently, plants emerged a promising target as well [17]. Finally, synthetic genomes for cell and gene therapies will be targeting human cells [26]. Researchers interested in imbuing these organisms with synthetic chromosome functionality must shuttle them from the assembly host to the desired target. This represents an additional step in the protocol, and a technical challenge.

Advances in DNA synthesis, screening, and engineering have enabled assembly of synthetic chromosomes that are regularly in the megabasepair range, and have even exceeded 10 million basepairs (Mbp) [24]. Chromosome transfer, on the other hand, continues to represent a larger challenge (Table 6.1). Despite early advances in transferring sequences as large as 2.3 Mbp, these approaches were limited to native DNA, suffered from poor DNA retention, and had low efficiency [33]. Whole wild-type genomes as large as 1.8 Mbp have been transferred from bacteria to yeast [36], and a number of genomes has been reconstituted in yeast with the transformation-associated recombination (TAR) cloning approach (e.g. [38,39]).

These approaches efficiently leverage yeast's large genetic plasticity and high HR efficiency. However, going in the opposite direction, i.e. transferring chromosomes built in yeast to other organisms, remains challenging, with the largest shuttled chromosome to date 1.08 Mbp [34,41]. Finally, isolating chromosomes for use in vitro has not been reported, with studies reserved to PFGE characterization of exogenous genomes assembled in yeast [11,40].

Genome Source	Genome Size	Isolation Method	Target Cell	Transfer Method	Efficiency	Ref.
A) Isolation-based genome transfer						
Fragment from human chrY on YAC (L438) †	2.3 Mbp	Agarose plug	*HT1080 (human)*	PEI assisted lipofection‡	Low	[33]
Mycoplasma mycoides	1.08 Mbp	Agarose plug (kit)	*Mycoplasma capricolum*	PEG-mediated transformation	Low	[25]
Mycoplasma mycoides †	1.08 Mbp	Agarose plug (kit)	*Mycoplasma capricolum*	PEG-mediated transformation	Medium	[25,34]
B) Spheroplast fusion based genome transfer						
Haemophilus influenzae	1.83 Mbp	NA	*S. cerevisiae*	Spheroplast fusion	Medium	[35,36]
Mycoplasma mycoides †	1.08 Mbp	NA	*HEK293 (human)*	Spheroplast fusion	Medium	[37]
YAC-Mm-4q21 LacO †,1	0.75 Mbp	NA	*HT1080, U2OS (human)*	Spheroplast fusion	High	[26]
C) Transformation-associated recombination cloning (examples)						
Spiroplasma chrysopicola	1.12 Mbp	Agarose plug	*S. cerevisiae*	PEG-mediated transformation	High	[38]
Prochlorococcus marinus	1.66 Mbp	Agarose plug	*S. cerevisiae*	PEG-mediated transformation	Low / Medium	[39]
D) Exogenous chromosome assembly in yeast (examples)						
Mycoplasma genitalium †	0.58 Mbp	Agarose plug (kit)	NA	NA	NA	[11,40]
Mycoplasma pneumoniae †	0.82 Mbp	Agarose plug (kit)	NA	NA	NA	[40]

Table 6.1: Examples of genome assembly and transfer. †Genomes cloned in *S. cerevisiae*, ‡PEI – polyethylenimine induced DNA compaction, [1]YAC cmbining 550kb of Mycoplasma mycoides genome and 4q21^{LacO} BAC). YAC – yeast artificial chromosome, BAC – bacterial artificial chromosome, TAR - transformation-associated recombination.

Type of DNA damage	Frequency	Repair mechanisms
Oxidative	1×10^{4} /cell/day	BER, NER
Single-strand break (nick)	$1 \times 10^{4} - 1 \times 10^{5}$ /cell/day	BER
Double-strand break	$1 \times 10^{1} - 5 \times 10^{1}$ /cell/day	NHEJ, HR
Mutation	$1 \times 10^{-9} - 1 \times 10^{-10}$ /bp/division	MMR

Table 6.2: Common sources of DNA damage in eukaryotic cells. BER – base excision repair, NER – nucleotide excision repair, NHEJ – non-homologous end joining, HR – homologous recombination, MMR – mismatch repair.

One major challenge in transferring large DNA molecules is their fragility, as large DNA molecules tend to break upon handling. DNA is continuously exposed to causes of damage. While DNA is relatively stable in the cellular milieu, it owes its stability to number of active repair processes (Table 6.2). In the absence of these processes, DNA will accumulate damage, both chemical, and mechanical. While this is relatively less important for short sequences, where it can be additionally compensated by amplification or abundance, it does represent a major bottleneck for manipulations of large low-copy number molecules. Delivering genetic payload only partially, or with breaks that lead to partial translation products, can have unpredictable consequences, and clearly reduces the utility the synthetic genomic approach. It is therefore vital to minimize this damage during genome extraction and transfer.

Given the relatively modest mutation rate of DNA (Table 6.2) the importance of mechanical damage overshadows that of the chemical one, even for sequences in the megabase-pair range. The central requirement for transfer of synthetic chromosomes is therefore the need to limit the amount of mechanical disruption the DNA experiences. Two main strategies have been developed to this end, spheroplast fusion and agarose-plug based transfer. We discuss these in detail next.

SPHEROPLAST FUSION

Spheroplast fusion is a technique that eliminates any direct DNA manipulation. Notably, it does not include any DNA pipetting step. Briefly, it relies on fusion of a donor and a target cell. As preliminary step, the cell wall of the yeast donor must be digested. If the target cell has a wall, that wall too needs to be digested. Cells with digested cell wall are commonly referred to as spheroplasts. Here, "sphero" refers to their spherical shape which their membrane assumes after the shape-defining wall has been digested. Spheroplasts are commonly created by incubating cells with a cell-wall degrading enzyme. Cell-wall composition differs between organisms and so does the corresponding used enzyme. Bacterial walls are commonly degraded with lysozyme. A mixture of enzymes including laminaripentao-hydrolase and glucanase is used to degrade the cell wall of yeast. If the target cell is of mammalian origin, there is no cell wall to degrade. Mammalian cell lines do, however, have a nuclear envelope that separates their genomic DNA from the cytoplasm. The nuclear envelope limits the transfer of genomic material

between the nucleus and the cytoplasm. Arresting the cells in the mitotic phase of the cell cycle, when the envelope is broken down, has been shown to yield improvements in chromosome transfer efficiency [37].

For any two membranes to fuse, they first need to come sufficiently close to each other in the bulk of the solution. This is generally unfavorable, as membranes tend to be hydrophilic. While efficiency could be greatly increased by traditional transfection approaches such as electroporation or mechanical disruption, these are not compatible with delivery of DNA beyond few tens of kilobasepairs. The energy barrier for membrane fusion can be reduced by adjusting buffer composition to depolarize the membranes. If membranes are sufficiently depolarized and hydrophilic, at short distances they will be attracted and held together by van der Waals interactions. This initially happens locally, at a small membrane patch, as it is usually just a small region, not the entirety of any of the two membranes, that satisfies the hydrophobicity and charge conditions.

The mechanism by which fusion proceeds from the contact of these two patches is not entirely known. Both the inner and outer membrane leaflets must merge for a successful fusion event. While outer leaflets fuse easily [42], such a hemifusion will not yet result in mixing of the cells' contents. For the mixing to occur, the fusion of inner leaflets is also required. It is thought that the fusion of the inner leaflets requires a temporary and matching defect and that transmembrane proteins can facilitate these defects [43].

The frequency of fusions has traditionally been low, reducing the efficiency of spheroplast fusion as a method of chromosome transfer. Some strategies are available to improve this. Addition of PEG, calcium, diacylglycerol and peptides have all been shown to increase the fusion success rate [43,44]. PEG for example, acts by volume exclusion effect that dehydrates the membrane interface making hemifusion more energetically favorable [45]. The details of the experimental protocol for addition of these factors, including their concentration, may have to be optimized on case by case basis to prevent cell toxicity.

Spheroplast fusion is a conceptually simple method, which is also experimentally straightforward to implement. Unfortunately, so far it has been limited to at most 1.1 Mbp (cf. Table 6.1), and suffered from low efficiency. More studies are needed to understand what sets the size limit and how to overcome it.

AGAROSE-PLUG BASED TRANSFER

Another method used to transfer chromosomes uses agarose plugs. It, too, aims to reduce sheering and pipetting of the isolated DNA. It achieves so by embedding cells, and eventually the isolated DNA, in agarose matrix. The plugs are prepared by heating and dissolving agarose powder in a buffer of choice. Traditionally, low-melting point agarose has been used owing to it requiring lower temperatures for melting which is eventually required. The cells are mixed into the solution after allowing it to cool down to around 37 - 42 °C. The agarose-cells suspension is then cast to plug molds and allowed to solidify at 4 °C, after which the plugs remain solid at room temperature and are ready for further

processing.

As in the case of spheroplast-based fusion, the wall of the donor cell must first be degraded. This is done in incubation with one or more corresponding cell-wall degrading enzyme(s). This can happen either before or after embedding the cells in the agarose plugs. Next, the cells in the plugs are gently lysed by treatment with proteinase and detergents. The chromosomal DNAs is released from the cell, but remains trapped in the agarose matrix. The plugs, and the therein contained DNA, can then be subject to number of handling and washing steps.

These treatment steps can include target-specific DNA modification (e.g. methylation) [41], enzymatic cleavage (e.g. for linearization), digestion of background genomic DNA, or cleanup by (pulse-field) gel electrophoresis. The handling of agarose plugs is generally low throughput, requiring a number of manipulations that are hard to automate. It has however historically been the go-to way for genomic transfer. It is also currently de-facto the only way for DNA isolation for cell-free applications.

Virtually all applications will eventually require the DNA to be released from the agarose matrix. To do so, the plugs are melted and enzymatically digested. It is clearly beneficial to keep the melting temperatures as low as possible to limit melting of the DNA double-strand, and low-melting point agarose has been used to achieve that. Yet, the melting process still leads to inevitable mechanical and temperature stresses. Although it is difficult to quantify the degree of damage, this clearly is associated with some reduction of DNA quality. Finally, the thus isolated DNA is ready to transformed into recipient cells or studied in-vitro.

Previous studies following this protocol have shown the ability to transfer 1.1 Mbp circular mycoplasma genome from yeast to another species of mycoplasma [41]. DNA cleanup, either by digestion of the background yeast DNA or by gel electrophoresis, did not have influence on the transformation success rate. The absolute quantity of the transplanted DNA, however, did influence the success rate, with efficiency peaking at 2.9 µg of *M. mycoides* genomic DNA per transplantation. No comparison was carried out for efficiency of transplantation of circular versus linear chromosome. To which extent the chromosome topology, content, or identity of the host and target species impact transfer efficiency remains an open question.

DEALING WITH CONTAMINATION

Above we highlighted the importance of minimizing mechanical damage during the chromosome transfer. Another challenge that can complicate the transfer process is the presence of contaminating DNA. Haploid *S. cerevisiae* natively carries 16 linear chromosomes, totaling 12 Mbp. Additionally, mitochondrial (mt) DNA is present in about 50-100 copies per a haploid cell [46], with each mtDNA nucleoid having a length of 86 kbp [47]. Finally, virtually every S. cerevisiae cell will carry 2-micron plasmid in the nucleus, on average in 40-60 copies per haploid cell. Each 2-micron plasmid is a highly stable 6.3 kbp long DNA element carrying only four genes. The sole known function of these genes

is to assure plasmid's propagation. Taken together, a haploid cell will carry over 13 Mbp of background DNA. This DNA will be lost through rounds of replication in a host that naturally selects for intact chromosomes with appropriate selection markers and chromosomal replication elements. However, this DNA will remain present in non-selective (e.g. cell-free) systems, and can scavenge system's limited resources and impede the synthetic chromosome's function. Isolation and transfer protocols for non-selective systems must account for the presence of background DNA and incorporate steps that reduce its abundance to minimum.

6.1.6. NEED FOR NEW APPROACHES

Despite the current limitations on size and low efficiency, as well as number of open questions about what sets them and how they can be alleviated, both spheroplast-fusion and agarose-plug based transfer have proven useful for cell-to-cell transfer of chromosomes up to 1 Mbp. However, there is currently no method that has been successfully used to isolate chromosomes for use in cell-free systems.

Simple approaches using off-the-shelf kits are limited only to small plasmids. Additionally, they suffer from background contamination by 2-micron plasmid and mtDNA which frequently compromises sustained in vitro expression (Céline Cleij, personal communication). Similarly, the spheroplast fusion method carries the drawback of transferring full chromosomal and intra-cellular background along the synthetic chromosome. The only way that this method could become relevant for in vitro studies would require fusion with a lipid vesicle, and some mechanism that would drive the loss or digestion of the background DNA. No such approach has been demonstrated, and establishing it will require multi-group multi-year effort. The agarose plug method, on the other hand, is more suitable for cell-free biology. It however carries number of drawbacks. It is laborious, requires specialized equipment, and includes inevitable DNA shearing. Most importantly, previous attempts to use this method to isolate synthetic chromosomes for use in cell-free systems have not been successful (Andrei Sakai – doctoral thesis thesis [48], Céline Cleij – personal communication). A clear understanding of what caused the failures is lacking, but likely has to do with excessive DNA damage, and the presence of contamination, including background DNA.

In other chapters of this thesis, we established tools and protocols for microfluidics-based chromosome isolation. This approach paves the way to single molecule whole-chromosome studies directly on the microfluidic chip. The technological toolbox thus established could be useful for on-chip chromosome isolation with optional washes with detergents and enzymes that could substantially decrease DNA background while preserving the synthetic chromosome's integrity. The isolated chromosomes could become subjects of mechanistic single molecule studies, potentially offering insights on the organization and function of synthetic chromosome organization and function. This, or similar, microfluidic system could also be employed to study the mechanical and molecular details of cell-to-cell chromosome transfer by spheroplast fusion. While promising, the microfluidic approach requires specialized training and equipment. Before microflu-

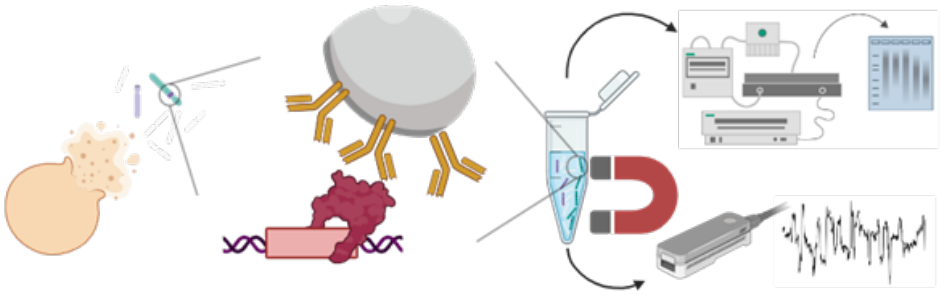

Figure 6.3: Overview of the immunoprecipitation based approach. Cells are engineered to contain array of binding sites on the target chromosome, and to express a protein (here in red) that bind at the sites. The cells are lysed and the target chromosome can be enriched for with an antibody-mediated pulldown (center). The degree of enrichment can be quantified with pulse field gel electrophoresis (PFGE), nanopore sequencing, or quantitative PCR (qPCR, not shown).

idics and surrounding hardware become more commonplace and easier to operate, different methods are needed to isolate synthetic genomes.

6.1.7. AIM OF THIS WORK

In the remainder of this chapter we describe early experiments towards the development of chromosome isolation with an approach that is potentially quick to execute, does not require specialized equipment, and can enrich for a specific chromosome independently of its topology.

Motivated by needs of researchers in cell-free and bottom-up biology, we sought to develop a method for isolation of synthetic chromosomes that is relatively easy to execute and does not require specialized equipment [49]. We hypothesized that an immunoprecipitation-based approach could achieve this, while also allowing to preserve chromosome integrity and reduce the DNA background. Immunoprecipitation is a routine, inexpensive, and widely used technique that isolates molecules from solution based on their interaction with a matching antibody coupled to solid support.

6.2. RESULTS AND DISCUSSION

We designed a strategy where the engineered chromosome harbors one or multiple regions that recruit a specific protein. Carrying out immunoprecipitation with an antibody for this protein could enable for target chromosome's enrichment (Fig. 6.3). It is important that this protein binds DNA tightly, and that the immunoprecipitation is carried out in a buffer that does not compromise this binding. Additionally, it is critical to use buffers that will cause DNA to compact, to minimize any damage that could experience by shear. To carry out the immunoprecipitation, the cells are gently lysed and their chro-

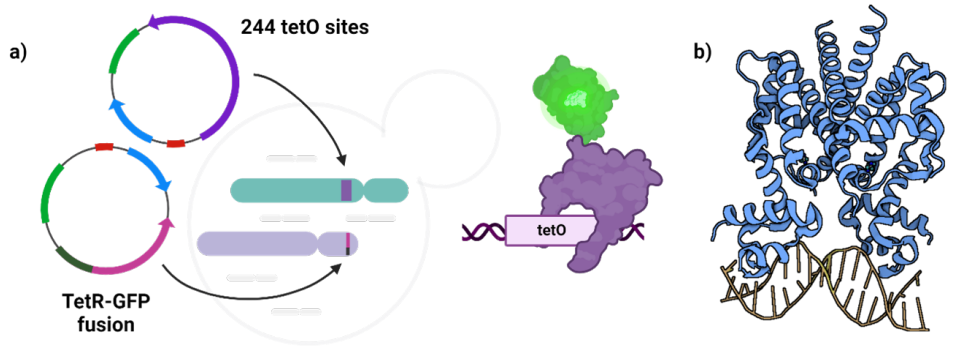

Figure 6.4: Cloning and DNA-recruitment of tetO array and TetR-GFP fusion. a) TetO binding sites were cloned near the synII chromosome centromere, and TetR-GFP fusion was ectopically expressed on chrXV. Illustration of binding of the fusion protein to DNA (center). b) 3D structure of the TetR homodimer in complex with DNA (PDB: 1QPI).

mosomes are released. The designer chromosome carrying the recruitment repeat array can be enriched for by immunoprecipitation with the antibody-functionalized magnetic beads. This in principle marks the end-point of the proposed protocol.

In the process of establishing the protocol, we however required some diagnostic readout. Here, we evaluated the degree of enrichment and intactness of the chromosome with pulse field gel electrophoresis (PFGE). We propose that future endeavors can obtain similar or more quantitative readout with long-read sequencing, or quantitative PCR (qPCR).

6.2.1. DNA SEQUENCE DESIGN

We reasoned that the localization of the binding sites on the chromosome could impact the isolation efficiency. We therefore opted for two extreme strategies for distributing them, either as a dense array at one location, or randomly throughout the chromosome. The second approach has the added benefit of requiring minimal genetic modification on top of the existing synthetic design. In both cases we benefited from parallel research lines ongoing at the Cai lab at the time of the fellowship, which allowed us to make use of available synthetic strains already carrying these modifications.

For the case of a concentrated array at one location, we chose a tandem repeat array of *tetO* binding sites that recruit the Tet repressor protein (TetR, Fig. 6.4). TetR is a transcriptional regulator conferring tetracycline antibiotic resistance in large number of bacterial species. The protein functions as a homodimer, with one DNA binding region per monomer, and it recognizes a 15 base pair palindromic sequence TCCCTATCAGTGATAGAGA. The dissociation constant for TetR is commonly reported in the single-digit nM range, indicating a strong binding affinity [50,51]. TetR has been engineered to either

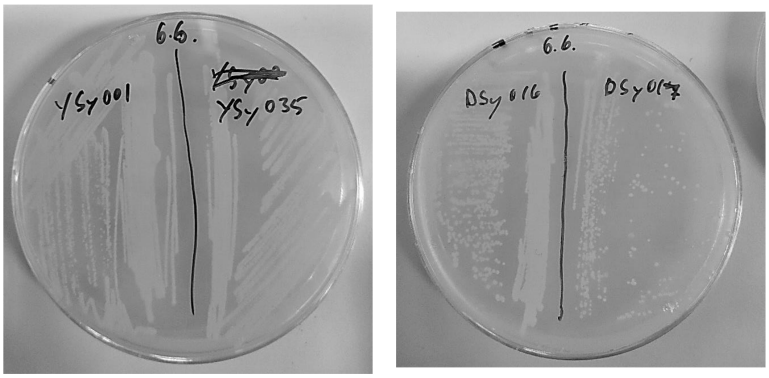

Figure 6.5: Confirmation of viability for the synthetic strains. See Table 6.3 for strain nomenclature.

bind DNA natively, or in complex with tetracycline and its analogues. This forms the basis of the commonly used tetracycline-inducible expression system.

Name	Other Name	Description
MHy001	YSy001	synII + *tetO* array
MHy002	YSy035	BY4741 + *tetO* array
MHy003	DSy016	neo-tRNA + *tetO* array, clone 2
MHy004	DSy017	neo-tRNA + *tetO* array, clone 1
MHy005	-	BY4741

Table 6.3: Overview of strains used for the TetR-GFP based pulldown.

Specifically, we worked with two synthetic chromosome strains, one carrying 770 kbp synII chromosome [32], and one with 190kb tRNA neo-chromosome [52]. The synII strain carried a 11 kbp long locus harboring 224 repeats of the *tetO* integrated 15 kbp to the right of the synII chromosome centromere region (CEN2), with URA3 used as selective marker for integration. The repeat array was practically implemented as 32 repeats of the tetracycline responsive element (TRE). Each TRE contained 7 tetO sites. The strain simultaneously carried a TetR-GFP expression cassette under the control of URA3 promoter integrated on chrXV, with LEU2 as a selective marker. The second strain carried a linearized version of the tRNAneo chromosome, with the same *tetO* array integrated randomly on the neochromosome, and the TetR-GFP expression cassette integrated to HO locus (ChrIV). We confirmed the strains' viability by plating it on selective media plates (Fig. 6.5). A detailed description of the strain and their generation is in the Materials & Methods. For the approach of dispersed binding sites, we leveraged the fact the synthetic chromosomes that we worked with included *loxPsym* sites – see the discussion of this approach in the Section "Pulldown by Targeting Dispersed Chromosomal Loci".

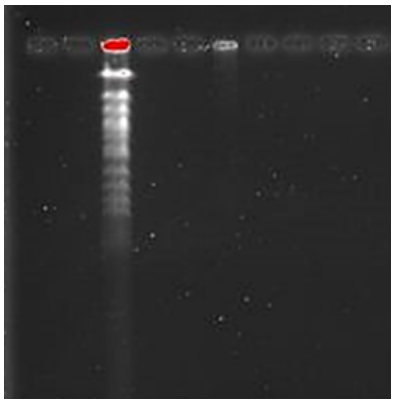

Figure 6.6: First successful PFGE only with the yeast genomic ladder.

6.2.2. PULSE-FIELD GEL ELECTROPHORESIS VISUALIZATION OF CHROMOSOMES

We sought a facile and low-cost method for assessing the pulldown quality. Here, we opted for PFGE for number of reasons. Aside from being readily available and cost-effective, we also considered previous experience with it, and the interest of other host-laboratory members in establishing the technique. Finally, it also offers the convenience of a visual output and straightforward interpretation. The downside of this approach is that it requires embedding the isolated chromosomes into agarose plugs, which is in fact something we were trying to avoid in the first place. However, this is different from agarose-plug isolation methods where cells are embedded before lysis, and where the embedding would be necessary for the chromosome isolation. Here, we carry out immunoprecipitation first and embed the isolate to plug only in order to run diagnostic PFGE. Once established, the protocol would not require running PFGE anymore. Looking ahead, we suggest that long-read sequencing, or qPCR, ultimately represent more suitable techniques for such a diagnostic readout, as they could be more standardized, require less hands-on time, and yield more quantitative output.

The PFGE protocol was not fully in place at the moment of our arrival to the host laboratory. Establishing a working version of it was therefore the first priority in the project. To do so, we had to optimize number of steps. In the first instance this included standardizing gel running conditions, and the way the gel is cast. After few tests this resulted in the ability to reproducibly resolve the *S. cerevisiae* genomic ladder on the pulse-field gel (Fig. 6.6).

6.2.3. ESTABLISHMENT OF THE AGAROSE-PLUG PROTOCOL

As a next step, we aimed to run the genomes of the synthetic and wild type strains on the gel. Running a wild type (MHy005) would remove the need for using genomic lad-

der, as well as having the advantage of using a reference that is closely related to the synthetic strain. Running synthetic strains (MHy001-004) would then allow to have a one-to-one comparison of before and after-pulldown conditions for the targeted chromosome. In order to be able to run these strains on the PFGE, we had to establish the agarose plug protocol. This included standardizing cell grow conditions and the number of cells used, as well as optimizing the cell wall digestion, the procedure of embedding cells into agarose plugs, and their lysis.

Briefly, we first grew cells overnight and resuspended them in an osmo-protective buffer. Next, we treated them with zymolyase, and casted them into agarose plugs. The plugs were then washed in a low osmolarity buffer and could be stored for up to a week. Finally, plugs were treated with Proteinase K. The samples were run on an 1% agarose pulse-field gel (Fig. 6.7 and Table 6.4). The detailed protocol is given in the Materials & Methods.

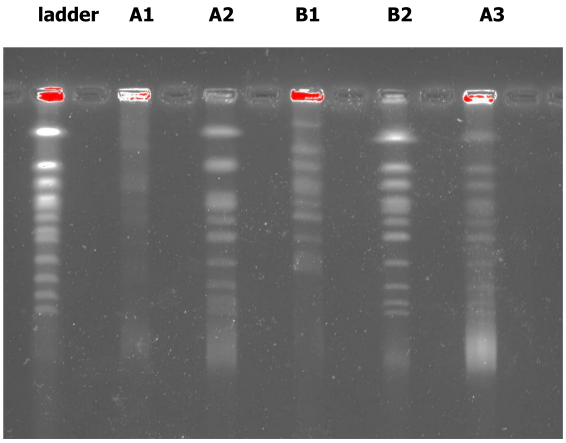

Figure 6.7: First successful PFGE with WT and two different synthetic strains.

ID	ladder	A1	A2	B1	B2	A3
Treatment	/	+ZL ON	+ZL ON +PK 3h	+ZL ON	+ZL ON +PK 3h	+ZL 3h +PK ON

Table 6.4: Sample annotation for Figure 6.7. A corresponds to MHy005 and B to MHy001. Further strain details are given in Table 6.3 and Table 6.6.

The best results were obtained when treating samples overnight with zymolyase, followed by 3 hours of Proteinase K treatment (A2 and B2 in Fig. 6.7). Proteinase K was necessary for good resolution of the chromosomes on the gel and together with extended zymolyase treatment, it contributed to cell lysis (B1 vs B2). Treating cells with Proteinase K overnight was shown to lead to some extra DNA fragmentation (A2 vs A3). Notably, we saw a difference in results between the two used strains A (MHy005) and B (MHy001)

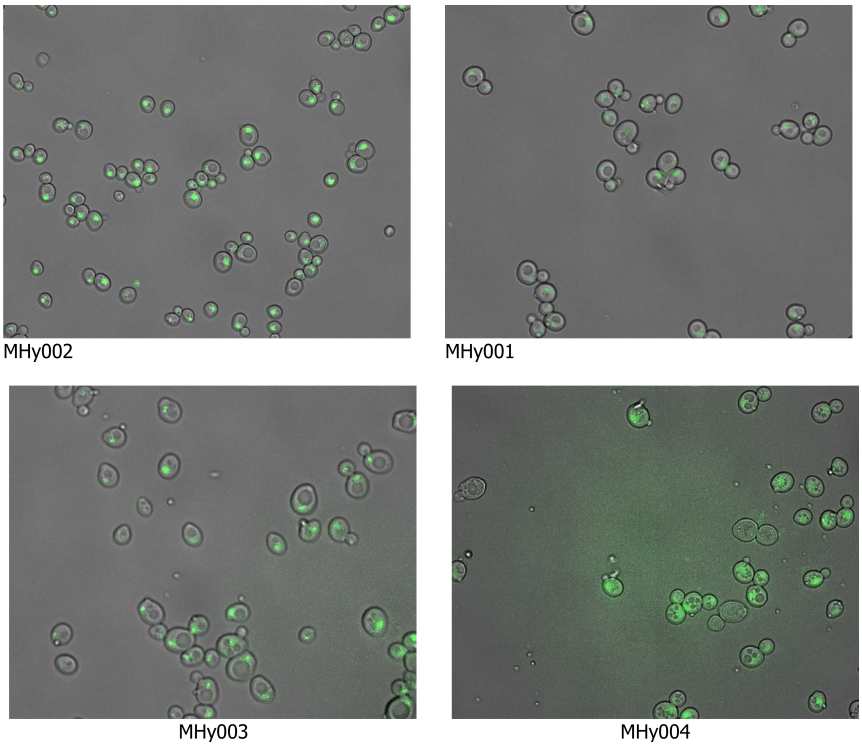

Figure 6.8: Synthetic strains showing TetR-GFP recruitment to a single genomic location.

(cf. A1 vs B1 and A2 vs B2), that we also recapitulated in later experiments. The treatment conditions may have to be optimized if different strains are to be used.

6.2.4. IMAGING TETR-GFP RECRUITMENT TO CHROMOSOMES

Next, we sought to confirm that the strains expressed TetR-GFP that can be recruited to a single genomic locus. To do so, we plated the strains on selective plates (Table 6.3), inoculated an overnight culture in corresponding selective media, and imaged it on a fluorescence microscope. We observed a bright fluorescent spot for all strains that we tested (Fig. 6.8). This confirmed that the expression cassettes were functional, and that the binding array was stably integrated. Altogether, these experiments provided a good starting point for establishment of the immunoprecipitation-based pulldown. We describe this next.

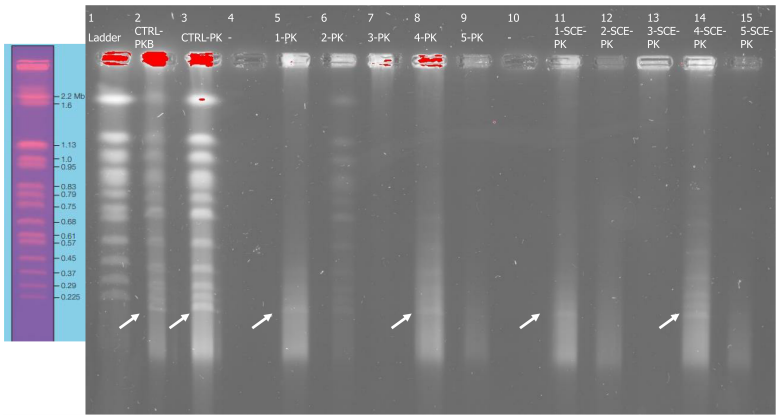

Figure 6.9: Enrichment of synthetic chromosome (tRNA-neo) after pulldown. White arrows indicate the bands corresponding to tRNA-neo (~190 kb) before (rows 2 and 3) and after pulldown (rows 5, 8, 11 and 14).

6.2.5. ANTIBODY-MEDIATED PULLDOWN OF SYNTHETIC CHROMOSOMES

In the preceding sections, we demonstrated that the necessary prerequisite for antibody-mediated pulldown were in place. This included strain construction, PFGE protocol establishment, as well as agarose plug embedding standardization. Next, we proceeded with development of the antibody mediated pulldown of the synthetic chromosomes. Here we provide an overview, with the detailed protocols in the Materials & Methods section.

Briefly, cells were inoculated from -80 °C stocks and grown for 16-20 hours under auxotrophic selection. The cultures were concentrated to a target OD, washed, and resuspended in osmo-protective buffer. Next, cells were incubated with zymolyase while shaking, after which they were lysed by osmotic shock. The lysate was immediately used for pulldown.

In order to carry out the immunohistochemistry mediated pulldown, we first coupled the anti-GFP antibody to a solid support. Here we used magnetic beads as they allowed for facile buffer exchange. We always prepared the functionalized beads fresh, generally carrying out the coupling reaction while the cells were undergoing zymolyase treatment. Specifically, we incubated GFP monoclonal antibody with the magnetic beads in weight-to-weight ratio 1:750. We used the buffer specified by the beads' manufacturer, but extended the reaction time to 30 min. The unbound antibody was then washed away, and the beads stored at 4 °C until use (for at most 2 hours).

We carried out the immunoprecipitation reaction by combining functionalized beads with the cell lysate. To be able to do so, we optimized buffer composition so as to obtain conditions that cause cell lysis, while also being favorable to protein-DNA binding. We dubbed the final composition "binding and lysis buffer" (BLB, details in Materials & Methods). First, we resuspended the functionalized beads in this buffer. Next, we lysed

cells by resuspending spheroplast in the same buffer. Next, we carried out the immunoprecipitation reaction for at least 15 minutes at room temperature while gently tumbling. After the incubation period, we washed away unbound sample and eluted the DNA from the beads by using a custom elution buffer. Instead of using a buffer provided by manufacturer that would decouple the antibody from the solid support, we chose a gentler composition that should compromise only the protein-DNA binding. This would effectively release the isolated chromosome from the antibodies, as the TetR-GFP fusions unbind from DNA. The optimized buffer was dubbed "ProtK & PEG buffer" (PKPB, for details see the Materials & Methods). Importantly, both BLB and PKPB contained 5% PEG-8000, a concentration at which we expect chromosomes to be compacted, and thus protected from shear stress. Fig. 6.9 shows results from a pulldown run that resulted in enrichment of the ~190 kbp tRNA neo-chromosome. (Fig. 6.9, white arrows). Specifically, we quantified the enrichment as mean band intensity of the target tRNA-neo chromosome against the mean band intensity of the ~1Mbp chromosome VII, resulting in enrichment factor ranging between 6 to 15-fold (lane 8 and 14 respectively).

6.2.6. SCREENING CELL LYSIS CONDITIONS

While we thus demonstrated the successful enrichment for the target chromosome, we observed variability in the robustness of the cell lysis. We sought to understand this better, and hypothesized that variability could be to some degree attributed to the cell growth stage. This was corroborated by earlier result that suggested that lysis and treatment conditions were strain dependent (Fig. 6.7). We therefore proceeded to screen different approaches for cell lysis. Notably, we were not merely looking for a protocol that would cause cells to lyse, which would have been straightforward. Instead, we sought a protocol, that lyses cells robustly, and across different strains, without disrupting protein-DNA binding that is necessary for immunoprecipitation-based enrichment that we were pursuing.

Unfortunately, resuspension in BLB buffer alone was not sufficient to guarantee consistent lysis, unless immediately followed by ProtK treatment (Fig. 6.10). However, we reasoned that using ProtK is undesirable, as this would lead to need for additional protocol step for its inactivation. Its concentration would need to be tightly selected so as to avoid intracellular protein digestion. Proteinase has also previously been seen to exhibit some degree of DNA-digestion activity (Anthony Birnie, personal communication). We therefore tested further buffer compositions that could potentially remove the need the use of a protease.

Advantageously, resuspension of zymolyase treated spheroplasts in PK(P)B (Proteinase K buffer, optionally with 5% PEG 8000) led to some degree of cell lysis even in the absence of ProtK (Fig. 6.10-11). Curiously, we observed that combining a PK(P)B treatment with ProtK treatment resulted in some loss of DNA signal, pointing to potential DNA-digestion activity of the protease. Taken together, we were able to identify conditions that cause robust cell lysis. Unfortunately, we were not able to test evaluate the efficiency of the antibody-mediated pulldown in these conditions due the limited duration

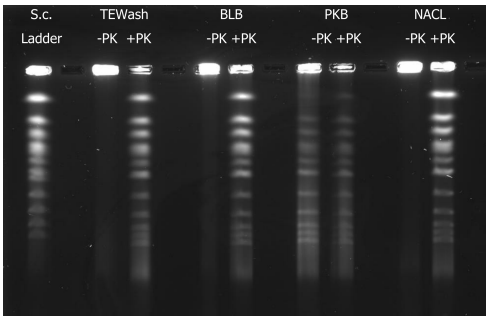

Figure 6.10: Different approaches to cell lysis. Only PKB (Proteinase K buffer) is able to lyse cells without the help of Proteinase K treatment (-PK).

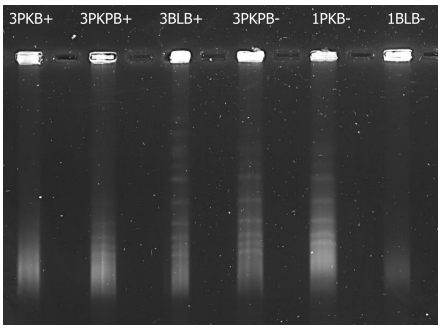

Figure 6.11: In addition to PKB and PKPB, also BLB immediately followed by ProtK treatment (BLB+) leads to cell lysis. 1 – MHy001 (synII), and 3 – MHy003 (tRNA-neo) strains.

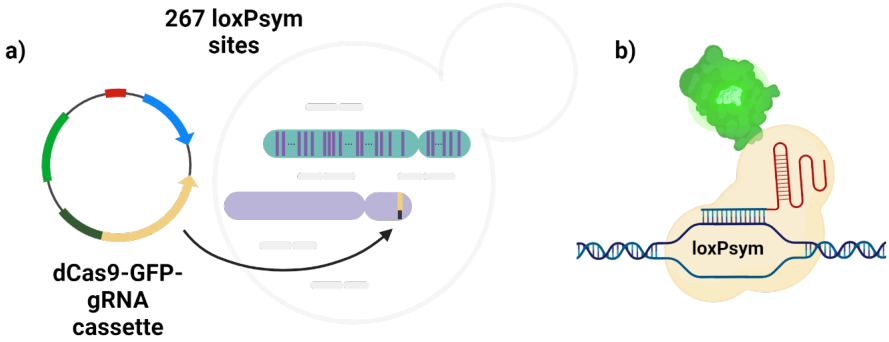

Figure 6.12: Targeting loxPsym sites on the synthetic chromosome. a) Synthetic chromosomes will contain several hundreds of dispersed *loxPsym* sites. SynII (770 kbp) used here contains 267 sites. dCas9-GFP-gRNA cassette is expressed ectopically. b) gRNA guides the dCas9-GFP fusion to *loxPsym* sites.

of the fellowship.

6.2.7. PULLDOWN BY TARGETING DISPERSED CHROMOSOMAL LOCI

Above, we described an approach of targeting a single extended locus on a designer chromosome, a tandem repeat array of *tetO* sites. Many synthetic chromosome designs will already include a similar array natively. Artificial telomeres and centromeres in mammalian chromosome designs are the most common examples [53]. In other instances, however, a tandem repeat array will not be part of the design. In these cases, researchers would usually prefer to target a different, dispersed, repeat site to avoid additional cloning. Additionally, it is not clear whether target a single concentrated locus, or multiple dispersed loci, would lead to a higher pulldown efficiency.

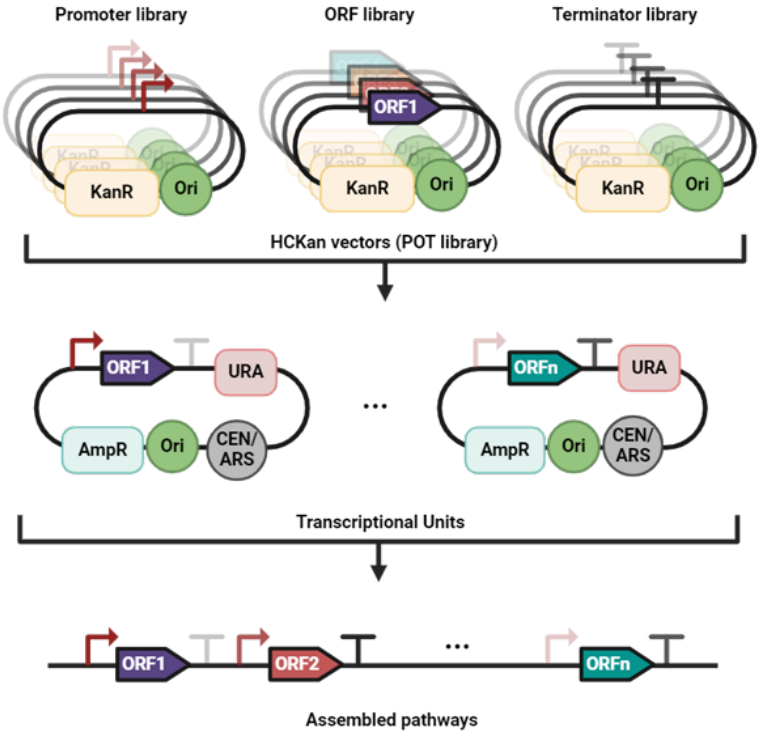

Figure 6.13: Overview of the YeastFab cloning strategy. The standard parts (promoters, open reding frames, and terminators) are cloned into standard vectors. These parts are modular and can be assembled into transcriptional units (Tus) on standard (shuttle) vectors. It is possible to include counterselection against self-closed backbone (not shown). Finally, TUs can be assembled into pathways on a plasmid, or integrated into genome. ORF – open reading frame, URA – auxotrophy selection marker.

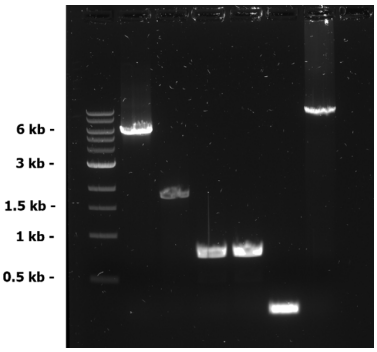

Figure 6.14: PCR amplification of individual TUs as well as linearization of the pYAC10 backbone.

Lane	Size	Sequence Annotation
1	5'642 bp	dCas9-GFP(1-155)
2	1'871 bp	MCP-GFP(156-173)
3	798 bp	gRNA loxP (AGT)
4	798 bp	gRNA loxP (CGA)
5	250 bp	Empty fragment
6	10'421 bp	pYAC10

Table 6.5: Size and identity annotation for fragments in Fig. 6.14.

Consequently, and in parallel to the research described above, we pursued the establishment of a strategy targeting randomly spread out sites on the designer chromosome (Fig. 6.12). There are multiple choices for a dispersed site design. These can include promoter sites, or sites for restriction enzymes and recombinases.

Here we opted for targeting the *loxPsym* sites (viz. the 'Synthetic Chromosome Design' section). We designed an expression unit for dCas9-GFP-(2xgRNA) on a pYAC10 plasmid. Transforming this plasmid to a strain harboring a synthetic chromosome with *loxP(sym)* sites should lead to expression of dCas9-GFP fusion that will be guided to these sites by the co-expressed guide RNAs. Notbaly, to target a different site, e.g. Dre recombinase cutting site *rox* present in 276 copies on tRNA neo-chromosome [52], one only needs to adjust the gRNA sequence.

6.2.8. dCAS9-GFP-GRNA PLASMID DESIGN AND ESTABLISHMENT OF STRAINS

To build the dCas9-GFP-(2xgRNA) expression unit, we made use of the YeastFab assembly (Fig. 6.13). YeastFab is modular cloning strategy devised specifically for rapid pathway construction in *S. cerevisiae* [54]. Briefly, each gene is split into three parts: promoter, the open reading frame (ORF) and a transcriptional terminator. Each of these is then cloned into a high-copy plasmid carrying a kanamycin resistance marker (HCKan). Next, these vectors are combined in a single reaction ('one-pot') to assemble individual transcriptional units (TUs) on shuttle vectors (POT vectors). These vectors can be maintained and propagated in bacteria, and eventually used for final construct assembly in-yeast.

To improve the stability of the inserts, we first transferred the TUs onto Gateway cloning compatible vectors. These vectors allow for selecting against clones that carry the backbone, but not the insert of interest. Finally, we sought to assemble the individual fragments onto a backbone in the MHy005 (BY4741) strain. We initially attempted transformation with gel-isolated fragments, but could not obtain any viable clones. PCR-amplifying the individual TUs (Fig. 6.14, Table 6.5) and co-transforming them with pYAC10 shuttle vector resulted in number of colonies. Randomly picking four colonies and amplifying an internal overhang of the assembly indicated a 50% success rate of the transformation (Fig. 6.15).

As a next step, we continued by isolating these plasmid and transforming them to a synII yeast strain (this strain did not carry TetR-GFP expression unit). Imaging the resulting cultures under the microscope showed green fluorescence and confirmed that the transcriptional units were successfully expressed (Fig. 6.16). Notably, only about 40% of the cells showed green signal, which is less favorable than what we observed for the TetR-GFP strains (Fig. 6.8). Additionally the signal appeared more dispersed within the cell volume. It is not clear whether this signals potential issues with the dCas9-GFP protein recruitment to DNA, or faithfully represents the organization of the chromosome at this stage. Next step should confirm the plasmid sequence by sequencing. Finally, and to

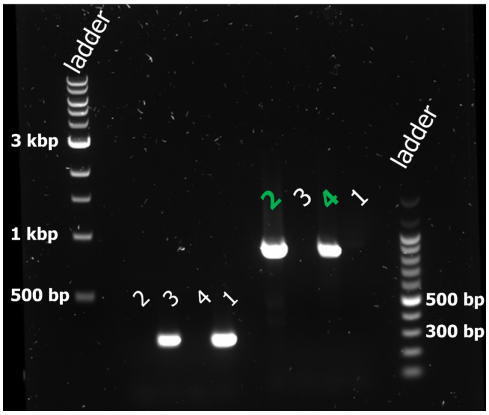

Figure 6.15: PCR on either self-closed backbone (first four) or an overhang internal to the assembly (second four) suggesting correct assembly of two out of four colonies assayed. Amplification on self-closed backbone with primers MHp001/MHp004 (297 bp). Amplification on internal assembly overhang with primers MHp013/MHp016 (937 bp). Ladders: left 1 kb, right 50bp. Details in Materials & Methods.

improve the rate at which cells carry the insert, we suggest to integrate the insert directly into a genomic locus. Unfortunately, we were not able to pursue these steps due to the limited duration of the fellowship.

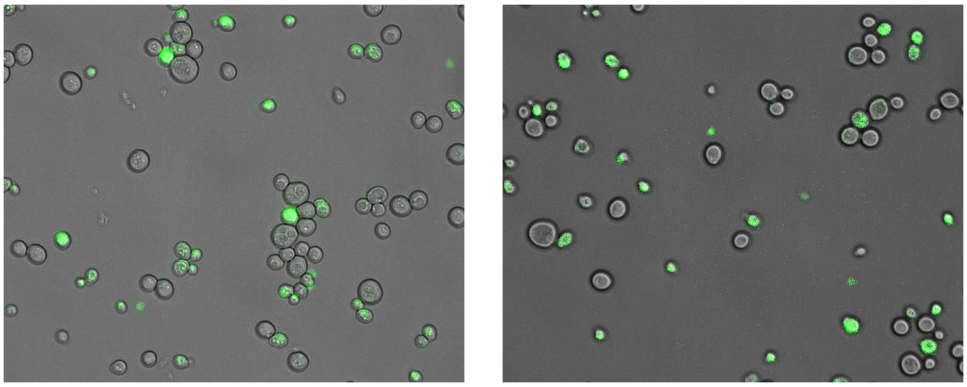

Figure 6.16: Result of transformation of plasmid #2 and #4 (from Fig. 6.15) to a synII strain.

6.3. SUMMARY AND OUTLOOK

In summary, we obtained initial proof of feasibility for an antibody-mediated yeast synthetic chromosome pulldown. We achieved that by targeting a single compact site on the tRNA-neo chromosome. Given the results presented in this chapter we conservatively anticipate that the immunoprecipitation approach we developed will allow for enrichment of <200 kb chromosomes.

The short duration of the fellowship made it impossible to carry out further experiments and improve the robustness of the method. Further research should characterize the isolated product in more detail. Specifically, long read sequencing and quantitative PCR can be used to validate the chromosome integrity. This should also be done in the context of evaluation whether targeting a single dense array or dispersed sites along the chromosome yields the more complete product. Next, using the chromosome in cell-free expression system could highlight whether the method sufficiently removes background DNA contamination. Finally, comparative studies of pulldown of differently sized chromosomes, both larger and smaller, as well as chromosomes with different topologies (linear vs circular) will enable further exploring the applicability and limitations of this method.

6.4. MATERIALS AND METHODS

6.4.1. SYNTHETIC YEAST STRAINS AND GROWTH CONDITIONS

All yeast strains were derivatives of BY4741 (which is in turn derived from S288C) and grown at 30 °C unless otherwise specified. Yeast cells were grown in either YPD media (10 g/L yeast extract, 20 g/L peptone) or Synthetic Complete (SC-8) media with auxotrophic selection, both supplemented with 2% glucose. Standard solid media contained 2% agar and cultures were incubated at 30 °C.

Name	Other Name	Description	Genotype	Reference
MHy001	YSy001	synII + *tetO* array	MATa his3Δ1 leu2Δ0 LYS2 met15Δ0 ura3Δ0 synII::tetOarray::URA3 chrXV::tetR-GFP:LEU2	[32]
MHy002	YSy035	BY4741 + *tetO* array	MATa his3Δ1 leu2Δ0 LYS2 met15Δ0 ura3Δ0 chrII::tetOarray::URA3 chrXV::tetR-GFP:LEU2	[32]
MHy003	DSy016	Linearized neo-tRNA + *tetO* array, clone 2	MATa leu2Δ0 met15Δ0 ura3Δ0 his3Δ1 HO::leu2::tetR-gfp [Syn.tRNA-neo(HIS3)-Ter3Δ::TeSS-ura3 ::tetOarray] [pRS415(LEU2)-I-SceI] AMp327	[52]
MHy004	DSy017	Linearized neo-tRNA + *tetO* array, clone 1	ditto	[52]
MHy005	-	BY4741	MATa his3Δ1 leu2Δ0 LYS2 met15Δ0 ura3Δ0	[55,56]

Table 6.6: Overview of yeast strains used in this work.

6.4.2. FLUORESCENCE MICROSCOPY

Yeast cultures were grown overnight in corresponding selection medium. The cultures were then pelleted, washed with PBS, diluted 1000x in MQ water, pipetted onto a microscopy glassed and imaged under coverslip on an oil immersion fluorescence microscope at 60x or 100x magnification.

6.4.3. PREPARATION OF SPHEROPLASTS

Cells were grown overnight in corresponding SC selection medium at 30 °C. The next day, their OD was measured at 100-fold dilution. Cells corresponding to OD equivalent of 20 were spun down at 4'000 x g for 5 min and the pellet was washed once in PBS. The pellet was then resuspended in 500 µL SCEM (1M sorbitol, 0.01 M EDTA pH 8, 100 mM $Na_3C_6H_5O_7$ pH 5.8) solution, and 300 units of zymolyase (E1004, Zymo Research Corporation, CA, USA) and 2% β-mercaptoethanol were added. The sample was incubated at 37 °C, most commonly for 1 to 3 hours. The spheroplasts were then collected by gentle centrifugation at 400 x g for 4-5 min at 4 °C, and resuspended in either SCEM (control) or an immunoprecipitation sample preparation buffer (BLB, described later).

6.4.4. PREPARATION OF AGAROSE PLUGS

2% low-melting point (LMP) agarose (V2831, Promega, WI, USA) solution was prepared by first homogenizing the agarose powder in SCEM solution at 75-80 C for 10+ minutes, and then equilibrating it at 40-42 °C until needed. Plug molds (Bio-Rad Laboratories, Inc., CA USA) were cleaned with 70% ethanol, then washed with water and allowed to dry. Samples, either control or after treatment, were incubated for 10+ minutes at 40 °C, and combined with 2% agarose solution in volume to volume ratio of 1:1 using a cut 1 mL tip and pipetting slowly. The mixture was loaded to agarose plug molds that were sealed at the bottom with parafilm, and incubated at 4 °C for 30+ min to allow plugs to solidify.

6.4.5. PROTEINASE K TREATMENT

Plugs were incubated individually in 1.5 mL eppendorf tubes in 1 mL ProtK buffer (10 mM EDTA pH 8, 0.2% sodium deoxycholate, 1% N-Lauryl Sarcosine, 20 mM Tris-HCl pH 8) and 0.5 mg of Proteinase K (P8107S, New England Biolabs, MA, USA) for 1 hour at 55 °C without shaking. The plugs were then washed three times in TE buffer, with 10+min incubation at 4 °C on a tumbler between each wash. The plugs could be stored in TE buffer at 4 °C for up to two weeks.

6.4.6. PULSE FIELD GEL ELECTROPHORESIS (PFGE)

CHEF-DR® III PFGE (Bio-Rad Laboratories, Inc., CA USA) system was used to resolve chromosomes on a gel. Before every use, the system was washed with 3 L of MQ for 30 minutes. Next the PFGE instrument was washed twice with 3 L of 1 x TAE buffer for 30+ min, while cooling to 14 °C. In the meantime, 1% agarose (BP1356-500, Fisher Scientific Inc., PA, USA) gel was cast in 0.5x TAE, and allowed to settle at 4 °C, after which it was ready to use. Special attention was taken to position the comb so as to cast wells deep enough into the gel, leaving about 1 mm of clearance. Once the instrument was ready to use, wells in the gel were prefilled with 20-50 µL of running buffer. Next the agarose plugs were halved and loaded into the wells with flat tweezers and a coverslip using ster-

ile technique. CHEF DNA Size Marker (0.2–2.2 Mb) was used as a ladder (#1703605, Bio-Rad Laboratories, Inc., CA USA). The PFGE was then started using a single block program of 4.5 V, 120° angle, initial settling time 20.2 s, final settling time 175 s, and runtime of 20.2 hours. After the program finished, the gel was stained by washing it in 200 mL of 3x staining dye (GelRed® Nucleic Acid Gel Stain #41001, Biotium Inc., Fremont, CA, USA) and 100 mM NaCl added for 30 minutes on a tumbler. Finally, the gel imaged in ChemiDoc imaging system (Bio-Rad Laboratories, Inc., CA USA). The PFGE system was flushed with 3 L MQ for 30+ min after each use.

6.4.7. IMMUNOPRECIPITATION-BASED CHROMOSOME PULLDOWN

Anti-GFP antibody (#14-6674-82, eBioscience™, CA USA) was coupled to magnetic beads as per manufacturers protocol (#10007D, Dynabeads™Protein G Immunoprecipitation Kit, Thermo Fisher Scientific Inc, MA, USA), using 12.5 µL of beads solution, and 2 ug of antibody per 200 µL of total buffer volume. The antibody-bound beads were resuspended in 200 µL Ab-binding and lysis buffer (BLB, 100 mM NaCl, 5% PEG 8000, 1 mM DTT, 20 mM Tris-HCl pH 8, 10 mM EDTA ph8). Spheroplast were prepared as described above. Spheroplast pellet was gently resuspended in 400 µL of BLB buffer and combined with 50 µL of magnetic beads. The samples were incubated for 15 minutes at room temperature while tumbling. The beads were then separated on magnet and supernatant was collected for further analysis. Next, the beads were gently resuspended in 300 µL of elution buffer (PKPB, 10 mM EDTA pH 8, 0.2% sodium deoxycholate, 1% N-Lauryl Sarcosine, 20 mM Tris-HCl pH 8, 5% PEG 8000, 1 mM DTT) and incubated at room temperature for 5 minutes while tumbling. Finally, the beads were separated on a magnet and supernatant was collected for analysis, which most commonly consisted of preparation of agarose plugs and PFGE.

6.4.8. YEAST TRANSFORMATION PROTOCOL

Yeast cultures were inoculated into 10 mL YPD and grown for 16-20 hours at 30 °C with rotation. The next day, the cells were diluted to 25 mL at OD_{600} of 0.2 and incubated at 30 °C until OD_{600} reached 0.8. Cells were then transferred to 50 mL falcon tubes and washed twice with 12.5 mL of MQ, with each centrifugation at 4'000 x g for 5 minutes. Next, cells were resuspended in 1 mL 100 mM lithium acetate, transferred into 1.5 mL Eppendorf tube, centrifuged at 13'000 x g for 15 seconds, and resuspended in 250 µL of 100 mM lithium acetate. 50 µL of cells were combined with 350 µL of transformation mix (final concentrations in 400 µL: 33.3% PEG 8000, 100 mM lithium acetate, 125 ug salmon sperm DNA) that was vortexed for 30+ s until fully combined and included 100+ ng of to-be-transformed DNA. After addition of cells, the transformation mix was again vortexed for 30+ s and incubated with rotation at 30 °C for 1 hour. Next, the cells were heat-shocked at 42 °C for 20 minutes, centrifuged at 4'000 x g for 3 minutes, resuspended in 200 µL of MQ. Finally, the cells were plated on selection media agar plates and incubated at 30 °C.

6.4.9. DNA RESTRICTION DIGESTION & GEL PURIFICATION

Typical restriction digestion and gel purification protocol included 0.25 – 1 ug of plasmid DNA, 0.5 – 1 µL of restriction enzyme, and 1 µL or rCutSmart 10x Buffer (B6004S, New England Biolabs, MA, USA) per 10 µL of volume. The volume was frequently scaled to 30 - 50 µL per reaction, and the reaction time and temperature was as specified by the enzyme's manufacturer. Each reaction was loaded on 1% agarose gel with 1x of loading dye (B7024S, New England Biolabs, MA, USA). After the gel was run, and imaged, the bands were excised, and the contained DNA was purified using Zymoclean Gel DNA Recovery kit (D4007/D4008, Zymo Research Corporation, CA, USA), using <10 µL of elution buffer prewarmed to 60 °C. Samples concentration was measured using the NanoDrop instrument (NanoDrop 2000, Thermo Fisher Scientific Inc, MA, USA).

6.4.10. PLASMID PURIFICATION DIRECTLY FROM YEAST

Yeast cultures were inoculated into 25 mL selective media and grown for 16-20 hours at 30 °C with rotation. The next day, the 20 mL of culture was spun down at 4'000 x g, resuspended in 500 µL of P1 buffer (Qiagen Plasmid Prep Kit #12125, Qiagen, DE) and vortexed until the pellet dissolved. Next, 200 µL of 2000 units/mL lyticase (ICN Biomedicals, OH, USA) was added and the suspension incubated at 37 °C for 60 min. Cells were lysed by addition of 500 µL P2 buffer (Qiagen Plasmid Prep Kit #12125, Qiagen, DE), gentle inversion, and incubation at 22 °C for 10 min. Next, 700 µL of buffer N3 (Qiagen Plasmid Prep Kit #12125, Qiagen, DE) was added to the lysate, the tube content was mixed gently by inversion and incubated on ice for 30 min. Each tube was spun down at 10'000 x g for 10 min at 4 °C, after which the remainder of the Qiagen Plasmid Prep Kit protocol was carried out, yielding purified plasmid sample. Finally, concentration of DNA in samples was measured using the NanoDrop instrument (NanoDrop 2000, Thermo Fisher Scientific Inc, MA, USA). The reaction ususally yielded 100 – 300 ng/µL of plasmid DNA. 10 µL was generally used per each 100 µL of cells in a transformation reaction.

6.5. REFERENCES

[1] Martufi, M. *et al.* Single-Step, High-Efficiency CRISPR-Cas9 Genome Editing in Primary Human Disease-Derived Fibroblasts. *The CRISPR Journal* **2**, 31–40 (2019).

[2] Ruffolo, J. A. *et al.* Design of highly functional genome editors by modeling the universe of CRISPR-Cas sequences. Preprint at (2024).

[3] Postma, E. D. *et al.* A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Research* **49**, 1769–1783 (2021).

[4] Caruthers, M. H. *et al.* Total Synthesis of the Gene for an Alanine Transfer Ribonucleic Acid from Yeast. **227**, (1970).

[5] Stemmer, W. P. C., Crameri, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49–53 (1995).

[6] Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391 (1994).

[7] Itakura, K. *et al.* Expression in *Escherichia coli* of a Chemically Synthesized Gene for the Hormone Somatostatin. *Science* **198**, 1056–1063 (1977).

[8] Goeddel, D. V. *et al.* Expression in Escherichia coli of chemically synthesized genes for human insulin. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 106–110 (1979).

[9] Goeddel, D. V. *et al.* Direct expression in Escherichia coli of a DNA sequence coding for human growth hormone. *Nature* **281**, 544–548 (1979).

[10] Cello, J., Paul, A. V. & Wimmer, E. Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template. **297**, (2002).

[11] Gibson, D. G. *et al.* Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome. *Science* **319**, 1215–1220 (2008).

[12] Fredens, J. *et al.* Total synthesis of Escherichia coli with a recoded genome. *Nature* **569**, 514–518 (2019).

[13] Ostrov, N. *et al.* Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016).

[14] Richardson, S. M. *et al.* Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017).

[15] Boeke, J. D. *et al.* The Genome Project-Write. *Science* **353**, 126–127 (2016).

[16] Dai, J. *et al.* A spotlight on global collaboration in the Sc2.0 yeast consortium. *Cell Genomics* **3**, 100441 (2023).

[17] Yu, W. *et al.* Designing a synthetic moss genome using GenoDesigner. *Nat. Plants* (2024).

[18] Lindeboom, T. A. *et al.* An Optimized Genotyping Workflow for Identifying Highly SCRaMbLEd Synthetic Yeasts. *ACS Synth. Biol.* **13**, 1116–1127 (2024).

[19] Zhao, Y. *et al.* Debugging and consolidating multiple synthetic chromosomes reveals combinatorial genetic interactions. *Cell* **186**, 5220-5236.e16 (2023).

[20] Brooks, A. N. *et al.* Transcriptional neighborhoods regulate transcript isoform lengths and expression levels. *Science* **375**, 1000–1005 (2022).

[21] Zhang, H. *et al.* Systematic dissection of key factors governing recombination outcomes by GCE-SCRaMbLE. *Nat Commun* **13**, 5836 (2022).

[22] Luo, Z. *et al.* Compacting a synthetic yeast chromosome arm. *Genome Biol* **22**, 5 (2021).

[23] Zhang, W. *et al.* Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science* **355**, eaaf3981 (2017).

[24] Shao, Y. *et al.* A single circular chromosome yeast. *Cell Res* **29**, 87–89 (2019).

[25] Lartigue, C. *et al.* Genome Transplantation in Bacteria: Changing One Species to Another. *Science* **317**, 632–638 (2007).

[26] Gambogi, C. W. *et al.* Efficient formation of single-copy human artificial chromosomes. (2024).

[27] Koster, C. C., Postma, E. D., Knibbe, E., Cleij, C. & Daran-Lapujade, P. Synthetic Genomics From a Yeast Perspective. *Front. Bioeng. Biotechnol.* **10**, 869486 (2022).

[28] Annaluru, N., Ramalingam, S. & Chandrasegaran, S. Rewriting the blueprint of life by synthetic genomics and genome engineering. *Genome Biol* **16**, 125 (2015).

[29] Shen, Y. *et al.* SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. *Genome Res.* **26**, 36–49 (2016).

[30] Zhang, W. *et al.* Manipulating the 3D organization of the largest synthetic yeast chromosome. 17.

[31] Shen, Y. *et al.* Dissecting aneuploidy phenotypes by constructing Sc2.0 chromosome VII and SCRaMbLEing synthetic disomic yeast. *bioRxiv* 2022.09.01.506252 (2022).

[32] Shen, Y. *et al.* Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science* **355**, eaaf4791 (2017).

[33] Marschall, P., Malik, N. & Larin, Z. Transfer of YACs up to 2.3 Mb intact into human cells with polyethylenimine. *Gene Ther* **6**, 1634–1637 (1999).

[34] Gibson, D. G. *et al.* Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **329**, 52–56 (2010).

[35] Karas, B. J. *et al.* Transferring whole genomes from bacteria to yeast spheroplasts using entire bacterial cells to reduce DNA shearing. *Nat Protoc* **9**, 743–750 (2014).

[36] Karas, B. J. *et al.* Direct transfer of whole genomes from bacteria to yeast. *Nat Methods* **10**, 410–412 (2013).

[37] Brown, D. M. *et al.* Efficient size-independent chromosome delivery from yeast to cultured cell lines. *Nucleic Acids Res* gkw1252 (2016).

[38] Mizutani, M. *et al.* Cloning and sequencing analysis of whole Spiroplasma genome in yeast. *Front. Microbiol.* **15**, 1411609 (2024).

[39] Tagwerker, C. *et al.* Sequence analysis of a complete 1.66 Mb Prochlorococcus marinus MED4 genome cloned in yeast. *Nucleic Acids Research* **40**, 10375–10383 (2012).

[40] Benders, G. A. *et al.* Cloning whole bacterial genomes in yeast. *Nucleic Acids Research* **38**, 2558–2569 (2010).

[41] Lartigue, C. *et al.* Creating Bacterial Strains from Genomes That Have Been Cloned and Engineered in Yeast. *Science* **325**, 1693–1696 (2009).

[42] Cevc, G. & Richardsen, H. Lipid vesicles and membrane fusion. *Advanced Drug Delivery Reviews* **38**, 207–232 (1999).

[43] Lentz, B. R. PEG as a tool to gain insight into membrane fusion. *Eur Biophys J* **36**, 315–326 (2007).

[44] Sáez, R., Alonso, A., Villena, A. & Goñi, F. M. Detergent-like properties of polyethyleneglycols in relation to model membranes. *FEBS Letters* **137**, 323–326 (1982).

[45] Boni, L. T., Stewart, T. P. & Hui, S. W. Alterations in phospholipid polymorphism by polyethylene glycol. *J. Membrain Biol.* **80**, 91–104 (1984).

[46] Göke, A. *et al.* Mrx6 regulates mitochondrial DNA copy number in *Saccharomyces cerevisiae* by engaging the evolutionarily conserved Lon protease Pim1. *MBoC* **31**, 527–545 (2020).

[47] Mortimer, R. K. & Johnston, J. R. GENEALOGY OF PRINCIPAL STRAINS OF THE YEAST GENETIC STOCK CENTER. *Genetics* **113**, 35–43 (1986).

[48] Sakai, A. Towards the bottom-up construction of a minimal synthetic cell.

[49] Blount, B. A., Driessen, M. R. M. & Ellis, T. GC Preps: Fast and Easy Extraction of Stable Yeast Genomic DNA. *Sci Rep* **6**, 26863 (2016).

[50] Huang, H.-H., Seeger, C., Helena Danielson, U. & Lindblad, P. Analysis of the leakage of gene repression by an artificial TetR-regulated promoter in cyanobacteria. *BMC Res Notes* **8**, 459 (2015).

[51] Aleksandrov, A., Schuldt, L., Hinrichs, W. & Simonson, T. Tetracycline-Tet Repressor Binding Specificity: Insights from Experiments and Simulations. *Biophysical Journal* **97**, 2829–2838 (2009).

[52] Schindler, D. *et al.* Design, construction, and functional characterization of a tRNA neochromosome in yeast. *Cell* **186**, 5237-5253.e22 (2023).

[53] Dawe, R. K. Engineering better artificial chromosomes. *Science* **383**, 1292–1293 (2024).

[54] Garcia-Ruiz, E., Auxillos, J., Li, T., Dai, J. & Cai, Y. YeastFab: High-Throughput Genetic Parts Construction, Measurement, and Pathway Engineering in Yeast. in *Methods in Enzymology* vol. 608 277–306 (Elsevier, 2018).

[55] Winston, F., Dollard, C. & Ricupero-Hovasse, S. L. Construction of a set of convenient *saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast* **11**, 53–55 (1995).

[56] Baker Brachmann, C. *et al.* Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115–132 (1998).

7

BIOFOUNDRIES AND CITIZEN SCIENCE CAN ACCELERATE DISEASE SURVEILLANCE AND ENVIRONMENTAL MONITORING

A biofoundry is a highly automated facility for processing of biological samples. In that capacity it has a major role in accelerating innovation and product development in engineering biology by implementing design, build, test and learn (DBTL) cycles. Biofoundries bring public and private stakeholders together to share resources, develop standards and forge collaborations on national and international levels. In this paper we argue for expanding the scope of applications for biofoundries towards roles in biosurveillance and biosecurity. Reviewing literature on these topics, we conclude that this could be achieved in multiple ways including developing measurement standards and protocols, engaging citizens in data collection, closer collaborations with biorefineries, and processing of samples. Here we provide an overview of these roles that despite their potential utility have not yet been commonly considered by policymakers and funding agencies and identify roadblocks to their realization. This document should prove useful to policymakers and other stakeholders who wish to strengthen biosecurity programs in ways that synergize with bioeconomy.

This chapter has been written as part of my engagement with the iGEM Policy & Governance Network. It has been been published as Holub M. and Agena E. (2023), Biofoundries and citizen science can accelerate disease surveillance and environmental monitoring. *Front. Bioeng. Biotechnol.* doi:10.3389/fbioe.2022.1110376.

7.1. INTRODUCTION

Humans have always been at prey to natural pathogens. There have been at least fifteen epidemics with a death toll over 1 million in the last 500 years (one every thirty three years on average). Two occurrences of bubonic plague, a bacterial respiratory infection, in the 6th and 14th century wiped out an estimated half of the worldwide population. Spanish flu, a viral respiratory infection, caused tens of millions of deaths in the early 20th century. More recently, the coronavirus pandemic caused millions of deaths worldwide. While the most shocking due to their rapid development, pandemics are only one of major global health risks. Another global health risk is due to antibiotic resistance. Increasingly prevalent among pathogens, it is causing an increase in the number of deaths due to bacterial infections globally [1]. Furthermore, as we become increasingly able to edit and engineer living organisms, man-made pathogens could be at the source of future health threats as well.

Driven to protect ourselves from the often-lethal forces of nature, we as humans have learnt to shape our environments in many ways early on. From building shelters to growing crops, these efforts have paid out wildly, testified by how well we have done as a species. It has been only very recently, however, that we are developing more appreciation for how we have influenced and continue to influence the natural environment around us in this process. Environmental pollution, temperature change and biodiversity loss are just some examples. One of the less known consequences is an emergence of novel urban ecosystems that give rise to novel species [2].

Risks to the health of humans and our environment must be monitored, as any attempts to manage and contain it in the future will have to rely on data to be effective. Biosurveillance (detection of biological threats to human health) and environmental monitoring (observation and characterization of the natural environment) are both processes of provisioning this data. In the recent case of coronavirus pandemic, biosurveillance through routine testing and contact tracing on the level of individuals has proved to be crucial to the coronavirus pandemic response worldwide. Additionally, aggregate monitoring of coronavirus through wastewater sampling has proved to be a predictive signal to case counts and hospital load independent of direct diagnostic data [3–6]. In a similar fashion, the benefits of biological monitoring have been seen for targets other than infectious disease such as tracking of bacterial antibiotic resistance in the environment [7], and even conservation efforts through the analysis of environmental DNA [8].

7.1.1. CURRENT BOTTLENECKS IN BIOLOGICAL MONITORING

Despite some successes, biological monitoring programs generally fall short on a multitude of levels when it comes to preparedness for detection and prevention of future biological risks. While, to our knowledge, there is no resource comprehensively reviewing and comparing biosecurity programs across the world, Nuclear Threat Initiative (NTI) has compared 195 countries in terms of preparedness for pandemics and epidemics in Global Health Security Index (`www.ghsindex.org`). The United States ranked number

one in 2021 and this, together with its being relatively well researched in academic literature, is one of the main reasons why we use it as an illustrative example. It is likely that the US system is average or above-average compared to biodefense systems across the world and that its shortcomings will reflect common shortcomings worldwide.

The main shortcomings of the US biodefense as reviewed by the Bipartisan Commission on Biodefense [9] are lack of quick response capability [10] and general lack of structured investment, lack of adequate data interoperability and data collection standards, and poorly developed regulatory structure. Several biosurveillance bottlenecks, such as insufficient testing and processing capacity, where at one point a single facility was responsible for handling samples nation-wide, became manifest during the COVID-19 pandemic and limited the speed of delivering public health interventions [11]. This ultimately encouraged establishment of more distributed testing sites and accessing unconventional sequencing facilities for diagnostic work, such as academic laboratories [12].

Coupled with the increased public awareness of biosecurity as result of the pandemic, along with the identification of bottlenecks in current biosurveillance programs, the question arises: Is there a different way to structure biosurveillance programs that could improve outcomes? In this paper we argue for options to do so by considering the newly emerging infrastructure of highly automated facilities for processing of biological samples, biofoundries (cf. Tools for Rapid and Robust Biological Surveillance, Fig. 7.1). In the following sections we discuss how this infrastructure can be exploited to benefit not only response to disease outbreaks, but also the response to more subtle targets in health, ecology, and biosecurity. We identify several opportunities at this interface, most of which have not been commonly considered by policymakers and funding agencies. These include developing measurement standards and protocols, engaging citizens in data collection, decentralized manufacturing (cf. Tools for Rapid and Robust Biological Surveillance), and processing of samples. We then finish by highlighting roadblocks to their realization. In this vision we focus on biofoundries that are run and funded by the public sector. While industry-owned biofoundries exist and undoubtedly deliver value, they may be subject to unique agendas of their owners and we do not see them as a suitable foundation of national biosecurity. In contrast, we believe that less-formal infrastructure for biological experimentation, such as bio-hack spaces and bio-DIY labs, can contribute to these ends in various ways, including increasing the impact of citizen scientists, as well as encouraging safe practices, through collaboration with biofoundries and community engagement. However, due to specific challenges these spaces currently face, including lack of appropriate regulatory schemes, issues with securing suitable lab space and equipment, as well as negative sentiment among broad public, we anticipate that their contribution will develop only as they mature on medium and long term. We leave them therefore out of scope of the present discussion and refer interested reader to recent reviews on the topic [13–16].

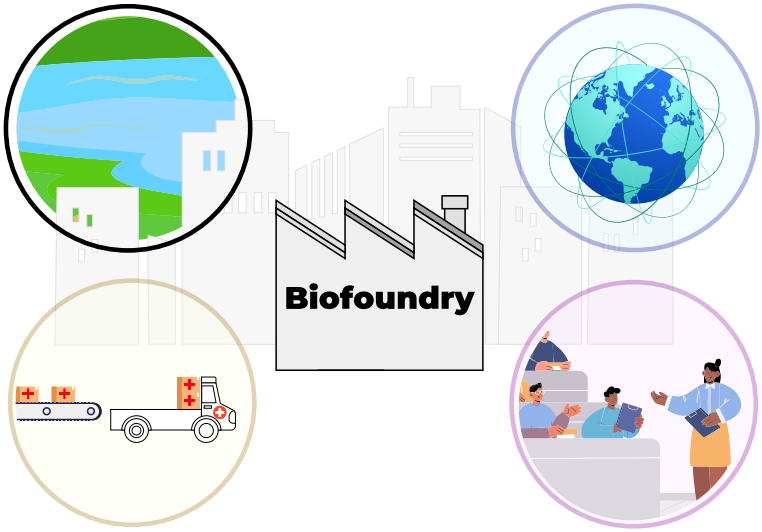

Figure 7.1: A vision for the future role of biofoundries and citizen science. Biofoundries are local hubs that are close to urban areas, and foster citizen engagement through citizen science (top left) and education (bottom right). Global network of biofoundries cooperates to share protocols and data (top right), which further strengthens the capacity of individual biofoundries to safeguard biosecurity and implement interventions.

7.2. TOOLS FOR RAPID AND ROBUST BIOLOGICAL SURVEILLANCE

7.2.1. BIOFOUNDRIES

A biofoundry is a highly automated facility for processing of biological samples. In that capacity it has a major role in accelerating innovation and product development in engineering biology by implementing design, build, test and learn DBTL cycles [17] (Fig. 7.2). The equipment in biofoundries typically include automated liquid handling systems, high-throughput sequencing and chemical analysis equipment, and a software ecosystem for data and personnel management [17]. For example, one of the largest biofoundries and synthetic biology companies in operation today, Ginkgo Bioworks, has leveraged their integrated system of automated bioengineering to evaluate on the order of tens of thousands of engineered strains [18] — a quantity that cannot be achieved with bench-scale workflows alone. Dropping costs of DNA synthesis and sequencing, development of facile technologies for genome editing, lab-on-chip microfluidics, and expanding ecosystem of hardware and software automation tools are some of the main factors that contribute to synthetic biology as an engineering discipline. The growth of bioeconomy enabled by these technological advances goes hand in hand with the increasing popularity of biofoundries. The establishment of the Global Biofoundry Alliance (GBA), which has grown to over 30 members since 2019 [17], including 14 biofoundries in Australia and Asia, 9 in North America and 10 in Europe, is a sign of the

continued growth of this sector. Importantly, first steps towards establishment of biofoundries in Latin America [19] and Africa [20] are already underway.

Aside from their direct role in biological experimentation, biofoundries serve as platforms that bring public and private stakeholders together to share resources, develop standards and forge collaborations on national and international level [10] . In that capacity they can gather sufficient momentum to realize collaborative projects that may need top-down incentive or broader consensus for economic viability (e.g. projects contributing to environmental sustainability), contribute to development of legal and ethical frameworks by shaping governance of emerging fields [21] and manage the relationship with the public. Despite their obvious utility, the high establishment, personnel and overall running costs make the business case for biofoundries difficult. While there is early evidence that biofoundries deliver high added value through innovation and knowledge creation [22], it is useful to consider additional roles for biofoundries that could strengthen their business case, which could further rationalize their establishment in countries with lower research budgets.

7.2.2. CITIZEN SCIENCE

Citizen science, which is the involvement of the public in scientific research, can range from collecting and analyzing data to prototyping low-cost sensing devices. Digitalization of our society and adoption of open-data and open-innovation paradigms are the main contributors to its rise in recent two decades [23]. The main benefits of citizen science are two-fold: 1) citizen science contributes to and expands research, and 2) it shapes the relationship between scientists and the public in an engaging, two-way interaction [24,25]. The first benefit enables a larger breadth of research than what is achievable by an academic laboratory alone, e.g. collection of data at higher spatial resolution, or making measurements of completely new parameters. The latter allows citizens to familiarize themselves with the scientific method and gain insight on interpretability and accuracy of collected data, as well as reciprocally provide feedback on collected data and the process of its acquisition. Recent incorporation of citizen science concepts into university [26,27] and high-school [28] curricula suggest that its impact will continue to rise.

7.2.3. CELL-FREE SYNTHETIC BIOLOGY

Standardization could be facilitated by adoption of cell-free systems (CFS). CFS could also contribute to a shift towards decentralization of manufacturing. Cell-free gene expression is gaining popularity in synthetic biology and bioengineering [29]. Diverse applications including protein production, therapeutics manufacturing and biosensing all can benefit from by-passing living cells. Benefits include facilitated rapid prototyping and condition screening, reduction of reaction volumes, higher predictability and amenability to mathematical modelling. Consequently, cell-free biomanufacturing is of imminent interest also beyond academia. Furthermore, engineered cell-free systems are not classified as genetically engineered organisms [30,31], which simplifies biosafety and

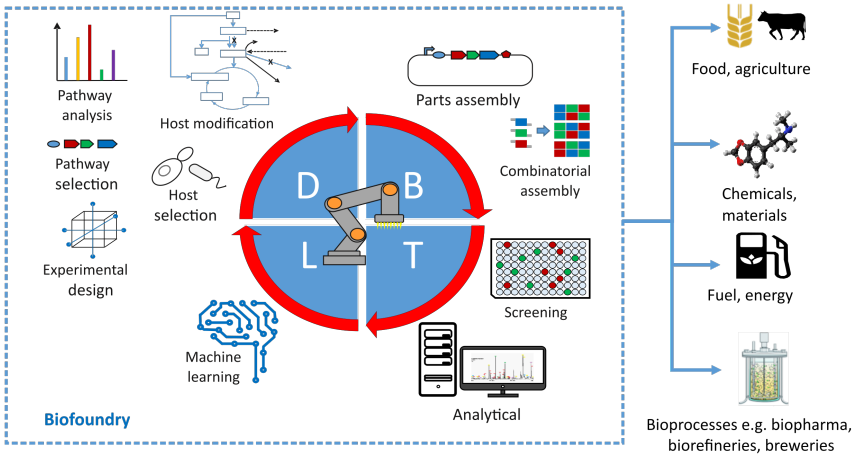

Figure 7.2: Overview of major processes in a biofoundry happening at design (D), build (B), test (T) and learn (L) stages of the development cycle. Reprinted with permission from ref. [22].

biosecurity of their application.

Adoption of cell-free systems further decreases batch-to-batch variability [32], reduces sample volumes and lowers regulatory barriers. Biofoundries are particularly suited to drive the transition to decentralized biomanufacturing through adoption of cell-free systems. The integrated design, build, test, learn cycle, the automation facilities for liquid handling, and the standardization in biofoundries are all vital to rapid, scalable and reproducible processes. Geographical distribution of biofoundries allows them to serve as local hubs [17], out of which products based on CFSs can be rapidly deployed, for instance in the case of response to health and environmental crises.

7.3. STANDARDIZED AND AUTOMATED MEASUREMENT WORKFLOWS FACILITATE BIOSURVEILLANCE

The cornerstone of biofoundry operations is the melding of automation of standardized bioengineering workflows and the design, build, test, learn cycle. Without these principles implemented, the difference in throughput achieved by biofoundries as compared to typical laboratories would not be possible (cf. Tools for Rapid and Robust Biological Surveillance for a general introduction to biofoundries). The outcomes of engineering biology can be variable due to the complexity of biological systems and the magnitude of unknowns and confounding factors. Thus, by leveraging automation technologies throughout sampling, processing, and analysis, as much human variability is removed from the process which allows for gains in consistency of results while shortening the timescale of workflows. This approach is also suitable for processing many samples at the same time, allowing to explore unprecedented breadth of genetic variability.

While mainly employed for sample processing, experimentation, and analysis by academics and researchers, biofoundries are also well suited to boost our ability to rapidly collect and analyze samples originating from patients, or the environment. In a recent example, Ginkgo Bioworks has used its high-throughput sequencing capabilities to support nation-wide efforts in COVID-19 testing, as well as supported vaccine manufacturers in optimizing their products [33]. On a similar note, automatized routines adopted at biofoundries, as well as their equipment, make them good candidates for handling samples with pathogenic potential.

Aside from automated processing of high numbers of samples, biofoundries are particularly suited for development of measurement standards and standardized calibration samples. Their nature as a collaborative platform, that can interface with governmental entities, further facilitates encouragement and adoption of so developed standards [21]. In the context of biosurveillance, adoption of these standards enables comparison of results across time and geographical regions and enables their users to harmonize interventions. An example is provided by London Biofoundry, which developed a rapid automated SARS-CoV-2 testing platform that was deployed and scaled in national diagnostic labs and could be adopted also by other biofoundries [34].

7.4. BIOSURVEILLANCE ENABLED BY BIOFOUNDRIES AND CITIZEN SCIENTISTS

Areas that can benefit from citizen science (cf. section 7.2.2) are diverse. With an aging population and increasing obesity rates on one hand, and ongoing prevalence of malnutrition, in both developed and developing countries, on the other [35], public health monitoring emerged as an important area for application of citizen science. In The American Gut project [36] scientists receive stool samples from the public with the aim of identifying the relationships between health and lifestyle and the microbiome. The 100 For Parkinson's project [37] invited people across the UK and United States to track their health for 100 days with a mobile app, with the aim of understanding how technology can support Parkinson patients. The Seattle Flu Study [38] focuses on studying seasonal influenza, aiming to understand how it develops and spreads in the Seattle area. Participants are typically asked to regularly answer simple survey questions and if they are identified as high-risk, they are sent a testing kit and asked to submit the swab back by post or to report the result of a self-test. Thanks to the high number and broad distribution of samples, The Seattle Flu Study was among the first to discover and identify COVID-19 in the Seattle area [39], clearly highlighting the utility of citizen science in public health monitoring and protection. Overall, these examples demonstrate the utility citizen science programs outside of conventional academic and medical studies on assessing healthcare outcomes and impacts.

Synthesizing the capabilities of biofoundry facilities with the breadth of sampling possible with citizen-based science programs described above brings to light a new conception for biological monitoring and surveillance. When considering the limitations of

citizen science programs, in terms of the input variability and the magnitude of samples collected, leveraging the processing pipeline of a biofoundry may allow more consistent results to be obtained. Furthermore, biofoundries could act as formal knowledge hubs which if engaged appropriately with the local community could facilitate the quality of input from citizen scientists. Both these aspects could encourage the establishment of more citizen science programs as biofoundries can effectively reduce some of the technical hurdles associated with citizen science. As another consideration, the automation technologies leveraged in biofoundries also enables the incorporation of additional engineering controls in the handling of hazardous samples that could be-risk many hazardous biosurveillance targets. Overall, the synergies between biofoundry automation and standardization, and the collaborative nature of biofoundries as interface between public and private sectors are all factors that point to utility and feasibility of expanding the applications of citizen science to more elusive biosurveillance targets that could strengthen existing biodefense programs and could have positive impacts our ability to monitor the environment and public health.

7.5. BIOSECURITY-RELATED ACTIVITIES ARE SOURCE OF FUNDING AND DIRECTION OF DEVELOPMENT FOR BIOFOUNDRIES

Biofoundries are useful to the communities of their users as hubs with dedicated instrumentation and support of skilled staff. Sample handling and processing can be automatized and standardized, carried out at small and medium scale rapidly and reproducibly. Resulting data are appropriately stored and processed, often in cooperation with trained bioinformaticians. However, acquiring and maintaining dedicated equipment carries cost. Equally importantly, salaries of highly-skilled employees, together with costs for consumables for experiments, contribute to high running costs of a biofoundry [40]. Consequently, putting together a viable business model for biofoundry is challenging.

Above we have outlined how biofoundries can foster and support biosecurity, bio- and environmental-surveillance efforts by various means including standardization of samples and protocols, engagement with citizen scientists, and interface with decentralized manufacturing facilities. We believe that these further strengthen rationale for structural public investment into biofoundries and that national security agencies, environmental protection agencies, and related institutions can reap substantial benefits from channeling some of their financial resources into their operations. Aside from enabling biosurveillance, such effort contributes to training of staff at the forefront of biological engineering and biorisk and environmental monitoring, which is a valuable asset for national economy and security both long and short term. Furthermore, such trained staff, at the disposition of biofoundry infrastructure, will be instrumental to establishment of biosecurity training programs for professionals across the fields of security, intelligence and law reinforcement. This was recently exemplified by hands-on introductory to synthetic biology developed in collaboration between the Federal Bureau of Investigation

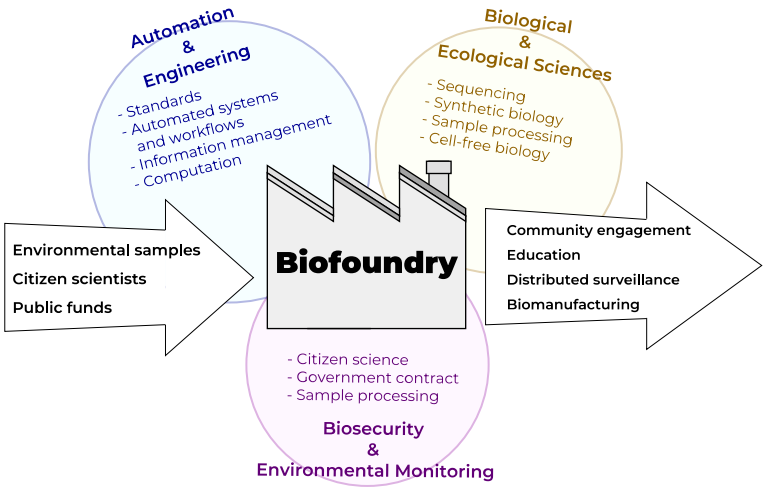

Figure 7.3: Biofoundries at the nexus of automation technologies, bioengineering, and biological/ecological monitoring interfacing with citizen science programs.

(FBI) and the Colorado State University (CSU) [41].

Recent years have seen growing interest in and implementation of decentralized biomanufacturing facilities (also called biorefineries) [42]. While decentralized manufacturing will likely develop infrastructure separate from biofoundries, there is potential for their synergy in bioeconomy as well as in health and biosecurity targets. Biofoundry-enabled surveillance would likely lead to shorter feedback cycles, earlier risk detection and ability to respond more locally to potential outbreaks. Such response could be further sped up by access to localized biomanufacturing facilities that would have the ability to develop therapeutic or other responses.

Similarly as the ability to produce crops locally contributes to food supply chain security and sustainability, so will decentralized biomanufacturing contribute to local security and sustainability. The rise of the bioeconomy suggests that this contribution will play out on multitude of levels including therapeutics, materials, fuels and food.

7.6. CONCLUSIONS

Biological risks, including pandemics or rapid rise of antimicrobial resistance, are commonly regarded as potentially existential to humanity [43]. Even if not fatal, biological catastrophes and engineered attacks have the potential to significantly impact lives of many, spreading rapidly to large geographical areas. Biosecurity therefore should be a critical priority for national security agencies (NSAs) worldwide. Similarly, climate change leads to gradual change of environmental conditions impacting ecosystems globally that are also existential threats to humanity. Accurate, wide-spread and time-

resolved monitoring is crucial for effective interventions and policy making in these scenarios.

Biosurveillance at the required level of spatial and temporal resolution remains challenging. Required number of samples and collection points is usually high. Moreover, samples may be perishable or pathogenic, complicating transport. In this paper, we have argued that biofoundry facilities can support several ways to improve our ability to carry out biosurveillance. They can function as distributed hubs of data collection and analysis, empowering biosurveillance by reducing transport times. Their distributed nature further confers the system with robustness, e.g. in case of targeted attack. They can play a key role in developing standard protocols and standardized samples and work with citizens to develop new sample collection schemes. Finally, they can collaborate with biorefineries for small scale rapid production of therapeutic compounds.

While there is potential for the vision presented in this paper (Fig. 7.3), biofoundries worldwide are still in their early stages of development and such biosurveillance programs have challenges barring implementation. We have identified some key barriers, as well as some directions to address these below:

- **Develop Biosecurity Policy to Leverage Biofoundries.** Foremost, biofoundries may not be eligible for biosurveillance related operations and or funding as they may not qualify for the correct biosafety clearance in their jurisdiction. Regulatory frameworks and granting programs, which differ jurisdiction to jurisdiction, should be reviewed with biofoundries in mind so that appropriate amendments, that support the biosecurity capacity of biofoundries, can be identified. Additionally, with the continued creation of biofoundries worldwide, it is imperative that a unified development of standards be created and adopted such that the benefit of standardization can be preserved between nations.

- **Design Biofoundries with Sufficient Biosafety Level**. Bifoundries are currently mostly designed and classified at the biosafety level 1. In order to be able to use their capacities for broad-spectrum pathogen monitoring, they will have to classify for biosecurity level 2 clearance. There is a need for collaboration between biofoundries and biosafety regulators to apply and adapt the regulations to biofoundry use cases.

- **Expand Use Cases for Biofoundries to Include Citizen Science.** Citizen science programs may not be currently considered as a part of a biofoundry's use cases. Thus, a biofoundry's engagement with citizens and citizen science groups may not be adequate and could preclude their use by these groups. Therefore, it is recommended that established, and up and coming biofoundries, ensure that citizens and citizen science groups are included in the development of their facilities and invited to participate in biofoundry operations.

- **Create Incentives to Encourage Biofoundry Establishment.** As biofoundries are at the confluence of automation and biological technologies, they have the potential to closely cooperate with decentralized biomanufacturing facilities, and cat-

alyze their further emergence. With the increasing growth in this sector, incentives for the establishment of biofoundries should be put forth as it could not only enable efforts in engineering biology, but could also help drive the transition to a circular bioeconomy.

- **Equip Future Biologists with Quantitative and Engineering Skills.** While many universities have adapted their study programs and include increasing amounts of quantitative, programming and even hardware skills in their curricula, these efforts require broader adoption to build a future workforce that can effectively work at the nexus of technology and biology and continue to push it forward. As biofoundry operations and related facilities become more common, the need for such skills will continue to rise.

Biofoundries are growing in prevalence year over year, and this growth highlights the importance of assessing the role biofoundries can play in a nation's biosecurity program. Synergies with citizen science could potentially extend the breadth of biosurveillance to more subtle targets than before by leveraging biofoundry facilities. Should the concepts in this paper be implemented, it could have transformative impacts on the way we monitor health, ecology, and biosecurity, by distributing the load among a network of biofoundries.

7.7. REFERENCES

[1] Zhang, Z. *et al.* Assessment of global health risk of antibiotic resistance genes. *Nat Commun* **13**, 1553 (2022).

[2] Danko, D. *et al.* A global metagenomic map of urban microbiomes and antimicrobial resistance. *Cell* **184**, 3376-3393.e17 (2021).

[3] Venugopal, A. *et al.* Novel wastewater surveillance strategy for early detection of coronavirus disease 2019 hotspots. *Current Opinion in Environmental Science & Health* **17**, 8–13 (2020).

[4] Zhu, Y. *et al.* Early warning of COVID-19 via wastewater-based epidemiology: potential and bottlenecks. *Science of The Total Environment* **767**, 145124 (2021).

[5] Calderón-Franco, D., Orschler, L., Lackner, S., Agrawal, S. & Weissbrodt, D. G. Monitoring SARS-CoV-2 in sewage: Toward sentinels with analytical accuracy. *Science of The Total Environment* **804**, 150244 (2022).

[6] Calderón-Franco, D., van Loosdrecht, M. C. M., Abeel, T. & Weissbrodt, D. G. Free-floating extracellular DNA: Systematic profiling of mobile genetic elements and antibiotic resistance from wastewater. *Water Research* **189**, 116592 (2021).

[7] Huijbers, P. M. C., Flach, C.-F. & Larsson, D. G. J. A conceptual framework for the environmental surveillance of antibiotics and antibiotic resistance. *Environment International* **130**, 104880 (2019).

[8] Thomsen, P. F. & Willerslev, E. Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation* **183**, 4–18 (2015).

[9] Bipartisaon Commision on Biodefense. *The Athena Agenda: Advancing The Apollo Program for Biodefense. Bipartisan Commission on Biodefense.* https://biodefensecommission.org/wp-content/uploads/2021/01/Apollo_report_final_v7_031521_web.pdf (2022).

[10] Vickers, C. E. & Freemont, P. S. Pandemic preparedness: synthetic biology and publicly funded biofoundries can rapidly accelerate response time. *Nat Commun* **13**, 453 (2022).

[11] Boeckh, M. *et al.* The Seattle Flu Study: when regulations hinder pandemic surveillance. *Nature Medicine* **28**, 7–8 (2022).

[12] Vanuytsel, K. *et al.* Rapid Implementation of a SARS-CoV-2 Diagnostic Quantitative Real-Time PCR Test with Emergency Use Authorization at a Large Academic Safety Net Hospital. *Med* **1**, 152-157.e3 (2020).

[13] Seyfried, G., Pei, L. & Schmidt, M. European do-it-yourself (DIY) biology: Beyond the hope, hype and horror. *BioEssays* **36**, 548–551 (2014).

[14] Meyer, M. Domesticating and democratizing science: A geography of do-it-yourself biology. *Journal of Material Culture* **18**, 117–134 (2013).

[15] Landrain, T., Meyer, M., Perez, A. M. & Sussan, R. Do-it-yourself biology: challenges and promises for an open science and technology movement. *Syst Synth Biol* **7**, 115–126 (2013).

[16] Keulartz, J. & van den Belt, H. DIY-Bio – economic, epistemological and ethical implications and ambivalences. *Life Sci Soc Policy* **12**, 7 (2016).

[17] Hillson, N. *et al.* Building a global alliance of biofoundries. *Nat Commun* **10**, 2040 (2019).

[18] Building New Pathways. *Ginkgo Bioworks* https://www.ginkgobioworks.com/our-ecosystem/building-new-pathways/.

[19] Crea y desarrolla tu bioemprendimiento | The Bridge Biofoundry. *TheBridgeBiofoundry* https://www.tbbiofoundry.org.

[20] Thimiri Govindaraj, D. B. CSIR Synthetic Biology and Precision Medicine Centre Bio-foundry Program. (2022).

[21] Mao, N. *et al.* Future trends in synthetic biology in Asia. *Advanced Genetics* **2**, (2021).

[22] Winickoff, D. E., Kreiling, L., Borowiecki, M. & Garden, H. COLLABORATIVE PLATFORMS FOR EMERGING TECHNOLOGY: CREATING CONVERGENCE SPACES. 85.

[23] Maccani, G. *et al. Scaling up Citizen Science: What Are the Factors Associated with Increased Reach and How to Lever Them to Achieve Impact.* (2020).

[24] *Citizen Science: Innovation in Open Science, Society and Policy.* (UCL Press, 2018).

[25] Den Broeder, L., Devilee, J., Van Oers, H., Schuit, A. J. & Wagemakers, A. Citizen Science for public health. *Health Promot. Int.* daw086 (2016).

[26] MOOC: Open Science: Sharing Your Research with the World | TU Delft Online. *TU Delft Online Learning* https://online-learning.tudelft.nl/courses/open-science-sharing-your-research-with-the-world/.

[27] UZH - Citizen Science School. http://www.citizenscienceschool.uzh.ch/en.html.

[28] Developer. Community Science Field Trips (Fall 2022). *Gowanus Canal Conservancy* https://gowanuscanalconservancy.org/communityscience/ (2018).

[29] Garenne, D. *et al.* Cell-free gene expression. *Nat Rev Methods Primers* **1**, 49 (2021).

[30] Khambhati, K. *et al.* Exploring the Potential of Cell-Free Protein Synthesis for Extending the Abilities of Biological Systems. *Front. Bioeng. Biotechnol.* **7**, 248 (2019).

[31] Sheahan, T. & Wieden, H.-J. Emerging regulatory challenges of next-generation synthetic biology. *Biochem. Cell Biol.* **99**, 766–771 (2021).

[32] Dondapati, S. K., Stech, M., Zemella, A. & Kubick, S. Cell-Free Protein Synthesis: A Promising Option for Future Drug Development. *BioDrugs* **34**, 327–348 (2020).

[33] Cho, J. Scaling COVID-19 Testing to Millions Per Day. *Ginkgo Bioworks* https://www.ginkgobioworks.com/2020/06/23/scaling-covid-19-testing-to-millions-per-day/ (2020).

[34] Crone, M. A. *et al.* A role for Biofoundries in rapid development and validation of automated SARS-CoV-2 clinical diagnostics. *Nat Commun* **11**, 4464 (2020).

[35] Jain, B., Bajaj, S. S., Noorulhuda, M. & Crews, R. D. Global health responsibilities in a Taliban-led Afghanistan. *Nat Med* (2021).

[36] The Microsetta Initiative – Researching Global Microbiomes. *The Microsetta Initiative* https://microsetta.ucsd.edu/.

[37] 100 For Parkinson's - Blog - The dataset from 100 For Parkinson's is now larger than 2.2 million data points! https://www.100forparkinsons.com/blog/research-dataset-from-100-for-parkinson-s-exceeds-2-2-million-data-points-to-power-new-insights.

[38] Seattle Flu Alliance. *Seattle Flu Alliance* https://seattleflu.org.

[39] Chu, H. Y. *et al.* Early Detection of Covid-19 through a Citywide Pandemic Surveillance Platform. *N Engl J Med* **383**, 185–187 (2020).

[40] Holowko, M. B., Frow, E. K., Reid, J. C., Rourke, M. & Vickers, C. E. Building a biofoundry. *Synthetic Biology* **6**, ysaa026 (2021).

[41] Adames, N. R., Gallegos, J. E., Hunt, S. Y., So, W. K. & Peccoud, J. Hands-On Introduction to Synthetic Biology for Security Professionals. *Trends in Biotechnology* **37**, 1143–1146 (2019).

[42] Kritharis, A., Tamer, I. M. & Yadav, V. G. Vaccine production and supply need a paradigm change. *Can J Chem Eng* **100**, 1670–1675 (2022).

[43] FHI, F. of H. I.-. Future of Humanity Institute. *The Future of Humanity Institute* http://www.fhi.ox.ac.uk/.

8

ENHANCING BIOSECURITY WITH LANGUAGE MODELS: DEFINING RESEARCH DIRECTIONS

This report explores the potential of large language models (LLMs) to enhance biosecurity. We conducted interviews with nine biosecurity experts to understand their daily tasks, and how LLMs could be more useful for their work. Our findings indicate that approximately 50% of our interviewees' biosecurity-related tasks, such as gathering information from papers and reports, reviewing safety forms, and writing memos and summaries, have high potential for automation with LLMs. Skills critical for biosecurity work, like processing information and communicating effectively, could also be augmented by LLMs. However, current LLMs have limitations, such as often providing shallow or incorrect information. We provide suggestions for LLM-based tools that could significantly advance biosecurity efforts and list field-specific datasets to facilitate their development.

This chapter is a result of independent research. The manuscript has been submitted to peer-reviewed journal and the preprint is available at: Chen, Michael and Holub, Martin and Tice, Cameron, Enhancing Biosecurity with Language Models: Defining Research Directions (March 25, 2024). Available at SSRN:http://dx.doi.org/10.2139/ssrn.4772574

8.1. INTRODUCTION

The COVID-19 pandemic brought with it a staunch reminder that not only individuals, but also societies at large, are subject to threats from natural pathogens. While it caught the majority of us by surprise, the occurrence of a global pandemic was, and continues to be, something we must learn to expect. In the last 500 years, there have been at least 15 epidemics with a death toll in excess of 1 million (that is one every 33 years on average), and historically, pandemics were able to nearly wipe out whole civilizations. Even if the risk of a global pandemic occurring in any given year is small, it is not zero or negligible, resulting in a substantial threat to the future of humanity in the long run. In fact, there are other contributors to the gravity of this risk. While in the past we may have had to deal with natural pathogens only, the technological developments in biological engineering we have seen over past decades create a tangible risk of malicious actors designing and building synthetic pathogens. AI has been seen as an enabling technology across a range of domains, and it is a matter of concern to which extent it can contribute to biological risk.

Recent empirical evaluations of large language models (LLMs) have found that AI assistance can improve human attempts to plan biological attacks, at least to some extent [1,2] although other reports have claimed no effect [3,4]. Even though responsible AI labs may release biologically capable AI models only when they have been trained to be safe (for example, generally refusing to assist with bioterrorism), skilled actors could overcome these guardrails through jailbreaks or other means [5,6]. Moreover, capable AI models released with public weights can have these safeguards entirely removed, eliminating the ability to prevent malicious use[7]. Several groups are actively researching these risks of AI-assisted bioterrorism including the use of LLMs, but model evaluations have limitations as a means for ensuring safety [8]. In contrast, some perspectives suggest that carefully guiding technological advancements could lead to a more secure world [9–11]. In this vein, the 2023 Executive Order on Artificial Intelligence calls for an assessment of "the ways in which AI applied to biology can be used to reduce biosecurity risks, including recommendations on opportunities to coordinate data and high-performance computing resources."[12] Despite this potential, there is limited research on how thoughtful development and scaffolding of AI models may be able to differentially advance biosecurity, without contributing to their pre-existing biological risk potential.

This report centers on LLMs as their usefulness in various applications, including language processing, data analysis, and decision-making, has increased markedly over the last year, creating a gap in understanding how this new technology could be leveraged to enhance biosecurity. We note that various special-purpose machine learning algorithms have been developed to make biosecurity-relevant predictions [13–15]; continued developments in classic machine learning for biosecurity will be valuable, although it is not the focus of our report. The use of LLMs and LLM agents in biotechnology has been perceived as offense-dominant, in which defensive improvements in biosecurity may not keep pace[7]. Given the disproportionate number of individuals working on biosecurity relative to bioterrorism, we hypothesized that assisting scientists with relevant LLM-based assistants could accelerate progress in developing safety measures that prevent

and tackle a range of biological threats, including pandemics, without assisting biorisk [16].

While a substantial portion of biological research that can be accelerated by LLMs has dual-use potential, for many roles in the biosecurity field, there need not be overlap with the type of work being done by potential bad actors. Many helpful interventions seem to be purely beneficial such as research on public health, far-UVC, and metagenomic sequencing, among others. However, the application of AI in these domains is under-explored and we are still far behind in being able to prevent or even respond to an engineered pandemic. Without strategic application and control, more advanced AIs make this risk increasingly likely. Therefore, there is a significant motivation for developers of AI to aid biosecurity researchers as soon as possible.

This report outlines our findings and progress from five weeks of research and interviews. Here, we attempt to define the specific roles LLMs could play in supporting biosecurity researchers to tip the scales toward defense, laying the groundwork for future research. Our interviews with biosecurity experts (Table 1 and Materials and Methods) highlighted current limitations of AI in biosecurity, such as issues with data hallucinations and lack of domain-specific functionalities. This report aims to lay the conceptual groundwork for future development in this area.

8.2. MATERIALS AND METHODS

The initial phase of the project involved conducting interviews with nine external biosecurity experts in various industries (Table 8.1). These interviews were designed to gather insights into the daily tasks of professionals in the field and to understand how LLMs could be developed to become more useful in their work. The interviewees were asked questions about their biosecurity-relevant tasks and the expertise required to execute them, levels of LLM usage, and their outlook on the impact of LLMs on biosecurity (Table 8.2).

Number of interviewees per industry type	
Nonprofits and Charitable Organizations	5
Academic and Educational Institutions	1
Corporate and Think Tank Entities	2
Governmental and Regulatory Bodies	1
Number of interviewees in a leadership position	3

Table 8.1: Overview of the dataset (n = 9)

- What is your role at your current institution?
- What are biosecurity-relevant tasks that you regularly carry out in your work?
- What expertise is needed to execute these tasks?
- Do you use AI LLM services (e.g. ChatGPT) in your work?
  - Have you experienced any shortcomings?
  - Why not? Are there any perceived inadequacies or shortcomings?
- What do you think that an AI trained to be the most helpful to you and other biosecurity researchers should do very well? Why would this be helpful?
  - What do you think would be especially hard for it to do safely?
- What is your outlook on how LLMs will affect biosecurity? Positive or negative?

Table 8.2: Interview questions

8.3. FINDINGS

A) Gathering and analyzing information:

- Reading papers and reports*
- Safety form review*
- Monitoring news and current events*
- Interviewing stakeholders (e.g. policymakers)

B) Synthesizing information: (in audience-specific style)

- Writing summaries and reports* (internal and external)
- Writing op-eds, memos, blog-posts*
- Setting scientific priorities

C) Communication:

- Messaging and social media*
- E-mails*
- Networking and forming alliances
- Leadership (communicating goals and purpose)

D) Operations:

- Calling people
- Website content update
- Meeting preparation* (agenda, structure, logistics)
- Meetings (one on one, teams)

Table 8.3: Biosecurity-related tasks. * - Tasks with high potential for LLM impact.

8.3.1. BIOSECURITY-RELATED TASKS HAVE HIGH POTENTIAL FOR AUTOMATION

We surveyed a range of biosecurity professionals to find out the tasks they routinely carry out (Table 8.3). Broadly, the tasks can be categorized into "gathering and analyzing information", "synthesizing information", "communication", and "operations". We estimate that 50% of these tasks have a high potential to be impacted by LLMs (denoted as * in Table 8.3). For a more rigorous analysis, future research can comprehensively list biosecurity tasks and their exposure to LLM automation [17,18].

8.3.2. INTERPERSONAL SKILLS AND INFORMAL RULES ARE CRITICAL FOR POLICY-MAKING

In order to find out what skills are specifically important for biosecurity-related work, we asked our interviewees what skills they improved the most since entering the field. Remarkably, a large portion of the critical skills are interpersonal and rely on an understanding of subtle cues and non-codified knowledge (Table 8.4). However, a number of skills have a high potential to be augmented by increased adoption of LLMs (denoted with * in Table 8.4).

8.3.3. CURRENT LEVELS OF USE OF LLMS AND BARRIERS TO THEIR ADOPTION

Clearly, LLMs have the potential to benefit a number of tasks that are carried out in biosecurity-related professions and augment some biosecurity-relevant skills. However, LLMs are recent technological developments (OpenAI's ChatGPT and Google's Bard/Gemini were released 14 and 9 months ago, as of the time of writing, respectively) and far from having reached maturity. We thus wondered what is their current use among biosecurity professionals. The majority (7/9) of interviewees do use LLMs in their work, and a third do so frequently (Fig. 8.1)

Interviewees listed experiencing a range of shortcomings using the current versions of LLMs in their work (Table 8.5). Most notably, LLMs are seen as providing low-insight information which contains inaccuracies and is poorly referenced. LLMs are unable to correctly intuit the relative importance of a range of parameters and stakeholders' views that influence the information in its inputs (either user-provided or in the data it has been trained on). This is a particularly important consideration in the area of policy-making, where unwritten rules, informal relationships, and diverging interests are important and common. Finally, and somewhat surprisingly, current versions of LLMs also struggle to fully obey user's instructions, although this is likely to improve as the technology develops.

- Processing information (high volume and density)*
- Assimilating new information (within and outside of one's field of expertise)*
- Communication: Writing (range of forms and audiences)*
- Communication: Interviewing (listening, asking follow-up questions, ...)
- Navigating environments with different perspectives and political views
- Networking and forming alliances
- Leadership: Setting scientific and organizational priorities
- Leadership: Conveying explainable mission and impact
- Procedural knowledge of policy-making (codified)*
- Procedural knowledge of policy-making (informal)

Table 8.4: Skills important for biosecurity. * - Skills with high potential benefit from LLM augmentation.

- Poor trustworthiness (hallucinations, excessive creativity)
- Missing and incorrect referencing
- Insufficient information depth and lack of insight
- Inability to correctly intuit relative weights of input and external parameters
- Inability to fully follow instructions (disregarding some of the input, not respecting boundaries)

Table 8.5: Perceived shortcomings of LLMs

8

8.3.4. DEVELOPMENT OF NEW LLM-BASED TOOLS CAN ADVANCE BIOSECURITY

The adoption of LLMs is poised to transform a number of work-related tasks and increase worker productivity across a wide range of occupations. Knowledge workers in particular are more likely to see a larger portion of their work-related tasks exposed to the effects of LLMs [18]. As an example, users adopting GitHub Copilot for programming report less time spent on coding tasks, increased productivity, improved code understanding, and a greater sense of satisfaction [19]. Among our interviewees, the majority report using LLM tools in their work (Fig. 8.1), despite their perceived shortcomings (Table 8.5).

Broadly speaking, adapting LLMs to specific domains can improve their usefulness. When LLMs can read relevant information before responding to user queries (retrieval-augmented generation), they can provide more factual answers that avoid hallucination [20]. Pretraining models on code is well-known to be necessary for coding performance [21], and likewise for math [22]. Fine-tuning language models to follow user instructions

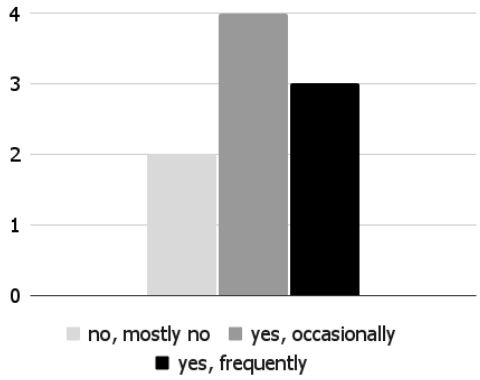

Figure 8.1: Current levels of LLM adoption among surveyed biosecurity professionals. Prompt: Do you use LLM services in your work?

makes them substantially easier to use [23], but available chatbots are fine-tuned with human preference data that is not adapted for biosecurity.

It was our goal to better understand the needs of biosecurity professionals and arrive at concrete suggestions for future development of LLM-enabled tooling that would be the most useful in their work. To do so, we directly prompted the interviewees for their needs and wishes, as well as conducted independent research and ideation. This led us to identify a range of directions for future development including a DURC potential evaluation tool, AI lab safety officer, and chatbots to aid generating and proposing implementable policies (the top five are summarized in Table 8.6). These LLM applications could be conceivably prototyped through a combination of high-quality prompting, retrieval-augmented generation, and excellent user interface design.

8.3.5. BIOSECURITY-SPECIFIC DATASETS TO AID THE DEVELOPMENT OF NEW LLM-TOOLS

As our research highlights, biosecurity is a niche with specific user profiles and requirements that determine how the users interact with and benefit from LLMs. A large part of the perceived LLM shortcomings (Table 8.5) reflect the tools' lack of appreciation for the specifics and subtleties of research and policy-making. While we have suggested a few examples of LLM-based tools to help with biosecurity (Table 8.6), we also realize that any LLM is only as good as the data it has been trained on. In view of this, we created a list of 75+ publicly available datasets (`https://github.com/martinholub/awesome-biosecurity-datasets`). This repository also includes a community wish list for datasets and resources that do not yet exist.

- General-purpose research assistant for biosecurity
  - Integrated with a search tool to retrieve relevant biosecurity articles and papers, perhaps similar to Consensus (`https://consensus.app/`) but with better curation of sources
  - Suggest diverse stakeholders that may be affected and relative importance/impact
  - Suggest and help defuse counter-arguments
- Chatbot for assisting biosecurity policy
  - Attempt to generate implementable policy ideas
  - List stakeholders, open calls, related funding, and so on
  - Given this and user-provided input data, generate a draft
  - Train on previous successful and failed policy proposals
- Dual Use Research of Concern (DURC) potential evaluation tool
  - Red-team research proposals and, if applicable, suggest ways that they could be made safer from a biosecurity perspective
  - Grant applications could be a natural place to integrate this tool, as many biology researchers may not seriously think through dual-use concerns, though it is important that the tool does not inspire bioterrorism
- Biosecurity text style adjustment tool
  - Reflect the role, status, and political views of the target audience
  - Adjust style to appeal to a target audience or medium (such as report, memo, blog post, e-mail)
- Virtual lab safety officer
  - Review experimental plans and lab books based on safety protocols
  - Review video footage to highlight potential unsafe practices (assuming vision-language model)

Table 8.6: Examples of LLM-based tools that could be helpful to biosecurity.

8.4. CONCLUSION

As a dual-use technology, advanced LLMs could have the potential to exacerbate biological risk by aiding bioterrorists, especially given concerns around the rapid advancement in general LLM capabilities. To help mitigate this risk, it is valuable to explore the development of biosecurity-focused LLM assistants that can accelerate biosecurity research. In our exploratory interviews with biosecurity professionals, we have found that the majority use LLMs like ChatGPT, at least on occasion, but to a limited extent, in their work.

Current LLM tools in use have a number of shortcomings, such as poor trustworthiness and lack of insight, which mean that they have a limited impact on the interviewees' workflows. We suggest various LLM-based assistants that could be developed for biosecurity tasks. These could be created as AI agents with access to custom biosecurity tools, or as specially fine-tuned models, for example. We also contribute a list of datasets that could be integrated with AI for biosecurity purposes.

Our report is an initial foray into investigating how LLM-based technologies could be leveraged to improve biosecurity. It is worth noting that improvements in general AI capabilities can improve their usefulness for biosecurity, but can also increase their potential to assist with bioterrorism. We emphasize efforts that differentially improve biosecurity, especially ones that are neglected by standard market forces. We propose several topics for future research:

- **Comprehensively outlining biosecurity tasks**: Biosecurity is a broad and interdisciplinary field, and our selection of experts interviewed does not cover all promising areas of biosecurity. Similar to the O*NET database (`https://www.onetonline.org/`) [17], it would be valuable to create a thorough inventory of tasks involved in biosecurity and evaluate their susceptibility to automation by LLMs or LLM agents. After prioritizing these tasks by what is most impactful for reducing catastrophic biological risk, this could serve as a basis for developing AI biosecurity assistants.

- **Prototyping and evaluating AI biosecurity assistants**: One direction for future research is to develop AI assistants that have improved performance for biosecurity – for example, through tools that allow them to retrieve relevant information from trusted sources, or fine-tuning data that trains them to produce higher-quality responses. Similar to how OpenAI recruits expert AI trainers to improve an LLM's software engineering capabilities, AI labs could have dedicated efforts to curate data for biosecurity performance. Iterative development in response to user feedback and ongoing model evaluations are important to ensure that the AI assistants are actually useful for biosecurity.

- **Forecasting how AI advancements aid bioterrorism vs biosecurity**: As our interviewed experts leaned towards believing that AI advancements would generally increase biorisk rather than decrease it, it is valuable to have advance understanding of the most likely ways they would do so, and which developments in biosecurity could alleviate this.

- **Clarifying safe and unsafe biological capabilities for AI**: Some types of expertise (e.g., in virology) may have applicability with biosecurity but also could aid with bioterrorism. While recommending the development of AI biosecurity assistants, we do not recommend generally accelerating AI capabilities, especially with respect to expertise that presents dual-use national security risks. Relatedly, we recommend the creation of safety standards for AI models that could increase biological risk [9].

- **Field building for innovation in AI for biosecurity**: In the realm of cybersecurity, the White House has partnered with major AI companies to launch a two-year competition for AI to improve the software security of critical infrastructure [24]. We believe that biosecurity could benefit from analogous efforts to discover ways that AI could help mitigate biological threats.

As foundation models have a growing impact on society, it is essential to steer their development to ensure the potential to increase biological risk is minimized. To complement ongoing evaluations of potential LLM-assisted biorisk and research into safety mitigations, we recommend developing LLM assistants specialized for biosecurity tasks. If this topic is adequately studied, a society with advanced AI capabilities could be prepared for biological threats through two means: by preventing the release of AIs with unsafe biological capabilities, as well as through the deployment of AIs continually working to improve biosecurity.

8.5. REFERENCES

1. OpenAI. Building an early warning system for LLM-aided biological threat creation. https://openai.com/research/building-an-early-warning-system-for-llm-aided-biological-threat-creation (2024).

2. Soice, E. H., Rocha, R., Cordova, K., Specter, M. & Esvelt, K. M. Can large language models democratize access to dual-use biotechnology? *arXiv* (2023) doi:10.48550/arXiv.2306.03809.

3. Mouton, C. A., Lucas, C. & Guest, E. *The Operational Risks of AI in Large-Scale Biological Attacks: Results of a Red-Team Study*. https://www.rand.org/pubs/research_reports/RRA2977-2.html (2024).

4. National Security Commission on Biotechnology. *AIxBio White Paper 4: Policy Options for AIxBio*. https://www.biotech.senate.gov/press-releases/aixbio-white-paper-4-policy-options-for-aixbio/ (2024).

5. Mazeika, M., Phan, L., Yin, X., & others. HarmBench: a standardized evaluation framework for automated red teaming and robust refusal. *arXiv* (2024).

6. Qi, X., Zeng, Y., Xie, T., & others. Fine-tuning aligned language models compromises safety, even when users do not intend to! *arXiv* (2023) doi:10.48550/arXiv.2310.03693.

7. Gopal, A., Helm-Burger, N., Justen, L., & others. Will releasing the weights of future large language models grant widespread access to pandemic agents? *arXiv* (2023) doi:10.48550/arXiv.2310.18233.

8. UK Department for Science, Innovation & Technology. *Frontier AI: Capabilities and Risks - Discussion Paper*. https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper (2024).

9. Sandbrink, J., Hobbs, H., Swett, J., Dafoe, A. & Sandberg, A. Differential technology development: An innovation governance consideration for navigating technology risks. *SSRN* (2022) doi:10.2139/ssrn.4213670.

10. Hendrycks, D., Carlini, N., Schulman, J. & Steinhardt, J. Unsolved problems in ML safety. *arXiv* (2021) doi:10.48550/arXiv.2109.13916.

11. Buterin, V. d/acc: Defensive (or decentralization, or differential) acceleration. https://vitalik.eth.limo/general/2023/11/27/techno_optimism.html#dacc (2023).

12. House, T. W. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. *The White House* https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/ (2023).

13. Alley, E. C., Turpin, M., Liu, A. B., & others. A machine learning toolkit for genetic engineering attribution to facilitate biosecurity. *Nature Communications* **11**, 6293 (2020).

14. Syrowatka, A., Kuznetsova, M., Alsubai, A., & others. Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases. *NPJ Digital Medicine* **4**, 1–14 (2021).

15. Sykes, A. L., Silva, G. S., Holtkamp, D. J., & others. Interpretable machine learning applied to on-farm biosecurity and porcine reproductive and respiratory syndrome virus. *Transboundary and Emerging Diseases* **69**, e916–e930 (2022).

16. Shavit, Y., Agarwal, S. & Brundage, M. *Practices for Governing Agentic AI Systems*. https://openai.com/research/practices-for-governing-agentic-ai-systems (2023).

17. O*NET OnLine. https://www.onetonline.org/ (2024).

18. Eloundou, T., Manning, S., Mishkin, P. & Rock, D. GPTs are GPTs: An early look at the labor market impact potential of large language models. *arXiv* (2023) doi:10.48550/arXiv.2303.10130.

19. Dohmke, T., Iansiti, M. & Richards, G. Sea change in software development: Economic and productivity analysis of the AI-powered developer lifecycle. *arXiv* (2023) doi:10.48550/arXiv.2306.15033.

20. Lála, J. *et al.* PaperQA: Retrieval-augmented generative agent for scientific research. *arXiv* (2023) doi:10.48550/arXiv.2312.07559.

21. Rozière, B., Gehring, J., Gloeckle, F., & others. Code llama: Open foundation models for code. *arXiv* (2023) doi:10.48550/arXiv.2308.12950.

22. Shao, Z., Wang, P., Zhu, Q., & others. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv* (2024) doi:10.48550/arXiv.2402.03300.

23. Ouyang, L., Wu, J., Jiang, X., & others. Training language models to follow instructions with human feedback. *arXiv* (2022) doi:10.48550/arXiv.2203.02155.

24. The White House. Biden-harris administration launches artificial intelligence cyber challenge to protect america's critical software. https://www.whitehouse.gov/briefing-room/statements-releases/2023/08/09/biden-harris-administration-launches-artificial-intelligence-cyber-challenge-to-protect-americas-critical-software/ (2023).

ACKNOWLEDGEMENTS

There are too many who impacted me as a doctoral candidate. I will still try to mention as many of you, keeping it brief. If I forgot you, shame on me—I will get you a drink.

Thank you to Cees, for creating a welcoming and friendly group, and for maintaining a productive working relationship; to Patrick, for treating me as a member of his group from day one; to Martin, Maxime and Dita for help with mass spectrometry; to Ethan, Cam, and Michael for building ad-hoc research teams together; to the iGEM engineering committee, particularly Jake, Ian, Paul, and Alejandro, for joining forces in making engineering biology easier and more accessible; to Tessa, Janvi, and Gurpreet for making biosecurity research more fun.

Thank you to all members of the CD lab. Thanks to Anthony for teaching me about supervision (and reviewing my Samenvatting); to Alex, without whom we would not have been able to make the devices shown in this thesis; to Nicola, for stepping in to give some supervision; to my office mates: Henry, for laying out the university bureaucracy; Roman, for setting high bar for productivity; Bis, for teaching me more about microscopy; and Justas, who helps me stay excited about science. Thank you to Eva, for her social skills; to JK, for accurately labeling some of our work as "high risk, low gain"; to Ian, for showing me that American friendliness needn't be just surface-level; to Sandro, for staying in touch despite the busyness of founding and fathering; to Rafa, who should quadruple himself because this group needs more people with your qualities (both technical and personal); to Sonja and Eugene, for being scientist role models; to Ale, for keeping it cool; to Mitasha, for missing Switzerland together; to Oskar, who introduced me to Danish music; to Wayne, for sharing many bits of life wisdom, despite me clearly being a poor receptacle; to Pinyaho, who became friend even in this short time; to Brian, for good vibes and fun trivia; to Ash, for the live music; to Paola, for being a fabulous host; to Richard, who made the department feel more like home; to Nils, for proactively pursuing experiments together; to Alejandro, for ad-hoc beach trips; to Allard, for AFM wizardry; to Eli, for exchanging numerous paper and music recommendations; to Jacob, for constructive and helpful feedback; to Jaco, for all-around help with molecular biology; to Tisma, for getting me to go to the gym; to Alberto and Fede, who gave me hope that I, too, could get through this; to Sabrina, for encouragement and helping me stay grounded; to Chenyu and Xin, who I am sure will be great supervisors; to Shuo, for good chats about science and being open to convincing that Netherlands is, after all, a nice country; to Anders and Angel, for their sense of humor; to Tor, for being a non-stereotypically friendly Swede; to Xiuqi, for discussions on the US; Charu for kindness; to Bert, for actually putting our syncell lab in order; and to Michel, for showing me what strong willpower looks like.

Thanks to everyone who made my time in BN special. The students I had the pleasure of working with and learning from, including Rafael, Marco, and Leander—good luck with your careers, I am confident you will do well. Thank you to the iGEM '20 & '21 teams and supervisors. It was fun, made me enjoy my time at BN much more, and gave me opportunities to learn. I was inspired by the quality of your work! Special thanks to Elisa, for convincing me to sign up, Essie for keeping the wheels turning, and to Christophe, for being a truly exceptional supervisor and mentor. Thanks to Daniel, for fittingly comparing the Netherlands to a mediocre relationship; to Tanja, for hosting; to Sophie, for making my time in BN more interesting; and to Michal, for career advice and hair-coloring skills. Thanks to Amber, Noa, Jeanette, and Myrna, for reminding me that the best students distinguish themselves by their work, not credentials; to Vadim and Lori for sharing the journey from their respective institutes.

Thank you to everyone at MIB who made my stay in the North great. Thank you, Kewin and James, for setting the bar for lab productivity high; Ravi, for the help with PFGE; and Daniel, for fantastic documentation. Thank you to Leandro, from whom I would be learning things even if I had stayed there for three years; to Ray, Smruti, and Ellie, for helping me to settle. Foremost, thank you to Stefan for being instrumental in making this visit possible! Thanks to the MIB PhD student gang too. :)

Outside of work, many people had an impact on me during my PhD years. Thank you to my adoptive families. First, all my fellow Forcies— with a special thanks to Francesca, whose life seems to be chaos and who still manages to check off all the boxes; and to Martin, for being a cool prof and great friend. Thank you to Jasmine, Annemirijn, Bram, Dilge, Cheron, Nick, Farkas, Jop, Jasper, Sara, Rosa, TQ, Filip, Karen, Paula, Niké, Camillo, Reuben, Vicky, Melissa, Isa, Markus, Dommi, Lennart, Berni, Dave, Bert, Elize, Friso, Patrick, Ravi, Klaas, Lorenzo, Lucas, Andrew, Meg, Valentina, Leo, Pietro, and others for the great times on and off the field. Thank you to my housemates, as well as all the special people I could meet and befriend while skating and doing cross-fit. Thanks to Bea, Akhila, Ebele, and Samuel for the great culinary experiences and chats in Manchester and beyond.

Thank you so much, Elena and Samuel, for the hard work you put into bringing Nucleate to the Netherlands, and the way we balanced this with networking across biotech hubs and hiking in California. I am truly grateful we got to do this together—it has been one of the most useful learning experiences I've had so far. Thank you to everyone on our team - I am grateful to have met you and have worked together!

Finally, thank you to my family. To my aunt, uncle, cousins, and grandfather, for always creating a warm welcome whenever I briefly turned up back in our hometown. Thank you to my sister for becoming a champion for my happiness. My final and biggest thanks goes to my mum—without whose dedication, hard work, and love, none of this would have been possible. Finishing a PhD pales in comparison to you building a family despite the circumstances!

CURRICULUM VITÆ

Martin HOLUB

EDUCATION

12/2019–12/2024	PhD. in Biophysics Delft University of Technology, the Netherlands Faculty of Applied Sciences & Kavli Institue of Nanoscience *Thesis:* Single-Chromosome Biophysics *Promotor:* Prof.dr. C. Dekker
09/2016–09/2019	M. Sc. Mechanical Engineering, Bioengineering Eidgenössische Technische Hochschule Zürich, Switzerland
09/2013–08/2016	B. Sc. in Mechanical Engineering Brno University of Technology, Czech Republic

EXPERIENCE

12/2023–12/2024	Co-Managing Director Nucleate Netherlands
06/2023–09/2023	EMBO Scientific Exchange, Synthetic Genomics University of Manchester, United Kingdom Manchester Institute of Biotechnology, Patrick Cai lab
03/2018–10/2018	Intern in Software Development and Bioinformatics NEBION AG, Zürich, Switzerland
11/2016–02/2018	Research Assistant in Image Analysis University of Zürich, Switzerland Institute of Pharmacology and Toxicology Laboratory for Experimental Imaging and Neuroenergetics

LIST OF PUBLICATIONS

10. Joesaar, A.; **Holub, M.**; Lutze, L.; Emanuele, M.; Kerssemakers, J.; Pabst, M.; Dekker, C. A Microfluidic Platform for Extraction and Analysis of Bacterial Genomic DNA *bioRxiv* 2024. https://doi.org/10.1101/2024.10.17.618837

9. George, I.; Ross, P.; Yang, Y.; **Holub, M.**; Rajpurohit, N.; Aldulijan, I.; Beal, J.; Vignoni, A.; Mishler, D. An Integrated Engineering Worldview of Synthetic Biology Education through the Lens of Webinar Based Pedagogy. *Front. Bioeng. Biotechnol.* 2024, *12*, 1431374. https://doi.org/10.3389/fbioe.2024.1431374.

8. Chen, M.; **Holub, M**; Tice, C. Enhancing Biosecurity with Language Models: Defining Research Directions (March 25, 2024). Available at SSRN: http://dx.doi.org/10.2139/ssrn.4772574

7. **Holub, M.**; Agena, E. Biofoundries and Citizen Science Can Accelerate Disease Surveillance and Environmental Monitoring. *Front. Bioeng. Biotechnol.* 2023, *10*, 1110376. https://doi.org/10.3389/fbioe.2022.1110376.

6. Aldulijan, I.; Beal, J.; Billerbeck, S.; Bouffard, J.; Chambonnier, G.; Ntelkis, N.; Guerreiro, I.; **Holub, M.**; Ross, P.; Selvarajah, V.; Sprent, N.; Vidal, G.; Vignoni, A. Functional Synthetic Biology. *Synthetic Biology* 2023, *8* (1), ysad006. https://doi.org/10.1093/synbio/ysad006.

5. **Holub, M.**; Birnie, A.; Japaridze, A.; der Torre, J. van; Ridder, M. den; de Ram, C.; Pabst, M.; Dekker, C. Extracting and Characterizing Protein-Free Megabase-Pair DNA for in Vitro Experiments. *Cell Reports Methods* 2022, 100366. https://doi.org/10.1016/j.crmeth.2022.100366.

4. **Holub, M.**; Birnie, A.; Japaridze, A.; der Torre, J. van; Ridder, M. den; de Ram, C.; Pabst, M.; Dekker, C. Extracting and Characterizing Protein-Free Megabase-Pair DNA for in Vitro Experiments. *Cell Reports Methods* 2022, 100366. https://doi.org/10.1016/j.crmeth.2022.100366.

3. Fisch, P.; **Holub, M**.; Zenobi-Wong, M. Improved Accuracy and Precision of Bioprinting through Progressive Cavity Pump-Controlled Extrusion. *Biofabrication* 2021, *13* (1), 015012. https://doi.org/10.1088/1758-5090/abc39b.

2. **Holub, M.**; Adobes-Vidal, M.; Frutiger, A.; Gschwend, P. M.; Pratsinis, S. E.; Momotenko, D. Single-Nanoparticle Thermometry with a Nanopipette. *ACS Nano* 2020, *14* (6), 7358–7369. https://doi.org/10.1021/acsnano.0c02798.

1. Barrett, M. J. P.; Ferrari, K. D.; Stobart, J. L.; **Holub, M.**; Weber, B. CHIPS: An Extensible Toolbox for Cellular and Hemodynamic Two-Photon Image Analysis. *Neuroinformatics* 2018, *16* (1), 145–147. https://doi.org/10.1007/s12021-017-9344-y.