Chemical Probes for Mechanistic Enzymology

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Pascal Engi

Aus Safiental, Schweiz

Basel, 2019

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

Auf Antrag von

Prof. Dr. Florian P. Seebeck

PD Dr. Daniel Häussinger

Basel, den 20.06.2017

Prof. Dr. Martin Spiess

Dekan der Philosophisch-

Naturwissenschaftlichen Fakultät

“Don’t Try”

-Charles Bukowski

Table of Contents

Abstract

Scientists have been fascinated by enzymatic catalysis for centuries and still marvel at the efficiency and selectivity with which these highly evolved molecular machines are able to catalyze complex reactions. For chemists, understanding the precise catalytic mechanisms can open up new possibilities in synthetic chemistry. From a medicinal standpoint, mechanistic knowledge provides valuable information on how enzymes can be targeted by drugs, as the molecular basis for many diseases lies in enzymatic malfunction. Among the different techniques used by researchers to investigate enzymatic catalysis, chemical probes are an indispensable tool. By synthetically inserting chemical changes into the substrate of a target enzyme, a wealth of information can be gained on structural and mechanistic features.

In this thesis we investigate three enzymes with the help of chemical probes. The first and second chapters focus on the first two steps in the biosynthetic pathway of ergothioneine, an essential thiohistidine with antioxidant properties. The third chapter focuses on a key enzyme in the maturation of sulfatases, inserting a unique and crucial catalytic residue into the latter's active site.

In the first chapter, we shed light on the molecular toolbox of the SAM-dependent methyltransferase EgtD. This enzyme catalyzes the first step in the synthesis of ergothioneine, the processive methylation of histidine to N^{α},N^{α},N^{α}-trimethyl histidine. By labelling the substrates with stable isotopes, we have created probes enabling us to decipher the molecular strategies which allow this enzyme to carry out three consecutive methylation steps with nearly equal efficiency.

In the second chapter we investigate the sulfoxide synthase EgtB, a non-heme iron dependent enzyme catalyzing the second step in ergothioneine biosynthesis. Interestingly, this system can potentially catalyze two distinct reactions, namely C-S bond formation and thiol dioxygenation. The factors governing this bifurcation in reactivity were investigated using substrate analogs with inhibitory properties.

In the third and final chapter we examine the formylglycine generating enzyme (FGE), a copper-dependent enzyme catalyzing oxidative C-H bond cleavage of a cysteine residue in the active site of sulfatases. By synthesizing substrate peptides carrying stereoselective deuterium-labels we have identified the rate limiting step of this reaction and have gained valuable insight into the geometry of the enzyme active site.

1 Introduction

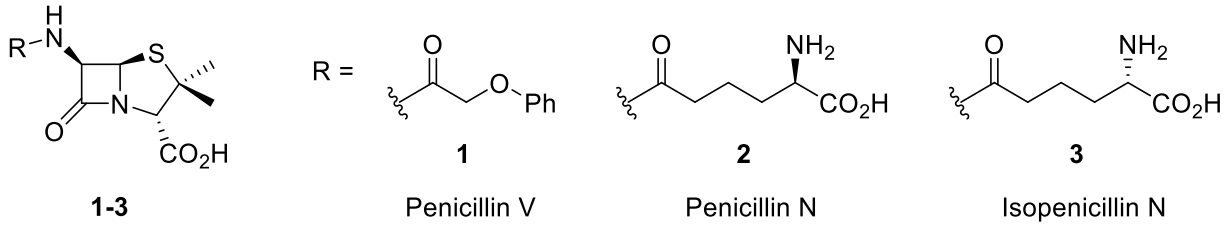
1.1 The Enigma of Enzymatic Catalysis

Enzymes can be considered the chemists of Nature, they are able to accelerate reactions by many orders of magnitude. It is hard to overstate the importance of the enhancement in rate that can be achieved by enzymes over the uncatalyzed reaction, as it is ultimately what makes life possible. Some of these reactions are so slow in the absence of enzymes, that the rates are hard to fathom and are indeed some of the lowest ever measured. For example, the half-time for arginine decarboxylation is estimated to be 1.1 billion years in neutral solution (2×10^{-17} s^{-1}).[1] The pyridoxal-dependent arginine decarboxylase from *Escherichia coli* accomplishes the same reaction with a k_{cat} of 1.4×10^{3} s^{-1}, resulting in an enhancement in rate of a baffling 19 orders of magnitude ($k_{cat}/k_{non} = 7 \times 10^{19}$).[2] While this enzyme uses a cofactor (PLP) to achieve this amazing feat, others do not and can catalyze reactions solely by interactions with their amino acid side-chains. But what is the chemical basis for this catalytic efficiency and how can it be elucidated? In order to understand this, enzymologists make use of a broad array of methods, with which it is possible to gain insights into the precise chemical steps in a given catalytic cycle. One of the most important advances has been the emergence of X-ray crystal structures of proteins, which give an accurate picture of an enzyme and its active site(s) in particular. While X-ray crystal structures give a wealth of information on the coordination of substrates and reactive groups, they ultimately provide a rather static view of the target enzyme. In order to reveal the dynamics of these processes, researchers turn to kinetic analysis, tracking enzymatic turnover as a function of time. One very powerful method with which one can gain deeper insights into the mechanistic details is by altering the substrate of the enzyme of interest. By decorating the basic scaffold of a substrate with different functional groups or isotopic labels, one can alter its reactivity. Subsequent analysis of reaction rates and alternative products offers keys for a detailed understanding of the activity of these truly amazing catalysts. The knowledge gained from these data is of utmost importance in the design of new drugs as well as new synthetic catalysts used by chemists around the world on a daily basis.

In the following we will highlight one classic and one more recent example of enzyme mechanisms, which have been elucidated with the help of brilliantly designed experiments involving synthetically modified substrates.

1.2 Isopenicillin N Synthase

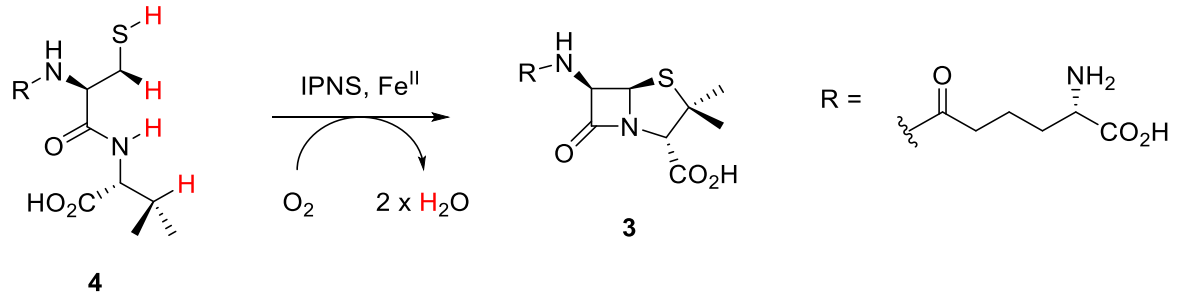
The discovery of penicillin in 1928 by Alexander Fleming was arguably one of the most significant advances in modern medicine, paving the way for the introduction of antibiotics into our health system. While penicillins constitute a wide array of structures, all contain the central β-lactam moiety, which is responsible for the inhibition of cell wall biosynthesis in the target bacterial organism. The precise structure of penicillin was the subject of much debate and the β-lactam ring was only conclusively shown to be a structural feature of the molecule by crystallographic data put forth by Dorothy Crowfoot Hodgkin in 1945 for which, among other achievements, she received the Nobel Prize in 1964.[3] For chemists the bicyclic core structure containing the thiazolidine ring and the highly strained four-membered β-lactam ring is of high interest and a challenging target for synthesis as it contains multiple stereocenters along with the labile, distorted amide bond in the four-membered ring. The first total synthesis of penicillin V (**1** in Figure 1) was published by John Sheehan in 1957 after the development of the *N*,*N*'-dicyclohexylcarbodiimide (DCC)-mediated coupling of carboxylic acids with amines to form amides, which solved the problem of closing the β-lactam ring.[4,5]



1-3 | **1** Penicillin V | **2** Penicillin N | **3** Isopenicillin N

Figure 1 Structure of the bicyclic core of penicillin containing the four-membered β-lactam ring and the five membered thiazolidine ring.

What chemists had struggled with for many years and required multiple chemical transformations to achieve, Nature had worked out how to accomplish in one single step: the enzyme isopenicillin N synthase (IPNS) catalyzes both ring closures starting with the linear tripeptide *L*-*δ*-(*α*-aminoadipoyl)-*L*-cysteinyl-*D*-valine (*LLD*-ACV, **4**). This non-heme iron-dependent oxidase harnesses the full four-electron oxidizing power of dioxygen without the need for additional electron donors or oxidizable co-substrates.[6] Four hydrogen atoms are thus abstracted from the linear peptide, converting one equivalent of oxygen into two equivalents of water, as shown in Figure 2.



4 **3**

Figure 2 Double ring-closure of the linear tripeptide *L*-*δ*-(*α*-aminoadipoyl)-*L*-cysteinyl-*D*-valine (*LLD*-ACV) forming isopenicillin N (**3**) as catalyzed by the iron-dependent oxidase isopenicillin N synthase (IPNS) with the four abstracted hydrogens highlighted in red.

One of the pioneers elucidating the biosynthesis of penicillins by IPNS is Sir Jack Baldwin, who with his research group at the University of Oxford, published a plethora of research articles on the topic.[7] By designing brilliantly crafted, chemically modified tripeptide substrates they were able to gain many important insights into the reaction mechanism before any crystal structures of IPNS was reported with or without substrate bound to the active site.[8,9] Different segments of the substrate were altered, and the countless variants were tested for their viability as substrates, with precise analysis of the rate of formation and structures of the corresponding products, if indeed the substrates were accepted by the enzyme. A selection of the experiments conducted, along with the respective findings, shall be highlighted in the following sections.

1.2.1 Mapping of the Active Site of IPNS with Aminoadipoyl Sidechain Variants

Variations of the *δ*-*L*-*α*-aminoadipoyl chain, which is not directly involved in the formation of the bicyclic system, proved to be generally very well tolerated by the enzyme.[10] The circumstance that most of the accepted variants contained a carboxylate group six carbons removed from the ε-carbonyl group led to the hypothesis that this group must be stabilized by H-bonding interactions with a remote Arg or Lys residue of the active site of the enzyme. Even quite bulky substrates containing aromatic “linkers” between carboxylate and ε-carbonyl were accepted for turnover (**7** and **8**). On the other hand, the acceptance of both *L*- and *D*-isomers (**5** and **6** respectively) at the aminoadipoyl terminus and the acceptance of adipic acid suggested that the amino group is not significantly involved in binding to the residues of the catalytic site of the enzyme. In contrast, substrates lacking the carboxylate group (**9** and **10**) or consisting of longer (**11**) or shorter (**12**) carbon chains were not recognized by the enzyme.

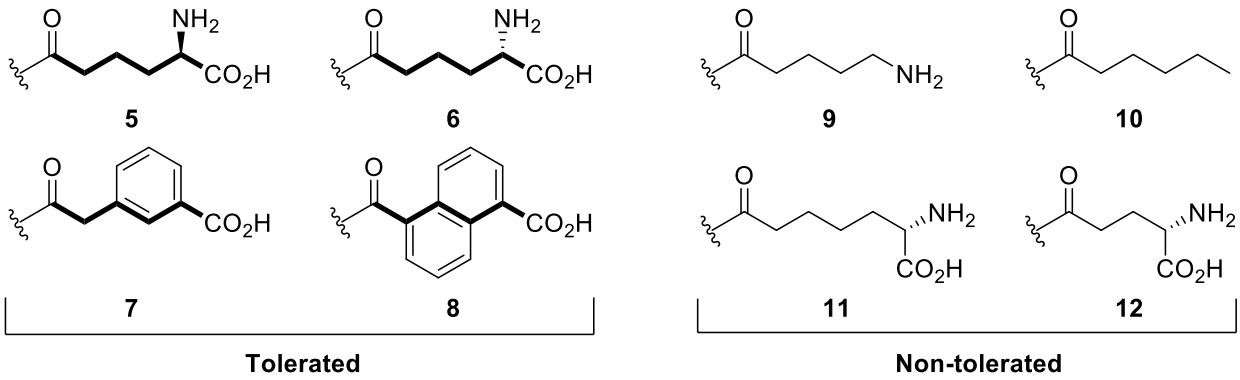


Figure 3 A selection of side chain variations of the *δ-L-α*-aminoadipoyl moiety in the linear peptide *LLD*-ACV, the rest of the peptide is analogous to that shown in the previous figure. The side chain on the top left is that of the natural substrate.

When the crystal structure of IPNS in complex with the substrate was eventually published roughly ten years later, it corroborated these findings nicely, showing that the carbon chain is in an extended conformation and that Arg87 hydrogen bonds to the carboxylate group at a distance of 2.8 Å. The closest polar group to the α-amine is the hydroxyl group of Thr331, the two groups are however well outside the range for H-bonding (5.6 Å).[9]

1.2.2 Variations on *L*-Cysteinyl and Kinetic Isotope Effects

Modifications of the second residue, the *L*-cysteinyl moiety, were generally poorly tolerated or not tolerated at all. This was to be expected as this key residue is crucially involved in both ring-closure steps and led the authors to conclude that many interactions between the enzyme and this portion of the substrate took place. A set of experiments with stereospecific deuterium-labels on Cβ of the cysteine however revealed some compelling findings (Figure 4).[11]

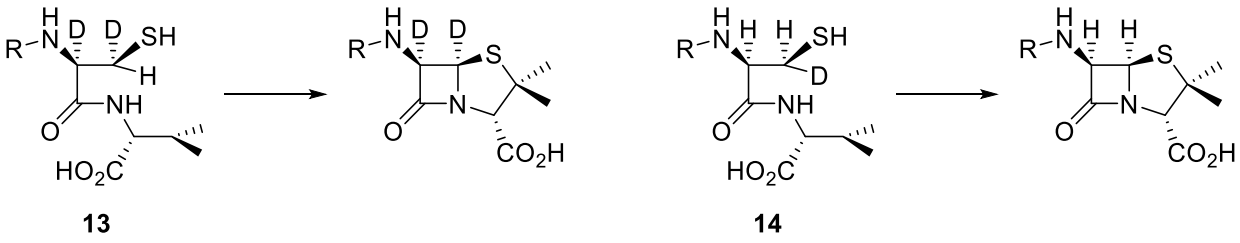


Figure 4 Conversion of stereospecifically deuterated variants of *LLD*-ACV to the corresponding products by IPNS with retention of the stereochemistry on Cβ. The R group is the same as that depicted in **Figure 2**.

Firstly, complete retention of the stereochemistry on Cβ was observed demonstrating that hydrogen abstraction is stereospecific. Secondly, the specificity of the abstraction was unaffected by a kinetic isotope effect (KIE) meaning that although this hydrogen abstraction is

rate limiting, the enzyme still preferentially abstracts the *pro-S* hydrogen.[12] These findings too were later corroborated by a quaternary crystal structure (IPNS-Fe^{II}-ACV-NO), where the *pro-S* hydrogen points in the direction of a molecule of NO bound to the iron center and serving as a placeholder for the abstracting oxygen species.[9]

1.2.3 Sequence of Ring Closures Elucidated by Kinetic Isotope Effects

Initially, the question of which ring closure takes place first was addressed by synthesizing the two most likely intermediates for each case and testing their viabilities as substrates or inhibitors. If C-H abstraction from Cβ of valine is the first step (pathway **a** in Figure 5) one would expect a seven-membered ring as an intermediate (**15**). This potential intermediate was synthesized, but behaved neither as a substrate nor as an inhibitor.[13] The second possible intermediate (**16**), where the β-lactam ring is formed first was more difficult to synthesize and could not be assessed as a substrate due to its inherent instability at physiological pH.[14]

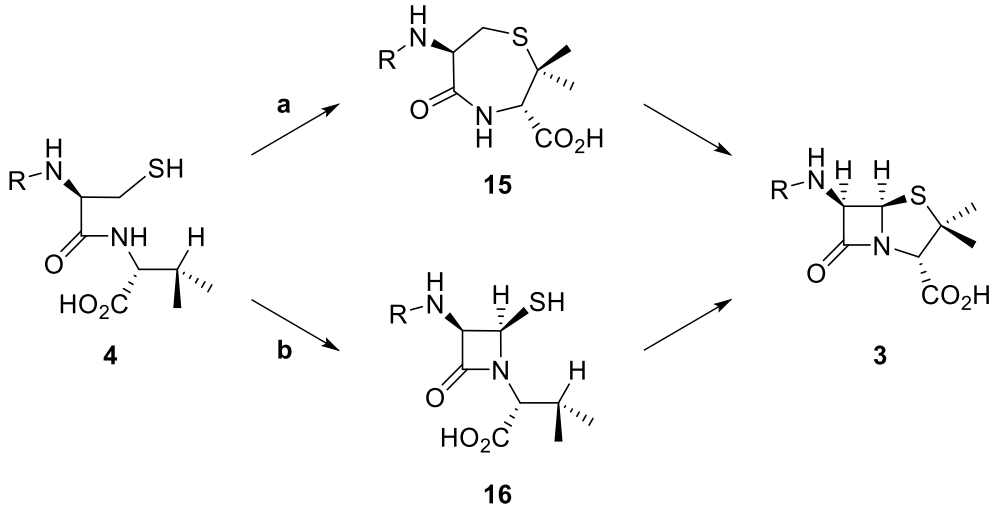


Figure 5 Reaction scheme for the order of ring closure(s) catalyzed by IPNS. The R group is the same as that depicted in **Figure 2**.

Since all other attempts of characterizing possible intermediates failed,[15,16] other techniques needed to be applied in order to gather mechanistic information. As discussed above, a KIE had been observed upon hydrogen abstraction from Cβ of the cysteinyl residue. Additionally a second, larger KIE was observed for the second C-H bond cleavage step from Cβ of the *D*-valinyl segment.[17]

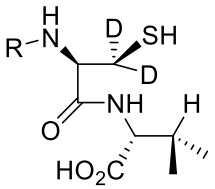
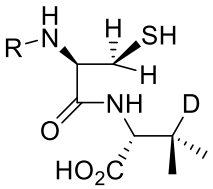
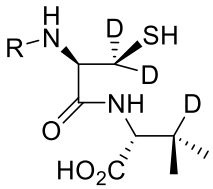
Substrate			
	17	**18**	**19**
KIE	**5.6 ± 0.3**	**13 ± 2**	**18 ± 2**

Figure 6 Deuterated substrates and the kinetic isotope effects observed for the conversion of each into the respective products. The R group is the same as that depicted in **Figure 2**.

In a first series of experiments a 1:1 mixture of deuterated tripeptide (either on Cβ of the cysteinyl or valinyl residue) and unlabeled peptide were reacted with IPNS and the ratio of labelled and unlabeled starting material were compared at different time points. Only in the case where the deuterium label was on the cysteinyl residue was an enrichment of the deuterated species observed; the KIE is hence an effect on k_{cat}/K_M.[18] In contrast to this finding, all three labelled substrates shown in Figure 6 showed significant KIEs on k_{cat} where a single substrate was used at saturating concentration. These findings led the researchers to form three major conclusions:

a) The reaction occurs stepwise and not in one concerted step.
b) Cleavage of the C-H bond on the cysteinyl residue takes place first, but must be similar in energy to the activation barrier for the second C-H bond cleavage step.
c) The intermediate formed remains bound to the enzyme active site.

This last point was consistent with the finding of Baldwin *et al.*[15] as well as others[16] that no intermediate could be detected, let alone isolated.

1.2.4 Variations on the *D*-Valinyl Moiety

The most exhaustive studies were conducted on the third amino acid in the tripeptide substrate, yielding a large amount of data and offering further valuable insights into the active site and the mechanism of ring closure. Some of the most important results shall be presented hereinafter.

When valine was substituted by α-aminobutyrate (**20**) a mixture of three products (**21**, **22**, **23** – ratio: 1:7:3) was formed, at first prompting the researchers to speculate whether a mixture of enzymes was responsible for this multiplicity of products. It was however shown that this was not the case and that IPNS can indeed catalyze branched pathways; the products **21** and **23** were both later discovered to be formed naturally in a species of *Streptomyces*.[19]

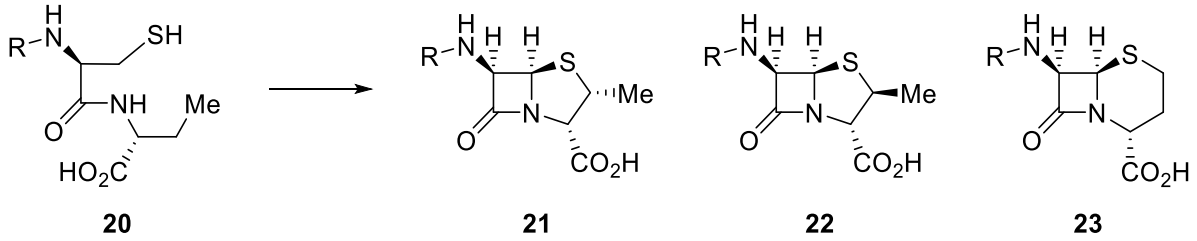


Figure 7 Distribution of products formed from the tripeptide *LLD*-ACV where valine was substituted by α-aminobutyrate. The R group is the same as that depicted in **Figure 2**.

When the stereochemistry of C-S bond formation was determined, a further surprising result was obtained. Reaction of both the (3*R*)- and (3*S*)-deuterated variants of **20** (**24** and **25**, respectively) with IPNS both resulted in the same monodeuterated product **26** (Figure 8).

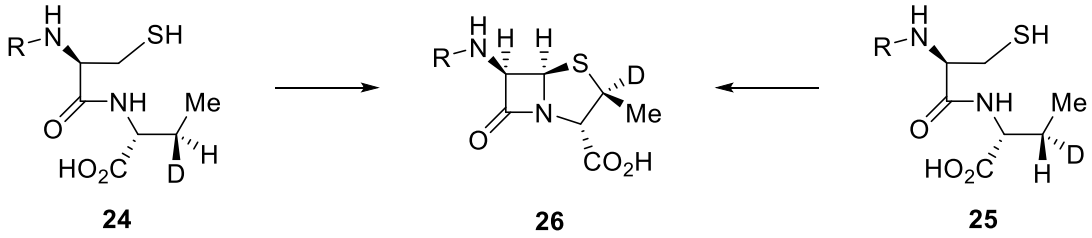


Figure 8 Single monodeuterated product formed from both deuterated α-aminobutyrate variants of the linear tripeptide substrate. The R group is the same as that depicted in Figure 2.

This behavior could be accounted for by a carbon radical forming on C3 after a substantial KIE, making hydrogen abstraction much more favorable than deuterium abstraction. A subsequent rapid equilibration (rotation around C2-C3) and slower ring-closure would result in the observed stereochemistry. Since a radical mechanism had been proposed previously,[20] it was attempted to validate this hypothesis using the cyclopropylcarbinyl test. As radicals of the latter mentioned species undergo rapid ring opening to the corresponding homoallyl radical, a tripeptide containing such a group (**27**) was synthesized and tested as a substrate. As could be predicted by a radical mechanism, a mixture of products (**28** and **29**) was formed, thus confirming what had been surmised from the results of the experiments described above.

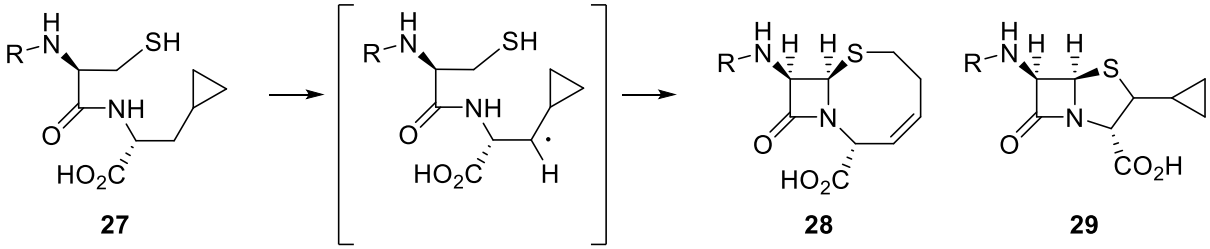


Figure 9 Distribution of products formed from the tripeptide *LLD*-ACV where valine was substituted by a cyclopropyl-containing moiety. The R group is the same as that depicted in Figure 2.

1.2.5 Proposed Catalytic Mechanism of IPNS, Then and Now

Along with some additional findings not presented in this summary, a catalytic cycle was proposed for the two cyclization reactions, where a highly reactive Fe^{IV}-oxo species is formed during the first ring closure.[7,17] This species would then be of sufficient energy to perform the less favorable H-abstraction from Cβ of the *D*-valinyl segment as shown below in Figure 10.

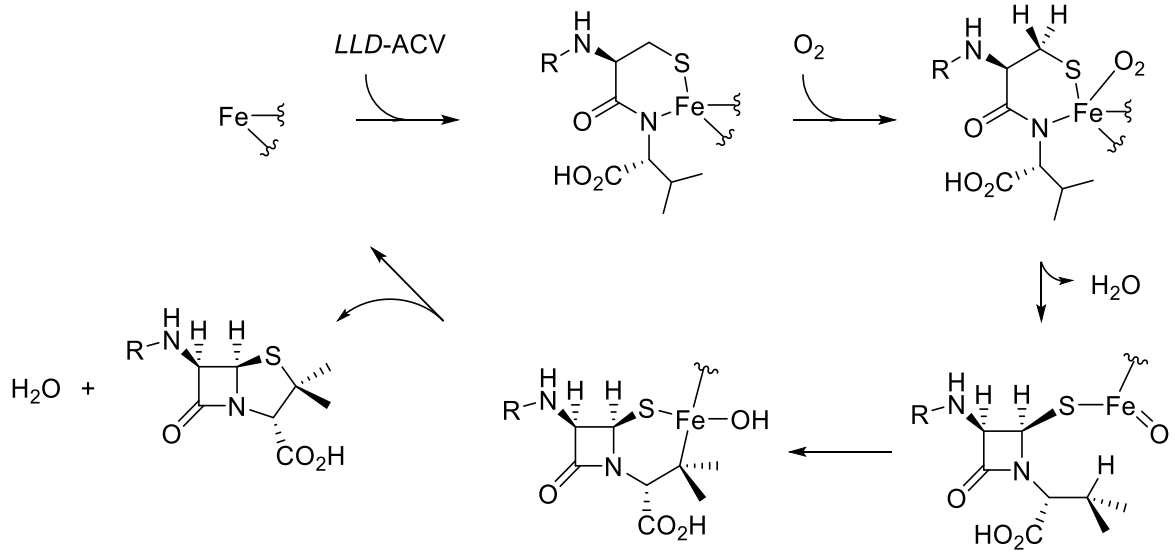


Figure 10 The catalytic mechanism of both ring closures by IPNS as proposed by Baldwin *et al.*[17] The R group is the same as that depicted in **Figure 2**.

In the decades that followed, countless more experiments were conducted by many research groups including the solving of X-ray crystal structures, DFT and QM/MM calculations as well as spectroscopic studies. These contributions further refined the proposed mechanism, but to this day the basic consensus remains unchanged. Most recently, by a combination of rapid-freeze-quench Mössbauer spectroscopy and UV-Vis spectroscopy, both abstracting iron intermediates could be detected.[21] The newest iteration of the catalytic mechanism is shown in Figure 11, the two hydrogen abstracting Fe^{III}-superoxo and Fe^{IV}-oxo species are displayed in structures **B** and **F** respectively.

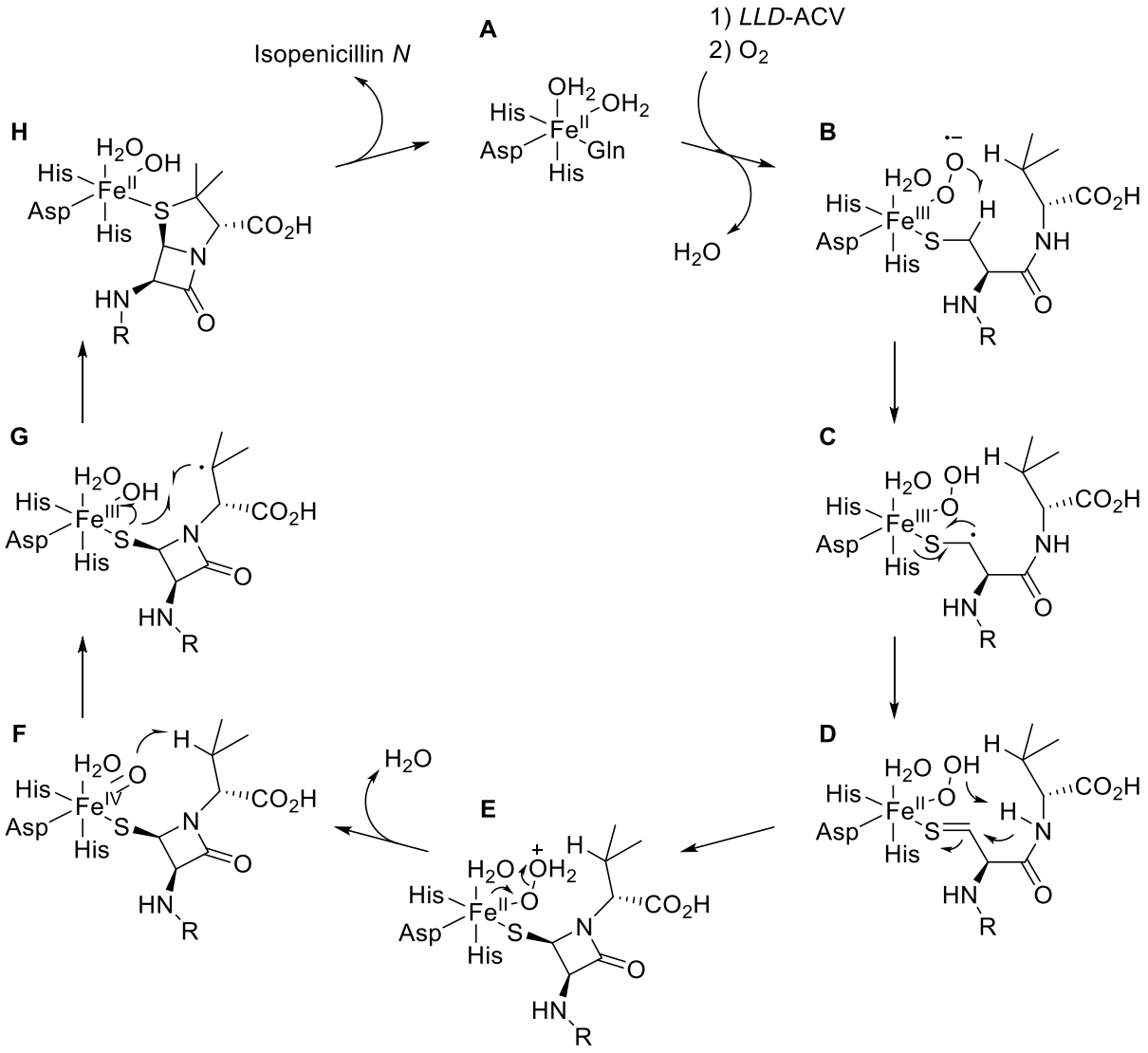


Figure 11 Proposed mechanism for the catalytic cycle of IPNS. The R group is the same as that depicted in **Figure 2**.

After sequential binding of the tripeptide and oxygen to the iron center, the first part of the catalytic cycle is initiated by hydrogen abstraction from Cβ of cysteine (**B**). After an inner-sphere electron transfer from the thiyl radical to Fe^{III} (**C**) an intermediate thioaldehyde coordinated to an Fe^{II}-hydroperoxo species is formed (**D**). This species can then abstract the amide hydrogen from valine; stereocontrolled attack of this deprotonated amide onto the iron-coordinated thioaldehyde forms the four-membered β-lactam ring and completes the first half of the catalytic cycle. Loss of the first equivalent of water with the first two abstracted hydrogen atoms leads to the highly reactive Fe^{IV}-oxo (ferryl) complex (**F**), which in turn initializes the second half of the catalytic cycle *via* abstraction of the Cβ hydrogen of valine. The generated radical then attacks the iron-coordinated sulfur (**G**), forming the thiazolidine ring which remains bound to the iron center (**H**) before being released upon displacement by the original Gln ligand, returning the enzyme to its resting state (**A**).

A complete picture of this catalytic cycle could not have been obtained without the help of crystallographic, spectroscopic and computational techniques. However, it was the assiduous work of the chemists and enzymologists who designed and crafted the molecular tools, analyzed the data and came up with hypotheses based on chemical rationale that contributed an indispensable amount of knowledge to the understanding of this enzyme.

1.3 Terpene Cyclases

For a more recent example of synthetically altered substrates used as probes for enzyme catalysis, we will turn to the catalysis performed by terpene cyclases and particularly the research group of Jeroen Dickschat. Along with his co-workers, he has developed some ingenious approaches of using stable isotope labelling to elucidate natural product structures as well as the mechanism of the enzymes involved in their biosyntheses. While this technique has been applied to a broad variety of natural products,[22] terpenes are of particular interest as they are a large class of natural products and have very intricate, stereospecific, often (poly)cyclic structures, which arise from some of the most complex enzymatic conversions found in Nature.[23] Terpenes were originally discovered in plants, starting in the late 19th Century with the pioneering work of Otto Wallach, who also noted that all products of this class are made up of repeating units of isoprene.[24] Only in the second half of the 20th Century did it become evident that terpenes were also produced by fungi and bacteria. Even more recently, this ubiquitous class of natural product has been discovered in certain eukaryotic microorganisms such as social amoebae.[25]

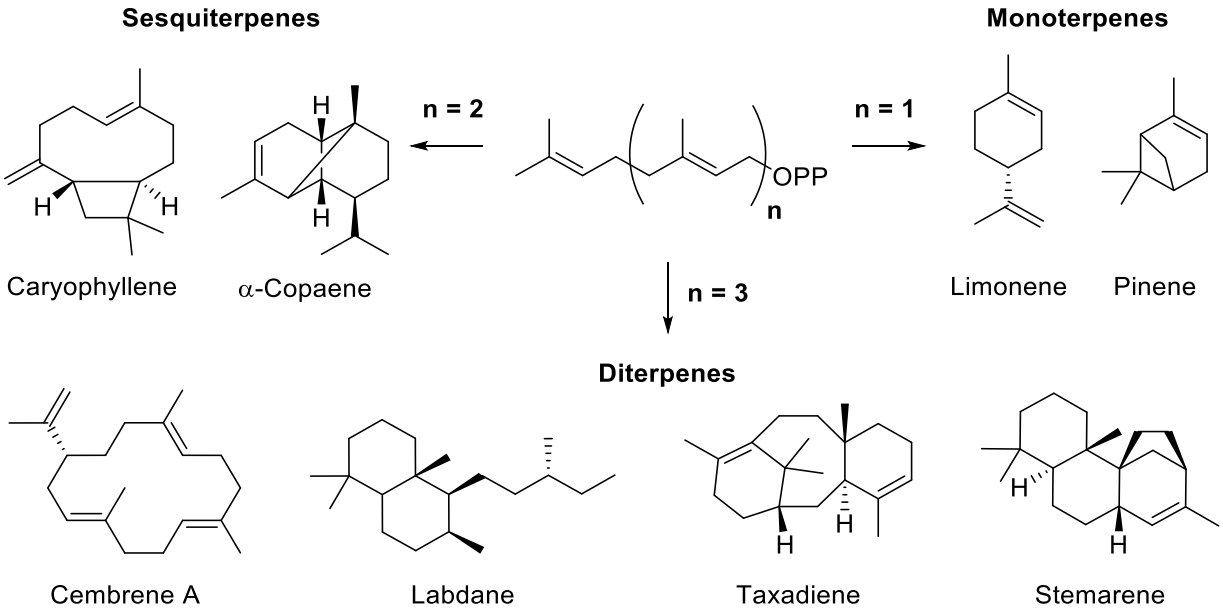


Figure 12 Chemical structures of a selection of naturally occurring terpenes. Monoterpenes are biosynthesized from geranyl pyrophosphate (n = 1), sesquiterpenes are biosynthesized from farnesyl pyrophosphate (n = 2) and diterpenes are biosynthesized from geranyl geranyl pyrophosphate (n = 3).

In a recent publication by Dickschat *et al.* the products and the mechanisms of cyclization were elucidated for two different bacterial diterpene cyclases using a combination of stable isotope labelling and NMR spectroscopy (Figure 13).[26] A putative terpene cyclase from *Streptomyces violens* was recombinantly produced and shown to convert geranyl geranyl pyrophosphate (GGPP) into a diterpene hydrocarbon, while geranyl (GPP) and farnesyl pyrophosphate (FPP) were not turned over. The structure of the product was identified as (-)-spiroviolene (**30**) by NMR spectroscopy, albeit with unclear absolute configuration, and the enzyme was termed spiroviolene synthase (SvS).

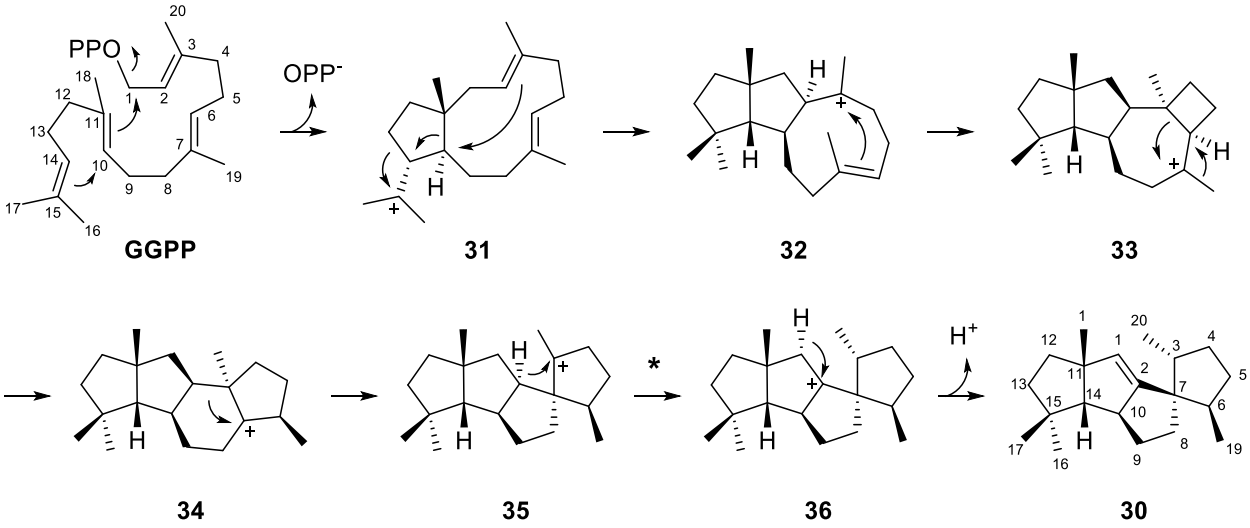


Figure 13 Proposed cyclization of GGPP to form **1** as catalyzed by spiroviolene synthase (SvS). The numeration of all carbon atoms in **1** corresponds to their origin from GGPP. The star (*) indicates a 1,3-hydride shifts.

In the proposed biosynthetic pathway, the reaction is initiated by two concerted ring closures of GGPP, forming the bicyclic cationic intermediate **31**. A subsequent double Wagner-Meerwein rearrangement and further cyclization step results in the tricyclic species **32**. Further cyclization and rearrangements lead to **33** and **34** respectively, followed by a ring contraction to form the spirocyclic cation **35**. Finally a 1,3-hydride migration leads to **36**, deprotonation of which results in the final product **30**.

Using a clever combination of synthetic and biosynthetic techniques, all 20 isotopomers of ($^{13}C_1$)GGPP were synthesized and converted to the corresponding products with SvS in order to assign where the different carbons originating from the starting material were located in the product. All carbons were determined to be in the expected positions predicted by the proposed mechanism shown in Figure 13. Next, the absolute stereochemistry of the product was assigned with GGPP carrying a stereoselectively introduced deuterium label on a ^{13}C-labelled methylene-group. These in turn were synthesized *in situ* by the GGPP synthase (GGPPS) from isoprene pyrophosphate (IPP) and (*R*)- or (*S*)-(1-^{13}C, 1-^{2}H)FPP and directly converted to the spiroviolene product by SvS as shown in Figure 14. Considering that elongations of IPP by oligoprenyl diphosphate synthases are known to invert the configuration at C1,[27] the absolute stereochemistry of the product could be deduced based on an HSQC analysis of the product. The same experiment was repeated with the ^{2}H/^{13}C-label in a different position, resulting in the same absolute stereochemistry.

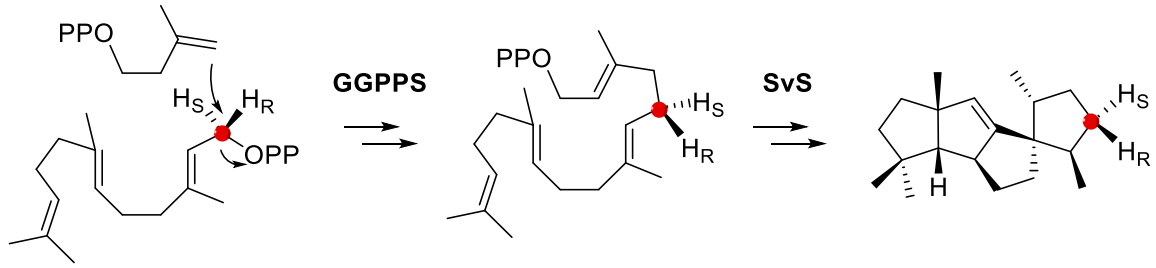


Figure 14 Synthesis of (*R*)- and (*S*)-(5-^{13}C, 5-^{2}H)GPP from (*R*)- or (*S*)-(1-^{13}C, 1-^{2}H)FPP and IPP by GPPS, followed by cyclization to **30** by SvS. The red dot indicates the position of the ^{13}C-label.

Next, the proposed 1,3-hydride shift (**35** to **36** in Figure 13) was investigated using (3-^{13}C, 2-^{2}H)GGPP as shown in Figure 15. With this starting material, the deuterium would be located on the ^{13}C-labelled carbon atom in the final product if indeed the purported shift took place. As predicted, this was the case, which was underlined by a triplet in the ^{13}C NMR spectrum, caused by ^{2}H-^{13}C spin coupling in the product.

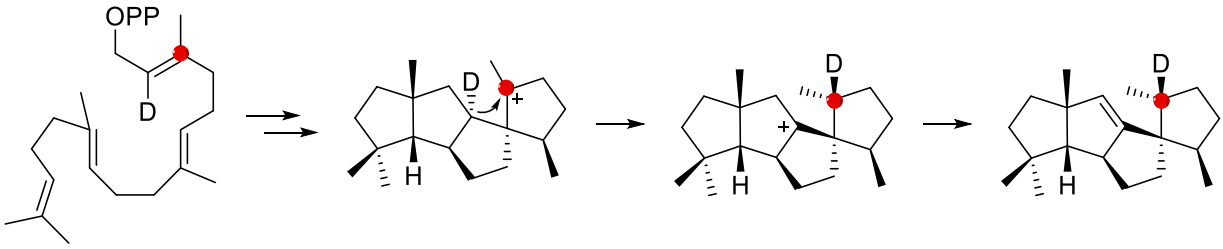


Figure 15 1,3-Hydride shift of deuterium onto the ^{13}C-labelled carbon, starting from (3-^{13}C, 2-^{2}H)GGPP

Lastly, the stereochemistry of the final deprotonation step was examined, using stereoselectively deuterium-labelled (*R*)- and (*S*)-(1-^{2}H)GGPP (Figure 16). In the case of the *S*-enantiomer, loss of deuterium was detected by GC/MS whereas loss of protium was detected for the *R*-enantiomer as a mass difference of 1 Da. It is thus the hydrogen *syn* to the migrating hydride that is abstracted and the authors suggest that the abstracting base is the diphosphate anion that is eliminated from C1 of GGPP in the initial ring closure (Figure 13). They argue that since it leaves C1 towards the back of the projection plane, it is ideally positioned for this final proton abstraction, the stereochemistry of which has been confirmed.

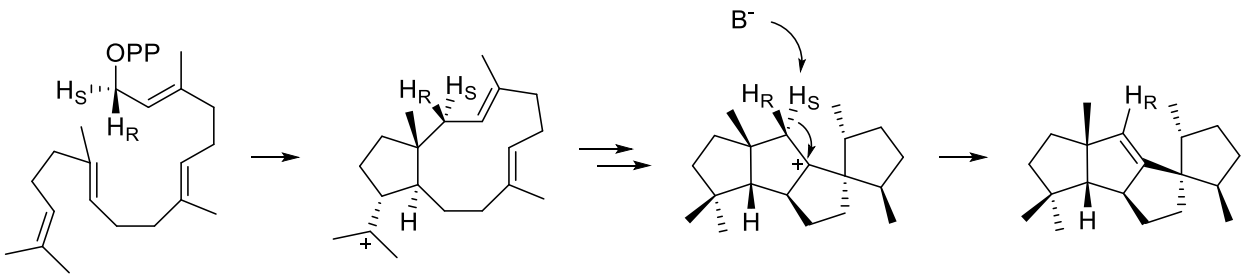
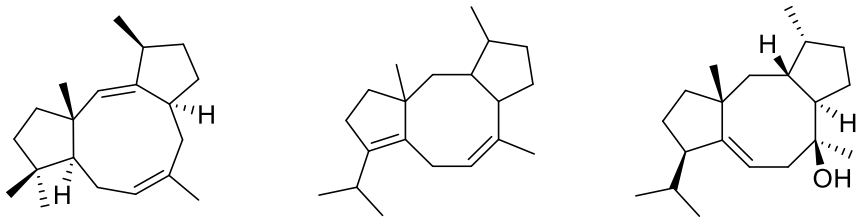


Figure 16 Stereochemistry of the final deprotonation step, showing how H_S is abstracted by an unknown base (B^-).

With the sequence of SvS, a BLAST search was conducted, aiming to identify closely related enzymes. Among others these included the cyclases for tsukubadiene (**37**) and cyclooctat-7(8),10(14)-diene (**38**), the latter of which displayed a similar structure to cyclooctat-9-en-7-ol (**39**) from a phylogenetically distant cyclase



37 **38** **39**

Figure 17 Structures of tsukubadiene (**37**), cyclooctat-7(8),10(14)-diene (**38**) and cyclooctat-9-en-7-ol (**39**). The first two terpenes are produced by enzymes closely related to SvS, while the last is not.

Since the cyclization mechanisms for **38** and **39** can be assumed to be quite similar and the cyclization of **39** is known to proceed through a unique cyclopropyl intermediate,[28] the mechanism for cyclization of **37** was investigated starting from a similar hypothesis.

The enzyme tsukubadiene synthase (TdS) from *Streptomyces tsukubanensis* was recombinantly produced and an analogous series of experiments was conducted as described above. After showing that the enzyme produces **37** from GGPP, the relative and absolute stereochemistry was determined by extensive NMR spectroscopy. A mechanism was suggested based on that proposed by Meguro *et al.* for **39**.[28]

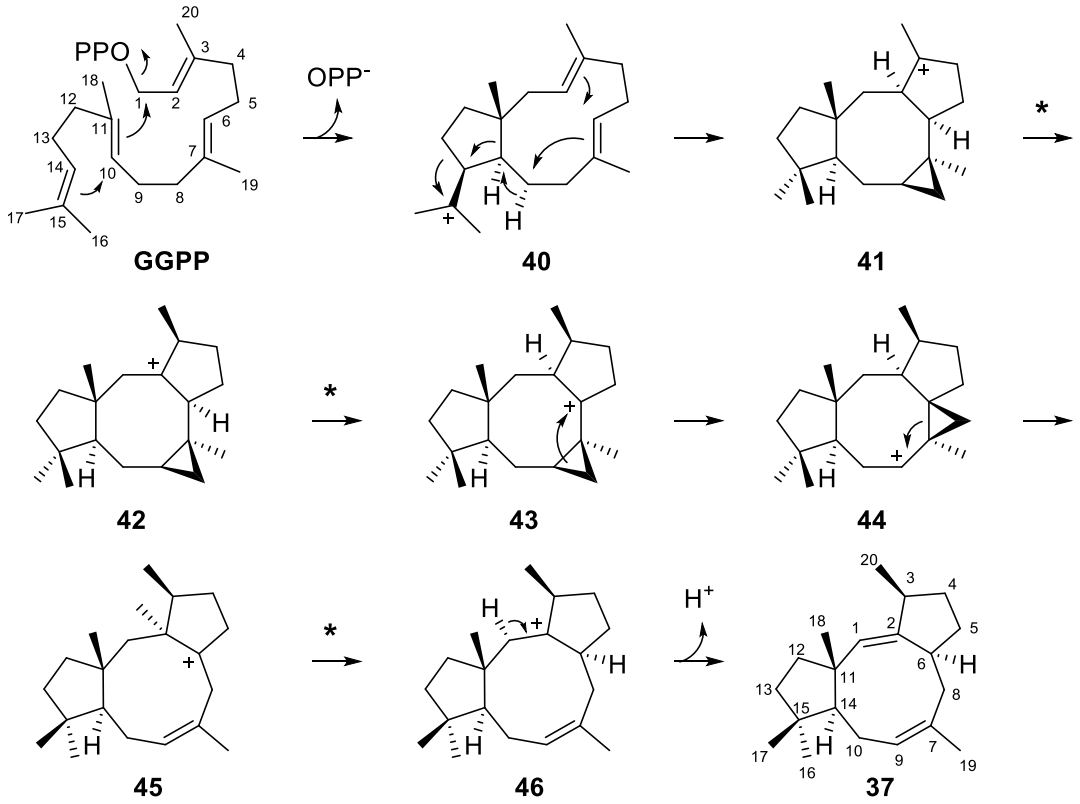


Figure 18 Proposed cyclization of GGPP to form **37** as catalyzed by tsukubadiene synthase (TdS). The numeration of all carbon atoms in **37** corresponds to their origin from GGPP. The stars (*) indicate 1,2-hydride shifts.

The catalytic cycle is initiated by the same two ring closures as described above, resulting in the conversion of GGPP to **40**, a diastereomer of **31**. Ring expansion and contraction by Wagner-Meerwein rearrangement, a 1,2-hydride migration and 7,9- as well as 2,6-cyclization lead to the formation of **41**. Two additional 1,2-hydride migrations take place via **42** to give intermediate **43**. A cyclopropane ring rearrangement produces **44**, similar to what was proposed for the biosynthesis of **39**.[28] A subsequent ring opening gives **45**, followed by a further 1,2-hydride migration forming **46** and finally deprotonation to yield the product **37**.

As for spiroviolene (**30**), the mechanism was confirmed by identifying the positions of each carbon in both the substrate and the product by converting all 20 isotopomers of ($^{13}C_1$)GGPP to the product. The absolute stereochemistry of the product was determined by HSQC using stereospecifically deuterated substrates. The 1,2-hydride shift from **40** to **41** was confirmed by $^{2}H/^{13}C$-labelling of the substrate and NMR analysis of the product and the two 1,2-hydride shifts from **41** to **42** and **42** to **43** were confirmed in the same manner. Finally, the stereochemistry of the final deprotonation step was investigated with (*R*)- and (*S*)-(1-^{2}H)GGPP, leading to the conclusion that H_S is abstracted just as in the case of spiroviolene (**30**). Again, the phosphate anion lost in the first cyclization step is an ideal candidate as a base for the same reasons described for the final deprotonation of **36** to **30**.

In summary, Dickschat *et al.* have shown just how useful stable isotope labelled substrates are for the elucidation of terpene cyclization. Detailed insights into the mechanism were obtained by tracking every single carbon atom from the linear precursor to the polycyclic products and the absolute stereochemistry was solved by incorporation of stereospecific deuterium labels. It is of particular interest to note that the intermediate bicyclic cations formed after the initial cyclization reaction (**31** and **40**) are in fact diastereomers. Only the further course of the reaction, determined by the folding and positioning of the intermediates in the active site, yields the structurally quite diverse products **30** and **37**. In the final deprotonation steps, leading from **36** to **30** and from **46 to 37**, nonetheless take place from the same face in which the diphosphate is eliminated from the starting material. This in turn provides further evidence that the latter does in fact act as the base in the final step.

As in the case of IPNS and the research conducted by Baldwin *et al*. a wealth of information was gained by intelligently crafting appropriate synthetic probes, followed by careful analysis of the results based on chemical rationale.

1.4 Biosynthetic Transformations Investigated in this Thesis

1.4.1 Biosynthesis of Thiohistidines; Ergothioneine and Ovothiol

Ergothioneine (**ET** in Figure 19) is a redox-active sulfur-containing betaine produced in cyanobacteria,[29] mycobacteria and fungi.[30] In humans it is actively taken up from dietary sources by the specific ergothioneine transporter protein OCTN1.[31] Polymorphisms in the gene *SLC22A4*, encoding for OCTN1, have been linked to rheumatoid arthritis,[32] ulcerative colitis,[33] Type I diabetes[34] and Crohn's disease.[35] In certain tissues such as erythrocytes, seminal fluid, liver and kidney it is accumulated to near millimolar concentrations and in millimolar concentration in the lens of the human eye. Although the physiological role of ergothioneine in humans has not yet been fully elucidated, it is thought to play a role in protection from oxidative stress, as it was shown to be an antioxidant *in vitro*.[36] The enzymes involved in the biosynthesis of ergothioneine by *Mycobacterium smegmatis* were discovered in 2010 and are summarized in Figure 19.[30] The first step is the processive trimethylation of histidine by the SAM-dependent methyltransferase EgtD, resulting in the histidine betaine hercynine (**TMH**). This is followed by the remarkable C-S bond formation between hercynine and γ-glutamyl cysteine on the 2-position of the imidazole ring by the iron dependent sulfoxidase EgtB (**47**). Subsequently, the glutamyl-rest is cleaved by EgtC (**48**) and the oxidized cysteinyl moiety is removed by the PLP-dependent EgtE, forming pyruvate and ammonia as leaving groups to give the final product

ergothioneine. In the fungus *Neurospora crassa*, a shorter synthetic route was discovered where an alternate variant of EgtB directly adds cysteine to the histidine scaffold.[37]

Ovothiol A (**OA** in Figure 19) is a related thiohistidine first discovered in sea-urchin eggs and thought to protect them from oxidative stress.[38] Later on it was also identified in several human pathogens such as *Leishmania donovani* and *Trypanosoma cruzi*.[39,40] A biosynthetic pathway for ovothiol A in *Erwinia tasmaniensis* was elucidated in 2011, involving two enzymes; OvoA and OvoB.[41] The first and key step in ovothiol biosynthesis is catalyzed by OvoA and involves C-S bond formation between the thiol of cysteine and C-5 of the imidazole ring of histidine (**49**). This step is analogous to the second step in ergothioneine biosynthesis, catalyzed by EgtB. In the second step, the PLP dependent OvoB cleaves the cysteinyl moiety resulting in 5-thiohistidine (**50**), which is finally methylated on N^{π} by the SAM-dependent methyltransferase domain of OvoA.

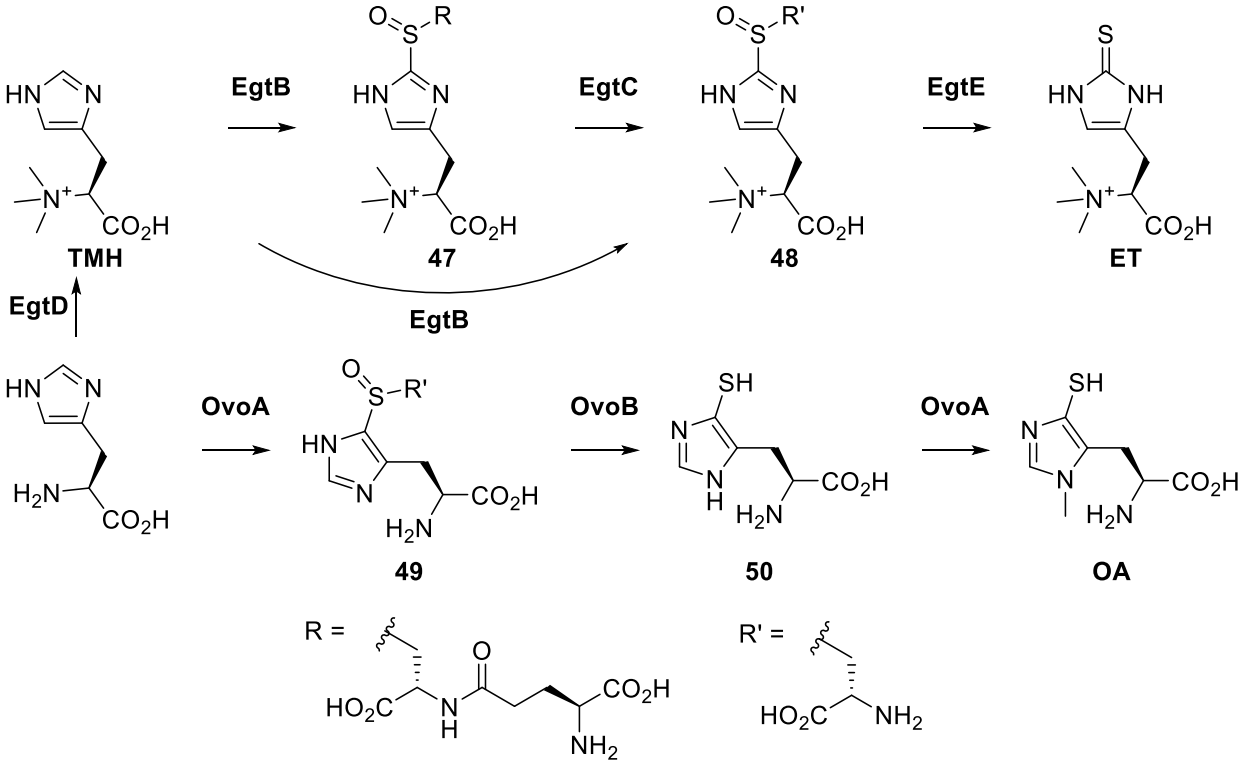


Figure 19 Biosynthetic pathways for the sulfur-containing antioxidants ergothioneine (**ET**) and ovothiol A (**OA**).

By using synthetically modified substrates we have furthered our understanding of two key enzymes in the biosynthesis of ergothioneine: EgtD and EgtB. In the first case, we synthesized ^{13}C-labelled substrates, which enabled us to investigate the binding mode of EgtD and characterize its strategies for substrate activation and catalysis. In the second case we synthesized unreactive substrate analogs with a modification on the 2-position of the imidazole ring, granting us novel insights into the reactivity of the catalytic iron center. Importantly, our findings shed new light on the role of a key active site residue and its role in catalysis.

1.4.2 Active Site Mapping of the Copper-Dependent Formylglycine Generating Enzyme

The formylglycine generating enzyme (FGE) is a copper-dependent enzyme, catalyzing the key step in the maturation of sulfatases, an enormously important class of enzymes in the metabolism of sulfate-esters. Mutations in the gene encoding FGE (*SUMF1*) have been linked to multiple sulfatase deficiency, a fatal orphan disease. Introduction of the unique formylglycine residue by FGE in the active site of sulfatases insures the latter's catalytic activity. Formylglycine is generated from a cysteine residue in a specific recognition sequence, which has also been engineered into other recombinant proteins, making FGE an attractive tool for biorthogonal conjugation reactions.

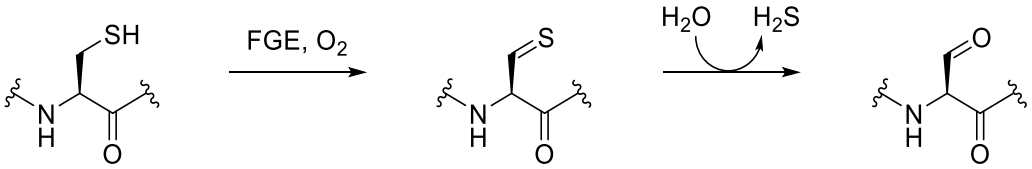


Figure 20 Schematic view of the generation of formylglycine from cysteine by FGE.

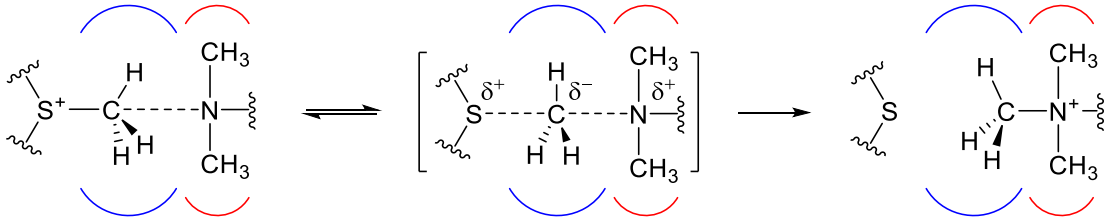
By synthesizing peptide substrates carrying a stereospecific deuterium label, we have created tools which have allowed us to characterize the rate limiting step of this reaction and effectively map the active site of this highly interesting enzyme.

2 Molecular Strategies of EgtD, a SAM-Dependent Methyltransferase

Abstract

Ergothioneine is an aromatic amino acid betaine derived from histidine, discovered over 100 years ago.[42,43] In this chapter we focus on the first step of the biosynthesis of ergothioneine, the cooperative trimethylation of histidine by the SAM-dependent methyltransferase EgtD.[30] A particular focus is placed on the binding mode of the methyl groups in its substrates and methyl donor. We aimed to elucidate the mechanism of this multistep reaction using a combination of X-ray crystallography, enzyme kinetics and NMR spectroscopy. By synthesizing compounds with specific ^{13}C-labels on the methyl groups of interest, we created powerful tools to analyze binding modes of the methyl donor as well as the respective acceptors via NMR.[44]

Herein we present a comprehensive description of the molecular mechanism that allows EgtD to methylate its primary, secondary and tertiary amine substrates with similar catalytic efficiencies.



2.1 Introduction

2.1.1 SAM-Dependent Methyltransferases

S-Adenosyl methionine (SAM) is a ubiquitous methyl group donor and is found in all kingdoms of life.[45] While some enzymes utilize SAM to form a 5'-deoxyadenosyl radical,[46] the largest group of SAM-dependent enzymes is that of SAM-dependent methyl transferases (MT), which catalyze methylation reactions. The targets of this reaction range from large macromolecules, such as proteins in post-translational modification, to small molecules.[47] In order to achieve maximal catalytic efficiency for this textbook S_N2-reaction, the coordination and positioning of the transferred methyl group in the active site of MTs is of particular interest.[48]

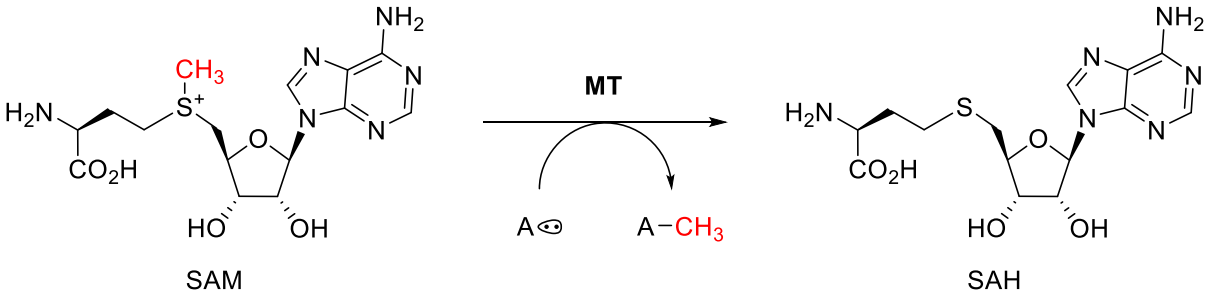


Figure 21 Basic methylation reaction catalyzed by a SAM–dependent methyltransferase. The methyl group from SAM is transferred onto a suitable acceptor (A) generating *S*-adenosyl homocysteine (SAH).

2.1.2 Catechol *O*-Methyltransferases

One system where the mechanism of methyl transfer from SAM onto the respective acceptor has been studied in great detail using a variety of experimental techniques is the enzyme catechol *O*-methyltransferase (COMT), which catalyzes methyl transfer from SAM onto dopamine and related catechols.[49]

Figure 22 Methyl group transfer from SAM onto dopamine, forming 3- or 4-O-methyl-dopamine, catalyzed by COMT.

Several theories exist on how this methyltransferase achieves its catalytic efficiency, speeding up the uncatalyzed reaction by an impressive factor of approximately 10^{16}. One of the first findings was that secondary KIEs with a deuterated methyl substrate were more inverse in the enzyme than in a model small molecule system in solution. These findings were interpreted as evidence that the transition state in the model system, e.g. in solution, was less organized than in the enzyme. An inverse KIE of this type is observed if the vibrational modes of the methyl hydrogens are more constrained in the transition state, leading to destabilization of the deuterated species. This led to the proposal by Schowen *et al.* of transition state (TS) compression as an explanation for increased catalysis in the active site.[50] It should however be noted, that this idea has been challenged when it could not be supported by QM/MM calculations put forth by Ruggiero *et al*.[51] More recently, Klinmann *et al.* conducted a combination of experiments with point mutations in the active site and their effect on secondary KIEs, also arguing for active site compaction.[52] They suggested that the enzyme creates transient active site configurations that are stiffer than the activated complex in solution by dynamic movements in the enzyme. In a second study mutants of Tyr68, a conserved active site residue located behind the reactive sulfur of SAM, were compared to the wild type enzyme. Binding isotope effects and computational models indicated stabilization by possible hydrogen bonding to the methyl group and increase in the methyl donor-acceptor distance in the order WT<Y68F<Y68A<solution, which in the opinion of the authors represented further evidence of active site compaction.[53] Other earlier theories, such as the concept of a cratic effect have been corrected to account for electrostatic factors.[54] Cratic refers to the free energy required to align reacting groups into a geometry enabling a facile reaction.[55] Subsequently however, Warshel *et al.* argued that indeed the influence of electrostatic preorganization is the dominant factor that accounts for the origin of the catalytic power of COMT.[56] By presenting a computational model that can reproduce all the previously obtained experimental observations, they contradicted the NAC, compression and entropic models. In what amounts to a veritable argument between the most respected female enzymologist and a Nobel Prize laureate accusing each other of unjustly dismissing experimental evidence[53] and misinterpreting results,[57] it seems as though the last word on this topic has hardly been spoken.

2.1.3 Lysine Methyl Transferases

2.1.3.1 SET Domain Lysine Methyltransferases

Another interesting case where the precise molecular mechanisms promoting catalysis have been studied in great depth are lysine methyl transferases (KMTs). Methylation of protein lysine is a prominent post translational modification and among other things responsible for the methylation of histones, the protein components of chromatin which allow for the dense packing of DNA in eukaryotic cells.[58] Of the characterized proteins responsible for lysine methylation, all are SAM-dependent and all but one contain a so-called SET-domain (**S**u(var)3-9, **E**nhancer-of-zeste and **T**rithorax, not to be confused with the acronym for single electron transfer).[59] A decade ago it was suggested that the coordination of the methyl group on SAM in the SET domain is reliant on C-H···O hydrogen bonds and that these facilitate methyl transfer from the donor to the acceptor.[60] The SAM-dependent lysine methyltransferase SET7/9 was shown to coordinate the methyl group of its co-factor SAM via C-H···O bonding by a combination of NMR, crystallographic data and computational studies.[44] In this study by Horowitz *et al.* the transferrable methyl group in ^{13}C-methyl labelled SAM displayed a significant downfield shift along the proton axis of 0.8 ppm from the unbound (3.0 ppm) to the bound (3.8 ppm) spectrum. Calculation of the expected bound shift matched the measured values with an accuracy of 0.1 ppm. Some clues pointing to the possibility that C-H···O bonds were involved in substrate binding could be found in the crystal structure.

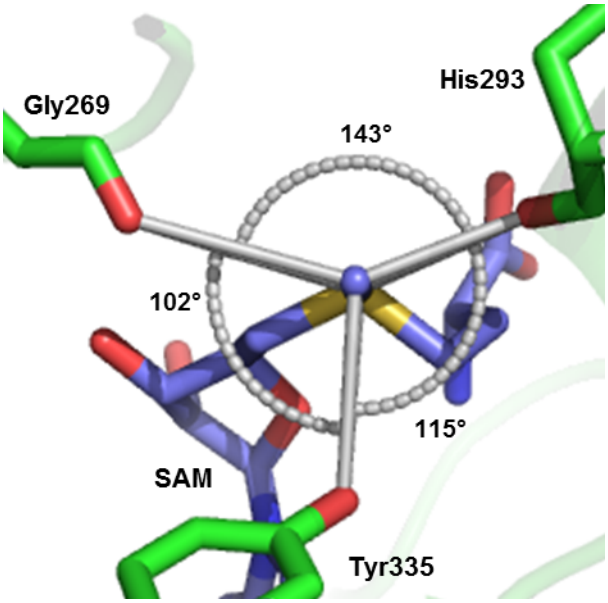


Figure 23 Crystal structure of SAM (in purple) bound to SET7/9 (PDB: 1N6A). The C···O distances between the methyl group on SAM and the coordinating oxygen atoms are 3.7 Å (Gly269), 3.0 Å (His293) and 3.5 Å (Tyr335)[61]

The C···O distances between the methyl group on SAM and two backbone carbonyls (Gly269 and His293) and an invariant tyrosine (Tyr335) were <3.7 Å and hence appropriate for

hydrogen bonding to occur.[60] Hydrogen bonds can form since the sulfonium cation in SAM is positively charged and withdraws electrons from the neighboring methyl group, polarizing the latter's C-H bonds. Finally, the angles between the three aforementioned residues were all close to 120°, approaching the CH_3-group in roughly trigonal coordination. It should be noted that in contrast to most other SET proteins, SET7/9 is exclusively a monomethylase.[62] This is largely due to steric constrictions in the active site, which prevent further methylation reactions from occurring.[63] Although a crystal structure of the SET7/9-SAH-MeLys ternary complex has been solved (PDB: 1O9S), no direct evidence for C-H···O hydrogen bonding to the methyl group on the acceptor exists besides the structure.

2.1.3.2 C-H···O Bonds

In addition to the more classical hydrogen bonds between N and O donors/acceptors, it was shown that carbon can act as hydrogen bond donors roughly 40 years ago.[64] Weak hydrogen bonding can even be observed for aliphatic carbons,[65] but when there is an adjacent polarizing atom the strength of these bonds can rival those formed by conventional donors/acceptors (N and O).[66] Since their discovery, this class of hydrogen bond has received more and more attention and has been shown to be important for enzyme stability and catalysis,[67] as well as stabilizing of nucleic acid structure[68] and interactions with methyl groups in small molecules.[69,70]

Methods of characterizing this type of interaction are somewhat limited, as crystal structures often fail to predict the exact position of hydrogen atoms. Many other current methods of identifying and characterizing C-H···O bonds are not easily applicable for protein-substrate interactions and obtaining experimental evidence remains a challenge.[71–73] One powerful tool for the description of these elusive connections is NMR-spectroscopy, where a substantial downfield ^1H chemical shift change can be observed upon binding.[74] In a number of systems C-H···O bonds have been characterized by NMR. Several recent examples include i) serine protease, where C-H···O bonds within the enzyme where described;[75] ii) maltose anomers, which were analyzed by solid state NMR and computational studies;[76] iii) Gly- and Ala-dipeptides, whose spectroscopic patterns were elucidated by purely computational studies.[77]

2.1.4 Glycine Betaine

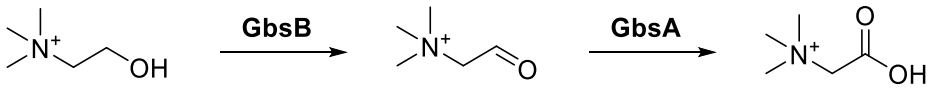
As discussed above, any polarized methyl group can potentially engage in C-H···O bonds. One ubiquitous class of natural products containing such polarized methyl groups is that of betaines. In the following section, some of the strategies of enzymes and proteins for betaine

binding shall be discussed based on the example of one of the most widespread exponents of this class of natural products.

Trimethyl glycine or glycine betaine was the first betaine to be discovered almost 150 years ago. Extracted from the common beet (*Beta vulgaris*, hence the name betaine) and from molasses in the year 1869 by Carl Scheibler,[78] it has since been discovered in most living organisms such as microorganisms, plants and mammals,[79,80] acting as an osmolyte,[81] a methyl donor in homocysteine metabolism[82] and possibly an antioxidant.[83] In humans, glycine betaine is taken up through an array of dietary sources such as seafood, wheat germ or bran and spinach.[79] There are two possible biosynthetic routes in organisms that produce it and these will be examined in more detail with particular emphasis on how the trimethylammonium group is bound by proteins. Importantly, the respective proteins utilize significantly different binding modes depending on whether transport of the target molecule or catalysis is the main function.

2.1.4.1 Glycine Betaine Biosynthesis by Choline Oxidation

The most common pathway for betaine biosynthesis is by oxidation of choline, either by a single enzyme or a two-enzyme system.[84] In order to accumulate choline in the cells, membrane-bound transporter proteins are necessary. The Gram-positive bacteria *Bacillus subtilis*, where glycine betaine serves as the preferred osmoprotectant,[85] takes up choline through the specific transporters OpuB and OpuC.[86] Following uptake, choline is oxidized to glycine betaine by the concerted actions of the alcohol dehydrogenase GbsB and the glycine betaine aldehyde dehydrogenase GbsA.[87] As shown in Figure 24, betaine aldehyde is the intermediate product.



51 **52** **53**

Figure 24 The biosynthetic pathway of glycine betaine (5**3**) from choline (5**1**) via the intermediate glycine betaine aldehyde (5**2**) by the alcohol dehydrogenase GbsB and the betaine aldehyde dehydrogenase GbSA.

The OpuB transporter (EC 3.6.3.32) from *B. subtilis* contains the two-domain trimethylcholine transporter protein OpuBC. The trimethylammonium group of the choline ligand is accommodated and coordinated by an aromatic cage, formed by two tyrosine residues from each domain, via cation-π interactions (Tyr71 and Tyr221 of domain I and Tyr117 and Tyr197 of domain II).[88]

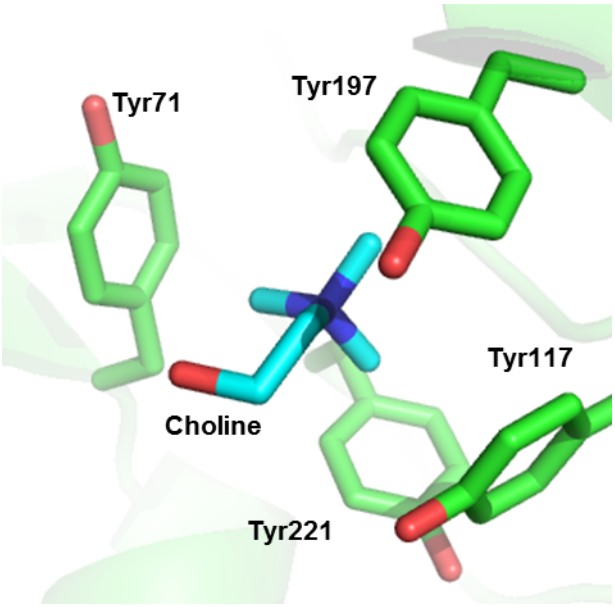


Figure 25 Active site of the choline transporter OpuB showing the trimethylammonium moiety of choline bound by the aromatic cage formed by four tyrosine residues. (PDB: 3R6U)

Aromatic cages are typically formed by the side-chains of 2-4 electron rich aromatic amino acids (Phe, Tyr, Trp) often supplemented by at least one acidic sidechain. The basis for the binding of trimethylammonium or other positively charged residues in aromatic cages are cation-π interactions. This was first demonstrated by Dougherty *et al.* by creating a completely synthetic receptor for the neurotransmitter acetylcholine, which bound the substrate with a comparable affinity (K_D = 50 μM) as the biological recognition sites.[89] Prior to these findings it was thought that anionic residues were largely responsible for binding of positively charged quaternary ammonium compounds. Cation- π interactions were subsequently found to be a widespread motif among proteins.[90]

2.1.4.2 Glycine Betaine Biosynthesis by Trimethylation of Glycine

An alternative synthetic pathway towards glycine betaine is the SAM-dependent trimethylation of glycine, which is rarely encountered due to the fact that it is an energetically much more costly synthetic path and has only been observed in a few organisms.[91,92] Crystal structures for sarcosine methyltransferase from *A. halophila* (PDB 5HIM)[93] and *G. sulphuraria* (PDB 2O57) have been solved, albeit without (trimethyl)glycine, making predictions of the binding mode difficult. Recently however, the crystal structure of glycine sarcosine *N*-methyltransferase (*Mp*GSMT) from *M. portucalensis* was solved, albeit in a catalytically inactive form.[94] Construction of a homology model revealed three key residues containing oxygen atoms arranged triangularly between the methyl group of SAM and the amino group of the substrate (the phenolic oxygen atoms of Tyr26 and Tyr185 and the main-chain carbonyl oxygen of Asn134).

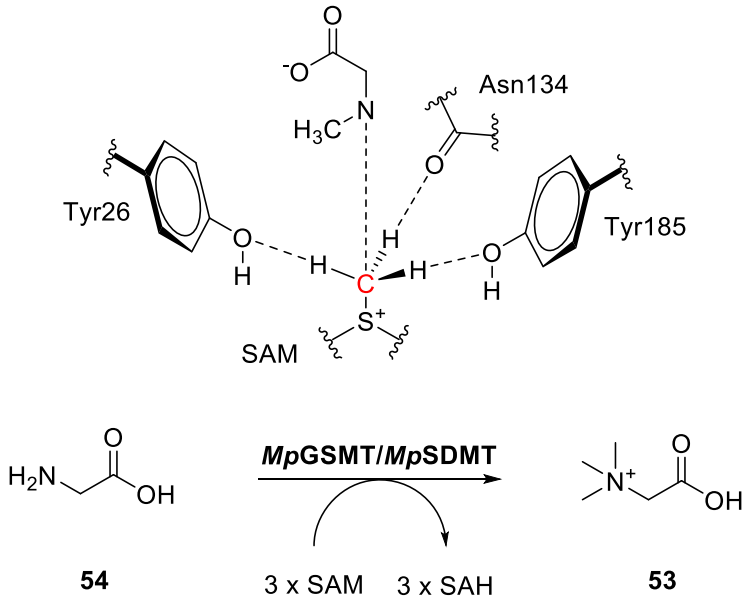


Figure 26 Top: schematic view of the homology model active site of MpGSMT and GNMT. Bottom: Biosynthetic pathway of glycine betaine (**53**) from glycine (**54**) in the halophilic archaeon *M. portucalensis*. *Mp*GSMT adds the first and second methyl groups to glycine and sarcosine, while *Mp*SDMT (sarcosine dimethylglycine N-methyltransferase) adds the second and third methyl groups to sarcosine and dimethylglycine. In total three equivalents of SAM are consumed, generating three equivalents of SAH.[95]

The distance between the methyl-accepting nitrogen and the SAM-methyl carbon was approximately 2.7 Å and these residues were thought to assist catalysis. Introduction of the point mutations Y26F and Y185F reduced the activity of the enzyme by 74- and 5-fold, respectively. The precise role of these residues in catalysis can only be speculated on, but it is quite possible that these interactions in fact constitute C-H···O hydrogen bonds as discussed in **Section 2.1.3.2**.

As an additional relevant piece of information, it should be noted that QM/MM simulations revealed that in aqueous solution water molecules surrounding the methyl groups of the trimethylammonium moiety are oriented in a way that suggests they sense the presence of the positively charged N atom and engage in C-H···O bonding.[96] On the one hand this confirms that the positive charge is delocalized among the three methyl groups, on the other hand the resolvation of this group may facilitate release of the product from the respective active site of the enzyme.

2.1.5 EgtD: The First Enzyme in the Biosynthesis of Ergothioneine

The first step in the biosynthesis of ergothioneine is the cooperative trimethylation of *L*-histidine by the SAM dependent methyl transferase EgtD, yielding hercynine (N^{α},N^{α},N^{α}-trimethyl-*L*-

histidine, TMH). This is then subsequently processed by the enzymes EgtB, EgtC and EgtE to give the final product ergothioneine (ET) as shown in Figure 27.

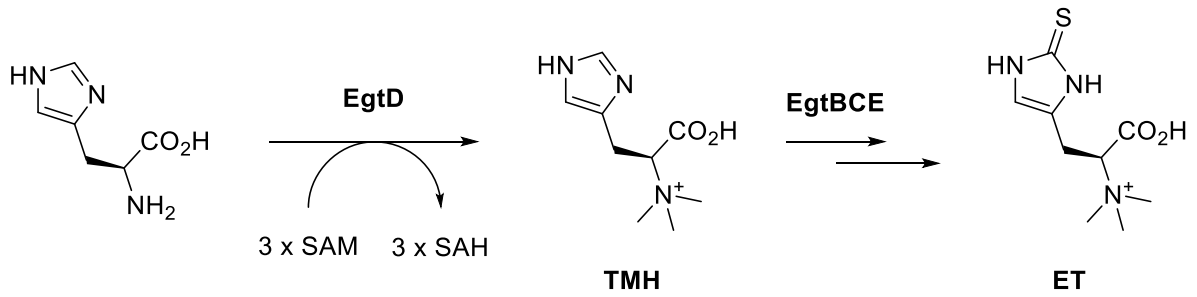


Figure 27 Biosynthetic pathway for the synthesis of ergothioneine from histidine.

At the outset of this project, the EgtD-SAH-DMH ternary complex had been solved at a resolution of 1.5 Å (Figure 8; PDB: 4PIO).[97] Of particular interest to us was the coordination of the N^{α}-methyl groups of DMH and their implications for the catalytic mechanism. Of the two N^{α}-methyl groups, one is in close proximity to the side-chain oxygen atoms of Thr163 (3.5 Å) and Tyr39 (3.6 Å) and the phenyl ring of Phe47 (3.9 Å), for the following discussion this position will be referred to as **Me1**.

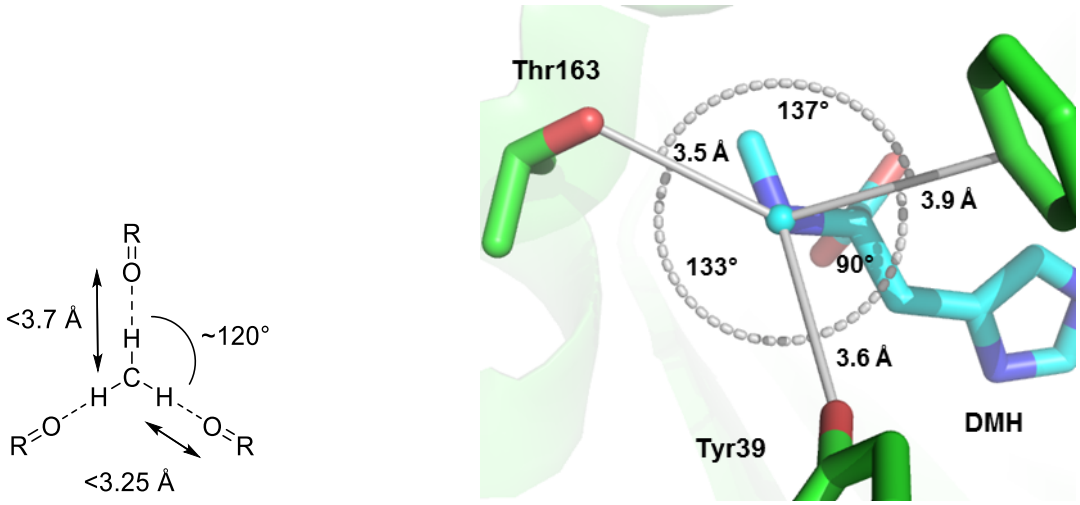


Figure 28 Left: ideal distances and angles for C-H···O hydrogen bonding to take place between three carbonyls and a methyl group. Right: distances and angles between **Me1** of DMH and the side-chain oxygen atoms of Thr163 (3.5 Å) and Tyr39 (3.6 Å) and the phenyl ring of Phe47 (3.9 Å). (PDB: 4PIO)

This methyl group points in the direction of the thioether-sulfur of SAH, the crystal structure thus represents a snapshot of the catalytic cycle when the second methyl group has been transferred from SAM to MMH, forming DMH. Upon closer inspection of the coordination of **Me1** it is evident that multiple conditions for C-H···O hydrogen bonding are fulfilled. The C···O distances of 3.5-3.9 Å are around the Van der Waals cutoff distance of ≥3.7 Å[98] and the angles between the hypothesized bonds are close to 120° as shown in Figure 28.

The second methyl makes close contact to the backbone carbonyl of Gly161 (3.0 Å) at an angle of almost 180°, this binding position will be referred to as **Me2**. A third position for methyl binding, named **Me3** hereinafter, was characterized in more detail in the dissertation of Laëtitia Misson, who demonstrated that an active site asparagine (Asn166) is crucial for catalysis. Orientation and coordination of Gly161 and Asn166 to DMH are shown on the left in Figure 29.[99] The nomenclature for all three methyl group is illustrated on the right in Figure 29 for clarity.

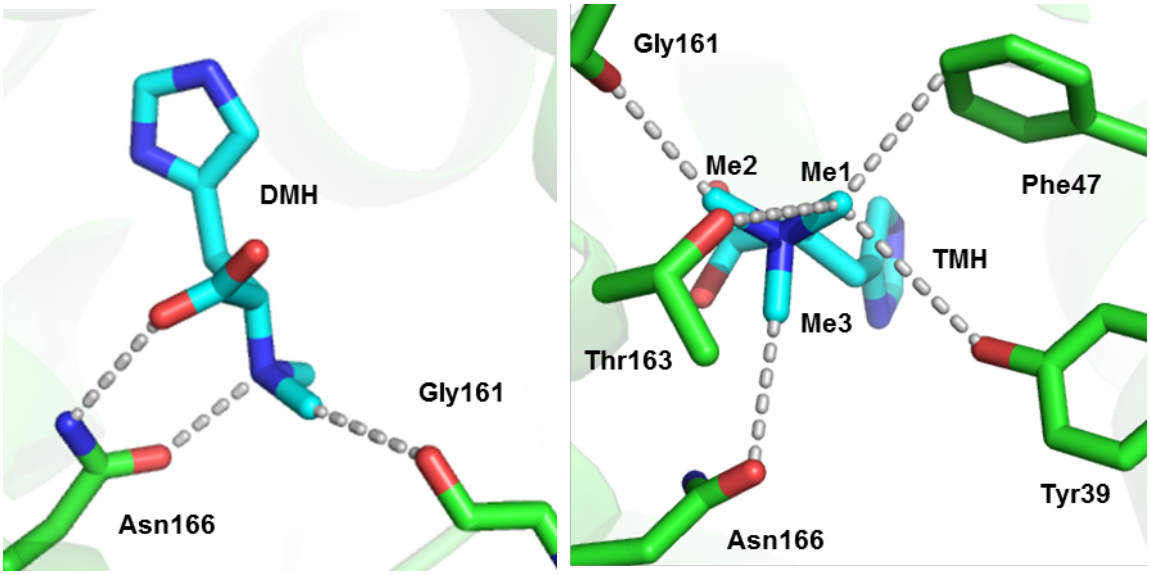


Figure 29 Left: Distances between **Me2** of DMH and the backbone carbonyl of Gly161 (3.0 Å) and between the amide and carbonyl of Asn166 and the carboxylate and N^{α}- of the substrate (2.9 Å and 2.7 Å respectively). (PDB: 4PIO) **Right**: Coordination of all three methyl groups of TMH in the active site of EgtD. (Unpublished structure)

In the last and most efficient methylation step from DMH to TMH (see also Table 1 in **Section 2.3.3**), Asn166 is proposed to have two essential roles. It acts both as the catalytic base, which deprotonates N^{α}, and coordinates the methyl group from position **Me1** to **Me3**, moving the latter away from the incoming methyl group on the final equivalent of SAM, which binds to position **Me1**. All the while Gly161 coordinates **Me2** as shown in Figure 30.

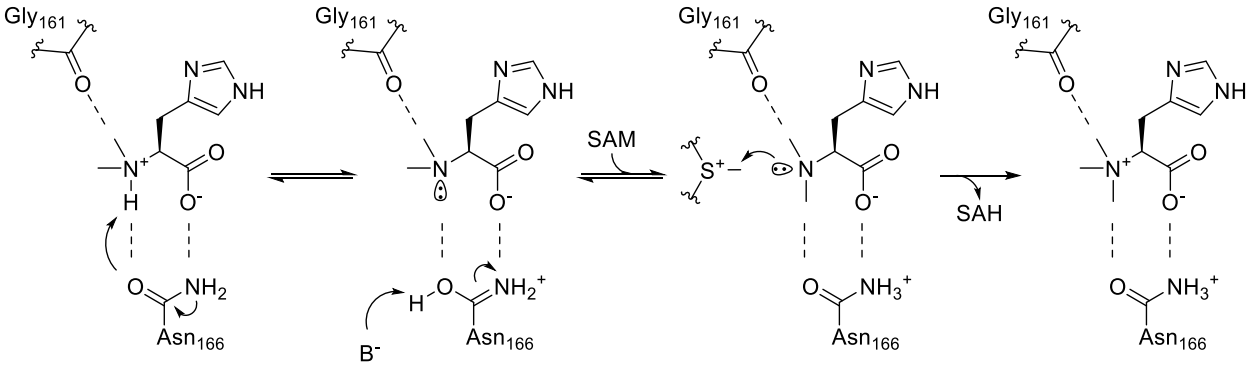


Figure 30 The proposed roles of Asn166 and the backbone carbonyl of Gly161 in the final methyl transfer from SAM onto DMH to form TMH. In the first step Asn166 acts as a catalytic base, deprotonating N^{α}. After passing on the proton to a different base (B^-) in the active site, it coordinates one methyl group in position **Me3** allowing for nucleophilic attack of the lone pair of N^{α} on the methyl group of SAM. During this process Gly161 constantly coordinates the other methyl group in position **Me2**, possibly increasing the nucleophilicity of N^{α} by a C=O···σ* interaction. The coordination in position **Me1** is omitted for clarity.

A concept which should be considered in this catalytic step especially is that of differential binding. This model predicts that binding affinity is higher for the transition state than for the substrate, thus facilitating catalysis.[100,101] While this is certainly a possibility, we propose an alternative hypothesis for the role of the carbonyl interactions with **Me2** and **Me3**, namely that they render N^{α} more nucleophilic. By drawing away electron density from the methyl carbon via n→σ* interactions, a partially positive charge is formed on the carbon center, which in turn creates a partial negative charge on the amine. This effect would increase the electron density on the nitrogen, making it a stronger nucleophile as shown with the example of Gly161 and **Me2** in Figure 31.

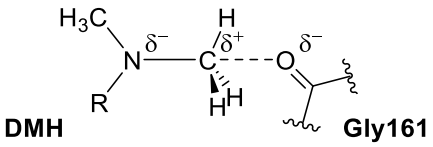


Figure 31 Coordination of **Me2** in DMH by Gly161 in the active site of EgtD, rendering the former more nucleophilic.

A comparable case is observed in Lewis acid-base complexation, where binding of a Lewis base to a Lewis acid enhances the electrophilic character of the latter as shown in Figure 32.[102]

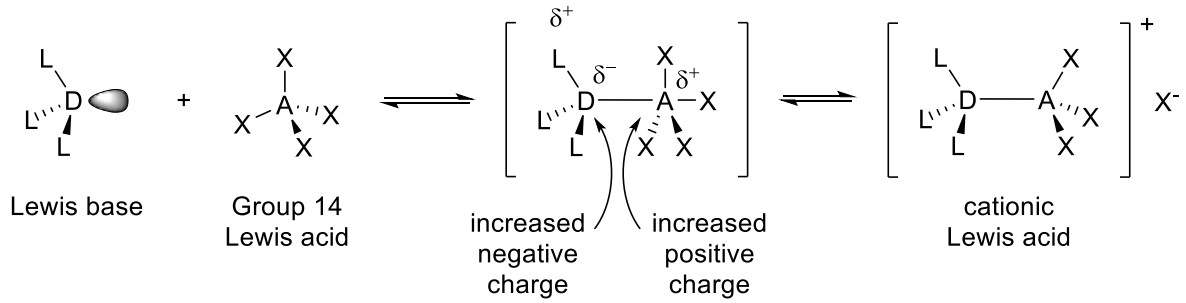


Figure 32 Electronic redistribution resulting from Lewis acid-base complexation.

2.2 Aim of this Chapter

Based on available kinetic and structural information, we hypothesized that two types of interactions are involved in the activation of the transferrable methyl group. The first of these is by solvation of said group by C-H···O bonds, in analogy to the elegant study by Horowitz *et al.*on SET7/9.[44] The second, more unorthodox proposal involves activation of the methyl acceptor by enhancing the nucleophilicity of the accepting *N*-center via carbonyl n→σ* interactions.[102] We thus designed a series of experiments to test if one or both of these hypotheses could be valid.

By synthesizing compounds with specific ^{13}C-labels on the methyl groups transferred form the donor to the acceptor, we created probes to analyze binding modes of the methyl donor as well as the respective receptors. The results we have obtained complement the existing crystal structures nicely and offer unique insights into the catalytic mechanism. The advantage of NMR spectra over crystal structures is that the former is measured in solution and allows for examination of electronic consequences of the structural observations. Our findings have helped us to gain a better understanding of how EgtD manages to methylate all its substrates with similar catalytic efficiency.

For the following discussion of methyl group binding sites we shall use the annotations introduced in the previous section and summarized again in Figure 33.

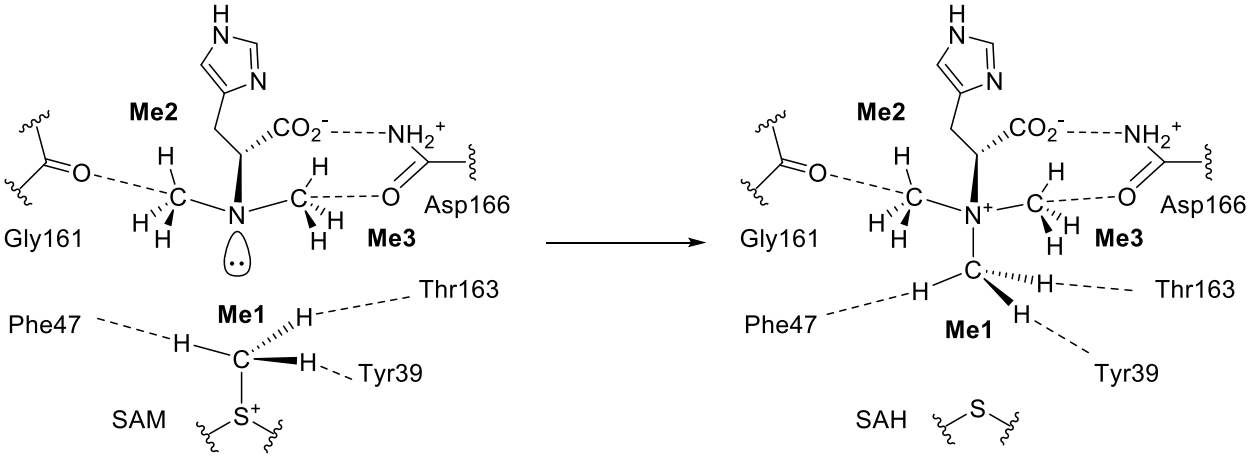


Figure 33 Proposed mechanism of the third and final methylation step catalyzed by EgtD. The three binding sites are annotated as follows: **Me1** is the binding site coordinating the transferred methyl group formed by the C-H···O bonding triangle consisting of Tyr39, Phe47 and Thr163. **Me2** is the binding site with coordination to the backbone carbonyl of Gly161. **Me3** is the binding site with coordination to the sidechain carbonyl of Asn166, which also acts as a catalytic base in the deprotonation steps.

2.3 Results and Discussion

2.3.1 Chemical Synthesis of Labelled Compounds

Synthesis of ^{13}C-labelled compounds was carried out using previously published strategies and commercially available starting materials. The ^{13}C-methyl groups were inserted either using ^{13}C-formaldehyde or ^{13}C-methyl iodide. ^{13}C-DMH (**55**) and ^{13}C-DMW (**56**) were synthesized by reductive amination with ^{13}C-formaldehyde and $NaCNBH_3$,[103] ^{13}C-TMH (**57**) was synthesized from ^{13}C-DMH with ^{13}C-methyl iodide.[104] For the synthesis of ^{13}C-MMH (**58**) it was necessary to first introduce a benzyl protecting group on N^{α}.[105] The resulting secondary amine could be methylated by the usual reductive amination procedure and the benzyl group subsequently removed by hydrogenation (**61** and **62** in Figure 34).

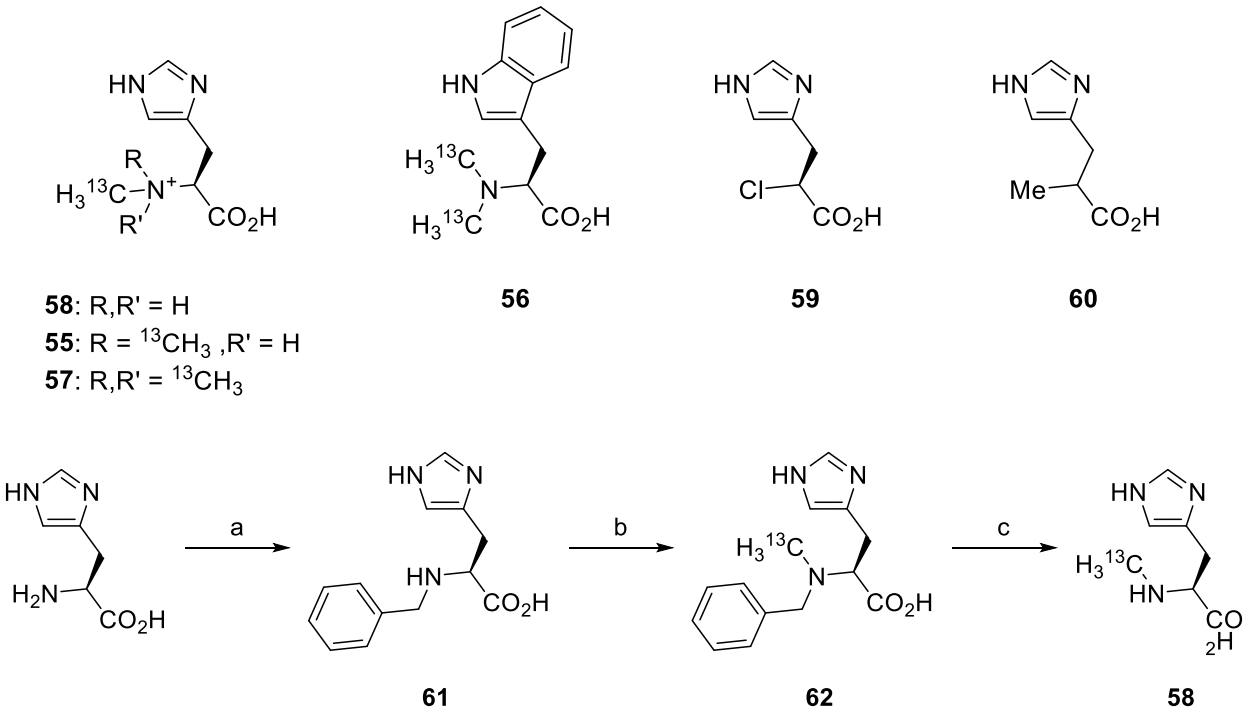


Figure 34 Top: Chemically synthesized substrates for EgtD. **Bottom**: Reagents and conditions for the synthesis of **57**: (a) benzaldehyde, NaOH, $NaBH_4$, MeOH, 0 °C, 2 h, 93 %; (b) ^{13}C-formaldehyde, $NaCNBH_3$, AcOH, water, 16 h, 82 %; (c) H_2, Pd/C, MeOH, 16 h, 88 %.

Of the two non-labelled compounds, α-chloro-*L*-histidine **59** was synthesized in one step by diazotization of the primary amine, followed by substitution with chlorine, similar to a Sandmeyer reaction. The second non-labelled compound α-methyl-*D/L*-histidine **60** was synthesized following published protocols[106,107] and was prepared and kindly supplied by Reto Burn.

All compounds were obtained in >98 % purity and stock solutions were prepared in D_2O, determination of the concentration was calculated from an internal standard of *t*-BuOH.

2.3.2 Enzymatic Synthesis of ^{13}C-Labelled SAM

SAM can be efficiently synthesized enzymatically on a preparative scale from methionine and ATP using SAM synthetase (EC 2.5.1.6).[108,109] This enzyme catalyzes the direct attack of the methionine sulfur onto the C-5' of ATP and has been used to synthesize a variety of labelled analogues of SAM such as Se-SAM,[110] ^{36}S-SAM[111] or ^{13}C-SAM.[44]

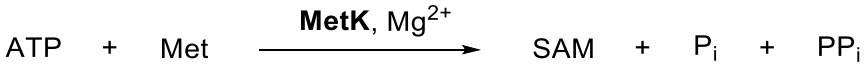


Figure 35 Basic reaction for the formation of SAM by the SAM synthase MetK from ATP and Met.

Using the reaction conditions detailed in the legend of Figure 36 it was possible to convert 40 % of the limiting starting material (^{13}C-Met) into ^{13}C-SAM. The reaction was monitored by IE-HPLC and seemed to reach a plateau after about 2-6 hours. Addition of more enzyme (t = 1000 min) did not result in additional production of SAM, which could be due to inhibition by pyrophosphate.[112]

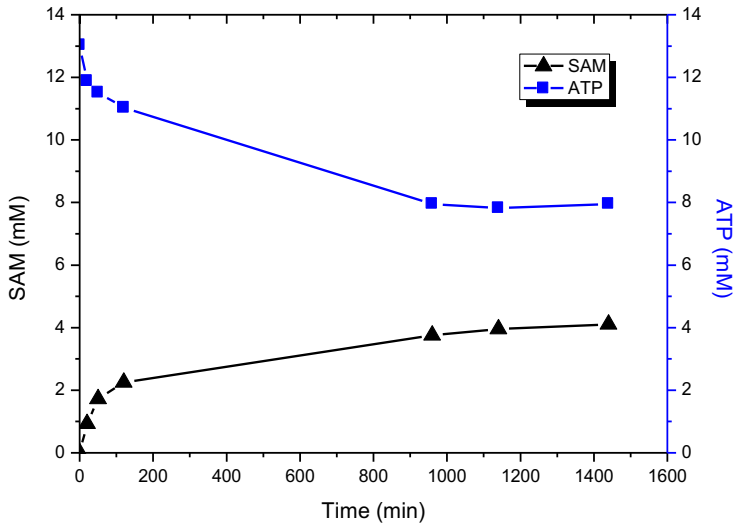


Figure 36 Consumption of ATP versus formation of SAM by the SAM synthase MetK. Conditions: 4 µM MetK, 10 mM Met, 13 mM ATP, 26 mM $MgCl_2$, 50 mM KCl, 1 mM EDTA, 8 % β-mercaptoethanol in 100 mM Tris at pH 8. Reactions were quenched by adding 40 µL of the reaction mixtures to 20 µL 1 % TFA before injecting 20 µL into the HPLC.

At a total reaction volume of 750 µL and after purification by RP-HPLC, the product was obtained in an overall yield of 34 %. Analysis by ^{1}H- and ^{13}C-NMR showed very high purity and

confirmed that only one ^{13}C-labelled species was present in the product. The shift of 23.5 ppm for the methyl group corresponds well with the reported literature value of 24.6 ppm, whereas no remaining ^{13}C-methionine could be detected at 15.2 ppm.[108]

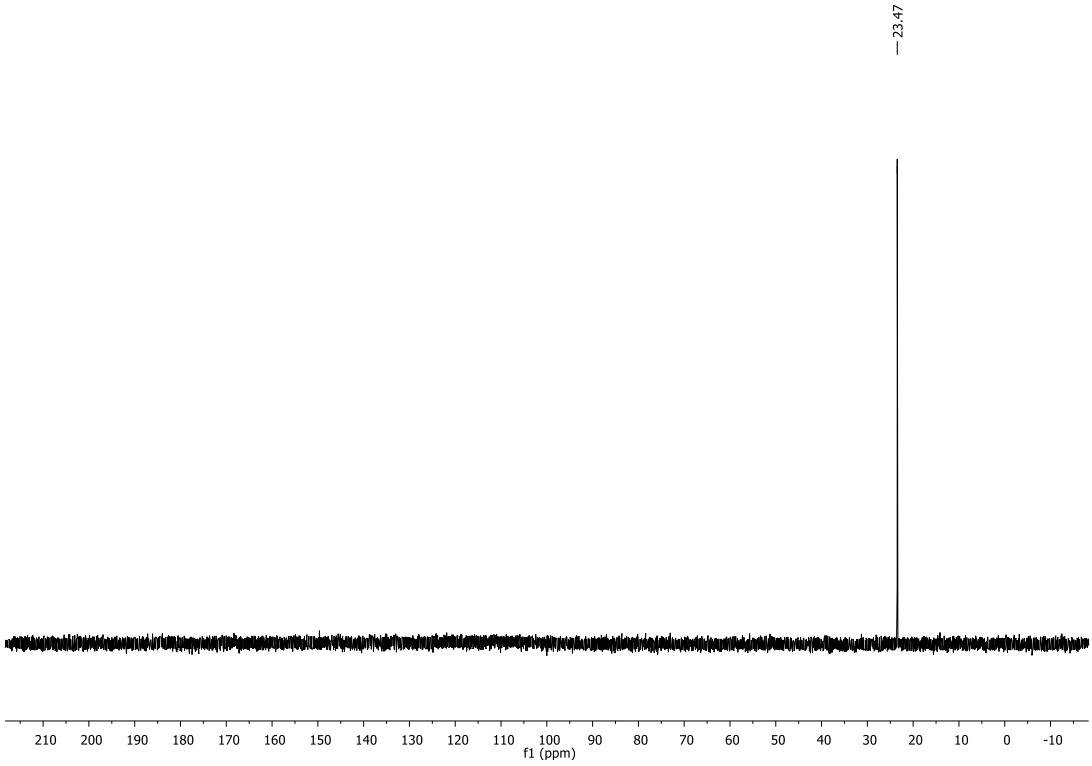


Figure 37 ^{13}C-NMR spectrum of ^{13}C-SAM, 500 MHz in D_2O (256 scans).

2.3.3 Binding of DMH to EgtD

EgtD produces almost 90 % TMH when incubated with one equivalent of histidine and one equivalent of SAM. This is explained by the increased affinity for the methylated histidine variants and not by the increased catalytic efficiencies, all shown in **Table 1**. The affinity for each of the three substrates increases again upon addition of a saturating amount of SAH, most drastically for histidine by 8-fold. The K_D-values for SAH and SAM were determined to be 210 ± 20 µM and 270 ± 20 µM respectively.[97]

Table 1 Kinetic parameters of EgtD for the three processive methylation steps and K_D values as determined by isothermal calorimetry titration.[97]

	k_{cat} [s^{-1}] [(a)]	k_{cat}/K_M [$M^{-1}s^{-1}$] [(a)]	K_D [µM] [(b)]	K_D [µM] [(c)]
His	0.58	5'300	290 ± 14	37 ± 1
MMH	0.23	13'000	70 ± 30	14 ± 7
DMH	0.43	17'000	4 ± 2	2 ± 1

Data represent averages from multiple measurements. The standard error is less than 20% of the average value. Reaction conditions: (a) 50 mM Tris·HCl pH 8.0, 26 °C, 200 µM SAM, 1 µM *S*-adenosylhomocysteine nucleosidase, 1 µM adenine deaminase. (b) 20 mM Tris·HCl pH 7.5, 150 mM NaCl, 25 °C, 100 µM EgtD in cell, 5 mM ligand in syringe. (c) EgtD solution contained 7mM SAH.

In our approach, we aimed to compare the ^{1}H-NMR shift of the two methyl groups of DMH free in solution with the respective shift(s) when bound to EgtD in presence and absence of SAH. For this purpose we synthesized N^{α},N^{α}-[^{13}C]dimethyl-*L*-histidine (^{13}C-DMH). First, a spectrum of the unbound substrate was measured at the concentration and under the same buffer conditions which we would later use to conduct the measurements with the enzyme. After measuring the shifts of the unbound ^{13}C-DMH (δ_H: 2.74 ppm, δ_C: 41.63 ppm), spectra were recorded at a substrate to enzyme ratio of 1:2.5 in order to bind most of the substrate (Figure 38). Judging by the peak areas, at least 70 % of the substrate was bound to the active site, while the other 30 % remained free in solution.

Under these conditions however, a distinct shift of each methyl group was detected, one moving into lower field along the proton axis and retaining its shift on the carbon axis (δ_H: 2.87 ppm, δ_C: 41.59 ppm; referred to in the following discussion as **Me1$_{DMH}$**), while the other shifted into lower field along the carbon axis, only moving downfield slightly along the proton axis (δ_H: 2.81 ppm, δ_C: 45.04 ppm; referred to in the following discussion as **Me2$_{DMH}$**).

These results were very encouraging as they demonstrated a significant difference between free and EgtD-bound DMH. While in both cases the dimethylammonium nitrogen is a

stereogenic center and, hence, the two methyl groups are diastereotopic, only the bound DMH displays two distinguishable methyl resonances. This may be due to the more pronounced anisotropy of the chiral environment in the active site of the enzyme but could also be interpreted as a sufficient increase of the rotational barrier along the C^{α}-N^{α} axis to shift the rotation into the slow exchange regime. Furthermore, the shift of **Me1$_{DMH}$** along the proton axis is consistent with C-H···O bonding.[44] The strong downfield shift of **Me2$_{DMH}$** on the carbon scale indicates a different deshielding interaction with the protein. One possibility is that this position is being coordinated by a carbonyl group, as proposed in **Section 2.1.5**. By creating a partial positive charge on the methyl carbon, the nucleus is deshielded accounting for the observed shift of Δ_C = 3.43 ppm.

We then continued by adding a saturating amount of SAH to the mixture (substrate/enzyme/SAH 1:2.5:5) and recorded another spectrum. Under these conditions >95 % of the substrate was bound, apparently “locked” into the active site by the addition of SAH. Most strikingly however, **Me1$_{DMH}$** shifted further downfield (δ_H: 3.06 ppm, δ_C: 42.90 ppm), while **Me2$_{DMH}$** remained more or less unchanged (δ_H: 2.73 ppm, δ_C: 44.61 ppm). Especially the shift of Δ_H = 0.32 ppm of **Me1$_{DMH}$** from its unbound to bound state suggests increased strengthening of the C-H···O bonds. This trend is also reflected in the downfield shift of **Me1$_{DMH}$** along the carbon-axis, possibly sensing the approach of the S-nucleophile and giving the methyl group a more electrophilic character. It thus seems likely that substrate binding becomes tighter and the hydrogen bonds stronger upon addition of SAH. Especially the former observation was to be expected, based on the increase in K_D for DMH when the system is saturated with SAH.

Alignment of these findings with the available crystal structures will be discussed later in **Section 2.3.6**.

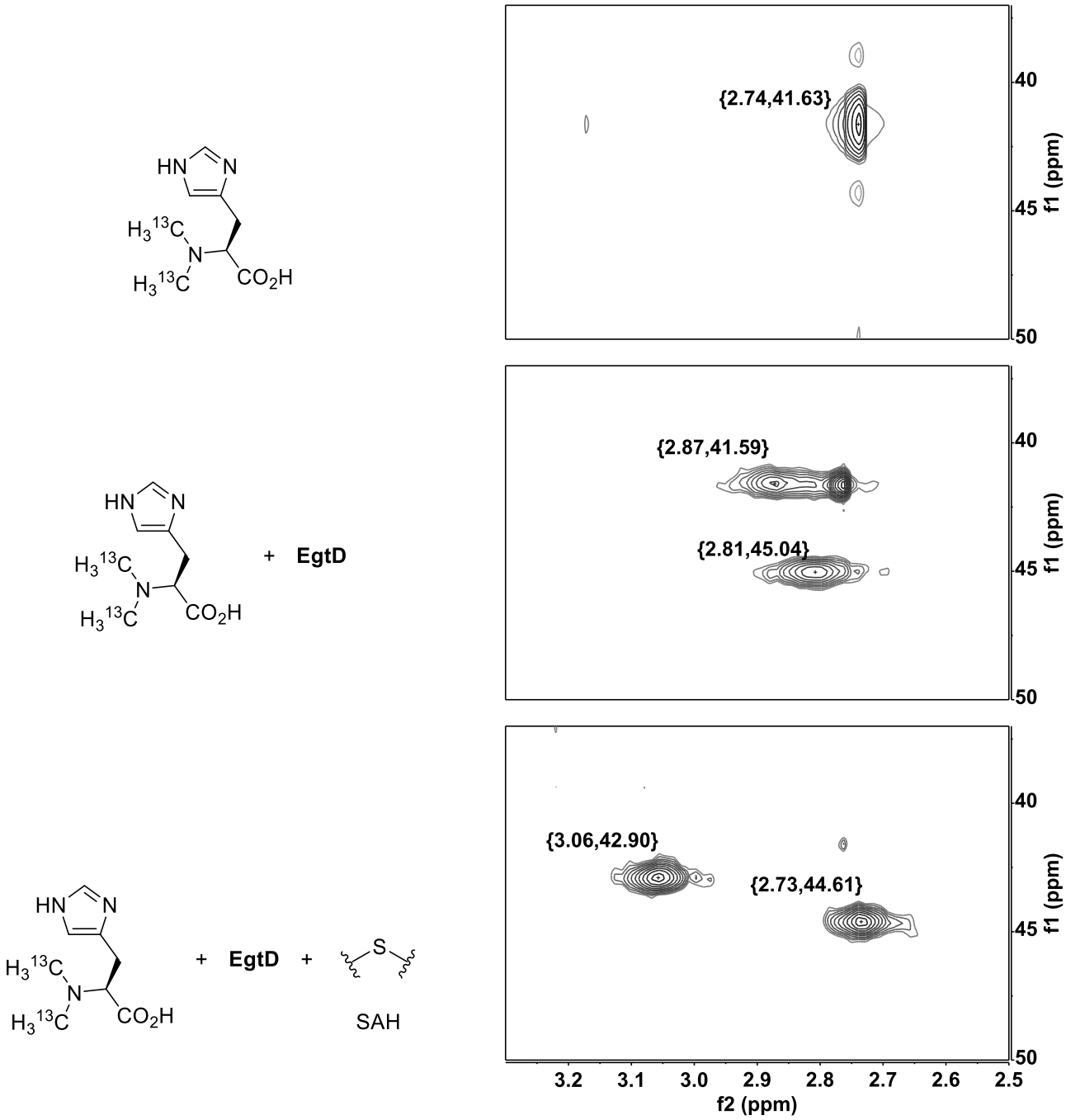


Figure 38 Top: HMQC-spectrum of ^{13}C-DMH [62 µM]. Middle: HMQC-spectrum of ^{13}C-DMH [62 µM] and EgtD [154 µM]. Bottom: HMQC-spectrum of ^{13}C-DMH [62 µM] and EgtD [154 µM] and SAH [300 µM]. Spectra were recorded in 50 mM phosphate buffer, pH 8 containing 50 mM NaCl and 5 % D_2O.

2.3.4 Binding of MMH to EgtD

A close examination of MMH binding produced corroborating evidence that SAH-binding increases occupation of the **Me1** position. We have demonstrated this by conducting the same set of experiments as described above, replacing ^{13}C-DMH with ^{13}C-MMH. After determining the shifts in the unbound state (δ_H: 2.60 ppm, δ_C: 32.05 ppm referred to in the following discussion as $\mathbf{Me_{MMH}}$), we repeated the measurement in the presence of 2.5 eq of EgtD. When bound to the enzyme, the methyl group of the substrate exhibits a similar change in chemical shift (δ_H: 2.55 ppm, δ_C: 33.28 ppm) as $\mathbf{Me2_{DMH}}$ does in the case of DMH. A slight downfield shift along the carbon axis, combined with little change – even a small upfield shift – along the proton axis. The downfield carbon shift suggests that this methyl group adopts the **Me2-** position, coordinating to Gly161.

After adding the same saturating concentration of SAH as previously, something rather remarkable occurred. We detected a distinct shift of $\mathbf{Me_{MMH}}$ to low field along the hydrogen axis, with an additional small downfield shift along the carbon axis (δ_H: 2.79 ppm, δ_C: 34.46 ppm). It seems as though $\mathbf{Me_{MMH}}$ is pointing into the binding pocket **Me2** when bound without co-substrate and moves into the position **Me1** once SAH is added.

For mechanistic implications, one needs to regard this process in reverse: after the methyl group is transferred from SAM onto histidine, producing MMH, SAH leaves the active site and the transferred methyl group on MMH rotates away from where the next methyl group will be bound when it enters the active site on the next equivalent of SAM (Figure 39).

^{13}C-MMH can thus be regarded as a molecular balance. Some more classical examples of molecular balances, where intramolecular interactions are examined,[113] include the study of C-H···O interactions by Ōki *et al.*,[114] or more recently the description of C-H···π interactions.[115] In contrast to these examples we have introduced ours into an enzyme active site and can observe reorientation of the methyl group upon addition of the co-ligand.

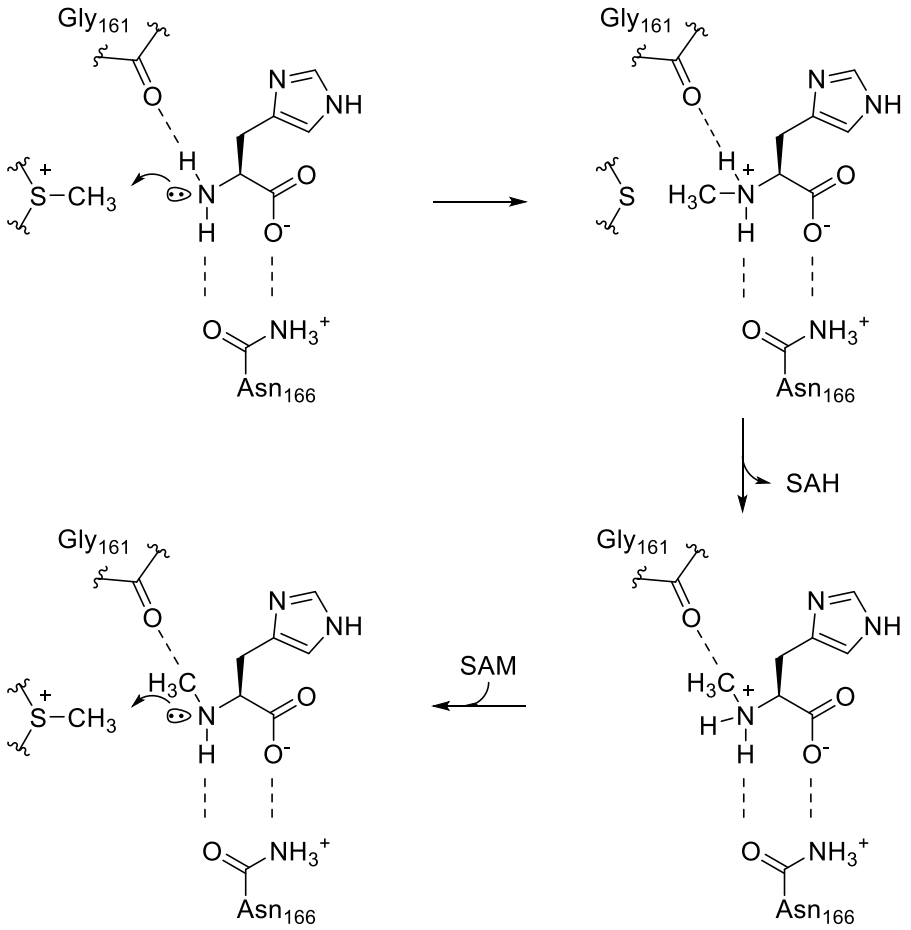


Figure 39 Methyl transfer from SAM onto histidine to form MMH, followed by rotation of the *N*$^{\alpha}$-methyl group into the Gly161 (**Me2**) position in preparation for subsequent methylation of MMH to DMH.

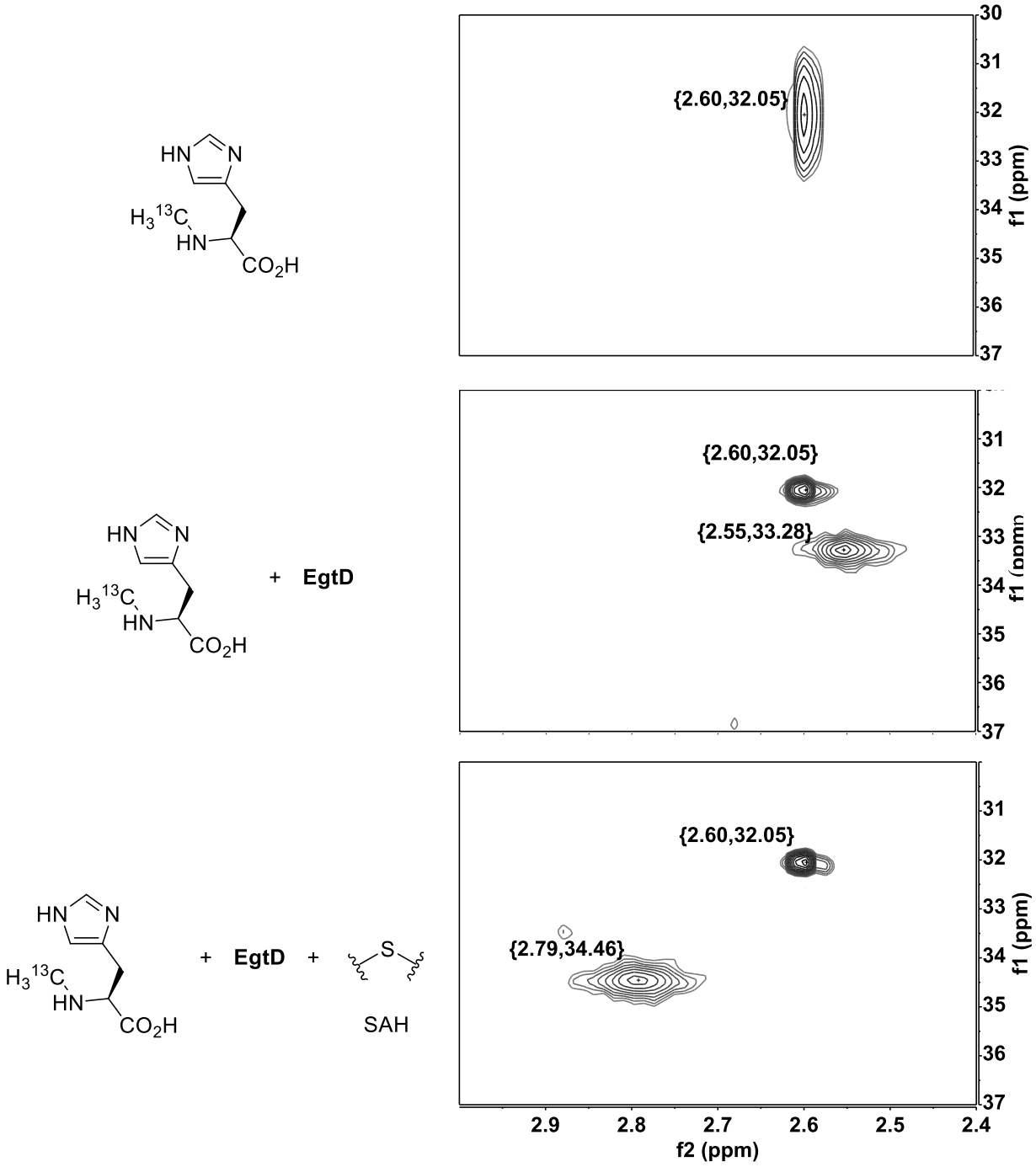


Figure 40 Top: HMQC-spectrum of ^{13}C-MMH [62 µM]. Middle: HMQC-spectrum of ^{13}C-MMH [62 µM] and EgtD [154 µM]. Bottom: HMQC-spectrum of ^{13}C-MMH [62 µM] and EgtD [154 µM] and SAH [300 µM]. Spectra recorded in 50 mM phosphate buffer, pH 8 containing 50 mM NaCl and 5 % D_2O.

2.3.5 Binding of TMH to EgtD

In order to obtain what we hoped would be the full picture of the active site after the final methylation step, we conducted an analogous set of experiments using ^{13}C-TMH. In particular, we were interested in position **Me3** and what effect coordination by Asn166 would have on the chemical shift of the respective methyl group. However, both upon addition of only enzyme and enzyme plus SAH, no differentiation between the three methyl groups could be recorded. Although we could see a difference between the unbound and the bound signal, the signal of the methyl groups could not be resolved into three distinct resonances. The rotational barrier along the C^{α}-N^{α} bond was too low to slow down the rotational frequency relative to that of the NMR frequency. In order to determine the energetic barrier of rotation it would be necessary to measure the samples at temperatures well below 0 °C. This in turn is not possible when working in an aqueous solvent system.

Based upon the available crystal structure, it is safe to assume that position **Me3** is indeed coordinated to Asn166, as shown in Figure 41.

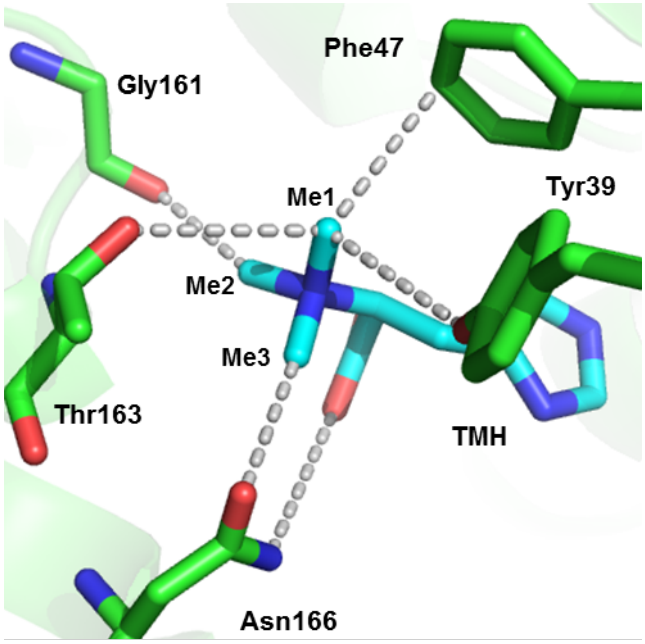


Figure 41 Active site of EgtD in complex with TMH (cyan), showing the coordination of all three methyl N^{α}-groups. The distances between the TMH-methyl groups and the side chain residues are as follows: **Me1** 3.8 Å (Thr163), 3.6 Å (Phe47) and 3.9 Å (Tyr39); **Me2** 3.0 Å (Gly161); **Me3** 3.1 Å (O-C, Asn166) and 4.0 Å (N-O, Asn166) (Unpublished data)

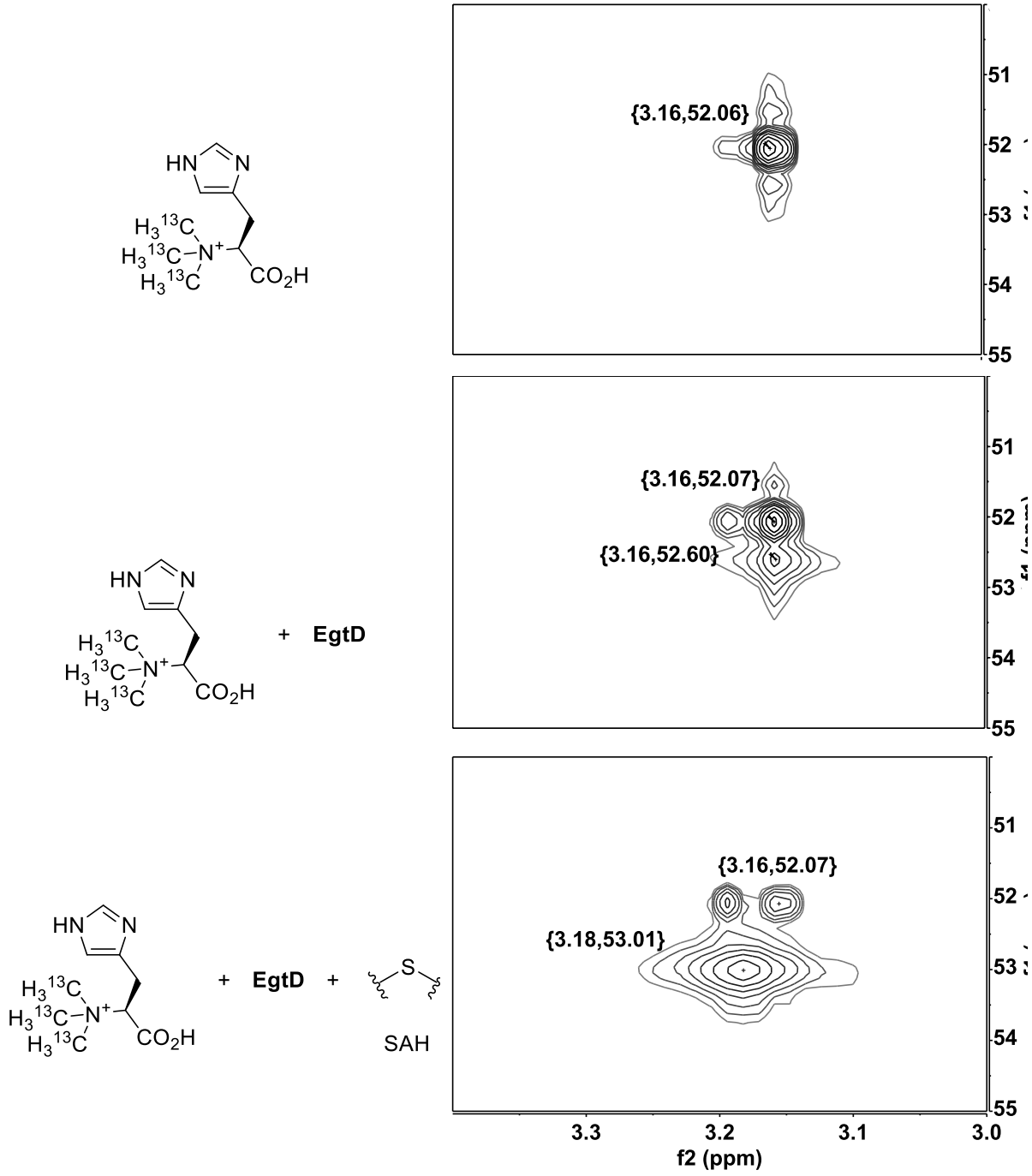


Figure 42 Top left: HMQC-spectrum of ^{13}C-TMH [40 µM]. Middle: HMQC-spectrum of ^{13}C-TMH [40 µM] and EgtD [100 µM]. Bottom: HMQC-spectrum of ^{13}C-TMH [40 µM], EgtD [100 µM] and SAH [200 µM]. All spectra recorded in 50 mM phosphate buffer, pH 8 containing 50 mM NaCl and 5 % D_2O.

2.3.6 Binding of SAM to EgtD with α-Chlorohistidine

To examine whether solvation of the transferrable methyl group forms the same C-H···O bonds before and after methyl transfer we also investigated EgtD in complex with SAM and the non-nucleophilic histidine analog αCl-*L*-histidine (αClHis). For this purpose ^{13}C-labelled SAM was synthesized as described in **Section 2.3.2** and measured in combination with the enzyme and a histidine analog, which did not possess a suitable acceptor and was known to bind to the enzyme with high affinity. For this purpose we chose αClHis (compound X in Figure 34), a derivative of histidine where the N^{α} is replaced by a chloride atom, which was shown to be a potent inhibitor of EgtD, competitive to histidine with a K_I of 6.2 ± 1.5 µM.[99]

An X-ray crystal structure for the EgtD:SAM:αClHis complex has been solved by Allegra Vit.[116] Alignment of this structure with the EgtD:SAH:DMH complex gives an RMSD of 0.245 Å (270 atoms). Comparison of the two ternary complexes revealed some important changes within the active site, which are immediately apparent and are relevant to catalysis. Firstly, the distance between the methyl group on SAM and the Cl-atom of αClHis (Figure 43, left) is 0.3 Å shorter than that of the methyl group on DMH and the thioether of SAH (Figure 43, right). The distance thus becomes slightly larger after methyl transfer. Secondly, the angles for C-H···O bonding are nearly 90° before and only 70° after methyl transfer relative to the S-CH_3-N/Cl axis (see also **Table 2**). According to this model, the methyl group is bound in a more strained conformation before transfer from SAM to the acceptor.

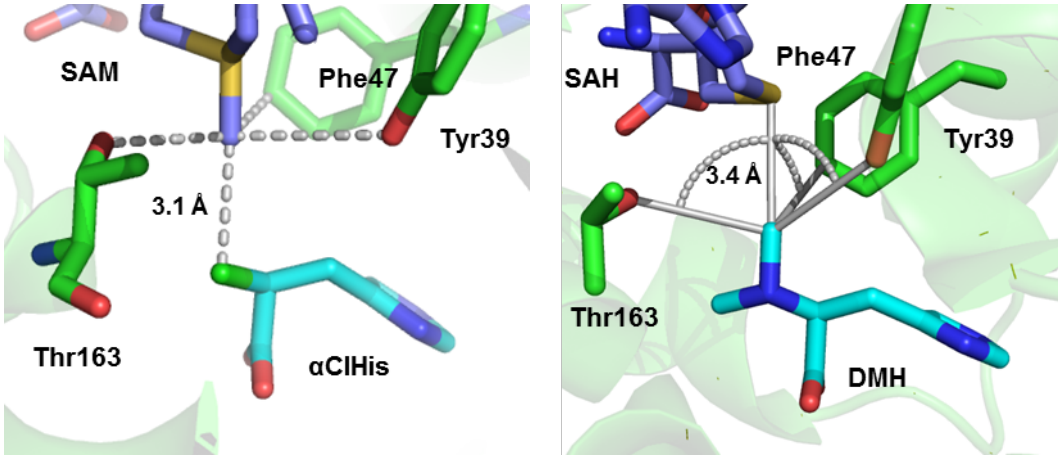


Figure 43 Left: View of the binding pocket of the ternary complex EgtD:SAM:αClHis. The distances between the SAM-methyl group and the three side chain residues are 3.2 Å (Thr163), 3.6 Å (Phe47) and 3.8 Å (Tyr39). **Right**: View of the binding pocket of the ternary complex EgtD:SAH:DMH.[97] The distances between the DMH-methyl group and the three side chain residues are 3.5 Å (Thr163), 3.9 Å (Phe47) and 3.6 Å (Tyr39). The solid black lines indicate distances and the dashed black lines indicate angles.

Table 2 Angles for C-H···O bonding relative to the S-CH_3-N/Cl axis in the ternary structures EgtD:SAM:αClHis and EgtD:SAH:DMH.

	EgtD:SAM:αClHis	**EgtD:SAH:DMH**
Residue	Angle	
Tyr39	91°	69°
Phe47	116°	55°
Thr163	80°	79°
Average	96° ± 18°	67° ± 12°

SAM free in solution and in combination with 2.5 eq of EgtD shows no substantial shift of its labelled methyl group (δ_H: 2.84 ppm, δ_C: 23.28 ppm and δ_H: 2.84 ppm, δ_C: 23.31 ppm, respectively, referred to in the following discussion as **Me_{SAM}**), allowing for the assumption that only a non-detectable fraction of SAM is bound to the enzyme when no histidine analog is present. This is in line with the K_D of 270 ± 20 µM determined for SAM, which preferentially binds to the EgtD:His complex, as well as the findings that EgtD binds its substrates in a sequential order, first binding histidine and only then binding SAM.[99]

Upon addition of αClHis a significant downfield shift of **Me_{SAM}** on the hydrogen axis could be observed (δ_H: 3.02 ppm, δ_C: 29.66 ppm), which was however less (Δ_H = 0.17 ppm) than that observed for **$Me1_{DMH}$** (Δ_H = 0.32 ppm). In comparison to the crystal structure with bound DMH, EgtD in complex with SAM and chlorohistidine reveals very similar distances between the methyl carbon in position **Me1** and the coordinating residues (**Table 3**).

Table 3 Distances between the transferred methyl group on the donor (SAM) and acceptor (DMH) and the three coordinating residues in position **Me1**.

Substrate:	**SAM**	**DMH**
Tyr39	3.8 Å	3.6 Å
Phe47	3.6 Å	3.9 Å
Thr163	3.2 Å	3.5 Å

The strong downfield shift along the carbon axis of Δ_C = 6.38 ppm is remarkable and the effect of significant deshielding and thus activation as an electrophile. It is however possible that the

nearby chlorine atom (3.1 Å from the methyl group of SAM) contributes to the deshielding effect, warranting additional experiments.

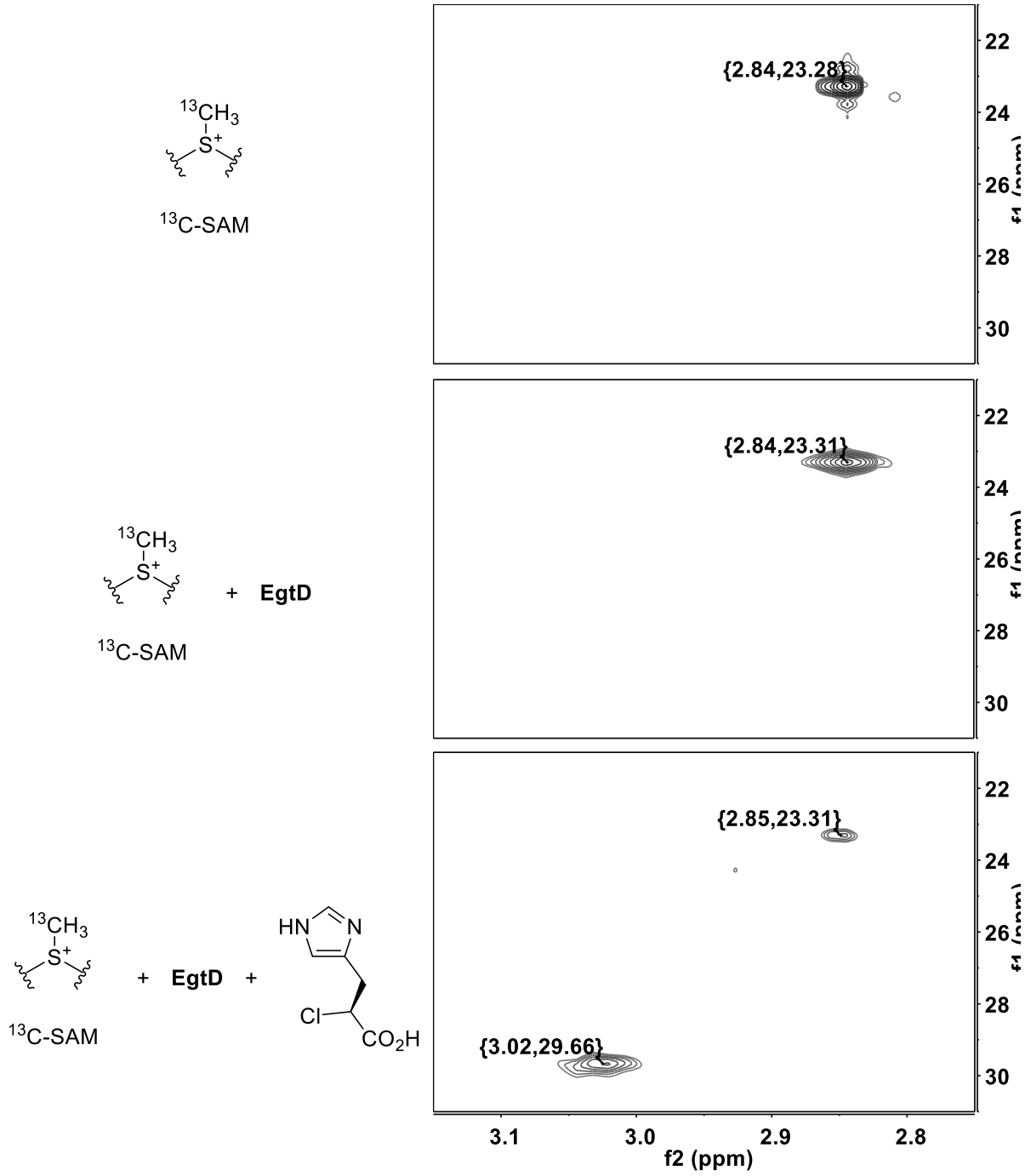


Figure 44 Top: HMQC-spectrum of ^{13}C-SAM [70 µM]. Middle: HMQC-spectrum of ^{13}C-SAM [39 µM] and EgtD [95 µM].Bottom: HMQC-spectrum of ^{13}C-SAM [39 µM], EgtD [95 µM] and αCl-*L*-His [190 µM]. All spectra recorded in 50 mM phosphate buffer, pH 8 containing 50 mM NaCl and 5 % D_2O.

2.3.7 Binding of SAM to EgtD with α-methylhistidine

The binding studies shown in the previous section were repeated using a different histidine analog in order to rule out any substantial influence of the α-chloro group. To do so, the chloro substituent was replaced with a methyl group (**59** and **60** respectively in Figure 34). The K_I-value of α-methyl *D/L*-histidine (αMeHis) is comparable to that of the chlorinated analog and was determined to be 5.4 ± 1.6 µM.

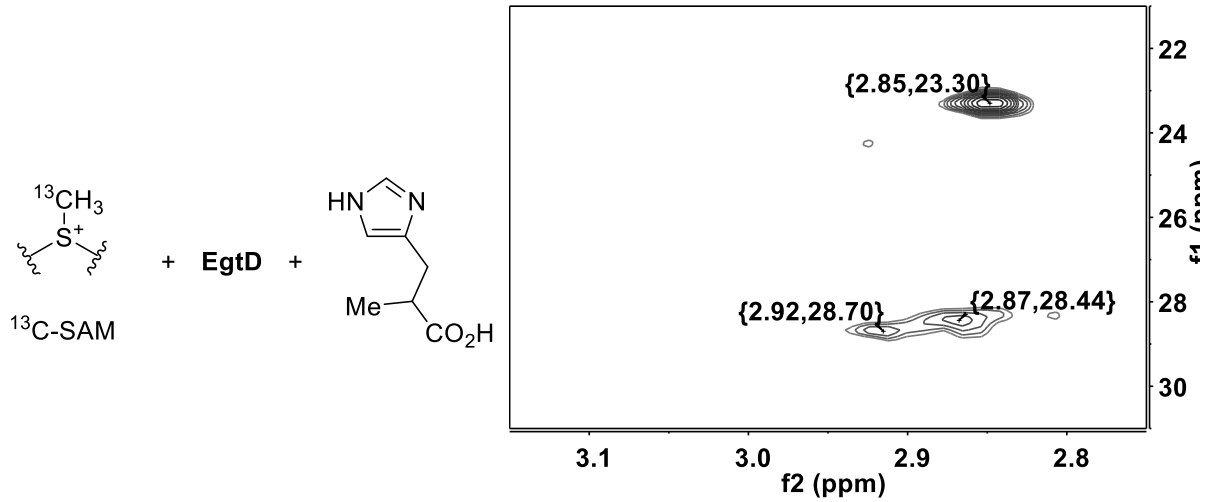


Figure 45 HMQC-spectrum of ^{13}C-SAM [39 µM], EgtD [95 µM] and αMe-*DL*-His [400 µM] All spectra recorded in 50 mM phosphate buffer, pH 8 containing 50 mM NaCl and 5 % D_2O.

Interestingly, this experiment produced two signals for the bound substrate with slightly different shifts, which could arise from the fact that αMeHis is a racemic mixture due to the synthesis (see **Section 2.6.2**). It is reasonable to assume that the weaker signal (δ_H: 2.92 ppm, δ_C: 28.70 ppm) is SAM in complex with EgtD and αMe-*D*-His, while the stronger signal (δ_H: 2.87 ppm, δ_C: 28.44 ppm) is in complex with αMe-*L*-His due to the different affinities for *D*- and *L*-histidine.[99] The downfield shifts along the carbon coordinate (Δ_C = 5.16 and 5.42 ppm) are comparable to the previously measured shift of Δ_C = 6.38 ppm with the chlorinated substrate. Along the hydrogen coordinate the downfield shift is decreased for both signals. In conclusion, it appears that the binding of SAM and the interactions of the enzyme with the methyl group are not greatly influenced by the nature of the substitution in the α-position of the co-substrate. It also looks like EgtD can bind both the *D*- and *L*-form of **60** at a ratio of roughly 1:3, an assumption that could not be made based on previously determined inhibition kinetics.[99]

2.3.8 Binding of DMW to Engineered EgtD$_{M252V,E282A}$

To examine whether solvation of the *N*-methyl group is influenced by the side chain of the substrate we also studied a variant of EgtD in complex with N^{α},N^{α}-dimethyl tryptophane (DMW).

By introducing two point mutations in the histidine binding pocket (M252V and E282A) one can change the substrate specificity of EgtD from histidine to tryptophan by a factor of almost 10^7. Comparing the wild type methylation of histidine with the mutant methylation of tryptophan reveals that k_{cat}/K_M remains unperturbed yet k_{cat} drops by roughly six-fold, indicating that binding affinity remains the same but catalysis is slightly slower (see also **Table 1**). The mutant enzyme showed virtually no more activity for histidine and diminished activity for DMW over tryptophan (**Table 4**). Despite this dramatic change in structure and activity, the binding affinity for tryptophan rivals that of the wild type for histidine. There is however one major change in the activity, namely the significantly less efficient final methylation step from DMW to TMW, which is in stark contrast to the high efficiency of the last methylation step from DMH to TMH in the wild type. It thus seems reasonable to assume that the transition state is only weakly stabilized in the final methylation step of this variant.

Table 4 Kinetic parameters of the engineered EgtD$_{M252V,E282A}$ for histidine, tryptophan and DMW.[99]

	k_{cat} [s^{-1}]	K_M [µM]	k_{cat}/K_M [$M^{-1}s^{-1}$]
His	-	-	2
Trp	0.11	20	5'500
DMW	0.01	11	830

Reaction conditions: 25 °C, 50 mM Tris·HCl pH 8.0, 50 mM NaCl, 200 µM Mn^{II}, 500 µM SAM, 10-500 µM tryptophan or DMW, 5 µM SAH nucleosidase, 10 µM adenine deaminase and 6.6 µM of EgtD$_{M252V,E282A}$.

In addition to the NMR experiments done on the wild type variant of EgtD, we hoped to gain some insights into the binding mode of tryptophan to the engineered EgtD with the point mutations M252V and E282A. We speculated that the 40-fold decrease in k_{cat} observed for this variant (compared to EgtBwt:DMH) could be due to weaker solvation of the N^{α}-methyl group and a decrease in nucleophilicity of the amine. In order to test this hypothesis, we measured the chemical shifts of ^{13}C-labelled DMW in complex with EgtD$_{M252V,E282A}$ with and without SAH as well as free in solution.

The results are similar to those obtained from the analogous measurements with EgtD and DMH. The chemical shift (δ_H: 2.79 ppm, δ_C: 41.54 ppm) is only marginally different than that of

DMH, as could be expected. Upon binding of DMW to the enzyme, a resolution of the single signal into two distinct binding modes can be observed. A slight shift downfield along the hydrogen and almost no shift along the carbon axis hints towards one methyl group pointing into the characteristic C-H···O hydrogen bond pocket **Me1** (δ_H: 2.93 ppm, δ_C: 41.79 ppm; referred to in the following discussion as **Me1$_{DMW}$**). As is the case with DMH, the second shift corresponds to the orientation **Me2**, with little shift along the hydrogen and a substantial shift along the carbon axis (δ_H: 2.85 ppm, δ_C: 44.47 ppm; referred to in the following discussion as **Me2$_{DMW}$**). This downfield carbon shift of Δ_C = 2.9 ppm is slightly less than what was measured for DMH, which was Δ_C = 3.4 ppm.

Once the system is saturated with SAH, **Me1$_{DMW}$** moves downfield along both the hydrogen and carbon axes (δ_H: 3.07 ppm, δ_C: 43.49 ppm), again demonstrating that C-H···O bonding is stronger in the presence of SAH. The shift along the hydrogen axis of Δ_H = 0.28 ppm is nearly the same as the 0.32 ppm for DMH. The carbon shift was slightly more with Δ_C = 1.95 ppm versus 1.27 for DMH. For the second binding position **Me2$_{DMW}$** (δ_H: 2.79 ppm, δ_C: 44.05 ppm) we observed a shift which was slightly less than **Me2$_{DMH}$** (Δ_C = 2.5 and 3.0 ppm, respectively). These shifts are both roughly 0.5 ppm less than the equivalent shifts in the EgtD:DMH and EgtD:DMH:SAH complexes, as summarized in **Table 5**.

Table 5 comparison of the shift along the carbon axis of the position Me2 for DMW and DMH in complex with EgtD and SAH.

	[X]	**[EgtD:X]**	**[EgtD:X:SAH]**
X		δ_C (ppm)	
DMW	41.54	44.47	44.05
DMH	41.63	45.04	44.61
Δ	0.09	0.57	0.56

While the reduction in reaction rate for this mutant versus the wild type enzyme does not seem to correlate with diminished C-H···O bonding in position **Me1**, it could be explained by the weaker coordination to Gly161 in position **Me2**. If this interaction indeed increases the nucleophilicity of N^{α}, this would cause the overall reaction rate to drop.

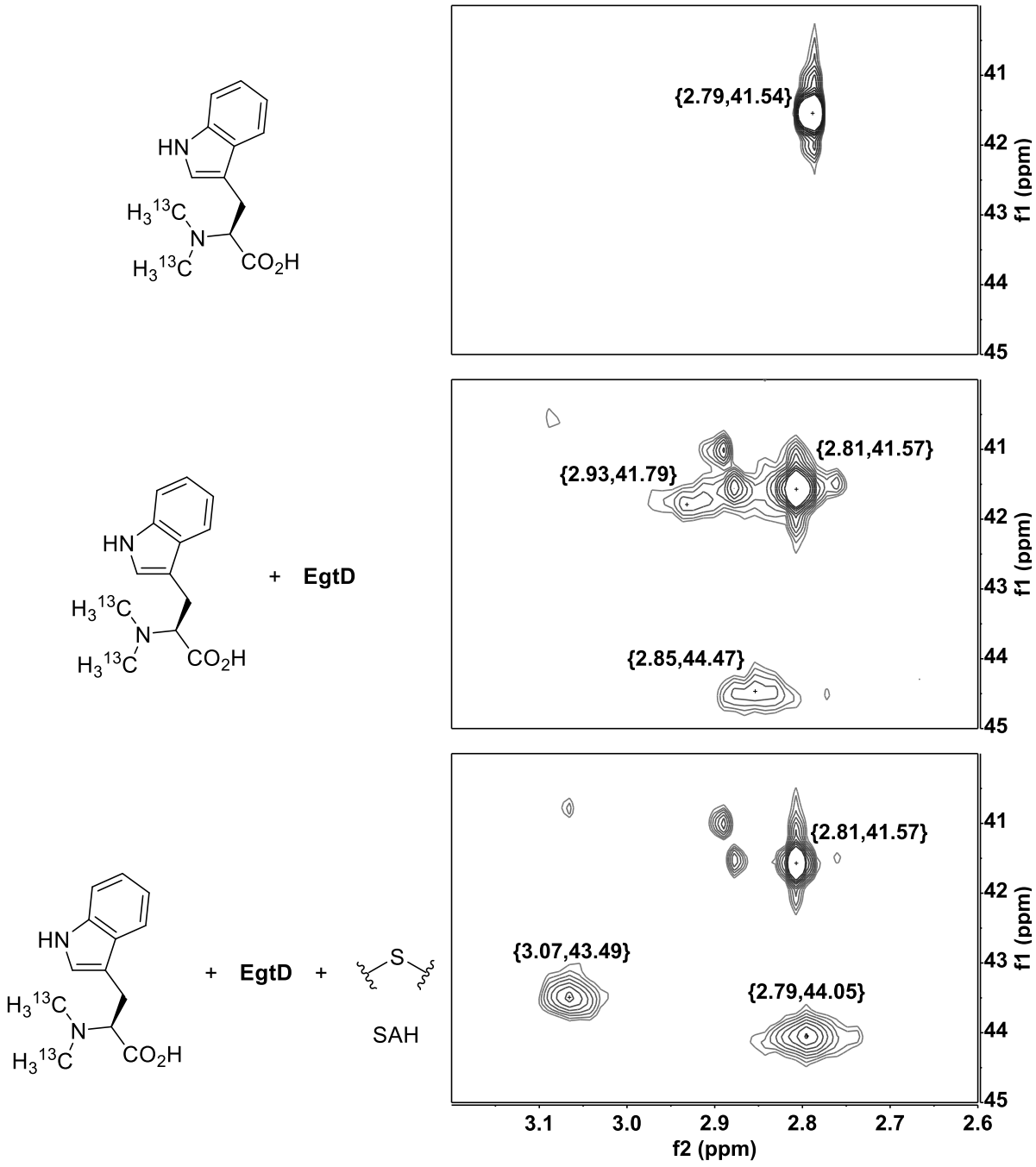


Figure 46 Top: HMQC-spectrum of ^{13}C-DMW [55 µM]. Middle: HMQC-spectrum of ^{13}C-DMW [55 µM] and $EgtD_{M252V,E282A}$ [138 µM]. Bottom: HMQC-spectrum of ^{13}C-DMW [55 µM], $EgtD_{M252V,E282A}$ [138 µM] and SAH [276 µM]. All spectra recorded in 50 mM phosphate buffer, pH 8 containing 50 mM NaCl and 5 % D_2O.

2.3.9 Binding of ^{15}N-Labelled DMH

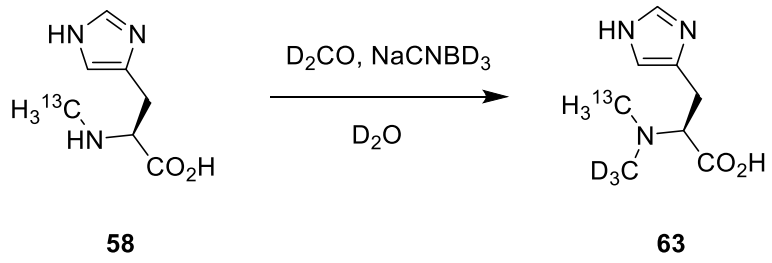
As a final indication that the nucleophilicity of N^{α} is increased upon binding to EgtD, we attempted to show that electron density was transferred onto the amine. In order to measure this, we synthesized the DMH-variant ^{15}N-DMH that contains a ^{15}N-label in the appropriate position by the usual procedure of reductive amination. As starting material we used histidine completely labelled with ^{15}N and ^{13}C. Due to the low sensitivity of the ^{15}N-nucleus, a pseudo 3D-NMR experiment was necessary. By irradiating H^{α} and transferring the magnetization via C^{α} to N^{α} and then back to H^{α} via C^{α} we were able to generate a ^{1}H/^{15}N-HMQC spectrum of ^{15}N-DMH. The signal however was much weaker than in the preceding ^{1}H/^{13}C-HMQC experiments, due to the very small coupling constant between $^{13}C^{\alpha}$ and $^{15}N^{\alpha}$ of J = 4.5 Hz. This resulted in a long transfer delay time from C^{α} to N^{α} and back of around 50 ms, which was larger than the (^{15}N)t_2 relaxation time.

While we were able to obtain a clear, albeit weak signal of the compound free in solution, the signal was lost once the substrate was bound to the enzyme both with and without SAH. It is possible that the signal was hidden under the water peak, in which case it would be impossible to detect under all conditions. If the signal was simply weaker than the signal to noise ratio, one could run a different magnetization-transfer experiment with an appropriate direct detection cryoprobe for ^{15}N. In such an experiment only one chemical shift evolution from H^{α} to C^{α} to N^{α} would be necessary and the magnetization could be directly detected on N^{α}. This however, was not realizable with the instruments at our disposition.

Quantum chemical computations, based on the X-ray structures of three EgtD-substrate complexes, provided some support for our hypothesis of enhanced nucleophilicity of N^{α}(His) upon binding. By calculating the effect of methyl group coordination by Gly161, Corminboeuf *et al.* were able to show that the positive charge of N^{α} in the unbound state was slightly higher than in the bound state. (Unpublished results kindly provided by Prof. Clémence Corminboeuf) This would mean that coordination by Gly161 does indeed increase the electron density on N^{α}, thus rendering it more nucleophilic. Since the nucleus would be more shielded, in the NMR spectrum one would expect an upfield shift away from the water peak. Additional NMR experiments with a cryoprobe which is more sensitive to ^{15}N would certainly be necessary to make a final statement on this matter.

2.3.10 Binding Isotope Effects of Asymmetrically ^{2}H-Labelled DMH

Another way to test our hypothesis of the presence of C-H···O interactions was via a binding isotope effect. Since hydrogen bonds are sensitive to changes in distance,[117] we reasoned that by labelling DMH with one $–CH_3$ and one $–CD_3$ methyl group we might be able to see an asymmetric distribution. We believed that position **Me1**, where C-H···O bonds are formed, would preferentially bind the $–CH_3$ group over the $–CD_3$ group, since C-D bonds are slightly shorter than C-H bonds due to their lower zero point energy.[118] We thus synthesized an asymmetrically labelled variant of DMH (X) from ^{13}C-MMH by the usual reductive amination procedure using deuterated formaldehyde and $NaCNBD_3$, as shown in Figure 47.



58 **63**

Figure 47 Reagents and conditions for the synthesis of **63** from **58**.

Unfortunately, no asymmetric distribution in binding could be observed in our measurements. If there is indeed a preferential binding of the non-deuterated methyl group in the **Me1** position, it is below the limit of detection. The error of peak integration was higher than the observable effect.

The difference in binding energies between C-H···O and C-D···O should be expected to be quite low in any case. Even arguably stronger hydrogen bonds show only very small deuterium isotope effects. For example, in the system $(H_2O···H/D···O_2H)^+$ the H-bonded complex was calculated to be favored over the D-bonded complex by only 0.5 kcal/mol, which is a difference of less than 2 %.[119]

2.3.11 Determination of the Geminal Angle of Methyl C-H Bonds

By creating a stereogenic center on each methyl group, we hoped to be able to compare the germinal coupling constants between the two hydrogen atoms on each methyl group. In case of a decrease in the coupling constant upon substrate binding to the enzyme, one could deduce that the hydrogen atoms are being "pushed" into an S_N2-like transition state where the C-N bond is being weakened and the hydrogen atoms are closer to perpendicular to the N-C···O-axis. We speculated that this effect would be especially pronounced in position **Me2**, where coordination to Gly161 would start to fill the σ*-orbital of the methyl carbon (Figure 48). Correlations between geminal H-H coupling constants and H-C-H bond angle have been determined on systems such as CH_2DCN (mono-deuterioacetonitrile).[120]

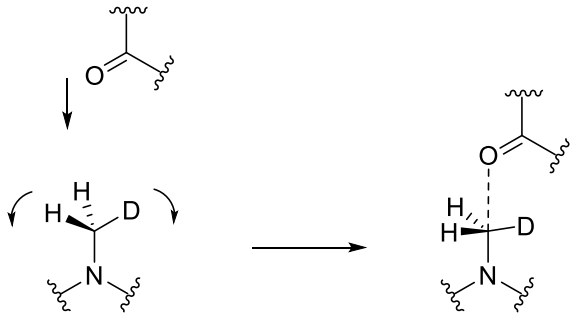
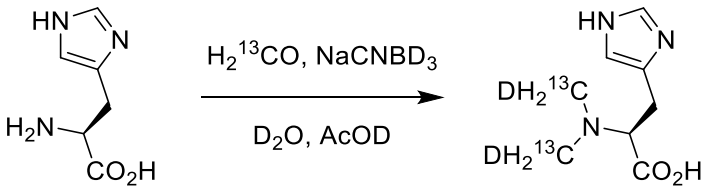


Figure 48 Possible increase in the geminal C-H angle upon binding of the substrate to EgtD.

For our purpose a derivative of DMH was synthesized containing two ^{13}C-methyl groups with two hydrogens and one deuterium each, thus creating methyl groups with diastereotopic hydrogens.



64

Figure 49 Synthesis of N^{α},N^{α}-bis(methyl-^{13}C-*d*)-*L*-histidine (**64**) from *L*-histidine.

The labelled methyl groups showed coupling constants of 144.2 Hz and 1.8 Hz, the first arising from the coupling to the ^{13}C-atom, the smaller one could possibly represent an H-C-D ^2J-coupling as it is similar to that observed in the molecule DMF (1.9 Hz).[121] Multiplication by the ratio of the gyromagnetic constants of hydrogen and deuterium (γH/γD = 6.55) gives a geminal H-C-H ^2J-coupling value of 11.8 Hz, which is just under that determined for methane (12.4 ± 0.6)[122] and well under that determined for acetonitrile (16.9 ± 0.3).[120] The coupling constants

in complex with the enzyme were unfortunately not of sufficient precision to be able to observe minute changes of the H-C-H angle in the methyl group.

2.4 Mechanistic Implications

The molecular toolbox of EgtD, which it uses to catalyze methyl transfer, can be divided into three parts: C-H···O bonding, n→σ* interactions and acid-base catalysis. The last of these was characterized comprehensively by Laëtitia Misson in her dissertation,[99] the former two were the focus of this work. In this last section, we would like to summarize our findings and what we have been able to learn with respect to the enzyme mechanism and its overall strategy.

By a combination of NMR and crystallography, we have characterized some of the molecular strategies employed by EgtD for catalysis. On the one hand, we have added another example to the group of methyltransferases where C-H···O bonding plays a role in the methyl transfer reaction from SAM to the acceptor.[44,123,124] On the other hand we propose a unique method of nucleophilic activation by unusual attractive coordination of two carbonyl residues to the substrate(s).[102] Binding via these unusual interactions has been shown both on the methyl donor as well as on different acceptors (MMH and DMH).

2.4.1.1 C-H···O Bonding

The first binding site we could observe was **Me1**, between the sulfur of SAM/SAH and the nitrogen acceptor, featuring the C-H···O bonding triangle formed by Phe47, Tyr39 and Thr163. Upon binding to this site, the chemical shift of one methyl group in DMH moves downfield slightly by 0.13 ppm. After addition of SAH an additional downfield shift is observed leading to a total shift of 0.32 ppm from the unbound to the bound state with SAH. The strength of the hydrogen bonds thus increases once the substrate is locked into the active site by SAH. In the binding of MMH this site is only occupied when SAH is present, leading to a downfield shift of 0.19 ppm. When this binding site is occupied by the methyl group on SAM, the downfield shift is less pronounced. When αClHis is used as the co-substrate the shift is 0.17 ppm, when methylhistidine is used the shift is only 0.02-0.07 ppm. Coordination of the transferred methyl group by C-H···O bonding is thus weaker when it is on the donor and stronger on the acceptor.

However, the downfield shift of the methyl group on the carbon scale was much larger on SAM, both in complex with methylhistidine (5.4 ppm) and chlorohistidine (6.4 ppm), than on the acceptors MMH (2.4 ppm with SAH) or DMH (1.3 ppm with SAH). The electrophilic character of the methyl group is thus increased on the donor versus the acceptor, as expected.

Analysis of the crystal structures of EgtD in complex with either SAM:αClHis or SAH:DMH provides a comparison of the binding before and after methyl transfer, respectively. In the first structure where C-H···O bonds are formed with the methyl group on SAM, the angle of these bonds relative to the S-C axis is nearly perpendicular. In the second structure where C-H···O bonds are formed with the methyl group on DMH, the angle becomes smaller. According to these models there would be more strain on the C-H bonds in the first structure, before methyl

transfer takes place. By these interactions it is conceivable that EgtD brings the substrate methyl-group closer to the geometry of the transition state and enhances its reactivity in this fashion. This in turn is observable as the strong downfield shift along the carbon axis, indicating increased electrophilicity of the carbon center.

Compared with a similar system, the lysine methyl transferase SET7/9, the proton chemical shift we measured and hence the hydrogen bonding effect is weaker. The main difference between the C-H···O bonding motives in EgtD and SET7/9 is that the former consists of two C-H···O and one C-H···π interaction, while the latter forms three C-H···O bonds. In contrast to this enzyme however, EgtD catalyzes three methylation steps instead of only one.

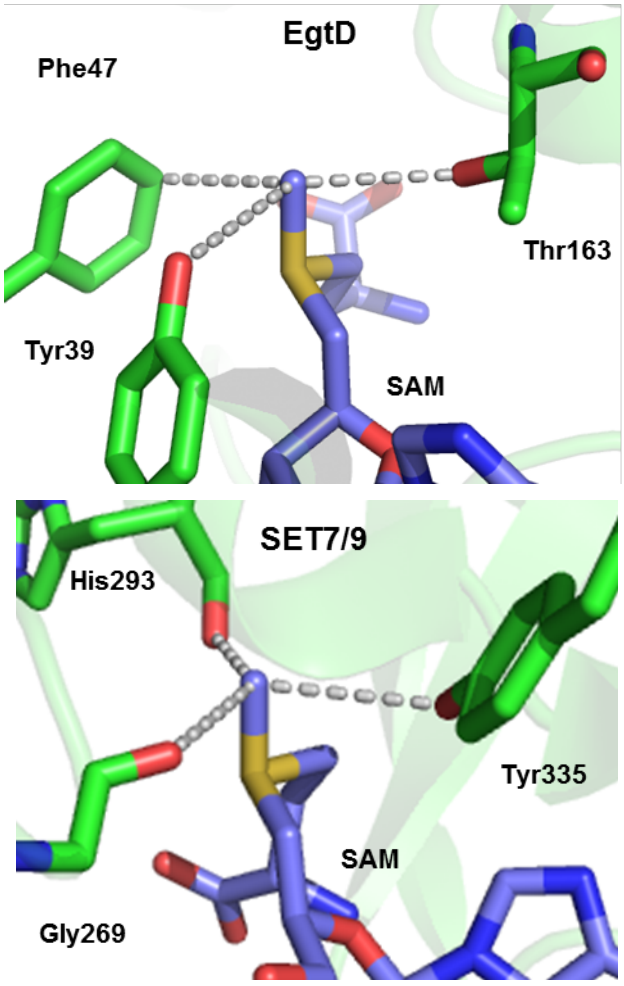


Figure 50 Active sites of EgtD (left, unpublished structure) and SET7/9 (right, PDB: 1N6A) showing the C-H···O/π interactions between the methyl group of SAM (purple) with the respective residues (green). The indicated distances to the carbons of each methyl group on SAM are as follows. EgtD: Tyr39 (3.9 Å), Phe47 (3.6 Å), Thr163 (3.8 Å); SET7/9: Gly269 (3.7 Å), His293 (3.0 Å), Tyr335 (3.5 Å).

Since the average distance to the coordinating residues is very similar (EgtD 3.8 ± 0.2 Å, SET7/9 3.4 ± 0.4 Å), this can hardly be used as an argument to justify this observation. The most obvious possibility for the weaker bonding in EgtD than in SET7/9 is the substitution of

one C-H···O bond with a C-H···π interaction. The latter are generally weaker with an enthalpy of 0.5-2 kcal/mol,[125] while the former has been calculated to be as high as around 9 kcal/mol for strongly polarized systems.[126] Testing this hypothesis would be a bit of a challenge, since introducing a point mutation replacing Phe47 with a potential C-H···O bond acceptor could significantly reduce the catalytic efficiency of the enzyme.

In any case, the main role of these interactions seems to be the coordination of the transition state (TS). In a recent study on human DNA methyltransferase 1 (DNTM1), Schramm *et al.* conducted a number of KIE measurements, which they combined with QM calculations to generate a model of the TS at subangstrom resolution.[127] Their exhaustive analysis resulted in a picture of a relatively loose TS, in contrast to the compressed TS suggested by some[50,52] and disputed by others[51,56] for COMT (see also **Section 2.1.2**).

Similar KIE/computational studies to those of Schramm *et al.* have also been done recently on different histone KMTs. In a study on SET8, Linscott *et al.* suggest an early S_N2 TS with a long N-C distance (2.35-2.40 Å) and a short C-S distance (2.00-2.05 Å), with the methyl hydrogens coordinated by noncanonical C-H···O bonding with Cys270, Arg295, and Tyr336.[128] In NSD2, a related histone KMT, Poulin *et al.* described the opposite case, with a late S_N2 TS with a short N-C distance (2.02-2.10 Å) and a long C-S distance (2.50 Å).[123]

Since there seems to be a more or less general consensus that enzymes catalyze reactions by stabilization of the transition state,[129] it makes sense to discuss our own findings in the light of Hammond's original postulate on transition state theory.[130] According to this, when reactant and product are of equal energy, the transition state should be exactly in the middle of the energy diagram and its structure should be equally similar to both (ii in Figure 51). If the reactant is of higher energy, the structure of the transition state should be more similar to that of the reactant and is termed an early transition state (i in Figure 51). In the opposite case, where the product is higher in energy than the reactant, the transition state is more similar to the product and one speaks of a late transition state (iii in Figure 51).

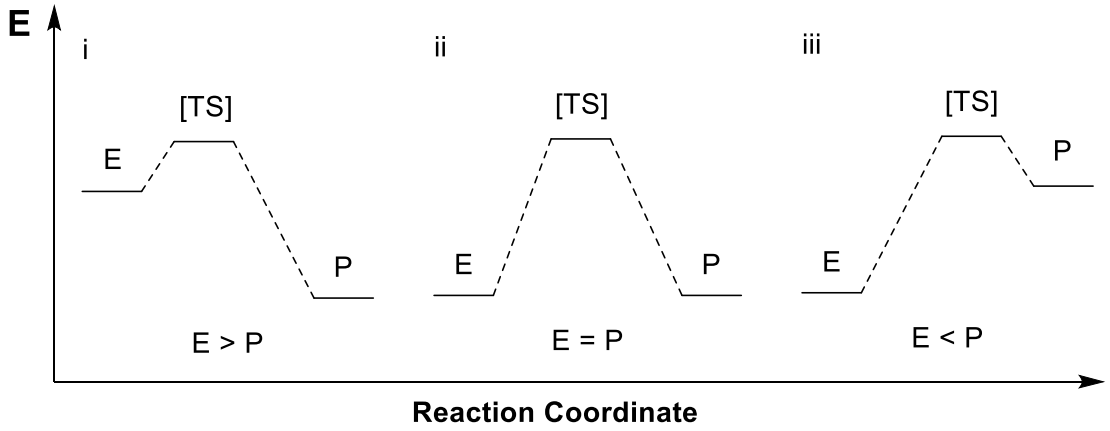


Figure 51 Graphical representation on Hammond's postulate on transition state theory.

Because the methyl group on SAM is clearly more electrophilic – as we have shown in our NMR analysis – and reactive than on MMH or DMH, an early transition state should be favored. Hence, stabilization of this state might be a strategy for EgtD to facilitate methyl transfer. In this unobservable transition state we would expect C-H···O bonding to be strongest due to the planar nature of the intermediate, where the H···O distance would be at a minimum. However, as discussed above, this is a slightly oversimplified model and many more subtle effects may play a role in defining whether a TS is early or late and tight or loose.

Hence, one option for elucidating the character of the EgtD TS would be to conduct the same combination of KIE determination with computational modelling.

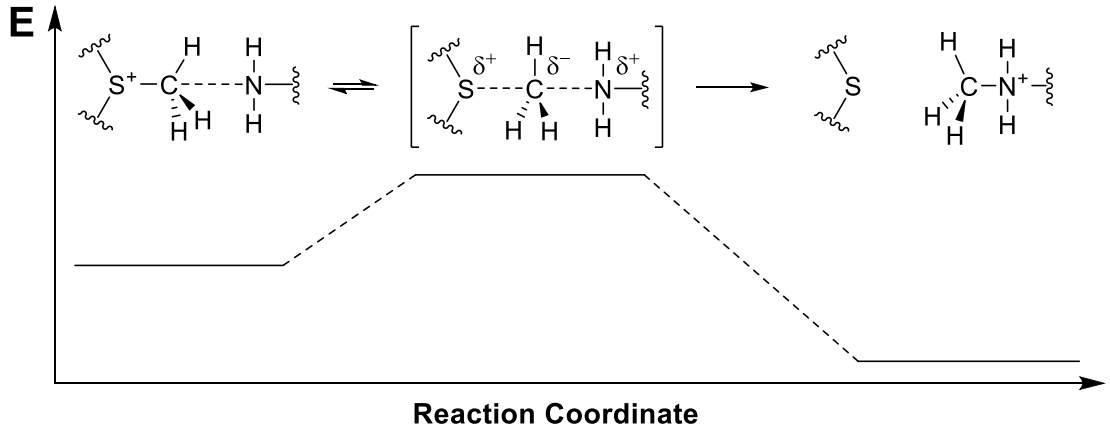


Figure 52 Transition state in the S_N2-type methyl transfer from SAM onto histidine.

This hypothesis could be tested with kinetic experiments using isotopically labelled SAM. When using SAM containing a methyl group with hydrogen substituted for deuterium or tritium one would expect an inverse secondary KIE for the methyl transfer step in the case that the TS is stabilized by equatorial C-H···O bonds.[131] Others may however argue that the only true

way of elucidating what exactly is the cause of the catalytic efficiency is by computational methods.

2.4.1.2 Carbonyl Coordination

The second binding site we could observe was **Me2**, where the methyl group is coordinated to the backbone carbonyl of Gly161. We propose that n→σ* interactions between the lone pair orbital of the carbonyl oxygen and the antibonding σ-orbital of the methyl carbon can be observed in the ^{13}C-NMR shift of the methyl carbon coordinated to this position. Upon binding, the resulting deshielding of the methyl carbon is observable as a downfield shift of 2.4-3.4 ppm, depending on the substrate. After the first methylation of histidine, this is the position occupied by the N^{α}-methyl of MMH, making room for the next methyl group to enter the active site on the second equivalent of SAM.

By demonstrating that the methyl carbon in the **Me2** position becomes electron deficient upon coordination to the enzyme we can deduce that the adjacent N^{α} becomes more nucleophilic (see also **Section 2.1.5**). This may also be the reason for the drop in catalytic efficiency of the EgtD M252V E282A mutant, which catalyzes the (di-)methylation of tryptophan. In our NMR experiments with DMW we observed a smaller downfield shift of the methyl carbon in position **Me2** than for DMH, both with and without SAH. Although the computational model of Corminboeuf *et al*. (see **Section 2.3.9**) supports our assumption, solidification of this hypothesis would warrant some sort of direct observation, for example on the NMR shift of N^{α}.

The second carbonyl coordination position **Me3** could not be observed in our NMR experiments. The methyl group of MMH preferentially coordinates to **Me1** (with SAH) and **Me2** (without SAH) and both methyl groups in DMH are coordinated to **Me1** and **Me2** (with and without SAH). Only before the final methylation step is **Me3** occupied by one of the DMH methyl groups and only during SAM-binding. While we were not able to observe this position in our NMR experiment with TMH, we can see coordination of the third methyl group to Asn166 in the crystal structure (Figure 41).

The complete catalytic mechanism we propose is shown in Figure 53. In **A**, histidine, the first substrate, is bound and deprotonated by Asn166 in **Me3** when SAM enters the active site in **B**. After methyl group transfer from SAM onto MMH in **C**, SAH leaves the active site and the methyl group on MMH rotates into **Me2** in **D**. The substrate is once again deprotonated by Asn166 upon binding of the second SAM equivalent in **E**. The second transferred methyl group remains coordinated in **Me2** in the presence (**F**) and absence (**G**) of SAH. Only after deprotonation of N^{α} and binding of the third equivalent of SAM does one methyl group occupy the **Me3** position during methyl transfer (**H**). Evidence for all intermediate ground states is provided by either NMR (**C**, **D**), crystal structures (**I**) or both (**F**, **G**).

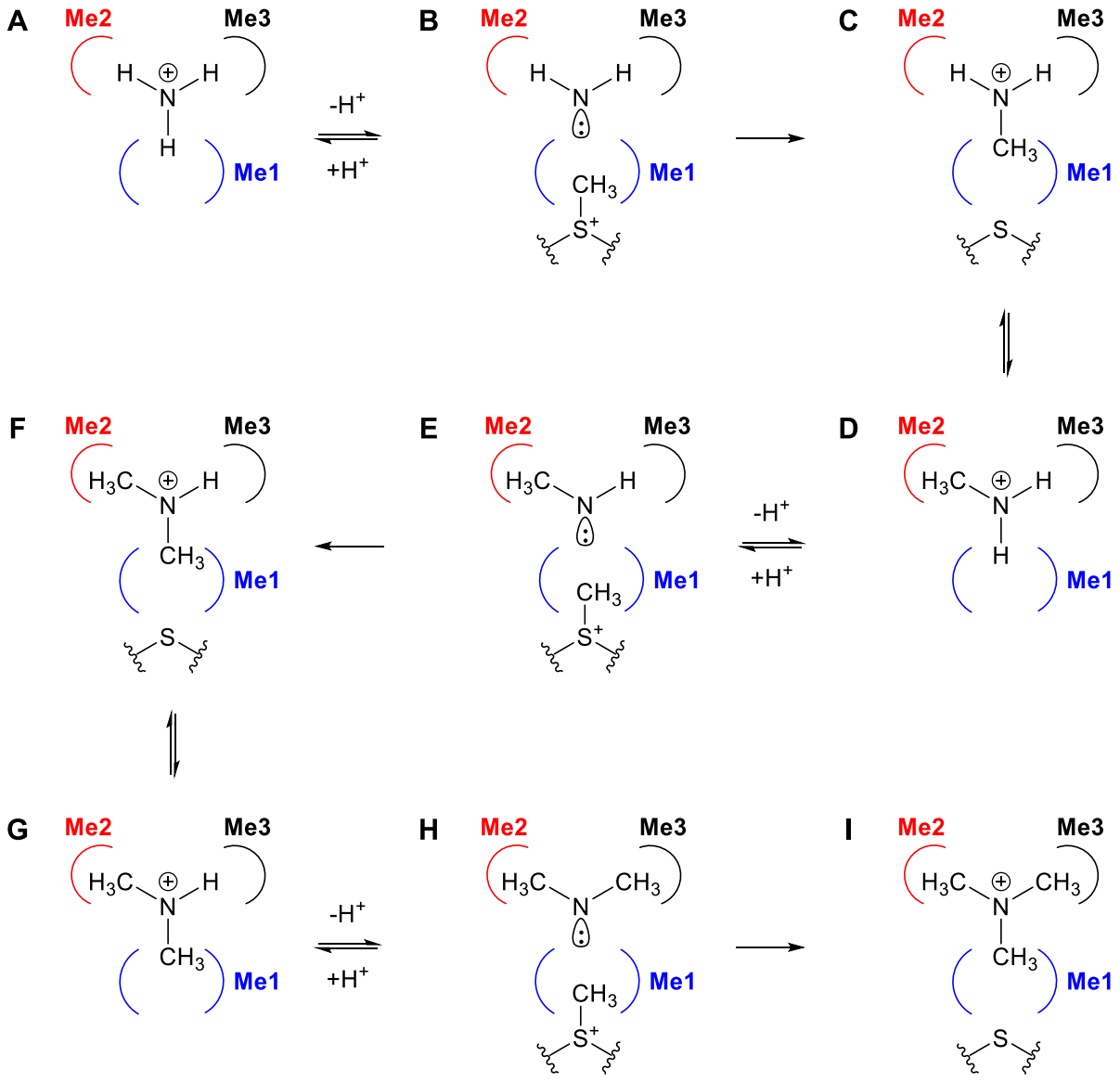


Figure 53 Proposed mechanism of the three methylation steps catalyzed by EgtD. **Me1** is the binding site formed by the C-H···O bonding triangle consisting of Tyr39, Phe47 and Thr163. **Me2** is the binding site with coordination to the backbone carbonyl of Gly161. **Me3** is the binding site with coordination to the sidechain carbonyl of Asn166, which also acts as a catalytic base in the deprotonation steps.

2.5 Conclusion

In conclusion, we have shown direct evidence for C-H···O bonding, both on the methyl donor as well as on a selection of acceptors via NMR. These findings are in excellent agreement with the available crystal structures and in the case of MMH provide new insights into methyl group coordination in the catalytic mechanism. We propose that C-H···O bonding may be a more widespread strategy of SAM-dependent methyltransferases to facilitate methyl group transfer than has previously been shown. We hope that our contribution will inspire research into similar systems, such as the glycine sarcosine *N*-methyltransferase discussed in **Section 2.1.4.2**.

We have also shown that a second, novel type of interaction is at play during catalysis, namely the transfer of electron density from two carbonyl groups to N^{α} via n→σ* interactions with the adjacent methyl carbons. Although direct experimental evidence for increased electron density on the nitrogen of interest could not yet be obtained, computational models support our assumption. Furthermore, all the indirect evidence obtained from our experiments strongly backs our hypothesis.

In contrast to the discussion surrounding the reasons for methyltransferase efficiency outlined in **Section 2.1.2**, where the transition state is the main focus, all our measurements represent snapshots of ground states. Further experiments would be warranted if the transition state were to be addressed more directly, either via kinetic isotope effects, QM/MM or a combination of the two.

2.6 Experimental

2.6.1 NMR Spectroscopy

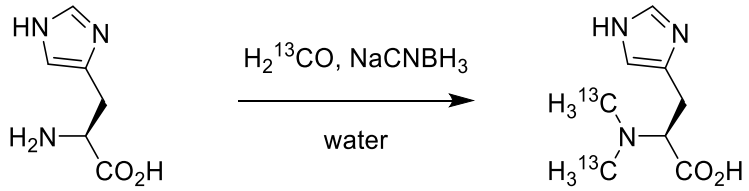
The ^{1}H-^{13}C HSQC experiments were performed using a sensitivity enhanced pulse sequence with watergate solvent suppression.[132] Typically, 512 increments were performed in the indirect carbon dimension and 2048 data points were acquired in the proton dimension. The acquisition time was 141 ms in F2 and 11 ms in F1 dimensions, the total experiment time varied between 40 min and 12 hours. Processing was performed using the Bruker Topspin 3.2 software and zero-filling to 1k data points was applied in F1.

2.6.2 Synthesis

Reductive amination using formaldehyde and $NaCNBH_3$ – Procedure A: The starting material (1 eq) was dissolved in water (10 ml per mmol) and the pH was adjusted to 5.5 – 6 by addition of glacial AcOH. Formaldehyde (2-5 eq) was added and the solution was stirred at room temperature for 15 min before adding $NaCNBH_3$ (2-3 eq), readjusting the pH if needed and stirring for 2-18 h. Conversion of starting material was checked by either ^{1}H-NMR or IE-HPLC (using the method shown in **Table 14** of **Section 3.6.2.1**) and if starting material was still detected, another portion of formaldehyde (1-2 eq) was added and the reaction was stirred for an additional 2 h. When the reaction was complete it was quenched by acidifying to pH 1 with concentrated HCl and stirred vigorously for 30 min. Unless noted otherwise, products were purified by procedure Xb.

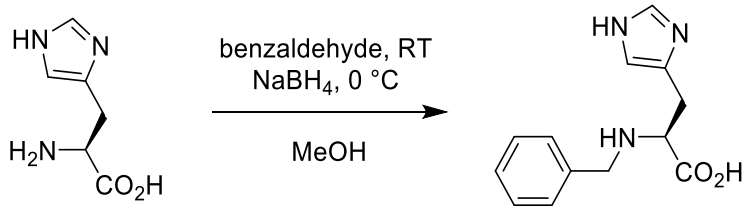
Purification by cation exchange chromatography – Procedure B: The crude sample was dissolved in water and the pH was adjusted to 1 by addition of HCl (either concentrated or 1 M, depending on sample size) and the volume was increased so that the total concentration of NaCl or other unwanted inorganic salts was <50 mM. The sample was loaded onto a flash column packed with Dowex® cation exchange resin (50WX-8) and the loaded resin was washed with several volumes of water. Product elution was carried out using 100 or 250 mM NH_4OH and fractions containing product were identified by TLC (1-BuOH/AcOH/water 3:1:1), staining with ninhydrin. All fractions containing product were combined and dried either by rotary evaporation or by lyophilization.

Synthesis of N^{α},N^{α}-di(methyl-^{13}C)-*L*-histidine (55)



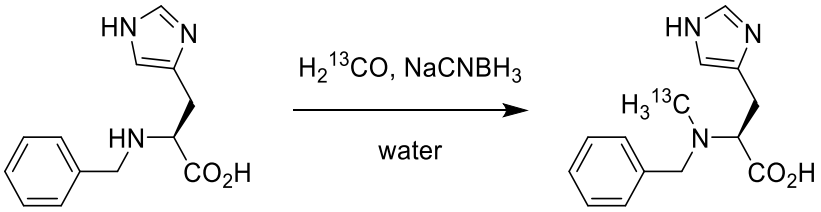
Using procedure A and starting with *L*-histidine (225 mg, 1.45 mmol, 1 eq), ^{13}C-formaldehyde (20 %, 1 ml, 7.26 mmol, 5 eq) and $NaCNBH_3$ (273 mg, 4.35 mmol, 3 eq), this product was obtained as white solid (222 mg, 1.21 mmol, 84 %) after purification according to procedure B and lyophilization. **^{1}H NMR (400 MHz, D_2O)** δ 7.63 (s, 1H), 6.94 (s, 1H), 3.71 (tt, $^3J_{HH}$ = 6.6 Hz, $^3J_{CH}$ = 2.5 Hz, 1H), 3.17 – 3.07 (m, 2H), 2.74 (dd, $^1J_{CH}$ = 142.2, $^3J_{CH}$ = 4.1 Hz, 6H). **HRMS** *m/z* calcd. for $C_6{}^{13}CH_{14}N_3O_2$ [M+H$^+$]: 186.1148, found: 186.1147

Synthesis of N^{α}-benzyl-*L*-histidine (61)



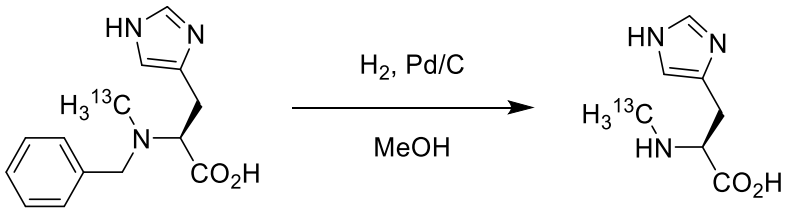
To a suspension of *L*-histidine (1.551 g, 10.00 mmol, 1 eq) in dry MeOH (25 ml) was added finely ground NaOH (420 mg, 10.50 mmol, 1.05 eq) and the solution was stirred at room temperature for 30 min, in which time the starting material dissolved. Benzyladehyde was added (3.55 ml, 34.99 mmol, 3.5 eq) and the clear solution was cooled to 0 °C before adding $NaBH_4$ (945 mg, 24.99 mmol, 2.5 eq) and the solution was stirred for 2 h at 0 °C. The reaction was acidified to pH 1 with conc. HCl and the organic solvent was removed under reduced pressure. The residue was dissolved in water and the product was purified by cation exchange chromatography according to procedure B, obtaining the pure product (2.284 g, 9.31 mmol, 93 %) as white solid after lyophilization. **^{1}H NMR (400 MHz, MeOD)** δ 8.92 (d, *J* = 1.4 Hz, 1H), 7.63 – 7.58 (m, 3H), 7.48 – 7.45 (m, 3H), 4.48 (dd, *J* = 8.4, 5.2 Hz, 1H), 4.42 – 4.35 (m, 2H), 3.67 – 3.61 (m, 1H), 3.56 – 3.50 (m, 1H). **MS** *m/z* calcd. for $C_{13}H_{16}N_3O_2$ [M+H$^+$]: 246.12, found: 246.03;

Synthesis of N^{α}-benzyl-N^{α}-(methyl-^{13}C)-*L*-histidine (62)



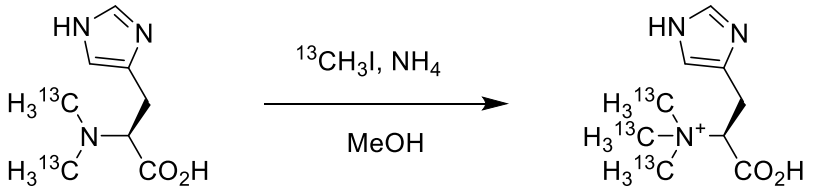
Using procedure A and starting with **61** (200 mg, 0.82 mmol 1 eq), ^{13}C-formaldehyde (20 %, 225 µL, 1.63 mmol, 2 eq) and $NaCNBH_3$ (102 mg, 1.63 mmol, 2 eq) this product was obtained as off-white solid (174 mg, 0.67 mmol, 82 %) after purification according to procedure B and lyophilization. **^{1}H NMR (400 MHz, D_2O)** δ 7.77 (s, 1H), 7.48 – 7.41 (m, 2H), 7.02 (s, 1H), 4.25 – 4.16 (m, 1H), 3.84 (td, *J* = 7.0, 6.9, 2.7 Hz, 0H), 3.28 – 3.18 (m, 1H), 2.69 (d, $^1J_{CH}$ = 142.0 Hz, 3H). **MS** *m/z* calcd. for $C_{13}{}^{13}CH_{18}N_3O_2$ $[M+Na]^+$ 261.14, found 261.01.

Synthesis of N^{α}-(methyl-^{13}C)-*L*-histidine (58)



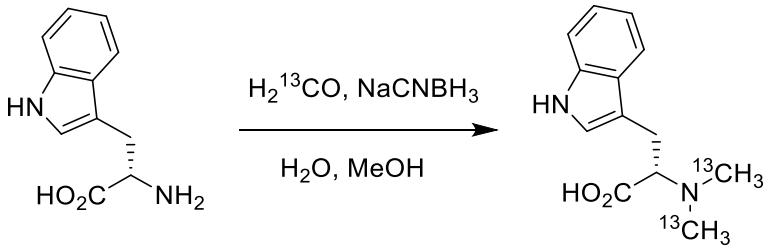
To a solution of **62** (174 mg, 0.67 mmol, 82 %) in dry MeOH (2 ml) was added 10 % Pd/C (71 mg, 0.07 mmol, 0.1 eq) an atmosphere of H_2 (1 atm) was applied and the reaction mixture was stirred at room temperature overnight. After reaction control by ^{1}H-NMR and completion of the reaction, the mixture was filtered over a pad of celite and evaporated to dryness. The product was obtained as an off white solid (100 mg, 0.59 mmol, 88 %) after purification by procedure B. **^{1}H NMR (400 MHz, D_2O)** δ 8.67 (d, *J* = 1.4 Hz, 1H), 7.42 (d, *J* = 1.1 Hz, 1H), 4.22 (ddd, *J* = 7.4, 5.4, 2.8 Hz, 1H), 3.53 – 3.47 (m, 1H), 3.45 – 3.39 (m, 1H), 2.79 (d, *J* = 144.2 Hz, 4H). **HRMS** *m/z* calcd. for $C_6{}^{13}CH_{12}N_3O_2$ $[M+H]^+$ 171.0958, found 171.0958.

Synthesis of N^{α},N^{α},N^{α}-tri(methyl-^{13}C)-*L*-histidine (57)



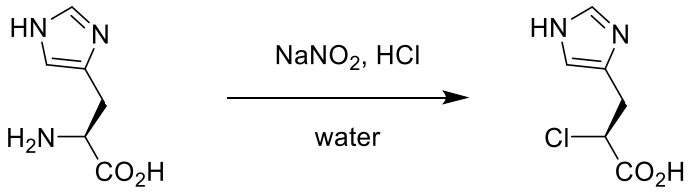
To a solution of **55** (120 mg, 0.65 mmol, 1 eq) in dry MeOH (3 ml) was added conc. NH_4OH (40 µL, 0.65 mmol, 1 eq) and $^{13}CH_3I$ (58 µL, 0.92 mmol, 1.4 eq) and the solution was stirred at room temperature overnight. Analysis by IE-HPLC (**Table 14**, in **Section 3.6.2.1**) showed 14 % of the starting material had not been converted, thus an additional portion of $^{13}CH_3I$ (16 µL, 0.26 mmol, 0.4 eq) was added and the reaction mixture was stirred overnight. The solvent was removed under reduced pressure and the residue was dissolved in a minimal amount of warm MeOH. To this, Et_2O was added and the crude product was allowed to precipitate in the fridge overnight. The solvent was decanted and the remaining solid was dried, dissolved in water and the product was purified by anion exchange chromatography, using Amberlite® IRA-410, where the product does not bind to the resin. After lyophilization of product-containing fractions (stained with ninhydrin), the pure product (99 mg, 0.50 mmol, 76 %) was obtained as colorless solid. **^{1}H NMR (400 MHz, D_2O)** δ 7.68 (d, *J* = 1.2 Hz, 1H), 6.97 (d, *J* = 1.2 Hz, 1H), 3.88 (ddt, $^3J_{HH}$ = 10.7, $^3J_{HH}$ = 4.4, $^3J_{CH}$ = 2.3 Hz, 1H), 3.25 (dt, $^1J_{CH}$ = 145.1, $^3J_{CH}$ = 3.4, $^3J_{CH}$ = 3.4 Hz, 3H), 3.28 – 3.14 (m, 3H). **HRMS** *m/z* calcd. for $C_6{}^{13}C_3H_{16}N_3O_2$ $[M+H]^+$ 201.1338, found 201.1339.

Synthesis of N^{α},N^{α}-di(methyl-^{13}C)-*L*-tryptophan (56)



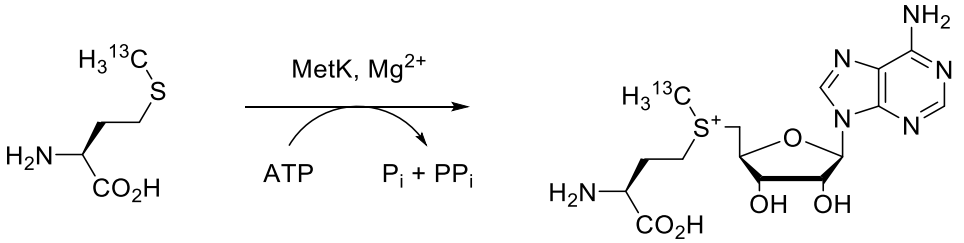
Using procedure A and starting with *L*-tryptophan (20 mg, 0.10 mmol, 1 eq), ^{13}C-formaldehyde (20 %, 54 µL, 0.39 mmol, 4 eq) and $NaCNBH_3$ (15 mg, 0.24 mmol, 2.5 eq), this product was obtained as white solid (29 mg, 0.10 mmol, quant.) after purification according to procedure B and lyophilization. **^{1}H NMR (500 MHz, D_2O)** δ 7.72 (d, *J* = 7.9 Hz, 1H), 7.51 (d, *J* = 8.1 Hz, 1H), 7.27 – 7.24 (m, 1H), 7.21 (s, 1H), 7.20 – 7.16 (m, 1H), 3.44 (dt, *J* = 14.0, 4.2, 4.2 Hz, 1H), 3.20 – 3.16 (m, 1H), 3.10 – 3.06 (m, 1H), 2.43 (dd, *J* = 135.8, 4.8 Hz, 9H). **HRMS** *m/z* calcd. for $C_{11}{}^{13}C_2H_{17}N_2O_2$ $[M+H]^+$ 235.1352, found 235.1349.

Synthesis of (*S*)-2-chloro-3-(1*H*-imidazol-4-yl)propanoic acid (X)



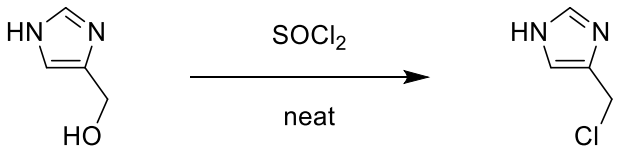
To a stirred solution of *L*-histidine (200 mg, 1.29 mmol, 1 eq) in conc. HCl (3 ml) was slowly added a cold solution of $NaNO_2$ (320 mg, 4.64 mmol, 3.6 eq) in water (1 ml) at 0 °C. The reaction mixture was stirred at 0 °C for 1 h and then at RT for 5 h. The solvent was removed by evaporation and the product was extracted from the white precipitate with 3 x acetone followed by evaporation of the organic solvent to give the HCl salt of the product (95 mg, 0.45 mmol, 35 %) as clear oil.**^{1}H NMR (400 MHz, D_2O)** δ 8.62 (d, *J* = 1.5 Hz, 1H), 7.37 (s, 1H), 4.82 (dd, *J* = 6.7, 5.6 Hz, 1H), 3.51 – 3.40 (m, 2H). **HRMS** *m/z* calcd. for $C_6H_8ClN_2O_2$ $[M+H]^+$ 175.0269, found 175.0269.

Synthesis of *S*-adenosyl ^{13}*C* methionine



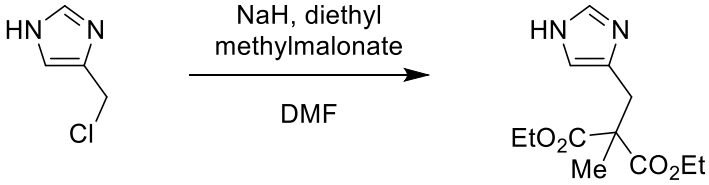
SAM was synthesized enzymatically using SAM synthetase metK, which was kindly provided by Alice Harnacke. The reaction was carried out in a total volume of 750 µL containing 100 mM Tris buffer (pH 8), 10 mM ^{13}C-Met, 13 mM ATP, 1 mM EDTA, 26 mM $MgCl_2$, 50 mM KCl and 8 % β-mercaptoethanol. The reaction was initiated by adding 3 µM of purified metK, after which the reaction was shaken for 5 h. The reaction was monitored by IE-HPLC until no more consumption of ATP could be detected, at which point the reaction mixture was flash frozen and lyophilized. After resuspension and centrifugation, the supernatant was purified by RP-HPLC and pure fractions were lyophilized to give the product as the triple TFA salt (2.5 mg, 3.4 µmol, 34 %) as white powder. **^{1}H NMR (500 MHz, D_2O)** δ 8.47 (d, *J* = 2.7 Hz, 2H), 6.19 (d, *J* = 4.0 Hz, 1H), 4.88 (dd, *J* = 5.5, 4.0 Hz, 1H), 4.64 (t, *J* = 5.7, 5.7 Hz, 1H), 4.60 – 4.57 (m, 1H), 4.06 – 4.01 (m, 1H), 3.96 – 3.93 (m, 1H), 3.86 (t, *J* = 6.6, 6.6 Hz, 1H), 3.74 (s, 2H), 3.73 – 3.67 (m, 1H), 3.52 – 3.45 (m, 1H), 3.02 (d, *J* = 146.6 Hz, 3H), 2.40 – 2.35 (m, 2H). **HRMS** *m/z* calcd. for $C_{14}{}^{13}CH_{23}ClN_6O_5S_1$ $[M]^+$ 400.1479, found 400.1482.

Synthesis of 4-(chloromethyl)-1*H*-imidazole hydrochloride (65).



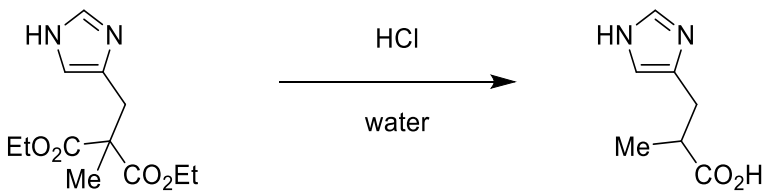
To 4-(hydroxymethyl)imidazole (613 mg, 6.25 mmol, 1 eq.) was added thionyl chloride (2.2 ml, 31,3 mmol, 5 eq.). The mixture was stirred for 1 day then the solvent was removed at the HV to obtain the title compound (950 mg, 6.2 mmol, 99%) as a beige solid. **^{1}H NMR (400 MHz, DMSO-d6)** δ 14.81 (bs, 2H), 9.16-9.10 (m, 1H), 7.88 – 7.64 (m, 1H), 4.88 (s, 2H).

Synthesis of diethyl 2-((1*H*-imidazol-4-yl)methyl)-2-methylmalonate (66).



To a stirred solution of diethyl methylmalonate (0.15 ml, 0.87 mmol, 1 eq.) in DMF (3ml) was added sodium hydride (60% mineral oil dispersion, 60 mg, 1.5 mmol, 2.3 eq.) at 0°C. To this mixture was added **65** (100 mg, 0.65 mmol, 1.0 eq.) portion wise. The mixture turned yellow. Stirred for 24 h at room temperature. The reaction was diluted with ethylacetate (10 ml), washed with 1 M sodium hydroxide (3x 10 ml) and dried over sodium sulfate. The solvent was removed *in vacuo* and the crude was purified by column chromatography (DCM/MeOH 94:4) to yield the title compound as white solid (67 mg, 0.26 mmol, 40%). **^{1}H NMR (400 MHz, DMSO-d6)** δ 11.80 (s, 1H), 7.49 (d, J = 1.2 Hz, 1H), 6.73 (s, 1H), 4.35 – 3.85 (m, 4H), 3.01 (s, 2H), 1.25 (s, 3H), 1.16 (t, J = 7.1 Hz, 6H). **ESI-MS** *m/z* calcd. for $C_{12}H_{19}N_2O_4$ $[M+H]^+$ 255.13, found 255.0

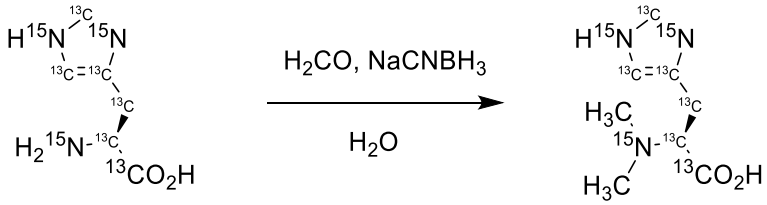
Synthesis of 3-(1*H*-imidazol-4-yl)-2-methylpropanoic acid (60).



To a stirred solution of **66** (30 mg, 0.12 mmol, 1 eq.) in water (0.35 ml) was added concentrated HCl (0.35 ml). The mixture was heated to reflux and stirred for 48 hours. The solvent was

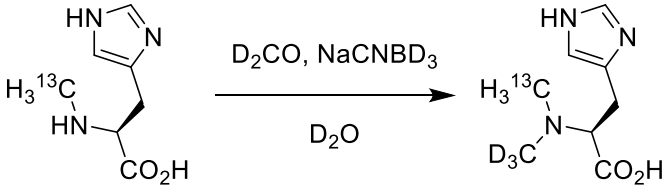
removed *in vacuo* to yield the title compound as colorless oil (22 mg, 0.12 mmol, 98%). **^{1}H NMR (400 MHz, D_2O)** δ 8.59 (d, J = 1.4 Hz, 1H), 7.29 – 7.26 (m, 1H), 3.09 – 2.76 (m, 3H), 1.25-1.15 (m, 3H). **HRMS** *m/z* calcd. for $C_7H_{11}N_2O_2$ $[M+H]^+$ 155.0815, found 155.0816.

Synthesis of N^{α}-dimethyl-*L*-histidine-$^{13}C_6$,$^{15}N_3$ (67)



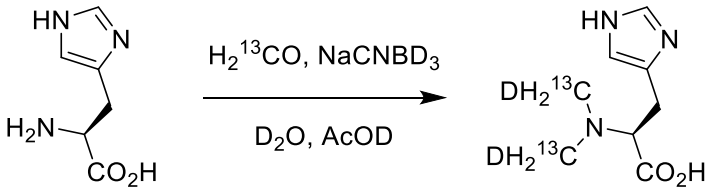
Using procedure A and starting with *L*-histidine-$^{13}C_6$,$^{15}N_3$ (10 mg, 0.05 mmol, 1 eq), formaldehyde (37 %, 17 µl, 0.23 mmol, 5 eq) and $NaCNBH_3$ (7 mg, 0.11 mmol, 2.5 eq), this product was obtained as white solid (7 mg, 0.04 mmol, 82 %) after purification according to procedure B and lyophilization. **^{1}H NMR (600 MHz, D_2O)** δ 7.72 (dtd, $^1J_{CH}$ = 209.6 Hz, $^2J_{NH}$ = 9.5 Hz, $^4J_{HH}$ = 5.0 Hz, 1H), 7.02 (ddd, $^1J_{CH}$ = 190.9 Hz, $^2J_{NH}$ = 9.1 Hz, $^4J_{HH}$ = 5.0 Hz, 1H), 3.85 (ddt, $^1J_{CH}$ =146.9, $^2J_{NH}$ = 5.6 Hz, $^3J_{HH}$ = 6.2 Hz, 1H), 3.24 (d, $^1J_{CH}$ = 131.2 Hz, 2H), 2.89 (d, $^2J_{NH}$ = 3.1 Hz, 6H). **MS** *m/z* calcd. for $C_2{}^{13}C_6H_9{}^{15}N_3O_2$ $[M+H^+]$: 188.08, found: 188.09.

Synthesis of N^{α}-(methyl-*d_3*)-N^{α}-(methyl-*^{13}C*)-*L*-histidine (63)



Using procedure A and starting with **58** (46 mg, 0.27 mmol, 1 eq), d_2-formaldehyde (20 %, 119 µL, 0.81 mmol, 3 eq) and $NaCNBD_3$ (34 mg, 0.54 mmol, 2 eq) adjusting to pH 5 with *d*4-acetic acid, this product was obtained as white solid (30 mg, 0.16 mmol, 59 %) after purification according to procedure B and lyophilization. **^{1}H NMR (400 MHz, D_2O)** δ 7.64 (s, 1H), 6.93 (s, 1H), 3.68 – 3.57 (m, 1H), 3.17 – 3.01 (m, 2H), 2.66 (d, *J* = 140.7 Hz, 3H). **MS** *m/z* calcd. for $C_7{}^{13}CH_{10}D_3N_3O_2$ $[M+H^+]$: 188.12, found: 188.10.

Synthesis of N^{α},N^{α}-di(methyl-^{13}C-^{2}H)-*L*-histidine (64)



Using procedure Xa and starting with *L*-histidine (50 mg, 0.32 mmol, 1 eq), ^{13}C-formaldehyde (20 %, 222 µL, 1.61 mmol, 5 eq) and $NaCNBD_3$ (61 mg, 0.97 mmol, 3 eq) in D_2O (1 ml), this product was obtained as white solid (55 mg, 0.29 mmol, 91 %) after purification according to procedure X and lyophilization. **^{1}H NMR (400 MHz, D_2O)** δ 8.67 (d, *J* = 1.4 Hz, 1H), 7.43 (d, *J* = 1.3 Hz, 1H), 4.16 (tt, *J* = 6.6, 2.5 Hz, 1H), 3.57 – 3.47 (m, 1H), 3.48 – 3.35 (m, 1H), 2.98 (ddnb-t, $^{1}J_{CH}$ = 144.2 Hz, $^{3}J_{CH}$ = 4.0 Hz, $^{2}J_{HD}$ = 5.0 Hz, 4H) **HRMS** *m/z* calcd. for $C_6{}^{13}C_2H_{12}{}^{2}H_2N_3O_2$ $[M+H]^+$ 188.1273, found 188.1271.

2.6.3 Protein Production

$EgtD_{wt}$ and $EgtD_{E282A,M252V}$ were produced and kindly provided by Laëtitia Misson. All enzymes were expressed in *E. coli* BL21 cells and purified using Ni^{2+} NTA agarose beads (Qiagen) following standard procedures.[99]

Sequence of $EgtD_{wt}$:

GHHHHHHAENLYFQGHMALSLANYLAADSAAEALRRDVRAGLTATQKSLPPKWFYDAVGSDLFDQITR
LPEYYPTRTEAQILRTRSAEIISAAGADTLVELGSGTSEKTRMLLDAMRDAELLRRFIPFDVDAGVLR
SAGAAIGAEYPGIEIDAVCGDFEEHLGKIPHVGRRLVVFLGSTIGNLTPAPRAEFLSTLADTLQPGDS
LLLGTDLVKDTGRLVRAYDDAAGVTAAFNRNVLAVVNRELSADFDLDAFEHVAKWNSDEERIEMWLRA
RTAQHVRVAALDLEVDFAAGEEMLTEVSCKFRPENVVAELAEAGLRQTHWWTDPAGDFGLSLAVR

m/z (EgtD WT): calc.: 36989.6 Da, measured: 36990.0 Da

Ɛ280(EgtD) = 36440 $M^{-1}cm^{-1}$

Sequence of $EgtD_{E282A,M252V}$:

GHHHHHHAENLYFQGHMALSLANYLAADSAAEALRRDVRAGLTATQKSLPPKWFYDAVGSDLFDQITR
LPEYYPTRTEAQILRTRSAEIISAAGADTLVELGSGTSEKTRMLLDAMRDAELLRRFIPFDVDAGVLR
SAGAAIGAEYPGIEIDAVCGDFEEHLGKIPHVGRRLVVFLGSTIGNLTPAPRAEFLSTLADTLQPGDS
LLLGTDLVKDTGRLVRAYDDAAGVTAAFNRNVLAVVNRELSADFDLDAFEHVAKWNSDEERIEVWLRA
RTAQHVRVAALDLEVDFAAGEEMLTAVSCKFRPENVVAELAEAGLRQTHWWTDPAGDFGLSLAVR

m/z ($EgtD_{E282A,M252V}$): calc.: 36899.5 Da, measured: 36899.1 Da

$\varepsilon 280(EgtD_{E282A,M252V}) = 36440\ M^{-1}cm^{-1}$

MetK was produced and kindly provided by Alice Harnacke. The enzyme was expressed in *E. coli* BL21 cells and purified using Ni^{2+} NTA agarose beads (Qiagen) following standard procedures.[133]

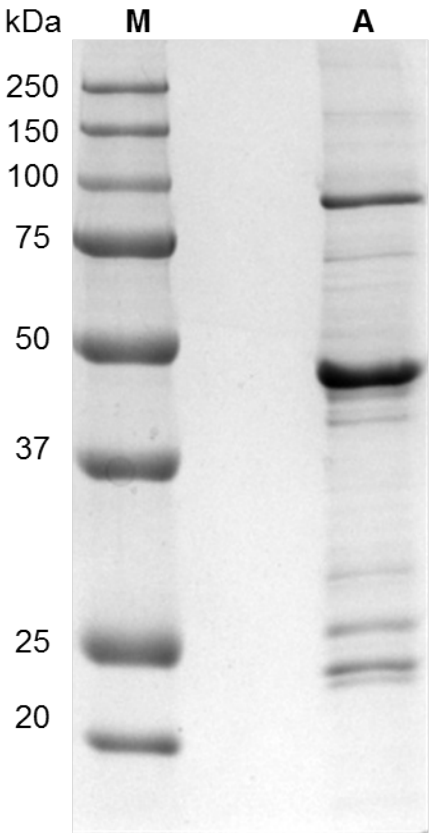


Figure 54 SDS-PAGE picture of MetK after Ni-NTA affinity chromatography **M** - Molecular weight marker; **A** – MetK.

Sequence of MetK:

MGSSHHHHHHSSGLVPRGSHMAKHLFTSESVSEGHPDKIADQISDAVLDAILEQDPKARVACETYVKT
GMVLVGGEITTSAWVDIEEITRNTVREIGYVHSDMGFDANSCAVLSAIGKQSPDINQGVDRADPLEQG

AGDQGLMFGYATNETDVLMPAPITYAHRLVQRQAEVRKNGTLPWLRPDAKSQVTFQYDDGKIVGIDAV
VLSTQHSEEIDQKSLQEAVMEEIIKPILPAEWLTSATKFFINPTGRFVIGGPMGDCGLTGRKIIVDTY
GGMARHGGGAFSGKDPSKVDRSAAYAARYVAKNIVAAGLADRCEIQVSYAIGVAEPTSIMVETFGTEK
VPSEQLTLLVREFFDLRPYGLIQMLDLLHPIYKETAAYGHFGREHFPWEKTDKAQLLRDAAGLK

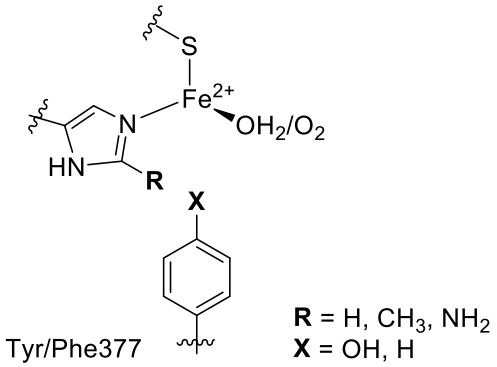
m/z (MetK): calc.: 44114.9 Da, measured: 44291.7 Da (The mass difference of 177 Da corresponds to N^{α}-6-phosphogluconoylation of the His-tag)[134]

Ɛ280 (MetK) = 39880 $M^{-1}cm^{-1}$

3 Probing the Bifurcation Mechanism in the Non-Heme Iron Enzyme EgtB

Abstract

The mycobacterial sulfoxide synthase EgtB belongs to the family of mononuclear non-heme iron (NHI) oxygenases and catalyzes C-S bond formation in the second step of ergothioneine biosynthesis between hercynine (TMH) and γGC. Introduction of a single point mutation on the proposed catalytic base (Y377F) effectively converts EgtB from a sulfoxide synthase to a thiol dioxygenase, consuming γGC and oxygen at the same rate as the wild type reaction. Astoundingly, this efficiency rivals that of naturally occurring cysteine dioxygenases (CDO), which share no evolutionary background with sulfoxide synthases. The precise role of Tyr377 is currently subject to debate. Specifically, it is not clear at which point in the catalytic cycle it helps to direct the reactivity towards the sulfoxide synthase pathway in the wild type reaction. Herein we show that substitution of TMH by 2-amino N^{α}-dimethyl-*L*-histidine (2A-DMH) effectively suppresses dioxygenase activity in both the wild type and the mutant variant of EgtB. X-Ray crystal structures of 2A-DMH in complex with both variants of EgtB demonstrate that the added amino-function provides a functional substitute for the phenol group of Tyr377. In combination with kinetic data from this and previous studies these findings provide strong evidence that active site acidity is key to determining the product specificity of this remarkable enzyme.



3.1 Introduction

3.1.1 Non-Heme Iron Enzymes

Non-heme iron (NHI) enzymes constitute an important class of biocatalysts that capitalize on the power of molecular oxygen to catalyze four-electron oxidation of their substrates.[135] While some rely on cofactors, such as α-ketoglutarate, others have no need for any apparent cosubstrates.[136] One important subset of NHI enzymes acts on substrates containing a thiol group, which is coordinated to the iron center during catalysis. This group include isopenicillin N synthase, cysteine dioxygenase (**A** in Figure 55) and the sulfoxide synthases involved in the biosynthesis of the thiohistidine antioxidants ovothiol and ergothioneine, OvoA and EgtB respectively (**B** and **C** in Figure 55).[17,137,138] While IPNS has been discussed extensively in the introduction, the latter two classes of enzymes will be examined in greater detail in the following.

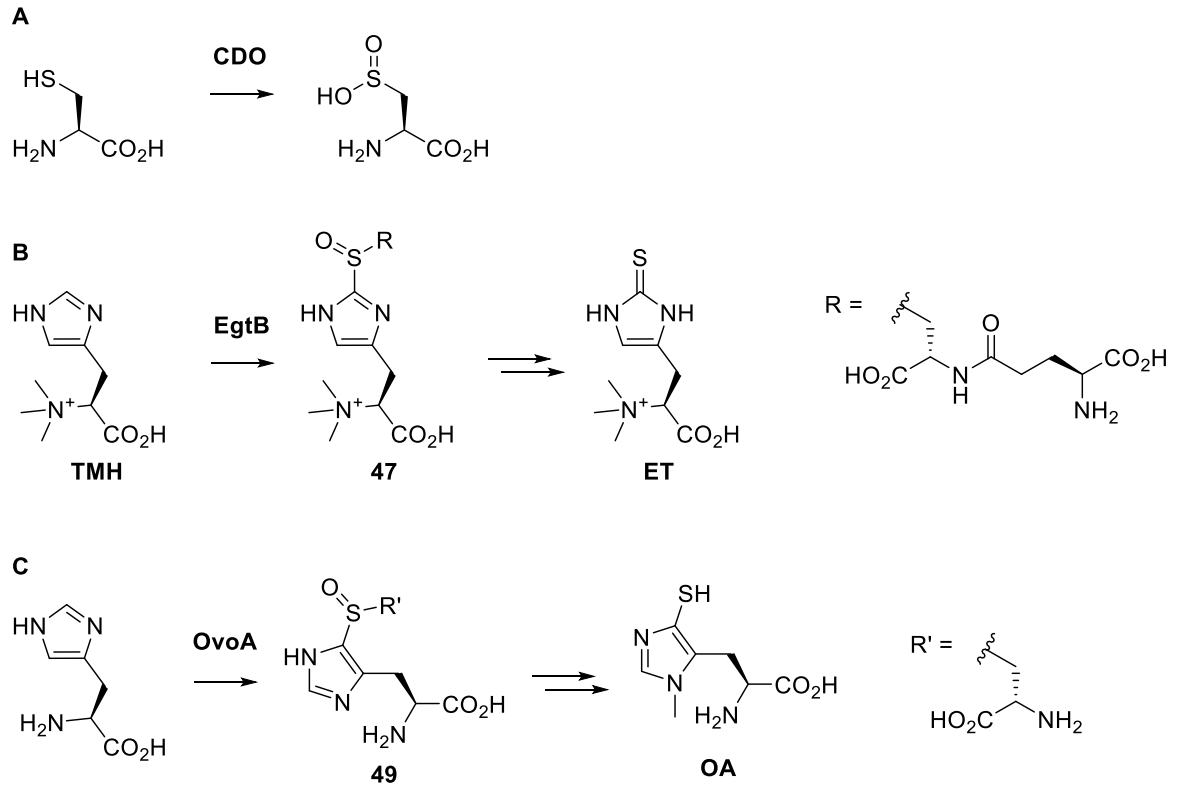


Figure 55 Substrates and products of the three non-heme iron enzymes CDO, EgtB and OvoA.

3.1.2 Cysteine Dioxygenase

Oxidation of cysteine to cysteine sulfinic acid (CSA) is the first step in the biochemical pathway used to oxidize cysteine to either taurine or sulfate as shown in Figure 56.[137] In the first pathway, desulfurization of cysteine produces inorganic sulfate and pyruvate, which is metabolized in the citric acid cycle. The initial dioxygenation is carried out by cysteine dioxygenase (CDO), an enzyme which has stirred much interest due to the association between elevated cysteine levels and Parkinson's as well as Alzheimer's disease among other pathological conditions.[139] As mentioned above, this enzyme is an NHI oxygenase with iron coordinated in a 3-His facial triad, which incorporates both atoms of oxygen into its product.[140] An additional, interesting structural feature of eukaryotic CDO is the presence of a tyrosine-cysteine crosslink, analogous to that found in galactose oxidase.[141]

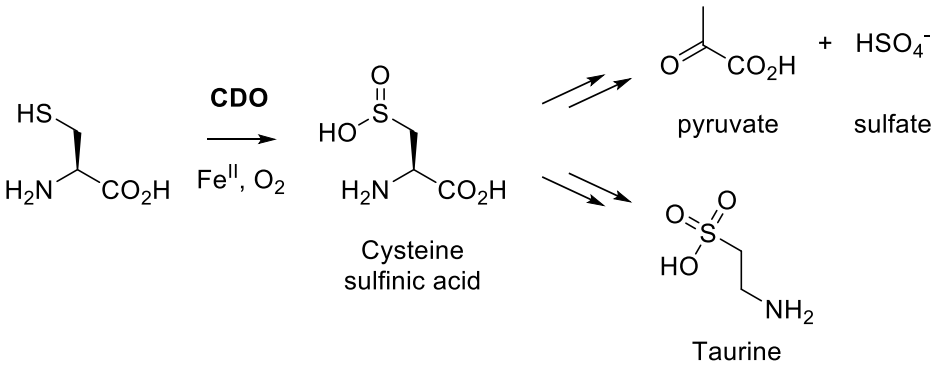


Figure 56 Pathways for cysteine oxidation leading either to sulfate via 3-sulfinyl pyruvate and pyruvate/sulfite or to taurine via hypotaurine.[137] CSD: Cysteine sulfonate dioxygenase.

In the proposed mechanism shown in Figure 57 cysteine binds to the iron center first (**A**), followed by oxygen (**B**), a binding order consistent with findings from EPR studies with nitrous oxide and cyanide used as mimics for oxygen.[142,143] The radical cation resonance form of the thiolate (**C**) is then attacked by the superoxo-species forming a four-membered ring between the sulfur, iron and both oxygen atoms (**D**). Although this intermediate has not been directly detected thus far, DFT calculations support its existence.[144] Heterogeneous cleavage of the O-O bond results in an Fe^{IV}-oxo species (**E**), which provides the second oxygen to the oxidized sulfur (**F**).

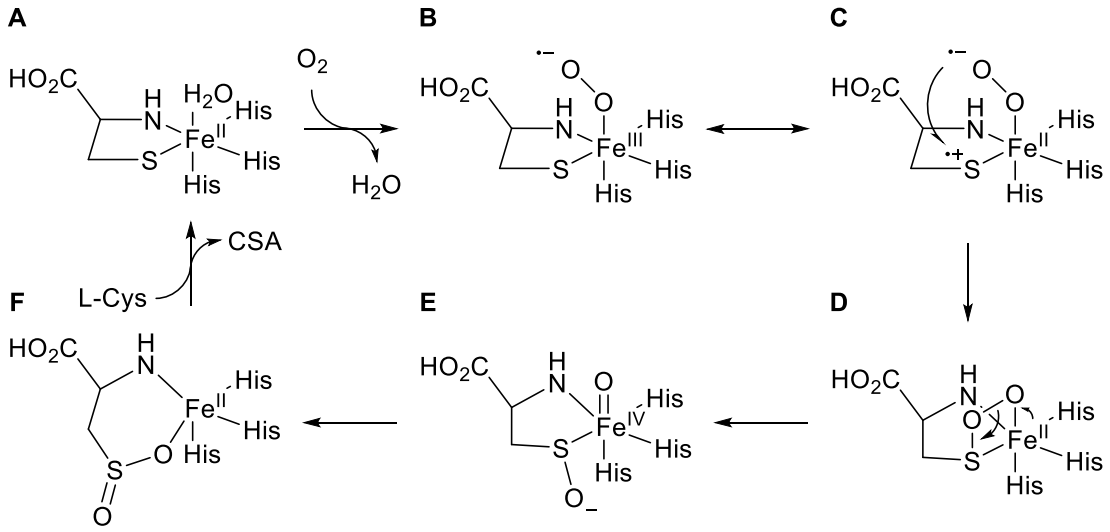


Figure 57 Proposed catalytic mechanism for the dioxygenation of cysteine by CDO.[140]

3.1.3 Sulfoxide Synthases OvoA and EgtB

As described in more detail in **Section 1.4.1**, ergothioneine and ovothiol are both thiohistidines with antioxidant properties.[138] The key step in both their biosynthetic pathways is the formation of a C-S bond between a histidinyl and a cysteinyl moiety. The enzymes catalyzing this step are OvoA for ovothiol and EgtB for ergothioneine and both are Fe^{II}- and O_2-dependent sulfoxide synthases.

Kinetic studies by Mashabela *et al.*[145] and Song *et al*.[104] showed that OvoA displays remarkable substrate promiscuity. Interestingly, when OvoA is supplied with N^{α}-mono-, di- and trimethyl histidine, C-S bond formation on the C-2 instead of the C-5 position of the imidazole ring becomes increasingly dominant. This in turn is essentially the same reaction that is catalyzed by EgtB. Although no X-ray crystal structure has been solved as of yet for OvoA, kinetic studies along with the obvious similarities to EgtB allow for the assumption that the two catalytic mechanisms follow the same pathway.

The X-ray crystal structure for EgtB from *Mycobacterium thermoresistibile* has been solved in its apo form and with TMH, binding iron as well as in complex with manganese and the substrates γGC and DMH.[146] The iron center was revealed to be coordinated by a 3-His facial triad, bearing strong resemblance to the iron coordination sphere of CDO, even though the two enzymes are not evolutionarily related (see also Figure 79 in **Section 3.4**). As is the case in CDO, the three other positions of the hexagonal coordination sphere of the iron center are occupied by the substrates for catalysis. In contrast to CDO however, where the iron coordinates the thiolate and the α-amine of cysteine along with one water molecule, in EgtB

two positions are occupied by N^{τ} of TMH and the thiolate of γGC along with one water. In each case the water molecule is thought to serve as a placeholder for oxygen, which is bound upon initiation of catalysis. As is the case for OvoA, in the course of said catalysis three main events take place: C-S bond formation, C-H bond cleavage and oxygen atom transfer from activated oxygen to sulfur. Although different opinions on the sequence of these steps exist, there is a consensus that an Fe^{III}-superoxo species is involved in the catalytic cycle.[41,146–148] There is however some considerable disagreement concerning the reactivity of the superoxo species during catalysis.[148]

One proposed mechanism, based on structural and kinetic analysis done in our laboratory is shown in Figure 58. After binding of TMH (**A**) followed by γGC and oxygen, an Fe^{III}-superoxo species is formed (**B**). A thiyl radical is generated after donating one electron to oxygen, generating a peroxide anion which is stabilized after protonation by Tyr377 (**C**). The thiyl radical then attacks the imidazole ring (**D**) in a manner resembling a thiol-ene type reaction and forming an iminyl radical.[145] As of yet it remains unclear at which point the distal oxygen atom of the peroxide is transferred to the sulfur (shown here in step **E**). It should be noted that the current data is consistent with both an early[148] and late[146,149] peroxidation step. Lastly, deprotonation of the C-2 hydrogen by the phenolate of Tyr377 (**F**) yields the final product (**G**).

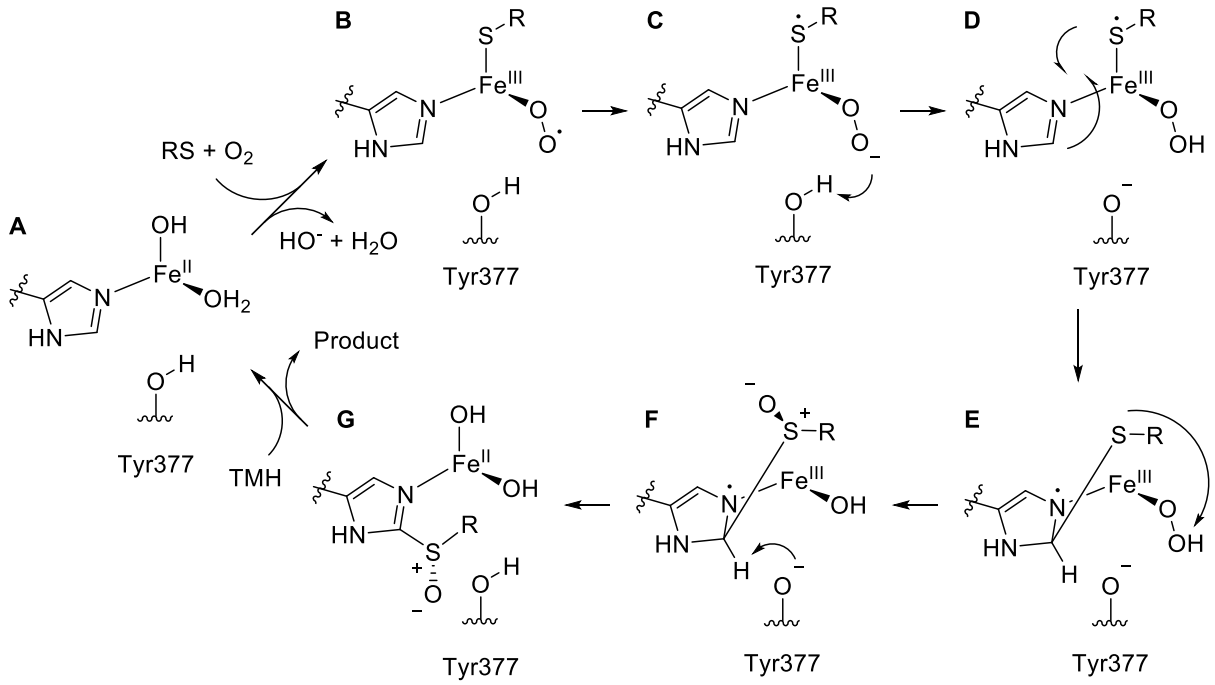


Figure 58 Proposed mechanism for the sulfoxidation of TMH by EgtB.[146]

In order to examine the role of Tyr377, a variant carrying the single point mutation Y377F was produced. This mutant still consumed γGC, but only generated traces of the usual sulfoxide product, and instead catalyzed the dioxygenation of γGC as the main product. By removal of

this crucial active site hydroxyl group EgtB could effectively be converted into a CDO with a kinetic efficiency that rivals naturally occurring CDOs.[150] These findings are consistent with the idea that the role of Tyr377 is to protonate the Fe^{III}-superoxo species and prevent early sulfoxidation of γGC by EgtB. This however is in stark contrast to DFT-calculations by Wei *et al.* who propose that sulfoxidation in fact occurs before C-S bond formation.[148] As shown in Figure 59 the authors argue that the Fe^{III}-superoxo species reacts to sulfoxidized γGC, generating an Fe^{IV}-oxo ferryl species which in turn is protonated by Tyr377 and induces C-S bond formation. The following C-H bond cleavage step, where Tyr377 abstracts the C-2 hydrogen of the imidazoyl ring was calculated to be the rate limiting step of the reaction. Although the calculated activation energy is in agreement with data from published kinetic studies, some difficulties with this explanation arise upon closer inspection.

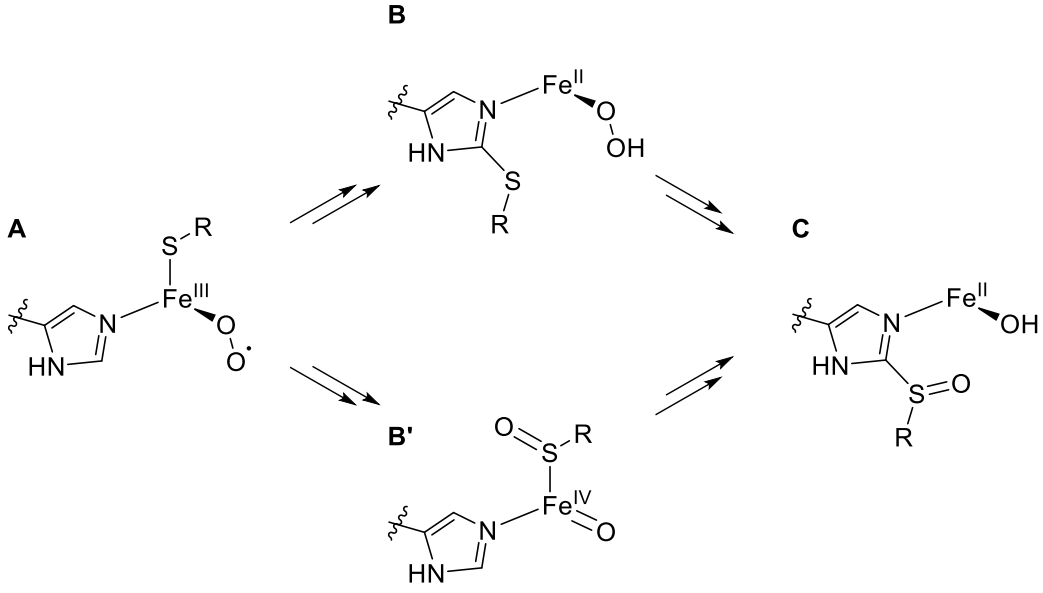


Figure 59 Alternative mechanisms for EgtB where C-S bond formation either takes place before (**B**) or after (**B'**) sulfoxidation.[148]

More recent experiments in our laboratory have shown that C-H bond cleavage is not the rate limiting step, as no KIE was observed using C-2 deuterated TMH.[149] It should be noted that this is also not the rate limiting step in the synthesis of ovothiol by OvoA, where the analogous experiment with C-5 deuterated histidine did not yield an observable KIE.[41] Furthermore, when using C-2 deuterium-labelled substrate on $EgtB_{Y377F}$, no redistribution of products was observed.[151] If indeed this step was isotope-sensitive one would expect that the enzyme would generate more γGC dioxide relative to sulfoxide. In order to illustrate this argument, it is necessary to consider the proposed energy profile of the reaction as calculated by Wei *et al.*[148] According to this proposal, with C-H bond cleavage as the rate limiting step, the point of bifurcation between the sulfoxide synthase and cysteine dioxygenation pathway must lie somewhere before this step. If hydrogen is replaced by deuterium, the energy barrier for C-H

bond cleavage would be raised, even if no KIE is observable. In this case, the dioxygenation pathway would have to be slightly favored over the same pathway with a non-deuterated substrate and thus lead to a redistribution of products. This however is not supported by our kinetic data.[151]

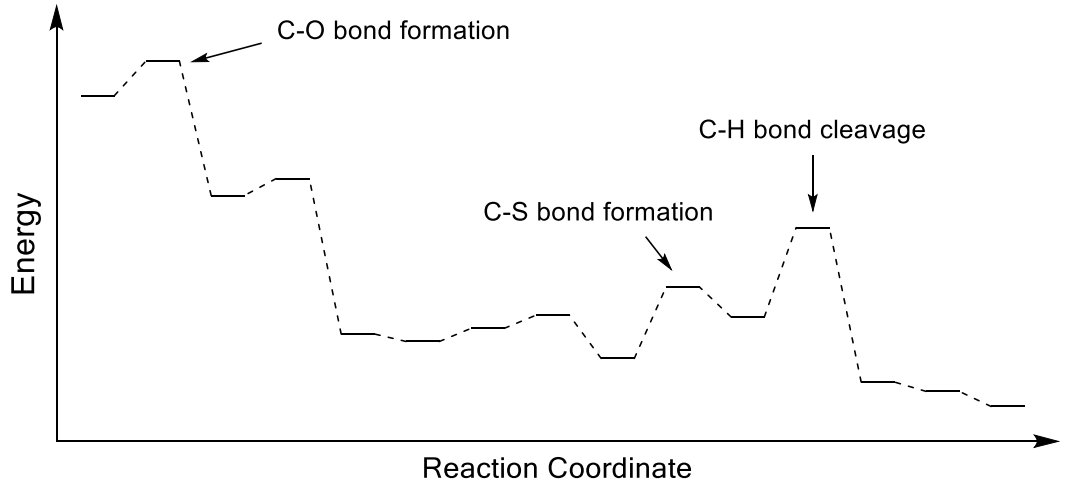


Figure 60 Qualitative representation of the energy profile calculated by Wei *et al.*[148] showing the three key steps of C-O and C-S bond formation, as well as the proposed rate limiting step of C-H bond cleavage. The full mechanism will be discussed in detail in **Section 3.4.3**.

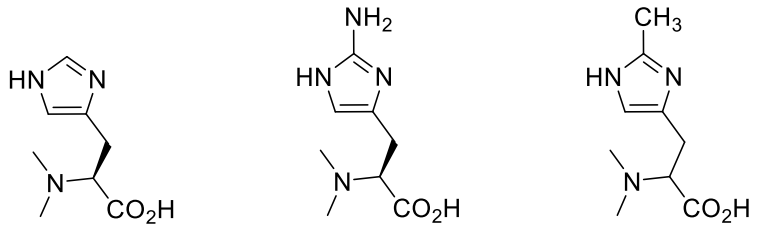
3.2 Aim of this Chapter

One of the main controversies in the catalytic mechanism of EgtB is the role of Tyr377, which functions as a catalytic acid/base. While it seems clear that it promotes sulfoxidation it is conceivable that it also actively suppresses dioxygenation of γGC. The question hence arises of how to disentangle sulfoxide synthase and γGC-dioxygenase activity. Since both the wild type and the Y377F variant of EgtB are hardly active as dioxygenases in the absence of TMH, an alternative co-ligand or activator was necessary. We chose to synthesize TMH-analogues with a substitution on the C-2 position, which cannot form a sulfoxide product and allow for analysis of the dioxygenation pathway in isolation. Comparison of kinetic and structural data of the enzyme in complex with these probes provided insights into the factors governing the specificity of EgtB.

3.3 Results and Discussion

3.3.1 Design and Synthesis of Histidine Analogues

Since formation of the C-S bond takes place on the C-2 position of the imidazole ring of histidine, a series of histidine analogues was synthesized with a modification in this position. In order to increase their affinities to EgtB, the N^{α}-dimethylated variants were synthesized since the binding affinity for DMH is similar to that of TMH.[146]



68 **69** **70**

Figure 61 Structures of N^{α},N^{α}-dimethyl histidine (DMH, **68**), 2-amino N^{α},N^{α}-dimethyl histidine (2A-DMH, **69**) and 2-methyl N^{α},N^{α}-dimethyl histidine (2M-DMH, **70**)

3.3.2 Synthesis of 2-Amino *L*-Histidine (2A-His)

Synthesis of 2-amino *L*-histidine (**71**) was carried out according to published procedures,[152,153] with a few slight modifications.[154] Starting from appropriately protected histidine (**72**), the desired nitrogen was inserted in the 2-position of the imidazole ring in a diazonium reaction in the form of diazo-4-methylbenzoate. Reduction of the resulting diazonium compound **73** was carried out by hydrogenation at 70 bar using PtO_2 as the catalyst, yielding an unstable intermediate which could be stabilized by Ts-protection of N^{τ}.[154] This approach had the additional advantage that **74** could be purified by flash chromatography, since subsequent deprotection essentially yielded the final product **71** in one step. The pure final product was obtained in a total yield of 8 % over 4 steps.

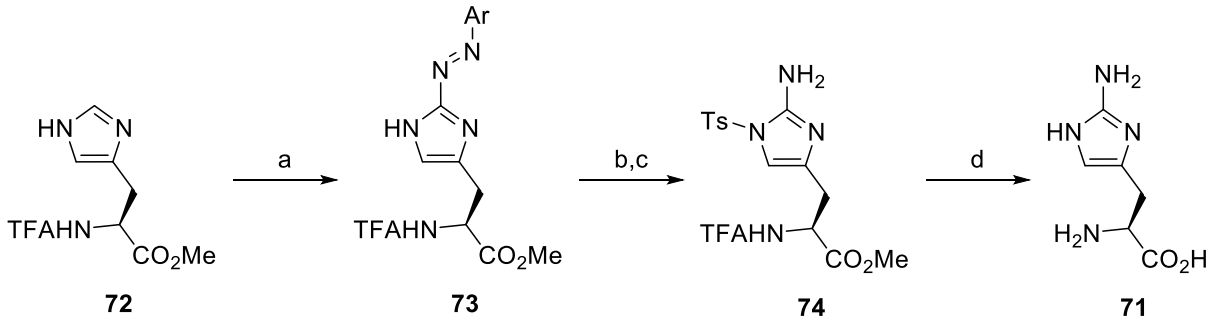


Figure 62 Reagents and conditions for the synthesis of **X**: a) methyl 4-aminobenzoate, $NaNO_2$, HCl, $Na_2B_4O_7$, NaOH, 0 °C, 1 h, 72 %; b) PtO_2, AcOH, H_2, 144 h; c) TsCl, Et_3N, MeCN, 2 h, 12 %; d) 1 M NaOH, 2.5 h, 94 %.

3.3.3 Synthesis of 2-Amino N^{α},N^{α}-Dimethyl *L*-Histidine (2A-DMH)

Originally it was attempted to convert 2-amino histidine to 2-amino N^{α},N^{α}-dimethyl *L*-histidine by reductive amination using formaldehyde and $NaCNBH_3$.[103] This approach however proved to be unsuccessful and a number of unwanted side-reactions took place. We therefore designed an alternative synthesis involving direct insertion of the nitrogen into the desired 2-position of the imidazole ring in the form of an azide, which could subsequently be reduced to the final amine **69**.

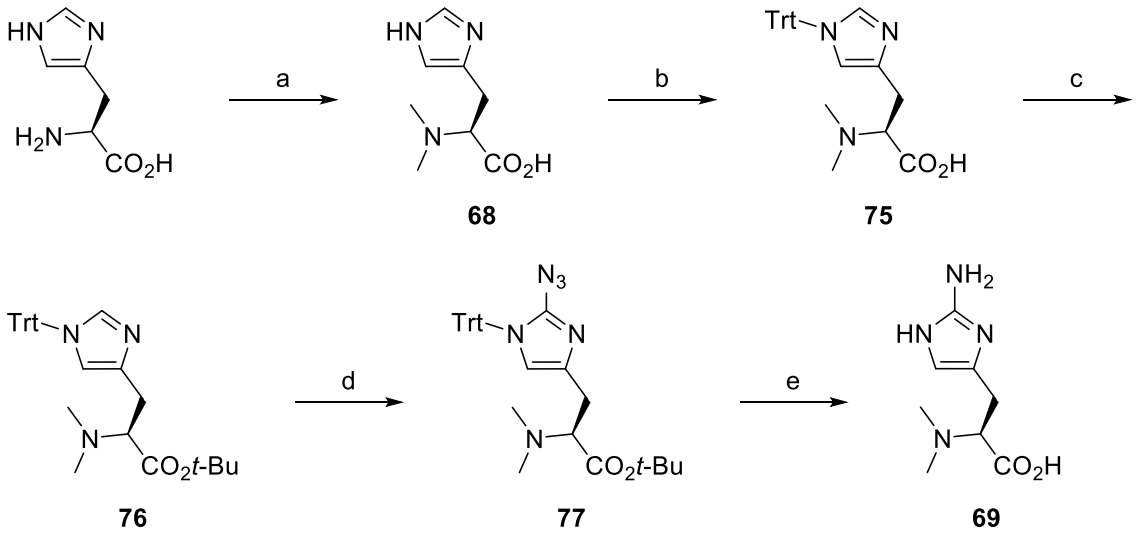


Figure 63 Reagents and conditions for the synthesis of **X**: (a) formaldehyde, $NaCNBH_3$, AcOH, water, 16 h, 99 %; (b) Trt-Cl, Et_3N, DCM, reflux, 18 h, 57 %; (c) *tert*-butyl 2,2,2-trichloroacetimidate, AcOH, DCM, 144 h, 66 %; (d) i) LDA, ii) TsN_3, THF, -78 °C, 1 h, 45 %; (e) i) TFA/TES/DCM, ii) H_2, 10 % Pd/C, water, 6 h, 91 %.

Starting with *L*-histidine, dimethylation of the primary amine was achieved by reductive amination using formaldehyde and $NaCNBH_3$, yielding **68** in nearly quantitative yield. After protection of the N^{π}-position by triphenylmethyl (**75**)[155] and the carboxylate by *t*-Bu (**76**),[156] the

azide functionality could be inserted (**77**). This was achieved by deprotonating and lithiating the 2-position of the imidazole ring using LDA followed by addition of TsN_3, an electrophilic azide donor.[157] Removal of both protecting groups was carried out in one step using TFA/TES/DCM,[158] during which a part of the azide was reduced to the amine (**69** see **Section 3.6.1**). Complete reduction of the azide to the desired amine was then performed with H_2 and Pd/C.[159] The pure final product was obtained in a total yield of 15 % over 5 steps.

3.3.4 Synthesis of 2-Methyl *D/L*-Histidine (2M-His) and 2-Methyl N^{α},N^{α}-Dimethyl *D/L*-Histidine (2M-DMH)

For the histidine derivative containing a methyl group in the 2-position of its imidazole ring a different synthetic approach was chosen, based on work done by Madsen *et al.* and Hirohato *et al.*[106,160] The sole disadvantage to this approach was that the final product was formed as a racemic mixture of the *D*- and *L*-forms as can be seen in Figure 64.

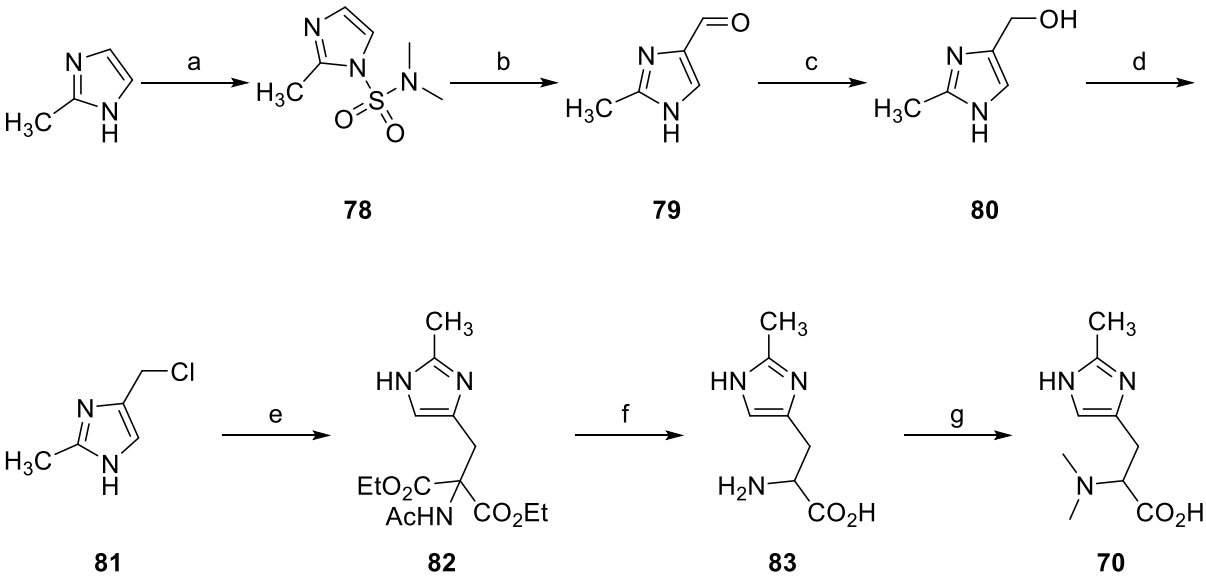


Figure 64 Reagents and conditions for the synthesis of **83** and **70**: (a) DMAS-Cl, Et_3N, 1,2-dichloroethane, 16 h, 98 %; (b) i) n-BuLi, DMF, ii) HCl, THF -78 °C – RT, 3 h, 61 %; (c) $NaBH_4$, EtOH, 4 h, 45 %; (d) i) $SOCl_2$, 16 h, 91 %; (e) NaH, diethyl acetamidomalonate, EtOH, 0 °C – RT, 2 h, 72 %; (f) 6 M HCl, 100 °C, 19 h, 97 %; (g) formaldehyde, $NaCNBH_3$, AcOH, water, 16 h, 69 %.

The first step involved protection of the N^{τ} position of 2-methylimidazole with *N,N*-dimethylsulfamoyl chloride (DMS-Cl) in 1,2-dichloroethane, using Et_3N as base (**78**). The next step consisted of insertion of the aldehyde functionality in the 4-position of the imidazole ring with subsequent deprotection of N^{τ} with concentrated acid at temperatures between -78 °C and room temperature (**79**). Aldehyde insertion was achieved by deprotonating the 4-position with *n*-BuLi, followed by nucleophilic attack on DMF. By acidifying the reaction (pH 1 with conc.

HCl), release of dimethylamine as well as deprotection of N^{τ} was accomplished, leading to the desired product. Reduction of the aldehyde (**79**) to the alcohol (**80**) was carried out using $NaBH_4$ in EtOH. The following chlorination of the alcohol was achieved by reacting **80** neatly in $SOCl_2$ to give **81** in high yield as the HCl salt. For the S_N2 reaction which followed to form **82**, diethyl acetamidomalonate had to be deprotonated by a strong base in order to generate the nucleophile. Although the literature suggested using Na in EtOH, NaH in ethanol was used instead, due to its easier handling. Subsequent deprotection of the amino acid function as well as decarboxylation was achieved by refluxing in 6 M HCl overnight. Non-reflux conditions did not show any conversion after 1 hour. The product 2-methyl *D/L*-histidine **83** was obtained as the di-HCl salt as light brown oil, which crystallized in the fridge overnight. Finally, the N^{α}-dimethylated product **70** was synthesized by reductive amination as described in the previous section. The pure final product was obtained in a total yield of 12 % over 7 steps.

3.3.5 Michaelis-Menten Parameters of *L*- and *D*-DMH

Due to the fact that 2M-DMH was synthesized as a racemic mixture, it was necessary to assess the preference of EgtB for the *L*- over the *D*-form. A good approximation could be given by comparing the kinetic parameters of *L*-DMH with *D*-DMH.

Table 6 Michaelis-Menten parameters of $EgtB_{Th1}$ for *L*- and *D*-DMH. The values for k_{cat} are given in s^{-1}, K_M in µM and k_{cat}/K_M in $M^{-1}s^{-1}$.

	k_{cat} [s^{-1}]	K_M [µM]	k_{cat}/K_M [$M^{-1}s^{-1}$]
***L*-DMH**	2.3 ± 0.1	85 ± 5	26'600
***D*-DMH**	0.9 ± 0.1	1'300 ± 150	700

Reaction conditions were: (12.5-1000) µM *L*- or *D*-DMH respectively, 1 mM γGC, 0.06 µM $EgtB_{Th1}$, 2 mM TCEP, 0.4 µM $FeSO_4$, 2 mM ascorbate, 100 mM NaCl in 100 mM phosphate pH 8.

The binding for the *D*-form is hence almost 15-fold lower than for the *L*-form, meaning that in a racemic mixture roughly 93 % of the bound substrate is in the *L*-form. Extrapolating these findings for the racemic 2M-DMH, we will discuss binding affinities and other data as that of the *L*-form only.

3.3.6 2A-DMH and 2M-DMH as Inhibitors of Sulfoxide Synthase Activity in $EgtB_{wt}$ and $EgtB_{Th2}$

With the two modified DMH variants in hand we sought out to test their respective potentials as inhibitors of sulfoxide synthase activity in two variants of EgtB. The first originates from *M. thermoresistibile*, uses γGC as the thiol-substrate and is termed $EgtB_{wt}$. The second variant is from *Candidatus chloracidobacterium thermophilum B*, uses cysteine as the thiol-substrate and is termed $EgtB_{Th2}$. Using an HPLC-based assay to determine the kinetics, conversion of DMH to DMH-sulfoxide was measured in the presence and absence of 1 mM of each respective inhibitor and 0.1 mM DMH. The methyl variant 2M-DMH showed no significant inhibition in either enzyme, meaning that it binds to the active site at least five-fold worse than DMH. On the other hand 2A-DMH showed strong inhibition of product formation by a factor of over 30 in $EgtB_{wt}$, meaning that this variant is bound at least as well as DMH.

Table 7 Rates of sulfoxide synthase activity of $EgtB_{wt}$ and $EgtB_{Th2}$ for 0.1mM DMH with and without 1 mM 2A-DMH or 2M-DMH

Enzyme	$EgtB_{wt}$		$EgtB_{Th2}$	
Substrate	Rate (s^{-1})	Activity (%)	Rate (s^{-1})	Activity (%)
DMH	1.23	100	0.037	100
w/ 2M-DMH	1.19	97	0.035	94
w/ 2A-DMH	0.04	3	0.030	82

Reaction conditions were: 0.1 mM DMH, 1 mM γGC, 0-1 mM 2A-DMH or 2M-DMH, 0.06 µM $EgtB_{wt}$(left) or 1.5 µM $EgtB_{Th2}$(right), 2 mM TCEP, 4 µM $FeSO_4$, 2 mM ascorbate, 100 mM NaCl in 100 mM phosphate pH 8.

These preliminary results prompted us to further investigate the details of sulfoxide inhibition by 2A-DMH in $EgtB_{wt}$.

3.3.7 2A-DMH is a Competitive Inhibitor to DMH for Sulfoxide Synthase Activity in $EgtB_{wt}$

After determining the Michaelis-Menten parameters for DMH the measurements were repeated with a fixed concentration of DMH (0.1 mM) and increasing concentrations of inhibitor (12.5 µM – 1 mM). While the K_M increased substantially upon addition of inhibitor, the values for k_{cat} remain unaffected, indicating competitive inhibition. Plotting 1/V against 1/S (Lineweaver-Burk plot)[161] leads to the typical form one would expect for competitive inhibition, where the linear fits of the inverse substrate concentration (1/S) relative to the inverse velocity (1/V) converge on the Y-axis.

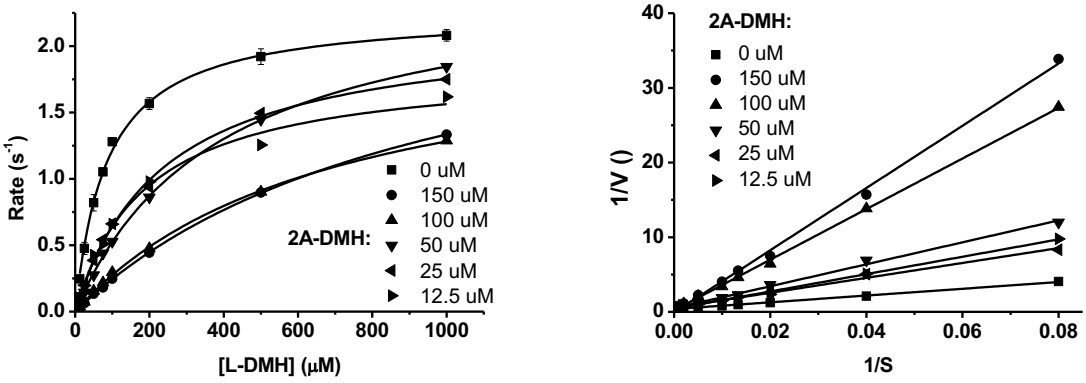


Figure 65 Left: Michaelis-Menten parameters for sulfoxide formation between DMH and γGC, catalyzed by $EgtB_{wt}$ with different concentrations of inhibitor. **Right**: Lineweaver-Burk plot of the same data. Reaction conditions were: 0.1 mM DMH, 1 mM γGC, (12.5-1000) µM 2A-DMH, 0.06 µM $EgtB_{wt}$, 2 mM TCEP, 0.4 µM $FeSO_4$, 2 mM ascorbate, 100 mM NaCl in 100 mM phosphate pH 8.

Using the apparent K_M-values obtained from these measurements and equation (i), a K_I of **14 ± 2 µM** was calculated, which is a value roughly five-fold lower than the K_M of DMH.

$$K_I = \frac{K_M[I]}{K_{M,app} - K_M} \qquad \text{(i)}$$

The same reaction was then repeated at a fixed concentration of the substrate DMH (0.2 mM) and a varying concentration of γGC (12.5 µM – 1 mM). At all concentrations a decrease in K_M as well as k_{cat} could be observed as can be seen in Figure 66.

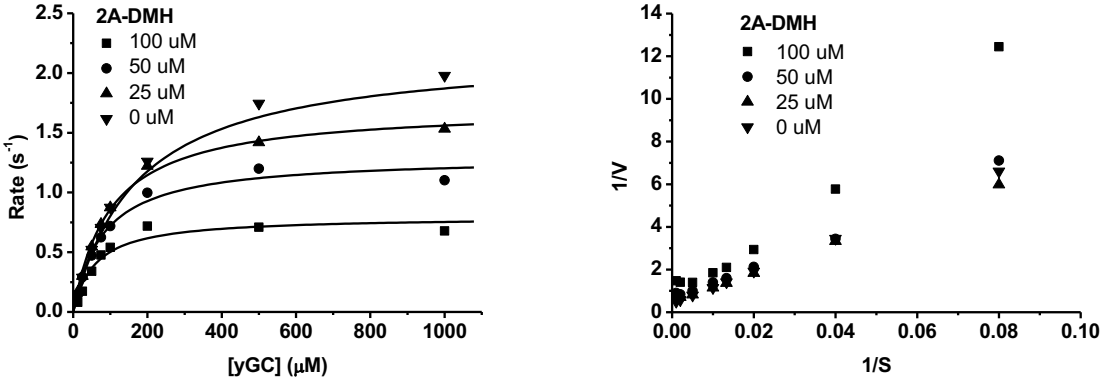


Figure 66 Left: Michaelis-Menten parameters for sulfoxide formation between DMH and γGC, catalyzed by $EgtB_{Th1}$ with different concentrations of inhibitor 2A-DMH. Right: Lineweaver-Burk plot of the same data. Reaction conditions were: 0.2 mM DMH, (12.5-1000) µM γGC, (25-100) µM 2A-DMH, 0.06 µM $EgtB_{wt}$, 2 mM TCEP, 0.4 µM $FeSO_4$, 2 mM ascorbate, 100 mM NaCl in 100 mM phosphate pH 8.

When the data is plotted in a Lineweaver-Burk plot, the data cannot be fitted linearly. Dixon and Webb describe two possible cases for a substrate analog A' that binds to an enzyme E in

the same site as the natural substrate A (Figure 67).[162] Furthermore, the binding of A' does not disturb the binding of the second substrate B. In case of a sequential binding order (i in Figure 67), binding of A' leads to a dead-end complex EA'B when the second substrate B is bound. In this case, plotting 1/[B] versus 1/V leads to a non-linear curve. In an alternative case (ii in Figure 67), where the binding order is non-sequential, a linear correlation between 1/[B] and 1/V is predicted.

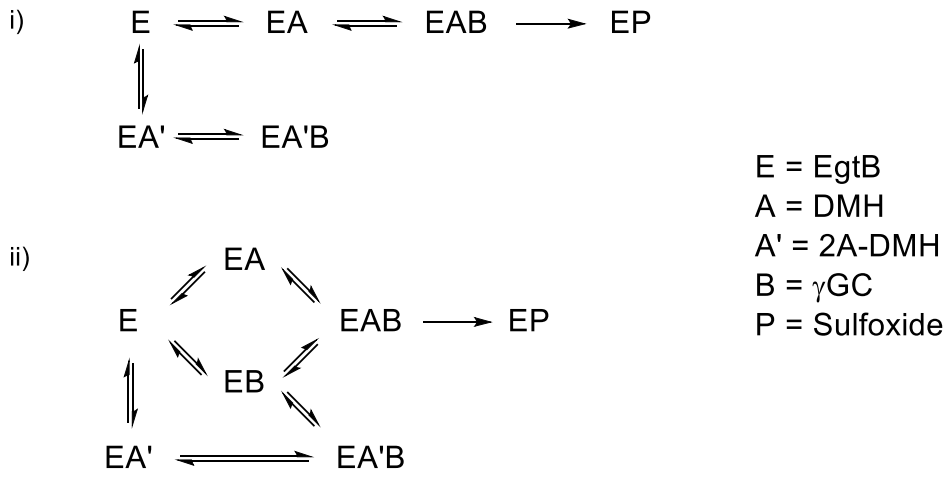


Figure 67 Alternative schemes for sequential and non-sequential binding orders, as proposed by Dixon and Webb.[162]

Thus, the observed non-linear correlation of 1/[B] (where B is γGC) versus 1/V suggests a sequential binding order, which is in agreement with previous kinetic findings on $EgtB_{Th2}$ by Kristina Goncharenko.[150]

3.3.8 Dioxygenase Activity of EgtB$_{Y377F}$ with DMH, 2A-DMH and 2M-DMH

As detailed previously a single point mutation of tyrosine to phenylalanine (Y377F) in the active site of EgtB results in a complete shift in product specificity from the usual sulfoxide to the sulfinic acid of γGC. We sought out to compare the effects of a modification in the 2-position of the imidazole ring on the activator in this reaction. Before determining the dioxygenase activity of the mutant variant EgtB$_{Y377F}$ in complex with 2A-DMH and 2M-DMH, we analyzed the reaction by ^{1}H-NMR to ensure that no unexpected products were formed. When EgtB$_{wt}$ is reacted with DMH, all γGC is incorporated into the sulfoxide product (Figure 68), and the dioxygenase pathway is too slow to form detectable levels of γGC-sulfinic acid. If instead DMH is incubated with EgtB$_{Y377F}$, barely any sulfoxide is formed (<2 %) and the γGC is instead dioxygenated (Figure 69).

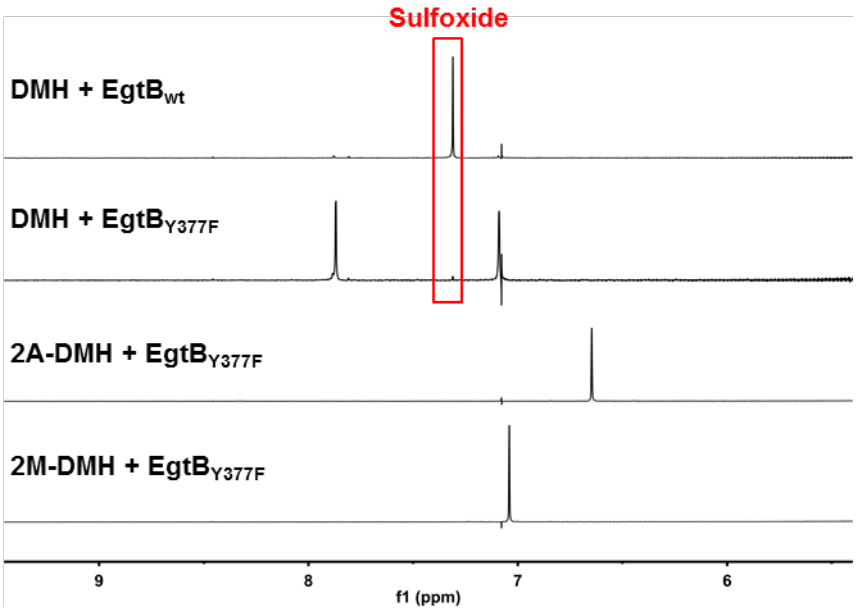


Figure 68 Aromatic region of the reaction mixtures containing EgtB$_{wt}$ with DMH and EgtB$_{Y377F}$ with DMH, 2A-DMH or 2M-DMH (top to bottom). Reaction conditions: 0.3 µM EgtB$_{wt}$ or EgtB$_{Y377F}$, 1 mM DMH, 2A-DMH or 2M-DMH, 0.8 mM γGC, 4 µM $FeSO_4$, 2 mM ascorbate, 2 mM TCEP, 100 mM NaCl in 100 mM phosphate buffer at pH 8 in a total volume of 2000 µL. Reactions were quenched by freezing the samples at -80 °C, followed by lyophilization, resuspension in 500 µL D_2O after which the spectra were recorded.

When EgtB$_{Y377F}$ is incubated with either 2A-DMH or 2M-DMH, no sulfoxidation product is observed judging by the aromatic region of the ^{1}H-NMR spectra (Figure 68). The enzyme is however still active and instead converts γGC to the dioxygenation product (Figure 69), albeit at a reduced rate. In order to precisely quantify this rate, we turned again to a kinetic assay.

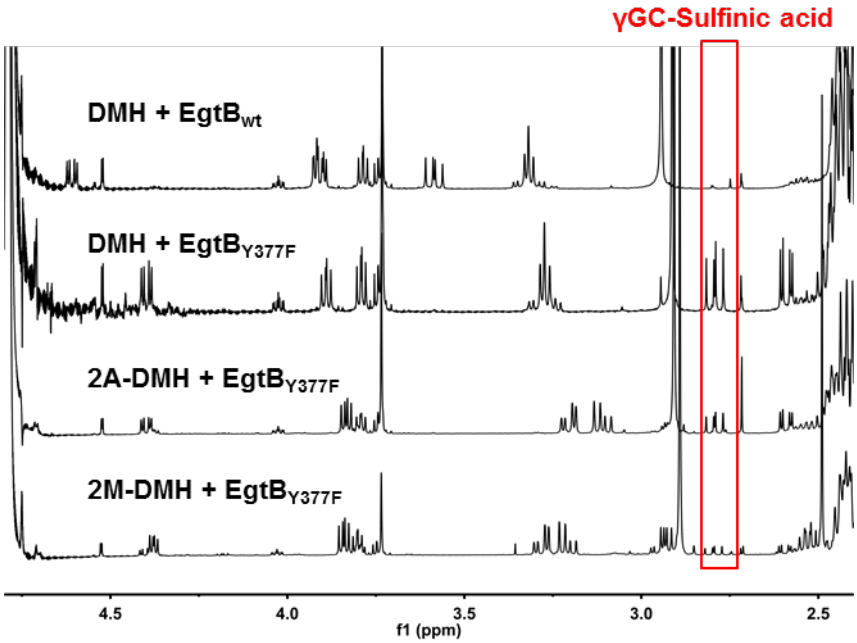


Figure 69 Aliphatic region of the reaction mixtures containing $EgtB_{wt}$ with DMH and $EgtB_{Y377F}$ with DMH, 2A-DMH or 2M-DMH (top to bottom). Reaction conditions: 0.3 µM $EgtB_{wt}$ or $EgtB_{Y377F}$, 1 mM DMH, 2A-DMH or 2M-DMH, 0.8 mM γGC, 4 µM $FeSO_4$, 2 mM ascorbate, 2 mM TCEP, 100 mM NaCl in 100 mM phosphate buffer at pH 8 in a total volume of 2000 µL. Reactions were quenched by freezing the samples at -80 °C, followed by lyophilization, resuspension in 500 µL D_2O after which the spectra were recorded.

Since sulfoxidation takes place to a negligible extent in the $EgtB_{Y377F}$:DMH system and is not a factor in complex with either 2A-DMH or 2M-DMH, the rate at which γGC is consumed can be assumed as the rate of γGC-sulfinic acid formation. Hence, the consumption of γGC was measured instead of the formation of product. Since neither γGC nor γGC-sulfinic acid is detectable at low concentrations, the former was derivatized with a chromophore. In our case we chose to use 4-BMC allowing us to detect the substrate at concentrations as low as 10 µM (Figure 70).

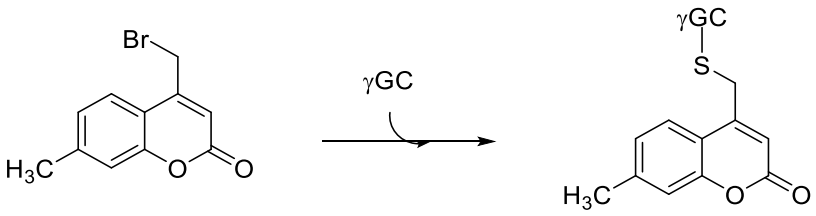


Figure 70 Reaction between γGC and 4-BMC.

Production of sulfinic acid was hence extrapolated from the decrease in substrate as a function of time. In all the described cases DMH, 2A-DMH and 2M-DMH act more like an activator than a substrate in the classical sense since they are not consumed in the course of the reaction. First, the binding affinities of all compounds were determined at a low concentration of γGC

(0.1 mM), where the binding affinity is not influenced by saturating conditions of the second substrate. Since EgtB was shown to follow an ordered substrate binding mode, with γGC as the second substrate bound to the Egtb:TMH complex, the concentration of γGC affects the observed K_M of TMH.[151]

Table 8 Michaelis-Menten parameters of $EgtB_{Y377F}$ for DMH, 2A-DMH and 2M-DMH at [γGC] = 0.1 mM.

	$k_{cat,app}$ [s^{-1}]	$K_{A,app}$ [μM]	k_{cat}/K_M [$M^{-1}s^{-1}$]
DMH	0.4 ± 0.1	20 ± 1	21'000
2A-DMH	0.2 ± 0.1	50 ± 5	4'600
2M-DMH	0.03 ± 0.1	1'500 ± 300	20

Reaction conditions were: (12.5 – 1000) μM DMH, (25 – 2000) μM 2A-DMH, (25 – 5000) μM 2M-DMH , 0.5 μM (DMH and 2A-DMH) or 2.5 μM (2M-DMH) $EgtB_{Y377F}$, 0.1 mM γGC, 4 μM $FeSO_4$, 2 mM ascorbate, 2 mM TCEP, 100 mM NaCl in 100 mM phosphate buffer at pH 8.

Interestingly, under the chosen conditions DMH and 2A-DMH catalyze the reaction at a similar k_{cat} and also bind within the same order of magnitude (herein we refer to the affinity of the activator as K_A). On the other hand the binding affinity and k_{cat} for 2M-DMH are strongly reduced. Binding affinities of γGC were determined next at saturating conditions (at least 3 x $K_{A,app}$) of the respective DMH variant. Under these conditions reliable values for k_{cat} could be calculated since the enzyme was saturated with both substrates.

Table 9 Michaelis-Menten parameters of $EgtB_{Y377F}$ for γGC at saturating concentrations of DMH, 2A-DMH and 2M-DMH.

	$k_{cat,\gamma GC}$ [s^{-1}]	$K_{M,\gamma GC}$ [μM]	k_{cat}/K_M [$M^{-1}s^{-1}$]
DMH	1.4 ± 0.2	185 ± 45	7'400
2A-DMH	0.08 ± 0.01	36 ± 9	2'200
2M-DMH	0.11 ± 0.1	290 ± 45	380

Reaction conditions were: DMH (1 mM), 2A-DMH (1 mM) or 2M-DMH (7.5 mM), 0.5 μM (DMH and 2A-DMH) or 2.5 μM (2M-DMH) $EgtB_{Y377F}$, 12.5 – 500 μM γGC, 4 μM $FeSO_4$, 2 mM ascorbate, 2 mM TCEP, 100 mM NaCl in 100 mM phosphate buffer at pH 8.

Under these conditions the difference becomes more apparent between the native and the unnatural substrates. Specifically, V_{max} of the enzyme with 2A-DMH as co-ligand is roughly the same as for 2M-DMH, which is almost 15-fold lower than for DMH. The binding affinity of the system for γGC seems to increase for 2A-DMH compared to both other ligands. Apparently

the presence of either an amino- or a methyl-group in the 2-position of the imidazole ring similarly disrupt the oxidation of γGC.

Additionally, the reaction was investigated for possible pH-dependence, measuring the rate at fully saturated conditions. While the reaction with DMH is most efficient at pH 8.0-8.5, no significant deviation from this pH-dependence could be observed for either substrate in the range 7.0-9.0.

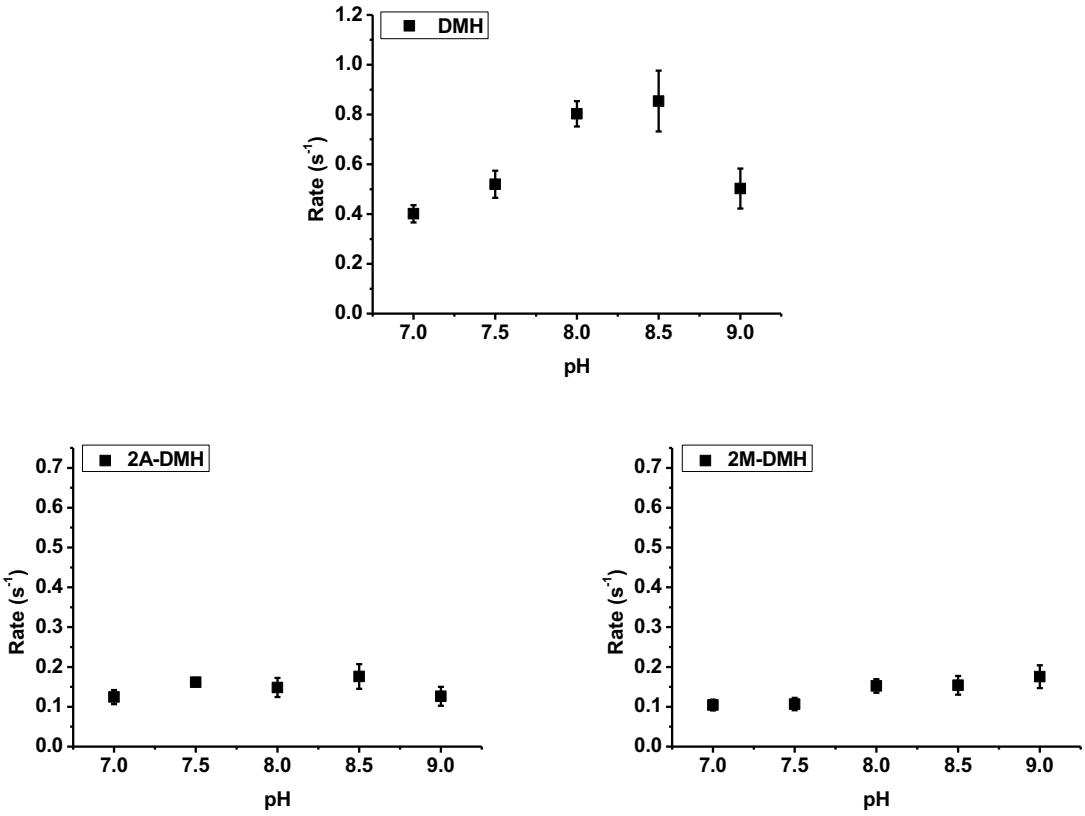


Figure 71 Comparison of the rates of γGC-dioxygenation by $EgtB_{Y377F}$ with DMH (top), 2A-DMH (bottom left) and 2M-DMH (bottom right) in the pH range 7.0-9.0. Reaction conditions: DMH (1 mM), 2A-DMH (1 mM) or 2M-DMH (5 mM), 0.5 µM (DMH) or 1 µM (2A-DMH) or 2.5 µM (2M-DMH) $EgtB_{Th1}$Y377F, 1 mM γGC, 4 µM $FeSO_4$, 2 mM ascorbate, 2 mM TCEP, 100 mM NaCl in 100 mM phosphate buffer (pH 7.0-8.0) or Tris buffer (pH 8.5-9.0).

3.3.9 Dioxygenase Activity of $EgtB_{Th1}$ with 2A-DMH and 2M-DMH

Next we tested whether binding of the unnatural substrates to the wild type enzyme would force the system to proceed along the dioxygenation pathway. By performing analogous experiments as described in **Section 3.3.8**, we concluded that no sulfoxide product was formed in the reaction and that instead the enzyme was producing γGC-sulfinic acid.

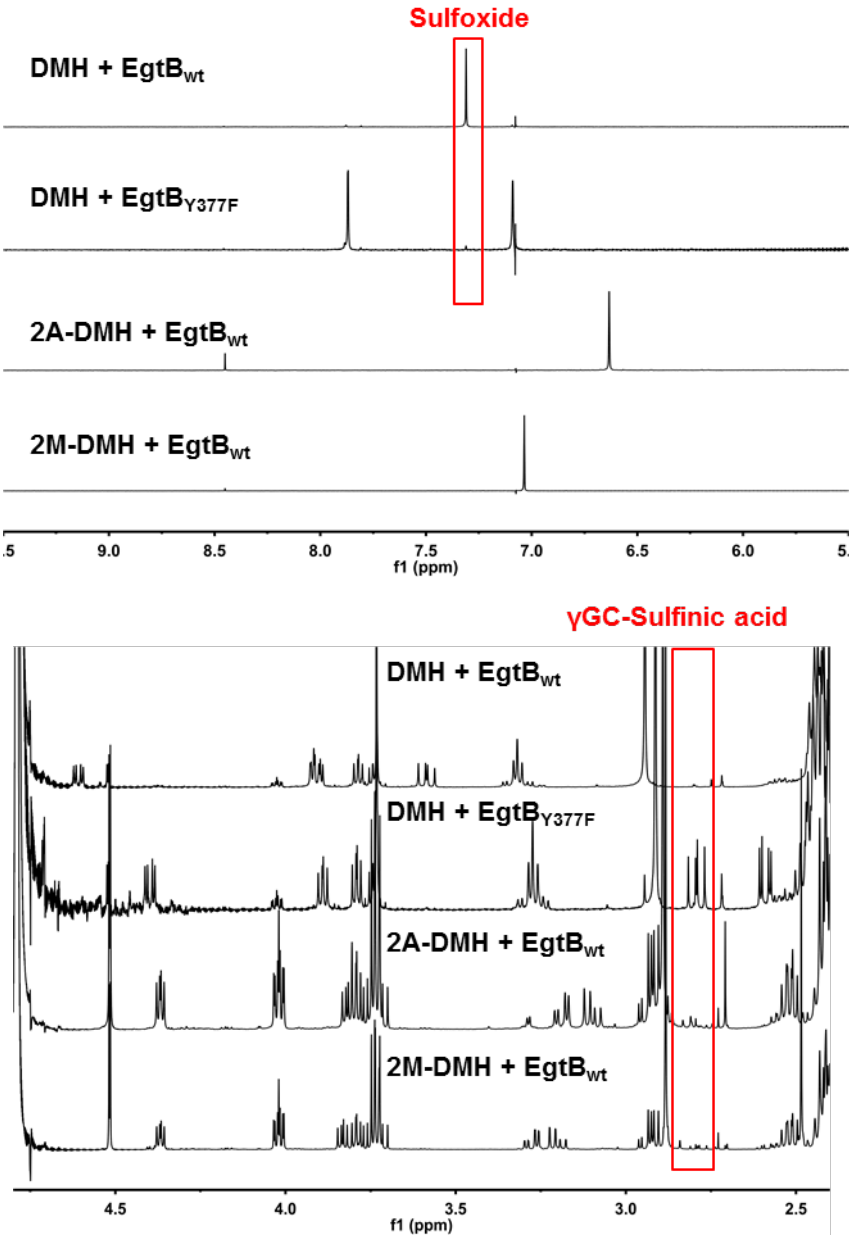


Figure 72 Aromatic (top) and aliphatic region (bottom) of the reaction mixtures containing $EgtB_{wt}$ with DMH and $EgtB_{Y377F}$ as well as $EgtB_{wt}$ with 2A-DMH or 2M-DMH (top to bottom). Reaction conditions: 0.3 µM $EgtB_{wt}$ or $EgtB_{Y377F}$, 1 mM DMH, 2A-DMH or 2M-DMH, 0.8 mM γGC, 4 µM $FeSO_4$, 2 mM ascorbate, 2 mM TCEP, 100 mM NaCl in 100 mM phosphate buffer at pH 8 in a total volume of 2000 µL. Reactions were quenched by freezing the samples at -80 °C, followed by lyophilization, resuspension in 500 µL D_2O after which the spectra were recorded.

Only traces of the dioxygenation product could be observed and the same kinetic assay as described in **Section 3.3.8** was employed in order to monitor consumption of γGC. Again, the measurements to determine K_A were conducted at a concentration of [γGC] = 0.1 mM.

Table 10 Michaelis-Menten parameters of $EgtB_{Th1}$ for 2A-DMH and 2M-DMH at 0.1 mM γGC. The values for k_{cat} are given in s^{-1}, K_M in µM and k_{cat}/K_M in $M^{-1}s^{-1}$.

	$k_{cat,app}$ [s^{-1}]	$K_{A,app}$ [µM]	k_{cat}/K_M [$M^{-1}s^{-1}$]
2A-DMH	0.014 ± 0.001	30 ± 10	400
2M-DMH	0.014 ± 0.002	1'100 ± 300	10

Reaction conditions were: 12.5-500 µM 2A-DMH or 25-1000 µM 2M-DMH, 0.1 mM γGC, 2.4 µM $EgtB_{wt}$, 4 µM $FeSO_4$, 2 mM ascorbate, 2 mM TCEP, 100 mM NaCl in 100 mM phosphate buffer at pH 8.

The measured binding affinity of $EgtB_{wt}$ for 2A-DMH (30 + 10 µM) is in the same range as the K_I of 14 ± 2 µM we had previously determined in our inhibition experiments of the sulfoxide synthase activity of the enzyme. Binding of 2M-DMH is nearly 20 times less favorable, which is also in line with our previous findings of its poor inhibition of sulfoxide synthesis.

In a second series of experiments the same parameters were determined for the dioxygenase activity of $EgtB_{wt}$, first determining the values for 2A-DMH and 2M-DMH at a non-saturating concentration of γGC (0.1 mM). After measuring the parameters for γGC at saturated conditions of 2A-DMH (1 mM) however, it turned out that 0.1 mM was well over the K_M of γGC in this system. The main point of these measurements though was to determine V_{max}, which is another 10-fold lower than the mutant with the same co-ligand.

Table 11 Michaelis-Menten parameters of $EgtB_{Th1}$ for γGC at saturated concentrations of 2A-DMH and 2M-DMH. The values for k_{cat} are given in s^{-1}, K_M in µM and k_{cat}/K_M in $M^{-1}s^{-1}$.

	$k_{cat,\gamma GC}$ [s^{-1}]	$K_{M,\gamma GC}$ [µM]	k_{cat}/K_M [$M^{-1}s^{-1}$]
2A-DMH	0.012 ± 0.001	10 ± 5	1'200
2M-DMH	0.07 ± 0.01	350 ± 50	200

Reaction conditions were: 1 mM 2A-DMH or 5 mM 2M-DMH, 0.1 mM γGC, 2.4 µM $EgtB_{Th1}$, 4 µM $FeSO_4$, 2 mM ascorbate, 2 mM TCEP, 100 mM NaCl in 100 mM phosphate buffer at pH 8.

As can be seen in **Table 11**, γGC possesses a greatly inflated K_M when in complex with 2M-DMH rather than 2A-DMH, which is in agreement with what we had previously measured when comparing the two as inhibitors of the sulfoxide synthesis. Interestingly however, V_{max} is six times higher for the methylated derivative than its amine counterpart. This indicates that the additional protons supplied by the protonated amino-group (the pK_A of 2-amino imidazole is 8.46[163]) perturb the reaction more severely than the added steric bulk of the methyl group.

These effects are seemingly not as relevant in the Y377F mutant, where the two aforementioned compounds display similar V_{max} values once the system is saturated. Again however, the methylated version is bound significantly worse than the amino variant. The overall catalytic efficiency of the system containing 2A-DMH is roughly five to six-fold higher for both the wild type and the mutant. Albeit at this point it should once again be noted that since 2M-DMH was measured as a racemic mixture, the K_M values can effectively be halved since the *L*-form is bound 15-times more efficiently than its enantiomer, as was shown with the example of *L*- versus *D*-DMH in **Section 3.3.5**.

As a further point, it should be noted that no spike in consumption of TCEP or ascorbate, the antioxidants present in all reaction mixtures, was detected in any of the assays. This indicates that, although 2A-DMH and 2M-DMH bind to EgtB and elicit dioxygenase activity, no detectable excess of activated oxygen species is formed by the enzyme. Specifically, no dehydroascorbate could be observed in the ^{1}H-NMR spectra of the reaction mixtures containing any combination of wild type or mutant enzyme and either activator (**Figure 73** and **Figure 74**). If dehydroascorbate was formed during the reaction, one would expect to see the following signals in the displayed range: δ 4.59 (m), 4.28 (dd, *J* = 10.34 Hz, *J* = 5.34 Hz), 4.28 (dd, *J* = 10.34 Hz, *J* = 2.59 Hz).[164]

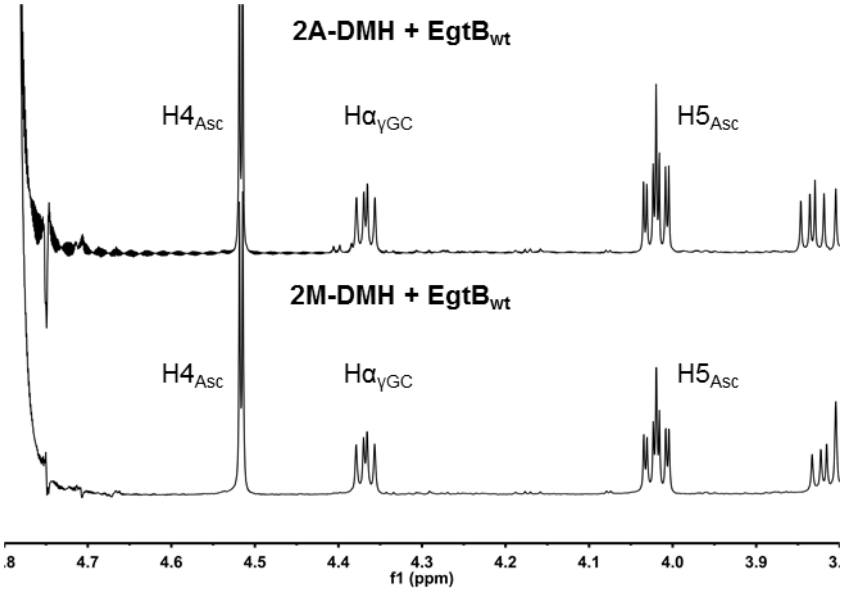


Figure 73 ^{1}H-NMR (500 MHz) spectra of 2M-DMH and 2A-DMH after incubation with EgtB$_{wt}$. Reaction conditions were: 0.3 µM EgtB$_{wt}$, 1 mM 2A-DMH or 2M-DMH, 0.8 mM γGC, 4 µM $FeSO_4$, 2 mM ascorbate, 2 mM TCEP, 100 mM NaCl in 100 mM phosphate buffer at pH 8 in a total volume of 2 mL. Reactions were quenched by freezing the samples at -80 °C, followed by lyophilization, resuspension in 500 µL D_2O after which the spectra were recorded.

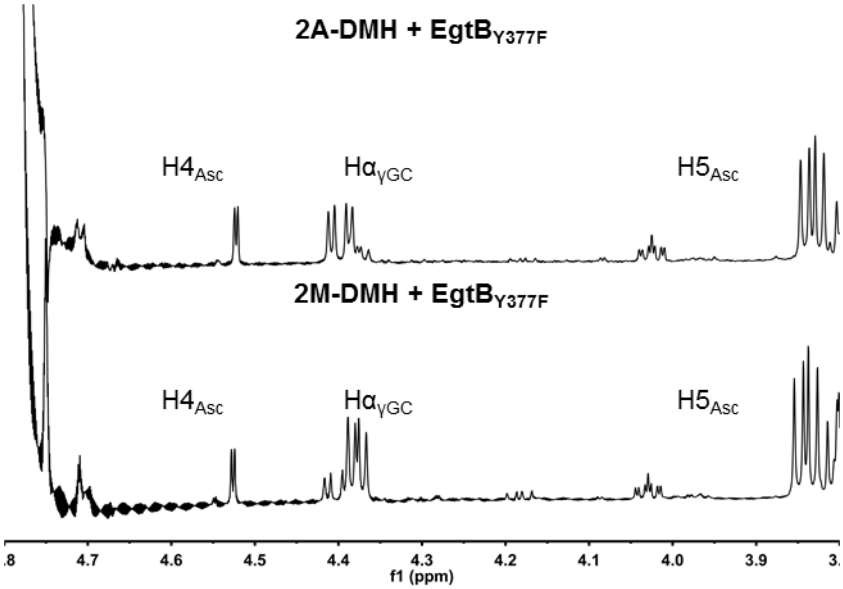


Figure 74 ^{1}H-NMR (500 MHz) spectra of 2M-DMH and 2A-DMH after incubation with EgtB$_{Y377F}$. Reaction conditions were: 0.3 µM EgtB$_{wt}$, 1 mM 2A-DMH or 2M-DMH, 0.8 mM γGC, 4 µM $FeSO_4$, 0.2 mM ascorbate, 2 mM TCEP, 100 mM NaCl in 100 mM phosphate buffer at pH 8 in a total volume of 2 mL. Reactions were quenched by freezing the samples at -80 °C, followed by lyophilization, resuspension in 500 µL D_2O after which the spectra were recorded.

Since TCEP also reacts with 4-BMC, it is possible to monitor TCEP-oxidation as a function of time in the same assay described in **Section 3.3.8**.

Table 12 Consumption of TCEP during the γGC-dioxygenation reaction as observed in a decrease in absorption at 310 nm (mAU/min).

	EgtB$_{wt}$	**EgtB$_{Y377F}$**
DMH	1.6 ± 1.7	
2A-DMH	2.2 ± 2.4	0.4 ± 0.2
2M-DMH	0.3 ± 0.5	1.3 ± 0.6

No significant deviation from the normal consumption of TCEP, could be detected either (see **Table 12**). It must be considered however, that these measurements did not deliver precise, quantitative values, but rather offer a qualitative analysis of this aspect. In order to more accurately quantify a possible increase in the rate of oxygen activation, measurements at higher enzyme concentrations would be required.

3.3.10 Crystal Structures

The kinetic data described thus far suggests that 2A-DMH is a competitive inhibitor towards DMH and binds in the same pocket as the latter. Furthermore, in the $EgtB_{wt}$:2A-DMH complex, dioxygenation activity is suppressed compared to the same enzyme in complex with 2M-DMH. We surmised that 2A-DMH with its acidic protons may engage in hydrogen bonding to the Fe^{II}-superoxo species, lowering the reactivity of this species much like Tyr377 does in the mechanism proposed in Figure 58. To examine the structural basis of this idea, we crystallized the EgtB:2A-DMH complex for both the wild type and the Y377F mutant. Crystal structures of EgtBY377F in complex with DMH or 2M-DMH were solved as well. Each structure was solved as the quaternary complexes EgtB:DMH:γGC with manganese as a substituent for iron. Crystals were grown and their structures solved by Peer Lukat at the Helmholtz Centre for Infection Research in Braunschweig, using conditions described in **Section 3.6.2.7**. All crystal structures were consistent with those reported previously for $EgtB_{wt}$,[146] belonging to the same space group and superimposing with an RMSDs < 0.150 Å (380 atoms). Due to anisotropies of the $EgtB_{Y377F}$:2M-DMH crystal and because of low occupancy, the electron density of 2M-DMH is incomplete. Given the low stability of the EgtB:2M-DMH complex this result is not surprising (**Table 8**).

The basic topology of the active site remains unchanged in all five quaternary complexes, with the metal center (manganese), γGC and the key active site residues (His51, Gln55, His 134, His 138, Tyr380) in identical positions. In each structure the histidine ligand is bound to the metal center by N^{τ}-coordination, while N^{π} forms a water-mediated H-bridge to the backbone carbonyl of Tyr380. In this binding mode the C-2 position of all three histidine ligands point towards residue 377. In $EgtB_{Y377F}$ the C-2 amino group makes van der Waals contact with the phenyl ring of Phe377 (3.3 Å). In $EgtB_{wt}$ the same group is within hydrogen bonding distance to the phenol function of Tyr377 (2.9 Å), dislocating the latter by 0.7 Å from its usual position to avoid steric clash.

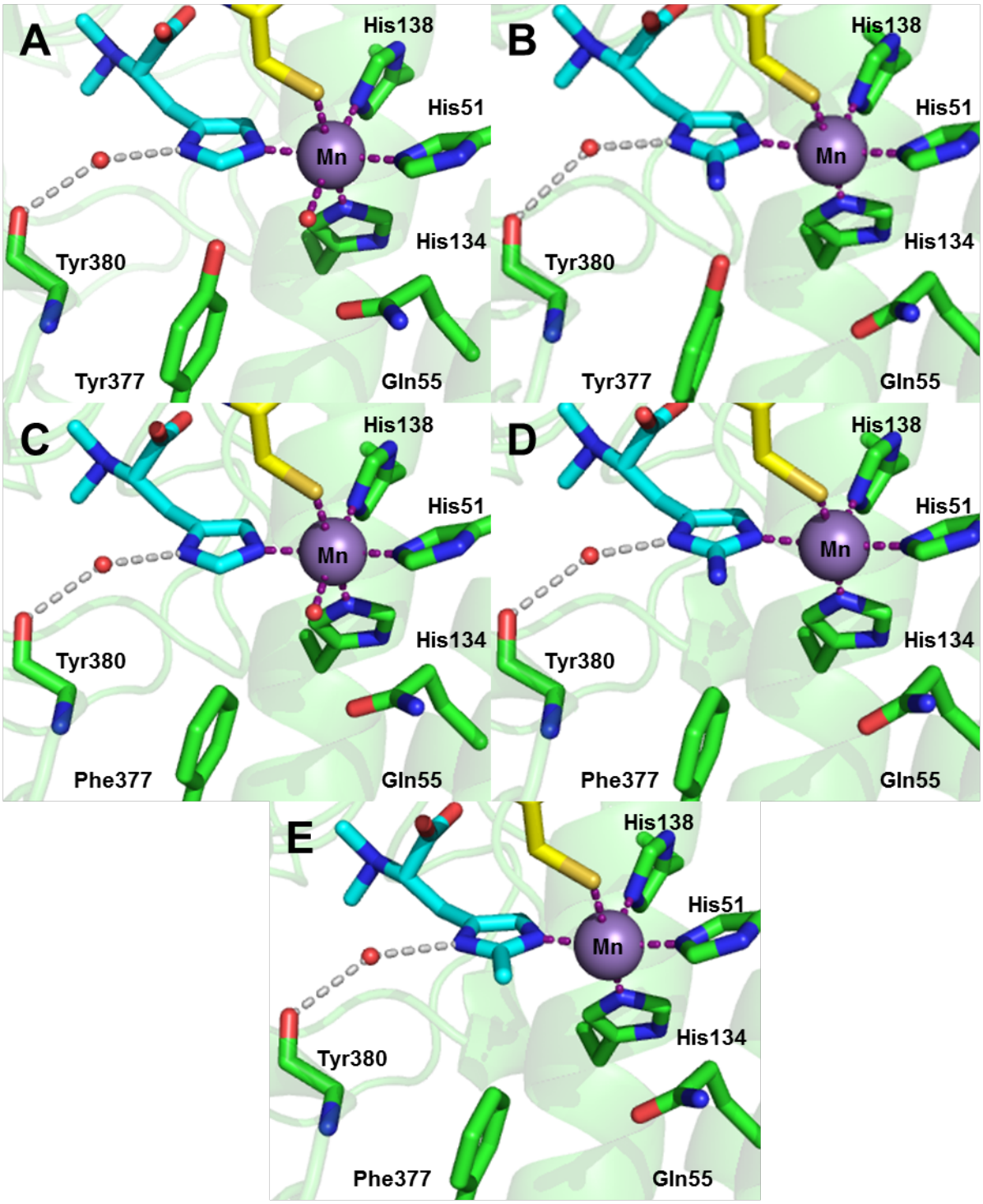


Figure 75 Top: Active site of $EgtB_{wt}$ in complex with γGC and DMH (**A**) or 2A-DMH (**B**). **Middle:** Active site of $EgtB_{Y377F}$ in complex with γGC and DMH (**C**) or 2A-DMH (**D**). Bottom: Active site of $EgtB_{Y377F}$ in complex with γGC and 2M-DMH (**E**). The side chain of Tyr380 is omitted for clarity.

Aquo Ligand

Besides the dislocation of the crucial tyrosine residue, a second major difference is apparent upon binding of 2A-DMH and 2M-DMH: the aquo ligand in the last coordination site of the metal center is not present in either structure. In all other crystal structures of EgtB (PDB: 4X8B, 4X8D, 4X8E) this ligand is present with an occupancy approaching 100 %, while in the two structures containing 2A-DMH and the structure containing 2M-DMH it is found in less than 10 % of the protein chains.[165] This observation is rather puzzling as the N-O distance of 2.7 Å is within the van der Waals cutoff distance of 3.7 Å for an N-H···O hydrogen bond.[166] In addition, the C2-N···O angle of 107° (the O-H···O angle in tetrahedral coordinated water is 109°)[167] would also allow H-bonding to take place with the expected aquo ligand as shown in the model in Figure 76 (left). Furthermore, there is no structural indication that the 2-amino group directly competes for the sixth coordination position or for steric clashes being the reason for its absence. A crystal structure of the binuclear metalloenzyme Human Arginase I in complex with 2-amino histidine can be used as a point of reference (PDB: 3MFW). Herein, the 2-amino group of interest hydrogen bonds with the oxygen anion of a sulfate (2.5 A) and forms an N-O-S angle of 103° as shown in Figure 76 (right).[168]

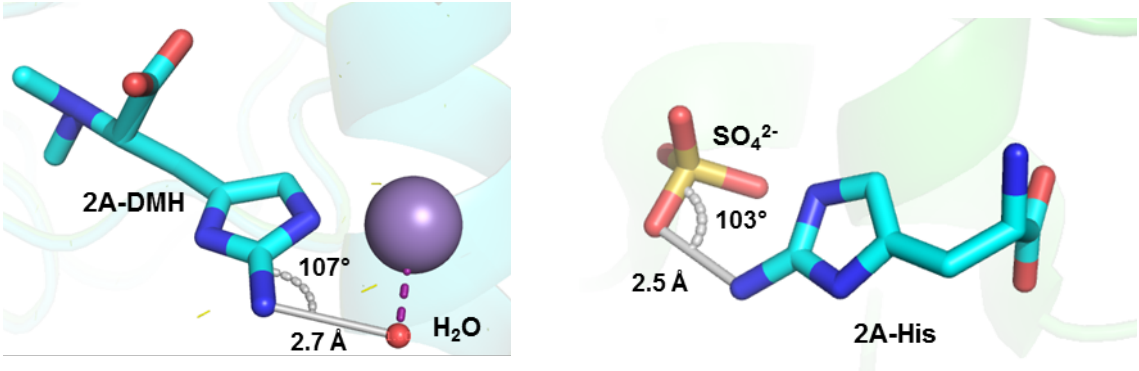


Figure 76 Left: Aligned structures of the complexes $EgtB_{wt}$:2A-DMH:γGC (cyan) and $EgtB_{wt}$:DMH:γGC (green, PDB: 4X8D).[146] The angle and distance between the 2-amino group and the aquo-ligand missing in the first structure are displayed in black. **Right**: Coordination of the 2-amino group of 2-amino *L*-histidine to the oxygen anion of a sulfate molecule in the substrate binding pocket of the binuclear manganese metalloenzyme human arginase I (PDB: 3MFW).[168]

As steric clash does not seem to be responsible for the displacement of the aquo ligand, it seems more plausible that protonation thereof by the 2-amino group of the co-ligand destabilizes coordination to the metal. This theory is in line with our finding that 2A-DMH essentially quenches dioxygenase activity in the wild type enzyme, even though it binds to the active site with similar affinity as the native substrate DMH.

It is also possible that 2A-DMH weakens the affinity of the enzyme for O_2 by stabilizing the metal-bound aquo ligand. In this case however we would expect this position to be more highly

occupied in the crystal structure, which is not the case. Every other structure of EgtB in the absence or presence of either DMH or TMH (PDB: 4X8B, 4X8D, 4X8E) contain a fully occupied aquo ligand. In addition, it has been shown that upon lowering of the pH from 8.0 to 6.0, the kcat of the dioxygenase activity of $EgtB_{Y377F}$ remains basically unchanged ($k_{cat,pH8.0}$ = 1.2 s^{-1}; $k_{cat,pH6.0}$ = 0.9 s^{-1}). Under the same conditions, the sulfoxide synthase activity of the mutant increases by 3.5-fold ($k_{cat,pH8.0}$ = 0.9 x 10^{-4} s^{-1}; $k_{cat,pH6.0}$ = 3.2 x 10^{-3} s^{-1}) while the same activity of the wild type increases by only 1.4-fold ($k_{cat,pH8.0}$ = 1.2 s^{-1}; $k_{cat,pH6.0}$ = 0.85 s^{-1}). If indeed the protonation state of the aquo-ligand was the factor decreasing the affinity for O_2, one would expect to see a considerable drop in activity at lower pH, leading us to conclude that active site acidity does not lead to decreased oxygen-binding.

Hydrogen-Bonding Network

In both crystal structures containing 2A-DMH, the 2-amino group hydrogen bonds to a crystallographic water (**a** in Figure 77) at distances of 3.1 Å ($EgtB_{wt}$; H_2O_160) and 3.0 Å ($EgtB_{Y377F}$; H_2O_112). This water is conserved in all other crystal structures of EgtB, except in the $EgtB_{Y377F}$:2M-DMH complex. In the first structure ($EgtB_{wt}$:2A-DMH, **A** in Figure 77), the 2-amino function makes direct contact to **a** and the phenol of Tyr377, which also seems to compete for the water ligand. Furthermore, H_2O_112 makes contact to the backbone carbonyl of Arg380, as well as the bridging water (**b** in Figure 77, H_2O_61) between the backbone carbonyl of Tyr380 and N^{π}. When bound to the mutant variant in $EgtB_{Y377F}$:2A-DMH (**B** in Figure 77), the same basic coordination to the water **a** (H_2O_160) is observed, albeit without the interaction with the missing phenol group. Instead, the water is in contact with an additional molecule of water (**c** in Figure 77, H_2O_107) above the coordination plane. This type of tetrahedral coordination in turn is found in both crystal structures with the native ligand DMH. In the $EgtB_{wt}$:DMH complex (**C** in Figure 77), the water molecule **a** (H_2O_729) is coordinated by the phenolic group of Tyr377, the backbone carbonyl of Arg380, the bridging water **b** (H_2O_719) between the backbone carbonyl of Tyr380 and N^{π} and the water above this coordination plane **c** (H_2O_766). Finally, in the complex with the mutant $EgtB_{Y377F}$:DMH (**D** in Figure 77), the missing contact with the phenol group is replaced by recruitment of a different water (**d** in Figure 77, H_2O_198), which in turn is in direct contact to the manganese-coordinated water (**e** in Figure 77, H_2O_304). In summary, the 2-amino group of 2A-DMH neatly fits in to the hydrogen-bonding network of the enzyme active site, which could be an explanation for its high affinity for both the native and the mutant variant. Not only does it make direct contact with Tyr377 in the wild type, but it seems to emulate some important aspects of this missing residue in the mutant. As noted above, in the $EgtB_{Y377F}$:2M-DMH complex (**D** in Figure 77) this key water molecule is not present and the imidazole ring is only stabilized by

the bridging water **b** (H_2O_117) between the backbone carbonyl of Tyr380 and N^{π}. Displacement of water **a** by the 2-methyl group is quite likely energetically unfavorable, a fact which may provide one explanation for the decreased affinity of this ligand.

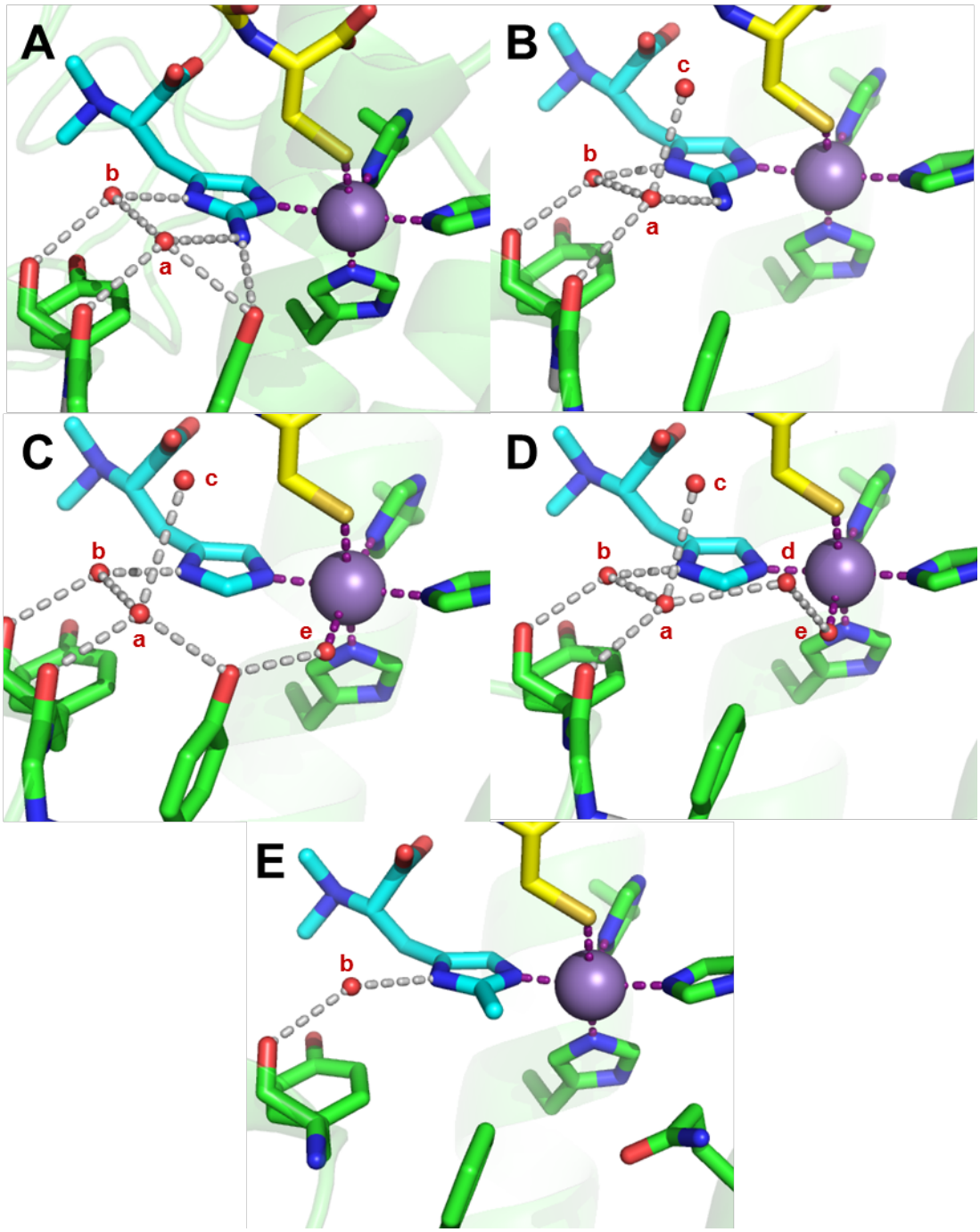


Figure 77 Hydrogen bonding system in the active site of EgtBwt and EgtBY377F with either DMH or 2A-DMH. The interactions of some key water residues are shown in black, for details on the H-bonding network see text.

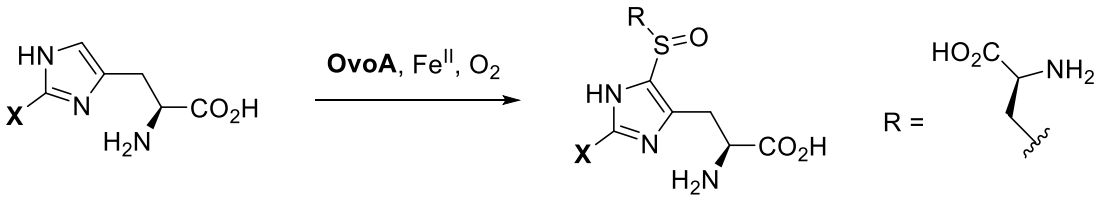
Based upon literature pK_A values for imidazole (7.0),[169] 2-methyl imidazole (7.9)[170] and 2-amino imidazole (8.5),[163] we expect the pK_A values of 2M-DMH and 2A-DMH to be 1 and 1.5 units higher than histidine.[171] Based upon the high binding affinity of 2A-DMH, which is

comparable to that of DMH, in contrast to the weak affinity for the 2-methyl variant, it is evident that the electronic properties of the imidazole ring do not dictate binding strength. Instead it seems more likely that the increased hydrophobic steric bulk of the 2-methyl group accounts for the decreased affinity of 2M-DMH, which cannot be compensated for by H-bonding, as is the case for 2A-DMH.

Since the 2-amino group in 2A-DMH has higher pK_A than its methyl counterpart or DMH, we can assume that it is a stronger nucleophile and thus a stronger donor ligand. We hypothesized that binding of both more electron-rich variants may decrease oxygen activation by the metal center. To test this, we investigated the viability of analogous compounds (2-amino *L*-histidine and 2-methyl *D/L*-histidine) as substrates for the related sulfoxide synthase OvoA.

3.3.11 Expanded Substrate Scope of OvoA

The sulfoxide synthase OvoA is a distant relative to EgtB and additionally contains a second, SAM-dependent methyltransferase domain. It thus catalyzes the first and last step of ovothiol biosynthesis (see also **Section 3.1.3** and **Section 1.4.1**). Different homology models of OvoA suggest that it contains a similar active site to that of EgtB.[135,148] Furthermore, as mentioned in **Section 3.1.3**, OvoA accepts a range of modified or truncated substrate analogues, among others 2-fluoro *L*-histidine with a turnover rate similar to the native substrate *L*-histidine.



X = H, F, CH_3, NH_2

Figure 78 Top: Conversion of C2-modified histidine derivatives to the corresponding sulfoxide by OvoA.

Since OvoA catalyzes C-S bond formation between the cysteine and the C-5 position of histidine, this enzyme provides an opportunity to test whether the electronic properties of the histidine analogues affect oxygen activation. Although the affinity of the enzyme for both substrates was quite low, the reaction rates were remarkably similar for both substrates and only marginally lower than for the native substrate as shown in **Table 13**. As was the case in the rather electron-poor substrate 2-fluoro *L*-histidine, the electronic structure and pK_A of the histidine-substrate does not seem to influence the catalytic efficiency of OvoA. Due to the similarities between OvoA and EgtB, we can extrapolate that this is indeed not a factor in the sulfoxide synthase activity of either enzyme.

Table 13 Rates of conversion of C2-modified histidine derivatives to the corresponding sulfoxide by OvoA.

	$k_{cat,app}$ [s^{-1}]	$K_{M,app}$ [µM]	k_{cat}/K_M [$M^{-1}s^{-1}$]
2A-His	0.33 ± 0.02	2'000 ± 200	170
2M-His	0.83 ± 0.07	5'300 ± 900	160
His	1.4 ± 0.1	200 ± 10	7'000

Reaction conditions were: 10 mM cysteine (2A-His) or 1 mM cysteine (2M-His), 1 µM OvoA (2A-His) or 1 µM OvoA (2M-His), 10 µM $FeSO_4$, 1 mM ascorbate, 2 mM TCEP, 20 mM NaCl in 50 mM phosphate buffer at pH8.

3.4 Discussion

One of the greatest challenges for enzymes is how to channel their reactivity in order to control product specificity. They must be able to efficiently promote the forward reaction with minimal contributions toward alternate, unproductive pathways. Enzymes which catalyze energetically challenging reactions rely on reactive intermediates to achieve this feat, much like synthetic chemists rely on harsh conditions or highly reactive catalysts to drive forward similar transformations. The reactions catalyzed by NHI enzymes are notoriously tricky to control, as this class makes use of species so reactive, that the enzyme must prevent them from reacting with the active site residues themselves. Within the protein core of a given enzyme, first and second order coordination spheres around the reactive site are key for keeping enzymes selective and productive. Naturally, as enzymes have evolved their unique strategies over millennia, the rational design of enzyme catalysts for single or multistep reactions remains extremely challenging.[172,173]

Derailing enzymes from their "natural" pathway is one way to gain understandings of their precise mechanism of action. This has been done successfully in the past by either modifying the enzyme itself by site-directed mutagenesis, modifying the enzyme substrate(s) or a combination of the two.[7,174] Careful analysis of the products formed and comparison of the reaction rates offers a deeper understanding even of more complex, multistep reactions.

In our case, we have chosen to study a thiol dioxygenase pathway in the NHI enzyme EgtB, which becomes dominant over the natural reactivity after introduction of a single point-mutation in the active site.[150] Similarly, in the related sulfoxide synthase OvoA it has been shown that the same reactivity is achieved when the enzyme is incubated with non-physiological substrates.[37,175] By pairing both the wild type and the mutant variant of EgtB with substrates that allow for only the dioxygenation pathway to take place, we have been able to shed light onto the factors governing the enzymes selectivity. This in turn gives us additional information on what intermediates are formed during the reaction and in which sequence.

Major structural differences between sulfoxide synthases and cysteine dioxygenases, as shown in Figure 79, point towards two diverse evolutionary backgrounds. On the one hand, CDO assumes the characteristic β-barrel fold of the cupin superfamily, which is typical for NHI enzymes.[140] In contrast, EgtB consists of two domains, with the active site located at the interface between the two. A first, N-terminal four-helical bundle forms the DinB_2 domain, which is reminiscent of Zn-dependent glutathione *S*-transferases.[149] The second, C-terminal domain

adopts a less ordered fold, which is closely related to the copper dependent formylglycine generating enzyme (FGE) and is thus termed the FGE-like domain.[176,177]

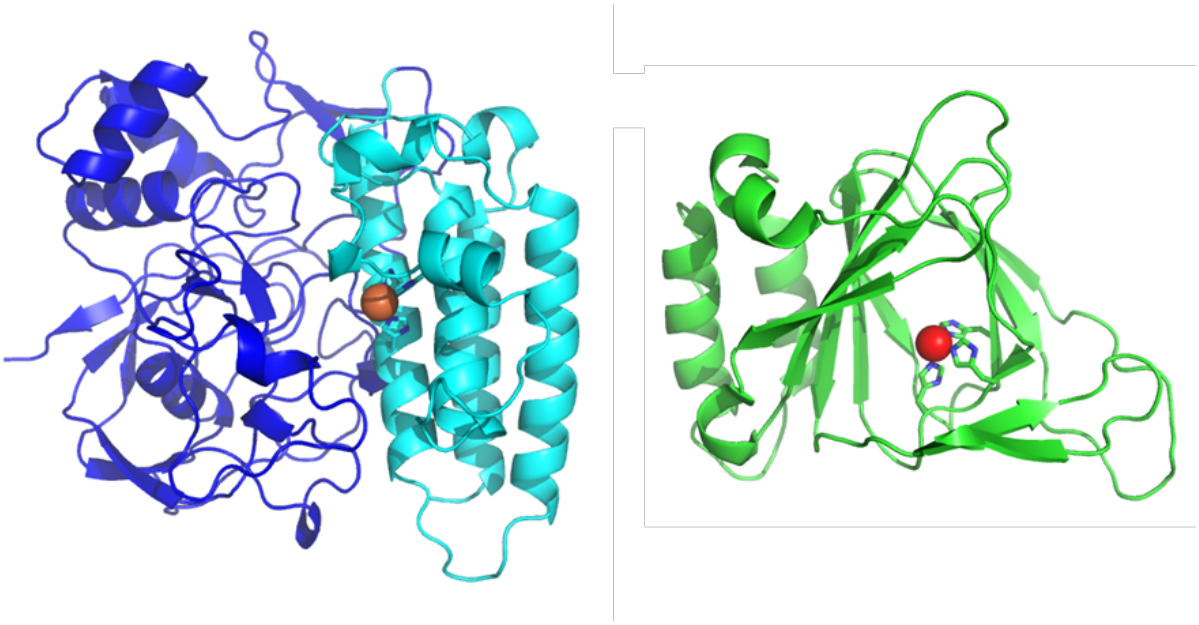


Figure 79 X-Ray crystal structures of EgtB from *M. thermoresistibile* (left, PDB: 4X8E)[146] and CDO from *Mus musculus* (right, PDB: 2ATF).[140] Showing the 3-His facial triad coordinating the respective metals in the active site (EgtB: Fe in brown; CDO: Ni in red). In the EgtB structure the DinB_2 domain is shown in cyan, while the FGE-like domain is shown in blue.

Regardless of these differences in evolutionary origin, EgtB and CDO display some remarkable commonalities. Both active sites harbor a 3-histidine facial triad coordinating Fe^{II}, leaving the other three sites open for substrate binding. Both enzymes catalyze four-electron oxidation reactions where at least one atom of the molecular oxygen used is inserted into the product.[135] Neither enzyme is in need of an auxiliary reducing agent or co-substrate in order to perform catalysis. And finally, as mentioned above, the sulfoxide synthase OvoA can also catalyze dioxygenation of cysteine under appropriate conditions.[37,175] All these factors, along with the fact that EgtB can effectively be turned into a thiol dioxygenase by the point mutation Y377F, support the assumption that sulfoxide synthesis and thiol dioxygenation share one or multiple reactive intermediates and apply similar strategies for catalysis. In recent years, controversy has arisen over the precise role of Tyr377 in determining which synthetic pathway EgtB follows during catalysis.

3.4.1 The Sulfoxide Synthase Pathway of EgtB

In previous studies it was shown that by mutating Tyr377 to Phe, sulfoxide synthase activity was reduced by a factor of 1000 from 8.5 x 10^{-1} s^{-1} to 0.9 x 10^{-4} s^{-1}. When the pH was lowered from 8.0 to 6.0, the same activity was increased by a factor of 3.5 to 3.2 x 10^{-3} s^{-1}. Furthermore this activity displayed a kinetic solvent isotope effect (KSIE) of 1.9 ± 0.1, showing that there is involvement of exchangeable protons in the rate limiting step. This is in contrast to both the sulfoxide synthase activity of $EgtB_{wt}$ and the dioxygenase activity of $EgtB_{Y377F}$, both of which showed KSIEs near unity (1.2 ± 0.2 and 0.9 ± 0.1 respectively).[150] These findings suggest that the role of Tyr377 is to regulate the reactivity of the Fe^{III} superoxo species by protonation. In absence of this crucial protonation step, dioxygenation is favored over sulfoxide synthesis. This would be analogous to the protonation of superoxide to the peroxyl radical, changing its reactivity from a reductant to an oxidant.[178]

The question arises, whether the role of Tyr377 is to support sulfoxide synthase activity or to quench possible dioxygenation of the thiol substrate. To this end we investigated the enzyme in combination with substrates that allow only for the dioxygenation route to take place. By adding additional protons into the active site, essentially rendering the catalytic center more acidic, dioxygenase activity was suppressed.

The crystal structure of the EgtB:DMH:γGC complex reveals that the C-2 hydrogen points towards the phenol group of Tyr377, indicating that a polar group such as an amine would be well-tolerated. Indeed, kinetic and structural data showed that 2A-DMH is an excellent ligand for both $EgtB_{wt}$ and $EgtB_{Y377F}$ with an affinity that rivals the natural ligand DMH. The orientation of the amino group in the active site is such that it is conceivable that protons from the amine could be transferred to the Fe^{III} superoxo species, leading to reduced dioxygenation of γGC (**B** → **B'** in Figure 80). As 2-aminoimidazoles have a pKA of 8.5,[163] one would expect 2A-DMH to be fully protonated under assay conditions (pH 8.0). Due to the fact that the C-2 position of this compound can hardly form a C-S bond – the intermediates would be highly unstable (**C** → **D** in Figure 80) – the only route for the enzyme to take in this case is that towards dioxygenation. As the related compound 2-aminobenzimidazole is characterized by a low oxidation potential of E_{ox} = 0.6 V,[179] the redox potential of 2A-DMH is also likely to be quite low. It thus seems plausible that 2A-DMH could alternatively act as an H-atom donor in a hydrogen atom transfer (HAT) instead of a protonation step (**A** → **A'** in Figure 80). In a recent computational study by Faponle *et al*,[180] Tyr377 was considered as a proton and electron donor to reduce the Fe^{III} superoxo species. The resulting tyrosyl radical is then reduced concomitantly to C-S bond formation. This proton-coupled electron

transfer pathway would prevent the Fe^{III} superoxo species from reacting with the thiol and producing the unwanted dioxygenation product. If this is the case, it is conceivable that 2A-DMH could act as a substitute for Tyr377 in this role. The observation that no spike in ascorbate or TCEP consumption is observed during the reaction suggests that any reactive species formed are quenched within the enzyme.

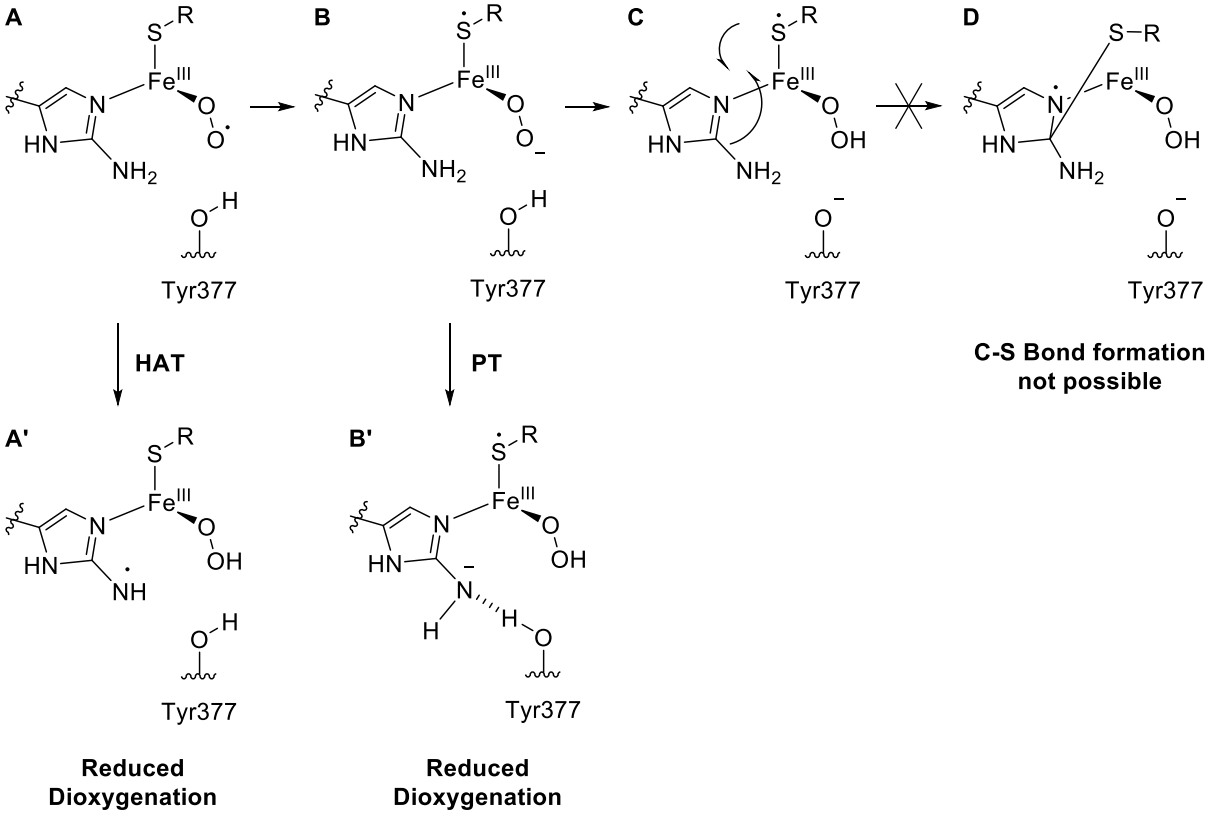


Figure 80 Conceivable intermediates in the complex $EgtB_{wt}$:2A-DMH.

In the mutant variant $EgtB_{Y377F}$, dioxygenation is reduced by a factor of 13 and 17 when adding either an amino or methyl group in the C-2 position of DMH. In the wild type enzyme, dioxygenation takes place at a 6-fold lower rate with the 2-amino than the 2-methyl variant. If this was a purely steric effect, one would expect the rates to be similar, or possibly even lower for 2M-DMH as a methyl group is slightly larger than an amino group.[181] In summary, $EgtB_{wt}$ and $EgtB_{Y377F}$ in complex with 2A-DMH are roughly 20 and 100 fold less active thiol dioxygenases compared to the system $EgtB_{Y377F}$:DMH. The complexes of $EgtB_{wt}$ and $EgtB_{Y377F}$ with 2M-DMH are roughly 10 and 20 fold less active thiol dioxygenases compared to the system $EgtB_{Y377F}$:DMH. Based on the observation that no dioxygenation product can be detected in the $EgtB_{wt}$:DMH complex, it can be

assumed that this process is at least as inefficient in the $EgtB_{wt}$:2A-DMH and $EgtB_{wt}$:2M-DMH system.

3.4.2 The Thiol Dioxygenase Pathway of EgtB

The consensus mechanism for thiol dioxygenation involves a highly reactive Fe^{IV} oxoferryl species.[137,182,183] Based on the crystal structures of EgtB in complex with 2A-DMH, one would expect the N-H bond of the 2-amino group to point directly towards this strong oxidant. Furthermore, given the large difference in redox potentials between 2-aminoimidazoles and oxoferryls,[184,185] this hydrogen atom transfer would most likely have to be irreversible. If the N-H bond is cleaved it must then be reverted in a subsequent step as no decay of 2A-DMH is observed during the reaction. It is also possible that sulfur oxygenation is faster than N-H bond cleavage or that no oxoferryl species is formed during thiol dioxygenation.[186] The observation that the 2-amino group of 2A-DMH slows the reaction down 6-fold more than the 2-methyl group of 2M-DMH, demonstrates that a polar group is more detrimental to the reactivity of the oxoferryl species than a hydrophobic group. One possible explanation for this can be illustrated when comparing the two resonance structures of the oxoferryl complex after the first oxidation of the thiolate, shown in Figure 81. In the structure **A**, the thiolate has donated one electron to the oxo-ligand, giving the latter a more anionic character. This anionic group could then be stabilized by protonation by the 2-amino group, which in turn is stabilized by Tyr377 (**B**). Product formation (**C**) would be significantly slowed down by such a hydrogen bonding network, for which the crystal structure of the $EgtB_{wt}$:2A-DMH complex provides suitable evidence.

If on the other hand, as shown in structure **A'**, electron density remains on the sulfoxide, the oxo-ligand would have more radical character. As described above, in that case a HAT could take place (**B'**), which could lead to oxygenation of the 2-amino group (**C'**). No such hydroxyamino product or decomposition-products thereof[187] were observed in any of our assays however, leading us to conclude that species **A** in Figure 81 is the dominant resonance structure.

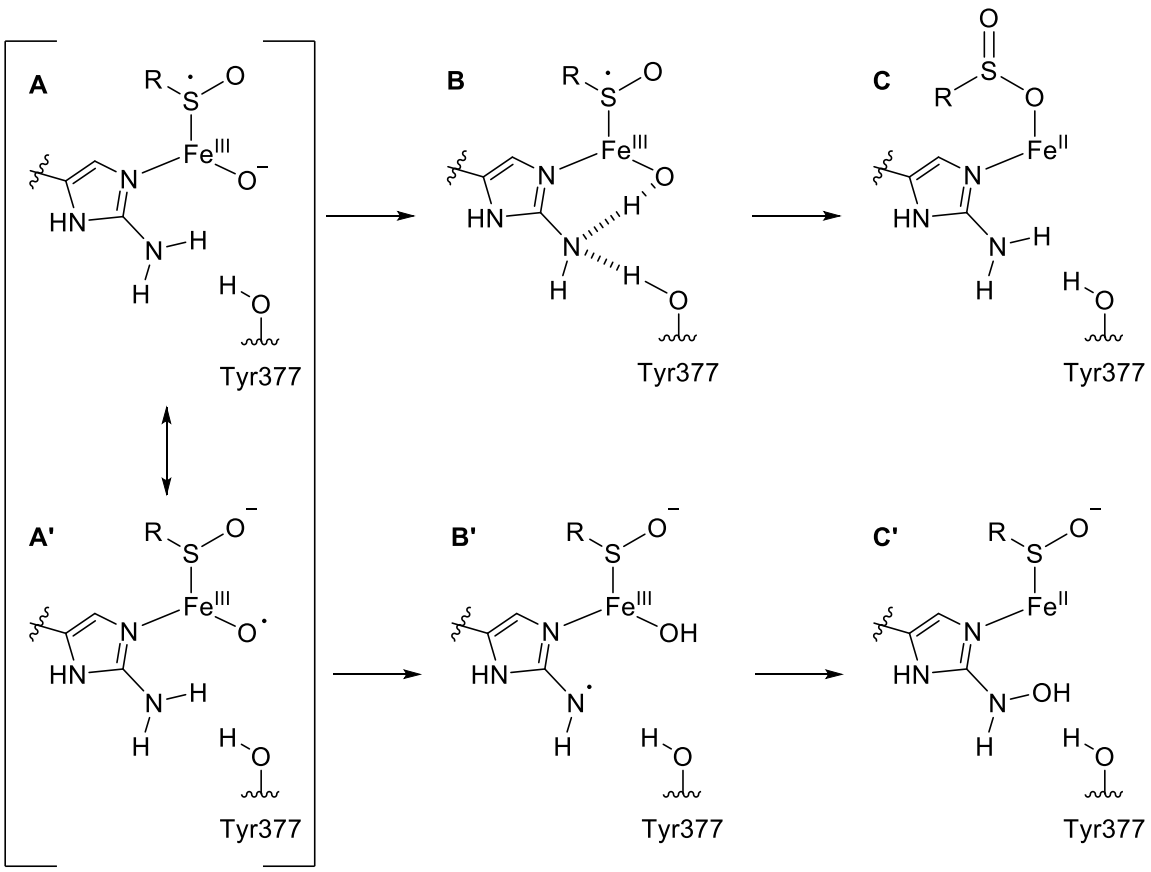


Figure 81 Alternative electron distribution in the oxoferryl species leading to γGC-dioxygenation. The top structure (**A**) positions higher electron density on the oxo-ligand, giving it a more anionic character amenable to proton transfer (**B**). This would lead to a reduction in activity of dioxygenation (**C**), which is in agreement with our kinetic data. The bottom structure (**A'**), describes an oxo-ligand with more radical character, which could engage in hydrogen atom transfer with 2A-DMH (**B'**). This however could lead to the *N*-hydroxylated product (**C'**), which was not observed in our assays and which are inherently instable and labile to attack by nucleophiles.[187]

3.4.3 Alternative Model for Timing of Proton Transfer

As discussed briefly in **Section 3.1.3** and shown in Figure 82, a computational study by Wei *et al*. using DFT calculations came to the conclusion that C-S bond formation occurs after irreversible O-atom transfer to the thiol.[148] According to this model, in both the dioxygenation and the sulfoxide synthase pathway this step (**A** and **B**) results in an Fe^{IV}=O intermediate (**C**, displayed here as the $RSO^{\bullet}$-Fe^{III}=O species instead of RSO^{-}-Fe^{IV}=O) which is protonated by Tyr377 to prevent the second oxygen atom being transferred to sulfur. Tyr377 then deprotonates N^{π} of the imidazole ring via two active site water molecules. Next, C-S bond formation occurs between the oxygenated

thiol and C-2 of the imidazole ring (**E**), followed by water-assisted deprotonation of the C-2 position by Tyr377 (**F**) and protonation of the imidazole ring (**G**).

This suggestion is in contrast to the model proposed by Goncharenko *et al.*[146,149,150] where the oxoferryl species is only formed in the dioxygenation pathway, which becomes dominant in the mutant enzyme $EgtB_{Y377F}$.

In this model, the 2-amino group of 2A-DMH would likely reduce the reactivity of the Fe^{IV} oxoferryl species and reduce dioxygenation of γGC. As mentioned above, the 2-amino group is in a position amenable to either proton or hydrogen atom transfer to the oxoferryl species and this step would almost certainly be irreversible due to the difference in reduction potential.

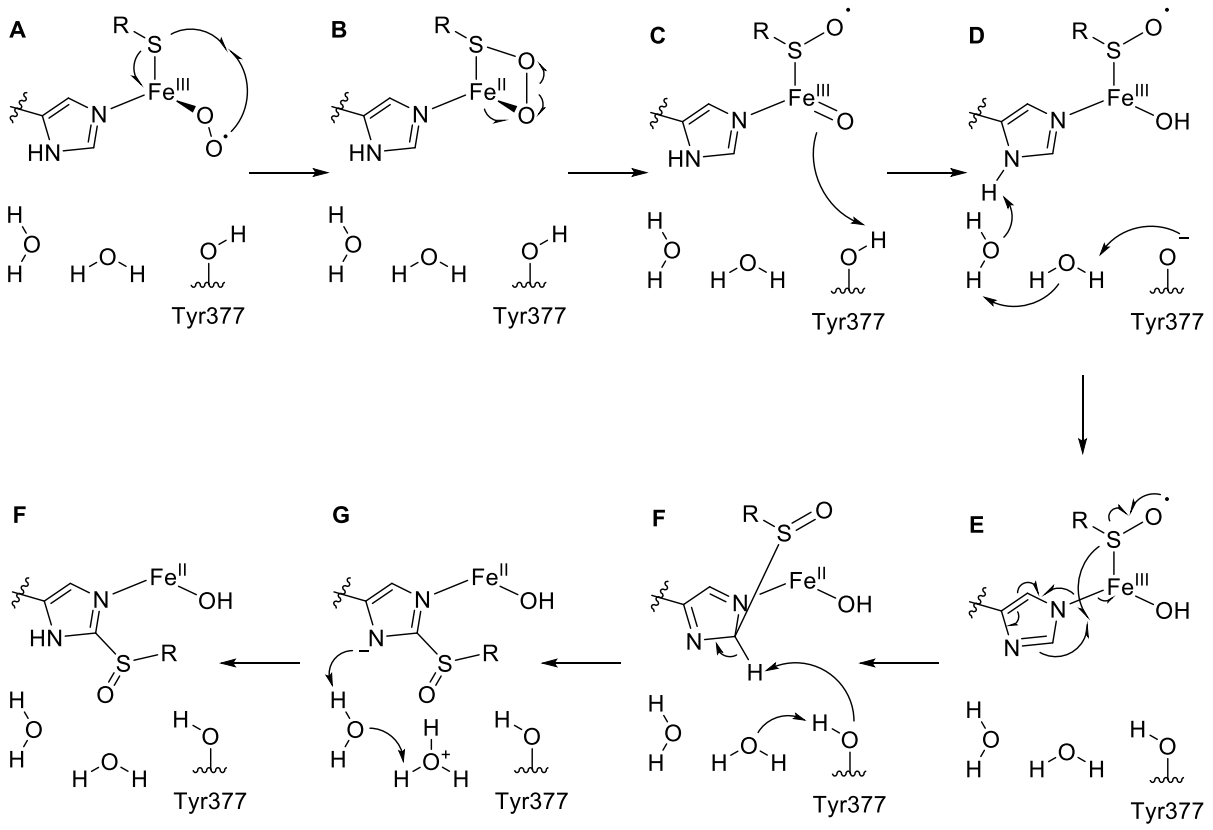


Figure 82 Suggested mechanism of sulfoxide synthesis by EgtB based on computational DFT calculations.[148]

3.4.4 Promiscuity of OvoA

Although the binding affinity of OvoA for both 2A-His and 2M-His is quite low, the enzyme is able to catalyze C-S bond formation on each substrate with a remarkably high turnover rate once saturation is reached. This finding underlines the assumption that the electronic structure of the imidazole ring does not play a large role in oxygen activation. In addition, these results further demonstrate the wide potential substrate scope of OvoA and may open new semisynthetic routes to novel bifunctionalized thioimidazole derivatives.

3.5 Conclusion

We have characterized two complementary probes for the reactivity of EgtBs two possible reactive pathways; 2A-DMH and 2M-DMH. The first inserts a hydrophilic and slightly acidic group towards the iron-bound oxygen necessary for catalysis, while the second introduces hydrophobic steric bulk into the active site. Crystallographic data provides a plausible explanation for the high affinity of EgtB for 2A-DMH and the low affinity for 2M-DMH. The neat incorporation into the hydrogen-bonding system of the former and the opposite for the latter not only explain the discrepancy in affinities, but also offer an explanation for their difference in reactivity in the wild type enzyme. While 2A-DMH could conceivably stabilize the system by its incorporation into the hydrogen bonding network, 2M-DMH does exactly the opposite, disrupting the geometry around the iron center with its hydrophobic steric bulk. Control experiments on the pH-sensitivity of these probes in combination with the viability of their non-methylated analogs as substrates for OvoA show that it is not the electronic structure of the imidazole substrate that reduces activity. To test whether oxygen binding is influenced, kinetic measurements at different oxygen concentrations will be necessary

Regarding the reactivity of the crucial active site residue Tyr377, these studies have raised some interesting points. In the mechanistic proposal by Goncharenko *et al.*[146,149,150] an Fe^{III} superoxo complex is formed and the role of this residue is to provide a proton to this species, slowing down oxygen transfer to sulfur. According to this model, bifurcation of the two possible pathways – sulfoxidation or dioxygenation – takes place early on in the catalytic cycle. In the alternative proposal by Wei *et al.*[148] the first step in the catalytic cycle is the same for both pathways and involves formation of an Fe^{IV} oxoferryl species. Herein, the role of Tyr377 is to protonate the oxo-ligand and prevent it from reacting with the sulfoxide and forming the dioxygenase product. Regulation of which pathway is chosen by the enzyme is thus achieved at a later stage in the catalytic cycle.

In the $EgtB_{wt}$:2A-DMH complex where only the dioxygenation pathway is available, the second oxygen transfer step is slowed down by acidic protons supplied by the co-ligand. This speaks for less radical character on the oxygen atom, since the anionic oxygen can hydrogen bond with the 2-amino group, which is further stabilized by interactions with the phenolic group of Tyr377. This observation applies to both models, albeit at different points in the catalytic cycle.

The low binding affinity of 2M-DMH additionally supports the hypothesis that deprotonation and re-aromatization of the imidazole ring after C-S bond formation do not precede sulfoxidation. This finding is in agreement with both proposed mechanisms and with kinetic data from our laboratory,

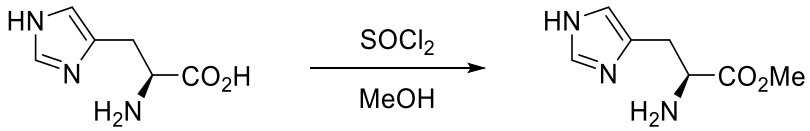
indicating that sulfoxidation and rearomatization is likely a concerted step. The lack of stabilization for a non-polar functional group in the C-2 position underlines that the thioether intermediate would be bound with a low affinity. In turn, a low binding affinity for an intermediate would not be in the interest of the enzyme as it only aims to release the fully synthesized product after completion of the transformation. Again, this is in agreement with both suggestions for the catalytic cycle although it does not address the question of when sulfoxidation takes place precisely.

We expect that our kinetic and structural contributions will help to further drive the discussion on the most likely mechanism for this highly interesting enzyme.

3.6 EgtB - Experimental

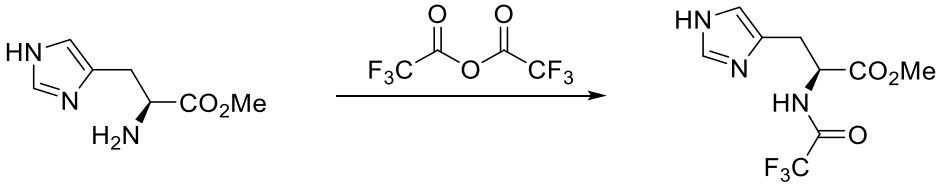
3.6.1 Synthesis

Synthesis of methyl *L*-histidinate



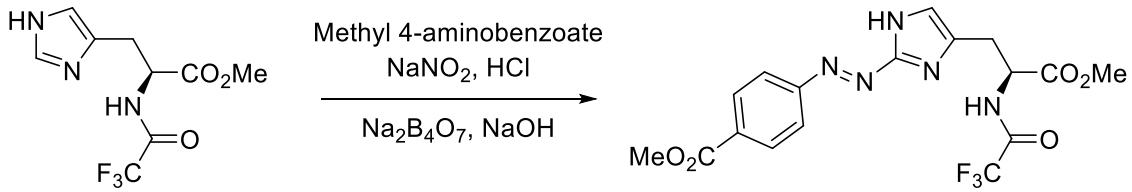
To a suspension of *L*-histidine (10.00 g, 64.45 mmol, 1 eq) in MeOH (100 ml) was carefully added $SOCl_2$ (7 ml, 96.67 mmol, 1.5 eq) at 0 °C, upon which the solution became clear. The solution was stirred at 0 °C for 30 min and then allowed to warm up to ambient temperature. The solution was refluxed for 24 h, after which no more starting material was detected by UPLC-MS. The solvent was evaporated to give the pure product (15.60 g, 64.44 mmol, 99 %) in quantitative yield as slightly off white powder as the 2 x HCl salt. **^{1}H NMR (400 MHz, MeOD)** δ 8.94 (d, *J* = 1.4 Hz, 1H), 7.53 (s, 1H), 4.47 – 4.43 (m, 1H), 3.86 (s, 3H), 3.49 – 3.35 (m, 2H). **MS** *m/z* calcd. for $C_7H_{12}N_3O_2$ [M+H$^+$]: 170.1, found: 170.1

Synthesis of methyl (2,2,2-trifluoroacetyl)-*L*-histidinate (72)



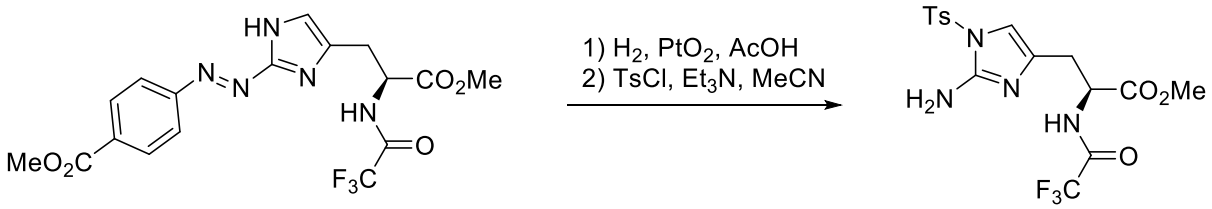
To trifluoracetic anhydride (27 ml, 193.35 mmol, 3 eq) cooled to 0 °C was added methyl *L*-histidinate (15.60 g, 64.44 mmol, 1 eq). The light brown suspension was allowed to warm up to room temperature and stirred for 17 h. Since some starting material was still not completely dissolved, dry MeOH (15 ml) was added and the reaction was stirred for an additional 5 h. The solvent was evaporated, to give the pure product (18.86 g, 55.78 mmol, 87 %) as off white solid as the 2 x HCl salt. **^{1}H NMR (400 MHz, MeOD)** δ 8.84 (s, 1H), 7.35 (s, 1H), 4.90 (m, 1H), 3.79 (s, 3H), 3.42 (m, 1H), 3.19 (m, 1H). **MS** *m/z* calcd. for $C_9H_{11}F_3N_3O_3$ [M+H]$^+$ 266.1, found 266.1

Synthesis of methyl (*S*,*E*)-4-((4-(3-methoxy-3-oxo-2-(2,2,2-trifluoroacetamido)propyl)-1*H*-imidazol-2-yl)diazenyl)benzoate (73)



A solution of NaNO2 (4.98 g, 58.57 mmol, 1.05 eq) in water (10 ml) was slowly added to an ice-cold solution of Methyl-4-aminobenzoate (8.432 g, 55.78 mmol, 1 eq) in water (150 ml) and conc. HCl (20 ml), whilst stirring vigorously at 0 °C for 1 h. During this time, a solution of **72** (18.86 g, 55.78 mmol, 1 eq) in water (500 ml) was brought to pH 7 by portion wise addition of solid sodium tetraborate and then to pH 8 by addition of 1 M NaOH. The solution of diazonium salt was gradually added to the aforementioned solution. The pH of the solution was maintained at 7-8 by adding 1 N sodium hydroxide. The dark orange suspension was stirred at 0 °C for 30 min. The dark red precipitated crude product was removed by centrifugation, dissolved in methanol, dried over Na_2SO_4 and concentrated to dryness *in vacuo*._The crude product was purified in 3 batches by absorbing the product onto silica and purifying by flash chromatography on silica gel (DCM/MeOH, 98.5:1.5), to give the pure product (17.25 g, 40.38 mmol, 72 %) as orange solid. **^{1}H NMR (400 MHz, MeOD)** δ 8.17 (d, *J* = 8.6 Hz, 1H), 7.98 (d, *J* = 8.6 Hz, 1H), 7.18 (s, 1H), 4.89 (m, 1H), 3.94 (s, 3H), 3.78 (s, 3H), 3.36 (m, 1H), 3.17 (m, 1H). **MS** *m/z* calcd. for $C_{17}H_{17}F_3N_5O_5$ $[M+H]^+$ 428.1, found 428.1

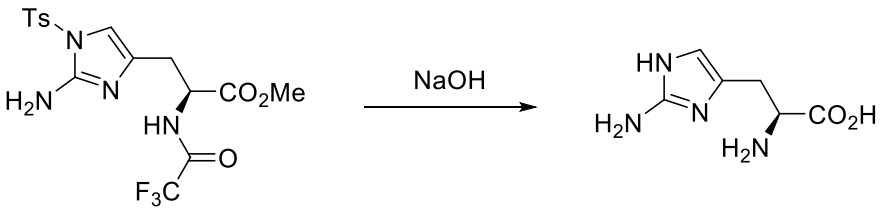
Synthesis of methyl (*S*)-3-(2-amino-1-tosyl-1*H*-imidazol-4-yl)-2-(2,2,2-trifluoroacetamido) propanoate (74)



To a solution/suspension of **PE2.7.A** (1 g, 2.34 mmol, 1 eq) in AcOH (10 ml), was added Pt_2O (266 mg, 1.17 mmol, 0.5 eq), an initial hydrogen pressure of 70 bar was applied in an autoclave and the dark red solution/suspension was stirred at ambient temperature. Since analysis by UPLC-MS showed that reaction was not yet complete, additional portions of PtO_2 (86 mg, 0.35 mmol,

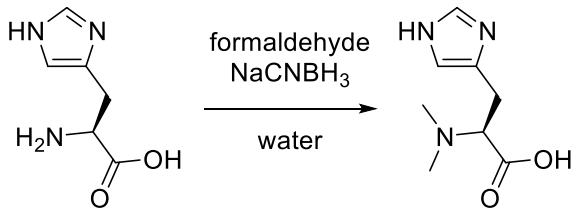
0.15 eq) were added followed by reloading with 70 bar H_2 after 40 and 64 h. After another 18 h reaction time, the colorless reaction mixture was filtered and evaporated to dryness. The residue was treated with water and washed with 4 x EtOAc, the aqueous layer was lyophilized and the product, which was confirmed by UPLC-MS and ^{1}H-NMR, was used in the next step without further purification. The intermediate product (950 mg, 1.70 mmol, ~50 % pure, 1 eq) was dissolved in MeCN (10 ml) and Et_3N (1.18 ml, 8.48 mmol, 5 eq) at 0 °C and was added TsCl (1.616 g, 8.48 mmol, 5 eq) in MeCN (5 ml) and the solution was stirred for 30 min. The reaction was allowed to warm up to room temperature and stirred for an additional 18 h. The solvent was evaporated under reduced pressure, the residue was treated with water and the aqueous layer was extracted with 3 x DCM. The combined organic layers were washed with sat. $NaHCO_3$ and brine, dried over Na_2SO_4 and evaporated to dryness to yield the crude product as brown oil. The crude product was purified by column chromatography (cHex/EtOAc 1:1) to give the product (123 mg, 0.28 mmol, 12 %) as orange oil. **Intermediate: ^{1}H NMR (400 MHz, MeOD)** δ 8.84 (s, 1H), 7.35 (s, 1H), 4.90 (m, 1H), 3.79 (s, 3H), 3.42 (m, 1H), 3.19 (m, 1H). **MS** *m/z* calcd. for $C_9H_{11}F_3N_3O_3$ $[M+H]^+$ 266.1, found 266.1 **Product: ^{1}H NMR (400 MHz, MeOD)** δ 7.83 (d, *J* = 8.4 Hz, 2H), 7.42 (d, *J* = 9.2 Hz, 2H), 6.72 (s, 1H), 4.67 (dd, *J* = 9.6, 4.8 Hz, 1H), 3.67 (s, 3H), 2.94 – 2.89 (m, 1H), 2.81 – 2.74 (m, 1H), 2.43 (s, 3H). **MS** *m/z* calcd. for $C_{16}H_{18}F_3N_4O_5S$ $[M+H]^+$ 435.1, found 435.2

Synthesis of (*S*)-2-amino-3-(2-amino-1*H*-imidazol-4-yl)propanoic acid (71)



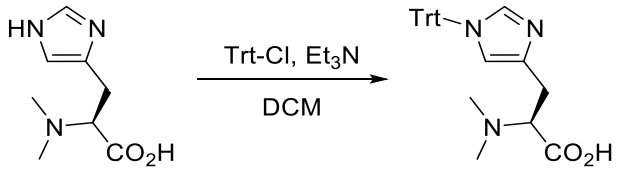
PE2.12.C (123 mg, 0.28 mmol, 1 eq) was dissolved in 1 M NaOH (1 ml) and the orange solution was stirred at room temperature for 2.5 h (start 09:40). The reaction mixture was cooled to 0 °C and brought to pH 2 using 1 M HCl (ca. 1 ml). Organic impurities were extracted with 3 x DCM and the aqueous layer was evaporated to dryness. The crude product was purified by procedure B (see **Section 2.6.2**) to give the product (45 mg, 0.26 mmol, 93 %) as green solid after lyophilization. **^{1}H NMR (400 MHz, D_2O)** δ 6.69 (s, 1H), 4.08 – 4.05 (m, 1H), 3.17 – 3.12 (m, 2H). **MS** *m/z* calcd. for $C_6H_{11}N_4O_2$ $[M+H]^+$ 171.0877, found 171.0876

Synthesis of N^{α},N^{α}-dimethyl-*L*-histidine (68)



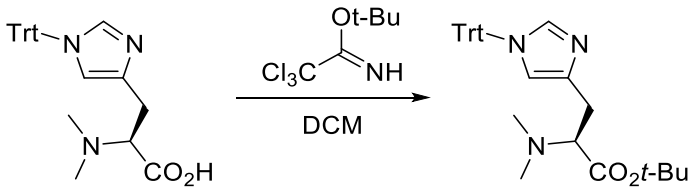
Using procedure A (see **Section 2.6.2**) and starting with *L*-histidine (2 g, 12.9 mmol, 1 eq), formaldehyde (37 % w/w, 4.8 ml, 64.5 mmol, 5 eq) and $NaCNBH_3$ (1.46 g, 38.7 mmol, 3 eq), the product **55** was obtained as white solid (2.103 g, 11.5 mmol, 89 %) after purification according to procedure B (see **Section 2.6.2**) and lyophilization. **^{1}H NMR (400 MHz, D_2O)** δ 7.66 (s, 1H), 6.95 (s, 1H), 3.63 (dd, *J* = 7.8, 6.0 Hz, 1H), 3.15 – 3.03 (m, 4H), 2.67 (s, 9H). **MS** *m/z* calcd. for $C_8H_{14}N_3O_2$ $[M+H^+]$: 184.11, found: 184.07

Synthesis of N^{α},N^{α}-dimethyl-N^{τ}-trityl-*L*-*histidine* (75)



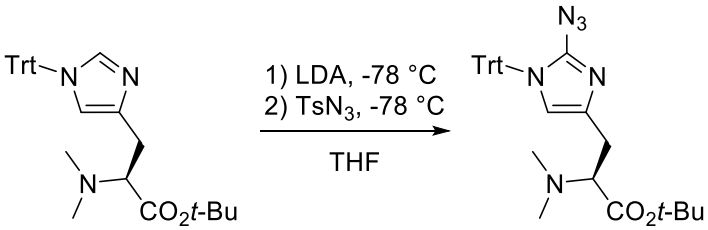
To a suspension of **68** (600 mg, 3.27 mmol, 1 eq) in DCM (40 ml) was added Et_3N (2.28 ml, 16.37 mmol, 5 eq) and the solution was stirred at reflux for 45 min before adding trityl-Cl (1.826 g, 6.55 mmol, 2 eq) and refluxing overnight. The reaction mixture was treated with water and extracted with 3 x DCM, the organic phases were combined, dried over Na_2SO_4 and evaporated to dryness to give the crude product as sticky white solid purified by flash chromatography (DCM/MeOH 9:1) to give the pure product (790 mg, 1.86 mmol, 57 %) as slightly off white solid. **^{1}H NMR** (400 MHz, MeOD) δ 7.42 – 7-36 (m, 10H), 7.18 – 7.15 (m, 6H) 6.92 (s, 1H), 3.81 – 3.78 (m, 1H), 3.31 – 3.22 (m, 1H), 3.15 – 3.10 (m, 1H), 2.88 (s, 6H). **MS** *m/z* calcd. for $C_7H_{12}N_3O_2$ $[M+H]^+$: 426.22, found: 426.05; $[M+Na]^+$: 428.20, found: 448.07; $[M+K]^+$: 464.17, found: 464.06

Synthesis of *tert*-butyl N^{α},N^{α}-dimethyl-N^{τ}-trityl-*L*-histidinate (76)



To a solution of **75** (780 mg, 1.83 mmol, 1 eq) in dry DCM (20 ml) was added *tert*-butyl 2,2,2-trichloroacetimidate (0.656 ml, 3.67 mmol, 2 eq) and acetic acid (0.105 ml, 1.83 mmol, 1 eq) at room temperature. The solution was stirred at reflux for 72 h, adding additional tert-butyl 2,2,2-trichloroacetimidate (0.328 ml, 1.84 mmol, 1 eq) in portions of one equivalent until no more starting material could be detected by TLC. The reaction mixture was treated with water and extracted with 3 x DCM, the organic phases were combined, dried over Na_2SO_4 and evaporated to dryness. The crude product was purified by flash chromatography (DCM + 5 % MeOH), to give the pure product (582 mg, 1.21 mmol, 66 %) as colorless oil. **^{1}H NMR** (400 MHz, MeOD) δ 7.38 – 7.36 (m, 11H), 7.13 – 7.11 (m, 7H), 6.71 (s, 1H), 3.41 (dd, *J* = 10.4, 4.8 Hz, 1H), 3.00 – 2.94 (m, 1H), 2.81 – 2.76 (m, 1H), 2.37 (s, 6H), 1.35 (s, 9H). **MS** *m/z* calcd. for $C_7H_{12}N_3O_2$ $[M+H]^+$: 482.28, found: 482.13.12; $[M+Na]^+$: 504.26, found: 504.12

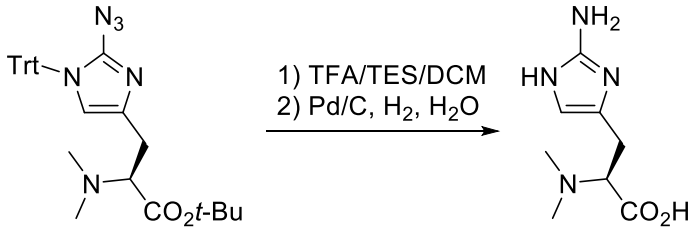
Synthesis of *tert*-butyl (*S*)-3-(2-azido-1-trityl-1*H*-imidazol-4-yl)-2-(dimethylamino) propanoate (77)



To a solution of **76** (562 mg, 1.17 mmol, 1 eq) in dry THF (12 ml) und Ar, cooled to -78 °C was added 2 M LDA (2.33 ml, 4.67 mmol, 4 eq) and the orange solution was stirred at -78 °C for 30 min. 10 % TsN_3 in toluene (9.66 ml, 4.90 mmol, 4.2 eq) was added and the red solution was allowed to warm up to room temperature and stirred for 1 h. The reaction was quenched by adding 1 M NH_4Cl (12 ml). The aqueous layer was extracted with 3 x EtOAc, the organic layers were combined, dried over $NaSO_4$ and evaporated to dryness. The crude product was purified by flash column chromatography (EtOAc/cHex 1:1) obtaining the pure product (272 mg, 0.52 mmol, 45 %)

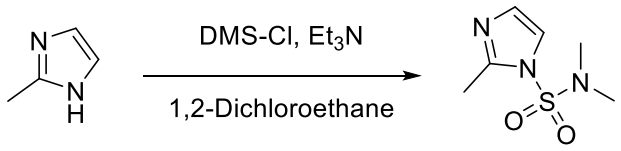
as slightly yellow solid. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.33 – 7.28 (m, 8H), 7.12 – 7.07 (m, 6H), 3.40 (dd, *J* = 10.5, 4.6 Hz, 1H), 2.94 (dd, *J* = 14.2, 10.5 Hz, 1H), 2.69 (dd, *J* = 14.5, 4.3 Hz, 1H), 2.36 (s, 6H), 1.38 (s, 9H). **MS** *m/z* calcd. for $C_7H_{12}N_3O_2$ $[M+H]^+$: 523.28, found: 523.12; $[M+Na]^+$: 545.26, found: 545.13

Synthesis of (*S*)-3-(2-azido-1*H*-imidazol-4-yl)-2-(dimethylamino)propanoic acid (69)



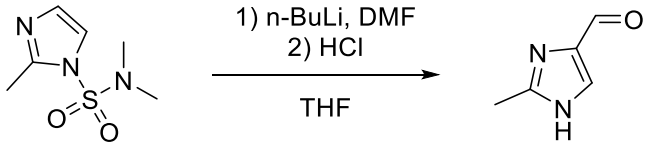
To a solution of **77** (272 mg, 0.52 mmol) in DCM (4.3 ml) was added TFA (2.1 ml) and TES (1.7 ml) and the orange solution was stirred overnight at room temperature. The solvent was evaporated and then co-evaporated 3 x with toluene, treated with water, washed with 3 x DCM and the aqueous layer was lyophilized to give the intermediate product as a slightly orange oil. Both the azide and the reduced amine were detected by ESI-MS and ^{1}H-NMR at a ratio of 55:45 respectively. The intermediate was dissolved in water (10 ml), 10 % Pd/C (49 mg, 0.05 mmol, 0.1 eq) was added and H_2 (1 atm) was applied by first bubbling through the solution for 5 min and then continued stirring for 6 h at room temperature. The Pd/C was removed by filtration through a pad of celite, and the filtrate was lyophilized. To the product was added 1 M HCl (3 ml) and evaporated to dryness (3 x), then treated with water and purified according to procedure B (see **Section 2.6.2**) and lyophilized to give the product (113 mg, 0.42 mmol, 91 %) as slightly orange solid. **Azide: ^{1}H NMR** (400 MHz, D_2O) δ 6.85 (s, 1H), 4.32 (t, *J* = 6.1, 6.1 Hz, 1H), 3.36 – 3.31 (m, 1H), 3.26 – 3.20 (m, 1H), 3.03 (s, 6H). **MS** *m/z* calcd. for azide $C_8H_{13}N_6O_2$ $[M+H]^+$: 225.11, found: 224.95 **Amine: ^{1}H NMR** (400 MHz, D_2O) δ 6.74 (s, 1H), 4.18 (dd, *J* = 8.7, 5.2 Hz, 1H), 3.36 – 3.31 (m, 1H), 3.26 – 3.20 (m, 1H), 2.98 (s, 6H). **MS** *m/z* calcd. for amine $C_8H_{15}N_4O_2$ $[M+H]^+$: 199.12, found: 198.94 **Final Product: ^{1}H NMR** (400 MHz, D_2O) δ 6.74 (s, 1H), 4.15 (dd, *J* = 8.7, 5.2 Hz, 1H), 3.36 – 3.31 (m, 1H), 3.26 – 3.20 (m, 1H), 2.99 (s, 6H). **^{13}C NMR** (101 MHz, D_2O) δ 169.82, 146.98, 120.06, 111.76, 67.02, 40.44, 22.53. **MS** *m/z* calcd. for $C_8H_{15}N_4O_2$ $[M+H]^+$: 199.1190, found: 199.1190

Synthesis of *N*,*N*,2-trimethyl-1*H*-imidazole-1-sulfonamide (78)



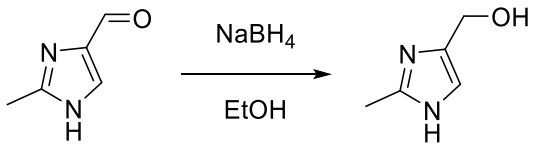
To a suspension of 2-methylimidazole (500 mg, 6.1 mmol, 1 eq) in 1,2-dichloroethane (6 ml) was added *N*,*N*-dimethylaminosulfonyl chloride (1.34 ml, 12.2 mmol, 2 eq) and Et_3N (0.98 ml, 7.0 mmol, 1.15 eq) and the turbid solution was stirred at room temperature for 17 h. The white precipitate was filtered off, the filter cake was washed with 1,2-dichloroethane and the filtrate was washed with saturated aqueous Na_2CO_3, dried over Na_2SO_4 and concentrated. The crude product was purified by flash chromatography (DCM to elute unreacted DMAS-Cl, then DCM/MeOH 4:1 to elute product) to give the pure product (1.135 g, 6.0 mmol, 98 %) as pale yellow oil. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.23 (d, *J* = 1.7 Hz, 1H), 6.96 (d, *J* = 1.7 Hz, 1H), 2.92 (s, 6H), 2.66 (s, 3H). **MS** *m/z* calcd. for $C_6H_{12}N_3O_2S$ $[M+H]^+$: 190.1, found: 190.1

Synthesis of 2-methyl-1*H*-imidazole-4-carbaldehyde (79)



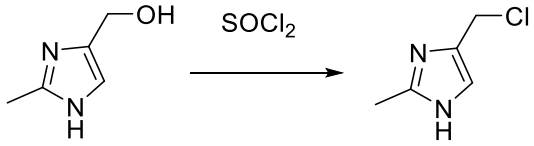
A solution of **78** (1.135 g, 6 mmol, 1 eq) in dry THF (40 ml) was cooled to -78 °C. 1.6 M *n*-BuLi in hexanes (4.5 ml, 7.2 mmol, 1.2 eq) was slowly added and the yellow solution was stirred at -78 °C for 30 min before adding DMF (3 ml, 39 mmol, 6.5 eq) and stirring for additional 1 h at -78 °C. The reaction mixture was allowed to gradually heat up to room temperature over the course of 1 h and then stirred for 30 min at room temperature. The pH was adjusted to 1 with concentrated HCl (aq) and the reaction mixture was stirred at room temperature for 2 h. The reaction mixture was poured into a saturated solution of $NaHCO_3$, insuring the pH was adjusted to 8, and the THF was evaporated. The residue was extracted with 3 x EtOAc, the combined organic phases were dried over Na_2SO_4 and evaporated to give the crude product (400 mg, 3.6 mmol, 61 %) as yellow crystals. **^{1}H NMR** (400 MHz, DMSO) δ 12.73 (d, *J* = 149.9 Hz, 1H), 9.60 (s, 1H), 7.87 (s, 1H), 2.32 (s, 3H). **MS** *m/z* calcd. for $C_5H_7N_2O$ $[M+H]^+$: 111.1, found: 111.1

Synthesis of (2-methyl-1*H*-imidazol-4-yl)methanol (80)



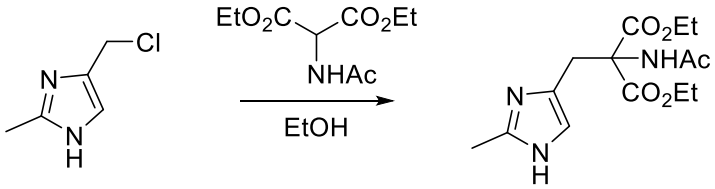
To a solution of **79** (400 mg, 3.63 mmol, 1 eq) in EtOH (20 ml) was added NaBH4 (137 mg, 3.63 mmol, 1 eq) and the yellow solution was stirred at room temperature for 4 h. Brine (10 ml) was added resulting in a white suspension. The solvent was evaporated and the crude material was suspended in MeCN and filtered. The filtrate was evaporated to give the pure product (182 mg, 1.62 mmol, 45 %) as white crystals. **^{1}H NMR** (400 MHz, DMSO) δ 6.66 (s, 1H), 4.29 (s, 2H), 2.21 (s, 3H). **MS** *m/z* calcd. for $C_5H_9N_2O$ $[M+H]^+$: 113.1, found: 113.2

Synthesis of 4-(chloromethyl)-2-methyl-1*H*-imidazole (81)



80 (182 mg, 1.62 mmol, 45 %) was dissolved in 1 ml $SOCl_2$ and stirred overnight at room temperature. The solvent was evaporated to give the pure product (246 mg, 1.47 mmol, 91 %) as yellow oil. **^{1}H NMR** (400 MHz, DMSO) δ 14.55 (d, *J* = 139.4 Hz, 1H), 7.59 (s, 1H), 4.84 (s, 2H), 2.57 (s, 3H). **MS** *m/z* calcd. for $C_5H_8ClN_2$ $[M+H]^+$: 131.0, found: 131.1

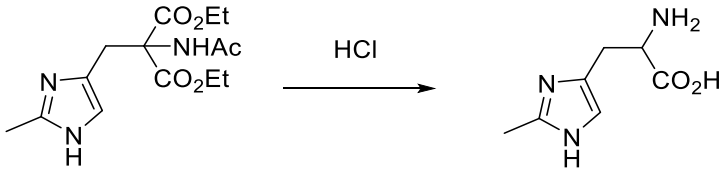
Synthesis of diethyl 2-acetamido-2-((2-methyl-1*H*-imidazol-4-yl)methyl)malonate (82)



To a solution of NaH (135 mg, 3.39 mmol, 2.3 eq, 60 % in mineral oil) in dry ethanol (5 ml), was slowly added a solution of diethyl acetamidomalonate (416 mg, 1.91 mmol, 1.3 eq) in dry ethanol (5 ml) at 0 °C under N_2. To this clear solution was slowly added **81** (246 mg, 1.47 mmol, 1 eq) in dry ethanol (5 ml) and the now slightly beige solution was stirred at 0 °C for 30 min and then at

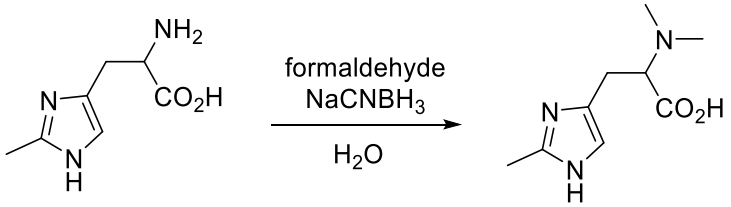
room temperature for 90 min. The solution was filtered and the filtrate was evaporated to dryness to give the crude product as oily brown solid. After purification by flash chromatography (DCM/MeOH 95:5), the product (330 mg, 1.96 mmol, 72 %) was obtained as a yellow oil. **^{1}H NMR** (400 MHz, DMSO) δ 11.51 (br s, 1H), 8.01 (s, 1H), 6.56 (s, 1H), 4.12 (t, *J* = 7.08 Hz 4H), 3.27 (s, 2H), 2.17 (s, 3H), 1.90 (s, 3H), 1.15 (t, *J* = 7.11 Hz, 3H). **MS** *m/z* calcd. for $C_{14}H_{22}N_3O_5$ $[M+H]^+$: 312.2, found: 312.2

Synthesis of 2-amino-3-(2-methyl-1H-imidazol-4-yl)propanoic acid (83)



To a suspension of **82** (330 mg, 1.06 mmol) in water (3 ml) was added conc. HCl (3 ml) and the clear yellow solution was heated to 100 °C and stirred under reflux for 19 h. The solvent was evaporated to dryness to give the product (250 mg, 1.03 mmol, 97 %) as light brown oil, which crystalized in the fridge. **^{1}H NMR** (400 MHz, DMSO) δ 14.66 (br s, 1H), 14.29 (br s, 1H), 8.66 (br s, 3H), 7.33 (s, 1H), 4.31 (s, 1H), 3.27 – 3.15 (m, 2H), 2.54 (s, 3H). **^{13}C NMR** (500 MHz, DMSO) δ 169.57 (s), 143.82 (s), 126.05 (s), 117.10 (s), 50.97 (s), 25.23 (s), 11.07 (s). **MS** *m/z* calcd. for $C_7H_{12}N_3O_2$ $[M+H]^+$: 170.1, found: 170.2

Synthesis of 2-(dimethylamino)-3-(2-methyl-1H-imidazol-4-yl)propanoic acid (70)



Using procedure A (see **Section 2.6.2**) and starting with **83** (50 mg, 0.296 mmol, 1 eq), formaldehyde (37 % in water, 88 µL, 1.18 mmol, 4 eq) and $NaCNBH_3$ (93 mg, 1.48 mmol, 5 eq), this product was obtained as white solid (40 mg, 0.203 mmol, 69 %) after purification according to procedure B (see **Section 2.6.2**) and lyophilization. **^{1}H NMR (400 MHz, D_2O)** δ 7.21 (s, 1H), 3.89 (dd, *J* = 9.4, 4.7 Hz, 1H), 3.44 – 3.39 (m, 1H), 3.29 – 3.22 (m, 1H), 2.97 (s, 6H), 2.60 (s, 3H). **^{13}C**

NMR (101 MHz, D_2O) δ 172.69, 145.52, 130.80, 116.00, 70.10, 41.48, 25.62, 12.07. **MS** *m/z* calcd. for $C_9H_{16}N_3O_2$ $[M+H]^+$ 198.1237, found 198.1239 *m/z* calcd. for $C_9H_{16}N_3NaO_2$ $[M+Na]^+$ 220.1056, found 220.1054

3.6.2 Enzyme Kinetics

3.6.2.1 Determination of Sulfoxide Formation

All reaction rates were determined by HPLC-based assays, using a cation exchange HPLC (IE-HPLC) to track product formation and a standard method (**Table 14**).

EgtB: Reactions containing 100 mM phosphate buffer (pH 8.0), 100 mM NaCl, 0.0125-1 mM γGC, 4 µM FeSO4, 2 mM tri-(2-carboxyethyl) phosphine (TCEP), 2 mM ascorbate and 0.0125-7.5 mM DMH/2A-DMH/2M-DMH in a final volume of 250 µL were incubated at 26 ℃. 40 µL of reaction mixture were quenched by adding to 20 µL 1 M phosphoric acid at different time intervals (4-6 samples) and analyzed by cation exchange HPLC to monitor the sulfoxide formation. Enzyme concentrations ranged from 0.06-2.5 µM, depending on the substrate.

OvoA: Reactions containing 50 mM phosphate buffer (pH 8.0), 20 mM NaCl, 10 µM FeSO4, 2 mM tri-(2-carboxyethyl) phosphine (TCEP), 1 mM ascorbate and 10 mM Cysteine for aminohistidine or 1 mM Cysteine for methylhistidine and histidine in a final volume of 250 µL were incubated at 26 ℃. 40 µL of reaction mixture were quenched by adding to 20 µL 1 M phosphoric acid at different time points (at least 4 samples) and analyzed by IE-HPLC to monitor the sulfoxide formation. Enzyme concentration of the aminohistidine reaction was 1 µM and the enzyme concentration of the methylhistidine and histidine reaction was 0.5 µM.

Table 14 Standard method for HPLC analysis of EgtB reaction products and starting materials; Solvent A: 20 mM phosphate pH 2.0; Solvent B: 20 mM phosphate pH 2.0 with 1 M NaCl.

Time (min)	Solvent A (%)	Solvent B (%)
0.01	85	15
2	80	20
8	1	99
9	1	99
11	85	15

3.6.2.2 Determination of Dioxygenase Activity via Thiol Consumption

Consumption of γGC or cysteine was followed using an HPLC-based assay on a reversed phase (RP) HPLC and a standard method (**Table 15**).

Since the thiol itself is not detectable at sufficiently low concentrations, it was derivatized with 4-bromo-7-methoxycoumarine (4-BMC) before analysis by RP-HPLC (as shown in **Figure 70**). The standard conditions were: 0.5-2.5 µM $EgtB_{wt}$ or $EgtB_{Y377F}$, 0.1 mM γGC, 4 µM $FeSO_4$, 2 mM ascorbate, 2 mM TCEP, 100 mM NaCl in 100 mM phosphate buffer at pH 8. Reactions were quenched by adding 40 µL of the reaction mixtures to 10 µL 30 mM 4-BMC in DMSO and 40 µL MeCN and incubating in the dark for 30 min. Each sample was diluted with 90 µL 0.1 % TFA and centrifuged (16'000 RPM, 15 min) before injecting 20 µL into the HPLC using the standard method.

Table 15 Standard method for HPLC analysis of EgtB reaction products and starting materials; Solvent A: water + 1 % MeCN + 0.1 % TFA; Solvent B: MeCN + 0.1 % TFA.

Time (min)	Solvent A (%)	Solvent B (%)
0.01	85	15
2	80	20
8	1	99
9	1	99
11	85	15

3.6.2.3 Inhibition of Sulfoxide Synthase Activity in $EgtB_{wt}$ and $EgtB_{Th2}$ by 2A-DMH and 2M-DMH

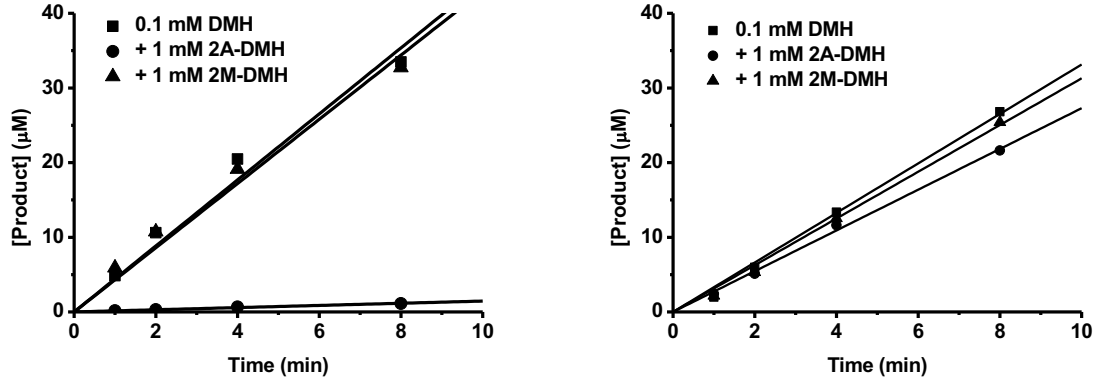


Figure 83 Inhibition kinetics of 2A-DMH and 2M-DMH with $EgtB_{wt}$ (left) and $EgtB_{Th2}$ (right). Reaction conditions were: 0.1 mM DMH, 1 mM γGC or cysteine, 0-1 mM 2A-DMH or 2M-DMH, 0.06 µM $EgtB_{wt}$(left) or 1.5 µM $EgtB_{Th2}$(right), 2 mM TCEP, 4 µM $FeSO_4$, 2 mM ascorbate, 100 mM NaCl in 100 mM phosphate pH 8.

3.6.2.4 Michealis-Menten Curves: *L*-DMH and *D*-DMH as Substrates for Sulfoxide Synthase

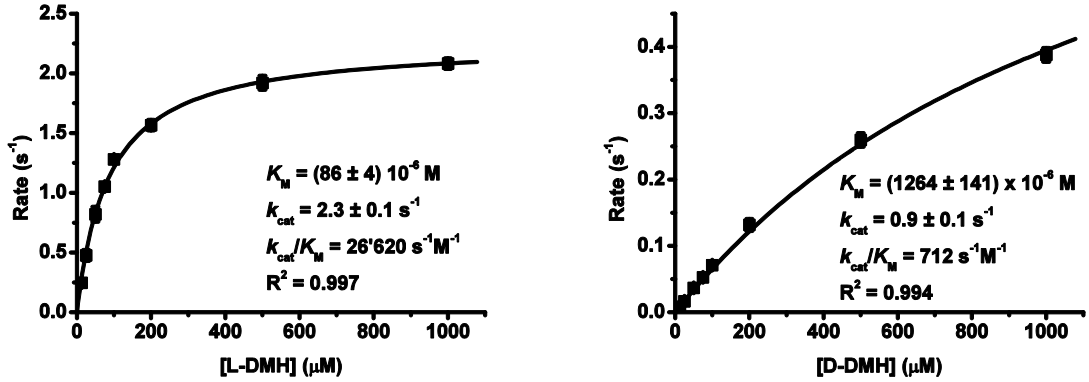
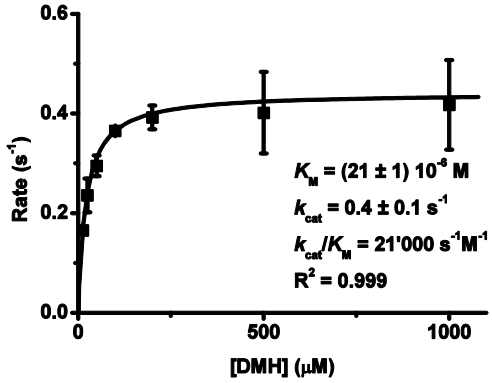
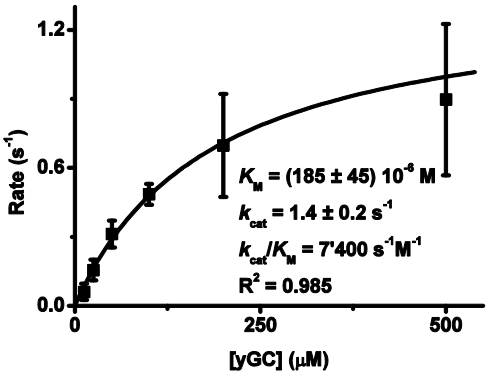


Figure 84 Michaelis-Menten parameters for sulfoxide formation between ***L*-DMH and *D*-DMH** and γGC, catalyzed by $EgtB_{Th1}$. Conditions: 12.5-1000 µM ***L*-DMH or *D*-DMH**, 1 mM γGC, 0.06 µM $EgtB_{wt}$, 2 mM TCEP, 0.4 µM $FeSO_4$, 2 mM ascorbate, 100 mM NaCl in 100 mM phosphate pH 8, final volume: 250 µL. The reaction mixture was incubated at 26 °C and at least 4 aliquots of 40 µL were quenched by addition of 20 µL 1 M phosphoric acid at intervals in a timeframe of 1-16 min. These were then analyzed by HPLC using the standard method.

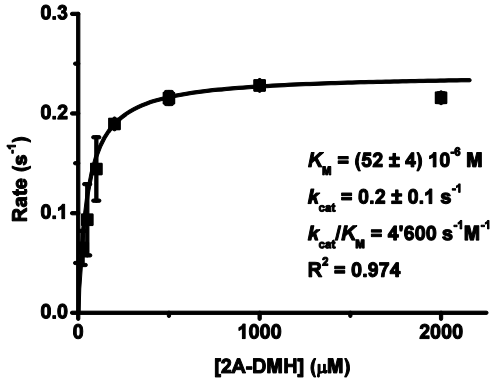
3.6.2.5 Michealis-Menten Curves: Dioxygenase Activity of $EgtB_{Y377F}$



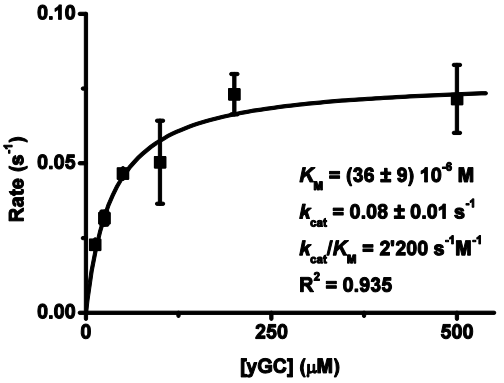
[γGC] = 0.1 mM



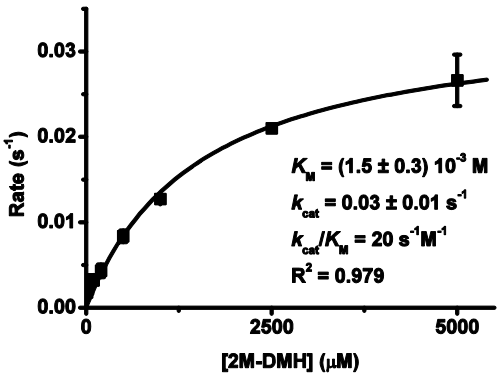
[DMH] = 1 mM



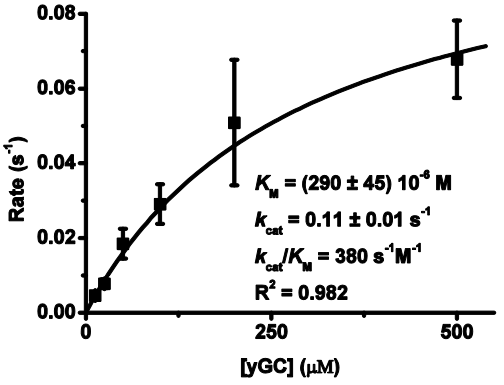
[γGC] = 0.1 mM



[2A-DMH] = 1 mM



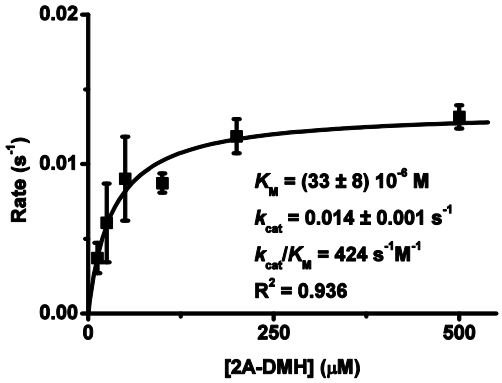
[γGC] = 0.1 mM



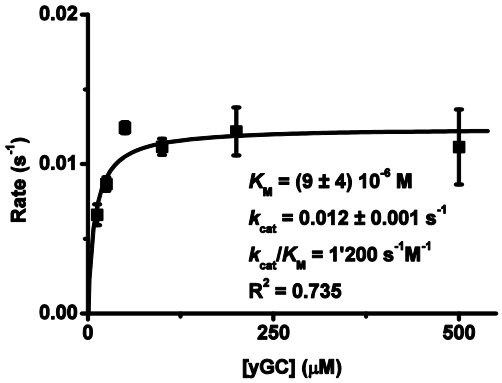
[2M-DMH] = 7.5 mM

Figure 85 Left: Michaelis-Menten parameters of $EgtB_{Y377F}$ for DMH (top), 2A-DMH (middle) and 2M-DMH (bottom) at [γGC] = 0.1 mM. Conditions were: (12.5 – 1000) µM DMH, (25 – 2000) µM 2A-DMH, (25 – 5000) µM 2M-DMH 0.5 µM (DMH and 2A-DMH) or 2.5 µM (2M-DMH) $EgtB_{Y377F}$, 0.1 mM γGC, 4 µM $FeSO_4$, 2 mM ascorbate, 2 mM TCEP, 100 mM NaCl in 100 mM phosphate buffer at pH 8. **Right**: Michaelis Menten Parameters of γGC at saturated conditions of DMH (top), 2A-DMH (middle) and 2M-DMH (bottom) Conditions were: 0.5 µM (top) and 2.5 µM (bottom) $EgtB_{Y377F}$, 12.5 – 500 µM γGC, 4 µM $FeSO_4$, 2 mM ascorbate, 2 mM TCEP, 100 mM NaCl in 100 mM phosphate buffer at pH 8. Reactions were quenched by adding 40 µL of the reaction mixtures to 10 µL 30 mM 4-BMC in DMSO and 40 µL MeCN and incubating for 30 min. Each sample was diluted with 90 µL 0.1 % TFA and centrifuged (16'000 RPM, 15 min) before injecting 20 µL into the HPLC.

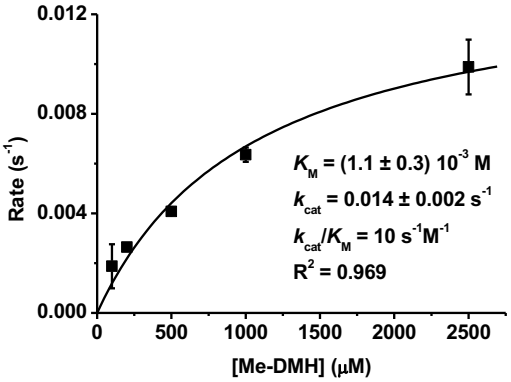
3.6.2.6 Michealis-Menten Curves: Dioxygenase Activity of $EgtB_{wt}$



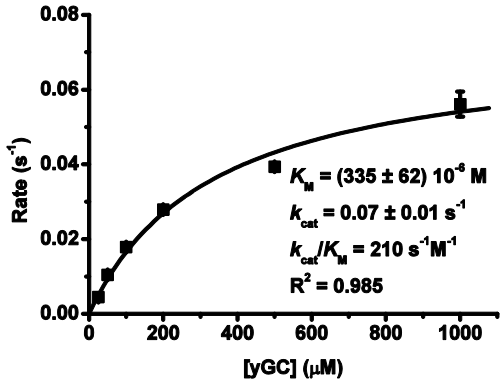
[γGC] = 0.1 mM



[2A-DMH] = 1 mM



[γGC] = 0.1 mM



[2M-DMH] = 5 mM

Figure 86 Top left: Michaelis Menten Parameters of 2A-DMH at [γGC] = 0.1 mM. Top right: Michaelis Menten Parameters of γGC at [2A-DMH] = 1 mM. Bottom left: Michaelis Menten Parameters of 2M-DMH at [γGC] = 0.1 mM. Bottom right: Michaelis Menten Parameters of γGC at [2M-DMH] = 5 mM. Reaction conditions were: 2.4 µM $EgtB_{wt}$, 4

μM $FeSO_4$, 2 mM ascorbate, 2 mM TCEP, 100 mM NaCl in 100 mM phosphate buffer at pH 8. Reactions were quenched by adding 40 μL of the reaction mixtures to 10 μL 30 mM 4-BMC in DMSO and 40 μL MeCN and incubating for 30 min. Each sample was diluted with 90 μL 0.1 % TFA and centrifuged (16'000 RPM, 15 min) before injecting 20 μL into the HPLC.

3.6.2.7 Crystal Structures

Crystals were grown and their structures solved by Dr. Peer Lukat at the Helmholtz Centre for Infection Research in Braunschweig, using conditions described in the experimental section. All crystals belong to the space group P43212 with cell constants a,b = 135 and c = 141 A. The crystal packing and conformation of all chains were essentially identical to those reported for $EgtB_{wt}$.[146] The structures superimpose with RMSDs < 0.150 Å (380 atoms), for a full table of all RMSD values, see **Table 16**. The resolution of the structures for $EgtB_{Y377F}$ in complex with DMH (2.2 Å), ADMH (2.5 Å) and MDMH (2.5 Å), and also $EgtB_{wt}$ in complex with 2ADMH (2.4 Å) is high. The electron densities of the complexes with DMH and 2ADMH allowed precise placement of the unnatural ligand in the active site. Due to anisotropies of the $EgtB_{Y377F}$:2M-DMH crystal and because of low occupancy, the electron density of 2M-DMH is incomplete. Given the low stability of the EgtB:2M-DMH complex this result is not surprising (see **Table 8**).

Analysis of the crystal structures (distances, angles, RMSDs etc.), as well as generation of all images thereof was done using PyMol software.

Table 16 RMSD values in Å for the EgtBwt and EgtBY377F complexes with DMH, 2A-DMH and 2M-DMH as well as γGC and Mn.

Crystal Structure	**EgtB:DMH:γGC**	**EgtB apo**	**EgtB:TMH**
PDB code	4X8D	4X8B	4X8E
	RMSD (Atoms)	RMSD (Atoms)	RMSD (Atoms)
$EgtB_{wt}$:2A-DMH	0.133 (383)	0.114 (392)	0.128 (393)
$EgtB_{Y377F}$:2A-DMH	0.130 (386)	0.104 (383)	0.116 (382)
$EgtB_{Y377F}$:2M-DMH	0.150 (380)	0.143 (380)	0.149 (370)
$EgtB_{Y377F}$:DMH	0.145 (378)	0.108 (372)	0.110 (362)

3.6.3 Protein Production

$EgtB_{wt}$, $EgtB_{Th2}$ and $EgtB_{Y377F}$ were produced and kindly provided by Kristina Goncharenko. All enzymes were expressed in *E. coli* BL21 cells and purified using Ni^{2+} NTA agarose beads (Qiagen) following standard procedures.[149]

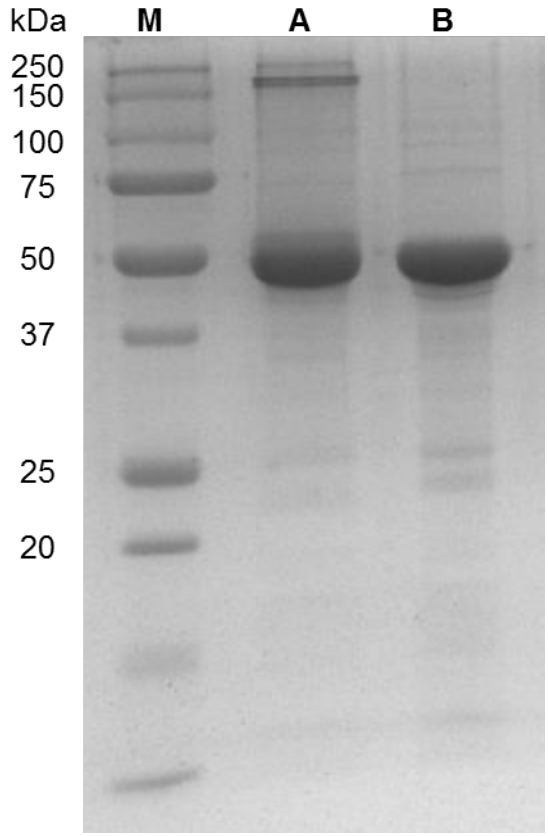


Figure 87 SDS-PAGE pictures of $EgtB_{wt}$ after Ni-NTA affinity chromatography. **M** – Molecular weight marker; **A** – $EgtB_{wt}$ without addition of a reducing agent; **B** - $EgtB_{wt}$ with addition of DTT;

Sequence of $EgtB_{wt}$:

GSSHHHHHHSSGLVPRGSHMGVAVPHRAELARQLIDARNRTLRLVDFDDAELRRQYDPLMSPLVWDLAHIG
QQEELWLLRGGDPRRPGLLEPAVEQLYDAFVHPRASRVHLPLLSPAQARRFCATVRSAVLDALDRLPEDAD
TFAFGMVVSHEHQHDETMLQALNLRSGEPLLGSGTALPPGRPGVAGTSVLVPGGPFVLGVDLADEPYALDN
ERPAHVVDVPAFRIGRVPVTNAEWRAFIDDGGYRQRRWWSDAGWAYRCEAGLTAPQFWNPDGTRTRFGHVE
DIPPDEPVQHVTYFEAEAYAAWAGARLPTEIEWEKACAWDPATGRRRRYPWGDAAPTAALANLGGDALRPA
PVGAYPAGASACGAEQMLGDVWEWTSSPLRPWPGFTPMIYQRYSQPFFEGAGSGDYRVLRGGSWAVAADIL
RPSFRNWDHPIRRQIFAGVRLAWDVDRQTARPGPVΓGC

m/z ($EgtB_{wt}$): calc.: 51343 Da, measured: 51374 Da (mass difference of 31 Da due to oxidation)

$\varepsilon 280(EgtB_{wt}) = 111630\ M^{-1}cm^{-1}$

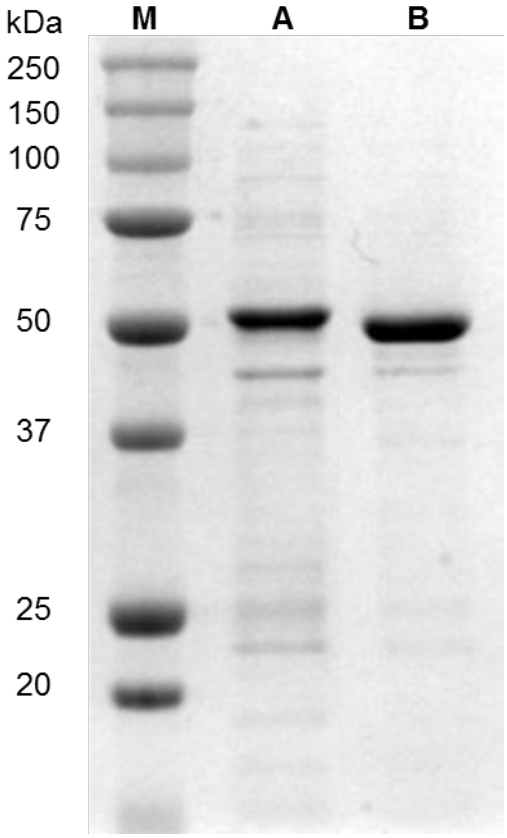


Figure 88 SDS-page picture of EgtB variants after Ni-NTA affinity chromatography. **M** - Molecular weight marker; **A** - $EgtB_{Th2}$; **B** - $EgtB_{Y377F}$.

Sequence of $EgtB_{Th2}$:

GSSHHHHHHSSGLVPRGSHMGVAVPHRAELARQLIDARNRTLRLVDFDDAELRRQYDPLMSPLVWDLAHIG
QQEELWLLRGGDPRRPGLLEPAVEQLYDAFVHPRASRVHLPLLSPAQARRFCATVRSAVLDALDRLPEDAD
TFAFGMVVSHEHQHDETMLQALNLRSGEPLLGSGTALPPGRPGVAGTSVLVPGGPFVLGVDLADEPYALDN
ERPAHVVDVPAFRIGRVPVTNAEWRAFIDDGGYRQRRWWSDAGWAYRCEAGLTAPQFWNPDGTRTRFGHVE
DIPPDEPVQHVTYFEAEAYAAWAGARLPTEIEWEKACAWDPATGRRRRYPWGDAAPTAALANLGGDALRPA
PVGAYPAGASACGAEQMLGDVWEWTSSPLRPWPGFTPMIYQRYSQPFFEGAGSGDYRVLRGGSWAVAADIL
RPSFRNWDHPIRRQIFAGVRLAWDVDRQTARPGPVΓGC

m/z ($EgtB_{Th2}$): calc.: 51131 Da, measured: 51130 Da.

$\varepsilon 280(EgtB_{Th2}) = 111380\ M^{-1}cm^{-1}$

Sequence of $EgtB_{Y377F}$:

GSSHHHHHHSSGLVPRGSHMGVAVPHRAELARQLIDARNRTLRLVDFDDAELRRQYDPLMSPLVWDLAHIG
QQEELWLLRGGDPRRPGLLEPAVEQLYDAFVHPRASRVHLPLLSPAQARRFCATVRSAVLDALDRLPEDAD
TFAFGMVVSHEHQHDETMLQALNLRSGEPLLGSGTALPPGRPGVAGTSVLVPGGPFVLGVDLADEPYALDN
ERPAHVVDVPAFRIGRVPVTNAEWRAFIDDGGYRQRRWWSDAGWAYRCEAGLTAPQFWNPDGTRTRFGHVE
DIPPDEPVQHVTYFEAEAYAAWAGARLPTEIEWEKACAWDPATGRRRRYPWGDAAPTAALANLGGDALRPA
PVGAYPAGASACGAEQMLGDVWEWTSSPLRPWPGFTPMIFQRYSQPFFEGAGSGDYRVLRGGSWAVAADIL
RPSFRNWDHPIRRQIFAGVRLAWDVDRQTARPGPVΓGC

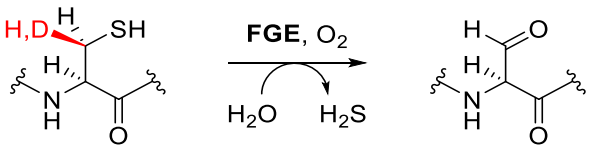
m/z ($EgtB_{Y377F}$): calc.: 51327 Da, measured: 51358 Da (mass difference of 32 Da due to oxidation)

Ɛ280($EgtB_{Y377F}$) = 109890 $M^{-1}cm^{-1}$

4 Enantioselectively Deuterated Peptide Substrates as Probes for the Formylglycine Generating Enzyme

Abstract

The formylglycine generating enzyme (FGE) catalyzes the oxidation of a cysteine to the unique aldehyde-containing amino acid formylglycine (fGly), a key functional residue in the active site of sulfatases. With the help of a copper cofactor, this enzyme harnesses the reactive potential of molecular oxygen in order to achieve C-H bond activation. In light of the electrophilic nature of aldehydes, FGE has become a staple workhorse for biorthogonal conjugation reactions between engineered proteins and a wide array of substrates. In order to investigate the catalytic mechanism and the active site geometry of the enzyme-substrate complex, we have synthesized a series of isotopologue peptide substrates, carrying either hydrogen or deuterium as shown in the example below. With these in hand, we have established that there is a significant, temperature-independent kinetic isotope effect (KIE) on C-H bond cleavage with contributions from primary and secondary KIEs. On the one hand this has confirmed that C-H bond cleavage is the rate determining step in the catalytic cycle, on the other hand the stereoselectivity of the reaction has provided valuable insights into the geometry of this step. This in turn has helped us to construct a novel model of the active site of this unique enzyme with all substrates bound, as well as to propose a catalytic mechanism which is consistent with all our previous findings.



4.1 Introduction

4.1.1 Formylglycine Generating Enzyme

Sulfated biomolecules are a large and diverse class of compounds and come in a variety of flavors in all living organisms. Ranging from small molecules such as steroid sulfates in humans[188] to large polymers like sulfated polysaccharides in marine algae[189] to signaling oligosaccharides in symbiotic bacteria,[190] sulfated molecules are truly omnipresent in nature. Sulfatases, along with sulfotransferases, are the key enzymes in sulfate metabolism. Their role is to cleave sulfate groups and can achieve this either by a hydrolytic or an oxidative mechanism.[191] The largest subgroup of sulfatases uses the former method and contain a key functional residue in the active site called formylglycine (fGly).[192] This unusual amino acid derivative is introduced in post-translational modification in the endoplasmatic reticulum (ER) in eukaryotes or in the cytosol of bacteria by the formylglycine generating enzyme (FGE).[193] It is most often generated from a highly conserved cysteine residue in the active site of sulfatases.[194] In certain prokaryotic sulfatases it can also be formed from a serine instead of a cysteine.[195,196] In all cases, an *N*-terminal [C/S]XPXR motif is recognized by FGE and the thiol or alcohol of the cysteine or serine is oxidized to an aldehyde. In sulfatases, the formyl group of fGly is hydrated to form a geminal diol on C^{β}, which is crucial for catalysis.[197]

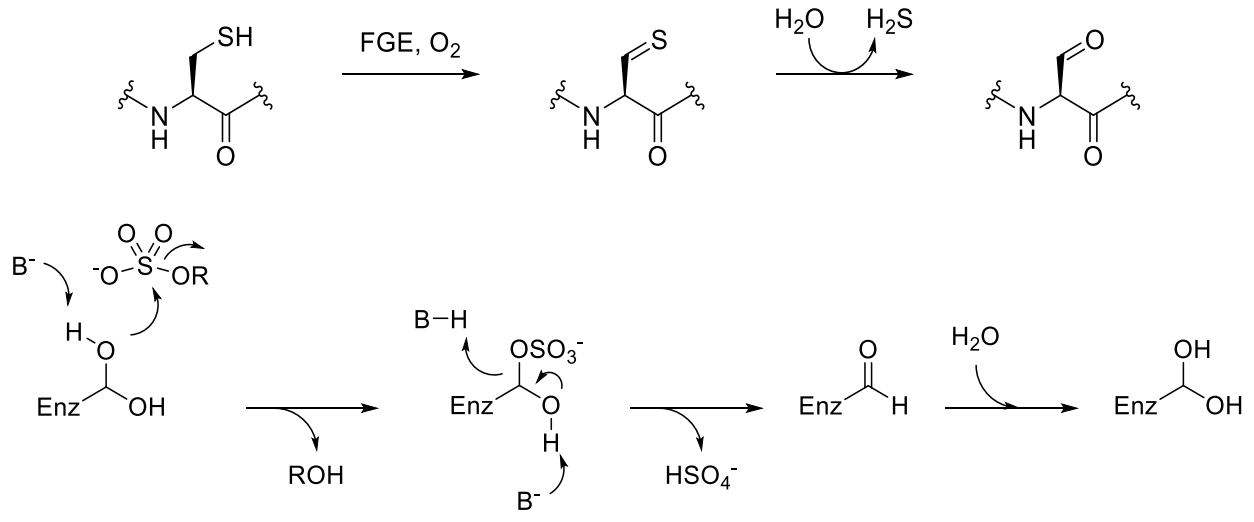


Figure 89 Top: Schematic view of the generation of formylglycine from cysteine by FGE. **Bottom:** Hydrolytic cleavage of a sulfate ester catalyzed by the fGly residue in sulfatases.

Mutations in the gene encoding FGE (*SUMF1*) was shown to be the basis for a severe pathological condition termed multiple sulfatase deficiency (MSD).[198,199] This rare orphan disease with a patient

population of only one in one million individuals[200] can come in different degrees of severity and the onset is usually at a very young age and often fatal.[201] Differences in the mutation pattern determine the residual activity of fGly and the resulting clinical phenotype.[202]

4.1.2 Bioconjugation Applications

Since aldehydes are not a part of the array of functional groups present in the 20 canonical amino acid side chains, FGE has become an attractive tool with which to insert such a unique residue into recombinant proteins.[203] Aldehydes are excellent electrophiles and are able to undergo condensation reactions with aminooxy and hydrazide reagents, forming the corresponding oxime and hydrazones respectively, allowing for very specific labelling of any protein containing them. This has been exploited by a number of research groups, aiming to introduce a bioorthogonal functional handle onto a protein of interest. One of the most proficient groups in this field is that of Bertozzi *et al*. After reporting the synthesis of a formylglycine building block suitable for Fmoc-based solid phase peptide synthesis (SPPS),[204] they went on to show that especially membrane-associated and secreted proteins are viable targets for aldehyde tagging, since eukaryotic FGE is localized in the ER.[205,206] Glycosylation of different proteins was also achieved using FGE-mediated fGly introduction into the engineered CXPXR sequence. Two examples are the human growth hormone (hGH) and one of the two distinct fragments of IgG, named the crystallizable fragment (Fc). The latter is a homodimeric glycoprotein, each monomer of which consists of two constant immunoglobulin domains from the C-terminal half of one heavy chain.[207] Introduction of the fGly-residue in these proteins was followed by oxime formation with aminooxy glycans, producing site-specifically glycosylated protein variants. These approaches represent alternatives to the natural glycosylation pathways.[208,209] Other research groups have introduced fGly into recombinant proteins to immobilize them on surfaces,[210] create conjugates between DNA and proteins,[211] or in order to attach fluorescent tags.[212] In our own laboratory we have fixated proteins on nanocellulose surfaces exploiting the unique chemistry of FGE.[213]

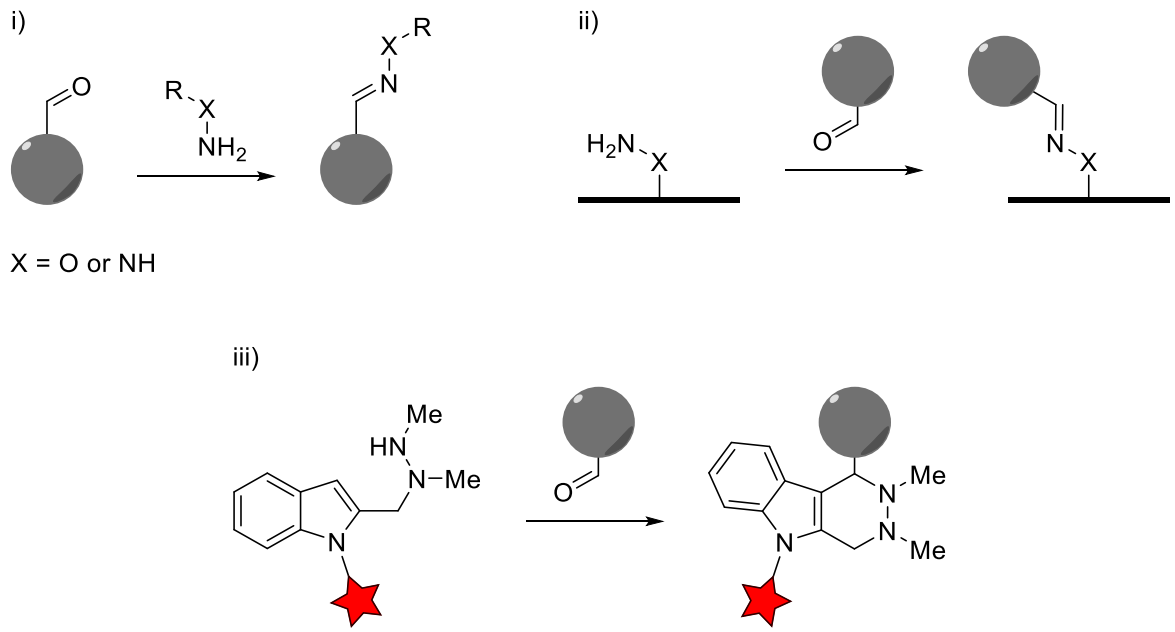


Figure 90 Several strategies for bioconjugations utilizing an aldehyde handle introduced onto a target protein (grey sphere) by FGE; i) condensation of the fGly aldehyde with an aminooxy or hydrazide reagent, forming an oxime- or hydrazine-linkage respectively; ii) immobilization of a target protein on a functionalized surface; iii) Pictet-Spengler type condensation reaction with the newest generation of ligands developed by Redwood Bioscience.

Due to the keen interest of the pharmaceutical industry for antibody-drug conjugates (ADC),[214] antibodies are especially attractive targets for fGly introduction and subsequent labelling.[215] However, since oximes and hydrazones are prone to hydrolysis at physiological pH, a more stable chemical linkage was developed based on the Pictet-Spengler reaction of aldehydes with tryptamine nucleophiles.[216] The newest generation of these reagents[217] has been used to create ADCs with cytotoxic payloads located in different areas of the respective antibody, resulting in different *in vivo* efficacies and pharmacokinetic behaviors in rhodents.[218] This technology, marketed under the name SMARTag™, has recently been acquired by the pharmaceutical company Catalent Biologics and is currently being employed in pre-clinical trials.[219]

4.1.3 FGE Mechanism

FGE-catalyzed C-H bond cleavage was recognized early on as being oxygen-dependent.[176] The question of how the enzyme is able to activate oxygen to a reactive state however remained a mystery. Reactive forms of oxygen utilized by enzymes can either be formed by exciting or activating oxygen from the unreactive triplet to the singlet state or by reduction by an electron-

donor. Originally, FGE was characterized as an enzyme without the need for any apparent inorganic or organic cofactors, besides a thiol reducing agent.[177,220] Only small traces of copper (0.09 mol/mol FGE) were detected and supplementation with additional Cu^{II} inhibited FGE activity.[176] In the meantime, our research group and others have established that copper is necessary for full activity and that Cu^{I} is the active species.[221–223] Most recently X-ray crystal structures of FGE from *Thermomonospora curvata* in complex with Ag^{I} or Cd^{II} were solved in our laboratory, for the first time presenting structural evidence of copper binding in the active site.[224] Based upon our results and the current literature we propose the catalytic mechanism depicted in Figure 91.

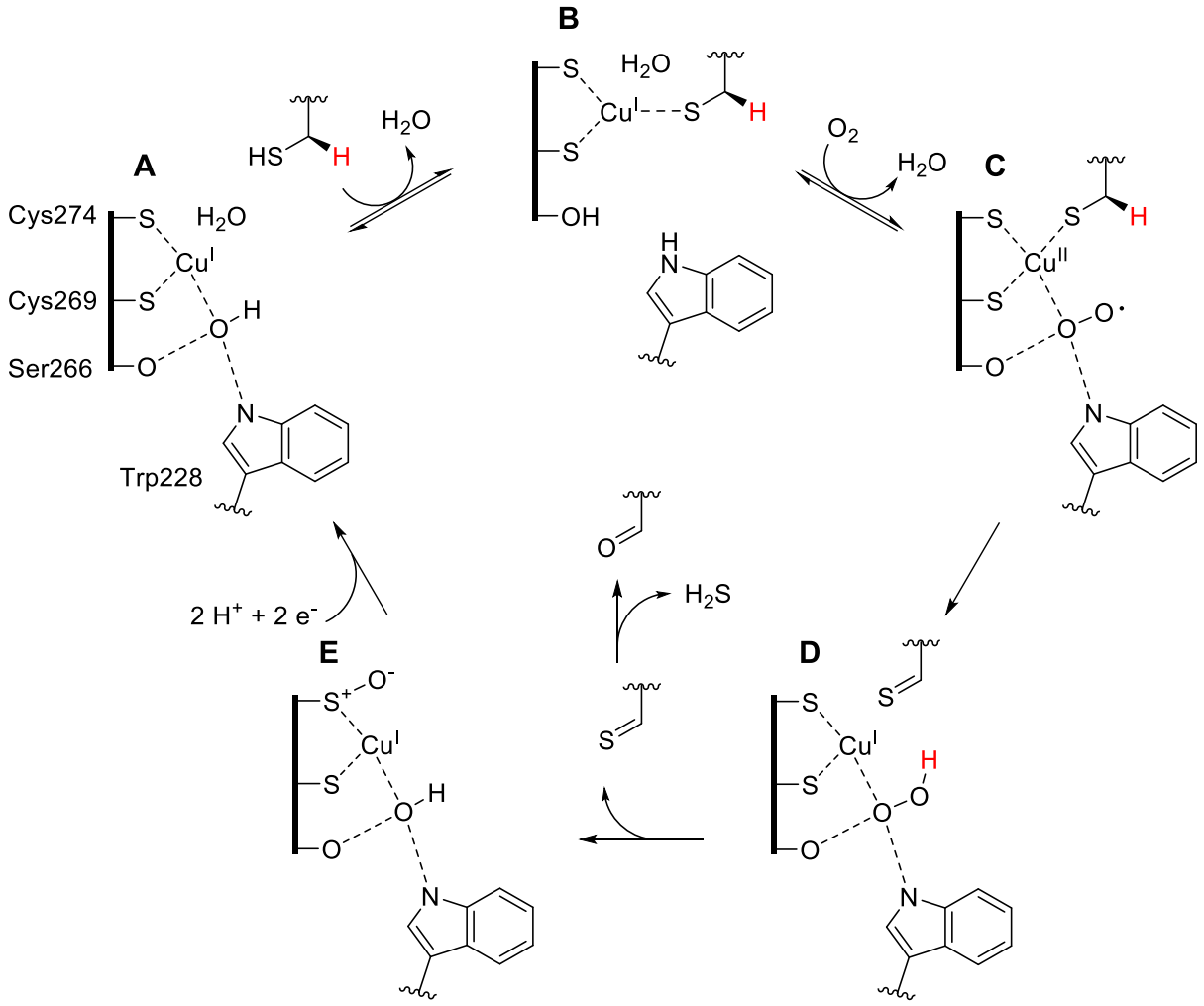


Figure 91 Proposed catalytic mechanism of cysteine oxidation by FGE

The resting state of the enzyme (**A**) binds Cu^{I} and loses one molecule of water upon substrate binding, forming an electron rich complex (**B**). Next, oxygen is bound leading to a reactive Cu^{II} superoxo complex (**C**) followed by hydrogen abstraction and electron transfer to the Cu^{II} center

forming Cu^{I} and a thioaldehyde of the substrate (**D**). This highly instable intermediate hydrolyzes spontaneously to the corresponding fGly-containing peptide, leaving the active site to decay into an inactive oxidized form (**E**), which is reduced by two electrons from an external source back to the reactive resting state (**A**).

In addition to the discussed FGE type, which relies on oxygen-dependent hydrogen abstraction, a second class exists using a 5'-deoxyadenosyl (5'-dA) radical generated from SAM by a redox-active iron-sulfur cluster in the active site to cleave the C-H bond. These anaerobic sulfatase maturating enzymes (anSME) can catalyze the oxidation of both cysteine and serine residues under anaerobic conditions.[225,226] Using a suitable peptide substrate carrying a deuterium label on C^{β} of the target cysteine residue, Berteau *et al*. established that hydrogen abstraction is the rate-limiting step (RLS) with a kinetic isotope effect (KIE) of $k_H/k_D = 5.6$ and that the 5'-dA radical is the abstracting species (Figure 92).[227]

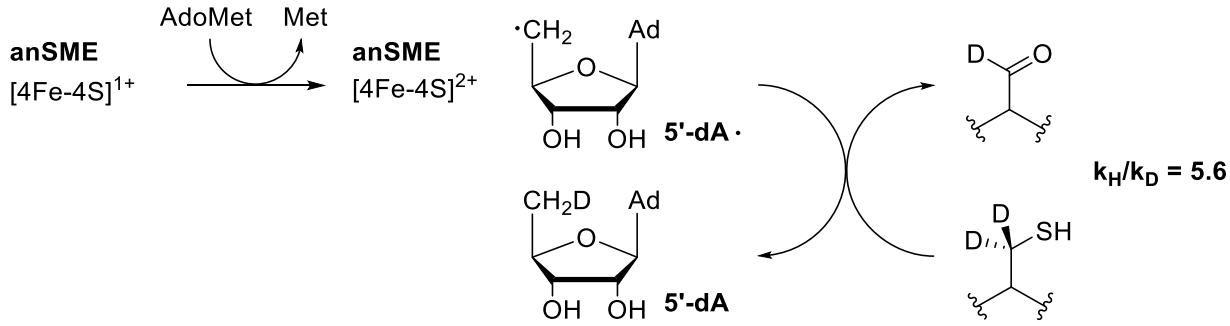


Figure 92 Oxidation of a cysteine-containing peptide by the 5'-dA radical produced by anSME.

By showing that anSME can also convert threonyl peptides to the corresponding ketone products, it was deduced that the *pro-S*-hydrogen is abstracted in the native substrates, as *allo*-threonine containing peptides were oxidized at much lower efficiency.[228]

4.1.4 Kinetic Isotope Effects

Among the various tools enzymologists use to elucidate catalytic mechanisms, kinetic isotope effects are extremely useful, since they not only give information on the nature of a systems RLS, but also of structural features and binding order.[229] Especially when bonds are cleaved, the difference in zero-point energies (ZPE) between two isotopic species can account for significant differences in rates. In the case of hydrogen and deuterium, primary KIEs are classically in the range of 2-10,[230] in extreme cases where hydrogen-tunneling is involved even over 80.[231]

Secondary KIEs are much lower, usually in the range of 0.7-1.4, and are observed during rehybridization of bonds e.g. $sp^3 \rightarrow sp^2$.[232] These secondary KIEs are associated with changes in the vibrational frequencies of the bonds in the donor and acceptor. Changes from high to low frequency will give rise to a normal KIE (>1), while the opposite results in inverse KIEs (<1) as shown in Figure 93. When hydrogen is abstracted from a methylene group and both hydrogens are substituted for deuterium, a primary and a secondary isotope effect are observed, which are multipicative.[233]

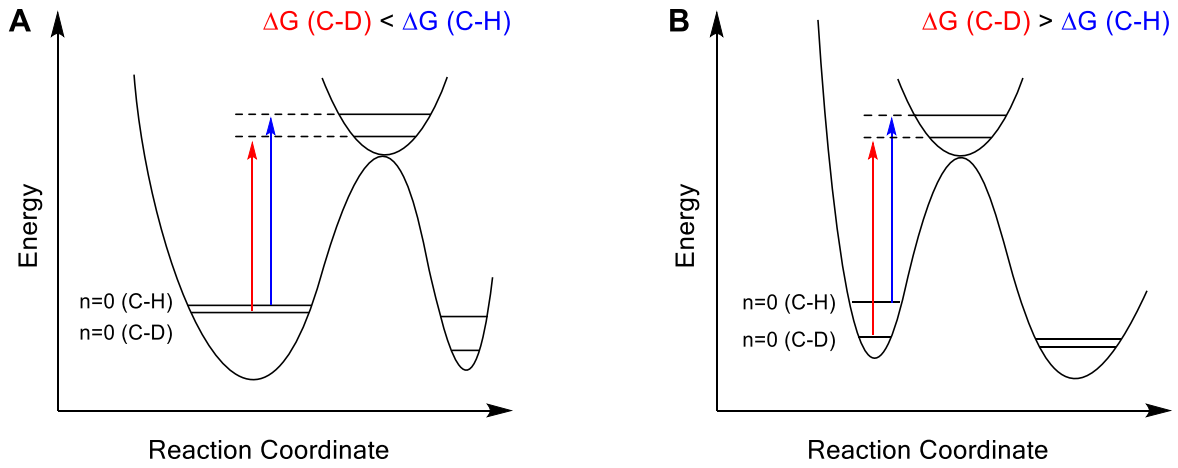


Figure 93 A: Rehybridization of C-H vs C-D bonds, where the vibrational frequencies of the product are higher than in the starting material (e.g. $sp^3 \rightarrow sp^2$). In this case the activation energy for the C-H bond will be higher than for the C-D bond and the KIE will be inverse. **B**: Rehybridization of C-H vs C-D bonds, where the vibrational frequencies of the product are lower than in the starting material (e.g. $sp^2 \rightarrow sp^3$). In this case the activation energy for the C-D bond will be higher than for the C-H bond and the KIE will be normal.

As shown in Figure 94 the observable overall rate of a chemical reaction is determined by the step with the energetically highest transition state, which is the rate limiting step (RLS). It follows that a KIE is directly observable if the isotope sensitive step is rate limiting. For this to be the case it must either occur on the initial RLS or render a different step in the reaction profile rate limiting. If the isotope effect does not increase, or is lower than the RLS itself it, will be obscured by the RLS but may still contribute to the overall rate.[234] In such cases, deconvolution of KIEs becomes vastly more difficult and many more factors must be taken into account.[235]

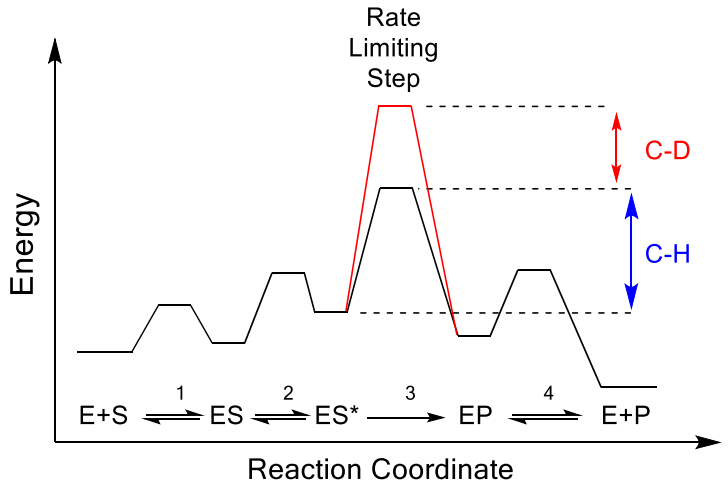


Figure 94 Hypothetical example of an observable KIE on a rate limiting C-H bond cleavage step.

The reaction catalyzed by FGE represents an ideal system to investigate by KIEs. Since C-H bond cleavage typically is not an easy task to accomplish in energetic terms, a KIE on this step is quite likely. Furthermore, since C-H bond cleavage takes place on a methylene group there is the possibility of observing a primary, secondary or cumulative KIE depending on the employed substrate. Indeed, preliminary measurements in our laboratory on FGE with a peptide substrate containing C^{β}-dideuterated *D*/*L*-cysteine revealed a cumulative KIE, establishing that C-H bond cleavage is the RLS in fGly-formation.[236] This prompted us to design compounds and experiments, which would enable us to differentiate between the primary and secondary KIEs and give us more information on the geometry and stability of the enzyme active site.

4.2 Aim of this Chapter

By synthesizing a peptide containing enantioselectively C^{β}-deuterated cysteine, we aimed to establish the primary KIE and determine the precise geometry of hydrogen abstraction. Comparison with a cysteine derivative carrying two deuterium in the C^{β}-position would allow us to resolve the cumulative KIE and thus calculate possible contributions from a secondary KIE, which one can expect to be significant since rehybridization of the carbon center takes place. Distinguishing additional factors such as temperature-dependence of the KIE and the fidelity of hydrogen abstraction could provide further hints towards the stability of the active site. Finally, measurement of the KIE in several mutant variants of FGE would give us insights into whether the in- or decrease of their respective reaction rates is associated with a shift in the RLS.

4.3 Results and Discussion

4.3.1 Synthesis of Enantioselectively Deuterated Cysteine

Enantioselectively deuterated cysteine was synthesized following published protocols, which start from *L*-cysteine and result in a deuterium label in the pro-*R* position.[237–239] If this hydrogen is abstracted by FGE a primary KIE should be observed, if instead the pro-*S* hydrogen is abstracted a secondary KIE would be observed instead.

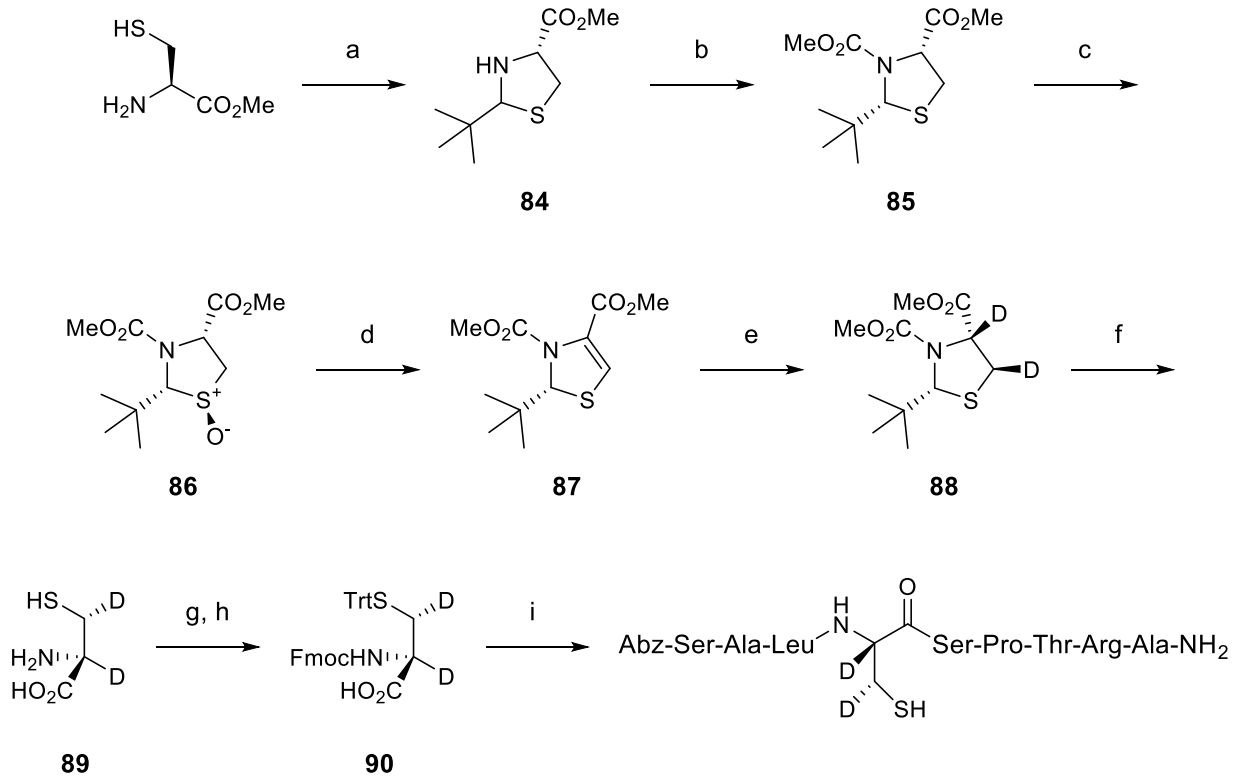


Figure 95 Reagents and conditions for the synthesis of the enantioselectively deuterated substrate and incorporation into the peptide substrate. (a) Pivalaldehyde, Et_3N, pentane, 50 °C, 24 h, 99 %; (b) methyl chloroformate, K_2CO_3, acetone, 60 °C, 18 h, 99 %; (c) H_2O_2, AcOH, 12 h, 89 %; (d) TBDMS-triflate, Et_3N, 72 h, 50 %; (e) D_2, 10 % Pd/C, d_4-AcOD, 144 h, 96 %; (f) 6 M HCl, 110 °C, 72 h, 63 %; (g) Trt-OH, TFA, 2 h, 55 %; (h) Fmoc-OSu, Et_3N, MeCN, H_2O, 3 h, 75 %; (i) SPPS: Rink Amide Resin, HCTU, DIPEA, DMF, 4 %.

Starting with *L*-cysteine, a cyclization is carried out with pivaldehyde, introducing the *tert*-butyl group in nearly quantitative yield. The product **84** is formed as a mix of diastereomers, which is cleanly converted into the desired conformation upon acylation of the amine by methyl chloroformate, again in nearly quantitative yield.[240] Formation of the single product **85** with the *tert*-butyl group in the proper conformation occurs via a ring opening mechanism,[241] as has also been

reported in penicillin derivatives[242] and similar compounds.[243] The configuration of **85** was confirmed by NOEs in the ^{1}H-NMR spectrum shown in Figure 107 in **Section 4.6.1**.

Sulfoxidation to **86** was then performed with hydrogen peroxide in high yield and stereoselectivity.[244] The following Pummerer rearrangement[245,246] to **87** was carried out using TBDMS triflate[247] at a medium yield of 50 %. At this point hydrogenation was performed using deuterium gas and palladium on activated carbon as catalyst. Deuterated acetic acid was used as the solvent, since non-deuterated solvent resulted in deuterium-hydrogen exchange between liquid and gas-phase and incorporation of unwanted hydrogen into the product **88**. Under optimized conditions <5 % hydrogen was incorporated into the product as can be seen in Figure 96. Very long reaction times were necessary for this reaction to run to completion. Subsequent opening of the thiazolidine ring and deprotection of the amine and carboxylic acid was carried out in one step under harsh conditions of 6 M HCl and reflux over three days to give the desired labelled cysteine **89** as a free amino acid in satisfying yield of 63 %. Introduction of protecting groups suitable for SPPS – Trt on the thiol and Fmoc on the amine to give **90** – constituted the final steps of this synthesis before introduction into the peptide.

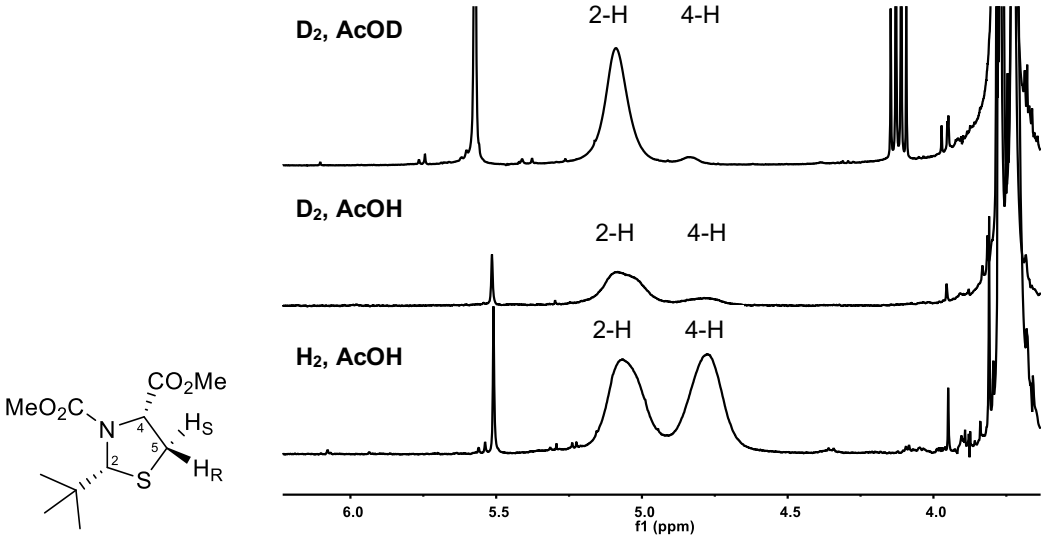


Figure 96 ^{1}H-NMR spectrum of **Xe**. Conditions: D_2, AcOD (top); D_2, AcOH (middle); H_2, AcOH (bottom).

The peptide sequence (Abz-SALCSPTRA-$CONH_2$) was chosen based on the recognition sequence of a sulfatase from *Mycobacterium smegmatis*[248] and included an *N*-terminal anthranillic acid (Abz) group as a chromophore. Synthesis was carried out using standard Fmoc solid phase peptide synthesis (SPPS), followed by purification by HPLC and characterization by ESI-HRMS and HPLC.

Besides the enantioselectively deuterated peptide, three additional peptides were synthesized for kinetic experiments. In the case of **Pep3** and **Pep4**, commercially available *D/L*-cysteine and *D/L*-[3,3-$^2H_2$]-cysteine respectively were protected in the same manner as **89** before incorporation into the respective peptides.

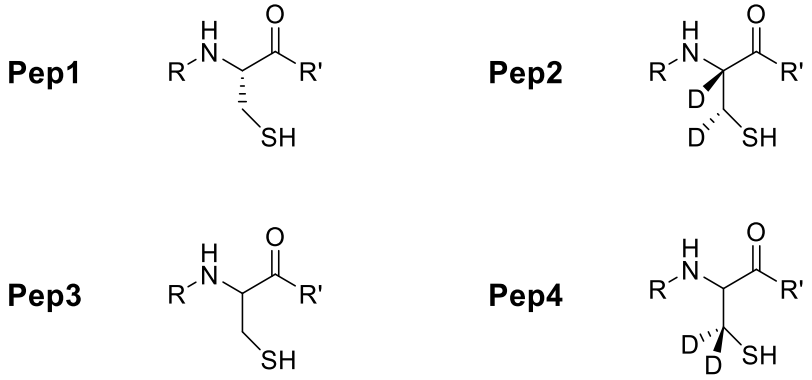


Figure 97 The relevant cysteine residue in the model peptides **Pep1-4**, where R = SerAlaLeuAbz-NH_2 and R' = SerProThrArgAla-$CONH_2$.

The enantiomeric purity of **Pep1** (100 % *L*-Cys) and **Pep2** (96 % *L*-Cys) was determined by RP-HPLC, as the peptides containing racemic *D/L*-cysteine could be separated. Incorporation of an *N*-terminal anthranillic acid (Abz) group as a chromophore insured that the starting materials as well as the product of the reaction were detectable at concentrations below 1 µM by HPLC (see also next section).

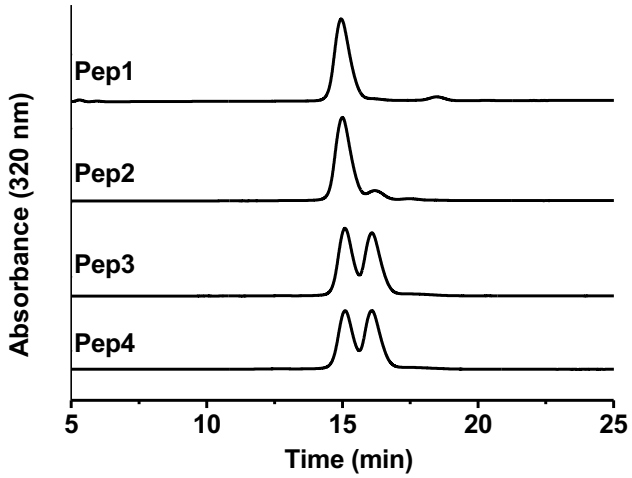


Figure 98 RP-HPLC traces for **Pep1**-**Pep4** at 320 nm, the peak at t_R = 15 min is the peptide containing *L*-Cys, the peak at t_R = 16 min contains *D*-Cys.

4.3.2 HPLC Based Activity Assay

Recombinant production of FGE was carried out in *E. coli* cell cultures and the protein was purified by Ni-NTA affinity chromatography. The mass of the enzyme was confirmed by HRMS and its homogeneity assessed by SDS-PAGE.[249] The FGE used in all the following assays was from *Thermomonospora curvata*, unless otherwise indicated and was kindly produced and provided by Matthias Knop. This FGE was chosen due to its thermostability, which can be advantageous in terms of production and isolation yield as well as increased resistance towards thermal stress. Additionally, the fact that it shares 58 % sequence identity with the human variant,[250] allows for comparison between the two. For all kinetic measurements we employed an HPLC based activity assay where the aldehyde of fGly formed from cysteine reacts with the *ortho*-amine of the *N*-terminal Abz group, forming a Schiff base under acidic conditions. This was doubly beneficial, since the Schiff base a) has a larger extinction coefficient due to the extended π-system and can thus be detected at lower concentrations and b) facilitates peak separation thanks to its cyclic structure.

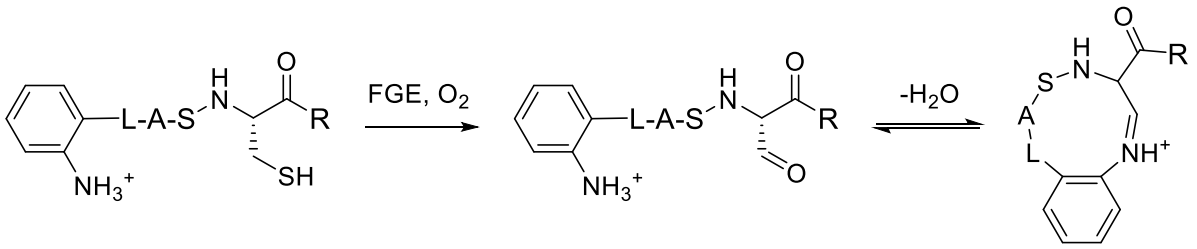


Figure 99 Structures of substrate, product and observed Schiff base in the kinetic assay. R = SerProThrArgAla-$CONH_2$

Assays were run in 50 mM phosphate buffer containing DTT as reducing agent and 1 equivalent of copper relative to the enzyme. Samples were taken at different time points quenched in 7.5 M urea, containing 2 % TFA and directly subjected to analysis by HPLC. From the peak area we could calculate the concentration of product and plot it as a function of time, the slope of this linear function, divided by the enzyme concentration, was equal to the rate of the reaction.

4.3.3 C-H Bond Cleavage by FGE Displays a KIE with Large Contribution from a Secondary KIE

With our array of differently deuterated peptides we sought out to determine the primary and cumulative KIEs by comparing **Pep1** with **Pep2** and **Pep3** with **Pep4** respectively. Dividing the cumulative by the primary KIE gives an estimate for the secondary KIE.[251] Using the kinetic assay described above, all four peptides were incubated with FGE, the rates of the respective reactions were measured and the respective KIEs calculated.

Table 17 A: Initial rates of conversion of **Pep1**-**Pep4** by FGE. **B**: Calculated KIEs from the measured rates, the secondary KIE was calculated by dividing the cumulative by the primary KIE.

A

Substrate	rate [s^{-1}]
Pep1	(2.4 ± 0.3) x 10^{-3}
Pep2	(0.7 ± 0.1) x 10^{-3}
Pep3	(2.4 ± 0.2) x 10^{-3}
Pep4	(0.4 ± 0.1) x 10^{-3}

B

KIE		
Prim.	**Cum.**	**Sec.**
3.4 ± 0.2	5.6 ± 0.2	1.7 ± 0.2

Reaction conditions: 50 µM **Pep1-4**, 2 mM DTT, 50 mM EDTA, 50 mM NaCl, 50 mM phosphate pH 8, 0.5 µM FGE (previously activated with 1 eq $CuSO_4$) at 30 °C.

Comparison of the reaction rates of **Pep1** and **Pep2** resulted in a primary KIE of k_H/k_D = 3.4 ± 0.2, demonstrating that H^R is the abstracted hydrogen.[222] The cumulative KIE was in turn calculated from the reaction rates of **Pep3** and **Pep4**, resulting in a larger effect of k_H/k_D = 5.6 ± 0.2. Such a large difference between the primary KIE and the cumulative KIE indicate considerable contributions to the latter from α-secondary KIEs in the rate limiting step. From our calculation we determined a large secondary KIE of 1.7 ± 0.2, which is above the semiclassical limit of 1.44.[232,252] A secondary effect of this magnitude is unusual, but the fact that it is a normal and not inverse KIE is compatible with rehybridization from $sp^3 \rightarrow sp^2$.[230] Although inflated secondary isotope effects have been linked to H-tunneling in enzymatic reactions, in these cases one also expects a primary KIE above the semiclassical limit.[253]

4.3.4 The KIE is an Effect on k_{cat} and not K_M

To determine whether hydrogen abstraction is fully or only partially rate limiting one can test if the KIE is observed on k_{cat} or k_{cat}/K_M. If C-H bond cleavage is only partially rate-limiting and takes place before the RLS, a small isotope effect is observable on k_{cat}/K_M, but not on k_{cat}. At conditions where the enzyme is saturated with substrate, an early C-H bond cleavage step would be less significant.[254] We measured the initial rates of the reaction at varying concentrations of **Pep1** and **Pep2** and calculated the respective KIEs, which remained unchanged at each concentration.

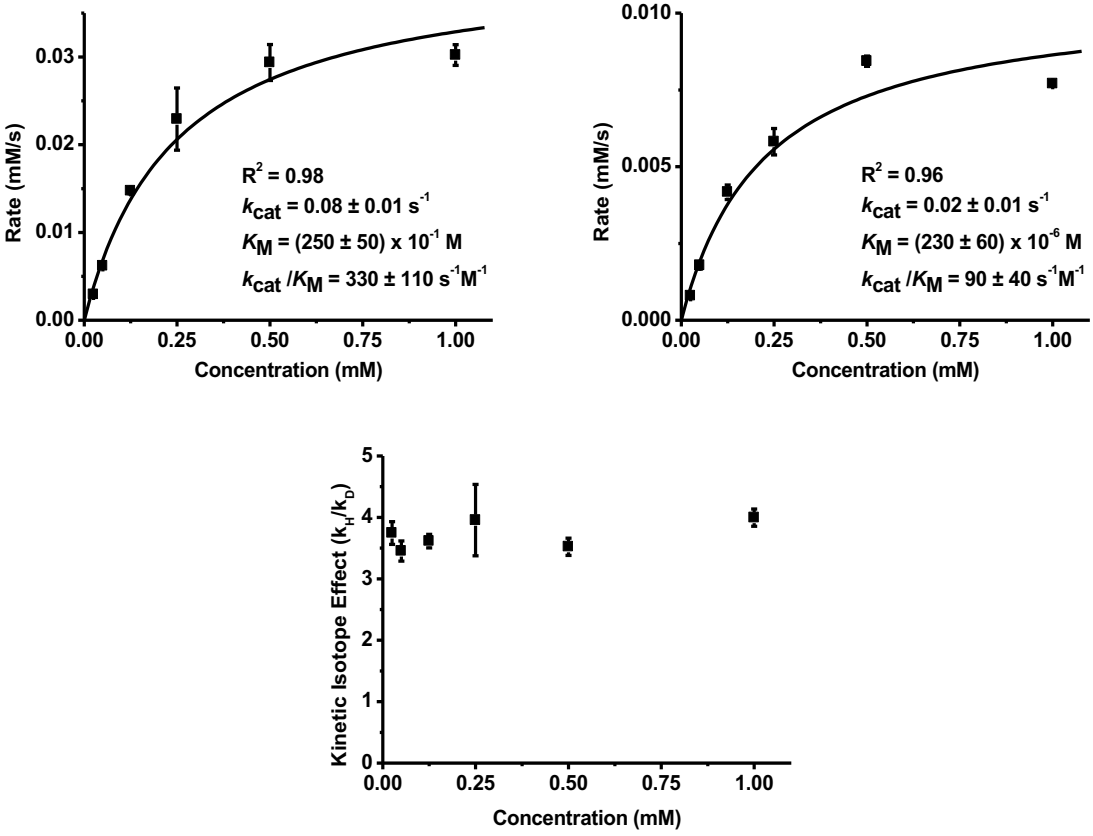


Figure 100 Michaelis-Menten parameters for the **Pep1** (top left) and **Pep2** (top right). Concentration dependence of KIE (bottom), Reaction conditions: 25 - 1000 µM **Pep1** or **Pep2**, 4 mM DTT, 50 mM EDTA, 50 mM NaCl, 50 mM phosphate pH 8, 0.5 µM FGE_{wt} (previously activated with 1 eq $CuSO_4$) at 30 °C.

The KIE is observable on k_{cat} as well as k_{cat}/K_M with an average of k_H/k_D = 3.7 ± 0.2, meaning that C-H bond cleavage is fully rate limiting. On K_M itself no effect is seen as one would expect, since the isotopic labelling should not have an observable effect on substrate binding.

4.3.5 The KIE Remains Unchanged in a Series of Mutant FGE Variants

Several mutant variants of FGE have been produced in our laboratory and we strove to show whether the increase or decrease in their respective rates is associated with a shift of the RLS. In case a different step than hydrogen abstraction, such as substrate-binding, becomes rate limiting one would expect the observed KIE to be reduced or even invisible. To address this question three mutants were tested with **Pep1** and **Pep2** in order to determine and compare the primary KIE to the wild type. In the first mutant, termed FGE_{4C}, all four non-active site cysteines are mutated to alanine or serine residues (C187A, C231A, C284S and C298A). In the presence of copper this first variant is roughly twice as active as the wild type, without copper it is even 10-fold more active.[222] The precise reason for this increase in activity has not been conclusively elucidated as of yet, but it has been shown that the point mutation C187A is largely the cause, resulting in an elevated k_{cat} and a reduced K_M.[249] The second mutation targeted an active site serine residue (S266A) in the immediate vicinity of the two active-site cysteines (C269 and C274). This led to a 60-fold decrease in activity, since Ser266 is thought to be involved in oxygen binding and activation (see **Section 4.1.3** and Figure 91). The last variant contained the mutation W84R, where a tryptophan residue involved in substrate binding is substituted by an arginine. This mutant was produced as part of an effort to create an orthogonal enzyme/substrate pair by inverting ionic bonds, allowing stepwise labelling of one protein with two different probes.[255] Its activity is reduced by roughly 10-fold compared to the wild type enzyme.

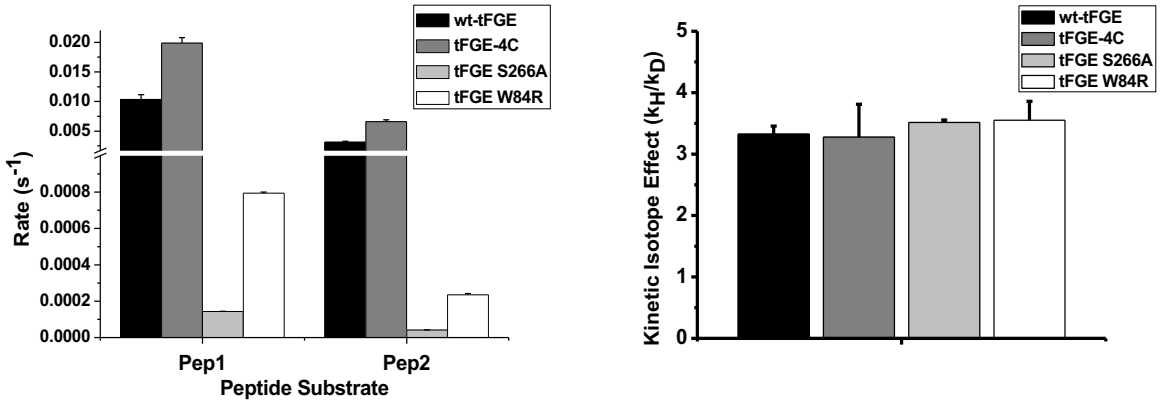


Figure 101 Left: Reaction rates measured on different mutants in comparison with the wild type enzyme with the substrates **Pep1** and **Pep2**. **Right:** calculated primary KIEs for different mutants in comparison with the wild type enzyme. Reaction conditions: 50 µM **Pep1** or **Pep2**, 4 mM DTT, 50 mM EDTA, 50 mM NaCl, 50 mM phosphate pH 8, 0.5 µM FGE (previously activated with 1 eq $CuSO_4$) at 30 °C.

While the values for the initial rate constants vary greatly between the different mutants, the KIEs are all within the same range of 3.4 ± 0.3. Hydrogen abstraction remains rate determining in each variant, allowing for the conclusion that the increase (FGE_{4C}) or decrease (FGE_{S266A} and FGE_{W84R}) is not associated with a different step becoming rate limiting, such as substrate binding or oxygen activation.

4.3.6 The KIE is Temperature Independent

In order to investigate a possible temperature dependence of either KIE, we determined the rate of conversion for each peptide substrate at different temperatures between 0 °C and 55 °C.

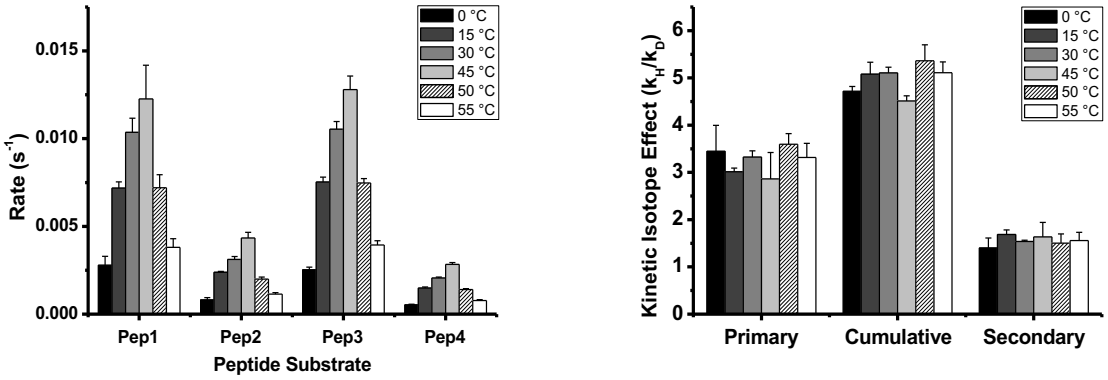


Figure 102 Temperature dependence of reaction rate (left) and kinetic isotope effect (right), Conditions: 50 µM **Pep1-4**, 2 mM DTT, 50 mM EDTA, 50 mM NaCl, 50 mM phosphate pH 8, 0.5 µM wtFGE (previously activated with 1 eq $CuSO_4$).

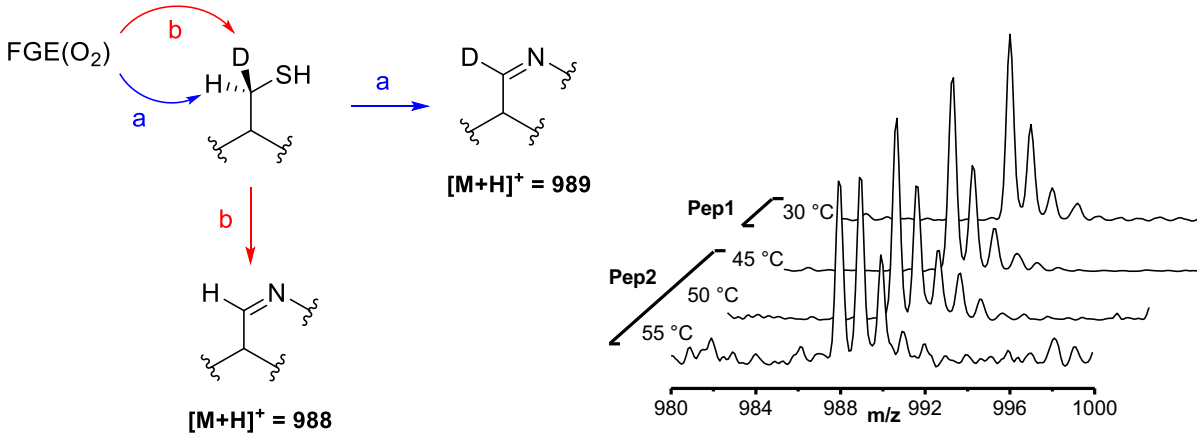
The rate of the reaction reaches a maximum at 45 °C before decreasing again approaching the highest temperature measured at 55 °C. Higher temperatures were not measured as the melting temperature of FGE was determined to be at 68 °C. Primary, cumulative and secondary KIEs remained constant over the full temperature range, displaying remarkable stability of the enzyme even at elevated temperatures. On average the primary KIE is 3.3 ± 0.5 and the cumulative is 5.0 ± 0.4, the calculated secondary KIE is 1.6 ± 0.2, which again is at the upper semiclassical limit of 1.1-1.4.[252]

4.3.7 Temperature Dependence of Selectivity

After establishing that hydrogen abstraction remains the RLS even at elevated temperatures, we aimed to test whether the selectivity of the enzyme was unperturbed. If the active site becomes less compact at temperatures close to the melting point, one would expect that the enzyme might start to abstract the *pro*-*S*-hydrogen instead of the usual *pro*-*R*-hydrogen (see Figure 103). This in turn would lead to an observable difference in mass, which can be quantified by comparing the relative peak intensities of the mass spectra since it is safe to assume that the ionizability of the deuterated and non-deuterated product are identical.[256] It is however noteworthy that racemization takes place on Cα of the formylglycine residue after cyclization and formation of the Schiff base at pH 2 (see Figure 99).

Reaction mixtures containing either **Pep1** or **Pep2** were incubated at different temperatures with FGE_{wt} overnight, after which the reaction mixtures were desalted by HPLC and product-containing fractions lyophilized before recording ESI-MS spectra. Up until 45 °C, where the reaction rate reaches a maximum, no difference in the hydrogen abstraction pattern could be detected when compared to control reactions at 15 °C and 30 °C. At temperatures of 50 °C and 55 °C however, a clear increase of the mass peak corresponding to pro-*S*-hydrogen abstraction was observable.

After comparison of peak intensities, a ratio of almost 1:1 was calculated for the two intensities. This means that after subtracting the base value for the m/z = 988 peak (obtained from **Pep1** at 30 °C), in 14 % and 32 % of **Pep2** the pro-*S*-hydrogen was abstracted at 50 °C and 55 °C respectively. The enzyme active site thus becomes increasingly instable at temperatures above 50 °C. Although the KIE remains unchanged even at 55 °C, binding of the substrate becomes less tight, allowing for abstraction of the pro-*S*-hydrogen.



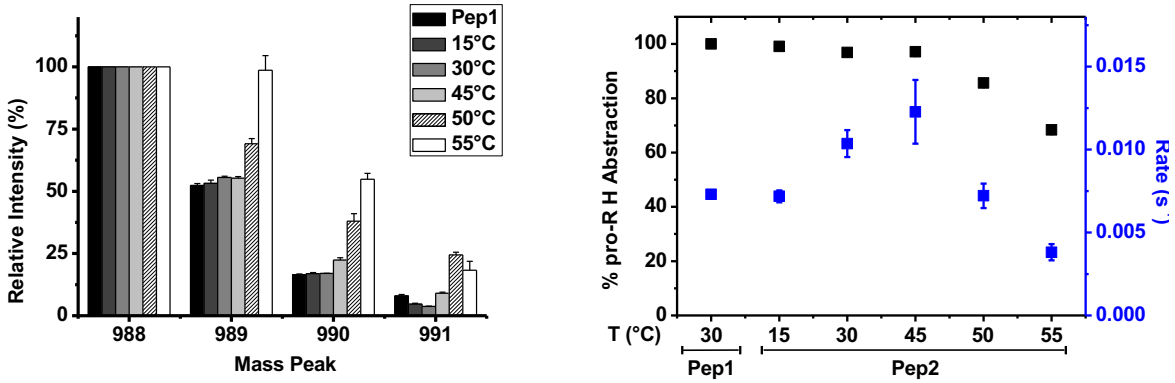


Figure 103 Top left: Two possible hydrogen abstractions from **Pep2** (a and b) with the corresponding products, which are formed in the reaction and their masses after cyclization. **Top right:** ESI-MS spectra of **Pep1** at 30 °C and **Pep2** at 45-55 °C, the range in which the hydrogen abstraction from the latter peptide becomes less selective. **Bottom left:** ratio of the mass peaks m/z = 988 to 989 in the measured temperature range using **Pep1** at 30 °C as a standard. **Bottom right:** relative intensity of the respective mass peaks of **Pep2** at increasing temperature, using **Pep1** at 30 °C as a standard. Reaction conditions were identical to those described above, apart from letting the reactions run to completion overnight, followed by desalting using a short RP-HPLC method and subsequent lyophilization.

These results underline that the large contribution from α-secondary KIEs is not inflated due to the enzyme cleaving the C-H^S bond instead of the energetically less favorable C-D bond in Pep1. If this had been the case the observed primary KIE would have been significantly reduced and the calculated α-secondary KIE an overestimate.

4.4 Mechanistic Implications

4.4.1 Geometry of the FGE Active Site

As mentioned in **Section 4.1.3**, FGE was initially characterized as a cofactor-free enzyme and to date all solved crystal structures seemed to support this hypothesis.[177,220,257] Subsequently however, it was shown by biochemical methods that this is not the case and FGE is indeed copper-dependent.[221–223] Most recently, our laboratory has been able to solve crystal structures of FGE from *Thermomonospora curvata* in complex with Ag^{I} and Cd^{II} (PDB: 5NXL and 5NYY respectively), both d^{10} systems mimicking copper in different coordination spheres.[224] The Ag^{I}-complex shows

nearly linear coordination between Cys269-Ag^{I}-Cys274 and most likely represents the resting state of the enzyme. On the other hand the Cd^{II}-complex displays tetrahedral coordination between Cys269, Cys274, a water molecule and an acetate molecule, giving a picture closer to the substrate binding mode. In particular, the metal-coordinating oxygen of the acetate ligand is in the exact same position as the substrate sulfur atom, when superpositioned with the crystal structure of human FGE bound with a short substrate analog (PDB 2AIJ).[220] Replacement of the abovementioned, metal-coordinated water molecule by oxygen places the latter in the direct vicinity of the pro-*R* hydrogen on C^{β} of the peptide substrate, which is consistent with the determined KIEs.

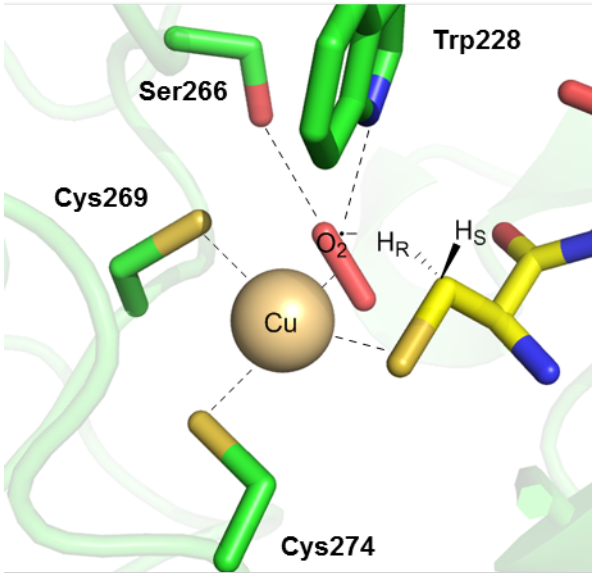


Figure 104 Model of the active site of $FGE_{curvata}$ (PDB: 5NYY)[224] as the Cu^{II}-superoxo-complex, with the peptide substrate analog (yellow) modelled based on the crystal structure of FGE_{human} (PDB 2AIJ). The proposed copper-center is coordinated by Cys269 and Cys274, the sulfur atom of the peptide substrate and activated oxygen (red). The latter is further coordinated by Ser266 and Trp228.

4.4.2 Possibility of a Single Electron Transfer Step in the Catalytic Mechanism of FGE

In galactose oxidase, a member of the radical copper oxidase family, a redox-cofactor consisting of Cu^{II} and tyrosylcysteine (TyrCys272) catalyzes the oxidation of galactose and other primary alcohols to the corresponding aldehydes as shown in Figure 105.[251,258] The substrate alcohol binds to the active site copper (**A**), where after proton transfer to the catalytic base Tyr495 (**B**) a single electron transfer (SET) step from the substrate to the metal center (**C**) and a hydrogen atom transfer from C^{β} of the substrate to the TyrCys cofactor (**D**) are thought to take place. Initial experiments by Whittaker *et al.* identified a large cumulative KIE of 22.5, which was proposed to

arise from a combination of factors.[259] If the abovementioned SET were to exhibit an equilibrium isotope effect, a fractionation factor of 0.8 per deuterium would suffice to correct the primary KIE to a value of 7-10. In subsequent studies primary, secondary and cumulative KIEs were determined with appropriately labelled substrates, resulting in observed secondary KIEs (1.19-1.44) ranging towards the upper limit of what theory would predict.[251] While it was speculated that these values may be indicative of H-tunneling, it was also considered that the cause could lie in a partially rate-limiting SET during substrate oxidation. Since oxidation of an alcohol to an alkoxyl free radical also affects the vibrational frequencies of the adjacent methylene group, an equilibrium deuterium isotope effect on electron transfer would contribute to the overall reaction when the methylene hydrogens are substituted for deuterium. Analysis of the data led to the calculation of a fractionation factor of 0.78-0.80 and the conclusion that equilibrium isotope effects may be involved in the overall observed effects, although conclusive evidence on this aspect of catalysis is still pending.

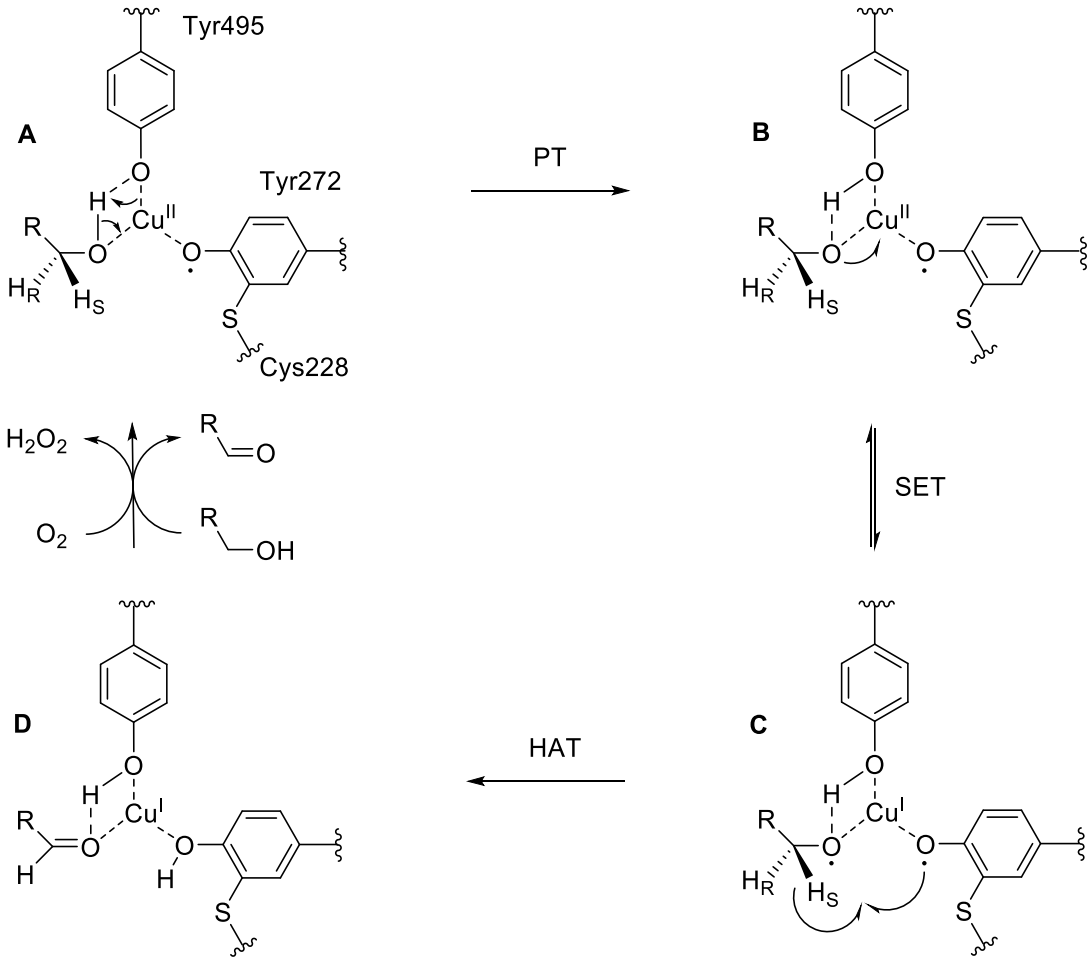


Figure 105 Catalytic mechanism for the oxidation of a primary alcohol to an aldehyde by galactose oxidase, as proposed by Whittaker *et al*. After proton transfer (PT) from the substrate alcohol to Tyr495 (**A** to **B**), an unfavorable single-electron transfer (SET) redox equilibrium generates an alkoxyl complex (**C**). The aldehyde is then formed after hydrogen atom transfer (HAT) to the phenoxyl free radical, which is the rate limiting step (**D**).[251] The oxidation state of copper is restored by one equivalent of dioxygen which reoxidizes the non-radical TyrCys species. In the crystal structure which this model is based on (PDB 1GOG), the copper center is additionally coordinated by two histidine residues (H496 and H581), which are omitted for clarity.[141]

This explanation for an inflated observed secondary KIE however, raises an interesting point concerning our proposed mechanism for FGE. As we too observed an unusually large secondary KIE, it could be that the observed cumulative KIE can be attributed to the primary KIE combined with an isotope fractionation factor (1/β) of 0.8 per deuterium on a prior unfavorable single electron transfer step. A correction by a factor of this magnitude results in a lower secondary KIE of 1.1 ± 0.2, which is much closer to what would be expected for bond rehybridization. This is however only a rough estimate, a precise value for β can only be calculated from the observed secondary isotope effect, something we did not measure directly. If this is indeed the case, the mechanism

shown in Figure 91 (**Section 4.1.3**) would need to be corrected (Figure 106) to account for an SET redox equilibrium as may also be the case in galactose oxidase.

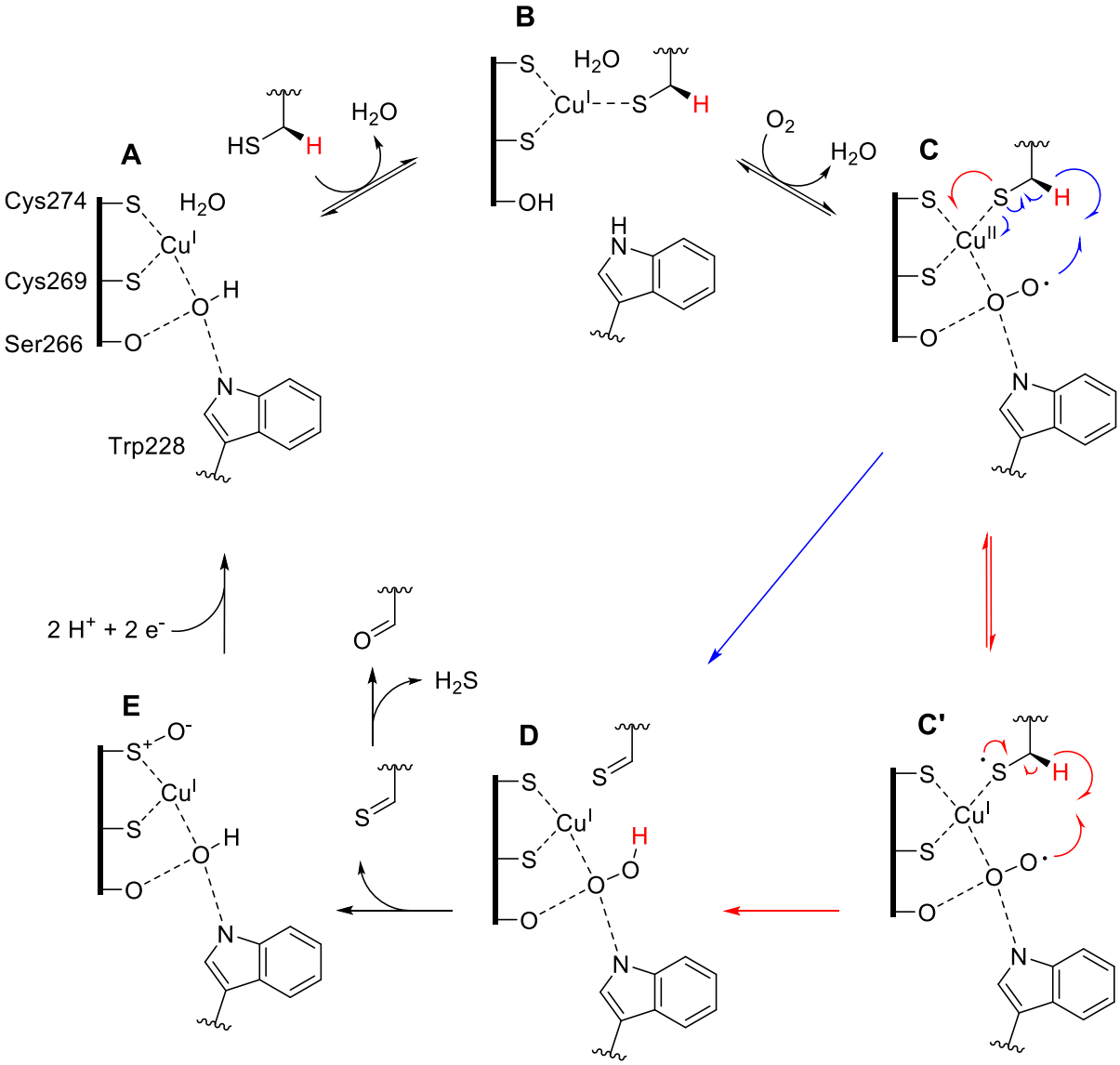


Figure 106 Accomodation of a possible single-electron transfer step in the mechanism of FGE. The blue pathway involves direct abstraction of H^R by the superoxide radical, leading directly from **C** to **D**. The red pathway involves a single electron transfer from the sulfur to the copper center (**C'**), which would result in an equilibrium deuterium isotope effect on this step contributing to the observed KIEs.

Incorporation of such an SET would make sense from an energetic standpoint, as the Cu^{II}-superoxo species is a rather weak hydrogen-abstracting species.[260] Weakening of the C-H bond by transferring one electron onto it might thus be a strategy of FGE to activate the substrate in addition to activating oxygen. In order to test this hypothesis one could envisage incorporating C^{β}-

dideuterated cysteine into the enzyme itself. The presence of deuterium in this position of Cys269 and Cys274 would destabilize a putative radical on the substrate cysteine, as the radical would be expected to be delocalized over all three copper-bound thiolates, which in turn would lead to an inverse KIE.

4.5 Conclusion

The KIEs we have measured in this work have been an important part of elucidating the mechanism and the active site geometry of FGE. Precise positioning of the reactive oxygen in the active site is what ultimately insures that hydrogen is abstracted only from the substrate. Although this may seem redundant, since the active site copper is coordinated by two cysteine residues of the enzyme and one of the substrate, all of which are essentially chemically equal, this feat is truly remarkable. All aspects of the modelled enzyme-substrate complex presented herein are fully consistent with the KIEs we have observed in this work. The remarkable thermostability of FGE has been demonstrated, with reduced stereochemical selectivity only at temperatures approaching the melting point of the enzyme. By showing that the KIE is on k_{cat}, we have confirmed that C-H bond cleavage is fully rate-limiting and remains the rate limiting step in a series of mutants with increased as well as decreased catalytic rates. Furthermore, the large values we have obtained for the secondary KIEs have prompted us to speculate whether an unfavorable SET step is involved in the catalytic mechanism. This could explain the observed inflation of the KIE and would also be consistent with our proposed mechanism with minor corrections as in the similar system of galactose oxidase, where comparable effects have been observed. We anticipate that our work presented herein will contribute to the increased understanding of copper enzymes and ultimately pave the way for the design of novel catalysts for organic synthesis.

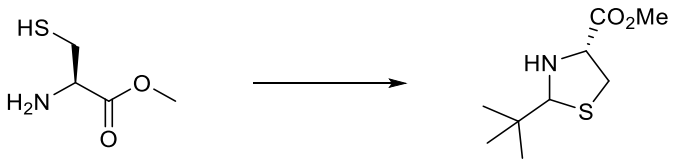
4.6 Experimental

4.6.1 Synthesis

Trityl protection of cysteine derivatives – Procedure C: The cysteine derivative (1 eq), was dissolved in TFA (1 ml per 40 mg of cysteine) and triphenylmethanol (1 eq) was added. The dark yellow solution was stirred for 2 hours at room temperature. The reaction mixture was then cooled to 0 °C and first set to pH 2 with 4 N NaOH (checking with pH paper), then diluted with diethylether (30 ml), set to pH 4 with 1 N NaOH and then to pH 5.5-6 with 10 % sodium acetate (checking with pH pmeter). The resulting precipitate was filtered, washed with Et_2O and dried to yield the product as white solid, which was used in the next step without further purification.

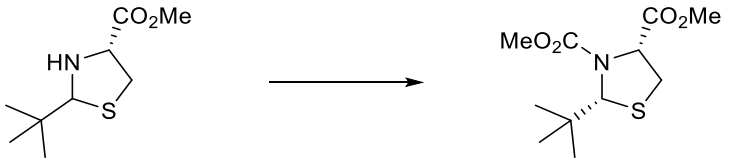
Fmoc protection of cysteine derivatives – Procedure D: The trityl-protected cysteine derivative from Procedure A (1 eq) was dissolved in water (1 ml per 25 mg of cysteine) and triethylamine (2.5 eq). To this was drop-wise added a solution of Fmoc-OSu (1.05 eq) in acetonitrile (1 ml per 50 mg of cysteine) and the colorless solution was stirred for 3 hours at room temperature. The organic solvent was removed *in vacuo* and the milky suspension was set to pH 3.5 with 15 % acetic acid (checking with pH paper) and extracted with 3 x EtOAc. The combined organic layers were dried over Na_2SO_4, concentrated and purified by flash column chromatography (DCM/MeOH 100:0 → 98:2), to give the pure product as colorless solid. Analysis of the pure compounds was limited to +ESI-MS and RP-UHPLC, no ^{1}H-NMR spectra were recorded for the fully protected cysteine derivatives.

Synthesis of (4*R*)-methyl 2-(*tert*-butyl)thiazolidine-4-carboxylate (84).



To a solution of *L*-cysteine methyl ester HCl (3.535 g, 20.59 mmol, 1 eq) in pentane (60 ml) was added pivalaldehyde (4.47 ml, 41.19 mmol, 2 eq) and Triethylamine (3.14 ml, 22.65 mmol, 1.1 eq) and the white suspension was heated to 50 °C under reflux and N_2-atmosphere in a Dean-Stark trap and the reaction was stirred for 20 h. The suspension was filtered, the filter cake was washed with ether and the filtrate was evaporated to dryness to give the desired product (4.368 g, 20.41 mmol, 99 %, purity: 95 %) as slightly yellow oil. **^{1}H NMR (400 MHz, $CDCl_3$)** Main isomer (70 %): *δ* 4.49 (s, 1H), 3.91 – 3.87 (m, 1H), 3.79 (s, 3H), 3.28 (dd, *J* = 10.4, 6.7 Hz, 1H), 2.81 – 2.69 (m, 1H), 1.09 (s, 9H). Second isomer (30 %): *δ* 4.56 (s, 1H), 4.26 – 4.16 (m, 1H), 3.77 (s, 3H), 3.19 – 3.10 (m, 1H), 3.09 – 3.01 (m, 1H), 1.00 (s, 9H). **MS** *m/z* calcd. for $C_9H_{18}NO_2S$ $[M+H]^+$ 204.1, found 204.1 .

Synthesis of (2*R*,4*R*)-dimethyl 2-(*tert*-butyl)thiazolidine-3,4-dicarboxylate (85).



To a solution of **84** (4.187 g, 20.41 mmol, 1 eq) and K_2CO_3 (16.926 g, 122.47 mmol, 6 eq) in dry acetone (35 ml), was added methyl chloroformate (6.32 ml, 81.64 mmol, 4 eq) and the reaction was stirred under reflux at 60 °C under N_2 for 18 h. The organic solvent was evaporated and the residue was treated with water. The aqueous layer was extracted with 3 x EtOAc, the organic fractions were combined, dried over Na_2SO_4 and concentrated to yield **85** (5.652 g, 20.33 mmol, 99 %, purity: 94 %) as clear orange oil. **^{1}H NMR (400 MHz, $CDCl_3$)** δ 5.07 (br s, 1H), 4.78 (br s, 1H), 3.77 (s, 3H), 3.73 (s, 3H), 3.33 – 3.20 (m, 2H), 1.02 (s, 9H). **MS** *m/z* calcd. for $C_{11}H_{20}NO_4S$ $[M+H]^+$ 262.1, found 262.1.

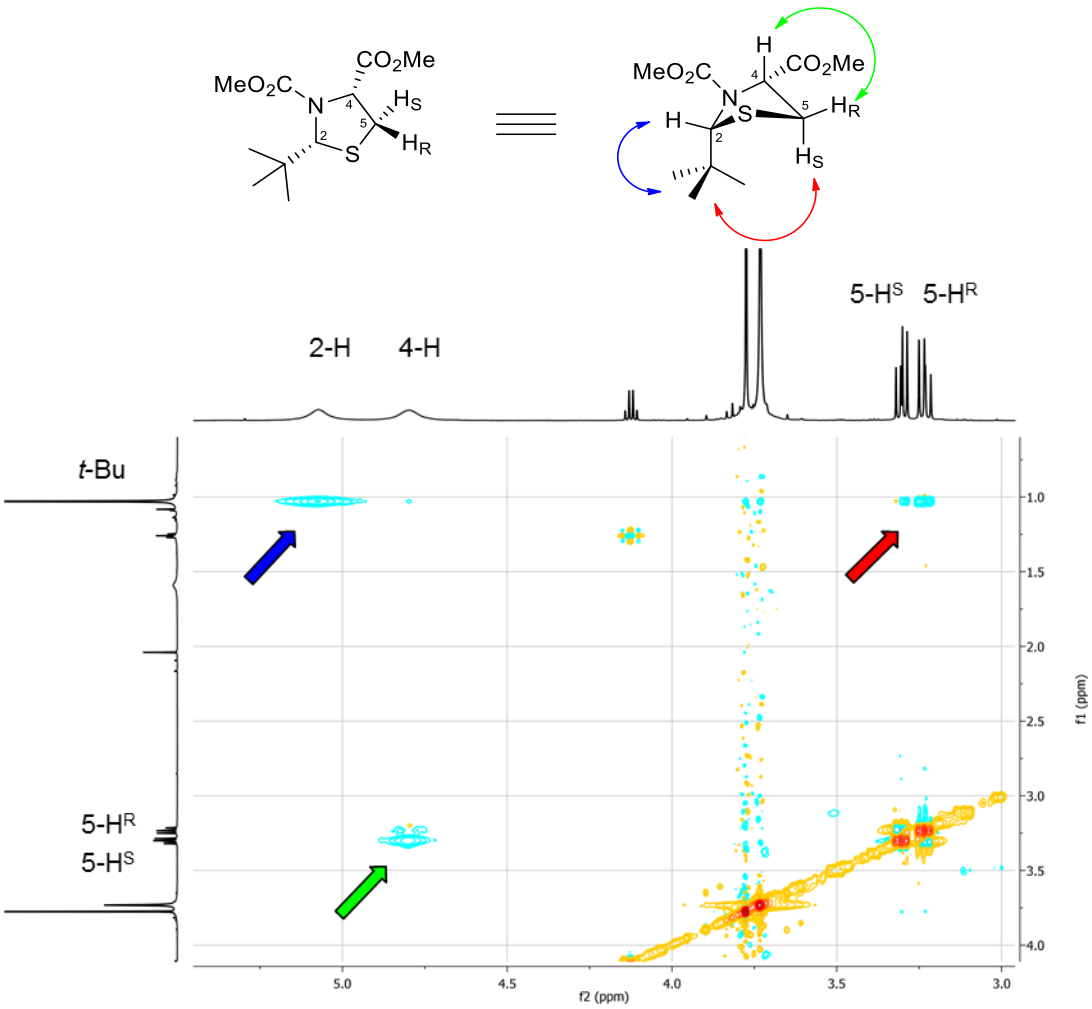
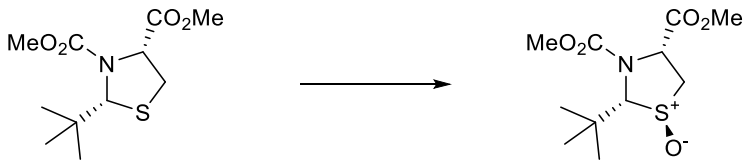


Figure 107 NOESY-NMR spectrum (600 MHz, $CDCl_3$) of the product with the desired stereochemistry, the arrows indicate the interaction between the different hydrogen atoms.

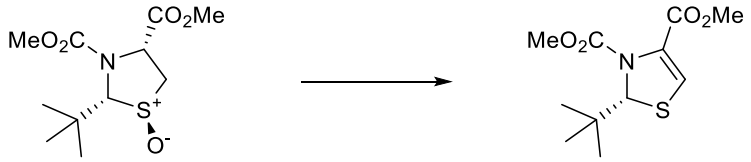
Synthesis of (2*R*,4*R*)-dimethyl 2-(tert-butyl)thiazolidine-3,4-dicarboxylate 1-oxide (86)



To a solution of **85** (5.334 g, 20.33 mmol, 1 eq) in acetic acid (25 ml), was carefully added hydrogen peroxide (2.29 ml, 22.36 mmol, 1.1 eq) and the light yellow solution was stirred at room temperature for 18 h. The solvent was evaporated and the crude product was purified by flash chromatography (pentane/EtOAc, 3:7) to give **86** as slightly yellow oil (4.996 g, 18.01 mmol, 89

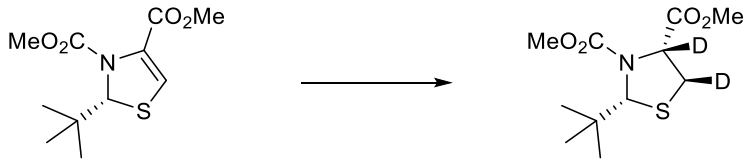
%). **^{1}H NMR (400 MHz, $CDCl_3$)** δ 5.34 (br s, 1H), 5.00 (br d, *J* = 50.6 Hz, 1H), 3.81 (s, 3H), 3.77 (s, 3H), 3.54 – 3.48 (m, 1H), 3.14 – 3.08 (m, 1H), 1.08 (s, 9H). **MS** *m/z* calcd. for $C_{11}H_{20}NO_5S$ $[M+H]^+$ 278.1, found: 278.1.

Synthesis of (*R*)-dimethyl 2-(*tert*-butyl)thiazole-3,4(*2H*)-dicarboxylate (87)



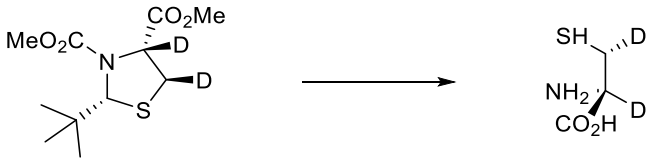
To a solution of **86** (4.996 g, 18.01 mmol, 1 eq) in dry DCM (60 ml) cooled to 0 °C, was carefully added TBDMS triflate (4.97 ml, 21.61 mmol, 1.2 eq) and the yellow solution was stirred under N_2. After 15 min, triethylamine (3.00 ml, 21.61 mmol, 1.2 eq) was added and the orange solution was stirred for 72 h at room temperature. Reaction control by TLC showed remaining starting material, hence an additional portion of TBDMS triflate (2.48 ml, 10.81 mmol, 0.6 eq) was added, followed after 15 min by triethylamine (1.50 ml, 10.81 mmol, 0.6 eq) and the solution was stirred for 24 h at room temperature. Reaction control by TLC showed remaining starting material, hence an additional portion of TBDMS triflate (1.24 ml, 5.40 mmol, 0.3 eq) was added, followed after 15 min by triethylamine (0.75 ml, 5.40 mmol, 0.3 eq) and the solution was stirred for 24 h at room temperature. Reaction control by TLC showed no remaining starting material. The solvent was evaporated and the oily residue was filtered over a small pad of silica using Et_2O as eluent and the filtrate was evaporated. The crude product was purified by flash column chromatography (pentane/EtOAc 95:5) to give **87** (2.316 g, 8.93 mmol, 50 %) as colorless oil. **^{1}H NMR (400 MHz, $CDCl_3$)** *δ* 6.89 (s, 1H), 5.52 (s, 1H), 3.79 (s, 3H), 3.77 (s, 3H), 0.95 (s, 9H). **MS** *m/z* calcd. for $C_{11}H_{18}NO_4S$ $[M+H]^+$ 260.1, found: 260.1.

Synthesis of (4*R*,5*R*)-4,5-[^{2}H]$_2$ (*R*)-dimethyl 2-(*tert*-butyl)thiazole-3,4(2*H*)-dicarboxylate (88)



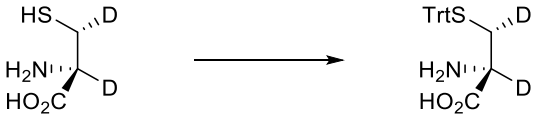
To a solution of **87** (1.105 g, 4.26 mmol, 1 eq) in AcOD-d_4 (8 ml), was added 10 % Pd/C (453 mg, 0.43 mmol, 10 mol %) and the black suspension was stirred under D_2 (50 bar) for 12 days at ambient temperature in a parr bomb. Reaction control by ^{1}H-NMR showed quantitative conversion of starting material. The reaction mixture was filtered over a pad of celite, the filter cake was washed with toluene and the filtrate was evaporated then co-evaporated to dryness twice with toluene to give the product (1.072 g, 4.07 mmol, 96 %) as light brown oil. **^{1}H-NMR (400 MHz, AcOD-d_4)** δ 5.09 (br s, 1H), 3.77 (s, 3H), 3.73 (s, 3H), 3.20 (s, 1H), 1.02 (s, 9H). **MS** *m/z* calcd. for $C_{11}H_{18}D_2NO_4S$ $[M+H]^+$ 264.1, found 264.1, $C_{11}H_{17}D_2NNaO_4S$ $[M+Na]^+$ 286.1, found 286.2.

Synthesis of (2*R*,3*R*)-2,3-[^{2}H]$_2$-cysteine (89)



A solution of **88** (1.052 g, 3.99 mmol, 1 eq) in 6 M HCl (30 ml) was stirred under reflux for 3 days under N_2-atmosphere. The solvent was evaporated to give the crude product (674 mg, ca. 60 % pure*, 2.53 mmol, 63 %) as light brown solid. **MS** *m/z* calcd. for $C_3H_6D_2NO_2S$ $[M+H]^+$ 124.0, found 124.1.

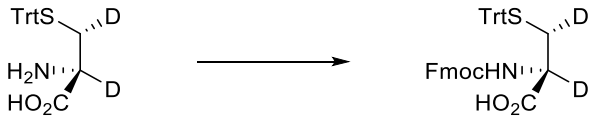
Synthesis of *S*-trityl-(2*R*,3*R*)-2,3-[^{2}H]$_2$-cysteine (91)



Using procedure C and starting with **89** (638 mg, 4.00 mmol), this product was obtained as off-white solid (804 mg, 2.20 mmol, 55 %). **^{1}H NMR (400 MHz, DMSO-d_6)** δ 7.36-7.23 (m, 15H), 2.37

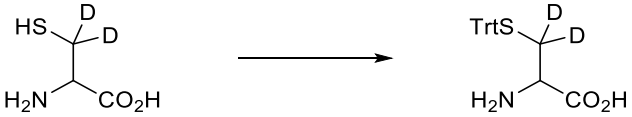
(s, 1H). **MS** *m/z* calcd. for $C_{22}H_{19}D_2NNaO_2S$ $[M+Na]^+$ 388.1, found 388.3. Analytical RP-UHPLC: t_R = 2.508 min (A/B 99:1 → 2:98 in 5.6 min, λ = 254 nm).

Synthesis of *S*-trityl-*N*-Fmoc-(2*R*,3*R*)-2,3-[^{2}H]$_2$-cysteine (90)



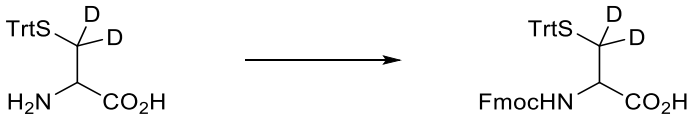
Using procedure D and starting with **91** (791 mg, 2.16 mmol), this product was obtained as slightly yellow solid (954 mg, 1.62 mmol, 75 %). **MS** *m/z* calcd. for $C_{37}H_{29}D_2NNaO_4S$ $[M+Na]^+$ 610.2, found 610.2. Analytical RP-UHPLC: t_R = 4.821 min (A/B 99:1 → 2:98 in 5.6 min, λ = 254 nm).

Synthesis of *S*-trityl-*DL*-3,3-[^{2}H]$_2$-cysteine (92)



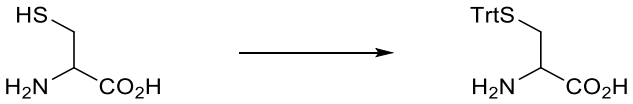
Using procedure C and starting with *D/L*-[3,3-^{2}H$_2$]-Cysteine (200 mg, 1.62 mmol), this product was obtained as white solid (590 mg, 1.61 mmol, 99 %). **^{1}H NMR (400 MHz, DMSO-d_6)** δ 7.36-7.24 (m, 15H), 2.92 (s, 1H). **MS** *m/z* calcd. for $C_{22}H_{21}NNaO_2S$ $[M+Na]^+$ 388.1, found 388.3. Analytical RP-UHPLC: t_R = 2.526 min (A/B 99:1 → 2:98 in 5.6 min, λ = 254 nm).

Synthesis of *S*-trityl-*N*-Fmoc-*DL*-3,3-[^{2}H]$_2$-cysteine (93)



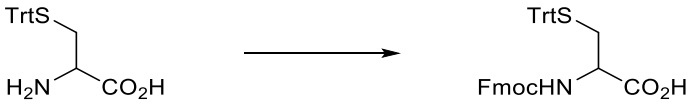
Using procedure D and starting with **92** (590 mg, 1.61 mmol), this product was obtained as white solid (700 mg, 1.19 mmol, 74 %). **MS** *m/z* calcd. for $C_{37}H_{29}D_2NNaO_4S$ $[M+Na]^+$ 610.2, found 610.2. Analytical RP-UHPLC: t_R = 5.058 min (A/B 99:1 → 2:98 in 5.6 min, λ = 254 nm).

Synthesis of *S*-trityl-*DL*-cysteine (94)



Using procedure C and starting with D/L-cysteine (242 mg, 2.00 mmol), this product was obtained as white solid (532 mg, 1.46 mmol, 73 %). **^{1}H NMR (400 MHz, DMSO-d_6)** δ 7.36-7.23 (m, 15H), 2.93-2.90 (m, 1H), 2.60-2.56 (m, 1H), 2.42-2.36 (m, 1H). **MS** *m/z* calcd. for $C_{22}H_{21}NNaO_2S$ $[M+Na]^+$ 386.1, found 386.3. Analytical RP-UHPLC: t_R = 2.524 min (A/B 99:1 → 2:98 in 5.6 min, λ = 254 nm).

Synthesis of *S*-trityl-*N*-Fmoc-*DL*-cysteine (95)



Using procedure D and starting with **94** (532 mg, 1.46 mmol), this product was obtained as white solid (703 mg, 1.20 mmol, 82 %). **MS** *m/z* calcd. for $C_{37}H_{31}NNaO_4S$ $[M+Na]^+$ 608.2, found 608.2. Analytical RP-UHPLC: t_R = 4.828 min (A/B 99:1 → 2:98 in 5.6 min, λ = 254 nm).

General Procedure for Peptide Synthesis (Abz-SALCSPTRA-NH_2)

Synthesis

Peptides were synthesized by standard Fmoc-SPPS procedure with a synthesis robot (Syro I, MultiSynTech). Rink amide resin (0.41 or 0.79 mmol/g) purchased from Novabiochem and Fmoc-protected amino acids purchased from Bachem were used. HCTU was used as the coupling agent. All amino acids were added as 4 eq. except lab synthesized Fmoc-Cys(Trt)-OH derivatives (marked in red), which were added as 1.3 eq. For the latter, reaction time was extended from one to two hours.

Cleavage and Purification

Cleavage and side-chain deprotection were carried out in TFA /TIPS/EDT/H_2O (94:2.5:2.5:1) during 5-12 h. The crude peptide was separated from the resin by filtration and precipitated in Et_2O, which was removed by centrifugation. The peptide was then resuspended in Et_2O, which was again removed by centrifugation. The crude product was allowed to dry, dissolved in water/MeCN and purified by preparative RP-HPLC and lyophilized. All yields are calculated as double TFA-salts.

Characterization

Peptides were characterized by analytical RP-HPLC and ESI-MS, measuring in the positive ion mode. For characterization by analytical RP-HPLC the following elution solutions were used: A = MeCN, B = mQ-deionized H_2O with 1 % MeCN and 0.1 % TFA.

(Abz-SALCSPTRA-NH_2) C = *L*-cysteine (Pep1)

Pep1 was obtained as fluffy white solid after RP-HPLC (7 mg, 6.84 µmol, 4 %). Analytical RP-HPLC: t_R = 15.547 min (A/B 15:85 → 20:80 in 23 min, λ = 310 nm). **MS** *m/z* calcd. for $C_{43}H_{71}N_{14}O_{13}S$ $[M+H]^+$ 1023.5040, found 1023.5030. *m/z* calcd. for $C_{43}H_{72}N_{14}O_{13}S$ $[M+2H]^{2+}$ 512.2557, found 512.2559.

(Abz-SALCSPTRA-NH_2) C = (2*R*,3*R*)-2,3-[^{2}H]$_2$-cysteine (Pep2)

Pep2 was obtained as fluffy white solid after RP-HPLC (150 mg, 119.7 µmol, 12 %). Analytical RP-HPLC: t_R = 15.404 min (A/B 15:85 → 20:80 in 23 min, λ = 310 nm). **MS** *m/z* calcd. for $C_{43}H_{69}D_2N_{14}O_{13}S$ $[M+H]^+$ 1025.5166, found 1025.5161. for $C_{43}H_{70}D_2N_{14}O_{13}S$ $[M+H]^+$ 513.2619, found 513.2621.

(Abz-SALCSPTRA-NH_2) C = *DL*-cysteine (Pep3)

Pep3 was obtained as fluffy white solid after RP-HPLC (145 mg, 115.7 µmol, 24 %). Analytical RP-HPLC: t_R = 15.511 min and 16.520 min (A/B 15:85 → 20:80 in 23 min, λ = 310 nm). **MS** *m/z* calcd. for $C_{43}H_{71}N_{14}O_{13}S$ $[M+H]^+$ 1023.5040, found 1023.5043. *m/z* calcd. for $C_{43}H_{72}N_{14}O_{13}S$ $[M+2H]^{2+}$ 512.2557, found 512.2562.

(Abz-SALCSPTRA-NH_2) C = *DL*-3,3-[^{2}H]$_2$-cysteine (Pep4)

Pep4 was obtained as fluffy white solid after RP-HPLC (128 mg, 102.1 µmol, 14 %). Analytical RP-HPLC: t_R = 15.453 min and 16.441 min (A/B 15:85 → 20:80 in 23 min, λ = 310 nm). **MS** *m/z* calcd. for $C_{43}H_{69}D_2N_{14}O_{13}S$ $[M+H]^+$ 1025.5166, found 1025.5163. for $C_{43}H_{70}D_2N_{14}O_{13}S$ $[M+2H]^{2+}$ 513.2619, found 513.2626.

4.6.2 Kinetic Assays

All reaction rates were determined by HPLC-based assays, using a RP-HPLC with a standard method (**Table 18**) to track product formation. Samples were quenched in appropriate intervals by adding 25 µL of the reaction mixture to 25 µL of 7.5 M Urea containing 1 % TFA. The samples were analyzed by RP-HPLC and the rate was determined by plotting product formation against the reaction time, from which the absolute rate was determined by extracting the slope of the plot and subsequent division by the enzyme concentration.

Table 18 Standard method for HPLC analysis of EgtB reaction products and starting materials; Solvent A: H_2O, 1 % MeCN, 0.1 % TFA; Solvent B: MeCN, 0.1 % TFA. Flow rate 1 mL/min.

Time (min)	Solvent A (%)	Solvent B (%)
0	85	15
0.5	85	15
12	80	20
17	5	95
20.7	5	95
22	85	15
25	Stop	Stop

Unless noted otherwise, conditions were: 50 µM **Pep1-4**, 2 mM DTT, 50 mM EDTA, 50 mM NaCl, 50 mM phosphate pH 8, 0.5 µM FGE. Prior to initiation of the assay with the enzyme, the latter was activated with 1 eq $CuSO_4$.

4.6.3 Protein Production

FGE_{wt}, FGE_{4C}, FGE_{S266A} and FGE_{W84R} were produced and kindly provided by Matthias Knop. All enzymes were expressed in *E. coli* BL21 cells and purified using Ni^{2+} NTA agarose beads (Qiagen) following standard procedures.[249]

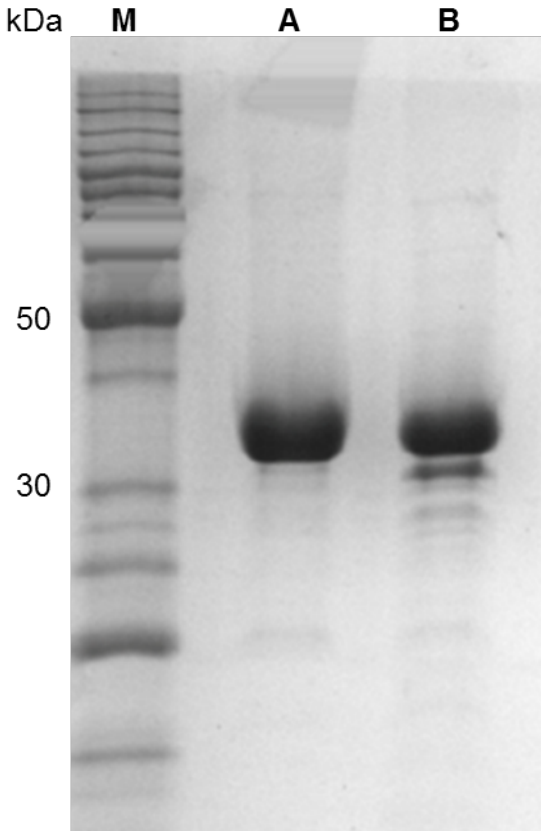


Figure 108 SDS-page picture of FGE variants after Ni-NTA affinity chromatography. **M** - Molecular weight marker; **A** - FGE_{wt}; **B** - FGE_{4C}.

Sequence of FGE_{wt}

GSSHHHHHHSSGLVPRGSHMPSFDFDIPRRSPQEIAKGMVAIPGGTFRMGGEDPDAFPEDGEGPVRTVRLS
PFLIDRYAVSNRQFAAFVKATGYVTDAERYGWSFVFHAHVAPGTPVMDAVVPEAPWWVAVPGAYWKAPEGP
GSSITDRPNHPVVHVSWNDAVAYATWAGKRLPTEAEWEMAARGGLDQARYPWGNELTPRGRHRCNIWQGTF
PVHDTGEDGYTGTAPVNAFAPNGYGLYNVAGNVWEWCADWWSADWHATESPATRIDPRGPETGTARVTKGG
SFLCHESYCNRYRVAARTCNTPDSSAAHTGFRCAADPL

m/z (FGE_{wt}): calc.: 35308.1 Da, measured: 35306.5 Da

Ɛ280(FGE_{wt}) = 93390 $M^{-1}cm^{-1}$

Sequence of FGE_{4C}

GSSHHHHHHSSGLVPRGSHMPSFDFDIPRRSPQEIAKGMVAIPGGTFRMGGEDPDAFPEDGEGPVRTVRLS
PFLIDRYAVSNRQFAAFVKATGYVTDAERYGWSFVFHAHVAPGTPVMDAVVPEAPWWVAVPGAYWKAPEGP
GSSITDRPNHPVVHVSWNDAVAYATWAGKRLPTEAEWEMAARGGLDQARYPWGNELTPRGRHRANIWQGTF
PVHDTGEDGYTGTAPVNAFAPNGYGLYNVAGNVWEWAADWWSADWHATESPATRIDPRGPETGTARVTKGG
SFLCHESYCNRYRVAARTSNTPDSSAAHTGFRAAADPL

m/z (FGE_{4C}): calc.: 35195.8 Da, measured: 35195.8 Da

Ɛ280(FGE_{4C}) = 93390 $M^{-1}cm^{-1}$

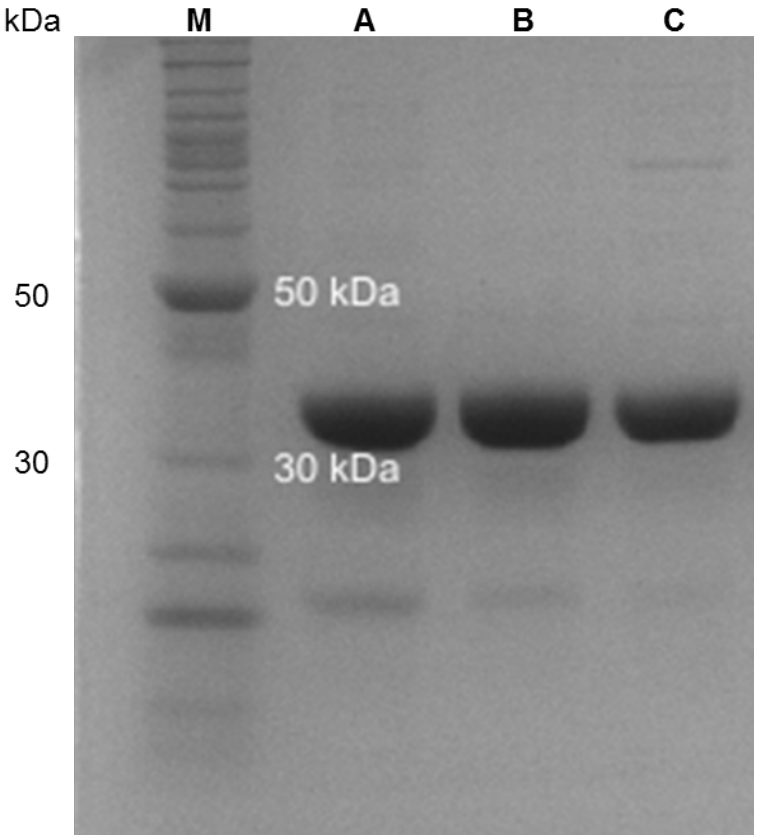


Figure 109 SDS-page picture of FGE variants after Ni-NTA affinity chromatography. **M** - Molecular weight marker; **A+B** - FGE_{wt}; **C** – FGE_{S266A}.

Sequence of FGE_{S266A}

GSSHHHHHHSSGLVPRGSHMPSFDFDIPRRSPQEIAKGMVAIPGGTFRMGGEDPDAFPEDGEGPVRTVRLS
PFLIDRYAVSNRQFAAFVKATGYVTDAERYGWSFVFHAHVAPGTPVMDAVVPEAPWWVAVPGAYWKAPEGP
GSSITDRPNHPVVHVSWNDAVAYATWAGKRLPTEAEWEMAARGGLDQARYPWGNELTPRGRHRCNIWQGTF
PVHDTGEDGYTGTAPVNAFAPNGYGLYNVAGNVWEWCADWWSADWHATESPATRIDPRGPETGTARVTKGG
AFLCHESYCNRYRVAARTCNTPDSSAAHTGFRCAADPL

m/z (FGE_{S266A}): calc.: 35292.1 Da, measured: 35290.4 Da

Ɛ280(FGE_{S266A}) = 93390 $M^{-1}cm^{-1}$

Sequence of FGE_{W84R}

GSSHHHHHHSSGLVPRGSHMPSFDFDIPRRSPQEIAKGMVAIPGGTFRMGGEDPDAFPEDGEGPVRTVRLS
PFLIDRYAVSNRQFAAFVKATGYVTDAERYGRSFVFHAHVAPGTPVMDAVVPEAPWWVAVPGAYWKAPEGP
GSSITDRPNHPVVHVSWNDAVAYATWAGKRLPTEAEWEMAARGGLDQARYPWGNELTPRGRHRCNIWQGTF
PVHDTGEDGYTGTAPVNAFAPNGYGLYNVAGNVWEWCADWWSADWHATESPATRIDPRGPETGTARVTKGG
SFLCHESYCNRYRVAARTCNTPDSSAAHTGFRCAADPL

m/z (FGE_{W84R}): calc.: 35278.1 Da, measured: 35276.9 Da

Ɛ280(FGE_{W84R}) = 87890 $M^{-1}cm^{-1}$

5 Final Conclusions

In this thesis, a variety of synthetic probes and a combination of analytical techniques have been used to investigate enzyme mechanisms.

Combining ^{13}C-labelled substrates with NMR spectroscopy and comparison of our findings to X-ray crystal structures has enabled the characterization of two molecular strategies of the SAM-dependent methyl transferase EgtD. This enzyme employs these in order to activate the methyl donor as well as the acceptor during catalysis. The first of which consists of interactions of active site enzyme residues with the transferred methyl group via C-H···O bonding, enhancing the electrophilicity of this group on the donor. The second strategy involves increasing the nucleophilicity of the accepting amine via unconventional $n \rightarrow \sigma^*$ interactions between active-site carbonyl groups and the methyl-group carbon of the acceptor. We anticipate that our measurements will inspire similar research into other methyltransferases and analogous non-enzymatic systems.

Investigation of the sulfoxide synthase EgtB in combination with two inhibitors has provided us with new insights into the factors determining the bifunctional nature of this enzyme. Specifically, by combining kinetic data with X-ray crystal structures we support a plausible hypothesis for the role of the key active site residue Tyr377. This residue has been the subject of much debate recently, with different propositions based on kinetic and computational studies. We suggest that active site acidity is key for the specificity of EgtB and that the role of Tyr377 is to suppress unwanted side-reactions. These results provide further points of discussion for this highly interesting system.

Finally, in the oxidation of cysteine to formylglycine by FGE, kinetic experiments using deuterium-labelled substrates have revealed a significant KIE on C-H bond cleavage, defining this as the rate limiting step of the reaction. By installing a deuterium label in an enantioselective manner, we have been able to elucidate the precise geometry of this cleavage step and create a model of the active site in the Michaelis-complex, consistent with all the current data. Furthermore, the presence of a surprisingly large secondary KIE has prompted us to re-examine the proposed catalytic mechanism and contemplate the integration of a single electron transfer step as a crucial part of catalysis.

6 References

(1) Snider, M. J.; Wolfenden, R. The Rate of Spontaneous Decarboxylation of Amino Acids. *J. Am. Chem. Soc.* **2000**, *122* (46), 11507–11508.

(2) Wolfenden, R.; Snider, M. J. The Depth of Chemical Time and the Power of Enzymes as Catalysts. *Acc. Chem. Res.* **2001**, *34* (12), 938–945.

(3) Pietzsch, J. Enhancing X-ray Vision http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1964/perspectives.html (accessed May 4, 2017).

(4) Sheehan, J. C.; Henery-Logan, K. R. The Total Synthesis of Penicillin V. *J. Am. Chem. Soc.* **1957**, *79* (5), 1262–1263.

(5) Sheehan, J. C.; Henery-Logan, K. R. The Total Synthesis of Penicillin V. *J. Am. Chem. Soc.* **1959**, *81*, 3089–3094.

(6) Feig, A.; Lippard, S. Reactions of Non-Heme Iron (II) Centers with Dioxygen in Biology and Chemistry. *Chem. Rev.* **1994**, *94* (3), 759–805.

(7) Baldwin, J. E.; Bradley, M. Isopenicillin N Synthase: Mechanistic Studies. *Chem. Rev.* **1990**, *90* (7), 1079–1088.

(8) Roach, P. L.; Clifton, I. J.; Fülöp, V.; Harlos, K.; Barton, G. J.; Hajdu, J.; Andersson, I.; Schofield, C. J.; Baldwin, J. E. Crystal Structure of Isopenicillin N Synthase Is the First from a New Structural Family of Enzymes. *Nature*. 1995, pp 700–704.

(9) Roach, P. L.; Clifton, I. J.; Hensgens, C. M.; Shibata, N.; Schofield, C. J.; Hajdu, J.; Baldwin, J. E. Structure of Isopenicillin N Synthase Complexed with Substrate and the Mechanism of Penicillin Formation. *Nature* **1997**, *387* (6635), 827–830.

(10) Baldwin, J. E.; Abraham, E. P.; Adlington, R. M.; Bahadur, G. A.; Bulbul Chakravarti; Domayne-hayman, B. P.; Field, L. D.; Flitsch, S. L.; Jayatilake, G. S.; Spakovskis, A.; et al. Penicillin Biosynthesis: Active Site Mapping with Aminoadipoylcysteinylvaline Variants. *J. Chem. Soc., Chem. Commun.* **1984**, *51* (3), 1225–1227.

(11) Baldwin, J. E.; Adlington, R. M.; Robinson, N. G.; Ting, H.-H. Stereospecificity of β-Lactam Formation in Penicillin Biosynthesis. *J. Chem. Soc., Chem. Commun.* **1986**, No. 5, 409–411.

(12) Baldwin, J. E.; Adlington, R. M.; Moroney, S. E.; Field, L. D.; Ting, H. Stepwise Ring Closure in Penicillin Biosynthesis. Initial β-Lactam Formation. *J. Chem. Soc., Chem. Commun.* **1984**, 984–986.

(13) Bahadur, G. A.; Baldwin, J. E.; Wan, T.; Jung, M.; Abraham, S. E.; Huddleston, J. A.; White, R. L. On the Proposed Intermediacy of β-Hydroxyvaline - and Thiazepinone - Containing Peptides in Penicillin Biosynthesis. *J. Chem. Soc., Chem. Commun.* **1981**, 1146–1147.

(14) Abraham, P.; Adlington, R. M.; Crimmin, M. J.; Field, L. D.; Usher, J. J. The Synthesis and Reactions of a Monocyclic β-Lactam Tripeptide, 1-[(1R)-Carboxy-2-Methylpropyl]- (3R - [(5S-5-Amino- 5-Carboxypentanamido]-(4R)-Mercaptoazetidin-2-One, a Putative Intermediate in Penicillin Biosynthesis. *Tetrahedron* **1984**, *40* (10), 1907–1918.

(15) Baldwin, J. E.; Johnson, B. L.; Usher, J. J.; Abraham, E. P.; Huddleston, J. A.; White, R. L. Direct N.m.r. Observation of Cell-Free Conversion of (L-a-Amino-D-Adipyl) - L-Cysteinyl-D-Valine into Isopenicillin N. *J. Chem. Soc., Chem. Commun.* **1980**, 1271–1273.

(16) Baxter, R. L.; McGregor, C. J.; Thomson, and G. A.; Scott, A. I. Intact Incorporation of D-(a-L- Aminoadipoyl) -L- [3- 13C]cysteinyl -D- [15N]valine into Lsopenicillin N. Observation of One-Bond 13C-15N Coupling. **1985**, 369–372.

(17) Baldwin, J. E.; Abraham, S. E. The Biosynthesis of Penicillins and Cephalosporins. *Nat. Prod. Rep.* **1988**, *5* (2), 129–145.

(18) Simon, H.; Palm, D. Isotope Effects in Organic Chemistry and Biochemistry. *Angew. Chemie Int. Ed.* **1966**, *5* (11), 920–933.

(19) Aldridge, D. C.; Carr, D. M.; Davies, D. H.; Hudson, A. J.; Nolan, R. D.; Poyser, J. P.; Strawson, C. J. Antibiotics 13285 A1 and A2 : Novel Cepham and Penam Metabolites from a Streptomyces Species. *J. Chem. Soc., Chem. Commun.* **1985**, 1513–1514.

(20) Wan, T. S.; Baldwin, J. E. Penicillin Biosynthesis . A Model for Carbon-Sulphur Bond Formation. *J. Chem. Soc., Chem. Commun.* **1979**, 249–250.

(21) Tamanaha, E.; Zhang, B.; Guo, Y.; Chang, W. C.; Barr, E. W.; Xing, G.; St Clair, J.; Ye, S.; Neese, F.; Bollinger, J. M.; et al. Spectroscopic Evidence for the Two C-H-Cleaving Intermediates of Aspergillus Nidulans Isopenicillin N Synthase. *J. Am. Chem. Soc.* **2016**, *138* (28), 8862–8874.

(22) Rinkel, J.; Dickschat, J. S. Recent Highlights in Biosynthesis Research Using Stable Isotopes. *Beilstein J. Org. Chem.* **2015**, *11*, 2493–2508.

(23) Dickschat, J. S. Bacterial Terpene Cyclases. *Nat. Prod. Rep.* **2016**, *33*, 87–110.

(24) Wallach, O. Zur Kenntnis Der Terpene Und Der ätherischen Öle. *Justus Liebigs Ann. Chem.* **1885**, *227* (3), 277–302.

(25) Chen, X.; Köllner, T. G.; Jia, Q.; Norris, A.; Santhanam, B.; Rabe, P.; Dickschat, J. S.; Shaulsky, G.; Gershenzon, J.; Chen, F. Terpene Synthase Genes in Eukaryotes beyond Plants and Fungi: Occurrence in Social Amoebae. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (43), 12132–12137.

(26) Rabe, P.; Rinkel, J.; Dolja, E.; Schmitz, T.; Nubbemeyer, B.; Luu, T. H.; Dickschat, J. S. Mechanistic Investigations of Two Bacterial Diterpene Cyclases: Spiroviolene Synthase and Tsukubadiene Synthase. *Angew. Chemie Int. Ed.* **2017**, 2776–2779.

(27) Thulasiram, H. V; Poulter, C. D. Farnesyl Diphosphate Synthase : The Art of Compromise between Stereoselectivity and Substrate Selectivity. *J. Am. Chem. Soc.* **2006**, *128*, 15819–15823.

(28) Meguro, A.; Motoyoshi, Y.; Teramoto, K.; Ueda, S.; Totsuka, Y.; Ando, Y.; Tomita, T.; Kim, S. Y.; Kimura, T.; Igarashi, M.; et al. An Unusual Terpene Cyclization Mechanism Involving a Carbon-Carbon Bond Rearrangement. *Angew. Chemie - Int. Ed.* **2015**, *54* (14), 4353–4356.

(29) Pfeiffer, C.; Bauer, T.; Surek, B.; Schömig, E.; Gründemann, D. Cyanobacteria Produce High Levels of Ergothioneine. *Food Chem.* **2011**, *129* (4), 1766–1769.

(30) Seebeck, F. P. In Vitro Reconstitution of Mycobacterial Ergothioneine Biosynthesis. *J. Am. Chem. Soc.* **2010**, *132* (19), 6632–6633.

(31) Gründemann, D.; Harlfinger, S.; Golz, S.; Geerts, A.; Lazar, A.; Berkels, R.; Jung, N.; Rubbert, A.; Schömig, E. Discovery of the Ergothioneine Transporter. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (14), 5256–5261.

(32) Tokuhiro, S.; Yamada, R.; Chang, X.; Suzuki, A.; Kochi, Y.; Sawada, T.; Suzuki, M.; Nagasaki, M.; Ohtsuki, M.; Ono, M.; et al. An Intronic SNP in a RUNX1 Binding Site of SLC22A4, Encoding an Organic Cation Transporter, Is Associated with Rheumatoid Arthritis. *Nat. Genet.* **2003**, *35* (4), 341–348.

(33) Waller, S.; Tremelling, M.; Bredin, F.; Godfrey, L.; Howson, J.; Parkes, M. Evidence for Association of OCTN Genes and IBD5 with Ulcerative Colitis. *Gut* **2006**, *55* (6), 809–814.

(34) Santiago, J. L.; Martínez, A.; de la Calle, H.; Fernández-Arquero, M.; Figueredo, M. A.; de la Concha, E. G.; Urcelay, E. Evidence for the Association of the SLC22A4 and SLC22A5 Genes with Type 1 Diabetes: A Case Control Study. *BMC Med. Genet.* **2006**, *7*, 54.

(35) Peltekova, V. D.; Wintle, R. F.; Rubin, L. a; Amos, C. I.; Huang, Q.; Gu, X.; Newman, B.; Van Oene, M.; Cescon, D.; Greenberg, G.; et al. Functional Variants of OCTN Cation Transporter Genes Are Associated with Crohn Disease. *Nat. Genet.* **2004**, *36* (5), 471–475.

(36) Hartman, P. E. Ergothioneine as Antioxidant. *Methods Enzymol.* **1990**, *186* (1958), 310–318.

(37) Hu, W.; Song, H.; Her, A. S.; Bak, D. W.; Naowarojna, N.; Elliott, S. J.; Qin, L.; Chen, X.; Liu, P. Bioinformatic and Biochemical Characterizations of C − S Bond Formation and Cleavage Enzymes in the Fungus Neurospora Crassa Ergothioneine Biosynthetic Pathway. *Org. Lett.* **2014**, *16* (20), 5382–5385.

(38) Turner, E.; Klevit, R.; Hopkins, P. B.; Shapiro, B. M. Ovothiol: A Novel Thiohistidine Compound from Sea Urchin Eggs That Confers NAD(P)H-O2 Oxidoreductase Activity on Ovoperoxidase. *J. Biol. Chem.* **1986**, *261* (28), 13056–13059.

(39) Steenkamp, D. J.; Spies, H. S. C. Identification of a Major Low-Molecular-Mass Thiol of the Trypanosomatid Crithidia-Fasciculata As Ovothiol-a - Facile Isolation and Structural-Analysis of the Bimane Derivative. *Eur. J. Biochem.* **1994**, *223* (1), 43–50.

(40) Ariyanayagam, M. R.; Fairlamb, A. H. Ovothiol and Trypanothione as Antioxidants in Trypanosomatids. *Mol. Biochem. Parasitol.* **2001**, *115* (2), 189–198.

(41) Braunshausen, A.; Seebeck, F. P. Identification and Characterization of the First Ovothiol Biosynthetic Enzyme. *J. Am. Chem. Soc.* **2011**, *133* (6), 1757–1759.

(42) Tanret, M. C. Sur Une Base Nouvelle Retireé Du Seigle Ergoté, L'Ergothionéine. *Ann. Chim. Phys.* **1909**, *18*, 114–124.

(43) Barger, G.; Ewins, A. J. The Constitution of Ergothioneine : A Betaine Related to Histidine. *J. Chem. Soc.* **1911**, *99*, 2336–2341.

(44) Horowitz, S.; Yesselman, J. D.; Al-Hashimi, H. M.; Trievel, R. C. Direct Evidence for Methyl Group Coordination by Carbon-Oxygen Hydrogen Bonds in the Lysine Methyltransferase SET7/9. *J. Biol. Chem.* **2011**, *286* (21), 18658–18663.

(45) Struck, A. W.; Thompson, M. L.; Wong, L. S.; Micklefield, J. S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications. *ChemBioChem* **2012**, *13* (18), 2642–2655.

(46) Fontecave, M.; Atta, M.; Mulliez, E. S-Adenosylmethionine: Nothing Goes to Waste. *Trends Biochem. Sci.* **2004**, *29* (5), 243–249.

(47) Copeland, R. a; Solomon, M. E.; Richon, V. M. Protein Methyltransferases as a Target Class for Drug Discovery. *Nat. Rev. Drug Discov.* **2009**, *8* (9), 724–732.

(48) O'Hagan, D.; Schmidberger, J. W. Enzymes That Catalyse SN2 Reaction Mechanisms. *Nat. Prod. Rep.* **2010**, *27* (6), 900–918.

(49) Axelrod, J.-. Noradrenaline: Fate and Control of Its Biosynthesis. *Science (80-.).* **1971**, *173*, 598–606.

(50) Mihel, I.; Knipe, J. O.; Coward, J. K.; Schowen, R. L. .alpha.-Deuterium Isotope Effects and Transition-State Structure in an Intramolecular Model System for Methyl-Transfer Enzymes. *J. Am. Chem. Soc.* **1979**, *101* (15), 4349–4351.

(51) Ruggiero, G. D.; Williams, I. H.; Roca, M.; Moliner, V.; Tuñón, I. QM/MM Determination of Kinetic Isotope Effects for COMT-Catalyzed Methyl Transfer Does Not Support Compression Hypothesis. *J. Am. Chem. Soc.* **2004**, *126* (28), 8634–8635.

(52) Zhang, J.; Klinman, J. P. Enzymatic Methyl Transfer: Role of an Active Site Residue in Generating Active Site Compaction That Correlates with Catalytic Efficiency. *J. Am. Chem. Soc.* **2011**, *133* (43), 17134–17137.

(53) Zhang, J.; Kulik, H. J.; Martinez, T. J.; Klinman, J. P. Mediation of Donor-Acceptor Distance in an Enzymatic Methyl Transfer Reaction. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (26), 7954–7959.

(54) Roca, M.; Mart, S.; Andrs, J.; Moliner, V.; Tun, I.; Bertrn, J.; Williams, I. H.; Roca, M.; Martı, S.; Andre, J.; et al. Theoretical Modeling of Enzyme Catalytic Power : Analysis of " Cratic " and Electrostatic Factors in Catechol O -Methyltransferase. *J. Am. Chem. Soc.* **2003**, *125* (2), 7726–7737.

(55) Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. Elucidating the Nature of Enzyme Catalysis Utilizing a New Twist on an Old Methodology: Quantum Mechanical-Free Energy Calculations on Chemical Reactions in Enzymes and in Aqueous Solution. *Acc. Chem. Res.* **2001**, *34* (1), 72–79.

(56) Lameira, J.; Bora, R. P.; Chu, Z. T.; Warshel, A. Methyltransferases Do Not Work by Compression, Cratic, or Desolvation Effects, but by Electrostatic Preorganization. *Proteins Struct. Funct. Bioinforma.* **2015**, *83* (2), 318–330.

(57) Warshel, A.; Bora, R. P. Perspective: Defining and Quantifying the Role of Dynamics in Enzyme Catalysis. *J. Chem. Phys.* **2016**, *144* (18).

(58) Wood, A.; Shilatifard, A. Posttranslational Modifications of Histones by Methylation. *Adv. Protein Chem.* **2004**, *67*, 201–222.

(59) Dillon, S. C.; Zhang, X.; Trievel, R. C.; Cheng, X. The SET-Domain Protein Superfamily: Protein Lysine Methyltransferases. *Genome Biol.* **2005**, *6* (8), 227.

(60) Couture, J. F.; Hauk, G.; Thompson, M. J.; Blackburn, G. M.; Trievel, R. C. Catalytic Roles for Carbon-Oxygen Hydrogen Bonding in SET Domain Lysine Methyltransferases. *J. Biol. Chem.* **2006**, *281* (28), 19280–19287.

(61) Kwon, T.; Chang, J. H.; Kwak, E.; Lee, C. W.; Joachimiak, A.; Kim, Y. C.; Lee, J. W.; Cho, Y. Mechanism of Histone Lysine Methyl Transfer Revealed by the Structure of SET7/9-AdoMet. *EMBO J.* **2003**, *22* (2), 292–303.

(62) Xiao, B.; Jing, C.; Wilson, J. R.; Walker, P. A.; Vasisht, N.; Kelly, G.; Howell, S.; Taylor, I. A.; Blackburn, G. M.; Gamblin, S. J. Structure and Catalytic Mechanism of the Human Histone Methyltransferase SET7/9. *Nature* **2003**, *421* (February), 652–656.

(63) Trievel, R. C.; Flynn, E. M.; Houtz, R. L.; Hurley, J. H. Mechanism of Multiple Lysine Methylation by the SET Domain Enzyme Rubisco LSMT. *Nat. Struct. Biol.* **2003**, *10* (7), 545–552.

(64) Ramachandran, G. N.; Sasisekharan, V. Refinement of the Structure of Collagen. *Biochim. Biophys. Acta - Biophys. Incl. Photosynth.* **1965**, *109* (1), 314–316.

(65) Steiner, T.; Desiraju, G. R.; R. Desiraju, G. Distinction between the Weak Hydrogen Bond and the van Der Waals Interaction. *Chem. Commun.* **1998**, *0* (8), 891–892.

(66) Cannizzaro, C. E.; Houk, K. N. Magnitudes and Chemical Consequences of R(3)N(+)-C-H...O[double bond]C Hydrogen Bonding. *J. Am. Chem. Soc.* **2002**, *124* (12), 7163–7169.

(67) Derewenda, Z. S.; Lee, L.; Derewenda, U. The Occurrence of C-H...O Hydrogen Bonds in Proteins. *J. Mol. Biol.* **1995**, *252* (2), 248–262.

(68) Berger, I.; Egli, M. The Role of Backbone Oxygen Atoms in the Organization of Nucleic Acid Tertiary Structure : Zippers, Networks, Clamps, and C-H O Hydrogen Bonds. *Chem. A Eur. J.* **1997**, *3* (9), 1400–1404.

(69) Qingzhong, L.; Wang, N.; Zhiwu, Y. Effect of Hydration on the CH···O Hydrogen Bond: A Theoretical Study. *J. Mol. Struct. THEOCHEM* **2007**, *847* (1-3), 68–74.

(70) Ratajczyk, T.; Czerski, I.; Kamienska-Trela, K.; Szymanski, S.; Wojcik, J. 1J(C,H) Couplings to the Individual Protons in a Methyl Group: Evidence of the Methyl Protons' Engagement in Hydrogen Bonds. *Angew. Chemie - Int. Ed.* **2005**, *44* (8), 1230–1232.

(71) Cordier, F.; Barfield, M.; Grzesiek, S. Direct Observation of Cα-Hα...O=C Hydrogen Bonds in Proteins by Interresidue h3JCαC' Scalar Couplings. *J. Am. Chem. Soc.* **2003**, *125* (51), 15750–15751.

(72) Sheppard, D.; Li, D.; Godoy-ruiz, R.; Bru, R. Variation in Quadrupole Couplings of R Deuterons in Ubiquitin Suggests the Presence of C R -H R · · · OdC Hydrogen Bonds. *J. Am. Chem. Soc.* **2010**, *132* (14), 7709–7719.

(73) Sukumar, N.; Mathews, F. S.; Langan, P.; Davidson, V. L. A Joint X-Ray and Neutron Study on Amicyanin Reveals the Role of Protein Dynamics in Electron Transfer. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (15), 6817–6822.

(74) Scheiner, S.; Gu, Y.; Kar, T. Evaluation of the H-Bonding Properties of CH ·· · O Interactions Based upon NMR Spectra. **2000**, *500*, 441–452.

(75) Ash, E. L.; Sudmeier, J. L.; Day, R. M.; Vincent, M.; Torchilin, E. V; Haddad, K. C.;

Bradshaw, E. M.; Sanford, D. G.; Bachovchin, W. W. Unusual 1H NMR Chemical Shifts Support (His) C(epsilon) 1-O=C H-Bond: Proposal for Reaction-Driven Ring Flip Mechanism in Serine Protease Catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (19), 10371–10376.

(76) Yates, J. R.; Pham, T. N.; Pickard, C. J.; Mauri, F.; Amado, A. M.; Gil, A. M.; Brown, S. P. An Investigation of Weak CH???O Hydrogen Bonds in Maltose Anomers by a Combination of Calculation and Experimental Solid-State NMR Spectroscopy. *J. Am. Chem. Soc.* **2005**, *127* (29), 10216–10220.

(77) Scheiner, S. Identification of Spectroscopic Patterns of CH ⋯ O H-Bonds in Proteins. *J. Phys. Chem. B* **2009**, *113* (30), 10421–10427.

(78) Scheibler, C. Über Das Betain, Eine Im Safte Der Zuckerrüben (Beta Vulgaris) Vorkommende Pflanzenbase. *Berichte Dtsch. Chem. Ges.* **1869**, *2*, 292–295.

(79) Craig, S. A. Betaine in Human Nutrition. *Am. J. Clin. Nutr.* **2004**, *80*, 539–549.

(80) Paleg, L. G.; Aspinall, D. *The Physiology and Biochemistry of Drought Resistance in Plants*; Academic Press, 1981.

(81) Yancey, P. H.; Clark, M. E.; Hand, S. C.; Bowlus, R. D.; Somero, G. N. Living with Water Stress: Evolution of Osmolyte Systems. *Science (80-.).* **1982**, *217* (4566), 1214–1222.

(82) Selhub, J. Homocysteine Metabolism. *Annu. Rev. Nutr.* **1999**, *19*, 217–246.

(83) Kwon, D. Y.; Jung, Y. S.; Kim, S. J.; Park, H. K.; Park, J. H.; Kim, Y. C. Impaired Sulfur-Amino Acid Metabolism and Oxidative Stress in Nonalcoholic Fatty Liver Are Alleviated by Betaine Supplementation in Rats. *J. Nutr.* **2009**, *139* (October 2008), 63–68.

(84) Fan, F.; Gadda, G. On the Catalytic Mechanism of Choline Oxidase. *J. Am. Chem. Soc.* **2005**, *127* (7), 2067–2074.

(85) Boch, J.; Kempf, B.; Bremer, E. Osmoregulation in Bacillus Subtilis: Synthesis of the Osmoprotectant Glycine Betaine from Exogenously Provided Choline. *J. Bacteriol.* **1994**, *176* (17), 5364–5371.

(86) Kappes, R. M.; Kempf, B.; Kneip, S.; Boch, J.; Gade, J.; Meier-Wagner, J.; Bremer, E. Two Evolutionarily Closely Related ABC Transporters Mediate the Uptake of Choline for Synthesis of the Osmoprotectant Glycine Betaine in Bacillus Subtilis. *Mol. Microbiol.* **1999**, *32* (1), 203–216.

(87) Boch, J.; Boch, J.; Kempf, B.; Kempf, B.; Schmid, R.; Schmid, R.; Bremer, E.; Bremer, E.; Dna, T.; Dna, T. Synthesis of the Osmoprotectant Glycine Betaine in Bacillus Subtilis: Characterization of the gbsAB Genes. *J. Bacteriol.* **1996**, *178* (17), 5121–5129.

(88) Pittelkow, M.; Tschapek, B.; Smits, S. H. J.; Schmitt, L.; Bremer, E. The Crystal Structure of the Substrate-Binding Protein OpuBC from Bacillus Subtilis in Complex with Choline. *J. Mol. Biol.* **2011**, *411* (1), 53–67.

(89) Dougherty, D. A.; Stauffer, D. A. Acetylcholine Binding by a Synthetic Receptor: Implications for Biological Recognition. *Science (80-.).* **1990**, *250* (4987), 1558–1560.

(90) Dougherty, D. A.; Gallivan, J. P. Cation-π Interactions in Structural Biology. *Biochemistry* **1999**, *96* (August), 9459–9464.

(91) Nyyssölä, A.; Kerovuo, J.; Kaukinen, P.; Von Weymarn, N.; Reinikainen, T. Extreme Halophiles Synthesize Betaine from Glycine by Methylation. *J. Biol. Chem.* **2000**, *275* (29), 22196–22201.

(92) McCoy, J. G.; Bailey, L. J.; Ng, Y. H.; Bingman, C. A.; Wrobel, R.; Weber, A. P. M.; Fox, B. G.; Phillips, G. N. Discovery of Sarcosine Dimethylglycine Methyltransferase from Galdieria Sulphuraria. *Proteins Struct. Funct. Bioinforma.* **2009**, *74* (2), 368–377.

(93) Kallio, J. P.; Jänis, J.; Nyyssölä, A.; Hakulinen, N.; Rouvinen, J. Preliminary X-Ray Analysis of Twinned Crystals of Sarcosine Dimethylglycine Methyltransferase from Halorhodospira Halochoris. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2009**, *65* (8), 805–808.

(94) Lee, Y.-R.; Lin, T.-S.; Lai, S.-J.; Liu, M.-S.; Lai, M.-C.; Chan, N.-L. Structural Analysis of Glycine Sarcosine N-Methyltransferase from Methanohalophilus Portucalensis Reveals Mechanistic Insights into the Regulation of Methyltransferase Activity. *Sci. Rep.* **2016**, *6* (May), 38071.

(95) Lai, M.-C.; Yang, D.-R.; Chuang, M.-J. Regulatory Factors Associated with Synthesis of the Osmolyte Glycine Betaine in the Halophilic MethanoarchaeonMethanohalophilus Portucalensis. *Appl. Environ. Microbiol.* **1999**, *65* (2), 828–833.

(96) Sironi, M.; Fornilia, A.; Fornili, S. L. Water Interaction with Glycine Betaine : A Hybrid QM / MM Molecular Dynamics Simulation. *Phys. Chem. Chem. Phys.* **2001**, *3*, 1081–1085.

(97) Vit, A.; Misson, L.; Blankenfeldt, W.; Seebeck, F. P. Ergothioneine Biosynthetic Methyltransferase EgtD Reveals the Structural Basis of Aromatic Amino Acid Betaine Biosynthesis. *Chembiochem* **2015**, *16* (1), 119–125.

(98) Desiraju, G. R. The C - H.... O Hydrogen Bond : Structural Implications and Supramolecular Design. *Acc. Chem. Res.* **1996**, *29* (95), 441–449.

(99) Misson, L. Insights into the Mechanism and Regulation of EgtD, a Novel Histidine Methyltransferase from Ergothioneine Biosynthesis, University of Basel, 2017.

(100) Albery, W. J.; Knowles, J. R. Evolution of Enzyme Function and the Development of Catalytic Efficiency. *Biochemistry* **1979**, *15* (25), 5631–5640.

(101) Burbaum, J. J.; Raines, R. T.; Albery, W. J.; Knowles, J. R. Evolutionary Optimization of the Catalytic Effectiveness of an Enzyme. *Biochemistry* **1989**, *28*, 9293–9305.

(102) Denmark, S. E. Asymmetric Catalysis with Chiral Lewis Bases: A New Frontier in Main Group Chemistry. *Chimia (Aarau).* **2008**, *62* (1-2), 37–40.

(103) Borch, R. F.; Bernstein, M. D.; Durst, H. D. The Cyanohydridoborate Anion as a Selective Reducing Agent. *J. Am. Chem. Soc.* **1971**, *93* (12), 2897–2904.

(104) Song, H.; Leninger, M.; Lee, N.; Liu, P. Regioselectivity of the Oxidative C-S Bond Formation in Ergothioneine and Ovothiol Biosyntheses. *Org. Lett.* **2013**, *15* (18), 4854–4857.

(105) Reinhold, V. N.; Ishikawa, Y.; Melville, D. B. Synthesis of a-N-Methylated Histidines. *J. Med. Chem.* **1968**, *11* (2), 258–260.

(106) Madsen, C.; Jensen, A. A.; Liljefors, T.; Kristiansen, U.; Nielsen, B.; Hansen, C. P.; Larsen, M.; Ebert, B.; Bang-andersen, B.; Krogsgaard-larsen, P.; et al. 5-Substituted Imidazole-4-

Acetic Acid Analogues : Synthesis , Modeling , and Pharmacological Characterization of a Series of Novel γ -Aminobutyric Acid C Receptor Agonists. *J. Med. Chem.* **2007**, *50* (17), 4147–4161.

(107) Matziari, M.; Bauer, K.; Dive, V.; Yiotakis, A. Synthesis of the Phosphinic Analogue of Thyrotropin Releasing Hormone. *J. Org. Chem.* **2008**, *73* (21), 8591–8593.

(108) Park, J.; Tai, J.; Roessner, C. A.; Scott, A. I. Enzymatic Synthesis of S-Adenosyl-L-Methionine on the Preparative Scale. *Bioorganic Med. Chem.* **1996**, *4* (12), 2179–2185.

(109) Park, J.; Tai, J.; Roessner, C. A.; Scott, A. I. Overcoming Product Inhibition of S-Adenosyl-L-Methionine (SAM) Synthetase: Preparation of SAM on the 30 mM Scale. *Bioorganic Med. Chem. Lett.* **1995**, *5* (19), 2203–2206.

(110) Walsby, C. J.; Hong, W.; Broderick, W. E.; Cheek, J.; Ortillo, D.; Broderick, J. B.; Hoffman, B. M. Electron-Nuclear Double Resonance Spectroscopic Evidence That S-Adenosylmethionine Binds in Contact with the Catalytically Active [4Fe-4S]+ Cluster of Pyruvate Formate-Lyase Activating Enzyme. *J. Am. Chem. Soc.* **2002**, *124* (12), 3143–3151.

(111) Poulin, M. B.; Du, Q.; Schramm, V. L. Chemoenzymatic Synthesis of 36S Isotopologues of Methionine and S-Adenosyl-L-Methionine. *J. Org. Chem.* **2015**, *80* (10), 5344–5347.

(112) Markham, G. D.; Hafner, E. W.; Tabor, C. W.; Tabor, H. S-Adenosylmethionine Synthetase from Escherichia Coli. *J. Biol. Chem.* **1980**, *255* (19), 9082–9092.

(113) Mati, I. K.; Cockroft, S. L. Molecular Balances for Quantifying Non-Covalent Interactions. *Chem. Soc. Rev.* **2010**, *39* (11), 4195–4205.

(114) Tamura, Y.; Yamamoto, G.; Oki, M. CH3···O Hydrogen Bond. Implications of Its Presence from the Substituent Effects on the Populations of Rotamers in 4-Substituted 9-Ethyl-1-Methoxytriptycenes and 9-(Substituted Phenoxymethyl)-1,4-Dimethyltriptycenes. 1987, pp 1781–1788.

(115) Carroll, W. R.; Zhao, C.; Smith, M. D.; Pellechia, P. J.; Shimizu, K. D. A Molecular Balance for Measuring Aliphatic CH-Pi Interactions. *Org. Lett.* **2011**, *13* (16), 4320–4323.

(116) Vit, A. *Unpublished*; 2017.

(117) Steiner, T. Unrolling the Hydrogen Bond Properties of C – H ··· O Interactions. *Chem. Commun.* **1997**, 727–734.

(118) Carpenter, B. K. Kinetic Isotope Effects: Unearthing the Unconventional. *Nat. Chem.* **2010**, *2* (2), 80–82.

(119) Scheiner, S.; Čuma, M. Relative Stability of Hydrogen and Deuterium Bonds. *J. Am. Chem. Soc.* **1996**, *118* (6), 1511–1521.

(120) Barfield, M.; Grant, D. M. The Dependence of Geminal Proton Spin-Spin Coupling Constants on Electron Delocalization in Molecules. *J. Am. Chem. Soc.* **1961**, *83* (23), 4726–4729.

(121) Cambridge Isotope Laboratories Inc. Deuterated Solvents for NMR https://www.isotope.com/uploads/File/NMRUNpriced-proof.pdf (accessed Apr 5, 2017).

(122) Karplus, M.; Anderson, D. H.; Farrar, T. C.; Gutowsky, H. S. Valence-Bond Interpretation of Electron Coupled Proton-Proton Magnetic Interactions Measured via Deuterium Substitution. *J. Chem. Phys.* **1957**, *27* (2), 597.

(123) Poulin, M. B.; Schneck, J. L.; Matico, R. E.; McDevitt, P. J.; Huddleston, M. J.; Hou, W.; Johnson, N. W.; Thrall, S. H.; Meek, T. D.; Schramm, V. L. Transition State for the NSD2-Catalyzed Methylation of Histone H3 Lysine 36. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (5), 1197–1201.

(124) Zhang, J.; Klinman, J. P. Convergent Mechanistic Features between the Structurally Diverse N- and O-Methyltransferases: Glycine N-Methyltransferase and Catechol O-Methyltransferase. *J. Am. Chem. Soc.* **2016**, *138* (29), 9158–9165.

(125) Aragay, G.; Hernández, D.; Verdejo, B.; Escudero-Adán, E. C.; Martínez, M.; Ballester, P. Quantification of CH-π Interactions Using Calix[4]pyrrole Receptors as Model Systems. *Molecules* **2015**, *20* (9), 16672–16686.

(126) Mourik, T. Van. Ab Initio Calculations on the C-H. . -0 Hydrogen-Bonded Systems CHd-HZO, CH3NH2-H20 and CH3NHl-H20*. *J. Mol. Struct. THEOCHEM* **1995**, *341*, 63–73.

(127) Du, Q.; Wang, Z.; Schramm, V. L. Human DNMT1 Transition State Structure. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (11), 2916–2921.

(128) Linscott, J. A.; Kapilashrami, K.; Wang, Z.; Senevirathne, C.; Bothwell, I. R.; Blum, G.; Luo, M. Kinetic Isotope Effects Reveal Early Transition State of Protein Lysine Methyltransferase SET8. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (5), 1197–1201.

(129) Lienhard, G. E. Enzymatic Catalysis and Transition-State Theory. *Science (80-.).* **1973**, *180* (4082), 149–154.

(130) Hammond, S. A Correlation of Reaction Rates. *J. Am. Chem. Soc.* **1955**, *77* (3), 334.

(131) Wilson, P. B.; Williams, I. H. Influence of Equatorial CH···O Interactions on Secondary Kinetic Isotope Effects for Methyl Transfer. *Angew. Chemie - Int. Ed.* **2016**, *55* (9), 3192–3195.

(132) Sklenar, V.; Piotto, M.; Leppik, R.; Saudek, V. Gradient-Tailored Water Suppression for 1H-15N HSQC Experiments Optimized to Retain Full Sensitivity. *Journal of Magnetic Resonance, Series A*. 1993, pp 241–245.

(133) Kitagawa, M.; Ara, T.; Arifuzzaman, M.; Ioka-Nakamichi, T.; Inamoto, E.; Toyonaga, H.; Mori, H. Complete Set of ORF Clones of Escherichia Coli ASKA Library (A Complete Set of E. Coli K-12 ORF Archive): Unique Resources for Biological Research. *DNA Res.* **2005**, *12* (5), 291–299.

(134) Geoghegan, K. F.; Dixon, H. B. F.; Rosner, P. J.; Hoth, L. R.; Lanzetti, A. J.; Borzilleri, K. A.; Marr, E. S.; Pezzullo, L. H.; Martin, L. B.; LeMotte, P. K.; et al. Spontaneous α-N-6-Phosphogluconoylation of a “His Tag” in Escherichia Coli: The Cause of Extra Mass of 258 or 178 Da in Fusion Proteins. *Anal. Biochem.* **1999**, *267* (1), 169–184.

(135) Peck, S. C.; van der Donk, W. A. Go It Alone: Four-Electron Oxidations by Mononuclear Non-Heme Iron Enzymes. *J. Biol. Inorg. Chem.* **2016**, 1–14.

(136) Kovaleva, E. G.; Lipscomb, J. D. Versatility of Biological Non-Heme Fe(II) Centers in Oxygen Activation Reactions. *Nat. Chem. Biol.* **2008**, *4* (3), 186–193.

(137) Joseph, C. a.; Maroney, M. J. Cysteine Dioxygenase: Structure and Mechanism. *Chem. Commun.* **2007**, No. 32, 3338–3349.

(138) Seebeck, F. P. Thiohistidine Biosynthesis. *Chimia (Aarau).* **2013**, *67* (5), 333–336.

(139) Heafield, M. T.; Fearn, S.; Steventon, G. B.; Waring, R. H.; Williams, A. C.; Sturman, S. G. Plasma Cysteine and Sulphate Levels in Patients with Motor Neurone, Parkinson's and Alzheimer's Disease. *Neurosci. Lett.* **1990**, *110* (1-2), 216–220.

(140) McCoy, J. G.; Bailey, L. J.; Bitto, E.; Bingman, C. a; Aceti, D. J.; Fox, B. G.; Phillips, G. N. Structure and Mechanism of Mouse Cysteine Dioxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (9), 3084–3089.

(141) Ito, N.; Phillips, S. E. V.; Stevens, C.; Ogel, Z. B.; McPherson, M. J.; Keen, J. N.; Yadav, K. D. S.; Knowles, P. F. Novel Thioether Bond Revealed by a 1.7 Å Crystal Structure of Galactose Oxidase. *Nature* **1991**, *350* (6313), 87–90.

(142) Pierce, B. S. .; Gardner, J. D. .; Bailey, L. J. .; Brunold, T. C. .; Fox, B. G. Characterization of the Nitrosyl Adduct of Substrate-Bound Mouse Cysteine Dioxygenase by Electron Paramagnetic Resonance: Electronic Structure of the Active Site and Mechanistic Implications. *Biochemistry* **2007**, *46*, 8569–8578.

(143) Li, W.; Blaesi, E. J.; Pecore, M. D.; Crowell, J. K.; Pierce, B. S. Second-Sphere Interactions between the C93-Y157 Cross-Link and the Substrate-Bound Fe Site Influence the O2 Coupling Efficiency in Mouse Cysteine Dioxygenase. *Biochemistry* **2013**, *52* (51), 9104–9119.

(144) Tchesnokov, E. P.; Faponle, A. S.; Davies, C. G.; Quesne, M. G.; Turner, R.; Fellner, M.; Souness, R. J.; Wilbanks, S. M.; de Visser, S. P.; Jameson, G. N. L. An Iron–oxygen Intermediate Formed during the Catalytic Cycle of Cysteine Dioxygenase. *Chem. Commun.* **2016**, *52* (57), 8814–8817.

(145) Mashabela, G. T. M.; Seebeck, F. P. Substrate Specificity of an Oxygen Dependent Sulfoxide Synthase in Ovothiol Biosynthesis. *Chem. Commun. (Camb).* **2013**, *49* (70), 7714–7716.

(146) Goncharenko, K. V.; Vit, A.; Blankenfeldt, W.; Seebeck, F. P. Structure of the Sulfoxide Synthase EgtB from the Ergothioneine Biosynthetic Pathway. *Angew. Chemie Int. Ed.* **2015**, *54*, 2821–2824.

(147) Bushnell, E. A. C.; Fortowsky, G. B.; Gauld, J. W. Model Iron-Oxo Species and the Oxidation of Imidazole: Insights into the Mechanism of OvoA and EgtB? *Inorg. Chem.* **2012**, *51* (24), 13351–13356.

(148) Wei, W.-J.; Siegbahn, P. E. M.; Liao, R.-Z. Theoretical Study of the Mechanism of the Nonheme Iron Enzyme EgtB. *Inorg. Chem.* **2017**, acs.inorgchem.6b03177.

(149) Goncharenko, K. The Catalytic Mechanism of the Iron-Dependent Sulfoxide Synthase EgtB, University of Basel, 2017.

(150) Goncharenko, K. V.; Seebeck, F. P. Conversion of a Non-Heme Iron-Dependent Sulfoxide Synthase into a Thiol Dioxygenase by a Single Point Mutation. *Chem. Commun.* **2016**, *52*, 1945–1948.

(151) Goncharenko, K.; Seebeck, F. P. *Unpublished*; 2017.

(152) Nagai, W.; Kirk, K. L.; Cohen, L. A. Synthesis of 2-Amino-L-Histidine and 2-Aminohistamine. *J. Org. Chem.* **1973**, *38* (71), 1971–1974.

(153) Kirk, K. L.; Nagai, W.; Cohen, L. A. Photochemistry of Diazonium Salts. II. Synthesis of 2-Fluoro-L-Histidine and 2-Fluorohistamine, and the Halogen Lability of 2-Fluoroimidazoles. *J. Am. Chem. Soc.* **1973**, *95* (25), 8389–8392.

(154) Jaganathen, A.; Ehret-sabatier, L.; Bouchet, M.; Goeldner, M. P.; Hirth, C. G. Synthesis and Properties of a 2-Diazohistidine Derivative: A New Photoactivatable Aromatic Amino-Acid Analog. *Helv. Chim. Acta* **1990**, *73*, 86–96.

(155) Hay, M. P.; Wilson, W. R.; Denny, W. A. Design , Synthesis and Evaluation of Imidazolylmethyl Carbamate Prodrugs of Alkylating Agents. *Tetrahedron* **2000**, *56*, 645–657.

(156) Thierry, J.; Yue, C.; Potier, P. 2-Phenyl Isopropyl and T-Butyl Trichloroacetimidates: Useful Reagents for Ester Preparation of N-Protected Amino Acids under Neutral Conditions. *Tetrahedron Lett* **1998**, *39*, 1557–1560.

(157) Mukherjee, S.; Sivappa, R.; Yousufuddin, M.; Lovely, C. An Approach to the Oxazoline-Containing Fragments of the Oroidin Dimers Nagelamide R and T. *Synlett* **2010**, *2010* (05), 817–821.

(158) Jaouhari, R. Improved Efficiency and Selectivity in Peptide Synthesis : Use of Triethylsilane as a Carbocation Scavenger in Deprotection of T-Butyl Esters and T-Butoxycarbonyl-Protected Sites. *Tetrahedron Lett* **1992**, *33* (37), 5441–5444.

(159) Koswatta, P.; Sivappa, R.; Dias, H.; Lovely, C. Total Synthesis of the Leucetta-Derived Alkaloid Calcaridine A. *Synthesis (Stuttg).* **2009**, *2009* (17), 2970–2982.

(160) Ono, T.; Hirohato, R. Synthese Des Ophidins. *Hoppe-Seyler's Zeitschrift für Physiol. Chemie* **1956**, *304*, 77–81.

(161) Lineweaver, H.; Burk, D. The Determination of Enzyme Dissociation Constants. *J. Am. Chem. Soc.* **1934**, *56* (3), 658–666.

(162) Dixon, M.; Webb, E. C. *Enzymes*, Third Edit.; Longman, USA, 1979.

(163) Storey, B. T.; Sullivan, W. W.; Moyer, C. L. The pKa Values of Some 2-Amino Imidazolium Ions. *J. Org. Chem.* **1964**, *29* (1), 3118–3120.

(164) The Human Metabolome Database http://www.hmdb.ca/spectra/nmr_one_d/1674 (accessed May 29, 2017).

(165) Peer Lukat. *Personal Communication*; 2017.

(166) Batsanov, S. S. Van Der Waals Radii of Elements. *Inorg. Mater.* **2001**, *37* (9), 871–885.

(167) Mason, P. E.; Brady, J. W. "Tetrahedrality" and the Relationship between Collective Structure and Radial Distribution Functions in Liquid Water. *J. Phys. Chem. B* **2007**, *111* (20), 5669–5679.

(168) Ilies, M.; Di Costanzo, L.; North, M. L.; Scott, J. a; Christianson, D. W. 2-Aminoimidazole Amino Acids as Inhibitors of the Binuclear Manganese Metalloenzyme Human Arginase I. *J. Med. Chem.* **2010**, *53* (10), 4266–4276.

(169) Walba, H.; Isensee, R. W. Acidity Constants of Some Arylimidazoles and Their Cations. *J. Org. Chem.* **1961**, *26*, 2789–2791.

(170) Kirby, A. H. M.; Neuberger, A. Glyoxalines: The Determination of Their pK Values and the Use of Their Salts as Buffers. *Biochem. J.* **1938**, *32* (7), 1146–1151.

(171) Dawson, R. M. C. et al. *Data for Biochemical Research*; Clarendon Press: Oxford, 1959.

(172) Hilvert, D. Nearer to Nature: Design and Optimization of Artificial Enzymes. *FASEB J.* **2016**, *30* (1 Supplement), 395.1–395.1.

(173) Obexer, R.; Godina, A.; Garrabou, X.; Mittl, P. R. E.; Baker, D.; Griffiths, A. D.; Hilvert, D. Emergence of a Catalytic Tetrad during Evolution of a Highly Active Artificial Aldolase. *Nat. Chem.* **2016**, *9* (August 2016).

(174) Toscano, M. D.; Woycechowsky, K. J.; Hilvert, D. Minimalist Active-Site Redesign: Teaching Old Enzymes New Tricks. *Angew. Chemie - Int. Ed.* **2007**, *46* (18), 3212–3236.

(175) Song, H.; Her, A. S.; Raso, F.; Zhen, Z.; Huo, Y.; Liu, P. Cysteine Oxidation Reactions Catalyzed by a Mononuclear Non- Heme Iron Enzyme (OvoA) in Ovothiol Biosynthesis. *Org. Lett.* **2014**, *16* (1), 2122–2125.

(176) Preusser-Kunze, A.; Mariappan, M.; Schmidt, B.; Gande, S. L.; Mutenda, K.; Wenzel, D.; Von Figura, K.; Dierksl, T. Molecular Characterization of the Human Ca-Formylglycine-Generating Enzyme. *J. Biol. Chem.* **2005**, *280* (15), 14900–14910.

(177) Dierks, T.; Dickmanns, A.; Preusser-Kunze, A.; Schmidt, B.; Mariappan, M.; von Figura, K.; Ficner, R.; Rudolph, M. G. Molecular Basis for Multiple Sulfatase Deficiency and Mechanism for Formylglycine Generation of the Human Formylglycine-Generating Enzyme. *Cell* **2005**, *121* (4), 541–552.

(178) Hayyan, M.; Hashim, M. A.; Alnashef, I. M. Superoxide Ion: Generation and Chemical Implications. *Chem. Rev.* **2016**, *116* (5), 3029–3085.

(179) Jornet, D.; Bartovský, P.; Domingo, L. R.; Tormos, R.; Miranda, M. A. Experimental and Theoretical Studies on the Mechanism of Photochemical Hydrogen Transfer from 2-Aminobenzimidazole to Nπ* and ππ*aromatic Ketones. *J. Phys. Chem. B* **2010**, *114* (36), 11920–11926.

(180) de Visser, S. P.; Faponle, A. S. *Unpublished*; 2017.

(181) Tsai, J.; Taylor, R.; Chothia, C.; Gerstein, M. The Packing Density in Proteins: Standard Radii and Volumes. *J. Mol. Biol.* **1999**, *290* (1), 253–266.

(182) Aluri, S.; De Visser, S. P. The Mechanism of Cysteine Oxygenation by Cysteine Dioxygenase Enzymes. *J. Am. Chem. Soc.* **2007**, *129* (48), 14846–14847.

(183) Kumar, D.; Thiel, W.; De Visser, S. P. Theoretical Study on the Mechanism of the Oxygen Activation Process in Cysteine Dioxygenase Enzymes. *J. Am. Chem. Soc* **2011**, *133* (Scheme 1), 3869–3882.

(184) Wang, D.; Zhang, M.; Bühlmann, P.; Que, L. Redox Potential and C-H Bond Cleaving Properties of a Nonheme Fe IV=O Complex in Aqueous Solution. *J. Am. Chem. Soc.* **2010**, *132* (22), 7638–7644.

(185) Shiga, T.; Imaizumi, K. Electron Spin Resonance Study on Peroxidase- and Oxidase-Reactions of Horse Radish Peroxidase and Methemoglobin. *Arch. Biochem. Biophys.* **1975**, *479* (12), 469–479.

(186) Driggers, C. M.; Cooley, R. B.; Sankaran, B.; Hirschberger, L. L.; Stipanuk, M. H.; Karplus, P. A. Cysteine Dioxygenase Structures from pH 4 to 9: Consistent Cys-Persulfenate Formation at Intermediate pH and a Cys-Bound Enzyme at Higher pH. *J. Mol. Biol.* **2013**, *425* (17), 3121–3136.

(187) Bolton, J. L.; McClelland, R. A. Kinetics and Mechanism of the Decomposition in Aqueous Solutions of 2-(hydroxyamino)imidazoles. *J. Am. Chem. Soc.* **1989**, *111* (21), 8172–8181.

(188) Reed, M. J.; Purohit, A.; Woo, L. W. L.; Newman, S. P.; Potter, B. V. L. Steroid Sulfatase: Molecular Biology, Regulation, and Inhibition. *Endocr. Rev.* **2005**, *26* (2), 171–202.

(189) Popper, Z. A.; Michel, G.; Herve, C.; Domozych, D. S.; Willats, W. G.; Tuohy, M. G.; Kloareg, B.; Stengel, D. B. Evolution and Diversity of Plant Cell Walls: From Algae to Flowering Plants. *Annu Rev Plant Biol* **2011**, *62*, 567–590.

(190) Lerouge, P.; Roche, P.; Faucher, C.; Maillet, F.; Truchet, G.; Promé, J. C.; Dénarié, J. Symbiotic Host-Specificity of Rhizobium Meliloti Is Determined by a Sulphated and Acylated Glucosamine Oligosaccharide Signal. *Nature*. 1990, pp 781–784.

(191) Müller, I.; Kahnert, A.; Pape, T.; Sheldrick, G. M.; Meyer-Klaucke, W.; Dierks, T.; Kertesz, M.; Usón, I. Crystal Structure of the Alkylsulfatase AtsK: Insights into the Catalytic Mechanism of the Fe(II) ??-Ketoglutarate-Dependent Dioxygenase Superfamily. *Biochemistry* **2004**, *43* (11), 3075–3088.

(192) Barbeyron, T.; Brillet-Gu??guen, L.; Carr??, W.; Carri??re, C.; Caron, C.; Czjzek, M.; Hoebeke, M.; Michel, G. Matching the Diversity of Sulfated Biomolecules: Creation of a Classification Database for Sulfatases Reflecting Their Substrate Specificity. *PLoS One* **2016**, *11* (10), 1–33.

(193) Dierks, T.; Schmidt, B.; von Figura, K. Conversion of Cysteine to Formylglycine: A Protein Modification in the Endoplasmic Reticulum. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (October), 11963–11968.

(194) Knaust, A.; Schmidt, B.; Dierks, T.; Von Bülow, R.; Von Figura, K. Residues Critical for Formylglycine Formation And/or Catalytic Activity of Arylsulfatase A. *Biochemistry* **1998**, *37* (40), 13941–13946.

(195) Miech, C.; Dierks, T.; Selmer, T.; Von Figura, K.; Schmidt, B. Arylsulfatase from Klebsiella Pneumoniae Carries a Formylglycine Generated from a Serine. *J. Biol. Chem.* **1998**, *273* (9), 4835–4837.

(196) Marquordt, C.; Fang, Q.; Will, E.; Peng, J.; Von Figura, K.; Dierks, T. Posttranslational Modification of Serine to Formylglycine in Bacterial Sulfatases: Recognition of the Modification Motif by the Iron-Sulfur Protein AtsB. *J. Biol. Chem.* **2003**, *278* (4), 2212–2218.

(197) Bojarová, P.; Williams, S. J. Sulfotransferases, Sulfatases and Formylglycine-Generating Enzymes: A Sulfation Fascination. *Curr. Opin. Chem. Biol.* **2008**, *12* (5), 573–581.

(198) Dierks, T.; Schmidt, B.; Borissenko, L. V; Peng, J.; Preusser, A.; Mariappan, M.; von Figura, K. Multiple Sulfatase Deficiency Is Caused by Mutations in the Gene Encoding the Human

C(alpha)-Formylglycine Generating Enzyme. *Cell* **2003**, *113*, 435–444.

(199) Cosma, M. P.; Pepe, S.; Annunziata, I.; Newbold, R. F.; Grompe, M.; Parenti, G.; Ballabio, A. The Multiple Sulfatase Deficiency Gene Encodes an Essential and Limiting Factor for the Activity of Sulfatases. *Cell* **2003**, *113* (4), 445–456.

(200) NIH - U.S. National Library of Medicine, Genetics Home Reference https://ghr.nlm.nih.gov/condition/multiple-sulfatase-deficiency (accessed Mar 28, 2017).

(201) Schlotawa, L.; Radhakrishnan, K.; Baumgartner, M.; Schmid, R.; Schmidt, B.; Dierks, T.; Gärtner, J. Rapid Degradation of an Active Formylglycine Generating Enzyme Variant Leads to a Late Infantile Severe Form of Multiple Sulfatase Deficiency. *Eur. J. Hum. Genet.* **2013**, *21* (9), 1020–1023.

(202) Schlotawa, L.; Steinfeld, R.; von Figura, K.; Dierks, T.; Gärtner, J. Molecular Analysis of SUMF1 Mutations: Stability and Residual Activity of Mutant Formylglycine-Generating Enzyme Determine Disease Severity in Multiple Sulfatase Deficiency. *Hum. Mutat.* **2008**, *29* (1), 205.

(203) Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R. Introducing Genetically Encoded Aldehydes into Proteins. *Nat. Chem. Biol.* **2007**, *3* (6), 321–322.

(204) Rush, J.; Bertozzi, C. R. An Alpha-Formylglycine Building Block for Fmoc-Based Solid-Phase Peptide Synthesis. *Org. Lett.* **2006**, *8* (1), 131–134.

(205) Wu, P.; Shui, W.; Carlson, B. L.; Hu, N.; Rabuka, D.; Lee, J.; Bertozzi, C. R. Site-Specific Chemical Modification of Recombinant Proteins Produced in Mammalian Cells by Using the Genetically Encoded Aldehyde Tag. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (9), 3000–3005.

(206) Rabuka, D.; Rush, J. S.; deHart, G. W.; Wu, P.; Bertozzi, C. R. Site-Specific Chemical Protein Conjugation Using Genetically Encoded Aldehyde Tags. *Nat. Protoc.* **2012**, *7* (6), 1052–1067.

(207) Lobner, E.; Traxlmayr, M. W.; Obinger, C.; Hasenhindl, C. Engineered IgG1-Fc - One Fragment to Bind Them All. *Immunol. Rev.* **2016**, *270* (1), 113–131.

(208) Hudak, J. E.; Yu, H. H.; Bertozzi, C. R. Protein Glycoengineering Enabled by the Versatile Synthesis of Aminooxy Glycans and the Genetically Encoded Aldehyde Tag. *J. Am. Chem. Soc.* **2011**, *133* (40), 16127–16135.

(209) Smith, E. L.; Giddens, J. P.; Iavarone, A. T.; Godula, K.; Wang, L. X.; Bertozzi, C. R. Chemoenzymatic Fc Glycosylation via Engineered Aldehyde Tags. *Bioconjug. Chem.* **2014**, *25* (4), 788–795.

(210) Jian, H.; Wang, Y.; Bai, Y.; Li, R.; Gao, R. Site-Specific, Covalent Immobilization of Dehalogenase ST2570 Catalyzed by Formylglycine-Generating Enzymes and Its Application in Batch and Semi-Continuous Flow Reactors. *Molecules* **2016**, *21* (7).

(211) Liang, S. I.; McFarland, J. M.; Rabuka, D.; Gartner, Z. J. A Modular Approach for Assembling Aldehyde-Tagged Proteins on DNA Scaffolds. *J. Am. Chem. Soc.* **2014**, *136* (31), 10850–10853.

(212) Shi, X.; Jung, Y.; Lin, L.-J.; Liu, C.; Wu, C.; Cann, I. K. O.; Ha, T. Quantitative Fluorescence Labeling of Aldehyde-Tagged Proteins for Single-Molecule Imaging. *Nat. Methods* **2012**, *9*

(5), 499–503.

(213) Lemnaru, R. A Chemoenzymatic Strategy for Protein-Nanocellulose Conjugates, University of Basel, 2016.

(214) Chari, R. V. J.; Miller, M. L.; Widdison, W. C. Antibody-Drug Conjugates: An Emerging Concept in Cancer Therapy. *Angew. Chemie - Int. Ed.* **2014**, *53* (15), 3796–3827.

(215) Agarwal, P.; Bertozzi, C. R. Site-Specific Antibody-Drug Conjugates: The Nexus of Bioorthogonal Chemistry, Protein Engineering, and Drug Development. *Bioconjug. Chem.* **2015**, *26* (2), 176–192.

(216) Agarwal, P.; van der Weijden, J.; Sletten, E. M.; Rabuka, D.; Bertozzi, C. R. A Pictet-Spengler Ligation for Protein Chemical Modification. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (1), 46–51.

(217) Agarwal, P.; Kudirka, R.; Albers, A. E.; Barfield, R. M.; De Hart, G. W.; Drake, P. M.; Jones, L. C.; Rabuka, D. Hydrazino-Pictet-Spengler Ligation as a Biocompatible Method for the Generation of Stable Protein Conjugates. *Bioconjug. Chem.* **2013**, *24* (6), 846–851.

(218) Drake, P. M.; Albers, A. E.; Baker, J.; Banas, S.; Barfield, R. M.; Bhat, A. S.; De Hart, G. W.; Garofalo, A. W.; Holder, P.; Jones, L. C.; et al. Aldehyde Tag Coupled with HIPS Chemistry Enables the Production of ADCs Conjugated Site-Specifically to Different Antibody Regions with Distinct in Vivo Efficacy and PK Outcomes. *Bioconjug. Chem.* **2014**, *25* (7), 1331–1341.

(219) Catalent Biologics http://biologics.catalent.com (accessed Apr 7, 2017).

(220) Roeser, D.; Preusser-Kunze, A.; Schmidt, B.; Gasow, K.; Wittmann, J. G.; Dierks, T.; von Figura, K.; Rudolph, M. G. A General Binding Mechanism for All Human Sulfatases by the Formylglycine-Generating Enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (1), 81–86.

(221) Holder, P. G.; Jones, L. C.; Drake, P. M.; Barfield, R. M.; Bañas, S.; De Hart, G. W.; Baker, J.; Rabuka, D. Reconstitution of Formylglycine-Generating Enzyme with copper(II) for Aldehyde Tag Conversion. *J. Biol. Chem.* **2015**, *290* (25), 15730–15745.

(222) Knop, M.; Engi, P.; Lemnaru, R.; Seebeck, F. P. In Vitro Reconstitution of Formylglycine-Generating Enzymes Requires Copper(I). *ChemBioChem* **2015**, *16* (15), 2147–2150.

(223) Knop, M.; Dang, T. Q. Q.; Jeschke, G.; Seebeck, F. P. Copper Is a Cofactor of the Formylglycine Generating Enzyme. *ChemBioChem* **2016**, 161–165.

(224) Meury, M.; Knop, M.; Seebeck, F. P. Structural Basis for Copper-Oxygen Mediated C-H Bond Activation by the Formylglycine-Generating Enzyme. *Angew. Chemie Int. Ed.* **2017**.

(225) Benjdia, A.; Dehò, G.; Rabot, S.; Berteau, O. First Evidences for a Third Sulfatase Maturation System in Prokaryotes from E. Coli aslB and ydeM Deletion Mutants. *FEBS Lett.* **2007**, *581* (5), 1009–1014.

(226) Benjdia, A.; Subramanian, S.; Leprince, J.; Vaudry, H.; Johnson, M. K.; Berteau, O. Anaerobic Sulfatase-Maturating Enzymes, First Dual Substrate Radical S-Adenosylmethionine Enzymes. *J. Biol. Chem.* **2008**, *283* (26), 17815–17826.

(227) Benjdia, A.; Leprince, J.; Sandström, C.; Vaudry, H.; Berteau, O. Mechanistic Investigations of Anaerobic Sulfatase-Maturating Enzyme: Direct Cbeta H-Atom Abstraction Catalyzed by

a Radical AdoMet Enzyme. *J. Am. Chem. Soc.* **2009**, *131* (24), 8348–8349.

(228) Grove, T. L.; Ahlum, J. H.; Qin, R. M.; Lanz, N. D.; Radle, M. I.; Krebs, C.; Booker, S. J. Further Characterization of Cys-Type and Ser-Type Anaerobic Sulfatase Maturating Enzymes Suggests a Commonality in the Mechanism of Catalysis. *Biochemistry* **2013**, *52* (17), 2874–2887.

(229) Klinman, J. P. The Power of Integrating Kinetic Isotope Effects into the Formalism of the Michaelis-Menten Equation. *FEBS J.* **2014**, *281* (2), 489–497.

(230) Klinman, J. P. Kinetic Isotope Effects in Enzymology. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1978**, *46*, 415–494.

(231) Rickert, K. W.; Klinman, J. P. Nature of Hydrogen Transfer in Soybean Lipoxygenase 1: Separation of Primary and Secondary Isotope Effects. *Biochemistry* **1999**, *38* (38), 12218–12228.

(232) Wiberg, K. B. The Deuterium Isotope Effect. *Chem. Rev.* **1955**, *55*, 713–743.

(233) Glickman, M. H.; Wiseman, J. S.; Klinman, J. P. Extremely Large Isotope Effects in the Soybean Lipoxygenase-Linoleic Acid Reaction. *J. Am. Chem. Soc.* **1994**, *116*, 793–794.

(234) Simmons, E. M.; Hartwig, J. F. On the Interpretation of Deuterium Kinetic Isotope Effects in C-H Bond Functionalizations by Transition-Metal Complexes. *Angew. Chemie - Int. Ed.* **2012**, *51* (13), 3066–3072.

(235) Northrop, D. B. The Expression of Isotope Effects of Enzyme-Catalysed Reactions. *Annu. Rev. Biochem.* **1981**, *50*, 103–131.

(236) Gao, J. Personal Communication. p Unpublished data.

(237) Jeanguenat, A.; Seebach, D. Stereoselective Chain Elongation at C-3 of Cysteine through 2,3-Dihydrothiazoles, without Racemization. Preparation of 2-Amino-5-Hydroxy-3-Mercaptoalkanoic Acid Derivatives. *J. Chem. Soc. Perkin Trans. 1* **1991**, No. 10, 2291–2298.

(238) Pattenden, G.; Thom, S. M.; Jones, M. F. Enantioselective Synthesis of 2-Alkyl Substituted Cysteines .*. *Tetrahedron* **1993**, *49* (10), 2131–2138.

(239) Strauss, E.; Begley, T. P. Stereochemical Studies on Phosphopantothenoylcysteine Decarboxylase from Escherichia Coli. *Bioorg. Med. Chem. Lett.* **2003**, *13* (3), 339–342.

(240) Corey, E. J.; Bock, M. G.; Kozikowski, A. P.; Floyd, D.; Lipshutz, B. A Key Intermediate for the Synthesis of Matansine and Related Antitumor Agents. *Tetrahedron Lett* **1978**, No. 12, 1051–1054.

(241) Ando, W.; Igarashi, Y.; Huang, L. Acylation of 2, 4-Disubstituted Thiazolidines and Oxazolidines. Concomitant Epimerization at C-2. *Chem. Lett.* **1987**, No. C, 1361–1364.

(242) Davis, A. M.; Layland, N. J.; Page, M. I.; Martin, F.; O'Ferrall, R. M. Thiazolidine Ring Opening in Penicillin Derivatives. Part 2. Enamine Formation. *J. Chem. Soc. Perkin Trans. 2* **1991**, No. 8, 1225.

(243) Fife, T. H.; Natarajan, R.; Shen, C. C.; Bembi, R. Mechanism of Thiazolidine Hydrolysis. Ring Opening and Hydrolysis of 1, 3-Thiazolidine Derivatives of P-(dimethylamino)

Cinnamaldehyde. *J. Am. …* **1991**, *113* (8), 3071–3079.

(244) Ando, W.; Huang, L. Asymmetric Induction to Sulfur Atom: Stereocontrolled S-Oxidation of Thiazolidines. *Tetrahedron Lett* **1986**, *27* (29), 3391–3394.

(245) Pummerer, R. Über Phenyl-Sulfoxyessigsäure. *Chem. Ber.* **1909**, *42* (2), 2282–2291.

(246) Pummerer, R. Über Phenylsulfoxy-Essigsäure. *Chem. Ber.* **1910**, *43* (2), 1401–1412.

(247) Ando, W.; Tokitoh, N.; Igarashi, Y. Silicon Pummerer Reaction of Thiazolidine S-Oxides; A New Method For Stereospecific C-5 Functionalization of Thiazolidines. *Tetrahedron Lett* **1987**, *28* (47), 5903–5906.

(248) Mougous, J. D.; Green, R. E.; Williams, S. J.; Brenner, S. E.; Bertozzi, C. R. Sulfotransferases and Sulfatases in Mycobacteria. *Chem. Biol.* **2002**, *9* (7), 767–776.

(249) Knop, M. Identification and Characterization of a Novel Copper Dependent Enzyme, University of Basel, 2017.

(250) Del Rio, T. G.; Tice, H.; Cheng, J.; Goodwin, L.; Pitluck, S. Complete Genome Sequence of Thermomonospora Curvata Type. *Stand. Genomic Sci.* **2011**, *4*, 13–22.

(251) Minasian, S. G.; Whittaker, M. M.; Whittaker, J. W. Stereoselective Hydrogen Abstraction by Galactose Oxidase. *Biochemistry* **2004**, *43* (43), 13683–13693.

(252) Wallace Cleland, W. Use of Isotope Effects to Elucidate Enzyme Mechanism. *Crit. Rev. Biochem.* **2008**, *13* (4), 385–428.

(253) Frantom, P. a; Pongdee, R.; Sulikowski, G. a; Fitzpatrick, P. F. Intrinsic Deuterium Isotope Effects on Benzylic Hydroxylation by Tyrosine Hydroxylase. *J. Am. Chem. Soc.* **2002**, *124* (16), 4202–4203.

(254) Bugg, T. D. H. *Introduction to Enzyme and Coenzyme Chemistry*, Third Edit.; John Wiley & Sons Limited: Chichester UK, 2012.

(255) Dang, T. Q. Engineering towards the Susbtrate Specificity of the Formylglycine Generating Enzyme (FGE), University of Basel, 2014.

(256) Chooi, K. P.; Galan, S. R. G.; Raj, R.; McCullagh, J.; Mohammed, S.; Jones, L. H.; Davis, B. G. Synthetic Phosphorylation of p38α Recapitulates Protein Kinase Activity. *J. Am. Chem. Soc.* **2014**, *136* (5), 1698–1701.

(257) Carlson, B. L.; Ballister, E. R.; Skordalakes, E.; King, D. S.; Breidenbach, M. a; Gilmore, S. a; Berger, J. M.; Bertozzi, C. R. Function and Structure of a Prokaryotic Formylglycine-Generating Enzyme. *J. Biol. Chem.* **2008**, *283* (29), 20117–20125.

(258) Whittaker, M. M.; Whittaker, J. W. The Active Site of Galactose Oxidase. *J. Biol. Chem.* **1988**, *263* (13), 6074–6080.

(259) Whittaker, M. M.; Ballou, D. P.; Whittaker, J. W. Kinetic Isotope Effects as Probes of the Mechanism of Galactose Oxidase †. *Biochemistry* **1998**, *2960* (37), 8426–8436.

(260) Fujii, T.; Yamaguchi, S.; Hirota, S.; Masuda, H. H-Atom Abstraction Reaction for Organic Substrates via Mononuclear Copper(ii)-Superoxo Species as a Model for DβM and PHM. *Dalt. Trans.* **2008**, No. 1, 164–170.

(261) Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities In the Course of the Routine Use of NMR as an Aid for Organic Chemistry , a Day-to-Day Problem Is the Identifica- Tion of Signals Deriving from Common Contaminants Literature , but the. *J. Org. Chem.* **1997**, *3263* (3), 7512–7515.

7 Appendix

7.1 General Experimental

All reagents used were purchased from commercial sources without further purification. All solvents used in reactions were purchased in HPLC-grade quality and used as such. Dry solvents were purchased in HPLC-grade quality and used as such. Chromatographic purifications (flash) were performed with SiliaFlash P60 from Silicycle (40-63 um; (230-400) mesh). Preparative RP-HPLC of peptides was performed on a Shimadzu system (SIL-10AD injector, LC-10AT liquid chromatograph, DGU-14A degasser, SDP-10AV UV-detector, SCL-10A system controller) with a Purospher® STAR RP-18 endcapped LiChroCART® column (250 x 10 mm, particle size 5 µm, Merck), flow rate 4.9 ml/min. Eluents were A: acetonitrile and B: water/acetonitrile/TFA (1000:10:1), peptides were detected by UV absorption at 310 nm. Analytical RP-HPLC was performed on an Agilent 1100 Series system with a LiChrospher® 100 RP-18e-column (250 x 4 mm, particle size 5 µm, Merck), flow rate was 1 ml/min. Eluents were A: acetonitrile and B: water/acetonitrile/TFA (1000:10:1), peptides were detected by UV absorption at 310 nm. UHPLC/MS-measurements were carried out on an Agilent 1290 Infinity system with a Zorbax Eclipse Plus C-18-column (2.1 x 50 mm; particle size 1.8 µm; Agilent), flow rate 0.4 ml/min, coupled to an Agilent 6130 quadrupol-MS. Eluents were A: acetonitrile and B: water/acetonitrile/TFA (1000:10:1). Compounds were detected at 254 nm. Reaction mixtures containing highly polar compounds were analyzed by cation-exchange HPLC (20 mM phosphoric acid pH 2 as the mobile phase) on a Luna 5u SCX column (100 Å, 150 x 4 mm, Phenomenex). The compounds were eluted in NaCl gradient. ESI-MS measurements were performed on a Bruker Esquire 3000plus, measuring in the positive ion mode. Centrifugation during peptide synthesis was carried out on an Allegra X-15R from Beckmann Coulter. pH values were measured on a Seven Easy pH meter from Mettler Toledo. Absorbance was measured on a Nanodrop 2000 Spectrophotometer from Thermo Scientific. Solid Phase Peptide Synthesis was performed on a Syro I from Biotage. NMR spectra were acquired on a Bruker 400 MHz or a Bruker 500 MHz instrument. ^{1}H and ^{13}C chemical shifts are quoted relative to solvent signals.[261] Multiplicity is reported as follows: s – singlet, d – doublet, dd – doublet of doublet, t – triplet, q – quartet, quin – quintet, s br – broad singlet nbd – non-binomial doublet and coupling constant *J* in Hz.

NMR Experiments 400 MHz

All NMR experiments were performed on a Bruker Avance III NMR spectrometer operating at 400 MHz proton frequency. The instrument was equipped with a direct observe 5-mm BBFO smart probe. The experiments were performed at 295 K and the temperature was calibrated using a methanol standard showing accuracy within +/- 0.2 K.

NMR Experiments 500 MHz

All NMR experiments were performed on a Bruker Avance III NMR spectrometer operating at 500 MHz proton frequency. The instrument was equipped with a indirect 5-mm BBI probe. The experiments were performed at 298 K and the temperature was calibrated using a methanol standard showing accuracy within +/- 0.2 K.

NMR Experiments 600 MHz cryo

All NMR experiments were performed on a Bruker Avance III HD four-channel NMR spectrometer operating at 600.27 MHz proton frequency. The instrument was equipped with a cryogenic 5mm four-channel QCI probe (H/C/N/F). with self-shielded z-gradient. The experiments were performed at 298 K and the temperature was calibrated using a methanol standard showing accuracy within +/- 0.2 K.

7.2 Abbreviations

2A-DMH	2-amino N^{α},N^{α}-dimethyl histidine
2M-DMH	2-methyl N^{α},N^{α}-dimethyl histidine
4-BMC	4-Bromomethyl-7-methoxycoumarin
Abz	Anthranilic acid
ATP	Adenosine triphosphate
CDO	Cysteine dioxygenase
DCM	Dichloromethane
DFT	Density functional theory
DIPEA	Diisopropylethylamine
DMF	Dimethylformamide
DMH	N^{α},N^{α}-dimethyl histidine
EDT	1,2-Ethanedithiol
EDTA	Ethylenediaminetetraacetic acid
$EgtB_{Th2}$	Sulfoxide synthase from *Candidatus chloracidobacterium thermophilum B*
$EgtB_{wt}$	Sulfoxide synthase from *Mycobacterium thermoresistibile*
ESI	Electrospray ionization
ESI-MS	Electrospray ionization mass spectrometry
ET	Ergothioneine
FGly	Formylglycine
Fmoc	Fluorenylmethyloxycarbonyl
HAT	Hydrogen atom transfer

HCTU	O-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
IE-HPLC	Ion exchange high performance liquid chromatography
KIE	Kinetic isotope effect
LDA	Lithium diisopropylamide
MMH	N^{α}-methyl histidine
NHI	Non-heme iron (enzyme)
NMR	Nuclear magnetic resonance
OA	Ovothiol A
QM/MM	Quantum mechanics/molecular modelling
SAH	*S*-Adenosyl homocysteine
SAM	*S*-Adenosyl methionine
SET	Single electron transfer
SPPS	Solid phase peptide synthesis
TBDMS	*tert*-Butyldimethylsilyl
TCEP	Tris(2-carboxyethyl)phosphine
TES	Triethylsilane
TFA	Trifluoroacetic acid
THF	Tetrahydrofurane
TIPS	Triidopropylsilane
TLC	Thin layer chromatography

TMH	N^{α},N^{α},N^{α}-trimethyl histidine
Trt	Triphenylmethyl
Ts	Tosyl
γGC	γ-Glutamyl cysteine

7.3 Acknowledgments

First and foremost I would like to thank Prof. Florian Seebeck for placing his trust in me during this truly great learning experience over the course of the last four and a half years. His creativity, enthusiasm and humor were always inspirational and made for a wonderful research environment.

Secondly I would like to thank PD Dr. Daniel Häussinger accepting to co-examine my thesis and especially for all the help with everything NMR-related and for making the project on EgtD the success it is. I am more than happy to have had him as my co-examiner and appreciate his patience in answering my countless questions.

I would also like to thank Prof. Dennis Gillingham for chairing my PhD defense and making fun of my New York Mets T-shirt.

A huge thank you goes out to all my lab mates during my time in Basel. This thesis would not have been possible without the high-quality enzymes I was supplied with by Dr. Matthias Knop, Dr. Kristina Goncharenko, Dr. Laëtitia Misson, Alice Harnacke and Dr. Cangsong Liao. Thanks is also due to Anja Stämpfli and Dr. Marcel Meury for answering every conceivable question I had about X-ray crystallography. Many thanks also to all my former students, Thanh Dang, David Kühne, Claudio Meyer and Julia Hildesheim, who did excellent work even when the project was not easy (or successful). And a general big thank you to the rest of the group, past and present: Dr. Jian Gao, Dr. Gabriel Mashabela, Dr. Sebastien Coyne, Dr. Ali Al-Alkaabi, Dr. Roxana Lemnaru, Alma Idrizovic, Dr. Davey Lim, Florian Leisinger and Dzmitry Miarzlou. Your support in all things chemical (and especially biochemical) was greatly appreciated. Despite my occasional onsets of grumpiness I had a great time and really enjoyed working in such an international group.

A very special thank you to everyone who provided me with constructive and critical feedback on my thesis. Sebastian Flückiger, Reto Burn, Dr. Marcel Hollenstein, Dr. Kristina Goncharenko and my father all gave me invaluable feedback, without which this thesis would not exist in its present form.

Thanks also to everyone else in the department who helped me with anything from chemistry to learning Greek (ευχαριστώ πολύ Νταλί!) to drinking beer and generally making for a good working atmosphere.

It goes without saying that my time in Basel would not have been nearly as enjoyable without my amazing friends Domi, Rafi and Oli, who helped me to relax and laugh when I needed it. I had

some of the best times of my life during our time together and look forward to more of the same. The same goes for my brother and best friend Phil, no one makes me laugh like he does and I could always count on his support and advice.

More than anything, I am grateful to my parents, from whom I learned the art of critical thinking and who formed the person I have become. All the love and support they have given me is more than anyone in their right mind could ask for and I proudly dedicate this thesis to both of them.

And last but not least, I would like to thank Domenika. My life would not be the same without her love, support and wonderful presence in it and I cannot imagine what I would do without her. She stood with me in trialing times and I am forever in her debt.