

Technical University of Denmark



## Mutants of *Micromonospora viridifaciens* sialidase have highly variable activities on natural and non-natural substrates

**Jers, Carsten; Guo, Yao; Kepp, Kasper Planeta; Mikkelsen, Jørn Dalgaard**

*Published in:*  
Protein Engineering Design and Selection (Print)

*Link to article, DOI:*  
[10.1093/protein/gzu054](https://doi.org/10.1093/protein/gzu054)

*Publication date:*  
2015

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Jers, C., Guo, Y., Kepp, K. P., & Mikkelsen, J. D. (2015). Mutants of *Micromonospora viridifaciens* sialidase have highly variable activities on natural and non-natural substrates. *Protein Engineering Design and Selection (Print)*, 28(2), 37-44. DOI: 10.1093/protein/gzu054

## DTU Library

Technical Information Center of Denmark

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Original Article

# Mutants of *Micromonospora viridifaciens* sialidase have highly variable activities on natural and non-natural substrates

C. Jers<sup>1,\*</sup>, Y. Guo<sup>1</sup>, K.P. Kepp<sup>2</sup>, and J.D. Mikkelsen<sup>1</sup>

<sup>1</sup>Center for BioProcess Engineering, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 229, 2800 Kgs. Lyngby, Denmark, and <sup>2</sup>Department of Chemistry, Technical University of Denmark, Building 206, 2800 Kgs. Lyngby, Denmark

\*To whom correspondence should be addressed. E-mail: cj@kt.dtu.dk

Edited by Matthew Cordes

Received 15 June 2014; Revised 30 October 2014; Accepted 8 December 2014

## Abstract

This study aimed to improve the hydrolase activity of the well-characterised bacterial sialidase from *Micromonospora viridifaciens*. The enzyme and its mutated versions were produced in *Bacillus subtilis* and secreted to the growth medium. Twenty amino acid positions in or near the active site were subjected to site-saturation mutagenesis and evaluated on the artificial sialidase substrate 2-*O*-(*p*-nitrophenyl)- $\alpha$ -D-*N*-acetylneuraminic acid and on the natural substrate casein glycomacropeptide. A considerably higher fraction of the mutants exhibited increased activity on the artificial substrate compared with the natural one, with the most proficient mutant showing a 13-fold improvement in  $k_{cat}/K_m$ . In contrast, no mutants displayed more than a 2-fold increase in activity on the natural substrate. To gain further insight into this important discrepancy, we analysed the stability of mutants using the PoPMuSiC software, a property that also correlates with the potential for introducing chemical variation, after validating the method with a set of experimental stability estimates. We found a significant correlation between improved hydrolase activity on the artificial substrate and reduced apparent stability. Together with the minor improvement on the natural substrate this shows an important difference between naturally evolved functionality and new laboratory functionality. Our results suggest that when engineering sialidasases and potentially other proteins towards non-natural substrates that are not optimized by natural evolution, major changes in chemical properties are advantageous, and these changes tend to correlate with decreased stability, partly explaining commonly observed trade-offs between stability and proficiency.

**Key words:** *Micromonospora viridifaciens*, mutagenesis, protein evolution, sialidase, stability–activity trade-off

## Introduction

In protein evolution an important characteristic of enzymes (and other proteins) is their ability to evolve to adapt to new functions (Tokuriki and Tawfik, 2009). This encompasses both altered substrate specificity (but same reaction type) and catalysis of new reactions. In fact, it is clear that many proteins beside their main activity can also catalyse other reactions, normally with lower efficiency. These enzymes appear to have arisen from ‘generalist’ ancestors (enzymes with broad

specificity) after adaptation towards high specificity and activity for a given reaction (Khersonsky and Tawfik, 2010).

A prerequisite for obtaining and maintaining functionality is that specific amino acids are present at important positions in the structure. These functionally important residues have been shown to generally reduce protein stability (Tokuriki *et al.*, 2008). Enzyme stability is normally not maintained at a level substantially higher (typically 5–15 kcal/mol) than that required to function in its natural setting (Branden

and Tooze, 1999). Trade-offs between stability and activity are a major issue in protein engineering, and several causes for these trade-offs have been suggested (Tokuriki et al., 2008; Kepp and Dasmeh, 2014). As adaptation towards new functionality often proceeds via destabilizing mutations, an evolutionary trajectory appears to proceed via initial accumulation of stabilizing mutations to allow incorporation of functional, but often destabilizing mutations (Tokuriki et al., 2008). This trade-off between stability and activity has been observed in both laboratory and natural evolution (Nagatani et al., 2007; Tokuriki et al., 2012; Studer et al., 2014).

Laboratory evolution of enzymes has demonstrated that in many cases latent promiscuous activities can be improved with a minimal amount of mutations, often with a relatively low effect on the main activity (Khersonsky et al., 2006). When approaching the optimum, a diminishing return is seen, that is, additional mutations will lead to relatively smaller improvement in proficiency (Tokuriki et al., 2012). This would indicate that improvement of an enzyme with respect to its natural function (for which it has been specialized) would be more difficult and less substantial than improving side activities. However, it should be noted that many natural enzymes appear to be far from their proficiency optimum because of additional constraints on the proteins (Tokuriki et al., 2012).

In the laboratory evolution of enzymes, the construction of smaller libraries is beneficial as this will reduce the amount of screening needed. To do so, semi-rational mutant designs in which select residues are subjected to site-saturation mutagenesis can be useful. As the observations on stability–function trade-off indicate that most functional mutations are destabilizing, this might be a useful parameter for selection of residues for mutagenesis, given that the target enzyme is stable enough to accommodate them.

In this study, a sialidase (EC 3.2.1.18) from *Micromonospora viridifaciens* was used as a model enzyme. Sialidases are enzymes that hydrolyze sialic acids or sialyl groups of glycans and are found in all domains of life. Bacterial sialidases are classified in glycosyl hydrolase family 33 and their structures are of the six-bladed  $\beta$ -propeller type (Crennell et al., 1993). As retaining glycosidases, they employ a double-displacement mechanism and use a Tyr residue as catalytic nucleophile (Watts et al., 2003) while a conserved Asp-Glu pair is involved in general acid/base catalysis (Chan et al., 2012). Interestingly, this enzyme can retain function upon mutation of either of the catalytic residues (D92, E260 and Y370) (Watson et al., 2003; Chan et al., 2012). A conserved arginine triad (R68, R276 and R342) stabilizes the carboxyl group of sialic acid (Crennell et al., 1993). The *M. viridifaciens* sialidase is the best characterized bacterial sialidase. It has a broad substrate specificity hydrolysing both  $\alpha$ -2,3-,  $\alpha$ -2,6- and  $\alpha$ -2,8-glycosidic linkages and exhibits maximum activity at pH 5 (Aisaka et al., 1991). The  $k_{\text{cat}}/K_m$  of the enzyme is  $8.5 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$  for hydrolysis of the artificial substrate 2-O-(*p*-nitrophenyl)- $\alpha$ -D-N-acetylneuraminic acid (pNP-Neu5Ac) and  $0.36 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$  for hydrolysis of the natural substrate 3'-sialyllactose (Watson et al., 2003).

Sialidases have gained significant interest in relation to synthesis of sialoglycans, due to their trans-sialidase activity (Schmidt et al., 2000; Jers et al., 2014). Several studies have demonstrated potential uses of sialidases as therapeutics targeting sialoglycans, most notably in the case of nervous system injuries. Upon injury, inhibitors of axon regeneration, some of which bind to sialylated receptors, accumulate to halt axon outgrowth. In rats, the administration of sialidase improved spinal axon outgrowth and recovery after spinal cord contusive injury (Yang et al., 2006; Mountney et al., 2010). While several protein-engineering efforts have been directed towards improving

trans-sialidase activity (Paris et al., 2005; Jers et al., 2014; Pierdominici-Sottile et al., 2014), to our knowledge no attempts have been made to improve hydrolase activity of sialidases despite their importance in disease and inflammation. With potential applications in clinical settings emerging, we set out to probe the potential for improving the hydrolase activity of the well-characterized sialidase from *M. viridifaciens* by mutagenesis.

In this study, we subjected selected active site-amino acids to site-saturation mutagenesis and tested the mutant libraries on both a natural substrate, casein glycomacropeptide (cGMP), containing sialic acid  $\alpha$ -2,3-linked to galactose and  $\alpha$ -2,6-linked to N-acetylgalactosamine (Saito and Itoh, 1992), but also on the widely used non-natural test substrate pNP-Neu5Ac. This substrate differs from its natural counterparts by having Neu5Ac linked to a *para*-nitrophenyl group opposed to most commonly galactose in natural glycans. When using pNP-Neu5Ac, a high proportion of mutants displayed activities substantially higher than the wild type and overall up to 25-fold higher activity. In contrast, screening on the natural substrate cGMP led to much less frequent and limited improvements.

The improved activities on pNP-Neu5Ac correlated with increased chemical variation and computed stability changes, confirming the hypothesis that the natural substrates resemble an evolved proficiency optimum, whereas activity on the non-natural substrate is associated with major adaptation upon variation in chemical properties. Thus, we conclude that for optimizing a protein towards non-natural substrates not optimized by natural evolution, major changes in chemical properties that tend to correlate with decreased stability are more likely to successfully improve functionality, consistent with a trade-off between stability and proficiency as widely debated in protein engineering (Tokuriki et al., 2008; Dellus-Gur et al., 2013).

## Materials and methods

### Substrates

pNP-Neu5Ac and colominic acid were purchased from Sigma-Aldrich (Steinheim, Germany). 3'-Sialyllactose and 6'-sialyllactose were obtained from Carbosynth (Compton, UK). The commercial cGMP product LACPRODAN CGMP-20 with a sialic acid content of ~9% (w/w) was supplied by Arla Foods (Viby, Denmark).

### Mutant selection and bioinformatics analysis

Pymol v1.3 (Schrödinger) was used to identify amino acids within 10 Å of the sialic acid binding site, based on the crystal structure with PDB code 1EUS (Gaskell et al., 1995). Individual sites were selected based on manual inspection of the structure. Since the number of mutations prevented experimental stability assessment, we analysed the stabilities of the mutants using the PoPMuSiC software (Dehouck et al., 2011) with the structure templates PDB 1EUS, 1EUT, 1EUU and 1EUR. This method has been shown previously to have a correlation coefficient between predicted stability changes and experimental measurements of ~0.7 and to give good accuracy especially for batches of mutants (Worth et al., 2011; Kepp, 2014). PoPMuSiC computes the  $\Delta\Delta G$  (change in the free energy of protein folding) for a mutant based on parameterized, environment-dependent empirical substitution frequencies in orthologues, i.e. it measures the propensity of a given substitution based on empirical substitution patterns. Rare substitutions are thus associated with large variations in chemical properties that are likely to change the protein function more, and these correspondingly tend to exhibit larger  $\Delta\Delta G$ , i.e. tend to be more destabilizing.

### Bacterial strains and growth conditions

*Escherichia coli* NM522 was used for plasmid propagation. *Escherichia coli* was grown in LB medium (10 g/l tryptone, 5 g/l yeast extract and 10 g/l NaCl) shaking at 37°C. *Bacillus subtilis* strains SCK6 (Zhang and Zhang, 2011) and WB800 (Wu *et al.*, 2002) were used for protein synthesis. *Bacillus subtilis* was grown shaking at 37°C in LB medium and C-medium (Martin-Verstraete *et al.*, 1990) supplemented with 5 g/l glucose and 2% LB medium (CG-LB medium). The growth media were supplemented with kanamycin at 25 µg/ml for *E. coli* and 50 µg/ml for *B. subtilis*.

### DNA manipulation and strain construction

The protein sequence of *M. viridifaciens* sialidase (Uniprot ID Q02834) contains a signal sequence constituted by amino acids 1–37 (Sakurada *et al.*, 1992). A synthetic gene encoding the mature protein (38–647) initiated by methionine and with a C-terminal 6xHis-tag was codon optimized for expression in *B. subtilis* and synthesized by DNA2.0 (Menlo Park, CA, USA). The gene fragment was inserted in the vector pDP66K-PME (Øbro *et al.*, 2009) between the NcoI and HindIII restriction sites replacing the PME-encoding gene and placing the sialidase gene in frame with the CGTase signal peptide and under control of the P32 promoter (Penninga *et al.*, 1996). The use of the NcoI restriction site led to the mutation of I38 to glycine. Amino acid numbering will follow that of the wild-type protein. Site-saturated mutagenesis of 20 positions was done by GeneArt (Regensburg, Germany). Two point mutations, I91E and S201L, were introduced in the wild-type gene by overlapping primer mutagenesis using the oligonucleotides displayed in Table I. The PCR products were restricted with NcoI and HindIII and inserted in pDP66K-PME. The 20 libraries of plasmids and the two constructed single mutant-encoding plasmids were used to transform *B. subtilis* SCK6 (Zhang and Zhang, 2011). This strain contains the *comK* gene under control of a xylose-inducible promoter and was transformed with plasmid DNA as described previously (Silva *et al.*, 2013). *Bacillus subtilis* WB800 was transformed with pDP66K containing the gene encoding the wild-type sialidase using a one-step transformation protocol (Jarmer *et al.*, 2002).

### Screening of site-saturation libraries

For each of the site-saturation libraries, screening was done in 96-well microtitre plate format with 90 transformants (4.5-fold coverage) and six wild-type enzymes per plate. *Bacillus subtilis* was transformed with the site-saturation library and wild-type plasmid, and after overnight incubation at 37°C, the transformants were transferred to 125 µl LB medium with kanamycin in a 250-µl round bottom microtiter plate (Nunc), covered with Airpore sheet (Qiagen) and incubated at 37°C with shaking at 250 rpm for 6 h. The preculture step reduced well-to-well variation from growth. From this preculture, 5 µl was used to inoculate 500 µl CG-LB medium with kanamycin in a 2-ml deepwell plate (Eppendorf). The plate was covered with airpore sheet and incubated at 37°C with shaking at 250 rpm for 16 h. At this point, the supernatants containing the secreted enzymes were harvested by centrifugation for 15 min at 5000 and 4°C and kept on ice or stored at –20°C in 10% glycerol until used in the assays.

### Protein production

For larger scale production of selected enzyme variants, 50 ml cultures were grown by inoculating 50 ml LB medium with kanamycin to OD<sub>600</sub> of 0.05 and incubation with shaking at 37°C for 23 h. At this point, the supernatant was harvested by centrifugation for

15 min at 5000 g and 4°C and passed through first a 0.50 and then a 0.20 µm filter. The ~48 ml supernatant was then concentrated to less than 1 ml using a Vivaspin20 10 MWCO concentrator (Sartorius AG), mixed with 10 ml purification buffer (20 mM sodium phosphate, 500 mM NaCl, pH 7.4) with 10 mM imidazole and concentrated to 1.5 ml. The concentrated sample was applied to a His spin trap column (GE Healthcare) and washed once with 600 µl of the same buffer, then four times with 600 µl purification buffer containing 30 mM imidazole and finally eluted in 400 µl purification buffer with 300 mM imidazole. To reduce imidazole concentration to less than 0.5 mM, several rounds of concentration and dilution in storage buffer (20 mM sodium phosphate, 100 mM NaCl and 10% glycerol, pH 7.4) were done using Vivaspin6 10 MWCO concentrator (Sartorius AG). Final volume was adjusted to 500 µl. Protein concentration was estimated using the BCA assay (Thermo Scientific) with BSA as standard.

### Sialidase activity assays

Two different enzyme assays were used for testing sialidase activity: one using the substrate pNP-Neu5Ac and another for testing activity on various natural substrates. The pNP-Neu5Ac hydrolase assay was done in a 50 mM phosphate-citrate buffer at pH 7 with 0.75 mM pNP-Neu5Ac at 30°C using various amounts of enzyme. In the screen, 5 µl of the culture supernatant was used in a 100-µl reaction. Reactions were started by addition of the substrate and monitored continuously in a spectrophotometer at 410 nm absorbance. Reactions were followed for 1 h, and the initial reaction rates were calculated as the slope of the linear portion of the curves and presented in the following as a percentage of the wild-type enzyme activity. For testing the purified enzymes, 1.5 µg/ml His-tagged wild-type enzyme was used, while for the mutant enzymes, a 5–14 times lower concentration was used. To assess thermostability, 60 µg/ml enzyme was incubated at 60°C for 0, 7.5, 15 and 30 min and residual activity was measured as described above. To estimate  $k_{cat}$  and  $K_m$ , substrate velocity curves were generated with purified enzyme in a discontinuous assay. Samples were terminated by addition of Na<sub>2</sub>CO<sub>3</sub> to a final concentration of 0.25 M. The values of  $k_{cat}$  and  $K_m$  were estimated by non-linear regression using the software Prism 6 (GraphPad software, inc.).

For measurement of hydrolysis of natural substrates, reactions were performed with either 1 mM 3'-SL, 1 mM 6'-SL or 1 mM cGMP-bound sialic acid in 50 mM phosphate-citrate buffer (pH 5 or 7) using 10 µl of 25-fold diluted culture supernatant in 50 µl reactions or 5 µg/ml purified enzyme. Reactions were started by addition of enzyme and stopped by adding H<sub>2</sub>SO<sub>4</sub> to a final concentration of 45 mM. Quantification of free sialic acid was done using a modified thiobarbituric acid assay (Jers *et al.*, 2014).

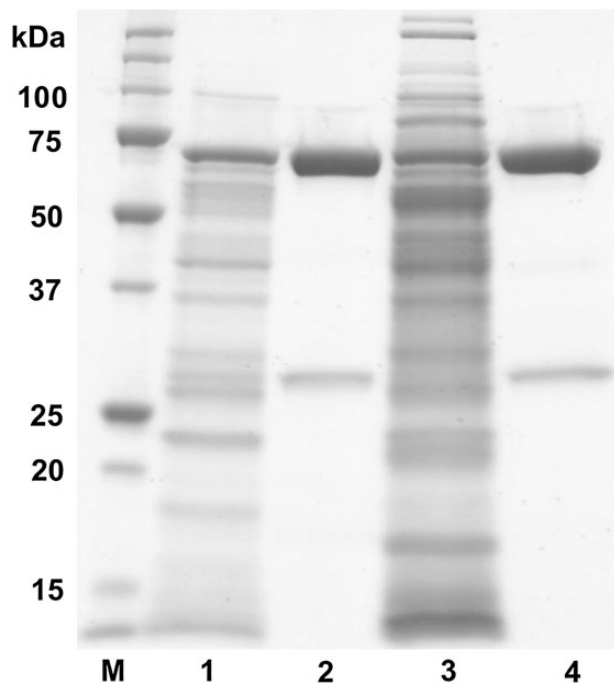
## Results

### Cloning, expression and purification of *M. viridifaciens* sialidase in *B. subtilis*

The sialidase from *M. viridifaciens* was previously produced in bacterial hosts *Streptomyces lividans* and *E. coli* (Sakurada *et al.*, 1992; Watson *et al.*, 2003). To allow efficient secretion to the culture medium, we attempted to use *B. subtilis* as a production host since this would simplify screening in microtiter plate format. Expression of a codon-optimized gene in *B. subtilis* SCK6 (a strain devoid in two of the major secreted proteases) allowed production and secretion of the 6xHis-tagged sialidase (Fig. 1). Besides the full length protein, a smaller form of ~28 kDa was observed similar to the size of the

**Table I.** List of oligos. Mutated codons are shown in bold. MCS is the multiple cloning site

Name	Sequence	Description
pDP66K fwd	CGAATTCGGTCCTCGGGATATGATAAG	Anneals upstream MCS
pDP66K rev	GCCTTCTGCGTCCGGTCGATCAGGGAGGATATCG	Anneals downstream MCS
Mv I91E Fwd	CCCAACAGGAGAAGATGCACCTGGACCGAACTC	I91E
Mv I91E Rev	CAGGTGCATCTTCTCCTGTGGGCGGCCGTC	I91E
Mv S201L Fwd	TGGATGGCGCTTGAGATTTGCTGCTAGCGGTGAAGG	S203L
Mv S201L Rev	CAGCAAATCTCAAGCGCCATCCAGGATCCGG	S203L



**Fig. 1.** SDS-polyacrylamide gel of *M. viridifaciens* sialidase produced in *B. subtilis*. M is a molecular weight marker, Lanes 1 and 2 shows supernatant and purified sialidase produced in *B. subtilis* strains SCK6 and lanes 3 and 4 shows supernatant and purified sialidase produced in *B. subtilis* WB800.

C-terminal, non-catalytic form reported previously (Sakurada et al., 1992). It was attempted to produce the enzyme in *B. subtilis* WB800, a strain devoid of eight proteases, but this did not reduce degradation (Fig. 1) and hence the strain *B. subtilis* SCK6 was used for the reported work.

### Mutations of active site amino acids

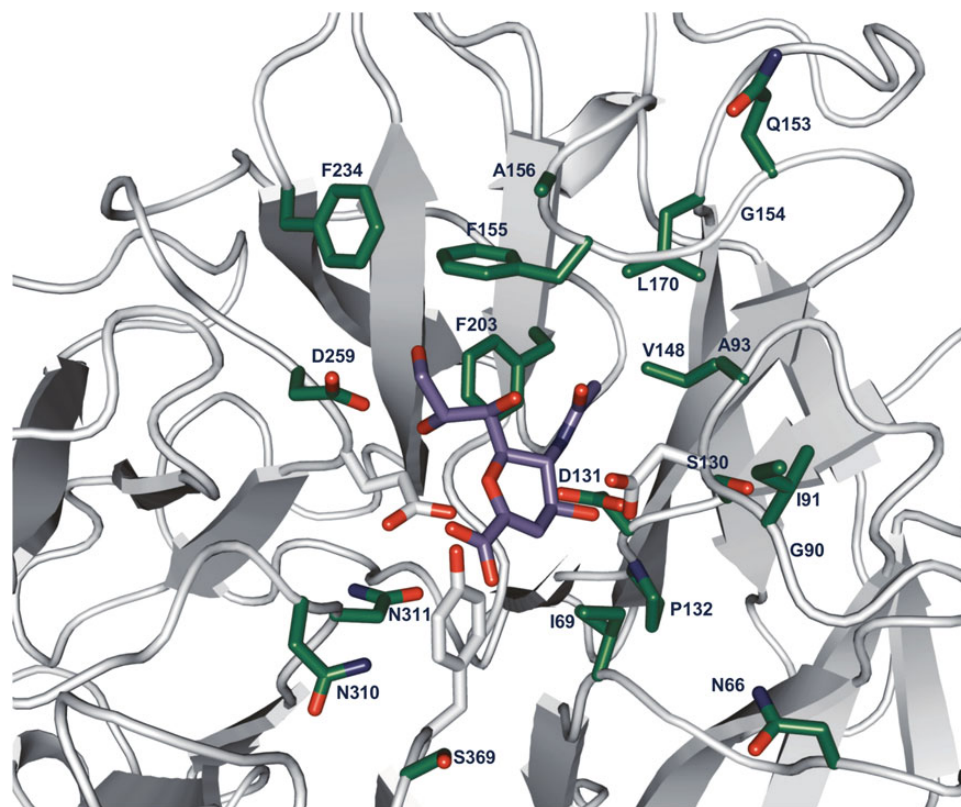
In order to investigate the possibility of improving the catalytic proficiency of sialidases, we mutated a number of positions in or near the active site. To maximize the possibility of identifying relevant mutations, we performed single-amino acid site-saturation mutagenesis in each of 20 selected positions. Positions were selected based on inspection of the solved structure (Gaskell et al., 1995; 1EUS). These included nine sites comprising part of the sialic acid binding pocket: I69, A93, S130, D131, P132, V148, F155, L170, F203 and D259. Of these, D131 and D259 form hydrogen bonds with the inhibitor DANA in the crystal structure and P132 is buried underneath the pocket. Three amino acids N310, N311, and S369 are located close to the catalytic tyrosine, and we envisioned that they might modulate

the covalent intermediate or transition state and thus affect turnover. A number of surface-exposed amino acids were selected because they might change the interaction with the substrate molecules to which sialic acid is bound. These included G90, I91, G154, Q153, A156, and F234. Finally, N66 a second-sphere, surface-exposed residue in close vicinity to conserved R68 was selected. None of the active site amino acids strictly conserved in all sialidases (R68, D92, E260, R276, R342, and Y370) were targeted. The sites selected for mutagenesis are shown in Fig. 2.

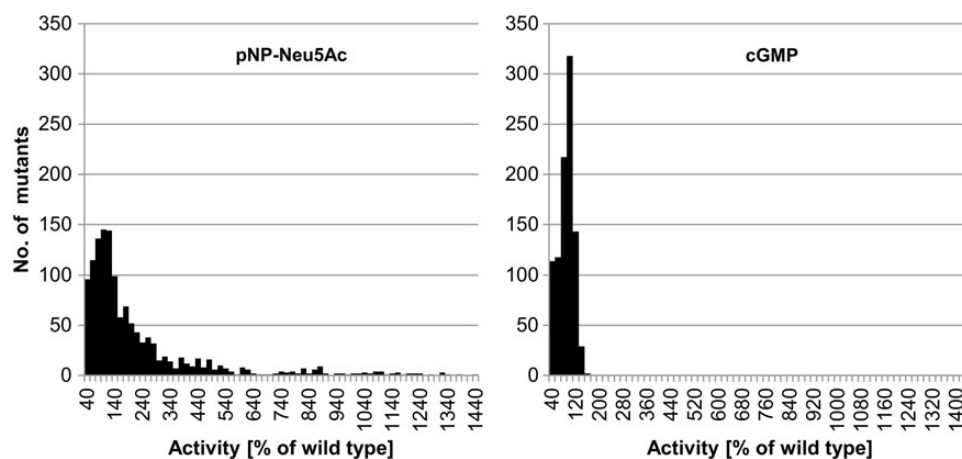
### Screening of site saturation libraries reveals variable activities on pNP-Neu5Ac and cGMP

The mutant libraries were synthesized to contain genes encoding all 20 amino acids in a given position. The DNA was used to transform *B. subtilis* and transformants were picked to a 96-well microtiter plate. Since the secreted sialidase only constituted a small fraction of total secreted protein (Fig. 1), it was not possible to normalize activity with respect to the amount of the sialidase. To compare the mutants, it was therefore important to assure similar growth in all wells. In initial experiments, a relatively high well-to-well variation was observed due to differences in inoculum used. That was alleviated by synchronizing growth by pre-culturing the cells to early stationary phase, and using this culture for inoculating the enzyme production plate. Cells were removed by centrifugation and supernatant containing the secreted enzyme was assayed for enzyme activity on the substrates pNP-Neu5Ac and cGMP. To validate the method, this pipeline was tested using DNA encoding the wild-type enzyme. This was repeated three times, and the activities of individual wells were between 80–125% of the average activity when tested on pNP-Neu5Ac. Next, the libraries were tested on the substrates pNP-Neu5Ac and cGMP. For each amino acid 90 transformants were tested and activities (initial reaction rates) were normalized with respect to the activity of the six replicates of the wild-type enzyme. Since the same enzyme preparation was used, any systematic error related to enzyme concentration would cancel out in the comparison of relative activities on the two substrates. The pooled results for all transformants tested are presented in a histogram (Fig. 3). Results for individual positions are presented in Supplementary Fig. S1. When screened on cGMP, a larger number of mutants showed no or low activity than on pNP-Neu5Ac. Further, the majority of active mutants exhibited wild-type-level activity on cGMP and very few mutants had an increased activity on this substrate. No mutants had an activity >160%. When tested on the artificial substrate pNP-Neu5Ac, the majority of mutants also exhibited wild-type-level activity but the number of mutants exhibiting increased activity was considerably higher, and the most active mutants displayed more than 10-fold increase in activity under the conditions used.

Considering the individual positions mutated, it was apparent that irrespective of the substrate, changing the amino acid D131 was



**Fig. 2.** Amino acid positions in the active site selected for mutagenesis. Selected amino acids are shown in green, catalytic residues D92, E260, and Y370 are in grey while the structure-bound ligand is in blue.



**Fig. 3.** Screening of mutant libraries. Histogram presenting the activity (% compared with wild type) of 1800 mutants tested on (A) pNP-Neu5Ac and (B) cGMP with a bin size of 20%. 475 and 859 mutants showed no or low activity (0–20% wild-type activity) on pNP-Neu5Ac and cGMP respectively and are not shown in the histograms.

accompanied by loss of activity in the majority of substitutions. The amino acids D259, on the other hand, appeared to be essential for hydrolysis of cGMP but not pNP-Neu5Ac. Mutation of the two positions I69 and F203 also led to significant decrease in the hydrolysis of cGMP, whereas on pNP-Neu5Ac, most mutations of F203 strongly improved activity. The largest fraction of mutants improved towards pNP-Neu5Ac (79 of 90 mutants) was observed for F155, whereas most mutations in this position led to a reduction in activity on cGMP.

### Mutations improving the hydrolysis activity on pNP-Neu5Ac

To provide a foundation for interpreting the reasons underlying the difference in improvements observed, we sequenced the variants that showed the largest increase in activity towards pNP-Neu5Ac (Supplementary Table SI). Further, we wanted to verify the screening results on a set of the variants exhibiting the largest increase in activity on pNP-Neu5Ac. To this end, plasmids were purified and entire genes were sequenced, and the corresponding strains were grown in larger

scale to allow production and purification of the 6xHis-tagged enzymes. We found that the most proficient mutant in the screen possessed both mutation I91E but also S201L. The corresponding enzymes were produced and tested on pNP-Neu5Ac (Supplementary Figs S2 and 4). This largely confirmed the screening results except it appeared that the relative activities of most improved variants were underestimated in the screen, and mutants with more than 20-fold improvement in activity was observed (Fig. 4). The mutant I91E/S201L exhibited the highest activity with a 27-fold improvement. When testing the individual single mutants, I91E had wild-type like activity whereas S201L could explain most of the improvement (Supplementary Fig. S3).

As indicated in the screen, neither of the mutants showed significant improvement on the natural substrate cGMP. To confirm the generality of this observation, we tested the purified enzymes on three other natural substrates 3'- and 6'-sialyllactose (Neu5Ac  $\alpha$ -2,3- and  $\alpha$ -2,6-linked to galactose respectively), and colominic acid (Neu5Ac polymer linked by  $\alpha$ -2,8 bonds) (Supplementary Fig. S4). The results confirmed the overall observation, with some variants showing slight improvements and others showing significantly reduced activity on these substrates. To allow detection of the chromophore pNP in a continuous assay, the screening using pNP-Neu5Ac was done at pH 7, while the cGMP screen was done at pH 5, the optimum pH for the wild-type enzyme (Aisaka et al., 1991). To assess whether this discrepancy could account for the differences observed between the two substrates, we analysed the purified mutant enzymes on 3'-sialyllactose at pH 5 and 7 (Supplementary Fig. S4). This confirmed that the observed improvements were not due to a change in pH optimum or range.

### Structure analysis and hypothesis for effect of mutations

After having showed that the mutagenesis strategy preferentially yielded mutants with improved activity on pNP-Neu5Ac, we investigated whether the number of improving mutations would correlate with the chemical variation and stability effects in the library. The wild-type enzyme showed no significant loss of activity after 30 min at 50°C (Supplementary Fig. S5). With the enzyme assays being performed at 30°C, only strongly destabilized/non-folding mutants would be affected at these conditions, and since the typical effect on stability of a mutation is  $\sim 1$  kcal/mol (Tokuriki et al., 2008), stability was not likely to affect yield and protein concentration. To assess chemical variation and stability effects, we used PoPMuSiC where the computed stability change  $\Delta\Delta G$  for each mutation provides a measure of the environment-dependent substitution propensity

based on empirical substitution frequencies of orthologues, and these can then be summed for each of the 20 sites. A large positive value of  $\Delta\Delta G$  thus implies both a highly destabilizing mutant and a natural-evolution-wise unlikely substitution. To validate the application of PoPMuSiC, we experimentally determined the half-lives of 17 mutants and observed decreased half-lives for the 17 mutants predicted to have destabilizing mutations (Supplementary Fig. S6).

We then assessed whether there were correlations between these  $\Delta\Delta G$  values for all 20 individual sites and the number of mutants with higher, similar or lower activity compared with the wild-type enzyme. This approach gives a site-specific estimate of the evolvability and a high value for a site implies that most mutations in the site are unlikely to occur in natural evolution and are associated with loss of stability due to major changes in the chemical properties of the protein. We found no correlations between activity on cGMP and the PoPMuSiC results, indicating that the protein is nearly optimal (by evolution) for activity on natural substrates, and any changes in the sites are likely to have small or negative effect on activity irrespective of chemical properties and stability (Supplementary Fig. S7).

In contrast, there was a markedly different and statistically significant correlation between the average PoPMuSiC  $\Delta\Delta G$  of a site and the site-specific activities of the most active mutants on the non-natural substrate pNP-Neu5Ac. This indicates much more room for improvement, and importantly, improved activity is associated with sites that exhibit major chemical changes and decreased stability (Fig. 5).

We finally considered the amino acid positions that yielded the largest increases in activity on pNP-Neu5Ac. For the most improved variants, the mutations in V148, F155, L170, and F203 were all located in the sialic acid-binding pocket within 4–5 Å of the *N*-acetyl group of sialic acid (Fig. 6). The mutation S201L that was functionally important in the most improved variant was located in the second sphere in close vicinity to these four sites. We further determined the  $k_{cat}$  and  $K_m$  of the wild-type sialidase and the most improved mutants I91E/S201L and F155T. Both mutant enzymes had a slightly increased  $K_m$  whereas the improvement observed in the screen was due to an increase in  $k_{cat}$  (Table II and Fig. S8). Sequencing of mutants of the two aspartic acid residues involved in sialic acid binding indicated that D131 only remained active when mutated to glutamic acid. However, this reduced the activity on pNP-Neu5Ac, and no activity was observed on cGMP, indicating its functional importance. For D259, mutation to glutamic acid increased activity towards pNP-Neu5Ac while other mutations led to reduced activity and in this case, we found no mutants active on cGMP.

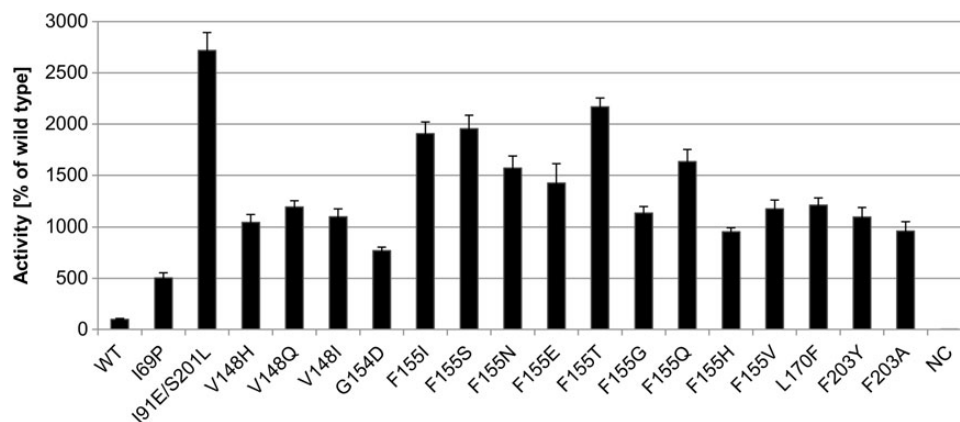
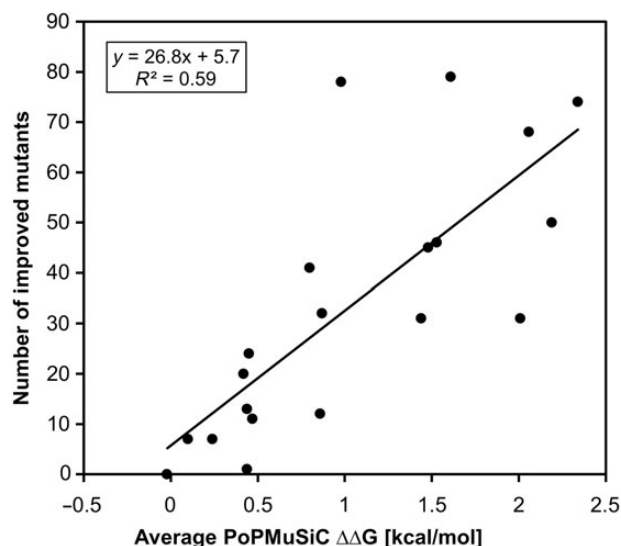
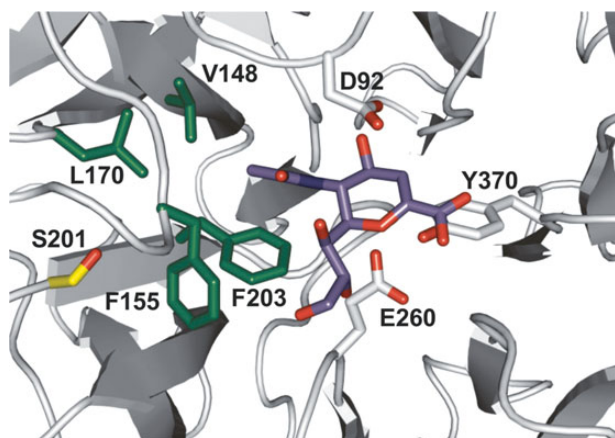


Fig. 4. Activity of the most proficient mutants selected on pNP-Neu5Ac. Activity increments compared with activity of the wild-type sialidase. Negative control (NC) was a reaction without enzyme added. Average of four replicates.



**Fig. 5.** Correlation between potential for chemical variation and improvement towards hydrolysis of pNP-Neu5Ac. PoPMuSiC predictions were used as a measure of potential chemical variation and were for each amino acid position plotted against the number of mutants that exhibited improved activity towards pNP-Neu5Ac.



**Fig. 6.** Representation of the positions with most improved variants. The mutated positions V148, F155, L170 and F203 are shown in green, the site S201 that yielded the most improving mutation is shown in yellow while the catalytic residues D92, E260 and Y370 are shown in grey. The ligand is shown in blue.

**Table II.** Kinetic constants for wild-type and most proficient mutants at pH 7 with pNP-Neu5Ac as substrate

	$k_{\text{cat}}$ [ $\text{s}^{-1}$ ]	$K_{\text{m}}$ [mM]	$k_{\text{cat}}/K_{\text{m}}$ [ $\text{mM}^{-1} \text{s}^{-1}$ ]
Wild type	$0.30 \pm 0.01$	$0.25 \pm 0.03$	1.2
I91E/S201L	$5.4 \pm 0.1$	$0.44 \pm 0.04$	12.3
F155T	$4.2 \pm 0.1$	$0.31 \pm 0.02$	13.6

## Discussion

In this work, we aimed to assess the potential for improvement of a well-characterized sialidase towards both artificial and natural substrates. To this end, we designed, cloned, expressed, and characterized

mutants resulting from saturated mutagenesis at 20 positions near the active site. Positions were selected based on inspection of the solved structure with PDB code 1EUS (Gaskell *et al.*, 1995). None of the strict catalytic residues (R68, D92, E260, R276, R342 and Y370) were selected, but except from that, conservation within the sialidase family was not considered since we wanted to optimize *in vitro* proficiency, and these conserved sites are likely to add additional constraints on the protein relating to, e.g. *in vivo* regulation and modification.

The potential for improving the catalytic efficiency of an enzyme largely depends on how well it is optimized for its particular function. In a previous study, the catalytic efficiency  $k_{\text{cat}}/K_{\text{m}}$  was determined for the substrate pNP-Neu5Ac and a natural substrate 3'-sialyllactose indicating a 10-fold higher efficiency of the enzyme towards the pNP-substrate (Watson *et al.*, 2003). For other classes of enzymes, large rate accelerations have been observed when using artificial pNP-substrates (Indurugalla *et al.*, 2006), reflecting elevated intrinsic reactivity of the substrate rather than the enzyme being optimized for this particular substrate. Consequently, we wanted to investigate if and how the potential for optimizing the enzyme towards non-natural substrates is higher than for natural substrates.

To test this, we assayed the mutant libraries on two substrates, the artificial test substrate pNP-Neu5Ac and the natural substrate cGMP derived from casein. This demonstrated a much higher proportion of beneficial mutations with respect to hydrolysis of the artificial substrate compared with the natural one and the best mutant (I91E/S201L) had a 27-fold improvement in activity whereas for cGMP only slight improvement was observed. This reflects substrate-specific differences and consequently mutations selected specifically to favour action on the artificial substrate. This is in line with observations on promiscuous enzymes, where it is well established that while changing main function can be difficult, a weaker side activity often can be substantially improved with one or few mutations often without significantly affecting the main activity (Kheronsky *et al.*, 2006).

Both in laboratory and natural evolution there is a trade-off between stability and function (Nagatani *et al.*, 2007; Tokuriki *et al.*, 2012; Studer *et al.*, 2014). Since we in this study analysed single mutants only, such effects would not be readily apparent. Destabilizing mutations would be expected to be tolerated by our model enzyme, since it appears reasonably stable (temperature optimum at 52°C) (Watson *et al.*, 2003). When analysing the stability change of mutations, there was a significant correlation between the average PoPMuSiC  $\Delta\Delta G$  and activity on the artificial substrate for the most improved mutants while no correlation with activity on the natural substrate was observed. This would be consistent with the view that the protein is nearly optimal (by evolution) for activity on natural substrates: Any changes are then likely to have small or negative effect on activity regardless of chemical properties and stability, i.e. we have probed a proficiency optimum (in evolutionary terms, resembling a partial fitness optimum, if other selection pressures on the protein are ignored).

In contrast, when optimizing a protein for a new, non-natural substrate, there is much more room for improvement since the substrate is no longer optimal, and importantly, improved activity is associated with sites that exhibit major chemical changes and thus, decreased stability. The observation that the activity towards non-native substrates can be readily improved and that the improvement correlates with the potential for introducing chemical variation in a given amino acid position has direct implications in protein engineering. On the other hand, none of the single mutants were markedly improved towards the



natural substrate. In nature, the enzyme is likely to have been selected for efficient hydrolysis of galactose-linked sialic acid and could potentially, in sequence space, have reached a local optimum. In that case, it might require multi-mutants to enable significant improvements in activity which would not be captured by our strategy (Tracewell and Arnold, 2009).

When analysing the effect in individual positions, the functional importance of especially D141 and D259 that interact directly with sialic acid, was apparent with most mutants showing no activity. For the reaction on pNP-Neu5Ac, mutation to the chemically similar but larger amino acid glutamic acid was tolerated (2-fold reduction and 2.5-fold increase in activity respectively). On cGMP, both mutants had no detectable activity indicating a difference in the position of sialic acid in the binding pocket. Further, it appeared that the largest improvements were observed with amino acids (V148, F155, L170, F203 and S201L) located in the sialic acid-binding pocket in close vicinity to the *N*-acetyl group of sialic acid. This might indicate that positioning of sialic acid in the binding pocket was changed likely due to secondary effects of the nitrophenyl group. This would imply that the major functional role of the mutations would be to allow an alternative positioning of sialic acid that would lead to a more optimal positioning of the nitrophenyl group. For the two most improved mutants, hardly any effect was observed on  $K_m$ , while  $k_{cat}$  was 12–13 times higher (Table II and Supplementary Fig. S8). This increased turnover, however, was substrate-dependent therefore not translating into increased overall catalytic activity of the enzyme as was observed for the cGMP data set.

pNP-Neu5Ac is one of the most frequently used test substrates for sialidases. While the use of artificial test substrates allows easy and fast screening, it is not necessarily indicative of the activity towards the natural substrate (Lee, 2010). This is also apparent in this data set, namely that there was not a clear correlation between activity of mutants on the artificial and natural substrate. The use of pNP-Neu5Ac as an indicator of potential sialidase activity may be reasonable, but with respect to characterization of biological functions of sialidases, more relevant, natural substrate should be chosen.

In conclusion, using a well-studied sialidase as an example, we used site-saturation mutagenesis of 20 sites close to the active site and subsequent activity characterization of mutants to investigate the ability of the enzyme to improve activity towards both non-natural and natural substrates. We show that while there is little room for improvement towards a natural substrate, consistent with the protein being near the proficiency optimum, there is substantial room for producing highly proficient sialidases with activity towards pNP-Neu5Ac. Furthermore, this tendency of improvement correlated with chemical variation introduced in the mutated sites, which again correlates with stability decrease, as computed by the PoPMuSiC algorithm and validated experimentally for selected mutants and the wild type. This study thus provides a mechanistic explanation for stability–activity trade-offs found in protein engineering when the investigated activities are not towards the natural substrate, which is commonly the case.

## Supplementary data

Supplementary data are available at *PEDS* online.

## Funding

This work was supported by the Danish Strategic Research Council (09-067134).

## References

- Aisaka, K., Igarashi, A. and Uwajima, T. (1991) *Agric. Biol. Chem.*, **55**, 997–1004.
- Branden, C. and Tooze, J. (1999) *Introduction to Protein Structure*. Garland, New York.
- Chan, J., Watson, J.N., Lu, A., Cerda, V.C., Borgford, T.J. and Bennet, A.J. (2012) *Biochemistry*, **51**, 433–441.
- Crennell, S.J., Garman, E.F., Laver, W.G., Vimr, E.R. and Taylor, G.L. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 9852–9856.
- Dehouck, Y., Kwasigroch, J.M., Gilis, D. and Rooman, M. (2011) *BMC Bioinformatics*, **12**, 151.
- Dellus-Gur, E., Toth-Petroczy, A., Elias, M. and Tawfik, D.S. (2013) *J. Mol. Biol.*, **425**, 2609–2621.
- Gaskell, A., Crennell, S. and Taylor, G. (1995) *Structure*, **3**, 1197–1205.
- Indurugalla, D., Watson, J.N. and Bennet, A.J. (2006) *Org. Biomol. Chem.*, **4**, 4453–4459.
- Jarmer, H., Berka, R., Knudsen, S. and Saxild, H.H. (2002) *FEMS Microbiol. Lett.*, **206**, 197–200.
- Jers, C., Michalak, M., Larsen, D.M., Kepp, K.P., Li, H., Guo, Y., Kirpekar, F., Meyer, A.S. and Mikkelsen, J.D. (2014) *PLoS One*, **9**, e83902.
- Kepp, K.P. (2014) *J. Phys. Chem. B*, **118**, 1799–1812.
- Kepp, K.P. and Dasmeh, P. (2014) *PLoS ONE*, **9**, e90504.
- Khersonsky, O. and Tawfik, D.S. (2010) *Annu. Rev. Biochem.*, **79**, 471–505.
- Khersonsky, O., Roodveldt, C. and Tawfik, D.S. (2006) *Curr. Opin. Chem. Biol.*, **10**, 498–508.
- Lee, C.C. (2010) *Biofuels*, **1**, 575–588.
- Martin-Verstraete, I., Débarbouillé, M., Klier, A. and Rapoport, G. (1990) *J. Mol. Biol.*, **214**, 657–671.
- Mountney, A., Zahner, M.R., Lorenzini, I., Oudega, M., Schramm, L.P. and Schnaar, R.L. (2010) *PNAS*, **107**, 11561–11566.
- Nagatani, R.A., Gonzalez, A., Shoichet, B.K., Brinen, L.S. and Babbitt, P.C. (2007) *Biochemistry*, **46**, 6688–6695.
- Øbro, J., Sørensen, I., Derckx, P., Madsen, C.T., Drews, M., Willer, M., Mikkelsen, J.D. and Willats, W.G.T. (2009) *Proteomics*, **9**, 1861–1868.
- Paris, G., Ratier, L., Amaya, M.F., Nguyen, T., Alzari, P.M. and Frasch, A.C. (2005) *J. Mol. Biol.*, **345**, 923–934.
- Penning, D., van der Veen, B.A., Knegtel, R.M., van Hijum, S.A., Rozeboom, H. J., Kalk, K.H., Dijkstra, B.W. and Dijkhuizen, L. (1996) *J. Biol. Chem.*, **271**, 32777–32784.
- Pierdominici-Sottile, G., Palma, J. and Roitberg, A.E. (2014) *Proteins*, **82**, 424–435.
- Saito, T. and Itoh, T. (1992) *J. Dairy Sci.*, **75**, 1768–1774.
- Sakurada, K., Ohta, T. and Hasegawa, M. (1992) *J. Bacteriol.*, **174**, 6896–6903.
- Schmidt, D., Sauerbrey, B. and Thiem, J. (2000) *J. Org. Chem.*, **65**, 8518–8526.
- Silva, I.R., Larsen, D.M., Jers, C., Derckx, P., Meyer, A.S. and Mikkelsen, J.D. (2013) *Appl. Microbiol. Biotechnol.*, **97**, 9727–9735.
- Studer, R.A., Christin, P.A., Williams, M.A. and Orengo, C.A. (2014) *Proc. Natl. Acad. Sci. USA*, **111**, 2223–2228.
- Tokuriki, N. and Tawfik, D.S. (2009) *Science*, **324**, 203–207.
- Tokuriki, N., Stricher, F., Serrano, L. and Tawfik, D.S. (2008) *PLoS Comput. Biol.*, **4**, e1000002.
- Tokuriki, N., Jackson, C.J., Afriat-Jurnou, L., Wyganowski, K.T., Tang, R. and Tawfik, D.S. (2012) *Nat. Commun.*, **3**, 1257.
- Tracewell, C.A. and Arnold, F.H. (2009) *Curr. Opin. Chem. Biol.*, **13**, 3–9.
- Watson, J.N., Dookhun, V., Borgford, T.J. and Bennet, A.J. (2003) *Biochemistry*, **42**, 12682–12690.
- Watts, A.G., Damager, I., Amaya, M.L., Buschiazio, A., Alzari, P., Frasch, A.C. and Withers, S.G. (2003) *J. Am. Chem. Soc.*, **125**, 7532–7533.
- Worth, C.L., Preissner, R. and Blundell, T.L. (2011) *Nucleic Acids Res.*, **39**, W215–W222.
- Wu, S.C., Yeung, J.C., Duan, Y., Ye, R., Szarka, S.J., Habibi, H.R. and Wong, S.L. (2002) *Appl. Environ. Microbiol.*, **68**, 3261–3269.
- Yang, L.J.S., Lorenzini, I., Vajn, K., Mountney, A., Schramm, L.P. and Schnaar, R. L. (2006) *PNAS*, **103**, 11057–11062.
- Zhang, X. and Zhang, Y.P. (2011) *Microb. Biotechnol.*, **4**, 98–105.