Nano-Science Student Journal

Vol. 2. 2018
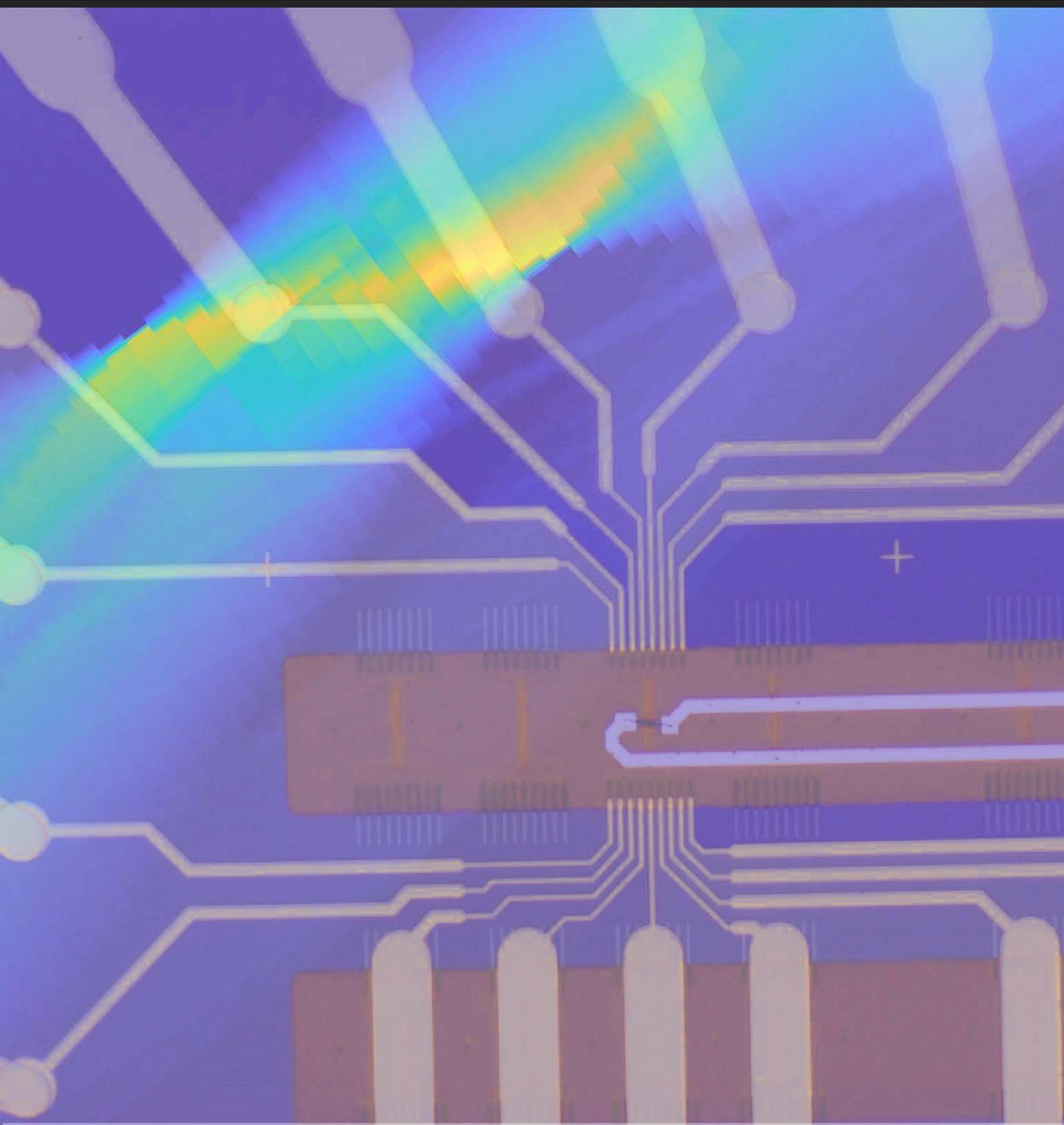
UCPH Nanoscience

a student research journal

About the journal
UCPH Nanoscience is not a peer reviewed journal, it is a journal that published opinion pieces and news from the Nano-Science Center and serves as a repository for student research performed as part of the nanoscience degree programme at the University of Copenhagen.

Editorial

Welcome to the third issue of UCPH NanoScience, the student journal of the Nano-Science Center at the University of Copenhagen and the journal of all Nanoscience students. This issue contains a selection of the bachelor reports handed in as part of obtaining a bachelor's degree in Nanoscience. All contributors have obtained their bachelor's degree in 2018.

All beginnings are difficult, and apparently starting a student journal in the framework of a University administration does not help. We are still struggling with getting the copyright transfer or thesis forms filled out and send our way, even though it is a natural part of handing in a bachelor's thesis in nanoscience. We hope to have better processes in place before the class of 2019 hands in the fruits of their labour before the summer holiday.

UCPH Nanoscience is an online archive of bachelor thesis and other reports from the nanoscience students at the University of Copenhagen. If you are reading a physical copy, you will find that the actual reports are not included, those you can find online. If you are reading this on a screen, you will find this third issue of the nanoscience student journal and all the previous issues a few clicks back.

Thomas Just Sørensen
Editor-in-chief

Competition

UCPH Nanoscience need a logo
Can you do better? The journal needs a logo to replace the temporary logo designed by the editor in chief. All suggestions are welcome, but you are encouraged to design a logo that reflects spirit of the Copenhagen flavour of nanoscience and the nanoscience student body. The prize is two vintage bottles of Verdi sparkling vine.

Suggestions should be sent to the editor-in-chief by December 1st 2019 at tjs@chem.ku.dk.

People Wanted!

UCPH Nanoscience is looking for staff
We are looking for 4 junior editors/reporters determined to shape the future of the student journal. The junior editors will be responsible for running the journal and provide student relevant content for the News & Views section.

Table of contents

Thesis in this issue

The following students have graduated with a BA in nanoscience in 2018. Their theses have been included in this issue of UCPH Nanoscience.

Thesis marked by an asterix (*) are under embargo and will be published in a later issue of UCPH Nanoscience.

Chemical Fabrication and Characterization of Graphene for Water-Based Inks used in Screen-Printing by Mathias K. Laursen

Mathias K. Laursen has conducted his bachelor's project in the NanoChemistry group. The group focuses on developing nano-materials, such as graphene and organic dyes, and employing these in devices for e.g. molecular electronics. The focus for the project was on graphene and how to utilize the properties of graphene in composite materials. During the project Mathias fabricated a conductive ink for use in screen printing applications.

Mathias did the first two years of his bachelor's studying in Aarhus and was transferred to Nanoscience in Copenhagen for his final year.

Supervisor: Associate Professor Kasper Nørgaard

Thesis abstract: Graphene exhibits a range of interesting properties such as high optical transmittance, good thermal conductivity, high Young's modulus, and good electrical conductivity. Graphene can serve as a component in conductive applications and potentially replace the expensive conductive materials, such as indium tin oxide (ITO), that are used today. In this article the graphene derivatives graphene oxide (GO) and reduced graphene oxide (rGO) are synthesized. The reduction of GO is done by either microwave assisted reduction, intense light reduction or chemical reduction with hydrogen iodide. The structures of the synthesized compounds are characterized and analyzed using X-ray photoelectron spectroscopy (XPS) and Raman spectroscopy. Based on the synthesized GO, a GO-ink is formulated and screen-printed on a polymer-based polyethylene terephthalate (PET) substrate, from which a sheet resistance of 0.4 kΩ s$-$q1 is measured. This work suggests that the synthesized GO can be used in production of conductive circuits or serve as a precursor for future scalable conductive films.

Characterization of the formation of magic-sized {Bi$_{38}$O$_{45}$} clusters by PDF and SAXS by Andy Anker

Andy Anker has always been a curious outdoor person, and he loves nature, which is plentiful and stunning at the small island, Bornholm, where he is born. Therefore, some of his holidays are dedicated to live in the wild.

Andy's fascination of nature has been the driving force to understand science and especially how materials are built by design. Therefore, Andy has worked with material chemistry. Andy's interest is in techniques as Total Scattering with Pair Distribution Function (PDF) and Small-Angle X-ray Scattering (SAXS), which is essential in order to study nanoparticles and structures in solution. Andy has applied advanced computer modelling to combine information of both the local order from PDF and the particle order from SAXS, which overcome problems that the methods cannot overcome themselves.

In the group there are 8 people, 2 bachelors-, 1 master-, 3 Ph.D. students and supervisor Kirsten Jensen. All young people, which gives a unique atmosphere where the members both work hard and learn from each other, but also do a range of activities outside the study together. **Supervisor** Assistant Professor Kirsten Marie Ørnsbjerg Jensen

Thesis abstract:

Bismuth oxido clusters exist in a range of sizes, all built up by octahedral {Bi6O8} units. While the atomic structure of various clusters has been solved by single crystal diffraction, it is much more challenging to study clusters directly in solution. Here, we use *in situ* X-ray total scattering with Pair Distribution Function (PDF) analysis to study the formation of the {Bi38O45}a cluster from [Bi6O5(OH)3(NO3)5]·(H2O)3 crystals dissolved in DMSO. The implementation of PDF analysis provides a unique insight into the structural rearrangements on the atomic scale. By combining with Small Angle X-ray Scattering (SAXS) we can furthermore investigate the size, morphology and size

10 dispersion of the clusters taking place in the process. Consequently, the combination of these two complementary techniques provides a mean of bridging the local atomic and macroscopic characteristics of the material.

In the presented study, the results obtained show that the reaction goes through several stable intermediates before the magic-sized product {Bi38O45} is reached. Through an associated temperature study, the intermediate was furthermore found to be capable of stabilization of up to days by varying the resulting reaction rate. The present studies show how powerful a tool the Debye Equation is in combination with 15 SAXS and PDF, which in this field is a new development that has a large potential for unravelling important questions in nanochemistry in solution.

Fabrication and measurement of hybrid quantum dot devices featuring Yu-Shiba-Rusinov sub-gap states by Jacob Ovesen
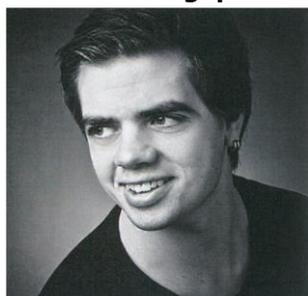
Jacob Ovesen studies electron transport and behaviour in hybrid superconductor-quantum dot devices based on nanotubes.
Supervisor: Professor Jesper Nygård

Thesis abstract: The aim of this bachelor project is to fabricate and study hybrid quantum dot (QD) devices featuring Yu-Shiba-Rusinov (YSR) sub-gap states, made on special InAs nanowires with 7 nm epitaxial aluminum coating on 3 facets. The advantage with these nanowires is the epitaxial aluminum, which can be etched away in a small window of semiconducting nanowire for creating QDs. These dots can be used for measuring the sub-gap states in superconducting materials due to the high degree of tunability. A device with a single QD featuring YSR sub-gap states was fabricated and analyzed, and a double QD system was afterwards made from the device using superconducting leads in the normal state, but not yet featuring YSR sub-gap states. The device does show the possibility for making an superconductor double dot-superconductor device featuring YSR sub-gap states.

Structural dynamics of Cpf1 by FRET by Freja Jacobsen Bohr

Freja decided to study nanoscience because she thinks that combining physics, chemistry and biology will give a great advantage in understanding scientific problems. Freja's interest primarily lies in using physics and chemistry to understand the mechanism underlying enzymatic structure and function relation. In her project Freja combines advanced microscopy with Förster Resonance Energy Transfer to understand the mechanism of the CRISPR/Cpf1 system. Freja is a part of the Hatzakis group located at Frederiksberg Campus. The main scope of the group is to decipher biomolecular recognition and decode how structural dynamics encode proteins capacity to accelerate chemical transformations while maintaining the functional plasticity of accepting structurally diverse substrates. The group approaches this formidable challenge by an eclectic mix of techniques - borrowed from whichever area of experimental science that promises to shed light on protein behaviour with an emphasis on microscopy techniques at the single molecule level. **Supervisor**: Associate Professor Nikos Hatzakis **Thesis abstract:** The RNA-guided endonuclease of class V Cpf1 *(Clustered Regularly Interspaced Short Palindromic Repeats* from *Prevotella and Francisella)* is a central element in prokaryotic immune mechanism, which use a CRISPR-RNA (crRNA) to locate and cleave viral DNA. Cpf1 allow genome-editing at a specific position specified by synthetic crRNA, hence a promising therapeutic agent to treat gene deficiencies. Our current understanding of Cpf1 structure and function primarily 10 relies on crystal structures and cryoEM data, providing unique and invariant structures. In contrast to recent methods, the scope of this project is to expand our understanding of the dynamic structure of Cpf1, and understand how conformational changes and catalysis are related. To do this I used FRET (Förster Resonance Energy Transfer). Here, the work on this projects is presented, with the aim of 1) understanding the conformational changes of Cpf1 in free-form, binary complex (crRNA-bound) and

tertiary complex (DNA-bound) in bulk, and correlating the conformational changes to intramolecular 15 distances, thus providing insight to the mechanism of DNA-cleavage and 2) calibrate the distance using dsDNA as a rigid scaffold, preparing for single molecule FRET measurements on Cpf1. Ensemble measurements revealed conformational changes of Cpf1 upon binding DNA, however the assay needs to be optimized further to extract distinct distances. Spectrometric experiments revealed that excess dye in solution was a general problem in ensemble measurements, interfering with the results. To validate and setup the calibration I did both ensemble and smFRET measurements on dual labelled dsDNA to address and eliminate the role of excess dye in solution. The inter-dye distance of dual-labelled dsDNA was determined to be 62.72 ± 0.93 Å. The simulated distance with Monte Carlo simulations was found to be 61.6 Å. This illustrates smFRET as a method to probe enzymatic motion, and thus could provide novel information about the mechanism of DNA-cleavage in Cpf1, paving the way for future genome-editing.

UCPH NanoScience, 2018, 2, 201805 5p
Effects of Hydrogenphosphate Quenching on Fluorescent Lifetime of Hydroxyphenyl Substituted Diazaoxatriangulenium by Stine G. Stenspil

Stine is studying nanoscience because it is an interesting field where connections are made across the various established disciplines to discover new phenomena and materials. In the Laursen group Stine worked on solving the little mystery of "disappearing" fluorescence and shortening of lifetimes in some solutions by doing various measurements of the photophysical properties of the phenol derivative of a triangulenium dye.

The Laursen group has its main focus on the design, synthesis and properties of fluorescent dyes. The majority of dyes used in this group are based on the triangulenium family as they are very rigid and therefore are excellent fluorescent emitters. Some of the synthetized dyes can be used as fluorescent probes and optical sensors, which is of high importance in biological research.

Supervisor: Professor Bo W Laursen
Thesis abstract: Fluorophores are used in many research applications such as cell staining and as probes for ions, DNA, proteins or membranes. By designing and synthesising the fluorophore to have specific photophysical properties it is possible to probe the surrounding environment of the fluorophore. For example, by utilising the property of photoinduced electron transfer (PET) which quenches the fluorescence of the fluorophore. A hydroxyphenyl substituted diazaoxatriangulenium (DAOTA+) designed to have the properties as a pH probe using photoinduced electron transfer showed another pH dependency in addition to the expected quenching. Results show that hydrogenphosphate is the reason for the additional quenching and other ions is also shown to have quenching abilities. Further understanding of this mechanism could result in a new triangulenium derivative that is able to detect hydrogenphosphate or other ions in solution. The detection of hydrogenphosphate is especially interesting as it is a key component in physiological solutions and cells..

UCPH NanoScience, 2018, 2, 201806 41p
From Breast Cancer to Computational Science by Catalina Cocan

Catalina is interested in the combination of clinical biology and scientific computing, which she believes must be further investigated. The group uses scientific computing methods to investigate the dynamics and control of molecular perturbations, solvent effects, chemical reactions in homogeneous and heterogeneous environments.

Supervisor: Professor Kurt Valentin Mikkelsen
Thesis abstract: Computational science focuses on the development of predictive computer models of the world around us. Scientific computing methods have become more accurate in the discovering of tumor treatment possibilities, than laboratory experiments have ever proven. This thesis will provide the mathematical modeling of breast tumor growth, including different aspects of how the immune system together with cycle-phase-specific chemotherapy can defeat the disease. The delay differential equations

representing the tumor environment are solved using the Runge Kutta fourth method, taking in consideration different phases of the cell-cycle.

"Turing mechanisms under pressure": Mesenchymal condensations could be involved in digit patterning of mice by Henrik Pinholt, Heura Cardona Blaya, James Harvey Swoger and James Sharpe

Paper abstract:
In recent years, some attention has been directed towards the possibility that mechanical forces could play a role in creating the expression patterns observed in vertebrate body parts. By compressing regions of cultured embryonic mesenchymal stem cells from mouse autopods, this study provides a novel way of testing effects of mechanochemical stimulation of the Turing mechanism responsible for digit patterning in mice. The study found important features to be included in these compression systems for 10 mesenchymal cells: Com- pressing a small area of the culture, using transparent pistons, and maintaining a stable environment from vibrations and disturbances. The results suggest that mechanical forces upregulates Sox9 in the boundary between compressed and uncompressed regions, but doesn't seem to change the compressed pattern in comparison to uncompressed cells. However incomplete, these results suggest that both chemical and mechanochemical regulation of Sox9 might be involved in patterning the 15 digits in mice.

Effects of environmental conditions on Thermomyces lanuginosus lipase behavior by single particle tracking on a native substrateby Jacob Kæstel-Hansen*

Jacob decided to study nanoscience because it combines physics, chemistry and biology. Jacob finds this interdisciplinary approach very interesting as it offers a new perspective on many scientific problems. Especially the combination of physics and chemistry to investigate enzyme structure and function relations, which is traditionally a biochemistry question. Specifically, this thesis seeks to deconvolute masked protein dynamics to understand distinct enzyme behavior, as well as the effect of short Debye length on TLL function on its native substrate using bulk- and especially single-particletracking

Jacob is a part of the Hatzakis group located at Frederiksberg Campus. The main scope of the group is to decipher biomolecular recognition and decode how structural dynamics encode proteins capacity to accelerate chemical transformations while maintaining the functional plasticity of accepting structurally diverse substrates. We approach this formidable challenge by an eclectic mix of techniques - borrowed from whichever area of experimental science that promises to shed light on protein behaviour with an emphasis on microscopy techniques at the single molecule level.

Supervisor: Associate Professor Nikos Hatzakis

Thesis abstract: Thermomyces Lanuginosus Lipase (TLL) is a carboxylester hydrolase that catalyses the hydrolysis of medium- to longchain triacylglycerol (TAG) into glycerol and fatty acids (FA). TLL sparks great industrial interest due to its innate thermostability and functionality at alkaline pH. For this reason TLL is especially important for the detergent industry. However, TLL's dynamics and how dynamics are correlated to function is poorly understood, as current understanding relies on bulk measurements, to which critical parameters, such as structural and functional dynamics are not easily accessible. Furthermore, environmental conditions such as Debye length (in short: a measure of electrical screening in a solution) is

hypothesized to be important, but has not yet been investigated despite the high impact it could pose for the detergent industry. This work will present fluorescence microscopy in a two-pronged approach to study TLL on its native substrate with minimally invasive methods. Firstly, activity of non-labelled TLL is measured by observing surface decay created by TLL over time. Secondly, single particle tracking (SPT) of thousands of individual enzymes simultaneously in real time to allow quantification of functional dynamics as well as observation of spatial localization of TLL in different environmental conditions. Herein, I show by single particle tracking (SPT) that TLL has five intrinsic diffusion states that correlate to activity. Additionally, the SPT readout allowed direct observation of how changes in environmental conditions such as pH and salt concentration not only decrease enzyme activity greatly, but also alter the mobility by shifting the occupancy of states towards the slower states. This information is crucial as it provides a new insight into the connection between protein mobility and function. It is anticipated that this work can be the starting point of assays, that will elucidate how protein dynamics and function correlate and thereby enabling future development of specialized biocatalysts.

Multilayer Reflection Tomography from Monte Carlo Simulations by Rasmus B. Nielsen

Rasmus studied nanoscience to get a general understanding of many underlying processes in nature. Now he uses his knowledge in his further studies in computational physics to create models of these processes. Rasmus likes realizing how complex phenomena can be modelled and analyzed using various computational methods. His current focus is development of algorithms which can accurately simulate and analyze novel methods within X-ray physics. To this end he has worked with Xnovo for a few years. The thesis was done as a collaboration with Xnovo Technology ApS.

Xnovo does research and development of new X-ray techniques and software, focusing primarily on capturing 3D crystallographic information.

Supervisor: Professor Kell Mortensen

Thesis abstract: X-ray Reflection Tomography (XRT) is a tool for imaging of buried layers and interfaces in multilayer thin-films. The method combines X-ray reflectivity and computerized tomography to determine spatially dependent reflectivity curves. Simulations of X-ray reflection tomography experiments make it possible to asses to which accuracy properties can be determined and evaluate various reconstruction methods. It is thus important that such simulations mimic real experiments. In this paper it is shown that a Monte Carlo approach can be used to simulate XRT experiments and test reconstruction techniques. This approach takes into account the statistical properties of an experimental X-ray setup and allows for simulation of diverse experimental configurations. The currently used analytical simulations based on projections do not include such statistics and are limited in scope. The Monte Carlo approach will facilitate further development of the applications of XRT.

Chemical Fabrication and Characterization of Graphene for Water-Based Inks used in Screen-Printing

Mathias Kirkholt Laursen

Graphene exhibits a range of interesting properties such as high optical transmittance, good thermal conductivity, high Young's modulus, and good electrical conductivity. Graphene can serve as a component in conductive applications and potentially replace the expensive conductive materials, such as indium tin oxide (ITO), that are used today. In this article the graphene derivatives graphene oxide (GO) and reduced graphene oxide (rGO) are synthesized. The reduction of GO is done by either microwave assisted reduction, intense light reduction or chemical reduction with hydrogen iodide. The structures of the synthesized compounds are characterized and analyzed using X-ray photoelectron spectroscopy (XPS) and Raman spectroscopy. Based on the synthesized GO, a GO-ink is formulated and screen-printed on a polymer-based polyethylene terephthalate (PET) substrate, from which a sheet resistance of $0.4 \, \text{k}\Omega \, \text{sq}^{-1}$ is measured. This work suggests that the synthesized GO can be used in production of conductive circuits or serve as a precursor for future scalable conductive films.

1 Introduction

Graphene has been the subject of intense research in the last decade due to its application possibilities which includes chemical sensors,[1] micro-supercapacitors,[2] semi-transparent conductive films,[3-4] and many more. The application possibilities with graphene are vast due to several factors such as the large theoretical specific surface area, high Young's modulus, good thermal conductivity, high optical transmittance and good electrical conductivity. These attributes arise from the molecular structure of graphene, which is shown in figure 1d.[5] Graphene consists of a two dimensional carbon structure, which is build up by sp^2-hybridized carbon atoms positioned in a honeycomb lattice. A Nobel price was given in 2010 to Andre Geim and Kostya Novoselov for formulating graphene through mechanical exfoliation.[6] Although, graphene was first synthesized in 1860[7] by Brodie, it is today still almost impossible to synthesize without lattice defects and functional groups.[5] Even though chemical vapor deposition produces a high quality of graphene the downsides to this method is the cost and that it is not scalable. In this article the graphene derivative graphene oxide(GO) is synthesized using the improved Hummers method.[8]

To synthesize GO, a precursor of natural graphite is needed to be separated into single sheets. This is done in two steps, first intercalation of the graphite sheets using sulfuric acid, then exfoliation of the sheets with a sonicator. GO is a single layer of graphene with functional
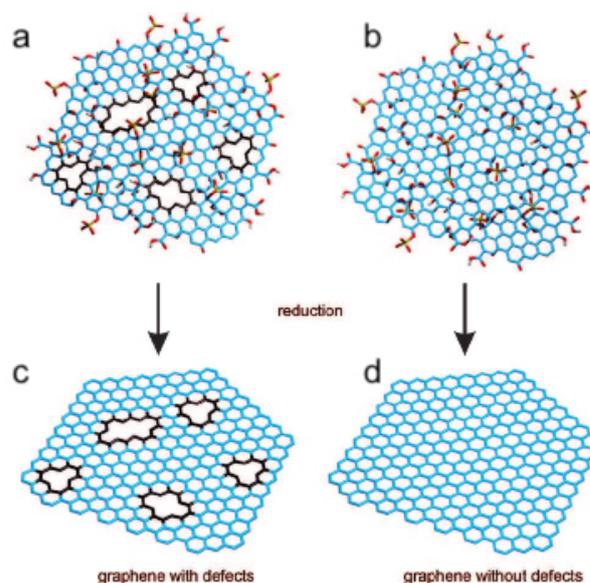
Fig. 1 Theoretical molecular structures of a) graphene oxide (GO) b) GO-without lattice defects, c) reduced graphene oxide (rGO), with lattice defects d) graphene. The figure is taken from article [5].
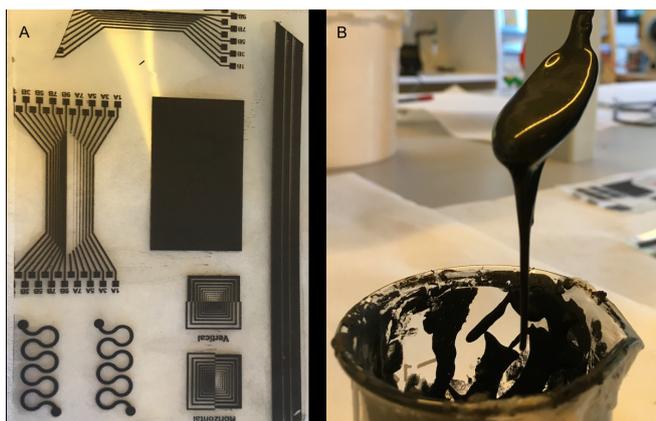
Fig. 2 In picture a), the synthesized GO-ink is shown. The print is made using the printing setup described in the experimental section. In picture b), the viscosity of the GO-ink used in picture a) is presented. The GO-ink was made using a water-based ink and the synthesized GO. The precise mixture of the ink can be found in the supporting information.

groups such as epoxy, hydroxyl and carbonyl groups some of which can be seen in figure 1a and 1b. These functional groups on GO disturb the electronic properties of graphene, and to enhance the conductive nature it is advantageous to remove these groups. This challenge is thoroughly researched and there are several reports that attempt to remove the functional groups and synthesize the reduced-graphene-oxide(rGO).[9-10] The rGO samples synthesized in this article are achieved with either microwave assisted reduction (MW-rGO),[11] intense-light reduction (IL-rGO)[1] or chemical reduction using hydrogen iodide (HI-rGO).[10] The goal of these reduction processes is to remove any functional groups from the GO, and restore the hybridized carbon lattice as seen in figure 1b to 1d. The rGO exhibit characteristic of the perfect graphene lattice and single rGO flakes can become visible in an optical microscope. However, rGO is often synthesized only to find a defective carbon lattice due to holes, additional atoms or lattice frameshifts. The reasons behind these defects are often either an over-oxidation in the initial synthesis, or a too hard reduction resulting in decarboxylation with the release of CO_2 from the GO. This article presents a high quality of rGO from different reduction methods in addition to discussing the effectiveness of the reduction agents.

Printed films have many applications such as flexible touch screens and displays, printable electronics and film heaters.[4] The conductive films in production today, use metal nanoparticles, metal nanowires or transparent conductive oxides such as indium tin oxide.[12] These materials, comes at a high price, and cannot meet the global consumer demands.[3] A water-based ink is preferable as

it is less toxic and less expensive than commercial available inks.[13] Therefore, a GO-ink application has attracted a great deal of interest as it could either directly replace the films in production today or be a component in the production of conductive films.[3] The GO synthesized in this paper, is formulated into a graphene oxide ink (GO-ink) using water as the solvent. In figure 2 the formulated GO-ink is shown. This ink-application can be transferred to a roll-to-roll (R2R) production line used on a flexible and transparent polymer substrate. This technique allows for mass production of low-cost fully printed devices.

2 Results and discussion

The synthesis method used in this article is referred to as the improved Hummers method,[8] as first described by Eigler.[14] Here natural graphite is the starting material, for a scalable production of GO. Sulfuric acid and permanganate is used to synthesize the GO, where it is desirable to prevent an overoxidation of the graphite as it leads to the release of CO_2. The decarboxylation of the GO leaves holes or defects in the graphene lattice, creating irreversible damage to the lattice, ultimately reducing the conductivity of the GO material. To prevent this, the reaction is kept at a low temperature, and the permanganate is added slowly to prevent local heating. To synthesize GO one needs to overcome the Van der Waals $\pi - \pi$ stacking interaction of the carbon sheets. In this article exfoliation with a sonicator, as described in the experimental section, is used to overcome these interactions and create monolayers of graphene sheets. This procedure of producing GO, which is fully described in the experimental section, allows for an almost intact carbon lattice throughout the synthesis, and serves as the precursor for further processing. A theoretical picture of the synthesized GO is shown in figure 1b.

To investigate and further understand the effects of different reduction techniques X-ray photoelectron spectroscopy (XPS) measurements were performed. The GO examined by XPS consists of several peaks as seen in figure 3a. By decomposing the XPS C1 spectral lines, each peak correspond to specific bonds. The C=O, C-O, C-C and C=C each has a specific binding energy and is presented in figure 3. The peaks provide details about the binding state of the carbon in the sample. The two very intense peaks at energies 287 eV and 285 eV correspond respectively to the C-O and C-C bonds. Furthermore, peaks at binding energies 289 eV and 284 eV correspond to the bonds C=O and C=C. This allows for calculating the composition of the carbon in the GO sample and each of the rGO samples. Using the XPS measurements the chemical binding states of the GO and rGO samples was determined and compared. All XPS measurements are shown
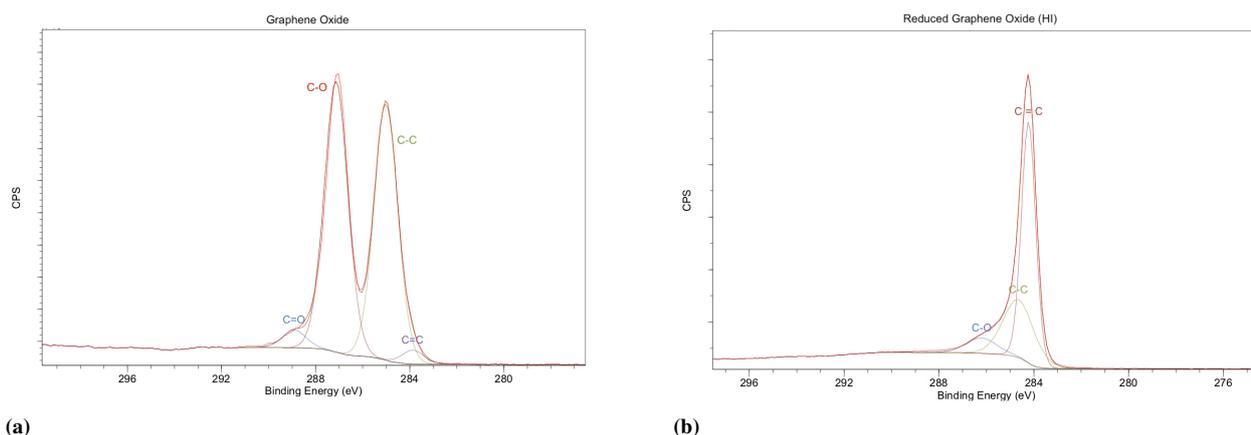
Fig. 3 In each figure the CPS corresponds to the intensity in arbitrary units. a) In this figure, the XPS data from the synthesized GO is shown. b) In this figure, the XPS data, from the HI-rGO is presented. The reduction process for this figure was done using hydrogen iodide(HI). In table 1, all data from the XPS measurements can be found.

Table 1 XPS spectroscopy data

Samples	Composition of the sample in %				Composition of the carbon in %			
	C	O	N	S	C-O	C-C	C=O	C=C
GO On SiO$_2$	67	27.1	2.5	3.4	46.5	42.1	5.0	6.4
rGO-MW powder	15.8	56.6	11.2	16.4	42.7	49.3	8.0	<1.0
rGO-IL On SiO$_2$	86.3	13.1	<1.0	<1.0	6.1	30.5	<1.0	63.41
	C	O	I	Traces of N and S	C-O	C-C	C=O	C=C
rGO-HI On SiO$_2$	90.2	6.5	2.3	<1.0	8.3	31.5	<1.0	60.2

in table 1. Each C-O peak corresponds to carbon atoms bound to either a hydroxide group (C-OH) or a epoxide group (C-O-C). After the chemical reduction with hydrogen iodide(HI), a major reduction in the C-O level is observed. Furthermore, the content of C=O, and thus the carbonyl groups (C=O), is no longer measureable and is determined to be less than 1%. These measurements indicate the effectiveness of the chemical reduction technique, as the amount of carbon-oxygen functional groups have decreased drastically. Additionally, the amount of C=C bonds and thus the sp^2 hybridization of carbon increase significantly, which is expected to be caused by the reconstruction of the carbon lattice network. The chemical reduction with HI is believed to follow an E1 elimination reaction mechanism.[15] This would remove functional groups from the graphene oxide and would explain the reconstruction of hybridized carbon lattice. However, as graphene is a big two-dimensional carbon network, with each carbon atom affecting the next one, the mechanism for this reaction is not completely understood and further research is needed in that area.

In this article the reduction of GO is done with by either MW-reduction, IL-reduction, or chemical reduction using HI. In the case of MW-reduction and IL-reduction, it is believed that the energy supplied, either by the microwave or the light introduced on the sample, forces the chemical binding to the functional groups to become unstable. However, if too much energy is dispersed to the carbon lattice, an irreversible decarboxylation of the carbon lattice would occur followed by the release of CO$_2$.

The XPS measurements of the synthesized GO and reduced GO derivatives are displayed in table 1. The traces of sulfur are believed to be from the sulfuric acid used in the synthesis method[8] and the nitrogen is estimated to be due to contamination in each sample. However, it is worth noticing that the contamination levels decrease after reduction of the sample, except for the MW-rGO. When synthesizing the MW-rGO, arcs were observed when the sample was exposed to microwaves, which indicates a violent reduction of the sample. It can be argued that the microwaves energy was too high, and a lot of the GO was reduced to CO$_2$ due to decarboxylation. This would also

Table 2 An average of the Raman Data

Sample	D-Band Peak, cm^{-1}	G-Band Peak, cm^{-1}	2D-Band FWHM, cm^{-1}	Intensity ratio, I_D/I_G
GO-MeOH	1406.1	1559.6	432.4	1.439
GO-MiliQ	1404.9	1560.8	404.1	1.411
rGO-MeOH	1405.4	1551.0	164.1	1.770
rGO-MiliQ	1407.8	1551.6	177.6	1.799

explain the high oxygen level in the MW-rGO sample and that is believed to have completely ruined the graphene lattice.

It is believed that the contamination in the samples decrease after reduction because several functional groups are removed from the GO. As seen in table 1, a significant decrease in both oxygen, nitrogen and sulfur is observed when comparing rGO-HI, and rGO-IL with the GO sample. Furthermore, after both the chemical- and light reduction, the carbon-oxygen bonds have decreased significantly as well. In addition, the C=C bonds after reduction shows a major increase. These indications both demonstrate that the reduction of GO was successful. The carbon lattice in these samples should have a low density of defects and serve effectively as a conducting material. The best and most efficient reduction technique is determined to be the intense light reduction. This is evident from the XPS data in table 1 showing a high decrease in C-O bonding compared to the other reduction methods. Additionally, the intense light reduction technique could be further refined by targeting the binding of specific functional groups. The energy supplied by the intense light is determined by the distance to the sample, the exposure time, and the wavelength of the mercury lamp. A way to improve the IL-reduction method would be to target specific functional groups by exposing the sample with light at a specific energy. Another possible way to improve the reduction of GO is to combine the chemical reduction with the light reduction. Both methods could potentially create an even higher quality of rGO. In this work the IL-reduction is performed with a mercury lamp for about 1-2 seconds and the sample was about 1 cm from the light source.

The improved Hummers method[8] is used to synthesize the GO in this article. To determine the quality and whether the chemical reduction had any influence on the quality, Raman spectroscopy was performed on the synthesized compounds. The spectroscopy measurements were done on a GO and a chemically reduced rGO sample. The full width at half maximum (FWHM) of the 2D band and the intensity ratio I_D/I_G, in the Raman spectroscopy, are correlated to the amount of defects in the carbon lattice. Typical Raman spectra of graphene displays three characteristic peaks; a D-peak, a G-peak and a 2D-peak.[5] All
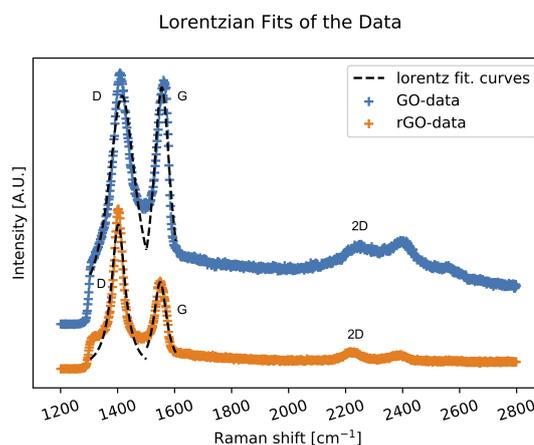
Fig. 4 The Raman spectroscopy data represented here is from both of the GO and the chemically reduced rGO sample. There is an offset between the two curves. The characteristic peaks of D, G and 2D for GO and rGO, are indicated on each data sets and a Lorentzian function have been fitted to the D and G peak. Furthermore, a FWHM calculation was made on the 2D peak.

peaks are visible in figure 4 where the D-peak is around 1400 cm^{-1}, the G-peak is approximately at 1555 cm^{-1}, and the 2D-peak around 2200 cm^{-1}. The intensity ratio, I_D/I_G, indicates the quality of the carbon lattice in the sample and is displayed in table 2. As described by Eigler[16] the FWHM of the 2D peak and intensity ratio I_D/I_G are correlated (this trend can be seen in supplementary information figure S4). At the measured FWHM of the 2D peak in this work, the intensity ratio I_D/I_G is believed to increase with a lower density of defects.

The Raman measurements done on the samples reports a small increase in the intensity ratio after chemical reduction with HI on GO. This increase in intensity ratio I_D/I_G can be explained by the restructure of the carbon lattice. The HI is believed to restore the conjugated system in the carbon lattice structure when removing functional groups. This is also indicated by the drop in the 2D-peak FWHM-calculation when comparing GO and rGO samples. After chemical reduction the peak is more defined and resembles the peak of pure single sheet graphene. Thus, it can be derived from the Raman spectroscopy data that the chemi-

Table 3 Sheet resistance of the different GO-inks

Sample	R_s kΩ sq^{-1}
GO-ink	10.4
rGO-ink	0.7
rGO-ink pressed	0.4
rGO-(Cellulose)-ink	20.2

cally reduced rGO decreases the density of defects.

The synthesized graphene oxide inks (GO-ink) were printed with a woven polyester screen onto a polymer based polyethylene terephthalate (PET) substrate. The GO and rGO were each applied, to a water based graphene ink and tested with a screen-printing method. The synthesized GO proposes a challenge, as the oxygen levels in the GO sample measured to almost 30%. In addition, 50% of the carbon in the GO sample is bound either to a hydroxide group, a epoxide group or a carbonyl group. Since there are many functional groups and defects in the lattice structure of the synthesized GO, the GO is expected to have a high resistivity. The conductivity of the all GO-inks were tested using a probe station which measures the resistance of a sample. This is used to calculate the sheet resistance (R_s) of each sample. In table 3, an average of the various sheet resistances of all synthesized inks are presented. As expected, the GO has a very high sheet resistance which is deemed due to the high oxygen level in that sample.

In an attempt to optimize the GO-ink different GO derivatives were tried. Firstly, the GO was chemically reduced using HI (rGO-ink), then the sample was pressed using a pressing drum (rGO-ink pressed). This resulted in a low sheet-resistance of 0.4 kΩ sq^{-1}. However, the pressure applied to the ink, resulted in a very fragile ink which fractured at a low force. Thus, another ink was synthesized using a mixture of rGO and hydroxyethyl cellulose which acted as a binding agent. The print results were visibly of much higher quality, which can be seen in figure 2A, where the print with hydroxyethyl cellulose is displayed. Even though the hydroxyethyl cellulose print was of high quality, the downside of using hydroxyethyl cellulose is proven to be a high sheet resistance. The increase in sheet resistance observed in the last print is believed to be caused by the insulating nature of the hydroxyethyl cellulose. To see whether the surface tension of the substrate had any effect on the quality of the prints, the polymer substrate was treated with oxygen plasma before applying the ink. This made a major visual improvement on the last print where the oxygen plasma was applied. This illustrates that each composition of ink, should be paired with a specific substrate, as surface tension of the substrate is vital to the finished printing result. However, it is believed that the oxygen plasma applied to the substrate only helps with the visual result, not the conductivity of the print.
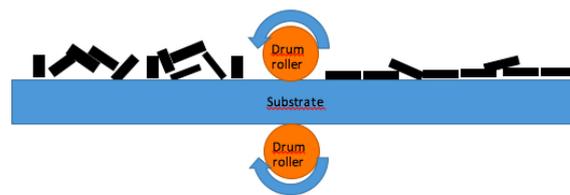
Fig. 5 Here the theoretical effect of pressing the GO-ink is shown. The black boxes represent the GO-flakes in the ink and before applying the pressure the GO-flakes are randomly dispersed. After the pressure is applied the GO-flakes are flat and connected.

The print with hydroxyethyl cellulose shows that if a correct binding agent is added to the GO-ink, it could result in a visually high quality print, and the result with the pressed rGO-ink shows that a high conductivity can be achieved. If a binding agent can be found that combines these results, then it would be possible to synthesize a fully conductive and flexible ink. The low sheet resistance measured from the pre-pressured rGO-ink is believed to be caused by the alignment of graphene flakes in the sample. This pressure effect is presented in figure 5.

To potentially improve the conductivity of the ink one could try and reduce the GO-ink by either thermal reduction or light reduction after printing. Another improvement could be to try to induce some conductive properties with dopants. Furthermore, the optimal print conditions are yet to be found so there is room for improvement in that area as well.

It is difficult to directly compare the commercially available ITO-ink[12] with the synthesized GO-ink because of the difference in film thickness.

3   Conclusion

In conclusion, the synthesized GO in this article works as a precursor for water-based conductive screen-prints. The GO was synthesized using the improved Hummers method,[8] for a high quality GO and then reduced using either chemical, microwave, or light-reduction. As in previous work,[1] the light reduction was very successful with the lowest measured oxygen level as shown in table 1. The Raman spectroscopy measurements of the samples with HI indicated a high quality of GO with a clear 2D-peak after chemical reduction of GO. This showed that with chemical reduction of GO, the GO would reconstruct the graphene lattice through an E1-like elimination reaction mechanism.

The synthesized GO was mixed with ink and printed onto a polymer substrate. Both the GO, rGO, and pre-pressed rGO, was measured for their conductivity using a probe station. The print with the lowest sheet resistance

proved to have a sheet resistance of 0.4 kΩ sq^{-1}. This is in line with previous work.[4] However, in this article the prints are scalable and inexpensive to produce. Furthermore, there are several options to improve the conductivity of the prints by either dopants or either thermal or light reduction of the prints.

This work proves that the synthesized GO can either serve as a precursor for other conductive applications, or be printed with the screen-printing method shown in this work.

4 Experimental

The synthesis of GO was done using a method described by Eigler.[8] The starting point of the synthesis is natural graphite. Here 2.01 g of large flaked graphite (<300µm), was loaded into a large round beaker. The beaker was put in an isolating box, where it was constantly cooled by a mixture of dry ice and water, and stirred at a moderate tempo. 50 mL of a 95% sulfuric acid solution was added to the beaker. When the graphite was completely suspended in the acid, 4 g of potassium permangate was added over 4 hours. The solution was continuously stirred and cooled for 12 hours while the reaction with potassium permangate occurred. Then 100 mL of diluted 20% sulfuric acid was added at a rate of 40 mL per hour. This was followed by the addition of 240 mL of miliQ water at a rate of 8 mL per hour. Then, 160 mL of 20% hydrogen peroxide was added over 12 hours. The beaker was kept cool and with a moderate stirring during every step. The obtained suspension of GO was sonicated with a Hielscher UPS400 (400 W) ultrasonic processor at an amplitude of 55% with 0.5 cycle for 4 minutes total. Finally, the GO solution was washed 4 times using cold (5°C) miliQ water and centrifugation at 9000 RPM for an hour. After each cycle the supernatant was replaced and the precipitate was resuspended. The pH was measured to be approximately 7 with pH paper after the washing. To remove small particles from the GO solution it was centrifuged 3 times at 2000 RPM for 16 minutes, and the precipitate was removed after each cycle. To further clean the solution and remove small graphene-like particles, the GO suspension was centrifuged at 9000 RPM 4 times for 45 minutes, each time replacing the supernatant.

Chemical reduction of GO was made using HI. The GO was placed on a silicon oxide wafer in drops of about 50 µL. The drops of GO were added each day for a total of 4 straight days, each time the GO layer was allowed to dry in atmosphere. Before adding the GO, the silicon oxide wafer was cleaned using acetone, ethanol and isopropanol, and then prepared using a base piranha solution which consist of 1/4 ammonia, 1/4 hydrogen peroxide and 1/2 miliQ water. The silicon-wafer, was now placed in a gas chamber with a single drop of HI. Next, the gas chamber was heated to 60°C for 30 minutes. This should reduce the GO and create reduced graphene oxide (rGO).

Physical reduction of GO was done using microwave.[11] GO was centrifuged at 9000 RPM for one hour. The excess water was removed from the GO solution and about 10 mL of thick GO solution remained. This was now freeze dried using a cold trap (for details on the method see supplementary information). The dry GO was now mildly reduced by thermal annealing at 300°C for one hour under a flow of nitrogen. The mildly reduced GO was placed in a glass beaker, which was filled with nitrogen or argon and sealed. The beaker with GO was placed inside a microwave oven, and heated at 800 W for 1-2 seconds. Arcs from the GO were observed which suggests an extremely fast annealing process where the GO is heated to a couple of thousands degrees Celsius within a few milliseconds. The MW-rGO was finally allowed to rest for some minutes.

Light reduction of GO was done using light from a mercury lamp.[1] A drop of GO from the stock solution was added to a silicon oxide wafer. The wafer was cleaned and prepared as in the chemical reduction part above. The mercury lamp was allowed to heat up for 5 minutes before exposing the samples to the light. The GO-samples on the wafers were held using a tweezer about 1 cm from the light source for 1-2 seconds. This should reduce the GO and form the intense light reduced graphene oxide (IL-rGO).

Graphene oxide was printed on a PET substrate. The print was done using a mixture of the stock GO solution and a water based ink. The stock GO solution was high speed centrifuged at 9000 RPM and excess water was removed. The remaining GO was then mixed with the ink and stirred for 10 minutes, then a highly viscous solution was achieved. About 15 mL of the GO-ink solution was needed for each print.

The screen-printing with GO was done using a mesh consisting of a tightly woven polyester yarn. The mesh used in this article had mesh openings between 57 µm and 77 µm. For each print the GO-ink solution was added to the mesh and a squeegee was used to press it through the mesh and onto the substrate. The print thickness could be varied by changing the number of print passes, varying the pressure from the squeegee, varying the angle from the squeegee to the mesh, and changing the viscosity of the GO-ink solution. Further information about the screen print can be found in supplementary information.

Raman spectroscopy was performed on GO and the chemically reduced rGO. A silicon oxide wafer was prepared as described earlier. 5 µL of the stock GO solution was spin coated on the silicon wafer. The spin coating was done at 3000 RPM for 30 seconds. Some of the GO samples were then reduced using the chemical reduction technique as described previously. The sample was placed

upside down on an Olympus IX71 microscope. The laser was aimed directly at flakes on the sample. The measurement performed by the Raman setup were done with a 543 nm laser. For the precise Raman setup and all Raman data see the supplementary information

X-Ray photoelectron spectroscopy (XPS) was performed using a Kratos Axis Ultra-DLD instrument. Samples was prepared on the silicon wafers as described previously. XPS measurements were performed on both the GO and different rGO samples. The rGO samples were produced either by microwave assisted reduction, intense-light reduction, or by chemical reduction using HI. The MW-rGO was a fluffy material which was compressed to a powder and then XPS was performed on that sample. For further information about the X-Ray see supplementary information.

Vapor deposition was performed on the finished screen prints to form gold electrodes. The bottom side of the screen printed polymer substrates, were fastened to a silicon wafer. These were then lowered into a vacuum tank where gold electrodes were deposited on the ink using a mask. A picture of the gold electrodes can be seen in the supplementary information.

The sheet resistance of each print was measured using a probe station, where a current was induced from the outer electrodes. The voltage drop over the inner electrodes was then measured. The sheet resistance for the material could then be calculated using the equation [17] $R = R_s \cdot L/W$ where R_s is the sheet resistance, R is the resistance measured, L and W is the length and width of the sample.

5   Acknowledgements

References

1 S.-J. Choi, S.-J. Kim and I.-D. Kim, *NPG Asia Materials*, 2016.

2 Y. M. Shulga, S. A. Baskakov, Y. V. Baskakova, A. S. Lobach, E. N. Kabachkov, Y. M. Volfkovich, V. E. Sosenkin, N. Y. Shulga, S. I. Nefedkin, Y. Kumar and A. Michtchenko, *Journal of Alloys and Compounds*, 2018, **730**, 88–95.

3 M. H. Overgaard, M. Kuhnel, R. Hvidsten, S. V. Petersen, T. Vosch, K. Norgaard and B. W. Laursen, *Advandes Materials Technologies*, 2017.

4 L. He and S. C. Tjong, *Journal of Materials Chemistry C*, 2016, **4**, 7043–7051.

5 S. Eigler and A. Hirsch, *Angewandte Chemie-International Edition*, 2014, **53**, 7720–7738.

6 A. K. G. . K. S. Novoselov, *Nature Materials* ,, 2007, **Vol. 6**, year.

7 B. C. Brodie, *Proceedings of the Royal Society of London*, 1860, **10**, 11–12.

8 E. Siegfried, E. Michael, G. Stefan, H. Philipp, K. Wolfgang, G. Andreas, D. Christoph, R. Michael, X. Jie, P. Christian, L. Ole, S. HansPeter, M. Paul and H. Andreas, *Advanced Materials*, 2013, **25**, 3583–3587.

9 F. T. Johra, J.-W. Lee and W.-G. Jung, *Journal of industrial and Engineering Chemistry*, 2014, **20**, 2883–2887.

10 S. Eigler, S. Grimm, M. Enzelberger-Heim, P. Muller and A. Hirsch, *Chem. Commun.*, 2013, **49**, 7391–7393.

11 D. Voiry, J. Yang, J. Kupferberg, R. Fullon, C. Lee, H. Y. Jeong, H. S. Shin and M. Chhowalla, *SCIENCE*, 2016, **353**, 1413–1416.

12 M. sung Hwang, B. yong Jeong, J. Moon, S.-K. Chun and J. Kim, *Materials Science and Engineering: B*, 2011, **176**, 1128 – 1131.

13 H. 1 2 3 4 Kipphan, *Handbook of print media:*, technologies and production methods (Illustrated ed.), 2001, vol. pp. 130–144, ISBN 3-540-67326-1.

14 S. Stankovich, D. A. Dikin, R. D. Piner, K. A. Kohlhaas, A. Kleinhammes, Y. Jia, Y. Wu, S. T. Nguyen and R. S. Ruoff, *Carbon*, 2007, **45**, 1558 – 1565.

15 J. E. McMurry, *Organic Chemistry*, Brooks/Cole, 8th edn, 2012.

16 S. Eigler, F. Hof, M. Enzelberger-Heim, S. Grimm, P. Mueller and A. Hirsch, *Journal of Physical Chemistry C*, 2014, **118**, 7698–7704.

17 J. G. Webster, *Electrical Measurement, Signal Processing, and Displays, Chapter 7 "Electrical Conductivity and Resistivity*, CRC Press, 2003.

Thesis Paper

Characterization of the formation of magic-sized {Bi$_{38}$O$_{45}$ clusters by PDF and SAXS

Andy Sode Anker,[*][a]

5

Bismuth oxido clusters exist in a range of sizes, all built up by octahedral {Bi$_6$O$_8$} units. While the atomic structure of various clusters has been solved by single crystal diffraction, it is much more challenging to study clusters directly in solution. Here, we use *in situ* X-ray total scattering with Pair Distribution Function (PDF) analysis to study the formation of the {Bi$_{38}$O$_{45}$}a cluster from [Bi$_6$O$_5$(OH)$_3$(NO$_3$)$_5$]·(H$_2$O)$_3$ crystals dissolved in DMSO. The implementation of PDF analysis provides a unique insight into the structural rearrangements on the atomic scale. By combining with Small Angle X-ray Scattering (SAXS) we can furthermore investigate the size, morphology and size dispersion of the clusters taking place in the process. Consequently, the combination of these two complementary techniques provides a mean of bridging the local atomic and macroscopic characteristics of the material.

In the presented study, the results obtained show that the reaction goes through several stable intermediates before the magic-sized product {Bi$_{38}$O$_{45}$} is reached. Through an associated temperature study, the intermediate was furthermore found to be capable of stabilization of up to days by varying the resulting reaction rate. The present studies show how powerful a tool the Debye Equation is in combination with SAXS and PDF, which in this field is a new development that has a large potential for unravelling important questions in nanochemistry in solution.

Introduction

In recent years, bismuth oxido clusters have been widely studied due to their applications ranging from medicin[1, 2], for designing radiopaque materials[2, 3] and building blocks for advanced catalysts[4], which all is possible due to their low toxicity[3, 5].

Bismuth oxido clusters exist in a range of sizes, all built up by octahedral {M$_6$O$_X$} (M = Bi) units, which is commonly accepted as a building block in inorganic chemistry of large metal cations [6-8]. The octahedral {Bi$_6$O$_8$} unit can be seen in Figure 1, right. With mass spectroscopy and single crystal diffraction, it has been shown that a range of large metal cations form stable clusters of {M$_{22}$O$_X$} and {M$_{38}$O$_X$}[8-11] both built by octahedral {M$_6$O$_X$} units. Both clusters are, for the bismuth oxido clusters, illustrated in Figure 2. The {M$_{38}$O$_X$} cluster is sometimes referred to as magic-sized[8, 11], which means that they are atomically monodisperse[12].

Despite the large potential of using this as a model system in studies of fundamental cluster chemistry of large cations, the chemical processes involved in the cluster formation are not well understood[8], neither 22- or 38 bismuth
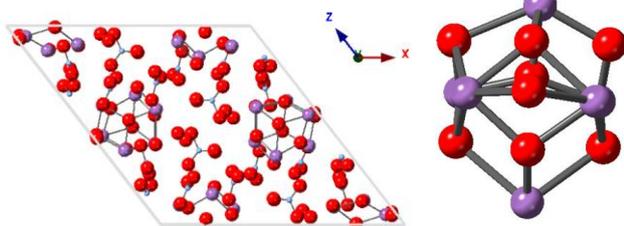
atoms are equally dividable by 6, so how does the building blocks create the {Bi$_{22}$O$_X$}- and the {Bi$_{38}$O$_{45}$} cluster?

While single crystal diffraction is restricted to samples of high crystallinity, it cannot be used to study materials in solution, as a result of the missing global order in the structure. Mass spectroscopy has been used to investigate the chemical composition of various clusters[13] but does not give any structural information. Therefore, it has been found challenging to study clusters directly in solution. This challenge can be generalized to other cluster systems[14].

If we were able to characterize the growth of the clusters on a molecular scale, we would get an understanding of the fundamental chemistry of clusters, and thereby be able to tailor-made materials with specific properties[8, 15]. This opens a new field of research, determining structures between molecules and particles[16, 17], whereas methods in the cross border of these fields must be used.

Figure 1: Left) The [Bi$_6$O$_5$(OH)$_3$(NO$_3$)$_5$]·(H$_2$O)$_3$ crystal which is built by the Right) octahedral {Bi$_6$O$_8$} unit. Purple) bismuth, red) oxygen, blue) nitrogen. Hydrogens are not included in the figure.
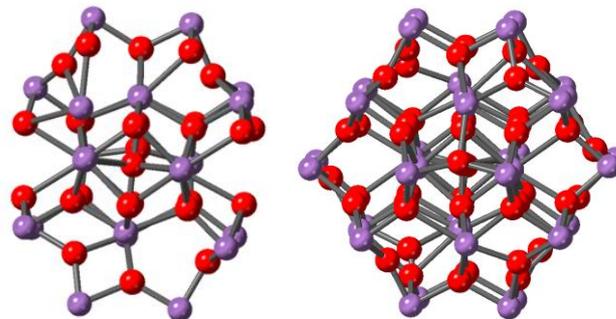

Figure 2: Left) The {Bi$_{22}$O$_X$} cluster seen with single crystal diffraction and mass spectroscopy. Right) The {Bi$_{38}$O$_{45}$} cluster seen with single crystal diffraction.

This study presents *in situ*[15] X-ray total scattering with Pair Distribution Function[14] (PDF) analysis of the formation of the {$Bi_{38}O_{45}$} cluster in solution. {$Bi_{38}O_{45}$} clusters are synthesized by dissolving crystals of [$Bi_6O_5(OH)_3(NO_3)_5$]·(H_2O)$_3$ in DMSO.

The PDF analysis has given a unique insight into the structural rearrangements on an atomic scale, but PDF gives limited information on the nanometer scale. Therefore, Small-Angle X-ray Scattering[18] (SAXS) is used to investigate the size, morphology and size dispersion of the clusters taking place in the process. These two techniques complement each other allowing us to follow the cluster formation in both the local- and global order of the particle. The combination of SAXS and PDF bridge the local atomic and macroscopic characteristics of materials. Further work can be done in order to combine these techniques with complex modelling[19], which has a large potential for unravelling important questions in nanochemistry in solution[19, 20].

Since the bismuth oxido cluster does not consist of a periodic order, the Debye Scattering Equation[21, 22] is used to calculate the theoretical scattering. The Debye Equation is a sum of the scattering contribution from every atom in an isotropic sample which does not necessarily have to contain a periodicity in the structure.

Equation 1: $I(Q) = N_j f_j^2 + N_i f_i^2 + f_i f_j \sum_i^N \sum_{j \neq i}^N \frac{sin(Q \cdot r_{ij})}{Q \cdot r_{ij}}$

With the exponential increase in computational power the last decades[23], and the focus on nanomaterials, the Debye Scattering Equation becomes essential to characterize materials.

Further details of the methods are available in Supplementary S2.

Experimental section

Synthesis of {$Bi_{38}O_{45}$} clusters for total scattering experiments

To characterize the formation of {$Bi_{38}O_{45}$} clusters 131.4 mg of crystalline [$Bi_6O_5(OH)_3(NO_3)_5$]·(H_2O)$_3$ was dissolved in 2 mL DMSO at room temperature. Additionally, a ligand exchange experiment was performed to investigate the effect of the ligand on the cluster structure. Therefore, 100 mg of [$Bi_{38}O_{45}(NO_3)_{24}(DMSO)_{25}$] and 37 mg of $NaOOCC(CH_2)CH_3$ was dissolved in 2.5 mL DMSO. These solutions were used for *in situ* PDF[15] experiments at DESY, P02.1. The solutions were loaded in Kapton tubes with an inner diameter of 1.5 mm and measured at temperatures ranging from 30°C to 80°C with X-ray wavelength of 0.207170 Å.

In the *ex situ* PDF experiments, 32.85 mg of crystalline [$Bi_6O_5(OH)_3(NO_3)_5$]·(H_2O)$_3$ was dissolved in 0.5 mL DMSO at room temperature and maintained undisturbed until measurements after 2 days, 4 days, 11 days and 100 days.

The experiments were performed at DESY, P07. The solutions were loaded in Kapton tubes with an inner diameter of 1.5 mm and measured at room temperature with X-ray wavelength of 0.123506 Å.

All total scattering data were integrated using the programme Fit2D[24], and Fourier transformed with PDFgetx3[25, 26] to obtain PDFs. The modelling was done using DiffPy-CMI[19]. When analysing the intermediate, only the bismuth atoms were modelled, since the radial distribution function is directly proportional to the

form factor squared as well as the Debye Equation (Equation 1) (Supplementary S2.2.2).

Synthesis of {$Bi_{38}O_{45}$} clusters for SAXS experiments

In the SAXS experiments 32.85 mg of crystalline [$Bi_6O_5(OH)_3(NO_3)_5$]·(H_2O)$_3$ was dissolved in 0.5 mL DMSO at room temperature. Afterwards, the samples were stirred until no precipitate was left (a few hours) and maintained undisturbed until measurement after 5 hours and 45 hours.

Additionally, a sample was heated at 40°C, in order to investigate the kinetic behavior.

The SAXS experiments were done at the NBI Institute, Copenhagen. The solutions were loaded in glass capillaries tubes with an inner diameter of 0.6 mm and measured at room temperature with X-ray wavelength of 1.5418 Å.

The data were integrated using the programme Fit2D[24] and analyzed with SASfit[27] for the SAXS form factor analysis, and with a Python script for the calculations of the Debye equation (This is further elaborated in Supplementary S2.3.5).

Further details of the experiments are available in Supplementary S1, and details about the PDF- and SAXS analysis is in Supplementary S2.

Results and Discussion

Formation of the {$Bi_{38}O_{45}$} cluster from the crystalline [$Bi_6O_5(OH)_3(NO_3)_5$]·(H_2O)$_3$ dissolved in DMSO

Figure 3, left, shows the time-resolved PDFs obtained from the dissolution of [$Bi_6O_5(OH)_3(NO_3)_5$]·(H_2O)$_3$ in DMSO at 80°C. The beginning of the reaction can from the long-range order be identified as a crystalline phase in suspension.

Figure 1, left, shows the crystalline [$Bi_6O_5(OH)_3(NO_3)_5$]·(H_2O)$_3$ structure which was determined with single crystal diffraction[28] in 1978. It contains bismuth oxido clusters in the octahedral building block {Bi_6O_8} as well as nitrate groups and water.
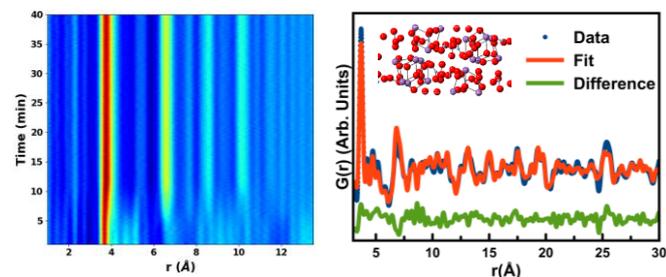
Figure 3: Left) The *in situ* measurement of the 80°C PDF measurement clearly indicates a structure change during the reaction. Right) The crystalline [$Bi_6O_5(OH)_3(NO_3)_5$]·(H_2O)$_3$ phase describes the first 7 minutes of the reaction well. : Graph-description: Left) Time is plotted versus the interatomic distances in the cluster with the intensity shown with color code from light blue to dark red. Right) The probability of finding pairs of atoms separated by a distance "r" is plotted versus the interatomic distances, "r". R_w = 29.0 %.

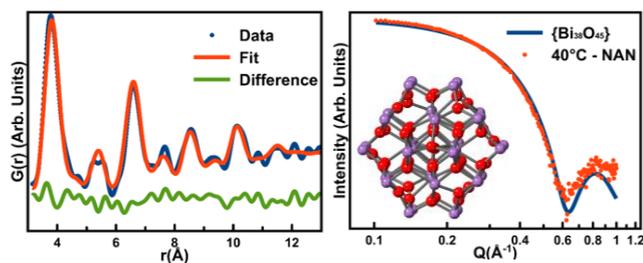Figure 4: The last frames of the 80°C *in situ* PDF measurement are described well by the {Bi₃₈O₄₅} structure. Left) The calculated structure of {Bi₃₈O₄₅} with the Debye Equation describes the data after 40 minutes of the beginning of the experiment in the *in situ* PDF measurement well, right) The calculated structure of {Bi₃₈O₄₅} with the Debye Equation describes the SAXS data from the sample prepared at 40°C well. Inset) The structure of {Bi₃₈O₄₅}. Graph-description: Right) The intensity is plotted versus the scattering vector **Q**.

At the beginning of the reaction, the PDFs can be fitted by the $Bi_6O_5(OH)_3(NO_3)_5 \cdot (H_2O)_3$ structure[28], as seen in Figure 3, right. Furthermore, a PDF was obtained by the crystalline $[Bi_6O_5(OH)_3(NO_3)_5] \cdot (H_2O)_3$ solid, which also fits well to the $[Bi_6O_5(OH)_3(NO_3)_5] \cdot (H_2O)_3$ structure from the literature (Supplementary S3.1.1 Figure 18). Since the molecular crystals have stronger intramolecular forces than intermolecular forces, the refinement was improved significantly by including two isotropic displacement factors[17] (Further details are available in Supplementary S3.1.2).

Within the first 7 minutes of the *in situ* PDF experiment performed at 80°C, the long-range order disappears as illustrated by the absence of the initial crystalline phase. This provides evidence of a structural change, where the crystalline suspension dissolves to small nanoparticles without long-range order.

At the last frame, which is 40 minutes after the beginning of the experiment, the reaction is expected to be in equilibrium and have formed {Bi₃₈O₄₅}, since a similar reaction has been seen to create {Bi₃₈O₄₅} clusters[13]. This is furthermore confirmed by fitting a model of the calculated Debye Scattering from the {Bi₃₈O₄₅} structure to the PDF from the last frame as seen in Figure 4, left.

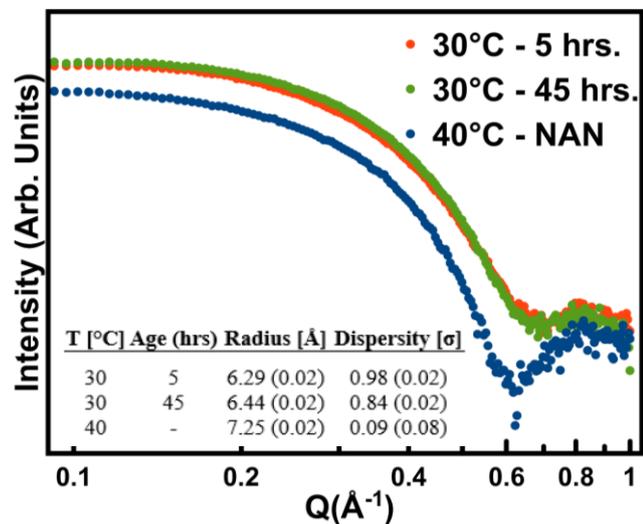| T [°C]	Age (hrs)	Radius [Å]	Dispersity [σ]
30	5	6.29 (0.02)	0.98 (0.02)
30	45	6.44 (0.02)	0.84 (0.02)
40	-	7.25 (0.02)	0.09 (0.08)

Figure 5: The SAXS measurements shows that the clusters will grow with time and temperature. Also the large cluster is characterized as magic-sized (fits are available in Supplementary S3.2.1).
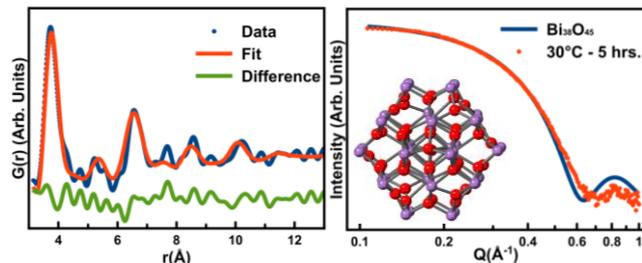

Figure 6: The data of the intermediate is not described well by the {Bi₃₈O₄₅} structure. Left) The calculated structure of {Bi₃₈O₄₅} with the Debye Equation does not describe the data of the 30°C *in situ* PDF measurement after 3 hours well. Right) The calculated structure of {Bi₃₈O₄₅} with the Debye Equation does not describe the *ex situ* SAXS data of the 30°C measurement after 5 hours after the beginning of the experiment, well.

To support the *in situ* PDF data, *ex situ* SAXS data were obtained from samples prepared under similar conditions. Only the time and temperature of the reaction were different. Figure 4, right shows the Debye Scattering of the {Bi₃₈O₄₅} structure fitted to the SAXS data of the sample prepared at 40°C. Since the model also describes the SAXS data very well, it indicates strongly that the dissolution of $[Bi_6O_5(OH)_3(NO_3)_5] \cdot (H_2O)_3$ in DMSO creates the {Bi₃₈O₄₅} cluster at equilibrium. It has earlier been shown that similar reactions create stable {Bi₃₈O₄₅} clusters[13] (Supplementary S3.6). From an analysis of the SAXS form factor, the radius of the spherical {Bi₃₈O₄₅} cluster was estimated to be 7.25 Å and the size dispersity to be 0.09 Å (Figure 5), which is the first time that it has been quantified that {Bi₃₈O₄₅} is magic-sized in solution.

Investigating the PDF's obtained at lower temperatures, 30°C-60°C, the {Bi₃₈O₄₅} cluster model does not describe the data well, indicating that the {Bi₃₈O₄₅} cluster has not yet been formed. Figure 6, left, shows the calculated Debye scattering fitted to the *in situ* PDF measurement at 30°C after 3 hours of the experiment. The structure presented by the measured data does exhibit similarities to the utilized {Bi₃₈O₄₅} cluster model. However, for a few of the peaks, such as at r = 7.8 Å, the intensity is not fully described with the given model. Concludingly, the given structure must be described as a structure with some similar structural features of the known {Bi₃₈O₄₅} structure, but it cannot solely be described with this model. Furthermore, *ex situ* SAXS data of the sample prepared at room temperature and measured 5 hours after initialization of the experiment was modelled with the {Bi₃₈O₄₅} cluster in Figure 6, right. It can be seen that the model in the high Q-value regime does not fit the data well, which supports the conclusion that this is an intermediate. Both the PDF and SAXS data collected from samples prepared at low temperatures cannot
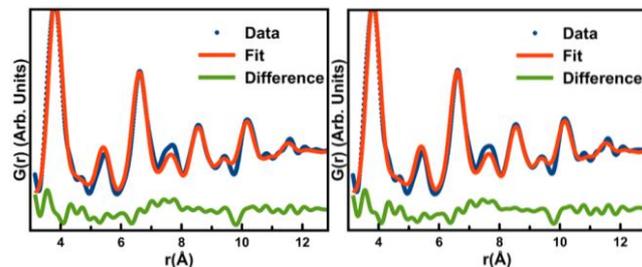

Figure 7: The {Bi₃₈O₄₅} structure describes the *ex situ* PDF data of the samples prepared at room temperature well after – left) 2 days, right) 100 days.

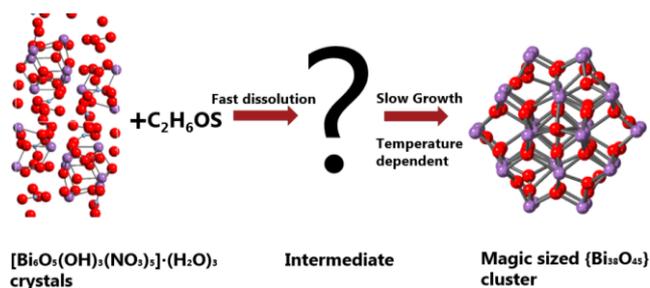Figure 8: The $[Bi_6O_5(OH)_3(NO_3)_5]\cdot(H_2O)_3$ crystals dissolved in DMSO have shown to go through an intermediate before growing to a magic-sized $\{Bi_{38}O_{45}\}$ cluster.

be described with the $\{Bi_{38}O_{45}\}$ cluster model, which is identified by the poor fits obtained. The full analysis can be read in Supplementary S3.1.4.

Ex situ measurements of samples prepared at room temperature
5 were done to characterize the reaction after 2 days, 4 days, 11 days and 100 days, which are illustrated in Figure 7. The structures of all of these samples correspond to $\{Bi_{38}O_{45}\}$ (Supplementary S3.1.4). Therefore, the equilibrium of the reaction seems to go towards $\{Bi_{38}O_{45}\}$ but is both temperature dependent and time-
10 dependent. At 80°C, the reaction goes to equilibrium in minutes, where it takes days at room temperature.

Analysis of intermediate structures

As presented above, it has been established that the first 7 minutes of the *in situ* PDF measurement at 80°C in Figure 3, left,
15 corresponds to the crystalline phase of $[Bi_6O_5(OH)_3(NO_3)_5]\cdot(H_2O)_3$ and the frames after 40 minutes corresponds to the $\{Bi_{38}O_{45}\}$ cluster. Figure 8 illustrates the reaction pathway, in which the intermediate structure is still to be elucidated. This is supported by Figure 9, which shows that the
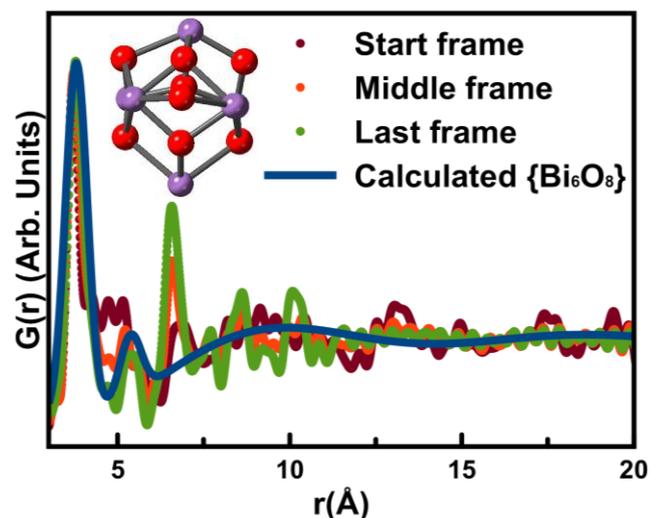20 middle frame and last frame are not the same.

Figure 9: Data from frames through the reaction compared to the calculated structure of the octahedral $\{Bi_6O_8\}$ unit, which shows that the octahedral $\{Bi_6O_8\}$ unit is not stable in solution. Inset: The structure of the octahedral $\{Bi_6O_8\}$ unit.

---

[1] Troels Lindahl Christiansen found that the octahedral building block octahedral $\{M_6O_X\}$ was never stable in solution.
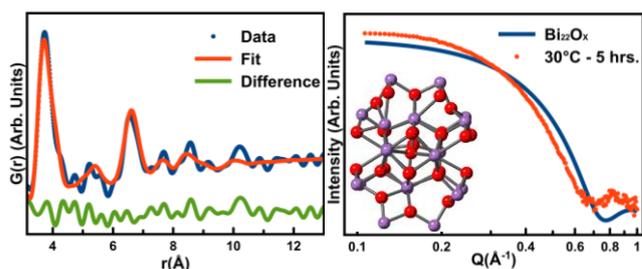

Figure 10: The frames corresponding to the intermediate structure in the *in situ* PDF measurement is not described well by the $\{Bi_{22}O_X\}$ structure. Left) The calculated structure of $\{Bi_{22}O_X\}$ with the Debye Equation fit the data of the frames after 3 hours after the beginning of the experiment of the 30°C *in situ* measurement partially, right) The calculated structure of $\{Bi_{22}O_X\}$ with the Debye Equation fitted to the SAXS data from the sample prepared at 30°C after 5 hours. Inset) The structure of $\{Bi_{22}O_X\}$. Refined parameters can be seen in Supplementary S3.1.4.

But from the *in situ* PDF measurement in Figure 3, left, it seems like the crystalline phase dissolves into an intermediate, before the formation of $\{Bi_{38}O_{45}\}$. When the long-range order disappears after 7 minutes, the peaks at r = 3.8 Å, r = 6.3 Å, r = 8.2 Å and r =
25 10.2 Å increases in intensity until it is similar to the $\{Bi_{38}O_{45}\}$ cluster. This must represent the intermediate, which is assumed to be the same structure in all experiments despite the varying temperature (Figure 6 + Figure 17).

Since the octahedral building block, $\{Bi_6O_8\}$, is referred to as
30 particularly stable in the literature[6-8], it is expected that the crystals will dissolve into clusters of building blocks with nitrate functioning as ligands. By comparing the calculated PDF of the octahedral building block $\{Bi_6O_8\}$ with an early frame, a middle frame and the last frame of the 80°C *in situ* PDF data, it can be
35 seen that $\{Bi_6O_8\}$ is never stable in solution[1] (Figure 9), since the structure of the presented data contains distances significantly longer than the octahedral $\{Bi_6O_8\}$ does.

From Figure 9 it can clearly be seen that the middle frame and last frame are different, especially in the peak at r = 6.5 Å, which is an
40 intermolecular distance between two $\{Bi_6O_8\}$ building blocks. Therefore, the intermediate must be a smaller cluster than the product. This agrees with what is chemically expected as we have a cluster growth proceeding in the reaction. Cluster Growth is furthermore verified through analysis of the SAXS form factor in
45 Figure 5 since the particles prepared at 30°C and measured after 5 hours has a radius of 6.25 Å, which is smaller than the $\{Bi_{38}O_{45}\}$ cluster with a radius of 7.25 Å.

While the octahedral building block, $\{Bi_6O_8\}$, is not seen stable in solution, it is instead utilized to build an intermediate structure. A
50 similar structure composed of 22 bismuth atoms as the one presented in Figure 2, left, has been identified with single crystal diffraction[29]. Additionally, based on mass spectroscopy, a cluster of 22 bismuth atoms has been observed from a similar reaction[13]. Therefore, a $\{Bi_{22}O_X\}$ structure was built simply by putting the
55 $\{Bi_6O_8\}$ units together, which can be seen in Figure 2, left. In Figure 2, it is seen that the $\{Bi_{22}O_X\}$ structure is similar to the $\{Bi_{38}O_{45}\}$ structure, both built by the octahedral building blocks,
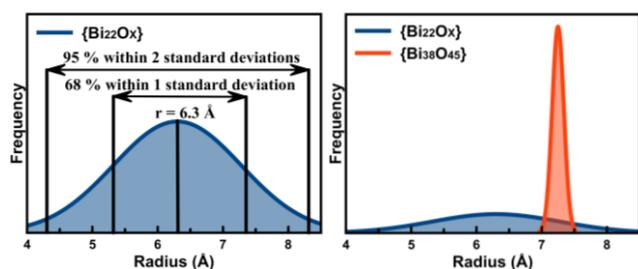
Figure 11: The size distributions of the clusters found with SAXS form factor analysis of the SAXS data of the sample prepared at 30°C and measured after 5 hours after the beginning of the experiment and the sample prepared at 40°C. Left) The size distribution of the intermediate, {Bi$_{22}$O$_X$} cluster. Right) The comparison of the size distributions of the intermediate, {Bi$_{22}$O$_X$}, and the product, {Bi$_{38}$O$_{45}$}.

{Bi$_6$O$_8$}, but the two clusters are different in size. Afterwards, the constructed {Bi$_{22}$O$_X$} cluster was fitted to the data from the intermediate. However, as observed in Figure 10, the data is only partially described, with good agreement between the peaks at r = 3.8 Å and r = 6.3 Å, while the smaller peaks at r = 8.2 Å and r = 10.2 Å cannot be fully described with the current model.

Apart from the cluster growth indications provided through the SAXS measurements, the measurements furthermore displayed a larger size distribution (0.94 Å) than in the final product as illustrated in Figure 5. Eventually, this indicates that the intermediate is disordered. The SAXS form factor analysis has been done by fitting a spherical SAXS form factor with a Gaussian curve to the data. Thereby, a size distribution of clusters can be illustrated as seen in Figure 11. Here the frequency of clusters with radius "R" is plotted versus the radius "R". In Figure 11, left, it can be seen that 68 % of the clusters in the intermediate has a radius in range R = 5.3 Å – 7.3 Å, while this is compared to the {Bi$_{38}$O$_{45}$} cluster in Figure 11, right, where 68 % of the clusters has a radius of R = 7.16 Å – 7.34 Å.

These results indicate that the reaction goes through a spherical polydisperse intermediate before growing to the spherical magic-sized {Bi$_{38}$O$_{45}$} cluster (further analysis and results of the SAXS data can be found in Supplementary S3.2.1).

Based upon the above-stated results, it has been shown that all the clusters are build up by the octahedral building block, {Bi$_6$O$_8$}, and that the product is the well-defined structure of {Bi$_{38}$O$_{45}$}.
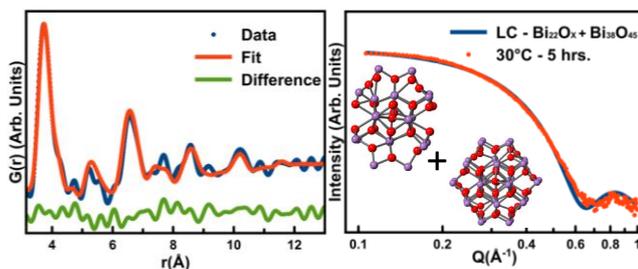

Figure 12: Data corresponding to the intermediate in the *in situ* PDF measurement is described well by a linear combination of the {Bi$_{38}$O$_{45}$} cluster and the {Bi$_{22}$O$_X$} cluster. Left) The calculated structure of a linear combination of the calculated Debye scattering of the {Bi$_{38}$O$_{45}$} and the {Bi$_{22}$O$_X$} describes the data of the frames after 3 hours from the beginning of the experiment of the 30°C *in situ* PDF measurement well, right) The calculated structure of a linear combination of the calculated Debye scattering of the {Bi$_{38}$O$_{45}$} and the {Bi$_{22}$O$_X$} describes the SAXS data from the sample prepared at 30°C after 5 hours well. Inset) The linear combination of the {Bi$_{22}$O$_X$} and the {Bi$_{38}$O$_{45}$} cluster.
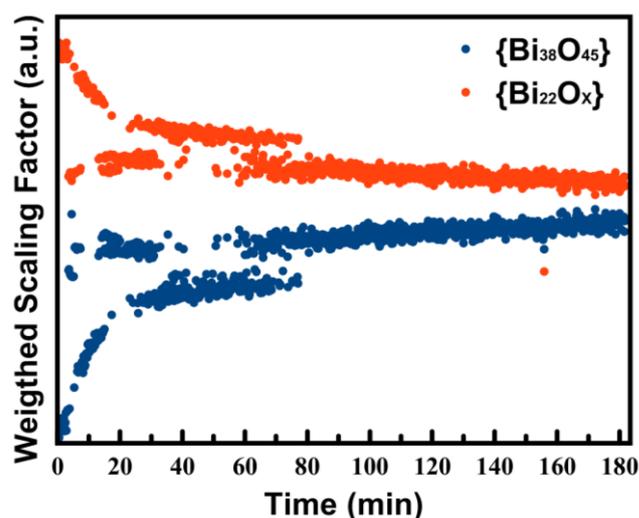

Figure 13: The scaling factor versus time of the reaction from the beginning of the experiment is plotted from the 30°C *in situ* PDF measurement, which shows that the {Bi$_{22}$O$_X$} cluster is growing to the {Bi$_{38}$O$_{45}$} cluster.

On the search for a complete solution; introducing a multi-phase refinement

From the previous investigations, it was found that single-phase refinement with either the {Bi$_{22}$O$_X$}- or {Bi$_{38}$O$_{45}$} cluster as the model did provide a description of some structural features, but still with severe discrepancies between model and experimental data. Therefore, it was motivated to investigate if the reaction instead should be considered as consisting of two phases, the {Bi$_{22}$O$_X$}- and {Bi$_{38}$O$_{45}$} cluster. Consequently, a two-phase refinement was performed on both the PDF and SAXS data. When a linear combination of the calculated Debye scattering from the {Bi$_{22}$O$_X$} cluster and the {Bi$_{38}$O$_{45}$} cluster is fitted to the 30°C *in situ* PDF data after 3 hours, and subsequently to the SAXS data which were prepared at low temperature, it can be seen that they describe the data very well (Figure 12). The SAXS form factor analysis can equally well be described by a two-phase refinement of magic-sized structures, with similar dimensions as the {Bi$_{22}$O$_X$} cluster and the {Bi$_{38}$O$_{45}$} cluster, as a polydisperse intermediate (Further details are available in Supplementary S3.2.1). Sequential refinement was done on the *in situ* PDF data, where the weighted scaling factor of the individual clusters can be seen as an indicator of how much the cluster is present in the sample (Further details in Supplementary S3.4). In Figure 13, it can be seen that the weighted scaling factors of the two clusters are changing through the reaction. The reaction can, in the beginning, be described as a large fraction of the {Bi$_{22}$O$_X$} cluster and small amounts of {Bi$_{38}$O$_{45}$}, but with time the {Bi$_{38}$O$_{45}$} cluster becomes more dominant. In Figure 14 the ratio between the two clusters is plotted versus the time after the beginning of the experiment. High values mean a high ratio of the {Bi$_{22}$O$_X$} cluster, and low values mean a high ratio of the {Bi$_{38}$O$_{45}$} cluster. It can be seen that the reaction can be divided into two regions; at the beginning of the reaction, the cluster ratio rapidly changes, which is followed by a rather constant change in cluster ratio throughout the reaction. Furthermore, as illustrated in Figure 14, the temperature is also observed to have an effect on the reaction rate and therefore dictates how rapid the reaction is growing from the intermediate to the {Bi$_{38}$O$_{45}$} cluster.
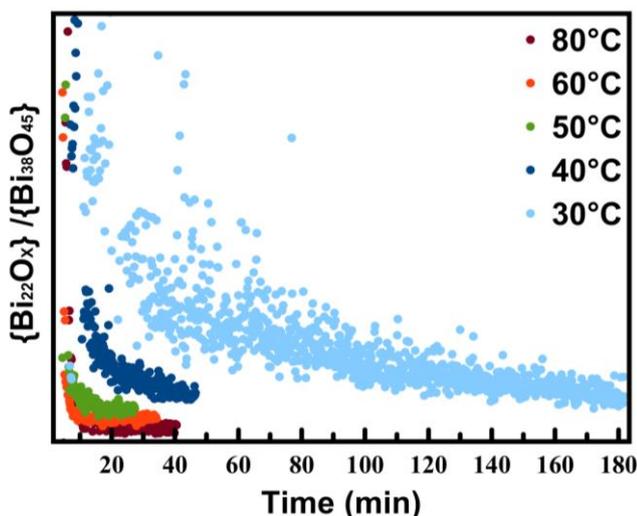
Figure 14: The ratio of {$Bi_{22}O_X$} cluster and {$Bi_{38}O_{45}$} cluster is plotted versus time of the reaction. It can be seen that the intermediate is growing until it becomes the {$Bi_{38}O_{45}$} cluster. The rate of ratio change is highly temperature- and time-dependent.

How fast the conversion occurs between the {$Bi_{22}O_X$}- and {$Bi_{38}O_{45}$} cluster has been further investigated through *ex situ* PDF measurements.

The data were measured at 30°C after 2 days, 4 days, 11 days and 100 days. By introducing the two-phase model for refinement, the resulting fit was significantly improved (Figure 15).

Furthermore, it can be seen that the ratio between the {$Bi_{22}O_X$} - and {$Bi_{38}O_{45}$} cluster is lower than for the high temperature *in situ* PDF measurements for 40 minutes (Supplementary S3.4), which indicates that both the time dependency and temperature dependency have a significant effect.

This conventional modelling technique, which has so far been introduced, provides a reasonable idea of the overall cluster growth process, where the intermediate is growing until it becomes the {$Bi_{38}O_{45}$} cluster. However, limited insight can be achieved, when considering how the cluster growth is happening. To gain such information, a new approach must be introduced.

Predicting the average structure through computational permutations

A new way of modelling the PDF data is introduced in order to characterize the average structure through the reaction. All possible models, with the same motif as the {$Bi_{22}O_X$}- and the {$Bi_{38}O_{45}$} cluster must be tried in order to eventually obtain the best possible structure (Supplementary S3.3 Figure 37).

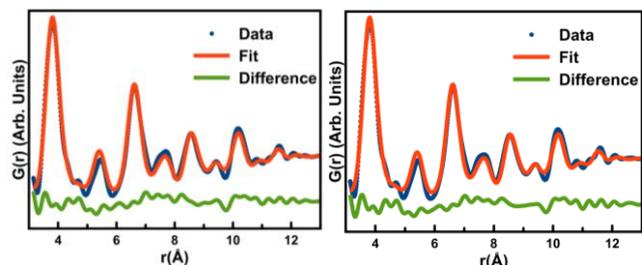The procedure of this approach is initialized by fitting a structure

Figure 15: The linear combination of the {$Bi_{22}O_X$} cluster and {$Bi_{38}O_{45}$} cluster describes the *ex situ* data of the samples well after – left) 2 days, right) 100 days.
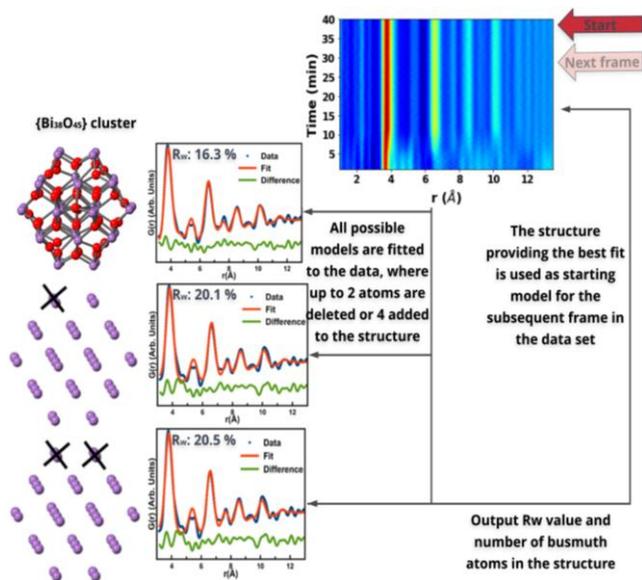
Figure 16: The permutations method follows the steps:
1. Fit the data of the last frame to all permutated structures of the {$Bi_{38}O_{45}$} cluster.
2. Output the structure which describes the last frame best.
3. Go 1 frame backwards and fit the data to all permutated structures derived from the outputted structure.
4. Output the structure which describes the data of the second to last frame best.

to a frame of the *in situ* PDF measurement. Proceeding from this, atoms are added or deleted to the structure individually in order to find a structure, which describes the data better. When the best possible structure is found, this is used as a starting point for the next frame, where atoms are deleted or added again. (Full description of this modelling method is available in Supplementary 2.4 & 3.3).

Be aware that the method can only be used when the structural motif is the same through the reaction, in which only the size of the cluster may vary.

The reaction can be followed by starting from the product and using this method backwards in the reaction by deleting atoms or adding atoms to the structure in order to find the structure, which describes the data best. This method is illustrated by a flowchart in Figure 16.

To avoid the unphysical values, the method was restricted. The cluster cannot contain more than 38 bismuth atoms and it cannot be smaller than the octahedral building block. Only one larger cluster is reported in the literature [$Bi_{50}Na_2O_{64}(OH)_2(OSiMe)_3)_{22}$][30], which does not fit the data well (Supplementary S3.5). The correlated motion, delta2, was restrained to between 0 and 7, the zoomscale between 0.9 and 1.1 and the atomic displacement parameter should be between 0 and 3. Permutations are done by adding 2 atoms per frame, which were the most inner atoms of the ones missing and deleting up to 6 atoms, which were always from the outer sphere of the particle. Ideally, all 38 bismuth atoms had to be included in the permutations, but this gives 2^{38} structures that have to be fitted to the data. Due to limited computational power, this is not possible. By experience, it was seen that 6 atoms are sufficient to follow the reaction process when 1 frame corresponds to 4 seconds.

Figure 17 shows the results of the time-resolved permutation. The

Figure 17: The reaction goes through several intermediates, which's stability is dependent of the temperature, before ending as the $\{Bi_{38}O_{45}\}$ cluster. Graph-description: The number of bismuth atoms in the best possible structure of the frame is plotted versus reaction time.

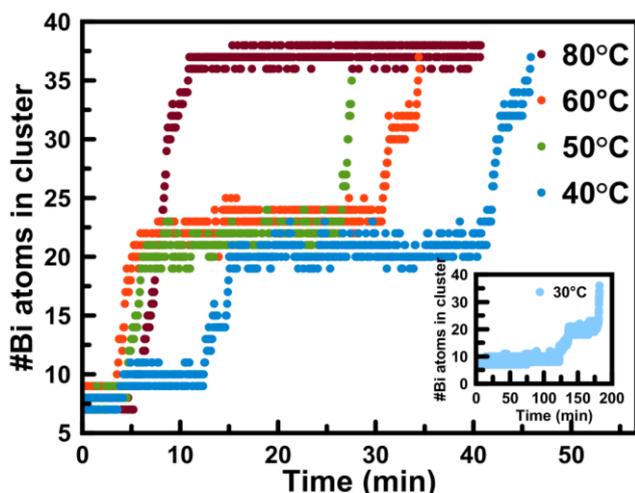number of bismuth atoms is plotted versus the time of the reaction; thus a reaction process of the average structure in the sample is shown. It is evident that the reaction goes through multiple intermediates.

5 From Figure 17, it is seen that an intermediate of 19-23 bismuth atoms appears, which upon both lower temperatures and longer reaction-time, grows to an intermediate of 30-34 bismuth atoms before the stable $\{Bi_{38}O_{45}\}$ structure is seen. Both intermediates are only stable with low temperature and the $\{Bi_{38}O_{45}\}$ cluster is 10 only seen at 80°C. Furthermore, the 30-34 bismuth intermediate is only seen at measurements longer than 30 minutes (Figure 17). In order to confirm the results, the residuals and the physical structure of every frame can be shown by this method. In Figure 18, left, the residuals are shown as function of the number of bismuth atoms in 15 the structure. It can be seen that all intermediate states before the 19-23 atom intermediate have a high R_w value. In Figure 18, right, a selected structure with less than 19 bismuth atoms is illustrated. This illustrates that when the structures have less than 19 atoms, they are seen to be unphysical and fit the data with high R_w values, 20 which probably are caused by some crystalline material in the sample.

In Figure 19, left, all structures larger than 19 bismuth atoms are plotted as function of how frequent they appear. In Figure 19, right, the most frequent structures through the reaction are shown. They
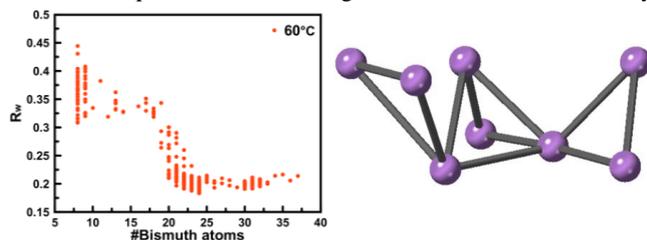

Figure 18: Left) The R_W values plotted against the number of bismuth atoms in the best possible structure of the frame. Right) One of the structures with high R_W value, which shows that these frames in the beginning of the reaction does not give a physical output.
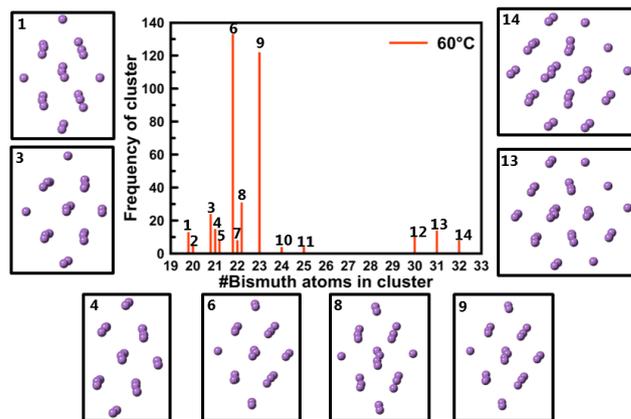
Figure 19: Left) The frequency of the clusters are plotted versus how many bismuth atoms they contain. Boxes) Illustrates which structures the reaction goes through.

25 all seem physical and consisting of the octahedral building block, $\{Bi_6O_8\}$.

For the 60°C *in situ* PDF measurement, the intermediate structures that were made this way by deleting atoms from the $\{Bi_{38}O_{45}\}$ structure, were investigated. 14 different structures were found 30 between the crystalline phase and the $\{Bi_{38}O_{45}\}$ structure (Figure 19), where the most frequent structures are denoted as number 6 and 9 (Further analysis of the intermediates of the 30°C *in situ* PDF measurement are available in Supplementary S3.3.1.1). It is also showed that the PDF's cannot distinguish all these structures, but 35 they are all physical following the building block principle (Supplementary S3.3.1 figure 38).

It can be seen that characterization of the cluster formation is very complex. Traditional refinements only show that the cluster is growing, but does not tell anything about how it is growing. 40 Therefore, a new modelling technique was designed, where the structure was permuted. This means that the model is chosen relative to the earlier frame. With this technique, the average cluster growth process has been characterized. Both techniques are essential in order to understand the reaction process (Further 45 comparison of the two techniques are available in Supplementary S3.8).

Exchange of ligands from $[Bi_{38}O_{45}(NO_3)_{24}(DMSO)_{25}]$ to $[Bi_{38}O_{45}(OMc)_{24}(DMSO)_{25}]$

In order to control the synthesis, one must not only be able to 50 control the synthesis with nitrate as a ligand, but with all kind of ligands. To characterize the effect of the ligand, a replacement of the nitrate ligand with $NaOOC(CH_2)CH_3$ was measured *in situ* (Figure 20) at 30°C.

It is evident that changing the ligand directly changes the structure 55 of the bismuth clusters in the reaction[2]. Comparing the first frame after dissolution and the last frame of the 30°C *in situ* ligand exchange PDF measurement with the $\{Bi_{38}O_{45}\}$ cluster reveals that the model describes the data reasonably just after dissolution, which is 5 minutes after the beginning of the experiment. However, 60 following the introduction of ligand exchange, the fit is significantly improved, showing good agreement between data and

---

[2] I want to credit Troels Lindahl Christiansen for making this observation

Figure 20: The 30°C *in situ* PDF measurement of the ligand exchange experiment clearly indicates a structure change during the reaction.

Figure 22: The $[Bi_6O_5(OH)_3(NO_3)_5] \cdot (H_2O)_3$ crystals dissolved in DMSO have shown to go through an $\{Bi_{22}O_X\}$ cluster as intermediate before growing to a magic-sized $\{Bi_{38}O_{45}\}$ cluster.
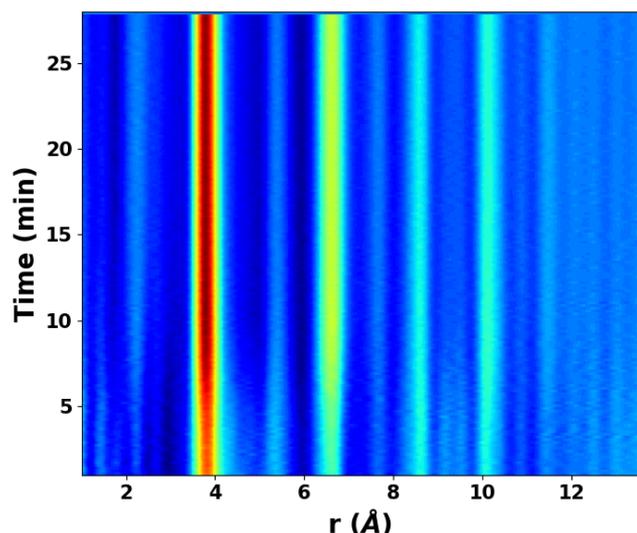
the utilized $\{Bi_{38}O_{45}\}$ cluster model. Consequently, this indicates that $NaOOC(CH_2)CH_3$ is found to stabilize the core built by the building blocks (Figure 21) (Further details is available in Supplementary S3.7).

5 However, the reasoning of why this ligand has a stabilizing effect is yet unknown, but interesting it is that the ligand has such a great effect on the structure of the clusters. Additionally, for further studies, it would be interesting to use $NaOOC(CH_2)CH_3$ in the cluster growth process to investigate if the intermediates could be

10 further stabilized.

The full analysis is available in Supplementary S3 with a description of the individual contributions from Troels Lindahl Christiansen, Martin Schmiele and Andy Sode Anker.

Conclusion

15 The experiments reported here demonstrate how challenging it is to characterize clusters directly in solution. Meanwhile, it has been demonstrated how powerful a tool the Debye Equation is, and how it can be used with both PDF and SAXS to characterize complex systems of multiple phases and size distributions of structures. The

20 combination of PDF and SAXS has shown its large potential in material chemistry of nanomaterials.

In the presented study, crystalline $[Bi_6O_5(OH)_3(NO_3)_5] \cdot (H_2O)_3$ was dissolved in DMSO, in which it was found that the material directly forms into a bismuth oxido cluster containing about 22

Figure 21: Left: The first frame after dissolution of the *in situ* PDF measurement of the ligand exchange fitted to the $\{Bi_{38}O_{45}\}$ cluster. Right: The last frame, which corresponds to 29 minutes after beginning of the experiment, fitted to the $\{Bi_{38}O_{45}\}$ cluster.

25 bismuth atoms, but not through a stable octahedral $\{Bi_6O_8\}$ unit as assumed in earlier studies[13]. The product was characterized as the atomically stable $\{Bi_{38}O_{45}\}$ compound, which could be further stabilized by varying the ligand. It is the first time that it has been quantified that the $\{Bi_{38}O_{45}\}$ structure is magic-sized in solution.

30 In order to characterize the intermediate two modelling approaches were used. The traditional PDF methods were used with a two-phase refinement to follow the reaction process from the $\{Bi_{22}O_X\}$ cluster to the $\{Bi_{38}O_{45}\}$ cluster. This method showed that the intermediate was found to grow until it reached the $\{Bi_{38}O_{45}\}$

35 cluster. However, the current method does not provide any additional information regarding the unique structure(s) through the reaction.

Therefore, a new method was designed in order to characterize the average structure in the sample through the reaction. This method

40 showed that the average structure of the cluster grows step-wise, and goes through two stable intermediates before growing to the $\{Bi_{38}O_{45}\}$ cluster. The general reaction is illustrated in Figure 22, where it can be seen that the $[Bi_6O_5(OH)_3(NO_3)_5] \cdot (H_2O)_3$ crystal dissolves directly into an intermediate, which grows until the

45 $\{Bi_{38}O_{45}\}$ cluster is reached.

The reaction rate was varied by varying the temperature. At room temperature, the reaction took several days, while at 80°C the reaction took about an hour.

This project is an outstanding improvement of the understanding

50 of "real materials". We have shown that bismuth oxido clusters in solution are affected by temperature and ligands. They are not simply found in one monodisperse phase, but a mixture of phases where the ratio between the phases is constantly changing. Additionally, the phases may exist of size distributions of clusters,

55 where we need new techniques to characterize a unique structure through the reaction.

Further work will be done to combine PDF and SAXS in order to unravel the full potential of these techniques. *In situ* SAXS measurements are currently in progress, which would greatly

60 contribute to the understanding of the cluster growth of bismuth oxido clusters, and other similar systems as uranium-, cerium- and plutonium oxido clusters.

65

Acknowledgements

Notes and references

*Address: Nano-Science Center & Department of Chemistry, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen East, Denmark.; *E-mail: Andy@nano.ku.dk*

1. L. Miersch, T. Ruffer, M. Schlesinger, H. Lang and M. Mehring, *Inorg Chem*, 2012, **51**, 9376-9384.
2. F. Y. Du, J. M. Lou, R. Jiang, Z. Z. Fang, X. F. Zhao, Y. Y. Niu, S. Q. Zou, M. M. Zhang, A. H. Gong and C. Y. Wu, *Int J Nanomed*, 2017, **12**, 5973-5992.
3. L. Miersch, T. Ruffer and M. Mehring, *Chem Commun (Camb)*, 2011, **47**, 6353-6355.
4. M. Schlesinger, S. Schulze, M. Hietschold and M. Mehring, *Dalton T*, 2013, **42**, 1047-1056.
5. L. Miersch, T. Rüffer, H. Lang, S. Schulze, M. Hietschold, D. Zahn and M. Mehring, *European Journal of Inorganic Chemistry*, 2010, **2010**, 4763-4769.
6. C. Falaise, C. Volkringer, J. F. Vigier, A. Beaurain, P. Roussel, P. Rabu and T. Loiseau, *Journal of the American Chemical Society*, 2013, **135**, 15678-15681.
7. I. L. Malaestean, A. Ellern, S. Baca and P. Kogerler, *Chem Commun (Camb)*, 2012, **48**, 1499-1501.
8. M. Mehring, in *Clusters – Contemporary Insight in Structure and Bonding*, ed. S. Dehnen, Springer International Publishing, Cham, 2017, DOI: 10.1007/430_2016_4, pp. 201-268.
9. C. Falaise, C. Volkringer, J. F. Vigier, A. Beaurain, P. Roussel, P. Rabu and T. Loiseau, *J Am Chem Soc*, 2013, **135**, 15678-15681.
10. K. J. Mitchell, K. A. Abboud and G. Christou, *Nat Commun*, 2017, **8**, 1445.
11. L. Soderholm, P. M. Almond, S. Skanthakumar, R. E. Wilson and P. C. Burns, *Angew Chem Int Ed Engl*, 2008, **47**, 298-302.
12. K. M. O. Jensen, P. Juhas, M. A. Tofanelli, C. L. Heinecke, G. Vaughan, C. J. Ackerson and S. J. L. Billinge, *Nature Communications*, 2016, **7**.
13. D. Sattler, M. Schlesinger, M. Mehring and C. A. Schalley, *ChemPlusChem*, 2013, **78**, 1005-1014.
14. S. J. L. Billinge and M. G. Kanatzidis, *Chem Commun*, 2004, DOI: 10.1039/b309577k, 749-760.
15. J. K. M. Ø., T. Christoffer, B. Martin and I. B. B., *ChemSusChem*, 2014, **7**, 1594-1611.
16. D. Prill, P. Juhas, S. J. L. Billinge and M. U. Schmidt, *Acta Crystallographica Section A*, 2016, **72**, 62-72.
17. D. Prill, P. Juhas, M. U. Schmidt and S. J. L. Billinge, *Journal of Applied Crystallography*, 2015, **48**, 171-178.
18. T. Li, A. J. Senesi and B. Lee, *Chem Rev*, 2016, **116**, 11128-11180.
19. P. Juhas, C. L. Farrow, X. Yang, K. R. Knox and S. J. Billinge, *Acta Crystallogr A Found Adv*, 2015, **71**, 562-568.
20. C. L. Farrow and S. J. L. Billinge, *Acta Crystallographica Section A*, 2009, **65**, 232-239.
21. D. P., *Annalen der Physik*, 1915, **351**, 809-823.
22. P. Scardi, S. J. L. Billinge, R. Neder and A. Cervellino, *Acta Crystallographica Section A*, 2016, **72**, 589-590.
23. in *Encyclopedia of the Sciences of Learning*, ed. N. M. Seel, Springer US, Boston, MA, 2012, DOI: 10.1007/978-1-4419-1428-6_2276, pp. 2332-2332.
24. C. L. Farrow, P. Juhas, J. W. Liu, D. Bryndin, E. S. Božin, J. Bloch, P. Th and S. J. L. Billinge, *Journal of Physics: Condensed Matter*, 2007, **19**, 335219.
25. P. Juhas, T. Davis, C. L. Farrow and S. J. L. Billinge, *Journal of Applied Crystallography*, 2013, **46**, 560-566.
26. P. J. Xiaohao Yang, Christopher L. Farrow, Simon J. L. Billinge, *xPDFsuite: an end-to-end software solution for high throughput pair distribution function transformation, visualization and analysis, arXiv:1402.3163*.
27. I. Bressler, J. Kohlbrecher and A. F. Thunemann, *Journal of Applied Crystallography*, 2015, **48**, 1587-1598.
28. L. F., *Acta Crystallographica Section B*, 1978, **34**, 3169-3173.
29. M. Dirk, M. Michael and S. Markus, *Angewandte Chemie International Edition*, 2005, **44**, 245-249.
30. M. Michael, M. Dirk, P. Sanna and S. Markus, *Chemistry – A European Journal*, 2006, **12**, 1767-1781.

Fabrication and measurement of hybrid quantum dot devices featuring Yu-Shiba-Rusinov sub-gap states

Jacob Ovesen, Kasper Grove-Rasmussen and Jesper Nygård

The aim of this bachelor project is to fabricate and study hybrid quantum dot (QD) devices featuring Yu-Shiba-Rusinov (YSR) sub-gap states, made on special InAs nanowires with 7 nm epitaxial aluminum coating on 3 facets. The advantage with these nanowires is the epitaxial aluminum, which can be etched away in a small window of semiconducting nanowire for creating QDs. These dots can be used for measuring the sub-gap states in superconducting materials due to the high degree of tunability. A device with a single QD featuring YSR sub-gap states was fabricated and analyzed, and a double QD system was afterwards made from the device using superconducting leads in the normal state, but not yet featuring YSR sub-gap states. The device does show the possibility for making an superconductor-double dot-superconductor device featuring YSR sub-gap states.

1 Introduction

Quantum dots (QDs) are sub-micron structures, confined in all dimensions, forming an island with a discrete energy spectrum working as an artificial atom[1][2]. If made from a semiconductor, the island has most of its electrons bound, except for a small number of free electrons, which can be added or removed from the dot[1]. By taking advantage of Coulomb repulsion, an electron can be added to the quantum dot only if sufficient energy is provided to the electron, giving single electron transport. This provides the opportunity for studying energy spectrum for ad-vanced electronic systems, for instance with superconducting electrodes (S), also known as hybrid-dot systems. By taking advantage of the single-electron control from a quantum dot, macroscopic quantum phenomenon in superconductors, involv-ing large electron numbers, can be studied[3], such as Yu-Shiba-Rusinov sub-gap states (YSR)[4].

YSR Aub-gap Atates Are A Ay-brid Ahenomenon Aappening in A-QD Aystems Aith Aeak Aou-pling Aetween Ahe electrodes. Ahe Atates Are Aeen Aithin Ahe Auperconducting gap Δ[5].

In this project, the focus is a fabrication and analysis of a double quantum dot system with superconducting leads of aluminum, to which the purpose is to show Yu-Shiba-Rusinov (YSR) sub-gap states. The quantum dots are based upon special InAs nanowires[6] with 7 nm of epitaxial aluminum coating on 3 of the facets, allowing them superconducting properties. This coating was then etched away in a region, after which the two quantum dots are defined, giving single electron transport at low temperatures.
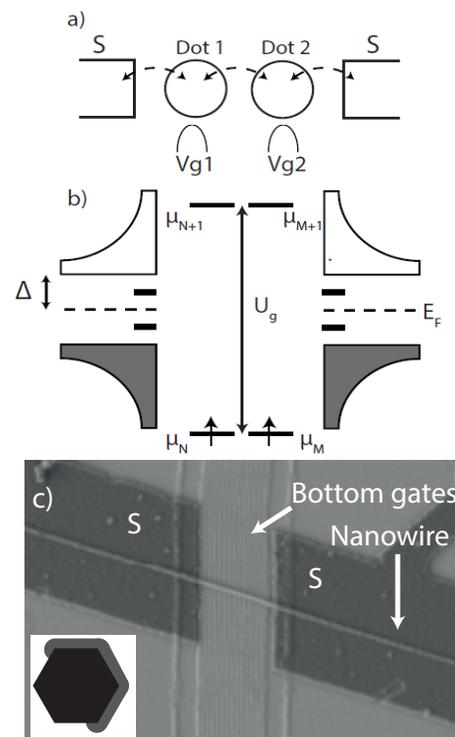
Fig. 1 a) Schematic view of a S-double dot-S system, showing tunneling behavior with a gate potential Vg coupled electrostatically to each dot. b) Schematic energy diagram of system showing S-state density of states on each side of the two dots, here just showing a single level. A single spin is added to each dot in order to show the ideal configuration for observing YSR sub-gap states. The next available states on the dots are also shown above Δ to show the condition $U_g \gg \Delta$. c) An SEM-image of system showing bottom gates and two Al-leads. Inset schematic: A cross-section of the InAs nanowire showing the epitaxial aluminum on three of the facets.

As shown in Fig. 1a the quantum dots will be in series and weakly coupled, creating a double-dot potential, where each dot has a plunger gate. The leads to the dots are superconducting, and will therefore have a density of states with a superconducting gap of width 2Δ[7], shown i Fig. 1b. The figure also shows that the quantum dots have a charging energy $U_g \gg \Delta$, and they have a single spin on each dot, which is needed to see YSR sub-gap states[5]. To avoid tunneling across the double dots, the thermal energy has to be much smaller than the charging energy, $U \gg k_B T$.

A SEM image of the finished device is shown in Fig. 1c, where both the bottom gates and the nanowire with the aluminum leads can be seen.

2 Fabrication

The devices were fabricated on Silicon wafers with 200nm SiO_2[a]. The bonding pads shown in Fig. 2a were the first structure to be made by optical lithography using a Heidelberg LED-writer and evaporation of 5 nm titanium (Ti) and 100 nm gold (Au). The 5 nm Ti is used for sticking layer. Afterwards the alignment marks, also shown in Fig. 2a, were made by electron beam lithography, also with 5 nm Ti and 100 nm Au. The next step was the bottom gates, which were made by electron beam lithography and evaporation of 5 nm Ti and 25 nm Au in the middle of the bonding pads pattern. The bottom gates were made in a grid of 8x4 sets of bottom gates. Each set has four bottom gate units, each unit with 17 bottom gates. They are 20 nm wide and spaced 40 nm apart, which gives a full bottom gate unit width of around a micron. An 20 nm oxide layer of HfO_2 was then applied on top of the bottom gates for both protection and for making sure they are gate coupled to the device only. The oxide was made by optical lithography and Atomic Layer Deposition (ALD). This was done in several steps in order to avoid the fencing effect[8] formed by the ALD (see Fig. 9 in supplementary for example). The device was also developed in MF321 with stirring to make a undercut, and after ALD, the device was left in N-Methyl-2-pyrrolidone (NMP) thermalized at $85°C$ for an hour for better lift-off. An SEM image showing 3 sets of 17 bottom gates with as oxide layer is shown in Fig. 2c.

Afterwards, the InAs nanowire with epitaxial aluminum was deposited on top and across the bottom gates. The nanowires epitaxial aluminum are 7 nm thick, and were grown by P. Krogstrup at NBI[9]. A window of 380 nm, corresponding to 7 bottom gates, was wet-etched on the nanowire in the middle of the bottom gates, removing the aluminum coating in the re-

gion, leaving only the semiconducting nanowire. Aluminum leads were then deposited by electron beam lithography and metal evaporation of 5 nm Ti and 100 nm Aluminum (Al) on the ends of the nanowire. A schematic of a finished device is shown in figure 2b. The final step was connecting the bonding pads with the leads and bottom gates, which was done by electron beam lithography and metal evaporation of 5 nm Ti and 100 nm Au. An optical image of the contacts is shown in Fig. 2d.
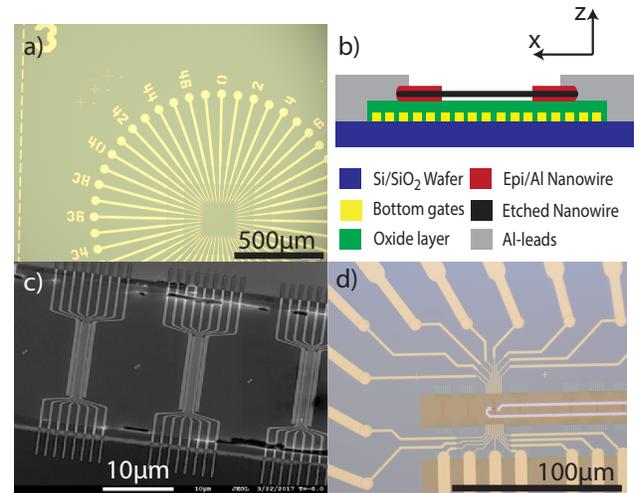
Fig. 2 a) An optical image of bonding pads and alignment marks, b) Schematic of cross-section of a finished device. There are 17 bottom gates for making the double-dot potential and 7 gates under the etch window. Not for scale. c) SEM image of bottom gates (here after oxide layer for protection). d) An optic image of the aluminum leads, the nanowire and the gold the bottom gates.

In Table 1 an overview of the devices fabricated and discontinued is shown. All measurements were made on one of the two devices from batch A2.

Table 1 Fabrication log: Overview of the generations of devices. Each generation included 7 lithography steps (counting oxide layer as one step). A device from batch A2 is the device in focus in this project, but others were also fabricated.

Generation	Working devices	Comment
A1	0	Misalignment between optical and e-beam lithography
A2	2	Same procedure as A1
A3	0	Bad lift-off for bottom-gates, batch was discontinued
A4	0	Last oxide layer went bad, batch was discontinued
A5	1	New design changes for the bottom gates for better gating

3 Measurement and analysis of the device

The setup measures dI/dV_{SD} and V^{AC} via lock-in techniques.[b] A AC signal from Lock-in Amplifier 1 and a DC signal from the DAC is sent through a 10000:1 and 1000:1 voltage divider respectively and sent into the sample. With a current amplifier, the AC signal is measured with Lock-in Amplifier 1 and the DC signal is measured with a Digital Multi-meter 1 (DMM1). The voltage across the sample is amplified with a voltage amplifier and the AC signal is measured with Lock-in Amplifier 2 and the DC signal is measured with Digital Multi-meter 2 (DMM2).

Electron transport measurements were performed on one of the finished devices from batch A2 (see table 1) at low temperatures (T= 30 mK). The setup can be seen in Fig. 3.

The first measurement was to see if the aluminum, both the epitaxial one on the nanowire and the deposited leads, was truly superconducting. Figure 4 shows the differential conductance as a function of bias voltage and magnetic field strength.

Figure 4 shows enhanced conductance around zero magnetic field and zero bias, which arises from Andreev reflections[10]. There is no supercurrent due to an asymmetric coupling to the leads. This indicates a good interface between the superconductor and the normal leads. The enhanced conductance is lost at lower field strength for the perpendicular magnetic field (Fig. 4a) than the magnetic field across the nanowire (Fig. 4b) as expected[7].

Fig. 4 Measurements of the differential conductivity as a function of bias voltage and magnetic field in two situations: a) out of plane for the device (z-direction) and b) in plane of the device (x-direction). See fig. 2b for directions

Fig. 3 Schematic of the measurement setup. Both Lock-in amplifiers are Stanford Research System model SR830. Lock-in Amplifier 1 measures the conductance $\frac{dI}{dV_{SD}}$, while Lock-in Amplifier 2 measures V^{AC}. The boxes labeled DMM are HP digital multi-meters model 34401A. The DAC connected to the bottom gates is the same as the one connected to the voltage divider, and is a Marcus decaDAC.

The critical field strengths for this device were estimated to be $B_C^{\perp} \approx 60$ mT for the B-field out of plane (B_Z) and $B_C^{\parallel} \approx 140$ mT for the B-field in plane (B_X). The superconducting gap was estimated in the open regime to be $\Delta = 200$ μeV $\pm$ 30 μeV. According to former research, Δ has not exceeded 250 μeV (see table 2[c]), which fits well with the observed in Fig. 4.

---

b The division of labor for measurements was shared between myself, Anders Jellinggaard and Morten Hels. Anders and I made the setup, while Morten and I tuned and measured the device.

c It is not specified which angle the critical field for the first reference is measured at, or which device of the five measured in the article[10].

Table 2 Overview of former research papers values for similar devices and nanowires. $B^\perp$ is the z-direction and $B^\parallel$ is the x or y-direction (see directions in Fig. 2).
NB: The dashed line means that the value was not mentioned or not measured.

Number of facets covered	Thickness (nm)	Δ^*	$B_C^\parallel$ or $B_C^\perp$
2	-	220 μeV [10]	$B_C \approx 2.2$ T^c
2	10	190 μeV [9]	$B_C^\perp = 1.9$ T
Full	-	190 μeV [11]	$B_C^\parallel = 75$ mT
Bulk	120	110 μeV [12]	$B_C^\perp = 90$ mT
This work	7	200 μeV	140/60 mT

The critical field strength also fits with some of the former research results. Figure 4 also shows some unexpected transitions (black lines). These lines shows values for B_C and Δ much higher than expected (table 2) and are therefore not understood.

After reassuring that the leads are superconducting, a quantum dot was tuned at zero field to keep the S-leads in the superconducting state. Bias spectroscopy of a QD-S system with a tunnel probe is expected to show YSR sub-gap states, due to alignment of the filled states in the probe with a QD state and the sub-gap states of the S-lead. Such a bias spectroscopy of the device was performed around the single-defined quantum dot and is shown in Fig. 5a. However, the results from Fig. 5a depends on the state of the probe. Figure 5b shows the energy in the N-QD-S system as a function of gate potential (V_{g1}), and shows the even-odd symmetric filling of the Coulomb Diamonds for the YSR sub-gap states. The figure also shows the excitation energy (ζ) for a specific gate potential. The excitation energy is the energy needed for exciting the electron from the ground state to the excited state (here the YSR sub-gap state). The excitation energy for this system can achieve any energy between zero and the superconducting gap Δ, due to the continuum of states in a normal lead. For a S-QD-S system, which is shown in Fig. 5c, the excitation energy can only achieve values between Δ and 2Δ. The reason can be seen in Fig. 5d, which shows the density of states for the two S-leads with a QD in between with asymmetric barriers. To align a filled state from the continuum of states in the probe (left S-lead) with a empty state in the S-lead (right S-lead), the probe's superconducting gap Δ need to be overcome by the applied bias, V_{SD}. At $eV_{SD} = \Delta$, states inside the gap of the S-lead strongly coupled to the QD (right S-lead) are being probed. If $eV_{SD} > 2\Delta$, the empty states over the gap Δ in the S-lead is being probed. So to see transition, the continuum of filled states in the probe has to align with either the empty sub-gap states at $|eV_{SD}| = \Delta + \zeta$ or the empty continuum of states in the S-lead at $|eV_{SD}| > 2\Delta$, as shown in Fig. 5c. Therefore, the YSR sub-gap states for a S-QD-S device is expected to have the gate dependency as illustrated in Fig. 5c.
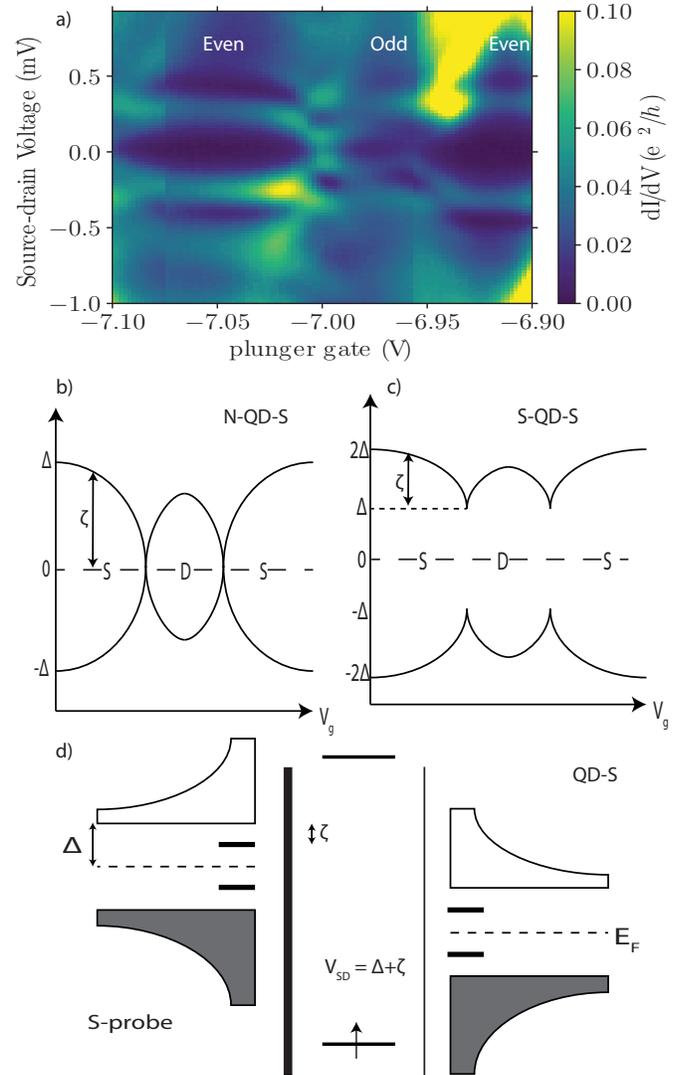
Fig. 5 a) A measurement of differential conductance as a function of gate voltage and bias voltage for a QD-S system with a S-probe at zero field. Clear YSR sub-gap states in Coulomb diamonds are visible. A even-odd symmetry can be seen. The states do not cross at $V_{SD} = 0$, but the gap is not clear. b) Schematic of the ideal QD-S configuration with a N-probe, showing YSR sub-gap states. The electron configuration show even-odd symmetry of singlets and doublets, changing with gate voltage. c) Schematic of the ideal QD-S configuration with a S-probe, showing YSR sub-gap states. The electron configurations show even-odd symmetry of singlets and doublets, changing with gate voltage. d) A schematic energy diagram of the S-QD-S system measured in a) and c) with a applied bias voltage and asymmetric tunneling barriers. The quantum dot shows a level with a single spin, and that the charging energy is much higher than Δ.

The YSR sub-gap states are clearly visible in Fig. 5a,

together with the even-odd symmetry. The YSR sub-gap states follows the illustration in Fig. 5b and 5c well within the Coulomb diamonds. The measurement also gives an unexpected result. The sub-gap states show the even-odd symmetry, but there is no gap at zero bias. Figure 5a shows an intermediate of the two situations illustrated in Fig. 5b and 5c, which makes an extraction of the superconducting gap Δ troubling. The states do not cross each other, since there is no increased conductance at zero bias. The height of the YSR sub-gap states are around 0.26 mV $\pm$ 0.06 mV, so Δ can be either 0.26 meV $\pm$ 0.06 meV for the N-QD-S situation of 0.13 meV $\pm$ 0.03 meV for the S-QD-S situation. For this device, Δ should be the second option, since both leads to the QD are made from aluminum. The asymmetric coupling does that the sub-gap states in the probe does not make transition to the lead, so there should only be transition when the bias exceeds Δ. If the sub-gap states were probing areas of negative differential conductance would be seen in Fig. 5a. Therefore, it is assumed that the probe is superconducting with a soft gap, which means it has a continuum of sub-gap states, thus the bias-symmetric sub-gap states meet at zero bias.

After confirming that the device shows YSR sub-gap states for a single dot and two S-leads, a double dot was the next step to examine. Since it is unclear what one is supposed to see, a magnetic field of 200 mT was applied to the device to make the leads normal conducting for the sake of simplicity. This will make the tuning of the double-dot potential simpler. Figure 6b shows the result of the double dot measurement of differential conductance as a function of the two plunger gate voltages, Vg1 and Vg2. Figure 6a shows the variables obtainable from the double dot charge stability diagram (honeycomb diagram).

The four voltage differences shown in Fig. 6a can be obtained from a honeycomb measurement, and with them, parameters like the capacitances and the charging energy for the system can be calculated. The gate capacitances for the dots can be obtained with Eq. 1, while the inter-dot capacitances is obtained with Eq. 2:

$$C_{g1(2)} = \frac{|e|}{\Delta V_{g1(2)}} \left(1 + \frac{\Delta E}{E_{C1(2)}} \right) \tag{1}$$

$$C_m = \frac{\Delta V_{g1(2)}}{\Delta V_{g1(2)}^m} * C_{2(1)} \left(1 + \frac{\Delta E}{E_{Cm}} \right) \tag{2}$$

Where ΔE is the energy spacing between the levels, E_C is the charging energies for a given dot and $C_{1(2)}$ is the sum of all capacitances attached to dot 1(2):
$C_{1(2)} = C_{L(R)} + C_{g1(2)} + C_m$.
$C_{L(R)}$ is the capacitances to the leads.
For degenerate levels, ΔE is zero. For the sake of simplicity it is assumed that this is the case. From the honeycomb plot (Fig.

6b), the four variables were estimated to: $\Delta V_{g1} = 44$ mV ± 2 mV, $\Delta V_{g2} = 54$ mV ± 2 mV, $\Delta V_{g1}^m = 12.0$ mV ± 0.5 mV and $\Delta V_{g2}^m = 8.0$ mV ± 0.5 mV.
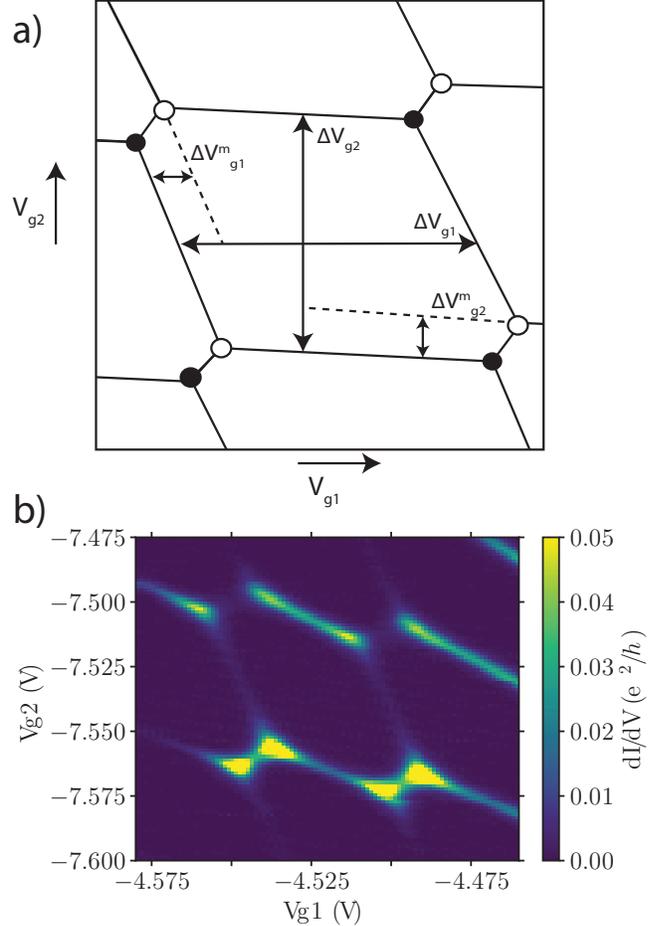
Fig. 6 a) A schematic of the double dot charge stability diagram (also known as honeycomb) with analysis possibilities. The small circles represent the triple point, where three charge states coexists. The filled circle represent electron transfer progress and the hollow circle represent hole transfer progress. b) Measurement of Coulomb honeycomb from a double dot system at $V_{SD}^{DC} = 0$, $B_Z = 200$ mT and with AC voltage with amplitude $V^{AC} = 5$ μV and frequency $f = 127$ Hz. NB: The z-axis is limited to range [0:0.05] to highlight the transition lines. Some transitions e.g. the triple points have higher differential conductivity than 0.05 e^2/h.

From the variables, ΔV_{g1} and ΔV_{g2}, the two gate capacitances were estimated for this measurement to be $C_{g1} = 3.6$ aF ± 0.2 aF and $C_{g2} = 3.0$ aF ± 0.1 aF.
To get the inter-dot capacitance and the capacitances to the leads, a third equation was needed for the lever-arms, which

is Eq. 3:

$$\alpha_{1(2)} \delta V_{g1(2)} = \frac{C_{g1(2)}}{C_{1(2)}} |e| \delta V_{g1(2)} = |eV_{SD}| \quad (3)$$

where δV_{g1} and δV_{g2} can be determined from a Coulomb honeycomb measurement with a finite bias[d]. Such a measurement can be seen on Fig. 7.
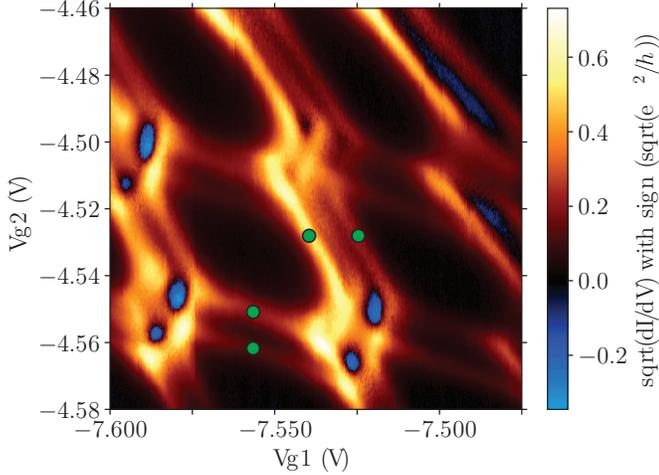
Fig. 7 Measurement of Coulomb honeycomb from a double dot system at $V_{SD}^{DC} = -1$ mV, $B_Z = 200$ mT and with AC voltage with amplitude $V^{AC} = 5$ µV and frequency $f = 127$ Hz. The color scale is made of the square root of difference conductance to highlight the small transitions. The green dots shows there the measurement has been made for determining δV_{g1} and δV_{g2}.

The two parameters δV_{g1} and δV_{g2} were determined to be $\delta V_{g1} = 12.0$ mV ± 0.5 mV and $\delta V_{g2} = 12.0$ mV ± 0.5 mV. By using Eq. 3 and Eq. 2 together, the inter-dot gate capacitance was estimated to $C_m = 6$ aF ± 1 aF and the two lead capacitances were estimated to $C_L = 34$ aF ± 4 aF and $C_R = 14$ aF ± 4 aF. These capacitances can be used to determine charging energies for the three dots, with help from Eq. 4 and Eq. 5:

$$E_{C1(2)} = \frac{e^2}{C_{1(2)}} \left(\frac{1}{1 - \frac{C_m^2}{C_1 C_2}} \right) \quad (4)$$

$$E_{Cm} = \frac{e^2}{C_m} \left(\frac{1}{\frac{C_1 C_2}{C_m^2} - 1} \right) \quad (5)$$

The charging energies were estimated to: $E_{C1} = 3.8$ meV ± 0.5 meV, $E_{C2} = 7$ meV ± 2 meV and $E_{Cm} = 1.0$ meV ± 0.4 meV.

d Here we assumed that the onset of the honeycomb diamond appeared at zero bias even though the superconductors have a gap

From Eq. 3 the two lever-arms α_1 and α_2 were determined to be $\alpha_1 = 0.083 \pm 0.002$ and $\alpha_2 = 0.13 \pm 0.04$ respectively. Table 3 gives a overview of the different parameters measured and calculated throughout this double dot analysis.

Table 3 Overview of values estimated and calculated throughout this paper. The uncertainties on the measured values are estimated from the precision of the peaks positions.

| Measured values:		
ΔV_{g1}	44 mV	± 2.0 mV
ΔV_{g2}	54 mV	± 2.0 mV
ΔV_{g1}^m	12.0 mV	± 0.5 mV
ΔV_{g2}^m	8.0 mV	± 0.5 mV
δV_{g1}	12.0 mV	± 0.5 mV
δV_{g2}	12.0 mV	± 0.5 mV
V_{SD}	-1 mV	
Calculated values:		
C_{g1}	3.6 aF	± 0.2 aF
C_{g2}	3.0 aF	± 0.1 aF
C_L	34 aF	± 4 aF
C_R	14 aF	± 4 aF
C_M	6 aF	± 1 aF
C_1	44 aF	± 5 aF
C_2	24 aF	± 5 aF
E_{C1}	3.8 meV	± 0.5 meV
E_{C2}	7 meV	± 2 meV
E_{CM}	1.0 meV	± 0.4 meV
α_1	0.083	± 0.002
α_2	0.13	± 0.04

As a final note, another measurement of the stability diagram for the double dot system was made with a broader range in gate voltages in order to see the honeycomb pattern, seen in Fig. 8:

The figure shows different sizes of the honeycombs for different number of electrons. This can for instance be seen for honeycomb (2,1), which $Vg2$ range is smaller than for honeycomb (2,3) or which $Vg1$ range is smaller than for honeycomb (1,1). This shows the assumption $\Delta E = 0$ is not met, for either dot. The figure shows that the addition energy for dot 2 is larger than for dot 1.

4 Conclusions

The measurements on the device presented in this paper have revealed two interesting features: Tunability for a double quantum dot and how Yu-Shiba-Rusinov sub-gap states for a single quantum dot evolve with plunger gate voltage. This makes it probable for the device in future studies to combine the two features and show YSR sub-gap states in a double QD configuration with a hybrid quantum dot superconductor-semiconductor nanowire. However, for the present double dot device, we do not see clear Yu-Shiba-Rusinov sub-gap states.
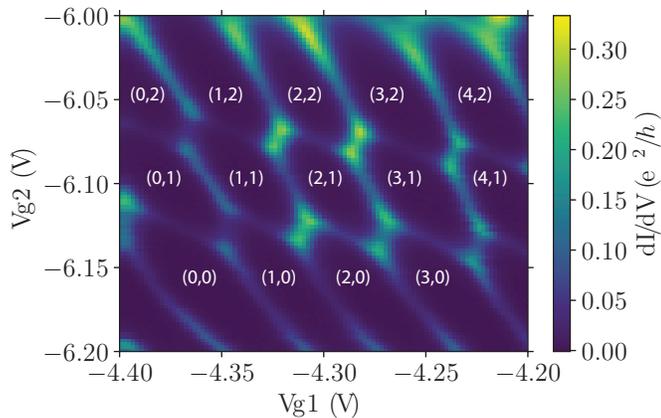
Fig. 8 A measurement of Coulomb honeycomb for the double dot system with a broader gate range for showing several honeycombs. The configuration $(0,0)$ is taken as the zero-point with electron configuration (N_1,N_2). The figure shows that the assumption of $\Delta E = 0$ is not met.

5 Acknowledgement

Bibliography

1 W. G. van der Wiel, S. De Franceschi, J. M. Elzerman, T. Fujisawa, S. Tarucha, and L. P. Kouwenhoven, "Electron transport through double quantum dots," *Reviews of Modern Physics*, vol. 75, no. 1, pp. 1–22, 2002.

2 G. Kiršanskas, *Electron Transport in Quantum Dots and Heat Transport in Molecules*. Thesis, 2014.

3 S. De Franceschi, L. Kouwenhoven, C. Schonenberger, and W. Wernsdorfer, "Hybrid superconductor-quantum dot devices," *Nat Nano*, vol. 5, no. 10, pp. 703–711, 2010.

4 A. Jellinggaard, K. Grove-Rasmussen, M. H. Madsen, and J. Nygård, "Tuning yu-shiba-rusinov states in a quantum dot," *Phys. Rev. B*, vol. 94, p. 064520, Aug 2016.

5 K. Grove-Rasmussen, T. Jespersen, A. Jellinggaard, and J. Nygård, *Hybrid Superconducting Devices Based on Quantum Wires*, book section 16. Oxford University Press, 2017.

6 M. T. Bjork, "Few-electron quantum dots in nanowires," *Nano Lett.*, vol. 4, pp. 1621–1625, 2004.

7 M. Tinkham, *Introduction to Superconductivity: Second Edition*. Dover Publications, 2004.

8 A. Jellinggard, *Quantum Dots Coupled to a Superconductor*. Thesis, 2016.

9 P. Krogstrup, N. L. B. Ziino, W. Chang, S. M. Albrecht, M. H. Madsen, E. Johnson, J. Nygård, C. M. Marcus, and T. S. Jespersen, "Epitaxy of semiconductor–superconductor nanowires," *Nat Mater*, vol. 14, pp. 400–406, 04 2015.

10 M. T. Deng, S. Vaitiekenas, E. B. Hansen, J. Danon, M. Leijnse, K. Flensberg, J. Nygård, P. Krogstrup, and C. M. Marcus, "Majorana bound state in a coupled quantum-dot hybrid-nanowire system," *Science*, vol. 354, no. 6319, pp. 1557–1562, 2016.

11 W. Chang, S. M. Albrecht, T. S. Jespersen, F. Kuemmeth, P. Krogstrup, J. Nygård, and C. M. Marcus, "Hard gap in epitaxial semiconductor–superconductor nanowires," *Nat Nano*, vol. 10, pp. 232–236, 03 2015.

12 H. I. Jørgensen, K. Grove-Rasmussen, K. Flensberg, and P. E. Lindelof, "Critical and excess current through an open quantum dot: Temperature and magnetic-field dependence," *Phys. Rev. B*, vol. 79, p. 155441, Apr 2009.

Structural dynamics of Cpf1 by FRET

Freja Jacobsen Bohr[a*] and Nikos Hatzakis[b]

The RNA-guided endonuclease of class V Cpf1 *(Clustered Regularly Interspaced Short Palindromic Repeats* from *Prevotella and Francisella)* is a central element in prokaryotic immune mechanism, which use a CRISPR-RNA (crRNA) to locate and cleave viral DNA. Cpf1 allow genome-editing at a specific position specified by synthetic crRNA, hence a promising therapeutic agent to treat gene deficiencies. Our current understanding of Cpf1 structure and function primarily

10 relies on crystal structures and cryoEM data, providing unique and invariant structures. In contrast to recent methods, the scope of this project is to expand our understanding of the dynamic structure of Cpf1, and understand how conformational changes and catalysis are related. To do this I used FRET (Förster Resonance Energy Transfer). Here, the work on this projects is presented, with the aim of 1) understanding the conformational changes of Cpf1 in free-form, binary complex (crRNA-bound) and tertiary complex (DNA-bound) in bulk, and correlating the conformational changes to intramolecular

15 distances, thus providing insight to the mechanism of DNA-cleavage and 2) calibrate the distance using dsDNA as a rigid scaffold, preparing for single molecule FRET measurements on Cpf1. Ensemble measurements revealed conformational changes of Cpf1 upon binding DNA, however the assay needs to be optimized further to extract distinct distances. Spectrometric experiments revealed that excess dye in solution was a general problem in ensemble measurements, interfering with the results. To validate and setup the calibration I did both ensemble and smFRET measurements on dual-

20 labelled dsDNA to address and eliminate the role of excess dye in solution. The inter-dye distance of dual-labelled dsDNA was determined to be 62.72 ± 0.93 Å. The simulated distance with Monte Carlo simulations was found to be 61.6 Å. This illustrates smFRET as a method to probe enzymatic motion, and thus could provide novel information about the mechanism of DNA-cleavage in Cpf1, paving the way for future genome-editing.

1 Introduction

25 In recent years RNA-guided endonucleases like CRISPR-Cas9 *(Clustered Regularly Interspaced Short Palindromic Repeats associated protein 9* from *Streptococcus pyogenes)*, has been of great interest as a powerful genome editing tool. Cas9 is a large, multi-domain protein that

30 undergoes conformational changes induced by single-guided RNA and trans-activating RNA, followed by cleaving of viral target DNA.[1] Providing the Cas9-complex with synthesized guide RNA, allows you to edit the genome at a specific position. The CRISPR technology is

35 applicable to many fields, such as stem cell- and gene therapy as a promising therapeutic option to treat inherited deficiencies such as cystic fibrosis or even cancer.[2,3,4]

In this thesis, the conformational changes of a similar RNA-guided endonuclease of class V Cpf1 *(CRISPR* from

40 *Prevotella and Francisella 1)* has been investigated. Cpf1 is a central element in the prokaryotic immune mechanism, that uses crRNA to locate and cleave viral DNA. In order to deconvolute the DNA-targeting mechanism and hence understand the molecular details in the system, the

45 structure of Cpf1 has recently been solved by our collaborating group at the Panum Institute using protein crystallization and cryoEM.[5]
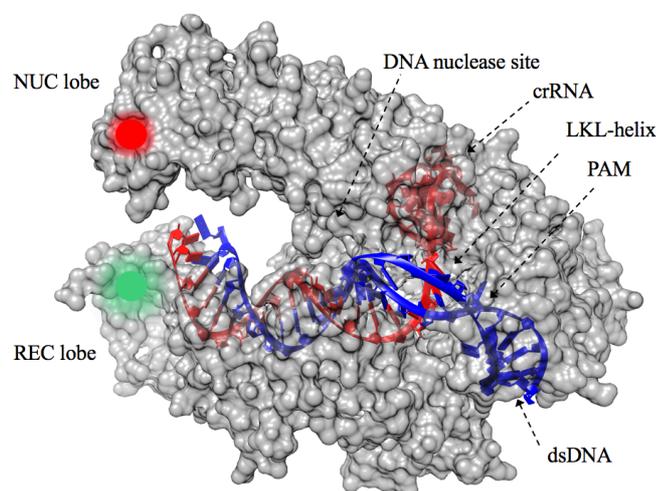
50

Figure 1: Crystal structure of Cpf1 after DNA cleavage. The transparent surface allows visualization of the crRNA-DNA complex. The NUC and REC lobes are shown, as well as the active site, PAM and LKL-helix. An acceptor fluorophore (red) is

55 at position CYS1190, and a donor fluorophore is at position CYS473. These fluorophores are used for FRET. (PDB ID: 5MGA)

Cpf1 contains two specific hinge regions, the NUC and REC lobe, and displays an oval 'sea conch' structure, as shown in Fig. 1. crRNA-induced conformational changes primes the enzyme to a DNA-binding competent state in presence of magnesium, that acts as a cofactor.[6] Target DNA induces further conformational changes to the Cpf1-crRNA complex, allowing recognition by the PAM (protospacer adjacent motif). PAM is essential for target binding as it is recognized directly by Cpf1. PAM is not found in the bacterial CRISPR locus, but is a component of the invading DNA, this way Cpf1 can differentiate between host and viral DNA. After PAM recognition a LKL helix (loop-lysine helix–loop) is inserted into the target DNA, thus beginning to unzip the dsDNA-strand. Due to electrostatics the ssDNA pairs with the crRNA, forming a crRNA-target-DNA heteroduplex. If crRNA and target-DNA are complementary, the target DNA is cleaved with PAM at 5' end. DNA nuclease activity takes place in a pocket at the interface between the RuvC and NUC domains (Fig. 1).

Current consensus indicates, that the free form of Cpf1 (from now on Cpf1-apo or apo) is primarily in a relaxed fully open conformation. Upon binding to crRNA the enzyme is expected to transit to a semi closed state. Finally, binding of target DNA and recognition by PAM makes the enzyme go towards a more closed compact conformation. As to unwind the DNA-strand, the conformational equilibrium changes to a more open conformation, thus cleaving the strand (Fig. S1). Hence, DNA-bound/DNA-cleaved complex samples many different conformations. From CryoEM and crystal structures the estimated distance between the two lobes in each conformational state has been.[7]

Cpf1 recognizes a T-rich PAM, while Cas9 targets a G-rich PAM, which provides alternate targeting sites. Cpf1 cleaves DNA via staggered cut approximately 18-23 basepairs downstream from the PAM site. By contrast, Cas9 cuts DNA proximal to the PAM site via blunt cut, hence allowing disruption to the recognition sequence after repair. Consequently, Cpf1 has proven to be a smaller and much more simple endonuclease than Cas9, expanding genome editing capabilities.[8]

The scope of this project is to investigate conformational dynamics of Cpf1 using FRET, to provide insight to the mechanism of DNA-cleavage.

1.1 Using FRET to record conformational changes

Förster Resonance Electron Transfer (FRET) can be used as a spectrometric ruler by reporting distance-related molecular changes between two fluorophores (donor and acceptor).[9] In this thesis, two cyanine dyes, Cy3 and Cy5, was chosen as FRET dye pair, since their donor emission spectrum and acceptor absorbance spectrum overlap, thus allowing energy transfer between donor and acceptor, as seen in Fig. 2. The magnitude of energy transfer is distance dependent and happens through intermolecular dipole-dipole coupling. In occurrence of FRET, the donor signal
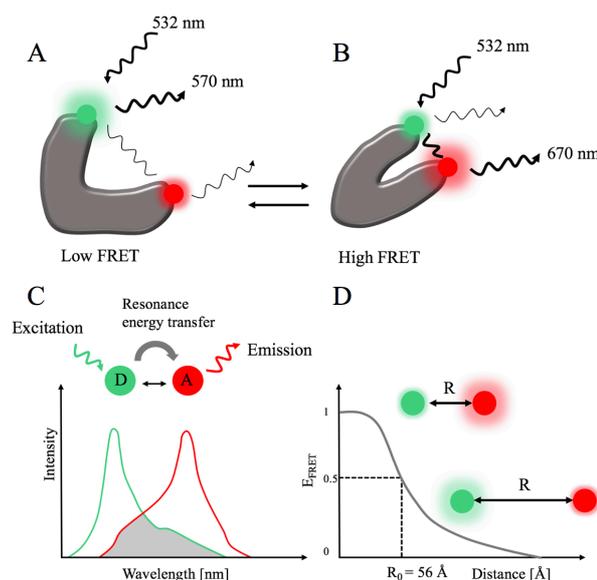
Figure 2: Illustration of how FRET works. **A:** Cy3 (green) and Cy5 (red) are attached to the enzyme at two flexible regions. **B:** Upon conformational changes and thus movement, the inter-dye distance between the fluorophores will change. The closer the two fluorophores are, the more energy will get transferred from Cy3 to Cy5, hence increasing the FRET signal. **C:** If donor emission spectrum (green) and acceptor excitation spectrum (red) overlap, the donor can transfer its resonance energy to the acceptor within a given distance. **D:** FRET efficiency as a function of distance. The Förster radius (R_0) can be found at 50 % FRET. When the dyes are in close proximity the FRET efficiency is high, and if they are far apart, the FRET efficiency is low.

is quenched, transferring energy to the acceptor, increasing the acceptor intensity (Fig. 2).[10] From the intensities of donor and acceptor (I_D and I_A) the FRET efficiency can be calculated as,

$$E_{FRET} = I_A / (I_A + I_D)$$

where I_D and I_A is the background and cross-talk corrected intensities for donor and acceptor, respectively. The FRET efficiency is given as the fraction of energy transfer occurring per donor excitation event. From E_{FRET} the intermolecular distance between donor and acceptor can be calculated using the Förster Radius (R_0), defined as the donor-acceptor distance obtained at 50 % E_{FRET} (here 56 Å[11]),

$$R = R_0 \times (1/E_{FRET} - 1)^{1/6}$$

The Förster radius is proportional to the donor quantum yield (Φ) or the emission efficiency of a given fluorophore, the overlap integral (J) between donor emission and acceptor excitation, the relative dipole orientation factor k^2 ($k^2 = 2/3$) and the refractive index of the medium (n)[12],

$$R_0 \sim k^2 \times \Phi \times J \times n^{-4}$$
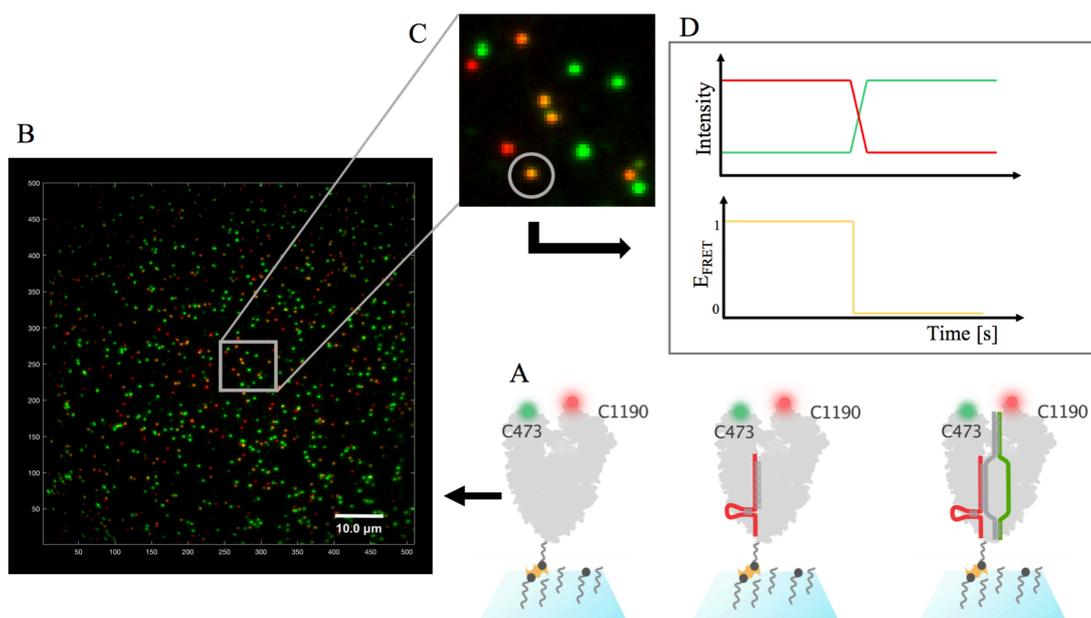
Figure 3: Experimental setup using TIRF microscopy for smFRET analysis. **A:** PLL-PEG surface with passivated Cpf1 through neutravidin-biotin linker. Here three different experimental conditions are shown: free-form (apo), Cpf1-crRNA binary complex and tertiary DNA-bound complex. **B:** Snapshot of the overlay of the red and green channel from dsDNA control experiment with same experimental setup (Fig. 6). The presented TIRF microscope setup provides data similar to what is shown. **C:** Zoom in focussing on donor-only (green), acceptor-only (red) and donor-acceptor (yellow). Colocalizing the D-only and A-only spots reveal FRET events as shown. **D:** Each colocalized donor and acceptor is related to an (idealized) FRET trace.

FRET is most sensitive to distance changes when the donor and acceptor are separated by a distance near the Förster Radius.[13] Thus, if fluorophores are attached to known sites within the enzyme, measurements of the energy transfer efficiency provide insight to inter- and/or intramolecular distances in the macromolecular length scale.[14]

1.2 Motivation for doing FRET studies on Cpf1

To investigate the dynamics and mechanism by which Cpf1 cleaves DNA, FRET studies are a valuable tool, as it is highly distance-sensitive. Ensemble FRET assays allow extraction of intermolecular distances between donor and acceptor in different environments, thereby providing crucial information about the conformational states. Protein dynamics are easily captured using FRET spectroscopy, as it has become a staple technique in recent years.[13] Additionally, ensemble assays are important in determining if a system is suited for smFRET measurements, and establishing the needed controls.

Shortcomings to the ensemble approach are that the extracted FRET distances provide only an average picture of the system, thereby potentially masking different states that might occur. Conformational dynamics can be observed in real-time by tracking changes in single molecule FRET efficiencies over time.[15] Hence, the single molecule approach allows investigation of the dynamic heterogeneity and subsequently disentangle the mechanism underlying conformational states of proteins. With single molecule techniques it is also possible to differentiate each molecule, yielding invaluable insight into distinctive molecular properties. The single molecule approach require complicated microscopy techniques, thus can be very challenging and expensive. Furthermore, single molecule data analysis is very time consuming, and requires numerous measurements to provide reliable and robust statistics, as the single molecule traces are hardly ever as the idealized ones portrayed in Fig. 3.

1.3 Total Internal Reflection Fluorescence microcopy

TIRF microscopy allow observation of thousands of single fluorescent events on surface-immobilized molecules, in a specific time-frame.[16] Immobilization of each molecule is necessary for real-time tracking using TIRF microscopy, as the diffusion of unrestricted molecules away from the field-of-view will make FRET measurements impossible.[17] Experimental setup for such TIRF smFRET experiment can be seen in Fig. 3. The enzymes, labelled with Cy3 and Cy5, are tethered to a PLL-PEG surface. Measuring smFRET through TIRF microscopy yields raw data as presented, that can be converted to FRET traces and thus providing FRET efficiencies and information about protein dynamics. For smFRET studies, it is crucial that the signal-to-noise ratio is sufficiently high to detect a single fluorescent molecule from the background. Because the evanescent field from the TIRF is restricted to a small volume, the signal-to-noise ratio is significantly improved than for other microscopes, and therefore yields the

required sensitivity for single molecule detection.[16] Consequently, the fluorescently labelled molecules must be tethered to the surface, so as they are located within the evanescent field.

1.4 Using ALEX to determine intermolecular distances

In order to resolve the conformational equilibrium changes of Cpf1, all experiments on the TIRF microscope were conducted using alternating laser excitation (ALEX). Using ALEX in smFRET studies has proven successful, as ALEX enables additional information and thus categorization of fluorescent traces by rapid switching between the two excitation wavelengths.[18] By doing so, it is possible to observe both fluorophores almost sequentially. Colocalizing donor and acceptor fluorescent signals in the raw image, reveal FRET events (Fig. 3). ALEX provides two fluorescence ratios; the FRET efficiency (E) and the stoichiometry (S) between the donor and acceptor fluorophores. The FRET efficiency is calculated from the background corrected intensity of donor and acceptor when exciting the donor, as previously described. The stoichiometry helps to ensure that a FRET signal arises from only one donor and one acceptor, and is calculated,

$$S = (I_A + I_D)/(I_A + I_D + I_A^*)$$

where I_A^* is the background corrected acceptor intensity when exciting the acceptor. Traces with only one acceptor and one donor are selected, defined by a bleaching step. Additionally, FRET measurements with ALEX allows for thermodynamic and kinetic analysis of the conformational changes.[19] The average number of background photons are calculated based on the uncorrected FRET traces, and subtracted from each channel separately.

Spectral cross-talk is usually a key problem in FRET studies using only one laser, since a part of the emitted donor photons are often detected in the acceptor-emission channel, due to spectral overlap. Therefore, the calculated energy transfer from donor to acceptor might be incorrect, thus resulting in wrong distances. ALEX provides a convenient way of accessing correction factors required for determining precise molecular distances.[20] Correction factors obtained are α and δ, to correct for blead-through between donor and acceptor channel and direct excitation of acceptor by donor-excitation laser, respectively. Differences in excitation intensities, quantum yields and detection efficiencies are accounted for by using the correction factor γ and β. Moreover, ALEX allows molecule sorting to exclude acceptor blinking and fluorophore bleaching.[21] α correction factor is calculated using background corrected FRET efficiency for donor-only traces (E^{DO}) and δ correction factor is determined from the stoichiometry of acceptor-only traces (S^{AO}),

$$\alpha = E^{DO} / (1 - E^{DO})$$
$$\delta = S^{AO} / (1 - S^{AO})$$

Incorporating α and δ correction factors allow calculation of the corrected acceptor fluorescence after donor excitation (F_A),

$$F_A = I_A - \alpha \times I_D - \delta \times I_A^*$$

The apparent FRET efficiency (E_{app}) and stoichiometry (S_{app}) is calculated, based on F_A. Correction factor γ is calculated as the normalization of fluorescence quantum yield and β is calculated as the normalization to equal excitation rates. In a homogenous approximation, γ and β can be determined by linearly fitting FRET populations to E_{app} and $1/S_{app}$ histograms, with y-intercept a and slope b as (Fig. 8A),

$$\beta = a + b - 1$$
$$\gamma = (a - 1)/(a + b - 1)$$

The corrected FRET efficiency and stoichiometry histogram is then calculated based on these four correction factors.[20] The expected FRET efficiency, <E> is deduced as the centre of a Gaussian fit to the corrected FRET efficiency. From the expected FRET efficiency, the distance is calculated as previously described.

Where single molecule measurements allow extraction of distinct heterogenetic behaviours from thousands of individual molecules, ensemble experiments yield the combined and hence average signal for millions of molecules. Even though this may mask distinct patterns within protein behaviour, the overall trend is observed - a trend that should be visible when comparing vast amounts of single molecule data with ensemble measurements. Hence a combination of both methods would allow validation and comparison of the achieved results.

1.5 Challenges when doing FRET

The major challenge in designing such FRET assays is to site-specifically label the enzyme with fluorophores, with high efficiency and precision to ensure that the reported FRET efficiency translates to the actual inter-dye distance. Cysteines are frequently used for site-specific labelling of proteins, because the thiol residues can react with a maleimide reagent (such as a fluorophore linker), resulting in a thioether.[20] Cpf1 was chemically modified, and all cysteines except three was removed.[7] The native surface exposed cysteines CYS1190 and CYS473 on the REC and NUC domains respectively, was used to monitor the conformational changes, linking the fluorophores to the protein. A third cysteine, C882, which is not solvent accessible, and hidden within the protein, could not be removed, because removal resulted in protein denaturation. Introducing multiple cysteines can contribute to protein misfolding due to non-native disulfide bridge formation, thus disrupt the DNA cleaving mechanism of Cpf1. Moreover, purification of labelled enzyme and separation from excess dye is crucial to ensure that reported FRET efficiencies are due to conformational changes and not to excess dyes in solution. Optimization of the purification

step can be done using e.g. size-exclusion chromatography to isolate enzyme from free dyes, however the yield might decrease drastically as the purity increases.[22]

Understanding the conformation and specific activity of proteins immobilized to surfaces are imperative, when doing smFRET experiments. Research has shown, that enzymes immobilized on the inert PLL-PEG/PEG-biotin surface remain fully active and that nonspecific adsorption is insignificant.[23]

1.6 Using dsDNA scaffolds as distance calibration

DNA labelled with Cy3 and Cy5 can act as a calibration control when doing ensemble and smFRET studies, as DNA can act as a rigid scaffold with no transitions. The 5' Cy3 is attached to the hydroxyl group of the ribose via a phosphodiester bond. The internal Cy5 dye is attached to the backbone via the phosphodiester bond of the base.[24] The primary amino group at the 3'-end can be used to attach a variety of modifiers, such as biotin for smFRET measurements. Moreover, labelling efficiency and purity are high compared to the enzyme. Additionally, DNA is inexpensive and fairly easy to work with, thus providing a technique to access the difficulties when doing FRET. The nature and dynamics of DNA is well-studied, and with its rigid scaffold and thus no transitions, dual labelled DNA has proven to be a great calibration of FRET efficiencies between donor and acceptor, and consequently intermolecular distance.[15]

2 Materials and Methods

For materials see supporting information.

2.1 Labelling of Cpf1

Labelled Cpf1 was provided by Montoya group. The enzyme is site-specifically labelled at positions CYS1190, CYS473 and CYS882, which can bind the fluorophores Cy3 and Cy5. The third labelling position CYS882 has shown not to have any effect on the FRET experiments, as it is hidden within the enzyme.[7] Unpublished (submitted) mass-spectroscopy data supports that the labelling efficiency on the third label is negligible. To model the actual distance between the two dyes, Monte Carlo simulations were performed. Simulations were done on crystal structures of Cpf1 (PDB ID: 5MGA), in FPS FRET positioning screening software.[25] The Monte Carlo simulations sample all possible dye-positions, and finds the average inter-dye distance (Fig. S3, A).

2.2 Ensemble FRET assay

2.2.1 Cpf1 experiment

1 μL 2.7 mg/mL Cpf1 (labelled with Cy3/Cy5) was divided into smaller eppendorf tubes. All fluorescence measurements were conducted at room temperature in reaction buffer (50 mM bicine, 150 mM KCl, pH = 8). The binary and ternary complexes between Cpf1, crRNA and DNA were assembled in reconstitution buffer by incubating Cpf1 with 200 nM RNA and 200 nM DNA for

at least 10 min at 25 °C prior to measuring, which resulted in full complex formation. When RNA and/or DNA was present 5 mM MgCl$_2$ was added to the buffer before incubation.

2.2.2 dsDNA control experiment

Dual-labelled dsDNA (45 basepairs oligonucleotide) with Cy3/Cy5 was made from ssDNA/Cy3 (end-biotinylated) and ssDNA/Cy5 mixed in 1:1 molar ratio, to ensure high annealing efficiency. The oligonucleotides were annealing using the assigned annealing protocol.[26] The two dyes are separated by 17 basepairs DNA sequence, corresponding to a distance of 58.3 Å (measured using Chimera[27]). To model the actual distance between the two dyes, Monte Carlo simulations were performed on the crystal structure of dsDNA (Make-Na server[28]), in FPS FRET positioning screening software.[25] (Fig. S3, B).

Bulk fluorescence measurements were carried out on a T-format spectrofluorometer (Jasco FP-6200; Jasco, Easton, MD). For each FRET construct, a donor-only (Cy3-labelled) and acceptor only (Cy5-labelled) sample was prepared and its emission spectra at 530 nm and 640 nm excitation were collected, to correct for cross excitation an emission. Each experimental condition was measured at least 3 times for statistical robustness. Excitation at 532 nm uncovers spectra with emission peaks observed at 560 nm (donor) and 670 nm (acceptor). The FRET efficiency was calculated based on donor and acceptor intensity, for each experiment and thereby allow extraction of the distance between the fluorophores in each condition.

2.3 smFRET assay

2.3.1 dsDNA control experiment

The surface for smFRET experiments was made on passivated and cleaned cover glass[29] with a 1:100 mixture of 80 μL PLL-PEG and PLL-PEG-biotin. The surface was functionalized with 80 μL 0.1 g/L neutravidin. Prior to each experiment, 1 μM dsDNA was diluted 50,000 times and introduced to the microscope chamber and incubated for 2 min at 25 °C to ensure immobilization via biotin-neutravidin linker. Excess dsDNA was washed away with buffer (50 mM bicine, 150 mM KCl, 5 mM MgCl$_2$, pH = 8). An oxygen scavenging system / imaging buffer (1 U/mL PCD, 2.5 mM PCA, 2 mM Trolox) was prepared, and flushed into the chamber just before each measurement, to prevent fluorophore-blinking.[30]

All smFRET experiments were conducted on the TIRF microscope equipped with two EMCCD cameras. Donor (Cy3) and acceptor (Cy5) were excited using 532 nm (green) and 640 nm (red) solid state laser lines. A dual-cam setup was used to split the signal into two distinct channels. All experiments were conducted using alternating laser excitation (ALEX), with 100 ms exposure time, 300 EM gain and frame rate of 2 s^{-1}.[9,20]

2.3.2 Data treatment

Image analysis was performed using iSMS software package for smFRET, to extract traces for each co-localized donor/acceptor FRET pair. α and δ correction factors were extracted from iSMS.[31] Further data analysis to extract correction factors γ and β, and trace categorization was made on an in-house software written in Python by Johannes Thomsen.

3 Results and discussion

3.1 Ensemble FRET

All fluorescence ensemble measurements were obtained as intensity as a function of wavelength. The intensity is measured in arbitrary units, and therefore relative to each sample, hence the following results are normalized to 1.
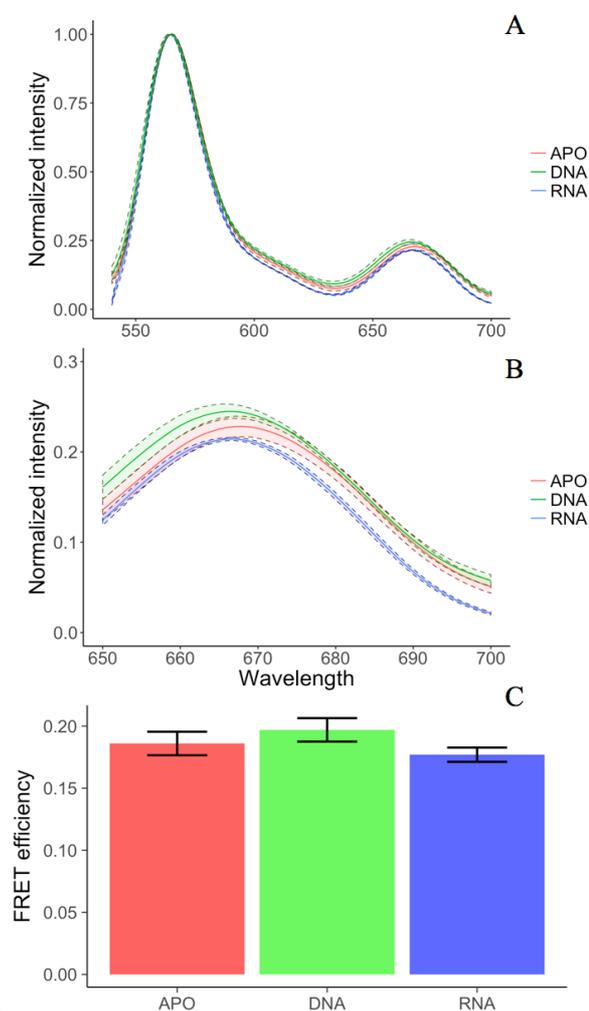
Figure 4: Results from ensemble FRET measurements on Cpf1. **A:** Normalized and corrected spectra for ensemble FRET measurements on free-form (apo) Cpf1, crRNA-bound and DNA-bound/DNA-cleaved. **B:** Zoom in of A in acceptor region. The spectrum indicates a transition from a more relaxed open conformation to a more compact conformation upon DNA binding. Excitation wavelength is 532 nm. **C:** Bar-chart of the FRET efficiencies.

Table 1: Ensemble FRET results for Cpf1

	Apo	**crRNA**	**DNA**
E_{FRET}	0.186 ± 0.009	0.177 ± 0.006	0.197 ± 0.009
R [Å]	71.64 ± 0.75	72.39 ± 0.48	70.80 ± 0.71

FRET efficiency and radius from ensemble FRET measurements on Cpf1 on free-form (apo), binary complex (crRNA) and tertiary complex (DNA). The results indicate that DNA bound form is in the most compact conformation. Errors are reported as S.D. of at least three measurements.

Exciting the donor fluorophore at 532 nm, and measuring emission intensity from 540 nm to 700 nm allows extraction of FRET efficiencies for each experimental condition (apo, crRNA/RNA and DNA). The acceptor intensity differs between apo, crRNA and DNA, hence the FRET efficiency and thereby distance varies as seen in table 1. Uncertainties in Fig. 4 are reported as the standard deviation from three-to-six experiments. The FRET efficiencies and distances were extracted as previously described, and the error determined with error propagation. The resulting FRET efficiency and distance can be seen in table 1. The ensemble approach allows investigation of the multiple conformations of Cpf1, thus recording conformational equilibrium shifts. Each sample was incubated for at least 10 minutes before measuring to ensure equilibrium between different conformational states, and consequently the reported distance will be a sum of distances for all possible conformations.

To quantify if distances are significantly different, I performed a Welch's t-test for unequal variance. This test assumes normality, and compares central tendency for two populations.[32] Performing the test with a significance threshold of α = 0.05 revealed only one significant transition (Table S1.), showing that the conformational change, when DNA is added to the sample, is significant, with a p-value 3.37e-04. Thus, binding of DNA results in conformational changes of Cpf1, and the enzyme transits to a more closed state, with a 15 % increase in acceptor intensity compared to crRNA-bound. The inter-dye distance for the tertiary DNA-bound/DNA-cleaved state was found to be 70.80 ± 0.71 Å. I then went on to use an orthogonal method to further evaluate the distances using Monte Carlo (MC) simulations on crystal structure of Cpf1 after DNA-cleavage, which reveals an average dye-dye distance of 68.2 Å (Fig. S3, A), in good agreement with the experimental results, and thus I could further back the hypothesis from FRET measurement. The difference between the DNA-bound and the DNA-cleaved state cannot be distinguished in ensemble measurements, as the distinct states are masked when averaging. However, since the samples were incubated, I hypothesize that conformational equilibrium was reached prior to measuring. The distance of 70.80 ± 0.71 Å corresponds to an open conformation, thus represents the DNA unwinding and the DNA-cleaved state of Cpf1, and indicates that binding of DNA shifts the conformational equilibrium of

Cpf1 compared to crRNA-bound conformation. The results from table 1 indicate that the enzyme upon binding crRNA shifts the conformational equilibrium towards a more open conformational state, with a greater distance on average, although this difference was found not to be significant (Table S1.). It seems that the different conformational states are masked in ensemble FRET, making it impossible to distinguish between apo-form and crRNA-bound conformation. The low-FRET state in the apo protein ($E_{FRET} = 0.186 \pm 0.009$) indicates the large conformational landscape of Cpf1 in the absence of crRNA and DNA and is in agreement with the hypothesized large conformational heterogeneity of the apo protein.

3.2 Evaluation of FRET

In order to ensure that the changes in FRET efficiencies arise from conformational changes and therefore converted accurately to distance, a calibration-control experiment using dual-labelled dsDNA with known inter-dye distance was performed. DNA is very rigid, hence a very prevailing measure for inter-dye distance in FRET experiments. The experiment was executed equivalently, and the emission spectrum recorded (Fig. S4). The hypothesis was further supported by MC simulations to extract dye-dye distances in the dual-labelled dsDNA. Different linker sequences can shift the distance and orientation of the fluorescent protein, causing changes in FRET efficiencies, and accentuating the importance in knowing the exact dye-dye distance.[33] MC simulations revealed a mean inter-dye distance of 61.6 Å (Fig. S3, B).

Converting FRET efficiencies into distance resulted in an inter-dye distance of 72.53 ± 0.46 Å, greater than the simulated dye-dye distance of 61.6 Å extracted from MC simulations, indicating that either the DNA-strands were annealed uncomplete, or that some additional fluorescent contribution is present when measuring.
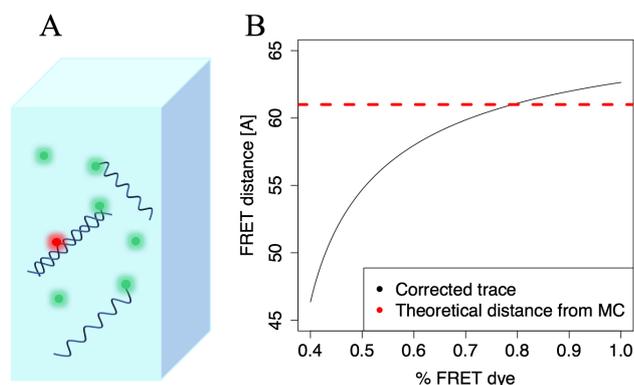
Figure 5: Correction for excess dye. **A:** Different types of dyes in solution. 1) Un-annealed ssDNA with Cy3-fluorophore, 2) free (unbound) Cy3-fluorophores and 3) dual-labelled dsDNA (so-called "FRET" dye) **B:** Correction for excess dye. The black line is the calculated distance assuming a given ratio of "FRET" dye (the percentage that actual contributes to the FRET efficiency and not due to excess dye). Intercept (red line = simulated distance) indicates the proportion of excess dye in solution that would correspond to the simulated distance of 61.6 Å.

According to the annealing protocol[26], the annealing is expected to be sufficiently high, so as the signal should not be disturbed from un-annealed ssDNA labelled with fluorophores.

Excess Cy3-fluorophores (Cy3) in solution will result in an increased donor intensity, hence less FRET efficiency and longer distance. This contribution to donor intensity could arise from unbound and thus free Cy3, or it could be un-annealed ssDNA with Cy3 (Fig. 5, A). Excess Cy5-fluorophores should not contribute to the intensity, as only laser excitation at 532 nm occurs. Using the equation:

$$I_D^{measured} = P_{cy3} \times I_{cy3} + P_{FRET} \times I_D^{actual}$$

it is possible to correct for excess dye in solution. I hypothesise that $I_D^{measured}$ (the measured donor intensity for dual-labelled dsDNA) corresponds to a sum of the contribution from excess dye, including unbound and un-annealed dyes ($P_{cy3} \times I_{cy3}$), and the contribution from actual FRET signal ($P_{FRET} \times I_D^{actual}$). Based on the equation, the actual donor intensity arising from FRET (I_D^{actual}) can be calculated. Calculating E_{FRET} and distance from I_D^{actual} instead of I_D allow to correct for excess dye (Fig. S5). Using the above equation, our findings suggest approximately 20 % excess dye in solution and thus 80 % "FRET" dye (Fig. 5, B). Consequently, having excess dye in solution whilst measuring FRET, may reveal the overall trend in the conformational changes, but fails to translate correctly into inter-dye distance, hence providing an incorrect depiction of the intermolecular distances of the different enzymatic states. Further on, excess dye could mask the smallest and most sensitive conformational equilibrium changes of Cpf1 in the ensemble approach.

These findings may provide an explanation as to why the emission spectrum for apo, crRNA- and DNA bound in Fig. 4 are very similar. I hypothesize that excess dye in solution contributes to the donor signal, masking the distinct difference between the different conformational states of Cpf1 in ensemble FRET measurements. It is impossible to differentiate between the relative population of multiple distributions and the shift of the distribution as a whole in ensemble measurements, consequently preventing extraction of the distinctive conformational equilibrium changes and corresponding intermolecular distances.

3.3 smFRET

As a setup for additional experiments with Cpf1, single molecule FRET is ideal to uncover distinct conformational changes otherwise masked in ensemble approach. To further understand the system, and calibrate for excess dyes in solution, single molecule FRET measurements on dsDNA were done. Dual-labelled dsDNA was passivated on a PLL-PEG surface using neutravidin-biotin linker (Fig. 6). Flushing with buffer ensures that only tethered DNA is present in the sample, thus removing unbound dyes and DNA.
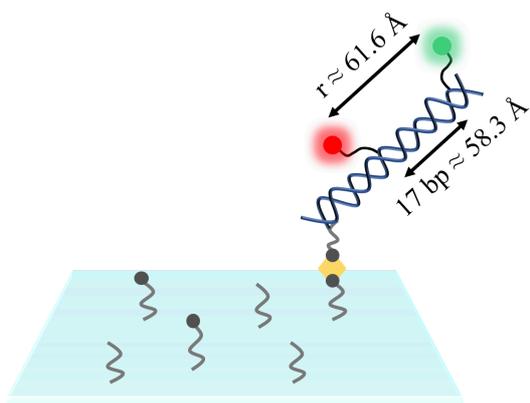
Figure 6: Experimental setup for the single molecule TIRF experiments on dual-labelled dsDNA. DNA is tethered to the PLL-PEG surface through biotin-neutravidin linker. The fluorophore linkers are separated by 17 bp corresponding to 58.3 Å. And the distance between the attached fluorophores is 61.6 Å found from MC.

smFRET experiments were conducted on a TIRF microscope with a dual-cam setup using ALEX. smFRET trajectories of hundreds of individual dsDNA molecules in steady state conditions were captured, and data treatment was finalized as presented earlier.

Counting each Cy3 and Cy5 in all raw movies, revealed 11.197 traces consisting of colocalized Cy3 and Cy5 (table S2). Comparing this number with the total number of Cy3 revealed 60.68 % non-colocalized Cy3. This number is even greater than the 20 % excess dye found in the distance calibration for ensemble experiments (Fig. 5, B). Surprisingly, the raw data showed equal proportions of Cy3 and Cy5. These findings indicate that Cy3-labelled ssDNA (end-biotinylated) and/or single-labelled dsDNA is tethered to the surface (Fig. S2). Emphasizing that labelling efficiency and annealing should be considered an important factor when doing ensemble measurements, as it can highly interfere with the results.

Using iSMS the raw traces were converted into distinct FRET traces (~ 200 traces) (Fig. 7). Further classification reveal 66 evident FRET traces. The traces are primarily categorized based on presence of a single bleaching step as well as a stoichiometry at 0.5, indicating a single FRET pair. A typical FRET trace can be seen in Fig. 7, where the green and red signal are cross-correlated, demonstrating that decrease in acceptor intensity leads to donor intensity increasing. For additional traces see Fig. S6.

The acceptor excitation channel verifies bleaching event for the acceptor, thus resulting in increasing donor intensity. The acceptor bleaching in combination with the donor bleaching and the stoichiometry reveals a certain FRET signal. Other factors taken into account are the standard deviation and median of stoichiometry, intensities and FRET efficiency. These guidelines are based on prior expertise in looking at FRET traces. The corrected FRET efficiency histogram (Fig. S7, B) is determined based on
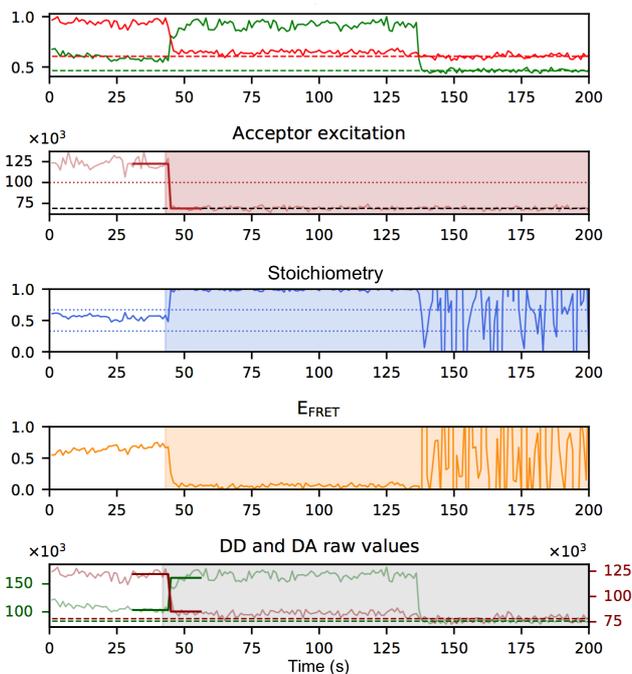
Figure 7: Representative single molecule trace using ALEX single molecule FRET measurements. ALEX provides a read-out for both the green and red channel, allowing for simultaneously observing both donor and acceptor. As the acceptor bleaches in a single bleaching step, the donor intensity increases, hence validating the trace as actual FRET with only a single FRET pair (S = 0.5). As the donor bleaches, the acceptor intensity is unchanged, indicating that both fluorophores are bleached.

the 66 FRET traces, from which the expected FRET efficiency is computed, and later converted to distance.

Correction factors were extracted from the 66 FRET traces (Fig. 8A). The α- and δ-correction factor is determined through iSMS, based on the baseline of each trace. The background can be set manually in iSMS, such that α and δ correct for blead-through and direct excitation. Additionally, β- and γ-correction factors are calculated as previously described, correcting for detection and excitation (table 2, Fig. 8A). Taking the correction factors into account, allow calculation of the expected FRET efficiency as the mean of a Gaussian fit to the non-binned FRET efficiency data, and thus distance is found (Fig. 8B). Although FRET efficiencies are not theoretically normally distributed, it has in practice been shown to be a robust method, with little discrepancy.[34]

Table 2: Correction factors

| α	δ	β	γ
0.060	0.020	2.317	0.644

Correction factors from smFRET analysis on dual-labelled dsDNA. The correction factors correct for blead-through, direct excitation, differences in excitation intensities, quantum yield and detection efficiencies.

Fitting the probability density with a normal distribution results in a reduced chi-squared statistics $\chi^2/\nu = 2.27$. The reduced chi-square statistics is defined as chi-squared per degree of freedom.[35] Given the found chi-square the

5 Gaussian fit seems to explain the data well. The residuals follow a normal distribution (Fig. S8) as expected from the central limit theorem, with error $\sigma = 0.17$, and mean $\mu = 0.00$. The standard error of the mean (SEM) for the distance is calculated to be 0.93 Å, which demonstrates

10 that the calculated mean is in close proximity to the true mean of the population.[36] Our single molecule data reveals 0.34 ± 0.02 FRET, which predicts an inter-dye distance of 62.72 ± 0.93 Å (Fig. 8B), in great agreement with theory proposing an inter-dye distance of 61.6 Å from MC.

15 Individual smFRET trajectories are analyzed by hidden Markov modelling[37] (Fig. 9), to investigate transitions and revealed only a single populated state with no transitions, as expected due to the rigidity of the DNA scaffold that was investigated.
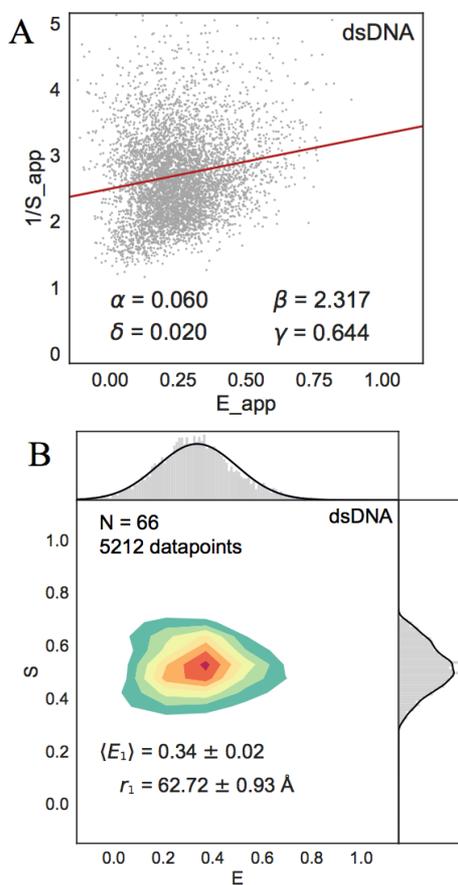
Figure 8: Results from the smFRET experiment **A:** Correction factors β and γ determined by linearly fitting E_{app} to $1/S_{app}$ histograms. **B:** Gaussian distributions were used to fit the histograms to identify individual populations of E_{FRET} states.

25 Using the four correction factors to determine the FRET efficiency reveals the corrected two dimensional histogram showing stoichiometry and FRET efficiency and thus the expected efficiency and distance with standard error of the mean.
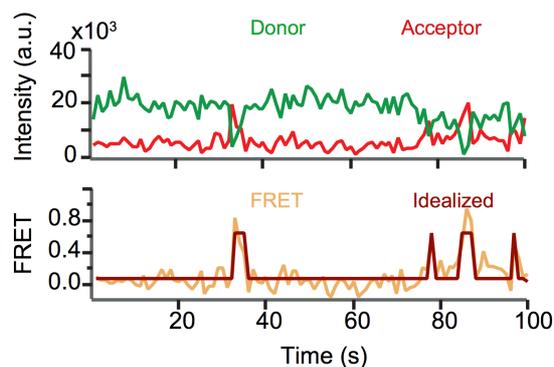
30 **Figure 9:** Individual smFRET traces fitted with hidden Markov modelling (HMM). The raw intensity traces for donor and acceptor can be seen at the top. The intensities are only shown until a bleaching event occurs. At the bottom is the calculated FRET trace, as well as the idealized (modelled from HMM). Each

35 frame corresponds to one data point. This trace is from smFRET studies on Cpf1 (done by Johannes Thomsen and Simon Bo Jensen), showing multiple transitions.[7] Traces from dsDNA smFRET measurements show only one state, and no transitions as expected.

40 The stoichiometry populates at around 0.5 indicating that only a single FRET pair is present. The smFRET results are concluded based on 66 traces and 5212 data points in total, and thus considered statistical robust. Furthermore,

45 the results illustrate that correction factors can be extracted, and consequently FRET efficiency and distance determined with great accuracy. This method can easily be translated to investigation of conformational changes of Cpf1, using the experimental setup shown in Fig. 3.

50 smFRET analysis would provide further knowledge to the conformational equilibrium shifts of Cpf1, without the problem of excess dye, allowing for extraction of intramolecular distances at each conformational state.

55 In conclusion, FRET can be used as a spectroscopic ruler to report distance related conformational changes of molecules. The ensemble FRET measurements provided an easy way to access information about the conformational changes of Cpf1, showing that only the DNA-bound state

60 is significantly different from the others. A control with dual-labelled dsDNA revealed that excess dye in solution complicates the ensemble FRET measurements, and thus may interfere with the results. smFRET experiments on the same dual-labelled dsDNA allowed us to extract dye-dye

65 distance of 62.72 ± 0.93 Å in close proximity to MC simulated distance of 61.6 Å, thus demonstrating smFRET as a precise method to reveal conformational changes, and thus extract intermolecular distances without interference of excess dye.

70

3.2 Future perspectives

Detailed quantification of the conformational changes of Cpf1 was not carried out in the smFRET assay within the time frame of this project, especially since the microscope was not working the last two months of the project. However, could be implemented, using the described assay above (Fig. 3). Additional information about the mechanism of DNA binding and cleavage could be achieved with three-color FRET imaging. Labelling the target-DNA with an additional fluorophore, would allow single molecule tracking to ensure that the DNA is bound and consequently when it is cleaved. This would provide further knowledge to the mechanism by which Cpf1 binds and cleaves DNA, as the DNA-bound and DNA-cleaved state can now be differentiated. Understanding, how the DNA binds and ultimately when it is cleaved provide knowledge about the structure and function of Cpf1 and thus adds another dimension to our understanding of modern genome editing.

4 Conclusion

Investigations of the conformational changes of Cpf1 using Förster Resonance Energy Transfer to quantify the intramolecular dynamics were carried out. Firstly, ensemble FRET measurements of the conformational changes of Cpf1 upon adding crRNA and DNA were executed. Cpf1 was labelled with Cy3 and Cy5 at the two hinge regions (NUC and REC lobe). The assay revealed a significant conformational change when adding DNA, translating into a dye-dye distance of 70.80 ± 0.71 Å. This corresponds to the MC simulated distance of 68.2 Å, showing that the enzyme is in a DNA-bound/DNA-cleaved state. However, in ensemble assay it was not possible to directly differentiate between DNA-bound and DNA-cleaved conformation. Further examination of dual-labelled dsDNA as a rigid scaffold displayed approximately 20 % excess dye in solution, revealing difficulties with the ensemble assay. smFRET based approach was utilized to directly observe and quantify the conformation of dsDNA, and eliminated the role of excess dye. smFRET assay further supported that labelling efficiency and annealing of dsDNA should be considered an important factor in ensemble measurements. The inter-dye distance of dual-labelled dsDNA was proven to be 62.72 ± 0.93 Å – which resembles the simulated distance of 61.6 Å modelled from Monte Carlo simulations. Only one populated state with no transitions was found using hidden Markov model analysis. In this thesis I show, that FRET is a confident method to extract intermolecular distances, thereby could prove novel information about the conformational states of different biomolecules. This single molecule assay could be expanded to study the conformational states of Cpf1.

6 References

[a] Thorvaldsensvej 40, 1871 Frederiksberg C, University of Copenhagen, Denmark.; E-mail: frejabohr@hotmail.com
[b] T554, Thorvaldsensvej 40, 1871 Frederiksberg C,
University of Copenhagen, Denmark.

1. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550,** 407–410 (2017).
2. Booth, C., Gaspar, H. B. & Thrasher, A. J. Treating Immunodeficiency through HSC Gene Therapy. *Trends Mol. Med.* **22,** 317–327 (2016).
3. Schwank, G. *et al.* Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* **13,** 653–658 (2013).
4. Kampmann, M. A CRISPR Approach to Neurodegenerative Diseases. *Trends Mol. Med.* **23,** 483–485 (2017).
5. Stella, S., Alcón, P. & Montoya, G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* **546,** 559–563 (2017).
6. Fonfara, I., Richter, H., BratoviÄ, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532,** 517–521 (2016).
7. Montoya, G., Hatzakis, N. & et. al. Conformational Activation of CRISPR-Cpf1 Catalysis and Endonuclease Function Recycling. *Submitted* (2018).
8. Fran, R. Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 Digvijay. 443–447 (2005). doi:10.1016/B978-0-12-397169-2.00038-X
9. Rahul, R., Hohng, S. & Ha, T. A Practical Guide to Single Molecule FRET. *Nat. Methods* **5,** 507–516 (2008).
10. Sekar, R. B. & Periasamy, A. Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J. Cell Biol.* **160,** 629–633 (2003).
11. Koch, A. *et al.* High-precision FRET analysis of the G-protein coupled receptor TGR5 in live cells. *Eur. J. Med. Res.* **19,** S12 (2014).
12. Iqbal, A. *et al.* Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proc. Natl. Acad. Sci.* **105,** 11176–11181 (2008).
13. Ma, L., Yang, F. & Zheng, J. Application of fluorescence resonance energy transfer in protein studies. *J. Mol. Struct.* **1077,** 87–100 (2014).
14. Hussain, S. An Introduction to Fluorescence Resonance Energy Transfer (FRET). *arXiv Prepr. arXiv0908.1815* (2009).
15. Di Fiori, N. & Meller, A. The Effect of dye-dye interactions on the spatial resolution of single-molecule FRET measurements in nucleic acids. *Biophys. J.* **98,** 2265–2272 (2010).
16. Tinoco, I., Gonzalez, R. L., Jr, I. T. & Jr, R. L. G. Biological mechanisms , one molecule at a time. 1205–1231 (2011). doi:10.1101/gad.2050011
17. Bavishi, K., Li, D., Eiersholt, S., Hooley, E. N. & Petersen, T. C. Direct observation of multiple conformational states in Cytochrome P450 oxidoreductase and their modulation by membrane environment and ionic strength. 1–9 (2018). doi:10.1038/s41598-018-24922-x
18. Kapanidis, A. N. *et al.* Alternating-laser excitation of single molecules. *Acc. Chem. Res.* **38,** 523–533 (2005).
19. Kapanidis, A. N. *et al.* Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci.* **101,** 8936–8941 (2004).
20. Hellenkamp, B. *et al.* Precision and accuracy of single-molecule FRET measurements - a worldwide benchmark study. 1–27 (2017).
21. Hohlbein, J., Craggs, T. D. & Cordes, T. Alternating-laser excitation: single-molecule FRET and beyond. *Chem. Soc. Rev.* **43,** 1156–1171 (2014).
22. van Vught, R., Pieters, R. J. & Breukink, E. Site-Specific Functionalization of Proteins and Their Applications To Therapeutic Antibodies. *Comput. Struct. Biotechnol. J.* **9,** e201402001 (2014).
23. Zhen, G. *et al.* Immobilization of the enzyme β-Lactamase on biotin-derivatized poly(L-lysine) -g-poly (ethylene glycol)-Coated sensor chips: A study on oriented attachment and surface activity by enzyme kinetics and in situ optical sensing. *Langmuir* **20,** 10464–10473 (2004).
24. Integrated DNA Technologies. *https://eu.idtdna.com/pages/Support/FAQs/where-on-a-nucleotide-are-cy-dyes-attached-*
25. Kalinin, S. *et al.* A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods* **9,** 1218–1225 (2012).
26. Bioneer. Anneal complementary pairs of oligonucleotides. *Tech. Tip* **0747,** 1–2 (2009).
27. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
28. Stroud, J. Make-Na server. (2004).
29. Moses, M. E., Hedegård, P. & Hatzakis, N. S. Quantification of Functional Dynamics of Membrane Proteins Reconstituted in Nanodiscs Membranes by Single Turnover Functional Readout. *Methods Enzymol.* **581,** 227–256 (2016).
30. Vogelsang, J. *et al.* A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angew. Chemie - Int. Ed.* **47,** 5465–5469 (2008).
31. Preus, S., Noer, S. L., Hildebrandt, L. L., Gudnason, D. & Birkedal, V. ISMS: Single-molecule FRET microscopy software. *Nat. Methods* **12,** 593–594 (2015).
32. Kim, T. K. T-test as a Parametric Statistic. *Korean J. Anesthesiol.* **68,** 540–546 (2015).
33. Bajar, B. T., Wang, E. S., Zhang, S., Lin, M. Z. & Chu, J. A guide to fluorescent protein FRET pairs. *Sensors (Switzerland)* **16,** 1–24 (2016).
34. Yang, M. *et al.* The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. *Cell Rep.* **22,** 372–382 (2018).
35. Andrae, R., Schulze-Hartung, T. & Melchior, P. Dos and don'ts of reduced chi-squared. 1–12 (2010). doi:10.1088/0004-637X/733/1/66
36. Nagele, P. Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *Br. J. Anaesth.* **90,** 514–516 (2003).
37. McKinney, S. A., Joo, C. & Ha, T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **91,** 1941–1951 (2006).

Thesis Paper

Effects of Hydrogenphosphate Quenching on Fluorescent Lifetime of Hydroxyphenyl Substituted Diazaoxatriangulenium

Stine Grønfeldt Stenspil*

5

Fluorophores are used in many research applications such as cell staining and as probes for ions, DNA, proteins or membranes. By designing and synthesising the fluorophore to have specific photophysical properties it is possible to probe the surrounding environment of the fluorophore. For example, by utilising the property of photoinduced electron transfer (PET) which quenches the fluorescence of the fluorophore. 10 A hydroxyphenyl substituted diazaoxatriangulenium (DAOTA$^+$) designed to have the properties as a pH probe using photoinduced electron transfer showed another pH dependency in addition to the expected quenching. Results show that hydrogenphosphate is the reason for the additional quenching and other ions is also shown to have quenching abilities. Further understanding of this mechanism could result in a new triangulenium derivative that is able to detect hydrogenphosphate or other ions in solution. The detection 15 of hydrogenphosphate is especially interesting as it is a key component in physiological solutions and cells.

Introduction

Triangulenium dyes are organic fluorophores consisting of carbon rings linked together either with oxygen,[1] nitrogen[2] or carbon[3] bridges forming a triangular planar and rigid structure with a
20 carbocation in the centre. The rigidity of these dyes gives them advantage of long lifetimes of 10-20 ns and large quantum yields compared to other well-known fluorophores.[4] The longer lifetimes give the possibility of high contrast images of cells using time gating. This is possible because the triangulenium fluorescence is
25 still measurable after the auto luminescence of the cell has fallen in intensity.[3] The triangulenium dyes can have many different applications depending on the substitution groups on the nitrogen bridges. Using a biolinker, the fluorophore can bind to proteins and then be used to study protein motion and interactions.[5] If an
30 "antenna" in the form of a dye with high absorption coefficient is used the fluorescence signal of the triangulenium will be enhanced with the help of energy transfer. These types of fluorophores are known as dyads and can be used in cellular imaging with the advantage of high fluorescence intensity and long lifetime.[6]
35 Many fluorescent compounds are synthesized and used as probes in biological and medical research. There are probes designed for the detection of ions such as Ca^{2+}, Mg^{2+}, Na^+, K^+ and as pH sensors.[7] Fluorescein and its charged derivatives are examples of pH responsive fluorescent probes, which are among some of the
40 first fluorescent pH sensors used.[8] However, they undergo rapid photobleaching making them less useful for measurements over longer periods of time. Tuning the triangulenium fluorophore using a phenol Figure 1A as substitution group makes the dye an efficient pH probe.[9] This is possible by taking advantage of
45 intramolecular photoinduced electron transfer (PET). In Figure 1C
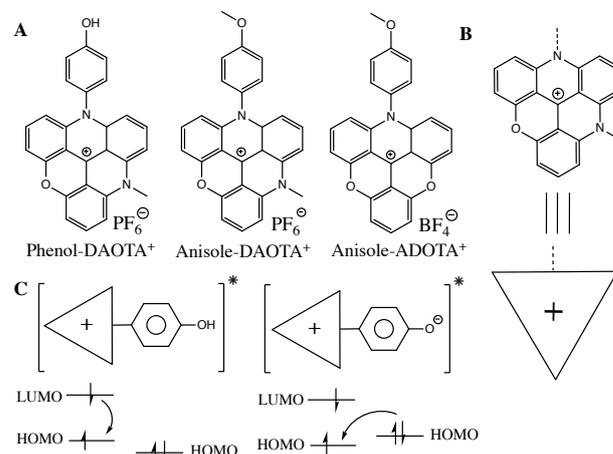
Figure 1: A: Molecular structure of hydroxyaryl diazaoxatriangulenium (Phenol-DAOTA$^+$), methoxyaryl diazaoxatriangulenium (Anisole-DAOTA$^+$) and methoxyaryl azadioxatriangulenium (Anisole-ADOTA$^+$).
50 **B**: Schematic representation of the diazaoxatriangulenium. **C**: Schematic drawing of photoinduced reductive electron transfer. The exited electron in the LUMO for the phenol form is free to deactivate by photon emission. For the phenolate the HOMO of the donor is raised and the HOMO of the fluorophore is occupied due to the larger PET efficiency preventing the
55 exited electron in the LUMO to deactivate.

the mechanism is shown. An electron in the HOMO of the fluorophore is exited to the LUMO. Another electron is transferred from the donor (the phenol) to the fluorophore and thus occupying the HOMO and preventing the exited electron to return to ground
60 state while emitting a photon.[10] In this case when the fluorophore is in acidic solution it is in its phenol form and will fluoresce, while in basic solution the phenolate form will dominate and the dye is

quenched. Using this fluorophore as pH sensor gives the possibility to measure the pH over increased periods of time due to its good photostability.[3, 11]

Previous work has characterized the phenol substituted DAOTA$^+$ in meta, ortho or para positions.[9] During photophysical characterization a small fluorescence quenching was observed before the major PET quenching as a result of pH change. Analysis of the lifetimes also showed a clear pH dependent decrease. The change in fluorescent lifetime showed that it was not related to the phenol/phenolate reaction as only the lifetime of the unquenched fluorophore should be present. Due to this small quenching not being near the phenol/phenolate equivalence point it must be a property of the solvent interacting with the dye and not connected to the pH related deprotonation. As the measurements were made in a phosphate buffered saline (PBS) and dimethylsulfoxide (DMSO) solution it could be hydrogen bonds to the hydroxyl group lowering its HOMO resulting in an extra fluorescence "on" state. There is also the possibility it is the salts in the solution which are shielding or deprotonating the fluorophore through some form of mechanism.

In this paper the compound of interest hydroxyphenyl diazaoxatriangulenium (phenol-DAOTA$^+$) Figure 1A was examined in various solvent compositions. The most similar derivative without possibility of deprotonation is the methoxyphenyl derivative shown in Figure 1A. This donor was examined with two different triangulenium acceptors shown in Figure 1A: diazaoxatriangulenium (anisole-DAOTA$^+$) and azadioxatriangulenium (anisole-ADOTA$^+$). Change of the solution parameters showed that it is hydrogenphosphate deprotonating the exited state of the phenol-DAOTA$^+$ when colliding with it and thus quenching the dye. Other ions were also investigated for their quenching abilities showing both hydroxide and acetate ions having an effect on the phenol-DAOTA$^+$ lifetime however not in the same degree as hydrogenphosphate.

With some improvement this property could be exploited for the detection of hydrogenphosphate in solution. Other dyes have been synthesized for sensing of hydrogenphosphate or other anions by using the chelate effect or an anion receptor linked to the fluorophore among many other approaches.[12-14] The influence of hydrogenphosphate is also important to have in mind if the fluorophore is used in biological research detecting the pH as there is significant phosphate concentration in solution under physiological conditions and in cells.

In the paper by Zelent et al. they have observed the same quenching properties of phosphate on pyrene-1-carboxylate when dihydrogenphosphate protonates the fluorophore.[15] They find a correlation of ΔpK and quenching efficiency between fluorophore and quencher which is as predicted by the Marcus theory. Initial results found during this project show the similar relationship between quenching and ΔpK.

Results and discussion

pH dependency of the phenol-DAOTA$^+$ and anisole-DAOTA$^+$

The phenol-DAOTA$^+$ in PBS and DMSO has previously shown a pH dependency of fluorescence and lifetime decay related to the deprotonation of the hydroxy group. This quenching is due to
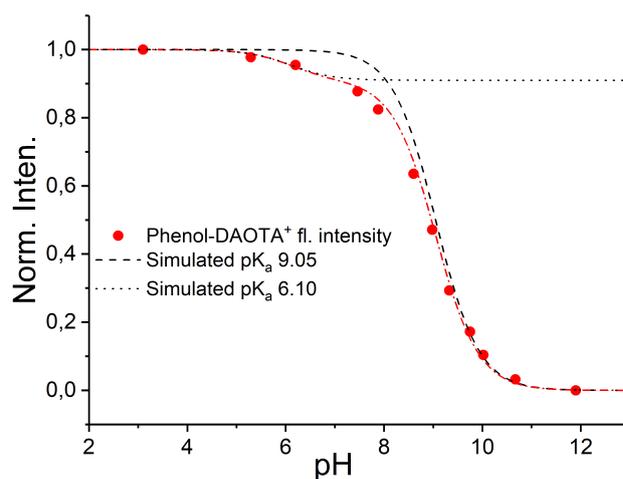
Figure 2: Normalized fluorescence intensity for the phenol-DAOTA$^+$ in PBS (red circles) showing a weak (black dotted) and a strong (black dashed) forms of quenching with pK$_a$ 9.05 and 6.10 respectively. The red line is the combination of the two simulations.

photoinduced reductive electron transfer (PET). But as described another 9% of the total quenching with a pK$_a$ lower than that of the quenching by PET was also observed. In this project these results were repeated with a slightly higher DMSO concentration as precipitation of the fluorophore otherwise was observed in basic pH. This is described further in the supplementary information section S3. The quenching of the fluorophore is shown in Figure 2 where the fluorescence, normalized with respect to the maximum intensity, shows the deprotonation related quenching with pK$_a$ 9.05 and the smaller quenching with an approximate pK$_a$ 6.10. Analysis of the lifetime Figure S8 left showed that the smaller quenching also resulted in a decrease in fluorescent lifetime. The absorbance spectra Figure S5 shows an isosbestic point which means that there is a pure conversion from the phenol to phenolate and no other reaction takes place as a side effect of the altered pH conditions. The normalized absorption intensities figure S6 were used to determine the pK$_a$ of the pH related deprotonation to be 9.05 as the absorption is only influenced by this and not the unknown quenching.

All the lifetime measurements showed monoexponential decays (section S4) indicating that there is only one fluorescent compound in solution. But due to the lifetime decrease being quite small at only 9% it can be challenging to choose the correct model for fitting because the difference will only be visible by single photon counts at times over 30 ns after excitation. These small variances are often hidden in the background noise. Another pitfall to be aware of is the fitting of biexponential decays, as the many variables give the possibility of fitting different amplitudes and lifetimes to the same dataset with little variance.[8, 16] This issue is elaborated further in the SI section S2. Having this in mind, it is not possible to conclude on the quenching mechanism being either dynamic or static using just the initial measurements.

As the pK$_a$'s of the quenching found from the lifetime data and the pH induced deprotonation are not similar it is possible that it is a quenching related to the solvent composition. The quenching of the lifetime cannot be related to the hydroxide concentration as [OH$^-$] keeps increasing with increasing pH. This would result in a continuous decrease of the lifetime but the data show that the
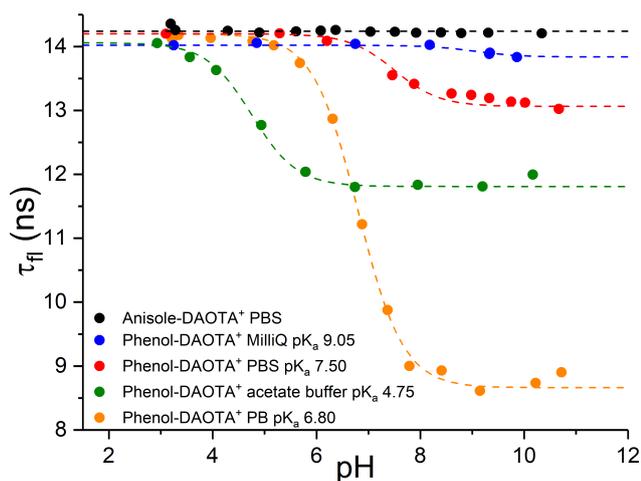
Figure 3: Fluorescent lifetimes of the phenol-DAOTA$^+$ in PB (orange), PBS (red), MilliQ (blue) and acetate buffer (green) showing the pH dependency in the different solutions. The anisole-DAOTA$^+$ (black) is shown as a constant lifetime reference.
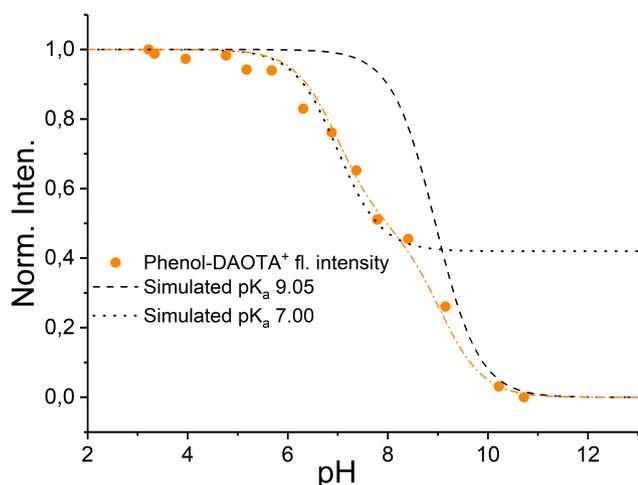
Figure 4: Normalized fluorescence intensity for the phenol-DAOTA$^+$ in PB (orange circles) showing a quenching with pK$_a$ 7.0 (black dotted) and another with pK$_a$ 9.05 (black dashed) forms of quenching. The orange line is the combination of the two simulations.

quenching reaches a saturation point at ~13 ns and does not decrease further. The anisole-DAOTA$^+$ was examined under the same conditions but did not show the quenching found for the phenol-DAOTA$^+$ both for the fluorescence intensity and lifetime as seen in Figure S9 and Figure 3. As there was a possibility of the methoxy group having lowered the donor HOMO orbital interfering with the mechanism, a more electron accepting fluorophore the anisole-ADOTA$^+$ was also examined in acidic and basic solution where the results are shown in Table S1 but no pH dependent lifetime was observed. Increasing the DMSO concentration to 30v% and measuring the fluorescent intensities and lifetimes (Table S1) did not show any effect either. These results rule out the possibility of the small quenching being hydrogen bonding, lowering the donor molecular orbital and thereby the PET efficiency which would result in a more fluorescent state than without hydrogen bonding.

Changing the buffer from PBS to phosphate buffer (PB) showed a significantly larger lifetime decrease at 36% of the longest lifetime
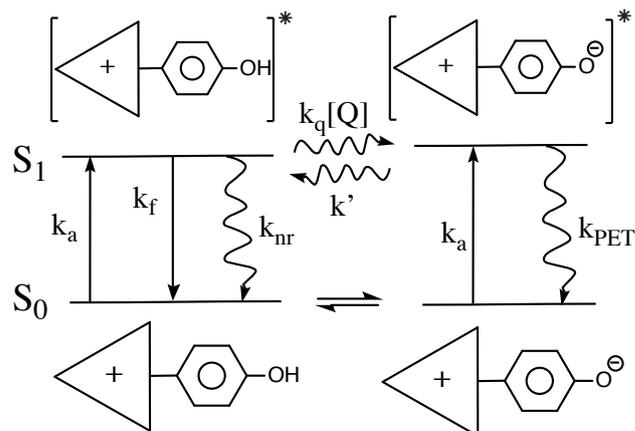
Figure 5: Jablonski diagram for the phenol-DAOTA$^+$ illustrating the ground state S$_0$ and first exited state S$_1$ and the processes affecting fluorescence. The constants k$_f$ and k$_{nr}$ are assumed constant and can be describes as one rate constant for deactivation k$_d$. As k$_{PET}$>>k' the phenolate is not protonated while in exited state which means a deprotonation by the quencher Q results in fluorescence quenching.

(Figure 3). This effect had a similar percentage fluorescence intensity quenching (Figure 4). The difference between the two buffers were a higher phosphate concentration and with no NaCl in solution giving reason to believe that it is an ion related quenching.

Ion related quenching

A control experiment was performed using no buffer but only MilliQ water adjusted to various pH values. Figure 3 shows only a very small decrease in lifetime with pK$_a$ 9.05. This determines that it is phosphate quenching the phenol-DAOTA$^+$ in a collisional process by deprotonating the fluorophore in its exited state resulting in immediate PET. In Figure 5 a Jablonski diagram shows the rate constants related to the excitation and deactivation of the fluorophore all having an influence on the fluorescent lifetime. All deactivation not related to the quencher is assumed constant as the triangulenium remains unchanged. The fluorescent lifetime can therefore be described as $\tau = 1/(k_d + k_q[Q])$, where k$_d$ is the rate of deactivation of the fluorophore $k_d = k_f + k_{nr}$. The pK$_a$ at approximately 6.80 shows that it must be hydrogenphosphate having the major quenching effect and not dihydrogenphosphate. By using the Henderson-Hasselbalch equation to calculate the concentration of HPO$_4^{2-}$ when the pH value in solution is known, it was possible to plot the data and fit it according to the Stern-Volmer equation 1.

$$\frac{\tau_0}{\tau} = 1 + k_q \tau_0 [Q] \qquad\qquad 1$$

The lifetimes are used, because the fluorescence intensity is also affected by the deprotonation quenching due to pH. The data show a good linear relationship as shown in Figure 6 which confirms that it is a dynamic quenching of the phenol-DAOTA$^+$ by HPO$_4^{2-}$. The total quenching of the phenol-DAOTA$^+$ can therefore be described as a combination of dynamic quenching by collision of ions with the phenol-DAOTA$^+$ in exited state and static quenching by pH dependent deprotonation to the phenolate in ground state.
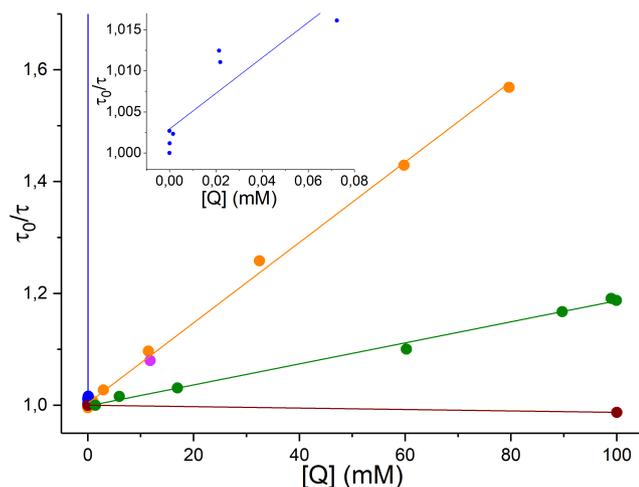
Figure 6: Stern-Volmer plot for the quenching ions. OH^- (blue) with a very steep slope has strong quenching abilities. The inset shows the data points for OH^- as the concentration is significantly smaller than for the other ions. HPO_4^{2-} (orange) and CH_3COO^- (green) also shows quenching properties. Cl^- (darkred) show no quenching with no change in lifetime between 0mM and 100mM, the third point from PBS at basic pH 11.8mM HPO_4^{2-} and 100mM Cl^- (violet).

Table 1: Properties of the anions in the examined solutions

Anion	k_q[a] ($M^{-1}s^{-1}$)	Diffusion rate[b] (m^2s^{-1})	pK$_a$ of conjugate acid[b]	ΔpK[d]
Cl^-	-	$2.032 \cdot 10^9$	-7/strong acid	-16.05
$H_2PO_4^-$	-	$0.959 \cdot 10^9$	2.16	-6.89
CH_3COO^-	$1.34 \cdot 10^8$	$1.089 \cdot 10^9$	4.756	-4.294
HPO_4^{2-}	$5.10 \cdot 10^8$	$0.759 \cdot 10^9$	7.21	-1.84
OH^-	$1.54 \cdot 10^{10}$	$5.273 \cdot 10^9$	13.995	4.945

[a] Rate constants calculated from linear regression of the Stern-Volmer plot. [b] Constants found in the Handbook of chemistry and physics 96th edition.[17] [d] Calculated as ΔpK= pK$_{a,acid}$ - pK$_{a,phenol-DAOTA}^+$, where pK$_{a,phenol-DAOTA}^+$=9.05

Selectivity

PBS, PB, MilliQ water or acetate buffer containing different ions were used as solvents to examine the quenching effects and efficiency of the various ions. The pH dependencies of the phenol-DAOTA$^+$ fluorescence lifetime are shown in Figure 3. Using the lifetime data acquired from the pH titrations a Stern-Volmer plot was generated. This is shown in Figure 6 and rate constants for quenching k_q are calculated and shown in Table 1. The initial lifetimes of the phenol-DAOTA$^+$ in the PBS (total phosphate concentration 11.8 mM) and PB (total phosphate concentration 100 mM) solutions at acidic pH where $H_2PO_4^-$ is the dominating species, were on the same order at ~14 ns as the lifetime measured in MilliQ water and the constant lifetime for the anisole-DAOTA$^+$. This indicates that $H_2PO_4^-$ have little to no quenching effect on the phenol-DAOTA$^+$. The large k_q for OH^- shows that the quenching by OH^- is diffusion controlled and that it is a highly efficient quencher, quenching the fluorophore at every collision. This could be due to the very high diffusion rate of the ion or the ability of transferring the proton through a chain of water molecules as described by the Grotthuss mechanism.[18] The reason the fluorescent lifetime only decreases by 1% is due to the very low concentration of OH^- in solution even at higher pH. Cl^- is shown by rough estimate to have no influence on the quenching of the

dye. $H_2PO_4^-$ have no influence on the phenol-DAOTA$^+$ as determined previously. It is therefore possible using the lifetime from PBS at very acidic pH and comparing to the lifetime in MilliQ water. Correlating the lifetime in PBS at more basic pH where $[HPO_4^{2-}]$ is near the total phosphate concentration to the Stern-Volmer plot of the PB titration at 11.8 mM also supports the claim that Cl^- has no shielding influence as the violet point Figure 6 shows no significant deviation from the phosphate titration. But a more thorough investigation is needed to establish certainty. The absence of quenching by Cl^- can be due to the ion having a pK$_a$ so low that it is too weak a base to be able to deprotonate the exited state phenol-DAOTA$^+$. Acetate also shows a 16% quenching of maximum for both fluorescence intensity Figure S16 and lifetime Figure 3. It is however smaller than the quenching by hydrogenphosphate which is also visible with the smaller slope in the Stern-Volmer plot and lower k_q constant. This variance could be due to the CH_3COO^- being singly charged and therefore not as attracted to the positively charged dye by electrostatic forces compared to HPO_4^{2-}. It could also be related to ΔpK$_a$ Table 1 between the base and dye. The diffusion constants mentioned in Table 1 show no correlation to the observed k_q's. However it is possible, that there is a relationship between the ion quenching efficiency and pK$_a$. The stronger the base, the larger k_q is observed. This is the same observation by Zelent et al.[15] where they found a linear relationship between ΔpK and log(k_q) corresponding to the Marcus theory. Using the k_q's (not taking the diffusion rates into account) and ΔpK's from Table 1 to create a preliminary log(k_q) vs ΔpK plot Figure S16 showed a good linear relationship.

Further work

To understand this mechanism further and to confirm the possible relation to the Marcus theory it would be beneficial to examine the fluorophore under other solution conditions containing other types of anions. This could be other organic or inorganic anions such as citrate, nitrate, sulphate or lactate. Competitional studies between hydrogenphosphate and another base would also be interesting as it could give more information on the mechanism and help determine if it is an actual collisional quenching or if the deprotonation also occurs by protons moving through a chain of bridging molecules. Another way to examine the mechanism and maybe improve the selectivity is adding sidechains next to the hydroxyl group either shielding the group or introducing a negative charge with deprotonating capabilities.

Conclusion

The photophysical properties of hydroxyphenyl DAOTA$^+$ were investigated to determine the cause for the extra quenching not related to the pH dependent PET mechanism. Results show that the fluorescence and lifetime quenching of the fluorophore previously observed were due to HPO_4^{2-} deprotonating the fluorophore in exited state in a collisional process. Other ions' quenching properties were investigated in connection with this finding showing that both OH^- and CH_3COO^- also quenches the dye with respectively higher or lower efficiency and that Cl^- has no effect. So far the data do not tell the exact reason for this difference in quenching abilities but an explanation could be the ΔpK$_a$ is having an influence which is also consisting to the Marcus theory. However further work is needed to reach a final conclusion.

Experimental section

All chemicals and solvents were used as received from Sigma-Aldrich. The buffer solutions were made following recipes from Cold Harbor Springs protocol or from AAT Bioquest.

5 The pH was measured on a calibrated Mettler-Toledo SevenEasy pH-meter after addition of DMSO to the pH adjusted buffer as the chemical caused the pH to shift unpredictably.

Absorption measurements were carried out on a Perkin-Elmer Lambda spectrometer in the 300-750 nm range. Fluorescence
10 measurement were carried out on an Agilent Technologies Cary Eclipse Fluorescence Spectrophotometer. Lifetime decay measurements were carried out on a FluoroTime 300 instrument from PicoQuant using a pulsed solid-state laser at 560 nm. The decays were analyzed using the FluoFit software from PicoQuant.
15 Further information can be found in the supplementary information.

30 ## References

Address: Nano-Science Center & Department of Chemistry, University of Copenhagen, Universitetsparken 5, Building 2 CS14, 2100 København Ø, Denmark.; E-mail: txd908@alumni.ku.dk

35 1. J. C. Martin and R. G. Smith, *Journal of the American Chemical Society*, 1964, **86**, 2252-2256.
2. L. B. W. and K. F. C., *Angewandte Chemie International Edition*, 2000, **39**, 3432-3434.
3. M. Rosenberg, K. R. Rostgaard, Z. Liao, A. O. Madsen, K. L.
40 Martinez, T. Vosch and B. W. Laursen, *Chemical Science*, 2018, **9**, 3122-3130.
4. I. Bora, S. A. Bogh, M. Santella, M. Rosenberg, T. J. Sørensen and B. W. Laursen, *European Journal of Organic Chemistry*, 2015, **28**, 6351-6358.
45 5. T. J. Sorensen, E. Thyrhaug, M. Szabelski, R. Luchowski, I. Gryczynski, Z. Gryczynski and B. W. Laursen, *Methods and Applications in Fluorescence*, 2013, **1**, 1-6.
6. BWLgroup, unpublished work.
7. R. Haugland, *The Molecular Probes Handbook*, Life
50 Technologies, 11th edn., 2010.
8. J. R. Lakowicz, *Principles of fluorescence spectroscopy*, Springer, 3rd edn., 2006.
9. M. Rosenberg, A. K. R. Junker, T. J. Sørensen and B. W. Laursen, unpublished work.
55 10. B. Wardle, *Principles and applications of photochemistry*, Wiley, 1st edn., 2009.
11. Z. Liao, S. A. Bogh, M. Santella, C. Rein, T. J. Sørensen, B. W. Laursen and T. Vosch, *The Journal of Physical Chemistry A*, 2016, **120**, 3554-3561.
60 12. M. E. Huston, E. U. Akkaya and A. W. Czarnik, *Journal of the American Chemical Society*, 1989, **111**, 8735-8737.
13. T. Gunnlaugsson, A. P. Davis and M. Glynn, *Chemical Communications*, 2001, **24**, 2556-2557.
14. P. D. Beer and P. A. Gale, *Angewandte Chemie - International Edition*, 2001, **40**, 486-516.
65 15. B. Zelent, J. M. Vanderkooi, R. G. Coleman, I. Gryczynski and Z. Gryczynski, *Biophysical Journal*, 2006, **91**, 3864-3871.
16. A. Grinvald and I. Z. Steinberg, *Analytical Biochemistry*, 1974, **59**, 583-598.
70 17. W. Haynes and D. Lide, *CRC Handbook of Chemistry and Physics*, Boca Raton, 92nd edn., 2015-2016.
18. C. J. T. de Grotthuss, *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 2006, **1757**, 871-875.

Paper

From Breast Cancer to Computational Science

Catalina Cocan

5 Computational science focuses on the development of predictive computer models of the world around us. Scientific computing methods have become more accurate in the discovering of tumor treatment possibilities, than laboratory experiments have ever proven. This thesis will provide the mathematical modeling of breast tumor growth, including different aspects of how the immune system together with cycle-phase-specific chemotherapy can defeat the disease. The delay differential equations
10 representing the tumor environment are solved using the Runge Kutta fourth method, taking in consideration different phases of the cell-cycle.

Submitted as full paper.

From Breast Cancer to Computational Science

Stud. Scient. Catalina Cocan,
Supervisor: Professor, Ph.D., Dr. Scient. Kurt V. Mikkelsen
Department of Chemistry,
University of Copenhagen, Denmark

June 15, 2018

Abstract

Computational science focuses on the development of predictive computer models of the world around us. Scientific computing methods have become more accurate in the discovering of tumor treatment possibilities, than laboratory experiments have ever proven. This thesis will provide the mathematical modeling of breast tumor growth, including different aspects of how the immune system together with cycle-phase-specific chemotherapy can defeat the disease. The delay differential equations representing the tumor environment are solved using the Runge Kutta fourth method, taking in consideration different phases of the cell-cycle.

Keywords : Computational science, Delay differential equation, Breast cancer, Immune system, Cell cycle, Cycle-specific chemotherapy, Quiescent cells, Proliferating cells, Paclitaxel(Taxol)

This thesis is dedicated to my grandmother, who lost the battle against cancer.

Acknowledgements

This thesis becomes a reality with the kind support and help of many individuals. I would like to extend my sincere thanks to all of them.

I would like to express my gratitude towards my beloved parents for the encouragement which helped me in the realization of this thesis.

I wish to express my sincere gratitude to my supervisor Professor, Ph.D., Dr. Scient. Kurt V. Mikkelsen, for providing me the great opportunity of being a part of his group.

I sincerely show appreciation for the guidance and persistence Professor, Ph.D., Dr. Scient. Kurt V. Mikkelsen provided with, for my principal research interest.

I would like to express my special gratitude to my adviser, Ph.D Student, Freja Eilsø Storm.

I also want to thank the rest of the group members for providing assistance that was needed to accomplish this thesis.

My thanks and appreciation also go to my colleague and people who have willingly helped me out with their abilities.

Contents

1 Introduction

Cancer is one of the leading causes of morbidity and mortality worldwide. The number of cancer patients has been alarming for the past years. In 2012 there were about 14 million new cases, and it is expected that a rise of about 70% will take place over the next 2 decades. In 2015 the number of deaths caused by cancer was around 8.8 million. The disease is taken seriously in various aspects of science, for the reason that it is categorized as the second leading cause of death in the world [1].

The most abundant cancer form among women is breast cancer. Because of the large interest in the aspect of breast cancer, the thesis will focus on this form of cancer [2]. Roughly one out of three Danes will experience some type of cancer in their lifetimes. About 60 percent of cancer patients will survive for more than five years after they have been diagnosed. The Danish reviews based on cancer patients do not look promising. Danish women got the Nordic record in all forms of cancer for the past 20 years, where breast cancer is one of the leading condition as mentioned above. Looking at the survival rates, about 61 percent of women are alive five years after their cancer diagnoses and this is unsatisfactory compared to the other Nordic countries. According to the World Cancer Research Fond from 2012, Denmark has 329 women cancer patients for every 100,000 residents. In 2015 there have been 64.405 breast cancer cases [3] [4] [5].

This thesis will focus on the numerical and mathematical analysis of data within cancer. It aims to gain understanding of the process in which cancer develops and to design and improve better treatment strategies to eliminate the disease or to improve the life of the patient. Numerical analysis implements the study of algorithm, which deals with numerical approximations for the problems of mathematical analysis[6].

The leading interest is to model the effects and interactions between tumor cells and immune cells, when cycle-phase-specific drugs, such as chemotherapy, are given to the patient. The systems will be subdivided into different cycle phases of tumor cells to control and model a more explicit method of how chemotherapy works.

To understand the complexity of biological systems, different mathematical models have been applied and manipulated throughout the years. Each new model and method contribute in its own way to a better understanding of cancer development and gives new insight on how a disease can be defeated. Some specific models will be discussed and taken into consideration throughout the thesis.

To understand the design of a system in which delay differential equations are solved, a simulation of real life is established and supported within the bounds of fundamental knowledge [7] [8] [9] [10].

The main topic of research in this thesis is how a tumor behaves at a cellular level. A rudimentary introduction about the disorder will give a better understanding of the mathematical model in a biological context. The biological systems are described and established into mathematical delay differential equations and solved by the iterative Runge Kutta fourth order method [6].

2 Theory

2.1 Breast Cancer, Immunological Response and Previous Studies

Cancer is a common term for a lot of types of diseases that can affect any part of the human body. It arises from the transformation of normal cells into tumor cells. The developing step has multiple phases in which a pre-cancerous lesion can evolve into a malignant tumor and can therefore be life-threatening. One way to define a tumor is the rapid formation of abnormal cells that grow beyond their usual boundaries [11].

Cancer cells only become apparent when the immune system weakens and the body's normal cells stop their usual cycle and take an abnormal path instead. When a cell changes its path, the immune system no longer recognizes and defeats it. Another consequence caused by the spontaneous change of path cycle is the fast reproduction of these new organisms, which have no boundaries. The phenomenon described is often called a malignant growth, in which cancer cells during their

division mutate even more so that the immune cells no longer recognize their genetic material and the cancer cells invade the whole body. Cancerous cells do not necessarily divide faster than the other cells in the body, but they lose the ability to regulate the cell cycle. A consequence of this is that proliferation of these cells cannot be controlled [11] [12].

During the last years a considerable number of drugs, therapies, researches and other analysis have been studied, among which chemotherapy has been the most efficient. Chemotherapy is a way of treating specific cancer types by injecting it in the veins of the patient. Chemotherapy may be a curative goal treatment, it may aim to prolong the life of the patient or to reduce the wide variety of symptoms which are caused by cancer. As any other medication chemotherapy has its undesirable side-effects. Chemotherapy not only attacks the cancer cells, but in time it also attacks the body's own inborn cells, which are necessary in recognizing and defeating cancer [13] [14] .

The system discussed in this thesis is very complex, including the body's own concentration of immune cells and cancer cells regard to the dose of chemotherapy a patient receives. The system has four individual paths of a cells cycle, the 'pre-synthethic phase' – G1 phase, the 'synthetic phase' – S phase, the 'post-synthetic phase' – G2 phase and mitosis. To make the model more manageable the G1, G2 and S phase are categorized as the interphase stage in which the division of a cell is prepared. Interphase is the phase in which a cell lives most of its life. It is the stage in which the cell copies its DNA and prepares for mitosis. In this phase the cell develops, obtains nutrients and metabolizes them so it can control other cell functions. Once mitosis is completed each daughter cell can enter the cycle again or shift into a quiescent phase [8] [15] [16].

Malignant cells as well as benign cells can become quiescent. Quiescence is an important phase in the development of cancer cells, also called G0 phase. The quiescent phase describes a cellular state outside of the replicative cell cycle. The definition of quiescent cells could be a reversible absence of proliferation. This means a cell is not dividing, but it can start dividing if the appropriate cell conditions are going to be present. Quiescent cells must be stimulated to reenter the cell cycle. Some of the cells never enter this phase, while others never leave this phase. Cells in quiescent phase are resistant to cytotoxic agents such as chemotherapy, and this is an important attribute because chemotherapy is cytotoxic to both immune cells and tumor cells [15].

Quiescent tumor cells have shown to complicate the diagnosis and treatment of cancer patients, because of their resistance to chemotherapy. Even after a patient receives chemotherapy, quiescent cells persist in developing and rapidly resume their cell cycle. It has been discovered that non-proliferating cells, such as quiescent tumor cells, need another medication than chemotherapy to be defeated [17].

Another significant part of the cell cycle system are the cytokines. Cytokines are very important hormones in cell signaling. They are secreted by cells and can change the properties of the cell which secretes them or of other cells around. Cytokines are known for their substantial functions in the immune system, where they are used by B-lymphocytes to activate T-lymphocytes. Lymphocytes originate from the bone marrow and migrate later in different parts of the lymphatic system. There are different type of T-lymphocytes, but the ones which directly have an effect on cancer cells and are supposed to eliminate them, are the cytotoxic T-cells [18].
Tumor cells have however the competence to deactivate lymphocytes, such as cytotoxic T-cells, so in the end, the absence of cytotoxic T-cells can be fatal, because the immune system cannot combat the tumor cells anymore. These cytotoxic T-cells mentioned are taken into consideration throughout the thesis, because of their significant importance in defending the human body from foreign organisms. However one should be aware that the immune system is much more complex than the role of cytotoxic T-cells. A simpler approach often gives a better understanding of the whole assembly [19] [20] [21].

2.2 Numerical Solutions of Delay Differential Equations

Differential equations provide a mathematical language for describing continuous change. Most of the fundamental laws of science are expressed as differential equations. Even models of systems that do not change over time are often best understood as being in an equilibrium state of a rele-

vant differential equation. The state of a system at any given time t, is described by some vector function $y(t)$, where $y\colon \Re \to \Re^n$. The components of $y(t)$ represent the concentrations or amount of chemical substances or number of cells. Differential equations therefore describe a relationship between this unknown state function $y(t)$ and one or more of its derivatives with respect to t that must hold at any given time [6].

In order to solve a differential equation the objective is to determine a differentiable function $y(t)$ that satisfies the prescribed relationship. Finding a solution of a specific differential equation can predict the future evolution of a system in time. However it is not necessary to determine the differentiable function. In this thesis the values of the different populations were calculated by differential equations to represent the graphic of biological systems.

There are two main models which have been studied during the years and will also be discussed and analyzed in this report. Both models only have one independent variable, and that is time, t. The two models mentioned are described entirely in the following papers: Heuristic Design of Cancer Chemotherapies by M. Villasana and G. Ochoa and Dynamics Analysis and Limit Cycle in a Delayed Model for Tumor Growth with Quiescence by R. Yafia [8] [9]. Beside the two models mentioned above a third model is established, where these models are combined to improve the complexity of the system [7] [8] [9] .

The models which have been studied in the past few years describe either how tumor cells will behave in different stages of their life or how the application of the chemotherapy drug Paclitaxel helps in threating cancer patients [16] [22] [23].
To have an outcome of the problem, a mathematical interpretation is required to simulate reality. The mathematical design behind the formation of a computational implementation will be established with the help of delay differential equations.

A delay differential equation, also called DDE is a differential equation in which the derivative of the unknown function at a certain time is given in terms of the values of the function at previous times. Delay differential equations are applied for time-delay systems, in which the system is composed of some type of aftereffect. [7] [6]
Simple Delay differential equations have the form:

$$y'(t) = f(t, y(t), y(t - \tau_1), y(t - \tau_2), ..., y(t - \tau_k))\qquad(1)$$

where the time delays τ_k are positive constants.

A fundamental technique to solve delay differential equations is to reduce them to a sequence of ordinary differential equations. It is clear that time delays equations have made a huge impact in the biological science, such as the dynamics of a population, because of their predictive information. When considering time delay systems, the best example to pick, is the predator-prey model, which is the building block for biological-systems. The classical predator-prey model was suggested by Lotka and Volterra in 1926 for an ordinary differential equation [24] and has the form:

$$\frac{dx(t)}{dt} = a1 \cdot x(t) - b1 \cdot x(t) \cdot y(t)\qquad(2)$$

$$\frac{dy(t)}{dt} = -a2 \cdot x(t) + b2 \cdot x(t) \cdot y(t)\qquad(3)$$

The predator-prey model is composed of two populations, $x(t)$ which represents the population of prey and $y(t)$ which represents the population of predators. The sizes of $x(t)$ and $y(t)$ depend on the time t and have the following initial conditions:

$$x(0) = x_0 \ and \ y(0) = y_0.\qquad(4)$$

In the two equations (2) and (3) there are some positive constants: a1, a2, b1, b2 which represent different attributes of a system. These parameters can describe different significant parts of a system, such as natural birth rates, death rates or division of cells. If there is more than one

population $y(t)$ in the system, some of the parameters can indicate the interactions between the populations.

These systems are taken into considerations when the birth rate of predators is affected by prior levels of predators or prey, rather than by only the current levels in a predator-prey model. [24] Additional information is required to specify a system of delay differential equation. If the value of $x(t)$ does not immediately affects the value of $y(t)$ throughout the change in time and the other way around, then a delay differential equation must be taken into consideration.

$$\frac{dy(t)}{dt} = a1 \cdot x(t) - b1 \cdot x(t) \cdot y(t - \tau_1) \tag{5}$$

$$\frac{dy(t)}{dt} = -a2 \cdot x(t) + b2 \cdot x(t - \tau_2) \cdot y(t) \tag{6}$$

For DDE's, instead of simple initial conditions, initial history functions must be taken into consideration. The derivative in eq. (1) depends on the solution at a previous time $t - \tau_j$ and therefore it is necessary to provide an initial history function to specify the value of the solution before time $t = 0$.

$$x(0) = x_0, \quad x(s) = \phi(t), \quad y(0) = y_0, \quad y(s) = \varphi(t), \quad -\tau < t < 0 \tag{7}$$

where $\tau_1 > 0$ and $\tau_2 > 0$ and the functions $\varphi(s)$ and $\phi(s)$ are the initial past history functions. In this thesis an initial value of the functions is given in such a way that the program takes the first values in time and provides an initial history function to specify the value of the solution before time $t = 0$. The initial conditions are properly described in the three mathematical models [7] [6].

2.3   Mathematical Model - Runge Kutta Fourth Order

The mathematical method applied to solve the differential equations is Runge Kutta of fourth order [6]. Runge Kutta methods are single step methods which replace higher derivatives by finite difference approximations based on values of f at points between t_n and t_{n+1}. Runge Kutta method of 4'th order is derived by applying numerical integration to the following interval:

$$y_{n+1} - y_n = \int_{t_n}^{t_{n+1}} (f(t, y(t))dt) \tag{8}$$

Derivation of the Runge Kutta method happens with the help of the Taylor Series Methods [6], where the second derivative of y' is given by:

$$y'' = f_t + f_y \cdot f, \tag{9}$$

and every function will now be evaluated at (t, y), where f_t is the derivative of time, t, and f_y is the derivative of y.

Runge Kutta of 4'th order uses four estimates of the slope to give an approximation of greater accuracy than using a single slope at the beginning of an interval. The more estimates of a slope are taken into considerations, the more accurate the result are expected to be .

Therefore Runge Kutta of fourth order is a good approximation method for the complex system evaluated in this thesis. Runge-Kutta methods are easy to program, (for the reason that they do not provide any error estimate based on the step size chosen for the mathematical system) and to develop in time they do not need any history, and are therefore self-starting, so the value of step-size can be changed during the integration. Even if Runge-Kutta methods do not need history functions, the system can have another approach for which it is develops dependent of a time delay [6]. The method approximates the solution to a first order differential equation given by,

$$\frac{dy(t)}{dt} = y'(t) = f(y(t), t) \tag{10}$$

with initial condition

$$y(t_0) = y_0. \tag{11}$$

An initial value of the function must be specified to start the algorithm. To estimate the slope at time t_0, the following four slope approximations will be evaluated. The approximation to $y(t_0)$ will be defined as $y^*(t_n)$, for $n = 0, 1, 2...$.

$$k1_{rk4} = f(y^*(t_n), t_n) \tag{12}$$

$$k2_{rk4} = f(y^*(t_n) + k1_{rk4} \cdot \frac{h}{2}, t_n + \frac{h}{2}) \tag{13}$$

$$k3_{rk4} = f(y^*(t_n) + k2_{rk4} \cdot \frac{h}{2}, t_n + \frac{h}{2}) \tag{14}$$

$$k2_{rk4} = f(y^*(t_n) + k3_{rk4} \cdot \frac{h}{2}, t_n + h), \tag{15}$$

here h is the time step used for the development of $y(t_0)$ in time. Each one of the four slopes to the function, describes something specific [6].

The first approximation k1, is the slope at the beginning of the time step. The slope $k1_{rk4}$ will be further used to step halfway through the time step, and $k2_{rk4}$ will be the new estimate of the slope at the midpoint. Then if $k2_{rk4}$ is used to step halfway through the time step, $k3_{rk4}$ is then another estimate of the slope at the midpoint. The last slope used is $k3_{rk4}$ to step all the way across the time step, $k4_{rk4}$ will then be another estimate of the slope. To get the final estimate of the approximations, following equation is applied:

$$y^*(t_n + h) = y^*(t_n) + \frac{(k1_{rk4} + 2 \cdot k2_{rk4} + 2 \cdot k3_{rk4} + k4_{rk4})}{6} h = (y^*(t_n) + m \cdot h), \tag{16}$$

where m is a weighted average slope approximation [6].

3 Mathematical and Biological Methods

3.1 Cancer Method 1 - Tumor Cells Behavior during Mitosis & Interphase

The first model is presented by M. Villasana, G. Ochoa and A. Radunskaya in multiple papers and does not include quiescent tumor cells [7] [8].

For the mathematical models to work, the best method is to have a specific amount of data for each cancer patient and analyze the system of each patient individually. The set of parameters mentioned throughout the thesis belong to a cancer patient with uncontrollable tumor. The parameters are non-dimensionalized and are taken into consideration in all the simulations (*Appendices B-E*) . These parameters will be explained throughout the thesis [7] [8].

When considering a large populations of cells, ordinary differential equation can be applied to build a mathematical model of a certain cell cycle in the system. This model is a system of ODE's that describes the lymphocytes division.

$$\frac{N_0(t)}{dt} = -(\alpha_0 + \beta_0) \cdot N_0(t) \tag{17}$$

$$\frac{N_j(t)}{dt} = -2 \cdot \alpha_{j-i} \cdot N_{j-i} - (\alpha_j + \beta_j) \cdot N_j(t), \tag{18}$$

where $j = 1, 2...J$.
Where for j=1,2.., J, let $N_j(t)$ be the population of the lymphocytes at time t, after j divisions. J is the amount of cell divisions that are going to take place. α_j indicates the rate of cell proliferation and β_j indicates the rate of cell death. The first term in the previous equation: $2 \cdot \alpha_{j-i}$ indicates the fact that a cell during mitosis splits into two new daughter cells. Cell division is a discrete system and therefore very complex, in account of that there is involved a discrete time delay starting at the time the cell is born until it divides [10].

One of the drugs which has been used for treating cancer is Paclitaxel, also named Taxol. Paclitaxel interferes with mitosis, disabling the cell from continuing in the cell cycle. This happens by blocking the cell from entering the normal cell cycle and just stops the cell from proliferating and permits the immune system to eliminate cancerous cells in a natural way.

Chemotherapeutic medication not only defeats cancer cells, but also kills the body's own immune cells. Paclitaxel is relevant in this model because it is involved in the cell cycle which targets the S-phase, and is a part of the interphase. Paclitaxel has been shown to attack the tumor cells while they are in the interphase, but they only die at the time where they are in the mitoses phase, some days later. The resident time of cells in interphase is described by τ [22].

The time delay τ is only taken into consideration when talking about interphase. Mitosis is a very short period of the cell cycle, in which the time delay will be insignificant and that is the reason why it is not studied in this chase.

The system created separates the tumor cells into interphase cells $T_I(t)$ and mitosis cells $T_M(t)$. Beside these another term, $I(t)$ is introduced. $I(t)$ represents the population of the body's own immune cells, the cytotoxic T-lymphocytes. Because cancer cells do not die immediately, but only after several divisions, chemotherapy is given in repeated doses in different forms [23].

At first we have the three equations the systems is built upon [8]:

$$\frac{dT_1(t)}{dt} = 2 \cdot a_4 \cdot T_M(t) - c_1 \cdot T_I(t) \cdot I(t) - d_2 \cdot T_I - a_1 \cdot T_I(t - \tau) \tag{19}$$

$$\frac{dT_M(t)}{dt} = a_1 \cdot T_I(t-\tau) - d_3 \cdot T_M(t) - a_4 \cdot T_M(t) - c_3 \cdot T_M(t) \cdot I(t) - k_1(1 - e^{-k_2 \cdot w(t)}) \cdot T_M(t) \tag{20}$$

$$\frac{dI(t)}{dt} = k + \frac{\rho \cdot I(t) \cdot (T_I(t) + T_M(t))^n}{\alpha + (T_I(t) + T_M(t))^n} - c_2 \cdot I(t) \cdot T_I(t) - c_4 \cdot T_M(t) \cdot I(t) - d_1 \cdot I(t) - k_3((1 - e^{-k_4 \cdot w(t)}) \cdot I(t)) \tag{21}$$

With the two states $w_1(t)$ and $w_2(t)$ which are given by:

$$\frac{dw_1(t)}{dt} = -\lambda_1 \cdot w_1(t) + c(t) \tag{22}$$

$$\frac{dw_2(t)}{dt} = -\lambda_2 \cdot w_2(t) + c(t) \tag{23}$$

Which together collaborate to the second and third function, respectively $T_M(t)$ and $I(t)$:

$$w(t) = r_1 \cdot w_1(t) + r_2 \cdot w_2(t) \tag{24}$$

$w(t)$ is a linear combination of the two terms $w_1(t)$ and $w_2(t)$, and $c(t)$ is the term which describes the concentration of the chemotherapeutic drug, Paclitaxel, which is administered by the patient at time t. The drug decay is assumed to be exponential and λ is a coefficient that includes both the elimination and absorption impacts of the drug. The terms $d_2 \cdot T_I$, $d_3 \cdot T_M$ and $d_1 \cdot I$ represent proportions of natural cell death(apoptosis), while a_1 and a_4 represent the different rates at which cells reproduce. c_i terms represent the losses from conflict of immune cells with tumor cells [25] [26].

If high concentrations of chemotherapy are given, the drug hinders the tumor cells in mitosis phase, so the tumor cell is taken out of the cycle and cannot proliferate.
This system can be modeled by following terms: $k1(1 - e^{-k_2 \cdot w(t)})$ constitutes the impact of the drug on mitosis and $k3(1 - e^{-k_4 \cdot w(t)})$ constitutes the impact of the drug on the cytotoxic T cells. This means that high concentrations of chemotherapy will destroy the immune cells. Once the drug encounters the tumor cell, the tumor cell is taken out of the cycle and can no longer proliferate. If the drug dosage is too high the value of these terms will have another significance, because the drug will also eliminate the body's own immune cells [27].

Given the initial conditions, the previous functions can be calculated with help of delay differential equations:

$$T_I(t) = \phi_1(t), for \ t \in [-\tau, 0] \tag{25}$$

$$T_M(t) = \phi_2(t), for \ t \in [-\tau, 0] \tag{26}$$

$$I(t) = \phi_3(t), for \ t \in [-\tau, 0] \tag{27}$$

$$w_1(0) = 0 \tag{28}$$

$$w_2(0) = 0 \tag{29}$$

$T_M \cdot I$ and $T_I \cdot I$ are standard competition terms that represent losses due to encounters among the different cell types. The time decay taken into consideration is based on the drug Paclitaxel. There will be one decay rate in the bloodstream and one decay rate in the peripheral tissues when Paclitaxel is given [8] [22].
Following term:

$$\frac{\rho \cdot I(t) \cdot (T_I(t) + T_M(t))^n}{\alpha + (T_I(t) + T_M(t))^n}, \tag{30}$$

is the nonlinear growth of immune cells population due to the presence of a tumor [8]. The form in which this term is established is by a Michaelis-Menten model [28]. The reason for which this model is applied, is because proliferation of tumor-specific effector cells, such as immune cells, are stimulated by the presence of tumor cells. Immune cells will at a certain time reach a saturation level at tumor populations.

Another important term is the constant k which represents the birth rate of the immune cells in the absence of cancer. Tumor cells reside in interphase for a certain period of time τ before entering the mitosis phase. The term $T_I(t - \tau)$ accounts for the stage in which tumor cells leave the interphase. Assuming that cells reside in interphase τ units of time, then cells that enter the mitosis stage at time t are those cells that entered interphase at time τ.

The parameters ρ, α and μ depend on which type of tumor the system consists of and at which state the cancer exists. The growth of tumor cell population is obtained through the mitosis term and is given by the constants a_1, a_4 and τ which regulate the rate of cell division present in T_M. ϕ describes the increase of immune cells suitable to a stimulant. α represents the half value of the immune response. The two parameters are dependent on each other in such a way, that when the tumor level is equal to α the immune response is half way to its maximum value π.

The parameter n determines the shape of the response term and was examined in different *invitro* studies. Based on these studies the best value of the exponent was found to be $n = 3$ [8]. The same value of n is applied for all three models in this thesis. Large values of n means that the immune system has difficulties recognizing the tumor and it takes the immune system longer then expected. τ describes the time delay of the number of days in which the tumor cells reside in the interphase stage [7] [7] [9].

3.2   Cancer Method 2 - Quiescent Tumor Cells Behavior

The second model is based on the article *Dynamics Analysis and Limit Cycle in a Delayed Model for Tumor Growth with Quiescence* which generates a delay differential equation model for the interactions of proliferating and quiescent tumor cells, without including the immunological or medicational components [9].

Drugs that have been used for threating cancer have also affected the other cells in the organism and prevented their development and functions. As mentioned earlier quiescence is the state of a cell, in which the cell does not divide. It is not guaranteed that it will improve the understanding of cancer treatment, but some organ have a higher amount of quiescent cells than proliferating cells, therefore it is important to understand the development of different types of cancers situated in different sections of the human body [27] [29].

All existing drugs mainly affect the proliferating cells but also, to a certain extent, the quiescent cells.

The idea of not including the drug terms in this model is to compute some fixed points of the cancerous environment in which the chemotherapeutic drugs are not present. This contributes to the understanding and further analysis of the drug-free system. Beside the absence of drugs, the model will not take account of the immune response either [9].

The mathematical model used for this approach is the following:

$$\frac{dP(t)}{dt} = b \cdot P(t - \tau) - r_P \cdot (N(T)) \cdot P(t) + r_Q \cdot (N(t)) \cdot Q(t) \tag{31}$$

$$\frac{dQ(t)}{dt} = r_p \cdot (N(t)) \cdot P(t) - \mu_Q \cdot Q(t) - r_Q \cdot (N(t)) \cdot Q(t) \tag{32}$$

$$P(t) \equiv P_0, \ Q(t) \equiv Q_0, \ -\tau < t < 0 \tag{33}$$

Where $P(t)$ indicates the number of proliferating tumor cells and $Q(t)$ indicates the number of quiescent tumor cells. $N(t)$ indicates the total number of tumor cells, which is build up by $N(t) = P(t) + Q(t)$. Some other parameters are taken into consideration, where if $\beta > 0$ is the division rate of the proliferating tumor cells, while $\mu_p > 0$ is the death rate of the proliferating cells, then $b = \beta - \mu_p > 0$ is the intrinsic growth rate of the proliferating cells.

As mentioned quiescent cells do not divide, so there is no need to have a division parameter for them. $u_Q \leq 0$ represents the mortality rate of these cells, $r_P(N)$ incorporates the transition of proliferating cells to quiescent cells and $r_Q(N)$ incorporates the transition of quiescent cells to proliferating cells. The time delay describes here the time it takes proliferating cells to divide and there is of course no time delay for the quiescent cells because they are in a state with no division. The term $\mu_Q \cdot Q(t)$ is the term which describes the natural death of quiescent cells. The system has been tested by Yafia [9] both with and without time delay. If there is no time delay the system will become an ordinary differential equation model, where the time delay is $\tau = 0$ [9].

3.3 Cancer Method 3 - Combination of Method 1 & 2 and the Impact of Paclitaxel

The last model is a combination of the two models used in the two papers written by Villasana together with Ochoa and Yafia. Some additional terms are taken into consideration to ascertain the impact of the chemotherapy drug, Paclitaxel, on the quiescent cells [22] [23].

This system is based on the data from the same patient but with perspective on the body's own defense, immune cells, quiescent cells together with the chemotherapeutic drug treatment. Another important detail included is the combination of cytotoxic T-cells with quiescent tumor cells. Quiescent tumor cells are resistant to cytotoxic agents, so a specific parameter representing the resistance of these tumor cells must be established.

The term $T_Q(t)$ describes the population of tumor cells in the quiescent stage of their cell cycle. The term $N(t)$ describes the total cancer cell population. In the equations from model 2 there are few terms which are kept under control in this model, among other the transition states $r_P(N) = a5$ and $r_Q(N) = a6$ which in this chase are constants. $d4$ is another constant which represents the natural death rate of the quiescent tumor cells, while $c5$ represents the kill rate of quiescent cells by the cytotoxic T-cells.

As in the first model a chemotherapeutic drug treatment has to be represented in the mathematical model of the delay differential equations. Paclitaxel is a cycle specific chemotherapeutic drug, which works by stopping cancer cells from separating into two new cells. This blocks the growth of the cancer, because it targets the tumor cells in their interphase stage. Like any other medicine, Paclitaxel can have side effects, and one of the most dangerous side effects is drug overdoses. This will lead to damaging and killing of the healthy inborn and immune cells [21] [29].

The final mathematical model simulates the following:

$$\frac{dT_Q(t)}{dt} = a_5 \cdot T_I(t-\tau) - a_6 \cdot T_Q(t) - d_4 \cdot T_Q(t) - c_5 \cdot I(t) \cdot T_Q(t) - u_1(t) \cdot T_Q(t) \qquad (34)$$

$$\frac{dT_I(t)}{dt} = 2 \cdot a_4 \cdot T_M(t) - a_5 \cdot T_1(t-\tau) + a_6 \cdot T_Q(t) - c_1 \cdot T_I(t) \cdot I(t) - d_2 \cdot T_I(t) - a_1 \cdot T_I(t-\tau) \quad (35)$$

$$\frac{dT_M(t)}{dt} = a_1 \cdot T_I(t-\tau) - d_3 \cdot T_M(t) - a_4 \cdot T_M(t) - c_3 \cdot T_M(t) \cdot I(t) - u_2 \cdot T_M(t) \qquad (36)$$

$$\frac{dI(t)}{dt} = k + \frac{\rho \cdot I(t) \cdot (T_Q(t) + T_I(t) + T_M(t))^n}{\alpha + (T_Q(t) + T_I(t) + T_M(t))^n} - c_2 \cdot I(t) \cdot T_I(t) - c_4 \cdot T_M(t) \cdot I(t) - c_6 \cdot T_Q(t) \cdot I(t) - d_1 \cdot I(t) - u_3(t) \cdot I(t)$$
$$(37)$$

Same as in model 1, $w(t)$ (eq.24) is a linear combination of the states $w_1(t)$ (eq.22) and $w_2(t)$ (eq.23). For a better understanding of model 3, the individual parameters are elaborated earlier in *Cancer method* 1.

Parameters a_1 and a_4 represent the fraction of cells which cycle from interphase to mitosis and from mitosis to interphase. Both constants need to have a value between 0.2 and 1.0 per day (usually between 0.7 and 1.0).

Constants d_1, d_2 and d_3 represent fractions of natural death, also called apoptosis, and should be between the value of 0.1 and 0.3. Constants c_i model the losses of the cells due on the encounter with another cell (usually between 0.1 and 0.3).

The function must be evaluated together with some initial conditions,

$$T_Q(t) \equiv 0.8, for\ t \in [-\tau, 0] \qquad (38)$$

$$T_I(t) \equiv 1.3, for\ t \in [-\tau, 0] \qquad (39)$$

$$T_M(t) \equiv 1.2, for\ t \in [-\tau, 0] \qquad (40)$$

$$I(t) \equiv 0.9, for\ t \in [-\tau, 0] \qquad (41)$$

The function describing the importance of cytotoxic T-cells, $I(t)$, has a similar term as in model 1, but here the quiescence is taken into account. Here the term is a nonlinear growth of immune population, when cancer is present in the patient:

$$\frac{\rho \cdot I(t) \cdot (T_Q(t) + T_I(t) + T_M(t))^n}{\alpha + (T_Q(t) + T_I(t) + T_M(t))^n} \qquad (42)$$

To make sure the last model can be applied for further studies, some controls from already existing literature [8] are taken into consideration, where the following applies:

$$u_1(t) = k_5 \cdot (1 - e^{k_6} \cdot w(t)) \qquad (43)$$

$$u_2(t) = k_1 \cdot (1 - e^{k_2} \cdot w(t)) \qquad (44)$$

$$u_3(t) = k_3 \cdot (1 - e^{k_4} \cdot w(t)) \qquad (45)$$

$u_1(t)$ is a term used in eq.(34) and describes the tumor cells in quiescent stage. $u_2(t)$ is a term used in eq.(36) and describes the change of tumor cells in interphase stage. While $u_3(t)$ is the term used in eq.(37) to describe the population of the immune cells, which means the cytotoxic T-cells. All the following constants are positive and under the value 1: a_1, a_4, a_5, a_6, d_1, d_2, d_3, d_4, c_1, c_2, c_3, $c4$, c_5 and c_6. Some of the parameter values are the same as in some of the previous studies. There is no unique set of parameters values for any given model, because every parameter varies between tumor types and from patient to patient [7]. Some of the constants listed here under are used for the numerical values, but there are some other parameters that have been chosen without benefit of data, this is the reason why I allowed myself to vary some of the values for purpose of analysis, leading to a deeper understanding of the behavior of the model [8] [9].

Like discussed in previous studies, it has been discovered that quiescent tumor cells are resistant to drugs, but up till a particular moment, where they exit the quiescent stage and reenter the cell-cycle [29] [32].

$u_1(t)T_Q(t)$ is the term which takes into account the loss of quiescent cells. If $k_5 = 0$ it means that the quiescent cells present in the system are not affected by the medicine, because they are cancerous. Even if the human body is attacked by the tumor, the immune system still works in some way, so the immune cells will still destroy some of the quiescent tumor cells.

$c_5T_I(t)T_Q(t)$ is the term which describes the loss of quiescent tumor cells caused by the immune system. $c_5I(t)T_Q(t)$ describes the deactivation of immune cells caused by the quiescent tumor cells. If $c_6 = 0$ it is possible to see how the systems develops if the quiescent cells do not deactivate the immune cells. Looking back at Yafia's model [9] it includes another term $bP(t-\tau)$, where the term b is the intrinsic rate of proliferation, this means $b = birth - death$.

The time delay corresponds to the time it takes the cells to proliferate. The birth rate and death rate are getting separated, because death has logically no time delay. $a_6T_I(t-\tau)$ is the term which describes the rate of change of tumor cells during mitosis. This models the proliferation of cells, together with the delay corresponding to the time the cells spend in the interphase stage before they replicate. $d_3T_M(t)$ is the term which represents the instantaneous death of the mitotic tumor cells, that is when the tumor cells are in the mitosis stage of the cell cycle [7] [8] [9].

4 Computational methods

4.1 Script and Considerations

One of the challenges for this thesis was to acquire a knowledge of programming in python 2.7 [30] [31].

I managed to create a script that calculates the detailed mathematical methods mentioned in the theory section (*Appendix B-D*) and the new method based on the latest discoveries (*Appendix E*).

Four different codes were implemented, one for each model. All the four scripts have the same course of action and make use of the mathematical method Runge Kutta of fourth order.

The way in which the code has been designed, is by creating user-defined functions for the differential equations of the different populations. Such a function describes a block of organized, reusable code that is used to perform a single, related action. Depending on which mathematical method is used, diverse functions were analyzed.

Every function depends on some specific parameters within the area of interest. One of the functions created, states the computation of how Runge Kutta of fourth order operates on the different cell population functions. Every iteration starts from a specific initial value of the cell populations. The program specifies that the first values of the iterations, n_{steps}, remain untouched for a short period of time, so a history function of the delay can be created as it is requested in the differential equations for the term $(t - \tau)$.

Every population has an initial condition from which a specific differential equation is evaluated. The script returns values that are determined by the number of iterations the program runs. The number of iterations describe the number of days in which the populations develop in time. Another specification used by the program, is the time-step 0.01 chosen for the program. The time-step is extremely important for the design of a program, because it says something about the time interval for which a simulation will progress during next "step".

The initial values applied for the systems have either been calculated mathematically or chosen arbitrary to see how sensitive the script is.

The parameters on which the script operates depend on the patient, but the initial values are chosen in order to understand the influence of tumor cells, immune system and chemotherapy in relevance to each other.

The parameters used in the scripts are non-dimensional, for this reason the y axis varies from one

graph to another, especially in Fig. 1.a and 1.b. For a better understanding of the population values, the individual parameters could be multiplied with a whole number, to receive higher values. In this way the population could be compared with a concentration of the number of cells.

5 Numerical Results & Discussion

5.1 Results for Cancer Method 1 - *Appendix B*

This section aims to focus on model 1, which has been studied among the years by various scientists. To make sure model 1 functions as instructed, the values of the three populations, interphase tumor cells $T_I(t)$, mitosis tumor cells $T_M(t)$ and cytotoxic T-cells $I(t)$ are the same as for the results from past research. The reason for which I chose to use same values as in the earlier studies is to check if the mathematical methods I apply have same output [8]. With the initial function values $[T_I(0), T_M(0), I(0)] = [0, 0, 0.9]$, where if $t \leq 0$ the cancer cells are not present, therefore interphase cancer cells or mitosis cancer cells should remain zero, while the immune cells approach a steady state. For the first analysis the script will run through 100 days.

The first analysis, (fig. 1.a), gives an overview of the system when neither cancer cells nor medication are present. Only the cytotoxic T cells develop during the first analyses approaching a steady state value of roughly 0.12 after 10 days, while the tumor cells in interphase and mitosis remain at zero. The cytotoxic T-cells depend on the constant k, which is the birth rate constant of the body's immune cells when the tumor is absent. I can hereby conclude that a low value of tumor cells will let the cytotoxic T cells develop. Because if $T_I(0)$ and $T_M(0)$ equal zero, then these terms will simplify $I(0)$ eq. (21) and only take into account the number of the immune cells. The reason for which the cytotoxic T-cells decrease is because of the operational sign, because for the initial values in fig 1.a, $dI(t)/dt = k - (d_1 * I(t))$.
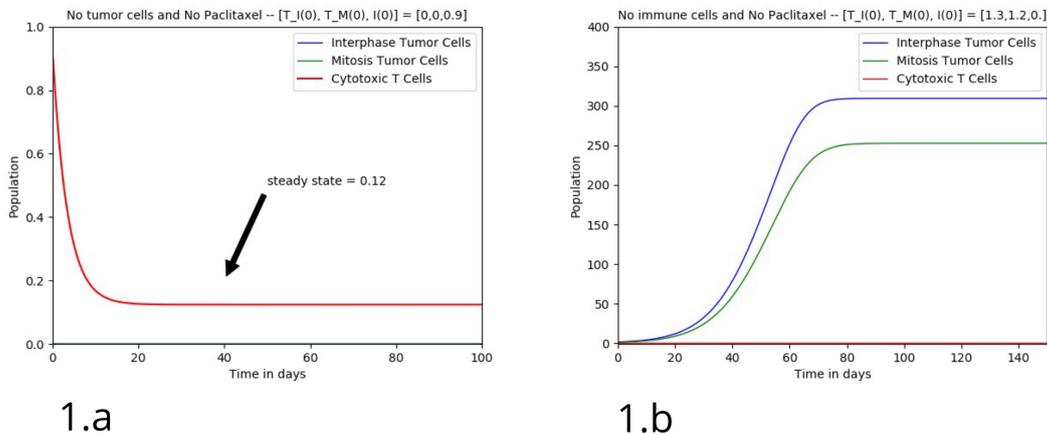
Figure 1: 1.a No tumor cells & no drug, 1.b No immune cells & no drug.

To understand the development of the tumor alone, another script (fig. 1.b) was established in which the immune cells are absent and the patient does not receive any medication.
As expected the cytotoxic T-cells do not develop throughout the analysis, while the interphase and mitosis tumor cells develop relatively fast in the beginning and then after 60 days both populations approach a steady state. Even if it is not conventional to have such a weak immune system, it gives a great understanding of the importance the immune system has in defeating a disease.

The next analysis (fig. 2.a) is based on the primary models studied by Villasanna & Ochoa, in which all three populations decrease rapidly the first 10 days, for after to approach a steady state between 0.1 and 0.2 depending on the population.

In this analysis no medication is administered to the patient [8].

In the simulations where no drug is administered the following terms are taken out of the calculations: $(-k_1(1 - e^{-k_2 \cdot w(t)})) \cdot T_M(t)$ and $(-k_3((1 - e^{-k_4 \cdot w(t)}) \cdot I(t)))$ because they represent the kill terms of the drug on mitosis tumor cells and the cytotoxic T-cells, respectively. So by removing these terms the states $w_1(t)$ and $w_2(t)$ will not be taken into consideration either.

Figure 2: 2.a No drug - $I(0) = 0.9$, 2.b No drug - $I(0) = 0.43$, 2.c No drug - $I(0) = 0.48$

To understand the function of the immune system, the start values for both tumor cells populations is kept the same, but the start value of cytotoxic T-cells will be changed to 0.43 and 0.48 (fig 2.b and fig 2.c). With a start value of 0.9 for the cytotoxic T-cells population (fig. 2.a), the tumor is reduced for a shorter period of time, but a lower initial value of the immune system (fig 2.b & 2.c) may influence the patient's capability to defeat the tumor.
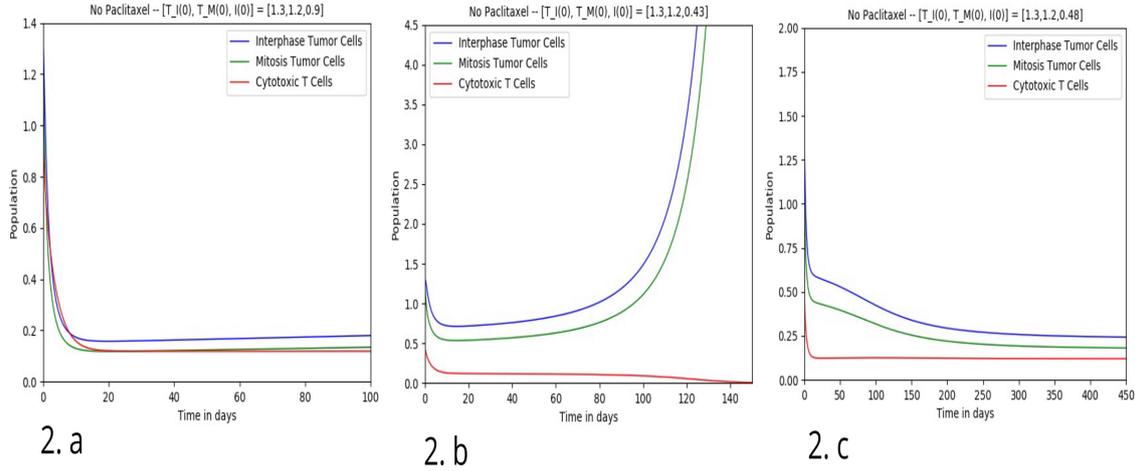The start value of the cytotoxic T-cells population for the first analysis is 0.43 (fig. 2.b) and for the other analysis 0.48 (fig. 2.c). The reason for which fig. 2.c spans over 450 days is to see the development for a longer period, cause for shorter periods the steady states of the three functions were not visible. For figure 2.b it was not necessary to show the development over a longer period of time, because the tumor cells kept increasing. The importance of the days is not significant here as much as it is for the simulations in which chemotherapy is given.

It is expected that with low start concentrations of cytotoxic T-cells, the tumor cells develop much faster, because the immune system is not capable to defeat the tumor cells. Fig. 2.b is a simulation of the system over 150 days, in which the tumor cells increase in population quite fast, while the immune cells decrease more and more.

Comparing the three simulations (fig. 2.a - 2.c) I can observe that patients with a weak immune system are more likely to die from cancer, while patients with a stronger immune system are able to battle the disease for some amount of time. A patient with a cytotoxic T-cell start population of 0.9 is categorized to be in a balanced condition, while a patient with an initial value of 0.48 or under is classified as a person with a weakened immune system.

To understand the function of chemotherapy on tumor cells, a new system is established(fig. 3.a - 3.c), in which Paclitaxel is given to the patient for certain days. The time in which the patient receives medication is given by, $c(t) = 1$ if $0 \leq t \leq 10, 20 \leq t \leq 30$ and $50 \leq t \leq 60$ and $c(t) = 0$ in the other chases. Around time $t = 0, 20, 50$ the influence of Paclitaxel is apparent(fig. 3.a & 3.c), where the population of both interphase and mitosis tumor cells decrease for some time, while the cytotoxic T-cells approach a steady state.

The reason for which the mathematical methods simulate short periods of time in which chemother-

apy is given, has to do with the damaging characteristics of chemotherapy on long term.
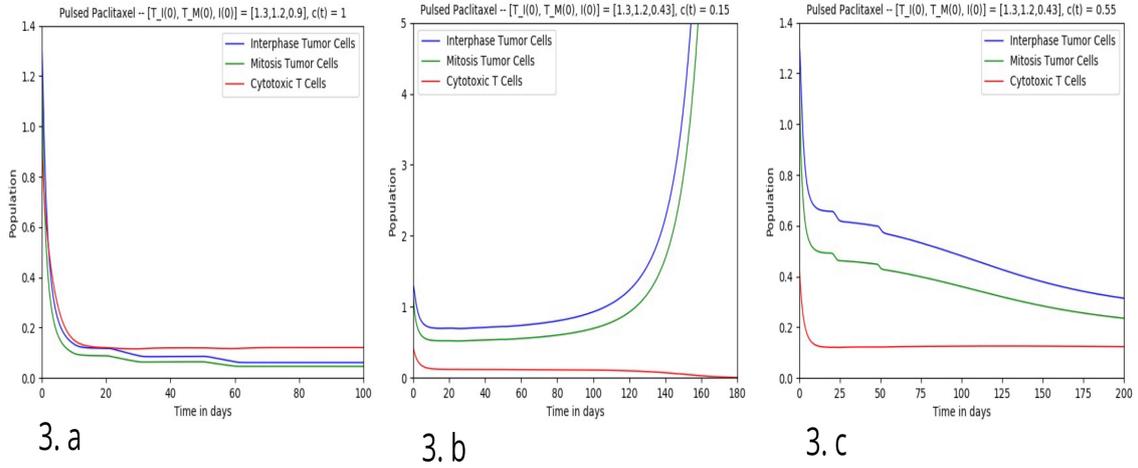
Figure 3: 3.a Pulsed drug - c(t) = 1, 3.b Pulsed drug - c(t) = 0.15, 3.c Pulsed drug - c(t) = 0.55

To have in mind, chemotherapy not only defeats the tumor cells, but also kills the healthy inborn cells [16]. Fig. 3.b represents the case in which the immune system is extremely low in comparison to the population of the tumor cells, and the amount of drug received is only $c(t) = 0.15$. In fig. 3.c the initial value of cytotoxic cells is 0.43, with the following inputs: $c(t) = 0.55$ for $0 \leq t \leq 3, 20 \leq t \leq 23, 48 \leq t \leq 50$ and $c(t) = 0$ in the other chases (eq. 22-23). It is clear in fig. 3.b that even when the patient's immune system is very weak, the low dosage of chemotherapy will still make a difference and defeat the tumor cells for a time.

Another thing to consider is the importance of the chemotherapeutic concentration that the patient receives. Will chemotherapy still work if the dosage is smaller, because we know that if the dosage gets too high it only damages the body more. What if the dosage gets smaller then $c(t) = 0.55$, will it still work?

The time in which the patient receives chemotherapy should be considered as well. Chemotherapy is damaging for the body and spreading the treatment for a longer evenly distributed period could have an impact [16]. Like any other disease the sooner you discover it, the easier and more approachable it is to cure. So the importance of receiving medication right after the disease is discovered, is essential. To prove this statement another analysis is established (fig. 4.a - 4.c), in which additional low level pulses are added. For this analyses the drug concentration is either $c(t) = 0.15$ or $c(t) = 0.30$, and beside the earlier three pulses an additional pulse is taken into consideration. First, a early additional pulse $7 \leq t \leq 9$ is considered for which the drug concentration equals $c(t) = 0.15$.

Figure 4: 4.a Early pulsed Paclitaxel & c(t) = 0.15, 4.b late pulsed Paclitaxel & c(t) = 0.15, 4.c late pulsed Paclitaxel & c(t) = 0.30
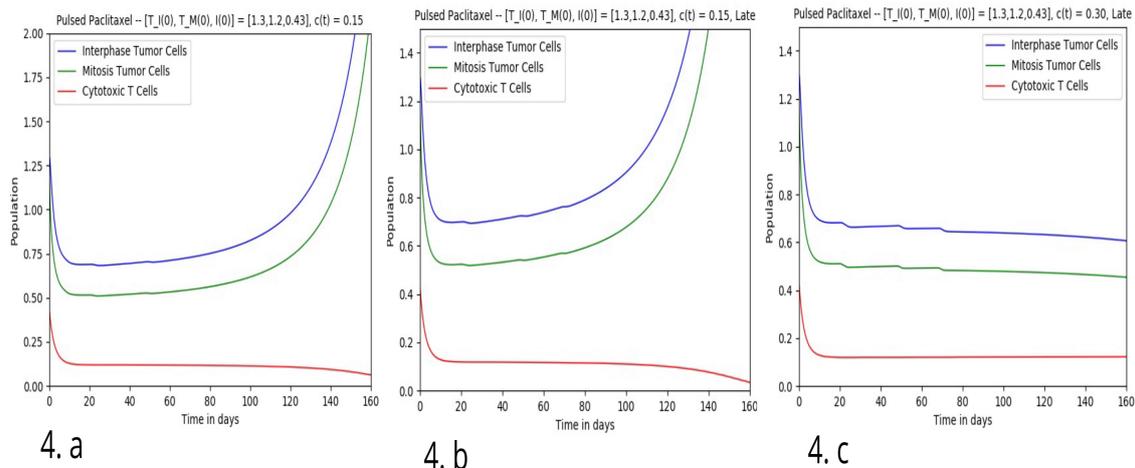
Tumor cells react to the chemotherapy received, but only for a short period of time of about 10 days. To prove that it is significant to threat the disease as soon as possible, the low level pulse added before is replaced with a later low level pulse at time $68 \leq t \leq 70$ with a dosage of $c(t) = 0.15$ (fig 4.b) and $c(t) = 0.30$ (fig 4.c). Like expected additional chemotherapy pulses will reduce the population of tumor cells, but even if only a small additional dosage is prescribed to the patient, it can be damaging for the patient's healthy cells, which are sensitive to chemotherapy.
Fig. 4.c shows clearly that a higher, more frequent dosage of chemotherapy kills the tumor cells, but may kill some other cells too.

Chemotherapy dosage is as important as the amount of times the drug is administered to the patient. A frequent low dosage of chemotherapy may be more efficient then a high rare dosage of chemotherapy.

5.2   Results for Cancer Method 3 - *Appendix D*

I chose to present the results for *model 3* and only mention *model 2* in *Appendix A and C*, because of the outcome from the numerical results. *Model 2* displays unphysical results, because of the insufficiency and lack of precision that the method provides.

As for model 1 (fig 1.a), a simulation only including the cytotoxic T-cells was implemented (fig. 5.a) and the results appeared to be the same. The way the immune cells developed and the steady state value showed to be the same.
The simulations for model 3 which I am going to discuss have nearly same onset as model 1, but to compare the two models I need to have same pattern of the immune cells and cancer cells alone when neither the drug nor the immune cells are present.

To make sure the code works as it should, to be compared to real life, there are some initial values that have to be tested. For the first run(fig 5.a) all three populations have same initial values as for model 1 in which $[T_Q(0), T_I(O), T_M(0), I(0)] = [0, 0, 0, 0.9]$ when chemotherapy is absent.

For further analysis I consider what influence the quiescent tumor cells have on the simulation by either given it the initial value 0 (fig. 5.b) or 1.8 (fig. 5.c). The quiescence component of the differential equations is present in the interpretation of the other tumor cells, as well as the immune systems cytotoxic T-cells, therefore it is important to see how the population of quiescent tumor cells develops in time.
This simulation takes into account chemotherapy with a dosage of $c(t) = 0.55$ at times $0 \leq t \leq 10, 20 \leq t \leq 30, 60 \leq 70$ and $c(t) = 0$ in the other cases, first where quiescence is absent

14

$[T_Q(0), T_I(O), T_M(0), I(0)] = [0, 1.3, 1.2, 0.4]$ and second where quiescence is present $[T_Q(0), T_I(O), T_M(0), I(0)] = [1.8, 1.3, 1.2, 0.4]$.
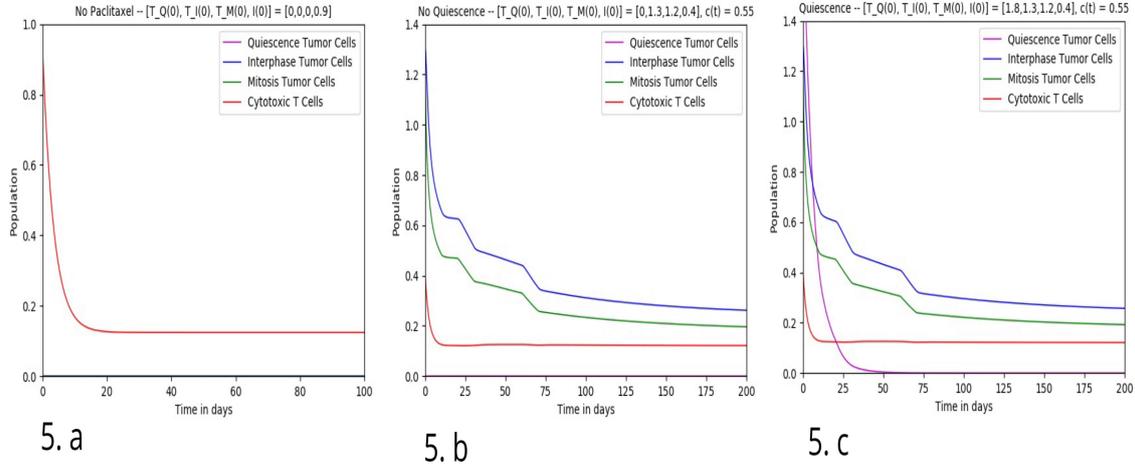

Figure 5: 5.a No Paclitaxel - only Cytotoxic T-cells, 5.b No quiescence cells - c(t) = 0.55, 5.c Quiescence present - c(t) = 0.55

In model 1 the initial value for the cytotoxic T-cells population was 0.9, but it did not show any significant change, because the value represented a patient with an extremely strong immune system. For this reason the initial value 0.4 is based on a patient with a weak immune system like in fig. 3.a - 3.c. From fig. 5.b and 5.c there is no significant change to be observed, beside the value of quiescent cells. This is because of the way the differential equations are build. The quiescent term in the equations is not of big importance, so only looking at the graphs will not give a wide perspective of the importance of the quiescent tumor cells, but for further research the small change in the values may be significant. A more detailed system based on quiescent tumor cells could be a step forward in discovering the importance of the quiescent stage.

To understand the meaning of how high or low the initial values of the different population concentrations should be, three simulations with same initial function values are tested (fig. 6.a - 6.c). Where $[T_Q(0), T_I(O), T_M(0), I(0)] = [1, 1, 1, 1]$, $[T_Q(0), T_I(O), T_M(0), I(0)] = [0.1, 0.1, 0.1, 0.1]$ and $[T_Q(0), T_I(O), T_M(0), I(0)] = [0.01, 0.01, 0.01, 0.01]$. As shown in the previous examples and model 1, the value of the cytotoxic T-cells $I(0)$ is quite important. A high starting value of tumor cells population and a low starting value of cytotoxic T-cells population will only make the tumor cells develop more. While high initial values of cytotoxic T-cells population will decrease the tumor for some time and then find a steady state for the system. In the first case where all the initial values are 1 (fig. 6.a), all three populations go towards zero. This means that even if the starting population of the tumor cells is high, the high initial value of cytotoxic T cells is able to defeat the tumor for some time and stabilize it afterwards.

By lowering the initial values to 0.1 (fig. 6.b), the quiescent cells will go towards zero, the tumor cell populations in mitosis are defeated for a very short period of time from day 0 to day 3, but then start growing again, while the interphase tumor cells will increase in population for most of the time. The cytotoxic T-cells population increase and eventually find a steady state, because they are still able to develop and divide when the population of tumor cells is so low. In the last case where all the initial values equal 0.01 (fig. 6.c), even if the starting value of cytotoxic T-cells is low, the immune system value is still able to develop in time and increase, while the tumor cells population will increase a little and find a steady state or decrease towards zero.

6. a

6. b

6. c
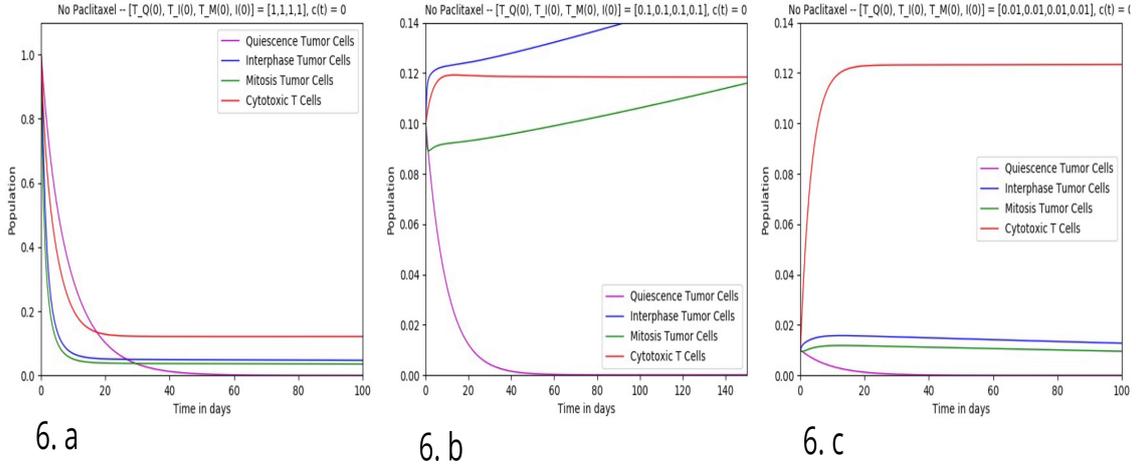
Figure 6: 6.a No Paclitaxel - all initial values 1, 6.b No Paclitaxel - all initial values 0.1, 6.c No Paclitaxel - all initial values 0.01.

From these three analyses (fig. 6.a, 6.b and 6.c) I can observe that the steady state of the cytotoxic T-cells is about 0.12. Hereby I can conclude that a normal healthy person will have an initial value of cytotoxic T-cells population around 0.12. For this I can now test how Paclitaxel will help the patient, depending on the patient's immune system.

If 0.12 is the initial value of cytotoxic T-cells for a normal healthy person, I would like to investigate how a person with an extreme immune system and a weakened immune system will react to the chemotherapy, since chemotherapy and the immune system depend on each other.

At first if the patient has an outstanding immune system the initial value of the cytotoxic T-cells population is set to 0.1488 (fig. 7.a & 7.b), while for the patient with a fragile immune system the initial value is set to 0.0912 (fig. 8.a 8.b). Fig. 7.a and 7.b show the development of the system for a healthy patient in the presence and absence of chemotherapy, respectively. In both chases the tumor cells decrease for some time. In the first case where Paclitaxel is absent the immune system defeats the tumor for a very short time and then develops by some means slowly, but when chemotherapy is present the tumor cells are defeated much faster and go towards zero.

7. a

7. b
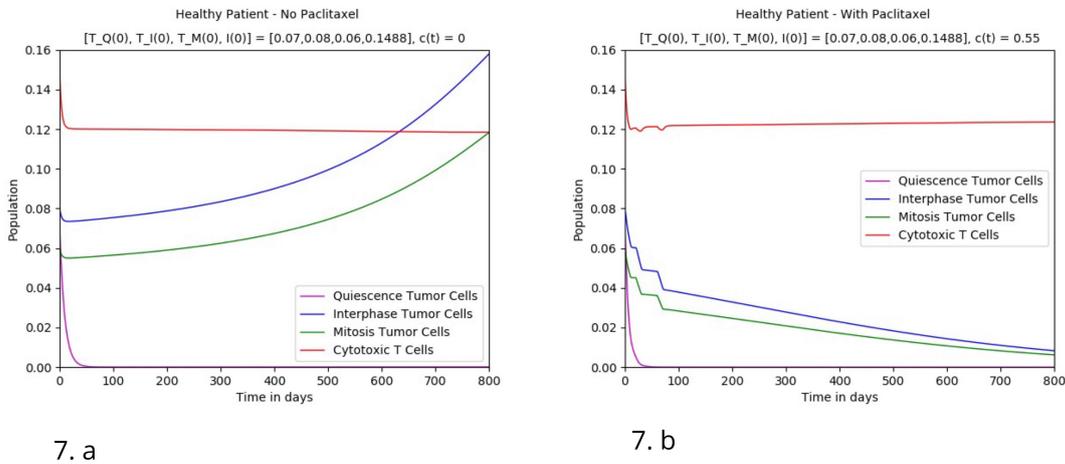
Figure 7: 7.a Patient with an extreme high immune system - No Paclitaxel, 7.b Patient with an extreme high immune system - With 3 pulses of Paclitaxel.

16

In fig 8.a and 8.b it is visible that the unhealthy patient cannot defeat the tumor without the presence of chemotherapy. Again, it proves how important a patient's immune system is, before any medication is taken into consideration.
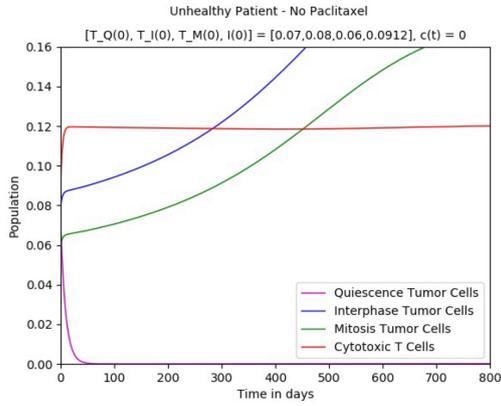
8. a
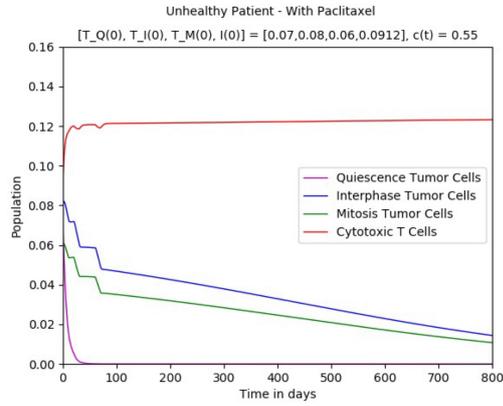
8. b

Figure 8: 8.a Patient with a weak immune system - No Paclitaxel, 8.b Patient with a weak immune system - With 3 pulses of Paclitaxel

To further understand the importance of immune system and immune deficiency the initial value of cytotoxic T-cells population is set to 0.06 (fig. 9.a & 9.b). Such a low cytotoxic T-cells population value represents an immune deficient patient, whose immune response no longer functions properly. The initial function values for the tumor cells are raised for this analysis to represent a cancerous patient in a critical stage, where $[T_Q(0), T_I(O), T_M(0), I(0)] = [0.23, 0.23, 0.12, 0.6]$ Fig 9.a and 9.b give a survey on how immune deficient patients struggle in defeating cancer with chemotherapy treatment.
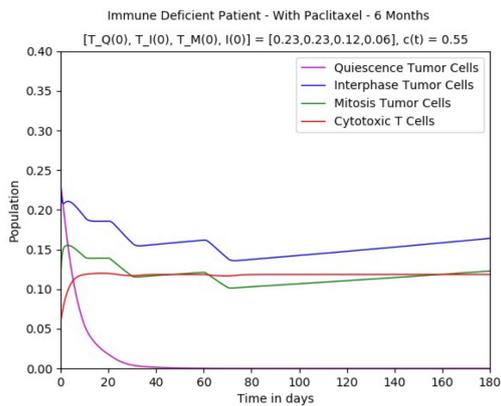
9. a
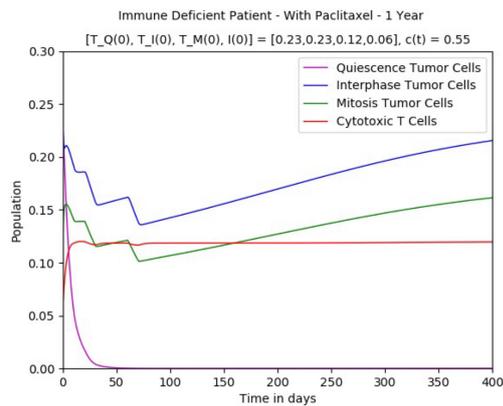
9. b

Figure 9: 9.a Immune deficient patient - 6 months of Paclitaxel, 9.b Immune deficient patient - 1 year of Paclitaxel.

As we see in both analysis Paclitaxel helps for a short period of time, but because of the extremely low value of cytotoxic T-cells, the patient is not able to cooperate with the chemotherapy, because in this case the chemotherapy starts influencing the body's own cells.

17

5.3 Discussion based on New Studies

The models involved in this thesis have been tested for the last years and have provided numerous answers concerning tumor cells. Recently some strange characteristics of the quiescent state of tumor cells have been noticed. In almost all analysis I acquire similar response from my simulations as the previous studies mentioned throughout the thesis [7] [8] [9]. Referring to the new discoveries, none of the graphs in which quiescent tumor cells decrease consist with recent observations in patients.
It has been shown that quiescent cells should not be dependent of chemotherapeutic drugs, since they are resistant to them and no change will occur.

Studies from 2015 up to 2018 have shown that human solid tumor growth depends not only on proliferating cancer cells but also on the continuous production of slow proliferations, such as quiescent cancer cells. The second mathematical method, "Cancer Method 2", mentioned in this thesis was only based on proliferating and quiescent cells. Based on the new discoveries, method 3 is misleading. The reason for which I conclude this among other things, is because of eq. (34) that should not depend on the kill-term $u_1(t) = k_5 \cdot (1 - e^{k_6} \cdot w(t))$ caused by chemotherapeutic medication, because quiescent tumor cells are not controlled by chemotherapy. Based on the new studies, eq. (35) should not be dependent of quiescence population T_Q, but eq. (36) should. The reason for this is because quiescence cells are produced after the mitosis-phase of the proliferating tumor cells is over. Most of cancer cells proliferate very slowly or not at all within human tumors. Quiescent tumor cells should not decrease in population immediately, but remain stable for some amount of time and eventually start to increase if they escape the quiescent state and resume their cell cycle.

The quiescent tumor population in model 3 is influenced by chemotherapy due to the way the differential equation is designed. The presence of quiescent cell populations in tumors represents a major challenge in treating the disease. Quiescent tumor cells show limited sensitivity to chemotherapeutical drugs, and tend to resume proliferation, resulting in tumor reseeding and growth. It is very important to develop therapies that target these quiescent cell populations, to achieve long-lasting remission. Most human tumors are thought to depend on AKT kinase signaling, which promotes tumor growth, survival and progression. It has been discovered that proliferating cancer cells, divide into $AKT1^{low}$ daughter cells, which are slowly proliferating, tumor-initiating and chemotherapy resistant. These daughter cells are considered to be in the quiescent state of a cell-cycle. Selective depletion of $AKT1^{low}$ slow-proliferator-cells have proved to reduce the tumor growth and further studies containing AKT mutations and AKT inhibitors are expanding [32] [17] [33]. Not only that these slow-proliferator-cells are resistant to cytotoxic agents, they also promote cytotoxic stress resistance of proliferating neighbors, through non-cell autonomous communication. Quiescent tumor cells have shown survival over 4 to 6 months of chemotherapeutic drugs. When considering this, quiescent cells may be the reason why metastasis occurs [34] [35] [36]. Based on the new information, I managed to assemble another cancer method, for which I take into account the new theories mentioned above.

$$\frac{T_{QS}(t)}{dt} = a_5 \cdot T_I(t - \tau) - a_6 \cdot T_{QS}(t) - d_4 \cdot T_{QS}(t) - c_5 \cdot I(t) \cdot T_{QS}(t) \tag{46}$$

$$\frac{T_{QR}(t)}{dt} = a_5 \cdot T_I(t - \tau) - a_6 \cdot T_{QR}(t) - d_4 \cdot T_{QR}(t) - c_5 \cdot I(t) \cdot T_{QR}(t) \tag{47}$$

$$\frac{T_I(t)}{dt} = 2 \cdot a_4 \cdot T_M(t) - a_5 \cdot T_1(t - \tau) - c_1 \cdot T_I(t) \cdot I(t) - d_2 \cdot T_I(t) - a_1 \cdot T_I(t - \tau) \tag{48}$$

$$\frac{T_M(t)}{dt} = a_1 \cdot T_I(t-\tau) - d_3 \cdot T_M(t) - a_4 \cdot T_M(t) - c_3 \cdot T_M(t) \cdot I(t) + a_6 \cdot T_{QS}(t) + a_6 \cdot T_{QR}(t) - k_1 \cdot (1 - e^{k_2} \cdot w(t)) \cdot T_M(t) \tag{49}$$

$$\frac{I(t)}{dt} = k + \frac{\rho \cdot I(t) \cdot (T_{QS}(t) + T_{QR}(t) + T_I(t) + T_M(t))^n}{\alpha + (T_{QS}(t) + T_{QR}(t) + T_I(t) + T_M(t))^n} - c_2 \cdot I(t) \cdot T_I(t) - c_4 \cdot T_M(t) \cdot I(t)$$

$$- c_6 \cdot T_{QS}(t) \cdot I(t) - c_6 \cdot T_{QR}(t) \cdot I(t) - d_1 \cdot I(t) - k_3 \cdot (1 - e^{k_4} \cdot w(t)) \cdot I(t) \tag{50}$$

$$\frac{C_M(t)}{dt} = (r_m \cdot C_M(t) \cdot (1 - \frac{C_M(t)}{M_{max}})) + (q_m \cdot C_k \cdot (1 - ds) \cdot T_{QS}(t) \cdot N_t) + (q_m \cdot C_k \cdot T_{QR}(t) \cdot N_t) + (q'_m \cdot C_k \cdot C_M(t) \cdot N_t) \tag{51}$$

$w(t)$ is calculated as mentioned in model 1 (eq. 24).

The new method has substantial changes in the differential equations. Instead of one differential equation based on quiescence cells, I have now formed two equations, one containing sensitive quiescent cells (eq. 48)and the other containing resistant quiescent cells (eq. 49). The whole system includes now both the sensitive and the resistant quiescent cells.

Relative to *model* 3, three new differential functions are taken into consideration. T_{QS} describes the sensitive quiescent tumor cells, T_{QR} describes the resistant quiescent tumor cells and C_M is the metastasis tumor cells population. The drug term: $-k_5 \cdot (1 - e^{k_6} \cdot w(t))$ is not present in eq. (48) and (49), because the quiescent cells do not respond to chemotherapy or other cytotoxic agents. Even if in method 3, the patient receives chemotherapy, the mathematical model can be changed in such a way that the patient received some other medication for the quiescent tumor cells. Another thing I changed is the term $a_6 \cdot T_{QS}(t)$ and $a_6 \cdot T_{QR}(t)$, which are taken into considerations in the differential equation for mitosis(eq. 51) and not in the differential equation for interphase tumor cells(eq. 52) as in model 3(eq. 35). The reason for this is because quiescent cells have shown to go in this non-proliferating stage after mitosis and should not be dependent of the interphase, even if interphase is the longest stage. The cytotoxic T-cell differential equation (eq. 52) has been changed with two additional terms $T_{QS}(t)$ and $T_{QR}(t)$.

The last differential equation (eq.53) is a part of a study from 2016, for which I take into consideration the development of quiescent cells combined with chemotherapy and the issue that resistant quiescent cells provide. Eq. 53 has some new parameters which I obtained from the studies mentioned above, while other parameters have been determined and approximated from method 3. C_M is the number of new metastatic cells after therapy. r_m is the metastatic growth rate. M_{max} is the maximal carrying capacity of the new metastatic cells. qm is the dissemination rate of the drug for the sensitive cancer cells, while q'_m is the dissemination rate of the drug for the resistant cancer cells. The rates depend on the existing sensitive and resistant quiescent tumor sizes. C_k is the angiogenic cell number, which describes the formation of new blood vessels, but is set to be a constant in this case. The term $(1 - ds) \cdot T_{QS}(t)$ describes the drug effect on sensitive metastatic cells. ds was also set to be a constant [37].

I only considered the last two cases analyzed for method 3 (fig 7.b and 8.b), in which we have a patient with a very strong immune system and a patient with a very weak immune system. From previous studies I achieved information for the initial values of quiescent sensitive and resistant tumor cells populations [37]. The new system established has not been examined in detail, but some of the developments of the different populations look promising. There is a slightly difference in the development of the different populations from the two patients (fig 11.a and 11.b).
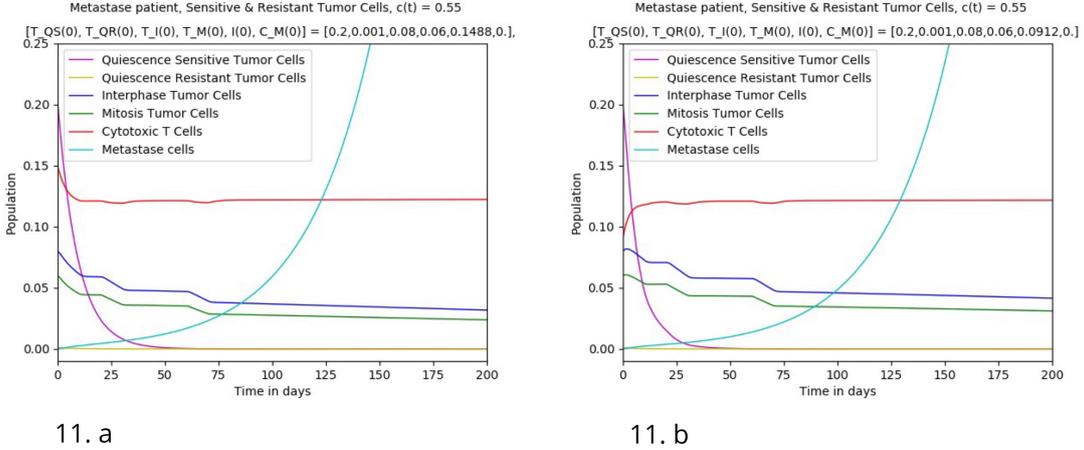
Figure 10: 11.a Patient with strong immune system & 11.b Patient with weak immune system

It is clear that the method can be considered for further research. A lot of biological perspective can be discussed based on the two figures, one of the most significant developments is the one of the quiescent sensitive tumor cells. The population of sensitive quiescent cells decreases in time, this may have a significance according to the fact that quiescent tumor cells escape the quiescent phase after some time. In association with this, the metastases cells population increases about the same time as the quiescent sensitive tumor cells reach steady state. So when the quiescent sensitive tumor cells escape the quiescent phase, a patient will experience metastasis. The reason why the metastasis cell population keeps increasing (fig 11.a and 11.b), is because the development process of the system has not stopped. Here another medicational interpretation should be applied, because metastasis T-cells have to be defeated as well. The development of interphase and mitosis tumor cells look similar, but there is already a little difference in the beginning of the graph, where the tumor cells decrease immediately for the patient with a stronger immune system. The cytotoxic T-cells find a steady state, as shown in previous models. Quiescent resistant tumor cells remain unchanged and this proves that they do not respond to chemotherapeutical drugs.

6 Conclusion & Suggestions for Future Research

There are many choices for how a cancer system should be modeled since there are numerous elements to be investigated for this topic. I observed that the early detection of cancer and a high immune system mean a lot when receiving chemotherapy. The simulations I established proved that even if chemotherapy is given, the human body cannot defeat the disease without a strong immune system. To have a deeper understanding of the way cancer cells develop, quiescent tumor cells were included in the system. The development of quiescent cells suggested in *method* 3, does not agree with the new discoveries of quiescent tumor cells growth. For this reason I experimented with a combination of *method* 3 and a new mathematical modeling of therapy-induced cancer drug resistance from 2016 [37]. The differential equation system behind the biological understanding is acceptable, but the complexity of the oncological structure is not accurate in earlier studies. This was the reason for why I attempted to describe a different system(fig 11.a and 11.b). Experimental data is necessary for a precise and complete outcome of the new model. Further research for this topic is needed in order to understand the implications of the model that I have developed. Quiescent tumor cells complicate the diagnosis and treatment of cancer patients in real life. Based on the differential equation I observed the same obstacle. A model of two different systems could be considered in further research. One of the systems containing quiescent, proliferating and Cytotoxic T cells while chemotherapy is given and the other system includes the metastases conditions mentioned above, which only start processing the data achieved from the first system after some time. The model including the tumor populations should depend both on chemotherapy, but also on terms that affect the quiescent cells, such as inhibitors.

References

[1] Jemal, A.; Bray, F.; Center, M.M.; Ferlay, J.; Ward, E.; Forman, D. *"Global Cancer Statistics"*. CA-A Cancer Journal for Clinicians , 61, (2011), 69-90.

[2] Hart, D., Shochat, E., Agur, Z. *"The growth law of primary breast cancer as inferred from mammography screening trials data"*. British Journal of Cancer, 78, (1998), 382-387.

[3] Torre, LA.; Bray, F.; Siegel, RL.; Ferlay, J.; Lortet-Tieulent, J.; Jemal, A. *"Global Cancer Statistics, 2012."* CA-A Cancer Journal for clinicians, 65, (2015), 87-108

[4] *Breast Cancer cases - Nordcan database - Northern Countries.* http://www-dep.iarc.fr/NORDCAN/DK/frame.asp & http://www-dep.iarc.fr/NORDCAN/DK/StatsFact.asp?cancer=200&country=208

[5] By:Walters, S (Walters, S. 1)[1] ; Maringe, C.; Butler, J.; Rachet, B.; Barrett-Lee, P.; Bergh, J.; Boyages, J.; Christiansen, P.; Lee, M.; Warnberg, F.; Allemani, C.; Engholm, G.; Fornander, T.; Gjerstorff, ML.; Johannesen, TB.; Lawrence, G.; McGahan, CE.; Middleton, R.; Steward, J.; Tracey, E.; Turner, D.; Richards, MA.; Coleman, MP. *"Breast cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK, 2000-2007: a population-based study"*. British Journal Club of Cancer, 108, (2013), 1195-1208.

[6] Michael T. Health. *"Scientific Computing, An Introductory Survey, Second Edition"*. The McGraw-Hill Companies, Inc.; 2nd edition (July 17, 2002)

[7] M. Villasana and A. Radunskaya. *"A delay differential equation model for tumor growth"*. Journal of Mathematical Biology, 47, (2003), 270-294.

[8] M. Villasana and G. Ochoa. *"Heuristic Design of Cancer Chemotherapies"*. IEEE Transactions of Evolutionary Computation, 8,(2004), 513-521.

[9] R. Yafia. *"Dynamics Analysis and Limit Cycle in a Delayed Model for Tumor Growth with Quiescence"*. Nonlinear Analysis: Modeling and Control, 11, (2006), 95-110.

[10] Baker, C.T.H, Bocharov, G.A., Paul, C.A.H., Rihan, F.A. *"Modeling and analysis of time-lags in some basic patterns of cell proliferation"*. Journal of Mathematical Biology, 37, (1998), 341-371.

[11] Reya, T.; Morrison, SJ.; Clarke, MF.; Weissman, IL. *"Stem cells, cancer, and cancer stem cells"*. Nature, 414, (2001), 105-111.

[12] Chen, DS.; Mellman, I. *"Oncology Meets Immunology: The Cancer-Immunity Cycle"*. Immunity, 39, (2013), 1-10.

[13] Hanahan, D.; Weinberg, RA. *"Hallmarks of Cancer: The Next Generation"*. Cell, 144, (2011), 646-674.

[14] Birkhead, B., Rankin, E., Gallivan, S., Dones, L., Rubens, R. *"A mathematical development of drug resistance to cancer chemotherapy"*. European journal of Cancer and Clinical Oncology, 23, (1987), 1421-1427.

[15] Kathleen Collins, Tyler Jacks, and Nikola P. Pavletich *"The cell cycle and cancer"*. PNAS, 94, (1997), 2776-2778.

[16] Kirschner, D., Panetta, J. *"Modeling immunotherapy of the tumor-immune interaction"*. Journal of Mathematical Biology, 37, (1998), 235-252.

[17] Alves, CP.; Dey-Guha, I.; Kabraji, S.; Yeh, AC.; Talele, NP.; Sole, X.; Chowdhury, J.; Mino-Kenudson, M.; Loda, M.; Sgroi, D.; Borresen-Dale, AL.; Russnes, HG.; Ross, KN.; Ramaswamy, S. *"AKT1(low) Quiescent Cancer Cells Promote Solid Tumor Growth."* Molecular Cancer Therapeutics, Volume 17 , JAN 2018, 254-263

[18] Smyth, MJ.; Cretney, E.; Kershaw, MH.; Hayakawa, Y. *"Cytokines in cancer immunity and immunotherapy"*. Immunological Reviews, 202, (2004), 275-293.

[19] Dranoff, G. *"Cytokines in cancer pathogenesis and cancer therapy"*. Nature Reviews Cancer, 4, (2004), 11-22.

[20] Ben-Baruch, A. *"Inflammatory cells, cytokines and chemokines in breast cancer progression: reciprocal tumor-microenvironment interactions."* Breast Cancer Research, 5, (2002), 31-36.

[21] Zitvogel, L.; Apetoh, L.; Ghiringhelli, F.; Kroemer, G. *"Immunological aspects of cancer chemotherapy"*. Nature Reviews Immunology, 8, (2008), 59-73.

[22] Javeed, A.; Ashraf, M.; Riaz, A.; Ghafoor, A.; Afzal, S.; Mukhtar, MM. *"Paclitaxel and immune system"*. European Journal of Pharmaceutical Sciences, 38, (2009), 283-290.

[23] Tang, W.; Yang, JB.; Yuan, Y.; Zhao, ZB.; Lian, ZX.; Liang, GL. *"Paclitaxel nanoparticle awakens immune system to fight against cancer"*. Nanoscale, 9, (2017), 6529-6536.

[24] Nicolas Bacaër. *"Lotka, Volterra and the predator–prey system (1920–1926)"*. A Short History of Mathematical Population Dynamics. Springer, London, (2011), 71-76.

[25] Panetta, J.C., Adam, J. *"A mathematical model of cycle specific chemotherapy"*.. Mathematical and Computer Modeling, 22, (1995), 67-82.

[26] Elmore, S. *"Apoptosis: A review of programmed cell death"*. Toxicologic Pathology, 35, (2007), 495-516.

[27] Zoli, W., Flamigni, A., Frassineti, Gl., Bajorko, P., Depaola, F., Milandri, C., Amadori, D., Gaspericampani, A. *"In-vitro activity of taxol and taxotere in comparison with doxorubicin and cisplatin on primary-cell cultures of human breast cancer"*. Breast Cancer Research and Treatment, 34, (1995), 63-69

[28] Tallarida R.J., Murray R.B. *"Enzyme Kinetics I: Michaelis—Menten Equation"*. Manual of Pharmacologic Calculations. Springer, New York, NY, (1987)

[29] Masunaga, S.; Ono, K.; Sakurai, Y.; Takagaki, M.; Kobayashi, T.; Suzuki, M.; Kinashi, Y.; Akaboshi, M. *"Response of quiescent and total tumor cells in solid tumors to neutrons with various cadmium ratios"*. International Journal of Radiation Oncology Biology Physics, 41, (1998), 1163-1170.

[30] G. van Rossum *"Python tutorial, Technical Report CS-R9526"* Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995.

[31] *Python Software Foundation. Python Language Reference, version 2.7.* http://www.python.org

[32] Kabraji, S.; Sole, X.; Huang, Y.; Bango, C.; Bowden, M.; Bardia, A.; Sgroi, D.; Loda, M.; Ramaswamy, S. *"AKT1(low) quiescent cancer cells persist after neoadjuvant chemotherapy in triple negative breast cancer "*. Breast Cancer Research, 17, (2017), Vol.88.

[33] Salony.; Sole, X.; Alves, CP.; Dey-Guha, I.; Ritsma, L.; Boukhali, M.; Lee, JH.; Chowdhury, J.; Ross, KN.; Haas, W.; Vasudevan, S.; Ramaswamy, S. *"AKT Inhibition Promotes Nonautonomous Cancer Cell Survival."* Molecular Cancer Therapeutics, Volume 15 , 2016, 142-153

[34] Yeh, AC.; Ramaswamy, S. *"Mechanisms of Cancer Cell Dormancy-Another Hallmark of Cancer?"* Cancer Research, Volume 75 , 2015, 5014-5022

[35] Dey-Guha, I.; Wolfer, A.; Yeh, AC.; Albeck, JG.; Darp, R.; Leon, E.; Wulfkuhle, J,; Petricoin, EF.; Wittner, BS.; Ramaswamy, S. *"Asymmetric cancer cell division regulated by AKT."* Proceedings of the national academy of sciences of the United States of the America, Volume 108 , 2011, 12845-12850

[36] Dey-Guha, I.; Alves, CP.; Yeh, AC; Salony.; Sole, X.; Darp, R.; Ramaswamy, S. *"A Mechanism for Asymmetric Cell Division Resulting in Proliferative Asynchronicity."* Molecular Cancer Research, Volume 13 , 2015, 223-230

[37] Xiaoqiang Sun, Jiquang Bao, Yongzhao Shao *"Mathematical Modeling of Therapy-induced Cancer Drug Resistance: Connecting Cancer Mechanisms to Population Survival Rates."* Nature - Scientific Reports, Volume 22498 , 2016, 22498

Appendices

A   Results for Cancer Method 2

Model 2 was mentioned and taken into consideration to show the development of proliferating tumor cells in combination with quiescent tumor cells. Neither immune cells nor chemotherapy is present in this model. This means no parameters will prevent the development of cancer cells.
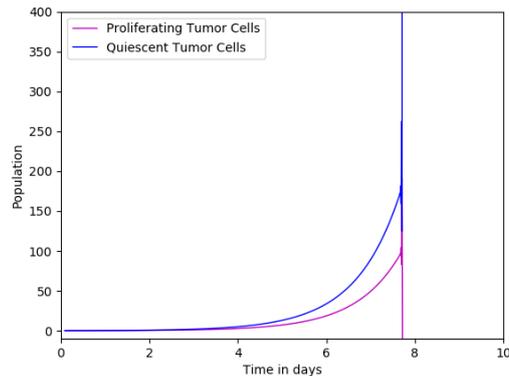

Figure 11: Model of Quiescent & Proliferating Tumor Cells

It is not clear that including quiescence cells in the model improves the understanding impact of a particular chemotherapy treatment that attacks only proliferating cells. In model 1 & model 3 chemotherapy specific parameters related to the proliferating tumor cells were present. Figure 10 does not show considerable information, beside that the combination of proliferating and quiescent cells in the absence of both the immune system and chemotherapy, only makes the two populations develop aberrant after 7 days. In biological interpretation it tells us the patient will collapse, because the tumor cells develop too fast. The figure as it looks reports an unphysical system from day 7 and forward, because the development of the quiescent tumor cells gives no biological understanding.

B   Cancer Method 1 - Script

```
import numpy as np
import matplotlib.pyplot as plt
import math

# MODEL 1

# Definitions for the different population of cells

def dT_interphase_dt(list1):
# Constants dependent of the development of Interphase tumor cells
a4 = 0.8
c1 = 0.9
d2 = 0.11
a1 = 0.98

# Placement of the new calculated values
T_mitosis = list1[0]
T_interphase = list1[1]
I_cytotoxic = list1[2]
T_delay = list1[3]
```

```python
# Differential equation
dT_interphase_dt = 2 * a4 *T_mitosis - c1 * T_interphase * I_cytotoxic - d2 * T_interphase - a1 *

return dT_interphase_dt

def dT_mitosis_dt(list2):

a1 = 0.98
d3 = 0.4
a4 = 0.8
c3 = 0.9
k1 = 0.47
k2 = 0.57

T_interphase = list2[0]
T_mitosis = list2[1]
I_cytotoxic = list2[2]
w_t = list2[3]
T_delay = list2[4]


dT_mitosis_dt = a1*T_delay - d3*T_mitosis - a4*T_mitosis - c3*T_mitosis * I_cytotoxic - k1*(1 - ma
return dT_mitosis_dt



def dI_cytotoxic_dt(list3):

k = 0.036  #constant birth rate of immune cells without presence of tumor cells
rho = 0.1
alfa = 0.2
c2 = 0.085
c4 = 0.085
d1 = 0.29
k3 = 0.49
k4 = 0.061
n = 3

I_cytotoxic = list3[0]
T_interphase = list3[1]
T_mitosis = list3[2]
w_t = list3[3]

#dI_cytotoxic_dt = 0

dI_cytotoxic_dt = k + ((rho*I_cytotoxic * (T_interphase + T_mitosis)**n)/(alfa+(T_interphase + T_m
return dI_cytotoxic_dt

def dw1_dt(w1):

global c

lambda1 = 126.12

dw1_dt = - lambda1 * w1 + c
return dw1_dt
```

```python
def dw2_dt(w2):

global c
lambda2 = 0.85

dw2_dt = - lambda2 * w2 + c
return dw2_dt

def w_t(w1,w2):
r1 = 0.73
r2 = 0.27


w_t = r1*w1 + r2*w2
return w_t

# Runge Kutta fourth method with which we solve all the other equations

def rk4(T, T_list, y, h):

# T is the function to be propagated, feks: dT_Interphase_dt, dT_mitosis_dt etc.
# T_list is the list of results created as time passes by
# y is the value on which the T function works on
# h is the time-step

y = np.array(y)

k1_rk4 = h * T (y)
k2_rk4 = h * T (y + 0.5 * k1_rk4)
k3_rk4 = h * T (y + 0.5 * k2_rk4)
  k4_rk4 = h * T (y + k3_rk4)


  runge_kutta = (k1_rk4 + k2_rk4 + k2_rk4 + k3_rk4 + k3_rk4 + k4_rk4) / 6

return runge_kutta

# Iteration times and delay constant
time_list = []
n_steps = 16000
delay = 1


#Specify start values of the population, and taking the delay values in consideration, so for the
T_interphase_list = [1.3 for i in range(10)]  #interphase cells #y(0)
T_mitosis_list = [1.2 for i in range(10)]  #mitosis cells
I_cytotoxic_list = [0.43 for i in range(10)]  #cytotoxic cells
w1_list = [0. for i in range(10)]    #linear combination of states
w2_list = [0. for i in range(10)]    #linear combination of states
w_t_list = [0. for i in range(10)]


# Iteration in time, for which the functions are only solved after the ninth place in the list
for i in range(9,n_steps):


if 0 <= i * 0.01 <= 3 or 20 <= i * 0.01 <= 23 or 48 <= i * 0.01 <= 50 or 68 <= i * 0.01 <= 70:
c = 0.30
```

```
else:
c = 0

#c(t) = 0 if no drug
#c(t) = 1 if drug present
# h = 0.1, and represents the time step

T_interphase_rk = rk4(dT_interphase_dt, T_interphase_list,[T_mitosis_list[i], T_interphase_list[i]
T_interphase_list.append(T_interphase_list[i]+T_interphase_rk)

T_mitosis_rk = rk4(dT_mitosis_dt,T_mitosis_list,[T_interphase_list[i], T_mitosis_list[i], I_cytoto
T_mitosis_list.append(T_mitosis_list[i]+T_mitosis_rk)

I_cytotoxic_rk = rk4(dI_cytotoxic_dt, I_cytotoxic_list, [I_cytotoxic_list[i], T_interphase_list[i]
I_cytotoxic_list.append(I_cytotoxic_list[i]+I_cytotoxic_rk)

w1_rk = rk4(dw1_dt, w1_list,w1_list[i],0.01)
w1_list.append(w1_list[i]+w1_rk)

w2_rk = rk4(dw2_dt,w2_list,w2_list[i],0.01)
w2_list.append(w2_list[i]+w2_rk)

w_t_list.append(w_t(w1_list[i],w2_list[i]))

time_list.append(0.01*i)



Interphase = plt.plot(time_list,T_interphase_list[10:], 'b-', label = 'Interphase Tumor Cells', li
Mitosis = plt.plot(time_list,T_mitosis_list[10:], 'g-', label = 'Mitosis Tumor Cells', linewidth=1
Cytotoxic = plt.plot(time_list,I_cytotoxic_list[10:], 'r-', label = 'Cytotoxic T Cells', linewidth
plt.legend()
plt.title('Pulsed Paclitaxel -- [T_I(0), T_M(0), I(0)] = [1.3,1.2,0.43], c(t) = 0.30, Late', fonts
plt.xlabel('Time in days', fontsize = 10)
plt.xlim((0,160))
plt.ylabel('Population', fontsize = 10)
plt.ylim((0,1.5))
plt.savefig('model1_x4LatePulsedDrug043c030.png')
plt.show()
```

C   Cancer Method 2 - Script

```
import numpy as np
import matplotlib.pyplot as plt
import math

# MODEL 2

# Definitions for the different population of cells

def dT_proliferating_dt(list1):
# The tumor becomes a malignant tumor for b > 0 and becomes benign for b < 0.
# Constants dependent of the development of Interphase tumor cells
beta = 2
u_p = 1
```

26

```python
    b = beta - u_p
    r_p = 0.9
    r_q = 0.5

    # Placement of the new calculated values
    T_proliferating_delay = list1[0]
    T_proliferating = list1[1]
    T_quiescence = list1[2]

    # Differential equation
    dT_proliferating_dt = (b * T_proliferating_delay) - (r_p * (T_proliferating + T_quiescence) * T_pr
    return dT_proliferating_dt


def dT_quiescence_dt(list2):

    r_p = 0.9
    u_q = - 0.9
    r_q = 0.5


    T_proliferating = list2[0]
    T_quiescence = list2[1]


    dT_quiescence_dt = (r_p *(T_proliferating + T_quiescence) * T_proliferating) - (u_q * T_quiescence
    return dT_quiescence_dt

    # Runge Kutta fourth method with which we solve all the other equations

def rk4(T, T_list, y, h):

    # T is the function to be propagated, feks: dT_Interphase_dt, dT_mitosis_dt etc.
    # T_list is the list of results created as time passes by
    # y is the value on which the T function works on
    # h is the time-step
    y = np.array(y)


    k1_rk4 = h * T (y)
    k2_rk4 = h * T (y + 0.5 * k1_rk4)
    k3_rk4 = h * T (y + 0.5 * k2_rk4)
      k4_rk4 = h * T (y + k3_rk4)


      runge_kutta = (k1_rk4 + k2_rk4 + k2_rk4 + k3_rk4 + k3_rk4 + k4_rk4) / 6

    return runge_kutta



    # Iteration times and delay constant
    time_list = []
    n_steps = 1000
    delay = 1

    # Specify start values of the population, and taking the delay values in consideration, so for the
    T_proliferating_list = [0.1 for i in range(10)]
    T_quiescence_list = [0.1 for i in range(10)]
```

```
# Iteration in time, for which the functions are only solved after the ninth place in the list
for i in range(9,n_steps):

# h = 0.1, and represents the time step

T_proliferating_rk = rk4(dT_proliferating_dt, T_proliferating_list,[T_proliferating_list[i-delay],
T_proliferating_list.append(T_proliferating_list[i] + T_proliferating_rk)

T_quiescence_rk = rk4(dT_quiescence_dt, T_quiescence_list, [T_proliferating_list[i], T_quiescence_
T_quiescence_list.append(T_quiescence_list[i] + T_quiescence_rk)


time_list.append(0.01*i) # h steps


plt.plot(time_list,T_proliferating_list[10:], 'm-', label = 'Proliferating Tumor Cells', linewidth
plt.plot(time_list,T_quiescence_list[10:], 'b-', label = 'Quiescent Tumor Cells', linewidth=1.0)
plt.legend()
plt.xlabel('Time in days', fontsize = 10)
plt.xlim((0,10))
plt.ylabel('Population', fontsize = 10)
plt.ylim((-10,400))
plt.savefig('model2_delay.png')


plt.show()
```

D Cancer Method 3 - Script

```
import numpy as np
import matplotlib.pyplot as plt
import math

# MODEL 3

# Definitions for the different population of cells

def dT_quiescence_dt(list1):
# Constants dependent of the development of Interphase tumor cells
a5 = 0.0001
a6 = 0.00015
d4 = 0.1
c5 = 50 * 10**(-3)
k5 = 0.47
k6 = 0.57


# Placement of the new calculated values
T_delay = list1[0]
T_quiescence = list1[1]
I_cytotoxic = list1[2]
w_t = list1[3]


# Differential equation
dT_quiescence_dt = (a5 * T_delay) - (a6 * T_quiescence) - (d4 * T_quiescence) - (c5 * I_cytotoxic
```

```
        return dT_quiescence_dt


    def dT_interphase_dt(list2):

    a4 = 0.8
    c1 = 0.9
    d2 = 0.11
    a1 = 0.98
    a5 = 0.0001
    a6 = 0.00015



    T_mitosis = list2[0]
    T_interphase = list2[1]
    I_cytotoxic = list2[2]
    T_delay = list2[3]
    T_quiescence = list2[4]

    dT_interphase_dt = (2 * a4 *T_mitosis) - (a5 *T_delay) + (a6 * T_quiescence) - (c1 * T_interphase
    return dT_interphase_dt


    def dT_mitosis_dt(list3):

    a1 = 0.98
    d3 = 0.4
    a4 = 0.8
    c3 = 0.9
    k1 = 0.47
    k2 = 0.57



    T_interphase = list3[0]
    T_mitosis = list3[1]
    I_cytotoxic = list3[2]
    w_t = list3[3]
    T_delay = list3[4]

    dT_mitosis_dt = (a1*T_delay) - (d3*T_mitosis) - (a4*T_mitosis) - (c3*T_mitosis * I_cytotoxic) - (k
    return dT_mitosis_dt



    def dI_cytotoxic_dt(list4):

    k = 0.036 #constant birth rate of immune cells without presence of tumor cells
    rho = 0.1
    alfa = 0.2
    c2 = 0.085
    c4 = 0.085
    c6 = 85 * 10**(-5)
    d1 = 0.29
    k3 = 0.49
    k4 = 0.061
    n = 3 # 1 or 2

    I_cytotoxic = list4[0]
```

```
T_interphase = list4[1]
T_mitosis = list4[2]
w_t = list4[3]
T_quiescence = list4[4]


dI_cytotoxic_dt = k + ((rho*I_cytotoxic * (T_quiescence + T_interphase + T_mitosis)**n)/(alfa+(T_q
return dI_cytotoxic_dt

def dw1_dt(w1):

global c
lambda1 = 126.12

dw1_dt = - lambda1 * w1 + c
return dw1_dt

def dw2_dt(w2):

global c
lambda2 = 0.85

dw2_dt = - lambda2 * w2 + c
return dw2_dt

def w_t(w1,w2):
r1 = 0.73
r2 = 0.27

w_t = r1*w1 + r2*w2
return w_t

#Runge Kutta fourth method with which we solve all the other equations

def rk4(T, T_list, y, h):

# T is the function to be propagated, feks: dT_Interphase_dt, dT_mitosis_dt etc.
# T_list is the list of results created as time passes by
# y is the value on which the T function works on
# h is the time-step
y = np.array(y)


k1_rk4 = h * T (y)
k2_rk4 = h * T (y + 0.5 * k1_rk4)
k3_rk4 = h * T (y + 0.5 * k2_rk4)
  k4_rk4 = h * T (y + k3_rk4)


  runge_kutta = (k1_rk4 + k2_rk4 + k2_rk4 + k3_rk4 + k3_rk4 + k4_rk4) / 6

return runge_kutta

# Iteration times and delay constant
time_list = []
n_steps = 18000
delay = 1
```

```
#Specify start values of the population, and taking the delay values in consideration, so for the
T_quiescence_list = [0.23 for i in range(10)]
T_interphase_list = [0.23 for i in range(10)]  #interphase cells #y(0)
T_mitosis_list = [0.12 for i in range(10)]  #mitosis cells
I_cytotoxic_list = [0.06 for i in range(10)]  #cytotoxic cells
w1_list = [0. for i in range(10)]    #linear combination of states
w2_list = [0. for i in range(10)]    #linear combination of states
w_t_list = [0. for i in range(10)]


# Iteration in time, for which the functions are only solved after the ninth place in the list
for i in range(9,n_steps):


if 0 <= i * 0.01 <= 10 or 20 <= i * 0.01 <= 30 or 60 <= i * 0.01 <= 70:
c = 0.55


else:
c = 0

#c(t) = 0 if no drug
#c(t) = 1 if drug present
#h = 0.1, and represents the time step

T_quiescence_rk = rk4(dT_quiescence_dt, T_quiescence_list, [T_interphase_list[i-delay], T_quiescen
T_quiescence_list.append(T_quiescence_list[i]+T_quiescence_rk)

T_interphase_rk = rk4(dT_interphase_dt, T_interphase_list,[T_mitosis_list[i], T_interphase_list[i]
T_interphase_list.append(T_interphase_list[i]+T_interphase_rk)

T_mitosis_rk = rk4(dT_mitosis_dt,T_mitosis_list,[T_interphase_list[i], T_mitosis_list[i], I_cytoto
T_mitosis_list.append(T_mitosis_list[i]+T_mitosis_rk)

I_cytotoxic_rk = rk4(dI_cytotoxic_dt, I_cytotoxic_list, [I_cytotoxic_list[i], T_interphase_list[i]
I_cytotoxic_list.append(I_cytotoxic_list[i]+I_cytotoxic_rk)

w1_rk = rk4(dw1_dt, w1_list,w1_list[i],0.01)
w1_list.append(w1_list[i]+w1_rk)

w2_rk = rk4(dw2_dt,w2_list,w2_list[i],0.01)
w2_list.append(w2_list[i]+w2_rk)

w_t_list.append(w_t(w1_list[i],w2_list[i]))

time_list.append(0.01*i) # h steps

Quiescence = plt.plot(time_list,T_quiescence_list[10:], 'm-', label = 'Quiescence Tumor Cells', li
Interphase = plt.plot(time_list,T_interphase_list[10:], 'b-', label = 'Interphase Tumor Cells', li
Mitosis = plt.plot(time_list,T_mitosis_list[10:], 'g-', label = 'Mitosis Tumor Cells', linewidth=1
Cytotoxic = plt.plot(time_list,I_cytotoxic_list[10:], 'r-', label = 'Cytotoxic T Cells', linewidth
plt.legend()
plt.title('[T_Q(0), T_I(0), T_M(0), I(0)] = [0.23,0.23,0.12,0.06], c(t) = 0.55', fontsize = 10)
plt.suptitle('Immune Deficient Patient - With Paclitaxel - 6 Months', fontsize = 10)
plt.xlabel('Time in days', fontsize = 10)
plt.xlim((0,180))
plt.ylabel('Population', fontsize = 10)
plt.ylim((0,0.4))
```

```
plt.savefig('model3_ImmuneDeficient6Months.png')
plt.show()
```

E New Cancer Method - Script

```
import numpy as np
import matplotlib.pyplot as plt
import math

# NEW MODEL

# Definitions for the different population of cells

def dT_quiescenceS_dt(list1):

# Constants dependent of the development of Interphase tumor cells
a5 = 0.0001
a6 = 0.00015
d4 = 0.1
c5 = 50 * 10**(-3)
k5 = 0.47
k6 = 0.57



# Placement of the new calculated values
T_delay = list1[0]
T_quiescenceS = list1[1]
I_cytotoxic = list1[2]
# w_t = list1[3]
# dT_quiescence_dt = 0

# Differential equation
dT_quiescenceS_dt = (a5 * T_delay) - (a6 * T_quiescenceS) - (d4 * T_quiescenceS) - (c5 * I_cytotox
return dT_quiescenceS_dt


def dT_quiescenceR_dt(list2):

a5 = 0.0001
a6 = 0.00015
d4 = 0.1
c5 = 50 * 10**(-3)
k5 = 0.47
k6 = 0.57

T_delay = list2[0]
T_quiescenceR = list2[1]
I_cytotoxic = list2[2]
#w_t = list2[3]


#dT_quiescence_dt = 0
dT_quiescenceR_dt = (a5 * T_delay) - (a6 * T_quiescenceR) - (d4 * T_quiescenceR) - (c5 * I_cytotox
return dT_quiescenceR_dt
```

```python
def dT_interphase_dt(list3):

a4 = 0.8
c1 = 0.9
d2 = 0.11
a1 = 0.98
a5 = 0.0001
a6 = 0.00015


T_mitosis = list3[0]
T_interphase = list3[1]
I_cytotoxic = list3[2]
T_delay = list3[3]

#T_quiescence = list2[4]

dT_interphase_dt = (2 * a4 *T_mitosis) - (a5 *T_delay)- (c1 * T_interphase * I_cytotoxic) - (d2 *
return dT_interphase_dt


def dT_mitosis_dt(list4):

a1 = 0.98
d3 = 0.4
a4 = 0.8
c3 = 0.9
k1 = 0.47
k2 = 0.57
a6 = 0.00015

T_interphase = list4[0]
T_mitosis = list4[1]
I_cytotoxic = list4[2]
w_t = list4[3]
T_delay = list4[4]
T_quiescenceS = list4[5]
T_quiescenceR = list4[6]

dT_mitosis_dt = (a1*T_delay) - (d3*T_mitosis) - (a4*T_mitosis) - (c3*T_mitosis * I_cytotoxic) + (a
return dT_mitosis_dt



def dI_cytotoxic_dt(list5):

k = 0.036 #constant birth rate of immune cells without presence of tumor cells
rho = 0.1
alfa = 0.2
c2 = 0.085
c4 = 0.085
c6 = 85 * 10**(-5)
d1 = 0.29
k3 = 0.49
k4 = 0.061
n = 3 # 1 or 2
```

```python
I_cytotoxic = list5[0]
T_interphase = list5[1]
T_mitosis = list5[2]
w_t = list5[3]
T_quiescenceS = list5[4]
T_quiescenceR = list5[5]


dI_cytotoxic_dt = k + ((rho*I_cytotoxic * (T_quiescenceS + T_quiescenceR + T_interphase + T_mitosi
return dI_cytotoxic_dt


def C_metastase_dt(list6):

rm = 0.03
qm = 0.035
q_m = 0.015
N_t = 1
M_max = 100
C_k = 0.1
k6 = 0.57
k5 = 0.47
d_s = 0.5

C_metastase = list6[0]
T_quiescenceS = list6[1]
T_quiescenceR = list6[2]
# w_t = list6[3]


C_metastase_dt = (rm * C_metastase * (1-(C_metastase/M_max))) + (qm * C_k * (1-d_s) * T_quiescence

# (- k5*(1 - math.exp(-k6*w_t)))
return C_metastase_dt

def dw1_dt(w1):
global c
lambda1 = 126.12

dw1_dt = - lambda1 * w1 + c
return dw1_dt

def dw2_dt(w2):

global c
lambda2 = 0.85

dw2_dt = - lambda2 * w2 + c
return dw2_dt

def w_t(w1,w2):
r1 = 0.73
r2 = 0.27

w_t = r1*w1 + r2*w2
return w_t
```

```
# Runge Kutta fourth method with which we solve all the other equations
def rk4(T, T_list, y, h):

# T is the function to be propagated, feks: dT_Interphase_dt, dT_mitosis_dt etc.
# T_list is the list of results created as time passes by
# y is the value on which the T function works on
# h is the time-step
y = np.array(y)


k1_rk4 = h * T (y)
k2_rk4 = h * T (y + 0.5 * k1_rk4)
k3_rk4 = h * T (y + 0.5 * k2_rk4)
  k4_rk4 = h * T (y + k3_rk4)


  runge_kutta = (k1_rk4 + k2_rk4 + k2_rk4 + k3_rk4 + k3_rk4 + k4_rk4) / 6

return runge_kutta

# Iteration times and delay constant
time_list = []
n_steps = 20000
delay = 1

#Specify start values of the population, and taking the delay values in consideration, so for the
T_quiescenceS_list = [0.2 for i in range(10)]
T_quiescenceR_list = [0.001 for i in range(10)]
T_interphase_list = [0.08 for i in range(10)]  #interphase cells #y(0)
T_mitosis_list = [0.06 for i in range(10)]   #mitosis cells
I_cytotoxic_list = [0.1488 for i in range(10)]   #cytotoxic cells
w1_list = [0. for i in range(10)]    #linear combination of states
w2_list = [0. for i in range(10)]    #linear combination of states
w_t_list = [0. for i in range(10)]
C_metastase_list = [0. for i in range(10)] # Metastase cells

# Iteration in time, for which the functions are only solved after the ninth place in the list
for i in range(9,n_steps):

if 0 <= i * 0.01 <= 10 or 20 <= i * 0.01 <= 30 or 60 <= i * 0.01 <= 70:
c = 0.55


else:
c = 0

#c(t) = 0 if no drug
#c(t) = 1 if drug present
# h = 0.1, and represents the time step

T_quiescenceS_rk = rk4(dT_quiescenceS_dt, T_quiescenceS_list, [T_interphase_list[i-delay], T_quies
T_quiescenceS_list.append(T_quiescenceS_list[i]+T_quiescenceS_rk)

T_quiescenceR_rk = rk4(dT_quiescenceR_dt, T_quiescenceR_list, [T_interphase_list[i-delay], T_quies
T_quiescenceR_list.append(T_quiescenceR_list[i]+T_quiescenceR_rk)

T_interphase_rk = rk4(dT_interphase_dt, T_interphase_list,[T_mitosis_list[i], T_interphase_list[i]
T_interphase_list.append(T_interphase_list[i]+T_interphase_rk)
```

```
T_mitosis_rk = rk4(dT_mitosis_dt,T_mitosis_list,[T_interphase_list[i], T_mitosis_list[i], I_cytoto
T_mitosis_list.append(T_mitosis_list[i]+T_mitosis_rk)

I_cytotoxic_rk = rk4(dI_cytotoxic_dt, I_cytotoxic_list, [I_cytotoxic_list[i], T_interphase_list[i]
I_cytotoxic_list.append(I_cytotoxic_list[i]+I_cytotoxic_rk)

w1_rk = rk4(dw1_dt, w1_list,w1_list[i],0.01)
w1_list.append(w1_list[i]+w1_rk)

w2_rk = rk4(dw2_dt,w2_list,w2_list[i],0.01)
w2_list.append(w2_list[i]+w2_rk)

w_t_list.append(w_t(w1_list[i],w2_list[i]))

C_metastase_rk = rk4(C_metastase_dt, C_metastase_list, [C_metastase_list[i], T_quiescenceS_list[i]
C_metastase_list.append(C_metastase_list[i]+C_metastase_rk)

time_list.append(0.01*i) # h steps


QuiescenceS = plt.plot(time_list,T_quiescenceS_list[10:], 'm-', label = 'Quiescence Sensitive Tumo
QuiescenceR = plt.plot(time_list,T_quiescenceR_list[10:], 'y-', label = 'Quiescence Resistant Tumo
Interphase = plt.plot(time_list,T_interphase_list[10:], 'b-', label = 'Interphase Tumor Cells', li
Mitosis = plt.plot(time_list,T_mitosis_list[10:], 'g-', label = 'Mitosis Tumor Cells', linewidth=1
Cytotoxic = plt.plot(time_list,I_cytotoxic_list[10:], 'r-', label = 'Cytotoxic T Cells', linewidth
MetastaseSR = plt.plot(time_list,C_metastase_list[10:],'c-', label = 'Metastase cells', linewidth=
plt.legend()
plt.title('[T_QS(0), T_QR(0), T_I(0), T_M(0), I(0), C_M(0)] = [0.2,0.001,0.08,0.06,0.1488,0.], ',
plt.suptitle('Metastase patient, Sensitive & Resistant Tumor Cells, c(t) = 0.55', fontsize = 10)
plt.xlabel('Time in days', fontsize = 10)
plt.xlim((0,200))
plt.ylabel('Population', fontsize = 10)
plt.ylim((-0.01,0.25))
plt.savefig('metastase_healthypatient.png')
plt.show()
```

Paper

"Turing mechanisms under pressure": Mesenchymal condensations could be involved in digit patterning of mice.

Henrik Pinholt,[*][a] **Heura Cardona Blaya,**[b] **James Harvey Swoger and James Sharpe**[b]

5

In recent years, some attention has been directed towards the possibility that mechanical forces could play a role in creating the expression patterns observed in vertebrate body parts. By compressing regions of cultured embryonic mesenchymal stem cells from mouse autopods, this study provides a novel way of testing effects of mechanochemical stimulation of the Turing mechanism responsible for digit patterning

10 in mice. The study found important features to be included in these compression systems for mesenchymal cells: Com- pressing a small area of the culture, using transparent pistons, and maintaining a stable environment from vibrations and disturbances. The results suggest that mechanical forces upregulates Sox9 in the boundary between compressed and uncompressed regions, but doesn't seem to change the compressed pattern in comparison to uncompressed cells. However incomplete, these results

15 suggest that both chemical and mechanochemical regulation of Sox9 might be involved in patterning the digits in mice.
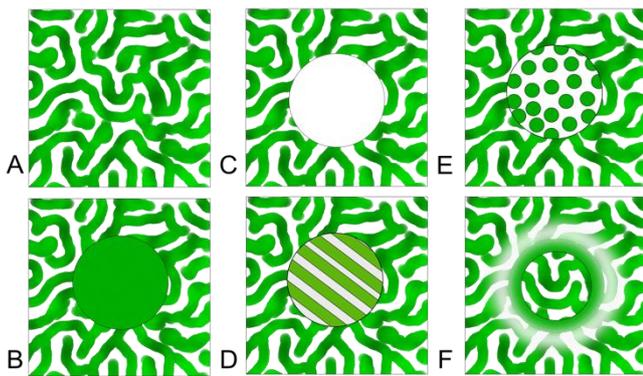
Figure 1: A) In micromass cultures of E11,5 mesenchymal tissue Sox9 is maintained in a stable Turing pattern regulated by Wnt and BMP. B) A compression might upregulate the entire region through mechanically stimulated Sox9 synthesis. C) It could also lead to down regulation through Wnt and BMP signaling. Or maybe the compres- sion changes the Turing mechanism and leads to new patterns like stripe formation (D), dot formation (E) or maybe the interactions happen only in the border zone between the compressed mechanism and the uncompressed (F).

Introduction

The attempt to understand why our body is shaped the way it is,

20 seems to be one of the most challenging questions in the field of developmental biology. An important distinction to make in this regard is the difference between local up- and down regulation of genes, and the spatial patterning of genetic expression to control the morphology of resulting tissue[1]. The study of the latter took a

25 new turn a few years ago, when experiments, conducted by Raspopovic et al., showed how a Turing mechanism[2,3] controls digit patterning in mice: In their model, Bmp and Wnt act under influence of Fgf signaling and Hox13, to position expression of bone-inducing Sox9[4] in digital regions, and the apoptosis

30 inducing Bmp[5] is localized in interdigital mesenchyme with apoptotic cell faiths[6].

This study falls in an increasingly large category of studies investigating the possibility that a Turing mechanism, originally proposed by Alan Turing in 1952, could be driving patterning of

35 some vertebrate body features[7-13]. These studies are different from most standard ways of approaching vertebrate development in that they investigate the global expression patterns of genes, and not local up- and downregulation.

Many great discoveries have been made from looking at the

40 chemical signaling of morphogens[14,15]. However, in recent years, it has been found that One such environmental problem is the fact that mechanical forces can induce genetic expression[16-19]. This fact has led to an increased attention towards the effects of mechanics on patterning the embryo[20-24]. The same applies for

45 digit patterning, and the biochemical pathways used in the Turing mechanism proposed by Raspopovic et al. is not the only way to regulate bone formation. Past studies show that mechanical forces alone, can activate expression of Sox9 and several other chondrogenic factors[25]. The question is therefore if this Sox9

50 activation pathway is relevant for patterning in vertebrate digit formation. Indeed, in recent years, theories and simulations on more general mechano- chemical regulation of morphology has indicated that mechanochemical pathways could be regulating Turing mechanisms[26,27]. A great deal of work has been put into

55 this computationally[28-31], but to our knowledge, not much experimental work has been done to investigate these mechanochemical effects in relation to Turing mechanisms. This might, in part, be due to the difficulty of working with physical effects when you go from testing local genetic activation or

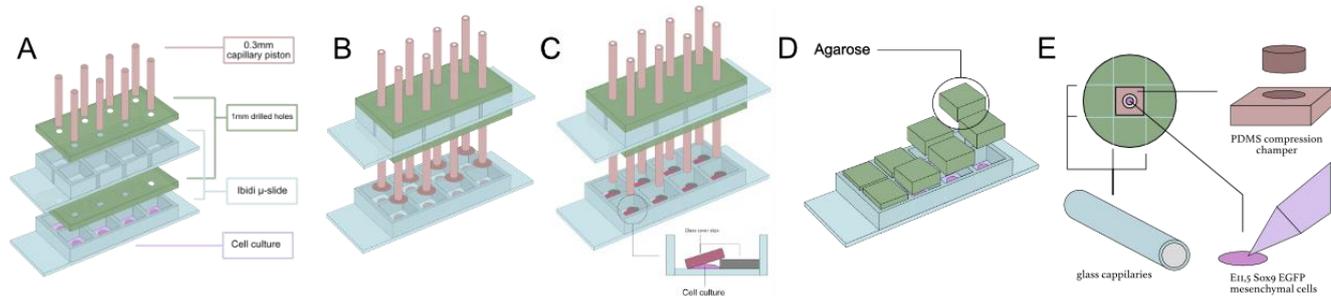60 inactivation, to testing global effects of mechanic stimuli to expression patterning.

Figure 2: Compression chamber designs. A) The first compression system used metal capillary pistons to compress the cells. B) The second device was made by gluing glass cover slips on 1,5mm glass capillaries. C) A variation of this was done by placing the glass cover slips on the culture and then pushing it down with glass capillaries. The two glasses were laid on top of each other, producing a gradient pressure. D) Hydrostatic pressure was created by casting blocks of agarose gels in μ-slides (Ibidi), and put- ting weights on top. E) We also tried using compression champers cut from blocks of PDMS placed in an Ibidi μ-dish to create hydrostatic pressure.

Here we present an attempt to address this issue. However incomplete, this study seeks to provide a fresh perspective to the topic of testing mechanochemical effects on Turing mechanisms, and provides evidence to support that these mechanochemical interactions could perhaps be involved in the Turing mechanism operating in mouse digit formation.

Digit formation in a petri dish

Raspopovic et al. showed that E11.5 mesenchymal cells in micromass culture express Sox9 in a Turing pattern. This pattern is created in the same fashion as in the digits and could therefore be used as a way of studying the digit patterning process in vitro. By culturing mesenchymal stem cells, we had a way of investigating the global effects of mechanical stimuli on the patterning process rather than the local up- or downregulation of genes. By introducing a compressive force to the culture, we could directly observe how the mechanical forces changed the pattern (fig 1). If the compression just up- or downregulated Sox9, the mechanical forces would not play a role in the spatial patterning of the mechanism (fig 1A-C), but if the pattern, on the other hand, changed un- der compression (fig 1D-F), the mesenchymal condensations during bone formation might play a role in the Turing mechanism. This idea provided the basis for the experimental work in this article.

Materials and methods

Compression chamber design

The mesenchymal condensations in digit formation was simulated using physical compression devices (fig 2). The systems were fabricated by drilling holes through two lids and a μ-slide from Ibidi. These where then glued together using Araldite Cristal. The holes where 1mm in diameter for the capillary pistons (fig 2A) and 1,6mm for the 1,5mm glass capillaries used in the other two devices (fig 2B-C). The capillary pistons were put directly on the culture, and for the glass capillary devices, a 5mm circular glass cover slip was either glued on the capillary (fig 2B) or placed directly on the culture (fig 2C). Weights were put on the top of the pistons and were made by attaching Alpino Plastilina to the top of the pistons.

Micromass culture & imaging

Sox9-EGFP transgenic mice embryos were removed and micro dissected in ice cold PBS to re- move limbs and hindlimbs. E11.5 mice were used for the capillary piston devices, and E12.5 mice were used for the glass slide devices. Autopods were micro dissected, and then digested in 0,5% Trypsin EDTA. The ectoderm was removed, and cells where filtered in a 35 μm cell strainer cap and resuspended in PBS. Cells where plated as10μL drops with a cell concentration of $2.5 \cdot 10^7$ cells/mL in 8 well μ-slides (Ibidi). Cells were allowed to attach for one hour at 37°C, 5% CO2 before adding DMEM/F12 with L-Glut +10%FBS +1% Penicillin/Streptomycin to each well. Hereafter the compression was introduced. Cells where incubated at 37°C, 5% CO2 for 24 hours before imaging results.

Following incubation, medium was removed, and cells were stained with DAPI for 15 minutes, samples were washed with PBS and fixed in Performaldehyde for 30 minutes and then washed in PBS. Results were viewed on a Leica TCS SP5 for closeup images, and mosaics where done on a Zeiss Cell Observer HS with 5x objective. Stitching was done using MosaicJ[32], and all other editing was done using Fiji.

Results

Device design

The metal pistons had a smaller surface area than the culture (fig 3). This allowed for an internal control in each well, since the compressed region was next to an uncompressed Turing pattern (fig 2A). The cells adhered to the bottom of the pistons, and the culture was ripped when pistons were removed. This stopped us from viewing the compressed pattern under the piston (fig 3B-C). The results therefore did not, shed light on the behavior of the system under compression since the compressed region was not visible.

This problem was solved partly by using glass cover slips. Gluing the coverslips to the capillaries killed the cells (results not shown), but by placing the cover slips on the cell culture, you could see the compressed region (fig 4). The problem was lack of stability. Since the glass slides were not attached, but merely laid on top, there was no way of knowing if the slides moved, and the system needed greater stability to function properly (fig 4A). It did, however, prove to be a system that easily allowed for visualization and compression of the sample. DAPI staining did not diffuse properly, but bright-field images prove cell density to be uniform underneath the glass, and Turing patterns was could be found underneath the slides (fig 4B).

Pattern changes
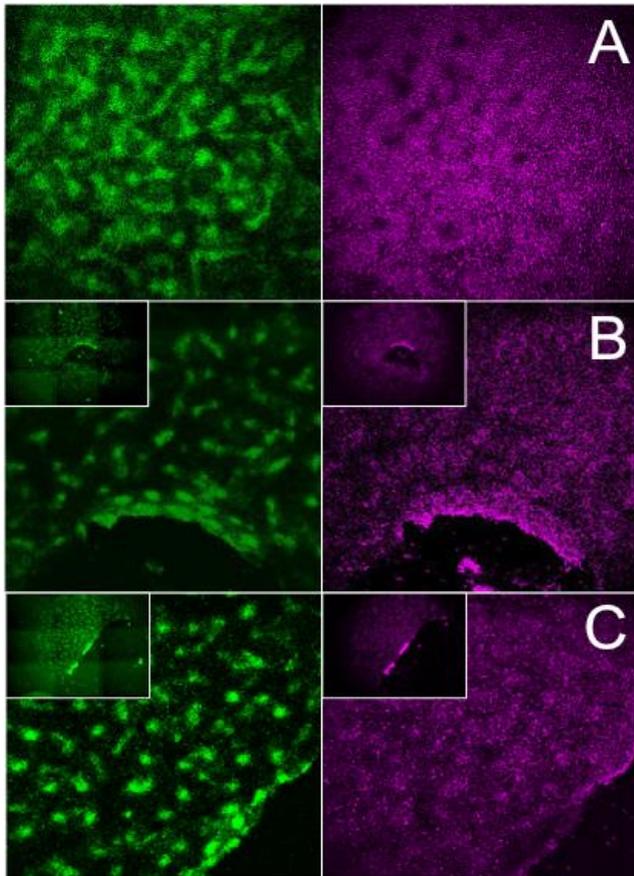
Figure 3: CLSM images and 3x3 Zeiss mosaics (inserts in B and C). Green is EGFP-Sox9, and magenta is DAPI. A) Control, showing the normal Turing pattern in culture. B) 0.5g compressed sample with capillary piston. The stimulus is surrounded by a control pattern, which allows for comparison C) 2.5g compressed sample with capillary piston. Due to cell adhesion, the compressed cells were removed with the pistons upon imaging, and the compressed pattern could not be viewed in this system.

Even though the two devices had very different physical impact on the culture, they sometimes showed the same results (fig 5). Some of the cultures had significant upregulation of Sox9 on the
5 border between uncompressed and compressed areas. In some cultures, there was also a down-regulation in the nearby surrounding uncompressed region (fig 5A-B). The glass devices allowed for visualization of the under- lying compressed region in these border phenomena (fig 5C). This revealed that the
10 compressed region still expressed a Turing pattern, and that the uncompressed region seized to express the pattern. This would make sense since the upregulation of Sox9 would lead to the induction of the diffusing Sox9 inhibitors Wnt and BMP, and ultimately inhibit Sox9 in surrounding tissue. Since the
15 compression upregulates Sox9, this might counteract the effects of the inhibitor in compressed regions, but the effects are still visible in uncompressed parts of the culture since there is no mechanical compression to induce Sox9, and the inhibitors dominate these regions.

20
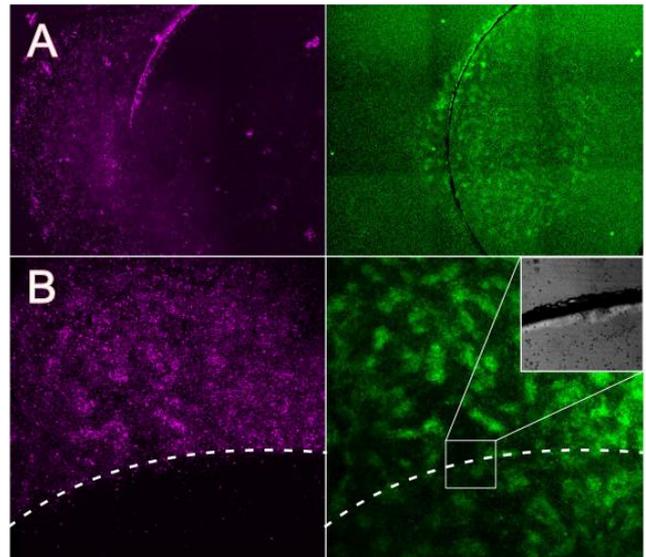
Figure 4: A) Mosaic of homogenous 0.5g pressure on glass cover slides. The DAPI-staining clearly shows the disruptive forces of the unstable system. B) Confocal images of another well showing the boarder (white dots) between compressed (bottom) and uncompressed (top). The brightfield insert shows cells distributed both under and outside the glass.

Discussion

The compression systems used in these experiments was prototype constructs, and therefore the results cannot be seen as concrete evidence for an interaction between mesenchymal
25 condensations and the Turing mechanism. The pistons and cover glasses were not completely fixed, and perturbations lead to several uncertainties as to what were Sox9 induced mesenchymal condensations, and what were cell movements from moving the de- vices and pistons. An important consideration for future work
30 with testing mechanochemical effects on patterning through a cell culture system like this, is therefore to reduce vibrations, and maybe even do imaging, compression and culturing at the same place. Another lesson to learn from these experiments is the importance of only compressing a fraction of the cell culture. As
35 seen from the results, a smaller area of compression allows for an internal control in each well. This is very important since Turing patterns are very hard to compare across cultures, and every environment is different. This leads to the third consideration when compression mesenchymal cell cultures: The importance of
40 using transparent pistons. Using the glass cover slips, we were able to observe both compressed and uncompressed cells. This allowed for comparative studies, and the interesting border phenomena could be observed. By working with these three principles: Small area of compression, stability and transparency,
45 one will obtain the best results when experimenting with the digit Turing mechanism.

As mentioned in the results, some wells where characteristic in that they had an upregulation of Sox9 in the border between compressed and uncompressed cells. Sometimes this border was
50 followed by a Sox9 inhibited region. These results suggest that compression could play a role in regulating the morphology of bone patterning. A possible explanation for these observations
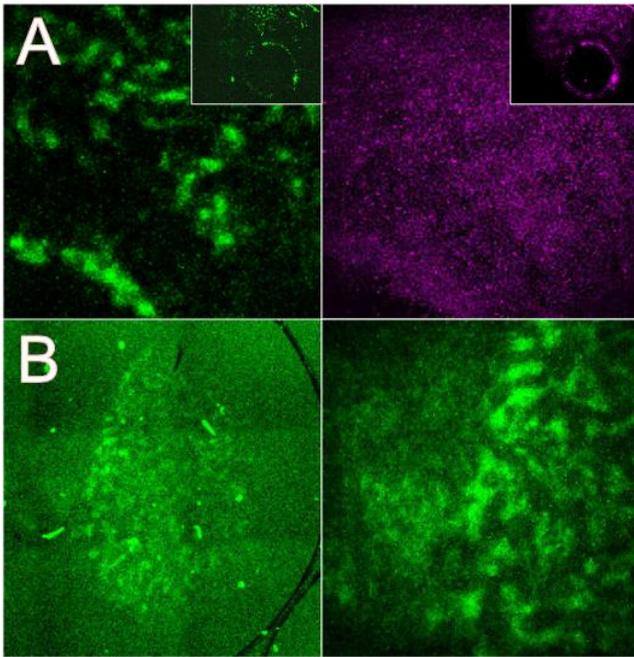
Figure 5: A) Confocal images showing the boarder region between the piston and the culture. GFP (left) and Dapi (right) is clearly upregulated next to the piston, but an area with no expression separates the Turing Pattern from this upregulated region. Inserts shows Zeiss mosaics showing that this behavior exist in the entire region facing the culture. B) Zeiss mosaic GFP (left) and CLSM (right)

could be that since compression is somewhat uniform under the piston, Sox9 is evenly upregulated under the glass. This allows for the formation of a periodic Turing pattern. However, in the border region, mechanically induced Sox9 diffuse outside the glass and creates an upregulated border region. Sox9-induced inhibitors diffuse into the surrounding uncompressed cells, and create a region of Sox9 cells. This wave of inhibition would stop due to the Turing mechanism upregulating local Sox9 at a distance to the compression depending on how much force was applied to the cells. This diffusion of inhibitors would not influence the compressed pattern, since Sox9 is produced through mechanochemical induction, and therefore the overall pattern stays the same under compressed conditions (fig 6). This could mean that the Turing mechanism in the digits helped guide early patterning of Sox9, Bmp and Wnt. Hereafter, Sox9-induced condensations in digit areas would up regulate Sox9 further in these areas, but in interdigital mesenchyme, the Sox9 expression would be downregulated tremendously. It is possible that this mechanism could be factor that helped stabilizing the initial pattern defined by the Turing mechanism discovered by Raspopovic et al.

Conclusions

These experiments serve, as a small explorative examination of how mechanochemical interactions could be relevant in patterning processes of vertebrate development. The study has many associated uncertainties and should be thought as more of as a new perspective, than a definitive proof. In this study, we show new methods for testing the global expression of genes and put forth data suggesting the possibility of interplay between
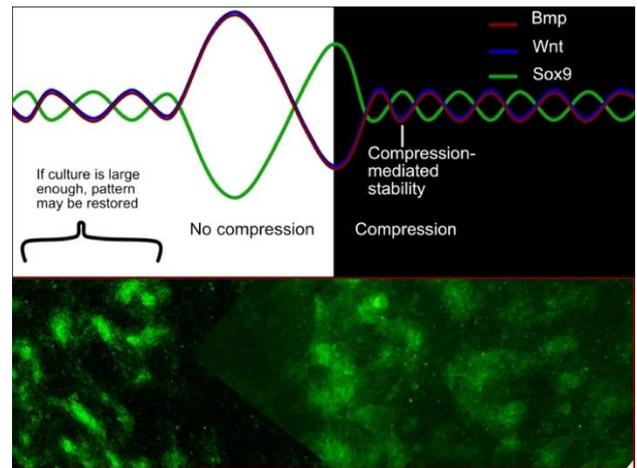
mesenchymal condensations and patterning of the digits. Hopefully these results will give rise to new ideas and approaches to the rather tricky topic of testing the effects of mechanical forces in developmental patterning processes.

Notes and references

ᵃ: Nano-Science Center, University of Copenhagen, Universitetsparken 5, 2100 København Ø, Denmark.; E-mail: henrikpinholt@hotmail.com

ᵇ: Systems Biology Program, Centre for Genomic Regulation (CRG), and Universitat Pompeu Fabra (UPF), Dr. Aiguader 88, 08003 Barcelona, Spain.

1   AJ. Koch, H. Meinhardt, *Rev Mod Phys*, 1994, **66**, 1481-507;
2   AM. Turing, *Phil Roy Biol*, 1952, **237(641)**, 37-72;
3   S. Kondo and T. Miura, *Science,* 2010, **329(5999),** 1616-20
4   E. Wright, MR. Hargrave, J. Christiansen, L. Cooper, J. Kun and T. Evans et al., *Nat Genet,* 1995, **9(1)**, 15-20;
5   E. Wright, MR. Hargrave, J. Christiansen, L. Cooper, J. Kun, T. Evans et al., *Nat Genet,* 1995, **9(1)**, 15-20
6   MM. Kaltcheva, MJ. Anderson, BD. Harfe and M. Lewandoski, *Dev Biol,* 2016, **411(2)**, 266-76

Figure 6: A possible theory on how mechanical forces interacts could interact with the Turing Pattern in the digits.

7   MM. Kaltcheva, MJ. Anderson, BD. Harfe and M. Lewandoski, *Dev Biol,* 2016, **411(2)**, 266-76
8   J. Raspopovic, L. Marcon, L. Russo and J. Sharpe, *Science,* 2014, **345(6196)**, 566-70
9   J. Guiu-Souto, J. Carballido-Landeira,V. Pérez-Villar and AP. Muñuzuri, *Phys Rev, E.* 2010, **82(6)**, 066209
10   R. Sheth, L. Marcon, MF. Bastida, M. Junco, L. Quintana, R. Dahn et al., *Science,* 2012, **338(6113)**, 1476-80
11   AD. Economou, A. Ohazama, T. Porntaveetus, PT. Sharpe, S. Kondo, MA. Basson et al., *Nat Gene,* 2012, **44(3)**, 348-51
12   P. Müller, KW. Rogers, BM. Jordan, JS. Lee, D. Robson, S. Ramanathan et al., *Science,* 2012, **336(6082)**, 721-4
13   S. Sick, S. Reinker, J. Timmer and T. Schlake, *Science,* 2006, **314(5804)**, 1447-50
14   J. Lefèvre, J-F. Mangin, *Com Biol* 2010, **22;6(4)**, e1000749;
15   RO. Prum and S. Williamson, *Proc Roy Biol,* 2002, **22;269(1493)**, 781-92
16   K. Unsicker and K. Krieglstein, *Wiley-VCH,* 2006, 523-578
17   SF. Gilbert and MJF. Baressi, *Sinauer Associates,* 2016, **11**, 810
18   NJ. Steinmetz, EA. Aisenbrey, KK. Westbrook, HJ. Qi and SJ. Bryant, *Acta Biomater,* 2015, **15(21)**, 142-53
19   AJ. Steward, DJ. Kelly, *J Anat,* 2014, **227(6)**, 717-31
20   P. Strzyz, *Nat Rev Genet,* 2016, **17(9)**, 505

21   KM. Warren, MM. Islam, PR. LeDuc and R. Steward, *App Mat Int,* 2016, **8(34)**, 21869-82.

22   T. Mammoto and De. Ingber, *Development,* 2010, **137(9)**, 1417-20

23   J. Foolen, T. Yamashita and P. Kollmannsberger, *J Phys D,* 2016, **49(5)**, 053001

24   P. McMillen and Sa. Holley, *Gen Dev,* 2015, **32**, 106-11

25   LV Beloussov, *Eur phys J E,* 2013, **36(11)**, 132

26   JH. Henderson and DR. Carter, *Bone,* 2002, **31(6)**, 645-53

27   I. Takahashi, GH. Nuckolls, K. Takahashi, O. Tanaka, I. Semba, R. Dashner et al., *J Cell Sci,* 1998, **30(111)**, 2067-76

28   J. Howard, SW. Grill and JS Bois, *Nat Rev Mol,* 2011, **12(6)**, 400-6

29   NW. Goehring and SW. Grill, *Trends in Cell Biol,* 2013, **23(2)**, 72-80

30   M. Mercker, F. Brinkmann, A. Marciniak-Czochra and T, Richter, *Biol Dir,* 2016, **11(22)**, 72-80

31   LG. Morelli, K. Uriu, S. Ares and AC. Oates, *Comp Appr Dev Pat,* 2012, **336(6078)**, 187-91

32   Ma. Wyczalkowski, Z. Chen, BA. Filas, VD. Varner and LA. Taber, *Emb Tod Rev,* 2012, **96(2)**, 132-52

33   L. Truskinovsky, G. Vitale and TH. Smit, *Engi Sci,* 2014, **83**, 124-37

34   P. Thévanaz and M. Unser, *Mic Res Tech,* 2007, **70(2)**, 135-46

Multilayer Reflection Tomography from Monte Carlo Simulations

Rasmus B. Nielsen[*a]

X-ray Reflection Tomography (XRT) is a tool for imaging of buried layers and interfaces in multilayer thin-films. The method combines X-ray reflectivity and computerized tomography to determine spatially dependent reflectivity curves.[1–3] Simulations of X-ray reflection tomography experiments make it possible to asses to which accuracy properties can be determined and evaluate various reconstruction methods.[2] It is thus important that such simulations mimic real experiments. In this paper it is shown that a Monte Carlo approach can be used to simulate XRT experiments and test reconstruction techniques. This approach takes into account the statistical properties of an experimental X-ray setup and allows for simulation of diverse experimental configurations. The currently used analytical simulations based on projections do not include such statistics and are limited in scope. The Monte Carlo approach will facilitate further development of the applications of XRT.

1 Introduction

X-ray reflectivity is an outstanding tool for non-destructive determination of the structure of thin multilayer structures. The main advantage is that the reflected intensity depends on features of the layer thickness at nanoscale. These multilayer thin films are used e.g. for energy storage, in photonics and in many semiconductor devices such as batteries, LEDs, MEMS, superlattices, etc. Deposition techniques allow the multilayer materials to be tailor-made with layer thicknesses at the atomic scale. To analyze and asses the quality of such materials a technique with sensitivity at a scale of similar length is required.[4]

Reflectivity is measured by the intensity of specular reflection as a function of wavevector transfer (Q). This is done by keeping the X-ray energy constant and varying the incident angle close to grazing incidence as shown in Figure 1. Q is a vector equal to the difference between the vector describing the incoming wave and the scattered wave, for measurements of specular reflection only the magnitude of Q is of interest:[4]

$$Q = \frac{4\pi}{\lambda} sin(\theta) \tag{1}$$

Up to a certain angle i.e. the critical angle, X-rays experience total external reflection because the index of refraction in the X-ray regime is lower in solids than in air.

Beyond this angle the reflectivity is determined by Kiessig fringes caused by interference of the scattered X-rays from the different interfaces in the sample.[4,5] As a function of Q, the reflectivity depends on electron density, the layer thickness and roughness of interfaces between layers. The measured reflectivity curve can thus be used to determine all these multilayer properties. For normal X-ray reflectivity (XRR), however, only the average reflectivity is recorded so the determined properties are also averages of the illuminated portion of the sample.[6]
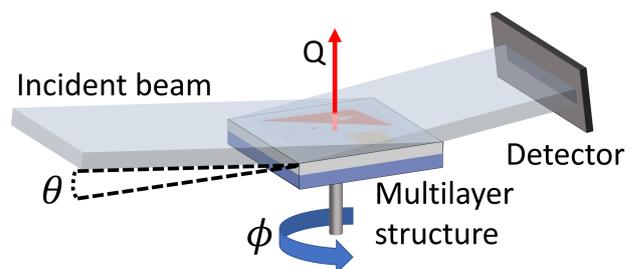

Fig. 1 XRT setup: A sinogram at constant Q is created by keeping the incident angle θ constant while changing the in-plane angle ϕ.

1.1 Parratt's recursive method

To calculate the reflectivity as a function of Q for a given multilayer structure, the kinematical approximation or Par-

[a] *Copenhagen, Denmark. E-mail: fbg450@alumni.ku.dk*

ratt's recursive method is usually applied. While being easier to use the kinematical approximation is only valid at a high Q and fails completely near and below the critical wavevector. Since reflectivity in this case is measured from below and past the critical wavevector, Parratt's recursive method is required. This method uses the refractive index and the Fresnel relation to determine the reflectivity from a single interface. The reflectivity is then calculated from the bottom interface and up. Since there are no reflections from below the bottom interface, the reflected intensity can be determined as a function of the intensity reaching this interface, the refractive index of the materials at this interface and Q. The reflectivity of the next layer can now be determined, since the reflectivity of the layer below is known. This can be continued recursively until the reflectivity of the entire multilayer structure is known.[4,7]
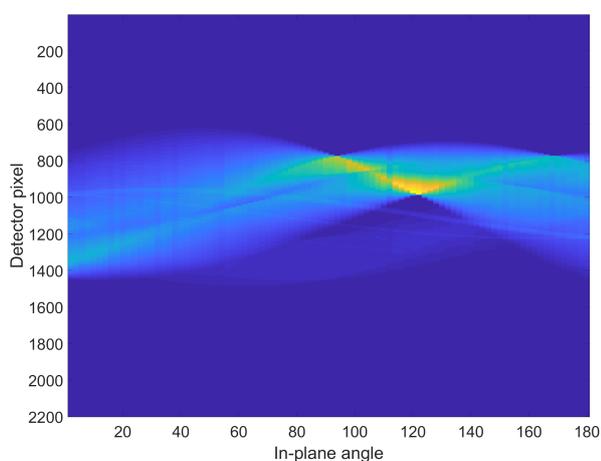
Fig. 2 Sinogram at $Q = 0.0976\ Å^{-1}$ generated from 90 measurements in McXtrace. It was used to reconstruct the spatially dependent reflectivity at this Q-value, as shown on Figure 4.

1.2 X-ray reflection tomography

X-ray reflection tomography (XRT) combines many reflectivity measurements from various sample orientations by applying computerized tomography.[1–3] This enables determination of the spatial dependence of the reflectivity curve.[1] X-ray reflectivity is typically measured at incident angles of a few milliradians, thus the height of the reflected beam along the Q-vector is in the 10 - 100 μm range for typical sample sizes. This is too small to separate intensity from different parts of the sample along the direction of X-ray propagation. The measured intensity is thus the sum of all reflected intensities along the propagation direction. With a collimated beam, the width of the reflected beam running parallel to the sample surface is equal to the width of the illuminated area of the sample.

By measuring the reflectivity with a collimated monochromatic beam at different in-plane angles (ϕ), it is possible to combine the results into a sinogram of the sample at constant Q as shown in Figure 2. By using computerized tomography on the sinogram, it has been shown that the spatially dependent reflectivity can be determined at constant Q.[1] By measuring sinograms at different incident angles, the entire spatially resolved reflectivity curves can be determined. By applying XRT in this manner it has been shown that the tomography can be reconstructed and used to determine heterogeneous layer thickness locally. This is done by fitting analytical solutions from Parratt's recursive method to the measured reflectivity curves.[1]

A spatially dependent reconstruction from a sinogram can be achieved by the methods listed below.

1.2.1 Filtered back projection

Reconstruction can be obtained through filtered back projection (FBP) since the Fourier transform of projections in real space is equal to slices of the Fourier space. By combining multiple slices and applying a high-pass filter the real space tomography of the sample can be reconstructed.[2–4] The FBP method has the advantage that it is computationally efficient, but the required use of a high-pass filter results in increased noise. This noise increase is troublesome if the reflectivity in a small area of the sample is to be determined, such as by micro X-ray reflectivity (μXR).

1.2.2 Algebraic reconstruction techniques

Algebraic reconstructions techniques (ART) are based on solving the inverse problem of the system of linear equation given by:

$$\mathbf{A}\vec{x} = \vec{b} \tag{2}$$

Where $\vec{x}$ is the spatially dependent reflectivity of the sample. $\vec{b}$ contains the observed intensities in the form of a vector representation of the sinogram. $\mathbf{A}$ describes the dependence of observations on the reflectivity of the sample. The components of $\mathbf{A}$ are so-called line-pixel coefficients, which describe how an X-ray beam interacts with the sample before hitting the detector. To determine $\vec{x}$ it is necessary to calculate an approximation of the inverse of $\mathbf{A}$. $\mathbf{A}$ can be calculated analytically, but determining the inverse of $\mathbf{A}$ is not a straightforward task for systems using detectors operating with thousands of pixels. ART methods are thus more demanding to compute than FBP but allow for more corrections, such as constraining the reflectivity so it doesn't use unrealistic values.[8,9] ART is especially useful when only limited data are available to reconstruct the tomography.
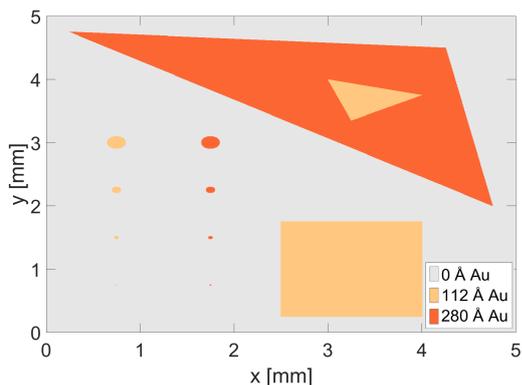
Fig. 3 Map of the heterogeneous multilayer phantom used in simulations. The map assigns each position a reflectivity curve corresponding to the multilayer structure, this means reflectivity is a function of x, y and Q. A map with sharp edges was chosen because these tend to be harder to reconstruct. The circles to the left side of the map have a diameter of 200μm, 100μm, 50μm and 20μm (from the top). They have been included to explore the achievable resolution.

2 Simulations

Simulations were based on simulating a XRT experimental setup with Monte Carlo algorithms. The Monte Carlo methods allow for simulation of systems with statistics corresponding to realistic experiments if the probability distributions of these systems are known. McXtrace[10] is a Monte Carlo environment for simulating X-ray experiments by using a ray-tracing technique where every single photon is created with position and wavelength based on the probability distribution of the source. The propagation of each photon is determined by calculating how it interacts with the components it passes.[10] For the purpose of using McXtrace for XRT simulations, the author has created a new heterogeneous multilayer component. McXtrace already included a homogeneous multilayer component that determined the specularly reflected intensity from kinematic calculations or from a reflectivity curve.[10] The new component mapped different positions on the sample to different reflectivity curves. This was done by initially reading an input map file by extending the embedded C-code in McXtrace. Whenever a photon hit the sample the impact position was used to determine which reflectivity curve to use. Subsequently, the existing multilayer component was called by using McXtrace's meta-language, but with the reflectivity curve corresponding to the position on the map.

To determine whether the simulations correlated with real experiments, the materials from "Interface-sensitive imaging by an image reconstruction aided X-ray reflectivity technique"[1] were simulated in a synchrotron setup

matching the one used in the experiment. The simulated multilayer was a Si substrate with an Au layer of varying thickness underneath a layer of Ti. The exact layout of the multilayers of the sample is shown in Figure 3. Reflectivity curves of the simulated multilayer structures were first simulated in GenX[11], then the reflectivity curves were used as input for McXtrace. The simulated reflectivity curves are shown in Figure 5. The Parratt recursion formula was used in GenX since the kinematic approach does not take into refraction account, which is needed for total external reflection.[4]

A complete reflectivity tomography dataset was recorded by measuring sinograms with in-plane angle increments of $2°$ from $2°$ to $180°$. Only a half circle was measured since the angles from $182°$ to $360°$ would be essentially the same measurements with mirrored results. For each of these 90 measurements in a sinogram, 10^8 photons were simulated. 90 sinograms were simulated at incident angles varying linearly from $0.108°$ to $0.464°$, corresponding to a Q from 0.0308 $Å^{-1}$ to 0.1317 $Å^{-1}$ at 16 keV. This means that the angle of total external reflection for Au is within the simulated angles as it is at Q = 0.08 $Å^{-1}$.[4] The simulated setup was an X-ray beam from a wiggler monochromized to 16 keV with an energy resolution of 10^{-4} keV.[1] This small energy interval was necessary to assume measurement of the reflectivity from each reconstruction was done at constant Q. A reconstruction was made for each of the 90 sinograms since each sinogram had a single corresponding Q value. This made it straightforward to compare the simulated reflectivity with that from the analytical solution in GenX.

The original synchrotron experiment was performed by using a detector with a pixel size of 6.45 μm and a beam with a horizontal divergence of 0.02 mrad.[1] Simulation of the experiment with McXtrace was performed by assuming a collimated beam. Subsequently, the reconstructions were compared with a reconstruction from a simulation using a fan beam with a horizontal divergence of 0.02 mrad in McXtrace. The purpose of this was to show that using a collimated beam was a good assumption. The fan beam was approximated as an infinite number of point-sources with a horizontal divergence given by a Gaussian with an uncertainty of 0.02 mrad. The exact beam and divergence profile depend on the source and optics used in the experimental setup. However, just like in the original experiment, it was assumed that there was no vertical divergence.

The reconstructions were achieved by means of both filtered back projection (FBP) and a sub-class of the algebraic reconstruction technique (ART) termed simul-
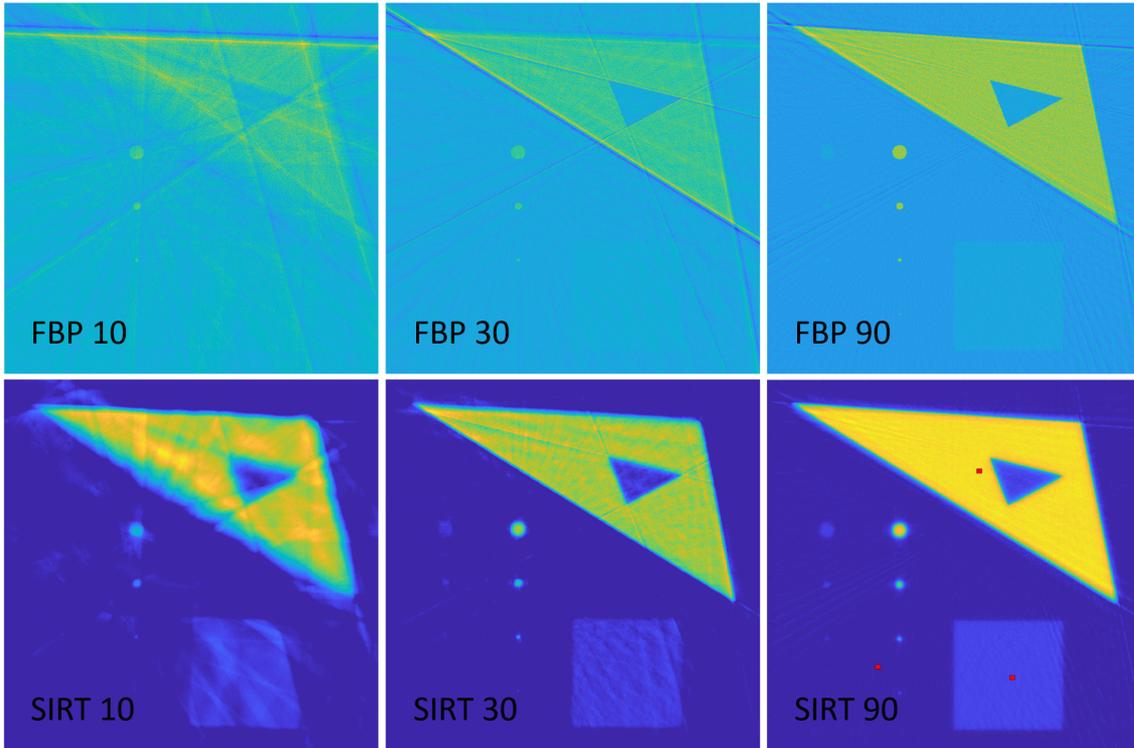
Fig. 4 Reconstructions of the phantom at Q = 0.0976 $\mathring{A}^{-1}$ using FBP or SIRT with 10, 30 or 90 equally spaced projections. The reconstructions performed by means of SIRT show improved signal to noise ratio and fewer artifacts when compared with the reconstructions performed with FBP. The reflectivity curves for Figure 5 are the values at the three red dots in the image of the SIRT reconstruction from 90 projections. Note the pixel size is 6.45 μm, but the red dots are considerably larger to make them visible.

taneous iterative reconstruction technique (SIRT). SIRT is a method that updates every iteration by multiplying the error in the previous iteration with the sums of the columns and rows of **A** simultaneously.

$$\vec{x}_{k+1} = \vec{x}_k + \omega_k \cdot \mathbf{D}^{-1} \cdot \mathbf{A}^T \cdot \mathbf{M}^{-1} \cdot (\vec{b} - \mathbf{A} \cdot \vec{x}_k) \qquad (3)$$

ω_k is a scalar, **D** and **M** are dependent on the ART method used. For SIRT, **D** is a diagonal matrix with the sum of each column of **A** as the diagonal elements, and **M** is a diagonal matrix with the sum of each row of **A** as the diagonal elements. For each iteration the artifacts from reconstruction are reduced, but noise from the measurements also becomes more significant. The optimal number of iterations to use is dependent on the tomography, the experimental setup, the nature of the noise and the exact reconstruction algorithm used. A good stopping rule will terminate the reconstruction when a minimum in noise is obtained.[8,12–16] Determining a stopping rule is a challenging task so a trial and error approach was used to select 150 iterations in the reconstructions: at this number of iterations the exact amount of iterations made minute difference.

3 Results and Discussion

To compare the effect on reconstruction quality from the number of projections in each sinogram, the phantom was reconstructed from 10, 30 and the full 90 projections, as shown in Figure 4. All reconstructed phantoms show the patterns from the map. For the full number of projections the reconstruction produces a clear match with the original map of the sample. It is thus shown that the simulated photons have successfully interacted with the heterogeneous sample. The pixel size is 6.45 μm, but a multilayer structure needs to extend to a larger area before the reflectivity matches the original reflectivity curve. Pixels in the center of the circles with a diameter of 200 μm match the reflectivity curves corresponding to the multilayer structure. The structure of the smaller circles was observable only as irregularities in the reflectivity of the surrounding structure. To create better reconstructions, the number of projections or the pixel density of the detector could be increased. Both of these would lead to more measurements, which would produce a reconstruction in a better quality.

To show that the method can be used to test reconstruction techniques, the reconstructions from the different
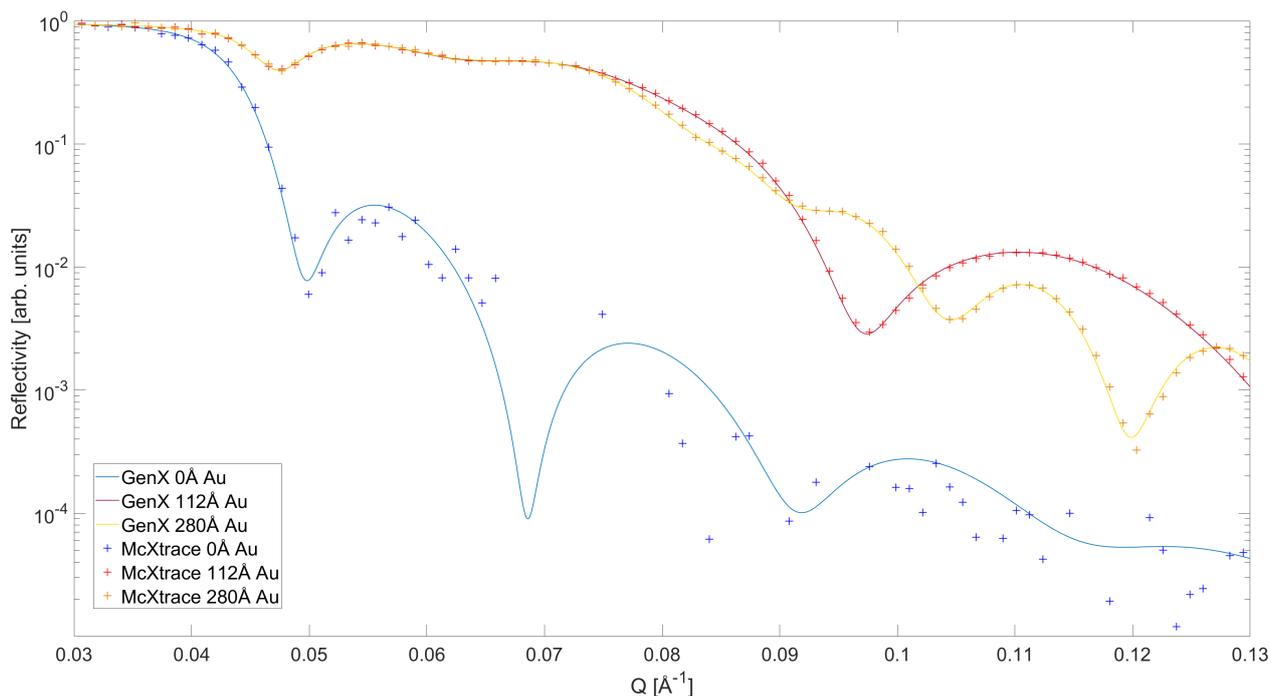
Fig. 5 The reconstructed reflectivities in single pixels of sidelength 6.45 μm show a high degree of correlation with original reflectivity curves. The position of the three pixels can be seen in Figure 4. All the reflectivity data from McXtrace has been multiplied with the same value to obtain the same reflectivity as for GenX for Q = 0.0308 $\mathring{A}^{-1}$. A similar value could also be obtained by extrapolating the curves to Q = 0. This is justified because the reflectivity should always approach 1 as Q approaches 0. The critical angle of Au is clearly visible as a high reflectivity until Q = 0.08 $\mathring{A}^{-1}$. At low reflectivity the uncertainty from artifacts in the reconstruction and the limited photon count in the Monte Carlo method dominates.

amount of projections were compared. As expected, missing information caused the greatest effect on the FBP reconstructions, as can be seen in the number of artifacts. Especially the thinner 112 $\mathring{A}$ layer of Au becomes hard to distinguish since areas with higher contrast create artifacts with more intensity.

The reflectivity at individual positions (micro X-ray reflectivity, μXR) was determined by plotting the reflectivity as a function of Q.[6] This is compared with the initial reflectivity curves in Figure 5. It is evident from the reconstruction that it is possible to determine which of the three multilayer structures the chosen pixels belong to.

To get a quantitative overview of the differences between the qualities of the reconstructions, Mander's overlap coefficient (MOC) was calculated. The purpose was to show that using SIRT with 150 iterations delivered a good reconstruction as shown in Figure 6. Another purpose was to show that the quality of the SIRT reconstruction was superior to that of the FBP. MOC is given

by:[17]

$$r = \frac{\sum_i s1_i \cdot s2_i}{\sqrt{\sum_i (s1_i)^2 \cdot \sum_i (s2_i)^2}} \qquad (4)$$

$s1_i$ was the reflectivity value of a pixel on the map as determined from the reflectivity curve corresponding to the material assigned to that position on the map. $s2_i$ was a weighted average of the reflectivities in the reconstruction that shared position with the map pixel. The weights were equal to the amount of overlap. A weighted average was required because the reconstructions might not have the same pixel position or size as the map.

The reflectivity of all pixels for all 90 incident angles θ were summed up and compared using MOC. A MOC of 1 means the reconstructions correlate perfectly with the map and the reflectivity curves, whereas a MOC of 0 means there is no correlation.[17] This was possible because the perfect result was known in the form of the reflectivity curves of the materials from GenX. The map used as input for McXtrace then assigned these materials to known positions. This was an advantage of using simulation since a reconstruction from experiment would require another

type of experiment with higher precision to determine the quality of a reconstruction. Using all 90 projections, the SIRT reconstruction with 150 iterations had a MOC of 0.993 whereas the reconstruction using FBP had 0.953. Comparison between SIRT and FBP for a different number of projections is shown in Figure 7. The SIRT reconstruction from a beam with a horizontal divergence of 0.02 mrad also had a MOC of 0.993, which shows that a collimated beam produced a good approximation.
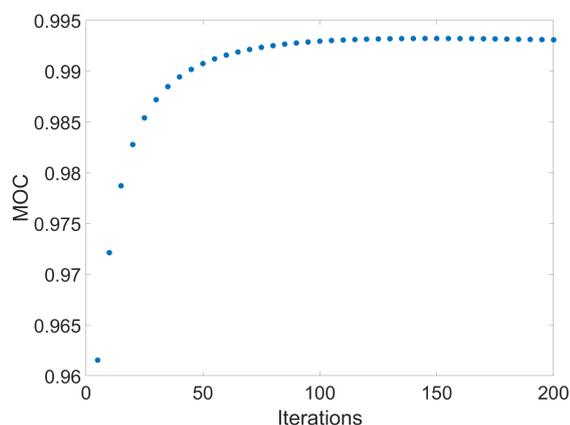
Fig. 6 SIRT reconstruction quality as a function of iterations shows that at 150 iterations the change in quality per iteration is minimal. MOC begins decreasing after approximately 150 iterations, which is why this number of iterations was chosen. MOC is calculated at every 5 iterations using all 90 projections.
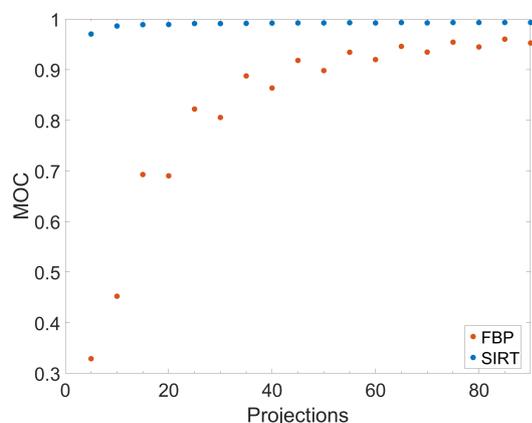
Fig. 7 MOC as a function of the number of equally spaced projections used in the reconstruction. SIRT is clearly the better match to the original reflectivity curves, especially with a small number of projections. Oscillatory behavior occurs because some directions are more important than others for calculating the best reconstruction. MOC is calculated at every 5 equally spaced projections.

McXtrace returns errors on the detector intensities

from the necessary use of a limited number of photons. Reconstruction of these errors using SIRT gave an estimate of errors arising from the Monte Carlo approach. This resulted in an uncertainty in the μXR curves in the order of 10^{-2} decreasing to 10^{-4} as Q is increased. This together with the artifacts from the reconstruction and the bandwidth of the beam is the source of any uncertainties in the μXR curves. Error bars are omitted since the exact contribution from artifacts is undetermined and depends on the positions as well as the exact reconstruction algorithm used. In real experiments uncertainty also arises from dispersion of the beam, which affects the incident angle and thus the magnitude of Q for the interaction. The lateral size of the evanescent wave means that a photon hitting the sample in a heterogeneous area will interact with several multilayer structures. This effect was not included in the simulations since only one reflectivity curve was used for any one photon. The effect of diffuse scattering from rough interfaces to other areas of the detector was not included in the simulations. However, diffuse scattering contributes significantly less than specular reflection. The reconstructed reflectivity curves for areas with low reflectivity might be affected by diffuse scattering from other areas of the sample.

4   Conclusions

Based on the full reflectivity dataset it has been shown that the Monte Carlo approach produces results which can be used to test reconstruction techniques. Since the optimal result is known, it is possible to calculate measures to quantitatively evaluate the reconstruction quality. The reconstruction has shown that there is a contrast between different areas and that it is possible to determine where the thickness of a layer changes. The μXR curves show that it is possible to determine the reflectivity as a function of both position and wavevector transfer. This means that the tomography of the phantom can be determined and it shows that the simulations return the expected results. The use of Monte Carlo simulations produce more realistic noise, and it is possible to add more factors, as deemed necessary. By using the McXtrace environment, it is possible to change the source characteristics, the optics and the detector. This can be achieved by using the already included components or by adding new ones.

4.1   Outlook

During the continued development of these simulations it will be important that the noise generated corresponds to real experimental noise. That way, the simulations can be used to determine optimal experimental setups and reconstruction techniques and to get an idea of the image quality

to be expected from these experiments. When this is considered, it will be a stepping stone towards creating optimal XRT experiments on lab-source experimental setups.

5   Acknowledgments

References

1   J. Jiang, K. Hirano and K. Sakurai, *Journal of Applied Crystallography*, 2017, **50**, 712-721.

2   V. A. Innis-Samson, M. Mizusawa, K. Sakurai, *Adv. X-Ray. Chem. Anal., Japan* 2012, **43**, 391-400.

3   V. A. Innis-Samson, M. Mizusawa and K. Sakurai, *Analytical chemistry*, 2011, **83**, 7600-7602.

4   J. Als-Nielsen and D. McMorrow, *Elements of Modern X-ray Physics*, John Wiley & Sons Ltd, Chichester, 2011.

5   H. Kiessig, *Annalen der Physik*, 1931, **402**, 769-788.

6   S. Kenji, M. Mari, I. Masashi, K. Shun-ichi and I. Yasuhiko, *Journal of Physics: Conference Series*, 2007, **83**, 012001.

7   L. G. Parratt, *Physical Review*, 1954, **95**, 359-369.

8   P. C. Hansen and J. S. Jœrgensen, *Numerical Algorithms*, 2017, DOI: 10.1007/s11075-017-0430-x.

9   R. Gordon, R. Bender, G. T. Herman, *Journal of theoretical biology*, 1970, **29**, 471-481.

10   E. Bergbäck Knudsen, A. Prodi, J. Baltser, M. Thomsen, P. Kjær Willendrup, M. Sanchez Del Rio, C. Ferrero, E. Farhi, K. Haldrup, A. Vickery, R. Feidenhans'l, K. Mortensen, M. Meedom Nielsen, H. Friis Poulsen, S. Schmidt and K. Lefmann, *Journal of Applied Crystallography*, 2013, **46**, 679-696.

11   M. Björck and G. Andersson, *Journal of Applied Crystallography*, 2007, **40**, 1174-1178.

12   P. Gilbert, *Journal of theoretical biology*, 1972 **36**, 105-117.

13   A. van Der Sluis and H. A. van Der Vorst, *Linear Algebra and its Applications*, 1990, **130**, 257-303.

14   W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg and J. Sijbers, *Optics Express*, 2016, **24**, 25129-25147.

15   W. van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg and J. Sijbers, *Ultramicroscopy*, 2015, **157**, 35-47.

16   W. J. Palenstijn, K. J. Batenburg and J. Sijbers, *Journal of Structural Biology*, 2011, **176**, 250-253.

17   E. M. Manders, F. J. Verbeek, J. A. Aten, *Journal of Microscopy*, 1993, **169**, 375-382.