

# An Artificial Cofactor catalyzing the Baylis-Hillman Reaction using Designed Streptavidin as Protein Host

## Abstract

An artificial cofactor based on an organocatalyst embedded in a protein was used to conduct the Baylis-Hillman reaction in a buffered system. As protein host we chose streptavidin, since it can be easily crystallized and thereby supports the design process. The protein host around the cofactor was rationally designed based on high-resolution crystal structures obtained after each variation of the amino acid sequence. Additionally, DFT-calculated intermediates and transition states were used to rationalize activity. Finally, repeated cycles of structure determination and redesign led to a system with 24 to 35-fold increased activity over the bare cofactor and to the most active proteinogenic catalyst for the Baylis-Hillman reaction known today.

## Authors

Horst Lechner<sup>\*[a]</sup>, Vincent R. Emann<sup>[a]</sup>, M. Breuning<sup>[b]</sup> and Birte Höcker<sup>\*[a]</sup>

## Affiliations

[a] University Bayreuth, Department of Biochemistry, Universitaetsstr. 30, 95447 Bayreuth, Germany

[b] University Bayreuth, Organic Chemistry, Universitaetsstr. 30, 95447 Bayreuth, Germany

## Correspondence

[horst.lechner@uni-bayreuth.de](mailto:horst.lechner@uni-bayreuth.de), [birte.hoecker@uni-bayreuth.de](mailto:birte.hoecker@uni-bayreuth.de)

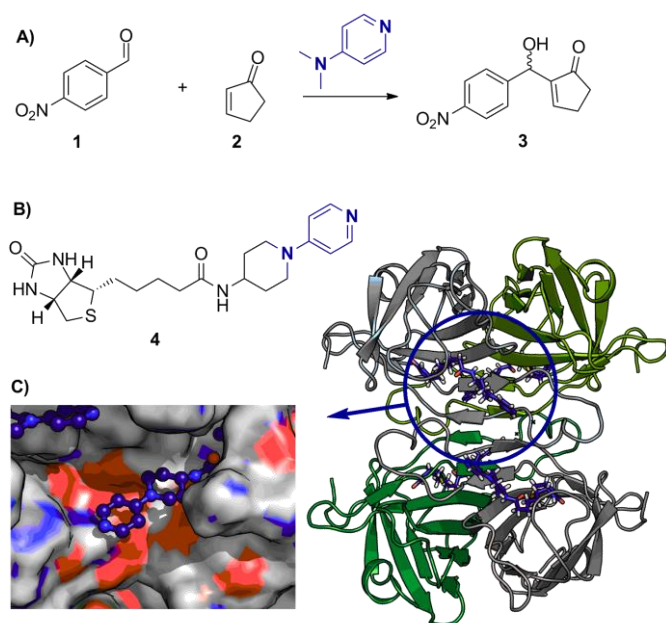
## Keywords

biocatalysis, artificial cofactor, protein design, Baylis-Hillman reaction, streptavidin

The design of proteins that display new catalytic activities is still a major challenge. Although several successful examples were reported<sup>[1–3]</sup>, these *de-novo* cases are limited to a small set of reactions, such as the Kemp elimination<sup>[4,5]</sup> and the Retro-aldol reaction<sup>[6]</sup> as well as a bimolecular Diels-Alder reaction.<sup>[7]</sup> All the initial designs provided a sufficient starting point but had to be strongly improved through directed evolution to show reasonable rate enhancement.<sup>[8–10]</sup>

Of all the natural enzyme catalyzed reactions half of them require cofactor(s) as part of their catalytic machinery.<sup>[11]</sup> Hence it might not be surprising that far more new protein-based catalysts were reported using artificial (metal-based) cofactors embedded in host-proteins.<sup>[12–16]</sup> These cofactors bear the advantage of intrinsic activity, which is usually low without surrounding protein. The protein can provide a different environment for the catalysis – increasing in most cases activity and maybe even facilitating (stereo)selective reactions.

There are a number of examples using streptavidin as host for these cofactors, since its natural ligand biotin binds strongly ( $K_d \sim 10^{-15}$  M) and is easy to modify with catalysts at the carboxylic acid group. The location of the catalyst is a shallow cavity at the surface of the protein. Streptavidin is a known thermo- and solvent stable protein, which is another advantage for developing new catalysts.



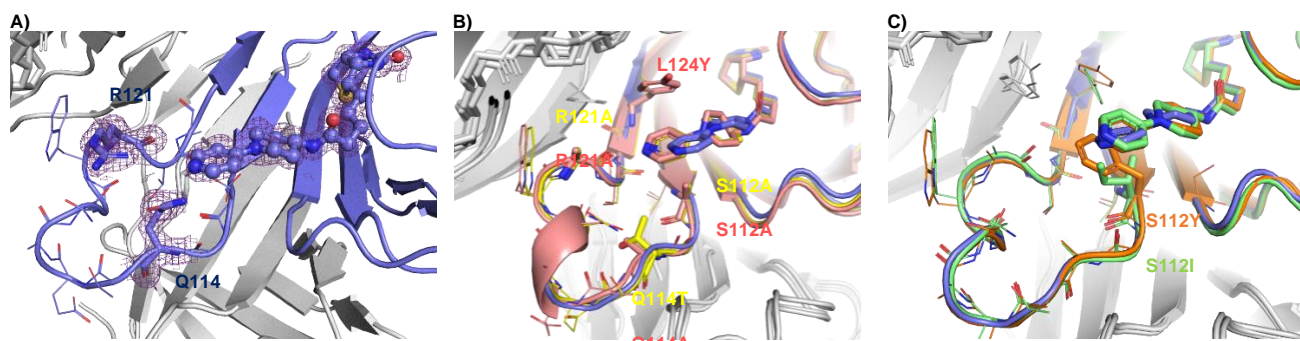
**Figure 1.** Concept of the artificial cofactor - protein host system as catalyzing the Baylis-Hillman reaction. **A)** DMAP used as catalyst for the reaction of an activated alkene **1** with an aldehyde **2** to yield the racemic alcohol **3**, **B)** Artificial cofactor **4** consisting of biotinylated DMAP-derivative, **C)** The crystal structure of the tetrameric *wt*-streptavidin harboring **4** and a closeup of the binding site of **4**.

Contrary to metal-based cofactors, organocatalysts are rarely used in this context. There are only few examples: proline<sup>[17]</sup> and an imidazolium salt<sup>[18]</sup> were used in wildtype (*wt*)-Streptavidin as host for Michael additions and Aldol-reactions, respectively.

No enzyme is known to naturally catalyze the Baylis-Hillman reaction<sup>[19,20]</sup>, a very versatile C-C bond forming reaction for the production of various compounds<sup>[21]</sup> and intermediates of APIs<sup>[22]</sup>. This reaction provides in an atom-economic way a very functionalized product. There were attempts to design a *de-novo* enzyme but with only minor success.<sup>[23]</sup> Nevertheless it was recognized that this reaction can be catalyzed using nucleophilic amine catalysts in aqueous systems, although with low rates and high catalyst concentrations. (Imid)azoles<sup>[24,25]</sup>, DABCO<sup>[26,27]</sup>, 3-hydroxyquinuclidine<sup>[27]</sup> and 4-dimethylaminopyridine (DMAP)<sup>[27]</sup> were reported as catalysts for this purpose.

We chose a derivative of the latter to be used as an artificial cofactor due to its low steric expansion, and supposedly easier to include in a protein scaffold. Thus, the DMAP derivative 4-(4-aminopiperidino)pyridine was used to synthesize the artificial cofactor **4** (Figure 1, B).

As protein host streptavidin was used due to its aforementioned advantages. The wt-protein was expressed, purified and after adding **4**, the activity for the reaction tested. *p*-Nitrobenzaldehyde **1** and cyclopentenone **2** were used as model substrates (Figure 1, A). No activity above the activity seen in a blank reaction without cofactor could be observed. Consequently, the x-ray structure was solved (Figure 1C, PDB: 6T1E) to reveal the causes. It clearly visualized that **4** is placed nicely at the entrance of the biotin-binding pocket. The density of the ligand is very well resolved (see Composite Omit maps, Figure S1). As depicted in Figure 1, C and Figure 2, A, the catalyst is sterically hindered by the sidechains of Q114 and R121. Especially position 121 is known to influence the outcome of metal-based catalysts to a great extent.<sup>[28]</sup> Thus, the sterically demanding Q114 was changed to A or T, and R121 and S112 were substituted by A as well. Additionally, the variant S112A, R121A was tested, but none of these variants showed activity higher than the background (Table 1, Entry 4 and 5). Exemplarily the activity of streptavidin S112A Q114A R121A is shown in Table 1, Entry 6. The derived crystal structures of these variants revealed that **4** was moved to a position close to the amide carbonyl oxygen of A121, leading to a non-accessible nucleophilic nitrogen atom (Figure 2, B yellow). An attempt of introducing Y as a bulkier amino acid at position L124 behind the catalyst to push it forward was not successful either (Figure 2, B pink). The next residue that caught our attention was S112 beneath the catalyst. By increasing its size from A to its original S and further increasing this residue to M, F and I we reasoned it might force the catalyst in a better position. The variant harboring S112 led to no activity and a similar structure was yielded (PDB: 6T30) as in round 1 and 2 of our design approaches. In contrast, all variants harboring sterically more demanding residues at position 112 (I, F, M) displayed activity (Table 1, Entries 7-11). Our reasoning was supported by the x-ray structures of these variants (Figure 2, C) where the position of the nucleophilic nitrogen of **4** superimposes with the one from the initial structure. The S112I variant displayed the highest activity among the three variants (Table 1, Entry 9). The S112F variant, although structurally very similar to S112I (Figure 2, C) catalyzed roughly half of the conversion as the S112I variant (Table 1, Entry 10). Reasons for this might be the  $\pi$ - $\pi$  interaction of F112 and the catalyst, keeping it too tightly together as explained below. S112M (PDB: 6T2Z) led to two different conformations of **4** (Figure S1) and very low activities (Table 1, Entry 11). While all these constructs are active, they however lack enantioselectivity.



**Figure 2.** Crystal structures obtained during the course of catalysts development. Chain A of corresponding structure is colored, chain B, C and D are in grey. **A)** wt-streptavidin (violet; PDB: 6T1E) displaying **4**, blocked by residues Q114 and R121. Density obtained from the 2mFo-DFc map. **B)** First and second round of design: S112A, Q114T and R121A (yellow, PDB: 6T1G) as well as S112A, Q114A, R121A, L124Y (pink, PDB: 6T2Y). Both variants have space at the original positions of the catalyst, but the catalyst moved tilted backwards in a non-accessible position. **C)** Third round of design where additional to S112A, Q114A, R121A and L124Y position S112 was changed to F (orange, PDB: 6T31) and I (green, PDB: 6T32) leading to a good position of the catalyst within the protein environment.

**Table 1.** Activities and selectivities of catalysts and selected streptavidin constructs for the model substrates **1 + 2**.

Entry	Catalyst	Mol% Cat.	Reaction time [h]	Conv. [%] <sup>[a]</sup>	e.e. [%] <sup>[a]</sup>
1	-	-	48	0	
2	DMAP	2	48	3	
3	<b>4</b>	2	48	1	
4	Streptavidin <sup>[b]</sup>		24	6	
5	Streptavidin		48	11	<5
6	<b>S112A</b> Q114A R121A <sup>[b]</sup> + <b>4</b>	1	24	8	
7	<b>S112I</b> Q114A R121A <sup>[b]</sup> L124Y + <b>4</b>	1	24	13	
8	<b>S112I</b> Q114A R121A L124Y + <b>4</b>	2	24	16	
9	<b>S112I</b> Q114A R121A L124Y + <b>4</b>	2	48	35	<5
10	<b>S112F</b> Q114A R121A L124Y + <b>4</b>	2	48	17	<5
11	<b>S112M</b> Q114A R121A L124Y + <b>4</b>	2	48	12	<5
12	K49N <b>S112I</b> Q114A R121A L124Y <b>H127Q</b>		48	9	
13	K49N <b>S112I</b> Q114A R121A L124Y <b>H127Q</b> + <b>4</b>	2	48	23	

Reaction conditions: HEPES buffer (10 mM, pH 7.0), streptavidin (2.6 mol% of monomer), DMSO (20 vol%), 4-nitrobenzaldehyde (50 mM), cyclopentenone (100 mM), 30 °C, orbital shaking 1000 rpm. [a] Determined by HPLC in duplicates [b] 1.3 mol% of monomer.

Two major questions arise with these results in hand: What is the cause for the high “background” reaction displayed by streptavidin alone? And why does the active protein-cofactor system lack enantioselectivity?

We want to address the question regarding the high background first. In this regard it is worth mentioning that others<sup>[29]</sup> reported some proteins at high concentrations (30 mg/ml), such as the carrier protein BSA, being able to catalyze the Baylis-Hillman reaction. The mechanism of catalysis was however not elucidated further. We considered histidine residues at the surface of streptavidin as the cause for catalysis, since it was described that (imid)azoles<sup>[24,30]</sup> can serve as catalysts under certain conditions. Streptavidin has two histidine residues per monomer located at the surface (H87, H127) (Figure S2). H127 is known to be tolerant against mutations.<sup>[31]</sup> Therefore we created a double mutant of the S112I variant with a H127Q and K49N exchange. The latter exchange removes a lysine close to the catalyst, which could also have an effect on the outcome of the reaction. While the conversions dropped somewhat (Table 1, Entry 12 and 13), they were still comparable to the results obtained without these mutations. Thus, H127 has minor importance for the background reaction. Exchange of H87 was already recognized to disturb the integrity of the protein.<sup>[32]</sup> After careful inspection of the structures we suspect that this histidine forms a hydrogen bond to an aspartic acid of a neighboring chain (D61) keeping loops of neighboring subunits of streptavidin together and creating a “catalytic diad” by activating the histidine, which would explain the outcome of our experiments. Our efforts to change this H87 to a D, N or S unfortunately lead to misfolded protein, only the double mutant H87Y and D61I could be refolded with very low yields and diminished biotin binding ability. As a consequence, this variant could not be characterized further.

Next we addressed the issue of enantioselectivity in the following way. Increasing the size of the aldehyde by using Isatin<sup>[33]</sup> instead of nitrobenzaldehyde **1** with the aim to achieve stereoselectivity via “substrate engineering” led to conversions up to 87% and using the even bigger N-methyl-isatin conversion could be measured again but to no enhanced enantioselectivity was detected (Table S4).

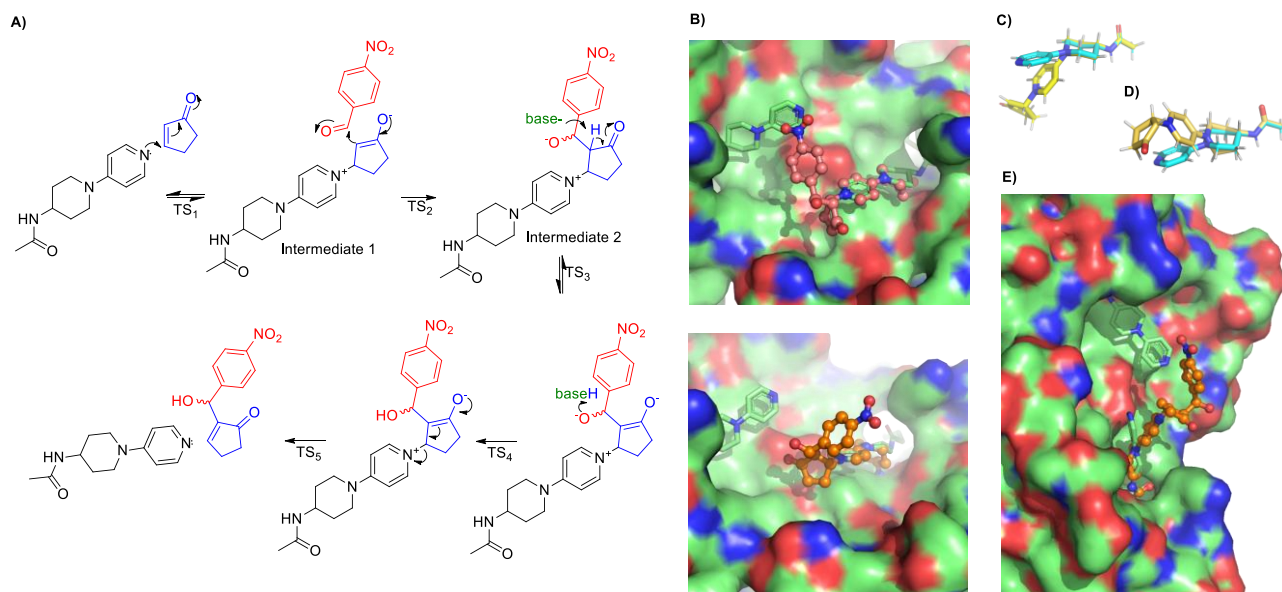
The lack of enantioselectivity of the catalyst was more deeply investigated using DFT methods. We employed ORCA<sup>[34]</sup> to calculate the reaction pathway until Intermediate 2, where the chiral center is already formed<sup>[35]</sup> (Figure 3, A) at a B3LYP level of theory using a dispersion correction<sup>[36]</sup> and a CPCM model for water. As basis set ma-def2-SVP<sup>[37]</sup> was employed.

By overlaying the positions of the pyridine ring of the catalyst of DFT-derived structures and the crystal structure, one could argue that the reasons for the missing enantioselectivity of the reaction should be clarified. As depicted in Figure 2, B the formation of both enantiomers is possible since both intermediates are not sterically hindered by surrounding amino acids although the intermediate leading to the (*S*) product (Figure 2 B, bottom) seems to undergo more favorable interactions with the protein (cyclopentenone **1** oxygen to Y124). Additionally, there is less spatial proximity of the cofactor **4** of chain B to -NO<sub>2</sub> of **2** as compared to the intermediate leading to the (*R*)-product (Figure 2 B, top). This should at least lead to a moderate enantioselectivity, which we could not observe. If the proposed mechanism of action using DMAP as catalyst is considered in more detail, we notice that during the intermediate steps the positive charge of the pyridinium migrates to the amine-nitrogen atom (Figure S3), which leads to some consequences concerning cofactor **4**. Since cofactor **4** possesses at this position, in contrast to DMAP, a 6-membered ring the conformation of this piperidine ring has to change due to the need for accommodation of a sp<sup>2</sup> hybridized amine nitrogen atom eventually bearing the positive charge. This conformational change propagates through **4** leading to a kinked molecule. Two possible conformations of the piperidine ring are possible – either the energetically favored chair conformation<sup>[38]</sup> or the unfavored (twisted) boat conformation.

The energetically lower conformation of the piperidine possessing a chair conformation (Figure 3 C) is hindered by the surrounding protein. On the other hand, fits the twisted boat conformation of piperidine (Figure 3, D), around 3.8 kcal/mol higher in energy for intermediate 1, quite well into into the given protein environment. The conformational change places the reactive nitrogen at a very different spatial position as seen in almost all crystal structures (Figure 3, C-E). Since the protein crystallized in all cases, except the wildtype, under basic conditions with a neutral cofactor this conformation was never observed. Interestingly, the piperidine boat conformation of the catalyst is observed once in the solved crystal structures, namely in variant S112M Q114A R121A L124Y (PDB: 6T2Z, Figure S1). But only one of the two chains of the assymmetric unit harbours **4** in this conformation. The second chain has electron density for **4** in a piperidine chair conformation. Surprisingly, this variant has the lowest activity of all active variants.

Since the catalyst is according to this theoretical considerations and calculations now pointing towards the solvent as depicted in Figure 3E, any chiral induction by the protein environment is lost, leading to achiral product formation.

The catalytic rate enhancement of protein with **4** over the rate of **4** is most presumably due to the hydrophobic pocket around the cofactor, increasing locally the concentration of the substrates and facilitating the formation of Intermediate 1.



**Figure 3.** Explanation of stereoselective outcome of reaction using DFT calculated intermediates and the protein crystal structure of streptavidin S112I Q114A R121A L124Y. **A)** Proposed reaction mechanism of the Baylis-Hillman reaction **B)** Both possible enantiomers of Intermediate 2 overlaid on the pyridine ring of the catalyst **4**. Top: (*R*)-product (pink), Bottom: (*S*)-product (orange). **C)** DFT structure of N-(1-(pyridin-4-yl)piperidin-4-yl)acetamide (turquoise) as catalyst and Intermediate 1 in chair conformation (yellow, right) and **D)** boat conformation (gold, left), respectively, overlaid on amide bond **E)** Intermediate 2 in twisted boat conformation overlaid on the amid bond from **4** in the crystal structure of the host protein. The cofactor is kinked and points towards the solvent.

In conclusion, a new artificial cofactor was developed, which utilizes a known organocatalyst and the biotin-streptavidin technology to successfully catalyze the Baylis-Hillman reaction. This protein-cofactor system yields a ~35-fold higher conversion compared to the bare artificial cofactor alone. The system was evolved via repeated cycles of mutagenesis, activity tests and structure determination. Hence, all protein design steps are rationalized. The reaction catalyzed by the artificial cofactor was further elucidated using DFT calculations to show major conformational changes of the catalyst, preventing the formation of chiral products. Here we demonstrate, how carefully considering the chemistry of a reaction, if fully understood, as well as deep knowledge of the (structural) properties of the target proteins are key to success in enzyme design.

## Acknowledgements

H.L. gratefully acknowledges financial support by the Austrian Science Fund (FWF, Erwin Schrödinger fellowship, J3994 B21). We thank Sooruban Shanmugaratnam for excellent technical assistance, all X-ray team members at the department for trips to Berlin, HZB (Berlin, Germany) for the allocation of synchrotron radiation beamtime at BESSY II as well as the Keylab HPC (University of Bayreuth) for computer time and the Central Analytics of the Organic Chemistry Department (University of Bayreuth) for HRMS measurements.

## Conflict of interest

The authors declare no conflict of interest.

- [1] H. Lechner, N. Ferruz, B. Höcker, *Curr. Opin. Chem. Biol.* **2018**, *47*, 67–76.
- [2] H. Kries, R. Blomberg, D. Hilvert, *Curr. Opin. Chem. Biol.* **2013**, *17*, 221–8.
- [3] G. Kiss, N. Çelebi-Ölçüm, R. Moretti, D. Baker, K. N. Houk, *Angew. Chemie Int. Ed.* **2013**, *52*, 5700–5725.
- [4] D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, et al., *Nature* **2008**, *453*, 190–195.
- [5] H. K. Privett, G. Kiss, T. M. Lee, R. Blomberg, R. A. Chica, L. M. Thomas, D. Hilvert, K. N. Houk, S. L. Mayo, *Proc. Natl. Acad. Sci.* **2012**,

- 109, 3790–3795.
- [6] 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, et al., *Science* **2008**, *319*, 1387–91.
- [7] 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, et al., *Science* **2010**, *329*, 309–313.
- [8] R. Blomberg, H. Kries, D. M. Pinkas, P. R. E. Mittl, M. G. Grütter, H. K. Privett, S. L. Mayo, D. Hilvert, *Nature* **2013**, *503*, 418–421.
- [9] O. Khersonsky, G. Kiss, D. Röthlisberger, O. Dym, S. Albeck, K. N. Houk, D. Baker, D. S. Tawfik, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 10358–63.
- [10] R. Obexer, A. Godina, X. Garrabou, P. R. E. Mittl, D. Baker, A. D. Griffiths, D. Hilvert, *Nat. Chem.* **2017**, *9*, 50–56.
- [11] J. D. Fischer, G. L. Holliday, S. A. Rahman, J. M. Thornton, *J. Mol. Biol.* **2010**, *403*, 803–824.
- [12] M. E. Wilson, G. M. Whitesides, *J. Am. Chem. Soc.* **1978**, *100*, 306–307.
- [13] M. Dürrenberger, T. R. Ward, *Curr. Opin. Chem. Biol.* **2014**, *19*, 99–106.
- [14] T. Heinisch, T. R. Ward, *Acc. Chem. Res.* **2016**, *49*, 1711–1721.
- [15] A. D. Liang, J. Serrano-Plana, R. L. Peterson, T. R. Ward, *Acc. Chem. Res.* **2019**, *52*, 585–595.
- [16] H. J. Davis, T. R. Ward, *ACS Cent. Sci.* **2019**, *5*, 1120–1136.
- [17] A. R. Nödling, K. Świderek, R. Castillo, J. W. Hall, A. Angelastro, L. C. Morrill, Y. Jin, Y.-H. Tsai, V. Moliner, L. Y. P. Luk, *Angew. Chemie Int. Ed.* **2018**, *57*, 12478–12482.
- [18] V. Gauchot, M. Branca, A. Schmitzer, *Chem. - A Eur. J.* **2014**, *20*, 1530–8.
- [19] Y. Wei, M. Shi, *Chem. Rev.* **2013**, 6659–90.
- [20] D. Basavaiah, B. S. Reddy, S. S. Badsara, *Chem. Rev.* **2010**, *110*, 5447–5674.
- [21] D. Basavaiah, A. J. Rao, T. Satyanarayana, *Chem. Rev.* **2003**, *103*, 811–891.
- [22] T. Narendar Reddy, V. Jayathirtha Rao, *Tetrahedron Lett.* **2018**, *59*, 2859–2875.
- [23] S. Bjelic, L. G. Nivón, N. Çelebi-Ölçüm, G. Kiss, C. F. Rosewall, H. M. Lovick, E. L. Ingalls, J. L. Gallaher, J. Seetharaman, S. Lew, et al., *ACS Chem. Biol.* **2013**, *8*, 749–757.
- [24] S. Luo, P. G. Wang, J. P. Cheng, *J. Org. Chem.* **2004**, *69*, 555–8.
- [25] S. Luo, X. Mi, P. G. Wang, J. P. Cheng, *Tetrahedron Lett.* **2004**, *45*, 5171–5174.
- [26] D. J. Maher, S. J. Connon, *Tetrahedron Lett.* **2004**, *45*, 1301–5.
- [27] V. K. Aggarwal, D. K. Dean, A. Mereu, R. Williams, *J. Org. Chem.* **2002**, *67*, 510–514.
- [28] H. Mallin, M. Hesticová, R. Reuter, T. R. Ward, *Nat. Protoc.* **2016**, *11*, 835–852.
- [29] M. T. Reetz, R. Mondière, J. D. Carballeira, *Tetrahedron Lett.* **2007**, *48*, 1679–81.
- [30] S. Luo, X. Mi, P. George Wang, J.-P. Cheng, *Tetrahedron Lett.* **2004**, *45*, 5171–4.
- [31] G. O. Reznik, S. Vajda, C. L. Smith, C. R. Cantor, T. Sano, *Nat. Biotechnol.* **1996**, *14*, 1007–1011.
- [32] T. Hohsaka, N. Muranaka, C. Komiyama, K. Matsui, S. Takaura, R. Abe, H. Murakami, M. Sisido, *FEBS Lett.* **2004**, *560*, 173–177.
- [33] J. C. Gomes, J. Sirvent, A. Moyano, M. T. Rodrigues, F. Coelho, *Org. Lett.* **2013**, *15*, 5838–5841.
- [34] F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1327.
- [35] R. E. Plata, D. A. Singleton, *J. Am. Chem. Soc.* **2015**, *137*, 3811–3826.
- [36] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- [37] J. Zheng, X. Xu, D. G. Truhlar, *Theor. Chem. Acc.* **2011**, *128*, 295–305.
- [38] A. Goldblum, O. Deeb, G. H. Loew, *J. Mol. Struct. THEOCHEM* **1990**, *207*, 1–14.