Protein Engineering

DOI: 10.1002/anie.201309591

Protein Engineering from "Scratch" Is Maturing

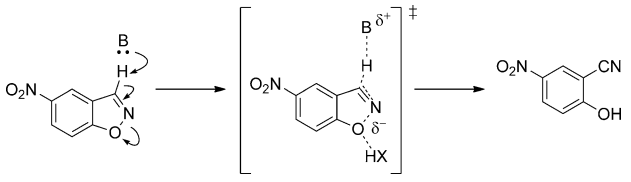
*Matthias Höhne and Uwe T. Bornscheuer**

computational chemistry · directed evolution · enzyme catalysis · Kemp eliminase · protein engineering

Biocatalysis uses natural enzymes to perform a desired reaction, and numerous examples covering almost all enzyme classes are well-documented in the literature. If a natural enzyme does not show the desired performance, then methods of protein engineering[1] can be used to adapt its properties to meet a given target such as stereoselectivity, desired substrate scope, or stability to heat or organic solvents. Particularly over the past two decades, many success stories based on various tools of protein engineering—ranging from random mutagenesis combined with high-throughput screening to rational protein design methods—have been written. This also includes various industrial processes based on specifically designed biocatalysts.[2]

A recent concept is the computational design of enzymes.[3] This design approach uses quantum mechanical models of the reaction transition state to create "theozymes", which are then docked into the enormous number of possible protein structures by means of an efficient algorithm. The in silico design takes advantage of the rapid increase in protein structure coordinates deposited in the RCSB Protein Data Bank (http://www.pdb.org), which has grown in the past decade about sixfold to now contain almost 100000 entries of protein structures.

The first successful design of an unnatural model reaction was the Kemp elimination of 5-nitrobenzisoxazole (Scheme 1).[4] The authors assumed that a His-Asp dyad should be able to ensure deprotonation of the carbon atom next to the N–O bond by a base and at the same time the aspartate group would regulate the pK_a value of the His residue. The first predicted proteins had detectable catalytic activity ($k_{cat} = 0.0009–0.29\ s^{-1}$) and a few multiple turnovers. In a further step, the authors were able to improve the activity by subsequent directed evolution to a k_{cat} value of $9.5\ s^{-1}$ and a k_{cat}/K_M value of $60430\ M^{-1}s^{-1}$ (variant KE 59.13).[5] On the positive side, Baker and co-workers demonstrated that catalytically active proteins can indeed be designed from scratch. Unfortunately, their efficiency was not in the range of synthetically useful catalysts.



Scheme 1. The Kemp elimination.

In an outstanding recent contribution, the Hilvert research group have now succeeded in creating an efficient Kemp eliminase (KE).[6] The starting scaffold (HG3, based on a xylanase structure) originated from an earlier study and had a k_{cat}/K_M value of $1300\ M^{-1}s^{-1}$. First, a combination of global methods such as the error-prone polymerase chain reaction (epPCR) and DNA shuffling paired with local analysis using rationally and experimentally derived hot spots led to a four-fold improvement (HG3.3b, Figure 1).

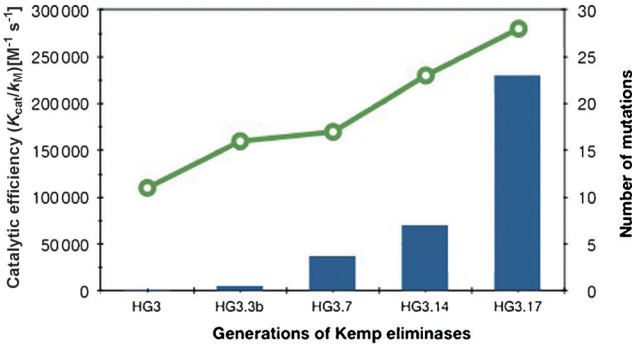


Figure 1. Improvement of the catalytic efficiency k_{cat}/K_M (blue columns) and number of mutations (green line) introduced in the generations of Kemp eliminases.

The best variants from both approaches were then combined and subjected to further rounds of shuffling and epPCR, which led to HG3.7 being about 29-fold more active. Further improvements were achieved by focusing on mutations in the active-site tunnel followed again by several rounds of shuffling, which led to the best variant with $k_{cat}/K_M = 230000\ M^{-1}s^{-1}$, an astonishing k_{cat} value of $700\ s^{-1}$ (exceeding

[*] Jun. Prof. M. Höhne
Abt. Proteinbiochemie
Institut für Biochemie, Universität Greifswald
Felix-Hausdorff-Strasse 4, 17487 Greifswald (Germany)
Prof. U. T. Bornscheuer
Abt. Biotechnologie & Enzymkatalyse
Institut für Biochemie, Universität Greifswald
Felix-Hausdorff-Strasse 4, 17487 Greifswald (Germany)
E-mail: uwe.bornscheuer@uni-greifswald.de
Homepage: http://biotech.uni-greifswald.de

many natural enzymes), and negligible product inhibition. This Kemp eliminase contained 28 mutations (11 were already introduced by computational design into the starting scaffold), thus on average one amino acid mutation was kept per directed evolution round.

The high number of mutations required to turn a slow Kemp eliminase into an efficient one matches other examples where extensive protein engineering led to success: directed evolution of a halohydrin dehalogenase for the production of the atorvastatin (Lipitor) side chain required alteration of at least 14% of the amino acids (35 of a total of 254 amino acids).[7] Protein engineering of an amine transaminase to make the antidiabetic drug sitagliptin changed 8.2% of all residues.[8] In a retro-aldolase, also created by using the computational design approach, 4–6% of the starting scaffold (a xylanase) needed to be exchanged.[9]

How can the extraordinary improvement in the Kemp eliminase be explained? Three decisive factors have been identified:

1) An extraordinarily high shape complementarity: whereas two orientations were found for substrate binding in HG3, only one binding mode existed in HG3.17, thus enabling more efficient catalysis. This perfect binding, however, resulted in a narrower substrate scope.
2) The ligand alignment with Asp127, which serves as the catalytic base, was optimized by multiple rounds of directed evolution and resulted in an unusually short hydrogen bond.
3) Gln50 was identified to be very well positioned and served as a new catalytic residue that stabilizes the negative charge in the transition state.

Technically, the achievement by the Hilvert research group still required the laborious creation of 17 libraries combined with substantial high-throughput screening. This situation is similar to that in the early 1990s, when directed evolution was discovered as a tool for protein engineering. Our lack of sufficient knowledge requires current computational design programs to neglect protein dynamics and requires the use of the fixed backbone assumption, thereby preventing the sampling of backbone structural states, which are needed for the precise positioning of catalytic sites. Only with this input, can the extensive rounds of protein engineering to create an efficient catalyst be avoided or reduced to become more practical.

One issue can also be the choice of the right starting scaffold. The initial design by Baker and co-workers (KE) started from an indole-3-glycerolphosphate synthase (an α/β-barrel structure), whereas the HG3 proteins were based on a xylanase, a triose-phosphate isomerase (TIM) barrel scaffold. In another example, Baker and co-workers designed an enzyme to catalyze the Morita–Baylis–Hillman reaction between 2-cyclohexenone and *p*-nitrobenzaldehyde.[10] They reasoned that a cysteine or histidine residue could serve as the catalytic residue, and this led to the identification of two designs out of 48 proteins that showed very low, but detectable activity above the background. Interestingly, several years before, the Reetz research group had already shown[11] that several esterases and lipases—all having a Ser-His-Asp catalytic triad—displayed very good activity (up to 10% conversion at 200 mM substrate concentration) in this promiscuous reaction. This demonstrates that more than one solution can exist to solve a challenge in enzyme catalysis, and raises the question whether the hydrolase mechanism would have been the better starting point here for the computational design of a Morita–Baylis–Hillman enzyme. Nature also uses different designs to solve a given synthetic problem, even when the same type of reaction is catalyzed. For example, type I aldolases, which occur in plants and animals, have a catalytic Lys residue, whereas type II aldolases from bacteria or fungi are Zn-dependent enzymes.

The Kemp elimination is a rather easy one-step reaction in which proton transfer is coupled to irreversible cleavage of the N–O bond. From the extensive protein engineering study and structural characterization of the improved variant, the Hilvert research group recognized factors that are essential for improving this computationally designed enzyme. Together with improved computational power, this might lead to optimized design algorithms. This in turn would allow a more efficient engineering from "scratch" right from the start (and would represent a true maturation of computational design). However, many of the important factors for gaining highly efficient catalysts for other catalytic problems are not yet revealed and, hence, extensive protein engineering is both challenging and also a fascinating possibility to not only improve an enzyme, but to gain further knowledge for future designs. Additionally, it remains to be answered whether a combination of computational design and protein engineering will also enable the creation of efficient enzymes that catalyze (stereoselective) multisubstrate reactions and preferentially have a broader substrate scope (e.g. the optimization of the Diels–Alderase[12] designed by Baker and co-workers). Only then will this concept become competitive to enzymes from nature.

Received: November 4, 2013
Published online: December 11, 2013

[1] a) *Protein Engineering Handbook*, *Vol. 1–3* (Eds.: S. Lutz, U. T. Bornscheuer), Wiley-VCH, Weinheim, **2009/2013**; b) R. J. Kazlauskas, U. T. Bornscheuer, *Nat. Chem. Biol.* **2009**, *5*, 526–529; c) M. T. Reetz, *Angew. Chem.* **2011**, *123*, 144–182; *Angew. Chem. Int. Ed.* **2011**, *50*, 138–174.
[2] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185–194.
[3] a) G. Kiss, N. Celebi-Olcum, R. Moretti, D. Baker, K. N. Houk, *Angew. Chem.* **2013**, *125*, 5810–5836; *Angew. Chem. Int. Ed.* **2013**, *52*, 5700–5725; b) H. Kries, R. Blomberg, D. Hilvert, *Curr. Opin. Chem. Biol.* **2013**, *17*, 221–228; c) D. N. Bolon, S. L. Mayo, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14274–14279; d) B. I. Dahiyat, S. L. Mayo, *Science* **1997**, *278*, 82–87.
[4] a) D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. De Chancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, *Nature* **2008**, *453*, 190–195; b) for an alternative design resulting in an allosterically controlled Kemp eliminase, see I. V. Korendovych, D. W. Kulp, Y. Wu, H. Cheng, H. Roder, W. F. DeGrado, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6823–6827.

[5] O. Khersonsky, G. Kiss, D. Röthlisberger, O. Dym, S. Albeck, K. N. Houk, D. Baker, D. S. Tawfik, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 10358 – 10363.
[6] R. Blomberg, H. Kries, D. M. Pinkas, P. R. Mittl, M. G. Grutter, H. K. Privett, S. L. Mayo, D. Hilvert, *Nature* **2013**, *503*, 418 – 421.
[7] R. J. Fox, S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam, S. Muley, J. Grate, J. Gruber, J. C. Whitman, R. A. Sheldon, G. W. Huisman, *Nat. Biotechnol.* **2007**, *25*, 338 – 344.
[8] C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman, G. J. Hughes, *Science* **2010**, *329*, 305 – 309.
[9] L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas III, D. Hilvert, K. N. Houk, B. L. Stoddard, D. Baker, *Science* **2008**, *319*, 1387 – 1391.
[10] S. Bjelic, L. G. Nivon, N. Celebi-Olcum, G. Kiss, C. F. Rosewall, H. M. Lovick, E. L. Ingalls, J. L. Gallaher, J. Seetharaman, S. Lew, G. T. Montelione, J. F. Hunt, F. E. Michael, K. N. Houk, D. Baker, *ACS Chem. Biol.* **2013**, *8*, 749 – 757.
[11] T. Reetz, R. Mondière, J. D. Carballeira, *Tetrahedron Lett.* **2007**, *48*, 1679 – 1681.
[12] J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael, D. Baker, *Science* **2010**, *329*, 309 – 313.

