EXPLORING PROTEIN FITNESS LANDSCAPES WITH NEW HIGH-THROUGHPUT TECHNOLOGIES

Paul Jannis Zurek

Department of Biochemistry and Darwin College,
University of Cambridge

Under the supervision of Florian Hollfelder and Ahir Pushpanath

A thesis submitted for the degree of *Doctor of Philosophy*

March 2021

DECLARATION

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared hereafter and specified in the text. Work done in section B.2.1 and B.2.2 was carried out in collaboration with Raphaëlle Hours. Parts of the work done in section B.2.3 was carried out in collaboration with Elizabeth Broom under my direct supervision. This thesis is not substantially the same as any work that has already been submitted before for any degree or other qualification and it does not exceed the prescribed word limit of 60,000 words.

Part of the work presented in this thesis has been published:

- Work presented in chapter A.3
    - Neun S*, Zurek PJ*, Kaminski TS* & Hollfelder F (2020) Ultrahigh throughput screening for enzyme function in droplets. *Methods in Enzymology*, **643:**317-343. *: equal contribution.
- Work presented in chapter C
    - Zurek PJ, Hours R, Schell U, Pushpanath A & Hollfelder F (2021) Growth amplification in ultrahigh-throughput microdroplet screening increases sensitivity of clonal enzyme assays and minimizes phenotypic variation. *Lab on a Chip*. **21**:163-173.
- Work presented in chapter D
    - Zurek PJ, Knyphausen P, Neufeld K, Pushpanath A & Hollfelder F (2020) UMI-linked consensus sequencing enables phylogenetic analysis of directed evolution. *Nature Communications*, **11**:6023.

Paul Jannis Zurek, 17th of March 2021

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to Prof. Florian Hollfelder, who has crucially aided me in following my scientific interests by providing valuable guidance and support during the PhD. I have enjoyed the flexibility and strong enthusiasm for all kinds of ideas and projects very much; I have learned so much during my time in Cambridge.

Further appreciation goes to Dr. Ursula Schell, who was a great help to me planning my arrival and organising the starting months in Cambridge, as well as setting up the collaboration between Johnson Matthey and the University of Cambridge. Additionally, I would like to express my appreciation to Dr. Ahir Pushpanath for the seamless transition at JM and the comfortable supervision with highly valued industrial insights; and to Dr. Katharina Neufeld for all the hands-on work around the practical aspects of industrial biocatalysis.

I would like to thank Dr. Raphaëlle Hours and Dr. Stéphane Emond for their great help in introducing me to the lab, on the one side showing me all the little useful tips when it comes to droplet microfluidics and AADS, and on the other side setting me up with inspiring protein evolutionary insights and concepts.

A big recognition is due Dr. Philipp 'Dodo' Knyphausen, who has enabled a large part of my PhD from the small but elegant thought of 'barcode funnelling'. His relentless literature research, practical lab tricks and boundless enthusiasm for new and daring projects have greatly enabled and advanced my and many other's projects.

I have enjoyed many constructive discussions with everyone in the Hollfelder lab, initially being intimidated by your collective knowledge but eventually growing into collaborative and increasingly adventurous theories and projects. I am very grateful for having had the opportunity to have worked, studied and socialised alongside all of you! For their further practical involvement in the project, I would especially like to thank Elizabeth 'Liz' Broom and Dr. Eleanor Campbell for the entertaining crystallography work and Joachim de Jonghe and Dr. Tomasz Kaminski for teaching me how to fabricate microfluidic masters while wearing

those sweaty blue overalls in the clean rooms flooded by orange-y neon lights. Thank you, Maximilian Gantz, for taking on many of the follow-up projects, there couldn't be a more motivated or brighter student to teach.

An honourable mention must go out to Remkes' little sister's laptop, which was selflessly made available to me for the early days of tinkering with bioinformatic tools and scientific programming. Furthermore, I would like to thank everyone in the ES-Cat network for the instructive and entertaining workshops and lessons at the network meetings, as well as the EU's Horizon 2020 research program for funding all of it with the Marie-Curie grant 722610.

Finally, I have really enjoyed working and living in Cambridge, especially thanks to all the banter during boardgames with Remke, Fryer, Jojo, Joel, Elena and Wolfgang. Cambridge is an amazing place full of opportunity and great people. At this point I would also like to thank the Cambridge University Volleyball Club for all the fun (and fitness) I had with them, as well as the Darwin College Sports Bursary for funding it. GDBO! I'd also like to thank Liisa for organising some super fun lab skiing trips and teaching me how to quickly advance from skiing pizza on the baby slopes.

Zuletzt und *zuwichtigst*, ein ganz großes Dankeschön an meine Familie und Freunde! Danke an Mike, Viktor, Eugen und Marten für die jährlichen Ausflüge und die Knuspertaler; Danke an die Heidelberg-Crew Lukas und Andy, immer wieder gut euch zu sehen! Ich bin froh, eine so großartige Familie zu haben, ohne euch wäre es nicht möglich gewesen. Barbara, Willi, Kara, Finnja und Ole, ich freue mich immer wieder in Stade und bei euch zu sein, vielen Dank für eure Unterstützung. Danke Jonas und Noah, dass ihr immer grenzenlos begeistert seid, wenn ich zurückkehre. Jana, du bist ein wundervoll lieber und mitfühlender Mensch, danke für all deine Ausdauer; danke für Alles.

ABSTRACT

The concept of a protein's fitness landscape – an abstract space in which related sequences are close together and matched with their fitness – is a useful tool to visualize core principles of protein evolution. Acquiring a new function, for example the laboratory evolution of an enzyme to convert an industrially relevant substrate, can be understood as a stepwise climb through a fitness landscape, reaching higher fitness (or activity) with each step (or mutation). The valleys of such a space relate to the starting points of protein engineering campaigns. Understanding this area could enlighten principles of how proteins quickly adapt in nature and help to identify starting points with a high potential for evolution, a high 'evolvability', speeding up protein engineering. In this study, high-throughput technologies will be developed that enable the read-out of directed evolution on a large scale, tracking the exploration of the valley of a fitness landscape: the conversion of an amino acid- to amine dehydrogenase will be investigated as a model of enzyme evolvability with a drastic change of substrate specificity. A sensitive high-throughput screening assay as well as a comprehensive sequencing read-out will be required to establish the identity of selected variants during evolution.

I will first generate and characterize three different but related starting points and test their initial evolvability. Stabilizing the starting point results in increased mutational robustness, broadening the range of accepted mutations. However, increased initial stability does not necessarily correlate to higher functional improvement, hinting at a nuanced view of evolvability. A sensitive high-throughput assay is necessary to verify the full potential of the starting points and study the early steps of evolution comprehensively. Broadly applicable ultrahigh-throughput assays of enzyme function, such as absorbance-activated droplet sorting, currently lack the sensitivity of more specific fluorescence-based or low-throughput counterparts. A universal approach to increase detectability in single cell-lysate microfluidic enzyme assays is established by amplifying the enzyme content per droplet more than 10-fold via homogeneous clonal cell growth. Clonal amplification enables the sensitive and precise detection of newly introduced amine dehydrogenase activities, a feat restricted in conventional assays by low initial activity and stability. To generate a truly complete view of directed evolution in a fitness landscape, however, an equally powerful sequencing read-out is

necessary to identify all selected variants. Here, unique molecular identifiers are used to increase the accuracy of nanopore sequencing to levels that can reliably distinguish point mutations. I establish an inexpensive and straightforward long read amplicon sequencing workflow which is then applied to map the trajectories of two comparative long-term directed evolution campaigns. In the parallel evolution campaigns, initial beneficial mutations are exclusive to each starting point and lead to incompatible trajectories. Beneficial mutations are scarce and large improvements are unavailable until recombination occurs and a jump through the fitness landscape is realized. The recombined variant holds high evolvability and quickly evolves to take over the population and form the most successful lineages, indicating the power of recombination as a means to innovation in protein evolution.

The tools established in this thesis can help protein engineers explore fitness landscapes more economically and comprehensively. Their application to mapping full trajectories of early adaptation uncovers differences in the evolvability of homologs, potentially aiding the identification of evolvable starting points as well as strategies to increase evolvability for efficient protein engineering in the future.

TABLE OF CONTENTS

DECLARATION **1**

ACKNOWLEDGEMENTS **2**

ABSTRACT **4**

TABLE OF CONTENTS **6**

A. INTRODUCTION **1**

A.1. DIRECTED EVOLUTION AND EVOLVABILITY 2

A.1.1. From Protein Engineering to Fitness Landscapes *2*

A.1.2. The Potential for Adaptation *6*

A.2. AMINE DEHYDROGENASES 11

A.2.1. Overview of Enzymatic Ways to Chiral Amines *11*

A.2.2. Generation of Amine Dehydrogenases *13*

A.3. ULTRAHIGH-THROUGHPUT SCREENING FOR ENZYME FUNCTION IN DROPLETS 15

A.3.1. Detection Modes *17*

A.3.2. Available Enzymatic Assays *20*

A.3.3. Consequences of Ultrahigh-Throughput Screening *20*

A.4. MAPPING SEQUENCE-FUNCTION RELATIONSHIPS 23

A.4.1. Deep Mutational Scanning of New Functions *25*

A.4.2. Applied Sequencing Technologies *27*

A.5. STRUCTURE AND AIM OF THIS THESIS 27

B. STARTING POINTS **29**

B.1. INTRODUCTION 30

B.2. RESULTS & DISCUSSION 32

B.2.1. Computational Stabilization *32*

B.2.2. Stabilization by Directed Evolution *33*

B.2.3. Comparison of Evolvability – Generation of AmDHs *36*

B.3. CONCLUSION & OUTLOOK 41

C. INCREASING THE SENSITIVITY OF SINGLE CELL MICROFLUIDIC ASSAYS 44

C.1. INTRODUCTION 45

C.2. RESULTS & DISCUSSION 48

C.2.1. A Device for Homogeneous Cell Growth in Droplets 49

C.2.2. Induction of Recombinant Protein Expression 52

C.2.3. Active Delivery of Assay Components 55

C.2.4. A Sorting Algorithm to Exclude Doublets 56

C.2.5. Detection and Sorting of Low Enzyme Activities 59

C.2.6. Quantification of Improvements 61

C.3. CONCLUSION & OUTLOOK 62

D. ACCURATE LONG-READ SEQUENCING FOR THE ANALYSIS OF DIRECTED EVOLUTION 66

D.1. INTRODUCTION 67

D.2. RESULTS & DISCUSSION 71

D.2.1. The UMIC-seq Workflow and Tools 71

D.2.2. An Exemplary AmDH Evolution 75

D.2.3. Determination of UMI Efficiency and Sequencing Error-Rate 75

D.2.4. A Phylogeny of Directed Evolution 80

D.2.5. Characterisation of Epistasis in Founder Variants 84

D.3. CONCLUSION & OUTLOOK 86

E. ESCAPING LOCAL OPTIMA IN COMPARATIVE LONG-TERM EVOLUTION 89

E.1. INTRODUCTION 90

E.2. RESULTS & DISCUSSION 92

E.2.1. Long-Term Microfluidic Directed Evolution of Two AmDHs 92

E.2.2. UMIC-seq Reveals the AmDH Sequence Space 94

E.2.3. Characterization and Transplantation of Key Mutations 98

E.2.4. End-Point Variants 100

E.2.5. Recombination Escapes Local Optima 102

E.2.6. Characterization of Chimeric Variants 105

E.3. CONCLUSION & OUTLOOK 108

F. FINAL CONCLUSION AND PERSPECTIVE 112

G. MATERIALS & METHODS 117

G.1. MATERIALS 118

G.2. MOLECULAR BIOLOGY 118

G.3. LYSATE ASSAYS 121

G.4. PROTEIN PURIFICATION AND CHARACTERIZATION 122

G.5. Microfluidics 124
G.6. Nanopore Sequencing 127
H. APPENDIX 131
H.1. Table of Abbreviations 131
H.2. Table of Figures 132
H.3. Table of Tables 134
H.4. Supplementary Figures 135
H.5. Supplementary Tables 140
H.6. Supplementary Sequences 141
I. REFERENCES 142

A. INTRODUCTION

What are the tools necessary for directed evolution and what can we learn about proteins from performing it? In this chapter I will provide a general introduction to directed evolution and the high-throughput methods making it possible, with a focus on droplet microfluidics. I will also summarize the current state of literature around the lessons that can be derived from large scale directed evolution experiments combined with an equally powerful readout in high-throughput sequencing, collectively known as deep mutational scanning.

The content and figures of chapter A.3 were published as Neun et al*, 2020.*

A.1. Directed Evolution and Evolvability

A.1.1. From Protein Engineering to Fitness Landscapes

Directed evolution is a mimic of natural Darwinian evolution in the laboratory. Mutations are introduced to a protein and variants are assayed under defined conditions in a test tube, leaving the experimenter to direct the protein's function outside of its original biological context towards the relevant task at hand (Arnold, 1996). For example, selection pressure can be set to evolve enzymes useful in industrial biocatalysis or to select new binding specificities of antibodies for pharmaceutical application. This has been proven with seminal work on enzyme evolution by Frances Arnold, and phage display by George Smith and Sir Gregory Winter; recognized with the Nobel Prize in Chemistry in 2018. Protein engineering by directed evolution has been established in these fields for many years thanks to its agnostic approach. Already in 1993, Chen and Arnold were able to mould the protease subtilisin E to work at high efficiencies in an artificial environment. The mutant enzyme, obtained after screening a total of ~4000 variants in three cycles of directed evolution, enabled catalysis in organic solvent at a 256-fold increased catalytic efficiency compared to the wild-type (Chen & Arnold, 1993). In such early studies, directed evolution of enzymes was commonly applied as iterative plate screening in which multiple rounds of mutagenesis (e.g. random mutagenesis, site-directed mutagenesis and/or recombination) were coupled with screening in 96-well plates for improved function in a variety of assays (Arnold, 1996, Fig. A-1). Selections, in contrast, were in parallel beginning to be employed to fish out functional variants of large pools of diversity by competition, traditionally affording much higher throughput but less assay flexibility compared to the individual measurements taken during screening (Winter *et al*, 1994; Yano *et al*, 1998). However, the lines between these approaches have started to blur with the application of emerging ultrahigh-throughput screening technologies (see chapter A.3) and non-iterative continuous directed evolution (Esvelt *et al*, 2011). In any case, all these approaches can be used to direct protein function to industrially useful properties by applying defined selection pressures while not requiring much prior knowledge about the protein.

Directed evolution, however, is much more than a blind tool used for protein engineering. In fact, as directed evolution is an approximation of natural evolution in defined conditions and selection pressures, it can be used to test evolutionary theories and establish fundamental properties of proteins themselves (Peisajovich & Tawfik, 2007). Here, it helps to imagine directed evolution as a stepwise journey through a fitness landscape. First, the enormous

diversity in protein sequences can be visualized by imagining a space in which each sequence is neighboured by all the sequences which can be reached by one mutation. This concept of a sequence space was first introduced to protein evolution by John Maynard Smith in 1970, visualizing the principles that allow natural evolution to overcome the insurmountable combinatorial diversity of protein sequences (Maynard Smith, 1970). For a theoretical protein of two amino acids a sequence space is easily visualized in two dimensions (Fig. A-2a), but the dimensionality of sequence space and the contained possible combinations of amino acids get unimaginably large with longer sequences: There are 20 possible amino acids per position, resulting in a simple 20 × 20 space for 400 possible di-peptides, but rapidly expanding to more than 10^{130} members in the 100 dimensional space of a theoretical small protein of 100 amino acids. This diversity in possible protein sequences is thought to result in a globally extremely sparsely populated sequence space. Random proteins will likely not be functional and functional protein superfamilies, therefore, exist as island of active protein sequences in a vastness of non-functional sequences (Keefe & Szostak, 2001; Romero & Arnold, 2009).

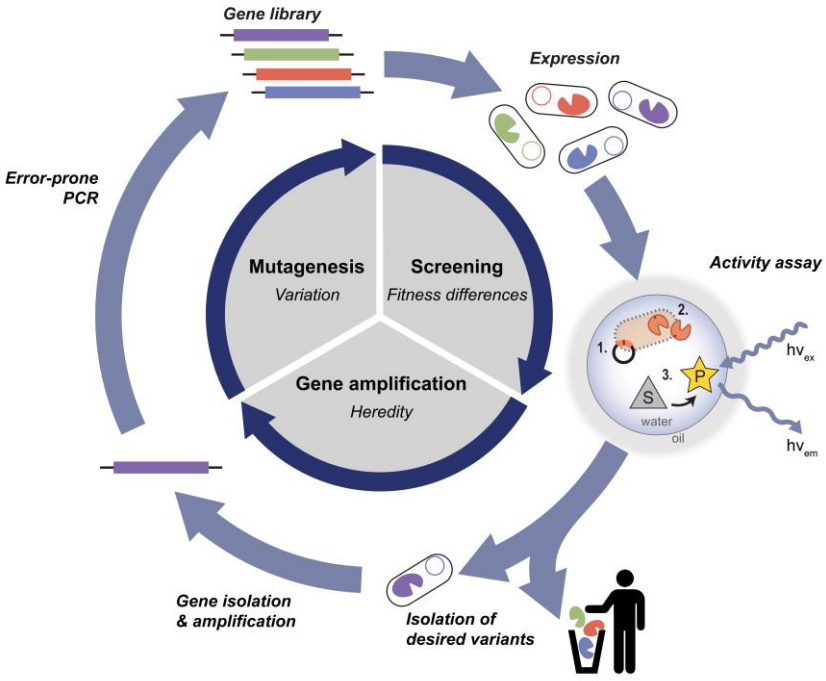


Fig. A-1: Cycles of directed evolution. Directed evolution follows a Darwinian process in three stages: introduction of random variation, selection for fitness differences, and amplification of successful variants. These three stages complete a cycle of directed evolution in the lab, with assays and molecular biological tools approximating natural evolution: Screening for fitness can be performed by measuring enzyme activity in a droplet microfluidic assay, the DNA of successful variants can be amplified recombinantly in cells, random variation can be introduced via error-prone PCR.

Directed evolution results in a stepwise accumulation of mutations conceptualised as a "walk through sequence space" from neighbouring sequence to neighbouring sequence, along an additional axis of protein fitness. It explores the local context of a protein sequence by randomly sampling many adjacent sequences for increased function, selecting a path towards higher activity visually represented as an ascend towards a peak in the fitness landscape. A fitness landscape reduces the overwhelming number of dimensions of the underlying sequence space into two dimensions, still arranging close sequences together, and adding a third dimension corresponding to variant fitness (Fig. A-2b). The concept of a fitness landscape was used in 1932 by Sewall Wright to conceptualize Natural evolution (Wright, 1932), but it also is immensely useful to visualize concepts in directed evolution and protein engineering. In the context of protein engineering, fitness is commonly defined as a function of interest, e.g. the enzymatic activity in defined reaction conditions, instead of organismal fitness in the sense of the production of viable offspring. A conventional directed evolution campaign then follows a single trajectory, at each point testing multiple randomly generated variants for higher fitness and selecting the very best variant to repeat the cycle (Fig. A-2b). It thus explores many variants at each step but follows only one evolutionary trajectory to reach higher fitness.

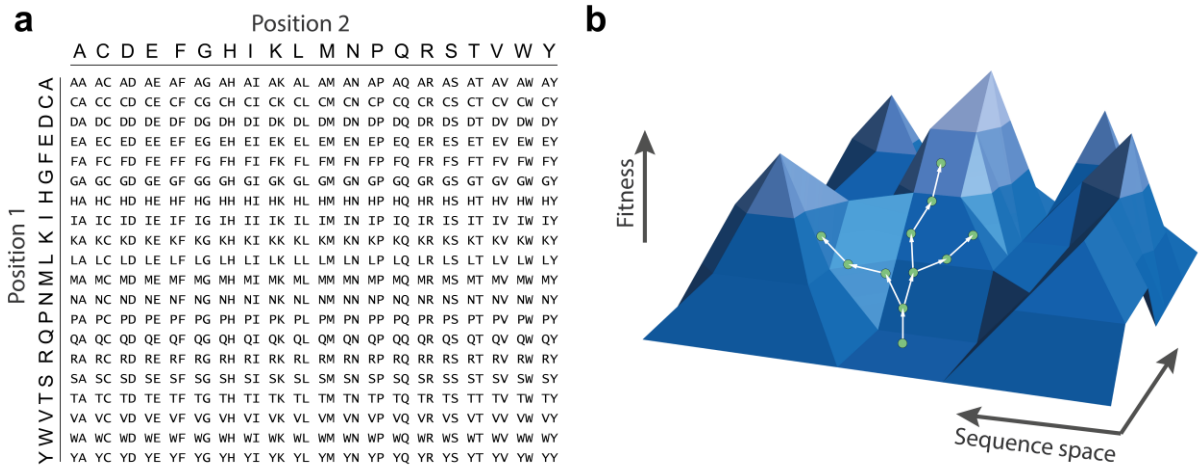


Fig. A-2: Protein sequence space and fitness landscape. (a) *A small protein sequence space.* A sequence space of a theoretical two amino acid protein. All 400 possible combinations can be listed in a two-dimensional graph. **(b)** *Fitness landscape showing multiple possible trajectories.* The high dimensional sequence space of a full-length protein is projected into two dimensions, where similar sequences are close together. The fitness of each variant is added in a third dimension, shaping a 'landscape'. Directed evolution can be conceptualized as a stepwise exploration of this fitness landscape. Variants are tested for increased activity, the best variant taken into the next round of randomization and screening.

The shape of a protein's fitness landscape is not smooth leading to just one peak. Fitness landscapes have extensively been characterized as "rugged"; a result of pervasive epistatic

interactions (Sarkisyan *et al*, 2016; Starr *et al*, 2018). Epistasis describes non-linear interactions, or in the case of a protein sequence, an unexpected effect if two or more mutations are combined (Fig. A-3). Such intra-gene epistasis between mutations results in a fitness landscape with many distinct local optima that are reachable by different evolutionary trajectories. This context-dependence of mutation limits protein engineering by reducing predictability and transferability of beneficial mutations (Miton & Tokuriki, 2016).

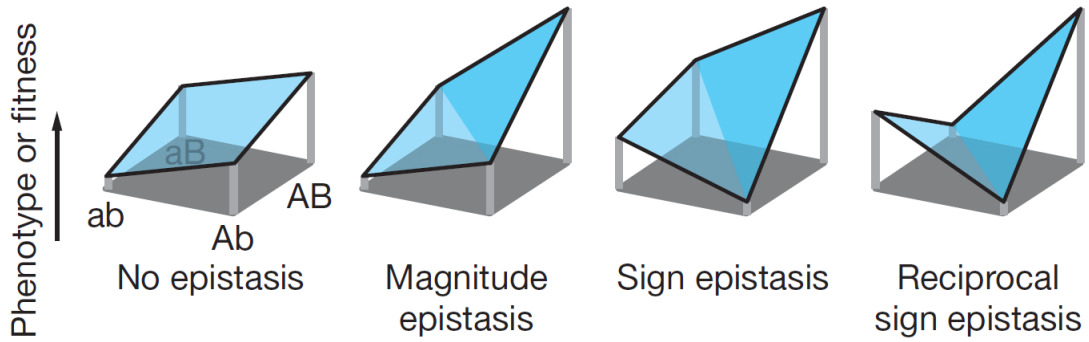


Fig. A-3: Epistasis results in the context-dependence of mutational effects. If no epistasis is present, mutation 'a' to 'A' yields the same effect indifferent of genetic background (being either 'b' or 'B') and 'AB' is the linear combination of both mutations. In magnitude epistasis, the extent of the effect is increased when both mutations are combined, while in sign epistasis the sign of the fitness effect changes. If the sign of fitness effect changes for both mutations, the effect is termed reciprocal sign epistasis, which in the context of protein engineering would mean that two deleterious mutations combined have a beneficial effect. Such reciprocal sign epistasis could result in a fitness landscape with many peaks, as the reversal of just one mutation would leave the protein unfunctional. Figure from Poelwijk *et al*, 2007.

The engineering of proteins, such as directed evolution to alter substrate-specificity of an enzyme and generate a valuable biocatalyst, is dependent on the choice of starting point. Usually, an enzyme is poorly adapted to an industrial process and needs to be engineered to perform in un-natural conditions and with new substrates. But how does a protein engineer choose a starting point to use for further engineering that has a high potential for adaptive evolution? In the context of fitness landscapes, this problem can be translated into identifying the points in valleys that have the steepest access to high peaks of the new function. Most studies of protein fitness landscapes, however, focus on already adapted enzymes and thus study the peaks of fitness landscapes (see also chapter A.4). Current knowledge about the factors influencing a protein's potential for adaptation – its evolvability – is outlined in the following chapter.

A.1.2. The Potential for Adaptation

The classical view of enzymes with one structure related to one substrate, as a key fits into a lock (Fischer, 1894), has since been amended with the realization that enzymes exhibit catalytic promiscuity. The ability of enzymes to catalyse other reactions than their main activity has been postulated to provide stepping stones for the evolution of new functions (Jensen, 1976; O'Brien & Herschlag, 1999). Such secondary activities can shorten the evolutionary distance to a variant that provides a selective advantage, such as the breakdown of a new xenobiotic, increasing overall evolvability. This functional promiscuity has been linked to a dynamic view of proteins exhibiting conformational flexibility (Tokuriki & Tawfik, 2009a).

Evolvability – defined here as the potential for adaptive evolution of a protein – is high when a fitness peak is close in sequence space, while other sequences might need to acquire multiple permissive mutations before large changes in fitness can be attained. Such permissive mutations are either stabilizing mutations, which can permit functionalizing but destabilizing mutations to become fixed – which is often termed global or threshold epistasis – or specific epistasis for example prohibiting the mutation of a key active site residue (Starr & Thornton, 2016). Evolvability thus manifests itself experimentally as a high proportion of improved variants during directed evolution, e.g. as a result of a 'stability buffer', as well as a high amplitude of functional improvement per step, e.g. by enabling epistatically productive combinations of mutation. These effects are quantifiable and permit an estimation of the evolvability of an enzyme from sampling variants of a library (Boyer *et al*, 2016).

Together, there are three properties of enzymes that relate to their evolvability: Stability, flexibility, and epistatic entrenchment. Investigation of these factors can lead to principles guiding the identification of good starting points for protein engineering. An evolvable starting point has to be identified in any protein engineering campaign, with general guidelines being rare (Trudeau & Tawfik, 2019).

A.1.2.1. Mutational robustness as a result of stability

Proteins tend to be marginally stable and mutations, on average, are destabilizing (Tokuriki *et al*, 2007). A protein, however, must be able to accept mutations and maintain its core structure and function in order to evolve. Consequently, excess stability should convey mutational robustness and thus increase evolvability by enabling acquisition of more functionalizing but destabilizing mutations (Tokuriki & Tawfik, 2009b). A stable protein can acquire more mutations before its stability is compromised in a way that diminishes activity and fold, thus

having greater access to potential functionalizing mutations (Fig. A-4). The stability threshold is the reason for prevalent unspecific epistasis in proteins, which occurs when a functionalizing but destabilizing mutation cannot be accepted on its own but must be compensated by any generally stabilizing yet functionally neutral mutation (Starr & Thornton, 2016; Bershtein *et al*, 2006).

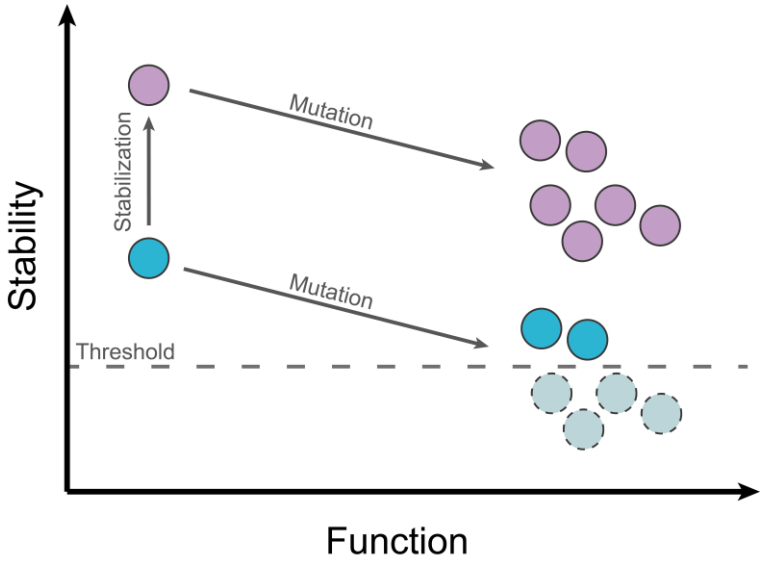


Fig. A-4: Stability and mutational robustness. A 'stability buffer' can convey mutational robustness leading to a greater number of accepted, generally destabilizing, mutations. By accepting more mutations, the chance of acquiring functionalizing mutations also increases, resulting in heightened evolvability.

Mutational robustness due to stability was proven by early protein engineers (Bershtein *et al*, 2006; Bloom *et al*, 2006), leading to a general rule-of-thumb to start engineering with very stable enzymes, such as proteins from thermophile organisms (Besenmatter *et al*, 2007; Finch & Kim, 2018). However, stabilizing mutations do not seem necessary to evolve active enzymes (Knies *et al*, 2017) and enzymes can be both stable *and* active (Giver *et al*, 1998). Furthermore, stability does not always correlate to mutational robustness, which can be seen with highly evolvable but instable and disordered viral proteins (Tokuriki *et al*, 2009). Nonetheless, some form of stabilization is usually employed in successful protein engineering campaigns (Romero & Arnold, 2009).

Protein stability itself can be grouped into thermodynamic stability and kinetic stability, where thermodynamic stability relates to 'conformational stability' or more specifically to the free energy difference between native and unfolded states (Tokuriki & Tawfik, 2009b). It is typically measured by proxy via the melting point in heat-denaturation experiments. Kinetic stability, on the other hand, relates to foldability or the probability of folding intermediates

leading to aggregation in the folding pathway (Sanchez-Ruiz, 2010). Kinetic stability is often measured by implication in soluble expression level compared to aggregated protein at expression temperature. In most proteins, thermodynamic and kinetic stability are connected (Mayer *et al*, 2007).

A.1.2.2. The role of flexibility in the adaptation to new substrates

The hypothesis that functional promiscuity of proteins stems from protein dynamism was formalized by Tokuriki and Tawfik in 2009. Proteins are conformationally dynamic and transition between an ensemble of substates. A less frequently occupied substate might show a productive binding pose with a non-native ligand and thus enable the catalysis of a promiscuous substrate (Fig. A-5). Mutations along an evolutionary trajectory could freeze out unproductive substates, increasing the overall activity with the new substrate (Tokuriki & Tawfik, 2009a). This model of evolution of new functions via shifting the equilibrium of protein conformational substates also finds an explanation for the weak negative trade-offs previously observed between existing and new function (Aharoni *et al*, 2005b; Khersonsky *et al*, 2006). While the increase in frequency of a scarcely populated substate from 1% to 10% can increase the corresponding promiscuous activity 10-fold, the relative occupancy of the main substate would only be marginally affected, e.g. reducing occupancy from 60% to 51%, equalling a loss in main activity of 15%.

Experimental evidence for this theory has accumulated in recent years (Campbell *et al*, 2018). Mutations far from the active site can have a great effect on enzyme activity, as shown in the earliest experiments of directed evolution, and can do so by changing the protein dynamics (Buller *et al*, 2018). Experimental directed evolution or ancestral sequence reconstruction can be used to study the effect of flexibility in protein evolution. For example, a reconstructed ancestral sequence of a non-catalytic solute-binding protein as well as ancestral intermediates along the evolutionary trajectory towards present-day cyclohexadienyl dehydratases was investigated with a variety of methods. A wide-open state was only present in the non-catalytic binding proteins and was steadily frozen out by remote mutation in the trajectory to the extant catalytic relatives (Kaczmarski *et al*, 2020). A similar effect can be seen in experimental directed evolution: The directed evolution of a designed retro-aldolase resulted in a 4,400-fold activity increase (Giger *et al*, 2013), which was further improved 30-fold by ultrahigh-throughput droplet microfluidic screening (Obexer *et al*, 2017). Long molecular dynamics simulations revealed the populations of substates showing a productive binding pose shift with

the evolutionary intermediates until all substates provided a catalytically productive pose in the final evolved variant (Romero-Rivera *et al*, 2017).

Together, the increasing body of work on the importance of protein dynamism in the evolution of new function culminates in a somewhat difficult challenge: identifying a flexible starting scaffold that remains robust in the overall fold, leading to a stable and evolvable protein.

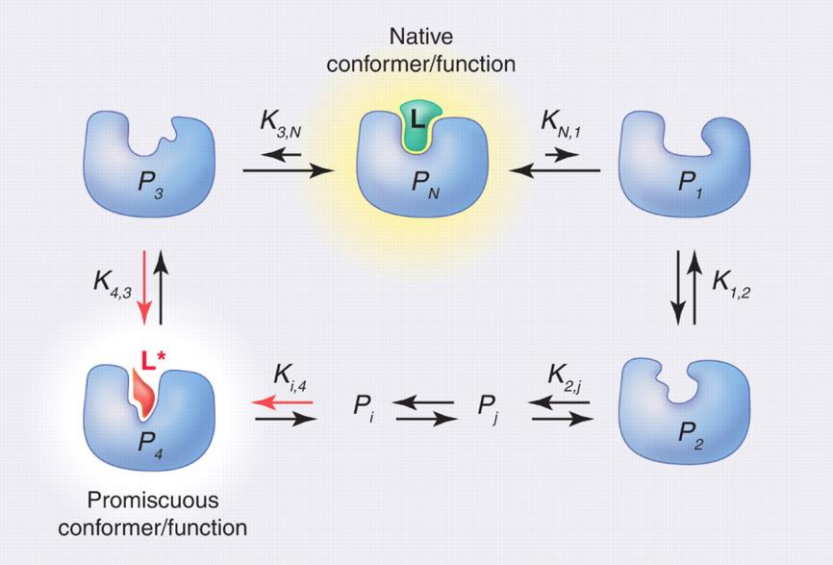


Fig. A-5: Protein dynamism and the evolution of new function. A protein exists as an ensemble of conformations, with each substate or conformer being a structural variation of different probability. The native state P_N is most frequently assumed and binds the native ligand *L*. Less frequent substates might enable promiscuous functions, such as the promiscuous interaction of P_4 with *L**. The equilibrium of substates shifts during the evolution of a promiscuous function so that the new productive substate is assumed more frequently, increasing the protein's promiscuous function. Figure from Tokuriki & Tawfik, 2009a.

A.1.2.3. Entrenchment by cryptic genetic variation and epistatic ratchets

Epistasis shapes the fitness landscapes of proteins and determines which paths to higher functions are accessible. Epistatic interactions among mutations are very frequent and thus each new mutation is likely contingent on the previous set of mutations, which itself might become entrenched by further subsequent mutations (Starr *et al*, 2018). The extent to which mutations in proteins interact is immense: 30% of multiple random mutations in GFP showed epistatic interaction (Sarkisyan *et al*, 2016) and 75% of mutations along the historical evolutionary trajectory of Hsp90 were either contingent on the previous state or entrenched by a later state (Starr *et al*, 2018). The implication of permissive mutations becoming entrenched by later contingent mutations leads to an irreversibility in evolution. Entrenchment and contingency continue to daisy chain with more mutations leading to a network of epistatic

interactions, so that experimental reverse evolution is likely to lead to a different solution instead of sequence reversal (Kaltenbach *et al*, 2015).

By which mechanism does the current epistatic state relate to a protein's evolvability specifically? In a recent example, differential evolvability of two homologue enzymes was related back to specific epistatic interactions instead of general differences in stability or flexibility. While one homologue enzyme quickly acquired a new function due to a crucial active site mutation, this active site mutation could not be acquired in the other homolog. Slightly different active site geometry resulted in steric hindrance with the strongly activating mutation, resulting in only marginal improvements over the course of 10 rounds of directed evolution (Baier *et al*, 2019). The evolvability of one starting point was thus much higher than the other, with the reason for it being hard to identify '*cryptic'* epistatic effects emerging from the underlying sequence context, a concept known in genetics as *'cryptic genetic variation'* (Paaby & Rockman, 2014).

The high complexity and frequency of interaction in protein mutations renders the evolvability of a protein, or even just the impact of single mutations, very hard to impossible to predict, although recent deep neural networks have shown potential by extracting patterns of millions of natural sequences for state-of-the-art performance (Alley *et al*, 2019).

For the time being, limits to evolvability by cryptic epistatic states can be circumvented by increased screening capacity and reduction of selection pressure. In so-called neutral drift, the selection pressure is removed or reduced and many variants with wild-type–like mutations are selected, which tend to accumulate stabilizing consensus mutations (Bershtein *et al*, 2008). This pool of neutral variation can be used as a collective starting point for adaptive evolution, increasing the chances to enable high fitness improvements from at least one of the variants contained in the pool, effectively broadening the available starting points for evolution. Small but intense neutral drift libraries with many mutations (Bershtein *et al*, 2008) or larger but less mutated pools of variants (Zheng *et al*, 2019) can be used to increase the likelihood of adaptive evolution and circumvent epistatic ratchets. Epistatic ratchets arise from seemingly innocuous variation preventing mutation in key residues (Bridgham *et al*, 2009).

While neutral drift has been experimentally shown to generate more robust and evolvable proteins, it requires an ultrahigh-throughput screening to select many variants, which is not always available in protein engineering. Thus, one of the biggest questions in protein engineering remains: what determines the evolvability of an individual starting point and can

it be predicted? A good model to study the evolvability of a new function could be found in the engineering of amine dehydrogenases from amino acid dehydrogenases.

A.2. Amine Dehydrogenases

Enzymes as industrial biocatalysts provide certain advantages: not only do they work in sustainable and energy-efficient conditions, such as mild aqueous solutions at low temperatures, but also enable conversions with remarkable selectivity and specificity (Walsh, 2001). High regio- and enantioselectivity have led to the wide adoption of biocatalysts in the synthesis of many pharmaceuticals as well as a recent expansion into the fine and speciality chemicals field (Wu *et al*, 2020; Woodley, 2020).

Much of the rising use of biocatalysts in the chemical industry is due to improvements in the methods used to discover and engineer new enzymes (Devine *et al*, 2018). One such example is found in the recent engineering of an imine reductase to catalyse a key step in the synthesis of a lysine-specific demethylase-1 inhibitor, a drug currently in Phase II clinical trials (Schober *et al*, 2019). An expensive multi-step chemical synthesis was replaced with a simple enzyme catalysed step to yield a sustainable process at reduced cost. For the process to reach desired performance, however, enzyme engineering was necessary. After three rounds of screening single-site saturation and recombination libraries, an enzyme with 36,000-fold improvement over wild-type was found which could be used in an efficient hydrogen borrowing cascade (Fig. A-6). The hydrogen borrowing cascade makes smart use of a keto-reductase to prepare the substrate for the engineered imine reductase, removing further chemical steps as well as regenerating the co-factor NADPH (Schober *et al*, 2019).

A.2.1. Overview of Enzymatic Ways to Chiral Amines

As seen in the example above, chiral amines are one of the most important groups in active pharmaceutical ingredients and fine chemicals that can well be made by biocatalysts. Chiral amines are found in 40% of active pharmaceutical ingredients with an estimated total revenue of 30 billion USD (Breuer *et al*, 2004; Schrittwieser *et al*, 2015). This trend has established itself with more than 90% of newly approved or current top-selling small molecules including amine groups (Wu *et al*, 2020). Enzymatic synthesis here again impresses with better atom economy, stereoselectivity and sustainable reaction conditions (Grogan, 2018).

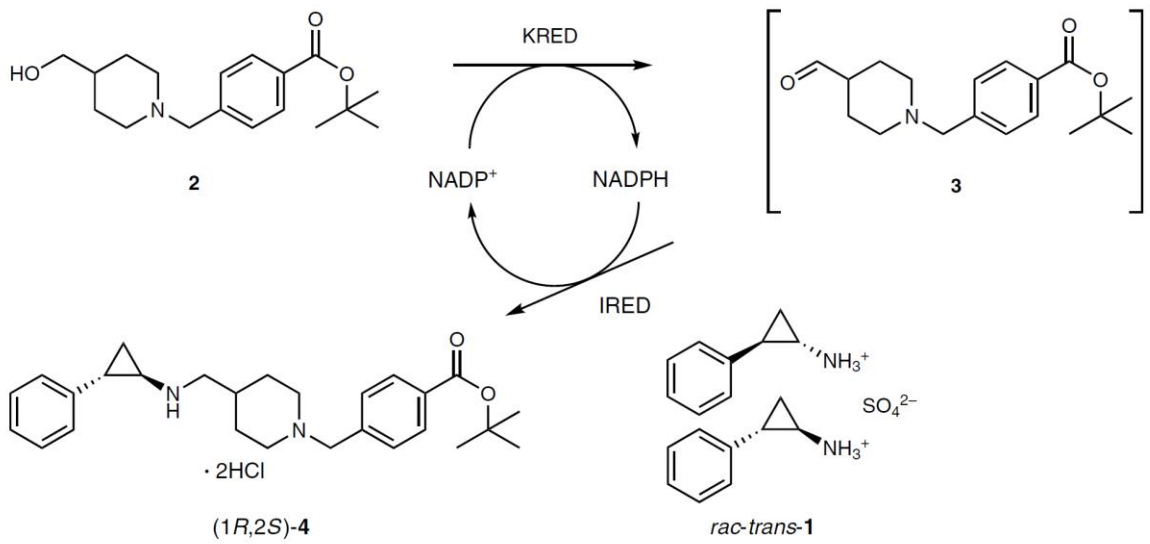


Fig. A-6: A hydrogen borrowing cascade for the synthesis of a leucine-specific demethtylase-1 inhibitor. The redox-neutral imine reductase (IRED) / keto-reductase (KRED) cascade enables the efficient and sustainable synthesis of the secondary amine 4 from the alcohol 2, generating a precursor for the synthesis of the drug GSK2879552. Figure from Schober *et al*, 2019.

However, there is not one enzymatic way to chiral amines. Instead, many different approaches have been realized in recent years (Fig. A-7). Monoamine oxidases can catalyse the enantioselective oxidation of amines to imines. This is employed in the deracemization of amines, effectively enriching the wanted chiral amine (Ghislieri *et al*, 2013). Kinetic resolution of racemic amines can also be performed with lipases, but it is usually less favoured than direct chiral synthesis because expensive metal catalysts and harsh conditions are needed to continuously racemize the substrate and circumvent the otherwise maximal yield of 50% (Paetzold & Bäckvall, 2005). Thus, transaminases have been used extensively for the direct asymmetric synthesis of primary amines from ketones (Wu *et al*, 2020). Additionally, a sacrificial amine donor is needed in high excess to drive the thermodynamic equilibrium which again complicates product separation and drives up cost (Höhne & Bornscheuer, 2009). Ammonia lyases can directly create chiral amines by addition of ammonia to an unsaturated substrate, circumventing many of the above-mentioned issues. However, ammonia lyases possess a very narrow substrate scope for a few amino acids and are mainly used to produce aspartic acid derivates or for the generation of aspartic acid itself from ammonia and fumarate as a precursor of the sweetener aspartame (Wu *et al*, 2020; Parmeggiani *et al*, 2018).

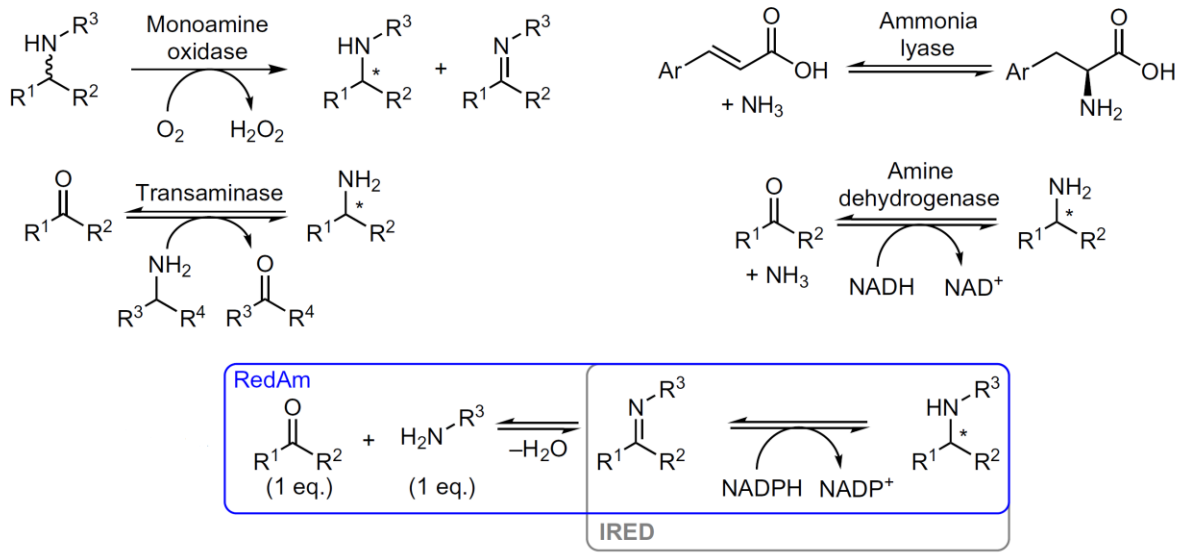


Fig. A-7: Selection of enzymatic routes to chiral amines. Many different approaches to generate chiral amines have been pursued, with reductive aminases (RedAm), a subgroup of imine reductases (IRED), showing a lot of promise (Marshall *et al*, 2020). Amine dehydrogenases, on the other hand, were engineered from amino acid dehydrogenases in 2012 and have since shown a high potential as biocatalysts with a simple reaction set up and high efficiencies for primary amines (Grogan, 2018; Abrahamson *et al*, 2012). Figure from Aleku *et al*, 2017.

Imine reductases have been the focus of much recent biocatalysis literature. Imine reductases can catalyse the NADPH-dependent asymmetric reduction of prochiral cyclic imines, but a recent finding showed that some variants can also catalyse the imine formation, leading to the full reductive amination of ketones with broad substrate specificity (Aleku *et al*, 2017). These novel variants were called reductive aminases and have since attracted considerable attention (Wu *et al*, 2020; Grogan, 2018). Structure-activity relationships within the imine reductase family are just now starting to be studied, aiming to understand the link between variants that can and cannot catalyse full reductive amination (Montgomery *et al*, 2020).

Finally, amine dehydrogenases (AmDHs) have shown great versatility for the generation of primary amines, promising direct chiral synthesis from a cheap nitrogen donor (ammonia) and many prochiral ketone substrates (Knaus *et al*, 2017).

A.2.2. Generation of Amine Dehydrogenases

Amino acid dehydrogenases (AADHs) catalyse the formation of amino acids from keto acids in the presence of ammonia and NADH (Brunhuber *et al*, 2000). Investigating the substrate specificity in more detail, Abrahamson and colleagues have found that AADHs can be engineered to accept the corresponding amine substrate via active site mutagenesis

(Abrahamson *et al*, 2012). Two key active site mutations in the carboxylate binding pocket are required to completely change the substrate specificity and generate an AmDH (Fig. A-8).

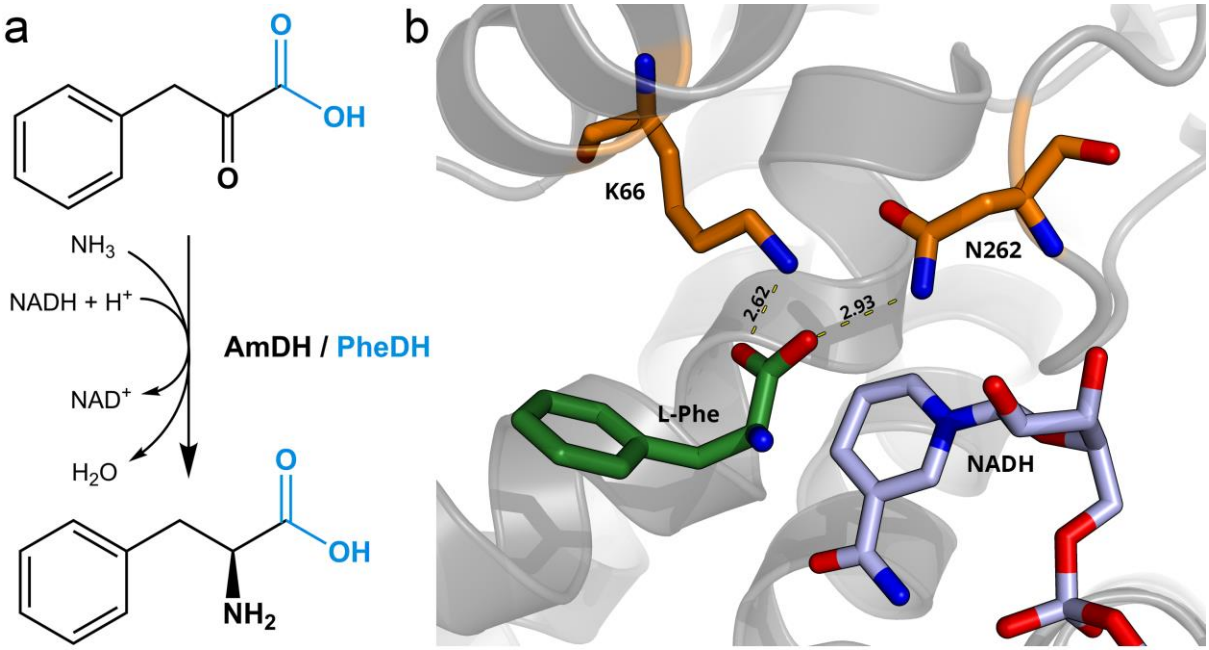


Fig. A-8: Engineering of amine dehydrogenases from amino acid dehydrogenases. **(a)** *Reductive amination.* Phenylalanine dehydrogenases (PheDH) catalyse the reductive amination of the keto-acid phenylpyruvic acid to the amino-acid L-phenylalanine. The corresponding AmDH is engineered to accept a substrate without the carboxylate shown in blue, i.e. converting phenylacetone to (*R*)-amphetamine. **(b)** *Phenylalanine dehydrogenase active site.* Crystal structure of *Rhodococcus sp.* M4 PheDH in complex with phenylalanine and NAD^+ (PDB ID 1C1D). Residues K66 and N262 are highlighted, which were shown to coordinate the carboxyl group of the natural substrate and on mutation enable the conversion to an AmDH.

Since then, this concept has been applied to leucine dehydrogenases (Abrahamson *et al*, 2012), phenylalanine dehydrogenases (Abrahamson *et al*, 2013; Ye *et al*, 2015) and chimeric versions (Bommarius *et al*, 2014) to broaden the substrate scope of these catalysts. The success of engineered AmDHs has also prompted a search for natural AmDHs, which came to fruition recently (Mayol *et al*, 2019). Nevertheless, the engineering of AmDHs from AADHs poses an interesting challenge to protein evolution, generating a complete shift to a new function while producing an industrially relevant enzyme class. AmDHs would especially benefit from further increases in activity which is likely to be limited by their dynamic properties, further posing a challenge to traditional protein engineering (Pushpanath *et al*, 2017).

A.3. Ultrahigh-Throughput Screening for Enzyme Function in Droplets[1]

Although directed evolution has established itself as a key method for protein engineering, to some it remains a source of frustration. Not every attempt leads to certain success and achieving improvements relies on a chance step within a vast and rugged fitness landscape (Romero & Arnold, 2009). Consequently, directed evolution is a numbers game: not only assay conditions and quality are critical for the success of a directed evolution campaign, but also the number of variants screened. Testing more variants increases the chance of finding improved variants. Ultrahigh-throughput screening thus makes directed evolution more successful by sheer force of explorative power. A prominent example of this is the evolution of an artificially designed aldolase. Directed evolution by iterative screening in 96-well plates had initially led to a more than 4400-fold improvement of aldolase activity (Giger *et al*, 2013). However, further improvement was challenging due to increasingly rare beneficial mutations. This optimization plateau – a common feature of protein evolution achieving large improvements with few mutations early on, followed by diminishing returns and low improvements by subsequent mutation (Tokuriki *et al*, 2012) – could be overcome with ultrahigh-throughput screening. Droplet microfluidic screening of millions of variants generated a further 30-fold improvement resulting in a final evolved artificial aldolase that rivals natural enzymes (Obexer *et al*, 2017). Ultrahigh-throughput screening is best applied in such scenarios where hit rates are low: the higher throughput allows increasingly daring projects and library designs to be completed, such as metagenomics (Colin *et al*, 2015a), active site remodelling (Debon *et al*, 2019) or even large sequence-function mapping experiments (Romero *et al*, 2015).

One method to speed up assays and bring down cost is the miniaturization of reaction compartments to form separate water-in-oil emulsion droplets. Droplet microfluidics can scale down reaction volumes to the picolitre range, corresponding to a more than 10^7-fold reduction in assay volume compared to regular 200 µL wells in the 96-well format (Agresti *et al*, 2010; Kintses *et al*, 2012). Each compartment contains a single genotype, e.g. a single cell expressing an enzyme variant, combined with a substrate allowing for the enzymatic reaction to be monitored as the corresponding phenotype (Fig. A-9). For example, a library of gene variants can be used to transform *E. coli* cells, which are then singly encapsulated into droplets. In the

[1] This chapter is based on the publication Neun*, Zurek* *et al*, 2020. *: Equal contribution.

encapsulation step, substrate and lysis solution is added to each droplet so that the single cell can release the contained enzyme variant and reaction can progress to accumulate product as a detectable phenotype (Kintses *et al*, 2012). A directed evolution workflow using droplet microfluidics includes library generation as the first step, e.g. via error-prone PCR, followed by encapsulation, reaction progress and sorting of the most active variants. The genetic information of sorted variants is recovered and verified in a secondary assay, such as a 96-well plate lysate activity assay. Improved variants can then be subjected to another round of directed evolution, if necessary (Fig. A-9). Overall, this miniaturization of reaction volumes to picolitre droplets leads to reduced cost by decreasing reagent consumption and dramatic assay speed-up. Droplets can be generated at rates of 1-15 kHz, leading to anywhere between 28-430 million droplets per 8 h working day. An equally impressive number of 28-86 million droplets can be assayed and sorted in the same time frame, with common frequencies of 1-3 kHz for droplet sorters (Baret *et al*, 2009; Debon *et al*, 2019).

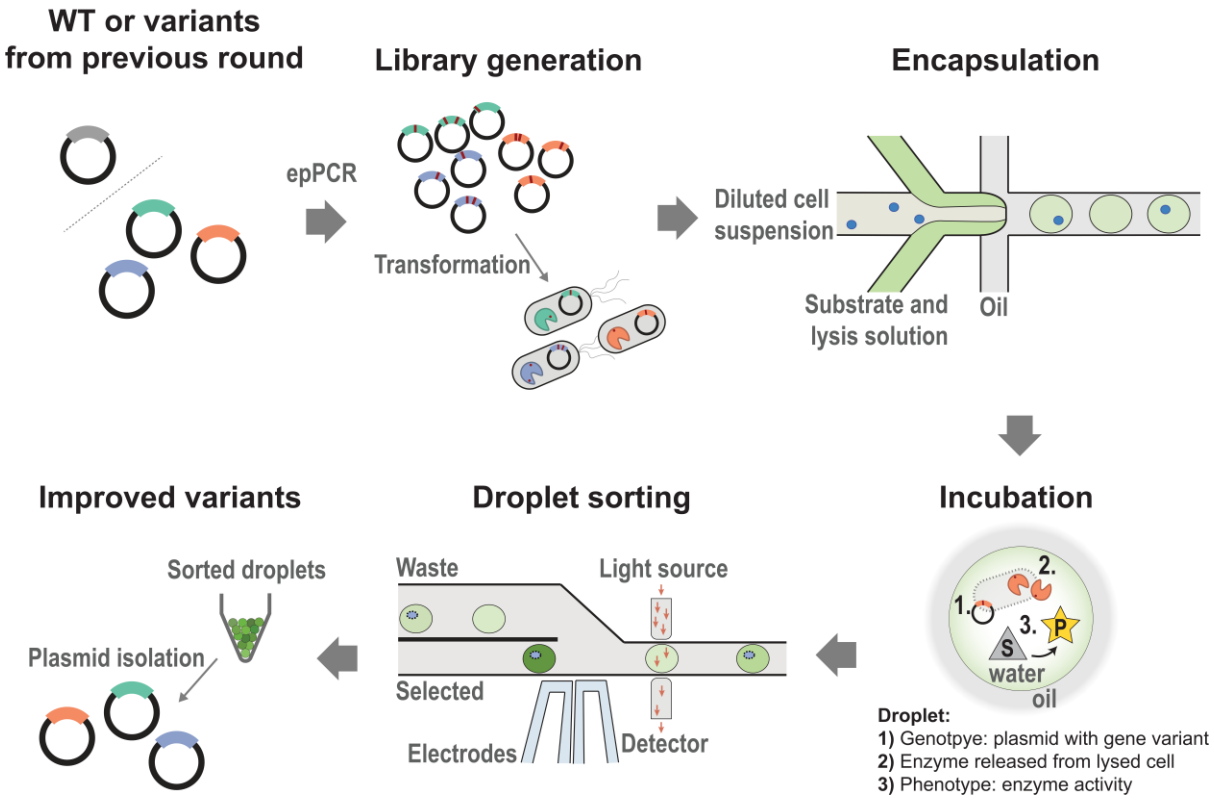


Fig. A-9: Droplet microfluidic workflow for enzyme engineering. The wild-type (WT) gene sequence or selected variant(s) from previous rounds of directed evolution are chosen as starting material. A library of variants is generated e.g. randomly by error-prone PCR and used to transform cells for expression. Each cell takes up one plasmid and thus expresses one enzyme variant. A diluted cell suspension, regulated so that no more than one cell is added per droplet, is co-encapsulated with substrate and lysis solution into monodisperse picolitre droplets. In each droplet, the cell is lysed and releases the enzyme variant. If the variant is active, reaction will progress and a detectable product accumulates within the droplet. A droplet sorter selects droplets with product concentration reaching a user-defined threshold, resulting in the enrichment of active variants. Figure from Zurek *et al*, 2020c.

A.3.1. Detection Modes

Droplet microfluidic screening distinguishes itself from selection by enabling the direct monitoring and quantification of a reaction product of individual variants. Selections, in which competition leads to enrichment of desired properties, have also been greatly enabled by water-in-oil emulsion droplets. The earliest experiments with droplets were not performed with microfluidic-generated monodisperse droplets, but used polydisperse emulsions. These artificial compartments of varying sizes still enabled assays that were not possible in conventional formats by removing crosstalk between variants. Now, enzymes that modify (Tawfik & Griffiths, 1998) or amplify (Ghadessy *et al*, 2001) their own gene in their separate compartment could selectively be enriched. Polydisperse droplets are attractive for their simple and fast generation, but they lack more gradual and quantitative readouts due to the varying droplet sizes, which might preclude detection of smaller activity differences. Still, polydisperse droplets are a useful tool for binary selection, in which the figurative needle is lost in the haystack of inactive enzymes (Colin *et al*, 2015b).

The microfluidic generation of monodisperse emulsions adds a level of control by generating droplets of equal volume. Additionally, as all droplets have the same size, further manipulation can easily be undertaken: monodisperse droplets have been used in elaborate multi-step workflows including steps such as droplet fusion (Mazutis *et al*, 2009), splitting (Link *et al*, 2004) or injection of a controlled volume of aqueous solution (Abate *et al*, 2010).

The first successful screening of an enzyme reaction in monodisperse droplets was established as fluorescence-activated droplet sorting by Baret *et al* in 2009. The set-up allowed the reliable assaying of millions of droplets for their fluorescence intensity in a day, sorting active or improved enzyme variants. There are three main experiments performed for enzyme activity: (i) Active enzyme is sorted from a dilution in inactive variant in an enrichment experiment, which is a stringent test of the assay and helps evaluate the readiness of the technology (Baret *et al*, 2009). (ii) In functional metagenomics active enzyme is sorted from an environmental library, increasing demands on assay sensitivity to detect low promiscuous function of sub-optimally expressed enzymes (Colin *et al*, 2015a). (iii) In a directed evolution campaign a gradual increase in activity is selected, needing the most reliable and quantitative assays (Kintses *et al*, 2012). Fluorescence-activated droplet sorting has been used for both, directed evolution and functional metagenomics, in many different ways and assays (Colin *et al*, 2015b), and technical development still proceeds. Multi-laser set-ups have been developed, enabling the detection of multiple properties at once (Ma *et al*, 2018; Hung *et al*, 2020; Hengoju

et al, 2020) or pushing sorting speeds up to 30 kHz in more elaborate layouts (Sciambi & Abate, 2015).

Most often, reaction progress is monitored with direct fluorogenic substrates. However, not all reactions are assayable with a fluorescent readout and the need for chemical synthesis of custom fluorescent substrates could preclude its use in many molecular biology labs. Furthermore, the non-natural substrate is only a mimic of the real-world substrate. The usually large hydrophobic fluorophores can lead to specialization to the assay substrate instead of the substrate of interest, an unwanted effect resulting from basic rule of directed evolution: "*you get what you screen for*" (You & Arnold, 1996). In some cases, indirect assays can enable an unbiased fluorescent readout, e.g. by detection of a co-factor in a cascade reaction (Debon *et al*, 2019). Cascade detections require extensive setup and finetuning of compatible reaction conditions, such as buffer composition, temperature and pH. Consequently, alternative detection modes were developed to extend the number of detectable reactions and enable more unlabelled detections. A summary of all detection modes that were miniaturized to the scale of droplet microfluidics is shown in Table 1.

Table 1: Detection modes for droplet microfluidic enzyme assays. Only the main or most important technical innovation is stated as reference. SERS: Surface-Enhanced Raman Scattering. Optical methods grouped as light blue. Passive methods grouped as light yellow. Low: 1-10. Mid: 10-100. High: 100-1000. Table from Neun, Zurek *et al*, 2020.

Detection mode	Sensitivity	Frequency	Droplet size	Applied to screening?[a]	Introduced by
Fluorescence intensity	Low nM	Low kHz	Low pL	☑	Baret *et al*, 2009
Fluorescence polarization	Low nM	n.a.	Mid nL	✕	Gielen *et al*, 2017
Fluorescence lifetime	Low nM	Low Hz	High pL	✕	Hasan *et al*, 2019
Absorbance	Low µM	High Hz	High pL	☑	Gielen *et al*, 2016
SERS	n.a.	<< 1 Hz	Mid nL	✕	Hassoun *et al*, 2018
Light scattering	n.a.	High Hz	High pL	✕	Liu *et al*, 2016
Image analysis	n.a.	Low Hz	Low nL	✕	Girault *et al*, 2017
Mass spectrometry	Mid µM	Low Hz	Mid nL	✕	Holland-Moritz *et al*, 2020
Electrochemical	Low µM	Low Hz	Mid nL	☑	Goto *et al*, 2020
Surface tension	High nM	Mid Hz	High pL	✕	Horvath *et al*, 2019
Density	n.a.	n.a.[b]	Mid nL	✕	Schmitt *et al*, 2018

[a]Applied in a monoclonal functional metagenomics or directed evolution experiment.
[b]This passive selection is theoretically only limited by the droplet formation frequency.

A lot of the assays common in molecular biology labs are based on absorbance measurements. The detection of absorbance in microfluidic droplets was enabled by on-chip fibre optics (Mao *et al*, 2015) and developed into absorbance-activated droplet sorting (AADS, Gielen *et al*, 2016). Compared to fluorescence sorters, larger droplets were necessary to find a compromise between speed and sensitivity, resulting in a working path length of 50 µm. The resulting sorter enabled the label-free detection and directed evolution of phenylalanine dehydrogenase activity by measuring the absorbance of the co-factor NADH down to a 10 µM detection limit at 0.3 kHz (Gielen *et al*, 2016). The layout of the flow-focussing device and sorter are shown as the photomask and manufactured chip images in Fig. A-10. Since the development of AADS, the only alternative detection mode that has been successfully applied within the stringent requirements of directed evolution has been the electrochemical detection of NADH for the evolution of an isocitrate dehydrogenase (Goto *et al*, 2020).

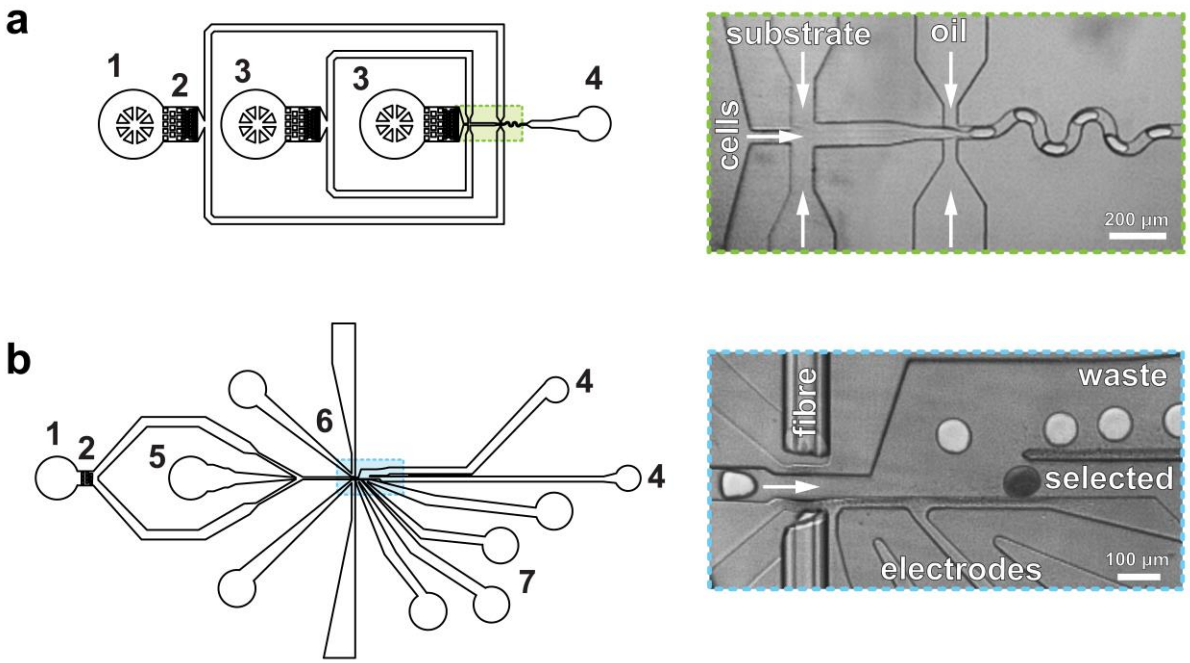


Fig. A-10: Design of microfluidic devices for AADS. (a) *Flow-focussing device for droplet generation.* Droplets are generated as a co-encapsulation of diluted cell suspension and substrate in fluorinated oil. Droplets of ~300 pL size can be generated at > 1000 Hz. Chip dimensions at the flow focussing site are 50 µm width and 80 µm height. 1: Inlet for fluorinated oil phase. 2: Passive filters. 3: Inlet for two aqueous phases. 4: Outlet. **(b)** *Absorbance-activated droplet sorter.* Optical fibres are inserted into the chip to measure the absorbance of passing droplets (50 µm path length) at up to 300 Hz. The image shows a representative snapshot of a sorting event with false colours superimposed to illustrate sorted and unsorted droplets. 5: Inlet for droplets. 6: Channel for optical fibre. 7: Salt-water (5 M NaCl) electrodes. Figure from Zurek *et al*, 2020c.

Other detection modes have shown proof-of-concept abilities to expand the scope of detectable assays in droplet microfluidics, but currently lack the versatility or sensitivity for application

to directed evolution and functional metagenomics. For example, selections based on buoyancy of droplets impacted by catalase-mediated gas formation (Schmitt *et al*, 2018), changes in surface tension due to enzymatically altered pH values (Horvath *et al*, 2019) and label-free mass spectrometric detection of a transaminase substrate (Holland-Moritz *et al*, 2020) have been able to detect enzymatic activity in microfluidic droplets. For these methods, however, increases in sensitivity are necessary to make them applicable to directed evolution and more stringently enable the detection of low promiscuous activities in functional metagenomics or the evolution of novel enzyme activities.

A.3.2. Available Enzymatic Assays

With all the available detection modes discussed above, a lot of different enzyme assays have been performed in droplets. A literature overview of chemical transformations that are detectable in droplets is shown in Table 2. Most assays use a fluorescence-activated droplet sorter or, less frequently, a fluorescence-activated cell sorter. If no dedicated droplet sorter is available, commercial flow cytometric cell sorters can be used when droplets are re-encapsulated into an aqueous solution, forming water-oil-water double emulsions (Zinchenko *et al*, 2014). Most enzymes that have been assayed in droplets are hydrolases, as fluorogenic substrates are often available or easily synthesized by replacing the leaving group with a fluorophore. Creative assay design has further enabled the fluorescent detection of a wide range of different enzyme activities, e.g. by detecting co-product with an enzymatic detection cascade (Debon *et al*, 2019), coupling enzyme activity directly to cell growth (Femmer *et al*, 2020) or read-out to a reporter strain (Siedler *et al*, 2017). Here, only assays which have been shown to work monoclonally, i.e. containing one genetic element per droplet, and have been applied to an enrichment experiment proving technology readiness and applicability, are listed (Table 2).

A.3.3. Consequences of Ultrahigh-Throughput Screening

Primarily, droplet microfluidics perform the same assays as they are used in conventional formats but much faster, with less reagent consumption, and in some cases more sensitively, making biocatalyst discovery and engineering more efficient. However, the great increase in screening throughput afforded by droplet microfluidics enables completely novel approaches in protein evolution and engineering too. Four such directions are:

Table 2: Types of enzyme reactions detected in droplets. Only assays that worked monoclonally and at least in an enrichment experiment are shown.

Catalytic reaction		Screening host				Format				Screening type			Emulsion				Sorting						Reference
Type	**Reaction**	**Bacteria**	**Yeast**	**Filamentous fungi**	***In vitro***	**Whole cell**	**Lysate**	**Display**	**Secretion**	**Enrichment**	**Directed evolution**	**Functional metagenomi**	**polydisperse**	**mono-disperse**	**single**	**double**	**DNA amplification**	**FACS**	**FADS**	**AADS**	**Electrochemistry**	**Density**	
Hydrolase	β-Galactosidase				✓						✓		✓			✓		✓					Mastrobattista *et al*, 2005
		✓				✓				✓				✓	✓				✓				Baret *et al*, 2009
					✓					✓				✓	✓				✓				Fallah-Araghi *et al*, 2012
	β-Glucosidase	✓					✓				✓			✓	✓				✓				Romero *et al*, 2015
	α-Amylase		✓						✓		✓			✓	✓				✓				Sjostrom *et al*, 2014
				✓					✓		✓			✓	✓				✓				Beneyton *et al*, 2016
	Cellulase	✓				✓				✓				✓	✓				✓				Najah *et al*, 2014
			✓					✓		✓				✓	✓				✓				Ostafe *et al*, 2014
	Phosphonate hydrolase	✓					✓				✓			✓	✓			✓					Kintses *et al*, 2012
	Phosphotriesterase				✓						✓		✓		(✓)			✓					Griffiths & Tawfik, 2003
		✓					✓				✓			✓	(✓)			✓					Fischlechner *et al*, 2014
		✓					✓					✓		✓	✓				✓				Colin *et al*, 2015a
	Organophosphate hydrolase	✓				✓					✓			✓		✓			✓				Gupta *et al*, 2011
	Esterase	✓				✓						✓		✓	(✓)			✓					Hosokawa *et al*, 2015
			✓					✓			✓			✓		✓		✓					Terekhov *et al*, 2017
		✓					✓				✓			✓	✓				✓				Ma *et al*, 2018
	Sulfatase	✓					✓				✓			✓		✓		✓					Zinchenko *et al*, 2014
		✓						✓			✓			✓	✓				✓				van Loo *et al*, 2019
	Thiolactonase	✓				✓					✓		✓			✓		✓					Aharoni *et al*, 2005a
Oxidoreductase	Peroxidase		✓					✓			✓			✓	✓				✓				Agresti *et al*, 2010
	Catalase	✓				✓				✓				✓	(✓)							✓	Schmitt *et al*, 2018
	Laccase	✓				✓				✓				✓	✓				✓				Beneyton *et al*, 2014
	Cyclohexylamine oxidase	✓					✓				✓			✓	✓				✓				Debon *et al*, 2019
	Phenylalanine DH	✓					✓				✓			✓	✓					✓			Gielen *et al*, 2016
	Isocitrate DH				✓						✓			✓	✓						✓		Goto *et al*, 2020
Aldolase	Retro-aldolase	✓					✓				✓			✓	✓				✓				Obexer *et al*, 2017, 2016
Transferase	Methyltransferase				✓					✓			✓		✓		✓						Tawfik & Griffiths, 1998
					✓						✓		✓		✓		✓						Lee *et al*, 2002
	DNA/XNA polymerase	✓					✓				✓		✓		✓		✓						Ghadessy *et al*, 2001; Pinheiro *et al*, 2012; Loakes *et al*, 2009; d'Abbadie *et al*, 2007
		✓					✓			✓				✓	✓				✓				Vallejo *et al*, 2019
		✓					✓				✓			✓		✓		✓					Larsen *et al*, 2016
		✓					✓				✓			✓	✓	✓		✓	✓				Nikoomanzar *et al*, 2019
Isomerase	Ornithine racemase	✓				✓					✓			✓	✓			✓					Femmer *et al*, 2020
Ribozyme	DNA ligase				✓						✓			✓	✓		✓						Paegel & Joyce, 2010
	RNA-cleavage				✓						✓			✓	✓				✓				Ryckelynck *et al*, 2015
					✓					✓				✓	✓				✓				Matsumura *et al*, 2016

(i) Synthetic compartments created in the form of droplets resemble artificial cells and can be used for screening completely *in vitro*. Single genes can be encapsulated in droplets, amplified and the protein of interest expressed via *in vitro* transcription and translation (Fallah-Araghi *et al*, 2012). These approaches currently remain underrepresented (Table 2), but de-coupling assays from host survival should enable the screening of even larger libraries and the evolution of toxic proteins.

(ii) Screening with ultrahigh-throughput methods leads to increasingly daring projects and library designs. Droplet microfluidics enable the sensitive and reliable recovery of rare '*one in a million*' events (Zinchenko *et al*, 2014), and thus support experiments that would otherwise require unreasonable luck. In a functional metagenomics experiment, less than 10 enzymes that showed promiscuous activities for new-to-nature reactions were identified out of million member libraries (Colin *et al*, 2015a). The low hit rate and low activities of promiscuous variants would preclude such efforts with slower and less sensitive technologies. In directed evolution, increasingly 'risky' and large library designs become feasible. Deleterious libraries of insertions and deletions or libraries with multiple mutations at once will have very low hit rates due to strong destabilizing effects (Emond *et al*, 2020; Sarkisyan *et al*, 2016). Such libraries, however, are likely to contain epistatic interactions that increase the functional improvements to more than what would be achievable in successive additive approaches. For example, simultaneous mutations were necessary to remodel the active site of a cyclohexylamine oxidase and enable a 960-fold increase in catalytic efficiency in just one round of screening (Debon *et al*, 2019).

(iii) Droplet microfluidics can speed up kinetic characterization. Conventionally, the pool of sorted variants after droplet screening has to be characterized in detail in microtiter plates, slowing down the overall process. Staying in droplets could automate obtaining kinetic parameters of many selected variants. The generation of concentration gradients in droplets was used to measure detailed Michaelis-Menten and inhibition enzyme kinetics (Gielen *et al*, 2015). Using droplets to speed up enzyme characterization could find more routine application in the future.

(iv) Combining ultrahigh-throughput screening with next-generation sequencing. By having a sequence readout of all selected variants, a map of the fitness landscape can be generated. Large datasets of sequence-function relationships based on microfluidic assays and next-generation sequencing have for example been established for single point mutations in a β-glucosidase (Romero *et al*, 2015) and a 48 amino acid domain of a DNA polymerase (Nikoomanzar *et al*,

2019). Generation of large datasets of sequence-function relationships in this way holds great promise for understanding and predicting protein evolution, as explained in the following chapter.

A.4. Mapping Sequence-Function Relationships

The combination of a high-throughput assay of protein function with high-throughput sequencing as a read-out is collectively known as 'deep mutational scanning' (Fowler & Fields, 2014). It enables the large-scale sequence to function mapping of proteins such as binders (Fowler *et al*, 2010), enzymes (Firnberg *et al*, 2014), or viral capsids (Ogden *et al*, 2019) as well as regulatory DNA elements such as promoters (Sharon *et al*, 2012). Deep mutational scanning works by generating a library comprising variation in the gene of interest, which is then fully sequenced before and after functional selection (Fig. A-11). Changes to the relative frequency in sequence count, e.g. an enrichment of a particular mutation post-selection, is directly related to the functional effect of that mutation in the assay (Fowler & Fields, 2014). In this way, very large datasets of mutational effects can be generated, with up to 600,000 variants measured in one assay (Fowler *et al*, 2010).

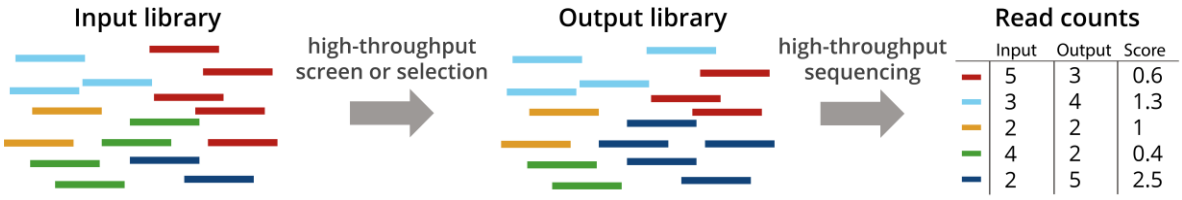


Fig. A-11: Deep mutational scanning generates functional values for many variants. A library of variants is generated and subjected to a high-throughput assay. Pre- and post-selection libraries are then quantified via high-throughput sequencing. Changes to relative frequencies in read counts of a mutation, such as enrichment, reflect the functional importance of that mutation.

The libraries of deep mutational scanning are often generated as comprehensive single site saturation libraries, i.e. mutating every codon to every other codon individually (Hietpas *et al*, 2011). However, multi-site saturation libraries (Wu *et al*, 2016), pairwise mutations (Olson *et al*, 2014), as well as multiple completely random mutations (Sarkisyan *et al*, 2016) have also been studied in deep mutational scanning. The functional assay is equally flexible: fluorescence-activated cell sorters (Sarkisyan *et al*, 2016) and droplet microfluidic screens (Romero *et al*, 2015) have been applied, although cell survival (Firnberg *et al*, 2014) or growth competition (Wrenbeck *et al*, 2017a) selections have been most frequently used for enzymes

(Table 3). The restricted availability of high-throughput enzyme assays has limited the application of deep mutational scanning to study enzyme evolution, with only 1 of 19 early deep mutational scanning studies targeting enzyme activity directly (Fowler & Fields, 2014). An overview of deep mutational scanning studies performed on enzyme activity is listed in Table 3. In cell survival selections, cells with active enzyme are selected at different concentrations of antibiotic, and cell survival and growth is dependent on the enzyme activity (Firnberg *et al*, 2014). Growth competition experiments, on the other hand, rely on auxotrophic strains or minimal media in which cell growth can only proceed when a metabolite is produced by an active enzyme (Wrenbeck *et al*, 2017a; Chan *et al*, 2017).

Table 3: Deep mutational scanning studies performed on enzymes.

Assay	Target(s)	Activity	Library design	Approx. library size	Epistasis	Reference
survival	TEM-1 β-lactamase	ampicillin resistance (*main*)	single-site saturations	5000	no	Firnberg *et al*, 2014
	TEM-1 β-lactamase	ampicillin (*main*) and cefotaxime (*promiscuous*) resistance	single-site saturations	5000	no	Stiffler *et al*, 2015
	TEM-1 β-lactamase: three mutants	ampicillin resistance (*main*)	single-site saturations	3x 5000	yes	Steinberg & Ostermeier 2016
	TEM-1 β-lactamase	ampicillin resistance (*main*)	sequential double mutations	12000	yes	Gonzalez & Ostermeier 2019
	TEM-1 β-lactamase	ampicillin resistance (*main*)	single insertions & deletions	6000	no	Gonzalez *et al*, 2019
	APH(3')II kinase	resistance to six aminoglycoside antibiotics (*main and promiscuous*)	single-site saturations	5000	no	Melnikov *et al*, 2014
FACS or microfluidic screening	PhoQ protein kinase	phosphorylation of response regulator (*main*)	multi-site saturation	160,000	yes	Podgornaia & Laub, 2015
	Bgl3 β-glucosidase	β-glucosidase activity (*main*), before and after heat treatment	random	3000	no	Romero *et al*, 2015
	DNA polymerase	activity with threose nucleic acids (*promiscuous*)	single-site saturations	1000	no	Nikoomanzar *et al*, 2019
FACS and survival	Levoglucosan kinase and TEM-1 β-lactamase	two assays determining folded expression (*no activity*)	single-site saturations	5000 and 9000	no	Klesmith *et al*, 2017
phage display	E3 ubiquitin ligase	auto-ubiquitination (*main*)	random	100,000	no	Starita *et al*, 2013
growth competition	indole-3-glycerole phosphate synthase: three homologs	tryptophan synthesis (*main*)	single-site saturations	3x 5000	no	Chan *et al*, 2017
	AmiE amidase	three substrates (*main and promiscuous*)	single-site saturations	7000	no	Wrenbeck *et al*, 2017a
	AmiE amidase: two variants with low stabilities	amidase activity (*main*)	single-site saturations	2x 7000	yes	Faber *et al*, 2019
	Levoglucosan kinase	levoglucosan consumption (*main*)	single-site saturations	8000	no	Klesmith *et al*, 2015

A.4.1. Deep Mutational Scanning of New Functions

What information can we gather from these large datasets of sequence-function relationships and how can we use this information for enzyme engineering? Most experiments are performed as single step explorations of mutational impact on the main activity, not necessarily representing the engineering of a new function (Table 3). In a recent example, however, Wrenbeck *et al* (2017a) studied the local fitness landscape of an amidase with three substrates of differing adaptation. A comprehensive single-site saturation library was generated, meaning that every residue in the 341 amino acid protein is mutated to every other amino acid individually resulting in a library of 20 × 341 = 6,820 members. A growth competition selection was then performed with three substrates of decreasing initial activity: acetamide (k_{cat}/K_M = 12.6 mM^{-1} s^{-1}), propionamide (k_{cat}/K_M =2.7 mM^{-1} s^{-1}) and isobutyramide (k_{cat}/K_M = 0.04 mM^{-1} s^{-1}), followed by deep sequencing to establish the impact of every single mutation on activity. The main conclusions drawn from this dataset are (i) fitness landscapes are highly substrate dependent and (ii) substrate-specific beneficial mutations are globally distributed (Wrenbeck *et al*, 2017a).

A re-analysis of this dataset provides some further interesting views on protein evolution: when the mutational impact is averaged per position, new conclusions emerge. Fig. A-12 shows the average impact of all mutations to a position versus the highest impact a mutation in this position can have. Evolutionary conservation is additionally calculated and overlain for visualization (Ashkenazy *et al*, 2016). The conservation of a position is related to its average impact upon mutation: when a conserved position is mutated the average impact on enzyme fitness is likely to be negative, while non-conserved positions are more likely to be functionally neutral upon mutation (Fig. A-12). Such a correlation of functional impact and evolutionary conservation is expected; evolutionary conservation has long been used to predict the pathogenicity of missense mutations in human genomes (Thusberg & Vihinen, 2009).

Interestingly, the percentage of beneficial mutations in all mutations decreases with the initial activity of the substrate. The proportion of positions that can generate a fitness score ≥ 0.3 increases from acetamide (2.6%) to propionamide (7.9%) and isobutyramide (10.3%). This is experimental proof of the protein engineer's motto '*innovation is easy, optimization is complicated*' which is based on often seen large improvement in the initial rounds of directed evolution followed by diminishing returns (Goldsmith & Tawfik, 2017; Newton *et al*, 2018). Additionally, the percentage of conserved positions in beneficial positions also increases along the promiscuity spectrum. Positions with a conservation score ≥ 7 increase from acetamide

(33.3%) to propionamide (40.7%) and isobutyramide (74.3%) in the respective set of beneficial positions. Consequently, conserved positions which on average decrease fitness are nonetheless important keys to unlocking new functions in proteins. Great functional improvements for new activities stem from positions where mutation is on average negative.

Thus, this deep mutational scanning dataset provides a view on protein engineering that is intuitive – a new function evolves from conserved residues – but not yet broadly applied due to restrictive screening assays. If throughput is limiting, many approaches rely on engineering by mutating non-conserved residues around the active site (Jochens & Bornscheuer, 2010; Qu *et al*, 2020; Pavelka *et al*, 2009), which is likely to only provide marginal improvements.

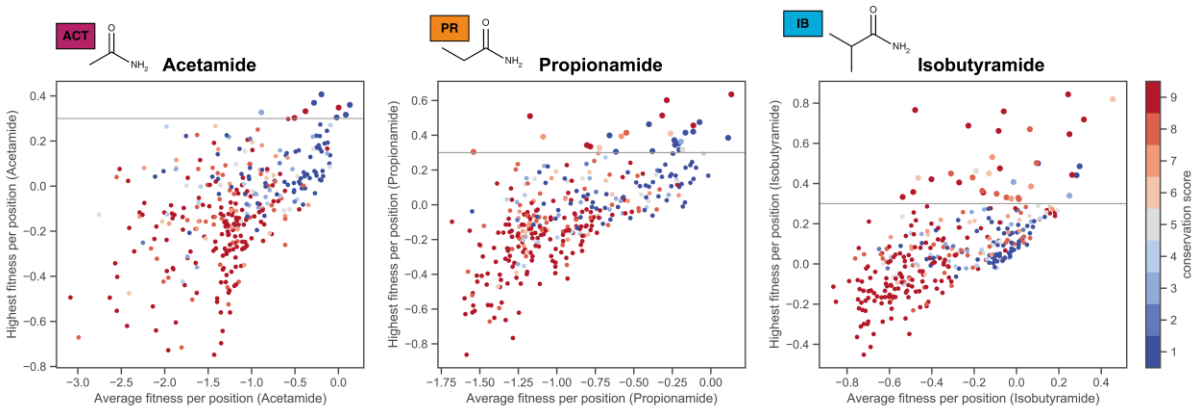


Fig. A-12: Deep mutational scanning data can inform protein engineering. A re-analysis of deep mutational scanning data obtained by Wrenbeck *et al*, 2017 for an amidase with three different substrates shows the importance of conserved positions in the evolution of new function. Analysis performed with custom python scripts: impact of individual mutations was grouped by amino acid position and averages and highest values are computed and plotted per substrate. Conservation scores were calculated with default settings on the ConSurf webserver (Ashkenazy *et al*, 2016, conservation calculated from alignment with 150 homologs of the UniRef90 database) and used as colour-coding.

This study (in fact, all studies of Table 3) provide a view of the local fitness landscapes of enzymes. Variants with merely single or at best few mutations are tested, providing a comprehensive and valuable yet short-sighted view of protein evolution, as many more mutations are usually needed to fully optimize an enzyme. For example, diminishing returns or closing in towards a fitness peak was observed for the evolution of a phosphotriesterase for arylesterase activity with 18 final mutations (Tokuriki *et al*, 2012). One limitation holding back long-term studies of directed evolution towards more global and multipeaked fitness landscapes is an unsuitable high-throughput sequencing technology.

A.4.2. Applied Sequencing Technologies

With the knowledge of remote mutations contributing significantly to enzyme activity, excluding positions from analysis is unreasonable (Wilding *et al*, 2019; Wrenbeck *et al*, 2017a). Ideally, multiple randomly distributed mutations throughout the whole protein will be analysed to fully understand how trajectories form in longer-term protein evolution. All studies of Table 3 sequence only part of the gene with either the deprecated Roche 454 or more recent Illumina sequencing-by-synthesis platforms. While Sanger sequencing remains a mainstay of modern molecular biology with highly accurate reads that can achieve full coverage of the average gene (0.9-1.4 kb, Xu *et al*, 2006), it is not cost-effective at high throughput. Illumina sequencing on the other hand offers incredibly high throughput and accuracy for two reads of 75 to 300 b, resulting in its regular application for deep mutational scanning (Wrenbeck *et al*, 2017b).

The longest primary sequencing reads used for deep mutational scanning have thus been 600 b, obtained by the direct association of the 300 b forward and reverse reads in Illumina sequencing (Yoo *et al*, 2020). More and more approaches are being established to link short sequencing reads into longer synthetic reads (Wrenbeck *et al*, 2017b). One such example was used to establish the fitness landscape of GFP: multiple random mutations were introduced simultaneously throughout the full 714 b gene, which were then sequenced by associating multiple reads of the gene to the same molecular barcode (Sarkisyan *et al*, 2016).

The method used by Sarkisyan *et al* (2016) involved successive restriction and ligation steps to pair the reads of one molecule to its barcode, which is work intensive, scales poorly to longer read lengths and risks the formation of chimeras by mispairing. Longer primary reads would not require assembly, making the developments of third-generation sequencers by PacBio and Oxford Nanopore interesting for future deep mutational scanning studies. Currently, these third-generation sequencing methods suffer from high direct error rates of ~15%, preventing the reliable identification of point mutations and their application in deep mutational scanning (van Dijk *et al*, 2018).

A.5. Structure and Aim of this Thesis

To understand protein evolution and evolvability on a larger scale, experimental directed evolution of a new function needs to be tracked for multiple rounds. The aim of this study is to

establish new technologies enabling this endeavour by (i) providing a general and sensitive ultrahigh-throughput screening method and (ii) increasing the accuracy of full-length nanopore amplicon sequencing. These methods combined are then used to establish the evolutionary trajectories in the directed evolution of two amine dehydrogenases, comparing their evolvabilities.

First, starting points for such a comparative directed evolution campaign are generated. Stabilized amine dehydrogenases are prepared as a model for the acquisition of a new function and their initial evolvability from amino acid dehydrogenases is studied. Next, a general approach to increase the sensitivity of lysate-based ultrahigh-throughput microfluidic enzyme assays is developed, enabling the screening of the low activities of the newly generated amine dehydrogenases. To pair the ultrahigh-throughput screening with a suitable sequencing read-out, a strategy to increase the accuracy of third-generation long-read sequencing methods is proposed. The new protocol enables the acquisition of many full-length reads at low cost by pairing gene variants with unique molecular identifiers to generate highly accurate consensus sequences. Finally, the two stabilized amine dehydrogenase starting points are subjected to parallel directed evolution and sequencing to study the contingencies and trajectories in protein evolution and hopefully derive principles for the engineering of novel activities in enzymes.

B. STARTING POINTS

In this chapter, stabilized amine dehydrogenases (AmDHs) will be generated and characterized. First, phenylalanine dehydrogenase (PheDH) starting points will be generated via directed evolution for increased stability or by using computational tools based on energy calculations of consensus mutations. The stabilized PheDHs will subsequently be turned into AmDHs by active site saturation as a first assay of evolvability. The stabilized PheDHs accept more functionalizing mutations in their active site compared to the wild-type, establishing a stability-mutational robustness link. Mutational robustness, however, is not necessarily correlated to functional improvement.

B.1. INTRODUCTION

A good starting point for directed evolution is a stable enzyme. Protein stability is related to robustness to mutation, thus enabling the acquisition of more functional – and on average destabilizing – mutations during engineering or evolution (Tokuriki *et al*, 2007; Bloom *et al*, 2006). More accepted mutations increase the chance of finding improved mutants, as reflected in the use of stable enzymes from thermophile organisms or consensus and ancestral approaches as starting points for protein engineering (Trudeau & Tawfik, 2019).

The stabilization of an existing enzyme helps to provide a 'stability buffer' (Fig. A-4), useful as the first step of a protein engineering campaign to make the protein more evolvable (Tokuriki *et al*, 2008), although stability is in all likelihood not the sole contributing factor to evolvability (see A.1.2). Stability can be generated itself by directed evolution or by utilizing many available computational approaches.

Directed evolution is a straightforward and successful way to increase protein stability. Diversity is introduced to the protein of interest either in a random manner, e.g. via error-prone PCR (epPCR), or in a more focused manner via site saturation. The screening step is then usually performed with a focus on stability by selecting with stringent expression conditions or after heat incubation. Directed evolution to drive stability has certain advantages: the rate of false positives is low, because the selection step is based on a test for residual activity. Additionally, the method is sequence and structure agnostic, no homologous sequences or crystal structures are required to perform directed evolution. It is thus useful when not a lot of knowledge about the protein is available (Zeymer & Hilvert, 2018; Eijsink *et al*, 2005). Directed evolution does require a significant experimental investment, however, as the number of screened variants is high, often even requiring multiple rounds of diversification and selection to reach large effects (Table 4).

Computer-aided methods for enzyme stabilization were established to reduce screening effort and increase the stabilization success. Here, two large sub-types can be distinguished: sequence-based approaches and energy-based approaches. Ever since it was shown that consensus mutations yield stabilized proteins (Bershtein *et al*, 2008), a back-to-consensus approach to protein stabilization was introduced. Consensus mutations from homologue sequences can be transplanted to the protein of interest, yielding a stabilized protein (Jones *et al*, 2017). A step further, the phylogenetic ancestral state can be computed from a set of homologue sequences, which has been shown to be a robust and promiscuous protein – in line

with the likely environmental requirements of an ancestral protein. Ancestral sequence reconstruction has thus also been used as a tool for protein stabilization (Gumulya *et al*, 2018).

Energy-based models, on the other hand, try to compute the impact of mutations on a protein via physical force fields. Most notably, the FoldX and Rosetta energy functions can be used to predict stabilizing effects in proteins (Alford *et al*, 2017; Guerois *et al*, 2002) and are employed in many automated tools for stabilization (Musil *et al*, 2019). A drawback of purely energy function-based approaches is a high rate of false positives, however. Combination approaches have thus been most successful in recent years. A prevalent example of this is the 'protein repair one stop shop', or PROSS, which combines consensus information with energy calculations. Only mutations supported by evolutionary precedent from homologs will be subjected to energy calculation, reducing the rate of false positives. This has led to high stabilizations while activity is retained with only few experimentally characterized variants (Goldenzweig *et al*, 2016).

Table 4: Examples for different enzyme stabilization methods.

Method			Variants screened	Δ Tm (°C)	Reference
Computational	Sequence	Consensus	15	7	Xie *et al*, 2019
		Ancestor	1	16-30	Gumulya *et al*, 2018
	Structure	Energy-function	52	6	Luo *et al*, 2016
	Combined	PROSS	6	10-15	Goldenzweig *et al*, 2016
Directed evolution			~ 9000	14	Giver *et al*, 1998
			> 70,000	35	Palackal *et al*, 2004

Most recently, machine-learning approaches try to speed up the protein stabilization workflow and increase its successes. These statistical methods benefit from large datasets of natural sequences from which patterns can be extracted (Alley *et al*, 2019) or are trained on available experimental datasets (Musil *et al*, 2019). Their gaining focus in current research as well as not requiring physical models or 3D structures could lead to more frequent adoption of these algorithms for stabilization in the future.

In this study, AmDHs are used as a model for innovation in protein evolution. An ultrahigh-throughput screen is available (Gielen *et al*, 2016) and the AmDH function is completely new to PheDHs, which have almost no prior AmDH activity without the mutation of key active site residues (Abrahamson *et al*, 2012; Ye *et al*, 2015). They are thus a suitable model to study

evolvability in protein evolution and will be stabilized as the first step of a protein engineering campaign in this chapter.

B.2. Results & Discussion

With the stability-robustness link in mind, different ways to generate a stable enzyme were pursued. A general starting point for the engineering of AmDH activity is found in the PheDH of *Rhodococcus sp.* M4 (Brunhuber *et al*, 2000). This enzyme has been turned into an AmDH previously (Ye *et al*, 2015). However, it has also been shown to be only moderately stable and poorly expressed in *E. coli* (Gielen *et al*, 2016). Now, the generation of a stabilized starting point could circumvent evolvability issues due to marginal stability and afford greater AmDH activity.

To verify this hypothesis, two approaches for protein stabilization will be used on the wild-type (WT) PheDH. First, the above-mentioned combined approach of consensus and energy-function information in the PROSS algorithm will be applied and resulting designs characterized.[2] Second, a directed evolution approach will be used to stabilize the PheDH, continuing the work of Gielen *et al* (2016). The different ways of stabilization can then be compared for their benefit to engineering and evolvability.

B.2.1. Computational Stabilization

Computational stabilization of the wild-type PheDH was done using the PROSS algorithm (Goldenzweig *et al*, 2016). The PROSS webserver was challenged with the *Rhodococcus sp.* M4 PheDH crystal structure (PDB 1C1D), while the active site was excluded from mutation by the algorithm by restricting all residues less than 8 Å from any of the two bound substrates. The webserver produced six potential designs with consecutively more permissive mutational threshold, here termed P1 to P6. The designs show an increasing number of mutations being included from 22 mutations in P1 to 57 mutations in P6, potentially achieving higher stabilization yet risking false positive mutations. All six designs were ordered as gene strings and cloned into the anhydrotetracycline (aTc)-inducible expression vector pASK-IBA63b+ containing a C-terminal strep-tag. Successful stabilization was tested by determining the

[2] Please note that the work on stabilization was initiated by Raphaëlle Hours, with whom some of the experiments in chapter B.2.1 were performed together, as indicated again in the relevant sections.

relative PheDH activity in cell lysate as well as densitometric analysis of SDS-PAGE gels to determine the soluble expression strength (Table 5). The soluble expression strength increases with the PROSS design's number of mutations until design P4, reaching almost completely soluble expression at 94% compared to the 44% soluble expression of the wild-type PheDH. Designs with higher number of mutations than P4 show reduced soluble expression, indicating an accumulation of false positive destabilizing mutations. The same trend is observed in lysate activity, which is directly influenced by the amount of soluble expression. Here, P3 and P4 show the highest activities with 80% and 77% of the activity of the non-mutated wild-type PheDH. Consequently, the design P4 (Supplementary Sequences) was chosen for all further characterization and evolution due to an indication of improved kinetic stability while retaining high levels of activity.

Curiously, all PROSS designs exhibit a decreased lysate activity, prompting a deeper characterization of their kinetic parameters as well as the generation of a PheDH variant stabilized by directed evolution.

Table 5: Computational stabilization with the PROSS algorithm. Data courtesy of Raphaëlle Hours. Wild-type (WT) PheDH as well as six PROSS designs with increasing mutational load were tested.

Variant	WT	P1	P2	P3	P4	P5	P6
Number of mutations	0	22	27	37	43	52	57
Soluble expression[a] (%)	44	59	60	62	94	73	55
Lysate activity[b] (relative to WT)	1.00 ± 0.24	0.66 ± 0.12	0.46 ± 0.02	0.80 ± 0.24	0.77 ± 0.13	0.55 ± 0.09	*n. d.*

[a]Protein expressed at 20 °C for 16 h. Expression cultures were lysed and cleared by centrifugation. Soluble fraction compared to insoluble fraction via SDS-PAGE gel densitometry to determine the amount of solubly expressed protein. Experiment performed by Raphaëlle Hours in one biological replicate.
[b]Initial rate of PheDH activity was measured in lysate (5 mM L-phenylalanine, 2 mM NAD^+ in 100 mM Glycine-KOH pH 10). Experiment performed by Raphaëlle Hours in three biological replicates, showing mean ± standard deviation.

B.2.2. Stabilization by Directed Evolution

As an alternative to computational stabilization, enzyme stabilization by directed evolution was performed. An ultrahigh-throughput assay for PheDH activity was found in the detection of the co-product NADH coupled to the formazan dye WST-1 in absorbance-activated droplet sorting (AADS, Gielen *et al*, 2016). The wild-type PheDH was evolved towards increased activity for two rounds by Gielen *et al*, however, no improvements to catalytic efficiency were achieved. The selected variants were improved in expression strength and stability, resulting in increased lysate activity and selection frequency in the assay. The most improved variant (V28I N124S L195M T341I) was submitted to another round of directed evolution by Raphaëlle Hours,

yielding five improved variants in the third round of directed evolution (adding the mutations S6R V71I; N132Y; Q202R; L203V; T334K). These variants were not on par with the stabilization afforded by the PROSS algorithm (94% soluble expression), with the variants from round 3 reaching maximally 68% soluble expression (Fig. B-1).

This prompted the continuation of the directed evolution efforts of Raphaëlle Hours, as a variant with comparable stabilization to the P4 variant was desired to study alternate effects other than stability on the evolution of a new function. A fourth round of directed evolution was thus performed specifically with stabilization in mind. A library was generated by shuffling the mutations of the five variants identified by and inherited from Raphaëlle Hours, using the staggered extension process (StEP, Zhao *et al*, 1998). In this modified PCR protocol, very short extension times (30 s) and low annealing temperatures (55 °C) combined with a high cycle number (100 x) result in the frequent recombination of similar template molecules. Each of the five variants from round 3 were subjected to DNA shuffling via StEP, generating a library of 1.3×10^7 transformants. To put emphasis on stabilization in the screening, selection pressure was put on expression strength by changing expression conditions from 16 h at 20 °C to a more stringent regiment, expressing the library at 37 °C for 2 h. The most active variants were selected via AADS as done in the previous rounds by Gielen *et al* (2016). Single cells were co-encapsulated with PheDH substrate (5 mM L-phenylalanine and 2 mM NAD^+) and a lysis solution, releasing the enzyme from the cell in the droplet to start PheDH reaction progress (Fig. A-9). Sorting was performed at ~120 Hz and a droplet occupancy $\lambda = 0.35$, selecting ~100 droplets from a total of ~400,000 droplets. The DNA from sorted droplets was isolated and used to transform cells for clonal recovery. The recovered sorted variants were subjected to a secondary activity assay in 96-well plates to confirm their activity. The four best variants were chosen for further analysis and characterization.

All four final selected variants, as well as the non-mutated wild-type and the best computationally stabilized variant P4, were purified and characterized in regard to their steady state kinetics, their soluble expression strength and their relative activity in lysate (Fig. B-1, Table S1). Of the four tested variants in round 4, a variant with a high level in all properties was selected to continue directed evolution with. This final evolutionarily stabilized PheDH variant (S6R V28I V71I N124S L195M Q202R T341I, from now on termed E4) acquired seven mutations in four rounds of evolution and shows comparable stabilization properties to the P4 design, making it a suitable candidate for further comparative investigation of evolvability.

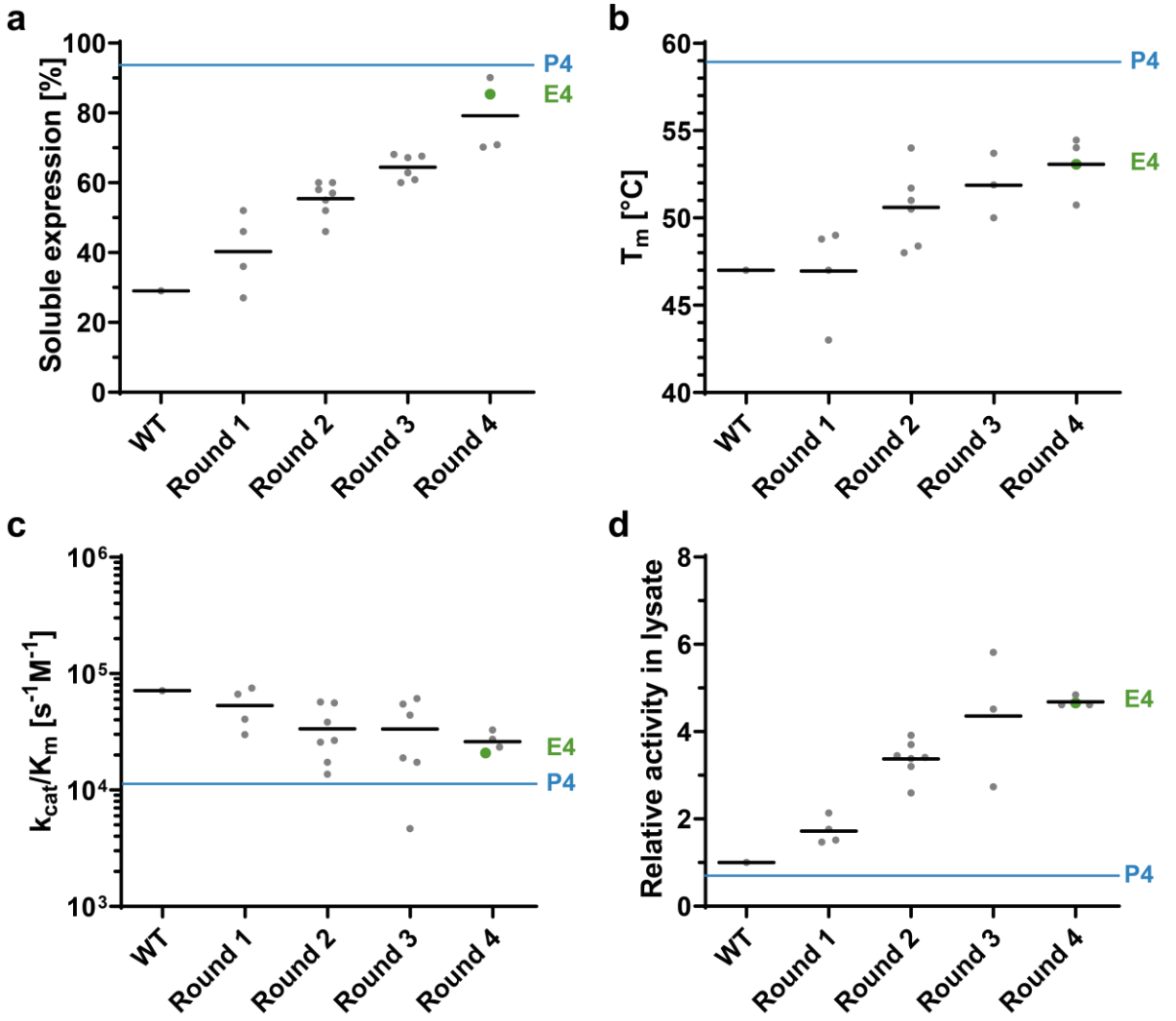
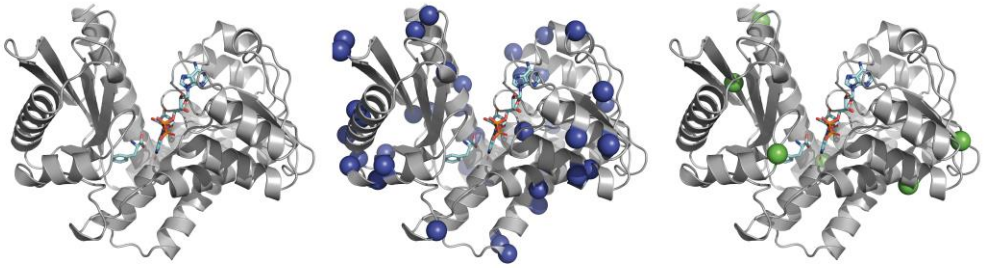


Fig. B-1: Characterization of PheDH stabilization success via directed evolution. The wild-type PheDH (WT) was subjected to directed evolution in four rounds with AADS. Data of round 1 to round 3 were obtained from Raphaëlle Hours. Library generation via epPCR (round 1 and round 3) or StEP (round 2 and round 4). The variants showing the highest activity in lysate in each round were selected for protein purification and characterization. Detailed values with standard deviations are shown in Table S1. Directed evolution continues with P4 and E4 variants, which are highlighted in blue and green, respectively. **(a)** *Soluble expression strength*. Soluble expression strength was determined via densitometric analysis of SDS-PAGE gels as the ratio of recombinant protein in the soluble lysate fraction compared to the insoluble pellet fraction. **(b)** *Thermal stability*. The thermal stability was determined via differential scanning fluorimetry with SYPRO orange. **(c)** *Catalytic efficiency*. Steady state kinetics were acquired with 0-15 mM L-phenylalanine and 5 mM NAD^+ in 100 mM Glycine-KOH pH 10 at 22 °C. **(d)** *Relative lysate activity*. Initial rate of activity in cell lysate compared to the non-mutated wild-type PheDH.

The final evolutionarily stabilized PheDH variant E4 achieves a similar level of improvement in soluble expression level than the computationally stabilized variant P4 (Fig. B-1a). This stabilization of E4 is achieved with seven mutations in the course of four rounds of evolution, whereas P4 is stabilized with 44 mutations in one single design step. While P4 adds another 5 °C of thermal stability (P4 reaching 59 °C compared to 54 °C for E4 and 47 °C for WT), E4

on the other hand has an only marginally reduced catalytic efficiency (Fig. B-1c) and improved activities in lysate (Fig. B-1d).

These three variants, WT, P4 and E4, will now be compared for their evolvability towards a new function. WT is the non-stabilized non-mutated PheDH, whereas P4 and E4 both are stabilized to similar extents but in very different ways (Fig. B-2). Introducing a new function to all three variants, turning the AADHs to AmDHs, might provide an insight into the importance of stability over other factors for the acquisition of new function in protein evolution.



	WT	P4	E4
Stabilization	non-stabilized	computational	directed evolution
Mutations	0	44	7
T_m	47 °C	59 °C	54 °C
Soluble expression	28%	98%	91%

Fig. B-2: Overview of starting points. Three PheDH starting points are available. WT: non-mutated wild-type PheDH *Rhodococcus sp*. M4. P4: Computationally stabilized variant of WT, design four of the PROSS algorithm. E4: Evolutionarily stabilized variant of WT, obtained after four rounds of directed evolution. WT structure (PDB ID 1C1D) with positions of mutation in stabilized variants highlighted.

The increase in thermal stability afforded by the PROSS algorithm of 12 °C is similar to the achievements in stabilization of 8 to 20 °C seen for four different enzymes using the same algorithm (Goldenzweig *et al*, 2016). The stabilization by directed evolution reached an increase in T_m of 7 °C with just 7 mutations, bringing it considerably closer in thermal and folding stability to the P4 variant than the non-mutated WT variant.

B.2.3. Comparison of Evolvability – Generation of AmDHs

All three potential starting points were turned into AmDHs. The initial step to generate AmDH function in an AADH is the mutation of two key active site residues – K66 and N262 – coordinating the carboxyl group of the amino acid substrate of the AADHs, in this case L-phenylalanine (Fig. A-8). A library of simultaneous randomization in these two positions was

screened in a 96-well plate lysate assay for activity with the amine substrate (*R*)-1-methyl-3-phenylpropylamine to generate an AmDH.

The WT PheDH from *Rhodococcus sp.* M4 had previously been screened and turned into an AmDH in this manner, resulting in the active site mutations K66Q N262C (Ye *et al*, 2015). Here, this assay is independently repeated for the WT in comparison to the stabilized variants E4 and P4. Screening the three starting points in parallel will provide a view on the impact of active site mutation and a comparison of accepted mutational diversity as a measure of active site plasticity and robustness. These first values can hopefully provide an indication towards the evolvability of starting points for a new function in a comparative manner.

The WT PheDH shows very little promiscuous activity for the amine substrate (k_{cat}/K_M = 3.4 s^{-1} M^{-1}), providing catalytic efficiencies of < 0.01% of the AADH activity (k_{cat}/K_M = 7.1 × 10^4 s^{-1} M^{-1}, data courtesy of Raphaëlle Hours), thus requiring a very sensitive screening assay to pick out low activating mutations. Consequently, a 96-well plate assay measuring the initial slope of deamination activity of raw cell lysate expressing AmDH variants was performed, as the deamination activity is the more active reaction direction. The reduced throughput in lysate assay compared to NADH-cofactor detection in AADS was mitigated by using a reduced library design. Positions K66 and N262 were simultaneously mutated in all three starting points with the 22c-trick (Kille *et al*, 2013). Site saturation mutagenesis is performed with a mixture of degenerate primers instead of conventional NNN or NNK mutagenesis. If the codon of interest is exchanged with NNK, all 20 possible amino acids along with one stop codon are encoded in the library of 32 possible codons. Thus, higher oversampling in screening is required due to codon redundancy. This problem is reduced by applying the 22c-trick: A mixture of primers containing the codon NDT, VHG and TGG are added at a ratio of 12:9:1, encoding all 20 amino acids with 22 codons, thus requiring minimal additional screening oversampling (Kille *et al*, 2013).

The K66X N262X libraries were generated and thus contained 400 amino acid variants in 484 codon variants. To account for sufficient oversampling, twelve 96-well plates were screened, ensuring screening of the library to 90% completeness by analysing 1104 variants (Patrick *et al*, 2003). The activities of all 1104 randomly tested variants relative to the non-mutated starting point control are shown in Fig. B-3. Improvements in AmDH activity can be found for all three variants, however, the distribution of fitness effects differs between the starting points. The number of neutral and beneficial mutations (as defined by relative activity > 0.8) in the WT background amounts to 20.9%, while this number is considerably higher in the two

stabilized backgrounds: 38.5% and 35.6% for P4 and E4 backgrounds, respectively (Fig. B-3d). The same trend is observed for beneficial mutations alone (as defined by relative activity > 1.5): 9.3% of variants in the WT background show beneficial effects, while the stabilized backgrounds achieve 22.2% and 23.9% (P4 and E4, respectively). This increase in the amount of neutral and beneficial mutations in the stabilized background provides evidence to support that stability confers mutational robustness.

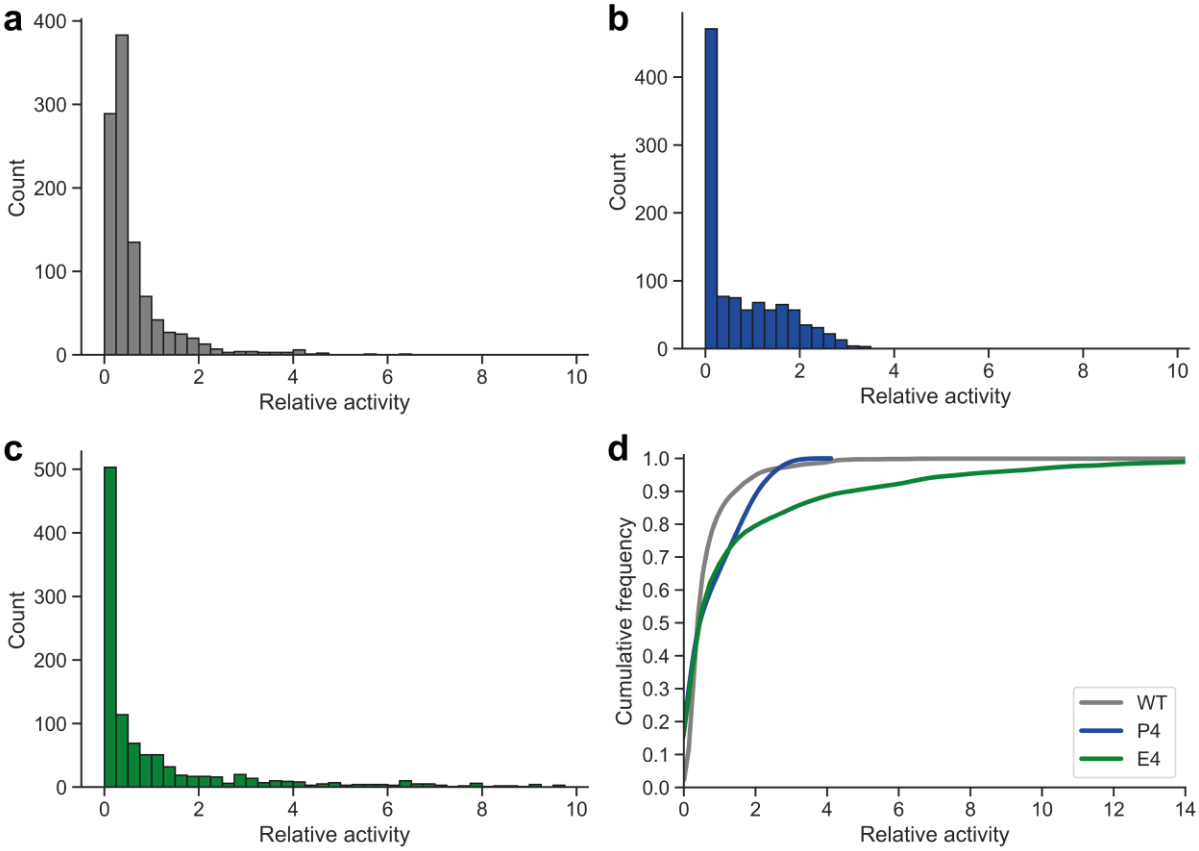


Fig. B-3: Distribution of fitness effects of amine dehydrogenase key mutations in three starting points. Mutations K66X and N262X were introduced to three PheDH starting points (WT, P4, E4). The resulting libraries were screened in 96-well plate lysate assay for AmDH activity (5 mM *R*-1-methyl-3-phenylpropylamine, 1 mM NAD^+ in 100 mM Glycine-KOH pH 10). The initial rate of product formation (measured via absorbance at 340 nm) of 1104 variants relative to the respective non-mutated parent are shown in histograms. Axes clipped to show meaningful part of the histogram. **(a)** *non-mutated wild-type PheDH. Rhodococcus sp*. M4 PheDH shown in grey. **(b)** *computationally stabilized PheDH P4*. PheDH stabilized by the PROSS algorithm, shown in blue. **(c)** *evolutionarily stabilized PheDH E4*. PheDH stabilized in four rounds of directed evolution, shown in green. **(d)** *cumulative histogram.* Cumulative frequency shown for all three starting points.

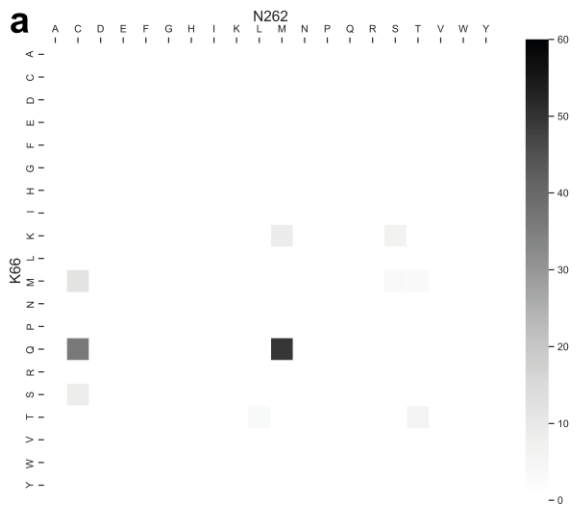
Interestingly, the evolvability – if defined by the highest achieved improvement – differs between the two stabilized backgrounds. While the P4 stabilized background increases mutational robustness, as established, it does not achieve high total improvements to the AmDH activity. The best AmDH variants observed for the three backgrounds provided relative increases to activity of 3.4-fold in the P4 background, while both WT and E4 achieved much

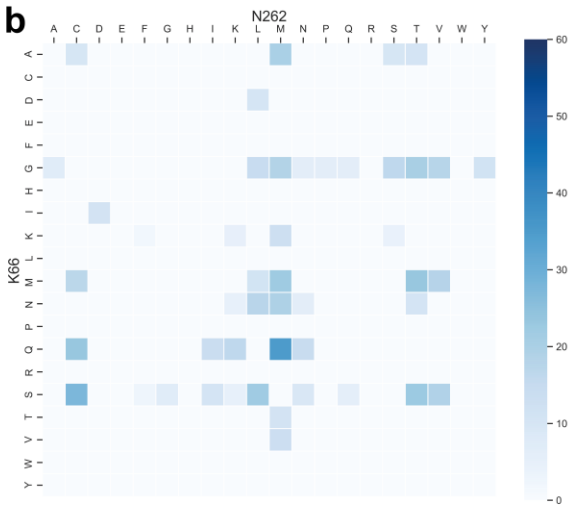
higher increases (15.9-fold and 29.8-fold, WT and E4 respectively). More quantitatively, the number of variants providing a very high increase in AmDH activity (as defined by relative activity > 5) are 0% for P4, 1.2% for WT and 11.3% for the E4 background. While both stabilized backgrounds show a high number of active variants, indicated by the increased numbers of neutral and slightly beneficial variants, only the E4 genetic background achieves an increased number of highly active variants. In the respective histograms of relative activity (Fig. B-3) these effects are reflected in a long and drawn-out tail of the distribution for E4 and a condensed and compacted tail for P4.

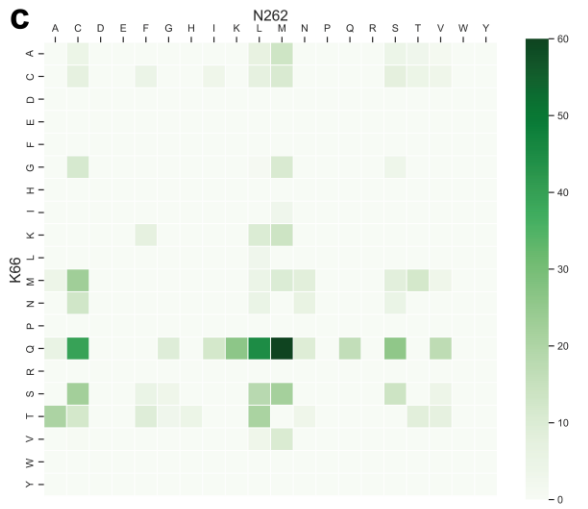
These findings hint towards a more nuanced view of evolvability than merely relying on stability. Potential confounding factors, such as direct epistasis with the active site mutations by the acquired stabilizing mutations, might hinder the adoption of greatly beneficial mutations in the P4 background, while these are accepted in the more closely related WT and E4 backgrounds. To assess how the differences in evolvability affect the tolerated mutations on a sequence level, all variants improving the AmDH activity at least 1.2-fold were isolated, their activity improvements verified in triplicates in a more sensitive assay and the corresponding mutations identified via Sanger sequencing. The sensitivity of the lysate assay is increased by measuring the produced NADH indirectly. NADH reduces a chromogenic dye, the absorbance of which is measured instead with higher sensitivity (Fig. S1). The resulting heatmap for the combination of mutations generating AmDH activity is shown in Fig. B-4. The best combination of mutations for AmDH activity is K66Q N262M for all three starting points (relative improvements: 44.7±2.7-fold, 34.9±2.4-fold, 59.9±8.2-fold, for WT, P4, E4, respectively, SD shown, n=3), confirming the differences in highest achieved improvement in activity seen in the initial screening. The active site mutations in the positions K66 N262 are shortened for convenience from here on, e.g. $P4_{QM}$ referring to the P4 stabilized background with the mutations K66Q N262M.

Curiously, the most activating mutations at these positions in previous AmDH engineering studies showed S/L as the best combination of mutations for PheDHs from *Bacillus badius* (Abrahamson *et al*, 2013) and *Caldalkalibacillus thermarum* (Pushpanath *et al*, 2017). Furthermore, Ye *et al* performed a similar screening of the two active site mutations with the same non-mutated WT PheDH from *Rhodococcus sp.* M4 and identified Q/C as the best combination of mutations (Ye *et al*, 2015), which is outperformed by Q/M in this study (Fig. B-4a). This speaks to prevalent interactions between the mutations accepted in the different

homologs as well as to simple chance of missing a mutation when screening a randomized library with low throughput methods.







d

Criterion	WT	P4	E4
Total combinations	400	400	400
> 1.2-fold initial rate	10	47	62
> 20-fold initial rate	2	8	10
Best variant	Q/M	Q/M	Q/M
Relative activity	44.7±2.7	34.9±2.4	59.9±8.2

Fig. B-4: Combinations of activating mutations in positions K66 and N262 for three PheDH starting points. All variants that registered > 1.2-fold relative activity compared to their respective non-mutated parent in the initial library screening were chosen for sequencing and verification. AmDH activity verification was performed in independent biological replicates (n=3) in an indirect initial rate assay in lysate for deamination activity. Produced NADH was detected sensitively with the chromogenic dye WST-1 at 455 nm. **(a)** *WT background.* Active mutations introduced to non-mutated WT PheDH from *Rhodococcus sp.* M4 and measured for AmDH activity. **(b)** *P4 background.* Computationally stabilized PheDH background. **(c)** *E4 background.* Evolutionarily stabilized PheDH background. **(d)** *Table of combinations.* Combinations of mutations that are activating were counted. Activity with SD shown for the best variant.

The diversity of accepted mutations, accepted here meaning that the mutations are neutral or beneficial, is increased for the two stabilized backgrounds (Fig. B-4) compared to the non-stabilized WT. The pattern of mutations that are accepted are similar; many highly beneficial mutations are overlapping (K66/N262 mutated to Q/M, Q/C, S/C or M/C). Thus, there is no strong specific epistasis detectable between the most beneficial active site mutations and

mutations introduced during stabilization. As mutations around the active site were excluded from mutation during PROSS stabilization, a lowered impact on the active site is expected. The few mutations that are exclusive to one of the two stabilized backgrounds, such as mutation Q/L, might just not have been sampled in the initial screening, making conclusions from missing data difficult without additional verification.

Stability-mediated threshold epistasis seems to be the major factor constricting the access to different mutations in the active site. Of the 400 possible combinations of K66/N262 residues, only 10 were found to be slightly beneficial (>1.2-fold relative initial rate) in the non-stabilized WT background, while 47 and 62 were slightly beneficial in P4 and E4 respectively (Fig. B-4d). Mutations are on average destabilizing (Tokuriki *et al*, 2007, 2008) and the active site mutations of AmDHs have been shown to be especially detrimental to stability, decreasing Tm by 2-11°C in different variants (Pushpanath *et al*, 2017). Such a strong negative effect could deplete active molecules of the non-stabilized WT, destabilizing it to an extent where it becomes completely unfunctional, while the same effect is buffered by the higher global stability of the P4 and E4 variants (Tokuriki & Tawfik, 2009b; Bloom *et al*, 2006).

B.3. Conclusion & Outlook

Two different modes of enzyme stabilization, computational stabilization and directed evolution, were successfully employed to create stabilized starting points for AmDH evolution. Both the P4 and E4 variants show increased stability, especially improving on soluble expression compared to the WT, increasing from 28% to more than 90% solubly expressed enzyme. These stabilized starting points are good candidates to introduce and evolve AmDH activity, and the first step of creating an AmDH was performed by mutating key active site residues K66 N262 in all three variants comparatively.

The data on mutations affecting AmDH activity in these three starting points again shows convincingly that stability does confer mutational robustness. Less mutations are accepted in the WT compared to P4 and E4, the two stabilized starting points. This effect has been well established (Bloom *et al*, 2006; Tokuriki & Tawfik, 2009b; Besenmatter *et al*, 2007) and its use in protein engineering is prevalent, with many protein engineers turning towards stable starting points from thermophile organisms (Finch & Kim, 2018).

More recently, however, a more nuanced view of evolvability has emerged. Factors other than mutational robustness seem to contribute to evolvability (see A.1.2). One such factor is conformational flexibility. Flexibility allows the enzyme to sample more conformational sub-states, some of which might be productive with new substrates, contributing to a broad promiscuity profile (Campbell *et al*, 2018). Conformational flexibility has been shown to be high in ancestral protein states and to become more restricted with specialization (Risso *et al*, 2017; Romero-Rivera *et al*, 2017; Campbell *et al*, 2016). Additionally, pervasive epistatic effects or so-called cryptic genetic variation can drastically affect which mutations are accessible to a protein and thus commit starting points to pre-determined trajectories and limit accessible fitness optima (Starr & Thornton, 2016; Baier *et al*, 2019).

These effects may be evident in the data of this chapter: while less mutational diversity is accepted in the WT background, a high beneficial effect of the accepted mutations is observed, nonetheless. The best WT AmDH (WT_{QM}) improves AmDH activity 45-fold, while the best P4 AmDH ($P4_{QM}$) improves by only 35-fold (Fig. B-5). This may hint at differential evolvability that is not necessarily correlated to mutational robustness and stability, but to the other factors mentioned above. Furthermore, differential evolvability is indicated in the two stabilized starting points: While both E4 and P4 are robust to mutation, the relative improvements afforded by mutation is higher for $E4_{QM}$ (60-fold) compared to $P4_{QM}$ (35-fold).

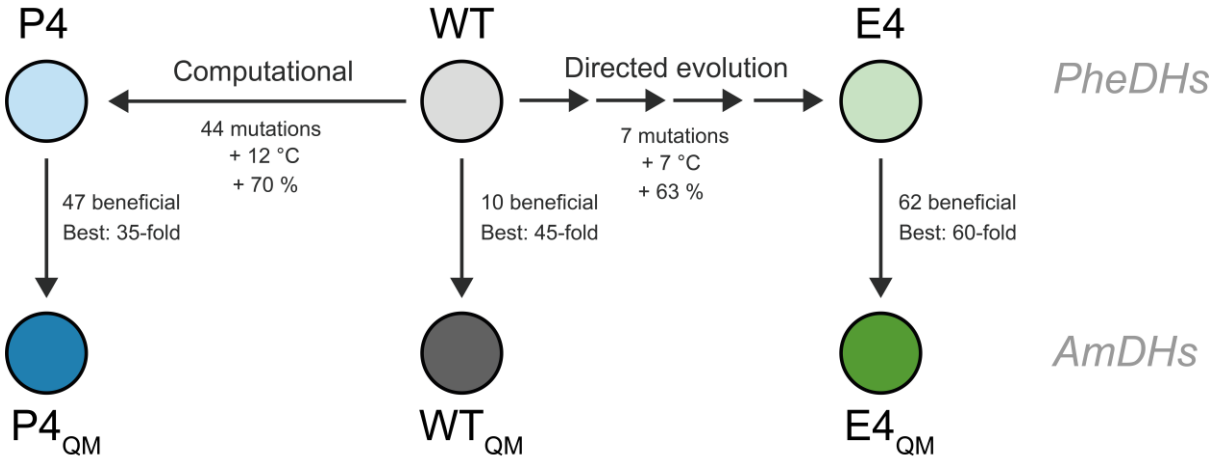


Fig. B-5: Overview of work performed in this chapter. Wild-type (WT) PheDH is computationally stabilized (P4, Tm + 12 °C, soluble expression + 70%) and evolutionarily stabilized (E4, Tm + 7 °C, soluble expression + 63%). All stabilized PheDHs are then turned into AmDHs by active site mutation and their initial evolvabilities are compared. A different number of mutations is beneficial for each of the starting points, with WT accepting few combinations of mutations (10) and P4 and E4 accepting many (47 and 62, respectively). The best variants all carried the K66Q N262M mutations with different activity improvements of 45-fold, 35-fold and 60-fold for WT_{QM}, $P4_{QM}$ and $E4_{QM}$, respectively. Interestingly, the increase in mutational robustness of P4 (many different mutations accepted) does not correlate to higher activity improvements ($P4_{QM}$ not as active as WT_{QM} or $E4_{QM}$), hinting at a nuanced view of evolvability.

These apparent differences in evolvability are likely to stem from specific epistatic interactions, even though the active site was excluded from mutation in the stabilization process. Such epistatic interactions have been shown to extend over long ranges in unexpected manners (Sarkisyan *et al*, 2016; Miton *et al*, 2021). Furthermore, a study of protein flexibility and confirmational dynamics might be interesting in the future, for example by generating and comparing crystal structures of the variants, as protein flexibility has been shown to influence evolvability in the past (Campbell *et al*, 2016). In any case, these currently still anecdotal findings should be solidified with a large-scale comparative evolution, ideally further evolving both stabilized AmDHs created here in parallel towards higher AmDH function. A high throughput screening assay and multiple rounds of directed evolution could prove more convincingly that accessible fitness peaks differ for the two related enzymes.

C. Increasing the Sensitivity of Single Cell Microfluidic Assays

In this chapter, limitations to the current workflow for single cell lysate assays in microfluidic droplets will be introduced and a solution will be proposed: monoclonal and homogeneous cell growth can increase the recombinant protein content per compartment, thus increasing the reactivity. A workflow is established that uses modified reaction tubes as robust devices for droplet manipulation during growth as well as a new 'pico-fusion' chip design enabling the addition of large amounts of substrate and lysis buffer. The sensitivity of droplet microfluidic screening of AmDHs is increased > 10-fold and the variation in the assay is reduced by half. This enables the detection of wild-type AmDH activity, a feat not possible in the corresponding single cell microfluidic assays.

The content and figures of this chapter were published as Zurek et al*, 2021.*

C.1. INTRODUCTION

Biocatalysis enables the increasing field of sustainable 'green' chemistry, due to an enzyme's remarkable specificity and selectivity in mild reaction conditions (Bornscheuer *et al*, 2012). However, the reactions to be performed are often not known to be enzyme catalysed, requiring efficient routes to generate enzymes with novel activities and properties. Here, secondary 'promiscuous' enzyme activities broaden the range of available reactions for biocatalysis (Nobeli *et al*, 2009; Kazlauskas, 2005). Furthermore, large samples of natural biodiversity can be sieved for novel enzymes in functional metagenomics (Colin *et al*, 2015a; Uchiyama & Miyazaki, 2009) or computational design can be employed to generate new active site geometries and functions (Zanghellini, 2014; Röthlisberger *et al*, 2008). However, in all these cases, the initial reactivity is low. Directed evolution is commonly employed to increase the initially weak activities of enzymes for new reactions or mend activities to industrial reaction conditions and has led to many successful biocatalysts (Porter *et al*, 2016; Zeymer & Hilvert, 2018).

The prospects of a directed evolution campaign increase with a higher screening capacity, as larger diversity can be explored by high and ultrahigh-throughput screening and finding a hit thus becomes more likely. A valuable tool for ultrahigh-throughput enzyme screening has been the emergence of many different droplet microfluidic enzyme assays in recent years (see A.3). Here, single library members, e.g. single cells expressing individual enzyme variants, are co-encapsulated with substrate and lysis solution to form small reaction compartments of few pL. In each compartment, the enzymes are released from the lysing cell and a detectable reaction progresses (Fig. A-9). Individual library members are then selected at kHz frequencies by interrogation of the reaction progress in droplets, enabling the screening of millions of enzyme variants per day with minimal reagent consumption (Agresti *et al*, 2010; Kintses *et al*, 2012; Neun *et al*, 2020).

A crucial requirement for droplet microfluidic assays is the necessity to compartmentalize just one entity per droplet compartment. Multiple variants per compartment would confound measurements, thus requiring dilution of the input solution to achieve mostly single occupancy. Encapsulation events follow a Poisson distribution, enabling the calculation of expected single-, double- and multi-encapsulations from the dilution of the input cell solution (Fig. C-1). A trade-off is thus created between throughput, decreased by the number of empty droplets, and false positives due to double encapsulations. An acceptable compromise often committed

to in this study is found in a λ, the expected rate of occurrence, of 0.2, resulting in 81.9% empty droplets to 16.4% single- and 1.7% multi-encapsulations.

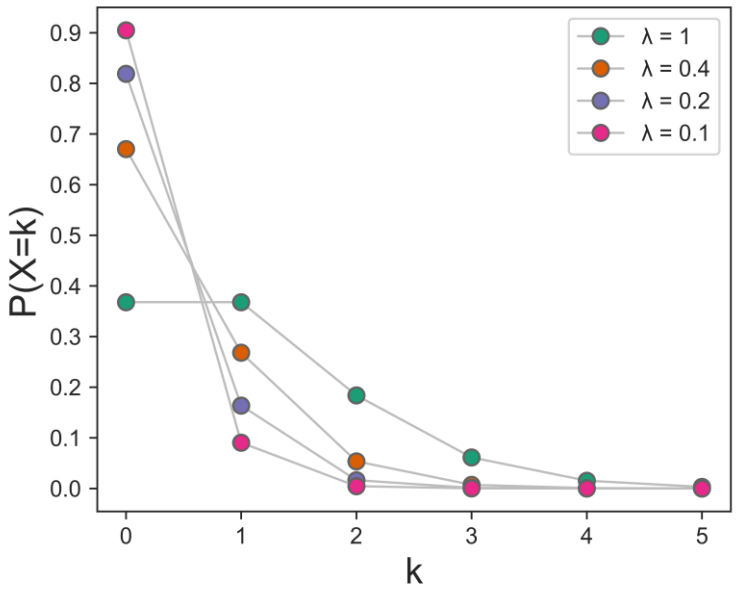


Fig. C-1: Poisson distribution for different values of λ. The probability mass function of the Poisson distribution is given by $f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$. The index *k* is the number of occurrences or, in the context of droplet microfluidics, the number of cells per droplet. The probability of *k* occurrences P(X=k) given different values of λ is plotted, with λ being the expected rate of occurrence. In this study, the droplet experiments are often set up to reach a λ of 0.2, i.e. 81.7% empty droplets, 16.4% single encapsulations and 1.7% multi-encapsulations.

After release of enzymes from the cell in the droplet reaction compartment, a detectable reaction progresses. Here, optical detection with a fluorescent reaction product is not only the earliest and most common detection mode, but also provides the highest sensitivity, e.g. enabling the detection of as little as 1 nM Amplex Ultrared in the directed evolution of a horseradish peroxidase, corresponding to ~3500 molecules in a 6 pL droplet (Agresti *et al*, 2010). However, very low enzyme activities e.g. of promiscuous metagenomic hits with low expression strength, could still be precluded even with this superb detection sensitivity. Additionally, the detection sensitivity of other optical assays such as absorbance (Gielen *et al*, 2016) or fluorescence anisotropy (Gielen *et al*, 2017) fall far lower than fluorescence intensity measurements, with detection limits of 10 µM and 0.1 µM, respectively. Sensitive detection is similarly challenging in recent non-optical detection modes for droplet microfluidics, such as electrochemistry (1 µM, Goto *et al*, 2020) or mass-spectrometry (30 µM, Holland-Moritz *et al*, 2020).

These newer detection modes promise the label-free detection of a manifold of reactions (Neun *et al*, 2020). Using fluorescently labelled substrates for directed evolution poses the risk of

specialization towards the assay substrate, instead of the actual industrially relevant substrate, in line with the directed evolution principle '*you get what you screen for*' (You & Arnold, 1996). Here, novel detection modes can relieve this issue, e.g. by absorbant (Gielen *et al*, 2016) or electrochemical (Goto *et al*, 2020) detection of the reduced co-factor NAD(P)H in a dehydrogenase reaction, although at the cost of decreased sensitivity.

While droplet monoclonality, e.g. the compartmentalization of a single cell, is necessary for accurate detection of the corresponding enzyme variant's activity, it limits the available amount of enzyme molecules per reaction compartment. Expression can achieve > 10^5 molecules per cell (Gielen *et al*, 2016) or > 30,000 *in vitro* (Courtois *et al*, 2009), while surface display achieves 20,000 molecules on yeast cells (Agresti *et al*, 2010) or > 10^5 on *E. coli* cells (Jose, 2006). These expression strengths can afford detection of less than one substrate turnover per enzyme molecule in small droplets for fluorescence detection, however, the droplet size and detection limits increase dramatically for absorbant as well as the new non-optical detection modes. These combined effects of limited enzyme supply per compartment with large droplet volumes and higher detection limits results in more drastic requirements to enzyme activity and stability for successful detection. Assuming a recombinant expression strength of ~ 8×10^5 molecules per cell as identified by Gielen *et al* (2016), each enzyme molecule needs to convert 1350 substrate molecules for absorbant detection (10 µM detection limit, 180 pL droplet volume, Gielen *et al*, 2016), or 22600 substrate molecules per enzyme molecule for electrochemical detection (1 µM detection limit, 30 nL droplet volume, Goto *et al*, 2020) and even 565000 substrate molecules per enzyme molecule for mass-spectrometric detection (30 µM detection limit, 25 nL droplet volume, Holland-Moritz *et al*, 2020). These high requirements to turnover per enzyme molecule prohibit screening and evolution of less active or less stable enzymes. For context, novel biocatalysts engineered towards new functions over multiple rounds of directed evolution have recently been improved from < 100 turnovers per catalyst to reach final total turnovers numbers of < 10000 (Chen & Arnold, 2020; Jia *et al*, 2020). The high turnover per enzyme molecule requirements of the novel microfluidic detection modes thus exclude even these final engineered biocatalysts from further evolution in droplets with non-optical detection modes, let alone being able to detect new function in the early rounds.

A general solution, theoretically applicable to all existing single cell lysate-based microfluidic assays, is proposed here: The recombinant protein content per compartment is multiplied by achieving cell growth and protein expression in droplets with the goal of increasing sensitivity

for droplet microfluidic assays and enabling the detection of the newly introduced AmDH functionalities in chapter B. AmDH activity is detected in droplets via AADS, measuring the absorbance of NADH via a coupled reaction with the formazan dye WST-1 (Fig. C-2, Fig. S1).

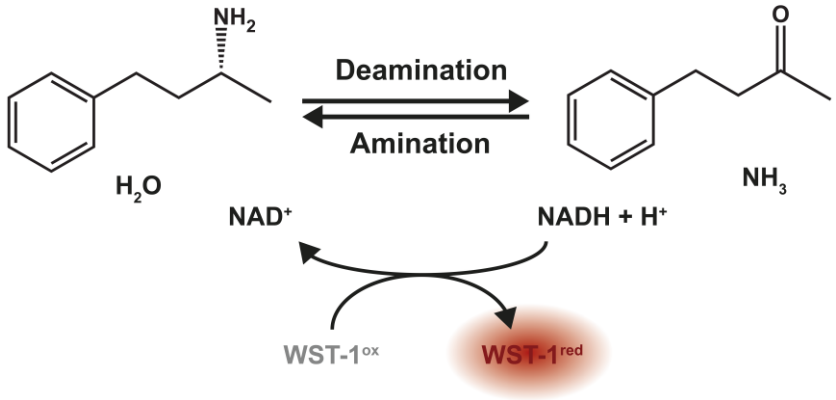


Fig. C-2: AmDH detection in AADS. The deamination of *R*-1-methyl-3-phenylpropylamine to 1-phenyl-3-butanone by AmDHs can be detected in droplets via the coupled detection of NADH in AADS. NADH is recycled to NAD^+ by reducing the tetrazolium dye WST-1 to its absorbing form, detectable at 455 nm. The full detection cascade and dye structures are shown in Fig. S1.

C.2. Results & Discussion

A larger enzyme concentration per compartment would remedy the issues of high turnover requirements and make it easier to detect product in the described novel detection modes or enable the directed evolution of low promiscuous activities of weakly expressed enzymes with fluorescent detection. To address this problem, the enzyme content per compartment can be increased by monoclonal and homogeneous cell growth in droplets (Fig. C-3). In a new workflow, single cells are encapsulated in growth medium. Cell growth and protein expression are performed in modified reaction tubes that enable homogeneous growth and oxygenation. Now, substrate and lysis reagents are added in excess via ‘pico-fusion’ in a new chip design utilizing selective droplet coalescence. The whole workflow is demonstrated by the successful sorting of a newly introduced, low and industrially relevant activity, establishing the detection of AmDH activity by AADS.

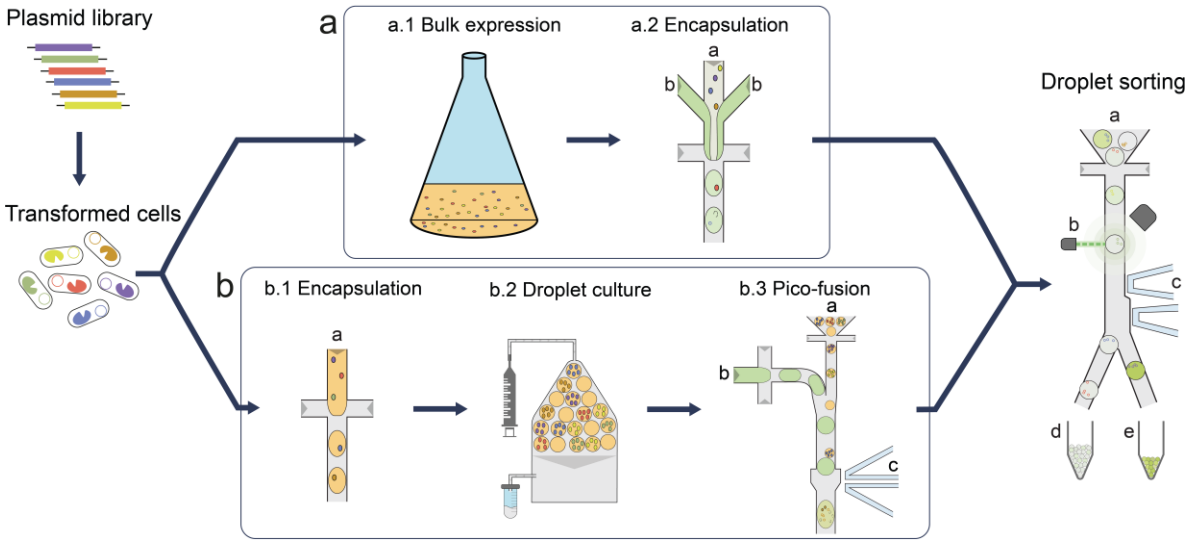


Fig. C-3: Overview of microfluidic cell lysate assays. *E. coli* cells are transformed with the plasmid library coding for enzyme variants of interest. Cells are Poisson-distributed to microfluidic droplets, so that the majority of droplets are monoclonal, containing no or one cell per droplet. **(a)** *Conventional single cell lysate assay*. Cell growth and protein expression is performed in bulk. Single cells (a.2a) are directly co-encapsulated with substrate and lysis solution (a.2b). The enzyme molecules expressed by each single cell need to generate enough substrate turnover for detection and droplet sorting. **(b)** *Clonal amplification in droplets to increase enzyme content*. Single cells are encapsulated in growth medium (b.1) and cultured for growth and protein expression in droplets (b.2). Substrate and lysis solution are added in a second step by 'pico-fusion' and subsequently sorted on chip.

C.2.1. A Device for Homogeneous Cell Growth in Droplets

For the first step, a robust device for the storage and handling of large volumes of droplet emulsion needs to be prepared. Handling droplets, such as re-injection into sorting or fusion devices as well as incubation, was performed in modified 0.5 mL reaction tubes (Fig. C-4). These incubation chambers are easily manufactured by opening access holes at the top and the side of a conventional plastic reaction tube, using biopsy punches and attaching tubing with high-viscosity plastic glue. Now, droplets can directly be collected from the droplet generation device into an oil-filled incubation chamber via the top access. Excess oil from flow focussing is extracted via the side access and the droplets pack tightly on top of the oil. These practically easy droplet incubation chambers provide benefits over other incubation approaches (such as incubation in unmodified reaction tubes or syringes): droplets are compacted, making re-injection more stable and regular, and destabilizing contact to air is minimized. Another benefit is the potential of pushing oil through the droplet emulsion from the top access, as done previously with custom 3D-printed devices for oxygenation of growing cells (Mahler *et al*, 2015).

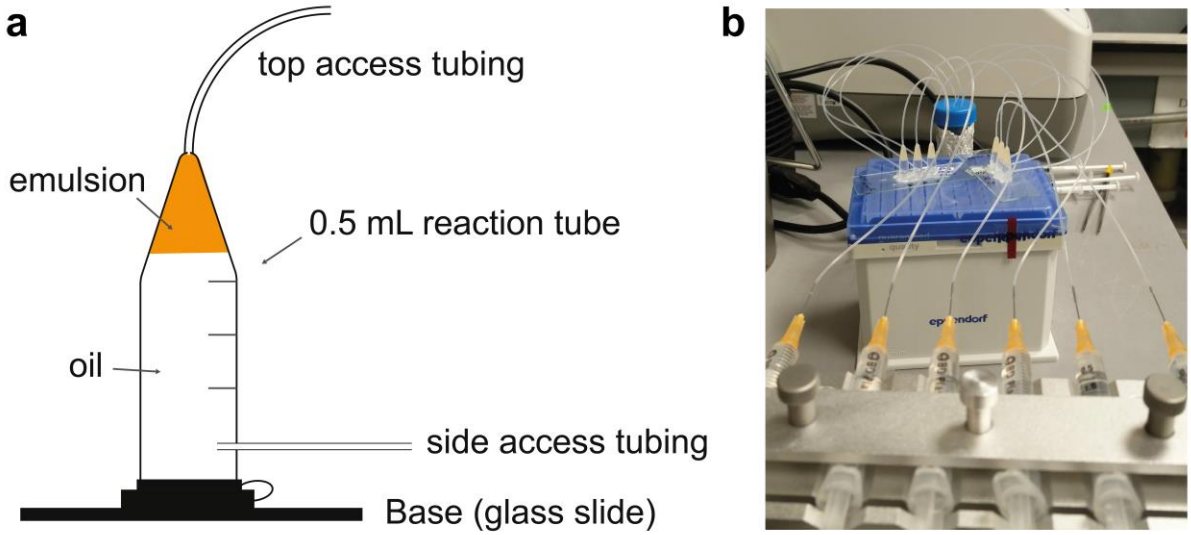


Fig. C-4: Droplet incubation chambers. (a) *Schematic droplet incubation chamber*. Incubation chambers are built from 0.5 mL reaction tubes by attaching polyethylene tubing with high viscosity plastic glue to openings made with biopsy punches. These chambers are quickly built and enable robust storage and convenient handling of droplet emulsions, which float and pack on top of the oil. Collection, storage, oxygenation and ejection of droplet emulsion is possible. **(b)** *Photograph of multiple droplet incubation chambers during oxygenation*. Oxygenation is performed by pushing oil through the emulsion via the top access tubing.

Now, cell growth was established in droplets in these incubation chambers. Static and oxygenated growth conditions were compared to test the hypothesis that oxygen availability is a crucial factor to achieve strong and homogeneous growth in each droplet. To measure cell density in droplets NADH-detection coupled to the chromogenic dye WST-1 was employed in AADS as a proxy of live cell count. First, single cells were encapsulated in growth medium and incubated in static conditions, without oxygenation or mixing, at 37 °C for 16 h. The live cell stain WST-1 was injected and the absorbance of 2500 droplets was measured. In the histogram of absorbances of occupied droplets (empty droplets are ignored), a main peak at low absorbance is indicated with a very strong tail (Fig. C-5a). This measurement shows that most droplets contain little cell growth while few droplets provide conditions that enable very strong cell growth as seen by high absorbances in the tail of the distribution. To establish growth homogeneity in form of similar levels of growth in all droplets, oil was pushed through the droplet emulsion in the incubation chambers thus providing oxygen and gentle mixing. Even with relatively low oil flow rates (4 µl min^{-1} per 100 µl of emulsion), homogeneous cell growth was achieved. The droplet population now shows a much narrower distribution (as indicated by the corresponding box plots in Fig. C-5a), as well as increased median absorbance, indicating higher total growth per droplet.

To quantify the effects of oxygenation compared to static growth conditions on the distribution of absorbances, the interquartile range (IQR) is used. The IQR is a robust measure of scale and is appropriate to compare the spread of data in this case, as the distribution of absorbances are non-normal. By convention, outliers are defined as droplets with an absorbance greater than the third quartile plus 1.5-fold IQR (indicated with an asterisk for the static culture in Fig. C-5a), expected to qual 0.35% in case of a reference normal distribution. Outliers can be interpreted as droplets with extreme growth compared to the bulk of droplets. Such extreme growth prohibits qualitative activity read-out, which is correlated to the enzyme content and thus to the cell count. In static growth conditions, 12.1% of all occupied droplets are high outliers, or rather, droplets with extreme growth, as indicated by the strong tail of the distribution in Fig. C-5a. When oxygenation is applied during cell growth, the percentage of outliers is reduced to 1.1% of all occupied droplets, thus quantifying the beneficial effect of cell growth in droplets under oxygen supply by providing growth homogeneity.

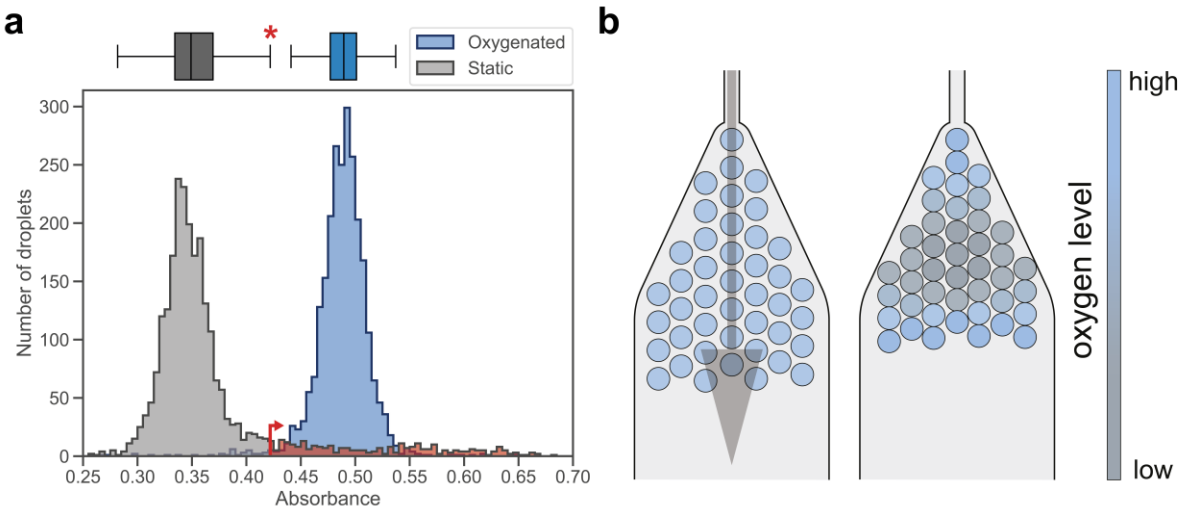


Fig. C-5: Gentle oxygenation via the oil phase enables strong and homogeneous cell growth in droplets. (a) *Histograms of WST-1 absorbance as an indicator of cell growth*. 2500 droplets measured. Cells were grown from initially singly compartmentalized *E. coli* cells in either static (grey) or oxygenated (blue) conditions for 16 h at 37 °C. Static cell growth was performed without oxygenation or mixing, while oxygenated cell growth was performed by perfusing the droplet emulsion with 4 µl/min HFE-7500 with 1% RAN surfactant. 10 mM WST-1 was injected into the droplets after growth as an indicator for cell density by detection of NADH in AADS. Moderate outliers in the static conditions, as defined with greater absorbance than the third quartile plus 1.5-fold interquartile range, are highlighted in red and indicated with an arrow and asterisk. **(b)** *Illustrative scenarios of oxygen availability during static and oxygenated cell growth in droplets*. Oil perfusion (*left*) provides oxygen and gentle mixing, thus leading to a homogeneous oxygen level and cell growth throughout the droplet emulsion. In static incubation (*right*), on the other hand, little mixing and exchange happens resulting in an oxygen gradient from the oil phase to the centre of the droplet emulsion, where the growing cells are depleting oxygen.

Static growth conditions have been used in previous studies (Hammar *et al*, 2015; Siedler *et al*, 2017; Beneyton *et al*, 2017), but do not lead to fast growth amplification *and* homogeneity,

a crucial requirement for precise assays. Identical conditions in every droplet are necessary to enable quantitative readouts of enzymatic activity, as the number of cells in every droplet directly determines the amount of enzyme per compartment. Previous studies using cell growth in droplets were based on qualitative selections, e.g. for antimicrobial resistance (Liu *et al*, 2016), not requiring differentiation of degrees of activity, or employed initial filtering for average cell growth via scattering properties in flow cytometry (Duarte *et al*, 2017). These previous assays might not have shared the same sensitivity to growth homogeneity as enzymatic turnover.

The heterogeneity in static growth conditions possibly stems from differential oxygen availability during incubation. Oxygen can diffuse into the emulsion from the oil phase; however, it is rapidly depleted by growing cells. This results in a gradient of oxygen availability within the emulsion in the device, with a lot of oxygen resulting in high cell growth closer to the bulk oil phase and a lack of oxygen limiting growth in most of the emulsion in the centre of the device (Fig. C-5b). Heterogeneity is thus introduced dependent on a droplets position in the device. Oxygenation during growth remedies this by supplying additional oxygen with the fresh supply of oil phase and provides a gentle mixing of droplets in the device, enabling stronger growth on average as well as more homogeneous growth.

A similar conclusion is found in the work of Mahler *et al* (2015). In their work, a custom 3D-printed incubation device is used with very strong oil flow (40 µl/min compared to 4 µl/min in this study), requiring a peristaltic pump in a closed circular setup, to provide a maximal amount oxygen to growing cells (Mahler *et al*, 2015). The set-up established in this chapter does not achieve maximal growth rates comparable to those in shaking flasks, yet oxygenation with comparatively low oil flow rates provides crucial growth homogeneity in a simple setup. Only regular syringe pumps and self-made incubation chambers from conventional reaction tubes are needed.

C.2.2. Induction of Recombinant Protein Expression

In the proposed cell growth workflow, recombinant protein expression must be induced after monoclonal growth in droplets (Fig. C-3). In the conventional single cell workflow, protein expression is conducted in bulk culture, as single cells containing the intracellularly expressed protein of interest are encapsulated subsequently. Here, growth and protein expression both must be performed in droplets. In many commonly used protein expression systems, such as the T7 promoter system with the inducer lactose or IPTG, external inducer is added after initial

growth allowing for fine control of cell density by induction timepoint and strength. Adding an external inducer poses a challenge to the microfluidic system, as the same amount of inducer needs to be added to each droplet, resulting in an increased workload and potentially reducing emulsion quality. Consequently, protein expression systems not requiring such an additional droplet manipulation step but offering similar control over expression strength and cell density were investigated. As a model system for protein expression, the oxygen-independent green fluorescent protein iLOV (Chapman *et al*, 2008) was used. This allowed facile measurement of protein content by determining fluorescence intensity, direct visualization in microscopy and heightened comparability of culturing conditions due to the oxygen-independent maturation.

Firstly, auto-induction of the aforementioned T7 promoter system was investigated, as it does not require the external addition of inducers such as IPTG after an initial growth phase. Auto-induction medium for the T7 promoter system contains lactose and glucose, as well as glycerol as the main carbon source (Studier, 2005). Glucose inhibits the recombinant protein expression from T7 promoters and is metabolised first by the growing *E. coli* cells. Once the available glucose is metabolised after an initial growth phase, the protein expression is freely induced by the lactose in the medium. Further cell growth and the recombinant protein expression is maintained by the available glycerol. To test the protein expression strength with T7 auto-induction, single cells were encapsulated in 50 pl auto-induction medium and grown for 16 h at 37 °C. The expression strength and cell count per droplet were estimated by breaking the droplet emulsion and determining the iLOV fluorescence intensity and the number of colony forming units of 1 µl aqueous phase, as well as acquiring microscopy images of the droplet emulsion (Fig. C-6). Successful protein expression and cell growth were determined with this expression system, however, adjusting the medium composition (amount and ratio of glucose and lactose) to achieve strong expression proved challenging, as culture conditions successful for bulk expression were not directly transferable to droplet culture. This is likely a result of different growth dynamics in droplets compared to bulk culture due to limited oxygen availability and suspension mixing in droplets.

Thus, a more straightforward expression system was tested next: constitutive expression. The *iLOV* gene was set under control of the constitutive lacI promoter. Now, continuous expression of the iLOV protein forgoes the problem of induction. In the same test, single cells grown in 50 pl droplets, expression from the constitutive promoter results in increased fluorescence intensity. However, the cell count per droplet rises drastically, reaching the highest number of

all tested expression systems with an average of 628 cells per occupied droplet. While the increased cell density might not be an issue, it can influence many downstream analyses. Many assays are influenced by unspecific background activity of cellular components. In fact, the model assay studied in this chapter, the detection of AmDH activity via to reduction of the dye WST-1 by NADH, is negatively affected by high cell counts due to the presence of intracellular reducing compounds (Huang *et al*, 2019).

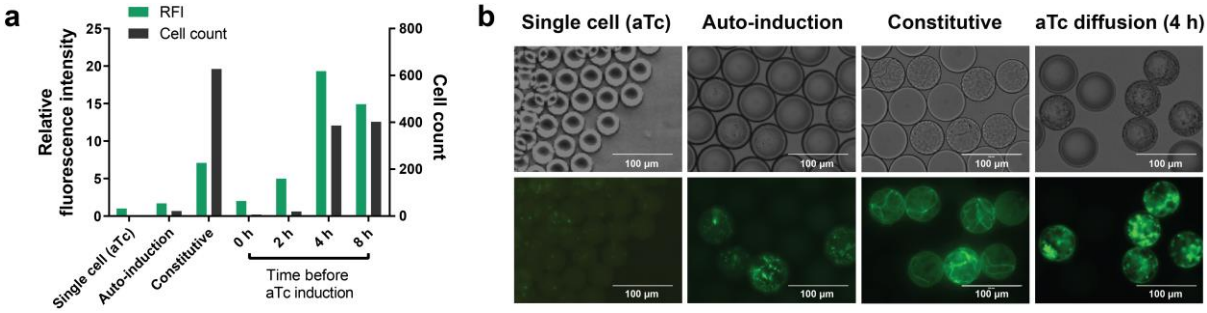


Fig. C-6: Protein expression in monoclonal droplet culture with different induction systems. Encapsulation of single cells in droplets of 50 pl growth medium, expressing the oxygen-independent green fluorescent protein iLOV under control of different promoter systems, incubated at 37 °C for 16 h. Auto-induction performed in ZYP-5052 medium, otherwise growth in TB medium. Induction via aTc is performed at different time points via solubilization of 400 ng ml^{-1} of aTc in the carrier oil used for oxygenation. **(a)** *Expression strength and cell counts*. Aqueous phase was generated from droplets by breaking the emulsion with an anti-static gun (Karbaschi *et al*, 2017), of which the iLOV fluorescence intensity was measured at 475 nm/510 nm as a proxy for recombinant protein expression strength. Cell count was determined by plating dilutions of the aqueous phase and counting the colony forming units. Measurements taken from ~4000 droplets in one biological replicate. **(b)** *Microscopy images*. Droplets were visualized in a fluorescence microscope, containing cells expressing iLOV under control of different induction systems. Strong cell growth visible in the bright-field image for constitutive expression, while strong fluorescence is visible in the constitutive expression and the aTc diffusion samples. Scale bar indicates 100 µm.

Consequently, an expression system resulting in strong protein production from few cells was sought next. The tetracycline promoter/operator system is another expression system frequently used for tight regulation and strong protein expression in *E. coli* strains (Terpe, 2006). Protein expression can be induced with low concentrations of anhydrotetracycline (aTc), which shows stronger repressor binding and lower antibiotic effect than the natural inducer tetracycline (Skerra, 1994). As an external inducer, it was shown that aTc is barely soluble in the oil phase used in droplet microfluidics, the fluorocarbon oil HFE-7500, prompting aTc to diffuse into the aqueous droplets. Thus, cells can be grown in droplets as described and at any point inducer can be added by changing the oil phase used for oxygenation to oil containing inducer. This induction approach via diffusion from the oil phase is a practically convenient way to achieve strong protein expression from few cells by controlling the induction time point (Fig. C-6) and does not require any additional droplet manipulation step. The WST-1 assay

used to detect AmDH activity in droplets is affected by high cellular background, preventing the use of long incubation times before induction, which would result in increased expression strength at the cost of high cell counts. Consequently, induction times of 2 h were used (resulting in 20 cells per droplet at 5-fold increased florescence) instead of longer incubation times such as 4 h (386 cells per droplet at 19-fold increased fluorescence). However, in assays without influences from cellular background, these longer incubation times could be used, potentially resulting in higher improvements to assays sensitivity than found for the WST-1 assay.

C.2.3. Active Delivery of Assay Components

To detect enzymatic turnover after cell growth, further assay components have to be delivered to the droplet. While for single cell assays lysate and substrate reagents can be added during encapsulation immediately leading to enzymatic reaction progress, the initial steps of cell growth and protein expression in droplets requires a delayed addition of reaction components. For the detection of intracellularly expressed enzyme, cell lysis is usually necessary to release the enzyme after cell growth. Additionally, a precise timing of the reaction start by the addition of substrate after growth allows for the adjustment of reaction time scales, potentially resulting in optimized signal to noise ratios and increasing the dynamic range for reactions with high chemical background. A high chemical background in form of a non-specific chemical reactivity that reduces available assay components can decrease assay sensitivity over long incubation times. Lastly, a step adding reaction components after cell growth can adjust the buffer composition of the droplet from the cell growth medium, widening the scope of detectable reactions.

Consequently, active delivery systems of reaction components to droplets were investigated next. Reagent addition is an important step in bioassays, which in microfluidics is most commonly achieved by electro-coalescence (Weng & Spoonamore, 2019; Payne *et al*, 2020). In this study, pico-injection was tested initially as a simple and robust way to add reagent to droplets: in a pico-injection device, small volumes of new reagents can be added to droplets at high speeds with a continuous aqueous flow adding fluid to passing droplets destabilized in an electric field (Abate *et al*, 2010). While pico-injection proved reliable and fast in experiment, volumes larger than 50% of the droplet could not be injected (data not shown). However, substrate and buffer component stability did not allow for high enough concentration for

successful pico-injection. Thus, a new device based on selective droplet coalescence was designed and manufactured.

With selective droplet coalescence, two populations of differently sized droplets can selectively be fused together (Mazutis & Griffiths, 2012). Thus, large droplets containing substrate and lysis solution could be generated on chip and fused with re-injected droplets after cell growth, allowing for the addition of large reagent volumes and thus buffer adjustment. Using AutoCAD 2018, a new device design was generated, combining features of previous droplet coalescence approaches. While a previous approach required a specific oil/surfactant composition resulting in passive droplet fusion due to interface destabilization in special channel geometries (Mazutis *et al*, 2009), a more generally applicable approach was pursued utilizing the electro-coalescence of droplets (Zagnoni & Cooper, 2009).

The new 'pico-fusion' device (Fig. C-7) combines flow focussing functionalities to generate large droplets of substrate and lysis solution on-chip with a droplet injector, spacing incoming droplets containing grown cells. Droplet speeds are synchronized and stacked in a delay channel, followed by electro-coalescence in a fusion chamber. Following multiple generations of device geometry optimization and fabrication (e.g. fusion chamber size and delay channel length), a final master was generated that allowed for reliable fusion of droplet at 200 Hz in conventional HFE-7500 1% RAN surfactant conditions. The final device reached non-detectable levels of multi-fusions and < 5% non-fused droplets. This additional step of droplet manipulation should not increase the sorting efficiencies, due to the very high rate of one-to-one fusions. However, the comparatively high number of unfused droplets poses a challenge to the sorting algorithm, needing to reliably exclude small non-fused droplets.

C.2.4. A Sorting Algorithm to Exclude Doublets

The detection signal in AADS is continuously monitored by an Arduino Due microcontroller. A simple point-over-threshold detection algorithm for droplets is currently employed (Gielen *et al*, 2016), which compares the detection voltage to a fixed threshold and actuates the sorting electronics should this threshold be crossed by an absorbant droplet. This approach is limiting the accuracy and applicability of AADS in two ways: Firstly, a negative sort is not possible: low absorbant droplets cannot be sorted out of a population of highly absorbant droplets. Secondly, droplet size cannot be distinguished potentially resulting in the erroneous sorting of small satellite droplets or large unwanted fusions in samples of bad emulsion quality, culminating in increased false positive rates. Droplet emulsion quality is reduced by many

factors, including the detergent based lysis agent used in microfluidic cell lysate assays as well as prolonged incubation and droplet manipulation. With the increased complexity of the cell-growth workflow compared to the single cell assay, the emulsion quality is slightly reduced. Additionally, while the pico-fusion step is reliably leading to no more than one-to-one fusions, a comparatively high number of non-fused droplets is produced. These non-fused droplets must be distinguished from fused droplets in the sorting algorithm to reduce the false positive rate.

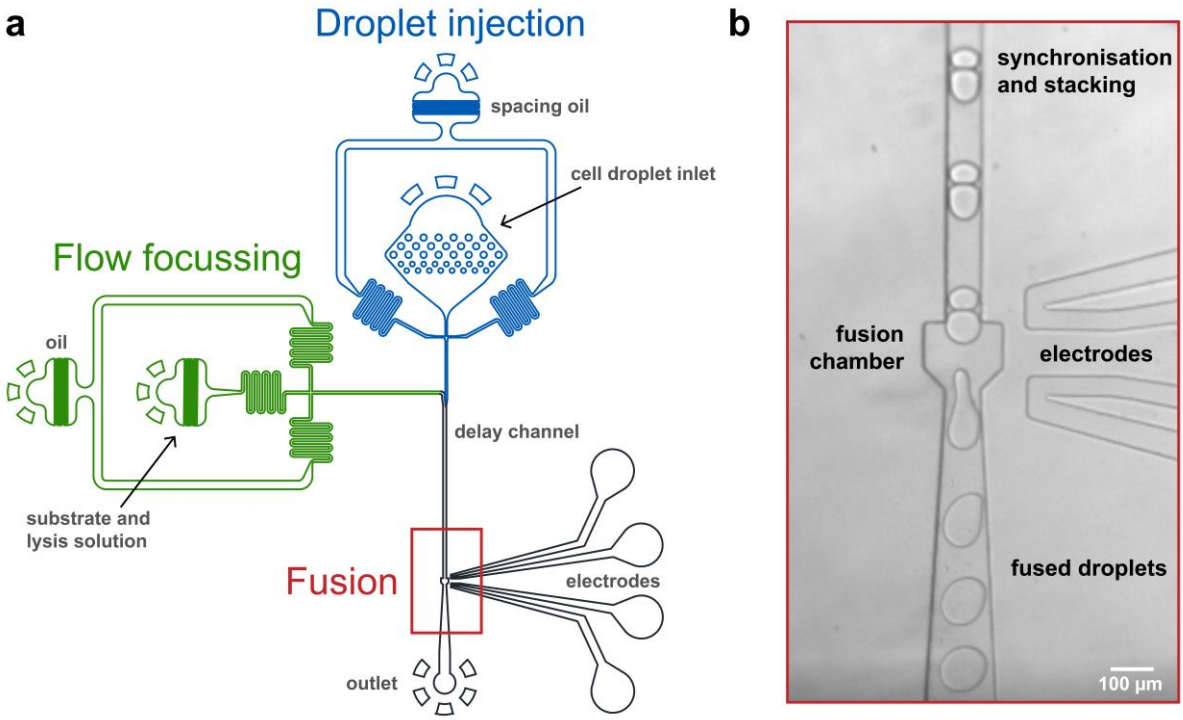


Fig. C-7: Pico-fusion device for selective droplet coalescence. (a) *Layout of the pico-fusion device*. Droplets (e.g. containing grown cells) are injected and spaced in the droplet injection part (*blue*), while droplets containing substrate and lysis solution are generated on-chip in the flow focussing section (*green*). Both droplet types are synchronized and stacked in the delay channel leading up to the fusion chamber, where an electric field causes selective droplet coalescence. A microscopy image of the fusion section (*red*) during operation is shown in b). **(b)** *Bright field image of the pico-fusion device in operation.* Two populations of droplets are synchronized and stacked in a delay channel. An electric field is applied to the fusion chamber via salt-water electrodes, causing droplet coalescence.

Thus, a new detection algorithm was designed, employing a true peak detection algorithm. The rising and falling edges of an event are identified, enabling the calculation of the true peak height as the maximum value in the event window. This allows the user to establish a sorting window and thus enables the selection of low absorbent values in the case of negative assays. More importantly however, the droplet size can be approximated as the droplet residence time in front of the detector by measuring the peak width as the time between the peak rising and falling edge. An intuitive visualization of the differences in the previously employed point-

over-threshold and the new true peak detection algorithms is shown in Fig. C-8. While the visualization shows absorbance as the measured variable for intuitive understanding, the true measurement in AADS is light transmission, inverting the algorithm: A highly absorbent droplet blocks a lot of light from the detector, resulting in a decrease of transmitted light and thus a decreased detection voltage. The full negative algorithm developed for AADS is available at https://github.com/pauljannis/AADSutility. This algorithm is employable for any other type of peak detection, e.g. for the detection of fluorescence in fluorescence-activated droplet sorting after sign inversion.

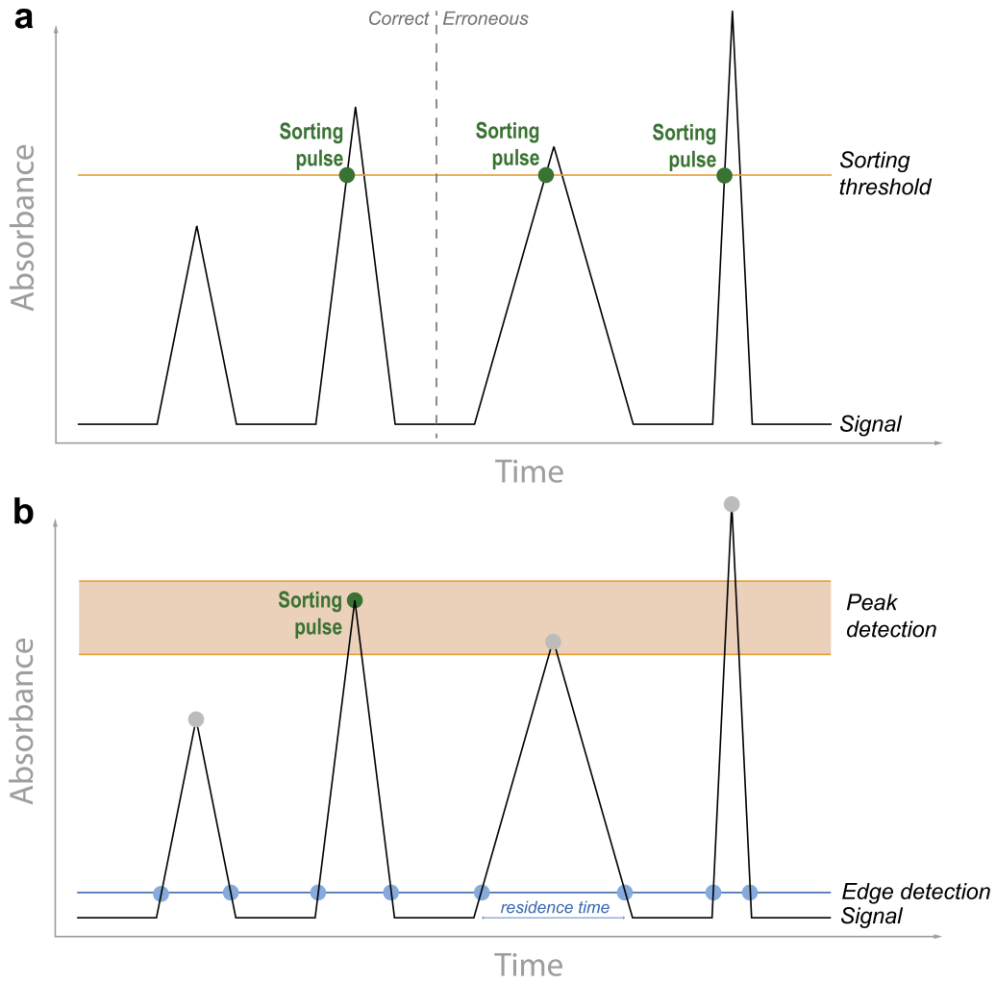


Fig. C-8: Comparison of detection algorithms employed in AADS. Four droplet types are exemplified, from left to right: Droplet with low absorbance, droplet with high absorbance, droplet fusion and droplet satellite. **(a)** *Point-over-threshold algorithm used by Gielen et al, 2016.* Sorting of droplets is actuated when the signal crosses a set threshold absorbance (green circles), potentially resulting in the erroneous sorting of fused large droplets (increased peak width) or satellites (lower peak width). **(b)** *True peak detection employed in this study*. Events are detected by rising and falling edges (blue circles), enabling the calculation of peak width corresponding to the droplet size. This allows for the exclusion of unwanted events such as large, fused droplets or small satellites. The true peak height can be calculated as the highest point within the event, enabling the user to set a selection window e.g. selecting low absorbant droplets.

C.2.5. Detection and Sorting of Low Enzyme Activities

As a model system for a novel enzyme activity, AmDH activity was chosen. When an AADH is converted to an AmDH, this novel activity is too low to be conveniently detected. AADH and AmDH activity can, however, both be detected in the same way in AADS via the detection of the co-factor NADH (Fig. C-2). In the published WT-background AmDH (K66Q S149G N262C), the novel activity lacks behind the original AADH activity with a low catalytic efficiency of 0.5 s^{-1} mM^{-1} (Ye *et al*, 2015), corresponding to < 1% of the parental AADH activity with a k_{cat}/K_M of 71 s^{-1} mM^{-1} (Gielen *et al*, 2016). Consequently, AmDH activity was not detectable in conventional single cell lysate AADS assay (Fig. C-9a). All droplets shown in the histogram have a low absorbance, indicating that droplets containing a single Poisson-distributed cell cannot generate any increased signal over the background compared to the unoccupied droplets without any cell. This is likely due to AmDH enzyme molecules supplied by a single cell not reaching the minimal turnover required for detection, determined as 1300 substrate conversions per enzyme molecule (corresponding to 10 µM product) in the formazan-dye coupled assay (Gielen *et al*, 2016). Insufficient product is produced in the case of the low AmDH activity and limited WT enzyme stability, preventing detection and sorting and, consequently, hindering directed evolution and engineering of AmDH variants with ultrahigh-throughput droplet screening.

This changes when the novel cell growth workflow is employed. Now, single cells are encapsulated in LB medium and grown for 2 h at 37 °C. Protein expression is then induced by oxygenation with inducer oil for 16 h at 20 °C: HFE-7500 containing 1% RAN surfactant and 400 ng ml^{-1} aTc is pushed through the droplet emulsion at 4 µl min^{-1} per 100 µl of emulsion to provide oxygen to the growing cells and inducer for protein expression. An excess of substrate and lysis buffer is added via pico-fusion, adjusting buffer conditions and enabling precise reaction timing. The absorbance of droplets is again measured via AADS, corresponding to the AmDH activity, after 2 h incubation to allow for sufficient reaction turnover. Cells grown in droplets and expressing the AmDH now show increased absorbance compared to the bulk peak of droplets not containing any cell (Fig. C-9c). For easy visualization, absorbance values greater than the background of 0.02 are highlighted in blue. However, as the formazan-dye assay also detects cellular NADH background, an increase in signal stemming from an increase in cell count has to be excluded. Cells containing an empty plasmid (not expressing AmDH) were used as a negative control. Droplets containing cells are not distinguishable from the

background of empty droplets, although the peak width increased indicating a low amount of background activity (Fig. C-9b).

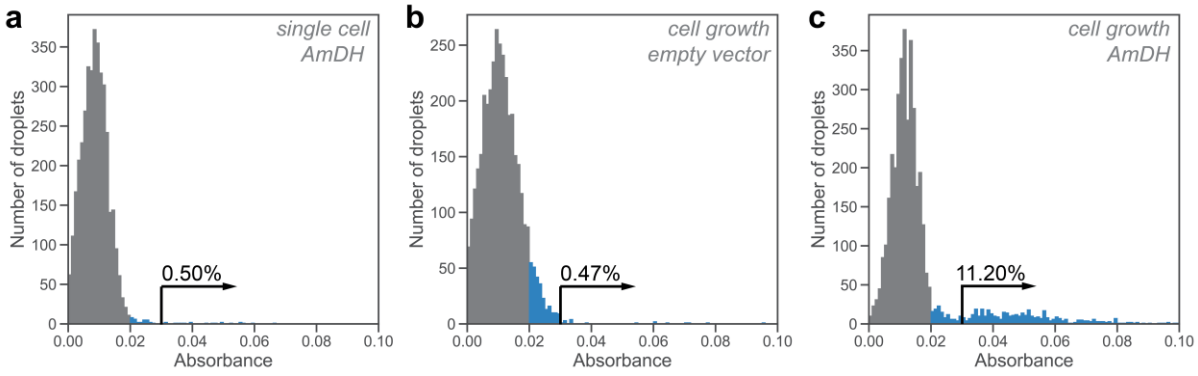


Fig. C-9: WST-1 absorbance as a readout for AmDH activity in AADS with different growth conditions. Absorbance of WST-1 measured at 455 nm. 4000 droplet absorbances per histogram. Values greater than 0.02 are indicated in blue to highlight detectable activity over background. Values greater than 0.03 are indicated with an arrow and the percentage of droplets higher than this selection threshold is shown. **(a)** *Single-cell AmDH measurement*. Co-encapsulation of single Poisson-distributed AmDH expressing cells with substrate and lysis solution, measured after 16 h of incubation. **(b)** *Cell-growth background*. Single non-expressing cells (containing an empty plasmid) were encapsulated in growth medium as a background control. Growth in 100 pl droplets was performed for 2 h at 37 °C, followed by oxygenation with inducer oil for 16 h at 20 °C (HFE-7500 with 1% RAN surfactant and 400 ng ml^{-1} aTc). For detection, droplets are then fused with 200 pl of substrate and lysis solution and absorbances measured after 2 h of incubation. **(c)** *Cell-growth AmDH measurement*. Single cells harbouring the AmDH plasmid were submitted to cell growth and protein expression in droplets in the same way as the background control in b). A large peak of droplets with low absorbances is visible in the histogram, corresponding to unoccupied droplets. Additionally, a tail of droplets with higher absorbance is indicative of detectable AmDH activity.

As the truest test to the technology, an enrichment experiment was performed. In an enrichment experiment, a dilution of active variants in inactive variants is recovered. It puts all components of a microfluidics assay together under biologically relevant time scales. Highly absorbent droplets were selected from a 0.5% dilution of AmDH expressing cells in non-expressing cells, following the cell growth protocol as described above. After sorting, a secondary lysate plate assay confirmed AmDH activity in 41% of tested variants (or 38 of 92 data points), corresponding to an enrichment factor of 82 achieved with the AADS AmDH cell growth assay.

This enrichment was only achieved when the new peak detection algorithm (described in C.2.4) was employed to exclude erroneous large fusions and smaller non-fused droplets. To exemplify this issue, Fig. C-10 shows the 2D plot of absorbance values with corresponding droplet sizes (approximated by peak width) of the positive control also shown in Fig. C-9c. The size selection for the main droplet population is indicated by red lines, highlighting the importance

of segregating the majority of correct one-to-one fusions from erroneous multi-fusions (larger peak width) and non-fused droplets (small peak width) for a successful enrichment.

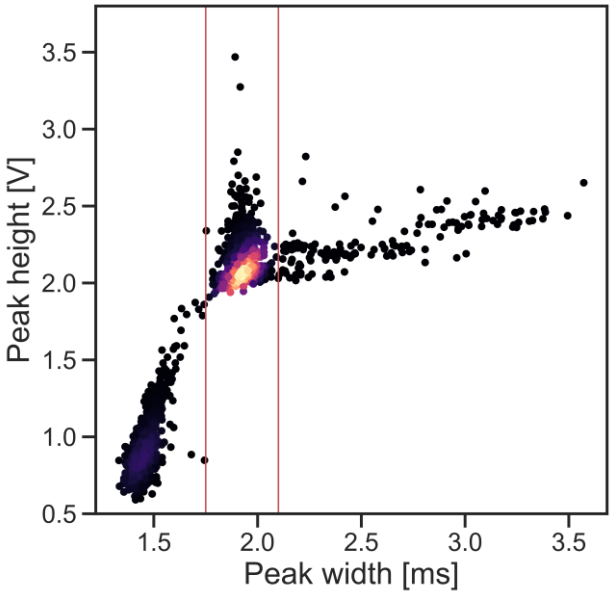


Fig. C-10: Droplet size (as peak width) and absorbance (as peak height) measured of AmDH expressing cells in the growth workflow. Gating of droplet size (red lines) is applied in the sorting to exclude large multi-fusions and non-fused droplets from correct one-to-one fusions. The filtered peak heights (converted to absorbance values) of this experiment are shown in Fig. C-9c. The large main population of droplets corresponds to non-occupied droplets, whereas a tail of droplets with increased peak height is distinguishable, corresponding to AmDH expressing cells.

Consequently, all parts of the cell growth workflow (homogeneous growth via oxygenation, addition of large volumes via pico-fusion, sorting with doublet exclusion) together enable the detection and selection of AmDH activity from the non-stabilized WT background, which is unselectable in conventional single cell lysate assays.

C.2.6. Quantification of Improvements

To quantify the improvements achieved in the AmDH assay when the in-droplet cell growth workflow is applied, an AmDH with detectable levels of activity in the conventional single cell assay is required. While the low activity of the WT-AmDH was made detectable via the cell growth workflow in the previous chapter, the improvement to activity comparing the two assays is quantified now with the stabilized $P4_{QM}$ AmDH. The stabilization of $P4_{QM}$ allows it to reach detectable levels of turnover in the conventional single cell AADS assay. Single cells expressing the stabilized AmDH $P4_{QM}$ are either directly encapsulated with substrate and lysis solution after bulk aTc induction or encapsulated in growth medium and subjected to in-droplet cell growth and induction, followed by addition of substrate and lysis solution via pico-fusion.

The product formation is measured at multiple time points via interrogation of absorbance at 455 nm in AADS. The difference in average detection voltage of occupied to unoccupied droplet peak at all time points is shown in Fig. C-11a, with the full histogram of exemplary time points shown in panels b and c. A stark increase in detection signal is recorded for droplets subjected to the cell growth assay: after 30 minutes, the detection voltage difference between occupied and unoccupied droplets approaches 4.0 V, corresponding to ~2.5 mM reduced WST-1. With progressing time the curve flattens indicating a quick conversion of most of the available 3 mM substrate. Contrastingly, the detection signal in the single cell assay starts plateauing after 2 h of incubation time at a signal of 1.3 V, corresponding to 1.1 mM reduced WST-1. The initial rate of product formation is thus increased 12-fold when using the in-droplet cell growth workflow compared to the conventional single cell assay.

Furthermore, the variation in droplet absorbances was measured. The distribution of measured absorbances for cells expressing the same enzyme variant ($P4_{QM}$) in many droplets gives an immediate measurement of the intrinsic assay variation. Normalizing the absorbances of occupied droplets for both approaches shows that the in-droplet cell growth assay not only increases the activity per droplet, but also reduces signal variation (Fig. C-11d). The peak of occupied droplets repeatedly showed 20-25 % relative standard deviation, whereas when cell growth is applied, the relative standard deviation dropped to 5-15 %. The variation in the assay was cut in half by measuring a signal from many cells instead of from a single cell per droplet. When a population of many cells is responsible for recombinant protein supply and thus activity readout, idiosyncratic effects such as variation of expression level per cell and time to lysis are averaged, resulting in a more consistent signal and reduced variation in the assay.

C.3. Conclusion & Outlook

The developed workflow utilising in-droplet cell growth and protein expression to amplify the signal in microfluidic lysate-based enzyme assays improves upon the conventional single cell lysate assays by increasing its sensitivity, enabling the detection of previously unselectable low activities from unstable starting points, such as the WT AmDH (see C.2.5). The development of the in-droplet growth workflow delivers multiple improvements to key steps in its process:

First, a new device for the incubation of droplets is used. Droplet emulsions are incubated and oxygenated for cell growth in modified reaction tubes that allow for the convenient and robust

handling of large emulsion volumes. These chambers allow access to densely packed emulsion in a much higher throughput and simpler set-up than previously used for cell analysis (Kleine-Brüggeney *et al*, 2019) or metagenomic screening (Mahler *et al*, 2015). Excess oil is drained from incubation chambers, resulting in a highly stable and compact droplet emulsion with easy access for manipulation such as oxygenation or ejection for pico-fusion and sorting.

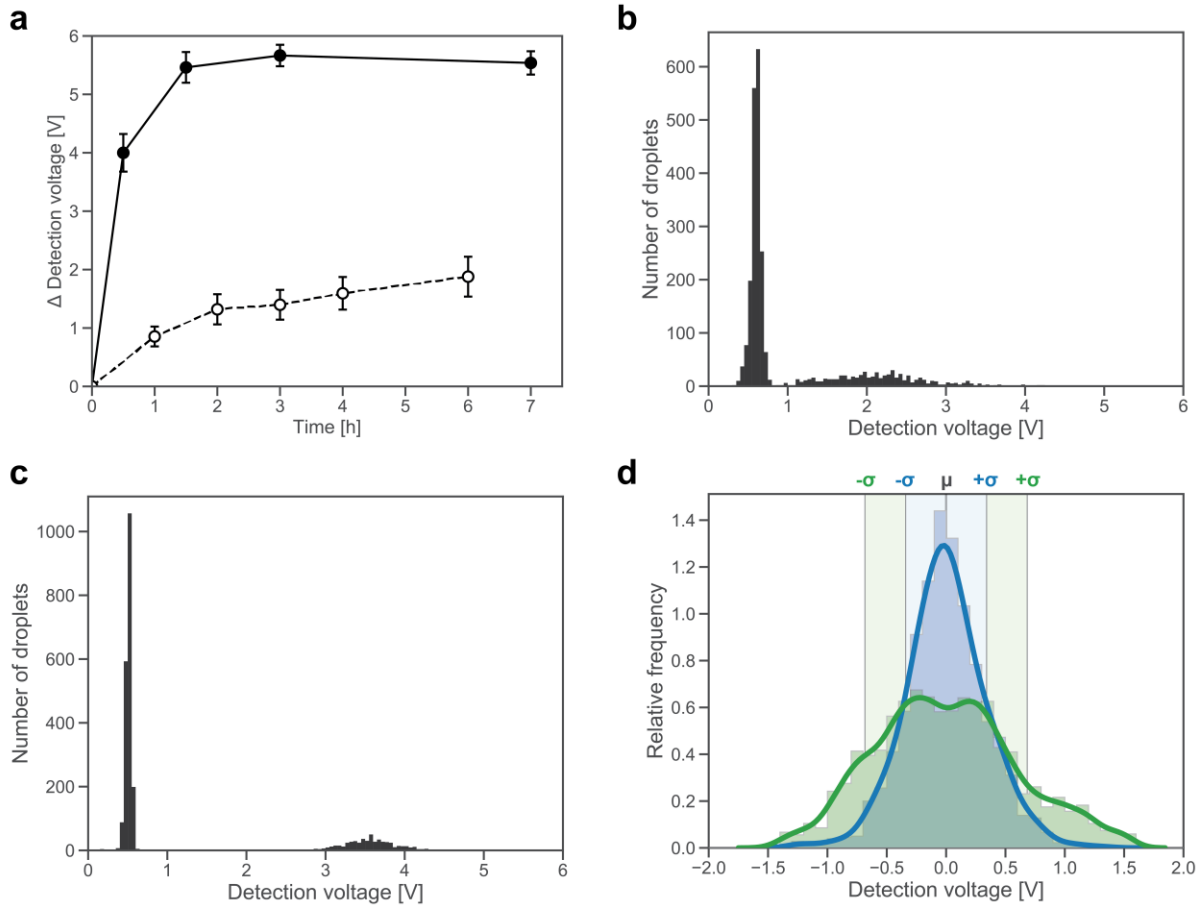


Fig. C-11: Quantification of improvements achieved by in-droplet cell growth. The stabilized AmDH P4$_{QM}$ was assayed in comparable conditions in conventional single cell and the proposed cell-growth workflows. **(a)** *Signal strength comparison.* The signal difference of occupied to unoccupied droplet peak is shown at multiple time points for the single cell (*dashed line*) and cell growth (*solid line*) workflows. Error bars represent standard deviation, n = 2500 droplets. **(b)** *Example histogram for single cell assay.* Single cells expressing the stabilized AmDH P4$_{QM}$ are co-encapsulated with lysis and substrate solution. Detection signal of n = 2500 droplets are shown after 4 h of incubation. A large peak of unoccupied droplets at a low detection signal is clearly separable from a broad peak of occupied droplets at higher detection signal. **(c)** *Example histogram for cell growth workflow.* Single cells are encapsulated in growth medium and subjected to the in-droplet cell growth workflow, inducing expression via aTc diffusion and adding substrate and lysis solution via pico-fusion. Detection signal of n = 650 droplets are shown 30 minutes after addition of the substrate and lysis solution. A narrow and strongly separated peak of occupied droplets is visible. **(d)** *Comparison of peak widths.* Assay variation of the single cell (green) and cell growth assay (blue) is quantified by isolating the occupied droplet peak, calculating relative frequencies and normalizing to a mean of one. Representative time points were chosen in 4 h (single cell assay, shown in panel b) and 0.5 h (cell growth assay, shown in panel c). Assay variation in the conventional single cell assay is higher (22.7% relative standard deviation) than in the cell growth workflow (10.7% relative standard deviation), as indicated by sigma intervals corresponding to 68% of all droplets.

Second, a passive delivery of assay components via the oil phase is set up. Continuous oil flow increases oxygen supply and mixing, resulting in increased cell growth and growth homogeneity between all droplets. Additionally, the inducer of recombinant protein expression aTc can be delivered passively (not requiring any extensive droplet manipulation) at any time point (decoupling cell growth from protein expression, e.g. for toxic proteins or high cellular background assays). In a similar approach the pH inside droplets was recently regulated by diffusion of acetic acid or diethylamine from the oil phase (Tovar *et al*, 2020), strengthening this practically easy and robust approach to supplying droplets. The achieved clonal amplification in droplets by cell growth and the control over expression conditions leads to reproducible signal amplification in each droplet.

Third, a device for the active delivery of an excess of assay components was established. The pico-fusion device enables the controlled addition of an excess of substrate and lysis buffer, favourably adjusting the reaction conditions and enabling precise timing of the reaction start. Now, reactions can take place outside of the growth medium conditions and the optimal ratio of enzymatic reaction to noise from uncatalyzed background reactions can be chosen, increasing the dynamic range for reactions with high chemical backgrounds.

Fourth, clonal amplification leads to more product readout. The cell amplification leads to more expressed enzyme per droplet, thus increasing the product formation per droplet. The improvement in initial assay activity was 12-fold for AmDH activity, reducing the limit of detection from originally 1300 substrate turnovers per enzyme molecule (Gielen *et al*, 2016) to around 100 molecules. This reduction in turnover per enzyme molecule necessary for detection enables the screening of less stable and less strongly expressed enzymes, such as the WT AmDH in this case. This workflow will thus prove important not only in engineering novel enzyme activities but also in future metagenomic screenings or in the establishment of less sensitive detection modes in microfluidics, such as mass spectrometry (Holland-Moritz *et al*, 2020) or electrochemistry (Goto *et al*, 2020). In all these cases, an amplification of enzyme content per droplet can resolve issues e.g. of low expression from natural metagenomic promoters or the strong turnover requirements in the large droplets of the novel detection modes. In assays with less cellular background than the used WST-1 detection, higher cell growth per droplet could be achieved potentially resulting in up to 48-fold signal amplification, based on quantification of GFP expression (Fig. C-6).

Fifth, the measurement of activity from multiple cells reduces variation compared to single cell assays. Strong variability in protein expression levels per individual cell (Elowitz *et al*, 2002;

Raj & van Oudenaarden, 2008) and stochasticity of cell lysis introduce variability to single cell assays that is not present when a population of multiple cells is measured and thus averaged. A reduction in assay variability will result in less false positive events and the recovery of more genuine hits. The problem of substantial cell-to-cell variability has been considered before (Duarte *et al*, 2017; Beneyton *et al*, 2017), but is here quantified for the first time and achieved in the context of enzyme activity. This advance to assay quality and reliability could furthermore prove useful in the context of deep mutational scanning (A.4), in which enrichment values are calculated by oversampling and reliable binning of variants, thus having special requirements to assay variability (Peterman & Levine, 2016).

Sixth, the recovery of sorted variants is improved. After droplet sorting, plasmid DNA is recovered and used for subsequent cell transformation. In the case of a single cell per droplet, the contained amount of DNA is low, resulting in a low recovery efficiency. With an ultrahigh copy plasmid, a recovery efficiency of 87% was calculated in an early study on single cell lysate screening (Kintses *et al*, 2012). The number of colonies after recovery from the cell-growth assay was usually 10 to 100 times the number of sorted droplets, varying with the actual time of cell growth applied, thus solidifying a higher chance of recovering all sorted variants with the medium-high copy number aTc-inducible plasmid.

Future work might include the combination of lysis independent assays with cell growth, such as employing yeast-display (Agresti *et al*, 2010), *E. coli* autodisplay (van Loo *et al*, 2019) or protein secretion (Beneyton *et al*, 2017) with the cell-growth approach, potentially reducing the cellular background and consequently increasing the dynamic range and achieved signal amplification.

Ultimately, the here presented workflow provides a general approach to increasing the sensitivity of microfluidic enzyme assays and will consequently make enzyme discovery and engineering campaigns more successful, bringing more challenging substrates and novel detection modes within reach of application.

D. Accurate Long-Read Sequencing for the Analysis of Directed Evolution

Mapping directed evolution and its trajectories in a fitness landscape requires a high-throughput readout of the identity of selected variants. With the ultrahigh-throughput activity assay for AmDHs established in the previous chapter, this chapter aims at developing an equally powerful sequencing readout. Accurate and full-length gene sequences are needed to provide the complete picture of a directed evolution phylogeny, containing mutations at any position and in any combination within the gene. The recent advent of third-generation sequencing technologies enables long reads at low quality, which in this chapter will be polished to accurate consensus sequences with the help of distinct molecular tags, so-called unique molecular identifiers (UMIs).

The content and figures of this chapter were published as Zurek et al*, 2020c.*

D.1. INTRODUCTION

The importance of directed evolution for the generation of novel biocatalysts is undisputable, as recognized by the Nobel Prize in Chemistry in 2018 (Arnold, 2019). The principles governing this intrinsically random process are yet to be fully understood. Mapping the sequence-function relationships of proteins (often called 'deep mutational scanning') has enabled great advances in our understanding of the local fitness landscape of proteins (see chapter A.4). Identifying the effect of many different mutations to a protein via the combination of high-throughput selection and sequencing has, for example, enabled the study of epistasis in the local fitness landscape of GFP (Sarkisyan *et al*, 2016), illuminated trade-offs between enzyme solubility and fitness (Klesmith *et al*, 2017) or investigated the determinants of substrate specificity of an amidase (Wrenbeck *et al*, 2017a). However, these studies are limited to the local exploration of sequence space and do not follow comprehensive random multistep evolutionary trajectories. Technical limits to assay throughput and sequencing read length result in many studies analysing either all single point mutations (e.g. Klesmith *et al*, 2017; Wrenbeck *et al*, 2017a) or defined multipoint mutations in short stretches or subdomains of the protein (e.g. Melamed *et al*, 2013; Olson *et al*, 2014; Araya *et al*, 2012). Additionally, most of these studies are performed on proteins for their native function, increasingly confining the studied effects to an existing fitness optimum (Table 3). Comprehensive multistep trajectories following function from initial to optimized activity would uncover previously unseen characteristics of protein fitness landscapes: by randomly distributing mutations throughout a whole protein and adding such mutations to variants consecutively in multiple steps, a comprehensive phylogenetic analysis of directed evolution is generated (Fig. D-1). Such a phylogenetic analysis of directed evolution would experimentally prove the extent of lineage formation in protein evolution, as expected by pervasive epistasis, and provide a solution to the question of the number of available distinct fitness optima (or the 'ruggedness' of a fitness landscape). These insights could inform future protein engineering campaigns regarding selection regimes or engineering approaches, by explaining the biochemical foundation of the distinct optimization solutions.

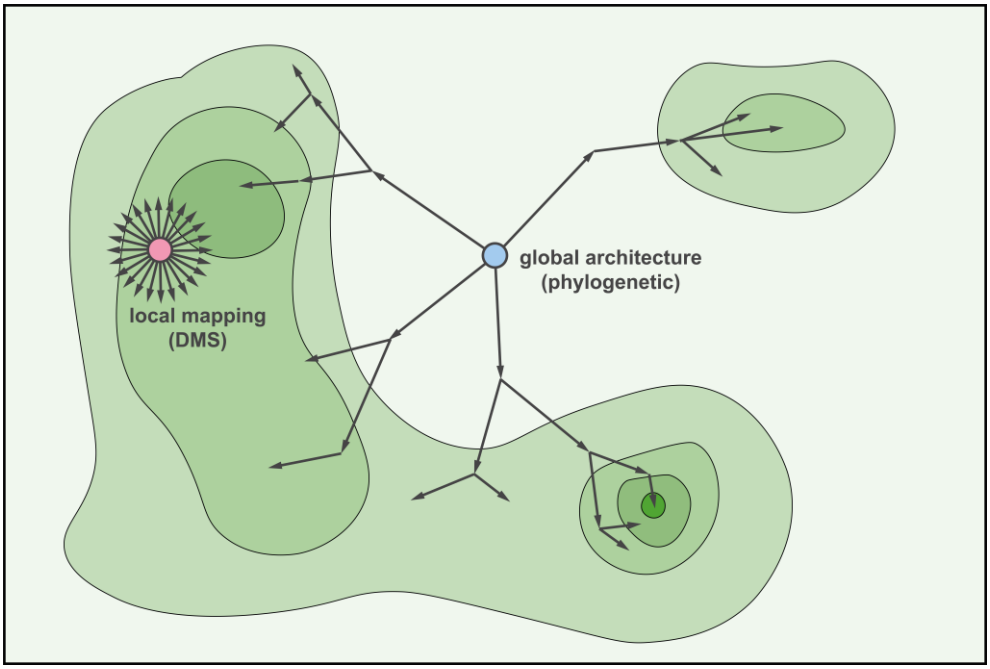


Fig. D-1: A topographical view of a fitness landscape. Deep mutational scanning (DMS, red dot) is usually (i) performed for the native protein function, thus being confined to mostly highly active variants near a fitness peak and (ii) performed with small steps, generally evaluating all single point mutations (Table 3). A phylogenetic analysis of directed evolution (blue dot), however, would start with a promiscuous low enzyme function and include many trajectories via many mutations in multiple rounds of evolution. The deep mutational scanning approach lays out a comprehensive view of the local fitness landscape, whereas a phylogenetic analysis of multiple rounds of directed evolution with long reads would capture the global architecture of the fitness landscape.

Studying random mutations throughout a full protein, however, requires the determination of the complete gene sequence, which ranges from 0.9 to 1.4 kb for typical genes (Xu *et al*, 2006). While Sanger sequencing has remained the standard for accurate verification of individual variant sequences in protein engineering, its low throughput cannot match the explorative power afforded e.g. by ultrahigh-throughput droplet microfluidic assays as described in the previous chapter. To map sequence-function relationships, a high-throughput sequencing technology is required. Sequencing-by-synthesis platforms such as Illumina sequencing have established themselves for cost-efficient high-throughput sequencing at high accuracy. With Illumina sequencing, however, an individual primary sequencing read length of less than 300 b falls far short of the average gene length, making it unsuitable to study multiple randomly occurring mutations across a full gene. There are practical innovations overcoming this limitation by associating multiple short reads into long sequences, so-called ‘synthetic long reads’. These synthetic long reads are well established in the field of genomics and have been applied there to resolve haplotypes and structural variations (Kuleshov *et al*, 2014; Zheng *et al*, 2016). When a single template molecule is present in many variants, e.g. in amplicon sequencing of 16 S rRNA variants or genes mutated by directed evolution, however, these

methods are not applicable as multiple molecules are still pooled per barcoding compartment. Synthetic long read methods specifically designed for amplicon sequencing currently suffer from uneven read coverage (Hong *et al*, 2014; Stapleton *et al*, 2016), high rates of chimeric reads (Lundin *et al*, 2013; Borgström *et al*, 2015) or are still limited in maximal read length (Burke & Darling, 2016; Yoo *et al*, 2020) due the chemistries used; limiting overall suitability, efficiency and throughput.

Third generation approaches, such as those developed by Pacific Biosciences and Oxford Nanopore Technologies, are capable of sequencing single molecules at immense read lengths. Because direct read lengths in nanopore sequencing are only limited by the DNA molecule length and not by the applied chemistry, impressive reads of more than 800 kb have recently been reported in a modified long-read protocol, while the read N50 in the standard protocol still exceeded 10 kb (Jain *et al*, 2018). A striking drawback of the technology, however, is the high raw read error rate of 5-15% (e.g. at least 50 bases in a 1 kb read are assigned incorrectly, van Dijk *et al*, 2018), preventing the accurate detection of true single nucleotide mutations in individual reads. Consequently, the innate high read length of third generation sequencing technologies would allow for the direct association of co-occurring mutations in gene variants to study epistasis in directed evolution, but low primary accuracy prohibits its use.

Accuracy is commonly increased by the generation of consensus sequences, in which multiple read copies are acquired to generate accuracy by averaging out sequencing errors. For genome assemblies, this process is straightforward because every sequencing read maps to a unique position in a draft genome assembly and overlaps with many others, facilitating stacking and consensus polishing (Loman *et al*, 2015; Vaser *et al*, 2017, Fig. D-2a). By contrast, template diversity in an amplicon sequencing library is typically low: many copies of rather closely related members must be sequenced accurately e.g. in metagenomic 16 S rRNA samples, immune repertoires or medical diagnostics. Also in protein evolution gene variants differing in few or even single point mutations need to be identified, as they can show drastically altered function. Primary sequencing reads with high error rates, however, cannot reliably be assigned to their template molecules when many template molecule candidates differing merely in point mutations are possible (Fig. D-2b). Consensus-based polishing is thus not feasible for unliked reads in all these cases, calling for the development a suitable solution.

Current approaches to increase nanopore amplicon sequencing accuracy rely on physically linking multiple copies of the same template molecule into one extended read (Fig. D-2c). Template molecules are circularized and amplified into one extended concatemerized fragment

containing many template copies in a single DNA molecule (Calus *et al*, 2018; Volden *et al*, 2018; Li *et al*, 2016). This physical link of template copies allows for the generation of consensus sequences with up to 99.5% accuracy, depending on gene length and resulting copy number in the concatemer. The remaining lowered error rate (at least 5 bases in a 1 kb gene assigned incorrectly) is not sufficient to distinguish individual point mutations and the applicability is limited by the necessity of very high molecular weight DNA molecules, which are difficult to generate and handle.

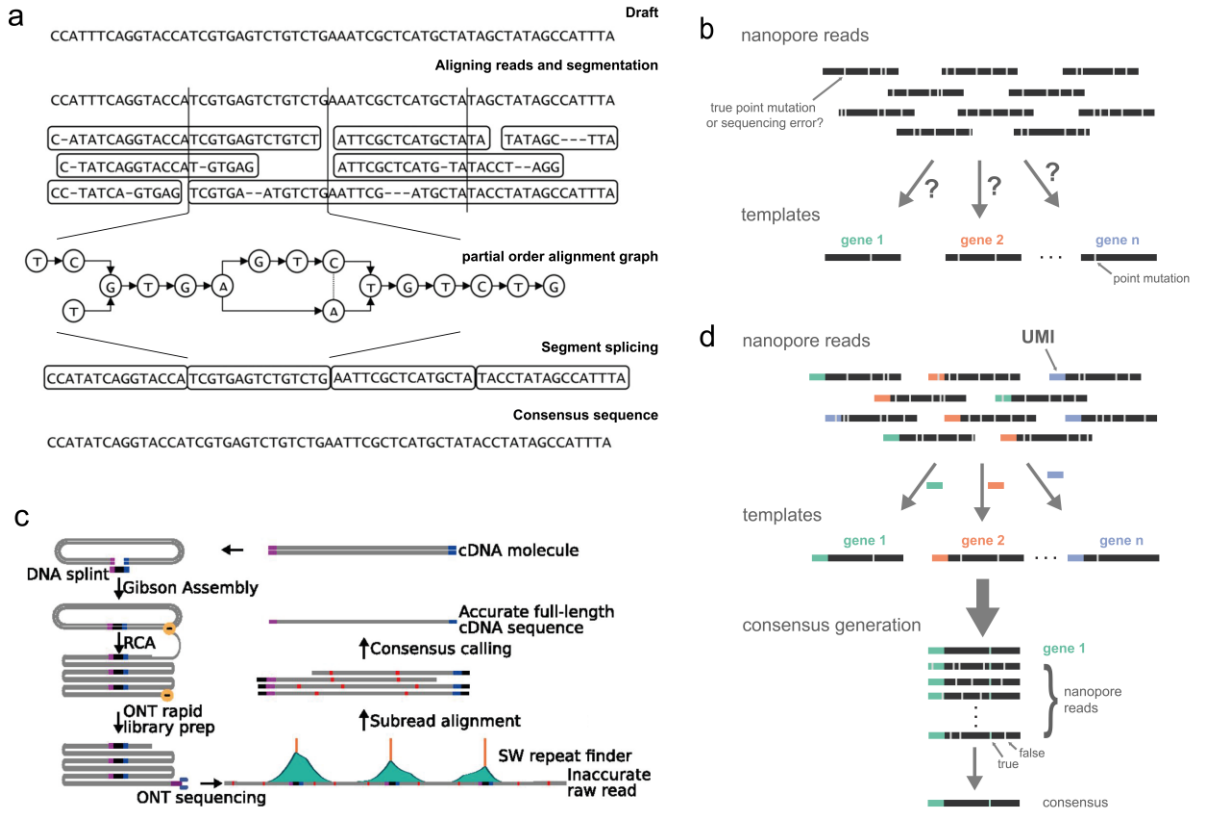


Fig. D-2: Consensus approaches to increase the accuracy of nanopore sequencing. (a) *Draft genome assembly available for alignment.* Long reads can be uniquely aligned to a draft genome assembly and used for consensus generation. Figure adapted from Vaser *et al*, 2017. **(b)** *The amplicon challenge.* Aligning is not possible when the template diversity is low: erroneous reads cannot be assigned reliably to a parent from a pool of highly similar variants, such as gene variants with single point mutations in directed evolution. **(c)** *Gene repeats in one molecule.* If alignment is not possible, accuracy can be increased by generating concatemerized repeats in a single molecule. Figure adapted from Volden *et al*, 2018. **(d)** *Unique tags could identify the template molecule.* Unique molecular identifiers (UMIs) could be used to generate distinct sequence tags for each template molecule that allow the assignment of erroneous sequencing reads to their origin. The clustering of those tags would allow for consensus generation at a more scalable level than repeats in one molecule.

A bioinformatic link between reads would remedy these issues: A random sequence tag, a so-called unique molecular identifier (UMI), added to each template molecule could be sufficiently distinct to allow the correct assignment of erroneous reads to their template molecule, even when these template molecules are very similar (Fig. D-2d). Such a UMI-

tagged sample would be easier to handle experimentally and provide scalability in sequencing throughput, allowing for generation of deeper consensus sequences, potentially increasing accuracy even further. UMIs were first used to mark the identity of template molecules prior to PCR amplification for quality control (McCloskey *et al*, 2007), but have since been extensively applied in various high-throughput sequencing approaches, such as to remove bias in molecule counts of RNA-seq expression data (Kivioja *et al*, 2012; Islam *et al*, 2014) and increase the accuracy of rare mutation quantification (Kinde *et al*, 2011; Schmitt *et al*, 2012). In this chapter, an accurate nanopore amplicon sequencing workflow will be established based on UMI-links, which allows for the tracking of gene variants over three rounds of microfluidic directed evolution of an amine dehydrogenase.

D.2. Results & Discussion

D.2.1. The UMIC-seq Workflow and Tools

Linking nanopore reads to their template molecule with unique sequence tags, UMIs, enables scalable and reliable bioinformatic clustering for consensus generation. The proposed workflow of UMI-linked consensus sequencing ('UMIC-seq') is shown in Fig. D-3 and allows individual raw reads of high error rate to be reliably assigned to their template molecule, which differs only by few nucleotide exchanges from other template molecules, by assigning unique barcodes.

A pool of similar variants – in this case selected gene variants during directed evolution – are identified by random sequence barcodes. These UMIs are tagged to the gene of interest by running two cycles of PCR with a stretch of random nucleotides in the PCR primers (Fig. D-3 step 2). This deliberately chosen low amplification in PCR reduces bias compared to other methods (Aird *et al*, 2011; Karst *et al*, 2021). Because the number of possible UMI sequences, e.g. $4^{50} \approx 10^{30}$ possible sequences for a 50 b UMI, far exceeds the number of molecules in a PCR reaction (e.g. 10 ng of a 2 kb gene $\approx 10^9$ molecules), each molecule will be tagged with a unique sequence. However, this number of uniquely tagged sequences is still too high to achieve deep sequencing coverage for consensus generation and the sample complexity needs to be further reduced. The sequencing capacity of a single MinION nanopore flow cell is 10-30 Gb, corresponding to a potential of ~10^5 clusters for 2 kb genes of 100 reads each. To reduce the number of uniquely tagged gene variants to a defined number suitable for consensus

sequencing, a transformation step is performed (Fig. D-3 step 3). After assembly of the UMI-tagged PCR product to a plasmid, the following transformation step serves two purposes: The number of assembled DNA molecules is defined in the colony count after transformation and easily controlled by plating dilution series. Additionally, the UMI-gene constructs are clonally amplified in the growing colonies. Thus, a small subset of UMI-gene variant combinations is selectively and homogeneously amplified, controlling the sample diversity with a simple molecular biology step.

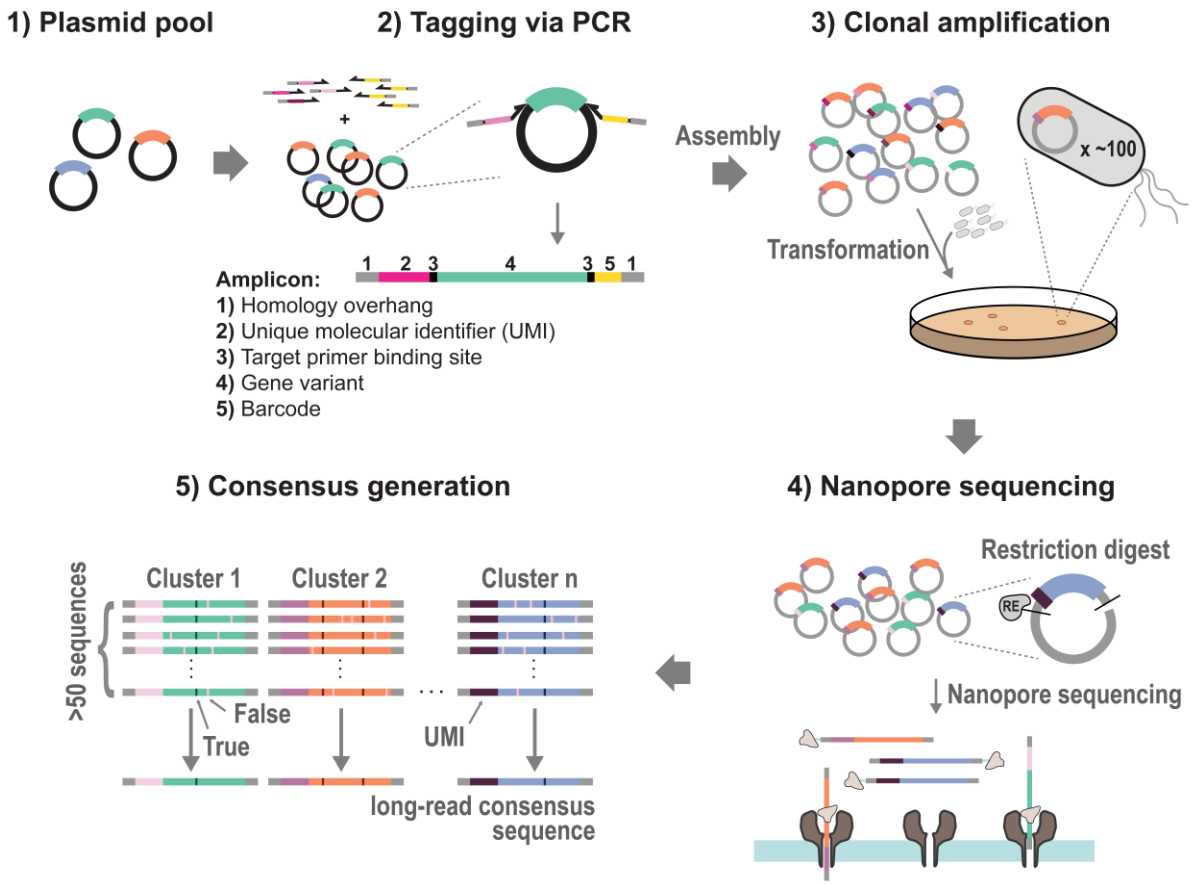


Fig. D-3: UMI-linked consensus sequencing (UMIC-seq) strategy for accurate amplicon sequences from nanopore long reads. (1) An amplicon library consisting of similar members, such as a plasmid pool of gene variants after directed evolution, is used as the input to the workflow. (2) UMIs are tagged to the gene variants in two cycles of PCR and re-assembled into plasmids, after which (3) the molecule diversity is restricted by transformation, resulting in the clonal amplification of a defined number of UMI-gene variant constructs. (4) Conventional nanopore sequencing is performed after excision of the UMI-gene construct, enabling the (5) clustering of reads based on UMI identity and the generation of accurate long-read consensus sequences.

This step reduces bias compared to the alternative cell-free approach. Diluting the template concentration in the initial PCR to the desired molecule count, e.g. ~1 pg of DNA to achieve 10^5 tagged sequences, could achieve a similar result. However, it relies strongly on accurate DNA quantification and requires many PCR reactions to amplify the low DNA molecule

numbers to required levels for sequencing, which will introduce strong bias due to PCR stochasticity (Kebschull & Zador, 2015). Consequently, the diversity restriction and clonal amplification in UMIC-seq is performed via cell transformation, combining the avoidance of PCR bias with experimental simplicity, rendering the molecule count easily identifiable in the number of colonies. To complete the practical aspects of the workflow, the plasmid is isolated from the grown colonies, the tagged gene of interest is excised via restriction digest and subjected to conventional nanopore sequencing with the SQK-LSK109 ligation sequencing kit.

Computational analysis was performed with custom-made python scripts (Zurek *et al*, 2020a) in multiple steps as shown in Fig. D-4. First, experiment-specific barcodes can be used to demultiplex the sequencing reads into bins, e.g. one bin for each round of directed evolution. Next, clusters according to the UMI identity are generated. For this, the UMI needs to be extracted from the sequencing reads and assigned to groups based on similarity, representing erroneous UMI reads from the same parent molecule. The implementation in the custom-made python scripts uses a fast and 'greedy' agglomerative clustering algorithm, similar to the implementation of CD-HIT (Fu *et al*, 2012) but chooses a random sequence as the first cluster representative (instead of the longest sequence) and uses relative alignment scores as similarities (instead of word dictionaries). These changes were made to adapt a greedy agglomerative clustering to the challenges of UMI reads from nanopore sequencers, for which CD-HIT – programmed to reduce redundancies of highly similar sequences in large databases – was unsuitable due to the short and constant sequence length with many insertion and deletion errors as well as the general high error rate in nanopore reads. The UMIC-seq clusterer was implemented with efficiency in mind by using: (*i*) python parallelism with the *multiprocessing* package, (*ii*) a fast striped Smith Waterman alignment implementation of the *scikit-bio* package, (*iii*) memory efficient read parsing with the *biopython* package and (*iv*) an option to ignore processing for cluster outliers. Most alternative unsupervised clustering protocols require all-vs-all comparisons, computationally inaccessible due to the immense number of UMI sequences. In the UMIC-seq implementation, a randomly selected UMI is compared to all sequences by alignment, of which UMIs above a user defined similarity threshold are removed. This reduces the pool of remaining UMIs and thus comparisons necessary linearly. A tool to approximate a suitable threshold is supplied within the UMIC-seq python package. Threshold approximation is performed by calculating one-vs-all alignment score histograms for 25 randomly selected UMIs, which provide an overview of the distribution of alignment scores (Fig. S2a). Then, clusters are generated for these 25 randomly selected UMIs over a

range of thresholds, leading to saturation of within-cluster similarities at a suitable threshold (Fig. S2b).

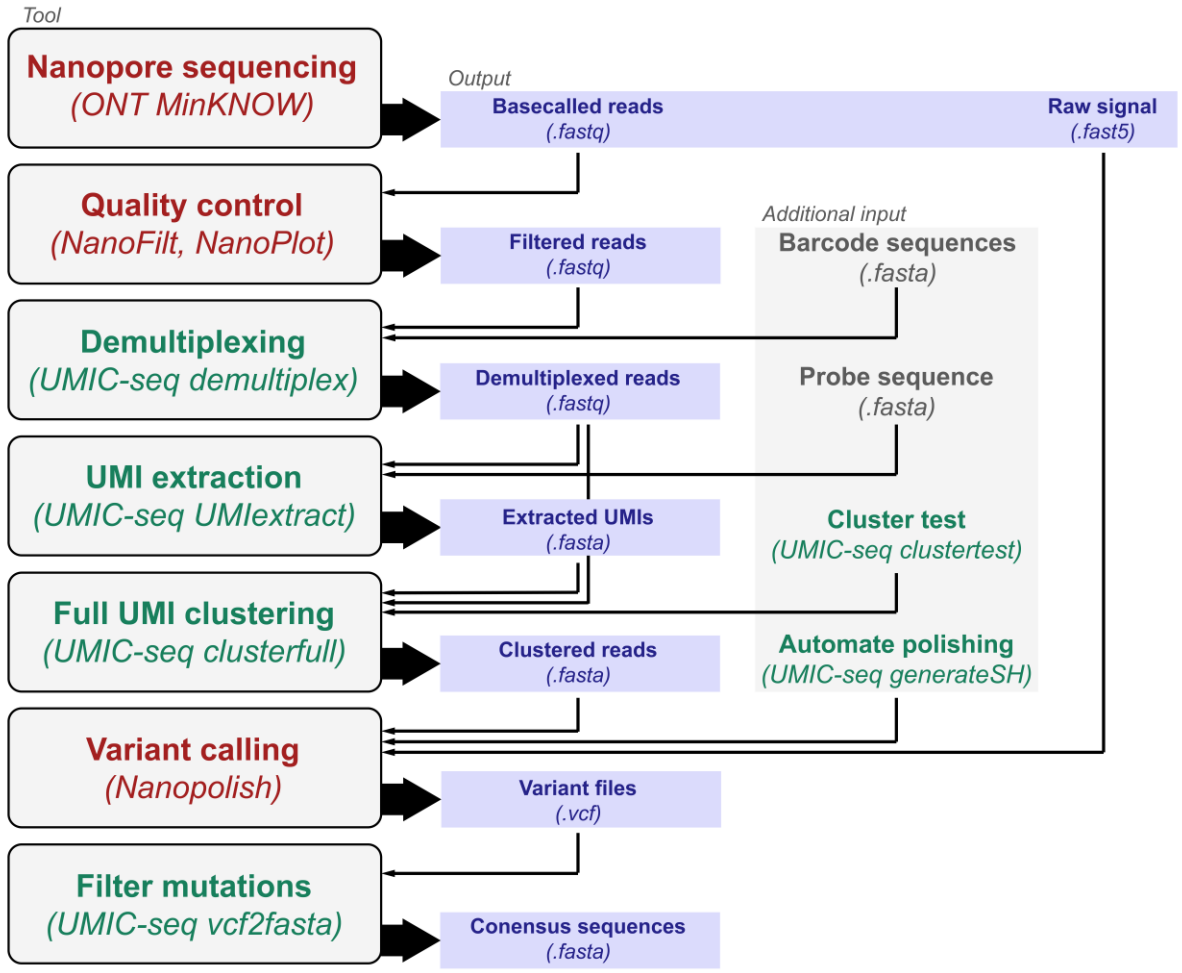


Fig. D-4: Overview of the bioinformatic tools used to generate accurate consensus sequences from UMI-linked nanopore reads. Tools that were custom made for the purpose of UMIC-seq are highlighted in green and deposited on GitHub (Zurek *et al*, 2020a). External programs used are highlighted in red: MinKNOW (Oxford Nanopore Technologies, ONT), NanoPack (De Coster *et al*, 2018) and nanopolish (Loman *et al*, 2015). Additional inputs that are required to complete the full bioinformatic workflow are barcode sequences used to demultiplex pooled experiments and a probe sequence representing a constant region adjacent to the UMI for UMI extraction. Additionally, a clustering threshold can be approximated with the supplied *clustertest* functionality and the nanopolish consensus generation can be automated with the supplied *generateSH* and *vcf2fasta* tools.

Finally, accurate consensus sequences can be generated form the read identities per cluster with the help of nanopolish (Loman *et al*, 2015). A full view of the individual steps taken in the bioinformatic pipeline, including all custom-made tools and external resources, are shown in Fig. D-4. A complete example of the UMIC-seq workflow is deposited on protocol exchange (Zurek *et al*, 2020b) with computational tools made available on GitHub (Zurek *et al*, 2020a).

D.2.2. An Exemplary AmDH Evolution

With an ultrahigh-throughput droplet microfluidic enzyme assay established (Fig. A-9, Fig. C-2), a short directed evolution of the $P4_{QM}$ starting point for improved AmDH was performed. With the goal of showing the applicability of the UMIC-seq workflow to establish the phylogeny emerging from directed evolution, mutations were introduced over the full length of the gene and a permissive selection threshold was set for three rounds of directed evolution. Specifically, each round consisted of the introduction of 1-3 random mutations via epPCR to the pool of previously selected variants (or the non-mutated $P4_{QM}$ gene in the first round) and the selection of the best 1000 variants of a total of 1 million droplets (or ~250,000 variants at $\lambda = 0.25$) in the single cell droplet microfluidic AmDH assay (Fig. A-9).

To verify sorting success, individual variants of the unsorted library as well as the output of each round of selection were tested for their activity. For each pool of variants, a lysate activity assay was performed for 46 randomly picked individual variants and the initial deamination rate in lysate was plotted relative to the non-mutated $P4_{QM}$ in Fig. D-5. The mean activity of sorted variants increases over the course of the directed evolution campaign from 0.17 to 0.80, showing successful application of droplet microfluidics to the screening of AmDHs. While a significant proportion of sorted variants is inactive (here defined as showing less than 20% of non-mutated $P4_{QM}$ activity) for all sorts, resulting from a high rate of false positives in droplet microfluidics, this number also drastically decreases from un-sorted library (76% inactive) to sorted library (48% inactive) in round 1. Thus, all sorted variants were next subjected to sequence analysis with the new UMIC-seq workflow.

D.2.3. Determination of UMI Efficiency and Sequencing Error-Rate

The UMI in the UMIC-seq workflow has to be distinct enough to enable reliable clustering of reads originating from one template molecule at the error rates common to nanopore sequencing. Simulations were performed to establish the necessary UMI length and feasibility on a large scale prior to committing to experiment. Increasing the UMI length increases its potential diversity and thus increases the dissimilarities between two UMIs randomly attached to template molecules. High dissimilarities between UMIs enable greater confidence in cluster assignment but increased UMI length also heightens the risk of introducing unwanted sequence biases such as secondary structures or self-complementarities.

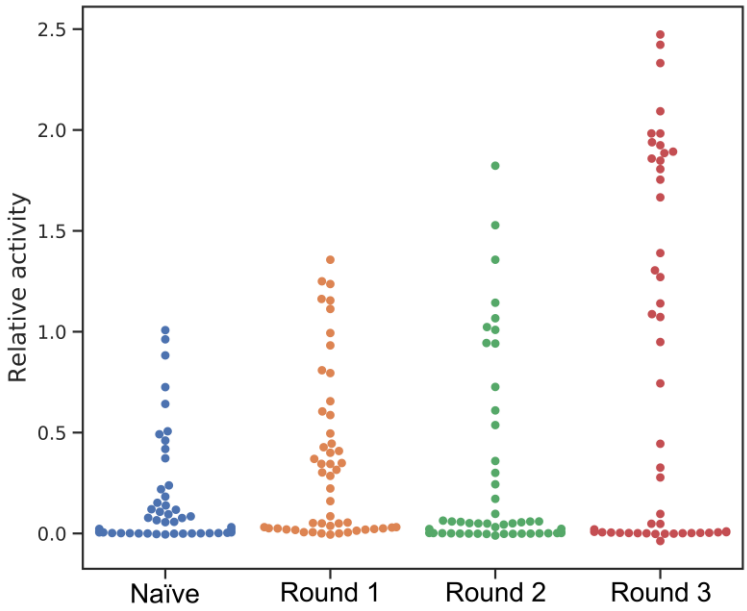


Fig. D-5: Lysate activity of sorted variants. Non-sorted library (naïve) and post-sort pools after each of the three rounds of directed evolution was tested for activity. 46 individual variants were randomly selected and assayed for AmDH activity as initial deamination rate relative to the non-mutated $P4_{QM}$ parental variant. The increase in mean activity confirms the successful enrichment of increasingly active variants over the course of the microfluidic directed evolution.

Clustering efficiencies were thus estimated with different UMI lengths at a range of error rates common to nanopore sequencing (Fig. D-6a). UMIs were modelled at different error rates as 1000 completely random sequences with 50±5 reads each. Errors were modelled as completely random substitution errors only. This set of 50,000 sequences was then clustered with the UMIC-seq clusterer (see D.2.1) and resulting clusters were compared to the original read identity. The clustering stringency was set to a consistently high level, automatically determined with via the 'threshold approximation' functionality, to exclude false assignments at the risk of not finding all members of a certain cluster. False assignments would increase error rates, but incomplete clusters could be compensated by increased sequencing depth. The comparison of UMIC-seq cluster assignment and information about the true sequence identity allows for the calculation of two clustering metrics: homogeneity and completeness. Homogeneity is a measure of correct assignment in clusters, a high homogeneity relates to a cluster containing only members of a single class. Completeness is a measure of efficiency in clustering, a high completeness relates to all members of a class being assigned to a single cluster (Rosenberg & Hirschberg, 2007). The high clustering stringency results in high homogeneities throughout, ideal to obtain high consensus accuracies from reliable clusters. Homogeneity is > 90% for all UMI lengths and often > 98% at lower error rates and higher

UMI lengths. At high error rates or for low UMI lengths, however, completeness of clustering is drastically reduced. This led to the conclusion of 50 bases as a suitable UMI length for experiment, as a good trade-off between sequence length and clustering efficiencies (>90% completeness at 14% error-rate).

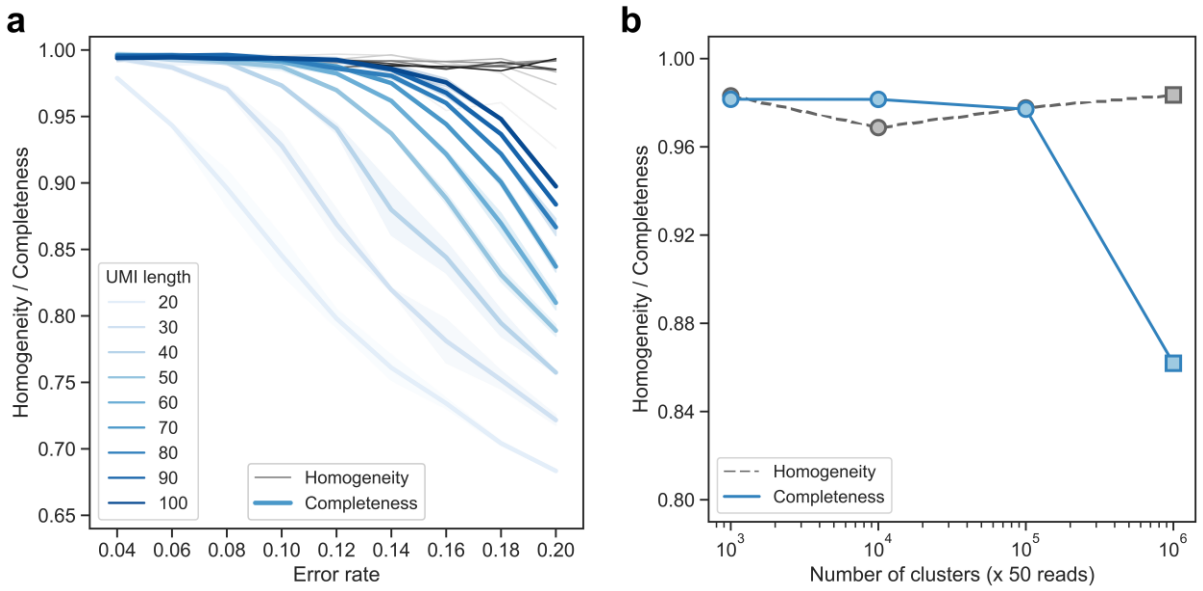


Fig. D-6: Influence of UMI length, error rate and library size on clustering efficiency. Cluster homogeneity relates to a cluster containing only reads of a single UMI. Cluster completeness relates to a single cluster containing all reads of one UMI. UMIs were generated as perfectly random sequences and sequencing errors were modelled as unbiased substitutions errors. **(a)** *Influence of UMI length and sequencing error rate.* 1000 UMIs were simulated with 50±5 reads each at different error rates and clustered using the UMIC-seq clusterer in triplicates. UMI lengths were varied from 20 b to 100 b indicated from light to dark. Completeness shown in blue with bands indicating standard deviation. A UMI of 50 bases was chosen for experiment as a trade-off between UMI length leading to unwanted biases and high clustering completeness (>90% at 14% error rate) **(b)** *Influence of library size.* Clusters of 10^3 to 10^6 UMIs with 50±5 reads each were generated at 8±1% error-rate and clustered using the UMIC-seq clusterer (circles) or MMseqs2 Linclust (squares). While the clustering completeness drops to 86% for libraries of 10^6 variants, it still shows the applicability for homogeneous clustering even beyond current flow cell capacity.

To verify that these properties would scale to the sequencing of large libraries, the simulation was repeated with 50 b UMI length but at increasing sequence count. Now, 10^3 to 10^6 UMIs were generated with 50±5 erroneous reads each at an error-rate of 8±1% and subjected to clustering. The results show very efficient clustering of libraries of up to 10^5 variants, which proves applicability for larger future projects. For even larger libraries of 10^6 variants (corresponding to $10^6 \times 50$ erroneous reads), the exponentially increasing number of pairwise comparisons necessary for the UMIC-seq clustering algorithm reached its limit and a very fast linear clustering algorithm, MMseqs2 Linclust, was used instead (Steinegger & Söding, 2018). This resulted in a drop in completeness but still allowed for homogeneous clusters, showing

applicability of the UMI clustering approach for the generation of reliable clusters even in extreme cases greater than current flow cell capacity[3].

Sorted variants from the three rounds of directed evolution (D.2.2) were thus tagged with 50 b UMIs for clustering and consensus generation as well as experiment-specific barcodes indicating the round of directed evolution. The plasmid DNA of ~3000 UMI-tagged colonies was isolated from a dilution series after transformation to account for oversampling of the expected ~1000 sorted variants per round of directed evolution. Sequencing and base-calling followed the standard amplicon sequencing kit instructions (SQK-LSK109, Oxford Nanopore Technologies) on one MinION R9.4 flow cell. Sequencing was stopped when approximately 100-fold sequencing oversampling was reached (~9000 consensus sequences of 1.2 kb at 100-fold sequencing coverage: 1-2 Gb), to preserve the remaining flow cell capacity. Quality filtering, clustering and consensus generation was performed as described in D.2.1. The isolation of sequencing DNA from growing colonies did not seem to introduce bias, as the cluster sizes were almost normally distributed with an IQR from 57 to 129 centred around a median of 94 reads per cluster (Fig. S3). This highlights the robustness of this method: UMIC-seq does not require accurate quantification of DNA concentrations, the molecule count is directly mirrored in the number of countable colonies which are uniformly amplified by the growing cells. This is in direct contrast to the recent results of Karst *et al.*, in which the PCR amplification of very few initial molecules requires guessing due to unreliable DNA quantification and a very wide spread of consensus coverage with only few sequences showing high coverage due to strong PCR bias after a total of three PCR reactions (Karst *et al*, 2021).

Potential mutations in the consensus sequences are identified with the non-mutated gene sequence as reference with nanopolish (Loman *et al*, 2015) and compared to 180 pairs of Sanger sequences (a total of 173 kb) for the calculation of error rates. Nanopolish conveniently associates each identified mutation with a confidence score, the so-called support fraction. Interestingly, filtering mutations based on their support fraction results in the reduction of false positive mutations (mainly insertions and deletions) while maintaining true mutations (mainly substitutions). When disregarding mutations with support in less than 60% of reads, errors which consistently occur in one of the two possible read directions are discarded, increasing the overall sequencing accuracy (Fig. D-7a), as shown in a recent study in a different context (Gilpatrick *et al*, 2020).

[3] A library of 10^6 gene variants of 2 kb length would require 100 Gb of sequencing output at 50-fold coverage. This lies far beyond current MinION flow cell capacities of ~10 Gb.

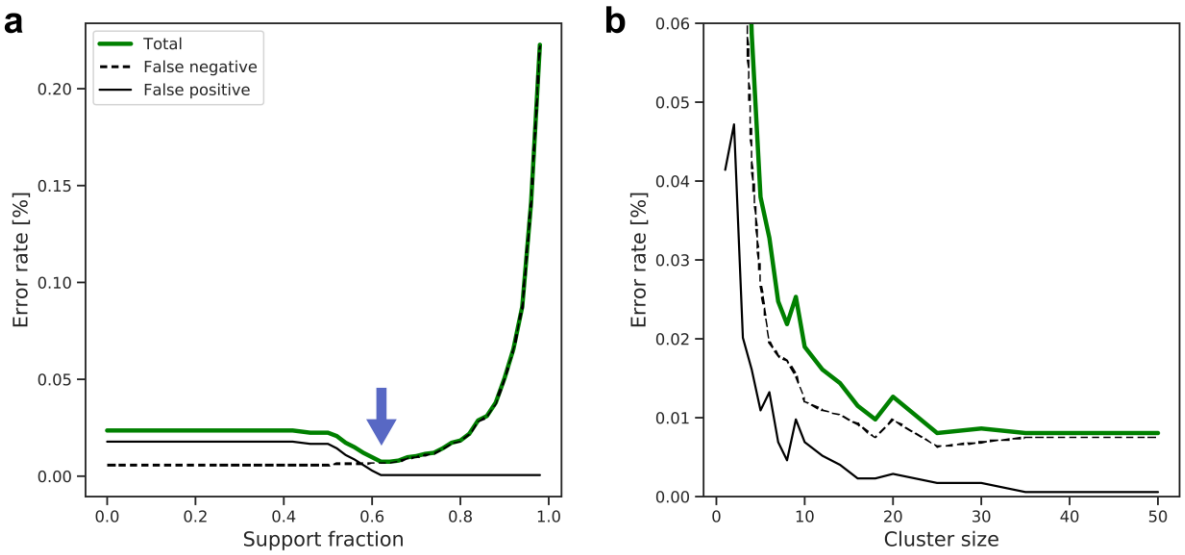


Fig. D-7: Analysis of errors in nanopore consensus sequences. Sanger sequences (180 pairs, 173 kb) can be paired to their respective nanopore consensus sequence via the UMI identity and were consequently used to confirm the error rate in nanopore sequences. **(a)** *The support fraction of mutations distinguishes false positives from true mutations.* The consensus mutations identified by nanopolish are associated with a read support fraction. Filtering of reads based on minimum accepted support fraction influences false negative (dashed line) and false positive (solid line) mutation calls. By accepting only mutations with a support fraction of at least 0.6, false positives are removed without increasing the false negative rate. **(b)** *Influence of sequencing depth on error rate.* Nanopore consensus sequences with a matching Sanger control are sampled and different depths and evaluated. The error rate is reduced to less than 0.01% when more than 35 reads per consensus sequence are available.

Applying this filtering threshold resulted in a final average error rate per base of 0.008% calculated in reference to 173 kb of Sanger sequencing control. To investigate the effect of cluster size on error rate, nanopore clusters with a matching Sanger sequencing control were sampled at different cluster sizes from 1 to 50 reads. Calculation of consensus error rates for those sub-sampled clusters shows that 35 sequencing reads suffice to calculate an accurate consensus sequence (Fig. D-7b). The Sanger sequencing control identified a total of 654 true mutations, of which 98% were correctly identified in the nanopore consensus sequences. The average error rate per mutation mainly stemmed from false negatives (1.99%), with false positive mutations being the minority error in the nanopore consensus sequences (0.16%). The higher rate of false negatives is likely to arise from clustering errors, as mutations in a mixed cluster would not reach necessary support fraction and be ignored, resulting in non-mutated wild-type-like sequences.

While a slight inflation of non-mutated sequences does not interfere with further downstream analysis, its cause was investigated. For some clusters containing false negatives, long homopolymer stretches in the UMI were identified, leading to a comprehensive analysis of the UMI composition. In fact, an AT-rich homopolymer bias was identified in all sequenced UMIs

(Fig. D-8), potentially leading to the reduction of the UMI diversity and in extreme cases leading to mixed clusters and consequently non-confident mutations and false negatives. This issue could be remedied by using a higher quality oligonucleotide and employing a non-continuous UMI design, i.e. short stretches of fully randomised bases interspaced by few defined bases, so that long homopolymers are not allowed to form. In the end, a slight inflation of reference sequence is not an issue for downstream analysis and crucial variants will not be missed thanks to the oversampling on variant level.

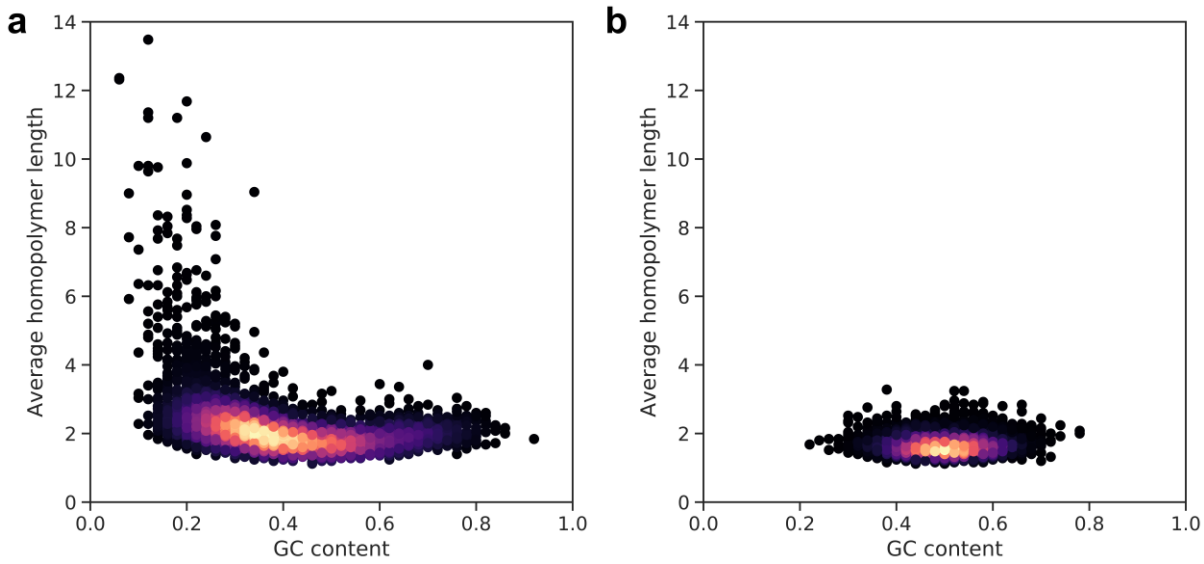


Fig. D-8: Analysis of UMI sequence composition. (a) *Sequenced UMIs.* All sequenced UMIs (n=6575) visibly show AT-bias and outliers with strong homopolymer composition. AT-homopolymers could reduce dissimilarity between UMIs and lead to clustering errors. A non-continuous UMI design, e.g. interspacing fixed bases, could reduce homopolymer content in future experiments. **(b)** *Random UMIs.* Perfectly random UMIs are simulated and analysed (n=6575) for comparison.

D.2.4. A Phylogeny of Directed Evolution

The sequence information obtained in this workflow can identify multiple simultaneously occurring mutations throughout the whole gene. The number of mutations per variant increases per round of directed evolution, as expected, from a median of 1 to 3 to 5 mutations per variant (Fig. D-9). With multiple mutations randomly distributed throughout the gene and building on each other over multiple rounds of directed evolution, epistatic interactions could be prevalent. Epistatic interactions are difficult or impossible to predict but are the driving force of lineages and phylogenies (Miton & Tokuriki, 2016; Starr & Thornton, 2016) and can now be identified thanks to the long-read sequencing information.

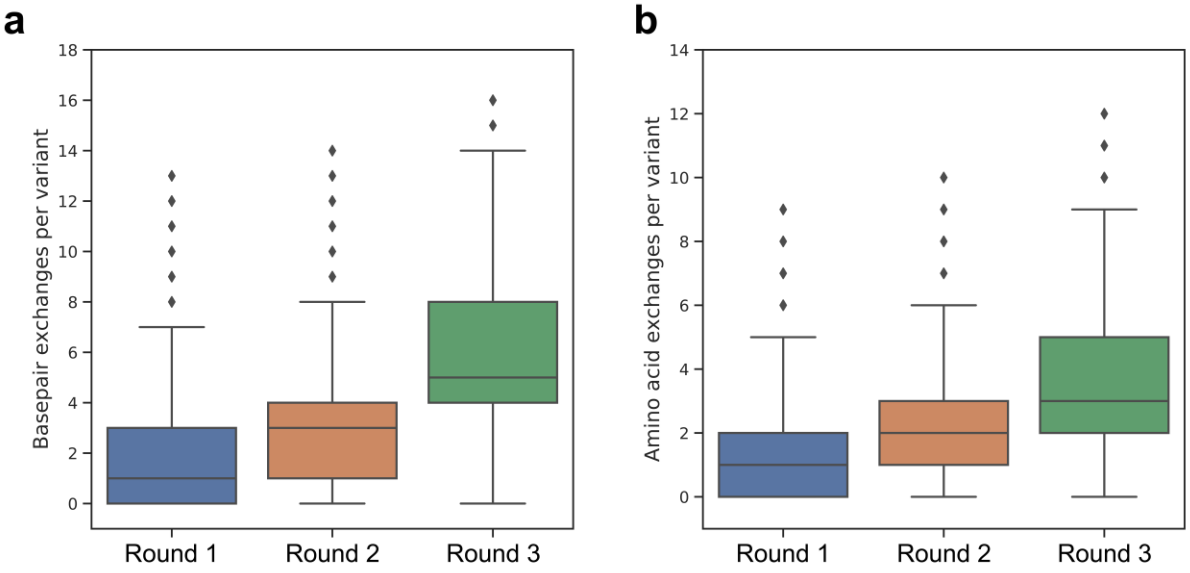


Fig. D-9: Mutations per variant over the course of three rounds of directed evolution. The distribution of **(a)** base pair and **(b)** amino acid mutations per variant are shown for each round (n=1665, n=1620, n=2207). Box plots shown are defined by the three quartiles, whiskers as 1.5 IQR from the upper and lower quartile and outliers greater than the whiskers are highlighted independently. The median number of amino acid exchanges increases from 1 through 3, resulting in most variants having multiple mutations at the end of the evolution.

A visually interpretable representation of the high dimensional sequence space can be achieved with dimensionality reduction tools, such as t-distributed stochastic neighbour embedding (tSNE, Maaten & Hinton, 2008). Such an approach, where similar sequences are presented in a lower dimensional space (e.g. 2D) based on their distance in the fully high-dimensional sequence space, are often called sequence similarity networks (SSN) in biological context (Atkinson *et al*, 2009). As similar sequences are expected to cluster around their starting points, lineages and phylogeny can be identified. In case of this experimental directed evolution, a phylogeny is identifiable as variants start to cluster and evidently form lineages in the later rounds of the experiment (Fig. D-10a). Interestingly, 'founder variants' of these clusters can be identified, in this context defined as the shared set of mutations that all variants in a cluster have in common (Fig. D-10b). Branches of a phylogeny can be distinguished even after just three rounds of directed evolution, e.g. as seen in the formation of sub-lineages by the acquisition of E323V or D308V to the previous founder variant A64E R102S.

The identified founder variants are likely to contain beneficial mutations, resulting in more frequent selection in the microfluidic assay and the production of more related sequences in the following rounds of evolution. To verify this hypothesis and prove the biological relevance inferred from the SSN, all founder variants shown in Fig. D-10b were assayed for their activity. Additionally, as the information on co-occurring mutations could inform potentially important

epistatic interactions, all individual single and possible combinations of mutations in the founder variants were tested (Fig. D-11). Activity was tested as the initial reaction slope of cell lysate in 96-well plates, as these reaction conditions are a close representation of the conditions in the single cell lysate microfluidic assay that led to the selection of these variants.

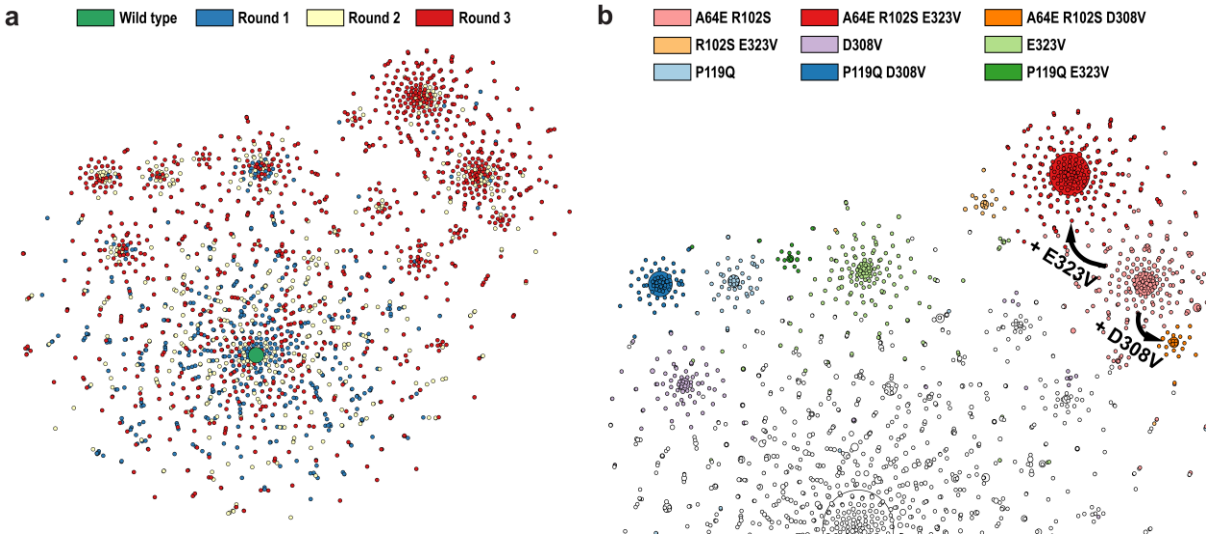


Fig. D-10: Phylogenetic analysis of directed evolution. (a) *Full sequence similarity network with variants coloured by their round of first appearance.* A graphical representation of sequence space in form of a tSNE illustrates the experimental phylogeny of three round of directed evolution. Each spot represents a unique variant with multiple mutations resulting in clustering of similar sequences. Non-mutated starting point sequence (wild type, green) highlighted with an increased spot size. **(b)** *Cluster analysis.* The core set of mutations shared between all variants within one cluster was identified and termed 'founder variant'. A part of the SSN is shown with clusters coloured by their founding variant and spot sizes corresponding to total variant sequence count. The formation of clusters is uniquely defined by the long-read information of multiple simultaneously occurring mutations.

Individual variants selected for testing were re-created: Mutations were introduced to the wild-type sequence via mutagenic primers in a whole plasmid PCR. This was necessary as individual variants are not retained after microfluidic sorting, but a pool of selected variants is collected (and subjected to sequencing). The protocol used for this was *in vivo* assembly (IVA), which generates point mutated variants in a very quick and straightforward manner (García-Nafría *et al*, 2016). If a variant with many mutations is needed for testing, however, IVA cloning reaches its limits. In this case, a UMI dial-out PCR can generate any sequenced variant from the UMI-tagged sequencing pool simply by amplifying DNA with a UMI specific primer (Schwartz *et al*, 2012).

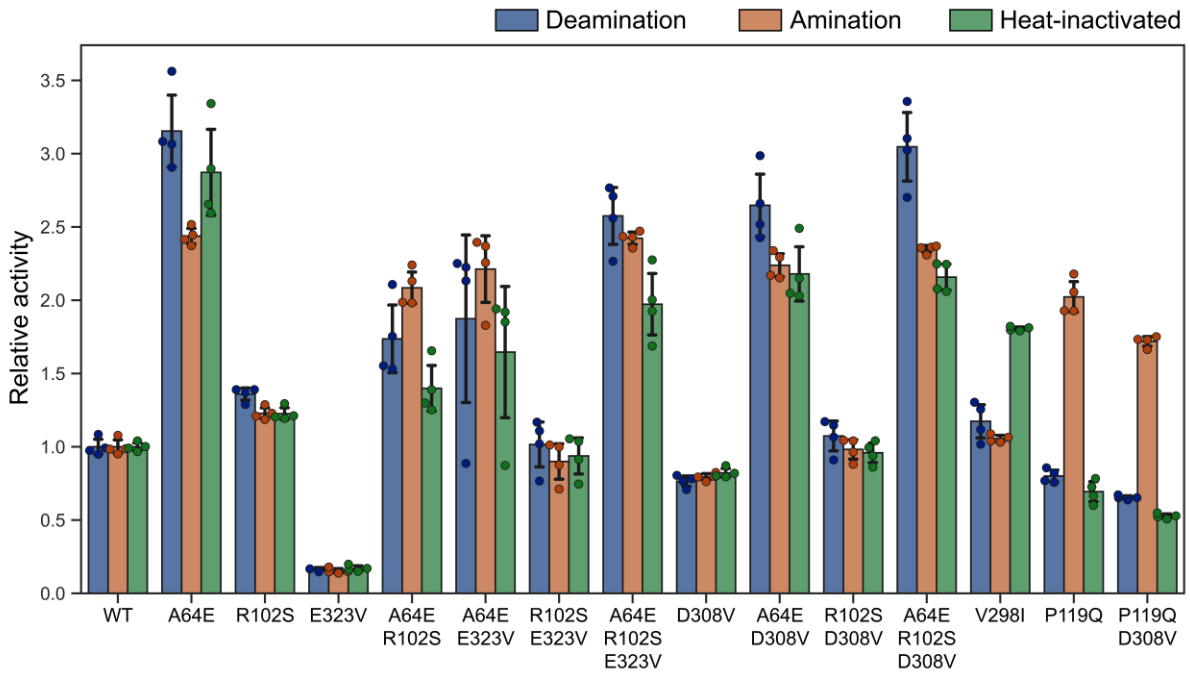


Fig. D-11: Lysate activity of founder variants relative to non-mutated reference. Identified founder variants (and contributing mutations individually) were created and tested in lysate activity for their initial reaction rate v_0. Mean value of n=4 independent biological replicates with standard deviations are shown. Reaction was tested in both reaction directions (Fig. C-2) and after heat-inactivation (10 min at 50°C) at 5 mM substrate concentration (deamination: *R*-1-methyl-3-phenylpropylamine. Amination: 4-phenyl-2-butanone). Sign epistasis can be identified in founder variant A64E R102S E323V: When E323V is introduced to previous variant A64E R102S, it provides a beneficial effect (121%-187% activity, 95% confidence interval, p = .012). However, when E323V is introduced into the wild-type background, its effect is strongly deleterious (14.6%-16.8% activity, 95% confidence interval, p = .012). Normal distribution was assumed and an ordinary one-way ANOVA ($F(6,21)$ = 42.6, p < .0001) with post-hoc multiple comparison via Tukey's test was performed.

Interestingly, an example of sign epistasis is found in the founder variant A64E R102S E323V (Fig. D-12a). In the phylogeny, the acquisition of mutation E323V to the lineage with A64E R102S as founding variant becomes apparent (Fig. D-10b), generating a large new lineage of A64E R102S E323V containing variants. The mutation E323V has a beneficial effect on variant A64E R102S, increasing its lysate deamination activity to 154±33% (95% CI), thus leading to its selection and propagation. When the mutation E323V is introduced to the wild-type sequence, however, its effect is strongly deleterious. The mutation E323V individually decreases activity to 15.7±1.1% (95% CI) of the non-mutated starting point. The basis of interaction of these three mutations is not obvious, as they are located far apart in the protein structure (at least 18 Å between the three positions, Fig. D-12b). More importantly, however, this interaction would have been overlooked by conventional hotspot analysis. Hotspot analysis is based on short-read sequencing and will thus often look at each positions individually, without the context of other mutations, as seen in the comparable study of glycosidase function

with microfluidic deep mutational scanning (Romero *et al*, 2015). While hotspot analysis is useful to show the distribution of beneficial mutations throughout the full protein (Fig. D-12b), again highlighting the importance of distal sites to protein function and engineering, as seen before (Wrenbeck *et al*, 2017a; Wilding *et al*, 2019), it cannot report on the compatibility of mutation. Conventional hotspot analysis would have predicted *all* high frequency mutations as beneficial, yet E323V is only beneficial in the context of founder variant A64E R102S.

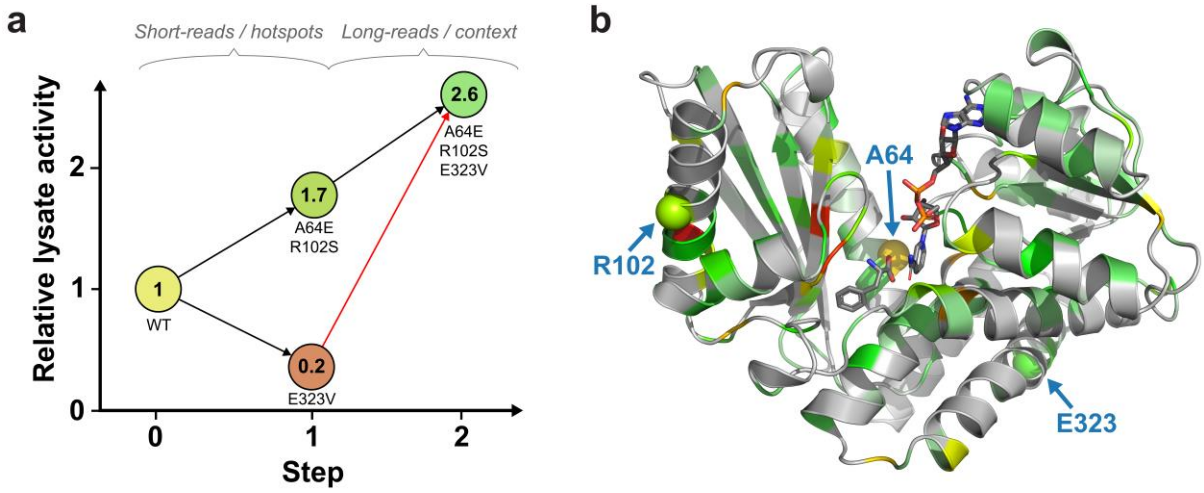


Fig. D-12: Sign epistasis in founder variant A64E R102S E323V. (a) *Lysate activities.* Initial rates of deamination activity in lysate are shown (see also Fig. D-11). When mutation E323V is introduced in the context of previous founder variant A64E R102S it is beneficial. When it is introduced to the non-mutated wild-type sequence, however, it is strongly deleterious. With conventional short-read sequencing, E323V would have been misclassified as a beneficial mutation, although it is only beneficial in the sequence context of variant A64E R102S. **(b)** *Hotspot analysis.* Enrichment of mutations per position from round 1 to round 2 of directed evolution colour coded from green to yellow to red are mapped onto the AmDH structure (PDB 1C1D). Positions with enriched mutations are found throughout the whole protein. Positions exhibiting sign epistatic interaction are highlighted.

The identification of random co-occurring mutations throughout a full protein can only be uncovered with long and accurate sequencing reads, highlighted by sign epistasis extending across 780 b in the gene and 18 Å in the protein in the most active and prevalent lineage identified here.

D.2.5. Characterisation of Epistasis in Founder Variants

The sign epistatic interaction in founder variant A64E R102S E323V was further investigated. All mutations individually and their possible combinations were purified and characterized regarding their catalytic parameters (Table 6). Additionally, performance as biocatalysts was measured as total turnover number in the industrially relevant amination direction.

Table 6: Biochemical characterization of A64E R102S E323V variants.

Variant	Steady state kinetics[a]			T_m[b] (°C)	Soluble expression[c] (%)	Total turnover number[d]
	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (s^{-1} mM^{-1})			
WT	1.55 ± 0.02	0.40 ± 0.02	3.84 ± 0.06	50.5 ± 0.17	97 ± 0.25	20,100 ± 1,180
A64E	6.12 ± 0.05	2.17 ± 0.06	2.83 ± 0.03	50.4 ± 0.08	99 ± 0.40	57,400 ± 730
R102S	1.72 ± 0.03	0.33 ± 0.03	5.27 ± 0.08	49.7 ± 0.09	96 ± 0.11	22,000 ± 330
E323V	1.20 ± 0.01	0.37 ± 0.02	3.28 ± 0.05	51.0 ± 0.09	79 ± 3.34	16,400 ± 700
A64E R102S	5.38 ± 0.07	1.86 ± 0.07	2.90 ± 0.04	48.1 ± 0.08	97 ± 0.22	51,500 ± 720
A64E E323V	4.94 ± 0.09	1.62 ± 0.09	3.05 ± 0.06	49.9 ± 0.05	94 ± 0.47	66,300 ± 2,990
R102S E323V	1.36 ± 0.02	0.38 ± 0.02	3.62 ± 0.05	49.9 ± 0.07	98 ± 0.18	16,200 ± 2,030
A64E R102S E323V	4.23 ± 0.07	1.27 ± 0.07	3.34 ± 0.06	48.9 ± 0.09	98 ± 1.45	65,900 ± 380

All values show standard deviation from three independent technical replicates.
[a]Buffer: 100 mM Glycine-KOH pH 10. Co-substrate: 2.5 mM NAD^+. Temperature 22 °C. Substrate: *R*-1-methyl-3-phenyl-propylamine from 0 mM to 12.8 mM. Michaelis-Menten curves in Fig. S4.
[b]Measured via differential scanning fluorimetry with SYPRO orange.
[c]Protein expressed at 20 °C for 16 h. Soluble fraction compared to insoluble fraction via SDS-PAGE gel densitometry.
[d]Small scale transformations with 50 mM 4-phenyl-2-butanone and 0.1 µM enzyme at 30 °C for 72 h. Glucose dehydrogenase used for cofactor recycling.

Mutation A64E contributes to a strongly altered catalytic profile, increasing both the K_M and k_{cat} in steady state kinetics. While the catalytic efficiency is thus not increased over the non-mutated wild-type variant, the more than 4-fold increased k_{cat} results in a very good biocatalyst, seen in a high total turnover number. The increases in both k_{cat} and K_M indicates a strong selection for fast substrate turnover at high concentrations, which is in line with the conditions in the droplet assay and thus follows the principle of directed evolution "*you get what you screen for*" (You & Arnold, 1996). The droplet assay, which requires high substrate turnovers for detection (Gielen *et al*, 2016), puts selection pressure on high k_{cat} values due to long incubation times at high substrate concentrations. This selection regime, however, makes the absorbance sorter a very suitable tool for the engineering of biocatalysts, which are usually required to work at high rates and high substrate concentrations for long times (Ma *et al*, 2010).

Furthermore, mutation A64E increases soluble expression strength, which is an important factor in lysate assays such as the droplet assay used for selection and the plate assay used for verification. High expression strength increases the amount of available protein in the compartment and thus increases the product detection in the compounded measurements in lysate. The effect of mutation on expression strength also forms the basis of the epistatic interaction between the mutations: The mutation E323V individually strongly reduces activity in lysate (Fig. D-11), which is a result of a reduction in expression. This reduction in expression strength, however, is compensated when E323V is combined with A64E, in which case the

negative effects of E323V are removed. The combined change in catalytic parameters, a reduction of K_M by E323V and R102S and an increase in k_{cat} by A64E, results in a rescued initial rate in lysate as well as increased performance as a biocatalyst for the triple mutant (Table 6).

The triple mutant outperforms not only the WT but also the A64E single mutant in the biocatalytic assay, showing the applicability of phylogenetic analysis and consideration of epistatic interactions in protein engineering to solve an applied biocatalytic challenge. The study of such interactions will benefit from a suitably accurate and inexpensive high-throughput sequencing strategy, as introduced in this chapter, opening the door to understanding of fundamental constrains in protein evolution (Sarkisyan *et al*, 2016; Diss & Lehner, 2018) and eventually training more and more efficient models for protein engineering (Alley *et al*, 2019; Yang *et al*, 2019).

D.3. Conclusion & Outlook

Current deep mutational scanning relies on Illumina short-read sequencing as a functional read-out of small focused or comprehensive single position libraries (see A.4). The short read length limitation is reflected in larger studies of pairwise epistatic interactions in proteins, which tend to focus on very small domains: blocks of 25 amino acid in the second RRM domain of Pab1 (Melamed *et al*, 2013), 32 amino acids at the interface of Fos and Jun (Diss & Lehner, 2018), 37 amino acids of the WW domain of Yap1 (Araya *et al*, 2012) or 56 amino acids of the IgG-binding domain of protein G (Olson *et al*, 2014). Capturing effects of mutations throughout entire proteins would be necessary to reveal the presumably more complex and difficult-to-understand long-range interactions. The biggest such experiment, the study of multiple simultaneously and randomly distributed mutations in a whole protein, was the study of the 238 amino acid fitness landscape of GFP (Sarkisyan *et al*, 2016). Here, the read length of Illumina sequencing was extended by combining two pairs of reads with the same UMI after restriction and ligation to cover the full GFP gene. The impact of mutation on GFP function could thus be studied over the full ~700 b gene range. However, this method scales badly to longer gene lengths as multiple restriction and ligation steps would be required, amplifying the experimental difficulty and any potential imperfections by chimera formation in the ligation step or by mutation to any of the many necessary restriction sites.

UMIC-seq removes these ambiguities, a simple primary long read does not require assembly but only comparison to the reference gene for consensus generation. Different levels sequencing depth can easily be adjusted and always generating full length reads results in great scalability over possible template lengths. True long reads result in uniform sequencing coverage and the same workflow for any the target length. Practically, UMIC-seq is easy to implement: Only a benchtop sequencer and common molecular biology equipment is required. The workflow can be performed by a single operator in short time, requiring only straightforward steps such as PCR and Gibson assembly with intuitive control over UMI diversity in colony count after transformation.

The sequencing workflow presented here affords pristine sequencing accuracy of amplicons of any length at reasonably low cost. Assuming a single MinION flow cell produces 10 Gb of sequencing output, approximately 100,000 accurate consensus sequences at 50-fold coverage can be produced of an amplicon of 2 kb length. This results in a cost per mutant sequence of less than 1 penny[4], a 500-fold cost reduction compared ~£5 for comparable paired Sanger sequences. Other applications that could benefit from a long-read amplicon sequencing workflow could be the sequencing of immune repertoires, viral populations or metagenomic 16S rRNA samples. However, part of the high accuracy of UMIC-seq can be expected to stem from a close and known reference sequence to which mutations are called. For the above-mentioned applications involving unknown references or larger sequence deviation between variants, a reference sequence could be assembled from raw reads prior to polishing with tools such as Miniasm (Li, 2016), but final accuracy remains to be determined and will likely require higher sequencing depth than 50x. Consensus accuracy could be increased, on the other hand, if UMI design was adjusted. Sequenced UMIs showed an AT and homopolymer bias (Fig. D-8), which is hypothesized to contribute to the remaining consensus sequencing errors by facilitating incorrect clusters. This issue could be overcome by (i) choosing a higher quality oligo synthesis method and (ii) applying a non-continuous UMI design, such as regular constant bases interspaced the random UMI bases or using NNNYR-repeats (Karst *et al*, 2021).

Finally, an affordable and accurate high-throughput long-read amplicon sequencing workflow will enable many more studies of sequence-function relationships in proteins. The increased

[4] At the time of writing, one R9.4 flow cell is sold for £720 and a library preparation kit (SQK-LSK109, containing 6 reactions) is sold for £480 (£80 per reaction) by Oxford Nanopore Technologies. The NEBNext companion module (E7180S, 24 reactions) is sold for £890 (~£37 per reaction) by New England Biolabs. Thus, the total cost per sequencing run is ~£837. Prices accessed on 11/12/2020 at https://store.nanoporetech.com/ and https://www.neb.uk.com/.

information content afforded by long reads will push phylogeny-based approaches and enable much more advanced machine-learning frameworks (Mazurenko *et al*, 2020), making use of true cooperative fitness effects instead of merely adding hotspots in protein engineering. In the next step, a larger dataset pushing to the limits of both the microfluidic assay and the nanopore sequencing capacity should be generated. Both technologies still hold untapped potential which will be used for a comparative study of enzyme evolvability in the next chapter.

E. Escaping Local Optima in Comparative Long-Term Evolution

With an accurate high-throughput sequencing method and a suitable ultrahigh-throughput screening assay at hand, the trajectories of two large directed evolution campaigns will be investigated. The two similarly stabilized AmDH starting points, $P4_{QM}$ and $E4_{QM}$, will be evolved for higher activity in parallel, prompting the comparison of their evolvabilities. Curiously, one starting point achieves higher relative and absolute improvements, expanding explored sequence space by means of recombination.

E.1. INTRODUCTION

Protein engineering is a major force of the modern green chemistry revolution (Woodley, 2020; Bornscheuer *et al*, 2012). Unfortunately, general strategies and best principles for the efficient engineering of proteins are difficult to establish due to the idiosyncrasy of each example: functional effects of mutation are strongly dependent on the current genetic context (Miton & Tokuriki, 2016). Each engineering starting point has a different potential for adaptation – a different 'evolvability' – meaning that improvements might be more or less difficult. These differences stem from pervasive cryptic epistasis (the many unexpected and non-additive interactions between mutations during protein evolution) that force evolutionary trajectories into different paths and thus towards different optima (Starr & Thornton, 2016). Evolvability manifests itself as the number of mutations necessary to reach an optimization plateau or a fitness peak as well as the achieved improvement or the height of the fitness peak (Goldsmith & Tawfik, 2017).

Consequently, very different evolvabilities for seemingly innocuous differences between homologous starting points can be identified. For example, very large effect differences of 50- to 770-fold were found for the first single step towards increasing promiscuous activity in nine homologous γ-glutamyl phosphate reductases, with no correlation to the initial promiscuous activity (Khanal *et al*, 2015). However, the full evolutionary potential might only come to fruition after multiple rounds of accumulating mutations lead to a functional plateau or fitness peak (Bloom & Arnold, 2009). Following a single trajectory of 10 rounds of directed evolution of two homologous metallo-β-lactamases for promiscuous phosphonate monoester hydrolase activity, Baier *et al* show striking differences in evolvability also for long-term evolution (Baier *et al*, 2019). Again, no correlation of initial activity with final activity as well as vastly different evolvabilities between the two homologous starting points were found. In this case, differential evolvability is related back to specific epistasis hindering the acquisition of a strongly activating mutation in one of the two starting points, convincingly showing the importance and unpredictability of cryptic genetic variation implicated in evolvability and thus protein engineering (Baier *et al*, 2019).

A possible way to increase the chances of successful directed evolution is neutral drift. If no direct optimization is possible, neutral mutations can be introduced to produce a pool of variation (Bershtein *et al*, 2008). If selection is then applied to this pool, chances of success are improved as a wider range of sequence space is covered and one of the many starting points in

the pool could lead to high function (Fig. E-1). This principle has for example been shown to be effective for the engineering of fluorescent proteins (Zheng *et al*, 2019).

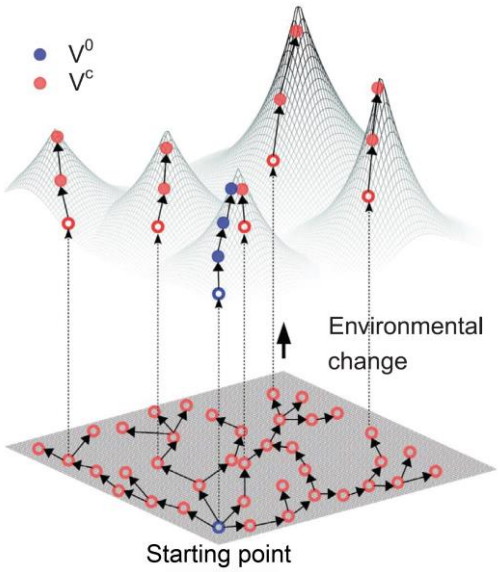


Fig. E-1: Neutral variation can increase evolvability. By diversifying the initial starting point (blue open circle) to a pool of variants with neutral mutations (red open circles), a lot of diverse adaptive peaks might become accessible. If selection is performed on just the initial starting point, a local optimum is found (V^0). By starting the selection with the pool of diversified variants, cryptic epistatic effects could lead directed evolution towards different optima, which might reach higher function (V^C). Figure adapted from Zheng *et al*, 2019.

With the introduction of the UMIC-seq protocol and an ultrahigh-throughput microfluidic AmDH screen in the previous chapters, full-length trajectories of pools of randomly mutated genes can be tracked during directed evolution for multiple rounds. This technology could generalize the anecdotal findings of previously short (Khanal *et al*, 2015) or single comparative trajectories (Baier *et al*, 2019) and expand the knowledge gained from neutral drift experiment on single starting points (Zheng *et al*, 2019) by testing its reproducibility and overlap comparatively. Evolving two similarly stabilized AmDH starting points, $P4_{QM}$ and $E4_{QM}$, in a large comparative long-term evolution would lead to a contrast of evolvabilities focused on cryptic genetic variation, while permitting more generalized conclusions due to the tracking of many trajectories in parallel, providing many evolutionary possibilities.

E.2. Results & Discussion

The methods established in chapter D, the microfluidic screening combined with accurate long-read sequencing of AmDH variants, were successfully used to map the trajectories of a short directed evolution campaign. This directed evolution test case, however, did not reach the full potential of either technology. Now, two AmDHs will be evolved for increased function in parallel at a larger scale.

E.2.1. Long-Term Microfluidic Directed Evolution of Two AmDHs

Two initial AmDH variants, the stabilized variants $P4_{QM}$ and $E4_{QM}$, will be subjected to directed evolution in parallel with the aim of comparing their evolutionary trajectories. First, new plasmids with different antibiotic resistance genes were generated for both variants. The $pASK_{amp}$ plasmid used throughout this thesis contains an ampicillin resistance gene, which was replaced with a chloramphenicol resistance cassette and a spectinomycin resistance cassette to yield $pASK_{ca}$ and $pASK_{sm}$, respectively. The resulting expression plasmids $pASK_{ca}$-$P4_{QM}$ and $pASK_{sm}$-$E4_{QM}$ were subsequently used for the comparative directed evolution campaigns.

The long-term directed evolution was performed for eight rounds as follows: (i) introduction of 2-3 random mutations throughout the full gene with error-prone PCR (ii) absorbance-activated droplet sorting selecting the most active 2% of 250,000 variants, thus sorting 5000 droplets from 1,000,000 droplets at 25% occupancy ($\lambda = 0.25$) and (iii) recovery of plasmid DNA from sorted droplets and preparation as input for the next cycle. Dilution series were plated to verify transformation efficiencies so that library coverage and recovery efficiencies stayed comparable between the two starting points. Sorting success was verified by comparing the lysate activities of few randomly selected variants before and after the first selection (Fig. E-2). Randomly selected variants from the error-prone PCR libraries before sorting show a large percentage of inactive variants (54% and 82% for $P4_{QM}$ and $E4_{QM}$, respectively) as well as a low mean relative activity compared to the non-mutated reference (0.3 and 0.1 for $P4_{QM}$ and $E4_{QM}$, respectively). After the selection via absorbance-activated droplet sorting, only few variants remain inactive (15% and 36% for $P4_{QM}$ and $E4_{QM}$, respectively) and the average activity is much closer to the non-mutated reference (0.7 for both $P4_{QM}$ and $E4_{QM}$, respectively). Interestingly, some sorted variants of the $E4_{QM}$ library show increased activity, reaching up to 3-fold relative lysate activity, while sorted variants from the $P4_{QM}$ library are largely wild-type-like (Fig. E-2).

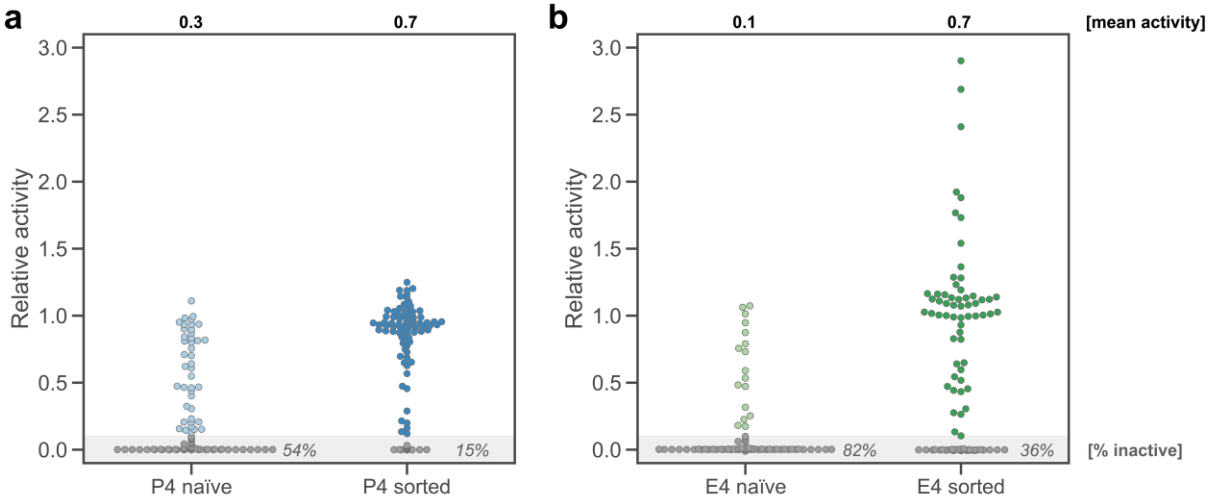


Fig. E-2: Relative lysate activities of randomly selected library variants before and after the first screening. 92 variants were randomly picked from the initial error-prone PCR library before ("naïve") and after the first screening ("sorted") and assayed for their lysate deamination activity (5 mM *R*-1-methyl-3-phenylpropylamine, 2.5 mM NAD^+ in 100 mM Glycine-KOH pH 10). Activities were calculated from the initial slopes relative to the respective non-mutated reference. The percentage of inactive variants (relative activity < 0.1) and the mean activity of all variants is shown. **(a)** *Tested variants of the $P4_{QM}$ directed evolution.* **(b)** *Tested variants of the $E4_{QM}$ directed evolution.*

Tracking the progress of the directed evolution campaigns on a functional level was performed similarly: 92 randomly selected variants were assayed after each of the eight rounds of sorting (Fig. E-3). Curiously, the $P4_{QM}$ directed evolution follows a neutral progression with the mean relative activity of functional variants fluctuating between 0.5 and 1.5 throughout the eight rounds. The directed evolution of $E4_{QM}$, on the other hand, assumes a more traditional progression with the mean relative activity of functional variants increasing from 1.1 to 17.5. These two trajectories of functional improvement show a great difference in evolvability between the two starting points.

No significant (Mann-Whitney $U = 1669$, two-tailed $p > .05$) differences in activity were measured, comparing the randomly selected variants after the first round (median 0.89) and the eighth round (median 0.91) of directed evolution for the $P4_{QM}$ campaign, while the $E4_{QM}$ variants show significant (Mann-Whitney $U = 68$, two-tailed $p < .0001$) differences in activity after round one (median 1.07) and eight (median 16.60). The striking differences in observed evolvability warrants a further investigation of the directed evolution on a sequence level.

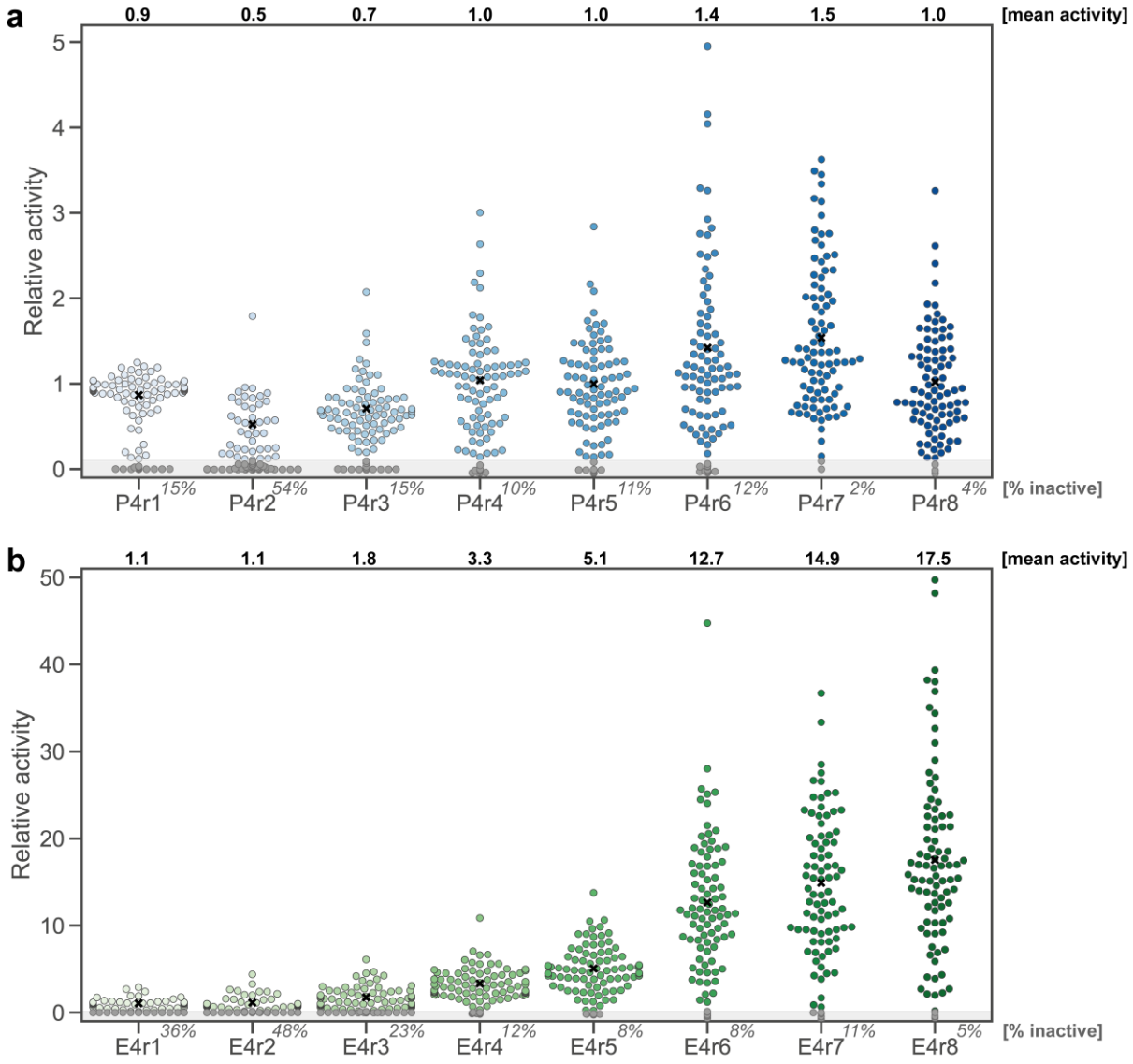


Fig. E-3: Relative lysate activities of randomly selected variants after each round of directed evolution. After each round of directed evolution (r1 to r8), 92 sorted variants were randomly selected and assayed for their lysate deamination activity (5 mM *R*-1-methyl-3-phenylpropylamine, 2.5 mM NAD^+ in 100 mM Glycine-KOH pH 10). Activities were calculated relative to the respective non-mutated reference. Additional values shown represent the percentage of inactive variants (relative activity < 0.1) and the mean activity of all active variants (relative activity > 0.1). **(a)** *Tested variants of the* $P4_{QM}$ *directed evolution.* **(b)** *Tested variants of the* $E4_{QM}$ *directed evolution.*

E.2.2. UMIC-seq Reveals the AmDH Sequence Space

Targeting the retrieval of sequence information for all ~5000 selected variants per round of directed evolution for each of the two starting points requires a lot of sequencing data. Each variant will be tagged with a UMI, conferring unique variant identity, and a barcode identifying the round of directed evolution the variant occurred in. The resulting tagged sequencing library is transformed to reduce UMI complexity according to the UMIC-seq workflow (see chapter D.2.1). A 3-fold colony transformation oversampling as well as a 100-fold read sequencing

depth is required to achieve high completeness and accuracy. Consequently, each of the two directed evolution campaigns will require 15.6 Gb[5] of sequencing data and thus be sequenced on an individual MinION R9.4 flow cell, which should be able to provide 10-30 Gb per sequencing run. The UMIC-seq workflow was performed just as the previous time with one exception: due to the observed AT-homopolymer bias in the UMIs (Fig. D-8) the UMI-tagging is performed with the UMI2-F primer containing $[NNNYR]_{10}$-repeats instead of a $[N]_{50}$ UMI, which breaks potential homopolymers to a maximal length of 5 bases (Karst *et al*, 2021).

Nanopore sequencing was performed for 48 h yielding 23.76 Gb and 18.96 Gb for the $E4_{QM}$ and $P4_{QM}$ runs, respectively. After quality control (filtering for reads with length of 1250-1400 b and a quality score greater than 8) and demultiplexing by round of directed evolution, 1,418,506 ± 169,080 and 1,029,109 ± 64,749 reads were obtained for each round of evolution in the $P4_{QM}$ and $E4_{QM}$ campaign, respectively. The number of UMI clusters containing more than 30 reads ranged from 2826 to 13,813 (mean 9,066) and from 4,690 to 9,669 (mean 6,530), resulting in a total of 72,531 and 52,238 consensus sequences for the $P4_{QM}$ and $E4_{QM}$ campaigns, respectively.

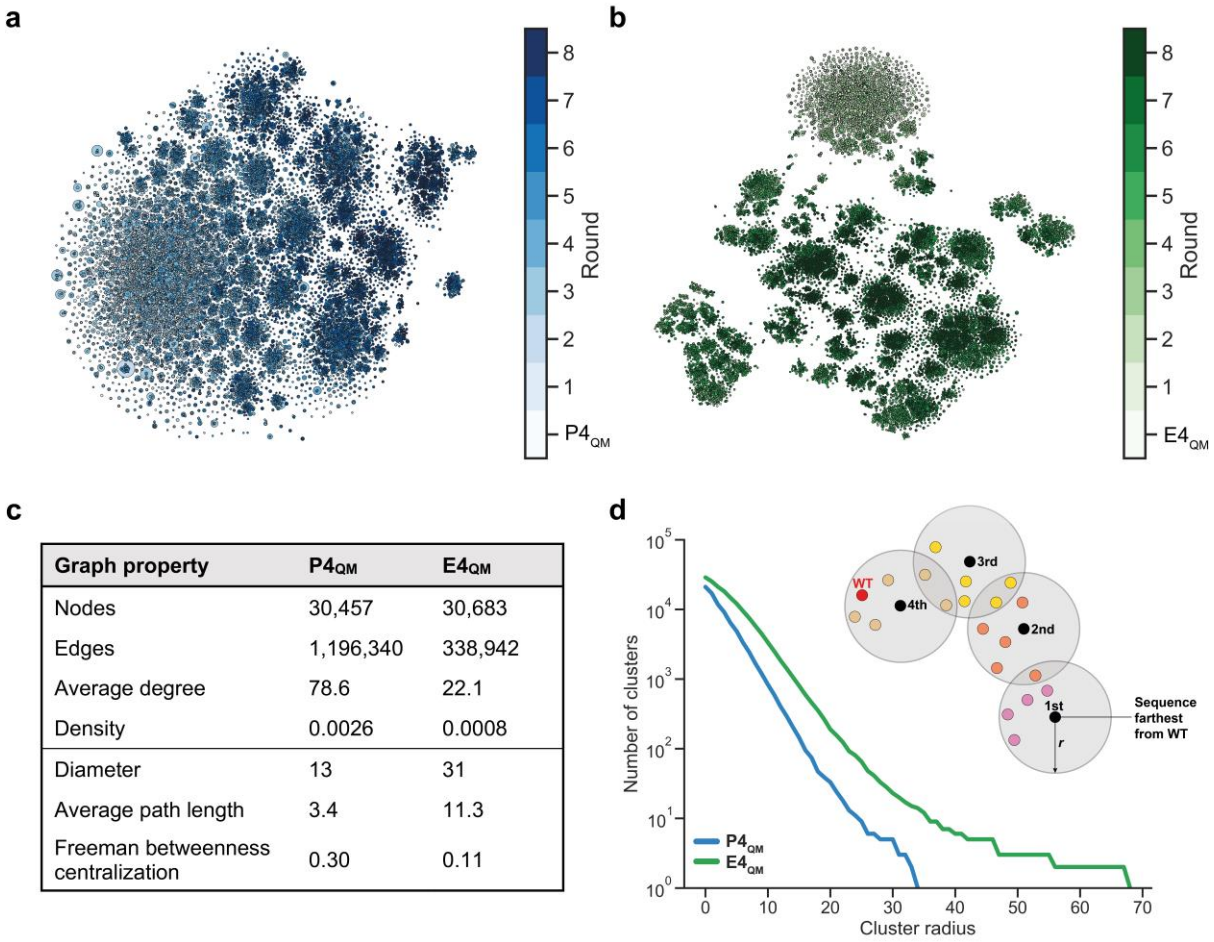
Unique variants, their corresponding total frequency (a variant can appear multiple times per round as well as in multiple rounds) and the round in which the variant first appeared were extracted. With this data, an overview of the phylogenetic architecture of the explored sequence space was established in sequence similarity networks (Fig. E-4a,b). Again, a structure in the directed evolution can be seen with many smaller and larger clusters forming with increasing progression. Interestingly, the explored sequence space in the directed evolution of $P4_{QM}$ appears well connected with large clusters forming in close proximity. The directed evolution of $E4_{QM}$ on the other hand shows more distant and less connected clusters. While the tSNE representations remain useful visualizations of the explored sequence space during directed evolution, their interpretation can be misleading for example in regard to absolute cluster sizes or distances between clusters, which might not be well preserved (Maaten & Hinton, 2008; Wattenberg *et al*, 2016). Consequently, more quantitative measures of the sequence space architecture are needed. In fact, most studies using sequence similarity networks have only used them for visualization and exploration (chapter D.2.4; Copp *et al*, 2018; Starr *et al*, 2017; Atkinson *et al*, 2009; Akiva *et al*, 2017). However, mathematical graph theory provides

[5] 3-fold transformation oversampling × 5000 variants × 8 rounds × 1300 b gene length × 100-fold read oversampling for accuracy = 15.6 Gb

quantitative statistical tools for the analysis of large networks, e.g. previously applied to the networks intrinsic to antibody repertoires (Miho *et al*, 2019).

Thus, network graphs were generated in which each node represents a unique variant and edges connect nodes within an edit distance of two or lower (two variants are connected in the graph if they can be interchanged by one or two mutations). Consequently, the resulting graphs contain 30,457 and 30,683 nodes with 1,196,340 and 338,942 edges for the $P4_{QM}$ and $E4_{QM}$ evolution, respectively. For visualization purposes, a simplified sub-graph containing only edges with an edit distance of one and only nodes with at least two connections (the 2-core) is shown in Fig. S5. Again, a densely populated $P4_{QM}$ sequence space is visually distinguishable from the disconnected wide $E4_{QM}$ sequence space, confirming the tSNE representation. More quantitatively, however, each node has multiple imputable properties such as its degree and betweenness. The degree of a node corresponds to the number of edges connecting to it and the betweenness of a node is inferred from its participation in all possible shortest paths between any two nodes. These two properties were computed for all nodes as they could potentially aid in the identification of important founder mutations or bridging variants between clusters, respectively. Key mutations as identified by their statistical properties in the network graph were chosen for functional verification, as their impact is likely beneficial to the protein (chapter E.2.3).

To assess the overall differences in the sequence space architecture for the two directed evolution campaigns, global network properties were calculated (Fig. E-4c). First, a comparison of the average degree (78.6 for $P4_{QM}$ evolution, 22.1 for $E4_{QM}$ evolution) shows that the density of the $P4_{QM}$ sequence space is in fact very high, while the $E4_{QM}$ sequence space is more loosely connected. These findings are underlined by the differences in network diameter (13 and 31 for $P4_{QM}$ and $E4_{QM}$, respectively) and average path length (3.4 and 11.3 for $P4_{QM}$ and $E4_{QM}$, respectively), quantifying a wide and distributed $E4_{QM}$ sequence space in contrast to a centralized and densely connected $P4_{QM}$ sequence space. Quantifying the level of sequence diversity explored by the two evolution campaigns was done by calculating the number of clusters the variants can be grouped into at different edit (Levenshtein) distances (Fig. E-4b, Byrant *et al*, 2021). The $E4_{QM}$ evolution consistently covers a larger volume of sequence space than the $P4_{QM}$ campaign. The cluster number of the $E4_{QM}$ sequence space increases with the cluster radius from 1.4- to 14-times the cluster number of $P4_{QM}$, being on average 5.3±2.4-times higher. This higher diversity in the explored sequence space of $E4_{QM}$ is consistent with the wider network graphs.



Graph property	$P4_{QM}$	$E4_{QM}$
Nodes	30,457	30,683
Edges	1,196,340	338,942
Average degree	78.6	22.1
Density	0.0026	0.0008
Diameter	13	31
Average path length	3.4	11.3
Freeman betweenness centralization	0.30	0.11

Fig. E-4: Visualization and global properties of the $P4_{QM}$ and $E4_{QM}$ directed evolution sequence space. A tSNE (for visualization) and a network graph (for quantification) was calculated for unique variants of both directed evolution campaigns. Similar sequences are placed close together in the tSNE, resulting in clusters representing phylogenetic clades. While clusters in the $P4_{QM}$ evolution seem closer and more connected, the clusters in the $E4_{QM}$ evolution show a more separate and wide distribution. Colour coding indicates the round of first appearance of each unique variant. **(a)** *tSNE of $P4_{QM}$ evolution sequence space.* 30,457 unique variants. **(b)** *tSNE of $E4_{QM}$ evolution sequence space.* 30,683 unique variants. **(c)** *Graph properties.* The information of sequenced variants (30,457 and 30,683 unique sequences for $P4_{QM}$ and $E4_{QM}$, respectively) was translated to a network graph, in which each unique variant constitutes a node which is connected to others if the variant is interchangeable by one or two mutations. Properties of this network were calculated: The average degree is the average numbers of connections of each node while the network density is the ratio of edges to all theoretically possible edges. The network diameter is the longest of all shortest paths between any two nodes while the average path length is the average of all shortest paths between any two nodes. The freeman centralization measures how central the most central node is compared to all other nodes, where in this case centralization is measured as node betweenness, which calculates the participation of a node in all possible shortest paths. The calculation of these properties provides a quantitative view on the differences of the two directed evolution campaigns, in which the $P4_{QM}$ evolution explores a dense and narrow sequence space and $E4_{QM}$ explores a wide and less-connected sequence space. **(d)** *Diversity analysis.* The number of clusters as a function of cluster radius is shown for the $P4_{QM}$ and $E4_{QM}$ evolution. To enable comparison, 40,000 sequences were randomly sampled for both evolution campaigns. The insert shows the method of clustering which was implemented as described by Bryant *et al*, 2021. Clusters are generated from the furthest sequence to the wild type consecutively including all sequences within an edit (Levenshtein) distance less than the radius. The $E4_{QM}$ sequence space consistently produces more clusters than the $P4_{QM}$ sequence space, speaking to a higher level of explored sequence diversity.

For the first time, measurable differences in the explored sequence space architecture are quantified here in the case of the directed evolution of two homologous enzymes.

E.2.3. Characterization and Transplantation of Key Mutations

Key mutations in the directed evolution of both starting points (founder variants) were functionally tested. Variants central to big clusters show a high degree in the network, meaning that those variants are connected to many related sequences and thus constitute a founding variant of a cluster. Ten mutations were selected in each evolution based on their degree and betweenness values, making them likely founder variants (Table S2). These ten mutations were re-introduced to the non-mutated parental sequence individually via *in vivo* assembly (Materials & Methods G.2) and tested for their functional impact on lysate deamination activity (Fig. E-5). Again, most founder variants are beneficial, explaining their successful enrichment and the foundation of a lineage in the sequence similarity networks. On average, the impact the founder mutations have on relative activity is higher for the $E4_{QM}$ variants than for the $P4_{QM}$ variants (average relative activity 1.6±0.4 and 1.2±0.6, respectively). The directed evolution of $P4_{QM}$ remained neutral overall (Fig. E-3), which is in line with the finding that most founder variants also have a neutral effect on activity. $E4_{QM}$, on the other hand, shows a high evolvability form the first round with many variants improving the activity in lysate (Fig. E-5).

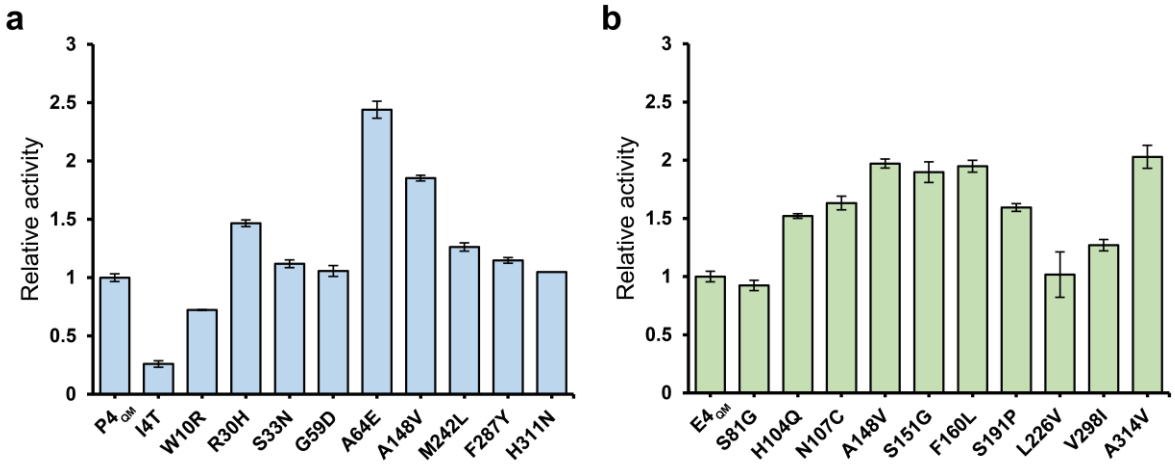


Fig. E-5: Relative lysate activity of founder variants. Initial rates of lysate deamination activities were measured in biological replicates (n=4, SD shown) for ten selected founder variants for both starting points. Activities are relative to the respective non-mutated parental variant. **(a)** *Founder variants of $P4_{QM}$ directed evolution.* Previously identified founder variant A64E is identified again. **(b)** *Founder variants of $E4_{QM}$ directed evolution.* The average impact of founder variants is higher than that for the $P4_{QM}$ founder variants.

The most beneficial founder variant in the $P4_{QM}$ evolution is the mutation A64E, reproducing the findings of the short directed evolution in the previous chapter (Fig. D-11). In the $E4_{QM}$

evolution the beneficial mutation S151G is identified, a mutation that was found to increase the activity of the non-stabilized WT_{QC} AmDH in a previous study (Ye *et al*, 2015). Together, these results hint at the reproducibility of finding individual beneficial mutations even in large library sizes when an ultrahigh-throughput screening is employed.

Only one mutation (A148V) is shared between the two starting points in the ten tested founder variants, which also leads to similar improvements in both starting points (1.85±0.02-fold and 1.97±0.04-fold improvement in $P4_{QM}$ and $E4_{QM}$, respectively). Thus, the question whether the founder mutations differ by stochasticity or by functional exclusiveness arises. Although the throughput of the microfluidic screen should suffice, in theory, to assay every possible individual point mutation more than 30-fold[6] at each round of directed evolution, confounding factors such as true library composition make predictions of the real oversampling difficult. A founder mutation of one starting point might still not have occurred and thus not have been screened in the other starting point. Consequently, beneficial founder variants were tested for their compatibility in the other starting point experimentally (Fig. E-6).

Beneficial $P4_{QM}$ founder mutations R30H and A64E were chosen for transplantation as well as beneficial $E4_{QM}$ founder mutations H104Q, N107C, S151G and F160L. The mutations were introduced individually to the respective other background and again tested for their relative lysate activity (Fig. E-6). Curiously, all four founder mutations that were beneficial in the $E4_{QM}$ starting point were neutral or slightly deleterious in the $P4_{QM}$ background. Beneficial $P4_{QM}$ mutations were either neutral to $E4_{QM}$ in case of A64E or highly deleterious in case of R30H.

These results show that the different trajectories and founder variants between the two lineages are not a result of differential sampling of mutations but rather the result of verifiable differences in function. Beneficial mutations in one starting point are not beneficial in the other, although both variants are similarly stable (reducing the importance of threshold epistasis) and similar in sequence composition (86% sequence identity). This prevalence of context-dependent mutations speaks to the pervasiveness of specific epistatic interactions and the difficulty of predicting the effect of individual mutations let alone the evolvability of a starting point.

[6] The AmDH consists of 357 amino acids, each of which would be mutated to the other 19 amino acids equally in a perfectly balanced library. Thus, a complete single mutation library would comprise 6783 variants, which could be oversampled 37-fold with a throughput of 250,000 variants assayed per round of directed evolution.

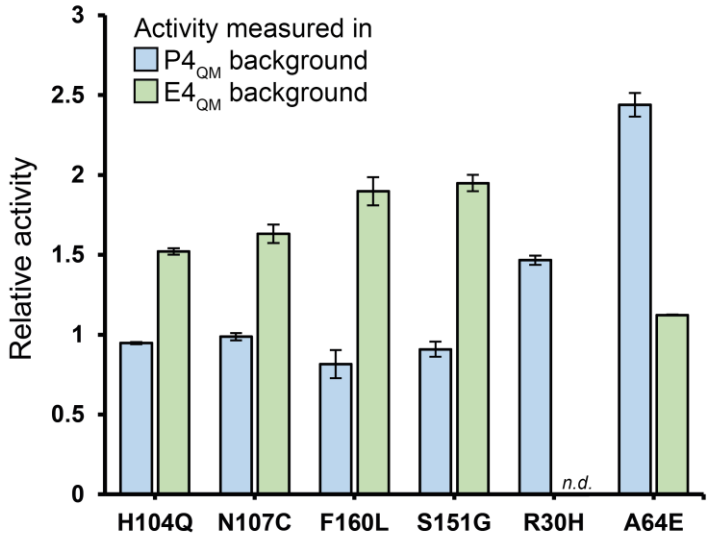


Fig. E-6: Activity of beneficial founder mutations measured in both $P4_{QM}$ and $E4_{QM}$ backgrounds. The initial slope of lysate deamination activity was measured in biological replicates (n = 4, SD shown) and normalized to the respective non-mutated background. Mutations show a highly differential effect depending on the genetic context they are introduced into, even in the similarly stable homologs $P4_{QM}$ and $E4_{QM}$. Activity of R30H was below detection limit in the $E4_{QM}$ background.

E.2.4. End-Point Variants

To narrow down the search space for improved variants a final stringent round of directed evolution was performed. Again, an error-prone PCR was performed to introduce 2-3 mutations per gene on the ~5000 selected variants of round 8. Now, a more stringent selection sorting the top 0.5% of 250,000 variants followed. Additionally, the sort was performed at a lower droplet occupancy ($\lambda = 0.15$) to reduce the false positive rate, assaying 250,000 variants in 1.7 million droplets. A traditional secondary screen was performed to identify the overall most active variants: individual colonies recovered after sorting were analysed in a 96-well plate lysate assay, in total screening 552 variants for each campaign. The five variants with the highest activity of both campaigns were isolated, their sequence identified, and their lysate activity verified in quadruplicates. The variant with the highest relative lysate activity (clone $P4_{QM}$-G8: 5.0±0.2, clone $E4_{QM}$-D1: 27.5±0.5) was chosen for purification and further characterization (Table 7). The $E4_{QM}$-D1 variant not only shows the highest relative improvements in activity, but also the highest absolute lysate activity. Again, the final variants obtained from directed evolution show great increases in k_{cat}, in agreement with the selection conditions. $E4_{QM}$-D1 outperforms not only the non-mutated $E4_{QM}$ variant but also the $P4_{QM}$-G8 variant considerably, its k_{cat} being 27.9- and 2.4-fold increased, respectively. A similar trend

transpires through the catalytic efficiencies, making these improved AmDHs serious candidates for biocatalysts.

Analysing the Sanger sequencing data obtained from the five most improved variants for each directed evolution provides an interesting view of the mechanism with which activity was improved. Curiously, conversion mutations can be identified in all five final variants for both campaigns. The most improved $P4_{QM}$-G8 variant contains a total of 11 mutations: 7 unique mutations (I4T A148V E198D L215Q M242L F287Y H311N) and 4 conversion mutations (A21R A22E V28I Q202R), which convert its state partially to the $E4_{QM}$ sequence state. Three of the best five $P4_{QM}$ variants show a similar pattern of combining founder mutations with the conversion mutations around A21R A22E. While the overall best variant $P4_{QM}$-G8 does not contain the most active founder mutation A64E, one of the best five variants has the mutations D5H N9D R30H A64E V298I H311N, indicating the possibility of non-additive interactions between the founder mutations as well as overall distinct solutions for activity improvement and multiple fitness peaks.

Table 7: Characterization of the most active $P4_{QM}$ and $E4_{QM}$ variants.

Variant	Steady state kinetics[a]			T_m[b] (°C)	Soluble expression[c] (%)	Lysate activity[d] (abs_{360nm}/h)
	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (s^{-1} mM^{-1})			
$P4_{QM}$	0.71 ± 0.02	0.56 ± 0.06	1.26 ± 0.11	49.8 ± 0.1	97.8 ± 4.1	56.7 ± 4.2
$E4_{QM}$	0.65 ± 0.01	0.59 ± 0.02	1.10 ± 0.04	51.3 ± 0.2	96.4 ± 2.0	19.0 ± 0.4
$P4_{QM}$-G8	7.71 ± 0.19	1.15 ± 0.10	6.71 ± 0.09	53.7 ± 0.1	100.0 ± 0.1	285.1 ± 11.8
$E4_{QM}$-D1	18.12 ± 0.21	2.09 ± 0.07	8.68 ± 0.04	54.3 ± 0.2	99.5 ± 0.1	522.5 ± 8.8

All values show standard deviation from three independent technical replicates, except lysate activity which is measured in biological independent quadruplicates.
[a]Buffer: 100 mM Glycine-KOH pH 10. Co-substrate: 2.5 mM NAD^+. Temperature 22 °C. Substrate: *R*-1-methyl-3-phenyl-propylamine from 0 mM to 12.8 mM. Michaelis-Menten curves in Fig. S6.
[b]Measured via differential scanning fluorimetry with SYPRO orange.
[c]Protein expressed at 20 °C for 16 h. Soluble fraction compared to insoluble fraction via SDS-PAGE gel densitometry.
[d]Lysate activity measured for deamination activity (5 mM *R*-1-methyl-3-phenylpropylamine, 2.5 mM NAD^+ in 100 mM Glycine-KOH pH 10).

The most active $E4_{QM}$ variant, $E4_{QM}$-D1, contains a total of 20 mutations: 6 unique mutations (H104Q N107C A148V V298I E337K A339S) and 14 conversion mutations (S124N S194K M195L A196L T248D S257K A260C I267L D269H S273A N330K I341E G344E A347M), which convert its state partially to the $P4_{QM}$ sequence state. The unique mutations, again, fix many of the activating founder mutations in different combinations between the five most active $E4_{QM}$ variants. Curiously, all highly improved $E4_{QM}$ variants show a partial conversion

to the $P4_{QM}$ sequence state, with conversion mutation count ranging from 14 to 19 in the five best $E4_{QM}$ variants, at an average of 15 mutations.

$P4_{QM}$ and $E4_{QM}$ differ by a total of 50 mutations, which raises the question of potential recombination between the two directed evolution campaigns, as the most improved variants for both campaigns show a high number of converting mutations. An acquisition of ~30% of mutations converting to the other state (i.e. 15 of 50), additionally given that many of those mutations are achieved by swapping a complete 3 b codon. Variant $E4_{QM}$-D1, for example, requires a total of 30 base pair mutations for its 14 amino acid conversion mutations, which is unlikely to naturally occur in an error-prone PCR library and thus warrants the investigation of recombination frequency during the parallel directed evolution.

E.2.5. Recombination Escapes Local Optima

To identify possible recombination between the two starting points, the sequencing dataset was analysed by counting the number of conversion mutations (mutations that would convert one starting point to the other, 50 in total) in each variant. This information is overlain on the tSNE sequence similarity networks (Fig. E-7a,b) and shows a cluster with many conversion mutations appearing isolated from the rest of the sequence space of $E4_{QM}$. Variants with less conversion mutations are more tightly integrated and form part of many clusters towards the later rounds (as seen by the final variants having less conversion mutations as well as the colour-coding in Fig. E-4). To identify which conversion mutations are enriched, the total count of each conversion mutation in all rounds for each evolution campaign was calculated. In the $P4_{QM}$ evolution this count shows conversion mutations mainly effecting the start of the protein (amino acids 6 to 32, Fig. E-7c). For $E4_{QM}$, contrastingly, all possible conversion mutations are detectable with a striking increase in count for the mutations in the latter half of the protein (amino acids 194 to 347, Fig. E-7d). This preference for conversion mutations in the second half of the protein supports the hypothesis of a contamination of the $P4_{QM}$ starting point into the $E4_{QM}$ evolution, which then quickly recombines with the $E4_{QM}$ starting point to yield a chimeric protein. The first half of the chimera is formed by $E4_{QM}$ and the second half by $P4_{QM}$, crossing over between positions 157 and 194. During following rounds of directed evolution of the chimeric protein, some conversion mutations such as the stretch from positions 215 to 227 are again removed, returning partly to the $E4_{QM}$ state.

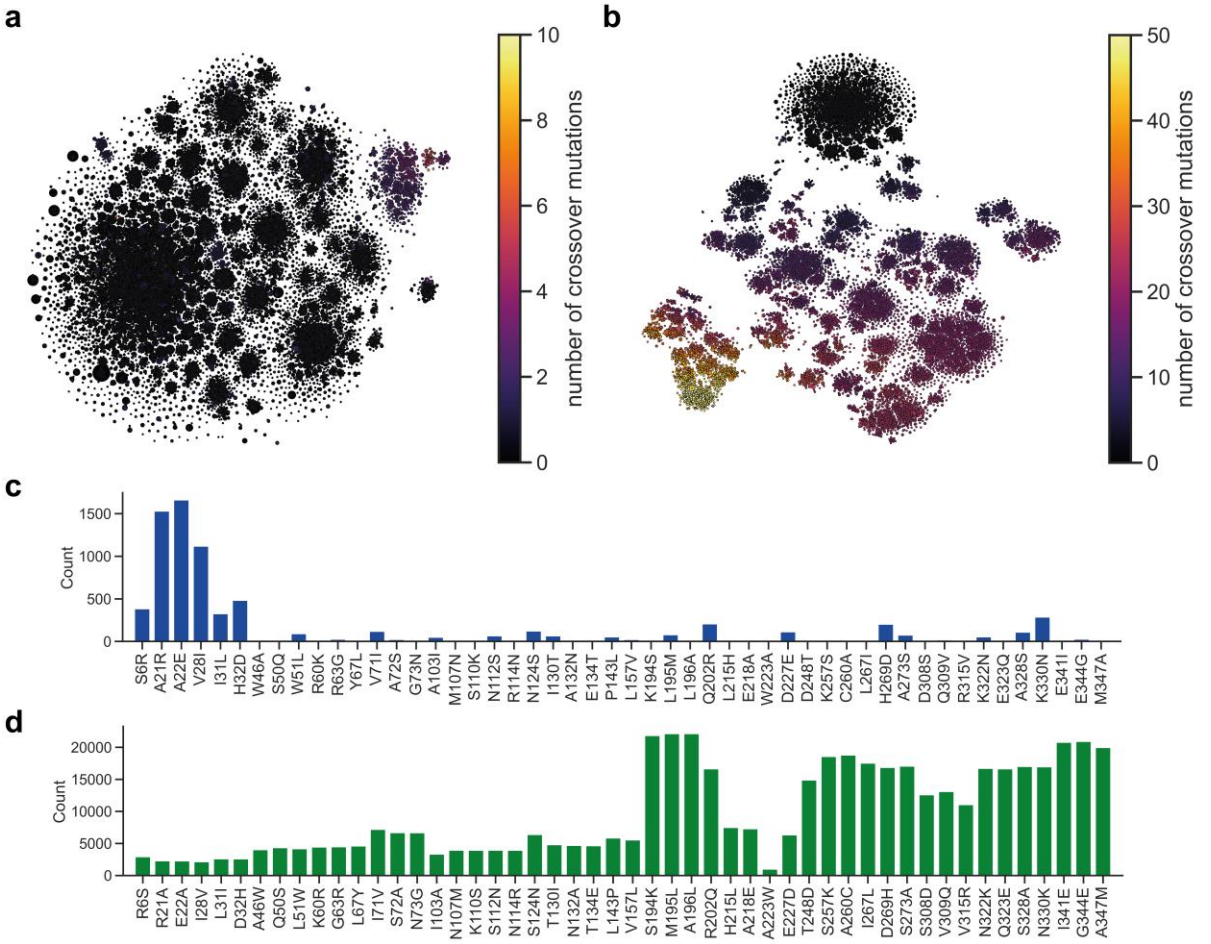


Fig. E-7: Conversion mutations in the directed evolution of P4$_{QM}$ and E4$_{QM}$. A tSNE representation of the explored sequence space was calculated for unique variants of both directed evolution campaigns in Fig. E-4. Here, it is colour coded to indicate the number of conversion mutations, mutations which convert one starting point to the other, in each unique variant. Furthermore, the total number of occurrences of each conversion mutation was counted. **(a)** *Sequence space of P4$_{QM}$ evolution.* 30,457 unique variants. **(b)** *Sequence space of E4$_{QM}$ evolution.* 30,683 unique variants. **(c)** *Conversion mutation frequency in P4$_{QM}$ evolution.* Conversion mutations accumulate at the N-terminus of the protein. **(d)** *Conversion mutation frequency in E4$_{QM}$ evolution.* While all conversion mutations are detectable in the E4$_{QM}$ evolution, a clear preference towards the C-terminal half of the protein is observable.

A similar recombination is likely to have occurred later in the P4$_{QM}$ directed evolution, where a chimeric protein with N-terminal E4$_{QM}$ parts appears in a detached cluster (Fig. E-7a). Again, a recombination is more likely than the accumulation of individual substitutions through error-prone PCR, as full codons are mutated e.g. using 3 base mutations (GCA to CGT) to fulfil the A21R amino acid conversion mutation of the P4$_{QM}$-G8 variant. Consecutive mutations are statistically much less likely to occur than non-consecutive mutations in random libraries such as those generated by error-prone PCR, with the probability of exactly mutating the 3 bases of a codon instead of producing three dispersed mutations in a 357-codon gene being

approximately 1 in 500,000[7]. Conversely, even minute contamination in a PCR reaction can lead to recombination, an effect known to be due to template switching of the polymerase (Potapov & Ong, 2017; Pääbo *et al*, 1990).

Further evidence supporting the hypothesis of recombination is found by analysing the frequency of all conversion mutations in each selected round (Fig. E-8). The frequency of the conversion mutations in the latter half of $E4_{QM}$ (S194K to A347M) increase dramatically during the directed evolution from round 2 to round 4, as highlighted by the lighter trajectories in Fig. E-8a. A slight contamination of the $P4_{QM}$ starting point into the $E4_{QM}$ evolution could have recombined, joining the first half of $E4_{QM}$ and the second half of $P4_{QM}$ between positions 157 and 194. This chimeric version might have been much more active or evolvable than either original starting point, leading to its quick invasion of the $E4_{QM}$ evolution. The variant quickly propagates through the population before some of the more deleterious conversion mutations are removed (dashed curves of S308D, V309Q, V315R, H215L, A218E and E227D in Fig. E-8a). A second block of conversion mutations (dashed curves of I71V S72A N73G, Fig. E-8a) propagates through the population slowly on a different scale. Interestingly, a similar event can be seen for the evolution of $P4_{QM}$, where a short block of mutations at the N-terminus of the protein (A21R A22E, Fig. E-8b) starts getting fixed quickly in the later rounds of the directed evolution campaign.

As characterized in the previous chapter and verified on the sequence level in this chapter, the best final variants of both directed evolution campaigns were a product of recombination. However, it is still unclear whether the initial recombined variant improves activity to an extent that leads to it becoming quickly fixed in the population, or if its evolvability is higher than the original starting points, leading to the quick accumulation of mutations and the propagation through many subsequent lineages.

[7] 357 codons are constructed from 1071 base pairs. While there are $\frac{n!}{r!\,(n-r)!}$ possible combinations in which three mutations can be introduced to the gene, here equalling to 204,173,655 combinations, only 357 form codons. The probability of generating a completely mutated codon thus is 357 / 204,173,655 = 1.75×10^{-6} for a random library containing all three mutation variants.

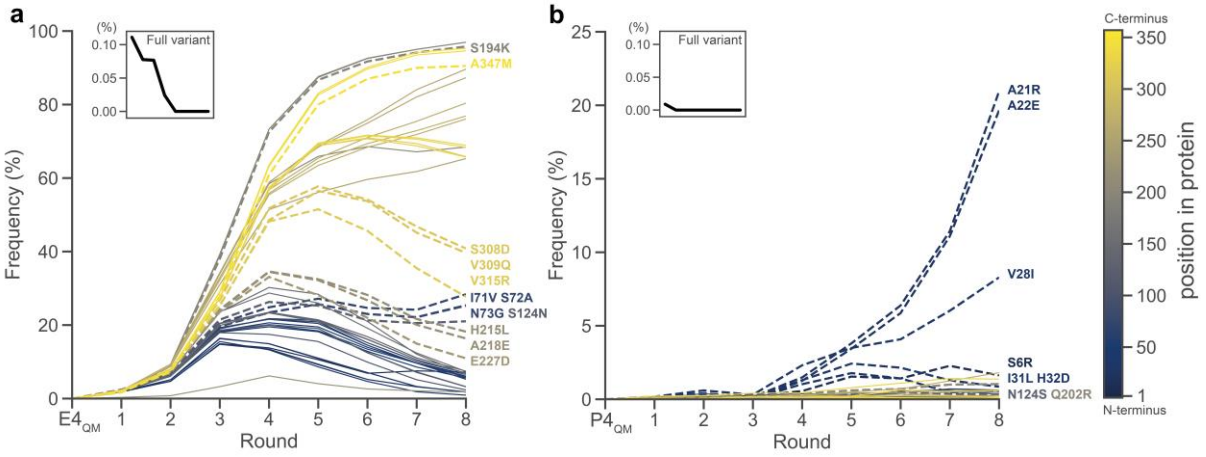


Fig. E-8: Frequency of conversion mutations in each round of directed evolution. The frequency of each of the 50 mutations converting one starting point into the other were calculated for each round of directed evolution. A frequency of 100% would indicate each variant containing that mutation, whereas at 0% no variant contained that mutation in the given round of directed evolution. Mutations are colour-coded according to their position in the protein. An insert additionally shows the frequency of the contaminant starting point per round. **(a)** *Conversions identified in the E4$_{QM}$ evolution.* All conversion mutations are present at low frequency from the beginning. The contaminant sequence (P4$_{QM}$) can be detected at ~0.1% frequency in round one of the E4$_{QM}$ evolution, but it is quickly removed from the population (see insert). Few variants, namely most mutations between S194K and A347M (dashed) are strongly enriched from round to round until reaching near full adoption. Some mutations in the latter half of the gene are initially co-enriched, but are phased out in later rounds, such as the cluster around mutations S308D V309Q V315R and H215L A218E E227D. Conversion mutations at the beginning of the gene (I71V S72A N73G) show a slower dynamic of adoption. **(b)** *Conversions identified in the P4$_{QM}$ evolution.* Conversion mutations at the beginning of the gene, especially A21R and A22E (dashed) strongly increase in frequency in the later rounds of directed evolution. Conversion mutations at other positions in the gene, such as N124S and Q202R, show a slow yet steady adoption.

E.2.6. Characterization of Chimeric Variants

Four chimeric variants were recreated, as schematized in Fig. E-9: (i) The initial recombination in the E4$_{QM}$ evolution is likely to have happened between mutations V157L and S194K, as a clear split in frequencies indicates (Fig. E-7d and Fig. E-8a). A chimera is thus generated that combines the beginning of the E4$_{QM}$ variant and the end of the P4$_{QM}$ variant, split before the S194K mutation. Each gene half is amplified via PCR and assembled in the pASK$_{amp}$ plasmid via Golden Gate assembly to yield a construct termed pASK$_{amp}$-Full-Chimera. (ii) The inverse chimera, first half P4$_{QM}$ and second half E4$_{QM}$, is created equivalently and called Inverse-Chimera. It should provide a meaningful test of the biological relevance of the orientation and localisation of the recombination crossover. (iii) After the recombination, some conversion mutations are removed from the Full-Chimera variant (Fig. E-7d and Fig. E-8a). Consequently, a recombination variant with a more selected set of conversion mutations is generated by creating a variant containing only the most frequent conversion mutations (S194K M195L A196L S257K A260C I341E G344E A347M) with the conversion mutations that accumulate

in later rounds (I71V S72A N73G S124N). This variant is generated via *in vivo* assembly and termed Selected-Chimera. (iv) The conversion mutations strongly enriched in the $P4_{QM}$ evolution, A21R A22E, as well as late addition Q202R, were combined via *in vivo* assembly to a small conversion variant termed SmallP4-Chimera.

All four chimeric variants were purified and characterized alongside the non-mutated variants (Table 8). The Full-Chimera, likely to have recombined initially, provides only a small improvement to enzyme activity. It thus has a decent chance of being selected in the assay, but more strikingly bears a high evolutionary potential. When the non-beneficial conversion mutations are removed again from the full chimera and the conversion mutations selected in the later rounds are added, the resulting variant (Selected-Chimera) readily improves in catalytic performance. From here, only few mutations are necessary to yield moderate further improvements in the best final variant, $E4_{QM}$-D1. Consequently, the recombination event in the directed evolution of $E4_{QM}$ opened a completely new region of sequence space for exploration, providing high evolvability to subsequent variants. The recombined variant escapes the local optima of the $E4_{QM}$ starting point, leading to an easy access to overall greatly improved final variants. The inversed chimera, in contrast, is non-functional highlighting the biological relevance of the chosen set of mutations.

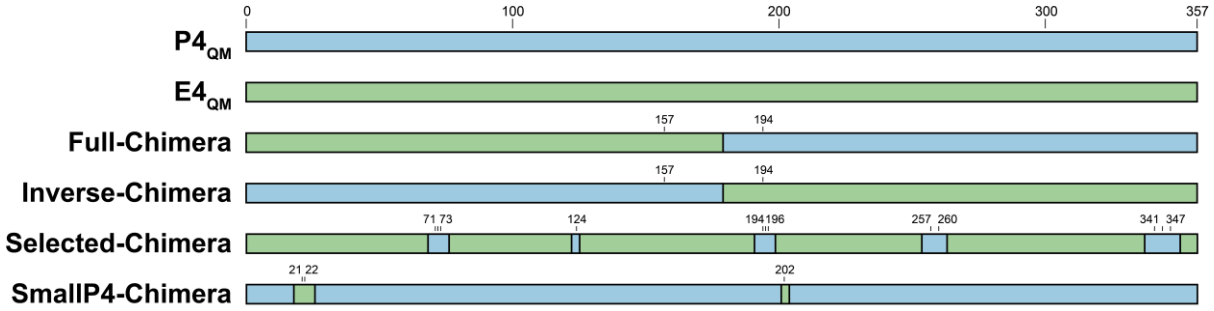


Fig. E-9: Overview of $P4_{QM}$-$E4_{QM}$-chimeras generated for characterization. The Full-Chimera and Inverse-Chimera are split between the conversion mutations at position 157 and 194, with the first half being formed by $E4_{QM}$ and the second half being formed by $P4_{QM}$ for the Full-Chimera. The inverse chimera has this orientation reversed. The Selected-Chimera is based on $E4_{QM}$ with only the highly enriched $P4_{QM}$ conversion mutations in the second half (S194K M195L A196L S257K A260C I341E G344E A347M) as well as conversion mutations enriched during the later rounds (I71V S72A N73G S124N). The SmallP4-Chimera is based on $P4_{QM}$ with the addition of enriched $E4_{QM}$ conversion mutations (A21R A22E and Q202R).

In the directed evolution of $P4_{QM}$, a recombination event likely occurred later in the campaign. Consequently, $P4_{QM}$ variants could not improve in activity as much as the $E4_{QM}$ variants, as reflected in the lysate activities of variants sampled in each round (Fig. E-3). Nonetheless, a recombination event likely yielded variants with the mutations A21R A22E that turned into

very active variants with merely three mutations (SmallP4-Chimera). Interestingly, the mutations R21A E22A were introduced to the WT variant by the PROSS algorithm as stabilizing mutations when the P4 starting point was generated (chapter B.2.1). These mutations clearly were false positives, as their removal is beneficial to the enzyme and quickly fixed in the population during directed evolution. Other conversion mutations showing enrichment during the $P4_{QM}$ evolution (V28I, N124S, Q202R), on the other hand, are mutations that $E4_{QM}$ acquired during stabilization (chapter B.2.2). The acquisition of those mutations thus speaks to the reproducibility of directed evolution with ultrahigh-throughput screening and, to some extent, the additivity found in stabilizing mutations.

Table 8: Characterization of recombined variants.

Variant	Steady state kinetics[a]			T_m[b] (°C)	Soluble content[c] (%)	Lysate activity[d] (abs_{360nm}/h)
	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (s^{-1} mM^{-1})			
$P4_{QM}$	0.71 ± 0.02	0.56 ± 0.06	1.26 ± 0.11	49.8 ± 0.1	97.8 ± 4.1	56.7 ± 4.2
$E4_{QM}$	0.65 ± 0.01	0.59 ± 0.02	1.10 ± 0.04	51.3 ± 0.2	96.4 ± 2.0	19.0 ± 0.4
Full-Chimera	1.15 ± 0.01	0.61 ± 0.02	1.89 ± 0.04	51.9 ± 0.2	99.5 ± 0.2	47.2 ± 3.6
Inverse-Chimera	0.05 ± 0.01	0.54 ± 0.11	0.09 ± 0.21	*n.d.*	52.6 ± 11.7	15.1 ± 1.2
Selected-Chimera	2.62 ± 0.12	0.50 ± 0.09	5.29 ± 0.19	52.4 ± 0.4	99.4 ± 0.6	117.1 ± 4.7
SmallP4-Chimera	2.00 ± 0.04	0.49 ± 0.04	4.10 ± 0.08	53.5 ± 0.4	99.1 ± 1.6	55.2 ± 5.6

All values show standard deviation from three independent technical replicates, except lysate activity which is measured in biological independent quadruplicates. Data was obtained together with the data of Table 7, resulting in the same values for $P4_{QM}$ and $E4_{QM}$ used here.
[a]Buffer: 100 mM Glycine-KOH pH 10. Co-substrate: 2.5 mM NAD^+. Temperature 22 °C. Substrate: *R*-1-methyl-3-phenyl-propylamine from 0 mM to 12.8 mM. Michaelis-Menten curves in Fig. S6.
[b]Measured via differential scanning fluorimetry with SYPRO orange.
[c]Protein expressed at 20 °C for 16 h. Soluble fraction compared to insoluble fraction via SDS-PAGE gel densitometry.
[d]Lysate activity measured for deamination activity (5 mM *R*-1-methyl-3-phenylpropylamine, 2.5 mM NAD^+ in 100 mM Glycine-KOH pH 10).

This serendipitous finding of recombination in both directed evolution campaigns speaks to the power it holds in enabling a vast exploration of sequence space and thus the generation of potentially highly evolvable enzymes. Local optima can be escaped by "jumping" through sequence space with recombination, acquiring a large set of mutations at once. A combinatorial library of chimeric proteins from homologous sequences is likely to cover a diverse area of sequence space and could provide very valuable starting points for directed evolution. Such an idea has been applied to the generation of new variants with diverse substrate scopes (Li *et al*, 2007) or increased thermostability (Heinzelman *et al*, 2009). The so-called SCHEMA recombination is a structure-guided recombination approach that smartly generates a library of enzyme chimeras at crossover points that are unlikely to disrupt the protein structure (Voigt *et*

al, 2002). The mixed successes of this approach might be due to recombined variants not necessarily being very active. However, they might hold high evolvability and could be used as a pool of variation that can generate powerful enzymes with few additional cycles of directed evolution.

Optimization plateaus and diminishing returns are regularly observed in the traditional directed evolution of proteins, which often proceeds by few substitution mutations at a time (Newton *et al*, 2018; Goldsmith & Tawfik, 2017). Here, the introduction of recombination with homologous enzymes might also hold the key to unlocking wider sequence space, in contrast to established shuffling methods that are generally used for the recombination of selected variants with few mutations (Stemmer, 1994; Zhao *et al*, 1998). Such an approach could again, as seen in this chapter, prove useful to escape local optima which cannot simply be escaped by accumulation of single mutations.

E.3. Conclusion & Outlook

The comparative evolution of two stabilized AmDH starting points in this chapter revealed a striking incompatibility of founder variants. Only one mutation was identified as a founder variant in both evolutions and while founder variants were beneficial in their original starting points, they were neutral or deleterious in the other. This shows that the extent of epistasis determining the outcome of evolution is high: differential mutations are acquired because of specific interactions with the genetic context they appear in. It is unlikely that differential mutations are seen because of stability effects, as both starting points are similarly stable, or that differential mutations appear because of the stochasticity in the screening process, as the throughput of droplet screening covers all single point mutations sufficiently.

However, even with the power of ultrahigh-throughput droplet microfluidic assays large improvements to enzyme activity might be precluded for variants isolated on fitness plateaus when conventional mutagenesis methods are applied. $P4_{QM}$ seems to be un-evolvable: improvements to activity in raw cell lysate, a metric similar to the selection in droplets, could not be increased more than 3.5-fold (Fig. D-11) during the three rounds of directed evolution of chapter D in which a total of 750,000 variants were screened to establish the nanopore sequencing workflow. Now, an approach borrowing ideas from neutral drift, the evolution of a large population of variants at once, could not generate much improved variants either.

Screening a total of more than 2 million $P4_{QM}$ variants over the course of nine rounds of directed evolution yielded a final variant showing 5-fold improved activity in lysate in this chapter (Table 7). Analysing the explored sequence space provides a view of a narrow and confined fitness landscape: the $P4_{QM}$ evolution is restricted to a local plateau that neutral drift, the accumulation of individual neutral substitutions, cannot bridge (Fig. E-4). $E4_{QM}$, on the other hand, is a more evolvable starting point, as mutations confer a larger benefit to it (Fig. E-6). Most of the evolvability in the $E4_{QM}$ evolution, however, can be attributed to a recombination event taking place early in the campaign, manifesting itself as a quantifiably large and diverse sequence space for the $E4_{QM}$ evolution. The initial recombination is not very beneficial to activity, however, the recombined variant quickly evolves to take over the population and form the basis of the most active lineages and variants (Fig. E-8). This shows the power of recombination introducing a large amount of neutral variation at once, thus being able to escape local optima and subsequently find novel fitness peaks.

These results show the importance of sexual recombination in evolution, a paradigm that is not used to its full potential in current approaches, which are mainly focused on few substitution mutations at once. It was previously shown that intense neutral drift could not escape the local optima in the directed evolution of RNA ligases (Petrie & Joyce, 2014). The authors concluded that adaptation, not neutral drift, is the primary driver of genetic change. However, their evolution campaign was based on substitution changes alone. While these results agree with the data observed for the $P4_{QM}$ evolution, it is also shown here that recombination can enable jumps between isolated islands of function in a sparsely populated fitness landscape, especially if they cannot be bridged by substitution errors alone. Consequently, the use of chimeric libraries might be better suited as collective starting points for directed evolution than a pool of neutral variants containing point mutations. Such libraries could be generated from a few homologous structures with a high probability of being functional using the SCHEMA approach (Voigt *et al*, 2002) and then be subjected to further directed evolution. Afterall, recombination and gene rearrangement are common ways of how natural enzyme families have evolved (Bornberg-Bauer & Albà, 2013; Lees *et al*, 2016) and the experimental mimicking of such processes in the lab has been possible (Peisajovich *et al*, 2006). Seeing their importance in the experimental evolution performed here as well as their prevalence in natural evolution, the use of recombination and gene rearrangement as tools for directed evolution should be considered and expanded.

Finally, a hypothesis of why the PROSS-stabilized AmDH $P4_{QM}$ cannot evolve by substitutions might be found in the method used for stabilization. By optimizing many of the possible interactions between amino acids in the protein to yield the most stable variant, a counterproductive overstabilization might have occurred. The introduction of 44 mutations optimized for stability introduces not only a restriction of their mutation, which would lead to a loss of stability and thus decrease of function, but also renders the mutation of residues interacting with the stabilizing mutations deleterious. The introduction of the P4 mutations to the WT in fact increases the total number of contacts in the structure by 145, as calculated by Arpeggio (Jubb *et al*, 2017). Thus, by removing accessible mutations the explorable sequence space and evolvability of the $P4_{QM}$ variant is reduced. Moreover, the enzyme could be rigidified and not be able to mould freely into new conformational substates, further reducing its evolvability (also see chapter A.1.2.2). This hypothesis is strengthened by analysis of the introduced mutations for their impact on flexibility: Using the normal mode analysis webserver DynaMut (Rodrigues *et al*, 2018) to analyse all 44 mutations that are introduced for the generation of the P4 variant and the 7 mutations for the E4 variant, a clear trend of rigidification for the PROSS mutations is observable. The total impact on vibrational entropy by the mutations introduced to form E4 is +0.10 kcal mol^{-1} K^{-1}, not altering molecule flexibility meaningfully, while the P4 mutations together are decreasing molecule flexibility (-1.37 kcal mol^{-1} K^{-1} vibrational entropy) leading to potential rigidification. A potential overstabilization could also explain the reduced activity of the P4 starting point (20% of the catalytic efficiency of the WT starting point, Fig. B-1). These effects should be further investigated with the generation and analysis of protein and variant 3D structures and the impact of PROSS-stabilization on protein flexibility should be investigated in depth with long molecular dynamics simulations, nuclear magnetic resonance studies or similar tools.

As high stability leading to rigidity might only be circumstantial and differential evolvability could be manifested by chance through specific epistatic interactions, further investigation of homologs should be conducted. This could for example be tested by investigating further homologs or different PROSS designs, such as P1 to P3, to verify if an association between overstabilization and evolvability is detectable. A binary library introducing all possible combinations of P4 mutations to E4 could be analysed via AADS to identify the optimal set of recombined mutations.

These approaches combined could narrow down on the mechanisms by which the evolvability is increased after recombination in this study. In the end, this data could inform future protein

engineering campaigns to be more efficient, diverting from the current go-to procedure of high stabilization. The most stable starting point might not be the most evolvable after all.

F. Final Conclusion and Perspective

To close, I will summarize the conclusions of the previous chapters and emphasize their applicability to protein engineering. Furthermore, I will propose perspectives of how the tools developed for this thesis can be used to challenge fundamental properties of protein evolution in future projects.

Understanding the mechanisms and constraints of protein evolution is not only of fundamental scientific interest, but it will also undoubtedly benefit the approaches applied to protein engineering. Making protein engineering more efficient will lead to the broader adoption of biocatalysts for the sustainable production of bulk chemicals and expand its foothold in fine chemical synthesis, such as the production of pharmaceuticals or fragrances, as well as simplify the engineering of biologics as therapeutics or diagnostic sensors. In this study, two technologies were developed that enable the efficient exploration of fitness landscapes in protein evolution.

First, a general strategy increasing the sensitivity of droplet microfluidic enzyme assays was created. The development of new microfluidic detection modes (Table 1) forebode improved assays for enzymatic reactions, such as the label-free screening of underrepresented enzyme classes (Table 2). However, high detection limits and large droplet volumes prevent the routine application of alternative detection modes to enzyme engineering, which requires monoclonal droplet populations, and thus imposes steep demands on enzyme total turnover numbers. A general workflow circumventing this issue is developed in chapter C. By amplifying clonal cell populations in droplets, enzyme content is increased leading to the detection of previously undetectable enzyme activities in droplets. The approach is developed with fluorescent read-outs (Fig. C-6) and shown to be applicable to one such new detection mode, absorbance-activated droplet sorting. By simply amplifying the enzyme content per compartment through monoclonal cell growth, the workflow is universally applicable to any previous lysate-based assay. This versatility should help it see speedy adoption with other detection modes, enzyme classes and reactions. Screening enzyme libraries at higher throughput directly increases the chances of engineering success by directly assaying more variants. Furthermore, ultrahigh-throughput screening provides the confidence necessary for increasingly daring library designs, which have lower rates of positives but unlock greater overall potential (see chapter A.4.1).

Second, an accurate long-read amplicon sequencing workflow meeting the requirements of protein engineering was developed. Knowing the sequence identity in conjunction with the function of many assayed variants fundamentally establishes the structure of the local fitness landscape (Sarkisyan *et al*, 2016) and can be used to train statistical models of protein function, enabling a computational exploration of sequence space for higher engineering outcomes (Ogden *et al*, 2019; Whitehead *et al*, 2012; Saito *et al*, 2018; Jenson *et al*, 2018). Long-read sequencing is necessary if the association of multiple co-occurring mutations throughout a full-length gene is required, e.g. when epistasis effects are to be considered. Many current deep

mutational scanning studies do not investigate the effects of epistasis (Table 3) although it has been shown to be deeply pervasive to protein evolution (Starr *et al*, 2018; Sarkisyan *et al*, 2016). Long-read amplicon sequencing workflows currently remain inaccurate or require the assembly of short reads, which can be error-prone (van Dijk *et al*, 2018). In the UMIC-seq workflow, inaccurate long-reads are matched to their template molecule with unique molecular identifiers to yield accurate long reads, with one flow cell providing around 100,000 consensus sequences of 1 kb. The practical simplicity and affordability of the workflow and its components will make it likely to be adopted as a quick verification of amplicon libraries in many laboratories, not only applicable to gene libraries in protein engineering but also to viral populations, immune repertoires, cancer diagnostics or metagenomics.

These two technologies are used together to sketch out the sequence space through which evolution progresses from two different starting points. It is unclear to which extent evolutionary outcomes are predetermined by the current sequence state and its epistatic restrictions. To investigate this issue, two stabilized starting points were generated and compared for their initial adoption of AmDH activity (chapter B). Both stabilized starting points accepted more mutations than the non-stabilized wild type, in agreement with mutational robustness hypothesis for stable proteins (Besenmatter *et al*, 2007), however, the evolutionarily stabilized variant E4 showed greater improvements to activity than the computationally stabilized variant P4 (Fig. B-4).

Both starting points were subjected to eight rounds of directed evolution and sequencing (chapter E). Interestingly, completely different outcomes are observed for both evolution campaigns, indicating a multi-peaked fitness landscape shaped by pervasive epistasis, in which two closely related enzymes (86% sequence identity) do not evolve towards the same solution. Furthermore, fitness plateaus cannot be escaped by sequential point mutations in the evolution of these AmDHs. For challenging evolutions in some starting points there might just not be any ‘paths that lead uphill, one step at a time’ (Tracewell & Arnold, 2009). A simple hill-climbing approach following consecutive small steps in one trajectory cannot find the optimal solution and is at best slow at producing local optima (Yang *et al*, 2019). Recombination, on the other hand, can enable jumps through sequence space and is the method by which innovation is facilitated in the AmDH evolution (Fig. E-7). Local exploration by conventional substitutions is transported to vast regions of sequence space by recombining existing parts. Consequently, the use of recombined starting point libraries could greatly benefit evolvability and thus engineering success.

The findings of this study highlight the importance of choosing a suitable starting point for directed evolution, going against the conventional approach of highly stabilizing the starting point. Computational stabilization seems to have confined the $P4_{QM}$ variant into a local optimum, not being able to further evolve AmDH activity. The fact that not only stability is crucial for high evolvability is highlighted again (chapter A.1.2). However, if the necessary assay throughput for the evolution of a pool of recombined starting points, as suggested above to increase evolvability, is unavailable, how can a suitable starting point be identified?

Recent developments in mathematical modelling have shown great ability to generalize to unseen regions of sequence space (Russ *et al*, 2020; Alley *et al*, 2019; Bryant *et al*, 2021). Deep neural networks can extract patterns of millions of sequences and thus could be able to generate some basic understanding of the epistatic state a protein is in. A recent model called UniRep, for example, was able to predict stability and function from natural sequence diversity alone (Alley *et al*, 2019). When deep learning approaches are combined with deep mutational scanning data, the engineering of a diverse set of proteins distinct from any naturally occurring variants is possible, as performed to generate viral capsids distinct form all known natural isolates (Bryant *et al*, 2021). It is not unreasonable to assume that such approaches could be used to predict the evolvability of a starting point for protein engineering in the future. While current model-guided approaches show promise in extending natural diversity, requiring some level of epistatic understanding, predicting protein evolvability or direct quantitative functional values loom on the horizon.

Large datasets are required to train these model-guided approaches and remain a bottleneck for their broad usage. Here, UMIC-seq could be seamlessly integrated in long-read deep mutational scanning. When deep mutational scanning is focused on a small region, as currently common practice (Bryant *et al*, 2021, see chapter D.3), a large part of the accessible diversity is disregarded. By using long reads for deep mutational scanning, model-driven protein engineering will have a larger training ground. Furthermore, long-read deep mutational scanning datasets could be used for the construction of 3D structures when conventional methods fail or no information on homologous sequences is available for the construction of homology models (e.g. applicable to metagenomics). Small structures have recently been accurately derived from multi-point deep mutational scanning datasets (Schmiedel & Lehner, 2019; Rollins *et al*, 2019). To scale these methods to full length proteins, mutational effects covering the full length of the gene must be identified (Chiasson & Fowler, 2019). UMIC-seq could provide such datasets for any of those use cases, be it for structure prediction, model-

guided protein engineering or the exploration of long-range epistatic effects. A naïve library could be tagged with UMIs and sequenced according to the UMIC-seq protocol, providing a clear association between long-read variant information and unique identifier. This library can then be subjected to a suitable high-throughput assay and UMI frequencies before and after selection would be read out using conventional short-read sequencing. Thus, reliable UMI enrichment factors could be related back to the original variant information, providing functional values.

Protein engineering methods are not very efficient yet and a lot of sequence space remains untapped by current approaches (chapter A.4.1). The combination of the here presented improvements to ultrahigh-throughput screening and long-read sequencing will not only inform future protein engineering campaigns but also help establish new and improved protein engineering methods themselves.

G. MATERIALS & METHODS

Here, I will list the materials and methods used in this thesis. I will cite references to published methods where applicable, but nonetheless detail published methods to provide a complete experimental description.

G.1. MATERIALS

All materials were bought from Sigma-Aldrich, unless otherwise noted. Enzymes were obtained from New England Biolabs (NEB) or Thermo Fisher. Oligonucleotides were purchased from Sigma-Aldrich. Plasmids were gifted from Raphaëlle Hours ($pASK_{amp}$-WT, $pASK_{amp}$-P4, $pASK_{amp}$-E3, $pASK_{amp}$-*empty*), ordered from Addgene (pET-iLOV #63723), purchased from Novagen (pRSFDuet-1) or constructed as described later on.

G.2. MOLECULAR BIOLOGY

General procedures. Common laboratory and maintenance procedures were performed according to manufacturer's specification and are thus only briefly summarized here. Antibiotic stocks were used where appropriate and applied to growth medium or agar plates at final concentrations of 100 µg/mL ampicillin, 33 µg/mL chloramphenicol, 50 µg/mL kanamycin or 50 µg/mL spectinomycin. Inducer stock was used at 1000x concentration (200 µg/mL aTc), prepared in ethanol and stored at -20 °C. PCR was performed using the Q5 master mix (NEB M0492) according to the manufacturer's recommendation. DNA quantification was performed with a nano spectrophotometer (NanoDrop 2000, Thermo Fisher). Transformation of *E. coli* cells was done with commercial cells. For routine low-efficiency transformation, chemically competent cells were used, whereas libraries and recovery experiments were done with high efficiency via electroporation. For DNA amplification and maintenance, as well as sparingly for protein expression, subcloning efficiency DH5α chemically competent cells (Invitrogen #18265017) or E. cloni 10G ELITE electrocompetent cells (Lucigen #60051) were used. Both strains are abbreviated as DH5α. For protein expression, *E. coli* BL21 (DE3) chemically competent cells (NEB C2527) were used, abbreviated as BL21. DNA purification of plasmids was performed with the GeneJET Plasmid Miniprep Kit (Thermo Fisher K0503), PCR products and assemblies were directly purified with the Zymo Clean & Conentrator-5 (Zymo Research D4004) or purified from agarose gel with the addition of Agarose Dissolving Buffer (Zymo Research D4001-1). For purification of many DNA fragments in parallel or in the sequencing library preparation, SPRI beads were used (AMPure XP, Beckman Coulter A63881).

Error-prone PCR libraries. Error-prone PCR (epPCR) libraries are generated using the GeneMorph II random mutatgenesis kit (Agilent). Conditions were adjusted to consistently yield error rates of 2-4 mutations per gene, as controlled via Sanger sequencing of randomly picked variants. Usually, the gene of interest is pre-amplified for 20 cycles using the high-fidelity Q5 polymerase (NEB), 10 ng plasmid template DNA and the outer binding IF and IR primers (Table 9). The resulting PCR product is then used as a template for epPCR. Now, 600 to 900 ng of target DNA fragment was used as a template for the low-fidelity Mutazyme II polymerase (Agilent), together with the Golden Gate cloning compatible inner primers epIN-F and epIN-R and amplified for 25 cycles. The resulting PCR product was assembled in a Golden Gate reaction with the backbone fragment, amplified from pASK-empty with primers epBB-F and epBB-R, as described below, to yield the final library.

Table 9: Primers used for the generation of error-prone PCR libraries. Features: Restriction enzyme site (green).

Name	Use	Sequence
IF	Sanger sequencing, Pre-amplification	GAGTTATTTTACCACTCCCT
IR	Sanger sequencing, Pre-amplification	CGCAGTAGCGGTAAACG
epIN-F	Golden Gate cloning, epPCR	TTTGGTCTCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACC
epIN-R	Golden Gate cloning, epPCR	TTTGGTCTCTGGCTCCAAGCGCTCTCGAG
epBB-F	Golden Gate cloning, epPCR	TTTGGTCTCGAGCCACCCGCAGTTCG
epBB-R	Golden Gate cloning, epPCR	TTTGGTCTCTCTAGATTTTTGTCGAACTATTCATTTCACTTTTC

DNA shuffling. Shuffled libraries were generated via StEP (Zhao *et al*, 1998). A PCR with BioTaq polymerase (Bioline) was run at 55 °C annealing temperature and 30 s extension time for 100 cycles. The gene variants of interest were added equally to the PCR of 50 µL to a final DNA content of 400 ng. The PCR fragment was gel excised, purified and used for plasmid construction as described below.

Plasmid construction. Plasmids were constructed by Golden Gate assembly (Engler *et al*, 2008), following recommendations specified by the supplier NEB. Usually, a single insert fragment containing a gene or variants of a gene obtained via epPCR was combined with a backbone vector fragment containing the regulatory elements, origin of replication and resistance cassette in a molar ratio of 3:1. Reactions were set up in a PCR reaction tube and processed in a thermal cycler as follows:

Reaction:		Cycling:		
100 ng	Insert	5 min	16 °C	30 cycles
100 ng	Backbone	5 min	37 °C	
1 µL	BsaI-v2	20 min	55 °C	
1 µL	T4 Ligase	5 min	80 °C	
0.5 µL	*Dpn*I			
ad 20 µL	H_2O			

Following the assembly reaction, DNA was purified and used for the transformation of electrocompetent *E. coli*. All plasmids gifted or constructed for this study are listed in Table 10.

Table 10: List of plasmids gifted or constructed for this study.

Name	From	Feature
$pASK_{amp}$-empty	Raphaëlle Hours	Empty pASK-IBA63b+
$pASK_{amp}$-WT	Raphaëlle Hours	Expression of *Rhodococcus sp. M4* PheDH
$pASK_{amp}$-P4	Raphaëlle Hours	PROSS stabilized PheDH variant
$pASK_{amp}$-E3	Raphaëlle Hours	PheDH variant stabilized in three rounds of evolution
$pASK_{amp}$-WT_{qgc}	Raphaëlle Hours	AmDH variant from Ye *et al* 2015 (K66Q S149G N262C)
$pASK_{amp}$-E4	This study	PheDH variant stabilized in four rounds of evolution
$pASK_{amp}$-WT_{QM}	This study	*Rhodococcus sp.* M4 AmDH (K66Q N262M)
$pASK_{amp}$-$P4_{QM}$	This study	PROSS stabilized AmDH (K66Q N262M)
$pASK_{amp}$-$E4_{QM}$	This study	Evolutionarily stabilized AmDH (K66Q N262M)
pAB094a	Addgene (#79241)	Template for spectinomycin cassette
BPK2101	Addgene (#65770)	Template for chloramphenicol cassette
$pASK_{ca}$-$P4_{QM}$	This study	pASK-IBA63b+ with chloramphenicol resistance
$pASK_{sm}$-$E4_{QM}$	This study	pASK-IBA63b+ with spectinomycin resistance
pRSFDuet-1	Invitrogen	Template for pLacI, subcloning of nanopore inserts
pET-iLOV	Addgene (#63723)	iLOV expression under T7 promoter control
pLacI-iLOV	This study	Weak constitutive iLOV expression
$pASK_{amp}$-iLOV	This study	Anhydrotetracycline-inducible iLOV expression
$pASK_{amp}$-FullChimera	This study	Chimeric variant of $E4_{QM}$ and $P4_{QM}$
$pASK_{amp}$-InverseChimera	This study	Chimeric variant of $P4_{QM}$ and $E4_{QM}$

Generation of point mutation variants. *In vivo* assembly (IVA cloning) was used to quickly generate variants with single and multi-point mutations (García-Nafría *et al*, 2016). Primers were designed as outlined by García-Nafría *et al*, 2016, examples used for the generation of chimeric variants are shown in Table 11. Primers containing point mutations were used for whole-plasmid PCR with 1 ng template DNA. The template DNA was removed after PCR by adding 1 µL of *Dpn*I and incubating for 30 min at 37 °C. This mixture was used to transform chemically-competent DH5α cells and the point mutation was verified by plasmid isolation and Sanger sequencing.

Table 11: Primers used for IVA cloning of chimeric variants Selected-Chimera and SmallP4-Chimera. Features: Mutated codon (red). Binding (black). Binding and homology (grey).

Name	Sequence
P4_A21R-A22E-f	GACCGTTACACGTTTTGATCGTGAAACCGGTGCACATTTTGTTATTCGTATTCATAG
P4_A21+-r	ATCAAAACGTGTAACGGTCATTTCACCATC
P4_Q202R-f	GCCGAAGCGGGTGCACGGCTGCTGGTTGCAGATACCGATACCG
P4_Q202+-r	TGCACCCGCTTCGGCCAGC
E4_I71V-S72A-N73G-f	ACCCTGCAGATGGCAGTTGCAGGTCTGCCGATGGGTGGTGGTAAAAG
E4_I71+-r	TGCCATCTGCAGGGTCATTGCG
E4_S124N-f	AGGTCCGGATGTTAATACCAATAGCGCAGATATGGATACCCTGAATGATAC
E4_S124-r	GGTATTAACATCCGGACCTGTCCAATAATTGC
E4_S194K-M195L-A196L-f	TGGTGGTTCACTGGCAAAACTGCTGGCCGAAGCGGGTGCACG
E4_S194+-r	TGCCAGTGAACCACCAACGGC
E4_S257K-A260C-f	CACGTACCCTGGATTGTAAAGTTGTTTGTGGTGCAGCAATGAATGTTATTGCAGATG
E4_S257+-r	ACAATCCAGGGTACGTGCAACTTCTG
E4_I341E-G344E-A347M-f	GATGAAGCCGCACGCGAACTGGCAGAACGTCGTATGCGTGAAGCATCAACCACCACCG
E4_I341+-r	GCGTGCGGCTTCATCCGGTG

Sanger sequencing. Sequencing to control individual constructs or verify epPCR error rates was performed at the Department of Biochemistry's sequencing facility. Analysis was performed by alignment in Benchling or with custom python scripts in the case of calculation of nanopore sequencing error rates.

G.3. Lysate Assays

Preparation of cell lysates. Protein expression for lysate activity assays was performed in 96-well plates. Cultures (300 µL LB medium per well) were inoculated from single colonies and grown overnight at 37 °C and 900 rpm. Expression cultures were inoculated by adding 25 µL of the grown overnight culture to 425 µL fresh LB medium in a new 96-well deep well plate. At the same time, the overnight culture can be used to create glycerol stocks for long-term storage of variants at -80 °C (150 µL overnight culture added to 50 µL of 50% (v/v) glycerol) or be stored at 4 °C for short term storage. Expression cultures were growth at 37 °C for 2 h followed by addition of inducer (50 µL LB containing 10x aTc (2 µg/mL)) and protein expression at 22 °C for 16 h at 900 rpm. The cultures were sedimented (20 min, 3220 × *g*, 4 °C) and the supernatant discarded. Cell pellets were frozen for at least 1 h at -80 °C to aid cell lysis. Lysis was performed by addition of 200 µL 25 mM Tris-HCl pH 8.0 containing 0.1% (v/v) Triton X-100, 100 µg/mL lysozyme and 0.8 U/mL benzonase and incubation at 600 rpm and

RT for 30 min. Lysates were cleared (60 min 3220 × *g* 4 °C) and used in the lysate activity assays as described below.

Lysate activity assay. Cleared lysate was diluted as appropriate to give a linear activity profile in the assay. Diluted cleared lysate was distributed at 20 µL in a clear 96-well plate. Reaction was started by addition of 180 µL substrate solution. PheDH activity was measured with 5 mM L-phenylalanine and 2 mM NAD^+ in 100 mM Glycine-KOH pH 10. AmDH deamination activity was measured by addition of 5 mM (*R*)-1-methyl-3-phenylpropylamine and 2 mM NAD^+ in 100 mM Glycine-KOH pH 10. AmDH amination activity was measured by addition of 5 mM 4-phenyl-2-butanone and 0.5 mM NADH in 0.5 M NH_4Cl/NH_4OH pH 9.6. The absorbance of NADH at 340 nm was measured in a spectrophotometer and the slope of the linear initial rate was extracted with a custom python script to ease data analysis, available at: https://github.com/pauljannis/96well_auto.

G.4. Protein Purification and Characterization

Protein purification. Expression cultures in 50 mL LB were inoculated to an absorbance at 600 nm (OD_{600}) of 0.05 from a well-grown overnight culture and grown at 37 °C 200 rpm for approximately 2 h or until an OD_{600} of 0.4-0.8 was reached. Protein expression was induced by addition of 200 ng/mL aTc and continued for 16 h at 20 °C 200 rpm. Protein purification was performed using Strep-Tactin Sepharose (IBA Lifesciences), largely as described by the manufacturer. All steps continued on ice with pre-chilled buffers. The expression culture is sedimented (20 min, 3220 × *g*, 4 °C) and lysed by suspension of the cell pellet in 5 mL lysis buffer (1x BugBuster (Merck #70921), 1 cOmplete Mini protease inhibitor tablet (Roche) per 10 mL and 3 µL/mL lysonase (Merck) in 100 mM Glycine-KOH pH 9) and incubation at RT and 600 rpm for 1 h. The crude lysate is cleared by centrifugation (1 h, 17,000 × *g*, 4 °C). The Strep-Tactin Sepharose was prepared in gravity flow columns to a final column volume (CV) of 1 mL. The column was equilibrated with 4 CV water followed by three times 4 CV wash buffer (150 mM NaCl in 100 mM Glycine-KOH pH 9). Next, the cleared lysate is loaded and washed two times with 3 CV wash buffer. Bound protein is eluted 6 times with 0.5 CV elution buffer (2.5 mM desthiobiotin and 150 mM NaCl in 100 mM Glycine-KOH pH 9), fractions collected separately. Elution fractions and samples taken throughout the purification process are analysed for purity via SDS-PAGE (see below). Clean elution fractions are pooled, buffer exchanged to 100 mM Glycine-KOH pH 9 with a desalting column (PD-10, GE Healthcare)

and concentrated (Amicon Ultra-4 10 kDa, Merck). Protein concentrations are measured with a NanoDrop 2000 (Thermo Fisher) and adjusted to 50 µM. Purified protein is immediately for all further analyses.

Analysis of soluble expression via SDS-PAGE. Small samples were taken at all steps during the purification process for quality control on SDS-PAGE (NuPAGE Bis-Tris gels, Thermo Fisher). Additionally, the remaining pellet after clearing the crude lysate is used to establish the amount of recombinant protein expression in soluble form. The soluble fraction is transferred to a new tube and the remaining pellet is suspended in diluted lysis buffer (1:5 dilution with water). The pellet fraction is then washed twice (spun down 1 min at 14,100 × g and resuspend in diluted lysis buffer). The crude extract, soluble fraction and pellet fraction are prepared by denaturation at 95 °C for 10 min in sample buffer (NuPAGE LDS Sample Buffer containing 200 mM dithiothreitol) and run in an SDS-Gel and analysed after Coomassie staining (InstantBlue, Expedeon) for intensity of the recombinant protein band. The gel is scanned and gel densitometry is performed in ImageJ Fiji after conversion to grey scale, comparing the intensities of the recombinant protein bands in the different fractions.

Steady state kinetics. Steady state kinetics were acquired for amination and deamination activities after determining suitable enzyme concentrations, usually 10 µM for amination and 1 µM for deamination. Substrate concentrations of 0 – 12.8 mM were presented in a 96-well plate and the reaction was started by adding purified enzyme. Amination was performed with 0.5 mM NADH in 0.5 M NH_4Cl/NH_4OH pH 9.6, whereas deamination was performed with 2 mM NAD^+ in 100 mM Glycine-KOH pH 10. Initial rates were obtained by measuring NADH absorbance at 340 nm in a spectrophotometer.

Thermal stability. Thermal stability was measured as the melting point in differential scanning fluorimetry with SYPRO orange (Invitrogen). Concentrations of 1-10 µM protein with 1-10× SYPRO orange were prepared in 100 mM Glycine-KOH pH 9. The heat denaturation and fluorescence measurements were recorded every 30 s from 25 °C to 95 °C in 0.5 °C steps with a Corbett Rotor-Gene 6000 (Qiagen). Combinations of protein and SYPRO orange concentrations were measured and the combination yielding the greatest signal was used for a repeated measurement in triplicates. The melting point was determined by extracting the relevant signal, normalizing the fluorescence and fitting a four-parameter logistic curve in GraphPad Prism.

Total turnover number. Small scale biotransformations were set up as 500 µL reactions with a concentration of 0.1 µM enzyme and 10 mM 4-phenyl-2-butanone in 0.5 M NH_4Cl/NH_4OH pH 9.6. Co-factor was added (2 mM NAD^+) and recycled over the course of the reaction with a glucose dehydrogenase system (50 mM glucose and 10 mg/mL GDH-101, Johnson Matthey). After reaction progress for 36 h at 30 °C and 800 rpm, reactions were extracted with 1 mL ethyl acetate and analysed via HPLC. A water-acetonitrile gradient with 0.1% (v/v) trifluoroacetic acid was applied (10% acetonitrile for 30 s, increase to 90% acetonitrile over 90 s, hold for 30 s) at 20 °C with 0.3 mL/min on a 1290 Infinity II LC (Agilent) fitted with a XBridge Phenyl column (4.6x30 mm, Waters). Substrate (4-phenyl-2-butanone, 1.93 min) and product (1-methyl-3-phenylpropylamine, 1.43 min) were detected at 210 nm and conversions were determined from observed uncorrected peak areas. The number of substrate turnovers per enzyme molecule is calculated from the relative conversion.

G.5. MICROFLUIDICS

Chip design, fabrication and replication. Chip masters were designed in AutoCAD and fabricated using standard soft lithography procedures (Xia & Whitesides, 1998). Silicon wafers of 3-inch diameter (Siegert Wafers) were coated with photoresist (SU-8 2050, Microchem) by spinning at the specified speeds according to the manufacturer's instructions. Chip designs were imprinted on plastic film photomasks (Microlitho) and used to pattern the photoresist via UV exposure in a MJB4 mask aligner (SUSS MicroTech). After development and hard baking as in the photoresist's specifications, these chip masters were coated by evaporating two drops of trichloro(1H,1H,2H,2H-perfluorooctyl)silane. The hardened chips can now be used to produce many replicas in polydimethylsiloxane (PDMS, Sylgard 184, Dow Corning). PDMS is poured onto the master, degassed (30 min 25 mPa) and hardened for 4 h at 65 °C. Now, the inlet and outlets of the PDMS replica device are opened (1 mm biopsy punch) and the device is bonded onto a glass slide after being activated by exposure to oxygen plasma (Diener Femto SLS). The device is flushed with 1% trichloro(1H,1H,2H,2H-perfluorooctyl)silane in HFE-7500 (3M Novec) to achieve a hydrophobic surface modification of the channels and is ready to use in subsequent microfluidic experiments.

Preparation of droplet incubation chambers. Droplet emulsions were incubated in modified reaction tubes. The chambers (Fig. C-4) were built using a conventional 0.5 mL reaction tube, to which polyethylene tubing was attached (0.38 mm inner diameter, 1.09 mm outer diameter,

Smiths Medical). The tubing was inserted into access holes opened to the reaction tube with a 1 mm biopsy punch and fixed in place with a high viscosity cyanoacrylate glue (PR1500, Scotch-Weld).

Preparation of cell suspensions. To generate monoclonal microfluidic droplets, i.e. droplets containing predominantly single cells, regular cell suspensions have to be diluted. The encapsulation of a cells into a droplet follows a Possion distribution and in most experiments performed here the expected rate of occurrence λ is set to 0.2, thus resulting in 81.7% empty droplets, 16.4% single encapsulations and 1.7% multiple cells per droplets (Fig. C-1). To achieve the corresponding cell dilution, a cell density of 5×10^8 cells per mL is assumed for *E. coli*, multiplied by the targeted λ and divided by the generated droplet volumes. In this study, cells were encapsulated in either 100 pL or in 300 pL droplets, for cell growth or direct assay, respectively. The resulting dilutions are a final OD_{600} of 0.005 and 0.0017.

Single cell lysate assay. For library sorts, all colonies of a fresh transformation are washed off the agar plate with 3 mL LB medium. A 4 mL expression culture is inoculated from this stock to an OD_{600} of 0.8 and immediately induced (200 ng/mL aTc). After 16 h of protein expression at 20 °C, a cell suspension is prepared at the required OD_{600} for encapsulation, as described above, by dilution in reaction buffer (100 mM Glycine-KOH pH 9) containing 25% (v/v) Percoll to prevent cells from sedimentation. Droplets of 300 pL volume are generated in a conventional droplet generator (80 µm height and 50 µm width at the flow focussing junction) by co-encapsulating the diluted cell suspension and the assay solution (6 mM WST-1 (NBS Biologicals), 6 mM (*R*)-1-methyl-3-phenylpropylamine, 2 mM NAD^+, 5 µg/mL 1-methoxy-5-methylphenazinium methyl sulfate (mPMS), 1 µL/mL rLysozyme (Merck) and 0.8x CelLytic B in 100 mM glycine-KOH pH 9) in fluorous oil (HFE-7500 containing 1% (v/v) 008-FluoroSurfactant (RAN biotechnologies)). Flow rates for encapsulation were 8 µL/min for each of the aqueous phases and 30 µL/min for the oil phase and the resulting emulsion was collected in a droplet chamber for incubation.

Oxygenation for cell growth. To achieve homogeneous cell growth, single cells were encapsulated in 100 pL droplets containing TB medium on a conventional droplet generator (50 µm height and 50 µm width at the flow focussing junction) by flowing the diluted cell suspension at 8 µL/min and the oil (HFE-7500 with 1% (w/v) 008-FluoroSurfactant) at 24 µL/min. Droplets were collected in droplet incubation chambers. During growth, the emulsion was oxygenated by pushing 4 µL/min oil through the emulsion from the top (Fig.

C-4). If protein expression was induced via aTc, the oil phase was changed to inducer oil, additionally containing 400 ng/mL aTc.

Cell growth lysate assay. For the cell growth assay, single cells were encapsulated in TB medium in droplet incubation chambers as described above. Oxygenation was performed for 2 h at 37 °C for cell growth, before switching to inducer oil (HFE-7500 with 1% (w/v) 008-FluoroSurfactant and 400 ng/mL aTc), enabling protein expression for 16 h at 20 °C. The substrate solution was supplied to these droplets via 'pico-fusion'. A flow focussing junction (50 µm width, 50 µm height) on the pico-fusion chip (Fig. C-7) generates substrate droplets of 200 pL containing 4.5 mM WST-1, 4.5 mM (*R*)-1-methyl-3-phenylpropylamine, 3 mM NAD^+, 7.5 µg/mL mPMS, 200 µg/mL streptomycin, 2 µL/mL rLysozyme and 0.6x CelLytic B in 100 mM glycine-KOH pH 9 at flow rates of 2 µL/min and 3 µL/min for the aqueous and oil phase, respectively. The droplets containing the grown cells were injected at 200 Hz or 1-2 µL/min and synchronized to the generated substrate droplets in a delay channel. The smaller cell droplets pack tightly behind the larger substrate droplets and are fused in the fusion chamber under application of an electric field (400 V, 10 kHz square wave). Fused droplets are collected in a droplet incubation chamber and sorted via AADS as described below.

Absorbance-activated droplet sorting (AADS). Absorbance of droplets was measured in AADS, as described previously (Gielen *et al*, 2016). The droplet emulsion was injected into the sorting chip at 100-200 Hz or 1-3 µL/min. Spacing oil (HFE-7500) was added at 30-40 µL/min and the absorbance of droplets passing through the interrogation point was analysed. At the interrogation point, optical fibers (M14, Thorlabs) are fixed (Fig. A-10). The fibers were inserted after de-cladding and cutting to an even edge and are coupled to a LED light-source (455 nm, M455F3 Thorlabs) and a detector (PDA100A2, Thorlabs). The signal is analysed live on an Arduino Due microcontroller, as well as recorded for later analysis and visualization on a computer. Sorting and analysis scripts written for this thesis are available online at https://github.com/pauljannis/AADSutility. The microcontroller actuates sorting by activating salt-water electrodes filled with 5 M NaCl (Sciambi & Abate, 2014), pulling droplets with high absorbance into the sorting channel (600 V, 10 kHz square wave, 5 ms pulse). Sorted droplets are collected in a DNA low-binding reaction tube (DNA LoBind 0.5 mL, Eppendorf) and their contained plasmid DNA is isolated and purified (Kintses *et al*, 2012). To recover plasmid DNA from the sorted droplets, the same volume 1H,1H,2H,2H-perfluorooctanol and 100 µL of nuclease-free water containing 100 ng/mL junk DNA (salmon-sperm) were added to the collected droplets. This mixture was vigorously mixed for 1 min and then spun down

(14,000 × *g* for 1 min) to enable extraction of the aqueous phase. This extraction was repeated two times with 100 µL water containing salmon-sperm DNA. The total collected 300 µL aqueous phase containing the recovered plasmids from sorted droplets was purified and concentrated using the Zymo Clean & Concentrator-5 (Zymo Research) and used to transform electrocompetent E. cloni 10G ELITE (Lucigen).

G.6. NANOPORE SEQUENCING

UMI complexity control. Plasmids with variants recovered from sorting were used as templates for UMI-tagging and subsequent nanopore sequencing with the primers listed in Table 12. The UMI and an experiment-specific barcode, here labelling the round of evolution, are introduced by tagging of the gene variants in two cycles of PCR. The experiment specific barcode is a unique 24 base stretch, distinctly identifiable after sequencing. Barcode sequences from the PCR Barcoding Expansion Pack 1-96 (EXP-PBC096, Oxford Nanopore Techologies) were used in custom primers (BCX-F). The UMI is introduced via the reverse primer (UMI-R), containing a total of 50 randomized bases thus introducing a unique sequence to each molecule. Both primers contain 25 base homology overhangs for Gibson assembly in a later step. Two cycles of PCR are performed with the Q5 High-Fidelity 2X Master Mix (NEB M0492) to attach a barcode and a UMI to each DNA molecule in the following reaction:

	Reaction:		Cycling:		
	25 µL	Q5 2X master mix	98 °C	1 min	
	2.5 µL	BC-F primer (10 µM)	98 °C	10 s	
	2.5 µL	UMI-R primer (10 µM)	60 °C	30 s	2 cycles
	500 ng	Template DNA	72 °C	1 min	
ad	50 µL	H_2O	72 °C	5 min	

The PCR product is purified using the Zymo Clean & Concentrator-5 (Zymo Research) and used as input for a second PCR amplifying the UMI- and barcode-tagged DNA fragments via the common 25 base Gibson overhangs. The total purified DNA from the previous step is used as template in a PCR with Gibson-F and Gibson-R primers, amplified for 15 cycles. This PCR product is used in a Gibson assembly reaction with *Bam*HI/*Kpn*I linearized pRSFDuet-1 as acceptor. The linearized acceptor fragment is generated by restriction digest, followed by agarose gel purification, and the Gibson assembly is performed with the Gibson Assembly Master Mix (NEB E2611).

	Reaction:		Cycling:	
	100 ng	Linearized pRSFDuet-1	50 °C	60 min
	100 ng	PCR fragment		
	10 µL	Gibson Assembly Master Mix (2X)		
ad	20 µL	H_2O		

Next, the Gibson Assembly reaction is purified and used for the transformation of electrocompetent E. cloni 10G ELITE (Lucigen). Crucially, a dilution series is plated from this transformation and the required colony number is selected. Here, if 1000 variants were sorted and thus contained in the plasmid pool for sequencing, 3000 – 5000 colonies would be selected in the next step to account for statistical sampling of each of those variants. The colonies are used as input for plasmid isolation (GeneJET, Thermo Fisher K0503) to obtain the final UMI-tagged and complexity restricted variant library.

Table 12: Primers used for nanopore sequencing. Primers for the first nanopore run (chapter D, nano1) were slightly altered in the second run (chapter E, nano2). Features: Homology overhang for Gibson assembly (yellow). UMI (red). Restriction enzyme site (green). Experiment-specific barcode (blue). Binding (black). Spacer (grey).

Name	Sequence
Nano1-UMI-R	GTTTCTTTACCAGACTCGAGGGTACCGATACNNNNNNNNNNNNNNNNNNNNNNNNGATACNNNNNNNNNNNNNNNNNNNNNNNNGCTCCAAGCGCTCTCGAG
Nano2-UMI-R	TACATCGAACTGGATCTCAACAGCGNNNYRNNNYRNNNYRNNNYRNNYRGATACNNNYRNNNYRNNNYRNNNYRNNYRTGGCTCCAAGCGCTCTCGAG
Nano2-BC1-F	CTTACTTCTGACAACGATCGGAGGACAAGAAAGTTGTCGGTGTCTTTGTGTCGACAAAAATCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATG
Nano2-BC2-F	CTTACTTCTGACAACGATCGGAGGACTCGATTCCGTTTGTAGTCGTCTGTTCGACAAAAATCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATG
Nano2-BC3-F	CTTACTTCTGACAACGATCGGAGGACGAGTCTTGTGTCCCAGTTACCAGGTCGACAAAAATCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATG
Nano2-BC4-F	CTTACTTCTGACAACGATCGGAGGACTTCGGATTCTATCGTGTTTCCCTATCGACAAAAATCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATG
Nano2-BC5-F	CTTACTTCTGACAACGATCGGAGGACCTTGTCCAGGGTTTGTGTAACCTTTCGACAAAAATCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATG
Nano2-BC6-F	CTTACTTCTGACAACGATCGGAGGACTTCTCGCAAAGGCAGAAAGTAGTCTCGACAAAAATCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATG
Nano2-BC7-F	CTTACTTCTGACAACGATCGGAGGACGTGTTACCGTGGGAATGAATCCTTTCGACAAAAATCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATG
Nano2-BC8-F	CTTACTTCTGACAACGATCGGAGGACTTCAGGGAACAAACCAAGTTACGTTCGACAAAAATCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATG
Nano1-BC1-F	ACCATCATCACCACAGCCAGGATCCGATACAAGAAAGTTGTCGGTGTCTTTGTGCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATG
Nano1-BC2-F	ACCATCATCACCACAGCCAGGATCCGATACTCGATTCCGTTTGTAGTCGTCTGTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATG
Nano1-BC3-F	ACCATCATCACCACAGCCAGGATCCGATACGAGTCTTGTGTCCCAGTTACCAGGCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATG
Nano1-Gibson-F	ACCATCATCACCACAGCCAG
Nano1-Gibson-R	GTTTCTTTACCAGACTCGAGGGTAC
Nano2-Gibson-F	ATACCATGGGCAGCAGCCATCACCATCATCACCACAGCCAGGATCCCTTACTTCTGACAACGATCGGAGGAC
Nano2-Gibson-R	GTTCAAATTTCGCAGCAGCGGTTTCTTTACCAGACTCGAGGGTACCTACATCGAACTGGATCTCAACAGCG

Library preparation and sequencing. The region of interest (gene variant with UMI and barcode) is excised from the plasmid library obtained in the steps above. This is achieved by conventional restriction digest with *Bam*HI and *Kpn*I, followed by purification of the excised fragment after agarose gel electrophoresis. This DNA fragment is used at 200 fmol concentration as input for nanopore sequencing using the SQK-LSK109 kit (Oxford Nanopore Technologies). Further steps are performed as described by the manufacturer, using the Short Fragment Wash Buffer and sequencing with a FLO-MIN106 (R9.4.1 pore) flow cell (Oxford Nanopore Technologies) until an expected variant coverage of 100x is achieved. Base calling of raw reads is done with guppy 3.5.2 (Oxford Nanopore Technologies) on a GeForce GTX 1050 with default settings. Sequencing metrics were visualized with NanoPlot and base called reads were filtered for length of 1250 – 1400 b and a minimum average base call quality score of 8 using NanoFilt (De Coster *et al*, 2018).

Bioinformatic pipeline 'UMIC-seq'. Filtered reads were further analysed with custom scripts written in Python, making use of open-source libraries such as biopython (Cock *et al*, 2009) and scikit-bio (http://scikit-bio.org/). Script are available at https://github.com/fhlab/UMIC-seq. A workflow diagram using the custom scripts is shown in Fig. D-4. Firstly, the experiments are demultiplexed by aligning the available 24 b barcode sequences and assigning one barcode to each read, using the 'UMIC-seq demultiplex' functionality with default settings. Next, the UMI is extracted from the reads by identifying a 100 b constant region adjacent to the UMI, using the 'UMIC-seq UMIextract' function. The extracted UMIs are then clustered, i.e. placed into groups based on sequence similarity. Reads with the same UMI will stem from the same variant and can be used to generate accurate consensus sequences, however, a sequencing error cut-off needs to be found for the groupings, because reads of the same UMI will have slight alterations due to the high raw read error-rate in nanopore sequencing of approximately 10%. This cut-off, or rather clustering threshold, can be approximated by generating clusters at different cut-offs and estimating their diversity. This progression should yield a saturation curve, as once true clusters are established, a further increase in clustering stringency will not increase cluster homogeneity. Such a threshold approximation can be performed using the 'UMIC-seq clustertest' functionality, and a suitable threshold for full clustering can be identified as a value beginning saturation. This value will usually be similar to the length of the UMI. Full clustering is then performed with the 'UMIC-seq clusterfull' functionality, generating all clusters of the full reads based on their UMI similarities. The reads of each cluster are then used to generate a polished consensus sequence from the raw signal with Nanopolish

0.10.2 (Loman *et al*, 2015), using the non-mutated gene as a reference sequence and the nanopolish variants module with parameters *--consensus --min-flanking-sequence 3 --max-round 100 --methylation-aware dam,dcm*.

Variant analysis. The consensus variants obtained from nanopore sequencing were analysed using custom python scripts. The main analysis was a t-SNE dimensionality reduction (Maaten & Hinton, 2008), as implemented in the scikit-learn python library (Pedregosa *et al*, 2011). For this analysis, unique variants and their frequencies in the total pool of sequences are analysed based on their similarity, here as a metric of Levenshtein distance (Yujian & Bo, 2007). Unique variants are represented as dots, closely placed to other similar variants, establishing patterns and phylogenies. The custom-made implementation in Python generated for this study is available at https://github.com/pauljannis/pySSN.

H. APPENDIX

H.1. TABLE OF ABBREVIATIONS

AADH	amino acid dehydrogenase
AADS	absorbance-activated droplet sorting
AmDH	amine dehydrogenase
aTc	anhydrotetracycline
BL21	*E. coli* BL21 (DE3)
b	base
CI	confidence interval
CV	column volume
DH5a	*E. coli* DH5a or E. cloni 10G
E4	evolutionarily-stabilized PheDH variant of round 4
epPCR	error-prone PCR
GFP	green fluorescent protein
IQR	interquartile range
mPMS	1-methoxy-5-methylphenazinium methyl sulfate
NEB	New England Biolabs
OD_{600}	absorbance at 600 nm
ONT	Oxford Nanopore Technologies
P4	PROSS-stabilized PheDH design 4
PCR	polymerase chain reaction
PDMS	polydimethylsiloxane
PheDH	phenylalanine dehydrogenase
RT	room temperature
SD	standard deviation
SSN	sequence similarity network
StEP	staggered extension process
tSNE	t-distributed stochastic neighbor embedding
UMI	unique molecular identifier
UMIC-seq	UMI-linked consensus sequencing
WST-1	water-soluble tetrazolium salt 1
WT	wild type

H.2. TABLE OF FIGURES

FIG. A-1: CYCLES OF DIRECTED EVOLUTION. 3
FIG. A-2: PROTEIN SEQUENCE SPACE AND FITNESS LANDSCAPE. 4
FIG. A-3: EPISTASIS RESULTS IN THE CONTEXT-DEPENDENCE OF MUTATIONAL EFFECTS. 5
FIG. A-4: STABILITY AND MUTATIONAL ROBUSTNESS. 7
FIG. A-5: PROTEIN DYNAMISM AND THE EVOLUTION OF NEW FUNCTION. 9
FIG. A-6: A HYDROGEN BORROWING CASCADE FOR THE SYNTHESIS OF A LEUCINE-SPECIFIC DEMETHTYLASE-1 INHIBITOR. 12
FIG. A-7: SELECTION OF ENZYMATIC ROUTES TO CHIRAL AMINES. 13
FIG. A-8: ENGINEERING OF AMINE DEHYDROGENASES FROM AMINO ACID DEHYDROGENASES. 14
FIG. A-9: DROPLET MICROFLUIDIC WORKFLOW FOR ENZYME ENGINEERING. 16
FIG. A-10: DESIGN OF MICROFLUIDIC DEVICES FOR AADS. 19
FIG. A-11: DEEP MUTATIONAL SCANNING GENERATES FUNCTIONAL VALUES FOR MANY VARIANTS. 23
FIG. A-12: DEEP MUTATIONAL SCANNING DATA CAN INFORM PROTEIN ENGINEERING. 26
FIG. B-1: CHARACTERIZATION OF PHEDH STABILIZATION SUCCESS VIA DIRECTED EVOLUTION. 35
FIG. B-2: OVERVIEW OF STARTING POINTS. 36
FIG. B-3: DISTRIBUTION OF FITNESS EFFECTS OF AMINE DEHYDROGENASE KEY MUTATIONS IN THREE STARTING POINTS. 38
FIG. B-4: COMBINATIONS OF ACTIVATING MUTATIONS IN POSITIONS K66 AND N262 FOR THREE PHEDH STARTING POINTS. 40
FIG. B-5: OVERVIEW OF WORK PERFORMED IN THIS CHAPTER. 42
FIG. C-1: POISSON DISTRIBUTION FOR DIFFERENT VALUES OF λ. 46
FIG. C-2: AMDH DETECTION IN AADS. 48
FIG. C-3: OVERVIEW OF MICROFLUIDIC CELL LYSATE ASSAYS. 49
FIG. C-4: DROPLET INCUBATION CHAMBERS. 50
FIG. C-5: GENTLE OXYGENATION VIA THE OIL PHASE ENABLES STRONG AND HOMOGENEOUS CELL GROWTH IN DROPLETS. 51
FIG. C-6: PROTEIN EXPRESSION IN MONOCLONAL DROPLET CULTURE WITH DIFFERENT INDUCTION SYSTEMS. 54
FIG. C-7: PICO-FUSION DEVICE FOR SELECTIVE DROPLET COALESCENCE. 57
FIG. C-8: COMPARISON OF DETECTION ALGORITHMS EMPLOYED IN AADS. 58
FIG. C-9: WST-1 ABSORBANCE AS A READOUT FOR AMDH ACTIVITY IN AADS WITH DIFFERENT GROWTH CONDITIONS. 60
FIG. C-10: DROPLET SIZE (AS PEAK WIDTH) AND ABSORBANCE (AS PEAK HEIGHT) MEASURED OF AMDH EXPRESSING CELLS IN THE GROWTH WORKFLOW. 61
FIG. C-11: QUANTIFICATION OF IMPROVEMENTS ACHIEVED BY IN-DROPLET CELL GROWTH. 63
FIG. D-1: A TOPOGRAPHICAL VIEW OF A FITNESS LANDSCAPE. 68
FIG. D-2: CONSENSUS APPROACHES TO INCREASE THE ACCURACY OF NANOPORE SEQUENCING. 70
FIG. D-3: UMI-LINKED CONSENSUS SEQUENCING (UMIC-SEQ) STRATEGY FOR ACCURATE AMPLICON SEQUENCES FROM NANOPORE LONG READS. 72
FIG. D-4: OVERVIEW OF THE BIOINFORMATIC TOOLS USED TO GENERATE ACCURATE CONSENSUS SEQUENCES FROM UMI-LINKED NANOPORE READS. 74
FIG. D-5: LYSATE ACTIVITY OF SORTED VARIANTS. 76

Fig. D-6: Influence of UMI length, error rate and library size on clustering efficiency. ... 77
Fig. D-7: Analysis of errors in nanopore consensus sequences. ... 79
Fig. D-8: Analysis of UMI sequence composition. ... 80
Fig. D-9: Mutations per variant over the course of three rounds of directed evolution. ... 81
Fig. D-10: Phylogenetic analysis of directed evolution. ... 82
Fig. D-11: Lysate activity of founder variants relative to non-mutated reference. ... 83
Fig. D-12: Sign epistasis in founder variant A64E R102S E323V. ... 84
Fig. E-1: Neutral variation can increase evolvability. ... 91
Fig. E-2: Relative lysate activities of randomly selected library variants before and after the first screening. ... 93
Fig. E-3: Relative lysate activities of randomly selected variants after each round of directed evolution. ... 94
Fig. E-4: Visualization and global properties of the $P4_{QM}$ and $E4_{QM}$ directed evolution sequence space. ... 97
Fig. E-5: Relative lysate activity of founder variants. ... 98
Fig. E-6: Activity of beneficial founder mutations measured in both $P4_{QM}$ and $E4_{QM}$ backgrounds. ... 100
Fig. E-7: Conversion mutations in the directed evolution of $P4_{QM}$ and $E4_{QM}$. ... 103
Fig. E-8: Frequency of conversion mutations in each round of directed evolution. ... 105
Fig. E-9: Overview of $P4_{QM}$-$E4_{QM}$-chimeras generated for characterization. ... 106

Fig. S1: NADH detection reaction in AADS. ... 135
Fig. S2: Threshold approximation for UMI clustering. ... 136
Fig. S3: Distribution of cluster sizes. ... 136
Fig. S4: Michaelis-Menten kinetics of founder variants. ... 137
Fig. S5: Network graphs of 10 rounds of (A) $P4_{QM}$ and (B) $E4_{QM}$ directed evolution. ... 138
Fig. S6: Michaelis-Menten kinetics of long-term evolution variants. ... 139

H.3. TABLE OF TABLES

TABLE 1: DETECTION MODES FOR DROPLET MICROFLUIDIC ENZYME ASSAYS. 18

TABLE 2: TYPES OF ENZYME REACTIONS DETECTED IN DROPLETS. 21

TABLE 3: DEEP MUTATIONAL SCANNING STUDIES PERFORMED ON ENZYMES. 24

TABLE 4: EXAMPLES FOR DIFFERENT ENZYME STABILIZATION METHODS. 31

TABLE 5: COMPUTATIONAL STABILIZATION WITH THE PROSS ALGORITHM. 33

TABLE 6: BIOCHEMICAL CHARACTERIZATION OF A64E R102S E323V VARIANTS. 85

TABLE 7: CHARACTERIZATION OF THE MOST ACTIVE $P4_{QM}$ AND $E4_{QM}$ VARIANTS. 101

TABLE 8: CHARACTERIZATION OF RECOMBINED VARIANTS. 107

TABLE 9: PRIMERS USED FOR THE GENERATION OF ERROR-PRONE PCR LIBRARIES. 119

TABLE 10: LIST OF PLASMIDS GIFTED OR CONSTRUCTED FOR THIS STUDY. 120

TABLE 11: PRIMERS USED FOR IVA CLONING OF CHIMERIC VARIANTS SELECTED-CHIMERA AND SMALLP4-CHIMERA. 121

TABLE 12: PRIMERS USED FOR NANOPORE SEQUENCING. 128

TABLE S1: CHARACTERIZATION OF EVOLUTIONARILY STABILIZED PHEDH VARIANTS. 140

TABLE S2: WEIGHTED DEGREE AND BETWEENNESS CENTRALITY CALCULATED FOR FOUNDER VARIANTS OF $P4_{QM}$ AND $E4_{QM}$ DIRECTED EVOLUTION. 140

H.4. SUPPLEMENTARY FIGURES

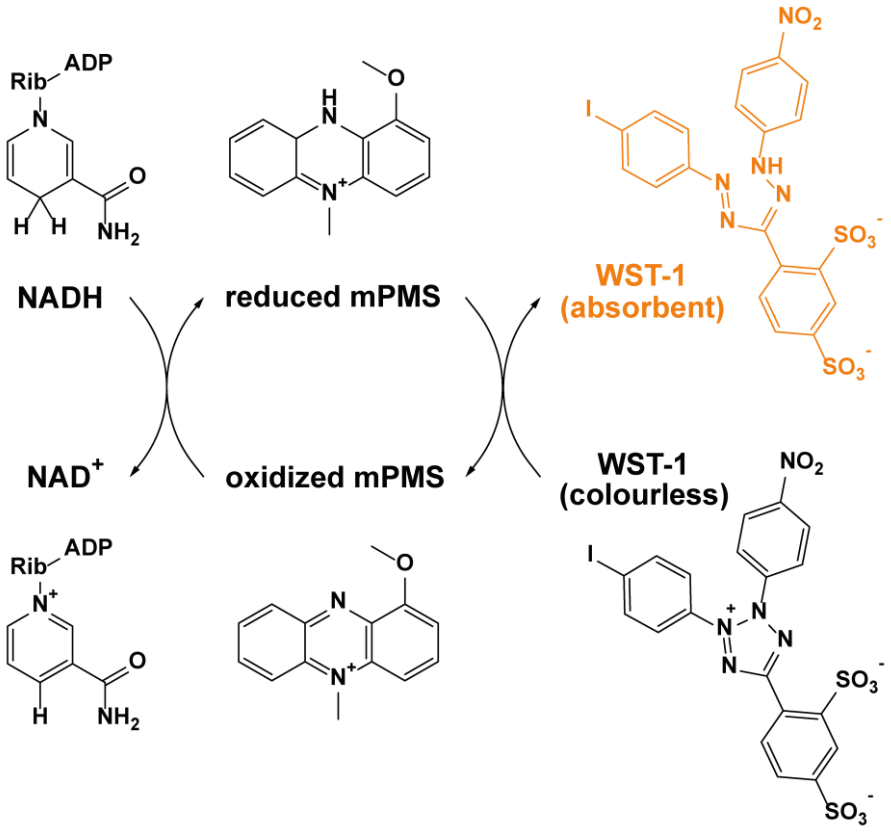


Fig. S1: NADH detection reaction in AADS. NADH, e.g. detected produced by the deamination activity of AmDHs or by live cells, is recycled to NAD^+ by the mPMS-coupled detection reaction in AADS. NAD: Nicotinamide adenine dinucleotide. mPMS: 1-methoxy-5-methylphenazinium methyl sulfate. WST-1: water-soluble tetrazolium salt 2-(4-iodophenyl)-3-(4-nitrophenyl)-5-(2,4-disulfophenyl)-2H-tetrazolium. Rib: Ribose. ADP: Adenosine diphosphate. Figure adapted from Gielen *et al* (2016).

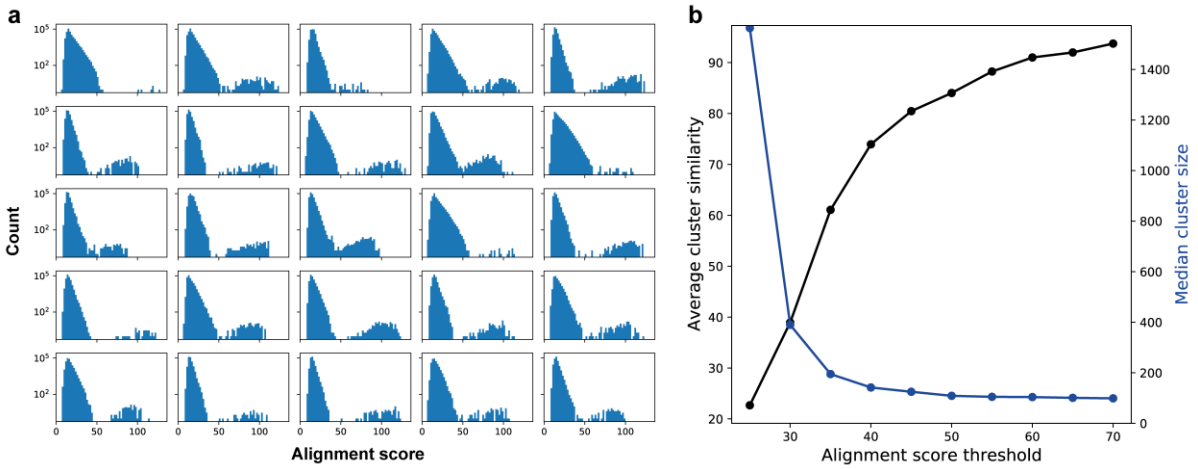


Fig. S2: Threshold approximation for UMI clustering. Exemplary output of the UMIC-seq tools 'clustertest' functionality shown. **(a)** *Alignment score histograms.* 25 randomly selected UMIs are aligned against all other UMIs and the distribution of resulting alignment scores is shown. **(b)** *Threshold approximation.* 25 randomly selected UMIs are clustered with increasingly stringent alignment score threshold. The average internal similarity as well as the median cluster size of the 25 resulting clusters is shown for each selected alignment score threshold.

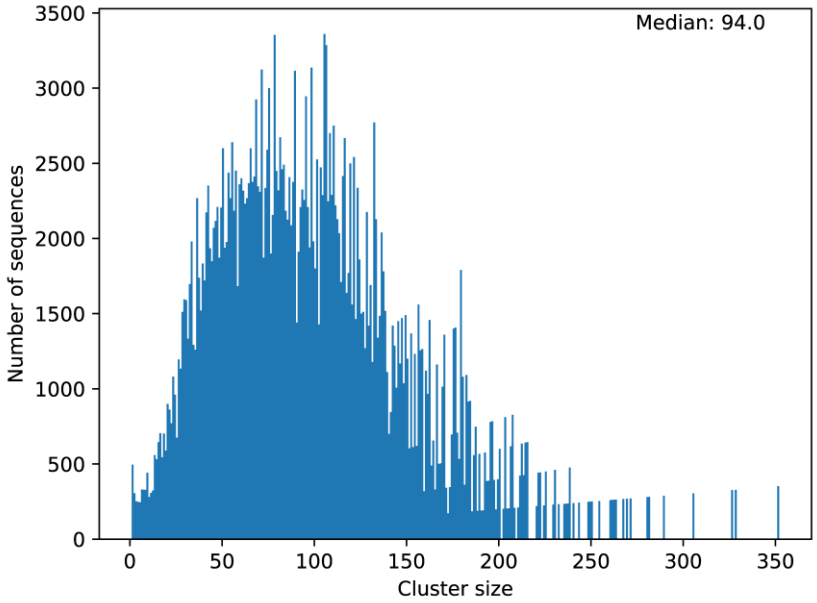


Fig. S3: Distribution of cluster sizes. The total number of sequences as a function of cluster size is shown. The median number of sequences per cluster is 94.

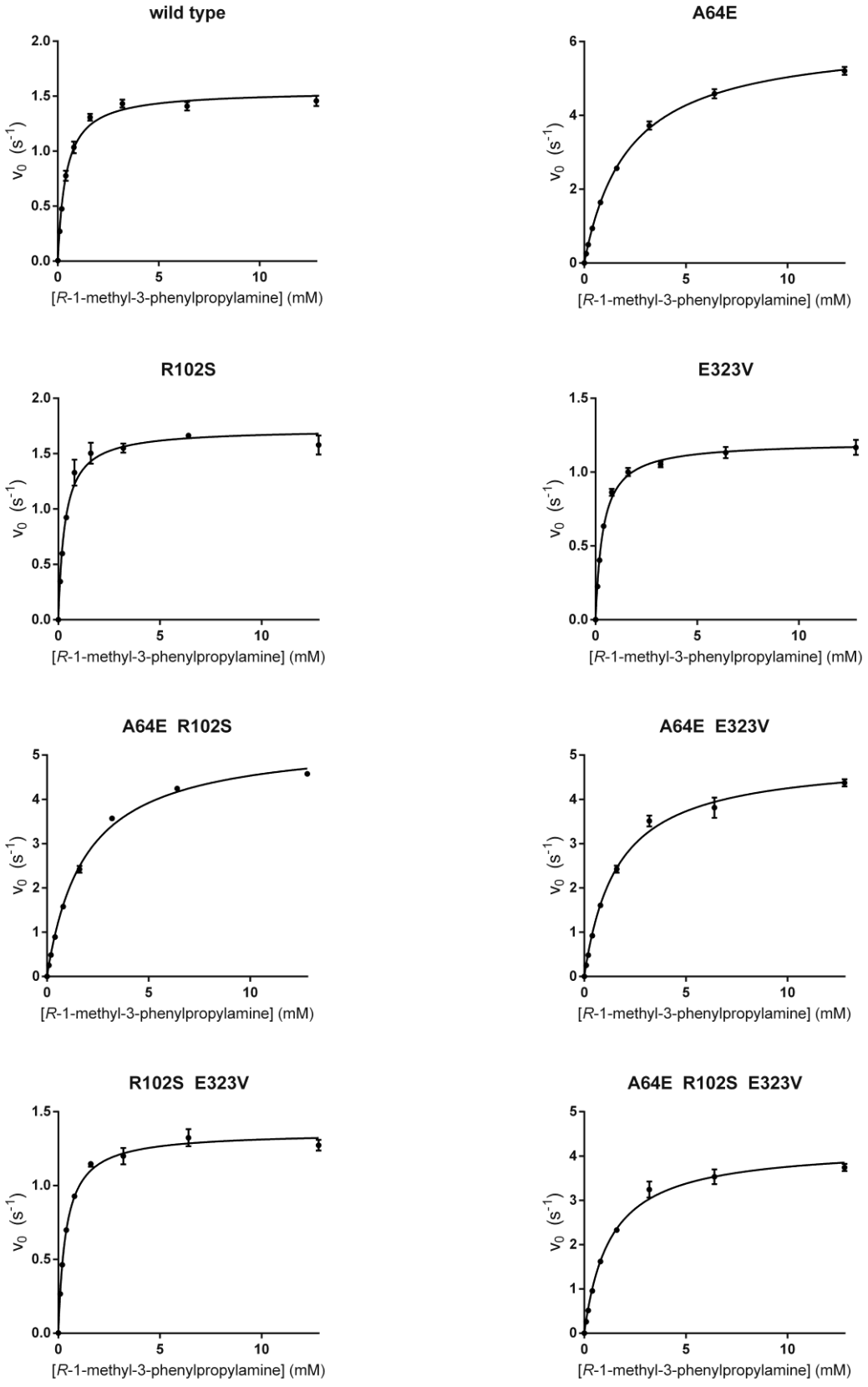


Fig. S4: Michaelis-Menten kinetics of founder variants. Wild type is the non-mutated $P4_{QM}$ AmDH. Reactions performed with 0 to 12.8 mM *R*-1-methyl-3-phenylpropylamine and 2.5 mM NAD^+ in 100 mM Glycine-KOH pH 10 at 22 °C. Independent technical replicates (n = 3, error bars show SD).

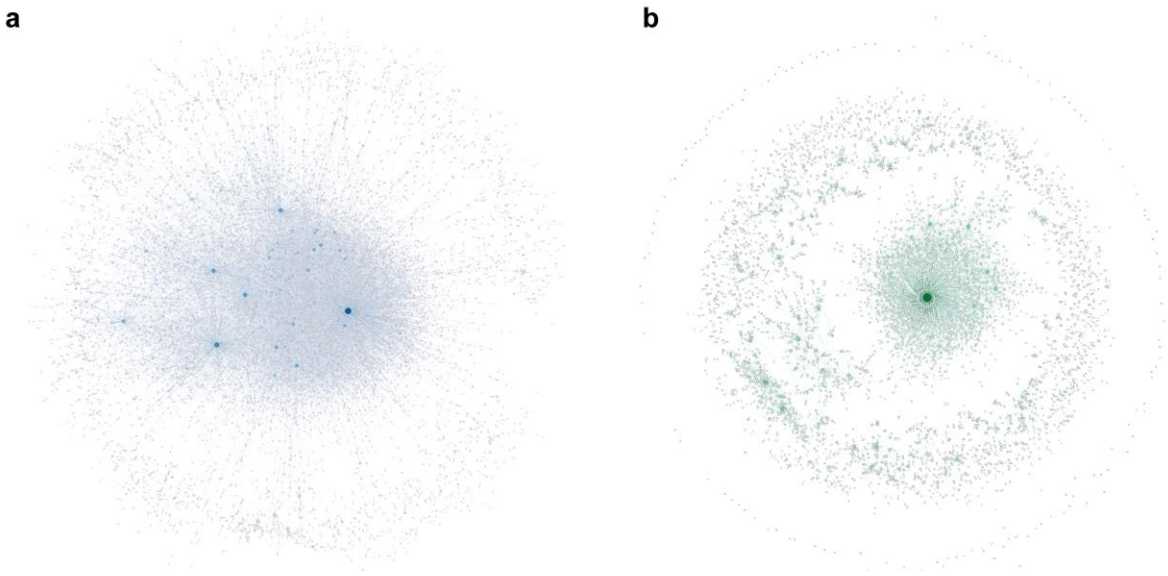


Fig. S5: Network graphs of 10 rounds of (a) P4$_{QM}$ and (b) E4$_{QM}$ directed evolution. Each node represents a unique variant which is connected via edges to each other node that is reachable by one mutation. Nodes are sized relative to their degree. Only nodes connected by at least 2 edges are shown. Force-directed layout generated via ForceAtlas 2 in Gephi 0.9.2.

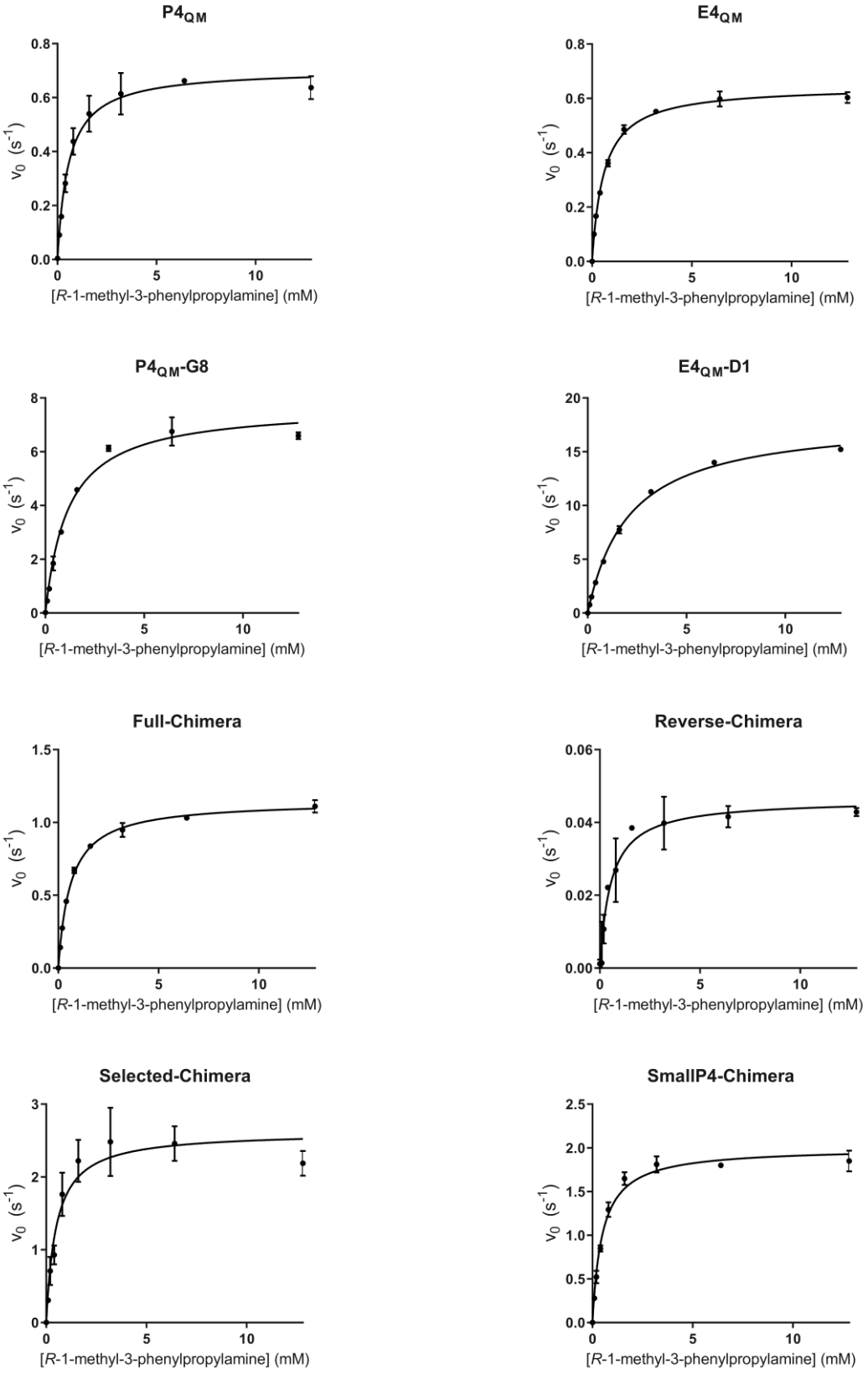


Fig. S6: Michaelis-Menten kinetics of long-term evolution variants. Reactions performed with 0 to 12.8 mM *R*-1-methyl-3-phenylpropylamine and 2.5 mM NAD^+ in 100 mM Glycine-KOH pH 10 at 22 °C. Independent technical replicates (n = 3, error bars show SD).

H.5. SUPPLEMENTARY TABLES

Table S1: Characterization of evolutionarily stabilized PheDH variants. Wild-type (WT) PheDH from *Rhodococcus sp.* M4. Var4 was chosen as the E4 variant for further study, as it showed a good trade-off of activity, thermal and kinetic stability. Variables showing error were determined in independent technical triplicates and show standard deviation.

	Var1	Var2	Var3	Var4 (E4)	P4*	WT
k_{cat} [s^{-1}]	26.4 ± 1.3	35.1 ± 1.6	30.0 ± 1.7	32.2 ± 1.6	2.20	18.7 ± 1.3
K_M [mM]	0.80 ± 0.10	1.28 ± 0.14	1.28 ± 0.19	1.55 ± 0.19	0.19	0.28 ± 0.05
K_i [mM]	28.3 ± 5.2	53.4 ± 12.2	72.1 ± 26.3	79.7 ± 25.9	n/a	8.1 ± 1.4
k_{cat}/K_M [s^{-1} M^{-1}]	32766 ± 0.1	27348 ± 0.1	23396 ± 0.2	20801 ± 0.1	11579	67292 ± 0.7
T_m [°C]	50.7 ± 0.12	54.5 ± 0.13	54.0 ± 0.14	53.1 ± 0.13	59.0	49.0± 0.18
Soluble expression [%]	90.2	70.9	70.2	85.4	94.0	28.9

*P4 data courtesy of Raphaëlle Hours.

Table S2: Weighted degree and betweenness centrality calculated for founder variants of $P4_{QM}$ and $E4_{QM}$ directed evolution. A network graph was generated in which each node represents a unique variant which is connected by edges to any other node within an edit distance of 2 or less.

	Weighted Degree	Betweenness Centrality
$P4_{QM}$	**4226**	**0.16039458**
R30H	1442	0.027926664
A148V	1167.5	0.020809441
W10R	1149	0.014050167
A64E	1103	0.014616032
S33N	946	0.006645116
G59D	908	0.006986279
I4T	854	0.005236744
M242L	787	0.003778797
F287Y	761	0.002479989
H311N	745.5	0.002567105

	Weighted Degree	Betweenness Centrality
$E4_{QM}$	**1493**	**0.017597789**
S151G	463.5	0.001399597
F160L	445.5	0.001012375
A314V	442	0.000938563
A148V	418.5	0.000560819
S191P	406.5	0.000341824
L226V	392.5	0.000213912
H104Q	389.5	0.000125628
V298I	387	0.000097647
S81G	380.5	0.000021612
N107C	379.5	0.000028374

H.6. SUPPLEMENTARY SEQUENCES

The amino acid sequences of the *Rhodococcus sp.* M4 wild-type PheDH (WT), the PROSS-stabilized variant PheDH (P4) and the evolutionarily-stabilized PheDH (E4) are shown. The two key active site residues turning the PheDH into AmDH (K66 N262) are highlighted in purple. Mutations relative to the WT reference are highlighted in blue for P4 and green for E4.

```
>WT
MGSIDSALNWDGEMTVTRFDRETGAHFVIRLDSTQLGPAAGGTRAAQYSQLADALTDAGKLA
GAMTLKMAVSNLPMGGGKSVIALPAPRHSIDPSTWARILRIHAENIDKLSGNYWTGPDVNTN
SADMDTLNDTTEFVFGRSLERGGAGSSAFTTAVGVFEAMKATVAHRGLGSLDGLTVLVQGLG
AVGGSLASLAAEAGAQLLVADTDTERVAHAVALGHTAVALEDVLSTPCDVFAPCAMGGVITT
EVARTLDCSVVAGAANNVIADEAASDILHARGILYAPDFVANAGGAIHLVGREVLGWSESVV
HERAVAIGDTLNQVFEISDNDGVTPDEAARTLAGRRAREASTTTATA

>P4
MGSIDSALNWDGEMTVTRFDAATGAHFVIRIHSTQLGPAAGGTRAWQYSSWADALTDAGRLA
RAMTYKMAVAGLPMGGGKSVIALPAPRHSIDPSTWARILRAHAEMIDSLNGRYWTGPDVNTN
SADMDILADETEFVFGRSPERGGAGSSAFTTALGVFEAMKATVAHRGLGSLDGLTVLVQGLG
AVGGSLAKLLAEAGAQLLVADTDTERVALAVELGHTWVALDDVLSTPCDVFAPCAMGGVITD
EVARTLDCKVVCGAANNVLAHEAAADILHARGILYAPDFVANAGGAIHLVGREVLGWSEDQV
HERARAIGDTLKEVFEIADKDGVTPDEAARELAERRMREASTTTATA

>E4
MGSIDRALNWDGEMTVTRFDRETGAHFIIRLDSTQLGPAAGGTRAAQYSQLADALTDAGKLA
GAMTLKMAISNLPMGGGKSVIALPAPRHSIDPSTWARILRIHAENIDKLSGNYWTGPDVNTS
SADMDTLNDTTEFVFGRSLERGGAGSSAFTTAVGVFEAMKATVAHRGLGSLDGLTVLVQGLG
AVGGSLASMAAEAGARLLVADTDTERVAHAVALGHTAVALEDVLSTPCDVFAPCAMGGVITT
EVARTLDCSVVAGAANNVIADEAASDILHARGILYAPDFVANAGGAIHLVGREVLGWSESVV
HERAVAIGDTLNQVFEISDNDGVTPDEAARILAGRRAREASTTTATA
```

I. REFERENCES

Abate AR, Hung T, Mary P, Agresti JJ & Weitz DA (2010) High-throughput injection with microfluidics using picoinjectors. *Proc Natl Acad Sci U S A* 107: 19163–19166

d'Abbadie M, Hofreiter M, Vaisman A, Loakes D, Gasparutto D, Cadet J, Woodgate R, Pääbo S & Holliger P (2007) Molecular breeding of polymerases for amplification of ancient DNA. *Nat Biotechnol* 25: 939–943

Abrahamson MJ, Vázquez-Figueroa E, Woodall NB, Moore JC & Bommarius AS (2012) Development of an Amine Dehydrogenase for Synthesis of Chiral Amines. *Angew Chem Int Ed* 51: 3969–3972

Abrahamson MJ, Wong JW & Bommarius AS (2013) The Evolution of an Amine Dehydrogenase Biocatalyst for the Asymmetric Production of Chiral Amines. *Adv Synth Catal* 355: 1780–1786

Agresti JJ, Antipov E, Abate AR, Ahn K, Rowat AC, Baret J-C, Marquez M, Klibanov AM, Griffiths AD & Weitz DA (2010) Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc Natl Acad Sci* 107: 4004–4009

Aharoni A, Amitai G, Bernath K, Magdassi S & Tawfik DS (2005a) High-Throughput Screening of Enzyme Libraries: Thiolactonases Evolved by Fluorescence-Activated Sorting of Single Cells in Emulsion Compartments. *Chem Biol* 12: 1281–1289

Aharoni A, Gaidukov L, Khersonsky O, Gould SM, Roodveldt C & Tawfik DS (2005b) The 'evolvability' of promiscuous protein functions. *Nat Genet* 37: 73–6

Aird D, Ross MG, Chen W-S, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C & Gnirke A (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* 12: R18

Akiva E, Copp JN, Tokuriki N & Babbitt PC (2017) Evolutionary and molecular foundations of multiple contemporary functions of the nitroreductase superfamily. *Proc Natl Acad Sci*: 201706849

Aleku GA, France SP, Man H, Mangas-Sanchez J, Montgomery SL, Sharma M, Leipold F, Hussain S, Grogan G & Turner NJ (2017) A reductive aminase from Aspergillus oryzae. *Nat Chem* 9: 961–969

Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, Park H, Shapovalov MV, Renfrew PD, Mulligan VK, Kappel K, *et al* (2017) The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J Chem Theory Comput* 13: 3031–3048

Alley EC, Khimulya G, Biswas S, AlQuraishi M & Church GM (2019) Unified rational protein engineering with sequence-based deep representation learning. *Nat Methods* 16: 1315–1322

Araya CL, Fowler DM, Chen W, Muniez I, Kelly JW & Fields S (2012) A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proc Natl Acad Sci* 109: 16858–16863

Arnold FH (1996) Directed evolution: Creating biocatalysts for the future. *Chem Eng Sci* 51: 5091–5102

Arnold FH (2019) Innovation by Evolution: Bringing New Chemistry to Life (Nobel Lecture). *Angew Chem Int Ed* 58: 14420–14426

Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T & Ben-Tal N (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* 44: W344–W350

Atkinson HJ, Morris JH, Ferrin TE & Babbitt PC (2009) Using Sequence Similarity Networks for Visualization of Relationships Across Diverse Protein Superfamilies. *PLOS ONE* 4: e4345

Baier F, Hong N, Yang G, Pabis A, Miton CM, Barrozo A, Carr PD, Kamerlin SC, Jackson CJ & Tokuriki N (2019) Cryptic genetic variation shapes the adaptive evolutionary potential of enzymes. *eLife* 8: e40789

Baret J-C, Miller OJ, Taly V, Ryckelynck M, El-Harrak A, Frenz L, Rick C, Samuels ML, Hutchison JB, Agresti JJ, *et al* (2009) Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity. *Lab Chip* 9: 1850–1858

Beneyton T, Coldren F, Baret J-C, D. Griffiths A & Taly V (2014) CotA laccase: high-throughput manipulation and analysis of recombinant enzyme libraries expressed in E. coli using droplet-based microfluidics. *Analyst* 139: 3314–3323

Beneyton T, Thomas S, Griffiths AD, Nicaud J-M, Drevelle A & Rossignol T (2017) Droplet-based microfluidic high-throughput screening of heterologous enzymes secreted by the yeast Yarrowia lipolytica. *Microb Cell Factories* 16: 18

Beneyton T, Wijaya IPM, Postros P, Najah M, Leblond P, Couvent A, Mayot E, Griffiths AD & Drevelle A (2016) High-throughput screening of filamentous fungi using nanoliter-range droplet-based microfluidics. *Sci Rep* 6: 1–10

Bershtein S, Goldin K & Tawfik DS (2008) Intense Neutral Drifts Yield Robust and Evolvable Consensus Proteins. *J Mol Biol* 379: 1029–1044

Bershtein S, Segal M, Bekerman R, Tokuriki N & Tawfik DS (2006) Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444: 929–932

Besenmatter W, Kast P & Hilvert D (2007) Relative tolerance of mesostable and thermostable protein homologs to extensive mutation. *Proteins Struct Funct Bioinforma* 66: 500–506

Bloom JD & Arnold FH (2009) In the light of directed evolution: Pathways of adaptive protein evolution. *Proc Natl Acad Sci* 106: 9995–10000

Bloom JD, Labthavikul ST, Otey CR & Arnold FH (2006) Protein stability promotes evolvability. *Proc Natl Acad Sci* 103: 5869–5874

Bommarius BR, Schürmann M & Bommarius AS (2014) A novel chimeric amine dehydrogenase shows altered substrate specificity compared to its parent enzymes. *Chem Commun* 50: 14953–14955

Borgström E, Redin D, Lundin S, Berglund E, Andersson AF & Ahmadian A (2015) Phasing of single DNA molecules by massively parallel barcoding. *Nat Commun* 6: 1–6

Bornberg-Bauer E & Albà MM (2013) Dynamics and adaptive benefits of modular protein evolution. *Curr Opin Struct Biol* 23: 459–466

Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC & Robins K (2012) Engineering the third wave of biocatalysis. *Nature* 485: 185–194

Boyer S, Biswas D, Soshee AK, Scaramozzino N, Nizak C & Rivoire O (2016) Hierarchy and extremes in selections from pools of randomized proteins. *Proc Natl Acad Sci* 113: 3482–3487

Breuer M, Ditrich K, Habicher T, Hauer B, Keßeler M, Stürmer R & Zelinski T (2004) Industrial Methods for the Production of Optically Active Intermediates. *Angew Chem Int Ed* 43: 788–824

Bridgham JT, Ortlund EA & Thornton JW (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461: 515

Brunhuber NMW, Thoden JB, Blanchard JS & Vanhooke JL (2000) Rhodococcus l-Phenylalanine Dehydrogenase: Kinetics, Mechanism, and Structural Basis for Catalytic Specifity,. *Biochemistry* 39: 9174–9187

Bryant DH, Bashir A, Sinai S, Jain NK, Ogden PJ, Riley PF, Church GM, Colwell LJ & Kelsic ED (2021) Deep diversification of an AAV capsid protein by machine learning. *Nat Biotechnol*: 1–6

Buller AR, van Roye P, Cahn JKB, Scheele RA, Herger M & Arnold FH (2018) Directed Evolution Mimics Allosteric Activation by Stepwise Tuning of the Conformational Ensemble. *J Am Chem Soc* 140: 7256–7266

Burke CM & Darling AE (2016) A method for high precision sequencing of near full-length 16S rRNA genes on an Illumina MiSeq. *PeerJ* 4: e2492

Calus ST, Ijaz UZ & Pinto AJ (2018) NanoAmpli-Seq: a workflow for amplicon sequencing for mixed microbial communities on the nanopore sequencing platform. *GigaScience* 7: 1–16

Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, Afriat-Jurnou L, Buckle AM, Weik M, Hollfelder F, *et al* (2016) The role of protein dynamics in the evolution of new enzyme function. *Nat Chem Biol* 12: 944–950

Campbell EC, Correy GJ, Mabbitt PD, Buckle AM, Tokuriki N & Jackson CJ (2018) Laboratory evolution of protein conformational dynamics. *Curr Opin Struct Biol* 50: 49–57

Chan YH, Venev SV, Zeldovich KB & Matthews CR (2017) Correlation of fitness landscapes from three orthologous TIM barrels originates from sequence and structure constraints. *Nat Commun* 8: 14614

Chapman S, Faulkner C, Kaiserli E, Garcia-Mata C, Savenkov EI, Roberts AG, Oparka KJ & Christie JM (2008) The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proc Natl Acad Sci* 105: 20038–20043

Chen K & Arnold FH (1993) Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc Natl Acad Sci* 90: 5618–5622

Chen K & Arnold FH (2020) Engineering Cytochrome P450s for Enantioselective Cyclopropenation of Internal Alkynes. *J Am Chem Soc* 142: 6891–6895

Chiasson M & Fowler DM (2019) Mutagenesis-based protein structure determination. *Nat Genet* 51: 1072–1073

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, *et al* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25: 1422–1423

Colin P-Y, Kintses B, Gielen F, Miton CM, Fischer G, Mohamed MF, Hyvönen M, Morgavi DP, Janssen DB & Hollfelder F (2015a) Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat Commun* 6: 10008

Colin P-Y, Zinchenko A & Hollfelder F (2015b) Enzyme engineering in biomimetic compartments. *Curr Opin Struct Biol* 33: 42–51

Copp JN, Akiva E, Babbitt PC & Tokuriki N (2018) Revealing Unexplored Sequence-Function Space Using Sequence Similarity Networks. *Biochemistry* 57: 4651–4662

Courtois F, Olguin LF, Whyte G, Theberge AB, Huck WTS, Hollfelder F & Abell C (2009) Controlling the Retention of Small Molecules in Emulsion Microdroplets for Use in Cell-Based Assays. *Anal Chem* 81: 3008–3016

De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C & Berger B (2018) NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34: 2666–2669

Debon A, Pott M, Obexer R, Green AP, Friedrich L, Griffiths AD & Hilvert D (2019) Ultrahigh-throughput screening enables efficient single-round oxidase remodelling. *Nat Catal* 2: 740–747

Devine PN, Howard RM, Kumar R, Thompson MP, Truppo MD & Turner NJ (2018) Extending the application of biocatalysis to meet the challenges of drug development. *Nat Rev Chem* 2: 409–421

van Dijk EL, Jaszczyszyn Y, Naquin D & Thermes C (2018) The Third Revolution in Sequencing Technology. *Trends Genet* 34: 666–681

Diss G & Lehner B (2018) The genetic landscape of a physical interaction. *eLife* 7: e32472

Duarte JM, Barbier I & Schaerli Y (2017) Bacterial Microcolonies in Gel Beads for High-Throughput Screening of Libraries in Synthetic Biology. *ACS Synth Biol* 6: 1988–1995

Eijsink VGH, Gåseidnes S, Borchert TV & van den Burg B (2005) Directed evolution of enzyme stability. *Biomol Eng* 22: 21–30

Elowitz MB, Levine AJ, Siggia ED & Swain PS (2002) Stochastic Gene Expression in a Single Cell. *Science* 297: 1183–1186

Emond S, Petek M, Kay EJ, Heames B, Devenish SRA, Tokuriki N & Hollfelder F (2020) Accessing unexplored regions of sequence space in directed enzyme evolution via insertion/deletion mutagenesis. *Nat Commun* 11: 3469

Engler C, Kandzia R & Marillonnet S (2008) A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE* 3: e3647

Esvelt KM, Carlson JC & Liu DR (2011) A system for the continuous directed evolution of biomolecules. *Nature* 472: 499–503

Faber MS, Wrenbeck EE, Azouz LR, Steiner PJ & Whitehead TA (2019) Impact of In Vivo Protein Folding Probability on Local Fitness Landscapes. *Mol Biol Evol* 36: 2764–2777

Fallah-Araghi A, Baret J-C, Ryckelynck M & D. Griffiths A (2012) A completely in vitro ultrahigh-throughput droplet-based microfluidic screening system for protein engineering and directed evolution. *Lab Chip* 12: 882–891

Femmer C, Bechtold M, Held M & Panke S (2020) In vivo directed enzyme evolution in nanoliter reactors with antimetabolite selection. *Metab Eng* 59: 15–23

Finch AJ & Kim JR (2018) Thermophilic Proteins as Versatile Scaffolds for Protein Engineering. *Microorganisms* 6: 97

Firnberg E, Labonte JW, Gray JJ & Ostermeier M (2014) A Comprehensive, High-Resolution Map of a Gene's Fitness Landscape. *Mol Biol Evol* 31: 1581–1592

Fischer E (1894) Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte Dtsch Chem Ges* 27: 2985–2993

Fischlechner M, Schaerli Y, Mohamed MF, Patil S, Abell C & Hollfelder F (2014) Evolution of enzyme catalysts caged in biomimetic gel-shell beads. *Nat Chem* 6: 791–796

Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D & Fields S (2010) High-resolution mapping of protein sequence-function relationships. *Nat Methods* 7: 741–746

Fowler DM & Fields S (2014) Deep mutational scanning: a new style of protein science. *Nat Methods* 11: 801–807

Fu L, Niu B, Zhu Z, Wu S & Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28: 3150–3152

García-Nafría J, Watson JF & Greger IH (2016) IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. *Sci Rep* 6: 27459

Ghadessy FJ, Ong JL & Holliger P (2001) Directed evolution of polymerase function by compartmentalized self-replication. *Proc Natl Acad Sci* 98: 4552–4557

Ghislieri D, Green AP, Pontini M, Willies SC, Rowles I, Frank A, Grogan G & Turner NJ (2013) Engineering an Enantioselective Amine Oxidase for the Synthesis of Pharmaceutical Building Blocks and Alkaloid Natural Products. *J Am Chem Soc* 135: 10863–10869

Gielen F, Buryska T, Vliet LV, Butz M, Damborsky J, Prokop Z & Hollfelder F (2015) Interfacing Microwells with Nanoliter Compartments: A Sampler Generating High-Resolution Concentration Gradients for Quantitative Biochemical Analyses in Droplets. *Anal Chem* 87: 624–632

Gielen F, Butz M, Rees EJ, Erdelyi M, Moschetti T, Hyvönen M, Edel JB, Kaminski CF & Hollfelder F (2017) Quantitative Affinity Determination by Fluorescence Anisotropy Measurements of Individual Nanoliter Droplets. *Anal Chem* 89: 1092–1101

Gielen F, Hours R, Emond S, Fischlechner M, Schell U & Hollfelder F (2016) Ultrahigh-throughput–directed enzyme evolution by absorbance-activated droplet sorting (AADS). *Proc Natl Acad Sci* 113: E7383–E7389

Giger L, Caner S, Obexer R, Kast P, Baker D, Ban N & Hilvert D (2013) Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat Chem Biol* 9: 494–498

Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, Downs B, Sukumar S, Sedlazeck FJ & Timp W (2020) Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat Biotechnol* 38: 433–438

Girault M, Kim H, Arakawa H, Matsuura K, Odaka M, Hattori A, Terazono H & Yasuda K (2017) An on-chip imaging droplet-sorting system: a real-time shape recognition method to screen target cells in droplets with single cell resolution. *Sci Rep* 7: 1–10

Giver L, Gershenson A, Freskgard P-O & Arnold FH (1998) Directed evolution of a thermostable esterase. *Proc Natl Acad Sci* 95: 12809–12813

Goldenzweig A, Goldsmith M, Hill SE, Gertman O, Laurino P, Ashani Y, Dym O, Unger T, Albeck S, Prilusky J, *et al* (2016) Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Mol Cell* 63: 337–346

Goldsmith M & Tawfik DS (2017) Enzyme engineering: reaching the maximal catalytic efficiency peak. *Curr Opin Struct Biol* 47: 140–150

Gonzalez CE & Ostermeier M (2019) Pervasive Pairwise Intragenic Epistasis among Sequential Mutations in TEM-1 β-Lactamase. *J Mol Biol* 431: 1981–1992

Gonzalez CE, Roberts P & Ostermeier M (2019) Fitness Effects of Single Amino Acid Insertions and Deletions in TEM-1 β-Lactamase. *J Mol Biol* 431: 2320–2330

Goto H, Kanai Y, Yotsui A, Shimokihara S, Shitara S, Oyobiki R, Fujiwara K, Watanabe T, Einaga Y, Matsumoto Y, *et al* (2020) Microfluidic screening system based on boron-doped diamond electrodes and dielectrophoretic sorting for directed evolution of NAD(P)-dependent oxidoreductases. *Lab Chip* 20: 852–861

Griffiths AD & Tawfik DS (2003) Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *EMBO J* 22: 24–35

Grogan G (2018) Synthesis of chiral amines using redox biocatalysis. *Curr Opin Chem Biol* 43: 15–22

Guerois R, Nielsen JE & Serrano L (2002) Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. *J Mol Biol* 320: 369–387

Gumulya Y, Baek J-M, Wun S-J, Thomson RES, Harris KL, Hunter DJB, Behrendorff JBYH, Kulig J, Zheng S, Wu X, *et al* (2018) Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat Catal* 1: 878–888

Gupta RD, Goldsmith M, Ashani Y, Simo Y, Mullokandov G, Bar H, Ben-David M, Leader H, Margalit R, Silman I, *et al* (2011) Directed evolution of hydrolases for prevention of G-type nerve agent intoxication. *Nat Chem Biol* 7: 120–125

Hammar P, Angermayr SA, Sjostrom SL, van der Meer J, Hellingwerf KJ, Hudson EP & Joensson HN (2015) Single-cell screening of photosynthetic growth and lactate production by cyanobacteria. *Biotechnol Biofuels* 8: 193

Hasan S, Geissler D, Wink K, Hagen A, J. Heiland J & Belder D (2019) Fluorescence lifetime-activated droplet sorting in microfluidic chip systems. *Lab Chip* 19: 403–409

Hassoun M, Rüger J, Kirchberger-Tolstik T, Schie IW, Henkel T, Weber K, Cialla-May D, Krafft C & Popp J (2018) A droplet-based microfluidic chip as a platform for leukemia cell lysate identification using surface-enhanced Raman scattering. *Anal Bioanal Chem* 410: 999–1006

Heinzelman P, Snow CD, Wu I, Nguyen C, Villalobos A, Govindarajan S, Minshull J & Arnold FH (2009) A family of thermostable fungal cellulases created by structure-guided recombination. *Proc Natl Acad Sci* 106: 5610–5615

Hengoju S, Wohlfeil S, Munser AS, Boehme S, Beckert E, Shvydkiv O, Tovar M, Roth M & Rosenbaum MA (2020) Optofluidic detection setup for multi-parametric analysis of microbiological samples in droplets. *Biomicrofluidics* 14: 024109

Hietpas RT, Jensen JD & Bolon DNA (2011) Experimental illumination of a fitness landscape. *Proc Natl Acad Sci* 108: 7896–7901

Höhne M & Bornscheuer UT (2009) Biocatalytic Routes to Optically Active Amines. *ChemCatChem* 1: 42–51

Holland-Moritz DA, Wismer MK, Mann BF, Farasat I, Devine P, Guetschow ED, Mangion I, Welch CJ, Moore JC, Sun S, *et al* (2020) Mass Activated Droplet Sorting (MADS) Enables High-Throughput Screening of Enzymatic Reactions at Nanoliter Scale. *Angew Chem* 132: 4500–4507

Hong LZ, Hong S, Wong HT, Aw PP, Cheng Y, Wilm A, de Sessions PF, Lim SG, Nagarajan N, Hibberd ML, *et al* (2014) BAsE-Seq: a method for obtaining long viral haplotypes from short sequence reads. *Genome Biol* 15: 517

Horvath DG, Braza S, Moore T, Pan CW, Zhu L, Pak OS & Abbyad P (2019) Sorting by interfacial tension (SIFT): Label-free enzyme sorting using droplet microfluidics. *Anal Chim Acta* 1089: 108–114

Hosokawa M, Hoshino Y, Nishikawa Y, Hirose T, Yoon DH, Mori T, Sekiguchi T, Shoji S & Takeyama H (2015) Droplet-based microfluidics for high-throughput screening of a metagenomic library for isolation of microbial enzymes. *Biosens Bioelectron* 67: 379–385

Huang R, Chen H, Upp DM, Lewis JC & Zhang Y-HPJ (2019) A High-Throughput Method for Directed Evolution of NAD(P)+-Dependent Dehydrogenases for the Reduction of Biomimetic Nicotinamide Analogues. *ACS Catal* 9: 11709–11719

Hung S-T, Mukherjee S & Jimenez R (2020) Enrichment of rare events using a multi-parameter high throughput microfluidic droplet sorter. *Lab Chip* 20: 834–843

Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, Lönnerberg P & Linnarsson S (2014) Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods* 11: 163–166

Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, *et al* (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* 36: 338–345

Jensen RA (1976) Enzyme Recruitment in Evolution of New Function. *Annu Rev Microbiol* 30: 409–425

Jenson JM, Xue V, Stretz L, Mandal T, Reich L "Luther" & Keating AE (2018) Peptide design by optimization on a data-parameterized protein interaction landscape. *Proc Natl Acad Sci* 115: E10342–E10351

Jia Z-J, Gao S & Arnold FH (2020) Enzymatic Primary Amination of Benzylic and Allylic C(sp3)–H Bonds. *J Am Chem Soc* 142: 10279–10283

Jochens H & Bornscheuer UT (2010) Natural Diversity to Guide Focused Directed Evolution. *ChemBioChem* 11: 1861–1866

Jones BJ, Lim HY, Huang J & Kazlauskas RJ (2017) Comparison of Five Protein Engineering Strategies for Stabilizing an α/β-Hydrolase. *Biochemistry* 56: 6521–6532

Jose J (2006) Autodisplay: efficient bacterial surface display of recombinant proteins. *Appl Microbiol Biotechnol* 69: 607–614

Jubb HC, Higueruelo AP, Ochoa-Montaño B, Pitt WR, Ascher DB & Blundell TL (2017) Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *J Mol Biol* 429: 365–371

Kaczmarski JA, Mahawaththa MC, Feintuch A, Clifton BE, Adams LA, Goldfarb D, Otting G & Jackson CJ (2020) Altered conformational sampling along an evolutionary trajectory changes the catalytic activity of an enzyme. *Nat Commun* 11: 5945

Kaltenbach M, Jackson CJ, Campbell EC, Hollfelder F & Tokuriki N (2015) Reverse evolution leads to genotypic incompatibility despite functional and active site convergence. *eLife* 4: e06492

Karbaschi M, Shahi P & Abate AR (2017) Rapid, chemical-free breaking of microfluidic emulsions with a hand-held antistatic gun. *Biomicrofluidics* 11: 044107

Karst SM, Ziels RM, Kirkegaard RH, Sørensen EA, McDonald D, Zhu Q, Knight R & Albertsen M (2021) High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nat Methods*: 1–5

Kazlauskas RJ (2005) Enhancing catalytic promiscuity for biocatalysis. *Curr Opin Chem Biol* 9: 195–201

Kebschull JM & Zador AM (2015) Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Res* 43: e143–e143

Keefe AD & Szostak JW (2001) Functional proteins from a random-sequence library. *Nature* 410: 715–718

Khanal A, Yu McLoughlin S, Kershner JP & Copley SD (2015) Differential Effects of a Mutation on the Normal and Promiscuous Activities of Orthologs: Implications for Natural and Directed Evolution. *Mol Biol Evol* 32: 100–108

Khersonsky O, Roodveldt C & Tawfik DS (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr Opin Chem Biol* 10: 498–508

Kille S, Acevedo-Rocha CG, Parra LP, Zhang Z-G, Opperman DJ, Reetz MT & Acevedo JP (2013) Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth Biol* 2: 83–92

Kinde I, Wu J, Papadopoulos N, Kinzler KW & Vogelstein B (2011) Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci* 108: 9530–9535

Kintses B, Hein C, Mohamed MF, Fischlechner M, Courtois F, Lainé C & Hollfelder F (2012) Picoliter Cell Lysate Assays in Microfluidic Droplet Compartments for Directed Enzyme Evolution. *Chem Biol* 19: 1001–1009

Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S & Taipale J (2012) Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods* 9: 72–74

Kleine-Brüggeney H, Vliet LD van, Mulas C, Gielen F, Agley CC, Silva JCR, Smith A, Chalut K & Hollfelder F (2019) Long-Term Perfusion Culture of Monoclonal Embryonic Stem Cells in 3D Hydrogel Beads for Continuous Optical Analysis of Differentiation. *Small* 15: 1804576

Klesmith JR, Bacik J-P, Michalczyk R & Whitehead TA (2015) Comprehensive Sequence-Flux Mapping of a Levoglucosan Utilization Pathway in E. coli. *ACS Synth Biol* 4: 1235–1243

Klesmith JR, Bacik J-P, Wrenbeck EE, Michalczyk R & Whitehead TA (2017) Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. *Proc Natl Acad Sci* 114: 2265–2270

Knaus T, Böhmer W & G. Mutti F (2017) Amine dehydrogenases: efficient biocatalysts for the reductive amination of carbonyl compounds. *Green Chem* 19: 453–463

Knies JL, Cai F & Weinreich DM (2017) Enzyme Efficiency but Not Thermostability Drives Cefotaxime Resistance Evolution in TEM-1 β-Lactamase. *Mol Biol Evol* 34: 1040–1054

Kuleshov V, Xie D, Chen R, Pushkarev D, Ma Z, Blauwkamp T, Kertesz M & Snyder M (2014) Whole-genome haplotyping using long reads and statistical methods. *Nat Biotechnol* 32: 261–266

Larsen AC, Dunn MR, Hatch A, Sau SP, Youngbull C & Chaput JC (2016) A general strategy for expanding polymerase function by droplet microfluidics. *Nat Commun* 7: 1–9

Lee Y, Tawfik DS & Griffiths AD (2002) Investigating the target recognition of DNA cytosine-5 methyltransferase Hha I by library selection using in vitro compartmentalisation. *Nucleic Acids Res* 30: 4937–4944

Lees JG, Dawson NL, Sillitoe I & Orengo CA (2016) Functional innovation from changes in protein domains and their combinations. *Curr Opin Struct Biol* 38: 44–52

Li C, Chng KR, Boey EJH, Ng AHQ, Wilm A & Nagarajan N (2016) INC-Seq: accurate single molecule reads using nanopore sequencing. *GigaScience* 5: 1–11

Li H (2016) Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 32: 2103–2110

Li Y, Drummond DA, Sawayama AM, Snow CD, Bloom JD & Arnold FH (2007) A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. *Nat Biotechnol* 25: 1051–1056

Link DR, Anna SL, Weitz DA & Stone HA (2004) Geometrically Mediated Breakup of Drops in Microfluidic Devices. *Phys Rev Lett* 92: 054503

Liu X, Painter RE, Enesa K, Holmes D, Whyte G, G. Garlisi C, J. Monsma F, Rehak M, F. Craig F & A. Smith C (2016) High-throughput screening of antibiotic-resistant bacteria in picodroplets. *Lab Chip* 16: 1636–1643

Loakes D, Gallego J, Pinheiro VB, Kool ET & Holliger P (2009) Evolving a Polymerase for Hydrophobic Base Analogues. *J Am Chem Soc* 131: 14827–14837

Loman NJ, Quick J & Simpson JT (2015) A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods* 12: 733–735

van Loo B, Heberlein M, Mair P, Zinchenko A, Schüürmann J, Eenink BDG, Holstein JM, Dilkaute C, Jose J, Hollfelder F, *et al* (2019) High-Throughput, Lysis-Free Screening for Sulfatase Activity Using Escherichia coli Autodisplay in Microdroplets. *ACS Synth Biol* 8: 2690–2700

Lundin S, Gruselius J, Nystedt B, Lexow P, Käller M & Lundeberg J (2013) Hierarchical molecular tagging to resolve long continuous sequences by massively parallel sequencing. *Sci Rep* 3: 1–7

Luo X-J, Zhao J, Li C-X, Bai Y-P, Reetz MT, Yu H-L & Xu J-H (2016) Combinatorial evolution of phosphotriesterase toward a robust malathion degrader by hierarchical iteration mutagenesis. *Biotechnol Bioeng* 113: 2350–2357

Ma F, Chung MT, Yao Y, Nidetz R, Lee LM, Liu AP, Feng Y, Kurabayashi K & Yang G-Y (2018) Efficient molecular evolution to generate enantioselective enzymes using a dual-channel microfluidic droplet screening platform. *Nat Commun* 9: 1030

Ma SK, Gruber J, Davis C, Newman L, Gray D, Wang A, Grate J, Huisman GW & Sheldon RA (2010) A green-by-design biocatalytic process for atorvastatin intermediate. *Green Chem* 12: 81–86

Maaten L van der & Hinton G (2008) Visualizing Data using t-SNE. *J Mach Learn Res* 9: 2579–2605

Mahler L, Tovar M, Weber T, Brandes S, Rudolph MM, Ehgartner J, Mayr T, Figge MT, Roth M & Zang E (2015) Enhanced and homogeneous oxygen availability during incubation of microfluidic droplets. *RSC Adv* 5: 101871–101878

Mao Z, Guo F, Xie Y, Zhao Y, Lapsley MI, Wang L, Mai JD, Costanzo F & Huang TJ (2015) Label-Free Measurements of Reaction Kinetics Using a Droplet-Based Optofluidic Device. *J Lab Autom* 20: 17–24

Marshall JR, Yao P, Montgomery SL, Finnigan JD, Thorpe TW, Palmer RB, Mangas-Sanchez J, Duncan RAM, Heath RS, Graham KM, *et al* (2020) Screening and characterization of a diverse panel of metagenomic imine reductases for biocatalytic reductive amination. *Nat Chem*: 1–9

Mastrobattista E, Taly V, Chanudet E, Treacy P, Kelly BT & Griffiths AD (2005) High-Throughput Screening of Enzyme Libraries: In Vitro Evolution of a β-Galactosidase by Fluorescence-Activated Sorting of Double Emulsions. *Chem Biol* 12: 1291–1300

Matsumura S, Kun Á, Ryckelynck M, Coldren F, Szilágyi A, Jossinet F, Rick C, Nghe P, Szathmáry E & Griffiths AD (2016) Transient compartmentalization of RNA replicators prevents extinction due to parasites. *Science* 354: 1293–1296

Mayer S, Rüdiger S, Ang HC, Joerger AC & Fersht AR (2007) Correlation of Levels of Folded Recombinant p53 in Escherichia coli with Thermodynamic Stability in Vitro. *J Mol Biol* 372: 268–276

Maynard Smith J (1970) Natural Selection and the Concept of a Protein Space. *Nature* 225: 563–564

Mayol O, Bastard K, Beloti L, Frese A, Turkenburg JP, Petit J-L, Mariage A, Debard A, Pellouin V, Perret A, *et al* (2019) A family of native amine dehydrogenases for the asymmetric reductive amination of ketones. *Nat Catal*: 1

Mazurenko S, Prokop Z & Damborsky J (2020) Machine Learning in Enzyme Engineering. *ACS Catal* 10: 1210–1223

Mazutis L, Baret J-C & D. Griffiths A (2009) A fast and efficient microfluidic system for highly selective one-to-one droplet fusion. *Lab Chip* 9: 2665–2672

Mazutis L & Griffiths AD (2012) Selective droplet coalescence using microfluidic systems. *Lab Chip* 12: 1800–1806

McCloskey ML, Stöger R, Hansen RS & Laird CD (2007) Encoding PCR Products with Batch-stamps and Barcodes. *Biochem Genet* 45: 761–767

Melamed D, Young DL, Gamble CE, Miller CR & Fields S (2013) Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. *RNA* 19: 1537–1551

Melnikov A, Rogov P, Wang L, Gnirke A & Mikkelsen TS (2014) Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res* 42: e112–e112

Miho E, Roškar R, Greiff V & Reddy ST (2019) Large-scale network analysis reveals the sequence space architecture of antibody repertoires. *Nat Commun* 10: 1321

Miton CM, Buda K & Tokuriki N (2021) Epistasis and intramolecular networks in protein evolution. *Curr Opin Struct Biol* 69: 160–168

Miton CM & Tokuriki N (2016) How mutational epistasis impairs predictability in protein evolution and design. *Protein Sci* 25: 1260–1272

Montgomery SL, Pushpanath A, Heath RS, Marshall JR, Klemstein U, Galman JL, Woodlock D, Bisagni S, Taylor CJ, Mangas-Sanchez J, *et al* (2020) Characterization of imine reductases in reductive amination for the exploration of structure-activity relationships. *Sci Adv* 6: eaay9320

Musil M, Konegger H, Hon J, Bednar D & Damborsky J (2019) Computational Design of Stable and Soluble Biocatalysts. *ACS Catal* 9: 1033–1054

Najah M, Calbrix R, Mahendra-Wijaya IP, Beneyton T, Griffiths AD & Drevelle A (2014) Droplet-Based Microfluidics Platform for Ultra-High-Throughput Bioprospecting of Cellulolytic Microorganisms. *Chem Biol* 21: 1722–1732

Neun S, Zurek PJ, Kaminski TS & Hollfelder F (2020) Ultrahigh throughput screening for enzyme function in droplets. *Methods Enzymol* 643: 317–343

Newton MS, Arcus VL, Gerth ML & Patrick WM (2018) Enzyme evolution: innovation is easy, optimization is complicated. *Curr Opin Struct Biol* 48: 110–116

Nikoomanzar A, Vallejo D & Chaput JC (2019) Elucidating the Determinants of Polymerase Specificity by Microfluidic-Based Deep Mutational Scanning. *ACS Synth Biol* 8: 1421–1429

Nobeli I, Favia AD & Thornton JM (2009) Protein promiscuity and its implications for biotechnology. *Nat Biotechnol* 27: 157–167

Obexer R, Godina A, Garrabou X, Mittl PRE, Baker D, Griffiths AD & Hilvert D (2017) Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat Chem* 9: 50–56

Obexer R, Pott M, Zeymer C, Griffiths AD & Hilvert D (2016) Efficient laboratory evolution of computationally designed enzymes with low starting activities using fluorescence-activated droplet sorting. *Protein Eng Des Sel* 29: 355–366

O'Brien PJ & Herschlag D (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol* 6: R91–R105

Ogden PJ, Kelsic ED, Sinai S & Church GM (2019) Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design. *Science* 366: 1139–1143

Olson CA, Wu NC & Sun R (2014) A Comprehensive Biophysical Description of Pairwise Epistasis throughout an Entire Protein Domain. *Curr Biol* 24: 2643–2651

Ostafe R, Prodanovic R, Lloyd Ung W, Weitz DA & Fischer R (2014) A high-throughput cellulase screening system based on droplet microfluidics. *Biomicrofluidics* 8: 041102

Pääbo S, Irwin DM & Wilson AC (1990) DNA damage promotes jumping between templates during enzymatic amplification. *J Biol Chem* 265: 4718–4721

Paaby AB & Rockman MV (2014) Cryptic genetic variation: evolution's hidden substrate. *Nat Rev Genet* 15: 247–258

Paegel BM & Joyce GF (2010) Microfluidic Compartmentalized Directed Evolution. *Chem Biol* 17: 717–724

Paetzold J & Bäckvall JE (2005) Chemoenzymatic Dynamic Kinetic Resolution of Primary Amines. *J Am Chem Soc* 127: 17620–17621

Palackal N, Brennan Y, Callen WN, Dupree P, Frey G, Goubet F, Hazlewood GP, Healey S, Kang YE, Kretz KA, *et al* (2004) An evolutionary route to xylanase process fitness. *Protein Sci* 13: 494–503

Parmeggiani F, Weise NJ, Ahmed ST & Turner NJ (2018) Synthetic and Therapeutic Applications of Ammonia-lyases and Aminomutases. *Chem Rev* 118: 73–118

Patrick WM, Firth AE & Blackburn JM (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng Des Sel* 16: 451–457

Pavelka A, Chovancova E & Damborsky J (2009) HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Res* 37: W376–W383

Payne EM, Holland-Moritz DA, Sun S & Kennedy RT (2020) High-throughput screening by droplet microfluidics: perspective into key challenges and future prospects. *Lab Chip* 20: 2247–2262

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, *et al* (2011) Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 12: 2825–2830

Peisajovich SG, Rockah L & Tawfik DS (2006) Evolution of new protein topologies through multistep gene rearrangements. *Nat Genet* 38: 168–174

Peisajovich SG & Tawfik DS (2007) Protein engineers turned evolutionists. *Nat Methods* 4: 991–994

Peterman N & Levine E (2016) Sort-seq under the hood: implications of design choices on large-scale characterization of sequence-function relations. *BMC Genomics* 17: 206

Petrie KL & Joyce GF (2014) Limits of Neutral Drift: Lessons From the In Vitro Evolution of Two Ribozymes. *J Mol Evol* 79: 75–90

Pinheiro VB, Taylor AI, Cozens C, Abramov M, Renders M, Zhang S, Chaput JC, Wengel J, Peak-Chew S-Y, McLaughlin SH, *et al* (2012) Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science* 336: 341–344

Podgornaia AI & Laub MT (2015) Pervasive degeneracy and epistasis in a protein-protein interface. *Science* 347: 673–677

Poelwijk FJ, Kiviet DJ, Weinreich DM & Tans SJ (2007) Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445: 383

Porter JL, Rusli RA & Ollis DL (2016) Directed Evolution of Enzymes for Industrial Biocatalysis. *ChemBioChem* 17: 197–203

Potapov V & Ong JL (2017) Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLOS ONE* 12: e0169774

Pushpanath A, Siirola E, Bornadel A, Woodlock D & Schell U (2017) Understanding and Overcoming the Limitations of Bacillus badius and Caldalkalibacillus thermarum

Amine Dehydrogenases for Biocatalytic Reductive Amination. *ACS Catal* 7: 3204–3209

Qu G, Li A, Acevedo-Rocha CG, Sun Z & Reetz MT (2020) The Crucial Role of Methodology Development in Directed Evolution of Selective Enzymes. *Angew Chem Int Ed* 59: 2–30

Raj A & van Oudenaarden A (2008) Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* 135: 216–226

Risso VA, Martinez-Rodriguez S, Candel AM, Krüger DM, Pantoja-Uceda D, Ortega-Muñoz M, Santoyo-Gonzalez F, Gaucher EA, Kamerlin SCL, Bruix M, *et al* (2017) De novo active sites for resurrected Precambrian enzymes. *Nat Commun* 8: ncomms16113

Rodrigues CH, Pires DE & Ascher DB (2018) DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res*

Rollins NJ, Brock KP, Poelwijk FJ, Stiffler MA, Gauthier NP, Sander C & Marks DS (2019) Inferring protein 3D structure from deep mutation scans. *Nat Genet* 51: 1170–1176

Romero PA & Arnold FH (2009) Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* 10: 866–876

Romero PA, Tran TM & Abate AR (2015) Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proc Natl Acad Sci* 112: 7159–7164

Romero-Rivera A, Garcia-Borràs M & Osuna S (2017) Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal* 7: 8524–8532

Rosenberg A & Hirschberg J (2007) V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)* pp 410–420. Prague, Czech Republic: Association for Computational Linguistics

Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, *et al* (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453: 190–195

Russ WP, Figliuzzi M, Stocker C, Barrat-Charlaix P, Socolich M, Kast P, Hilvert D, Monasson R, Cocco S, Weigt M, *et al* (2020) An evolution-based model for designing chorismate mutase enzymes. *Science* 369: 440–445

Ryckelynck M, Baudrey S, Rick C, Marin A, Coldren F, Westhof E & Griffiths AD (2015) Using droplet-based microfluidics to improve the catalytic properties of RNA under multiple-turnover conditions. *RNA* 21: 458–469

Saito Y, Oikawa M, Nakazawa H, Niide T, Kameda T, Tsuda K & Umetsu M (2018) Machine-Learning-Guided Mutagenesis for Directed Evolution of Fluorescent Proteins. *ACS Synth Biol* 7: 2014–2022

Sanchez-Ruiz JM (2010) Protein kinetic stability. *Biophys Chem* 148: 1–15

Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV, Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, *et al* (2016) Local fitness landscape of the green fluorescent protein. *Nature* 533: 397–401

Schmiedel JM & Lehner B (2019) Determining protein structures using deep mutagenesis. *Nat Genet* 51: 1177–1186

Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB & Loeb LA (2012) Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A* 109: 14508–14513

Schmitt S, Walser M, Rehmann M, Oesterle S, Panke S & Held M (2018) Archimedes' principle for characterisation of recombinant whole cell biocatalysts. *Sci Rep* 8: 1–9

Schober M, MacDermaid C, Ollis AA, Chang S, Khan D, Hosford J, Latham J, Ihnken LAF, Brown MJB, Fuerst D, *et al* (2019) Chiral synthesis of LSD1 inhibitor GSK2879552 enabled by directed evolution of an imine reductase. *Nat Catal* 2: 909–915

Schrittwieser JH, Velikogne S & Kroutil W (2015) Biocatalytic Imine Reduction and Reductive Amination of Ketones. *Adv Synth Catal* 357: 1655–1685

Schwartz JJ, Lee C & Shendure J (2012) Accurate gene synthesis with tag-directed retrieval of sequence-verified DNA molecules. *Nat Methods* 9: 913–915

Sciambi A & Abate AR (2014) Generating electric fields in PDMS microfluidic devices with salt water electrodes. *Lab Chip* 14: 2605–2609

Sciambi A & Abate AR (2015) Accurate microfluidic sorting of droplets at 30 kHz. *Lab Chip* 15: 47–51

Sharon E, Kalma Y, Sharp A, Raveh-Sadka T, Levo M, Zeevi D, Keren L, Yakhini Z, Weinberger A & Segal E (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat Biotechnol* 30: 521–530

Siedler S, Khatri NK, Zsohár A, Kjærbølling I, Vogt M, Hammar P, Nielsen CF, Marienhagen J, Sommer MOA & Joensson HN (2017) Development of a Bacterial Biosensor for Rapid Screening of Yeast p-Coumaric Acid Production. *ACS Synth Biol* 6: 1860–1869

Sjostrom SL, Bai Y, Huang M, Liu Z, Nielsen J, Joensson HN & Svahn HA (2014) High-throughput screening for industrial enzyme production hosts by droplet microfluidics. *Lab Chip* 14: 806–813

Skerra A (1994) Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in Escherichia coli. *Gene* 151: 131–135

Stapleton JA, Kim J, Hamilton JP, Wu M, Irber LC, Maddamsetti R, Briney B, Newton L, Burton DR, Brown CT, *et al* (2016) Haplotype-Phased Synthetic Long Reads from Short-Read Sequencing. *PLOS ONE* 11: e0147229

Starita LM, Pruneda JN, Lo RS, Fowler DM, Kim HJ, Hiatt JB, Shendure J, Brzovic PS, Fields S & Klevit RE (2013) Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. *Proc Natl Acad Sci* 110: E1263–E1272

Starr TN, Flynn JM, Mishra P, Bolon DNA & Thornton JW (2018) Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. *Proc Natl Acad Sci* 115: 4453–4458

Starr TN, Picton LK & Thornton JW (2017) Alternative evolutionary histories in the sequence space of an ancient protein. *Nature* 549: 409–413

Starr TN & Thornton JW (2016) Epistasis in protein evolution. *Protein Sci* 25: 1204–1218

Steinberg B & Ostermeier M (2016) Shifting Fitness and Epistatic Landscapes Reflect Trade-offs along an Evolutionary Pathway. *J Mol Biol* 428: 2730–2743

Steinegger M & Söding J (2018) Clustering huge protein sequence sets in linear time. *Nat Commun* 9: 1–8

Stemmer WP (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci* 91: 10747–10751

Stiffler MA, Hekstra DR & Ranganathan R (2015) Evolvability as a Function of Purifying Selection in TEM-1 β-Lactamase. *Cell* 160: 882–892

Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif* 41: 207–234

Tawfik DS & Griffiths AD (1998) Man-made cell-like compartments for molecular evolution. *Nat Biotechnol* 16: 652

Terekhov SS, Smirnov IV, Stepanova AV, Bobik TV, Mokrushina YA, Ponomarenko NA, Belogurov AA, Rubtsova MP, Kartseva OV, Gomzikova MO, *et al* (2017) Microfluidic droplet platform for ultrahigh-throughput single-cell screening of biodiversity. *Proc Natl Acad Sci* 114: 2550–2555

Terpe K (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* 72: 211–222

Thusberg J & Vihinen M (2009) Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. *Hum Mutat* 30: 703–714

Tokuriki N, Jackson CJ, Afriat-Jurnou L, Wyganowski KT, Tang R & Tawfik DS (2012) Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat Commun* 3: ncomms2246

Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN & Tawfik DS (2009) Do viral proteins possess unique biophysical features? *Trends Biochem Sci* 34: 53–59

Tokuriki N, Stricher F, Schymkowitz J, Serrano L & Tawfik DS (2007) The Stability Effects of Protein Mutations Appear to be Universally Distributed. *J Mol Biol* 369: 1318–1332

Tokuriki N, Stricher F, Serrano L & Tawfik DS (2008) How Protein Stability and New Functions Trade Off. *PLOS Comput Biol* 4: e1000002

Tokuriki N & Tawfik DS (2009a) Protein Dynamism and Evolvability. *Science* 324: 203–207

Tokuriki N & Tawfik DS (2009b) Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* 19: 596–604

Tovar M, Mahler L, Buchheim S, Roth M & Rosenbaum MA (2020) Monitoring and external control of pH in microfluidic droplets during microbial culturing. *Microb Cell Factories* 19: 16

Tracewell CA & Arnold FH (2009) Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr Opin Chem Biol* 13: 3–9

Trudeau DL & Tawfik DS (2019) Protein engineers turned evolutionists—the quest for the optimal starting point. *Curr Opin Biotechnol* 60: 46–52

Uchiyama T & Miyazaki K (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* 20: 616–622

Vallejo D, Nikoomanzar A, Paegel BM & Chaput JC (2019) Fluorescence-Activated Droplet Sorting for Single-Cell Directed Evolution. *ACS Synth Biol* 8: 1430–1440

Vaser R, Sovic I, Nagarajan N & Sikic M (2017) Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res*: gr.214270.116

Voigt CA, Martinez C, Wang Z-G, Mayo SL & Arnold FH (2002) Protein building blocks preserved by recombination. *Nat Struct Biol* 9: 553–558

Volden R, Palmer T, Byrne A, Cole C, Schmitz RJ, Green RE & Vollmers C (2018) Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. *Proc Natl Acad Sci* 115: 9726–9731

Walsh C (2001) Enabling the chemistry of life. *Nature* 409: 226–231

Wattenberg M, Viégas F & Johnson I (2016) How to Use t-SNE Effectively. *Distill* 1: e2

Weng L & Spoonamore JE (2019) Droplet Microfluidics-Enabled High-Throughput Screening for Protein Engineering. *Micromachines* 10: 734

Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, *et al* (2012) Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol* 30: 543–548

Wilding M, Hong N, Spence M, Buckle AM & Jackson CJ (2019) Protein engineering: the potential of remote mutations. *Biochem Soc Trans*: BST20180614

Winter G, Griffiths AD, Hawkins RE & Hoogenboom HR (1994) Making Antibodies by Phage Display Technology. *Annu Rev Immunol* 12: 433–455

Woodley JM (2020) New frontiers in biocatalysis for sustainable synthesis. *Curr Opin Green Sustain Chem* 21: 22–26

Wrenbeck EE, Azouz LR & Whitehead TA (2017a) Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nat Commun* 8: ncomms15695

Wrenbeck EE, Faber MS & Whitehead TA (2017b) Deep sequencing methods for protein engineering and design. *Curr Opin Struct Biol* 45: 36–44

Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc Sixth Int Congr Genet* 1: 356–366

Wu NC, Dai L, Olson CA, Lloyd-Smith JO & Sun R (2016) Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* 5: e16965

Wu S, Snajdrova R, Moore JC, Baldenius K & Bornscheuer UT (2020) Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angew Chem Int Ed* n/a

Xia Y & Whitesides GM (1998) Soft Lithography. *Annu Rev Mater Sci* 28: 153–184

Xie D-F, Yang J-X, Lv C-J, Mei J-Q, Wang H-P, Hu S, Zhao W-R, Cao J-R, Tu J-L, Huang J, *et al* (2019) Construction of stabilized (R)-selective amine transaminase from Aspergillus terreus by consensus mutagenesis. *J Biotechnol* 293: 8–16

Xu L, Chen H, Hu X, Zhang R, Zhang Z & Luo ZW (2006) Average Gene Length Is Highly Conserved in Prokaryotes and Eukaryotes and Diverges Only Between the Two Kingdoms. *Mol Biol Evol* 23: 1107–1108

Yang KK, Wu Z & Arnold FH (2019) Machine-learning-guided directed evolution for protein engineering. *Nat Methods* 16: 687–694

Yano T, Oue S & Kagamiyama H (1998) Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc Natl Acad Sci* 95: 5511–5515

Ye LJ, Toh HH, Yang Y, Adams JP, Snajdrova R & Li Z (2015) Engineering of Amine Dehydrogenase for Asymmetric Reductive Amination of Ketone by Evolving Rhodococcus Phenylalanine Dehydrogenase. *ACS Catal* 5: 1119–1122

Yoo JI, Daugherty PS & O'Malley MA (2020) Bridging non-overlapping reads illuminates high-order epistasis between distal protein sites in a GPCR. *Nat Commun* 11: 690

You L & Arnold FH (1996) Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide. *Protein Eng Des Sel* 9: 77–83

Yujian L & Bo L (2007) A Normalized Levenshtein Distance Metric. *IEEE Trans Pattern Anal Mach Intell* 29: 1091–1095

Zagnoni M & Cooper JM (2009) On-chip electrocoalescence of microdroplets as a function of voltage, frequency and droplet size. *Lab Chip* 9: 2652–2658

Zanghellini A (2014) de novo computational enzyme design. *Curr Opin Biotechnol* 29: 132–138

Zeymer C & Hilvert D (2018) Directed Evolution of Protein Catalysts. *Annu Rev Biochem* 87: 131–157

Zhao H, Giver L, Shao Z, Affholter JA & Arnold FH (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol* 16: 258–261

Zheng GXY, Lau BT, Schnall-Levin M, Jarosz M, Bell JM, Hindson CM, Kyriazopoulou-Panagiotopoulou S, Masquelier DA, Merrill L, Terry JM, *et al* (2016) Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat Biotechnol* 34: 303–311

Zheng J, Payne JL & Wagner A (2019) Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. *Science* 365: 347–353

Zinchenko A, Devenish SRA, Kintses B, Colin P-Y, Fischlechner M & Hollfelder F (2014) One in a Million: Flow Cytometric Sorting of Single Cell-Lysate Assays in Monodisperse Picolitre Double Emulsion Droplets for Directed Evolution. *Anal Chem* 86: 2526–2533

Zurek PJ, Hours R, Schell U, Pushpanath A & Hollfelder F (2021) Growth amplification in ultrahigh-throughput microdroplet screening increases sensitivity of clonal enzyme assays and minimizes phenotypic variation. *Lab Chip* 21: 163–173

Zurek PJ, Knyphausen P, Neufeld K, Pushpanath A & Hollfelder F (2020a) UMI-linked consensus sequencing enables phylogenetic analysis of directed evolution. *Github Fhlab UMIC-Seq*

Zurek PJ, Knyphausen P, Neufeld K, Pushpanath A & Hollfelder F (2020b) UMI-linked nanopore consensus sequencing (UMIC-seq) of highly similar gene variants. *Protoc Exch*

Zurek PJ, Knyphausen P, Neufeld K, Pushpanath A & Hollfelder F (2020c) UMI-linked consensus sequencing enables phylogenetic analysis of directed evolution. *Nat Commun* 11: 6023