

Handicap-Recover Evolution Leads to a Chemically Versatile, Nucleophile-Permissive Protease

Thomas Shafee,^[a, b, c] Pietro Gatti-Lafranconi,^[a] Ralph Minter,^[b] and Florian Hollfelder^{*[a]}

Mutation of the tobacco etch virus (TEV) protease nucleophile from cysteine to serine causes an approximately $\sim 10^4$ -fold loss in activity. Ten rounds of directed evolution of the mutant, TEV^{Ser}, overcame the detrimental effects of nucleophile exchange to recover near-wild-type activity in the mutant TEV^{Ser}X. Rather than respecialising TEV to the new nucleophile, all the enzymes along the evolutionary trajectory also retained the ability to use the original cysteine nucleophile. Therefore the adaptive evolution of TEV^{Ser} is paralleled by a neutral trajectory for TEV^{Cys}, in which mutations that increase serine nucleophile reactivity hardly affect the reactivity of cysteine. This apparent nucleophile permissiveness explains how nucleophile switches can occur in the phylogeny of the chymotrypsin-like protease PA superfamily. Despite the changed key component of their chemical mechanisms, the evolved variants TEV^{Ser}X and TEV^{Cys}X have similar activities; this could potentially facilitate escape from adaptive conflict to enable active-site evolution.

Enzymes achieve efficient catalysis through the precise orientation of a key set of active-site residues. This arrangement is dependent on chemical constraints, to the extent that some active-site geometries have convergently evolved many times.^[1] Consequently, active-site residues are the most evolutionarily conserved within enzyme families. Phylogenetic analysis of extended protein superfamilies suggests that even residues that are crucial for activity are exchanged during evolution over sufficiently long timescales.^[2] The evidence for such exchanges raises the question of how a gene coding for an inefficient enzyme can persevere in the transition from one type of active site to another. There is no evolutionary advantage for maintaining a gene coding for a catalytically impaired or inactive protein, thus creating the scenario of “adaptive

conflict”.^[3] We know from studies on enzymes^[4] and enzyme models^[5] that precise positioning is easily disturbed. Minute disturbances down to the picometer scale cause substantial rate reductions.^[6] Given the delicacy of catalytic arrangements, it is unknown how the evolution of active sites avoids unfit variants.

Serine and cysteine proteases are textbook examples of enzymes employing covalent, nucleophilic catalysis (Figure 1A)^[7] that leads to substantial rate acceleration of a difficult reaction (with a half-life of ≈ 500 years).^[8] The sophisticated interplay of the multiple active-site residues involved, including for example, the charge relay system of the catalytic triad,^[7c] suggests that any deviation from such a highly efficient arrangement is likely to be penalised.^[9] However, phylogenetic analysis of proteases suggests that nucleophile exchanges do occur during evolution. The PA clan of chymotrypsin-like proteases^[10] encompasses both serine and cysteine proteases evolved from a hypothetical common ancestor.^[11] Constructing a phylogeny of this clan of proteases (Figure 1B) shows that—within a highly conserved structure—nucleophile switches must have occurred by divergent evolution at least once: cellular proteases use a serine nucleophile, but both cysteine and serine protease families are found in viruses.^[12]

When the active-site nucleophiles of serine or cysteine protease are interconverted, the single atomic change typically leads to a $> 10^4$ -fold reduction in k_{cat}/K_M .^[9a, 14] Although both thiol and hydroxy groups can act as nucleophilic catalysts, their positioning is likely to be suboptimal after mutation due to structural differences between oxygen and sulfur: oxygen's smaller atomic radius (by ~ 0.4 Å)^[14a] and the formation of shorter bonds (decreasing d_{C-X} and d_{X-H} by ~ 1.3 -fold each) would be expected to disturb the precise nucleophile positioning. In addition there are reactivity differences: sulfur is softer, and its different pK_a (4–5 units lower for RSH compared to ROH) provides a larger fraction of the active form of the nucleophile at physiological pH; this explains why the reactivity of the hydroxy side chain of serine that replaces cysteine would be compromised.^[14d]

These considerations of chemical reactivity and structure raise the question of how such nucleophile transitions have occurred in proteases, despite the enzyme inactivation typically associated with mutating a key active-site residue. Handicap-recover experiments can be used to find if any mutations can epistatically offset a known deleterious mutation or the replacement of a native cofactor.^[15] Here we use this approach to demonstrate a scenario that could satisfy the fitness requirements of protein evolution by mutating the crucial nucleophile of tobacco etch virus cysteine protease (TEV^{Cys}) to serine

[a] Dr. T. Shafee, Dr. P. Gatti-Lafranconi, Dr. F. Hollfelder
Department of Biochemistry, University of Cambridge
80 Tennis Court Road, Cambridge, CB2 1GA (UK)
E-mail: fh111@cam.ac.uk

[b] Dr. T. Shafee, Dr. R. Minter
Antibody Discovery and Protein Engineering, MedImmune
Granta Park, Cambridge, CB21 6GH (UK)

[c] Dr. T. Shafee
Present address:
La Trobe Institute of Molecular Sciences, La Trobe University
Science Drive, Melbourne, Victoria 3086 (Australia)

Supporting information for this article is available on the WWW under
<http://dx.doi.org/10.1002/cbic.201500295>.

© 2015 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

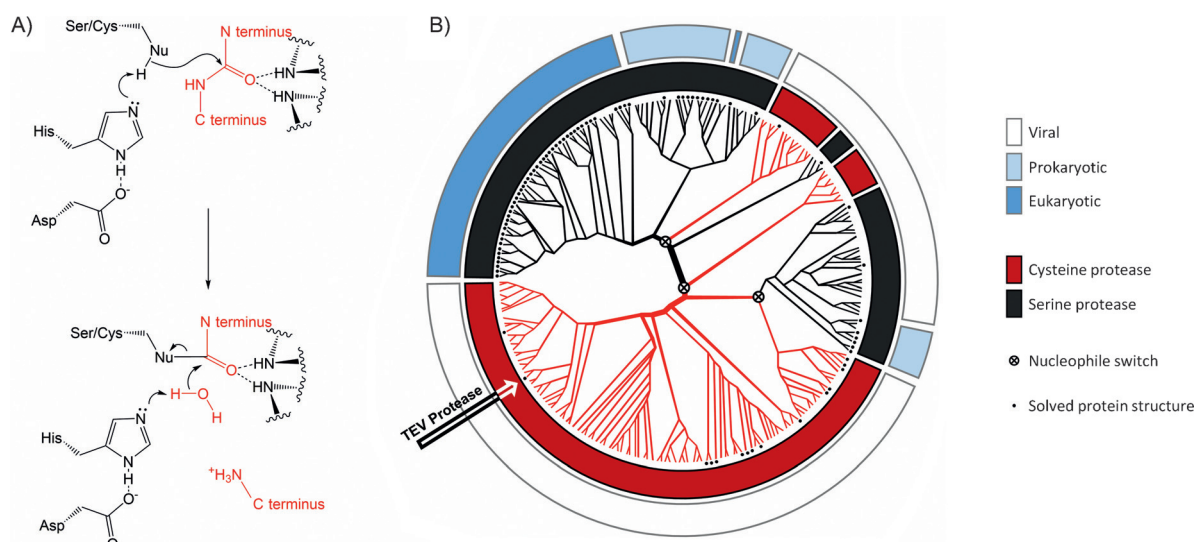


Figure 1. A) Simplified catalytic mechanism of cysteine and serine proteases (Nu = S, e.g., in TEV^{Cys}; or Nu = O, e.g., in TEV^{Ser}).^[7] In the enzyme's catalytic triad (black) aspartate aligns and polarises histidine, which reduces the nucleophile's pK_a and positions it for reaction. The activated nucleophile attacks the carbonyl of the peptide bond of the substrate (red), a tetrahedral intermediate is stabilised by backbone amide hydrogens ("oxanion hole"), and the C terminus of the substrate is ejected, leaving a covalent adduct behind. This acyl-enzyme intermediate (lower panel) is hydrolysed by histidine-activated water to release the N terminus of the substrate (aided by proton transfer to histidine), and this regenerates the free enzyme. B) The phylogeny of the PA clan (MEROPS classification)^[11] of proteases with nucleophile and possible nucleophile-switching events indicated. Proteases with known structures were aligned to TEV^{Cys} by structural comparison using DALI,^[13] sequences of unknown structure were added by sequence alignment with BLASTp (see Table S1 and Figure S1 for details).

(TEV^{Ser}), recovering activity by directed evolution (DE) and measuring trade-offs in response to nucleophile switches.

The nucleophile mutation compromised the centrepiece of the catalytic mechanism and consequently had a much greater effect on activity than in previous handicap-recover experiments.^[15] A TEV^{Ser} mutant (C151S mutation) was constructed, and the substitution of the cysteine thiol nucleophile by a serine alcohol was found to reduce activity by four orders of magnitude (Figure 2 and Figure S4 in the Supporting Information). The effect of this deliberately introduced handicap was then quantified by measuring the reaction kinetics (by monitoring the hydrolysis of the C-Y FRET-pair substrate,^[16] Figure S2). Conversion of the catalytic nucleophile from sulfur to oxygen resulted in biphasic kinetics (Figures 2A and S3, Table S2); this is consistent with the formation and breakdown of an acyl-enzyme intermediate (i.e., a fast first step followed by a slower, rate-limiting step, Figure S5). The second-order rate constants k_2^{obs1} and k_2^{obs2} of TEV^{Ser} were found to be 80 and 20 000 times lower, respectively, than the measured second-order rate constant of TEV^{Cys} (representing k_{cat}/K_M).

In order to investigate how the enzyme can compensate for the handicap of using a non-native nucleophile and altered reaction chemistry, ten rounds of DE for activity recovery were performed (numbered TEV^{Ser} to TEV^{SerX}). Each round of DE consisted of error-prone PCR (1.3 ± 0.4 amino acid mutations per gene), activity screening of 350 enzyme variants by destruction of FRET in cell lysate (Figure S2), and selection of the single best variant. Any S151C revertants were discarded to force evolution to follow a forward pathway. Measurement of turnover rates in cells during screening reflects enzyme fitness as the product of both chemical reactivity and catalyst concentra-

tion (determined by biophysical properties, such as folding and solubility). The same FRET-pair substrate was used for both in vivo screening and in vitro kinetics to describe the enzyme-substrate interactions that were relevant for the selection (and avoid unique effects of the recognition of for example, small-peptide substrates with different reactivity and affinity).

During the rounds of evolution, no mutations in residues that make direct contacts with the triad (or are within a radius of 4 Å) resulted from experimental selections. Conversely, nine of the 13 point mutations accumulated 4–8 Å from the catalytic triad, in the second shell of residue interactions (Figure 3). In vitro kinetics of purified variants showed that the process of directed evolution recovered proteolytic activity by an improvement in both k_2^{obs1} (2×10^3 -fold) and k_2^{obs2} (3×10^3 -fold) and also changed the burst amplitude (Figure 2A, Tables S2 and S3), with diminishing improvements in later rounds. The accumulation of mutations around the enzyme active site (second shell) that increased catalysis reduced soluble expression sixfold; however, in rounds V and VI, surface mutations (W130C and E194D) and a C-terminal truncation (due to a frameshift) were selected that improved both solubility and activity (Figure 2C).

In addition to the adaptive trajectory of TEV^{Ser} (Figure 2B, front), the identified adaptive mutations were examined in the parental background, by reverting the nucleophile to the original cysteine at each step of the evolutionary trajectory (Figure 2B, back). The kinetics of these TEV^{Cys} variants could be fitted to a monophasic model with good correlation coefficients. Rather than respecialising the active site to use serine, as is typical of directed-evolution experiments, the 14 mutations accumulated by DE proved nearly neutral to activity with

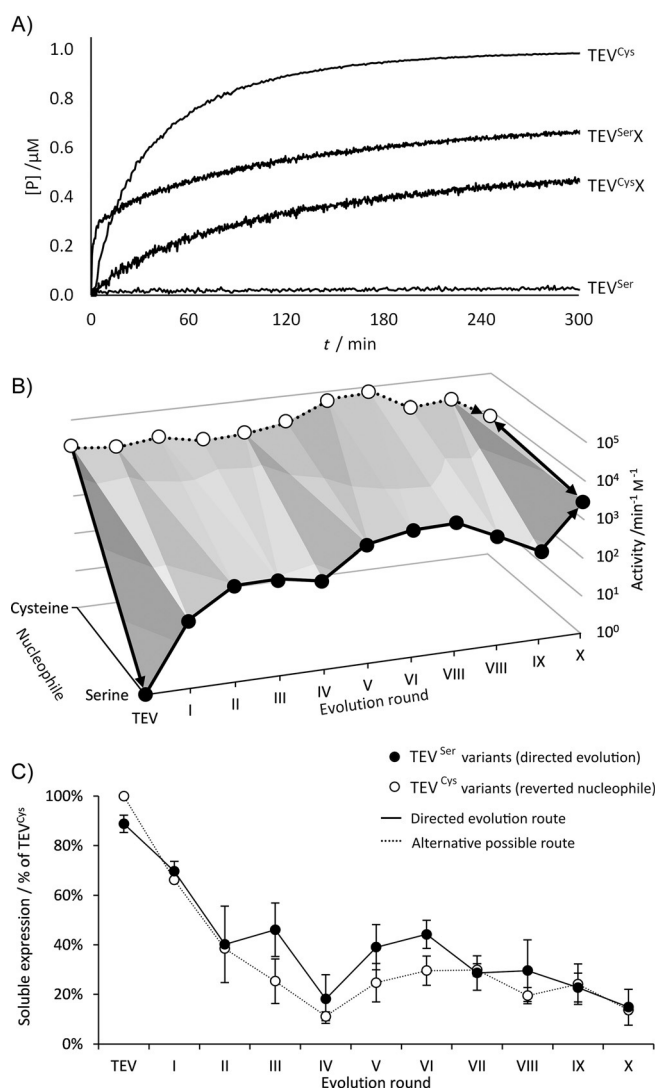


Figure 2. Activity recovery of the directed evolution lineage from TEV^{Ser} to TEV^{Ser}X and the corresponding nucleophile revertants, in which the original Cys nucleophile was restored. A) Kinetic traces of enzyme activity for the wild type-type enzyme (TEV^{Cys}), the nucleophile mutant (TEV^{Ser}), the evolved enzyme from the tenth round (TEV^{Ser}X) and the evolved mutant with a reverted nucleophile (TEV^{Cys}X). Conditions: [E] = 1 μM, [S] = 1 μM, pH 8, 25 °C. B) Development of the activity of purified TEV mutants as a function of the rounds of directed evolution. Activities are plotted as the second order rate constants k_2^{obs2} of the ten TEV^{Ser} variants and k_{cat}/K_M of the ten TEV^{Cys} nucleophile revertants. Solid arrows indicate the DE route (starting with the deleterious TEV^{Cys} → TEV^{Ser} mutation, then activity recovery in ten rounds from TEV^{Ser} to TEV^{Ser}X) to arrive at the neutral twin enzymes TEV^{Ser}X and TEV^{Cys}X, capable of using either nucleophile. The dotted arrow indicates an alternative, nearly neutral TEV^{Cys} → TEV^{Cys}X route. Conditions: [E] = 1–8 μM, [S] = 1 μM, pH 8, 25 °C. Standard deviations of four repeats were below 15% (Figure S5). C) Soluble expression of evolved variants and TEV^{Cys} back-mutants. Error bars represent standard deviation of two biological repeats.

the original cysteine nucleophile (i.e., did not trade-off). Whereas the k_2^{obs1} and k_2^{obs2} of TEV^{Ser}X are 1000-fold improved over those of TEV^{Ser}, the nucleophile revertants retained activity within fourfold of TEV^{Cys} (Figures 2B and S6). The evolutionary trajectory therefore results in twin enzymes, differing only in their nucleophile (TEV^{Ser}X ↔ TEV^{Cys}X) and with only a small, 2.3-fold difference in activity upon nucleophile exchange.

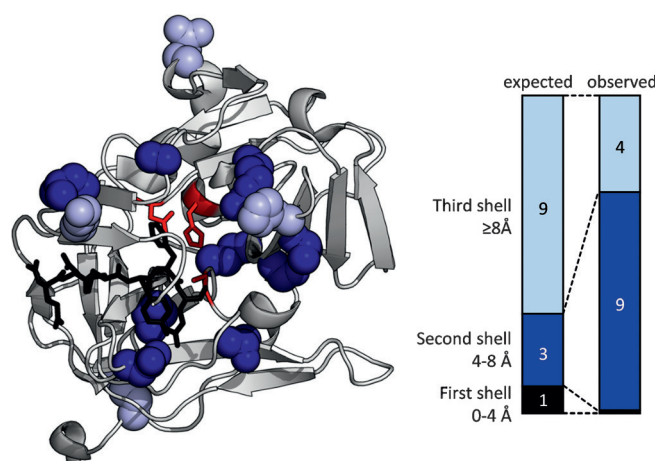


Figure 3. Positions of mutations accumulated during directed evolution mapped onto the structure of TEV protease (PDB ID: 1LVM). The catalytic triad is shown in red, the substrate peptide is in black. There are no mutations within 4 Å of the catalytic triad (the first shell), mutations 4–8 Å from the triad (the second shell) are shown in dark blue, and mutations 8–12 Å from the triad (the third shell) are shown in light blue. Mutations in the second shell around the catalytic triad are enriched compared to what would be expected from a random distribution.

By forcing TEV into a local fitness valley (TEV^{Ser}) and experimentally evolving for activity recovery, we mapped out an uphill trajectory (TEV^{Ser} → TEV^{Ser}X) that lies parallel in sequence space to a nearly flat, neutral trajectory of mutants with constant Cys nucleophile (TEV^{Cys} → TEV^{Cys}X; Figure 2B). Therefore, despite deliberately evolving TEV^{Cys} by using a fitness valley, the nucleophile-permissive TEV^{Ser}X can also be accessed from TEV^{Cys} by nearly neutral mutations without any large drops in activity. The most closely related natural serine protease only retains 15% sequence identity to TEV^{Cys}. However, the mutant TEV^{Ser}X shows 92.4% identity (Table S2, Figure 3), thus suggesting that only a few mutations are necessary to accommodate a nucleophile generalist. The >1000-fold improvement to generate a nucleophile generalist with only 13 mutations explains how divergent evolution of core catalytic machinery can occur within evolutionarily superfamilies such as the PA clan (Figure 1B). It also emphasises the power of directed evolution to find solutions for the challenge of retuning chemical reactivity.^[17]

The challenge of nucleophile permissiveness is conceptually similar to that of catalytic promiscuity (the ability of an enzyme to accept different substrates): how can an enzyme make and break bonds different from those it has evolved for?^[18] The evolution of promiscuous activities has been previously observed to pass through catalytic generalists, which were able to promote a new reaction, while retaining some activity on their original substrate. Generalist enzymes are proposed to perform an important role in the evolution of new functions by being particularly evolvable.^[19] Although enzyme promiscuity towards different substrates is well documented,^[18,20] the ability to use different residues for nucleophilic, covalent catalysis represents an alternative kind of chemical versatility in the core catalytic machinery.^[21]

Even though the lynchpin of catalysis in the protease active site—the nucleophile—was mutated, the large activity drop was readily recoverable by evolution. Quantitatively, both the handicap introduced and the extent of the recovery exceed those previously observed by approximately two orders of magnitude.^[15] Specifically, a 120-fold reduction triggered by co-factor exchange was followed by a 70-fold recovery in a study by Miller et al.^[15a] and a 400-fold reduction caused by mutation of a conserved residue was followed by a 40-fold recovery observed by Wellner et al.,^[15b] compared with a 20 000-fold reduction and 3000-fold recovery in this work.

What is more, the similar rates of TEV^{CysX} and TEV^{SerX} (3×10^3 vs. $7 \times 10^3 \text{ min}^{-1} \text{ M}^{-1}$) represent a rare example of catalytic promiscuity at high, wild-type levels (in contrast to promiscuous, yet low-activity catalytic generalists).^[22] Paradoxically, a nucleophile mutation would be predicted to be more difficult to recover from, when compared to evolution to accommodate promiscuous binding of multiple substrates, as two different types of bonds (O–C vs. S–C) are being formed and cleaved. However, our data suggest that the differences in nucleophile reactivity and structure can be readily accommodated by TEV protease with apparently little trade-off between rates for different nucleophiles. The unexpected nucleophile tolerance suggests that chemically versatile intermediates such as TEV^{Ser/CysX} exist that could facilitate the phylogenetically observed switch between protease clans that differ in their nucleophiles prior to specialisation. The protein framework of TEV^{Ser/CysX} allows two nucleophiles to execute their function with good efficiency and constitutes a molecular solution to escape from adaptive conflict.

Acknowledgements

We thank several colleagues, particularly Sean Devenish for insightful comments on the manuscript. We acknowledge financial support from the Biotechnology and Biological Sciences Research Council and MedImmune. F.H. is an ERC Starting Investigator.

Keywords: directed evolution • nucleophilic catalysis • PA clan • proteases • tobacco etch virus

- [1] a) N. Furnham, G. L. Holliday, T. A. de Beer, J. O. Jacobsen, W. R. Pearson, J. M. Thornton, *Nucleic Acids Res.* **2014**, *42*, D485–D489; b) A. R. Buller, C. A. Townsend, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E653–661.
[2] M. Y. Galperin, D. R. Walker, E. V. Koonin, *Genome Res.* **1998**, *8*, 779–790.

- [3] a) C. T. Hittinger, S. B. Carroll, *Nature* **2007**, *449*, 677–681; b) D. L. Des Marais, M. D. Rausher, *Nature* **2008**, *454*, 762–765; c) T. Sikosek, H. S. Chan, E. Bornberg-Bauer, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14888–14893.
[4] D. A. Kraut, K. S. Carroll, D. Herschlag, *Annu. Rev. Biochem.* **2003**, *72*, 517–571.
[5] a) A. J. Kirby, F. Hollfelder, *From Enzyme Models to Model Enzymes*, Royal Society of Chemistry, Cambridge, **2009**; b) A. J. Kirby, *Adv. Phys. Org. Chem.* **1980**, 183–278.
[6] a) P. A. Sigala, D. A. Kraut, J. M. Caaveiro, B. Pybus, E. A. Ruben, D. Ringe, G. A. Petsko, D. Herschlag, *J. Am. Chem. Soc.* **2008**, *130*, 13696–13708; b) A. J. Kirby, F. Hollfelder, *Nature* **2008**, *456*, 45.
[7] a) E. S. Radisky, J. M. Lee, C. J. Lu, D. E. Koshland, Jr., *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 6835–6840; b) L. Polgár, *Cell. Mol. Life Sci.* **2005**, *62*, 2161–2172; c) L. Hedstrom, *Chem. Rev.* **2002**, *102*, 4501–4524.
[8] A. Radzicka, R. Wolfenden, *J. Am. Chem. Soc.* **1996**, *118*, 6105–6109.
[9] a) A. J. Beveridge, *Protein Sci.* **1996**, *5*, 1355–1365; b) J. N. Higaki, B. W. Gibson, C. S. Craik, *Cold Spring Harbor Symp. Quant. Biol.* **1987**, *52*, 615–621.
[10] E. Di Cera, *IUBMB Life* **2009**, *61*, 510–515.
[11] N. D. Rawlings, A. J. Barrett, A. Bateman, *Nucleic Acids Res.* **2012**, *40*, D343–D350.
[12] a) P. S. Goldfarb, *Nature* **1988**, *336*, 429; b) S. Brenner, *Nature* **1988**, *334*, 528–530; c) A. E. Gorbalenya, V. M. Blinov, A. P. Donchenko, *FEBS Lett.* **1986**, *194*, 253–257; d) M. Irwin, *Nature* **1988**, *336*, 429–430.
[13] L. Holm, P. Rosenström, *Nucleic Acids Res.* **2010**, *38*, W545–W549.
[14] a) M. E. McGrath, M. E. Wilke, J. N. Higaki, C. S. Craik, R. J. Fletterick, *Biochemistry* **1989**, *28*, 9264–9270; b) J. P. Turkenburg, M. B. Lamers, A. M. Brzozowski, L. M. Wright, R. E. Hubbard, S. L. Sturt, D. H. Williams, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2002**, *58*, 451–455; c) K. E. Neet, D. E. Koshland, Jr., *Proc. Natl. Acad. Sci. USA* **1966**, *56*, 1606–1611; d) L. Polgár, B. Asboth, *J. Theor. Biol.* **1986**, *121*, 323–326; e) K. C. Cheah, L. E. Leong, A. G. Porter, *J. Biol. Chem.* **1990**, *265*, 7180–7187; f) C. S. Hahn, J. H. Strauss, *J. Virol.* **1990**, *64*, 3069–3073.
[15] a) S. P. Miller, M. Lunzer, A. M. Dean, *Science* **2006**, *314*, 458–461; b) A. Wellner, M. Raitses Gurevich, D. S. Tawfik, *PLoS Genet.* **2013**, *9*, e1003665.
[16] A. W. Nguyen, P. S. Daugherty, *Nat. Biotechnol.* **2005**, *23*, 355–360.
[17] H. Renata, Z. J. Wang, F. H. Arnold, *Angew. Chem. Int. Ed.* **2015**, *54*, 3351–3367; *Angew. Chem.* **2015**, *127*, 3408–3426.
[18] P. J. O'Brien, D. Herschlag, *Chem. Biol.* **1999**, *6*, R91–R105.
[19] a) A. Aharoni, L. Gaidukov, O. Khersonsky, S. McQ. Gould, C. Roodveldt, D. S. Tawfik, *Nat. Genet.* **2004**, *37*, 73–76; b) R. A. Jensen, *Annu. Rev. Microbiol.* **1976**, *30*, 409–425.
[20] a) I. Nobeli, A. D. Favia, J. M. Thornton, *Nat. Biotechnol.* **2009**, *27*, 157–167; b) O. Khersonsky, D. S. Tawfik, *Annu. Rev. Biochem.* **2010**, *79*, 471–505; c) S. D. Copley, *Trends Biochem. Sci.* **2015**, *40*, 72–78.
[21] A. Babbie, N. Tokuriki, F. Hollfelder, *Curr. Opin. Chem. Biol.* **2010**, *14*, 200–207.
[22] a) I. Matsumura, A. D. Ellington, *J. Mol. Biol.* **2001**, *305*, 331–339; b) N. Tokuriki, C. J. Jackson, L. Afriat-Jurnou, K. T. Wyganowski, R. Tang, D. S. Tawfik, *Nat. Commun.* **2012**, *3*, 1257.

Manuscript received: June 11, 2015

Accepted article published: June 19, 2015

Final article published: July 14, 2015