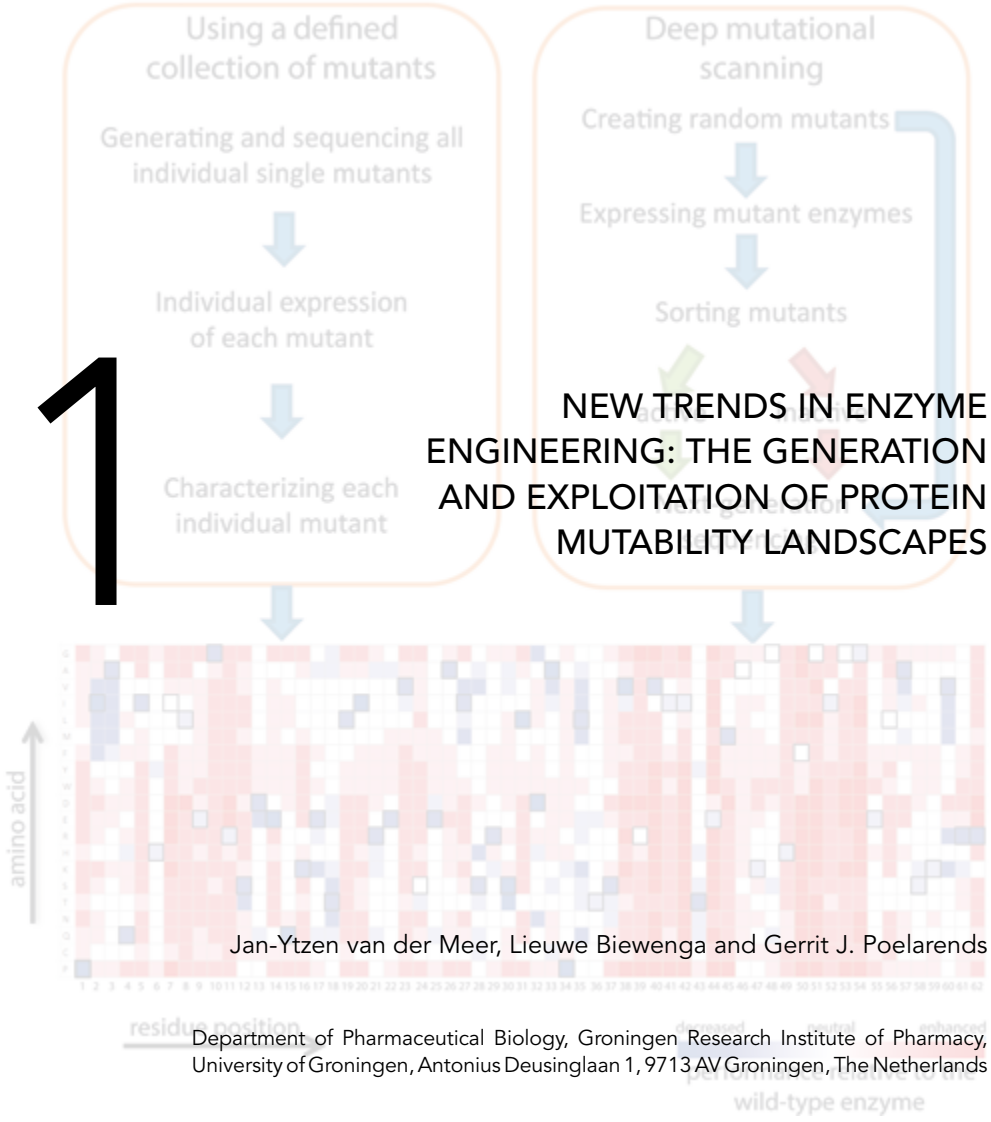
Mutability-landscape guided enzyme engineering

van der Meer, Jan Ytzen



1 NEW TRENDS IN ENZYME ENGINEERING: THE GENERATION AND EXPLOITATION OF PROTEIN MUTABILITY LANDSCAPES

Jan-Ytzen van der Meer, Lieuwe Biewenga and Gerrit J. Poelarends

Department of Pharmaceutical Biology, Groningen Research Institute of Pharmacy, University of Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands

In press: Chembiochem 2016, DOI: 10.1002/cbic.201600382

ABSTRACT

The increasing number of enzyme applications in chemical synthesis calls for new engineering methods to develop the biocatalysts of the future. To efficiently engineer enzymes, the creation of small libraries, which contain a high percentage of positive hits, is highly desirable. Such focused libraries typically only cover ''hotspot'' positions at which mutations are likely to improve the desired property. Predicting "hotspot" positions, however, requires extensive knowledge of sequence-function relationships of the target enzyme. Owing to the technical advances in high-throughput screening and DNA sequencing, it is now possible to rapidly chart entire mutability landscapes by determining the functional consequences of a great number of protein variants. These mutability landscapes allow the important discrimination between beneficial mutations and those that are neutral or detrimental, providing detailed insight into sequence-function relationships. As such, mutability landscapes may have important implications for understanding and engineering enzymes. In this review, we provide an overview of the recent advances in the generation and exploitation of protein mutability landscapes for enzyme engineering.

ADVANTAGES OF ENZYME CATALYSIS

Application of enzymes as catalysts in the production of chemicals has the potential of being a sustainable and efficient alternative to traditional catalysts used in organic synthesis. Enzymes are nature's catalysts and therefore generally function under mild reaction conditions (i.e., ambient temperatures in aqueous solvent systems). Furthermore, enzymes are biodegradable, non-toxic, readily available, and their production is not dependent on any rare elements. These features underline the sustainable potential of using enzymes as catalysts. Another important feature of enzymes is their excellent catalytic properties which can make them highly efficient catalysts. Enzymes are known for their high catalytic rates and excellent regio-, chemo- or stereoselectivity. Enantioselectivity is still a major challenge in traditional catalysis and is highly desirable for the production of pharmaceuticals. Finally, enzymes can be optimized for their application in industrial biocatalysis by means of protein engineering. Owing to these advantages, the number of applications for enzyme catalysts in the production of valuable chemicals, especially pharmaceuticals and agrochemicals, is increasing.[1-3]

WHY IS ENZYME ENGINEERING REQUIRED?

Typical goals of engineering projects in the field of biocatalysis can be divided into three topics. The first topic has a focus on the catalytic properties of enzymes and includes engineering projects which aim to improve catalytic activity, alter substrate scope or improve (enantio-)selectivity. As a result of engineering projects in this theme, there are now many examples of enzymes which carry out industrially relevant transformations, with practical turnover rates.[1-3] The second topic covers enzyme engineering projects which aim to improve enzyme stability. Enzymes can be unstable under process conditions, which may include high temperatures, extreme pH values, high substrate (and product) concentrations and/or the presence of organic solvents. Major improvements in enzyme stability can be achieved using enzyme engineering.[4] Alternatively, solvent engineering or enzyme immobilization can be used to address these stability issues. These methods have recently been reviewed elsewhere.[4-7] The third topic in enzyme engineering is the generation of enzymes which catalyze unnatural chemical transformations. Creating enzymes with new enzymatic activities is currently one of the frontiers in biocatalysis and there are two main approaches to achieve this. Firstly, the *de novo* computational design of enzymes, which involves the computational design of an active site and placing it in a suitable protein scaffold.[8-10] Enzyme engineering is required to improve the activity of the initial *de novo* designed protein to a practical level. The second approach to create new enzymatic activities is to exploit the catalytic promiscuity of existing enzymes. Promiscuous activities are enzymatic activities other than the activity for which an enzyme has evolved and that are not part of the organism's physiology.[11] It has been long recognized that promiscuous activities can serve as a starting point for

natural evolution of new enzymatic functions.[12, 13] Using nature's approach, enzyme engineering can be applied to improve promiscuous activities for the generation of novel biocatalysts for unnatural chemical transformations.[14]

HOTSPOT IDENTIFICATION FOR ENZYME ENGINEERING

Enzyme engineering can be viewed as an iterative procedure which starts with generating diversity in the wild-type (WT) enzyme and screening a collection of mutants for the desired properties. To efficiently engineer enzymes, researchers try to identity hotspot positions in an enzyme where mutations are likely to be beneficial.[15] Targeting these sites for combinatorial mutagenesis leads to relatively small libraries with a high percentage of positive hits. The identification of these hotspots requires extensive knowledge on the sequence-function relationships of an enzyme and the main ways to obtain this information is by analyzing: the (crystal)structure of the enzyme, multi-sequence alignments (MSAs) of homologues proteins or empirical mutational data.

Hotspot identification based on the structure of an enzyme is the most commonly used method in enzyme engineering. Damborski and co-workers recently published an extensive review on *in silico* hotspot identification methods which are available as web tools.[16] The majority of these tools are structure-based and therefore require a crystal structure of the enzyme. The computational tools then identify hotspot positions, based on predicted protein-ligand interactions, binding pockets or residues present in access tunnels of enzymes with a buried active site. Computational tools for the identification of hotspots to improve enzyme stability are mainly based on crystallographic B-factors, although computational protein design and consensus methods are gaining momentum in this area.[4, 16, 17] Besides these *in silico* approaches, several experimental, semi-rational, structure-based enzyme engineering methods have been developed, which apply targeted site-saturation mutagenesis on active site residues. These methods include the highly successful CASTing method and derivatives thereof.[3, 18, 19]

Homology-based hotspot identification tools require a MSA of homologues proteins to identify the evolutionary conservation of specific amino acid residues in a protein. High conservation scores suggest that a specific residue is important for the structure or function of the protein, whereas low conservation suggests that this residue may be mutated without the loss of function. Targeting these positions with mutational robustness, therefore, increases the chance of obtaining viable mutant enzymes and thereby increases the quality of the library.[16]

The third basis on which hotspot identification can be conducted is empirical data. This data can be generated by screening libraries which were created using random mutagenesis methods such as error prone PCR. The hotspots identified in these libraries can be targeted by combinatorial site saturation mutagenesis.[20] The main advantage of this approach is that it does not require extensive prior knowledge of the target enzyme.

Obviously, there are tools available which combine information from all three sources. A successful example of this is the PROSAR method. Here, a collection of enzyme variants which carry multiple mutations per sequence are generated and empirically tested for the desired activity. The initial pool of enzyme variants covers mutations which are selected based on a combination of structural information, analysis of MSAs and random mutagenesis.[21] By statistical analysis of the screening results, the PROSAR software tool then evaluates the contribution of each individual mutation in each enzyme variant with multiple mutations. The identified residue positions with beneficial mutations are used for the subsequent rounds of diversification and screening. This cycle is repeated until the engineering goal is met.

PROTEIN MUTABILITY LANDSCAPES

A new trend in enzyme engineering is the use of mutability landscapes. For this type of analysis, a large number of protein variants are analyzed to determine the effect of each single amino acid substitution on protein function, stability or pathogenicity. In contrast to other systematic mutagenesis approaches such as gene site saturation mutagenesis (GSSM), mutability landscapes do not only provide information on beneficial mutations but also on detrimental and neutral mutations. This gives valuable information on sequence-function relationships by revealing regions in the enzyme with mutational robustness as well as functionally important residues and hotspot positions.

The term 'mutability landscape' was first used by Rost and co-workers, who developed the screening for non-acceptable polymorphisms (SNAP) algorithm to predict the effect of single amino acid substitutions in disease related proteins.[22] The predictions of this SNAP algorithm are based on information from both a MSA and structural features of the protein of interest.[23] Alternatively, the sorting intolerant from tolerant (SIFT) algorithm can be used to make similar predictions based on residue conservation.[24] Both methods predict whether an amino-acid substitution will be neutral or lead to a functional effect but do not distinguish between detrimental or beneficial effects. This is sufficient when merely looking at pathogenicity because both gain-of-function and loss-of-function-mutations can lead to disease. However, it is of limited use when this mutability landscape is generated for enzyme engineering purposes.

Hecht *et al.* argue that the lack of comprehensive experimental mutagenesis data seems a crucial problem for the development of better computational tools and that the generation of such experimental data is constrained by the amount of required resources.[22] Indeed, available data from experimental protein mutability landscapes is scarce and the majority of these available studies cover protein-protein interactions or protein-DNA interactions.[25-27] In the last few years, however, there have been several reports on experimental mutability landscapes of enzymes. Here we present an overview of the recent advances in experimental mutability landscapes of enzymes

to illustrate how these mutability landscapes were generated and used to gain insight in sequence-function relationships or exploited for enzyme engineering.

GENERATING MUTABILITY LANDSCAPES USING A DEFINED COLLECTION OF SINGLE MUTANTS

There are two approaches for generating experimental protein mutability landscapes. The first approach involves the characterization of a defined collection of single mutants and the second approach is called deep mutational scanning (Fig. 1). To construct a defined collection of mutant enzymes, which covers (nearly) all possible single amino acid substitutions of an enzyme, requires significant effort and resources, but the characterization of the mutants can be relatively easy as it does not require any oversampling. Therefore, the screening methods are not limited to a high-throughput assay, giving more flexibility in the design of the assays and providing access to a broader range of analyses (e.g. HPLC, UV-spectroscopy). The following examples of mutability landscapes are generated using this approach.

PROTEASE ACTIVITY AND STABILITY

The usage of 'site evaluation libraries' described in a patent of Estell and Aehle was basically the first example where a mutability landscape of an enzyme was generated and applied in enzyme engineering.[28] The inventors used a defined collection of single mutants of a serine protease from *Cellulomonas* strain 69B4 (ASP), which covered at least 12 variants on each of its 189 residue positions. All members of this collection were screened for protease activity on three substrates (keratin, casein and succinyl-alanine-alanine-proline-phenylalanine-p-nitroanilide), for thermostability and for stability in the presence of 0.06% dodecylbenzenesulfonate (LAS). The performance of each mutant was scored as the apparent change of free energy in the process of interest, relative to WT ASP ($\Delta\Delta G_{app}$).

This value was calculated using the following formula:

$$\Delta\Delta G_{app} = -RT\ln\frac{Pvar}{Pwt}$$

where *Pvar* is the performance value of the variant and *Pwt* is the performance value of WT ASP. Therefore, negative $\Delta\Delta G_{app}$ values indicate improved performance of the variant, relative to WT ASP. The majority (84% - 94%) of the 2851 analyzed single mutants performed worse compared to WT ASP based on activity or stability. Interestingly, 5-10% of the positions in ASP contained mutations that were deleterious for all analyzed properties. As the residues at these positions were also highly conserved among 20 non-redundant homologs of ASP, the authors concluded that these residues are required for the structural fold of the enzyme. Another remarkable

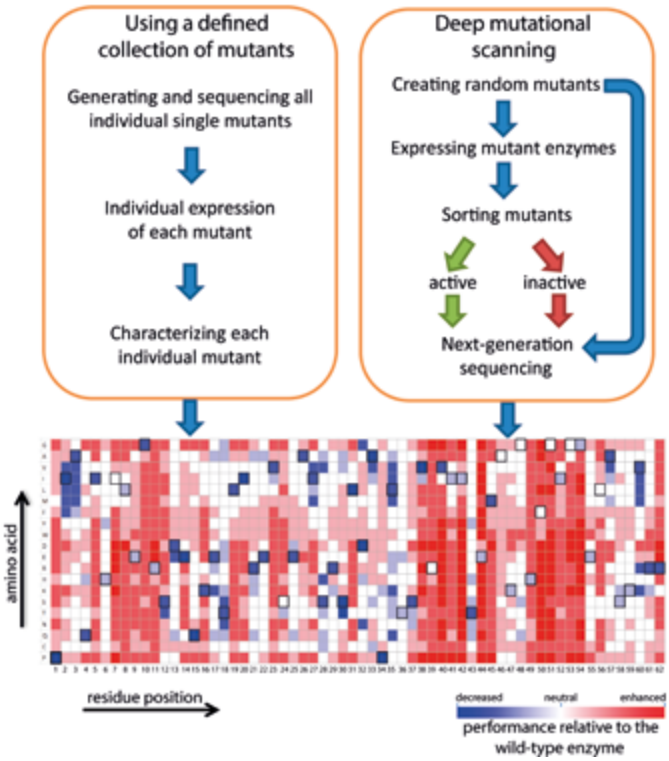


Figure 1 – General methods for generating mutability landscapes.

finding was that most mutations which led to improved protease activity, where outside of the enzyme's active site. For example, the closest residue position at which mutations led to improved protease activity on keratin (Arg-14) was 13 Å away from the catalytic Ser-137. Therefore, targeted saturation mutagenesis on active site residues would most likely not have led to the identification of improved mutants for this reaction.

One unique advantage of this mutability landscape analysis is that it provides information on mutations which lead to the simultaneous improvement of multiple properties. For example, four positions were identified at which mutations led to both improved protease activity towards keratin and improved stability in the presence of LAS. These four positions were simultaneously randomized and the quality of the resulting library was determined based on the performance of 64 randomly picked mutants in both the activity and the LAS stability assay. The average observed performance of these mutants exceeded the expected average performance of the library members, which was calculated based on the assumption of additive effects of single mutations at the four sites. This indicated that information from the mutability landscape of an enzyme can provide valuable guidance for enzyme engineering.

MUTABILITY LANDSCAPES FOR IMPROVED DETERGENT STABILITY

The large α/β-hydrolase fold superfamily includes a broad range of synthetically useful enzymes.[29] Fulton *et al.* generated complete mutability landscapes of *Bacillus subtilus* lipase A (BSLA), which is an α/β-hydrolase fold superfamily member, for stability in the presence of different detergents.[30] Therefore, the authors constructed a defined collection of single mutants, covering each amino acid substitution at each residue position of BSLA. This collection was constructed by performing site saturation mutagenesis at each of the 181 residue positions in BSLA. The resulting 181 libraries were subsequently used to transform *E. coli* cells. From each library, plasmid DNA was isolated from 102 randomly picked colonies and sequenced to determine whether all 19 possible single mutants per residue position were present. Missing single mutants were separately constructed to ensure that the collection of mutants covered all possible 3439 single mutants of BSLA. Subsequently, the residual activity of each mutant was assessed after incubation with varying concentrations of four detergents with different physicochemical properties (i.e. cationic, anionic, zwitterionic and non-ionic). The enzymatic activity of the BSLA mutants was measured using the screening substrate *p*-nitrophenyl butyrate (**1**), which after enzymatic hydrolysis yields *p*-nitrophenol (**2**) which can be detected using UV spectroscopy (Fig. 2a). By plotting the differences in residual activity of each mutant relative to WT BSLA, the authors could identify residue positions at which mutations led to increased tolerance or increased sensitivity towards detergents. By comparing this data to the crystallographic B-factors of BSLA, the authors observed that only two of the five regions in BSLA with high B-factors contained SDS tolerant variants, suggesting that B-factors are not a good predictor for hotspot positions which can be targeted to enhance detergent stability. Additionally, the authors observed that 84% of the hotspots for detergent tolerance were located on surface-exposed sites and that mainly substitutions to aromatic or charged residues, along with cysteine, improved detergent tolerance. This prompted the authors to suggest an optimized mutagenesis strategy using degenerate codons to introduce only those amino acids at solvent exposed sides, for efficiently improving the stability of other (BSLA) α/β-hydrolase fold enzymes.

NEW CATALYTIC FUNCTIONS AND ENANTIOSELECTIVITY

Poelarends and co-workers recently reported the use of mutability landscapes of the promiscuous enzyme 4-oxalocrotonate tautomerase (4-OT) to guide the engineering of novel biocatalysts for Michael-type additions.[31] The enzyme 4-OT is extremely promiscuous and its small monomer size of only 62 residues makes it an ideal template for mutability-landscape guided enzyme engineering.[32] One of 4-OT's promiscuous activities is the Michael-type addition of unmodified aldehydes to nitroalkenes yielding chiral γ-nitroaldehydes, which are valuable precursors for γ-aminobutyric acid (GABA)-based pharmaceuticals.[33-35] To generate the mutability landscapes, a defined collection of 4-OT genes was constructed which encoded at

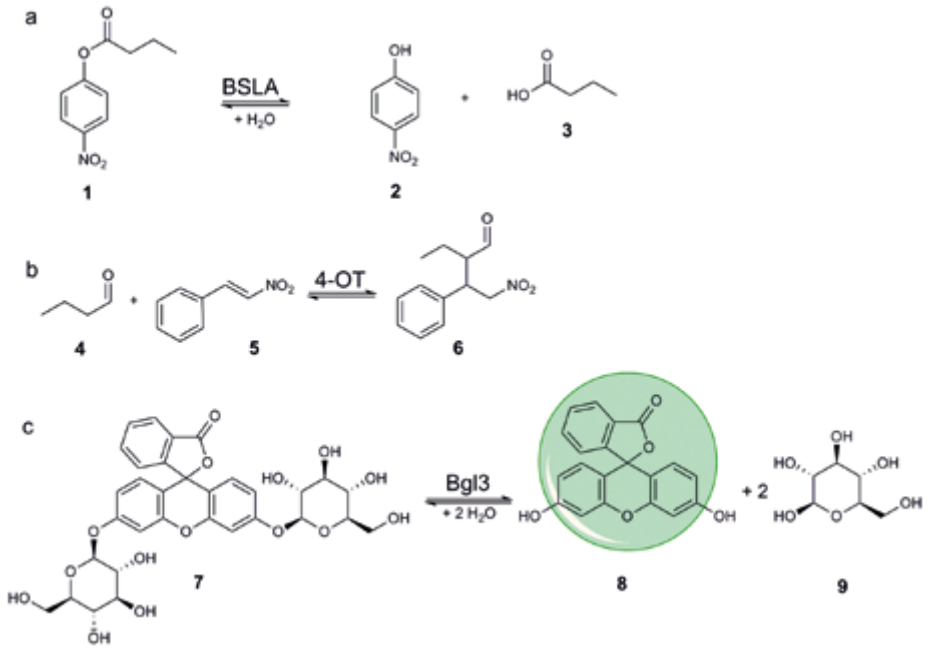


Figure 2 – Screening reactions used to generate mutability landscapes for enzymatic activity. a, The BSLA catalyzed hydrolysis of *p*-nitrophenyl butyrate (1). b, The 4-OT catalyzed Michael-type addition of butanal (4) to β-nitrostyrene (5), yielding chiral γ-nitroaldehyde 6. c, The Bgl3 glycoside-bond cleavage of fluorescein di-(β-D-glucopyranoside) (7), yielding fluorescein (8) and β-D-glucopyranoside (9).

least 15 of the 19 possible variants at each residue position. Each member of this collection was individually characterized for the level of soluble protein expression, tautomerase and 'Michaelase' activities, and enantioselectivity. The level of soluble protein expression was determined for each mutant by using quantitative densitometry on SDS gels. After the 4-OT concentrations in the cell free extracts were quantified, the cell-free extracts were used in the activity and enantioselectivity assessments. All the activities were related to the amount of soluble 4-OT enzyme, yielding the specific activities of each mutant. An overview of the effect of each single mutant on both the tautomerase and 'Michaelase' activity (Fig. 3a), provides insight in the number of neutral amino acid substitutions, essential residues for one or both activities, and beneficial mutations. The positions where mutations lead to improved 'Michaelase' activity (His-6, Ala-33, Met-45 and Phe-50) were simultaneously varied in a focused library, which covered only those amino acid substitutions at each position that improved activity. This led to the identification of a triple mutant (H6M/A33E/F50V), which had an ~15-fold improvement in 'Michaelase' activity.

To screen for enantioselectivity, the authors assayed the enzymatic Michael-type addition of butanal (**4**) to *trans*-β-nitrostyrene (**5**) (Fig. 2b). After following the progress of the reaction with UV-spectroscopy, the reaction mixtures were cleared

by ultrafiltration and directly injected on a RP-HPLC system with a chiral stationary phase. Each single mutant was individually analyzed in this way, which allowed for the determination of both the 'Michaelase' activity and the enantiomeric ratio of the enzymatically produced 2-ethyl-3-phenyl-4-nitrobutanal (**6**) (Fig. 2b). When the activity data is plotted versus the enantioselectivity data (Fig. 3b) it becomes apparent that single amino acid substitutions can have significant effects on improving, inverting or losing the enantioselectivity. In the case of 4-OT, an inversion in enantioselectivity was required to produce precursors for the biologically most active enantiomer of the GABA-analogues. Therefore, the authors made combinations of the single mutants which had the most pronounced inversion in enantioselectivity (H6I, M45Y and F50A) leading to the identification of 4-OT M45Y/F50A which produced the *2S3R*-enantiomer of 2-ethyl-3-phenyl-4-nitrobutanal (**6**) with a e.r. of 96:4. This double mutant also had inverted enantioselectivity in the acetaldehyde addition to various nitroalkenes compared to WT 4-OT, producing the pharmaceutically relevant enantiomers of GABA precursors in enantiomeric ratios up to 97:3. The Michaelase activity of M45Y/F50A was also improved relative to WT 4-OT, which was not surprising because the mutability landscape already indicated that single mutations at these positions led to improved activity (Fig. 3b). Structural analysis of the M45Y/F50A mutant revealed the opening of a hydrophobic pocket in the active site of 4-OT which could accomodate the phenyl group of *trans*-β-nitrostyrene (**5**). It seems likely that this new binding pocket is related to the inverted enantioselectivity of M45Y/F50A. The simultaneous improvement in activity and enantioselectivity underlines the usefulness of mutability landscapes in enzyme engineering.

DEEP MUTATIONAL SCANNING FOR MUTABILITY LANDSCAPES

As mentioned above, it requires significant effort and resources to generate a defined gene collection encoding all single mutants of an enzyme. This bottleneck can be circumvented by using deep mutational scanning. For this, diversity in the WT enzyme is created followed by high-throughput sorting of active mutants from inactive mutants (e.g. by flow cytometry, microfluidics, phage display or growth selection). This allows for the enrichment of active mutants. Conducting next-generation sequencing enables the comparison of the DNA read counts in the sorted library relative to the unsorted (or pre-selected) library (Fig. 1).[36, 37] Using this approach, the enrichment factor (E-factor) of each mutant can be determined and compared to the E-factor of the WT enzyme. A mutability landscape can be generated based on these E-factors, which maps the beneficial, neutral and detrimental effects of (nearly) all single amino acid substitutions of an enzyme. However, to obtain full coverage a large oversampling is required, demanding high throughputs for both the functional sorting and sequencing. Several examples of mutability landscapes using deep mutational scanning to investigate protein-DNA or protein-protein interactions can be found in the literature. [25-27, 36, 37] Recently, the first studies on the generation of mutability landscapes of enzymes by using deep mutational scanning have been published, which are discussed below.

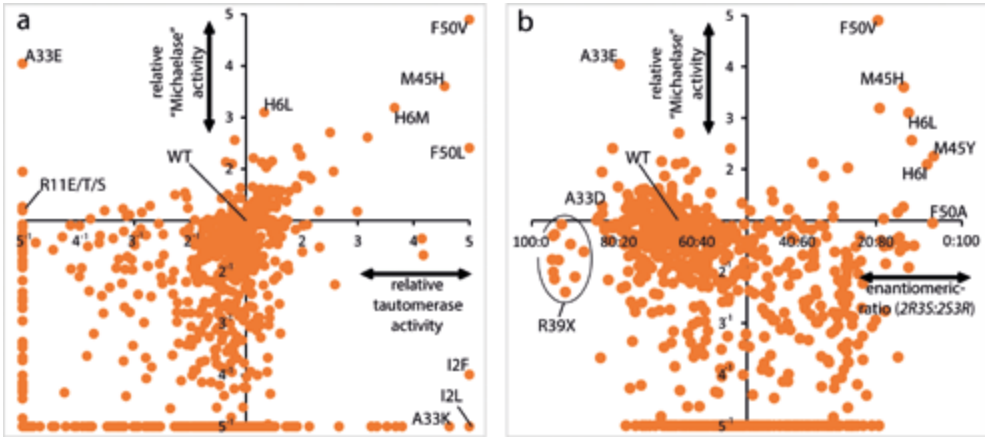


Figure 3 – Mutability landscape data derived from van der Meer *et al.*[31] a, mutational effects on 4-OT's tautomerase activity, plotted versus the mutational effects on 4-OT's promiscuous Michael-type addition activity. b, mutational effects on 4-OT's enantioselectivity in the Michael-type addition reaction, plotted versus the mutational effects on 4-OT's activity in the Michael-type addition reaction.

MUTABILITY LANDSCAPE GENERATION USING MICROFLUIDICS

β-glucosidases are enzymes which cleave β-D-glucosidic bonds by hydrolysis, which can be an important step in the conversion of biomass into fermentable sugars.[38] Romero *et al.* have generated mutability landscapes of a β-glucosidase from *Streptomyces* sp. (Bgl3) using a deep mutational scanning approach in combination with a micro-fluidic based sorting system.[39] For this, they generated a random mutant library of Bgl3 using error-prone PCR with an average of 3.8 mutations per Bgl3 gene. This library was first analyzed using high-throughput sequencing to establish the DNA read counts in the unsorted library. After expressing this library in *E. coli* BL21(DE3), single *E. coli* cells were encapsulated in a micro droplet containing lysing agents and fluorescein di-(β-D-glucopyranoside) (**7**), which is a fluorogenic substrate for Bgl3 (Fig. 2c). Micro droplets containing an active Bgl3 variant were sorted based on fluorescence, using a microfluidic device. This way the authors achieved a throughput of 100 s^{-1}. DNA was retrieved from the sorted micro droplets and sequenced using Illumina sequencing. After analyzing 10^7 variants, the effects of the mutations were determined based on changes in the frequency of occurrence of each mutation before and after the functional sorting. Because of the disadvantage of working with an error-prone library, mainly those amino-acid substitutions which require one nucleotide mutation per codon were accessed in this study. Therefore, only 31% of all possible amino acid substitutions were analyzed. Nevertheless, the generated mutability landscape gave important insights in sequence-function relationships of the enzyme. For example, two essential residues (Lys-461 and Asn-307), which are located outside of the enzyme's active site, were identified in this study. Crystal structure analysis of Bgl3 revealed that Lys-461 is part of a network of salt bridges, which suggests that this residue plays a role in structural stability of the enzyme.

Asn-307 is in hydrogen bonding distance with Glu-178, which is the catalytic acid/base in Bgl3 (Fig. 4). It was therefore suggested that Asn-307 induced a crucial shift in the pK_a of this catalytic residue. Single mutations that improve the thermostability of Bgl3 have been identified in a slightly modified microfluidic screening protocol, which included a heat challenge (65°C for 10 min). Again, 10^7 enzyme variants were analyzed, revealing several single mutants with improved thermostability including mutant S325C. Further characterization of this mutant revealed a 5.3°C increase in T_{50} relative to WT Bgl3.

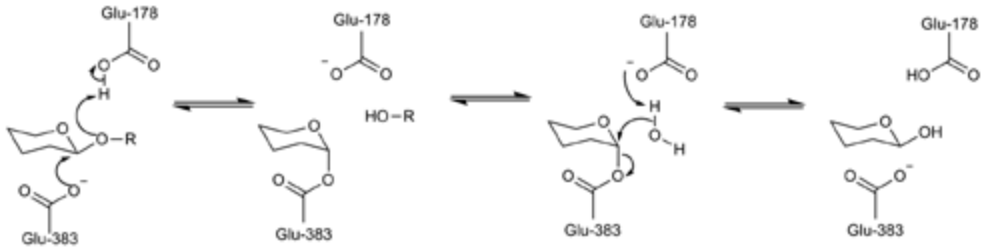


Figure 4 – General mechanism of glucosidic-bond cleavage by Bgl3. Mechanism is derived from Zechel *et al*.[40]

MUTABILITY LANDSCAPE GENERATION USING GROWTH SELECTION

Aminoglycoside-3'-phosphotransferase II (APH(3')II) is a kinase involved in antibiotic resistance that catalyzes the phosphorylation of aminoglycoside antibiotics leading to their inactivation. Melnikov *et al*. performed a single-substitution mutational scan on APH(3')II by analyzing the effect of these mutations on the enzyme's activity and substrate specificity, using kanamycin and five other aminoglycoside antibiotics.[41] For this, the genes coding for each single mutant were individually prepared using a microarray-based DNA synthesis (MITE) approach. All synthesized genes were pooled in equimolar amounts and used to transform *E. coli* cells. These cells were cultured in liquid medium in the presence of aminoglycoside antibiotics, thereby selecting for cells which express an active APH(3')II mutant. After this selection, DNA was isolated from the surviving cells and sequenced using an Illumina sequencing approach to determine the frequency of occurrence of each mutant. By determining the difference in abundance of each mutant before and after selection, the authors could map the effect of all single amino acid substitutions on activity on six aminoglycoside antibiotics. Based on these maps, amino acid substitutions were identified which led to a shift in substrate specificity either towards kanamycin or towards one of the five other tested aminoglycoside antibiotics. By making combinations of these specific amino acid substitutions, the authors engineered five pairs of APH(3')IIs which either favor or disfavor any of the tested antibiotics over kanamycin. For example, Paro+

and Paro- is a pair of APH(3')IIs, engineered to either favor or disfavor paromomycin over kanamycin. Paro+ showed unaltered activity for paromomycin (MIC= 2000-4000 µg/ml) relative to WT APH(3')II but a decreased activity for kanamycin (MIC 31.3 µg/ml). Paro- had a decreased activity for paromomycin (MIC= 62.5 µg/ml) relative to WT APH(3')II but unaltered activity towards kanamycin (MIC= 2000 µg/ml) relative to WT APH(3')II. This remarkable shift in substrate specificity underlines the applicability of mutability landscapes to identify hotspots for enzyme engineering.

MUTABILITY LANDSCAPE GENERATION USING PHAGE DISPLAY

E3-ubiquitin ligases are enzymes which catalyze an ubiquitin transfer from E2-ubiquitin conjugating enzymes to lysines of substrate proteins. This ubiquitination promotes degradation of the substrate protein, which is a crucial process for homeostasis. Ube4b, for example, functions as an E3-ubiquitin ligase, which has been linked to cancer pathogenesis as it ubiquitinates the p53 tumor suppressor *in vivo*.[42] A mutability landscape for the activity of the Ube4b enzyme has been generated and analyzed in order to identify the molecular determinants that modulate the ligase activity of these E3 ligases.[43] A deep-mutational scanning approach was conducted on the U-box domain of Ube4b. This is the active domain of the enzyme which can perform an auto-ubiquitination. Libraries with on average two random nucleotide mutations per gene were generated, sequenced and subsequently displayed on bacteriophages. Bacteriophages which display active (auto-ubiquitinated) U-box domains were then enriched using antibodies against (FLAG)-ubiquitin. Because these antibodies were immobilized on agarose beads, unbound bacteriophages could be washed away (Fig. 5). The enriched bacteriophages were subsequently sequenced using Illumina technology. By comparing the DNA-read counts of each mutation before and after the enrichment, an E-factor was calculated. In this way, 98,289 unique mutant enzymes were characterized, of which 932 single mutants. Mapping the E-factors of these single mutants revealed that some regions (*e.g.* loop 1,2 and helix 1) were less tolerant to mutations than other portions of the U-box domain. Interestingly, several single mutants could be identified from this mutability landscape with improved activity relative to WT. Combining these beneficial single mutations had a synergistic effect and resulted in two double mutants (M1124V/N1142T and D1139N/N1142T) which each have a 22-fold enhanced ubiquitin ligase activity relative to the WT U-box domain. Mechanistic studies on these improved single and double mutants revealed that all beneficial mutations either enhanced the ligase activity by improving binding of the U-box domain to the E2-ubiquitin complex or by improving allosteric activation of the E2-ubiquitin complex. This illustrates that beneficial mutations can be useful both for the generation of superior enzymes and to provide useful insight in enzyme mechanisms.

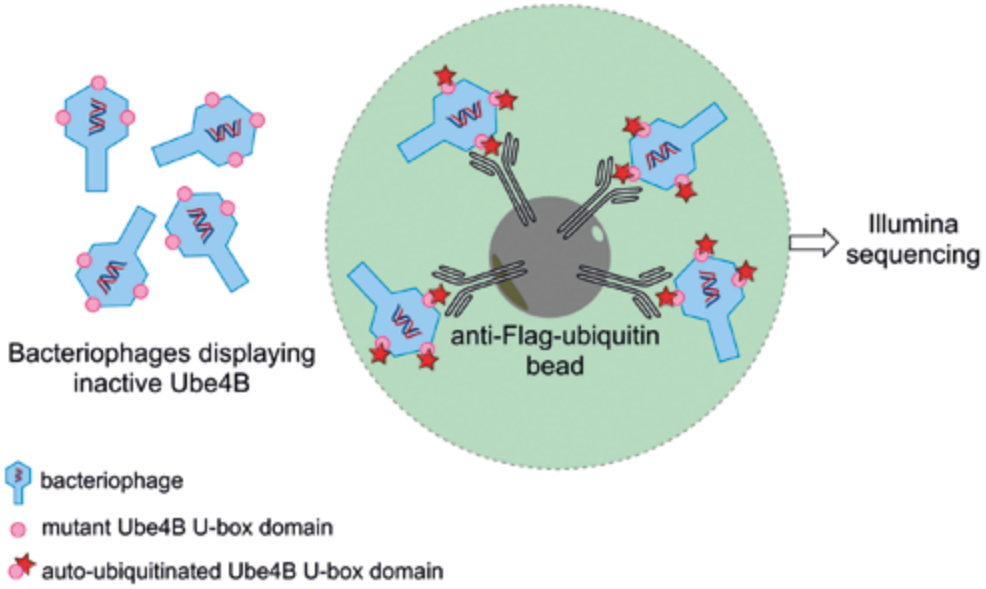


Figure 5 – Enrichment procedure for bacteriophages displaying active (auto-ubiquitinated) U-box domains of Ube4b. Bacteriophages displaying inactive Ube4B U-box domains do not bind to anti-Flag-ubiquitin beads and are washed away. Only the bacteriophages which display active auto-ubiquitinated Ube4B U-box domains bind to the anti-Flag-ubiquitin beads and are sequenced.[43]

CONCLUSIONS AND FUTURE PERSPECTIVES

Currently, most studies on enzyme mutability landscapes have focused on small enzymes (Table 1), which is related to the required costs and effort to generate a mutability landscape. When using a defined collection of single mutants, the bottleneck lays in the generation of this defined mutant gene collection. Currently, PCR-based site-directed mutagenesis techniques are mostly used for the generation of the mutants. Other more recently developed mutagenesis techniques include chemo-enzymatic methods (e.g. SeSaM),[44] micro-array based DNA synthesis (e.g. MIRE)[41] or nonsense suppressor t-RNA methods.[25] The development of these methods might reduce the required amount of effort and costs to generate a defined collection of single mutants. Moreover, because of the ever decreasing costs of commercially available synthetic DNA, the most economical way to obtain a defined collection of single mutants of an enzyme might be DNA synthesis.[45] In the case of deep-mutational scanning the bottleneck for generating mutability landscapes lays in the high-throughput sequencing and high-throughput screening. Both these techniques are rapidly evolving [36, 37, 46] which might facilitate the generation of mutability landscapes using deep mutational scanning.

In conclusion, mutability landscapes are a powerful tool to identify "hotspots" at any place in the amino acid sequence of an enzyme. These "hotspots" can be used as targets for combinatorial mutagenesis to yield superior enzymes with improved

Table 1 – Available studies on experimental mutability landscape analyses of enzymes.

Type of enzyme	Defined mutant collection	Deep mutational scanning	Investigated enzymatic property[a]	Used for hotspot identification[b]	Size of enzyme	Ref.
Protease	X		A, S, SS	X	189	28
Lipase	X		A, S		181	30
Tautomerase / 'Michaelase'	X		A, E, ES, SS	X	62	31
Glucosidase		X	A, S		500	39
Kinase		X	A, SS	X	263	41
Ligase		X	A	X	102[c]	43

[a] S, stability; A, activity; E, expression; ES, enantioselectivity; SS, substrate specificity. [b] The box is checked when combinatorial mutagenesis was conducted on hotspots which were identified in the mutability landscape. [c] Only the active U-box domain of Ube4b was analyzed.

catalytic properties, stability or even new enzymatic activities. The generation of mutability landscapes for several properties of one enzyme (for example, stability and activity or activity and enantioselectivity) provides the unique opportunity to select mutations, which are beneficial for either one or both these properties. Furthermore, mutability landscapes can be used to advance our understanding of sequence-function relationships in enzymes since they provide systematic information on neutral, beneficial and detrimental amino acid substitutions. Both detrimental and beneficial mutations can be extremely helpful to elucidate enzyme mechanisms. Owing to these advantages, combined with the technical advances in high-throughput screening and DNA sequencing, we expect that mutability landscape analysis will become accessible for larger enzymes, and more commonly used for enzyme engineering in the coming years.

ACKNOWLEDGEMENTS

The authors acknowledge funding from the Division of Earth and Life Sciences of the Netherlands Organisation of Scientific Research (ALW grant 820.02.021), the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement n° 242293, and the European Union's Horizon 2020 research and innovation programme under grant agreement No 635595.

REFERENCES

1. Bornscheuer, U. T. *et al*. Engineering the third wave of biocatalysis. *Nature* **485**, 185-194 (2012).
2. Nestl, B. M., Hammer, S. C., Nebel, B. A. & Hauer, B. New generation of biocatalysts for organic synthesis. *Angew. Chem. Int. Ed. Engl.* **53**, 3070-3095 (2014).
3. Reetz, M. T. Laboratory Evolution of Stereoselective Enzymes: A prolific source of catalysts for asymmetric reactions. *Angew. Chem. Int. Ed.* **50**, 138-174 (2011).
4. Bommarius, A. S. & Paye, M. F. Stabilizing biocatalysts. *Chem. Soc. Rev.* **42**, 6534-6565 (2013).
5. Homaei, A. A., Sariri, R., Vianello, F. & Stevanato, R. Enzyme immobilization: an update. *J. of Chem. Biol.* **6**, 185-205 (2013).
6. Illanes, A., Cauerhff, A., Wilson, L. & Castro, G. R. Recent trends in biocatalysis engineering. *Bioresour. Technol.* **115**, 48-57 (2012).
7. Sheldon, R. A. & van Pelt, S. Enzyme immobilisation in biocatalysis: why, what and how. *Chem. Soc. Rev.* **42**, 6223-6235 (2013).
8. Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D. & Houk, K. N. Computational enzyme design. *Angew. Chem. Int. Ed. Engl.* **52**, 5700-5725 (2013).
9. Kries, H., Blomberg, R. & Hilvert, D. De novo enzymes by computational design. *Curr. Opin. Chem. Biol.* **17**, 221-228 (2013).
10. Zanghellini, A. De novo computational enzyme design. *Curr. Opin. Biotechnol.* **29**, 132-138 (2014).
11. Tawfik,Olga Khersonsky and Dan S. Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **79**, 471-505 (2010).
12. Jensen, R., A. Enzyme recruitment in evolution of new function. *Ann. Rev. Microbiol.* **30**, 409-425 (1976).
13. O'Brien, P.,J. & Herschlag, D. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **6**, R91-R105 (1999).
14. Renata, H., Wang, Z. J. & Arnold, F. H. Expanding the enzyme universe: Accessing non-natural reactions by mechanism-guided directed evolution. *Angew. Chem. Int. Ed. Engl.* **54**, 3351-3367 (2015).
15. Turner, N. J. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **5**, 567-573 (2009).
16. Sebestova, E., Bendl, J., Brezovsky, J. & Damborsky, J. in *Methods in Molecular Biology - Directed Evolution Library Creation* (eds Gillam, E. M. J., Copp, J. N. & Ackerley, D.) 291-314 (Springer New York, 2014).
17. Wijma, H. J., Floor, R. J. & Janssen, D. B. Structure-and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Curr. Opin. Struct. Biol.* **23**, 588-594 (2013).
18. Reetz, M. T., Wang, L. & Bocola, M. Directed evolution of enantioselective enzymes: Iterative cycles of CASTing for probing protein-sequence space. *Angew. Chem. Int. ed. Engl.* **45**, 1236-1241 (2006).
19. Reetz, M. T., Kahakeaw, D. & Sanchis, J. Shedding light on the efficacy of laboratory evolution based on iterative saturation mutagenesis. *Molecular BioSystems* **5**, 115-122 (2009).
20. Chica, R. A., Doucet, N. & Pelletier, J. N. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* **16**, 378-384 (2005).
21. Fox, R. J. *et al*. Improving catalytic function by ProSAR-driven enzyme evolution. *Nat Biotechnol.* **25**, 338-344 (2007).
22. Hecht, M., Bromberg, Y. & Rost, B. News from the protein mutability landscape. *J. Mol. Biol.* **425**, 3937-3948 (2013).

23. Hecht, M., Bromberg, Y. & Rost, B. Better prediction of functional effects for sequence variants. *BMC Genomics* **16 Suppl 8**, S1-2164-16-S8-S1. Epub 2015 Jun 18 (2015).

24. Sim, N. *et al*. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40**, W452-W457 (2012).

25. Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S. & Miller, J. H. Genetic studies of the lac repressor. XIV. Analysis of 4000 altered Escherichia coli lac repressors reveals essential and non-essential residues, as well as "spacers" which do not require a specific sequence. *J. Mol. Biol.* **240**, 421-433 (1994).

26. Whitehead, T. A. *et al*. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol.* **30**, 543-548 (2012).

27. McLaughlin, R. N.,Jr, Poelwijk, F. J., Raman, A., Gosal, W. S. & Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* **491**, 138-142 (2012).

28. Estell, D., A. & Aehle, W. Systematic evaluation of sequence and activity relationships using site evaluation libraries for engineering multiple properties. **12/778,915** (2011).

29. Jochens, H. *et al*. Protein engineering of α/β-hydrolase fold enzymes. *ChemBioChem.* **12**, 1508-1517 (2011).

30. Fulton, A. *et al*. Exploring the Protein Stability Landscape: Bacillus subtilis lipase A as a model for detergent tolerance. *ChemBioChem.* **16**, 930-936 (2015).

31. van der Meer, J. Y. *et al*. Using mutability landscapes of a promiscuous tautomerase to guide the engineering of enantioselective Michaelases. *Nat. Commun.* 7:10911 doi: 10.1038/ncomms10911 (2016).

32. Baas, B. J., Zandvoort, E., Geertsema, E. M. & Poelarends, G. J. Recent advances in the study of enzyme promiscuity in the tautomerase superfamily. *ChemBioChem.* **14**, 917-926 (2013).

33. Geertsema, E. M. *et al*. Biocatalytic Michael-type additions of acetaldehyde to nitroolefins with the proline-based enzyme 4-oxalocrotonate tautomerase yielding enantioenriched gamma-nitroaldehydes. *Chem. Eur. J.* **19**, 14407-14410 (2013).

34. Miao, Y., Geertsema, E. M., Tepper, P. G., Zandvoort, E. & Poelarends, G. J. Promiscuous catalysis of asymmetric Michael-type additions of linear aldehydes to beta-nitrostyrene by the proline-based enzyme 4-oxalocrotonate tautomerase. *ChemBioChem.* **14**, 191-194 (2013).

35. Zandvoort, E., Geertsema, E. M., Baas, B. J., Quax, W. J. & Poelarends, G. J. Bridging between organocatalysis and biocatalysis: asymmetric addition of acetaldehyde to beta-nitrostyrenes catalyzed by a promiscuous proline-based tautomerase. *Angew. Chem. Int. Ed. Engl.* **51**, 1240-1243 (2012).

36. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat Meth.* **11**, 801-807 (2014).

37. Shin, H. & Cho, B. Rational protein engineering guided by deep mutational scanning. *Int. J. Mol. Sci.* **16**, 23094-23110 (2015).

38. Perez-Pons, J. A. *et al*. A β-glucosidase gene (bgl3) from *Streptomyces sp.* strain QM-B814. *FEBS J.* **223**, 557-565 (1994).

39. Romero, P. A., Tran, T. M. & Abate, A. R. Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proc. Natl. Acad. Sci.* **112**, 7159-7164 (2015).

40. Zechel, D. L. & Withers, S. G. Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.* **33**, 11-18 (2000).

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

42. Wu, H. *et al*. UBE4B promotes Hdm2-mediated degradation of the tumor suppressor p53. *Nat. Med.* **17**, 347-355 (2011).

43. Starita, L. M. *et al*. Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. *Proc. Natl. Acad. Sci.* **110**, E1263-E1272 (2013).

44. Ruff, A.,J., Kardashliev, T., Dennig, A. & Schwaneberg, U. in *Methods in Molecular Biology - Directed Evolution Library Creation* (eds Gillam, E. M. J., Copp, J. N. & Ackerley, D.) 45-68 (Springer New York, 2014).

45. Carlson, R. The changing economics of DNA synthesis. *Nat. Biotechnol.* **27**, 1091-1094 (2009).

46. Wójcik, M., Telzerow, A., Quax, W. J. & Boersma, Y. L. High-throughput screening in protein engineering: Recent advances and future perspectives. *Int. J. Mol. Sci.* **16**, 24918-24945 (2015).