University of Groningen

Multidrug Resistance Regulators as Artificial Enzymes

Gutiérrez de Souza, Cora

DOI:
[10.33612/diss.168445268](https://doi.org/10.33612/diss.168445268)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Gutiérrez de Souza, C. (2021). *Multidrug Resistance Regulators as Artificial Enzymes*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.168445268>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Multidrug Resistance Regulators as Artificial Enzymes

Cora Gutiérrez de Souza

The research described in this thesis was carried out at the Stratingh Institute for Chemistry, University of Groningen, The Netherlands.

This work was financially supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) in the Netherlands.

Cover design by Marzia Munafò (munafomarzia.com)

Printed by Ipskamp Drukkers B.V.

ISBN (printed version): 978-94-6421-294-5

university of
 groningen

Multidrug Resistance Regulators as Artificial Enzymes

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Friday 14 May 2021 at 14.30 hours

by

Cora Gutiérrez de Souza

born on 2 March 1990
in Barcelona, Spain

Supervisor

Prof. J.G. Roelfes

Co-supervisor

Dr. S. Otto

Assessment Committee

Prof. M.W. Fraaije

Prof. W.C. Szymanski

Prof. J.-P. Mahy

"I'm a scientist and I know what constitutes proof. But the reason I call myself by my childhood name is to remind myself that a scientist must also be absolutely like a child. If he sees a thing, he must say that he sees it, whether it was what he thought he was going to see or not. See first, think later, then test. But always see first. Otherwise, you will only see what you were expecting. Most scientists forget that."

Douglas Adams, So Long, and Thanks for all the Fish

TABLE OF CONTENTS

Chapter 1

1

General Introduction

Chapter 2

29

Optimisation of the Second Coordination Sphere of an Artificial Metallo-Hydratase bearing an Unnatural Amino Acid

Chapter 3

57

Development of Porphyrin-LmrR based Artificial Metalloenzymes for Carbene-Transfer Reactions

Chapter 4

93

Artificial Metalloenzymes based on TetR Proteins and Cu(II) for Enantioselective Friedel-Crafts Alkylation Reactions

Chapter 5

119

Unexpected Catalytic Activity of the Regulatory Protein QacR

Chapter 6

137

Conclusions and Perspectives

Samenvatting

147

Acknowledgements

153

CHAPTER 1

General Introduction

1.1. INTRODUCTION

Enzymes are extraordinary catalysts. Their well-defined active sites allow to selectively bind cofactors and substrates in the ideal orientation to decrease the activation energy of a specific reaction. Enzymes exhibit excellent activities and enantioselectivities while working in water and at ambient temperature, under environmentally friendly conditions. However, due to their specificity the repertoire of transformations that enzymes are able to perform is limited.¹ Therefore, developments in the field of enzyme engineering, in which enzymes can be tuned to incorporate reactivities differing from their natural ones, are crucial for the progress of biocatalysis.^{2,3} On the other hand, homogeneous transition metal catalysts are able to achieve high activities and selectivities in a range of challenging reactions.⁴ These catalysts consist of a metal ion bound to one or several ligands, which form the 1st coordination sphere of the metal. The ligands can influence the reactivity of the metal and impart stereo- and site selectivity. Ligand substituents distal to the metal centre form the 2nd coordination sphere and can contribute to the reactivity and selectivity. However, the control over the orientation of the substituents is insufficient for an efficient design. Moreover, performing these chemical processes in a green manner still remains a challenge. The field of artificial metalloenzymes (ArMs) emerged from the idea of combining features of both enzymes and homogeneous metal catalysts.

ArMs consist of a bioscaffold (DNA or protein) in which a catalytically active metal complex is embedded.⁵⁻⁸ The bioscaffold constitutes the 2nd coordination sphere, interacts with the metal complex and the substrates, and stabilises the transition state of the reaction, thus contributing to the rate acceleration. The scaffold also provides a well-controlled chiral environment that can induce enantioselectivity. The transition metal complex forms the 1st coordination sphere and contributes to broadening the scope and to the site- and stereo selectivity.^{5,9}

The first ArMs were created by Kaiser and co-workers¹⁰ and Wilson and Whitesides¹¹ in the 1970's. Kaiser described a carboxypeptidase A, in which the original Zn(II) was replaced by Cu(II) endowing unnatural catalytic activity towards the oxidation of ascorbic acid. Whitesides and co-workers

had developed a methodology to attach an achiral diphosphinerhodium(I) moiety at a specific site in a protein to enantioselectively catalyse the asymmetric hydrogenation of α -acetamidoacrylic acid to *N*-acetylalanine. The high affinity of avidin for biotin was used in this study to incorporate a chemically modified biotin containing a diphosphine-rhodium (I) complex (1-RhNBD⁺Tf⁻), constituting the first example of “Trojan-horse” strategy (Scheme 1). While this pioneer artificial metalloenzyme only led to modest enantioselectivities and showed a limited applicability, it opened the path to the idea of combining a metal complex to provide new reactivities to an enzyme and a protein scaffold to create a chiral environment to induce enantioselectivity.

Scheme 1. a) Schematic representation of the avidin-biotin technology developed by Whitesides, with avidin depicted in blue. b) Biotin derivatised with the diphosphine-rhodium (I) complex (1-RhNBD⁺Tf⁻). NBD: norbornadiene; Tf⁻: triflate. c) Scheme of the hydrogenation reaction.

The next milestone for the field was the Darwinian approach, first investigated by Chen and Arnold,¹² and Stemmer¹³ in the 1990's. This approach consists in the introduction of sequential random mutations in protein scaffolds to modulate the activity of enzymes in unnatural environments or towards unnatural reactivities.

From the early 2000's, the development of protein engineering techniques has led to a rapid growth of the field of ArMs. Nowadays there are many examples of extremely efficient ArMs with turnover numbers (TONs) comparable to natural enzymes.¹⁴⁻¹⁶ A good example of the great potential of the field is the work of Hartwig, in which a P450 enzyme reconstituted with

an iridium porphyrin catalyses intramolecular carbene insertions into inactivated C-H bonds of diazoesters and intermolecular carbene insertions into C-H bonds with up to 98% ee and 35,000 TON (Scheme 2).¹⁷

Scheme 2. a) P450 variant reconstitution with iridium porphyrin. b) Model intra- and intermolecular carbene insertion reactions.

1.2. ARTIFICIAL ENZYMES DESIGN

Nature has always been an inspiration for scientists. In chemistry, one such approach is illustrated in biomimetic chemistry or artificial enzymes, i.e. synthetic materials that resemble enzymes.¹⁸ A high percentage of all enzymes known are metalloenzymes, which has driven a especial focus to engineering artificial metalloenzymes (ArMs). Learning from the binding of metal complexes to natural enzymes, and how those evolve to adapt to changing environments by selecting for beneficial mutations is key for the design of ArMs.

Selection of bioscaffolds

The first step in the design of an ArM is the selection of a bioscaffold-cofactor pair for catalysis of the targeted reaction (Scheme 3). Determining the protein to be used as scaffold is not an easy choice, considering the large number of possibilities (more than 150,000 protein structures available in the Protein Data Bank).¹⁹ The selection process typically requires computational design of *de novo* active sites,²⁰ screening large protein libraries, or rationally selecting a scaffold with potentially advantageous binding or catalytic characteristics. Since design from scratch is still too challenging, a common approach for creating ArMs is to engineer natural enzymes to enable new-to-

nature reactions via rational,^{2,21} computational,^{22,23} and/or evolutionary techniques.^{24–27}

Scheme 3. Approaches to selecting bioscaffolds for creation of artificial metalloenzymes.

Notably, many enzymes are extremely selective, since they evolved for a specific catalytic activity. Hence, not every enzyme can readily be repurposed for any reaction. The enzymes to be engineered have to show basal activity for the target reaction and fulfil certain characteristics regarding, for instance, stability.²⁸ Despite the challenge of having to screen many protein variants (either *in silico* or *in vitro*), several research groups have now shown successful examples of evolved ArMs. For instance, the reactive heme iron of cytochrome P450s and cytochrome *c* can be used to catalyse a range of new-to-nature carbene and nitrene transfer reactions when the protein is submitted to rational and evolutionary optimisation.^{29,30} A conceptually different approach is to use bioscaffolds without inherent reactivity to either repurpose the protein cofactor for catalysis or introduce a catalytically competent, abiotic element. For example, myoglobin, which, despite being a transport protein instead of an enzyme, has shown similar catalytic behaviour as P450s and cytochromes *c*.³¹ Moreover, substitution of the catalytic heme by other porphyrinoid cofactors has also given access to different reactivities.³² Additionally, proteins without natural metal cofactor but with

promiscuous or engineered binding sites, can be used to embed metal complexes through a number of available methodologies. Albumins, for instance, have no catalytic activity in nature, but contain a lysine residue that can be used as catalytic residue.³³ Multidrug Resistance Regulator proteins have also shown that their binding promiscuity allows to incorporate a range of metal complexes and access a variety of reactivities.^{34,35} As shown by the above mentioned examples, the metal cofactor is a rich source of novel activity. Although altering the metal complex can be a way to modulate the activity of ArMs, the efforts to chemically optimise the cofactor are considerably higher than genetic optimisation of the scaffold.

Introducing transition metal complexes into protein scaffolds has proven to be an advantageous way to create new artificial enzymes. Despite being less common, proteins can also be repurposed towards abiotic reactivities without the need of a metal,^{33,36} as exemplified by artificial transaminases, created from small lipid-binding proteins with covalently attached pyridoxamine analogues.³⁷ These non-metal containing constructs rely on a hydrophobic binding site to effectively bind the substrates; a chiral environment to provide enantioselectivity; and amino acid residues or functional groups attached to the protein scaffolds to catalyse the targeted reaction.

Therefore, binding promiscuity to be able to incorporate different metal complexes and/or substrates, as well as versatility towards the use of certain protein residues or cofactors (natural or otherwise), are key features to take into account in the search of ArM's scaffolds.

Anchoring strategies

There are several methodologies available for introducing a metal complex into a bioscaffold. ArMs can be built by using dative bonds between the metal ion and functionalities present on the scaffold; covalently attaching a metal complex to a residue of the protein through the ligand; biosynthetically incorporating an unnatural amino acid to use it as metal-binding or catalytic residue; or taking advantage of supramolecular protein-cofactor interactions.³⁸ Each methodology involves certain requirements and provides specific characteristics to the system, making the choice of the approach of assembly an important factor of the design of an ArM.⁸

The dative approach (Scheme 4) consists of expressing the apo form of a metalloprotein, or removing the metal ion or cofactor after expression, and reconstituting it with a new metal. Active sites of several proteins have been reconstituted with exogenous metals to obtain unnatural reactivities.^{39,40}

Scheme 4. Dative approach to ArM formation.

This is the case for carbonic anhydrases, which have a well-defined tris-histidine-ligated Zn(II) binding site that has been reconstituted with metals like Mn(II) and Rh(I) to catalyse alkene hydrogenation,⁴¹ epoxidation⁴² and hydroformylation⁴³ reactions (Scheme 5). Binding of these metals lead to a very similar overall structure of the active site compared to the original one.

a) Hydrogenation

b) Epoxidation

c) Hydroformylation

Scheme 5. Scheme of reactions catalysed by reconstituted human carbonic anhydrases (hCA).

Despite that the use of reconstituted natural binding sites facilitates the formation of ArMs, it also limits the approach to metals fitting in the 1st coordination sphere of the scaffold; proteins that can be expressed in the apo form or that the metal can be removed while remaining stable; and proteins that can be efficiently reconstituted. The use of proteins without a natural metal binding site, like albumins, which showed to be able to bind catalytically active metal complexes⁴⁴ or *de novo* designed binding sites have also been explored.⁴⁵

The covalent approach (Scheme 6) comprises the orthogonal bioconjugation of a functionalised metal ligand to an amino acid side-chain of the protein with unique reactivity. In this approach, usually a cysteine is selectively conjugated to the metal ligand, derivatised with an electrophilic moiety. This approach requires that the protein does not contain any cysteine residue, so that the metal ligand can bind selectively to the cysteine mutated onto the desired position, and a sufficiently large vacant space to accommodate the linker, the metal complex and the substrates.

Scheme 6. Covalent approach to ArM formation.

In a seminal work, Kaiser showed that covalent modification of a cysteine residue in papain with bromoacetyl derivatives of flavin converted the protein in a highly effective oxidoreductase (Scheme 7).⁴⁶ The precise site selectivity of the covalent approach allows for a good control over the active site, facilitating the rational prediction of beneficial mutations. Moreover, the well-established and vast array of bioconjugation techniques⁴⁷ offers the possibility of incorporating an uncountable amount of diverse metal ligands.³⁸ This methodology, however, requires post-expression chemical modification of the scaffold and further purification to remove the excess of ligand, which makes screening protein variants very laborious and limits the possibility of optimisation of the ArM.

Scheme 7. a) Papain conjugated to a flavin derivative. b) Oxidation of *N*-alkyl-1,4-dihydronicotinamides by flavopapain.

The supramolecular approach (Scheme 8) involves non-covalent interactions between the protein scaffold and the metal complex, which eliminates the need of post-expression chemical modifications of the scaffold. This results in a straightforward preparation of the ArMs that readily self-assemble upon mixing the protein and the metal complex, vastly facilitating the screening and optimisation processes.

Scheme 8. Supramolecular approach to ArM formation.

The first example of this methodology was based on the Trojan Horse strategy, established by Whiteside and co-workers in 1978, where they exploited the high affinity between avidin and its natural cofactor biotin. In this work, they described the creation of an asymmetric hydrogenase by embedding an achiral diphosphinerhodium(I) moiety attached to a biotin cofactor into avidin.¹¹ Thereafter, many research groups have built ArMs relying on a ligand with affinity for a scaffold.^{48,49} The scope of scaffolds that can be used in the Trojan Horse strategy is, however, limited, as strong and specific interactions between the protein and the cofactor are required. Supramolecular assembly of ArMs has been particularly successful for heme-proteins, like myoglobin, P450s and cytochrome *c*.⁵⁰ The affinity of these proteins for heme and the high range of reactions that porphyrins can

perform, create a very suitable scenario for developing new ArMs⁵¹⁻⁵⁶ (Scheme 9).

Scheme 9. Examples of reactions catalysed by engineered heme proteins with iron carbenoids and nitrenoids as intermediate.^{50,57,58}

This approach has been applied to a variety of protein-metal complex pairs.²⁶ Proteins with promiscuous binding sites, like multidrug resistance regulators are also a good target to introduce different metal complexes and substrates in their scaffolds using hydrophobic interactions with the binding pocket. As shown by LmrR with hemin⁵⁹ and Cu(II)-phenanthroline⁶⁰ complexes, as well as RamR, CgmR and QacR.⁶¹ This method leads to versatile systems where different metal complexes can be incorporated to the same scaffold to catalyse several reactions.

Lastly, unnatural amino acids can be biosynthetically incorporated into protein scaffolds *via* the expanded genetic code methodology.⁶² To date, a number of different unnatural amino acids with different functional groups have been incorporated into proteins.^{63,64} The expanded genetic code methodology has been used for attaching fluorophores as labelling strategy, reactive groups as handles for biorthogonal chemistry and, lately, they have

shown to be a technique useful for incorporating catalytic activity into proteins (Scheme 10).⁶⁵

Scheme 10. *In vivo* incorporation of metal-binding UAA approach to ArM formation.

a) METAL BINDING UAA

LmrR_Bipyridylalanine

LmrR_Hydroxyquinolinalanine

Assymetric hydration

Friedel-Crafts alkylation

b) NON-CANONICAL HEME-PROXIMAL LIGAND

Ascorbate peroxidase_N δ -methyl histidine

Myoglobin_N δ -methyl histidine

Guaiacol oxidation

c) REACTIVE UAA

LmrR_p-aminophenylalanine

Hydrazone formation

Oxime formation

Scheme 11. Unnatural reactions catalysed by engineered proteins bearing non-canonical amino acids.^{65–72}

Integrating functional groups beyond those of the 20 canonical amino acids into a protein scaffold allows to incorporate catalytic activity following several

strategies (see Scheme 11): i) using metal binding unnatural amino acids, like (2,2'-bipyridin-5-yl)alanine^{65,66} or (8-hydroxyquinolin-3-yl)alanine⁶⁷ to chelate catalytically active transition metal ions; ii) substitution of natural amino acids with structural analogues to tune the coordination site of metalloenzymes, like *N*₈-methyl histidine,^{68,69,71} iii) the use of unnatural amino acids as catalytic residues, like *p*-aminophenylalanine for aniline catalysis.^{70,72,73} However, this technique is restricted to the number of tRNA/aa-tRNA synthetase pairs available for encoding the specific amino acids. Lower expression yields than those of wild-type proteins and expensive cost or commercial unavailability of some UAA can occasionally hamper the applicability of this approach.

Optimisation

Once the ArM has been assembled and presents some activity towards the targeted reaction, protein engineering tools can be used to optimise its activity and selectivity (see Scheme 12).

Scheme 12. Approaches to artificial metalloenzymes optimisation.

Directed evolution is the laboratory equivalent of natural selection, which consists in iteratively creating gene variability with random mutagenesis and screening for a particular characteristic to accumulate performance-enhancing mutations. This revolutionary approach obtained full credit for its great utility when Frances Arnold, George Smith and Gregory Winter were awarded in 2018 the Nobel Prize for pioneering and refine the directed

evolution of enzymes and proteins.⁷⁴ This technique has led to many examples of evolved enzymes with high activities and selectivities.^{75,76} Arnold transformed cytochrome P450s into carbene and nitrene transferases. These heme containing enzymes showed basal level of the abiotic activity and were further evolved to become highly active and selective for the targeted abiotic reactivities.⁵⁰ In another work, covalent linkage of a di-rhodium complex to a propyl oligopeptidase containing the UAA L-4-azidophenylalanine was active in an olefin cyclopropanation reaction. The ArM was evolve through several rounds of mutagenesis to achieve high enantioselectivities, improved water tolerance and broader substrate scope.⁷⁷ These protein variants, however, have to be purified prior to activity screening, since the ArMs are formed *in vitro*, which makes high throughput screening difficult. The groups of Hayashi and Schwaneberg have applied the directed evolution approach in artificial metalloenzymes, for instance in a $\text{Cp}^*\text{Rh}^{\text{III}}$ -nitrobinding for the cycloaddition of acetophenone oximes with alkynes.⁷⁸ In this work, Hayashi and Schwaneberg developed an *in vitro* high throughput screening platform that allowed to screen more than 4000 variants to obtain an evolved highly active ArM.

A successful approach to overcome some limitations when screening is to create ArMs *in vivo*, which avoids the protein purification step and allows to increase the number of variants that can be screened. For instance, Rudi Fasan, created a myoglobin-based ArM able to work in whole cells, by substituting the heme cofactor for a more oxygen tolerant porphyrin;⁵¹ and Frances Arnold developed many artificial heme-enzymes that were used *in vivo*.⁷⁹ Because many metal cofactors are inhibited by cellular components, the group of Tomas Ward, developed an *in vivo* compartmentalisation strategy in which streptavidin is secreted to the periplasm where a complex with the biotinylated metal cofactor can be formed. This ArM was active in an olefin metathesis reaction and was improved by directed evolution *in vivo*.⁸⁰

The possibility of applying the Darwinian strategy *in vivo* presents a great advantage and paves the way towards engineered metabolism. This approach, however, is still limited to enzymes that exhibit basal level of activity and to the possibility of applying high-throughput screening in order to efficiently assign the fitness of the numerous produced mutants. Nowadays, these assays

are not general for all kind of substrates and the large screening efforts are not viable in every laboratory.^{6,24,25,81,82}

1

On the other hand, computational tools can be used to both identify suitable scaffold-cofactor pairs and select potentially beneficial mutations, reducing the number of variants that have to be screened. Although there are a few examples of *de novo* computationally designed ArMs,^{20,83} computational algorithms are not robust enough yet to be generalised as effective prediction methods, but rather as a way to understand experimental data. For instance, using quantum mechanical (QM) calculations and Rosetta algorithms, the active site of an ArM able to catalyse intermolecular Diels-Alder reactions was designed. However, the activity reached was very low compared to natural enzymes and further evolution studies were required for optimisation of the artificial Diels-Alderase.⁸⁴ The examples of *de novo* design show that computational tools often require previous mechanistic or binding information, as well as further evolution of the enzymes to improve the activities.^{20,22,83,85}

Lastly, by inspecting the crystal structure and considering previous knowledge on the reaction mechanism, rational design can be used to judiciously predict how certain mutations will affect the enzyme behaviour. This approach often allows to identify single beneficial mutations without the need of an extensive production and screening of protein variants. However, structural information about the protein is required and usually only mutations on the proximity of the active site can be predicted. Therefore, this approach is usually complemented by directed evolution techniques. Moreover, the lack of deep knowledge on the structure-function relationship of proteins, limits our potential to rationally design artificial enzymes.^{21,86}

Overall, future developments in screening technology, computational analysis and knowledge on protein behaviour, will most likely contribute to improve the applicability of each of these approaches. To date, a combination of these complementary tools constitutes the ideal scenario for shaping a protein towards a specific catalytic activity.^{87,88}

1.3. MULTIDRUG RESISTANCE REGULATORS

Multidrug resistance regulators (MDRs) are a class of proteins involved in conferring antibiotic resistance to bacteria by regulating efflux pumps expression. These proteins are usually homodimers that consist of a DNA binding domain and a binding pocket for recognition of structurally diverse aromatic, planar drugs.⁸⁹ Their intrinsic multispecificity has raised interest within our group for their use as versatile artificial enzymes.

In this thesis, four scaffolds from this class of proteins are explored: LmrR, CgmR, RamR and QacR. They were selected because X-ray structures of these proteins with aromatic guest molecules bound are available. Structural and binding information of the protein scaffold are key for the design and optimisation of artificial metalloenzymes, therefore, characteristics of the four proteins are described in this section.

1.3.1. LmrR

LmrR (Lactococcal Multidrug Resistance Regulator) is a PadR family transcriptional repressor in *Lactococcus lactis*. It regulates the production of LmrCD, a major multidrug ABC transporter. The N-terminus of LmrR, as other characterised proteins from the PadR family, contains a conserved DNA-binding β -winged helix-turn-helix domain⁹⁰ (α 1, α 2 and α 3 helices, and β 1 and β 2 strands). The C-terminus comprises a helix (α 4) involved in dimerisation by interaction with the DNA-binding domain of the other monomer subunit.⁹¹ The large and flat hydrophobic pore formed in the dimer interface, about 22 Å in width and 6 Å in height, serves as multidrug binding domain. All the X-ray structures available of LmrR with different guest molecules bound show a consistent stoichiometry of one molecule bound per dimer of protein. In all cases, the guest molecules appear wedged in between W96 and W96' by π -stacking interactions, with some additional stabilising interactions with hydrophobic residues of the protein. Comparison of X-ray structures of LmrR bound to different guest molecules (Figure 1b) evidence the plasticity of the hydrophobic pore, which is responsible for the promiscuous drug recognition nature of this protein. The main differences between the structures result from different orientations of α 4 helix. Small differences can also be observed in helices α 1 and α 2, whereas no other significant shifts are observed in the rest of the β -winged helix-turn-helix domain.

Figure 1. Ribbon representation of: a) LmrR apo form (3F8B) with secondary structure elements highlighted in colour: α 1, yellow; α 2, grey; α 3, magenta, β 1 and β 2, green; α 4, blue. b) Superposition of X-ray structures of LmrR with Hoechst 33342 (3F8C, yellow), daunomycin (3F8F, blue), riboflavin (4ZZD, purple), hemin (6FUU, green) and Cu(II)-1,10'-phenanthroline (6R1L, orange).

1.3.2. TetR family of proteins

CgmR, RamR and QacR are transcriptional repressors from the TetR family of MDRs. They are homodimeric proteins of ~20 kDa per monomer, with one multisite-binding pocket per monomer. Each monomer contains 9 α -helices divided in a highly conserved helix-turn-helix DNA-binding domain at the N-terminus, and a regulatory core involved in drug binding and dimerisation. Despite the global structural similarity within proteins of the TetR family, the low sequence identity between the regulatory domains results in response of the repressors to different drugs as inducers.^{92,93}

CgmR

CgmR has been crystallised in the presence of the guest molecules ethidium bromide (1:2 stoichiometry guest molecule:monomer) and methylene blue (2:2 stoichiometry) (Figure 2a). Studies of the structural changes of CgmR with these two guest molecules bound, suggest that the difference in stoichiometry is due to the different size of the bound molecules. Ethidium occupies a bigger volume, of approximately 1400 Å³, than methylene blue, which diminishes the size of the other pocket (300 Å³ with ethidium and 700 Å³ with methylene blue), leaving no room for another guest molecule to bind. Nevertheless, both guest molecules were found to bind in a similar manner to the inducer-binding pocket between Trp63 and Phe147 by π - π interactions. Other residues from the regulatory domain (Leu59, Leu88, Val92, Leu100 and Trp113), as well as Ile152' from the neighbouring subunit, also contribute to the hydrophobic interactions with the guest molecules (Figure 2b). The

overall structure of CgmR with both guest molecules is quite similar (Figure 2a), however ethidium and methylene blue form different water-mediated hydrogen networks, which provides certain degree of binding promiscuity.⁹⁴

Figure 2. a) Superposition of ribbon representation of CgmR with ethidium (2ZOZ, brown) and methylene blue (2YVE, blue). b) Ribbon representation of CgmR bound to ethidium (2ZOZ), with Trp63 and Phe147 depicted in blue; Leu59, Leu88, Val92, Leu100 and Trp113, depicted in orange; and Ile152' in green.

RamR

RamR was found to bind several antimicrobial drugs with 1:2 stoichiometry (berberine, ethidium bromide and rhodamine 6G) and with 1:1 stoichiometry (crystal violet and dequalinium). For all guest molecules studied, π - π interactions with Phe155 were identified (Figure 3a). In addition, each guest molecule interacts with a different set of amino acids.⁹⁵ Later on, it was found that RamR was also able to bind bile acids like cholic and chenodeoxycholic acid, involving hydrogen bonds with Tyr59, Thr85, Ser137 and Asp152 residues (Figure 3b).⁹⁶ Interestingly, these molecules did not interact with Phe155, which further illustrates the promiscuity of the multisite-binding pocket of RamR.

Figure 3. a) Ribbon representation of RamR in complex with ethidium (3VYV), with Phe155 depicted in blue. b) Ribbon representation of RamR in complex with cholic

acid (6IE8), with drug-interacting residues depicted in green.

QacR

To date, QacR seems to have the most broad criteria for drug recognition, since it binds a large range of structurally diverse mono and bivalent cationic lipophilic compounds involving different amino acid side chains.⁹⁷ Binding of guest molecules typically happens in a 1:2 (guest molecule:monomer) stoichiometry. The pockets of QacR are rich in negatively charged residues (Glu57, 58, 90 and 120), responsible for attracting different cationic molecules and neutralising their charges; as well as in aromatic residues (Trp61, Tyr93, Tyr103, Tyr123 and Phe162) which are involved in binding.^{97–100} Tyr92 plays an important structural role in stabilising the inducer-binding pocket in the absence of guest molecules. Upon binding, Tyr92 rotates towards the solvent, creating an extensive binding-site.¹⁰¹ Each binding site of QacR consist of two separate but partially overlapping pockets, corresponding to the centre of ethidium and Rhodamine 6G (R6G) binding sites.^{98,101} The X-ray structure of QacR has also been elucidated bound simultaneously to two guest molecules, ethidium and proflavine (one in each of the overlapping pockets), which is possible due to the smaller size of proflavin.¹⁰²

Figure 4. a) Ribbon representation from two different perspectives of QacR with ethidium bound (1QVU). Glutamates are depicted in blue, aromatic residues in

magenta, Tyr92 in orange and ethidium in green. b) Superposition of ribbon representation of QacR bound to ethidium with blue ribbon, and to R6G (IJUS) with yellow ribbon and pink R6G.

1.3.3. Binding mechanism in LmrR and QacR

The obvious differences in drug recognition between LmrR and QacR –the most studied of the TetR proteins described in this thesis– inevitably lead to differences in catalytic outputs: in the reaction scope, the substrate acceptance and the selectivity.

LmrR consist of one unique hydrophobic pocket in the dimer interface and uses W96 and W96' as main binding residues. The hydrophobicity of the pocket serves to attract apolar molecules, which are then captured by π - π stacking interactions in between W96/W96', and further stabilised by polar interactions with other residues in the pocket. The flexibility of the α 4 and α 4' helixes allow to adapt the size of the pocket to different aromatic planar molecules. The wide and promiscuous pocket is capable of accommodating aromatic metal complexes (and linkers when necessary), hydrophobic substrates, and even has room for them to adopt conformations suitable for catalysis. This has allowed to use LmrR as biological host for a variety of organometallic cofactors to create new ArMs.³⁴ Nevertheless, the boundaries of the diversity of LmrR based ArMs lie in the need of hydrophobicity of metal complexes and substrates in order to specifically bind to LmrR. The fact that the active site is embedded in the cavity between the two monomers causes that every mutation in the active site happens twice in the pocket. This observation has to be taken into account when introducing mutations or creating active sites in LmrR's pocket by covalent binding of metal ligands to residues of the protein or biosynthetic incorporation of metal-binding unnatural amino acids.

On the other hand, QacR has two smaller hydrophobic pockets, one per monomer. The drug recognition pattern of QacR involves hydrophobic interactions with several aromatic residues and cationic molecules attracted by negatively charged residues. Studies on the effect of acid residues of QacR in binding affinity concluded that, whereas these residues individually played a role in attracting and orientation of the guest molecules in the pocket, their absence did not affect the binding affinity.⁹⁸ This finding highlights the joint

use of several residues of QacR's active site for drug recognition and binding, in contrast to LmrR which relays mainly on tryptophans 96 for binding. To the best of our knowledge, there are no structures of QacR where it simultaneously binds a guest molecule in each of the monomer pockets. The occupancy of one of the monomers induces certain allosteric changes that block the entrance of the unoccupied monomer, preventing binding of guest molecules in both monomers simultaneously.¹⁰¹ This fact also presents a significant difference between QacR and LmrR, since in the former, mutations in the pocket will affect only once to the active site. Additionally, each binding site of QacR consists of two overlapping sub-pockets, which contribute to binding promiscuity, since each guest molecule shows preference for one of them. Binding in one sub-pocket excludes binding in the other one, except for smaller guest molecules that can occupy both sub-pockets at the same time.^{101,102}

A good example of the effect of these differences in binding between LmrR and QacR are the ArMs created with these proteins and Cu(II)-phenanthroline for the Friedel-Crafts alkylation of unsaturated enones. LmrR needs the aromatic phenanthroline ligand to drive the reaction to happen within the hydrophobic environment of the protein.⁶⁰ QacR, on the other hand, was able to catalyse the reaction both with Cu(II)-phenanthroline and with Cu(II) ions alone, due to the smaller size of the pocket of QacR and its preference for binding cationic molecules.¹⁰³

In conclusion, MDR proteins constitute a good versatile selection of scaffolds for creation of novel ArMs. Knowledge of the binding characteristics of the scaffolds, as well as structural information of the protein in drug-bound and drug-free states are beneficial for successful design.

1.4. AIM AND OUTLINE OF THE THESIS

This thesis describes the work on the design, production and optimisation of artificial (metallo)enzymes based on multispecific proteins without native catalytic activity. Ultimately, the goal was to deepen our understanding of the potential and versatility of the studied scaffolds as artificial enzymes.

Chapters 2 and 3, discuss artificial metalloenzymes based on the LmrR protein scaffold. The large and flexible hydrophobic pore of this protein allows to introduce metal complexes *via* different assembly methodologies.

In **Chapter 2**, attempts towards optimising the catalytic performance of an artificial metallohydratase by mutating rationally selected residues of the second coordination sphere are described. To create the ArM, the expanded genetic code methodology is used to incorporate the metal binding unnatural amino acid 2,2'-bipyridylalanine (bpyA). In this chapter, the limitations of this approach are explored and an alternative methodology for protein expression is investigated.

Chapter 3 presents our efforts on the design, production and characterisation of an artificial heme-enzyme for the abiotic cyclopropanation reaction. Here, the ArM is created using the supramolecular approach of assembly, a very straightforward method that greatly facilitates the ArM's optimisation. A range of strategies are used to further characterise and understand the system: crystallographic and computational data, as well as UV-Vis analysis and binding affinity by fluorescence quenching. The potential of the heme-LmrR system as artificial enzyme is explored by investigating the reaction scope, binding of different porphyrins to LmrR, and engineering of an axial ligand for the iron ion.

In **Chapters 4 and 5**, artificial enzymes for the Friedel-Crafts alkylation reaction (F-C) and the tandem Friedel-Crafts/enantioselective protonation (F-C/EP) are created using three proteins from the TetR family of multidrug resistance regulators: CgmR, RamR and QacR. In these two reactions very similar substrates are used, however, the results described in **Chapters 4 and 5**, evidence that they can follow different reaction pathways.

In **Chapter 4**, we present ArMs based on CgmR, RamR and QacR able to catalyse the Friedel-Crafts alkylation of β -substituted α,β -unsaturated enones to indoles, using Cu(II) ions without the need of a ligand for the metal. Moreover, the substrate scope for the F-C reaction with CgmR, RamR and QacR based ArMs is investigated.

Chapter 5 describes the versatility of the Cu(II)-CgmR/RamR/QacR ArMs by showing their ability to also catalyse the F-C/EP reaction of α -substituted

α,β -unsaturated enones to indoles. The key mechanistic difference between these two reactions relies on the fact that, in the F-C/EP, the protonation step of the reaction is responsible for creating the chiral centre, which makes achieving enantioselectivity more challenging. The main finding investigated in this chapter, however, is the unexpected catalytic activity of QacR at pH 5 on the F-C/EP reaction without the need of any metal. Mutagenesis and computational studies are used to examine the activity of the system.

Finally, **Chapter 6** offers the overall conclusions of the thesis as well as perspectives on the field of artificial metalloenzymes and possible continuation of the research presented.

1.5. REFERENCES

- 1 M. T. Reetz, *Chem. Rec.*, 2016, **16**, 2449–2459.
- 2 S. M. Basheer and S. Chellappan, in *Bioresources and Bioprocess in Biotechnology*, 2017.
- 3 M. Leisola, J. Jokela, O. Pastinen, O. Turunen and H. Schoemaker, in *Physiology and Maintenance*, 2002.
- 4 E. P. Beaumier, A. J. Pearce, X. Y. See and I. A. Tonks, *Nat. Rev. Chem.*, 2019, **3**, 15–34.
- 5 Y. Okamoto, V. Köhler, V. Lebrun, R. Reuter, T. R. Ward, F. Schwizer, T. Heinisch, M. M. Pellizzoni, Y. Gu and J. C. Lewis, *Chem. Rev.*, 2017, **118**, 142–231.
- 6 E. M. Brustad and F. H. Arnold, *Curr. Opin. Chem. Biol.*, 2011, **15**, 201–210.
- 7 J. Bos and G. Roelfes, *Curr. Opin. Chem. Biol.*, 2014, **19**, 135–143.
- 8 J. C. Lewis, *ACS Catal.*, 2013, **3**, 2954–2975.
- 9 J. C. Lewis, *Curr. Opin. Chem. Biol.*, 2015, **25**, 27–35.
- 10 K. Yamamura and E. T. Kaiser, *J. Chem. Soc. Chem. Commun.*, 1976, 830–831.
- 11 G. M. Wilson, M. E.; Whitesides, *J. Am. Chem. Soc.*, 1978, **100**, 306–307.
- 12 K. Chen and F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 5618–5622.
- 13 W. P. C. Stemmer, *Nature*, 1994, **370**, 389–391.

- 14 M. Bordeaux, V. Tyagi and R. Fasan, *Angew. Chem. Int. Ed.*, 2015, **54**, 1744–1748.
- 15 O. F. Brandenburg, D. C. Miller, U. Markel, A. Ouald Chaib and F. H. Arnold, *ACS Catal.*, 2019, **9**, 8271–8275.
- 16 O. F. Brandenburg, C. K. Prier, K. Chen, A. M. Knight, Z. Wu and F. H. Arnold, *ACS Catal.*, 2018, **8**, 2629–2634.
- 17 P. Dydio, H. M. Key, A. Nazarenko, J. Y. E. Rha, V. Seyedkazemi, D. S. Clark and J. F. Hartwig, *Science*, 2016, **354**, 102–106.
- 18 Y. Murakami, J. I. Kikuchi, Y. Hisaeda and O. Hayashida, *Chem. Rev.*, 1996, **96**, 721–758.
- 19 https://cdn.rcsb.org/rcsb-pdb//general_information/news_publications/annual_reports/annueal_report_year_2019.pdf.
- 20 H. Kries, R. Blomberg and D. Hilvert, *Curr. Opin. Chem. Biol.*, 2013, **17**, 221–228.
- 21 Y. W. Lin, *Coord. Chem. Rev.*, 2017, **336**, 1–27.
- 22 D. Hilvert, *Annu. Rev. Biochem.*, 2013, **82**, 447–470.
- 23 V. Muñoz Robles, E. Ortega-Carrasco, L. Alonso-Cotchico, J. Rodriguez-Guerra, A. Lledós and J. D. Maréchal, *ACS Catal.*, 2015, **5**, 2469–2480.
- 24 F. H. Arnold, *Angew. Chem. Int. Ed.*, 2018, **57**, 4143–4148.
- 25 R. E. Cobb, R. Chao and H. Zhao, *AIChE J.*, 2013, **59**, 1432–1440.
- 26 A. Ilie and M. T. Reetz, *Isr. J. Chem.*, 2015, **55**, 51–60.
- 27 C. Zeymer and D. Hilvert, *Annu. Rev. Biochem.*, 2018, **87**, 131–157.
- 28 R. B. Leveson-Gower, C. Mayer and G. Roelfes, *Nat. Rev. Chem.*, 2019, **3**, 687–705.
- 29 S. T. Jung, R. Lauchli and F. H. Arnold, *Curr. Opin. Biotechnol.*, 2011, **22**, 809–817.
- 30 S. B. J. Kan, R. D. Lewis, K. Chen and F. H. Arnold, *Science*, 2016, **354**, 1048–1051.
- 31 G. Sreenilayam and R. Fasan, *Chem. Commun.*, 2015, **51**, 1532–1534.

- 32 K. Oohora, A. Onoda and T. Hayashi, *Acc. Chem. Res.*, 2019, **52**, 945–954.
- 33 D. C. M. Albanese and N. Gaggero, *RSC Adv.*, 2015, **5**, 10588–10598.
- 34 G. Roelfes, *Acc. Chem. Res.*, 2019, **52**, 545–556.
- 35 M. Bersellini and G. Roelfes, *Org. Biomol. Chem.*, 2017, **15**, 3069–3073.
- 36 D. Qi, C. M. Tann, D. Haring and M. D. Distefano, *Chem. Rev.*, 2001, **101**, 3081–3111.
- 37 D. Häring and M. D. Distefano, *Bioorganic Med. Chem.*, 2001, **9**, 2461–2466.
- 38 N. Stephanopoulos and M. B. Francis, *Nat. Chem. Biol.*, 2011, **7**, 876–884.
- 39 A. L. Pinto, H. W. Hellinga and J. P. Caradonna, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 5562–5567.
- 40 D. E. Benson, M. S. Wisz and H. W. Hellinga, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 6292–6297.
- 41 Q. Jing, K. Okrasa and R. J. Kazlauskas, *Chem. - A Eur. J.*, 2009, **15**, 1370–1376.
- 42 A. Fernández-Gacio, A. Codina, J. Fastrez, O. Riant and P. Soumillion, *ChemBioChem*, 2006, **7**, 1013–1016.
- 43 Q. Jing and R. J. Kazlauskas, *ChemCatChem*, 2010, **2**, 877–877.
- 44 T. Kokubo, T. Sugimoto, T. Uchida, S. Tanimoto and M. Okano, *J. Chem. Soc. - Ser. Chem. Commun.*, 1983, 769–770.
- 45 J. Podtetenieff, A. Taglieber, E. Bill, E. J. Reijerse and M. T. Reetz, *Angew. Chem. Int. Ed.*, 2010, **49**, 5151–5155.
- 46 E. T. Kaiser and D. S. Lawrence, *Science*, 1984, **226**, 505–511.
- 47 G. T. Hermanson, *Bioconjugate Techniques: Third Edition*, 2013.
- 48 C. Lo, M. R. Ringenberg, D. Gnanadt, Y. Wilson and T. R. Ward, *Chem. Commun.*, 2011, **47**, 12065–12067.
- 49 T. R. Ward, *Acc. Chem. Res.*, 2011, **44**, 47–57.
- 50 O. F. Brandenburg, R. Fasan and F. H. Arnold, *Curr. Opin. Biotechnol.*, 2017, **47**, 102–111.
- 51 G. Sreenilayam, E. J. Moore, V. Steck and R. Fasan, *ACS Catal.*, 2017, **7**, 7629–7633.

- 52 E. W. Reynolds, M. W. McHenry, F. Cannac, J. G. Gober, C. D. Snow and E. M. Brustad, *J. Am. Chem. Soc.*, 2016, **138**, 12451–12458.
- 53 H. M. Key, P. Dydio, D. S. Clark and J. F. Hartwig, *Nature*, 2016, **534**, 534–537.
- 54 G. Sreenilayam, E. J. Moore, V. Steck and R. Fasan, *Adv. Synth. Catal.*, 2017, **359**, 2076–2089.
- 55 S. I. Ozaki, T. Matsui and Y. Watanabe, *J. Am. Chem. Soc.*, 1996, **118**, 9784–9785.
- 56 S. I. Ozaki, M. P. Roach, T. Matsui and Y. Watanabe, *Acc. Chem. Res.*, 2001, **34**, 818–825.
- 57 S. B. Jennifer Kan, X. Huang, Y. Gumulya, K. Chen and F. H. Arnold, *Nature*, 2017, **552**, 132–136.
- 58 K. Chen, X. Huang, S. B. Jennifer Kan, R. K. Zhang and F. H. Arnold, *Science*, 2018, **360**, 71–75.
- 59 L. Villarino, K. E. Splan, E. Reddem, L. Alonso-Cotchico, C. Gutiérrez de Souza, A. Lledós, J.-D. Maréchal, A.-M. W. H. Thunnissen and G. Roelfes, *Angew. Chem. Int. Ed.*, 2018, **57**, 7785–7789.
- 60 L. Villarino, S. Chordia, L. Alonso-Cotchico, E. R. Reddem, A.-M. Thunnissen, J.-D. Maréchal and G. Roelfes, *ACS Catal.*, 2020, **10**, 11783–11790.
- 61 C. Gutiérrez de Souza, M. Bersellini and G. Roelfes, *ChemCatChem*, 2020, **12**, 3190–3194.
- 62 L. Wang, A. Brock, B. Herberich and P. G. Schultz, *Science*, 2001, **292**, 498–500.
- 63 D. D. Young and P. G. Schultz, *ACS Chem. Biol.*, 2018, **13**, 854–870.
- 64 J. W. Chin, *Nature*, 2017, **550**, 53–60.
- 65 I. Drienovská, A. Rioz-Martinez, A. Draksharapu and G. Roelfes, *Chem. Sci.*, 2015, **6**, 770–776.
- 66 I. Drienovská, L. Alonso-Cotchico, P. Vidossich, A. Lledós, J.-D. Maréchal and G. Roelfes, *Chem. Sci.*, 2017, **8**, 7228–7235.
- 67 I. Drienovská, R. A. Scheele, C. Gutiérrez de Souza and G. Roelfes, *ChemBioChem*, 2020, **21**, 3077–3081.
- 68 A. P. Green, T. Hayashi, P. R. E. Mittl and D. Hilvert, *J. Am. Chem. Soc.*, 2016,

138, 11344–11352.

- 69 T. Mori, D. Hilvert, M. Pott, T. Hayashi, P. R. E. Mittl and A. P. Green, *J. Am. Chem. Soc.*, 2018, **140**, 1535–1543.
- 70 C. Mayer, C. Dulson, E. Reddem, A. M. W. H. Thunnissen and G. Roelfes, *Angew. Chem. Int. Ed.*, 2019, **58**, 2083–2087.
- 71 T. Hayashi, D. Hilvert and A. P. Green, *Chem. - A Eur. J.*, 2018, **24**, 11821–11830.
- 72 I. Drienovská, C. Mayer, C. Dulson and G. Roelfes, *Nat. Chem.*, 2018, **10**, 946–952.
- 73 Z. Zhou and G. Roelfes, *Nat. Catal.*, 2020, **3**, 289–294.
- 74 R. Ranganathan, *Cell*, 2018, **175**, 1449–1451.
- 75 T. K. Hyster and T. R. Ward, *Angew. Chem. Int. Ed.*, 2016, **55**, 7344–7357.
- 76 U. Markel, D. F. Sauer, J. Schiffels, J. Okuda and U. Schwaneberg, *Angew. Chem. Int. Ed.*, 2019, **58**, 4454 – 4464.
- 77 P. Srivastava, H. Yang, K. Ellis-Guardiola and J. C. Lewis, *Nat. Commun.*, 2015, **6**, 1–8.
- 78 S. Kato, A. Onoda, N. Taniguchi, U. Schwaneberg and T. Hayashi, *ChemBioChem*, 2020, **22**, 679–685.
- 79 J. A. McIntosh, C. C. Farwell and F. H. Arnold, *Curr. Opin. Chem. Biol.*, 2014, **19**, 126–134.
- 80 M. Jeschek, R. Reuter, T. Heinisch, C. Trindler, J. Klehr, S. Panke and T. R. Ward, *Nature*, 2016, **537**, 661–665.
- 81 M. Jeschek, S. Panke and T. R. Ward, *Trends Biotechnol.*, 2018, **36**, 60–72.
- 82 M. T. Reetz, *J. Org. Chem.*, 2009, **74**, 5767–5778.
- 83 G. Kiss, N. Çelebi-Ölçüm, R. Moretti, D. Baker and K. N. Houk, *Angew. Chem. Int. Ed.*, 2013, **52**, 5700–5725.
- 84 J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St.Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael and D. Baker, *Science*, 2010, **329**, 309–313.
- 85 T. Heinisch, M. Pellizzoni, M. Dürrenberger, C. E. Tinberg, V. Köhler, J. Klehr, D. Häussinger, D. Baker and T. R. Ward, *J. Am. Chem. Soc.*, 2015, **137**, 10414–

- 10419.
- 86 F. Van De Velde, L. Könemann, F. Van Rantwijk and R. A. Sheldon, *Biotechnol. Bioeng.*, 2000, **67**, 87–96.
- 87 I. D. Petrik, J. Liu and Y. Lu, *Curr. Opin. Chem. Biol.*, 2014, **19**, 67–75.
- 88 T. Heinisch and T. R. Ward, *Curr. Opin. Chem. Biol.*, 2010, **14**, 184–199.
- 89 G. Chang, *FEBS Lett.*, 2003, **555**, 102–105.
- 90 L. Aravind, V. Anantharaman, S. Balaji, M. M. Babu and L. M. Iyer, *FEMS Microbiol. Rev.*, 2005, **29**, 231–262.
- 91 P. K. Madoori, H. Agustiandari, A. J. M. Driessen and A.-M. W. H. Thunnissen, *EMBO J.*, 2009, **28**, 156–166.
- 92 J. L. Ramos, M. Martinez-Bueno, A. J. Molina-Henares, W. Teran, K. Watanabe, X. Zhang, M. T. Gallegos, R. Brennan and R. Tobes, *Microbiol. Mol. Biol. Rev.*, 2005, **69**, 326–356.
- 93 J. L. Ramos, M. Martí, A. J. Molina-henares, W. Tera, R. Brennan and R. Tobes, *Society*, 2005, **69**, 1–31.
- 94 H. Itou, N. Watanabe, M. Yao, Y. Shirakihara and I. Tanaka, *J. Mol. Biol.*, 2010, **403**, 174–84.
- 95 D. Fujiwara, S. Yamasaki, K. Sakurai, K. Nishino, I. Fujii, R. Nakashima and E. Nikaïdo, *Nat. Commun.*, 2013, **4**, 1–7.
- 96 S. Yamasaki, R. Nakashima, K. Sakurai, S. Baucheron, E. Giraud, B. Doublet, A. Cloeckert and K. Nishino, *Sci. Rep.*, 2019, **9**, 1–8.
- 97 S. Grkovic, K. M. Hardie, M. H. Brown and R. A. Skurray, *Biochemistry*, 2003, **42**, 15226–36.
- 98 K. M. Peters, B. E. Brooks, M. A. Schumacher, R. A. Skurray, G. Richard, M. H. Brown, R. G. Brennan and M. H. Brown, *PLoS One*, 2011, **6**, e15974.
- 99 M. A. Schumacher, M. C. Miller, S. Grkovic, M. H. Brown, R. A. Skurray and R. G. Brennan, *Science*, 2001, **294**, 2158–2164.
- 100 K. M. Peters, J. T. Schuman, R. A. Skurray, M. H. Brown, R. G. Brennan and M. A. Schumacher, *Biochemistry*, 2008, **47**, 8122–9.
- 101 M. A. Schumacher, M. C. Miller, S. Grkovic, M. H. Brown, R. A. Skurray and R. G. Brennan, *Science*, 2001, **294**, 2158–2164.

- 102 M. A. Schumacher, M. C. Miller and R. G. Brennan, *EMBO J.*, 2004, **23**, 2923–2930.
- 103 C. Gutiérrez de Souza, L. Alonso-Cotchico, M. Bersellini and G. Roelfes, *ChemRxiv*, *Prepr.*

CHAPTER 2

Optimisation of the Second Coordination Sphere of an Artificial Metallo-Hydratase bearing an Unnatural Amino Acid

Abstract

The unique activity and selectivity of metalloenzymes are due to a complex network of second coordination sphere interactions, which play a key role in the structure-function relationship. These crucial interactions, however, are difficult to predict and implement in metalloenzyme design. Here, we describe an attempt toward a better understanding of the role of the second coordination sphere by studying a previously designed artificial metallo-hydratase based on the LmrR protein with the metal-binding unnatural amino acid (UAA) (2,2'-bipyridin-5-yl)alanine incorporated in the scaffold. Introducing UAA via the expanded genetic code methodology appears as a suitable approach for this study, since it allows for a good control over the first and second coordination sphere without requiring further chemical modifications or purification of the artificial metalloenzyme. Furthermore, natural hydratases heavily rely on the second coordination sphere to perform chemically challenging water addition reactions. Therefore, hydration of enones was selected as benchmark reaction for this study. In this chapter, efforts to optimise the second coordination sphere of this artificial metalloenzyme by rational design, as well as to improve the expression protocol for unnatural amino acid incorporation are described.

2.1. INTRODUCTION

Advances in the biocatalysis field have made possible the use biological systems as catalysts. The high efficiencies, regio-, chemo- and stereoselectivities, as well as the mild reaction conditions make enzymes highly attractive both for research and industrial applications. One of the main disadvantages of enzymes over traditional synthetic chemistry is their typically intrinsically limited reaction and substrate scope, which narrows down the range of products that can be obtained.¹ In this context, artificial metalloenzymes (ArMs) have emerged as an attractive approach to access new-to-nature reactions, aiming to combine the high efficiency of enzymes and the broad substrate and reaction scope of homogeneous metal catalysts.²

2

Enzyme engineering tools are a key technology to the thriving of the ArMs field. The Darwinian approach of directed evolution for instance, comprising iterative cycles of random mutagenesis, expression, and selection has proven a powerful tool. A number of ArMs with high activities and selectivities towards non-natural reactions have thereby been evolved, emphasising the great potential of ArMs.³⁻⁶ However, this approach is limited to enzymes exhibiting basal levels of the targeted activity and the possibility of applying high-throughput screening assays. Therefore, a judicious design of the ArM is usually required to identify potential scaffold-cofactor pairs suitable for the targeted activity and to reduce the number of variants to be screened by selecting potentially beneficial mutations.⁷⁻¹⁴

The design of ArMs can be divided in first and second coordination sphere. In general, the most prominent effect on the activity is given by the first coordination sphere, that is, the functional groups that interact directly with the metal centre, i.e. the metal-coordinating ligand or metal-coordinated amino acid residues.¹⁵ The metal is responsible for conferring reactivity, while the ligand modulates the metal centre's electronic properties and can contribute to the site- and stereoselectivity. Additionally, the protein scaffold is not only involved in the direct interactions with the metal, but also provides an environment that endows the enzyme with activity and selectivity beyond the first coordination sphere. As exemplified by many natural metalloproteins, the second coordination sphere provides non-covalent interactions that can be more subtle, and therefore more difficult to predict,

but are equally important for catalysis.^{16,17} These supramolecular interactions (hydrogen bond networks, π - π interactions, and steric effects) have an important influence on: orientation of the substrate, generation of the active species, stabilisation of the transition state, promotion of electron transfer, and modulation of specificity.^{15,16,18}

Another important factor of the design is the choice of the assembly methodology by which the metal ion or complex is embedded into the protein scaffold. The approach used can be based on i) dative bonds between a residue of the protein and the metal, ii) the covalent anchoring of the ligand to the scaffold, iii) supramolecular interactions between the metal complex and the protein, or iv) introduction of an unnatural amino acid (UAA) in the protein sequence to act as ligand or catalytic residue.^{19,20} The *in vivo* incorporation of a metal-binding non-canonical amino acid into the protein *via* genetic code expansion is particularly promising: like the covalent approach, it achieves a very precise site selectivity, but does not require chemical post-expression modifications or purification, thus facilitating the optimisation of the artificial metalloenzyme.²¹

The expanded genetic code methodology consists of an engineered orthogonal tRNA/aminoacyl-tRNA synthetase (aa-tRNA) pair, which recognises an amber stop codon, the least used stop codon in *E. coli*, and incorporates the specific UAA in response.²² Whereas natural proteins are limited to the functional groups contained within the 20 canonical amino acids (occasionally post-translationally modified), this approach allows to introduce a broad diversity of functionalities that have proven suitable for a range of applications. To date, more than 150 UAA have been successfully incorporated into proteins using this methodology.^{22–28}

In the search for versatile scaffolds that can ultimately be used to create artificial enzymes towards any targeted reaction, the Lactococcal Multidrug Resistance Regulator (LmrR) has proven to be an attractive choice. LmrR's large hydrophobic pore contains two parallel tryptophans that are able to bind structurally diverse planar aromatic drugs. This characteristic has been exploited to successfully incorporate several catalytically active metal complexes using different assembly methodologies to form new ArMs.²⁹

In order to create a well-controlled active site within the protein's hydrophobic pocket, a metal binding unnatural amino acid was introduced into LmrR. Previous work of our group demonstrated the potential of incorporating (2,2-bipyridin-5-yl)alanine (bpyA) into LmrR to bind metal ions and catalyse the Friedel-Crafts alkylation reaction of indoles with conjugated enones.³⁰ The potential of this methodology was later investigated for the more challenging hydration of enones. This asymmetric water addition remains an unsolved long-standing challenge in traditional chemistry due to the difficulties involved in the use of water as nucleophile, which leads to an unfavourable equilibrium of the reaction, and the difficulty of controlling the enantioselectivity of the small and mobile water molecule.³¹ In nature, hydratases are capable of providing the right activation and environment to perform the asymmetric water addition to C=C bonds with high efficiencies and selectivities. Hydrations by both natural as well as engineered enzymes were shown to heavily rely on the second coordination sphere.^{30,32–35} Therefore, the water addition reaction was chosen as a benchmark to investigate the effect of modulating the second coordination sphere of this artificial metalloenzyme.

In previous work, rational design based on the knowledge of the reaction mechanism and computational modelling allowed to create an LmrR_bpyA-based enantioselective artificial metallo-hydratase for conjugated enones.³⁰ The design included the coordinated activation of the enone substrate by the metal-bpyA complex and the activation of the water molecule by a residue of the protein acting as general base. Computational predictions of suitable dual-active sites to be introduced into LmrR were performed and then experimentally tested. Locating the UAA in position M89 and a glutamic acid as general base in position V15, resulted in the best performing combination leading up to 75% conversion and 64% ee.³⁶ With the catalytic machinery in place, we could now focus on fine-tuning the interactions of the protein scaffold with the pre-engineered active site to optimise both the activity and enantioselectivity of this artificial metallo-hydratase.

2.2. AIM

The aim of this chapter was to rationally optimise the second coordination sphere of an enantioselective artificial metallo-hydratase based on the LmrR

protein containing the metal-binding amino acid 2,2'-bipyridylalanine. Since the enzymatic addition of water to enones is known to heavily rely on the second coordination sphere,^{30,32–35} this reaction was selected as a benchmark to investigate the effect on catalysis.

2.3. DESIGN

In this work, a strategy for optimisation of the second coordination sphere of this ArM was investigated. The optimisation involved performing site directed mutagenesis on rationally selected amino acid residues.

Figure 1. Design of the optimisation strategy of the LmrR_M89bpyA_V15E_Y based artificial metallo-hydratase. a) Model of the LmrR-bpyA complex created by manually adding the two bipyridyl moieties (pink) in the LmrR crystal structure (PDB 3F8B) with Chimera. b) Double activation of the substrate (pale yellow) by bpyA-Cu²⁺ complex (pink) and of the water molecule by V15E (cyan).³⁶ c) Close-up of the active site with docked Cu²⁺ ion (blue) coordinated to the bipyridyl moiety (pink) and the conjugated enone substrate (pale yellow). The positions selected for mutagenesis are depicted in orange and green. d) Chemical structure of the side chain of the amino acids selected for the mutagenesis studies in orange for position A11 and green for position A92.

The available structure from previous work of LmrR with docked M89bpyA-Cu²⁺-substrate (**1a**) complex (further optimised by DFT calculations)³⁶ was visually inspected and two positions were selected: All, which is in the proximity of the activating base, and A92, located between the protein scaffold and the metal complex. We hypothesised that placing residues of different sizes in these positions (see Figure 1) would influence the distance between the metal complex, the substrate and the activating base and reveal a more productive orientation for catalysis of the components of the active site. Additionally, bulkier residues might hinder the accessibility to the pocket through one of the entrances and therefore affect the enantioselectivity of the reaction. In position All, charged amino acids were also selected to investigate possible effects of electronic interactions with the activating base VI5E.

2.4. RESULTS AND DISCUSSION

Preparation of LmrR_M89bpyA_VI5E_Y mutants

We started by synthesising the unnatural amino acid bpyA according to previously reported procedures with minor optimisations.^{30,37} The six-steps synthesis in average leads to an overall yield of up to 12%, obtaining up to 6 g of final product of the racemic bpyA (see Materials and Methods for more information about the synthesis and its optimisation).

The unnatural amino acid bpyA was incorporated at position M89 of LmrR using amber stop codon suppression. The LmrR variant used contains a C-terminal Strep-tag and has two mutations in the DNA binding domain (K55D and K59Q) that improve purification.³⁴ From a LmrR_M89bpyA_VI5E gene available from previous work, the selected mutants were prepared using QuikChange mutagenesis and subsequently expressed in *E. coli* BL21(DE3) in the presence of 1 mM bpyA. The proteins were purified by affinity chromatography using a Strep-tactin Sepharose column and characterised by SDS-PAGE and exact mass.

The expression yield of LmrR_M89bpyA_VI5E was around 5 mg per litre of culture, while the yield of the triple mutants varied between 0 and 3.5 mg. When investigating the cause of this unsatisfactory yield by performing SDS-PAGE, the expected band at around 15 kDa corresponding to LmrR was found,

but also a prominent band at around 10 kDa. This size corresponds to truncated LmrR, which occurs when the amber stop codon is recognised by release factors, which leads to termination of translation. Since the truncated protein is cut at the C-terminus, it does not contain the Strep-tag, yet, it is still found after purification mixed with fully expressed LmrR. This can be explained by the dimerisation of one truncated with one full length LmrR monomers, which are purified together (Figure 2).

Figure 2. Schematic explanation of the mixed protein sample containing full length LmrR and truncated LmrR obtained after expression and purification due to failure on incorporating the UAA in position M89.

Optimisation of the expression protocol

One of the drawbacks of large-scale *in vivo* incorporation of unnatural amino acids is the requirement for high concentrations of UAA (typically 1 mM or higher), which are frequently expensive or not commercially available at all, as is the case of bpyA. This has to be prepared in a time consuming multistep synthesis.³⁸ Due to the low expression yield and the presence of truncated protein in the system, only a small fraction of the media-added bpyA was finally incorporated into LmrR (less than 0.1% of the added UAA) and a considerable amount of the amino acid thus remained as waste in the expression culture media. The residual UAA was also clearly noticeable by the dark pink colour of the culture after expression, which can be attributed to the presence of free bpyA coordinated to metal ions. Therefore, we considered recovering the remaining UAA by reusing the culture media after expression. As the recycled media would then already contain the inducers, it would not allow the cells to grow before starting expression. To circumvent this issue, an auto-induced expression by inducer repression system was used, following a protocol reported by the Lee group.³⁸

For incorporation of UAA we typically use a two-plasmid system comprising pET (containing LmrR with the amber stop codon at the desired position) and pEVOL (containing the tRNA/aa-tRNA synthetase orthogonal pair for incorporation of the UAA under the control of *araBAD* promoter derived from pBAD). Expression systems like pET and pBAD (or pEVOL, in this case), contain a T7 and an *araBAD* promoters, respectively, and are usually meant to produce high levels of protein expression in presence of the adequate inducers (IPTG for pET and L-arabinose for pBAD). The vectors used here contain a *lac* operator that avoids basal levels of expression in the absence of IPTG and arabinose. Because these promoters also require cyclic adenosine monophosphate (cAMP) and the cAMP receptor protein complex (CRP) for transcription initiation, their expression can be regulated by the amount of glucose present. In a glucose-controlled expression, glucose and lactose (in pET/*lac*) or glucose and L-arabinose (in *araBAD*) are present in the media. High levels of glucose decrease the amount of cAMP by catabolite repression, which blocks the uptake of other sugars and inducers and thus prevents induction. When glucose is used up, the expression is induced by lactose (in pET) and arabinose (pBAD).^{39–41} Although pBAD is typically used in strains lacking L-arabinose metabolism to avoid depletion of the inducer, BL21(DE3) was here used as expression host and the issue was counteracted by using a higher concentration of L-arabinose.

In Lee's work, the pET/pEVOL plasmid system was used for the production of several proteins containing different UAAs (including bpyA). Minimal media (R-media) was used for auto-induced glucose-controlled expression from which the UAAs were recycled.³⁹ After each expression, the media containing the UAA was sterile filtered and reused for a new cycle of expression upon being supplemented with an optimised amount of essential nutrients. With this approach, it was possible to do up to five rounds of expression of proteins with different UAAs incorporated without a significant decrease in yield.³⁸

Therefore, R-media contains optimised amounts of glucose that allow the delay of the expression until the cells are optimally grown and glucose is consumed. Glycerol as the major carbon source, $(\text{NH}_4)_2\text{SO}_4$ to supply nitrogen and sulphur, MgSO_4 to supply Mg^{2+} , and a mixture of trace metals. The media also contains a mixture of amino acids to facilitate cell growth and

phosphate buffer to avoid acidification of the media. Finally, the UAA (only in the first cycle), inducers of the pET and the pEVOL systems (lactose and arabinose respectively),³⁸ and the required antibiotics (ampicillin and chloramphenicol) are added prior to inoculation of the culture.

Table 1. Composition of R-media for recycling UAA approach slightly adapted from Lee's paper.³⁸

Entry		Na ₂ HPO ₄ (mM)	KH ₂ PO ₄ (mM)	(NH ₄) ₂ SO ₄ (mM)	MgSO ₄ (mM)	Trace metal mixture ^a (%)	Glycerol (%)	Glucose (%)	amino acids ^b (%)	Lactose and arabinose (%)
1	Initial composition	50	50	25	2	0.1	0.5	0.05	5	0.2
2	Supplement 1	50	50	25	2	0.1	0.5	0.05	5	0.2
3	Supplement 2	10	10	5	0.4	0.01	0.25	0.05	1	0.02
4	Supplement 3	10	10	5	0.4	0.01	0.25	0.05	1	0.1
5	Supplement 4 _≤	10	10	5	0.4	0.01	0.25	0.05	2.5	0.02
^a Trace metal mixture stock: 10 mM FeCl ₃ , 4 mM CaCl ₂ , 2 mM each of MnCl ₂ and ZnSO ₄ , and 0.4 mM each of CoCl ₂ , CuCl ₂ , NiCl ₂ , Na ₂ MoO ₄ , Na ₂ SeO ₃ , and H ₃ BO ₃ . ^b 19 amino acids stock of about 4 mg/mL of each amino acid.										

In order to establish the best conditions for our LmrR-bpyA system, the best expressing mutant, LmrR_M89bpyA_VI5E_A92I, was chosen for the optimisation study. A cell growth experiment was conducted in order to establish the right duration of each expression cycle. The cells were grown in R-media with the initial composition (Table 1 entry 1) and the growth was monitored by measuring the OD_{600nm} over time (Figure 3).

Since the cells grow approximately 25-30 hours before reaching stationary phase, we decided that each cycle of expression would last a convenient 24 hours, after which the cells would be harvested, and the media sterile filtered and used for a new round of expression.

Figure 3. Cell growth of *E. coli*. BL21(DE3) producing LmrR_M89bpyA_V15E_A92I in R-media. The black and grey lines correspond to two independent experiments.

The effects of three parameters (temperature, arabinose and bpyA concentration) were tested in order to optimise the expression for the LmrR-bpyA system. The cell growth of *E. coli* containing LmrR_M89X_V15E_A92I under four different conditions (Table 2) was monitored (Figure 4b). After 4, 7 and 22 hours a sample of each experiment was taken and the soluble proteins were loaded onto an SDS-PAGE gel to compare expression levels under these conditions (Figure 4a).

Table 2. Expression conditions.

Condition	Temperature (°C)	Arabinose (% w/v)	bpyA (mM)
Initial	37	0.2	1
1	37	0.2	0.5
2	37	0	1
3	37	0.02	1
4	30	0.02	1

Figure 4. a) SDS-PAGE showing the soluble fraction from LmrR_M89bpyA_V15E_A92I grown under conditions 1 to 4 at 5, 7, and 22 hours. b) OD_{600nm} measured over time for the experiments under the 4 described conditions.

Whereas the cell growth was not influenced by the decrease of UAA concentration, more fragmented than full length LmrR was found, indicating that higher amount of bpyA results in a better incorporation of the UAA. The control in absence of the inducer arabinose led to, as expected, no incorporation of bpyA, evidenced by the lack of the full length LmrR band in the gels. On the other hand, with 0.02% arabinose a band a significantly more prominent at 15 kDa than at 10 kDa was observed. However, the cell growth seemed to decrease compared to the initial arabinose concentration (10 times higher) and, thus, an arabinose concentration of 0.2% was maintained.

Additionally, the effect of the temperature was studied. At 30 °C the cell growth was slower and a higher proportion of the truncated protein compared to full length LmrR was observed. Whereas at 37 °C the culture reached the plateau in 25-30 hours leading to a higher proportion of full length LmrR.

With this information at hand we proceeded to test the expression protocol with the LmrR-bpyA mutants using the best conditions, 37 °C, 0.2% arabinose and 1 mM bpyA. LmrR_M89bpyA_V15E_A92I was expressed for 7 cycles. The overall expression yield was 13 mg/L, almost four times more than with LB. Therefore, the protocol with the other mutants that presented lower expression yields in LB media were tested. Regrettably, however, after three cycles of expression the overall yield was 5 to 20 times lower than with one expression in LB (Figure 6) for these protein variants.

Figure 5. Electrophoresis gel of the purification of LmrR_M89bpyA_V15E_A92I expressed in LB and for 7 cycles with R-media.

These results led us to conclude that, whereas recycling the UAA from the culture media can be an alternative to reduce the amount of UAA wasted during expression, the suitability of this approach is strongly protein dependent.

The mutants that were obtained in sufficient quantities (LmrR_M89bpyA_V15E, LmrR_M89bpyA_V15E_A92I and LmrR_M89bpyA_V15E_A11Q), were then tested in catalysis.

Catalysis

The catalytic activity of the LmrR-bpyA based artificial metalloenzymes with Cu(II) ions, was evaluated in the asymmetric 1,4-addition of water to α,β -unsaturated 2-acyl pyridines. The artificial metalloenzymes were prepared by incubating Cu(NO₃)₂ (90 μ M) with a slight excess of the LmrR_bpyA mutants

(112.5 μ M) for 30 minutes prior to catalysis. The reactions were initiated by addition of 1 mM of the corresponding substrate (**1a** or **1b**). The experiments were carried out for 3 days at 4 °C in 20 mM MOPS buffer, pH 6.5 and 150 mM NaCl. Afterwards the products were extracted and analysed by chiral HPLC.

The artificial metallo-hydratase was first tested on the *tert*-butyl substituted enone **1a**, which in previous work led to 64% conversion and 50% ee using LmrR_M89bpyA_VI5E.³⁶

As expected, due to the poor nucleophilicity of water, the uncatalysed reaction led to insignificant yields, corroborating previously reported results.^{30,34} The reaction in presence of Cu(NO₃)₂, but without protein led to 80% yield and racemic product (see Table 3 entry 2), illustrating that the electrophilic activation of the enone substrate by Cu²⁺ ions is necessary. Interestingly, the bulkier isoleucine in position A92 led to a similar rate acceleration than the previously reported LmrR_M89bpyA_VI5E and an increase of the enantioselectivity (49% yield and 72% ee, see Table 3 entry 4). This result could support our design hypothesis suggesting that a bulkier residue can partially block one of the entrances of the pocket and hence favour the formation of one of the product enantiomers (**2a**) or force a position of the ligand-Cu²⁺-substrate complex in which one of the enantiotopic faces of the substrate is more accessible than the other. On the contrary, mutating All to glutamine resulted in a decrease of the yield to 12% and the ee to 42% (see Table 3 entry 5), which might suggest that a bulkier group on this position could hinder the entrance of the substrate in the hydrophobic pore and thus be prejudicial for the reaction. To investigate whether the positive effect on ee of mutation A92I was substrate depended, the reaction was also tested with substrate **1b**. A similar effect of the mutation A92I was observed; although the yield was not affected by the mutation, the enantioselectivity increased from 58 to 80% (Table 3 entries 6 and 7). However, additional experiments with a different batch of protein, led to inconsistent results with both substrates **1a** and **1b**. The protein used in this case contained a higher amount of truncated LmrR and, therefore, these results were discarded. Nevertheless, this evidenced the difficulty introduced by varying proportions of truncated protein in different expression batches, which hampers an accurate determination of protein concentration.

Table 3. Results of catalysis for the hydration reaction of substrates **1a** and **1b**.

Entry	Catalyst	Substrate	Yield (%)	ee (%)
1 ^a	-	1a	4 ± 1	n.d.
2	Cu(NO ₃) ₂	1a	80 ± 12	0 ± 1
3 ³⁶	Cu ²⁺ -LmrR_M89bpyA_V15E	1a	64 ± 8 ^b	50 ± 2
4	Cu ²⁺ -LmrR_M89bpyA_V15E_A92I	1a	49 ± 15	72 ± 8
5	Cu ²⁺ -LmrR_M89bpyA_V15E_A11Q	1a	12 ± 5	42 ± 1
6	Cu ²⁺ -LmrR_M89bpyA_V15E	1b	78 ± 10	58 ± 3
7	Cu ²⁺ -LmrR_M89bpyA_V15E_A92I	1b	74 ± 3	80 ± 3

Conditions: 90 μM Cu(NO₃)₂ (9 mol %) loading with 1.3 eq of protein (120 μM) and 1 mM of substrates **1** in 20 mM MOPS buffer pH 7.0, 150 mM NaCl, for 3 days at 4 °C. All the results listed correspond to the average of two independent experiments, each carried out in duplicate. Errors listed are standard deviations. Yields were determined by HPLC and using 2-phenylquinoline as internal standard. ^aAverage of two experiments. ^bConversion (%) reported instead of yield.

2.5. CONCLUSIONS AND PERSPECTIVES

Here we have presented that the LmrR_bpyA based artificial metallo-hydratase has potential for rationally optimising the second coordination sphere. This claim is exemplified by mutant A92I, which exhibited a significantly increased ee of up to 80%. For a deeper understanding of the system, more mutants should be tested. However, a long and expensive synthesis, as well as the poor expression yield of the mutants impaired a more complete study and therefore the limitations of the system were investigated.

2,2'-(bipyridin-5yl)alanine's feasibility as UAA for creating ArMs has been proven by its incorporation into LmrR- and QacR-based ArMs.^{30,36,37} The limitation with bpyA based ArMs is the lack of a fast and economic synthetic route to obtain the amino acid. Step 3 (see synthesis in materials and methods) of the synthesis used here is particularly challenging since it

involves bromination *via* a radical reaction, which cannot be further scaled up and suffers from poor reproducibility. Moreover, the original protocol described the use of CCl_4 as solvent, which is nowadays difficult to purchase due to its toxicity. For that reason, CCl_4 was changed to benzene, which resulted in a slower reaction (more than 24 hours in average compared to the 6 hours with CCl_4) and led to lower reaction yields. On the other hand, targeting the L-amino acid instead of a racemic mixture, would increase the efficiency of the reaction since the tRNA/aa-tRNA synthetase could incorporate all the fed product rather than only half of it. Therefore, an alternative and more effective synthetic pathway to prepare pure L-bpyA would be advisable to facilitate conducting experiments with this unnatural amino acid in an efficient way.

The introduction of one extra mutation required for the optimisation of the artificial metallo-hydratase, decreased the expression yield to a level that complicated the production of enough protein to perform catalysis. The protein obtained after purification contained a fraction of truncated LmrR where bpyA was not incorporated. This recognition of the stop codon by release factors led to low expression yields and impaired the estimation of protein concentration of this ArM, producing variations of the results with some batches of protein.

In an effort to tackle low synthesis yields and incorporation efficiencies, the potential of recycling the UAA from the culture media was investigated by optimising an auto-induced glucose regulated media. Whereas this approach allowed to obtain almost 4 times more protein than with a regular expression in LB for the best expressing mutant A92I, the success of this methodology proved to be protein dependent. The other mutants—which showed lower expression yield with LB—led to even less amount of protein after several cycles of expression with the auto-induced media.

Several alternatives to try to circumvent the incorporation and expression difficulties could be explored. For example, the use of BL21(AI) as expression strain, which contains a T7 RNA polymerase controlled by the tightly regulated arabinose-inducible P_{BAD} promoter.⁴² With BL21(AI) both pET and pBAD plasmids could be induced by arabinose, and, unlike BL21(DE3), this strain does not use arabinose as nutrient and thus the cell growth does not

depend on the arabinose concentration. Another option could be the use of *E. coli* strains in which the translation release factor RFI has been knocked out and is hence less likely to recognise TAG as stop codon.⁴³ Moreover, strains with fully synthetic genome that do not contain any amber stop codon have recently been developed.⁴⁴ These strains ensure that the UAA is not incorporated in any host proteins and hence facilitate cell growth.

Overall, despite the difficulties experienced due to the synthesis of the UAA and the low protein expression, the work presented in this chapter suggests a potential for rationally selecting a small number of mutations to gain insights into the structure-function relationship of the ArM and, hence, to contribute to expand the knowledge of this field which is necessary both for a better understanding of biological systems and for the design of new enzymes.

2

2.6. EXPERIMENTAL SECTION

General remarks

Chemicals and reagents were purchased from commercial suppliers (Sigma Aldrich, Acros or TCI europe) and used without further purification. Flash chromatography was performed using Reveleris® X2 Flash Chromatography System. ¹H-NMR spectra were recorded on a Varian 400 (400 MHz) spectrometer in CDCl₃. Chemical shifts values (δ) are denoted in ppm using residual solvent peaks as the internal standard (CHCl₃: δ 7.26 for 1H). ¹H-NMR results are reported in the conventional form: chemical shifts, multiplicity (d = doublet, dd = doublet of doublet, m=multiplet, q=quartet, s=singlet, t=triplet), coupling constants (Hz), and integration. Enantiomeric excess determinations were performed by HPLC analysis using UV-detection (Shimadzu SCL-10Avp) on Chiralpak ADH, n-heptane:iPrOH 99:1, 0.5 ml/min. UPLC-MS on protein samples was performed on a Acquity TQ Detector (ESITQD- MS) coupled to Waters Acquity Ultra Performance LC using a Acquity BEH C8 (1.7 μm 2.1 x 150 mm). Water (solvent A) and acetonitrile (solvent B) containing 0.1% formic acid by volume, were used as the mobile phase at a flow rate of 0.3 mL/min. Gradient: 90% A for 2 min, linear gradient to 50% A in 2 min, linear gradient to 20% A in 5 min, followed by 2 min at 5% A and re-equilibration of the column with 2 min at 90% A. *E. coli* strain NEB5α-T1 Phage Resistant (Stratagene) was used for cloning and *E. coli* BL21(DE3)-T1 Phage Resistant (Stratagene) strain was used for protein

expression. Primers were synthesised by Eurofins MWG Operon (Ebesberg, Germany). Pfu Turbo polymerase was purchased from Agilent and DpnI was purchased from New England Biolabs. Plasmid isolation Kit was purchased from Roche. DNA sequencing was carried out by Eurofins Genomics. The concentration of the proteins was measured with Nanodrop 2000 (Thermo Scientific). The amino acid stock solution used for R-media was prepared from a MEM amino acid solution (50X), MEM Non-essential amino acid solution (100X) and L-glutamine solution (200 mM), all purchased from Thermo Fisher Scientific. Trace metals solution was purchased from Teknova.

Synthesis

(2,2-bipyridin-5-yl)alanine (bpyA) was synthesised according to literature procedures. A typical experiment started from 24 mL of 2-acetyl pyridine (213 mmol) and led to an overall yield of up to 12% of the racemic product. Step 3 was the most problematic since, being a radical reaction, it is difficult to scale up and the outcome substantially varied from batch to batch. In the previously reported reaction,^{45,46} CCl_4 was used as solvent, which commercially availability decreased due to its harmful properties. Therefore, benzene was used instead. This solvent usually led to longer reaction times (from 6 to about 24 h) and lower reaction yields.

1-[2-oxo-2-(pyridin-2-yl)ethyl]pyridinium iodide⁴⁵

¹H-NMR (400 MHz, DMSO-*d*₆) δ 9.11 – 8.66 (m, 3H), 8.64 – 8.43 (m, 1H), 8.37 – 8.01 (m, 3H), 7.99 – 7.79 (m, 1H), 7.77 – 7.48 (m, 1H), 6.36 (s, 2H).

5-methyl-2,2'-bipyridine⁴⁵

¹H-NMR (400 MHz, Chloroform-*d*) 8.63 – 8.68 (m, 1H), 8.50 (s, 1H), 8.34 (d, *J* = 8.1 Hz, 1H), 8.27 (d, *J* = 8.1 Hz, 1H), 7.82 – 7.75 (m, 1H), 7.62 (d, *J* = 8.1 Hz, 1H), 7.29 – 7.24 (m, 1H), 2.38 (s, 3H).

5-(bromomethyl)-2,2'-bipyridine^{45,46}

¹H-NMR (400 MHz, Chloroform-*d*) δ 8.69 – 8.64 (m, 2H), 8.42 – 8.36 (m, 2H), 7.88 – 7.76 (m, 2H), 7.34 – 7.28 (m, 1H), 4.53 (s, 2H).

5-(dibromomethyl)-2,2'-bipyridine⁴⁶

¹H-NMR (400 MHz, Chloroform-*d*) δ 8.78 (d, *J* = 2.4 Hz, 1H), 8.72 – 8.68 (m, 1H), 8.45 (dd, *J* = 22.2, 8.2 Hz, 2H), 8.10 (dd, *J* = 8.4, 2.5 Hz, 1H), 7.87 – 7.80 (m, 1H), 7.38 – 7.31 (m, 1H), 6.73 (s, 1H).

diethyl 2-([2,2'-bipyridin]-5-ylmethyl)-2-acetamidomalonate⁴⁷

¹H-NMR (400 MHz, Chloroform-*d*) δ 8.62 – 8.58 (m, 1H), 8.33 – 8.23 (m, 3H), 7.80 – 7.71 (m, 1H), 7.44 – 7.24 (m, 1H), 7.29 – 7.21 (m, 1H), 6.51 (s, 1H), 4.25 – 4.15 (m, 4H), 3.66 (s, 2H), 1.99 (s, 3H), 1.26 – 1.20 (m, 6H).

3-([2,2'-bipyridin]-5-yl)-2-ammoniopropanoate⁴⁷

¹H-NMR (400 MHz, D₂O) δ 8.70 – 8.66 (m, 1H), 8.62 – 8.58 (m, 1H), 8.48 – 8.38 (m, 2H), 8.17 (d, J = 8.3 Hz, 1H), 8.00 (d, J = 8.3, 1H), 7.88 – 7.83 (m, 1H), 4.14 (t, J = 6.7 Hz, 1H), 3.28 (m, 2H).

Synthesis of substrates for the water addition reaction

Several substrates were synthesised with the aim of exploring the substrate scope of the reaction, however, only the substrates mentioned in the catalysis section were tested. The procedure was adapted from the literature.⁴⁸ To 30 mL of dry toluene 1.83 g (10.4 mmol) of 3-(dimethylamino)-1-(2-pyridyl)-2-propen-1-one was added and the solution was cooled to 0 °C under nitrogen. 12.5 mmol (1.25 eq) of a 2 M solution of the corresponding alkylmagnesium chloride in THF was added slowly and the reaction mixture was stirred for 16 h at room temperature and the solvent was then evaporated under vacuum. The oil obtained was dissolved in 150 mL dichloromethane and hydrolysed with 250 mL water under vigorous stirring. 2 g of sodium bicarbonate was added and the organic phase was collected. The aqueous phase was extracted three times with dichloromethane (50 mL). The combined organic phases were washed successively with a saturated bicarbonate solution, water and a saturated sodium chloride solution and then dried over sodium sulphate. The solvent was concentrated under vacuum and purified by silica flash column chromatography with a EtOAc:pentane gradient (0:100 10 min; 30:70 40 min) yielding between 50 and 76% of the corresponding products.

(E)-1-(pyridin-2-yl)oct-2-en-1-one

$^1\text{H-NMR}$ (400 MHz, Chloroform-*d*) δ 8.72 – 8.67 (m, 1H), 8.13 – 8.08 (m, 1H), 7.87 – 8.80 (m, 1H), 7.61 – 7.84 (m, 1H), 7.47 – 7.41 (m, 1H), 7.29 – 7.2 (m, 1H), 2.37 – 2.29 (m, 2H), 1.58 – 1.48 (m, 2H), 1.38 – 1.26 (m, 4H), 0.96 – 0.81 (m, 3H).

(E)-3-cyclohexyl-1-(pyridin-2-yl)prop-2-en-1-one

$^1\text{H-NMR}$ (400 MHz, Chloroform-*d*) δ 8.72 – 8.66 (m, 1H), 8.14 – 8.09 (m, 1H), 7.87 – 7.81 (m, 1H), 7.55 (dd, $J = 15.7$, 1.4 Hz, 1H), 7.8 – 7.43 (m, 1H), 7.27 – 7.16 (m, 1H), 2.56 – 2.66 (m, 1H), 1.18 – 1.11 (m, 10H).

(E)-4-methyl-1-(pyridin-2-yl)pent-2-en-1-one (1b)⁴⁸

$^1\text{H-NMR}$ (400 MHz, Chloroform-*d*) δ 8.69 – 8.62 (m, 1H), 8.05 – 7.94 (m, 2H), 7.83 – 7.77 (m, 1H), 7.47 – 7.40 (m, 1H), 7.20 – 7.14 (m, 1H), 2.39 – 2.32 (m, 1H), 1.11 (d, $J = 6.7$ Hz, 6H).

(E)-4,4-dimethyl-1-(pyridin-2-yl)pent-2-en-1-one (1a)

Procedure adapted from literature.⁴⁹ To 75 mL of water was added 1.5 g (12 mmol) 2-acetylpyridine and 1.00 g (11.6 mmol) pivaldehyde at 0 °C. 7.5 mL of a 10% NaOH solution was then added and the mixture was shaken vigorously and left undisturbed overnight at 4 °C. The aqueous phase was extracted three times with dichloromethane (50 mL). The combined organic phases were washed successively with a saturated bicarbonate solution, water and a saturated sodium chloride solution and then dried over sodium sulphate. The solvent was evaporated under vacuo, and the product was purified by flash column chromatography with an EtOAc:pentane gradient 0:1-3:7 yielding pure **1a**. $^1\text{H-NMR}$ (400 MHz, Chloroform-*d*) δ 8.70 (m, 1H), 8.11 (m, 1H), 7.84

(m, 1H), 7.54 (d, $J = 15.9$ Hz, 1H), 7.46 (m, 1H), 7.24 (d, $J = 15.9$ Hz, 1H), 1.16 (s, 9H).

General procedure for the preparation of the water addition products as reference compounds^{34,49}

To 1 L of 20 mM MOPS buffer pH 6.5 was added 200 mg of the substrate (pre dissolved in 5 mL CH₃CN) and 60 mg of Cu(1,10-phenanthroline)(NO₃)₂. The reaction was stirred for 12 hours at room temperature and the product was extracted with diethyl ether (3 x 50 mL), washed with saturated sodium chloride and dried over sodium sulphate. The solvent was evaporated under vacuum and the crude product was purified by column chromatography.

3-hydroxy-4,4-dimethyl-1-(pyridin-2-yl)pentan-1-one (2a)

The ee was determined by HPLC analysis (Chiralpak-ADH, n-heptane:iPrOH 99:1, 0.5 ml/min). Retention times: 66.3 min and 75.6. min.

3-hydroxy-4-methyl-1-(pyridin-2-yl)pentan-1-one (2b)

3-hydroxy-1-(pyridin-2-yl)octan-1-one

3-cyclohexyl-3-hydroxy-1-(pyridin-2-yl)propan-1-one **Molecular biology: Site-directed mutagenesis**

The mutants were prepared by site-directed mutagenesis (QuikChange) from available pET17b plasmid encoding for LmrR_LM_M89X_V15E (where LM refers to K55D and K59Q mutations for purification purposes and X to the stop codon TAG at position 89). The primers used for mutagenesis are listed in Table 4. *Pfu* Turbo DNA polymerase was used. After an initial denaturation at 95 °C for 1 min, the following cycles were repeated 16 times: denaturation at 98 °C for 30 s, annealing at 52–63 °C for 30 s, elongation at 72 °C for 4 min and 30 s. The final elongation was performed at 72 °C for 10 min. The resulting PCR products were digested with restriction endonuclease DpnI for 2 h at 37 °C and 5 µL of the mixture were directly transformed into *E. coli* NEB5α competent cells. After overnight growth, single colonies were used to inoculate 5 mL of LB media. Plasmids were isolated and successful mutagenesis were confirmed by Sanger sequencing (T7 forward primer).

Table 4. Primers used for site-directed mutagenesis.

Primer	Sequence (5' → 3')
LmrR_LM_M89X_V15E_A11F_fw	GAAATGCTGCGTTTCCAAACCAATGAA
LmrR_LM_M89X_V15E_A11F_rv	TTCATTGGTTTGGAACGCAGCATTTC
LmrR_LM_M89X_V15E_A11V_fw	GAAATGCTGCGTGTTCAAACCAATGAA
LmrR_LM_M89X_V15E_A11V_rv	TTCATTGGTTTGAACACGCAGCATTTC
LmrR_LM_M89X_V15E_A11R_fw	GAAATGCTGCGTCGTCAAACCAATGAA
LmrR_LM_M89X_V15E_A11R_rv	TTCATTGGTTTGACGACGCAGCATTTC
LmrR_LM_M89X_V15E_A11Q_fw	GAAATGCTGCGTCAACAAACCAATGAA
LmrR_LM_M89X_V15E_A11Q_rv	TTCATTGGTTTGTTGACGCAGCATTTC
LmrR_LM_M89X_V15E_A11H_fw	GAAATGCTGCGTCATCAAACCAATGAA

LmrR_LM_M89X_VI5E_A11H_rv	TTCATTGGTTTGTATGACGCAGCATTTC
LmrR_LM_M89X_VI5E_A92F_fw	AACTAGCGCCTGTTCTTTCGAATCCTGG
LmrR_LM_M89X_VI5E_A92F_rv	CCAGGATTCTGAAGAACAGGCGCTAGTT
LmrR_LM_M89X_VI5E_A92R_fw	AACTAGCGCCTGCGTTTTCGAATCCTGG
LmrR_LM_M89X_VI5E_A92R_rv	CCAGGATTCTGAAACGCAGGCGCTAGTT
LmrR_LM_M89X_VI5E_A92V_fw	AACTAGCGCCTGGTTTTTCGAATCCTGG
LmrR_LM_M89X_VI5E_A92V_rv	CCAGGATTCTGAAAACCAGGCGCTAGTT
LmrR_LM_M89X_VI5E_A92S_fw	AACTAGCGCCTGTCTTTTCGAATCCTGG
LmrR_LM_M89X_VI5E_A92S_rv	CCAGGATTCTGAAAGACAGGCGCTAGTT
LmrR_LM_M89X_VI5E_A92I_fw	AACTAGCGCCTGATCTTTCGAATCCTGG
LmrR_LM_M89X_VI5E_A92I_rv	CCAGGATTCTGAAGATCAGGCGCTAGTT

Protein expression and purification

pETI7b plasmids containing the LmrR_LM_M89X_VI5E_Y mutants (where Y refers to extra mutations) and pEVOL plasmid containing the tRNA/aa-tRNA synthetase for incorporation of bpyA, kindly provided by Schultz group, were co-transformed into *E. coli* BL21(DE3). A single colony was used to inoculate 5 mL of fresh LB media with 100 µg/mL of ampicillin (Amp) to select for the pETI7b plasmid and 34 µg/mL of chloramphenicol (Chl) to select for the pEVOL plasmid. The culture was grown at 37 °C overnight. In a typical expression in LB media, 2 mL of the overnight culture were used to inoculate 500 mL of fresh LB medium with 100 µg/mL Amp and 34 µg/mL Chl that were incubated at 37 °C. When the adequate optical density at 600 nm of 0.8-0.9 was reached, the synthesised racemic mixture of bpyA to a final concentration of 1 mM was added and the expression of the targeted proteins was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) (final concentration of 1 mM) and L-arabinose (final concentration 0.02%). Protein expression was performed at 30 °C overnight. Cells were harvested by centrifugation (6000 rpm, JA-10, 20 min, 4 °C, Beckman) and the pellet was resuspended in 50 mM NaH₂PO₄, pH 8.0, 150 mM NaCl and protease inhibitor cocktail (Complete, Roche, 1 tablet, for 50 mL of resuspension buffer). Cells were sonicated (70% (200 W) for 15 min (10 sec on, 10 sec off)

and afterwards incubated for 30 min on ice with PMSF solution (final concentration 0.1 mM), DNaseI (0.1 mg/mL) and 10 mM MgCl₂. After centrifugation (16000 rpm, JA-17, 45 min, 4 °C, Beckman), the supernatant was equilibrated with 5 mL of pre-equilibrated Strep-tag Tactin column material for 1 h (mixed at 200 rpm on a rotatory shaker) at 4 °C. The column was washed with 3 x CV (column volume) of the re-suspension buffer, and the protein was eluted with five IX CV of resuspension buffer containing 5 mM desthiobiotin. Elution fractions were analysed on a 12% polyacrylamide SDS-Tris Tricine gel followed by Coomassie staining (InstantBlue, Expedeon). Fractions containing protein were collected and concentrated using Vivaspin Turbo (5000 MWCO, Sartorius) centrifugation filters and they were dialysed twice against the reaction buffer containing 20 mM MOPS and 250 mM NaCl at pH 6.5. Protein concentration was measured on a Nanodrop 2000 (Thermo Scientific), using the calculated extinction coefficient $\epsilon_{280}=25,440 \text{ M}^{-1}\text{cm}^{-1}$. Typical expression yields were between 0 and 3.5 mg/L.

The expression when bpyA was recycled from the culture media was performed using the following protocol: 2 mL of overnight culture in LB with 100 µg/mL Amp and 34 µg/mL Chl was used to inoculate 500 mL of the starting R-media (Table 1 entry 1), containing 1 mM bpyA final concentration and the appropriate antibiotics. The culture was grown for 24 h at 37 °C. Subsequently, the cells were harvested by centrifugation (6000 rpm, JA-10, 20 min, 4 °C, Beckman), the pellet was stored at -20 °C for further purification following the above-mentioned protocol and the culture media was sterile filtered for use in the next cycle. Before each cycle of expression, the corresponding components listed in Table 1 were added, as well as fresh antibiotics and 2 mL of LB overnight culture.

Monitoring growth and expression under different conditions

20 mL of the initial media (Table 1 entry 1) were inoculated with 40 µL of overnight culture in LB media with ampicillin and chloramphenicol and the specified amount of bpyA. This experiment was conducted 4 times in parallel, each of them under one of the conditions described in Table 2. After 5, 10 and 22 hours, the OD_{600nm} was measured and 500 µL of sample of each culture were taken. The samples were centrifuged and the cell pellets were frozen. The cells were then thawed, resuspended with 50 µL of Bugbuster and

incubated for 1 hour at room temperature. Subsequently, the cell debris were removed by centrifugation and the supernatants were loaded onto an SDS-PAGE gel (see Figure 4).

Characterisation

SDS-PAGE

Figure 6. SDS-PAGE of LmrR_M89bpyA_V15E mutants after Strep-Tag purification. L: protein ladder (Unstained PageRuler Broad Range Protein Ladder), FT: flow through, sol: soluble fraction; ins: insoluble fraction W: wash fraction; E: elution fraction.

Exact mass

Protein	MW _{observed} (Da)	MW _{calculated} (Da)
LmrR_LM_M89X_V15E	15094.63	15094.59
LmrR_LM_M89X_V15E_A92I	15136.69	15136.13
LmrR_LM_M89X_V15E_A11F	15170.70	15170.69
LmrR_LM_M89X_V15E_A11Q	14999.66 (-Met)	15001.44 (-Met)
LmrR_LM_M89X_V15E_A11E	15155.10	15152.63

LmrR_LM_M89X_VI5E_AIIH	15160.6	15160.66
------------------------	---------	----------

Mutants A92F, A92R, A92S, A92V and AIIV were not obtained in a sufficient amount and purity to acquire a good mass spectrum.

Catalysis

The corresponding amount of reaction buffer (20 mM MOPS buffer, 150 mM NaCl at pH 6.5), 10 μ L of a 2.7 mM stock solution of $\text{Cu}(\text{NO}_3)_2 \cdot 3\text{H}_2\text{O}$ and the protein solution (0.11 mM final concentration) were added to a final volume of 140 μ L. The reaction mixture was incubated under continuous inversion for 30 minutes at 4 $^\circ\text{C}$ to ensure binding of the metal into the bpyA active site. Subsequently, 10 μ L of a 30 mM stock solution of substrate **1a** or **1b** in 5% acetonitrile was added and the reaction was incubated for three days. Thereafter, 50 μ L of a 1 mM solution of 2-phenylquinoline (internal standard) were added. The reaction mixture was extracted with 1 mL of diethyl ether three times and the organic layer was dried over MgSO_4 and evaporated in vacuo. The mixture was then re-dissolved in 150 μ L of heptane:isopropanol 9:1 and analysed by chiral HPLC. The yields of the reactions were determined using calibration curves based on peak area at 268 nm using 2-phenylquinoline as internal standard.

2.7. REFERENCES

- 1 M. T. Reetz, *J. Am. Chem. Soc.*, 2013, **135**, 12480–12496.
- 2 H. J. Davis and T. R. Ward, *ACS Cent. Sci.*, 2019, **5**, 1120–1136.
- 3 M. Jeschek, S. Panke and T. R. Ward, *Trends Biotechnol.*, 2018, **36**, 60–72.
- 4 M. T. Reetz, *J. Org. Chem.*, 2009, **74**, 5767–5778.
- 5 F. H. Arnold, *Angew. Chem. Int. Ed.*, 2018, **57**, 4143–4148.
- 6 R. E. Cobb, R. Chao and H. Zhao, *AIChE J.*, 2013, **59**, 1432–1440.
- 7 Y. W. Lin, *Coord. Chem. Rev.*, 2017, **336**, 1–27.
- 8 G. Kiss, N. Çelebi-Ölçüm, R. Moretti, D. Baker and K. N. Houk, *Angew. Chem. Int. Ed.*, 2013, **52**, 5700–5725.
- 9 D. Hilvert, *Annu. Rev. Biochem.*, 2013, **82**, 447–470.
- 10 E. M. Brustad and F. H. Arnold, *Curr. Opin. Chem. Biol.*, 2011, **15**, 201–210.

- 11 I. D. Petrik, J. Liu and Y. Lu, *Curr. Opin. Chem. Biol.*, 2014, **19**, 67–75.
- 12 T. Heinisch and T. R. Ward, *Curr. Opin. Chem. Biol.*, 2010, **14**, 184–199.
- 13 H. Kries, R. Blomberg and D. Hilvert, *Curr. Opin. Chem. Biol.*, 2013, **17**, 221–228.
- 14 F. Van De Velde, L. Könemann, F. Van Rantwijk and R. A. Sheldon, *Biotechnol. Bioeng.*, 2000, **67**, 87–96.
- 15 M. L. Zastrow and V. L. Pecoraro, *Coord. Chem. Rev.*, 2013, **257**, 2565–2588.
- 16 M. Zhao, H. B. Wang, L. N. Ji and Z. W. Mao, *Chem. Soc. Rev.*, 2013, **42**, 8360–8375.
- 17 S. P. de Visser, *Chem. - A Eur. J.*, 2020, **26**, 5308–5327.
- 18 C. M. Bunney, P. E., Zink, A. N., Holm, A. A., Billington, C. J., & Kotz, *Nat Chem*, 2017, **176**, 139–148.
- 19 N. Stephanopoulos and M. B. Francis, *Nat. Chem. Biol.*, 2011, **7**, 876–884.
- 20 J. C. Lewis, *ACS Catal.*, 2013, **3**, 2954–2975.
- 21 T. Hayashi, D. Hilvert and A. P. Green, *Chem. - A Eur. J.*, 2018, **24**, 11821–11830.
- 22 L. Wang, A. Brock, B. Herberich and P. G. Schultz, *Science*, 2001, **292**, 498–500.
- 23 C. C. Liu and P. G. Schultz, *Annu. Rev. Biochem.*, 2010, **79**, 413–444.
- 24 J. W. Chin, *Annu. Rev. Biochem.*, 2014, **83**, 379–408.
- 25 Y. Ryu and P. G. Schultz, *Nat. Methods*, 2006, **3**, 263–5.
- 26 A. Dumas, L. Lercher, C. D. Spicer and B. G. Davis, *Chem. Sci.*, 2015, **6**, 50–69.
- 27 J. W. Chin, *Nature*, 2017, **550**, 53–60.
- 28 D. D. Young and P. G. Schultz, *ACS Chem. Biol.*, 2018, **13**, 854–870.
- 29 G. Roelfes, *Acc. Chem. Res.*, 2019, **52**, 545–556.
- 30 I. Drienovská, A. Rioz-Martinez, A. Draksharapu and G. Roelfes, *Chem. Sci.*, 2015, **6**, 770–776.
- 31 V. Resch and U. Hanefeld, *Catal. Sci. Technol.*, 2015, **5**, 1385–1399.
- 32 M. Engleder and H. Pichler, *Appl. Microbiol. Biotechnol.*, 2018, **102**, 5841–

5858.

- 33 J. Jin and U. Hanefeld, *Chem. Commun.*, 2011, **47**, 2502–2510.
- 34 J. Bos, A. García-Herraiz and G. Roelfes, *Chem. Sci.*, 2013, **4**, 3578–3582.
- 35 A. J. Boersma, D. Coquire, D. Geerdink, F. Rosati, B. L. Feringa and G. Roelfes, *Nat. Chem.*, 2010, **2**, 991–995.
- 36 I. Drienovsk, L. Alonso-Cotchico, P. Vidossich, A. Lleds, J.-D. Marchal and G. Roelfes, *Chem. Sci.*, 2017, **8**, 7228–7235.
- 37 M. Bersellini and G. Roelfes, *Org. Biomol. Chem.*, 2017, **15**, 3069–3073.
- 38 W. Ko, S. Kim, K. Jo and H. S. Lee, *Amino Acids*, 2016, **48**, 357–363.
- 39 F. W. Studier, *Protein Expr. Purif.*, 2005, **41**, 207–234.
- 40 L.-M. Guzman, D. Belin, R. T. Cartee, M. J. and J. Beckwith, *J. Bacteriol.*, 2000, **177**, 4121–4130.
- 41 F. W. Studier, *Methods Mol. Biol.*, 2014, **1091**, 17–32.
- 42 N. Bhawsinghka, K. F. Glenn and R. M. Schaaper, *Microbiol. Resour. Announc.*, 2020, e00009–20.
- 43 I. U. Heinemann, A. J. Rovner, H. R. Aerni, S. Rogulina, L. Cheng, W. Olds, J. T. Fischer, D. Sll, F. J. Isaacs and J. Rinehart, *FEBS Lett.*, 2012, **586**, 37616–3722.
- 44 J. Fredens, K. Wang, D. de la Torre, L. F. H. Funke, W. E. Robertson, Y. Christova, T. Chia, W. H. Schmied, D. L. Dunkelmann, V. Bernek, C. Uttamapinant, A. G. Llamazares, T. S. Elliott and J. W. Chin, *Nature*, 2019, **569**, 514–518.
- 45 R. Ballardini, V. Balzani, M. Clemente-Len, A. Credi, M. Teresa Gandolfi, E. Ishow, J. Perkins, J. Fraser Stoddart, H. R. Tseng and S. Wenger, *J. Am. Chem. Soc.*, 2002, **124**, 12786–12795.
- 46 P. Liu, Y. Chen, J. Deng and Y. Tu, *Synthesis*, 2001, **14**, 2078–2080.
- 47 S. L. Hyun and P. G. Schultz, *J. Am. Chem. Soc.*, 2008, **130**, 13194–13195.
- 48 C. Fontenas, H. At-Haddou, E. Bejan and G. G. A. Balavoine, *Synth. Commun.*, 1998, **28**, 1743–1753.
- 49 G. Roelfes, A. J. Boersma and B. L. Feringa, *Chem. Commun.*, 2006, 635–637.

CHAPTER 3

Development of Porphyrin-LmrR based Artificial Metalloenzymes for Carbene-Transfer Reactions

Abstract

Inspired by the diverse functions that heme proteins perform in nature, several artificial heme-enzymes towards new-to-nature reactivities have been developed. The broad reactivities of porphyrins and the possibility of tuning their properties by introducing them into engineered proteins offer the possibility of accessing a variety of catalytic activities. Here, we present an artificial heme enzyme for abiotic carbene transfer reactions based on the flexible LmrR scaffold. The supramolecular self-assembly of the artificial metalloenzyme (ArM) was confirmed by electronic absorption spectroscopy and protein crystallography, where, interestingly, the heme iron appears inaccessible for substrates. However, good catalytic activity and moderate enantioselectivity were observed in an abiological cyclopropanation reaction. We propose that the flexibility of the scaffold is key for allowing binding of the substrates and the formation of pre-catalytic structures. This hypothesis is supported by molecular dynamics simulations and by comparison of catalytic performance with binding affinity data. In this chapter, the outreach of the porphyrin-based ArMs is further investigated by expanding the reaction scope to N-H insertion reactions; studying the binding of different porphyrins to LmrR; and attempting to engineer an axial ligand.

Part of the results published in:

L. Villarino, K. E. Splan, E. Reddem, L. Alonso-Cotchico, C. Gutiérrez de Souza, A. Lledós, J.-D. Maréchal, A.-M. W. H. Thunnissen and G. Roelfes, *Angew. Chem. Int. Ed.*, 2018, **57**, 7785–7789.

3.1. INTRODUCTION

Artificial heme proteins

Heme proteins are metalloproteins with versatile functions. In nature, they are involved in respiratory electron-transfer processes (e.g. cytochrome *c*), oxygen binding and transport (e.g. myoglobin and haemoglobin), cellular signalling and substrate oxidation (e.g. cytochrome P450).¹ A key contributor to all these diverse activities is the heme prosthetic group—a porphyrin ligand coordinating an iron ion (found in its ferric or ferrous form)—. In heme proteins, the iron is usually hexacoordinated, and it is chelated by the four nitrogens of the porphyrin and by one or two amino acid residues of the protein acting as axial ligands. Axial ligands in heme proteins play an important role in shaping the electronic properties, the spin states, and the geometry of the active site.²

3

Inspired by the importance of heme proteins in biological systems, both naturally occurring and synthetic metalloporphyrins have been extensively used in catalysis, especially in oxidation reactions and as photocatalysts.^{3,4} The diverse functionalities of heme proteins have also attracted attention of different research fields and instigated the use of this class of proteins for different purposes. In particular, several heme enzymes have been engineered for accessing new-to-nature reactions.^{5,6} Notably, cytochrome P450s, cytochrome *c* and myoglobin have been transformed into artificial metalloenzymes (ArMs) for carbene and nitrene transfer reactions (like cyclopropanations and aziridinations),^{7–11} olefinations^{12,13} and N-H insertions,^{14,15} as well as S-H,¹⁶ Si-H¹⁷ and B-H¹⁸ insertions. Various strategies have been followed to create these artificial heme enzymes. On the ligand side, the native heme has been replaced with functionalised metalloporphyrins^{19,20} or with porphyrins containing metals such as Co, Rh, Mn, Ir and Zn.^{21,22} On the scaffold side, site-directed mutagenesis of proteins based on a rational^{23,24} or computational design,^{25,26} and on application of evolutionary strategies has been performed.²⁷ These engineered heme enzymes have in common a large hydrophobic binding pocket and the presence of an axial ligand, which has again been engineered by some research groups in order to obtain the desired properties.^{24,28,29} Moreover, *de novo* design of heme proteins has been explored: these proteins are usually

based on either a tight, heme ligand-binding site within a 4-helix bundles structure,^{30–33} antibodies, or different protein scaffolds.^{34,35} *De novo* heme proteins generally involve computational design and, to date, the reported examples result in rigid structures lacking a large substrate binding site, which limits their application in catalysis. Here we present an artificial heme enzyme based on the lactococcal multidrug resistance regulator (LmrR) protein, which presents a large promiscuous substrate binding site, capable of catalysing abiotic enantioselective cyclopropanation reactions. Efforts towards understanding its mechanism and exploring the possibilities of porphyrin-LmrR based ArMs are discussed in this chapter.

LmrR-based artificial metalloenzymes

LmrR is a multidrug resistance regulator from *Lactococcus Lactis*. It is a homodimer with an unusually large hydrophobic pore at the dimer interface. LmrR contains two parallel tryptophans (W96 and W96') in the middle of the pore, which are mainly responsible for binding hydrophobic drugs by π - π stacking interactions. LmrR's highly flexible scaffold allows to bind structurally different aromatic planar compounds, like Hoechst 33342 ($K_d = 21 \pm 8$ nM) or daunomycin ($K_d = 236 \pm 53$ nM).³⁶ In previous work, our group showed that LmrR can also bind Cu^{II}-phenanthroline (phen) with high affinity (2.6 ± 2 μ M)³⁷ and later on the crystal structure of Cu^{II}(phen)-LmrR was elucidated (PDB 6RIL),³⁸ showing binding of the Cu^{II}(phen) complex in between tryptophans 96 and 96'. Taking advantage of the high binding affinity of this complex for LmrR, an active and selective artificial metalloenzyme for a Friedel-Crafts alkylation reaction was created.³⁷ The supramolecular approach is highly convenient because it allows for a straightforward formation of the ArM by self-assembly. Application of this methodology results in a dynamic protein-cofactor system, and hence, it typically relies on strong hydrophobic interactions to push the equilibrium towards the desired conformation. The binding promiscuity of LmrR and its versatility as scaffold for ArMs³⁹ encouraged us to investigate the possibility of creating a LmrR-based artificial heme enzyme using this approach.

3.2. RESULTS AND DISCUSSION

Formation of heme-LmrR based artificial metalloenzyme

The design was based on a previously reported LmrR variant containing a C-terminal Strep-tag and two extra mutations for purification purposes (K55D and K59Q).⁴⁰ We started our investigation by creating the ArMs by addition of hemin (Fe(III) chloroproporphyrin IX) to the corresponding LmrR solution in 50 mM KHPO₄ buffer, pH 7.0 and 150 mM NaCl. To assess the successful assembly, we then investigated the interactions between heme and LmrR by electronic absorption spectroscopy. The visible spectrum of hemin in aqueous solution at pH 7.0 showed a broad Soret band at around 400 nm and a small peak at around 610 nm corresponding to the Q band. This spectral pattern is characteristic for the π - π porphyrin dimers that hemin forms under aqueous conditions.⁴¹ In presence of LmrR these two peaks sharpened up and a spectrum more similar to that of hemin in organic solvents was obtained (Figure 1). This observation indicates that the dimers are disrupted in the presence of LmrR, and hemin is in its monomeric form. Additionally, monitoring spectral changes upon addition of increasing amount of protein, evidenced that a dimer of protein is necessary to bind one hemin molecule (Figure 1a). This stoichiometry was confirmed by monitoring the quenching of tryptophan fluorescence at 280 nm, which was completed after one equivalent of hemin per dimeric LmrR (Figure 1b).

Figure 1. a) Electronic absorption spectra upon addition of increasing amounts of LmrR to 5 μ M hemin. b) Quenching of fluorescence of 1 μ M LmrR dimer upon consecutive addition of hemin. Conditions: 20 mM phosphate buffer, 150 mM NaCl, pH 7.0.

A dissociation constant of 38 ± 27 nM was calculated by fitting of the curves obtained from the fluorescence quenching data. This value indicated a

comparable, or even somewhat stronger affinity of LmrR to heme than that calculated for other planar aromatic molecules.

We examined the importance of W96 on binding of the heme moiety by mutating it to alanine. UV-Vis spectra upon titration of LmrR_W96A to hemin showed the peak around 400 nm present in the heme-LmrR spectra with a decrease in intensity (Figure 8), attributed to non-specific binding of heme on the protein surface. Together, these results evidence the formation of a heme $\subset$ LmrR complex in which one heme molecule strongly binds in between the tryptophans of the hydrophobic pocket of LmrR.

These findings were further confirmed by the X-ray structure of LmrR co-crystallised with hemin (PDB 6FUU). In this structure, electron density in between the indole rings of the central W96/W96' pair attributed to bound heme was found. Whereas the porphyrin ring showed a clear density, the substituents were not well resolved, indicating a lack of specific stabilising interactions with the peripheral heme and mobility of the heme molecule within the hydrophobic pore. In all the four possible conformations modelled in this structure, the heme was sandwiched by cation- π interactions in between W96 and W96' and further stabilised by Van der Waals interactions with hydrophobic residues of the active site; namely M8, A11 and V15. Notably, none of these structures seemed to be suitable for the observed catalysis, since the catalytic iron was shielded, hampering coordination of the substrates. However, the strong dynamics of heme $\subset$ LmrR indicated by the disordered crystal structure as well as the versatility of LmrR observed in previous work, led us to hypothesise that the substrates could be accommodated in the pocket in an orientation that would allow catalysis to happen.

Catalysis

Pioneering work of the Arnold group showed the potential of using engineered cytochrome P450_{BM3} in the cyclopropanation of styrenes.⁸ Subsequently, other engineered heme-binding enzymes, like myoglobin, were found active in this kind of reactions.^{7,42,43} These results led us to select the cyclopropanation of *o*-methoxystyrene (**1a**) with ethyldiazoacetate (EDA, **2**) as model reaction for testing the potential of the novel ArM.

The ArM was created in anaerobic conditions by addition of hemein (10 μ M) to a 1.1 fold excess of LmrR (11 μ M) in a solution containing 50 mM KPi buffer at pH 8.0 and sodium dithionite, used as a reducing agent to generate Fe(II) needed for carbene formation. The reaction conditions were selected based on previous literature reports.^{7,8} 30 mM of substrate **1a** and 10 mM of substrate **2**, were added to initiate the reaction. In absence of LmrR, diethyl fumarate (**4**) was obtained as major product, formed by dimerisation of EDA. The reaction with hemeC-LmrR, led to the cyclopropanation *trans* product (**3a**) as major product, with 25% yield (98:2 *trans*:*cis* ratio) and 17% ee (Table 1). Favourable formation of the *trans* product was expected, since it is typical for the concerted, non-radical mechanism follow by iron (II) cyclopropanation catalytic reactions.^{44–46} Only an insignificant amount of **4** (less than 5% yield) was obtained as side product. These encouraging results clearly indicated the role of the LmrR scaffold on enhancing the reaction rate as well as on the chemoselectivity of the reaction.

3

The evidence that this new ArM, without further modifications of the scaffold, was catalytically active despite the lack of an apparent coordination site for the substrates, led us to further investigate LmrR's active site. In order to assess the role of amino acid residues of the active site in catalysis, an alanine scanning was performed (Figure 2). Positions around the heme binding site, either facing towards the inside of the pocket (W96, M8, M89, V15, A92) or positioned at the entrance of the pocket (F93 and D100), were mutated to alanine, except A92 that was mutated to glutamic acid. Firstly, W96A was tested. As expected, since it is responsible of heme binding, W96 plays an important role in catalysis and, when removed, the enantioselectivity was completely lost. However, W96A still led to 28% yield, most likely due to unspecific interactions of the heme with the protein scaffold. The most prominent effect was given by V15A and M8A (Table 1). V15A led to a loss of activity, indicating that V15 is important for catalysis. In contrast, with M8A an increase of both yield (up to 36%) and ee (up to 44%) was observed. D100A also led to an increase of yield (38%) and a less significant increase of ee (24%), and A92E to an increase of activity (31% yield) and a loss of ee. Finally, the reaction with F93A and M89A resulted in a similar activity than LmrR but a decrease or loss of ee (Table 1). These results show that alanine scanning is

a useful approach that allows to identify important residues for catalysis by mutating few selected positions.

Figure 2. X-ray structure of hemeC-LmrR (PDB 6FUU) with positions selected for alanine scanning highlighted. Overlapping alternate binding modes of heme are shown.

Table 1. Results of the cyclopropanation reaction of styrene **1a** with EDA (**2**) in presence of hemeC-LmrR variants. TON: turnover number.

Entry	Protein	Yield (%)	TON (3a)	ee _E (%)
1	-	5 ± 2	51	-
2	LmrR	25 ± 1	247	17 ± 5

3	LmrR_W96A	23 ± 3	232	11 ± 1
4	LmrR_D100A	38 ± 8	375	24 ± 5
5	LmrR_A92E	31 ± 11	314	< 2
6	LmrR_F93A	28 ± 13	276	< 5
7	LmrR_V15A	2 ± 1	15	17 ± 1
8	LmrR_M89A	22 ± 9	224	4 ± 2
9	LmrR_M8A	36 ± 13	359	44 ± 12

Conditions: **1a** (30 mM), **2** (10 mM), hemin (1 mol %; 10 μM), protein (1.1 mol %; 11 μM) and 10 mM Na₂S₂O₄ in 50 mM phosphate buffer (pH 8.0), under Ar, at 4 °C for 18 h. Results are the average of at least two independent experiments, both carried out in duplicate. eeE corresponds to that of the *trans* product.

The reaction conditions were then optimised for the most enantioselective mutant M8A to evaluate the effect of the styrene:EDA ratio (Table 2) and of the pH (Table 3). The reactions with excess of styrene and in 1:1 ratio led to similar yields and ee's. As expected, excess of styrene led to the lowest amount of dimerisation product **4**, which is highest in excess of EDA. Interestingly, with excess of EDA a slightly higher yield was obtained, suggesting that the carbene formation might be the rate limiting step, as well as a slight loss of ee. As a compromise between activity, and chemo- and enantioselectivity, the following experiments were performed in a ratio styrene:EDA 3:1.

Table 2. Results of the cyclopropanation reaction of **1a** and **2** in presence of hemeC-LmrR_M8A with different substrates ratios. TON: total turnover number.

Entry	Styrene:EDA	Yield (%)	TON (3a)	<i>trans:cis</i>	eeE (%)	ratio 3a/4
1	3:1	36 ± 13	359	92:8	44 ± 12	15

2	1:1	36 ± 11	357	93:7	42 ± 12	7
3	1:3	44 ± 19	436	93:7	35 ± 13	1.5

Conditions: hemin (1 mol %; 10 µM), protein (1.1 mol %; 11 µM) and 10 mM Na₂S₂O₄ in 50 mM phosphate buffer (pH 8.0), under Ar, at 4 °C for 18 h. Results are the average of at least two independent experiments, both carried out in duplicate. ee_E corresponds to the ee of the *trans* product.

On the other hand, decreasing the pH from 8.0 to 7.0 led to an increase of yield and ee, but also to an increase of dimeric secondary product **4**. The results at pH 6.0 did not change with respect to pH 7.0.

Table 3. Results of the cyclopropanation reaction of **1a** and **2** in presence of hemeC-LmrR_M8A at different pHs.

Entry	pH	Yield (%)	TON (3a)	<i>trans:cis</i>	ee _E (%)	ratio 3a/4
1	8.0	36 ± 13	359	92:8	44 ± 12	15
2	7.0	45 ± 9	449	93:7	51 ± 14	6
3	6.0	49 ± 20	494	92:8	50 ± 16	5

Conditions: **1a** (30 mM), **2** (10 mM), hemin (1 mol %; 10 µM), protein (1.1 mol %; 11 µM) and 10 mM Na₂S₂O₄ in 50 mM phosphate buffer, under Ar, at 4 °C for 18 h. Results are the average of at least two independent experiments, both carried out in duplicate. ee_E corresponds to the ee of the *trans* product.

Next, the substrate scope was explored. The styrene derivatives **1a-c** were tested under the optimised conditions. All substrates showed an enhanced activity compared to the reaction with hemin without protein and led to ee's ranging from 25% to 51% (**3c**) yield (Table 4). These results showed that, despite the apparent lack of a coordination site of the heme-iron for the substrates, these ArMs can successfully perform abiotic cyclopropanation reactions with a selection of styrene substrates.

Table 4. Substrate scope of the cyclopropanation reaction of styrene derivatives **1a-c** and EDA (**2**).

$\text{1a-c} + \text{2} \longrightarrow \text{3a-c}$

a: R = OMe
b: R = H
c: R = Cl

Entry	Styrene	Product	LmrR_M8A (μM)	Yield (%)	TON	<i>trans</i> : <i>cis</i>	ee _E (%)	ratio (3/4)
1	1a	3a	-	5 $\pm$ 2	51	93:7	< 5	0.8
2	1a	3a	II	45 $\pm$ 9	449	93:7	51 $\pm$ 14	6
3	1b	3b	-	6 $\pm$ 0	59	88:12	< 5	n.d.
4	1b	3b	II	39 $\pm$ 13	391	86:14	38 $\pm$ 5	n.d.
5	1c	3c	-	1 $\pm$ 0	12	89:11	< 5	0.2
6	1c	3c	II	30 $\pm$ 16	301	93:7	33 $\pm$ 19	3

Conditions: **1a-c** (30 mM), **2** (10 mM), hemin (1 mol %; 10 μM), protein (1.1 mol %; 11 μM) and 10 mM Na₂S₂O₄ in 50 mM phosphate buffer, pH 7.0, under Ar, at 4 °C for 18 h. Results are the average of at least two independent experiments, both carried out in duplicate. ee_E corresponds to the ee of the *trans* product.

To investigate whether the effect of the mutations was related to the binding of heme, the K_d of two mutants of amino acids located inside of the hydrophobic pocket of LmrR that led to different catalytic outputs (improved or diminished activity and selectivity) were measured. The K_d calculated for M8A was 280 $\pm$ 55 nM and for M89A 781 $\pm$ 15 nM. Interestingly, the K_d of M8A, which led to an increase of yield and ee, was an order of magnitude higher than that of wild-type LmrR. This fact might be explained by the smaller size residues favouring mobility of the heme inside the pocket, and, hence, decreasing its binding affinity. On the other hand, M89A, which led to a loss of enantioselectivity, presents an even higher K_d . These results are in

contrast with the general tendency in enzyme design, where high binding affinities are usually desired, since they result in a higher amount of effective catalyst. The results presented here, illustrate once more on the unique characteristics of the LmrR scaffold and evidence that certain degree of mobility in the appropriate location within the protein can facilitate the cofactor adopting a more suitable conformation for catalysis.

Computational analysis⁴⁷ allowed to shed light on the intriguing ability of heme $\subset$ LmrR to perform this reaction regardless of the absence of a clear coordination site for the substrates close to the catalytic iron. The results obtained from heme and its adduct with the carbene intermediate docked into LmrR's crystal structure (**5** and **6** respectively) and further submitted to molecular dynamic (MD) simulations, revealed a potential conformation in which the α' helix appeared displaced and W96' facing towards the outside of the pocket, leaving enough room for accommodating the co-substrate. Additionally, the transition state (TS) geometry leading to the 1R,2R cyclopropane product (**7**) was calculated by quantum mechanics (QM). The obtained structures were docked into the LmrR structure and submitted to MD simulations under the same conditions than before. Interestingly, the majority of the structures showed a broader space between the two helices where the TS could be accommodated. Moreover, W96' contributed to stability of the TS by π -stacking interactions with the phenyl group of the styrene substrate. To understand the role of the mutation leading to the best performing catalyst, MD simulations of the TS in LmrR_M8A structure were also performed. The simulations showed that the beneficial effect of this mutation was mainly due to steric effects. Change of the larger methionine by a smaller alanine leaves a space that can be occupied by the styrene co-substrate. Moreover, W96' appears rotated towards the solvent and thus contributes to the binding of the TS in the dimer interface. The analysis performed gives a plausible explanation to how the cyclopropanation reaction can occur in the hydrophobic pocket of LmrR and highlights the beneficial effect of a flexible scaffold for the catalytic activity of the ArM.

Figure 3. Representative structures resulting from MD simulations. a) The LmrR-heme system (cluster 3, RMSD with crystal 1.634 Å). b) The LmrR-heme-carbene system (cluster 0, RMSD with crystal of 0.548 Å). c) The cyclopropanation transition state with LmrR. d) The cyclopropanation transition state with LmrR_M8A.

Reaction scope

3

In order to further explore the potential of the heme C LmrR system, we decided to test the ArM for other reactivities. The cyclopropanation of styrenes typically proceeds *via* formation of a reactive iron-carbenoid intermediate between the heme iron and EDA. The Arnold group described a P450 variant that was not only able to perform the abiotic cyclopropanation of styrenes, but also to use the versatile carbene intermediate to catalyse N-H insertion reactions.¹⁴ Later on, the reactivity of engineered P450s as well as myoglobin was extended to other carbene and nitrene transfer reactions.⁵ These findings prompted us to investigate the activity of the heme C LmrR complex in the N-H insertion with aniline (**5**) and EDA (**2**). Performed in organic solvents with porphyrins such as iron tetraphenylporphyrin (FeTPP) and in water with Cu(I) catalyst, this reaction yields a mixture of single and double alkylated product when primary amines are used.^{48,49} With evolved enzymes, however, the protein scaffold allows to sterically select for the single addition product.^{14,15}

The reaction was run under anaerobic conditions for 18 hours in presence of $\text{Na}_2\text{S}_2\text{O}_4$, 50 mM phosphate buffer, at pH 8.0 and room temperature, 30 mM **5** and 10 mM **2**. The reaction without protein led to a mixture of single (**6**) and double alkylated product (**7**), with 78% yield of **6**. The presence of LmrR led to an increase of yield up to 93% and only the single alkylated product was

obtained, evidencing the participation of the LmrR scaffold in the reaction. Interestingly, the reaction in presence of the protein proceeds even with an excess of aniline, which is known to bind to the heme iron of some heme proteins and could therefore prevent the formation of the carbene with the diazo substrate.^{50,51} The yield without protein, however, was already too high to allow for a significant optimisation of the ArM, even when the reaction was performed in shorter periods of time and at lower temperature (30 minutes and 4 °C), as shown in Table 5.

Table 5. Results N-H insertion of aniline (**5**) and EDA (**2**). TON: turn over number; rt: room temperature.

Reaction scheme: Aniline (**5**) + Ethyl diazoacetate (**2**) → N-ethyl-N-phenylglycinate (**6**) + N,N-diethyl-N-phenylglycinate (**7**)

Entry	Catalyst	Time (h)	T (°C)	Yield (6) (%)	TON (6)	Product
1	hemin	18	rt	78 ± 1	780	6 + 7
2	hemin◌LmrR	18	rt	93 ± 1	931	6
3	hemin	4	rt	82 ± 1	820	6 + 7
4	hemin◌LmrR	4	rt	90 ± 0	900	6
5	hemin	0.5	4	64 ± 0	638	6 + 7
6	hemin◌LmrR	0.5	4	70 ± 5	698	6

Conditions: **5** (30 mM), **2** (10 mM), hemin (1 mol %; 10 μM), protein (1.1 mol %; 11 μM) and 10 mM Na₂S₂O₄ in 50 mM phosphate buffer (pH 8.0), under Ar. Results are the average of at least two independent experiments, both carried out in duplicate.

Therefore, some preliminary studies with a less reactive diazo substrate, ethyl 2-diazopropanoate (**8**), were performed. This reaction would lead to the

chiral product **9**, which would allow to judge the performance of the ArM also by the enantioselectivity obtained. Initial studies showed low yields of **9** both with hemin and heme $\subset$ LmrR. Since no significant enantioselectivity was obtained, the role of the protein in this reaction could not be assessed.

Scheme 1. Reaction scheme of the N-H insertion of aniline and 2-diazopropanoate.

Porphyrins

Metalloporphyrin activities are easily diversified by the use of different metals and modifications of the porphyrin core. Heme proteins are, therefore, a common target to be engineered by replacing heme with artificial cofactors.^{52–}

54

The binding and catalytic promiscuity of LmrR brought us to investigate whether other porphyrins could be incorporated into its scaffold and ultimately create a library of porphyrin $\subset$ LmrR based ArMs for different reactivities. The structural requirements for binding of the porphyrins was first investigated. Based on the available crystal structures of drugs bound to LmrR, we anticipated that planar porphyrins would preferably bind in the pocket. This hypothesis was confirmed by titrating two structurally different porphyrins to LmrR and measuring the binding constants (Figure 4b,c). The K_d for the planar Fe(III) mesoporphyrin IX chloride (FeMP) was 44 ± 16 nM, similar to that of hemin. On the other hand, the binding affinity of Fe(III)tetrakis (1-methyl-4-pyridyl) porphyrin pentachloride (FeTMPyP), which contain four phenyl groups in *meso*-position that can freely rotate, was lower (500 ± 79 nM). Subsequently, introduction of another porphyrin with useful characteristics for catalysis was investigated. The Fasan group described a myoglobin variant reconstituted with FeCe6 porphyrin that became highly active and selective for olefin cyclopropanation under aerobic conditions. The effect of this porphyrin is attributed to the presence of an extra carboxylic group in *meso* position compared to hemin: electron-

withdrawing groups at this position are known to stabilise the iron porphyrin carbene^{55,56} and to decrease the affinity of the iron atom for O₂.⁵⁷ Consequently, the binding of FeCe6 was explored. The commercially available apo-porphyrin was metallated with FeCl₃ following literature procedures.⁵⁵ The calculated K_d for FeCe6 was 816 ± 376 nM (Figure 4d). These porphyrin-LmrR complexes can be further tested in catalysis to explore the combined potential of a multi-binding protein scaffold and the functional promiscuity of porphyrin skeletons.

Figure 4. Porphyrin structures and the corresponding binding constants calculated for LmrR.

Engineering an axial ligand

Natural heme-binding enzymes have the presence of an axial ligand in common, typically a histidine or cysteine residue that coordinates the heme iron and to plays an important role on binding and reactivity of the heme group.² In fact, most engineered heme enzymes also contain an axial ligand that can be key for accessing certain reactivities.^{29,58} Remarkably, heme-LmrR is catalytically active without an obvious axial ligand for the heme. In order to further explore the versatility of the system, we investigated the possibility of engineering an axial ligand for heme binding in the LmrR scaffold, which would allow us to tune the electronic properties of the metal and thus expand the range of reactivities of this ArM.

In an attempt to introduce an axial ligand into LmrR, we selected position D100 to be mutated to amino acids that are known to bind heme in natural enzymes, like cysteine, methionine, glutamic acid and histidine. We judged D100 to be a promising position, since it is at the entrance of the pocket but

relatively close to the tryptophans. We hoped that a coordinating residue at this position might be able to displace the heme from the pocket requiring little movement. However, D100 and D100' are facing each other in LmrR's dimeric structure and therefore—if they would bind to the heme iron—they might occupy both coordination sites, preventing the substrates to bind. We prepared these mutants and tested them in cyclopropanation reaction to investigate if they were still active. Whereas methionine and glutamic acid at this position led to a significant decrease of the yield, the cysteine and histidine mutants led to an increase of the yield and cysteine also of the ee compared to wild type LmrR (Table 6). Since D100C was the best performing mutant within the heme binding amino acids tested, we went to investigate whether the cysteine at this position of LmrR could be acting as axial ligand. Additionally, position V15 was also explored as possible heme coordinator. V15 and V15' are located in opposite corners of the active site, leaving enough space to bind a heme molecule in each monomer residue without cross-interacting, and leaving a coordination site free for binding of the substrates in each heme iron. V15 was mutated to cysteine for comparison with D100C.

Table 6. Results for the cyclopropanation reaction of **1a** with **2**.

Entry	protein	Yield (%)	ee (%)
1	-	5 ± 2	-
2	LmrR	25 ± 11	17 ± 5
3	LmrR_D100C	14 ± 1	27 ± 4
4	LmrR_D100M	2 ± 0	n.d.
5	LmrR_D100E	4 ± 5	n.d.
6	LmrR_D100H	33 ± 5	5 ± 0

Conditions: **1** (30 mM), **2** (10 mM), hemin (1 mol %; 10 μM), protein (1.1 mol %; 11 μM) and 10 mM Na₂S₂O₄ in 50 mM phosphate buffer (pH 8.0), under argon. Results are the average of at least two independent experiments. For yields < 5%, ee was not determined (n.d.).

To interrogate the iron heme, we used Resonance Raman (RR) spectroscopy, which allows to monitor vibrational normal modes upon photon irradiation. When the frequency of the excitation laser coincides with that of an allowed electronic transition, the signal can be selectively enhanced.⁵⁹ Excitation of the π - π^* Soret band leads to enhancement of the Raman vibrational modes of the porphyrin without interference of the protein, allowing a study of changes in the heme moiety.^{60–62} In particular, it has been proven that bands in the high-frequency spectrum of heme proteins are related to the porphyrin skeletal vibrations that are indicative of the oxidation, coordination, and spin states of the heme iron. The reduction of Fe(III) to Fe(II) is known to lead to a small lowering of these frequencies, whereas a change in axial coordination and spin state lead to significant frequency shifts. Typically, the most pronounced band at around 1370 cm^{-1} is considered as the main oxidation marker in heme proteins. On the other hand, the main characteristic bands in the low-frequency region are associated to vibrations of the proximal ligand of heme.^{60,61,63–65} The RR spectrum of LmrR and the cysteine mutants in presence of hemin were compared.

The samples containing the pre-assembled heme $\subset$ protein complexes were excited at 405 nm. Since the heme $\subset$ LmrR complex is in a binding equilibrium, we expected to observe the Raman signal of heme both inside the pocket of LmrR and in solution. Therefore, the spectra of heme $\subset$ LmrR and free hemin were compared.

Figure 5. Stacked Raman spectra of heme $\subset$ LmrR (blue) and hemin in solution (purple).

As shown in Figure 5, both spectra are quite similar, with some differences in relative intensity for the bands at higher frequencies. This can be the result of heme maintaining the same coordination state and similar geometry both inside the pocket and in solution.

In the low-frequency region of the hemeC LmrR spectrum, three weak peaks between 300 and 400 cm^{-1} can be observed. However, the accuracy in the area below 400 cm^{-1} was considered too low and these bands were not taken into account for the analysis. The first intense band appears at 676 cm^{-1} . A peak around this frequency is common in many heme proteins.^{61,63} Several peaks were found between 1300 and 1700 cm^{-1} in the hemeC LmrR spectra, which appeared in a similar pattern than the most characteristic peaks corresponding to vibrations of the heme ring (Figure 6). In Table 7, bands in the high frequency region for several ferric heme proteins and the hemeC LmrR ArM are compared.

Figure 6. Raman spectra of ferric heme-LmrR (blue) and ferrous heme-LmrR (green), formed by addition of sodium dithionite to the ferric solution.

The band at 1372 cm⁻¹ was tentatively assigned as the oxidation state marker by comparison to literature reports. To further justify this assignment, the spectra upon addition of Na₂S₂O₄ were recorded. Na₂S₂O₄ reduces the heme iron to its ferrous state, as evidenced by the role of the thiosulphate complex in catalysis, which is required to reduce the iron to the reactive iron (II). We expected that this change in oxidation state could be clearly monitored by comparison of the Raman spectrum. In fact, the significant increase in intensity of this band upon addition of the reducing agent supports this assignment. A slight decrease of Raman shift can also be observed, which is characteristic in reduced heme-proteins, although the shifts are typically higher (around 10 cm⁻¹, rather than 2 cm⁻¹ found here).

Table 7. Heme skeletal vibrations of natural ferric heme proteins compared to hemeC-LmrR.^{61,63}

eNOS (5c/hs)	b-NOS (5c/hs)	P450 (5c/hs)	Ferri cytochrome c (ls)	CPO (5c/hs)	eNOS (6c/ls)	b- NOS (6c/ls)	P450 (6c/ls)	CPO (6c/ls)	HemeC- LmrR
1370	1370	1368	1345	1369	1374	1371	1372	1373	1372
									1419
1489	1487	1488	1432	1490	1503	1502	n.d.	1504	1491
1563	1563	1570	1552	1567	1579	1575	1582	n.d.	1568

An amino acid residue coordinating the iron heme would cause significant changes in geometry, coordination and/or spin state of the iron, reflecting on substantial shifts of bands on the 1300-1700 cm⁻¹ region as well as the low frequency region. However, the spectra of both mutants were fairly similar compared to that of LmrR (Figure 7), indicating that no apparent change in coordination occurred by mutating these residues. The small differences observed between the spectra can be due to changes in the environment caused by introducing a different amino acid into the protein scaffold or changes in the affinity of heme for the hydrophobic pocket of LmrR. Overall,

the intrinsic complexity of the system and differences with reported spectra of heme-proteins hinder a detailed analysis. Nevertheless, Raman spectroscopy allowed us to exclude a change of coordination to an axial ligand occurring in the mutants.

Figure 7. Stacked spectra of hemeCLmrR, hemeCLmrR_D100C and hemeCLmrR_V15C.

3.3. CONCLUSIONS AND PERSPECTIVES

In conclusion, here we show a hemeCLmrR based artificial metalloenzyme for carbene transfer reactions. The supramolecular formation of this ArM with heme bound in between W96 and W96' of LmrR was confirmed by electronic absorption spectroscopy and protein crystallography. Titration of heme to LmrR indicated tight binding with a calculated $K_d = 38 \pm 27$ nM. Despite the heme being shielded by W96/W96', this ArM afforded good activity and moderate enantioselectivity for the abiotic cyclopropanation of

styrenes. The novelty of this heme enzyme relies on the absence of an axial ligation of the heme by an amino acid residue of the protein and the apparent lack of a free coordination site on the heme iron for binding of the substrates. Both MD simulations and the X-ray structure evidenced that the system was extremely dynamic, and one of the helices of LmrR was able to move and open up the structure to accommodate the substrates. Alanine scanning of selected amino acids of the active site proved to be a useful approach that allowed to easily point-out important residues for catalysis (V15 which caused a loss of activity when it was changed by alanine) and beneficial mutations (M8A, which gave the highest yield and ee).

The binding affinity of heme with some mutants was then calculated and the relation between K_d and activity was investigated. Mutant M89A increased the K_d one order of magnitude, which resulted in a decrease of activity. This result was expected, since the supramolecular approach typically relies on strong supramolecular interactions between scaffold and cofactor. On the contrary, mutant M8A, which also leaves more space for binding of the substrates, as shown by the MD simulations, and thus led to lower binding affinity, favoured the activity of this ArM. These interesting findings show that whereas binding affinity is necessary to recruit the metal complex, allowing some mobility of the cofactor in certain positions of the pocket can be key for the reactivity of the ArM.

The combined results of catalysis and binding affinity prompted us to reflect on general considerations for the design of ArMs. Usually, a rigid structure, which allows for a better control of the active site, and high binding affinity for the metal complex of study, which ensures the presence of more active enzyme-metal complexes in the reaction mixture, are preferred. However, the results with LmrR constitute an example of the importance of the flexibility of the scaffold and dynamics of the system to enable catalysis. The potential of this ArM, was further explored by testing other reactions, studying the possibility of using different porphyrins, and attempting to engineer an axial ligand.

The potential of expanding the reaction scope of the heme-LmrR based ArM, was demonstrated with the N-H insertion reaction of aniline with EDA. Whereas the reaction without protein led to a mixture of double and single

alkylated product, in presence of LmrR the single alkylated product was exclusively obtained. The effect of decreasing the protein concentration should be explored in order to further investigate the potential of this ArM. Preliminary studies were also performed with aniline and a more challenging diazo substrate, 2-diazopropanoate, which would lead to a chiral product and thus the performance of the ArM can be judged by both yield and enantioselectivity. Although low conversions were found in initial studies in presence and absence of protein, further investigations are required in order to assess the role of the protein in the reaction. Exploring the reaction outcome with more reactive aniline derivatives, like substrates with electron-withdrawing groups in *para* position, might be a promising approach to overcome the low conversion rates. Furthermore, other carbene and nitrene transfer reactions could be investigated to expand the range of activities of hemeC-LmrR ArMs.

3

Additionally, binding studies showed that LmrR was able to bind several porphyrins other than heme (FeTPyM, FeMS, FeCe6), demonstrating the potential of LmrR's binding promiscuity for incorporating exogenous reactive complexes, and opening a window for combining the vast array of activities of porphyrins and the advantageous hydrophobic and chiral environment provided by LmrR. Testing these porphyrins in catalysis would be interesting since, particularly FeTPyM and FeCe6, have been shown to be able to catalyse the cyclopropanation of styrenes under aerobic conditions.^{19,66}

Lastly, the possibility of engineering an axial ligand into the protein, was investigated by Resonance Raman spectroscopy. The analysis of the Raman spectrum of the two tested positions (D100C and V15C) compared to the one of LmrR suggested the absence of axial ligation with these two mutants. Acquiring the Raman spectra of the catalytically inactive D100 mutants could tell us whether the loss of activity is due to a double coordination to the heme iron. Mutating a heme binding residue in position W96 would be interesting, since the axial ligation would then substitute the tryptophan binding instead of competing with it. However, removing of the tryptophans could potentially lead to structural issues. Overall, this ArM constitutes an interesting system to be studied from the spectroscopic point of view. The heme-proteins reported in literature usually present axial ligation, so that the heme is anchored in a particular position. In contrast, heme and LmrR are in a

dynamic binding equilibrium, where the heme moiety has a certain degree of mobility. Comparison of the Raman spectra of hemeC-LmrR to that of reported tetra-⁶⁷ and pentacoordinated iron porphyrins, suggest the presence of a high spin pentacoordinated iron in the hemeC-LmrR system. This hypothesis could be in agreement with, for instance, a water molecule being coordinated to the iron in solution. A water molecule is labile enough to be easily displaced by the substrates, and despite not appearing in the crystal structure, might be present in solution. However, as mentioned above, this system is extremely dynamic and different than other previously characterised heme-proteins. Hence, a more detailed study is required to be able to establish solid claims from the Raman spectra. In conclusion, studying this system from the spectroscopic point of view with techniques like Raman or redox potential could give us valuable information about the ArM and be a straightforward means to screen for axial ligation.

3.4. EXPERIMENTAL SECTION

General remarks

Chemicals and reagents were purchased from commercial suppliers (Sigma Aldrich, Acros or TCI Europe) and used without further purification. Flash chromatography was performed using Reveleris® X2 Flash Chromatography System. ¹H-NMR spectra were recorded on a Varian 400 (400 MHz) spectrometer in CDCl₃. Chemical shift values (δ) are denoted in ppm using residual solvent peaks as the internal standard (CHCl₃: δ 7.26 for 1H). ¹H-NMR results are reported in the conventional form: chemical shifts, multiplicity (d = doublet, dd = doublet of doublet, m = multiplet, q = quartet, s = singlet, t = triplet), coupling constants (Hz), and integration. Yields were determined by GC analysis (HP-5 (30 m x 250 μm x 0.25 μm) or CP Chirasil-Dex CB (25 m x 250 μm x 0.25 μm) using FID detection (Hewlett Packard 6890 Series II chromatograph). The injector temperature was 225 °C and the FID temperature was 250 °C. Enantiomeric excess determinations were performed by HPLC analysis using UV-detection (Shimadzu SCL-10Avp) on Chiralpak ODH, 140 min, 100_0 Heptane_isopropanol, 0.5 ml/min. UPLC-MS on protein samples was performed on a Acquity TQ Detector (ESITQD-MS) coupled to Waters Acquity Ultra Performance LC using a Acquity BEH C8 (1.7 μm 2.1 x 150 mm). Water (solvent A) and acetonitrile (solvent B)

containing 0.1% formic acid by volume, were used as the mobile phase at a flow rate of 0.3 mL/min. Gradient: 90% A for 2 min, linear gradient to 50% A in 2 min, linear gradient to 20% A in 5 min, followed by 2 min at 5% A. Re-equilibration of the column with 2 min at 90% A.

Molecular biology

General remarks

E. coli strain NEB5 α -T1 Phage Resistant (Stratagene) was used for cloning and *E. coli* BL21(DE3)-C43-T1 Phage Resistant (Stratagene) strain was used for protein expression. Primers were synthesised by Eurofins MWG Operon (Ebesberg, Germany). Pfu Turbo polymerase was purchased from Agilent and DpnI was purchased from New England Biolabs. The plasmid isolation kit was purchased from Roche. DNA sequencing was carried out by Eurofins Genomics. The concentration of the proteins was measured with Nanodrop 2000 (Thermo Scientific).

Site-directed mutagenesis

All the LmrR mutants were prepared by site-directed mutagenesis from available pET17b plasmid encoding for LmrR (previously optimised with mutations K55D and K59Q for purification purposes). The primers used for the mutagenesis are summarised in Table 8. The following PCR cycle was used: initial denaturation at 95 °C for 1 min, denaturation at 95 °C for 30 s, annealing at 58-65 °C for 1 min (depending on the T_m of the particular mutant) and extension at 68 °C for 5 min. The thermal cycle was repeated 16 times. The resulting PCR product was digested with restriction endonuclease DpnI for 1 h at 37 °C. Chemically competent *E. coli* NEB5 α -T1 Phage resistant cells were transformed with 5 μ L of the DpnI digested sample.

Table 8. Primers used for site-directed mutagenesis.

Primer	Sequence (5' $\rightarrow$ 3')
LmrR_M8A_fw	ATC CCG AAA GAA GCG CTG CGT GCT CAA
LmrR_M8A_rv	TTG AGC ACG CAG CGC TTC TTT CGG GAT
LmrR_F93A_fw	CGC CTG GCG GCG GAA TCC TGG
LmrR_F93A_rv	CCA GGA TTC CGC CGC CAG GCG

LmrR_VI5A_fw	GCT CAA ACC AAT GCG ATC CTG CTG AAT
LmrR_VI5A_rv	ATT CAG CAG GAT CGC ATT GGT TTG AGC
LmrR_DI00A_fw	TGG AGT CGT GTG GCG AAA ATC ATT GAA
LmrR_DI00A_rv	TTC AAT GAT TTT CGC CAC ACG ACT CCA
LmrR_W96A_fw	GCG TTC GAA TCC GCC AGT CGT GTG GAC
LmrR_W96A_rv	GTC CAC ACG ACT GGC GGA TTC GAA CGC
LmrR_DI00E_fw	TGG AGT CGT GTG GAG AAA ATC ATT GAA
LmrR_DI00E_rv	TTC AAT GAT TTT CTC CAC ACG ACT CCA
LmrR_DI00M_fw	TGG AGT CGT GTG ATG AAA ATC ATT GAA
LmrR_DI00M_rv	TTC AAT GAT TTT CAT CAC ACG ACT CCA
LmrR_DI00H_fw	TGG AGT CGT GTG CAC AAA ATC ATT GAA
LmrR_DI00H_rv	TTC AAT GAT TTT GTG CAC ACG ACT CCA
LmrR_DI00C_fw	TGG AGT CGT GTG TGC AAA ATC ATT GAA
LmrR_DI00C_rv	TTC AAT GAT TTT GCA CAC ACG ACT CCA
LmrR_M89A_fw	GGC CAT GAA AAC GCG CGC CTG ATT TTC
LmrR_M89A_rv	GAA AAT CAG GCG CGC GTT TTC ATG GCC
LmrR_VI5C_fw	GCT CAA ACC AAT TGC ATC CTG CTG AAT
LmrR_VI5C_rv	ATT CAG CAG GAT GCA ATT GGT TTG AGC

Protein expression and purification

pET17b expression vector carrying the corresponding LmrR variant constructs were transformed into *E. coli* BL21(DE3)-C43-T1 Phage Resistant cells. A starter culture of 10 mL of fresh LB medium containing 100 µg/mL of ampicillin was inoculated with a single colony. The starter culture was then used to inoculate 500 mL of fresh LB medium containing 100 µg/mL of ampicillin. When the culture reached an optical density at 600 nm between 0.8–0.9, isopropyl β-D-1- thiogalactopyranoside (IPTG) at final concentration of 1 mM was added to induce the expression of the target protein. Expressions were carried out at 30 °C, overnight. Cells were harvested by centrifugation (6000 rpm, JA10, 20 min, 4 °C, Beckman), re-suspended in 25 mL of 50 mM NaH₂PO₄, pH 8.0, 150 mM NaCl along with half a tablet of mini complete EDTA-free protease inhibitor cocktail (Roche). The re-suspended cells were sonicated (75% (200W) for 8 min (10 s on, 15 s off). The lysed cells were incubated with DnaseI (final concentration, 0.1 mg/mL) and MgCl₂ (final concentration, 10 mM) for 30 min on ice. After centrifugation (16000 rpm,

JA-17, 45 min, 4 °C, Beckman), 12.5 mL of the supernatant was equilibrated with 2.5 mL of Strep-tag Tactin column material for 1 h (mixed at 200 rpm on a rotary shaker) at 4 °C. The column was washed with 3 x 2 CV (column volume) of the above-mentioned buffer, and eluted with 0.6 CV, 1.6 CV and 0.8 CV of buffer containing 5 mM desthiobiotin. Fractions were analysed on a 12% polyacrylamide SDS-Tris Tricine gel followed by Coomassie Blue staining. Fractions containing protein were dialysed against phosphate buffer (50 mM KH_2PO_4 , pH 8.0) overnight at 4 °C. The concentration of LmrR and the various mutants were determined by using the calculated extinction coefficient $\epsilon_{280} = 25440 \text{ M}^{-1} \text{ cm}^{-1}$ (monomer), and LmrR_W96A $\epsilon_{280} = 19940 \text{ M}^{-1} \text{ cm}^{-1}$ (monomer). Calculations were done by Protparam on the ExPASy server. Typical expression yields obtained were 20-30 mg/L media.

Protein characterisation

UV-Vis experiments

Representative procedure:

A concentrated stock solution of iron (III) chloroproporphyrin IX complex (hemin) (3.58 mM) was prepared in DMSO and was subsequently diluted to 5 μM (2 mL) in PBS buffer (50 mM NaH_2PO_4 , 150 mM NaCl, pH 7.0). To this solution, 1 μM aliquots of the corresponding LmrR protein variants (dimer) were added, and the absorbance spectrum was recorded on a JACSO V-660 spectrophotometer. Protein stock concentrations were determined using Nanodrop 2000, Thermo Scientific.

Figure 8. Electronic absorption spectra upon the addition of a) LmrR and b) LmrR_W96A to 5 μM hemin. Buffer: 50 mM phosphate buffer, 150 mM NaCl, pH 7.0.

Fluorescence titration

Representative procedure:

Titration experiments were performed with 1-4 μM solutions of LmrR (2 mL) titrated with 80 μM solution of iron (III) chloroproporphyrin IX complex (hemin) in 50 mM NaH_2PO_4 , pH 7.0, containing 150 mM NaCl. 8 mM stock solution of hemin was prepared by dissolving the solid in DMSO. By subsequent dilutions with DMSO, 80 μM solution of hemin was prepared. 40 μM protein solutions (dimer) in buffer was prepared and then diluted x10. Emission spectra were recorded at room temperature after 2 min incubation. The sample was excited at 292 nm and emission spectra were recorded from 310 nm to 450 nm. The excitation and emission slit widths were 5 nm. The maximum of the emission spectra (345 nm) were plotted against the concentration ratio between hemin and protein dimer. The data obtained from the titration of the corresponding hemin with LmrR were fitted using non-linear regression analysis (program OriginPro 8.5), employing equation 1.⁶⁸

$$\text{FQ} = \frac{k}{2} (C_P + C_L + K_d - \sqrt{(C_P + C_L + K_d)^2 - 4C_L C_P}) \quad (1)$$

where FQ is the tryptophan fluorescence quenching (ΔQF) by hemin (expressed as $(I_{\text{max}} - I) / (I_{\text{max}} - I_0)$), with I_{max} the intensity of fluorescence in the absence of quencher, I the intensity of fluorescence upon addition of quencher and I_0 the intensity of fluorescence at saturation). C_P and C_L are the total concentrations of protein and hemin, respectively, and k is a constant.

Figure 9. Fluorescence titration experiment. The reported apparent dissociation constants are average of at least two independent experiments performed with independent batches of protein and hemin.

Resonance Raman spectroscopy

405 nm laser (ONDAX, 50 mW) was used in 180° back scattering mode with a Semrock 405 nm dichroic mirror. The laser was focused on the sample by a 7.5 cm focal length 2.5 cm diameter planoconvex lens and backscattered light collected and colimated by the same lens and passed through the dichroic and a second Rayleigh line rejection filter before entering the spectrometer (Shamrock300i, Andor Technology, 1200 l/mm 500 nm blaze grating) via a 7.5 cm focal length 2.5 cm diameter planoconvex lens with 12 micron slits. The dispersed light was collected by an ANDOR iDus-420-OE CCD using Andor Solis for spectrometer control and data acquisition.

Catalysis

Cyclopropanation

3

Representative procedure:

50 mM potassium phosphate buffer and a solution containing sodium dithionate (80 μ L, 100 mM stock solution) at pH 8.0 were combined in a sealed vial equipped with a stir bar and degassed by bubbling argon into the mixture for 10 min. Afterwards, a buffered solution containing the corresponding LmrR variant solution (11 μ M final concentration) was added into the reaction vial with a final volume of 720 μ L and continued degassing by flushing argon without bubbling. In a separate vial 80 μ L of a 100 μ M solution of hemin complex (pre-dissolved in buffer/DMSO, final fraction of DMSO 5% v/v) to a final hemin concentration of 10 μ M was degassed by bubbling argon through the solution for 10 min, and added into the reaction mixture *via* cannulation. After incubation for 30 min, a fresh stock solution of styrene (20 μ L, 1.2 M in DMSO) and EDA (20 μ L, 0.4 M in DMSO) were added to the catalyst solution and the reaction was stirred at 4 °C for 18 h under positive argon pressure. The reactions were stopped by extraction with 800 μ L of ethyl acetate containing 100 μ L of internal standard (2-methylanisole, 1 mM in EtOAc). The organic layers were dried over Na₂SO₄ and transferred to an autosampler vial for analysis. Calibration curves for quantification of the different cyclopropane products were constructed using available pure reference compounds and 2-methylanisole as internal

standard. For products **3a** and **3c**, the yield and stereoselectivities were analysed by GC (HP-5-50 °C (0 min), 10 °C/min 100 °C (6 min), 20 °C/min 280 °C (2 min); Total time: 22 min) and the enantioselectivity by HPLC (Chiralcel-OBH, n-heptane 100%, 0.5 ml/min). Product **3b** was analysed by GC (Chirasil-Dex CB isotherm 125 °C).

N-H insertion

The procedure described above for the cyclopropanation reaction was followed using aniline (**5**) instead of the styrene substrates, and the reaction times and temperatures indicated in Table 5.

Determination of absolute configuration of product **3a**

The absolute configuration of the favoured *trans*-product **3a** was assigned according to reference reactions carried out in the presence of a reported chiral Ruthenium(II)/Phenyloxazoline complex. The complex was synthesised following procedures adapted from the literature.⁶⁹

A mixture of (*S*)-(+)-2-phenylglycinol (1207.2 mg, 8.80 mmol) and Et₃N (4.5 mL, 32 mmol) were dissolved in 15 mL of dichloromethane. A solution of benzoylchloride (0.9 mL, 8 mmol) in 4 mL DCM was added to the mixture at 0 °C and it was left stirring overnight at room temperature. Subsequently, the mixture was concentrated under vacuum and re-dissolved in 20 mL of CHCl₃. 3 mL of SOCl₂ (40 mmol) were added at 0 °C and the solution was left stirring at rt for 24 h. The solvent was then removed under vacuum and to the remaining product 50 mL of saturated NaHCO₃ was added and vigorously stirred for 5 minutes. The organic product was extracted with DCM, dried over Na₂SO₄ and concentrated under reduced pressure. The solid was dissolved in 1,4-dioxane and 32 mL of 2.5 M NaOH was added slowly at 0 °C. The reaction mixture was stirred for 4 h at rt and the solvent was evaporated under vacuum, followed by addition of 25 mL of water and extraction with CH₂Cl₂. The solvent was evaporated under reduced pressure and crude

product was purified by silica column chromatography (EtOAc:pentane 10:1) to afford (S)-4,5-dihydro-2,4-diphenyloxazole as a yellow oil.

A mixture of the synthesised (S)-4,5-dihydro-2,4-diphenyloxazole (67.0 mg, 0.30 mmol), $[\text{RuCl}_2(\text{benzene})]_2$ (75.0 mg, 0.15 mmol), KPF_6 (220.9 mg, 1.2 mmol), and NaOH(aq) (0.3 mL, 0.30 mmol, 1 M) in 5 mL of degassed CH_3CN was stirred for 48 h at 80°C . The reaction mixture was left cool down and extracted with CH_2Cl_2 after addition of 20 mL of NaHCO_3 for phase separation. The organic layer was dried over Na_2SO_4 and concentrated under vacuum. The product was purified by silica gel column chromatography ($\text{CH}_3\text{CN}:\text{CH}_2\text{Cl}_2$ 1:20). ^1H NMR (400 MHz, Chloroform-*d*) δ 7.55 (d, $J = 7.4$ Hz, 1H), 7.52 (s, 1H), 7.28 – 7.35 (m, 5H), 7.20 (m, 1H), 6.97 (m, 1H), 5.30 (m, 1H), 5.10 (m, 1H), 4.55 (dd, $J = 8.8, 7.3$ Hz, 1H), 2.53 – 2.47 (m, 3H), 2.25 (s, 3H), 2.20 (s, 3H), 2.04 (s, 3H).

Porphyrin synthesis

Under argon atmosphere, chlorin e6 (100 mg, 0.1676 mmol) was dissolved in acetone (100 mL) and to the solution L-ascorbic acid (2.952 g, 100 eq.) in acetone (100 mL) and $\text{Fe(II)Cl}_2 \cdot 4\text{H}_2\text{O}$ (4.665 g, 140 eq.) were added. The reaction mixture was refluxed at 60°C for 4 h. The reaction mixture was cooled down at rt and 100 mL of dichloromethane were added. The solution was washed with saturated brine, then with HCl (100 mL, 0.01 M), and finally with 200 mL of H_2O . The organic layer was dried over MgSO_4 and the solvent

was evaporated down to 20 mL. The solution was centrifuged for 30 minutes to precipitate most of the salts presents. The supernatant was filtered out and dialysed to remove the excess of apo-porphyrin. The solvent was evaporated to yield 73% of Fe(Cl)-chlorin e6 as a dark green solid.

Metallation of the porphyrin was confirmed by inspection of the UV-Vis absorption spectrum. The new bands obtained after metallation of the porphyrin correspond to those reported in the literature for the fully characterised Fe(II)-Ce6.¹⁹

Figure 10. Electronic absorption spectra of the apo-porphyrin Chlorin e6 (left) and the metallated porphyrin with Fe(II), Fe(II)-Chlorin e6 (right).

3.4. ACKNOWLEDGEMENTS

The authors would like to acknowledge Dr. Lara Villarino who designed the hemeC-LmrR system, performed the alanine scanning experiments and contributed to expansion of the reaction scope. Prof. Kathryn E. Splan for performing some of the spectroscopic measurements, Dr. Lur Alonso Cotchico for the computational analysis and Dr. Eswar Reddem for the crystallographic experiments. Prof. Wesley R. Browne is acknowledged for the Raman experiments.

3.6. REFERENCES

- 1 J. Everse, in *Encyclopedia of Biological Chemistry: Second Edition*, 2013, pp. 532–538.
- 2 P. Rydberg, E. Sigfridsson and U. Ryde, *J. Biol. Inorg. Chem.*, 2004, **9**, 203–223.

-
- 3 J. C. Barona-Castaño, C. C. Carmona-Vargas, T. J. Brocksom, K. T. De Oliveira, M. Graça, P. M. S. Neves, M. Amparo and F. Faustino, *Molecules*, 2016, **21**, 310.
- 4 J. F. Longevial, S. Clément, J. A. Wytke, R. Ruppert, J. Weiss and S. Richeter, *Chem. - A Eur. J.*, 2018, **24**, 15442–15460.
- 5 O. F. Brandenburg, R. Fasan and F. H. Arnold, *Curr. Opin. Biotechnol.*, 2017, **47**, 102–111.
- 6 C. K. Prier and F. H. Arnold, *J. Am. Chem. Soc.*, 2015, **137**, 13992–14006.
- 7 M. Bordeaux, V. Tyagi and R. Fasan, *Angew. Chem. Int. Ed.*, 2015, **54**, 1744–1748.
- 8 P. S. Cohelo, E. M. Brustad, A. Kannan and F. H. Arnold, *Science.*, 2012, **3747**, 307–311.
- 9 J. G. Gober, A. E. Rydeen, E. J. Gibson-O’Grady, J. B. Leuthaeuser, J. S. Fetrow and E. M. Brustad, *ChemBioChem*, 2016, **17**, 394–397.
- 10 M. J. Weissenborn and R. M. Koenigs, *ChemCatChem*, 2020, **12**, 2171–2179.
- 11 C. C. Farwell, R. K. Zhang, J. A. McIntosh, T. K. Hyster and F. H. Arnold, *ACS Cent. Sci.*, 2015, **1**, 89–93.
- 12 V. Tyagi and R. Fasan, *Angew. Chem. Int. Ed.*, 2016, **55**, 2512–2516.
- 13 M. J. Weissenborn, S. A. Löw, N. Borlinghaus, M. Kuhn, S. Kummer, F. Rami, B. Plietker and B. Hauer, *ChemCatChem*, 2016, **8**, 1636–1640.
- 14 Z. J. Wang, N. E. Peck, H. Renata and F. H. Arnold, *Chem. Sci.*, 2014, **5**, 598–601.
- 15 G. Sreenilayam and R. Fasan, *Chem. Commun.*, 2015, **51**, 1532–1534.
- 16 V. Tyagi, R. B. Bonn and R. Fasan, *Chem. Sci.*, 2015, **6**, 2488–2494.
- 17 S. B. J. Kan, R. D. Lewis, K. Chen and F. H. Arnold, *Science.*, 2016, **354**, 1048–1051.
- 18 S. B. J. Kan, Y. Huang, K. Gumulya, K. Chen and F. H. Arnold, *Synfacts*, 2017, **552**, 132–136.
- 19 G. Sreenilayam, E. J. Moore, V. Steck and R. Fasan, *ACS Catal.*, 2017, **7**, 7629–

- 7633.
- 20 E. W. Reynolds, M. W. McHenry, F. Cannac, J. G. Gober, C. D. Snow and E. M. Brustad, *J. Am. Chem. Soc.*, 2016, **138**, 12451–12458.
- 21 H. M. Key, P. Dydio, D. S. Clark and J. F. Hartwig, *Nature*, 2016, **534**, 534–537.
- 22 G. Sreenilayam, E. J. Moore, V. Steck and R. Fasan, *Adv. Synth. Catal.*, 2017, **359**, 2076–2089.
- 23 S. I. Ozaki, T. Matsui and Y. Watanabe, *J. Am. Chem. Soc.*, 1996, **118**, 9784–9785.
- 24 S. I. Ozaki, M. P. Roach, T. Matsui and Y. Watanabe, *Acc. Chem. Res.*, 2001, **34**, 818–825.
- 25 H. Kries, R. Blomberg and D. Hilvert, *Curr. Opin. Chem. Biol.*, 2013, **17**, 221–228.
- 26 T. Heinisch, M. Pellizzoni, M. Dürrenberger, C. E. Tinberg, V. Köhler, J. Klehr, D. Häussinger, D. Baker and T. R. Ward, *J. Am. Chem. Soc.*, 2015, **137**, 10414–10419.
- 27 A. Ilie and M. T. Reetz, *Isr. J. Chem.*, 2015, **55**, 51–60.
- 28 A. P. Green, T. Hayashi, P. R. E. Mittl and D. Hilvert, *J. Am. Chem. Soc.*, 2016, **138**, 11344–11352.
- 29 T. Mori, D. Hilvert, M. Pott, T. Hayashi, P. R. E. Mittl and A. P. Green, *J. Am. Chem. Soc.*, 2018, **140**, 1535–1543.
- 30 C. T. Choma, J. D. Lear, M. J. Nelson, P. L. Dutton, D. E. Robertson and W. F. DeGrado, *J. Am. Chem. Soc.*, 1994, **116**, 856–865.
- 31 I. V. Korendovych, A. Senes, Y. H. Kim, J. D. Lear, H. C. Fry, M. J. Therien, J. K. Blasie, F. A. Walker and W. F. DeGrado, *J. Am. Chem. Soc.*, 2010, **132**, 15516–15518.
- 32 C. J. Reedy and B. R. Gibney, *Chem. Rev.*, 2004, **104**, 617–650.
- 33 Y. W. Lin, E. B. Sawyer and J. Wang, *Chem. - An Asian J.*, 2013, **8**, 2534–2544.
- 34 J. P. Mahy, J. D. Maréchal and R. Ricoux, *Chem. Commun.*, 2015, **51**, 2476–2494.

- 35 A. D'Souza, X. Wu, E. K. L. Yeow and S. Bhattacharjya, *Angew. Chem. Int. Ed.*, 2017, **56**, 1–6.
- 36 H. Agustindari, J. Lubelski, H. B. Van Den Berg Van Saparoea, O. P. Kuipers and A. J. M. Driessen, *J. Bacteriol.*, 2008, **109**, 759–763.
- 37 J. Bos, W. R. Browne, A. J. M. Driessen and G. Roelfes, *J. Am. Chem. Soc.*, 2015, **137**, 9796–9799.
- 38 L. Villarino, S. Chordia, L. Alonso-Cotchico, E. R. Reddem, A.-M. Thunnissen, J.-D. Maréchal and G. Roelfes, *ACS Catal.*, 2020, **10**, 11783–11790.
- 39 G. Roelfes, *Acc. Chem. Res.*, 2019, **52**, 545–556.
- 40 J. Bos, A. García-Herraiz and G. Roelfes, *Chem. Sci.*, 2013, **4**, 3578–3582.
- 41 K. A. De Villiers, C. H. Kaschula, T. J. Egan and H. M. Marques, *J. Biol. Inorg. Chem.*, 2007, **12**, 101–117.
- 42 A. Tinoco, V. Steck, V. Tyagi and R. Fasan, *J. Am. Chem. Soc.*, 2017, **139**, 5293–5296.
- 43 T. Heel, J. A. McIntosh, S. C. Dodani, J. T. Meyerowitz and F. H. Arnold, *ChemBioChem*, 2014, **15**, 2556–2562.
- 44 A. Caballero, A. Prieto, M. M. Díaz-Requejo and P. J. Pérez, *Eur. J. Inorg. Chem.*, 2009, 1137–1144.
- 45 R. Fasan, V. Steck, Y. Wei, A. Tinoco and Y. Zhang, *J. Am. Chem. Soc.*, 2017, **140**, 1649–1662.
- 46 C. Damiano, P. Sonzini and E. Gallo, *Chem. Soc. Rev.*, 2020, **49**, 4867–4905.
- 47 L. Villarino, K. E. Splan, E. Reddem, L. Alonso-Cotchico, C. Gutiérrez de Souza, A. Lledós, J.-D. Maréchal, A.-M. W. H. Thunnissen and G. Roelfes, *Angew. Chem. Int. Ed.*, 2018, **57**, 7785–7789.
- 48 K. Tishinov, N. Fei and D. Gillingham, *Chem. Sci.*, 2013, **4**, 4401–4406.
- 49 L. K. Baumann, H. M. Mbuvi, G. Du and L. K. Woo, *Organometallics*, 2007, **26**, 3995–4002.
- 50 H. Rein, S. Maricic, G. R. Jänig, S. Vuk-Pavlovic, B. Benko, O. Ristau and K. Ruckpaul, *BBA - Protein Struct.*, 1976, **446**, 325–330.

- 51 J. T. Szilagyi, V. Mishin, D. E. Heck, Y. H. Jan, L. M. Aleksunes, J. R. Richardson, N. D. Heindel, D. L. Laskin and J. D. Laskin, *Toxicol. Sci.*, 2016, **151**, 150–159.
- 52 K. Oohora and T. Hayashi, in *Methods in Enzymology*, 2016, pp. 439–454.
- 53 L. Fruk, C. H. Kuo, E. Torres and C. M. Niemeyer, *Angew. Chem. Int. Ed.*, 2009, **48**, 1550–1574.
- 54 T. Hayashi, in *Coordination Chemistry in Protein Cages: Principles, Design, and Applications*, 2013, pp. 87–110.
- 55 R. L. Khade and Y. Zhang, *J. Am. Chem. Soc.*, 2015, **137**, 7560–7563.
- 56 Y. Zhu and R. B. Silverman, *J. Org. Chem.*, 2007, **72**, 233–239.
- 57 T. Shibata, S. Nagao, M. Fukaya, H. Tai, S. Nagatomo, K. Morihashi, T. Matsuo, S. Hirota, A. Suzuki, K. Imai and Y. Yamamoto, *J. Am. Chem. Soc.*, 2010, **132**, 6091–6098.
- 58 P. S. Coelho, Z. J. Wang, M. E. Ener, S. A. Baril, A. Kannan, F. H. Arnold and E. M. Brustad, *Nat. Chem. Biol.*, 2013, **9**, 485–487.
- 59 D. W. Shipp, F. Sinjab and I. Notingher, *Adv. Opt. Photonics*, 2017, **9**, 315.
- 60 A. Warshel, *Annu. Rev. Biophys. Bioeng.*, 1977, **6**, 273–300.
- 61 T. G. Spiro and T. C. Strekas, *J. Am. Chem. Soc.*, 1974, **96**, 338–345.
- 62 A. L. Verma and H. J. Bernstein, *J. Raman Spectrosc.*, 1974, **2**, 163–173.
- 63 J. P. M. Schelvis, V. Berka, G. T. Babcock and A. L. Tsai, *Biochemistry*, 2002, **41**, 5695–5701.
- 64 S. Todorovic and M. Teixeira, *J. Biol. Inorg. Chem.*, 2018, **23**, 647–661.
- 65 S. Hu, K. M. Smith and T. G. Spiro, *J. Am. Chem. Soc.*, 1996, **118**, 12638–12646.
- 66 A. Rioz-Martínez, J. Oelerich, N. Ségaud and G. Roelfes, *Angew. Chem. Int. Ed.*, 2016, **55**, 14136–14140.
- 67 L. A. Andersson, M. Mylrajan, E. P. Sullivan and S. H. Strauss, *J. Biol. Chem.*, 1989, **264**, 19099–19102.
- 68 F. Zsila, Z. Bikádi and M. Simonyi, *Bioorganic Med. Chem.*, 2004, **12**, 3239–

3245.

- 69 A. M. Abu-Elfotoh, K. Phomkeona, K. Shibatomi and S. Iwasa, *Angew. Chem. Int. Ed.*, 2010, **49**, 8439–8443.

CHAPTER 4

Artificial Metalloenzymes based on TetR Proteins and Cu(II) for Enantioselective Friedel-Crafts Alkylation Reactions

Abstract

The supramolecular approach is among the most convenient methodologies for creating artificial metalloenzymes (ArMs). Usually, this approach involves the binding of a transition metal ion complex to a biomolecular scaffold via its ligand, which also modulates the catalytic properties of the metal ion. Herein, we report ArMs based on the proteins CgmR, RamR and QacR from the TetR family of multidrug resistance regulators (MDRs) and Cu²⁺ ions, assembled without the need of a ligand. These ArMs catalyse the enantioselective vinylogous Friedel-Crafts alkylation reaction with up to 75% ee. Competition experiments with ethidium and rhodamine 6G confirm that the reactions occur in the chiral environment of the hydrophobic pocket. It is proposed that the Cu²⁺-substrate complex is bound via a combination of electrostatic and π -stacking interactions provided by the second coordination sphere. This approach constitutes a fast and straightforward way to assemble metalloenzymes and may facilitate future optimisation of the protein scaffolds via mutagenesis or directed evolution approaches.

Published in:

C. Gutiérrez de Souza, M. Bersellini and G. Roelfes, *ChemCatChem*, 2020, **12**, 3190–3194.

4.1. INTRODUCTION

The field of artificial metalloenzymes (ArMs) has the potential to expand dramatically the possibilities of enzyme engineering towards achieving biocatalysis of reactions that have no equivalent in nature.¹⁻³ ArMs aim to combine the high efficiency of natural enzymes, with the broad reaction scope characteristic of homogeneous transition metal catalysts. These biocatalysts consist of a biological scaffold (DNA or protein) in which a catalytically active transition metal complex is embedded. The bioscaffold provides the second coordination sphere interactions that are envisioned to contribute to the rate acceleration of the reaction and a chiral environment that allows catalytic transformations to proceed in an enantioselective manner. The metal complex is responsible for the catalytic activity. A key role is held by the ligand, which modulates the electronic properties of the metal ion and, hence, attenuates the reactivity. Additionally, the ligand can be used to anchor the metal complex inside the biomolecular scaffold *via* a tethered “Trojan horse” moiety, as applied successfully in the (strept-)avidin/biotin based systems, or directly by making use of supramolecular interactions between the ligand and the protein scaffold.⁴⁻¹¹

4

The supramolecular approach is attractive because of its simplicity, since the system self-assembles upon mixing the protein and the metal complex. This approach does not require covalent modifications of the scaffold and thus facilitates ArM optimisation. Nevertheless, it is limited in the scope of proteins that can be used as scaffold, as strong and specific interactions between the protein and the cofactor are generally required.¹¹⁻¹⁴

In our group, we have applied the supramolecular approach to develop ArMs based on the protein LmrR (Lactococcal Multidrug Resistance Regulator) from the PadR family of Multidrug Resistance Regulators (MDRs).¹⁵⁻¹⁷ LmrR is a homodimeric protein with a spacious hydrophobic pocket at the dimer interface that captures aromatic compounds by π -stacking interactions in between the indole moieties of tryptophans W96 and W96', located in the middle of the pore (Figure 1a). The affinity of these two tryptophans for binding planar metal complexes, e.g. the Cu^{2+} complex of 1,10-phenanthroline ($\text{Cu}(\text{phen})$), was exploited to create LmrR-based ArMs for the

enantioselective Friedel-Crafts alkylation of α - β unsaturated acyl imidazole with indole.¹⁸

Figure 1. Cartoon representations of a) LmrR with Hoechst 33342 (PDB: 3F8C). b) CgmR with ethidium bromide (PDB: 2ZOZ). c) RamR with ethidium bromide (PDB: 3VVY). d) QacR with ethidium bromide (PDB: 1JTY).

In the current study, we aimed to expand the array of bioscaffolds to other MDRs, in particular proteins belonging to the TetR family, which are well-studied bacterial transcriptional repressors.¹⁹ These are homodimeric proteins of ~20 KDa per monomer that, unlike LmrR, have one binding pocket per monomer. We hypothesised that differences in the structures of the hydrophobic pockets could give rise to new ArMs with reactivities and selectivities complementary to that of LmrR. Here, we show that these novel ArMs are able to catalyse the enantioselective Friedel-Crafts alkylation of α , β -unsaturated imidazoles with indoles using $\text{Cu}(\text{NO}_3)_2$ alone, without the need of any ligand for the metal ion.

CgmR, RamR and QacR were selected as scaffolds for this study because crystal structures of these proteins with different hydrophobic and cationic drugs were available.^{20–27} The ability of these proteins to bind aromatic and cationic molecules suggested their capability to capture hydrophobic metal complexes. Moreover, these protein scaffolds have previously proven their utility in artificial metalloenzyme design when combined with a metal-binding unnatural amino acid.²⁸

4.2. RESULTS AND DISCUSSION

The ArMs were prepared by self-assembly of CgmR, RamR and QacR with $\text{Cu}(\text{phen})(\text{NO}_3)_2$ complex or $\text{Cu}(\text{NO}_3)_2$. In case of QacR, a variant containing C72A and C141S mutations was used.²⁸ UV-Vis spectroscopy studies suggested interaction between the proteins and these copper species (Figure 3), albeit no quantitative binding data could be obtained. The Friedel-Crafts alkylation

(F-C reaction) of 2-methyl-(1H)-indole (**2a**) with 1-(1-methyl-1H-imidazole-2-yl)but-2-en-1-one (**1a**) was selected as the model reaction. This reaction involves the conjugate addition of a heteroaromatic nucleophilic indole to an α,β -unsaturated 2-acyl-(1-methyl)imidazole substrate and it is typically catalysed by a Lewis acid, e.g. Cu^{2+} .²⁹

Initial studies were performed using $\text{Cu}(\text{phen})(\text{NO}_3)_2$ or $\text{Cu}(\text{NO}_3)_2$ (90 μM) with a slight excess of 1.3 equivalents of each protein (120 μM in protein dimer), 1 mM **1a** and 1 mM **2a** in 20 mM MOPS pH 7.0, containing 500 mM NaCl. The high salt concentrations were required to avoid precipitation of the proteins during the reaction. The reactions were incubated under continuous inversion at 4 °C for 72 hours. No product formation was observed in absence of $\text{Cu}(\text{NO}_3)_2$ or $\text{Cu}(\text{phen})(\text{NO}_3)_2$. The reaction in presence of $\text{Cu}(\text{NO}_3)_2$ alone, without protein, resulted in formation of a racemic mixture of the Friedel-Crafts product in 22% yield, indicating that the Cu^{2+} ion is required for activation of the imidazole substrate. In addition, the proteins alone led to low yields (up to 11%) and ee's. The yields are somewhat higher than observed without any catalyst, which might be due to non-specific interactions of the substrates with the protein, resulting in a higher effective molarity.

4

The reaction catalysed by $\text{Cu}(\text{phen})(\text{NO}_3)_2$ gave a significantly higher yield compared to using $\text{Cu}(\text{NO}_3)_2$ as catalyst (Table 1, entries 2 and 3). Unexpectedly, the reactions performed with the MDR proteins in combination with $\text{Cu}(\text{phen})(\text{NO}_3)_2$ resulted in lower yields (ranging from 21 to 36% depending on the protein scaffold used) than with MDR/ $\text{Cu}(\text{NO}_3)_2$ (52 to 78% yield). The best results were obtained with QacR and $\text{Cu}(\text{NO}_3)_2$, which led to an increase of yield up to 78% and similar enantioselectivity, 34%, compared to QacR with $\text{Cu}(\text{phen})(\text{NO}_3)_2$ (Table 1). The inferior yields obtained with the TetR ArMs prepared with $\text{Cu}(\text{phen})(\text{NO}_3)_2$ compared to those prepared from $\text{Cu}(\text{NO}_3)_2$, without ligand, was a surprising finding, since it is in marked contrast to what was previously reported for LmrR. For the latter, the presence of the ligand for the metal ion was required.¹⁸

The enantioselectivities observed indicate that the catalytic reaction is occurring within the chiral environment of the proteins. Moreover, the higher yield obtained with Cu^{2+} and the proteins together compared to the reaction

performed only with Cu²⁺ salt suggests that all the protein scaffolds also contribute to acceleration of the reaction.

Table 1. Vinylogous Friedel-Crafts alkylation reactions catalysed by MDR/Cu(II) or MDR/Cu(phen).

Reaction scheme: 1a + 2a $\xrightarrow[\text{20 mM MOPS, 500 mM NaCl, pH 7.0, 4 °C, 72 h}]{\text{Cu(NO}_3)_2 \text{ or Cu(phen) (90 } \mu\text{M), protein (120 } \mu\text{M)}}$ 3a

Entry	Catalyst	Yield (%) ^a	ee (%) ^a
1	-	< 5	-
2	Cu ²⁺	22 ± 7	-
3	Cu(phen)	43 ± 2	-
4	RamR	5 ± 6	8 ± 6
5	Cu ²⁺ ⊂RamR	57 ± 9	29 ± 3
6	Cu(phen)⊂RamR	21 ± 13	34 ± 7
7	CgmR	6 ± 3	10 ± 21
8	Cu ²⁺ ⊂CgmR	52 ± 12	13 ± 3
9	Cu(phen)⊂CgmR	30 ± 3	15 ± 1
10	QacR	11 ± 7	13 ± 19
11	Cu ²⁺ ⊂QacR	78 ± 11	34 ± 3
12	Cu(phen)⊂QacR	36 ± 6	30 ± 2

^aYields and ee's were determined by HPLC. Yields were calculated using 2-phenylquinoline as internal standard. All the results listed correspond to the average of two independent experiments, each carried out in duplicate. Errors listed are

standard deviations. ^b In all cases the (-) enantiomer was obtained in excess, as determined by comparison of the elution order in chiral HPLC to literature reports.^{29,30}

To determine the substrate scope of these new ArMs, the F-C alkylation reaction was performed with the TetR/Cu²⁺ based ArMs with three α - β -substituted imidazoles (**1a–c**) and a selection of substituted indoles (**2a–f**, Table 2). The indole scope was tested on β -methyl substituted imidazole substrate (**1a**). 2-methyl-(1H)-indole (**2a**) was found to be the most reactive among the indoles, with yields and ee's of product **3a** ranging from 52 to 78% and from 13 to 34%, respectively. 5-methoxy-(1H)-indole (**2d**) also showed good reactivity, giving rise to yields of **3d** between 19 and 39% and ee's between 9 and 37%. Unsubstituted indole led to low yields of product **2b**, whereas indoles harbouring a phenyl moiety at position 2 and deactivating groups at position 5 of the indole ring, such as chlorine and bromine, were mostly unreactive (Table 2, entries 3, 5 and 6).¹⁸ Only in case of Cu²⁺⊂QacR, a slightly increased yield of the reaction with 5-chloro-(1H)-indole substrate (**2e**) with a moderate enantioselectivity was found.

4

2-methyl-(1H)-indole (**2a**) was then combined with β -phenyl substituted imidazole (**1b**), resulting in low yields and enantioselectivities of **3g** with Cu²⁺⊂RamR and Cu²⁺⊂CgmR. However, Cu²⁺⊂QacR led to moderate yield (54%) of product, which, unlike the other reactions where activity was observed, was similar to that obtained with Cu²⁺ alone. Nevertheless, the product was obtained in 75% ee, the highest enantioselectivity found in this study, which is clear proof for the protein being involved in the reaction, as it is the only source of chirality. β -tert-butyl substituted imidazole (**1c**) resulted in no product formation, regardless of the presence or absence of any of the proteins. In some cases, the appearance of an extra pair of peaks in the HPLC chromatograms was observed. Based on our previous work, we assigned these peaks to the two enantiomers of the product of conjugate addition of water to the enone (Figure 5).^{30,31} For the reaction with substrate **3h**, this was confirmed by comparison of the retention times with those of an independently prepared sample (Figure 4). The formation of this side product was outcompeted by the F-C alkylation reaction where higher yields were obtained, but was more prominent in the cases where F-C reaction yields were lower.

Table 2. Results of the substrate scope of the vinylogous Friedel–Crafts alkylation reactions.^a

Entry	Product	Y / ee (%) ^b Cu ²⁺ Cu ²⁺ -cRamR Cu ²⁺ -cCgmR Cu ²⁺ -cQacR	Entry	Product	Y / ee (%) ^b Cu ²⁺ Cu ²⁺ -cRamR Cu ²⁺ -cCgmR Cu ²⁺ -cQacR
1	3a	22 ± 7 / - 57 ± 9 / 29 ± 3 (-) 52 ± 12 / 13 ± 3 (-) 78 ± 11 / 34 ± 3 (-)	5	3e	<5 / n.d. <5 / n.d. <5 / n.d. 17 ± 8 / 38 ± 5 (+)
2	3b	7 ± 8 / - 5 / n.d. 7 ± 3 / 27 ± 3 (+)-R 27 ± 12 / 26 ± 2 (+)-R	6	3f	<5 / n.d. <5 / n.d. <5 / n.d. <5 / n.d.
3	3c	<5 / n.d. <5 / n.d. <5 / n.d. <5 / n.d.	7	3g	54 ± 9 / - <5 / n.d. 18 ± 4 / 6 ± 2 59 ± 7 / 75 ± 4
4	3d MeO	10 ± 6 / - 19 ± 3 / 35 ± 3 (+) 31 ± 14 / 37 ± 1 (+) 39 ± 10 / 9 ± 4 (+)	8	3h	<5 / n.d. <5 / n.d. <5 / n.d. <5 / n.d.

^aTypical conditions: 90 μM Cu(NO₃)₂ (9 mol%) loading with 1.3 equivalents of protein (120 μM). ^bYields and ee were determined by HPLC using 2-phenylquinoline as internal standard. For yields <5% ee's were not determined. All the results listed correspond to the average of two independent experiments, each of them carried out in duplicate. Errors listed are standard deviations. Signs of optical rotation and absolute configuration were assigned by comparison to the literature.^{29,30}

Overall, among the ArMs investigated here, $\text{Cu}^{2+}\text{-QacR}$ was found to be the most active and enantioselective catalyst, except with 5-methoxy-(1H)-indole (**2d**) where we observed slightly higher activity, but lower selectivity, with $\text{Cu}^{2+}\text{-QacR}$ than with $\text{Cu}^{2+}\text{-RamR}$ and $\text{Cu}^{2+}\text{-CgmR}$. Furthermore, $\text{Cu}^{2+}\text{-QacR}$, also showed a broader substrate scope compared to $\text{Cu}^{2+}\text{-RamR}$ and $\text{Cu}^{2+}\text{-CgmR}$.

The better performance of QacR in catalysis may be related to its high versatility with regard to drug recognition, i. e. it binds a wide range of structurally diverse cationic compounds.²⁴ QacR's binding flexibility, is believed to be due to several negatively charged residues (Glu57, 58, 90 and 120) within the pocket, responsible for attracting different cationic compounds and neutralising their charges; as well as to a number of aromatic residues (Trp61, Tyr93, Tyr103, Tyr123 and Phe162) which are involved in the binding of these compounds through general hydrophobic and π -stacking interactions. Each binding site of QacR consists of two overlapping mini-pockets, each of which involves interaction of the drugs with different residues.^{24,25,27} To confirm that these reactions catalysed by $\text{Cu}(\text{NO}_3)_2$, in absence of ligand, do take place inside the hydrophobic pocket of the protein and, for example, not on the protein surface, we performed inhibition studies in presence of ethidium bromide, which is known to bind within the pockets of CgmR, RamR and QacR, with reported binding constants of 610 nM,²¹ 14.6 μM ²³ and 2.35 μM ,²⁵ respectively. We anticipated that ethidium will compete with the substrate- Cu^{2+} complex for binding into the pockets of the proteins. Since the reaction of **1a** with **2a** is protein accelerated, expulsion of the substrate- Cu^{2+} complex from the pocket by ethidium bromide would cause a decrease in yield and ee. In the case of QacR, the inhibition experiments were also performed with rhodamine 6G, since it is known to bind in a different part of the binding pocket of QacR than ethidium bromide.²⁶

Using $\text{Cu}(\text{NO}_3)_2$ alone, without protein, a small decrease in yield was observed upon addition of the first equivalent of inhibitor, but the yield remained unchanged upon addition of further equivalents (Table 3). This makes it unlikely that the ethidium bromide or rhodamine 6G directly inhibit the reaction catalysed by Cu^{2+} alone. In the case of RamR, the presence of ethidium bromide caused a slight decrease in yield but did not affect the enantioselectivity (Figure 2). In contrast, for CgmR, a decrease in both yield

and ee was observed in presence of ethidium bromide. A strong effect was also observed for Cu^{2+} -QacR. The presence of increasing amounts of both ethidium bromide or rhodamine 6G resulted in a gradual decrease in yield and enantioselectivity (Figure 2). The observation that blocking the binding sites of CgmR and QacR reduced both activity and selectivity of the ArMs, led us to conclude that the catalysed reaction indeed takes place inside the pockets of these proteins

Figure 2. Results of the inhibition experiments for the vinyllogous Friedel-Crafts alkylation reaction. Typical conditions: 90 μM $\text{Cu}(\text{NO}_3)_2$ (9 mol%) loading with 1.3 eq of protein (120 μM) and indicated amount of dye per monomer of protein. All results correspond to the average of two independent experiments, each carried out in duplicate. Yields (Y) and ee were determined by HPLC using 2-phenylquinoline as internal standard. Error bars represent standard deviations.

A remarkable result from this study is that, in contrast to our earlier work on LmrR, the TetR based ArMs perform better in catalysis using Cu^{2+} ions alone compared to the $\text{Cu}(\text{phen})$ complex. This may be due to the $\text{Cu}(\text{phen})$

complex binding in an unfavourable orientation or to the smaller size of the pockets of these proteins compared to LmrR, causing there to be not enough space for the Cu(phen)-substrate complex. The question then is what is driving the reaction to occur inside the pocket in case of the Cu(NO₃)₂ catalysed reactions? Inspection of the X-ray crystal structures does not reveal an obvious Cu²⁺ binding site. Thus, it is unlikely that the activated complex is held in place solely by dative anchoring via the Cu²⁺ ion. This is further supported by the observation that assembly of the metalloprotein, followed by dialysis, resulted in a significant drop in activity in the reaction, suggesting the Cu²⁺ ion, in the absence of substrate, is only weakly bound to the protein (Table 4).

4 Instead, we hypothesise that the highly π -conjugated and cationic complex formed by Cu²⁺ coordinated to the imidazole substrate (**1**) does have enough affinity to bind in the pockets of these MDR proteins. This occurs without the need of a ligand such as phenanthroline. In this case π - π stacking interactions and possibly cation π interactions will play a key role. This has precedent in recent work by Onoda, Hayashi and co-workers, who used a pyrene conjugated β -barrel protein for the catalysis of enantioselective Diels-Alder reactions by Cu²⁺ salts.³² In their study π -stacking interactions between the substrate and the pyrene moiety were found to be key to the observed catalysis. The available X-ray crystal structures of these TetR proteins with drugs bound further support this hypothesis: they show that π - π interactions with phenylalanine and tryptophans play a key role in drug binding.^{21,22,26} Moreover, the pockets of, for example, QacR are rich in negatively charged residues (Glu57, 58, 90 and 120) which can contribute further to the binding of these cationic complexes.

Lastly, we tested whether we could take advantage of this simple assembly methodology to form the ArMs in more complex media like cell-free extracts (CFEs). The reaction was run in presence of lysates of the CFEs of CgmR, RamR and QacR, as well as LmrR and SUMO, as proteins known to not bind Cu²⁺. In all cases the reactions performed in presence of the different proteins but without Cu(NO₃)₂ did not lead to product formation. LmrR and SUMO, as well as CgmR and RamR, in combination with Cu(NO₃)₂ resulted in moderate yields between 39 and 68% of racemic mixture of product **3a**. The lack of enantioselectivity suggested that these proteins are not able to bind

Cu^{2+} ions in CFE media. On the contrary, the reactions performed in presence of QacR and $\text{Cu}(\text{NO}_3)_2$, resulted in moderate yields (61%) and modest ee (20%) (Table 5). These results indicate that despite CgmR and RamR interact with Cu^{2+} ions when the reactions are performed with the purified proteins in buffer solutions, they do not bind Cu^{2+} in the more complex media of the CFE. However, the observation that Cu^{2+} -QacR CFE was able to induce some enantioselective product formation, indicated that the interaction of this protein scaffold with Cu^{2+} is strong enough to be detectable even in complex media.

4.3. CONCLUSIONS

In this work, we have shown that CgmR, RamR and QacR, proteins from the TetR family of MDRs, are good scaffolds for the design of ArMs using the supramolecular approach. The side chains of amino acids located in their pockets create networks of hydrophobic interactions that allow the formation of ArMs for the enantioselective vinylogous Friedel-Crafts alkylation reaction using Cu^{2+} ions and an aromatic substrate without the need of an additional ligand for the metal ion. Interestingly, the reaction also led to modest ee in the lysate of cell-free extract of QacR with $\text{Cu}(\text{NO}_3)_2$, indicating that the interactions between Cu^{2+} and QacR are strong enough to partially prevail in the complex CFE media. The use of CFE to, for instance, screen for beneficial mutations is advantageous since it avoids the step of protein purification. However, the ee obtained with CFE of QacR is still too low to be able to detect significant differences between mutants. Therefore, further optimisation of the expression protocol with CFE to try to improve these results—for instance by increasing the CFE concentration—would be desired to evaluate whether further screenings could be performed in CFE.

Overall, the approach presented in this chapter constitutes a fast and straightforward way to assemble metalloenzymes which provides significant advantages compared to standard supramolecularly-assembled ArMs, as it avoids the need for an external ligand for the metal ion. This circumvents also the need for a ligand-screening step while designing new ArMs and it facilitates optimisation of the protein scaffold via mutagenesis or directed evolution approaches.

4.4. EXPERIMENTAL SECTION

Materials and methods

All the chemicals and reagents were purchased from commercial suppliers (Sigma Aldrich or Acros) and used without further purification. Flash chromatography was performed using Reveleris® X2 Flash Chromatography System. ¹H-NMR and ¹³C-NMR spectra were recorded on a Varian 400 (400 MHz) spectrometer in CDCl₃. Chemical shifts values (δ) are denoted in ppm using residual solvent peaks as the internal standard (CHCl₃: δ 7.26 for 1H). ¹H-NMR results are reported in the conventional form: chemical shifts, multiplicity (d = doublet, dd = doublet of doublet, m= multiplet, q=quartet, s=singlet, t=triplet), coupling constants (Hz), and integration. Data for ¹³C-NMR are reported in terms of chemical shift (δ ppm) and multiplicity. Enantiomeric excess determinations were performed by HPLC analysis using UV-detection (Shimadzu SCL-10Avp) on Chiralpak AD, n-heptane:iPrOH 90:10, 1.0 ml/min. UPLC-MS on protein samples was performed on a Acquity TQ Detector (ESITQD- MS) coupled to Waters Acquity Ultra Performance LC using a Acquity BEH C8 (1.7 μm 2.1 x 150 mm). Water (solvent A) and acetonitrile (solvent B) containing 0.1% formic acid by volume, were used as the mobile phase at a flow rate of 0.3 mL/min. Gradient: 90% A for 2 min, linear gradient to 50% A in 2 min, linear gradient to 20% A in 5 min, followed by 2 min at 5% A. Re-equilibration of the column with 2 min at 90% A. *E. coli* strain NEB5α-T1 Phage Resistant (Stratagene) was used for cloning and *E. coli* BL21(DE3)-C43-T1 Phage Resistant (Stratagene) strain was used for protein expression. Primers were synthesised by Eurofins MWG Operon (Ebesberg, Germany). Pfu Turbo polymerase was purchased from Agilent and DpnI was purchased from New England Biolabs. Plasmid isolation Kit was purchased from Roche. DNA sequencing was carried out by Eurofins Genomics. FPLC columns were purchased from GE Healthcare. Size exclusion chromatography was performed using AKTApurifier 10 apparatus. The concentration of the proteins was measured with Nanodrop 2000 (Thermo Scientific).

General procedure for the preparation of the Friedel-Crafts products as reference material

The procedure was adapted from literature.^{18,33} 50 mg of the corresponding α,β-unsaturated-2-acyl imidazole and 2 equivalents of the corresponding

indole, pre-dissolved in 10 mL acetonitrile, were added to 1 L of water containing 2.3 g sodium dodecyl sulfate (8 mM final concentration) and 15 mol% $\text{Cu}(\text{NO}_3)_2 \cdot 3\text{H}_2\text{O}$. The reaction was stirred at room temperature for 16 hours. 10 g of NaCl was added and the aqueous phase was extracted 3 times with 100 mL diethyl ether. Subsequently, the organic phase was washed with 100 mL brine and dried on Na_2SO_4 . The product was purified by column chromatography with a gradient heptane:ethyl acetate.

3a

3-(1H-indol-3-yl)-1-(1-methyl-1H-imidazol-2-yl)butan-1-one (3a).^{28,29} ^1H -NMR (400 MHz, Chloroform-*d*) δ

8.35 (s, 1H), 7.66 (d, $J = 7.9$ Hz, 1H), 7.30 (d, $J = 8.0$ Hz, 1H), 7.19 – 7.10 (m, 2H), 7.10 – 7.04 (m, 1H), 7.04 – 6.96 (m, 2H), 3.92 (d, $J = 1.2$ Hz, 3H), 3.84 (d, $J = 7.1$ Hz, 1H), 3.57 (dd, $J = 15.9, 6.3$ Hz, 1H), 3.46 (dd, $J = 15.9, 8.2$ Hz, 1H), 1.41 (dd, $J = 6.9, 1.3$ Hz, 3H).

3b

1-(1-methyl-1H-imidazol-2-yl)-3-(2-methyl-1H-indol-3-yl)butan-1-one (3b).^{28,34,35} ^1H -NMR (400 MHz, Chloroform-*d*) δ

7.68 – 7.60 (m, 2H), 7.24 – 7.20 (m, 1H), 7.09 (d, $J = 1.0$ Hz, 1H), 7.04 (td, $J = 6.9, 1.5$ Hz, 2H), 6.94 (s, 1H), 3.86 (s, 3H), 3.78 (m, $J = 7.2$ Hz, 1H), 3.60 (dd, $J = 7.4, 2.8$ Hz, 2H), 2.42 (s, 3H), 1.46 (d, $J = 7.1$ Hz, 3H).

3c

1-(1-methyl-1H-imidazol-2-yl)-3-(2-phenyl-1H-indol-3-yl)butan-1-one (3c). ^1H -NMR (400 MHz, Chloroform-*d*) δ

7.98 (s, 1H), 7.87 (d, $J = 7.9$ Hz, 1H), 7.62 – 7.57 (m, 3H), 7.44 (d, $J = 7.9$ Hz, 2H), 7.40 – 7.31 (m, 2H), 7.21 – 7.06 (m, 2H), 6.93 (s, 1H), 4.02 – 3.94 (m, 1H), 3.88 (s, 3H), 3.81 (dd, $J = 15.8, 8.5$ Hz, 1H), 3.58 (dd, $J = 15.8, 6.3$ Hz, 1H), 1.49 (d, $J = 7.0$ Hz, 3H). HRMS (ES⁺): Exact mass calculated for $\text{C}_{22}\text{H}_{21}\text{N}_3\text{O}$ $[\text{M}+\text{H}]^+$ 343.1684, found 342.1607.

3d

3-(5-methoxy-1H-indol-3-yl)-1-(1-methyl-1H-imidazol-2-yl)butan-1-one (3d).^{18,34,35} ^1H -NMR (400 MHz, Chloroform-*d*) δ

7.84 (s, 1H), 7.21 (d, $J = 8.8$ Hz, 1H), 7.16 (d, $J = 2.4$ Hz, 1H), 7.14 – 7.12 (m, 1H), 7.04 (d, $J = 2.4$ Hz, 1H), 6.99 (s, 1H), 6.82 (dd, $J = 8.8, 2.4$ Hz, 1H), 3.93 (s, 3H), 3.86 (s, 3H), 3.79 (m, $J = 7.0$ Hz, 1H), 3.62 – 3.32 (m, 2H), 1.41 (d, $J = 6.9$ Hz, 3H).

3-(5-chloro-1H-indol-3-yl)-1-(1-methyl-1H-imidazol-2-yl)butan-1-one (3e).^{28,34,35}

¹H NMR (400 MHz, Chloroform-*d*) δ 8.68 (s, 1H), 7.54 (d, J = 2.0 Hz, 1H), 7.22 – 7.14 (m, 2H), 7.07 (dd, J = 8.6, 2.0 Hz, 1H), 7.03 – 6.96 (m, 2H), 3.92 (s, 3H), 3.81 – 3.71 (m, 1H), 3.50 – 3.39 (m, 2H), 1.38 (d, J = 6.9 Hz, 3H).

3-(5-bromo-1H-indol-3-yl)-1-(1-methyl-1H-imidazol-2-yl)butan-1-one (3f).

¹H-NMR (400 MHz, Chloroform-*d*) δ 8.32 (s, 1H), 7.69 (d, J = 2.0 Hz, 1H), 7.21 (dd, J = 8.6, 1.8 Hz, 1H), 7.16 (d, J = 8.6 Hz, 2H), 7.04 – 6.98 (m, 2H), 3.92 (d, J = 1.3 Hz, 3H), 3.77 (m, J = 7.1 Hz, 1H), 3.44 (m, J = 6.2, 1.2 Hz, 2H), 1.39 (dd, J = 6.9, 1.2 Hz, 3H). ¹³C-NMR (101 MHz, Chloroform-*d*) δ 194.6, 145.8, 137.5, 131.5, 131.0, 129.6, 127.3, 124.4, 124.1, 123.8, 115.1, 115.0, 49.6, 38.8, 29.7, 24.2.

1-(1-methyl-1H-imidazol-2-yl)-3-(2-methyl-1H-indol-3-yl)-3-phenylpropan-1-one (3g).³⁵

¹H-NMR (400 MHz, Chloroform-*d*) δ 7.75 (s, 1H), 7.55 – 7.50 (m, 1H), 7.42 – 7.38 (m, 2H), 7.25 – 7.19 (m, 3H), 7.16 – 7.09 (m, 2H), 7.07 – 6.95 (m, 2H), 6.94 (d, J = 1.0 Hz, 1H), 5.11 (t, J = 7.6 Hz, 1H), 4.21 (dd, J = 16.6, 7.0 Hz, 1H), 4.04 – 3.95 (m, 1H), 3.79 (s, 3H), 2.40 (s, 3H).

4,4-dimethyl-1-(1-methyl-1H-imidazol-2-yl)-3-(2-methyl-1H-indol-3-yl)pentan-1-one (3h).

¹H-NMR (400 MHz, Chloroform-*d*) δ 7.76 – 7.68 (m, 2H), 7.21 – 7.17 (m, 1H), 7.08 – 7.06 (m, 1H), 7.02 – 6.97 (m, 2H), 6.87 (s, 1H), 3.98 (dd, J = 17.6, 9.1 Hz, 1H), 3.74 (s, 3H), 3.72 – 3.66 (m, 1H), 3.49 (dd, J = 9.1, 4.6 Hz, 1H), 2.41 (s, 3H), 1.02 (s, 9H).

General procedure for the preparation of the Friedel-Crafts substrate

(E)-1-(1-Methyl-1H-imidazole-2-yl)-but-2-en-1-one (**1a**) was synthesised according to a literature procedure.^{29,34} Starting from 861 mg of crotonic acid (10 mmol) **1a** was obtained in a 20% purified yield (300 mg).

$^1\text{H-NMR}$ (400 MHz, Chloroform-*d*) δ 7.42 (dd, J = 15.5, 1.7 Hz, 1H), 7.17 (d, J = 1.0 Hz, 1H), 7.13 (dd, J = 15.5, 6.9 Hz, 1H), 7.04 (d, J = 0.9 Hz, 1H), 4.04 (s, 3H), 1.99 (dd, J = 6.9, 1.7 Hz, 3H).

1b and **1c** used were available from previous work.

(E)-1-(1-methyl-1H-imidazol-2-yl)-3-phenylprop-2-en-1-one (1b).²⁹ $^1\text{H-NMR}$ (400 MHz, Chloroform-*d*) δ 7.94 (dd, J = 101.5, 16.0 Hz, 2H), 7.74 – 7.63 (m, 2H), 7.39 (dd, J = 4.9, 1.9 Hz, 3H), 7.21 (d, J = 1.0 Hz, 1H), 7.07 (s, 1H), 4.08 (s, 3H).

(E)-4,4-dimethyl-1-(1-methyl-1H-imidazol-2-yl)pent-2-en-1-one (1c).³⁶ $^1\text{H-NMR}$ (400 MHz, Chloroform-*d*) δ 7.29 – 7.18 (m, 1H), 7.11 – 6.97 (m, 2H), 6.95 (d, J = 3.7 Hz, 1H), 3.93 (s, 3H), 1.05 (s, 9H).

Water-addition product

3-hydroxy-4,4-dimethyl-1-(1-methyl-1H-imidazol-2-yl)pentan-1-one (4). $^1\text{H-NMR}$ (400 MHz, Chloroform-*d*) δ 7.16 – 7.13 (s, 1H), 7.04 (s, 1H), 4.01 (s, 3H), 3.73 (dd, J = 9.4, 2.2 Hz, 1H), 3.26 – 3.07 (m, 2H), 0.98 (s, 9H).

UV-Vis titrations

UV-visible titrations were performed with 4 μM solution of the dimeric proteins titrated with 400 μM solution of $\text{Cu}(\text{NO}_3)_2$ in 20 mM MOPS, pH 7.0 containing 500 μM NaCl. The experiments with $\text{Cu}(\text{phen})(\text{NO}_3)_2$ were performed with 80 μM solution of $\text{Cu}(\text{phen})(\text{NO}_3)_2$ titrated with 400 μM solution of the dimeric proteins.

Figure 3. UV-visible titrations.

QacR (C72A, C141S) 4 μM : x equivalents $\text{Cu}(\text{NO}_3)_2$

80 μM $\text{Cu}(\text{phen})(\text{NO}_3)_2$: x equivalents QacR (C72A, C141S)

RamR 4 μM : x equivalents $\text{Cu}(\text{NO}_3)_2$

80 μM Cu(phen)(NO₃)₂ : x equivalents RamR

CgmR 4 μM : x equivalents Cu(NO₃)₂

CgmR : Cu(phen)(NO₃)₂

Catalysis

Catalytic reactions were performed in 150 μL total volume containing 90 μM $\text{Cu}(\text{NO}_3)_2$ (9 mol%) and 120 μM proteins (monomer, 1.3 equivalents), 1 mM of substrate **1** and **2** in 20 mM MOPS buffer pH 7.0 (or 20 mM MES buffer pH 5.0), 500 mM NaCl. In a typical procedure, the protein and the metal solutions were incubated for an hour under continuous inversion at 4 $^\circ\text{C}$ prior to addition of the substrates. Reactions were then incubated under continuous inversion at 4 $^\circ\text{C}$ for 72 h after which 50 μL of a 1 mM solution of 2-phenylquinoline in 20 mM MOPS buffer pH 7.0, 500 mM NaCl, 20% CH_3CN were added. Reactions were extracted 3 times with 600 μL diethylether and the organic layers were dried over Na_2SO_4 and evaporated *in vacuo*. The resulting products were re-dissolved in 150 μL heptane:isopropanol 9:1 and analysed by chiral HPLC (Chiralpak AD, 90_10 heptane:isopropanol). Yields of all the catalytic reactions are based on peak areas at 275 nm using 2-phenylquinoline as internal standard.

Inhibition experiments

Typical procedure

4

The corresponding catalysts were incubated with the indicated amount of dye per equivalent of protein monomer (60 μM = 1 equivalent of dye). The reactions were performed with the same conditions as above mentioned. The results are average of two independent experiments each carried out in duplicate. ^aThe reactions were performed in duplicate.

Table 3. Inhibition experiments for catalytic vinylogous Friedel–Crafts alkylation reactions.

Entry	Catalyst	Drug/equivalents	Yield (%)	ee (%)
1	$\text{Cu}(\text{NO}_3)_2$	-	22 ± 7	-
2 ^a	$\text{Cu}(\text{NO}_3)_2$	Ethidium/1	13 ± 0	-

3 ^a	Cu(NO ₃) ₂	Ethidium/4	13 ± 1	-
4 ^a	Cu(NO ₃) ₂	Ethidium/8	12 ± 0	-
5 ^a	Cu(NO ₃) ₂	R6G/1	12 ± 1	-
6 ^a	Cu(NO ₃) ₂	R6G/4	10 ± 1	-
7 ^a	Cu(NO ₃) ₂	R6G/8	12 ± 1	-
8	Cu ²⁺ ⊂QacR	-	78 ± 11	34 ± 3 (-)
9	Cu ²⁺ ⊂QacR	Ethidium/1	60 ± 7	25 ± 1 (-)
10	Cu ²⁺ ⊂QacR	Ethidium/4	36 ± 9	5 ± 3 (+)
11	Cu ²⁺ ⊂QacR	Ethidium/8	27 ± 10	9 ± 2 (+)
12	Cu ²⁺ ⊂QacR	R6G/1	76 ± 9	32 ± 1 (-)
13	Cu ²⁺ ⊂QacR	R6G/4	44 ± 4	15 ± 4 (-)
14	Cu ²⁺ ⊂QacR	R6G/8	33 ± 17	7 ± 9 (-)
15	Cu ²⁺ ⊂RamR	-	57 ± 9	29 ± 3 (-)
16	Cu ²⁺ ⊂RamR	Ethidium/1	32 ± 8	32 ± 4 (-)
17	Cu ²⁺ ⊂RamR	Ethidium/4	26 ± 6	31 ± 0 (-)
18	Cu ²⁺ ⊂RamR	Ethidium/8	17 ± 3	29 ± 2 (-)
19	Cu ²⁺ ⊂CgmR	-	52 ± 12	13 ± 3 (-)
20	Cu ²⁺ ⊂CgmR	Ethidium/1	39 ± 11	3 ± 2 (-)
21	Cu ²⁺ ⊂CgmR	Ethidium/4	25 ± 9	6 ± 1 (+)
22	Cu ²⁺ ⊂CgmR	Ethidium/8	18 ± 4	11 ± 1 (+)

Control Friedel-Crafts reaction with dialysed Cu²⁺⊂protein

Typical procedure

The proteins were incubated with 1.1 excess of Cu(NO₃)₂ for 1 h. Subsequently, the 1 mL samples were dialysed against 1 L of 20 mM MOPS, 500 mM NaCl, pH 7.0 (reaction buffer) overnight and the recovered

samples were then used for catalysis as previously described. The reactions were performed in duplicates.

Table 4. Results of the vinylogous Friedel–Crafts alkylation reaction with dialysed Cu^{2+} -protein.

Entry	Protein	Dialysis	Yield (%)	ee (%)
1	Cu^{2+} -QacR	NO	78 ± 11	34 ± 3
2	Cu^{2+} -QacR	YES	29 ± 1	48 ± 0
3	Cu^{2+} -RamR	NO	57 ± 9	29 ± 3
4	Cu^{2+} -RamR	YES	13 ± 2	43 ± 4
5	Cu^{2+} -CgmR	NO	52 ± 12	13 ± 3
6	Cu^{2+} -CgmR	YES	24 ± 8	15 ± 2

4

Catalysis with cell-free extracts

5 mL of LB media containing ampicillin (AMP) were inoculated from glycerol stocks of the targeted proteins and grown overnight at 37 °C. 50 μL of the densely grown overnight cultures were added to 2 mL sterile Eppendorf tubes containing 1.150 μL of fresh LB media and Amp. Bacteria were cultured at 37 °C for 4.5 hours while shaking at 950 rpm. Subsequently, protein production was induced by addition of 50 μL LB media, containing IPTG (1.2 μL of a 1 M stock solution). Plates were then incubated at 30 °C for 16 hours while shaking (950 rpm) and harvested by centrifugation (3,500 rpm at 4 °C for 15 minutes). After removing the supernatant, cells were washed by addition of 500 μL of reaction buffer (20 mM MOPS, 500 mM NaCl, pH = 7.0) and subsequently harvested by centrifugation (3,500 rpm for 10 minutes). The supernatant was discarded and bacteria were resuspended in 300 μL reaction buffer, containing protease inhibitor (*Roche cOmplete*), lysozyme (1 mg/mL), DNase I (0.1 mg/mL) and MgSO_4 (10 mM) to assist cell lysis and prevent protein

degradation. Resuspended cells were incubated for 2 hours at 30 °C at 800 rpm and then stored overnight at -20 °C. The lysates were defrosted and cell debris was removed by centrifugation (4,000 rpm, 1 hour, 4 °C) and 135 μ L of the cleared lysate was transferred into new 1.5 mL tubes for catalysis.

5 μ L of substrate **1a** (final concentration 1 mM) and 5 μ L of Cu(NO₃)₂ (final concentration 90 μ M) were added to the lysates and incubated for 30 minutes at 4 °C under continuous inversion. Consecutively, **2a** (1 mM) was added and the reaction mixtures were left reacting for three days.

Table 5. Results of the vinylogous Friedel-Crafts alkylation reaction with cell-free extracts.

Entry	CFE	Yield (%)	ee (%)
1	CFE SUMO	4 $\pm$ 2	n.d.
2	CFE LmrR	4 $\pm$ 1	n.d.
3	CFE CgmR	3 $\pm$ 1	n.d.
4	CFE RamR	4 $\pm$ 2	n.d.
5	CFE QacR	3 $\pm$ 1	n.d.
6	CFE SUMO + Cu(NO ₃) ₂	53 $\pm$ 24	1 $\pm$ 7
7	CFE LmrR + Cu(NO ₃) ₂	68 $\pm$ 12	5 $\pm$ 2
8	CFE CgmR + Cu(NO ₃) ₂	58 $\pm$ 16	2 $\pm$ 1
9	CFE RamR + Cu(NO ₃) ₂	39 $\pm$ 19	-5 $\pm$ 5
10	CFE QacR + Cu(NO ₃) ₂	61 $\pm$ 14	20 $\pm$ 6

The results are average of at least two independent experiments, each carried out in duplicate. For yields < 5% the ee was not determined.

Chiral HPLC traces

a)

b)

Figure 4. a) Typical chiral HPLC traces obtained for biocatalytic reactions with product **3h**. b) HPLC traces obtained from an independently prepared sample of the water addition product to **1c**.

Figure 5. Typical chiral HPLC traces obtained for biocatalytic reactions with product **3a**.^{30,31}

4.5. REFERENCES

- 1 P. J. Deuss, R. den Heeten, W. Laan and P. C. Kamer, *Chemistry (Easton)*, 2011, **17**, 4680–4698.
- 2 J. Bos and G. Roelfes, *Curr. Opin. Chem. Biol.*, 2014, **19**, 135–143.
- 3 Y. Okamoto, V. Köhler, V. Lebrun, R. Reuter, T. R. Ward, F. Schwizer, T. Heinisch, M. M. Pellizzoni, Y. Gu and J. C. Lewis, *Chem. Rev.*, 2017, **118**, 142–231.
- 4 T. Heinisch and T. R. Ward, *Acc. Chem. Res.*, 2016, **49**, 1711–1721.
- 5 G. M. Wilson, M. E.; Whitesides, *J. Am. Chem. Soc.*, 1978, **100**, 306–307.
- 6 K. Oohra, Y. Kihir, E. Mizohata, T. Inoue and T. Hayashi, *J. Am. Chem. Soc.*, 2013, **135**, 17282–17285.
- 7 M. T. Reetz and N. Jiao, *Angew. Chem. Int. Ed.*, 2006, **45**, 2416–2419.
- 8 M. Ohashi, T. Koshiyama, T. Ueno, M. Yanase, H. Fujii and Y. Watanabe, *Angew. Chem. Int. Ed.*, 2003, **42**, 1005–1008.
- 9 J. C. Lewis, *ACS Catal.*, 2013, **3**, 2954–2975.
- 10 W. Ghattas, V. Dubosclard, A. Wick, A. Bendelac, R. Guillot, R. Ricoux and J. P. Mahy, *J. Am. Chem. Soc.*, 2018, **140**, 8756–8762.
- 11 G. Sreenilayam, E. J. Moore, V. Steck and R. Fasan, *Adv. Synth. Catal.*, 2017, **359**, 2076–2089.

- 12 T. R. Ward, *Acc. Chem. Res.*, 2011, **44**, 47–57.
- 13 J. M. Zimbron, T. Heinisch, M. Schmid, D. Hamels, E. S. Nogueira, T. Schirmer and T. R. Ward, *J. Am. Chem. Soc.*, 2013, **135**, 5384–5388.
- 14 O. F. Brandenburg, R. Fasan and F. H. Arnold, *Curr. Opin. Biotechnol.*, 2017, **47**, 102–111.
- 15 K. Takeuchi, Y. Tokunaga, M. Imai, H. Takahashi and I. Shimada, *Sci. Rep.*, 2014, **4**, 6922.
- 16 P. K. Madoori, H. Agustindari, A. J. M. Driessen and A.-M. W. H. Thunnissen, *EMBO J.*, 2009, **28**, 156–166.
- 17 G. Roelfes, *Acc. Chem. Res.*, 2019, **52**, 545–556.
- 18 J. Bos, W. R. Browne, A. J. M. Driessen and G. Roelfes, *J. Am. Chem. Soc.*, 2015, **137**, 9796–9799.
- 19 I. T. Paulsen, *Curr. Opin. Microbiol.*, 2003, **6**, 446–451.
- 20 J. L. Ramos, M. Martinez-Bueno, A. J. Molina-Henares, W. Teran, K. Watanabe, X. Zhang, M. T. Gallegos, R. Brennan and R. Tobes, *Microbiol. Mol. Biol. Rev.*, 2005, **69**, 326–356.
- 21 H. Itou, N. Watanabe, M. Yao, Y. Shirakihara and I. Tanaka, *J. Mol. Biol.*, 2010, **403**, 174–84.
- 22 S. Yamasaki, R. Nakashima, K. Sakurai, S. Baucheron, E. Giraud, B. Doublet, A. Cloeckert and K. Nishino, *Sci. Rep.*, 2019, **9**, 1–8.
- 23 D. Fujiwara, S. Yamasaki, K. Sakurai, K. Nishino, I. Fujii, R. Nakashima and E. Nikaido, *Nat. Commun.*, 2013, **4**, 1–7.
- 24 S. Grkovic, K. M. Hardie, M. H. Brown and R. A. Skurray, *Biochemistry*, 2003, **42**, 15226–36.
- 25 K. M. Peters, B. E. Brooks, M. A. Schumacher, R. A. Skurray, G. Richard, M. H. Brown, R. G. Brennan and M. H. Brown, *PLoS One*, 2011, **6**, e15974.
- 26 M. A. Schumacher, M. C. Miller, S. Grkovic, M. H. Brown, R. A. Skurray and R. G. Brennan, *Science*, 2001, **294**, 2158–2164.
- 27 K. M. Peters, J. T. Schuman, R. A. Skurray, M. H. Brown, R. G. Brennan and M. A. Schumacher, *Biochemistry*, 2008, **47**, 8122–9.
- 28 M. Bersellini and G. Roelfes, *Org. Biomol. Chem.*, 2017, **15**, 3069–3073.
- 29 D. A. Evans, K. R. Fandrick and H. J. Song, *J. Am. Chem. Soc.*, 2005, **127**,

8942–8943.

- 30 A. J. Boersma, D. Coquire, D. Geerdink, F. Rosati, B. L. Feringa and G. Roelfes, *Nat. Chem.*, 2010, **2**, 991–995.
- 31 J. Bos, A. Garca-Herraiz and G. Roelfes, *Chem. Sci.*, 2013, **4**, 3578–3582.
- 32 S. Kato, A. Onoda, T. Hayashi, N. Taniguchi, T. Himiyama, N. Taniguchi, S. Kato, A. Onoda and T. Hayashi, *Angew. Chem. Int. Ed.*, 2017, **56**, 13618–13622.
- 33 V. P. A. and I. P. Beletskaya, *Top Organomet Chem* 43 1–20, 2009, **48**, 11533–11542.
- 34 N. Duchemin, E. Benedetti, L. Bethge, S. Vonhoff, S. Klussmann, J. J. Vasseur, J. Cossy, M. Smietana and S. Arseniyadis, *Chem. Commun.*, 2016, **52**, 8604–8607.
- 35 A. J. Boersma, B. L. Feringa and G. Roelfes, *Angew. Chem. Int. Ed.*, 2009, **48**, 3346–3348.
- 36 J. H. Yum, S. Park, R. Hiraga, I. Okamura, S. Notsu and H. Sugiyama, *Org. Biomol. Chem.*, 2019, **17**, 2548–2553.

CHAPTER 5

Unexpected Catalytic Activity of the Regulatory Protein QacR

Abstract

Natural proteins often present certain degree of binding or functional promiscuity. In biocatalysis, this promiscuity has been exploited for accessing new-to-nature reactions. Here, we report an unexpected catalytic activity for the regulatory protein QacR from the TetR family of Multidrug Resistance Regulators. QacR_C72A_C141S is able to catalyse the tandem Friedel-Crafts/enantioselective protonation of indoles with α -substituted conjugated enones with 38% yield and up to 83% ee. Mutagenesis and computational studies support the hypothesis that an acidic residue in the binding pocket of the protein is responsible for protonating the reaction intermediate.

Preprint published:

C. Gutiérrez de Souza, L. Alonso-Cotchico, M. Bersellini and G. Roelfes, *ChemRxiv*, Prepr.

5.1. INTRODUCTION

In the quest for biocatalysis of abiotic reactions, an attractive approach involves taking advantage of enzyme catalytic promiscuity.¹ Basal levels of activity different than the native one confer enzymes the ability to adapt to environmental changes.² In enzyme engineering, this characteristic is exploited to create enzymes for new reactions by applying rational, computational or directed evolution techniques.^{3,4} A common approach is repurposing existing proteins that exhibit some level of promiscuity in substrate binding or activity to enable new-to-nature reactions. Such an approach is well illustrated by heme-dependent enzymes like P450s.^{5,6} The use of xenobiotic substrates allowed these proteins to access a wide range of unnatural reactions with high turnover numbers.⁵ Basal levels of catalytic activity for abiological transformations can sometimes also be observed in proteins which natural function is not catalysis. For example, heme-containing proteins like myoglobin, have been shown to catalyse non-natural reactions.^{7,8} Similarly, bovine serum albumin (BSA) does not have any natural catalytic activity, however, the presence of a lysine residue, originally responsible for ligand binding, grants it low levels of catalytic activity for a range of abiotic reactions.⁹ Alternatively, pre-existing binding sites of proteins have been repurposed for binding external abiological catalytic cofactors and thus introduce different reactivities into their scaffolds.¹⁰ Here, we present a regulatory protein from the TetR family, QacR, that showed unexpected catalytic activity for the tandem Friedel-Crafts/enantioselective protonation reaction of an indole with an α -substituted α,β -unsaturated thiazole.

Multidrug resistance regulators (MDRs) are a type of proteins involved in regulating efflux pump expression for antibiotic resistance. Their natural function grants these proteins the ability to bind structurally different aromatic planar drugs. In previous work, our group investigated the potential of the multispecificity of this class of proteins as a means for introducing catalytic activity into their scaffolds.¹¹⁻¹³ A series of artificial metalloenzymes (ArMs), hybrid systems in which a metal complex was embedded into a biological scaffold to introduce

abiotic reactivities, were created. Among the MDRs, a large amount of work has focused on the Lactococcal multidrug resistance regulator (LmrR) protein from the PadR family.^{14,15} Recently, the repertoire of MDR scaffolds for ArMs was expanded to three proteins from the TetR family: CgmR, RamR and QacR.¹⁶ These three proteins proved their catalytic potential in the Friedel-Crafts (F-C) alkylation reaction of indoles with β -substituted α,β -unsaturated 2-acyl-imidazoles when introducing the metal-binding unnatural amino acid (2,2'-bipyridin-5-yl)alanine (bpyA) into their scaffolds¹³ or simply by using $\text{Cu}(\text{NO}_3)_2$ and an aromatic substrate without the need for a ligand to anchor the metal into the protein scaffolds.¹⁷

This inspired us to further explore the catalytic promiscuity of CgmR, RamR and QacR in the tandem Friedel-Crafts/enantioselective protonation (F-C/EP) of indoles with α -substituted conjugated enones. ArMs based on DNA¹⁸ and LmrR¹⁹ were reported previously for this reaction, which encouraged the investigation of the F-C/EP with these proteins and Cu^{2+} ions. In this reaction, the stereochemical centre is not created in the conjugate addition step, but in the subsequent protonation (Scheme 1). This presents several challenges which are related to the small size of protons and their high mobility, especially in aqueous media, which makes them difficult to control and makes it hard to achieve control over the protonation step.²⁰⁻²²

a) Friedel-Crafts alkylation, β -substituted enone.

b) Friedel-Crafts/enantioselective protonation, α -substituted enone.

Scheme 1. Reaction scheme of Friedel-Crafts alkylation and tandem Friedel-Crafts/enantioselective protonation of an indole.

5.2. RESULTS AND DISCUSSION

The ArMs were self-assembled by incubation of QacR, CgmR or RamR with $\text{Cu}(\text{NO}_3)_2$. A double mutant of QacR (C72A and C141S) was used to avoid disulphide bonds and simplify the protein production protocol.²³ The ArMs were then tested in the conjugate addition of 2-methyl-1H-indole (**2**) to 2-methyl-1-(thiazol-2-yl)prop-2-en-1-one (**1**) (Table 1). Initially, the reactions were performed at pH 7.0. Experiments without catalysts, or with only $\text{Cu}(\text{NO}_3)_2$ did not give rise to product formation (entries 1, 2, 12, 13).

Interestingly, while in previous work describing DNA-catalysed F-C/EP, Cu^{2+} was strictly required for the reaction to proceed,¹⁸ CgmR, RamR and QacR in absence of $\text{Cu}(\text{NO}_3)_2$ led to low reaction yields and low or no enantioselectivity. The combination of protein with $\text{Cu}(\text{NO}_3)_2$ resulted in increased yields of product **3**, with a moderate ee of 54 % in case of RamR (entry 3) and 42% ee, for the opposite enantiomer, in case of QacR (entry 6). No significant difference was observed for CgmR in presence of $\text{Cu}(\text{NO}_3)_2$, which maintained a yield of around 13% and no enantioselectivity. These results highlight the importance of the combined effect of Cu^{2+} , acting as Lewis acid and thus accelerating the reaction, and the protein, needed to bring the substrates together and provide a chiral environment. Decreasing the pH to 5.0 led to an increase of the enantioselectivity to 77% in case of Cu^{2+} -QacR (entry 14). CgmR and RamR partially precipitated at this pH and this led to irreproducibility of the catalysis results.

Surprisingly, at pH 5.0 the reaction in presence of QacR without $\text{Cu}(\text{NO}_3)_2$ led to similar ee of 75% and an increased yield (38%). To ensure that no other metal ions were present in the reaction mixture, which could be responsible for the catalytic activity observed, the reactions were repeated after incubation of the proteins with 50 mM EDTA and double dialysis against 1:1000 20 mM MES, 500 mM NaCl, pH 5.0. No loss of activity was observed for any of the proteins, and a

further small increase in enantioselectivity to 83% for QacR at pH 5.0 was obtained (entries 5, 8, 11 and 16).

Table 1. Catalysis results for the Friedel-Crafts/enantioselective protonation reaction.

Entry	Catalyst	pH	Yield (%)	ee (%)
1	-	7.0	<5	n.d.
2	Cu ²⁺	7.0	<5	n.d.
3	Cu ²⁺ +RamR	7.0	24 ± 4	54 ± 5
4	RamR	7.0	15 ± 5	-13 ± 10
5 ^a	RamR	7.0	12 ± 8	4 ± 9
6	Cu ²⁺ +QacR	7.0	31 ± 5	-42 ± 8
7	QacR	7.0	13 ± 9	9 ± 3
8 ^a	QacR	7.0	15 ± 7	10 ± 3
9	Cu ²⁺ +CgmR	7.0	15 ± 1	-4 ± 2
10	CgmR	7.0	13 ± 6	-1 ± 4
11 ^a	CgmR	7.0	13 ± 10	-5 ± 9
12	-	5.0	<5	n.d.
13	Cu ²⁺	5.0	<5	n.d.
14	Cu ²⁺ +QacR	5.0	26 ± 3	-77 ± 4
15	QacR	5.0	38 ± 9	-75 ± 4
16 ^{a,b}	QacR	5.0	40 ± 6	-83 ± 0

Typical reaction conditions: 120 μM of QacR (dimer), 90 μM of Cu(NO₃)₂ and 1 mM of substrates **1** and **2** in 20 mM MES buffer pH 5.0 or 20 mM MOPS buffer for pH 7.0, 500 mM NaCl, at 4 °C for 72 h. The results are average of two independent experiments each of them performed in duplicate. Sign of ee

indicates the order of elution on chiral HPLC with + corresponding to the first-eluting enantiomer in excess. ^aProtein previously incubated with EDTA and subsequently submitted to dialysis against the reaction buffer. ^bAverage of two independent experiments.

The fact that QacR by itself, without any added metal ions, showed an inherent activity for catalysis of the F-C/EP reaction is a striking finding. In nature, enzymes such as decarboxylases and esterases use interactions of a combination of amino acid residues to both activate a prochiral proton acceptor and act as proton donor to effectively perform enantioselective protonations.²¹ We hypothesised that, analogous to natural enzymes, an acidic amino acid residue of the protein could be responsible for effectively protonating the prochiral C α of the reaction intermediate following a Brønsted acid catalysed mechanism.

Figure 1. Effect of pH in QacR and Cu²⁺-QacR-based artificial enzymes for the Friedel-Crafts/enantioselective protonation reaction. Typical conditions: 120 μ M QacR, 90 μ M Cu(NO₃)₂ (when applicable), 1 mM **1** and **2**, in 20 mM MES buffer for pH 5.0 and 6.0 or 20 mM MOPS buffer for pH 7.0 and 8.0, containing 500 mM NaCl, at 4 °C for 72 h. The results are average of at least two independent experiments, each of them performed in duplicate.

To gain further insights into the mechanism, the effect of the pH in the reaction was investigated. The F-C/EP reaction was performed in a range of pH's, with QacR in presence and absence of Cu(NO₃)₂. The data obtained showed a correlation between pH and both activity and selectivity in the reaction performed with QacR alone. A decrease of pH

translated into an increase of yield and ee, with pH 5.0 being optimal (Figure 1). This result was attributed to the relation between the protonation state of the proton donor amino acid and the proton transfer efficiency to the enolate intermediate. Moreover, the pronounced increase in ee at pH 5.0 indicates that at this pH the enantioselective protonation happens preferably *via* one of the two prochiral faces of the enolate intermediate. On the other hand, the effect of pH on the Cu^{2+} -QacR system was less clear. No significant trend in activity was observed in the pH range investigated. Higher enantioselectivities were observed at acidic pHs, but the ee was still significant at neutral or slightly basic pH (Figure 1). Notably, at pH > 5 the ee was generally higher for the reaction catalysed by Cu^{2+} -QacR compared to QacR alone. The different effect of pH in the reaction with QacR in presence or absence of $\text{Cu}(\text{NO}_3)_2$, suggested that the reaction can follow two pathways, which are either catalysed by the protein alone, or by Cu^{2+} embedded in the protein, as reported before for the Friedel-Crafts reaction.¹⁷

The crystal structure (PDB 1JTY) was inspected in order to identify residues that could be playing a role in donating a proton to the C α of the enolate intermediate. Initially, three potential proton donors (Tyr103, Tyr123 and Glu120) were identified in the binding pocket of QacR (Figure 2) and their pKa's in the environment of the protein were

Figure 2. X-ray structure of QacR (PDB 1JTY) with the docked intermediate of the reaction (yellow) and possible proton-donors highlighted.

calculated using the PROPKA algorithm, included in the Chimera software.

Tyrosines 103 and 123, with $pK_a = 16.0$ and 14.2 respectively, were discarded as possible proton donors, since their pK_a 's were too high to explain a difference in the protonation state between pH 5 and 7. In contrast, the calculated pK_a of Glu120 was approximated as 7.8 , meaning that this residue would be fully protonated at pH 5.0, whereas at pH 7.0 it would only be partially protonated. Thus, Glu120 could potentially act as Brønsted acid at lower pH explaining the higher activities and selectivities observed at pH 5 and 6 compared to $pH \geq 7$.

To investigate this hypothesis further, we mutated glutamic acid 120 to glutamine (E120Q), to maintain a similar size of the side chain and remove the acidic character. The mutant QacR_E120Q showed a similar yield (24%) compared to QacR, but a complete loss of ee (Scheme 2).

Scheme 2. a) Schematic representation of the enantioselective protonation of the enolate intermediate by the acidic proton of Glu120. The intermediate is depicted stabilised by hydrophobic interactions with residues of the protein scaffold. b) Results of the F-C/EP catalysis with QacR and QacR_E120Q highlighting the effect of the acidic residue.

This finding supports the hypothesis of Glu120 acting as Brønsted acid for the enantioselective protonation step at pH 5.0 as well as that it

suggests that the reaction is taking place inside the hydrophobic pocket of QacR. In addition, the presence of $\text{Cu}(\text{NO}_3)_2$ in the reaction with QacR_E120Q showed a comparable yield (26%) but an increase of the enantiomeric excess (22%) compared to that of QacR_E120Q without Cu^{2+} . These results suggested that Cu^{2+} bound to the thiazole substrate **1** could be driving the reaction to happen within the protein scaffold (and hence in an enantioselective manner), as shown in our previous work with Cu^{2+} -QacR for the Friedel-Crafts reaction.¹⁷

To gain further insight into the plausibility of a Brønsted acid-accelerated mechanism involving proton transfer by Glu120, computational studies were performed. The geometry of the enolate intermediate was optimised by quantum mechanics calculation (QM) and docked into QacR's active site (PDB 1JTY). The resulting structure was submitted to 100 ns MD simulation per triplicate and the interactions between the enolate intermediate and the active site were assessed. The intermediate appeared predominantly stabilised by hydrophobic interactions with aromatic active-site residues from both monomers A (F162) and B (Y103, I112, L119, Y123) (Figure 3b,c). These interactions were consistent with those found for the protein bound to ethidium bromide (ETB), which has been reported to bind within the pocket of QacR with a binding constant of $2.35 \mu\text{M}$ ²⁴ (Figure 3a).

Figure 3. Docking of the FC/EP intermediate (left) and the transition state structure of the first step of the reaction TSI (centre).

A clustering analysis of the MD simulation revealed three major binding modes of the intermediate into the active site, grouped into clusters c0,

c1 and c2. Cluster c1 showed structures in which the intermediate was displaced along the tunnel of the QacR active site; this would not permit catalysis to occur. Therefore, c1 was not taken into consideration. While for cluster c0 (33% of the simulation snapshots), structures that could allow catalysis according to our hypothesis were not found (Figure 3b), the binding modes in c2 (15% of the MD simulation snapshots) showed orientations of the intermediate in which C α could be protonated by Glu120 (Figure 3c). To assess whether the binding mode of this intermediate was also suitable for the conjugate addition step (TS1) of the reaction, TS1 was also docked into the active site. The data obtained suggested that the geometries of both TS1 and the enolate intermediate can fit in the active site in a similar arrangement that would allow both steps TS1 (conjugate addition of the indole) and TS2 (protonation by Glu120) of the reaction mechanism to occur without significant structural rearrangements of the intermediate (Figure 3b).

These results support the analysis from the experimental findings, suggesting Glu120 as a possible catalytic residue able to protonate the enolate intermediate of the F-C/EP reaction. The fact that pre-catalytic conformations were not highly represented during the MD simulation time scale, could explain why this system shows good enantioselectivity at pH 5.0, but only moderate yield.

5

5.3. CONCLUSIONS AND PERSPECTIVES

In conclusion, in this chapter we showed that QacR, a regulatory protein from the TetR family of multidrug resistance regulators with no natural catalytic function, exhibits unexpected catalytic activity for the Friedel-Crafts/enantioselective protonation reaction of 2-methyl indole with an α -substituted α,β -unsaturated 2-acyl thiazole achieving moderate yield (38%) and good enantiomeric excess, up to 83% ee, without requiring a metal cofactor. Mutagenesis and computational studies suggest the role of Glu120 in protonating the prochiral reaction intermediate and, thus, acting as Brønsted acid. The data gathered from the computational analysis showed that the conformation that allows Glu120 to be able to

protonate the enolate intermediate of the reaction is not the most represented during the MD simulation. This fact could explain why QacR system shows high enantioselectivity for this reaction at pH 5.0 but only moderate yield. Furthermore, knowledge of the beneficial structure for catalysis, offers a possibility of improving the catalytic proficiency of this artificial enzyme by, for instance, creating interactions to stabilise the structures showed in cluster2.

Moreover, CgmR, RamR and QacR proteins at pH 7.0 are able to catalyse the Friedel-Crafts/enantioselective protonation reaction using Cu^{2+} ions as Lewis acid, without the need of a metal ligand, leading to low yields and different enantiomeric outcomes (RamR and QacR to opposite enantiomers, and CgmR to a racemic mixture). These results highlight the versatility of the scaffolds and indicate that these three proteins are a well-suited small selection of scaffolds for ArM design. Overall, these findings provide a new example of the potential of exploiting binding multispecificity of natural proteins as a source of new-to-nature reactivities.

5.4. EXPERIMENTAL SECTION

Computational methods

The computational analysis was performed by Dr. Lur Alonso Cotchico, see the manuscript for details.²⁵

Materials and methods

All the chemicals and reagents were purchased from commercial suppliers (Sigma Aldrich or Acros) and used without further purification. Flash chromatography was performed using Reveleris® X2 Flash Chromatography System. ^1H -NMR and ^{13}C -NMR spectra were recorded on a Varian 400 (400 MHz) spectrometer in CDCl_3 . Chemical shifts values (δ) are denoted in ppm using residual solvent peaks as the internal standard (CHCl_3 : δ 7.26 for ^1H). ^1H -NMR results are reported in the conventional form: chemical shifts, multiplicity (s = singlet, d = doublet, dd = doublet of doublet, m= multiplet, q = quartet, s = singlet, t = triplet), coupling constants (Hz), and integration. Enantiomeric excess determinations were performed by HPLC analysis using

UV-detection (Shimadzu SCL-10Avp) on Chiralpak AD, n-heptane:iPrOH 90:10, 1.0 ml/min. UPLC-MS on protein samples was performed on a Acquity TQ Detector (ESITQD- MS) coupled to Waters Acquity Ultra Performance LC using a Acquity BEH C8 (1.7 μ m 2.1 x 150 mm). Water (solvent A) and acetonitrile (solvent B) containing 0.1% formic acid by volume, were used as the mobile phase at a flow rate of 0.3 mL/min. Gradient: 90% A for 2 min, linear gradient to 50% A in 2 min, linear gradient to 20% A in 5 min, followed by 2 min at 5% A. Re-equilibration of the column for 2 min at 90% A. *E. coli* strain NEB5 α -T1 Phage Resistant (Stratagene) was used for cloning and *E. coli* BL21C43(DE3)-T1 Phage Resistant (Stratagene) strain was used for protein expression. Primers were synthesised by Eurofins MWG Operon (Ebesberg, Germany). Pfu Turbo polymerase was purchased from Agilent and DpnI was purchased from New England Biolabs. Plasmid isolation Kit was purchased from Roche. DNA sequencing was carried out by Eurofins Genomics. FPLC columns were purchased from GE Healthcare. Size exclusion chromatography was performed using AKTApurifier 10 apparatus. The concentration of the proteins was measured with Nanodrop 2000 (Thermo Scientific).

General procedure for the preparation of the Friedel-Crafts/enantioselective protonation thiazole substrate

2-methyl-1-(thiazol-2-yl)prop-2-en-1-one was synthesised in two steps according to literature procedures.¹⁸

5

2-methyl-1-(thiazol-2-yl)prop-2-en-1-ol

2-(Trimethylsilyl)thiazole (1664 μ L, 10 mmol) and methacrolein (868 μ L, 10 mmol) were stirred for 4 h at room temperature.

The reaction mixture was diluted with THF (200 mL) and treated with

tetrabutylammonium fluoride solution 1.0 M in THF (10 mL, 10 mmol). After additional stirring for 1 h, the solvent was removed under vacuum. The crude residue was dissolved in 100 mL ethyl acetate and washed with saturated NaHCO₃. The organic layer was dried over MgSO₄ and the solvent removed under vacuum. The residue was purified by column chromatography with a gradient ether:pentane to give product the desired product as slightly yellow oil (503 mg, 37%). ¹H-NMR (400 MHz, Chloroform-*d*) δ 7.61 (d, *J* = 3.3 Hz, 1H), 7.24 (d, *J* = 3.3 Hz, 1H), 5.39 (s, 1H), 5.17 (s, 1H), 4.96 (t, *J* = 1.6 Hz, 1H), 4.87 (s, 1H), 1.65 (d, *J* = 1.5 Hz, 3H).

2-methyl-1-(thiazol-2-yl)prop-2-en-1-one (**1**)

10 mg of 2-methyl-1-(thiazol-2-yl)prop-2-en-1-ol (0.64 mmol) were dissolved in 1 mL dichloromethane. MnO₄ (55 mg, 63 mmol) was added to the solution and the reaction mixture was stirred for 45 min at room temperature. The crude product was filtered over celite and the solvent was removed under vacuum. 7.1 mg of product (72 % yield) were obtained as a colorless oil. Product **1** was used directly for catalysis; upon standing, **1** undergoes a hetero Diels-Alder reaction with itself.¹⁸ ¹H-NMR (400 MHz, Chloroform-*d*) δ 7.99 (dd, *J* = 2.8, 1.5 Hz, 1H), 7.64 (dd, *J* = 2.8, 1.4 Hz, 1H), 6.93 (s, 1H), 6.17 – 6.12 (m, 1H), 2.09 (d, *J* = 1.5 Hz, 3H).

Characterisation of the Friedel-Crafts/enantioselective protonation product as reference material

2-methyl-3-(2-methyl-1H-indol-3-yl)-1-(thiazol-2-yl)propan-1-one.

¹H-NMR (400 MHz, Chloroform-*d*) δ 8.00 (d, *J* = 3.0 Hz, 1H), 7.73 (s, 1H), 7.69 – 7.66 (m, 1H), 7.63 (d, *J* = 3.0 Hz, 1H), 7.25 – 7.21 (m, 1H), 7.12 – 7.06 (m, 2H), 4.28 – 4.15 (m, 1H), 3.30 (dd, *J* = 14.2, 5.7 Hz, 1H), 2.80 (dd, *J* = 14.2, 8.9 Hz, 1H), 2.41 (s, 3H), 1.24 (d, *J* = 6.9 Hz, 3H). NMR identical to that reported in the literature.¹⁸

Molecular biology

pET 17b plasmids encoding for wt-QacR, wt-CgmR and wt-RamR were available from a previous work.⁴ Plasmids were purchased from Genescript

(USA) as codon optimised sequences for *E. coli* expression and included a C-terminal Strep-tag for purification purposes.

Site-directed mutagenesis

Site directed mutagenesis for preparation of QacR mutant was performed on a previously optimised QacR variant without cysteines (C72A and C141S). The primers required for the mutagenesis are summarised in Table S1. Mutation at position E120Q was performed using partially overlapping primers, *Phu* polymerase and the following PCR cycles: 2' elongation, 35 cycles, 54°C annealing temperature. The resulting PCR product was digested with restriction endonuclease *DpnI* for 2-3 h at 37 °C and transformed into the chemically competent *E. coli* NEB5α cells. Cells were spread onto an agar plate containing 100 µg/mL of ampicillin. Single colonies were selected after overnight growth and used to inoculate 5 mL of LB medium containing the same antibiotic. Plasmid was isolated and successful mutagenesis was confirmed by DNA sequencing.

Table 2. Primers used for site directed mutagenesis (fw: forward primer, rv: reverse primer).

Primer	DNA sequence (5' → 3')
QacR_C72A_C141S_E120Q_fw	AACTGCAGAACAAAGTATATCGACGCGTACCAC
QacR_C72A_C141S_E120Q_rv	CTTGTTCTGCAGTTTGTTCATCTTCTCGTTGATG

5

Protein production, purification and characterisation

pETI7b plasmids encoding for wt-CgmR, wt-RamR, QacR_C72A_C141S and QacR_C72A_C141S_E120Q, were transformed into *E. coli* BL21C43(DE3) cells. The cells were spread onto an agar plate containing 100 µg/mL of ampicillin. 5 mL of LB media with µg/mL of ampicillin were inoculated with a single colony from the transformation and grown at 37 °C overnight. The entire

starter culture was used to inoculate 500 mL of fresh LB medium containing 100 µg/mL of ampicillin. When the culture reached an optical density at 600 nm between 0.8-0.9, isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM was added to induce the expression of the target proteins. Protein expression was performed at 30 °C overnight. Cells were harvested by centrifugation (6000 rpm, JA-10, 20 min, 4 °C, Beckman) and the pellet was resuspended in 50 mM NaH₂PO₄, pH 8.0, 150 mM NaCl and protease inhibitor cocktail (Complete, Roche, 1 tablet, for 50 mL of resuspension buffer). Cells were sonicated (70% (200 W) for 15 min (10 sec on, 10 sec off) and afterwards incubated for 30 min on ice with PMSF solution (final concentration 0.1 mM), DNaseI (0.1 mg/mL) and 10 mM MgCl₂. After centrifugation (16000 rpm, JA-17, 45 min, 4°C, Beckman), the supernatant was equilibrated with 5 mL of pre-equilibrated Strep-tag Tactin column material (Strep-Tactin Sepharose 50% suspension, IBA Lifescience) for 1 h (mixed at 200 rpm on a rotatory shaker) at 4°C. The column was washed with three times with 3XCV (column volume) of the resuspension buffer, and the protein was eluted with four fractions of 1X CV of resuspension buffer containing 5 mM desthiobiotin. Elution fractions were analysed on a 12% polyacrylamide SDS-Tris Tricine gel followed by Coumassie staining (InstantBlue, Expediton). Fractions containing protein were pulled and concentrated using Vivaspinn Turbo (5000 MWCO, Sartorius) centrifugation filters and they were dialysed twice against the reaction buffer containing 20 mM 2-(N-morpholino)ethanesulfonic acid (MES) and 500 mM NaCl at pH 5.0, unless otherwise noticed. Concentration of the proteins was measured on a Nanodrop 2000 (Thermo Scientific), using the calculated extinction coefficient ProtParam, ExPASy server). The proteins with A²⁶⁰/A²⁸⁰ higher than 0.7, were purified *via* cation exchange chromatography on a Hitrap Heparin HP column by a gradient of NaCl concentration from 0 to 1 M in 5 min with a flow of 1 mL min⁻¹. Typical expression yields were between 15 and 30 mg/L. Analytical size exclusion chromatography was performed on a Superdex 75 10/300 GL (GE Healthcare). 100 µL of the sample was injected using 20 mM MOPS, pH 7.0, 500 mM NaCl as buffer (flow 0.5 mL min⁻¹).

5.5. REFERENCES

- 1 H. Renata, Z. J. Wang and F. H. Arnold, *Angew. Chem. Int. Ed.*, 2015, **54**, 3351–3367.

- 2 O. Khersonsky and D. S. Tawfik, *Annu. Rev. Biochem.*, 2010, **79**, 471–505.
- 3 I. Nobeli, A. D. Favia and J. M. Thornton, *Nat. Biotechnol.*, 2009, **27**, 157–167.
- 4 R. B. Leveson-Gower, C. Mayer and G. Roelfes, *Nat. Rev. Chem.*, 2019, **3**, 687–705.
- 5 O. F. Brandenburg, R. Fasan and F. H. Arnold, *Curr. Opin. Biotechnol.*, 2017, **47**, 102–111.
- 6 P. S. Cohelo, E. M. Brustad, A. Kannan and F. H. Arnold, *Science*, 2012, **3747**, 307–311.
- 7 M. Bordeaux, V. Tyagi and R. Fasan, *Angew. Chem. Int. Ed.*, 2015, **54**, 1744–1748.
- 8 M. Bordeaux, R. Singh and R. Fasan, *Bioorganic Med. Chem.*, 2014, **22**, 5697–5704.
- 9 D. C. M. Albanese and N. Gaggero, *RSC Adv.*, 2015, **5**, 10588–10598.
- 10 Y. Okamoto, V. Köhler, V. Lebrun, R. Reuter, T. R. Ward, F. Schwizer, T. Heinisch, M. M. Pellizzoni, Y. Gu and J. C. Lewis, *Chem. Rev.*, 2017, **118**, 142–231.
- 11 J. Bos, F. Fusetti, A. J. M. Driessen and G. Roelfes, *Angew. Chem. Int. Ed.*, 2012, **51**, 7472–7475.
- 12 L. Villarino, K. E. Splan, E. Reddem, L. Alonso-Cotchico, C. Gutiérrez de Souza, A. Lledós, J.-D. Maréchal, A.-M. W. H. Thunnissen and G. Roelfes, *Angew. Chem. Int. Ed.*, 2018, **57**, 7785–7789.
- 13 M. Bersellini and G. Roelfes, *Org. Biomol. Chem.*, 2017, **15**, 3069–3073.
- 14 G. Roelfes, *Acc. Chem. Res.*, 2019, **52**, 545–556.
- 15 P. K. Madoori, H. Agustiandari, A. J. M. Driessen and A.-M. W. H. Thunnissen, *EMBO J.*, 2009, **28**, 156–166.
- 16 J. L. Ramos, M. Martinez-Bueno, A. J. Molina-Henares, W. Teran, K. Watanabe, X. Zhang, M. T. Gallegos, R. Brennan and R. Tobes, *Microbiol. Mol. Biol. Rev.*, 2005, **69**, 326–356.
- 17 C. Gutiérrez de Souza, M. Bersellini and G. Roelfes, *ChemCatChem*, 2020, **12**,

- 3190–3194.
- 18 A. García-Fernández, R. P. Megens, L. Villarino and G. Roelfes, *J. Am. Chem. Soc.*, 2016, **138**, 16308–16314.
- 19 L. Villarino, S. Chordia, L. Alonso-Cotchico, E. R. Reddem, A.-M. Thunnissen, J.-D. Maréchal and G. Roelfes, *ACS Catal.*, 2020, **10**, 11783–11790.
- 20 J. P. Phelan and J. A. Ellman, *Beilstein J. Org. Chem.*, 2016, **12**, 1203–1228.
- 21 J. T. Mohr, A. Y. Hong and B. M. Stoltz, *Nat. Chem.*, 2009, **1**, 359–369.
- 22 C. Fehr, *Angew. Chem. Int. Ed.*, 1996, **35**, 2566–2587.
- 23 S. Grkovic, M. H. Brown, M. A. Schumacher, R. G. Brennan and R. A. Skurray, *J. Bacteriol.*, 2001, **183**, 7102–7109.
- 24 K. M. Peters, B. E. Brooks, M. A. Schumacher, R. A. Skurray, G. Richard, M. H. Brown, R. G. Brennan and M. H. Brown, *PLoS One*, 2011, **6**, e15974.
- 25 C. Gutiérrez de Souza, L. Alonso-Cotchico, M. Bersellini and G. Roelfes, *ChemRxiv, Prepr.*

CHAPTER 6

Conclusions and Perspectives

6.1. INTRODUCTION

In the last 50 years, organometallic catalysis has had a great impact in synthetic chemistry by allowing the discovery of new chemical transformations. Nowadays, most of the synthetic routes used both in research and industry contain some catalytic steps.¹⁻³ The growth and great success of this field is due to the vast amount of reactions that organometallic catalysts have enabled, and the high yields and selectivities that they are able to achieve. However, this area is still expanding and lately the focus is increasingly shifting towards sustainability.⁴ Despite the great efforts and achievements in organometallic catalysis, the use of toxic solvents and reagents, harsh reaction conditions, high E factors and poor atom economy are still a significant challenge. On the other hand, enzymes have evolved over time to be proficient in a specific function, while working under environmentally friendly conditions.

Evolution is driven by the environment, and, therefore, by the reciprocal evolution of other species. Joining forces between specialists in different disciplines and combine efforts towards a common goal, is a natural evolution of scientific research. In this regard, the development of artificial (metallo)enzymes is a logical progression that aims to put together chemists and biochemists to get the best of both worlds,⁵ as it is commonly said in the field.

As it has been described in this thesis, several approaches can be followed to create artificial enzymes. Despite the new advances in this area, and the many examples of excellent protein catalysts artificially created, all these approaches still have limitations that hamper their application.⁶ Research on the growing field of ArMs is, therefore, important to keep perfecting these methodologies and could eventually lead us to have a deeper understanding of the structure-function relationship of enzymes, thus allowing scientists to shape them at will.

Overall, this research shows the utility of using multibinding proteins as versatile scaffolds for artificial metalloenzymes. The results discussed here indicate that the use of flexible scaffolds that can adapt to different anchoring approaches, incorporate a broad range of metal complexes, and accept an

array of substrates is a promising method to access a variety of activities. The relatively low efficiencies of these ArMs, however, need to be tackled by applying further protein engineering techniques. Nevertheless, this work shows the potential of multidrug resistance regulator proteins as promiscuous scaffolds for artificial (metallo)enzymes.

6.2. RESEARCH OVERVIEW

The large number of possibilities, both in the protein scaffold-cofactor pairs to be used and the further protein mutants to be produced and screened, is one of the biggest challenges when creating and optimising artificial enzymes. In the quest for an effective methodology that would facilitate the selection and optimisation of ArMs, this thesis describes a journey through different approaches to create ArMs, exploring their challenges and possibilities. Moreover, we aimed to study a small selection of versatile scaffolds for artificial metalloenzymes with similar properties that could be easily screened and optimised for any desired activity.

Binding multispecificity of multidrug resistance regulator proteins (MDRs), grants them the ability of incorporating a broad range of catalytic cofactors and being repurposed into ArMs. The fact that they do not have inherent catalytic activity gives the added advantage of not being biased by a pre-shaped active site that has been evolved for being substrate and reaction specific.

Firstly, the LmrR protein scaffold was investigated. In **Chapter 2**, this scaffold is assembled into an ArM for the challenging enantioselective hydration of enones, using a metal binding unnatural amino acid (UAA) for recruiting the catalytically active Cu^{2+} ion. The previously designed active site of this artificial metallo-hydratase evidenced the utility of incorporating the UAA (2,2'-biyridin-5-yl)alanine (bpyA) into LmrR's scaffold to create an enzyme that gave rise to good conversion and moderate enantioselectivity.⁷ This ArM left an opportunity for investigating the role of the second coordination sphere by rationally selecting point mutations in the hydrophobic binding site. An improved variant that was predicted to be beneficial, LmrR_M89bpyA_V15E_A92I, led indeed to an increase of yield and ee of the ArM. This result indicated that a few judiciously selected mutations can

provide valuable information about residues of the active site and even lead to variants with improved activity, reducing the screening effort. However, a limitation of rational design is that only active site mutations can be predicted. In this chapter, weaknesses of the UAA incorporation approach were identified and efforts towards overcoming them were described. A laborious synthesis of the UAA and a poor expression yield hindered the advance of the project. Therefore, an alternative auto-induced expression method to recycle the UAA from the culture media was investigated. Whereas with some mutants this methodology allowed to recycle the UAA left in the media after several expression cycles, the protein variants with poor expression in LB media, led to even lower expression yields when following this protocol. The fact that the effectiveness of the auto-induced media was protein dependent hampers its application as general expression protocol.

In **Chapter 3**, we described the use of the supramolecular approach for incorporating a metal iron porphyrin reactivity into the LmrR scaffold. The broad scope of reactivity of metalloporphyrins makes them an interesting group of molecules to be used as catalytic cofactor in ArMs. The supramolecular approach of assembly is a straightforward methodology that allows for an easy screening and optimisation. A heme-LmrR ArM was successfully applied in the abiotic enantioselective cyclopropanation of styrenes. A series of spectroscopic techniques, as well as X-ray crystallography, revealed the formation of the ArM, and provided evidence for the flexibility of the system. Molecular dynamics simulations indicated that the dynamics of the system are responsible for allowing catalytic activity of the ArM, even when lacking an apparent free coordination site or space proximal to the catalytic iron site to bind the substrates. The metalloporphyrin-LmrR based ArMs further showed their potential with the ability of LmrR to bind several iron porphyrins (as shown by the measured binding constants) and their catalytic activity in an N-H insertion reaction. Heme-LmrR based ArMs are unique compared to other heme-enzymes, since these usually bind heme through axial coordination to an amino acid residue of the protein. Axial ligation plays a key role on the electronic properties of the metal, therefore attempts to engineering an axial ligand by introducing heme binding residues at specific positions are also described in this chapter.

However, Resonance Raman spectra seemed to indicate that the mutants prepared did not provide an axial ligand to coordinate the metal.

In **Chapters 4** and **5**, three other proteins from the TetR family of MDRs (CgmR, RamR and QacR) were tested as potential scaffolds for ArMs. The initial design involved formation of the ArMs by supramolecular assembly with Cu(phenanthroline) complex. The enzymes were tested in the Friedel-Crafts alkylation reaction (F-C) of indoles with α -substituted α,β -unsaturated enones. **Chapter 4** describes how we serendipitously found that this scaffolds led to better results in presence of Cu(NO₃)₂ without ligand than with the phenanthroline complex. This was an interesting finding because in previous experience of this reaction with LmrR and DNA, the aromatic phenanthroline ligand was necessary to recruit the enzyme into the scaffold and afford enantioselectivity. Inhibition experiments suggested that the reaction happens inside the hydrophobic pockets of these proteins. We hypothesized that the aromatic and cationic complex formed by Cu²⁺ ions coordinated to the imidazole substrate has enough affinity to bind in the pockets of CgmR, RamR and QacR. These results show the potential of these proteins to be used as ArMs, with a simplified version of supramolecular approach that does not require a ligand. Moreover, Cu²⁺⊂QacR led to low enantioselectivity in cell-free extract (CFE), which was an encouraging step towards catalysis *in vivo*. However, further investigation and optimisation of the catalysis protocol in CFE is required.

Furthermore, in **Chapter 5** the Cu²⁺⊂TetR systems were tested in a more challenging reaction, the F-C/enantioselective protonation (F-C/EP). This reaction differs from the F-C in the position of the substituent in the α,β -unsaturated enone substrate (β position for F-C and α position for F-C/EP), which reflects on the protonation step of the reaction being responsible for creating a chiral centre. The Cu²⁺⊂TetR ArMs were also able to catalyse this reaction with low yields and low to moderate ee at pH 7.0. Interestingly, the three proteins led to different enantiomeric outputs: RamR and QacR gave rise to opposite enantiomers, and CgmR to a racemic mixture. Moreover, QacR also showed its versatility, since at pH 5.0 it was able to accelerate this reaction in absence of Cu(NO₃)₂. Molecular dynamics simulations (MD) and mutagenesis studies gave valuable information about the reaction and the

results suggested that the F-C/EP reaction can be accelerated by a proton donor amino acid residue of QacR, Glu120. These results indicated that in a pH dependent manner QacR follows different preferred mechanisms to catalyse this reaction: Lewis acid accelerated with Cu^{2+} at pH 7.0 or Brønsted acid accelerated at pH 5.0 with Glu120 acting as proton donor.

Overall, several ArMs based on MDR proteins able to catalyse a range of chemically relevant reactions are explored in this thesis, as well as, the strengths and limitations of the approaches used.

6.3. RESEARCH PERSPECTIVES

Artificial metalloenzymes constitute a growing field with a great impact in biocatalysis, being able to access new-to-nature reactivities. Although many successful examples of such enzymes have been developed during the last decades, several challenges (like more general application and optimisation) remain to be overcome. In fact, the laborious purification of multiple protein variants is one of the main challenges. Recent efforts have been focussed on creating ArMs able to function in cell-free extracts or *in vivo*, which drastically facilitates high throughput screening. Ultimately, incorporating artificial enzymes into novel metabolic pathways, would allow for sustainable production of abiotic chemicals.⁸ The supramolecular approach and the biosynthetic incorporation of UAA are the only assembly methodologies that can be applied *in vivo*. Therefore, the ArMs studied in this thesis are created using these two approaches.

The expanded genetic code methodology, developed by Schultz and co-workers,⁹ allows to incorporate unnatural amino acids into proteins. The use of metal-binding UAA to attract catalytically active metal ions has been successfully exploited to create LmrR based ArM, as described in **Chapter 2**. However, the limited number of available tRNA/aatRNA synthetase pairs for incorporation of metal-binding UAA reduces the possibilities of this approach.¹⁰⁻¹² An alternative use of UAA in ArMs, is the incorporation of amino acids bearing reactive groups that can be used as catalytic residues.¹³ Production of tRNA/aatRNA synthetase pairs for incorporation of more unnatural amino acids will contribute to the development of the field. The issues described in **Chapter 2** regarding low protein expression yields and

availability of UAAs also need to be addressed. Presence of truncated protein after purification due to lack of incorporation of the UAA could be circumvented by using *E. coli* strains which do not contain the translation release factor RFL, and, thus, do not recognise TAG as stop codon.^{14,15} On the other hand, alternative synthetic routes to obtain the UAA used, (2,2'-bipyridin-5-yl)alanine (bpyA), would be required for further studies with this UAA.

Supramolecular assembly with promiscuous proteins bearing a large hydrophobic pore like LmrR, presents numerous possibilities to form ArMs for new-to-nature reactivities. This is exemplified in **Chapter 3**, which describes how incorporating porphyrins into LmrR's scaffold can open the way to the broad range of reactivities inherent of metalloporphyrin moieties. Designing such a highly dynamic system is challenging. However, **Chapter 3** shows how flexible scaffolds and systems in dynamic equilibrium can rearrange into catalytically active conformation that could not have been predicted based on previous knowledge of the protein structure.

Crystallographic data of the protein scaffold or effects of mutations in similar systems are necessary for a rational or computational design. For instance, in **Chapter 2**, a visual inspection of the X-ray structure of LmrR, allowed to identify potential beneficial mutations in the active site. Despite that rationally or computationally predicting beneficial mutations can considerably reduce the number of variants to be screened, our limited knowledge of structure-function relationship of enzymes makes a fully rational design of ArMs not viable at present. In this thesis, we explored a small selection of versatile scaffolds with similar properties that could be easily screened for new reactivities and further optimised to achieve highly active and selective ArMs.

Three potential candidates from the TetR family of MDR proteins (CgmR, RamR and QacR), with X-ray structures with cationic aromatic drugs bound available were explored in **Chapters 4** and **5**. CgmR, RamR and QacR, proved their versatility by unexpectedly catalysing different reactions with metal-ligand complexes, metal ions and, in the case of QacR, using an amino acid residue of the same protein as proton donor (**Chapter 5**). The results obtained indicated that these three proteins, together with LmrR, are a good small

library of promiscuous scaffolds for ArM screening and represent a viable alternative to designing scaffold-cofactor pairs from scratch.

Moreover, these simplified versions of the supramolecular approach should facilitate the formation of the ArMs *in vivo*. Exploratory experiments described in **Chapter 4**, indicated that Cu²⁺⊂QacR ArM can assemble in cell lysates. Optimisation of this protocol, and testing these ArMs in cell-free extracts and *in vivo*, should be the next step in order to increase the possibilities of high throughput screening.

Unfortunately, the high versatility of a protein scaffold occasionally comes at expenses of lower catalytic efficiencies. Co-crystallising QacR with the Cu²⁺-substrates complex and calculations of the energy profile of this F-C/EP reaction would be advisable for a better understanding of the systems. Hopefully, further developments of protein engineering tools -like computational techniques, or analysis technologies suitable for high throughput screening-, as well as the integration of ArMs in cells, will facilitate achieving high efficiencies from versatile scaffolds.

REFERENCES

- 1 E. V. Gusevskaya, *Quim. Nova*, 2003, **26**, 242–248.
- 2 P. Sautet and F. Delbecq, *Chem. Rev.*, 2010, **110**, 1788–1806.
- 3 A. de Meijere and T. Kurahashi, *J. Organomet. Chem.*, 2007, **42**, 152–177.
- 4 J. D. Smith, F. Gallou and S. Handa, *Johnson Matthey Technol. Rev.*, 2017, **61**, 231–245.
- 5 M. R. Ringenberg and T. R. Ward, *Chem. Commun.*, 2011, **47**, 8470–8476.
- 6 H. J. Davis and T. R. Ward, *ACS Cent. Sci.*, 2019, **5**, 1120–1136.
- 7 I. Drienovská, L. Alonso-Cotchico, P. Vidossich, A. Lledós, J.-D. Maréchal and G. Roelfes, *Chem. Sci.*, 2017, **8**, 7228–7235.
- 8 M. Jeschek, S. Panke and T. R. Ward, *Trends Biotechnol.*, 2018, **36**, 60–72.
- 9 L. Wang, A. Brock, B. Herberich and P. G. Schultz, *Science*, 2001, **292**, 498–500.

-
- 10 Q. Zhou, M. Hu, W. Zhang, L. Jiang, S. Perrett, J. Zhou and J. Wang, *Angew. Chem. Int. Ed.*, 2013, **52**, 1203–1207.
 - 11 X. Liu, Y. Yu, C. Hu, W. Zhang, Y. Lu and J. Wang, *Angew. Chem. Int. Ed.*, 2012, **51**, 4312–4316.
 - 12 X. Liu, J. Li, C. Hu, Q. Zhou, W. Zhang, M. Hu, J. Zhou and J. Wang, *Angew. Chem. Int. Ed.*, 2013, **52**, 4805–4809.
 - 13 I. Drienovská, C. Mayer, C. Dulson and G. Roelfes, *Nat. Chem.*, 2018, **10**, 946–952.
 - 14 I. U. Heinemann, A. J. Rovner, H. R. Aerni, S. Rogulina, L. Cheng, W. Olds, J. T. Fischer, D. Söll, F. J. Isaacs and J. Rinehart, *FEBS Lett.*, 2012, **586**, 37616–3722.
 - 15 J. Fredens, K. Wang, D. de la Torre, L. F. H. Funke, W. E. Robertson, Y. Christova, T. Chia, W. H. Schmied, D. L. Dunkelmann, V. Beránek, C. Uttamapinant, A. G. Llamazares, T. S. Elliott and J. W. Chin, *Nature*, 2019, **569**, 514–518.

Samenvatting _____

Multidrug Resistance Regulators als Kunstmatige Enzymen

Translated by **Roos van Lier**

In de zoektocht naar groenere chemie om de vaak giftige, afval producerende en vervuilende processen van de chemische industrie te vervangen, wordt de natuur vaak als inspiratiebron gebruikt. Enzymen zijn de beste chemici op aarde, ze hebben zich ontwikkeld om extreem efficiënt en selectief te zijn voor een specifieke reactie. Kunstmatige (metaal)enzymen (ArM's) zijn hybride biochemische systemen die zich kenmerken door combinatie van het brede reactiespectrum van de organische chemie en de hoge efficiëntie en selectiviteit van enzymen, terwijl ze onder milieuvriendelijke omstandigheden werken.

De ontwikkeling van enzymen is gebaseerd op een mate van promiscuïteit (hetzij in activiteit, hetzij in binding), waardoor ze zich kunnen aanpassen aan veranderingen in de omgeving. Geïnspireerd door dit concept, werd gerichte evolutie gebruikt als een extreem krachtige techniek om verbeterde enzymen te vinden. Echter, net zoals in de natuur, moeten deze enzymen een minimaal niveau van de gewenste katalytische activiteit vertonen om te kunnen evolueren. Ook blijft de algemene toepassing van dit concept van gerichte evolutie uitdagend vanwege het grote aantal enzymvarianten dat gescreend moet worden, waarvoor vaak apparatuur nodig is die in de meeste laboratoria ontbreekt. Een andere manier om de promiscuïteit van eiwitten te benutten om nieuwe kunstmatige enzymen te maken is het gebruik van eiwitten met meerdere bindingsplaatsen om substraten en metaalcofactoren op te nemen die katalytisch actief zijn voor de gewenste reactiviteit. Deze methode maakt het mogelijk om de bindingspromiscuïteit van eiwitten te extrapoleren naar de promiscue reactiviteit van kunstmatige enzymen. In dit proefschrift worden een aantal multispecifieke eiwit-bioplatformen onderzocht en omgezet naar ArM's. De geselecteerde bioplatformen behoren tot de klasse van Multidrug Resistance Regulator eiwitten (MDR's). Deze eiwitten zijn dimeren die in staat zijn om diverse aromatische vlakke gastmoleculen, waaronder verschillende medicijnen, te herkennen en antibioticaresistentie te bevorderen. Hun intrinsieke promiscuïteit voor het binden van een verscheidenheid aan moleculen maakt ze een geschikte

eiwitklasse voor het creëren van ArM's via verschillende assemblagemethoden, die elk diverse kenmerken aan het systeem geven.

Het eerste eiwit dat wordt onderzocht is LmrR, afkomstig van *Lactococcus lactis*, dat een grote hydrofobe holte heeft op het dimeergrensvlak. LmrR is door onze groep gebruikt om katalytische metaalcomplexen te introduceren met behulp van verschillende strategieën. In dit proefschrift worden op LmrR gebaseerde ArM's beschreven die gecreëerd zijn door het opnemen van metaalbindende onnatuurlijke aminozuren (UAA's) of door supramoleculaire assemblage. In **Hoofdstuk 2** wordt het onnatuurlijke aminozuur 2,2'-(bipyridin-5,5')alanine (bpyA) geïntroduceerd als metaal chelator in LmrR om de enantioselectieve additie van water aan enonen te katalyseren. In de natuur zijn hydratatiereacties sterk afhankelijk van secundaire interacties met het enzym. Daarom zijn er pogingen gedaan om de activiteit en selectiviteit van een eerder ontworpen kunstmatig metaalhydratase te verbeteren door het rationeel optimaliseren van de tweede coördinatieomgeving. Deze resultaten zijn beschreven in **Hoofdstuk 2**. De vondst van een gunstige mutatie waarbij zowel de opbrengst als de enantioselectiviteit voor twee verschillende substraten werden verbeterd, liet zien dat het toepassen van rationele optimalisatie voor het ontwerp van een ArM gebaseerd op het LmrR-eiwit met een metaalbindend onnatuurlijk aminozuur mogelijk is. Deze methodologie vereist geen modificaties van het eiwit na de expressie en biedt een grote mate van controle over de reactie, wat de rationele optimalisatie van de kunstmatige metaalhydratase vereenvoudigt. Echter, de kosten van sommige UAA's, het feit dat sommige UAA's niet commercieel verkrijgbaar zijn, de bewerkelijke en dure synthese van een aantal UAA's, evenals de lage expressieopbrengsten die worden verkregen, beperken de toepasbaarheid van deze methode. In dit hoofdstuk wordt een alternatieve expressiemethode onderzocht waarbij het UAA kan worden gerecycled uit het expressiemedium. Een automatisch geïnduceerd glucosemedium wordt gefilterd onder steriele omstandigheden en hergebruikt voor een andere expressiecyclus, na toevoeging van geoptimaliseerde hoeveelheden van de vereiste voedingsstoffen. Deze

methode genereerde meer bpyA-bevattende LmrR uit dezelfde hoeveelheid onnatuurlijk aminozuur, voor de mutanten die het beste tot expressie kwamen. Niettemin waren de mutanten die lage expressieniveaus vertoonden in algemene LB media ook niet geschikt voor dit expressieprotocol.

Vervolgens werd de supramoleculaire strategie onderzocht. Deze methode maakt een eenvoudige assemblage mogelijk en vereenvoudigt de optimalisatie, maar leidt tot een dynamisch systeem dat moeilijker te controleren is. In **Hoofdstuk 3** hebben we uitgezocht of het gebruik van LmrR's grote hydrofobe holte potentie biedt voor het opnemen van verschillende porfyrynes. We hebben aangetoond dat LmrR, heem stevig kan binden en ook bindingsaffiniteit vertoont voor andere porfyrynes. Het heem-LmrR complex bleek gebruikt te kunnen worden voor een abiotische reactie: de cyclopropanering van styreen. Spectroscopische technieken, röntgenkristallografie, moleculaire dynamica-simulaties en mutagenesestudies maakten het mogelijk om het systeem beter te begrijpen en toonden het belang van een flexibel bioplatform aan om de substraten in een geschikte conformatie voor katalyse te kunnen rekruteren. Het heem-LmrR enzym bleek ook actief te zijn voor een N-H insertiereactie, wat aangeeft dat het mogelijk is om de toepasbaarheid van dit ArM uit te breiden. In tegenstelling tot natuurlijke heem-bevattende eiwitten, heeft dit unieke heem-LmrR enzym geen axiaal ligand om ijzer te binden. Het axiale ligand is belangrijk om de elektronische eigenschappen van het ijzer en de positie van de porfyryne cofactor te controleren. Daarom beschrijven we in dit hoofdstuk ook pogingen om een axiaal ligand in LmrR op te nemen en onderzoeken we de mogelijkheid om de porfyryne-LmrR gebaseerde ArM's te karakteriseren met behulp van spectroscopische technieken, zoals Resonance Raman.

In de volgende hoofdstukken werd de supramoleculaire strategie getest met drie andere eiwitten (CgmR, RamR en QacR) afkomstig uit de TetR familie van de MDR's. In tegenstelling tot de grote hydrofobe holte in het dimeergrensvlak van LmrR bevatten de eiwitten afkomstig uit de TetR familie twee kleinere hydrofobe holtes. Deze eiwitten hebben een regulerend

domein, welke sterk behouden is binnen de TetR familie, en een domein dat gastmoleculen, bijvoorbeeld medicijnen, kan binden. Deze verschilt sterk van eiwit tot eiwit en leidt tot de grote bindingsdiversiteit binnen deze familie. De verscheidenheid van substraten die kunnen worden gebonden in CgmR, RamR en QacR, evenals LmrR, maken dit scala van eiwitten tot een goede selectie van bioplatformen die gemakkelijk kunnen worden gescreend voor nieuwe activiteit met behulp van de supramoleculaire strategie.

Dit wordt geïllustreerd in **Hoofdstuk 4**, waar CgmR, RamR en QacR worden gebruikt als bioplatform voor ArM's voor de Friedel-Crafts alkylering van indolen met β -gesubstitueerde α,β -onverzadigde enonen (F-C). Deze reactie is eerder gerapporteerd met katalyse door een Cu(II)(fenanthroline)-LmrR gebaseerde ArM, maar de kleinere holtes van CgmR, RamR en QacR vertonen een betere enzymatische activiteit met alleen Cu^{2+} -ionen, zonder dat er een ligand nodig is. We denken dat het kationische en aromatische Cu^{2+} -substraatcomplex voldoende affiniteit heeft voor de holtes van deze eiwitten en dat dit daarom gebruikt kan worden voor katalyse. Bovendien wordt in **Hoofdstuk 5** het bereik deze TetR eiwitten als ArM's aangetoond door het gebruik van een meer uitdagende reactie: de Friedel-Crafts/enantioselectieve protonering van indolen met α -gesubstitueerde α,β -onverzadigde enonen (F-C/EP). Misleidend is dat een klein verschil in het substraat kan leiden tot een significant verschil in het mechanisme: in het geval van de F-C, wordt chiraliteit gecreërd in de alkyleringsstap, terwijl in de F-C/EP, de protonering leidt tot de vorming van het chirale centrum. In het geval van de F-C/EP zijn de kleine, wendbare protonen moeilijk controleerbaar, wat het bereiken van enantioselectiviteit moeilijker maakt. Interessant genoeg waren CgmR, RamR en QacR in staat om de enantioselectiviteit in de F-C/EP reactie te versnellen met Cu^{2+} -ionen bij pH 7.0. De drie eiwitten leidden tot verschillende uitkomsten qua enantioselectiviteit (tegenovergestelde enantiomeren voor RamR en QacR en een racemisch mengsel voor CgmR). Opmerkelijk was dat QacR deze reactie ook kon uitvoeren bij pH 5.0 zonder dat er een metaal nodig was. MD simulaties en mutagenese studies suggereerden dat een zuur aminozuur in het eiwit verantwoordelijk is voor de protonering van het

prochirale tussenproduct dat wordt gevormd in de reactie. Dit gebeurt bij voorkeur via een van de enantiotopische vlakken. Deze resultaten benadrukten nog maar eens weer dat deze eiwitten een veelzijdige selectie van bioplatformen voor ArM's zijn en toont aan dat de supramoleculaire strategie geschikt is voor een makkelijke screening naar nieuwe reactiviteit.

Samenvattend, het werk gepresenteerd in dit proefschrift beschrijft verschillende ArM's die zijn gebaseerd op Multidrug Resistance Regulator eiwitten die in staat zijn een reeks chemisch relevante reacties te katalyseren. De sterke kanten en beperkingen van de gebruikte assemblage methodes, evenals de verschillende kenmerken die worden verkregen bij het toepassen van elk van deze methodes, worden besproken. Over het algemeen bieden de bevindingen potentie voor het herontwikkelen van multispecifieke eiwitten tot nieuwe kunstmatige (metaal)enzymen.

Acknowledgements

This final bit of the thesis is a space of self-reflexion where I am given the chance to dedicate a thought to all the people who, in one way or another, contributed to fill the book of my PhD expedition with personal and scientific growth, great memories, and precious friendships.

First of all, I would like to thank **Gerard** for giving me the opportunity to be part of his group and embark on this PhD journey. Thank you for your support during all these years. I probably never fully believed myself that I was good enough, but you gave me the chance to try and believed in me even when things in the lab were not going as well as planned. It took me a while to get used to “the Dutch way” of working, but I think we finally found a good balance. Thanks also for supporting me when I got excited planning group activities and for acknowledging the importance of social interactions within the group. Sorry for scaring you when I entered your office saying “I have to tell you something...”, usually followed by information about some *little* accident...

Thanks to the members of my reading committee Prof. **Marco Fraaije**, Dr. **Wiktor Szymanski** and Prof. **Jean-Pier Mahy**, and my second supervisor Prof. **Sijbren Otto** for proofreading this thesis and the valuable comments.

I would like to express my gratitude to **Wesley** for helping me with the Raman measurements, which opened my eyes to a different way of looking at properties of metallo-proteins.

Dear **Katy**, it was a pleasure to work together. Thank you for the interesting discussions, scientific and otherwise.

I couldn't choose other paranymphs that the two people who paranymphed me from the very beginning without me even asking for it. Who guided me through the PhD and through life in the lovely Groningen.

Dear **Manu**, my bubbly paranymph. I immediately saw a shine behind the ‘toughManuwhodidn’twantmorefriends’ wall, and I am happy I climbed it up to discover the most dedicated scientist, colleague and friend. After probably quite some hesitation you offered me being neighbours and that was only the first of uncountable times that you unselfishly supported me through-out the whole PhD... and beyond! Thanks for the help and advice in the lab, for the long nights out, the climbing, the trips, the laughs, chats and friendship which helped me to push through. Thanks for sharing QacR's success! I am pleased to have landed in Cambridge to be your bubble (and continue to benefit from your wisdom). You have always been an example to admire and I am sure you will find your place in this crazy scientific world.

Dear **Ivana**, I can't thank you enough for taking me under your wing when I arrived and did not even know where to start. I had a tough beginning and your positivism and cheerfulness was exactly what I needed to drive forward. Thanks for teaching me the *bio stuff* and for motivating me to keep optimistic even when experiments didn't work, definitely the most valuable lesson. You are such a lovely and caring friend and colleague. Your passion and determination grew you into the successful scientist you dreamt of being and I am sure that many more achievements, scientific and otherwise, are there to come (which we'll celebrate together!). Don't ever lose your unconditional smile and *everything will be fine*!

I would like to thank the past and present members of the Roelfes group (and adopted kids), who altogether created an environment that made me happy to go to work every day. First, of course, **The Roelfes Girls**, definitely not over!

Sambi, my hugger source of energy. I always appreciated your transparency and honesty. You don't let everyone see your bright side, but the lucky ones that get to your heart know how a sensitive and loving person you are. Your superpower is your always welcoming smile and open arms ready to show your love and support. Your hugs have done a lot of good to the world (and to me!!). Let your best self shine, I am sure that there will be a lot of happiness in your life. **Lili**, you are the first responsible for me jumping on this PhD. You were so kind and nice when you showed me around during the interview and my first days that I immediately felt at home. Because that's what you do. You make everyone feel calm, unjudged and comfortable around you. Keep up your positive spirit and enjoy this new period of your life. **Lili**, it was an honour to be your paranymph and to share this task with **Sambi**. You two will always have a special spot in my heart wherever we are. **Fra**, I am sure you'll be delighted to be included in The Roelfes Girls' section. You are amazingly annoying and funny at the same time. Thanks for the laughs and take care of my Lili! **Lariña**, qué bien nos lo hemos pasado tú y yo. Fue un placer compartir tanto tiempo contigo en lab, con interminables cafés, y trabajando hasta las tantas a ritmo de Tina Turner. Siempre fuiste la que mejor me entendía, mamiña. Aprendí muchísimo de ti con las ciclopropanaciones y de tu forma de trabajar. Eres la persona más organizada, minuciosa y perfeccionista que conozco, quizás por eso sea más divertido aún cuando te desmadras. El mundo de la investigación no lo pone fácil, pero estoy segura de que encontrarás el sitio en el que se te valore como te mereces. Vete preparándome un pulpito a la gallega que en nada me tienes allí de visita. **Ana**, eres el carisma en persona. Gracias por los buenos momentos y por dejarme hacer de tía de ese bombón que tienes en casa. Tu amistad significó mucho para mí, y más aún en el último año, os echo mucho de menos! **Ruben**, I can't forget how confused I was my first Friday afternoon when you were playing "Sevilla tiene un color especial" full volume in the lab followed by some heavy metal song. I enjoyed sharing office with you and appreciate your efforts for keeping me awake with the loud music. I wish the four of you, **Ana**, **Ruben**, **Amaia** and **Mxxx**, a lot of happiness. **Sara**, fue divertido ver el insólito mundo de Sara desde dentro. Siempre admiré la pasión y determinación que pones en todo lo que haces, estoy segura de que te llevarán lejos. Me costó que salieras pero hiciste que valiera la pena la insistencia. Gracias por las risas, borrels, girls night

y ¡hacernos de guía en Bilbao! ¡Espero que haya muchos más viajes! **Raquel**, gracias por compartir mi pasión por el baile del pollo, qué pena no haber coincidido más tiempo! Me encantaba que fueras una más en las cenas de grupo, siempre añades unas cuantas risas a la noche. Eché mucho de menos tu energía positiva y buen humor cuando te fuiste. **Matt**, thanks for the good times, vuestro último mes en Groningen fue sin duda el más intenso y divertido. Os deseo lo mejor y espero poder conocer a la pequeña **Vera** pronto. **Dowine**, the impersonation of efficiency. It is always nice to be around you and your positivism is contagious. Thanks for the help with the autocolumn, somehow this machine never liked me... **Nathalie**, I am happy we shared some time together. Parties and reunions were always real fun with you! I hope we'll keep meeting around the world. **Annika**, you are extremely hard-working and genuinely good person. I wish you all the best in your future.

Yans, it was nice to share time in the lab with you, always open to scientific discussions. I wish you and your family a lot of happiness. **Reinder**, you are an extremely good colleague and I loved that I could always trust you as DJ in the lab. Thanks for creating a good working atmosphere, being a great lab assistant and for the good music. **Clemens**, thanks for always being willing to share your knowledge and to help out. Your scientific input was a great contribution to everyone in the group and I really appreciate your honesty, and straightforward and always helpful comments. I was happy to be a guinea pig with your cooking experiments, I so much love those burgers!

The second batch of Roelfiejes, who keep my hopes up for a fruitful continuity of social activities within the group. **Roos**, always kind and smiling. It's a joy being around you. Thanks for always caring, being a great lab and office mate and for translating my Dutch summary. **Lur**, fue un placer trabajar contigo, gracias por los cálculos con QacR. Fuiste una buena incorporación para el grupo, ¡y para fiestas y escalada! Mucha suerte con todo, ¡me muero de ganas de ir a ese bodorrio!

Linda, biggest smile on hearth. I enjoyed a lot having you as a friend and officemate. You are always cheerful and keep the mood up around you. My best wishes for you, **Mohamed** and the little cuty **Awindor**. **Zhi**, you and **Linda** were quite an entertainment in the office. I enjoyed our conversations, thanks for always being open to discuss science, politics or personal stuff. I am happy that your years in Groningen were quite successful (both in the lab and at home with your two little ones). All the best for your future career and for your lovely family. **Bart**, it was nice to have you back in the group. It's a shame that we didn't overlap more. I know you are going to do great in your PhD, best of luck! And sorry again for the blue eye... **Michela**, the lab really benefited from your organization skills. You are a great scientist, I wish you a lot of success in your PhD. My students, **Tom** and **Anke**. I was happy to be given the chance to supervise you in the lab, and I learnt a lot from it. I hope you did too and had a good time in the group.

Especial thanks to the baby Roelfes (adopted ones included), a cute and fun crew: **Rudy**, **Titia**, **Claudia**, **Andrea** and **Elyse**. **Andreita**, tienes un corazón de oro.

Gracias por tu apoyo en los momentos difíciles y por el perreo, no pierdas el buen humor! **Rudy**, you are a great guy. Smart, fun and with a big heart. I am sure your PhD will be a success, keep enjoying Groningen! **Titia**, I am happy I got to know you. I enjoyed a lot our conversations and the climbing. I am sure we would have become even closer if we had had more time. **Claudia**, I am sad I wasn't there anymore at the beginning of your PhD. You are a good scientist and a great person. Take it easy and enjoy, you are surrounded by wonderful people and I am sure you'll have the time of your life in Groningen. **Reuben**, you are a good scientist and a fun guy, good luck finishing your PhD. **Ginny**, I am happy you became part of the group. You two are a funny couple, thanks for the good times and parties!

Of course, next come the “downstairs people”, the biochemist part of the family. I first have to thank **Marco** and **Dick** for allowing us to use your facilities and sharing the labs with us. Dear **Marco**, thanks for always treating me as part of the group by allowing me to intrude in your group-coffee breaks and even inviting me to the Christmas dinner at your place. I feel flattered to have you in my committee. **Elvira**, la energía y pasión que pones en todo lo que haces es una fuente de inspiración. Sin buscarlo te convertiste en una de las personas más importantes durante estos años ¡Te echo muchísimo de menos, sesión Madre de José ya! Dear **Niko**, it's true, I didn't like you at the beginning. But when I got to know you, I realised that you are not only the self-confident leader personality that you appear to be, but also a wonderful and caring friend. It was a pleasure to share the lab with you. Laughs, deep conversations as well as your advice in the lab were the most helpful. Thanks for leading this group of crazy friends and for never letting me down. I always know I can count on you and that goes both ways. **Misuntje**, there's some unwrapping to do before reaching the real Misun. It has been a pleasure to go through this process, discovering you little by little. It was great to be party, climbing and life buddies. **Martin**, I am happy that you finally got to live together after all this time. Thanks for the invitation to your wedding, which was definitely one of the highlights of these years. Take care of her! **Ale**, es sorprendente lo bien que nos entendemos, pese al idioma. Me he divertido mucho contigo, gracias por estar siempre dispuesto a escuchar y a echarme una mano cuando lo necesitaba (¡y lo necesité!). **Mel**, gracias por las fiestas latinas, mucha suerte con el final de tu PhD. ¡Espero que nos veamos pronto! **Cate**, I always appreciated your clear mindset, and your energetic and straightforward attitude. I am happy that we ended up being good friends and had a lot of fun together. All the best for you and el loco de **Fermin**. **Eli**, I missed a lot our chatty climbing sessions and parties when you left. I am very happy we keep in touch and I hope to see you, **Malcom** and **Aki** very soon again. Crazy little **Marzy**, with you interesting conversations and a good laugh are for granted. Thanks for the great time and trips together, for sure there are more to come. **Mohi**, I know I was a bit pushy with the hugs, but it made me very happy when you started accepting them! I am looking forward to a trip to Egypt to visit you and your lovely family. **Milos**, it was great to have you in the lab, always kind and helpful, but also outside the lab as a friend. **Jasna**, you were a good acquisition for the girls nights, I miss you! **Sofija**, thanks for playing with me in the ball-pool and in the boring grownup parties. I hope we can play together again soon, maybe also with your little brother! **Mathias** and **Judith** thanks for the time we spent together

in Groningen, Graz and Barcelona (and zoom, I must add), it was always great fun! I am sad Covid didn't let us meet in BCN last summer, but we'll make up for it! **Simone**, it was great that you came to Groningen and joined the Crown crew. Thanks for the fun times! **Friso**, thanks for the nice conversations that always had a calming effect. Good luck in the Roelfes group! **Hemant**, thanks for your joyful attitude and all the time we spent together, especially playing sports and partying! **Suzan**, I really enjoyed the time you spent in Groningen and I was pleased to meet again in the Gordon. You have a lovely family and I had a lot of fun playing with your sons! **Cyntia**, coincidimos poco pero siempre me lo pasé muy bien contigo. ¡Mucha suerte con todo! **Hugo**, thanks for being a good labmate and always making sure that things kept running smoothly in the lab. Dear **Hanna**, we met after you had already left Groningen, but I enjoyed spending time with you during your visits. **Laura**, linda, qué pena que no coincidamos en tu siguiente etapa en Groningen, seguro que lo habríamos pasado bien. Disfrútalo y espero vernos pronto. **Marco** and **Silvia**, I am happy we met and spent some nice time together. Best wishes to you and your little ones. **Nikolas**, thanks for giving a try to crystallising QacR, I am sorry it didn't work out at the end. I wish you good luck at the end of your PhD and lots of crystals! **Ulises**, ¡muchas gracias por tu ayuda con los análisis de Raman! Fue un placer teneros a ti y a **Renata** unos meses en Groningen. Un abrazo para los dos. Overall, thanks to all the biochemistry crew for creating a lovely working environment: **Yiming**, **Jeroen**, **Qinglong**, **Estela**, **Fabiola**, **Elvira Bombino**, **Gautier**, **Brenda**, **Eduardo**, **Yapei**, **Dana**, **Arne**, **Ezequiel**, **Nina**.

Ilias, thank you for always being there to play any kind of sports and drink any kind of beer. Long talks with you are always interesting and I am glad I met you and **Kiki**. I wish the two of you all the best.

Dear **Balint**, thanks for being the best lab-neighbour one could hope for. You always managed to make me smile, with your kind and cheerful attitude and with your chocolates supply, especially when I had a broken foot! You obviously have a brilliant mind, but also a big heart. Good luck with everything to you and **Fanny**!

Thanks to the old Sussi's crowd **Ciccio**, **Simone**, **Tilde**, **Pablo**, **Juanfer** and **Mamen** for the borrels, parties at the Car (and houses), drinks in the Amigo and the basket matches! **Pablo**, gracias por hacerme apreciar el calimocho y por siempre encontrar un hueco para vernos cuando veníais de visita a Groningen con **Gerda** y **Alaia**. ¡La próxima os visitamos nosotros en España!

Dear **Giulia**, I loved sharing cultural activities with you and always appreciated your advice on beers. I am grateful for the nice time we spent together and the friendship that we very quickly developed. I am happy we met **Marlies** and **Michael**, who made us bond even more. **Marlies** and **Michael**, I can't thank you enough for all the spectacles and exhibitions that we attended; for sharing your interminable knowledge with us; inviting us to your beautiful home; and helping us discover the wonders of the nice Friesland. Your friendship meant a lot to me all these years in Groningen. I

wish you all the happiness and I hope we can soon meet again. **Michael (Lerch)**, it was nice to get together on all these lovely occasions. Best of luck with your carrier!

Ivana (Maric), this fight against the injustice of the world wouldn't have been as fun without you. It would have been an honour to defend the same day, but I am happy that at least we went through the last bit of the process together. You've been a great friend and I have to thank you for your support all these years.

Guille, siempre es un placer hablar contigo, tanto en los pasillos de la uni como en fiestas y bares cerveza en mano. Mucha suerte en tus próximos pasos. **Boris**, your DJ abilities at late hours in the night when everybody else's judgement is totally unreliable are really appreciated. Thanks for all the time we spent together and your friendship during these years. **Gaël**, thanks for the fun nights out and for your advice when I was reaching the end of my PhD. **Babis** and **Eirini**, it was nice to have some interesting talks in the Otto's parties.

Gracias a la segunda ronda de Sussies (y españoles adoptados) por los borrels, cenas y birras en el bar de "las tapas", nos echamos unas buenas risas: **Juani, Mariajo, Roxana, Paula, Jorge, Nacho. Juani**, gracias también por hacerme compañía los domingos en la uni e incluso jugar a squash de vez en cuando, se me hizo más llevadero gracias a tí. **Roxana**, un placer que te incorporaras a los Roelfes! Disfruta el tiempo que te queda (¡no solo trabajar!), te has esforzado mucho y estoy segura de que recibirás tu recompensa. ¡Suerte! **Jorge**, me encanta tu humor sarcástico e inteligente. ¡Gracias por los borrels y los partidos de básquet! **Nacho**, tu espontaneidad, tus rarezas adorables y tu buen humor te hacen único. Nunca te he oído decir que no a un plan, molas. No olvidaré que te presentaras en Barcelona para mi cumpleaños. Suerte en tu nuevo trabajo y en la vida.

Dusan, it's always nice to chat with you. Thank you for lending me your card for the green building, it speeded up a lot my data collection when I was very needy of time!

Thanks to the crowd of The Crown, who became our second family in the last years. Especial thanks to **Ashleigh** for your big smile and your ability to always know what we wanted and when. Also, thanks to **Osman, Miglè, Dimitris, Saikat, Erik, Hugo, Herman** and **Kris** for all the fun.

I would like to express my gratitude to all the members of the Stratingh Institute for the technical and administrative support, especially **Annette, Pieter, Monique, Theodora, Renze, Alphons, Hilda** and the **magazijn staff**, who make everybody's life easier in the institute. Dear **Jack Jager**, as safety officer of the department, our paths were meant to cross (several times...) and I owe you a debt of gratitude. Thanks for showing up as a life saver in the few occasions I required your help and, after making sure there was no risk, sharing a good laugh with me about it. Your positivity and sense of humour were a good companion in these "especial" occasions. Dear **Henk**, I missed you when you changed to another canteen. Thanks for always making sure that my favourite Milka Muffin was available and for not letting me get too

skinny by constantly feeding me muffins and ice creams. I always thought that either my mum or my grandmas had sent you to take care of that.

Silvia, you have been a mentor to me in life and science. Your good-will and generosity are inspiring, I couldn't be more grateful to have you as a friend. I wish life brings you a lot of success in your scientific career and a lot of happiness.

Maxi, amiguito. Si una tesis reflejara de verdad lo que es un PhD, tú estarías en cada página de la mía. Gracias por dejarlo todo para darme un abrazo cuando lo necesitaba. Por entenderme cuando hablo en una lengua inventada, y cuando no hablo. Por tus consejos, aunque soy ein bisschen stubborn y tuvieses que insistir para que te hiciera caso, tenías razón (a veces), y cuando tienes razón te la doy (a veces). Por cuidarme y dejarte cuidar. And for all the patoaventuras! Seguiremos explorando el mundo con nuestro ejército de eentjes.

And to end this at the beginning: gracias a mi **familia** y **amigos** en España, a quiénes llevo conmigo allá donde voy. El saber que os tengo cerca, caminando conmigo pese a la distancia, es la fuente de energía que me ayuda a avanzar, sin miedo. No podría estar más orgullosa y agradecida de teneros en mi vida. Os quiero.

Many thanks,

Cora