Prebiotic DNA replication and Analyzing a genetic disorder with thermophoresis

Evgeniia Edeleva

Dissertation

München, Oktober 2018

Prebiotic DNA replication and

Analyzing a genetic disorder with thermophoresis

Evgeniia Edeleva

Dissertation
zur Erlangung des Grades
Doktor der Naturwissenschaft (Dr. rer. nat.)

an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Evgeniia Edeleva
aus Yuzhno-Sakhalinsk, Russland

München, den 11.10.2018

Zusammenfassung

Die Arbeit, die in dieser Dissertation vorgestellt wird, besteht aus drei unabhängige Teilen. Die Teile I und III sind durch das Thema des Ursprungs des Lebens verbunden. Die Teile II und III basieren auf der gleichen biophysikalischen Analysemethode der Mikrothermophorese (MST).

Teil I konzentriert sich auf die präbiotische Replikation von DNA. Verschiedene Ansätze zur Replikation von RNA und DNA unter präbiotischen Bedingungen werden in der Einführung diskutiert. Ein nonenzymatisches DNA-Replikationssystem wird als einer der möglichen Ansätze für einen präbiotischen Replikator vorgestellt. Zwei DNA-Stränge mit 24 Basen kreuz-katalysieren die gegenseitige Synthese aus vier 12 Basen langen Substratsträngen, von denen zwei *in situ* durch ein Kondensationsmittel 1-ethyl-3-(3-dimethylamino-propyl)carbodiimid (EDC) aktiviert werden. Eine (2.2 ± 0.5)-fache Amplifikation wird in 6 Stunden erreicht. Eine stärkere Amplifikation wird durch EDC-induzierte Modifikation von Substratsträngen unterdrückt. Ein detailliertes kinetisches Model, das die Modifikationskinetik berücksichtigt, wurde mit einem System von gewöhnlichen Differentialgleichungen realisiert und erlaubte es, kritische Parameter der Reaktion zu verstehen. Weitere Forschungsarbeiten sollten sich darauf konzentrieren, Wege zu finden, wie das Substrat wiederverwendet werden kann oder wie seine Modifikation reduziert werden kann. Dies könnte zu DNA-Replikations- und Verlängerungsexperimenten führen, die von kurzen zufälligen DNA-Strängen ausgehen. Solche Experimente könnten mit next-generation Sequenzierung analysiert werden, da die nativen Phosphodiesterbindungen an den Bindungsstellen entstehen.

In Teil II wird MST verwendet, um einen diagnostischen Assay für α-Antitriypsin (AAT)-Mangel zu entwickeln, der direkt im Blutplasma funktioniert. Bei AAT-Mangel haben Einzelpersonen eine reduzierte Plasmakonzentration von AAT, was sie unter anderem anfällig für Lungenemphyseme macht. Überraschenderweise korreliert die Konzentration von AAT nicht mit dem Ausmass der Symptome. Bei einem Standard-MST-Ansatz ist die Amplitude der erhaltenen Bindungskurve für die Affinitätsanalyse nicht wichtig. In einem in dieser Arbeit vorgestellten MST-basierten kompetitiven Assay hängt die Amplitude der erhaltenen Bindungskurve sowohl von der Affinität der AAT-Bindung an ihr Target als auch von der Konzentration von AAT ab. In einem Proof-of-Principle-Experiment wurde das Blutplasma von Individuen mit AAT-Mangel zur Quantifizierung der Amplitude verwendet. Die Amplitude zeigte eine bessere Korrelation mit dem Ausmaß des Krankheitszustandes als die Konzentration von AAT allein, was darauf hindeutet, dass die Affinität von AAT zu seinem Target bei verschiedenen Individuen unterschiedlich ist. Die direkte Messung der AAT-Affinität zu ihrem Target im Plasma von Individuen mit starker *vs.* schwacher Manifestation der Symptome zeigte, dass die Affinität bei Individuen, die nur leicht betroffen waren, höher war. Dies deutet darauf hin, dass die Plasmazusammensetzung einen Einfluss auf die AAT-Affinität hat.

In Teil III werden experimentelle Systeme entwickelt, um die stereochemische Theorie der Herkunft des genetischen Codes zu testen. Der genetische Code definiert die Zuordnung von 20 Aminosäuren zu ihren Codons. Die Hypothesen darüber, was die Zuordnung bestimmt hat, wurden in der Liter-

atur ausführlich diskutiert, aber die Frage bleibt eine der experimentell anspruchsvollsten im Bereich des Ursprungs des Lebens. Die stereochemische Theorie besagt, dass Aminosäuren eine intrinsische Bindungsaffinität zu ihren Codons haben. Dies kann durch direkte Messung der Wechselwirkung zwischen den aktivierten Aminosäuren und codonhaltigen RNA-Strukturen getestet werden. Auf der Grundlage der Literaturrecherche wurden RNA-Strukturen entwickelt und die Affinität mit der MST-Methode gemessen. Proof-of-Principle-Experimente für drei RNA-Strukturen wurden durchgeführt, und ein Ansatz für die zukünftige Arbeit wird am Ende detailliert diskutiert.

Abstract

The work presented in this Thesis is divided into three independent parts. Parts I and III are connected by the topic of the Origin of Life. Parts II and III are based on the same biophysical analysis method of microscale thermophoresis (MST).

Part I focuses on prebiotic replication of DNA. Different approaches to replicate RNA and DNA under prebiotic conditions are discussed in the introduction. A nonenzymatic DNA replication system is presented as an approaches to a prebiotic replicator. Two DNA strands of 24 bases cross catalyze each other's synthesis from four 12 bases long substrate strands, two of which are *in situ* activated by a condensing agent 1-ethyl-3-(3-dimethylamino-propyl)carbodiimide (EDC). DNA amplified (2.2 ± 0.5)-fold in 6 hours. Stronger amplification is suppressed by EDC-dependent modification of substrate strands. Detailed kinetic modeling was realized with the system of ordinary differential equations and allowed to understand critical parameters of the reaction. Further research should be focused on finding ways to recycle substrate or to reduce its modification. This could lead to DNA replication and elongation experiments starting from short random DNA strands. Such experiments could be analyzed with next-generation sequencing due to the formation of the natural phosphodiester bonds at the ligation sites.

In part II, MST is used to develop a diagnostic assay for α1-antitriypsin (AAT) deficiency disorder working directly in blood plasma. In AAT deficiency, individuals have reduced plasma concentration of AAT, which makes them susceptible to lung emphysema among other symptoms. Surprisingly, concentration of AAT does not correlate with the severity of the symptoms manifestation. In a standard MST approach, the amplitude of the obtained binding curve is not important for the affinity analysis. In an MST-based competition assay introduced in this work, the amplitude of the obtained binding curve is sensitive to both the affinity of AAT binding to its target and the concentration of AAT in plasma. In a proof-of-principle experiment, blood plasma of individuals with AAT deficiency was used to quantify the amplitude. The amplitude showed a better correlation to the severity of the disease status than the concentration of AAT alone, suggesting that AAT affinity to its target is different in different individuals. Direct measurement of AAT affinity to its target in plasma of individuals with strong *vs* weak manifestation of symptoms showed that affinity was higher in individuals who were only mildly affected. This suggests that plasma composition has an effect on AAT affinity.

In part III, experimental systems are developed in order to test the stereochemical theory of the origin of the genetic code. The genetic code defines the assignment of 20 amino acids to their codons. The hypotheses to what determined the assignment have been widely discussed in literature but the question remains one of the experimentally most challenging in the origin of life field. The stereochemical theory states that amino acids have an intrinsic binding affinity towards their cognate codons. This can be tested by directly measuring the interaction between the activated amino acids and RNA structures containing codons. RNA structures were developed based on literature research, and the affinity was measured with the MST method. Proof-of-principle experiments for three RNA structures were performed, and the outline for future work is detailed in the end.

Contents

List of Figures

List of Tables

Part I

Nonenzymatic cross-replication of DNA strands

1 Motivation and Introduction

According to a current NASA definition of life, life is "a self-sustaining chemical system capable of Darwinian evolution". To be able to undergo Darwinian evolution, a living system has to be able to reproduce, and reproduction is based on the ability of the system to encode and transfer information through generations. The replication of information-carrying polymers (DNA and RNA) is central to this process. In modern organisms, the replication of DNA depends on complex proteins called enzymes. However, enzymes themselves are encoded in the DNA. This poses a well-known chicken and the egg problem of the origin of life: what came first, complex proteins or DNA? It seems unlikely, however, that enzymes were present early in evolution. Therefore, research has been focused on finding enzyme-independent pathways to replicate DNA-like polymers. Of high importance is to find pathways simple enough to be plausible on early Earth.

Base-by-base replication by a ribozyme   One of the widely popular hypothesis, coined the "RNA world" hypothesis, suggests that life initially was RNA based. RNA molecules on the prebiotic Earth, apart from storing information, could have catalyzed their own replication [6, 24, 61, 98]. And indeed RNA molecules, called ribozymes, are known to be able to catalyze reactions. To prove that RNA can catalyze its own replication, researchers have been developing polymerase ribozymes that mimic polymerase enzymes in that they could catalyze RNA replication in a base-by-base fashion.

The best base-by-base polymerase ribozyme developed to date can synthesize a variety of functional RNAs, including aptamers and other ribozymes [31]. Although, the polymerase ribozyme was shown to extend strands through difficult sequence features and to copy strands of 35-40 bases long to full length and in good yield, synthesis of longer sequences remains challenging. The longest structure synthesized by this polymerase was a tRNA molecule of 76 bases, however, in a very low yield. The synthesis of the 180 bases long polymerase ribozyme itself remains a non-resolved task.

Apart from a ribozyme-catalyzed replication by polymerization of monomers, ribozyme-catalyzed replication by successive polymerization of short oligonucleotides was also studied. Just recently, an *in vitro* evolved ribozyme was shown to successfully replicate structured RNA motifs starting from trimer oligonucleotide substrates instead of monomers [1]. Although 240-times slower than the most advanced base-by-base polymerase ribozyme discussed above, the triplet polymerase achieved a comparable fidelity and allowed to copy through difficult previously unaccessible regions of structured RNA. Replication of the triplet polymerase ribozyme itself was only achieved by segmenting the sequence into shorter sequences and resulted in low efficiencies of both synthesis and final assembly.

A base-by-base polymerase ribozyme able to copy a sequence longer than itself (206 nt sequence by a 202 nt-long ribozyme) was reported [2], however, it posed strong sequence limitations on the primer, from which it was able to start the polymerization reaction. It preferred C-rich sequences, absence of secondary structures, and reduced G-content, among others. Such limitations meant that copying of both "-" and "+" strands of its own sequence was impossible, and, thus, an exponential amplification of the polymerase ribozyme required for Darwinian evolution experiments was also impossible.

The work on triplet polymerase ribozyme stressed the importance of considering different substrates for replication than just monomers. Interestingly, it was discussed that conditions promoting rather than disfavoring base-pairing led to the ability of the ribozyme to copy through difficult structural motifs. The utilization of triplets resulted in lower entropic costs per binding position and cooperative invading effects that helped in unwinding the structured RNA motifs. Extra stability also came from the intra-triplet base stacking interactions. Low temperature conditions and high solute concentrations from working in eutectic ice phase aided in increasing the base pairing stability. Replication by polymerization of short oligonucleotides should be considered alongside the canonical base-by-base replication.

Ribozyme that achieved exponential replication-by-ligation   None of the above systems was capable of exponential amplification. An exponential amplification was achieved by a different kind of a ribozyme, a ligase that catalyzed replication through ligation of two substrate strands [47]. The product of ligation was another ligase ribozyme that could in turn ligate two different substrate strands to produce the initial ligase ribozyme. Thus, the system was replicating in a cross-catalytic fashion and could amplify exponentially. An exponential amplification behavior allowed the authors to perform selection experiments starting from the population of slightly mutated ligases. However, the dependence of the system on the preformed long substrate strands of a specific composition has a major drawback: evolution from random pools of short oligonucleotides as well as open-ended evolution are not allowed in this ligase-based system.

Nonenzymatic replication   Although it is an impressive achievement of science that ribozyme polymerases are routinely used in labs today, the fact that none of them can reproduce their own sequences in good yield to allow Darwinian evolution experiments suggests that less complex (though not less difficult to realize in lab) nonenzymatic mechanisms could have been at work early in evolution. These mechanisms should have been able to replicate and elongate short random oligonucleotides present on prebiotic Earth. To this end, ribozyme-free approaches to replication were also considered in literature on prebiotic DNA/RNA replication.

One approach is to use activated substrate monomers or oligomers that, upon base pairing with the template, can ligate to the primer or to each other releasing a leaving group that was storing the activation energy. For example, monomers pre-activated with a 2'-methylimidazol leaving group were shown to polymerize in a base-by-base fashion on an RNA template starting from a primer, when supported by activated trimers binding downstream of the reaction site [66]. Since trimers were activated with the same leaving group as monomers, it is not surprising that also ligation of these helper oligonucleotides was observed in the same reaction pot. Thus, here again, we see an advantage of using different substrates for replication: monomers and trimers.

In all of the above base-by-base replication experiments, the primary goal was to achieve polymerization leading to replication of the template. However, for Darwinian evolution experiments, replication of the template should be followed by replication of the product to gain a new molecule of the initial

template, similar to a PCR reaction. This - over many cycles - would lead to an exponential amplification of the initial template and could drive sequence evolution through random mutations.

A replication scheme, where initial template catalyzes the synthesis of a molecule of its complement, and then the complement catalyzes the synthesis of the new molecule of the initial template, is called a cross-catalytic scheme. On the other hand, if the initial template is self-complementary, its complement will be an initial template itself. Such a replication scheme is called an autocatalytic scheme.

Replication of a short self-complementary DNA strand by ligation of two shorter complementary strands was studied by the Kiedrowski group. They used chemical ligation catalyzed by a 1-ethyl-3-(3-dimethylamino-propyl)carbodiimide (EDC) under isothermal conditions [93]. In the reaction presented by the Kiedrowski group in 1986, two DNA trimers condensed in a template-directed fashion to form a hexamer oligonucleotide that served as a new template molecule. In the following work by the same group, a substitution of a 5′-hydroxyl group for a 5′-amino group at the ligation site significantly increased the efficiency of the ligation reaction [94]. However, this strategy led to a formation of a phosphoramidate bond instead of the naturally occurring phosphodiester bond at the ligation site and, thus, resulted in a product with heterogenous bonds. Independent of the type of bond formed at the ligation site, the product concentration saturated, presumably as a result of product inhibition.

The problem of product inhibition In product inhibition, self-complementary product strands form stable duplex complexes, thereby inhibiting the next reaction. To overcome product inhibition and to drive exponential amplification of longer strands, temperature cycling can be introduced to melt product duplexes and allow their reuse as templates. Conditions of temperature cycling on early Earth could have been present e.g. in the form of temperature gradients across porous rocks of hydrothermal vents deep in the oceans or hydrothermal fields with geysers.

Carbodiimides Carbodiimides, such as EDC, have been optimized for efficient condensation of carboxylic acids with nucleophilic amines. While they can also be used for the condensation of weaker nucleophiles, like carboxylic acids or pyrophosphates [43], these reactions suffer from an unwanted side-reaction as a consequence of the relatively low nucleophilicity. In the side reaction, the O-acylisourea intermediate rearranges into the stable non-reactive N-acylurea [80]. Therefore, side products are also expected in EDC-driven condensation of oligonucleotides. However, the high speed of ligation could compensate for the amount of modifications.

Although a prebiotic pathway for EDC synthesis has not been suggested, it shares its reactive group with a tautomer of cyanamide – a molecule presumably formed under prebiotic conditions [77, 90].

Project goal The goal of this project was to establish and characterize an exponential cross-catalytic DNA replication system. For this I aimed to:

- design a cross-catalytic DNA replication system with the following characteristics:

  – replication should be nonenzymatic, template-driven, occur by ligation of shorter comple-
    mentary oligonucleotide fragments, and lead to the formation of a product with the native
    backbone;

- employ temperature cycling in order to overcome product inhibition;

- establish protocols for serial dilution (in order to mimic environmental selection pressure) and
  feeding (in order to provide the reaction with fresh fragments for ligation);

- realize a detailed kinetic model of the system by solving a system of ordinary differential equations
  (ODEs) and use it to understand and fit experimental data.

2 Experimental Procedures and Data Analysis

2.1 Materials

Buffers

Control buffer: 100 mM MOPS at pH 6.5, adjusted with NaOH. Reaction buffer: 400 mM EDC freshly dissolved in the control buffer. MOPS was purchased from Sigma Life Science (M1254-100G) and EDC-HCl was purchased from Carl Roth GmbH (Art.-Nr. 2156.2). Only nuclease-free water from Ambion (AM9932) was used in all experiments.

DNA oligonucleotides

All DNA oligonucleotides were purchased from biomers.net GmbH with HPLC purification. The sequences were as follows. Template BA: TTT ATT ATT TTA TAT TAT TTA TTT, template ab: AAA TAA ATA ATA TAA AAT AAT AAA, replicate ~BA: CCC TTT ATT ATT TTA TAT TAT TTA TTT, replicate ~ab: CCC CCC CCC AAA TAA ATA ATA TAA AAT AAT AAA, fragment ~B: CCC TTT ATT ATT TTA (with or without a 3′-phosphate group for reaction or control experiments), fragment b: TAA AAT AAT AAA, fragment A: TAT TAT TTA TTT, fragment ~a: AAA TAA ATA ATA (with or without a 3′-phosphate group for reaction or control experiments), normalization strand: TAG TTA TGT CGA TAC GAG ACA ACA CCA TAT GCA TTT AAG TCG CTT GAA ATT GCT ATA AGC AGA GCA CTC GTA TCG ACA TAA CTA.

2.2 General reaction procedure

The DNA was mixed on ice at 1:1 ratio with the 2x reaction buffer in a maximum recovery PCR tube (Axygen, brand of Corning Life Sciences). Promptly, the tube was placed in a qTower3G cycler (Analytik Jena AG) at 20 ℃ to start temperature cycling (with 8 ℃/sec heating rate, 6 ℃/sec cooling rate). For serial dilution, the cycler was paused 2 seconds before the end of the corresponding 20 ℃ step and the tube was placed on ice. 10% of the reaction volume was precipitated for analysis and the volume was replaced with 5% volume of the fragment mix in water (each fragment at 20x its initial concentration) and with another 5% volume of the 2x reaction buffer. The tube was promptly placed back in the cycler, and the cycler program was continued. At the end of the reaction run time, the full sample volume was precipitated. The standard ethanol precipitation [70] was performed but with addition of 2 µl (for a sample volume of 30 µl) of glycogen (10 mg/ml), which is known to enhance the precipitation of short oligonucleotides. The DNA was allowed to precipitate overnight at 4 ℃ before centrifugation, cleaning the pellet with 70% ethanol, and air-drying it.

2.3   Reaction analysis method

Precipitated samples were redissolved in $x/50 * v$ volume of water, where x is the initial concentration of one of the templates and v is the volume of the sample used for precipitation. The samples were loaded onto the Bioanalyzer chip (Agilent Small RNA Kit) and the Bioanalyzer assay was performed according to the company's protocol.

Quantification of Bioanalyzer curves

The Bioanalyzer instrument provides an automated gel electrophoresis system running on a microfluidic chip. Per chip, a maximum of 12 samples could be analyzed. Although the Agilent Small RNA Kit was designed for detection of short RNA strands (6-150 nt), in this work a procedure for detection and quantification of short ssDNA strands (~10-100 nt) was established.

The full-timescale fluorescence traces of the example reaction (reaction is introduced in Section 3.1) and calibration samples are shown in Figure 1a. The area and position of the peak that belongs to a specific strand depends on its length, sequence, and present modifications (e.g. the 3′-phosphate group). Thus, calling the peaks of the reaction traces required studying traces of samples composed of various species combinations.

Calibration experiments showed that the area of the peak was proportional to the ssDNA strand concentration (Figure 1c). Using this, the concentrations of the templates and the emerging products were calculated from the two calibration traces ran on the same Bioanalyzer chip. The two calibration traces included a template or its complement and both product strands with the known concentration of 50 nM each (Figure 1b). Due to a varying gel/dye composition and subsequently slightly varying strength of the fluorescence signal between Bioanalyzer chips, the two calibration samples were included in every run.

In addition, a slight fluorescence signal variation between different wells of the same Bioanalyzer chip was observed. According to the company protocol, a commercially available "marker solution" with a known strand used for normalization should exclude this fluorescence variation. However, the marker strand overlaid with some of the fragment peaks and could not have been used as a fluorescence control in the analysis of the system from this project (Figure 1a). Instead, a 84 nt-long DNA strand was included as a normalization strand.

Quantification with the Agilent Bioanalyzer instrument established in this paper for ssDNA allowed to quantify many samples in a parallelized fashion using minimal amount of material per analysis (1 μl of 50 nM concentrated DNA strand per sample), roughly 100 times less than with HPLC.

Error quantification

To quantify the error that analysis on the Bioanalyzer instrument yielded, a control experiment was performed, where in addition to the marker solution, the same sample consisting of one template, two products, and the normalization strand was pipetted into every well. It showed that the area normalization
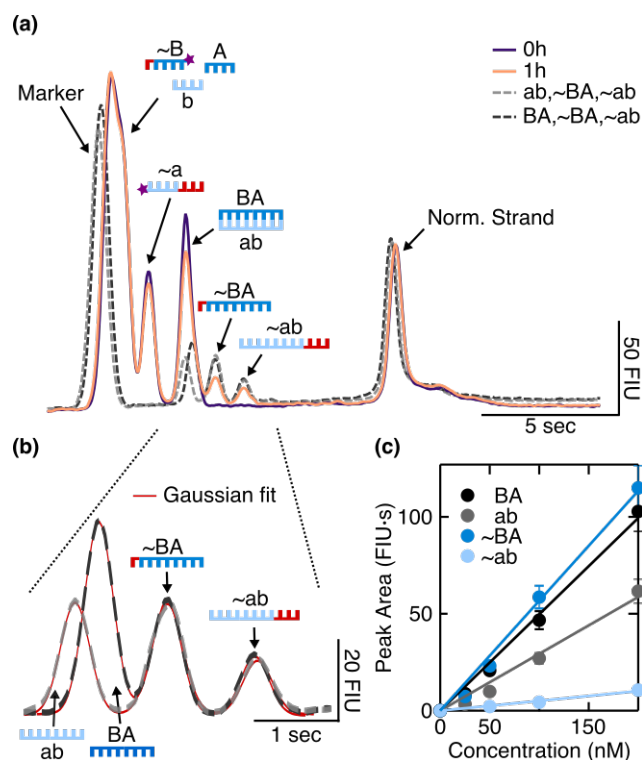
Figure 1: ssDNA analysis with Bioanalyzer gel electrophoresis.

As introduced in Section 3.1, ab and BA represent templates, ~ab and ~BA represent products of ligation, ~B, A, b, and ~a represent oligonucleotides that undergo templated ligation. (a) Full-timescale fluorescence traces of the reaction samples at 0 and 1 hours (solid lines) and calibration samples (dashed lines) from the microfluidic gel electrophoresis from Bioanalyzer. Calibration traces contained each strand at a fixed concentration of 50 nM and were required for quantification of the reactions. The normalization strand was included in each sample to correct for the fluorescence differences between different wells of the Bioanalyzer chip. A marker is used in standard Bioanalyzer applications instead of the additional normalization strand. In this work, the marker was not included, because the marker peak overlaid with some of the fragment peaks. (b) Zoom into the calibration traces (dashed lines) with additional Gaussian function fits (solid red lines). To facilitate the analysis of the products of the reaction, the 3 nt- and 9 nt-long 5′-polycytidine tags were included in the fragments ~B and ~a, respectively. This allowed to separate the products from the templates and individually quantify each product. Since the peaks of the two templates overlaid, two calibration samples were included in every Bioanalyzer run. The area of each peak was quantified by fitting a Gaussian function. (c) The increase in the peak area with the strand concentration showed a linear dependence for each template and product strand. Beyond 400 nM, a nonlinear dependence was observed and, therefore, the concentrations used in the analysis with Bioanalyzer were kept below this value.

between the curves of different wells could be performed by normalizing either to the peak area of the marker strand or to the peak area of the normalization strand introduced by us. The well-to-well error left after the normalization was 10%, which corresponds to the pipetting error. The 10% pipetting error was thus included as the predominant error source in the results.

2.4   Acquisition of melting curves

Melting curves were measured in a C1000TM Thermal Cycler with a CFX96TM detection system (Bio-Rad Laboratories) at a rate of 0.083 ºC/sec in either the control buffer or the reaction buffer. Eva Green dye (Biotium Inc.) was used at a 1x concentration for detection. Templates and replicates were measured at 1 μM concentration each, while all fragments were measured at 10 μM concentration each.

2.5   Quantification of the dissociation rates

Each melting curve was baseline corrected and normalized to depict the fraction of bound strands on the y-axis. The association constant, K_a, was quantified from the Arrhenius plot following the description in literature [56], using an in-house written LabVIEW routine (LabVIEW 2010, National Instruments). The dissociation rate, k_{off}, was then quantified from $K_a = k_{on}/k_{off}$, where the association rate, k_{on} (0.4 $\mu M^{-1} s^{-1}$), was taken from literature [62, 104].

2.6   Simulation and data fitting

The ODEs were solved in MATLAB using the solver ode15s. The experimental data was fitted with the *lsqcurvefit* tool with k_{mod} (introduced later in Section 3.3) as a single fit parameter.

3   Results

3.1   Experimental realization of the replication system

Experimental system

The initial mix of the experimental system (Figure 2a) consisted of two 24 bases-long DNA templates ab and BA and four shorter complementary DNA fragments A (12 bases), ~B (15 bases, with 12 binding to ab), b (12 bases), and ~a (21 bases, with 12 binding to BA). The templates were comprised only of adenine and thymine nucleobases and were designed to prevent interstrand hairpin formation. The base composition was limited to adenine and thymine only, in order to limit the sequence space for potential future projects when short random oligonucleotides could be analyzed (addressed in more detail in Section 4 "Discussion and Outlook"). The strands ~a (tagged a) and ~B (tagged B), apart from having parts complementary to the templates, also contained polycytidine tails at the 5′-ends important as tags for the detection with electrophoresis, and phosphate groups at the 3′-ends necessary for the ligation reaction.

When EDC was added to the initial mix, a nonenzymatic ligation of the fragments on the corresponding templates occurred ("Initial Ligation" in Figure 2a), producing the tagged replicates ~BA (27 bases with a 3nt tag) and ~ab (33 bases with a 9nt tag). Tagging the replicates with polycytidine tails allowed to distinguish products of the ligation reaction from the templates on the microfluidic gel electrophoresis Bioanalyzer instrument as explained in Section 2.3. Importantly, the reaction led to the formation of natural phosphodiester bonds at the ligation site. Once the replicates formed, they were cross-catalytically replicated through cycles of melting and annealing/ligation ("Replication" in Figure 2a).

Reaction conditions

The reaction was performed in the reaction buffer: 400 mM EDC freshly dissolved in 100 mM MOPS at pH 6.5. Under these conditions, EDC has a half-life in the range of tens of hours [89, 97, 100]. The reaction was cycled between 20°C for 0.5 hours and 65°C for 1 second. Additionally, every hour, 10% volume of the reaction was exchanged for a fresh solution of a fragment mix (b, ~a, ~B, and A) and EDC (Figure 2b). This step is further referred to as a feeding/dilution step, the goal of which was to apply an environmental pressure through an exponential dilution of the templates but at the same time to continuously provide the reaction with "food" for sustained replication. The EDC concentration was kept constant while the fragment concentration was increasing with each feeding step. I allowed the reactions to run for a total of up to 6 hours.
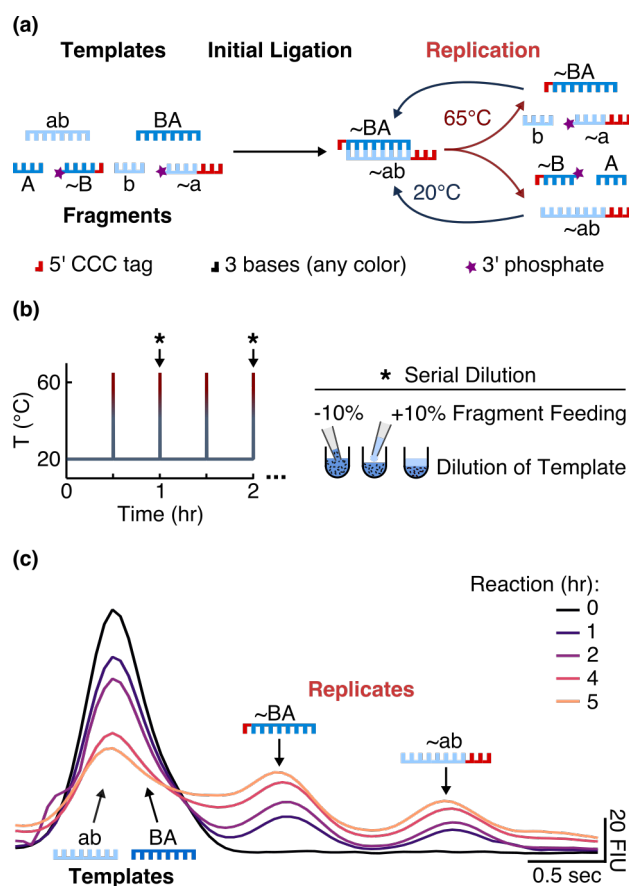
Figure 2: Replication system and detection.
(a) When templates were mixed with fragments and EDC at 20°C, an initial ligation occurred that formed the replicates ~BA and ~ab. Once the replicates were released from the templates during the melting step (not shown), both replicates could be further exponentially amplified. (b) The reaction was cycled between 20° C for 0.5 hours and 65°C for 1 second with the heating and cooling rates of 8 and 6 °C/sec, respectively. Every hour (black star and arrow), 10% of the reaction volume was replaced with 10% of the fragment mix in the reaction buffer. The original templates were thereby actively diluted, while the new fragments were available for the reaction. (c) Templates and tagged replicates separated on the microfluidic gel electrophoresis chip allowing individual quantification. Reactions were started with 1 μM of each template and 10 μM of each fragment in the reaction buffer and were stopped after 0, 1, 2, 4, and 5 hours. Over the course of the reaction, the template peak was decreasing due to the serial dilution, while the replicate peaks were increasing despite the serial dilution.
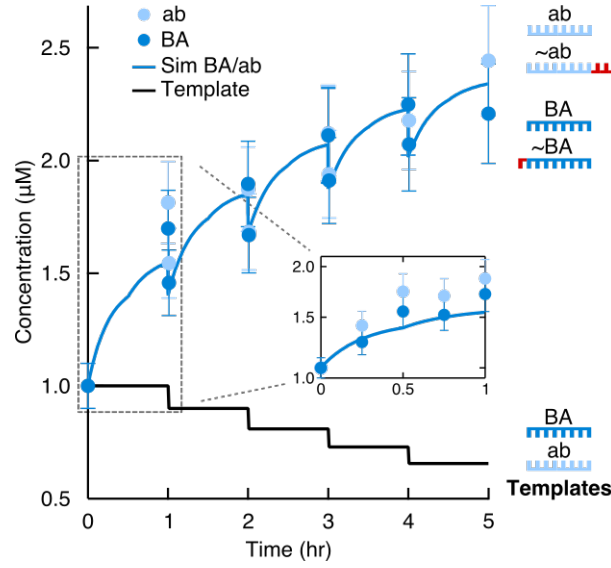
Figure 3: Replication by ligation.
Concentration of replicates from Figure 2c was quantified for the total ab and BA motifs (light and dark blue circles). The solid blue line shows the full kinetic simulation of the system (introduced in Section 3.3). The solid black line documents the calculated concentration of each template considering the dilution steps. The inset shows time points taken within the first hour and the simulation result for the first hour.

Surviving replicator

The reaction progression was monitored on the Bioanalyzer instrument. Quantification of the replicate concentrations required to

- fit a Gaussian function to each peak to obtain peak areas for each curve,

- normalize each curve to a peak area of an internal control strand, and

- calculate concentrations from peak areas, using peak areas of calibration curves that contained known concentrations of template and product strands as references (refer to Section 2.3 for more detail).

A consistent increase of replicate concentration in the reaction was observed while templates were being diluted out (Figure 2c, Figure 3). As a control, a reaction was initiated with 3′-dephosphorylated fragments ~a and ~B. As expected, no product was formed after 5 hours of monitoring the reaction, because a condensation reaction could not occur without a phosphate group at the ligation site (Figure 4).

In each experiment, the amplification factor was quantified separately for BA and ab motifs as $A_{BA} = c^{t=\Delta t}_{\sim BA+BA}/c^{t=0}_{BA}$ and $A_{ab} = c^{t=\Delta t}_{\sim ab+ab}/c^{t=0}_{ab}$, where A_{BA} and A_{ab} are the amplification factors for BA and ab, $c^{t=\Delta t}_{\sim BA+BA}$ is the BA concentration originating from template BA and the replicate ~BA at the end of the reaction ($c^{t=\Delta t}_{\sim ab+ab}$ represents the same for the ab concentration), and $c^{t=0}_{BA}$ is the BA concentration
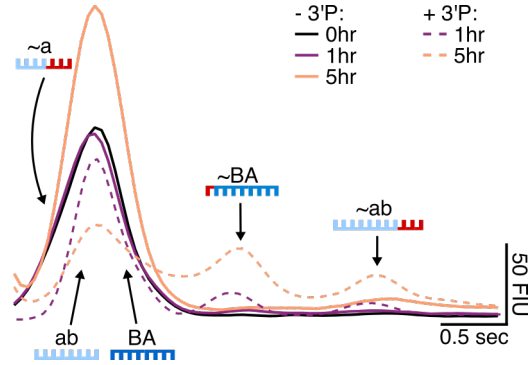
Figure 4: Control reaction in the absence of 3′-phosphate groups on the corresponding fragments. The gel electrophoresis traces in the presence (dashed lines) and absence (solid lines) of the 3′-phosphate groups. As expected, no replicates were formed in the absence of the 3′-phosphate group, since the condensation reaction between the 3′- and the 5′-ends of the corresponding fragments could not occur. Here, the area of the first peak of the 5-hour control reaction trace (solid orange line) is large because the peak of the ~a fragment without a phosphate group overlaid with the template peaks ab and BA. The area increase resulted from the hourly feeding of the ~a fragment.

originating only from the template at the start of the reaction ($c_{ab}^{t=0}$ represents the same for the ab concentration). The average amplification factor and its standard deviation were obtained from four different experiments where individual amplification factors for BA and ab were calculated (Table 1). Variation in the precipitation yields and pipetting errors are the likely sources of discrepancies between the amplification factors in different experiments. Taken together, this resulted in amplification fold for each template of 2.2 ± 0.5 within 6 hours. The original templates, meanwhile, were diluted such that 59% of the original templates was left after 6 hours (Figure 3).

Table 1: Reaction conditions and amplification factors for BA/ab motifs in four reactions.

Exp Nr.	$c_{BA}^{t=0}$ (μM)	$c_{ab}^{t=0}$ (μM)	Δt (hrs)	A_{BA}	A_{ab}
1	1	1	6	2.12	2.08
2	0.7	0.7	6	1.52	1.66
3	0.5	0.5	6	2.36	3.03
4	1	1	5	2.21	2.44

Both products were found to replicate symmetrically (dark and light blue marker points in Figure 3). The sequences for this reaction were chosen to have a symmetric ligation site on both sides: . . . TA-3′-PHO at the 5′-end ligating to 5′-TA. . . at the 3′-end.
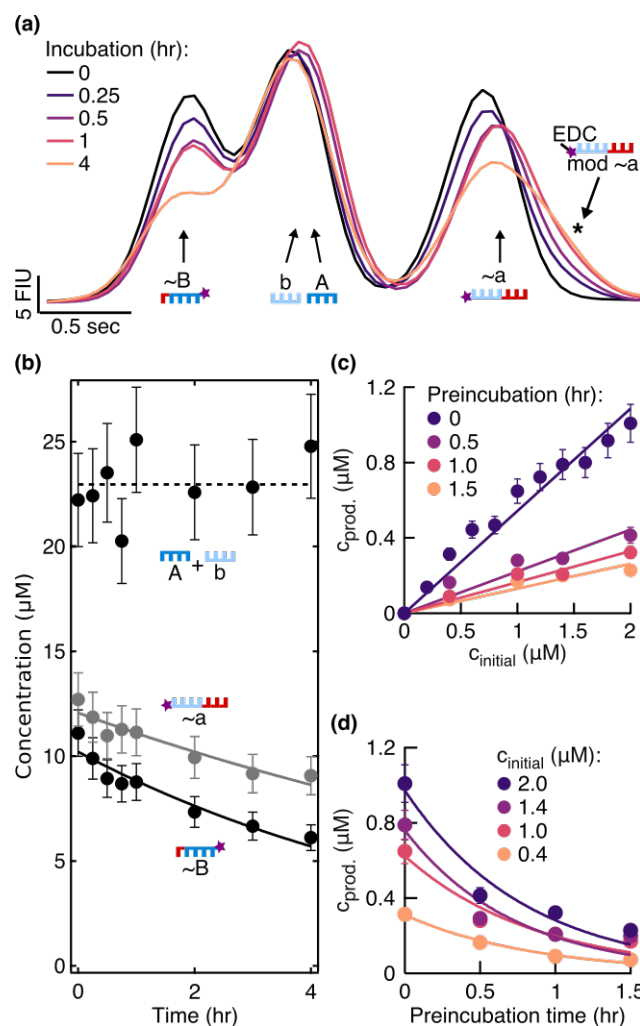
Figure 5: Effect of fragment modification on the ligation rate.
(a) Fragments were incubated in the reaction buffer under the temperature cycling conditions. A decrease in the signal from the phosphorylated fragments was observed and a shoulder formed on the right of the ~a peak (black star). (b) Quantification of concentrations from (a). The rates from exponential fits were 0.08 ± 0.01 and 0.14 ± 0.02 hr^{-1} for ~a and ~B, respectively. (c) Product concentrations, $c_{prod.}$, 0.5 hours after the start of the reaction plotted against initial template concentrations. Reactions were started with fragments preincubated in the reaction buffer for 0, 0.5, 1, and 1.5 hours. Linear fits gave ligation rates of 1.08 ± 0.04, 0.44 ± 0.04, 0.32 ± 0.02, and 0.26 ± 0.02 hr^{-1}, respectively. The respective replication yields were $(61 \pm 10)\%$, $(28 \pm 10)\%$, $(19 \pm 4)\%$, and $(15 \pm 3)\%$ in 0.5 hours. (d) Data from (c) were replotted to demonstrate the exponential decrease in product concentrations during the preincubation in EDC. The exponential fits yielded the rate constants of 1.14 ± 0.12, 1.14 ± 0.24, 1.38 ± 0.36, and 1.26 ± 0.24 hr^{-1} for the initial template concentrations of 2.0, 1.4, 1.0, and 0.4 μM, respectively.

24

3.2 Characterization of fragment modification

A closer look at the fragments incubated with EDC over time revealed that the fragments bearing 3′-phosphate groups decreased in concentration (Figure 5a). Additionally, a shoulder next to the peak assigned to fragment ~a started to form. Shoulder formation only occurred in the presence of the 3′-phosphate group (Figure 6a), suggesting that it originated from the fragment ~a modified by EDC on the 3′-phosphate group (Figure 6b). Since EDC-modified fragments are less negatively charged and 155 Da heavier than unmodified fragments, they were expected to run slower in a gel electrophoresis experiment, confirmed by the position of the shoulder on the right of the peak. The fragment ~B had a 3′-phosphate group and was decreasing in concentration over time as well (Figure 5a). This observation suggested that ~B was also modified by EDC on the 3′-phosphate group. The expected shoulder in this case was hidden by the dominant peak combined from the individual peaks of fragments b and A.

The position of each fragment was identified using a ladder (Figure 6c) and quantified (Figure 5b). Exponential fits, assuming full final degradation of phosphorylated fragments, produced degradation rates of 0.08 ± 0.01 hr^{-1} for ~a and 0.14 ± 0.02 hr^{-1} for ~B.

The effect of fragment modification on the ligation rate

According to our hypothesis, the 3′-phosphate group of the phosphorylated fragments was modified by EDC. If true, such modified fragments would still be able to bind to the templates and replicates but would not be able to participate in the ligation reaction. Therefore, a strong effect of a fragment incubation in EDC on the ligation rate was expected.

To test this hypothesis, I first measured the ligation rate under the reaction conditions. Reactions with varied initial template concentrations were followed for 0.5 hours and the product concentrations were quantified (Figure 5c). Product concentration showed a linear dependence on the initial template concentration, suggesting an underlying exponential behavior of the reaction. The ligation rate obtained from the linear fit was 1.08 ± 0.04 hr^{-1}. The experiment resulted in a replication yield of (61 ± 10) % in 0.5 hours, where the replication yield is defined as the replicate concentration at time t divided by the template concentration at time $t = 0$.

To test if EDC indeed formed a side product with the fragments and inhibited their ligation, I incubated the fragments in EDC for 0.5, 1.0, or 1.5 hours in the absence of templates under the temperature cycling conditions of the reaction. After the so-called preincubation time without templates, the template strands were added to start the ligation reaction. The ligation rate decreased to 0.44 ± 0.04, 0.32 ± 0.02, and 0.26 ± 0.02 hr^{-1} after 0.5, 1.0, and 1.5 hours of preincubation, respectively. Respective replication yields were $(28 \pm 10)\%$, $(19 \pm 4)\%$, and $(15 \pm 3)\%$ in 0.5 hours. In the experiment where the fragments were first incubated for 1.5 hours with EDC, both the replication rate and yield decreased 4-fold compared to no preincubation. The result suggested that EDC and fragments in the absence of templates were forming a side product that inhibited ligation.

To quantify the decrease in the ligation rate due to the preincubation of fragments, the data were
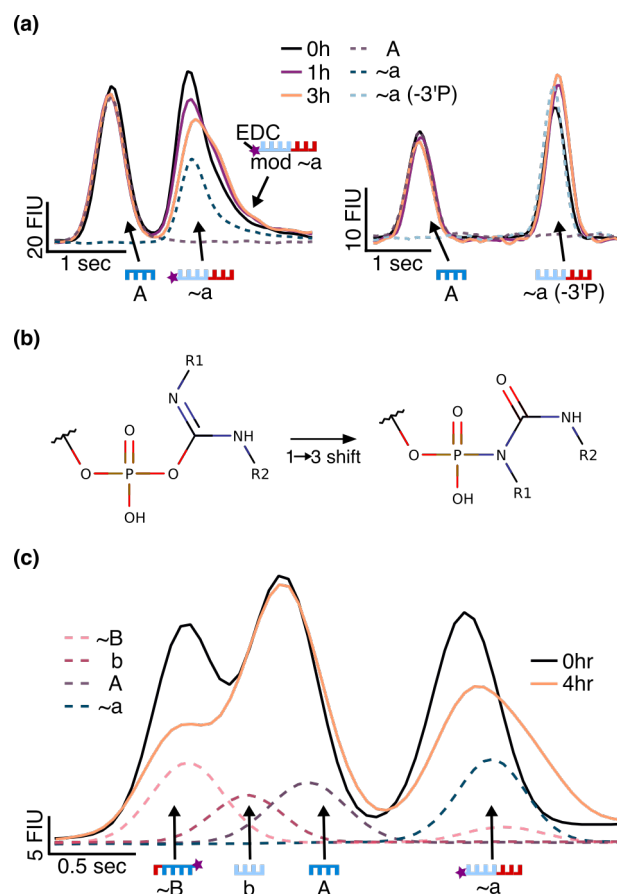
Figure 6: Fragment peaks on the gel electrophoresis traces, and 3′-phosphate group-dependent fragment modification.

(a) The gel electrophoresis traces of the samples where the fragments A and ~a were mixed together and incubated in the reaction buffer for 0, 1, or 3 hours (solid lines). Fragment ~a was used either with (left graph) or without (right graph) a 3′-phosphate group. Over time, a shoulder on the right of the ~a fragment peak emerged only in the case when the fragment ~a had a 3′-phosphate group. Running each strand separately (dashed lines) allowed the clear identification of the species peaks. (b) A likely pathway for the formation of a modified fragment is shown. The EDC-activated phosphate intermediate O-phosphorylisourea (left) could isomerize to produce N-phosphorylurea [43, 57] (right) that would not be able to support the condensation reaction. Since fragments with 3′-phosphate groups modified in such a way would still be able to bind to the corresponding templates, an inhibiting effect on the ligation reaction was expected. R1=C_2H_6, R2=CH_3–CH_2–CH_2–N–$(CH_3)_2$ (c) The gel electrophoresis traces of the samples where the four fragments ~B, b, A, and ~a were mixed together and incubated in the reaction buffer for 0 or 4 hours (solid lines). Running each strand separately (dashed lines) allowed the clear identification of the species peaks.

replotted with the final product concentration on the y-axis and the fragment preincubation time on the x-axis (Figure 5d). The exponential fits with assumed zero asymptote averaged to an effective first order rate constant of 1.23 ± 0.11 hr^{-1}.

3.3   Kinetic description and simulations

Simulation without fragment modification

To understand the replication behavior in the reaction, I modeled the experimental system with a set of rate equations:

$$BA + a + b \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [BA \cdot a] + b \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [BA \cdot a \cdot b] \tag{1}$$

$$BA + a + b \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [BA \cdot b] + a \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [BA \cdot a \cdot b] \tag{2}$$

$$ab + A + B \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [A \cdot ab] + B \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [B \cdot A \cdot ab] \tag{3}$$

$$ab + A + B \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [B \cdot ab] + A \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [B \cdot A \cdot ab] \tag{4}$$
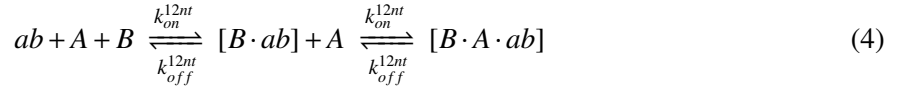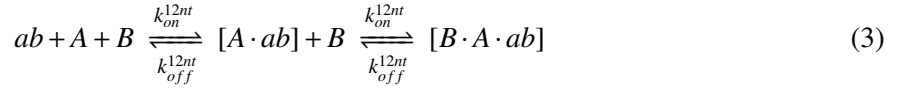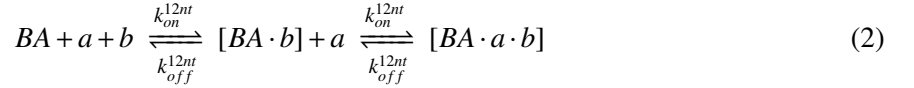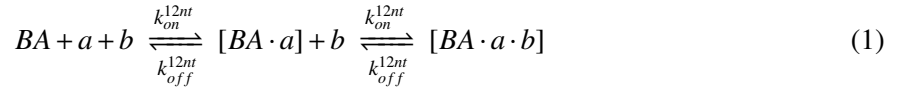
$$[BA \cdot a \cdot b] \overset{k_{lig}}{\longrightarrow} [BA \cdot ab] \tag{5}$$

$$[B \cdot A \cdot ab] \overset{k_{lig}}{\longrightarrow} [BA \cdot ab] \tag{6}$$

$$[BA \cdot ab] \underset{k_{off}^{24nt}}{\overset{k_{on}^{24nt}}{\rightleftharpoons}} BA + ab \tag{7}$$

$$A + a \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [A \cdot a] \tag{8}$$

$$B + b \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [B \cdot b] \tag{9}$$

Equations 1-4 describe the reversible formation of the termolecular complexes between templates and their complementary fragments with association and dissociation rate constants k_{on}^{12nt} and $k_{off}^{(12nt,T)}$, where T refers to the temperature dependence of the dissociation rate. Following the formation of the complex, an irreversible ligation reaction can occur with a ligation rate k_{lig} (equations 5-6). The resulting complex can reversibly dissociate with association and dissociation rate constants k_{on}^{24nt} and $k_{off}^{(24nt,T)}$ (equation 7) leading to the reshuffling of the strands such that the reaction can start again. Complementary frag-
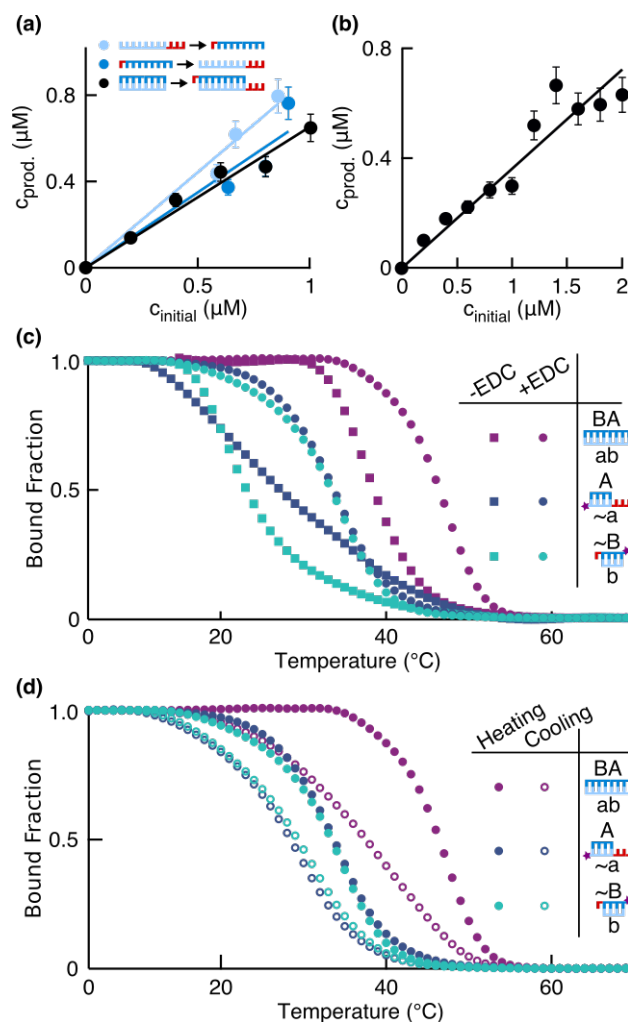
Figure 7: Determination of rates for the simulation.
(a) Product concentrations 0.5 hours after the start of the reaction. Templates were replaced with ~ab or ~BA (light and dark blue points) for the initial ligation, or initial ligation was started with templates as before (black points). The ligation rates from the linear fits were 1.78 ± 0.08, 1.40 ± 0.14, 1.32 ± 0.06 hr^{-1}, respectively. Respective replication yields were $(87 \pm 11)\%$, $(67 \pm 15)\%$, and $(55 \pm 6)\%$ in 0.5 hours. (b) Product concentrations quantified 10 min after the start of the reaction. The linear fit produced a ligation rate of 2.16 ± 0.12 hr^{-1}, with a replication yield of $(39 \pm 7)\%$ in 10 minutes. (c) Baseline-corrected melting curves of the complementary templates and complementary fragments in the control buffer *vs* in the reaction buffer. (d) Baseline-corrected melting and annealing curves of the complementary templates and complementary fragments in the reaction buffer used to quantify the dissociation rates.

ments can also reversibly bind to each other with association and dissociation rate constants k_{on}^{12nt} and $k_{off}^{(12nt,T)}$ (equations 8-9). I did not differentiate between ligation on templates and replicates that contained polycytidine tails since the tags were not found to significantly affect the ligation reaction as shown in Figure 7a.

Ligation rate k_{lig} I observed in the previous experiments that the replication rate slowed down during the first 30 minutes of the reaction (Figure 3, inset). I assumed that the product inhibition and fragment modification had already affected the reaction, and the ligation rate quantified from this data was being underestimated. To obtain the value k_{lig} that was independent of the product inhibition and fragment modification effects, product concentrations were measured 10 min after the start of the reaction for different initial template concentrations (Figure 7b). This ligation rate, which should be least affected by the modification reactions, averaged to k_{lig} of 2.16 ± 0.12 hr^{-1}, about two-fold higher as compared to measuring the product formation after 30 minutes.

Association and dissociation rates   I obtained the k_{on} value of 1440 μM^{-1} hr^{-1} for a short DNA oligonucleotide composed only of adenine and thymine nucleobases from literature [62, 104]. The rates k_{on}^{12nt} and k_{on}^{24nt} were assumed length independent and were both set equal to this k_{on} value. Since I wanted to explicitly model the temperature cycling, I needed to determine $k_{off}^{12nt,20°C}$, $k_{off}^{12nt,65°C}$, $k_{off}^{24nt,20°C}$, and $k_{off}^{24nt,65°C}$.

The values for $k_{off}^{12nt,20°C}$ and $k_{off}^{24nt,20°C}$ were quantified experimentally by analyzing melting curves. First, I showed that the duplexes in the reaction buffer were significantly more stable than in the control buffer (reaction buffer without EDC) (Figure 7c). Therefore, measuring the dissociation rates in the control buffer would have yielded overestimated values. Then, I measured annealing curves in the reaction buffer and observed a hysteresis between the melting and annealing curves for all the tested species combinations (Figure 7d). This meant that a slower temperature ramp was required to reach the dynamic equilibrium at each temperature point of data acquisition. However, I expected that the slower temperature ramp would increase the extent of EDC-induced unwanted modifications of the DNA bases in the single-stranded form [15]. Instead of slowing down the temperature ramp, the minimum and maximum $k_{off}^{x\,nt,20°C}$ values were estimated from the melting and annealing curves, respectively. I quantified $k_{off}^{24nt,20°C}$ and $k_{off}^{12nt,20°C}$ to be within the bounds of (2 to 64100) 10^{-4} hr^{-1} and (17 to 339) hr^{-1}, respectively. Although the range of $k_{off}^{x\,nt,20°C}$ values is large, these dissociation rates are slower than the melting kinetics of the implemented thermal cycling. The reaction model was not expected to be sensitive to variations of this parameter as indeed was the case (as shown below in Figures 8a, b).

The values $k_{off}^{24nt,65°C}$ and $k_{off}^{12nt,65°C}$ estimated in the same way yielded respective bounds of (0.4 to 9.6) 10^6 hr^{-1} and (0.4 to 1.1) 10^8 hr^{-1}. The result of the large dissociation rates is the absence of duplex strands at equilibrium at 65°C. Without significantly altering the simulation results, I approximated the rates $k_{off}^{24nt,65°C}$ and $k_{off}^{12nt,65°C}$ to be 100 times higher than the chosen k_{on} value. This allowed the dissociation of the majority of the DNA complexes during the simulated melting steps and reduced

Table 2: Two sets of rate constants that model temperature cycling in the system without and with modification (normal and bold terms).

Rate Constant	20°C for 30 min	65°C for 16 sec
Dissociation for 12nt long strands	$k_{off}^{12nt,20°C}$	$k_{off}^{12nt,65°C}$
Dissociation for 24nt long strands	$k_{off}^{24nt,20°C}$	$k_{off}^{24nt,65°C}$
Association for all strands	k_{on}	k_{on}
Ligation	k_{lig}	k_{lig}
Feeding	0	k_{feed}
Dilution	0	k_{dil}
Modification	$\mathbf{k_{mod}}$	$\mathbf{k_{mod}}$

the computational load for the simulation.

EDC hydrolysis, fragment feeding and dilution   The rate of EDC hydrolysis at the pH value 6.5 of the reaction was quantified from literature to be 0.047 hr^{-1} [80, 89, 97, 100], leading to 75% of the initial EDC concentration left after 6 hours of the reaction, maintaining the >2500-fold excess of EDC over the DNA oligonucleotides. Therefore, I did not explicitly model the hydrolysis of EDC in the simulation. Lastly, feeding and dilution were modeled to occur during the warm stage of the reaction. For this, the rates k_{feed} and k_{dil} were fixed to 0 for the reaction at 20°C and to 0.625 sec^{-1} and 0.00625 sec^{-1}, respectively, for the reaction at 65°C. The values were chosen in order to model the experimental feeding and dilution concentrations.

Simulation details   To simulate the reaction curves, 15 ODEs based on equations 1-9 (terms not in bold in the Table 3) were solved in MATLAB. To simulate temperature cycling, the ODEs were alternately solved with two sets of rate constants (Table 2). The ODEs were solved using an *ode15s* solver to obtain concentrations of all the reaction species at time points chosen by the solver. The absolute and relative tolerances of the solver were set to 10^{-4} and 10^{-2}, respectively.

   The simulation performed for upper and lower bounds of $k_{off}^{12nt,20°C}$ and $k_{off}^{24nt,20°C}$ values both produced curves that showed rapid exponential amplification behavior far from our experimental results (Figure 8a). Noteworthy, the simulation matched the experimental data of the first 0.5 hours, where the inhibition from the fragment modification was probably still weak.

Simulation with fragment modification

To test the hypothesis that the observed formation of the EDC-induced fragments could significantly affect the behavior of the system, I added a fragment modification pathway to the system of equations:
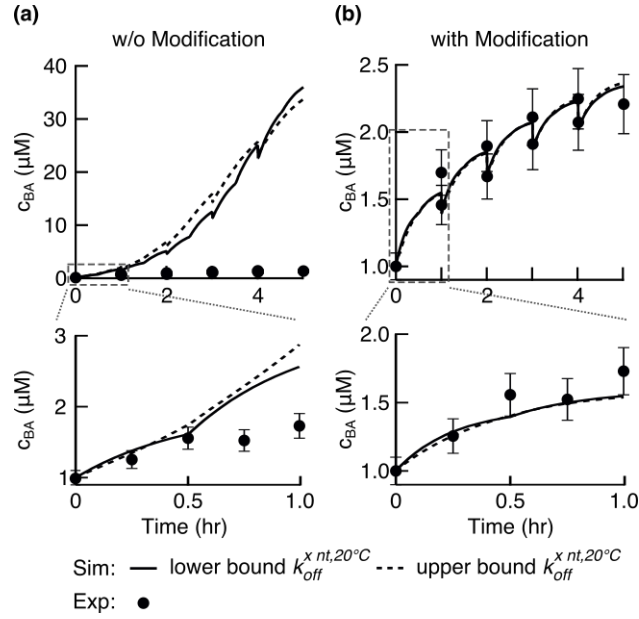
Figure 8: Simulation of the ligation reaction and fit of the experimental data.
(a) Cross-replication was modeled with a system of ODEs used to simulate the concentration of replicates. Without a modification pathway, the experimental rates predicted an exponential amplification of the templates that did not match the data. (b) By including a modification pathway, the deviations could be explained quantitatively.

$$B \xrightarrow{k_{mod}} B_{mod} \tag{10}$$

$$a \xrightarrow{k_{mod}} a_{mod} \tag{11}$$

$$[B \cdot ab] \xrightarrow{k_{mod}} [B_{mod} \cdot ab] \tag{12}$$

$$[BA \cdot a] \xrightarrow{k_{mod}} [BA \cdot a_{mod}] \tag{13}$$

$$[B \cdot A \cdot ab] \xrightarrow{k_{mod}} [B_{mod} \cdot A \cdot ab] \tag{14}$$

$$[BA \cdot a \cdot b] \xrightarrow{k_{mod}} [BA \cdot a_{mod} \cdot b] \tag{15}$$

$$[B \cdot b_{mod}] \xrightarrow{k_{mod}} [B_{mod} \cdot b] \tag{16}$$

$$[A \cdot a] \xrightarrow{k_{mod}} [A \cdot a_{mod}] \tag{17}$$

$$BA + a_{mod} + b \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [BA \cdot a_{mod}] + b \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [BA \cdot a_{mod} \cdot b] \tag{18}$$

$$BA + a_{mod} + b \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [BA \cdot b] + a_{mod} \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [BA \cdot a_{mod} \cdot b] \tag{19}$$

$$ab + A + B_{mod} \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [A \cdot ab] + B_{mod} \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [B_{mod} \cdot A \cdot ab] \tag{20}$$

$$ab + A + B_{mod} \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [B_{mod} \cdot ab] + A \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [B_{mod} \cdot A \cdot ab] \tag{21}$$

$$A + a_{mod} \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [A \cdot a_{mod}] \tag{22}$$

$$B_{mod} + b \underset{k_{off}^{12nt}}{\overset{k_{on}^{12nt}}{\rightleftharpoons}} [B_{mod} \cdot b] \tag{23}$$

Based on the experimental observation (Figure 5a, b), the fragments 'B' and 'a' were allowed to be modified with a modification rate k_{mod} independent of the fragments being in a free or a bound state (equations 10-17). The modification allowed the fragments to bind to the complementary fragments, the templates, and the replicates (equations 18-23) but inhibited the ligation reaction, sequestering templates and replicates from the reaction (equations 18-21).

Table 3: ODEs describing the experimental system without and with modification (normal and bold terms).

Left side	Right side of the equation
dB/dt	$-k_{dil} * B + k_{feed} - \mathbf{k_{mod}} * \mathbf{B} + k_{off}^{12nt} * ([B \cdot b] + [B \cdot ab] + [B \cdot A \cdot ab]) - k_{on}^{12nt} * (B * b + B * ab + B * [A \cdot ab])$
dA/dt	$-k_{dil} * A + k_{feed} + k_{off}^{12nt} * ([A \cdot a] + [A \cdot ab] + [B \cdot A \cdot ab] + [\mathbf{A \cdot a_{mod}}] + [\mathbf{B_{mod} \cdot A \cdot ab}]) - k_{on}^{12nt} * (A * a + A * ab + A * [B \cdot ab] + \mathbf{A * a_{mod}} + A * [\mathbf{B_{mod} \cdot ab}])$
db/dt	$-k_{dil} * b + k_{feed} + k_{off}^{12nt} * ([B \cdot b] + [BA \cdot b] + [BA \cdot a \cdot b] + [\mathbf{B_{mod} \cdot b}] + [\mathbf{BA \cdot a_{mod} \cdot b}]) - k_{on}^{12nt} * (b * B + b * BA + b * [BA \cdot a] + \mathbf{b * B_{mod}} + \mathbf{b * [BA \cdot a_{mod}]})$

da/dt	$-k_{dil} * a + k_{feed} - \mathbf{k_{mod}} * \mathbf{a} + k_{off}^{12nt} * ([A \cdot a] + [BA \cdot a] + [BA \cdot a \cdot b]) -$ $k_{on}^{12nt} * (a * A + a * BA + a * [BA \cdot b])$
dBA/dt	$-k_{dil} * BA + k_{off}^{12nt} * ([BA \cdot a] + [BA \cdot b] + [\mathbf{BA} \cdot \mathbf{a_{mod}}]) -$ $k_{on}^{12nt} * (BA * a + BA * b + \mathbf{BA} * \mathbf{a_{mod}}) + k_{off}^{24nt} * [BA \cdot ab] - k_{on}^{24nt} * BA * ab$
ab/dt	$-k_{dil} * ab + k_{off}^{12nt} * ([A \cdot ab] + [B \cdot ab] + [\mathbf{B_{mod}} \cdot \mathbf{ab}]) -$ $k_{on}^{12nt} * (ab * B + ab * A + \mathbf{ab} * \mathbf{B_{mod}}) + k_{off}^{24nt} * [BA \cdot ab] - k_{on}^{24nt} * ab * BA$
$d[BA \cdot ab]/dt$	$-k_{dil} * [BA \cdot ab] + k_{on}^{24nt} * ab * BA - k_{off}^{24nt} * [BA \cdot ab] + k_{lig} * ([B \cdot A \cdot ab] + [BA \cdot a \cdot b])$
$d[BA \cdot b]/dt$	$-k_{dil} * [BA \cdot b] + k_{off}^{12nt} * ([BA \cdot a \cdot b] + [\mathbf{BA} \cdot \mathbf{a_{mod}} \cdot \mathbf{b}] - [BA \cdot b]) +$ $k_{on}^{12nt} * (BA * b - [BA \cdot b] * a - [\mathbf{BA} \cdot \mathbf{b}] * \mathbf{a_{mod}})$
$d[BA \cdot a]/dt$	$-k_{dil} * [BA \cdot a] - \mathbf{k_{mod}} * [\mathbf{BA} \cdot \mathbf{a}] + k_{off}^{12nt} * ([BA \cdot a \cdot b] - [BA \cdot a]) +$ $k_{on}^{12nt} * (BA * a - [BA \cdot a] * b)$
$d[B \cdot ab]/dt$	$-k_{dil} * [B \cdot ab] - \mathbf{k_{mod}} * [\mathbf{B} \cdot \mathbf{ab}] + k_{off}^{12nt} * ([B \cdot A \cdot ab] - [B \cdot ab]) +$ $k_{on}^{12nt} * (B * ab - [B \cdot ab] * A)$
$d[A \cdot ab]/dt$	$-k_{dil} * [A \cdot ab] + k_{off}^{12nt} * ([B \cdot A \cdot ab] + [\mathbf{B_{mod}} \cdot \mathbf{A} \cdot \mathbf{ab}] - [A \cdot ab]) +$ $k_{on}^{12nt} * (A * ab - [A \cdot ab] * B - [\mathbf{A} \cdot \mathbf{ab}] * \mathbf{B_{mod}})$
$d[B \cdot A \cdot ab]/dt$	$-k_{dil} * [B \cdot A \cdot ab] - \mathbf{k_{mod}} * [\mathbf{B} \cdot \mathbf{A} \cdot \mathbf{ab}] + k_{on}^{12nt} * (B * [A \cdot ab] + A * [B \cdot ab]) -$ $2k_{off}^{12nt} * [B \cdot A \cdot ab] - k_{lig} * [B \cdot A \cdot ab]$
$d[BA \cdot a \cdot b]/dt$	$-k_{dil} * [BA \cdot a \cdot b] - \mathbf{k_{mod}} * [\mathbf{BA} \cdot \mathbf{a} \cdot \mathbf{b}] + k_{on}^{12nt} * (b * [BA \cdot a] + a * [BA \cdot b]) -$ $2k_{off}^{12nt} * [BA \cdot a \cdot b] - k_{lig} * [BA \cdot a \cdot b]$
$d[B \cdot b]/dt$	$-k_{dil} * [B \cdot b] - \mathbf{k_{mod}} * [\mathbf{B} \cdot \mathbf{b}] + k_{on}^{12nt} * B * b - k_{off}^{12nt} * [B \cdot b]$
$d[A \cdot a]/dt$	$-k_{dil} * [A \cdot a] - \mathbf{k_{mod}} * [\mathbf{A} \cdot \mathbf{a}] + k_{on}^{12nt} * A * a - k_{off}^{12nt} * [A \cdot a]$
$\mathbf{dB_{mod}/dt}$	$-\mathbf{k_{dil}} * \mathbf{B_{mod}} + \mathbf{k_{mod}} * \mathbf{B} + \mathbf{k_{off}^{12nt}} * ([\mathbf{B_{mod}} \cdot \mathbf{b}] + [\mathbf{B_{mod}} \cdot \mathbf{ab}] + [\mathbf{B_{mod}} \cdot \mathbf{A} \cdot \mathbf{ab}]) -$ $\mathbf{k_{on}^{12nt}} * (\mathbf{B_{mod}} * \mathbf{b} + \mathbf{B_{mod}} * \mathbf{ab} + \mathbf{B_{mod}} * [\mathbf{A} \cdot \mathbf{ab}])$
$\mathbf{da_{mod}/dt}$	$-\mathbf{k_{dil}} * \mathbf{a_{mod}} + \mathbf{k_{mod}} * \mathbf{a} + \mathbf{k_{off}^{12nt}} * ([\mathbf{A} \cdot \mathbf{a_{mod}}] + [\mathbf{BA} \cdot \mathbf{a_{mod}}] + [\mathbf{BA} \cdot \mathbf{a_{mod}} \cdot \mathbf{b}]) -$ $\mathbf{k_{on}^{12nt}} * (\mathbf{a_{mod}} * \mathbf{A} + \mathbf{a_{mod}} * \mathbf{BA} + \mathbf{a_{mod}} * [\mathbf{BA} \cdot \mathbf{b}])$
$\mathbf{d[B_{mod} \cdot ab]/dt}$	$-\mathbf{k_{dil}} * [\mathbf{B_{mod}} \cdot \mathbf{ab}] + \mathbf{k_{mod}} * [\mathbf{B} \cdot \mathbf{ab}] + \mathbf{k_{off}^{12nt}} * ([\mathbf{B_{mod}} \cdot \mathbf{A} \cdot \mathbf{ab}] - [\mathbf{B_{mod}} \cdot \mathbf{ab}]) +$ $\mathbf{k_{on}^{12nt}} * (\mathbf{B_{mod}} * \mathbf{ab} - [\mathbf{B_{mod}} \cdot \mathbf{ab}] * \mathbf{A})$
$\mathbf{d[BA \cdot a_{mod}]/dt}$	$-\mathbf{k_{dil}} * [\mathbf{BA} \cdot \mathbf{a_{mod}}] + \mathbf{k_{mod}} * [\mathbf{BA} \cdot \mathbf{a}] + \mathbf{k_{off}^{12nt}} * ([\mathbf{BA} \cdot \mathbf{a_{mod}} \cdot \mathbf{b}] - [\mathbf{BA} \cdot \mathbf{a_{mod}}]) +$ $\mathbf{k_{on}^{12nt}} * (\mathbf{BA} * \mathbf{a_{mod}} - [\mathbf{BA} \cdot \mathbf{a_{mod}}] * \mathbf{b})$
$\mathbf{d[B_{mod} \cdot A \cdot ab]/dt}$	$-\mathbf{k_{dil}} * [\mathbf{B_{mod}} \cdot \mathbf{A} \cdot \mathbf{ab}] + \mathbf{k_{mod}} * [\mathbf{B} \cdot \mathbf{A} \cdot \mathbf{ab}] +$ $k_{on}^{12nt} * (B_{mod} * [A \cdot ab] + A * [B_{mod} \cdot ab]) - \mathbf{2k_{off}^{12nt}} * [\mathbf{B_{mod}} \cdot \mathbf{A} \cdot \mathbf{ab}]$

$d[\mathbf{BA \cdot a_{mod} \cdot b}]/dt$	$-\mathbf{k_{dil}} * [\mathbf{BA \cdot a_{mod} \cdot b}] + \mathbf{k_{mod}} * [\mathbf{BA \cdot a \cdot b}] +$ $k_{on}^{12nt} * (b * [BA \cdot a_{mod}] + a_{mod} * [BA \cdot b]) - 2\mathbf{k_{off}^{12nt}} * [\mathbf{BA \cdot a_{mod} \cdot b}]$
$d[\mathbf{B_{mod} \cdot b}]/dt$	$-\mathbf{k_{dil}} * [\mathbf{B_{mod} \cdot b}] + \mathbf{k_{mod}} * [\mathbf{B \cdot b}] + \mathbf{k_{on}^{12nt}} * \mathbf{B_{mod}} * \mathbf{b} - \mathbf{k_{off}^{12nt}} * [\mathbf{B_{mod} \cdot b}]$
$d[\mathbf{A \cdot a_{mod}}]/dt$	$-\mathbf{k_{dil}} * [\mathbf{A \cdot a_{mod}}] + \mathbf{k_{mod}} * [\mathbf{A \cdot a}] + \mathbf{k_{on}^{12nt}} * \mathbf{A} * \mathbf{a_{mod}} - \mathbf{k_{off}^{12nt}} * [\mathbf{A \cdot a_{mod}}]$

I approximated the modification rate by using the average rate of the decrease of the ligation rate due to the fragment preincubation in EDC (1.23 ± 0.11 hr^{-1}) that was quantified above (Figure 5d). Since the new rate equations were introduced, 23 new ODEs were constructed (terms in bold in the Table 3). Since the solution of the kinetic model with fragment modification reproduced the experimental behavior of the system sufficiently well, I then used this kinetic model to fit the experimental data for the modification rate. The fit was realized with the *lsqcurvefit* tool of MATLAB with k_{mod} as the only fit parameter. For upper and lower bounds of $k_{off}^{12nt,20°C}$ and $k_{off}^{24nt,20°C}$ values, the fit produced the k_{mod} values of 2.69 hr^{-1} and 2.63 hr^{-1}, respectively. This was in good agreement with the experimental data (Figure 8b).

4   Discussion and Outlook

In this project, I established a DNA-based system that allows to run nonenzymatic exponential DNA replication-by-ligation experiments using EDC as a condensing agent for the first time over several cycles of ligation and melting to overcome product inhibition. Additionally, a selection pressure was mimicked by diluting the reaction. The cross-replicating DNA strands amplified 2.2-fold in 6 hours in a sustained fashion under the feeding conditions despite their serial dilution. This result indicates that DNA strands with the defined complementary sequences can survive under simple *in situ* activation conditions for well-chosen experimental parameters.

The problem of fragment modification

Faster amplification in the presented system was suppressed by the EDC-induced fragment modification, as was shown by detailed modeling of the experimental system with rate equations and simulation of the experimental data with ODE solutions. There are several possible mechanisms for the fragment modification. Activated oligonucleotides with a length as short as 4 nucleobases can effectively cyclize preventing them from base-pairing and consequently templated ligation reactions [32, 41]. Another source of fragment modification could have been the isomerization of the EDC-activated phosphate group intermediate, leading to the formation of a stable, non-reactive group, analogous to a known isomerization reaction of O-acylisourea to N-acylurea (Figure 6b). Fragments with non-reactive phosphate groups would still be able to base pair with templates but would not be able to support ligation reactions and, thus, would inhibit the replication by sequestering template strands.

Experimental and theoretical modification rates

Accounting for fragment modification in the kinetic model allowed to closely simulate the experimental behavior of the system with all the rate constants except the modification rate derived from the experiments. Importantly, I could obtain the modification rate k_{mod} from fitting the experimental data with the solution of the ODE system. The fitted value k_{mod}^{Sim} = 2.66 hr^{-1} matched the experimentally estimated value k_{mod}^{Exp} = 1.23 hr^{-1} within a factor of 2. The discrepancy likely originated from the temperature dependence of the modification rate or from the effects of the EDC hydrolysis product, its urea, on the DNA duplex stability. Both effects could not be explicitly modeled.

Effect of modulating reaction parameters

The kinetic model is central to finding reaction conditions that can support sustained replication. It predicts that the replication system is very sensitive to the modulation of the modification and ligation rates. Reduction in the modification rate by a factor of 1.5 should significantly speed up the replication (Figure 9a, orange solid line) with an amplification factor of four within 5 hours. Reduction by a factor
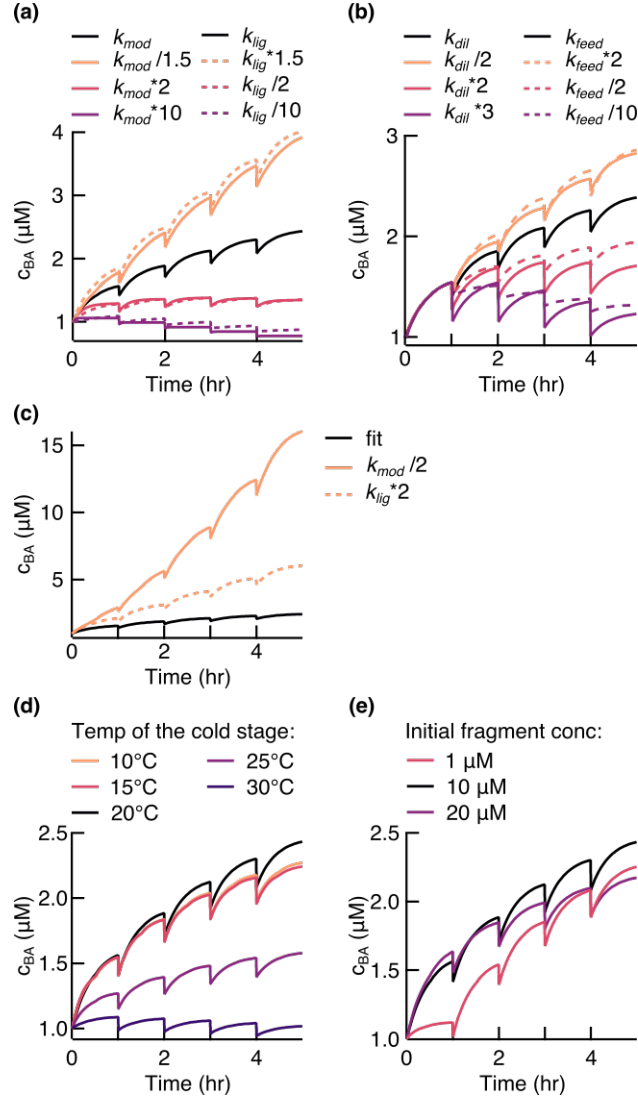
Figure 9: Simulation of the effect of the reaction rates modulation on the ligation reaction outcome. (a) The result of the modulation of rates k_{mod} and k_{lig} on the replication reaction. (d) The result of the modulation of dilution and feeding rates. (c) Additional curves to the modulation of rates k_{mod} and k_{lig} on the replication reaction: a factor two smaller k_{mod} or larger k_{lig} lead to a dramatically more effective replication system. (d) The result of the modulation of the temperature of the cold stage of the reaction on replication. (e) The result of the modulation of the initial fragment concentration on replication.

of two should result in a drastic increase in replication with an amplification factor of roughly fifteen within 5 hours (Figure 9c, orange solid line).

An increase in the ligation rate had a less significant effect on the amplification factor, i.e. the amplification factor reached four and six within 5 hours when the ligation rate was increased by factors of 1.5 and 2, respectively (Figure 9a and c, orange dashed line). Therefore, focus should be placed on finding ways to reduce the modification rate by, for example, finding a condensing agent with less side product formation, e.g. diamidophosphate [23], or a mechanism for a turnover of modified fragments. The effects of ligation and modification rates are, however, in direct competition. Thus, an improved activation protocol would need to maintain a high ligation rate. An increase in the modification rate and decrease in the ligation rate could easily result in a system with a negative fitness (Figure 9a).

Finding a balance between dilution and feeding is of equal importance. Higher dilution rate and less fragment feeding should result in a negative fitness of the system (Figure 9b). It could be of interest to decrease the dilution rate or increase feeding to drive stronger amplification. However, a too low dilution rate would lift the environmental pressure that is desired for fast evolution experiments, and too high feeding would result in insoluble fragment concentrations.

The effect of the modulation of other parameters was also estimated with the model, such as temperature of the cold reaction stage and the initial concentration of the fragments in the system. Lower temperature during the stage of the reaction when ligation occurs does not significantly affect the replication, while 5°C and 10 °C higher temperatures are predicted to significantly suppress the replication (Figure 9d). The effects of the temperature modulation originate from the higher dissociation rates of the duplexes at higher temperatures. The modification and ligation rates were assumed to be temperature independent in this analysis.

For the analysis of the effect of the initial fragment concentration on the replication reaction, the model allowed to conclude that the amount of fragment strands in the initial mix should be balanced between providing excess over template strands but not saturating the system with too many fragments (Figure 9e). Too many fragments would inhibit replication by sequestering template strands when modified.

Towards Darwinian-like evolution experiments starting from random sequence pools

The presented DNA cross-replicator works on a defined set of DNA strands. However, it is expected that any two strands brought together by a template in the presence of EDC and 3′-phosphate groups would ligate, producing longer molecules that can serve as templates in other reactions. The efficiency of templated ligation between random strands would likely depend on the geometry and, thus, the sequence composition of the ligation site [41]. Such sequence dependence in a pool of strands replicating by ligation could represent one way of how specific sequences were selected from random pools of short oligonucleotides likely present on the prebiotic Earth.

Replication-by-ligation of short strands could lead to Darwinian-like evolution experiments where longer strands with useful functions could be produced and selected from random pools of short oligonu-

cleotides. Strands in mixtures of short random oligonucleotides supplied with 3′-phosphate groups would serve at the same time as templates and fragments for ligation, creating a cascade of reactions. Cycles of ligation/melting in such systems could lead to effective strand elongation. Longer strands could then adopt secondary structures allowing them to perform useful functions. Since the replication in such a network could be achieved by chemical means by the demonstrated ligation chain reaction, the first oligonucleotide enzymes would not need to necessarily execute a replication function.

Evolution experiments from random sequence pools would require managing limited but large sequence spaces already at the sequence length of 12 bases. There are $2^{12} = 4096$ 12-nt long sequences composed of two types of bases. That implies that the feeding concentration for a replication reaction that started from random 12-nt long strands would be 4096*x μM, where x μM is the feeding concentration per strand. Thus, starting evolution experiments from shorter strands is of practical interest. Preliminary experiments show, however, that ligating the 6nt-long fragments on the 12nt-long templates does not result in sufficiently high yields at lower temperatures required for hybridization of shorter strands.

Short summary

In summary, in this project I realized a cross-replicating system composed of DNA oligonucleotides that are *in situ* activated with the carbodiimide-based condensing agent EDC that shares its reactive group with the tautomer of the prebiotically plausible cyanamide. The system achieved a positive fitness over several cycles of ligation/melting with a supply of fresh fragments every second cycle. I developed a detailed kinetic model in order to explain the experimental data. This model provided important insights into finding conditions for sustained replication. As an outlook, the realized reaction system serves as a stepping stone towards studying evolution of reaction systems composed of short random oligonucleotides. The goal will be to experimentally show how nonrandom reaction networks of longer oligonucleotides helping each other in preserving their sequence information might have originated.

Part II

Development of an assay to diagnose α1-antitriypsin deficiency disorder in blood plasma

5   Motivation and Introduction

Biological molecules in the context of living organisms interact inside complex environments, such as cell cytoplasms, extracellular matrices, and blood plasma. How strong two molecules interact with each other depends on the concentration of the interacting species, and on how likely the two molecules are to form a noncovalent bond with each other. The likelihood of a molecular interaction in solution, also called an affinity of interaction, is described quantitatively by an equilibrium dissociation constant K_D. A K_D value between the two interacting species A and B is given by $\frac{[AB]}{[A][B]}$, where $[AB]$ is the concentration of a molecular complex AB in equilibrium, $[A]$ is the concentration of free species A in equilibrium, and $[B]$ is the concentration of free species B in equilibrium.

The dissociation constant is commonly considered to be determined by the stereochemical properties of the interacting molecules. The properties of the environment, where interaction takes place, are often undermined. However, the various components of the complex environment inside living organisms could play a crucial role in determining the strength of the interaction between the molecules. In blood plasma a lot of parameters could affect the binding between the molecules of interest, such as the pH value, the salt conditions, the presence of additional small molecules, lipids, or proteins. These parameters could change the chemical properties of the molecules, such as their charge or hydration shell, hinder the interaction via competitive binding to the same binding site, or by changing the binding site via interacting with an allosteric binding pocket.

If this hypothesis is true, the ability to measure not only the concentration of the molecule of interest but also its affinity to the target in the same environment is of significant importance.

To test this hypothesis, I developed an assay sensitive to both the concentration of the interacting partners and the affinity of their interaction. Specifically, I focused on the interaction between the two proteins associated with a disease condition in the context of human blood plasma environment: the formation of the noncovalent encounter complex between the enzyme neutrophil elastase (NE) and its inhibitor α-1-antitrypsin (AAT).

α-1-antitrypsin deficiency disorder

The interaction between NE and AAT has been well studied *in vitro* using the two purified proteins in standard buffers [14]. AAT inhibits NE irreversibly in a two-step reaction [65]. The first and rate-determining step is the formation of a reversible encounter complex between NE and AAT, where AAT mimics NE's substrate [22, 40].

In a human body physiology, NE is known as an important regulator of the inflammatory processes. NE modulates the biological activity of cytokines and chemokines and activates receptors on neutrophils [64]. It is highly important that the body controls duration of the NE activity and, thus, the pro-inflammatory effect of NE, for example by AAT inhibition. Exactly how important the tight control over NE activity is, could be seen from a variety of diseases associated with low levels of AAT in blood plasma.

Low plasma levels of AAT predispose to a variety of diseases, most commonly recognized in patients suffering from lung emphysema and chronic obstructive pulmonary disease [25, 84]. The reason for decreased AAT levels is often genetical. A relatively common Z-mutation (Glu342Lys) in the AAT gene sequence [34] accounts for 95% of patients with clinically relevant manifestations. AAT with the Z-mutation is also called the Z-variant AAT.

Although all the homozygous carriers of the Z-variant AAT (ZZ-carriers) have reduced AAT levels, not everyone of them develops lung emphysema or other clinically relevant symptoms [25, 83, 84]. The reason for this is unclear and clinical tests predicting, who of the ZZ-carriers will develop symptoms do not exist.

During diagnostic tests in clinics, AAT concentration in plasma of ZZ-carries is determined. The standard methods are used, such as nephelometry or commercial ELISA. These assays, however, do not assess the functionality and inhibitory properties of the patient's AAT in the natural milieu of human plasma. Oligomers, inactive, cleaved, or oxidized forms of AAT, all contribute to the measurements, and the fraction of actually functional proteins remains unclear. Moreover, these approaches focus only on the concentration of AAT [25, 34, 83, 96], although kinetically efficient formation of the encounter complex also depends on the dissociation constant between NE and AAT [22, 40].

The problem of affinity determination in blood plasma

Functional assessment of the dissociation constant in native environments has been a major challenge to date [81, 85]. Immunoassays and surface plasmon resonance are the current standard approaches to quantify concentrations in plasma and determine affinities of known targets [54, 55]. However, both methods require surface immobilization of an antigen or an antibody, potentially imposing steric hindrance and molecular activity problems. Immobilization-free approaches based on fluorescence resonance energy transfer (FRET) require availability and labeling of both interacting partners [45, 92]. Therefore, FRET-based methods cannot be used directly in the plasma sample to determine the affinity of a person-specific endogenous protein to its target.

Microscale thermophoresis

Microscale thermophoresis (MST) is a recently established immobilization-free affinity measurement technique that has been applied to characterize ligand-binder interactions and affinity constants ranging from pM to mM under challenging conditions [37, 78, 96]. MST employs the physical phenomenon of thermophoresis – the movement of molecules in a temperature gradient. Biomolecules tend to move against the temperature gradient causing the depletion of a biomolecule in the heated spot. In MST experiment, the depletion of the fluorescently labeled molecule, called the binder, changes upon binding to another molecule, called the ligand. Measuring the depletion at increasing ligand concentrations results in a binding curve, which can be used to derive the dissociation constant.

In a previous work using an MST-based approach, it was shown how the affinity, as well as the

concentration of the ligand, can potentially be determined in complex matrices such as blood serum [49]. However, due to the limitations in the sensitivity of the assay, only an additionally spiked but not initially present ligand could have been detected in serum. Also, low affinity interactions remained a challenge, because for high K_D values, the approach requires very high and often unrealizable concentrations of the unlabeled binder in order to establish a complete binding curve.

Project goal

In this work, my goal was to develop an MST-based assay for measurements directly in blood plasma that would overcome the limitations of the previous approach. To test the assay, I aimed to apply it to study the formation of the NE-AAT encounter complex in blood plasma sample.

This project was conducted in a collaboration with Therese Dau from the laboratory of Dieter Jenne, both associated with Comprehensive Pneumology Center, Institute of Lung Biology and Disease (iLBD), University Hospital, Ludwig Maximilians University, and Helmholtz Zentrum München. My contribution to the project was in the assay development, its theoretical description, and evaluation of all the MST experimental data. All the MST experiments were conducted by both of us.

Since my contribution was based on the implementation of the MST technology, I will now give a more detailed introduction to a typical MST experiment and a literature overview of the currently established bioanalytical applications of MST.

5.1   A typical MST experiment

The biophysics-based technology of MST was first developed in 2010 [3] and stemmed from the research on why molecules undergo thermophoresis in solutions [16, 17]. Thermophoresis, or thermodiffusion, is a directed movement of molecules in a temperature gradient. The effect has first been described experimentally by Carl Ludwig in 1856 [51]. While thermophoresis in gases is a theoretically understood phenomenon, thermophoresis in liquids remains an active area of research. There are two common approaches to describing thermophoresis in liquids. One assumes that it is a local non-equilibrium effect and uses force fields, particle-particle potentials, and fluid dynamics among others to describe it [4, 59, 63, 76]. Another approach assumes a local thermodynamic equilibrium for moderate temperature gradients and describes thermophoresis applying equilibrium thermodynamics [10, 16].

In a typical thermophoresis experiment, the concentration of the molecule is depleted at higher temperatures. For a phenomenological description, a linear drift response is used [9]. The velocity of a molecule in a temperature gradient, v, depends linearly on the temperature gradient, ∇T, with a proportionality constant D_T:

$$v = -D_T \nabla T$$

D_T is called thermal diffusion coefficient or thermophoretic mobility. The thermal drift results in a concentration gradient, ∇c, leading to mass diffusion with a diffusion coefficient D. The flow density

considering the thermal drift and mass diffusion for a molecule of a concentration c is given by:

$$j = -cD_T\nabla T - D\nabla c$$

In steady state, the thermal drift and mass diffusion balance each other out, such that $j = 0$. This leads to:

$$\frac{\nabla c}{c} = -S_T\nabla T \tag{24}$$

where $S_T = D_T/D$ is a Soret coefficient of a molecule. Integrating equation 24 for small temperature differences dT and small concentration changes dc leads to:

$$\frac{c}{c_0} = e^{-S_T dT} \tag{25}$$

where c_0 is an initial concentration and c is a steady state concentration after temperature is changed by dT. Though theory continues to develop, the predominant parameters that affect the S_T of a molecule under physiological conditions were found to be the charge of the molecule, its size, and its interaction with the solvent [17]. Since at least one of the mentioned parameters is very likely to be affected when one molecule binds to another, the change in thermophoretic behavior upon binding is used as a readout to monitor and quantify binding interactions. The method was named microscale thermophoresis to highlight that the movement of molecules occurs within microscale temperature gradients.

In a typical MST experiment, a solution of a fluorescently labeled molecules is filled into a transparent glass capillary and placed under a fluorescence microscope (Figure 10a). A temperature gradient is created by an infrared (IR) laser that is coupled into the optical path of the fluorescence microscope. A small temperature increase of 4-12 K is obtained (Figure 10b). Most of the biologically relevant molecules have a positive S_T and undergo "positive" thermophoresis. This means that they move towards the colder area of the temperature gradient, resulting in a depletion of the fluorescence signal in the observed heated spot (Figure 10c). Depletion is calculated as the difference between the steady state fluorescence signal with the IR laser on and with the laser off. Since thermophoresis of free and bound states of a fluorescent molecule is different, the depletion value over the range of ligand concentrations varies (Figure 10c) and can be replotted against the ligand concentration to obtain a typical binding curve (Figure 10d, circles).

For two molecules A and B binding to each other with the dissociation constant K_D, one can write:

$$K_D = \frac{[A][B]}{[AB]} \tag{26}$$

$$[A]_{total} = [A] + [AB] \tag{27}$$

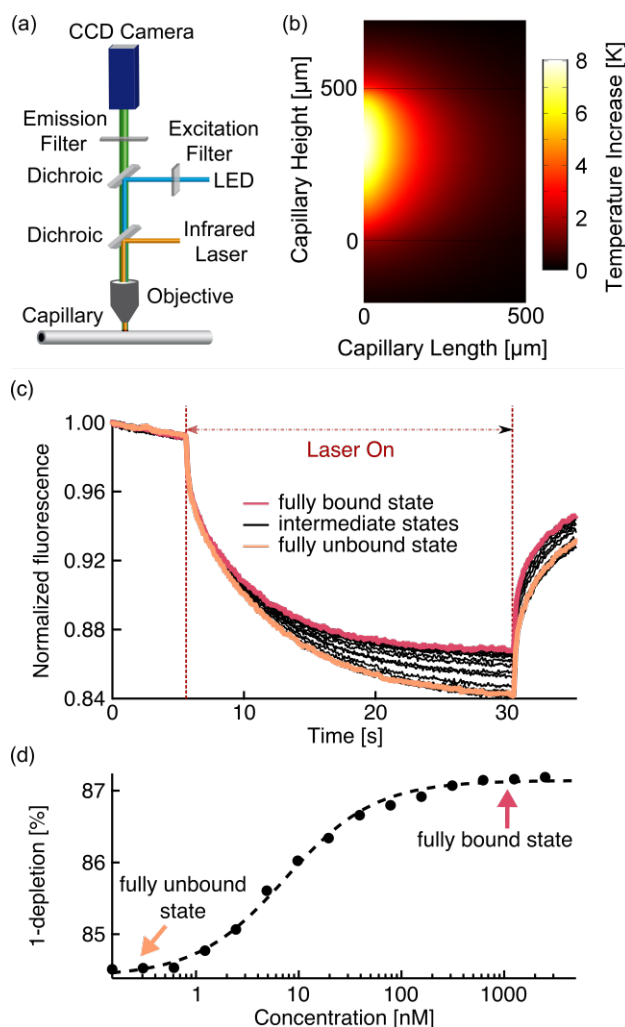$$[B]_{total} = [B] + [AB] \tag{28}$$

Figure 10: A typical MST experiment.

(a) Adopted from [49]. Sample with a fluorescent molecule is filled into a glass capillary. Illumination from the light-emitting diode (LED) is focused on the sample. The infrared (IR) laser for heating is coupled into the optical path of the microscope via the dichroic mirror and is focused on the sample with the standard objective. The detection is achieved with the charge-coupled device (CCD) camera. (b) Adopted from [49]. The established temperature profile calculated using finite element simulations and measurements with temperature sensitive dyes. (c) Typical time traces obtained in MST experiments. When the IR laser is switched on, the fluorescence in the observed spot is depleted due to the fast kinetics of the temperature dependence of the dye and the slow kinetics of thermophoresis. The steady state is achieved when thermophoresis and back-diffusion balance each other. Switching the IR laser off leads to the back diffusion of the molecules. The steady state depletion value depends on the state of the fluorescent molecule. Here, the bound molecule shows a smaller depletion value than a free molecule. (d) When $(1 - depletion)$ value is plotted against the concentrations of the titrated molecule, the binding curve is obtained (circles). It can be fitted with the solution of the two-body binding system between the fluorescent and the titrated molecules with a K_D value as the only fit parameter.

where $[A]$, $[B]$, and $[AB]$ are the equilibrium concentrations of molecules A, B, and the complex between A and B; $[A]_{total}$ and $[B]_{total}$ are total concentrations of molecules A and B (which includes both free and bound forms). Since total rather than equilibrium concentrations of molecules are typically known in the experiment, substituting $[A]$ and $[B]$ in equation 26 with $[A]_{total}$ and $[B]_{total}$ from equations 27 and 28 and solving the resulting equation for $[AB]/[A]_{total}$ results in:

$$\frac{[AB]}{[A]_{total}} = \frac{[A]_{total} + [B]_{total} + K_D - \sqrt{([A]_{total} + [B]_{total} + K_D)^2 - 4[A]_{total}[B]_{total}}}{2[A]_{total}} \tag{29}$$

$[AB]/[A]_{total}$ refers to a proportion of a molecule A that is found in a bound form in equilibrium. If molecule A is a labeled molecule in the experiment, $[AB]/[A]_{total}$ can be shown to be proportional to the fluorescence depletion at a given concentration of molecule B [49]. Equation 29 can thus be used to fit the binding curve with a K_D value as the only fit parameter (Figure 10d, dashed line).

5.2 Technical extensions of a typical MST experiment

The logical extension of the technique to a label free approach followed in 2012 with the incorporation of an ultraviolet (UV) lamp for visualization of molecules [79]. Proteins containing tryptophan residues can be monitored as long as neither the binding partner nor the buffer exhibits intrinsic fluorescence in the UV range.

Over the last years, the dynamic range of the technique in terms of the lowest K_D value that can be reliably resolved has significantly improved. K_D resolution is limited by the sensitivity of the detection of a fluorescent molecule of interest. A commercial device introduced in 2014 allowed to increase the sensitivity of fluorescence detection down to 10 pM concentration, resolving affinities as low as 1 pM [37].

Two different approaches to upgrade the technique to a high-throughput platform were successfully applied. A commercially available platform allows automatic pipetting of titration series and filling solutions into capillaries. The capillaries are positioned on the chips that can then be easily placed inside an automated MST machine for standard measurements [48].

In another high-throughput approach, the MST measurements are performed in the wells of 384-well microplates [27]. Nanoliter sample volumes are automatically dispersed onto the bottom of cone-shaped wells. Mineral oil is used to prevent evaporation of the samples. Measurements are then performed on a microscope setup, where the microplate is being automatically repositioned to obtain measurements in all the sample wells. Up to 24 sixteen-point binding curves can be measured per experiment.

Based on the variable formats of the available devices, a number of assays were proposed and successfully implemented over the last years and will now be discussed.

5.3   Literature overview of the bioanalytical applications of MST

Binding interactions between the purified molecules

The classical application of the MST method is to measure dissociation constants of interactions between two molecules available in a purified form. Typically, one of the molecules is fluorescently labeled and another is titrated against the first one to cover a concentration range of ~100 fold above and below the expected K_D value. The fluorescent label is commonly attached to the molecule, for which the labeling procedure is the easiest. The experimenter does not have to label the smaller of the two molecules, threatening to distort the binding, because as compared to FCS, the MST technique does not rely solely on the change in size of the fluorescent molecule upon binding. Where possible, also a label-free MST assay could be exploited for quantifying the bimolecular binding between two purified species.

Protein-protein binding   The possibility to quantify protein-to-protein interactions with an MST approach was established already in the first papers on the technique. For example, the dissociation constant of 10 nM was obtained for the binding between a human interferon and its specific antibody in a direct titration experiment [96].

An interaction of an even higher affinity with the dissociation constant of 620 pM was quantified for the therapeutic molecule Anakinra binding to its receptor, the interleukin-1. Therapeutic molecules tend to have affinities in the pM range. The ability to quantify dissociation constants of such interactions is of high importance when it comes to screening for modifications enhancing molecular stability and affinity.

The MST method can be applied to study bimolecular interactions in presence of other components of the binding system. For example, MST was successfully employed to study the influence of cofactors ATP/Mg^{2+} on the binding between protein kinase A (PKA) and its physiological inhibitor [38]. The dissociation constant dropped more than 200-fold in the absence of the cofactors. Only one component of the reaction, the inhibitor, had to be labeled for this study.

Sometimes, high concentrations of the titrated protein are not available due to protein's poor solubility, tendency to aggregate, or limited availability. A tendency to aggregate is not always easily discernible by eye. In the MST approach, such cases can be caught from the aberrations in the MST time trace signals (bumps in normally smooth curves) or in sudden drops of the fluorescence depletion values at high concentrations when no secondary binding between the molecules is expected [73].

Protein-peptide binding   An important advantage of using the MST technique for quantification of protein-peptide interactions lies in the possibility to place a label on a protein. The protein in case of protein-peptide interaction is the bigger of the two interacting partners. Labeling the protein minimizes the possibility of the label to interfere with the binding, which may occur when the label is placed on the peptide. FCS and fluorescence anisotropy techniques would, on the contrary, require labeling the smaller peptide to observe a stronger change in the translational or rotational diffusion upon binding.

In the method paper of 2013, for example, the binding between the malaria related protein AMA1

and RON2 peptide was quantified by MST [78]. The assay was performed in two different designs for comparison: with labeling the RON2 peptide and with labeling the AMA1 protein. The expected high affinity binding site could be quantified in both designs. However, only in the design with labeling the AMA1 protein, the second binding site of lower affinity could be observed, highlighting the importance of performing the assay in both configurations.

Binding between nucleic acids and proteins The quantification of binding interactions between nucleic acids and proteins was shown for a 50-bp long dsDNA oligonucleotide and a Ku70/80 protein complex [38]. Here, the MST technology was able to distinguish between the low- and high-affinity modes of binding and produced a complex biphasic signal that could be fitted independently with two binding curves.

A binding affinity with the dissociation constant of as low as 1 pM was successfully quantified by the MST technique for the interaction between the *Escherichia coli* single-strand binding (EcoSSB) protein and a ssDNA oligonucleotide of 70-nt in length [37]. Not just the affinity of the interaction was accessible but also the stoichiometry of the interaction was established in this study. Two 35-nt long ssDNA strands were binding per one EcoSSB protein molecule.

Binding to small molecules The affinities of interactions between the ATP aptamer and small molecules ATP and AMP were quantified in a proof-of-principle experiment in the first paper on MST [3]. This work was expanded by mapping the binding site of the ATP aptamer on the ATP molecule to a specific region of the adenine residue [18]. This study was done by scanning binding affinities of the aptamer to a range of ligands that were only slightly different from each other. Compared to the assays that required radiolabeling of either the AMP or the ATP aptamer for K_D determination [33, 39], the MST assay with a fluorescently labeled aptamer is simple and allows such ligand screenings to be performed within hours.

Compared to highly charged nucleic acids that have a strong thermophoretic response, it is a more challenging task to use thermophoresis to monitor binding between small molecules and proteins. As a proof of principle, the binding between the calmodulin protein and Ca^{2+} ions was successfully quantified [96]. Specificity of detection was proven by no observable binding to Mg^{2+} ions.

A binding of an ion to a protein induces a change in charge. To show that protein interactions with uncharged small molecules could also be quantified, a binding between an uncharged inhibitor quercetin and a PKA protein was successfully analyzed in the same paper.

Label-free MST approach is well suited for quantifying binding interactions between proteins and small molecules. If a protein contains more than two tryptophan residues, concentrations down to 100 nM could be detected. If a protein of interest lacks tryptophan residues, they could be introduced by mutation. To show feasibility of such an approach, the binding interactions between a mitogen-activated protein kinase p38α containing five tryptophan residues and three inhibitors were quantified [79]. Specificity of the detected interactions was controlled by no observable binding of the inhibitors to the denatured p38α.

In another study, inhibition of human neutrophil elastase (HNE) by their natural inhibitors penta-

cyclic triterpenes was studied. The study provides an example of how undesired protein aggregation can easily be controlled for during the MST experiments [86]. Pentacyclic triterpenes have a limited solubility in an aqueous medium. Thus, DMSO or ethanol had to be used with the disadvantage of conferring the possibility of a reduced HNE stability. When the labeled protein aggregated, the time traces recorded in an MST experiment showed characteristic bumps. By controlling the smoothness of the time traces, the authors observed HNE aggregation already in the presence of 5% DMSO but not in the presence of up to 20% of ethanol in the medium. For further quantifications of the dissociation constants, ethanol was used as a solvent of choice for the inhibitors.

Purification-free assays

Purification-free assays for protein-protein interaction studies can be divided into assays that utilize cell-free protein synthesis and assays that use cellular expression of proteins.

MST measurements on cell-free synthesized proteins were used to establish a hierarchy of the K_D values of the interactions between the human T-cell phosphoprotein ADAP and its potential ligands, various SH2 proteins, which were identified in a preceding pull-down assay [52]. ADAP and SH2 proteins were synthesized using commercially available cell-free systems. ADAP protein was SNAP-tagged, allowing a directed co-translational fluorescence labeling. All proteins were only partly purified by re-buffering to appropriate buffers. K_D values from 100 nM to 5 μM were then determined with the MST approach for different SH2 proteins, extending the results of the pull-down assay and allowing to hypothesize on the physiological relevance of the different ADAP interaction partners. Due to the high sensitivity of the fluorescence detection, only small amount of ADAP protein was required, showing that the assay could be used for proteins that are hard to obtain in high yields.

In a protocol published in the journal for visualized experiments [42], a purification-free approach was proposed for proteins expressed in cell lines. A signal transducer and activator of transcription (STAT) protein was expressed with a fused GFP molecule in HEK293 cells. The cells were lysed in non-denaturing conditions, and the cell lysate was diluted in physiological buffer to obtain GFP fluorescence intensity within the dynamic range of the MST instrument detector. Oligonucleotides of different sequences were mixed directly into the cell lysate for the MST measurements, making the method essentially protein purification-free. By adjusting the time for expression of the GFP-fused protein, one could adjust the required dilution factor of the lysate, allowing to measure the binding interactions in close-to-native conditions.

Binding in natural environments

As seen from the just mentioned purification-free assays, the MST method can quantify dissociation constants in complex environments, providing access to affinity information under close-to-native conditions.

This advantage of the MST approach was recognized already in the first publication on MST where

binding between the thrombin aptamer and thrombin was measured in the selection buffer as well as in 10% and 50% human serum [3]. The presence of serum shifted the observed K_D value more than 20-fold. The same tendency towards reduced affinity in serum was shown for the above-mentioned binding between the PKA protein and its inhibitor quercetin [96]. In that study, decrease in affinity was traced back to the binding between the quercetin and human serum albumin.

In contrast to using serum as a solvent for the fluorescent molecule, into which the ligand is titrated, one can use MST to perform quantitative analysis of the molecular interaction with the molecule of interest initially present in the serum/plasma of an individual. For example, as autocompetition assay developed by Lippok [49] was successfully used to quantify in one experiment both concentration and affinity of the autoimmune antibody against the cardiac β1-adrenergic receptor related to dilated cardiomyopathy. In this study, the serum containing the molecule of interest - the antibody L - was spiked with a labeled artificial antigen - the labeled binding partner B^* - such that all the labeled binding partner was in the bound form B^*L: $c(B^*)_{total} = c(B^*L)$, where c denotes concentration. The unlabeled form of the same binding partner B was then titrated in until it could completely compete out its labeled version: $c(B^*)_{total} \ll c(B)_{total}$ and $c(B^*)_{total} = c(B^*)_{free}$. The inflection point of the obtained binding curves reported on the concentration of the antibody in serum, while the amplitude reported on the affinity of the antibody to its antigen. As mentioned above, the limitations of this approach are its sensitivity and often an inaccessibility of low affinity interactions.

An assay developed in this Thesis

In this part of the Thesis, I developed a new MST-based competition assay that allowed to measure effects of both affinity and concentration of a molecule present directly in blood plasma [7]. The assay was tested by analyzing the NE-AAT binding in blood plasma of AAT deficient individuals. The dissociation constant of NE-AAT binding is on the order of μM. Applying the autocompetition assay introduced by Lippok (described in the paragraph above) would have required to obtain NE in mM concentration, which turned out to be impossible. The assay developed here managed to overcome this limitation by using another high affinity ligand of AAT, elafin.

In short in this assay, the plasma sample that contained the molecule of interest - AAT - was first spiked with the labeled binding partner of AAT - labeled neutrophil elastase (NE*). Differently from the previous approach, another binding partner of AAT - elafin - known to have a higher affinity for AAT compared to NE* was then titrated. Elafin competed out NE* and formed a complex with AAT.

The theoretical analysis of such an experimental setting showed that the amplitude of the obtained binding curve was sensitive to both concentration of AAT and its affinity to NE*. The amplitude, being a reporter of AAT concentration and affinity to NE*, correlated with the disease status of AAT deficient individuals better than the concentration of AAT alone.

In the next sections of this chapter, I will first give the theoretical analysis of the experimental system. I will then describe the experimental realization and application of the assay to the analysis of the NE-AAT binding in blood plasma of individuals with AAT deficiency disorder. I will finally discuss

interesting findings about the dependence of the AAT activity on the composition of the blood plasma environment.

6   Theoretical Considerations

Defining the system

In the competition assay introduced in this work, elafin competes for binding to AAT that is initially partially bound to NE* and partially is free in blood plasma. To describe such system theoretically, let us consider the chemical binding equilibrium between ligands I_1 and I_2 and a fluorescently labeled binder E, where I_1 is AAT in plasma, I_2 is elafin, and E is the labeled elastase (Figure 11a):

$$E + I_1 + I_2 \rightleftarrows EI_1 + EI_2$$
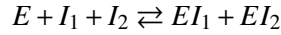
As described in Section 5.1, the binding is observed in MST experiments as the difference in the thermophoretic depletion of a molecule E in a free state and in the bound states EI_1 and EI_2. The thermophoretic depletion of a molecule is described by its Soret coefficient $S_T^{molecule}$. I assume that the Soret coefficients of the free and bound species of the molecule E in the sample are different:

$$S_T^E \neq S_T^{EI_1} \neq S_T^{EI_2}$$

Detected fluorescence

In a typical MST experiment, the temperature in the observed spot is increased by a small $\Delta T = T - T_0$ in the range of 5-10 K, where T_0 is a steady-state temperature before heating occurred and T is a steady-state temperature upon heating of the solution. From equation 25, the steady-state concentration of a molecule E, c^E, at the position where the temperature is increased by ΔT can be found as:

$$\frac{c^E}{c_0^E} = e^{-S_T^E \Delta T} \tag{30}$$

where c_0^E is the concentration of a molecule E when $\Delta T = 0$. (In Section 6, c_0^i refers to the concentration of a species i when $\Delta T = 0$).

For small enough ΔT, Taylor expansion can be used:

$$\frac{c^E}{c_0^E} \approx 1 - S_T^E \Delta T \tag{31}$$

Similarly for EI_1 and EI_2:

$$\frac{c^{EI_1}}{c_0^{EI_1}} = e^{-S_T^{EI_1} \Delta T} \approx 1 - S_T^{EI_1} \Delta T \tag{32}$$

$$\frac{c^{EI_2}}{c_0^{EI_2}} = e^{-S_T^{EI_2} \Delta T} \approx 1 - S_T^{EI_2} \Delta T \tag{33}$$

The fluorescence detected from a sample consisting of a mixture of E, EI_1, and EI_2 can be expressed as the sum of the fluorescence detected from the molecule E in free and bound states:

$$F = F_E + F_{EI_1} + F_{EI_2}$$

The detected fluorescence can further be expressed as a product of the quantum efficiency of the dye attached to the molecule $f^{molecule}$ and the concentration of a particular species:

$$F = f^E c^E + f^{EI_1} c^{EI_1} + f^{EI_2} c^{EI_2}$$

I assumed that quantum efficiencies of bound and unbound species are equal, because the dye on all the species is the same:

$$F = f c^E + f c^{EI_1} + f c^{EI_2} \tag{34}$$

By substituting c^E, c^{EI_1}, and c^{EI_2} in equation 34 with expressions derived from equations 31-33, the fluorescence detected after the sample was heated by a small ΔT can be expressed as:

$$F(\triangle T) = f c^E + f c^{EI_1} + f c^{EI_2} - (f c_0^E S_T^E \triangle T + f c_0^{EI_1} S_T^{EI_1} \triangle T + f c_0^{EI_2} S_T^{EI_2} \triangle T) \tag{35}$$

Thermophoretic depletion

By definition, the thermophoretic depletion is the fraction of fluorescence detected in the heated state and fluorescence detected in the cold state:

$$Depletion = \frac{F(\triangle T)}{F(\triangle T = 0)} \tag{36}$$

By substituting $F(\Delta T)$ in equation 36with equation 35 and $F(\Delta T = 0)$ with equation 34, after the algebraic transformations, follows that *Depletion* is given by:

$$Depletion = 1 - (S_T^E \frac{c_0^E}{c_0^E + c_0^{EI_1} + c_0^{EI_2}} + S_T^{EI_1} \frac{c_0^{EI_1}}{c_0^E + c_0^{EI_1} + c_0^{EI_2}} + S_T^{EI_2} \frac{c_0^{EI_2}}{c_0^E + c_0^{EI_1} + c_0^{EI_2}}) \triangle T \tag{37}$$

Amplitude

By definition, the amplitude of a binding curve in an MST experiment is the difference between the thermophoretic depletion at the negligible concentration of the titrated molecule I_2 ($Depletion_{start}$) and after adding the maximal concentration of the titrated molecule I_2 ($Depletion_{end}$) (also Figure 11b):

$$Amplitude = Depletion_{start} - Depletion_{end} \tag{38}$$

Assuming $c_0^{EI_2} = 0$ for $Depletion_{start}$, and $c_0^{EI_2} \gg c_0^E + c_0^{EI_1}$ for $Depletion_{end}$, equation 37 gives:

$$Depletion_{start} = 1 - (S_T^E \frac{c_0^E}{c_0^E + c_0^{EI_1}} + S_T^{EI_1} \frac{c_0^{EI_1}}{c_0^E + c_0^{EI_1}}) \triangle T \tag{39}$$

$$Depletion_{end} = 1 - S_T^{EI_2} \frac{c_0^{EI_2}}{c_0^E + c_0^{EI_1} + c_0^{EI_2}} \triangle T = 1 - S_T^{EI_2} \triangle T \tag{40}$$

From equations 38-40:

$$Amplitude = S_T^{EI_2} \triangle T - (S_T^E \frac{c_0^E}{c_0^E + c_0^{EI_1}} + S_T^{EI_1} \frac{c_0^{EI_1}}{c_0^E + c_0^{EI_1}}) \triangle T \tag{41}$$

The Soret coefficients $S_T^{EI_2}$, S_T^E, and $S_T^{EI_1}$ are intrinsic properties of the molecules EI_2, E, and EI_1. They define the behavior of these molecules in temperature gradients. $\frac{c_0^E}{c_0^E + c_0^{EI_1}}$ is the proportion of free E in steady state when only E and I_1 are present in the sample. Similarly, $\frac{c_0^{EI_1}}{c_0^E + c_0^{EI_1}}$ is the proportion of complex EI_1 in steady state when only E and I_1 are present in the sample.

Amplitude as a function of the I_1 concentration and the affinity between I_1 and E

As before in Section 5.1 and equations 26-29, $c_0^{EI_1}$ can be expressed in terms of total I_1, total E, and the dissociation constant between molecules I_1 and E that I denote here as $K_D^{EI_1}$:

$$c_0^{I_1 \ total} = c_0^{EI_1} + c_0^{I_1}$$

$$c_0^{E \ total} = c_0^{EI_1} + c_0^{E}$$

$$K_D^{EI_1} = \frac{c_0^E c_0^{I_1}}{c_0^{EI_1}}$$

$$c_0^{EI_1} = \frac{c_0^{I_1 \ total} + c_0^{E \ total} + K_D^{EI_1} - \sqrt{(c_0^{I_1 \ total} + c_0^{E \ total} + K_D^{EI_1})^2 - 4 c_0^{I_1 \ total} c_0^{E \ total}}}{2}$$

Thus, $c_0^{EI_1}$ is a function of total I_1, total E, and $K_D^{EI_1}$:

$$c_0^{EI_1} = f(c_0^{I_1 \ total}, c_0^{E \ total}, K_D^{EI_1}) \tag{42}$$

c_0^E is also a function of total I_1, total E, and $K_D^{EI_1}$:

$$c_0^E = c_0^{E \ total} - c_0^{EI_1} = f(c_0^{I_1 \ total}, c_0^{E \ total}, K_D^{EI_1}) \tag{43}$$

Finally, it follows from equations 41-43 that

$$Amplitude = f(c_0^{I_1\ total}, c_0^{E\ total}, K_D^{EI_1}) \tag{44}$$

$c_0^{I_1\ total}$ refers to the total concentration of AAT in plasma, $c_0^{E\ total}$ refers to the total concentration of labeled elastase NE that is being spiked into the plasma sample, and $K_D^{EI_1}$ refers to the dissociation constant between AAT and NE in the plasma sample. The total concentration of NE used in the developed competition assay is defined by the protocol. Therefore, the *Amplitude* can only differ if the total concentration of AAT in plasma samples or the affinity between AAT and NE is different between the samples. In other words, *Amplitude* is sensitive to the concentration of AAT and its affinity to NE directly in blood plasma.

7 Experimental Realization

In the following chapter, I will describe materials that were required for the assay development, details of the assay protocol and the binding model for simulations, and then the assay development itself on an example of AAT-NE binding in blood plasma.

7.1 Experimental Procedures and Data Analysis

Production of the recombinant proteins

All the recombinant proteins for this project were produced by my collaborator on the project Dr. Therese Dau associated with Comprehensive Pneumology Center, Institute of Lung Biology and Disease (iLBD), University Hospital, Ludwig Maximilians University and Helmholtz Zentrum München. The details can be found in the research article published based on this work [7].

Labelling of NE

The labeling was performed on the cysteine residues of the cysteine-tag of the recombinant NE protein. First, 1 mM of reducing agent dithiothreitol (DTT) was added to the recombinant NE in storage buffer (20 mM NaHPO$_4$, 300 mM NaCl pH 7.4) and the NE was incubated for two hours at room temperature to reduce all cysteine-tags. This step was important, as only reduced sulfur groups could be labeled. Then the DTT was removed by precipitating NE with 75% ammonium sulfate. This reduction and precipitation step was repeated once. Then, NE was re-dissolved in storage buffer, 5-fold molar excess of dye (Alexa Fluor® 647 NHS Ester, ThermoFisher Scientific) was added, and the solution was incubated at room temperature for one hour. To remove the excess dye, the solution was added to PD MiniTrap G-10 column (GE Healthcare) according to the manufacturer's protocol.

Determination of the concentration of the recombinant AAT and AAT in blood plasma

The concentration of AAT in blood plasma and the concentration of the recombinant AAT were determined by my collaborator using an activity assay with an active site titrated neutrophil elastase. The details on the assay can be found in the research article associated with this project [7]. The strong advantage of the assay is that allowed to measure the concentration of functional AAT in blood plasma.

Development of the binding model for simulations

The mass action law equations with two dissociation constants were used to describe the dependency of free NE (here: E) on the concentration of AAT (here: I_1) and on the affinity of AAT to NE in the presence of NE ligands AAT and elafin (here: I_2) (Figure 11a):

$$E + I_1 \rightleftharpoons [EI_1] \tag{45}$$

$$E + I_2 \rightleftharpoons [EI_2] \tag{46}$$

$$K_D^{I_1} = \frac{([E] - [EI_1] - [EI_2])([I_1] - [EI_1])}{[EI_1]} = [E](1-x)(\frac{[I_1]}{[EI_1]} - 1) \tag{47}$$

$$K_D^{I_2} = \frac{([E] - [EI_1] - [EI_2])([I_2] - [EI_2])}{[EI_2]} = [E](1-x)(\frac{[I_2]}{x[E][EI_2]} - 1) \tag{48}$$

where $K_D^{I_1}$ is the dissociation constant of NE-AAT interaction, $K_D^{I_2}$ is the dissociation constant of NE-elafin interaction, $[E]$, $[I_1]$, $[I_2]$ are the total concentrations of NE, AAT, and elafin in the reaction, respectively, $[EI_1]$ is the concentration of NE-AAT complex in the reaction, $[EI_2]$ is the concentration of NE-elafin complex in the reaction, and $x = \frac{[EI_1]+[EI_2]}{[E]}$ is the fraction of occupied NE.

Equation 48 was solved for $[E]$, which was then applied to equation 47 to yield a cubic equation for x. The cubic equation was solved numerically in Igor Pro 5.03 and the smallest real root of the equation was used to plot the fraction of bound NE, x, against different elafin concentrations, I_2, to simulate binding curves of the competition assay.

Determination of the binding curve amplitude

Human recombinant elafin (Sigma-Aldrich) was serially diluted (1:1) over five orders of magnitude in phosphate-buffered saline (PBS). Separately, plasma was diluted (final concentration: 7.5% v/v) in storage buffer substituted with 0.02% Tween 20 and anti-photobleaching enzyme and substrate components (Monolith Anti Photobleach Kit, NanoTemper Technologies) were added. This plasma solution was mixed 1:1 with PBS for background measurements. After adding fluorescently labeled NE (final concentration: 500 pM) to the plasma solution, it was mixed 1:1 with each concentration of the serial dilution of elafin. The samples were loaded into the Monolith™ NT.115 premium coated capillaries (NanoTemper Technologies) and incubated at 22°C for 2 hours. The samples were measured in the instrument (Monolith NT.115Pico, NanoTemper Technologies) at 22°C using 60% light-emitting diode and 40% infrared laser (IR) powers with IR laser on/off times of 25/5 seconds. Each dilution point was measured in triplicates. For each plasma sample, the whole procedure was repeated three times to yield independent triplicates.

In MATLAB, the time trace of the background signal was subtracted from the time traces of the sample signals. Then, the fluorescence after the temperature jump and equilibrated thermophoresis was normalized to the fluorescence before the IR laser heating, yielding fluorescence depletion values. Fluorescence depletion values of three technical replicates (mean ± S.D.) were plotted in per mille units (‰) on a linear y-axis against the concentration of the serially diluted elafin on the log10 x-axis resulting in binding curves. The binding curves were shifted so that the depletion values at the baseline where elafin concentration is saturating were the same. Several example binding curves are shown in Figure 11b.

To yield the amplitude A of each curve, a weighted with standard deviation fit was performed to the

solution of the quadratic equation that describes binding between two species (based on equation 29):

$$f(x) = A \frac{[E] + x + K_D - \sqrt{([E] + x + K_D)^2 - 4[E]x}}{2[E]} + t \tag{49}$$

where K_D is the dissociation constant of NE-elafin interaction, $[E]$ is the total concentration of NE, x is the total concentration of elafin, A is the amplitude of the curve, t is the y-offset of the curve, with A, t, and K_D as free fit parameters. The fitting was performed in Igor Pro 5.03.

Determination of the dissociation constant between AAT and NE

The amplitude from the competition binding assay is sensitive at the same time to AAT affinity to NE and AAT concentration. In order to quantify only the AAT affinity to NE in a plasma environment, the standard MST assay was also performed with purified AAT as follows. Freshly purified AAT was serially diluted (1:1) over five orders of magnitude in the storage buffer. Separately, plasma was diluted in the storage buffer substituted with 0.2% Tween 20 (final plasma concentration: 7.5% v/v) and anti-bleaching reagents were added. This plasma solution was mixed 1:1 with the storage buffer for background measurements. Fluorescently labeled NE (final concentration: 5 nM) was added to the plasma solution and mixed 1:1 with the serial dilution of AAT. The samples were incubated for 2 hours at 22°C and measured at 22°C in the instrument using 40% light-emitting diode and 40% infrared laser (IR) powers with IR laser on/off times of 20/5 seconds. Binding curves were obtained as described in section 7.1. They were then additionally normalized to 0 for the lowest AAT concentration and to 1 for the highest AAT concentration. Binding curves normalized this way display the proportion of bound NE at each titration point. Such curves are shown for example in Figures 13-14. The global fit of at least three replicates to equation 49 was performed to yield the dissociation constant of each curve.

Plasma samples and FEV1 data

Many AAT-deficient patients suffer from or are predisposed to develop lung emphysema and chronic obstructive pulmonary disease. In clinics, the lung function is assessed by the volume of air that can be forcibly exhaled during 1 second. This parameter is called the forced expiratory volume in 1 second or FEV1. It is expressed as the percentage (%) predicted for age, sex, height, and race. All plasma samples and the FEV1 values data were kindly provided by the collaborator Prof. Rob Stockley in Lung Investigation Unit, Queen Elizabeth Hospital Birmingham, UK.

Statistical analysis

Results are given as median values. A Mann-Whitney test without assuming Gaussian distribution was applied to all studies (***$p<0.001$). Statistical analysis was done with GraphPad Prism 6 software (GraphPad Software).

7.2   Assay development

In the competition assay, a low affinity inhibitor I_1 (here: AAT) competes with a high affinity inhibitor I_2 (here: elafin) for the catalytically inactive labeled enzyme E (here: NE) (Figure 11a). Inactivation of E was important to prevent the removal of I_1 from the equilibrium by the irreversible formation of a covalent complex.

The labeled E at a fixed final concentration of 500 pM was mixed with a 12-fold diluted plasma containing the molecule of interest I_1. To obtain binding curves, increasing amounts of I_2 were added and thermophoretic depletion of free E and E bound to I_1 (EI_1) or I_2 (EI_2) was measured for each concentration of I_2 (Figure 11a, b). From the obtained binding curves, one can observe that the depletion in the low end range of I_2 concentrations varied between plasma samples while the depletion in the high end range of I_2 concentrations did not change (Figure 11b). Thus, the amplitude of the binding curves varied between plasma samples (Figure 11b). This led to the hypothesis that I_1 concentrations and, more importantly, I_1 affinities to E are not necessarily constant in the person-specific complex environments of blood plasma (Figure 11b).

To characterize the system at different affinities between I_1 and E (Figure 11c) and different I_1 concentrations (Figure 11d), the system was described with mass action law equations and the corresponding binding curves were simulated as described in Section 7.1. The simulation results indicate that the amplitude increases with an increasing $K_D^{EI_1}$ (Figure 11c, small insertion graph) and decreases with increasing I_1 concentrations (Figure 11d, small insertion graph). Hence, the amplitude reports on both I_1 concentration and $K_D^{EI_1}$ as described mathematically in Section 6.

Analysis of the amplitude

The amplitude is defined as the difference between minimal and maximal depletion signals in the low and high end ranges of I_2 concentrations, respectively (Figure 11b). To get an intuitive understanding of what the amplitude of the binding curve in the introduced competition assay depends on, let us analyze the theoretical equilibrium for both situations separately in more detail.

In the low end range of I_2 concentrations, interactions between I_2 and E are almost non-existent. Therefore, the proportion of bound E depends solely on its interaction with I_1 (Figure 11b, dark grey box on top of the graph). The proportion of bound E at this point is a function of both, the concentration of I_1 as well as the dissociation constant between E and I_1 ($K_D^{EI_1}$) in plasma. Higher concentrations of I_1 shift the equilibrium towards EI_1 complexes, while a higher $K_D^{EI_1}$ results in a decrease of EI_1 complexes (Figure 11a, dark grey box on top of the graph). The thermophoretic depletion of I_1-bound E is considerably larger than that of the free E. Hence, the measured thermophoretic depletion reflects the proportion of E bound to I_1 in the mixture that varies between the plasma samples.

At maximal concentration of I_2, all E is bound to I_2 independent of the other parameters due to the high affinity between E and I_2 (Figure 11b, light grey box). The thermophoretic depletion is the depletion of EI_2 complexes and is the same for all the plasma samples and only depends on the amount of labeled
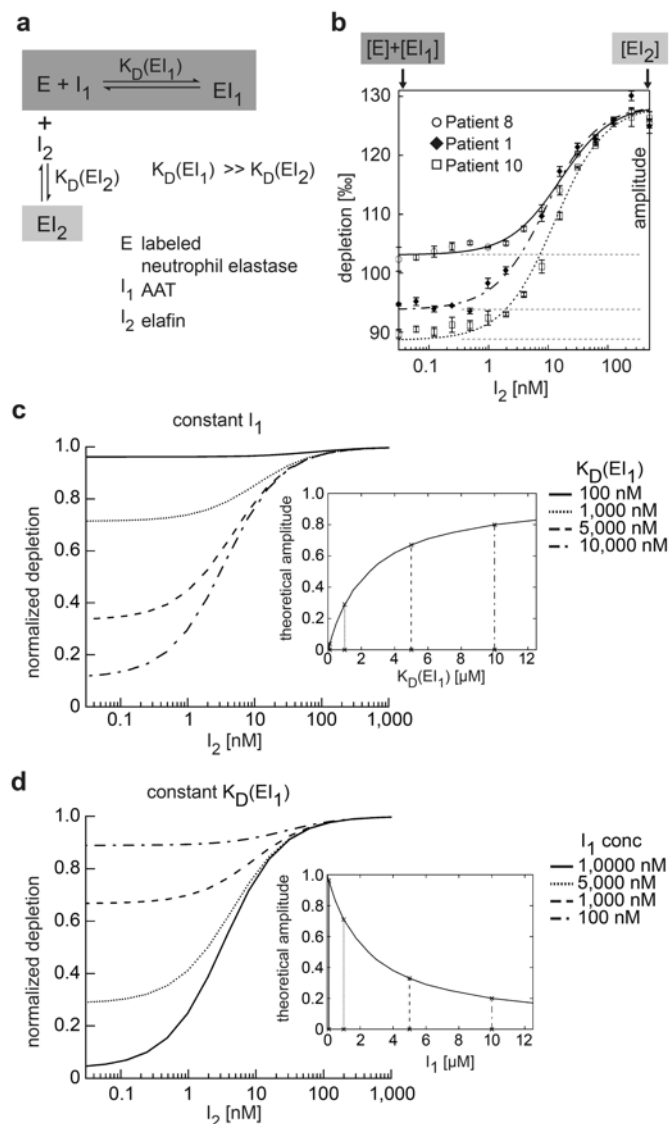
Figure 11: Dependency of the amplitude on the AAT concentration and the dissociation constant of NE-AAT binding.

(a) The theoretical background of the method. At each data point, the inhibitor I_1 (here: AAT) competes with I_2 (here: elafin) for the enzyme E (here: labeled inactive NE). The affinity between E and I_2 is much higher than the affinity between E and I_1. (b) To a constant amount of plasma (7.5%) containing I_1 and a constant amount of E (500 pM) increasing concentrations of I_2 were added and the thermophoretic depletion measured. Titration curves with theoretical fits are given for three individuals homozygous for the Z-variant AAT (patients 1, 8 and 10) as examples. Each depletion value is given as mean ± S.D. from three technical replicates. The amplitude is defined as the difference between the minimal depletion signal in the low end range (dark grey box with arrow) and maximal depletion signal in the high end range of I_2 concentrations (light grey box with arrow). The thermophoretic depletion of free E is smaller than that of E bound to I_1 or I_2, and depletion at each I_2 concentration indicates the proportion of bound E. (c) and (d) Simulations according to the mass action laws. (c) shows that the theoretical amplitude (simulated as the amount of bound E) increases when the dissociation constant between E and I_1 increases. (d) shows that the theoretical amplitude decreases with the increasing concentrations of I_1.
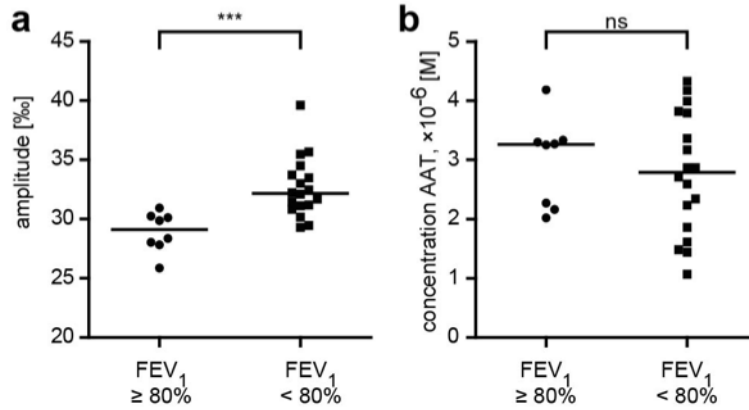
Figure 12: Amplitude, but not plasma concentration of AAT correlates with disease severity.
(a) ZZ-carriers (n = 26) were divided into two groups with high FEV1 ($\geq$ 80%, n = 8) and low FEV1 (< 80%, n = 18), which is a clinical parameter of airway obstruction. The amplitude was significantly lower (P = 0.0008) in samples with high FEV1 (***P<0.001; ns not significant). The amplitude is given as a mean value determined from at least two independent experiments. (b) The concentration of functional AAT was not significantly different (P = 0.64) between the groups with high FEV1 ($\geq$80 %) and low FEV1 (< 80%). The AAT concentration was measured twice by elastase titration in duplicates, and is given as a mean value here.

probe E spiked into a sample. Thermophoretic depletion of I_2-bound E is also considerably larger than that of the free E.

Taken together, the amplitude depends on the proportion of I_1-bound E in the sample at I_2 concentrations approaching zero. The largest amplitude is obtained when all E is free at lowest I_2 concentration in equilibrium. Accordingly, the amplitude decreases with decreasing $K_D^{EI_1}$ values and increasing I_1 concentrations (Figure 11c, d, smaller insertion graphs).

7.3 Amplitude and disease status

To evaluate whether the amplitude, which reports on AAT concentration and AAT affinity variations, represents a better lung function parameter than AAT concentration alone, both parameters were determined in plasma samples of individuals homozygous for the AAT Z-variant as described in Sections 7.1 and 7.1. They were then compared to FEV1 values (Section 7.1). FEV1 values allow to assess the lung function in AAT-deficient individuals [5, 83].

Comparison of the correlations between the AAT concentration and FEV1 and the amplitude and FEV1 showed that a significantly lower amplitude (P=0.0008) was observed for plasma samples from individuals with high FEV1 ($\geq$ 80%) than from those with low FEV1 (< 80%) (Figure 12a). The concentration of functional AAT did not differ between the two groups (P = 0.64) (Figure 12b). Hence, inclusion of affinity with the introduced assay resulted in a much better correlation to disease status and severity. To explain the discrepancy between the quantifying just the concentration and a parameter sensitive to
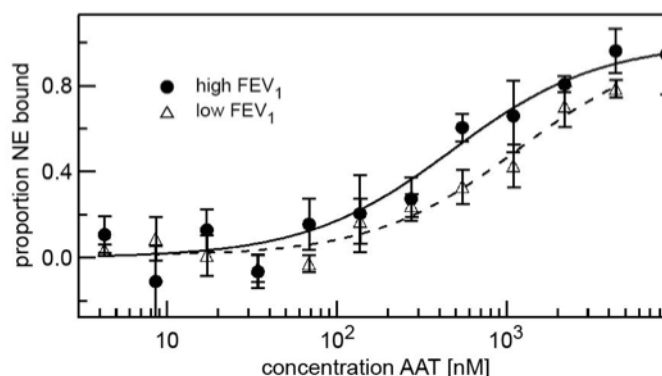
Figure 13: Plasma-dependent change in affinity of NE-AAT binding.
The dissociation constants of the interaction between AAT and NE in two plasma pools from individuals with high FEV1 ($\geq$ 80%, n = 8) and low FEV1 ($\leq$ 50%, n = 12) were compared. The affinity between Z-variant of AAT and NE was higher in plasma with high FEV1 ($\geq$ 80%) ($K_D^{EI_1}$ = 500 $\pm$ 100 nM) compared to the affinity in low FEV1 ($\leq$ 50%) ($K_D^{EI_1}$ = 1300 $\pm$ 250 nM). The presented binding curves represent example measurements where each measurement point (mean $\pm$ S.D.) was averaged from three technical replicates. The fitted binding curves and $K_D^{EI_1}$ values (mean $\pm$ S.D.) were derived from global fitting of four measurements (three independent protein expressions). The measurements were performed in 7.5% plasma and with 5 nM of labeled NE.

concentration and affinity, it was reasoned that $K_D^{EI_1}$ of the Z-variant is in fact varied between the plasma samples.

7.4   Sensitivity of Z-variant AAT to plasma environment

To investigate whether affinity is indeed sensitive to a given plasma environment, the $K_D^{EI_1}$ value of the AAT Z-variant binding to NE in two AAT-deficient plasma pools from persons with high ($\geq$ 80%) and low FEV1 ($\leq$ 50%) values was determined (Figure 13).

To facilitate the recombinant production of the aggregation-prone Z-variant, three stabilizing mutations derived from the work of [46] – Met374Ile, Ser381Ala, Lys387Arg – were introduced into the AAT sequence. As a control, AAT without the Z-mutation but containing the same stabilizing mutations was also expressed.

The affinity of the Z-variant to NE was more than two times better in the plasma pool from individuals with high FEV1 ($\geq$ 80%) ($K_D^{EI_1}$ = 500 $\pm$ 100 nM) than from individuals with low FEV1 ($\leq$ 50%) ($K_D^{EI_1}$ = 1300 $\pm$ 250 nM), confirming the plasma-dependent change in the affinity between NE and Z-variant AAT as observed with the developed assay.

Additionally, it was shown that the Z-mutation improves the encounter reaction of AAT with NE (Figure 14a) and Z-variant AAT is more susceptible to plasma components than control AAT (compare Figure 13 and Figure 14b).
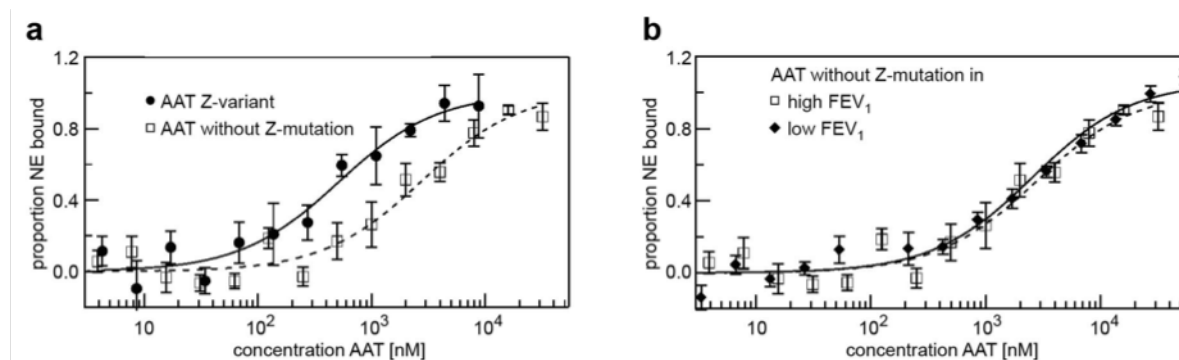
Figure 14: Z-variant AAT is more susceptible to plasma-dependent changes than AAT without Z-mutation.

(a) In high FEV1 ($\geq$ 80%) plasma pool, affinity between the stabilized Z-variant of AAT and NE ($K_D^{EI_1}$ = 500 $\pm$ 100 nM) was approximately five times better than between the stabilized AAT without Z-mutation and NE ($K_D^{EI_1}$ = 2760 $\pm$ 550 nM). The fitted binding curves and $K_D^{EI_1}$ values (mean $\pm$ S.D.) were derived from global fitting of four measurements (three protein expressions) with Z-variant of AAT and three measurements (three protein expressions) with AAT without Z-mutation. (b) There was no significant difference between the affinity of the stabilized AAT to NE in high FEV1 ($\geq$ 80%) ($K_D^{EI_1}$ = 2760 $\pm$ 550 nM) and low FEV1 ($\leq$ 50%) ($K_D^{EI_1}$ = 2540 $\pm$ 300 nM) plasma environments. The fitted binding curves and $K_D^{EI_1}$ values (mean $\pm$ S.D.) were derived from global fitting of three measurements (three protein expressions) in high FEV1 plasma and four measurements (two protein expressions) in low FEV1 plasma. The measurements were performed in 7.5% plasma and with 5 nM NE. The presented binding curves represent example measurements where each measurement point (mean $\pm$ S.D.) was averaged from three technical replicates.

8   Discussion and Outlook

In this project, I worked on developing an assay for analyzing concentration and affinity effects of a disease-relevant molecule present directly in human plasma samples. The assay is based on the biophysical technology of microscale thermophoresis that allows to separate molecules based on their thermophoretic mobility directly in human plasma solution. Analysis performed directly in the plasma sample allowed to take into account effects of different plasma components on the affinity of the interaction between the molecules of interest.

By incorporating a high affinity ligand in thermophoretic measurements, this work expanded the spectrum of applications towards the low affinity range where e.g. enzyme-substrate interactions take place. This new approach allowed to detect unexpected plasma-dependent interferences with AAT. These interferences suggest that the functioning of the Z-variant AAT depends on various plasma components that can modulate the affinity of AAT towards target enzymes. For example, it was reported [26, 87] that lipoproteins can bind to wild-type AAT, thereby influencing its inhibition of NE. Thus, lipoproteins present in the plasma environment could be suspects for affecting the reactivity of the Z-variant AAT.
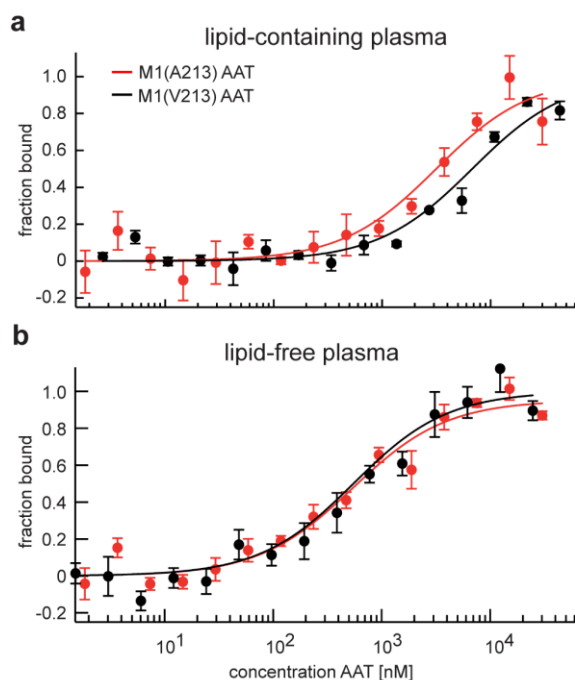
Figure 15: Functional differences between two M1 AAT variants in lipid-containing *vs* in lipid-free plasma.
Adopted from [53] (a) In the presence of lipids, the affinity NE binding to M1(A213) AAT (K_D = 3200 ± 50 nM) was substantially higher than the affinity of NE binding to M1(V213) AAT (K_D = 6800 ± 1000 nM). (b) In the absence of lipids, there was no observable difference between the affinity of NE binding to M1(A213) AAT (K_D = 540 ± 50 nM) or NE binding to M1(V213) AAT (K_D = 550 ± 60 nM). Fitted binding curves and K_D values (mean ± S.D.) were derived from global fitting of three measurements (three independent protein expressions). The measurements were performed in 7.5% plasma and with 5 nM NE. Presented binding curves represent example measurements where each measurement point (mean ± S.D.) was derived from three technical replicates.

Possible explanation for variability of symptoms in ZZ carriers   The analysis of the interaction between the Z-variant AAT and NE in plasma with the introduced assay provides a possible explanation of the variability in disease development in ZZ carriers. Normally, the reduced AAT plasma levels in ZZ carriers are compensated by the enhanced affinity of Z-variant AAT to NE (Figure 14a). However, the enhanced affinity between Z-variant AAT and NE can be disrupted by plasma factors (Figure 13), especially in individuals who have higher concentrations of the interfering plasma components.

Extending the developed technique to a different project   I took part in a different project where I measured the dissociation constant of AAT-NE binding in lipid-containing plasma *vs* in lipid-free plasma [53] for two other variants of AAT, the minor M1(A213) allele and the prevalent M1(V213) allele of AAT. The co-authors on this study found that M1(A213) AAT associates with large artery atherosclerotic stroke condition, however, no functional differences between the two alleles were found in the previous studies

where the affinity was tested in simple buffer systems [60]. For testing whether there is a functional difference between the two alleles interacting with NE within a complex matrix of plasma, the lipid-free plasma and proteins were prepared by my collaborator, Therese Dau [53], and the MST binding assay was performed as explained in Section 7.1. I saw that the M1(A213) variant had an almost two-fold higher affinity towards NE ($K_D = 3200 \pm 50$ nM) compared to M1(V213) variant ($K_D = 6800 \pm 1000$ nM) in the lipid-containing plasma (Figure 15a). The difference was not detectable in the lipid-free plasma (Figure 15b) with K_D values of (540 $\pm$ 50) nM and (550 $\pm$ 60) nM for M1(A213) and M1(V213) variants, respectively. The affinity of interaction between NE and AAT was enhanced for both variants in the lipid-free plasma. This experiment confirmed that plasma components play a crucial role when studying intermolecular interactions. This result also supports the idea that the lipoproteins present in plasma could affect the affinity of the Z-variant AAT to NE in the context of the AAT-deficiency disorder.

The results of the work presented in this part of the Thesis underscore the importance of the personally unique and distinct endogenous environments where enzymes interact with their substrates in a complex web of other components including proteins and lipids. The effects of such highly personalized environments have often been overlooked due to the absence of appropriate techniques. Therefore, the approach developed in this work may open a door to study how the components of native environments where interactions take place interfere with known biochemical reactions and could allow to discover new surprising elements in these reactions.

Part III

Stereochemical theory of the origin of the genetic code

9 Motivation and Introduction

The genetic code   The genetic code describes a set of rules that define which amino acids are encoded by which codons (sequences of three nucleobases) during the translation of RNA into proteins in the cells. It is, therefore, an essential mechanism of all life. The genetic code defines the mapping of 64 codons to 20 naturally occurring amino acids (Figure 16a). The code is highly conserved and is shared with only minor modifications by all known life forms on Earth [19].

The genetic code is highly nonrandom [20]. Specifically, the codons are clustered based on the properties of the amino acids. Amino acids that share a biosynthesis pathway tent to have the same first letter of the code [88]. Amino acids of the shikimate, pyruvate, aspartate, and glutamate pathway families tend to have at the first positions of their codons bases U, G, A, and C, respectively. This connection to the synthesis pathway could indicate that a more simple code with less amino acids had existed first and was extended to a larger set of amino acids by assigning new amino acids to the codons that were similar to the codons of the precursor amino acids.
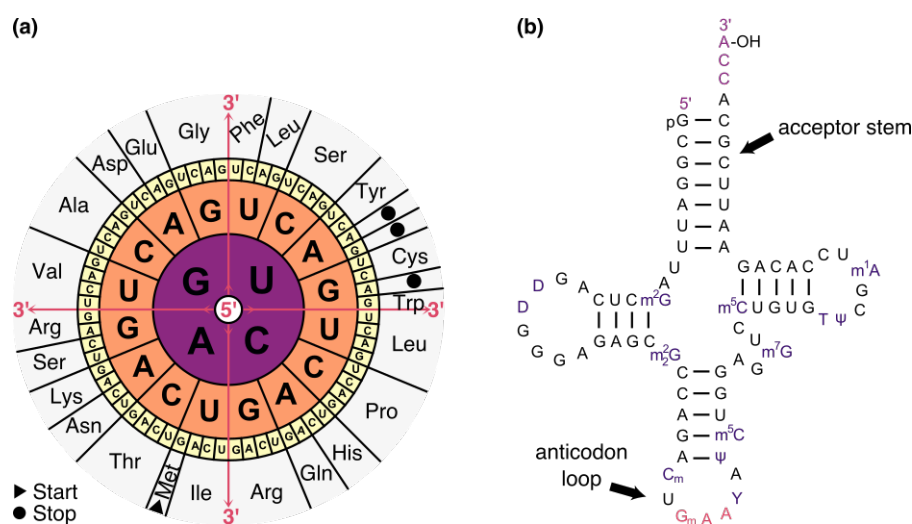

Figure 16: The genetic code and the tRNA.
(a) One of the representations of the genetic code. (b) The cloverleaf representation of the secondary structure of the tRNA molecule (modified from https://en.wikipedia.org/wiki/Transfer_RNA). "m" stands for methylation, "D" stands for 5, 6-Dihydrouridine, "Ψ" stands for pseudouridine, "Y" stands for wybutosine, and "p" stands for a phosphate group.

Amino acids of similar physiochemical properties, such as polarity and the structure of the side chain, also have similar codons [8, 12]. This property of the genetic code could help to minimize the deleterious effects of mistranslations and point mutations on the outcome of protein synthesis. The overall code in general was shown to be optimized to minimize the likelihood of mistranslations and point mutations [21].

Another perplexing property of the genetic code is a negative correlation between the molecular weight of the amino acid and the number of cognate codons encoding the amino acid [11]. This suggests

that early amino acids, which had a chance to establish the largest codon plurality, were simple and, thus, had side chains of smaller molecular weights.

Amino-acyl tRNA synthetases   In modern cells, the assignment of amino acids to their codons is executed by an enzyme, amino-acyl tRNA synthetase (aaRS) in several steps. aaRS first binds ATP and the corresponding amino acid (or its precursor) to form an aminoacyl-AMP (amino acid bound to AMP, or aa-AMP), releasing the inorganic pyrophosphate (PPi). The complex of aaRS and aa-AMP then binds the appropriate transfer RNA molecule (tRNA, Figure 16b) and the amino acid is transferred from the aa-AMP to either the 2'- or the 3'-OH of the last tRNA nucleotide (A76) at the 3'-end.

1. amino acid + ATP → aa-AMP + PPi

2. aa-AMP + tRNA → aa-tRNA + AMP

aaRS is a complex enzyme that recognizes the correct tRNA based on several tRNA sequence and structure signatures, including the codon loop and the acceptor stem of the tRNA. Interestingly, in some cases, the codon loop is not important for a correct charging of the tRNA by a corresponding aaRS (reviewed in [74]). A complex recognition process between tRNAs and aaRSs could only exist once a reliable copying of ~ 70 bases long RNA strands have evolved, which, as discussed in Part I of this Thesis, is not a trivial evolutionary invention. Therefore, the complex system of coded protein synthesis must have evolved from a more primitive version.

Theories on the origin of the genetic code   So how did the system of coded protein synthesis evolve, or what is the origin of the genetic code? An idea that the genetic code was the result of a "frozen" accident was introduced by Crick in 1968 [6]. In his hypothesis, the early assignments of tRNA-like ribozymes and amino acids were established randomly. Once functional peptides could have been produced from amino acids encoded in this early genetic code, any random change would have been lethal to the organism. Thus, the code became "frozen". However, the high degree of structure found in the genetic code made researchers suggest more function-related theories of the origin of the genetic code, including the adaptive theory, the coevolution theory, and the stereochemical theory.

The adaptive theory states that the code was developed under the selective pressure that made the code optimized against translational and mutational errors [21]. The coevolution theory postulates that the code's structure reflects the amino acids' biosynthesis pathways, and evolved in parallel with them [99]. Both theories do not deal with the question, how the very first assignments between primordial amino acids and their codons were established. This gap is covered by the stereochemical theory that posits that specific assignments were established based on the specific stereochemical interactions between amino acids and their cognate codons [30, 103]. All three theories do not have to be exclusive but could hold true for different amino acids' assignments and periods in evolution. The early amino acids could have been assigned based on their stereochemical properties, while the newer additions to the genetic code could have been assigned based on the similarity of their biochemical pathway to already existing amino acids.

The code could have additionally been checked by natural selection to be robust against mistranslations and point mutations. Especially the last mechanism could have modified the early "original" genetic code hiding its early signatures.

In this project, I focused on the stereochemical theory of the origin of the genetic code. The idea behind the stereochemical theory is to consider the early translation mechanism as being enzyme-free, and thus to remove aaRS as a player from the story. The question is then, could have a primordial tRNA itself executed the function of recognizing the correct amino acid? And could have this recognition been mediated by the intrinsic stereochemical interactions between the amino acid and the codon in the sequence of this primordial tRNA?

So far, the stereochemical theory was mainly tested in numerous SELEX experiments, where oligonucleotides were selected to have an amino acid specific binding affinity. Selected aptamers showed enrichment in amino acid specific codons/anticodons [101]. However, SELEX experiments suffer from sequence biases and accumulate PCR artifacts; additionally, significance of the enrichment depends on the applied statistical analysis showing the weaknesses of the experimental approach (reviewed in [44]). A more direct approach would be to measure experimentally the binding affinity between the codons and amino acids.

Project goals   The goal of this project was to device new experimental systems for testing the stereochemical theory of the origin of the genetic code and to lay the groundwork for future research on the topic. This was done with two approaches.

1. In one approach, an AMP-binding RNA aptamer was chosen and used as a testbed to measure binding affinities of amino acids modified with AMP to the aptamer modified to contain different codons/anticodons within its sequence.

2. In the other approach, two hypotheses for the structure of the primordial tRNA were taken from literature (Hopfield and Rodin hypotheses [30, 68]) and used to test their binding affinities to amino acids modified with AMP.

Primordial amino acids   Not all 20 naturally occurring amino acids were likely a part of the early genetic code. Pudritz and Higgs identified amino acids glycine (Gly), alanine (Ala), aspartate (Asp), glutamate (Glu), valine (Val), serine (Ser), isoleucine (Ile), leucine (Leu), and proline (Pro) as prebiotically plausible by systematizing various experiments on prebiotic amino acid syntheses, including hydrothermal synthesis, synthesis in icy grains, amino acids brought my meteorites, atmospheric synthesis, and other types of synthesis [29]. To these 9 amino acids Arg was added. Arg in SELEX experiments generated aptamers often enriched in Arg codons [102]. Also Arg is a highly charged amino acid, which makes it plausible to expect Arg to build non-covalent bonds with nucleic acids.
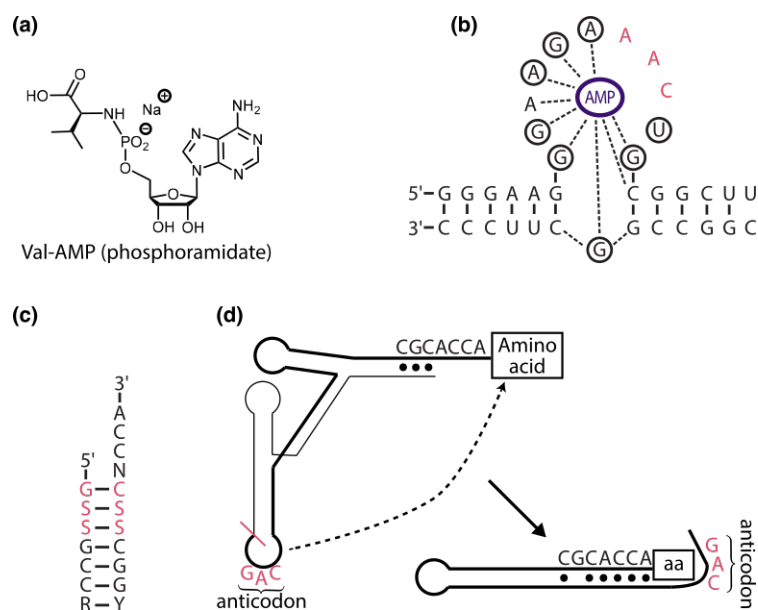
Figure 17: Structures used for analyzing binding affinities.
(a) Stable phosphoramidate form of an activated amino acid. Val-AMP is shown as an example. (b) Secondary structure of the AMP-binding RNA aptamer used for affinity measurements in this project. Bases essential for AMP binding are circled. Dashed lines indicate hydrogen bonds that coordinate AMP binding. In pink are the non-essential for binding bases that were modified to be codons/anticodons for different amino acids. (c) Rodin type pre-tRNA structure. In pink are the positions that Rodin considered to be coding for amino acids in the early days of coded translation. (d) Hopfied's hypothesis of how the pre-tRNA might have looked like. Just a few mutations and cutting the modern tRNA along the red line lead to refolding of a modern tRNA such, that the anticodon and the amino acid are brought close in space.

aminoacyl-AMPs As discussed above, aaRSs use aa-AMP molecules when charging tRNAs with amino acids, making aa-AMP an important intermediate. aa-AMPs are, however, highly unstable in aqueous environments being subjected to fast hydrolysis. aaRSs reaction center shields the aa-AMP intermediate from water and thus stabilizes it against hydrolysis. On the watery early Earth, on the other hand, the aa-AMP intermediates would have been subjected to fast hydrolysis. Whether or not aa-AMP were intermediates in the primordial translation system is debatable [50]. However, different stabilized versions of aa-AMPs exist. I used a stable phosphoramidate version of the AMP-activated amino acids, where the AMP moiety is attached to the N-end of the amino acid instead of to the C-end. This construct occurred as a natural product in an EDC-assisted polymerization reaction performed by Richert's group [36].

AMP-binding RNA aptamer To test the influence of the amino acid properties on the codon preference, literature was searched for a known AMP-binding RNA aptamer. The idea behind the search was the following: The aptamer should stably bind the AMP moiety of the aa-AMP. This would allow to test the preference of the amino acid part of the aa-AMP for codons introduced into the sequence of the

aptamer by mutating nucleotides that are nonessential for AMP binding.

The AMP-binding RNA aptamer that has an adenosine binding motif of 11nt in length was selected from literature [13, 72]. This motif was reproducibly isolated in different experiments starting from random sequence pools [71], suggesting it is a rather simple binding structure. Additionally, several aptamers with the same 11nt-long adenosine-binding motif were identified in humans in functional structures that sense ATP concentration [95]. Therefore, it is plausible that the motif could have evolved as an adenosine-binding RNA motif early in evolution.

Primordial tRNA hypotheses   For testing binding affinities to hypothetical pre-tRNA structures, two hypothesis were selected from literature. Rodin observed that the two tRNAs with complementary anticodons in the anticodon loops tend to have complementary second bases in the acceptor stems [67, 69]. He hypothesized that the code evolved in complementary pairs and was originally encoded in the 1-2-3 and 70-71-72 bases of the acceptor stem. The anticodon present in the anticodon loop of the modern tRNA has evolved later together with the tRNA evolution. The idea that the genetic code was first encoded in the acceptor stem was also suggested by the earlier experiments with the truncated tRNAs and truncated aaRSs. tRNAs truncated to acceptor mini-helix could be correctly charged with the corresponding amino acids. aaRSs truncated so that they couldn't probe the anticodon loop of the tRNA, still could place an amino acid on the correct tRNA with the same rate as before [75]. From the phylogenetic trees for Bacteria, Archaea, and Eukarya, Rodin reconstructed the common ancestor of the acceptor stem [67], which is a pre-tRNA of Rodin's type (Figure 17c). The coding bases according to Rodin's hypothesis are highlighted in red and are located in close proximity to the NCCA-3′ tail, where amino acid is supposed to attach. The pre-tRNA of Rodin's type was experimentally analyzed in this project.

Another pre-tRNA structure was proposed by Hopfield in 1978 [30]. He noticed that just a couple of mutations to the sequence of the modern tRNA leads to refolding of the tRNA into a Hopfield's pre-tRNA, where the acceptor stem and the anticodon loop are brought into close proximity. Therefore, the amino acid binding site appears close enough to the anticodon loop to allow to imagine that the amino acid could have been preselected by noncovalent binding to the anticodon sequences, followed by the attachment to the 3′-end of the acceptor stem. This way aaRSs only became important for recognition later, once the tRNA structure evolved to a modern version with the anticodon loop and the acceptor stem far apart.

10   Experimental Realization

10.1   Materials and Methods

Materials

Buffers   All experiments were performed in 20 mM Tris at pH 7.6 (adjusted with NaOH) supplemented with 300 mM NaCl and 5 mM $MgCl_2$. Nuclease-free components were purchased from Ambion. Only nuclease-free water from Ambion was used in all experiments.

DNA oligonucleotides   All DNA oligonucleotides were purchased from biomers.net GmbH with HPLC purification. Aptamer sequence with Val codon in bold: Cy5 - 5′ - GGG AAG GGA AGA **GUC** UGC GGC UUC GGC CGG CUU CCC. Two strands required for Rodin's pre-tRNA with Gly and Pro codons in bold: 5′ - **GGG** GCC G-3′ and Cy5 - 5′ - CGG **CCC C**AC CA - 3′. Two strands required for Hopfield's-like pre-tRNA with Gly codon in bold: 5′ - **GGG** UGG GGG GGG - 3′ and Cy5 - 5′ - CCC CCC CCA - 3′.

Amino acids and AMP   L-Val, L-Gly, and AMP were purchased from Sigma. Val-AMP and Gly-AMP were synthesized and kindly provided by the lab of Prof. Clemens Richert.

Microscale Thermophoresis Experiments

MST was introduced in detail in Sections 5.1-5.3. Here, only the details to the experiments performed in Part III are described.

AMP, amino acid, or aa-AMP (ligands) were serially diluted (1:1) over five orders of magnitude in the buffer. Separately, a fluorescently labeled RNA structure was diluted to the final concentration of 10 nM in the the same buffer substituted with 0.2% of Tween 20 and mixed 1:1 with the serial dilution of the ligand. The samples were measured at 22°C in the instrument using 10% light-emitting diode power and various infrared laser (IR) powers (mentioned in the caption to each figure containing the experimental data) with IR laser on/off times of 25/5 seconds.

To obtain binding curves, the fluorescence after the temperature jump and equilibrated thermophoresis was normalized to the fluorescence before the IR laser heating, yielding fluorescence depletion values. Fluorescence depletion values of three technical replicates (mean ± S.D.) were plotted in per mille units (‰) on a linear y-axis against the concentration of the serially diluted ligand on the log10 x-axis resulting in binding curves. In some cases where the baseline for the bound state could be clearly observed, the depletion values were additionally normalized to 0 at the lowest ligand concentration and to 1 at the highest ligand concentration. The fit to equation 49 (Part II) was performed to yield the dissociation constant of each curve.
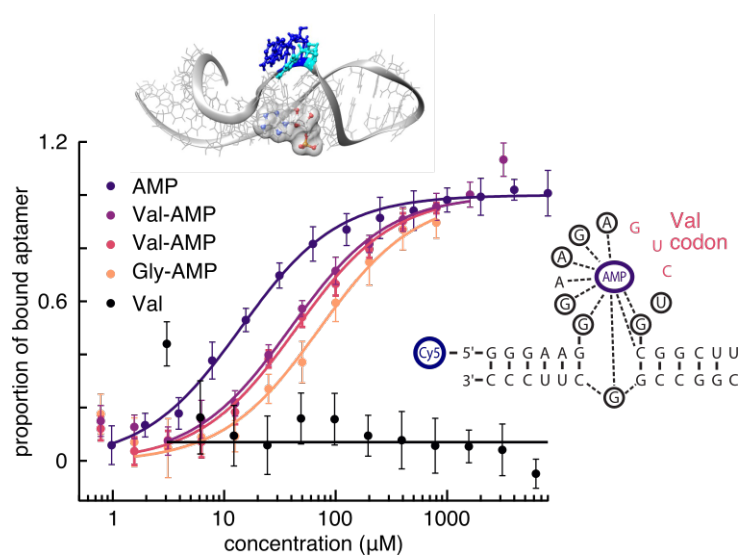
Figure 18: Preliminary affinity measurements between the AMP-binding aptamer and AMP, Gly, or aa-AMP ligands.
The crystal structure of the original AMP-binding RNA aptamer with AMP in the binding pocket (PDB: 1raw) in shown on the top. The bases colored in blue are the ones that were mutated. The graph displays the binding curves from the MST experiments (with 40% of the IR laser power). AMP, Val-AMP (two different synthesis batches), Gly-AMP, or Val were titrated against the constant concentration of the aptamer. For the experiment, the original aptamer was modified to include the Val codon (pink letters) and was labeled on the 5′-end with Cy5 dye. The error bars are standard deviations from three technical replicates of each experiment. Two experiments with two different synthesis batches of Val-AMP highlight the reproducibility of the MST approach.

10.2 Results and Discussion

A simple aptamer as a testbed

To make a first experimental step towards testing whether there is an intrinsic affinity between the codons/anticodons and the corresponding primordially plausible amino acids, the AMP-binding RNA aptamer with Cy5 dye at the 5′-end was synthesized for MST experiments. The three bases unimportant for AMP binding were modified to match the Val codon "GUC" (Figure 18, pink bases in the structure on the right). Before the experiments, the aptamer was melted at 95ºC for 2 minutes and allowed to slowly cool down to adapt the lowest energy conformation.

AMP (8 mM final highest concentration) or Val (100 mM final highest concentration) were titrated against the constant concentration of labeled aptamer (10 nM final concentration). The dissociation constant of 14 ± 3 μM was obtained for AMP binding, while Val showed no detectable binding (Figure 18). This result was expected as, according to our hypothesis, amino acids by themselves do not have a high enough affinity for binding. However, aa-AMPs can first bind the structure due to the presence of AMP, and then the amino acid part can build additional non-covalent bonds with the aptamer. The idea is that if there is an enhanced affinity of amino acids to their codons/ anticodons, the affinity of aa-AMP
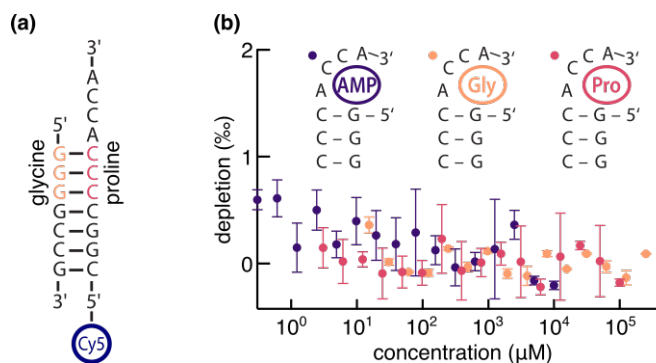
Figure 19: Preliminary tests of the Rodin's hypothesis.
(a) Selected Rodin type pre-tRNA with Gly and Pro codons at the coding positions. One of the strands was labeled with Cy5 dye at the position far away from the expected binding site. (b) The binding curves from the MST experiments (with 60% of the IR laser power). AMP, Gly, or Val were titrated against the constant concentration of the Rodin's pre-tRNA (10nM). The error bars are standard deviations from three technical replicates of each experiment.

to the aptamer with the corresponding codon/anticodon would be stronger than to the aptamer with any other codon/ anticodon.

To try this idea, Val-AMP (0.8 mM or 3.2 mM final highest concentration, two different Val-AMP syntheses) or Gly-AMP (0.8 mM final highest concentration) were titrated against the constant concentration of labeled aptamer (10 nM final concentration). The average dissociation constant of 42 ± 4 μM was obtained for Val-AMP binding, while Gly-AMP bound with the the dissociation constant of 71 ± 8 μM (Figure 18). Two binding curves for Val-AMP binding from two different synthesis batches of Val-AMP are shown to emphasize the reproducibility of the synthesis quality and the affinity measurement method. Here, Val-AMP bound indeed with the higher affinity than Gly-AMP, in correspondence with the Val codon "GUC" and not Gly codon "GGC" present in the aptamer sequence.

This proof of principle experiment suggests that the high affinity AMP-binding RNA aptamer could be a good system to start testing for signatures of the genetic code. A high-throughput approach to affinity measurements has to be implemented to be able to screen the binding between all the 10 selected amino acids (Section 9) to all the possible aptamers with various corresponding codons/anticodons. Especially considering that, due to the degeneration of the genetic code, more than 20 codons/anticodons correspond to the 10 selected amino acids. Also, to decrease the costs for the synthesis of labeled RNA strands, instead of labeling the aptamer strand itself, the 3′- or the 5′-end of the aptamer can be elongated to allow binding of a short labeled complementary oligonucleotide. This short oligonucleotide will be the only required labeling strand for all the aptamers with various codons/anticodons used for screening.

Rodin's hypothesis

To test Rodin's hypothesis of the pre-tRNA structure, Rodin's pre-tRNA with codons for Gly and respectively Pro was synthesized. One of the strands was labeled with Cy5 for detection (Figure 19a). The
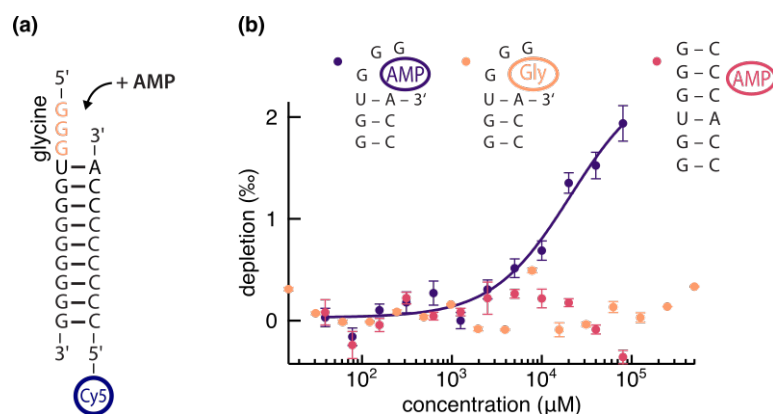
Figure 20: Preliminary tests of the Hopfield's hypothesis.
(a) The RNA structure with an overhang of three bases at the 5′-end and the CCA sequence at the 3′-end of the stem overlooking the overhang. The structure was used as an approximation to the Hopfield's pre-tRNA. One of the strands was labeled with Cy5 dye at the position far away from the expected binding site. (b) The binding curves from the MST experiments (with 80% and 60% of the IR laser power for binding to AMP and Gly, respectively). AMP or Gly were titrated against the constant concentration of the RNA structure from (a). Or AMP was titrated against the fully double-stranded RNA fragment with "C"s at the positions opposite of "G"s in the overhang. The measured dissociation constant for AMP to Hopfield's-like pre-tRNA binding was quantified to be 24 ± 4 mM. The error bars are standard deviations from three technical replicates of each experiment.

strands were mixed together at an equimolar concentrations and allowed to slowly anneal. The binding of such a structure (at 10 nM final concentration) to Gly or Pro was tested with titrated amino acids reaching final highest concentrations of 100 mM for Pro and 500 mM for Gly. No detectable binding was observed (Figure 19b).

Binding to the aa-AMP molecules should be tested next. However, as AMP by itself did not show any binding to the Rodin's pre-tRNA (Figure 19b), the expectation is that aa-AMPs will not show any binding either. According to our hypothesis, AMP is the guiding moiety of aa-AMP, which brings an amino acid to its potential binding site. If the amino acid moiety corresponds to the codon of the pre-tRNA structure, the affinity of aa-AMP could be enhanced compared to the binding of AMP alone. But if AMP does not bind, then aa-AMP would probably also show no binding.

Hopfield's hypothesis

For preliminary tests of Hopfield's hypothesis of the pre-tRNA structure, Hopfield's pre-tRNA-like structure with the codon for Gly was synthesized (Figure 20a). The codon was placed in the 3-bases overhang at the 5′-end of one of the strands. The other strand was labeled with Cy5 for detection at the position far away from the codon. The CCA sequence on the labeled strand corresponded to the CCA 3′-end tail of the modern tRNA.

MST experiments for the RNA structure and Gly alone didn't reveal any significant binding (Figure 20b). Binding to AMP provided a more interesting case. AMP bound to the overhang with the dissocia-

tion constant of 24 ± 4 mM in the presented experiment (Figure 20b) but affinities measured in repeats of the experiment ranged between 2 and 25 mM. The disperse value is probably the result of the difficulty of measuring dissociation constants of low affine interactions. Significantly, there was no binding between AMP and a fully double-stranded RNA fragment, where each G in the overhang had C to bind to. Thus, the binding of AMP was indeed specific to the overhang.

The question to answer next is whether the AMP-activated amino acids would bind to the overhang with an enhanced affinity in case of the overhang matching the codon/anticodon of the tested amino acid. The expectation is that the AMP part of the AMP-activated amino acid would serve as a guide for an amino acid, bringing it close to the overhang. Then the amino acid, by interacting with the overhang, could enhance the affinity of the aa-AMP binding.

11   Outlook

The results of this project provide a starting point for different screening approaches in order to test whether the prebiotically plausible amino acids have an intrinsic affinity towards their cognate codons/ anticodons. This idea is the main thesis of the stereochemical theory of the origin of the genetic code. For such tests, two components are of utmost importance: the RNA structures and activated amino acids.

One type of a stable activated amino acid, produced in experiments mimicking prebiotic Earth conditions, is known from the works of the laboratory of Clemens Richert [36]. They obtain activated amino acids, where an amino acid is attached to the AMP via its N-end and not via its C-end as in the aaRS reaction center. Activated amino acids in the experiments of Richert's group form directly in EDC-driven polymerization reactions starting from mixtures of amino acids, RNA monomers, and short RNA strands

A way to stabilize an AMP-activated amino acid molecule, where an amino acid is attached to AMP in a canonic orientation via its C-end, is to exchange the oxygen atom that bridges an AMP moiety and an amino acid with either CH_2 or NH. CH_2 substitution is likely to result in a very stable activated amino acid analog. However, two hydrogen atoms could result in a substantial steric hindrance when binding to RNA motifs. Thus, the O to NH substitution is preferred. The synthesis protocol for activated amino acids with an N-acyl phosphoramidate linkage is known from literature [58].

For the case of using the AMP-binding RNA aptamer as a testbed for affinity correlations, the following general directions could be explored. First of all, a high-throughput screening has to be conducted. For this, aptamers with codon/ anticodon sequences corresponding to the prebiotically plausible amino acids should be synthesized. Synthesizing each aptamer with a fluorescent label is expensive. Thus, an aptamer with an elongated 3′- or 5′-end should be synthesized, to which a short labeled oligonucleotide can stably bind, enabling fluorescent labeling of all the aptamers with just a single labeled RNA strand.

Analysis of the available crystal structure of AMP in the aptamer's binding pocket [13] can foster the use of the aptamer as a testbed for affinity interactions. For example, in figure 21, the aptamer is shown with AMP (in grey) and Val-AMP from Richert's group (purple) aligned such that the adenine part of Val-AMP maximally overlaps with the adenine of AMP co-crystallized with the aptamer. Such an overlap was chosen because the aptamer builds interactions with the adenine part and it is expected that Val-AMP would also be stabilized in the binding pocket via a similar adenine binding.

For some preliminary analysis, I highlighted in turquoise the bases essential for AMP binding. These are the bases that shouldn't be modified in the screening procedure. The bases in dark blue are the bases unimportant for AMP binding and they were the ones modified in the preliminary tests presented in this Thesis. However, these bases are positioned quite far away from the amino acid part of the Val-AMP. The question arises whether the aa-AMP can "flip" inside the binding pocket, using the flexible link between the ribose and adenine, such that the amino acid part comes into closer contact with the bases highlighted in the dark blue color. Since I saw significant modulations in the dissociation constant depending on the presence of the amino acid moiety (Figure 18), the expectation is that these bases could affect the binding.

On the other hand, the bases highlighted in pink are not important for AMP binding. They are
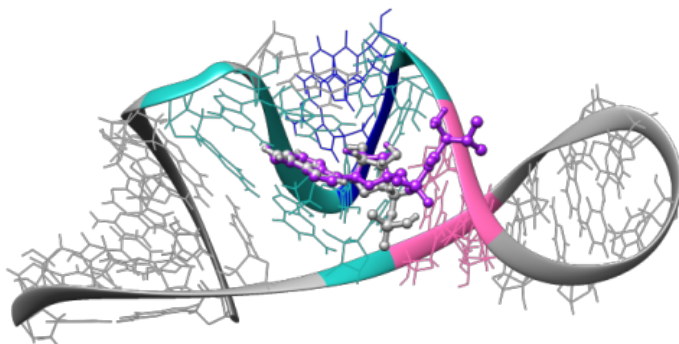
Figure 21: AMP and Val-AMP in the binding pocket of the AMP-binding RNA aptamer.
The crystal structure of the AMP-binding RNA aptamer with AMP (grey) in the binding pocket (PDB: 1raw) and Val-AMP (purple) aligned such that the adenine moieties of the two molecules are maximally overlapping. In turquoise are the bases essential for AMP-binding. In dark blue are the bases modified to Val codon sequence in the preliminary tests. In pink are the bases in close proximity to the Val part of the Val-AMP.

building the double stranded region of the aptamer that probably gives stability to the binding pocket. A possibility to modify the bases in this region to match codon/ anticodon sequences can also be explored, as this region is in close proximity to the amino acid part of the Val-AMP without any "flipping" involved.

Though the tests of the Rodin's hypothesis with AMP and amino acids were so far negative, one should explore the possibility that aa-AMPs could show preferential binding to stems containing codon/ anticodon sequences. The Hopfield's pre-tRNA structure showed binding to AMP, albeit with low affinity. That makes it a promising instrument for checking whether the affinity could be enhanced when the cognate amino acid is attached to AMP.

Another possibility for the experimental setting is to test binding of the discussed structures to small polypeptides rather than to individual amino acids. Richert's work shows that polypeptides are readily formed when the first amino acid is capture by AMP via its N-end [28]. Thus, AMP-polypeptides could be prebiotically plausible substrate for early translation mechanisms.

One should also not forget to test different buffer conditions, mono- and divalent ions dependence, temperature dependence, and other environmental parameters.

A fascinating outlook for the project would be to test a library of RNA sequences forming potentially enzymatic tetraloop-like structures of the Rodin's or Hopfield's type [82] for aminoacylation activity starting with aa-AMP substrates. The plausibility that small RNA structures can support aminoacylation reactions is based on the work by Yarus's group, where they isolated a 5-nt long ribozyme capable of *trans*-phenylalanylating a partially complementary 4-nt long substrate on its terminal 2'-ribose hydroxyl [91]. Moreover, the sequence of two bases - GU - that shows catalytic activity within the 5-nt long sequence of the ribozyme was found to show aminoacylation activity also within a structure of a long complex ribosomal RNA strand [35].

SELEX experiments for the aminoacylation activity should be combined with their quantitative anal-

ysis with next-generation sequencing platforms. A big advantage of this combination is that by stopping the reaction at different time points, also the kinetics data could be obtained. Using different amino acids and comparing the resulting structures, a common motif may emerge.

In summary, in this part of the Thesis, I tried to establish various models for testing the stereochemical theory of the origin of the genetic code. The models range between the pre-tRNA structures suggested in literature to simply using an aptamer that can bind AMP and, thus, could bind aa-AMP - an important intermediate in aminoacylation reaction. Only the preliminary proof-of-principle experiments were performed to show the feasibility of the approach to measure affinities directly between the activated amino acids and RNA structures with MST as the method of choice. A wide variety of experiments could be conducted from this point on to advance the project to the next stages. Hopefully, in a long term, new experimentally proven patterns of amino acids' associations with their cognate codons/ anticodons could be shown to strengthen or disprove the stereochemical theory of the origin of the genetic code.

Bibliography

[1] James Attwater, Aditya Raguram, Alexey S Morgunov, Edoardo Gianni, and Philipp Holliger. Ribozyme-catalysed RNA synthesis using triplet building blocks. *eLife*, 7, may 2018.

[2] James Attwater, Aniela Wochner, and Philipp Holliger. In-ice evolution of RNA polymerase ribozyme activity. *Nature Chemistry*, 5(12):1011–1018, dec 2013.

[3] Philipp Baaske, Christoph J Wienken, Philipp Reineck, Stefan Duhr, and Dieter Braun. Optical thermophoresis for quantifying the buffer dependence of aptamer binding. *Angewandte Chemie (International ed. in English)*, 49(12):2238–41, mar 2010.

[4] James R. Bielenberg and Howard Brenner. A hydrodynamic/Brownian motion model of thermal diffusion in liquids. *Physica A: Statistical Mechanics and its Applications*, 356(2-4):279–293, oct 2005.

[5] British Thoracic Society. Guidelines for the measurement of respiratory function, mar 1994.

[6] F.H.C. H.C. Crick. The origin of the genetic code. *Journal of Molecular Biology*, 38(3):367–379, dec 1968.

[7] T Dau, E V Edeleva, S A I Seidel, R A Stockley, D Braun, and D E Jenne. Quantitative analysis of protease recognition by inhibitors in plasma using microscale thermophoresis. *Scientific reports*, 6:35413, oct 2016.

[8] Oleg V. Davydov. Amino acid contribution to the genetic code structure: End-atom chemical rules of doublet composition. *Journal of Theoretical Biology*, 193(4):679–690, aug 1998.

[9] Sybren Ruurds De Groot and Peter Mazur. *Non-equilibrium thermodynamics.* Dover Publications, 1984.

[10] Jan K.G. Dhont, S. Wiegand, S. Duhr, and D. Braun. Thermodiffusion of charged colloids: Single-particle diffusion. *Langmuir*, 23(4):1674–1683, 2007.

[11] M Di Giulio. Some aspects of the organization and evolution of the genetic code. *Journal of molecular evolution*, 29(3):191–201, 1989.

[12] Massimo Di Giulio. The extension reached by the minimization of the polarity distances during the evolution of the genetic code. *Journal of Molecular Evolution*, 29(4):288–293, 1989.

[13] T Dieckmann, E Suzuki, G K Nakamura, and J Feigon. Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA (New York, N.Y.)*, 2(7):628–40, jul 1996.

[14] József Dobó and Peter G W Gettins. alpha1-Proteinase inhibitor forms initial non-covalent and final covalent complexes with elastase analogously to other serpin-proteinase pairs, suggesting a common mechanism of inhibition. *The Journal of biological chemistry*, 279(10):9264–9269, mar 2004.

[15] V.F. Drevitch, R.I. Salganik, D.G. Knorre, and E.G. Mal'Ygin. The reaction of DNA with water-soluble carbodiimide and enzyme sensitivity of modified DNA. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 123(1):207–209, jul 1966.

[16] Stefan Duhr and Dieter Braun. Thermophoretic depletion follows boltzmann distribution. *Physical Review Letters*, 96(16), 2006.

[17] Stefan Duhr and Dieter Braun. Why molecules move along a temperature gradient. *Proceedings of the National Academy of Sciences*, 103(52):19678–19682, dec 2006.

[18] Clemens Entzian and Thomas Schubert. Studying small molecule-aptamer interactions using MicroScale Thermophoresis (MST). *Methods*, 97:27–34, mar 2016.

[19] Thomas D Fox. NATURAL VARIATION IN THE GENETIC CODE. Technical report, 1987.

[20] Stephen J Freeland and Laurence D Hurst. The genetic code is one in a million. *Journal of Molecular Evolution*, 47(3):238–248, 1998.

[21] Stephen J Freeland, Tao Wu, and Nick Keulmann. The case for an error minimizing standard genetic code. *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life*, 33(4-5):457–77, oct 2003.

[22] Peter G W Gettins. Serpin structure, mechanism, and function. *Chemical reviews*, 102(12):4751–804, dec 2002.

[23] Clémentine Gibard, Subhendu Bhowmik, Megha Karki, Eun Kyong Kim, and Ramanarayanan Krishnamurthy. Phosphorylation, oligomerization and self-assembly in water under potential pre-biotic conditions. *Nature Chemistry*, 10(2):212–217, 2018.

[24] Walter Gilbert. The RNA world, feb 1986.

[25] B Gooptu, U I Ekeowa, and D A Lomas. Mechanisms of emphysema in alpha1-antitrypsin deficiency: molecular and cellular insights. *The European respiratory journal*, 34(2):475–88, aug 2009.

[26] Scott M Gordon, Benjamin McKenzie, Georgina Kemeh, Maureen Sampson, Shira Perl, Neal S Young, Michael B Fessler, and Alan T Remaley. Rosuvastatin Alters the Proteome of High Density Lipoproteins: Generation of alpha-1-antitrypsin Enriched Particles with Anti-inflammatory Properties. *Molecular & cellular proteomics : MCP*, 14(12):3247–57, dec 2015.

[27] Ferdinand Greiss, Franziska Kriegel, and Dieter Braun. Probing the Cooperativity of Binding Networks with High-Throughput Thermophoresis. *Analytical Chemistry*, 89(4):2592–2597, 2017.

[28] Helmut Griesser, Maren Bechthold, Peter Tremmel, Eric Kervio, and Clemens Richert. Amino Acid-Specific, Ribonucleotide-Promoted Peptide Formation in the Absence of Enzymes. *Angewandte Chemie - International Edition*, 56(5):1224–1228, jan 2017.

[29] Paul G Higgs and Ralph E Pudritz. A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. *Astrobiology*, 9(5):483–90, jun 2009.

[30] J J Hopfield. Origin of the genetic code: a testable hypothesis based on tRNA structure, sequence, and kinetic proofreading. *Proceedings of the National Academy of Sciences of the United States of America*, 75(9):4334–8, sep 1978.

[31] David P Horning and Gerald F Joyce. Amplification of RNA by an RNA polymerase ribozyme. *Proceedings of the National Academy of Sciences*, 113(35):9786–9791, 2016.

[32] Eric D Horowitz, Aaron E Engelhart, Michael C Chen, Kaycee A Quarles, Michael W Smith, David G Lynn, and Nicholas V Hud. Intercalation as a means to suppress cyclization and promote polymerization of base-pairing oligonucleotides in a prebiotic world. *Proceedings of the National Academy of Sciences*, 107(12):5288–5293, mar 2010.

[33] D E Huizenga and J W Szostak. A DNA aptamer that binds adenosine and ATP. *Biochemistry*, 34(2):656–665, 1995.

[34] D C Hutchison. Natural history of alpha-1-protease inhibitor deficiency. *The American journal of medicine*, 84(6A):3–12, jun 1988.

[35] Mali Illangasekare and Michael Yarus. Small aminoacyl transfer centers at GU within a larger RNA. *RNA Biology*, 9(1):59–66, jan 2012.

[36] Mario Jauker, Helmut Griesser, and Clemens Richert. Spontaneous Formation of RNA Strands, Peptidyl RNA, and Cofactors. *Angewandte Chemie (International ed. in English)*, 54(48):14564–9, nov 2015.

[37] Moran Jerabek-Willemsen, Timon André, Randy Wanner, Heide Marie Roth, Stefan Duhr, Philipp Baaske, and Dennis Breitsprecher. MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure*, mar 2014.

[38] Moran Jerabek-Willemsen, Chistoph J Wienken, Dieter Braun, Philipp Baaske, and Stefan Duhr. Molecular interaction studies using microscale thermophoresis. *Assay and drug development technologies*, 9(4):342–53, aug 2011.

[39] Sulay D. Jhaveri, Romy Kirby, Rick Conrad, Emily J. Maglott, Michael Bowser, Robert T. Kennedy, Gary Glick, and Andrew D. Ellington. Designed signaling aptamers that transduce molecular recognition to changes in fluorescence intensity. *Journal of the American Chemical Society*, 122(11):2469–2473, 2000.

[40] Un Beom Kang, Je Hyun Baek, Seung Hyun Ryu, Joon Kim, Myeong Hee Yu, and Cheolju Lee. Kinetic mechanism of protease inhibition by alpha- 1-antitrypsin. *Biochemical and Biophysical Research Communications*, 323:409–415, 2004.

[41] Kunio Kawamura and Fumitaka Okamoto. Condensation reaction of hexanucleotides containing guanine and cytosine with water soluble carbodiimide. *Nucleic Acids Symposium Series*, 44(44):217–218, 2000.

[42] Lyuba Khavrutskii, Joanna Yeh, Olga Timofeeva, Sergey G. Tarasov, Samuel Pritt, Karen Stefanisko, and Nadya Tarasova. Protein Purification-free Method of Binding Affinity Determination by Microscale Thermophoresis. *Journal of Visualized Experiments*, (78):e50541–e50541, aug 2013.

[43] H. G. Khorana and A. R. Todd. 465. Studies on phosphorylation. Part XI. The reaction between carbodi-imides and acid esters of phosphoric acid. A new method for the preparation of pyrophosphates. *Journal of the Chemical Society*, 0(0):2257–2260, jan 1953.

[44] Eugene V Koonin and Artem S Novozhilov. Origin and evolution of the genetic code: the universal enigma. *IUBMB life*, 61(2):99–111, feb 2009.

[45] V S Kraynov, C Chamberlain, G M Bokoch, M A Schwartz, S Slabaugh, and K M Hahn. Localized Rac activation dynamics visualized in living cells. *Science (New York, N.Y.)*, 290(5490):333–7, oct 2000.

[46] K N Lee, S D Park, and M H Yu. Probing the native strain iin alpha1-antitrypsin. *Nature structural biology*, 3(6):497–500, jun 1996.

[47] Tracey A. Lincoln and Gerald F. Joyce. Self-sustained replication of an RNA enzyme. *Science*, 323(5918):1229–1232, 2009.

[48] Pawel Linke, Kwame Amaning, Melanie Maschberger, Francois Vallee, Valerie Steier, Philipp Baaske, Stefan Duhr, Dennis Breitsprecher, and Alexey Rak. An Automated Microscale Thermophoresis Screening Approach for Fragment-Based Lead Discovery. *Journal of biomolecular screening*, 21(4):414–21, apr 2016.

[49] Svenja Lippok, Susanne A I Seidel, Stefan Duhr, Kerstin Uhland, Hans-Peter Holthoff, Dieter Jenne, and Dieter Braun. Direct detection of antibody concentration and affinity in human serum using microscale thermophoresis. *Analytical chemistry*, 84(8):3523–30, apr 2012.

[50] Ziwei Liu, Damien Beaufils, Jean Christophe Rossi, and Robert Pascal. Evolutionary importance of the intramolecular pathways of hydrolysis of phosphate ester mixed anhydrides with amino acids and peptides. *Scientific Reports*, 4:7440, dec 2014.

[51] Carl Friedrich Wilhelm Ludwig. Diffusion zwischen ungleich erwärmten Orten gleich zusammengesetzter Lösungen. *Sitzungsberichte der Akademie der Wissenschaften der mathematisch-naturwissenschaftliche Klasse (Wien)*, 20:539, 1856.

[52] Paul Majkut, Iris Claußnitzer, Helmut Merk, Christian Freund, Christian P R Hackenberger, and Michael Gerrits. Completion of proteomic data sets by Kd measurement using cell-free synthesis of site-specifically labeled proteins. *PLoS ONE*, 8(12), 2013.

[53] Rainer Malik, Therese Dau, Maria Gonik, Anirudh Sivakumar, Daniel J Deredge, Evgeniia V Edeleva, Jessica Götzfried, Sander W van der Laan, Gerard Pasterkamp, Nathalie Beaufort, Susana Seixas, Steve Bevan, Lisa F Lincz, Elizabeth G Holliday, Annette I Burgess, Kristiina Rannikmäe, Jens Minnerup, Jennifer Kriebel, Melanie Waldenberger, Martina Müller-Nurasyid, Peter Lichtner, Danish Saleheen, Peter M Rothwell, Christopher Levi, John Attia, Cathie L M Sudlow, Dieter Braun, Hugh S Markus, Patrick L Wintrode, Klaus Berger, Dieter E Jenne, and Martin Dichgans. Common coding variant in SERPINA1 increases the risk for large artery stroke. *Proceedings of the National Academy of Sciences*, 114(14):3613–3618, mar 2017.

[54] Stefano Mariani and Maria Minunni. Surface plasmon resonance applications in clinical analysis, apr 2014.

[55] Noel G McElvaney. Diagnosing α1-antitrypsin deficiency: How to improve the current algorithm, mar 2015.

[56] Jean-Louis Mergny and Laurent Lacroix. Analysis of Thermal Melting Curves. *Oligonucleotides*, 13(6):515–537, dec 2003.

[57] Marian Mikolajczyk, Piotr Kielbasinski, and Wlodzimierz Basinski. Reaction of carbodiimides with phosphorothioic, phosphorodithioic, and phosphoroselenoic acids: Products, intermediates, and steps. *Journal of Organic Chemistry*, 49(5):899–908, 1984.

[58] T Moriguchi, T Yanagi, M Kunimori, T Wada, and M Sekine. Synthesis and properties of aminoacylamido-AMP: chemical optimization for the construction of an N-acyl phosphoramidate linkage. *The Journal of organic chemistry*, 65(24):8229–38, dec 2000.

[59] Morozov K. I. Thermal diffusion in disperse systems. *Journal of Experimental and Theoretical Physics*, 88(5):944–946, may 1999.

[60] Toshihiro Nukiwa, Mark Brantly, Gerald Fells, Ken Satoh, Larue Stier, Michael Courtney, Ronald G. Crystal, and Fumitaka Ogushi. Characterization of the M1(Ala213) Type of α1-

Antitrypsin, a Newly Recognized, Common "Normal" α1-Antitrypsin Haplotype. *Biochemistry*, 26(17):5259–5267, aug 1987.

[61] L. E. Orgel. Evolution of the genetic apparatus. *Journal of Molecular Biology*, 38(3):381–393, dec 1968.

[62] Thomas E Ouldridge, Petr Šulc, Flavio Romano, Jonathan P.K. Doye, and Ard A Louis. DNA hybridization kinetics: Zippering, internal displacement and sequence dependence. *Nucleic Acids Research*, 41(19):8886–8895, 2013.

[63] A. Parola and R. Piazza. Particle thermophoresis in liquids. In *European Physical Journal E*, volume 15, pages 255–263. Società Italiana di Fisica, nov 2004.

[64] Christine T N Pham. Neutrophil serine proteases: specific regulators of inflammation. *Nature reviews. Immunology*, 6(7):541–50, jul 2006.

[65] J Potempa, E Korzus, and J Travis. The serpin superfamily of proteinase inhibitors: structure, function, and regulation. *The Journal of biological chemistry*, 269(23):15957–60, jun 1994.

[66] Noam Prywes, J. Craig Blain, Francesca Del Frate, and Jack W Szostak. Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides. *eLife*, 5:e17756, 2016.

[67] Andrei S Rodin, Eörs Szathmáry, and Sergei N Rodin. One ancestor for two codes viewed from the perspective of two complementary modes of tRNA aminoacylation. *Biology direct*, 4:4, 2009.

[68] S Rodin, S Ohno, and a Rodin. Transfer RNAs with complementary anticodons: could they reflect early evolution of discriminative genetic code adaptors? *Proceedings of the National Academy of Sciences of the United States of America*, 90(10):4723–7, 1993.

[69] S Rodin, a Rodin, and S Ohno. The presence of codon-anticodon pairs in the acceptor stem of tRNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 93(10):4537–42, 1996.

[70] Joseph. Sambrook, E. F. Fritsch, and Tom. Maniatis. *Molecular cloning : a laboratory manual.* Cold Spring Harbor Laboratory, 1989.

[71] Dayal Saran, Joseph Frank, and Donald H Burke. The tyranny of adenosine recognition among RNA aptamers to coenzyme A. *BMC evolutionary biology*, 3:26, 2003.

[72] M Sassanfar and J W Szostak. An RNA motif that binds ATP. *Nature*, 364(6437):550–3, 1993.

[73] Thomas H Scheuermann, Shae B Padrick, Kevin H Gardner, and Chad A Brautigam. On the acquisition and analysis of microscale thermophoresis data. *Analytical biochemistry*, 496:79–93, mar 2016.

[74] Paul Schimmel and Rebecca Alexander. Diverse RNA substrates for aminoacylation: Clues to origins? *Proc Nat Acad Sci*, 95(September):10351–10353, 1998.

[75] Paul Schimmel, R Giegé, D Moras, and S Yokoyama. An operational RNA code for amino acids and possible relationship to genetic code. *Proceedings of the National Academy of Sciences of the United States of America*, 90(19):8763–8768, 1993.

[76] Martin E Schimpf and Semen N Semenov. Mechanism of polymer thermophoresis in nonaqueous solvents. *Journal of Physical Chemistry B*, 104(42):9935–9942, 2000.

[77] Anneliese Schimpl, Richard M Lemmon, and Melvin Calvin. Cyanamide formation under primitive earth conditions. *Science*, 147(3654):149–150, jan 1965.

[78] Susanne A I Seidel, Patricia M Dijkman, Wendy A Lea, Geert van den Bogaart, Moran Jerabek-Willemsen, Ana Lazic, Jeremiah S Joseph, Prakash Srinivasan, Philipp Baaske, Anton Simeonov, Ilia Katritch, Fernando A Melo, John E Ladbury, Gideon Schreiber, Anthony Watts, Dieter Braun, and Stefan Duhr. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods (San Diego, Calif.)*, 59(3):301–15, mar 2013.

[79] Susanne A I Seidel, Christoph J Wienken, Sandra Geissler, Moran Jerabek-Willemsen, Stefan Duhr, Alwin Reiter, Dirk Trauner, Dieter Braun, and Philipp Baaske. Label-free microscale thermophoresis discriminates sites and affinity of protein-ligand binding. *Angewandte Chemie - International Edition*, 51(42):10656–10659, oct 2012.

[80] Velizar Shivarov, Milena Ivanova, and Elissaveta Naumova. Solid Phase Peptide Synthesis (SPPS). *PLoS ONE*, 9(6):e99769, 2014.

[81] Andrea Sinz. Investigation of protein-protein interactions in living cells by chemical crosslinking and mass spectrometry, aug 2010.

[82] Petr Stadlbauer, Jiri Sponer, Giovanna Costanzo, Ernesto Di Mauro, Samanta Pino, and Judit E. Sponer. Tetraloop-like geometries could form the basis of the catalytic activity of the most ancient ribooligonucleotides. *Chemistry - A European Journal*, 21(9):3596–3604, feb 2015.

[83] Robert A. Stockley and Alice M. Turner. α-1-Antitrypsin deficiency: clinical variability, assessment, and treatment. *Trends in Molecular Medicine*, 20(2):105–115, feb 2014.

[84] Zhifu Sun and Ping Yang. Role of imbalance between neutrophil elastase and alpha 1-antitrypsin in cancer development and progression. *The Lancet. Oncology*, 5(3):182–90, mar 2004.

[85] Syafrizayanti, Christian Betzen, Jörg D Hoheisel, and Damjana Kastelic. Methods for analyzing and quantifying protein-protein interaction, feb 2014.

[86] Fayad Syntia, Reine Nehmé, Bérengère Claude, and Philippe Morin. Human neutrophil elastase inhibition studied by capillary electrophoresis with laser induced fluorescence detection and microscale thermophoresis. *Journal of Chromatography A*, 1431:215–223, jan 2016.

[87] Philippa J Talmud, Steve Martin, George Steiner, David M Flavell, David B Whitehouse, Sylvia Nagl, Richard Jackson, Marja-Riitta Taskinen, M Heikki Frick, Markku S Nieminen, Y Antero Kesäniemi, Amos Pasternack, Steve E Humphries, Mikko Syvänne, and Diabetes Atherosclerosis Intervention Study Investigators. Progression of atherosclerosis is associated with variation in the alpha1-antitrypsin gene. *Arteriosclerosis, thrombosis, and vascular biology*, 23(4):644–9, apr 2003.

[88] F. J.R. Taylor and D. Coates. The code within the codons. *BioSystems*, 22(3):177–187, jan 1989.

[89] Marta Tena-Solsona, Benedikt Rieß, Raphael K. Grötsch, Franziska C. Löhrer, Caren Wanzke, Benjamin Käsdorf, Andreas R. Bausch, Peter Müller-Buschbaum, Oliver Lieleg, and Job Boekhoven. Non-equilibrium dissipative supramolecular materials with a tunable lifetime. *Nature Communications*, 8:15895, jul 2017.

[90] Maria Tsanakopoulou and John D. Sutherland. Cyanamide as a prebiotic phosphate activating agent-catalysis by simple 2-oxoacid salts. *Chemical Communications*, 53(87):11893–11896, oct 2017.

[91] Rebecca M Turk, Nataliya V Chumachenko, and Michael Yarus. Multiple translational products from a five-nucleotide ribozyme. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10):4585–9, mar 2010.

[92] Mark T Uhlik, Amy N Abell, Nancy L Johnson, Weiyong Sun, Bruce D Cuevas, Katherine E Lobel-Rice, Eric A Horne, Mark L Dell'Acqua, and Gary L Johnson. Rac-MEKK3-MKK3 scaffolding for p38 MAPK activation during hyperosmotic shock. *Nature cell biology*, 5(12):1104–10, dec 2003.

[93] Günter von Kiedrowski. A Self-Replicating Hexadeoxynucleotide. *Angewandte Chemie International Edition in English*, 25(10):932–935, oct 1986.

[94] Günter von Kiedrowski, Britta Wlotzka, Jörg Helbing, Matthias Matzen, and Stephan Jordan. Parabolic Growth of a Self-Replicating Hexadeoxynucleotide Bearing a 3'-5'-Phosphoamidate Linkage. *Angewandte Chemie International Edition in English*, 30(4):423–426, 1991.

[95] Michael M K Vu, Nora E Jameson, Stuart J Masuda, Dana Lin, Rosa Larralde-Ridaura, and Andrej Lupták. Convergent evolution of adenosine aptamers spanning bacterial, human, and random sequences revealed by structure-based bioinformatics and genomic SELEX. *Chemistry & biology*, 19(10):1247–54, oct 2012.

[96] Christoph J Wienken, Philipp Baaske, Ulrich Rothbauer, Dieter Braun, and Stefan Duhr. Protein-binding assays in biological liquids using microscale thermophoresis. *Nature communications*, 1:100, jan 2010.

[97] Andrew Williams and Ibrahim T. Ibrahim. Carbodiimide Chemistry: Recent Advances. *Chemical Reviews*, 81(6):589–636, dec 1981.

[98] Carl R Woese. *The genetic code*. New York : Harper and Row, 1967.

[99] J T Wong. A co-evolution theory of the genetic code. *Proceedings of the National Academy of Sciences of the United States of America*, 72(5):1909–12, may 1975.

[100] Nadia Wrobel, Manfred Schinkinger, and Vladimir M. Mirsky. A novel ultraviolet assay for testing side reactions of carbodiimides. *Analytical Biochemistry*, 305(2):135–138, 2002.

[101] Michael Yarus. The Genetic Code and RNA-Amino Acid Affinities. *Life*, 7(2):13, mar 2017.

[102] Michael Yarus, J Gregory Caporaso, and Rob Knight. ORIGINS OF THE GENETIC CODE: The Escaped Triplet Theory. *Review Literature And Arts Of The Americas*, 74:179–98, 2005.

[103] Michael Yarus, Jeremy Joseph Widmann, and Rob Knight. RNA-amino acid binding: A stereochemical era for the genetic code, nov 2009.

[104] David Yu Zhang and Erik Winfree. Control of DNA strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society*, 131(47):17303–17314, 2009.

Acknowledgments

Now, when the long-awaited end of my PhD is here, I would like to thank, first and foremost, Dieter - for letting me work in his lab, for motivating constantly and most enthusiastically, for giving me enough freedom to explore the projects, and for always having to-the-point ideas when I was running out of them. Also, I will never forget "learning to paraglide" as a stress-releasing activity for lab :)

Without the people of the lab, my PhD times would not have been the same. Throughout the years, some people left, many new came, but the lab atmosphere always stayed amazing. Despite so often failing experiments, I always wanted to show up at work, because of all of you.

First, I would like to thank the old lab crew. Simon for being knowledgable about whatever I asked you. Und natürlich dafür, dass wir auf Deutsch gesprochen haben. Christof for programming support and always being there when I needed an advice on a new camera, an objective, or any new tech device. Ike for being a role model on how to handle negotiations and combine hard work with fun afterwards. And for the TeaSchnapps recipe. Lonzo for the most entertaining LabVIEW support and teaching how to make "Dieter-proof" paper figures. Matzi for the relaxing in the middle of the rough day the handstand sessions, and the correcting of the failures in my use of the articles. Manu for showing me the world of bouldering, which kept me sane during the hardships of my PhD. Ferdi for always nice conversations. Julian for your dry sense of humor and always being there to talk to. I would like to thank Therese Dau for the greatest collaboration experience ever. I learnt a lot about project management and paper writing from you. With Georg I had my most memorable scientific discussions. Thank you for listening to all of my whining, keeping me strong and laughing, and always supporting me.

I would also like to thank the new lab crew. I was lucky to have Annalena as my Master student. You helped me collect my strength to finish simulations that I otherwise always postponed. Philipp for helping me when I was going through my lowest productivity months. If I could give you an award for the most trustworthy Bachelor student, I would. Patrick, perceptually uniform colormaps forever! And thank you for always being ready to jump on when someone needs help. Alex for the very useful corrections of my Thesis, and for reviving the lab bouldering sessions. Alan for your seldom but so good jokes spiced with an Italian accent. Christina for the passion and energy you brought to our lab. Newlian for your attention to detail even when many of us would have already lost patience. Thomas and David, it was very nice to have the two of you around.

I would like to thank Slava for supporting me when I decided to move to Munich for my PhD and during my first years in Munich.

I thank the Golden Coffee Machine for bringing the Braunis, Gambis, and Rädlers together for the discussions over the best self-roasted coffee in town. A lot of those discussions were scientific. During many of them, good ideas were born.
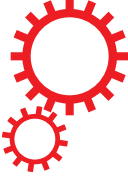
I would like to thank QBM for financial support during the first three years of my PhD, and its amazing team for answering all of the questions I had from the beginning to the end. QBM, CeNS, and NIM for the serious scientific business and all the fun during the retreats.

My mom, my dad, my brother, and Kanika - thank you for being there for me through it all. No matter what.

A   Associated First Author Publications

- T. Dau, **E. V. Edeleva (shared first author)**, S. A. I. Seidel, R. A. Stockley, D. Braun & D. E. Jenne. Quantitative analysis of protease recognition by inhibitors in plasma using microscale thermophoresis. Scientific Reports **6**, Article number: 35413 (2016)

- **E. Edeleva**, A. Salditt, J. Stamp, P. Schwintek, J. Boekhoven & D. Braun. Continuous nonenzymatic cross-replication of DNA strands with *in situ* activated short DNA oligonucleotides. Manuscript in preparation.

SCIENTIFIC REP◯RTS

OPEN

Quantitative analysis of protease recognition by inhibitors in plasma using microscale thermophoresis

T. Dau[1,*], E. V. Edeleva[2,3,*], S. A. I. Seidel[2], R. A. Stockley[4], D. Braun[2] & D. E. Jenne[1,5]

High abundance proteins like protease inhibitors of plasma display a multitude of interactions in natural environments. Quantitative analysis of such interactions *in vivo* is essential to study diseases, but have not been forthcoming, as most methods cannot be directly applied in a complex biological environment. Here, we report a quantitative microscale thermophoresis assay capable of deciphering functional deviations from *in vitro* inhibition data by combining concentration and affinity measurements. We obtained stable measurement signals for the substrate-like interaction of the disease relevant inhibitor α-1-antitrypsin (AAT) Z-variant with catalytically inactive elastase. The signal differentiates between healthy and sick AAT-deficient individuals suggesting that affinity between AAT and elastase is strongly modulated by so-far overlooked additional binding partners from the plasma.

One of the fundamental regulatory mechanisms in cell biology is the finely tuned interaction between enzymes and their substrates[1]. Binding between an enzyme and a substrate is most likely strongly influenced by the composition of the biological buffer system, for example by various plasma components[2]. However, functional assessment of protein-protein interactions in native environments has been a major challenge to date[3,4].

Immunoassays and surface plasmon resonance are the current standard approaches to quantitate concentrations in plasma and to determine affinities of potential known targets[5,6]. However, both methods require surface immobilization of an antigen or an antibody, potentially imposing steric hindrance and molecular activity problems. Immobilization-free approaches based on fluorescence resonance energy transfer (FRET) require availability and labeling of both interacting partners[7,8]. Therefore, FRET-based methods cannot determine the affinity of a person-specific endogenous protein to a known target of interest in a plasma sample.

Microscale thermophoresis (MST) is a recently established immobilization-free affinity measurement technique that has been applied to characterize ligand-binder interactions and affinity constants ranging from pM to mM under challenging conditions[9–11]. MST employs the physical phenomenon of thermophoresis – movement of molecules in a temperature gradient. Biomolecules tend to move against the temperature gradient causing the depletion of a biomolecule in the heated spot. In the MST binding measurement, the depletion of the fluorescently labeled binder changes upon ligand binding. Measurement of depletion at increasing ligand concentrations results in a binding curve, which can be used to derive the affinity.

In our previous work using an MST-based approach, we showed how the affinity as well as the concentration of the ligand can potentially be determined in complex matrices such as blood serum[12]. The first limitation of this method is the sensitivity as only an additionally spiked but not initially present ligand could be detected in serum. Secondly, low affinity interactions remain a challenge, as they often require unrealizable high concentrations of the unlabeled binder to establish a complete binding curve.

An important clinically relevant example of protein interactions in plasma is the inhibition of the enzyme neutrophil elastase (NE) by α-1-antitrypsin (AAT). Low plasma levels of AAT predispose to emphysema and chronic obstructive pulmonary disease[13,14]. Decreased AAT levels are often caused by the relatively common Z-mutation

[1]Comprehensive Pneumology Center, Institute of Lung Biology and Disease (iLBD), University Hospital, Ludwig Maximilians University and Helmholtz Zentrum München, Member of the German Center for Lung Research (DZL), Max-Lebsche-Platz 31, 81377 Munich, Germany. [2]Systems Biophysics and Functional Nanosystems, Ludwig Maximilians University München, Amalienstrasse 54, 80799 Munich, Germany. [3]Graduate School of Quantitative Biosciences, Ludwig Maximilians University, Feodor-Lynen-Str. 25, 81337 Munich, Germany. [4]Lung Investigation Unit, Queen Elizabeth Hospital Birmingham, Mindelsohn way, Edgbaston, Birmingham B15 2WB, UK. [5]Max-Planck-Institute of Neurobiology, Am Klopferspitz 18, D-82152 Martinsried, Germany. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to D.E.J. (email: dieter.jenne@helmholtz-muenchen.de)
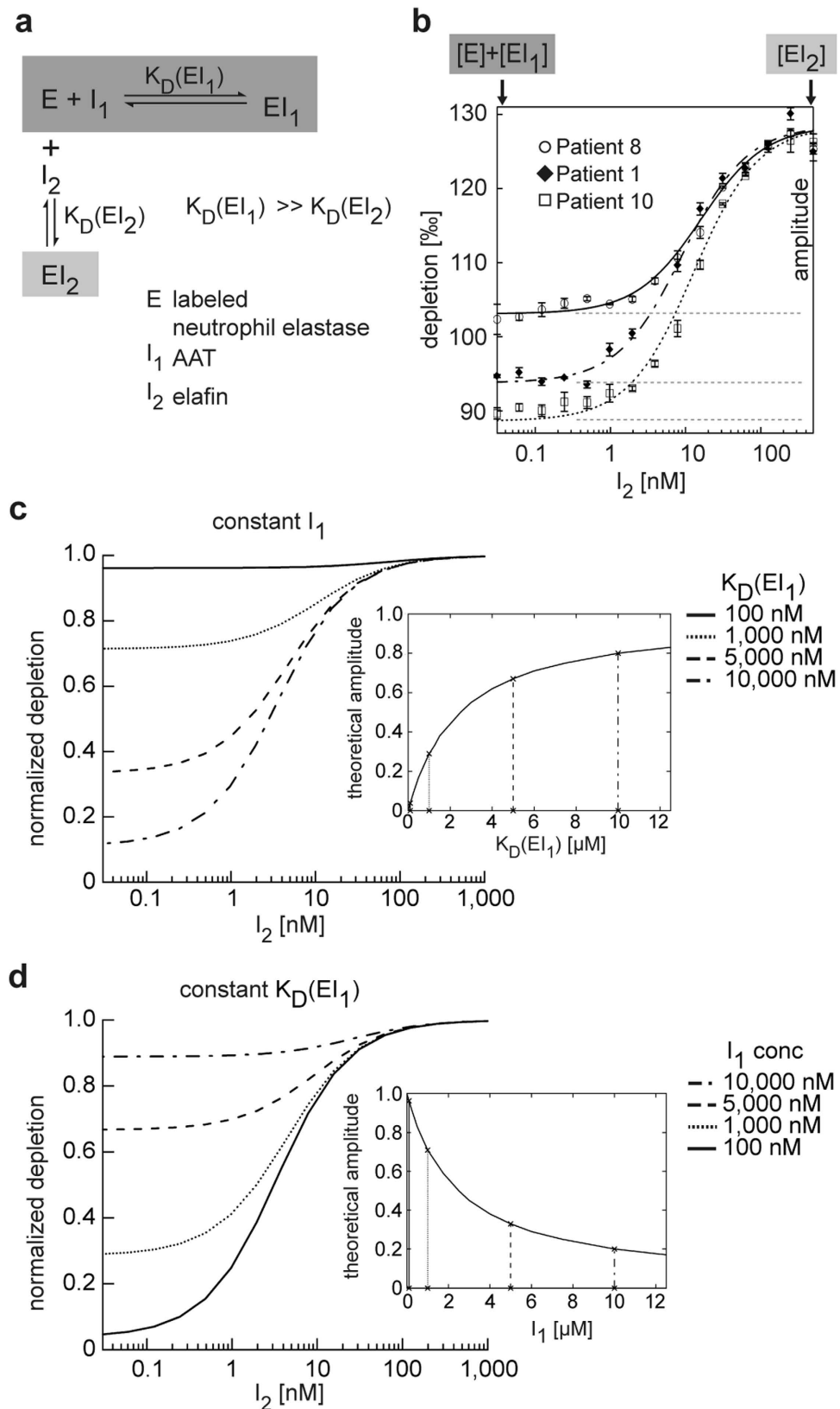
Figure 1. Dependency of the amplitude on the AAT concentration and the dissociation constant of NE-AAT binding. (**a**) Describes the theoretical background of the method. At each data point, the inhibitor I_1 (here: AAT) competes with I_2 (here: elafin) for the enzyme E (here: labeled inactive NE). The affinity between E and I_2 is much higher than the affinity between E and I_1. (**b**) To a constant amount of plasma (7.5%) containing I_1 and a constant amount of E (500 pM) increasing concentrations of I_2 were added and the thermophoretic depletion measured. Titration curves with theoretical fits are given for three individuals (patients 1, 8 and 10) as examples in (**b**). All individuals were homozygous for the Z-variant AAT. Each depletion value is given as

mean ± S.D. from three technical replicates. The amplitude is defined as the difference of minimal depletion signals in the low end range (dark grey box with arrow) and maximal depletion signals in the high end range of I_2 concentrations (light grey box with arrow). The thermophoretic depletion of free E is smaller than that of E bound to I_1 or I_2, and depletion at each I_2 concentration indicates the proportion of bound E. (**c,d**) Are simulations according to the mass action laws. (**c**) Shows that the theoretical amplitude (simulated as the amount of bound E) increases when the dissociation constant between E and I_1 increases. (**d**) On the other hand, the theoretical amplitude decreases with the increasing concentrations of I_1.

....

(Glu342Lys)[15] which accounts for 95% of patients with clinically relevant manifestations. Although all individuals with the Z-mutation have reduced AAT levels, development of lung emphysema is highly variable among ZZ carriers and some are minimally affected[13,14,16].

The interaction between NE and AAT has been well studied *in vitro* using the two purified interacting proteins in standard buffers[17]. AAT inhibits NE irreversibly in a two-step reaction[18]. The first and rate-determining step is the formation of a reversible encounter complex between NE and AAT, where AAT mimics a substrate[19,20]. Previous approaches focused only on the concentration of AAT[5,13,15,16], although kinetically efficient formation of this encounter complex also depends on the dissociation constant between NE and AAT[19,20].

In this work, we developed a novel assay that overcomes the limitations of the previous approaches and allowed us to characterize the formation of the NE-AAT encounter complex at the given AAT level of a blood plasma sample under constant equilibrium conditions.

Results

Method development: combining concentration and affinity. In our assay, a low affinity inhibitor I_1 (here: AAT) competes with a high affinity inhibitor I_2 (here: elafin) for a catalytically inactive labeled enzyme E (here: NE) (Fig. 1a). Inactivation of E was important to prevent the removal of I_1 from the equilibrium by the irreversible formation of a covalent complex, thereby allowing us to analyze the encounter reaction in the plasma environment.

The labeled E at a fixed final concentration of 500 pM was mixed with 12-fold diluted plasma containing the analyte I_1. To obtain binding curves, increasing amounts of I_2 were then added and thermophoretic depletion of free E and E bound to I_1 (EI_1) or I_2 (EI_2) was measured for each concentration of I_2 (Fig. 1a,b). We observed that the depletion in the low end range of I_2 concentrations varied between plasma samples while the depletion in the high end range of I_2 concentrations did not change (Fig. 1b).

To compare the binding curves, we introduced a normalized parameter, the thermophoretic amplitude, by calculating the difference between the depletion signals in these two I_2 concentration ranges (Fig. 1b, Supplementary Information 1). We observed that the thermophoretic amplitude varied between plasma samples. We hypothesized that I_1 concentrations and, more importantly, I_1 affinities to E are not necessarily constant in the person-specific proteomic environment of plasma samples (Fig. 1b).

To characterize the system at different affinities between I_1 and E (Fig. 1c) and different I_1 concentrations (Fig. 1d), we described it with the mass action law equations and simulated the corresponding binding curves. The simulation results indicate that the thermophoretic amplitude increases with an increasing $K_D(EI_1)$ (Fig. 1c small insertion graph) and decreases with increasing I_1 concentrations (Fig. 1d small insertion graph). Hence, the thermophoretic amplitude reports on both I_1 concentration and $K_D(EI_1)$ (a detailed mathematical analysis is provided in the Supplementary Information 2).

Thermophoretic amplitude is better correlated with disease status. To evaluate whether the thermophoretic amplitude that includes affinity variations represents a better lung function parameter than AAT concentration alone, we determined both parameters in plasma samples of individuals homozygous for the AAT Z-variant. We then compared these to FEV_1 (the forced expiratory volume in one second) expressed as the percentage (%) predicted for age, sex, height, and race, which is widely used to assess the level of airway obstruction and bronchoconstriction in AAT-deficient individuals[16,21].

In order to avoid measuring the concentration of inactive AAT, we titrated the concentration of functional plasma AAT with purified and active site-titrated NE. We observed a significantly lower thermophoretic amplitude (P = 0.0008) with plasma samples from individuals showing a high FEV_1 ($\geq$80%) than from those with low FEV_1 (<80%) (Fig. 2a), while the concentration of functional AAT did not differ between the two groups (P = 0.6368). Hence, inclusion of affinity with our method resulted in a much better relationship to disease status and severity. To explain the discrepancy between the two quantification methods, we reasoned that $K_D(EI_1)$ of the Z-variant in fact varied in the plasma samples.

Z-variant AAT is sensitive to plasma environments. To investigate whether affinity is indeed sensitive to a given plasma environment, we determined the $K_D(EI_1)$ of the AAT Z-variant in the two AAT-deficient plasma pools from persons with high ($\geq$80%) and low FEV_1 ($\leq$50%) values (Fig. 3). To facilitate the recombinant production of the aggregation-prone Z-variant, we combined three stabilizing mutations derived from the work of Lee *et al.*[22] – Met374Ile, Ser381Ala, Lys387Arg – and introduced them into the Z-variant. As a control, we expressed AAT without the Z-mutation but containing the same stabilizing mutations. The affinity of the Z-variant to NE was more than two times better in the plasma pool from individuals with high FEV_1 ($\geq$80%) ($K_D(EI_1) = 500 \pm 100$ nM) than from individuals with low FEV_1 ($\leq$50%) ($K_D(EI_1) = 1300 \pm 250$ nM), confirming the plasma-dependent change in the affinity between NE and Z-variant AAT as observed with our newly developed assay. Additionally, we found that the Z-mutation improves the encounter reaction of AAT with
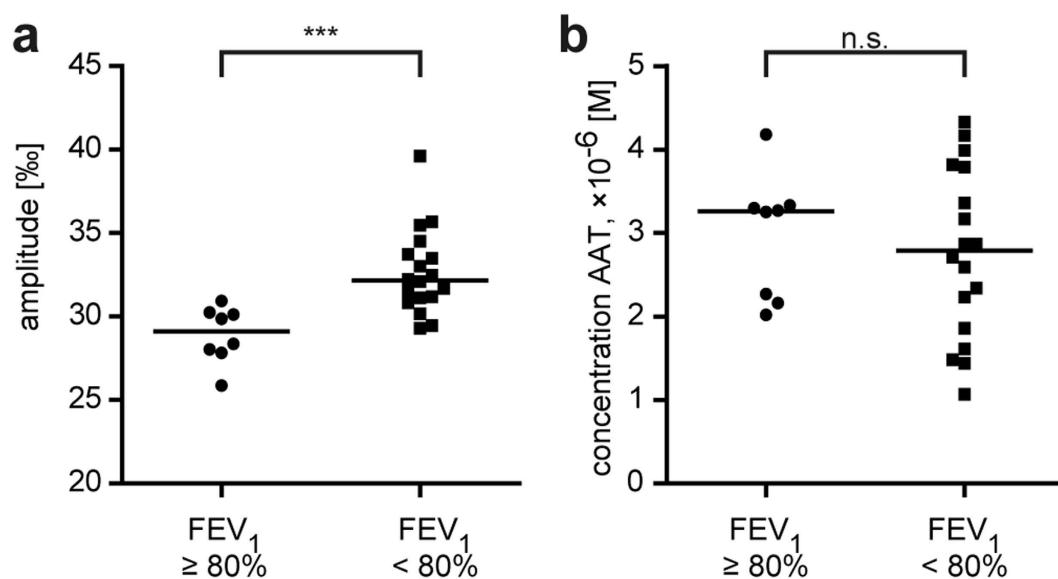
Figure 2. Thermophoretic amplitude, but not plasma concentration of AAT is associated with disease severity. (**a**) ZZ-carriers (n = 26) were divided into two groups with high FEV_1 (≥80%, n = 8) and low FEV_1 (<80%, n = 18), which is a clinical parameter of airway obstruction. The thermophoretic amplitude was significantly lower (P = 0.0008) in samples with high FEV_1 (***P < 0.001; ns not significant). The amplitude is given as a mean value determined from at least two independent experiments. (**b**) The concentration of functional AAT was not significantly different (P = 0.6368) between the groups with high FEV_1 (≥80%) and low FEV_1 (<80%). The AAT concentration was measured twice by elastase titration in duplicates, and is given as a mean value here.
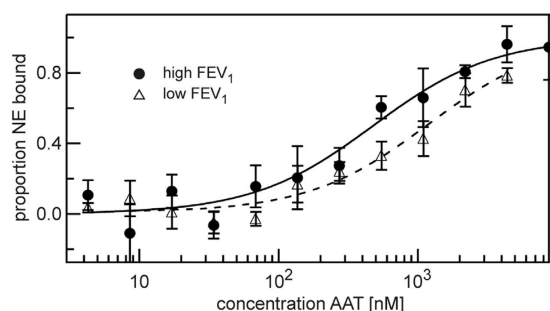

Figure 3. Plasma-dependent change in affinity of NE-AAT binding. We compared the affinity between AAT and NE in two pools of plasma from individuals with high FEV_1 (≥80%, n = 8) and low FEV_1 (≤50%, n = 12). The affinity between Z-variant of AAT and NE was higher in plasma with high FEV_1 (≥80%) ($K_D(EI_1) = 500 \pm 100$ nM) compared to low FEV_1 (≤50%) ($K_D(EI_1) = 1300 \pm 250$ nM). Presented binding curves represent example measurements where each measurement point (mean ± S.D.) was derived from three technical replicates. Fitted binding curves and $K_D(EI_1)$ values (mean ± S.D.) were derived from global fitting of four measurements (three independent protein expressions). The measurements were performed in 7.5% plasma and with 5 nM of labeled NE.

NE (Supplementary Fig. 1a) and Z-variant AAT is more susceptible to plasma components than control AAT (Supplementary Fig. 1b).

Discussion

By incorporating a high affinity ligand in thermophoretic measurements, we expanded the spectrum of applications towards the low affinity range where e.g. enzyme-substrate interactions take place. This new approach allowed us to detect unexpected plasma-dependent interferences with AAT, suggesting that the functional capacity of the Z-variant AAT is determined by the heterogeneity of other plasma components, which modulate the affinity of AAT towards target enzymes. For example, it was reported[23,24] and has been confirmed by our group (unpublished) that lipoproteins bind to wildtype AAT, thereby influencing its inhibition of NE and thus could also affect the reactivity of the Z-variant in the plasma environment.

The analysis of the interaction between AAT and NE in plasma with our assay provides a possible explanation of the variability in disease development in ZZ carriers. Normally, the reduced AAT plasma levels in

ZZ carriers are compensated by the enhanced affinity of Z-variant AAT to NE. However, the enhanced affinity between Z-variant AAT and NE can be disrupted by plasma factors, especially in individuals who have higher concentrations of interfering plasma components.

Our work underscores the importance of the endogenous personally unique and distinct environment where substrates interact with enzymes with relatively low affinity in a complex web of other proteins and lipid components. Such personalized variations and effects have often been overlooked. Our approach may open a way to study how other components from the native environments interfere with biochemical reactions.

Methods

Production of recombinant proteins. As the natural Z-variant is on a M1(Ala213) background, we amplified a cDNA fragment encoding the M1(Ala213) variant using the primer DJ3689 (5′-GACTTCCACG TGGACCAGG**C**GACCACCGTGAA-3′) and DJ3613 (5′-GATG*ACCGGT*TTTTTGGGTGGGATTCAC CACTT-3′) (*cursive*: restriction site; **bold**: mutation) and cloned it into the *Pml* I and *Age* I site of the previously described M1(Val213) AAT construct in pTT5 plasmid (Perera *et al.*[25]. To introduce the Z-mutation (Glu342Lys), a PCR fragment was generated with the primers DJ3557 (5′-CCAC*GATATC*ATCACCAAGTTCCT-3′) and DJ3558 (5′-GTATGG*CCTCGAGG*AACATGGCCCCAGCAGCTTCAGTCCCTTTCTTGTCGATGGT-3′) and inserted into the M1(Ala213) AAT pTT5 plasmid between the *Eco* RV and *Abs* I sites. The three stabilizing mutations (M374I, S381A, K387R) were introduced by PCR using the following forward and backward primers DJ 3646 (5′- CAC*GATATC*ATCACCAAGTTCCTGGAAAATGAAGACAGAAGGTCTGCCGACTTACATTTACC-3′) and DJ3696 (5′-ATG*ACCGGT*TTTTTGGGTGGGATTCACCACT**C**TTCCCATGAAGAGGGGAG**C** CTTGGT ATTTTGTTCAAT**A**ATTAAG-3′) respectively. The product was inserted into the *Eco* RV and *Age* I sites.

To produce a catalytically inactive variant of NE, we mutated the Ser195 to Ala195 by inserting an oligo duplex (DJ3532 (5′-GTGAACGTATGCACTCTGGTGCCACGTCGGCAGGCAGGCATCTGCTTCGG GGACGCT-3′) and DJ3533 (5′-CGTCCCCGAAGCAGATGCCTGCCTGCCGACGTGGCACCAGAGTG CATACGTTCACAC-3′)) into the *Alf I* site of the previously described wildtype mouse NE construct in pTT5 (Dau *et al.*[26]. To enable site-specific labeling of NE, we added a C-terminal short peptide with a cysteine flanked by three aspartate residues on each side by the insertion of an oligoduplex (DJ3632 (5′- CTAGCGACGACGATTG CGACGATG ATC-3′) and DJ3633 (5′- CTAGGATCATCGTCGCAATCGTCGTCG-3′)) into the *Avr* II site.

All proteins were expressed in HEK293 EBNA cells (Yves Durocher, National Research Council Canada, Montreal, Canada) in FreeStyleTM 293 expression medium (Thermo Fisher Scientific Inc.), 1% Pluronic and G418 (25 μg ml^{-1}) at 37 °C and 8% CO_2.

Labelling of NE. First, we added 1 mM DTT to the recombinant NE in storage buffer (20 mM Na_2HPO_4, 300 mM NaCl pH 7.4) and incubated it for two hours at room temperature to reduce all cysteine-tags. Then we removed DTT by precipitating NE with 75% ammonium sulfate. This reduction and precipitation step was repeated once. NE was dissolved in storage buffer and incubated at room temperature for one hour after adding a 5-fold molar excess of dye (Alexa Fluor® 647 NHS Ester, ThermoFisher Scientific). To remove the excess dye, the solution was added to PD MiniTrap G-10 column (GE Healthcare) according to the manufacturers instruction.

Determination of AAT concentration in blood plasma. After titrating trypsin (Sigma-Aldrich, St. Louis) with 4-nitrophenyl 4-guanidinobenzoate (Sigma-Aldrich, St Louis, USA) in veronal buffer, AAT (Athens Research & Technology, Athens, GA, USA) was titrated against trypsin. A dilution series of AAT was incubated with a constant amount of trypsin at 37 °C for 1 hour. The residual activity was measured using Boc-Gln-GLy-Arg-AMC. Thereafter, a dilution series of the titrated AAT was incubated with a constant volume of active human NE (Elastin Products Company, Inc., Owensville) at 37 °C for 1 hour. The remaining activity was measured with MCA-GEAIPTSIPPEVK(Dnp)-rr (EMC microcollections, Tübingen, Germany). All reactions were performed in 150 mM NaCl, 50 mM Tris, 0.01% Triton-X-100, pH 7.4.

Active site titrated neutrophil elastase was added to a dilution series of plasma and incubated at 37 °C for 1 hour. The plasma was diluted in 150 mM NaCl, 50 mM Tris, 0.01% Triton-X-100, pH 7.4. The residual activity was measured using MCA-GEAIPTSIPPEVK(Dnp)-rr (EMC microcollections, Tübingen, Germany).

Determination of the concentration of recombinant AAT. A dilution series of AAT-variants was incubated with 1.6 nM active site titrated human NE at 37 °C for 1 hour. The residual activity was measured using MCA-GEAIPTSIPPEVK(Dnp)-rr (EMC microcollections, Tübingen, Germany).

Statistical analysis. Results are given as median values. A Mann-Whitney test without assuming Gaussian distribution was applied to all studies (***P < 0.001). Statistical analyses were done with GraphPad Prism 6 software (GraphPad Software).

Development of the Binding Model for Simulations. Mass action law equations with two dissociation constants were used to describe the dependency of free NE on the concentration of AAT and on the affinity of AAT to NE in the presence of two NE ligands – AAT and elafin.

$$K^{I_1}{}_D = \frac{([E] - [EI_1] - [EI_2])([I_1] - [EI_1])}{[EI_1]} = [E](1 - x)\left(\frac{[I_1]}{[EI_1]} - 1\right) \tag{1}$$

$$K^{I_2}{}_D = \frac{([E] - [EI_1] - [EI_2])([I_2] - [EI_2])}{[EI_2]} = [E](1 - x)\left(\frac{[I_2]}{x[E] - [EI_2]} - 1\right),$$

(2)

where

$K^{I_1}{}_D$ – dissociation constant of NE – AAT interaction,

$K^{I_2}{}_D$ – dissociation constant of NE – elafin interaction,

$[E], [I_1], [I_2]$ – total concentrations of NE, AAT, and elafin in the reaction, respectively,

$[EI_1]$ – concentration of NE-AAT complex in the reaction,

$[EI_2]$ – concentration of NE-elafin complex in the reaction,

$x = \frac{[EI_1] + [EI_2]}{[E]}$ – fraction of occupied NE,

Equation (2) was solved for [E], which was then applied to equation (1) to yield a cubic equation for *x*. The cubic equation was solved numerically in Igor Pro 5.03 and the smallest real root of the equation was then used to plot the fraction of bound NE *x* against different elafin concentrations I_2 to simulate binding curves.

Determination of the thermophoretic amplitude.　Human recombinant elafin (Sigma-Aldrich) was serially diluted (1:1) over five orders of magnitude in PBS. Separately, plasma was diluted (final concentration: 7.5%) with storage buffer substituted with 0.02% Tween 20 and anti-photobleaching enzyme and substrate components (Monolith Anti Photobleach Kit, NanoTemper Technologies) were added. This plasma solution was mixed 1:1 with PBS for background measurements. After adding fluorescently labeled NE (final concentration: 500 pM) to the plasma solution, we mixed this 1:1 with each concentration of the serial elafin dilution. The samples were loaded into the Monolith™ NT.115 premium coated capillaries (NanoTemper Technologies), and incubated at 22 °C for 2 hours. The samples were measured in the instrument (Monolith NT.115Pico, NanoTemper Technologies) at 22 °C using 60% light-emitting diode and 40% infrared laser (IR) powers with IR laser on/off times of 25/5 seconds. Each dilution point was measured in triplicates. For each plasma sample, the whole procedure was repeated three times to yield independent triplicates.

In Matlab R2014a (8.3.0.532), the time trace of the background signal was subtracted from the time traces of the sample signals. Then, the fluorescence after the temperature jump and equilibrated thermophoresis was normalized to the fluorescence before the IR laser heating yielding fluorescence depletion values. Fluorescence depletion values of three technical replicates (mean ± S.D.) were plotted in per mille units (‰) on a linear y-axis against the concentration of the serially diluted elafin on the $\log_{10}$ x-axis resulting in binding curves. The binding curves were shifted so that the depletion value at the baseline, where elafin concentration is saturating, was the same. In Igor Pro 5.03, a weighted fit to the quadratic solution of the mass action law was performed to yield the amplitude *A* of each curve:

$$f(x) = A\frac{[E] + x + K_D - \sqrt{([E] + x + K_D)^2 - 4[E]x}}{2[E]} + t$$

(3)

where

K_D – dissociation constant of NE – elafin interaction,

$[E]$ – total concentration of NE,

x – total concentration of elafin,

A – the amplitude of the curve,

t – y-offset of the curve,

with A, t, and K_D as free fit parameters.

Determination of dissociation constant between AAT and NE.　Freshly purified AAT was serially diluted (1:1) over five orders of magnitude in storage buffer. Separately, plasma was diluted in storage buffer substituted with 0.2% Tween 20 (final concentration: 7.5%) and anti-bleaching reagents were added. This plasma solution was mixed 1:1 with storage buffer for background measurements. Fluorescently labeled NE (final concentration: 5 nM) was added to the plasma solution and mixed 1:1 with the serial dilution of AAT. The samples were incubated at 22 °C for 2 hours and measured at 22 °C in the instrument using 40% light-emitting diode and 40% infrared laser (IR) powers with IR laser on/off times of 20/5 seconds.

Binding curves were obtained as described for the determination of the thermophoretic amplitude. They were then additionally normalized to 0 at the lowest AAT concentration and to 1 at the highest AAT concentration. This corresponds to the proportion of bound NE at each titration point. The global fit of at least three replicates to the quadratic solution of the mass action law was performed to yield the dissociation constant of each curve.

Plasma samples.　FEV_1 was determined post bronchodilator treatment and according to the British Thoracic Society/Association of Respiratory Technicians and Physiologists (BTS/ARTP) guidelines as described previously[21]. All subjects provided written informed consent, and ethical approval was obtained for all aspects of this study (South Birmingham Research Ethics Committee LREC 3359).

References

1. Overall, C. M. & Blobel, C. P. In search of partners: linking extracellular proteases to substrates. *Nat. Rev. Mol. Cell Biol.* **8,** 245–257 (2007).
2. Adams, T. E. & Huntington, J. A. Thrombin-cofactor interactions: structural insights into regulatory mechanisms. *Arterioscler. Thromb. Vasc. Biol.* **26,** 1738–1745 (2006).

3. Syafrizayanti, Betzen, C., Hoheisel, J. D. & Kastelic, D. Methods for analyzing and quantifying protein-protein interaction. *Expert Rev. Proteomics* **11,** 107–120 (2014).
4. Sinz, A. Investigation of protein-protein interactions in living cells by chemical crosslinking and mass spectrometry. *Anal. Bioanal. Chem.* **397,** 3433–3440 (2010).
5. McElvaney, N. G. Diagnosing α1-antitrypsin deficiency: how to improve the current algorithm. *Eur. Respir. Rev.* **24,** 52–57 (2015).
6. Mariani, S. & Minunni, M. Surface plasmon resonance applications in clinical analysis. *Anal. Bioanal. Chem.* **406,** 2303–2323 (2014).
7. Kraynov, V. S. *et al.* Localized Rac activation dynamics visualized in living cells. *Science* **290,** 333–337 (2000).
8. Uhlik, M. T. *et al.* Rac-MEKK3-MKK3 scaffolding for p38 MAPK activation during hyperosmotic shock. *Nat. Cell Biol.* **5,** 1104–1110 (2003).
9. Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D. & Duhr, S. Protein-binding assays in biological liquids using microscale thermophoresis. *Nat. Commun.* **1,** 1–7 (2010).
10. Seidel, S. A. I. *et al.* Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods* **59,** 301–315 (2013).
11. Jerabek-Willemsen, M. *et al.* MicroScale Thermophoresis : Interaction analysis and beyond. *J. Mol. Struct.* **1077,** 101–113 (2014).
12. Lippok, S. *et al.* Direct detection of antibody concentration and affinity in human serum using microscale thermophoresis. *Anal. Chem.* **84,** 3523–3530 (2012).
13. Gooptu, B., Ekeowa, U. I. & Lomas, D. A. Mechanisms of emphysema in alpha1-antitrypsin deficiency: molecular and cellular insights. *Eur. Respir. J.* **34,** 475–488 (2009).
14. Sun, Z. & Yang, P. Role of imbalance between neutrophil elastase and alpha 1-antitrypsin in cancer development and progression. *Lancet. Oncol.* **5,** 182–190 (2004).
15. Hutchison, D. C. Natural history of alpha-1-protease inhibitor deficiency. *Am. J. Med.* **84,** 3–12 (1988).
16. Stockley, R. A. & Turner, A. M. α-1-Antitrypsin deficiency: clinical variability, assessment, and treatment. *Trends Mol. Med.* **20,** 105–115 (2014).
17. Dobó, J. & Gettins, P. G. W. alpha1-Proteinase inhibitor forms initial non-covalent and final covalent complexes with elastase analogously to other serpin-proteinase pairs, suggesting a common mechanism of inhibition. *J. Biol. Chem.* **279,** 9264–9269 (2004).
18. Potempa, J., Korzus, E. & Travis, J. The serpin superfamily of proteinase inhibitors: structure, function, and regulation. *J. Biol. Chem.* **269,** 15957–15960 (1994).
19. Kang, U. B. *et al.* Kinetic mechanism of protease inhibition by alpha-1-antitrypsin. *Biochem. Biophys. Res. Commun.* **323,** 409–415 (2004).
20. Gettins, P. G. W. Serpin structure, mechanism, and function. *Chem. Rev.* **102,** 4751–4804 (2002).
21. Guidelines for the measurement of respiratory function. Recommendations of the British Thoracic Society and the Association of Respiratory Technicians and Physiologists. *Respir. Med.* **88,** 165–194 (1994).
22. Lee, K. N., Park, S. D. & Yu, M. H. Probing the native strain iin alpha1-antitrypsin. *Nat. Struct. Biol.* **3,** 497–500 (1996).
23. Gordon, S. M. *et al.* Rosuvastatin alters the proteome of high density lipoproteins: Generation of alpha-1-antitrypsin enriched particles with anti-inflammatory properties. *Mol. Cell. Proteomics* **14,** 3247–3257 (2015).
24. Talmud, P. J. *et al.* Progression of atherosclerosis is associated with variation in the alpha1-antitrypsin gene. *Arterioscler. Thromb. Vasc. Biol.* **23,** 644–649 (2003).
25. Perera, N. C. *et al.* NSP4 is stored in azurophil granules and released by activated neutrophils as active endoprotease with restricted specificity. *J. Immunol.* **191,** 2700–2707 (2013).
26. Dau, T., Sarker, R. S. J., Yildirim, a. O., Eickelberg, O. & Jenne, D. E. Autoprocessing of neutrophil elastase near its active site reduces the efficiency of natural and synthetic elastase inhibitors. *Nat. Commun.* **6,** 6722 (2015).

Acknowledgements

Author Contributions

T.D., E.V.E. and S.A.I.S. designed and performed the experiments. D.E.J. and D.B. supervised the project. R.A.S. provided plasma samples and clinical data. All authors discussed the results and implications and commented on the manuscript at all stages.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Dau, T. *et al.* Quantitative analysis of protease recognition by inhibitors in plasma using microscale thermophoresis. *Sci. Rep.* **6**, 35413; doi: 10.1038/srep35413 (2016).