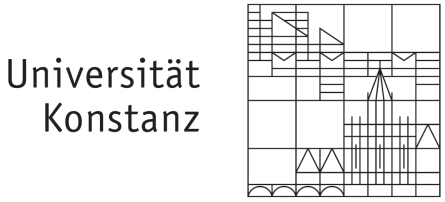
Bachelor Thesis

A droplet based technique to measure protein nucleation rates

presented by Daniel Gerz at



Department of Physics

First Supervisor: Prof. Dr. G. Maret

Second Supervisor: Prof. Dr. S. Fraden

Summer Semester 2014

The research was conducted at Brandeis University.

Abstract

In this work a revised method to determine the nucleation rate of protein crystals is presented.
In accordance with the microfluidic approach, large numbers of identical, sub-nanoliter drops are used. The protein supersaturated drops are all quenched simultaneously and at constant temperature, the number of drops without a crystal are counted regularly.
The system was adjusted to allow analyzing thousands of experiments. Through these improvements an automated image analysis was implemented in Matlab. Measurements were made with the the standard protein, Glucose Isomerase, using the common precipitants PEG and Ammonium Sulfate. The results are in agreement with the last carried out measurements of the protein lysozym.

Zusammenfassung

In dieser Arbeit wird eine verbesserte Methode zur Bestimmung der Nukleationsraten von Proteinkristallen vorgestellt. Dem verwendeten Mikrofluidikansatz entsprechend, wird beim Experimentieren eine große Anzahl von identischen, sub-nanoliter Tropfen verwendet. Diese mit Protein übersättigten Tropfen werden alle gleichzeitig abgeschreckt und bei konstanter Temperatur wird die Anzahl der Tropfen, die keinen Kristall besitzen, in regelmäßigen Zeitschritten gezählt.
Das System wurde so angepasst, dass das Analysieren großer Datenmengen möglich ist. Durch diese Verbesserungen konnte eine automatische Bildverarbeitung in Matlab implementiert werden.
Messungen wurden am Standardprotein Glucose Isomerase unter Verwendung der geläufigen Präzipitationsmittel PEG und Ammonium Sulfat durchgeführt. Die Ergebnisse stehen in guter Übereinstimmung mit den bisher durchgeführten Messungen an dem Protein Lysozym.

Contents

1 Introduction

Proteins play a vast role in living beings and the structure of proteins is essential for their function. The most common way to determine a protein structure is with X-ray diffraction from protein crystals. This is the reason why many people are researching protein nucleation.
Proteins are highly complex macromolecues so that predicting, whether or not there is even a crystalline phase in the vast chemical space. Even if there is a crystalline phase, there might not be a suitable way past other metastabil protein phases. In this field, where trial and error experiments still occur on a daily bases, a better understanding of the underlying processes will help to improve the success rate of nucleation tryouts.
The previous experiments wanted to answer the questions to what extent nucleation is homogeneous or heterogeneous and if the measured data can be explained with CNT.
This thesis has two other goals: optimizing the experimental setup, so that necessary statistics can be more easily gathered and the measurements of nucleation rates of Glucose Isomerase with two of the most often used precipitants: PEG and Ammonium Sulfate.

2 Background

This background is meant to help understand the previous work done with this setup and thus strongly orientates itself on [1, 2].

2.1 Nucleation

Nucleation marks the start of a first-order phase transition. Classical Nucleation Theory (CNT) describes the kinetics of this process under some fundamental simplifications. The first assumption is that microscopic properties and activities are identical with macroscopic ones, e.g. that pre-critical and post-critical clusters have an identical structure. The second major simplification is the treatment of molecules as hard spheres, discarding interactions due to shape or inhomogeneity of the molecule surface.

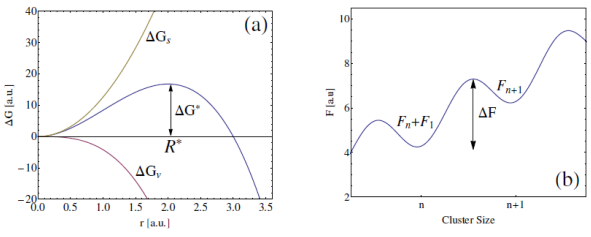


Figure 1: (a) The free energy ΔG as a function of radius r. "The energy barrier, ΔG^*, for nucleation is a result of the competition between the volume free energy, Δ_v and the surface free energy, ΔG_s." (b) Activation energy for addition of one monomer to a cluster of n [2].

Under these assumptions the total change in free energy ΔG for creating a nucleus from the bulk phase just consists out of two terms: the bulk term ΔG_v, representing the energy gain in transferring the free molecules from solution into the sphere shaped bulk phase and the surface term ΔG_s describing the free energy cost for creating the interface between the nucleus and the solution. These two competing terms add up to:

$$\Delta G = \Delta G_v + \Delta G_s = -\frac{4\pi r s}{3}\rho_c \Delta\mu + 4\pi r^2 \gamma. \tag{1}$$

Where ρ_c is the number density of nucleating phase, r is the radius of the spherical cluster, $\Delta\mu$ is the chemical potential difference between a molecule in the solution phase and in the nucleus phase, and γ is the interfacial tension. Due to the opposite signs of ΔG_v and ΔG_s, the different scaling, in respect to the radius of the cluster, leads to a finite maximum of ΔG, the activation barrier ΔG^* at the radius of the critical nucleus R^*. To lower their free energy, all nuclei with a radius smaller than R^* need to shrink, while all cluster bigger than R^* lower their free energy by growing. With equation 1 we find:

$$\frac{\partial \Delta G}{\partial r} = 0 \qquad \Rightarrow \qquad R^* = \frac{2\gamma}{\rho_c \Delta\mu} \tag{2}$$

$$\left.\frac{\partial \Delta G}{\partial r}\right|_{R^*} = 0 \qquad \Rightarrow \qquad \Delta G^* = \frac{16\pi}{3}\frac{\gamma^3}{(\rho_c \Delta\mu)^2}. \tag{3}$$

With ρ_N, the number density of nucleation sites and the Boltzmann weight $e^{-\Delta G^*/k_B T}$, the number of critical nuclei per unit volume is $\rho_N e^{-\Delta G^*/k_b T}$. Once a nucleus reaches the critical size, it can go further into a stable phase by incorporating molecules. The probability that this growth happens, rather than the nucleus going back into the metastable phase, is given by the Zeldovich factor Z. With the rate j, at which single molecules add to a critical nucleus, the nucleation rate can be written as:

$$J = \rho_N Z j e^{-\Delta G^*/k_B T}. \tag{4}$$

The required activation energy to attach a monomer to an existing cluster is ΔF as can be seen in figure 1. Therefore, $e^{-\Delta F/k_B T}$ is the probability that a collision between a monomer and a critical nucleus results in the attachment of the monomer. The frequency at which monomers impinge upon a critical nucleus is diffusion limited. With ρ as the number density of monomers and D as the diffusion constant for monomers, this diffusion limited rate can be represented by $4\pi\rho D R^*$. The rate j can be approximated as:

$$j \simeq 4\pi\rho D R^* e^{-\Delta F/k_B T}. \tag{5}$$

So far, only nucleation from homogeneous solution is considered and therefore homogeneous nucleation. In the case of heterogeneous nucleation, formation of the nucleus on a surface, e.g. a dust corn, the surface tension is reduced from γ to γ_{het}. A nucleation barrier still remains, so that for the heterogeneous barrier height only γ is replaced by γ_{het} in equation 3.
The nucleation rates of homogeneous and heterogeneous nucleation are similar in form,

$$J_{hom} \simeq 4\pi\rho^2 D R^* Z e^{-\Delta F/k_B T} e^{-\Delta G^*_{hom}/k_B T}$$
$$J_{het} \simeq 4\pi\rho\rho_N D R^* Z e^{-\Delta F/k_B T} e^{-\Delta G^*_{het}/k_B T}.$$

Here we assume that the free energy barrier for adding a monomer to the cluster ΔF is independent of concentration and whether or not heterogeneous or homogeneous nucleation occurs. These assumptions are based on the simplification of CNT, that the pre-critical and post-critical nucleus have the same form, as well as the crystal shape being independent of the nucleation mechanism.
The number density of nucleation sites is in homogeneous nucleation equal to the number density of monomers, because every individual monomer can act as a nucleation site, hence $\rho_N = \rho$. On the other hand, in heterogeneous nucleation the nucleation sites are impurities and their number density is generally much smaller, $\rho_N << \rho$.
To express the nucleation rate in terms of supersaturation σ and barrier height B, the protein solution is considered as an ideal solution. The chemical potential is set to $\sigma = \Delta\mu/k_BT = \ln(C/C_S)$, where C is the actual concentration of the monomer and $C_S(T)$ the equilibrium concentration of the monomer in the solution at the coexistence of both phases. The number density of monomers can be written as

$$\rho = \frac{C * N_A}{M_W} = \frac{C_S(T)e^{sigma} * N_A}{M_W}, \tag{6}$$

where M_W is the molecular weight of a monomer.
The nucleation rate can be rewritten with the supersaturation, defined as $\sigma = \Delta\mu/k_BT$:

$$J = \frac{4\pi\rho_N N_A C_S(T) DR * Z}{M_W} e^{-\frac{\Delta F}{k_B T}} e^{\sigma - \frac{\Delta G^*}{k_B T}}. \tag{7}$$

By measurements of J, σ, C_S and T, the barrier height ΔG^* can be estimated.

2.2 Nucleation Rates In Drops

A drop which contains a supersaturated protein solution, is considered instantly quenched, so that the nucleation rate changes from 0 to J. Moreover, it is assumed that nucleation is a Poisson process. The probability that the drop is nucleating is Jvt at constant volume and nucleation rate. The probability that there is no nucleation event happening in an uncrystallized drop in the time frame of Δt is $e^{-Jv\Delta t}$. The fraction of uncrystallized drops f_ϕ of a population of N drops is then

$$f_\phi = \frac{N_\phi(t)}{N} e^{-Jv\Delta t}. \tag{8}$$

Let's assume there are two different pathways S_1 and S_2 to form a crystal and consequentially two rates J_1 and J_2. In a system with N identical drops, containing the possibility for both pathways. The rate equation is $\frac{dN}{dt} = -(J_1 + J_2)vN$, which leads to $f_\phi = e^{-(J_1+J_2)vt}$, a simple exponential decay of uncrystallized drops.
Assuming the drops are heterogeneous with at least two different nucleation rates. The chemical rate equations result in a two exponential decay of the fraction of uncrystallized drops.

$$f_\phi(t) = f_1 e^{-J_1 k_B t} + f_2 e^{-J_2 k_B t}. \tag{9}$$

So only in a system with more than one kind of drops, multi-exponential decay rates are possible.

2.3 Inverse Laplace Transformation

A general approach to identify the decay rate distribution g(s) uses the fact that $f_\phi(t)$, the fraction of not crystallized drops, is the Laplace transform of g(s)

$$f_\phi(t) = \mathcal{L}g(s) = \int_0^\infty g(s)e^{-ts}ds. \tag{10}$$

The calculation of the inverse Laplace transform is an ill-posed problem. A numerical computation using the Tikhonov Regularization is done to obtain g(s). The following equation is minimized

$$g(s,\alpha) = \min_{g(s)} \left\{ \left\| f_\phi(t) - \int_0^\infty g(s)e^{-ts}ds \right\|^2 + \alpha \left\| g(s) \right\|^2 \right\}. \tag{11}$$

Here is α the regularization parameter that acts as a low pass for the distribution $g(s)$. The Matlab function fmincon can be used to minimize equation 11 under the constrain of non-negativity.

2.4 Pound and La Mer model

In [1] the decay rates stemming from the inverse laplace transformation have always two peaks. This indicates that the experimental data supports a model with three parameters (rate one, rate two and the ratio between them how strong a decay rate is present) at most. The Pound and La Mer model [3] is such a model with three parameters.
Pond and La Mer consider nucleation happening only from nucleation sites inside and not from solution. This model is slightly changed here, so that If k_0 is the rate at which nucleation occurs in a drop without any additional nucleation site and a drop with p nucleation sites has the rate (k_0 +p k). For $k_0 << k$ the situation is nearly the same. If we assume a Poisson distribution of the nucleation sites, with m as the average number of nucleation sites, the probability s that a drop contains p nucleation sites is

$$s(p) = \frac{e^{-m}m^p}{p!} \tag{12}$$

The fraction of drops without a crystal f_ϕ can now be written as:

$$f_\phi(t) = \sum_{p=0}^{i} nftys(p)e^{-(k_0+pk)t} = \sum_{p=0}^{i} nfty\frac{e^{-m}m^p}{p!}e^{-(k_0+pk)t} = e^{-m}e^{-k_0t}e^{me^{-kt}} \tag{13}$$

Monte-Carlo Simulation were performed to estimate the errors associated with the experimental system. Because the order of magnitude of the parameters that was used to perform the simulation is intrinsic to the setup, e.g. the measurement takes 1-5 days and only nucleation rates that are resolvable in this timeframe are suited to be found.
The simulation came to the conclusion that approximately 500 drops are needed to have the nucleation within a 10% confidence interval.

2.5 Materials

Glucose isomerase also known as Xylose isomerase, is an enzyme that participates in the metabolism of Saccharides. Because of its industrial use for converting Glucose into Fructose, it is well researched and can be easily obtained.
The protein (MW:173kDa) is a tetramer, composed out of four identical polypeptides of ≈43kDa. On a SDS-PAGE the pure protein should appear as a clear single band at the polypeptides weight.
Glucose isomerase is most stable between pH 6.0 and 8.0 and rapidly denatures below pH 5.0 [4]. The estimated isoelectric point is 3pH.

We bought Glucose isomerase purified from *Streptomyces rubiginosus* from Hampton Research in a Ammonium sulfate rich crystal suspension. The suspension was extensively dialysed at room temperature in a Spectra/Por 2 Dialysis Membrane (MWCO: 12-14kDa) against a 10mM Tris-HCl pH 6.9 buffer solution. An Amicon centrifugal filter unit (MWCO 30kDa) from Millipore was used to obtain the desired protein concentrations.
The UV absorbance was measured at 280nm of 2μl protein solution samples with the Nanodrop 2000c (Thermo Scientific). The concentration then was determined with the Bert-Lambert law, using a extinction coefficient of 45660 $M^{-1}cm^{-1}$ [4].
The pureness of the protein was examined with a SDS-PAGE (see fig.2). The expected 43kDa band is dominant but smeared out and there are several weak bands with a molecular weight under 43kDa. The protein is Therefore not 100% pure, but under the knowledge of these impurities good enough to obtain qualitative data.

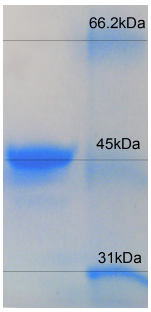


Figure 2: SDS-PAGE. Left: GI (0.5 mg/ml) Right: Broadrange standard ladder

Polyethylene glycol (PEG) ($H-(O-CH_2-CH_2)_n-OH$) is a hydrophilic, water soluble polymer, which only has mild effects on biological activity [5]. It is Therefore the most extensively used polymer in aqueous solutions with biological molecules.
Although the attractive depletion force, predicted by the Asakura-Oosawa model, usually is in agreement with observation, the effect can be system dependent as well [5]. It was shown in [6], that depending on the molecular weight of PEG, the cloud point of a lysozyme-PEG mixture can shift either to a higher or lower temperature. The authors suggest that an energetic attractive interaction between lyzozyme and PEG must be considered.
PEG 10kDa was purchased as flakes from Sigma-Aldrich, CAS: 25322-68-3 and used at 20% w/v as a precipitant.

Ammonium sulfate, $(NH_4)_2SO_4$ is an inorganic salt that is often used to precipitate protein. It is very water soluble, which allows a high ionic strength and consists out of two ions which are ranking high in the Hofmeister series. Ammonium sulfate is well suited to achieve a salting out of the protein. Additionally, there is no adverse effect on enzymatic activity, so that it well suited

for protein precipitation.
Ammonium sulfate was purchased from Sigma, CAS: 7783-20-2.

Tris HCl ($NH_2C(CH_2OH)_3 \cdot HCl$) is a buffer, widely used in biochemistry, because it does not inhibit the activity of most enzymes. It is a suitable buffer for a pH between 7 and 9.
As for a buffer it does exhibit a strong temperature dependence, more precisely, there is an pH increase of about 0.03 pH per °C decrease [7].

EA is a surfactant based on PFPE-PEG, which was especially designed for compartmentalizing aqueous drops in fluorocarbon oil. Those emulsions are compatible with both, PDMS microfluidic devices and biological systems [8]. The designed surfactant prevents adsorbtion of the protein to the drop surface and no difference between the enzyme kinetics in and out of the drops could be observed [9].
1.8% w/w EA surfactant (RainDance Technologies, Inc.) was used in HFE-7500, a fluorinated oil from 3M, to create a stable emulsion.

The following precipitant solutions were prepared before the experiment:

1. 40% w/v PEG 10kDa, buffered with 10mM Tris-HCl at pH 7.0
2. 20% w/v PEG 10kDa, 200mM Ammonium sulfate, buffered with 10mM Tris-HCl at pH 7.0

All the buffer, protein and salt solutions were filtered through sterile 0.2μm VWR cellulose acetate filters. For this experiment only the concentration of the protein solution was varied, while the precipitant concentrations stayed the same, to measure nucleation rates at different supersaturations, but identical precipitant conditions.

3 Experimental Set-up

We used a flow-focusing microfluidic polydimethylsiloxane (PDMS) device to create an emulsion of monodisperse aqueous droplets in oil [10]. To avoid any nucleation before starting the experiment, first proper crystallization conditions were established, directly within the drops on-chip. For this, a co-flow focusing nozzle is used, as shown in figure 3. Note the straight line in the middle of the channel between the protein stream and the precipitant stream.
Because both solutions in the streams possess the same viscosity, this indicates that the flow rates of both streams are the same. The fact that the line is straight, suggests that there is a laminar co-flow. For this case, it was previously shown via cloud point measurements, that the chemical composition within the drops varies at most 3% from each other [2].

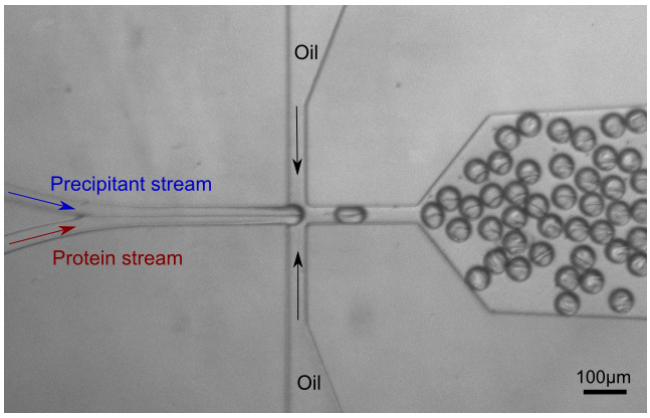


Figure 3: Generation of practically identical droplets using a co-flow microfluidic device. The contents of the drops mix rapidly over time, as can be seen by the border in the drops getting less defined, the closer the drops are getting to the exit. The oil flow rate was set at 550μm/h and to 100μm/h for the precipitant and protein solutions.
The final concentration in drops is 20% w/v PEG 10kDa, 100mM Ammonium sulfate with a quarter of the original protein concentration.

The protein stream contains a 1:1 mixture of the Glucose Isomerase solution and the first precipitant solution resulting into 20% w/v PEG 10kDa, 10mM Tris HCl and half the original protein concentration, while the precipitant stream contains only the second precipitant solution. Therefore, both streams have the same PEG concentration, which determines, in this case, the viscosity. The flow rates of those two streams were kept identical, resulting in a 1:1 mixture with 20% w/v PEG 10kDa, 100mM Ammonium sulfate, 10mM Tris-HCl and a quarter of the original protein concentration at approximately pH 7.0.
The volume of the drops is determined by the flow rates, the size of the nozzle and how hydrophilic the surface of the nozzle is. To get the desired size of droplets, the flow rates were adjusted.
The droplets leaving the chip are temporarily stored as an emulsion in an 200μl Eppendorf vial sealed with Parafilm.
The emulsion is then loaded into rectangular, hydrophobic glass capillaries via capillary forces. The inner dimensions are either 1mm x 50μm or 1mm x 100μm with an $\pm$ 10% uncertainty for each dimension.
A monolayer of drops is desirable to visually separate the drops into regions of interests. Consequentially the size of the drops should be chosen in a way that the diameter is bigger than the height of the capillary. A quick sealing of the capillary ends that has only a negligible influence on the droplets, is achieved with VALAP, a low melting wax consisting to equals parts out of **Va**seline, **La**nolin and low melting temperature **P**araffin wax.

The sealing fixes the capillaries on the center of a microscope glass slide (75 x 50 x 1 mm), which forms with another microscope slide and a poly-siloxane (Pr. No. 3788T24, McMaster-Carr) washer a water chamber. This ensures a more homogeneous temperature at the sample. On top of the water chamber, an air chamber is created with another microscope slide and washer. The air

chamber insulates the sample from the surroundings and prevents condensation of humid air in the optical path.
The sample is then mounted on a temperature controllable XYZ-scanner. For a better thermal contact, a thin Paraffin oil film is added between the sample and the aluminum plate of the scanner. The setup is shown in figure 4.

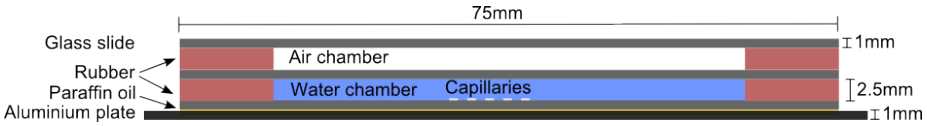


Figure 4: Longitudinal cross section of the mounted sample. The visible five capillaries are glued to the bottom glass slide with VALAP.

Images of defined positions in the capillary are taken at discrete time intervals and at multiple foci per position, until more than 80% of the drops contain at least one crystal.

3.1 Scanning Stages

The scanner was specially designed in a collaboration with Olin College for monitoring microfluidic devices over time at a controlled temperature. The stage in figure 5 is one out of three nearly identical stages.
Two stepping motors enable repeatable scanning of the sample in the xy-plane at a translation precision of less than 10μm. In z-direction a stepping motor moves the counterbalanced, optical setup in a precision of 0.625μm.
At both ends of the aluminum temperature stage (200 x 50 x 7 mm) water cooled, independently controlled Peltier element allows adjustment of the temperature.
The scanner is controlled with a field-programmable gate array (FPGA) via a Labview Interface.

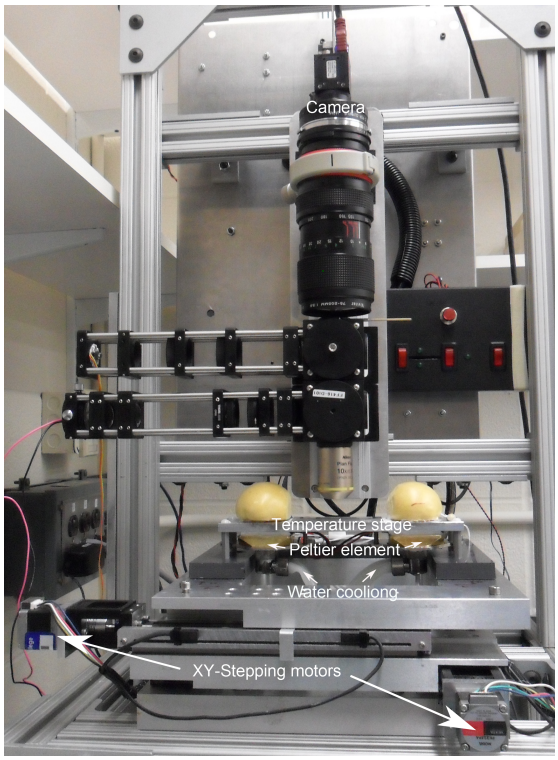


Figure 5: The original scanning stage designed by students of Olin College.

3.2 Temperature control

The solubility of Glucose isomerase changes exponentially with the temperature and the nucleation rate therefore depends strongly on the temperature. Consequently, a good, consistent temperature calibration is crucial for the experiment. We used the Fluke 52K Thermometer as a reference to get a consistent calibration for all used thermistors in our system. The Fluke 52K Thermometer is a thermocouple thermometer with an accuracy of $\pm 0.1°C$ [11]. Beside an ice bath, we used several water thermostats to calibrate the thermocouple and found the discrepancy in our temperature range (4-25 °C) to be at most 0.1 °C.
For each scanner two thermistors are needed to follow the stage temperature and a third thermistor is used to measure the room temperature. The measured temperature dependent resistance of the thermistors was fitted to the linear Steinhart-Hart equation:

$$\frac{1}{T} = a + b\ln(R). \tag{14}$$

The resulting fit deviated from the measured data in the temperature range 0.14°C at most. For the temperature at which nucleation rates were measured (6°C-15°C), the calibration error is only 0.05°C. The whole distance between the sensors is 16cm and the length of the maximum sample area is 3.3cm. As a result, an error of $\approx 0.01°C$ in the sample area is induced due to the callibration.

The temperature at the sample is, in general, not the same as the temperature at the thermo-controlled ends of the stage. The gradient between the room and the stage temperature ΔT_{RS} leads to an horizontal and vertical gradient in the sample.
Under the assumption that the thermal conductivity k of the used materials does not considerably change over the used range, an estimation of the gradients can be made with Fouriers's law of thermal conduction:

$$\vec{q} = -k\nabla T. \tag{15}$$

In the stationary case the local heat flux density $\vec{q}$ is constant. If we only consider the heat flux in z-direction, $\Delta_z T$ is directly proportional to $-1/k$. Consequently it holds:

$$\frac{\Delta T_{\Delta z}}{\Delta T_{tot}} = \frac{R_{\Delta z}}{R_{tot}} \tag{16}$$

where $R_{\Delta z} = \frac{\Delta z}{k}$ is the thermal resistance and the subscript tot is standing for the total change.

	Air	Water	Soda Lime Glass
k in $\frac{W}{m \cdot K}$	0.0026	0.5862	1.00

Table 1: Literature values for thermal conductivities from [12].

Using an average outside capillary height of 0.2mm and the setup as shown in figure 6, the vertical temperature gradient over the capillary amounts to only 3.3‰ of the gradient between 1 and 4.
To quantify the vertical gradient, the temperature T_T was measured with the calibration thermometer in the center of the sample at four different heights, see figure 6. The temperature was allowed 5min for each measurement to reach a state close to equilibrium. This plotted data in figure 7 shows the expected linear behavior. All y-interceptions of the linear fits are zero within the error margin, which indicates valid data. With the slopes of positions 2 and 4 and table 1, the slope of position 3 is calculated to be 0.081. This agrees with the slope from the linear fit for this position, so that it can be assumed that the material parameters are correct. Accordingly the estimation for vertical gradient should be close to the estimated value.

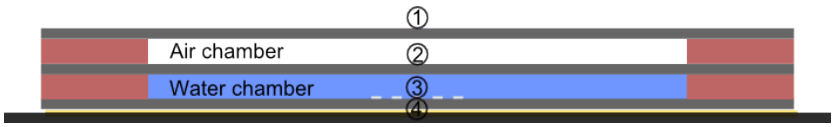


Figure 6: Positions for measuring the temperature. 1 Top of air chamber; 2 Top of water chamber; 3 Sample position; 4 Directly on metal stage.

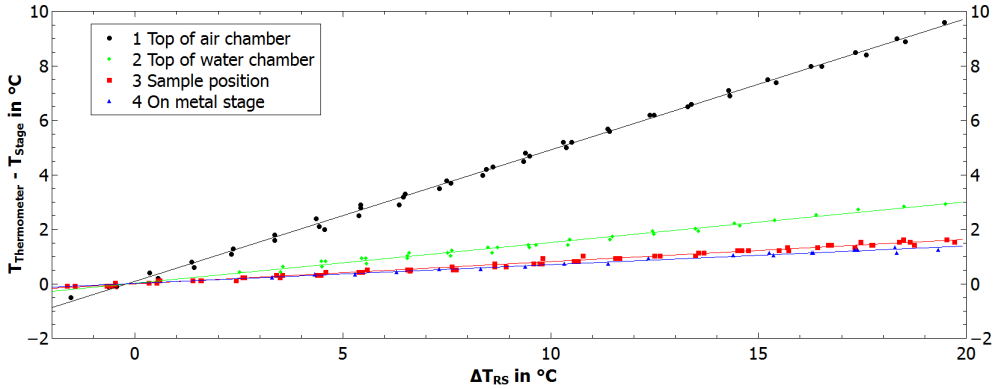


Figure 7: Measured gradients of in figure 6 marked positions, plotted versus the total temperature gradient.

position	A in 10^{-3}	B in 10^{-3} °C
1 Top of air chamber	482 ±3	95 ±30
2 Top of water chamber	149±3	22±3
3 Sample position	81 ±1	5 ±12
4 On metal stage	68 ±2	20 ±30

Table 2: Fits obtained from figure 7 using the function: A·x+B

The vertical gradient in the the capillary equals 3.3‰ of the temperature gradient between 1 and 4. Converted to a slope, in regard to ΔT_{RS}, this equals 1.37 10^{-3}. This means that even at a high temperature gradient of 20°C, the vertical gradient only amounts to ≈ 0.02.

To measure the horizontal gradients, five thermistors were calibrated in a stirred water bath and glued on a microscope slide with epoxy, see positions in figure 8. It allowed monitoring the temperature simultaneously at 5 different positions over time. For the measurement the room temperature was varied between 25 and 29°C and the sample was turned 180° in the middle, to expose possible systematic errors.

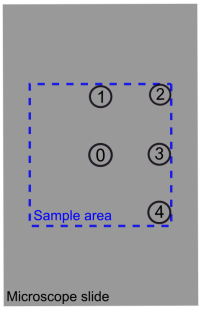


Figure 8: Positions used for temperature measurements in order to estimate the horizontal gradients in the sample area (3.3 x 3.3 cm).

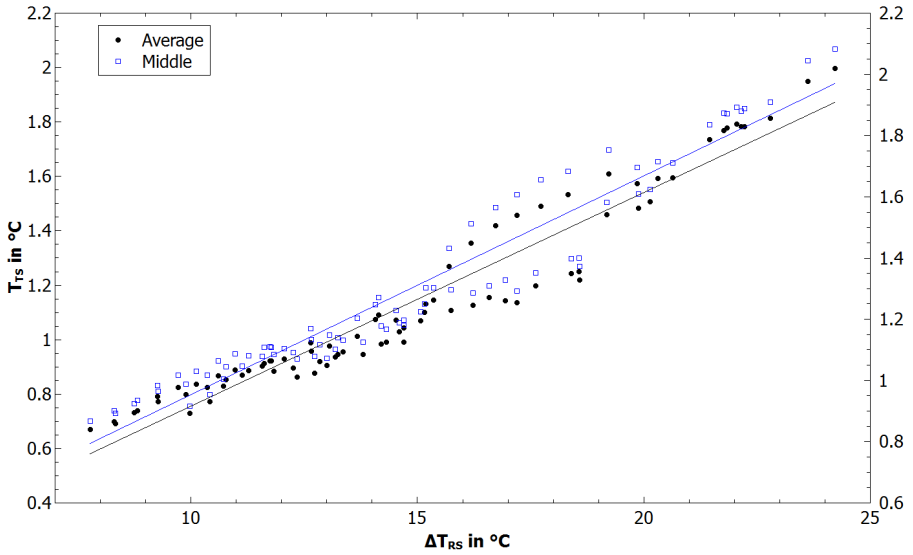


Figure 9: The temperature gradient between the stage and the thermistor T_{ST} is plotted versus the overall temperature difference. In black the average from the 5 positions was used, while in blue only the thermistor in the center (position 0) of the sample area was used.

There was no visible distinction between the data obtained at different room temperatures or in the mirrored position. Therefore, data was plotted together in figure 9.
The slope of the fit for the thermistor in the center is 0.0804 ± 0.0025 and consequently the same as the previously determined slope within the error margin. This and the fact that the fits start in the origin, indicate that the calibration of the thermistors is consistent.
For the average thermistor temperature the slope is with 0.0785 ± 0.002 slightly smaller. This can be explained with figure 10, where the deviation of the measured temperatures from the average temperature is shown. A clear linear trend is visible, although for the linear fits of positions 0, 3 and 4 the origin of the plot is outside the error interval of the y-intersection. This calibration error is 0.016 at most and still allows a quantitative statement.
The slope for the center position amounts to 0.0023 ± 0.0003 and explains the difference of 0.0019 ± 0.0045 in the slope in figure 9.
Figure 10 clearly shows that the gradient is bigger the farther the thermistor is from the Peltier elements. This indicates that the heat flux through the washer plays a minor role compared to the heat flux between the aluminum stage and the air.
With the fits from 10 the horizontal temperature error δT_{hor} in the sample area can be estimated with the steepest slope from figure 10 to be ≈ 0.009.

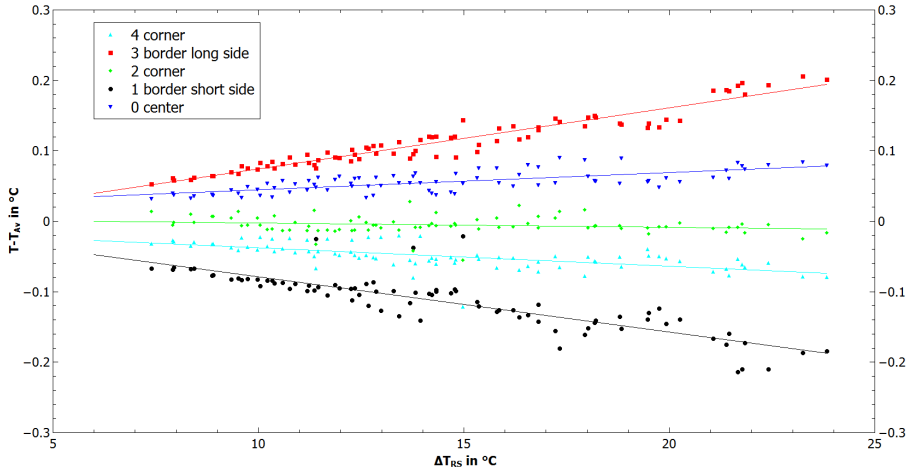


Figure 10: The difference between the temperature and the average temperature at the 5 positions in the figure 8 in plotted against the room - stage temperature gradient ΔT_{RS}.

position	A in 10^{-3}	B in 10^{-3} °C
0 center	2.4 ±0.3	20 ±4
1 border short side	-7.9 ±0.6	-0.8 ±1.8
2 corner	0.6±0.3	3.3 ±4.8
3 border long side	8.6±0.4	-12 ±0.9
4 corner	-2.6±0.4	-11±5.5

Table 3: Fits obtained from figure 10 using the function: A·x+B

To illustrate the horizontal temperature gradients in the sample area the slopes were treated as exact and a linear interpolation was applied, see figure 11.

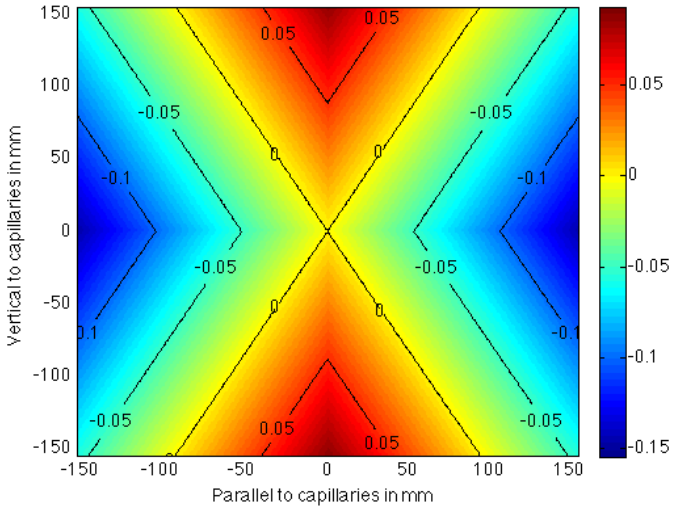


Figure 11: Linear interpolated temperature distribution in the sample area, where the center was set to zero, the ΔT_{RS} values 15°C and the slopes in the corners were averaged.

The temperature in the sample area depends on the stage temperature T_S and the ΔT_{RS}. While T_S is being controlled ΔT_{RS} varies with the room temperature approximately around 1°C during a day. Using 3 the temperature change at the sample can be estimated to be 0.08°C. To avoid this unnecessary temperature change, the air temperature is monitored and the stage temperature is adjusted to keep ΔT_{RS} stable.

With the temperature at the center position set as the exact temperature, the total temperature gradient δT_{tot} in the sample can now be estimated:

$$\begin{aligned}\delta T_{tot} &= \delta T_{hor} + \delta T_{vertical} + \delta T_{callibration} \\ &\approx (14.4) \cdot 10^{-3} \cdot \Delta T_{RS} + 0.01.\end{aligned}$$

3.3 Optics

The most labor-intensive step in this technique is to count the thousands of crystallized and uncrystallized drops over time. The automatization of this step with a computer program requires a good image quality in terms of contrast, depth of focus and homogeneity of the background. To fit those needs, the optical setup was modified, see figure 13.
The sample is mounted on an aluminum stage on the scanner, to control the temperature from one side. A simple reflective light microscope accounts for this restriction, but yields a very low contrast because the sample is nearly transparent. With a polished metal plate underneath the sample, transmitted light can be reflected back into the objective. In this case, a stronger contrast for the nearly transparent sample mainly is generated due to refraction.

The oil has a refractive index of 1.29 [13] and the refractive index of the drop content can be estimated, under the assumption that the protein, buffer and the ammonium sulfate concentrations are neglectable, to be ≈ 1.36 [14].

The aqueous drops work consequently as focusing lenses. In case that the focal length is bigger than the distance between the drop an the reflecting surface, bright disks in the center of the drops can be visible, see figure 12. To avoid those disks, a 1mm microscope slide was used to separate capillary metal plate. In addition, this resulted in a smoother background, because the partly scratchy metal surface is totally unfocused. In this transmission microscope the borders with a change of refraction index appear dark.

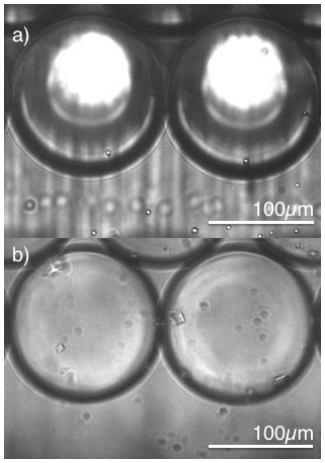


Figure 12: Drops in capillary working as a focusing lens. Capillary separated from reflacting surface by a) a thin cover slide (≈0.15mm) b) a thicker microscope slide (≈1mm)

To ensure illumination, the bright-field microscope was set up to work with Köhler illumination, where the light source is completely unfocused at the sample plane. The adjustable aperture diaphragm allows to alter the effective numerical aperture and the depth of focus.
The schematic of the optical paths of the used Köhler epi-illumination are shown in figure 13. The used components in arm a) ordered sequentially in a Köhler illumination are: A red LED (Philips LumiLED, 630 nm, 350 mA) as the light source, followed by a condenser lens (f_c=15mm), aperture and field diaphragm (Thorlabs SM1D12D), a transfer lens (f_t = 50 mm achromat), a beamsplitter cube (Thorlabs CM1-DCH) holding a dichroic beam splitter (Semrock FF416-Di01-25 × 36), an infinity corrected microscope objective lens (Nikon Plan Flour 10×, NA = 0.5 or Nikon 10×, NA = 0.25) connected with an CCD camera (Guppy 146B resolution: 1392×1040 or Marlin F131B IRF resolution 1280× 1024) via a transfer lens. To selectively increase the contrast in the images for crystals,another optical arm was added to include either the possibility of fluorescence microscopy or polarization microscopy.
For the fluorescence arm an aspheric condenser lens (Thorlabs A240TM-A, f_c= 8 mm, NA=0.5) was used for the UV LED (Nichia NCSU034A, 385 nm, 350 mW). Directly after the transfer lens (f_t = 30mm achromat) an excitation filter (387/11-25 Brightline single bandpass filter) is installed before the beam splitter (beam splitter (Semrock FF416-Di01-25 × 36)) with an edge wavelength of 416 nm. In addition, an emission filter (E 420nm Long Pass V2) is placed before the tube lens.
For the polarization microscope the UV LED is replaced with a red LED (Philips LumiLED, 630 nm, 350 mA) and the excitation and the emission filter with linear polarizers.

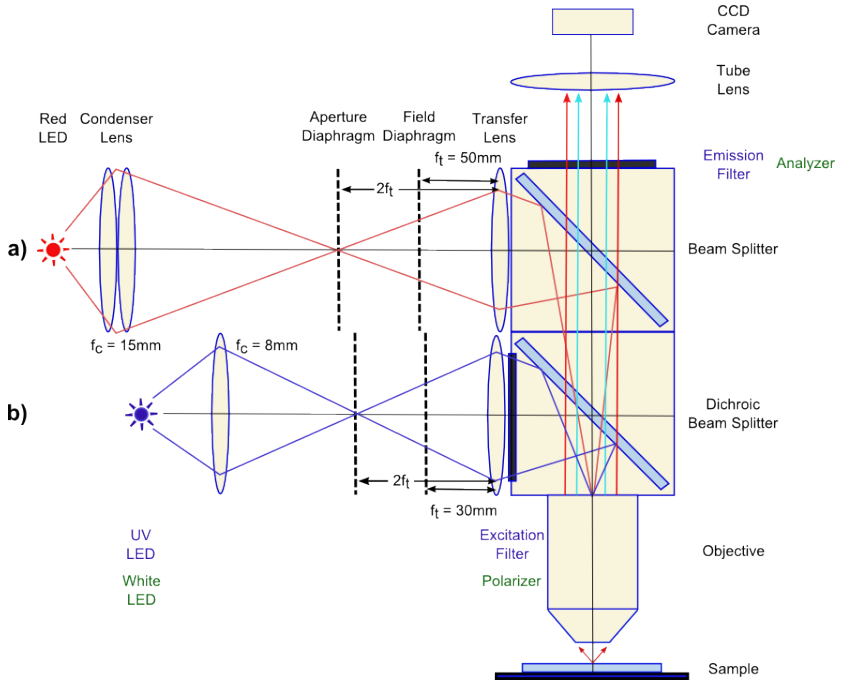


Figure 13: A schematic of the optical paths. Arm a) is for the common bright-field microscope and arm b) can be modified for either polarization microscopy (green) or fluorescence microscopy (purple).

Fluorescence microscopy The amino acid tryptophan in proteins absorbs ultraviolet light in the range of 260-320nm (peak at 280nm) and emits light in the range of 300 to 450nm. This native fluorescence could not be used because the absorption leads to damages of the protein and consequentially to an observable change in nucleation rate. Therefore, a fluorescence dye is needed. The non-covalent fluorescence dye 1-Anilino-8-Napthalene Sulfonic acid ammonium salt (1,8-ANS) (Sigma, BioChemika 10417-25G-F) has a high affinity to hydrophobic surfaces and is essentially non-fluorescent in water. Its fluorescence is only visible when attached to the hydrophobic parts of the protein or the drop membrane, resulting in a strong contrast, see figure 14. The peak absorption of 1,8-ANS is at 380nm, a wavelength where GI doesn't absorb any radiation, avoid radiation damage during excitation. Additionally, the dye shows no significant influence on the crystallization rates of proteins [15, 16]. The peak of the emission depends on its hydrophobic interaction with its environment and was, in our case, around 480nm.
The advantages of using this fluorescence dye setup are that no pre-preparations are needed to mark the proteins and only the crystal and drop surface can have activated 1,8 ANS around. This selectiveness reduces noise stemming from dirt. The biggest disadvantage is the small depth of focus where the fluorescence dye can be clearly resolved. Furthermore, having a different nucleation receipt, makes it more seems

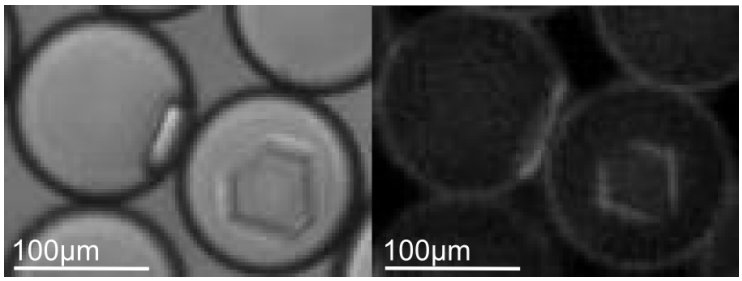


Figure 14: Left: light microscopy, right: fluorescence microscopy. The 1,8 ANS concentration in drops is $45\mu m$.

Polarization microscopyA commonly used contrast method for protein crystals is their birefringence. The observed Glucose Isomerase crystals are orthorhombic (space group I222) and accordingly biaxial [17]. In figure 15 it is clearly visible that some of the crystals have brighter centers with crossed polarizers, while others remain with a very similar intensity. For the crystals with no change in intensity, the optical path is parallel to one of the optical axes, or the birefringence in that direction is very weak.
This greatly reduces the advantage of polarization microscopy, the emphasize of crystals.

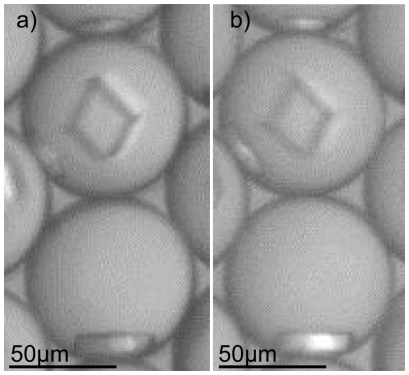


Figure 15: a) Polarizer and analyzer parallel, b) polarizer and analyzer crossed at 90° with shutter time 100× increased.

4 Image Processing

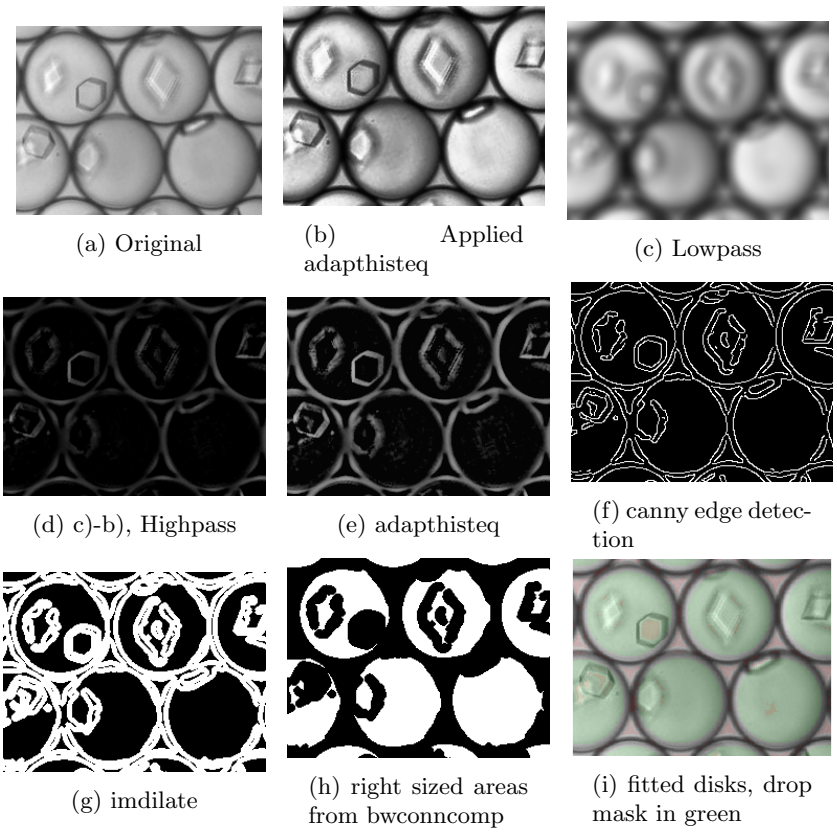
An automatized image analysis is needed to evaluate enough data for analysis based on statistics. In the following paragraph the mode of operation of the Matlab function, written for this purpose, is described. The goal is a fast and reliable analysis of the separated images to extract the total number of drops and the fraction of crystallized drops.
The biggest problems induced by the optics for image processing are heterogeneous illumination, uneven background, change of contrast over time and a change of focus over time. The heterogeneous illumination and the uneven background were minimized as much as possible through Köhler illumination and

the unfocusing of the metal underneath the capillaries, see 3.3. The change of contrast, especially the decrease, occurs due to condensation of water in the optical path.
One way to explain the change of focus over time is a mechanical relaxation processes of the clamped sample, which leads to a shift of the whole sample. Another possibility is a change of the thickness of the paraffin oil between the sample and the metal plate.

4.1 Drop Detection

The images are black-and-white and in the resolution of either 1280×1024 or 1392×1040. The drops in the capillary appear as black ringsordered in a two dimensional hexagonal lattice with several defects. One way to segment the image into drops, is to threshold on the approximate intensity value of the borders. Only darker pixel would be considered and an edge detection could define those borders. The problem of this method is to find the exact value for which the drop borders are as sharp and well defined as possible, even with a contrast change over time.
The approach which is chosen here, makes use of the local intensity gradient, in order to define the boundaries and its function can be seen in figure 16.

First of all, the contrast in the image is increased with Histogram Equalization, so that a narrow histogram of the intensity gets stretched to the full width of the intensity range. To consider an overall trend of intensity, the image is divided into 64 tiles and on each a histogram equalization is performed using Contrast-limited Adaptive Histogram equalization, which additionally limits the amplification of noise, figure 16a.
Each pixel is averaged over a disk with a 10 pixel radius, which is similar to a low pass, figure 16b. Subtracting the contrast enhanced image from the low pass filtered leaves only pixels which are darker than their surrounding unequal to zero. On this high pass, see figure 16d, adapthiseq is applied again to increase the contrast, figure 16e.
All these steps were taken to increase the contrast of the dark rings of the drops. Now an edge detection method which finds the maxima of the intensity gradients can be easily used. The edge-finding method 'canny' works with two thresholds, to detect strong and weak edges and only admits the weak edges if they are close to a strong edge. This leads to a detection of smooth edges with only few background noise, figure 16f.
To detect the area within the drop, it is necessary to have connected boundary. Small gaps in the edge image are overcome by a morphological dilation, in this case a disk with variable size, figure 16g. Connected areas in the negated binary image are sorted by size. The drop diameter itself is very homogeneous, so that an upper threshold can be chosen easily. The lower threshold, defining the minimum size, filters out the areas which are too small, e.g. the ones in between the drops. The crystals in the drops occupy some of the space in the drop and it is important to choose the minimum threshold in a way that all crystallized drops are not sorted out, figure 16h.
It is essential to identify the center and the radius of drops to ensure that the whole inside of the drop is searched for a crystal.



(a) Original

(b) Applied adapthisteq

(c) Lowpass

(d) c)-b), Highpass

(e) adapthisteq

(f) canny edge detection

(g) imdilate

(h) right sized areas from bwconncomp

(i) fitted disks, drop mask in green

Figure 16: Steps in drop detection.

In the case that there is no strong force which presses the drops together, the Laplace pressure inside the drops ensures that they appear circular. Therefore, the borders of those areas should be convex. With the option 'ConvexHull' of the function regionprops, the pixels forming the convex hull of the areas are identified. Those borders then are fitted to a circle and again compared to the known diameter of the drop to further sort out areas that are not a drop. This step filters out mainly the areas stemming from defects in the hexagonal lattice. A binary disk, identifying everything within as part of the drop, is created. The radius of the disk is chosen to be slightly smaller (≈90%) than the fitted one, to account for deviations of the drops from perfect circles and a shifted center of the drop. All disks are added to the same image, forming a binary drop mask, see figure 16i. The drop mask defines the region of interest in which crystals are searched for.

4.2 Crystal Detection

Crystals can be located randomly in the drop and are often positioned either flat in the middle of the drop or upright at the border of the drop. As a result the crystals are found in different focal planes.
Crystals in focus have clear, straight, black border. The crystals at the bottom of the drop only have a fuzzy, slightly darker border than the background, but a bright center.
This difference in contrast is the reason why simple putting a threshold for the intensity is not sufficient to identify all crystals.
The crystals at the border of the drop appears as a thin rectangle. To be able to detect those crystals the borders of the drop must be accurately determined.

After having a look at figure 16, it is clear that the method for finding the drops edges can resolve big crystals easily as well. The only difference is the value of the gradient threshold that should be chosen. For the drop detection it is mandatory to get smooth connected borders, while in the crystal detection resolving only crystals has more value. Therefore, the threshold for the edge detecting is chosen higher to filter out artifacts, e.g. edges from an uneven background.

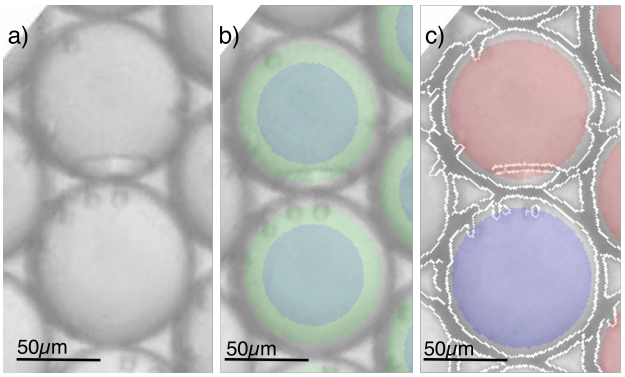


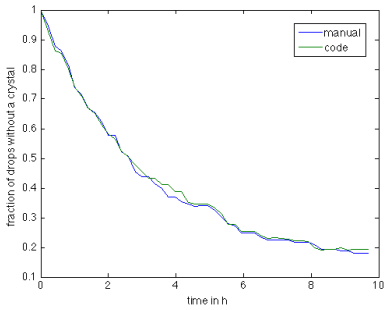
Figure 17: a) original, b) areas of interest c) red: crystal in drop; blue: no crystal in drop; white: detected edge

Another source of artifacts are very small droplets, called satellite droplets, that form in the dropmaker as a byproduct. They can move around in the capillary and if they are close to a drop, they look similar to crystals stuck on the wall, see figure 17a. These satellite droplets are smaller than a fully grown crystal, but more numerous. As a result, a filter is put in place to sort out small objects at the border of the detected drop. The drop is divided into two areas, see figure 17b. In the dark greenish center area the edge of even a very small object is interpreted as a crystal. Thus the crystal detection stays very sensible to small crystals.
In the outer, light green area of figure 17c, the found connected edges are sorted by size and only areas bigger than a predefined threshold are filtered out. The area threshold and the radius of the inner dark greenish area, are parameters that need to be chosen to optimize filtering out satellite drops.
With the current resolution, the smallest crystals that can be detected in an image with a high contrast are approximately $3\mu m$ in size. For the average image quality crystals get detected at a size of $\approx 5\mu m$.
In figure 18a the result of the image analysis for one position in a capillary with 131 drops is shown. The trend of the fraction of drops without a crystal is identical for the curves evaluated by the written Matlab code (c) and manually (m). The Euclidean distance between those curves is given by $d = \sqrt{\sum_i^N (c_i - m_i)^2}$, where N is the number of points in time at which the state of the drops is evaluated and c_i and m_i are the fraction values at time i. The Euclidean distance is 0.085, which is approximately 10% of the total change in f_ϕ.

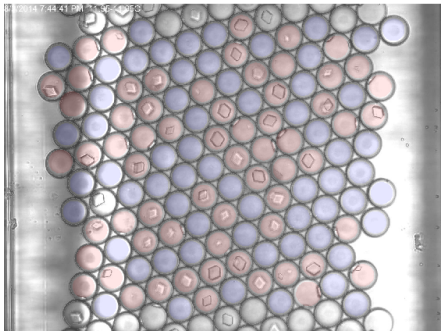
5 Results

5.1 Stochastically Independence of nearest neighbors

One of the key assumption is that the drops have identical conditions in terms of temperature, supersaturation and precipitant concentrations. The nucleation rate should not change over the course of the experiment so that the drop should be ideally isolated and therefore independent. To be independent the joint probability of set A and set B must equal the product of their probabilities.



(a) The decay of the population of uncrystallized drops over time evaluated by hand(blue) and with the code(green).



(b) Code evaluated image of drop emulsion in capillary.
Red: drop contains a crystal; Blue: drop contains no crystal.

Therefore the conditional probabilities are equal to their their own probability. Any big enough subset of drops should consequentially lead to the same decay rate as all drops together.
Too test the independence, subsets of nearest neighbors of crystallized drops are taken. In figure 19 it is clearly visible, that there is no dependence on the nucleation probability if there is a crystallized drop or not. The second and third nearest neighbors seem to have the same slope as the average of all drops, ups , but they are shifted slightly

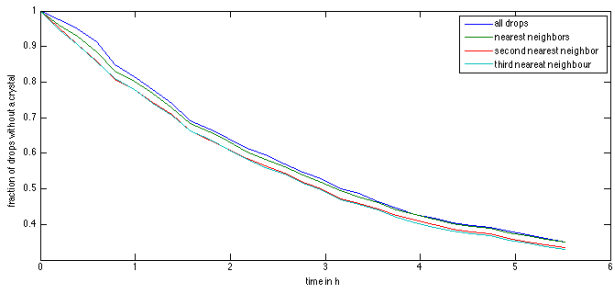


Figure 19: Rates of 4300 drops divided into subset depending on their distance to a crystallized drop.

5.2 Long range effects

Besides the direct neighbor interactions between the droplets long reaching gradients can influence the nucleation rate considerable. Besides temperature gradients there can be chemical gradients influencing nucleation. The sealing of the capillary is most important to avert chemical gradients within a capillary.
Three different sealing mechanism were tested.
The wax VALAP was melted and one drop was put onto the cold capillary and glass slide, solidifying and sealing the capillary rapidly. This is the most commonly used quick sealing method for capillaries.
Furthermore sealing was tested with 5^M Quik-cure epoxy (Bob Smith Indus-

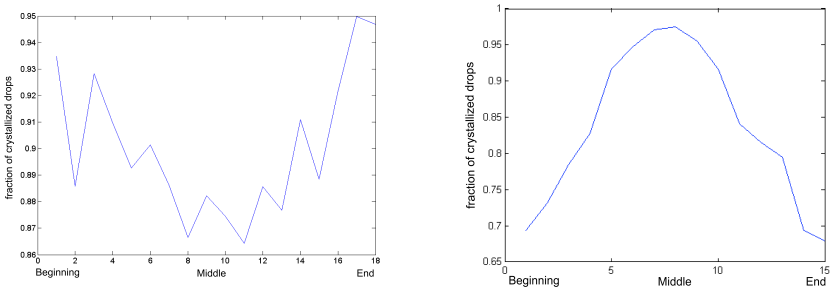


Figure 20: The crystal fraction on the last measured point plotted versus the position in the capillary. Left: wax; Right: epoxy.

tries) and Norland Optical Adhesive 81, of which both work via cross-linking. While it is enough for epoxy to mix two components that will cure, the optical glue needs an optical excitation to be able to cross-link. In both cases the capillary end is longer exposed to fluid chemicals as if wax is used.
To estimate if there is a connection between the nucleation rate and the position in the capillary, the last run was taken and and the crystal fraction was plotted versus the position in the capillary, see figure 20.

For epoxy a clear trend is visible. In the center the nucleation is the fastest while at the ends of the capillary the fraction of crystallized drops is more than 30% lower. This strong influence was as well visible for the optical glue. it is obvious that the ends of the capillary have a way lower nucleation rate than the ones in the middle of the capillary. Both sealing methods relaying on chemical cross-linking influence the interior of the capillary greatly. Additional the optical glue was cured under a UV lamp (302nm), which can denature the protein. Careful protection of the capillary was necessary to use the optical glue.
In the case of wax, the fraction of crystallized drops varies around 15%, but seems to show the opposite trend, having a slightly higher nucleation rate at the end of the capillary. The VALAP wax was chosen to seal the capillary, because of its smaller impact on nucleation.
Another sensible step is the loading of the capillary with the emulsion. The time needed to load one capillary is mainly determined by the capillary force, which drive the process. Loading the capillaries takes around 30min in which the emulsion is exposed to the room air. An exchange of chemicals, like water and Ammonia, can take place rapidly because of the high surface to volume ratio. A dramatic change in the nucleation rate between drops with different exposure times can be seen in figure 21.

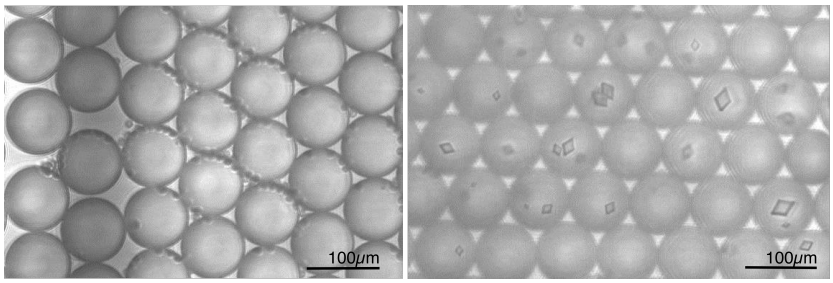


Figure 21: All drops were quenched at the same time and both pictures are taken within 2min from each other. The only difference the drops in the right capillary were exposed 30min longer to the air than the ones on the left.

Measurements of the drop size indicate that no measurable change in drop volume took place. A small change of the volume and therefore supersaturation should not lead to such a strong difference between the nucleation rates.
The isoelectric point of Glucose Isomerase is approximately at a pH of 3, [4]. The closer the pH is to the isoelectric point the lesser is the net charge on the protein and consequentially the repulsive interaction. Hence, a change in pH can influence the nucleation rate strongly. There are two speculations how this could explain the observed change in nucleation rate.
First CO_2 can dissolve in water and form the weak acid, carbonic acid, thus decreasing the pH. The TRIS-HCl buffer might not be able to fulfill its buffering function, because it is already one unit away from its pK ≈ 8.07.
The other process considers the volatile Ammonia from the Ammonium sulfate, leaving the drop. Ammonia is moderately basic, which increases the pH. When Ammonia is leaving the drop, the drop becomes more acidic, lowering the difference between the pH and the isoelectric point of Glucose Isomerase.
So far no conclusive evidence was found, what exactly causes the change in the nucleation rate.
Enclosing the emulsion with a Parafilm while loading the capillaries, decreased the difference in nucleation drastically. Future experiments are planned, that will show whether the change is due to Ammonia leaving the drop or not.

5.3 Qualitative Discussion

Unfortunately was it not possible to repeat experiments under the same conditions. Like in section 5.2 mentioned were several trends throughout the hole capillary visible. These changes make it impossible to just merge all the made measurements. But trends can be found out.
The Glucose Isomerase measurements resulted as well as the lysozyme [1, 2] in two decay rates shown by the inverse Laplace transform, see figure 22.

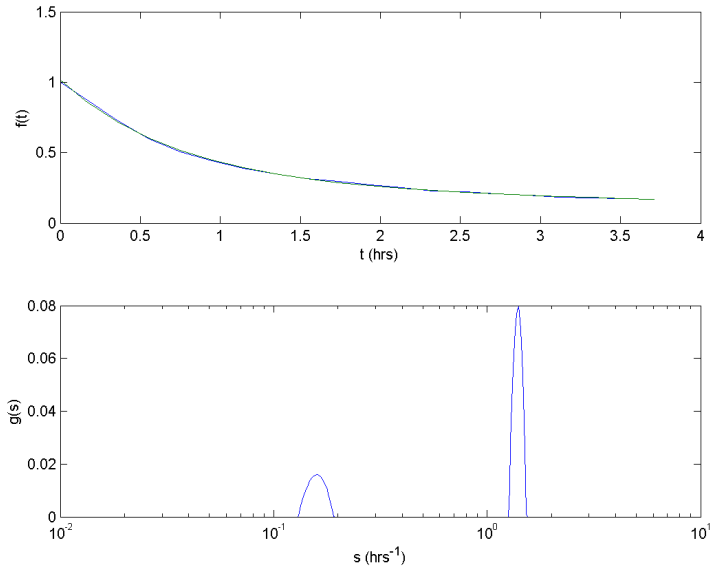


Figure 22: Inverse Laplace Tranformation.

In 23 a precipitant solution older than two weeks was used. There is no long reaching effect visible. This would support the theory that the volatile Ammonia is the culprit, because the loading time and did not change, but there is no effect visible anymore on how long the emulsion is exposed to air

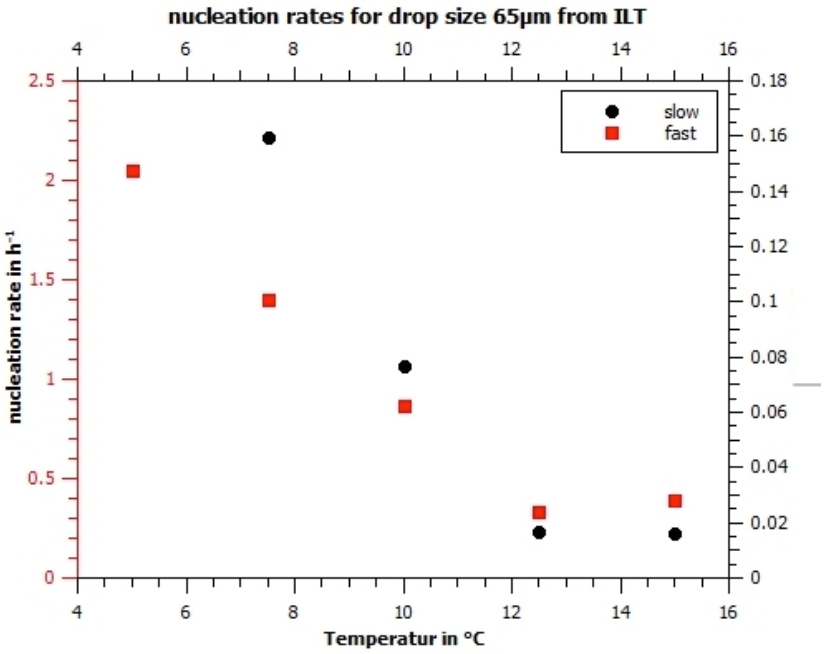


Figure 23: Decay rates from protein with old ammonium precipitant solution.

6 Conclusion

The optics of the system were successfully adjusted so that the contrast is strong enough to use the automated image analysis. The Matlab program works well, but it introduces more errors the dirtier the sample is. Depending on the image quality a lot more drops than 500 would be necessary to be within a 10% confident bound of the actual nucleation rate.
The measurements of the nucleation rate for Glucose Isomerase turned out to be more challenging than expected because of chemical gradients. A measurement looked at individually resembles the trend that was visible in Future steps involve clarifying if the ammonia sulfate is too volatile to use as a precipitant. The solubility and the second virial coefficient of the protein in the precipitant solution are measurements necessary to give the nuclation rate measurement more meaning.

References

[1] Sathish V Akella, Aaron Mowitz, Michael Heymann, and Seth Fraden. An emulsion based technique to measure protein crystal nucleation rates of lysozyme. 1(781), 2014.

[2] Sathish V. Akella. *Characterizing Protein Crystal Nucleation*. Dissertation, Brandeis University, 2014.

[3] Victor K La Mer and Guy M Pound. Kinetics of Crystalline Nucleus Formation i n Supercooled Liquid Tin1l2. *Journal of the american chemical society*, 74(10):2323—-2332, 1952.

[4] Hampton Research. Glucose Isomerase - Data sheet.

[5] J. Gunton, A. Shiryayev, and D. Pagan. *Protein Condensation*. Cambridge University Press, New York, first edit edition, 2007.

[6] J. Bloustine, T. Virmani, G. Thurston, and S. Fraden. Light Scattering and Phase Behavior of Lysozyme-Poly(Ethylene Glycol) Mixtures. *Physical Review Letters*, 96(8):087803, March 2006.

[7] Sigma-Aldrich. Trizma® hydrochloride, 2014.

[8] C Holtze, a C Rowat, J J Agresti, J B Hutchison, F E Angilè, C H J Schmitz, S Köster, H Duan, K J Humphry, R a Scanga, J S Johnson, D Pisignano, and D a Weitz. Biocompatible surfactants for water-in-fluorocarbon emulsions. *Lab on a chip*, 8(10):1632–9, October 2008.

[9] L Spencer Roach, Helen Song, and Rustem F Ismagilov. Controlling Nonspecific Protein Adsorption in a Plug-Based Microfluidic System by Controlling Interfacial Chemistry Using Fluorous-Phase Surfactants important for using microfluidic devices for biochemical. 77(3):785–796, 2005.

[10] Shelley L. Anna, Nathalie Bontoux, and Howard a. Stone. Formation of dispersions using "flow focusing" in microchannels. *Applied Physics Letters*, 82(3):364, 2003.

[11] Fluke. 51 & 52 Series II - Thermometer, 2001.

[12] David Lide. *CRC Handbook of Chemistry and Physics*. CRC Press, 86th edition, 2006.

[13] Sindy K Y Tang, Zhenyu Li, Adam R Abate, Jeremy J Agresti, David a Weitz, Demetri Psaltis, and George M Whitesides. A multi-color fast-switching microfluidic droplet dye laser. *Lab on a chip*, 9(19):2767–71, October 2009.

[14] H Modarress and H Rasa. Measurement and Modeling of Density, Kinematic Viscosity, and Refractive Index for Poly(ethylene Glycol) Aqueous Solution at Different Temperatures. pages 1662–1666, 2005.

[15] Matthew R Groves, Ingrid B Müller, Xandra Kreplin, and Jochen Müller-Dieckmann. A method for the general identification of protein crystals in crystallization experiments using a noncovalent fluorescent dye. *Acta crystallographica. Section D, Biological crystallography*, 63(Pt 4):526–35, April 2007.

[16] Andrea Hawe, Marc Sutter, and Wim Jiskoot. Extrinsic fluorescent dyes as tools for protein characterization. *Pharmaceutical research*, 25(7):1487–99, July 2008.

[17] a Echalier, R L Glazer, V Fülöp, and M a Geday. Assessing crystallization droplets using birefringence. *Acta crystallographica. Section D, Biological crystallography*, 60(Pt 4):696–702, April 2004.

A Matlab code for image processing

```
function [XF,ND,varargout] = xtal_HighPass2(ImgPath,lowerDiameterLimit, ...
    upperDiameterLimit,CircleSizeAdapt,DropFilterSize,EdgeFilterSize,varargin)
% This function opens an Image located at 'ImgPath' and detects circle shaped
% Areas, which have an Diameter between 'lowerDiameterLimit' and
% 'upperDiameterLimit'. These Areas are fitted to circles whose size is
% adapted by CircleSizeAdapt and form then a mask of circles.
% Edges found within this mask are considered as crystals.
% xtal_HighPass2 returns the Number of detected Drops 'ND' and
% the Fraction of Drops that contain crystals 'XF'.
% For the necessary edge detection a HighPass is applied, which shows
% Pixels which are darker than their surrounding Pixels in average.
% This HighPass is applied in two different magnitudes.
% For the Drop detection it is good to apply a weak HighPass, which allows
% most of the borders to be seen.
% Therefore the circles are complete and realatively easy to fit.
% For Edge Detection a stronger HighPass is used, so that noise or artefacts
% disturb the crystall detection less.

% various options are available:
% With the option 'show' the original image, the detected areas,
% the Dropmask, the Crystalmask, the crystal edge mask and the dropedge
% mask are shown.
% With the option 'savemasks' the Dropmask and the Crystalmask are shown
% and saved. An input String after 'savemasks' creates a new subfolder
% where the images are saved.
% With the option 'pinhole' a circular shaped mask can be overlayd the
% original image. This can be advantagous when the illumination is not
% homogenious.
% With the option 'connectivity' the connectivity of bwconncomp
% can be changed to 4 the default is 8.
% With the option 'dilation' the size of the neighbourhood for
% imdilate can be changed. The default is 1 and only Integers are accepted
% (imdilate(IMG,strel('disk',dilation))
%
% example:
% [XF,ND] = xtal_HighPass2('C:\Users\sathish\Dropbox\Images\...
% Test for time extraction\Run001-R00-C00-00.jpg'...
% ,30,65,0.85,10,10,'pinhole',0.8,'connectivity'...
% ,8,'dilation',1,'savemasks','Example','show')

original = imread(ImgPath);
```

```
if size(original,3)==3
    original = rgb2gray(original);
end

[ShowFlag,Pinhole,connectivity,dilation,SaveMasks,Shift,...
    showDrops,showCrystals,edgeInput,edgeInputCryst,BorderFlag...
    ,positionsflag,MaxNumberOfDropsInImage,minPixel] ...
    = parse_input(ImgPath,nargin-6,varargin{:});
% Image overlay for adjustment of focus change
if Shift(1)
    if Shift(2)>=0
        imgShift =imread([ImgPath(1:end-5) num2str(ceil(Shift(2))) '.jpg']);
    else
        imgShift =imread([ImgPath(1:end-5) num2str(floor(Shift(2))) '.jpg']);
    end
    original = uint8((1-abs(Shift(2))) .* single(original) ...
                + abs(Shift(2) .* single(imgShift)));
end

% additional restriction of Image for bad Illumination at border.
if Pinhole(1)
    s = size(original);
    original = uint8(drawDisk(zeros(s),[s(2),s(1)].*0.5,...
                0.5*Pinhole(2)*s(2))) .* original ;
end

%% Creating EdgeImage for Drop detection
H = fspecial('disk', DropFilterSize);    % defining a Filter
original2 = adapthisteq(original);       % enhancing contrast in Image
LowPass = imfilter(original2, H);        % applying a LowPass/ taking the mean
HighPass= imsubtract(LowPass,original2);% Only darker borders come through
HighPass = adapthisteq(HighPass);        % enhance contrast

EdgeImage = edge(HighPass,'canny',edgeInput(1),edgeInput(2));% detect edges

% overcome small gaps in detected lines
negativeEdgeImage = imdilate(EdgeImage, strel('disk',dilation));
% creating Binary negative Image for further processing
negativeEdgeImage = not(negativeEdgeImage);

%% creating Xtal_EdgeImage for Crystal Detection
H = fspecial('disk', EdgeFilterSize);
LowPass = imfilter(original, H);
HighPass = imsubtract(LowPass,original);
HighPass = adapthisteq(HighPass);
Xtal_EdgeImage = edge(HighPass,'canny',edgeInputCryst(1),edgeInputCryst(2));

%% Filtering detected Regions for drops
cc = bwconncomp(negativeEdgeImage,connectivity);% find connected components
LM = labelmatrix(cc);
% create necessary data for each region
stats = regionprops(LM, 'MajorAxisLength', 'MinorAxisLength','Area');
% which Regions have the right size
idx = find( (lowerDiameterLimit <=[stats.MinorAxisLength])...
            & ([stats.MajorAxisLength] <= upperDiameterLimit) );

BW = ismember(LM, idx); % Get the fitting regions into one Image
tic

cc = bwconncomp(BW);
props = regionprops(cc,'ConvexHull'); %get Coordinates of the border points
```

```
%% Each object is fitted with a cirlce and searched for crystals
ND_temp = cc.NumObjects;      % Number of Components is the number of drops
ND = 0;
NX = 0;                       % Number of Crystals
mask = zeros(size(BW));
XM   = zeros(size(BW));
CircleFitFail = zeros(size(BW));

if BorderFlag
    CapillaryBorder = getCapillaryBorder(original,upperDiameterLimit);
else
    CapillaryBorder = false(size(original));
end

dropPositions = zeros(ND_temp,2);
xtalPositions = zeros(ND_temp,2);

for j = 1:ND_temp
    % checking for minimum Number of points for accurate circle fit
    if ( size(props(j).ConvexHull, 1) <= 24 )
        continue
    end
    x = props(j).ConvexHull(:,1);
    y = props(j).ConvexHull(:,2);
    k = convhull(x,y);
    XY = [x(k) y(k)];
    [centerCircle, radiusCircle] = fitCircle(XY);%fit with a circle
    if ( (2*radiusCircle < lowerDiameterLimit) ||... % fit succesful?
            (2*radiusCircle > upperDiameterLimit)||...
            (2*radiusCircle < 0.5*upperDiameterLimit)...
            || centerCircle(1) > size(BW,2) - 0.7*upperDiameterLimit ||...
           centerCircle(1) < 0.7*upperDiameterLimit ...
            || centerCircle(2) > size(BW,1) - 0.7*upperDiameterLimit ||...
           centerCircle(2) < 0.7*upperDiameterLimit ...
            || CapillaryBorder(fix(centerCircle(2)),fix(centerCircle(1))) )
        % adjusting the radius of the circle
        radiusCircle = radiusCircle*CircleSizeAdapt;
         % make a Binary mask with circle
        tempMask = drawDisk(mask,centerCircle,radiusCircle);
        CircleFitFail = CircleFitFail | tempMask;
        continue
    end
    ND = ND +1;
    dropPositions(ND,:) = centerCircle;
     % adjusting the radius of the circle
    radiusCircle = radiusCircle*CircleSizeAdapt;
    % make a Binary mask with circle
    tempMask = drawDisk(mask,centerCircle,radiusCircle);
    mask = mask | tempMask;                  % updating drop mask

    xtal = tempMask & Xtal_EdgeImage;        % edges on current ROI
    xtal = imdilate(xtal,strel('disk',1));
    xtalcc = bwconncomp(xtal);
    xtalProps = regionprops(xtalcc,'Area','Centroid');
    centroids = cat(1, xtalProps.Centroid);
    if not(isempty(centroids))
        difference = centroids-ones(size(centroids))*diag(centerCircle);
        % filter bei position and size of connected components
        idx = find(((norm(difference(:,1:2),2)>= 0.7 * radiusCircle)&...
            (minPixel <=[xtalProps.Area]))| ...
            ((norm(difference(:,1:2),2)< 0.7 * radiusCircle)&...
            (minPixel/10 <=[xtalProps.Area])) );
```

```
        if not(isempty(idx))     % is there a crystal
            NX = NX +1;          % updating Crystal count
            XM = XM | tempMask   % updating mask of crystalized drops
            xtalPositions(NX,:) = centerCircle;
        end
    end

end

if ND > 0
    XF = NX/ND;
else
    XF = 0;
end
```