# On the emergence of the hemD-like fold and its use for fold-chimeragenesis

**Dissertation**

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Saacnicteh Toledo Patino
aus Morelia Mexiko

Tübingen 2019

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

| | |
|---|---|
| Tag der mündlichen Qualifikation: | 25.09.2019 |
| Dekan: | Prof. Dr. Wolfgang Rosenstiel |
| 1. Berichterstatter: | Prof. Dr. Birte Höcker |
| 2. Berichterstatter: | Prof. Dr. Thilo Stehle |

# Table of Contents

# Abstract (German)

Proteine sind für nahezu alle zellulären Prozesse verantwortlich. Selbst wenn ihre Vielfältigkeit an Grössen und Formen endlos zu sein erscheint, können Proteine jedoch in kleinere Einheiten namens Domäne unterteilt werden. Wie diese Domänen entstanden sind, ist zur Zeit nicht ganz klar. Mehrere Hypothesen gehen davon aus, dass sie aus der Kombination von kleineren Fragmenten resultieren., Obwohl Proteine global gesehen sehr unterschiedliche Formen annehmen, findet man tatsächlich lokale Ähnlichkeiten auf Sequenz- und Struktur-Ebene. Daher haben wir nach homologen Fragmenten unterschiedlicher Faltung gesucht und fanden, dass viele Domänen der α/β-Proteinklasse von SCOP homologe Fragmente von einem Dutzend bis >200 Aminosäuren teilen. Ein besonders interessanter Fall stellte die HemD-ähnliche Faltung dar, da ihre Sequenz-Profile auf Genduplikation, Insertion und Segment-Swap von einer Flavodoxin-ähnlichen Faltung hinwies. Um diese Hypothese zu testen, kehrte ich diese evolutiven Schritte experimentell um und fand, dass das mutmaßliche ursprüngliche Protein in der Tat die Flavodoxin-ähnliche Topologie aufwies. Diese Ergebnisse veranschaulichen die Strategie der Natur, bei der die Kombination einer reduzierten Anzahl an Fragmenten eine weitaus größere Vielfältigkeit an Strukturen ermöglicht. Unter der Annahme, dass homologe Fragmente sich leichter kombinieren lassen, stellte ich eine Faltungschimäre her. Dabei tauschte ich ein Fragment, welches Teil der Cobalamin bindenden Tasche der Flavodoxin-like Faltung enthielt und baute dieses in die HemD-ähnliche Faltung hinein. Die Proteinchimäre war gefaltet und wies native Eigenschaften auf, was die Aufklärung ihrer Struktur mittels Kristallographie ermöglichte. Bindungsanalysen wiesen darauf hin, dass die Faltungschimäre Cobalamin bindet. Allgemein zeigen die Ergebnisse, dass die Kombination homologer Fragmente ausgenutzt werden kann, um Bindungstaschen von komplexen Kofaktoren zu transferieren und damit das Design von neuen Katalysatoren zu designen.

# Abstract (English)

Whilst the structural diversity of proteins may appear endless, even large protein complexes can be decomposed into their independently folding units, the domains. Little is known about domain emergence. Structural and sequence evidence suggests they evolved through combination of subdomain-sized fragments. To investigate this hypothesis, we searched for homologous regions among domains with a broadly different topology (fold), employing the α/β-class as defined in the Structural Classification Of Proteins (SCOP), which is believed to contain the oldest domains. We compared their sequence profiles all-against-all and found that in spite of their globally different architectures a large number of them share local homologous regions ranging from a dozen to >200 amino acids. An interesting hit constitutes the hemD-like fold, whose profile alignments provide strong evidence for its emergence via flavodoxin-like gene duplication, insertion and segment-swapping. To test this hypothesis experimentally, we reverted these evolutionary events, finding that the obtained protein in fact adopts the canonical flavodoxin-like fold. These results illustrate a way how Nature recycles a limited repertoire of building blocks, which provides a successful strategy to reach diversity at a lower molecular cost than creating every unit *de novo*. Such building units may have overcome a selective pressure through the course of evolution due to their function and/or intrinsic stability that allowed them to be modified and extended. Inspired by this naturally occurring strategy, I designed a cobalamin-binding chimera, extracting a portion of the binding pocket of a cobalamin-binding domain and exchanged it against its homologous region in the hemD-like fold. The resulting chimera expresses solubly, is well folded and binds cobalamin, illustrating that mimicking Nature's combinatorial approach is a good source of soluble and well-folded proteins and may be employed as an alternative strategy to design novel functionalities.

# List of figures

## *Chapter 1*



## *Chapter 2*



## *Chapter 3*

## *Chapter 4*



## *Chapter 5*



## *Chapter 6*

# List of tables

### *Chapter 2*



### *Chapter 3*



### *Chapter 4*



### *Chapter 6*

# Abbreviations

| | |
|---|---|
| ***AdoCbl*** | adenosylcobalamin |
| ***BESSY*** | Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung |
| ***CarH*** | coenzyme B12-dependent photoreceptor |
| ***CATH*** | Protein Structure Classification Database |
| ***CD*** | circular dichroism |
| ***CNCbl*** | cyanocobalamin |
| ***CPs*** | circular permutations |
| ***DMB*** | dimethylbenzimidazole |
| ***FPLC*** | fast protein liquid chromatography |
| ***hemD*** | hemD-like fold |
| ***IPTG*** | Isopropyl β-D-thiogalactoside |
| ***ITC*** | isothermal titration calorimetry |
| ***MCoAM*** | methylmalonyl-CoA mutase |
| ***MeCbl*** | methylcobalamin |
| ***MeSyn*** | methionine synthase |
| ***MMACHC*** | methylmalonic aciduria cblC type with homocystinuria |
| ***OHCbl*** | hydroxocobalamin |
| ***PEG*** | polyethylenglycol |
| ***PSI*** | Paul Scherrer Institute |
| ***SCOP*** | Structural Classification Of Proteins |
| ***SSE*** | secondary structural elements |
| ***SSP*** | segment-swap |
| ***U3S*** | uroporphyrinogen III synthase |
| ***UShsMMap*** | chimeric protein built from human U3S and MCoAM from *aeropyrum pernix* |
| ***aR3P*** | α-ribazole-3-phosphate |

# Chapter 1

## Introduction

### 1.1. Protein emergence as RNA cofactors

Approximately 4.2 billion years ago, the first prebiotic molecules appeared to provide the chemical conditions that would initiate the most beautiful accident of nature: life (**Figure 1.1**). In the beginning of life, however, proteins did not play a major role as they do now. In fact, lacking complexity and self-replicating properties such these of RNA (Kiedrowski and von Kiedrowski 1986), small peptides may have assisted an RNA-organized world (Gilbert 1986; Gerald F. Joyce 1989), helping to expand the limited chemical functionalities of ribozymes (Roth and Breaker 1998; G. F. Joyce 1998) as for instance, chelating valuable metals to be used in catalysis. Thus, RNA and peptides may have cooperated closely to allow the synthesis of larger and more complex polymers, from which the most resistant overcame extinction through improvement of their replication/degradation rates. As simple as it may appear, via trial and error, proteins were optimized to become fundamental chemical agents that drive nearly all functions necessary to sustain life.

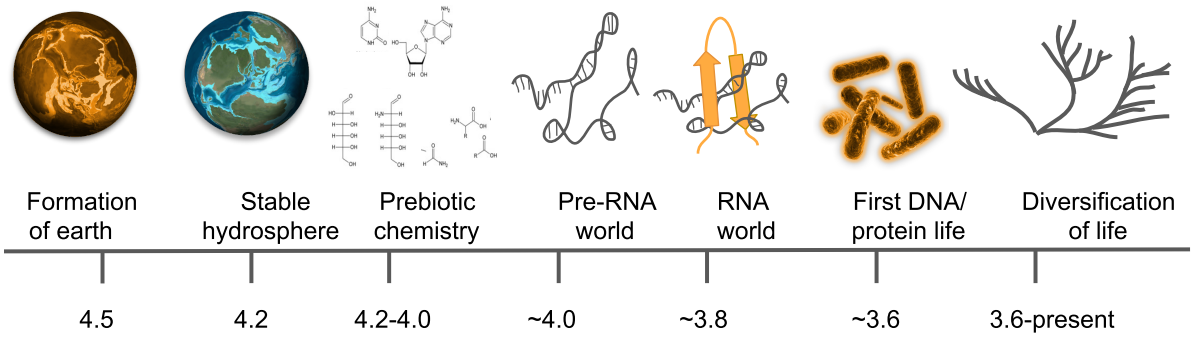



**Figure 1.1 Timeline of events pertaining to the early history of life on Earth, with approximate dates in billions of years before the present.** Adapted from "The antiquity of RNA-based evolution" by Gerald F. Joyce, Nature, 418, page 214 (2002).

## 1.2. From small peptides to proteins

The details on how small peptides evolved into larger and more sophisticated proteins remain mainly unsolved. However, footprints hidden in the sequences and structures of modern proteins provide hints about their origin. Margaret O. Dayhoff and Richard Eck were the first to recognize the internal symmetry of ferredoxin, suggesting that its periodic content may be the result of internal gene duplication of short peptides (Eck and Dayhoff 1966). Subsequently in 1973, repeating units ranging from 2 to 7 residues were described by Fraser and colleagues for proteins such as collagen, silk fibroin, keratin and tropomyosin (Jorda et al. 2010). Five years later, Andrew McLachlan described the three-fold repetition pattern in the soybean trypsin inhibitor (McLachlan 1979). These pioneer works established the groundwork for the following half century of studying the underlying principles of protein evolution, supporting the idea that early proteins had to be simple. Thus, in the absence of a fine-tuned translational machinery, the combination of small peptides may have been a good strategy to gain size at a low molecular cost. However, not only ancient proteins seem to have been created via this strategy. A recently evolved arctic cod fish antifreeze protein also emerged from repetition of a three-peptide repeat. These threonine-containing units allow glycosylation on the protein surface, preventing formation of ice latices (Baalsrud et al. 2017). Another type of recently evolved repeat proteins are solenoids (Kobe and Kajava 2000). They result from the duplication of structural α/α, β/β or α/β elements that belong to the so termed 'supersecondary structural elements', which have been described in detail (Murzin 1995, Koonin EV 2012, Thorton 1999, Russell 2002). The addition of tandem repeats usually yields linear arrays such as the *Iafp* protein or superhelicalical architectures as *HEAT*. The fact that these structures have large areas accessible to solvent, provide binding surfaces for large substrates. In some cases however, tandem repeats reach globular architectures such as these of the TIM-barrel and β-propellers (Andrade, Perez-Iratxeta, and Ponting 2001). The advantage of these circular arrangements is that they provide with relatively small binding areas for interaction with smaller ligands. Another advantage of globular proteins is that they are compact and commonly stable.

Previous studies have demonstrated that it is possible to recreate TIM-barrels by duplicating one of its halves (Höcker, Claren, and Sterner 2004) or even taking a fragment from the flavodoxin-like fold, suggesting that the TIM-barrel emerged through duplication and fusion of a smaller unit, which may have given rise to the flavodoxin-like fold in parallel (Farías-Rico, Schmidt, and Höcker 2014; Bharat et al. 2008; Coles et al. 2012). A recent study also recreated the evolution of β-propellers by duplicating and fusing one of its blades and subjecting it to divergence. The results demonstrated that the introduction of mutations had a positive impact on the protein stability and binding (Smock et al. 2016), which shows why perfect repeats are rarely found in Nature. However, other cases of proteins that evolved through gene duplication and kept high sequence identity of their repeating units can also be found. One example are the ferredoxins, which have been widely cited as evidence of gene duplication since their halves are almost identical and display periodic arrangements of cysteines (George et al. 1985, Bruschi and Guerlesquin 1988). Such proven cases of duplication are of great value since they indicate the existence of a vocabulary of fragments smaller than the protein domain from which many proteins may have derived and illustrate how their enlargement through duplication provided the opportunity to broaden cellular functions (**Figure 1.2**). These hypothesis correlates with the fact that even though repeats are common in all phylia, they are more represented in eukaryotic than in prokaryotic organisms (Marcotte et al. 1999).

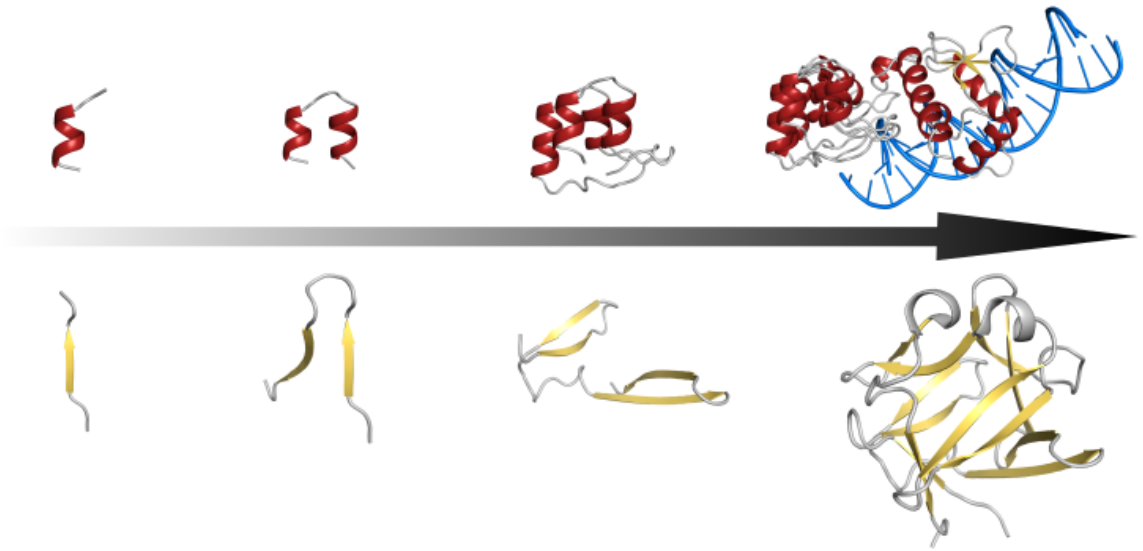


**Fig.1.2 Protein from pieces.** Duplication of small structural elements is a common evolutionary strategy to gain size and complexity, leading to larger architectures that are capable to perform novel functions not observed for their repeating units alone.

## 1.3. The domain as an evolutionary unit

A highly significant contribution to the understanding of proteins has been made by structural biology: X-Ray crystallography, nuclear magnetic resonance, and more recently Cryo-EM techniques. In the course of the last 60 years, more than 40 000 unique protein structures have been solved (Berman 2000), broadening our understanding on their functions and organization. In the light of this large amount of three-dimensional information, it was noticed that even large protein complexes could be decomposed into independently-folding units, termed domains (**Figure 1.3**). Each domain performs a certain function and this task is usually maintained even if the domain is extracted from its protein complex. In Nature, domains can be found as single units as well as part of larger polypeptide chains. A protein complex can contain two or more copies of the same unit or a combination of different ones. One example constitutes the vitamin B12-binding domain. This module expresses separately from its domain counterpart and later associates with it to build the glutamate mutase (Hoffmann et al. 1999). In contrast, in proteins such as the methionine synthase (Matthews, Koutmos, and Datta 2008) and the photoreceptor CarH (Z. Cheng, Yamamoto, and Bauer 2016; Jost et al. 2015), it is embedded in the same polypeptide chain adjacent to their coacting domains. Embedded or isolated, the vitamin B12-binding domain performs the same task in all these proteins, namely to bind cobalamin. However, the reaction outcome is different in all of them: methylation, isomerization and cobalamin-dependent oligomerization, respectively. The fact that a certain function can be reused and combined with others to reach diversity, postulated the domain as the evolutionary unit par excellence, and illustrated the principle of domain shuffling (Babushok, Ostertag, and Kazazian 2007; Chothia et al. 2003). After the discovery of domains and their crucial role in evolution, the next obvious question arose. Where do these domains come from? Are they related to each other or emerged independently? Recent research postulates that both events are possible, although the latter is much less common.

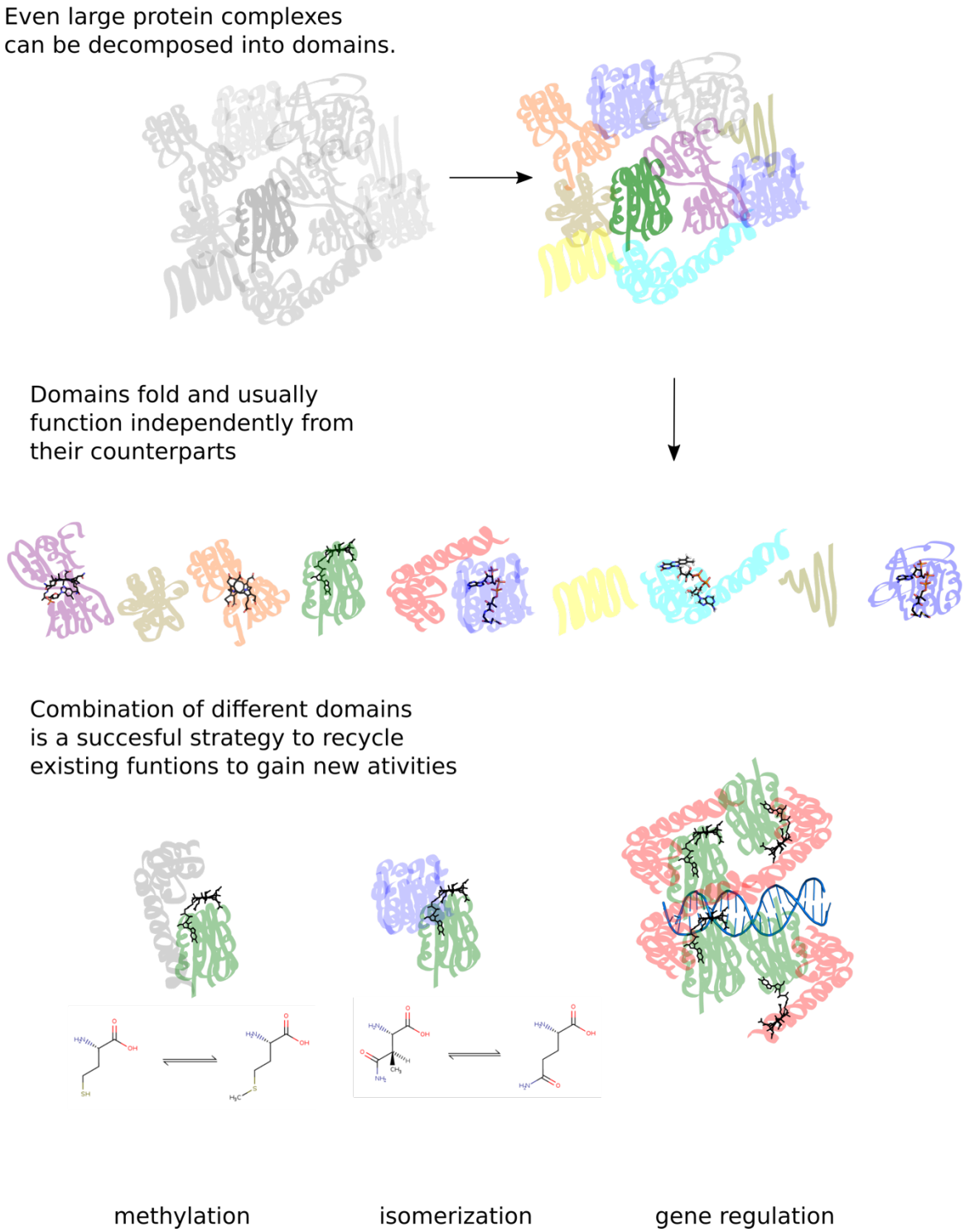



**Figure 1.3 The domain as the evolutionary unit of proteins.** Even large protein complexes can be decomposed into domains (represented in different colors). These domains usually function independently and can be combined with others to reach diversity. For instance, the cobalamin-binding domain (green) employs cobalamin to perform diverse functionalities.

## 1.4. Duplication and divergence as a common evolutionary strategy to achieve functional diversity

The most common type of evolutionary events through which domains emerge is gene duplication (Lupas, Ponting, and Russell 2001; Chothia and Gough 2009). Hereby, the duplicated unit is free to explore new functionalities without functional pressure, since the original copy maintains the essential function. Through mutations, the sequences of the derived domains change. Nevertheless, they usually keep similar architectures and topologies. This is the reason why structure is more conserved than sequence, which is also reflected in the large number of domains that adopt the same structure, even though they perform different activities. Particularly in bacteria and archaea, 35% of the enzymatic activities are carried out by only one structural domain (Orengo and Thornton 2005) and even though duplication events have been reported for all species, the animal kingdom is located at the top with 93% of estimated duplications, followed by fungi and bacteria with 85 and 50%, respectively (Chothia et al. 2003). Occasionally, duplicated domains experience circular permutations, a process by which a domain achieves a new topology while keeping its intact architecture. It is estimated that 5% of all domains resulted in this manner (Vogel and Morea 2006). Deletions and insertions also play an important role in the diversification of domains. For instance, adding decorations to a protein may facilitate the taking part in allosteric regulation and protein-protein interactions. However, insertions within the core of proteins are expected to affect their stability. Perhaps, this is the reason for why these events occur less frequently, accounting for one order of magnitude frequency reduction compared to mutations (Grishin 2001). Even though gene duplication and diversification are common processes in evolution, they do not provide an answer to: Where do domains come from? However, the recent identification of repeats and internal symmetry within domains provide hints about the existence of a subset of building blocks smaller than the domains and it has been postulated that these units may have been combined to create domains from ‘scratch’. I will discuss some of the most prominent examples in the following section.

## 1.5. Domain definition and classification in SCOP

The recognition of domains as structural building blocks is the basis of many protein classification methods such as the CATH Protein Structure Classification database (CATH) (Orengo et al. 1997), the Evolutionary Classification of Protein Domains (ECOD) (H. Cheng et al. 2014) and the Structural Classification Of Proteins (SCOP) (Murzin et al. 1995). In spite of their discrepancies in assignment, these classification methods agree upon that a domain constitutes an independently folding unit that can exist and evolve independently from the rest of the protein. A widely used compendium has been assessed by SCOP. Hereby proteins are divided into domains and each domain is sorted hierarchically into classes, folds, superfamilies, and families. Members of a given class are composed of the same secondary structural elements (SSE, e.g. all-α and all-β), or their combinations (α+β and α/β). Domains displaying the same SSE, architecture and topological connections are categorized in the same fold. Finally, domains within the same fold are further classified in superfamilies and families, thereby sharing similar functionalities and a common origin. The classification of domains is possible, because despite the large number of theoretical shapes they can adopt, only an infinitesimal fraction are indeed represented. On the one hand, the number of particles in the known universe would not be enough to try all structural combinations (even for a relatively small domain of ~100 amino acids) (Alva et al. 2010). On the other hand, rules of physicochemical nature govern their amino acid interactions, in that manner, limiting the 3D-structures they can adopt. Moreover, not all feasible structures are functional, therefore they would not pass the demanding selective functional pressure. Thus, it is not surprising that Nature has reused a set of optimized peptides as a good strategy to elaborate larger protein complexes. To date, SCOP recognizes 4919 superfamilies further sorted into 2026 families, which adopt only 1232 folds. Considering the much larger number of unique protein sequences on earth, this amount of shapes is still surprisingly small.

## 1.6. Difficulties in establishing domain boundaries

In order to sort domains, it is necessary to establish their boundaries. However, this is not an easy task. One complication is the discontinuous distribution of a protein chain along two domains, which in fact accounts for 20% of these units (Orengo 2007). In addition, domains can vary in their topology connections, resulting from circular permutation, which can lead to disagreement in their classification by different methods. These and other issues such as decorations and deletions of short regions have as consequence that domain structures can vary to an extent that they are classified in different categories. For that reason, defining domain boundaries has been described as 'an art more than a science' and several classification methods have been described in the literature (Csaba, Birzele, and Zimmer 2009; Day et al. 2003; Hadley and Jones 1999; Schaeffer et al. 2011: and Holland et al. 2006). For instance, in SCOP, Murzin and colleagues sorted domains mainly manually to give form to one of the most widely used databases. Manual classification has the advantage that it profits from the expertise of the protein scientists, however, it lacks the speed that is only offered by automated methods. Only recently, an automated algorithm was added to SCOP to assist the previously hand-curated database, creating the so called SCOPe database. Automated matches can be differentiated from the manual ones by the number zero placed at the end of their IDs. Other algorithms work entirely automated. Among them, CATH is one of the most commonly used (Orengo et al. 1997). SCOP and CATH converge to only 80% of their classified domains. Whereas CATH tends to break protein chains into smaller domains, SCOP keeps larger domains, leading to examples where a single domain in SCOP can be assigned to as many as six different smaller domains in CATH. To address this problem, studies suggest the usage of consensus sets (Day et al. 2003; Schaeffer et al. 2011). The advantage of employing such datasets is that they are useful to benchmark new methods for classifications but they lack the evolutionary knowledge included in SCOP.

## 1.7. Bioinformatic tools for the detection of remote homology

To confirm common ancestry between proteins with low sequence similarity is a challenging task. Although it is generally accepted that over the time structure varies more than sequence, it is also known that similar structures can be the result of convergence (Goldstein 2008; Xia and Levitt 2004). Therefore, to confirm homology, sequence evidence is necessary. Occasionally, sequence similarity cannot be detected with standard methods such as Blast or PSI-Blast (Altschul 1997). However, the development of more sensitive methods such as Hidden Markov Models (HMMs)-based alignments have also made it possible to detect homology between proteins with low sequence identity. This is because they profit from much more information than single sequence comparisons. In fact, HMMs store information of multiple sequences in a matrix, where the tendencies of mutations, deletions and extensions are calculated for each position of the aligned sequences to obtain a consensus profile that can be compared to other profiles. This method has been employed successfully in the past to detect homology among distantly related proteins and predict their previously unknown functions (Söding, Biegert, and Lupas 2005; Fidler et al. 2016; Bystroff and Krogh, n.d.). Another application of HMMs is the detection of subdomain related fragments among different folds. One example constitutes the flavodoxin-like, which shares an homologous fragment with other folds such as: (1) The TIM-barrel (Farías-Rico, Schmidt, and Höcker 2014), (2) the periplasmic-binding protein-like I (Farias-Rico J., Toledo-Patino S., Götz S. & Höcker B., unpublished data), and as later described in this work, (3) the hemD-like fold. Interestingly, these evolutionary relationships are not isolated cases. In fact, an extended search among other fold categories yielded a large number of folds that share homology with many others (unpublished, **Appendix 7.1**). These examples not only illustrate evolutionary paths (**Chapter 2**) but also provide an alternative for the protein design, as I will discuss in **Chapter 3**.

## 1.8. Protein Design: a young but promising field

The rational design of proteins is a young field. The first prototypes were built in the 70's, by Gutte and colleagues, who developed a minimalistic version of the bovine ribonuclease. The resulting protein is half of the size of the original version and shows 16% of the native activity and a similar specificity. The same group engineered a DTT-binding protein and a 34-residue peptide that was able to bind several trinucleotides and digest poly-C. These achievements were obtained via manual analysis of the amino acid composition and geometry of the desired structures. It was not until the late 90s that the first computational tools for protein design were developed and with them the first fully automated designs. In 1997 for instance, Mayo and colleagues designed a zinc finger protein with low sequence identity compared to known natural proteins (Dahiyat, Sarisky, and Mayo 1997) and shortly thereafter in 1998, Kim and coworkers developed a right-handed coil-coil, non-existent in Nature (Harbury et al. 1998). Baker et al, designed in 2003 the first protein with a fold not observed before in nature (Kuhlman et al. 2003) and in 2008 they implemented the unnatural Kemp elimination reaction onto a naturally occurring scaffold protein (Röthlisberger et al. 2008). Since then, advances in design algorithms have allowed the engineering of multiple novel functions and scaffolds, yet synthetic proteins are far from being able to compete with natural proteins in terms of speed and variety. Therefore, the development of new design approaches are desired.

## 1.9. Combination of related fold fragments as an alternative for design

The fact that Nature combines a limited set of units to create new architectures, inspired the development of a combinatorial approach for the design of proteins (Toledo-Patino S. (2013). "*Design of chimeric proteins using homologous fragments from different folds*" (Diploma thesis). Based on the premise that 'if nature can do it, we can do it too", this work carries out the design of a fold-chimera that differs from the previously designed HisF-CheY fold-hybrid (Shanmugaratnam, Eisenbeis, and Höcker 2012), in that the fusion regions are selected based on sequence similarity in addition to their structural resemblance. A further hybrid design that employs this method is the PBP-CheY chimera (pdb: 4qwv, associated publication to be submitted), which contains fragments from the flavodoxin-like and the periplasmic-binding protein-like I folds. The successful combination of fragments raised the question, whether using this approach, a function could be transfer, which is one of the aims of this doctoral thesis and will be discussed in **Chapter 4**.

## 1.10. Aims and objectives

This doctoral thesis contains three scientific projects that originated from my Master Studies. Inspired by the published evolutionary relationships between the TIM-barrel and flavodoxin-like superfolds (Farías-Rico, Schmidt, and Höcker 2014) and perusing the hypothesis that domains are unlikely to have arisen independently, I dedicated my PhD to search for additional examples that support the assumption that different folds result from a subset of fragments. I designated the first part of my research to analyze sequence-based profile-profile alignments generated for the α/β proteins based on the SCOP classification. The parsing of the data revealed that the majority of folds shared homologous regions with others. These findings brought me to the next question of my research, namely, whether these homologous fragments were selected during the course of evolution due to specific features, such as: stability or intrinsic plasticity that facilitated their combination. If that would be true, one should be able to combine these fragments, creating native-like fold-chimeras. We tested this hypothesis previously, creating a chimeric protein employing fragments from the flavodoxin-like and the periplasmic-binding protein-like I fold. The hybrid protein expressed solubly and displayed resemblance to its parental proteins (Farias-Rico J., Toledo-Patino S., Götz S. & Höcker B., unpublished data). This result encouraged me to find a binding pocket within homologous fragments that could be transplanted into an acceptor fold in a similar manner.

From an evolutionary point of view, the goal of this work consisted in selecting a pair of distant relatives to investigate their relationships in detail. I chose the hemD-like and flavodoxin-like folds, for which here I propose an emergence path (**Chapter 2**). From a design perspective, I employed their related fragments to transfer the vitamin B12 binding-pocket from a cobalamin-binding domain (flavodoxin-like) into the uroporphyrinogen III synthase (hemD-like) (**Chapter 3**). The functional characterization of the resulting hybrid showed that it retained the transferred function (**Chapter4**).

# Chapter 2

## The hemD-like fold encloses two flavodoxin-like fossils

### 2.1. Overview

A major evolutionary trigger constitutes domain duplication and divergence. In this process, the new copy is free to acquire a new function since the selective pressure is alleviated by the remaining copy. Usually, a duplicated domain retains a similar structure as the original template, even though their sequences diverge during the course of time. However, in a few cases, the duplicated domains fuse to create new scaffolds and continue to diverge to an extent where structural similarities to their single domain homologs are no longer evident. For that reason, proven cases of distant protein relatives that display low sequence similarities, different structures, and functions, are of great value and broaden our understanding on protein evolution.

In this chapter, I will present the results of a profile-profile sequence analysis among α/β proteins in SCOP, which yielded evidence to support the emergence of the hemD-like fold via insert-mediated segment-swapping and gene duplication of the ancient flavodoxin-like fold. To prove this hypothesis experimentally, I reverted these evolutionary events one by one, employing the intact C-terminal half of the hemD-like member, uroporphyrinogen III synthase (cU3S), and a variant lacking the insertion (cU3SΔ). The results were conclusive: cU3S builds multimers whereas cU3SΔ runs exclusively as monomer when subjected to an analytical gel filtration. Moreover, the NMR structure of cU3SΔ revealed the canonical flavodoxin-like architecture, strongly supporting a potential emergence path for hemD-like proteins. To the best of my knowledge, this is the first time an ancient fold is reconstructed employing the sequence of a different one without further mutations.

## 2.2 The uroporphyrinogen III synthase in a nutshell

The uroporphyrinogen III synthase (U3S) is an essential enzyme found in all kingdoms of life. It catalyzes the conversion of hydroxymethylbilane (HMB) to uroporphyrinogen III (U3) (**Figure 2.1**), which is the last common precursor of all tetrapyrrole cofactors, the so-called pigments of life. These essential molecules are involved in metabolic and catalytic processes such as oxygen transport (hem), photosynthesis (chlorophyll), methane production (coenzyme F430) and methionine synthesis (cobalamins). In the absence of U3S, HMB spontaneously cyclize to generate non-physiological uroporphyrinogen I (UI) (**Figure 2.1**). In humans, this reaction is favored in cases of erythropoietic porphyria, a rare but severe disease that causes hypersensitivity to sun exposure due to U3S malfunction (Fortian et al. 2009).

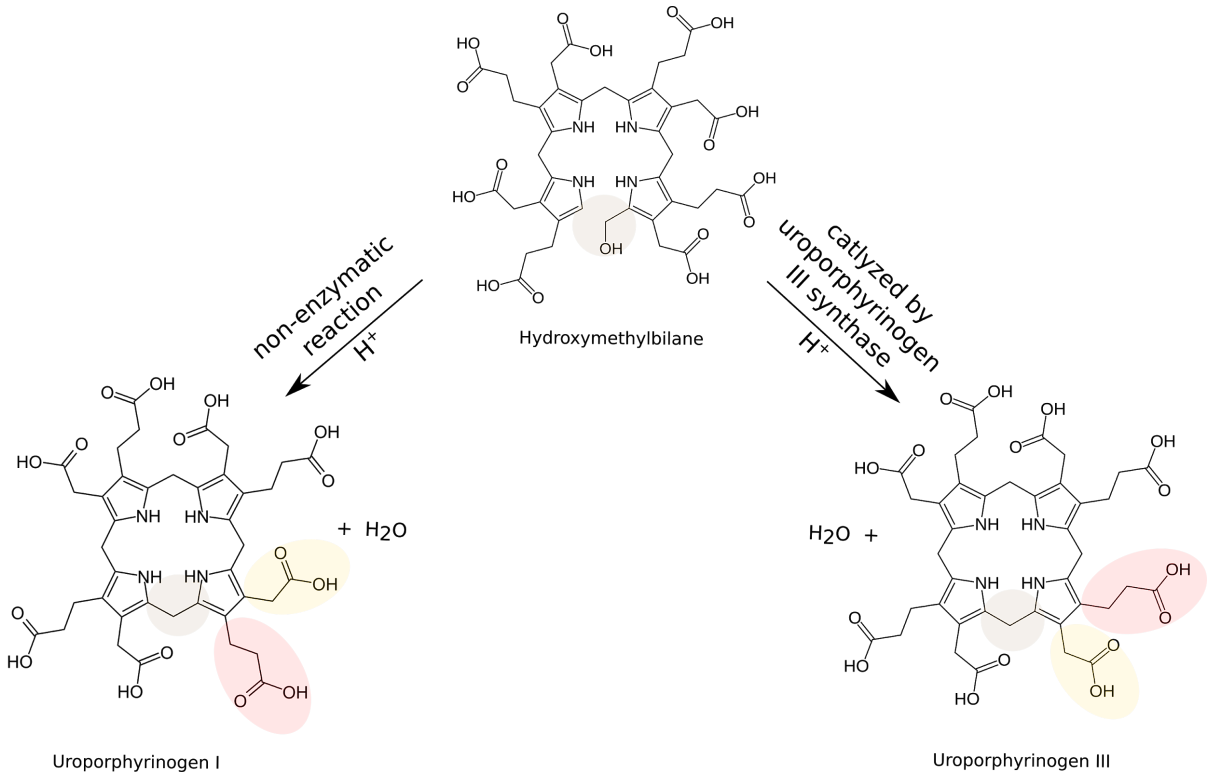



**Figure 2.1 Synthesis of uroporphyrinogen I and III from hydroxymethylbilane.** Spontaneous conversion of uroporphyrinogen I goes on non-enzymatically and has no physiological relevance, whereas conversion of its isomer uroporphyrinogen III requires the presence of uroporphyrinogen III synthase. Differences between the isomers are highlighted in red and yellow.

At a structural level, U3S is an α/β bi-lobular enzyme that displays two distinct symmetry axes, which leads to disagreement in domain annotation by different systems. For instance, lobular symmetry is weighted by CATH Protein Structure Classification database (Dawson et al. 2017), which assigns U3S into two structural domains (**Figure 2.2-A**). This classification has three important implications: 1) it generates domains with different topologies, 2) one of these domains is discontinuous (contains both: N- and C-termini), and 3) the binding pocket located between the lobes, is disrupted. In contrast, the Structural Classification of Proteins (SCOP) (Murzin et al. 1995), considers U3S as a single domain enzyme, acknowledging the fact that the interface between the lobes is required for catalysis. U3S is the only member of the so called hemD-like SCOP superfamily and adopts the homonymous hemD-like fold (**Figure 2.2-B**). An alternative symmetry axis can be defined vertically, where the polypeptide chain is cut once in its middle to generate two antiparallel halves with the same amount of secondary structural elements and topological connections (**Figure 2.2-C**). Yet, none of the systems for domain classification considers these units as single domains, since this type of symmetry leads to disruption of the hydrophobic core of the protein, a premise widely accepted for domain definition. However, as I will discuss in the following section, in terms of evolution, this symmetry hides meaningful details about the emergence of the hemD-like scaffold via gene duplication from a flavodoxin-like fold, a protein architecture that can be traced back to the last common ancestor of all kingdoms of life (Karlusich et al. 2014).

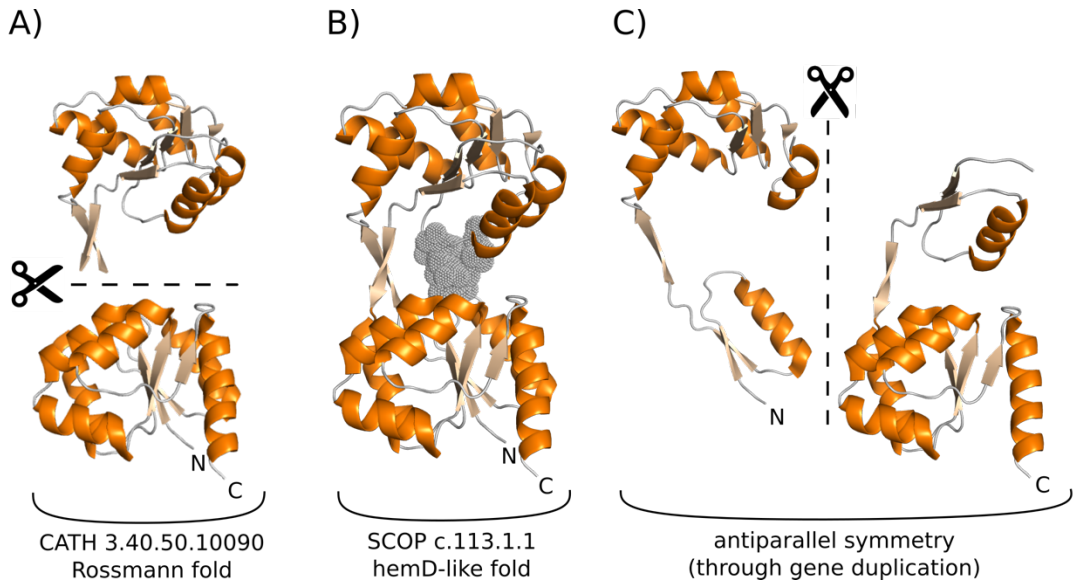



**Figure 2.2 Uroporphyrinogen III synthase (U3S) symmetry and domain assignment by CATH and SCOP.** CATH dismantles U3S into two domains **(A)**. In contrast, SCOP acknowledges U3S ligand binding and considers this enzyme as single domain (hemD-like fold) **(B)**. Finally, antiparallel symmetry of U3S results from isolation of its N- and C-terminal halves **(C)**. The above illustrated structures based on U3S structure from *Thermus thermophilus* pdb identifiers 3d8r and 3d8n. Uroporphyrinogen III (cognate ligand) is illustrate as spheres (gray).

## 2.3. Emergence of the HemD-like fold from flavodoxin-like fold gene duplication

Evolutionary paths that led to the emergence of structural domains left a footprint in the sequences of contemporary proteins, which can only be uncover with the help of state-of-the-art methods for remote homology detection. For instance, Hidden Markov Models (HMMs) are a powerful tool to detect distant homologs, since they enclose much more information than single-sequence comparisons. Recently, HMMs were employed to detect remote ancestry between the flavodoxin-like and TIM-barrel folds, which in spite of their overall different architectures, have been proposed to be related (Farías-Rico, Schmidt, and Höcker 2014). Inspired by these findings, we searched for further remote homologs within different fold categories, in particular among α/β proteins, which are believed to enclose some of the oldest architectures (Taylor 2006). Thus, we generated HMMs for the α/β-class of SCOP employing the HHsuite package (Söding 2005) and performed all-against-all sequence-based profile-profile alignments without using the structural prediction part of the program[1] (**Methods 6.1.1**). The search revealed that the majority of the fold members shared homologous fragments of around 30 residues in length. However, a few spanned considerably larger regions in domain-sized ranges (**Appendix 2**). One of these exceptions represents the hemD-like fold (circa 250 residues), which is hit twice by the flavodoxin-like fold (circa 115 residues) at its N- and C-terminal halves, respectively. These observations not only suggested a duplication event but also indicated flavodoxin-like as the duplicating template (**Figure 2.3**). Supporting this hypothesis, the N- and C-terminal halves of hemD-like also align to one another with probabilities nearly at 100%, clearly displaying their sequence similarity, which as expected corresponds also at a structural level. In contrast, the hemD-like halves do not superimpose with the flavodoxin-like structure. Whereas flavodoxin-like displays a globular architecture, the hemD-like halves present a segment-swap (**Figure 2.4**). This phenomenon has been previously studied by Szilágyi and colleagues with computational methods for molecular dynamics (Szilágyi, Györffy, and Závodszky 2017), proposing that the scaffold may have evolved via circular permutation of a mono-connected ancestor. However, in the following sections, I will provide sequence and experimental evidence that supports an alternative evolutionary scenario consisting of insertion, gene duplication and fusion.

[1] Profile-profile alignments were assessed by José Farías Rico

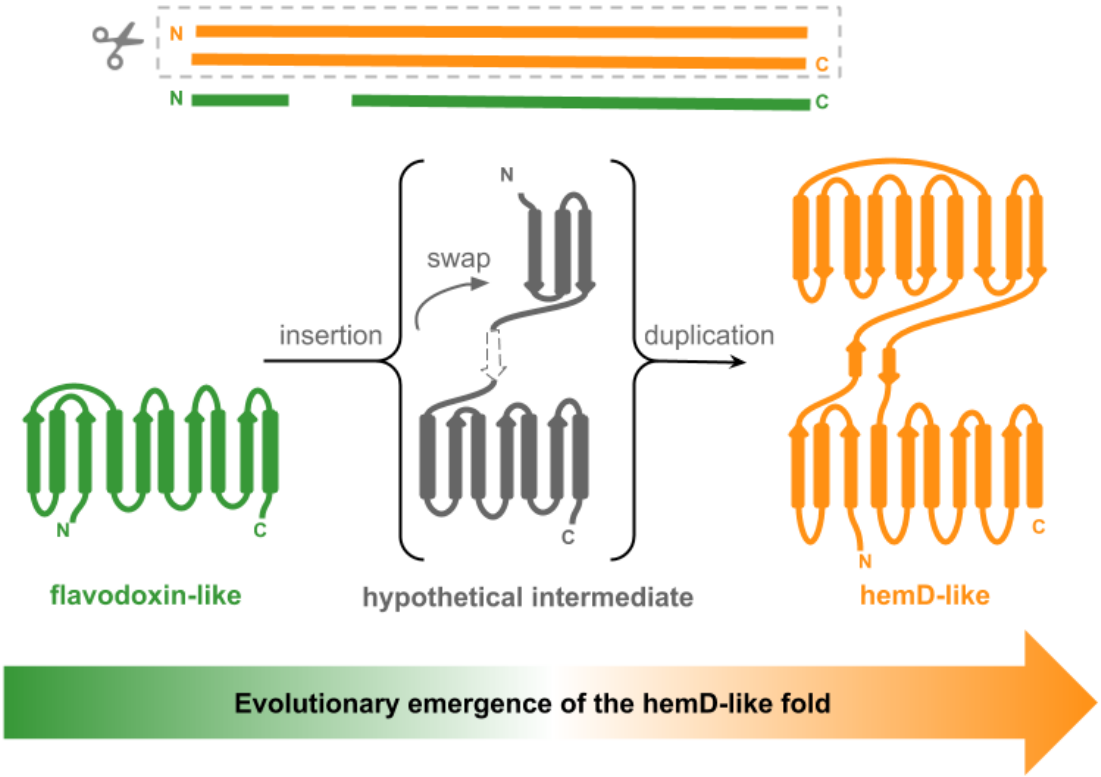



**Figure 2.3 HemD-like fold emergence from flavodoxin-like.** Profile-profile alignments showed that N- and C-terminal hemD-like halves match each other and flavodoxin-like with probabilities above 90%, suggesting that HemD-like arose via flavodoxin-like gene duplication. Structural and topological discrepancies between the folds can be explained by an insert-mediated segment-swapping, which after fusion gave rise to the current bi-lobular architecture of the uroporphyrinogen III synthase, which is the only enzyme known to adopt the hemD-like scaffold.

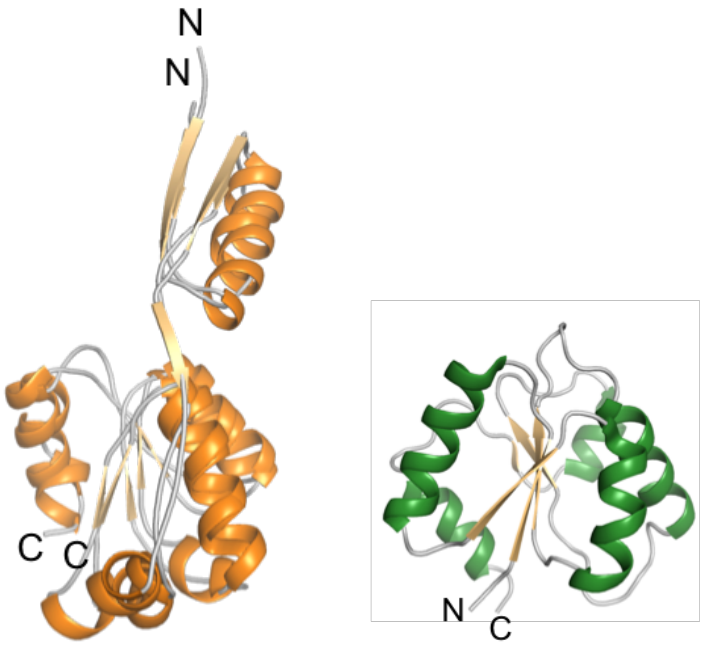
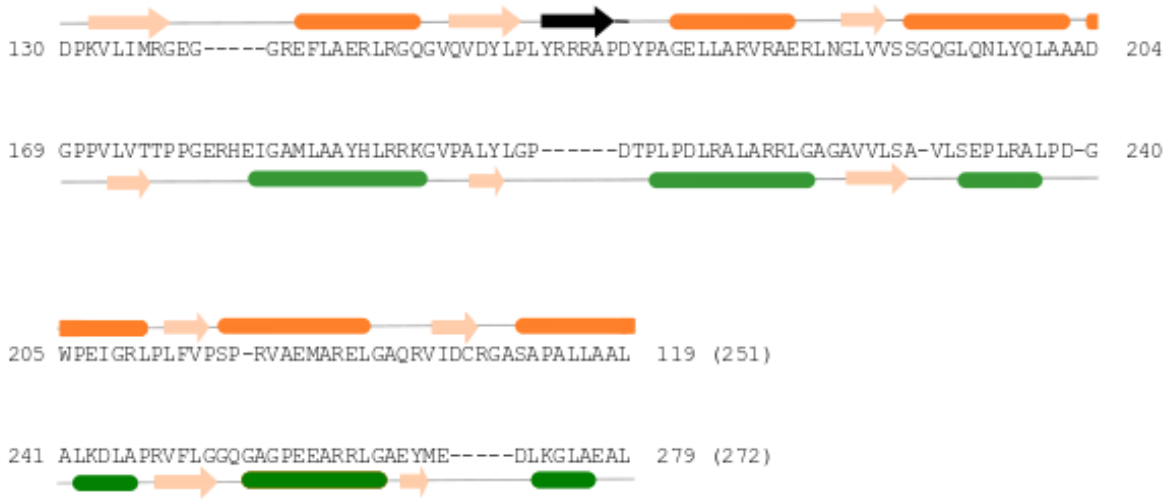



**Figure 2.4 Structural discrepancies between flavodoxin-like fold and hemD-like halves.** N- and C- terminal hemD-like halves correspond to each other in their topology and architecture (orange). However, they superpose only partially onto the flavodoxin-like fold (**Figure 2.5-B**), which displays a globular architecture (green). Pdb identifiers: 4es6 and 3whp, respectively.

## 2.4. An insertion in the flavodoxin-like fold mediated its transition to a bi-lobular architecture

When it comes to sustain homology, sequence identity is the most reliable source of information. However, at large evolutionary distances also similar structures or functions provide a solid hint for a common origin. In fact, even cases of proteins that display different functionalities, architectures and rather low sequence resemblance hide subtitle similarities in their sequence patterns. Thus, is it not surprising that scholars agree upon that convergent evolution is a rare phenomenon that accounts for only a small fraction of evolving proteins (Gough 2005; Forslund et al. 2008). While most evolutionary links are identified on the basis of sequence identity, the most fascinating discoveries resulted from searching in the so called 'twilight zone'. The hemD-like and the flavodoxin-like folds represent such an example of scaffolds that in spite of their overall differences, a handful of their members still retain sequence identity high enough to support a common origin. For instance, the C-terminal half of the uroporphyrinogen III synthase (U3S) from *Pseudomonas aeruginosa* (pdb: 4es6) matches the flavodoxin-like domain LitR from *Thermus thermophilus* (pdb: 3whp) with 30% sequence identity over 115 amino acids. These values are significant in terms of evolution according to a benchmarking published earlier (Sander and Schneider 1991). In this work, Sander and colleagues suggest that two sequences ≥ 80 amino acids in length can be considered homologous if they display at least 25.8% sequence identity. Thus, how is it possible that two sequences having high sequence similarity display different architectures? If we have a closer look at the U3S and LitR topologies, it becomes obvious that they share the same secondary structural elements (SSE) except by an insertion of six amino acids that form a beta strand (**Figure 2.5**). Moreover, their SSE have the same order, differing only in their orientation in space due to an αβα-swap. Interestingly, but perhaps not surprisingly, the predicted insertion is located right before the swapped elements in all hemD-like halves but in any of the flavodoxin-like members. These observations brought me to the conclusion that the insert may have mediated the swapping, parting from a flavodoxin-like precursor. If this was the case, truncation of one hemD-like half and removal of its insertion should lead to the reconstruction of a protein with a flavodoxin-like topology and architecture. Therefore, I employed the U3S C-terminal half from *P. aeruginosa* to revert in vitro the events that potentially led to the emergence of the hemD-like scaffold. I will review the obtained experimental results in the following sections.

**A)**




**B)**


**Figure 2.5 Sequence and structural alignment for the best-scored HHsearch hit.** The best ranked sequence-based profile alignment, according to sequence identity, constitutes the C-terminal half of the uroporphyrinogen III synthase (U3S) from *Pseudomonas aeruginosa* (pdb: 4es6) towards the quorum sensing transcriptional regulator (LitR) from *Thermus thermophilus* (pdb: 3whp). Secondary structural elements are highlighted in orange and green for U3S and LitR, respectively. The insertion forming a beta strand is shown in black. The values obtained by HHsearch are: 30% sequence identity, 88% probability and E-value = 0.11 over 105 residues **(A)**. Alignment of the U3S C-terminal half on the LitR structure reveals an αβα-segment-swap in U3S in respect to LitR, which adopts the globular architecture characteristic of flavodoxin-like members.

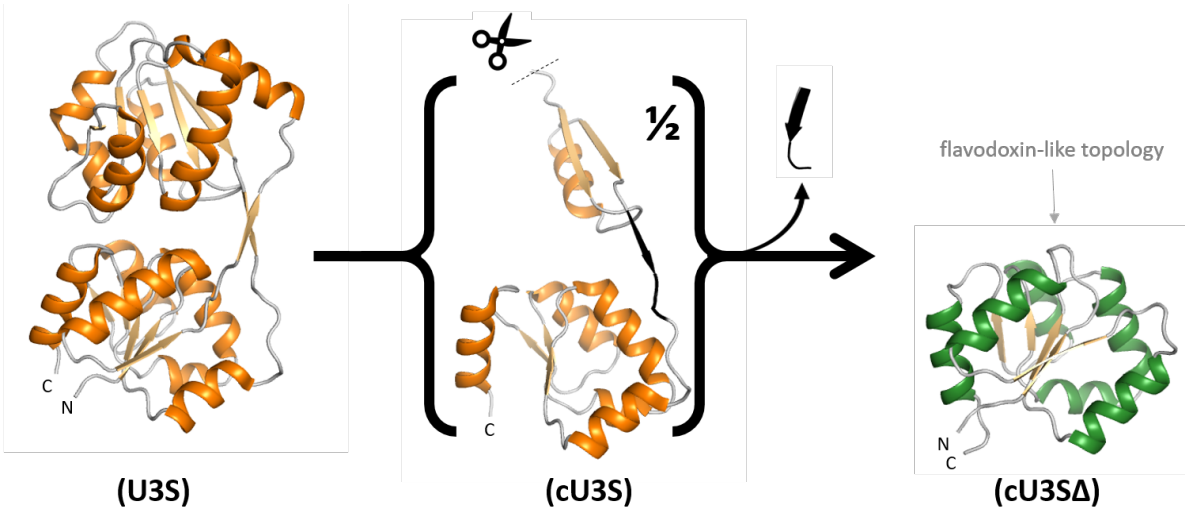
## 2.5. Experimental Reconstruction of flavodoxin-like from one hemD-like half

To reverse the gene duplication and insertion events experimentally, two different constructs were generated: 1) the intact C-terminal half of the uroporphyrinogen III synthase (cU3S) from *Pseudomonas aeruginosa* (residues H129 to A251), and 2) a derived variant (cU3SΔ) lacking the predicted insertion, which builds a beta strand with residues Y162 to P167 (**Figure 2.6-A**). As template, the PA5259 plasmid was employed, which has been described elsewhere (Schnell and Schneider 2013) and was kindly handed out by Prof. Gunter Schneider and Dr. Robert Schnell. cU3S and cU3SΔ were amplified by PCR (**Section 6.2.1**) and cloned into a pET21a vector system, yielding the constructs pET21a-cU3S and pET21a-cU3SΔ, whose sequences were validated at the sequencing facility of the Max Planck Institute in Tübingen. Protein expression for both constructs showed soluble expression. However, cU3S expressed mainly in the inclusion bodies compared to cU3SΔ (**Figure 2.7-A**). In addition, native cU3S was less stable and aggregated already at a concentration of 1 mg/ml, precipitating gradually during the purification process (**Figure 2.7-B**), whereas cU3SΔ remained stable at concentrations above 25 mg/ml, running exclusively as monomer when loaded onto a gel filtration column (**Figure 2.7-C**). In an attempt to obtain better yields of soluble protein, cU3S was refolded with guanidinium chloride (**Section 6.13**). However, it aggregated, forming multimers of diverse sizes, indicating its tendency to associate with other halves, whereas the deleted variant stayed as monomer even after refolding under the same conditions. The CD spectra for both constructs showed well-folded proteins with very similar αβ-content (**Figure 2.7-D**). However, the thermal denaturation indicated a lower thermal stability for cU3S (Tm=45°C) compared to cU3SΔ (60°C) (**Figure 2.7-E**). None of the variants crystallized in any of the screenings attempted (**Section 5.2.6**). However, the stability of cU3SΔ at high concentration and temperature made it possible to elucidate its NMR structure[2], which revealed the successful reconstruction of a compact monomeric architecture (**Figure 2.6-A**). The cU3SΔ structure clearly displays a flavodoxin-like topology and architecture. However, by definition, two domains belong to the same fold if their structures align with an RMSD ≲ 3Å. Therefore, I performed a structural alignment with cU3SΔ against the PDB90, employing the server for protein structure comparison DALI (Holm and Laakso 2016). The results corroborated that cU3SΔ adopts a flavodoxin-like fold, as it matches flavodoxin-like members with RMSD values down to 2.7Å (**Table 2.1**). As expected,

[2] NMR structure (unpublished) was solved by Murray Coles and Manish Chaubey at the Max Planck Institute in Tübingen, Germany.

cU3SΔ also hits its parental protein and other hemD-like members, showing RMSD values of 2.2Å and up to 2.5Å, respectively (**Table 2.1**). Moreover, a closer look to the cU3SΔ structure, revealed residue-residue interactions at the interface created between the originally-swapped region with the accepting fragment (**Figure 2.6-B**). These interactions promote the formation of a hydrophobic core.

**A)**




**B)**

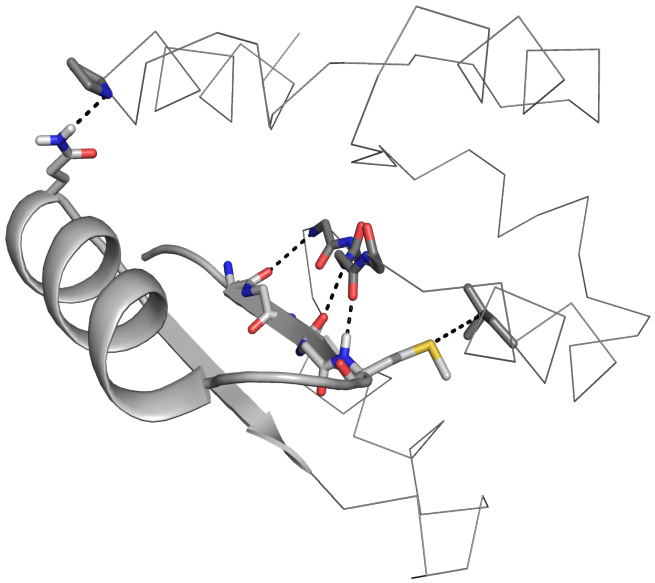


**Figure 2.6 experimental reconstruction of flavodoxin-like from the hemD-like C-terminal half.** The N-terminal half of uroporphyrinogen III synthase (U3S, orange, pdb: 4es6) from *Pseudomonas aeruginosa* was removed, leading to an unstable truncated C-terminal half (in brackets). After removing the insert (black) predicted by sequence-based profiles alignments, the resulting protein adopted the flavodoxin-like fold, whose structure was successfully solved by nuclear magnetic resonance (green) **(A)**. cU3SΔ NMR structure shows the originally swapped αβα-segment (cartoon) successfully flipped back, rebuilding residue-residue contacts with the accepting fragment (ribbon) **(B)**.

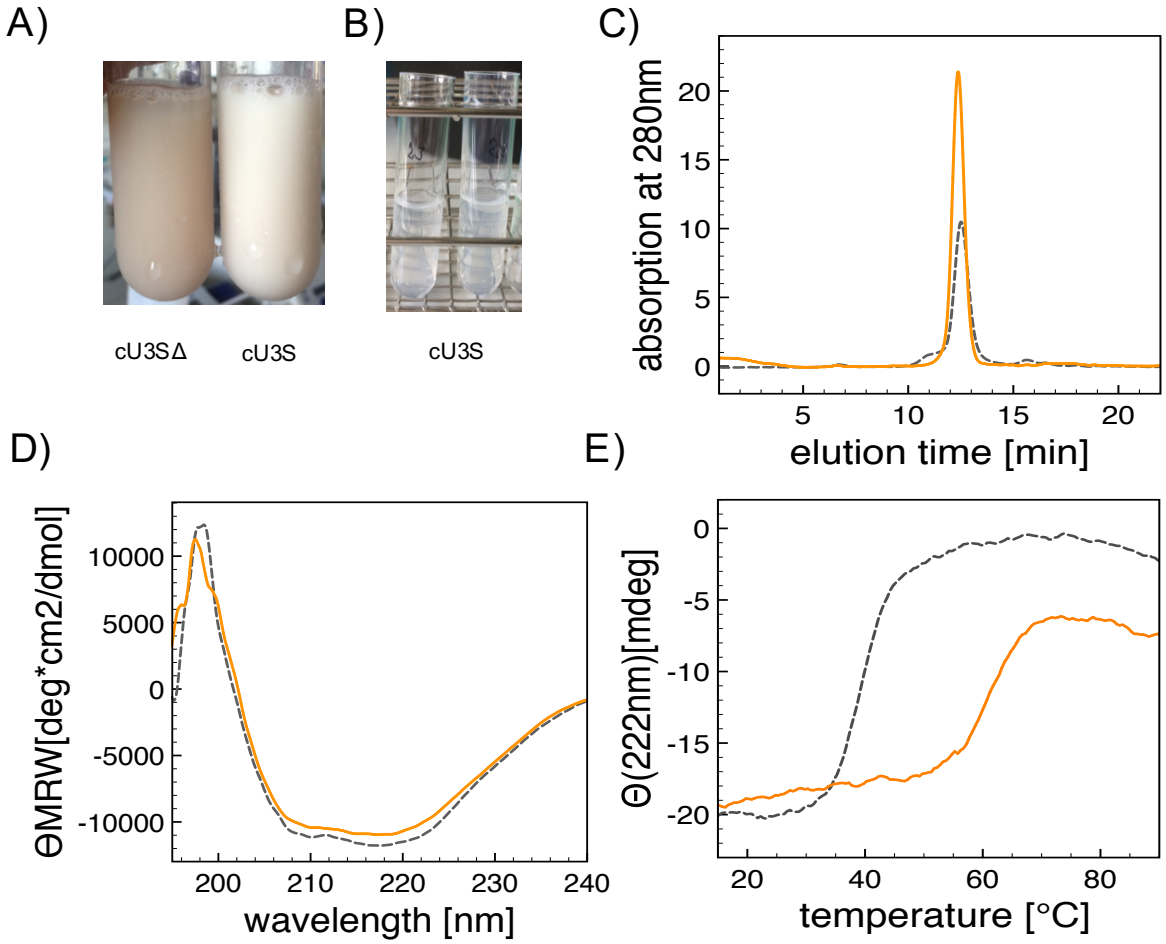



**Figure 2.7 Biophysical characterization of hemD-like half variants.** The characteristic white color that results from recombinant expression in the inclusion bodies was particularly observed for cU3S in contrast to the cU3SΔ crude extract **(A)**. Soluble expression of cU3S showed gradual protein precipitation during the purification process **(B)**. At low concentration, cU3S (black dashed line) remained as a monomer when loaded onto a gel filtration, whereas cU3SΔ (orange solid line) remained as monomer even at concentrations above 25 mg/ml **(C)**. CD signal showed similar secondary structural elements for both variants **(D)** however, the thermal denaturation indicated a higher stability for the deleted version **(E)**.

**Table 2.1 Structural alignment of cU3SΔ against the PDB90 employing DALI structure library.** The NMR structure matches its parental U3S (lower lobe) at 2.2Å and 14 Z score (orange). Proteins adopting a flavodoxin-like architecture are also hit with RMSDs up to 2.7Å and Z scores up to 11.2 (green).

| Rank | pdb | Z | RMSD | L | nres | ID | Description |
|---|---|---|---|---|---|---|---|
| 1: | 4es6-A | 14.0 | 2.2 | 111 | 249 | 70 | UROPORPHYRINOGEN-III SYNTHASE; |
| 2: | 3re1-B | 13.8 | 2.5 | 116 | 257 | 48 | UROPORPHYRINOGEN-III SYNTHETASE; |
| 3: | 3mw8-A | 11.7 | 2.5 | 110 | 237 | 28 | UROPORPHYRINOGEN-III SYNTHASE; |
| 4: | 1wd7-B | 11.3 | 2.7 | 117 | 255 | 14 | UROPORPHYRINOGEN III SYNTHASE; |
| 5: | 2j48-A | 11.2 | 3.1 | 109 | 119 | 13 | TWO-COMPONENT SENSOR KINASE; |
| 6: | 1jr2-A | 10.9 | 3.0 | 111 | 260 | 17 | UROPORPHYRINOGEN-III SYNTHASE; |
| 7: | 4qpj-D | 10.7 | 2.7 | 112 | 121 | 13 | PHOSPHOTRANSFERASE; |
| 8: | 1xhe-B | 10.3 | 2.7 | 109 | 122 | 15 | AEROBIC RESPIRATION CONTROL PROTEIN ARCA; |
| 9: | 2zwm-A | 10.1 | 2.9 | 111 | 120 | 17 | TRANSCRIPTIONAL REGULATORY PROTEIN YYCF; |
| 10: | 4lda-B | 9.9 | 3.5 | 117 | 128 | 18 | TADZ; |
| 11: | 3ilh-A | 9.9 | 2.8 | 114 | 133 | 11 | TWO COMPONENT RESPONSE REGULATOR; |
| 12: | 2qr3-A | 9.8 | 2.9 | 110 | 121 | 15 | TWO-COMPONENT SYSTEM RESPONSE REGULATOR; |
| 13: | 6br7-B | 9.8 | 3.2 | 113 | 125 | 17 | TWO-COMPONENT SYSTEM RESPONSE REGULATOR PROTEIN; |
| 14: | 2zay-A | 9.8 | 3.2 | 113 | 123 | 9 | RESPONSE REGULATOR RECEIVER PROTEIN; |
| 15: | 3jte-A | 9.8 | 3.3 | 113 | 126 | 12 | RESPONSE REGULATOR RECEIVER PROTEIN; |
| 16: | 3t6k-A | 9.8 | 3.2 | 111 | 122 | 18 | RESPONSE REGULATOR RECEIVER; |
| 17: | 5u8k-B | 9.7 | 3.5 | 113 | 230 | 11 | RESPONSE REGULATOR; |
| 18: | 5m7o-A | 9.7 | 3.2 | 115 | 448 | 17 | NITROGEN ASSIMILATION REGULATORY PROTEIN; |
| 19: | 1k7y-A | 9.7 | 2.8 | 113 | 577 | 12 | METHIONINE SYNTHASE; |
| 20: | 5xt2-B | 9.7 | 3.0 | 110 | 204 | 16 | RESPONSE REGULATOR FIXJ; |
| 21: | 3grc-A | 9.7 | 3.1 | 111 | 125 | 16 | SENSOR PROTEIN, KINASE; |
| 22: | 5m7n-B | 9.7 | 3.4 | 113 | 428 | 18 | NITROGEN ASSIMILATION REGULATORY PROTEIN; |
| 23: | 3dge-C | 9.6 | 3.1 | 110 | 122 | 18 | SENSOR PROTEIN; |
| 24: | 1y80-A | 9.6 | 2.7 | 104 | 125 | 16 | PREDICTED COBALAMIN BINDING PROTEIN; |
| 25: | 5wq0-D | 9.6 | 3.1 | 113 | 129 | 15 | STAGE 0 SPORULATION PROTEIN; |
| 26: | 4nic-A | 9.6 | 2.7 | 105 | 117 | 14 | DNA-BINDING TRANSCRIPTIONAL REGULATOR RSTA; |
| 27: | 1dc7-A | 9.5 | 2.9 | 109 | 124 | 17 | NITROGEN REGULATION PROTEIN; |
| 28: | 4jgi-B | 9.5 | 2.7 | 104 | 206 | 11 | PUTATIVE UNCHARACTERIZED PROTEIN; |
| 29: | 4s04-A | 9.5 | 3.0 | 111 | 219 | 12 | DNA-BINDING TRANSCRIPTIONAL REGULATOR BASR; |
| 30: | 5ep0-A | 9.5 | 3.1 | 112 | 380 | 13 | PUTATIVE REPRESSOR PROTEIN LUXO; |
| 31: | 1ys6-B | 9.4 | 3.2 | 112 | 227 | 15 | TRANSCRIPTIONAL REGULATORY PROTEIN PRRA; |
| 32: | 4q7e-A | 9.4 | 3.2 | 111 | 125 | 19 | RESPONSE REGULATOR OF A TWO COMPONENT REGULATORY |
| 33: | 3m6m-D | 9.3 | 3.0 | 105 | 118 | 16 | RPFF PROTEIN; |
| 34: | 2gwr-A | 9.3 | 3.3 | 109 | 225 | 16 | DNA-BINDING RESPONSE REGULATOR MTRA; |
| 35: | 2qzj-A | 9.3 | 3.1 | 108 | 121 | 14 | TWO-COMPONENT RESPONSE REGULATOR; |
| 36: | 4h60-A | 9.3 | 3.2 | 109 | 120 | 16 | CHEMOTAXIS PROTEIN CHEY; |
| 37: | 4uhk-A | 9.3 | 3.5 | 113 | 130 | 14 | TRANSCRIPTIONAL REGULATORY PROTEIN CPXR; |
| 38: | 4l4u-A | 9.3 | 3.3 | 113 | 388 | 16 | TRANSCRIPTIONAL REGULATOR (NTRC FAMILY); |
| 39: | 2qv0-A | 9.2 | 3.0 | 106 | 122 | 16 | PROTEIN MRKE; |
| 40: | 2nt4-A | 9.2 | 3.1 | 107 | 126 | 19 | RESPONSE REGULATOR HOMOLOG; |
| 41: | 3q9s-A | 9.2 | 3.2 | 113 | 210 | 17 | DNA-BINDING RESPONSE REGULATOR; |
| 42: | 3cfy-A | 9.2 | 3.7 | 115 | 130 | 10 | PUTATIVE LUXO REPRESSOR PROTEIN; |
| 43: | 5dcl-A | 9.2 | 2.9 | 108 | 117 | 15 | PHOB FAMILY TRANSCRIPTIONAL REGULATOR; |
| 44: | 2rjn-A | 9.2 | 3.8 | 114 | 135 | 12 | RESPONSE REGULATOR RECEIVER:METAL-DEPENDENT |
| 45: | 3luf-B | 9.1 | 2.8 | 111 | 246 | 14 | TWO-COMPONENT SYSTEM RESPONSE REGULATOR/GGDEF |
| 46: | 3hdv-B | 9.1 | 3.0 | 107 | 126 | 19 | RESPONSE REGULATOR; |
| 47: | 2jba-A | 9.1 | 3.3 | 111 | 125 | 14 | PHOSPHATE REGULON TRANSCRIPTIONAL REGULATORY PROT |
| 48: | 1nxp-A | 9.1 | 2.8 | 105 | 118 | 17 | DNA-BINDING RESPONSE REGULATOR; |
| 49: | 1sqs-A | 9.0 | 2.9 | 116 | 232 | 10 | CONSERVED HYPOTHETICAL PROTEIN; |
| 50: | 1zgz-A | 9.0 | 2.9 | 104 | 121 | 11 | TORCAD OPERON TRANSCRIPTIONAL REGULATORY PROTEIN |

*Abbreviations: Z-score **(Z)**, alignment length **(L)**, total number of residues **(nres)** and sequence identity **(ID)***

## 2.6. Segment-swaps and their role in the evolution of new folds

While three-dimensional domain swapping has been widely described in the literature (Bennett, Choe, and Eisenberg 1994; Bennett, Schlunegger, and Eisenberg 1995; Gronenborn 2009; Jaskólski 2001; Liu and Eisenberg 2002), only little attention has been paid to segment-swaps (SSPs) (Andreeva and Murzin 2006; Szilágyi, Zhang, and Závodszky 2012). In this process, only a few structural elements within a domain are flipped, creating residue-residue contacts with a complementary protein fragment. It has been estimated that at least 18 structural families present well-defined covalently linked swapped regions (Szilágyi, Zhang, and Závodszky 2012). However, SSPs are also observed as crystallization artifacts (Farías-Rico, Schmidt, and Höcker 2014), oligomerization of engineered proteins (Riechmann et al. 2005) and more interestingly, as dynamic transitions required for natural functions. One example constitutes the Glyoxilase I, which is assembled of two non-covalently linked domains that dislocate to build a functional protein (Cameron 1997). Interestingly, the dihydroxybiphenyl dioxygenase displays a similar architecture, however it lacks the SSP and its building units are embedded in a single chain. Scholars suggest that these two proteins are related based on their similar metal-binding positions and structures (2Å over 79 Cα atoms), suggesting that the SSP may have led to the different function in Glyoxylase I (Cameron 1997). Thus, SSPs illustrate the plastic nature of proteins to adopt alternative conformations. This is a successful strategy to reach complexity and size by joining fragments that do not require to be covalently attached. They provide new cavities for binding and motion restrains for catalysis, particularly if the involved units become fused in a single chain. This appears to be the case for the uroporphyrinogen III synthase (U3S), in which the insert-mediated SSP provided the necessary extension to create a cleft for binding and the interdomain linkers conferred movement restriction, perhaps improving catalysis. Previously, the impact of flexibility restriction on the U3S lobes has been studied (Szilágyi, Györffy, and Závodszky 2017). In this work, Szilágyi and colleagues simulated the conversion of U3S into a mono-connected bi-lobular enzyme by circular permutation (*in silico*), observing a considerable decrease in its reaction capacity. Based on these results, Szilágyi and colleagues suggest that circular permutation was a step in the U3S emergence path, which conferred it an evolutionary advantage over a hypothetical mono-connected scaffold due to improvements in entropy costs, thus achieving lower free energies for binding. I will discuss this potential scenario in the following section.

## 2.7. Circular permutation as potential emergence path for hemD-like

Circular permutations (CPs) are rare events, accounting for only 5% of proteins with known structure (Jung and Lee 2001). While CPs are easily visualized at a structural level, this process is less straightforward at a genetic level and several mechanisms have been proposed to explain their formation (Vogel and Morea 2006). As pointed out in the previous section, a circular permutation event has been proposed for the emergence of uroporphyrinogen III synthase (U3S). To test this hypothesis, I assessed an HHsearch employing a database of manually permuted U3S halves (**Section 5.4.1** and **6.1.4**) to search for high sequence identities towards flavodoxin-like proteins contained in the SCOP and pdb databases. The results showed that the permuted variants not only scored flavodoxin-like with lower probabilities than their intact halves (**Table 2.2**), but in addition none of the hits showed a statistically significant sequence identity towards flavodoxin-like members. These observations, speak in favor of the insertion-mediated segment-swapping and correlate with the fact that permutation is less common than domain duplication, which is the most common event in domain evolution (Lupas, Ponting, and Russell 2001; Chothia and Gough 2009).

**Table 2.2 Best-ranked HHsearch hits for hemD-like halves and their circular-permuted variants.**

| Q ID | S ID | Function | CATH | SCOP | P | SI | Cols | Q*s* | Q*e* | S*s* | S*e* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 4es6_ct | 3whp_A | Cobalamin-binding | 3.40.50.280 | c.23.1 | 88 | 30 | 105 | 2 | 119 | 169 | 279 |
| 3re1_ct | 4hh0_A | Cobalamin-binding | 3.40.50.280 | c.23.1 | 90 | 24 | 117 | 1 | 128 | 271 | 402 |
| 3re1_ct | d1yioa2 | Response regulator | 3.40.50.2300 | c.23.1 | 91 | 24 | 108 | 3 | 124 | 3 | 119 |
| 3re1_ct | 3whp_A | Cobalamin-binding | 3.40.50.280 | c.23.6 | 91 | 23 | 110 | 1 | 123 | 168 | 285 |
| 4es6_ct | 4r3u_C | Cobalamin-binding | 3.40.50.280 | c.23.6 | 94 | 23 | 114 | 1 | 121 | 15 | 139 |
| 4es6_ct | 4hh3_C | Cobalamin-binding | 3.40.50.280 | c.23.6 | 91 | 23 | 115 | 1 | 123 | 105 | 234 |
| d8r_L1 | 3ph3_A | Ribose-5P isomerase | 3.40.1400.10 | c.121 | 87 | 21 | 118 | 1 | 126 | 21 | 143 |
| 3d8r_L1 | d3he8a_ | Ribose-5P-isomerase | 3.40.1400.10 | c.121 | 95 | 21 | 118 | 1 | 126 | 1 | 123 |
| 3mw8_ct | d1ccwa_ | Cobalamin-binding | 3.40.50.280 | c.23.6 | 96 | 21 | 110 | 3 | 119 | 5 | 130 |
| 3mw8_ct | d1u0sy_ | Response regulator | 3.40.50.2300 | c.23.1 | 94 | 21 | 103 | 1 | 120 | 1 | 115 |
| 3mw8_ct | d2zwma_ | Response regulator | 3.40.50.2300 | c.23.1 | 91 | 21 | 101 | 2 | 119 | 2 | 112 |
| 3mw8_L2 | d1ccwa_ | Cobalamin-binding | 3.40.50.280 | c.23.6 | 90 | 21 | 110 | 84 | 199 | 2 | 127 |
| 3re1_ct | 5c8a_A | Cobalamin-binding | 3.40.50.280 | c.23.6 | 90 | 21 | 112 | 1 | 124 | 84 | 202 |
| 3re1_ct | d1fmfa_ | Cobalamin-binding | 3.40.50.280 | c.23.6 | 95 | 21 | 102 | 13 | 123 | 19 | 134 |
| 4es6_L1 | 4r3u_C | Cobalamin-binding | 3.40.50.280 | c.23.6 | 92 | 21 | 114 | 2 | 123 | 15 | 139 |
| 3d8r_nt | d1dxya2 | HI-dehydrogenase | 3.40.50.720 | c.23.12 | 94 | 20 | 104 | 1 | 117 | 1 | 107 |
| 3mw8_ct | d2jbaa_ | Response regulator | 3.40.50.2300 | c.23.1 | 92 | 20 | 106 | 2 | 119 | 2 | 115 |
| 3re1_ct | 4r3u_C | Cobalamin-binding | 3.40.50.280 | c.23.6 | 94 | 20 | 114 | 1 | 121 | 15 | 139 |
| 3re1_ct | d3to5a_ | Response regulator | 3.40.50.2300 | c.23.1 | 89 | 20 | 112 | 2 | 124 | 4 | 124 |
| 3re1_ct | d1ys7a2 | Response regulator | 3.40.50.2300 | c.23.1 | 89 | 20 | 107 | 4 | 123 | 3 | 117 |
| 3re1_ct | d3hdva_ | Response regulator | 3.40.50.2300 | c.23.1 | 86 | 20 | 110 | 4 | 124 | 3 | 120 |
| 3re1_ct | d1zh2a1 | Response regulator | 3.40.50.2300 | c.23.1 | 85 | 20 | 106 | 4 | 123 | 2 | 115 |
| 4es6_ct | d3q15c_ | Response regulator | 3.40.50.2300 | c.23.1 | 94 | 20 | 106 | 3 | 122 | 1 | 115 |
| 4es6_L1 | d1ccwa_ | Cobalamin-binding | 3.40.50.280 | c.23.6 | 96 | 20 | 115 | 1 | 123 | 1 | 132 |
| 4es6_L1 | d1xrsb1 | Cobalamin-binding | 3.40.50.280 | c.23.6 | 88 | 20 | 102 | 14 | 123 | 35 | 153 |

*Abbreviations:* ***Q ID****: query ID,* **S ID***: subject ID,* ***P****: probability,* ***SI****: sequence identity in percent,* ***Cols****: number of aligned residues (columns) in sequence profile* ***Qs****: residue where the query alignment starts,* ***Qe****: residue where the query alignment ends,* **S*s***: *residue where the subject alignment starts,* ***Se****: residue where the subject alignment ends,* ***ct****: C-terminus,* ***nt****: N-terminusl,* ***L1****: lower U3S lobe (circular permutation), and* ***L2****: upper U3S lobe (circular permutation).*

**Query (hemD-like) IDs**: **4es6**: *Pseudomonas aeruginosa*, **3re1**: *Pseudomonas syringae,* ***3mw8****: Shewanella amazonensis*, **3d8r**: *Thermus thermophilus***.**

**Subject IDs**: **3whp:** *Thermus thermophilus*, **d1yioa2**: *Pseudomonas fluorescens***, 4r3u_C**: *Aquincola tertiaricarbonis,* **4hh3_C**: *Rhodobacter phaeroides*, **d1ccw_A (1ccw_A)**: *Clostridium cochlearium***, d1s8na_**: *Mycobacterium tuberculosis*, **5c8a_A**: *Thermus thermophilus* **d2zwma_**: *Bacillus subtilis*, **d1fmfa_**: *Clostridium tetanomorphum***, d2jbaa_**: *Escherichia coli*, **d3to5a_**: *Vibrio cholerae,* **d1ys7a2**: *Mycobacterium tuberculosis,* **d3hdva_**: *Pseudomonas putida,* **d1zh2a1**: *Escherichia coli,* **d3nhma_**: *Myxococcus xanthus* **d2v0na2**: *Caulobacter vibrioides,* **d3q15c_**: *Bacillus subtilis,* **d1dxya2**: *Lactobacillus paracasei.*

## 2.8. HemD-like and periplasmic-binding protein like-I share a common origin

Another fold that shows a clear relationship to hemD-like constitutes the periplasmic-binding protein-like (PBP, SCOP ID c.93). This fold is particularly interesting because it aligns nearly all amino acids in hemD-like fold with probabilities up 93%. The sequence identities between the lobes are rather low (up to 17%), but their architectures and topologies are remarkably similar. In addition, hemD-like and PBPs share similar features such as a bi-lobular architecture and a binding site between their lobes (**Figure 2.8-A to B**). Initially, I hypothesized that one may have derived from the other. For instance, via circular permutation. However, the profile-profile alignments revealed that flavodoxin-like also hits PBP-I twice at its N- and C-terminal halves with sequence identities up to 27%. This suggests that in analogy to hemD-like, PBP-I may have evolved from flavodoxin-like gene duplication. Interestingly, a truncated half of a PBP-I member, the ribose-binding protein (RBP) from *Thermotoga maritima*, expresses *in situ* as a side product of its recombinant expression in *Escherichia coli* (Cuneo, Beese, and Hellinga 2008). At the DNA level, the RBP sequence bares a ribosomal-binding site at about the half of the gene, which can be recognized by *E. coli*. The resulting protein is about the half of the size of the full-length RBP protein, is soluble and thermostable, sustaining the hypothesis of a flavodoxin-like precursor. In fact, previous researches have pointed out the possibility of gene duplication for the emergence of PBPs (Fukami-Kobayashi et al. 1999). Thus, how is it possible that two folds that originated from duplication of the same protein display different topologies? Circular permutation may be the immediate or intuitive answer. However, the topological discrepancies between PBP and hemD-like can be additionally described as the result of an alternative flavodoxin-like segment-swap, namely a C-terminal alpha helix instead of the N-terminal αβα-element in hemD-like (**Figure 2.8-C**). However, whether there are additional steps missing, remains to be investigated.

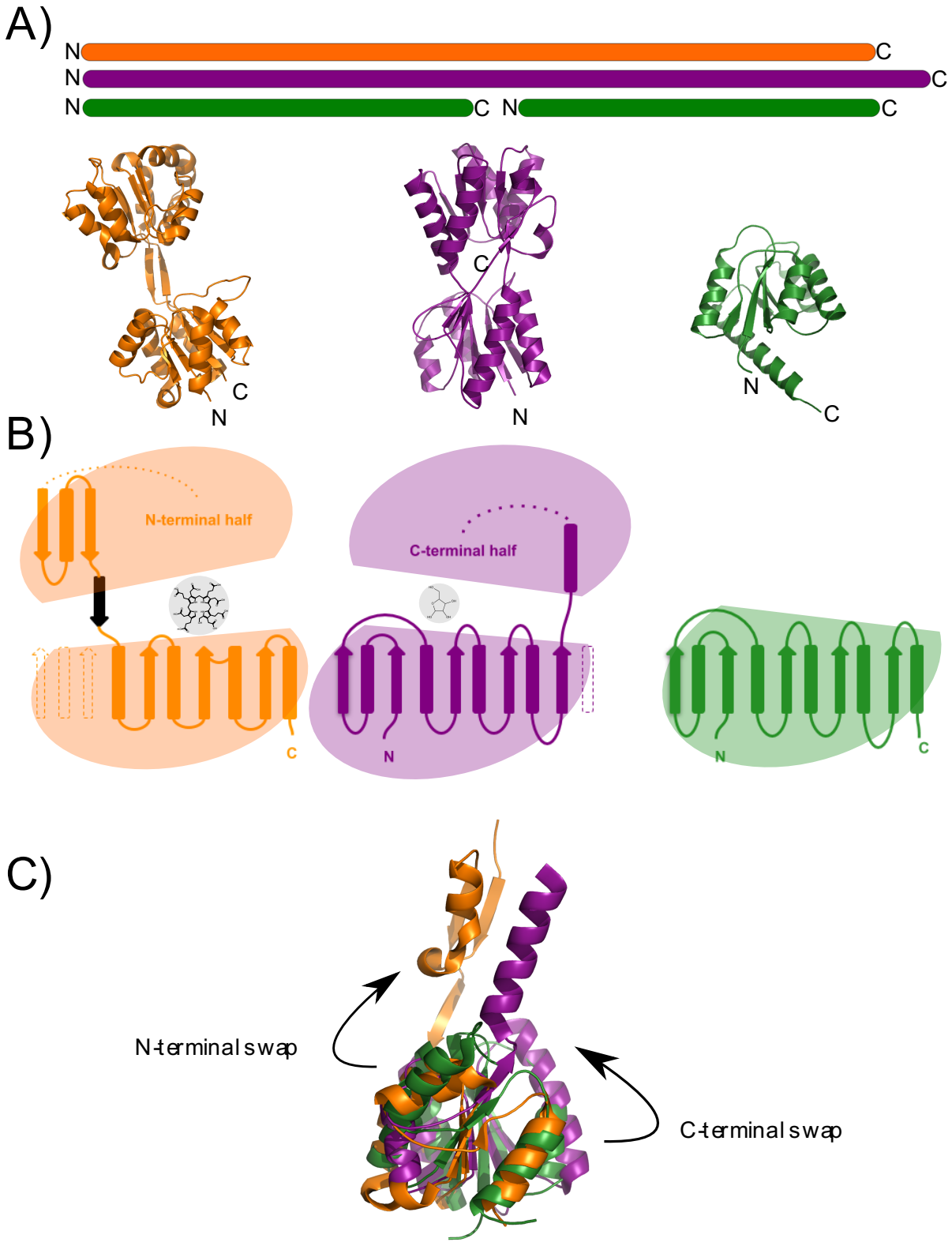



**Figure 2.8 Evolutionary implications of segment-swapping. The case of the flavodoxin-like fold.** Profile-profile comparisons suggest that hemD-like (orange) and periplasmic-binding protein-like I (PBP, purple) folds evolved via flavodoxin-like (green) gene duplication (**A**). HemD-like and PBP display a distinct topology but a similar architecture and both bind their cognate ligands (gray) between their protein lobes (**B**). Structural alignments of the flavodoxin-like fold vs. hemD-like and PBP halves reveal an N-terminal and C-terminal segment-swap, respectively, which appear to cause their different topological connections (**C**).

## 2.9. Discussion

Here we extend our initial search for homology links across domain boundaries (**Section 1.9**) by exploring a particular case of insert-mediated segment-swapping and gene duplication of a flavodoxin-like ancestor that lead to the hemD-like scaffold.

The sequence and structural resemblance between the hemD-like N- and C-terminal halves, suggest that they resulted through gene duplication (**Section 2.3**). Consolidating this hypothesis, the sequence alignment between the flavodoxin-like member LitR[3] and the C-terminal half of U3S[4] provides with statistically significant evidence that supports the flavodoxin-like fold is the duplicated precursor (**Section 2.4**). Finally, the extraction of an insertion presented only in the hemD-like halves, revealed the successful reconstruction of the canonical flavodoxin-like topology (**Section 2.5**).

Based on the gathered results, I suggest that a flavodoxin-like ancestor may have undergone an insertion that mediated a segment-swap. This domain dislocation facilitated the association of two units, leading to the hemD-like bi-lobular architecture. Similar phenomena have been observed as likely artifacts for other flavodoxin-like proteins (pdb identifiers: 4q375 and 3c85) but also playing a role in catalysis in natural enzymes (**Section 2.6**). U3S may have acted in a similar fashion. First the insertion may have mediated better association conditions. Finally, fusion contributed with motion restrains to lower the entropy costs for binding.

Overall, these results illustrate that evolutionary events can be reconstructed experimentally similar to the reconstruction of TIM-barrels from their halves (Höcker, Claren, and Sterner 2004). This time, however, the reconstruction of an ancestral superfold has been assessed, employing the sequence of a different fold, displaying the plastic nature of proteins and highlighting that segment-swaps presumably played a greater role in evolution than assumed before.

[3] Quorum sensing transcriptional regulator (LitR) from *Thermus thermophilus* (pdb: 3whp)

[4] Uroporphyrinogen III synthase (U3S) from *Pseudomonas aeruginosa* (pdb: 4es6)

# Chapter 3

## Design of a cobalamin-binding fold-chimera

### 3.1 Overview

Despite the advances in developing methods for protein design, since the first attempts back in the 70s, humans still cannot compete with Nature in creating highly efficient catalysts. Significant contributions to the field have been achieved through the use of directed evolution (Arnold 2015; Goldsmith and Tawfik 2012). The success of this method is based on mimicking the natural phenomenon of introducing mutations at random, with the eventual result of improving a protein feature. One limitation, however, is that novel functions will be hardly explored by this approach, since the high throughput selection methods that are usually coupled to it, require previous knowledge of the feature to be selected. This issue is partially addressed by *de novo* design, which probes sequence combinations beyond the ones explored by Nature. Nevertheless, this method also has its limitations. For instance, only a fraction of protein designs adopts stable folded structures when recombinantly expressed in *E. coli.* Insolubility and polydispersity are the most common reasons for failure, which most likely arise due to unanticipated hydrophobic interactions (Huang, Boyken, and Baker 2016; Khoury et al. 2014).

Here we pursue the long-standing goal of complex custom-made protein design, addressing some of the limitations of *de novo* design such as foldability by combining homologous fragments. These fragments appear to be a good source for well-folded and solubly-expressed protein chimeras (Höcker 2014). In addition, we address the question, whether pre-existing cofactor or binding pockets can be placed in different contexts from the ones observed for naturally-occurring enzymes, with the long-term goal of employing these cofactors to assist novel functionalities.

## 3.2. Searching for cofactor-binding pockets within homologous fragments

Hidden Markov Models were employed previously to find homologous regions within α/β proteins (**Chapter 1.9**). The gathered results yielded a large number of fragments shared by different folds (**Appendix 1-2**). However, from the obtained fragments, only a small fraction is large enough to bare whole cofactor-binding pockets for our design purposes. One example of a protein pair that shares large homologous regions constitute the flavodoxin-like and hemD-like folds (**Figure 2.3**). As discussed in **Chapter 2**, their sequence and structural resemblance strongly suggest that a flavodoxin-like fold duplication gave rise to the uroporphyrinogen III synthase (U3S). This is advantageous for our strategy, since it can be assumed that practically the whole hemD-like architecture is related to the flavodoxin-like one, meaning that any desired fragment can be replaced by its corresponding region from the other. Therefore, we looked for interesting functions among the 16 flavodoxin-like SCOP families, and identified the cobalamin-binding domain as a suitable target for design.

## 3.3. Functional considerations for the selection of parental proteins

Before proceeding to design the chimeric protein, a number of considerations were taken into account. Special attention was paid to the fact that members of the cobalamin-binding family only trap the sugar moiety of vitamin B12 derivatives (Martin Tollinger et al. 2001) whereas the β-axial (upper) face of the cofactor is usually protected by a second domain, which varies according to the enzymatic function (**Figure 3.1**). For instance, the methylmalonyl-CoA and glutamate mutases act with the help of a TIM-barrel counterpart, whereas both: The O-demethylase and methionine synthase require a four-helix bundle. To date, there is no structure of a cobalamin-binding domain bound to cobalamin in absence of a protecting domain, which may be an indication that a shielding protein is required to keep cobalamin in place. To fulfill these requirements in our design, we anticipated that the upper lobe in U3S could act as domain counterpart (**Figure 3.2**), whose large inner interface towards the binding region could be advantageous to introduce further mutations for design. In addition, the U3S scaffold displays a large cavity that provides enough space to accommodate the bulky cobalamin, once the binding pocket is transplanted (**Figure 3.2**).

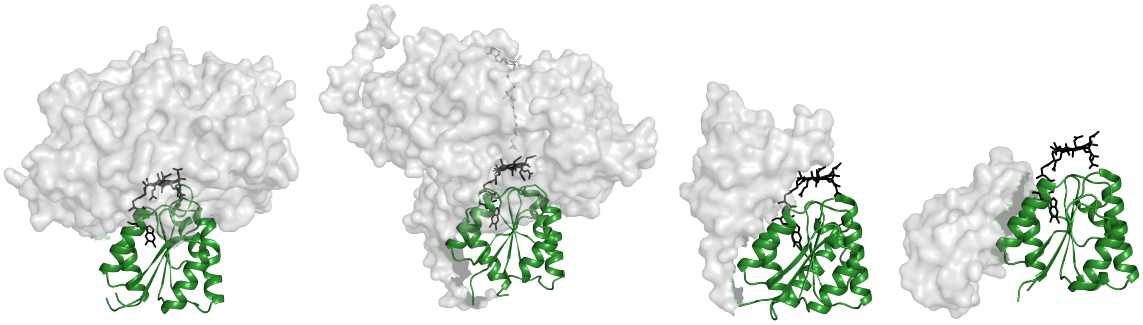


**Figure 3.1 Cobalamin-binding domain and interacting partners.** The structures illustrated above correspond to four cobalamin-dependent enzymes, whose cofactor-binding domain is shown in green, domain counterparts in gray and the bound cofactor in black. From left to right are displayed: glutamate mutase (pdb: 1ccw), methylmalonyl-CoA mutase (pdb: 7req), methionine synthase (pdb:1bmt) and O-demethylase (pdb: 4jgi). Even though the last example appears not to shield cobalamin with the helper domain, in the crystal structure the cofactor is covered by the helical bundle domain from a vicinal protein. These examples show that protecting the cofactor from solvent may be necessary for keeping it in place or for proper function.

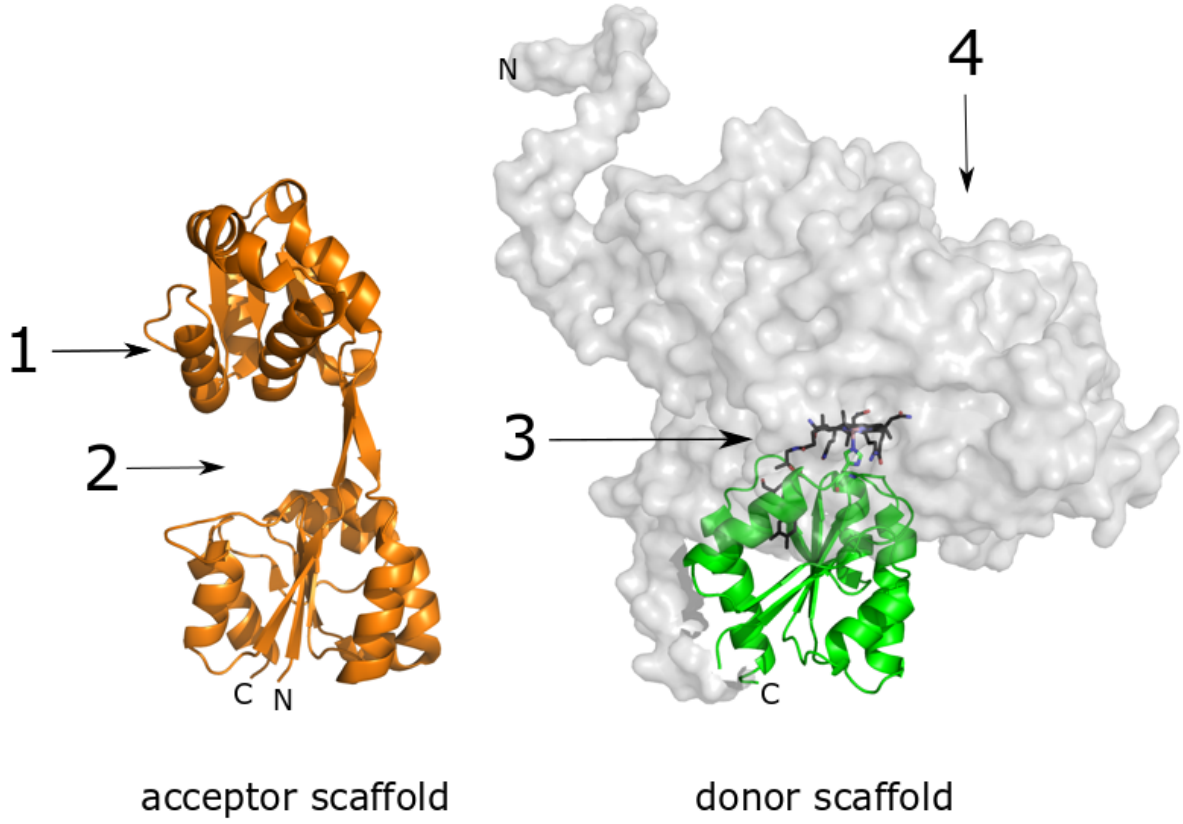



**Figure 3.2 Functional considerations for hybrid design.** The upper lobe of U3S provides a shield for the ligand from solvent as well as a large inner interface for further design mutations **(1)**. In addition, the cleft between its lobes provides enough space **(2)** to accommodate the bulky cobalamin **(3)**, which is shielded by a domain counterpart in all cobalamin-dependent enzymes **(4)**. Pdb identifiers: 1jr2 (orange) and 1req (green-gray).

Only a handful of uroporphyrinogen III synthase structures have been solved. Nevertheless, there are enough of them to illustrate the wide range of relative domain orientations their protein lobes can adopt. The structures display aperture differences of their lobes of up to 23.6 Å and rotations up to 90° from one structure to another (Schubert et al. 2008). We selected the human ortholog over the others, because it displays the largest opening between its lobes. This characteristic was highly desired, since a large opening may facilitate accommodation of a cobalamin molecule without encountering steric problems (**Figure 3.3**). Another crucial aspect to consider for the chimeric design is undoubtedly the sequence and structure similarity. We selected the methylmalonyl-CoA mutase from *Aeropyrum pernix* as cofactor donor, since it showed the highest sequence identity (23%) to the human U3S and high structural resemblance (**Figure 3.3**). In addition, this protein has been solved as a single domain and comes from a highly thermostable organism, which are both properties that we considered advantageous for design, as not all domains are stable when they are isolated from their counterparts.

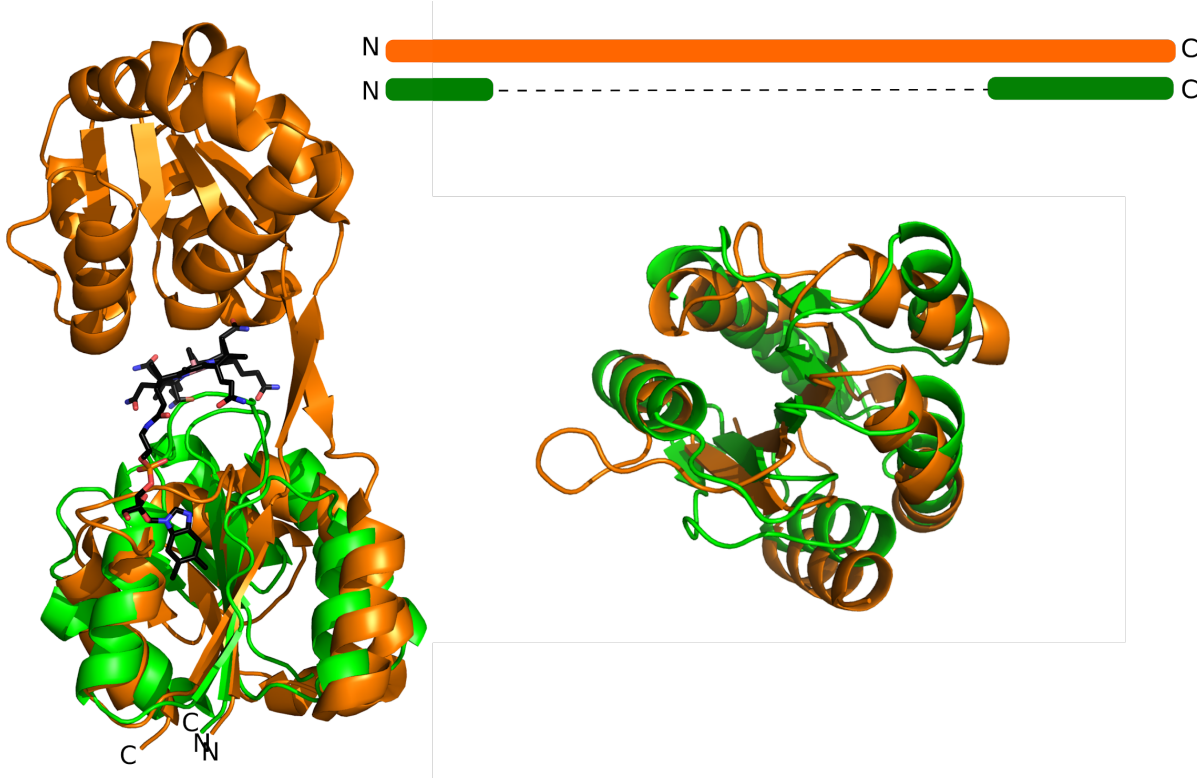



**Figure 3.3 Structure alignment of selected parental proteins.** The structure of the cobalamin-binding domain (green) of methylmalonyl-CoA mutase (MCoAM) from *Aeropyrum pernix* (pdb: 2yxb) superimposed onto the lower lobe of the human uroporphyrinogen III synthase (orange, pdb: 1jr2:A) shows high structural resemblance. The cobalamin molecule has been placed via structural alignment of the crystal structure of the MCoAM of *Propionibacterium freudenreichii* (pdb: 1req) superposed to the apo structure of MCoAM from *Aeropyrum pernix* (pdb: 2yxb) to illustrate that U3S confers enough space for placing cobalamin (black).

## 3.4. Fold-chimeragenesis

To define the chimeric crossovers, we searched for well-superposed areas, ideally containing similar residues. In addition, we aimed to cut in the middle of secondary structural elements instead of loops to build hybrid secondary structural elements. First, the N-terminal cut site was chosen where the residues at the desired transition were very similar and the corresponding structures well-aligned (**Figure 3.4**). Thus, we truncated the sequence of methylmalonyl-CoA mutase (MCoAM) at residue V52, joining the sequence of uroporphyrinogen III synthase (U3S) starting with L32, building a chimeric beta strand. The C-terminal crossover, however, was a bit more challenging due to an insertion of one amino acid in the sequence of U3S. This residue changes slightly the register of the helix in respect to the flavodoxin-like fold and its distance to the protein core (**Figure 3.4**). As a consequence, the sequence-based profile-alignment does not correspond to the structural one. For that reason, we looked for similar residues to build the transition, finding that we could replace L180 of U3S by V63 of MCoAM. The resulting chimeric sequence and protein model are shown in **Figure 3.5**.

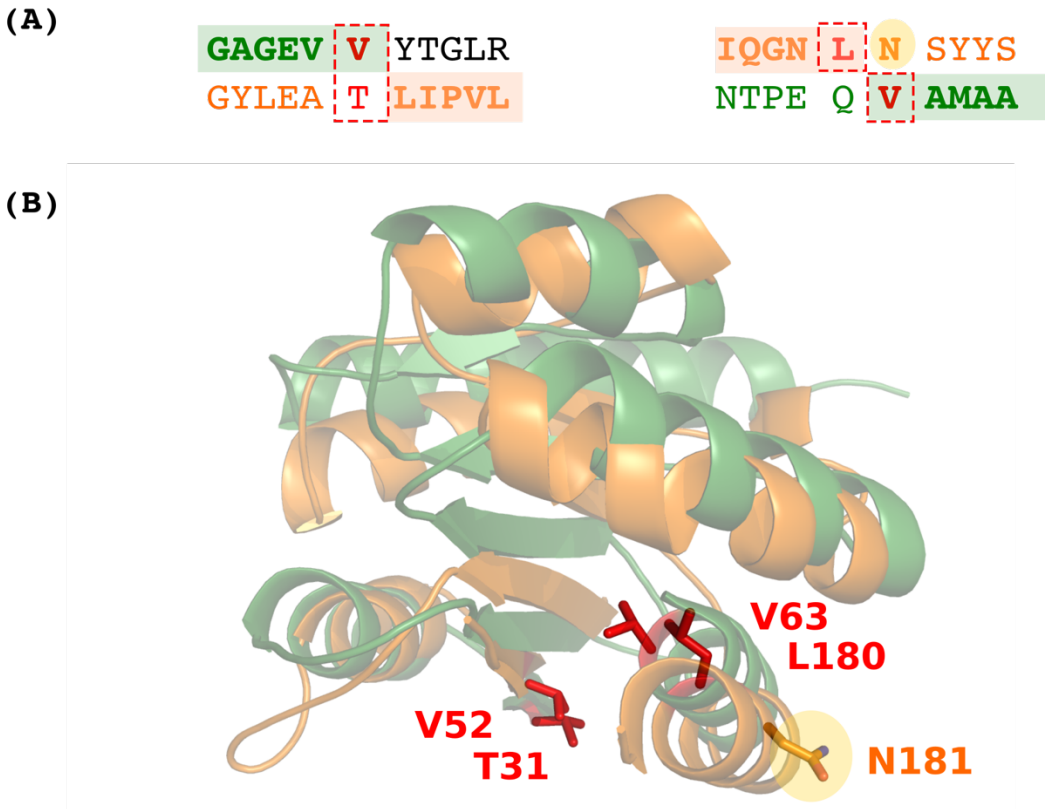



**Figure 3.4 Selection of chimeric crossovers.** The N-terminal crossover was performed according to the sequence alignment by HHsearch, replacing U3S-T31 by MCoAM-V52, which correspond also at a structural level (**A**-left). In contrast, the sequence alignment by HHsearch (A-right) differs from the structural alignment (**B**). Whereas MCoAM-V63 aligns in sequence with U3S-N181 (yellow), it aligns structurally with U3S-L180. This discrepancy is owed to an insertion in the U3S sequence, which causes a change in the helix register by one amino acid. Therefore, we followed the structural alignment, replacing U3S-L180 by MCoAM-V63.

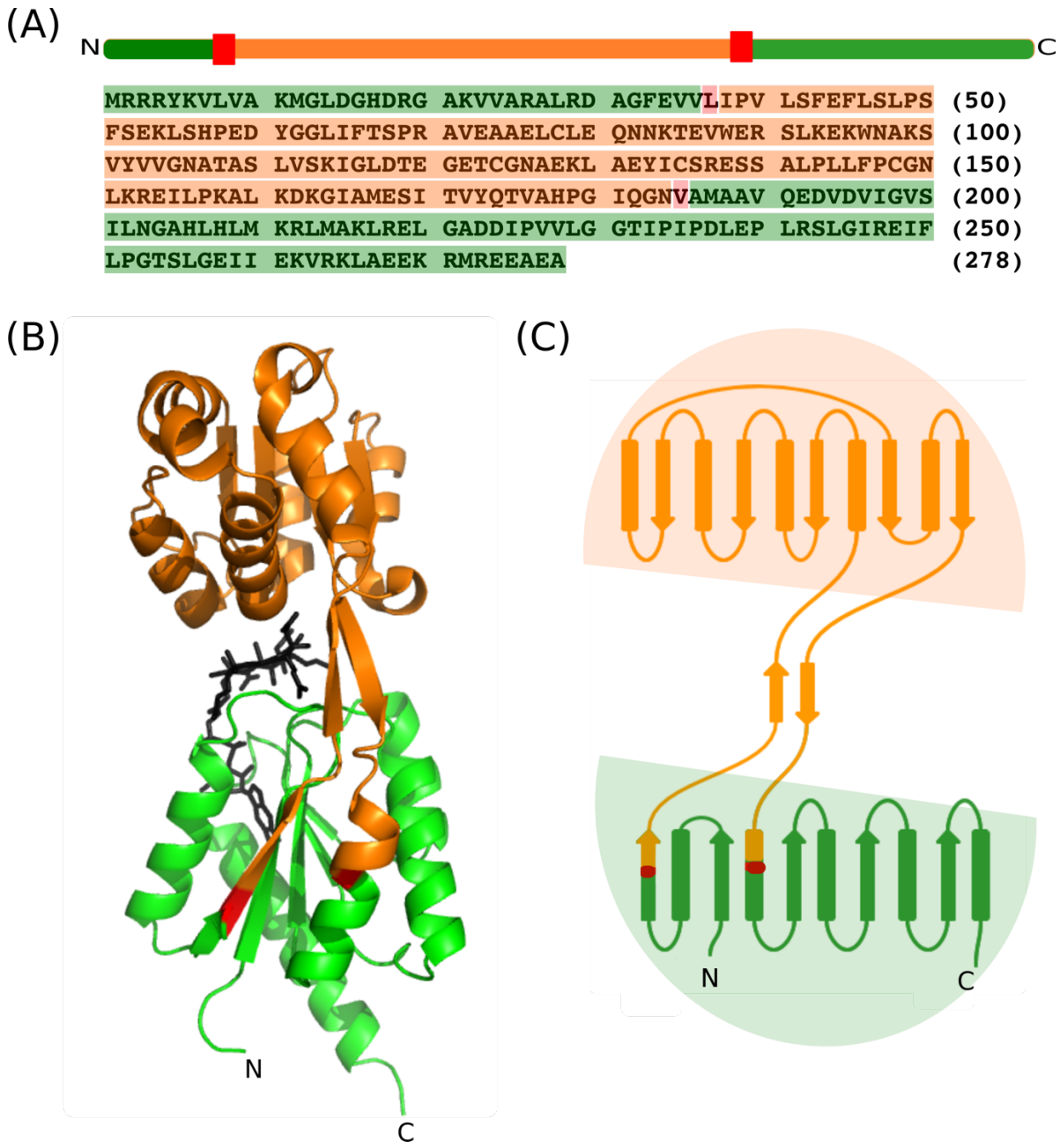



**Figure 3.5 Chimeric design and hybrid model.** Chimeric protein sequence **(A),** model **(B)** and topological connections **(C).** Crossovers are shown in red and cobalamin has been placed, superimposing the cobalamin from methylmalonyl-CoA mutase from *Propionibacterium freudenreichii* (pdb: 1req).

## 3.5. Cloning, protein expression and purification

The chimeric gene was assembled by PCR employing two fragments from MCoAM and one fragment from U3S, all fragments contained overlapping regions to facilitate annealing (**Figure 3.6**). The MCoAM fragments were amplified from the construct 13AN6UP_MMcM_pMA purchased from Life Technologies (**Section 5.5.1**). The U3S region was amplified from template pET16b-U3S, which is described elsewhere (Mathews 2001) and was kindly handed out by Prof. Heidi L. Schubert. All yielded fragments were loaded onto an agarose gel (1%) for purification by gel extraction. Purified fragments were all mixed for a final PCR reaction employing outer primers bearing restriction sites for *XhoI a*nd *NdeI*. The obtained gene was digested and cloned into a pET21a expression vector, yielding the construct pET21a-UShsMMap01. The final construct was sequenced at the sequencing facility of the Max Planck Institute in Tübingen, Germany.

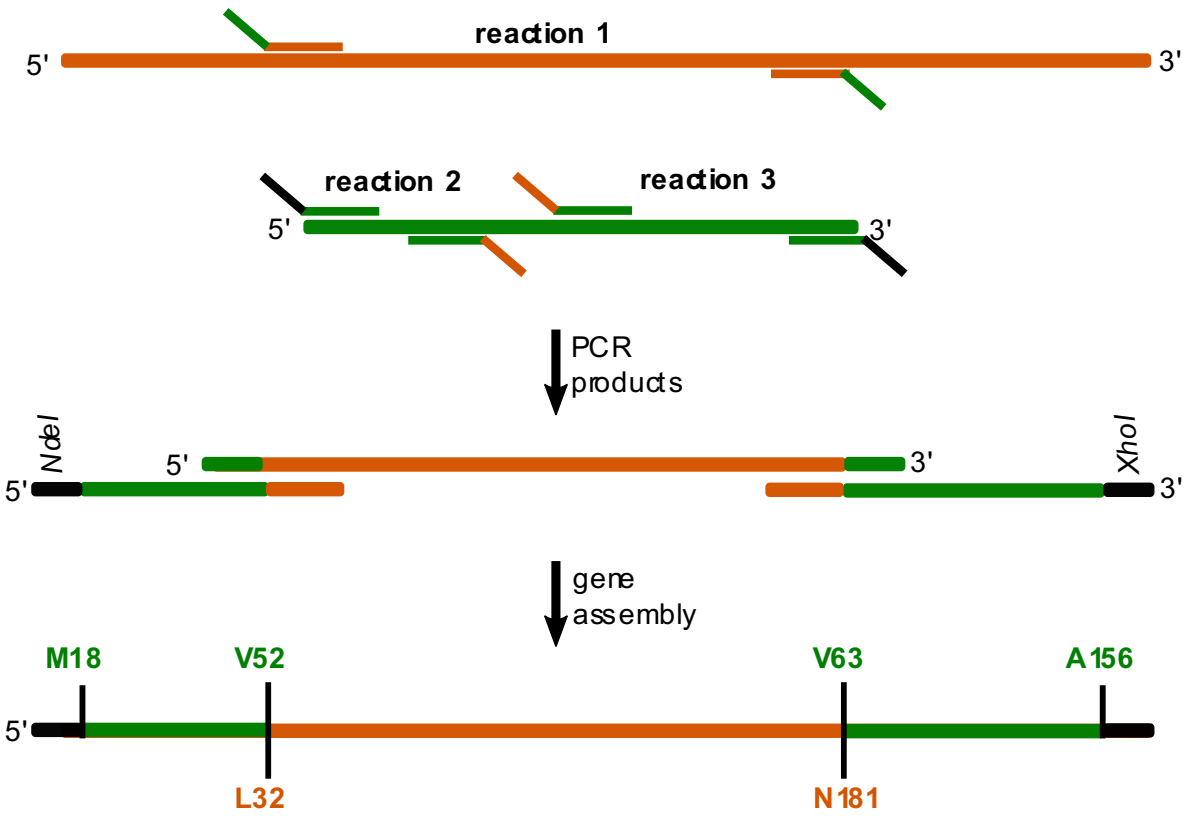



**Figure 3.6 Gene assembly** Chimeric gene construction was performed employing primers with complementary regions. In the first step, one fragment from the human U3S (orange) and two fragments from MCoAM (green) from *Aeropyrum pernix* were amplified in individual PCR reactions, using adequate primers (**Section 5.3.1**). These fragments were then assembled in a final PCR reaction, employing outer oligos to yield the hybrid gene, bearing restriction sites for *NdeI* and *XhoI*.

Protein expression of the parental and chimeric constructs was performed in BL21 cells at 20°C and in ArcticExpress cells (Stratagene®) at 11°C (**Section 6.11**). The obtained cell pellets were mixed with **Buffer-A** and disrupted by sonication (**Section 6.12**). All proteins expressed solubly. However, the hybrid protein additionally showed high amounts of insoluble protein in inclusion bodies. Protein purification was performed via nickel affinity and size exclusion chromatography. Refolding was also performed to make use of the large amount of proteins in the inclusion bodies (**Section 6.13**). The refolded protein corresponded to the solubly expressed hybrid according to its biophysical characterization. Therefore, we used both proteins, treating them as equivalent.

## 3.6. Biophysical characterization

All proteins run as monomers when loaded onto an analytical gel filtration column (**Figure 3.7-A**), and eluted as single peaks independent of their concentrations. Elution volumes and calculated molecular weight can be found in **Table 3.1**. Evaluation of the secondary structural content was carried out by circular dichroism (**Section 6.17.2**). As expected, the obtained curves corresponded to proteins containing α and β structural elements and the shapes of their curves were very close to one another (**Figure 3.7-B**). Fluorescence spectra corresponding to the hybrid and parental U3S proteins were compared. Both signals showed a maximum at 317 nm and a second shoulder at 333 nm (**Figure 3.7-C**). The presence of two peaks may result from two tryptophan residues that are more and less hidden from the solvent, or due to different conformers of the highly flexible proteins. Differences in intensity are likely due to the different content of chromophores, since the protein concentrations and buffers used were the same (**Section 6.17.3**). Thermal denaturation indicated that the hybrid protein is slightly less stable than the parental U3S and displays a two-step denaturation, perhaps owed to the fact that the fragments from MCoAM come from a thermostable organism and the one corresponding to U3S is from human (**Figure 3.7-D**).

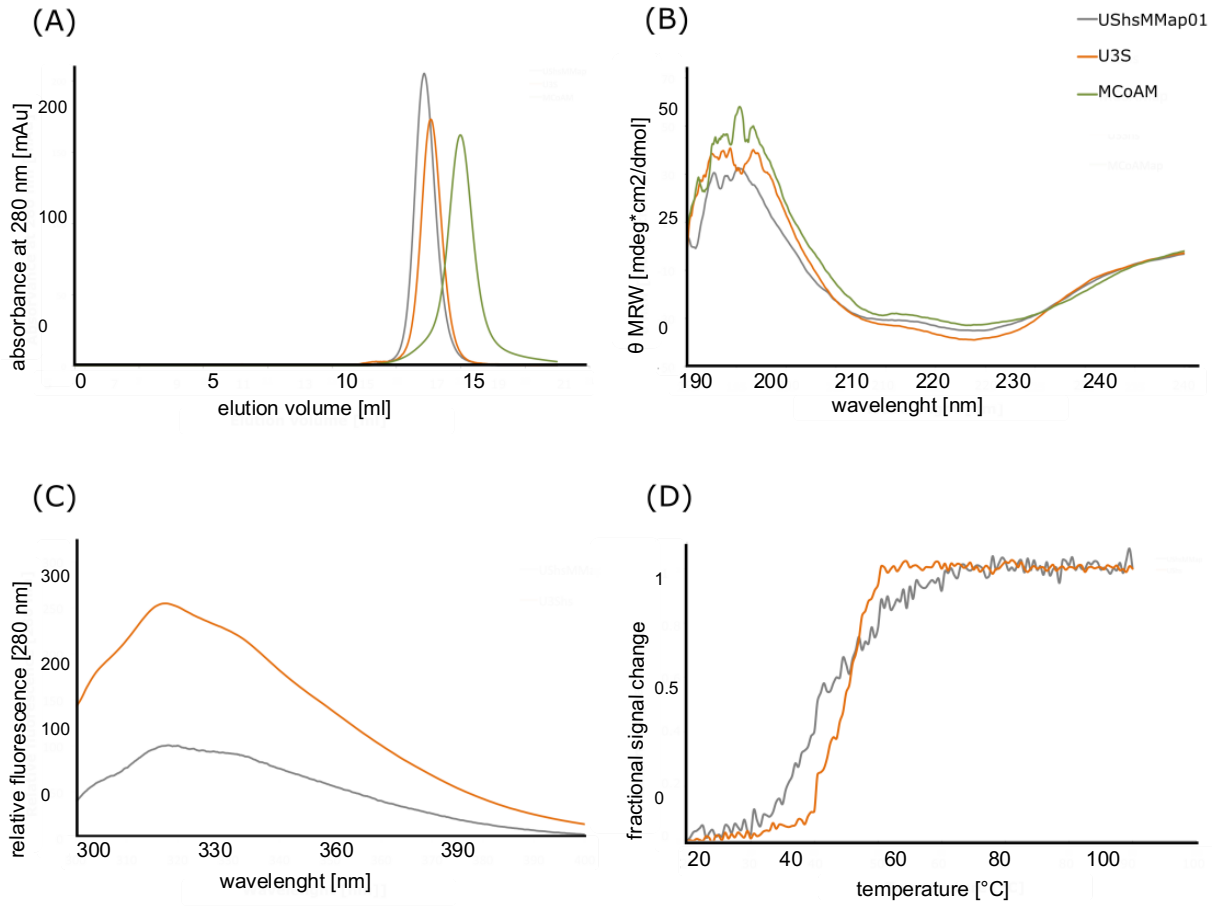



**Figure 3.7 Biophysical characterization of parental proteins and hybrid UShsMMap01.** Analytical gel filtration showed that all proteins eluted exclusively as monomers **(A)**. Comparison of their CD spectra revealed similar signals characteristic of proteins containing αβ secondary structural elements **(B)**. Fluorescence of parental U3S protein vs. the chimera showed different intensities due to distinct chromophore composition in their sequences. However, both displayed two defined peaks likely owed to the only two tryptophan residues, which are present in both constructs **(C)**. Thermal denaturation revealed that the hybrid protein displayed a two-step denaturation, perhaps due to different thermal stabilities of the parental fragments employed (extremophile vs. mesophilic). In addition, a decrease in thermal stability was observed in comparison to the parental protein **(D)**. Color coding: hybrid protein UShsMMap01 in gray, parental proteins U3S and MCoAM in orange and green, respectively. Please refer to buffer section (**Section 5.2.2**) for details of solvent composition for each experiment.

**Table 3.1 Protein characterization by analytical gel filtration.**

| Protein variant | Elution volume [ml] | Theoretical MW [kDa] | Experimental MW [kDa] |
|---|---|---|---|
| UShsMMap01 | 13.0 | 32.5 | 39.6 |
| U3S | 13.4 | 31.0 | 37.1 |
| MCoAM | 14.8 | 16.5 | 16.9 |

## 3.7. Structure determination and analysis

*Crystal growth.* Protein crystals for UShsMMap01 were obtained at a concentration of 12 mg/ml of a pure protein sample in a 96-well plate via the hanging drop method at 16°C. Each well contained 0.4 µl of protein sample and 0.4 µl of screening buffer. The best crystals grew in 0.1 M sodium cacodylate at pH 6.0 and 10 % PEG 4000. Crystal optimization was performed employing the same crystallization method, increasing the drop size to 1.5 µl under same buffer and temperature conditions. The obtained crystals were cryo-protected in 30% PEG 400 and frozen with liquid nitrogen before data collection.

*Data collection.* Crystallographic experiments were carried out on the X10SA beamline at the Swiss Light Source, Paul Scherrer Institute (PSI) in Villingen, Switzerland. The data corresponding to a single crystal was collected at 100K and 0.5 oscillation degrees. The obtained images were recorded on a Pilatus 6M at a detector distance of 270 mm and a wavelength of 1.000Å.

*Structure determination*. The obtained reflections were indexed, integrated and scaled with go.com. This software is a pipeline that includes: 1) XDS for indexing, integration, and data correction (Kabsch 2010), 2) Pointless for space group determination (Evans 2006) , 3) Scala for scaling (Evans 2006), 4) XDSCONV for conversion (Kabsch 2010) and 5) Phenix_xtriage (Adams et al. 2010) for data quality evaluation. Molecular replacement was performed with Phenix.phaser using pdb ids 2yxb and 1jr2 as assemble molecules, omitting the regions not present in the chimera. An initial model was built using Phenix.autobuild (Adams et al. 2010). Final refinement resulted in R-work and R-free values at 19% and 24%, respectively. Final statistics corresponding to refinement are summarized in **Appendix 7.4**.

*Structure analysis.* Elucidation of the hybrid structure revealed an unexpected orientation of the protein lobes (**Figure 3.8**), which behave as independent domains due to high flexibility at the connecting region. For instance, an antiparallel beta sheet connects the lobes in the parental U3S, which is shorter in the hybrid. In contrast, the fragments employed for chimeric building were highly similar to their parents, showing RMSD values of 1.0 Å over 129 Cα atoms for the MCoAM fragment and 0.8 Å over 129 Cα atoms for the U3S fragment, respectively. A closer look at the crossovers revealed that the chimeric beta-strand at the N-terminal cut site was successfully recreated and even elongated. However, the chimeric helix at the C-terminal crossover was disrupted right after the cut site (**Figure 3.9**). This issue appeared to result from a clash caused by residue Q198 as discussed in detail in **Section 3.8.1**, causing partial structural loss of the hybrid helix, conferring extra flexibility to the hybrid

lobes, which open and rotate 145 and 35°, respectively, compared to the U3S scaffold (**Figure 3.8**).

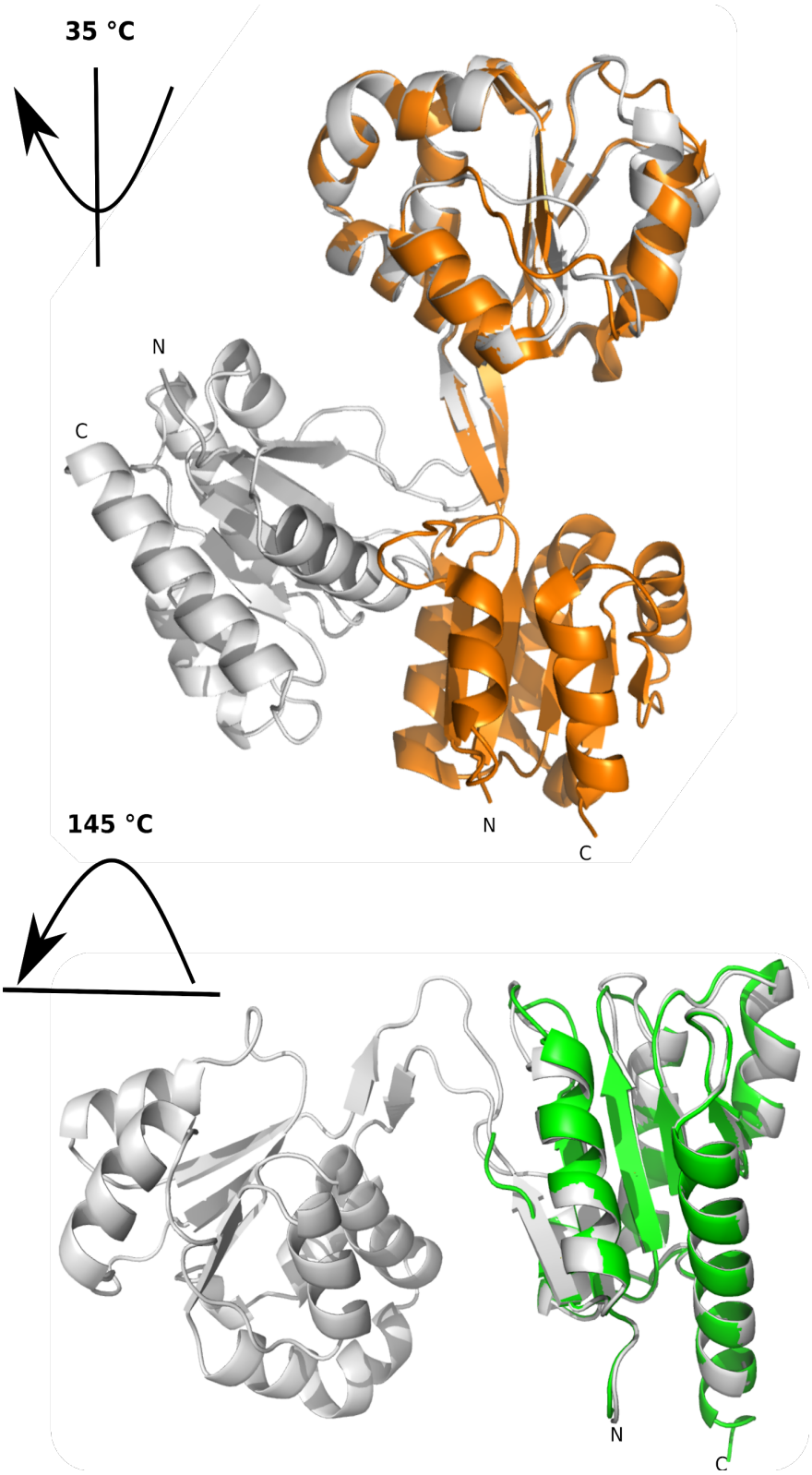



**Figure 3.8 Structural analysis of chimeric variant UShsMMap01.** Elucidation of the hybrid structure revealed an unexpected flexibility, which allows an increased opening and rotation angles in respect to the parental protein U3S of 145 and 35°, respectively.

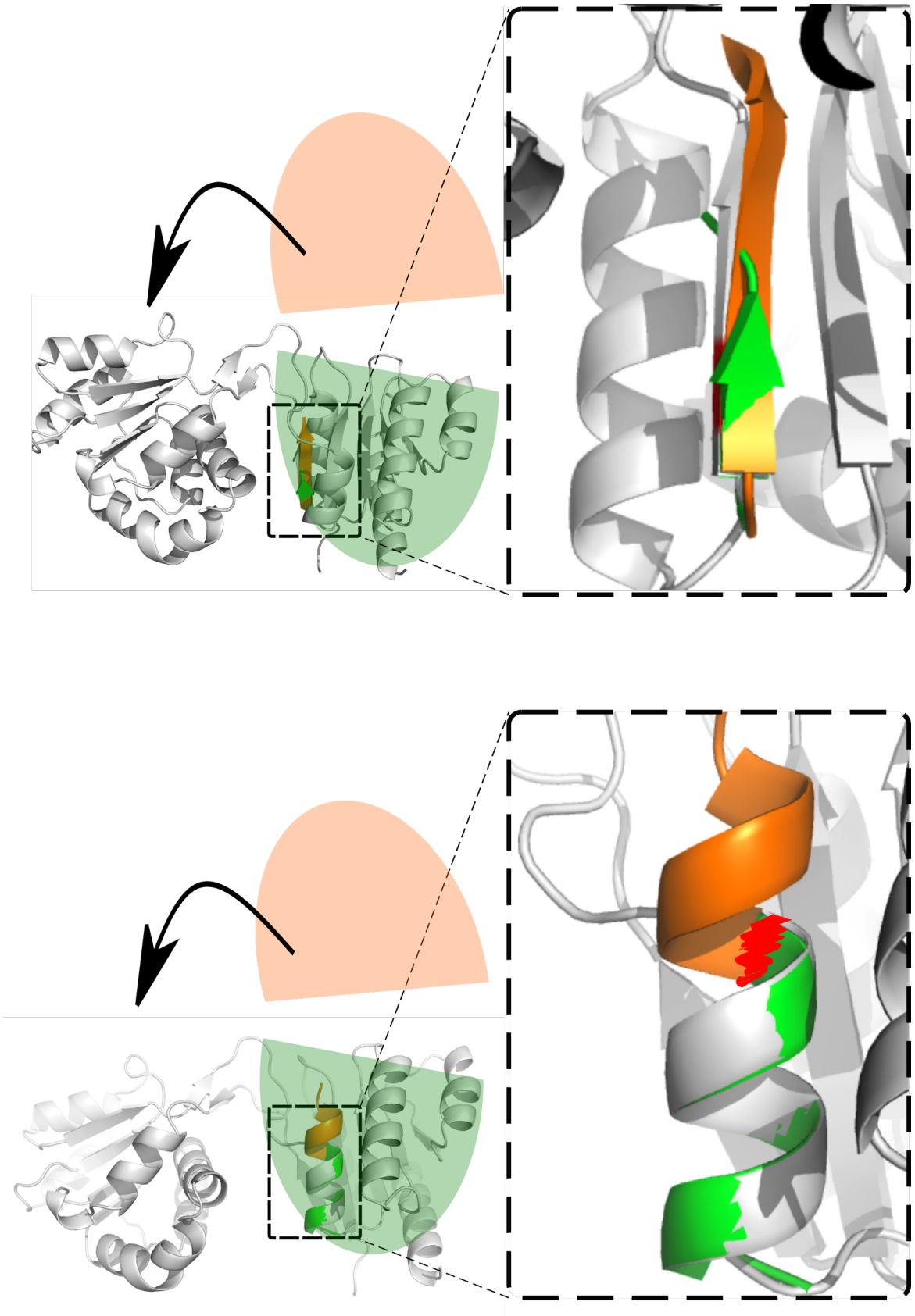


**Figure 3.9 Structural analysis of crossovers in UShsMMap01.** N-terminal crossover showed that the chimeric beta strand (gray) became elongated with respect to the parental MCoAM (green), following the directionality of the parental U3S chain (orange). In contrast, the C-terminal cut site displayed a disruption of the resulting chimeric helix (gray) right at the transition between the parental sequences (red).

## 3.8. Strategies for structural improvement

Four different strategies were followed to decrease the flexibility between the lobes of the chimeric protein, resulting in 5 additional variants that are summarized in **Figure 3.10**. First, a clashing residue was eliminated to elongate helix 2, yielding UShsMMap02 (**Section 3.8.1**). In addition, we searched for residues in the U3S parental protein that seem to contribute to rigidity to the beta meander of the parental protein and inserted a mutation according to this finding. This resulted in UShsMMap03 (**Section 3.8.2**). In parallel, we eliminated a glycine residue in helix 2, anticipating that it may contribute to its disruption. Therefore, we mutated this glycine to an alanine, which is known to facilitate helix formation, as well as to a glutamate, which was present in several structures at this position of the U3S sequence profiles, resulting in variants UShsMMap04 and UShsMMap05 (**Section 3.8.3**). Finally, a helix cap was introduced to attempt helix elongation and restore the highly conserved sequence that is present in cobalamin-binding proteins at the end of helix 2. This modification yielded UShsMMap06 (**Section 3.8.4**).

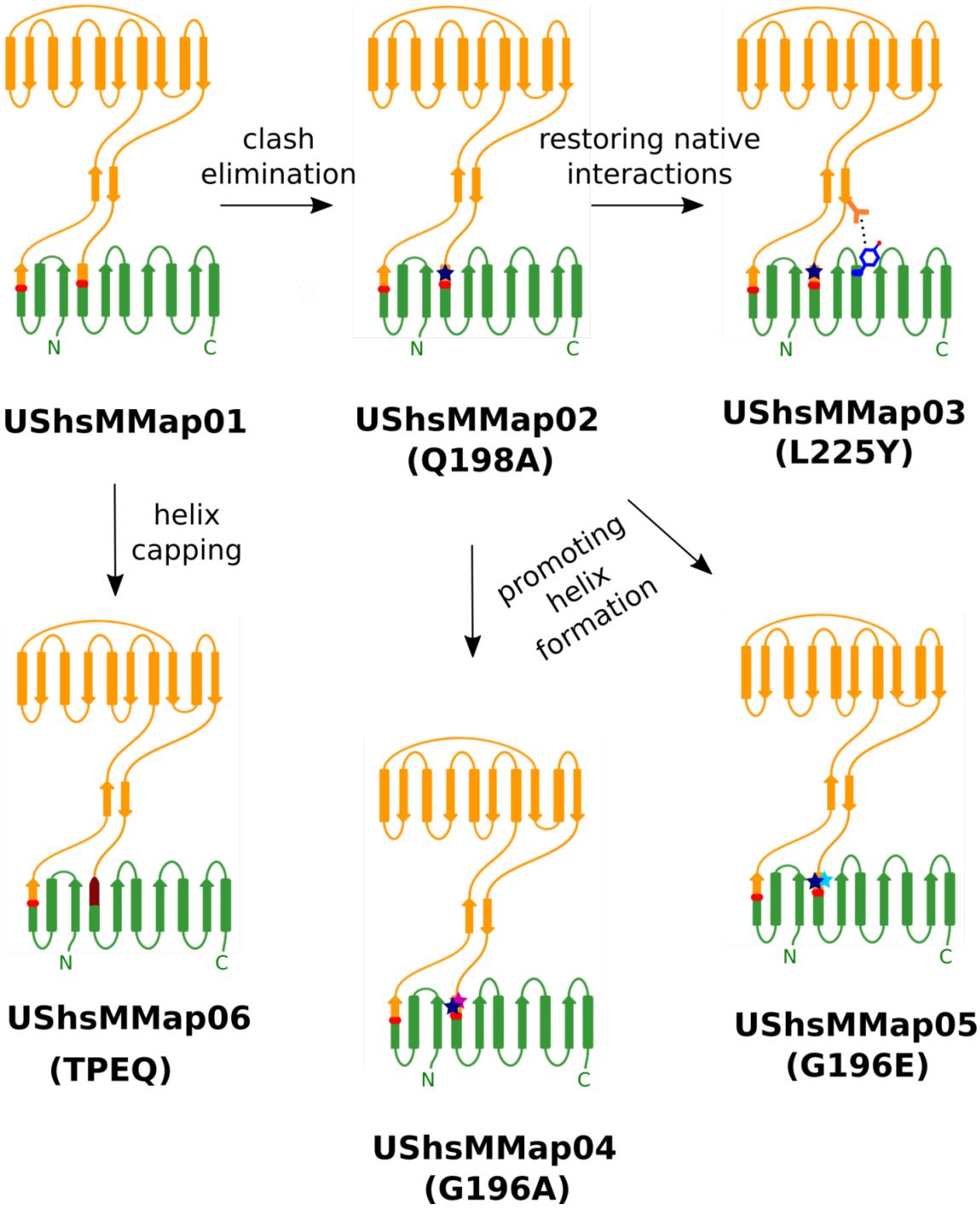



**Figure 3.10 Different strategies for structural improvement.** A clashing residue (Q198A) was eliminated (blue star), yielding UShsMMap02. Subsequently, an interaction in U3S parental protein that appears to join the connecting beta sheet to the lower lobe was introduced to the flavodoxin-like fragment (L225Y) resulting in variant UShsMMap03. Two additional variants were designed to elongate the helix at the crossover: First introducing G196A and finally G196E (pink and magenta stars, respectively). These mutations yielded variants UShsMMap04 and 05. A helix cap (TPEQ, purple cap) was introduced at the end of the transition helix to improve helicity, resulting in variant UShsMMap06. Fusion sites are shown as red squares.

### 3.8.1. Chimeric helix elongation via elimination of clashes

As described in **Section 3.4**, definition of the C-terminal crossover was not as obvious as the N-terminal one, since the register of the aligned helices change slightly due to an insertion in U3S compared to MCoAM. In addition, the distance between helix 2 and the vicinal helix 3 is larger for U3S than for MCoAM (**Figure 3.11**). Anticipating that Q198 may clash in the resulting chimera, the Q198A mutation was introduced via directed mutagenesis (**Section 6.2.4**). The yielded protein, termed UShsMMap02, behaved very similar to the original hybrid in terms of their biophysical characterization but the elucidation of its structure by X-ray crystallography (**Appendix 7.5**) revealed that the flexibility of the connecting loops was partially fixed (**Figure 3.12**). However, the two lobes remained distant, indicating that the Q198 residue contributed only partially to lobe aperture.

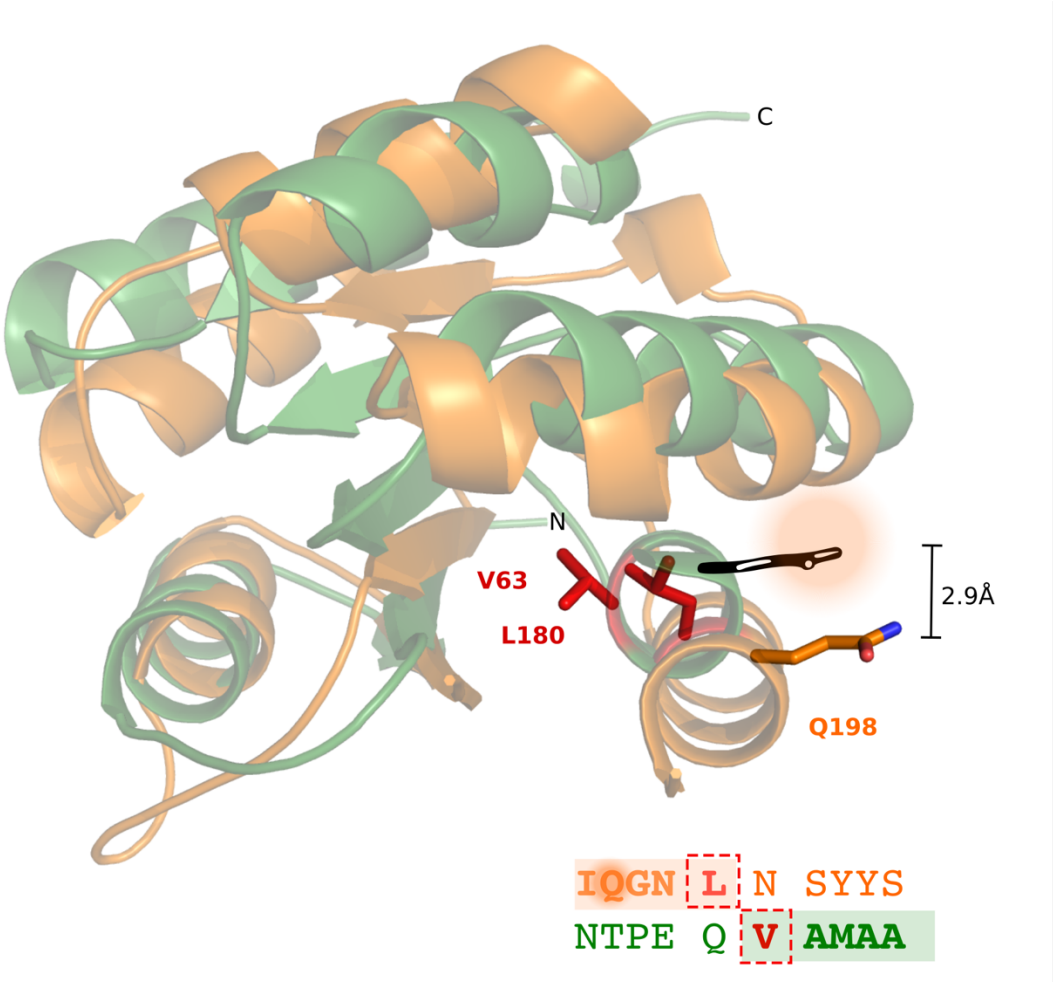



**Figure 3.11 Structural alignment of U3S onto MCoAM displays complications at C-terminal crossover.** The helix H2 in the flavodoxin-like fold (green) is 2.9 Å closer to the protein core compared to its corresponding helix in hemD-like fold (orange). As a consequence, the transition at this helix gets disrupted by Q177 (Q198 in the chimera), which clashes in its new position (black). Residues in red illustrate the chimeric cut sites.

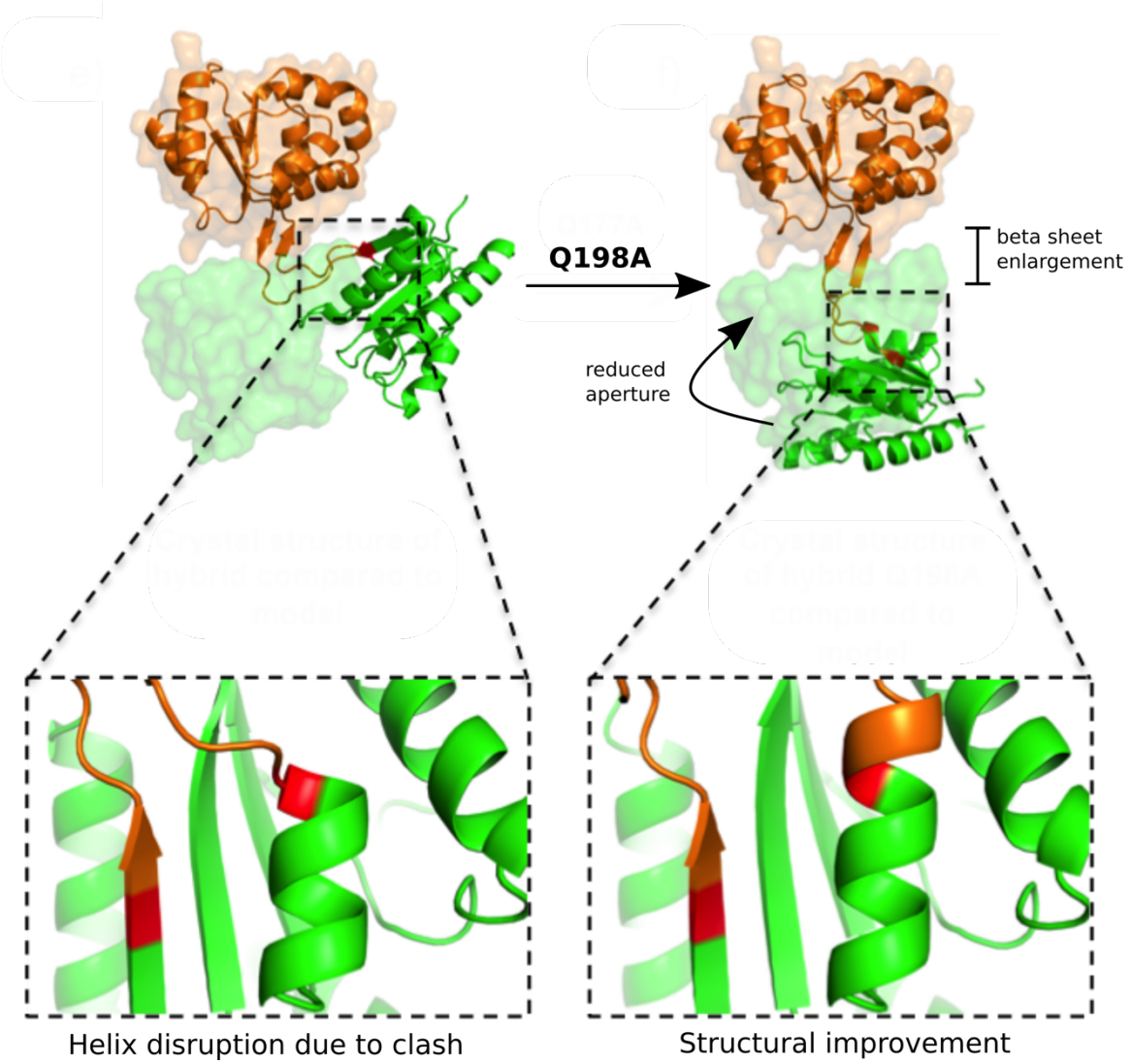



**Figure 3.12 Structural improvements by Q198A mutation.** The detected clashing glutamine residue was mutated to an alanine, extending the chimeric helix and decreasing the aperture distance between the lobes by about 45°. Protein model as surface and crystal structures as cartoon.

### 3.8.2. Searching for residue-residue contacts to favor interaction between lobes

As an alternative strategy we searched in the parental U3S structure for interactions that may induce its mono domain architecture. We found that in the connecting beta-bridge the strands interact via π–π stacking through residues R65, F38, A172 and Y203 (**Figure 3.13**). All these residues are present in the hybrid except for Y203, which is located in the region replaced by the MCoAM fragment. In an attempt to restore a similar interaction, we mutated the corresponding residue in the chimeric protein, leucine 225, to a tyrosine. Unfortunately, the elucidation of this variant's X-ray structure (**Appendix 7.6**) showed that the flexibility between the chimeric lobes could not be improved by this strategy.

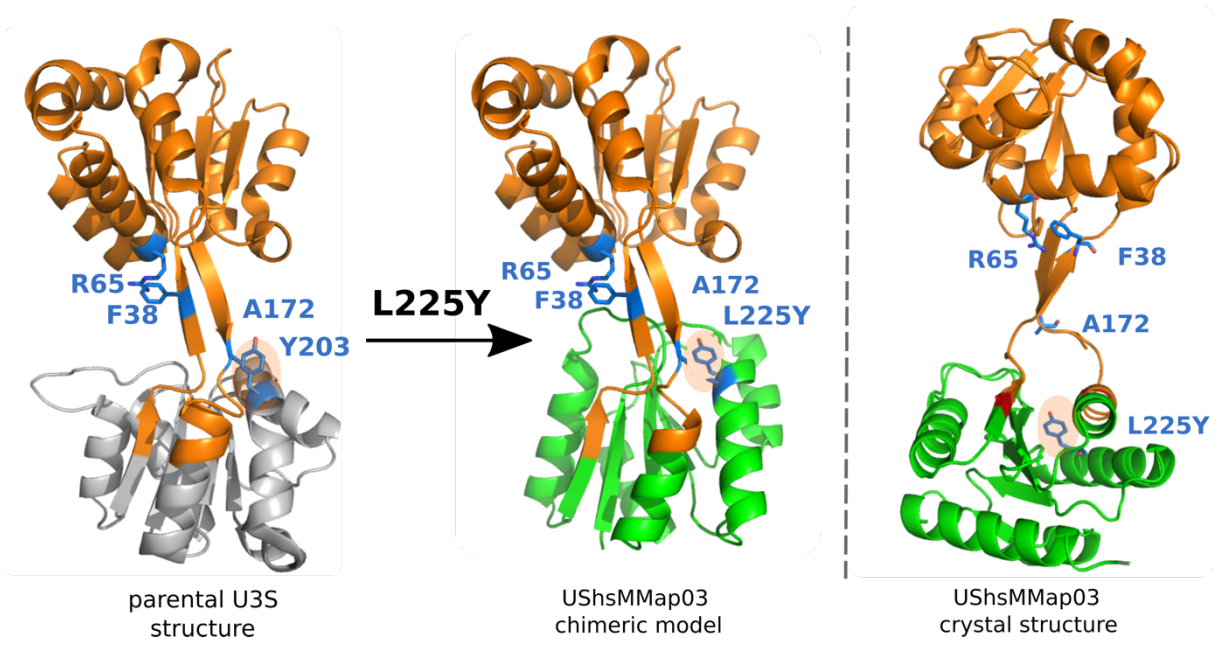



**Figure 3.13 Structural improvement attempted via aromatic interactions.** The U3S parental protein displays four residues that interact via π–π stacking (left). One of these residues, namely Y203 that was absent in the hybrid was introduced in a corresponding area, yielding UShsMMap03 (UShsMMap02-L225Y). Elucidation of its X-ray structure revealed that the introduced tyrosine did not interact as desired (right).

### 3.8.3. Further stabilization of the chimeric helix via mutations

Another attempt to enlarge helix 2 was performed by mutating the residue G196. Glycine residues are well known for disrupting helices and even though this residue builds a helix in the parental U3S structure, it may not contribute to build a helical structure in its new amino acid environment. Therefore, we mutated this residue to an alanine (UShsMMap04) or a glutamate (UShsMMap05) (**Figure 3.10**). It is well known that alanine residues promote helicity. Glutamate was selected, since this residue was highly populated in the U3S sequence-profile at this position. Biophysical characterization of both variants corresponded to the original hybrid, being well folded and solubly-expressed. Nevertheless, only the variant G196A yielded good-quality protein crystals to solve a structure by X-ray crystallography. The elucidation of UShsMMap04 (**Appendix 7.7**) revealed that its conformation was similar to this of UShsMMap02, in which the protein lobes behave independently (**Figure 3.14**). Interestingly, the conformation of the aperture in these variants is different (**Figure 3.15**), corroborating further the high flexibility of the connecting loops observed for UShsMmap02 as well. This property is also observed for the parental U3S proteins, which show apertures and rotation distances up to 90° and 26.6Å, respectively.

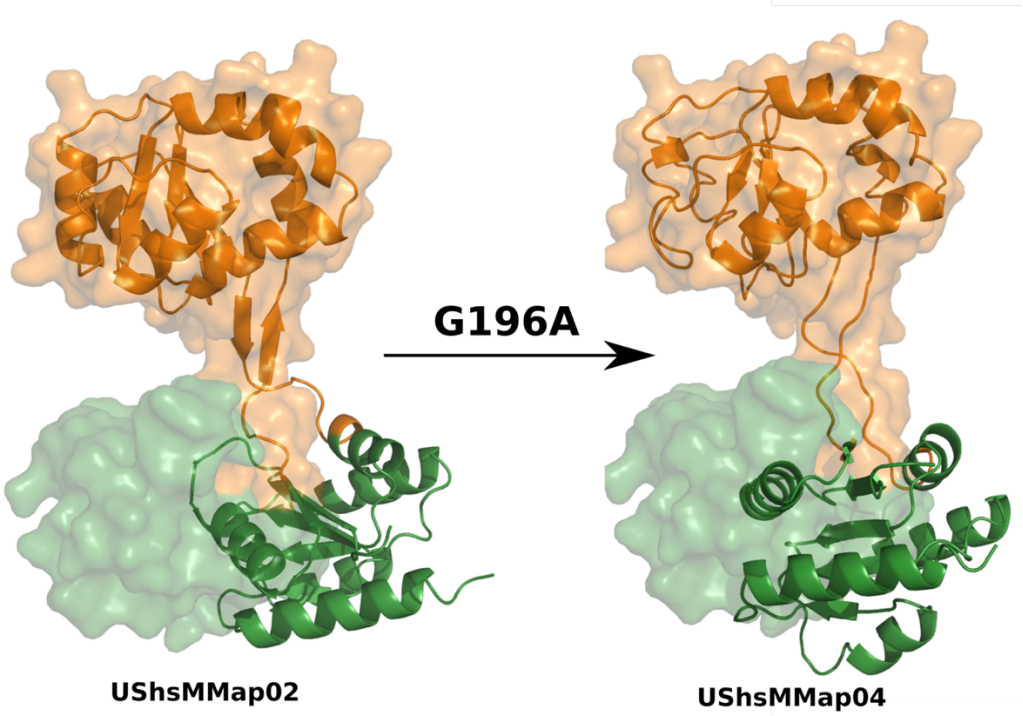



**Figure 3.14. Elucidation of the UShsMMap04 structure by X-ray crystallography.** The conformation of the protein lobes of UShsMMap04 is similar to UShsMMap02. The lobes also behave independently due to the high flexibility at the loops. Protein model is shown as surface and crystal structures as cartoon.

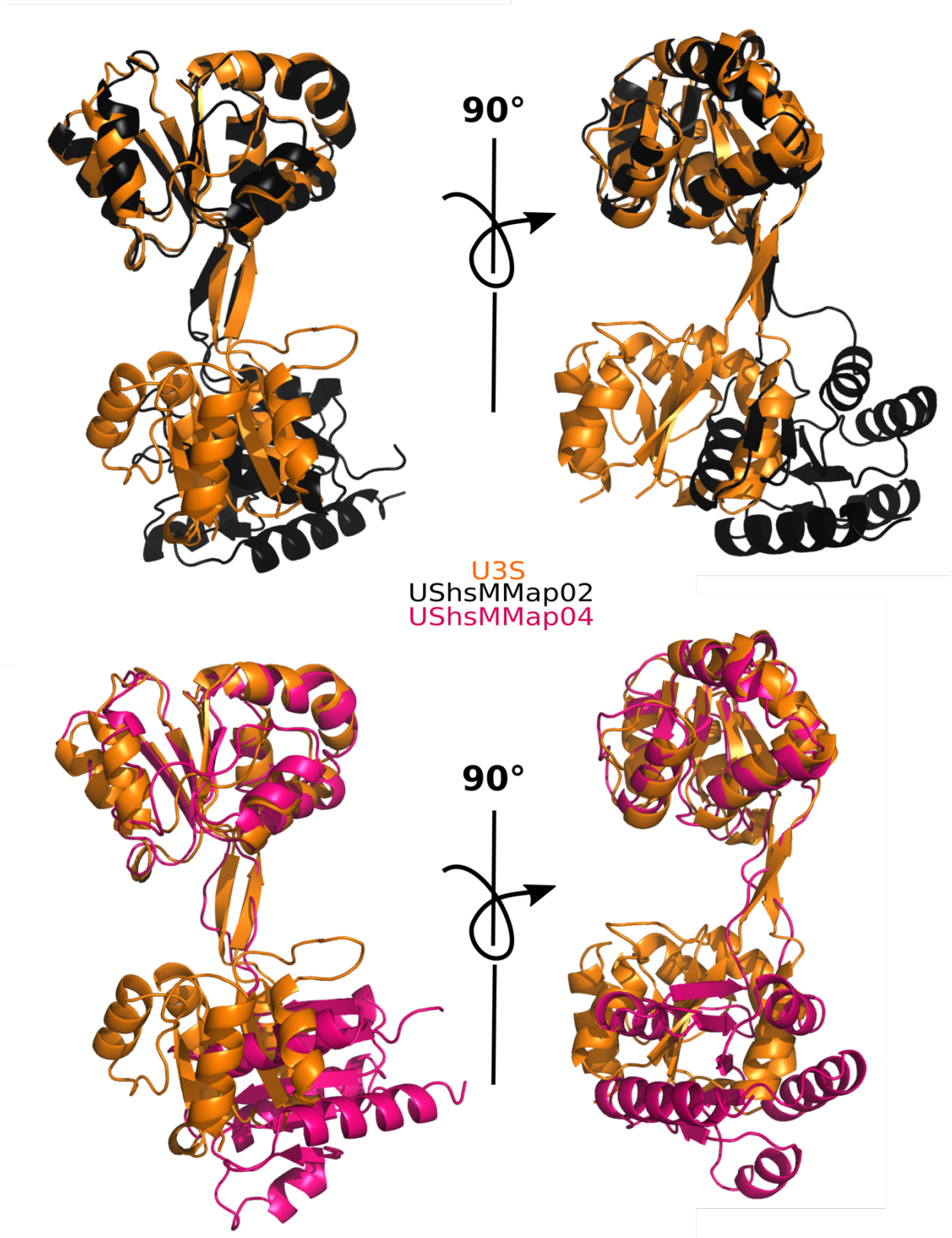



**Figure 3.15 Structural alignment of U3S parental structure vs. UShsMMap02 and UShsMMap04.** The crystal structures of UShsMMap02 (black) and 04 (pink) revealed a different relative orientation of their lobes, illustrating the high flexibility of their connecting loops.

### *3.8.4. Extension of chimeric helix through insertion of a helix cap*

The last attempt to extend the chimeric helix was through insertion of the helix-capping sequence QTPE. This sequence is highly conserved among the cobalamin-binding proteins. Therefore, we restored this sequence in the hybrid, yielding the UShsMMap06 variant, which in terms of its biophysical characterization corresponded to UShsMMap02. Attempts to solve the UShsMMap06 crystal structure failed, since the obtained crystals diffracted only at low resolution.

**Table 3.2 Crystallization conditions for chimeric variants**

| protein | buffer/salt | precipitant | salt /additives |
|---|---|---|---|
| UShsMMap02 | a) 0.2 M lithium chloride pH 7.4 | 10% (w/v) PEG 6000 | |
| | b) O.1 M HEPES | 15% (w/v) PEG 4000 | 10 mM spermine tetrahydrochloride |
| UShsMMap03 | a) 0.1 M MES pH 7.0 | 30% (v/v) Jeffamine ED 2001 | |
| | b) 0.1 M sodium acetate pH 4.6 | 25% (w/v) PEG 6000 | |
| UShsMMap04 | a) 0.1 M sodium HEPES pH 7.5 | 25% (w/v) PEG 3000 | |
| | b) 0.1 M BIS-TRIS pH 6.5 | 20% (w/v) PEG 5000 MME | |
| UShsMMap05 | a) 0.2 M potassium sodium tartrate | 20% (w/v) PEG 3035 | |
| | b) 0.1 M TRIS pH 8.0 | 1.5 M ammonium sulfate | |

## 3.9. Discussion

Pursuing the hypothesis that homologous fragments, shared by different folds, can be combined to create solubly-expressed hybrids, we designed a fold-chimera, that in addition, addresses the question whether a cofactor-binding function can be transplanted from one fold to a different one. The fact that sequence-based profiles are able to find reciprocal regions that appear to be reused multiple times in different architectures may be an indication that these units carry advantageous features coded in their sequences, from which we can profit to create artificial proteins. Thus, the recombination of these fragments constitutes an alternative and innovative tool that offers a good starting point in terms of solubility. These protein templates can be improved by rational design or directed evolution. For instance, the flexibility of UShsMMap02 can be advantageous to bind others cofactors. In fact, a UShsMMap02-derived variant showed that the incorporation of a cysteine residue, in cooperation with a native histidine residue (H17), are able to bind heme ("*Implementation and characterization of a heme-binding site in the chimera UShsMMap*" (2018) master's thesis by M.C. Hönle). This observation suggests that the independently-behaving lobes may associate to achieve binding, in a similar manner as the parental U3S traps its cognate ligand. Thus, despite donor and acceptor scaffolds may not be optimized for hybridization, they still yield well-behaving proteins, which is a good starting point for rational design. For instance, in this work the elimination of a clashing residue yielded significant structural improvements. Similarly, an earlier design attempting a chimeric TIM-barrel, showed the unanticipated formation of an additional beta strand (Bharat et al. 2008), an issue that was successfully solved with computational assistance (Eisenbeis et al. 2012). Overall, these unpredicted events teach us that computational algorithms must be employed to validate a protein model prior to its experimental characterization to improve the chances of success. Finally, we have proven that this approach not only yields proteins that can be expressed in *E. coli*, overcoming a major bottleneck of engineered proteins, but also provides the opportunity to combine existing functions that can be placed together economically and minimalistically for particular design purposes.

# Chapter 4

## Cobalamin-binding characterization of UShsMMap02

### 4.1. Overview

As described in the previous chapter, a cobalamin-binding protein was designed, testing the hypothesis that existing binding pockets can be transplanted to remote related proteins, retaining their functions in the resulting chimeras. The long-term goal of such chimeric design is the usage of natural cofactors in the future to assist complex reactions in synthetic enzymes. One example of cofactors that has awakened particular interest in organic synthesis are cobalamins, which can be employed to synthesize non-natural organic molecules in an environmentally friendly setup and non-hazardous reaction conditions. Only recently, and for the very first time, a cobalamin-dependent methyltransferase has been modified to perform the reversible methylation of organic molecules in mild conditions, showing that engineered cobalamin-dependent enzymes are a promising alternative to the currently available protocols for organic synthesis. In this chapter, I describe the binding characterization of the protein variant UShsMMap02 (**Section 3.8.1**). Hereby, I employed photometric, calorimetric, colorimetric and crystallization methods to confirm the successful transfer of the cofactor-binding function.

## 4.2. Cobalamins, ‘the most complex cofactors on earth’

The discovery of cobalamin (Cbl), also called vitamin B12 has its origins at the beginning of the 19th century. Back then, George H. Whipple, George R. Minot and William P. Murphy observed that cases of anemia pernicious, described already 100 years before, could be successfully treated with liver extract (Robscheit-Robbins and Whipple 1929; Richter 1933; Murphy 1932; Minot 1926). Based on this observation, this team of scientists predicted that liver may contain an ‘anti-anemia factor’ that prevented and healed this sickness. Thanks to this notable discovery, Whipple, Minot and Murphy were acknowledged with the Nobel Prize in Medicine and Physiology in 1934. About fifteen years later, Dorothy Hodgkin accomplished the elucidation of the vitamin molecule by X-ray crystallography (Hodgkin 1958), which in 1964 earned her the Nobel prize in Chemistry as well.

The cofactor crystal structure revealed an octahedral cobalt center coordinating four equatorial reduced pyrrole rings that build a corrin and seven amide chains attached to them. One of the macrocycle amides is bound to a nucleotide, whose moiety contains 5,6-dimethylbenzimidazole (DMB) as base (**Figure 4.1**). DMB coordinates the cobalt center from its α-axial (lower) face, whereas a cyano group coordinates the β-axial (upper) one.

It was not until the beginning of the 1970’s that scientists became aware that the cyano group attached to the cobalt was an artifact of its purification process and that the most biologically relevant molecules were methylcobalamin (MeCbl) and adenosylcobalamin (AdoCbl), which coordinate a methyl and an adenosyl group instead. Since then, several publications have described the chemical and biological properties of this molecule (Kenneth L. Brown 2005; Ruma Banerjee 1999; Matthews 2009), including the titanic labor of elucidating the required enzymes for its full biosynthesis in aerobic (Debussche et al. 1993) and anaerobic organisms (Debussche et al. 1993; Schulze, Vogler, and Renz 1998), projects that involved decades of work and the enormous effort of more than a hundred scientists.

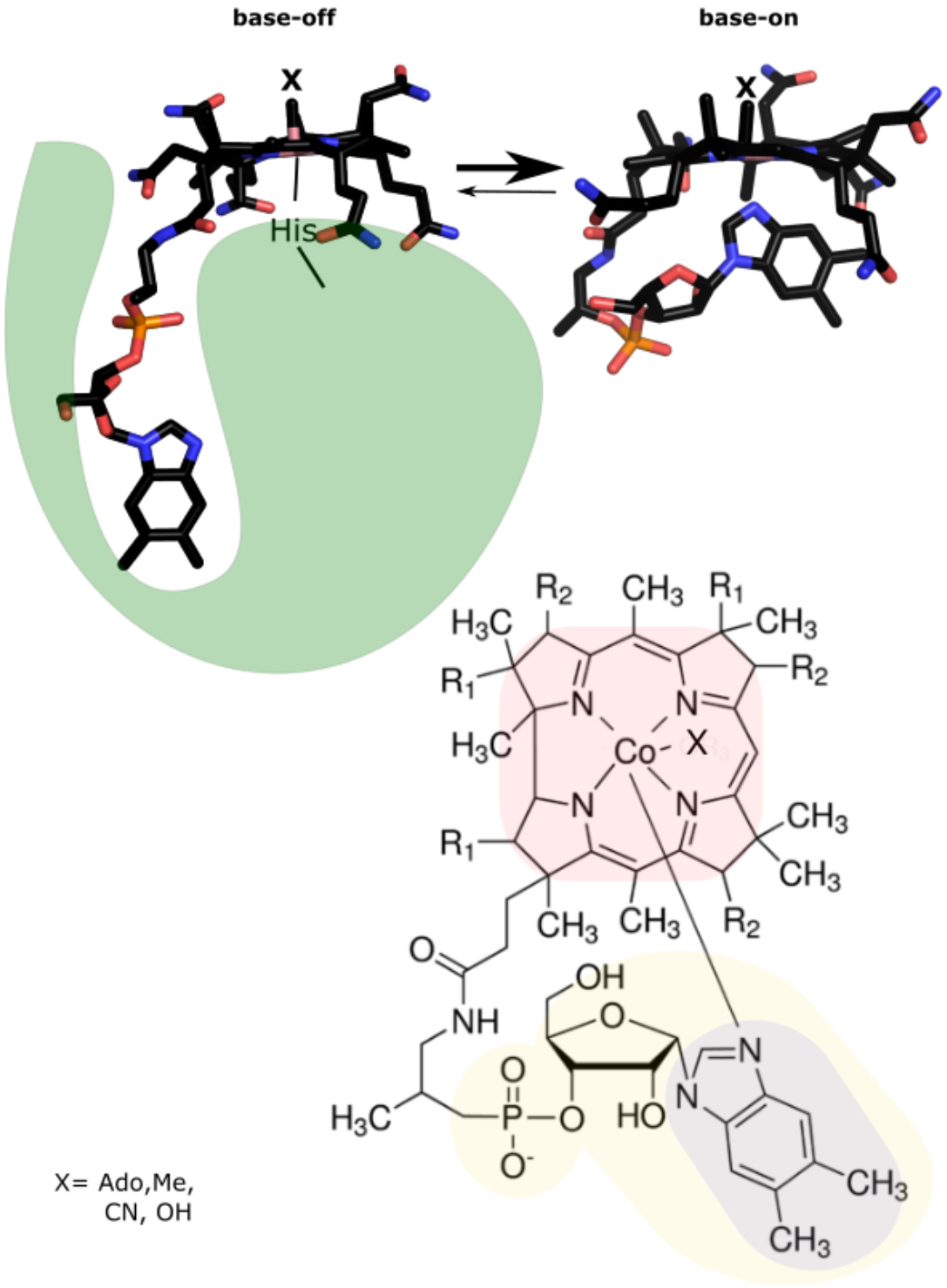



**Figure 4.1 Cobalamin cofactors and derived molecules.** Structurally, cobalamins are composed of a corrin ring (red), which is similar to the porphyrin ring of heme but contains cobalt as metal ion. This can be coordinated four to six times according to its redox state. In the hexa-coordinated form, the Co(III)balamin center displays four metallic bonds to the pyrrole rings, a β-axial coordination to a cyano (CN), methyl (Me), hydroxo (OH) or an adenosyl (Ado) group, and an α-axial coordination to the 5,6-dimethylbenzimidazole (DMB) intramolecular base. In the Co(II)balamin form, either the β or α coordination dissociates. Finally, Co(I)balamin presents no axial coordination. At physiological conditions, cobalamins display predominantly the so-called base-on conformation, in which DMB is coordinated to the cobalt. However, it is the base-off conformer that binds to cobalamin-dependent domains (**Figure 3.1**). By this process, the DMB-cobalamin coordination is displaced by a histidine residue to build the base-off his-on form.

## 4.3. The versatility of cobalamin-dependent catalysis

To date, three families of cobalamin-dependent enzymes have been discovered: the adenosylcobalamin-dependent isomerases (Mancia and Evans 1998), the methylcobalamin-dependent methyltransferases (Hilhorst et al. 1993; Ghosh et al. 2018) and dehalogenases (Payne et al. 2015). Here, only the most prominent examples will be described briefly, summarizing the chemical aspects of cobalamin that are commonly exploited for their enzymatic function.

### 4.3.1. Adenosylcobalamin-dependent isomerases

The largest subfamily of cobalamin-dependent enzymes are isomerases. They play an exclusive role in fermentation pathways in bacteria (R. Banerjee 1997, 2001; Marsh and Drennan 1976), except methylmalonyl-CoA mutase, which is also found in humans. Another function that employs vitamin B12 for isomerization is driven by the ribonucleotide reductase, which assists the conversion of ribonucleotides to deoxyribonucleotides, taking part in DNA replication and repair (Stubbe 1983). All these isomerases exploit the chemical properties of the organometallic linkage of Co-adenosyl in adenosylcobalamin (AdoCbl), which is stable in water and shows a dissociation energy of approximately 31.5 kcal/mol for the base-off form (Finke and Hay 1984; Hay and Finke 1987; K. L. Brown and Zou 1999). This confers an inherent lability of the Co-adenosyl bond, resulting in a latent radical pool. However, this radical source can be activated only in the presence of a substrate (K. L. Brown and Zou 1999). In fact, only when the substrate is bound, the reaction accelerates a trillion-fold, assisting the homolytic cleavage. Thus, the substrate-binding is the key energy control in all isomerase mechanisms that employ this cofactor.

### 4.3.2. Methylcobalamin-dependent methyltransferases

B12-dependent methyltransferases play an important role in one-carbon metabolism, carbon dioxide fixation and amino acid metabolism. All these reactions involve the transfer of a methyl donor to cobalamin and its further transfer to a methyl acceptor substrate. The most extensively studied B12-dependent methyltransferase constitutes the methionine synthase (MetH) from *E. Coli*, which catalyzes the methyl transfer from tetrahydrofolate to a Co(I)

form of Methylcobalamin yielding a CH3-Co(III) cobalamin intermediate from which the final methyl transfer to homocysteine is accomplished to synthesize methionine (Hall et al. 2001; Huennekens et al. 1974). These transitions are modulated by a highly conserved histidine-aspartate-serine triad, which is key in controlling the coordination state of cobalt (His-on/His-off) by modulating the protonation state of the histidine.

Another group of B12-dependent methyltransferases are O-demethylases, which are essential for acetogenic microbes. These enzymes catalyze the demethylation of aromatic methyl ethers. In this process, the ether oxygen remains in the phenolic product (DeWeerd et al. 1988), which is further excreted and the methyl group is further metabolized by the Wood-Ljungdahl pathway (Payne et al. 2015; Ragsdale 2008). Similar to MetH, O-demethylases employ a three-component system (Naidu and Ragsdale 2001), here however, the methyl group is donated by the phenylmethylether to cobalamin to form a CH3-Co(III) from which the methyl group is finally transferred to tetrahydrofolate (Naidu and Ragsdale 2001; Berman and Frazer 1992; el Kasmi, Rajasekharan, and Ragsdale 1994).

### 4.3.3. Cobalamin-dependent dehalogenases

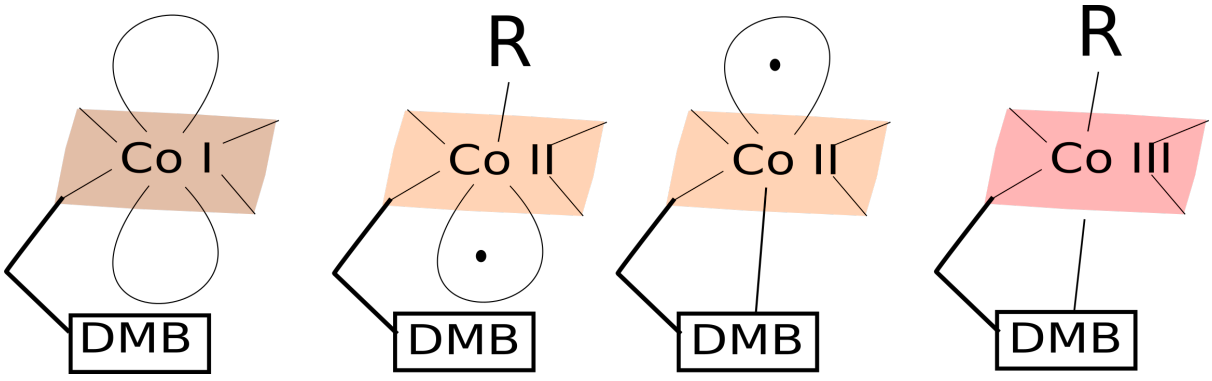
There are two main known classes of anaerobic dehalogenases: heme-containing (Ni, Fredrickson, and Xun 1995) and cobalamin-dependent enzymes (Christiansen et al. 1998; van de Pas et al. 1999). Reductive dehalogenation reactions play an important role in the detoxification of aliphatic and aromatic chlorinated molecules (Copley 1998; Janssen, Oppentocht, and Poelarends 2001). Cobalamin-dependent dehalogenases are poorly understood, however, it is known that the role of cobalamin in their reactions is significantly different from other cobalamin-dependent enzymes and two potential mechanism models have been proposed (Miller, Wohlfarth, and Diekert 1997; Neumann, Wohlfarth, and Diekert 1995; Krone, Thauer, and Hogenkamp 1989).

## 4.4. Employing cobalamin for artificial methylation

Synthetic reactions for methylation and demethylation are of great value. For instance, O-methylated phenols are indispensable building blocks for the synthesis of antioxidants, artificial flavors, dyes, fragrances, and agricultural chemicals (Law et al. 2016; M.-X. Zhang et al. 2015; Lee et al. 2017). Several reaction protocols are currently available for methylation and demethylation (Cody 2007). However, these reactions rely on toxic and corrosive reagents (Selva and Perosa 2008; W. E. Wymann et al. 1989; Walter E. Wymann et al. 1988) and are neither chemo nor regioselective. In addition, none of the currently available methods allow the reversible C-O-ether formation or demethylation in a large scale. Due to the high resistance of O-methyl ethers against hydrolysis, their cleavage usually requires the usage of strong Bronsted reactants (Fredriksson and Stone-Elander 2002; Hart, Aldous, and Harper 2017), Lewis acids (McOmie, Watts, and West 1968; Pasquini and Bassetti 2010) or metal catalysts (Cornella, Zarate, and Martin 2014; Yang et al. 2016). Only in the last decade, biocatalysts have gained attention as complementary or even as substitution to the conventional chemical approaches. They not only offer mild, sustainable and selective reactions (Patel 2011; Hönig et al. 2017; Sheldon and Woodley 2018), but also a broad repertoire of biosynthesis (Bornscheuer 2018; Turner and O'Reilly 2013; Chen et al. 2018), since reactions involving methylation are widely represented in Nature (Copeland, Solomon, and Richon 2009; Liscombe, Louie, and Noel 2012; Struck et al. 2012). Only recently, Farnberger and colleagues minimized the tetra-component system of a cobalamin-dependent methyltransferase to a bi-component biocatalyst. This system allows both methylation and demethylation in a reversible manner (Farnberger et al. 2018) and shows the utility of a methyl transfer concept based on corrinoid-assisted catalysis. The system is not only sustainable, safe and scalable but also not limited to natural substrates nor by cofactor turnover. We envision the cobalamin-dependent methylation as a promising add-on to organic synthesis. For instance, employing the fold-chimera approach described in **Chapter 3**, we may be able to place a desired binding pocket in close proximity to the cofactor, in that manner, assisting methylation of novel substrates by exploiting the highly nucleophilic nature of methylcobalamin.

## 4.5. Cobalamin- binding characterization of UShsMMap02

Under physiological conditions cobalamins exist in three different oxidation forms: Co(I), Co(II) and Co(III) (**Figure 4.2**). These forms are the basis of the chemical reactions assisted by this cofactor since the coordination of the axial ligands to the cobalt center depends on the formal oxidation state of the metal. As a rule, the number of ligands attached to cobalt decreases with the oxidation state of this metal. For instance, in cobalamin Co(III) is hexa coordinated, Co(II) penta coordinated, and Co(I) tetra coordinated, meaning the metal does not ligate beyond its coordinating corrin ring. Each oxidation state has a particular color. This physicochemical property is very useful, since the oxidation state can be validated simply by sight. In addition, the fact that cobalamins are natural chromophores can be exploited for binding analysis on proteins, since the protein-cofactor complex will display the characteristic color of the particular cobalamin form bound -a feature, which allows the application of UV-vis spectroscopy to define changes in the electronic environment of the corrin ring. Any change around the metal can be detected by this technique, such as the redox state but also the nature of a particular ligand attached to the cobalt.




**4.2 Cobalamin redox forms.** The cobalt center in cobalamins can adopt three oxidation states: (1) Co(I), a super nucleophile that has no α- or β-axial ligands and shows a characteristic brown color. (2) Co(II), which binds either an α- or a β-axial ligand, and displays an orange color. Finally, Co(III), which binds both, α- and β-axial ligands, displaying a red color. The α-axial coordination occurs intramolecularly to the molecule base, which is usually 5,6-dimethylbenzimidazole (DMB).

### 4.5.1. Analytical gel filtration in the presence of cobalamins

Taking advantage of the particular colors of cobalamins, we carried out an analytical gel filtration, using a sample of UShsMMap02 in the presence of four different cobalamins. Hereby, the individual cofactors were mixed in 20-fold excess with a pure protein sample and incubated for 30 minutes prior to loading. As positive control, the cobalamin-binding domain of methylmalonyl-CoA mutase from *Aeropyrum pernix* (parental protein) was also tested with each of the cobalamins under the same conditions. Both proteins eluates were red (indicative of cobalamin in its Co(III) redox state. This color remained even after dialysis against buffer without Cbl (**Figure 4.3-A**), indicating that Cbl does not easily dissociate from the hybrid upon dilution. In addition, concentration of the protein-Cbl complexes in a centrifugal filter (3 kDa) showed an increased red color in the concentrating eluates, whereas the flow through was transparent (**Figure 4.3-B**). Finally, the presence of Cbls in both protein eluates was confirmed by absorbance at 361 nm (**Figure 4.3-C-D**).

Despite the evident protein-cofactor formation, we were concerned that cobalamin could bind nonspecifically, for instance through displacement of the β-axial ligands by histidine or cysteine residues instead of proper insertion of its nucleotide moiety into the protein core, and ligation of the cis-face of the cobalamin with the putative native histidine residue (**figure 4.1**). Therefore, the cyanocobalamin (CNCbl) sample was of particular interest. Its strong Co-CN bond is highly stable compared to alkylcobalamins and does not dissociate under light exposure (Kozlowski et al. 2016). Thus, we co-refolded UShsMMap02 in presence of CNCbl and run an analytical gel filtration to get rid of the unbound cobalamin (**Figure 4.3-E**). The formation of a protein-CNCbl complex strongly suggests the proper cobalamin-binding, involving displacement of the DMB base, which can be corroborated photometrically as well (**Section 4.5.2**). To validate the stability of the protein-CNCbl complex, the eluted hybrid-CNCbl complex was loaded again onto an analytical gel filtration (without adding extra CNCbl). The second run showed that the majority of the cofactor remained bound to the protein (**Figure 4.3-F)**, suggesting that the binding affinity is at least in the low millimolar range.

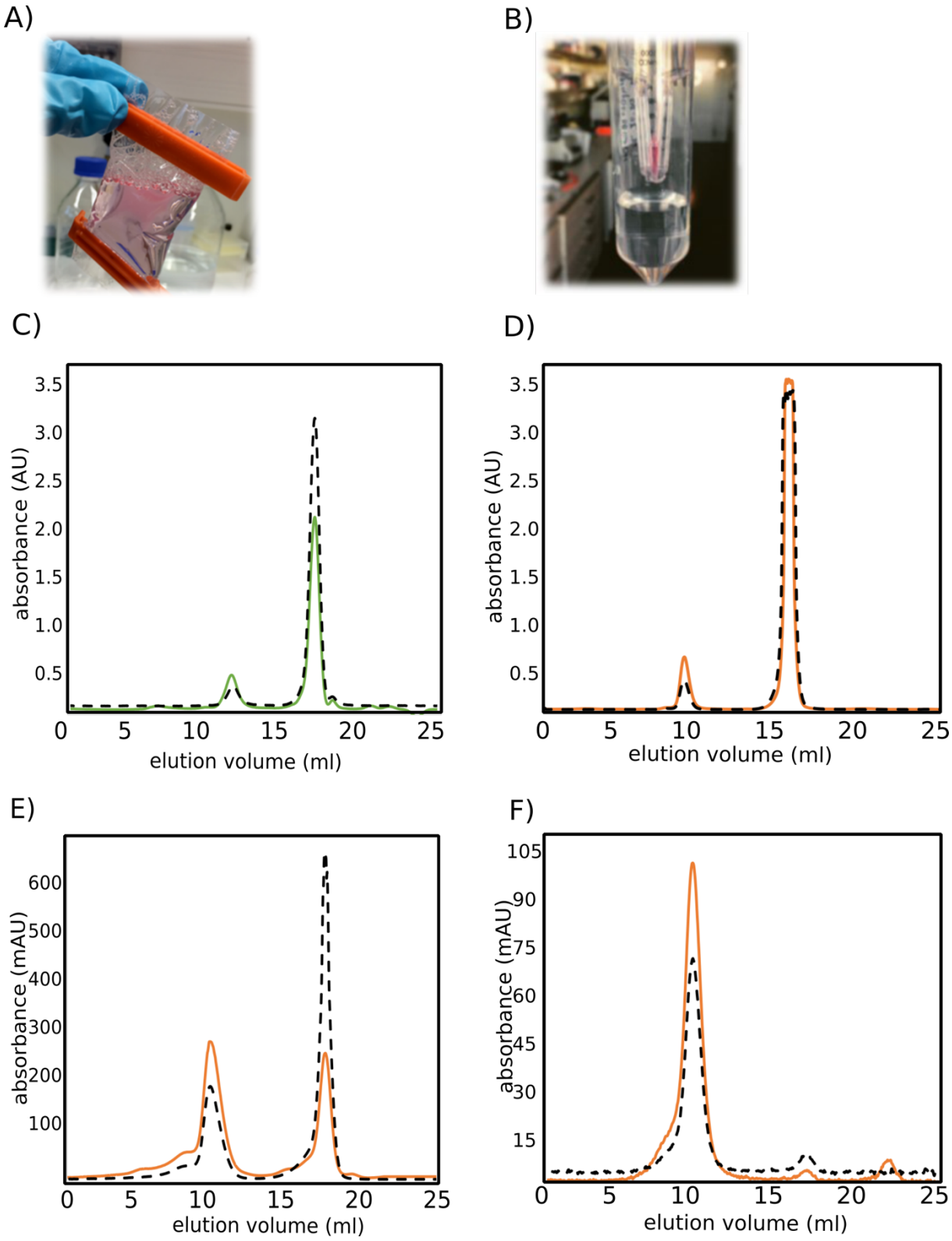



**Figure 4.3 Analytical gel filtration of MCoAM and UShsMMap02 in presence of cyanocobalamin.** The protein-cobalamin complex can be isolated via gel filtration as indicated by the red color of the sample after dialysis (**A**) or concentration (**B**). Analytical gel filtration of MCoAM-CNCbl (**C**) and UShsMMap02-CNCbl (**D**) samples. Analytical gel filtration of refolded UShsMMap02 in presence of CNCbl (**E**). A second run, employing the eluted hybrid-CNCbl sample showed poor CNCbl dissociation from the hybrid proteni (**F**). The protein-CNCbl samples were contained in 20 mM Tris at pH 7.4, 300 mM NaCl and 1 mM β-mercaptoethanol. CNCbl was tracked photometrically at 361 nm (black dashed line) and proteins at 280 nm (green and orange solid lines).

### 4.5.2. Photometrical analysis of cobalamins bound to the hybrid chimera

The UV-visible spectra of cobalamins are well described in the literature and provide a useful diagnostic for oxidation and ligation forms (Tsiminis et al. 2017; Dolphin 1971; Denis S. Salnikov et al. 2011; D. S. Salnikov and Makarov 2019; Dassanayake et al. 2016; Kim, Gherasim, and Banerjee 2008). Therefore, we performed UV-vis photometric studies on free and bound Cbls, resulting from the analytical gel filtration experiments described in the previous section. We found that adenosylcobalamin (AdoCbl) and methylcobalamin (MeCbl) lose their β-axial ligands upon binding to both, parental and hybrid proteins, leading to the typical spectrum of hydroxocobalamin (**Figure 4.4-A**). It is known that AdoCbl and MeCbl are less stable than CNCbl or OHCbl when exposed to light (Cole et al. 2002). However, the spectral changes observe here are not due to photolytic cleavage. As control, cobalamin samples without protein were incubated under the same conditions, showing no UV-vis spectral changes. In fact, the photolysis rates of base-off cobalamins are very different from their base-on analogs (Peng et al. 2010). The lack of the DMB coordination alters the electronic nature of the cofactor, opening a channel for fast non radiative decay, which is comparable to the Co-C bond photodissociation. This means that the cleavage of beta Co-coordinated ligands will dissociate faster from Cbl bound to the protein compared to free Cbl in solution. This is the reason why base-off Cbls are excellent agents for methylation.

As a parallel control, a titration with imidazole on free OHCbl was performed to test whether the observed photometric changes were due to imidazole bound to cobalamin as an artifact of the Ni-affinity chromatography instead of histidine ligation. In fact, the spectral changes were very similar (**Figure 4.4-B**). however, later binding studies with CNCbl corroborated that the cyano group remains on Cbl, speaking in favor of a proper insertion of the cofactor instead of β-axial coordination to imidazole.

Special attention has to be paid to the use of reducing agents such as β-mercaptoethanol, since this molecule reduces AdoCbl, MeCbl and OHCbl, and displaces the β-axial ligand (**Figure 4.4-C**). Thus, to avoid spectral changes owed to photo-lability and reduction, we proceeded in the photometric characterization exclusively with CNCbl. Only small spectral changes were observed in the presence of CNCbl (**Figure 4.4.-D**) compared to other CNCbl-binding proteins (**Figure 4.4-D**) (Kim, Gherasim, and Banerjee 2008) However, the spectrum of the eluted hybrid-CNCbl complex displayed the typical signal of CNCbl (**Figure 4.4-E to F**), indicating that binding to the protein likely occurs via insertion of its nucleotide moiety instead of replacement of the β-axial ligand, e.g. by residues such as histidine or cysteine.

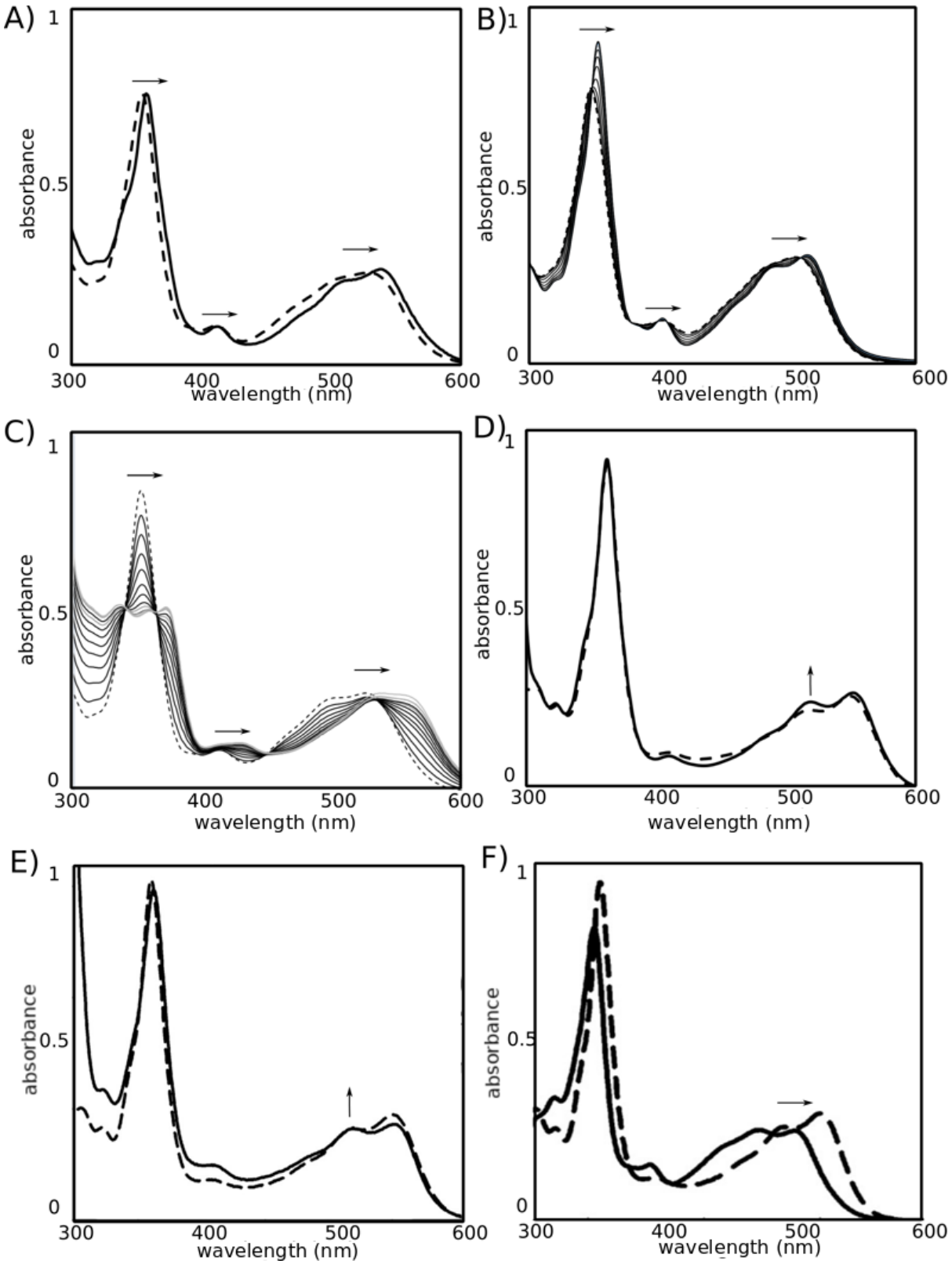



**Figure 4.4 UV-vis analysis of UShsMMap02 with cobalamins.** Spectral changes of free methylcobalamin (dashed) and bound to UShsMMap02 (solid line) after elution from gel filtration (**A**). Spectral changes of OHCbl (40 μM) upon addition of imidazole (**B**). Reductive titration of OHCbl (40 μM) with β-mercaptoethanol (**C**). Photometric changes upon binding of CNCbl (dashed line) to UShsMMap02 (solid line) (**D**). UV-vis absorbance spectra of free CNCbl (dashed line) and bound to MMACHC trafficking chaperone (**E**). Comparison between base-on (dashed line) and base-off (solid line) CNCbl in water and 6% HClO4, respectively (**F**). Captions (E) and (F) were adapted from "Decyanation of vitamin B12 by a trafficking chaperone" by Jihoe Kim, Carmen Gherasim, and Ruma Banerjee (2008).

### 4.5.3. Cobalamin binding analysis on UShsMMap02 in anaerobic conditions

In a parallel attempt to promote the base-off conformation of cobalamin and assist its binding to the hybrid protein, samples containing MeCbl, OHCbl and AdoCbl were reduced employing ascorbic acid. Reduction of Co(III)balamin to Co(II)balamin, in anaerobic conditions, results in a base-off form, which is stable in the absence of oxygen. Therefore, all buffers and samples employed were prepared in an anaerobic glove box under anoxic conditions. Hereby, the cofactor samples were added to the hybrid protein and incubated for 30 minutes. After this time, a bench gel filtration column was loaded with the protein-cofactor mixture to separate the unbound Cbl. Unexpectedly, the cobalamin bound to the protein had a different color than the free (reduced) cobalamin (**Figure 4.5**). It was anticipated that all Cbl molecules would adopt a base-off reduced conformation and, therefore, would display the characteristic orange color of Co(II)balamin. However, the protein-Cbl complex displayed the characteristic red color of Co(III)balamins, indicating that the cobalamin form attached to the protein is hexa-coordinated. How exactly the transition from Co(II) to Co(III) occurs, remains unclear. While the co(II) form of Cbl should be stable in anaerobic conditions, its association with the protein changes its chemical properties.

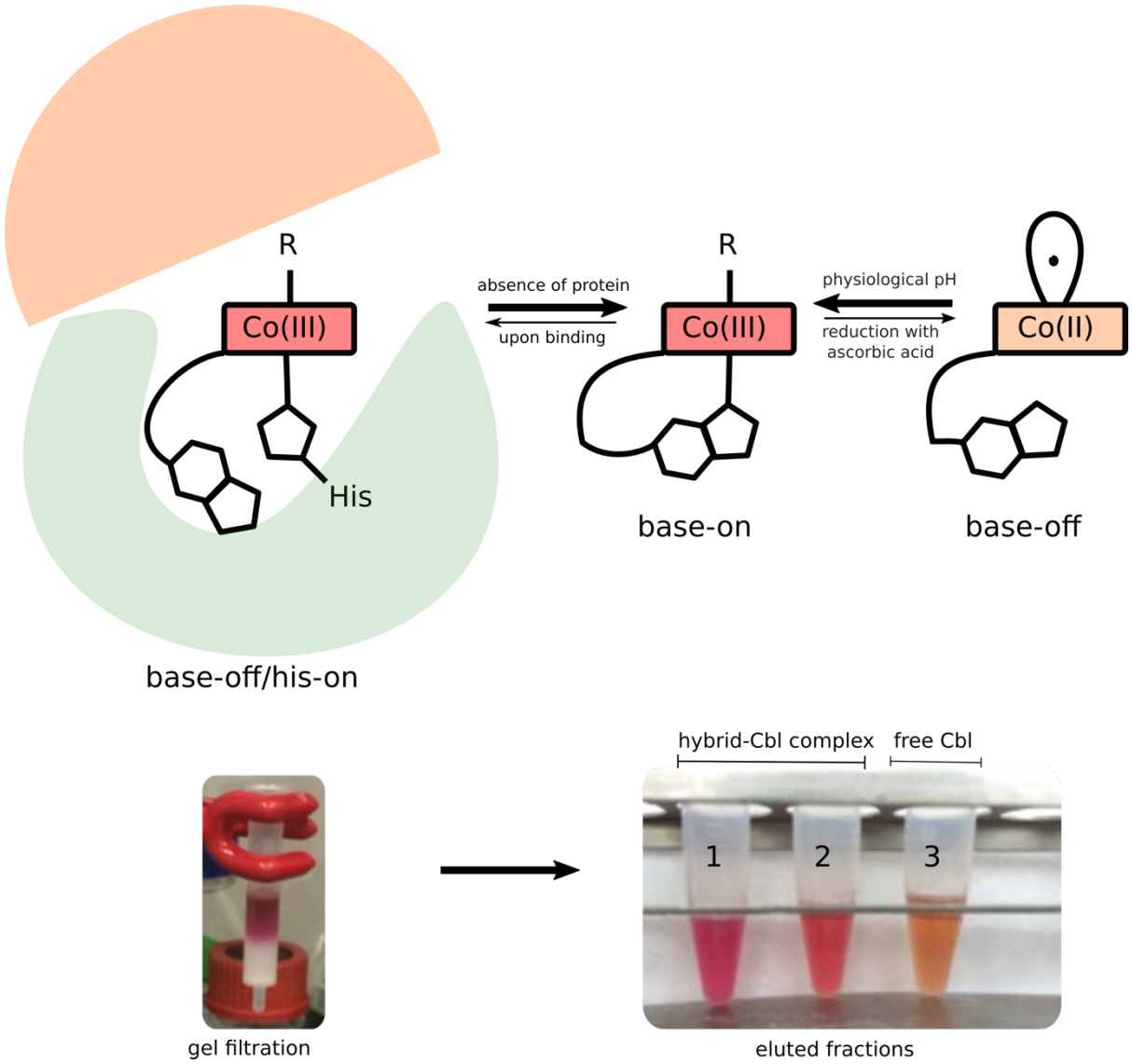



**Figure 4.5 Isolation of UShsMMap02-cobalamin complex under anaerobic conditions.** At physiological pH cobalamins can be converted to their base-off form via addition of reducing agents. For example, ascorbic acid reduces Co(III)balamins to their Co(II)balamin form (orange), which is stable in anaerobic conditions. Addition of the protein hybrid to a Co(II)balamin solution favours the transition to Co(III)balamin (red). The yielded protein-Co(III)balamin complex (fractions 1 and 2) can be isolated from free Co(II)balamin (fraction 3) via gel filtration as indicated by its red color, whereas the free Co(II)balamin displays the characteristic orange color of this form.

### 4.5.4. Isothermal titration calorimetry analysis of cyanocobalamin on fold-hybrid

To determine the binding affinity of cobalamin to the hybrid protein, a cyanocobalamin solution at 3 mM was titrated in steps of 2 µl to a protein solution (300 µl) of refolded UShsMMap02 at 282 µM (**Methods *6.19.3)*** Unexpectedly, the addition of cofactor did not show significant heat changes upon titrations (**Figure 4.6**). These observations do not correspond to the previous characterization methods via gel filtration and UV-vis spectrometry. However, it is evident that the protein-cofactor complex is formed since it can be isolated via size exclusion chromatography and remains even after a second gel filtration run. It remains unclear why no changes in protein-ligand thermodynamics cloud be detected; however, Cbl loading onto protein is complex. This process is often performed by chaperones in the cells and might take longer times, which may explain this observation.

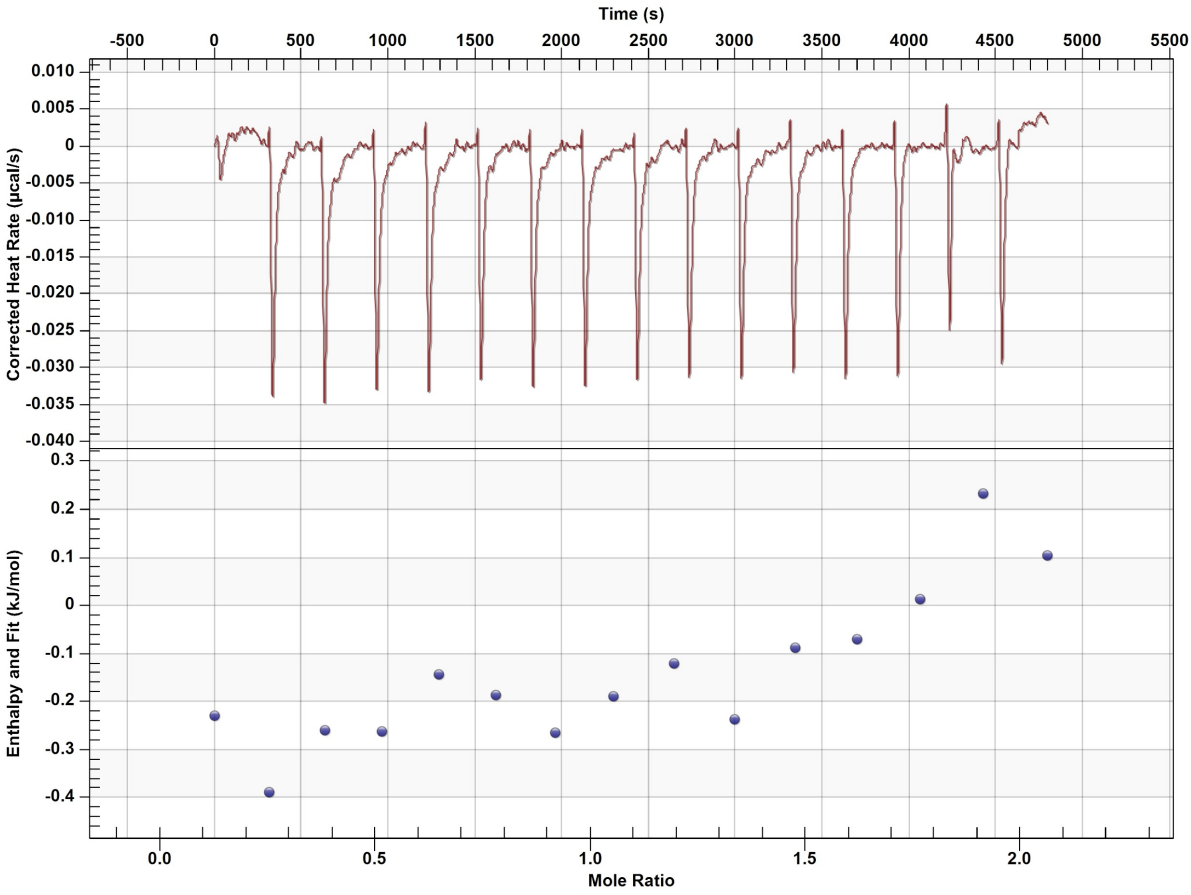



**Figure 4.6 Isothermal titration calorimetry measurements with cyanocobalamin.** A UShsMMap02 protein sample at 282 µM (300 µl) was titrated with 2 µl injections of a cyanocobalamin solution (3 mM) in 20 mM TRIS (pH 7.4), 100 mM NaCl, and 1 mM β-mercaptoethanol (**Methods 6.19.3**). No significant thermodynamic changes were detected upon addition of ligand. The measurements were carried out on a NanoITC isothermal titration calorimeter.

### 4.5.5. Crystallization in the presence of cobalamin and derived molecules

*Co-crystallization*

Crystal screenings were assessed for UShsMMap02 in the presence of six different vitamin B12 derivatives (**Table 4.1**.) First, four different cobalamins: methylcobalamin, hydroxocobalamin, adenosylcobalamin and cyanocobalamin were added (individually) in 20-fold excess to the protein sample at 12 mg/ml and incubated for at least 30 minutes. The samples were employed for co-crystallization trials via the hanging drop method as described in **Section 6.17.1**. None of the conditions yielded crystals, perhaps due to the large amount of cofactor added, considering that cobalamins are large molecules that may interfere with the crystallization process. To improve the chances of nucleation, samples of the four protein-Cbl complexes, previously isolated through size exclusion chromatography, were screened. Only one condition yielded one pink crystal (**Figure 4.7-B**). However, it diffracted at low resolution (up to 9Å) and improvement attempts (**Section 6.17.2**) did not succeed.

Since UShsMMap02 could not be crystallized in complex with any Cbl form, two derivatives were tried: (1) α-ribazole-3'-phosphate (αR3P) and (2) 5,6-dimethylbenzimidazole (DMB). The first constitutes the nucleotide moiety of cobalamins (**Figure 4.7-A**), which has been shown to bind to the cobalamin-binding domain of the glutamate mutase (Tollinger et al. 2001). Therefore, hanging drops experiments with this molecule were assessed. Since this molecule is not available for purchase, it was synthesized by hydrolysis of cyanocobalamin by my colleague Bruce Lichtenstein (**Methods 6.20**). In addition to αR3P, it was predicted that the DMB base may be able to bind as well and co-crystallization trays were also set in the presence of this ligand. Both experiments produced crystals (**Table 4.1**). However, crystals with αR3P were highly amorphous (**Figure 4.7-C**) and diffracted only at resolution above 4Å. In contrast, co-crystallization with DMB grew good quality crystals for structure determination. The crystals were handled and measured as described in **Section 6.17.3**. The best dataset was collected at 2.4Å and processed with the XDS package (Adams et al. 2010). Molecular replacement was performed with Phenix.phaser using pdb ids 2yxb and 1jr2 as separate assemble molecules. Final refinement (**Appendix 7.8**) resulted in R-work and R-free values at 25% and 26%, respectively. To corroborate the presence of the ligand, a difference 2Fo – Fc density map was calculated, showing density in the binding pocket region (**Figure 4.8-A**) at a sigma value of 1.0 (**Figure 4.8-B**). Finally, ligand-residue contacts were generated with LigPlot+ (Wallace et al. 1995) (**Figure 4.8-C to D**).

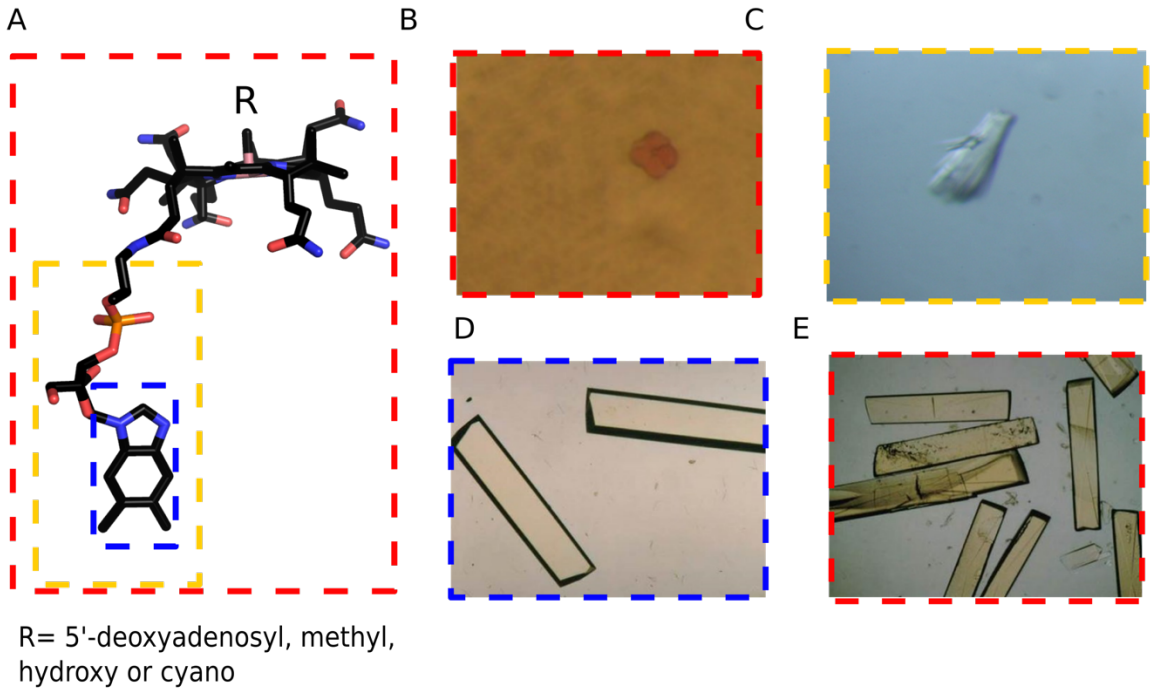



**Figure 4.7 Co-crystallization and soaking experiments.** Six cobalamin derivatives were employed for crystallization. Four cobalamins: adenosylcobalamin, methylcobalamin, hydroxocobalamin and cyanocobalamin; and two truncated molecules: alpha-ribazole-3'-phosphate (αR3P, yellow dashed) and 5,6-dimethylbenzimidazole (DMB, blue dashed) (**A**). Co-crystallization with cobalamins yielded only one amorphous crystal in the presence of OHCbl (**B**), which diffracted at low resolution (9Å). Co-crystallization with αR3P, also gave amorphous crystals, which diffracted up to 4Å (**C**). Apo crystals (**D**) were soaked with DMB, αR3P and Cbls. The latter led to disruption of the crystal lattices (**E**).

**Table 4.1 Crystallization attempts with cobalamin derivatives.**

| *Method* | *Ligand* | *Crystals* | *Successful condition* | *Res (Å)* | *Structure* | *Occupancy* |
|---|---|---|---|---|---|---|
| *Co-crystallization* | *MeCbl* | ✗ | *1 M Succinic acid, 0.1 M HEPES pH 7.0, 15 % (w/v) PEG MME 2000* | - | ✗ | |
| | *OHCbl* | ✓ | | - | ✗ | |
| | *AdoCbl* | ✗ | | - | ✗ | |
| | *CNCbl* | ✗ | | - | ✗ | |
| | *aR3P* | ✓ | *0.1 M HEPES, 20% PEG 400 pH 6.8* | *4.0* | ✗ | - |
| | *DMB* | ✓ | | *2.7* | ✓ | *100% (DMB)* |
| *Soaking* | *MeCbl* | ✗ | | - | ✗ | |
| | *OHCbl* | ✗ | | - | ✗ | |
| | *AdoCbl* | ✗ | | - | ✗ | |
| | *CHCbl* | ✗ | | - | ✗ | |
| | *aR3P* | ✓ | *0.1 M Tris, 15% PEG 4000 pH 5.7* | *2.2* | ✓ | *100% (PEG)* |
| | *DMB* | ✓ | | *2.4* | ✓ | *100% (PEG)* |

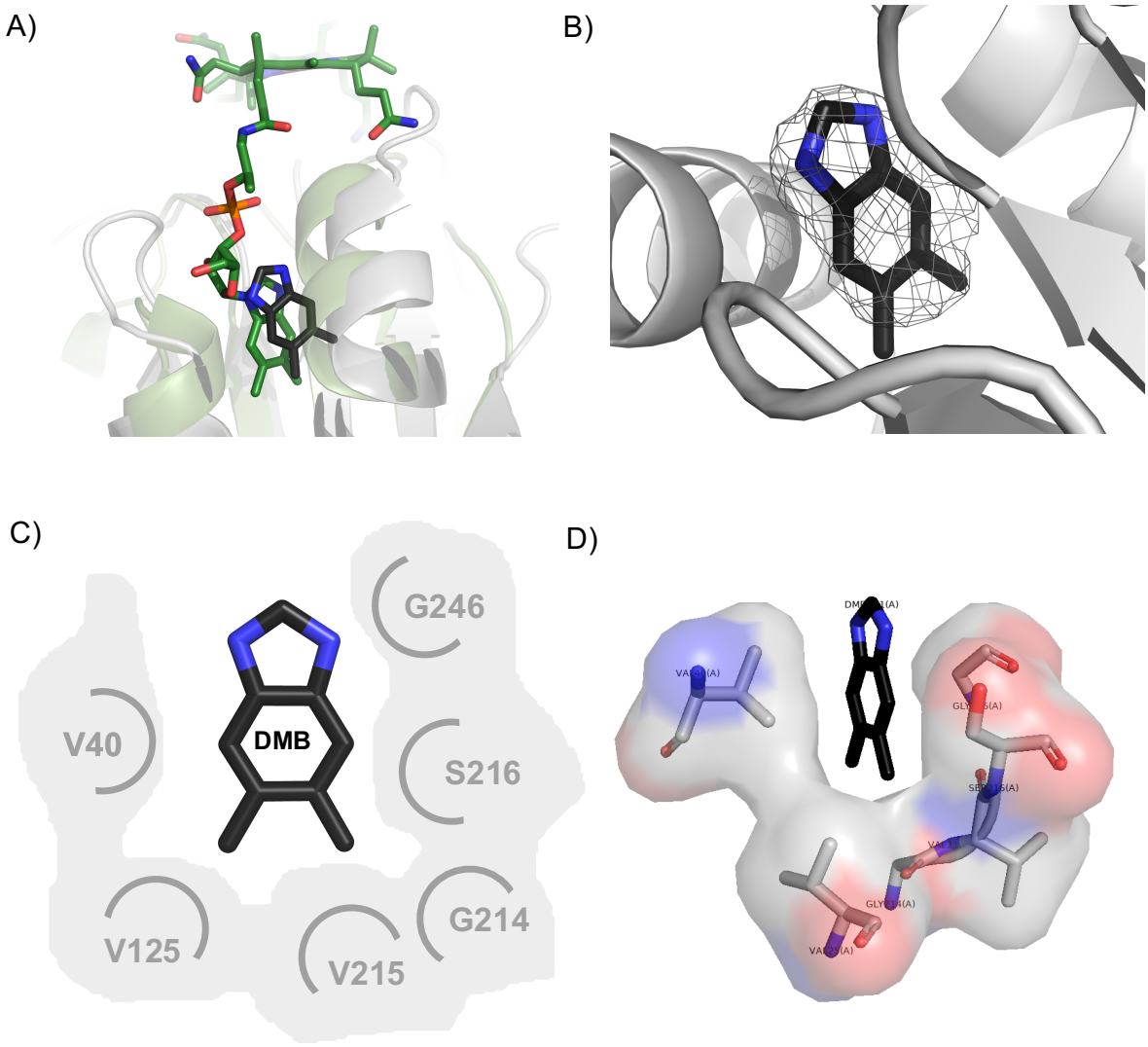



**Figure 4.8 Crystal structure of UShsMMap02 in complex with DMB.** The crystal structure of UShsMMap02 (gray) in complex with DMB (black) provides insights for cobalamin binding. This molecule binds to UShsMMap02 in the Cbl-binding area, as corroborated by the structural alignment to the cobalamin-binding domain of glutamate mutase from *E. coli* (green, pdb: 1bmt) (**A)** The difference 2Fo – Fc electron density map at 1,0σ shows electronic density in the expected binding region (**B**). Ligand-residue contacts were calculated with LigPlo+ (Wallace et al. 1995) **(C-D)**.

*Soaking*

The apo crystals showed a resolution of 2.0Å. Therefore, we soaked these with the same ligands mentioned in **Table 4.1.** The crystals soaked with cobalamins became damaged after few minutes of incubation (**Figure 4.7-E**), likely owed to the tight crystal packing in absence of ligand, which may become disrupted in the process of binding of the large ligand. In contrast, soaking with αR3P and DMB did not affected the crystal quality substantially, which remained intact after incubations of longer than 24 hours, as the apo crystals (**Figure 4.7-D**).

The obtained crystals were fished with X-ray mounting loops in adequate sizes and cryoprotected in 30% (v/v) PEG 400 dissolved in the crystal mother liquor employed for soaking. These crystals were finally cryocooled by plunging them into liquid nitrogen prior to data collection on the X10SA beamline at the Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung (BESSY). The data was collected at 100K and 0.5 oscillation degrees. The obtained images were recorded on a Pixel-detector Pilatus 2M at a distance of 270 nm and a wavelength of 1.000Å. Two datasets at 2.4 (DMB) and 2.2Å (αR3P) were collected. The obtained reflections were indexed, integrated and scaled with the XDS package (Kabsch 2010). Molecular replacement and refinement were performed with Phenix (Adams et al. 2010), as for UShsMMap02 and the final refinements are summarized in **Appendix 7.9 to 10**. Unexpectedly, the binding pocket in both structures was occupied by a PEG molecule instead of the desired molecules (**Figure 4.9**). This is likely due to the high concentration of PEG 400 (30%) added to the soaking conditions. We know from the co-crystallization attempts with DMB that this ligand successfully fills the Cbl binding pocket. Thus, improving the resolution of co-crystallized crystals with αR3P may yield a hybrid structure in complex with this ligand as well.

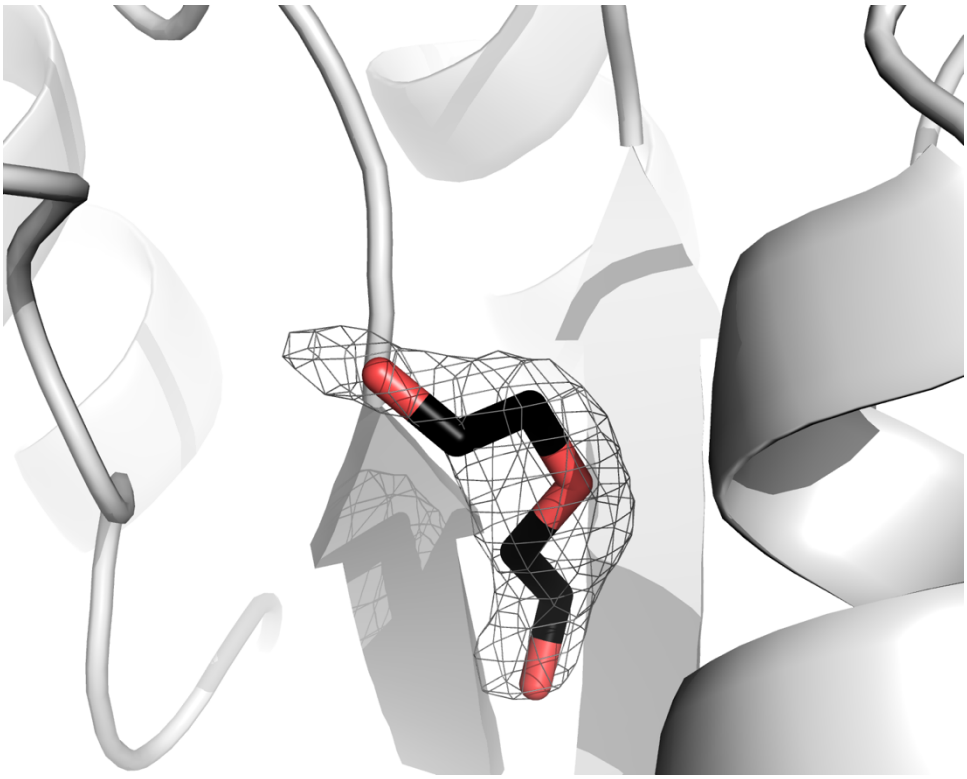


**Figure 4.9 UShsMMap02 crystal structure in complex with PEG.** Soaking attempts with αR3P and DMB failed to incorporate these ligands. The elucidation of both structures showed PEG in the cobalamin binding pocket instead of the soaked ligands. Difference 2Fo – Fc density electron density map observed at 1.0 σ.

## 4.6. Mutation of the cobalt-coordinating H17 in UShsMMap02 does not impair cobalamin binding

The first solved structure of cobalamin bound to a protein was the methionine synthase from *E. coli.* This structure revealed how the cofactor base is displaced by His579 to build its base-off/His-on conformation. Due to the high conservation of this residue among the cobalamin-binding units, we anticipated that this histidine (H17 in our construct), would be necessary for proper cobalamin binding. Thus, we generated a UShsMMap02-H17A mutant as control and performed its binding characterization by analytical gel filtration and UV-vis spectrophotometer. However, binding analysis by gel filtration showed that this variant still binds to cobalamin, showing a characteristic red coloration after elution. A deeper research in the literature revealed that previous mutagenesis analysis on the methionine synthase showed that mutation of any of the catalytic triad Ser-Asp-His has an impact on reaction rate but not on binding (Matthews R.G. (2009) "Cobalamin- and corrinoid-Dependent enzymes" Vol. 6: Roya Society of Chemistry). Thus, USHsMMap02-H17A is not a suitable negative control for binding.

## 4.7. Mimicking CarH activation mechanisms to limit UShsMMap02 flexibility

Only recently, the underlying principles of gene expression regulation by the cobalamin-dependent photoreceptor CarH were elucidated (Jost et al. 2015). This enzyme senses light with help of adenosylcobalamin (AdoCbl), whose Co-adenosyl bond is the basis for the regulatory function. For instance, in the presence of light, the Co-Ado bond dissociates and the cobalt center coordinates a histidine residue (**Figure 4.10-A**). This leads to a conformational change that induces dissociation of the CarH tetrameric assembly. Inspired by this sophisticated mechanism, I introduced five histidine mutations at the upper lobe interface of UShsMMap02 (**Figure 4.10-B to C**) to confer structural rigidity to the independently-behaving hybrid lobes upon Cbl binding. The mutations either replace existing residues (S84H and G121H) or are insertions located before residues S63, G141 and G165. These mutations may not only decrease the structural flexibility of UShsMMap02, but also facilitate double cobalt coordination (His-Co-His), which can be validated photometrically. However, the characterization of these constructs (**Section 5.4.3**) remains to be performed.

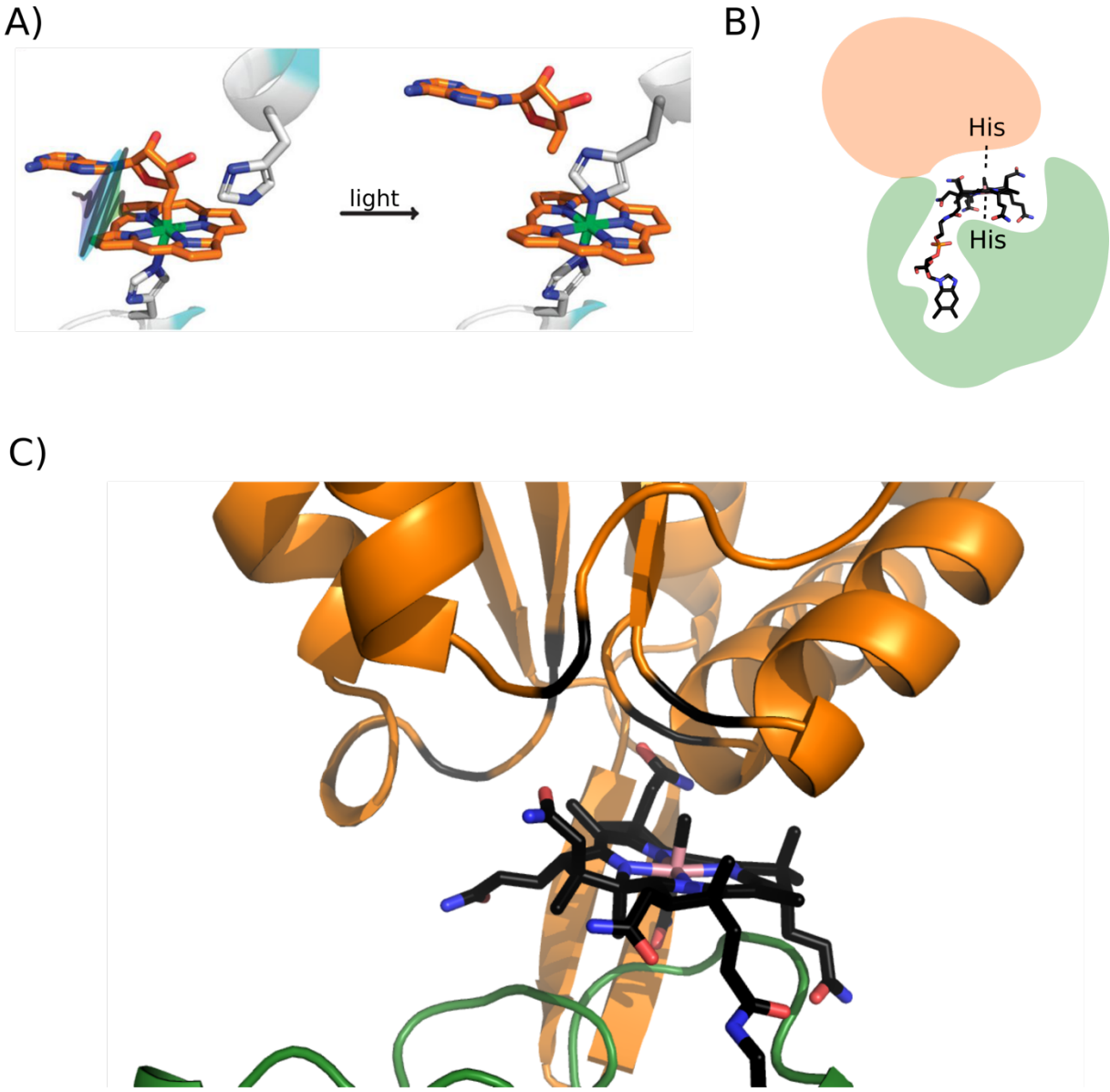



**Figure 4.10 CarH activation mechanism and its application to improve UShsMMap hybrids.** Activation of the adenosylcobalamin-dependent CarH photoreceptor is induced under light exposure, leading to dissociation of the adenosyl moiety from AdoCbl. By this process, its coordination to cobalt is replaced by a histidine residue (**A**). Five histidine mutations (black) (**Section 5.4.3**) were introduced at the upper UShsMMao02 interface to mimic the bis-his coordination of CarH (**B-C**). Caption (A) adapted from *"The photochemical mechanism of a B12-dependent photoreceptor protein" by* Jones A.R., Nature Communications, page 3 (2018).

## 4.8. Discussion

The goal of transplanting a cofactor-binding function from one fold into a different one was accomplished in this work. A hybrid-Cbl complex was successfully isolated via analytical gel filtration, which presents the red color characteristic of Co(III)balamins and was corroborated via UV-vis spectroscopy. Moreover, the hybrid-Cbl complex is relatively stable, since a second gel filtration run showed that only a small fraction of the cofactor dissociates from the designed protein. Despite these findings, the ITC experiments with CNCbl did not show any heat changes upon titration. However, the UV-vis spectra with the same ligand corroborated the cyano group remains axially bound to the cofactor, meaning that cobalamin-binding to the protein likely occurs via displacement of the DMB intramolecular interaction to Co. Supporting this hypothesis, the elucidation of the UShsMMap02 structure in presence of DMB showed binding to the protein in the expected B12-binding region.

To support the binding and structural evidence presented in this work, electron paramagnetic resonance (EPR) experiments can be performed, since EPR spectra are sensitive to the coordination state of Cbls and are helpful to differentiate base-off/His-on conformers from their base-on species. In addition, crystallization and ITC trials can be assessed with so-called pseudo cobalamins, which adopt exclusively a base-off conformation. Nevertheless, these compounds are not available for purchase and have to be synthesized.

Overall, these results illustrate that the recycling of prosthetic groups can be exploited for the design of proteins. Thus, the combinatorial protocol described in **Chapter 3**, may allow the reuse of naturally occurring cofactors, providing a wide range of reaction alternatives to be combined. Particularly the usage of large cofactors such as cobalamins, sulfur-iron clusters or heme are excellent catalysts that assist highly complex reactions. In fact, the later has been shown to bind to a UShsMMap-derived protein ("*Implementation and characterization of a heme-binding site in the chimera UShsMMap*" (2018) master's thesis by M.C. Hönle, University of Bayreuth), demonstrating that this approach is a good starting point for the design of novel functionalities in the future.

# Chapter 5

## Instruments and materials

### 4.1. Equipment

#### 5.1.1 Centrifuges

- Beckmann Coultier Avanti J-26xPI (Avanti)
- Biofuge table centrifuge 5425 (Eppendorf)
- Centrifuge 5810R (Eppendorf)

#### 5.1.2 Chromatography

- AEkta Pure 25M (GE healthcare)
- AEkta Pure P900 (GE healthcare)
- AEkta Pure Prime (GE healthcare)

#### 5.1.3 Crystallization

- Honeybee 963 crystallization robot (Genomic Solutions)
- Rock Imager 182 Imaging System with UV optics (Formulatrix)
- Mosquito crystallization robot (TTP)

#### 5.1.4 Isothermal titration calorimetry

- NanoITC isothermal titration calorimeter (TA Instruments)

#### 5.1.5 Scales

- Kern 572
- Kern Fine Scale ALS120-4

#### 5.1.6 Spectrophotometer

- Jasco J-810 CD-Spectrometer
- Jasco FP-6500 fluorescence-spectrometer
- Varian Cary 50 Scan UV-VIS spectrometer

5.1.7 Thermocycler

- Biometra T3000 Thermocycler
- BioRad MyCycler Thermocycler

5.1.8 Shakers

- IINOVA® 44 incubator (New Brunswick Scientific)
- Multitron II (Infors)

5.1.9 Miscellaneous

- Branson Digital sonifier W-250
- Hanna pH211 microprocessor pH meter
- Olaf Waase Thermoblock
- Anaerobic Glove Box Coy Lab (© MR)

## 5.2 Material

### 5.2.1 Bacterial strains

- Arctic Express (Stratagene)
- BL21 StarT M (Invitrogen)
- DH5α (Life Technologies)
- Top10 (Invitrogen)

### 5.2.2 Buffers and solutions

#### 5.2.2.1 Cloning

- TAE-Buffer (242g Tris, 57.1ml acetic acid, 37.2 EDTA in 1L)
- Agarose solution 1%: 1% agarose in TAE
- Agarose solution 2%: 2% agarose in TAE
- 1000× ampicillin stock solution: 100mg/ml Na-ampicillin in water
- BDT-mix: Obtained from in-house sequencing service
- Ethidium bromide stock solution: 10 mg/ml ethidium bromide in water
- 2x SDS sample buffer (100 mM TRIS pH 6.8, 20 % Glycerol, 4 % SDS, 200 mM)

#### 5.2.2.2 Protein expression

- 1000× IPTG stock solution: 1 M IPTG in water
- LB Medium For 1l: 10g pepton from bacteria, 5g NaCl, 5g yeast extract
- TB medium For 1l: 24g yeast extract, 12g tryptone, 4ml glycerol, 2.31g KH2PO4, 12.54g K2HPO4

#### 5.2.2.3 Protein purification

- **Buffer-A1** (20 mM Tris pH 7.4, 100 mM NaCl, 1 mM B-mcpto EtOH and 15 mM imidazole)
- **Buffer-B1** (20 mM Tris pH 7.4, 100 mM NaCl, 1 mM B-mcpto EtOH and 500 mM imidazole)

- **Buffer-A2** (50 mM HEPES pH 7.4, 100 mM NaCl, 1 mM B-mcpto EtOH and 15 mM imidazole)
- **Buffer-B2** (50 mM HEPES, 100 mM NaCl, 1 mM B-mcpto EtOH and 500 mM imidazole)
- **Buffer-A3 (**50 mM KP pH8, 150 mM KCl)
- **Buffer-B3 (**50 mM KP pH8, 150 mM KCl, 500 mM)

#### 5.2.2.4 Circular dichroism

- **CD-buffer** (10 mM phosphate buffer  pH 8)

#### 5.2.2.5 Isothermal titration calorimetry

- **ITC-buffer** (20mM Tris pH7.4, 100mM NaCl, 1mM β-mercaptoethanol)

#### 5.2.2.6 Refolding

- **R1-buffer** (50 mM Tris pH 7.4, 6M guanidinium chloride, 300mM sodium chloride)
- **R2-buffer** (50 mM Tris pH 7,5, 1 M guanidinium chloride, 300mM sodium chloride)
- **R3-buffer** (50 mM Tris pH 7,5, 2 M guanidinium chloride, 300mM sodium chloride)

#### 5.2.2.7 Gel filtration

- **GeFi-buffer-1** (20 mM Tris pH 7.4, 300mM sodium chloride, 1mM β-mercaptoethanol)
- **GeFi-buffer-2** (50 HEPES Tris pH 7.4, 300mM sodium chloride, 1mM β-mercaptoethanol)

5.2.3. Chemicals

**β-mercaptoethanol Brand**
Carl ROTH®
Formula C2H6OS
MW 78.13 g/mol
Purity 99%

**Cyanocobalamin (B12)**
Brand Sigma-Aldrich ®
Formula C63H88CoN14O14P
MW 1355.37 g/mol
Purity ≥98%

**Ethanol**
Brand Carl ROTH®
Formula C2H5OH
MW 46.07 g/mol
Purity 99.9%

**Guanidine hydrochloride**
Brand Carl ROTH®
Formula CH5N3 * HCL
MW 95.53 g/mol
Purity 99.5%

**Hydrochloric acid**
Brand Carl ROTH®
Formula HCL
WW 36.46 g/mol
Purity 37%

**Vitamin B12a hydrochloride**
Brand Sigma-Aldrich ®
Formula C62H89CoN13O15P · HCl
MW 1382.82 g/mol
Purity 96%

**Imidazol**
Brand Carl ROTH®
Formula C3H4N2
MW 68.08 g/mol
Purity 99%

**IPTG Brand Brand**
Sigma-Aldrich ®
Formula C9H18O5S
MW 238.30 g/mol
Purity 99%

**Methylcobalamin**
Brand Sigma-Aldrich ®
Formula
C63H91CoN13O14P
MW 1344.38 g/mol
Purity 97 %

**PEG 400**
Brand Carl ROTH®
Formula NA
MW 380-420 g/mol

**S-adenosylcobalamin**

Brand Sigma-Aldrich ®

Formula

C72H100CoN18O17P

MW 1579.58 g/mol

Purity ≥97%

**Sodium cacodylate**

Brand Sigma-Aldrich ®

Formula $(CH_3)_2AsO_2Na \cdot 3H_2O$

MW 214.03 g/mol

**Sodium chloride**

Brand Merck Millipore®

Formula NaCl

MW 58.44 g/mol

Purity NA

**Sodium hydroxide**

Brand Brand Carl ROTH®

Formula NaOH

MW 40.01 g/mol

Purity 44 – 46%

**Sodium L-ascorbate**

Brand Sigma-Aldrich ®

Formula C6H7NaO6

MW 198.11 g/mol

Purity ≥97%

5.2.4 Columns

- HisTrap HP column, 5ml Ni Sepharose (GE Healthcare)
- Superdex 75 10/300 GL (GE Healthcare)
- Superdex 200 10/300 GL (GE Healthcare)
- His GraviTrap™ (GE Healthcare)
- Sephadex G-25 in PD-10 Desalting Columns (GE Healthcare)
- illustra™ NAP™ Columns, NAP-5 (GE Healthcare)

5.2.5 Enzymes

- FastAP (Fermentas/Thermo Scientific)
- FD *NdeI* (Fermentas/Thermo Scientific)
- FD *XhoI* (Fermentas/Thermo Scientific)
- Q5 High fidelity DNA Polymerase (New England Biolabs)
- Taq polymerase (Fermentas/Thermo Scientific)

- T4 ligase (New England Biolabs)
- DNase I, RNase-free (Thermo Scientific)
- Alkaline phosphatase (Fermentas)

### 5.2.6 Kits

- Molecular weight calibration kit for gel filtration (GE Healthcare)
- Molecular weight calibration kit for SDS-PAGE (GE Healthcare)
- Miniprep kit (Quiagen)
- PCR-purification/Gel extraction kit (Quiagen)
- Protease inhibitors (SERVA)
- QIAGEN crystallization screenings
    - Classics I Suite
    - Classics II Suite
    - PEGs I Suite
    - PEGs II Suite
    - JCSG+ Suite
    - PACT Suite
    - ProComplex Suite

### 5.2.7 Miscellaneous

- Amicon Ultra-15ml and 50ml (Millipore)
- Dialysis membranes 3.5 kDa with glycerol (Spectra/Pore®)
- Gene ruler 50bp (Fermentas/Thermo Scientific)
- Gene ruler 100bp (Fermentas/Thermo Scientific)
- Gene ruler 1kbp (Fermentas/Thermo Scientific)
- Loading dye (6x) for DNA electrophoresis (Fermentas/Thermo Scientific)

## 5.3 Oligonucleotide sequences

### 5.3.1 Chapter 2

| Name | Sequence |
|---|---|
| OT10_U3Spa_Cter_fwd | ATATCGCATATGGATCCGAAAGTGCTGATCATGCGCG |
| OT10_U3Spa_Cter_rev | AATCTCGAGGGCGGCGCTCGT |
| OT10_U3Spa_Cter_FA_del_rev | CGGCCGGATAGTCGAGTGGCAGGTA |
| OT10_U3Spa_Cter_FB_del_fwd | TACCTGCCACTCGACTATCCGGCCG |

### 5.3.2 Chapter 3

| | |
|---|---|
| OT02_MCoAM_A_fwd | ATATCGCATATGCGTCGTCGTTATAAAGTT |
| OT02_MCoAM_A_rev | ATAAAACAGGGATCAAAACAACTTCAAAACCT |
| OT02_UShs_B_fwd | AGGTTTTGAAGTTGTTTTGATCCCTGTTTTA |
| OT02_UShs_B_rev | CCATTGCAACGTTCCCTTGGATT |
| OT02_MCoA_C_rev | AATCTCGAGTGCTTCTGCTTCTTCAC |
| OT02_MCoAM_C_fwd | CCAAGGGAACGTTGCAATGGCA |
| OT05_Q198A_UShsMMap01_fwd | ATCGCAGGGAACATAGTAAGAGCT |
| OT05_Q198A_UShsMMap01_rev | AGCTCTTACTATGTTCCCTGCGAT |
| OT12_UShsMMap_L225Y_fwd | CAT CTG CAT TAT ATG AAA CGT |
| OT12_UShsMMap_L225Y_rev | ACG TTT CAT ATA ATG CAG ATG |
| OT16_UShsMMap_hcap_fwd | CAGACACCGGAACAGGTTGCAATGGCAGCAGTTCAAGAG |
| OT16_UShsMMap_hcap_rev | CTGTTCCGGTGTCTGTGGGTGTGCAACTGTCTGATACAC |
| OT17_UShsMMap_G196A_fwd | GTTGCACACCCAGCAATCGCAGGGAACGTT |
| OT17_UShsMMap_G196A_rev | AACGTTCCCTGCGATTGCTGGGTGTGCAAC |
| OT18_UShsMMap_G196E_fwd | GTTGCACACCCAGAAATCGCAGGGAACGTT |
| OT18_UShsMMap_G196E_rev | AACGTTCCCTGCGATTTCTGGGTGTGCAAC |

### 5.3.1 Chapter 4

| Name | Sequence |
|---|---|
| OT15_UShsMMap_H17A_fwd | TCTGGATGGTGCTGATCGTGGT |
| OT15_UShsMMap_H17A_rev | ACCACGATCAGCACCATCCAGA |
| OT27_U3Shs_S84H_fwd | ATTTTTACCCACCCCAGAGCAGTG |
| OT27_U3Shs_i84H_fwd | ATTTTTACCCACAGCCCCAGAGCAGTG |
| OT27_U3Shs_rev | GAGTCCCCCGTAATCTTCAGG |
| OT28_U3Shs_G121H_fwd | TATGTGGTTCACAATGCTACTGCTT |
| OT28_U3Shs_G121H_rev | CACTGACTTGGCATTCCATTTTTCTTT |
| OT29_U3Shs_i141H_fwd | GAAACCTGTCACAATGCAGAAAAG |
| OT29_U3Shs_ rev | TCCTTCTGTATCCAGGCCAATTTT |
| OT30_U3Shs_i165H_fwd | CTATTTCCCTGTCATGGAAACCTCAAAAG |
| OT30_U3Shs_r | AAGAGGCAGTGCTGAGGA |

## 5.4 Protein sequences

### 5.4.1 Sequences corresponding to Chapter 2

| **U3S half variants for experimental characterization** | |
|---|---|
| cU3S | MDPKVLIMRGEGGREFLAERLRGQGVQVDYLPLYRRRAPDYPA<br>GELLARVRAERLNGLVVSSGQGLQNLYQLAAADWPEIGRLPLF<br>VPSPRVAEMARELGAQRVIDCRGASAPALLAALTSAALEHHHH<br>HH |
| cU3SΔ | MDPKVLIMRGEGGREFLAERLRGQGVQVDYLPL------DYPA<br>GELLARVRAERLNGLVVSSGQGLQNLYQLAAADWPEIGRLPLF<br>VPSPRVAEMARELGAQRVIDCRGASAPALLAALTSAALEHHHH<br>HH |
| **U3S full-length sequences** | |
| 1jr2 | MKVLLLKDAKEDDCGQDPYIRELGLYGLEATLIPVLSFEFLSL<br>PSFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKTEVWERS<br>LKEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAEKLAEYI<br>CSRESSALPLLFPCGNLKREILPKALKDKGIAMESITVYQTVA<br>HPGIQGNLNSYYSQQGVPASITFFSPSGLTYSLKHIQELSGDN<br>IDQIKFAAIGPTTARALAAQGLPVSCTAESPTPQALATGIRKA<br>LQ |
| 4es6 | MSGWRLLLTRPDEECAALAASLGEAGVHSSSLPLLAIDPLEET<br>PEQRTLMLDLDRYCAVVVVSKPAARLGLERLDRYWPQPPQQTW |

| | |
|---|---|
| | CSVGAATAAILEAYGLDVTYPEQGDDSEALLALPAFQDSLRVH<br>DPKVLIMRGEGGREFLAERLRGQGVQVDYLPLYRRRAPDYPAG<br>ELLARVRAERLNGLVVSSGQGLQNLYQLAAADWPEIGRLPLFV<br>PSPRVAEMARELGAQRVIDCRGASAPALLAALTSAA |
| 3re1 | MSAWRLLLTRPAEESAALARVLADAGIFSSSLPLLETEPLPLT<br>PAQRSIIFELLNYSAVIVVSKPAARLAIELIDEVWPQPPMQPW<br>FSVGSATGQILLDYGLDASWPEQGDDSEALLDHPRLKQAIAVP<br>GSRVLIMRGNEGRELLAEQLRERGVGVDYLPLYRRYLPQHAPG<br>TLLQRVEVERLNGLVVSSGQGFEHLLQLAGDSWPDLAGLPLFV<br>PSPRVASLAQAAGARNVIDCRGASAAALLAALRDQPQPAVKAY |
| 3d8r | MRIAYAGLRRKEEFKALAEKLGFTPLLFPVQATEKVPVPEYRD<br>QVRELAQGVDLFLATTGVGVRDLLEAGKALGLDLEGPLAKAFR<br>LARGAKAARALKEAGLPPHAVGDGTSKSLLPLLPQGRGVAALQ<br>LYGKPLPLLENALAERGYRVLPLMPYRHLPDPEGILRLEEAVL<br>RGEVDALAFVAAIQVEFLFEGAKDPKALREALNTRVKALAVGR<br>VTADALREWGVKPFYVDETERLGSLLQGFKRALQKEVA |
| 1wcw | MRRLEEDAVRVAYAGLRRKEAFKALAEKLGFTPLLFPVQATEK<br>VPVPEYRDQVRALAQGVDLFLATTGVGVRDLLEAGKALGLDLE<br>GPLAKAFRLARGAKAARALKEAGLPPHAVGDGTSKSLLPLLPQ<br>GRGVAALQLYGKPLPLLENALAERGYRVLPLMPYRHLPDPEGI<br>LRLEEALLRGEVDALAFVAAIQVEFLFEGAKDPKALREALNTR<br>VKALAVGRVTADALREWGVKPFYVDETERLGSLLQGFKRALQK<br>EVA |
| 3mw8 | MKLLLTRPEGKNAAMASALDALAIPYLVEPLLSVEAAAVTQAQ<br>LDELSRADILIFISTSAVSFATPWLKDQWPKATYYAVGDATAD<br>ALALQGITAERSPADSQATEGLLTLPSLEQVSGKQIVIVRGKG<br>GREAMADGLRLRGANVSYLEVYQRACPPLDAPASVSRWQSFGI<br>DTIVVTSGEVLENLINLVPKDSFAWLRDCHIIVPSARVETQAR<br>KKGLRRVTNAGAANQAAVLDALGM |
| **U3S N-terminal halves** | |
| 1jr2_nter | MKVLLLKDAKEDDCGQDPYIRELGLYGLEATLIPVLSFEFLSL<br>PSFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKTEVWERS<br>LKEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAEKLAEYI<br>CSRES |
| 4es6_nter | MSGWRLLLTRPDEECAALAASLGEAGVHSSSLPLLAIDPLEET<br>PEQRTLMLDLDRYCAVVVVSKPAARLGLERLDRYWPQPPQQTW<br>CSVGAATAAILEAYGLDVTYPEQGDDSEALLALPAFQDSLRV |
| 3re1_nter | MSAWRLLLTRPAEESAALARVLADAGIFSSSLPLLETEPLPLT<br>PAQRSIIFELLNYSAVIVVSKPAARLAIELIDEVWPQPPMQPW<br>FSVGSATGQILLDYGLDASWPEQGDDSEALLDHPRLKQAIAV |
| 1wcw_nter | MRRLEEDAVRVAYAGLRRKEAFKALAEKLGFTPLLFPVQATEK<br>VPVPEYRDQVRALAQGVDLFLATTGVGVRDLLEAGKALGLDLE<br>GPLAKAFRLARGAKAARALKEAGLPPHAVGDGTSKSLLPLLPQ<br>G |

| | |
|---|---|
| 3mw8_nter | MKLLLTRPEGKNAAMASALDALAIPYLVEPLLSVEAAAVTQAQ<br>LDELSRADILIFISTSAVSFATPWLKDQWPKATYYAVGDATAD<br>ALALQGITAERSPADSQATEGLLTLPSLEQVS |
| 3d8r_nter | MRIAYAGLRRKEEFKALAEKLGFTPLLFPVQATEKVPVPEYRD<br>QVRELAQGVDLFLATTGVGVRDLLEAGKALGLDLEGPLAKAFR<br>LARGAKAARALKEAGLPPHAVGDGTSKSLLPLLPQG |
| **U3S C-terminal halves** | |
| 1jr2_cter | SALPLLFPCGNLKREILPKALKDKGIAMESITVYQTVAHPGIQ<br>GNLNSYYSQQGVPASITFFSPSGLTYSLKHIQELSGDNIDQIK<br>FAAIGPTTARALAAQGLPVSCTAESPTPQALATGIRKALQ |
| 4es6_cter | HDPKVLIMRGEGGREFLAERLRGQGVQVDYLPLYRRRAPDYPA<br>GELLARVRAERLNGLVVSSGQGLQNLYQLAAADWPEIGRLPLF<br>VPSPRVAEMARELGAQRVIDCRGASAPALLAALTSAA |
| 3re1_cter | PGSRVLIMRGNEGRELLAEQLRERGVGVDYLPLYRRYLPQHAP<br>GTLLQRVEVERLNGLVVSSGQGFEHLLQLAGDSWPDLAGLPLF<br>VPSPRVASLAQAAGARNVIDCRGASAAALLAALRDQPQPAVKA<br>Y |
| 1wcw_cter | RGVAALQLYGKPLPLLENALAERGYRVLPLMPYRHLPDPEGIL<br>RLEEALLRGEVDALAFVAAIQVEFLFEGAKDPKALREALNTRV<br>KALAVGRVTADALREWGVKPFYVDETERLGSLLQGFKRALQKE<br>VA |
| 3mw8_cter | GKQIVIVRGKGGREAMADGLRLRGANVSYLEVYQRACPPLDAP<br>ASVSRWQSFGIDTIVVTSGEVLENLINLVPKDSFAWLRDCHII<br>VPSARVETQARKKGLRRVTNAGAANQAAVLDALGM |
| 3d8r_cter | RGVAALQLYGKPLPLLENALAERGYRVLPLMPYRHLPDPEGIL<br>RLEEAVLRGEVDALAFVAAIQVEFLFEGAKDPKALREALNTRV<br>KALAVGRVTADALREWGVKPFYVDETERLGSLLQGFKRALQKE<br>VA |
| **U3S permuted variants according to the protein lobes** | |
| 1jr2_L1 | MKVLLLKDAKEDDCGQDPYIRELGLYGLEATLIPVLSSITVYQ<br>TVAHPGIQGNLNSYYSQQGVPASITFFSPSGLTYSLKHIQELS<br>GDNIDQIKFAAIGPTTARALAAQGLPVSCTAESPTPQALATGI<br>RKALQ |
| 4es6_L1 | MSGWRLLLTRPDEECAALAASLGEAGVHSSSLPLLAIRRAPDY<br>PAGELLARVRAERLNGLVVSSGQGLQNLYQLAAADWPEIGRLP<br>LFVPSPRVAEMARELGAQRVIDCRGASAPALLAALTSAA |
| 3re1_L1 | MSAWRLLLTRPAEESAALARVLADAGIFSSSLPLLETRYLPQH<br>APGTLLQRVEVERLNGLVVSSGQGFEHLLQLAGDSWPDLAGLP<br>LFVPSPRVASLAQAAGARNVIDCRGASAAALLAALRDQPQPAV<br>KAY |
| 1wcw_L1 | MRRLEEDAVRVAYAGLRRKEAFKALAEKLGFTPLLFPVQATHL |

| | |
|---|---|
| | PDPEGILRLEEALLRGEVDALAFVAAIQVEFLFEGAKDPKALR<br>EALNTRVKALAVGRVTADALREWGVKPFYVDETERLGSLLQGF<br>KRALQKEVA |
| 3mw8_L1 | MKLLLTRPEGKNAAMASALDALAIPYLVEPLLSVRACPPLDAP<br>ASVSRWQSFGIDTIVVTSGEVLENLINLVPKDSFAWLRDCHII<br>VPSARVETQARKKGLRRVTNAGAANQAAVLDALGM |
| 3d8r_L1 | MRIAYAGLRRKEEFKALAEKLGFTPLLFPVQATEHLPDPEGIL<br>RLEEAVLRGEVDALAFVAAIQVEFLFEGAKDPKALREALNTRV<br>KALAVGRVTADALREWGVKPFYVDETERLGSLLQGFKRALQKE<br>VA |
| 1jr2_L2 | FEFLSLPSFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKT<br>EVWERSLKEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAE<br>KLAEYICSRESSALPLLFPCGNLKREILPKALKDKGIAMESIT<br>VYQT |
| 4es6_L2 | IDPLEETPEQRTLMLDLDRYCAVVVVSKPAARLGLERLDRYWP<br>QPPQQTWCSVGAATAAILEAYGLDVTYPEQGDDSEALLALPAF<br>QDSLRVHDPKVLIMRGEGGREFLAERLRGQGVQVDYLPLYR |
| 3re1_L2 | EPLPLTPAQRSIIFELLNYSAVIVVSKPAARLAIELIDEVWPQ<br>PPMQPWFSVGSATGQILLDYGLDASWPEQGDDSEALLDHPRLK<br>QAIAVPGSRVLIMRGNEGRELLAEQLRERGVGVDYLPLYRRYL<br>PQHA |
| 1wcw_L2 | TEKVPVPEYRDQVRALAQGVDLFLATTGVGVRDLLEAGKALGL<br>DLEGPLAKAFRLARGAKAARALKEAGLPPHAVGDGTSKSLLPL<br>LPQGRGVAALQLYGKPLPLLENALAERGYRVLPLMPYRH |
| 3mw8_L2 | EAAAVTQAQLDELSRADILIFISTSAVSFATPWLKDQWPKATY<br>YAVGDATADALALQGITAERSPADSQATEGLLTLPSLEQVSGK<br>QIVIVRGKGGREAMADGLRLRGANVSYLEVY |
| 3d8r_L2 | KVPVPEYRDQVRELAQGVDLFLATTGVGVRDLLEAGKALGLDL<br>EGPLAKAFRLARGAKAARALKEAGLPPHAVGDGTSKSLLPLLP<br>QGRGVAALQLYGKPLPLLENALAERGYRVLPLMPYR |
| **U3S permuted variants (αβα-swaps permuted)** | |
| 1jr2-Cαβα | SALPLLFPCGNLKREILPKALKDKGIAMESITVYQTFEFLSLP<br>SFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKTEVWERSL<br>KEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAEKLAEYIC<br>SRES |
| 1wcw-Cαβα | RGVAALQLYGKPLPLLENALAETEKVPVPEYRDQVRALAQGVD<br>LFLATTGVGVRDLLEAGKALGLDLEGPLAKAFRLARGAKAARA<br>LKEAGLPPHAVGDGTSKSLLPLLPQG |
| 3d8r-Cαβα | RGVAALQLYGKPLPLLENALAERGYRVLPLMPYRKVPVPEYRD<br>QVRELAQGVDLFLATTGVGVRDLLEAGKALGLDLEGPLAKAFR<br>LARGAKAARALKEAGLPPHAVGDGTSKSLLPLLPQG |

| | |
|---|---|
| 3mw8-Cαβα | GKQIVIVRGKGGREA?ADGLRLRGANVSYLEVYQRACPEAAAV<br>TQAQLDELSRADILIFISTSAVSFATPWLKDQWPKATYYAVGD<br>ATADALALQGITAERSPADSQATEGLLTLPSLEQVS |
| 3re1-Cαβα | PGSRVLIMRGNEGRELLAEQLRERGVGVDYLPLYRRYEPLPLT<br>PAQRSIIFELLNYSAVIVVSKPAARLAIELIDEVWPQPPMQPW<br>FSVGSATGQILLDYGLDASWPEQGDDSEALLDHPRLKQAIAV |
| 4es6-Cαβα | HDPKVLIMRGEGGREFLAERLRGQGVQVDYLPLYRIDPLEETP<br>EQRTLMLDLDRYCAVVVVSKPAARLGLERLDRYWPQPPQQTWC<br>SVGAATAAILEAYGLDVTYPEQGDDSEALLALPAFQDSLRV |

### 5.4.2 Sequences corresponding to Chapter 3

| | |
|---|---|
| MCoAMap | RRRYKVLVAKMGLDGHDRGAKVVARALRDAGFEVVYTGLRQTP<br>EQVAMAAVQEDVDVIGVSILNGAHLHLMKRLMAKLRELGADDI<br>PVVLGGTIPIPDLEPLRSLGIREIFLPGTSLGEIIEKVRKLAE<br>EKRMREEAEAHHHHHH |
| UShsMMap01 | MRRRYKVLVAKMGLDGHDRGAKVVARALRDAGFEVVLIPVLSF<br>EFLSLPSFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKTE<br>VWERSLKEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAEK<br>LAEYICSRESSALPLLFPCGNLKREILPKALKDKGIAMESITV<br>YQTVAHPGIQGNVAMAAVQEDVDVIGVSILNGAHLHLMKRLMA<br>KLRELGADDIPVVLGGTIPIPDLEPLRSLSIREIFLPGTSLGE<br>IIEKVRKLAEEKRMREEAEALEHHHHHH |
| UShsMMap02<br>(UShsMMap01_Q198A) | MRRRYKVLVAKMGLDGHDRGAKVVARALRDAGFEVVLIPVLSF<br>EFLSLPSFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKTE<br>VWERSLKEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAEK<br>LAEYICSRESSALPLLFPCGNLKREILPKALKDKGIAMESITV<br>YQTVAHPGI**A**GNVAMAAVQEDVDVIGVSILNGAHLHLMKRLMA<br>KLRELGADDIPVVLGGTIPIPDLEPLRSLGIREIFLPGTSLGE<br>IIEKVRKLAEEKRMREEAEALEHHHHHH |
| UShsMMap03<br>(UShsMMap02_L225Y) | MRRRYKVLVAKMGLDGHDRGAKVVARALRDAGFEVVLIPVLSF<br>EFLSLPSFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKTE<br>VWERSLKEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAEK<br>LAEYICSRESSALPLLFPCGNLKREILPKALKDKGIAMESITV<br>YQTVAHPGI**A**GNVAMAAVQEDVDVIGVSILNGAHLH**L**MKRLMA<br>KLRELGADDIPVVLGGTIPIPDLEPLRSLGIREIFLPGTSLGE<br>IIEKVRKLAEEKRMREEAEALEHHHHHH |
| UShsMMap04<br>(UShsMMap02_G196A) | MRRRYKVLVAKMGLDGHDRGAKVVARALRDAGFEVVLIPVLSF<br>EFLSLPSFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKTE<br>VWERSLKEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAEK<br>LAEYICSRESSALPLLFPCGNLKREILPKALKDKAIAMESITV<br>YQTVAHP**A**I**A**GNVAMAAVQEDVDVIGVSILNGAHLHLMKRLMA<br>KLRELGADDIPVVLGGTIPIPDLEPLRSLGIREIFLPGTSLGE<br>IIEKVRKLAEEKRMREEAEALEHHHHHH |
| UShsMMap05 | MRRRYKVLVAKMGLDGHDRGAKVVARALRDAGFEVVLIPVLSF |

| | |
|---|---|
| (UShsMMap02_G196E) | EFLSLPSFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKTE<br>VWERSLKEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAEK<br>LAEYICSRESSALPLLFPCGNLKREILPKALKDKAIAMESITV<br>YQTVAHP**E**I**A**GNVAMAAVQEDVDVIGVSILNGAHLHLMKRLMA<br>KLRELGADDIPVVLGGTIPIPDLEPLRSLGIREIFLPGTSLGE<br>IIEKVRKLAEEKRMREEAEALEHHHHHH |
| UShsMMap06<br>(UShsMMap02_HCap) | MRRRYKVLVAKMGLDGHDRGAKVVARALRDAGFEVVLIPVLSF<br>EFLSLPSFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKTE<br>VWERSLKEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAEK<br>LAEYICSRESSALPLLFPCGNLKREILPKALKDKGIAMESITV<br>YQTVAHP**TPEQ**VAMANAVQEDVDVIGVSILNGAHLHLMKRLMA<br>KLRELGADDIPVVLGGTIPIPDLEPLRSLGIREIFLPGTSLGE<br>IIEKVRKLAEEKRMREEAEALEHHHHHH |

5.4.3 Sequences corresponding to Chapter 4

| | |
|---|---|
| UShsMMap02_H17A | MRRRYKVLVAKMGLDG**A**DRGAKVVARALRDAGFEVVLIPVLSFE<br>FLSLPSFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKTEVW<br>ERSLKEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAEKLAE<br>YICSRESSALPLLFPCGNLKREILPKALKDKGIAMESITVYQTV<br>AHPGIAGNVAMAAVQEDVDVIGVSILNGAHLHLMKRLMAKLREL<br>GADDIPVVLGGTIPIPDLEPLRSLGIREIFLPGTSLGEIIEKVR<br>KLAEEKRMREEAEALEHHHHHH |
| UShsMMap02_G121H | MRRRYKVLVAKMGLDGHDRGAKVVARALRDAGFEVVLIPVLSFE<br>FLSLPSFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKTEVW<br>ERSLKEKWNAKSVYVV**H**NATASLVSKIGLDTEGETCGNAEKLAE<br>YICSRESSALPLLFPCGNLKREILPKALKDKGIAMESITVYQTV<br>AHPGIAGNVAMAAVQEDVDVIGVSILNGAHLHLMKRLMAKLREL<br>GADDIPVVLGGTIPIPDLEPLRSLGIREIFLPGTSLGEIIEKVR<br>KLAEEKRMREEAEALEHHHHHH |
| UShsMMap02_S84H | MRRRYKVLVAKMGLDGHDRGAKVVARALRDAGFEVVLIPVLSFE<br>FLSLPSFSEKLSHPEDYGGLIFT**H**PRAVEAAELCLEQNNKTEVW<br>ERSLKEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAEKLAE<br>YICSRESSALPLLFPCGNLKREILPKALKDKGIAMESITVYQTV<br>AHPGIAGNVAMAAVQEDVDVIGVSILNGAHLHLMKRLMAKLREL<br>GADDIPVVLGGTIPIPDLEPLRSLGIREIFLPGTSLGEIIEKVR<br>KLAEEKRMREEAEALEHHHHHH |
| UShsMMap02_i141H<br>(insertion before<br>G141) | MRRRYKVLVAKMGLDGHDRGAKVVARALRDAGFEVVLIPVLSFE<br>FLSLPSFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKTEVW<br>ERSLKEKWNAKSVYVVGNATASLVSKIGLDTEGETC**H**GNAEKLA<br>EYICSRESSALPLLFPCGNLKREILPKALKDKGIAMESITVYQT<br>VAHPGIAGNVAMAAVQEDVDVIGVSILNGAHLHLMKRLMAKLRE<br>LGADDIPVVLGGTIPIPDLEPLRSLGIREIFLPGTSLGEIIEKV<br>RKLAEEKRMREEAEALEHHHHHH |
| UShsMMap02_i165H<br>(insertion before<br>G165) | MRRRYKVLVAKMGLDGHDRGAKVVARALRDAGFEVVLIPVLSFE<br>FLSLPSFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKTEVW<br>ERSLKEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAEKLAE<br>YICSRESSALPLLFPC**H**GNLKREILPKALKDKGIAMESITVYQT |

| | |
|---|---|
| | VAHPGIAGNVAMAAVQEDVDVIGVSILNGAHLHLMKRLMAKLRE<br>LGADDIPVVLGGTIPIPDLEPLRSLGIREIFLPGTSLGEIIEKV<br>RKLAEEKRMREEAEALEHHHHHH |
| UShsMMap02_i84H<br>(ins before S63) | MRRRYKVLVAKMGLDGHDRGAKVVARALRDAGFEVVLIPVLSFE<br>FLSLPSFSEKLSHPEDYGGLIFT**H**SPRAVEAAELCLEQNNKTEV<br>WERSLKEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAEKLA<br>EYICSRESSALPLLFPCGNLKREILPKALKDKGIAMESITVYQT<br>VAHPGIAGNVAMAAVQEDVDVIGVSILNGAHLHLMKRLMAKLRE<br>LGADDIPVVLGGTIPIPDLEPLRSLGIREIFLPGTSLGEIIEKV<br>RKLAEEKRMREEAEALEHHHHHH |
| UShsMMap02_H17C | MRRRYKVLVAKMGLDG**C**DRGAKVVARALRDAGFEVVLIPVLSFE<br>FLSLPSFSEKLSHPEDYGGLIFTSPRAVEAAELCLEQNNKTEVW<br>ERSLKEKWNAKSVYVVGNATASLVSKIGLDTEGETCGNAEKLAE<br>YICSRESSALPLLFPCGNLKREILPKALKDKGIAMESITVYQTV<br>AHPGIAGNVAMAAVQEDVDVIGVSILNGAHLHLMKRLMAKLREL<br>GADDIPVVLGGTIPIPDLEPLRSLGIREIFLPGTSLGEIIEKVR<br>KLAEEKRMREEAEALEHHHHHH |

## 5.5 Vectors

### 5.5.1 13AN6UP_MMcM_pMA

This vector codes for the cobalamin-binding subunit of methylmalonyl-CoA mutase (MCoAMap) (**Section 2.9.1**). The gene was synthesized by Life Technologies and cloned into a pMA vector, employing *the Kpn*I and *SacI* cloning sites. The plasmid bears ampicillin resistance.

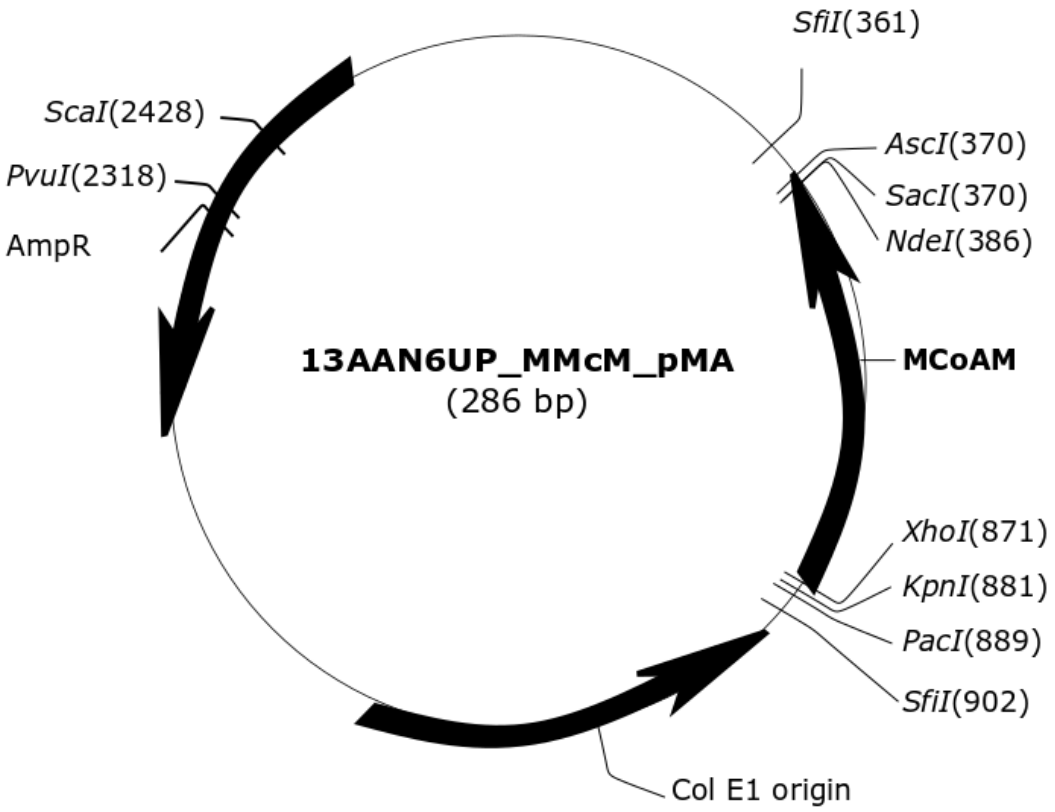



**Figure 5.1 Plasmid DNA 13AAN6UP_MMcM_pMA.** This vector contains the methylmalonyl-CoA mutase gene from *Aeropyrum pernix*.

### 5.5.2 pET21a(+) Novagen® ampicillin resistance

pET-21(+) is a transcription vector designed for expression from bacterial translation signals carried within a cloned insert. Generated genes in this work were clones into this vector employing *NdeI* and *XhoI* restriction sites.

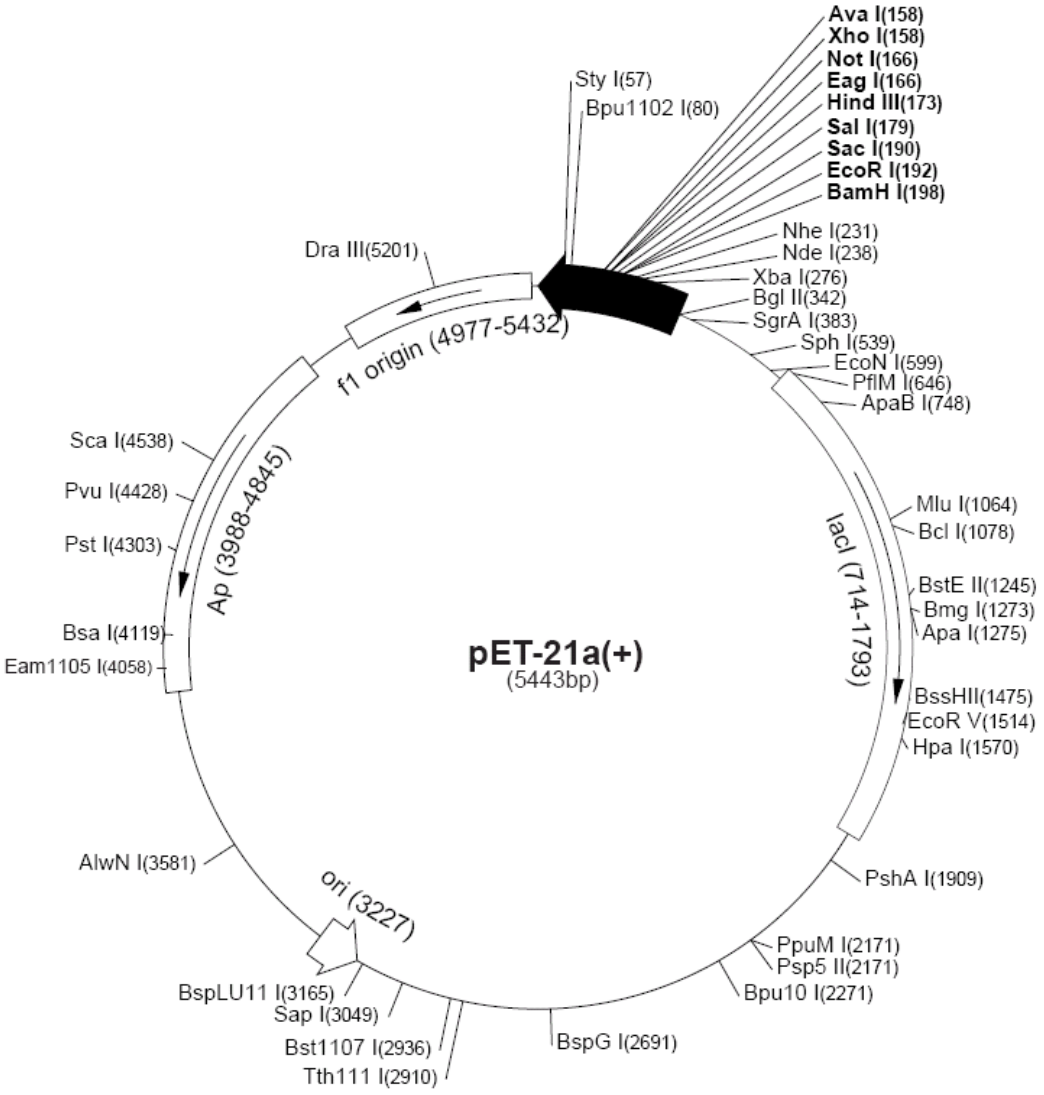



**Figure 5.2 Expression vector pET-21a(+)** [5]. This vector was employed to clone all the constructs created for this doctoral work employing restriction sites *NdeI* and *XhoI*.

[5] Adapted from https://www.addgene.org/12604/

5.5.3 pET16b(+) Novagen® ampicillin resistance.

The gene coding for the uroporphyrinogen III synthase from human was contained in this vector pET16b-U3S and has been described in the literature (Mathews 2001). This construct was kindly handed out by Prof. Heidi L. Schubert.

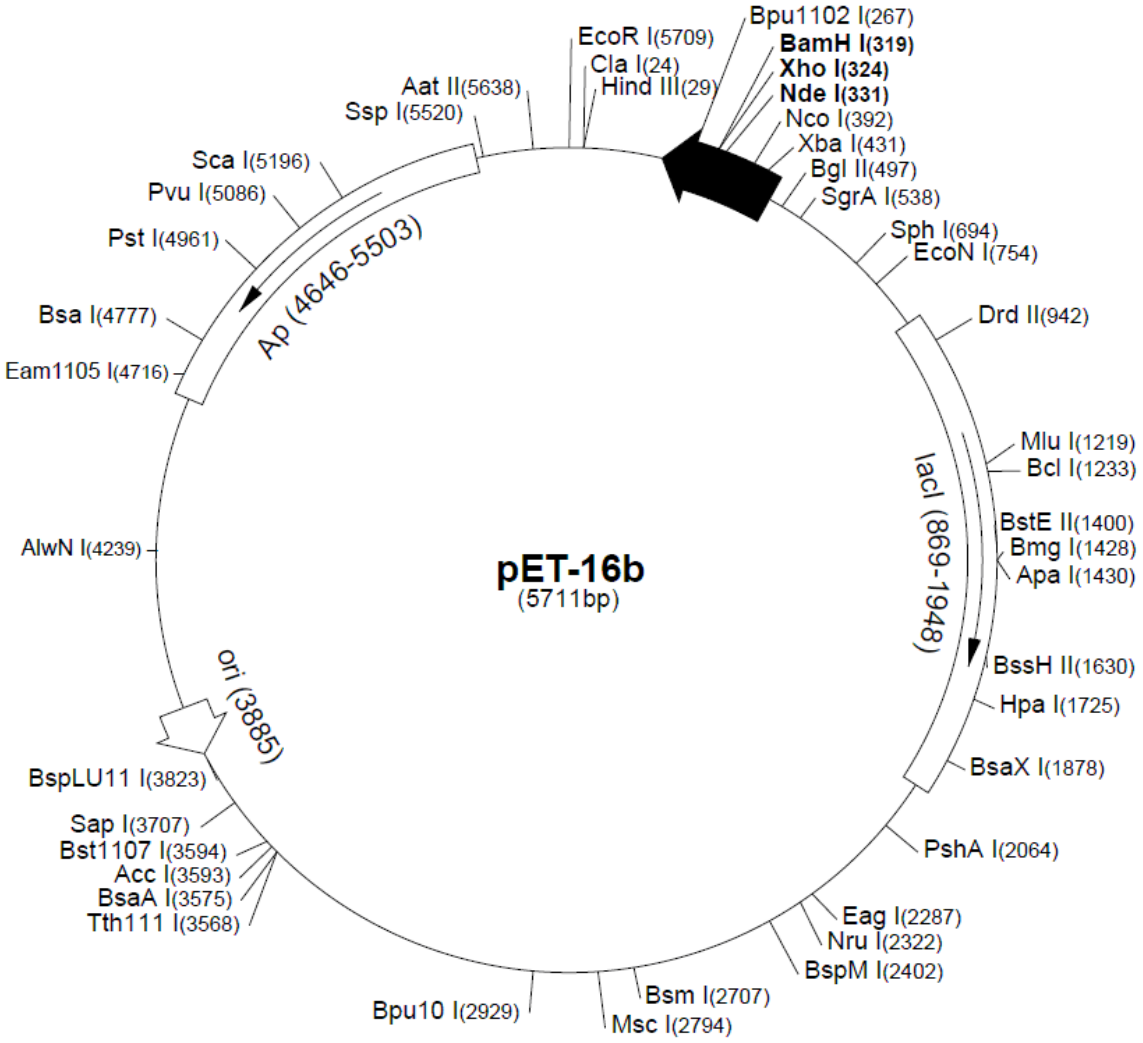



**Figure 5.3 Expression vector pET-16b(+)**[6]. This vectors contained the uroporphyrinogen III synthase from human described in **Chapter 3**.

[6] Adapted from http://www.biofeng.com/zaiti/dachang/pet16b.html

### 5.5.4 pET28a(+) Novagen® ampicillin resistance

The pET-28a-c(+) vector carries an N-terminal His•Tag®/thrombin/T7•Tag® configuration in addition to an optional C-terminal His•Tag sequence. Restriction sites are shown on the circle map. In this vector, the T7 expression region is reversed on the circular map. Construct PA5259 (Moynie et al. 2013) employed for the amplification of ctU3S and ctU3SΔ (**Chapter 2**) was contained in this vector and was kindly handed out by Prof. Gunter Schneider and Dr. Robert Schnell.

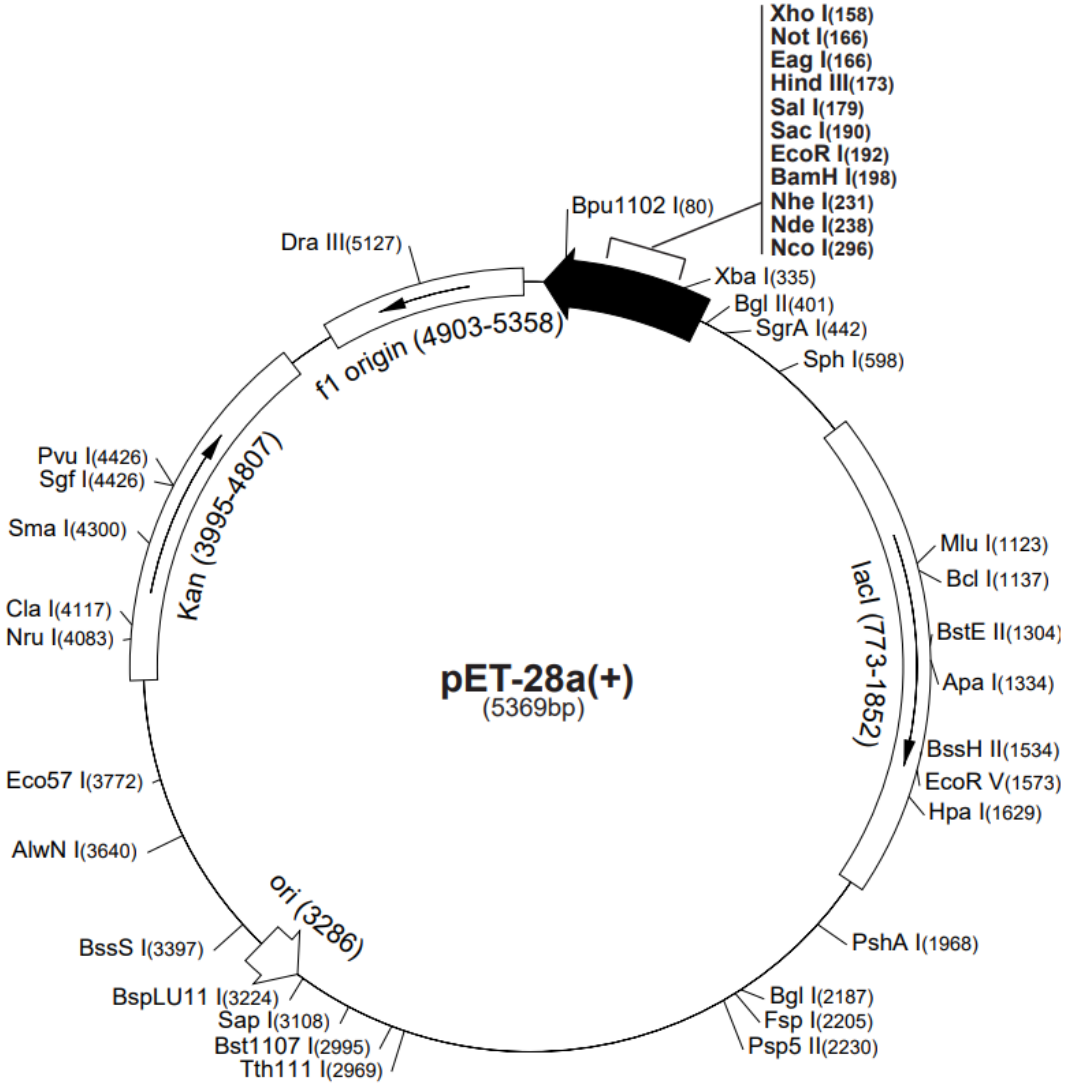



**Figure 5.4 Expression vector pET-28a(+) vector map by Novagen[7].**

[7] Adapted from (https://biochem.web.utah.edu/hill/links/pET28.pdf)

## 5.6 Software

- Ape-A Plasmid Editor (M. Wayne Davis)
- PyMOL (Molecular Graphics System, Version 1.7 and 2.0 Schrödinger, LLC.
- LigPlot+

# Chapter 6

## Methods

### 6.1. Bioinformatics

#### 6.1.1. Sequence-based profile-profile comparisons with HHsearch

To generate sequence-based Hidden-Markov-Models, we first built multiple sequence alignments for all domains of the α/β-class contained in the astral database release 2.07 (Chandonia, Fox, and Brenner 2019). The alignments were generated employing PSI-BLAST (Altschul 1997), which is a program included in the build.pl protocol described elsewhere (Söding 2005). This program was used to generate the profile-profile comparisons with default parameters. The secondary structure prediction was turned off (ssm=0) in order to perform a strictly sequence-based search. The gathered homologous regions were aligned with TM-align (Y. Zhang and Skolnick 2005). The obtained alignments were parsed and stored in a spreadsheet-based pivot table for further pair selection according to sequence identity and probability scores.

#### 6.1.2. Structural alignments

Structural alignments of individual candidates were assessed with PDBeFold (Krissinel and Henrick 2004), an online server for secondary-structure matching. In cases where only a part of the pdb structure was needed, the corresponding pdb coordinates were extracted manually employing The PyMOL Molecular Graphics System (by Schrödinger, LLC).

#### 6.1.3. Construction of a U3S database of protein halves

All U3S sequences with known structure were trimmed manually in half. Hereby, the cutsites were defined by eye, cutting in loop regions in the protein middle. N- and C-terminal regions kept exactly the same secondary structural elements.

### 6.1.4. Construction of a U3S dataset of permuted protein halves

Three different permuted variants were generated. Two of them by trimming the lower and upper lobes of U3S proteins, similar to the CATH domain assignment, but ignoring the connecting beta sheets between the lobes. Lower lobes were named as their pdb identifiers followed by the ending "_L1" for lower and "_L2" for upper lobes, e.g., 1jr2_L1 and 1jr2_L2 (**Section 5.4.1**). An additional variant can be generated by appending the N-terminal αβα-swap to the C-termini. These variants were named after their corresponding pdb identifiers followed by "_Cαβα", e.g, 1jr2_Cαβα (**Section 5.4.1**).

## 6.2. Cloning

### 6.2.1. Gene assembly of U3S and U3SΔ by PCR (Chapter).

The C-terminal half of the U3S (cU3S) gene from *Pseudomonas aeruginosa* was amplified by PCR, employing primers OT10_U3Spa_Cter_fwd and OT10_U3Spa_Cter_rev (**Section 5.3.1**), and PA5259 as template. PA5259 has been described elsewhere (Moynie et al. 2013) and was kindly handed out by Prof. Gunter Schneider and Dr. Robert Schnell. To delete the six amino acid insertion, two fragments A and B were amplified from cU3S employing the previously mentioned outer primers and adequate oligos (OT10_U3Spa_Cter_FA_del_rev and OT10_U3Spa_Cter_FB_del_fwd, respectively, **Section 5.4.1**), which flank the region to be deleted, creating instead sticky ends for a further annealing in a last PCR reaction. To finally assemble cU3S*Δ,* fragments A and B were mixed and amplified using the outer primers employed for cU3S. The assembled genes were cloned into a pET21a expression vector, yielding pET21a-cU3S and pET21a-cU3SΔ, respectively.

### 6.1.2. Gene assembly of chimeric hybrid UShsMMap01 (**Chapter 3**)

The chimeric gene was assembled in two steps as shown in **Figure 3.5**. Hereby, three flanking fragments termed A, B and C were amplified employing the parental gene corresponding to the cobalamin-binding subunit of Methylmalonyl-CoA mutase (MCoAM) from *Aeropyrum pernix* (**Section 5.5.1**). The middle fragment B, was amplified from pET16b-U3S described

elsewhere (Moynie et al. 2013), employing for each PCR reaction corresponding oligonucleotides as shown below:

*Fragment A:* *OT02_MCoA_A_fwd and OT02_MCoA_A_rev (**Section 5.3.1**)*

*Fragment B:* *OT02_UShs_B_fwd and OT02_UShs_B_rev (**Section 5.3.1**)*

*Fragment C:* *OT02_MCoA_C_fwd and OT02_McoA_C_rev (**Section 5.3.1***

### 6.1.3. PCR reaction conditions

| ***Vol. [µl]*** | ***Reagent*** | ***Stock conc.*** | ***Final conc*** |
|---|---|---|---|
| *10.0* | *Q5 Buffer* | *5x* | *1x* |
| *4.0* | *dNTPs (Eurogene®)* | *10 mM* | *0.8 mM* |
| *0.5* | *Template* | *100 ng/µl* | *50 ng* |
| *2.5* | *Forward primer* | *10 µM* | *0.5 µM* |
| *2.5* | *Reverse primer* | *10 µM* | *0.5 µM* |
| *1.0* | *Q5 DNA polymerase* | *2 u/µl* | *1 u* |
| *29.5* | *H2O* | | |

| ***Temperature*** | ***Cycles*** | ***Time*** |
|---|---|---|
| *98°C* | *1x* | *5 min* |
| *98°C*<br>*MT + 3°C*<br>*72°C* | *30x* | *30 sec*<br>*30 sec*<br>*120 sec* |
| *72°C* | *1x* | *5 min* |
| *4°C* | | *∞* |

### 6.1.4. Directed mutagenesis (Quick change)

Site-directed mutagenesis was carried out to insert a single mutation employing a method described elsewhere (Ho et al. 1989). The obtained products were purified with the Qiagen PCR purification method according to the manufacturer protocol before digestion with adequate restriction enzymes.

## *6.2. Molecular methods for purification*

### 6.2.2. Gel electrophoresis

All PCR products and digested vectors were mixed with a DNA-loading dye (Thermo scientific) and loaded onto a 1% agarose gel. DNA containing mixtures were carefully deposited into the gel pockets. A current of 70V was applied to the chamber to separate the PCR products. Running times varied depending on the size of the fragments. As running buffer **TAE-buffer** (**Section 5.2.2.1**) was employed.

### 6.2.3. DNA purification by gel extraction

PCR products and digested fragments were loaded onto an agarose gel (1%) for gel electrophoresis. The gel was observed under a UV lamp ECX-20M (Peqlab) to make the DNA-bands visible. Desired bands were extracted with a sterile cutter and deposited into a 1.5 ml Eppendorf® tube. Purification of the amplified fragments was performed according to the recommended protocol QIAquick gel extraction kit of QIAGEN® Handbook 2011. Elution volumes ranged from 35 to 50µl depending on the intensity of the bands.

## 6.3. DNA digestion with restriction enzymes

Digestion of DNA fragments and vectors was performed using *XhoI* and *NdeI* (Thermo Scientific) restriction enzymes. All digestion samples were incubated at 37°C for 30 minutes using the following reaction mix:

| ***Volume in [µl]*** | ***Component*** | ***Concentration*** |
|---|---|---|
| *5* | *Universal Buffer* | *10x* |
| *1* | *NdeI 100% activity* | *10u/µl* |
| *1* | *XhoI 100% activity* | *10u/µl* |
| *5* | *fragment* | *50 ng/* |
| *38* | *H2O* | |
| *50.00 (final vol.)* | | |

## 6.4. Vector dephosphorylation

To avoid vetor self-ligation, all digested vectors were phosphorylated directly after digestion without purifying the sample, by adding 1 µl of alkaline phosphatase (Fermentas) and incubating for 1 hour at 37°C.

## 6.5. DNA ligation with T4 DNA-ligase

The ligation of digested DNA fragments was performed using the T4 DNA-ligase from NE BioLabs®. All samples were incubated at 4°C for 24 hours and stored at -20°C if not used immediately after ligation. Different vector/insert molar ratios were tried (commonly 1 to 2, 4 and 6, respectively). The concentration of the digested vector was always 60 ng in 20 µl reaction volume.

## 6.6. Transformation of chemically competent cells

Three different chemically competent cells were employed: Top10™, BL21 and ArtictExpress™. All were prepared according to standard protocols (Dagert and Ehrlich 1979) and stored in 100 µl aliquots in 1.5 ml Eppendorf® tubes at -80°C until further usage. Competent cells were thawed on ice previous to addition of DNA. 50 ng of a plasmid solution or 10µl of a ligation reaction were added to the cells and incubated on ice for 15 min. A heat shock step was performed (45 seconds for Top10 and BL21 and 22 seconds for ArticExpress™) and incubated on ice for 15 min. 800 µl of LB medium (**Section 5.2.2.2**) were added to the cells for further incubation at 37°C for 1 hour. During this time, samples were shakenat 600 rpm using an Eppendorf® thermomixer model No.5436. After the incubation time transformed cells coming from ligation attempts were centrifuged at 4000 rpm using an Eppendorf® bench centrifuge model No.5425 for 5 min. Pellets were resuspended in 200 µl LB medium for their subsequent plating on agar Petri dishes with adequate antibiotics. Petri dishes were incubated at 37°C overnight or for 72 hours at room temperature.

## 6.7. Plasmid extraction

5 ml LB medium (**Section 5.2.2.2**) were inoculated with a single colony coming from Top10™ transformation and incubated overnight at 37°C. The extraction of the supercoiled DNA-plasmids from the cell culture was performed according to the recommended protocol of QIAGEN® Plasmid Mini Kit. Handbook 2011. Elution volumes ranged from 35 to 50µl.

## 6.8. Colony PCR

A screening for properly ligated target genes was assessed directly from freshly transformed colonies via PCR amplification. Single colonies were picked and added to the PCR-mix (**Section *6.2.3)*** using Taq-polymerase (Fermentas/Thermo Scientific). T7prom and T7ter oligonucleotides were employed as primers for the amplification. Resulting products were analysed by agarose gel electrophoresis.

## 6.9. DNA Sequencing

Vectors mentioned in this work were sequenced by the sequencing facility of the Max Planck Institute for Developmental Biology in Tuebingen or Eurofins Laboratory Scientific employing 50 ng of pure target DNA. In house sequencing required a manual sequencing reaction. Thus, 50 ng of DNA template were mixed with 2 µl of sequencing buffer, 0.5µl of BDT mix and the corresponding oligonucleotide to reach final concentration of 1µM. Water was added to a volume of 10µl and the reaction was set on a thermocycler employing the following sequencing program:

*96◦C 20 sec*
*50◦C 10 sec (30 cycles)*
*60◦C 4 min*

## 6.10. Test expression

An expression test was carried out to check solubility of modified proteins before scaling up expression. 15 ml of LB-medium were inoculated with freshly transformed colonies (either AEX or BL21) and incubated at 37°C until an optical density (OD600) of about 0.6 was reached. Immediately after the cell suspensions were split into two 50ml-flasks one of the cell solutions was induced with 1 mM final concentration IPTG and the other treated as a control to check for leaky expression. Special attention was paid to expression with AExpress. Once the cultures reached the optimal OD, they were cooled on ice to lower the temperature before addition of IPTG. Protein expression succeeded at 20°C (BL21) and 11-16°C (AExpress) overnight. Protein solubility was corroborated by Gel electrophoresis (**Section 6.15.3**)

## 6.11. Protein expression

Protein expression was usually performed in6 L batches.Single colonies resulting from plasmid transformation were employed to inoculate 50 ml LB medium enriched with 1 mM ampicillin and incubated at 37°C overnight. 10 ml of the starting culture were used to inoculate each 1 L TB medium (**Section 5.2.2.2**) enriched with 1 mM ampicillin and incubated at 37°C in 5L-sized flasks. As soon as an OD600 of ~0.4 was observed, the temperature of the incubator was reduced to 20°C. Expression of the protein was induced at ~0.6 OD600 by adding 1 mM IPTG incubating further for ~ 16 hours before harvesting. If not employed immediately, pellets were stored at -20/-80°C.

## 6.12. Cell disruption by sonication

After expression cells from 3 L culture were centrifuged at 4000 rpm using an Avanti J-26xPI (Beckmann Coultier) centrifuge for 20 min. The resulting pellet was resuspended in 10 ml of **Buffer-A**. 100 µl of a protease inhibitor cocktail (SERVA) were added to avoid proteolysis.
The cell suspension was disrupted by sonication with a Bandelin HD 3100 (Sonoplus) sonicator 3 times at 40% amplitude for 3 minutes. The lysate was centrifuged at 18000 rpm in an Avanti J-26xPI (Beckmann Coultier) centrifuge.

## 6.13. Refolding

Protein refolding was usually performed from inclusion bodies coming from 1 L culture. If not mentioned otherwise all incubation steps were performed at 4°C. The inclusion bodies-containing pellet was resuspended in 5 ml **R1-buffer** and incubated for at least 60 min. 5ml of **R2-buffer** were added to the previous mixture and mixed gently before further incubation for 1h. After this time, the protein suspension was centrifuged at 18000 rpm for 60 min to remove membrane and other non-protein cell components. The resulting protein solution was mixed with **R3-buffer** to reach a final volume of 25 ml and dialyzed 2 times against 5 L of **Buffer-A**. The resulting solution was centrifuged at 18000 rpm and 4°C for 1 hours using an Avanti J-26xPI (Beckmann Coultier) centrifuge to separate soluble proteins from their aggregates.

## 6.14. Protein purification

### 6.14.1. Nickel affinity chromatography

Protein samples coming from lysate or refolding were passed through a 0.2um filter prior to injection onto a 5ml-nickel column HisTrap HP™ (GE Healthcare Life Sciences) previously equilibrated with 50 ml of **Buffer-A**. After loading the protein, a wash step using 20 ml of **buffer-A** was performed. The elution of the target proteins succeeded with an imidazole linear gradient with **Buffer-B**.

### 6.14.2. Preparative Gel filtration

Protein samples were passed through a 0.2 µm filter or centrifuged at 18 000 rpm for 15 min before injection. Preparative Superdex™ columns S200 or S75 were equilibrated with 350 ml of **GeFi-buffer**. The protein sample was loaded and run at a constant flow rate of 1.5 ml/min.

## 6.15. Protein analysis by gel electrophoresis

To determine size and purity of a given protein, polyacrylamide gels (15%) were prepared. Protein samples were heated at 95°C for 5 min in the presence of denaturing SDS-buffer. 10 to 15 µl of the denatured sample were carefully deposited in the gel pockets and subjected to 70 mV for separation.

## 6.16. Biophysical characterization of proteins

### 6.16.1. Analytical gel filtration

Prior to injection the protein sample was either filtered with a 0.2 µm filter or centrifuged at 18 000 rpm using an Avanti J-26xPI (Beckmann Coultier) centrifuge. Analytical Superdex$^{TM}$ columns S200 or S75 were equilibrated with 50 ml of GeFi-buffer (**Section 5.2.7**). After injecting the protein sample, a 25 ml run was performed at a constant flow rate of 0.5 ml/min.

### 6.16.2. Circular dichroism and thermal denaturation

A pure protein sample was dialyzed overnight against **CD-buffer** to get rid of salts. 200µl of the dialyzed sample was adjusted to a concentration of 0.2 to 0.3 mg/ml and pipetted into a clean 1 mm quartz cuvette. Five instrumental replica were performed using a J-810 CD-spectrometer (Jasco) with following parameters:

| | |
|---|---|
| *Band width* | *1nm* |
| *Response* | *2sec* |
| *Sensitivity* | *standard* |
| *Measurement range* | *190 - 240 nm* |
| *Data pitch* | *0.1nm* |
| *Scanning speed* | *100 nm/min* |
| *Instrument replica* | *5* |

The same sample used for CD measurements was employed for the thermal denaturation on the same instrument using the following parameters:

| | |
|---|---|
| *Band width* | *1nm* |
| *Response* | *2sec* |
| *Sensitivity* | *standard* |
| *Measurement range* | *190 - 240 nm* |
| *Data pitch* | *0.1nm* |
| *Monitor wavelength* | *222nm* |
| *Temperature slope* | *0.5°C* |

The CD data in this work is reported in units of ellipticity and has been normalized for the molar concentration of the sample, employing the sample concentration, the cell pathlength, and the molecular weight.

### 6.16.3. Fluorescence

A pure protein sample (usually the same employed for CD) was adjusted to a concentration of 0.2 to 0.3 mg/ml and pipetted into a clean black fluorescence quartz cuvette with round bottom. Five instrumental replica were performed using a FP-6500 fluorescence-spectrometer (Jasco) with following parameters:

| | |
|---|---|
| *Measurement mode* | *Emission spectrum* |
| *Band width (ex)* | *3nm* |
| *Band width (em)* | *3nm* |
| *Response* | *0.2sec* |
| *Sensitivity* | *medium* |
| *Measurement range* | *300 - 400 nm* |
| *Data pitch* | *0.2nm* |
| *Excitation wavelength* | *280 nm* |
| *Scanning speed* | *200 nm/min* |

### 6.16.4. UV-Vis spectrometry

All measurements were carried out on a Cary 50 Scan UV-VIS spectrometer. Stock solutions for all cobalamins were prepared at 8 mM and diluted to 40 μM to take reference spectra, employing the corresponding buffer, depending on the measurement.

## 6.17. Structure determination by X-ray crystallography

### 6.17.1. Crystal screening

To screen for optimal conditions for crystallization 0.4 µl of a pure protein sample at 10 to 12.5 mg/ml were mixed with the same volume of a screening buffer (**Section 5.2.6**) and deposited in 96-well plates via hanging drop method, employing a Mosquito crystallization robot. Screening plates were stored at 16°C or 4°C.

### 6.17.2. Crystal improvement

Once crystals were obtained, the crystallization conditions were reproduced via hanging drop method, increasing the drop sized to 1 - 1.5 µl. In addition, the components of the screening buffer such as pH, salt concentration and precipitant concentration were varied in an attempt to improve crystal quality. All conditions were pipetted manually and all instruments needed were equilibrated at the desired incubation temperature before mixing with the protein sample.

### 6.17.3. Crystal handling and measurement

Obtained crystals were fished with X-ray mounting loops in adequate sizes and cryoprotected in a solution of 30% v/v PEG 400. This was contained in the same crystal mother liquor. Fished crystals were finally cryocooled by plunging them into liquid nitrogen prior to data collection on the X10SA beamline at the Swiss Light Source, Paul Scherrer Institute (PSI) in Villingen Switzerland or the Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung (BESSY). The data was collected at 100K and 0.5 oscillation degrees. The obtained images were recorded on a Pilatus 6M or a Pixel-detector Pilatus 2M at a distance of 270 nm and a wavelength of 1.000Å. The obtained reflexions were indexed, integrated and scaled with the XDS package (Kabsch 2010), employing 5% reflections for the test set.

### 6.17.4. Crystal soaking with cobalamin derivatives

Apo crystals grown in 0.1 M Tris at pH 5.7 and 15% PEG 4000 Were fished with appropriate crystal loops and deposited in a ligand solution with the same mother liquor for dilution. The concentration of the ligands varied according to **Table 6.1**. Soaking times ranged from 30 minutes to 24 hours. The best soaking time was 3 hours.

**Table 6.1 soaking conditions with different cobalamin derivatives**

| ***Ligand*** | ***Concentration*** |
|---|---|
| *Adenosylcobalamin, methylcobalamin, hydroxocobalamin and cyanocobalamin* | *400 µM/ 4mM* |
| *alpha-ribazole3-phosphate* | *10/20/40 mM* |
| *dimethylbenzimidazole* | *20 mM (5% DMSO)* |

## 6.18. Structure determination with NMR (By Murray Coles and Manish Chaubey)

2.9 mM of labelled cU3SΔ were contained in 50 mM KP and 150 mM KCl, 80% H2O/20% D2O at pH 8. All spectra were recorded at 288 K on Bruker AVIII-600 and AVIII-800 spectrometers. Backbone sequential assignments were completed using standard triple resonance experiments implemented using selective proton flip-back techniques for fast pulsing (Diercks, Daniels, and Kaptein 2005). Aliphatic side chain assignments were completed by a combination of CCH-COSY and CCH-TOCSY experiments, while aromatic assignments were made by linking aromatic spin systems to the respective CH2 protons in a 2D-NOESY spectrum, combined with a PLUSH-TACSY experiment (Carlomagno et al. 1996). Stereospecific assignments and the resulting rotamer assignments were determined from an HNHB experiment and NOESY cross-peak patterns. Distance data were derived from 3D15N-HSQC NOESY and 3D-NNH-NOESY spectra on a $^{15}$N-labeled sample, and 3D13C-HSQC-NOESY and 3D-CCH- and 3D-CNH-NOESY spectra (Diercks, Coles, and Kessler 1999) on the $^{15}$N,$^{13}$C-labeled sample. Aromatic contacts were observed in a $^{15}$N-filtered 2D-NOESY spectrum. Structural restraints were compiled using a protocol aimed at high local accuracy using expectation NOESY spectra to test local conformational hypotheses (in-house software). Chemical shift similarity searches using TALOS (Cornilescu, Delaglio, and Bax

1999) were used to generate hypotheses for backbone conformations, while sidechain rotamers were searched exhaustively. Conformations identified in this manner were applied via dihedral restraints, using the TALOS-derived tolerances for backbone and ±30° for sidechains. Further NOE contacts were assigned iteratively using back-calculation of expectation NOESY spectra from preliminary structures. NOESY crosspeaks in the three-dimensional spectra were converted into distance ranges after rescaling according to corresponding HSQC intensities. Crosspeaks were divided into four classes, which resulted in restraints on upper distances of 2.7, 3.2, 4.0 and 5.0 Å, respectively. Additional classes of 3.6 and 4.5 Å were used for medium and weak backbone contacts often affected by spin-diffusion. Lower distance restraints were included for very weak or absent sequential HN-HN crosspeaks using a minimum distance of 3.5 Å and medium intensity or weaker sequential and intraresidue HN-H# crosspeaks using a minimum distance of 2.7 Å. Allowances for the use of pseudoatoms (using averaging) were added for methyl groups and non-stereospecifically assigned methylene groups. Hydrogen bond restraints were applied for residues in secondary structure where donor–acceptor pairs were consistently identified in preliminary calculations. The restraints were applied via inclusion of pseudo-covalent bonds between heavy atom acceptors and hydrogen donors, with force constants of 14 kcal/Å2 and 8 kcal/rad2 on bond lengths and angles, respectively. Structures were calculated with XPLOR (NIH version 2.9.3 (C. Schwieters, Kuszewski, and Mariusclore 2006; C. D. Schwieters et al. 2003)) using a three-stage simulated annealing protocol. The second stage included a conformational database potential, while the third used a relaxed force constant on peptide bond planarity. Sets of 50 structures were calculated and a final set of 20 chosen on the basis of lowest restraint violations. An average structure was calculated and regularized to give a structure representative of the ensemble. Statistics for the final structure set are presented in **Appendix 7.3**. A preliminary structure for cU3SΔ was also calculated during validation of the recently published CoMAND method of NMR structure determination (ElGamacy et al. 2019). Briefly, this method uses spectral decomposition of CNH-NOESY spectra to derive local conformational parameters, providing input for de novo folding routines (in this case Rosetta (Das and Baker 2008)). Convergence is monitored by a quantitative R-factor expressing the match between back-calculated CNH-NOESY spectra and the 2 experimental spectrum. This structure therefore provides independent confirmation of the cU3SΔ fold, with an RMSD of 1.98 Å to the refined structure presented here.

## 6.19. Binding assays

### 6.19.1. Analytical gel filtration in presence of cobalamins

A pure UShsMMap02 protein sample at 80 μM was mixed 1:1 with a buffer-containing either OHClb, CNCbl, MeCbl or S-AdoCbl at 40 mM and incubated on ice for 30 min. After this time, the protein-ligand sample was centrifuged at 18000 rpm at 4C to eliminate possible aggregates and the supernatant was loaded onto an analytical gel filtration column. The observed peaks for free Cbl and Cbl-protein complex were collected and analyzed photometrically at 351 (Cbl) and 280 nm (protein), respectively. The samples were contained in 20 mM Tris at pH 7.4, 100 mM NaCl and 1 mM $\beta$-mcaptoEtOH.

### 6.19.2. Aerobic UV-vis monitoring of free cobalamin Vs. bound cobalamin

*Free cobalamin samples were prepared in the corresponding protein buffer at a concentration of 4 mM. An aliquot was diluted to 40* μM to record the spectra. Protein-Cbl complex was recorded right after eluting from analytical gel filtration, diluting when required. All spectra were recorded on a Varian Cary 50 Scan UV-VIS spectrometer.

### 6.19.3. Isothermal titration calorimetry

A UShsMMap02 protein sample (300 μl) at 282 μM was titrated with a cyanocobalamin solution at 3 mM in steps of 2 μl injections with spacing of 300 sec between each injection. The protein was contained in 20 mM TRIS (pH 7.4), 100 mM NaCl and 1 mM β-mercaptoethanol. The measurements were carried out on a TA Instruments nano ITC LV calorimeter in a 191μL sample cell. The sample cell was overloaded with 300μL of protein sample. All samples were degassed and temperature was equilibrated for 30min before the measurement.

### 6.19.4. Anaerobic size exclusion chromatography

Buffers and reagents were anoxically prepared in an anaerobic Glove Box from Coy Lab (© MR) in the laboratory of Holger Dobbek at HU Berlin. A Stock solution of 4 mM OHCbl and 200 mM ascorbic acid were prepared in 100 mM KP buffer at pH8 and 300 mM KCl. 30 µl of the ascorbic acid solution (20 mM end concentration) were added to 270 µl of the OHCbl solution (3.6 mM end concentration). The mixture immediately turned orange, indicating the successful reduction of cobalamin with ascorbic acid. The reduced cofactor sample was given to a protein sample at a concentration of 1.5 mM and loaded onto a 5 ml-bench gel filtration column for separation of the protein-cofactor complex. The eluting fractions were deposited 1 ml each in 1.5 ml Eppendorf containers.

### 6.19.5. Cobalamin titration with imidazole

40 µM OHCb contained in water were titrated in steps of 2 µM (2 µl injections) of imidazole (3 mM) and followed photometrically in a 200 - 700 nm range on a Varian Cary 50 Scan UV-VIS spectrometer.

### 6.19.6. Cobalamin titration with beta-mercaptoethanol

40 µM OHCb contained in water were titrated in steps of 4 µM (2 µl injections) with a beta-mercaptoethanol previously diluted in water. The measurement was followed photometrically in a 200 - 700 nm range on a Varian Cary 50 Scan UV-VIS spectrometer.

## 6.20. Synthesis of α-ribazole-3'-phosphate (by Bruce Lichtenstein)

The synthesis of αR3P was carried out by hydrolyzing cyanocobalamin (**Figure 6.1**) after protocols described in the literature (Lezius and Barker 1965; K. L. Brown and Hakimi 1986). Hereby, cyanocobalamin (CNCbl) (2 g, 1.48 mmol) was dissolved in 75 mL perchloric acid and stirred for 3 hours in the dark. The reaction was transferred to a large beaker in ice equipped with a magnetic stir bar. 75 mL ice cold water was added to the solution with stirring. The solution pH was adjusted to 0 – checked by pH paper – by dropwise addition of 10 M NaOH. Once at pH 0, approximately 1 g of Tris was added to aid in buffering at elevated pH. To remove the excess perchlorate as precipitate, the beaker is equipped with a

pH electrode and with rigorous stirring, aliquots of 2 M KOH were added carefully until the solution reached a pH of 8-8.5. The solution was stored overnight at 4 ºC to further the precipitation. The obtained precipitate was filtered and washed with ice cold water (approximately 50 mL) until no red color was observed. To remove additional potassium perchlorate, the filtrate was concentrated in vacuo, cooled overnight at 4 ºC and again filtered. The filtered solution was loaded onto a Dowex-22 formate column connected to an Akta FPLC. Weakly associated red hydrolysis products were washed from the column with 50 mM acetic acid (3.5 L) and crude phosphorylated products (including a majority α-ribazole 3'-phosphate) were eluted with 20 mM HCl (approximately 1-1.5 L). Elution with 20 mM HCl was monitored at 280 nm and fractions containing substantial amounts of absorbing species are combined. The combined eluant is freeze-dried. The slightly pink dry material was dissolved in water with 0.1% formic acid and loaded in batches onto pre-equilibrated Sep Pak C18 columns attached to an Akta FPLC. The Sep Pak C18 columns were washed with 0.1% formic acid in water and 0.1% formic acid in 95/5 water/acetonitrile while monitoring at 280 nm to ensure a return to baseline. αR3P was eluted with 0.1% formic acid in 90/10 water/acetonitrile. After recrystallization from boiling water and acetone, 213 mg αR3P were obtained (0.594 mmol, 40%) as the zwitterion. The compound validation was checked by NMR and mass spectrometry (**Appendix 7.11 to 12**).

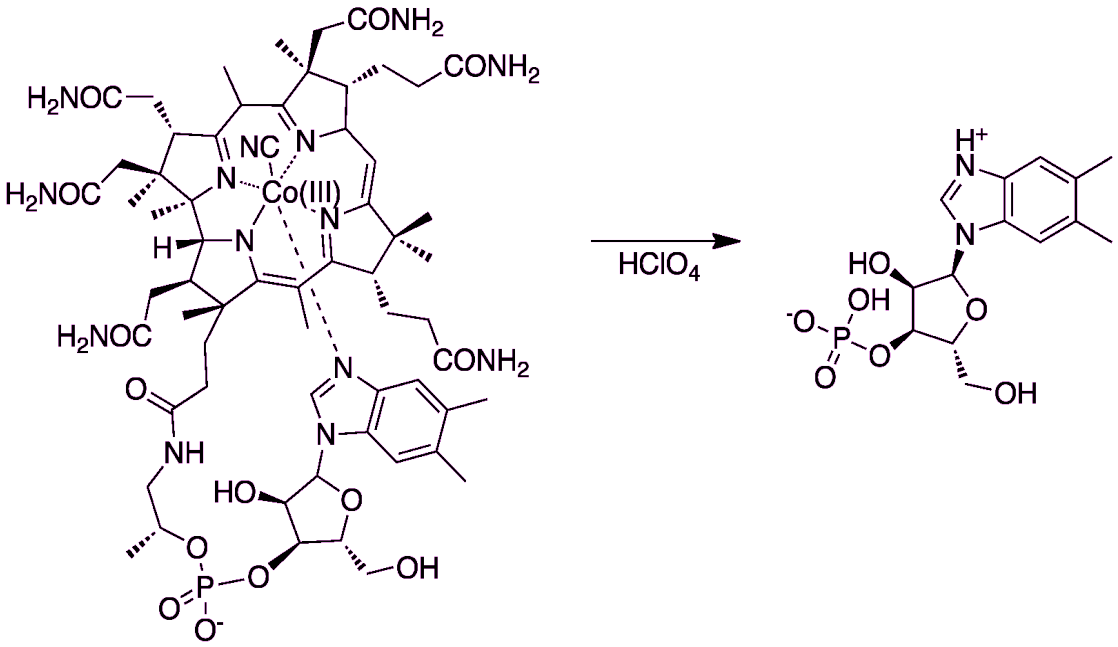
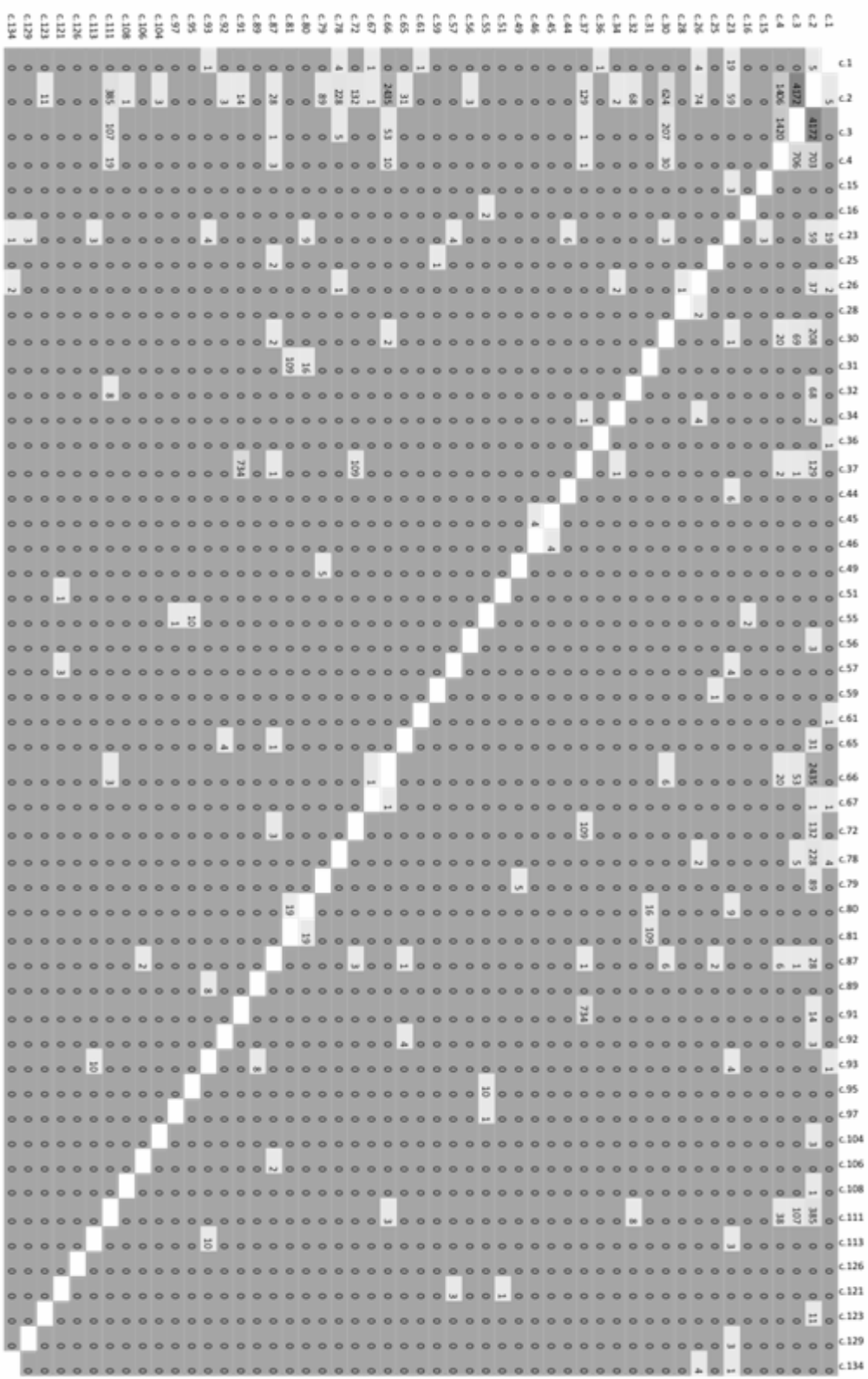



**Figure 6.1 Synthesis of α-ribazole-3'-phosphate from cyanocobalamin hydrolysis with perchloric acid ($HClO_4$).**

# 7. Appendix

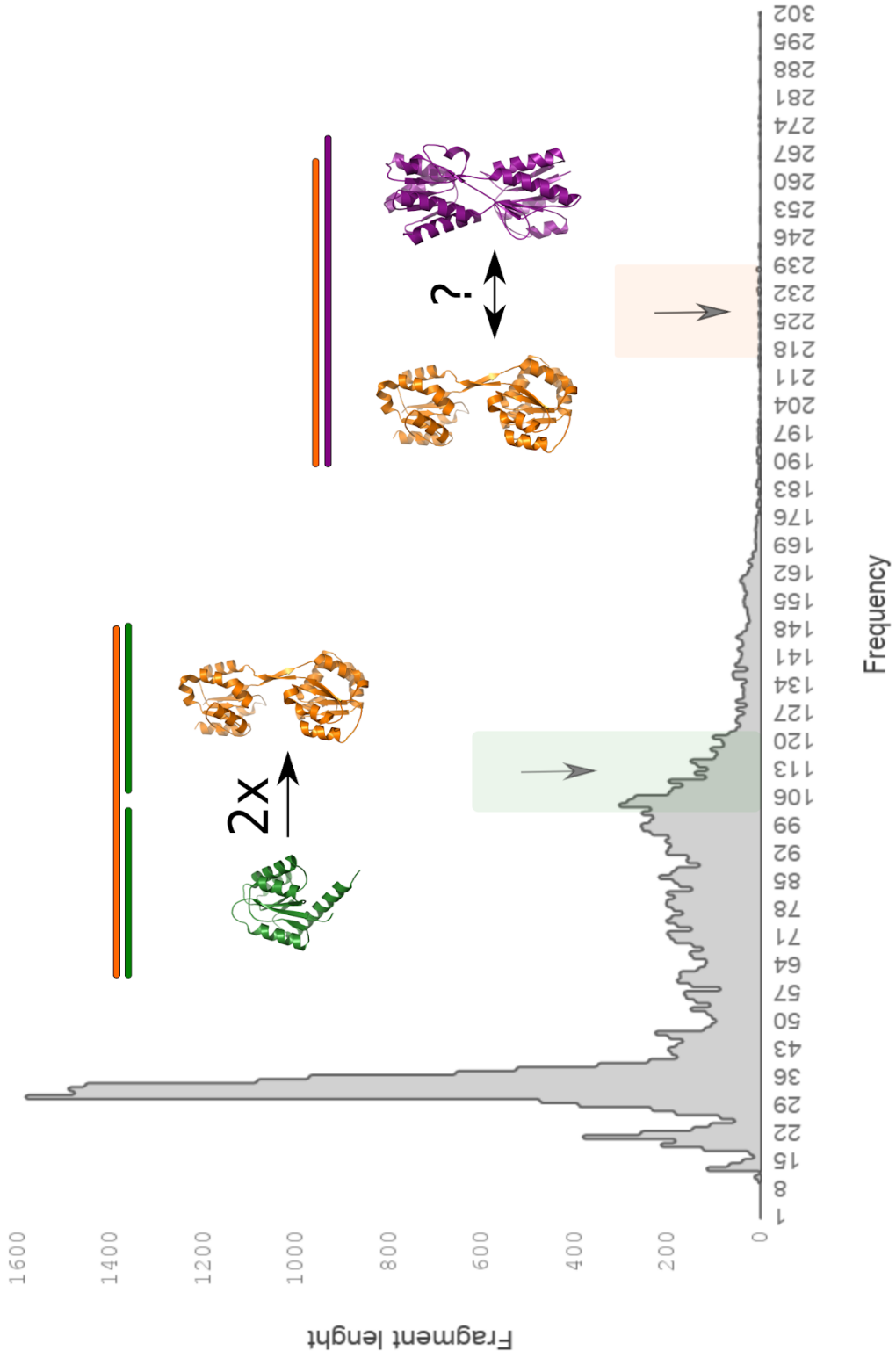
## 7.1. Homologous fragments across folds.

Folds sharing homologous regions with probability scores above 80% have been plotted for the α/β-class of SCOP. The number of hits (numbers in squares) is considerably higher for certain member than others because they are overrepresented in the pdb database.



## 7.2 Distribution of the HHsearch hits obtained for the α/β folds of SCOP.

The obtained hits were filtered at 80% probability. The most populated length is observed at around 30 residues. However fragments with larger lengths spread further to domain-sized regions. For instance, flavodoxin-like (green, circa 115 residues) matches twice to the hemD-like fold (orange, circa 250 residues) and the periplasmic-binding protein-like I (purple, circa 300 residues) also matches nearly the entire hemD-like domain.



### 7.3. NMR refinement statistics for cU3SΔ structure

| | SA | <SA>r |
|---|---|---|
| **Restraint Violations**[8] | | |
| Distance restraints (Å) | | |
| All (676) | 0.013 ± 0.001 | 0.013 |
| Intra-residue (98) | 0.004 ± 0.001 | 0.004 |
| Inter-residue sequential (192) | 0.016 ± 0.001 | 0.016 |
| Medium range (85) | 0.018 ± 0.001 | 0.016 |
| Long range (225) | 0.014 ± 0.001 | 0.015 |
| H-bond (76) | 0.000 ± 0.000 | 0.000 |
| Persistent viol. thres.2 | 0.075 | - |
| Dihedral restraints (°) | | |
| All (347) | 0.047 ± 0.005 | 0.045 |
| Persistent viol. thres.[9] | 0.25 | |
| H-bond restraints[10] | | |
| Distance (Å) (76) | 2.18 ± 0.11 | 2.13 ± 0.12 |
| Antecedent angle (°) | 13.0 ± 5.1 | 13.7 ± 5.7 |
| **Covalent Geometry** | | |
| Bonds (Å × 10E-3) | 2.66 ± 0.03 | 2.65 |
| Angles (°) | 0.65 ± 0.01 | 0.65 |
| Impropers (°) | 1.27 ± 0.03 | 1.22 |

[8] Violations are expressed as RMSD ± SD unless otherwise stated. Numbers in brackets indicate the number of restraints of each type.

[9] Persistent violations are defined as those occurring in at least 75% of all structures. The thresholds at which no persistent violations occur are tabulated.

[10] 3 Hydrogen bonds were treated as pseudo-covalent bonds. Deviations are expressed as the average distance/average deviation from linearity for restrained hydrogen bonds.

| **Structure Quality Indicators**[11] | | |
|---|---|---|
| Ramachandran Map (%) | 99.1 / 0.9 / 0.0 | |
| **Atomic R.M.S.D (Å)**[12] | | |
| | Backbone Heavy Atom | All Heavy Atom |
| SA vs <SA> | 0.27 ± 0.06 | 0.86 ± 0.06 |
| SA vs <SA>r | 0.36 ± 0.05 | 1.04 ± 0.08 |
| <SA> vs <SA>r | 0.24 | 0.72 |

[11] 3 Hydrogen bonds were treated as pseudo-covalent bonds. Deviations are expressed as the average distance/average deviation from linearity for restrained hydrogen bonds.

[12] Structures are labeled as follows: SA, the final set of 20 simulated annealing structures; <SA>, the mean structure calculated by averaging the coordinates of SA structures after fitting over secondary structure elements; <SA>r, the structure obtained by regularizing the mean structure under experimental restraints. RMSD values were obtained based on superimpositions over ordered residue (defined as P3-L118).

## 7.4. Refinement statistics for UShsMMap01

| **Protein** | **UShsMMap01** |
|---|---|
| Wavelength | 1.0000 |
| Resolution range | 33.0 - 2.0 (2.0 - 2.0) |
| Space group | P 1 21 1 |
| Unit cell (Å, °) | 49.589 43.196 65.664 90 108.218 90 |
| Total reflections | 52908 (4903) |
| Unique reflections | 17758 (1750) |
| Multiplicity | 3.0 (2.8) |
| Completeness (%) | 98.70 (97.55) |
| Mean I/sigma(I) | 12.60 (2.19) |
| Wilson B-factor | 24.96 |
| R-merge | 0.07722 (0.5379) |
| R-meas | 0.09426 (0.6626) |
| R-pim | 0.05337 (0.3822) |
| CC1/2 | 0.997 (0.757) |
| CC* | 0.999 (0.928) |
| Reflections used in refinement | 17752 (1750) |
| Reflections used for R-free | 888 (87) |
| R-work | 0.19 (0.26) |
| R-free | 0.24 (0.33) |
| CC(work) | 0.962 (0.829) |
| CC(free) | 0.933 (0.746) |
| Number of non-hydrogen atoms | 2331 |
| macromolecules | 2093 |
| solvent | 238 |
| Protein residues | 272 |
| RMS(bonds) | 0.016 |
| RMS(angles) | 1.18 |
| Ramachandran favored (%) | 97.41 |
| Ramachandran allowed (%) | 2.22 |
| Ramachandran outliers (%) | 0.37 |
| Rotamer outliers (%) | 0 |
| Clashscore | 6.09 |
| Average B-factor | 33.84 |
| macromolecules | 33.31 |
| solvent | 38.56 |
| Number of TLS groups | 3 |

## 7.5. Refinement statistics for UShsMMap02

| **Protein** | **UShsMMap02** |
|---|---|
| Wavelength | 1.0000 |
| Resolution range | 37.0 - 2.7 (2.8 - 2.7) |
| Space group | P 1 21 1 |
| Unit cell (Å, °) | 49.685 42.486 64.295 90 96.883 90 |
| Total reflections | 51018 (4944) |
| Unique reflections | 7723 (726) |
| Multiplicity | 6.6 (6.5) |
| Completeness (%) | 99.16 (94.63) |
| Mean I/sigma(I) | 11.61 (1.15) |
| Wilson B-factor | 78.28 |
| R-merge | 0.08823 (1.207) |
| R-meas | 0.096 (1.313) |
| R-pim | 0.03733 (0.5109) |
| CC1/2 | 0.998 (0.712) |
| CC* | 0.999 (0.912) |
| Reflections used in refinement | 7686 (723) |
| Reflections used for R-free | 385 (36) |
| R-work | 0.23 (0.37) |
| R-free | 0.28 (0.31) |
| CC(work) | 0.959 (0.736) |
| CC(free) | 0.907 (0.726) |
| Number of non-hydrogen atoms | 2087 |
| macromolecules | 2087 |
| Protein residues | 272 |
| RMS(bonds) | 0.015 |
| RMS(angles) | 1.5 |
| Ramachandran favored (%) | 90.74 |
| Ramachandran allowed (%) | 6.67 |
| Ramachandran outliers (%) | 2.59 |
| Rotamer outliers (%) | 0.44 |
| Clashscore | 19.51 |
| Average B-factor | 98.04 |
| macromolecules | 98.04 |
| Number of TLS groups | 3 |

| **Protein** | **UShsMMap03** |
|---|---|
| Wavelength | 1.0000 |
| Resolution range | 38.0 - 2.2 (2.3 - 2.2) |
| Space group | P 1 21 1 |
| Unit cell (Å, °) | 48.348 42.668 60.61 90 92.044 90 |
| Total reflections | 82631 (7117) |
| Unique reflections | 12602 (1168) |
| Multiplicity | 6.6 (6.1) |
| Completeness (%) | 98.98 (93.74) |
| Mean I/sigma(I) | 10.54 (1.70) |
| Wilson B-factor | 47.62 |
| R-merge | 0.09053 (1.2) |

## 7.6. Refinement statistics for UShsMMap03

| | |
|---|---|
| R-meas | 0.09857 (1.313) |
| R-pim | 0.03849 (0.5241) |
| CC1/2 | 0.998 (0.677) |
| CC* | 0.999 (0.898) |
| Reflections used in refinement | 12601 (1168) |
| Reflections used for R-free | 631 (59) |
| R-work | 0.22 (0.39) |
| R-free | 0.25 (0.41) |
| CC(work) | 0.965 (0.712) |
| CC(free) | 0.924 (0.568) |
| Number of non-hydrogen atoms | 2100 |
| macromolecules | 2100 |
| Protein residues | 273 |
| RMS(bonds) | 0.02 |
| RMS(angles) | 2.11 |
| Ramachandran favored (%) | 91.88 |
| Ramachandran allowed (%) | 7.01 |
| Ramachandran outliers (%) | 1.11 |
| Rotamer outliers (%) | 21.24 |
| Clashscore | 30.43 |
| Average B-factor | 68.57 |
| macromolecules | 68.57 |
| Number of TLS groups | 3 |

## 7.7. Refinement statistics for UShsMMap04

| **Protein** | **UShsMMap04** |
|---|---|
| Wavelength | 1.0000 |
| Resolution range | 41.0 - 2.4 (2.6 - 2.4) |
| Space group | P 1 21 1 |
| Unit cell (Å, °) | 70.546 47.294 88.102 90 103.354 90 |
| Total reflections | 138600 (11296) |
| Unique reflections | 21373 (1902) |
| Multiplicity | 6.5 (5.9) |
| Completeness (%) | 98.23 (88.25) |
| Mean I/sigma(I) | 11.12 (1.31) |
| Wilson B-factor | 51.23 |
| R-merge | 0.1161 (1.151) |
| R-meas | 0.1265 (1.259) |
| R-pim | 0.04958 (0.5036) |
| CC1/2 | 0.998 (0.659) |
| CC* | 1 (0.891) |
| Reflections used in refinement | 21313 (1901) |
| Reflections used for R-free | 1067 (95) |
| R-work | 0.24 (0.29) |
| R-free | 0.30 (0.37) |
| Number of non-hydrogen atoms | 3709 |
| macromolecules | 3709 |
| Protein residues | 521 |
| RMS(bonds) | 0.010 |
| RMS(angles) | 1.52 |
| Ramachandran favored (%) | 86.25 |
| Ramachandran allowed (%) | 9.43 |
| Ramachandran outliers (%) | 4.32 |
| Rotamer outliers (%) | 1.68 |
| Clashscore | 10.38 |
| Average B-factor | 75.61 |
| macromolecules | 75.61 |
| Number of TLS groups | 6 |

## 7.8. Refinement statistics for UShsMMap02 in complex with DMB

| **Protein** | **UShsMMap02 (DMB co-crystallization)** |
|---|---|
| Wavelength | 1.0000 |
| Resolution range | 49.0 - 2.7 (2.8 - 2.7) |
| Space group | P 1 21 1 |
| Unit cell (Å, °) | 49.717 42.698 64.935 90 96.948 90 |
| Total reflections | 26856 (2547) |
| Unique reflections | 7706 (736) |
| Multiplicity | 3.5 (3.4) |
| Completeness (%) | 99.03 (96.07) |
| Mean I/sigma(I) | 8.92 (1.33) |
| Wilson B-factor | 70.17 |
| R-merge | 0.07745 (0.8707) |
| R-meas | 0.0917 (1.041) |
| R-pim | 0.04845 (0.5626) |
| CC1/2 | 0.997 (0.747) |
| CC* | 0.999 (0.925) |
| Reflections used in refinement | 7683 (733) |
| Reflections used for R-free | 383 (36) |
| R-work | 0.23 (0.4015) |
| R-free | 0.26 (0.4342) |
| CC(work) | 0.947 (0.741) |
| CC(free) | 0.952 (0.707) |
| Number of non-hydrogen atoms | 2107 |
| macromolecules | 2096 |
| ligands | 11 |
| Protein residues | 273 |
| RMS(bonds) | 0.004 |
| RMS(angles) | 0.67 |
| Ramachandran favored (%) | 91.88 |
| Ramachandran allowed (%) | 7.75 |
| Ramachandran outliers (%) | 0.37 |
| Rotamer outliers (%) | 0 |
| Clashsscore | 7.69 |
| Average B-factor | 95.62 |
| macromolecules | 95.62 |
| ligands | 96.56 |
| Number of TLS groups | 3 |

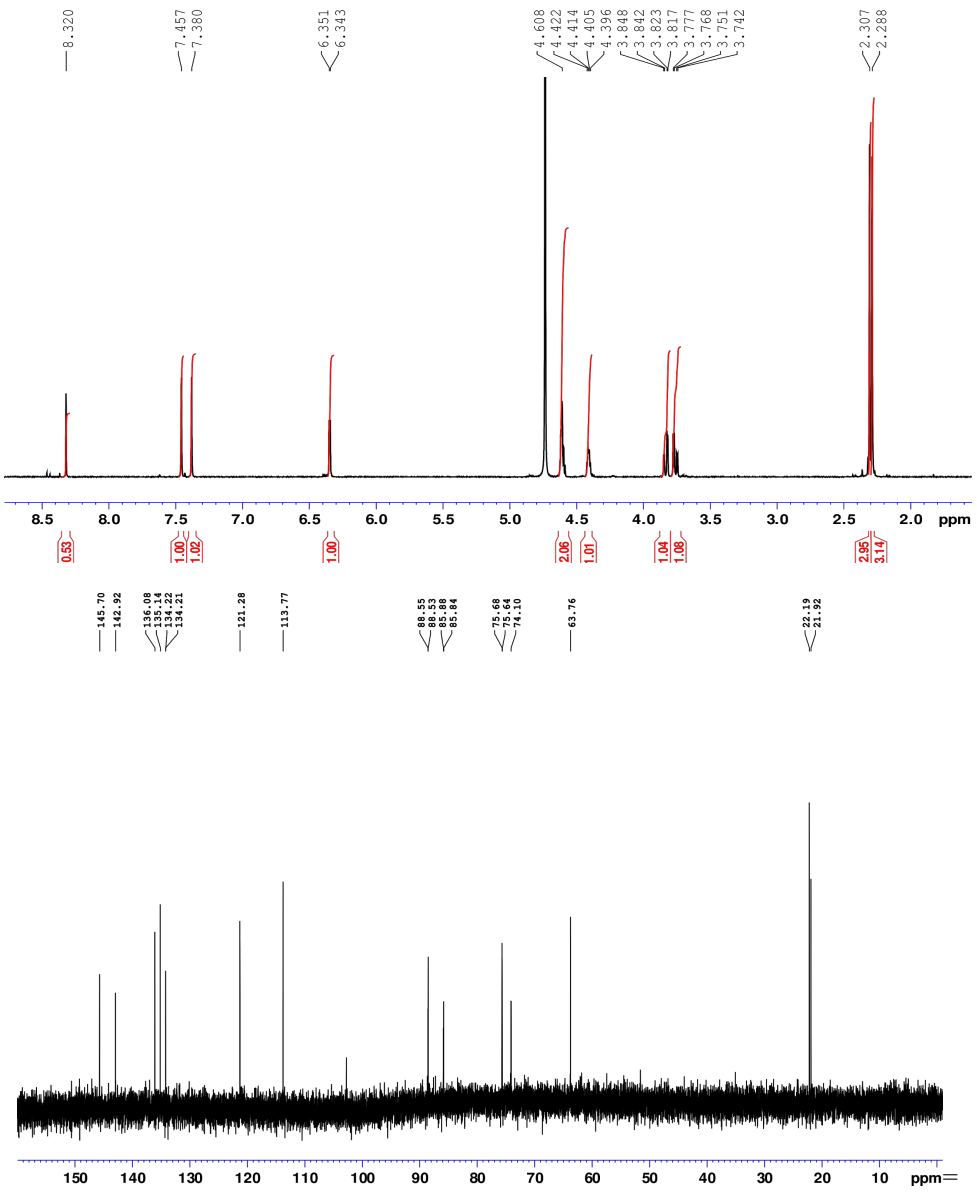
## 7.9. Refinement statistics for UShsMMap02 in complex with PEG

| **Protein** | **UShsMMap02-PEG (DMB soaking)** |
|---|---|
| Wavelength | 1.0000 |
| Resolution range | 40.0 - 2.4 (2.5 - 2.4) |
| Space group | P 1 21 1 |
| Unit cell | 48.412 42.826 63.12 90 94.921 90 |
| Total reflections | 47121 (4370) |
| Unique reflections | 10588 (963) |
| Multiplicity | 4.5 (4.5) |
| Completeness (%) | 98.25 (91.11) |
| Mean I/sigma(I) | 8.43 (0.35) |
| Wilson B-factor | 63.98 |
| R-merge | 0.1269 (3.547) |
| R-meas | 0.1445 (4.023) |
| R-pim | 0.06752 (1.862) |
| CC1/2 | 0.998 (0.169) |
| CC* | 0.999 (0.538) |
| Reflections used in refinement | 10557 (963) |
| Reflections used for R-free | 528 (48) |
| R-work | 0.25 (0.53) |
| R-free | 0.26 (0.51) |
| CC(work) | 0.960 (0.419) |
| CC(free) | 0.959 (0.476) |
| Number of non-hydrogen atoms | 2103 |
| macromolecules | 2096 |
| ligands | 7 |
| Protein residues | 273 |
| RMS(bonds) | 0.016 |
| RMS(angles) | 1.9 |
| Ramachandran favored (%) | 91.14 |
| Ramachandran allowed (%) | 7.01 |
| Ramachandran outliers (%) | 1.85 |
| Rotamer outliers (%) | 18.14 |
| Clashscore | 10.4 |
| Average B-factor | 90.74 |
| macromolecules | 90.81 |
| ligands | 68.22 |
| Number of TLS groups | 3 |

### 7.10. Refinement statistics for UShsMMap02 in complex with PEG

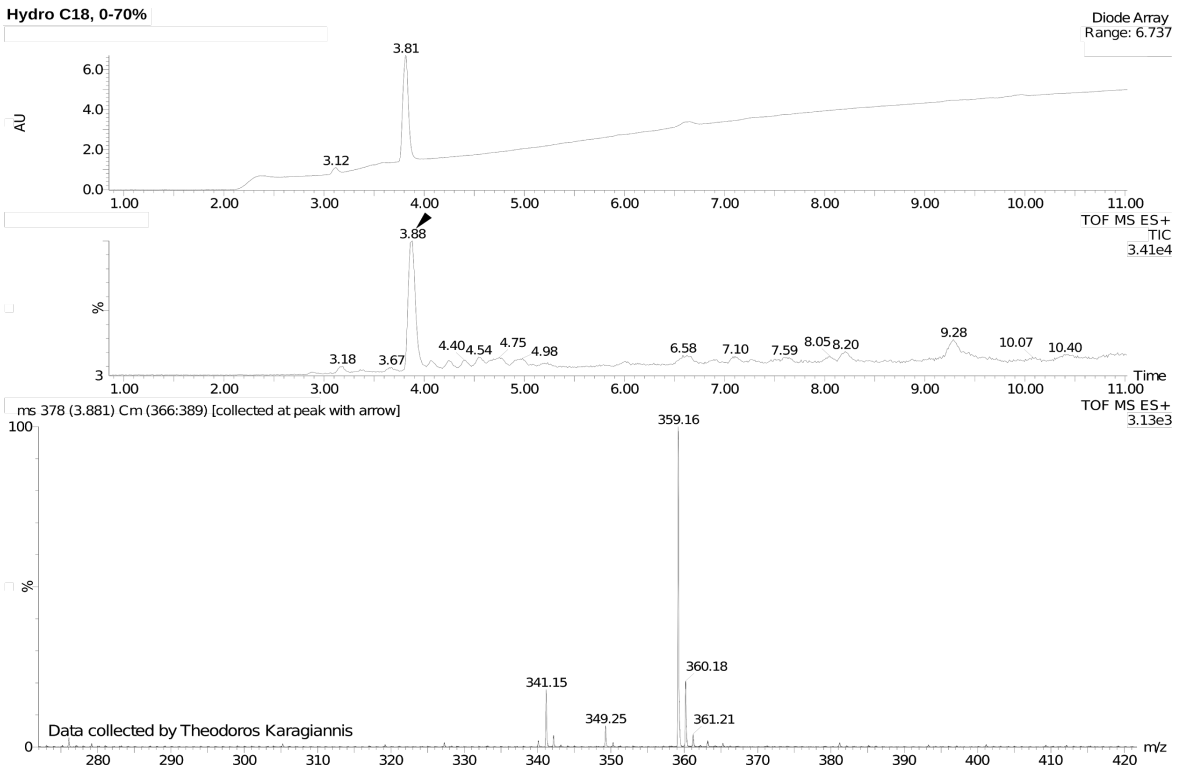
| **Protein** | **UShsMMap02-PEG (αR3P soaking)** |
|---|---|
| Wavelength | 1.000 |
| Resolution range | 39.0 - 2.3 (2.3 - 2.2) |
| Space group | P 1 21 1 |
| Unit cell (Å, °) | 48.437 42.42 62.022 90 93.308 90 |
| Total reflections | 52495 (3969) |
| Unique reflections | 11884 (1079) |
| Multiplicity | 4.4 (3.7) |
| Completeness (%) | 98.04 (90.28) |
| Mean I/sigma(I) | 10.69 (0.72) |
| Wilson B-factor | 51.28 |
| R-merge | 0.08584 (1.61) |
| R-meas | 0.09738 (1.877) |
| R-pim | 0.04495 (0.9398) |
| CC1/2 | 0.999 (0.481) |
| CC* | 1 (0.806) |
| Reflections used in refinement | 11862 (1077) |
| Reflections used for R-free | 594 (54) |
| R-work | 0.24 (0.42) |
| R-free | 0.26 (0.41) |
| CC(work) | 0.961 (0.646) |
| CC(free) | 0.898 (0.519) |
| Number of non-hydrogen atoms | 2094 |
| macromolecules | 2087 |
| ligands | 7 |
| Protein residues | 272 |
| RMS(bonds) | 0.016 |
| RMS(angles) | 2.03 |
| Ramachandran favored (%) | 93.7 |
| Ramachandran allowed (%) | 5.56 |
| Ramachandran outliers (%) | 0.74 |
| Rotamer outliers (%) | 17.78 |
| Clashscore | 18.74 |
| Average B-factor | 72.82 |
| macromolecules | 72.85 |
| ligands | 65.9 |
| Number of TLS groups | 3 |

## 7.11. Compound validation of alpha-ribazole-3'-phosphate by NMR (by Bruce Lichtenstein)[13]




[13] NMR $^{1}H$ (500 MHz, $D_2O$, neutral pH) 2.29 (s, 3H); 2.31 (s, 3H); 3.76 (dd,1H, $J_1$=12.70 Hz, $J_2$=4.40 Hz); 3.83 (dd, 1H, $J_1$=12.70 Hz, $J_2$=3.13 Hz); 4.41 (ap. q, 1H, J=4.46 Hz); 4.61 (ap. m, 2H); 6.35 (d, 1H, J=4.27 Hz); 7.38 (s, 1H); 7.46 (s, 1H); 8.32 (s, 0.5H owing to exchange); $^{13}C$ (125 MHz, $D_2O$, neutral pH)21.915; 22.189; 63.760; 74.104 (J=2.47 Hz); 75.662 (J=4.54 Hz); 85.864 (J=5.45 Hz); 85.864 (J=5.45 Hz); 113.771; 121.279; 134.218 (J=1.82 Hz); 135.218; 136.084; 142.924 (J=2.72 Hz); 145.699.

## 7.12. Compound validation of alpha-ribazole-3'-phosphate by mass spectrometry (LRMS)[14] (by Bruce Lichtenstein)




[14] LRMS: 359.16 m/z (Expected 359.09 m/z).

# Acknowledgements

I would like to express my most sincere gratitude to my supervising professor Birte Höcker for her constant support and motivation. Her valuable guidance accompanied me since I joined her lab as a student assistant.

Besides my advisor, I would like to thank the other members of my thesis advisory committee: professor Andrei Lupas and professor Thilo Stehle for their encouragement and insightful comments. I am particularly grateful for the valuable academic advice I have received from professor Lupas, who has supported and encouraged me to pursue the academic path.

I sincerely thank my labmates for the stimulating discussions and for the amazing working environment. I cannot imagine being part of a better team. I am especially grateful to Bruce Lichtenstein for the fruitful discussions; to Sooruban Shanmugaratnam for the excellent technical assistance; to Francisco Lobos, Noelia Ferruz, Josef Kynast and Sergio Romero for sharing valuable software and scripting tricks. Thanks to all of them, also at a personal level, for their honest friendship. Also thanks to my cousin Liliana Lopez for proofreading this thesis.

I appreciate the support received from professor Holger Dobbek (Humbold University) and professor Stefan Peiffer (Bayreuth University), for giving me the opportunity to make use of the anaerobic glove box facilities in their laboratories. I have a particular appreciation for Dr. Berta M. Martins, who assisted me in the experimental procedures.

Also, many thanks to all my friends, the ones I left behind and the ones I found on my journey to Europe, especially to my husband. Life next to him and our daughter is always full of adventure and joy.

Last but not least, I would like to thank my beloved parents: Mereli Patiño and Rubén Toledo for their unconditional love and support. This work is dedicated to them, they are my greatest example to follow.

# Contributions

The HHsearch calculations that yielded the first insights of homology across α/β-folds were carried out by Jose Farias Rico. The gathered raw data was parsed and filtered by myself and together with Matthias Schwer, I created an automatic approach for structural alignment and visualization for further manual analysis. Based on these preliminary results, I planned and executed the experimental procedures. However, Sooruban Shanmugaratnam carried out cloning of the variants UShsMMap02 and UShsMMap02_H17A and performed the isothermal titration calorimetry experiment described in Chapter 4, Muray coles and Manish Chambey solved the NMR structure of cU3SΔ NMR structure, and Bruce Lichtenstein contributed to the design of experimental procedures explored in Chapter 4 and synthesized the α-ribazole-3-phosphate ligand employed for the co-crystallization experiments.

# Literature


Adams, Paul D., Pavel V. Afonine, Gábor Bunkóczi, Vincent B. Chen, Ian W. Davis, Nathaniel Echols, Jeffrey J. Headd, et al. 2010. "PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution." *Acta Crystallographica. Section D, Biological Crystallography* 66 (Pt 2): 213–21.

Altschul, S. 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research*. https://doi.org/10.1093/nar/25.17.3389.

Alva, Vikram, Michael Remmert, Andreas Biegert, Andrei N. Lupas, and Johannes Söding. 2010. "A Galaxy of Folds." *Protein Science: A Publication of the Protein Society* 19 (1): 124–30.

Andrade, Miguel A., Carolina Perez-Iratxeta, and Chris P. Ponting. 2001. "Protein Repeats: Structures, Functions, and Evolution." *Journal of Structural Biology*. https://doi.org/10.1006/jsbi.2001.4392.

Andreeva, Antonina, and Alexey G. Murzin. 2006. "Evolution of Protein Fold in the Presence of Functional Constraints." *Current Opinion in Structural Biology* 16 (3): 399–408.

Arnold, Frances H. 2015. "The Nature of Chemical Innovation: New Enzymes by Evolution." *Quarterly Reviews of Biophysics* 48 (4): 404–10.

Baalsrud, Helle Tessand, Ole Kristian Tørresen, Monica Hongrø Solbakken, Walter Salzburger, Reinhold Hanel, Kjetill S. Jakobsen, and Sissel Jentoft. 2017. "De Novo Gene Evolution of Antifreeze Glycoproteins in Codfishes Revealed by Whole Genome Sequence Data." *Molecular Biology and Evolution*, December. https://doi.org/10.1093/molbev/msx311.

Banerjee, R. 1997. "The Yin-Yang of Cobalamin Biochemistry." *Chemistry & Biology* 4 (3): 175–86.

———. 2001. "Radical Peregrinations Catalyzed by Coenzyme B12-Dependent Enzymes." *Biochemistry* 40 (21): 6191–98.

Banerjee, Ruma. 1999. *Chemistry and Biochemistry of B12*. John Wiley & Sons.

Bennett, M. J., S. Choe, and D. Eisenberg. 1994. "Domain Swapping: Entangling Alliances between Proteins." *Proceedings of the National Academy of Sciences of the United States of America* 91 (8): 3127–31.

Bennett, M. J., M. P. Schlunegger, and D. Eisenberg. 1995. "3D Domain Swapping: A Mechanism for Oligomer Assembly." *Protein Science: A Publication of the Protein Society* 4 (12): 2455–68.

Berman, M. H., and A. C. Frazer. 1992. "Importance of Tetrahydrofolate and ATP in the Anaerobic O-Demethylation Reaction for Phenylmethylethers." *Applied and Environmental Microbiology* 58 (3): 925–31.

Bharat, Tanmay A. M., Simone Eisenbeis, Kornelius Zeth, and Birte Höcker. 2008. "A Beta Alpha-Barrel Built by the Combination of Fragments from Different Folds." *Proceedings of the National Academy of Sciences of the United States of America* 105 (29): 9942–47.

Bornscheuer, Uwe T. 2018. "The Fourth Wave of Biocatalysis Is Approaching." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. https://doi.org/10.1098/rsta.2017.0063.

Brown, Kenneth L. 2005. "Chemistry and Enzymology of Vitamin B12." *ChemInform*. https://doi.org/10.1002/chin.200537278.

Brown, K. L., and J. M. Hakimi. 1986. "Heteronuclear NMR Studies of Cobalamins. 4. .alpha.-Ribazole-3'-Phosphate and the Nucleotide Loop of Base-on Cobalamins." *Journal of the American Chemical Society* 108 (3): 496–503.

Brown, K. L., and X. Zou. 1999. "Thermolysis of Coenzymes B12 at Physiological Temperatures: Activation Parameters for Cobalt-Carbon Bond Homolysis and a Quantitative Analysis of the Perturbation of the Homolysis Equilibrium by the Ribonucleoside Triphosphate Reductase from Lactobacillus Leichmannii." *Journal of Inorganic Biochemistry* 77 (3-4): 185–95.

Bystroff, Christopher, and Anders Krogh. n.d. "Hidden Markov Models for Prediction of Protein Features." *Protein Structure Prediction, Second Edition*. https://doi.org/10.1385/1-59745-574-1:173.

Cameron, A. D. 1997. "Crystal Structure of Human Glyoxalase I_evidence for Gene Duplication and 3D Domain Swapping." *The EMBO Journal*. https://doi.org/10.1093/emboj/16.12.3386.

Carlomagno, T., M. Maurer, M. Sattler, M. G. Schwendinger, S. J. Glaser, and C. Griesinger. 1996. "PLUSH TACSY: Homonuclear Planar TACSY with Two-Band Selective Shaped Pulses Applied to

C?,C? Transfer and C?,Caromatic Correlations." *Journal of Biomolecular NMR*. https://doi.org/10.1007/bf00211162.
Chandonia, John-Marc, Naomi K. Fox, and Steven E. Brenner. 2019. "SCOPe: Classification of Large Macromolecular Structures in the Structural Classification of Proteins-Extended Database." *Nucleic Acids Research* 47 (D1): D475–81.
Cheng, Hua, R. Dustin Schaeffer, Yuxing Liao, Lisa N. Kinch, Jimin Pei, Shuoyong Shi, Bong-Hyun Kim, and Nick V. Grishin. 2014. "ECOD: An Evolutionary Classification of Protein Domains." *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1003926.
Cheng, Zhuo, Haruki Yamamoto, and Carl E. Bauer. 2016. "Cobalamin's (Vitamin B 12 ) Surprising Function as a Photoreceptor." *Trends in Biochemical Sciences*. https://doi.org/10.1016/j.tibs.2016.05.002.
Chen, Kai, Xiongyi Huang, S. B. Jennifer Kan, Ruijie K. Zhang, and Frances H. Arnold. 2018. "Enzymatic Construction of Highly Strained Carbocycles." *Science*. https://doi.org/10.1126/science.aar4239.
Christiansen, N., B. K. Ahring, G. Wohlfarth, and G. Diekert. 1998. "Purification and Characterization of the 3-Chloro-4-Hydroxy-Phenylacetate Reductive Dehalogenase of Desulfitobacterium Hafniense." *FEBS Letters* 436 (2): 159–62.
Cody, Wayne L. 2007. "Greene's Protective Groups in Organic Synthesis. Fourth Edition.By Peter G. M. Wuts and Theodora W. Greene. John Wiley & Sons, Inc., Hoboken, NJ. 2006. Xvii 1082 Pp. 16 × 24 Cm. ISBN 0-471-69754-0. $94.95." *Journal of Medicinal Chemistry*. https://doi.org/10.1021/jm068051f.
Cole, Allwyn G., Laurie M. Yoder, Joseph J. Shiang, Neil A. Anderson, Larry A. Walker, Mark M. Banaszak Holl, and Roseanne J. Sension. 2002. "Time-Resolved Spectroscopic Studies of B12Coenzymes: A Comparison of the Primary Photolysis Mechanism in Methyl-, Ethyl-,n-Propyl-, and 5'-Deoxyadenosylcobalamin." *Journal of the American Chemical Society*. https://doi.org/10.1021/ja011628s.
Coles, M., V. Truffault, S. Eisenbeis, W. Proffitt, J. Meiler, and B. Hocker. 2012. "Computational Design of an Eight-Stranded (beta/alpha)-Barrel from Fragments of Different Folds." https://doi.org/10.2210/pdb2lle/pdb.
Copeland, Robert A., Michael E. Solomon, and Victoria M. Richon. 2009. "Protein Methyltransferases as a Target Class for Drug Discovery." *Nature Reviews Drug Discovery*. https://doi.org/10.1038/nrd2974.
Copley, S. D. 1998. "Microbial Dehalogenases: Enzymes Recruited to Convert Xenobiotic Substrates." *Current Opinion in Chemical Biology* 2 (5): 613–17.
Cornella, Josep, Cayetana Zarate, and Ruben Martin. 2014. "Metal-Catalyzed Activation of Ethers via C–O Bond Cleavage: A New Strategy for Molecular Diversity." *Chem. Soc. Rev.* https://doi.org/10.1039/c4cs00206g.
Cornilescu, Gabriel, Frank Delaglio, and Ad Bax. 1999. "Journal of Biomolecular NMR." https://doi.org/10.1023/a:1008392405740.
Csaba, Gergely, Fabian Birzele, and Ralf Zimmer. 2009. "Systematic Comparison of SCOP and CATH: A New Gold Standard for Protein Structure Analysis." *BMC Structural Biology* 9 (April): 23.
Cuneo, Matthew J., Lorena S. Beese, and Homme W. Hellinga. 2008. "Ligand-Induced Conformational Changes in a Thermophilic Ribose-Binding Protein." *BMC Structural Biology*. https://doi.org/10.1186/1472-6807-8-50.
Dagert, M., and S. D. Ehrlich. 1979. "Prolonged Incubation in Calcium Chloride Improves the Competence of Escherichia Coli Cells." *Gene* 6 (1): 23–28.
Dahiyat, B. I., C. A. Sarisky, and S. L. Mayo. 1997. "De Novo Protein Design: Towards Fully Automated Sequence Selection." *Journal of Molecular Biology* 273 (4): 789–96.
Das, Rhiju, and David Baker. 2008. "Macromolecular Modeling with Rosetta." *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev.biochem.77.062906.171838.
Dassanayake, Rohan S., Mohamed M. Farhath, Jacob T. Shelley, Soumitra Basu, and Nicola E. Brasch. 2016. "Kinetic Studies on the Reaction of cob(II)alamin with Hypochlorous Acid: Evidence for One Electron Oxidation of the Metal Center and Corrin Ring Destruction." *Journal of Inorganic Biochemistry* 163 (October): 81–87.
Dawson, Natalie L., Tony E. Lewis, Sayoni Das, Jonathan G. Lees, David Lee, Paul Ashford, Christine A. Orengo, and Ian Sillitoe. 2017. "CATH: An Expanded Resource to Predict Protein Function through Structure and Sequence." *Nucleic Acids Research* 45 (D1): D289–95.

Day, Ryan, David A. C. Beck, Roger S. Armen, and Valerie Daggett. 2003. “A Consensus View of Fold Space: Combining SCOP, CATH, and the Dali Domain Dictionary.” *Protein Science: A Publication of the Protein Society* 12 (10): 2150–60.
Debussche, L., D. Thibaut, B. Cameron, J. Crouzet, and F. Blanche. 1993. “Biosynthesis of the Corrin Macrocycle of Coenzyme B12 in Pseudomonas Denitrificans.” *Journal of Bacteriology*. https://doi.org/10.1128/jb.175.22.7430-7440.1993.
DeWeerd, K. A., A. Saxena, D. P. Nagle Jr, and J. M. Suflita. 1988. “Metabolism of the 18O-Methoxy Substituent of 3-Methoxybenzoic Acid and Other Unlabeled Methoxybenzoic Acids by Anaerobic Bacteria.” *Applied and Environmental Microbiology* 54 (5): 1237–42.
Diercks, Tammo, Murray Coles, and Horst Kessler. 1999. “Journal of Biomolecular NMR.” https://doi.org/10.1023/a:1008367912535.
Diercks, Tammo, Mark Daniels, and Robert Kaptein. 2005. “Extended Flip-Back Schemes for Sensitivity Enhancement in Multidimensional HSQC-Type Out-and-Back Experiments.” *Journal of Biomolecular NMR*. https://doi.org/10.1007/s10858-005-3868-4.
Dolphin, D. 1971. “[205] Preparation of the Reduced Forms of Vitamin B12 and of Some Analogs of the Vitamin B12 Coenzyme Containing a Cobalt-Carbon Bond.” *Vitamins and Coenzymes*. https://doi.org/10.1016/s0076-6879(71)18006-8.
Eck, R. V., and M. O. Dayhoff. 1966. “Evolution of the Structure of Ferredoxin Based on Living Relics of Primitive Amino Acid Sequences.” *Science*. https://doi.org/10.1126/science.152.3720.363.
Eisenbeis, Simone, William Proffitt, Murray Coles, Vincent Truffault, Sooruban Shanmugaratnam, Jens Meiler, and Birte Höcker. 2012. “Potential of Fragment Recombination for Rational Design of Proteins.” *Journal of the American Chemical Society*. https://doi.org/10.1021/ja211657k.
ElGamacy, M., V. Truffault, H. Zhu, and M. Coles. 2019. “Solution NMR Ensemble for Human Ubiquitin at 298K Compiled Using the CoMAND Method.” https://doi.org/10.2210/pdb6qf8/pdb.
Evans, Philip. 2006. “Scaling and Assessment of Data Quality.” *Acta Crystallographica. Section D, Biological Crystallography* 62 (Pt 1): 72–82.
Farías-Rico, José Arcadio, Steffen Schmidt, and Birte Höcker. 2014. “Evolutionary Relationship of Two Ancient Protein Superfolds.” *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.1579.
Farnberger, Judith E., Nina Richter, Katharina Hiebler, Sarah Bierbaumer, Mathias Pickl, Wolfgang Skibar, Ferdinand Zepeck, and Wolfgang Kroutil. 2018. “Biocatalytic Methylation and Demethylation via a Shuttle Catalysis Concept Involving Corrinoid Proteins.” *Communications Chemistry*. https://doi.org/10.1038/s42004-018-0083-2.
Fidler, David R., Sarah E. Murphy, Katherine Courtis, Pantelis Antonoudiou, Rana El-Tohamy, Jonathan Ient, and Timothy P. Levine. 2016. “Using HHsearch to Tackle Proteins of Unknown Function: A Pilot Study with PH Domains.” *Traffic* 17 (11): 1214–26.
Finke, Richard G., and Benjamin P. Hay. 1984. “Thermolysis of Adenosylcobalamin: A Product, Kinetic, and Cobalt-Carbon (C5’) Bond Dissociation Energy Study.” *Inorganic Chemistry*. https://doi.org/10.1021/ic00188a002.
Fortian, Arola, David Castaño, Gabriel Ortega, Ana Laín, Miquel Pons, and Oscar Millet. 2009. “Uroporphyrinogen III Synthase Mutations Related to Congenital Erythropoietic Porphyria Identify a Key Helix for Protein Stability.” *Biochemistry* 48 (2): 454–61.
Fredriksson, Anna, and Sharon Stone-Elander. 2002. “Rapid Microwave-Assisted Cleavage of Methyl Phenyl Ethers: New Method for Synthesizing Desmethyl Precursors and for Removing Protecting Groups.” *Journal of Labelled Compounds and Radiopharmaceuticals*. https://doi.org/10.1002/jlcr.577.
Ghosh, Arghya Pratim, Abdullah Al Mamun, Piotr Lodowski, Maria Jaworska, and Pawel M. Kozlowski. 2018. “Mechanism of the Photo-Induced Activation of CoC Bond in Methylcobalamin-Dependent Methionine Synthase.” *Journal of Photochemistry and Photobiology. B, Biology* 189 (December): 306–17.
Gilbert, Walter. 1986. “Origin of Life: The RNA World.” *Nature*. https://doi.org/10.1038/319618a0.
Goldsmith, Moshe, and Dan S. Tawfik. 2012. “Directed Enzyme Evolution: Beyond the Low-Hanging Fruit.” *Current Opinion in Structural Biology* 22 (4): 406–12.
Goldstein, Richard A. 2008. “The Structure of Protein Evolution and the Evolution of Protein Structure.” *Current Opinion in Structural Biology*. https://doi.org/10.1016/j.sbi.2008.01.006.
Grishin, Nick V. 2001. “Fold Change in Evolution of Protein Structures.” *Journal of Structural Biology*. https://doi.org/10.1006/jsbi.2001.4335.

Gronenborn, Angela M. 2009. “Protein Acrobatics in Pairs—dimerization via Domain Swapping.” *Current Opinion in Structural Biology*. https://doi.org/10.1016/j.sbi.2008.12.002.
Hadley, C., and D. T. Jones. 1999. “A Systematic Comparison of Protein Structure Classifications: SCOP, CATH and FSSP.” *Structure* 7 (9): 1099–1112.
Hall, D. A., C. W. Vander Kooi, C. N. Stasik, S. Y. Stevens, E. R. Zuiderweg, and R. G. Matthews. 2001. “Mapping the Interactions between Flavodoxin and Its Physiological Partners Flavodoxin Reductase and Cobalamin-Dependent Methionine Synthase.” *Proceedings of the National Academy of Sciences of the United States of America* 98 (17): 9521–26.
Harbury, P. B., J. J. Plecs, B. Tidor, T. Alber, and P. S. Kim. 1998. “RH4 DESIGNED RIGHT-HANDED COILED COIL TETRAMER.” https://doi.org/10.2210/pdb1rh4/pdb.
Hart, William E. S., Leigh Aldous, and Jason B. Harper. 2017. “Cleavage of Ethers in an Ionic Liquid. Enhancement, Selectivity and Potential Application.” *Organic & Biomolecular Chemistry*. https://doi.org/10.1039/c7ob01096f.
Hay, Benjamin P., and Richard G. Finke. 1987. “Thermolysis of the Cobalt-Carbon Bond in Adenosylcorrins. 3. Quantification of the Axial Base Effect in Adenosylcobalamin by the Synthesis and Thermolysis of Axial Base-Free Adenosylcobinamide. Insights into the Energetics of Enzyme-Assisted Cobalt-Carbon Bond Homolysis.” *Journal of the American Chemical Society*. https://doi.org/10.1021/ja00260a011.
Hilhorst, Ellen, Atef S. Iskander, Tjoe B R, and Upendra K. Pandit. 1993. “Model Studies of the Cobalamin-Dependent Methionine Synthase Reaction.” *Tetrahedron Letters*. https://doi.org/10.1016/s0040-4039(00)60542-7.
Höcker, Birte. 2014. “Design of Proteins from Smaller Fragments-Learning from Evolution.” *Current Opinion in Structural Biology* 27 (August): 56–62.
Höcker, Birte, Jörg Claren, and Reinhard Sterner. 2004. “Mimicking Enzyme Evolution by Generating New (betaalpha)8-Barrels from (betaalpha)4-Half-Barrels.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (47): 16448–53.
Hodgkin, Dorothy Crowfoot. 1958. “X-Ray Analysis and the Structure of Vitamin B12.” *Fortschritte Der Chemie Organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products / Progrès Dans La Chimie Des Substances Organiques Naturelles*. https://doi.org/10.1007/978-3-7091-7162-2_4.
Hoffmann, B., R. Konrat, H. Bothe, W. Buckel, and B. Kräutler. 1999. “Structure and Dynamics of the B12-Binding Subunit of Glutamate Mutase from Clostridium Cochlearium.” *European Journal of Biochemistry / FEBS* 263 (1): 178–88.
Holland, Timothy A., Stella Veretnik, Ilya N. Shindyalov, and Philip E. Bourne. 2006. “Partitioning Protein Structures into Domains: Why Is It so Difficult?” *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2006.05.060.
Holm, Liisa, and Laura M. Laakso. 2016. “Dali Server Update.” *Nucleic Acids Research* 44 (W1): W351–55.
Hönig, Moritz, Philipp Sondermann, Nicholas J. Turner, and Erick M. Carreira. 2017. “Enantioselektive Chemo- Und Biokatalyse: Partner in Der Retrosynthese.” *Angewandte Chemie*. https://doi.org/10.1002/ange.201612462.
Ho, Steffan N., Henry D. Hunt, Robert M. Horton, Jeffrey K. Pullen, and Larry R. Pease. 1989. “Site-Directed Mutagenesis by Overlap Extension Using the Polymerase Chain Reaction.” *Gene*. https://doi.org/10.1016/0378-1119(89)90358-2.
Huang, Po-Ssu, Scott E. Boyken, and David Baker. 2016. “The Coming of Age of de Novo Protein Design.” *Nature* 537 (7620): 320–27.
Huennekens, F. M., P. M. Digirolamo, K. Fujii, G. B. Henderson, D. W. Jacobsen, V. G. Neef, and J. I. Rader. 1974. “Folic Acid and Vitamin B12: Transport and Conversion to Coenzyme Forms.” *Advances in Enzyme Regulation*. https://doi.org/10.1016/0065-2571(74)90011-9.
Janssen, D. B., J. E. Oppentocht, and G. J. Poelarends. 2001. “Microbial Dehalogenation.” *Current Opinion in Biotechnology* 12 (3): 254–58.
Jaskólski, M. 2001. “3D Domain Swapping, Protein Oligomerization, and Amyloid Formation.” *Acta Biochimica Polonica* 48 (4): 807–27.
Jorda, Julien, Bin Xue, Vladimir N. Uversky, and Andrey V. Kajava. 2010. “Protein Tandem Repeats - the More Perfect, the Less Structured.” *The FEBS Journal* 277 (12): 2673–82.

Jost, Marco, Jésus Fernández-Zapata, María Carmen Polanco, Juan Manuel Ortiz-Guerrero, Percival Yang-Ting Chen, Gyunghoon Kang, S. Padmanabhan, Montserrat Elías-Arnanz, and Catherine L. Drennan. 2015. “Structural Basis for Gene Regulation by a B12-Dependent Photoreceptor.” *Nature* 526 (7574): 536–41.

Joyce, Gerald F. 1989. “RNA Evolution and the Origins of Life.” *Nature*. https://doi.org/10.1038/338217a0.

Joyce, G. F. 1998. “Nucleic Acid Enzymes: Playing with a Fuller Deck.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (11): 5845–47.

Jung, J., and B. Lee. 2001. “Circularly Permuted Proteins in the Protein Structure Database.” *Protein Science: A Publication of the Protein Society* 10 (9): 1881–86.

Kabsch, Wolfgang. 2010. “XDS.” *Acta Crystallographica. Section D, Biological Crystallography* 66 (Pt 2): 125–32.

Karlusich, J. J. Pierella, J. J. Pierella Karlusich, A. F. Lodeyro, and N. Carrillo. 2014. “The Long Goodbye: The Rise and Fall of Flavodoxin during Plant Evolution.” *Journal of Experimental Botany*. https://doi.org/10.1093/jxb/eru273.

Kasmi, A. el, S. Rajasekharan, and S. W. Ragsdale. 1994. “Anaerobic Pathway for Conversion of the Methyl Group of Aromatic Methyl Ethers to Acetic Acid by Clostridium Thermoaceticum.” *Biochemistry* 33 (37): 11217–24.

Khoury, George A., James Smadbeck, Chris A. Kieslich, and Christodoulos A. Floudas. 2014. “Protein Folding and de Novo Protein Design for Biotechnological Applications.” *Trends in Biotechnology* 32 (2): 99–109.

Kiedrowski, G□nter von, and G□nter von Kiedrowski. 1986. “A Self-Replicating Hexadeoxynucleotide.” *Angewandte Chemie International Edition in English*. https://doi.org/10.1002/anie.198609322.

Kim, Jihoe, Carmen Gherasim, and Ruma Banerjee. 2008. “Decyanation of Vitamin B12 by a Trafficking Chaperone.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (38): 14551–54.

Kozlowski, Pawel M., Brady D. Garabato, Piotr Lodowski, and Maria Jaworska. 2016. “Photolytic Properties of Cobalamins: A Theoretical Perspective.” *Dalton Transactions* 45 (11): 4457–70.

Krissinel, E., and K. Henrick. 2004. “Secondary-Structure Matching (SSM), a New Tool for Fast Protein Structure Alignment in Three Dimensions.” *Acta Crystallographica. Section D, Biological Crystallography* 60 (Pt 12 Pt 1): 2256–68.

Krone, Ute E., Rudolf K. Thauer, and Harry P. C. Hogenkamp. 1989. “Reductive Dehalogenation of Chlorinated C1-Hydrocarbons Mediated by Corrinoids.” *Biochemistry*. https://doi.org/10.1021/bi00437a057.

Kuhlman, Brian, Gautam Dantas, Gregory C. Ireton, Gabriele Varani, Barry L. Stoddard, and David Baker. 2003. “Design of a Novel Globular Protein Fold with Atomic-Level Accuracy.” *Science* 302 (5649): 1364–68.

Law, Brian J. C., Matthew R. Bennett, Mark L. Thompson, Colin Levy, Sarah A. Shepherd, David Leys, and Jason Micklefield. 2016. “Effects of Active-Site Modification and Quaternary Structure on the Regioselectivity of Catechol-O -Methyltransferase.” *Angewandte Chemie International Edition*. https://doi.org/10.1002/anie.201508287.

Lee, Danbi, Hye Lin Park, Sang-Won Lee, Seong Hee Bhoo, and Man-Ho Cho. 2017. “Biotechnological Production of Dimethoxyflavonoids Using a Fusion Flavonoid O-Methyltransferase Possessing Both 3′- and 7-O-Methyltransferase Activities.” *Journal of Natural Products*. https://doi.org/10.1021/acs.jnatprod.6b01164.

Lezius, A. G., and H. A. Barker. 1965. “CORRINOID COMPOUNDS OF METHANOBACILLUS OMELIANSKII. I. FRACTIONATION OF THE CORRINOID COMPOUNDS AND IDENTIFICATION OF FACTOR 3 AND FACTOR 3 COENZYME.” *Biochemistry* 4 (March): 510–18.

Liscombe, David K., Gordon V. Louie, and Joseph P. Noel. 2012. “Architectures, Mechanisms and Molecular Evolution of Natural Product Methyltransferases.” *Natural Product Reports* 29 (10): 1238–50.

Liu, Yanshun, and David Eisenberg. 2002. “3D Domain Swapping: As Domains Continue to Swap.” *Protein Science: A Publication of the Protein Society* 11 (6): 1285–99.

Lupas, A. N., C. P. Ponting, and R. B. Russell. 2001. “On the Evolution of Protein Folds: Are Similar Motifs in Different Protein Folds the Result of Convergence, Insertion, or Relics of an Ancient Peptide World?” *Journal of Structural Biology* 134 (2-3): 191–203.

Mancia, Filippo, and Philip R. Evans. 1998. “Conformational Changes on Substrate Binding to Methylmalonyl CoA Mutase and New Insights into the Free Radical Mechanism.” *Structure*. https://doi.org/10.1016/s0969-2126(98)00073-2.

Marcotte, E. M., M. Pellegrini, T. O. Yeates, and D. Eisenberg. 1999. “A Census of Protein Repeats.” *Journal of Molecular Biology* 293 (1): 151–60.

Marsh, Claire, and Beverly Drennan. 1976. “Ego States and Egogram Therapy.” *Transactional Analysis Bulletin*. https://doi.org/10.1177/036215377600600207.

Mathews, M. A. A. 2001. “Crystal Structure of Human Uroporphyrinogen III Synthase.” *The EMBO Journal*. https://doi.org/10.1093/emboj/20.21.5832.

Matthews, Rowena G. 2009. “Cobalamin- and Corrinoid-Dependent Enzymes.” *Metal Ions in Life Sciences* 6 (January): 53–114.

Matthews, Rowena G., Markos Koutmos, and Supratim Datta. 2008. “Cobalamin-Dependent and Cobamide-Dependent Methyltransferases.” *Current Opinion in Structural Biology* 18 (6): 658–66.

McLachlan, A. D. 1979. “Three-Fold Structural Pattern in the Soybean Trypsin Inhibitor (Kunitz).” *Journal of Molecular Biology* 133 (4): 557–63.

McOmie, J. F. W., M. L. Watts, and D. E. West. 1968. “Demethylation of Aryl Methyl Ethers by Boron Tribromide.” *Tetrahedron*. https://doi.org/10.1016/0040-4020(68)88130-x.

Miller, Evelyn, Gert Wohlfarth, and G. Diekert. 1997. “Studies on Tetrachloroethene Respiration in Dehalospirillum Multivorans.” *Archives of Microbiology*. https://doi.org/10.1007/s002030050399.

Minot, George R. 1926. “TREATMENT OF PERNICIOUS ANEMIA BY A SPECIAL DIET.” *JAMA: The Journal of the American Medical Association*. https://doi.org/10.1001/jama.1926.02680070016005.

Moynie, Lucille, Robert Schnell, Stephen A. McMahon, Tatyana Sandalova, Wassila Abdelli Boulkerou, Jason W. Schmidberger, Magnus Alphey, et al. 2013. “The AEROPATH Project Targeting Pseudomonas Aeruginosa: Crystallographic Studies for Assessment of Potential Targets in Early-Stage Drug Discovery.” *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications* 69 (Pt 1): 25–34.

Murphy, William P. 1932. “THE PARENTERAL USE OF LIVER EXTRACT IN PERNICIOUS ANEMIA.” *Journal of the American Medical Association*. https://doi.org/10.1001/jama.1932.02730390005002.

Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia. 1995. “SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures.” *Journal of Molecular Biology* 247 (4): 536–40.

Naidu, D., and S. W. Ragsdale. 2001. “Characterization of a Three-Component Vanillate O-Demethylase from Moorella Thermoacetica.” *Journal of Bacteriology* 183 (11): 3276–81.

Neumann, Anke, Gert Wohlfarth, and G. Diekert. 1995. “Properties of Tetrachloroethene and Trichloroethene Dehalogenase of Dehalospirillum Multivorans.” *Archives of Microbiology*. https://doi.org/10.1007/s002030050204.

Ni, S., J. K. Fredrickson, and L. Xun. 1995. “Purification and Characterization of a Novel 3-Chlorobenzoate-Reductive Dehalogenase from the Cytoplasmic Membrane of Desulfomonile Tiedjei DCB-1.” *Journal of Bacteriology*. https://doi.org/10.1128/jb.177.17.5135-5139.1995.

Orengo, C. A., A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells, and J. M. Thornton. 1997. “CATH – a Hierarchic Classification of Protein Domain Structures.” *Structure*. https://doi.org/10.1016/s0969-2126(97)00260-8.

Pas, B. A. van de, H. Smidt, W. R. Hagen, J. van der Oost, G. Schraa, A. J. Stams, and W. M. de Vos. 1999. “Purification and Molecular Characterization of Ortho-Chlorophenol Reductive Dehalogenase, a Key Enzyme of Halorespiration in Desulfitobacterium Dehalogenans.” *The Journal of Biological Chemistry* 274 (29): 20287–92.

Pasquini, Chiara, and Mauro Bassetti. 2010. “One-Pot Desilylation/Dimerization of Terminal Alkynes by Ruthenium and Acid-Promoted (RAP) Catalysis.” *Advanced Synthesis & Catalysis*. https://doi.org/10.1002/adsc.201000347.

Patel, Ramesh N. 2011. “Biocatalysis: Synthesis of Key Intermediates for Development of Pharmaceuticals.” *ACS Catalysis*. https://doi.org/10.1021/cs200219b.

Payne, Karl Ap, Carolina P. Quezada, Karl Fisher, Mark S. Dunstan, Fraser A. Collins, Hanno Sjuts, Colin Levy, Sam Hay, Stephen Ej Rigby, and David Leys. 2015. “Reductive Dehalogenase Structure Suggests a Mechanism for B12-Dependent Dehalogenation.” *Nature* 517 (7535): 513–16.

Peng, Jian, Kuo-Chun Tang, Kaitlin McLoughlin, Yang Yang, Danika Forgach, and Roseanne J. Sension. 2010. “Ultrafast Excited-State Dynamics and Photolysis in Base-off B12 Coenzymes and Analogues: Absence of the Trans-Nitrogenous Ligand Opens a Channel for Rapid Nonradiative Decay.” *The Journal of Physical Chemistry. B* 114 (38): 12398–405.

Ragsdale, Stephen W. 2008. “Enzymology of the Wood-Ljungdahl Pathway of Acetogenesis.” *Annals of the New York Academy of Sciences*. https://doi.org/10.1196/annals.1419.015.

Richter, Oscar. 1933. “FURTHER OBSERVATIONS ON THE TREATMENT OF PERNICIOUS ANEMIA WITH PARENTERAL HORSE LIVER EXTRACT*.” *Annals of Internal Medicine*. https://doi.org/10.7326/0003-4819-7-3-353.

Riechmann, Lutz, Isabelle Lavenir, Stephanie de Bono, and Greg Winter. 2005. “Folding and Stability of a Primitive Protein.” *Journal of Molecular Biology* 348 (5): 1261–72.

Robscheit-Robbins, F. S., and G. H. Whipple. 1929. “BLOOD REGENERATION IN SEVERE ANEMIA : XIV. A LIVER FRACTION POTENT IN PERNICIOUS ANEMIA FED ALONE AND COMBINED WITH WHOLE LIVER, LIVER ASH AND FRESH BILE.” *The Journal of Experimental Medicine* 49 (2): 215–27.

Roth, A., and R. R. Breaker. 1998. “An Amino Acid as a Cofactor for a Catalytic Polynucleotide.” *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.95.11.6027.

Röthlisberger, Daniela, Olga Khersonsky, Andrew M. Wollacott, Lin Jiang, Jason DeChancie, Jamie Betker, Jasmine L. Gallaher, et al. 2008. “Kemp Elimination Catalysts by Computational Enzyme Design.” *Nature* 453 (7192): 190–95.

Salnikov, Denis S., Radu Silaghi-Dumitrescu, Sergei V. Makarov, Rudi van Eldik, and Gerry R. Boss. 2011. “Cobalamin Reduction by Dithionite. Evidence for the Formation of a Six-Coordinate cobalamin(II) Complex.” *Dalton Transactions* 40 (38): 9831–34.

Salnikov, D. S., and S. V. Makarov. 2019. “Kinetics and Mechanism of the Reaction of Cyanocobalamin with Potassium Hydroxide in Non-Aqueous Media.” *New Journal of Chemistry*. https://doi.org/10.1039/c9nj01361j.

Schaeffer, R. Dustin, Amanda L. Jonsson, Andrew M. Simms, and Valerie Daggett. 2011. “Generation of a Consensus Protein Domain Dictionary.” *Bioinformatics* 27 (1): 46–54.

Schnell, R., and G. Schneider. 2013. “Crystal Structure of HemD (PA5259) from Pseudomonas Aeruginosa (PAO1) at 2.22 A Resolution.” https://doi.org/10.2210/pdb4es6/pdb.

Schubert, Heidi L., John D. Phillips, Annie Heroux, and Christopher P. Hill. 2008. “Structure and Mechanistic Implications of a Uroporphyrinogen III Synthase-Product Complex.” *Biochemistry* 47 (33): 8648–55.

Schulze, Bettina, Bernhard Vogler, and Paul Renz. 1998. “Biosynthesis of Vitamin B12 in Anaerobic Bacteria . Experiments with Eubacterium Limosum on the Transformation of 5-Hydroxy-6-Methyl-Benzimidazole, Its Nucleoside, Its Cobamide, and of 5-Hydroxybenzimidazolylcobamide in Vitamin B12.” *European Journal of Biochemistry*. https://doi.org/10.1046/j.1432-1327.1998.2540620.x.

Schwieters, Charles D., John J. Kuszewski, Nico Tjandra, and G. Marius Clore. 2003. “The Xplor-NIH NMR Molecular Structure Determination Package.” *Journal of Magnetic Resonance*. https://doi.org/10.1016/s1090-7807(02)00014-9.

Schwieters, C., J. Kuszewski, and G. Mariusclore. 2006. “Using Xplor–NIH for NMR Molecular Structure Determination.” *Progress in Nuclear Magnetic Resonance Spectroscopy*. https://doi.org/10.1016/j.pnmrs.2005.10.001.

Selva, Maurizio, and Alvise Perosa. 2008. “Green Chemistry Metrics: A Comparative Evaluation of Dimethyl Carbonate, Methyl Iodide, Dimethyl Sulfate and Methanol as Methylating Agents.” *Green Chemistry*. https://doi.org/10.1039/b713985c.

Shanmugaratnam, Sooruban, Simone Eisenbeis, and Birte Höcker. 2012. “A Highly Stable Protein Chimera Built from Fragments of Different Folds.” *Protein Engineering, Design & Selection: PEDS* 25 (11): 699–703.

Sheldon, Roger A., and John M. Woodley. 2018. “Role of Biocatalysis in Sustainable Chemistry.” *Chemical Reviews*. https://doi.org/10.1021/acs.chemrev.7b00203.

Söding, Johannes. 2005. “Protein Homology Detection by HMM-HMM Comparison.” *Bioinformatics* 21 (7): 951–60.

Söding, Johannes, Andreas Biegert, and Andrei N. Lupas. 2005. “The HHpred Interactive Server for Protein Homology Detection and Structure Prediction.” *Nucleic Acids Research* 33 (Web Server issue): W244–48.
Struck, Anna-Winona, Mark L. Thompson, Lu Shin Wong, and Jason Micklefield. 2012. “S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications.” *ChemBioChem*. https://doi.org/10.1002/cbic.201200556.
Stubbe, J. A. 1983. “Mechanism of B12-Dependent Ribonucleotide Reductase.” *Molecular and Cellular Biochemistry* 50 (1): 25–45.
Szilágyi, András, Dániel Györffy, and Péter Závodszky. 2017. “Segment Swapping Aided the Evolution of Enzyme Function: The Case of Uroporphyrinogen III Synthase.” *Proteins* 85 (1): 46–53.
Szilágyi, András, Yang Zhang, and Péter Závodszky. 2012. “Intra-Chain 3D Segment Swapping Spawns the Evolution of New Multidomain Protein Architectures.” *Journal of Molecular Biology* 415 (1): 221–35.
Taylor, William R. 2006. “Topological Accessibility Shows a Distinct Asymmetry in the Folds of βα Proteins.” *FEBS Letters*. https://doi.org/10.1016/j.febslet.2006.08.070.
Tollinger, Martin, Christian Eichmüller, Robert Konrat, Marja S. Huhta, E. Neil G. Marsh, and Bernhard Kräutler. 2001. “The B12-Binding Subunit of Glutamate Mutase from Clostridium Tetanomorphum Traps the Nucleotide Moiety of Coenzyme B12.” *Journal of Molecular Biology*. https://doi.org/10.1006/jmbi.2001.4696.
Tollinger, M., C. Eichmüller, R. Konrat, M. S. Huhta, E. N. Marsh, and B. Kräutler. 2001. “The B(12)-Binding Subunit of Glutamate Mutase from Clostridium Tetanomorphum Traps the Nucleotide Moiety of Coenzyme B(12).” *Journal of Molecular Biology* 309 (3): 777–91.
Tsiminis, Georgios, Erik P. Schartner, Joanna L. Brooks, and Mark R. Hutchinson. 2017. “Measuring and Tracking Vitamin B12: A Review of Current Methods with a Focus on Optical Spectroscopy.” *Applied Spectroscopy Reviews*. https://doi.org/10.1080/05704928.2016.1229325.
Turner, Nicholas J., and Elaine O’Reilly. 2013. “Biocatalytic Retrosynthesis.” *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.1235.
Vogel, Christine, and Veronica Morea. 2006. “Duplication, Divergence and Formation of Novel Protein Topologies.” *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 28 (10): 973–78.
Wymann, Walter E., Roman Davis, John W. Patterson, and Jürg R. Pfister. 1988. “Selective Alkylations of Certain Phenolic and Enolic Functions with Lithium Carbonate/Alkyl Halide.” *Synthetic Communications*. https://doi.org/10.1080/00397918808078806.
Wymann, W. E., R. Davis, J. W. Jun. Patterson, and J. R. Pfister. 1989. “ChemInform Abstract: Selective Alkylations of Certain Phenolic and Enolic Functions with Lithium Carbonate/Alkyl Halide.” *ChemInform*. https://doi.org/10.1002/chin.198914124.
Xia, Yu, and Michael Levitt. 2004. “Simulating Protein Evolution in Sequence and Structure Space.” *Current Opinion in Structural Biology* 14 (2): 202–7.
Yang, Biao, Haiping Lin, Kangjian Miao, Pan Zhu, Liangbo Liang, Kewei Sun, Haiming Zhang, et al. 2016. “Catalytic Dealkylation of Ethers to Alcohols on Metal Surfaces.” *Angewandte Chemie* 55 (34): 9881–85.
Zhang, Ming-Xin, Xu-Hong Hu, Yun-He Xu, and Teck-Peng Loh. 2015. “Selective Dealkylation of Alkyl Aryl Ethers.” *Asian Journal of Organic Chemistry*. https://doi.org/10.1002/ajoc.201500196.
Zhang, Yang, and Jeffrey Skolnick. 2005. “TM-Align: A Protein Structure Alignment Algorithm Based on the TM-Score.” *Nucleic Acids Research* 33 (7): 2302–9.