Structural Base of Cyclic-Nucleotide Dependent Signalling in *Sinorhizobium meliloti*

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
(Dr. rer. nat.)

dem
Fachbereich Chemie
der Philipps-Universität Marburg

vorgelegt von

Laura Werel
M.Sc. Chemie

aus Amöneburg
geboren in Marburg

Marburg/Lahn im Februar 2021

Structural Base of Cyclic-Nucleotide Dependent Signalling in *Sinorhizobium meliloti*

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
(Dr. rer. nat.)

dem
Fachbereich Chemie
der Philipps-Universität Marburg
Hochschulkennziffer 1180

vorgelegt von

Laura Werel
M.Sc. Chemie

aus Amöneburg
geboren in Marburg

Marburg/Lahn im Februar 2021

Vom Fachbereich Chemie der Philipps-Universität Marburg (Hochschulkennziffer: 1180) als Dissertation am 17.02.21 angenommen.

Erstgutachter: Prof. Dr. Lars-Oliver Essen (FB Chemie)

Zweitgutachter: Prof. Dr. Anke Becker (FB Biologie)

Vorsitzender d. Prüfungskommission: Prof. Dr. Ulrich Tallarek (FB Chemie)

Tag der mündlichen Prüfung am 05.05.21.

Meiner Familie

Eidesstattliche Erklärung

Statutory Declaration

Hiermit versichere ich, dass ich die vorliegende Arbeit mit dem Titel

I hereby declare to have authored the thesis

Structural Base of Cyclic-Nucleotide Dependent Signalling in *Sinorhizobium meliloti*

selbstständig und ohne fremde Hilfe verfasst, nicht andere als die in ihr angegebenen Quellen oder Hilfsmittel benutzt, sowie alle vollständig oder sinngemäß übernommenen Zitate als solche gekennzeichnet habe.

on my own without external assistance. I did not use other than the indicated references and resources and I have appropriately identified all quotations as borrowed in full or in paraphrased form.

Diese Dissertation wurde in der vorliegenden oder einer ähnlichen Form noch bei keiner anderen in- oder ausländischen Hochschule anlässlich eines Promotionsgesuchs oder zu anderen Prüfungszwecken eingereicht.

The present dissertation in the current or a similar form has not been handed in to any other domestic or foreign university and did not serve for a doctoral application or any other examination purposes prior to this date.

Marburg/Lahn, den 02. Februar 2021

Marburg/Lahn, February 2nd, 2021

____

Laura Werel

Verlauf des wissenschaftlichen Werdegangs
Course of the academic career

Laura Werel, *M.Sc. Chemie*

geboren am 26.08.1989

born on August 26th 1989

in Marburg/Lahn

10/2017 – Gegenwart	**Promotion in der Arbeitsgruppe von Prof. Dr. Lars-Oliver Essen am Fachbereich Chemie der Philipps-Universität Marburg** Titel der Dissertation: Structural Base of Cyclic-Nucleotide Dependent Signalling in *Sinorhizobium meliloti* Schwerpunkte: Biochemie und Strukturbiologie
10/2017 – present	*PhD in Biochemistry performed in the group of Prof. Dr. Lars-Oliver Essen, Department of Chemistry, Philipps-Universität Marburg*
04/2014 – 09/2017	**Master of Science in Chemie am Fachbereich Chemie der Philipps-Universität Marburg** Schwerpunkte: Organische Chemie, Biochemie Masterarbeit in der Arbeitsgruppe von Prof. Dr. Gert Bange mit dem Titel: Studies on substrate binding and regulation of (p)ppGpp synthetases *Master studies in Chemistry, Department of Chemistry, Philipps-Universität Marburg, Supervisor of the master thesis: Prof. Dr. Gert Bange*
10/2009 – 03/2014	**Bachelor of Science in Chemie am Fachbereich Chemie der Philipps-Universität Marburg** Schwerpunkt: Organische Chemie Bachelorarbeit in der Arbeitsgruppe von Prof. Dr. Ulrich Koert mit dem Titel: Darstellung dihydroxy-funktionalisierter Bromsulfone über geminale Dibromide *Bachelor studies in Chemistry, Department of Chemistry, Philipps-Universität Marburg, Supervisor of the bachelor thesis: Prof. Dr. Ulrich Koert*
09/2002-07/2009	**Abitur am Gymnasium Stiftsschule St. Johann Amöneburg** *University Entrance Diploma at the Gymnasium Stiftsschule St. Johann Amöneburg*

Danksagung

Diese Arbeit wäre in der vorliegenden Form ohne viele Menschen nicht möglich gewesen. Allen voran möchte ich natürlich meinem Doktorvater, Prof. Dr. Lars-Oliver Essen, danken für die vielen wissenschaftlichen Anregungen, den bioinformatischen Support, konstruktiven Diskussion, aber auch die Freiheit auszuprobieren und eigene Ideen zu entwickeln. Sehr positiv wahrgenommen habe ich auch das unkomplizierte und herzliche Miteinander in den vergangenen drei Jahren.

Auch in den Meetings mit Prof. Dr. Anke Becker konnte ich viel vor allem über die biologischen und genregulatorischen Aspekte meines Projektes lernen und habe in Gesprächen immer wieder neue Denkanstöße für meine Arbeit gefunden. Ich danke ihr außerdem für die Übernahme des Zweitgutachtens. Neda Farmani und Dr. Elizaveta Krol danke ich für die gute Zusammenarbeit.

Sehr zu schätzen weiß ich, dass Prof. Dr. Ulrich Tallarek so freundlich war, als Drittprüfer für meine Dissertation zu fungieren.

Wieland Steinchen danke ich für das Vermessen der HPLC-Assays und die HDX-MS Messungen und den Wissensaustausch in Sachen Nukleotid-Messenger. Freundlicherweise durfte ich zur Erstellung meiner Arbeit ITC-Geräte aus den Arbeitsgruppen Bange und Klebe verwenden, für deren Benutzung mir Pietro Giammarinaro beziehungsweise Matthias Oebbeke mit Rat und Tat zur Seite standen.

Ein besonderer Dank gilt meinen Kollegen in der Arbeitsgruppe Essen für das tolle Miteinander, die zahlreichen Anregungen und große Hilfsbereitschaft. Besonders hervorheben möchte ich an dieser Stelle Petra Gnau, die mir vor allem den Anfang meiner Arbeit sehr erleichtert hat und auf deren Unterstützung ich mich auch in schwereren Zeiten immer verlassen konnte. Weiterhin Filipp Bezold, Maximilian Biermeier, Lukas Korf, Dr. Viktoria Reithofer und Dr. Marian Vogt, die mit mir über viele Kaffee- und Mittagspausen alle Aspekte meiner Arbeit und des Lebens diskutiert haben und von denen ich viel lernen konnte. Christin Geil danke ich für die Hilfe im Labor und die netten Gespräche.

Auch außerhalb meiner Gruppe haben mich viele Menschen unterstützt. Annika Brych und Dr. Stefan Brückner danke ich für die stets unterhaltsamen Mittagspausen. Anke Werner danke ich für die unkomplizierte Zusammenarbeit, das tolle Miteinander und ihren Enthusiasmus, der mich stets mit motiviert hat.

Zuletzt möchte ich natürlich meiner Familie danken, meinen Eltern, meiner Schwester, Dominik und Alexander, dass sie immer für mich da sind und mich in Allem unterstützen, was ich tue. Ich liebe euch!

Zusammenfassung

Die Symbiose von Hülsenfrüchten mit Stickstoff-fixierende Bakterien erlaubt diesen Zugang zu elementarem Stickstoff aus der Atmosphäre. Die Interaktion ist hauptsächlich durch die Pflanze kontrolliert. Zusätzlich wird das Infektionsverhalten auf der Seite des symbiontischen Bakteriums *Sinorhizobium meliloti* über cAMP-gesteuerte Transkriptionskontrolle beeinflusst. Hierfür besitzt es eine außerordentliche Anzahl von 28 Adenylyl- oder Guanylylzyklasen. Drei von diesen, stehen im Zusammenhang mit der Unterdrückung sekundärer Infektionen – CyaD1, CyaD2 und CyaK. Jede von ihnen enthält die regulatorische Domäne CHASE2, deren Funktion bisher unbekannt ist. Im Umfeld des *cyaD1* Gens finden sich weitere Gene, deren Überexpression oder Deletion einen Einfluss im regulatorischen Mechanismus gezeigt hat. Eines dieser Gene kodiert für den Transkriptionsfaktor Clr, ein cAMP-Rezeptorprotein-Vertreter (Crp), von dem eine Regulation sowohl durch cAMP als auch cGMP beobachtet wurde. Downstream von CyaD1 wurde außerdem ein Gen für eine CpdA-artige Phosphodiesterase identifiziert. Clr ist das erste bifunktionale Crp-Ortholog und der Hintergrund dieser Besonderheit ist derzeit unklar. Sie ist von besonderem Interesse, da cGMP-Signalling in Bakterien bisher nur wenig erforscht ist. Ein weiterer interessanter Aspekt ist, dass CpdA in der Literatur als 2‘,3‘-cAMP-spezifisch beschrieben wurde und daher unklar ist welche Rolle es innerhalb des Regulons spielt.

Um zum Verständnis der Funktion dieses neuen cAMP- und cGMP-regulierten Prozesses der Sekundärinfektionsunterdrückung und daraus resultierenden Genregulierung beizutragen, wurden in dieser Arbeit die biochemischen und strukturellen Eigenschaften der CyaD1-Lokus-Komponenten untersucht. In Affinitätsmessungen für die Bildung des Clr-Effektor-DNA-Komplexes konnte gezeigt werden, dass in der Anwesenheit von cAMP oder cGMP Target-DNA mit vergleichbarer Affinität an das Protein gebunden wird. In der Kristallstruktur des Komplexes zeigt sich, dass die Bindung beider Nukleotide dieselbe aktive Konformation hervorruft, aber das Effektormolekül in einer etwas anderen Ausrichtung bindet. Die Unterschiede zu anderen Crp-Orthologen sind nicht so groß, dass sie das besondere bifunktionale Verhalten von Clr erklären könnten. Vielmehr ist davon auszugehen, dass eine Veränderung im dynamischen Netzwerk des Proteins dazu führt, dass eine Aktivierung auch durch cGMP möglich wird. Zusätzlich zeigt die Kristallstruktur des Clr · cNMP · DNA-Komplexes, dass der Transkriptionsfaktor direkt mit dem konservierten DNA-Bindemotiv interagiert. HDX-MS-Messungen geben außerdem einen Einblick in die Regionen die in den apo zu holo Übergang involviert sind. Eine phylogenetische Analyse von Crp-Proteinen gibt

Hinweise darauf, dass Clr zu einer neuen, bisher unbekannten Unterklasse dieser Transkriptionsregulatoren gehören könnte, die sich möglicherweise vor allem durch ihre Fähigkeit von cGMP aktiviert zu werden hervorheben.

Die Struktur von CpdA aus *Sinorhizobium meliloti* und die Charakterisierung seiner Aktivität geben neue Einblicke in seine mögliche biologische Funktion. Entgegen vorangegangener Annahmen ist CpdA sehr vielseitig bezüglich seiner Substrate und in der Lage 3‘,5‘-cAMP mit hoher Aktivität zu hydrolysieren. Außerdem zeigt es Aktivität gegenüber 2’,3’-cAMP, 2’,3’-cGMP und 3’,5’-cGMP mit in dieser Reihe abnehmender Effizienz. CpdA ist die erste Phosphodiesterase der Klasse III die 3’,5‘-cAMP hauptsächlich zu 3‘-AMP und 2‘,3‘-cAMP primär zu 2‘-AMP hydrolysiert. Seine Aktivität wird durch die Zugabe von Mangan erhöht, was möglicherweise seinen nativen Kofaktor darstellt (allein oder heteronuklear mit Eisen). Die Kristallstruktur von CpdA ist erst die zweite Klasse III Struktur die bekannt ist und zeigt deutlich ein katalytisches Hydroxid in der aktiven Tasche. Damit gibt es einen wichtigen Hinweis bezüglich des Mechanismus. Dieser läuft vermutlich über einen S_N2-artigen Angriff des zwischen den Metallen koordinierten Hydroxid an das Phosphat des cNMPs ab und nicht wie in *E. coli* UshA beobachtet über einen asymmetrischen Übergangszustand mit einem weiteren Hydroxid-Nukleophil. Zusätzlich zu diesen Beobachtungen gibt diese Arbeit einen ersten bioinformatischen Einblick in die Struktur der CHASE2 Regulationsdomäne in CyaD1D2K und gibt eine allgemeine Einordnung der oben erläuterten strukturellen Aspekte in das cAMP-gesteuerte Regulon in *Sinorhizobium meliloti.*

Summary

Legume crops undergo a symbiotic relationship with nitrogen-fixing bacteria in order to make atmospheric nitrogen accessible to them. The interaction is mainly controlled by the plant, however the symbiotic bacterium *Sinorhizobium meliloti* mediates its infection behaviour using cAMP-mediated transcription regulation. To that end, its genome encodes an extraordinary number of 28 adenylyl or guanylyl cyclases, three of which have been linked with suppression of secondary infection. CyaD1, CyaD2 and CyaK each contain a CHASE2 regulatory domain of unknown function. In the proximity of the gene coding for CyaD1 a few other genes were shown to be involved in the regulatory mechanism via overexpression or deletion mutants. One of those is cAMP receptor protein-like (Crp) transcription factor Clr, that has been shown to be regulated by both cAMP and cGMP. Downstream from CyaD1 a gene coding for a CpdA-like phosphodiesterase was identified. Clr is the first bifunctional Crp ortholog to be reported and the underlying mechanism of this feature remains to be revealed. This is of particular interest as cGMP signalling in bacteria is a largely unexplored field. Additionally, CpdA has been reported to be specific for 2’,3’-cAMP. How it is linked to the rest of the regulon was therefore unclear.

To support the understanding of function of this novel cAMP and cGMP regulated process of secondary infection repression and downstream gene regulation, the biochemical and structural features of the CyaD1 locus components were investigated in this work. Affinity measurements for the formation of the Clr-effector-DNA complex reveal similar target DNA binding strengths in the presence of cAMP and cGMP. The crystal structure of the complex bound to each nucleotide shows that they elicit the same active conformation, but bind the effector molecules in different conformers. The differences from other Crp orthologs are not sufficient to explain the basis of Clr bifunctionality. Instead, a modified dynamic network likely results in the ability to be activated by cGMP as well. The crystal structure of the Clr · cNMP · DNA complex also shows how the transcription factor directly interacts with the conserved DNA binding motif. HDX-MS measurements give indications on the regions involved in the apo to holo transition. The phylogenetic analysis of Crp-like proteins indicates that Clr belongs to an unexplored subclass of these transcription factors, potentially differentiated by their ability to be activated by cGMP.

The structure of CpdA from *S. meliloti* combined with a characterization of its activity gives new insights in its potential biological role. Contrary to previous reports, CpdA is very promiscuous in its substrate utilization and able to degrade 3’,5’-cAMP with high activity. Aside from that, it also is capable to hydrolyse 2’,3’-cAMP, 2’,3’-cGMP and 3’,5’-cGMP with slightly efficiency. It is the first class III phosphodiesterase to mainly produce 3’-AMP or 2’-AMP from the hydrolysis of 3’,5’-cAMP or 2’,3’-cAMP respectively. Its activity is amplified by manganese addition, which is likely its native cofactor (alone or heteronuclear with iron). The crystal structure of CpdA is only the second class III PDE structure to be reported and nicely reveals a catalytic hydroxide within the active site. With this it provides convincing evidence for a catalytic mechanism involving a S_N2-like attack of the bimetallically coordinated hydroxide molecule on the cNMP phosphate, as opposed to the asymmetric transition state involving an additional hydroxide found in *E. coli* UshA. Additionally, this work provides a first bioinformatic analysis of the CHASE2 regulatory domain structure found in CyaD1D2K and gives a general classification of the aforementioned structural aspects in the cAMP-mediated regulon of *Sinorhizobium meliloti.*

Table of contents

Eidesstattliche Erklärung v

Verlauf des wissenschaftlichen Werdegangs vi

Danksagung vii

Zusammenfassung i

Summary iii

1 Introduction 1

1.1 Legume plants establish symbiotic interactions with soil bacteria 1

1.2 Cyclic mononucleotide signalling regulates *S. meliloti* infection 4

1.3 The significance of cyclic GMP is not limited to eukaryotes 5

1.4 cNMP receptor proteins are essential for many bacterial functions 7

1.5 Phosphodiesterases are in charge of the regulation of cNMP levels 10

1.6 Objectives 12

2 Materials 14

2.1 Chemicals 14

2.2 Consumables 16

2.3 Enzymes 16

2.4 Equipment 16

2.5 Oligonucleotides, Vectors and DNA 17

2.5.1 Oligonucleotides 17

2.5.1.1 Primers 17

2.5.1.2 Duplex DNA 18

2.5.2 Vectors 18

2.5.2.3 pET36b(+) vector 18

2.5.2.4 pET28a(+) vector 19

2.5.2.5 pET His_6 Mocr TEV LIC Cloning Vector 20

2.5.2.6 pET His_6 GST TEV LIC Cloning Vector 21

2.5.2.6 pET His$_6$ MBP TEV LIC Cloning Vector 22
2.6 Organisms 23
2.6.1 *Escherichia coli* DH5α 23
2.6.2 *Escherichia coli* BL21 (DE3) Gold 23
2.6.3 *Escherichia coli* T7 Shuffle Express 24
2.7 Software 24
2.7.1 Structure Visualisation 24
3 Methods 25
3.1 Molecular Biology 25
3.1.1 Polymerase Chain Reaction (PCR) 25
3.1.2 Preparation of plasmid DNA 25
3.1.3 PCR Purification 26
3.1.4 DNA Quantification 26
3.1.5 Restriction Digest 26
3.1.6 LIC Cloning 27
3.1.7 Ligation 28
3.1.8 Agarose Gel Electrophoresis 28
3.1.9 RbCl-Competent Cells 28
3.1.10 Transformation 29
3.1.11 Colony PCR 29
3.1.12 Sequencing 30
3.1.13 Hybridisation of DNA Oligonucleotides 30
3.2 Protein Biochemistry 30
3.2.1 Heterologous Overproduction in *E. coli* 30
3.2.2 Cell Lysis 31
3.2.3 Immobilized Metal Affinity Chromatography (IMAC) 31
3.2.4 Size Exclusion Chromatography (SEC) 32
3.2.5 Buffer Exchange 33

3.2.6 SDS-PAGE 34
3.2.7 Protein Quantification 35
3.2.8 Thermal Shift Assay 35
3.2.9 Western Blot 36
3.2.10 Fluorescence Anisotropy Assays 37
3.2.11 Isothermal Titration Calorimetry (ITC) 38
3.2.12 H/D-Exchange Mass Spectrometry (HDX-MS) 41
3.2.13 Nuclear Magnetic Resonance (NMR) 43
3.2.14 HLPC-Nucleotide Assays 43
3.3 Protein Crystallography 45
3.3.1 Protein Crystallisation 45
3.3.2 Soaking of Protein Crystals 47
3.3.3 X-ray Diffraction and Data Collection 47
3.3.4 Processing of Diffraction Data 49
3.3.5 Phasing 50
3.3.6 Structure Refinement 51
4 Results 53
4.1 *Sm*Clr: a bifunctional cNMP receptor protein 53
4.1.1 Expression and purification of *Sm*Clr 53
4.1.2 Clr affinity for cGMP and cAMP levels out in the presence of DNA 55
4.1.3 Crystallisation and structure determination cAMP-activated *Sm*Clr bound to DNA 59
4.1.4 Crystallisation and structure determination cGMP-activated *Sm*Clr bound to DNA 62
4.1.5 General features of the nucleotide-induced Clr-DNA complex 64
4.1.6 Both cAMP and cGMP are able to shift Clr to its active conformation 67
4.1.7 Clr promotor specificity can be explained by its nucleotide-bound structures 71
4.2 *Sm*CpdA: the misunderstood phosphodiesterase 74

4.2.1 Expression and purification of *Sm*CpdA 74

4.2.2 *Sm*CpdA is a manganese-dependent cyclic nucleotide phosphodiesterase 75

4.2.3 Crystallisation and structure determination of *Sm*CpdA 82

4.2.4 *Sm*CpdA is a Class III phosphodiesterase 85

4.2.5 *Sm*CpdA crystals seem to be incompatible with substrate binding 88

4.3 CyaD1 CHASE2 regulatory domain 90

5 Discussion 96

5.1 The CyaD1 locus regulates host infection 96

5.1 Clr is the first bifunctional cNMP receptor protein 97

5.1.1 The active site and binding mode of Crp-like proteins are highly conserved 99

5.1.2 The activation of Crp-Fnr like proteins can occur via direct and indirect interaction with the hinge 101

5.1.3 Clr binding behaviour differs from the one found in other Crp 104

5.1.4 The transformation from the open to closed form is characterized by decreased flexibility 105

5.1.5 Receptor-DNA interaction happens at a conserved motif 109

5.1.6 The RNA polymerase is recruited by Clr to regulate gene expression 111

5.2 CpdA might be a Clr antagonist 114

5.2.1 CpdA belongs to the novel class III of phosphodiesterases 114

5.2.2 *Sinorhizobium meliloti* contains 14 phosphodiesterases of varying function 116

5.2.3 CpdA is a metal-dependent cNMP phosphodiesterase 117

5.2.4 Class III phosphodiesterases are functionally diverse 119

6 Outlook 124

7 Bibliography 126

8 Appendix 154

8.1 Testexpressions 154

Test expression of Clr dN from S. americanum, S. Sp. A49 and S. saheli 154

Test expression of native phosphodiesterase constructs and Expression of pQE9 His_6 *Sm*CpdA (native) ... 154

Test expression of solubility tag fused phosphodiesterase constructs ... 155

8.2 Clr-DNA affinity measurement via fluorescence anisotropy ... 156

8.3 HDX-MS analysis of Clr substrate binding ... 159

8.4 NMR analysis of potential *Sm*CpdA substrates ... 191

8.5 CpdA kinetics in ITC measurements ... 198

8.5 CHASE2 homologs ... 199

1 Introduction

1.1 Legume plants establish symbiotic interactions with soil bacteria

Nitrogen availability is an essential factor for plant growth. While it is the most abundant element in our atmosphere, it is inaccessible to crops in its elemental form (Ibáñez et al. 2017). Rather, roots are able to absorb inorganic nitrogen compounds like nitrate and ammonium from soil (Xu et al. 2012). The element is an irreplaceable component of nucleic acids, proteins, small organic compounds like ATP, as well as chlorophyll, which plants utilize to absorb sunlight energy. Inorganic nitrogen is provided by fertilization in commercial agriculture leading to a whole string of environmental issues. Legume crops and actinorhizal plants however have developed an alternate method of nitrogen uptake. They are able to engage in a symbiotic interaction with nitrogen-fixing rhizobia or actinomycetes (Ibáñez et al. 2017). While the bacteria fix atmospheric nitrogen for the plant, they receive carbon compounds in return and dwell in the protected environment of the root. This is advantageous not only for the legume plant itself, as the nitrogen content of the soil increases during growth, benefiting subsequent crops as well (Murray et al. 2017). Inversely, nitrogen fertilizers inhibit both symbiosis establishment and nitrogen fixation, making overfertilized soils unfit for legume cultivation (Streeter and Wong 1988). Nitrogen fixation involves high energetic costs compared to nitrogen compound uptake, meaning legumes likely only have an advantage over other plants in nitrogen-depleted soil, in response to which the symbiosis probably evolved in the first place (Gibson and Harper 1985).

The different variations and stages of rhizobia interaction are illustrated in Figure 1. Rhizobia can be internalized by the plant via two distinct mechanisms, either formation of infection threads or intercellular invasion (Ibáñez et al. 2017). The former has been characterized more extensively and takes place in soybean, alfalfa, pea, bean and vetch (Sprent 2008). The two model organisms from this group are *Medicago truncatula* and *Lotus japonicus* (Oldroyd 2001). A regulated uptake is important, as plants constantly need to discriminate between beneficial and potentially harmful microorganisms. This is conducted by the exchange of diffusible signalling molecules between the symbiotic partners (Gamas et al. 2017). The infection initiation is characterized by an interplay between the symbiotic partners, but is ultimately under the control of the host (Murray 2011). In the presence of compatible rhizobia, plants will initiate an infection program that facilitates the internalization of the bacteria in specialized compartments within the root, called nodules (Via et al. 2016). Legume roots extrudate flavonoids and other related compounds into the soil to signal their need for nitrogen to compatible rhizobia. In nitrogen-fixing bacteria like *Sinorhizobium meliloti*, the transcription of nodulation genes is elevated in response to the plant signal, as the flavonoids are bound to NodD, a transcriptional regulator of the LysR family (Jones et al. 2007). The process results in the synthesis of so-called nodulation factors (Nod factors) (Ibáñez et al. 2017). These factors are secreted lipochitooligosaccharides, that in turn can be recognized by membrane LysM receptor-like kinases (LysM-RLKs) in the root epidermal cells (Via et al. 2016). The chemical modification of Nod factors varies depending on the rhizobial species and acts as a main determinant of host specificity (Limpens et al. 2015). They consist of a backbone of β-1,4-linked *N*-acetyl-D-glucosamine residues, which can differ in number (Jones et al. 2007). Nodulation factors have also been found to suppress host immune response (Gourion et al. 2015). Their perception triggers a complex signal transduction pathway resulting in the synthesis of phytohormone cytokine. In addition to that, the rhizobial exopolysaccharide transmembrane receptor kinase EPR3 is expressed, which is able to differentiate bacteria by their exopolysaccharides (Ibáñez et al. 2017). Cytokine production triggers the formation of nodules, compartments within the root optimized for symbiosis, by causing cortical cell division (Gamas et al. 2017). Additionally, a series of pre-infection events are triggered, including curling at the root hair tip, which is followed by bacterial invasion into the resulting infection pocket. At this point the root hair ceases to grow and a tubular invagination of cell wall and membrane extends from the pocket through the plant tissue until it reaches the nascent nodule. Here the bacteria are released and begin nitrogen fixation (Murray 2011). While the symbiosis is established, a complex regulation pathway takes place in both the plant and the bacterium. A comprehensive overview over the invasion of *Medicago* plants by *Sinorhizobia* was given by Jones et al. 2007.

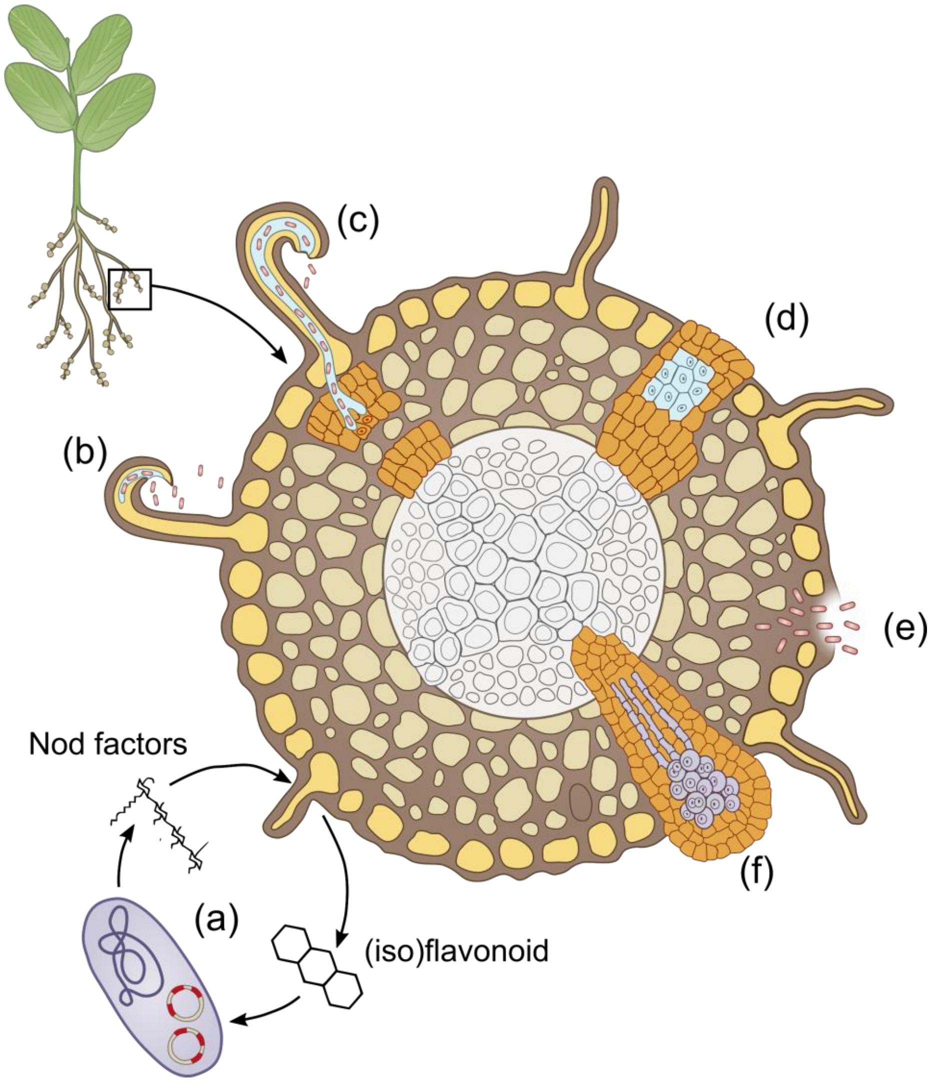


Figure 1: Schematic overview over the initiation of plant symbiosis with nitrogen-fixing bacteria. (a) Nitrogen-depleted plants secrete (iso)flavonoids, where they are perceived by compatible bacteria. Binding of these compounds to the transcription factor NodD initiates the synthesis and modification of specific oligosaccharides, called nodulation factors, which in turn can be sensed by the plant. (b) Root hairs, in which LysM receptor-like kinases have detected the presence of nodulation factors, cease growing and start curling inward to form characteristic “shepherds hook”-like structures with an infection pocket filled with bacteria at their centre. The signal also initiates the formation of nodules, partitions optimized for symbiosis within the root. (c, d) Outgoing from the infection pocket, the membrane and cell wall start growing inward to form a so-called infection thread. In this structure bacteria travel inwards towards the nascent nodule, where they are internalized and start fixing nitrogen. (e) An alternate mechanism of bacterium uptake, intercellular invasion is also possible (Ibáñez et al. 2017). (f) The plant provides the bacteria with energy via nodule vascular tissue. The image was adapted from Le Roux et al. 2017.

1.2 Cyclic mononucleotide signalling regulates *S. meliloti* infection

The genome of *Sinorhizobium meliloti* codes for an astonishing number of 28 class III adenylyl or guanylyl cyclases (Krol et al. 2016). Most proteins from this highly abundant class of cyclases are multi-domain proteins. They contain the catalytic cyclase homology domain (CHD) in combination with a regulatory domain and can be either soluble or membrane-bound (Linder 2006). The diversity of regulatory domains found in conjunction with them, indicates a broad spectrum of signals that can be sensed at different times (Linder 2006; Dass et al. 2008). Class III cyclases must form dimers to be active. Two catalytic centres are formed at the dimer interface of bacterial orthologs, while mammalian representatives form heterodimers with only one active site. Two metal cofactors, generally Mg^{2+} or Mn^{2+}, are required for the catalytic activity following a S_N2 mechanisms (Eckstein et al. 1981). Often multiple isoforms of class III cyclases are found in bacteria to respond to several different stimuli, as is the case in *S. meliloti* and in *Mycobacterium tuberculosis*, which encodes for 16 cyclases (Knapp and McDonough 2014). This is in striking contrast to *Escherichia coli*, *Corynebacterium glutamicum* or *Candida albicans*, in which only a single copy of a cyclase present (Dass et al. 2008; Knapp and McDonough 2014).

The cyclases in *S. meliloti* carry a number of accessory domains. Overexpression of *C*-terminally membrane-bound CyaJ resulted in high levels of cAMP in the culture supernatant, decreased swimming motility and increased succinoglycan synthesis, one of the exopolysaccharides sensed during plant invasion. Overexpression of *N*-terminally membrane-bound CyaB and CyaG1, for which no accessory domain has been annotated so far, resulted in the same phenotypes, albeit to a lesser extent (Krol et al. 2016). Three of the *S. meliloti* cyclases are *N*-terminally fused to CHASE2 domains, a regulatory domain found in cyclases as well as histidine kinases, serine/threonine kinases, predicted diguanylate cyclases and phosphodiesterases (Linder 2006). Two class III cyclases with the same architecture have been linked with sensing high osmotic pressure in *Myxococcus xanthus* (Kimura et al. 2002; Kimura et al. 2005). In *S. meliloti*, NsrA, an outer membrane protein involved in plant signal perception, has been suggested as potential interaction partner of CHASE2. Inactivation resulted in an abolishment of downstream gene regulation (Garnerone et al. 2018). Finding the interaction partner of the CHASE2 domain is of particular importance, as the three cyclases CyaD1 (*Sm*c02176), CyaD2 (*Sm*c04307) and CyaK (*Sm*b20776), that each contain said domain, have been suggested to be involved in the control of secondary infection (Zou et al. 2017). Enhanced cAMP levels in the organism stimulate succinoglycan production and represses swimming motility, both aspects important in the establishment of symbiosis (Krol et al. 2016). The extracellular CHASE2 domain is connected to the intracellular CHD domain via three trans-domain helices, so it is likely that the activating signal is given extracellularly as well (Zhulin et al. 2003). The gene locus of CyaD1 also contains a Crp-like transcription regulator Clr, a phosphodiesterase CpdA, as well as two other genes of unknown function (*Sm*c02177 and *Sm*c02178). Inactivation of the three cyclases, Clr or *Sm*c02178 resulted in a hyper-infection phenotype. Individual cyclase activation is not sufficient to cause this outcome (Tian et al. 2012). LacZ reporter fusion revealed, that the three cyclases are active in different stages of plant invasion *in vivo*. CyaK was active in both young and mature nodules, whereas CyaD1 and CyaD2 activity was only detected in the earlier stages (Garnerone et al. 2018). In *Pseudomonas aeruginosa* it has been suggested that cyclases might actually not only be regulated at a transcriptional level and by their regulatory domains, but also via posttranslational modifications (Nolan et al. 2012; Inclan et al. 2011). Another interesting aspect of bacterial adenylyl or guanylyl cyclases is their use as toxins in some organisms. As cyclic nucleotide signalling is so universal, bacteria like *Pseudomonas aeruginosa* and *Bacillus anthracis* secrete adenylyl cyclases to elevate the host cAMP levels and weaken the immune response. They also employ strategies to tamper with the host adenylyl cyclase regulation, as was observed for *Vibrio cholerae* and *E. coli* toxins (Zaveri and Visweswariah 2013).

1.3 The significance of cyclic GMP is not limited to eukaryotes

In higher eukaryotes high levels of cyclic GMP haven been found in brain, kidney and cardiovascular tissue, to name a few. Its main function seems to be the regulation of ion transport. As such, it is involved in the regulation of blood flow, phototransduction in the eye and potentially neural development and is an integral part of mammalian second messenger networks (Tsai and Kass 2009). Signals that elicit a cGMP-mediated regulation include light, nitric oxide and hormones (Beavo and Brunton 2002). As cGMP signalling seems to be impaired during heart failure, it has even been considered a drug target (Blanton 2020). But cyclic GMP is not only important in higher eukaryotes. *Plasmodium falciparum*, a parasite causing malaria, uses cGMP perception to adapt its morphology during its biphasic life cycle in mosquitos and humans (Brochet et al. 2020; Brochet 2018).

The presence of cGMP in *E. coli* was found quite early. But as its levels were two orders of magnitude below that of cAMP and no specific receptor protein was identified, it was concluded that the nucleotide had to be a by-product of the bacterial adenylyl cyclase *in vivo* (Shibuya 1977; Tao and Huberman 1970). The hypothesis was brought into question when a paper by Veltman et al. 2004 reported the finding of a cGMP signalling pathway in cellular slime mold

Dictyostelium discoideum. In 2008 Rauch et al. published the crystal structure of Cya2 from cyanobacterium *Synechocystis PCC6803*, a cyclase that shows guanylyl specificity almost exclusively, which made it clear that cGMP signalling is not limited to eukaryotes. For mammalian class III cyclases substrate-selecting residues have been identified and the specificity towards ATP or GTP can be predicted quite well. Bacterial cyclases however seem to not follow this pattern, so that no prediction is possible (Linder 2010). In a *Rhodospirillum centenum* strain, which contained a GC deletion, deficient cyst formation was observed. The effects could be reversed when cGMP was supplemented (Marden et al. 2011) Fellow cyst-forming Alphaproteobacteria *Azospirillum brasilense* and *Sinorhizobium meliloti* secrete cGMP as well as cAMP during cyst formation too (Gomelsky 2011). Figure 2 illustrates the different targets of cGMP signalling in eukaryotes and bacteria. In general, the effects of cGMP as second messenger on bacteria still remain poorly understood.

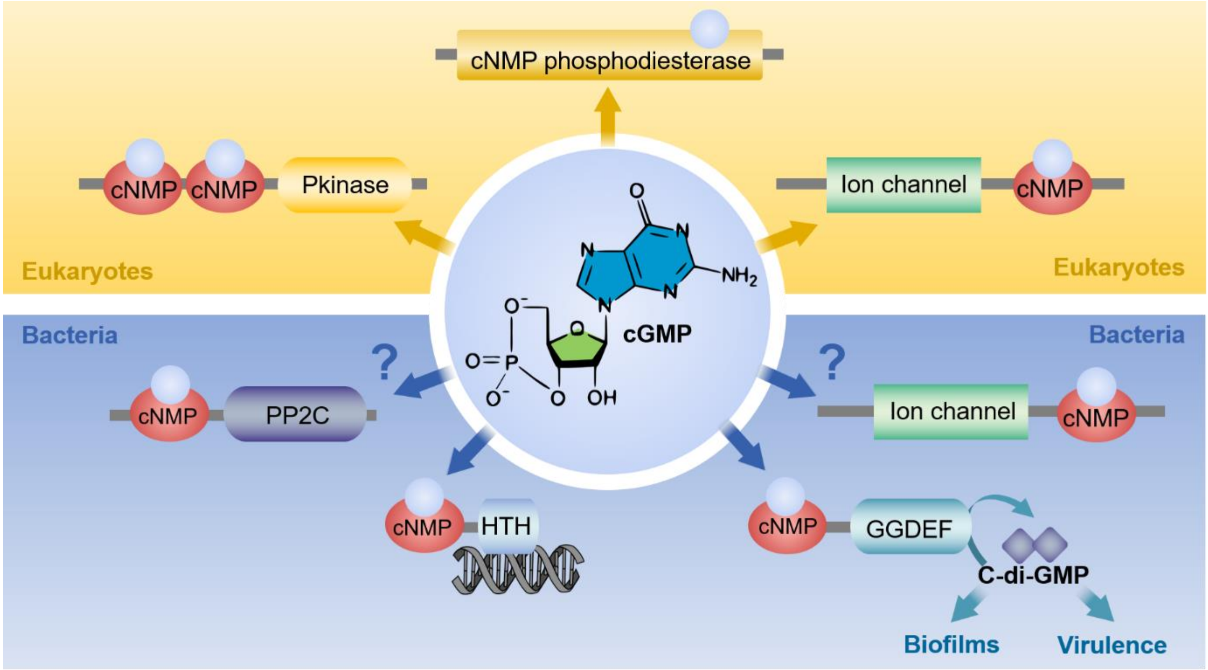

cNMP phosphodiesterase
cNMP
cNMP
Pkinase
Ion channel
cNMP
Eukaryotes
Eukaryotes
O
N
N
N
N
NH2
O⁻
O=P
O⁻
O
O
OH
cGMP
Bacteria
Bacteria
?
?
cNMP
PP2C
Ion channel
cNMP
cNMP
HTH
cNMP
GGDEF
C-di-GMP
Biofilms
Virulence

Figure 2: Cyclic nucleotide signalling in different areas of life. In eukaryotes cGMP mainly interacts with protein kinases, cyclic nucleotide-gated ion channels and phosphodiesterases. In bacteria a cGMP-regulated transcription factor has been identified in *Rhodospirillum centenum (Marden et al. 2011)*. Additionally, c-di-GMP synthesis is linked to cGMP levels via cNMP–GGDEF domain fusion protein XC_0249 from *Xanthomonas campestris*. Other potential targets are the cNMP-regulated ion channel BBta_5447 from *Bradyrhizobium sp.* and the protein serine phosphatase Cagg_2419 from *Chloroflexus aggregans*. The image was adapted from Gomelsky and Galperin 2013.

1.4 cNMP receptor proteins are essential for many bacterial functions

Bacteria need to respond and adapt to various intra- and extracellular signals in different stages of their life circle (Körner et al. 2003). The regulation of gene transcription to RNA is a vital process and conducted by transcription factors, that activate or repress transcription at specific binding sequences in response to cellular signals like small molecules or post-translational modifications. As these proteins often regulate the transcription of other factors as well, they form complex interlinked regulatory networks. Transcription in bacteria is catalysed by the RNA polymerase, which is recruited to promotors by a sigma factor or other transcription factor (Busby and Ebright 1999). Sigma factors are considered responsible for baseline transcription of all genes (Aravind et al. 2005). The affinities of other transcription factors to their target DNA were often found to be dependent on the interaction with small effector molecules like cyclic nucleotides. One of the best-studied bacterial transcription factors is the cAMP receptor protein (Crp) CAP from *E. coli*. The homo-dimeric protein consisting of a cNMP effector binding and a helix-turn-helix DNA binding domain shifts the bacterial metabolisms towards other carbon source like lactose or maltose when glucose is scarce (Weber and Steitz 1987). To that end, cAMP production in the bacterium is linked via cyclase phosphorylation to the phosphoenol pyruvate sugar-phosphotransfer system (PTS) and hundreds of genes are regulated in response to the Crp (Soberón-Chávez et al. 2017; Youn et al. 2006). In addition to metabolic adaption, CAP also affects amino acid metabolism, transport processes, protein folding, as well as toxin production and pilus synthesis (Körner et al. 2003; Bodero and Munson 2009). CAP has been found to have the largest non-sigma factor regulon in *E. coli* and interacts with multiple other transcription factors of the Fnr, Fur or ArcA family to name a few (Frendorf et al. 2019). Crp homologs are present in a multitude of bacterial species and in any case they regulate crucial functions of that particular organism without being part of their primary metabolism (Soberón-Chávez et al. 2017). The Crp-like ortholog of *Vibrio cholerae* has an exceptionally high identity of 95% with *E. coli* CAP and regulates metabolic adaption as well (Skorupski and Taylor 1997). However, it has additional functions as it activates genes involved with chitin degradation, competence, biofilm formation, motility, quorum sensing and other virulence-associated traits (Blokesch 2012; Fong and Yildiz 2008). For *Vibrio cholerae* the regulation is of particular importance, as it switches from an aquatic lifestyle to cause disease in humans. The transport of toxins in response to host colonisation is regulated by a Crp (Manneh-Roussel et al. 2018). Another well-investigated organism using Crp-like proteins for the regulation of virulence factors is *Pseudomonas aeruginosa*. Here the virulence regulation factor Vrf is involved in complex quorum sensing responses, twitching motility, but contrary to the general rule this Crp ortholog also affects global gene expression (Croda-García et al. 2011; Suh et al. 2002; Nolan et al. 2012). Notwithstanding these differences Vfr is able to partially complement a *E. coli* CRP mutant. This illustrates that the diversity of Crp-dependent signalling is not solely due to differences on the protein level, but also determined by the distribution of its cognate DNA binding sequence (West et al. 1994). In Actinobacteria GlxR from *C. glutamicum* regulates genes affecting amino acid synthesis and the *Mycobacterium tuberculosis* Crp is involved in the modulation of pathogenicity (Akhter et al. 2008). In *C. glutamicum* over 200 genes are regulated by GlxR and deletion mutants have severe growth defects (Schulte et al. 2017a; Townsend et al. 2014). Aside from *E. coli* Cap, Crp^{Mt} (Rv3676) has received the most scrutiny so far. Knockout mutants are impaired in growth in the macrophage *in vivo* and show attenuated virulence in mice (Rickman et al. 2005). The presence of cAMP increases Crp^{Mt} DNA affinity to a lesser degree than *E. coli* CAP (merely factor 2.5, instead of several thousand-fold) (Agarwal et al. 2006; Takahashi et al. 1989). This is likely an adaptation to the high levels of cAMP found in *Mycobacteria* from the Crp-dependent transcription in *E. coli* to a mere Crp-regulated one (Zaveri and Visweswariah 2013). For *M. tuberculosis* whiB, another transcriptional regulator, both negative and positive control has been observed (Knapp and McDonough 2014). There are two Crp^{Mt} binding sites upstream of the gene. Occupation of the first binding site activates transcription, while the binding to both inhibits it (Stapleton et al. 2010; Alberto et al. 2012). This means the gene is activated above a certain cAMP level, but then inhibited once the concentration increases further, a great example of regulatory fine tuning. Notably, Crp^{Mt} is unable to substitute for CAP in *E. coli*, as it fails to interact properly with the RNAP (Knapp and McDonough 2014).

Cyclic AMP has always been considered the only effector molecule of Crp-like proteins. Only recently, exceptions to this rule have been reported. The transcription factor CgrA from *Rhodospirillum centenum* is activated by cGMP instead of cAMP and in *S. meliloti*, the first Crp-like protein to be activated by both cAMP and cGMP was found (Tian et al. 2012; Roychowdhury et al. 2015). The bifunctional transcription factor from *S. meliloti* was termed Clr (Tian et al. 2012). Deletion of Clr results in a hyperinfection phenotype similar to that of the CyaD1, CyaD1, CyaK triple mutant. The phenotype is characterized by increased infection thread formation by a factor of three or four (Figure 3). The activation of the cAMP regulation cascade in endosymbiotic bacteria within the nodule seems to reduce root sensitivity to nodulation factors. As a results secondary infection is repressed (Sorroche et al. 2019). Clr is not involved in catabolite repression by succinate, but deletion mutants revealed that it plays a role in succinoglycan production and swimming motility regulation (Tian et al. 2012; Krol et al. 2016). Many of the genes that were affected by high cAMP levels, are also affected by changes in osmolarity. Repression was observed in genes linked to respiration, carbon and amino acid metabolism (Krol et al. 2016). The exogenous addition of cGMP results in a regulation profile very similar to that of cAMP. This highlights the ability of Clr to interact with both in a similar manner (Krol et al. 2016). Generally it seems that Clr is not limited to CyaD1D2K, but integrates different environmental signals via varying cyclases (Zou et al. 2017). 2’,3’-cNMP are unable to activate Clr however. The Clr-binding sequence has been reported as HGTYHC(N_4)GRWACW according to the IUPAC degenerate base nomenclature (Cornish-Bowden 1985; Zou et al. 2017). Transcriptomic studies on Clr have been conducted in great detail by Krol et al. 2016 and Zou et al. 2017, identifying a myriad of genes that are affected by Clr-regulation. Contrary to Crp^{Mt}, Clr is able to interact with the RNAP from *E. coli (Krol et al. 2016).*

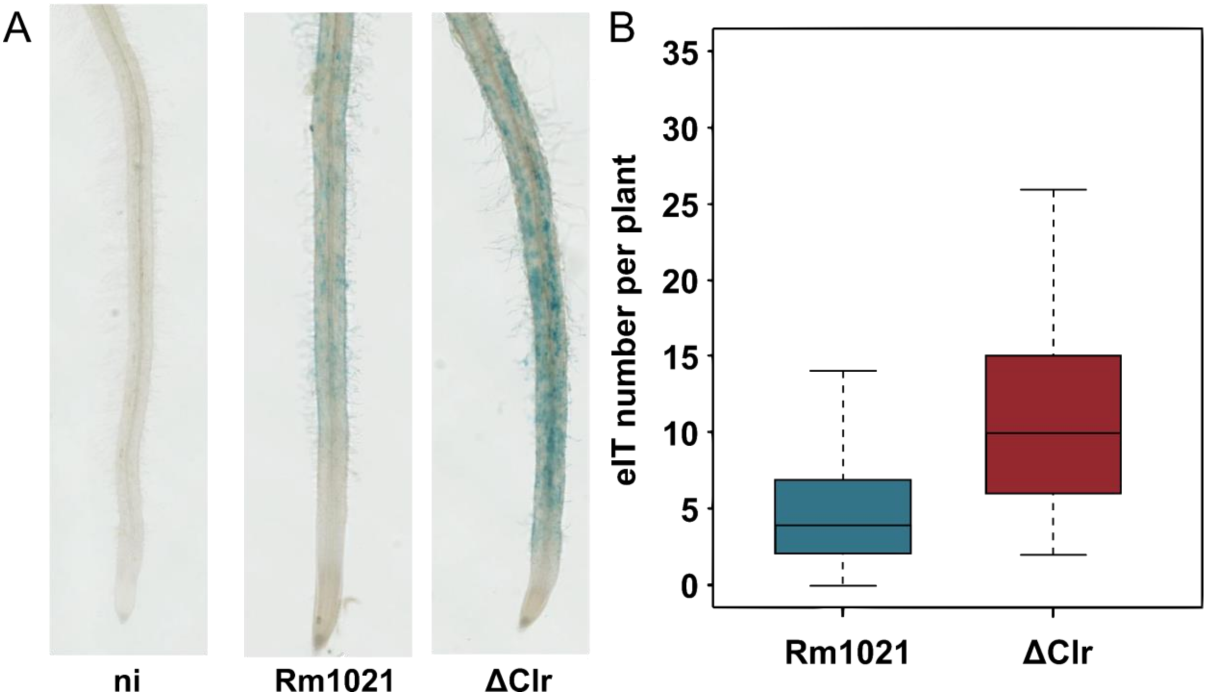


Figure 3: *clr* mutants result in a hyperinfection phenotype with largely increased numbers of infection threads. (A) Comparing noninocculated (ni) *Medicago truncatula* proximal roots, with those inoculated with either *Sinorhizobium meliloti* Rm1021 (wild type) or a *clr* mutant, the hyperinfection caused by the lack of Clr-dependent regulation becomes apparent. (B) Not only the number of incorporated rhizobia is increased, but the number of epidermal infection threads (eIT) per plant is multiplied. Some of these threads are disrupted due to the plant immune response. The image was adapted from Sorroche et al. 2019.

1.5 Phosphodiesterases are in charge of the regulation of cNMP levels

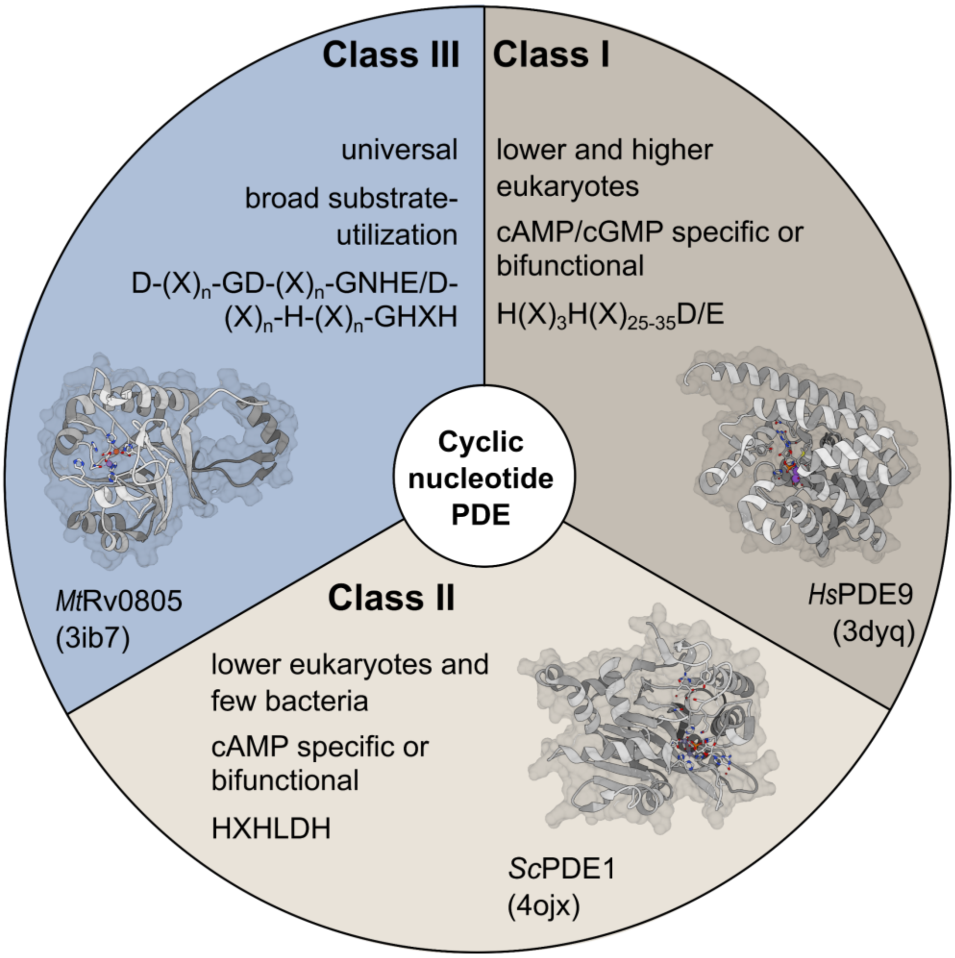

Class III
universal
broad substrate-utilization
D-(X)n-GD-(X)n-GNHE/D-(X)n-H-(X)n-GHXH
*Mt*Rv0805
(3ib7)
Class I
lower and higher eukaryotes
cAMP/cGMP specific or bifunctional
H(X)3H(X)25-35D/E
*Hs*PDE9
(3dyq)
Cyclic nucleotide PDE
Class II
lower eukaryotes and few bacteria
cAMP specific or bifunctional
HXHLDH
*Sc*PDE1
(4ojx)

Figure 4: To fulfil the highly important function of phosphodiester degradation three distinct protein classes developed independently. Class I is prevalent solely in eukaryotes and orthologs can be either nucleotide-specific or able to hydrolyse both cAMP and cGMP. Many members of the group are structurally resolved. The phosphodiesterase 9 from *Homo sapiens* in complex with cGMP is pictured as an example. Class II PDE are a rather narrow group of phosphodiesterases and only one member, phosphodiesterase 1 from yeast, has been structurally resolved. The same holds true for class III, where *M. tuberculosis* Rv0805 is the first structure to be reported. While the class can be found in archaea, bacteria and eukaryotes, it is still under-explored. A striking feature of class III PDE as a group is the varying substrate specificity of its members. PDB identifiers are indicated in the parenthesis. The image was inspired by Matange 2015a.

Phosphoesters are ubiquitous compounds in all domains of life. Their high solubility in water, the negative charge provided by the three ionizable groups and their high stability towards hydrolysis under physiological conditions make them remarkably versatile and it is no coincidence that genetic information is stored using phosphoesters (Hunter 2012). The high stability of these compounds makes degradation an issue however. Phosphoesterases, enzymes catalysing the hydrolysis of the phosphoester bond, have evolved independently several times during evolution underlining the importance of their function for biological systems (Matange et al. 2015b). For the hydrolysis of cyclic nucleotides, which are universal signalling molecules, three distinct phosphodiesterase classes have evolved (Matange 2015a). The features of the three classes are summarized in Figure 4. Class I phosphodiesterases (Pfam: Pf00233) are specific to eukaryotes and can be nucleotide-specific in regards to cAMP and cGMP or bifunctional (Richter 2002). Class II PDE are a rather small group of proteins found primarily in lower eukaryotes like yeast, but can be found in a few bacteria as well. An interesting feature found in several orthologs is the presence of membrane targeting or secretion signals (Callahan et al. 1995). The third class is the most widespread and present in bacteria, archaea and eukaryotes. It is therefore surprising, that class III phosphodiesterases were only formally defined by Richter in 2002 using their flagship member CpdA from *Escherichia coli*. The group is a member of the calcineurin-like phosphoesterase superfamily and phylogenetically related to purple acid phosphatases (Matange 2015a). Occasionally, bacteria were found to code for more than just one CpdA ortholog (Kimura et al. 2009). CpdA-like proteins can have varying degrees of product specificity and are closely related to the CpdB subclass, which contains specific 2’,3’-cyclic nucleotide phosphodiesterases (Matange et al. 2015b). Unlike CpdA homologs, CpdB-like proteins may also possess 3’-nucleotidase activity (Anraku 1964). The role of 2’,3’-cyclic nucleotide is largely unexplored. Suggestion like an involvement in RNA degradation and repurposing as carbon and energy source have been made, but experimental evidence is lacking. An indication might be that CpdB expression is increased when bacteria are grown on poor carbon sources (Trülzsch et al. 2001). The regulation of 3’,5’-cyclic nucleotide levels on the other hand has been shown to be highly important, as a high level of Crp activity was found lethal in *E. coli* (Youn et al. 2006).

Class III phosphodiesterases are essential to the control of intracellular nucleotide messenger levels and have been linked to the Crp regulon in multiple organisms. In *E. coli* for example overexpression of CpdA resulted in a modified CAP-dependent transcription pattern, in *H. influenzae* competence was strongly affected by a deletion and in *P. aeruginosa* twitching motility and virulence are CpdA-regulated (Imamura et al. 1996; Macfadyen et al. 1998; Fuchs et al. 2010). The level of class III PDE activity seems to have adapted to the hosts need. While the deletion of CpdA from *E. coli* increases intracellular cAMP levels by factor two to four, a similar deletion mutant of *Serratia marcescens* results in cAMP levels up to 50 fold (Imamura et al. 1996; Kalivoda et al. 2013). An interesting case is Rv0805 from *M. tuberculosis*, the only PDE to be identified in this organism to date (Zaveri and Visweswariah 2013). Overexpression of this CpdA ortholog seems to have little to no effect on the Crp signalling network, so the connection between Crp and CpdA might have been lost during evolution (Matange et al. 2013). Additionally, intracellular cAMP levels were reduced by mere 30% in this context (Matange et al. 2013). This is not surprising, as Rv0805 shows significantly higher activity towards 2',3'-cAMP than for 3',5'-cAMP (Podobnik et al. 2009). The *C*-terminal part of Rv0805 has been shown to have an additional nucleotide-independent function with the modulation of cell wall permeability (Podobnik et al. 2009). Additionally, the activity of Rv0805 seems to be downregulated in the full length protein compared to a *C*-terminally truncated version, meaning the *C*-terminus likely has a regulatory function as well (Matange et al. 2014). For CpdA from *S. meliloti* (SpdA in some publications) only activity towards 2',3'-cAMP was reported. They also observed a lowered level of Clr-dependent transcription after exogenous addition of 2',3'-cAMP, particularly in a ΔCpdA background and concluded that Clr might be inhibited by the nucleotide (Mathieu-Demazière et al. 2013).

1.6 Objectives

Their ability to enrich soils with inorganic nitrogen compound makes legume crops an important factor for sustainability in agriculture. They undergo a symbiotic relationship with nitrogen-fixing bacteria like *Sinorhizobium meliloti* in exchange for carbon-based nutrients and require no artificial nitrogen-based fertilization. The reduction in nitrogen-fertilizers also reduces the emission of greenhouse gases (Stagnari et al. 2017). But the effect is not limited to the legumes themselves as they leave the soil nutrient enriched and are already used in crop rotations for that reason (Ibáñez et al. 2017; Graham and Vance 2003). In this context, understanding the well-regulated interaction between plant host and nitrogen-fixing microbe is highly important for potential improvements via genetic engineering (Wang et al. 2012).

On a biochemical level, the cyclic nucleotide signalling in *S. meliloti* has garnered a lot of attention in recent years due to its high number in adenylyl or guanylyl cyclases. Particularly the cyclases involved in the control of secondary infection, namely CyaD1, CyD1 and CyaK, have been the focus of research. The presence of a CHASE2 domain in all three orthologs, a regulatory motif of unknown function, which has been found in many A/G cyclases and protein kinases, increases the informational value that can be gained even further. Additionally, the Crp-like transcription factor Clr in the CyaD1 locus is the first bifunctional ortholog to be reported. Recently Crp-members responding to c-di-GMP or cGMP have been described (Leduc and Roberts 2009; Roychowdhury et al. 2015). While the protein class has been investigated for some time now, its remarkable flexibility still leaves a lot to be discovered. The structural and functional characterization of the bifunctional transcription factor Clr was identified as worthwhile target for this reason.

Only one class III phosphodiesterase is structurally resolved to date. Additionally, Rv0805 from *Mycobacterium tuberculosis* seems to be more exception than norm. It possesses both cNMP-dependent and independent functions and a higher activity for 2’,3’-cyclic NMP isomers than their 3’,5’-counterparts (Matange 2015a). The locus of CyaD1 contains a CpdA-like protein, that was initially thought to control 3’,5’-cNMP levels as Clr antagonist. Subsequent reports of a 2’,3’-specificity led to confusion about the actual function of the hydrolase (Mathieu-Demazière et al. 2013). Shedding light on function and structure of *Sm*CpdA can contribute to understanding its function in a biological context and provide further information about the features of class III phosphodiesterases as a whole.

2 Materials

2.1 Chemicals

Name	Supplier
Acetic acid	*VWR*
Acrylamide/bis-acrylamide	*Carl Roth*
2’,3’-Adenosine monophosphate, mixed isomers	*Merck*
3’-Adenosine monophosphate (3’-AMP)	*Merck*
5’-Adenosine monophosphate (5’-AMP)	*Carbolution Chemicals*
Agar	*Carl Roth*
Agarose	*Thermo Fisher Scientific*
Ammonia	*VWR*
Ammonium acetate	*Merck*
Ammonium dihydrogen phosphate	*Fluka*
Ammonium peroxodisulfate (APS)	*Merck*
Ammonium sulfate	*Fluka*
Ampicillin, Sodium salt	*Carl Roth*
Angstrom Additive Crystallisation Screen	*Molecular Dimensions*
2-(Bis(2-hydroxyethyl)amino)acetic acid (BICINE)	*Merck*
1,3-bis(tris(hydroxymethyl)methylamino)propane (Bis-Tris propane)	*Merck*
Bovine serum albumin (BSA)	*Merck*
Bradford reagent	*Biorad*
Bromophenol blue	*Carl Roth*
1,4-Butanediol	*Merck*
1-Butanol	*Merck*
Calcium chloride	*Fluka*
Classics Suite Crystallisation Screen	*Nextal Biotech*
Cobalt chloride	*Fluka*
CutSmart Buffer	*New England Biolabs*
2’,3’-Cyclic adenosine monophosphate (cAMP), Sodium salt	*Merck*
2’,3’-Cyclic guanosine monophosphate (cGMP), Sodium salt	*Biolog*
3’,5’-Cyclic adenosine monophosphate (cAMP), Sodium salt	*Merck*
3’,5’-Cyclic guanosine monophosphate (cGMP), Sodium salt	*Merck*
2′-Deoxycytidine 5′-triphosphate (dCTP)	*Thermo Fisher Scientific*
2′-Deoxyguanosine 5′-triphosphate (dGTP)	*Jena Biosciences*
2’-Deoxynucleosine 5’triphosphate mix (dNTPs)	*New England Biolabs*
Diethylene glycol	*Merck*
Disodium hydrogen phosphate	*Grüssing*
Enhanced chemoluminescence (ECL) solution	*Biorad*
Ethylenediaminetetraacetic acid (EDTA)	*Merck*
Glycerol	*Carl Roth*
Glycine	*Carl Roth*
1,6-Hexanediol	*Merck*
Hydrochloric acid	*VWR*
4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	*Fluka*
Imidazole	*Merck*
Isopropyl β-d-1-thiogalactopyranoside (IPTG)	*Gerbu Biotechnik*
Kanamycin sulfate	*VWR*
Lysozyme	*PanReac Appli Chem*

Magnesium chloride	*Carl Roth*
Manganese chloride	*Merck*
MBC Suite Crystallisation Screen	*Nextal Biotech*
2-Mercaptoethanol	*Merck*
Methanol	*Merck*
2-Methyl-2,4-pentanediol (MPD)	*Merck*
Morpheus crystallisation screen	*Molecular Dimensions*
NEBuffer 2.1	*New England Biolabs*
Nickel chloride	*Merck*
2-(N-morpholino)ethanesulfonic acid (MES)	*Merck*
3-(N-morpholino)propanesulfonic acid (MOPS)	*Merck*
PACT Suite crystallisation screen	*Jena Bioscience*
Pentaethylene glycol	*Merck*
Penta His HRP Conjugate	*Qiagen*
Phenylmethylsulfonyl fluoride (PMSF)	*Merck*
Phusion HF Buffer	*New England Biolabs*
Polyethylene glycol 1000	*Merck*
Polyethylene glycol 1500	*Fluka*
Polyethylene glycol 3350	*Fluka*
Polyethylene glycol 4000	*Fluka*
Polyethylene glycol sorbitan monolaurate (Tween® 20)	*Serva*
Potassium acetate	*Merck*
Potassium sodium tartrate	*Merck*
1,2-Propanediol	*Thermo Fisher Scientific*
1,3-Propanediol	*Carl Roth*
2-Propanol (Isopropanol)	*VWR*
QIAquick PCR Purification Kit	*Qiagen*
QIAprep Spin Miniprep Kit	*Qiagen*
Roti GelStain	*Carl Roth*
Rp-Adenosine-3',5'-cyclic monophosphorothioate (Rp-cAMPS)	*Santa Cruz Biotech*
Rubidium chloride	*Merck*
Sodium acetate	*Fluka*
Sodium bromide	*Grüssing*
Sodium cacodylate	*Fluka*
Sodium chloride	*VWR*
Sodium dihydrogen phosphate	*Merck*
Sodium dodecyl sulfate (SDS)	*AppliChem*
Sodium fluoride	*Fluka*
Sodium formate	*Fluka*
Sodium hydroxide	*AppliChem*
Sodium iodide	*Merck*
Sodium nitrate	*Merck*
Sodium oxamate	*Merck*
Sodium propionate	*Merck*
Sp-Adenosine-3',5'-cyclic monophosphorothioate (Sp-cAMPS)	*Santa Cruz Biotech*
Sypro Orange	*Life Technologies*
Tetraethylene glycol	*Merck*
Tetramethylethylenediamine (TEMED)	*Carl Roth*
threo-1,4-Dimercapto-2,3-butanediol (DTT)	*Merck*
T4 DNA Ligase Reaction Buffer	*New England Biolabs*
Triethylene glycol	*Merck*

Tris(hydroxymethyl)aminomethane (Tris)	*Carl Roth*
Trisodium citrate	*Merck*
Tryptone	*Th. Geyer*
Unstained Protein MW Marker	*Thermo Fisher Scientific*
Yeast extract	*Th. Geyer*

2.2 Consumables

Type	Model	Supplier
Dialysis membranes	Slide-A-Lyzer 3.5 K 3-12 mL	*Thermo Fisher Scientific*
Desalting columns	PD-10 desalting columns	*GE Healthcare*
Spin Concentrators	Amicon (3/10 kDa MWCO)	*Millipore*

2.3 Enzymes

Name	Type	Supplier
*Nco*I	Endonuclease	*New England Biolabs*
*Nde*I	Endonuclease	*New England Biolabs*
*Ssp*I	Endonuclease	*New England Biolabs*
T4 DNA Ligase	Ligase	*New England Biolabs*
T4 DNA polymerase	DNA polymerase	*New England Biolabs*
*Xho*I	Endonuclease	*New England Biolabs*

2.4 Equipment

Type	Model	Supplier
Agarose Gel Casting Chamber	Mini-PROTEAN 3	*Bio-Rad*
Analytical Balance	KB 2400-2N	*Kern&Sohn*
	Explorer®	*Ohaus*
Autoclave	Tline Exor R&D	*Fedegari*
Benchtop Centrifuge	5810 R	*Eppendorf*
Centrifuge	J2-HS	*Beckmann*
	Sorvall Lynx 6000	*Thermo Scientific*
Centrifuge Rotors	F6S 6x1000Y	*Thermo Scientific*
	JA-20 Fixed Angle Rotor	*Beckmann*
	55.2 Ti	*Beckmann*
Chromatography Columns	Superdex 200pg 16/60 & 26/60	*GE Healthcare*
	Superdex Increase 10/300GL	*GE Healthcare*
	Protino® Ni-NTA	*Macherey-Nagel*
Chromatography Systems	NGC Quest	*Bio-Rad*
	Äkta Prime	*GE Healthcare*
Documentation System Agarose Gels	Herolab Smartblue	*Accuris*
Documentation System Crystallization	Rock ImagerTM	*Formulatrix*

Documentation System Western-Blot	Fusion SL, Camera serial-number: 279505	*Vilber Lourmat*
Electrophoresis Chamber for Agarose Gels	Mini-PROTEAN^{R} Tetra System	*Bio-Rad*
Electrophoresis Power Supply	EPS301	*Amersham*
Heated Stirring Waterbath	Y14	*Grant*
Heating Shaker	Thermomixer Comfort	*Eppendorf*
Homogenizer	FastPrep-24	*MP Biomedicals*
	EmulsiFlex-C5	*Avestin*
	FRENCH® Press with in-house produced pressure cell	*Thermo IEC*
Incubator	BFED-53	*Binder*
Incubator Shaker	Innova S44i	*Eppendorf*
	Ecotron	*Infors HT*
	Thermomixer comfort	*Eppendorf*
ITC-System	Microcal PEAQ-ITC	*Malvern*
Magnetic Stirrer	MR Hei-Mix L	*Heidolph*
Microcentrifuge	Heraeus Fresco21	*Thermo Scientific*
Microscope	SZ60	*Olympus*
	EC3	*Leica*
Peristaltic Pump	Pumpdrive 5201	*Heidolph*
pH Meter	HI2020 edge®	*Hanna*
Photometer	NanoDrop™ 2000/2000c	*Thermo Scientific*
	Ultrospec 3100 Pro	*Amersham*
	OD600 DiluPhotometer™	*Implen*
Pipettes	Eppendorf Research®	*Eppendorf*
Plate Reader	Spark	*Tecan*
Protein Crystallization System	HoneyBee 963™	*Digilab*
Real-Time PCR Cycler	Rotor Gene Q	*Quiagen*
Shaker & Mixer	Unimax 2010	*Heidolph*
Thermal Cycler	T-personal	*Biometra*
	Mastercycler® personal	*Eppendorf*
Ultra-Pure Water Systems	UV	*VWR*
X-ray Sources	Synchrotron Beamline P13	*EMBL Hamburg*
	Synchrotron Beamline PXI	*SLS Villigen*

2.5 Oligonucleotides, Vectors and DNA

2.5.1 Oligonucleotides

2.5.1.1 Primers

Name	Sequence (5'-3')	Restriction site included
Smc01798LIC_F	tac ttc caa tcc aat GCA AGA ATC GCG GTA ATC TCC GAT A	
Smc01798LIC_R	tta tcc act tcc aat GTT ATT AGT CGA GGA ATC CGG TCG CGA GC	
Smc02712LIC_F	tac ttc caa tcc aat GCA TTC AAA CTT GCG CAT ATA TCC G	
Smc02712LIC_R	tta tcc act tcc aat GTT ATT AGA AAT ACG TCG TTT CGG CGA GC	
Smc03964LIC_F	tac ttc caa tcc aat GCA TTC GGC ATC GAA CCG CTT GCG C	

Smc03964LIC_R	tac ttc caa tcc aat GTT ATT ATC CGA GAT CGA CGA CCA CAA TT	
Smb20081LIC_F	tac ttc caa tcc aat GCA AGC AAG TTG AAG GTT GCC GCC G	
Smb20081LIC_R	tac ttc caa tcc aat GTT ATT AAG AAG CCA AGT TCA GAG CTC GA	
EcCpdA_LIC_F	tac ttc caa tcc aat GCA TTG GAA AGC CTG TTA ACC CTT C	
EcCpdA_LIC_R	tac ttc caa tcc aat GTT AAT AGT AGC CTT CTG AAG CGG TAT CA	
CLR_sense	CTA GAT CAT ATG GCT GAA GTC ATC CG	*Nde*I
CLR_anti	CTC ATA CTC GAG ATC CTC CTC CG	*Xho*I
SsaCrp_F	ATATACCATGGCTAGCCGTAGCTTTCCGATT	*Nco*I
SsaCrp_R	TGGTGCTCGAGATCTTCTTCCGGATCGGCAA	*Xho*I
SamCrp_F	TATACCATGGGCCGTAGCTTTCCGATTTTTGA	*Nco*I
SamCrp_R	TGGTGCTCGAGATCTTCCGGCGGATCTGCAA	*Xho*I
SspCrp_F	ATATACCATGAGTTTCGCAATTTTCGAAGG	*Nco*I
SspCrp_R	TGGTGCTCGAGATCTTCTTCCGGATCTGCAA	*Xho*I

2.5.1.2 Duplex DNA

Name	Sequence (5’-3’)	Length
Clr_sym_19mer	CTAGGTAACATTACTCGCG	19 bp
Clr_sym_14mer	GCGAGTAATGTTAC	14 bp

2.5.2 Vectors

The pET vector system is the most used for cloning and expression of protein ins *E. coli*. It utilizes the strong bacteriophage T7 expression system implemented in the recombinant host combined with a *lac* repression. This allows low basal activity and high levels of protein expression upon induction, ranging up to 50% of the total protein content of the cell. In addition to the gene of interest, pET plasmids code for an antibiotic resistance used as selection marker and *lacI* encoding the lac repressor (Goeddel 1990; Rosenberg et al. 1987; Studier and Moffatt 1986).

2.5.2.3 pET36b(+) vector

The expression vector pET36b(+) differs from other vectors of the pET system by its N-terminal cellulose binding domain (CBD). It is also equipped with a *C*-terminal Histidine-Tag and a Thrombin cleavage site for removal of the CBD-Tag. The vector provides an export signal as well. During cloning of the Clr expression construct, all *N*-terminal features were removed. The resulting vector is illustrated in Figure 5.

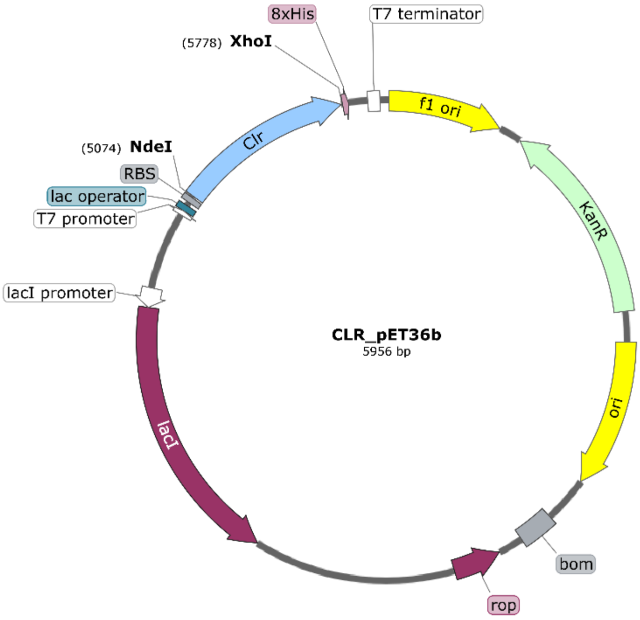

8xHis
T7 terminator
(5778) XhoI
f1 ori
Clr
(5074) NdeI
RBS
lac operator
T7 promoter
KanR
lacI promoter
CLR_pET36b
5956 bp
ori
lacI
bom
rop

Figure 5: Vector map of pET36b(+)-Clr. The plasmid was used for heterologous protein production and was a kind gift from Petra Gnau. The native sequence was cloned into the vector using the *Nde*I and *Xho*I restriction sites. The recombinant Clr protein is equipped with a *C*-terminal octahistidine tag and produced with kanamycin as selection marker.

2.5.2.4 pET28a(+) vector

The vectors pET28a-c(+) are among the most used expression vectors. pET28b and c merely deviate from each other in the removal of one base from each side beyond the *BamHI* restriction site. Both *N*- and *C*-terminal histidine tags are set up within the plasmid and can be added or removed depending on the choice of restriction enzyme. Cleavage of the *N*-terminal tag with thrombin is possible as well. The expression plasmid for *Sm*CpdA was purchased from *BioCat* containing a *C*-terminal hexahistidine tag (Figure 6).

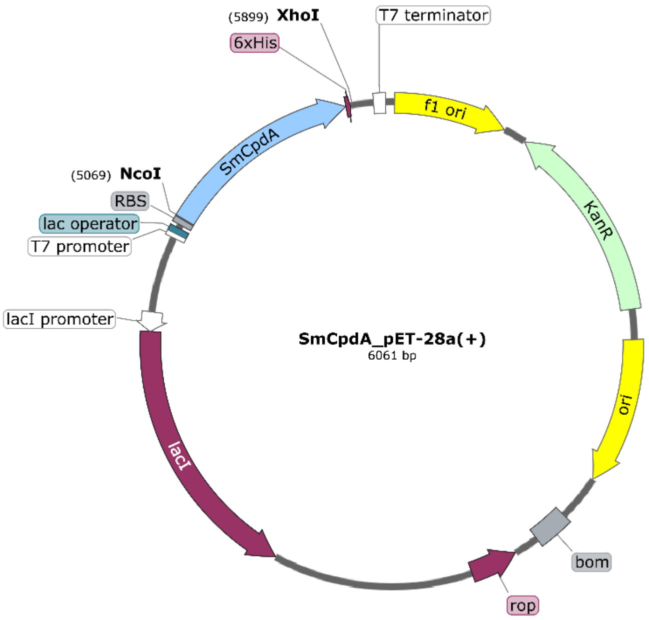

(5899) XhoI
6xHis
T7 terminator
f1 ori
KanR
ori
bom
rop
lacI
lacI promoter
T7 promoter
lac operator
RBS
(5069) NcoI
SmCpdA
SmCpdA_pET-28a(+)
6061 bp

Figure 6: Vector map of pET28a(+)-*Sm*CpdA: The construct was used for heterologous production of phosphodiesterase CpdA fused to a *C*-terminal hexahistidine tag. The plasmid codes for the aminoglycoside 3'-phosphotransferase from *Brucella abortus* lending the transformed organism a kanamycin resistance.

2.5.2.5 pET His$_6$ Mocr TEV LIC Cloning Vector

The pET His$_6$ Mocr TEV LIC cloning vector is optimized for ligation independent cloning (LIC; see methods). It contains the monomeric mutant of the Ocr (Mocr) protein of bacteriophage T7 as a solubility-enhancing fusion tag. This protein is rather small (13.8 kDa) and highly charged, as well as very acidic (pI = 3.8) and is overproduced very robustly (DelProposto et al. 2009). While the surface charge lends itself to a purification via DEAE resins, an additional His$_6$-Tag was included as well, to allow routine affinity purification. The fusion tag can be cleaved via TEV protease and kanamycin acts as selection marker. Figure 7 illustrates the plasmid containing phosphodiesterase *Sm*b200081 within the LIC cloning site.

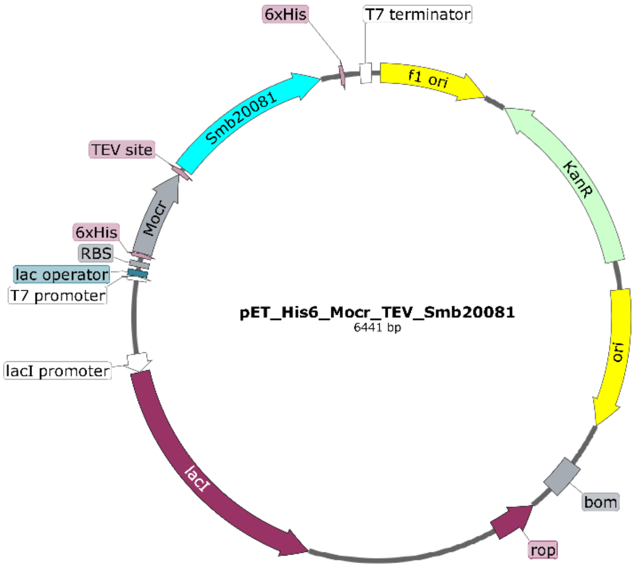

6xHis
T7 terminator
f1 ori
KanR
Smb20081
TEV site
Mocr
6xHis
RBS
lac operator
T7 promoter
pET_His6_Mocr_TEV_Smb20081
6441 bp
ori
lacI promoter
lacI
bom
rop

Figure 7: Vector map of pET His$_6$ Mocr TEV LIC cloning vector: The plasmid contains phosphodiesterase *Sm*b20081 and was used for in solubility optimisation. Monomeric Ocr has been reported to increase solubility and can be removed via TEV protease cleavage (DelProposto et al. 2009). The *N*-terminal hexahistidine tag allows conventional nickel affinity purification.

2.5.2.6 pET His$_6$ GST TEV LIC Cloning Vector

The glutathione S-transferase Tag included in the pET His$_6$ GST TEV LIC cloning vector is a 26 kDa protein derived from *E. coli*. It is known to fold rapidly into a stable and highly soluble protein upon translation and can therefore be used to promote greater expression and solubility in recombinant proteins. In addition to that, the affinity of the protein to its substrate glutathione (GSH, tripeptide ECG) can be used for affinity purification (Frangioni and Neel 1993). The vector also contains the *N*-terminal hexahistidine tag and TEV protease cleavage site present in its Mocr counterpart. The vector used for GST-fusion of *Sm*b20081 is shown in Figure 8.

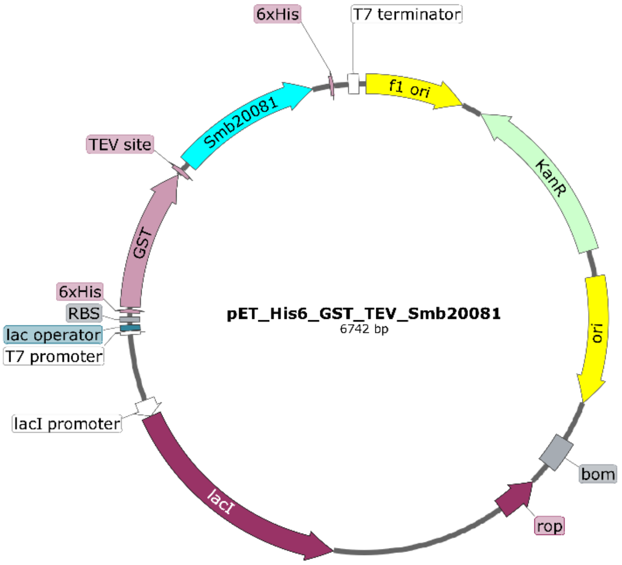

6xHis
T7 terminator
f1 ori
Smb20081
TEV site
KanR
GST
6xHis
RBS
lac operator
T7 promoter
pET_His6_GST_TEV_Smb20081
6742 bp
ori
lacI promoter
lacI
bom
rop

Figure 8: Vector map of pET His$_6$ GST TEV LIC cloning vector: To improve the solubility of *Sm*b20081, it was fused to a glutathione-S-transferase tag. This not only improves solubility in some cases, but also is an affinity tag in and of itself when used with a glutathione matrix. Alternatively, the *N*-terminal hexahistidine tag can be used as well. Like its Mocr counterpart the fusion construct is TEV protease cleavable.

2.5.2.6 pET His$_6$ MBP TEV LIC Cloning Vector

The most commonly used solubility and secretion enhancing tag is the maltose binding protein (MBP). It is considered exceptionally robust for increasing protein production and is tolerated in high concentrations in the cell. Moreover, proteins fused to MBP can be affinity purified via amylose resin and subsequently released by the addition of maltose (Reuten et al. 2016). In the pET His$_6$ MBP TEV LIC cloning vector, the protein of interest can then be cleaved via TEV protease and an *N*-terminal His$_6$ tag is included, facilitating the use of nickel affinity chromatography as alternative initial purification. Due to the large size (42.5 kDa) of the tag, some targets can be expressed soluble but then precipitate upon TEV cleavage. Figure 9 highlights the important features of this vector.

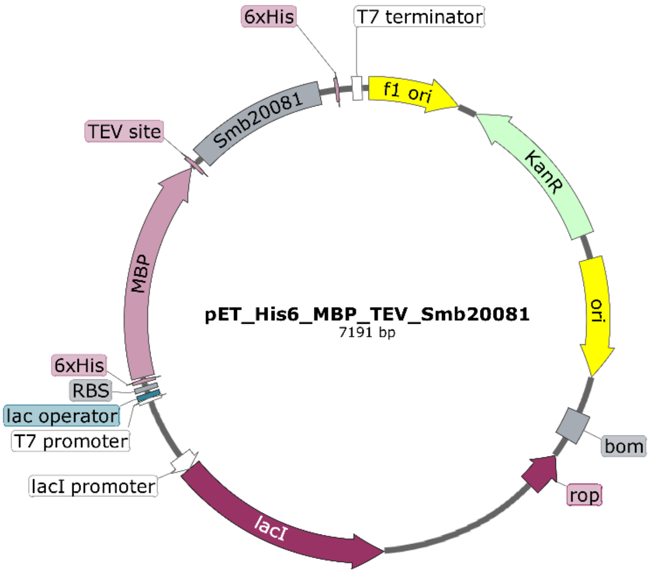

6xHis
T7 terminator
f1 ori
KanR
ori
bom
rop
lacI
lacI promoter
T7 promoter
lac operator
RBS
6xHis
MBP
TEV site
Smb20081
pET_His6_MBP_TEV_Smb20081
7191 bp

Figure 9: Vector map of pET His$_6$ MBP TEV LIC cloning vector: With about 42.5 kDa maltose-binding protein (MBP) is the largest solubility enhancing tag used in this work. It is part of the maltose uptake and catabolism system of *Escherichia coli* and its affinity for maltose can be utilized in affinity purification.

2.6 Organisms

2.6.1 *Escherichia coli* DH5α

Escherichia coli DH5α cells were purchased from *Invitrogen*. These cells do not contain the gene coding for T7 RNA polymerase and are defined by three mutations. *LacZΔM15* enables blue/white colour screening, a *recA1* mutation inactivates recombination to ensure genetic stability and the *endA1* mutation prevents the intracellular endonuclease from degrading the plasmid.

Genotype: *F^{-} φ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17(rk^{-}-, mk^{+}) phoA supE44 thi-1 gyrA96 relA1 λ^{-}*

2.6.2 *Escherichia coli* BL21 (DE3) Gold

Escherichia coli BL21(DE3) Gold cells were purchased from *Stratagene*. The strain utilizes the T7 RNA polymerase and is deficient in the Lon and OmpT proteases, which would otherwise degrade recombinant protein. They have been engineered towards high transformation efficiency and their endA1 gene is inactivated same as in DH5α, theoretically allowing direct cloning.

Genotype: *F^{-} ompT gal dcm lon hsdS$_B$(r$_B$$^{-}m_B$$^{-}$) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB^{+}]$_{K-12}$(λ^{S})*

2.6.3 *Escherichia coli* T7 Shuffle Express

Escherichia coli T7 Shuffle Express cells were purchased from *New England Biolabs*. Like *E. coli* BL21 (DE3) Gold the strain utilizes the T7 RNA polymerase and is deficient in the Lon and OmpT proteases to protect recombinant proteins from degradation. The strain has been engineered to form disulphide bonds in proteins expressed in the cytoplasm and contains DsbC to correct mis-oxidation and chaperone all recombinant proteins, even if they do not require disulphide bonds.

Genotype: fhuA2 lacZ::T7 gene1 [lon] ompT ahpC gal λatt::pNEB3-r1-cDsbC (Spec^R*, lacI*^{q}*) ΔtrxB sulA11 R(mcr-73::miniTn10--Tet*^{S}*)2 [dcm] R(zgb-210::Tn10 --Tet*^{S}*) endA1 Δgor Δ(mcrC-mrr)114::IS10*

2.7 Software

MicroCal PEAQ-ITC Analysis software	*Malvern Pananalytical, 2015*
mxCuBE v2	*Oscarsson et al. 2019*
DA+ SLS Data acquisition GUI	*Wojdyla et al. 2018*
XDS	*Kabsch 2010*
UCLA Diffraction Anisotropy Server	*Strong et al. 2006*
CCP4i2	*Potterton et al. 2018*
Swiss-Model Server	*Waterhouse et al. 2018*
Phenix	*Liebschner et al. 2019*
Coot 0.8.9.2	*Emsley and Cowtan 2004*
DynamX	*Waters, 2015*
Topspin	*Bruker, 2016*
TMHMM Server	*Krogh et al. 2001*
ProtParam	*Expasy, 2003*
Gremlin Server	*Ovchinnikov et al. 2014*
Robetta	*BakerLab, 2019*
Mega-X	*Kumar et al. 2018*

2.7.1 Structure Visualisation

Pymol 2.2.0	*Schrödinger, LLC, 2020*
Chimera 1.15	*Pettersen et al. 2004*
Inkscape 1.0.1	*Inkscape, 2020*
Paint.net 4.2.15	*dotPDN, 2020*
ChemBio Office 2010	*Perkin Elmer, 2010*

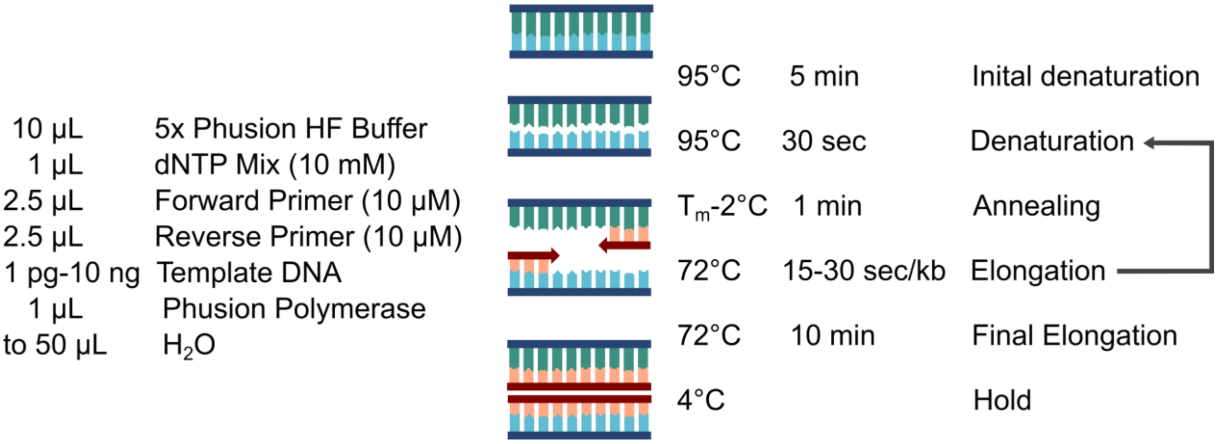
3 Methods

3.1 Molecular Biology

3.1.1 Polymerase Chain Reaction (PCR)

The polymerase chain reaction, a method developed by Kary B. Mullis in 1987, allows the enzymatic amplification of a specific nucleotide sequence (Mullis and Faloona 1987). The development enables the analysis of even trace amounts of DNA and resembles natural replication. Starting from short oligomers, called primers, RNA polymerase synthesizes a new DNA strand complementing a single stranded template. As the 3'-ends of the primers mark the replication start, multiple replication cycles result in a DNA fragment of defined length and sequence.

A PCR cycle starts with thermal denaturation of the double stranded template at 90-95°C followed by hybridisation of synthetic single stranded oligomer primers at temperatures $\geq$ 50°C to their complementary sequences within the template. Amplification by the DNA polymerase is typically conducted at 72°C. The use of polymerases originating from thermophilic organisms like *Thermus aquaticus* (Taq) circumvents the need for enzyme supplementation after each heating cycle. A typical PCR procedure is shown in Figure 10.



10 µL	5x Phusion HF Buffer
1 µL	dNTP Mix (10 mM)
2.5 µL	Forward Primer (10 µM)
2.5 µL	Reverse Primer (10 µM)
1 pg-10 ng	Template DNA
1 µL	Phusion Polymerase
to 50 µL	H_2O

95°C	5 min	Inital denaturation
95°C	30 sec	Denaturation
T_m-2°C	1 min	Annealing
72°C	15-30 sec/kb	Elongation
72°C	10 min	Final Elongation
4°C		Hold

Figure 10: Typical example of a PCR reaction: Elongation temperature and enzyme concentration are dependent on the kind of polymerase used in the experiment (here Phusion Polymerase from *New England Biolabs*). The annealing temperature can be deduced from the primers' melting temperature, whereas the elongation time is dependent on the size of the amplificate.

3.1.2 Preparation of plasmid DNA

Isolating plasmid DNA from bacterial cultures historically involved an alkaline lysis and the removal of cell debris via centrifugation, followed by phenol-extraction and precipitation (Birnboim and Doly 1979). Nowadays all reagents and consumables for the purification of plasmids are sold in commercial kits, like the *QIAprep Spin Miniprep Kit* from *Qiagen* used in this work. In the first step of the procedure RNA is digested via RNAse, whereas EDTA destabilizes membranes. This is followed by the alkaline and detergent-based lysis step, after which genomic DNA protein and cell debris are precipitated. The plasmid DNA is then bound to a silica column and washed with a buffer containing ethanol. Elution from the column is pH-dependent and can be conducted either via Tris buffer at pH 8.5 or just water.

3.1.3 PCR Purification

The aim of the purification step after a PCR reaction is the removal of residual primers, enzymes and dNTPs. Similar to the plasmid purification, PCR purifications are based on DNA binding to an anion-exchange column and elution at low salt concentrations either via Tris buffer at pH 8.5 or water. In this work the *QIAquick PCR Purification Kit* was used. The main difference between silica columns used for plasmid and PCR purification is the size of DNA they are optimized to bind to. Fragments ranging from 100 bp to 10 kbp are suitable for this particular kit.

3.1.4 DNA Quantification

The absorption maximum of DNA in UV/VIS is located around 260 nm, slightly higher or lower depending on pH, solvent as well as base sequence and pairing. The ratio between absorption at 260 and 280 nm is an indicator for protein contamination- a ratio OD_{260}/OD_{280} of 1.8-2.0 is considered to be ideal. The *Nanodrop2000* from *Thermo Fisher* used in this work to determine nucleic acid concentrations required very little sample (0.5-2 µL), however it is also quite error prone, meaning measured concentrations were treated as rough estimates. For crystallisation and ITC measurements synthetic DNA was weighed and its concentration verified by UV/VIS. RNA and DNA cannot be distinguished by this method.

3.1.5 Restriction Digest

Endonucleases or Restriction enzymes are part of the defence mechanism of many organisms against foreign DNA. They cleave either in response to a recognition sequence or modification. The actual cleavage can occur in varying distances from the recognized signal and can either be “blunt” or “sticky" depending on whether or not there is an overlap between one strand of the resulting fragments. For restriction cloning, complementary enzymes are used to cut both insert and vector (Figure 11). To avoid religation of the vector, the DNA can be dephosphorylated using *calf intestine alkaline phosphatase* (CIP). The enzyme needs to be heat inactivated before ligation. The restriction enzymes *Nco*I, *Nde*I and *Xho*I were purchased from

New England Biolabs. A typical restriction digest reaction contained 1 µg of vector or 1.5-2 µg of insert DNA with 10x CutSmart® buffer in a total volume of 50 µL and was incubated at 37°C for 15 min to 1 h.

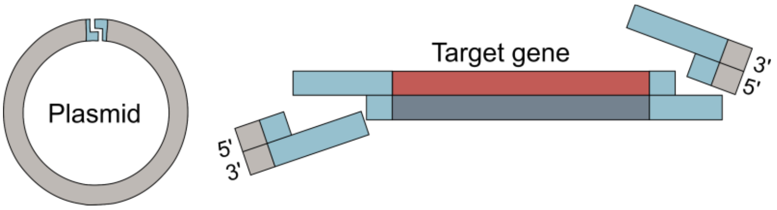


Figure 11: Schematic representation of a restriction digest: Plasmid and PCR amplificate are cleaved at a specific motif using the same endonucleases. This results in short complementary overhangs that can be annealed and connected using ligase.

3.1.6 LIC Cloning

Ligation-independent cloning was developed as an alternative to traditional restriction enzyme-mediated cloning. The upsides are that cleavage sites within the gene of interest are of no concern as well as the higher selectivity resulting from longer overhangs. In addition to that, the universal overhangs make construction of libraries easier. The technique uses the 3' exonuclease activity of T4 DNA polymerase to create complementary overhangs between vector and insert. Inserts are amplified via PCR using primers complementing the vector, that is linearized using a type II restriction enzyme like *Bsa*I or *Ssp*I. Both are then treated with T4 DNA polymerase, stopping the reaction at the first guanine or cysteine via inclusion of dGTP and dCTP respectively. The overhang length (typically 10-12 bp) is sufficient to allow ligation free construct assembly and direct transformation. Any LIC cloning steps in this work were conducted according to the Addgene ligation-independent cloning protocol. The procedure is outlined in Figure 12.

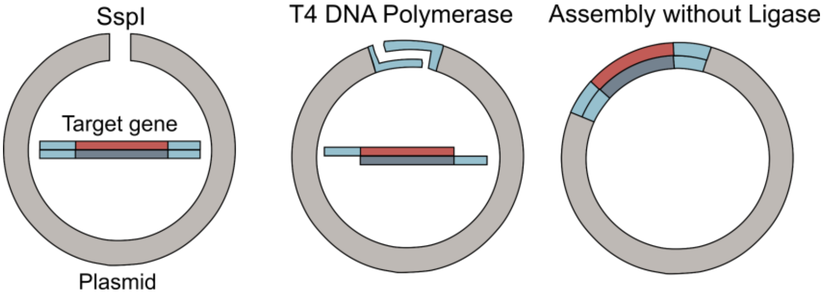


Figure 12: Typical procedure of ligation-independent cloning: The vector is linearized with a blunt end cutting endonuclease, the target gene is amplified using complementary primers via PCR. The exonuclease activity of T4 DNA polymerase is utilized to create long overhangs. The digest is limited

by the addition of 2'deoxynucleotides, as the incorporation by the polymerase makes further exonuclease attacks impossible. Since the overhangs are about 4-5 times longer than in classic restriction cloning, the assembly is stable enough to be transformed without prior ligation. The single strand breaks are then healed by *E. coli's* inherent DNA repair mechanisms.

3.1.7 Ligation

The ligase from bacteriophage T4 is most commonly used in laboratory research as it is capable of ligation of both cohesive and blunt ends of nucleic acids resulting from restriction enzyme digestion, but not of single-stranded oligonucleotides. The reaction is Mg^{2+} dependent and consumes ATP, which are both supplied in the ligation buffer (Gaastra and Hansen 1985). In this work commercially available *T4 DNA Ligase* and the *T4 DNA Ligase Reaction buffer* from *New England Biolabs* were used. A typical reaction would contain insert: vector at a 3:1 ratio in a nanomolar range, 10x DNA Ligase buffer and 1 µL T4 DNA Ligase in a reaction volume of 20 µL. The reaction was incubated at room temperature for 1 h.

3.1.8 Agarose Gel Electrophoresis

Agarose gel electrophoresis is commonly used to both analyse and purify DNA fragments in the range of 0.5 to 25 kbp. Fragments smaller than 100 bp are more effectively separated using polyacrylamide gel electrophoresis (Lee et al. 2012). The method utilizes the migration of the negatively charged phosphate backbone in an electric field in combination with the segregation by size within the agarose matrix. Agarose is a galactose-based polysaccharide derived from red algae, that is dissolved in a Tris borate EDTA (TBE buffer) by heating and forms a porous gel upon cooling (Serwer 1983). To stain the gel traditionally ethidium bromide was used, due to its intercalating properties it has now been replaced by other dyes that are considered non-cancerogenic. The dyes are excited under UV light, revealing the DNA bands within the gel, the respective size is determined by a standardized mix of fragments.

Any agarose gels used in this work were prepared at a concentration of 1% with the addition of 4 µL ROTI Gel Stain from *Carl Roth* and overlaid with TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA). For analytical gels 5 µL sample were used, whereas a preparative gel contained 50 µL sample per pocket. Runs were conducted at 120 V for 1 h. In case of preparative gel electrophoresis, the gel slices were dissolved with the QIAquick Gel Extraction Kit from *Qiagen*. The operating principle is highlighted under *PCR Purification*.

3.1.9 RbCl-Competent Cells

Bacteria take up external DNA as a means of obtaining new skills, called horizontal gene transfer (Findlay and Redfield 2009). The ability of the cells to bring DNA across their membrane is termed natural competence and can be enhanced artificially. The surface of bacteria such as *E. coli* is negatively charged due to phospholipids and lipopolysaccharides on its cell surface. It is argued that divalent cations like calcium and rubidium shield the phosphate charges of the oligonucleotide and membrane, making entry possible (Mandel and Higa 1970). Cells purchased from the suppliers listed under *Materials* were grown over night on LB agar (15% (w/v) agar, 10% (w/v) NaCl, 10% (w/v) tryptone and 5% (w/v) yeast extract). A single colony was grown in 4 mL LB medium (10% (w/v) NaCl, 10% (w/v) tryptone and 5% (w/v) yeast extract, pH 7.0 (800 µL 10 M NaOH per L)) over night, the main LB culture (50 mL) was inoculated with preculture (1:100) and grown to an optical density of 0.5 (595 nm). The cells were harvested by centrifugation (15 min, 4000 rpm, 4°C), resuspended in 15 mL TBFI buffer (100 mM $RbCl_2$, 50 mM $MnCl_2$, 30 mM potassium acetate, 10 mM $CaCl_2$, 13% (v/v) glycerol) and incubated on ice for 2 hours. After centrifugation (10 min, 4000 rpm, 4°C) they were suspended in 2 mL TBFII buffer (10 mM MOPS, 10 mM $RbCl_2$, 75 mM $CaCl_2$, 13% (v/v) glycerol) and immediately flash frozen in liquid nitrogen in 50 µL aliquots. The cells were stored at -80°C.

3.1.10 Transformation

Chemically competent cells are subjected to a heat shock to facilitate DNA uptake. It has been theorized that the thermal stress renders the cell membrane more permeable via pore formation, but the definite mechanism still remains to be explored (Rahimzadeh et al. 2016). For Transformations during this work, competent cells were thawed on ice for 10 min, about 10 ng of plasmid DNA or the entirety of a ligation reaction were added, gently mixed and incubated for 30 min. The mixture was subjected to a heat shock at 42°C for 45 seconds and cooled on ice for 5 minutes. Subsequently, 200 µL of LB medium were added and the samples were shaken vigorously at 37°C for one hour. 10 and 90 percent of the slurry were plated on different halves of LB agar plates containing the antibiotic complementing the plasmid used.

3.1.11 Colony PCR

When transforming freshly ligated plasmids, possible religation of vector needs to be excluded using either of two methods. In a test digest one would prepare plasmid from a culture of transformants, perform a restriction digest and analyse on agarose gel to verify the presence of insert. The alternative method is the so-called colony PCR, where a colony of transformants is used as template in a PCR reaction using the original primers and the amplified insert is visualized on agarose gel. The latter was used in this work. To perform a colony PCR the reaction mixture was set up as described under *polymerase chain reaction* and 5 µL of the finished amplificate were analysed via *agarose gel electrophoresis*.

3.1.12 Sequencing

A traditional method of DNA sequence determination is the so-called *Sanger Sequencing*. It is based on the incorporation of chain terminating dideoxynucleotides during in vitro replication. The resulting fluorescently labelled fragments are then separated via gel electrophoresis and the eluting fluorescent peaks recorded. The method is suitable for fragments up to 500 bp, for larger sequences one can either combine multiple reads or switch to *Next generation sequencing*, a real time fluorescence detection during elongation done in parallel short reads with hundreds of megabases to gigabases per instrument run. Newly cloned plasmids during this thesis were sent to *Eurofins* (previously *GATC*) for sequencing using the Sanger method.

3.1.13 Hybridisation of DNA Oligonucleotides

The stability of hybridisation of complementary DNA is essential for many of its functions. It is mainly determined by two factors, namely stacking between adjacent bases and pairing between complementary bases (Protozanova et al. 2004; Yakovchuk et al. 2006). Slow cooling of the complementary strands is preferred as it increases the specificity of the hybridisation, especially in a DNA fragment containing a double strand break und assembled from four single stranded oligonucleotides, like the one used in this thesis (SantaLucia and Hicks 2004).

3.2 Protein Biochemistry

3.2.1 Heterologous Overproduction in *E. coli*

Escherichia coli is the standard organism for heterologous production of protein in bacteria. It is fully sequenced and its replication and expression systems are well understood (Baneyx 1999). The production of proteins, in particular originating from higher organisms, still presents some major challenges mainly regarding codon usage, solubility and toxicity. The result is impacted by factors like vector, strains containing different features and culture conditions. The expression is regulated by the lac operon utilizing the T7 RNA polymerase and induced by the addition of lactose or isopropyl β-D-1-thiogalactopyranoside (IPTG), an allolactose homolog (Rosano and Ceccarelli 2014). Large scale expressions in this work were conducted in large (5 L), baffled flasks at a shaking frequency of 125 rpm. Test expressions were conducted to identify suitable expression conditions before scale-up.

3.2.2 Cell Lysis

The most common methods for breaking *E. coli* cells are either based on mechanic shear forces, chemical or ultrasound disruption (Kelemen and Sharpe 1979; MILNER et al. 1950). In a French Press pressure cell, cells are pressed through a fine needle repeatedly. The resulting rapid reduction in pressure is enough to disrupt the cell wall of *E. coli* reliably (Rogers et al. 2012). For lysis the cell suspension, e.g., Clr after addition of 50 µL lysozyme (50 mg/mL) and PMSF (200 mM) and incubation for 15 min) was added to a precooled French Press pressure cell and passed through the needle at 15.2 bar cm^{-2} three times. The raw lysate was centrifuged at 18000 rpm for 1 h in a Beckman JA-20 rotor.

In an analytical scale, bead beating was used (Verollet 2008). 1 mL of the sample and 500 mg of 0.2-0.5 mm glass beads were combined and homogenized four times for 30 s using a *FastPrep-24* from *MP Biomedicals* in intervals of 5 min. The sample was cooled in between. Sedimentation of the cell debris took 1 h at 14800 rpm (JA-20) and the sample was then analysed regarding the ratio of soluble protein in the supernatant versus insoluble in the pellet. Detailed descriptions can be found in the *Appendix*.

3.2.3 Immobilized Metal Affinity Chromatography (IMAC)

The high affinity of histidine residues to transition metal ions, Ni^{2+} in particular, can be used to purify proteins from raw lysate (J. Hearon 1948). The metal was initially fixed to resin via iminodiacetic acid (IDA) and has since been upgraded to nitrilotriacetic acid (NTA) as chelating ligand (Porath et al. 1975). The metal is coordinated with four of its valences by the NTA ligand, two valences remain for histidine interaction as shown in Figure 13 (Block et al. 2009). The method is pH-dependent due to the protonation state of the NTA ligand. Bound protein can therefore be either eluted by acidic pH, metal removal via ethylenediaminetetraacetic acid (EDTA) or a stronger binder, e.g., imidazole. Only the latter is leaving the metallized column matrix intact and is generally preferred. Due to its selectivity for tagged target proteins, IMAC is a powerful and robust first purification step from raw lysates and the resin can withstand quite high pressure while being very affordable (Spriestersbach et al. 2015). Another advantage of the method is the high binding capacity of about 50 mg/mL resin, however there are also some limitations. While small or medium-sized proteins have high binding capacities, the recovery rate decreases for large sized proteins (Tovar and Odunuga 2019). In these cases, the StrepII-tag becomes a viable alternative.

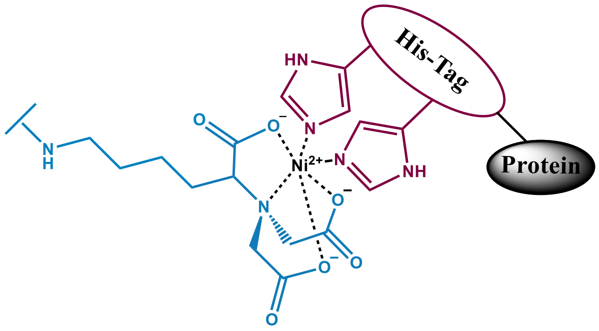


Figure 13: Schematic representation of the octahedral nickel complex formed during IMAC: The coordination of nickel by nitrilotriacetic acid (NTA) chelates the metal and leaves two free binding sites. These can be occupied by histidine ligands. The affinity is proportionate to the number of subsequent histidines and therefore higher for polyhistidine tags introduced in the protein than for native histidines present in the sequence. Imidazole has an even higher nickel affinity and can be used to disrupt the complex, releasing the protein of interest from the resin.

For IMAC purifications during this thesis the raw lysate was filtered at 0.45 nm, the pH was measured and the lysate was passed through a 5 mL *Protino Ni-NTA* FPLC column from *Macherey Nagel* for one hour. Once the whole solution had passed the column one last time, the column was washed with varying combinations of buffers containing low and high imidazole amounts, to first remove unspecific binders and subsequently elute the protein of interest. The elution was either monitored via UV/VIS using the *NGC* FPLC System from Biorad and/or *SDS-PAGE*. The column was washed with high imidazole buffer and water and stored under 20% ethanol.

Binding buffer Clr	**Elution buffer Clr**	**Binding buffer CpdA**	**Elution buffer CpdA**
50 mM NaH_2PO4	50 mM NaH_2PO4	50 mM NaH_2PO4	50 mM NaH_2PO4
1 M NaCl	1 M NaCl	300 mM NaCl	300 mM NaCl
10 mM imidazole	250 mM imidazole	10 mM imidazole	250 mM imidazole
pH 7.0	pH 7.0	pH 7.8	pH 7.8

3.2.4 Size Exclusion Chromatography (SEC)

Gel filtration or size exclusion chromatography is a common polishing step in protein purification. The column consists of cross-linked polyacrylamide, agarose or dextran beads acting as a molecular sieve that segregates biomolecules by size (Porath and Flodin 1959). While larger molecules cannot enter the pores and elute first, the passage of smaller molecules is delayed (Scopes 1994). Protein species larger than the cut-off size of the beads elute in the void volume (not occupied by matrix) of the column. For spherical beads the void volume equals to ~25-30% of the column volume (Duong-Ly and Gabelli 2014). The elution volume of a particular protein has a logarithmic relationship to its size and this can be used for size determination with a suitably calibrated column (Whitaker 1963). Different parameters determine the separation quality. The resolution of separation will decrease, if a large sample volume is injected - less than 3% of the column volume has been suggested to yield good results (Scopes 1994). The sample is diluted during purification and the dilution increases with larger elution volumes. Higher resolutions can be obtained with longer columns, whereas higher diameters allow for a larger volume of sample (Duong-Ly and Gabelli 2014). Proteins in their native state are segregated according to their Stokes radius by size exclusion chromatography. This allows for the separation of oligomeric species and complexes as well. A very comprehensive review of the method has been provided by Winzor 2003.

In this thesis Superdex Prep Grade (Dextran-Agarose) columns with 75 and 200 resin in 16/600 or 26/600 columns were used. Two to five millilitres of sample were purified using HEPES or TRIS based buffers on an *NGC* FPLC system from *BioRad*. A typical size exclusion chromatogram can be seen in Figure 14.

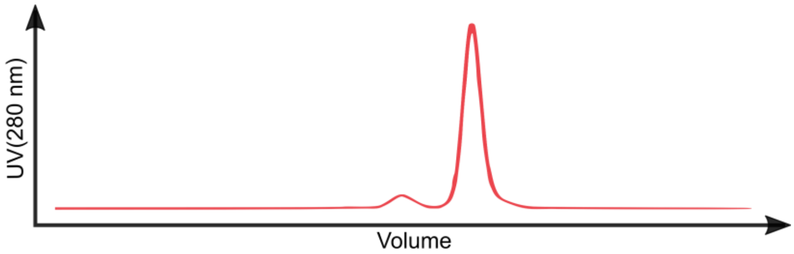


Figure 14: Schematic representation of a size exclusion chromatogram: Larger proteins are able to pass the column faster than smaller ones, which are retained in the pores of the resin. The pore size determines the optimal segregation range of the column. Proteins that are too large elute in the void volume. The elution can generally be monitored via UV/Vis spectroscopy at 280 nm corresponding to the aromatic amino acids phenylalanine, tryptophan, tyrosine and histidine.

SEC buffer Clr	**SEC buffer CpdA**
20 mM HEPES	50 mM Tris
300 mM NaCl	300 mM NaCl
pH 7.0	pH 7.8

3.2.5 Buffer Exchange

Desalting columns for buffer exchange operate by the same principle as size exclusion columns, however their pores are a lot smaller, meaning smaller compounds like ions and buffer salts are delayed whereas protein passes quickly (Duong-Ly and Gabelli 2014; Porath and Flodin 1959). This also means, that all proteins above the cut-off of typically 5000 Da elute close to the void volume and the elution can be done without fractionation or UV detection, simply by collecting 3.5 mL from the void volume. PD-10 Desalting columns from *GE Life Sciences* were used for buffer exchange, metal or EDTA removal during this work according to the manufacturers protocol.

3.2.6 SDS-PAGE

Denaturing gel electrophoresis is a powerful and quick tool for monitoring protein expression and purification. The addition of sodium dodecyl sulfate (SDS) unfolds any proteins within the sample and provides a strong negative charge (Pitt-Rivers and Impiombato 1968). As a result, proteins pass the polyacrylamide gel electrophoresis (PAGE) purely based on size and neither hydrodynamic volume nor isoelectric point. The polyacrylamide gels used in the method are biphasic to improve separation by using the stacking effect. The upper stacking gel contains the sample pockets, has larger pores due to a lower acrylamide content and a pH of 6.8. The lower separation gel has a higher acrylamide concentration (smaller pores) and a pH of 8.8. The stacking gel pH is close to the pI of glycine within the running buffer, giving it a low charge and a resulting low electrophoretic mobility (lagging ion). Also contained chloride ions with their high size to charge ratio have a high mobility and travel first (leading ion). Protein ions are trapped within field strength gradient of leading and lagging ions. They are focused on the interface by the delay resulting from smaller pore size in combination with the changing pH that deprotonates the lagging ion glycine. Through the introduction of the negative charge glycine overtakes the protein ions at the interface and forces the stacking even further. After entering as one sharp band, proteins are free to move within the separation gel without any restrictions and segregate based on size (Laemmli 1970). There are multiple ways of visualizing proteins within the gel. The two most common are Coomassie blue staining which is quite rapid and simple and silver staining which is a lot more sensitive (Wilson 1979). A more recent alternative is given by fluorescent staining using dyes like SYPRO orange as it combines the advantages of being easier to use than silver staining while providing just as much sensitivity (Sasse and Gallagher 2004). For the heterologous production of proteins in the milligram range, sensitivity is not the issue at hand and Coomassie staining still has its place here.

For any SDS-PAGE experiments performed in this thesis samples were mixed with equal parts sample buffer (4% SDS, 20% glycerol, 10% 2-mercaptoethnaol, 0.004% bromophenol blue and 0.125 M Tris pH 6.8). 2-Mercaptoethanol is contained to reduce any disulfide bridges. Samples consisting of intact cells or cell pellets were heated to 98°C for 10 min and centrifuged at 12.000 rpm for 2 min. 5-20 µL of sample were added to the stacking gel with *Pierce Unstained Protein Ruler* from *Thermo Fisher* as reference and run at 40 mA per gel (300 V) for 35 min. Gels were dyed with a solution containing 0.32 % (w/v) Coomassie blue, 40% (v/v) ethanol and 8% (v/v) acetic acid. Background staining was removed using a mix of 30% (v/v) ethanol and 10% (v/v) acetic acid.

3.2.7 Protein Quantification

Photometric methods are most commonly used to quantify protein in solution. They are of particular importance when aiming for kinetics and affinity measurements. A pitfall in all of these methods is their dependence on protein composition (Sapan et al. 1999). In the colorimetric Bradford assay, protein binding to Coomassie blue dye is used for quantification (Bradford 1976). The complex formation increases the molar absorbance which is measured at 595 nm, making the method a quite sensitive one. The reaction takes place under acidic conditions achieved generally by the addition of phosphoric acid. Coomassie blue mainly forms ion pair with lysine and arginine residues (Congdon et al. 1993; Compton and Jones 1985), therefore the choice of protein used in the calibration is quite important for the accuracy of the assay.

The Bradford solution used in this work was calibrated using BSA and 1 µL protein solution were incubated with 1 mL dye solution for 5 min prior to measurement at 595 nm. Protein dilutions were made as needed to stay within the linear range of the assay (up to 2 mg/mL).

A quick and easy way to quantify protein in solution is taking the UV/VIS absorbance of aromatic residues at 280 nm. It is also most likely to provide incorrect results. The method is not suitable for proteins containing few aromatic groups or cofactors absorbing at that particular wavelength. Tyrosine and tryptophan absorb at ~280 nm, whereas phenylalanine has its absorption maximum at ~260 nm. Higher order interactions can also play into the total absorbance, making the method sensitive to pH and high ionic strengths (Olson 2016; Noble and Bailey 2009).

To determine the protein concentration from the absorbance at 280 nm, the molar extinction coefficient ε was estimated using *ProtParam* from the *Expasy* server. After blanking against the buffer solution, 1 µL of protein sample were measured using the *Nanodrop2000* by *Thermo Scientific*. To account for the error susceptibility of the method, the measurement was done in triplicates.

3.2.8 Thermal Shift Assay

The thermal denaturation of proteins can be detected by suitable fluorescent dyes in real-time PCR instruments. This approach can be used to rapidly screen for buffer conditions that enhance stability as well as initial protein:ligand interaction screening (Huynh and Partch 2015). A sufficient stability is important for both biochemical activity and crystallisation (Crowther et al. 2010; Dupeux et al. 2011). The dye, generally SYPRO Orange, undergoes a significant increase in quantum yield upon binding to hydrophobic regions within the protein. These regions become solvent exposed upon denaturation (Figure 15). The excitation and emission wavelength of the dye are at 470 and 570 nm (Niesen et al. 2007). The stabilisation of the protein based on buffer conditions or ligand addition increases the melting temperature and the resulting *thermal shift* quantifies the degree of stabilisation. As real time PCR instruments allow the measurement of up to 96 solutions in parallel, thermal shift is a simple and rapid option for screening a multitude of conditions (Boivin et al. 2013).

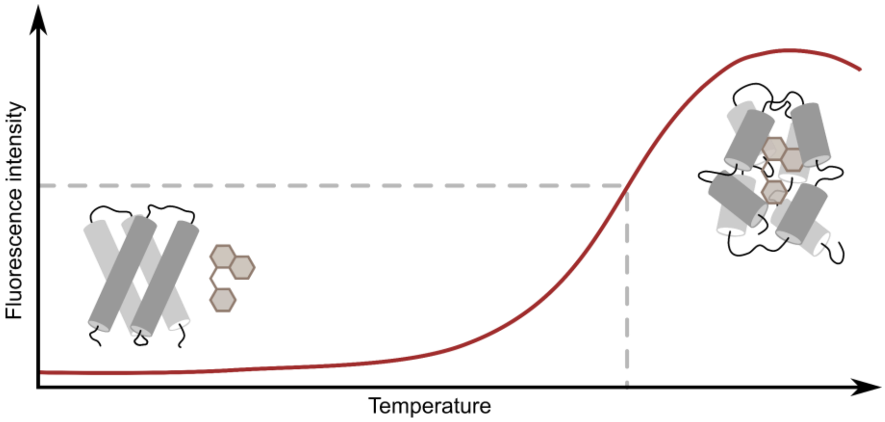


Figure 15: Course of a typical thermal shift assay: At the start of the experiment the protein is intact in its fold and only the fluorescence of free dye can be observed. The temperature is increased in small increments and the protein fold gradually disintegrates. This results in hydrophobic regions becoming accessible to dye binding increasing fluorescence. The melting temperature of the protein is defined as inflection point of its melting curve and can be simultaneously determined for multiple samples using a real-time PCR machine.

To optimize the buffer conditions in this work, a mixture of 20 µL buffer mix, 2.5 µL protein and 2.5 µL SYPRO orange was heated to 75°C in 0.25°C increments. The inflection point was determined to find the melting temperature T_M.

3.2.9 Western Blot

Even though heterologous expression of proteins generally yields significant amounts of the target peptide that are easily identified, occasionally additional methods are necessary to identify protein bands in *SDS-PAGE* gels. A very unambiguous identification can be achieved by analysing gel slices via Mass Spectroscopy, having said that, one is dependent on occupancy of an external core facility in this case. An alternative method that circumvents the need for external resources is the Western Blot. The polyacrylamide gel is electrophoretically transferred to a nitrocellulose membrane, the original band pattern is preserved with no loss in resolution albeit with some losses in quantity (Towbin et al. 1979). Free binding sites of the membrane are occupied with blocking solution containing another protein, usually casein. Thereafter the excess is washed away and the protein of interest can be detected by a suitable antibody (Burnette 1981).

A typical Western Blot detection would involve incubating the nitrocellulose membrane 30 min and subsequently both gel and Whatman paper for about a minute in transfer buffer (50 mM Tris, 40 mM glycine, 0.04% (w/v) SDS, 20% (v/v) methanol). Four layers of Whatman paper, membrane, gel and another four layers of filter paper are stacked and blotted for 1 h at 41 mA per gel (5 V). The membrane is then treated according to Table 1. After ink staining the gel is documented and ruler bands are pencil marked if no prestained marker was used.

Table 1: Western blot procedure: The experiment involves a sequence of incubation steps in alternating solutions summarized below.

Step	Time	Solution
Ink stain	Until stained	2% (w/v) royal blue ink, 1% (v/v) acetic acid
Destain	Until fully destained	0.2 M sodium hydroxide
Rinse	2-3 times	Water
TBS wash	10 min twice	100 mM Tris, 1.37 M NaCl, pH 7.3
Blocking	overnight	Blocking solution purchased from *Biorad*
TBST wash	10 min twice	100 mM Tris, 1.37 M NaCl, 0.05% (v/v) Tween, pH 7.3
Antibody binding	1 h	Anti His-HRP conjugate 1/1000 in blocking solution
TBST wash	10 min twice	
TBS wash	10 min	

The membrane is placed in the imaging system and coated with ECL solution. The resulting fluorescence is detected and overlaid with the full blot documented in the ink staining step.

3.2.10 Fluorescence Anisotropy Assays

When exciting a fluorophore with polarized light the emitted radiation usually is linear polarized as well. As the fluorophore is not fixed in space in solution, rather it moves around according to the concept of Brownian motion, the polarisation angle at the start and end of the fluorescence lifetime changes with it. The shift can be quantified as fluorescence anisotropy by measuring the fluorescence intensity parallel and perpendicular to the excitation polarisation (Gijsbers et al. 2016). Based on Perrins Law the anisotropy is dependent on the volume of the fluorophore which allows detection of binding to larger binding partners by the increase in stokes volume (Pei et al. 2016). Additional factors are fluorescence lifetime, temperature and viscosity of the solution.

5’-6-fluorescein labelled DNA in varying sizes was annealed with its unlabelled counterpart as described in 3.1.13. Varying dilutions of protein (0.001-100 nM) were added to DNA (5 nM) in the presence of cyclic adenosine monophosphate (250 µM) and magnesium chloride (250 µM) and the resulting fluorescence detected in a *Tecan Spark* plate reader.

3.2.11 Isothermal Titration Calorimetry (ITC)

Isothermal titration calorimetry allows for label-free determination of kinetic and thermodynamic parameters of reactions in solution by measuring the heat released during the reaction (Jelesarov and Bosshard 1999). The measurement can be conducted in real time and does not require time-, or dilution-series (Mazzei et al. 2016). ITC instruments operate based on the dynamic power compensation principle, i.e., they measure the amount of power required to maintain a constant temperature in relation to a reference cell. Each injection event triggers the binding or enzymatic reaction that is accompanied by release or absorption (exo-/endothermic) of heat and depends on the concentration of the reactants in the cell (Figure 16). For binding measurements, the system regains equilibrium before the next injection, whereas for kinetic measurements a new baseline is established but the conversion of significant amounts of substrate is avoided to achieve steady-state conditions (Velázquez-Campoy et al. 2004).

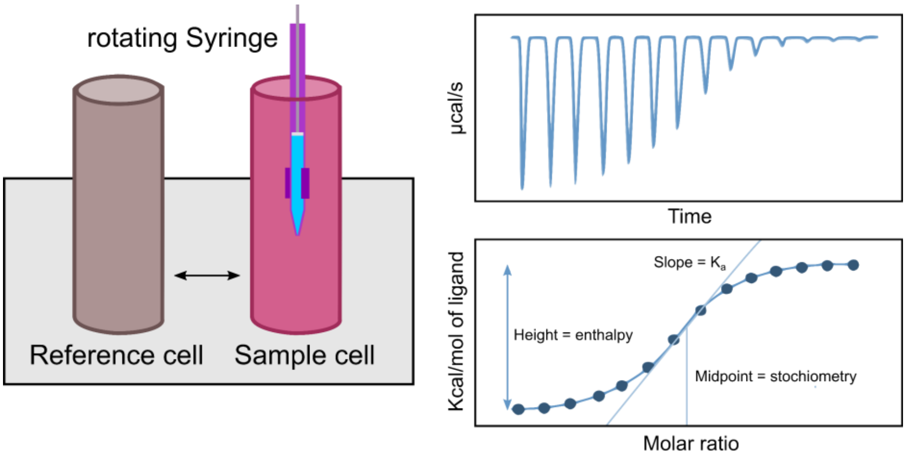


Figure 16: Schematic representation of an ITC setup and measured curves: Changes in heat relative to a reference cell are measured for the addition of small amounts of a binding substance to the reaction partner in the sample cell. From the resulting saturation curve multiple thermodynamic and kinetic parameters can be deduced.

For data analysis of binding events the peak area is integrated and plotted against the molar ratio of ligand and the resulting isothermal curve is then adjusted to fit a suitable binding model. The concentrations need to be adjusted to reach saturation, obtain sufficient signal intensities and minimize dilution effects. For that reason, ligand dilutions in buffer are subtracted as well. The Gibbs free energy ΔG and entropy ΔS are calculated from the fitted isotherm based on:

$$\Delta G = -RTlnK_b = \Delta H - T\Delta S \qquad (1)$$

K_b: binding affinity ($K_D = 1/K_b$); ΔH: enthalpy

Binding kinetics can be measured via ITC as well. The reaction starts by the enzyme E and the substrate S forming a complex ES, which undergoes a transition state ES^* to form the enzyme-product complex EP. The final step is the dissociation to free enzyme and product P. k_{cat} describes the catalytic rate constant, k_1 and k_{-1} the rate of association and dissociation, respectively.

$$\mathrm{E} + \mathrm{S} \underset{\mathrm{k}_{-1}}{\overset{\mathrm{k}_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{\mathrm{k}_{\mathrm{cat}}} \mathrm{E} + \mathrm{P}$$

The formation of an enzyme-substrate complex can be described via the Michaelis-Menten equation (Johnson and Goody 2011) as:

$$K_M = \frac{k_{-1}+k_{cat}}{k_1} \qquad (2)$$

$$v = \frac{V_{\max}\,[S]}{K_M+[S]} = \frac{k_{cat}[E]_{total}\,[S]}{K_M+[S]} \qquad (3)$$

Here v is the rate of the enzyme catalysed reaction, V_{max} describes the maximum velocity when all enzyme is present as ES, [S] is the substrate concentration and K_M is the Michaelis-Menten constant. For low K_M less substrate is required to achieve a certain rate, i.e., substrate affinity is higher. The K_M can be graphically determined as of half the maximum velocity V_{max}. The total heat measured in an ITC experiment is proportional to the enthalpy of all these molecular events, the apparent enthalpy (ΔH_{app}), and to the number of moles of product generated (n), which in turn is given by the total volume multiplied by the concentration of product:

$$Q = n\Delta H_{app} = V[P]\Delta H_{app} \qquad (4)$$

The reaction rate v (d[P]/dt), can be related to the amount of heat generated over the same time (dQ/dt) through:

$$\frac{dQ}{dt} = \frac{d[P]}{dt} V\Delta H_{app} \qquad (5)$$

$$v = \frac{d[Q]}{dt} \cdot \frac{1}{V\Delta H_{app}} \qquad (6)$$

As a result of these correlations the total molar enthalpy and the power generated *(d[Q]/dt)* need to be determined, which can be achieved using two different methods - single injection and multiple injection method.

Multiple injection measurements do not yield ΔH_{app}, which is therefore determined before by adding relatively small amounts of substrate to a large excess of protein and leaving sufficient gaps for the conversion to be completed and the baseline restored. ΔH is then determined according to the binding method. Thereafter the actual kinetic measurement can occur, in which higher amounts of substrate are added to a lower enzyme concentration leaving shorter gaps between injections. Less than 5% of the substrate is depleted prior to the next injection, keeping the reaction in steady-state conditions. The results can be displayed in a Michaelis-Menten plot as a relation of substrate concentration to reaction rate or a Lineweaver-Burk plot as the dependence of *1/v* on *1/[S]* (Lineweaver and Burk 1934; Figure 17).

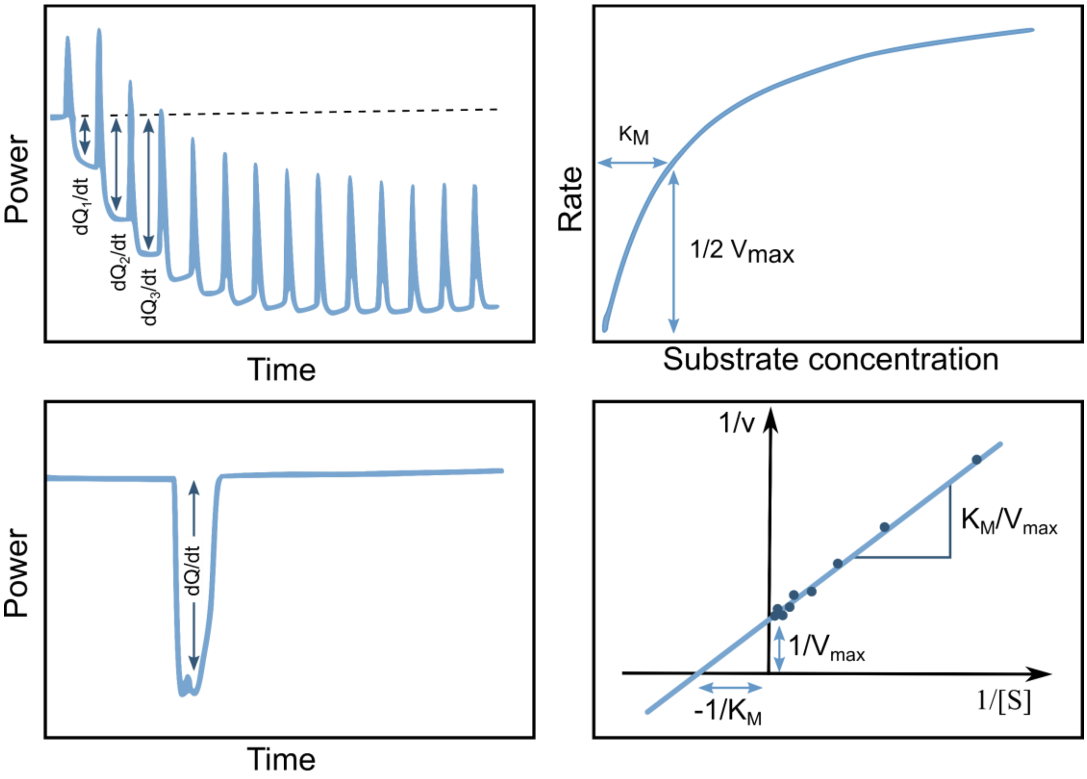


Figure 17: Multiple and single injection experiments for the determination of kinetic parameters can be mapped using either a Michaelis-Menten or Lineweaver Burk plot: Top left: In the multiple injection method small amounts of substrate are injected as soon as a new baseline is established; Bottom left: In the single injection method a large excess of substrate is injected at once and the turnover is monitored. The choice of method depends on the expected K_M range; Top right. Michaelis-Menten plot of an enzymatic reaction. Most automated data analysis programs use this representation; Bottom right: The Lineweaver-Burk plot has the advantage that all parameters can easily be deduced and non-linear behaviour is easily visible.

In single-injection measurements, concentrations of substrate greater than K_M are injected into the enzyme solution. The heat rate is monitored until the signal returns to baseline, indicating that the substrate is completely depleted, and the enzymatic reaction is complete. The ΔH_{app} is measured by the integration of the single peak, with $[S]_0$ being the total substrate concentration (Malvern Panalytical 2020b; Abis et al. 2019):

$$\Delta H_{app} = \frac{1}{[S]_0 V} \int_{t=0}^{t=\infty} \frac{dQ}{dt} dt \qquad (7)$$

$[S]_I$ can be extrapolated at any given time by integrating the heat evolved. Combining this with equation (6) and (7) leads to:

$$v = \frac{d[P]}{dt} = \frac{k_{cat}[E]_{tot}[S]}{K_M} + [S] \qquad (8)$$

The total concentration of the enzyme is quantified in $[E]_{tot}$ and equation (8) allows for the determination of Michaelis-Menten constant K_M, turnover rate k_{cat}, catalytic efficiency $K_{sp} = k_{cat}/K_M$ and ΔH_{app} in a single experiment. The single injection method is suitable for K_M lower than 10 µM, otherwise the multiple injection method needs to be used (Malvern Panalytical 2020a).

The binding affinities of transcription factor *Sm*Clr to cNMP and DNA were measured in this work. A typical titration consisted of injecting 2 µL aliquots of 5 mM of the ligand solution into 0.3 to 0.4 mM of the protein solution every 2.5 min to ensure that the titration peak returned to the baseline prior to the next injection. For the measurement of DNA affinities, the setup was reversed, titrating protein at 0.3 to 0.4 M containing 1 mM of cyclic nucleotide into DNA at 25 µM. The cell was temperature-controlled to 25°C. Titration curves for the dilution of the ligand solution were deducted from the data.

Additionally, the kinetics of the cyclic nucleotide hydrolysis activity of phosphodiesterase *Sm*CpdA were investigated. For single injection experiments 38 µL of 500 µM, 1 mM or 10 mM substrate were added to 1.25 and 6.1 µM protein and the heat was measured 20-70 min. For multiple injection measurements, 10 mM substrate was added in 13 1-2 µL injections to Protein in varying concentrations (615 nM, 1.23 µM or 29 µM). All measurements were conducted at 25°C.

3.2.12 H/D-Exchange Mass Spectrometry (HDX-MS)

Mapping of hydrogen-deuterium exchange in the protein backbone is a powerful tool to investigate protein conformation and dynamics in solution (Weis 2016). Deuterated buffer is incubated with the protein of interest and the reaction is quenched by a shift to acidic pH. Deuteration pattern can then be mapped using either NMR or mass spectrometry. Whereas sidechain proton exchange is way too swift for this approach that involves manual mixing of the sample and deuteration buffer, backbone amide exchange occurs in the seconds to minutes range. The degree of deuteration in a particular location is largely dependent on solvent accessibility. Both folded state of the protein and its dynamics factor in the accessibility for H/D exchange and can thus be monitored using this method (Masson et al. 2019). Other factors like temperature, pH and chemical environment are excluded by calculating the relative uptake compared to an undeuterated reference t_0 (Steinchen et al.). The mapping of deuteration sites involves protease digestion, HPLC peptide separation and mass spectrometric detection of peptide mass. Identical peptides of sample and reference are assigned according to their retention time. Mass peaks are largely assigned to the peptides automatically using the Dynamix HDX data analysis software from *Waters* and manually verified and polished. From the peak shift relative to t_0 the deuterium uptake can be calculated for each individual fragment. The course of a typical HDX-MS experiment is outlined in Figure 18. Averaging all peptides spanning a particular sequence a deuterium uptake heat map can be generated. Protein kinetics are reflected in the peak shape during deuteration. If a particular location is fully protonated in an opening event, two peaks can be observed, each corresponding to the unopened and opened fraction respectively. This exchange behaviour has been termed EX1 kinetics. If multiple opening events are necessary to fully protonate a particular region, a gradual transition to higher m/z over time can be observed. This kinetic behaviour is described as EX2 kinetics (Brown and Wilson 2017).

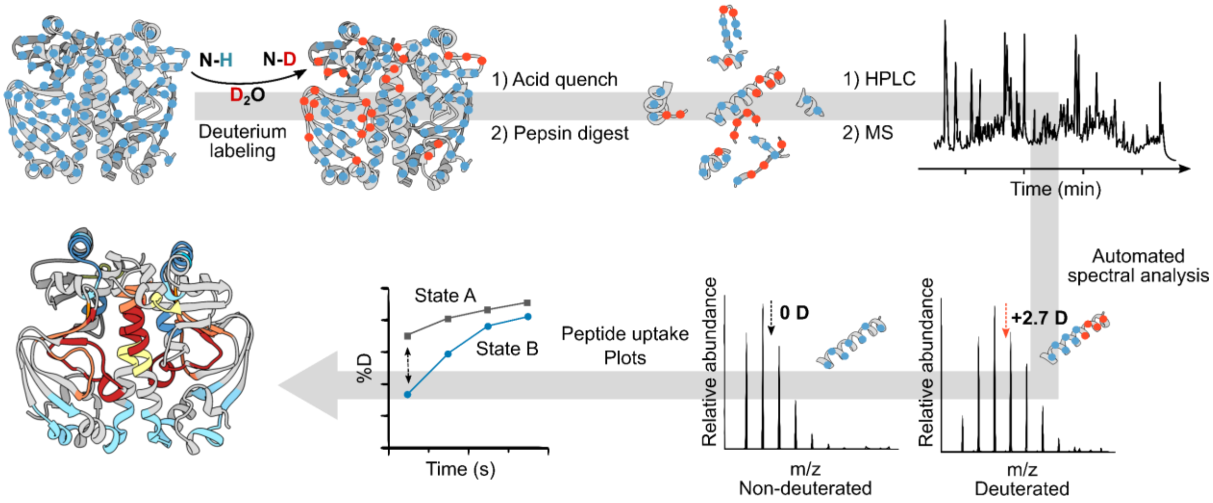


Figure 18: Classic workflow of a HDX experiment: The protein is incubated in deuterated buffer for multiple timepoints. The reaction is then quenched by bringing the solution to a pH of about 2 and the protein is protease digested. The resulting peptides are segregated over HPLC and peptide mass is determined by MS. The mass spectra of protonated and unprotonated species can be assigned according to their retention time on the HPLC column and the mass shift can be deduced and mapped to a protein model or structure. The figure was adapted from Masson et al. 2019.

3.2.13 Nuclear Magnetic Resonance (NMR)

Nuclear magnetic resonance utilizes the interaction of certain sample nuclei interacting with a weak oscillating magnetic field in the presence of a strong static magnetic field (Kubo and Tomita 1954; Rabi et al. 1938). Suitable isotopes have a nuclear spin different from 0, e.g., ^{1}H, ^{13}C, ^{19}F and ^{31}P. The resonance frequency is dependent on the chemical environment via electronic and magnetic interactions with neighbouring atoms. These interactions can also be probed in designated two-dimensional experiments like ^{1}H-COSY (hydrogen correlation spectroscopy), ^{1}H-^{13}C-HSQC (heteronuclear single quantum coherence), ^{1}H-^{13}C-HMBC (heteronuclear multiple bond correlation). The interaction between adjacent nuclei can be assigned based on identical coupling constants within the higher order resonance multiplets. The resonance is determined relative to a reference, usually derived from the solvent, and displayed as chemical shift. It is highly dependent on the chemical environment, e.g. buffer conditions and pH (Holyoak and Nowak 2001).

Phosphorus, hydrogen and fluorine are (nearly) monoisotopic, meaning substances can be investigated without any labelling. Due to the high abundancy of hydrogen and carbon in organic samples, only pure samples below a certain size (generally about 30 kDa) can be probed (Cukkemane and Baldus 2013). For this work nucleotide standards were measured at 5.8 mM in D_2O and assigned. The protein (~160 µM) was incubated with its substrate (2.7 mM) in aqueous solution and investigated as well. Manual measurements and spectra assignments were kindly performed by Dr. Xiulan Xie.

3.2.14 HLPC-Nucleotide Assays

Incubating an enzyme with its substrate, stopping the reaction at a certain time point and subsequent analysis via HPLC segregation is a very versatile method for activity determination. To avoid column clogging the protein preferably is removed prior to analysis. The step is generally conducted via precipitation. Five methods, i.e., heat shock, organic solvent, acid, salt or metal ion addition, are interchangeably used to disrupted the secondary fold. The choice of method depends on the stability of the substrate and products as well as compatibility with the subsequent method (Polson et al. 2003). The resulting mixture can then be segregated on a suitable column. In general, HPLC separations are based on one of three principles, i.e., polarity, electrical charge or molecular size. For separations based on polarity normal phase or reversed-phase chromatography can be used. In normal phase chromatography the stationary phase is hydrophilic while the mobile phase is hydrophobic. This is inverted for inversed-phase segregation. In order to segregate charged analytes on a column like that (n-octylsilyl- C8, noctadecylsilyl- C18), they are incubated with ion pairing reagents, which neutralize their charge (Molnár and Horváth 1976; Nwokeoji et al. 2019). A mobile phase needs to be chosen to retain the analytes of interest on the column but still elute them. The reproducibility and broad applicability of reversed-phase chromatography makes it the standard choice for many experiments.

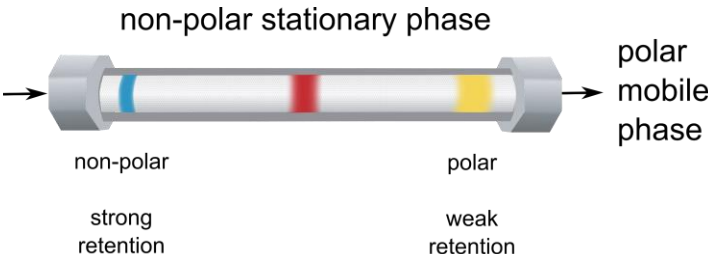


Figure 19: Principle of reversed-phase HPLC. As the stationary phase in RP HPLC is hydrophobic, analytes with little polarity interact with the matrix the most and are travelling less with the mobile phase resulting in long retention times. The opposite holds true for polar compounds that are eluted very quickly.

For analytes like nucleotides ion-exchange chromatography might be an even better choice. Here the segregation is taking place based on charge. Like charge are repelled and have low retention, whereas opposite charges attract and are retained. There are two kinds of stationary phases based on the charge of the analyte: anion exchangers with a positive surface and cation exchangers with a negative one. Another factor is the strength of the surface charge, depending on how strongly charged the analyte itself is (Dai et al. 2020). For nucleotides, strong anion exchangers with quaternary amines on the surface are the best choice. Displacement from the stationary phase is generally achieved by higher ionic strengths. Alternatively, changes in pH can modify the charges and disrupt the interaction, however in strong anion exchangers a pH below 3 would be required. This can also result in analyte degradation via hydrolysis and is therefore not preferred (Jungbauer and Hahn 2009; Williams and Frasca 2001).

Retention time and fragmentation of *Sm*CpdA substrates and products were kindly investigated at a concentration of 10 µM in reversed-phase HPLC-MS by Filipp Bezold and Josephine Menke. The samples were injected into a NanoElute HPLC (Bruker) equipped with a YMC Triart C18 (TA12S03-10Q1PTP) column and eluted with 1.25 mM DBAA, 10 mM NH_4HCO_2, pH 5.2 and acetonitrile/water (90/10), 1.25 mM DBAA, 10 mM NH_4HCO_2 according to Table 2 at a flow rate of 0.2 mL/min. The eluate was injected into an Agilent 6570 Triple Quadrupole for ion detection. All potential adenosine monophosphate isomer mixtures can be distinguished via retention time alone and the subsequent MS/MS detection was eliminated from the protocol as a result.

Table 2: Solvent composition for HPLC nucleotide segregation. Buffer A: 1.25 mM DBAA, 10 mM NH_4HCO_2, pH 5.2; Buffer B: acetonitrile/water (90/10), 1.25 mM DBAA, 10 mM NH_4HCO_2.

Time (min)	A (%)	B (%)
2.00	100.0	0.0
20.00	65.0	35.0
27.00	5.0	95.0
35.00	100.0	0.0

After incubating 20 µM protein with 1 mM nucleotide with or without 0.5 mM additional metal chloride for 30 min at a defined temperature, the protein was precipitated to avoid column clogging and quench the reaction. 150 µL chloroform were added to 50 µL reaction mixture, vortexed for 30 seconds, placed in a heating block at 95°C for the same amount of time and flash frozen in liquid nitrogen. The samples were thawed, spun at 17300 g for 15 minutes and the upper aqueous phase was decanted. For investigation on an anion exchange column, the sample was diluted 1/5 with water. The HPLC assay measurements were kindly conducted by Dr. Wieland Steinchen by high performance liquid chromatography (HPLC) on an Agilent 1260 Infinity system equipped with a Metrosep A Supp 5 - 150/4.0 column. 10 µl of sample were injected and nucleotides eluted isocratically with a flow rate of 0.7 ml/min of 100 mM $(NH_4)_2CO_3$ at pH 9.25 and detected at 260 nm wavelength in agreement with standards.

3.3 Protein Crystallography

3.3.1 Protein Crystallisation

Growing well-ordered protein crystals can be quite the bottleneck for protein structure determination, but when it has been successful, yield atomic resolution models, that are still considered a gold standard in structural biology. The function of proteins is governed by their three-dimensional structure and understanding their structure is essential to the development of therapeutic treatments and protein engineering (Chayen 2004). Factors making this particular task difficult are mainly size and conformational flexibility of proteins and even more so of biomolecular complexes (Gavira 2016). Many proteins are only ordered in the presence of an interaction partner, which means they might not crystallize by themselves (Dale et al. 2003). A lot of factors leading to well-diffracting crystal cannot be predicted and need to be optimized empirically (McPherson and Gavira 2014). This does not mean that no trends can be deduced from sequence similarity data, often times similar proteins can be crystallisation in akin conditions (Lu et al. 2012).

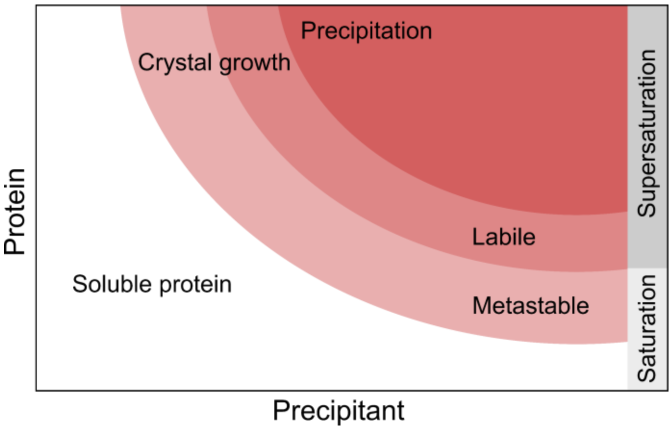


Figure 20: Principles of protein crystallisation: To enter the nucleation zone, where small crystal seeds appear the protein needs to be in a supersaturated state. The solubility of the protein can also be altered by the addition of precipitant. Too high concentrations lead to protein precipitation, which generally is irreversible and unwanted. To achieve well ordered crystals, nuclei should form in the metastable phase facilitating subsequent crystal growth.

To achieve crystal growth generally a homogenous pure sample in a stabilising buffer environment is necessary (Gavira 2016; Iwai et al. 2008). The crystallisation process is comprised out of three steps, namely nucleation, growth and growth termination (Figure 20). The two main techniques to protein crystallisation are vapour diffusion and batch crystallisation. In vapour diffusion a reservoir of precipitant solution is sealed in a well containing a mixture of precipitant and protein solution. Through diffusion the osmolarity is evening out over time, ideally reaching the nucleation zone of the protein, i.e., supersaturation. The concentration needed to reach this state depends on the solubility of the protein (Dessau and Modis 2011). Other factors that influence the success of the crystallisation attempt are protein/precipitant ratio, temperature, pH and the addition of various additives. To make things even more difficult, overshooting often results in a showering of small crystals and/or low-quality crystals. The conditions are then subsequently optimized to slow nucleation and crystal growth by reducing the concentrations involved. Very good overviews over the strategies that can be employed for crystallisation condition optimisation were given by McPherson and Cudney 2014 and Müller 2017 with the latter focussing particularly on protein-ligand complexes. An important approach to improving crystal quality is microseeding, the addition of a homogeneous or heterogeneous nucleant, that lowers the nucleation energy barrier (Fermani et al. 2013; Saridakis and Chayen 2009). Typically, subpar crystals from an initial hit are crushed and diluted (1:50-1:2000) and then added to a new crystallisation screen. Batch crystallisation starts in a slightly undersaturated state and crystallisation is achieved by either cooling, solvent evaporation or precipitant addition. The method allows for large scale crystal production and is therefore often used for material-consuming methods like X-Ray Free-Electron Laser measurements (XFEL) (Lewis et al. 2015). In some cases crystal quality can be further improved post-crystallisation by dehydration (Heras and Martin 2005).

All crystals in this work were grown in vapour diffusion setups. Initial screening was conducted in commercial 96-well screening solutions using a sitting drop approach in *Swissci MRC 2 Lens Microplates*. The drop volume was 2 µL of a 1:1 mixture of protein and crystallisation solution, the pipetting was conducted with a *Honeybee 963* from *Digilab*. The conditions were further optimized in a hanging drop approach in *NeXtal EasyXtal 15 plates* with a drop volume of 1-2 µL. Homogenous microseeds were prepared according to the seeding protocol from Hampton Research 2020 and dilutions ranging from 1:50 to 1:2000 were added in some cases. 96-well plates were screened at 4 and 18 or 8 and 20°C respectively in a *Formulatrix Rock Imager*, whereas *Nextal* plates were monitored manually. Crystals were harvested using *DualThickness Micromounts* from *Mitegen* or *CryoLoops* from *Hamptons Research* and cryoprotected with 20% glycerol as needed.

3.3.2 Soaking of Protein Crystals

Macromolecular crystals can be functionalized by soaking metal-ions or ligands into the pre-existing crystals. An advantage of this strategy is that it allows for screening a multitude of potential ligands and consumes relatively little material as the crystals can already be optimized and tested for diffraction before functionalisation. However, there are several factors to consider. Soaking time and ligand concentration need to be optimized and protein crystals can be sensitive to solvents used to dissolve the ligand, often DMSO. The ligand needs to be able to enter the crystal via interstitial spaces and solvent channels (Tari 2012). Ligand binding can be accompanied by larger structural changes, causing crystals to crack or dissolve as the complex dimensions change (Hassell et al. 2007). Heavy metals used for phase-determination (see *Phasing*) increase the radiation damage the crystal is subjected to during the diffraction experiment and crystals are often times “back-soaked” in metal-free buffer to remove excess metal (Dauter and Dauter 2017). To stabilize crystals during soaking buffers with increased concentrations of precipitants or gradually increased ligand content can be used. The cryo-protection of protein crystals by immersing them in a solution containing glycerol or polyethylene glycol in concentrations above 30% can be considered soaking as well.

3.3.3 X-ray Diffraction and Data Collection

Atoms scatter X-ray waves mainly through their electrons. In an ordered crystal lattice of identical protein molecules this results in specific diffraction pattern and the location of the scattering atoms can be deduced from the angles and intensities of the reflexes. A three-dimensional electron density map can be generated representing the mean position of the atoms. The atomic resolution makes X-ray crystallography one of the most prevalent means of structure determination. The atoms within protein crystals are not static, but oscillate about their mean positions. Additionally, protein conformations within the crystal can deviate and the data will reflect their mean position. Alternative methods are cryo-electron microscopy (cryo-EM) and small angle X-ray scattering (SAXS).

A single monocrystalline crystal is mounted on a goniometer that allows moving the crystals in different planes. Generally, the ω (omega) translation moves the crystal into the beam, the ω (omega) rotation moves perpendicular to the beam, κ (kappa) rotates at about 50° to ω and φ (phi) controls the torsion in loop direction. The crystal is kept frozen, which reduces radiation damage, dehydration and noise due to thermal motion. A monochromatic X-ray beam is scattered on the crystal lattice producing a diffraction pattern consisting of spots called reflexions. The stronger the diffraction and the higher the resolution, the further the reflexions deviate from the original beam direction. The detector distance is adapted accordingly, moving closer for high resolutions and further away for lower. During the measurement, the crystal is rotated to record different orientations of the lattice and avoid radiation damage within the data (Figure 21).

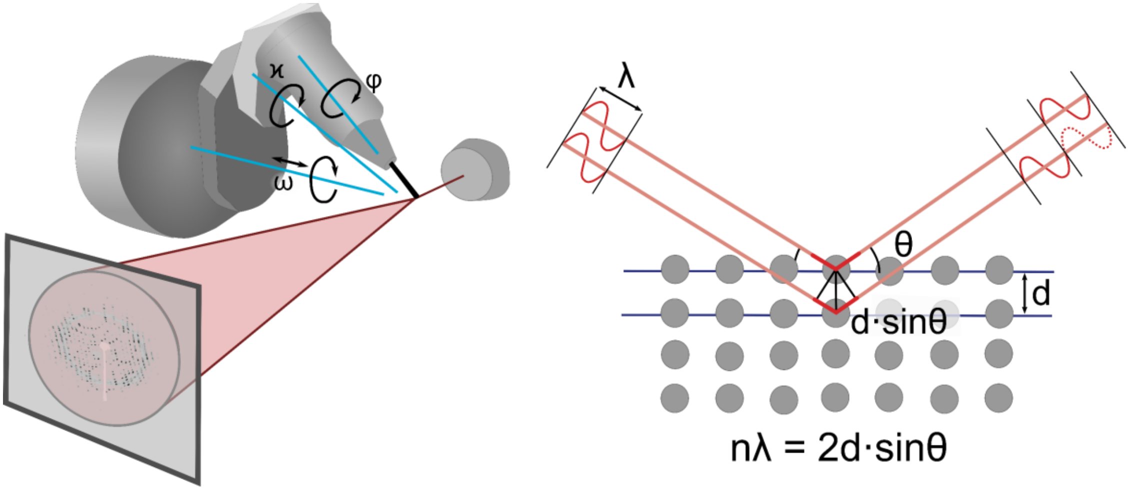


Figure 21: Data collection for protein crystals: Crystals are magnetically mounted on a goniometer allowing orientation in three different orientations. The beam is scattered on the crystal lattice following Bragg’s Law (n: integer; λ: wavelength, d: interplanar distance, θ: glacing angle) and the reflexions are captured on a detector in line with the beam. The crystal is rotated during data collection to avoid radiation damage and capture the lattice in different angles. The detector distance is adjusted according to the diffraction quality.

X-ray beams are used to produce the diffraction pattern, as their wavelength is in the same range as the diffraction plane spacing d (1-100 Å). Synchrotron radiation provides one of the brightest light sources available by accelerating electrons on a circular trajectory and focussing the tangentially emitted electromagnetic radiation to a single monochromatic and monophasic beam. The radiation is elastically scattered upon free electrons, that oscillate in the same frequency and produce a secondary wave out of phase by 180°. The resulting waves interfere constructively in specific directions, determined by Bragg's law, when their wavelength are multiples of each other:

$$n\lambda = 2d\sin\theta \qquad (9)$$

The same relationship has also been described by Laue using lattice vectors. The concepts are interchangeable.

Multiple crystal orientations need to be collected to reconstruct the whole crystal, the angle required depends on the crystal symmetry (usually 90-180°). The crystal lattice is determined from the position of reflexions, whereas the molecular structure needs information on intensity as well. The planes intersecting the unit cell of the lattice are defined by Miller Indices defining at which point the plane intersects the unit cell. Negative indices (marked with a line above the number) show a plane opposite to the unit cell. The diffraction pattern and the lattice are related by Fourier transformation, the diffraction pattern representing the corresponding reciprocal space. The lattice distances orthogonal to the beam are inversely proportional to the spot distances. Using these relations the unit cell parameters can be deduces from the spot locations (Parker 2003; Dauter 2017).

3.3.4 Processing of Diffraction Data

The processing of raw diffraction images involves finding and indexing recorded diffraction spots, determination of cell parameters and space group and integration of the reflexes accordingly. The automated pipelines like *XDS* or *iMosflm* also correct geometrical distortions and general noise on the detector. A comprehensive overview over the XDS package used in this thesis was given by Kabsch 2010.

The beamline-generated XDS.INP files were used to start the processing in XDS Version Mar15, 2019 and Jan31, 2020. H5-compressed files generated by the EIGER 16M detector at SLS were unpacked using the Neggia plugin from *Dectris*. The program sequence XYCORR, INIT, COLSPOT, IDXREF, DEFPIX, INTEGRATE and CORRECT was run, occasionally image range and space group were modified to improve processing and data quality. The data in the .HKL output file was scaled and anisotropically sharpened using Aniso Scale on the

Anisotropy Server (UCLA) based on the signal to noise ratio of the structure factor (F/sigma) and exported as .MTZ structure factor file. If very little anisotropy could be detected the datasets were truncated according to $CC_{1/2}$ (≥0.5) and the mtz files were generated using CCP4i2 (Karplus and Diederichs 2015). For new datasets a R_{free} set from 5% of the data was generated, to assess the agreement between crystallographic model and experimental data. For any further datasets the original R_{free}-set was inherited using CCP4i (The CCP4 suite: programs for protein crystallography 1994).

3.3.5 Phasing

The intensity of the diffracted beam is proportional to the square of the structure factor amplitude. Structure factor F describes the distribution of the electron density within the unit cell as:

$$F(hkl) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(xyz) \exp[2\pi i(hx + ky + lz)]\, dx\, dy\, dz \qquad (10)$$

x, y and z are fractional coordinates in the unit cell (0 ≤ x/y/z < 1) with a volume V, ρ(xyz) describes the electron density at a certain position and the term 2πi(hx+ky+lz) describes the phase of the resulting wave. However, the goal is not the prediction of a diffraction pattern from an electron density. F(hkl) is not only the Fourier transformation of ρ(xyz), but the opposite holds true as well:

$$\rho(hkl) = \frac{1}{V} \sum_h \sum_k \sum_l F(hkl) \exp[-2\pi i(hx + ky + lz)] \qquad (11)$$

According to the Laue conditions, i.e., diffraction only occurs in discrete directions depending on the phase angles α(hkl), the term has to be modified to:

$$\rho(hkl) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)] \qquad (12)$$

One information is inevitably lost within the scattering experiment: the phase of the wave upon crystal interference. There are a few strategies to reconstruct the missing information, generally termed “the phase problem”. Two of the most commonly used ones are molecular replacement and anomalous diffraction. The former is a pretty straight forward method but requires the availability of a suitably homologous model structure from which the phase information is acquired. If this does not exist, anomalous diffraction is usually the method of choice (Messerschmidt 2007).

If an anomalous scatterer, generally a heavy metal due to its high electron density, is present in the unit cell the intensities of the reflections are no longer following Friedel’s law of being centrosymmetrical, i.e., ($F(\mathrm{h\,k\,l}) \neq F(\bar{\mathrm{h}}\,\bar{\mathrm{k}}\,\bar{\mathrm{l}})$). From the structure factors of a native and a strong enough anomalous signal in the dataset of a suitable derivative the phases can be calculated. The anomalous dataset is collected as close to the absorption edge of the scatterer as possible, to obtain the strongest signal. Additionally, multiple datasets can be merged. The method is known as single-wavelength anomalous diffraction (SAD) (Drenth 1999; Messerschmidt 2007).

In molecular replacement a Patterson map is generated via Fourier transformation of the reflection intensities. This is then compared with a Patterson map derived from a known homologous structure and the orientation is optimized in regards to rotation and translation and the phases are derived from the final oriented phase model (Drenth 1999).

The truncated structure factors generated using Aniso Scale were used to start molecular replacement runs in Phenix 1.16 (Liebschner et al. 2019) with search models generated on the SwissModel server (Waterhouse et al. 2018).

3.3.6 Structure Refinement

Once the initial phases have been obtained, the model needs to be adjusted to better fit the experimental data. The process is called refinement. The agreement between calculated and observed structure factors can be assessed by the *R* factor:

$$R = \frac{\sum_{hkl} \left| |F_{obs}| - k|F_{calc}| \right|}{\sum_{hkl} |F_{obs}|} \times 100\% \qquad (13)$$

A *R*-factor of 59% would correspond to a random structure, but factors up to 50% aren't uncommon for low resolution structures (Wilson 1950). The refinement step involves the variation of coordinates within angle and distance restraints in real and reciprocal space, as well as B factors and occupancies. It is alternated with manual model building steps in COOT (Version 0.8.9.). In the refinement process the working model is parameterised in regards to bond lengths, binding and torsion angles (Headd et al. 2012). New phases are derived and therefrom new density maps are calculated and compared to the observed data via the *R*-factor (R_{work}). To avoid introducing bias into a structure when refining for the best fit and then assessing with the resulting density, about 5% of the experimental data is removed and not refined but only used in assessment - the resulting factor is called R_{free}. As an obvious result to the way they are calculated, R_{free} will always be higher than R_{work}. Another way to avoid bias in refinement is to always refine against the original experimental electron density map (Kleywegt and Alwyn Jones 1997).

The quality of a refined structure is also rated according to multiple conformational and model-dependent factors. Any angles and bond lengths outside the standard restraints are listed, as well as outliers in backbone torsion (Ramachandran outliers, Ramachandran et al. 1963). The clashscore highlights atomic overlaps. There are programs like MolProbity available (and recently implemented in the protein data bank PDB) to help quickly assess the quality of data (Williams et al. 2018).

All refinements were conducted in phenix refine with secondary structure restraints in case of the low-resolution structures of Clr. For structures of the same protein that crystallised with the same cell parameters, the original structure was refined firstly as a rigid body against the electron density. The PDB Redo server was finally utilized as a polishing step (Joosten et al. 2014).

4 Results

4.1 *Sm*Clr: a bifunctional cNMP receptor protein

To elucidate the structural basis for the unique feature of cAMP receptor protein Clr to initiate transcription in response to both cAMP and cGMP, a nucleotide-bound crystal structure in complex with promotor DNA was sought-after and the interaction was to be characterized. In previous experiments by Petra Gnau an increased stability of *C*-terminally tagged Clr protein was observed, likely by preventing proteolysis. This issue had been described for $\mathrm{FixK_2}$, a related transcription factor from *Bradyrhizobium japonicum*, as well (Bonnet et al. 2013). The properties of the recombinant Clr encoded in the pET36b-*Sm*Clr $\mathrm{His_8}$ plasmid kindly provided by Petra are listed below.

Name	**UniProt-ID**	**Length** (aa)	**Properties**	**pI**	**MW** (Da)	**ε** ($\mathrm{M^{-1}cm^{-1}}$)	**Sequence**
Clr	Q92SD2	244	C-terminal $\mathrm{His_8}$	5.36	27128	17210	native

4.1.1 Expression and purification of *Sm*Clr

For heterologous overproduction of Clr, cells were grown in LB/kanamycin (35 µg/mL) and expression was induced by addition of 0.1 mM IPTG at an optical density of 0.5. The cells were harvested after 3 h at 30°C via centrifugation at 4000 rpm (F6S 6x1000Y) for 15 min. The pellet was resuspended in binding buffer and flash frozen.

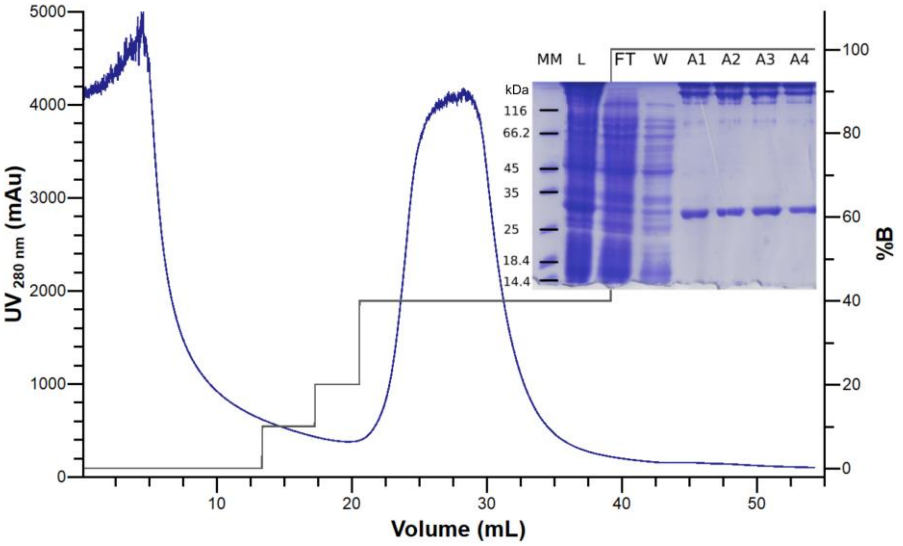


Figure 22: Affinity purification of *Sm*Clr: Filtered lysate was applied to a 5 mL Ni-NTA column using a peristaltic pump and eluted at a flow rate of 2 mL/min using a stepwise gradient in imidazole

concentration on a FLPC system. The elution was monitored using UV/VIS. The apparently high noise at absorptions above 4000 mAu is caused by detector overload. Different stages of the purification were applied a 12 % SDS-PAGE gel. Clr at about 30 kDa is initially present in the lysate (L). The column was overloaded slightly, as some protein remains in the flowthrough fraction (FT). In the washing step (W) quite a few non-specifically binding proteins eluted from the column. In the final elution (A1-4) Clr could be isolated, aside from a few impurities of high molecular weight (possibly aggregated protein), which could not be removed in any case.

Cells corresponding to 1.6 L culture were sheared in a French press pressure cell. The resulting lysate was spun at 18000 rpm (JA-20) for one hour and the supernatant was taken off the pellet. The supernatant obtained after lysis was initially purified using Ni-NTA chromatography (Figure 22) followed by a size exclusion chromatography polishing step (Figure 23). The shoulder in the elution profile results from small amounts of a dimeric species, which is not stable for more than 48 h. Only the monomeric peak was used in the subsequent experiments. The final yield per litre culture amounted up to 30 mg. The SDS-PAGE gel shows a higher molecular impurity, that in mass analysis was matched with being Clr as well. The addition of DTT to the SEC buffer shifts the equilibrium to the monomer peak in the chromatogram, yet the appearance on SDS-PAGE remains the same. As the impurity could not be removed and did not seem to impact behaviour in solution as well as crystallisation, further optimisation was omitted.

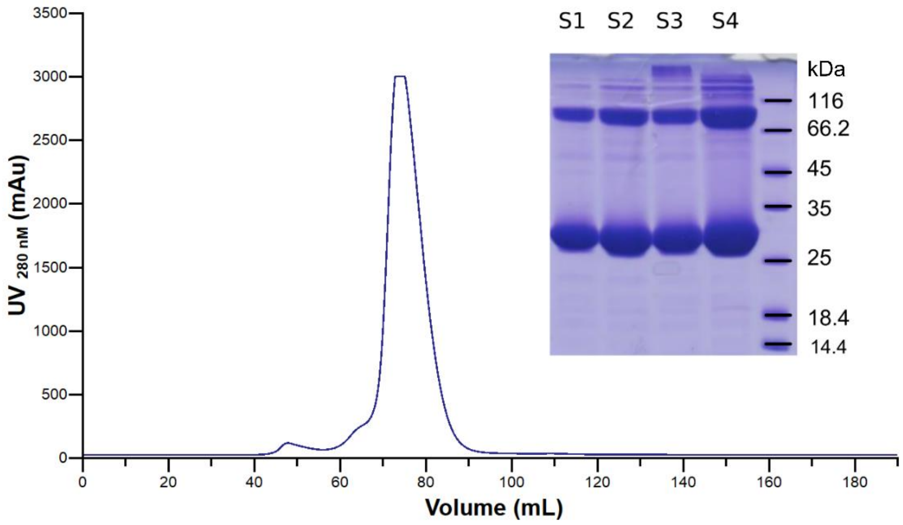


Figure 23: Size exclusion chromatography run for Clr: As polishing step, Clr was applied to a S200 16/600 Superdex column at a flow rate of 1 mL/min. The protein mainly elutes in a monomeric species, with a small amount of dimer corresponding to the shoulder left of the main peak. The main peak fractions were pooled and used in further experiments. An 12% SDS-PAGE gel of fractions from the

monomeric peak show purified Clr with the same high-molecular non-detachable impurities as in the IMAC step.

The stability of Clr in solution was optimized using a Thermal Shift Assay. A HEPES-based buffer containing 300 mM sodium chloride at a pH of 7 was chosen as new size exclusion buffer. The Ni-NTA binding and elution buffer pH were adjusted accordingly.

Buffer	**Content**	**Tm [°C]**
1	20 mM Tris, 200 mM NaCl, pH 8.0	57.75
2	20 mM Tris, 300 mM NaCl, pH 8.0	58.50
3	50 mM HEPES, 300 mM NaCl, pH 7.7	58.75
4	20 mM HEPES, 300 mM NaCl, 5% glycerol, pH 7.7	58.90
5	20 mM HEPES, 300 mM NaCl, pH 7.5	59.00
6	20 mM HEPES, 300 mM NaCl, pH 7.0	59.00
7	20 mM Tris, 300 mM NaCl, pH 7.5	58.75

4.1.2 Clr affinity for cGMP and cAMP levels out in the presence of DNA

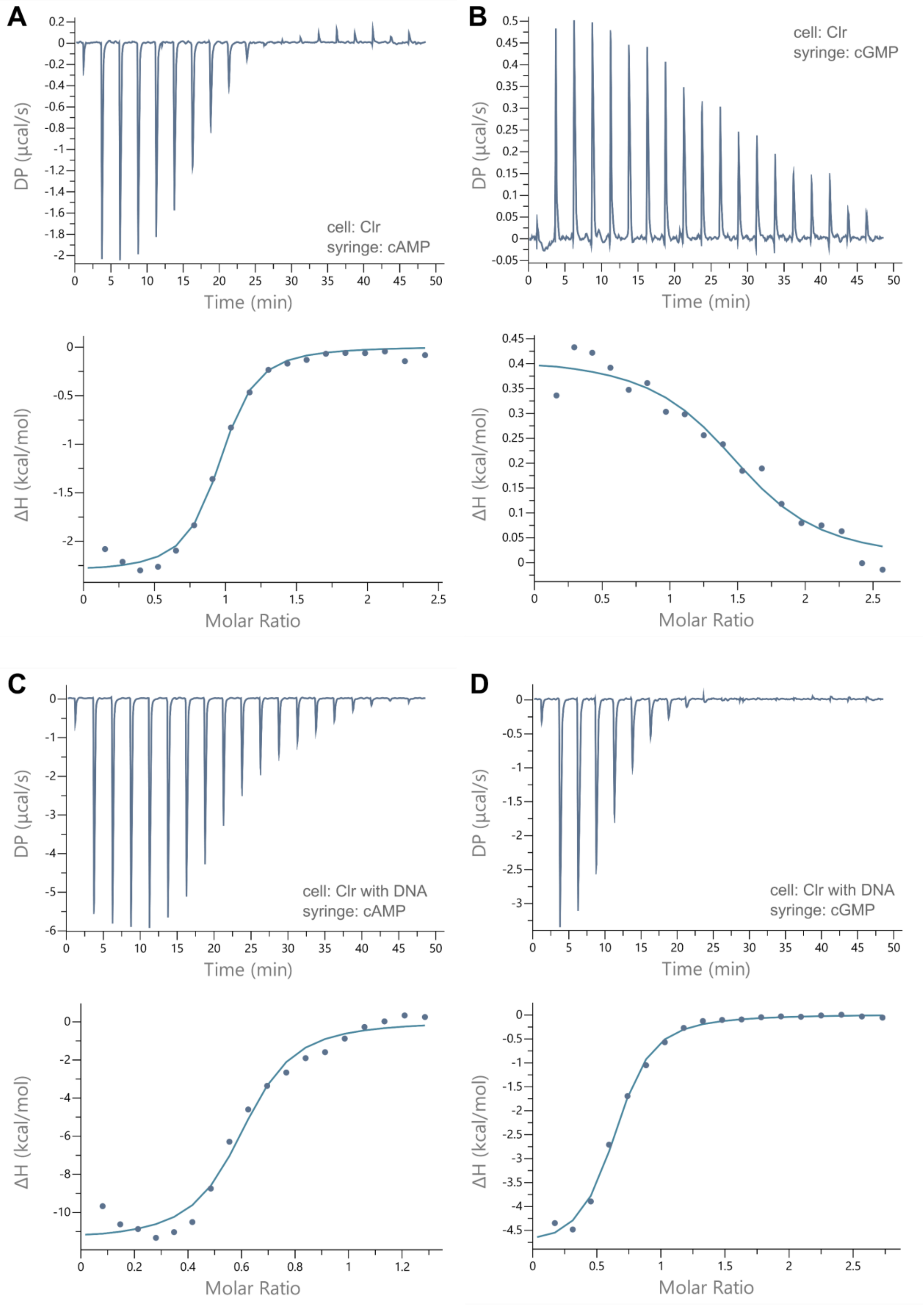
In order to assess the reason for different levels of cAMP and cGMP-dependent transcription regulation *in vivo*, the affinity of the different transcription activation complex components to each other was to be determined. The recognition sequence of Clr was purchased as two single-stranded DNA fragments, that are complementary and align to a symmetrical overlap (Krol et al. 2016). The DNA solution was heated to 98°C and then cooled overnight within the metal block of the thermocycler to avoid mismatches. The result is a 33 base pair duplex DNA containing a double strand break, which was also used in the crystallisation experiments. Initially, fluorescence anisotropy measurements with 5'-fluorescently labelled DNA fragments of varying length were conducted. No affinity at all could be observed using this method, likely due to label interference. The experiment is elaborated further in the Appendix under 8.2 Clr-DNA affinity measurement via fluorescence anisotropy. The binding behaviour of Clr could be investigated using isothermal titration calorimetry (ITC) instead.

The protein was transferred to the binding buffer containing 20 mM magnesium chloride via PD-10 column, as it had been found to be magnesium-dependent by Ankan Banerjee. The titration calorimetry measurements were performed at a Malvern MicroCAL PEAQ-ITC as previously described by Wiseman et al. (Wiseman et al. 1989). A typical titration consisted of 2 µL injections of 5 mM of the ligand solution into 0.3 to 0.4 mM of the protein solution every 2.5 min to ensure that the titration peak returned to the baseline prior to the next injection. For the measurement of DNA affinities, the setup was reversed, titrating protein at 0.3 to 0.4 mM containing 1 mM of cyclic nucleotide into DNA at 25 µM. The setup reversal reduces the amount of dilution heat for the very polar DNA and resulted in heat curves with less noise. The cell was temperature-controlled to 25°C. Titration curves for the dilution of the ligand solution were deducted from the data using the MicroCal PEAQ-ITC Analysis software from *Malvern Pananalytical*.

Table 3: Thermodynamic parameters for cNMP-mediated Clr-DNA interaction: The values were derived from triplicate measurements using the MicroCal PEAQ-ITC Analysis Software from *Malvern Pananalyticals*. The raw data were measured at 25°C. Further details regarding concentrations and setup can be found in Figure 24 below.

Cell	Syringe	K_D (µM)	n	ΔH (kcal/mol)	ΔG (kcal/mol)	-TΔS (kcal/mol)
Clr	cAMP	5.8 ± 1.4	1.02 ± 0.016	-2.22 ± 0.05	-7.16 ± 0.12	-4.94 ± 0.22
Clr, DNA	cAMP	6.7 ± 2.7	0.57 ± 0.019	-11.5 ± 0.77	-7.08 ± 0.16	-4.44 ± 0.36
DNA	Clr, cAMP	6.0 ± 4.7	0.39 ± 0.201	-11.0 ± 7.30	-7.13 ± 0.10	-3.84 ± 1.37
Clr	cGMP	23.8 ± 14.5	1.49 ± 0.081	0.29 ± 0.02	-6.31 ± 0.05	-6.60 ± 0.07
Clr, DNA	cGMP	10.7 ± 3.9	0.57 ± 0.022	-4.91 ± 0.02	-6.79 ± 0.09	-1.88 ± 0.19
DNA	Clr, cGMP	4.5 ± 12.4	0.45 ± 0.408	-3.47 ± 5.33	-7.30 ± 0.02	-3.83 ± 0.02

Every measurement was done in triplicates. After definition of the baseline, the peaks of the thermogram were integrated in the software and plotted against the ratio of ligand. The resulting data points can then be fitted to a curve consistent with the binding mode. In this case a model assuming one binding site per molecule fits the raw data, no cooperativity between the binding sites could be observed. The ratio n between the complex components can be deduced from the inflection point of the binding curve, while the slope at this position reflects the equilibrium dissociation constant K_D, which relates to the binding constant as $1/K_b$. The binding enthalpy ΔH can be deduced from the ordinate of the binding curve. Knowing these values, the *Gibbs-Helmholtz* equation can be used to determine the *Gibbs* free energy and entropy of the binding. While the presence of DNA did not seem to impact the affinity of cAMP to the protein, there are obvious changes in cGMP binding. While the effector affinity in the absence of the binding sequence seemed to be greater for cAMP, when the DNA oligomer was added, the difference decreased to equal strength in cAMP and cGMP binding. Clr might be able to interact unspecifically with the DNA, priming it for nucleotide interaction by a shift in the apo-holo equilibrium, pre-stabilising the latter to be more readily populated. The affinity of nucleotide bound Clr for DNA is in the same range for both cyclic AMP and GMP. For Clr alone no binding could be detected in ITC. The stoichiometry decreases when DNA is present in the experiment, however in the crystal structures both subunits bind a nucleotide molecule. Cyclic GMP binding to Clr alone is endothermic and therefore entropically driven.


A
DP (µcal/s)
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
-1.6
-1.8
-2
cell: Clr
syringe: cAMP
0
5
10
15
20
25
30
35
40
45
50
Time (min)
ΔH (kcal/mol)
0
-0.5
-1
-1.5
-2
0
0.5
1
1.5
2
2.5
Molar Ratio
B
DP (µcal/s)
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
cell: Clr
syringe: cGMP
0
5
10
15
20
25
30
35
40
45
50
Time (min)
ΔH (kcal/mol)
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.5
1
1.5
2
2.5
Molar Ratio
C
DP (µcal/s)
0
-1
-2
-3
-4
-5
-6
cell: Clr with DNA
syringe: cAMP
0
5
10
15
20
25
30
35
40
45
50
Time (min)
ΔH (kcal/mol)
0
-2
-4
-6
-8
-10
0
0.2
0.4
0.6
0.8
1
1.2
Molar Ratio
D
DP (µcal/s)
0
-0.5
-1
-1.5
-2
-2.5
-3
cell: Clr with DNA
syringe: cGMP
0
5
10
15
20
25
30
35
40
45
50
Time (min)
ΔH (kcal/mol)
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
-4.5
0
0.5
1
1.5
2
2.5
Molar Ratio

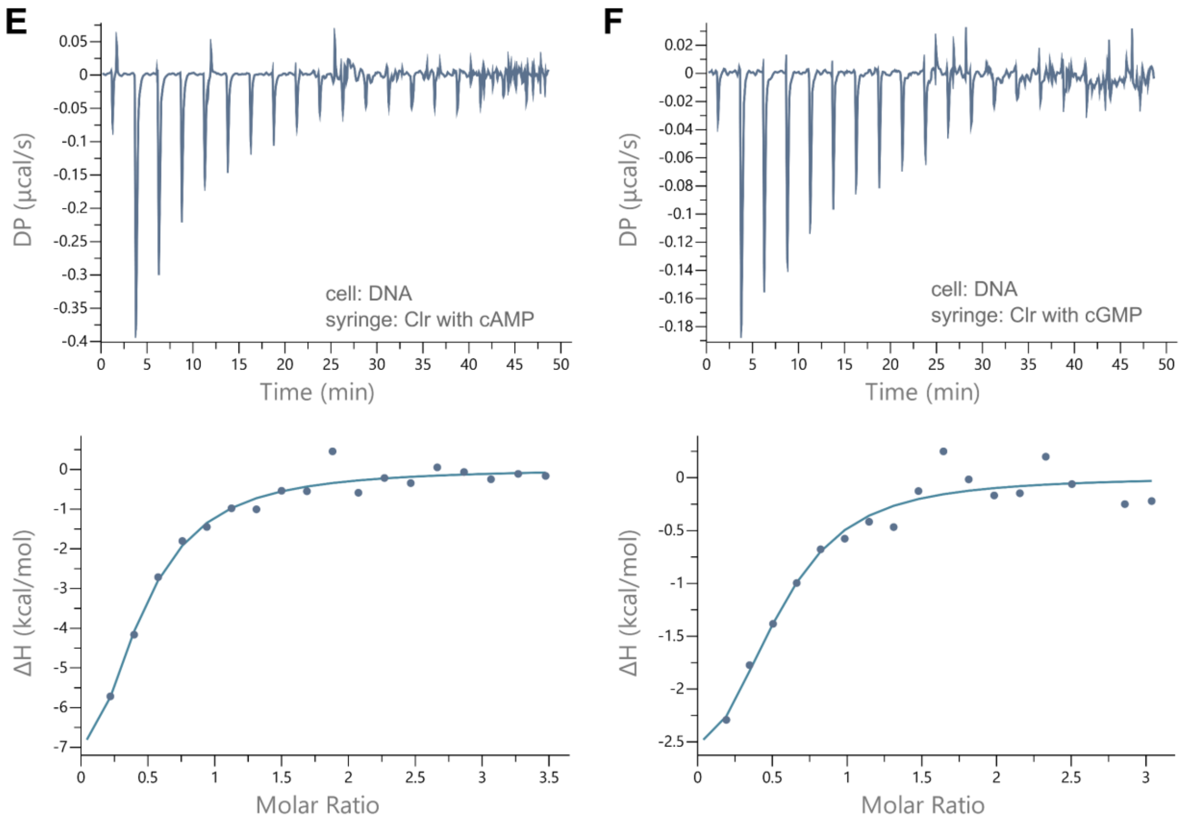


Figure 24: Raw and integrated ITC measurements for the interaction of Clr with cNMP and duplex DNA: (A, B) To probe the binding of cAMP and cGMP to Clr, the 280 µL of protein were placed into the cell at a concentration of 300-400 µM and a sequence of nucleotide injections at 5 mM with 2 µL each were performed. The injections were spaced 180 s apart and the cell was temperature-controlled to 25°C. For cAMP the binding is exothermic, whereas for cGMP it is actually endothermic. Measurements were conducted in triplicates and titration curves for ligand dilution into buffers were deducted. The integration and data analysis were performed using the *Malvern Pananalytical* MicroCal PEAQ-ITC Analysis software. (C, D) The binding of cNMP to Clr is influenced by the presence of DNA. The binding for cyclic GMP is switched from an endothermic to an exothermic process. To quantify this, Clr at 300-400 µM was preequilibrated with a 1.25 excess of the duplex DNA used in the crystallisation experiments and placed in the cell. The injections were conducted as in (A) and (B). Here again, triplicates were measured and a blank for the titrant dilution was deduced before data analysis according to the same protocol. (E, F) As the dilution of the polar duplex DNA results in a lot of overlaying heat signals, the setup was reversed for the measurement of DNA affinity. The protein at 300-400 µM was equilibrated with 1 mM cyclic nucleotide and titrated into the duplex DNA with a concentration of 25 µM. The signal for this measurement is weaker, which is also reflected in the higher noise. The triplicate measurements were processed as described above.

4.1.3 Crystallisation and structure determination cAMP-activated *Sm*Clr bound to DNA

Contrary to CAP, its *E. coli* homolog, Clr is activated by cGMP as well as cAMP. To shed light on the structural features this bifunctional behaviour stems from, co-crystals of the DNA complex were to be achieved. Clr is not overly stable in solution, which is why the protein was freshly purified for any crystallisation attempts. The purified protein was concentrated with a cut-off of 3 kDa to a concentration of 5.9 mg/mL (217 µM). Prior to crystallising the Clr transcription activation complex, said complex has to be reconstituted in solution. A slight excess (1.25x, 271 µM) of the same annealed duplex DNA, that had been used in the ITC measurements, as well as a 2.5-fold excess (543 µM) cAMP and magnesium chloride were added to the freshly purified protein to create the crystallisation batch. Different crystal morphologies for Clr-DNA co-crystals bound to cyclic AMP are shown in Figure 25.

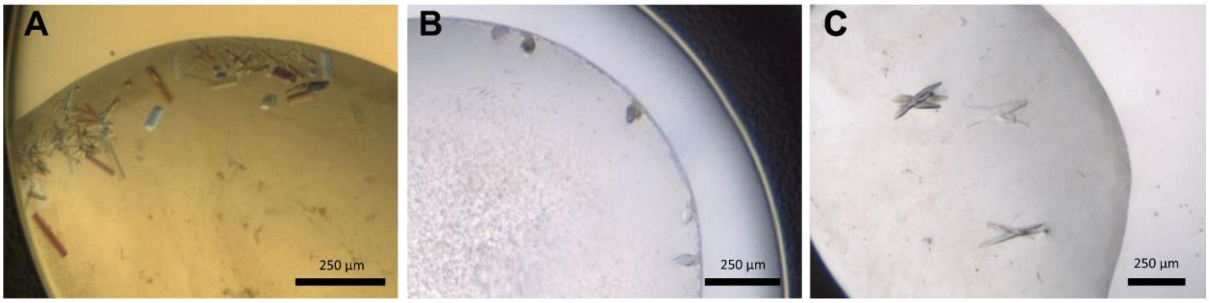


Figure 25: Morphologies of cAMP activated Clr-DNA cocrystals: Crystals appeared within hours in solutions containing the same MPD and PEG-based precipitant mix and alternating buffers and additives and had different morphologies. (A) Initial small needles diffracted anisotropically to 3.5, 3.8 and 2.9 Å and were grown in a solution of MOPS/HEPES at pH 7.5 and a mix of short chain alcohols and diols (*Molecular Dimensions* Morpheus Crystallisation Screen, D8). (B) Diamond shaped plates were grown in the presence of bicine/Tris pH 8.5 and a mix of halides (*Molecular Dimensions* Morpheus Crystallisation Screen, B12). These crystals yielded the final solution. (C) The addition of an ethylene glycol mix to a solution containing MOPS/HEPES at pH 7.5 produces clusters of plates (*Molecular Dimensions* Morpheus Crystallisation Screen, E8). These diffracted to 2.9, 3.5 and 2.9 Å.

To obtain Clr-DNA co-crystals a myriad of commercial 96 well screens, namely Morpheus I and II, MemStart, Midas Plus and ProPlex screens from *Molecular Dimensions*, JB Penta from *Jena Biosciences*, as well as the $AmSO_4$, Anions, Classics, Classics Lite, Cryos, JSCG Plus, Core I – IV, MBClass I and II and PACT Suite from *Nextal* and the Additive Screen from *Hampton Research* were used. In total 116 initial screening plates were produced. Protein crystals appeared in quite a few of them, however only *Molecular Dimensions* Morpheus Screen and the *Nextal* PACT Suite yielded crystals that diffracted below 4 Ångstrom. Two initial hits were obtained. In a solution containing 0.03 M diethylene glycol, triethylene glycol, tetraethylene glycol and pentaethylene glycol each, 0.1 M MOPS/HEPES buffer mix at a pH of 7.5 and 12.5% (w/v) PEG 1000, 12.5% (w/v) PEG 3350, as well as 12.5% (v/v) MPD (*Molecular Dimensions* Morpheus E8) at 18°C, clusters of plates appeared within two days. In addition to that, in a solution containing 0.02 M 1,6-hexanediol, 1-butanol, (*RS*)-1,2-propanediol, 2-propanol, 1,4-butanediol and 1,3-propanediol each, 0.1 M MOPS/ HEPES buffer mix at a pH of 7.5 and 12.5% (w/v) PEG 1000, 12.5% (w/v) PEG 3350, as well as 12.5% (v/v) MPD (*Molecular Dimensions* Morpheus D8) at 4°C rhombic microcrystals appeared within 6 days. The crystals were both highly anisotropic, the former diffracting to 2.9, 3.5 and 2.9 Å in *a*, *b* and *c* direction, the latter only reaching 3.5, 3.8 and 2.9 Å. No meaningful molecular replacement solution could be obtained.

While the crystal morphologies deviate slightly from one condition to the next, the crystal form is more or less the same. All crystals appear in the same space group $P2_12_12$ and have very close unit cell parameters. This is illustrated using crystals recorded at ID23-1 at the ESRF in Table 4.

Table 4: Data collection statistics of cAMP-bound Clr-DNA cocrystals. Although crystals one through five originate from a completely different crystallisation condition (*Nextal* PACT Suite C4, 18°C) than six and seven (*Molecular Dimensions* Morpheus G8, 4°C), they clearly have the same crystal form. All crystals have a rather low resolution, often paired with severe anisotropy.

No	**space group**	**unit cell parameters [a, b, c, α, β, γ]**	**resolution [a, b, c]**	**reflections**	**redundancy**	**completeness**	R_{merge}	I/σ
1	$P2_12_12_1$	50.3, 70.5, 201.0, 90, 90, 90	3.3, 3.3, 2.9	48829	3.0	77.7%	9.4%	7.7
2	$P2_12_12_1$	51.0, 73.3, 202.3, 90, 90, 90	3.6, 4.4, 3.7	33990	3.7	78.3%	5.6%	10.9
3	$P2_12_12_1$	50.0, 70.4, 200.6, 90, 90, 90	3.5, 3.7, 3.1	43723	3.3	73.8%	19.6%	5.4
4	$P2_12_12_1$	51.0, 73.5, 202.1, 90, 90, 90	3.7, 4.2, 3.3	38955	3.2	70.3%	6.0%	11.1
5	$P2_12_12_1$	51.7, 75.3, 201.5, 90, 90, 90	3.7, 3.8, 3.3	42365	3.4	77.6%	6.2%	11.1
6	$P2_12_12_1$	50.5, 71.9, 201.0, 90, 90, 90	3.7, 5.5, 3.7	27435	3.3	67.6%	7.3%	12.6
7	$P2_12_12_1$	50.1, 69.8, 199.9, 90, 90, 90	3.8, 4.1, 3.8	35347	4.7	91.8%	10.2%	9.4

Nextal PACT Suite C4: 0.1 M PCB buffer (sodium propionate, sodium cacodylate, and BIS-TRIS propane 2:1:2) pH 7, 25% (w/v) PEG 1500.

Molecular Dimensions Morpheus G8: 0.2 M sodium formate, 0.2 M ammonium acetate, 0.2 M sodium citrate, 0.2 M potassium sodium tartrate, 0.2 M sodium oxamate, 0.1 M MOPS/ HEPES pH 7.5, 12.5% (w/v) PEG 1000, 12.5% (w/v) PEG 3350, 12.5% (v/v) MPD.

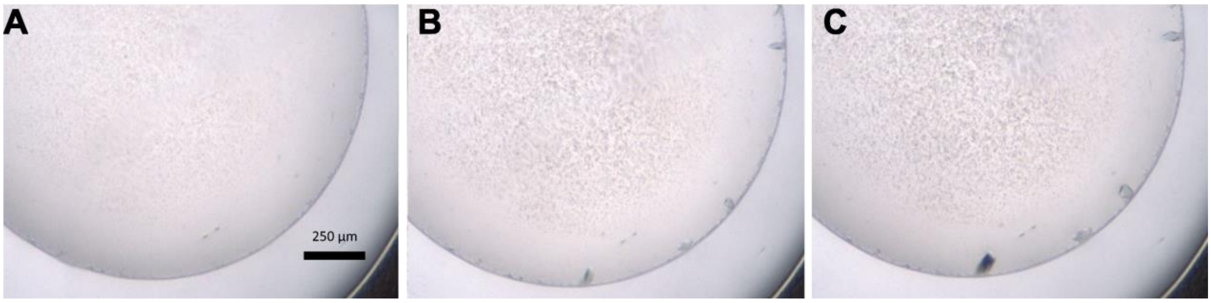


Figure 26: Cyclic AMP bound Clr-DNA cocrystals: The final crystals used for structure determination were grown in bicine/Tris pH 8.5 and a mix of halides in the presence of an MPD/PEG precipitant mix. They were purified using a HEPES based buffer and diffracted to 3.2, 3.1 and 2.7 Å. Crystals appeared within 17 hours and reached their final size two days thereafter.

The crystal quality could be marginally improved by switching from the Tris-based SEC buffer used in the initial screen to a HEPES-based one (see 4.2.1). In addition to that the cAMP and magnesium chloride concentration were increased to 25 mM. Additive screens, pH and precipitant concentration screening and even microseeding did not enhance crystal quality. As it is highly flexible and the resulting multiplicity in conformation leads to the anisotropy issue, the Clr transcription activation complex proved to be particularly difficult to crystallize. Diffraction was neither consistent from one reproduction of the crystallisation condition to another, nor within multiple crystals from the same well and a solvable crystal was obtained empirically over about 300 crystals.

The best crystals were finally grown at 18°C over two days in 0.03 M sodium fluoride, sodium bromide and sodium iodide each, 0.1 M bicine/Tris buffer mix at a pH of 8.5 and 12.5% (w/v) PEG 1000, 12.5% (w/v) PEG 3350, as well as 12.5% (v/v) MPD (*Molecular Dimensions* Morpheus B12) and diffracted to 3.2, 3.1 and 2.7 Å (Figure 26). Images recorded at P13 beamline at the DESY synchrotron in Hamburg were indexed and processed using XDS. The raw data was scaled anisotropically using the UCLA Diffraction Anisotropy Server and a free *R* set based on 5% of the data was generated. The dataset had a rather low completeness of 74.9%, an R_{merge} of 14.5% and an overall I/σ of 11.35. The crystals had the orthorhombic space group $P2_12_12_1$ and a Matthews coefficient of 2.84 $Å^3$/Da indicated the presence of one Clr-DNA complex dimer per unit cell at a solvent content of 56.7%. The cell content was estimated using CCP4i2.

To perform the molecular replacement, a hybrid model consisting the protein entity of GlxR from *Corynebacterium glutamicum* (PDB-ID: 4cyd) and the DNA moiety of the cAMP receptor protein from *Mycobacterium tuberculosis* (PDB-ID: 3mzh) was used as input for creating a search model on the Swiss-Model Server. GlxR has a sequence identity of 28.9% to Clr. After molecular replacement using Phaser in CCP4i2 and the hybrid model, the initial solution was further refined using simulated annealing of the torsion angles and secondary structure restraints. For the rigid body refinement, the model was divided into a few structural elements. These were the DNA fragments corresponding to one Clr molecule each, the helical *N*-terminal section, the effector binding domain with the bound cAMP, the hinge and finally the DNA binding domain. The Clr chains of the obtained coordinates were input into Modeller generating five variations each and combined to a total of 25 new models. These were tested in Phenix refine using the same refinement strategy with simulated annealing and secondary structure restraints. Model Clr 5.1 was identified as best with an R_{free} of 45.0% and used in the further refinement. After an autobuild step with Resolve, the solution was morphed into the data and further refined alternated with manual model building in Coot, until reaching a R_{free} of 31.5% and an R_{work} of 25.6%. The relevant statistics are listed in Table 5.

Table 5: Data collection and refinement statistics for Clr · cAMP in complex with DNA.

Data collection		**Refinement**	
X-ray source	DESY P13	Resolution range (Å)	57.34 - 2.72
Space group	$P2_12_12_1$	$R_{\text{work}}/R_{\text{free}}$ (%)	0.256/ 0.315
Unit-cell parameters (Å, °)	a=50.1, b=69.9, c=200.4, α=90, β=90, γ=90	Average B factor (Å^2)	63.3
Wavelength (Å)	0.976	No. of atoms	
Resolution range (Å)	57.34 - 2.72 (2.82 - 2.72)	Total	4588
Completeness (%)	74.89 (6.72)	Protein	3458
Observed reflections	137677 (927)	Ligand	1125
Unique reflections	14733 (129)	Solvent	5
Multiplicity	9.3 (7.2)	R.m.s.d., bond lengths (Å)	0.006
Wilson B factor (Å^2)	64.80	R.m.s.d., bond angles (Å)	1.14
R_{merge} (%)	0.145 (1.223)	Ramachandran plot	
R_{measure} (%)	0.154 (1.315)	Favoured (%)	88.81
Mean $I/\sigma(I)$	11.35 (1.96)	Allowed (%)	8.50
$CC_{1/2}$ (%)	0.994 (0.788)	Outliers (%)	2.68

4.1.4 Crystallisation and structure determination cGMP-activated *Sm*Clr bound to DNA

The cAMP receptor protein CAP from *E. coli* binds to cyclic guanosine monophosphate, but does not respond with the necessary conformational conversion to enable transcription activation. This is not true for Clr however. To explain the difference in behaviour, Clr was crystallized in complex with cGMP as well. To that end, freshly purified Clr was concentrated to a concentration of 5.95 mg/mL (219 µM) and the dsDNA oligonucleotide was added to a concentration of 274 µM. In addition to that the crystallisation batch contained 25 mM cGMP and magnesium chloride. Crystals appeared within 6 days in a solution containing 0.02 M 1,6-

hexanediol, 1-butanol, (*RS*)-1,2-propanediol, 2-propanol, 1,4-butanediol and 1,3-propanediol each, 0.1 M MOPS/HEPES buffer mix at a pH of 7.5 and 12.5% (w/v) PEG 1000, 12.5% (w/v) PEG 3350, as well as 12.5% (v/v) MPD at 8°C (*Molecular Dimensions* Morpheus D8). The crystals were strongly anisotropic again, only diffracting to 3.3, 4.2 and 3.1 Å in *a*, *b* and *c* direction (Figure 27).

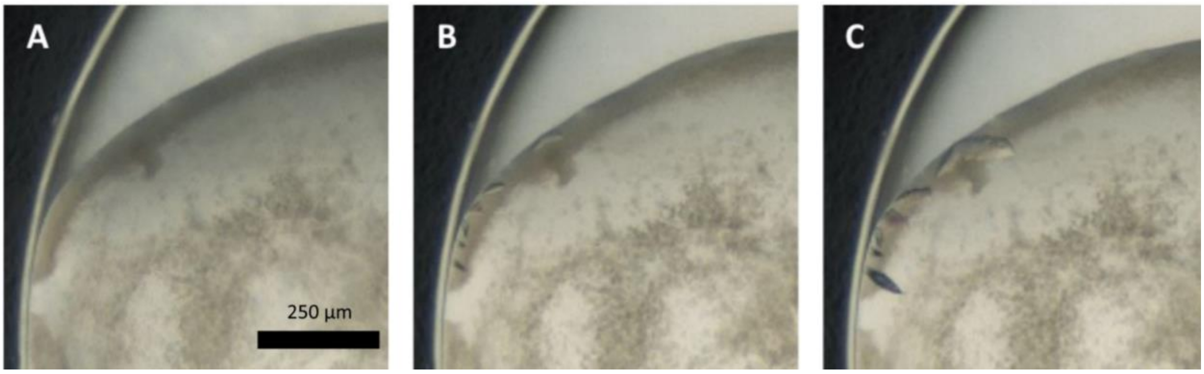


Figure 27: Crystals used for structure determination of the cGMP bound Clr-DNA complex: Crystals appeared within 12 hours and reached their maximum size within 6 days. The solution contained a MOPS/HEPES buffer mix at pH 7.5, various alcohols and diols as well as the MPD/PEG precipitant mix mentioned before.

Images collected at the PXI beamline of the Swiss Light Source (SLS) were processed using XDS. The first 120° proved to be best, yielding a dataset with only 70.2% completeness, an R_{merge} of 5.6% and a mean I/σ of 11.1. Like its cAMP bound counterpart, the crystals grew in the orthorhombic space group $P2_12_12_1$ and had a Matthews coefficient of 2.93 Å^3/Da, resulting in the presence of one Clr-DNA complex dimer per unit cell at a solvent content of 58.0%. The previous structure was used for phasing and the resulting coordinates were analogously refined to a final R_{free} of 35.1% and a R_{work} of 28.9%. The relevant statistics for the data collection and refinement can be found in Table 6.

Table 6: Data collection and refinement statistics for Clr · cGMP in complex with DNA.

Data collection		**Refinement**	
X-ray source	SLS PXI	Resolution range (Å)	49.15 - 3.12
Space group	$P2_12_12_1$	R_{work}/R_{free} (%)	0.291/ 0.353
Unit-cell parameters (Å, °)	*a*=50.7, *b*=72.9, *c*=200.0, α=90, β=90, γ=90	Average B factor (Å^2)	96.8
Wavelength (Å)	1.000	No. of atoms	
Resolution range (Å)	49.15 - 3.12 (3.23 - 3.12)	Total	4694
Completeness (%)	70.19 (5.53)	Protein	3466
Observed reflections	41837 (146)	Ligand	1225
Unique reflections	9719 (74)	Solvent	3
Multiplicity	4.3 (2.0)	R.m.s.d., bond lengths (Å)	0.003

Wilson *B* factor (Å^2)	100.50	R.m.s.d., bond angles (Å)	0.63
R_{merge} (%)	0.056 (0.481)	Ramachandran plot	
R_{measure} (%)	0.064 (0.621)	Favoured (%)	92.87
Mean $I/\sigma(I)$	11.08 (1.87)	Allowed (%)	4.90
$CC_{1/2}$ (%)	1 (0.61)	Outliers (%)	2.23

4.1.5 General features of the nucleotide-induced Clr-DNA complex

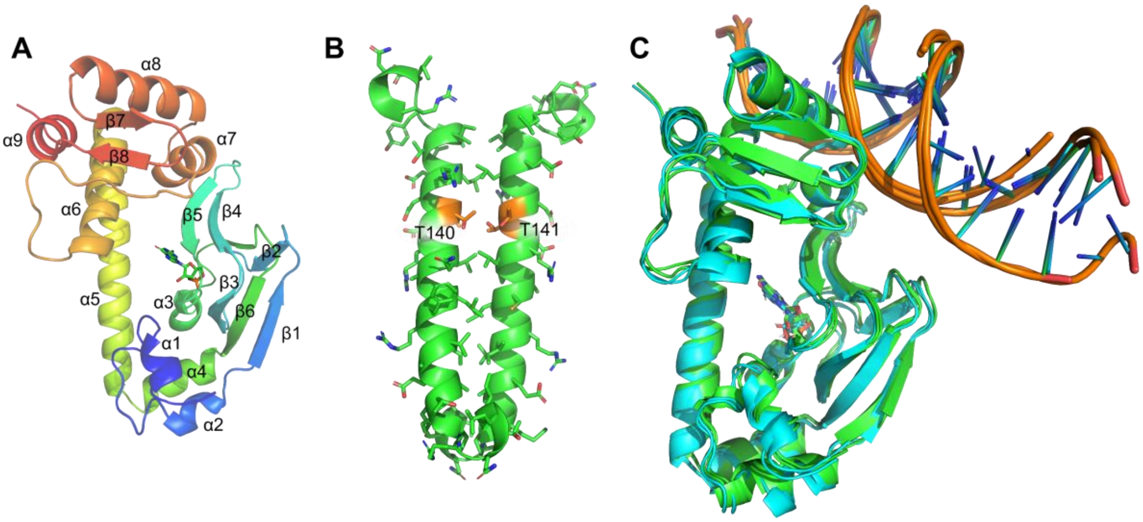


Figure 28: Structural overview of the Clr-DNA transcription activation complex: (A) Clr consists of two domains connected by a long hinge helix α5, which contains the dimerization interface. *N*-terminally located is the cNMP binding domain, whereas the DNA-binding helix-turn-helix motif follows *C*-terminally. Helix α8 is of particular importance for DNA interaction, as it inserts into the major groove and facilitates base readout. (B) Helices α5 form a zipper-like assembly upon dimer formation. The interaction is mainly driven by hydrophobic interactions, T140 and T141 being notable exceptions. Interaction of these amino acids with the nucleotide in the cNMP binding domain results in stabilisation of the upper helix. (C) Superimposition of the cAMP (green) and cGMP (cyan) bound structure in both chains. With the exception of the crystal packing-based asymmetry in DNA resolution, the conformation is perfectly aligned.

The crystal structure of DNA-bound Clr in its nucleotide-activated state shows the overall features of a cAMP receptor protein in the closed conformation. The individual subunits fold into two domains linked by helix α5 (residues R122-A148), which is also the dimerization interface. Residues 1-121 form the cNMP binding domain, the larger of the two. It is formed by four α helices and seven β strands in an antiparallel orientation of a β-barrel encasing the cNMP binding site. A helix-turn-helix (HTH) DNA binding domain consisting of four α helices and two β strands is located at the C-terminus from residue 149 to 234. The DNA interacts with the adjacent symmetry mate of the crystal lattice and is therefore quite well resolved (Figure

29). The subunits themselves, as well as the two nucleotide bound structures, show a high conformational congruity. The RMSD of the two entire subunits is 0.492 Å (0.563 Å for cGMP), the cNMP binding domain aligns with 0.498 Å (0.529 Å) and the HTH domain even with a RMSD of 0.332 Å (0.515 Å). As the crystal structures containing cAMP and cGMP can be aligned with an RMSD of 0.435 Å including the DNA, both nucleotides seem to be able to shift the protein fully towards the active conformation. The superimposition of chain A and B of both structures is shown in Figure 28 (C).

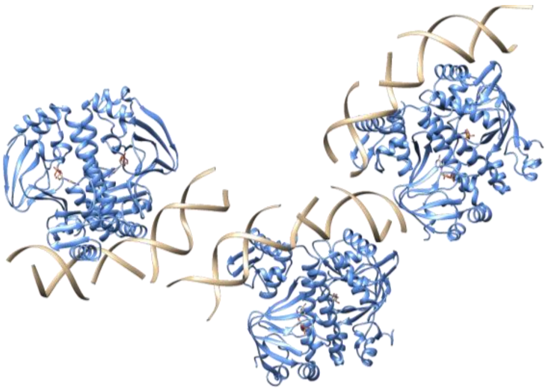


Figure 29: Crystal packing in Clr-DNA co-crystals: One side of the synthetic duplex DNA forms packing interactions with the symmetry mate *in crystallo*, resulting in a better resolved density in that particular direction. The side lacking these interactions has no sufficient electron density for the five outermost bases, whereas the other is fully resolved.

The cNMP binding domain is universally conserved throughout cNMP receptor proteins, as well as protein kinases and ion channels, with sequence identities above 20% even for mammal representatives (Shabb and Corbin 1992). The HTH motif of the DNA binding domain is equally prevalent with 116 so far deposited in the PDB according to InterPro classifications (Rohs et al. 2010; Finn et al. 2017). The overall fold of Clr is highly similar to its *E. coli* homologue, the catabolite activator protein CAP (PDB-ID: 1CGP; Schultz et al. 1991).

The transition to the active state is mainly dependent on the bZIP-like dimer interface formed by α5. The dimer interacts via hydrophobic interactions formed by P123, A126, V129 with I130, L133, L137, L144 and A147. T140 and T141, which are involved in nucleotide binding and initiate the switching to the active state, also interact hydrophobically with their methyl groups, while their hydroxy group is turned towards the active site (Figure 28 B).

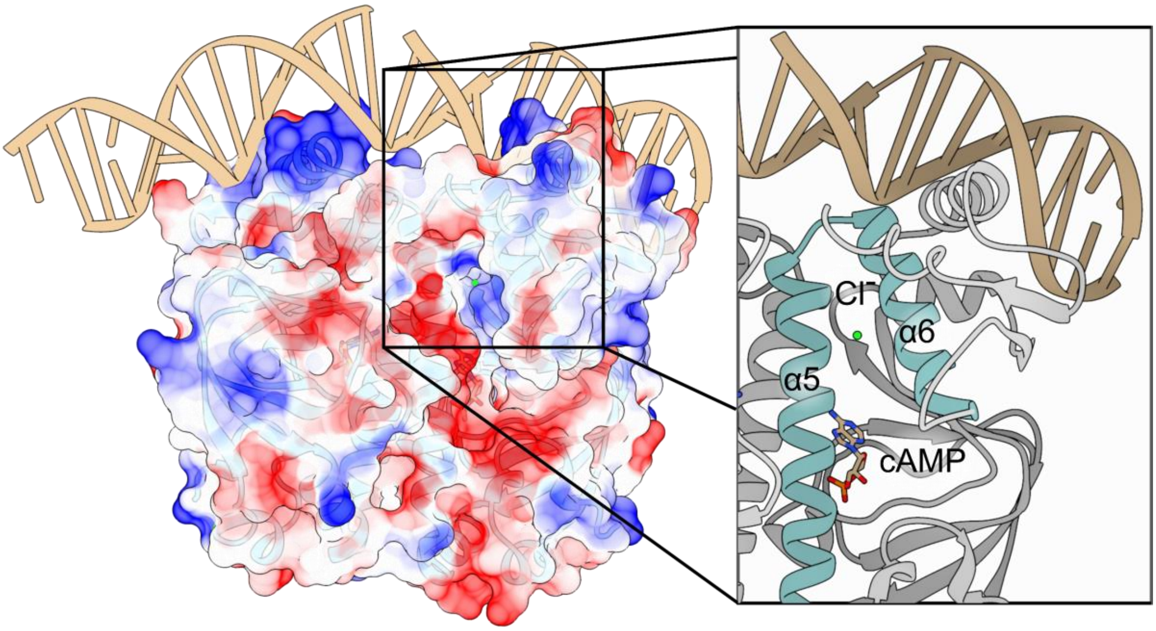


Figure 30: Clr contains a positively charged cavity between helix 5 and 6 occupied by Cl$^-$. The active conformation of Clr is defined by an ordered *C*-terminal section of the hinge α5 and a resulting alignment of helix 6 in an angle of about 40° (Won et al. 2009). In our structure we found a chloride binding site located in this exact position, which might be involved in the interaction network placing the HTH domain in its active closed position.

The open conformation of Clr is characterized by a high flexibility in the positioning of the DNA binding domain and a partial unwinding of the hinge helix. To adopt its closed conformation, the *C*-terminal part of the hinge needs to be fully ordered and the subsequent helix 6 has to be aligned to it in a 40° angle, bringing helix 8 in a suitable position for interaction with the DNA major groove. This change in conformation is also accompanied by a significant reduction in movement. The crystal structure of Clr reveals a positively charged cavity between the hinge and helix α6, which is occupied by a chloride ion in both monomers (Figure 30). Additionally, it was observed during purification, that Clr is rather unstable and prone to precipitation in solutions of low ionic strength like size exclusion buffer. Thermofluor buffer optimisation also revealed a stabilising effect of salt. These observations might suggest, that the chloride provides a stabilising effect on the hinge-helix 6 interaction and reduces the conformational degrees of freedom in the DNA binding domain.

4.1.6 Both cAMP and cGMP are able to shift Clr to its active conformation

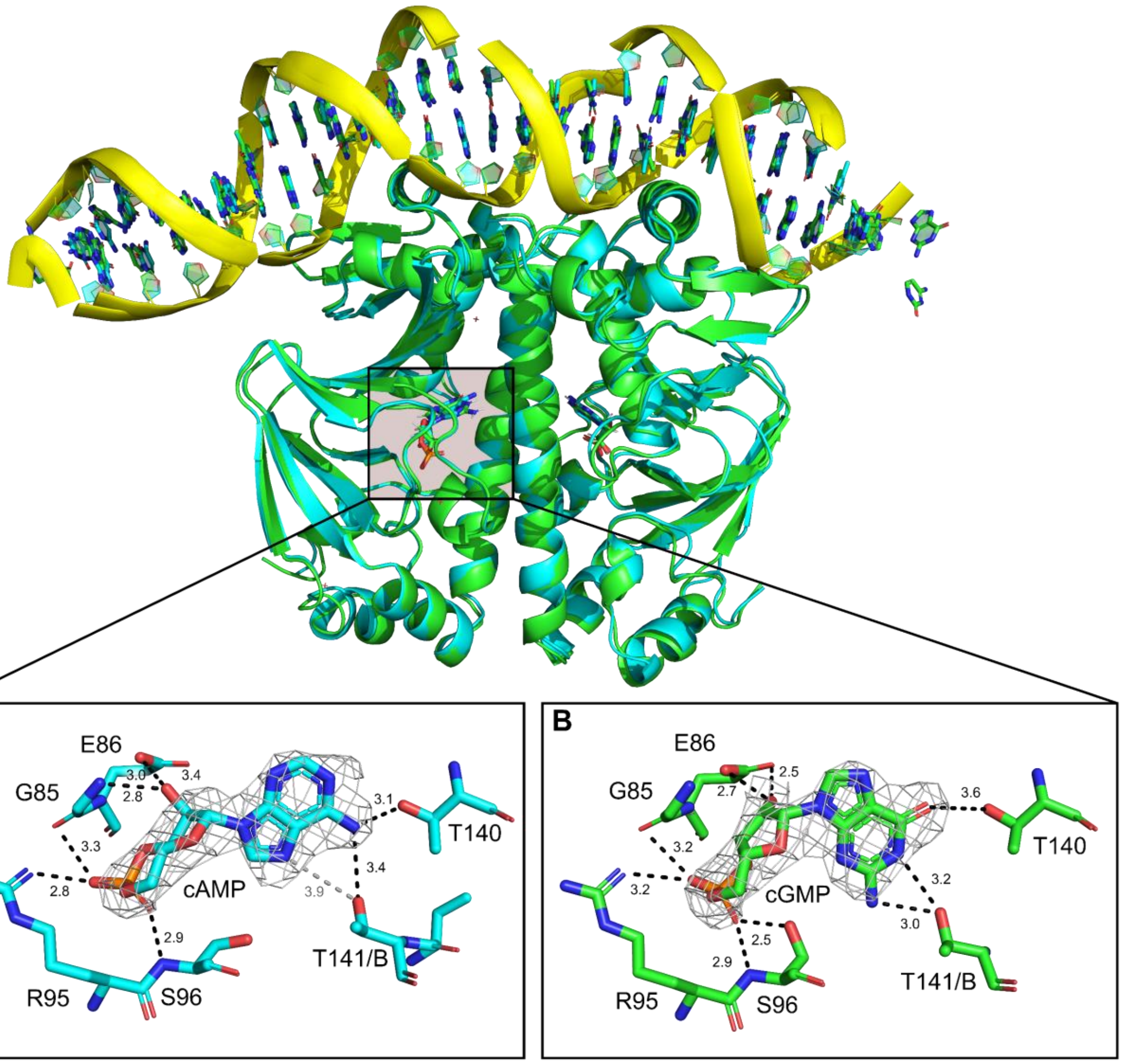


Figure 31: Nucleotide binding environment in Clr: (A) The sugar and phosphate moiety of cAMP interact with G85 and E86 of helix α4, as well as R95 and S96; the primary amine of the adenosine is coordinating T140 from α5 in the same subunit and T141 from the adjacent one; the ligand $2\,mF_{obs}{-}DF_{calc}$ experimental electron density map contoured at 1.3 σ is shown in grey. (B) The cGMP bound crystal structure reveals an identical binding mode with slightly deviating hydrogen bond lengths. The sidechain of S96 is tilted towards the ligand and T141 of the adjacent subunit is in closer proximity to it. The nucleotide is shifted to a *syn*-conformation.

The active site located in the N-terminal cNMP binding domain is occupied by cAMP and cGMP respectively. The active site orientation is illustrated in Figure 31. The binding mode of the phosphate and sugar moiety is identical, involving G85 and E86 within α4 and R95 as well as S96 of the following loop. The sidechain of S96 is flipped towards the phosphate in the cGMP-bound structure only. The nucleobase of cGMP is rotated towards a *syn*-conformation, while it occupies the *anti*-conformation in case of cAMP. This allows for interaction of both guanosine and adenosine bases with the helix α5 residues Thr140 of the same subunit and Thr141 of the adjacent subunit. While cAMP interacts with T140 and T141 with the primary amine, cGMP contacts T140 with its carbonyl group and T141 with the primary amine. The tertiary amine N1' of cGMP is able to establish one more hydrogen bond to T141 of the adjacent molecule than the secondary amine of cAMP, which is a substantial difference from Clr to *E. coli* CAP. Interaction of the nucleotide with both α5 helices has been described to trigger a coil-to-helix transition in the *C*-terminal part of the hinge and shift the HTH domain towards its active conformation for CAP [29]. cGMP had been found to be incapable to establish interactions with the second helix for the subsequent Crp activation [30]. In contrast to this, in case of Clr both the interaction and arrangement of the full α5 helix can be seen in the structure, which is in agreement with the activation of the Clr-DNA binding in vitro and promoter activity in vivo by both cNMPs.

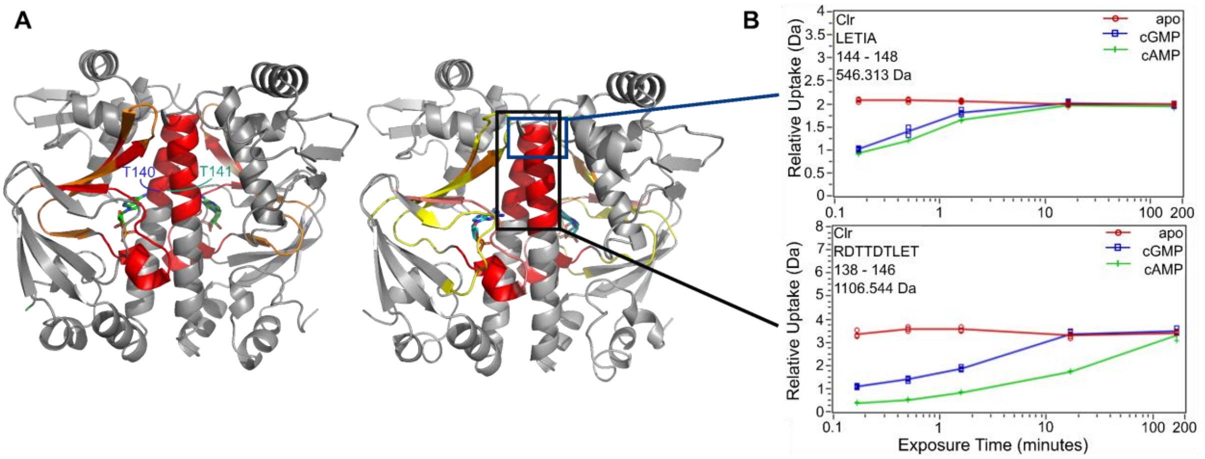


Figure 32: Nucleotide binding results in decreased H/D exchange within helix α5 in HDX-MS: (A) Areas of reduced relative hydrogen-deuterium exchange upon binding of cAMP (left) and cGMP (right). In addition to the expected displacement of solvent in the binding cavity, the *C*-terminal half of helix 5 shows a reduction in exchange, that can be attributed to a coil-to-helix transformation. (B) Time-dependent deuterium exchange graphs for two key peptides within helix 5 show, that the shielding cannot be solely due to interaction with T140 and T141, but is the result of a change in secondary structure. Peptides not including both exemplified by ^{144}LETIA (blue box) also show a decrease in H/D exchange.

Mapping the relative hydrogen-deuterium exchange between apo and holo state in HDX-MS experiments shows that upon nucleotide addition the *C*-terminal half of the helix is shielded from exchange. This further supports the notion, that both nucleotides are able to induce the coil-to-helix transfer in the linker to the following helix-turn-helix domain. The shielding is not only due to nucleotide interactions with T140 and T141, as even fragments containing the five

C-terminal residues of this region show a clear decline in solvent accessibility for both nucleotide states (Figure 32). In addition to that, the active site shows a slightly stronger decline in deuterium uptake for cAMP binding compared to cGMP.

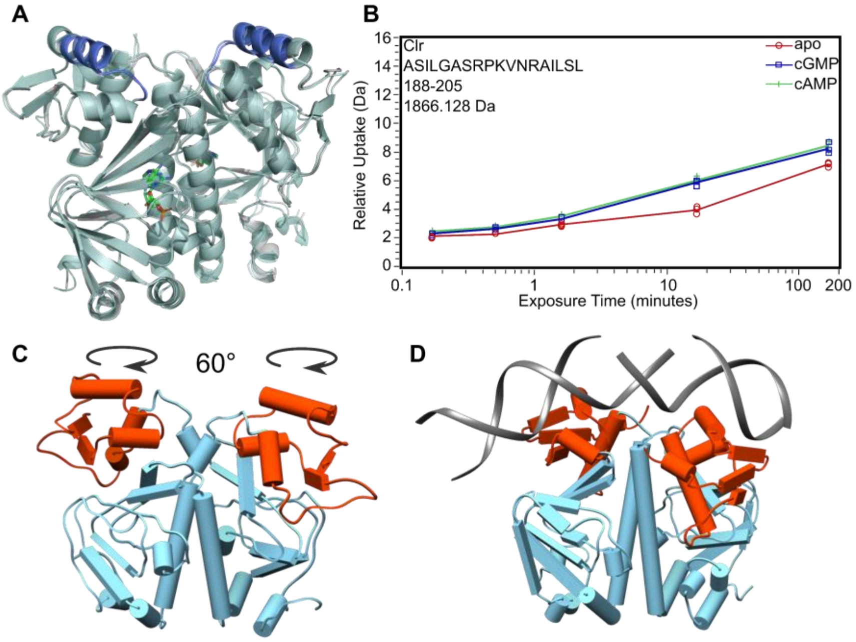


Figure 33: Conformational changes in the HTH domain upon ligand binding can be mapped via HDX-MS: (A) Mapping the relative deuterium uptake difference for apo- vs. nucleotide-bound states onto the crystals structure clearly reveals α8 as the area of interest. (B) The deuterium incorporation increases upon nucleotide binding, indication structural rearrangements involving this particular region. (C) As already described for *E. coli* CAP, α8 undergoes a rotation by about 60° from an open to a closed conformation, which is capable of DNA binding. This can be illustrated by creating a homology model of the open conformation of Clr based on *E. coli* CAP (PDB-ID: 2wc2) using the Swissmodel server in contrast to the closed conformation found in the crystal structure (D).

The C-terminal helix-turn-helix motif of Clr can be found in two conformations, termed open and closed conformation. Starting from the nucleotide-induced rearrangement of the *C*-terminal end of the helix α5 dimer, the HTC domain packs tighter to the dimer centre. The conformational transition is accompanied by a clockwise rotation of helix α8 enabling the helices to enter the DNA major groove. In HDX measurements increased deuterium uptake between α7 and α8 upon nucleotide binding can be observed. This likely stems from helix 8 no longer shielding the interface as a result of its rotation as shown in Figure 33.

Aside from the conformational changes, deuterium exchange is greatly reduced in the cAMP binding domain’s β-barrel upon ligand binding. The relative deuterium uptake difference for apo cNMP-bound Clr over the entirety of the protein is shown in Figure 34.

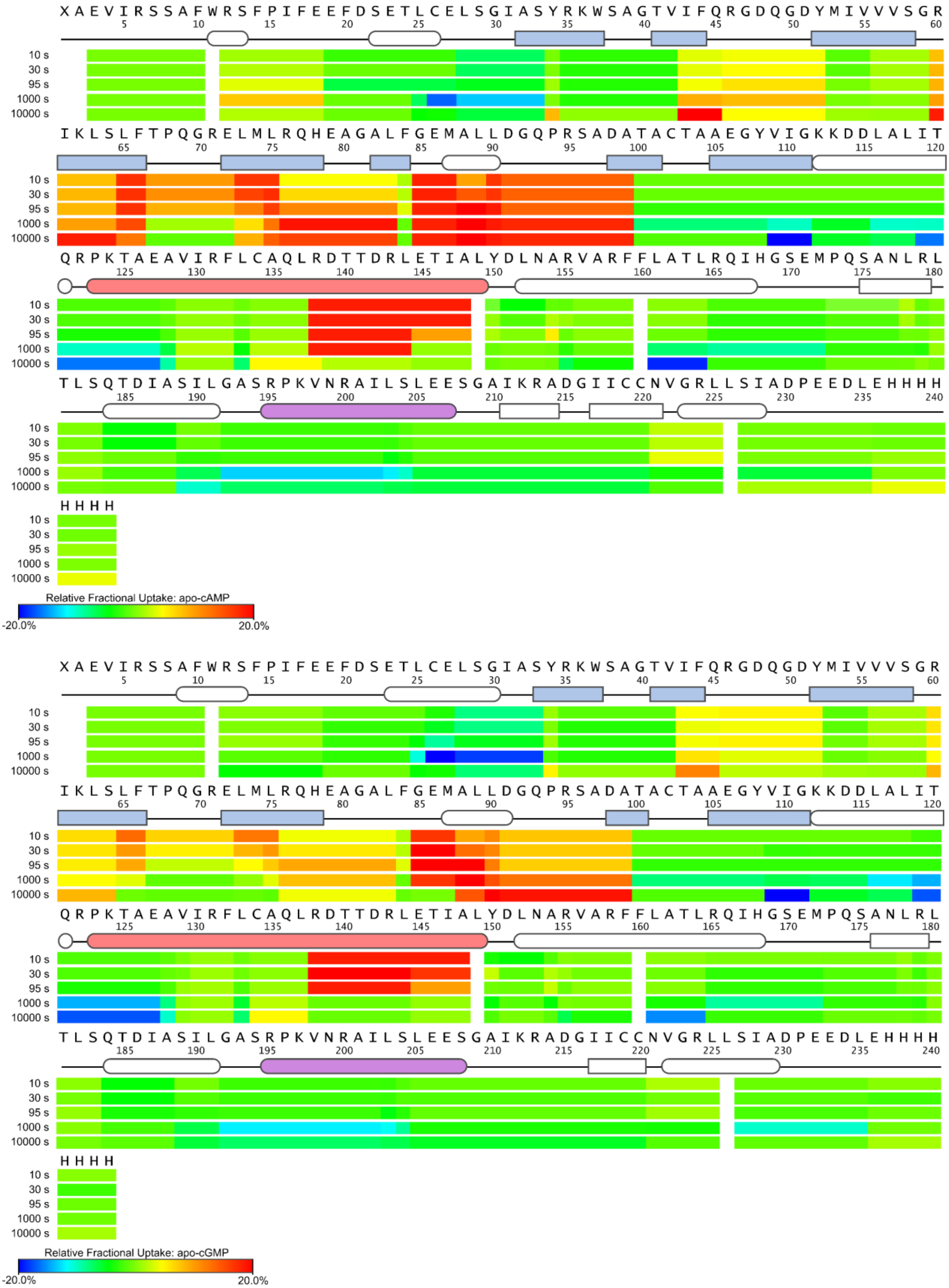


Figure 34: Relative fractional deuterium uptake: Top: The relative uptake of apo - cAMP-bound state was averaged over all peptides containing a particular amino acid, resulting in the shown heat map. Secondary structure motifs are indicated above. β-strands highlighted as grey boxes form the β-barrel of the cNMP binding domain. Upon nucleotide binding the active site is shielded from solvent, resulting

in a reduced uptake in these regions. Helix α5 is indicated in red and the *C*-terminal elongation of it can be clearly seen in reduced uptake upon nucleotide addition. The rearrangement of α8 (violet) makes the area more solvent accessible. The effect is not quite as obvious as the active site shielding, as it takes place on the surface on the protein, that is more solvent accessible anyways. Bottom: The relative uptake heat map for apo – cGMP-bound state shows the same arrangement with only slight deviations. It can be assumed that both show variations of the same state.

4.1.7 Clr promotor specificity can be explained by its nucleotide-bound structures

Clr binds to the target DNA of 33 bp length via electrostatic interactions and hydrogen-bonding to the major groove. The surface of the HTH DNA binding domain has a strongly positively charged patch that interacts with the DNA phosphate backbone. The protein has a likewise positively charged cavity in each subunit that contains the cyclic nucleotide. Helix α8 is inserted in the major groove and interacts with hydrogen bonds over six base pairs. The binding results in a 40° and 32° kinking of the DNA helix at the longer and shorter DNA oligomers respectively, which is eased by the double strand break (Figure 35). The bend is not a result of the double strand break, as it is retained in MD simulations after linking. The entropic cost associated with DNA bending is overcompensated by the establishment of further interactions. The observation of kinking implicates that the flexibility of the flanking regions of the recognition sequence might be distinguished by Clr, adding another layer of specificity (Dai et al. 2004).

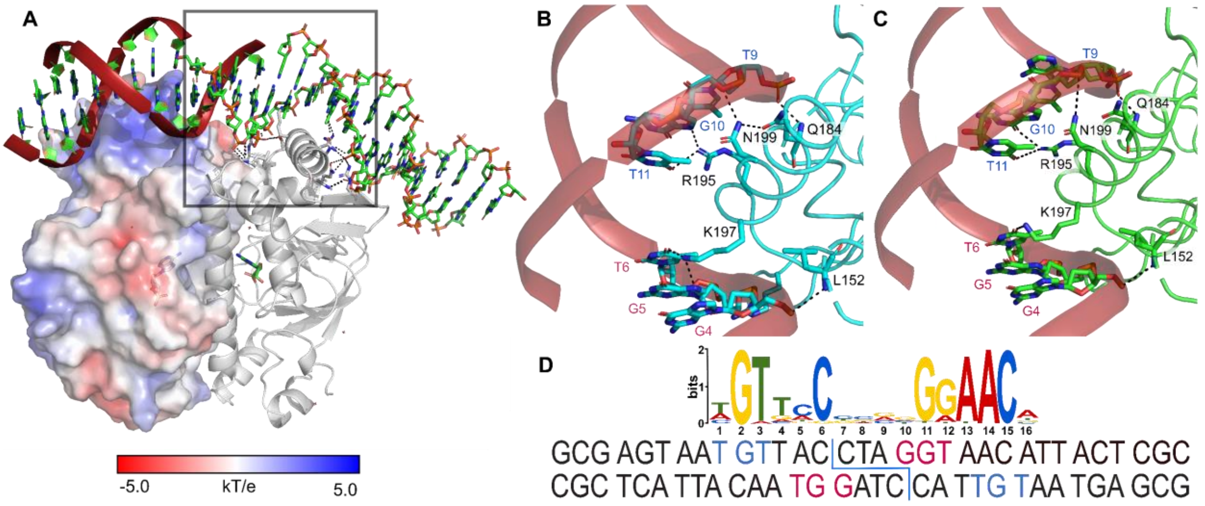


Figure 35: Interaction of Clr with its DNA recognition site. (A) The surface of the helix-turn-helix DNA binding domain is positively charged to interact with the phosphate backbone. Helix α8 is inserted into the major groove and provides base readout via direct hydrogen bonding interactions. The double strand break above the subunit interface facilitates the DNA kinking of about 40°. (B, C) Direct interactions of Clr with the DNA major groove upon cAMP (B) and cGMP (C) binding. The DNA bases

involved in the interaction are highly conserved. (D) The synthetic oligonucleotide used in crystallisation has a double strand break with sticky ends right above the subunit interface. The design was based on the consensus motif displayed above.

The electron density of the double-stranded DNA oligonucleotides (strand CD and EF) could be assigned for 17 (19) bases of strand C, 14 bases of strand D, 13 (14) bases of strand E and 14 bases for strand F for the cAMP and cGMP structure respectively. The base readout happens in the most conserved positions of the WGTTCC-N4-GGAACW binding motif, which explains the need for these residues to be retained. The ^{9}TGT motif of the 14-mer is in close proximity to the *N*-terminal part of α7 and on top of α8, which lays orthogonal to it. The interactions within this motif are established exclusively by the phosphate backbone. The sidechain of S183 in α7 is located 3.8 Å from ^{8}A of the 14-mer in the cAMP-bound structure (3.1 Å in the cGMP state) but in the other Clr monomer it is turned even further away from the phosphate at 4.8 Å in the cAMP-bound structure (4.3 Å for cGMP). As the sidechain of S183 is rotated away from the phosphate in one monomer this bond seems to be rather unstable. Q184 however is involved in two base contacts. Its peptide nitrogen is within 2.7 Å from the ^{9}T phosphate in the 14-mer, while the sidechain reaches towards the ^{10}G phosphate. The amine binds with a O-N distance of 3.1 Å, whereas the oxygens are spaced 3.4 Å apart. Another residue that might turn out to be involved in backbone interaction once a higher resolution structure of Clr is achieved is G192 in the hinge between α7 and α8. Its peptide oxygen is within 3.6 and 3.8 Å of the ^{5}G phosphate of the 19-mer for the cAMP- and cGMP-bound state respectively. Helix α8 has been found to provide base readout in previous Crp homologs and this can be observed for Clr as well. Its *N*-terminal part is held in place by two interactions of S194 with the ^{5}G phosphate of the 19-mer. Both the backbone nitrogen and the side chain oxygen are within 3 Å of it. R195 establishes the first contact with a DNA base. The distance between the sidechain nitrogen and the ^{10}G purine oxygen in the 14-mer is merely 2.7 Å. In the cGMP-bound structure the side chain is turned further towards the base, allowing it to contact the purine 7’N with 2.5 Å distance as well. The K197 sidechain interacts with the complementary 19 bp DNA strand at ^{6}T. The interaction takes place at the 4’O of the thymine base with a hydrogen bond length of 3.1 Å. Finally, N199 provides another two backbone contacts with the DNA. The sidechain nitrogen is within 2.6 and 3.1 Å of the phosphate moiety of ^{10}G in the 14 bp oligomer. Aside from helix 7 and 8, two additional interactions between Clr and the DNA can be seen in the crystal structure. The backbone nitrogen of L152 located in the loop C-terminally of the hinge α5 is within 3.1 Å of ^{4}G of the 19-mer and for the cAMP-bound structure R71 of the cNMP binding domain might be contacting ^{6}G of the 19-mer in a distance of 3.4 Å. In the cGMP-bound state

R71 is slightly turned away in one monomer, while it interacts with L149 in the other, which brings it in an averted orientation from ^{6}G. The binding network is illustrated in Figure 36.

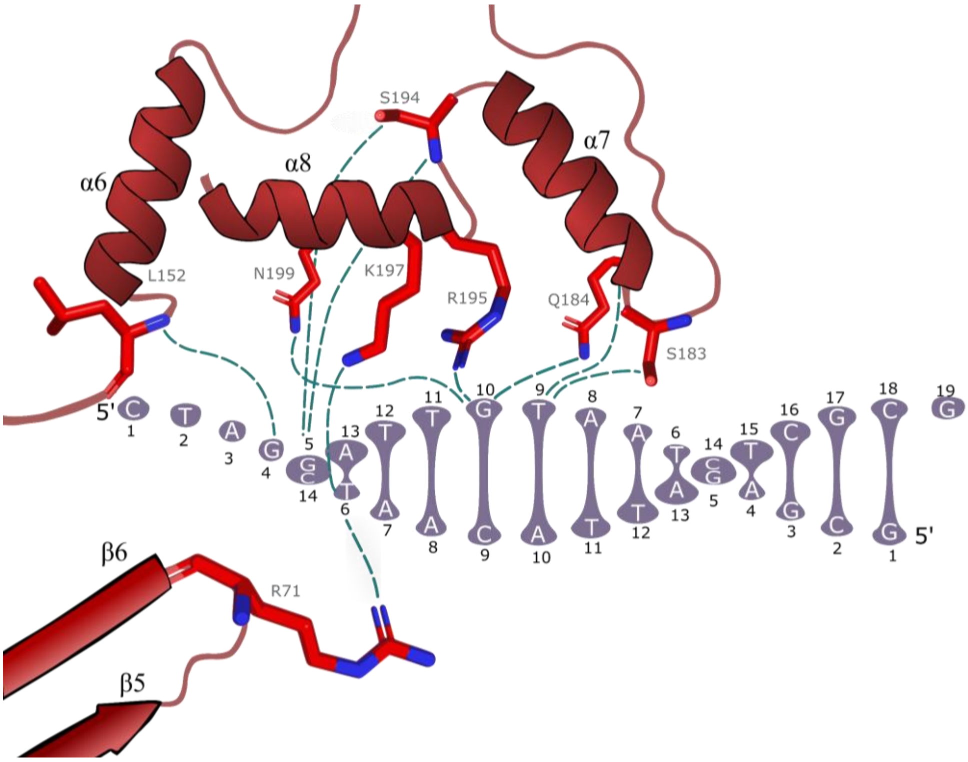

α6
α8
α7
S194
L152
N199
K197
R195
Q184
S183
5'
C 1
T 2
A 3
G 4
G 5
T 6
A 7
A 8
T 9
G 10
T 11
T 12
A 13
C 14
T 15
C 16
G 17
C 18
G 19
G 1
C 2
G 3
A 4
G 5
A 13
T 12
T 11
A 10
C 9
A 8
A 7
T 6
C 14
A 13
5'
β6
R71
β5

Figure 36: Schematic overview of Clr contacts with the corresponding DNA motif. Two areas of the Clr binding motif are particularly conserved. These correspond to ^{4}GGT on the 14-mer and ^{9}TGT on the 19-mer of the duplex DNA used in crystallisation. All of Clr's DNA contacts are with the aforementioned bases. Most of the hydrogen bonds established by Clr upon DNA binding involve oxygen from the phosphate backbone, exceptions are the interaction of R198 with ^{10}G and K197 with ^{6}T.

4.2 *Sm*CpdA: the misunderstood phosphodiesterase

*Sm*c02179 in the CyaD1 locus encodes for a bacterial class III phosphodiesterase and has been initially characterized by Mathieu-Demazière et al. 2013. They described the protein as metal-independent monomeric 2’,3’-cyclic nucleotide phosphodiesterase. *In vivo* inactivation of CpdA led to decreased *Sm*c02178 expression in their studies. As part of the CyaD1 locus and since only one other class III phosphodiesterase Rv0805 (PDB-ID: 3IB8) from *M. tuberculosis* has been structurally characterized so far, CpdA warranted a closer look. Surprisingly, our experiments resulted in quite a few preconceived notions about CpdA needing to be revised.

4.2.1 Expression and purification of *Sm*CpdA

*Sm*CpdA was heterologously produced from a codon optimized construct in *E. coli* using LB kanamycin (35 µg/mL) and expression was induced by the addition of 0.1 mM IPTG. The cells were harvested after 4h at 30°C via centrifugation (4000 rpm, 15 min, F6S 6x1000Y). The pellet was resuspended in binding buffer and flash frozen.

Cells corresponding to 0.8 L culture were passed once in a semi-frozen state without back pressure, and thrice at 15.2 bar cm^{-2} though a French press pressure cell. The solution was centrifuged at 18000 rpm (JA-20) for 1 h and the supernatant was decanted. Initial nickel affinity purification resulted in sample already containing very few impurities (Figure 37).

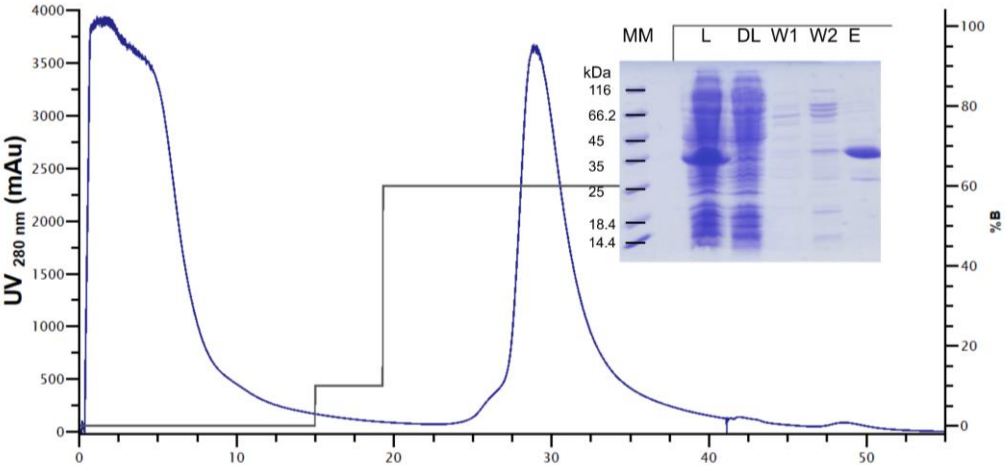


Figure 37: IMAC purification of CpdA: Raw lysate was applied to a 5 mL nickel NTA column using a peristaltic pump and eluted in a stepwise imidazole gradient on a FPLC system. The flow rate of both steps was 2 mL/min. Both the chromatogram and the SDS-PAGE at 12% show the excellent purification of CpdA at around 35 kDa (slightly above its actual weight). In the lysate sample (L) a strong overproduction of the protein of interest is visible. In the column flow through (FT) and wash fractions (W1-2) barely any was displaced from the column. The elution (E) contains CpdA in good purity.

After affinity purification, size exclusion chromatography was used as final polishing step shown in Figure 38. The protein elutes at a volume consistent with a monomeric species. X-ray fluorescence clearly showed that CpdA heterologously expressed in *E. coli* always contains iron as cofactor metal. The measurement is further explained in Figure 46. To obtain manganese bound protein, the nickel-NTA eluate was diluted 1:1 with a binding buffer containing 2 mM EDTA, incubated at 8°C for 2 h and the sample was diluted against size exclusion buffer containing 1 mM manganese chloride overnight prior to SEC. Manganese-bound *Sm*CpdA is extremely stable, lasting months at 4°C without noticeable precipitation. The protein can be frozen without additional cryoprotectant without loss in activity.

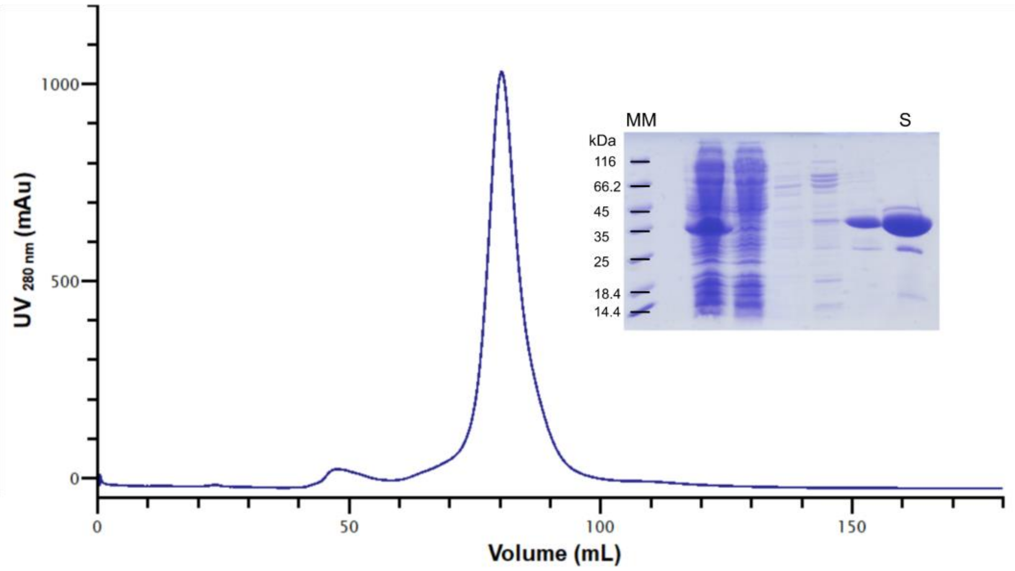


Figure 38: Size exclusion chromatography of *Sm*CpdA: The elution fractions from IMAC were concentrated and applied to a Superdex 200 16/600 column at a flow rate of 1 mL/min. The already great purity of the sample is reflected in a very clean elution profile and was verified in SDS-PAGE using 12% gels (S).

4.2.2 *Sm*CpdA is a manganese-dependent cyclic nucleotide phosphodiesterase

To evaluated substrate- and cofactor specificity of CpdA, the educt to product ratio needs to be quantified. The cheapest and most simple way in terms of setup are enzymatic assays, like *Promega*'s PDE-Glo for cAMP detection. The high specificity of the enzymes used in these assays can allow them to be used even in complex mixtures like crude cell lysate, but this also limits them in terms of potential substrates they can screen for. The same advantages and limitation are also present in ELISA Immunoassays. Traditional analytical methods like mass spectrometry and NMR on the other hand are less accessible but also a lot more versatile. Which of these methods is the most feasible for our particular problem, had to be explored.

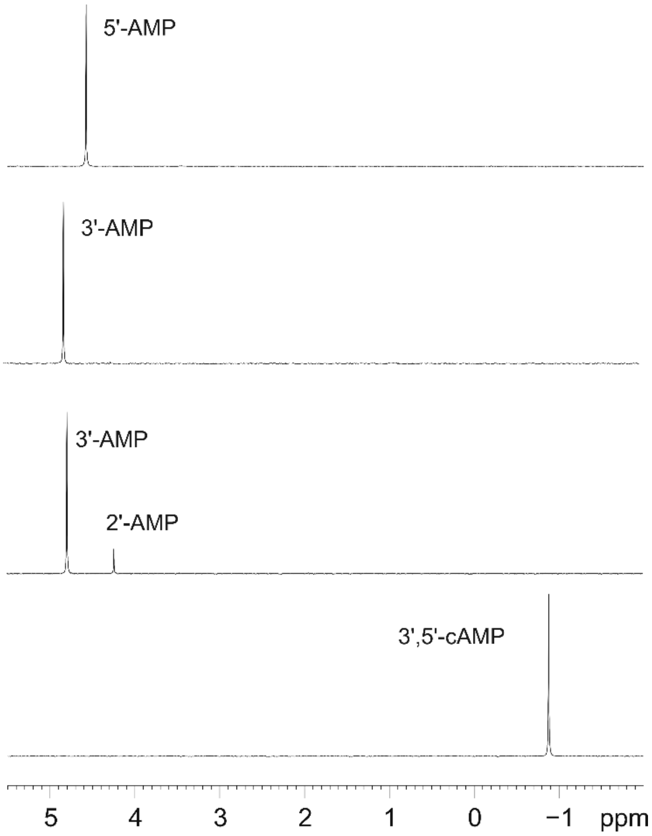


Figure 39: Stack plot of the ^{31}P spectra of all compounds in D_2O at room temperature. The chemical shift of the cyclic phosphate educt can be clearly distinguished from the potential products. As the chemical shift is highly dependent on buffer environment and pH an identification of the linear monophosphates solely based on ^{31}P is not possible and additional spectra need to be recorded.

All potential substrates of *Sm*CpdA contain phosphorus, an element naturally present in a single isotope ^{31}P. Due to its nuclear spin this isotope can be detected in NMR and the first experiment aimed to find out whether 3’5’-cAMP and the potential products can be distinguished in ^{31}P-NMR alone. The Nucleotide were dissolved in D_2O to a final concentration of 5.8 mM and NaOH was added to neutralize in case of the 3’-AMP free acid. Additionally, NMR spectra of ^{1}H, ^{1}H with suppression of the residual H_2O signal in the D_2O solution, ^{1}H-^{1}H DQF-COSY, ^{1}H-^{13}C HSQC, ^{1}H with ^{31}P decoupling, ^{31}P, and ^{1}H-^{31}P HMBC for 3’5’-cAMP, 3’-, 5’-AMP and a 2’-/3’-AMP mixture were recorded on a Bruker AVIII 500 MHz installed with 5 mm Cryo probe Prodigy BBO (broadband observation) with z-gradient and Bruker AVIII HD 500 MHz installed with 5 mm probe TBI (triple broadband inverse resonance) with z-gradient. The spectra were assigned in collaboration with Dr. Xiulan Xie and can be found in the Appendix. While the cyclic and linear monophosphates can be clearly distinguished in ^{31}P, the chemical shifts of the monophosphate isomers are too close to allow assignment only based on these spectra. This mean additional ^{1}H-based spectra need to be collected in the absence of water and any sample would have to undergo buffer exchange to deuterated buffer. This is feasible, but not overly cost- and time-effective and a HPLC-based assay was explored as an alternative.

For the HPLC assay of CpdA activity and specificity the protein with or without additional metal (II) chlorides was incubated with various substrates for 30 min at 25°C and subsequently precipitated. The sample was diluted 1:5 to decrease ionic strength and analysed using an ion exchange column as described in 3.2.14.

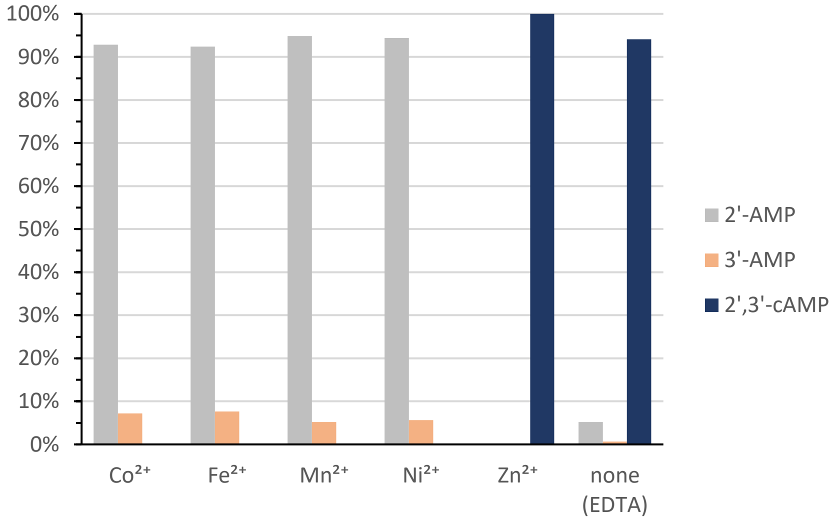


Figure 40: Cofactor-dependent 2',3'-cAMP hydrolysis of *Sm*CpdA: *Sm*CpdA expressed in *E. coli* consistently carries iron cofactors. To find the optimal metal for protein activity, said cofactor was removed and supplemented by the addition of metal chloride. The hydrolysis of 2',3'-cyclic AMP leads to 2'-AMP as the main product and the protein is quite active for any metal (Zn^{2+} being an exception). The highest activity and selectivity could be observed with manganese. The addition of EDTA greatly inhibits activity, but is unable to completely abolish it. This is likely due to incomplete metal removal.

CpdA converts 2',3'-cyclic AMP almost entirely to 2'-AMP with only about 6% of product yield corresponding to 3'-AMP (Figure 40). Manganese-bound protein had the highest selectivity with 95% 2'-AMP yield. This is remarkable, as its homolog *M. tuberculosis* Rv0805 favours 3'-AMP and hydrolysis to the 2'-product is rare (Keppetipola and Shuman 2008). The addition of zinc led to precipitation of the protein, whereas for any other metals, full turnover after 30 min was observed. Treatment with 1 mM EDTA largely decreased the activity, but some residual activity likely due to incomplete metal removal remained. For 3',5'-cAMP hydrolysis shown in Figure 41 the main product is 3'-AMP. Manganese bound CpdA had the highest selectivity with 92% main product again, whereas magnesium led to the lowest 3’-AMP ratio of 84%.

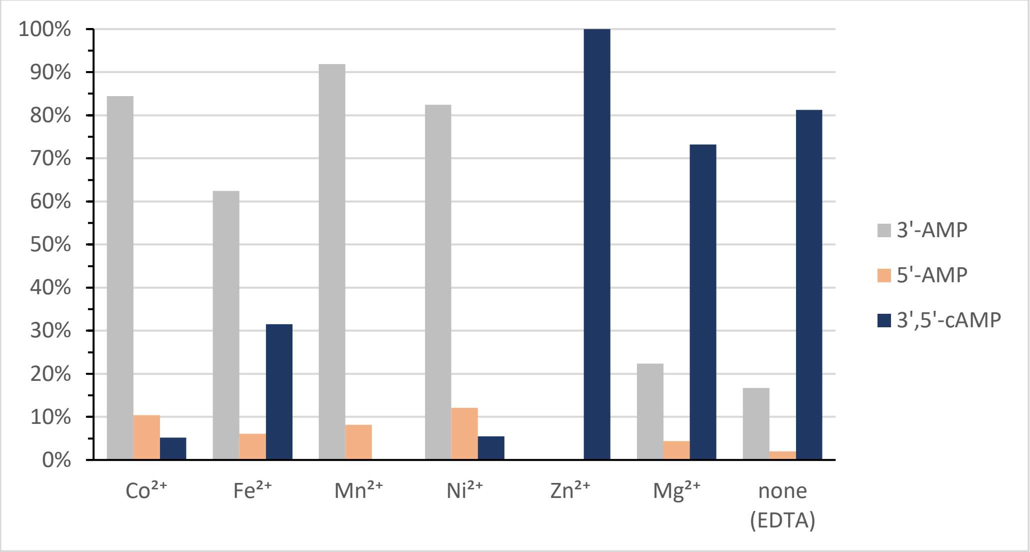


Figure 41: Cofactor-dependent 3',5'-cAMP hydrolysis of *Sm*CpdA: The activity of *Sm*CpdA towards 3’,5’-cAMP is generally lower than for its 2’,3’-counterpart. Manganese clearly is the most optimal cofactor followed by cobalt, nickel and iron. The main product is 3’-AMP. Zink leads to protein precipitation and no turnover is observed, whereas magnesium activity is barely above the EDTA inhibited level.

Defining the turnover rate by substrate consumption, only manganese resulted in full turnover after 30 min for 3’,5’-cAMP. The activity of metal bound *Sm*CpdA can be described as $Mn^{2+} > Co^{2+} > Ni^{2+} >> Fe^{2+} >> Mg^{2+}$. The protein precipitated upon zinc addition and no turnover could be observed, EDTA seems to be unable to fully remove the cofactor from *Sm*CpdA and some residual activity remains. The turnover rates are shown in the graph of Figure 42. The siderophore deferoxamine was also tested, but did come close to the inhibitory effects of EDTA.

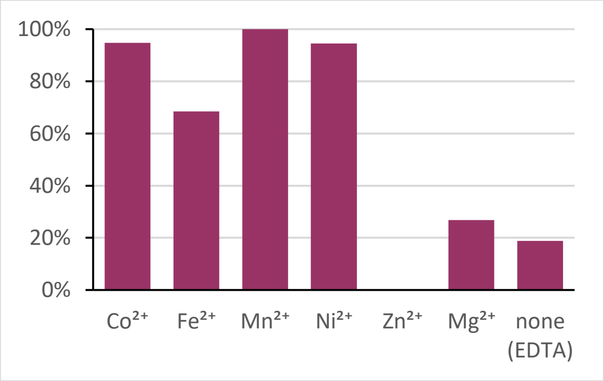


Figure 42: Cofactor dependent 3',5'-cAMP turnover: If the turnover is defined by the rate of educt consumption Mn^{2+} reached full turnover, followed by Co^{2+} and Ni^{2+}. Iron is able to hydrolyse merely

about 70% of the cyclic nucleotide after 30 min. Mg^{2+} shows very weak activity, while EDTA is not fully inhibiting the protein but rather reducing its activity by two thirds compared to the Fe^{2+}-bound complex it was added to.

Aside from the aforementioned cAMP isomers, other potential substrates were tested and the activity was plotted in Figure 43. CpdA is able to hydrolyse 3',5'-cGMP, albeit with less activity compared to adenosine homologs. The main product is 3' guanosine monophosphate with a selectivity of 95.8%, in line with the adenosine compound. For 2',3'-cGMP the turnover after 30 min is almost equal to that of the adenosine nucleotides also clearly favouring 2'cGMP with 86.2% of the product. CpdA is incapable of hydrolysing cyclic di-nucleotide mono-phosphates. For ATP a slight conversion can be detected, but at 0.43% conversion over 30 min the biological relevance seems questionable. The values deducted from the HPLC assay are listed in Table 7. *Sm*CpdA not only remains active after freezing, but no loss in activity could be observed for assays at 25, 35 and 45°C. Given the natural habitat of *S. meliloti* is soil in moderate climates, further experiments were conducted at 25°C.

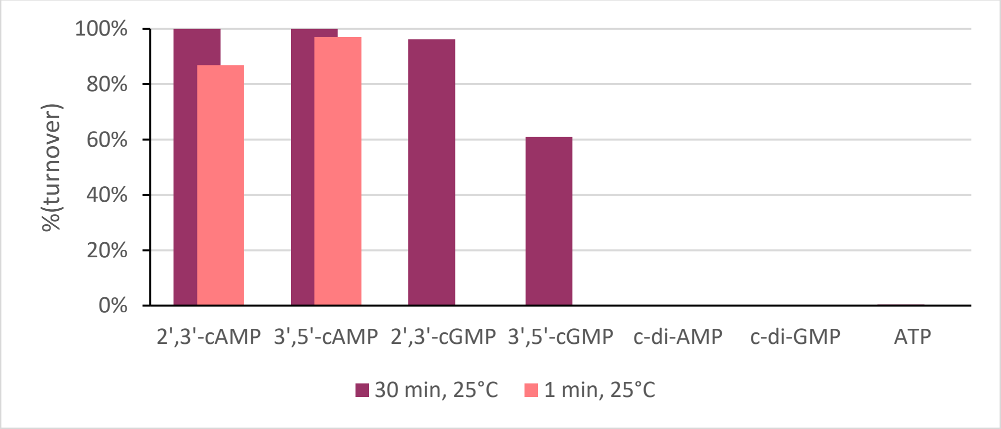


Figure 43: Substrate specificity of *Sm*CpdA-Mn: CpdA is a promiscuous phosphodiesterase able to hydrolyse cyclic adenosine and guanosine isomers at high to moderate rates. Even after as little as 1 min incubation time turnover rates above 85% are reached. The protein is unable to catalyse cyclic dinucleotide degradation nor does it possess catalytic triphosphatase or 5'hydrolase abilities as verified by MS.

Table 7: Results of the HPLC nucleotide conversion assay: Any values correspond to the final peak ratio at a given time and temperature. *Sm*CpdA is most active with manganese cofactor, whereas the iron cofactor bound in heterologous expression in E. coli is inferior in regards to 3',5'-cAMP hydrolysis.

2',3'-cAMP, 30 min, 25°C

	2'-AMP (%)	3'-AMP (%)	2',3'-cAMP (%)	Turnover (%)
Co^{2+}	92.8	7.2	0.0	100
Fe^{2+}	92.4	7.6	0.0	100
Mn^{2+}	94.8	5.2	0.0	100
Ni^{2+}	94.4	5.6	0.0	100
Zn^{2+}	Precipitated		100	0.0
None (EDTA)	5.2	0.7	94.1	5.9

3',5'-cAMP, 30 min, 25°C

	3'-AMP (%)	5'-AMP (%)	3',5'-cAMP (%)	Turnover (%)
Co^{2+}	84.4	10.4	5.2	94.8
Fe^{2+}	62.4	6.1	31.5	68.5
Mn^{2+}	91.9	8.1	0.0	100
Ni^{2+}	82.4	12.1	5.5	94.5
Zn^{2+}	Precipitated		100	0.0
Mg^{2+}	22.4	4.4	73.2	26.8
None (EDTA)	16.8	2.0	81.2	18.8

Substrate-dependent Turnover (%) at 25°C

	1 min	30 min
2',3'-cAMP	86.9	100
3',5'-cAMP	97.1	100
2',3'-cGMP	n.d.	96.3
3',5'-cGMP	n.d.	61.0
c-di-AMP	n.d.	0.0
c-di-GMP	n.d.	0.0
ATP	n.d.	0.42

To determine the kinetics of the hydrolysis reaction, ITC measurements for the degradation of 2'3'-cAMP were conducted. Two method can be employed to achieve this - the single injection method is suitable for K_M lower than 10 µM, otherwise the multiple injection method needs to be used. To get an initial appraisal of the kinetic parameters the single injection method was chosen. The reaction after the injection of 38 µL of substrate created a strong drop in heat, which would then be expected to return to a new baseline as the substrate is consumed. However, the reaction seemed to slow down over time never quite reaching a defined baseline of full consumption. The Michaelis-Menten plots display no levelling of the reaction rate towards $V_{\max,}$ and if the data are plotted in a Lineweaver-Burk plot, they also appear to be non-linear. The data are displayed in the Appendix. At this point the single injection method was abandoned in favour of the multiple injection method. The small spike after each injection is caused by the dissolution of 2',3'cAMP in buffer and a control experiment was deducted from the data prior to analysis. The exemplary Michaelis-Menten plot shown in Figure 44 gives a clue where the issue might be rooted. After an initial increase in reaction rate, higher substrate concentrations lead to a drop in speed. This is also reflected in the Lineweaver-Burk plot, which is hyperbolic, rather than linear. When the excess of substrate relative to the enzyme is increased, the effect is lessened, albeit still present. All these observations might suggest, that CpdA is product-inhibited.

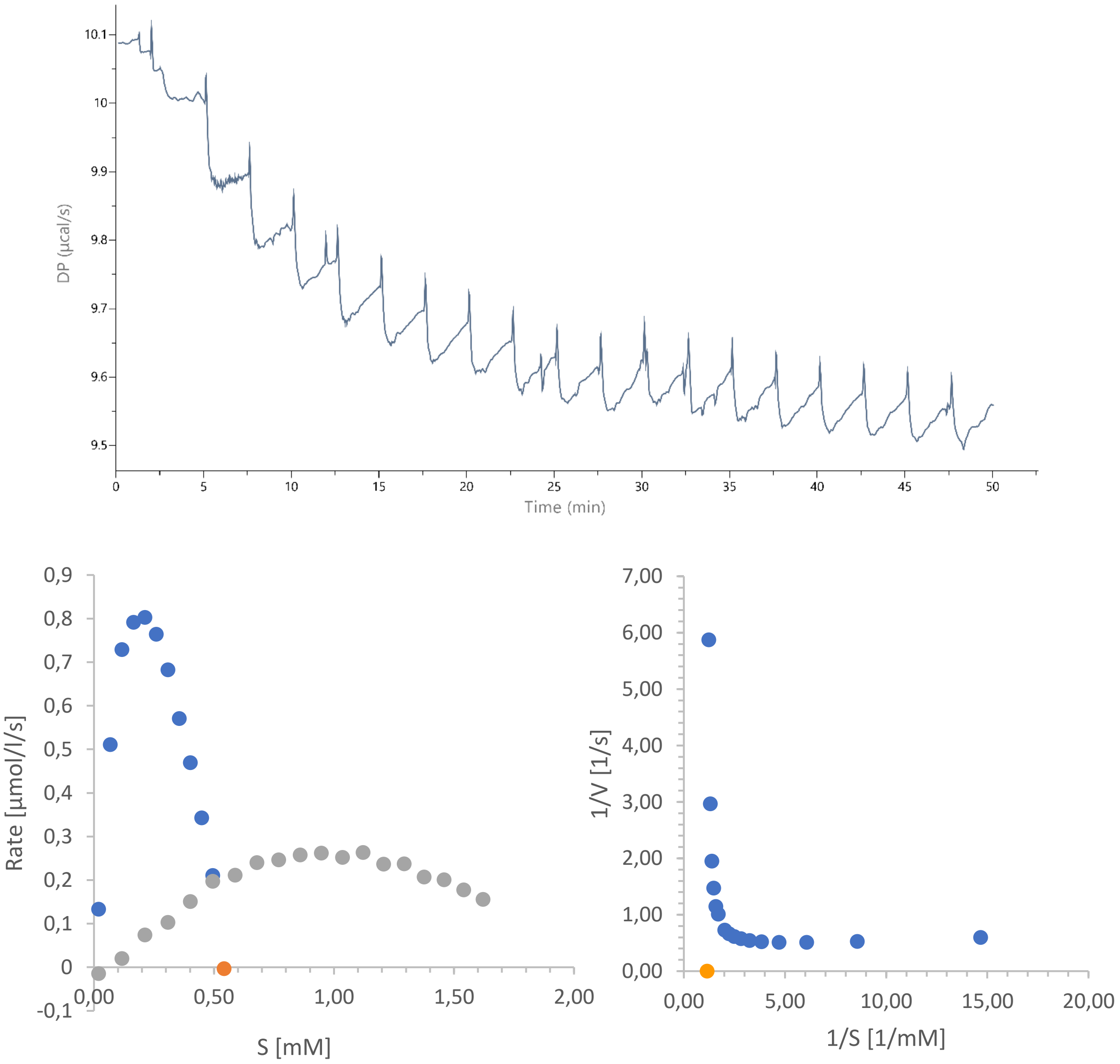


Figure 44: Kinetic measurements of 2’,3’-cAMP hydrolysis by *Sm*CpdA. Top: The raw thermogram of the kinetic measurement. Small amounts of substrate are injected each time a new baseline is established to keep the system in steady-state conditions. The blue dots correspond to 2’,3’-cAMP hydrolysis at 10 mM injected into 1.25 µM *Sm*CpdA in 1 µL increments, while the grey dots belong to 615 µM CpdA injected with 2 µL 10 mM 2',3'-cAMP each. The inhibition seems to be greater at higher protein to substrate ratios. Bottom left: In the Michaelis-Menten plot, the reaction rate is plotted onto the substrate concentration. After an initial increase in speed, the rate drops significantly. This is likely

due to product inhibition. Bottom right: The unusual hyperbolic shape of the Lineweaver-Burk plot is cause by the drop in reaction rate, through which 1/v points towards infinite.

4.2.3 Crystallisation and structure determination of *Sm*CpdA

Only one structure of a class III phosphodiesterases has been published so far. To gain more insight into the special features of this protein class, *Sm*CpdA was crystallized. Purified protein was concentrated using a cut-off of 3 kDa to a concentration of 17 mg/mL (556 µM). Crystals appeared within 3 days in a solution containing 0.2 M sodium acetate, 0.1 M Tris-HCl pH 8.5 and 30% (w/v) PEG 4000 (*Nextal* Classics Suite H6) at 8°C and diffracted to 1.4 Å. *Sm*CpdA can be crystallized in a solution containing 2 M ammonium phosphate and 0.1 M Tris adjusted to pH 8.5 using ammonia (MBClass II A3) as well, however the crystals have lower diffraction quality. Table 8 illustrates the difference in crystal quality for the two conditions. Different morphologies for phosphate bound *Sm*CpdA-Fe^{2+} crystals can be seen in Figure 45.

Table 8: Data collection statistics for phosphate-bound CpdA-Fe^{2+} crystals. Datasets one to three were obtained from crystals grown in the MBClass II condition, while the crystals corresponding to datasets four to seven were grown in the Classics Suite. While all crystals appear in the same space group, the unit cell parameters for the MBClass II crystals fluctuate a lot more, which is also reflected in crystal quality. Crystals grown in the Classics Suite were a lot more consistent in unit cell size and have a good resolution.

No	space group	unit cell parameters [a, b, c, α, β, γ]	resolution [a, b, c]	reflections	redundancy	completeness	R_{merge}	I/σ
1	P2	46.5, 82.1, 61.1, 90, 90.8, 90	2.5, 1.9, 2.2	85104	1.2	64.7%	12.3%	4.8
2	P2	78.0, 81.8, 112.6, 90, 97.5, 90	2.4, 2.4, 2.5	156539	2.9	93.0%	11.7%	4.8
3	P2	46.2, 82.0, 60.33, 90, 91.5, 90	2.7, 2.4, 3.2	37142	1.2	67.6%	11.1%	3.9
4	P2	46.8, 81.5, 63.0, 90, 91.6, 90	1.6	210839	1.6	88.3%	6.6%	5.4
5	P2	46.9, 81.9, 63.0, 90, 91.4, 90	2.0, 1,8, 1.7	105008	1.1	69.3%	12.3%	5.3
6	P2	46.8, 81.8, 62.8, 90, 91.2, 90	1.5	264854	1.7	90.1%	5.6%	7.3
7	P2	46.8, 81.7, 62.8, 90, 91.3, 90	1.5	245993	1.6	85.9%	5.0%	7.4

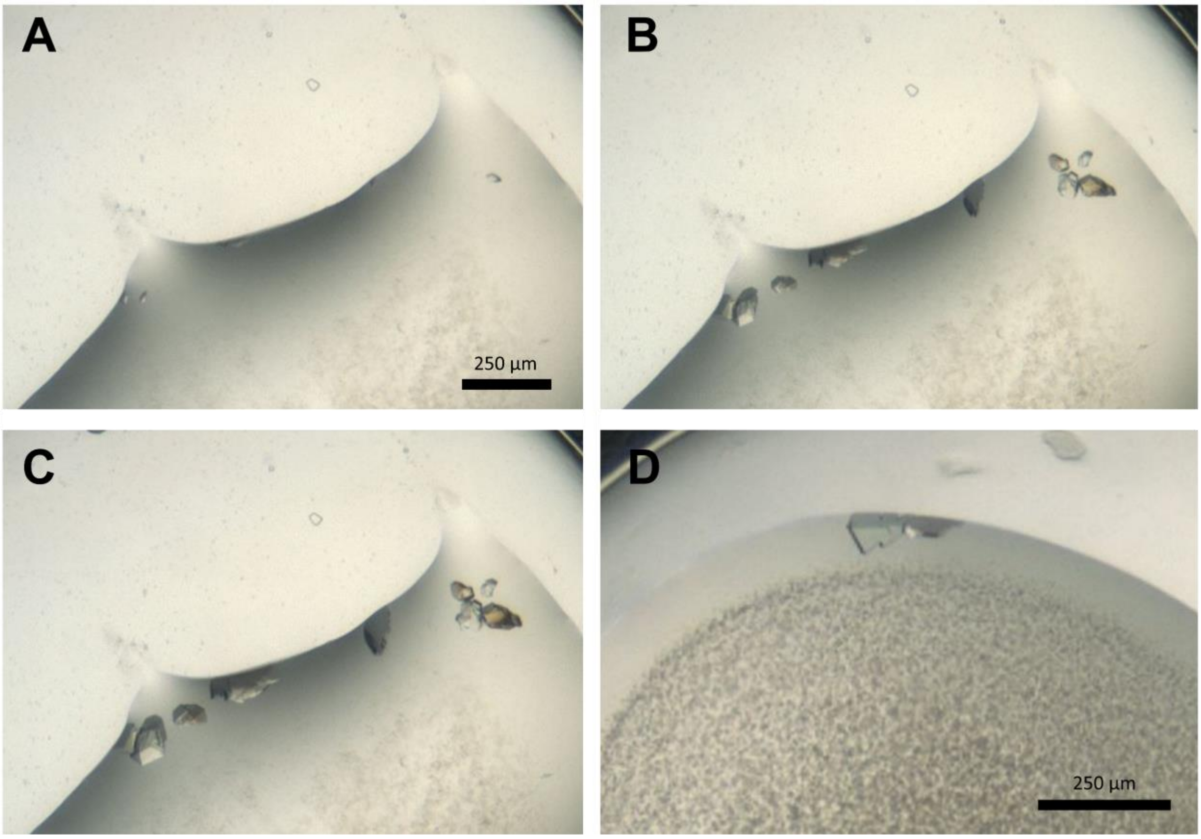


Figure 45: Crystal morphologies for phosphate-bound *Sm*CpdA crystals containing iron cofactors: (A, B, C) Crystals appeared within 12 hours and reached their final size after three days. They reached a resolution of 1.6 Ångstrom. (D) Many crystals are two-dimensional plates rather than the three-dimensional crystals shown in (C). This particular crystal contained some additional density aside from phosphate in the active site and diffracted to 1.6 Å as well. The significance of the differences within the active site are discussed in 4.3.5. The crystals appeared after 6 days and reached their final size after 22 days.

*Sm*CpdA crystallizes in thin plates with reproducible diffraction quality of 1.5-1.8 Å. Xray fluorograms of these crystals always result in three distinct peaks, which can be assigned using the NIST X-ray transition energies database. CpdA overproduced in *E. coli* always contains iron cofactors, as well as a bromide impurity. The third peak at 12.39 keV corresponds to the excitation energy. The origin of the bromide signal is obscure, as neither the crystallisation solution nor any solutions prior contain it. The only possible explanation is the contamination of a reagent used in the crystallisation solution, but as no apparent corresponding density appears in the crystal structure, it should be of no consequence. If the cofactor is removed with EDTA and supplemented with manganese chloride prior to size exclusion chromatography, no crystals appear in the crystallisation conditions and variations of it in regards to pH and precipitant concentration. The crystallisation can be facilitated by the addition of seed crystals;

however, the resulting crystals consist of overlaying plates (Figure 47). Even seed dilutions of 1/2000 do not resolve this issue. Thus, even though manganese is the most active cofactor of *Sm*CpdA, only the Fe^{2+}-bound crystal structure could be obtained. Superimpositions of phosphodiesterase structures (e.g., PDB-ID: 3IB8 and 3D03) show however, that the type of metal does not impact the active site architecture, so that the significance of the crystal structure should not be reduced by this issue.

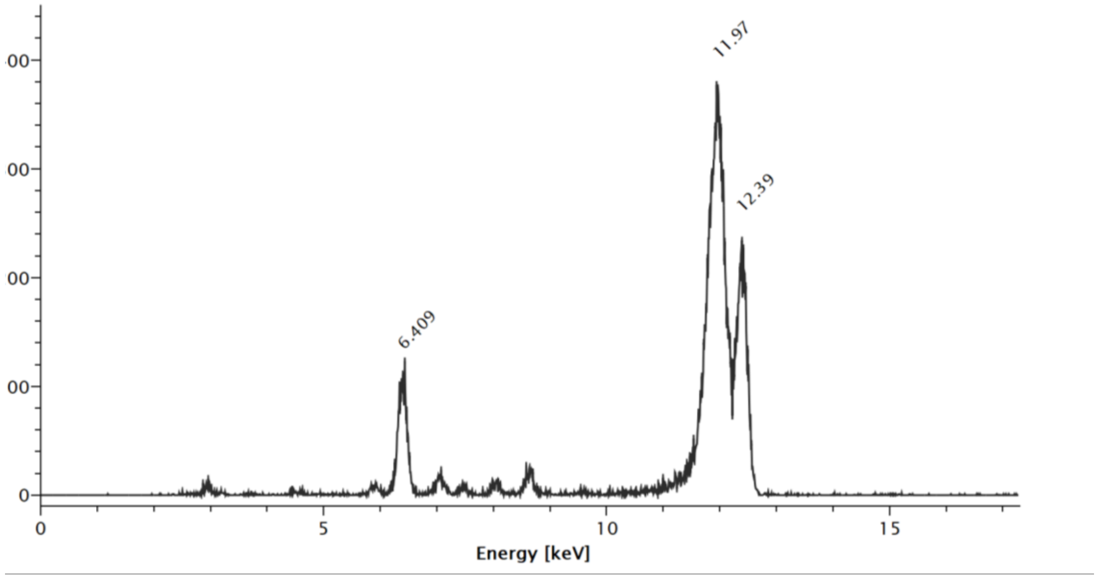


Figure 46: X-ray fluorescence measurement for *Sm*CpdA crystal. All crystals that were measured resulted in a spectrum with three distinct peaks. As a result only one example is pictured here. The X-ray transition energies can be asigned using the NIST database. The iron KL$_3$ transition has an energy of 6403 eV, which fits perfectly with the first peak. The peak found at 11.97 keV corresponds to the KL$_3$ transition of bromide. The excitation was conducted at 13.39 keV.

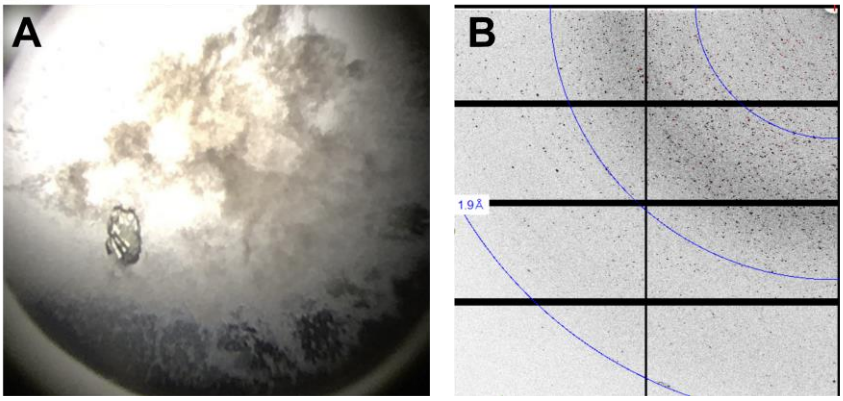


Figure 47: Microseeding in manganese-bound CpdA: (A) If the cofactor was exchanged to Mn^{2+}, seed crystals were required for nucleation. The crystals grow in imbricate very fine plates. (B) The overlapping of multiple plates results in a diffraction pattern, in which a myriad of lattices are inter-leaved. Even the edges of the crystals displayed this, while the resolution itself is below 2 Å.

Images from the best *Sm*CpdA crystal were recorded at PXI beamline at the Swiss Light Source (SLS) in Villigen (Switzerland) and indexed and processed using XDS. The raw data was scaled using the UCLA Diffraction Anisotropy Server and a free *R* set based on 5% of the data was generated. The dataset had a completeness of 89.60%, an R_{merge} of 6.8% and an overall I/σ of 12.9. The crystals had the orthorhombic space group $P12_11$ and a Matthews coefficient of 1.95 Å^3/Da indicated the presence of two *Sm*CpdA molecules per unit cell at a rather low solvent content of 37.08%. The cell content was estimated using CCP4i2. Molecular replacement was conducted using BALBES in CCP4 online, which generated a model based on Rv0805 from *M. tuberculosis* (PDB-ID: 3IB7). The protein structure has a sequence identity of 25.0% with *Sm*CpdA. After molecular replacement the coordinates were refined against the experimental data using Phenix Refine. After an initial Autobuild step, the refinement was alternated with manual model building in Coot, until reaching a R_{free} of 19.5% and an R_{work} of 17.4%. The relevant statistics are listed in Table 9.

Table 9: Data collection and Refinement statistics for phosphate-bound *Sm*CpdA crystals.

Data collection		**Refinement**	
X-ray source	SLS PXI	Resolution range (Å)	46.46 - 1.35
Space group	$P12_11$	$R_{\mathrm{work}}/R_{\mathrm{free}}$ (%)	0.153/ 0.177
Unit-cell parameters (Å, °)	a=46.5, b=81.7, c=63.0, α=90, β=91, γ=90	Average B factor (Å^2)	23.33
Wavelength (Å)	1.00	No. of atoms	
Resolution range (Å)	46.46 - 1.35 (1.40 - 1.35)	Total	4588
Completeness (%)	90.14 (42.60)	Protein	4061
Observed reflections	593280 (18572)	Ligand	9
Unique reflections	93623 (4374)	Solvent	518
Multiplicity	6.3 (3.6)	R.m.s.d., bond lengths (Å)	0.012
Wilson B factor (Å^2)	15.04	R.m.s.d., bond angles (Å)	1.5
R_{merge} (%)	0.0681 (1.077)	Ramachandran plot	
R_{measure} (%)	0.0739 (1.258)	Favoured (%)	95.89
Mean $I/\sigma(I)$	12.94 (1.21)	Allowed (%)	3.74
$CC_{1/2}$ (%)	0.998 (0.629)	Outliers (%)	0.37

Overall, 81 *Sm*CpdA crystals were measured and 31 datasets were obtained. Every single one of them contains phosphate in varying occupancies (0.6-0.9) within the active site. As *Sm*CpdA has been shown to be quite active, phosphate cannot be tightly bound. As high affinity of phosphate to the active site obviously is not the issue, the observation likely stems from a stabilisation of the crystal form by phosphate binding.

4.2.4 *Sm*CpdA is a Class III phosphodiesterase

Cyclic nucleotide monophosphate degradation is an essential function in bacteria and eukaryotes and is accomplished via cNMP phosphodiesterases (PDE). These proteins can be divided based on sequence similarly into three distinct classes. Class I phosphodiesterases are present only in eukaryotes with proteins like PDE1 in humans and structurally and functionally well characterized. Several inhibitors are available for this class as well. Class II phosphodiesterases are a relatively small group present mainly in lower eukaryotes and γ-proteobacteria. They have a β-lactamase fold and often possess signals for membrane targeting or secretion. The third and newest class of phosphodiesterases, class III, is found solely in bacteria (Matange 2015a). They were discovered by Richter 2002 when realizing that the *E. coli* phosphodiesterase CpdA did not fit either of the previous groups. Rather class III phosphodiesterases belong to the superfamily termed calcineurin-like metallophosphoesterases, which also comprises a diverse range of phosphodiesterases like nucleosidases and glycerophosphodiesterases. To this date, only one class III homolog, *M. tuberculosis* Rv0805, has been structurally characterized by Shenoy et al. 2007. It is now joined by *Sm*CpdA, giving further insight in this novel cNMP phosphodiesterase class.

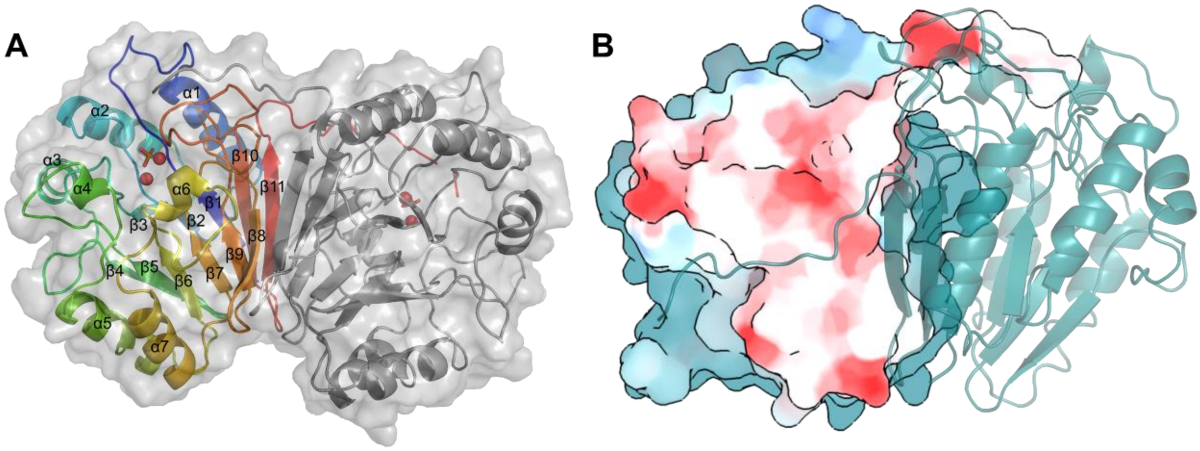


Figure 48: General features of the *Sm*CpdA crystal structure: (A) The protein forms dimers *in crystallo*, with the active site turned away from the interface. It is occupied by phosphate bound to two iron atoms. Secondary structure elements are indicated in the left chain going from blue colouring at the *N*-terminus to red on the *C*-terminus. (B) The dimer formation in the crystal assembly is stabilized by the *C*-terminus interlocking with the neighbouring molecule, reaching towards the active site.

*Sm*CpdA is a globular protein comprised of two central β sheets containing strands of mixed orientation surrounded by α helices. This βαβαβ fold is typical for metallophosphoesterases. The fold contains about equal amounts of helices (24%), strands (24%) and turns (29%), with 28% coil as analysed with STRIDE (Frishman and Argos 1995). A high number of proline residues are located within the turns giving them some rigidity. The active site is facing outwards on top of the central beta barrel motif and is very open to the solvent. While it elutes as a monomer on size exclusion, *Sm*CpdA crystallises as dimer. The assembly is stabilized by backbone peptide stacking in the central β-strands, as well as an interaction of H254 of β11 with the backbone of S165 in the *C*-terminal loop to α6 in the neighbouring molecule. The *C*-termini are wrapped around the adjacent molecule towards the outward facing active site. As the histidine tag is never fully resolved, this part seems to be rather flexible. This assembly appears to be driven by the interaction of the affinity tag histidines with the active site metals and might impair substrate binding in the present crystal form (Figure 48).

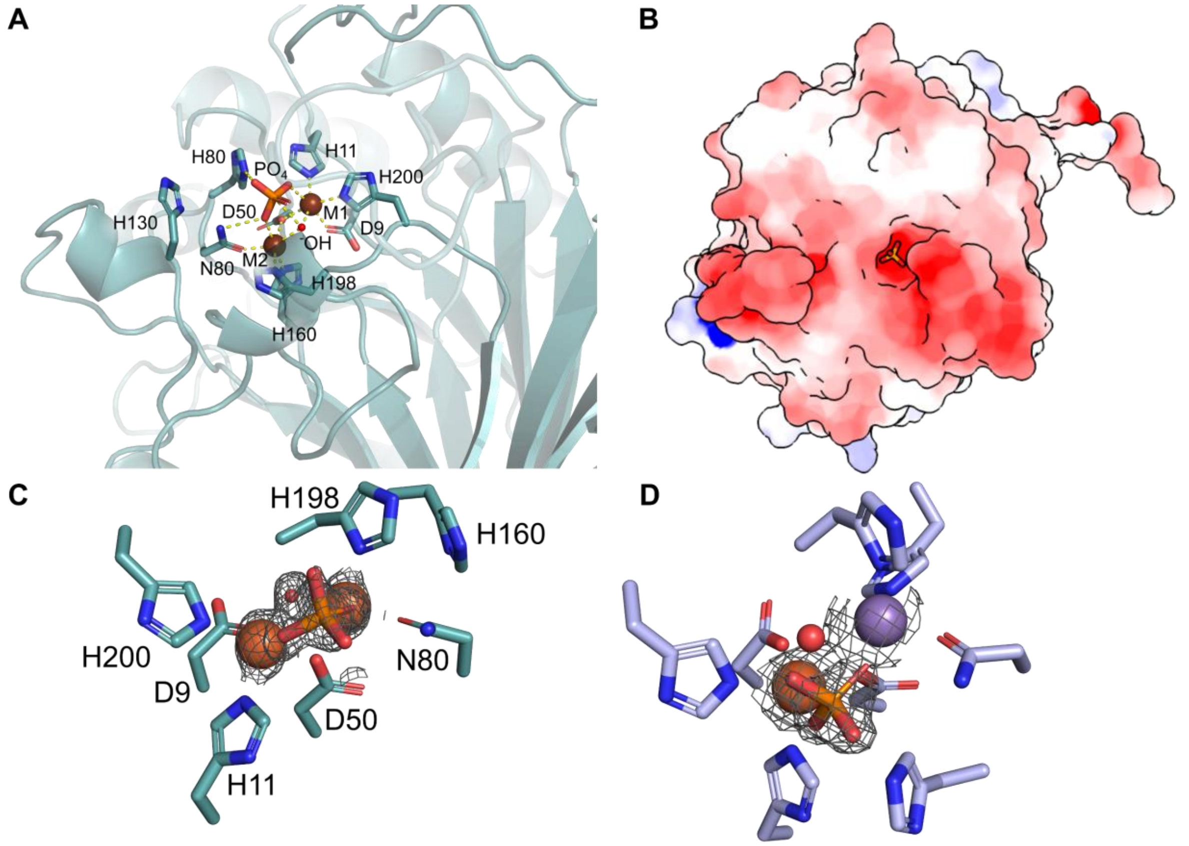


Figure 49: The active site of the phosphodiesterase: (A) Two iron molecules (M1, M2), a catalytic hydroxide, as well as the phosphate are each octahedrally coordinated by a cage of highly conserved histidines, asparagines and aspartic acids. (B) The binding site is quite open to the solvent. (C) The catalytic hydroxide is well resolved in the electron density and is situated perfectly for backside attack to the phosphorus. This is the first time the nucleophile is clearly visible and supports the proposed mechanism. (D) For Rv0805 from *M. tuberculosis* (PDB-ID: 2hy1) the density for the hydroxide is not resolved.

All crystals obtained for *Sm*CpdA have phosphate bound to the active site. The active site also contains two metal ions octahedrally coordinated by histidine and aspartic acid, which could unambiguously be identified as iron via X-ray fluorescence. D9 and H11 within the loop C-terminal to β1, as well as D50 located C-terminally to β2 and H202 N-terminal of β8 are coordinating iron 1 (M1). Iron 2 (M2) is in turn interacting with N80 C-terminal of α4, H159 near β6 and H200 N-terminal of β8, as well as the aforementioned D50. The phosphate interacts with the metals and makes additional contacts with N80 and H81. The metals have an occupancy of around 80% *in crystallo* and the assembly of the active site as well as the already discussed activity assays strongly support the notion that the activity of class III phosphodiesterases is strictly dependent on metal binding. A hydroxide is coordinated by these metals in a S_N2 trajectory to the phosphate. The identity of the nucleophile as hydroxide was established in Hybrid Quantum Mechanical/ Molecular Mechanical Calculations (Xiong et al. 2006; Chen and Zhan 2004). Previous structures of the class III phosphodiesterase Rv0805 from *M. tuberculosis* were lacking this critical active site component, while in our structure it is clearly resolved (Podobnik et al. 2009). Shenoy et al. 2007 claimed to see this catalytically active ion, but the experimental data reveals a lack of corresponding electron density (Figure 49).

4.2.5 *Sm*CpdA crystals seem to be incompatible with substrate binding

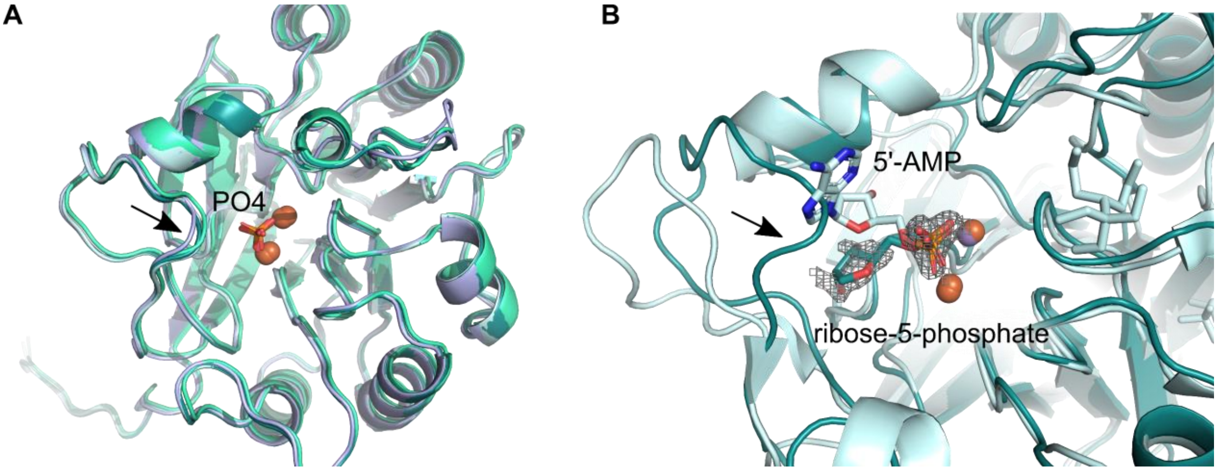


Figure 50: Potential substrate binding in the reported crystal form: (A) Multicrystal alignment of *Sm*CpdA shows very little flexibility in the loop located on top of the active site (indicated with an arrow). (B) This loop is incompatible with the 5’-AMP binding mode observed in Rv0805 (PDB-ID: 3ib8, light grey). In a single dataset additional density consistent with the ribose could be observed after soaking with 2’-/3’-AMP mix (teal). The ligand might have potentially displaced the loop in a small fraction of the crystal.

During this thesis a plethora of soaking experiments ranging up to 200 mM ligand solutions over 24 h were conducted. While the crystals still diffracted, ligand density remained elusive. This seems surprising, given how open the active site is to the solvent. An alignment of multiple *Sm*CpdA crystal structures in Figure 50 shows very little diversity in conformation from one crystal to another. A single crystal could be produced however, in which some density consistent with a ribose moiety was observed. The density is poorly resolved and building of 3’-AMP results in a clash with the loop N-terminal of β10. It is likely that the ligand was able to displace said loop only in a small fraction of the crystal, resulting in this state. The same clash with the loop can be observed for a superimposition of the 5’-AMP bound state of Rv0805 (PDB-ID: 3ib8). All of this suggests, that our crystal form is incompatible with substrate binding. This doesn’t hold true in solution, as clearly shown in the activity assays. The location of the loop blocking the active site is stabilized by backbone interactions α4 and even more so with the *C*-terminus of the protein. *C*-terminal truncation or co-crystallisation might be a solution to this issue in the future.

4.3 CyaD1 CHASE2 regulatory domain

CyaD1 is a membrane-anchored adenylate cyclase, that is activated to its multimeric form via a *N*-terminal CHASE2 regulatory domain located on the exoplasmic side. It is potentially able to convert 3’,5’-cyclic guanylate as well, as its cyclase homology domain (CHD) shows high congruity with bifunctional species. The inactive monomeric form of the CyaD1 CHD domain has been crystallized in our lab by Ankan Banerjee (unpublished). For switching to the active assembly, CyaD1 requires the regulatory domain, that had been truncated for solubility reasons in this instance. The dimeric active state can be reconstructed using CyaC from *Arthrospira platensis* as shown in Figure 51 and reveals two active sites in the point-symmetrical dimer interface. Two metal atoms are coordinated by aspartic acids and interact with the α and β phosphates, while the sugar and base are coordinated by the other inverse monomer. The CHASE2 regulatory domain has been found in many A/G cyclases and protein kinases, but structure and interaction partners remain elusive.

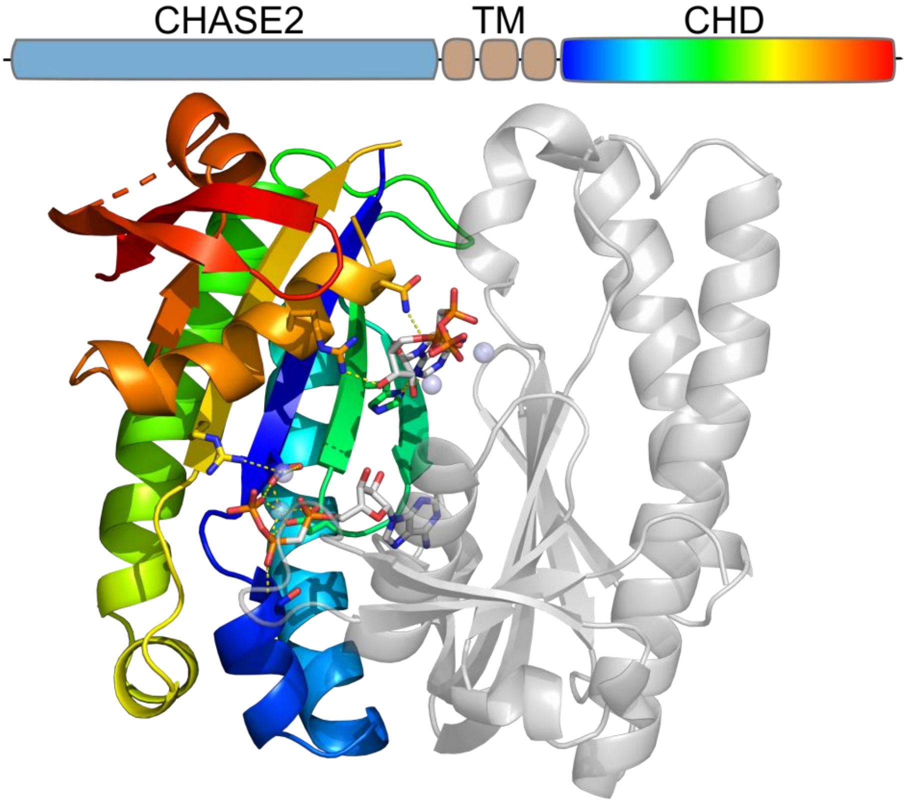


Figure 51: Crystal structure of the CyaD1 CHD domain: CyaD1 consists of a *N*-terminal regulatory domain called CHASE2, which is extracellularly localized. Three transmembrane helices (TM) link it

to the cytosolic cyclase homology domain; The crystal structure solved by Ankan Banerjee (unpublished, rainbow coloured) contains the monomeric inactive form of the CHD domain. It can be complemented using the structure of CyaC from *Arthrospira platensis* (PDB-ID: 1wc5, grey). The reconstructed assembly contains two active sites within the dimer interphase, in which two metal ions (calcium and magnesium in CyaC) coordinate an adenosine triphosphate (ATP) each and catalyse the cyclisation.

The CHASE2 domain in CyaD1 has been predicted to by exoplasmic and is *C*-terminally followed by three transmembrane domains. An analysis via the TMHMM Server for prediction of transmembrane helices reveals an additional *N*-terminal patch that might be membrane anchored. Residue 7-29 are predicted to be transmembrane helices, followed by the CHASE2 domain from L30 to R315, three transmembrane helices from G316 to A338, A345 to S367 and F372 to A394 respectively. The cyclase homology domain (CHD) is located at the cytoplasmic side. Taking the *N*-terminal patch into account, the potentially soluble part of CHASE2 in CyaD1 consists of 286 amino acids with a molecular weight of 30.0 kDa. Both the full length and a by 45 AA *N*-terminally truncated variation of CHASE2 were tested for solubility by Ankan Banerjee. Either are insoluble and do not respond to refolding. This is not surprising when the composition of the domain is analysed. Its major components are alanine (21.2%) and leucine (8.7%). The protein is rated unstable by Expasy ProtParam.

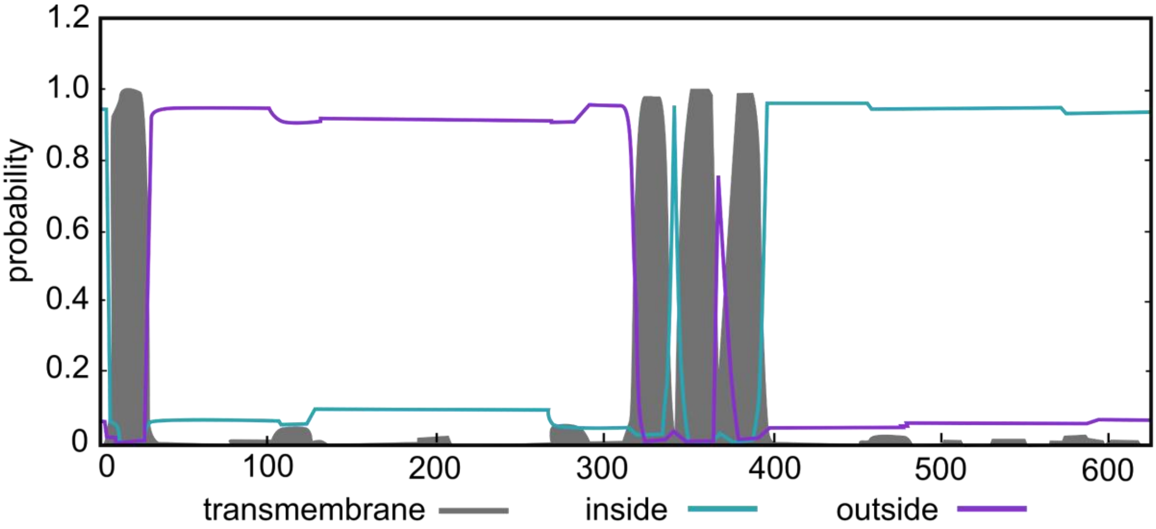


Figure 52: TMHMM prediction of transmembrane helices of CyaD1. The protein likely has a *N*-terminal membrane anchored region, followed by the exocellular CHASE2 regulatory domain, three transmembrane helices and the cytoplasmic cyclase homology domain. Based on the prediction the CHASE2 domain was *N*-terminally truncated for purification. The cyclase homology domain could by purified and crystallised in its inactive monomeric form by Ankan Banerjee.

CHASE2 homologs from the thermophilic organisms *Thermosynechococcus elongatus*, *Myxococcus xanthus*, *Synechococcus elongatus PCC 6301* and *Deinococcus radiodurans* were *N*-terminally truncated and testexpressed in the hope of finding an increase in solubility. The high content of hydrophobic amino acids already suggests, that CHASE2 might be a peripheric membrane protein after all. All constructs were codon-optimised and cloned in pET28a (Table 10). The sequences are listed in the Appendix.

Table 10: Properties of the CHASE2 homologs tested for solubility. The high ratio of hydrophobic amino acids already suggests, that CHASE2 might be a peripheric membrane protein.

No	Organism	UniProt-ID	Protein function	truncation	length [AA]	MW [Da]	% unpolar AA
C1	*Sinorhizobium meliloti*	Q92SD3	Adenylate/guanylate cyclase	45-270	225	25140	60.2
C2	*Thermosynechococcus elongatus*	Q8DIP1	Serine/threonine protein kinase	84-376	293	32340	54.1
C3	*Myxococcus xanthus*	Q1D4J2	Adenylate/guanylate cyclase	48-365	317	33495	63.1
C4	*Synechococcus elongatus PCC 6301*	A0A0H3K3Y5	Adenylate/guanylate cyclase	36-303	267	29472	56.3
C5	*Deinococcus radiodurans*	Q9RTW9	GGDEF family protein	38-272	234	24974	58.7

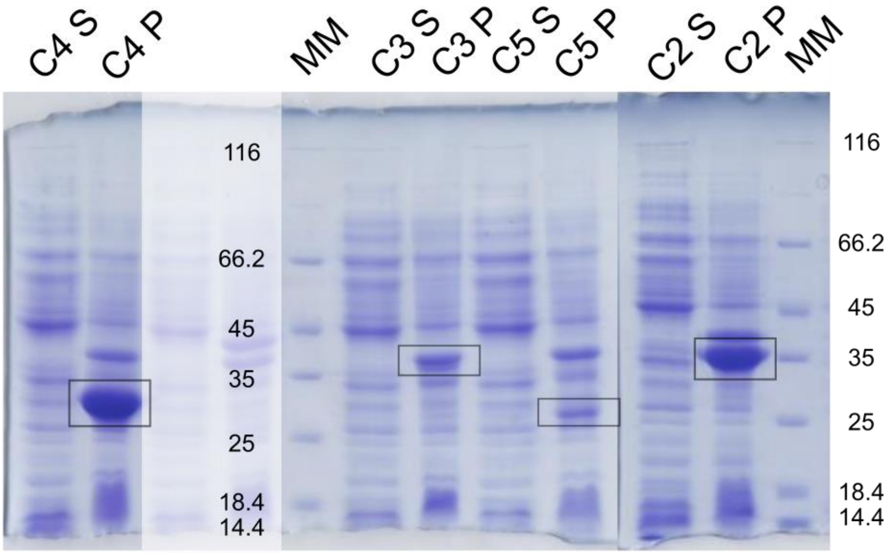


Figure 53: Expression of CHASE2 homologs at 30°C induced with 0.1 mM IPTG for 4 h. The CHASE2 homologs from *Thermosynechococcus elongatus* (C2), *Myxococcus xanthus* (C3), *Synechococcus elongatus PCC 6301* (C4) and *Deinococcus radiodurans* (C5) were expressed at 30°C after induction with 0.1 mM IPTG. While the overproduction is good for C2 and C4, only small amounts of C3 and C5 are present. All proteins appear solely in the insoluble pellet (P) and not in the supernatant (S).

BL21 (DE3) Gold cell were grown at 37°C in 50 mL LB medium to an OD_{600} of 0.5 and induced with 0.1 mM IPTG. The cells were harvested after 4 h at 30°C and sheared via bead beating. Insoluble components were pelleted at 18111g for 1 h, the supernatant was removed and the pellet was resuspended in SDS sample buffer in another bead beating circle. A 12% SDS-PAGE gel reveals that only the CHASE2 domains from Thermosynechococcus elongatus (C2) and *Synechococcus elongatus PCC 6301* (C4) are overproduced in high amounts (Figure 54). Nevertheless, all proteins are insoluble.

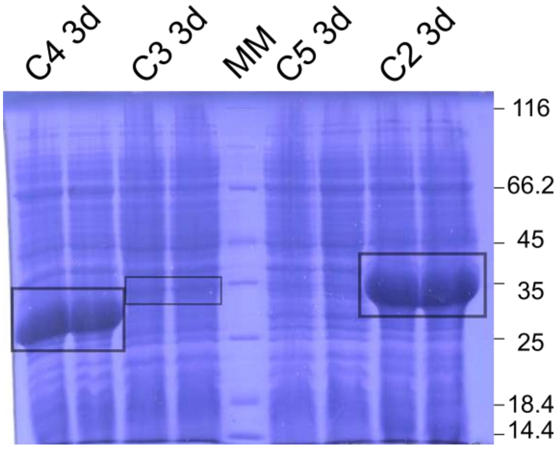


Figure 54: Overproduction of CHASE2 homologs at 12°C induced with 0.1 mM IPTG for three days. For the homolog from *Thermosynechococcus elongatus* (C2) and *Synechococcus elongatus PCC 6301* (C4) the protein amount seems to be unchanged, while the signal for *Myxococcus xanthus* (C3) seem less strong and for *Deinococcus radiodurans* (C5) no protein band could be assigned.

A second expression attempt was made using the same protocol outlined above, but reducing the temperature to 12°C with an expression time of three days. The cells were harvested and sheared and pellet and supernatant samples were analysed on 12% SDS-PAGE gels. The experiments pertaining CHASE2 were conducted by Alexander Lichtenberg. The overproduction of *Thermosynechococcus elongatus* C2 and *Synechococcus elongatus PCC 6301* C4 is unchanged, however *Myxococcus xanthus* C3 can be barely made out and *Deinococcus radiodurans* C5 is only present in invisible amounts, if at all. Like the experiment at higher temperatures, C2 and C4 are fully insoluble and C3 and C5 cannot be identified after lysis.

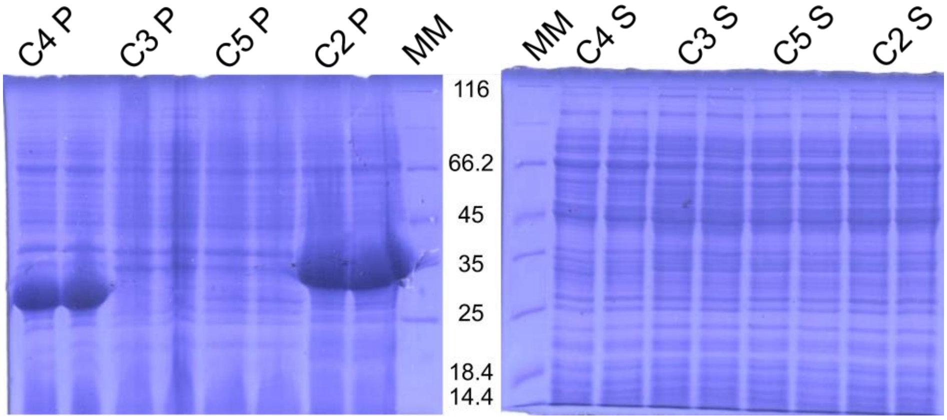


Figure 55: Solubility of CHASE2 homologs at 12°C induced with 0.1 mM IPTG for three days. The 12% SDS-PAGE gels containing the insoluble pellet (P, left gel) and supernatant (S, right gel) show no soluble protein and poor expression in case of C3 and C5.

A bioanalysis of the CHASE2 domain using the Gremlin Server by Bakerlab was conducted. The method uses a multiple sequence alignment of homologous protein sequences analyses the co-dependency of conservation and coevolution in residues. It produces a contact map with residues that show correlated mutations. The first 315 amino acids of CyaD1 were input into the server leading to 1681 homologous sequences. The coevolution analysis resulted in the contact map presented in Figure 56, which suggests two alpha-helical parts that interact with each other and flank a central unordered region. Based on the contacts, a model was generated using the Robetta protein structure prediction server from Bakerlab as well. The server is an extension of the Rosetta modelling software and allows the generation of models, when structure homology-based approaches like the SwissModel server fail due to a lack of homologous structure templates.

Overall, CHASE2 still leaves a lot to be explored. It’s structure and interaction partners remain elusive, however based on our experiments it is likely a membrane-associated protein in an alpha-helical barrel fold.

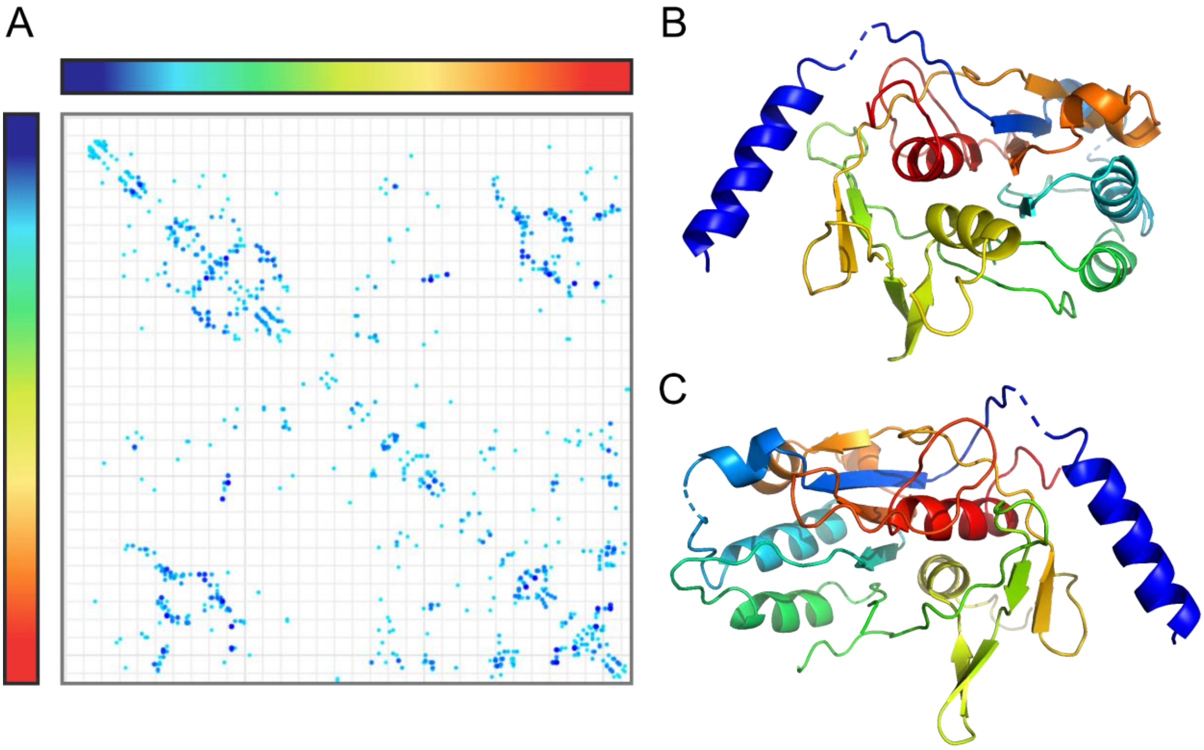


Figure 56: Bioanalysis of CHASE2 using the Servers from Bakerlab. (A) The Gremlin coevolution analysis results in a contact map of co-dependent residues. The map reveals alpha-helically ordered *N*- and *C*-termini that interact with each other. The central part of the domain shows very little order. (B, C) The resulting contacts were input in the Robetta modelling software to generate a model. The interaction of the light blue and green helices with the red and orange ones is consistent with the contact map and might form an alpha-helical barrel. The assignment of the *N*-terminal patch as alpha helical seems reasonable, but the placement of this and the central part of the domain need to be taken with a grain of salt.

5 Discussion

5.1 The CyaD1 locus regulates host infection

The importance of cyclic nucleotide monophosphate levels in the establishment of *Sinorhizobium meliloti* – *Medicago* symbiosis has been demonstrated in cyclase knockout mutants, resulting in hyperinfection phenotypes. In the same study by Tian et al. 2012 a cAMP receptor protein-like transcription regulator named Clr was reported in the CyaD1 locus within the chromosome. CyaD1 itself is one of astonishingly 28 different adenylate cyclase isozymes in *Sinorhizobium meliloti* (Figure 57). CpdA (*Sm*c02179, sometimes SpdA in literature) is a class III phosphodiesterase located in the same locus as well and has been discussed as Clr activation antagonist. In *E.* coli CpdA expression is positively regulated by the Crp homolog CAP (Kim et al. 2009), however the same could not be shown for *S. meliloti*. *Sm*c02177 is a FecR domain containing protein, a class that has been shown to function as regulators of iron transport (van Hove et al. 1990). This gene is strictly required for the control of secondary plant infection and was found to depend on Clr by Zou et al. 2017. *Sm*c02178 is upregulated by Clr (Mathieu-Demazière et al. 2013) and is likely a Sel1-like repeat protein, a class of α-helical proteins frequently involved in signal transduction. Their precise function and interaction partners are yet to be determined.

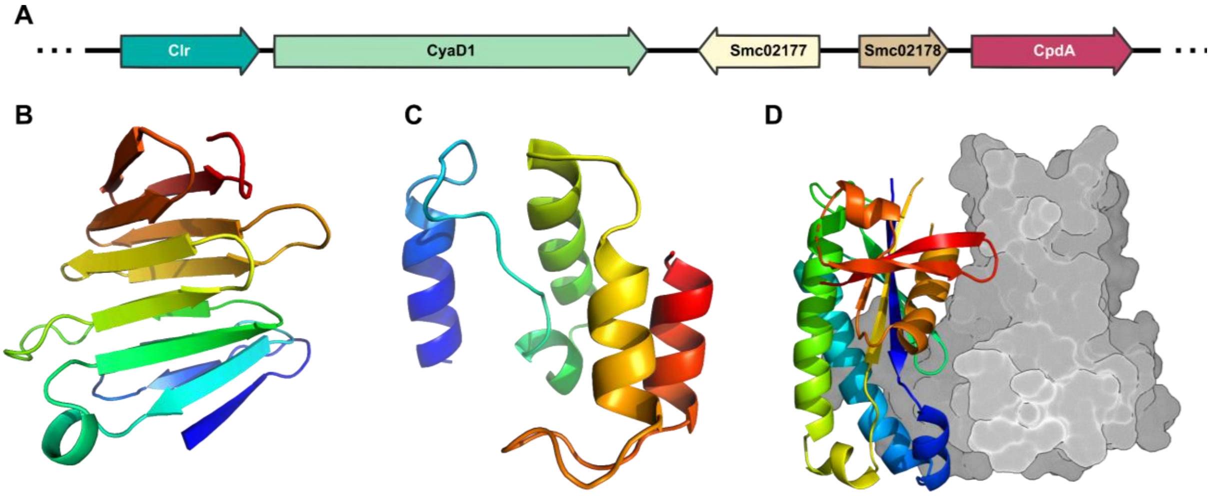


Figure 57: The CyaD1 locus regulates symbiosis establishment: (A) The gene orientation within the CyaD1 locus located in the *S. meliloti* chromosome shows that multiple genes shown to be involved in symbiosis establishment are located in very close proximity and likely dependent on each other. (B) The FecR domain containing *Sm*c02177 was modelled after PubR from *Pseudomonas capeferrum* (PDB-ID: 6ovk) using the SwissModel server. (C) The homology model of Sel1-like protein *Sm*c02178 based on Kpn_04481 from *Klebsiella pneumoniae* reveals a fully α-helical protein. (D) The crystal structure of

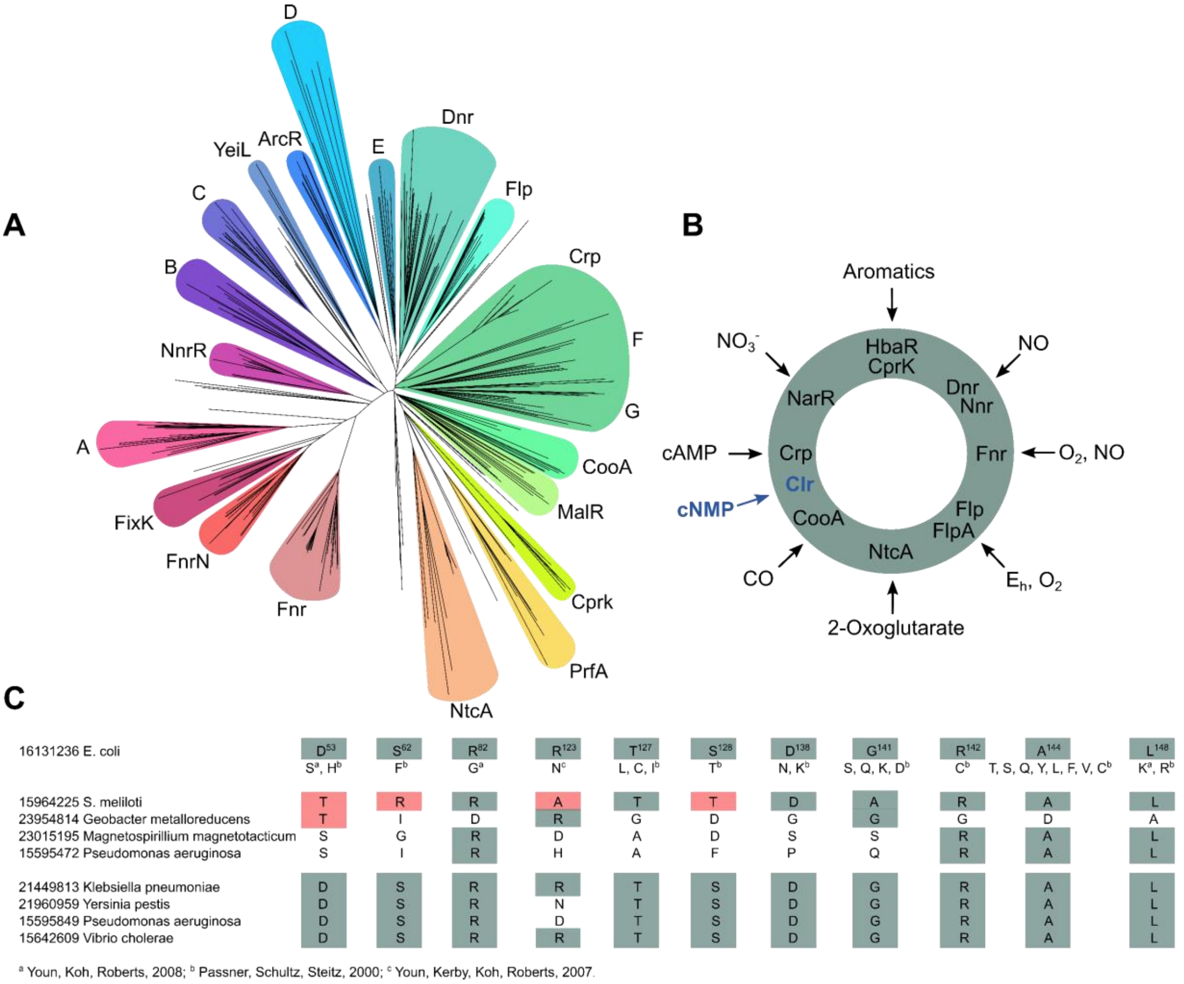
the monomeric CyaD1 could be solved by Ankan Banerjee. It complemented with the surface of the other dimer half using the structure of CyaC from *Arthrospira platensis* (PDB-ID: 1wc5).

CyaD1 is one of *S. meliloti's* 28 class III adenylate cyclases. Together with CyaD2 (*Sm*c04307) and CyaK (*Sm*b20776) the three receptor-like cyclases control secondary host infection (Zou et al. 2017). All three have the same domain structure consisting of a *N*-terminal CHASE2 regulatory domain, three transmembrane helices and a cyclase homology domain. While they appear to be activated by the same signal, it has been suggested they are active in different stages of host infection. Tian et al. 2012 proposed that CyaD1 possibly might have a guanylate activity as well. While in this publication they found that CHASE2 deletion led to significantly increased cAMP levels *in vivo*, no activity could be observed for the CyaD1 CHD domain *in vitro* in our lab. For NsrA an interaction of CHASE2 with FecR domains has been suggested (Garnerone et al. 2018), so a possible deduction might be, that *Sm*c02177 provides the signal for cAMP production by CyaD1. The fact that in *E. coli* ferric uptake genes are also regulated by the cAMP receptor protein, supports this notion (Lorenzo et al. 1988).

The transcription factor Clr has been shown to regulate in a bifunctional cAMP/cGMP dependent manner and has been structurally characterized in this work. The phosphodiesterase CpdA as a potential Clr antagonist has been investigated as well. We also examined several CHASE2 domain homologs and found them to be likely membrane associated.

5.1 Clr is the first bifunctional cNMP receptor protein

Bacteria need to adapt to their extremely diverse environment, which requires a fast response to changing conditions. The superfamily of Crp-Fnr transcription regulators provides a very versatile way of metabolic reprogramming. The class is named after its first members to be discovered, the cAMP receptor protein (Crp) and fumarate and nitrate reductase regulator (Fnr). It contains a myriad of different protein subclasses, usually defined by their most prominent representative like carbon monoxide sensing protein CooA, nitrogen-fixation regulator FixK or arginine catabolism regulatory protein ArcR (Körner et al. 2003). Each subclass has a distinctive mechanism of activation (Green et al. 2001). The diversity in activating signals also means that generally multiple Crp-Fnr transcription factors are present in an organism, for *E. coli* the estimate is as high as 314 representatives (Pérez-Rueda and Collado-Vides 2000). Clr belongs to the Crp subclass, that has been known to be activated by cyclic adenosine monophosphate. Körner et al. 2003 found three distinct clusters of Crp-like proteins in their phylogenetic analysis, of which the well-characterized metabolic regulator CAP from *E. coli* belongs to the main Crp group, whereas Clr can be assigned to subgroup G (Figure 58). The species within this cluster are mainly proteobacteria, but the presence of a few Actinobacteria led to the notion, that they have inherited this protein class via horizontal gene transfer (Soberón-Chávez et al. 2017). Quite a few class G Crp containing families are involved in nitrogen-fixation, e.g., rhizobiaceae, bradyrhizobiaceae and even some burkholderiaceae.



16131236 E. coli	D^{53} S^a, H^b	S^{62} F^b	R^{82} G^a	R^{123} N^c	T^{127} L, C, I^b	S^{128} T^b	D^{138} N, K^b	G^{141} S, Q, K, D^b	R^{142} C^b	A^{144} T, S, Q, Y, L, F, V, C^b	L^{148} K^a, R^b
15964225 S. meliloti	T	R	R	A	T	T	D	A	R	A	L
23954814 Geobacter metallroducens	T	I	D	R	G	D	G	G	G	D	A
23015195 Magnetospirillium magnetotacticum	S	G	R	D	A	D	S	S	R	A	L
15595472 Pseudomonas aeruginosa	S	I	R	H	A	F	P	Q	R	A	L
21449813 Klebsiella pneumoniae	D	S	R	R	T	S	D	G	R	A	L
21960959 Yersinia pestis	D	S	R	N	T	S	D	G	R	A	L
15595849 Pseudomonas aeruginosa	D	S	R	D	T	S	D	G	R	A	L
15642609 Vibrio cholerae	D	S	R	R	T	S	D	G	R	A	L

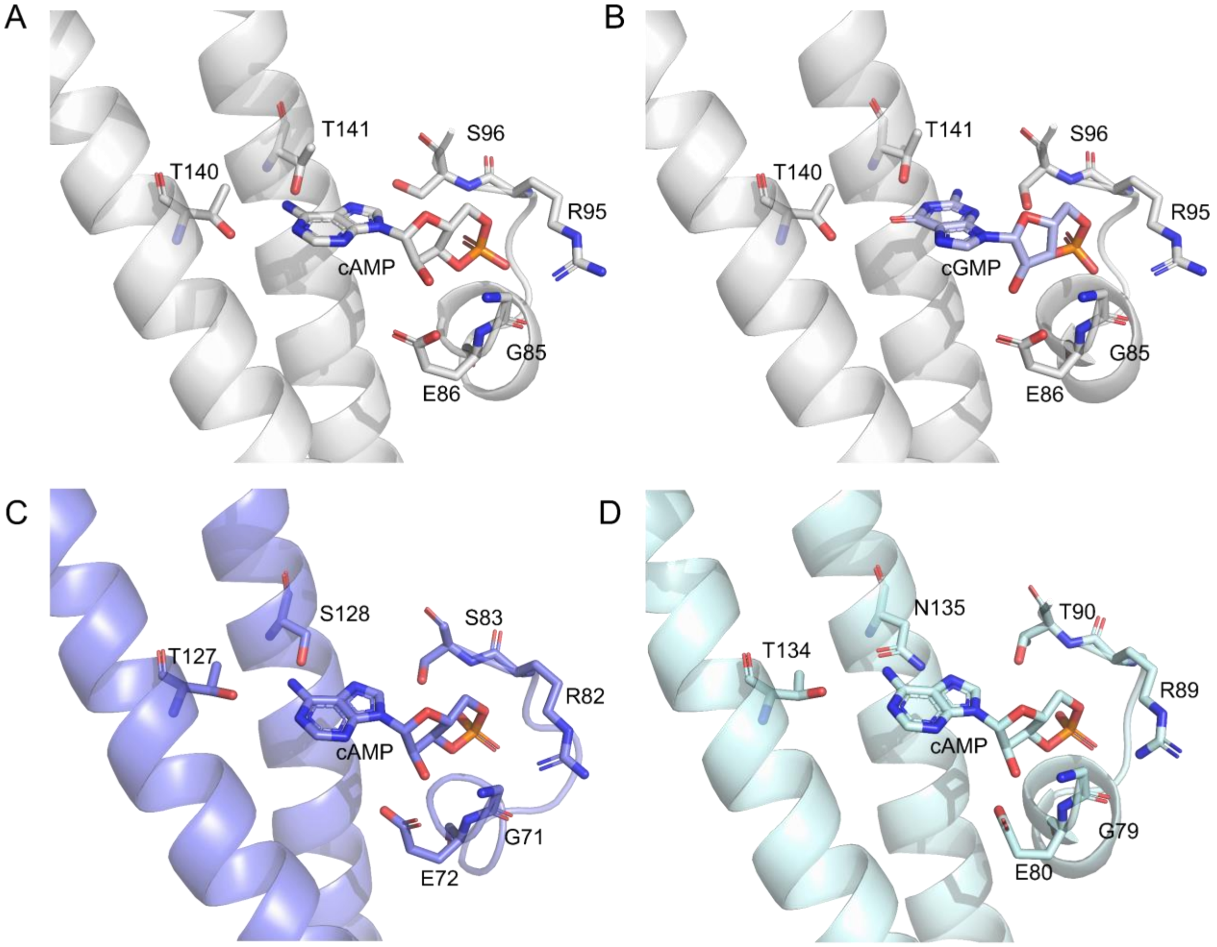
[a] Youn, Koh, Roberts, 2008; [b] Passner, Schultz, Steitz, 2000; [c] Youn, Kerby, Koh, Roberts, 2007.

Figure 58: Sequence analysis of the Crp-Fnr subfamily: (A) Phylogenetic tree of the Crp-Fnr super-family adapted from Körner et al. 2003. The Crp family is further divided into three subclusters, of which only representatives of the main subgroup have been structurally resolved (i.e., *E. coli* CAP, *M. tuberculosis* Rv3676 and *P. aeruginosa* Vfr). Clr is now the first subgroup G structure to be obtained. (B) While their overall appearance and function is quite similar, each Crp-Fnr subgroup acts as sensor to different activating signals. Crp are activated by cAMP only, whereas Clr is able to sense cyclic GMP as well. (C) Mutations in certain positions render *E. coli* CAP cGMP-sensitive. Almost all of those residues are conserved in the main Crp subclass (bottom), while the proteins from subclass G are a lot more diverse (top). As a result, it is unclear if the activation by both cNMP is a unique feature of Clr or a general property of the cluster.

Escherichia coli CAP has been shown to bind cyclic GMP without transitioning to its active form. Different cGMP activated mutants have been found. Generally, the mutation is either located in the cNMP binding site or the perimeter of the *C*-terminal hinge (D53 in β4; R123, T127 and S128 in hinge α5; D138, G141, R142, A144 and L148 in α6). It has been suggested that these mutations stabilize the active closed conformation of Crp, facilitating cGMP-dependent activation. Comparing these residues over several Crp homologs shows that the Crp subgroup is almost fully conserved and the fact, that all proteins, that have been characterized from this group, behaved very similarly, is unsurprising. The sequences in cluster G however, are a lot more diverse, especially around the hinge. The phylogenetic relation, function and sequence conservation of Crp-Fnr class proteins was analysed in Figure 58. Clr is the only cluster G Crp that has been structurally characterized so far and stands out by its bifunctional activation via both 3',5'-cAMP and 3',5'-cGMP. Whether this is a unique feature of the protein or a common trait for G cluster Crp proteins remains to be explored. CgrA from *Rhodospirillum centenum* has been described to be cGMP activated, with very little cAMP activity (Roychowdhury et al. 2015). As it belongs to class G too, the common trait might actually be the cGMP activation, regardless of whether cAMP activity is retained.

5.1.1 The active site and binding mode of Crp-like proteins are highly conserved

The activation of Crp-like proteins has been well studied mainly for *E. coli* CAP, but also *M. tuberculosis* Rv3676 and *C. glutamicum* GlxR. All of them belong to the main Crp subgroup and are activated by cyclic adenosine monophosphate only. The effector binding domains of Rv3676 and GlxR are extremely similar with a sequence identity of 71%, whereas CAP has a rather low identity of 22% and 24% with them. For the hinge region the identity between Rv3676 and GlxR increases to 82% and for CAP to 36% with both. Crp-like proteins are activated by cAMP binding in the effector binding domain by a set of highly conserved residues shown in Figure 59. In fact, the cNMP binding mode of Clr is the same as any of the other structurally resolved Crp-like protein and does not explain the bifunctional behaviour. The 2'-hydroxy group of the cNMP ribose is coordinated by a GE motif, which is present in all homologs. Additionally, the glycine peptide oxygen is within binding distance with the phosphate oxygen. The RS motif in Clr and *E. coli* CAP is mutated to the very similar RT in *M. tuberculosis* Rv3676 and *C. glutamicum* GlxR. The arginine sidechain coordinates on of the phosphate oxygens, whereas the serine or threonine interacts with the other. The peptide nitrogen of the residue always establishes the interaction, but the sidechain is turned away in the cAMP-bound Clr structure. The interaction of the effector molecule with the two residues in the hinge has been found essential for the activation of Crp. In Rv3676 and GlxR the hinge motif is mutated to TN, which allows for an additional interaction of the asparagine with the adenine N7. The ^{140}TT motif in Clr is highly similar to the ^{127}TS one present in CAP, but while it has been theorized for CAP, that cGMP is unable to interact with S128 and therefore fails to activate the transcription factor, the nucleotide is flipped from a *trans* to a *syn* conformation and interact with ^{141}T with two hydrogen bonds. NMR studies conducted with CAP have revealed that while cGMP is able to bind to its active site, it is unable to switch the protein to its active conformation (Popovych et al. 2009). Given that the binding mode in all of these homologs is highly conserved and cGMP is obviously able to activate Clr but not CAP without visible differences in the active site, additional factors likely play a role.


A
T141
S96
T140
R95
cAMP
G85
E86
B
T141
S96
T140
R95
cGMP
G85
E86
C
S128
S83
T127
R82
cAMP
G71
E72
D
N135
T90
T134
R89
cAMP
G79
E80

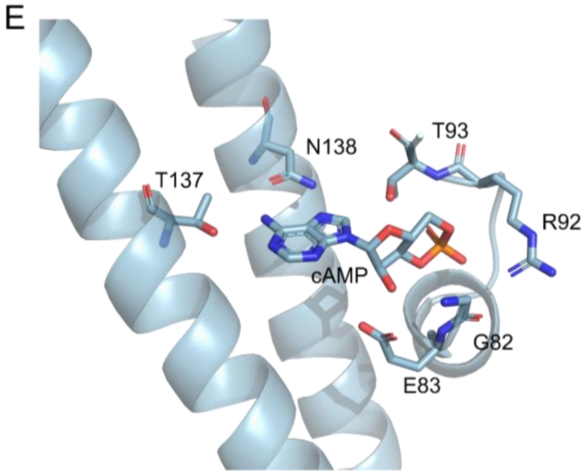


Figure 59: Nucleotide binding mode in different Crp-like proteins. (A) cAMP interacts with Clr via a series of hydrogen bonds, of which only two are actually with the adenine base. T140 and T141 are both within 3.2 Å of the primary amine, the ribose 2'OH is coordinated by G85 and E86. The oxygens of the phosphate moiety establish four to five additional interactions via G85, R96 and S96, depending on whether the ε-nitrogen is turned towards or away. (B) The binding mode for cGMP is identical for the ribose and phosphate backbone, however the guanine is present in a *syn* conformation. This brings the guanine in a more favourable position and allows for an additional bond to T141. T140 is within 3.6 Å of the carbonyl group, while T141 is not within 3.2 and 3.0 Å of the 1'N of the purine and the primary amine. (C) The active site of *E. coli* CAP only deviates in the ^{127}TS instead of a TT motif in the hinge helices. All interactions are preserved. (D) *M. tuberculosis* Rv3676 has an ^{89}RT motif, which is inconsequential for binding, however the hinge residues are T134 and N135. The asparagine reaches further into the active site and its nitrogen allows for the establishment of additional hydrogen bonds to the primary amine and 7'N of the purine. T134 is 3.4, the N135 oxygen is 2.5 and the nitrogen 3.3 Å of the primary amine. The nitrogen also contacts 7'N is a distance of 2.9 Å. (E) The *C. glutamicum* GlxR active site is identical to that of Rv3676 with binding lengths of 2.9, 3.0, 4.1 and 2.9 Å (listed in the same order).

5.1.2 The activation of Crp-Fnr like proteins can occur via direct and indirect interaction with the hinge

Without a doubt, Crp-like proteins are the most well researched Crp-Fnr superfamily representatives. Nevertheless, a few structures from other classes are available and can give an insight in the different modes of activation that have developed in the superfamily. The most similar to Clr are the transcription factors NtcA and CprK, which both interact with small organic molecules as well. NtcA is a sensor of nitrogen scarcity in cyanobacteria and interact with the metabolite 2-oxoglutarate, an intermediate in the Krebs cycle (Llácer et al. 2010). Like in Clr and other Crp the activation requires the interaction of the effector molecule with both hinges. As 2-oxoglutarate is a much smaller molecule than cAMP the active site has adapted slightly to accommodate. T132 is conserved, but too distant to interact directly with the ligand, while E133 is larger than T141 in Clr and hence capable to interact with the ligand. To make up for the lack of a T132 contact, N136 in Clr is here replaced by arginine R128. The sidechain reaches towards the ligand's carboxylate group and establishes two hydrogen bonds. Another difference between the two transcription factors is the location of helix α3. In NtcA α3 is closer positioned to the neighbouring hinge and interacts with S126 via V76. This interaction might further anchor the effector binding domain, making sure the rather small ligand stays close enough for activating interactions.

Some bacteria are able to use halo-organic compounds in a respiratory-type metabolism. Transcription factor CprK binds to *o*-chlorophenol ligands and regulates halorespiratory genes in response (Levy et al. 2008). Similar to NtcA, the effector is a lot smaller than cAMP and cannot interact with the hinge in the same position, which is therefore not conserved here. Instead, CprK has a lysine K133, which similarly to NtcA R128 reaches into the active site, where it interacts with the chlorine. The effector binding in CprK does not involve the neighbouring hinge, as it does in the other two. There is however a stabilising interaction between the hinge and α3 established by G85 and Y130. The binding mode for the different homologs is shown in Figure 60 below.

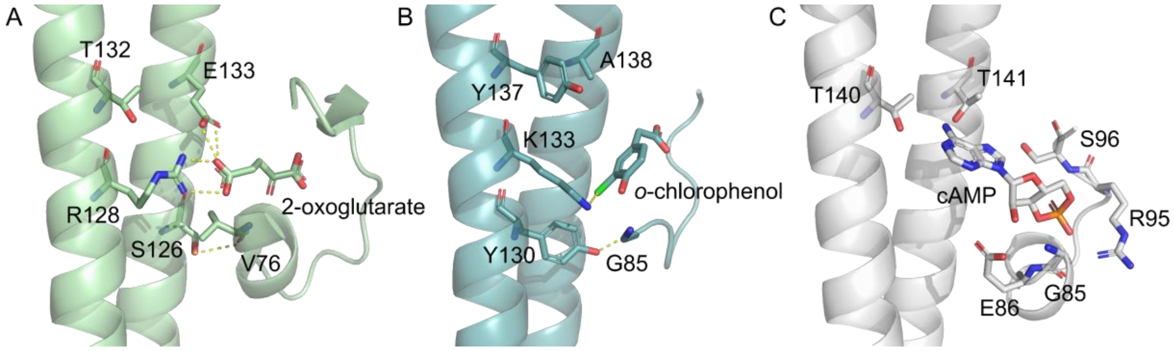


Figure 60: Hinge interaction in Crp-Fnr homologs binding to small organic molecules. (A) NtcA from *Synechococcus elongatus* (PDB-ID: 2xgx) regulates transcription upon binding of 2-oxoglutarate, a signal of nitrogen scarcity in cyanobacteria (Llácer et al. 2010). Instead of interacting with the conserved T132 it contacts R128 in the second hinge. A bond between S126 and V76 stabilises the effector binding domain. (B) The *o*-chlorophenol bound to halorespiratory control factor CprK from *Desulfitobacterium hafniense* (PDB-ID: 3e5x) only interacts with the hinge K133 in the same chain as the binding domain (Levy et al. 2008). The domain is stabilised to the hinge via G85-Y130 interaction. (C) cAMP in Clr interacts further *C*-terminally in the hinge, as it is a larger molecule. Its helix α3 is the further away from the hinge than in NtcA.

Not only small organic molecules are important signals in the cell, but inorganic molecules like oxygen, nitrogen monoxide, carbon monoxide and reactive oxygen species (ROS) are effector molecules for Crp-Fnr-like transcription factors as well. Their mode of activation is highly different for classic Crp-like Clr and illustrated in Figure 61. FixK$_2$ regulates many symbiotic nitrogen fixation genes in *rhizobiaceae*. While it retains the effector binding domain found in Clr, the transcription factor is cAMP-independent and negatively controlled by oxidation of its single cysteine C183 located in helix α7. This might in turn influence the positioning of α8 towards the active closed conformation. The oxidation of C183 is triggered by reactive oxygen species produced in the early stages of root infection, in the process of endosymbiotic respiration and during late nodule senescence, the *N*-terminal domain appears to have lost its function (Bonnet et al. 2013). The carbon monoxide oxidation activator protein CooA senses its effector via heme group bound in the *N*-terminal domain. In the off state the iron is octahedrally coordinated by the heme itself as well as H77 from the same domain and P2 from the neighbouring *N*-terminus. The loss of coordination to Pro2 results in a pentacoordinate heme that can now bind CO, which shifts the conformational equilibrium towards the closed state (Kuchinskas et al. 2006). Another special case is the anaerobic regulator Fnr. It switches from aerobic to anaerobic respiration through the assembly and degradation of its oxygen-sensitive [4Fe-4S] cluster located below the beta barrel at the *N*-terminus of the protein. The hinge region S148 is tightly coordinated by D102, I104 and T105. The activation likely happens over an intramolecular interaction network and is yet to be fully understood (Volbeda et al. 2015).

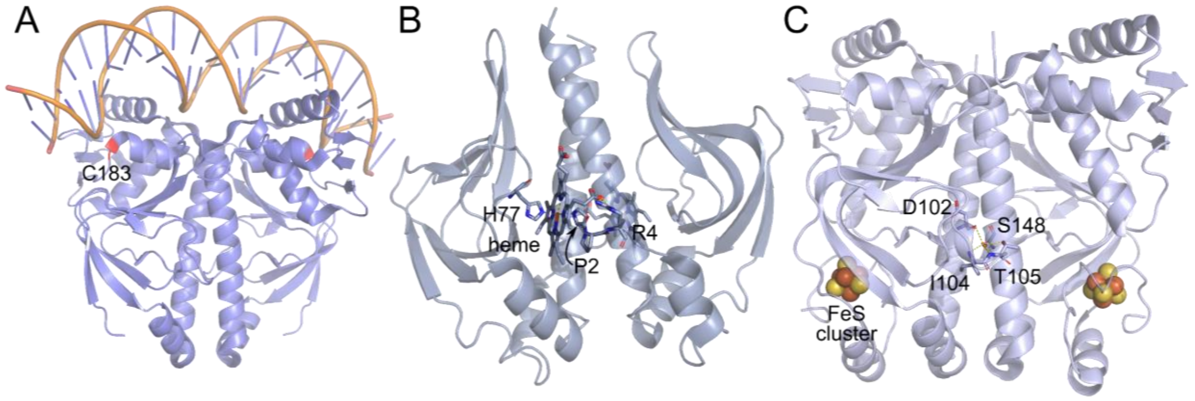


Figure 61: Other Crp-Fnr activation mechanisms. (A) Ligand-independent transcription regulator FixK$_2$ from *Bradyrhizobium japonicum* (PDB-ID: 4i2o) is negatively controlled by oxidation in position C183 caused by reactive oxygen species (Bonnet et al. 2013). (B) CooA from *Rhodospirillum rubrum* (PDB-ID: 4k8f) senses carbon monoxide via a heme group bound in its effector binding domain (Kuchinskas et al. 2006). The signal molecule displaces P2 from the heme coordination sphere and

activates the transcription factor. (C) The anaerobic transcription factor Fnr from *Aliivibrio fischeri* (PDB-ID: 5e44) senses oxygen via degradation of its [4Fe-4S] cluster.

5.1.3 Clr binding behaviour differs from the one found in other Crp

Clr binds cAMP and DNA with low micromolar affinity. The slightly lower affinity for cGMP is only observed in the absence of DNA. Once the binding sequence is present, the protein seems to be conformationally primed to bind cGMP and cAMP at almost equal affinities. The best studied Clr homolog is *E.* coli CAP, which is able to bind both cAMP and cGMP, but is only activated by the former (Popovych et al. 2009). The lack of activation in response to cGMP has also been found for *M. tuberculosis* Rv3676 and has been assumed to be a generic trait of Crp-like proteins so far (Reddy et al. 2009). Apo-Crp-like proteins bind DNA with low affinity and sequence specificity (Harman 2001; Fic et al. 2009). The activation of *E. coli* CAP has been proposed to be a model example of allosteric activation with effector binding far away from the DNA binding domain and the active state achieved via conformational changes throughout the protein, shifting the equilibrium towards its on-state (Harman 2001). As ligand binding is much faster (about 10 times) than the following conformational rearrangement, methods tracking the latter like fluorescence of labelled protein will observe a sequential mechanism (Małecki et al. 2000). ITC directly measures ligand binding and is therefore independent of this. Early crystal structures of CAP contained cAMP in a secondary binding site between the cNMP binding and the HTH domain, sometimes also termed “*syn*-site” in literature as the ligand was found in *syn* conformation. The site is only occupied at millimolar concentrations, which is above the physiological concentrations of up to 10 µM in *E. coli* and the theory of the negative cooperativity of a second binding site was discarded as biologically irrelevant after some debate in the field (Epstein et al. 1975; Fic et al. 2009). No secondary binding site could be observed for Clr. The affinities reported in the literature are in good agreement with the ones found for Clr as shown in Table 11. Some studies reported DNA affinities in the nanomolar range, however all of those were conducted using fluorescence anisotropy and longer 26-mer DNA duplexes (Gunasekara et al. 2015; Youn et al. 2006; Youn et al. 2007; Youn et al. 2008). The affinity might potentially increase for longer flanking regions. While the affinity for cGMP has been reported to be equal to that of cAMP based on NMR measurements, no actual values have been reported to our knowledge (Fic et al. 2009; Reddy et al. 2009).

Table 11: DNA and cAMP affinities for Crp homologs. The values measured for Clr using ITC are well in agreement with ones reported for CAP, the cAMP-independent CAP T127L S128I double mutant CAP*, *M. tuberculosis* Rv3676 and *C. glutamicum* GlxR.

		K_D (μM)	
Clr	cAMP	5.8 ± 1.4	
Clr, DNA	cAMP	6.7 ± 2.7	
DNA	Clr, cAMP	6.0 ± 4.7	
Clr	cGMP	23.8 ± 14.5	
Clr, DNA	cGMP	10.7 ± 3.9	
DNA	Clr, cGMP	4.5 ± 12.4	
CAP	cAMP	9.6 - 27.5	Małecki et al. 2000
CAP	cAMP	10.0	Harman 2001
CAP*	DNA	0.5	Tzeng and Kalodimos 2013
Rv3676	cAMP	17.0	Stapleton et al. 2010
GlxR	cAMP	17.0	Townsend et al. 2014

5.1.4 The transformation from the open to closed form is characterized by decreased flexibility

The crystal structures presented in this work are all in the active closed conformation of the Crp. Crystals for apo Clr could be obtained by Petra Gnau previously, but have yet to be solved. Apart from the same quality issues also observed for the cocrystals, the reason for the difficulties in finding a suitable model for refinement might be caused by the properties of Crp apo structures in general. While the first active structure of *E. coli* CAP was published in 1994 (Schultz et al. 1991), the structure of the unliganded state remained a mystery until 2009 were both a NMR study (Popovych et al. 2009) and the apo wild type crystal structure were released (Sharma et al. 2009). Since then, a few apo structures of Crp homologs have been reported. If all available apo Crp states are superimposed, as shown in Figure 62, and coloured by r.m.s.d., it becomes clear that the structures deviate strongly in their HTH domain. This high degree of flexibility makes finding a suitable molecular replacement model highly complex. If the new apo state is not accidentally in agreement, anomalous scatterers are needed. A feasible option taken by Sharma et al. 2009 is the use of seed crystals from CAP mutant D138L. It has been reported that this mutation stabilizes the inactive conformation of CAP and the mutant should therefore crystallize more readily (Ryu et al. 1993). The drawback is, that while these crystals should be easier to solve, as they likely have a similar conformation, bias also has been introduced and the structure possibly does not represent the protein in solution that well. The first crystal structure of apo CAP contains three dimers per asymmetric unit, all of which slightly deviate from one another (Sharma et al. 2009).

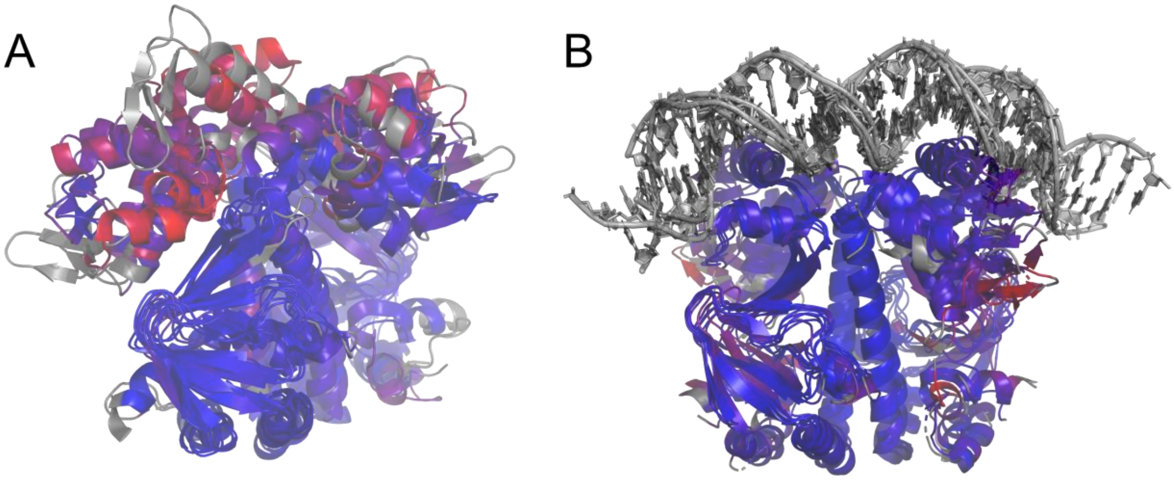


Figure 62: Crp apo and holo states. (A) The superimposition of apo structures of Crp transcription factors coloured by r.m.s.d. shows the high deviation within the DNA binding domain. This indicates that the unligated protein is highly flexible in solution. Included were PDB-ID 3d0s, 4byy, 2wc2, 3hif and 3fwe. (B) A similar analysis for the closed conformation shows that the holo form seems to be universally the same. The analysis contains the proteins also shown in Figure 63.

Apo and holo state of Crp are present in an equilibrium. Even in the apo state, a pre-existing small active population can be detected in NMR studies. If cNMP is added, the nucleotide selectively binds to the active conformation and as a results shifts the population towards the holo state (Gunasekara et al. 2015). The activation starts with the interaction of cNMP to one hinge residue in each helix, T140 and T141 in Clr. In its absence, the two resides disrupt the leucin-zipper in helix 5, keeping the *C*-terminal part unordered and the protein inactive (Youn et al. 2006). This is important as a high level of CRP activity has been found to be lethal to *E. coli* in the same study. The residues found in these positions likely evolved to provide optimal switching ratio between the *on* and *off* state. The interaction of the nucleotide with the hinge binding motif brings the *C*-terminal part of the zipper closer together and triggers a coil-to-helix transformation extending the hinge by three turns (Gunasekara et al. 2015; Popovych et al. 2009). The expulsion of W85 from the active site during activation might also contribute to the coil-to-helix transformation (Popovych et al. 2009). As a result the following helix 6 is shortened (Yu et al. 2012). The two helices are interacting at an angle of about 40° in the holo structure, which has been reported to be optimal for activity bringing the cNMP binding domain and the HTH domain in the correct orientation for DNA binding (Won et al. 2009). The β4-β5 hairpin region interacts with the elongated hinge region in the holo state and is located closer to α7 as a result, bringing the helices 7 and 8 in the correct orientation for DNA interaction. The movement of the hairpin towards the HTH domain is also driven by W85, that is displaced from the active site upon cNMP binding (Popovych et al. 2009).

The dynamic fluctuations of the allosteric network have been investigated using Molecular Dynamics (MD) simulations and Gaussian Network Model (GNM). In MD simulations for the apo and holo state, the latter was much more conformationally stable in the hinge helix. The effector binding also reduced fluctuations in the β4-β5 hairpin and the DNA binding helices 7 and 8. The hairpin movements correlated with α7 in the DNA binding domain in the holo state. In the apo state the cAMP and DNA binding domains appear as quasi-independent dynamic units, while the intra- and inter-subunit correlations increased in the holo state. Particularly L134-D138, corresponding to the *C*-terminal hinge in *E. coli* CAP, were correlated with only the cAMP binding and HTH domain of the same subunit. Binding of the nucleotide changed the correlation to include the neighbouring monomer as well. This is in agreement with the fact, that the nucleotide forms hydrogen bonds with both hinge helices and provides the connection (Aykaç Fas et al. 2013). Not only mutations within the active site influence the activity of Crp, as Youn et al. 2008 found significant activation for CAP D53S, G141K, A144T and L148K with only small increase in cAMP binding. As discussed in 5.1.1 they active site composition cannot explain Clr's unique ability to regulate in response to cGMP. The feature is likely a result of a shift in the apo-holo equilibrium pre-stabilising the latter to be more readily populated. This has been previously described for the CAP* cAMP-independent mutation T127L S128I (Aykaç Fas et al. 2013; Tzeng and Kalodimos 2013). In our structure a previously unknown Cl^- binding site is found between α5 and 6. The ion might stabilise the alignment of the two towards the active site. As the cavity is also present and positively charged in other Crp structures, this feature has either been overlooked or is unique to Clr, in which case it might also be involved in its changed activation behaviour.

The lowered degrees of conformational freedom are also reflected in highly similar holo structures. In Figure 63, the structures of the closed conformation for the CRP homologs *S. meliloti* Clr, *M. tuberculosis* Rv3676, *E. coli* CAP and *C. glutamicum* GlxR, as well as the NtcA 2-oxoglutarate sensor from *S. elongatus* are superimposed and clearly show the agreement in overall conformation. The statistics of the alignment show the contrast of the rather low identity and good r.m.s.d. agreement (Table 12).

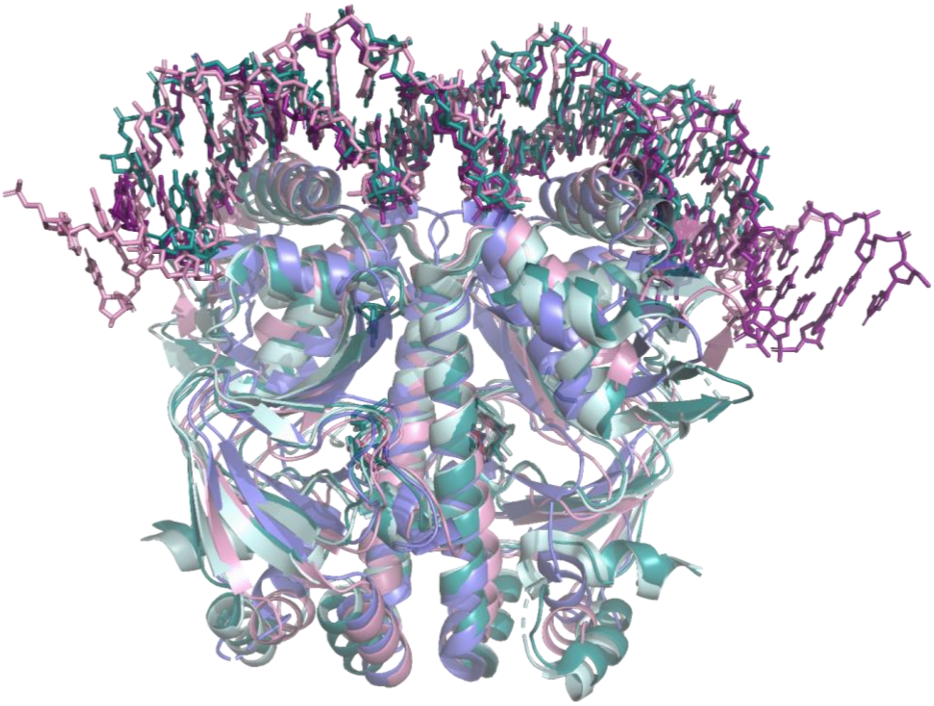


Figure 63: Superimposition of different closed conformations found in Crp-Fnr crystal structures. *S. meliloti* Clr (purple), *M. tuberculosis* Rv3676 (teal, PDB-ID: 3mzh) and *E. coli* CAP (lilac, PDB-ID: 2cgp) were co-crystallized with DNA, while the structures of *C. glutamicum* GlxR (mint, PDB-ID: 4cyd) and *S. elongatus* NtcA (slate, PDB-ID: 2xhk) from the Crp-Fnr superfamily contain only cAMP. The structures superimpose extremely well, given how little sequence similarity they have, reflecting the rigidity of the closed state of these transcription factors.

Table 12: Multiple alignment statistics for Crp-Fnr in a closed conformation. The analysis was generated by superimposing the protein chains in PDBe Fold and show that even though the proteins have a rather low sequence identity, their conformations superimpose well.

R.m.s.d.

	structure	**1**	**2**	**3**	**4**	**5**
1	Clr		1.973	1.901	2.736	3.037
2	Rv3676	1.973		0.655	2.378	2.390
3	GlxR	1.901	0.655		2.216	2.337
4	CAP	2.736	2.378	2.216		2.802
5	NtcA	3.037	2.390	2.337	2.802	

Sequence identity

	structure	**1**	**2**	**3**	**4**	**5**
1	Clr		28.3	28.9	21.7	20.6
2	Rv3676	28.3		80.6	28.3	22.8
3	GlxR	28.9	90.6		27.8	20.0
4	CAP	21.7	28.3	27.8		19.4
5	NtcA	20.6	22.8	20.0	19.4	

5.1.5 Receptor-DNA interaction happens at a conserved motif

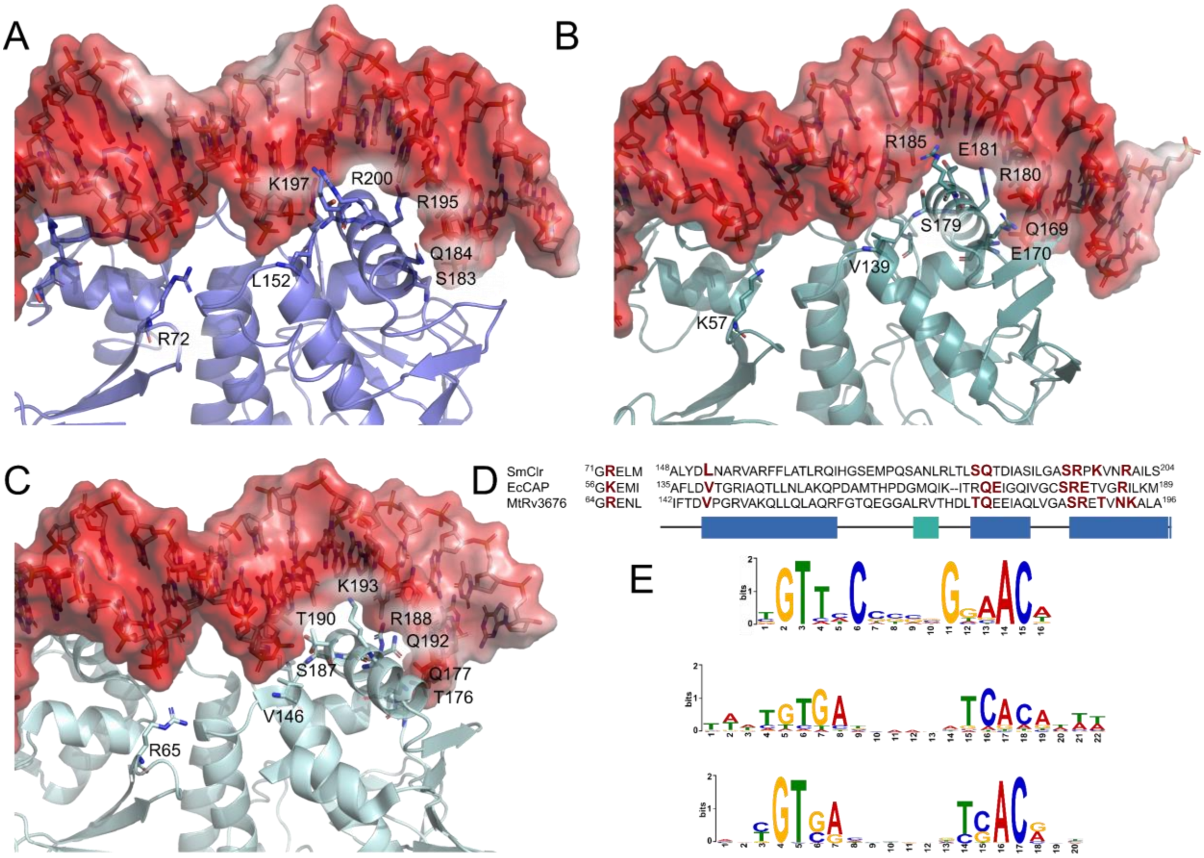
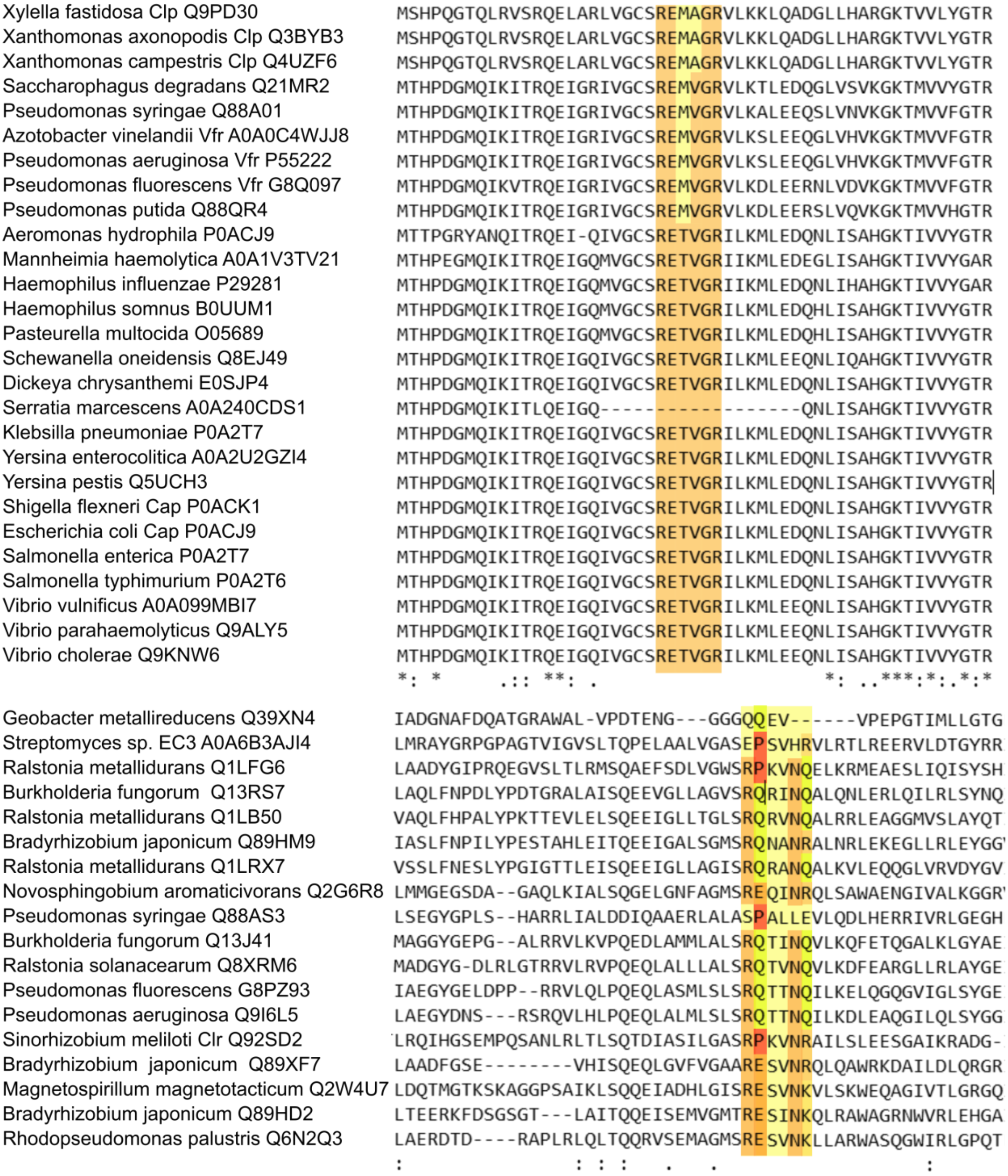

A
K197
R200
R195
Q184
S183
L152
R72
B
R185
E181
R180
S179
Q169
E170
V139
K57
C
K193
T190
R188
Q192
S187
Q177
T176
V146
R65
D
SmClr ^{71}GRELM ^{148}ALYDLNARVARFFLATLRQIHGSEMPQSANLRLTLSQTDIASILGASRPKVNRAILS^{204}
EcCAP ^{56}GKEMI ^{135}AFLDVTGRIAQTLLNLAKQPDAMTHPDGMQIK--ITRQEIGQIVGCSRETVGRILKM^{189}
MtRv3676 ^{64}GRENL ^{142}IFTDVPGRVAKQLLQLAQRFGTQEGGALRVTHDLTQEEIAQLVGASRETVNKALA^{196}
E
bits

Figure 64: DNA binding site interactions in different Crp homologs. (A-E) Crp interact with their conserved binding motif via major groove insertion of helix 8. The interaction is largely conserved only slightly deviating in the resides involved in the binding. (A) shows the binding interactions of Clr, (B) illustrates the ones for *E. coli* CAP (PDB-ID: 2cgp) and (C) displays the recognition in *M. tuberculosis* Rv3676. (D) An alignment of the residues involved in binding shows that while they are located at the same positions in the protein fold, they are not fully conserved. (E) This is also reflected in differences in the binding motif conservation adapted from Bai et al. 2005. It is worth mentioning, that while the three proteins have differences in their ideal binding site, they are also able to interact with imperfect ones to a certain degree.

The DNA binding site of the HTH domain has a positive charge complementing the highly negative phosphate backbone of the DNA. For sequence-specific interactions helix 8 is inserted into the major groove, where it forms multiple hydrogen bonds both with the phosphate backbone and the DNA bases. Figure 64 shows the binding environment of Clr and the two Crp class proteins *E. coli* CAP and *M. tuberculosis* Rv3676. While the overall orientation is very similar, the residues are not conserved. The five-base inverted repeat transcription sites in the three organisms deviate accordingly. Notably, the conserved motif of *E. coli* includes more residues than the others. This indicates a stronger bend of the DNA around the globular protein involving not only the major grooves directly above the dimer constituents, but also the following ones. The relation between motif conservation and binding interface has been studied by Holmquist et al. 2011. The importance of the residues in helix 8 for DNA specificity was illustrated in CAP mutant R180V/G184S. The mutant mimics the Fnr helix assembly and acquires the corresponding specificity for Fnr promotors (Cherfils et al. 1989; Desai et al. 2009).



```
Xylella fastidosa Clp Q9PD30                MSHPQGTQLRVSRQELARLVGCSREMAGRVLKKLQADGLLHARGKTVVLYGTR
Xanthomonas axonopodis Clp Q3BYB3           MSHPQGTQLRVSRQELARLVGCSREMAGRVLKKLQADGLLHARGKTVVLYGTR
Xanthomonas campestris Clp Q4UZF6           MSHPQGTQLRVSRQELARLVGCSREMAGRVLKKLQADGLLHARGKTVVLYGTR
Saccharophagus degradans Q21MR2             MTHPDGMQIKITRQEIGRIVGCSREMVGRVLKTLEDQGLVSVKGKTMVVYGTR
Pseudomonas syringae Q88A01                 MTHPDGMQIKITRQEIGRIVGCSREMVGRVLKALEEQSLVNVKGKTMVVFGTR
Azotobacter vinelandii Vfr A0A0C4WJJ8       MTHPDGMQIKITRQEIGRIVGCSREMVGRVLKSLEEQGLVHVKGKTMVVFGTR
Pseudomonas aeruginosa Vfr P55222           MTHPDGMQIKITRQEIGRIVGCSREMVGRVLKSLEEQGLVHVKGKTMVVFGTR
Pseudomonas fluorescens Vfr G8Q097          MTHPDGMQIKVTRQEIGRIVGCSREMVGRVLKDLEERNLVDVKGKTMVVFGTR
Pseudomonas putida Q88QR4                   MTHPDGMQIKITRQEIGRIVGCSREMVGRVLKDLEERSLVQVKGKTMVVHGTR
Aeromonas hydrophila P0ACJ9                 MTTPGRYANQITRQEI-QIVGCSRETVGRILKMLEDQNLISAHGKTIVVYGTR
Mannheimia haemolytica A0A1V3TV21           MTHPEGMQIKITRQEIGQMVGCSRETVGRIIKMLEDEGLISAHGKTIVVYGAR
Haemophilus influenzae P29281               MTHPDGMQIKITRQEIGQMVGCSRETVGRIIKMLEDQNLIHAHGKTIVVYGAR
Haemophilus somnus B0UUM1                   MTHPDGMQIKITRQEIGQMVGCSRETVGRILKMLEDQHLISAHGKTIVVYGTR
Pasteurella multocida O05689                MTHPDGMQIKITRQEIGQMVGCSRETVGRILKMLEDQHLISAHGKTIVVYGTR
Schewanella oneidensis Q8EJ49               MTHPDGMQIKITRQEIGQIVGCSRETVGRILKMLEEQNLIQAHGKTIVVYGTR
Dickeya chrysanthemi E0SJP4                 MTHPDGMQIKITRQEIGQIVGCSRETVGRILKMLEDQNLISAHGKTIVVYGTR
Serratia marcescens A0A240CDS1              MTHPDGMQIKITLQEIGQ------------------QNLISAHGKTIVVYGTR
Klebsilla pneumoniae P0A2T7                 MTHPDGMQIKITRQEIGQIVGCSRETVGRILKMLEDQNLISAHGKTIVVYGTR
Yersina enterocolitica A0A2U2GZI4           MTHPDGMQIKITRQEIGQIVGCSRETVGRILKMLEDQNLISAHGKTIVVYGTR
Yersina pestis Q5UCH3                       MTHPDGMQIKITRQEIGQIVGCSRETVGRILKMLEDQNLISAHGKTIVVYGTR
Shigella flexneri Cap P0ACK1                MTHPDGMQIKITRQEIGQIVGCSRETVGRILKMLEDQNLISAHGKTIVVYGTR
Escherichia coli Cap P0ACJ9                 MTHPDGMQIKITRQEIGQIVGCSRETVGRILKMLEDQNLISAHGKTIVVYGTR
Salmonella enterica P0A2T7                  MTHPDGMQIKITRQEIGQIVGCSRETVGRILKMLEDQNLISAHGKTIVVYGTR
Salmonella typhimurium P0A2T6               MTHPDGMQIKITRQEIGQIVGCSRETVGRILKMLEDQNLISAHGKTIVVYGTR
Vibrio vulnificus A0A099MBI7                MTHPDGMQIKITRQEIGQIVGCSRETVGRILKMLEEQNLISAHGKTIVVYGTR
Vibrio parahaemolyticus Q9ALY5              MTHPDGMQIKITRQEIGQIVGCSRETVGRILKMLEEQNLISAHGKTIVVYGTR
Vibrio cholerae Q9KNW6                      MTHPDGMQIKITRQEIGQIVGCSRETVGRILKMLEEQNLISAHGKTIVVYGTR
                                            *: *     .:: **: .                    *: ..***:*:.*:*

Geobacter metallireducens Q39XN4            IADGNAFDQATGRAWAL-VPDTENG---GGGQQEV------VPEPGTIMLLGTG
Streptomyces sp. EC3 A0A6B3AJI4             LMRAYGRPGPAGTVIGVSLTQPELAALVGASEPSVHRVLRTLREERVLDTGYRR
Ralstonia metallidurans Q1LFG6              LAADYGIPRQEGVSLTLRMSQAEFSDLVGWSRPKVNQELKRMEAESLIQISYSH
Burkholderia fungorum  Q13RS7               LAQLFNPDLYPDTGRALAISQEEVGLLAGVSRQRINQALQNLERLQILRLSYNQ
Ralstonia metallidurans Q1LB50              VAQLFHPALYPKTTEVLELSQEEIGLLTGLSRQRVNQALRRLEAGGMVSLAYQT
Bradyrhizobium japonicum Q89HM9             IASLFNPILYPESTAHLEITQEEIGALSGMSRQNANRALNRLEKEGLLRLEYGG
Ralstonia metallidurans Q1LRX7              VSSLFNESLYPGIGTTLEISQEEIGLLAGISRQRANQALKVLEQQGLVRVDYGV
Novosphingobium aromaticivorans Q2G6R8      LMMGEGSDA--GAQLKIALSQGELGNFAGMSREQINRQLSAWAENGIVALKGGR
Pseudomonas syringae Q88AS3                 LSEGYGPLS--HARRLIALDDIQAAERLALASPALLEVLQDLHERRIVRLGEGH
Burkholderia fungorum Q13J41                MAGGYGEPG--ALRRVLKVPQEDLAMMLALSRQTINQVLKQFETQGALKLGYAE
Ralstonia solanacearum Q8XRM6               MADGYG-DLRLGTRRVLRVPQEQLALLLALSRQTVNQVLKDFEARGLLRLAYGE
Pseudomonas fluorescens G8PZ93              IAEGYGELDPP--RRVLQLPQEQLASMLSLSRQTTNQILKELQGQGVIGLSYGE
Pseudomonas aeruginosa Q9I6L5               LAEGYDNS---RSRQVLHLPQEQLALMLSLSRQTTNQILKDLEAQGILQLSYGG
Sinorhizobium meliloti Clr Q92SD2           LRQIHGSEMPQSANLRLTLSQTDIASILGASRPKVNRAILSLEESGAIKRADG-
Bradyrhizobium  japonicum  Q89XF7           LAADFGSE--------VHISQEQLGVFVGAARESVNRQLQAWRKDAILDLQRGR
Magnetospirillum magnetotacticum Q2W4U7     LDQTMGTKSKAGGPSAIKLSQQEIADHLGISRESVNKVLSKWEQAGIVTLGRGQ
Bradyrhizobium japonicum Q89HD2             LTEERKFDSGSGT---LAITQQEISEMVGMTRESINKQLRAWAGRNWVRLEHGA
Rhodopseudomonas palustris Q6N2Q3           LAERDTD----RAPLRLQLTQQRVSEMAGMSRESVNKLLARWASQGWIRLGPQT
                                            :               : : :   .   .                  :
```

Figure 65: DNA binding residues differ from Crp class to Crp class G proteins. The residues in helix 8 of Crp are essential for DNA specificity and RNAP interaction and have been mutationally

studied (Hicks et al. 2017). The RETVGR motif found in *E. coli* CAP is largely conserved in the Crp cluster shown on top. The class G subgroup proteins aligned at the bottom show a lot less conservation or other residues are retained in the group.

The significance of helix 8 residues has been studied by mutagenesis. A multiple sequence alignment of Crp-like proteins reveals a highly conserved RE(T/M)(V/A)GR motif. If the same region is aligned in Crp class G proteins, it is apparent that they indeed form a new class including a new largely conserved DNA interaction motif R(Q/E/P)xxN(Q/R/K) shown in Figure 65. Notwithstanding its lack of a sidechain G184 has been identified to be essential for transcription activation activity in *E. coli*. T182, V183 and G184 have not warranted much attention until recently, as they are not pointing towards the consensus DNA or don't have a sidechain, their significance has become apparent in mutagenesis studies. An important feature of Crp-like transcription factors is the DNA binding they elicit. A kink of about 40° has been reported for Cap, which is in agreement with our Clr structure (Schultz et al. 1991). Here the better resolved DNA half is kinked by 40°C as well, while the less resolved other half has a bend of 32°. Crp-dependent promotors were proven to be converted to be Crp-independent by modifying upstream sequences, that introduced an inherent bent to the DNA (Bracco et al. 1989). Interestingly, the DNA kink is juxtaposed with G184 in CAP. The question remains whether the conservation in this position is connected to DNA bending. Mutations in position 184 revealed that not the size but the conformational flexibility in this location is important for the activation (Hicks et al. 2017). As the asparagine found in all class G orthologs has not been investigated, it will be an interesting target for future studies and might shed light on the importance of the position. An analysis of the Crp-Fnr superfamily (higher Cap identity than 25.2%) revealed for position 182 T in 41.8% of all sequences, followed by M (28.6%) and S (15.4%). In position 183 valine was highly conserved with 74.5% and in position 184 the most prevalent residues are G (34.5%), S (28.7%) and N (20.8%) (Hicks et al. 2017). Clr deviates from this in the first position, but contains a VN motif in the latter ones.

5.1.6 The RNA polymerase is recruited by Clr to regulate gene expression

For Crp dependent transcription activation, three different promotor types outlined in Figure 66 A have been described: Class I Crp dependent promotor require only the Crp for transcription activation. The CAP recognition site appears in various distances upstream of the RNAP site, provided both are spatially in line. The most prominent example of a class I system is the lac promotor. Class II promotors are similar; However, the Crp DNA site is overlapping the RNAP site, replacing the -35 element. The best-characterized class II promotor is the galP1 promotor.

An interesting case are class III promotors, that require two or more Crp molecules, but function via a combination of class I and II mechanisms (Busby and Ebright 1999; Epstein et al. 1975). Crp proteins interact with the RNA polymerase via the stacked beta strands within the HTH domain that has been termed AR1 in class I promotors (Hudson et al. 2009) or the top of the β-barrel within the cNMP binding domain, which has been labelled AR2 , in class II promotors (Liu et al. 2017).

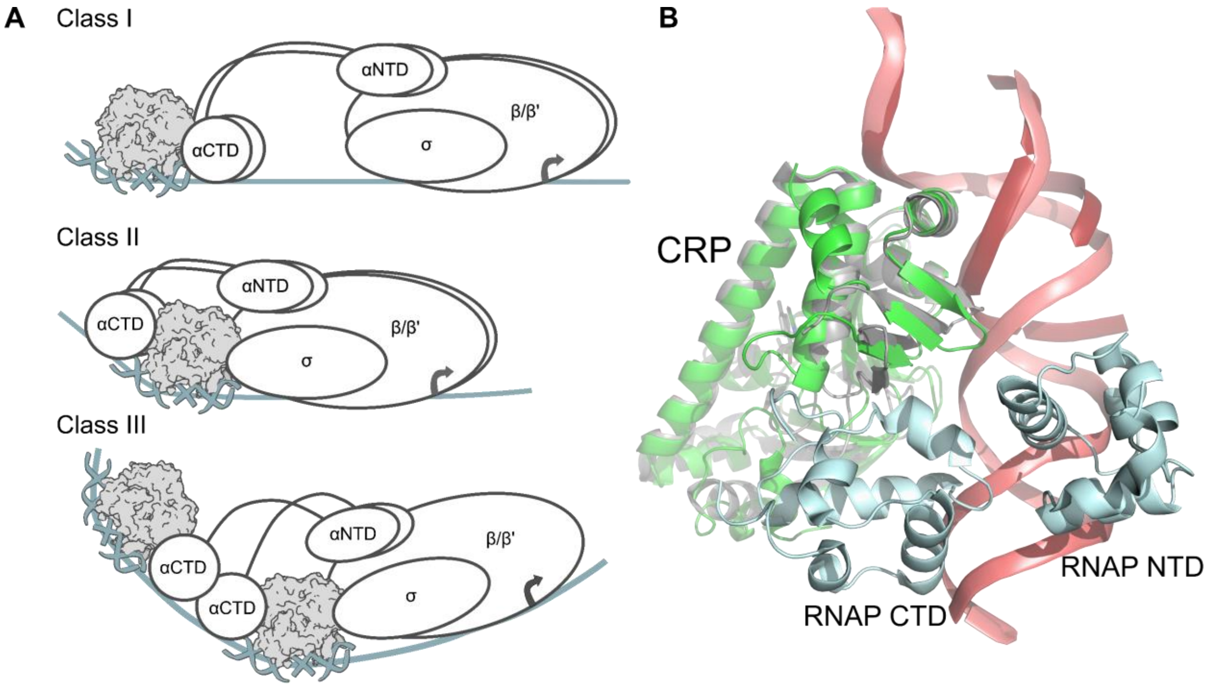


Figure 66: Interaction of DNA bound Clr to recruit the RNA polymerase: (A) Promoter architectures differ in the assembly sequence of transcription activation components. In Class I promoters the Crp interacts with the CTD of the RNAP only, whereas in class II promoters the Crp is enclosed by the RNAP. Class III promoters involve two Crp binding sites. (B) Structural overlay of a DNA-bound Clr monomer with CTD and NTD of the RNA polymerase of the E. coli class II transcription activation complex (PDB-ID: 6PB6). The interface shows slight deviations; However, Neda Farmani was clearly able to show, that Clr is capable of interacting with *E. coli*'s RNAP *in vivo* (publication pending).

For Clr AR1 three of the four β-strands are present, whereas the N-terminal loop of the motif is turned more towards the protein core. Crystal structures of CAP in complex with DNA and the αCTD domain of the RNA polymerase show very little flexibility within the interface and surprisingly only one hydrogen bonding interaction between T158 of CAP and T285 of the αCTD (PDB-ID: 5ciz, 3n4m and 1lb2; Figure 66 B). The interface is hydrophobic, a feature that is retained in Clr. The activating region AR1 has significantly divergent orientation in its

N-terminal part for Clr, which does not keep the protein from complementing the RNAP in *E. coli in vivo*. The overall conservation of the 22 residues comprising AR1 is low with three identical and three similar amino acids. In class II promotors the RNAP interacts with the Crp on two additional sites: AR1 of the second symmetry unit and AR2 of this unit (Busby and Ebright 1999). The interactions with AR2 involve the beta and omega subunit as shown in crystal and cryo-EM structures from *Thermus thermophilus* and *E. coli* (PDB-ID: 5i2d and 6pb6). The interactions within the omega unit are conserved as G47 and D48 (G14 and E15 in *T. thermophilus*), whereas the area involved in the interaction with the beta subunit in *T. thermophilus* (mainly through E8) deviates in a way that allows no deductions on whether it would interact in the same way.

5.2 CpdA might be a Clr antagonist

For cNMP-dependent regulation to be successful, the messenger levels need to be tightly controlled not only by its synthesis by adenylyl/guanylyl cyclases, but also via degradation. The degradation also ensures that the signal has a defined half-life and is performed by phosphodiesterases (PDE). As the transcription factor Clr and the phosphodiesterase CpdA are both located in the CyaD1 locus, it has been theorized that CpdA acts as an antagonist to the cAMP-dependent gene regulation. The hypothesis was abandoned when Mathieu-Demazière et al. found the protein to be inactive towards 3',5'-cyclic nucleotides. On the contrary, were able to show that CpdA is indeed active against both cNMP isomers. Another hint pointing towards the initial hypothesis would be a Clr-dependence in the expression of the CpdA gene. This has not been observed so far in *S. meliloti (Mathieu-Demazière et al. 2013)*, but for the highly similar Crp-CpdA regulation systems in *Vibrio vulnificus* (Kim et al. 2009), *Pseudomonas aeruginosa* (Fuchs et al. 2010) and *C. glutamicum* (Schulte et al. 2017b). It will be interesting to see, whether *S. meliloti* remains the exception to the rule, but its activity certainly complements Clr.

5.2.1 CpdA belongs to the novel class III of phosphodiesterases

The hydrolysis of phosphate is an important biological function and three distinct protein classes developed independently to fulfil that task. Class I and II phosphodiesterases are present in higher and lower eukaryotes respectively. Class III is found in bacteria, archaea and eukaryotes and was defined by Richter 2002 by its flagship member CpdA from *E.coli*. Class III phosphodiesterases are still a rather unexplored protein class with only one structure, Rv0805 from *M. tuberculosis*, available. The class belongs to the calcineurin-like phosphoesterase superfamily and is closely related to purple acid phosphatases (PAP) and orthologs of the Mre11 DNA-phosphodiesterase. The phylogeny of the superfamily was analysed in Figure 67. The closest ortholog to the CpdA family are YfcE-like proteins. They are manganese-dependent and have been structurally resolved, however their natural substrate is unknown, as they only show activity towards bis-*p*-nitrophenyl phosphate, a general substrate for phosphoesterases, but neither phosphodiesters or monoesters (Miller et al. 2007). Mre11-like proteins are exonucleases involved in end-processing during DNA double-strand break repair (Zha et al. 2009), whereas phosphoprotein proteases (PPP) dephosphorylate phosphoserine and threonine residues (Matange et al. 2015b). CpdB-like proteins were initially thought to complement CpdA, as they are 2',3'-cyclic nucleotide-specific. Since the CpdA homolog Rv0805 and *S.*

meliloti CpdA presented in this thesis are rather promiscuous in their choice of substrate, this might only be true in some organisms (Matange et al. 2015b). Purple acid phosphatases (PAP) catalyse the hydrolysis of a wide range of phosphomonoester and amide substrates. The mammalian orthologs are considered a potential target for anti-osteoporosis drugs, but there are also plant and some bacterial representatives. PAPs possess hydrolase and peroxidase properties (Schenk et al. 2013).

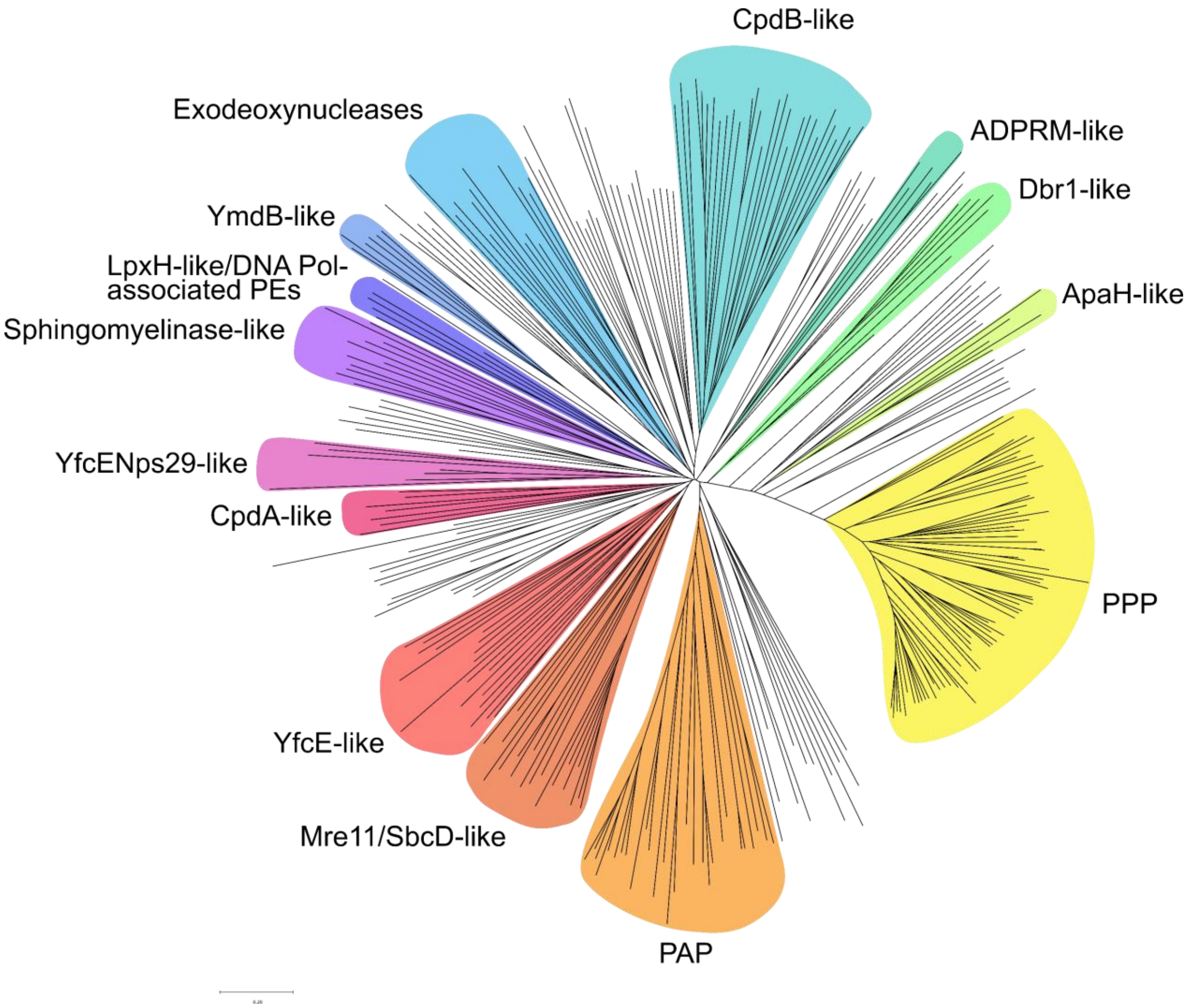

CpdB-like
ADPRM-like
Dbr1-like
ApaH-like
PPP
PAP
Mre11/SbcD-like
YfcE-like
CpdA-like
YfcENps29-like
Sphingomyelinase-like
LpxH-like/DNA Pol-associated PEs
YmdB-like
Exodeoxynucleases

Figure 67: Phylogenetic analysis of the calcineurin-like phosphoesterase family. A protein seed containing 324 sequences from the calcineurin-like phosphoesterase superfamily (Pfam: Pf00149) was aligned and analysed using the neighbour-joining algorithm in Mega-X. Clusters were manually assigned using RCSB blast and UniProt annotations when possible.

Quite a few protein classes from the calcineurin-like superfamily have been structurally resolved. Figure 68 shows a superimposition of orthologs from the Mre11, purple acid phosphatase, YfcE and CpdA class and clearly illustrates the diversity in their architecture – so much so that alignment and superimposition algorithms fail to properly position them. If the active sites are superimposed, the conservation of the DxH-D-N-H-HxH binding motif becomes apparent. The only protein deviating from this is the PAP with its DxY-D-N-H-HxH binding motif, in which the tyrosine results in its purple colour. While the active sites of calcineurin-like phosphoesterases are highly similar, their substrate preferences differ and seem to be influenced by residues in the proximity. This makes it highly difficult to predict the specificity even within a class. As the class III of phosphodiesterases is rather new and unexplored and only one representative has been structurally resolved, our studies on *Sm*CpdA are a valuable contribution to understanding how class III phosphodiesterases function.

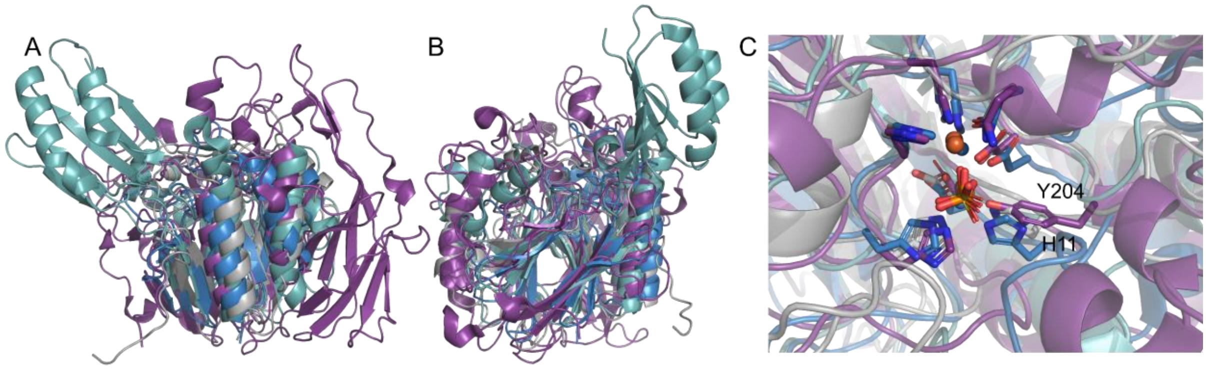


Figure 68: Comparison of Mre11, PAP, YfcE and CpdA homologs on a structural level. (A, B) The fold of proteins in the calcineurin superfamily is merely somewhat conserved in the protein core. The outer parts are very diverse. The crystal structures of Mre11 from *Pyrococcus furiosus* (teal, PDB-ID: 1ii7), the purple acid phosphatase from *Triticum aestivum* (purple, PDB-ID: 6gj2), YfcE from *E. coli* (blue, PDB-ID: 1su1 and CpdA from *S. meliloti* (grey) were superimposed by their active site residues. Their fold is too different to get satisfactory alignment via other algorithms. (A) and (B) are related to each other via 90°C rotation. (C) The active site assembly of all orthologs is nearly identical. The only binding residue not conserved is H11 (CpdA), which is Y204 in the PAP. This difference results in the purple colour in this protein group.

5.2.2 *Sinorhizobium meliloti* contains 14 phosphodiesterases of varying function

In *M. tuberculosis* Rv0805 has been described to be the only cNMP phosphodiesterase. Its promiscuity towards different substrates has been explained by the lack of other orthologs (Shenoy et al. 2007). We have found *Sm*CpdA to be multifunctional as well, even though it lacks the monoesterase activity of Rv0805. An analysis of the genome of *Sinorhizobium meliloti* strain 1021 by the Pfam annotation of its gene products reveals 14 metallophosphoesterases (Pf00149, Pf13277, Pf12850) listed in Table 13. As the group annotation in UniProt seemed questionable at times, the classes were newly assigned via Blast analysis. Three CpdA and one CpdB-like protein were found in the organism. Of these three phosphodiesterases only *Sm*CpdA is found in a cyclase containing locus, the CyaD1 locus. This might suggest, that the other three are housekeeping hydrolases keeping cNMP levels stable under normal conditions, whereas *Sm*CpdA functions specifically in response to Clr.

Table 13: Metallophosphoesterases in *S. meliloti*. Proteins were identified in the *S. meliloti* strain 1021 proteome by Pfam superfamily annotation and manually assigned using Blastp. In two cases marked with a * the assignment was conducted using HHblits. The analysis reveals that the bacterium codes for three CpdA-like proteins in addition to one CpdB ortholog.

gene	UniProt-ID:	length	Pfam-ID	Assignment
*Sm*c01909	Q92QQ0	430	Pf00149	exonuclease SbcD
*Sm*b20152	Q92X10	572	Pf00149	CpdA-like
*Sm*c03166	Q92LW6	361	Pf00149	closest to PAP*
*Sm*c02236	Q92S83	242	Pf00149	endonuclease PdeM
*Sm*c04018	Q926C1	628	Pf00149, Pf01476, Pf02872	ApaH-like 5'-hydrolase UshA
*Sm*b21213	Q92V37	219	Pf00149	phosphoprotein phosphatase
*Sm*c04449	Q926C9	657	Pf00149, Pf02872	CpdB-like
*Sm*b20081	Q92X81	236	Pf00149	closest to LpxH*
*Sm*c03964	Q92M93	300	Pf00149	LpxH-like protein Yael
*Sm*c02179	Q92SD5	268	Pf00149	*Sm*CpdA
*Sm*c02712	Q92K16	300	Pf00149	CpdA-like
*Sm*c00171	Q92PD2	281	Pf00149	LpxH-like
*Sm*c01798	Q92QS0	246	Pf12850	YfcE-like
*Sm*c03973	Q92M85	265	Pf13277	YmdB-like

5.2.3 CpdA is a metal-dependent cNMP phosphodiesterase

Contrary to earlier reports by Mathieu-Demazière et al., we were surprised to find that *Sm*CpdA actually is not metal-independent, but shares the same active site geometry found in *Mt*Rv0805 and *Klebsiella aerogenes* GpdQ. Two Fe^{2+} ions are coordinated by the conserved binding residues as shown in Xray fluorescence and the activity significantly decreases upon EDTA treatment and can be restored to different degrees by metal addition. The addition of Mn^{2+} provided the highest activity and specificity with a higher substrate diversity than initially described. The cause for the lack of activity in the initial study probably lies in a misfolded protein incapable of metal binding, as even metal addition did not restore the function in their case.

The observations made in this thesis also mean, that the general features present in all calcineurin-like phosphoesterases can be found in *Sm*CpdA as well. The active site consisting of seven conserved residues contains two divalent cations crucial for the binding of the catalytically important hydroxide molecule and anchoring of the substrate via its phosphate moiety (Matange 2015a). The metals M1 and M2 (or α and β) can be either homo- or heteronuclear and their occupancy seems to vary as well. In GpdQ the α site has a higher occupancy, whereas the β site is occupied equally only in the presence of substrate (Jackson et al. 2008). In *Sm*CpdA we found the metal occupancies to be in the same range with 0.8-0.9 for both. The metal occupancy and identity can alter the catalytic function, but does not impact the fold itself. For Rv0805 a metal dependence of dimerization has been observed, where the addition of Fe^{3+} or Mn^{2+} during cell growth shifted equilibrium towards the dimeric species (Shenoy et al. 2007). For *Sm*CpdA no metal was added to the LB medium and only a monomeric peak was observed. The impact of metalation has been nicely illustrated by Keppetipola and Shuman 2008, where they found that metal exchange in Pnkp from *Clostridium thermocellum* altered both specificity and affinity of the phosphoesterase for different substrates. Similarly, we find that the product specificity of *Sm*CpdA is the greatest in the presence of Mn^{2+}. Whether the resulting active species is a homonuclear Mn^{2+}-Mn^{2+} or a heteronuclear Fe^{2+}-Mn^{2+} is yet to be determined. The metals found in some of the CpdA and calcineurin-like orthologs are listed in Table 14. The dependence of substrate specificity and metal might also mean that phosphoesterases might function as metal sensors in some cases (Matange et al. 2015b). The affinity of class III PDE to their metals is rather high, as 10 mM EDTA was unable to fully strip Rv0805 of its cofactors (Shenoy et al. 2005), which is in line with our observations. The incorporation of metals during expression and the incomplete removal by chelators has at times led to the interpretation that there are metal independent CpdA-like PDE (Zheng et al. 2013), but those two factors are a much more likely explanation.

Table 14: Metal centres in a selection of calcineurin-like phosphoesterases. Entries marked with (a) were extracted from Matange et al. 2015b, entries marked with (b) refer to Keppetipola and Shuman 2008 and (c) to Schulte et al. 2017b. Values presented with the reference (d) are derived from Kimura et al. 2009 and (e) from Zheng et al. 2013. If only a single cation instead of a cation combination is listed, the data was gathered in activity assays and the question whether the protein utilizes a homo- or heteronuclear centre is unresolved.

enzyme	organism	substrates	catalytic metals (α-β)	class	ref.
CpdA	*S. meliloti*	cAMP, cGMP	**Mn^{2+}**, Co^{2+}, Ni^{2+}, Fe^{2+}, Mg^{2+} (Fe3^{+}-Me^{2+} or Me^{2+}-Me^{2+})	class III PDE	
Rv0805	*M. tuberculosis*	cGMP, 2’,3’-cAMP	**Fe^{3+}-Mn^{2+}**	class III PDE	A
GpdQ	*K. aerogenes*	*p*NPP	**Fe^{3+}-Fe^{2+}, Zn^{2+}-Zn^{2+}**	GPDQ	A
CpdA	*Arthrobacter sp.*	cAMP, 3’,5’-cGMP	**Fe^{3+}-Mn^{2+}, Fe^{3+}-Co^{2+}**	class III PDE	E
CpdA	*C. glutamicum*	3’,5’-cAMP, 3’,5’-cGMP	**Zn^{2+}-Zn^{2+}**	class III PDE	C
PdeA	*M. xanthus*	cAMP	**Mn^{2+}**, Co^{2+}, Ca^{2+}, Mg^{2+} (Fe3^{+}-Me^{2+} or Me^{2+}-Me^{2+})	class III PDE	D
PdeB	*M. xanthus*	cAMP	**Mn^{2+}**, Co^{2+}, Ca^{2+}, Mg^{2+} (Fe3^{+}-Me^{2+} or Me^{2+}-Me^{2+})	class III PDE	D
Pnkp	*C. thermocellum*	2’,3’-cAMP, 2’/3’-AMP	**Ni^{2+}-Mn^{2+}**	λ-phosphatase	B
Mre11	*H. sapiens*	ssDNA, dsDNA	**Mn^{2+}-Mn^{2+}**	Mre11-like	A
UshA	*E. coli*	ATP, ADP, AMP, UDP-glucose	**Mn^{2+}-Mn^{2+}**, Co^{2+}	5’-Nucleotidase	A

To achieve correct metalation, it has been suggested that bacteria might alter local metal concentrations (Waldron and Robinson 2009). The localisation of the respective phosphoesterase is very important in this context. CpdA has been found to be intracellular in both *E. coli* and *S. meliloti* (Matange et al. 2015b; Mathieu-Demazière et al. 2013), whereas Rv0805 is located between the cell envelope and cytosol (Podobnik et al. 2009).

5.2.4 Class III phosphodiesterases are functionally diverse

For both the calcineurin-like proteins in general and class III phosphodiesterases that have been functionally characterized so far, no clear rules on the substrate specificity have been observed. Rather, class III PDE seem to be a multifunctional tool that adapts to the organisms needs with few subtle changes. While our structure of *Sm*CpdA aligns perfectly with the active site of the only other phosphate bound structure, that of *M. tuberculosis* Rv0805 shown in Figure 69 A, the activity is not quite the same. Rv0805 is highly 2’,3’-NMP specific and produces 3’-AMP as the main product, while *Sm*CpdA is most active against 3’,5’-cAMP and also displays an inverse stereoselectivity in its products. An overview of all characterized class III PDE is given in Table 15. The ambiguity towards multiple substrates might be facilitated by the dynamic behaviour of class III phosphodiesterases. Jackson et al. conducted molecular dynamics simulations for GpdQ, the glycerophosphodiesterase from *K. aerogenes* and found a breathing motion of the structural motifs surrounding the active site. This was also reflected in increased B-factors in these more flexible regions (Jackson et al. 2007). Figure 69 B shows the structure of phosphate-bound *Sm*CpdA coloured by B-factors and highlights that the dynamic behaviour might be similar.

Table 15: Substrate specificity of all biochemically characterized class III PDE. While all of the listed enzymes are CpdA orthologs and are capable of hydrolyzing 3’,5’-cAMP, they are also active against other monophosphates. In Rv0805 the activity for 2’,3’-cAMP surpasses that for 3’,5’-cAMP by a factor of 150 (Keppetipola and Shuman 2008). *Sm*CpdA is the only known class III PDE that mainly hydrolyses 3’,5’-cAMP to 3’-AMP (91.9%) and 2’,3’-cAMP to 2’-AMP (94.8%).

Enzyme	**organism**	**PDE activity**	**Pase activity**	**3’,5’-cNMP product**	**2’,3’-cNMP product**	**ref**
CpdA	*S. meliloti*	3’,5’-cAMP > 2’,3’-cGMP > 2’,3’-cAMP >> 3’,5’-cGMP	no	mostly 3’-AMP	mostly 2’-AMP	
Rv0805	*M. tuberculosis*	2’,3’-cAMP ≈ 2’,3’-cGMP >> 3’,5’-cAMP ≈ 3’,5’-cGMP	no	mostly 5’-AMP	mostly 3’-AMP	A
CpdA	*E. coli*	3’,5’-cAMP	no	5’-AMP	N.D.	B
CpdA	*Arthrobacter sp.*	3’,5’-cGMP > 3’,5’-cAMP	N.D.	N.D.	N.D.	C
CpdA	*P. aeruginosa*	3’,5’-cAMP (other N.D.)	N.D.	5’-AMP	N.D.	D
PdeA	*M. xanthus*	2’,3’-cAMP ≈ 3’,5’-cAMP, no cGMP hydrolysis	yes	5’-AMP	3’-AMP	E

PdeB	*M. xanthus*	2’,3’-cAMP ≈ 3’,5’-cAMP, no cGMP hydrolysis	yes	5’-AMP	3’-AMP	E
CpdA	*C. glutamicum*	3’,5’-cAMP ≈ 3’,5’-cGMP		5’-AMP	N.D.	F

(a) Podobnik et al. 2009, Keppetipola and Shuman 2008; (b) Matange 2015a ; (c) Zheng et al. 2013; (d) Fuchs et al. 2010; (e) Kimura et al. 2009; (f) Schulte et al. 2017b

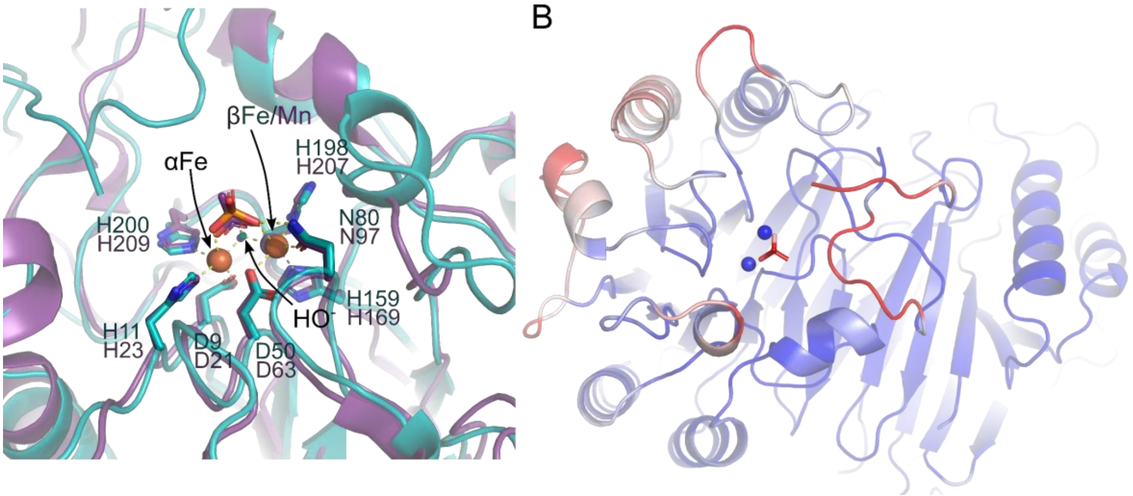


Figure 69: The active site of *Sm*CpdA is influenced not only by the conserved residues. (A) Superimposition of *Sm*CpdA (teal) with the only other phosphate-bound class III PDE structure available, *Mt*Rv0805 (PDB-ID: 3ib7, violet). Rv0805 has a heteronuclear Fe^{3+}-Mn^{2+} cofactor assembly, which might be identical to native *Sm*CpdA, as it is most active in the presence of Mn^{2+} as well. In the crystal form we obtained, both metals are iron. While the residues coordinating the phosphate are perfectly superimposed, Rv0805 differs in substrate specificity from *Sm*CpdA, which means additional factors in the proximity of the active site need to play a role. (B) Colouring the structure of *Sm*CpdA by B-factor ranging from 20 (blue) to 40 and above (red) reveals increased values surrounding the active site. This has been previously observed on a molecular dynamics simulation by Jackson et al. 2007 and been attributed to a breathing motion of the active site cleft potentially allowing the pocket to adjust for substrate binding. This dynamic behaviour seems to be conserved in *Sm*CpdA.

None of our attempts of soaking the 3’,5’-cAMP derivatives Sp- and Rp-cAMPS into *Sm*CpdA crystals were successful. Figure 70 shows the potential orientations of the different substrates. For 3’,5’-cAMP (C), both poses resulting in the favoured 3’-AMP, deduced from the phosphate orientation, result in clashes with parts of the protein (S232 as well as G267 from the adjacent molecule for pose A and H140 for pose B). Even in the orientation suitable for 5’-AMP formation the 2’OH of the ribose is too close to the H11 with a distance of 1.7 Å. This supports our earlier notion, that our crystal form is incompatible with 3’,5’-cAMP binding (see 4.3.5.). Aside from this, the simple docking also illustrates a few interesting points. The transition state of *Sm*CpdA seems to involve a pose, where the adenine moiety is pointing straight out from the binding pocket (A) and (C), rather than being tilted toward the surface in poses (B) and (D), which are consistent with the formation of 3’- and 5’-AMP respectively. H81 coordinates the scissile oxygen in all cases and has been found to be essential to 3’,5’-cAMP activity by Podobnik et al. 2009. Keppetipola and Shuman 2008 even proved, that a mutation of the corresponding C74 in the inactive YfcE protein from *E. coli* to a histidine, rendered it active towards 2’,3’-cAMP. Another residue found to be essential for PDE activity in Rv0805 is H140 located in the short alpha helical stretch close to pose B (H130 in *Sm*CpdA) (Shenoy et al. 2007). Whether this is the case for *Sm*CpdA as well, can only be fully elucidated once a substrate-bound structure is achieved. The inverted stereospecificity makes it seem unlikely.

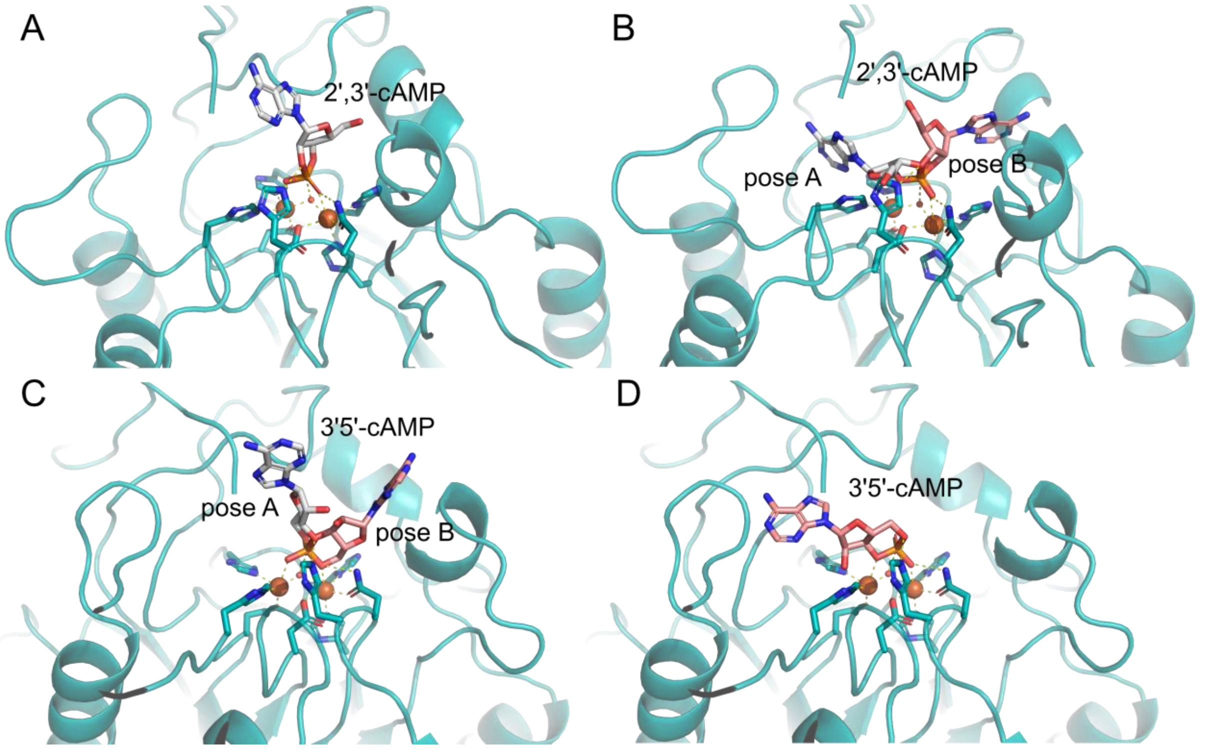


Figure 70: Possible substrate binding modes for *Sm*CpdA based on the phosphate orientation. (A) In the 2’,3’-cAMP-orientation consistent with 2’-AMP formation the adenine and ribose moieties point out from the binding pocket. (B) The poses A and B consistent with 3’-AMP formation on the other hand are tilted more towards the protein surface and result in significant clashes. (C) For the formation of 3’-AMP from 3’,5’-cAMP to possible poses can be generated. Pose A superimposes very well with the one found for 2’,3’-cAMP (A) and is similarly pointed outwards. This might be the actual trajectory. Pose B results in a lot more overlap with the protein fold and is therefore less likely. (D) The orientation of 3’,5’-cAMP resulting in 5’-AMP brings the 2’OH too close to H11.

The active site of *S*mCpdA shows additional density between its metals consistent with a hydroxide ion. This is the first time this catalytically important component has been clearly visible in a class III phosphodiesterase structure and has implications for the mechanistic discussion pertaining calcineurin-like phosphoesterases, that has been ongoing for some time.

All metallophosphoesterases contain a binuclear metal centre, that has a hydroxide coordinated in between. Whether this is also the nucleophile attacking the phosphate might differ from one ortholog to another. For GpdQ from *K. aerogenes* and Mre11 from *Pyrococcus furiosus* an additional metal-bound hydroxide has been suggested to play this role (Hadler et al. 2008; Hopfner et al. 2001). However, no compelling experimental evidence was found so far that these proteins follow the mechanism highlighted in Figure 71 B. A notable exception is the 5'-hydrolase UshA from *E. coli*, that was crystallized in an adenosine- and phosphate-bound complex by Knöfel and Sträter 2001 (PDB-ID: 1ho5). Here the alpha-phosphate is bound to an additional water or hydroxide (μ-hydroxide in literature) and the phosphate is coordinated by this. As there are other structures without the μ-hydroxide, the authors suggested that both scenarios are operational in UshA, although the former is preferred. In class III PDE, *Sm*CpdA is the first structure in which the bridging hydroxide is unambiguously resolved. The phosphate is coordinated by both metals with equal bond lengths, so it is likely the mechanism follows Figure 71 A.

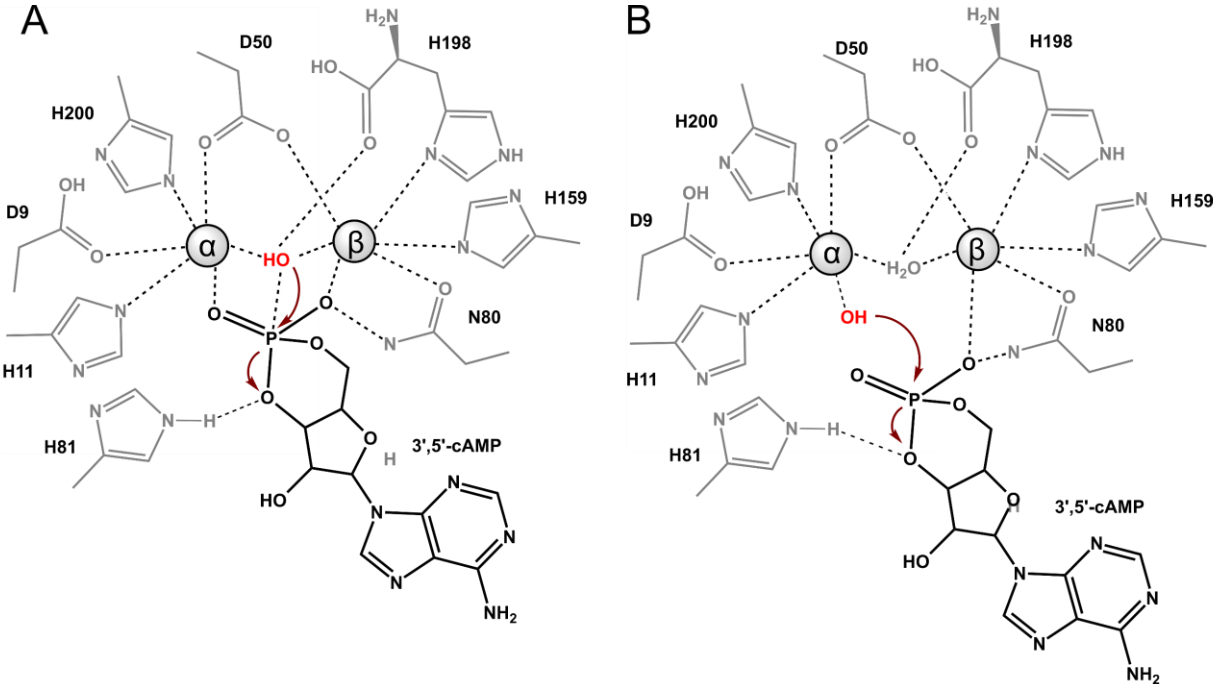


Figure 71: Scheme showing the proposed mechanisms of class III PDE in the context of SmCpdA. (A) The crystal structure reported in this thesis supports a mechanism assuming a bridging nucleophile. Both metals are octahedrally coordinated. The coordination sphere for metal α consists of D9, H11, D50 and H200 as well as the bridging hydroxide and a phosphate oxygen, while metal β interacts with D50, N80, H159, H198 and the ion as well as another phosphate oxygen. The hydroxide ion is held in place by the metals and the peptide oxygen of H198 in a perfect trajectory for a S_N2-like attack on the phosphorus, which then leads to the dissolution of the opposite bond. H81 coordinates the leaving

oxygen and acts as hydrogen donor. The residue has been found to be essential in mutagenesis studies. (B) An alternate mechanism has been proposed involving an additional μ-hydroxide bound to metal α. While this mechanism has structural evidence in other members of the calcineurin metallophosphatase superfamily like UshA from *E. coli*, there is no evidence pointing towards it for class III PDE to date. For both mechanisms the final phosphate release is going to be the rate-limiting step (Matange et al. 2015b).

6 Outlook

In the present work, the structural background of the cAMP-mediated feedback loop highlighted in Figure 72 was analysed, mainly focussing on the transcription factor Clr and the phosphodiesterase CpdA. Bioinformatic analysis of the CyaD1 locus revealed first indications, that the remaining genes *Sm*c02177 and *Sm*c02178 might code for a FecR-like iron transport regulator and a Sel1-like repeat signal transducing protein. Based on findings in *E. coli* that FecR and CHASE2 domains might interact, these two proteins remain very important targets to possibly decoding the elusive plant signal that activates CyaD1. As the recognition might involve iron, complementing *in vivo* studies involving iron availability would be promising. For CyaD1 Ankan Banerjee already had previously obtained a crystal structure of the cyclase homology domain in its inactive monomeric form. The generation of the active dimeric form, potentially by fusing the domain to a dimerization domain, will reveal the substrate specificity, i.e., whether it is an adenylate or guanylate cyclase or even both. For the regulative CHASE2 domain lingering solubility issues during purification point to it being a peripheral membrane protein. Sequence covariation analysis suggest two *N*- and *C*-terminal alpha-helical stretches with an unordered core, but a full structural and functional characterization is still lacking.

With the crystal structure and binding affinity measurements for Clr, the first bifunctional Crp-like transcription factor was characterized. It belongs to the class G Crp, that seem to have distinct features from classical Crp like *E. coli* CAP. The only other biochemical analysis of a class G Crp involved CgrA from *Rhodospirillum centenum*, which is cGMP-activated as well. No further structures aside of that of Clr are known for the group. Mapping all the common features of this class is going to be a very promising new field. For Clr in particular, Petra Gnau from our group was able to obtain first apo crystals and improving them potentially by seeding might help understand the on-off equilibrium of the transcription factor and its activation by cGMP better. *S. meliloti* codes for eight Crp-like proteins in total (*Sm*a0662, *Sm*a1067, *Sm*a1141, *Sm*a1207, *Sm*a1948, *Sm*b21079, *Sm*c01954 and *Sm*c02175 (Clr)). How these transcription factors interplay and how they differ from each other also still remains to be determined.

With CpdA we report only the second class III phosphodiesterase structure. While its activity in the presence of metals was thoroughly investigated, the question whether *Sm*CpdA contains a homonuclear Mn^{2+}-Mn^{2+} core or a heteronuclear Fe^{3+}-Mn^{2+} core in its native form remains to be elucidated. Its activity against 3’,5’-cAMP and 3’,5’-cGMP means it potentially complements Clr, however the kinetics of the hydrolysis remain to be quantified. On the *in vivo* side, finding possible regulative connections between CpdA and Clr would further support their pairing as antagonists. M. *tuberculosis* Rv0805, the first class III PDE to be structurally investigated is almost exclusively 2',3'-cNMP specific. The subtle differences leading to a different specificity in *Sm*CpdA can be further elucidated as well, potentially involving substrate-bound co-crystallisation. The results will also provide insight in the mechanism and the potential active site movement, we found indications for in our B-factor analysis.

Clr promotor assays conducted in the Becker group revealed a plethora of genes that are downstream regulated by Clr, most of them uncharacterized. A lot of work remains to be done here, until we can understand how the signal cascade ultimately prevents overpopulation of nodules in plant symbiosis.

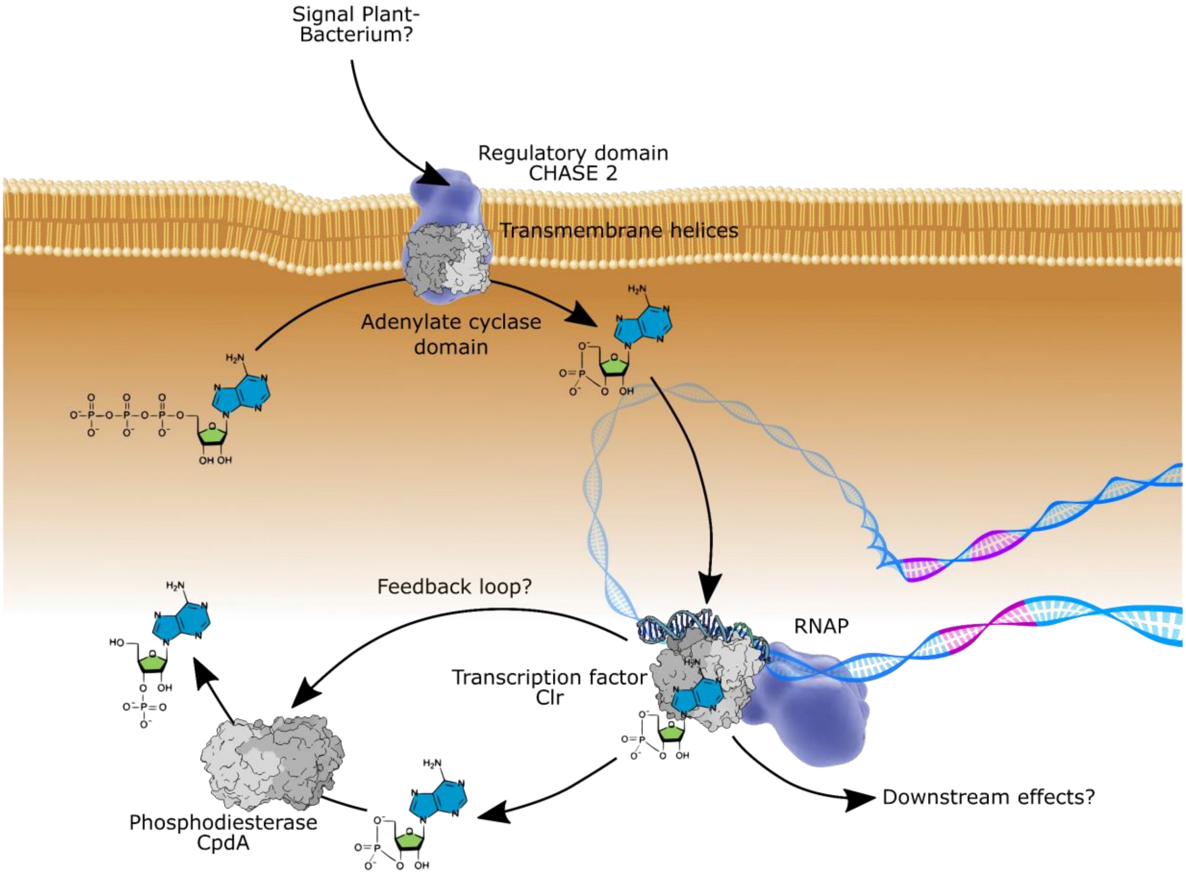

Signal Plant-Bacterium?
Regulatory domain CHASE 2
Transmembrane helices
Adenylate cyclase domain
Feedback loop?
RNAP
Transcription factor Clr
Phosphodiesterase CpdA
Downstream effects?

Figure 72: Many aspects of the cAMP-mediated transcription regulation in *S. meliloti* have now become clearer. While the structure of the CHD domain of CyaD1 has been solved, both the activating signal sensed by the extracellular CHASE2 domain and whether the cyclase is adenylyl or guanylyl-specific remain unknown. Clr and CpdA are both members of rather new protein groups and their characterisation in this work can be helpful in understanding these novel classes. Their interplay and the downstream effects of Clr-dependent regulation are interesting targets for future studies helping us understand the signal cascade better.

7 Bibliography

Abis, Giancarlo; Pacheco-Gómez, Raúl; Bui, Tam T. T.; Conte, Maria R. (2019): Isothermal Titration Calorimetry Enables Rapid Characterization of Enzyme Kinetics and Inhibition for the Human Soluble Epoxide Hydrolase. In *Anal. Chem.* 91 (23), pp. 14865–14872. DOI: 10.1021/acs.analchem.9b01847.

Agarwal, Nisheeth; Raghunand, Tirumalai R.; Bishai, William R. (2006): Regulation of the expression of whiB1 in Mycobacterium tuberculosis: role of cAMP receptor protein. In *Microbiology (Reading, England)* 152 (Pt 9), pp. 2749–2756. DOI: 10.1099/mic.0.28924-0.

Akhter, Yusuf; Yellaboina, Sailu; Farhana, Aisha; Ranjan, Akash; Ahmed, Niyaz; Hasnain, Seyed E. (2008): Genome scale portrait of cAMP-receptor protein (CRP) regulons in mycobacteria points to their role in pathogenesis. In *Gene* 407 (1-2), pp. 148–158. DOI: 10.1016/j.gene.2007.10.017.

Alberto, Mario; Barba, Jeannette; H., Angel (2012): Mycobacterium Tuberculosis Signaling via c-AMP. In Pere-Joan Cardona (Ed.): Lipid Surrounding of Mycobacteria: Lethal and Resuscitating Effects: INTECH Open Access Publisher.

Anraku, Y. (1964): A new cyclic phosphodiesterase having a 3'-nucleotidase activity from Escherichia coli B. In *The Journal of biological chemistry* 239, pp. 3412–3419.

Aravind, L.; Anantharaman, Vivek; Balaji, Santhanam; Babu, M. Mohan; Iyer, Lakshminarayan M. (2005): The many faces of the helix-turn-helix domain: Transcription regulation and beyond. In *FEMS Microbiol Rev* 29 (2), pp. 231–262. DOI: 10.1016/j.fmrre.2004.12.008.

Aykaç Fas, Burcu; Tutar, Yusuf; Haliloğlu, Türkan (2013): Dynamic fluctuations provide the basis of a conformational switch mechanism in apo cyclic AMP receptor protein. In *PLoS computational biology* 9 (7), e1003141. DOI: 10.1371/journal.pcbi.1003141.

Bai, Guangchun; McCue, Lee Ann; McDonough, Kathleen A. (2005): Characterization of Mycobacterium tuberculosis Rv3676 (CRPMt), a cyclic AMP receptor protein-like DNA binding protein. In *Journal of bacteriology* 187 (22), pp. 7795–7804. DOI: 10.1128/JB.187.22.7795-7804.2005.

Baneyx, F. (1999): Recombinant protein expression in Escherichia coli. In *Current Opinion in Biotechnology* 10 (5), pp. 411–421. DOI: 10.1016/S0958-1669(99)00003-8.

Beavo, Joseph A.; Brunton, Laurence L. (2002): Cyclic nucleotide research -- still expanding after half a century. In *Nature reviews. Molecular cell biology* 3 (9), pp. 710–718. DOI: 10.1038/nrm911.

Birnboim, H. C.; Doly, J. (1979): A rapid alkaline extraction procedure for screening recombinant plasmid DNA. In *Nucleic acids research* 7 (6), pp. 1513–1523. DOI: 10.1093/nar/7.6.1513.

Blanton, Robert M. (2020): cGMP Signaling and Modulation in Heart Failure. In *Journal of cardiovascular pharmacology* 75 (5), pp. 385–398. DOI: 10.1097/FJC.0000000000000749.

Block, Helena; Maertens, Barbara; Spriestersbach, Anne; Brinker, Nicole; Kubicek, Jan; Fabis, Roland et al. (2009): Chapter 27 Immobilized-Metal Affinity Chromatography (IMAC). In Richard R. Burgess, Murray P. Deutscher (Eds.): Guide to protein purification, vol. 463. 2. ed. San Diego, Calif.: Elsevier Acad. Press (Methods in Enzymology, 463), pp. 439–473.

Blokesch, Melanie (2012): Chitin colonization, chitin degradation and chitin-induced natural competence of Vibrio cholerae are subject to catabolite repression. In *Environmental microbiology* 14 (8), pp. 1898–1912. DOI: 10.1111/j.1462-2920.2011.02689.x.

Bodero, Maria D.; Munson, George P. (2009): Cyclic AMP receptor protein-dependent repression of heat-labile enterotoxin. In *Infection and Immunity* 77 (2), pp. 791–798. DOI: 10.1128/IAI.00928-08.

Boivin, Stephane; Kozak, Sandra; Meijers, Rob (2013): Optimization of protein purification and characterization using Thermofluor screens. In *Protein expression and purification* 91 (2), pp. 192–206. DOI: 10.1016/j.pep.2013.08.002.

Bonnet, Mariette; Kurz, Mareike; Mesa, Socorro; Briand, Christophe; Hennecke, Hauke; Grütter, Markus G. (2013): The structure of Bradyrhizobium japonicum transcription factor FixK2 unveils sites of DNA binding and oxidation. In *The Journal of biological chemistry* 288 (20), pp. 14238–14246. DOI: 10.1074/jbc.M113.465484.

Bracco, L.; Kotlarz, D.; Kolb, A.; Diekmann, S.; Buc, H. (1989): Synthetic curved DNA sequences can act as transcriptional activators in Escherichia coli. In *The EMBO Journal* 8 (13), pp. 4289–4296. DOI: 10.1002/j.1460-2075.1989.tb08615.x.

Bradford, Marion M. (1976): A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. In *Analytical biochemistry* 72 (1-2), pp. 248–254. DOI: 10.1016/0003-2697(76)90527-3.

Brochet, Mathieu (2018): cGMP Signalling: Malarial Guanylyl Cyclase Leads the Way. In *Current biology : CB* 28 (17), R939-R941. DOI: 10.1016/j.cub.2018.07.059.

Brochet, Mathieu; Balestra, Aurélia C.; Brusini, Lorenzo (2020): cGMP homeostasis in malaria parasites-The key to perceiving and integrating environmental changes during transmission to the mosquito. In *Molecular microbiology*. DOI: 10.1111/mmi.14633.

Brown, Kerene A.; Wilson, Derek J. (2017): Bottom-up hydrogen deuterium exchange mass spectrometry: data analysis and interpretation. In *The Analyst* 142 (16), pp. 2874–2886. DOI: 10.1039/c7an00662d.

Burnette, W.Neal (1981): "Western Blotting": Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. In *Analytical biochemistry* 112 (2), pp. 195–203. DOI: 10.1016/0003-2697(81)90281-5.

Busby, S.; Ebright, R. H. (1999): Transcription activation by catabolite activator protein (CAP). In *Journal of Molecular Biology* 293 (2), pp. 199–213. DOI: 10.1006/jmbi.1999.3161.

Callahan, S. M.; Cornell, N. W.; Dunlap, P. V. (1995): Purification and properties of periplasmic 3':5'-cyclic nucleotide phosphodiesterase. A novel zinc-containing enzyme from the marine symbiotic bacterium Vibrio fischeri. In *The Journal of biological chemistry* 270 (29), pp. 17627–17632. DOI: 10.1074/jbc.270.29.17627.

Chayen, Naomi E. (2004): Turning protein crystallisation from an art into a science. In *Current opinion in structural biology* 14 (5), pp. 577–583. DOI: 10.1016/j.sbi.2004.08.002.

Chen, Xi; Zhan, Chang-Guo (2004): Fundamental Reaction Pathways and Free-Energy Barriers for Ester Hydrolysis of Intracellular Second-Messenger 3‘,5‘-Cyclic Nucleotide. In *J. Phys. Chem. A* 108 (17), pp. 3789–3797. DOI: 10.1021/jp0371635.

Cherfils, J.; Gibrat, J. F.; Levin, J.; Batut, J.; Kahn, D. (1989): Model-building of Fnr and FixK DNA-binding domains suggests a basis for specific DNA recognition. In *J. Mol. Recognit.* 2 (3), pp. 114–121. DOI: 10.1002/jmr.300020303.

Compton, Steve J.; Jones, Clive G. (1985): Mechanism of dye response and interference in the Bradford protein assay. In *Analytical biochemistry* 151 (2), pp. 369–374. DOI: 10.1016/0003-2697(85)90190-3.

Congdon, R. W.; Muth, G. W.; Splittgerber, A. G. (1993): The binding interaction of Coomassie blue with proteins. In *Analytical biochemistry* 213 (2), pp. 407–413. DOI: 10.1006/abio.1993.1439.

Cornish-Bowden, A. (1985): Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. In *Nucleic acids research* 13 (9), pp. 3021–3030. DOI: 10.1093/nar/13.9.3021.

Croda-García, Gerardo; Grosso-Becerra, Victoria; Gonzalez-Valdez, Abigail; Servín-González, Luis; Soberón-Chávez, Gloria (2011): Transcriptional regulation of Pseudomonas aeruginosa rhlR: role of the CRP orthologue Vfr (virulence factor regulator) and quorum-sensing regulators LasR and RhlR. In *Microbiology (Reading, England)* 157 (Pt 9), pp. 2545–2555. DOI: 10.1099/mic.0.050161-0.

Crowther, Gregory J.; He, Panqing; Rodenbough, Philip P.; Thomas, Andrew P.; Kovzun, Kuzma V.; Leibly, David J. et al. (2010): Use of thermal melt curves to assess the quality of enzyme preparations. In *Analytical biochemistry* 399 (2), pp. 268–275. DOI: 10.1016/j.ab.2009.12.018.

Cukkemane, Abhishek; Baldus, Marc (2013): Characterization of a cyclic nucleotide-activated K(+) channel and its lipid environment by using solid-state NMR spectroscopy. In *Chembiochem : a European journal of chemical biology* 14 (14), pp. 1789–1798. DOI: 10.1002/cbic.201300182.

Dai, Jiayin; Lin, Shwu-Hwa; Kemmis, Carly; Chin, Anita J.; Lee, J. Ching (2004): Interplay between site-specific mutations and cyclic nucleotides in modulating DNA recognition by Escherichia coli cyclic AMP receptor protein. In *Biochemistry* 43 (28), pp. 8901–8910. DOI: 10.1021/bi0499359.

Dai, Kun; Peng, Xiaoqiang; Zhuang, Wei; Yang, Pengpeng; Jiao, Pengfei; Wu, Jinglan; Ying, Hanjie (2020): Mass transfer process and separation mechanism of four 5'-ribonucleotides on a strong acid cation exchange resin. In *Journal of chromatography. A* 1634, p. 461681. DOI: 10.1016/j.chroma.2020.461681.

Dale, Glenn E.; Oefner, Christian; D’Arcy, Allan (2003): The protein as a variable in protein crystallization. In *Journal of Structural Biology* 142 (1), pp. 88–97. DOI: 10.1016/S1047-8477(03)00041-8.

Dass, Bob Kennedy M.; Sharma, Ritu; Shenoy, Avinash R.; Mattoo, Rohini; Visweswariah, Sandhya S. (2008): Cyclic AMP in mycobacteria: characterization and functional role of the

Rv1647 ortholog in Mycobacterium smegmatis. In *Journal of bacteriology* 190 (11), pp. 3824–3834. DOI: 10.1128/JB.00138-08.

Dauter, Miroslawa; Dauter, Zbigniew (2017): Many Ways to Derivatize Macromolecules and Their Crystals for Phasing. In *Methods in molecular biology (Clifton, N.J.)* 1607, pp. 349–356. DOI: 10.1007/978-1-4939-7000-1_14.

Dauter, Zbigniew (2017): Collection of X-Ray Diffraction Data from Macromolecular Crystals. In *Methods in molecular biology (Clifton, N.J.)* 1607, pp. 165–184. DOI: 10.1007/978-1-4939-7000-1_7.

DelProposto, James; Majmudar, Chinmay Y.; Smith, Janet L.; Brown, William Clay (2009): Mocr: a novel fusion tag for enhancing solubility that is compatible with structural biology applications. In *Protein expression and purification* 63 (1), pp. 40–49. DOI: 10.1016/j.pep.2008.08.011.

Desai, Tasha A.; Rodionov, Dmitry A.; Gelfand, Mikhail S.; Alm, Eric J.; Rao, Christopher V. (2009): Engineering transcription factors with novel DNA-binding specificity using comparative genomics. In *Nucleic acids research* 37 (8), pp. 2493–2503. DOI: 10.1093/nar/gkp079.

Dessau, Moshe A.; Modis, Yorgo (2011): Protein crystallization for X-ray crystallography. In *Journal of visualized experiments : JoVE* (47). DOI: 10.3791/2285.

Drenth, Jan (1999): Principles of Protein X-ray Crystallography. Second Edition. New York, NY: Springer (Springer Advanced Texts in Chemistry).

Duong-Ly, Krisna C.; Gabelli, Sandra B. (2014): Gel filtration chromatography (size exclusion chromatography) of proteins. In *Methods in enzymology* 541, pp. 105–114. DOI: 10.1016/B978-0-12-420119-4.00009-4.

Dupeux, Florine; Röwer, Martin; Seroul, Gael; Blot, Delphine; Márquez, José A. (2011): A thermal stability assay can help to estimate the crystallization likelihood of biological samples. In *Acta crystallographica. Section D, Biological crystallography* 67 (Pt 11), pp. 915–919. DOI: 10.1107/S0907444911036225.

Eckstein, F.; Romaniuk, P. J.; Heideman, W.; Storm, D. R. (1981): Stereochemistry of the mammalian adenylate cyclase reaction. In *The Journal of biological chemistry* 256 (17), pp. 9118–9120.

Emsley, Paul; Cowtan, Kevin (2004): Coot: model-building tools for molecular graphics. In *Acta crystallographica. Section D, Biological crystallography* 60 (Pt 12 Pt 1), pp. 2126–2132. DOI: 10.1107/S0907444904019158.

Epstein, W.; Rothman-Denes, L. B.; Hesse, J. (1975): Adenosine 3':5'-cyclic monophosphate as mediator of catabolite repression in Escherichia coli. In *Proceedings of the National Academy of Sciences of the United States of America* 72 (6), pp. 2300–2304. DOI: 10.1073/pnas.72.6.2300.

Fermani, Simona; Vettraino, Chiara; Bonacini, Irene; Marcaccio, Massimo; Falini, Giuseppe; Gavira, Josè A.; Garcia Ruiz, Juan M. (2013): Heterogeneous Crystallization of Proteins: Is it a Prenucleation Clusters Mediated Process? In *Crystal Growth & Design* 13 (7), pp. 3110–3115. DOI: 10.1021/cg400501f.

Fic, E.; Bonarek, P.; Gorecki, A.; Kedracka-Krok, S.; Mikolajczak, J.; Polit, A. et al. (2009): cAMP receptor protein from escherichia coli as a model of signal transduction in proteins--a review. In *Journal of molecular microbiology and biotechnology* 17 (1), pp. 1–11. DOI: 10.1159/000178014.

Findlay, W. A.; Redfield, R. J. (2009): Coevolution of DNA uptake sequences and bacterial proteomes. In *Genome biology and evolution* 1, pp. 45–55. DOI: 10.1093/gbe/evp005.

Finn, Robert D.; Attwood, Teresa K.; Babbitt, Patricia C.; Bateman, Alex; Bork, Peer; Bridge, Alan J. et al. (2017): InterPro in 2017-beyond protein family and domain annotations. In *Nucleic acids research* 45 (D1), D190-D199. DOI: 10.1093/nar/gkw1107.

Fong, Jiunn C. N.; Yildiz, Fitnat H. (2008): Interplay between cyclic AMP-cyclic AMP receptor protein and cyclic di-GMP signaling in Vibrio cholerae biofilm formation. In *Journal of bacteriology* 190 (20), pp. 6646–6659. DOI: 10.1128/JB.00466-08.

Frangioni, J. V.; Neel, B. G. (1993): Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins. In *Analytical biochemistry* 210 (1), pp. 179–187. DOI: 10.1006/abio.1993.1170.

Frendorf, Pernille Ott; Lauritsen, Ida; Sekowska, Agnieszka; Danchin, Antoine; Nørholm, Morten H. H. (2019): Mutations in the Global Transcription Factor CRP/CAP: Insights from Experimental Evolution and Deep Sequencing. In *Computational and structural biotechnology journal* 17, pp. 730–736. DOI: 10.1016/j.csbj.2019.05.009.

Frishman, D.; Argos, P. (1995): Knowledge-based protein secondary structure assignment. In *Proteins* 23 (4), pp. 566–579. DOI: 10.1002/prot.340230412.

Fuchs, Erin L.; Brutinel, Evan D.; Klem, Erich R.; Fehr, Anthony R.; Yahr, Timothy L.; Wolfgang, Matthew C. (2010): In vitro and in vivo characterization of the Pseudomonas aeruginosa cyclic AMP (cAMP) phosphodiesterase CpdA, required for cAMP homeostasis and virulence factor regulation. In *Journal of bacteriology* 192 (11), pp. 2779–2790. DOI: 10.1128/JB.00168-10.

Gaastra, W.; Hansen, K. (1985): Ligation of DNA with t(4) DNA ligase. In *Methods in molecular biology (Clifton, N.J.)* 2, pp. 225–230. DOI: 10.1385/0-89603-064-4:225.

Gamas, Pascal; Brault, Mathias; Jardinaud, Marie-Françoise; Frugier, Florian (2017): Cytokinins in Symbiotic Nodulation: When, Where, What For? In *Trends in plant science* 22 (9), pp. 792–802. DOI: 10.1016/j.tplants.2017.06.012.

Garnerone, Anne-Marie; Sorroche, Fernando; Zou, Lan; Mathieu-Demazière, Céline; Tian, Chang Fu; Masson-Boivin, Catherine; Batut, Jacques (2018): NsrA, a Predicted β-Barrel Outer Membrane Protein Involved in Plant Signal Perception and the Control of Secondary Infection in Sinorhizobium meliloti. In *Journal of bacteriology* 200 (11). DOI: 10.1128/JB.00019-18.

Gavira, José A. (2016): Current trends in protein crystallization. In *Archives of biochemistry and biophysics* 602, pp. 3–11. DOI: 10.1016/j.abb.2015.12.010.

Gibson, A. H.; Harper, J. E. (1985): Nitrate Effect on Nodulation of Soybean by Bradyrhizobium japonicum1. In *Crop Sci.* 25 (3), pp. 497–501. DOI: 10.2135/cropsci1985.0011183X002500030015x.

Gijsbers, Abril; Nishigaki, Takuya; Sánchez-Puig, Nuria (2016): Fluorescence Anisotropy as a Tool to Study Protein-protein Interactions. In *Journal of visualized experiments : JoVE* (116). DOI: 10.3791/54640.

Goeddel, David V. (Ed.) (1990): Gene expression technology: Academic Press (Methods in Enzymology).

Gomelsky, Mark (2011): cAMP, c-di-GMP, c-di-AMP and now cGMP: bacteria use them all! In *Molecular microbiology* 79 (3), pp. 562–565. DOI: 10.1111/j.1365-2958.2010.07514.x.

Gomelsky, Mark; Galperin, Michael Y. (2013): Bacterial second messengers, cGMP and c-di-GMP, in a quest for regulatory dominance. In *The EMBO Journal* 32 (18), pp. 2421–2423. DOI: 10.1038/emboj.2013.193.

Gourion, Benjamin; Berrabah, Fathi; Ratet, Pascal; Stacey, Gary (2015): Rhizobium-legume symbioses: the crucial role of plant immunity. In *Trends in plant science* 20 (3), pp. 186–194. DOI: 10.1016/j.tplants.2014.11.008.

Graham, Peter H.; Vance, Carroll P. (2003): Legumes: importance and constraints to greater use. In *Plant physiology* 131 (3), pp. 872–877. DOI: 10.1104/pp.017004.

Green, Jeffrey; Scott, Colin; Guest, John R. (2001): Functional versatility in the CRP-FNR superfamily of transcription factors: FNR and FLP. In *Advances in microbial physiology* 44, pp. 1–34. DOI: 10.1016/s0065-2911(01)44010-0.

Gunasekara, Sanjiva M.; Hicks, Matt N.; Park, Jin; Brooks, Cory L.; Serate, Jose; Saunders, Cameron V. et al. (2015): Directed evolution of the Escherichia coli cAMP receptor protein at the cAMP pocket. In *The Journal of biological chemistry* 290 (44), pp. 26587–26596. DOI: 10.1074/jbc.M115.678474.

Hadler, Kieran S.; Tanifum, Eric A.; Yip, Sylvia Hsu-Chen; Mitić, Natasa; Guddat, Luke W.; Jackson, Colin J. et al. (2008): Substrate-promoted formation of a catalytically competent binuclear center and regulation of reactivity in a glycerophosphodiesterase from Enterobacter aerogenes. In *Journal of the American chemical society* 130 (43), pp. 14129–14138. DOI: 10.1021/ja803346w.

Hampton Research (Ed.) (2020): Seeding. Crystal growth 101. Available online at https://hamptonresearch.com/uploads/cg_pdf/CG101_Seeding_2020.pdf.

Harman, James G. (2001): Allosteric regulation of the cAMP receptor protein. In *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* 1547 (1), pp. 1–17. DOI: 10.1016/s0167-4838(01)00187-x.

Hassell, Anne M.; An, Gang; Bledsoe, Randy K.; Bynum, Jane M.; Carter, H. Luke; Deng, Su-Jun J. et al. (2007): Crystallization of protein-ligand complexes. In *Acta crystallographica. Section D, Biological crystallography* 63 (Pt 1), pp. 72–79. DOI: 10.1107/S0907444906047020.

Headd, Jeffrey J.; Echols, Nathaniel; Afonine, Pavel V.; Grosse-Kunstleve, Ralf W.; Chen, Vincent B.; Moriarty, Nigel W. et al. (2012): Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. In *Acta crystallographica. Section D, Biological crystallography* 68 (Pt 4), pp. 381–390. DOI: 10.1107/S0907444911047834.

Heras, Begoña; Martin, Jennifer L. (2005): Post-crystallization treatments for improving diffraction quality of protein crystals. In *Acta crystallographica. Section D, Biological crystallography* 61 (Pt 9), pp. 1173–1180. DOI: 10.1107/S0907444905019451.

Hicks, Matt N.; Gunasekara, Sanjiva; Serate, Jose; Park, Jin; Mosharaf, Pegah; Zhou, Yue et al. (2017): Gly184 of the Escherichia coli cAMP receptor protein provides optimal context for both DNA binding and RNA polymerase interaction. In *Journal of microbiology (Seoul, Korea)* 55 (10), pp. 816–822. DOI: 10.1007/s12275-017-7266-x.

Holmquist, Peter C.; Holmquist, Gerald P.; Summers, Michael L. (2011): Comparing binding site information to binding affinity reveals that Crp/DNA complexes have several distinct binding conformers. In *Nucleic acids research* 39 (15), pp. 6813–6824. DOI: 10.1093/nar/gkr369.

Holyoak, T.; Nowak, T. (2001): Structural investigation of the binding of nucleotide to phosphoenolpyruvate carboxykinase by NMR. In *Biochemistry* 40 (37), pp. 11037–11047. DOI: 10.1021/bi011374n.

Hopfner, Karl-Peter; Karcher, Annette; Craig, Lisa; Woo, Tammy T.; Carney, James P.; Tainer, John A. (2001): Structural Biochemistry and Interaction Architecture of the DNA Double-Strand Break Repair Mre11 Nuclease and Rad50-ATPase. In *Cell* 105 (4), pp. 473–485. DOI: 10.1016/S0092-8674(01)00335-X.

Hudson, Brian P.; Quispe, Joel; Lara-González, Samuel; Kim, Younggyu; Berman, Helen M.; Arnold, Eddy et al. (2009): Three-dimensional EM structure of an intact activator-dependent transcription initiation complex. In *Proceedings of the National Academy of Sciences of the United States of America* 106 (47), pp. 19830–19835. DOI: 10.1073/pnas.0908782106.

Hunter, Tony (2012): Why nature chose phosphate to modify proteins. In *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 367 (1602), pp. 2513–2516. DOI: 10.1098/rstb.2012.0013.

Huynh, Kathy; Partch, Carrie L. (2015): Analysis of protein stability and ligand interactions by thermal shift assay. In *Current protocols in protein science* 79, 28.9.1-28.9.14. DOI: 10.1002/0471140864.ps2809s79.

Ibáñez, Fernando; Wall, Luis; Fabra, Adriana (2017): Starting points in plant-bacteria nitrogen-fixing symbioses: intercellular invasion of the roots. In *Journal of experimental botany* 68 (8), pp. 1905–1918. DOI: 10.1093/jxb/erw387.

Imamura, R.; Yamanaka, K.; Ogura, T.; Hiraga, S.; Fujita, N.; Ishihama, A.; Niki, H. (1996): Identification of the cpdA gene encoding cyclic 3',5'-adenosine monophosphate phosphodiesterase in Escherichia coli. In *The Journal of biological chemistry* 271 (41), pp. 25423–25429. DOI: 10.1074/jbc.271.41.25423.

Inclan, Yuki F.; Huseby, Medora J.; Engel, Joanne N. (2011): FimL regulates cAMP synthesis in Pseudomonas aeruginosa. In *PloS one* 6 (1), e15867. DOI: 10.1371/journal.pone.0015867.

Iwai, Wakari; Yagi, Daichi; Ishikawa, Takuya; Ohnishi, Yuki; Tanaka, Ichiro; Niimura, Nobuo (2008): Crystallization and evaluation of hen egg-white lysozyme crystals for protein pH titration in the crystalline state. In *Journal of synchrotron radiation* 15 (Pt 3), pp. 312–315. DOI: 10.1107/S0909049507059559.

J. Hearon (1948): The Configuration of Cobaltodihistidine and Oxy-Bis(Cobaltodihistidine)[2]. In *JNCI: Journal of the National Cancer Institute* (9), pp. 1–11. DOI: 10.1093/jnci/9.1.1.

Jackson, Colin J.; Carr, Paul D.; Liu, Jian-Wei; Watt, Stephen J.; Beck, Jennifer L.; Ollis, David L. (2007): The structure and function of a novel glycerophosphodiesterase from Enterobacter aerogenes. In *Journal of Molecular Biology* 367 (4), pp. 1047–1062. DOI: 10.1016/j.jmb.2007.01.032.

Jackson, Colin J.; Hadler, Kieran S.; Carr, Paul D.; Oakley, Aaron J.; Yip, Sylvia; Schenk, Gerhard; Ollis, David L. (2008): Malonate-bound structure of the glycerophosphodiesterase from Enterobacter aerogenes (GpdQ) and characterization of the native Fe2+ metal-ion preference. In *Acta crystallographica. Section F, Structural biology and crystallization communications* 64 (Pt 8), pp. 681–685. DOI: 10.1107/S1744309108017600.

Jelesarov, Ilian; Bosshard, Hans Rudolf (1999): Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. In *J. Mol. Recognit.* 12 (1), pp. 3–18. DOI: 10.1002/(SICI)1099-1352(199901/02)12:1<3::AID-JMR441>3.0.CO;2-6.

Johnson, Kenneth A.; Goody, Roger S. (2011): The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. In *Biochemistry* 50 (39), pp. 8264–8269. DOI: 10.1021/bi201284u.

Jones, Kathryn M.; Kobayashi, Hajime; Davies, Bryan W.; Taga, Michiko E.; Walker, Graham C. (2007): How rhizobial symbionts invade plants: the Sinorhizobium-Medicago model. In *Nature reviews. Microbiology* 5 (8), pp. 619–633. DOI: 10.1038/nrmicro1705.

Joosten, Robbie P.; Long, Fei; Murshudov, Garib N.; Perrakis, Anastassis (2014): The PDB_REDO server for macromolecular structure model optimization. In *IUCrJ* 1 (Pt 4), pp. 213–220. DOI: 10.1107/S2052252514009324.

Jungbauer, Alois; Hahn, Rainer (2009): Chapter 22 Ion-Exchange Chromatography. In Richard R. Burgess, Murray P. Deutscher (Eds.): Guide to protein purification, vol. 463. 2. ed. San Diego, Calif.: Elsevier Acad. Press (Methods in Enzymology, 463), pp. 349–371.

Kabsch, Wolfgang (2010): XDS. In *Acta crystallographica. Section D, Biological crystallography* 66 (Pt 2), pp. 125–132. DOI: 10.1107/S0907444909047337.

Kalivoda, Eric J.; Brothers, Kimberly M.; Stella, Nicholas A.; Schmitt, Matthew J.; Shanks, Robert M. Q. (2013): Bacterial cyclic AMP-phosphodiesterase activity coordinates biofilm formation. In *PloS one* 8 (7), e71267. DOI: 10.1371/journal.pone.0071267.

Karplus, P. Andrew; Diederichs, Kay (2015): Assessing and maximizing data quality in macromolecular crystallography. In *Current opinion in structural biology* 34, pp. 60–68. DOI: 10.1016/j.sbi.2015.07.003.

Kelemen, M. V.; Sharpe, J. E. (1979): Controlled cell disruption: a comparison of the forces required to disrupt different micro-organisms. In *Journal of Cell Science* 35 (1), pp. 431–441.

Keppetipola, Niroshika; Shuman, Stewart (2008): A phosphate-binding histidine of binuclear metallophosphodiesterase enzymes is a determinant of 2',3'-cyclic nucleotide phosphodiesterase activity. In *The Journal of biological chemistry* 283 (45), pp. 30942–30949. DOI: 10.1074/jbc.M805064200.

Kim, Han-Suk; Kim, Sung-Min; Lee, Hyun-Jung; Park, Soon-Jung; Lee, Kyu-Ho (2009): Expression of the cpdA gene, encoding a 3',5'-cyclic AMP (cAMP) phosphodiesterase, is positively regulated by the cAMP-cAMP receptor protein complex. In *Journal of bacteriology* 191 (3), pp. 922–930. DOI: 10.1128/JB.01350-08.

Kimura, Yoshio; Mishima, Yukako; Nakano, Hiromi; Takegawa, Kaoru (2002): An adenylyl cyclase, CyaA, of Myxococcus xanthus functions in signal transduction during osmotic stress. In *Journal of bacteriology* 184 (13), pp. 3578–3585. DOI: 10.1128/jb.184.13.3578-3585.2002.

Kimura, Yoshio; Ohtani, Mika; Takegawa, Kaoru (2005): An adenylyl cyclase, CyaB, acts as an osmosensor in Myxococcus xanthus. In *Journal of bacteriology* 187 (10), pp. 3593–3598. DOI: 10.1128/JB.187.10.3593-3598.2005.

Kimura, Yoshio; Okazaki, Nozomi; Takegawa, Kaoru (2009): Enzymatic characteristics of two novel Myxococcus xanthus enzymes, PdeA and PdeB, displaying 3',5'- and 2',3'-cAMP phosphodiesterase, and phosphatase activities. In *FEBS letters* 583 (2), pp. 443–448. DOI: 10.1016/j.febslet.2008.12.044.

Kleywegt, Gerard J.; Alwyn Jones, T. (1997): [11] Model building and refinement practice. In Charles W. Carter (Ed.): Macromolecular crystallography, vol. 277. San Diego, Calif.: Acad. Press (Methods in Enzymology, 277), pp. 208–230.

Knapp, Gwendowlyn S.; McDonough, Kathleen A. (2014): Cyclic AMP Signaling in Mycobacteria. In *Microbiology spectrum* 2 (2). DOI: 10.1128/microbiolspec.MGM2-0011-2013.

Knöfel, T.; Sträter, N. (2001): Mechanism of hydrolysis of phosphate esters by the dimetal center of 5'-nucleotidase based on crystal structures. In *Journal of Molecular Biology* 309 (1), pp. 239–254. DOI: 10.1006/jmbi.2001.4656.

Körner, Heinz; Sofia, Heidi J.; Zumft, Walter G. (2003): Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. In *FEMS Microbiol Rev* 27 (5), pp. 559–592. DOI: 10.1016/S0168-6445(03)00066-4.

Krogh, A.; Larsson, B.; Heijne, G. von; Sonnhammer, E. L. (2001): Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. In *Journal of Molecular Biology* 305 (3), pp. 567–580. DOI: 10.1006/jmbi.2000.4315.

Krol, Elizaveta; Klaner, Christina; Gnau, Petra; Kaever, Volkhard; Essen, Lars-Oliver; Becker, Anke (2016): Cyclic mononucleotide- and Clr-dependent gene regulation in Sinorhizobium meliloti. In *Microbiology (Reading, England)* 162 (10), pp. 1840–1856. DOI: 10.1099/mic.0.000356.

Kubo, Ryogo; Tomita, Kazuhisa (1954): A General Theory of Magnetic Resonance Absorption. In *J. Phys. Soc. Jpn.* 9 (6), pp. 888–919. DOI: 10.1143/JPSJ.9.888.

Kuchinskas, Michael; Li, Huiying; Conrad, Mary; Roberts, Gary; Poulos, Thomas L. (2006): The role of the DNA-binding domains in CooA activation. In *Biochemistry* 45 (23), pp. 7148–7153. DOI: 10.1021/bi052609o.

Kumar, Sudhir; Stecher, Glen; Li, Michael; Knyaz, Christina; Tamura, Koichiro (2018): MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. In *Molecular biology and evolution* 35 (6), pp. 1547–1549. DOI: 10.1093/molbev/msy096.

Laemmli, U. K. (1970): Cleavage of structural proteins during the assembly of the head of bacteriophage T4. In *Nature* 227 (5259), pp. 680–685. DOI: 10.1038/227680a0.

Le Roux, Johannes J.; Hui, Cang; Keet, Jan-Hendrik; Ellis, Allan G. (2017): Co-introduction vs ecological fitting as pathways to the establishment of effective mutualisms during biological invasions. In *The New phytologist* 215 (4), pp. 1354–1360. DOI: 10.1111/nph.14593.

Leduc, Jason L.; Roberts, Gary P. (2009): Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of Xanthomonas axonopodis pv. citri. In *Journal of bacteriology* 191 (22), pp. 7121–7122. DOI: 10.1128/JB.00845-09.

Lee, Pei Yun; Costumbrado, John; Hsu, Chih-Yuan; Kim, Yong Hoon (2012): Agarose gel electrophoresis for the separation of DNA fragments. In *Journal of visualized experiments : JoVE* (62). DOI: 10.3791/3923.

Levy, Colin; Pike, Katharine; Heyes, Derren J.; Joyce, M. Gordon; Gabor, Krisztina; Smidt, Hauke et al. (2008): Molecular basis of halorespiration control by CprK, a CRP-FNR type transcriptional regulator. In *Molecular microbiology* 70 (1), pp. 151–167. DOI: 10.1111/j.1365-2958.2008.06399.x.

Lewis, Alison; Seckler, Marcelo Martins; Kramer, Herman; van Rosmalen, Gerda (2015): Batch crystallization. In Alison Emslie Lewis (Ed.): Industrial crystallization. Fundamentals and applications. First published. Cambridge: Cambridge University Press, pp. 178–191.

Liebschner, Dorothee; Afonine, Pavel V.; Baker, Matthew L.; Bunkóczi, Gábor; Chen, Vincent B.; Croll, Tristan I. et al. (2019): Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. In *Acta crystallographica. Section D, Structural biology* 75 (Pt 10), pp. 861–877. DOI: 10.1107/S2059798319011471.

Limpens, Erik; van Zeijl, Arjan; Geurts, Rene (2015): Lipochitooligosaccharides modulate plant host immunity to enable endosymbioses. In *Annual review of phytopathology* 53, pp. 311–334. DOI: 10.1146/annurev-phyto-080614-120149.

Linder, J. U. (2006): Class III adenylyl cyclases: molecular mechanisms of catalysis and regulation. In *Cellular and molecular life sciences : CMLS* 63 (15), pp. 1736–1751. DOI: 10.1007/s00018-006-6072-0.

Linder, Jürgen U. (2010): cGMP production in bacteria. In *Molecular and cellular biochemistry* 334 (1-2), pp. 215–219. DOI: 10.1007/s11010-009-0321-0.

Lineweaver, Hans; Burk, Dean (1934): The determination of enzyme dissociation constants. In *Journal of the American chemical society* 56 (3), pp. 658–666.

Liu, Bin; Hong, Chuan; Huang, Rick K.; Yu, Zhiheng; Steitz, Thomas A. (2017): Structural basis of bacterial transcription activation. In *Science (New York, N.Y.)* 358 (6365), pp. 947–951. DOI: 10.1126/science.aao1923.

Llácer, José L.; Espinosa, Javier; Castells, Miguel A.; Contreras, Asunción; Forchhammer, Karl; Rubio, Vicente (2010): Structural basis for the regulation of NtcA-dependent transcription by proteins PipX and PII. In *Proceedings of the National Academy of Sciences of the United States of America* 107 (35), pp. 15397–15402. DOI: 10.1073/pnas.1007015107.

Lorenzo, V. de; Herrero, M.; Giovannini, F.; Neilands, J. B. (1988): Fur (ferric uptake regulation) protein and CAP (catabolite-activator protein) modulate transcription of fur gene in Escherichia coli. In *European journal of biochemistry* 173 (3), pp. 537–546. DOI: 10.1111/j.1432-1033.1988.tb14032.x.

Lu, Hui-Meng; Yin, Da-Chuan; Liu, Yong-Ming; Guo, Wei-Hong; Zhou, Ren-Bin (2012): Correlation between protein sequence similarity and crystallization reagents in the biological macromolecule crystallization database. In *International journal of molecular sciences* 13 (8), pp. 9514–9526. DOI: 10.3390/ijms13089514.

Macfadyen, L. P.; Ma, C.; Redfield, R. J. (1998): A 3',5' cyclic AMP (cAMP) phosphodiesterase modulates cAMP levels and optimizes competence in Haemophilus influenzae Rd. In *Journal of bacteriology* 180 (17), pp. 4401–4405. DOI: 10.1128/JB.180.17.4401-4405.1998.

Małecki, J.; Polit, A.; Wasylewski, Z. (2000): Kinetic studies of cAMP-induced allosteric changes in cyclic AMP receptor protein from Escherichia coli. In *The Journal of biological chemistry* 275 (12), pp. 8480–8486. DOI: 10.1074/jbc.275.12.8480.

Malvern Panalytical (Ed.) (2020a): Enzyme kinetics assays with MicroCal ITC systems. Part 2: Methods. Available online at https://www.malvernpanalytical.com/de/learn/knowledge-center/technical-notes/TN200528EnzymeKineticAssaysITCPart2Methods.

Malvern Panalytical (Ed.) (2020b): Using Isothermal Titration Calorimetry to Characterize Enzyme Kinetics. Part 1: Principles. Available online at https://www.malvernpanalytical.com/en/learn/knowledge-center/whitepapers/WP280520ITCEnzymeKineticsPart1Principles.html.

Mandel, M.; Higa, A. (1970): Calcium-dependent bacteriophage DNA infection. In *Journal of Molecular Biology* 53 (1), pp. 159–162. DOI: 10.1016/0022-2836(70)90051-3.

Manneh-Roussel, Jainaba; Haycocks, James R. J.; Magán, Andrés; Perez-Soto, Nicolas; Voelz, Kerstin; Camilli, Andrew et al. (2018): cAMP Receptor Protein Controls Vibrio cholerae Gene Expression in Response to Host Colonization. In *mBio* 9 (4). DOI: 10.1128/mBio.00966-18.

Marden, Jeremiah N.; Dong, Qian; Roychowdhury, Sugata; Berleman, James E.; Bauer, Carl E. (2011): Cyclic GMP controls Rhodospirillum centenum cyst development. In *Molecular microbiology* 79 (3), pp. 600–615. DOI: 10.1111/j.1365-2958.2010.07513.x.

Masson, Glenn R.; Burke, John E.; Ahn, Natalie G.; Anand, Ganesh S.; Borchers, Christoph; Brier, Sébastien et al. (2019): Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. In *Nature methods* 16 (7), pp. 595–602. DOI: 10.1038/s41592-019-0459-y.

Matange, Nishad (2015a): Revisiting bacterial cyclic nucleotide phosphodiesterases: cyclic AMP hydrolysis and beyond. In *FEMS microbiology letters* 362 (22). DOI: 10.1093/femsle/fnv183.

Matange, Nishad; Hunt, Debbie M.; Buxton, Roger S.; Visweswariah, Sandhya S. (2013): Overexpression of the Rv0805 phosphodiesterase elicits a cAMP-independent transcriptional response. In *Tuberculosis (Edinburgh, Scotland)* 93 (5), pp. 492–500. DOI: 10.1016/j.tube.2013.05.004.

Matange, Nishad; Podobnik, Marjetka; Visweswariah, Sandhya S. (2014): The non-catalytic "cap domain" of a mycobacterial metallophosphoesterase regulates its expression and localization in the cell. In *The Journal of biological chemistry* 289 (32), pp. 22470–22481. DOI: 10.1074/jbc.M114.578328.

Matange, Nishad; Podobnik, Marjetka; Visweswariah, Sandhya S. (2015b): Metallophosphoesterases: structural fidelity with functional promiscuity. In *The Biochemical journal* 467 (2), pp. 201–216. DOI: 10.1042/BJ20150028.

Mathieu-Demazière, Céline; Poinsot, Véréna; Masson-Boivin, Catherine; Garnerone, Anne-Marie; Batut, Jacques (2013): Biochemical and functional characterization of SpdA, a 2', 3'cyclic nucleotide phosphodiesterase from Sinorhizobium meliloti. In *BMC microbiology* 13, p. 268. DOI: 10.1186/1471-2180-13-268.

Mazzei, Luca; Ciurli, Stefano; Zambelli, Barbara (2016): Isothermal Titration Calorimetry to Characterize Enzymatic Reactions. In *Methods in enzymology* 567, pp. 215–236. DOI: 10.1016/bs.mie.2015.07.022.

McPherson, Alexander; Cudney, Bob (2014): Optimization of crystallization conditions for biological macromolecules. In *Acta crystallographica. Section F, Structural biology communications* 70 (Pt 11), pp. 1445–1467. DOI: 10.1107/S2053230X14019670.

McPherson, Alexander; Gavira, Jose A. (2014): Introduction to protein crystallization. In *Acta crystallographica. Section F, Structural biology communications* 70 (Pt 1), pp. 2–20. DOI: 10.1107/S2053230X13033141.

Messerschmidt, Albrecht (2007): X-Ray Crystallography of Biomacromolecules. A Practical Guide. Hoboken: Wiley-VCH. Available online at http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=189980.

Miller, Darcie J.; Shuvalova, Ludmilla; Evdokimova, Elena; Savchenko, Alexei; Yakunin, Alexander F.; Anderson, Wayne F. (2007): Structural and biochemical characterization of a novel Mn2+-dependent phosphodiesterase encoded by the yfcE gene. In *Protein science : a publication of the Protein Society* 16 (7), pp. 1338–1348. DOI: 10.1110/ps.072764907.

MILNER, H. W.; LAWRENCE, N. S.; FRENCH, C. S. (1950): Colloidal dispersion of chloroplast material. In *Science (New York, N.Y.)* 111 (2893), pp. 633–634. DOI: 10.1126/science.111.2893.633.

Molnár, I.; Horváth, C. (1976): Reverse-phase chromatography of polar biological substances: separation of catechol compounds by high-performance liquid chromatography. In *Clinical Chemistry* 22 (9), pp. 1497–1502. DOI: 10.1093/clinchem/22.9.1497.

Müller, Ilka (2017): Guidelines for the successful generation of protein-ligand complex crystals. In *Acta crystallographica. Section D, Structural biology* 73 (Pt 2), pp. 79–92. DOI: 10.1107/S2059798316020271.

Mullis, Kary B.; Faloona, Fred A. (1987): [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In Ray Wu (Ed.): Recombinant DNA, vol. 155. San Diego, Calif.: Academic Press (Methods in Enzymology, 155), pp. 335–350.

Murray, Jeremy D. (2011): Invasion by invitation: rhizobial infection in legumes. In *Molecular plant-microbe interactions : MPMI* 24 (6), pp. 631–639. DOI: 10.1094/MPMI-08-10-0181.

Murray, Jeremy D.; Liu, Cheng-Wu; Chen, Yi; Miller, Anthony J. (2017): Nitrogen sensing in legumes. In *Journal of experimental botany* 68 (8), pp. 1919–1926. DOI: 10.1093/jxb/erw405.

Niesen, Frank H.; Berglund, Helena; Vedadi, Masoud (2007): The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. In *Nature protocols* 2 (9), pp. 2212–2221. DOI: 10.1038/nprot.2007.321.

Noble, James E.; Bailey, Marc J.A. (2009): Chapter 8 Quantitation of Protein. In Richard R. Burgess, Murray P. Deutscher (Eds.): Guide to protein purification, vol. 463. 2. ed. San Diego, Calif.: Elsevier Acad. Press (Methods in Enzymology, 463), pp. 73–95.

Nolan, Laura M.; Beatson, Scott A.; Croft, Larry; Jones, Peter M.; George, Anthony M.; Mattick, John S. et al. (2012): Extragenic suppressor mutations that restore twitching motility to fimL mutants of Pseudomonas aeruginosa are associated with elevated intracellular cyclic AMP levels. In *MicrobiologyOpen* 1 (4), pp. 490–501. DOI: 10.1002/mbo3.49.

Nwokeoji, Alison O.; Earll, Mark E.; Kilby, Peter M.; Portwood, David E.; Dickman, Mark J. (2019): High resolution fingerprinting of single and double-stranded RNA using ion-pair reverse-phase chromatography. In *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* 1104, pp. 212–219. DOI: 10.1016/j.jchromb.2018.11.027.

Oldroyd, G. (2001): Dissecting Symbiosis: Developments in Nod Factor Signal Transduction. In *Annals of Botany* 87 (6), pp. 709–718. DOI: 10.1006/anbo.2001.1410.

Olson, Bradley J. S. C. (2016): Assays for Determination of Protein Concentration. In *Current protocols in pharmacology* 73, A.3A.1-A.3A.32. DOI: 10.1002/cpph.3.

Oscarsson, Marcus; Beteva, Antonia; Flot, David; Gordon, Elspeth; Guijarro, Matias; Leonard, Gordon et al. (2019): MXCuBE2: the dawn of MXCuBE Collaboration. In *Journal of synchrotron radiation* 26 (Pt 2), pp. 393–405. DOI: 10.1107/S1600577519001267.

Ovchinnikov, Sergey; Kamisetty, Hetunandan; Baker, David (2014): Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. In *eLife* 3, e02030. DOI: 10.7554/eLife.02030.

Parker, M. W. (2003): Protein structure from X-ray diffraction. In *Journal of biological physics* 29 (4), pp. 341–362.

Pei, Xiaojing; Huang, Hongduan; Chen, Yang; Li, Chenxi; Liu, Feng; Li, Na (2016): Modulating fluorescence anisotropy of dye-labeled DNA without involving mass amplification. In *Talanta* 154, pp. 567–573. DOI: 10.1016/j.talanta.2016.01.016.

Pérez-Rueda, E.; Collado-Vides, J. (2000): The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. In *Nucleic acids research* 28 (8), pp. 1838–1847. DOI: 10.1093/nar/28.8.1838.

Pettersen, Eric F.; Goddard, Thomas D.; Huang, Conrad C.; Couch, Gregory S.; Greenblatt, Daniel M.; Meng, Elaine C.; Ferrin, Thomas E. (2004): UCSF Chimera--a visualization system for exploratory research and analysis. In *Journal of computational chemistry* 25 (13), pp. 1605–1612. DOI: 10.1002/jcc.20084.

Pitt-Rivers, R.; Impiombato, F. S. (1968): The binding of sodium dodecyl sulphate to various proteins. In *The Biochemical journal* 109 (5), pp. 825–830. DOI: 10.1042/bj1090825.

Podobnik, Marjetka; Tyagi, Richa; Matange, Nishad; Dermol, Urska; Gupta, Arun K.; Mattoo, Rohini et al. (2009): A mycobacterial cyclic AMP phosphodiesterase that moonlights as a modifier of cell wall permeability. In *The Journal of biological chemistry* 284 (47), pp. 32846–32857. DOI: 10.1074/jbc.M109.049635.

Polson, Cara; Sarkar, Pratibha; Incledon, Bev; Raguvaran, Vanaja; Grant, Russell (2003): Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography–tandem mass spectrometry. In *Journal of Chromatography B* 785 (2), pp. 263–275. DOI: 10.1016/S1570-0232(02)00914-5.

Popovych, Nataliya; Tzeng, Shiou-Ru; Tonelli, Marco; Ebright, Richard H.; Kalodimos, Charalampos G. (2009): Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. In *Proceedings of the National Academy of Sciences of the United States of America* 106 (17), pp. 6927–6932. DOI: 10.1073/pnas.0900595106.

Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G. (1975): Metal chelate affinity chromatography, a new approach to protein fractionation. In *Nature* 258 (5536), pp. 598–599. DOI: 10.1038/258598a0.

Porath, J.; Flodin, P. (1959): Gel filtration: a method for desalting and group separation. In *Nature* 183 (4676), pp. 1657–1659. DOI: 10.1038/1831657a0.

Potterton, Liz; Agirre, Jon; Ballard, Charles; Cowtan, Kevin; Dodson, Eleanor; Evans, Phil R. et al. (2018): CCP4i2: the new graphical user interface to the CCP4 program suite. In *Acta* *crystallographica. Section D, Structural biology* 74 (Pt 2), pp. 68–84. DOI: 10.1107/S2059798317016035.

Protozanova, Ekaterina; Yakovchuk, Peter; Frank-Kamenetskii, Maxim D. (2004): Stacked-unstacked equilibrium at the nick site of DNA. In *Journal of Molecular Biology* 342 (3), pp. 775–785. DOI: 10.1016/j.jmb.2004.07.075.

Rabi, I. I.; Zacharias, J. R.; Millman, S.; Kusch, P. (1938): A New Method of Measuring Nuclear Magnetic Moment. In *Phys. Rev.* 53 (4), p. 318. DOI: 10.1103/PhysRev.53.318.

Rahimzadeh, Maral; Sadeghizadeh, Majid; Najafi, Farhood; Arab, Seyed; Mobasheri, Hamid (2016): Impact of heat shock step on bacterial transformation efficiency. In *Molecular Biology Research Communications* 5 (4), pp. 257–261.

Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. (1963): Stereochemistry of polypeptide chain configurations. In *Journal of Molecular Biology* 7 (1), pp. 95–99. DOI: 10.1016/S0022-2836(63)80023-6.

Rauch, Annika; Leipelt, Martina; Russwurm, Michael; Steegborn, Clemens (2008): Crystal structure of the guanylyl cyclase Cya2. In *Proceedings of the National Academy of Sciences of the United States of America* 105 (41), pp. 15720–15725. DOI: 10.1073/pnas.0808473105.

Reddy, Manchi C. M.; Palaninathan, Satheesh K.; Bruning, John B.; Thurman, Cory; Smith, Danielle; Sacchettini, James C. (2009): Structural insights into the mechanism of the allosteric transitions of Mycobacterium tuberculosis cAMP receptor protein. In *The Journal of biological chemistry* 284 (52), pp. 36581–36591. DOI: 10.1074/jbc.M109.041343.

Reuten, Raphael; Nikodemus, Denise; Oliveira, Maria B.; Patel, Trushar R.; Brachvogel, Bent; Breloy, Isabelle et al. (2016): Maltose-Binding Protein (MBP), a Secretion-Enhancing Tag for Mammalian Protein Expression Systems. In *PloS one* 11 (3), e0152386. DOI: 10.1371/journal.pone.0152386.

Richter, Wito (2002): 3',5' Cyclic nucleotide phosphodiesterases class III: members, structure, and catalytic mechanism. In *Proteins* 46 (3), pp. 278–286. DOI: 10.1002/prot.10049.

Rickman, Lisa; Scott, Colin; Hunt, Debbie M.; Hutchinson, Thomas; Menéndez, M. Carmen; Whalan, Rachael et al. (2005): A member of the cAMP receptor protein family of transcription regulators in Mycobacterium tuberculosis is required for virulence in mice and controls transcription of the rpfA gene coding for a resuscitation promoting factor. In *Molecular microbiology* 56 (5), pp. 1274–1286. DOI: 10.1111/j.1365-2958.2005.04609.x.

Rogers, H. J.; Perkins, H. R.; Ward, J. B. (2012): Microbial Cell Walls and Membranes. Dordrecht: Springer Netherlands. Available online at https://ebookcentral.proquest.com/lib/gbv/detail.action?docID=3566819.

Rohs, Remo; Jin, Xiangshu; West, Sean M.; Joshi, Rohit; Honig, Barry; Mann, Richard S. (2010): Origins of specificity in protein-DNA recognition. In *Annual review of biochemistry* 79, pp. 233–269. DOI: 10.1146/annurev-biochem-060408-091030.

Rosano, Germán L.; Ceccarelli, Eduardo A. (2014): Recombinant protein expression in Escherichia coli: advances and challenges. In *Frontiers in microbiology* 5, p. 172. DOI: 10.3389/fmicb.2014.00172.

Rosenberg, Alan H.; Lade, Barbara N.; Dao-shan, Chui; Lin, Shu-Wha; Dunn, John J.; Studier, F.William (1987): Vectors for selective expression of cloned DNAs by T7 RNA polymerase. In *Gene* 56 (1), pp. 125–135. DOI: 10.1016/0378-1119(87)90165-X.

Roychowdhury, Sugata; Dong, Qian; Bauer, Carl E. (2015): DNA-binding properties of a cGMP-binding CRP homologue that controls development of metabolically dormant cysts of Rhodospirillum centenum. In *Microbiology (Reading, England)* 161 (11), pp. 2256–2264. DOI: 10.1099/mic.0.000172.

Ryu, S.; Kim, J.; Adhya, S.; Garges, S. (1993): Pivotal role of amino acid at position 138 in the allosteric hinge reorientation of cAMP receptor protein. In *Proceedings of the National Academy of Sciences of the United States of America* 90 (1), pp. 75–79. DOI: 10.1073/pnas.90.1.75.

SantaLucia, John; Hicks, Donald (2004): The thermodynamics of DNA structural motifs. In *Annual review of biophysics and biomolecular structure* 33, pp. 415–440. DOI: 10.1146/annurev.biophys.32.110601.141800.

Sapan, Christine V.; Lundblad, Roger L.; Price, Nicholas C. (1999): Colorimetric protein assay techniques. In *Biotechnology and Applied Biochemistry* 29 (2), pp. 99–108. DOI: 10.1111/j.1470-8744.1999.tb00538.x.

Saridakis, Emmanuel; Chayen, Naomi E. (2009): Towards a 'universal' nucleant for protein crystallization. In *Trends in biotechnology* 27 (2), pp. 99–106. DOI: 10.1016/j.tibtech.2008.10.008.

Sasse, Joachim; Gallagher, Sean R. (2004): Staining proteins in gels. In *Current protocols in immunology* Chapter 8, Unit 8.9. DOI: 10.1002/0471142735.im0809s58.

Schenk, Gerhard; Mitić, Nataša; Hanson, Graeme R.; Comba, Peter (2013): Purple acid phosphatase: A journey into the function and mechanism of a colorful enzyme. In *Coordination Chemistry Reviews* 257 (2), pp. 473–482. DOI: 10.1016/j.ccr.2012.03.020.

Schulte, Julia; Baumgart, Meike; Bott, Michael (2017a): Development of a single-cell GlxR-based cAMP biosensor for Corynebacterium glutamicum. In *Journal of biotechnology* 258, pp. 33–40. DOI: 10.1016/j.jbiotec.2017.07.004.

Schulte, Julia; Baumgart, Meike; Bott, Michael (2017b): Identification of the cAMP phosphodiesterase CpdA as novel key player in cAMP-dependent regulation in Corynebacterium glutamicum. In *Molecular microbiology* 103 (3), pp. 534–552. DOI: 10.1111/mmi.13574.

Schultz, S. C.; Shields, G. C.; Steitz, T. A. (1991): Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. In *Science (New York, N.Y.)* 253 (5023), pp. 1001–1007. DOI: 10.1126/science.1653449.

Scopes, Robert K. (1994): Protein Purification. Principles and Practice. Third Edition. New York, NY: Springer (Springer Advanced Texts in Chemistry).

Serwer, Philip (1983): Agarose gels: Properties and use for electrophoresis. In *Electrophoresis* 4 (6), pp. 375–382. DOI: 10.1002/elps.1150040602.

Shabb, J. B.; Corbin, J. D. (1992): Cyclic nucleotide-binding domains in proteins having diverse functions. In *The Journal of biological chemistry* 267 (9), pp. 5723–5726.

Sharma, Hitesh; Yu, Shaoning; Kong, Jilie; Wang, Jimin; Steitz, Thomas A. (2009): Structure of apo-CAP reveals that large conformational changes are necessary for DNA binding. In *Proceedings of the National Academy of Sciences of the United States of America* 106 (39), pp. 16604–16609. DOI: 10.1073/pnas.0908380106.

Shenoy, Avinash R.; Capuder, Maja; Draskovic, Petra; Lamba, Doriano; Visweswariah, Sandhya S.; Podobnik, Marjetka (2007): Structural and biochemical analysis of the Rv0805 cyclic nucleotide phosphodiesterase from Mycobacterium tuberculosis. In *Journal of Molecular Biology* 365 (1), pp. 211–225. DOI: 10.1016/j.jmb.2006.10.005.

Shenoy, Avinash R.; Sreenath, Nandini; Podobnik, Marjetka; Kovacevic, Miroslav; Visweswariah, Sandhya S. (2005): The Rv0805 gene from Mycobacterium tuberculosis encodes a 3',5'-cyclic nucleotide phosphodiesterase: biochemical and mutational analysis. In *Biochemistry* 44 (48), pp. 15695–15704. DOI: 10.1021/bi0512391.

Shibuya, M. (1977): A possible involvement of cya gene in the synthesis of cyclic guanosine 3′:5′-monophosphate in E. coli. In *Cell* 12 (2), pp. 521–528. DOI: 10.1016/0092-8674(77)90128-3.

Skorupski, Karen; Taylor, Ronald K. (1997): Sequence and functional analysis of the gene encoding Vibrio cholerae cAMP receptor protein. In *Gene* 198 (1-2), pp. 297–303. DOI: 10.1016/S0378-1119(97)00331-4.

Soberón-Chávez, Gloria; Alcaraz, Luis D.; Morales, Estefanía; Ponce-Soto, Gabriel Y.; Servín-González, Luis (2017): The Transcriptional Regulators of the CRP Family Regulate Different Essential Bacterial Functions and Can Be Inherited Vertically and Horizontally. In *Frontiers in microbiology* 8, p. 959. DOI: 10.3389/fmicb.2017.00959.

Sorroche, Fernando; Walch, Mathilda; Zou, Lan; Rengel, David; Maillet, Fabienne; Gibelin-Viala, Chrystel et al. (2019): Endosymbiotic Sinorhizobium meliloti modulate Medicago root susceptibility to secondary infection via ethylene. In *The New phytologist* 223 (3), pp. 1505–1515. DOI: 10.1111/nph.15883.

Sprent, Janet I. (2008): 60Ma of legume nodulation. What's new? What's changing? In *Journal of experimental botany* 59 (5), pp. 1081–1084. DOI: 10.1093/jxb/erm286.

Spriestersbach, Anne; Kubicek, Jan; Schäfer, Frank; Block, Helena; Maertens, Barbara (2015): Purification of His-Tagged Proteins. In *Methods in enzymology* 559, pp. 1–15. DOI: 10.1016/bs.mie.2014.11.003.

Stagnari, Fabio; Maggio, Albino; Galieni, Angelica; Pisante, Michele (2017): Multiple benefits of legumes for agriculture sustainability: an overview. In *Chem. Biol. Technol. Agric.* 4 (1). DOI: 10.1186/s40538-016-0085-1.

Stapleton, Melanie; Haq, Ihtshamul; Hunt, Debbie M.; Arnvig, Kristine B.; Artymiuk, Peter J.; Buxton, Roger S.; Green, Jeffrey (2010): Mycobacterium tuberculosis cAMP receptor protein (Rv3676) differs from the Escherichia coli paradigm in its cAMP binding and DNA binding properties and transcription activation properties. In *The Journal of biological chemistry* 285 (10), pp. 7016–7027. DOI: 10.1074/jbc.M109.047720.

Steinchen, Wieland; Linne, Uwe; Bange, Gert: HDX-MS in den Lebenswissenschaften. In *Biospektrum* 23 (7), pp. 772–775.

Streeter, John; Wong, Peter P. (1988): Inhibition of legume nodule formation and N 2 fixation by nitrate. In *Critical Reviews in Plant Sciences* 7 (1), pp. 1–23. DOI: 10.1080/07352688809382257.

Strong, Michael; Sawaya, Michael R.; Wang, Shuishu; Phillips, Martin; Cascio, Duilio; Eisenberg, David (2006): Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. In *Proceedings of the National Academy of Sciences of the United States of America* 103 (21), pp. 8060–8065. DOI: 10.1073/pnas.0602606103.

Studier, F.William; Moffatt, Barbara A. (1986): Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. In *Journal of Molecular Biology* 189 (1), pp. 113–130. DOI: 10.1016/0022-2836(86)90385-2.

Suh, Sang-Jin; Runyen-Janecky, Laura J.; Maleniak, Tricia C.; Hager, Paul; MacGregor, Carolyn H.; Zielinski-Mozny, Nicolette A. et al. (2002): Effect of vfr mutation on global gene expression and catabolite repression control of Pseudomonas aeruginosa. In *Microbiology (Reading, England)* 148 (Pt 5), pp. 1561–1569. DOI: 10.1099/00221287-148-5-1561.

Takahashi, Masayuki; Blazy, Bernadette; Baudras, Alain; Hillen, Wolfgang (1989): Ligand-modulated binding of a gene regulatory protein to DNA. In *Journal of Molecular Biology* 207 (4), pp. 783–796. DOI: 10.1016/0022-2836(89)90244-1.

Tao, Mariano; Huberman, Alberto (1970): Some properties of Escherichia coli adenyl cyclase. In *Archives of biochemistry and biophysics* 141 (1), pp. 236–240. DOI: 10.1016/0003-9861(70)90127-X.

Tari, Leslie W. (Ed.) (2012): Structure-based drug discovery. New York: Humanan Pr (Springer protocols, 841).

The CCP4 suite: programs for protein crystallography (1994). In *Acta crystallographica. Section D, Biological crystallography* 50 (Pt 5), pp. 760–763.

Tian, Chang Fu; Garnerone, Anne-Marie; Mathieu-Demazière, Céline; Masson-Boivin, Catherine; Batut, Jacques (2012): Plant-activated bacterial receptor adenylate cyclases modulate epidermal infection in the Sinorhizobium meliloti-Medicago symbiosis. In *Proceedings of the National Academy of Sciences of the United States of America* 109 (17), pp. 6751–6756. DOI: 10.1073/pnas.1120260109.

Tovar, Carmel N.; Odunuga, Odutayo O. (2019): Size of Protein is a Major Factor that Affects Retention on Preparative IMAC Columns. In *The protein journal* 38 (1), pp. 76–82. DOI: 10.1007/s10930-018-9803-9.

Towbin, H.; Staehelin, T.; Gordon, J. (1979): Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. In *Proceedings of the National Academy of Sciences of the United States of America* 76 (9), pp. 4350–4354. DOI: 10.1073/pnas.76.9.4350.

Townsend, Philip D.; Jungwirth, Britta; Pojer, Florence; Bußmann, Michael; Money, Victoria A.; Cole, Stewart T. et al. (2014): The crystal structures of apo and cAMP-bound GlxR from Corynebacterium glutamicum reveal structural and dynamic changes upon cAMP binding in CRP/FNR family transcription factors. In *PloS one* 9 (12), e113265. DOI: 10.1371/journal.pone.0113265.

Trülzsch, K.; Roggenkamp, A.; Pelludat, C.; Rakin, A.; Jacobi, C.; Heesemann, J. (2001): Cloning and characterization of the gene encoding periplasmic 2',3'-cyclic phosphodiesterase of Yersinia enterocolitica O:8. In *Microbiology (Reading, England)* 147 (Pt 1), pp. 203–213. DOI: 10.1099/00221287-147-1-203.

Tsai, Emily J.; Kass, David A. (2009): Cyclic GMP signaling in cardiovascular pathophysiology and therapeutics. In *Pharmacology & therapeutics* 122 (3), pp. 216–238. DOI: 10.1016/j.pharmthera.2009.02.009.

Tzeng, Shiou-Ru; Kalodimos, Charalampos G. (2013): Allosteric inhibition through suppression of transient conformational states. In *Nature chemical biology* 9 (7), pp. 462–465. DOI: 10.1038/nchembio.1250.

UCLA: Aniso Scale. Available online at http://services.mbi.ucla.edu/anisoscale/anisoscale_xds/.

van Hove, B.; Staudenmaier, H.; Braun, V. (1990): Novel two-component transmembrane transcription control: regulation of iron dicitrate transport in Escherichia coli K-12. In *Journal of bacteriology* 172 (12), pp. 6749–6758. DOI: 10.1128/jb.172.12.6749-6758.1990.

Velázquez-Campoy, Adrián; Ohtaka, Hiroyasu; Nezami, Azin; Muzammil, Salman; Freire, Ernesto (2004): Isothermal titration calorimetry. In *Current protocols in cell biology* Chapter 17, Unit 17.8. DOI: 10.1002/0471143030.cb1708s23.

Veltman, Douwe M.; Bosgraaf, Leonard; van Haastert, Peter J.M (2004): Unusual Guanylyl Cyclases and cGMP Signaling in Dictyostelium discoideum. In Gerald Litwack (Ed.): Vitamins and Hormones, vol. 69. 1. Aufl. s.l.: Elsevier textbooks (Vitamins and Hormones, v.69), pp. 95–115.

Verollet, Romain (2008): A major step towards efficient sample preparation with bead-beating. In *BioTechniques* 44 (6), pp. 832–833. DOI: 10.2144/000112893.

Via, Virginia Dalla; Zanetti, María Eugenia; Blanco, Flavio (2016): How legumes recognize rhizobia. In *Plant signaling & behavior* 11 (2), e1120396. DOI: 10.1080/15592324.2015.1120396.

Volbeda, Anne; Darnault, Claudine; Renoux, Oriane; Nicolet, Yvain; Fontecilla-Camps, Juan C. (2015): The crystal structure of the global anaerobic transcriptional regulator FNR explains its extremely fine-tuned monomer-dimer equilibrium. In *Science advances* 1 (11), e1501086. DOI: 10.1126/sciadv.1501086.

Waldron, Kevin J.; Robinson, Nigel J. (2009): How do bacterial cells ensure that metalloproteins get the correct metal? In *Nature reviews. Microbiology* 7 (1), pp. 25–35. DOI: 10.1038/nrmicro2057.

Wang, Dong; Yang, Shengming; Tang, Fang; Zhu, Hongyan (2012): Symbiosis specificity in the legume: rhizobial mutualism. In *Cellular microbiology* 14 (3), pp. 334–342. DOI: 10.1111/j.1462-5822.2011.01736.x.

Waterhouse, Andrew; Bertoni, Martino; Bienert, Stefan; Studer, Gabriel; Tauriello, Gerardo; Gumienny, Rafal et al. (2018): SWISS-MODEL: homology modelling of protein structures and complexes. In *Nucleic acids research* 46 (W1), W296-W303. DOI: 10.1093/nar/gky427.

Weber, Irene T.; Steitz, Thomas A. (1987): Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution. In *Journal of Molecular Biology* 198 (2), pp. 311–326. DOI: 10.1016/0022-2836(87)90315-9.

Weis, David D. (Ed.) (2016): Hydrogen exchange mass spectrometry of proteins. Fundamentals, methods, and applications. Chichester, West Sussex: John Wiley & Sons Inc. Available online at http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=1146891.

West, S. E.; Sample, A. K.; Runyen-Janecky, L. J. (1994): The vfr gene product, required for Pseudomonas aeruginosa exotoxin A and protease production, belongs to the cyclic AMP receptor protein family. In *Journal of bacteriology* 176 (24), pp. 7532–7542. DOI: 10.1128/jb.176.24.7532-7542.1994.

Whitaker, J. R. (1963): Determination of Molecular Weights of Proteins by Gel Filtration of Sephadex. In *Anal. Chem.* 35 (12), pp. 1950–1953. DOI: 10.1021/ac60205a048.

Williams, A.; Frasca, V. (2001): Ion-exchange chromatography. In *Current protocols in protein science* Chapter 8, Unit8.2. DOI: 10.1002/0471140864.ps0802s15.

Williams, Christopher J.; Headd, Jeffrey J.; Moriarty, Nigel W.; Prisant, Michael G.; Videau, Lizbeth L.; Deis, Lindsay N. et al. (2018): MolProbity: More and better reference data for improved all-atom structure validation. In *Protein science : a publication of the Protein Society* 27 (1), pp. 293–315. DOI: 10.1002/pro.3330.

Wilson, A. J. C. (1950): Largest likely values for the reliability index. In *Acta Cryst* 3 (5), pp. 397–398. DOI: 10.1107/S0365110X50001129.

Wilson, Curtis M. (1979): Studies and critique of amido black 10B, Coomassie blue R, and fast green FCF as stains for proteins after polyacrylamide gel electrophoresis. In *Analytical biochemistry* 96 (2), pp. 263–278. DOI: 10.1016/0003-2697(79)90581-5.

Winzor, Donald J. (2003): Analytical exclusion chromatography. In *Journal of Biochemical and Biophysical Methods* 56 (1-3), pp. 15–52. DOI: 10.1016/S0165-022X(03)00071-X.

Wiseman, Thomas; Williston, Samuel; Brandts, John F.; Lin, Lung-Nan (1989): Rapid measurement of binding constants and heats of binding using a new titration calorimeter. In *Analytical biochemistry* 179 (1), pp. 131–137. DOI: 10.1016/0003-2697(89)90213-3.

Wojdyla, Justyna Aleksandra; Kaminski, Jakub W.; Panepucci, Ezequiel; Ebner, Simon; Wang, Xiaoqiang; Gabadinho, Jose; Wang, Meitian (2018): DA+ data acquisition and analysis software at the Swiss Light Source macromolecular crystallography beamlines. In *Journal of synchrotron radiation* 25 (Pt 1), pp. 293–303. DOI: 10.1107/S1600577517014503.

Won, Hyung-Sik; Lee, Yoo-Sup; Lee, Sung-Hee; Lee, Bong-Jin (2009): Structural overview on the allosteric activation of cyclic AMP receptor protein. In *Biochimica et biophysica acta* 1794 (9), pp. 1299–1308. DOI: 10.1016/j.bbapap.2009.04.015.

Xiong, Ying; Lu, Hai-Ting; Li, Yongjian; Yang, Guang-Fu; Zhan, Chang-Guo (2006): Characterization of a catalytic ligand bridging metal ions in phosphodiesterases 4 and 5 by molecular dynamics simulations and hybrid quantum mechanical/molecular mechanical calculations. In *Biophysical Journal* 91 (5), pp. 1858–1867. DOI: 10.1529/biophysj.106.086835.

Xu, Guohua; Fan, Xiaorong; Miller, Anthony J. (2012): Plant nitrogen assimilation and use efficiency. In *Annual review of plant biology* 63, pp. 153–182. DOI: 10.1146/annurev-arplant-042811-105532.

Yakovchuk, Peter; Protozanova, Ekaterina; Frank-Kamenetskii, Maxim D. (2006): Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. In *Nucleic acids research* 34 (2), pp. 564–574. DOI: 10.1093/nar/gkj454.

Youn, Hwan; Kerby, Robert L.; Conrad, Mary; Roberts, Gary P. (2006): Study of highly constitutively active mutants suggests how cAMP activates cAMP receptor protein. In *The Journal of biological chemistry* 281 (2), pp. 1119–1127. DOI: 10.1074/jbc.M509421200.

Youn, Hwan; Kerby, Robert L.; Koh, Junseock; Roberts, Gary P. (2007): A C-helix residue, Arg-123, has important roles in both the active and inactive forms of the cAMP receptor protein. In *The Journal of biological chemistry* 282 (6), pp. 3632–3639. DOI: 10.1074/jbc.M606602200.

Youn, Hwan; Koh, Junseock; Roberts, Gary P. (2008): Two-state allosteric modeling suggests protein equilibrium as an integral component for cyclic AMP (cAMP) specificity in the cAMP receptor protein of Escherichia coli. In *Journal of bacteriology* 190 (13), pp. 4532–4540. DOI: 10.1128/JB.00074-08.

Yu, Shaoning; Maillard, Rodrigo A.; Gribenko, Alexey V.; Lee, J. Ching (2012): The N-terminal capping propensities of the D-helix modulate the allosteric activation of the Escherichia coli cAMP receptor protein. In *The Journal of biological chemistry* 287 (47), pp. 39402–39411. DOI: 10.1074/jbc.M112.404806.

Zaveri, Anisha; Visweswariah, Sandhya (2013): Cyclic AMP in Mycobacteria: The second messenger comes first. In *Current Science* 105, pp. 666–675.

Zha, Shan; Boboila, Cristian; Alt, Frederick W. (2009): Mre11: roles in DNA repair beyond homologous recombination. In *Nature structural & molecular biology* 16 (8), pp. 798–800. DOI: 10.1038/nsmb0809-798.

Zheng, Zhifang; Zhu, Mengzhu; He, Ying; Li, Nan; Guo, Ting; Chen, Yong et al. (2013): Gene cloning, expression, and characterization of a cyclic nucleotide phosphodiesterase from

Arthrobacter sp. CGMCC 3584. In *Applied biochemistry and biotechnology* 169 (8), pp. 2442–2456. DOI: 10.1007/s12010-013-0136-z.

Zhulin, Igor B.; Nikolskaya, Anastasia N.; Galperin, Michael Y. (2003): Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea. In *Journal of bacteriology* 185 (1), pp. 285–294. DOI: 10.1128/jb.185.1.285-294.2003.

Zou, Lan; Gastebois, Amandine; Mathieu-Demazière, Céline; Sorroche, Fernando; Masson-Boivin, Catherine; Batut, Jacques; Garnerone, Anne-Marie (2017): Transcriptomic Insight in the Control of Legume Root Secondary Infection by the Sinorhizobium meliloti Transcriptional Regulator Clr. In *Frontiers in microbiology* 8, p. 1236. DOI: 10.3389/fmicb.2017.01236.

8 Appendix

8.1 Testexpressions

Test expression of Clr dN from S. americanum, S. Sp. A49 and S. saheli

N-terminally truncated Clr homologs from *Sinorhizobium americanum* (dN11), *Sinorhizobium Species A49* (dN12) and *Sinorhizobium saheli* (dN11) in pET 28a were purchased from *BioCat*. A clone of BL21 (DE3) Gold cells containing these constructs was grown in 5 mL LB Kanamycin (35 µg/mL) medium overnight at 37°C. Subsequently, 4 mL overnight culture were added to 50 mL LB Kanamycin, grown to an OD_{600} of 0.6 and induced with 0.1 mM IPTG while reducing the temperature to 30°C. Three hours later, the cells were harvested via centrifugation (4000 rpm, 15 min, F6S 6x1000Y)) and resuspended in 2 mL Ni-NTA buffer A (50 mM NaH_2PO4, 1 M NaCl, 10 mM imidazole, pH 7.0). 0.5 g of glass beads (0.2-0.5 mm) were added to 1 mL of the slurry and subjected to 3 rounds of 30s in a benchtop homogenizer, with cooling on ice for 5 min in between. The lysate was spun at 14.800 rpm at 4°C for 1h and supernatant and pellet were analyses via *SDS PAGE*. All constructs were expressed nicely and soluble under the chosen expression conditions.

Test expression of native phosphodiesterase constructs and Expression of pQE9 His_6 *Sm*CpdA (native)

Constructs coding for multiple phosphodiesterase homologs were kindly provided by Neda Farmani. BL21 (DE3) Gold cells transformed with these constructs were grown in 5 mL LB Ampicillin (100 µg/mL) medium overnight at 37°C. 4 mL overnight culture were added to 50 mL LB Ampicillin, grown to an OD_{600} of 0.5 and induced with 0.1 mM IPTG switching the temperature to 30°C. The cells were harvested after 4 h via centrifugation (4000 rpm, 15 min, F6S 6x1000Y)) and resuspended in 2 mL Ni-NTA buffer A (50 mM NaH_2PO4, 300 mM NaCl, 10 mM imidazole, pH 7.8). After addition of 500 mg 0.2-0.5 mm glass beads to 1 mL suspension the cells were lysed using 3 rounds of 30s in a benchtop homogenizer, with cooling on ice for 5 min in between. The lysate was spun at 14.800 rpm at 4°C for 1h and supernatant and pellet were analyses via *SDS PAGE*. About half of the expressed protein could be found in the supernatant for CpdA from both *E. coli* and *S. meliloti*, the others proved to be well expressed but insoluble. Subsequent expressions in both Bl21 (DE3) Gold and Shuffle T7 Express at 12°C for 72 hours induced with 0.05 mM IPTG yielded slightly improved solubility for the remaining homologs.

Organism	**Vector**	**Gene**

E. coli	pQE9	CpdA
S. meliloti	pQE9	CpdA (SpdA in some sources)
S. meliloti	pWH844	*Sm*c01798
S. meliloti	pWH844	*Sm*c03964
S. meliloti	pWH844	*Sm*c2712
S. meliloti	pQE9	*Sm*b20081

The production of *Sm*CpdA was scaled up in response to the successful testing. Freshly transformed *E. coli* Bl21 (DE3) Gold was expressed at 30°C in cultures of 2 L LB Ampicillin (100 µg/mL) according to the procedure described for the test expression. The cells were resuspended in 30 mL Ni-NTA buffer A (50 mM NaH_2PO4, 300 mM NaCl, 10 mM imidazole, pH 7.8) and flash frozen in liquid nitrogen. The cells were stored at -80°C.

Cells corresponding to 1.6 L culture were thawed and treated with 50 µL Lysozyme (50 mg/mL) and PMSF (200 mM) for 15 min, then lysed by three passes through a cold French press pressure cell at 15.2 bar cm^{-2}. After centrifugation at 18000 rpm (JA-20) for 1 h the supernatant was removed to a new flask and filtered at 0.45 nm using a syringe filter. Filtered lysate was applied to a 5 mL Ni-NTA column using a peristaltic pump and eluted at a flow rate of 2 mL/min using a stepwise gradient in imidazole concentration on a FLPC system. The elution was monitored using UV/VIS. Size exclusion chromatography was used as a polishing step. The resulting sample was quite yellow and showed three distinct bands in a 12% SDS-PAGE gel. The impurities could not be removed and the construct was abandoned in favour of its codon-optimized counterpart, which offers much higher solubility and yield and can be readily purified.

Test expression of solubility tag fused phosphodiesterase constructs

E. coli CpdA as well as *S. meliloti Sm*c1798, *Sm*c2712, *Sm*c3964 and *Sm*b20081 were cloned in solubility enhancing tag LIC cleavable plasmids from *Addgene*. The tags included were maltose binding protein (MBP), monomeric Ocr (Mocr) and glutathione S-transferase (GST). 4 mL overnight culture were added to 50 mL LB Ampicillin, grown at 37°C to an OD$_{600}$ of 0.5 and induced with 0.1 mM IPTG at 30°C. The cells were harvested after 4 h via centrifugation (4000 rpm, 15 min, F6S 6x1000Y)) and resuspended in 2 mL Ni-NTA buffer A (50 mM NaH_2PO4, 300 mM NaCl, 10 mM imidazole, pH 7.8). After addition of 500 mg 0.2-0.5 mm glass beads to 1 mL suspension the cells were lysed using 3 rounds of 30s in a benchtop homogenizer, with cooling on ice for 5 min in between. The lysate was spun at 14.800 rpm at 4°C for 1h and supernatant and pellet were analyses via *SDS PAGE*. *Sm*c1798 and *Sm*c3964 seemed to improve in solubility when fused to Mocr, CpdA was soluble in all cases whereas *Sm*c2712 and *Sm*b20081 did not improve much.

Gene	**MBP**	**Mocr**	**GST**
*Ec*CpdA	soluble	mainly soluble	about half in supernatant
*Sm*c01798	some soluble	mainly soluble	insoluble
*Sm*c03964	insoluble	soluble	insoluble
*Sm*c2712	insoluble	insoluble	insoluble
*Sm*b20081	insoluble	cloned but untested	insoluble

8.2 Clr-DNA affinity measurement via fluorescence anisotropy

A simple method for characterising the interaction of proteins with each other or larger ligands is measuring the fluorescence anisotropy of a fluorescently labelled complex component. In our case one strand of the duplex DNA was either purchased from *IBA Lifesciences* with a 5-carboxyfluorescein label at the 5'-end in case of the 19-bp oligomer used in the crystallisation experiments, or the PCR product was fluorescently labelled in the same position. The complementing strand was annealed via gradual cooling from 98°C as described in 4.1.2 Clr affinity for cGMP and cAMP levels out in the presence of DNA.

Name	**Length [bp]**	**Features**
DNAsym	19/14	Synthetic oligo used for crystallisation, 19-mer 5'-labelled
Smc04190-FAM1	100	5'-labelled sense strand, gene Smc04190
Smc04190-FAM2	261	5'-labelled sense strand, gene Smc04190
Smc04190-FAM3	150	5'-labelled antisense strand, gene Smc04190
Smc04190-FAM4	100	5'-labelled antisense strand, gene Smc04190
Smc04190-FAM5	200	5'-labelled antisense strand, gene Smc04190
Smb20906-FAM1	100	5'-labelled sense strand, gene Smb20906
Smb20906-FAM2	284	5'-labelled sense strand, gene Smb20906
Smb20906-FAM3	140	5'-labelled antisense strand, gene Smb20906
Smb20906-FAM4	100	5'-labelled antisense strand, gene Smb20906
Smb20906-FAM5	185	5'-labelled antisense strand, gene Smb20906

The DNA was used in a concentration of 5 nM and incubated with 250 µM cAMP and $MgCl_2$. To this, varying concentrations of protein (0.1 nM, 1 nM, 10 nM, 100 nM, 1 µM, 10 µM) were added and the fluorescence polarisation was measured perpendicular and parallel to the excitation was measured using a *Tecan Spark 20M*. Excitation of the label was achieved at 485 nm and the resulting fluorescence measured at 535 nm. Plotting the logarithm of the protein concentration against the measured anisotropy should result in a sigmoidal binding curve in case of binding interactions. No increase in polarisation could be observed in our measurements (Figure 73 and Figure 74). This is likely due to the fact that the method actually measures fluorophore movement in solution. As most DNA fragments are fairly large and linear their flexibility might not be impaired enough to be detected in this method. The opposite is probably true for the DNAsym fragment, in which the label might interfere with binding. Fluorescence polarisation is therefore only suitable for midrange DNA lengths, a condition which our targets did not meet.

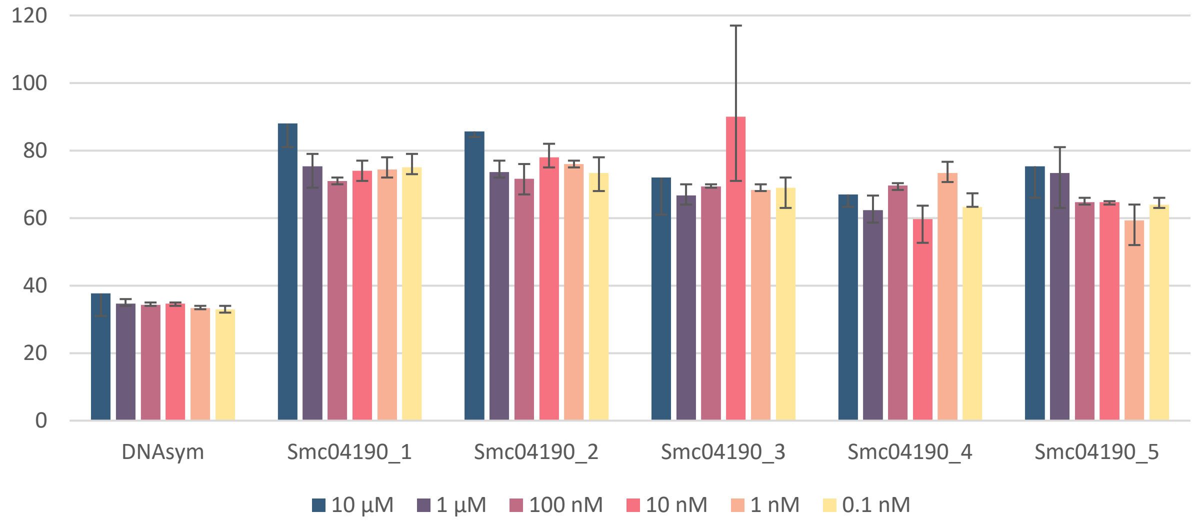


Figure 73: Fluorescence anisotropy of Clr interaction with the synthetic duplex DNA used in crystallisation and fragments based on gene Smc04190. The measurements were conducted in triplicates with deviations indicated in the error bars. 5 nM duplex DNA was incubated with 0.25 mM cAMP and $MgCl_2$ and measured with varying protein concentrations. No concentration-dependent increase in anisotropy could be observed for any fragment length and label position. The total anisotropy values for the fragments based on Smc04190 are higher than those for DNAsym, as they are much larger and their movement in solution much slower in consequence.

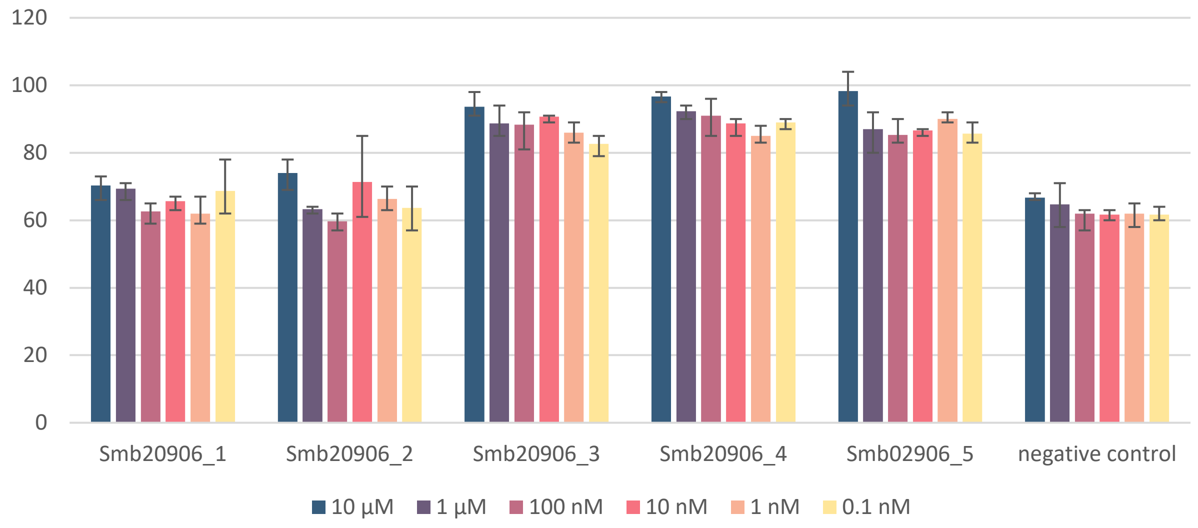
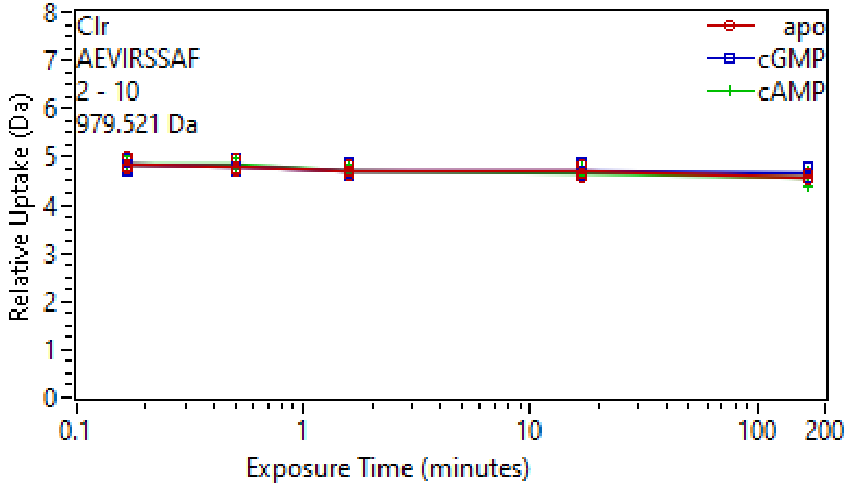
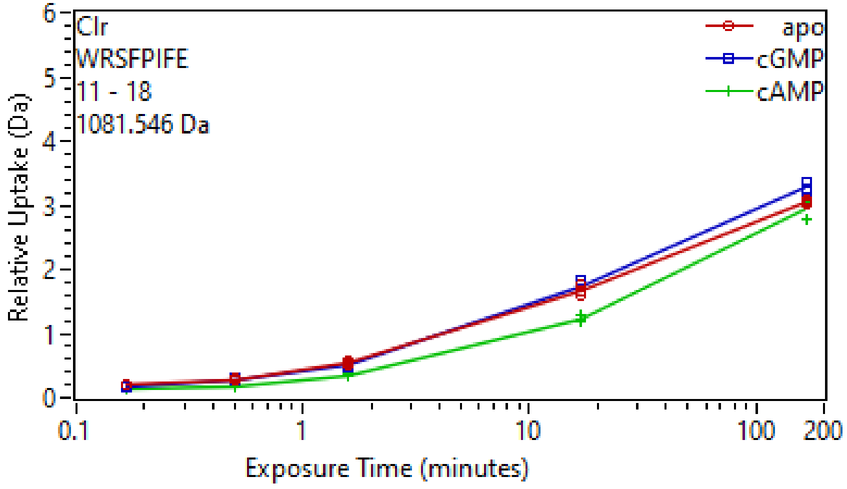
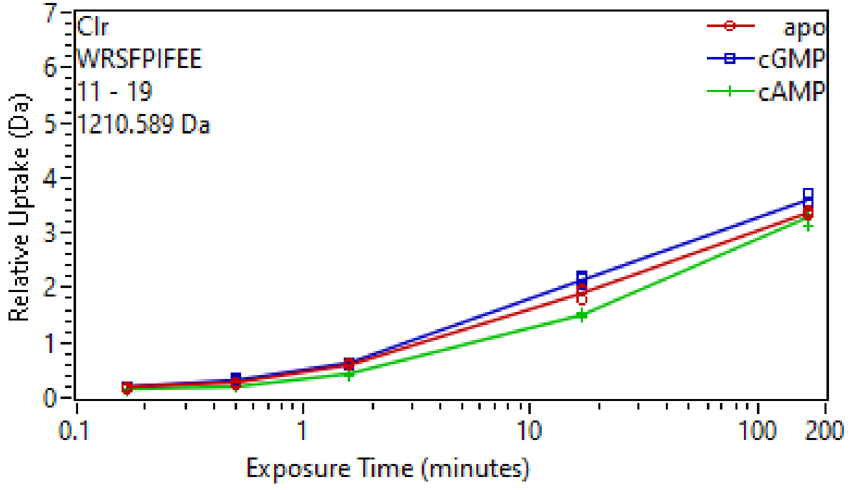


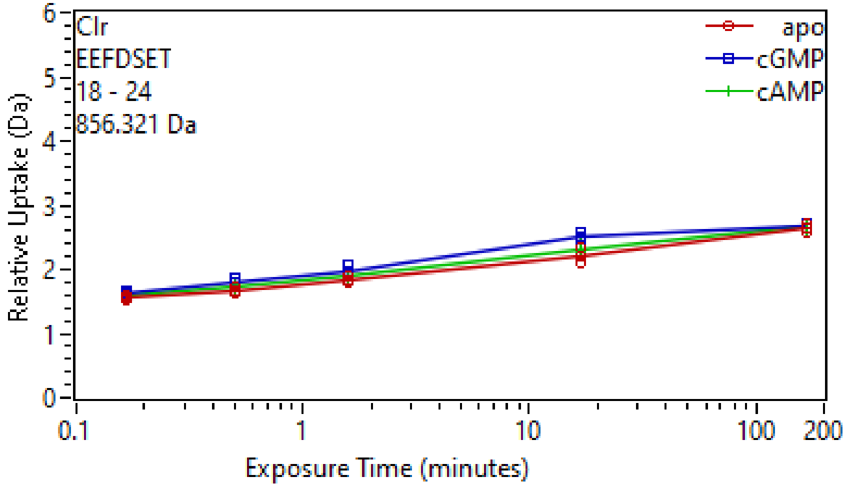
Figure 74: Fluorescence anisotropy measurements in the presence and absence of Clr for fluorescently labelled DNA fragments based on Smb20906. For the triplicate measurements of Smb20906-based DNA fragments with Clr, no interaction could be observed as well. For measurements in the absence of Clr the anisotropy is in the same range as in its presence, further supporting the notion that this method is simply not suitable for the DNA sizes we used.

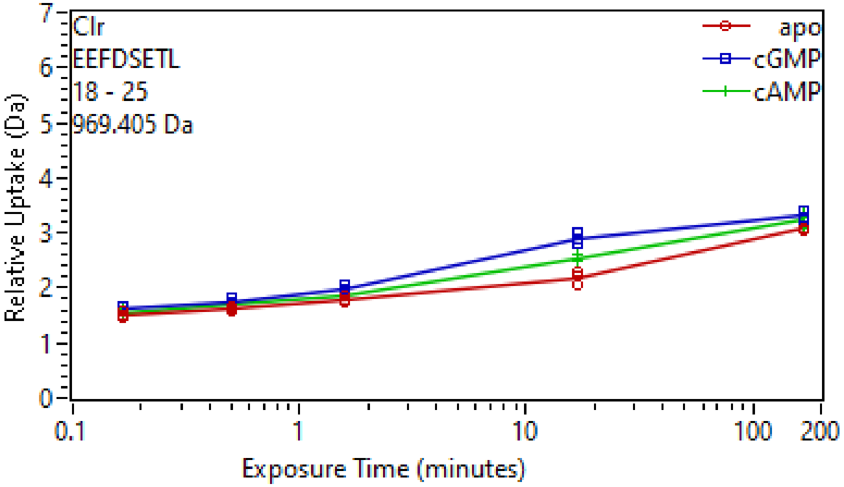
8.3 HDX-MS analysis of Clr substrate binding

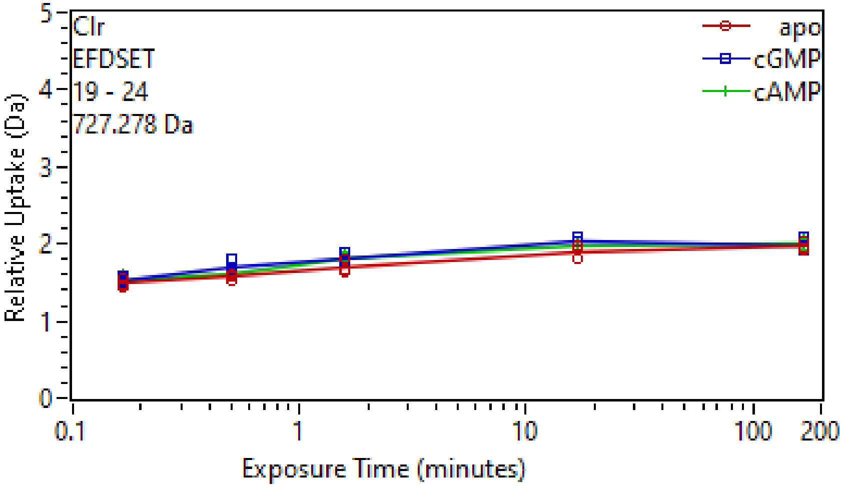


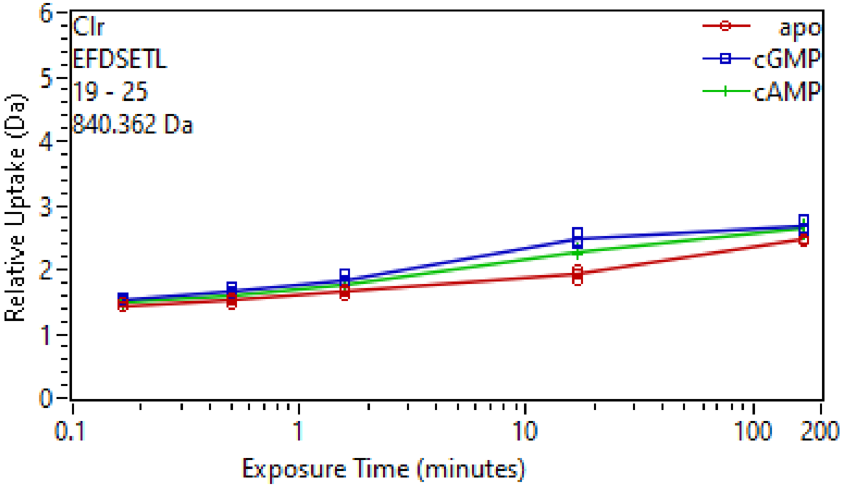


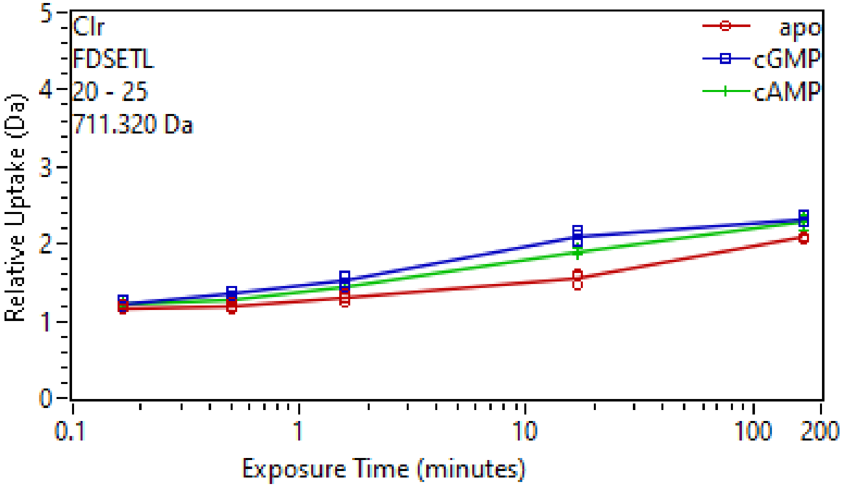


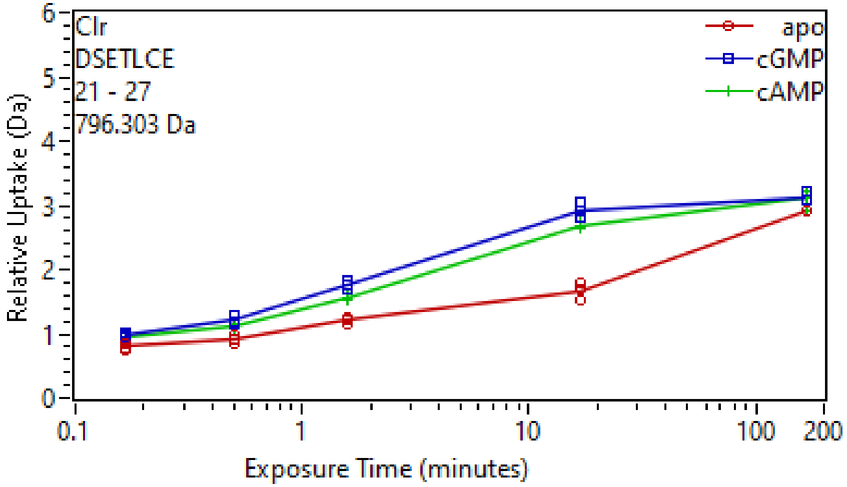


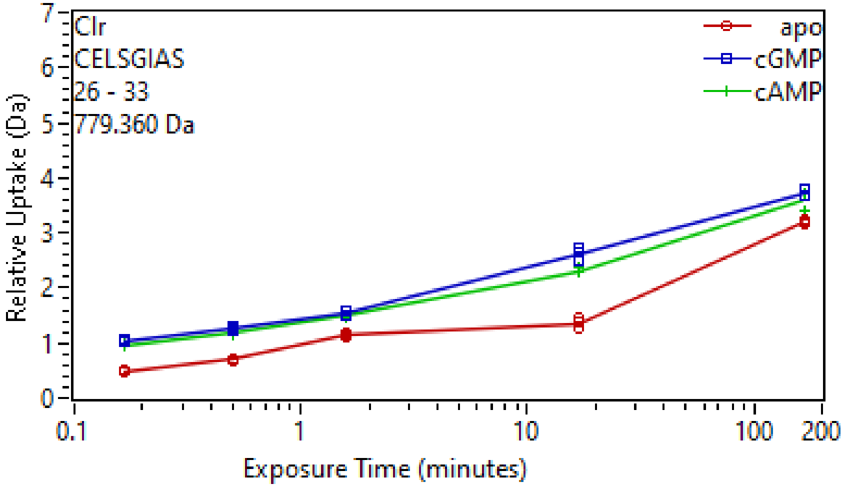

ClR
EEFDSETL
18 - 25
969.405 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
0.1
1
10
100
200
Exposure Time (minutes)

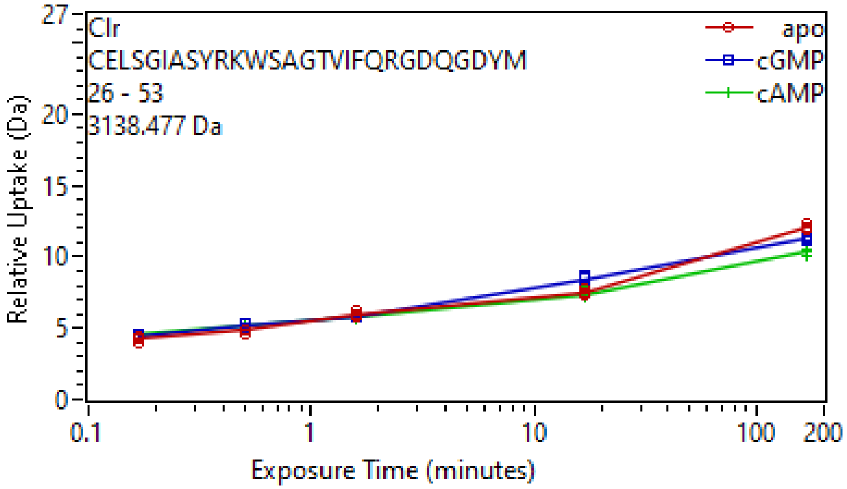

ClR
EFDSET
19 - 24
727.278 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
0.1
1
10
100
200
Exposure Time (minutes)

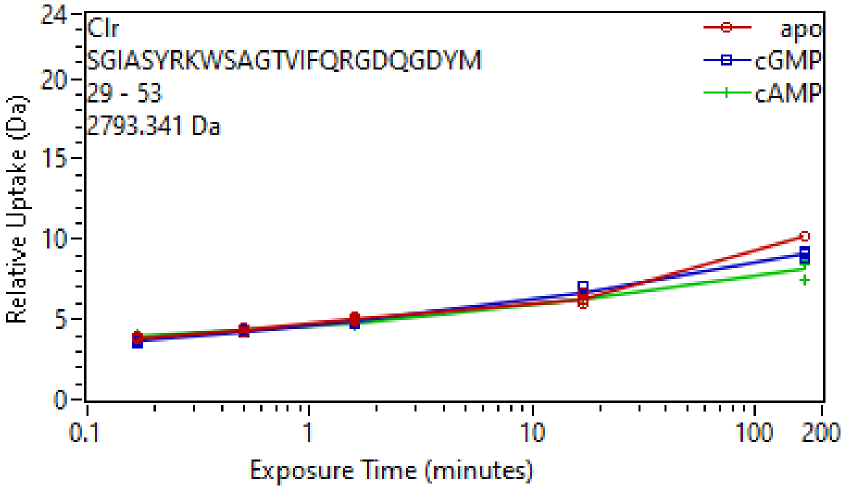

ClR
EFDSETL
19 - 25
840.362 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
0.1
1
10
100
200
Exposure Time (minutes)

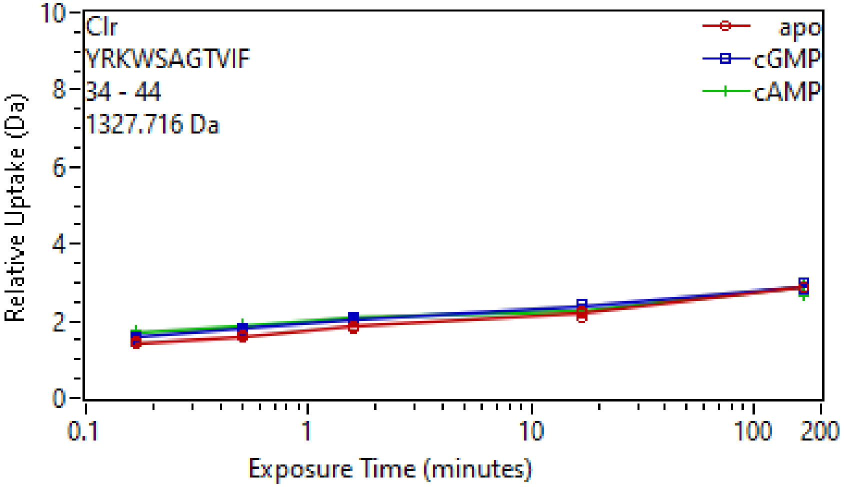

ClR
FDSETL
20 - 25
711.320 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
0.1
1
10
100
200
Exposure Time (minutes)

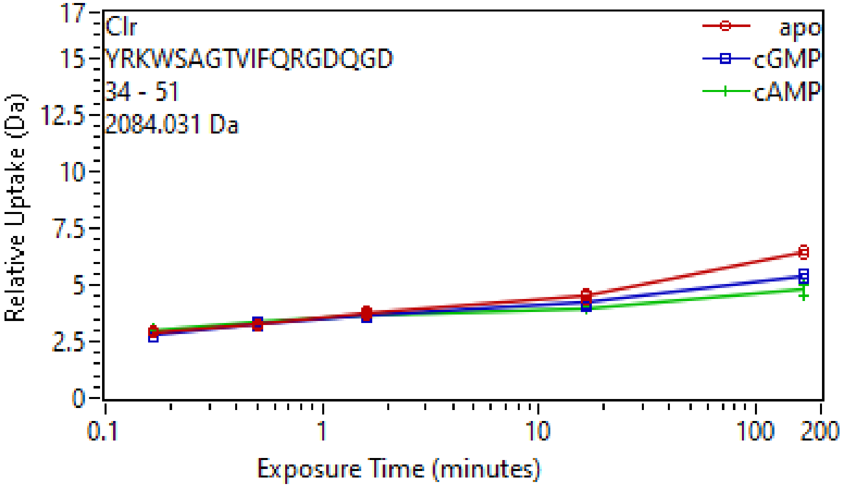

Clr
DSETLCE
21 - 27
796.303 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
0.1
1
10
100
200
Exposure Time (minutes)

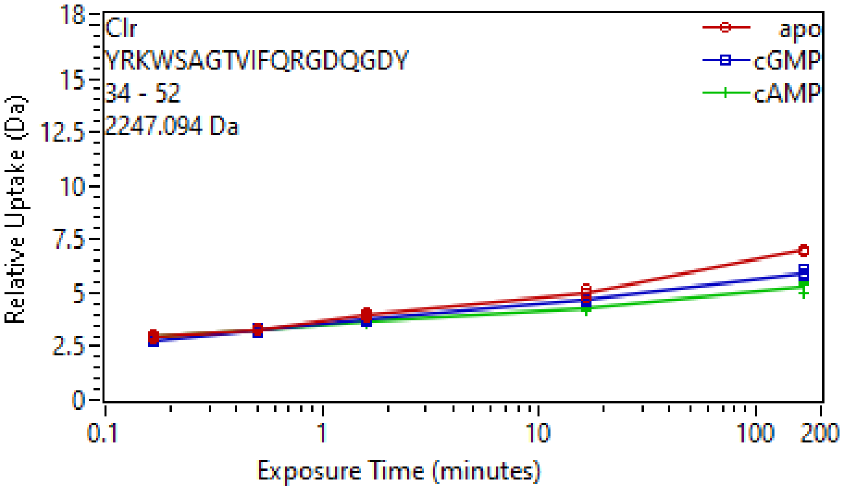

Clr
CELSGIAS
26 - 33
779.360 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
0.1
1
10
100
200
Exposure Time (minutes)

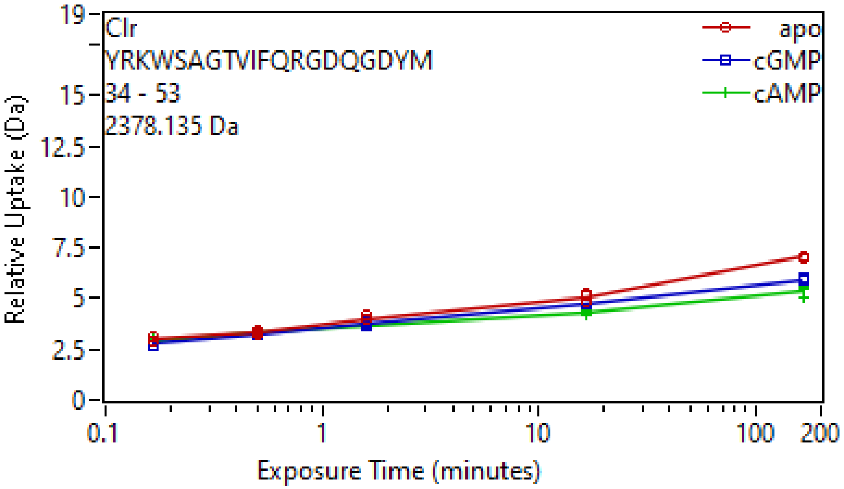

Clr
CELSGIASYRKWSAGTVIFQRGDQGDYM
26 - 53
3138.477 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
5
10
15
20
27
0.1
1
10
100
200
Exposure Time (minutes)

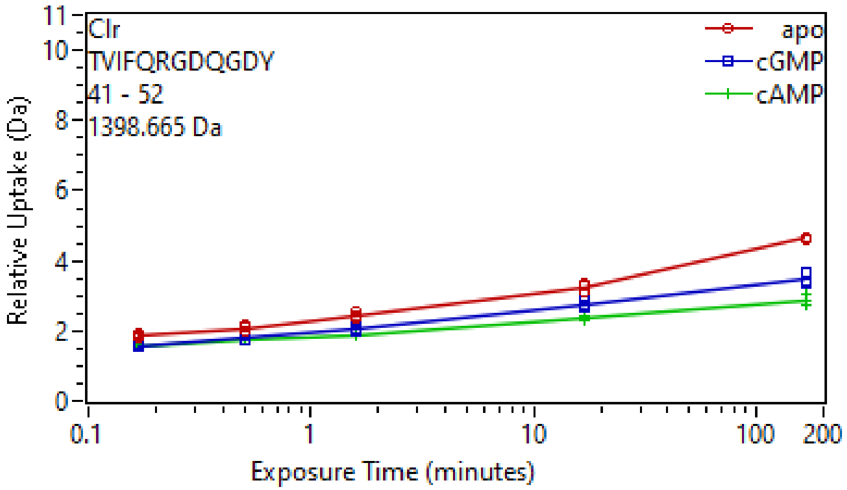

Clr
SGIASYRKWSAGTVIFQRGDQGDYM
29 - 53
2793.341 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
5
10
15
20
24
0.1
1
10
100
200
Exposure Time (minutes)

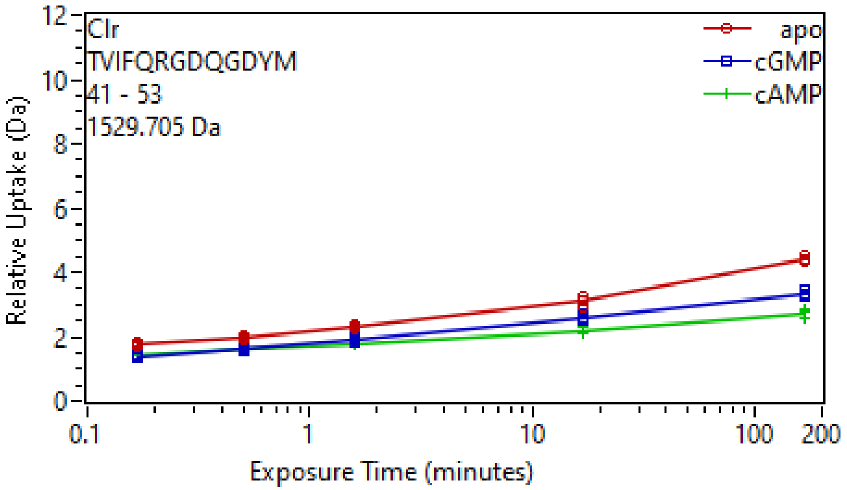

ClrYRKWSAGTVIF
34 - 44
1327.716 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
0.1
1
10
100
200
Exposure Time (minutes)

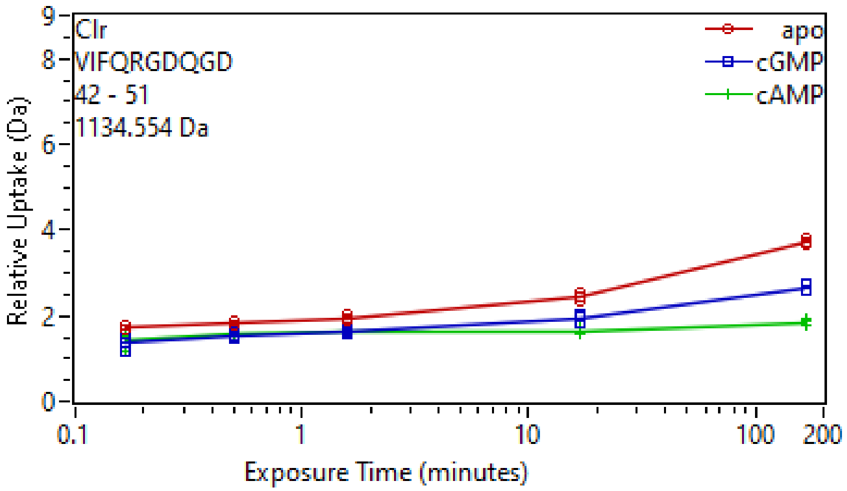

Clr
YRKWSAGTVIFQRGDQGD
34 - 51
2084.031 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
17
0.1
1
10
100
200
Exposure Time (minutes)

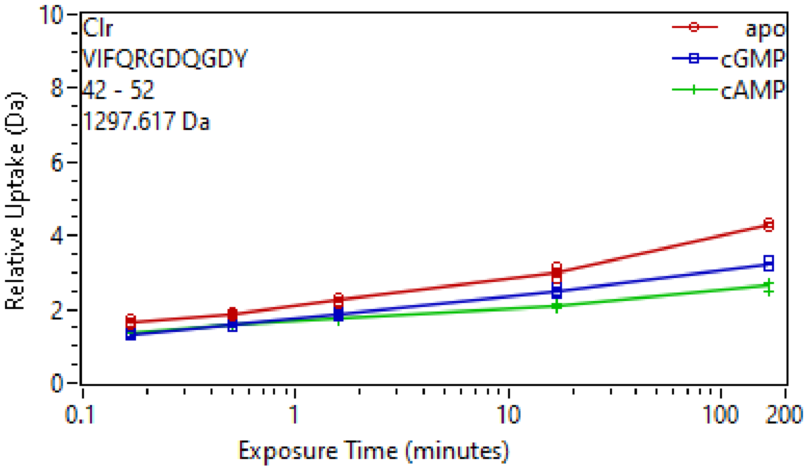

Clr
YRKWSAGTVIFQRGDQGDY
34 - 52
2247.094 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
18
0.1
1
10
100
200
Exposure Time (minutes)

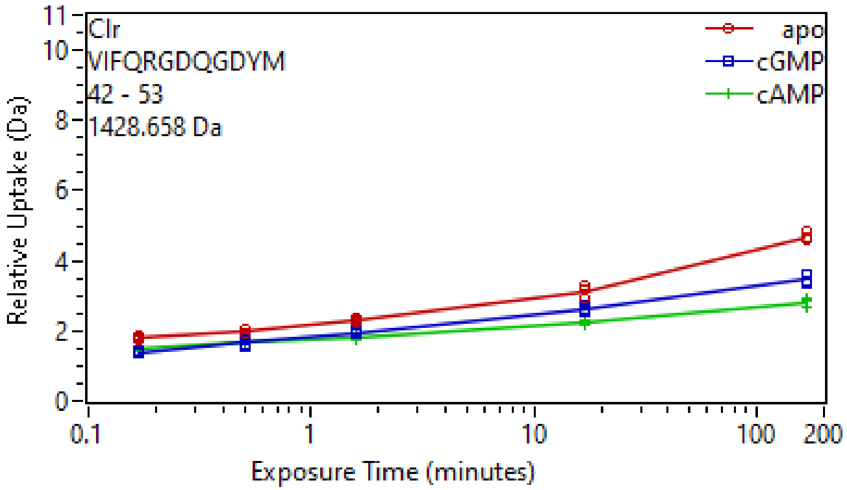

Clr
YRKWSAGTVIFQRGDQGDYM
34 - 53
2378.135 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
19
0.1
1
10
100
200
Exposure Time (minutes)

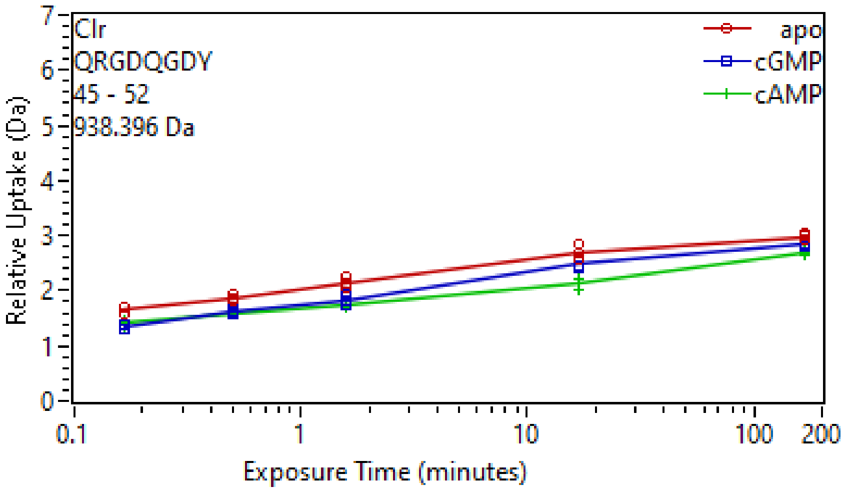

Clr
TVIFQRGDQGDY
41 - 52
1398.665 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
11
Exposure Time (minutes)
0.1
1
10
100
200

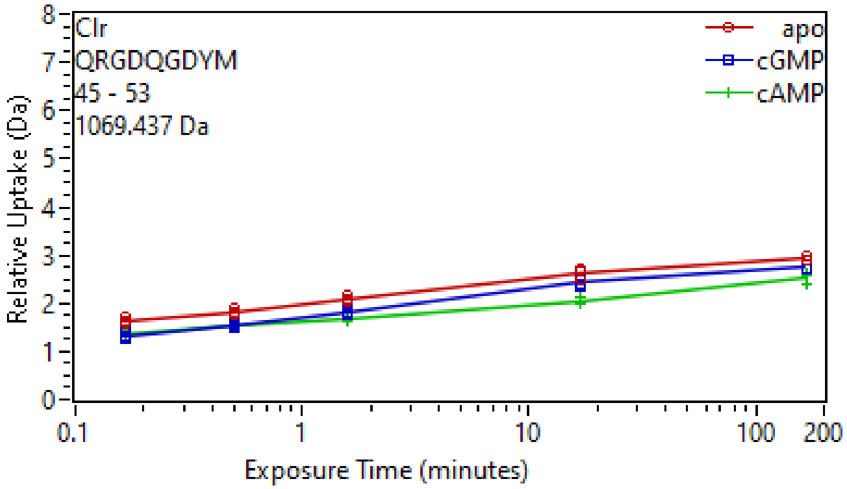

Clr
TVIFQRGDQGDYM
41 - 53
1529.705 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
Exposure Time (minutes)
0.1
1
10
100
200

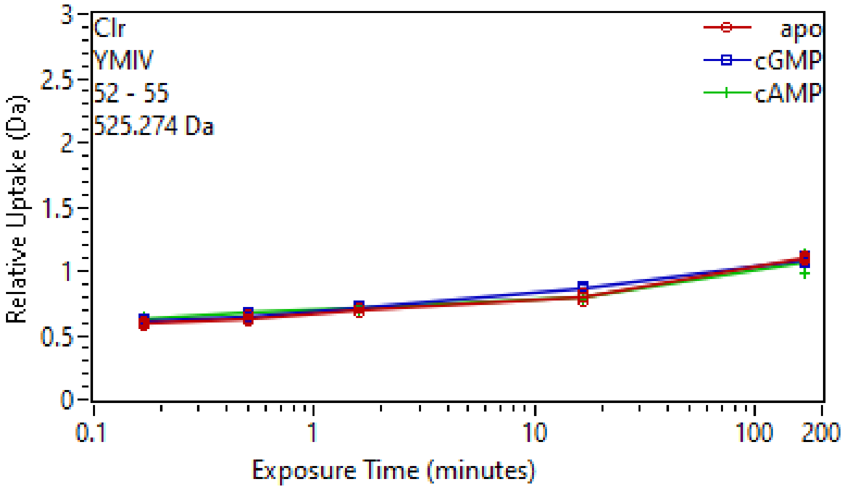

Clr
VIFQRGDQGD
42 - 51
1134.554 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
9
Exposure Time (minutes)
0.1
1
10
100
200

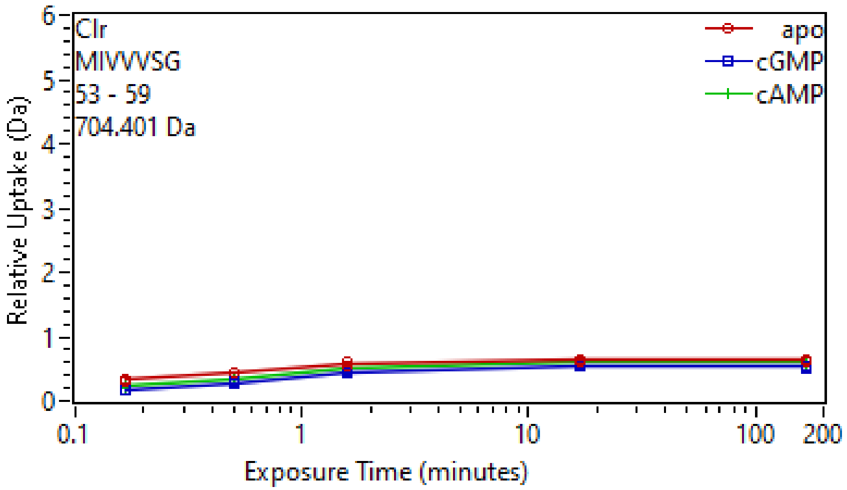

Clr
VIFQRGDQGDY
42 - 52
1297.617 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
Exposure Time (minutes)
0.1
1
10
100
200

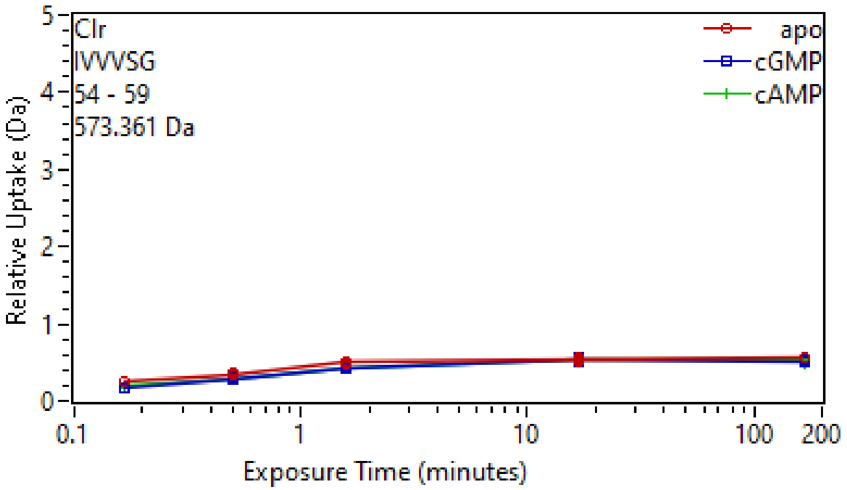

ClR
VIFQRGDQGDYM
42 - 53
1428.658 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
11
0.1
1
10
100
200
Exposure Time (minutes)

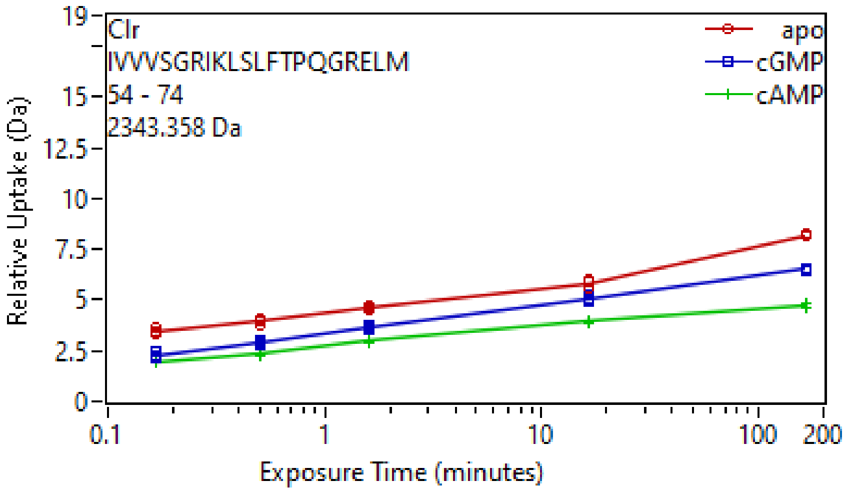

ClR
QRGDQGDY
45 - 52
938.396 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
0.1
1
10
100
200
Exposure Time (minutes)

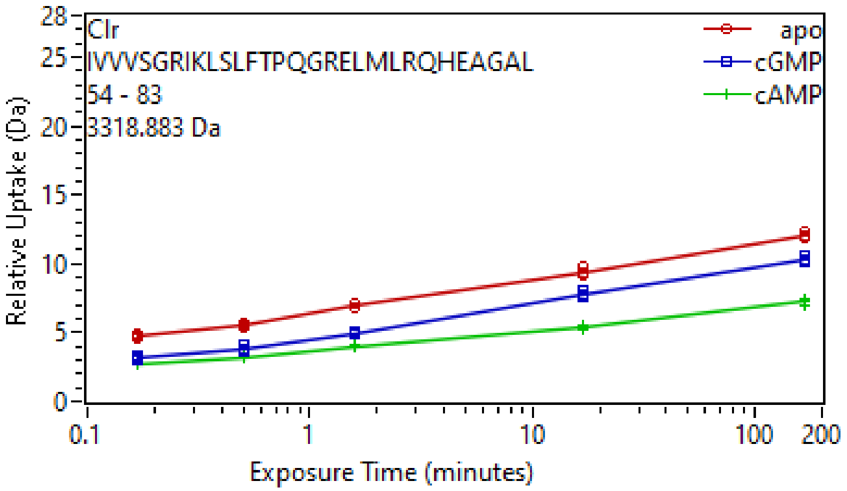

ClR
QRGDQGDYM
45 - 53
1069.437 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
8
0.1
1
10
100
200
Exposure Time (minutes)

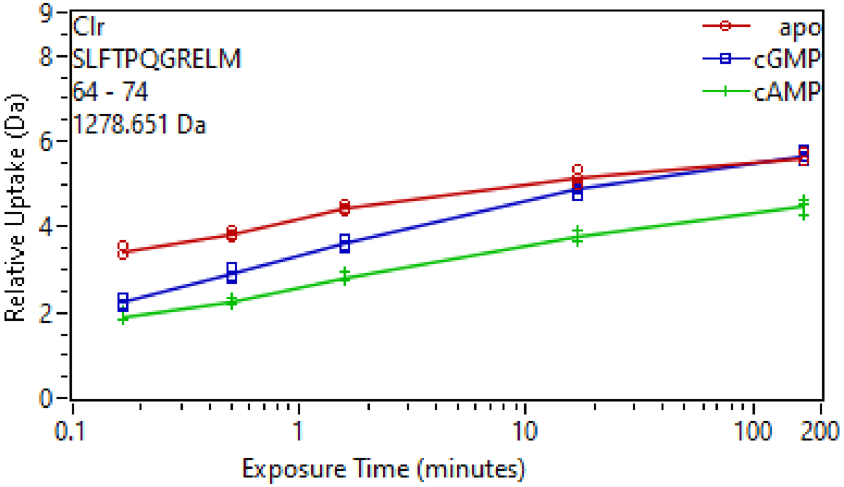

ClR
YMIV
52 - 55
525.274 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
0.5
1
1.5
2
2.5
3
0.1
1
10
100
200
Exposure Time (minutes)

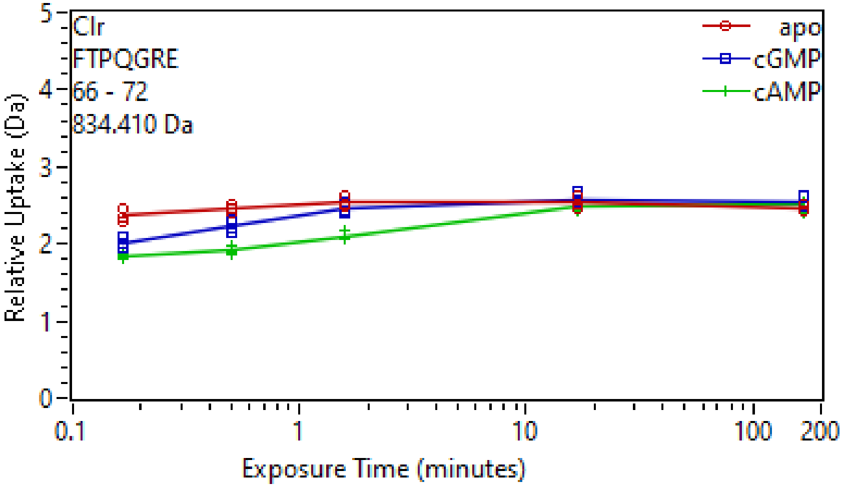

ClR
MIVVVSG
53 - 59
704.401 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
0.1
1
10
100
200
Exposure Time (minutes)

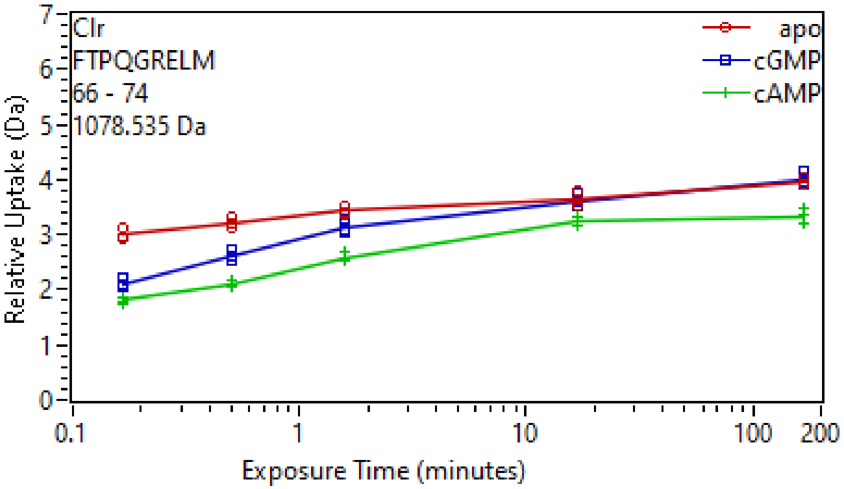

ClR
IVVVSG
54 - 59
573.361 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
0.1
1
10
100
200
Exposure Time (minutes)

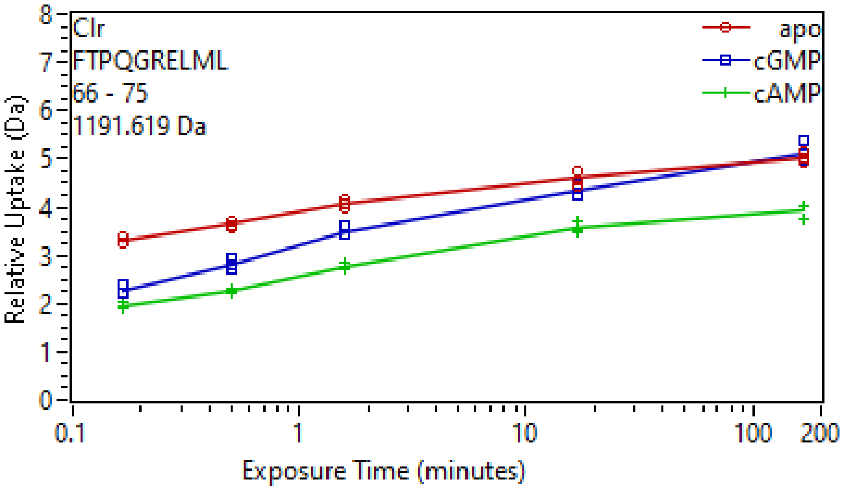

ClR
IVVVSGRIKLSLFTPQGRELM
54 - 74
2343.358 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
19
0.1
1
10
100
200
Exposure Time (minutes)

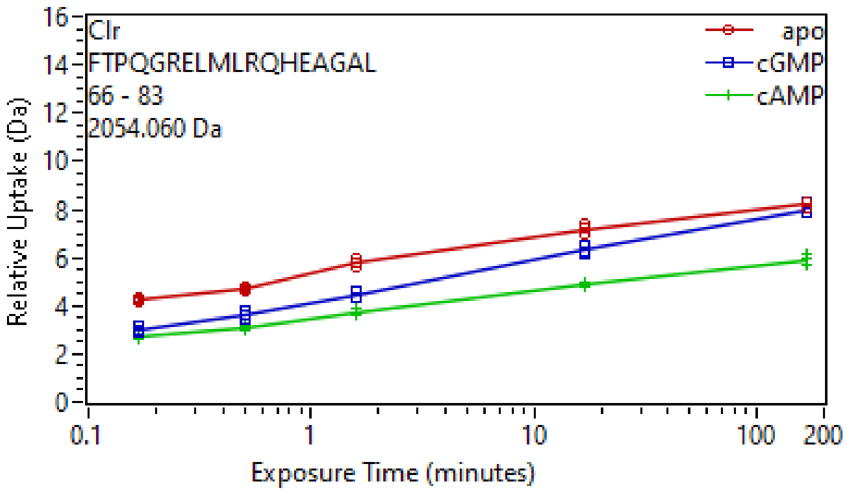

ClR
IVVVSGRIKLSLFTPQGRELMLRQHEAGAL
54 - 83
3318.883 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
5
10
15
20
25
28
0.1
1
10
100
200
Exposure Time (minutes)

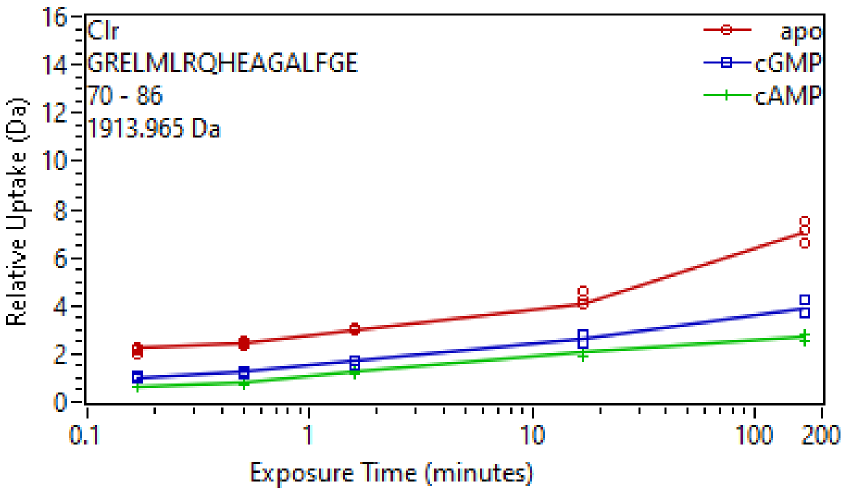

ClrSLFTPQGRELM
64 - 74
1278.651 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
9
0.1
1
10
100
200
Exposure Time (minutes)

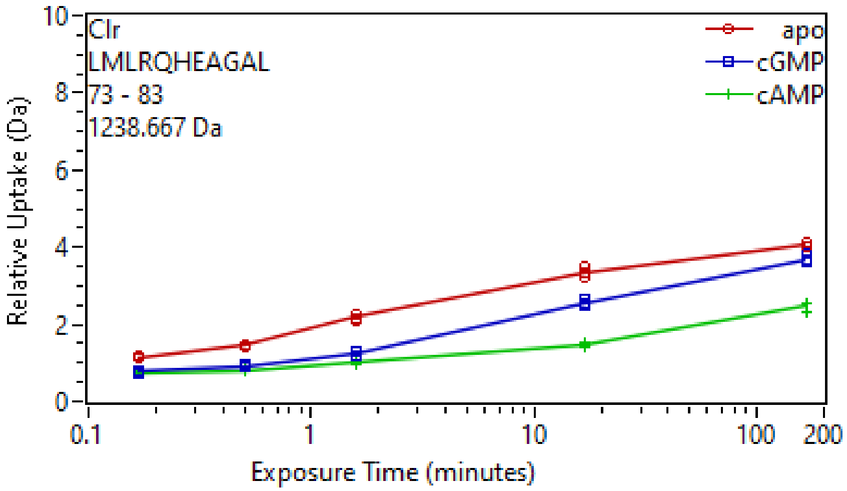

Clr
FTPQGRE
66 - 72
834.410 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
0.1
1
10
100
200
Exposure Time (minutes)

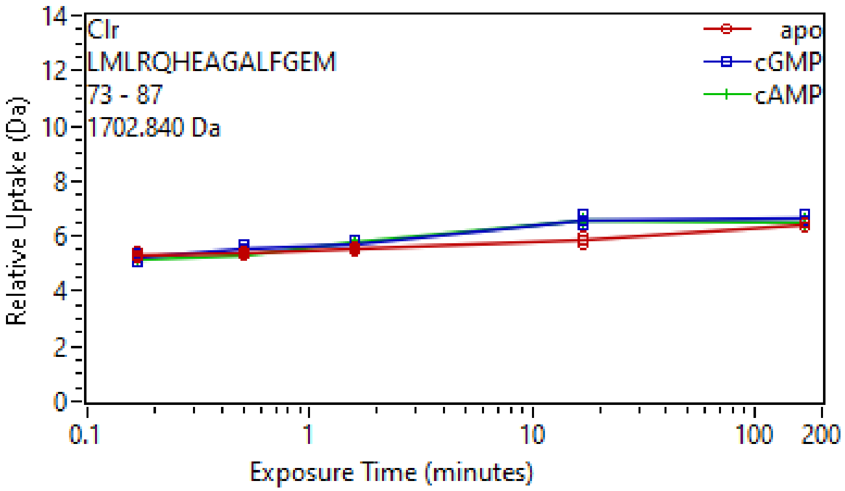

Clr
FTPQGRELM
66 - 74
1078.535 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
0.1
1
10
100
200
Exposure Time (minutes)

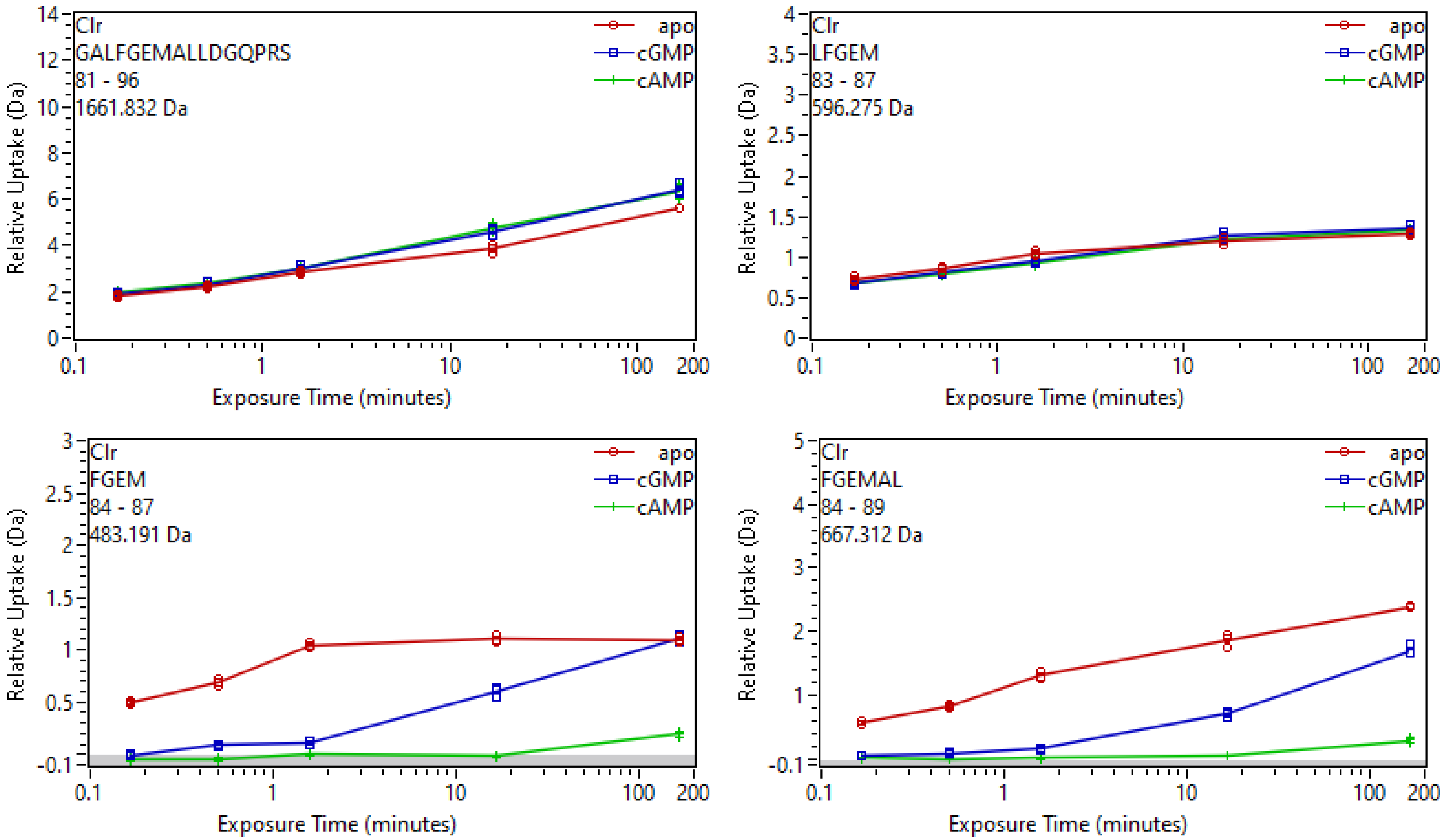

Clr
FTPQGRELML
66 - 75
1191.619 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
8
0.1
1
10
100
200
Exposure Time (minutes)

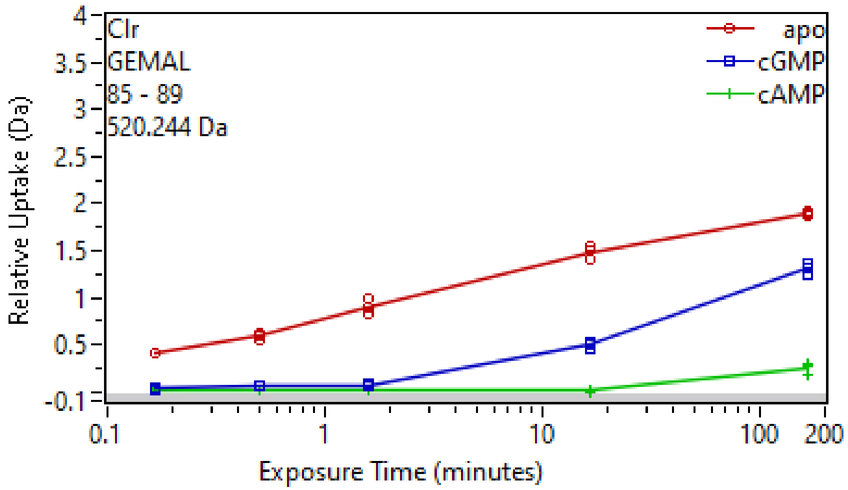

Clr
FTPQGRELMLRQHEAGAL
66 - 83
2054.060 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
14
16
0.1
1
10
100
200
Exposure Time (minutes)

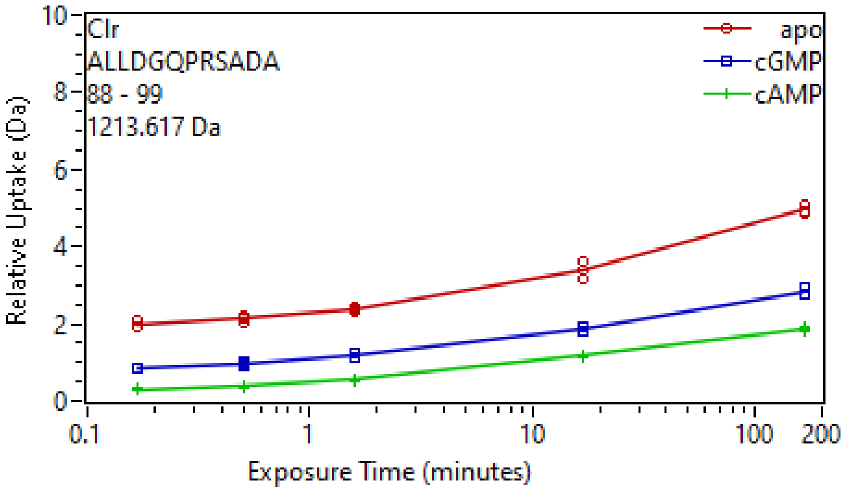

Clr
GRELMLRQHEAGALFGE
70 - 86
1913.965 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
14
16
0.1
1
10
100
200
Exposure Time (minutes)

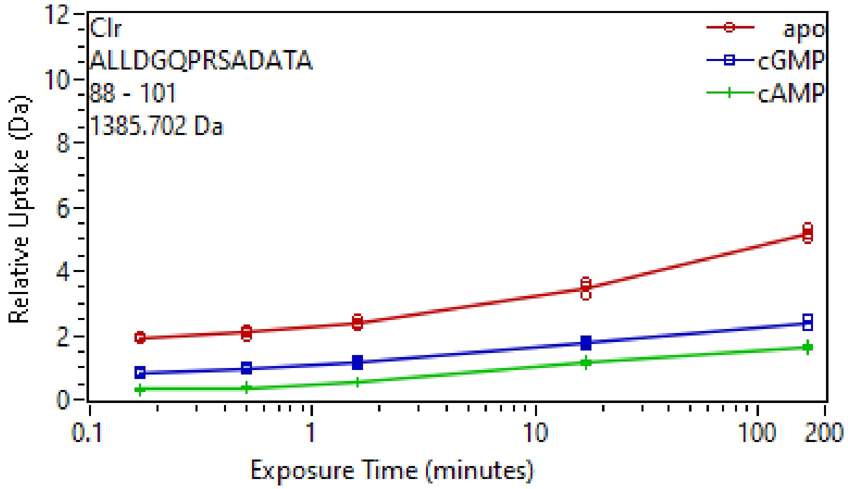

Clr
LMLRQHEAGAL
73 - 83
1238.667 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
0.1
1
10
100
200
Exposure Time (minutes)

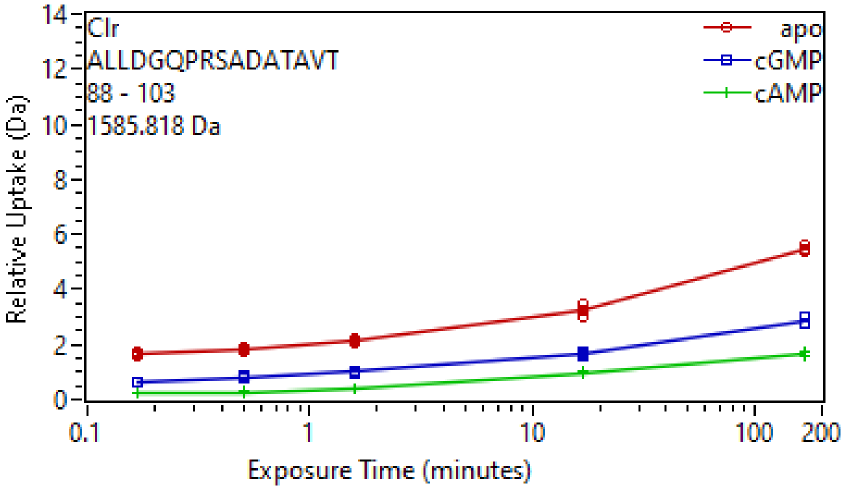

Clr
LMLRQHEAGALFGEM
73 - 87
1702.840 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
14
0.1
1
10
100
200
Exposure Time (minutes)

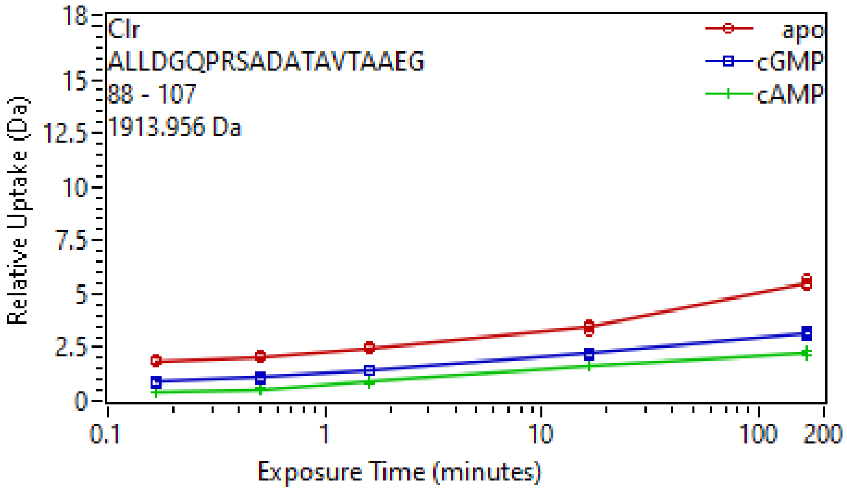

Clr
GALFGEMALLDGQPRS
81 - 96
1661.832 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
14
Exposure Time (minutes)
0.1
1
10
100
200
Clr
LFGEM
83 - 87
596.275 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
0.5
1
1.5
2
2.5
3
3.5
4
Exposure Time (minutes)
0.1
1
10
100
200
Clr
FGEM
84 - 87
483.191 Da
apo
cGMP
cAMP
Relative Uptake (Da)
-0.1
0.5
1
1.5
2
2.5
3
Exposure Time (minutes)
0.1
1
10
100
200
Clr
FGEMAL
84 - 89
667.312 Da
apo
cGMP
cAMP
Relative Uptake (Da)
-0.1
1
2
3
4
5
Exposure Time (minutes)
0.1
1
10
100
200

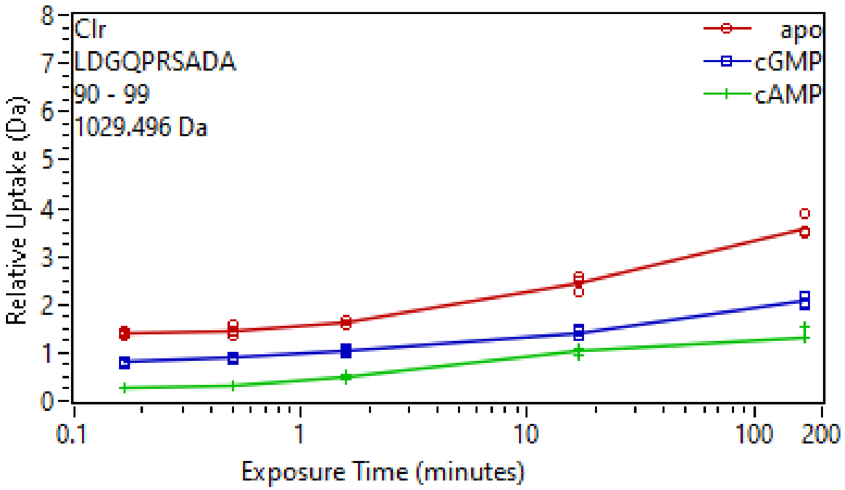

ClR
GEMAL
85 - 89
520.244 Da
apo
cGMP
cAMP
Relative Uptake (Da)
4
3.5
3
2.5
2
1.5
1
0.5
-0.1
0.1
1
10
100
200
Exposure Time (minutes)

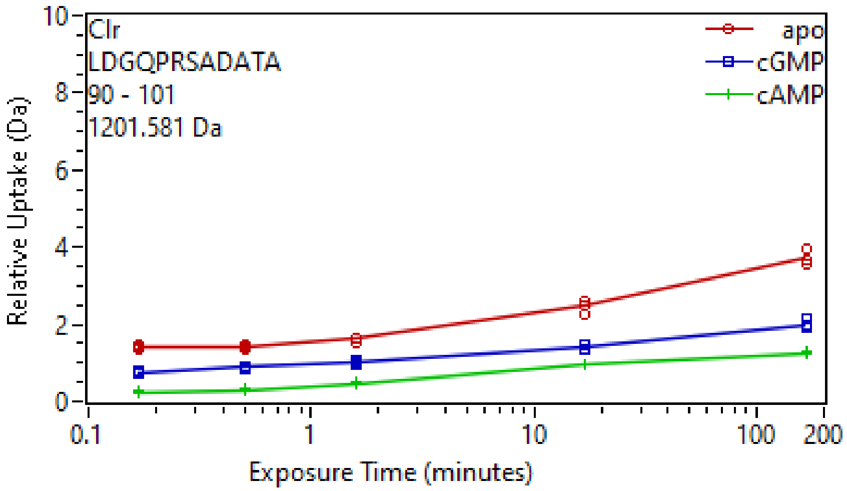

ClR
ALLDGQPRSADA
88 - 99
1213.617 Da
apo
cGMP
cAMP
Relative Uptake (Da)
10
8
6
4
2
0
0.1
1
10
100
200
Exposure Time (minutes)

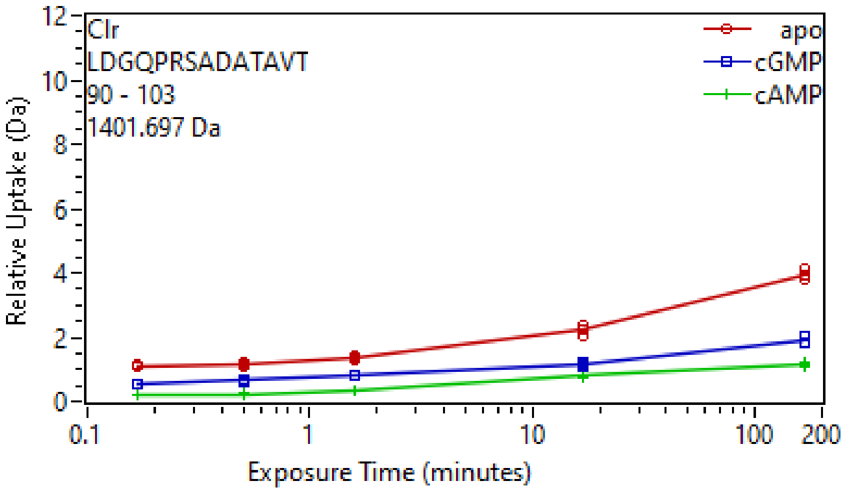

ClR
ALLDGQPRSADATA
88 - 101
1385.702 Da
apo
cGMP
cAMP
Relative Uptake (Da)
12
10
8
6
4
2
0
0.1
1
10
100
200
Exposure Time (minutes)

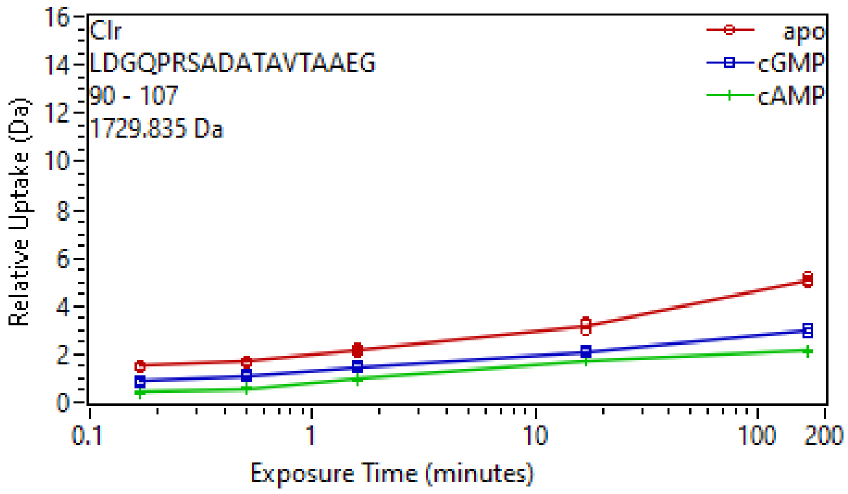

ClR
ALLDGQPRSADATAVT
88 - 103
1585.818 Da
apo
cGMP
cAMP
Relative Uptake (Da)
14
12
10
8
6
4
2
0
0.1
1
10
100
200
Exposure Time (minutes)

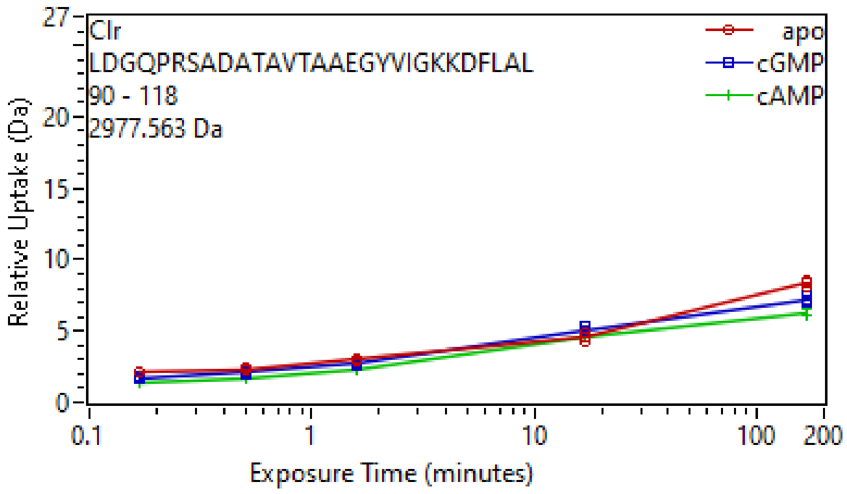

Clr
ALLDGQPRSADATAVTAAEG
88 - 107
1913.956 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
18
0.1
1
10
100
200
Exposure Time (minutes)

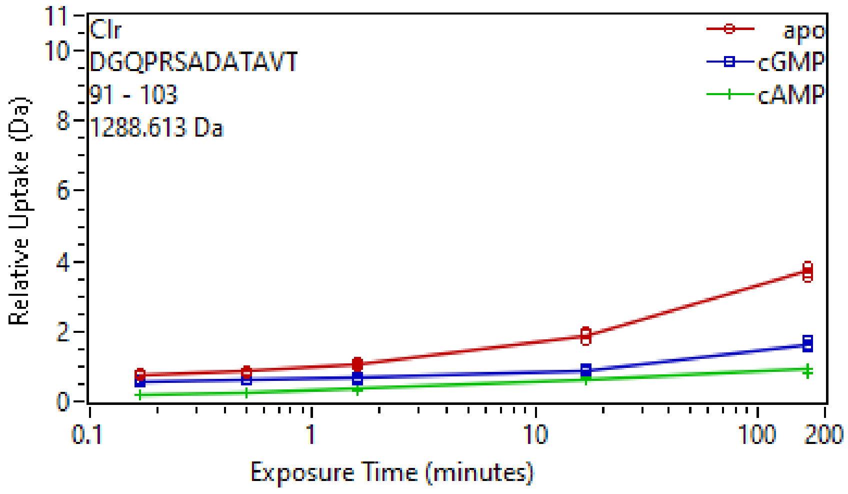

Clr
LDGQPRSADA
90 - 99
1029.496 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
8
0.1
1
10
100
200
Exposure Time (minutes)

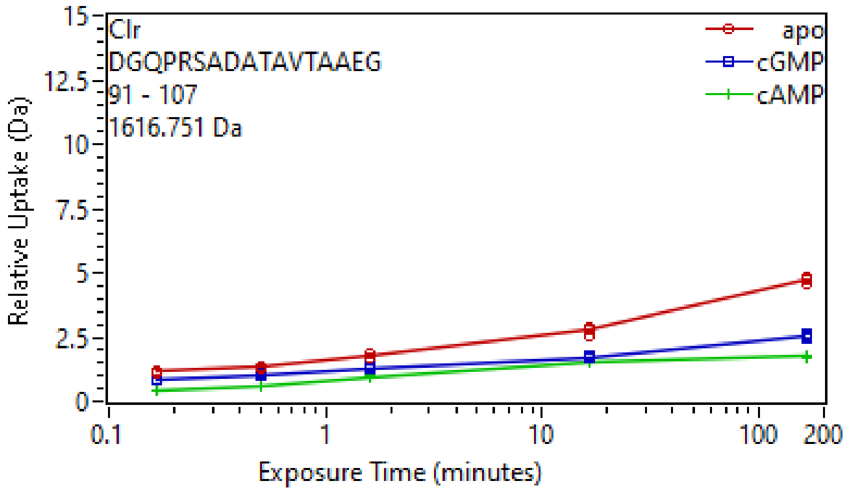

Clr
LDGQPRSADATA
90 - 101
1201.581 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
0.1
1
10
100
200
Exposure Time (minutes)

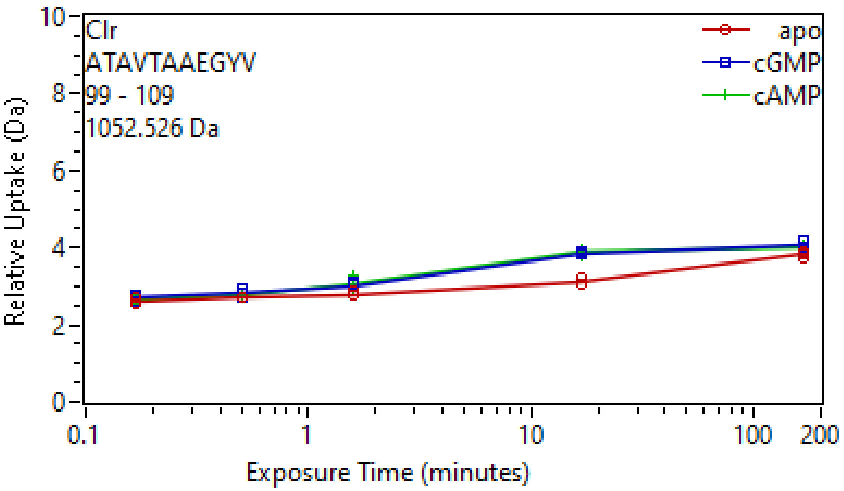

Clr
LDGQPRSADATAVT
90 - 103
1401.697 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
0.1
1
10
100
200
Exposure Time (minutes)

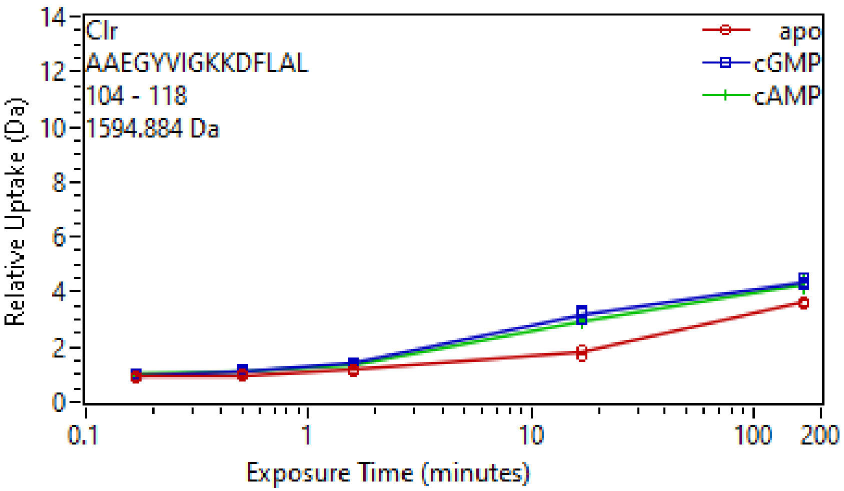

Clr
LDGQPRSADATAVTAAEG
90 - 107
1729.835 Da
apo
cGMP
cAMP
Relative Uptake (Da)
16
14
12
10
8
6
4
2
0
0.1
1
10
100
200
Exposure Time (minutes)

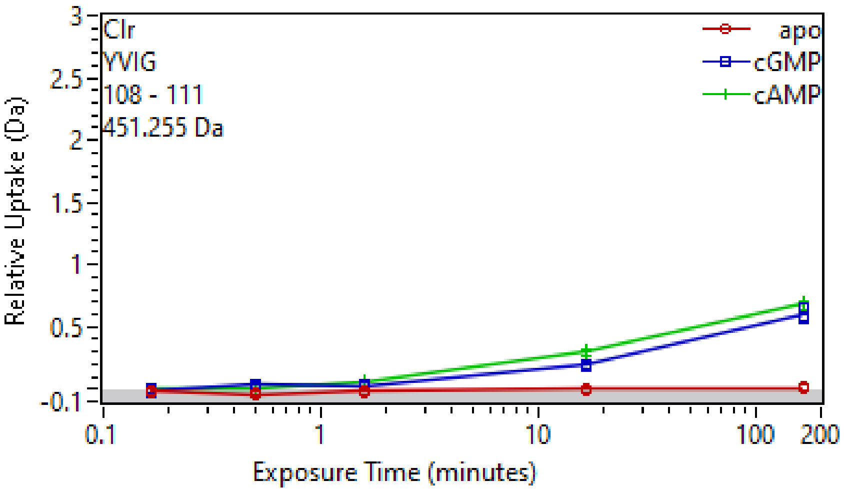

Clr
LDGQPRSADATAVTAAEGYVIGKKDFLAL
90 - 118
2977.563 Da
apo
cGMP
cAMP
Relative Uptake (Da)
27
20
15
10
5
0
0.1
1
10
100
200
Exposure Time (minutes)

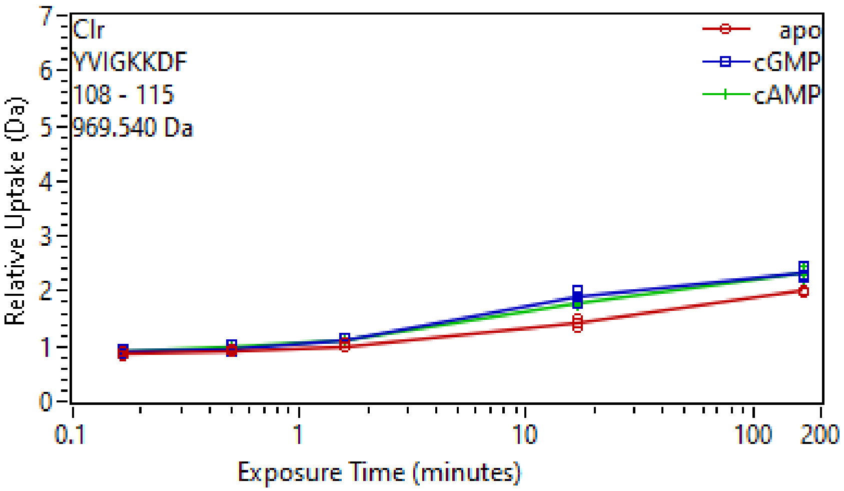

Clr
DGQPRSADATAVT
91 - 103
1288.613 Da
apo
cGMP
cAMP
Relative Uptake (Da)
11
10
8
6
4
2
0
0.1
1
10
100
200
Exposure Time (minutes)

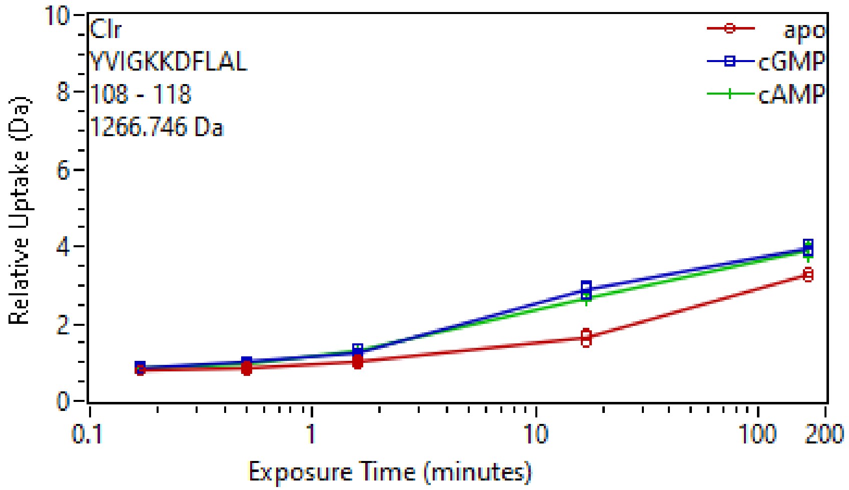

Clr
DGQPRSADATAVTAAEG
91 - 107
1616.751 Da
apo
cGMP
cAMP
Relative Uptake (Da)
15
12.5
10
7.5
5
2.5
0
0.1
1
10
100
200
Exposure Time (minutes)

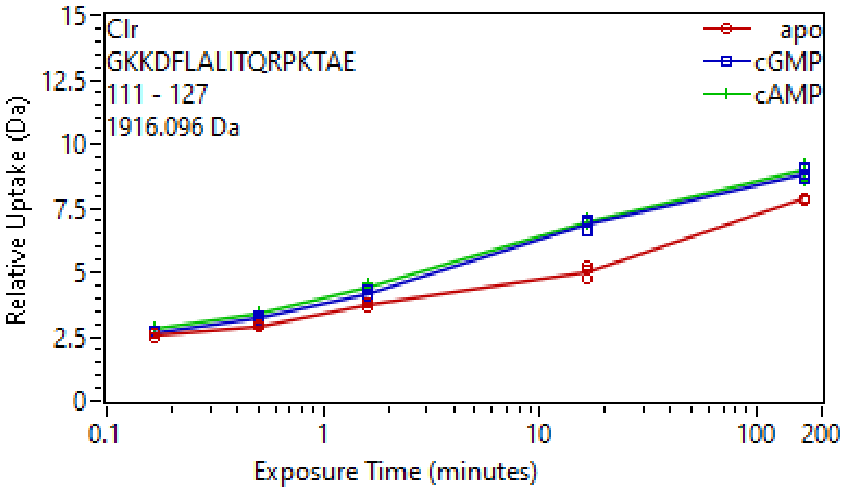

Clr
ATAVTAAEGYV
99 - 109
1052.526 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
0.1
1
10
100
200
Exposure Time (minutes)

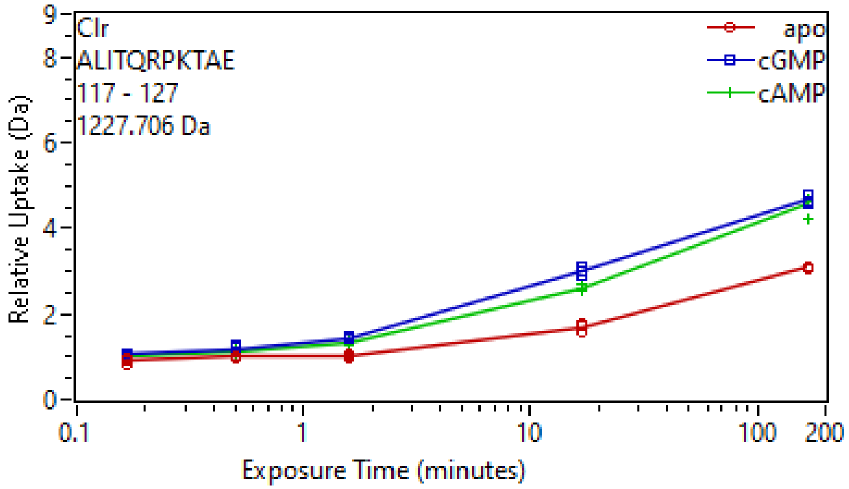

Clr
AAEGYVIGKKDFLAL
104 - 118
1594.884 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
14
0.1
1
10
100
200
Exposure Time (minutes)

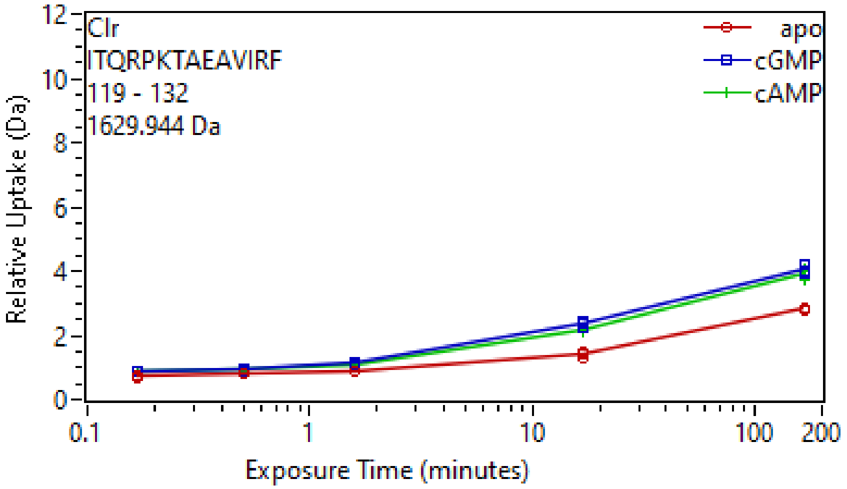

Clr
YVIG
108 - 111
451.255 Da
apo
cGMP
cAMP
Relative Uptake (Da)
-0.1
0.5
1
1.5
2
2.5
3
0.1
1
10
100
200
Exposure Time (minutes)

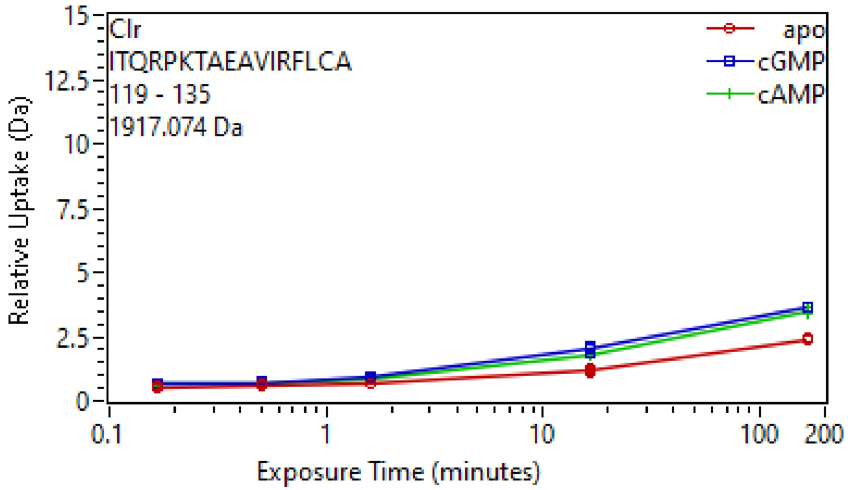

Clr
YVIGKKDF
108 - 115
969.540 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
0.1
1
10
100
200
Exposure Time (minutes)

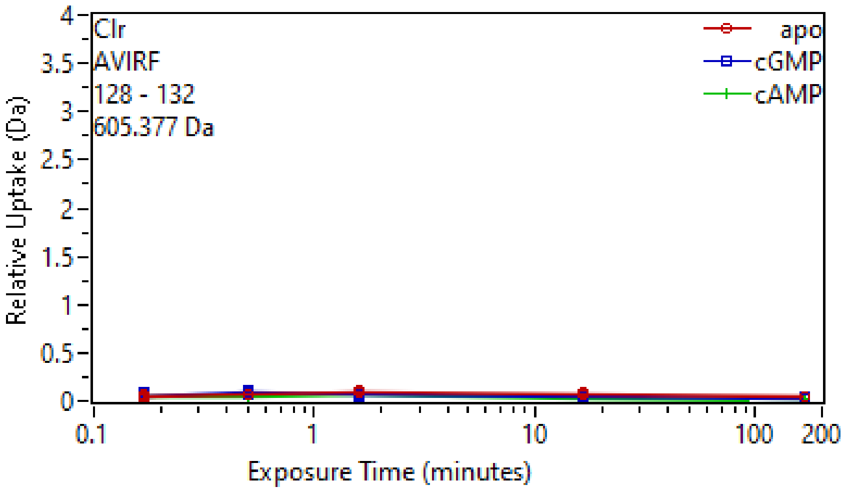

ClR
YVIGKKDFLAL
108 - 118
1266.746 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
0.1
1
10
100
200
Exposure Time (minutes)

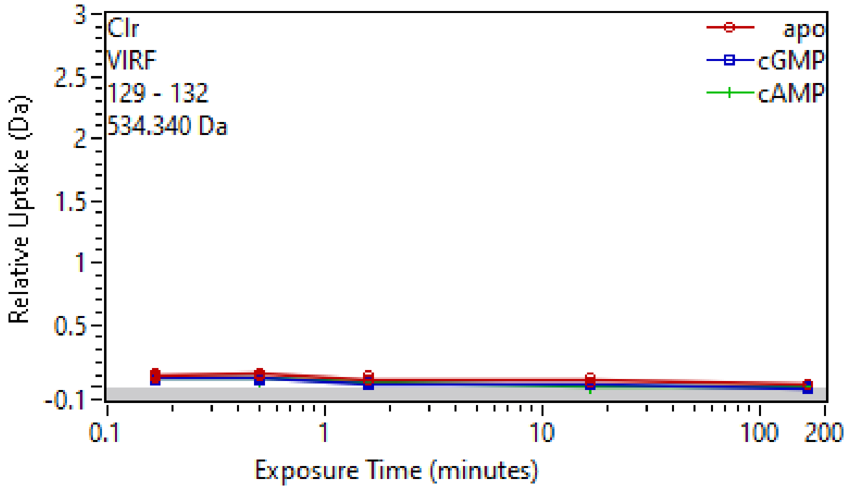

ClR
GKKDFLALITQRPKTAE
111 - 127
1916.096 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
0.1
1
10
100
200
Exposure Time (minutes)

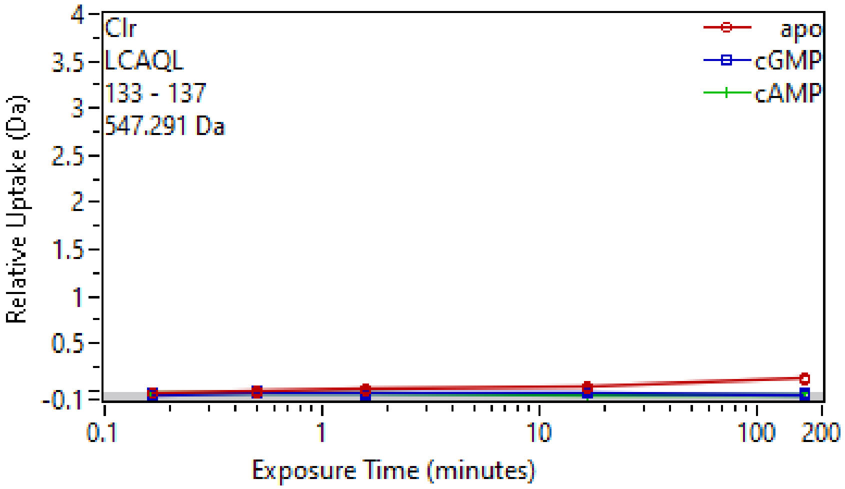

ClR
ALITQRPKTAE
117 - 127
1227.706 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
9
0.1
1
10
100
200
Exposure Time (minutes)

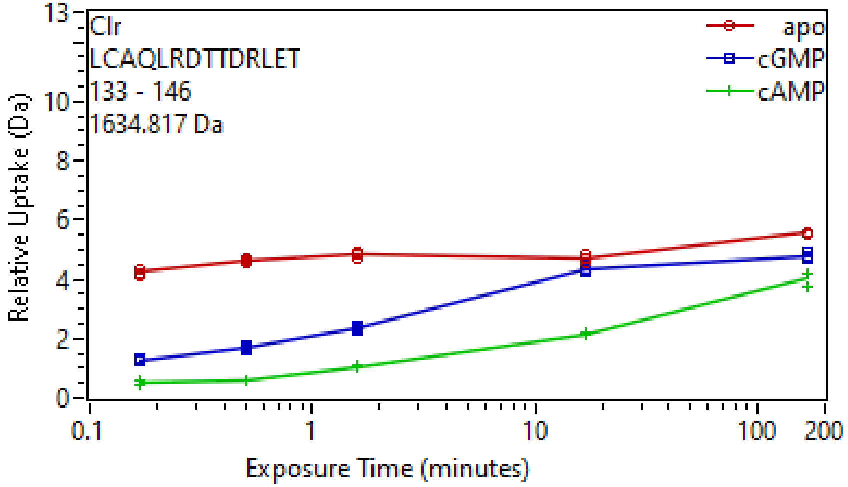

ClR
ITQRPKTAEAVIRF
119 - 132
1629.944 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
0.1
1
10
100
200
Exposure Time (minutes)

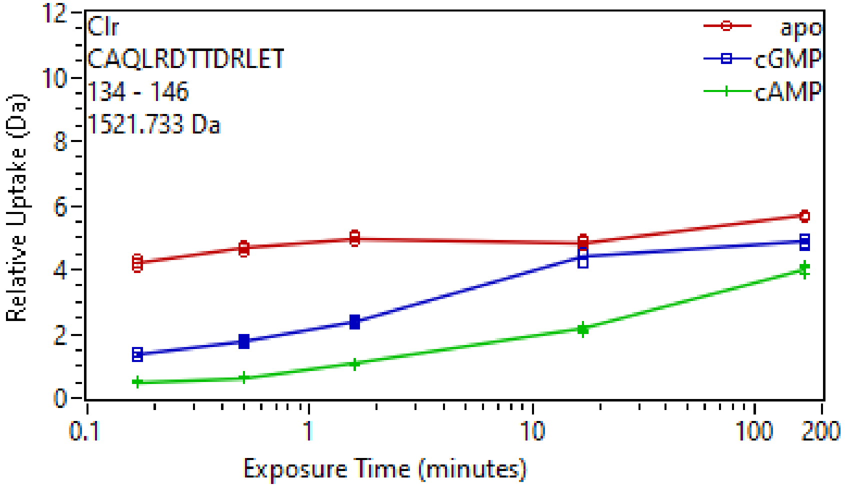

Clr
ITQRPKTAEAVIRFLCA
119 - 135
1917.074 Da
apo
cGMP
cAMP
Relative Uptake (Da)
15
12.5
10
7.5
5
2.5
0
0.1
1
10
100
200
Exposure Time (minutes)

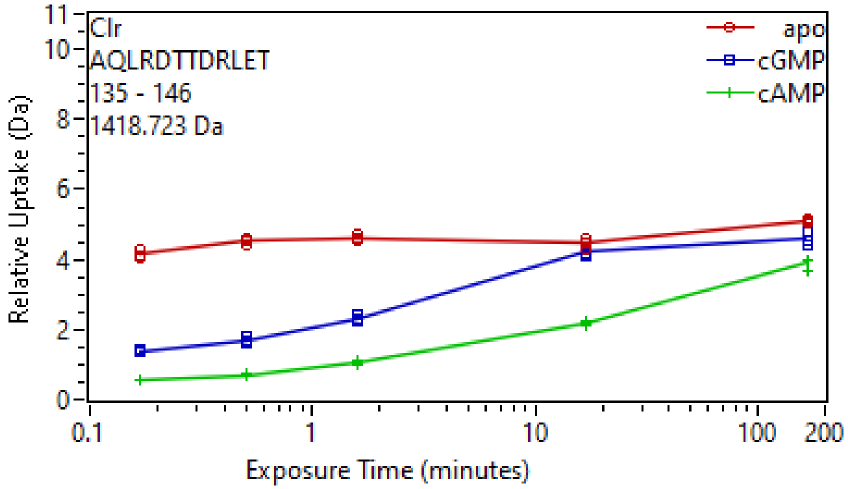

Clr
AVIRF
128 - 132
605.377 Da
apo
cGMP
cAMP
Relative Uptake (Da)
4
3.5
3
2.5
2
1.5
1
0.5
0
0.1
1
10
100
200
Exposure Time (minutes)

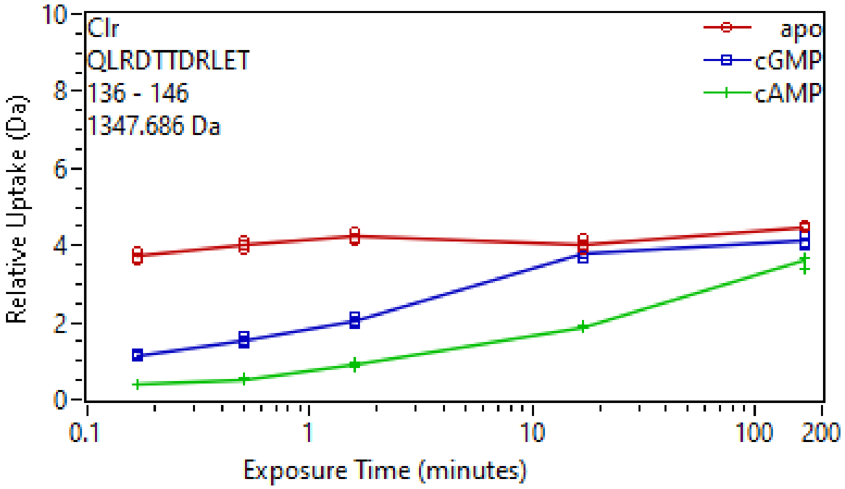

Clr
VIRF
129 - 132
534.340 Da
apo
cGMP
cAMP
Relative Uptake (Da)
3
2.5
2
1.5
1
0.5
-0.1
0.1
1
10
100
200
Exposure Time (minutes)

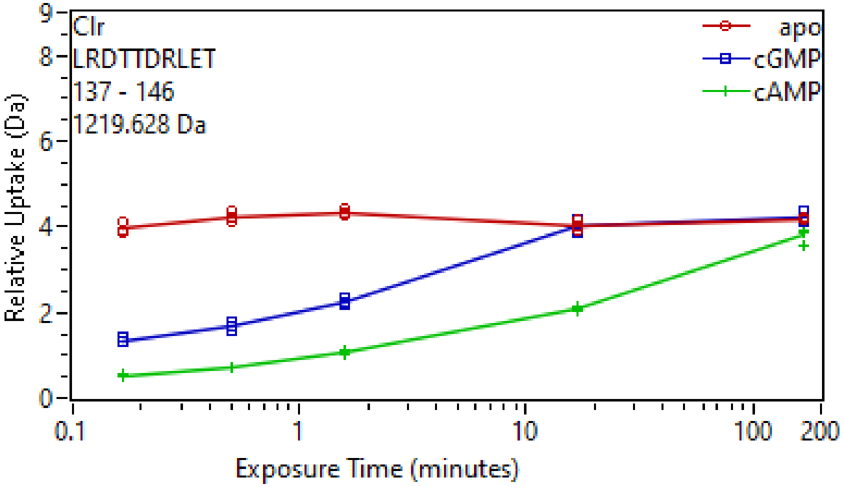

Clr
LCAQL
133 - 137
547.291 Da
apo
cGMP
cAMP
Relative Uptake (Da)
4
3.5
3
2.5
2
1.5
1
0.5
-0.1
0.1
1
10
100
200
Exposure Time (minutes)

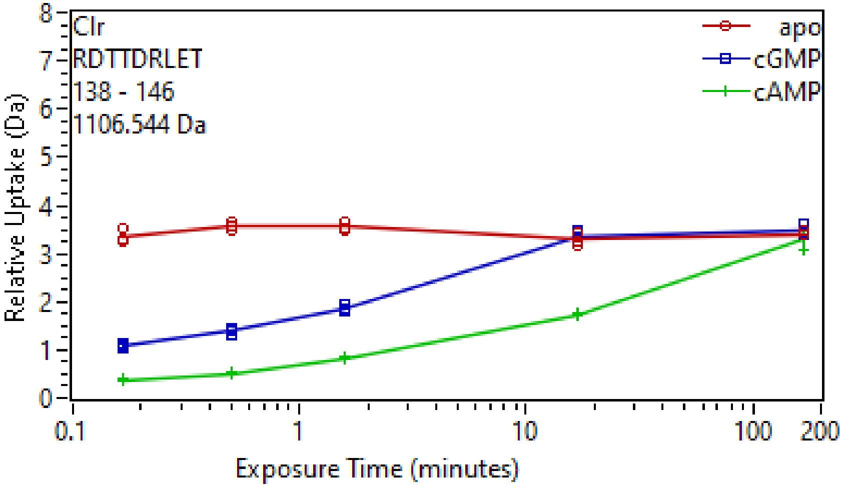

ClR
LCAQLRDTTDRLET
133 - 146
1634.817 Da
apo
cGMP
cAMP
Relative Uptake (Da)
13
10
8
6
4
2
0
0.1
1
10
100
200
Exposure Time (minutes)

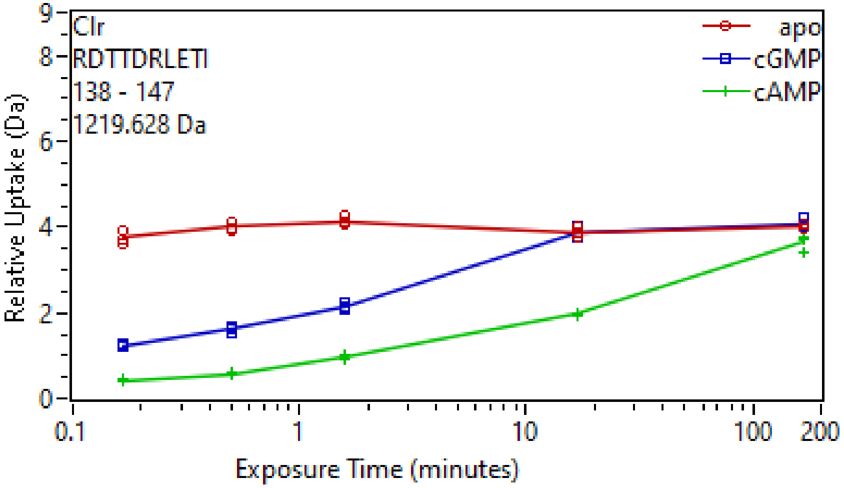

ClR
CAQLRDTTDRLET
134 - 146
1521.733 Da
apo
cGMP
cAMP
Relative Uptake (Da)
12
10
8
6
4
2
0
0.1
1
10
100
200
Exposure Time (minutes)

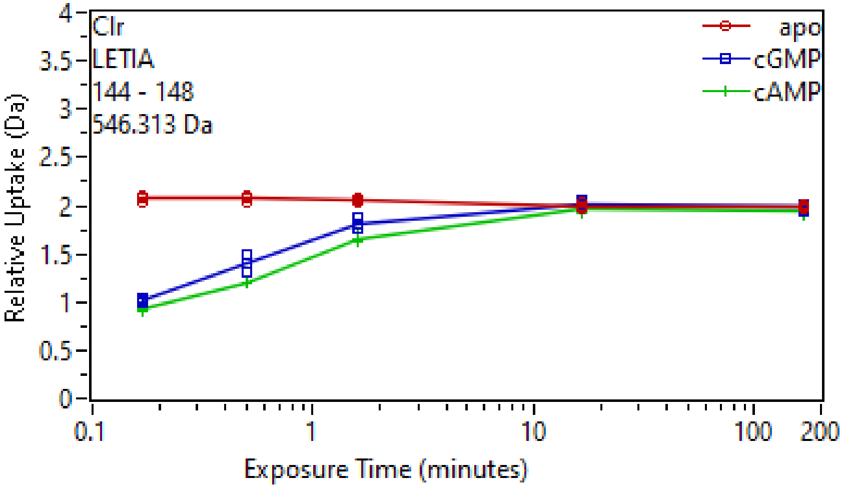

ClR
AQLRDTTDRLET
135 - 146
1418.723 Da
apo
cGMP
cAMP
Relative Uptake (Da)
11
10
8
6
4
2
0
0.1
1
10
100
200
Exposure Time (minutes)

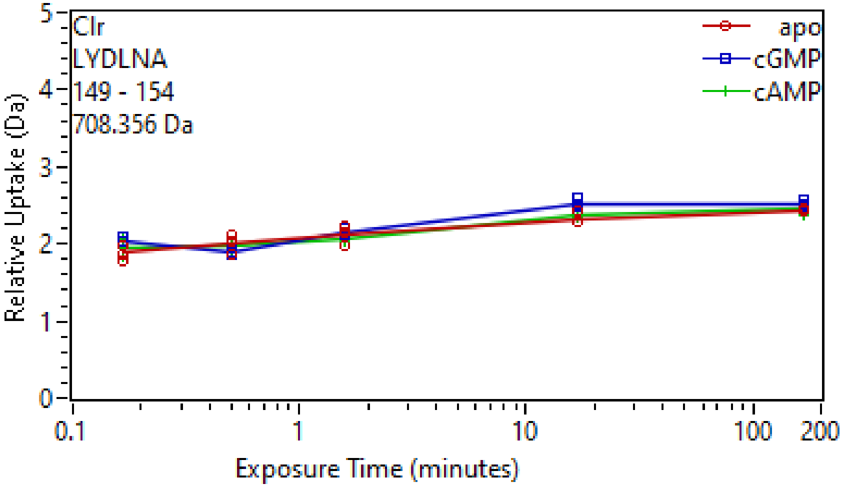

ClR
QLRDTTDRLET
136 - 146
1347.686 Da
apo
cGMP
cAMP
Relative Uptake (Da)
10
8
6
4
2
0
0.1
1
10
100
200
Exposure Time (minutes)

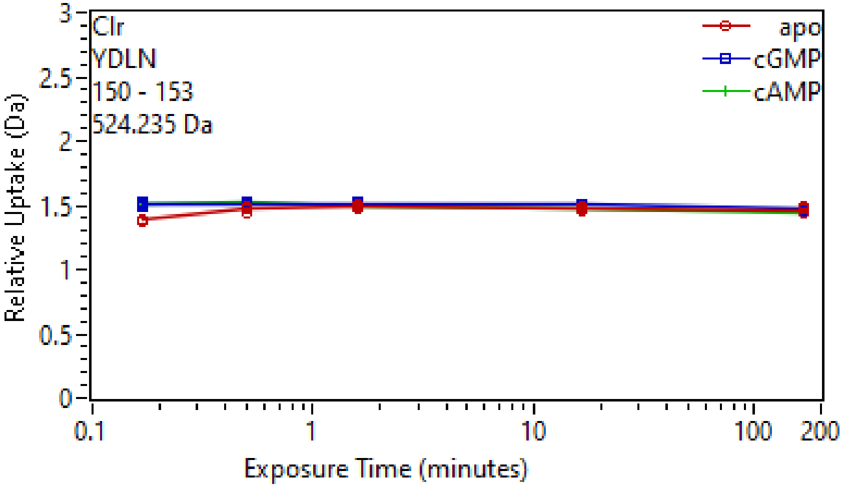

Clr
LRDTTDRLET
137 - 146
1219.628 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
9
0.1
1
10
100
200
Exposure Time (minutes)

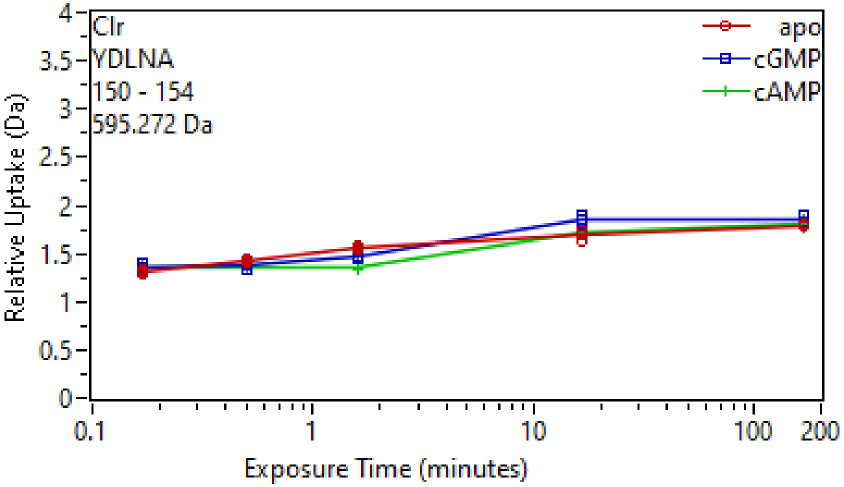

Clr
RDTTDRLET
138 - 146
1106.544 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
8
0.1
1
10
100
200
Exposure Time (minutes)

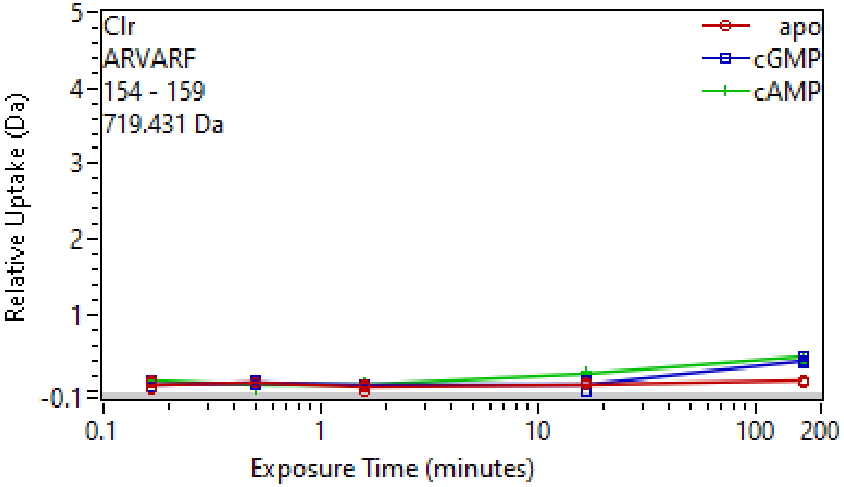

Clr
RDTTDRLETI
138 - 147
1219.628 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
9
0.1
1
10
100
200
Exposure Time (minutes)

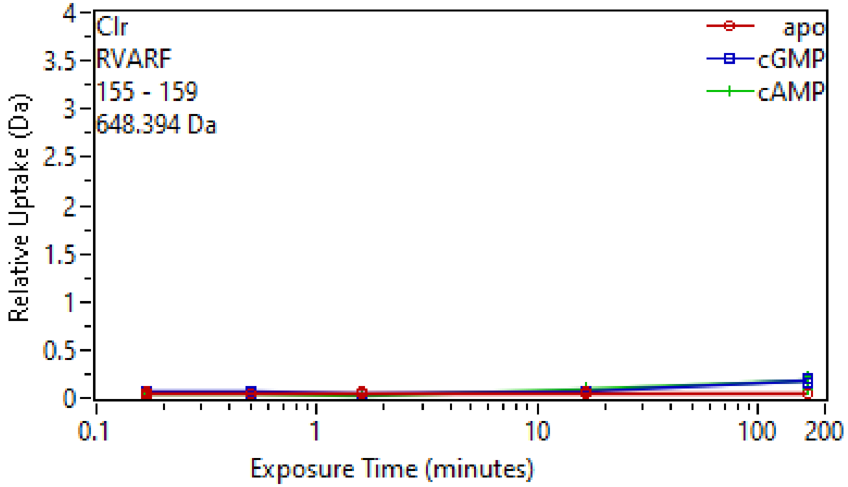

Clr
LETIA
144 - 148
546.313 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
0.5
1
1.5
2
2.5
3
3.5
4
0.1
1
10
100
200
Exposure Time (minutes)

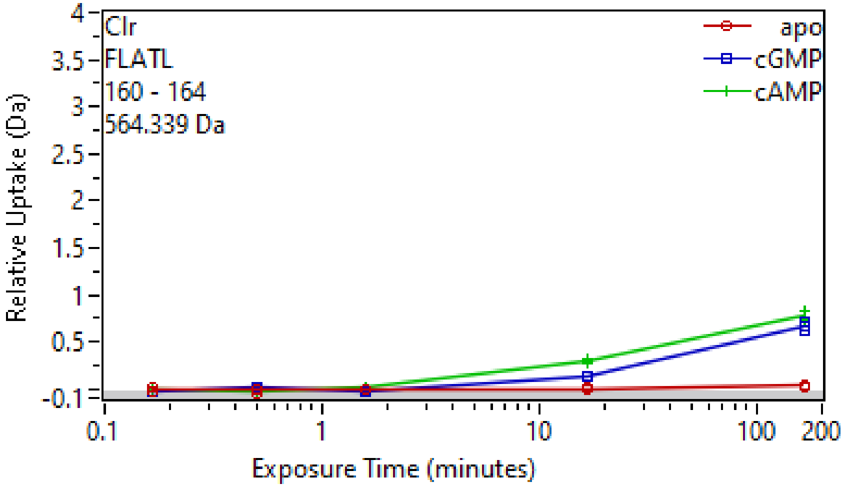

Clr
LYDLNA
149 - 154
708.356 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
0.1
1
10
100
200
Exposure Time (minutes)

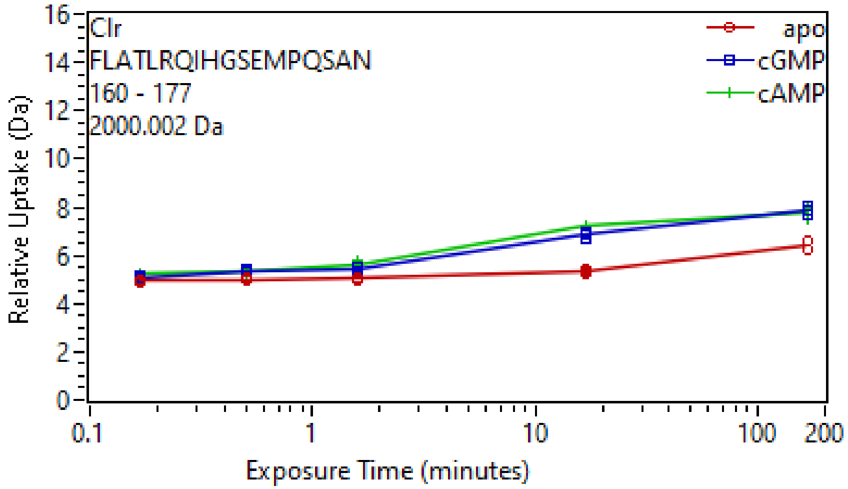

Clr
YDLN
150 - 153
524.235 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
0.5
1
1.5
2
2.5
3
0.1
1
10
100
200
Exposure Time (minutes)

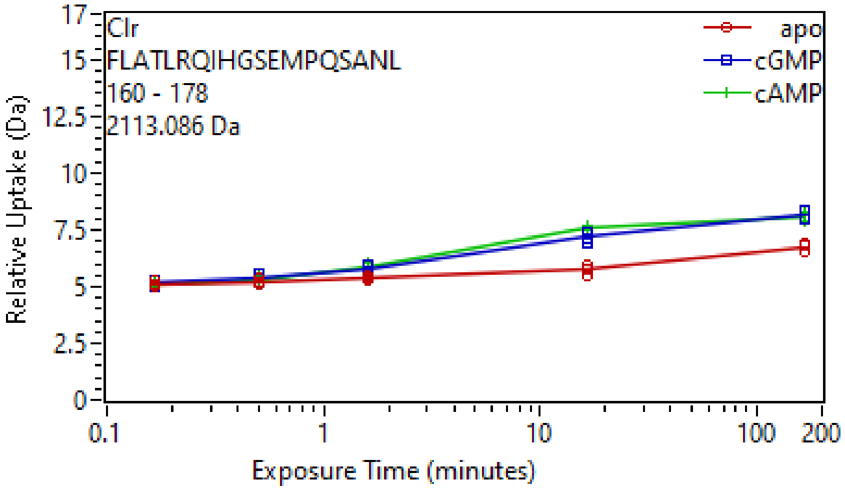

Clr
YDLNA
150 - 154
595.272 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
0.5
1
1.5
2
2.5
3
3.5
4
0.1
1
10
100
200
Exposure Time (minutes)

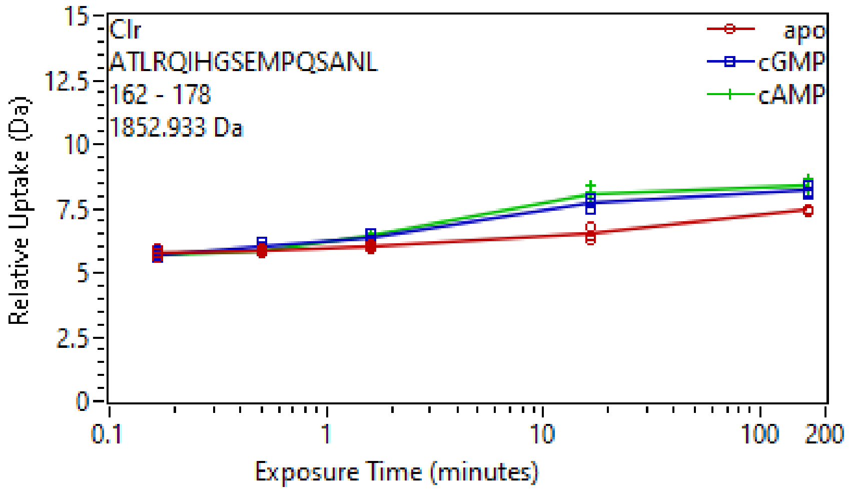

Clr
ARVARF
154 - 159
719.431 Da
apo
cGMP
cAMP
Relative Uptake (Da)
-0.1
1
2
3
4
5
0.1
1
10
100
200
Exposure Time (minutes)

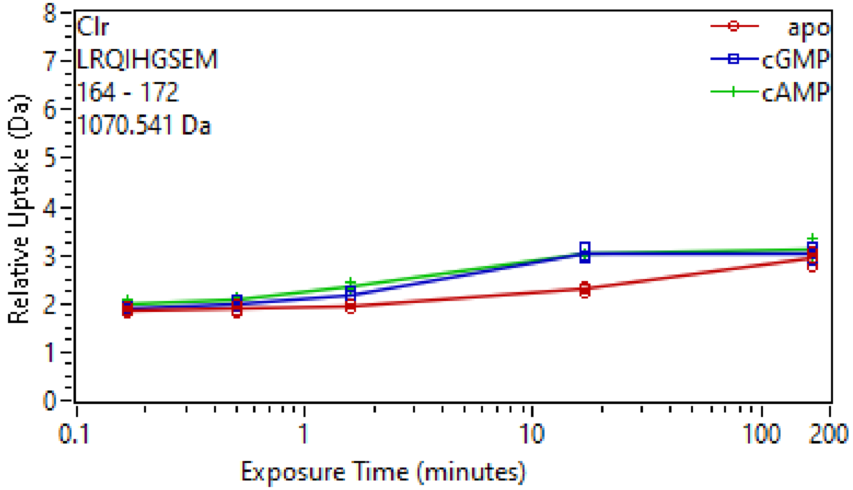

Clr
RVARF
155 - 159
648.394 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
0.5
1
1.5
2
2.5
3
3.5
4
0.1
1
10
100
200
Exposure Time (minutes)

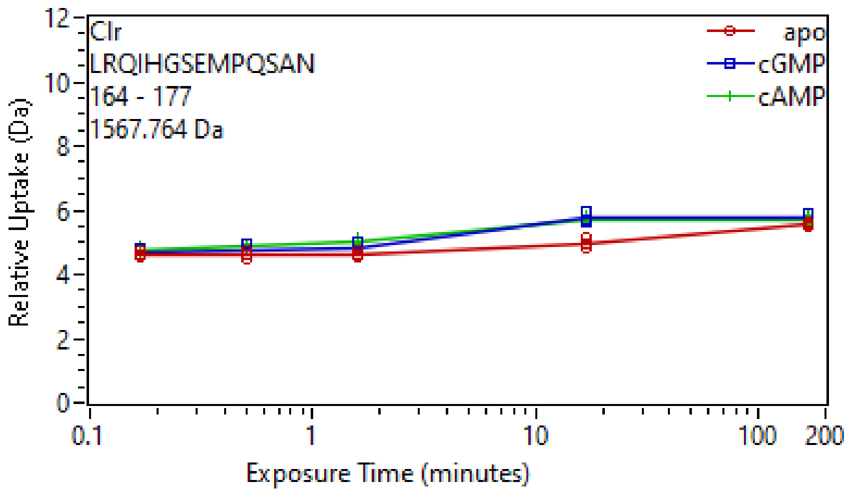

Clr
FLATL
160 - 164
564.339 Da
apo
cGMP
cAMP
Relative Uptake (Da)
-0.1
0.5
1
1.5
2
2.5
3
3.5
4
0.1
1
10
100
200
Exposure Time (minutes)

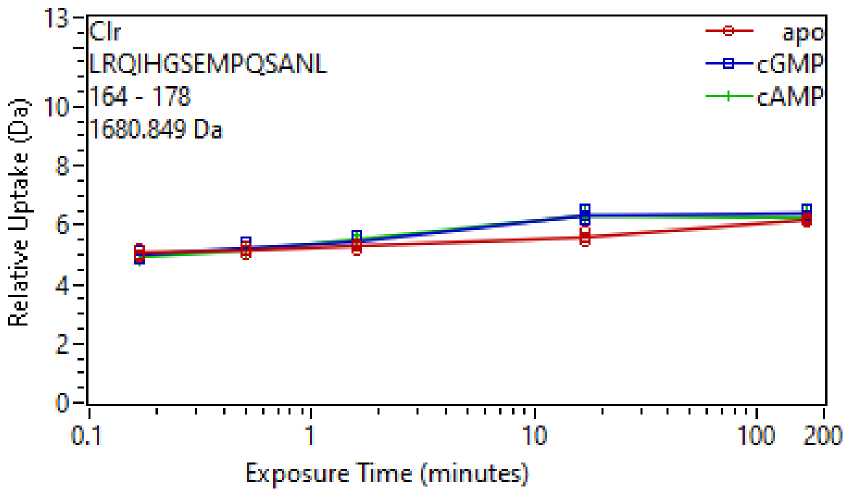

Clr
FLATLRQIHGSEMPQSAN
160 - 177
2000.002 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
14
16
0.1
1
10
100
200
Exposure Time (minutes)

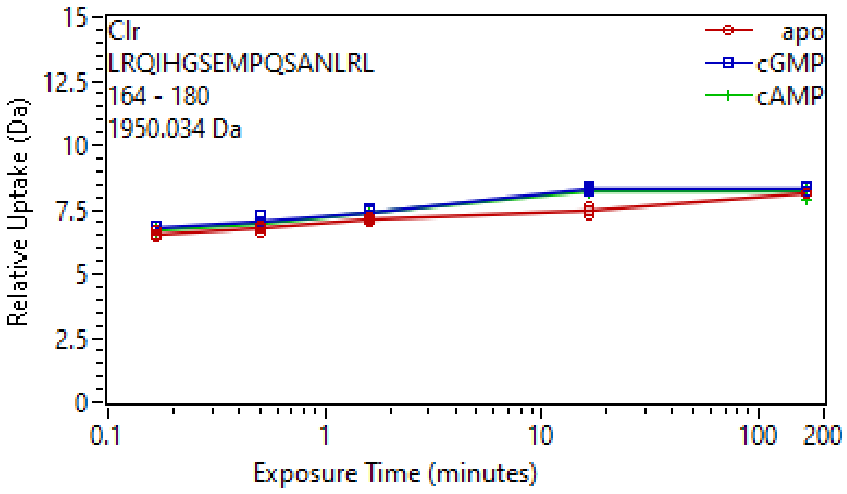

Clr
FLATLRQIHGSEMPQSANL
160 - 178
2113.086 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
17
0.1
1
10
100
200
Exposure Time (minutes)

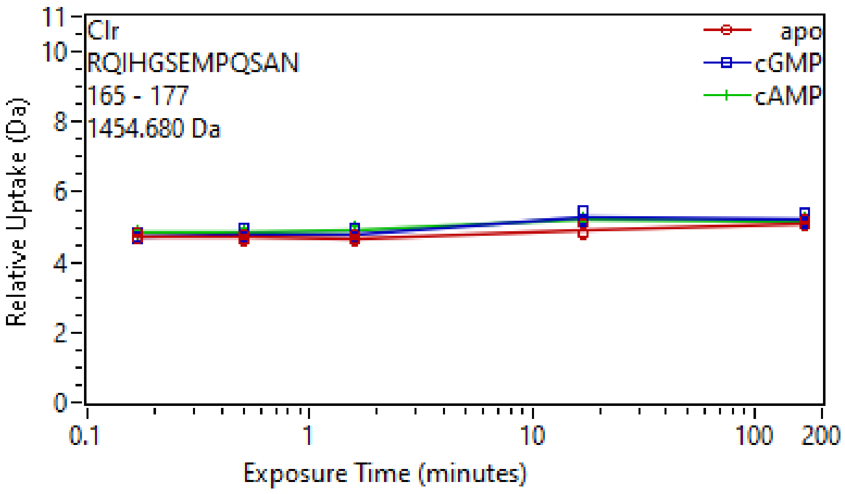

Clr
ATLRQIHGSEMPQSANL
162 - 178
1852.933 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
0.1
1
10
100
200
Exposure Time (minutes)

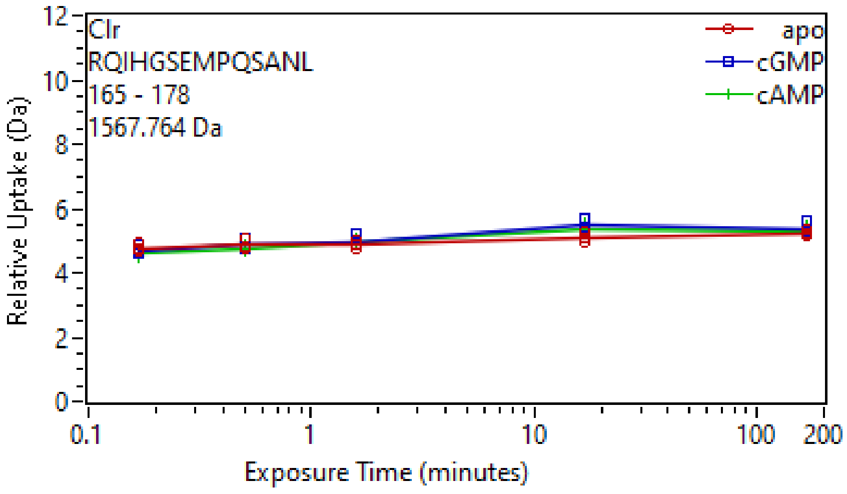

Clr
LRQIHGSEM
164 - 172
1070.541 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
8
0.1
1
10
100
200
Exposure Time (minutes)

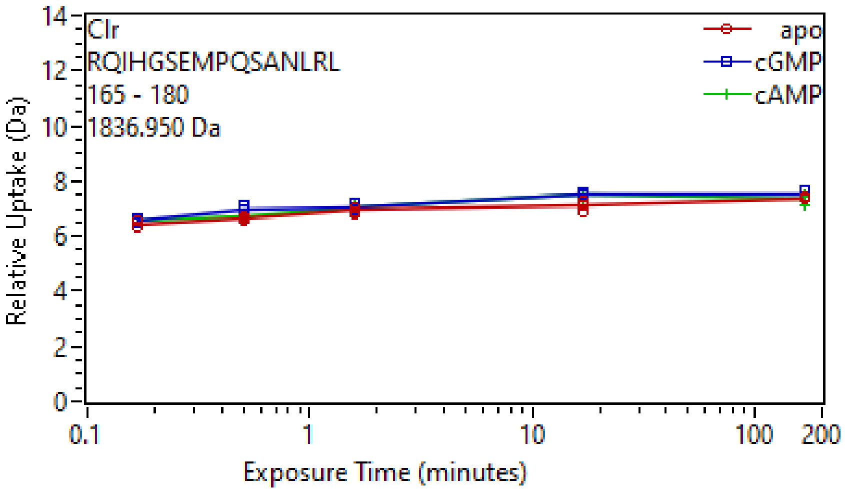

Clr
LRQIHGSEMPQSAN
164 - 177
1567.764 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
0.1
1
10
100
200
Exposure Time (minutes)

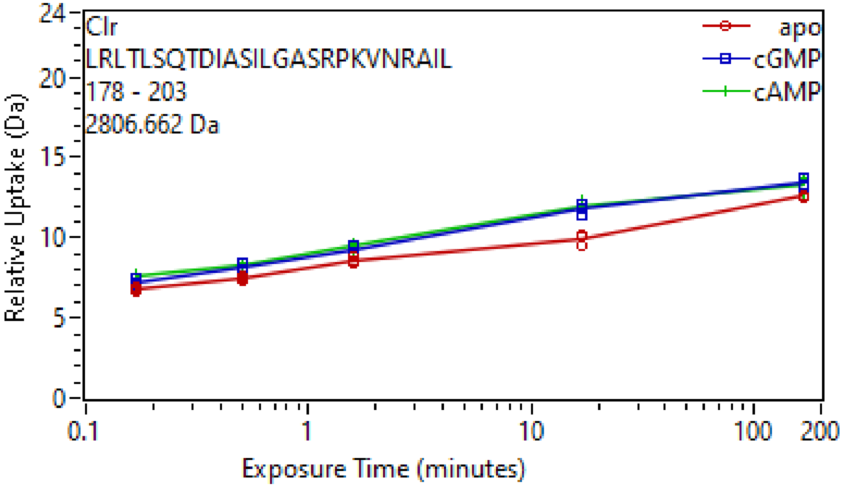

Clr
LRQIHGSEMPQSANL
164 - 178
1680.849 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
13
0.1
1
10
100
200
Exposure Time (minutes)

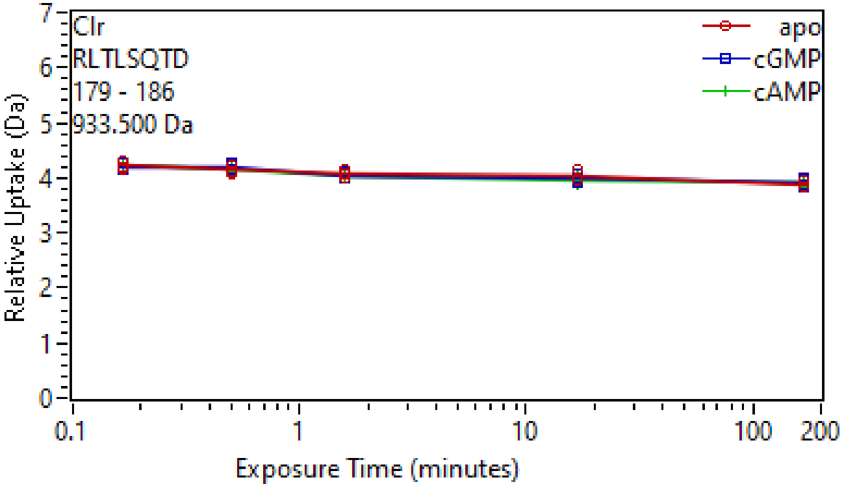

ClR
LRQIHGSEMPQSANLRL
164 - 180
1950.034 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
0.1
1
10
100
200
Exposure Time (minutes)

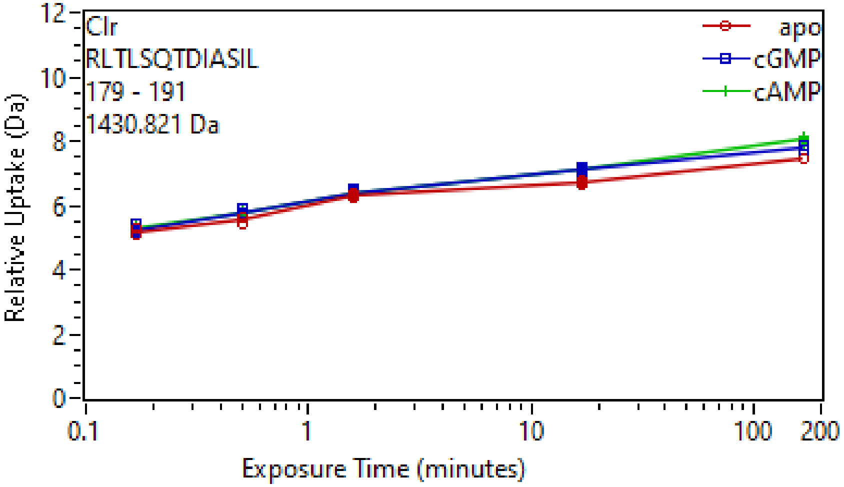

ClR
RQIHGSEMPQSAN
165 - 177
1454.680 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
11
0.1
1
10
100
200
Exposure Time (minutes)

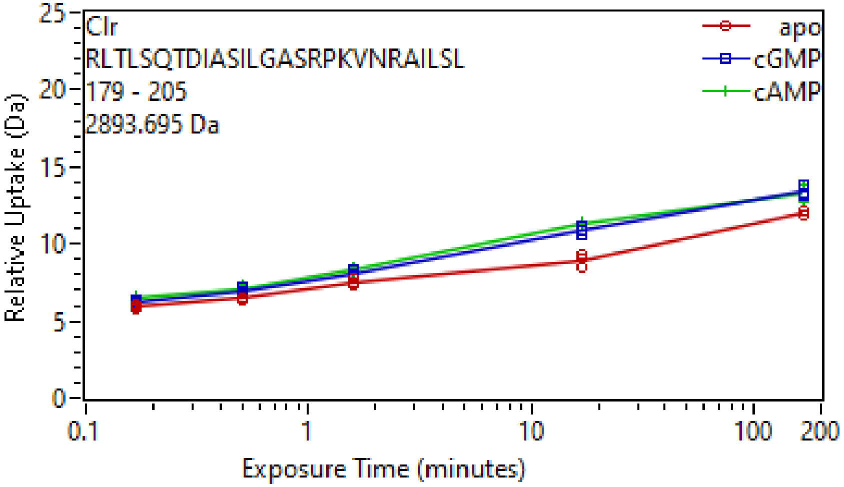

ClR
RQIHGSEMPQSANL
165 - 178
1567.764 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
0.1
1
10
100
200
Exposure Time (minutes)

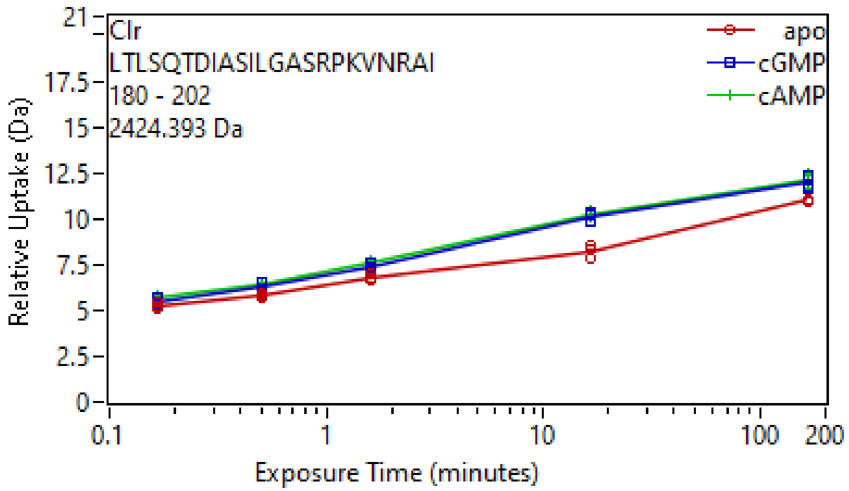

ClR
RQIHGSEMPQSANLRL
165 - 180
1836.950 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
14
0.1
1
10
100
200
Exposure Time (minutes)

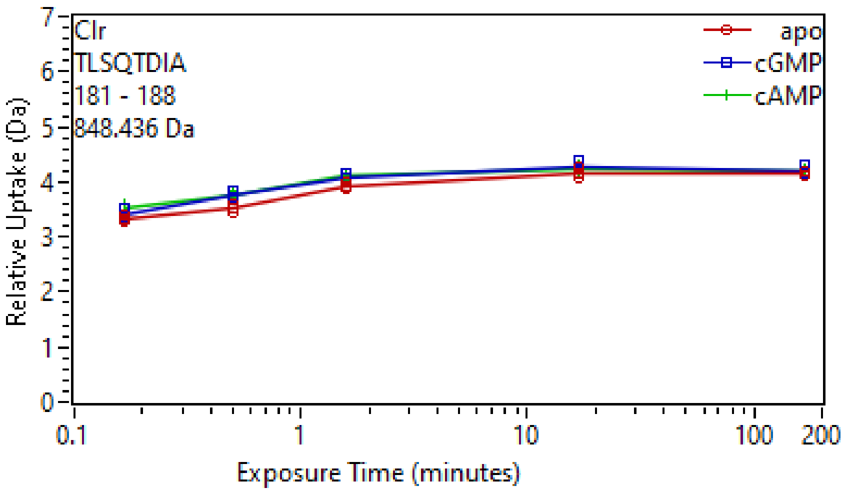

ClrLRLTLSQTDIASILGASRPKVNRAIL
178 - 203
2806.662 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
5
10
15
20
24
0.1
1
10
100
200
Exposure Time (minutes)

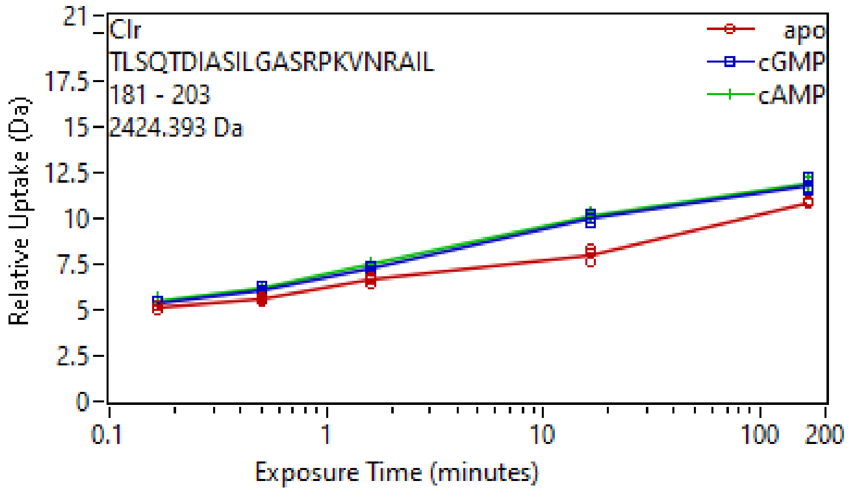

Clr
RLTLSQTD
179 - 186
933.500 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
0.1
1
10
100
200
Exposure Time (minutes)

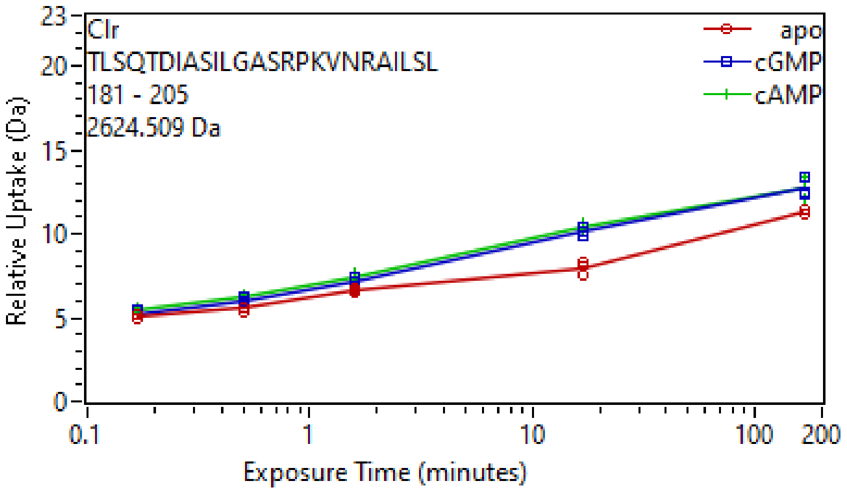

Clr
RLTLSQTDIASIL
179 - 191
1430.821 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
0.1
1
10
100
200
Exposure Time (minutes)

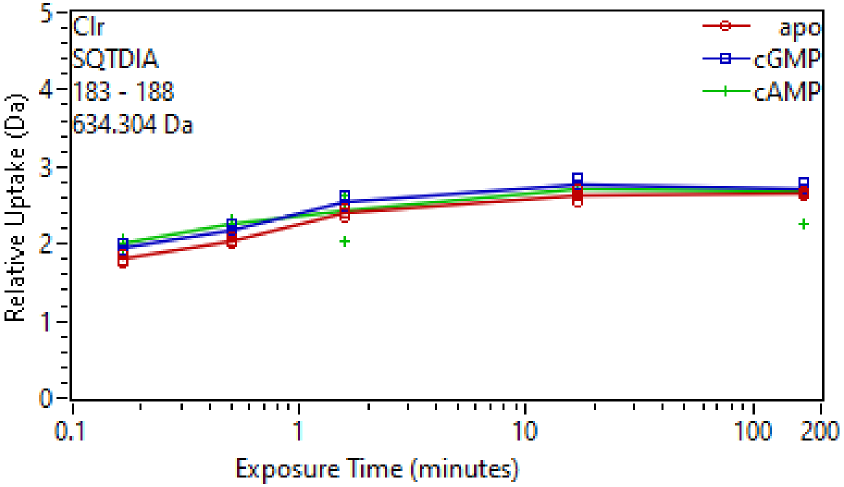

Clr
RLTLSQTDIASILGASRPKVNRAILSL
179 - 205
2893.695 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
5
10
15
20
25
0.1
1
10
100
200
Exposure Time (minutes)

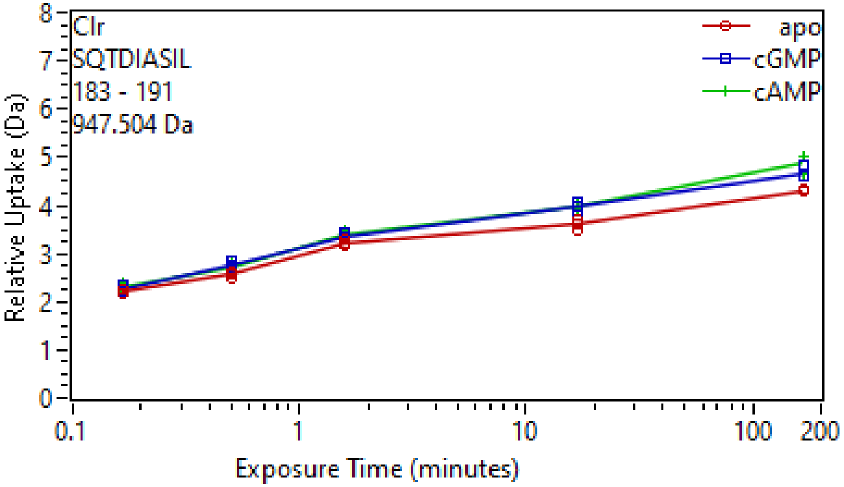

ClR
LTLSQTDIASILGASRPKVNRAI
180 - 202
2424.393 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
17.5
21
0.1
1
10
100
200
Exposure Time (minutes)

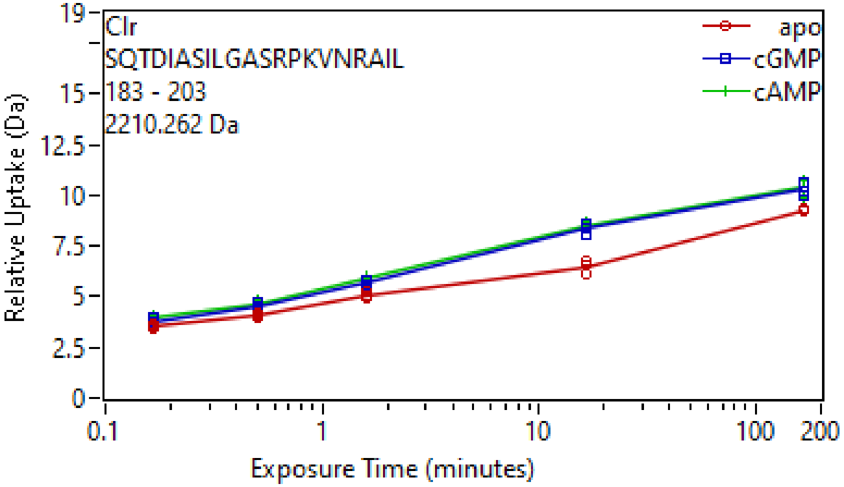

ClR
TLSQTDIA
181 - 188
848.436 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
0.1
1
10
100
200
Exposure Time (minutes)

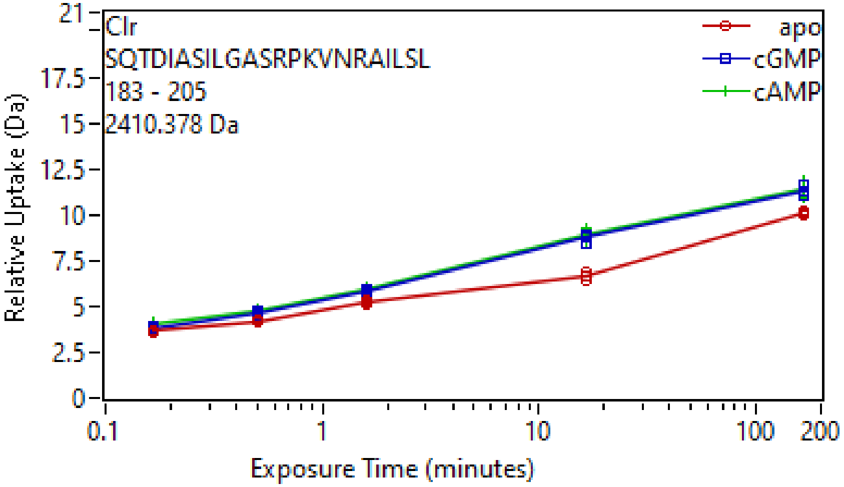

ClR
TLSQTDIASILGASRPKVNRAIL
181 - 203
2424.393 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
17.5
21
0.1
1
10
100
200
Exposure Time (minutes)

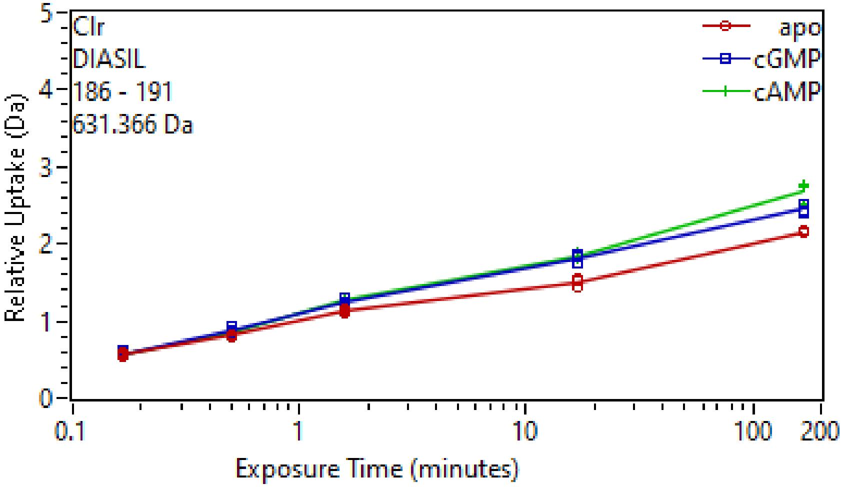

ClR
TLSQTDIASILGASRPKVNRAILSL
181 - 205
2624.509 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
5
10
15
20
23
0.1
1
10
100
200
Exposure Time (minutes)

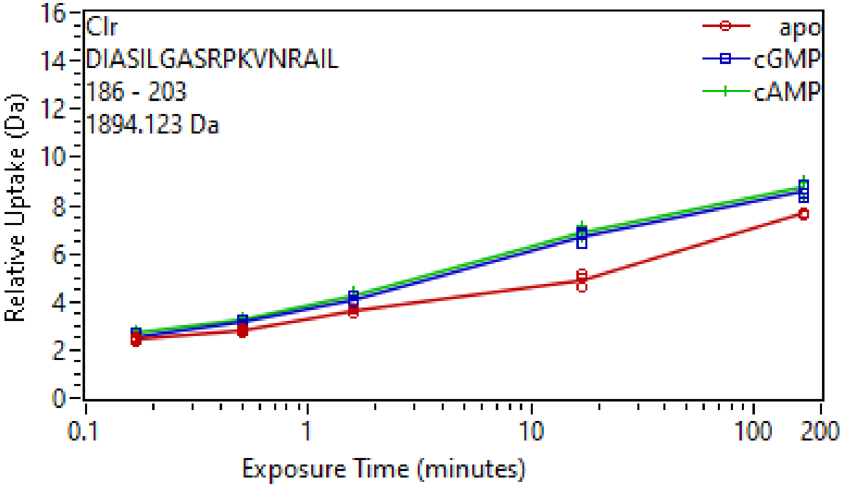

Clr
SQTDIA
183 - 188
634.304 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
0.1
1
10
100
200
Exposure Time (minutes)

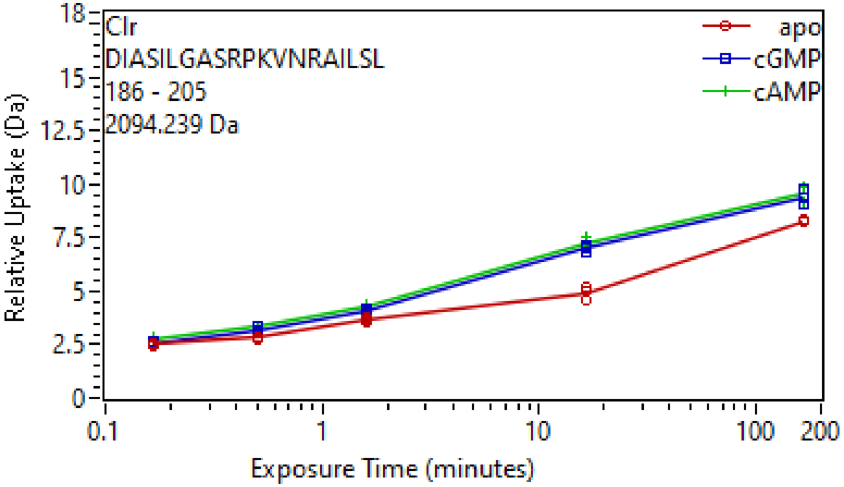

Clr
SQTDIASIL
183 - 191
947.504 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
8
0.1
1
10
100
200
Exposure Time (minutes)

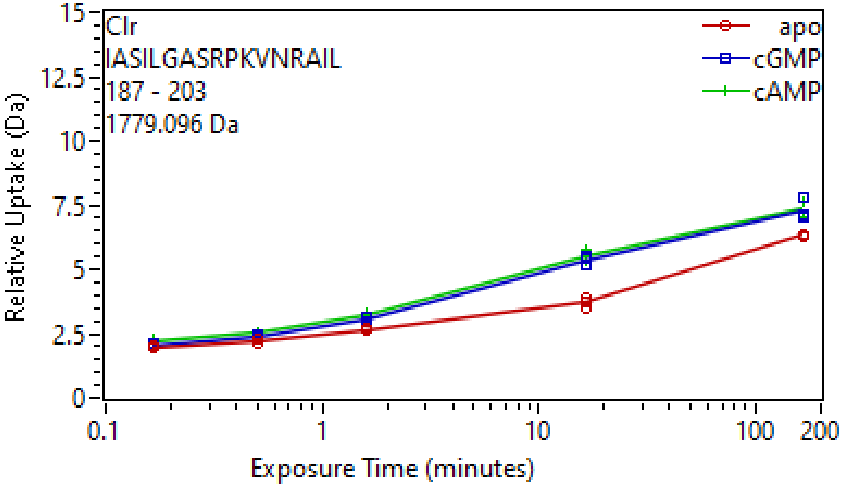

Clr
SQTDIASILGASRPKVNRAIL
183 - 203
2210.262 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
19
0.1
1
10
100
200
Exposure Time (minutes)

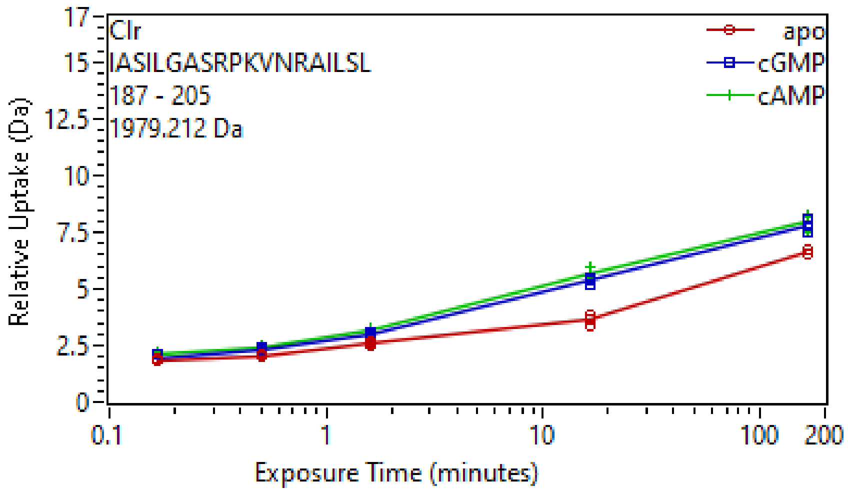

Clr
SQTDIASILGASRPKVNRAILSL
183 - 205
2410.378 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
17.5
21
0.1
1
10
100
200
Exposure Time (minutes)

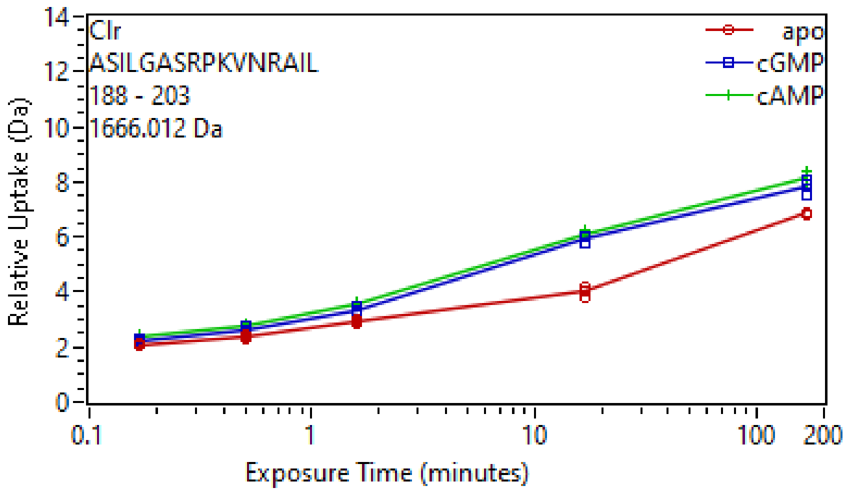

Clr
DIASIL
186 - 191
631.366 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
0.1
1
10
100
200
Exposure Time (minutes)

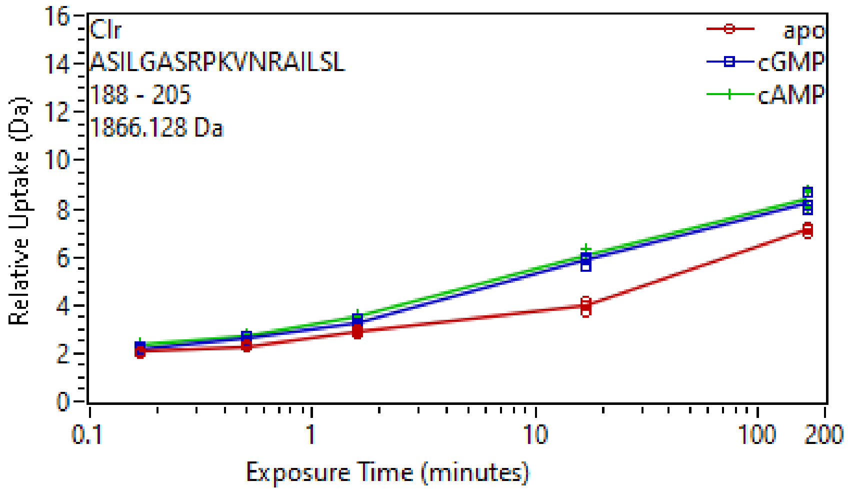

Clr
DIASILGASRPKVNRAIL
186 - 203
1894.123 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
14
16
0.1
1
10
100
200
Exposure Time (minutes)

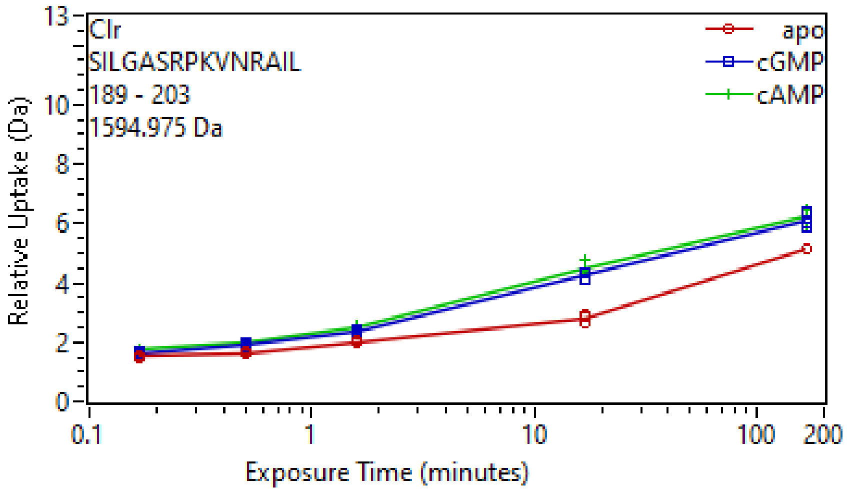

Clr
DIASILGASRPKVNRAILSL
186 - 205
2094.239 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
18
0.1
1
10
100
200
Exposure Time (minutes)

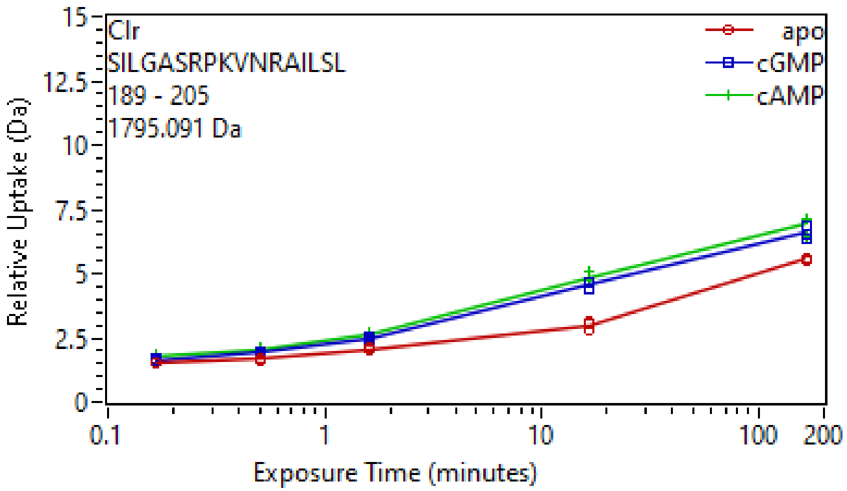

Clr
IASILGASRPKVNRAIL
187 - 203
1779.096 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
0.1
1
10
100
200
Exposure Time (minutes)

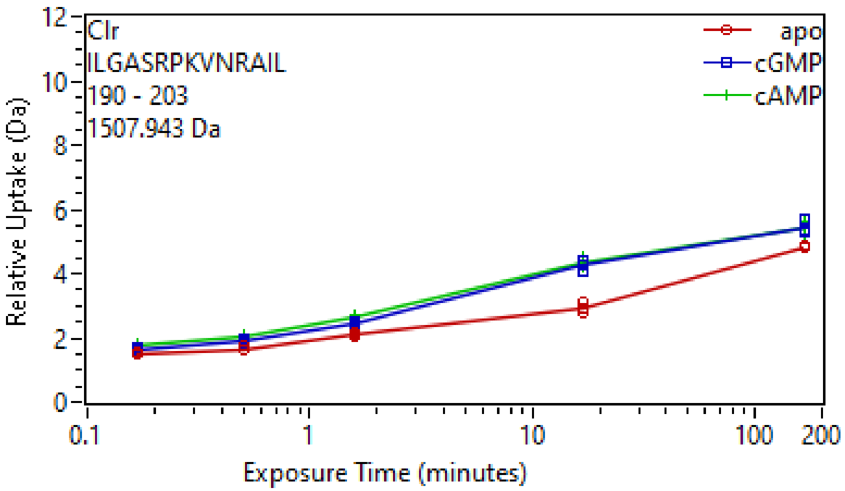

Clr
IASILGASRPKVNRAILSL
187 - 205
1979.212 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
17
0.1
1
10
100
200
Exposure Time (minutes)

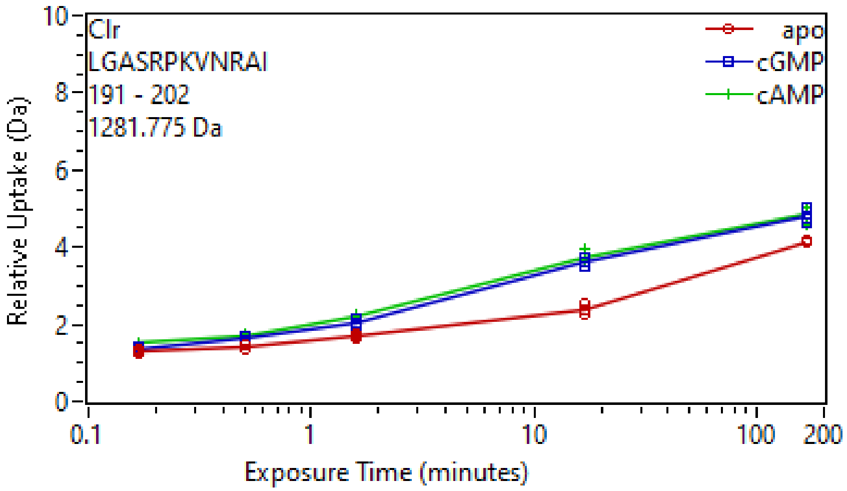

Clr
ASILGASRPKVNRAIL
188 - 203
1666.012 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
14
0.1
1
10
100
200
Exposure Time (minutes)

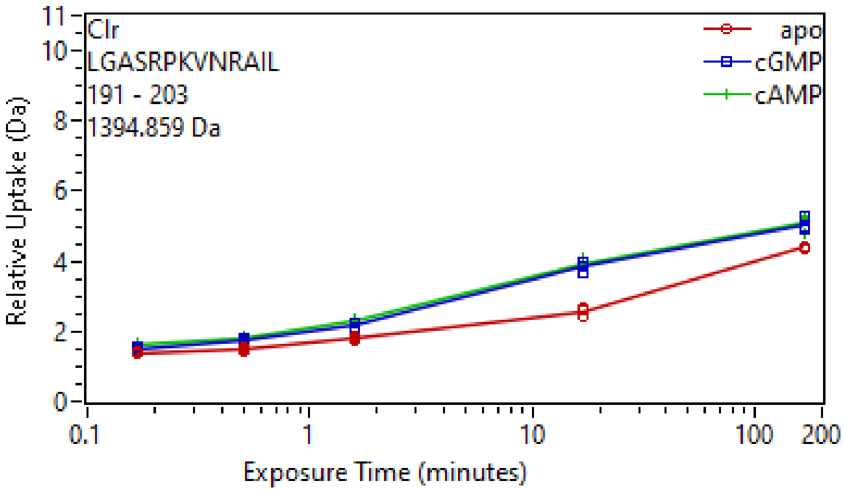

Clr
ASILGASRPKVNRAILSL
188 - 205
1866.128 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
14
16
0.1
1
10
100
200
Exposure Time (minutes)

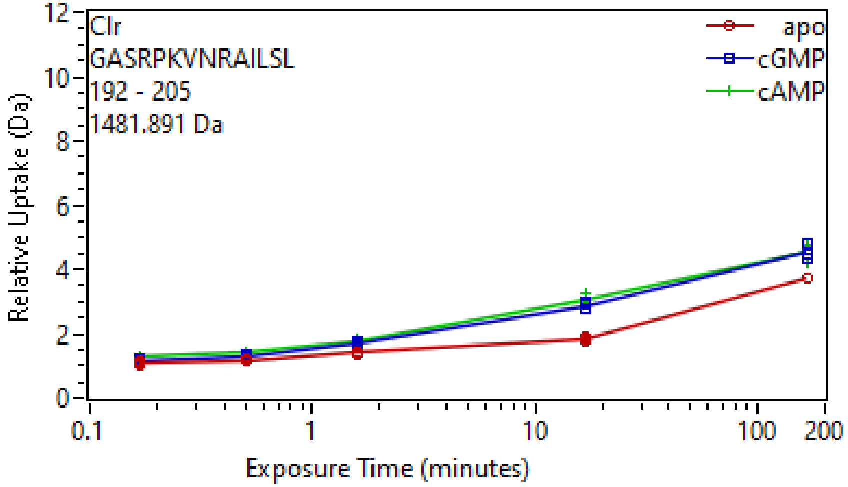

Clr
SILGASRPKVNRAIL
189 - 203
1594.975 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
13
0.1
1
10
100
200
Exposure Time (minutes)

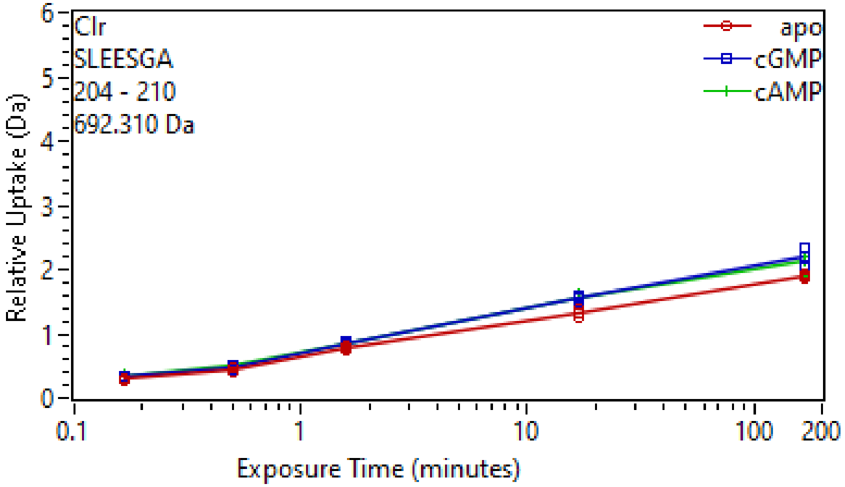

ClR
SILGASRPKVNRAILSL
189 - 205
1795.091 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
0.1
1
10
100
200
Exposure Time (minutes)

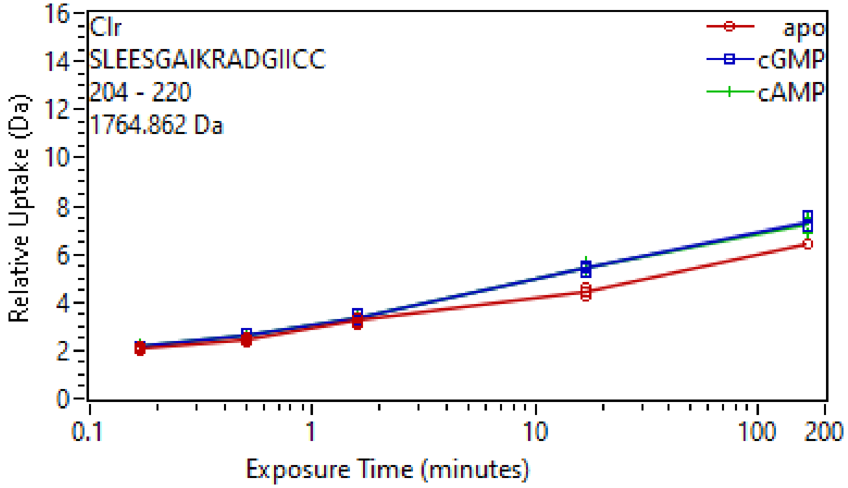

ClR
ILGASRPKVNRAIL
190 - 203
1507.943 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
0.1
1
10
100
200
Exposure Time (minutes)

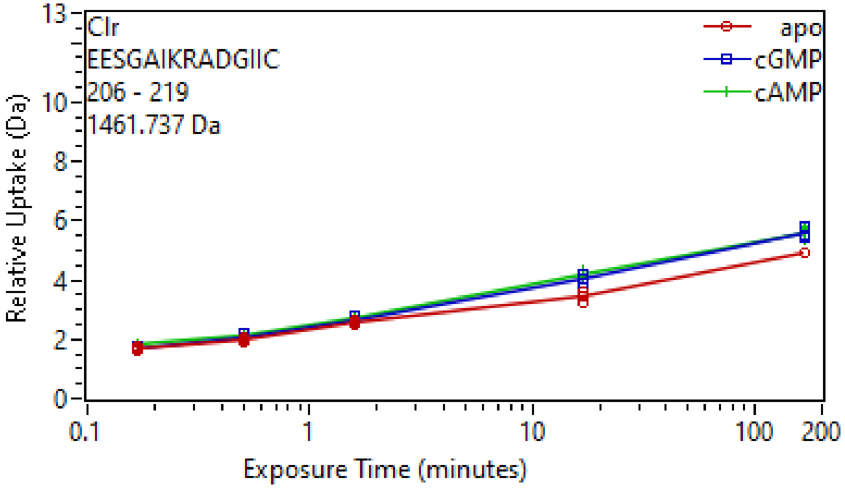

ClR
LGASRPKVNRAI
191 - 202
1281.775 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
0.1
1
10
100
200
Exposure Time (minutes)

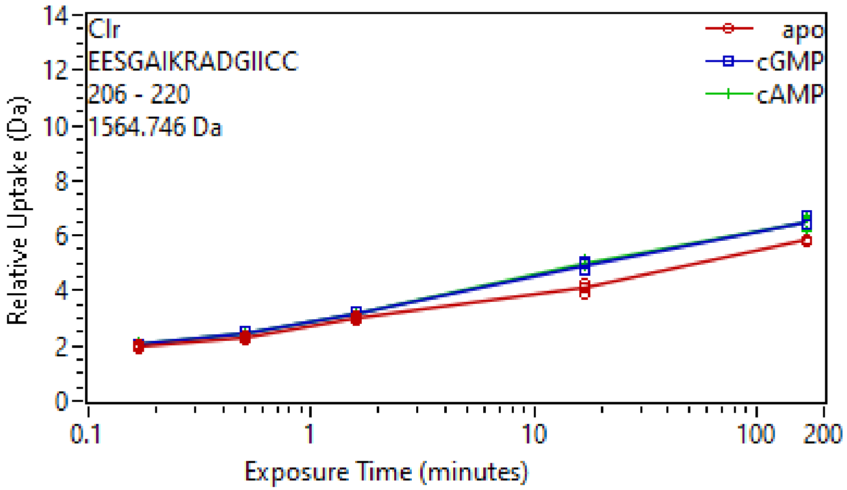

ClR
LGASRPKVNRAIL
191 - 203
1394.859 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
11
0.1
1
10
100
200
Exposure Time (minutes)

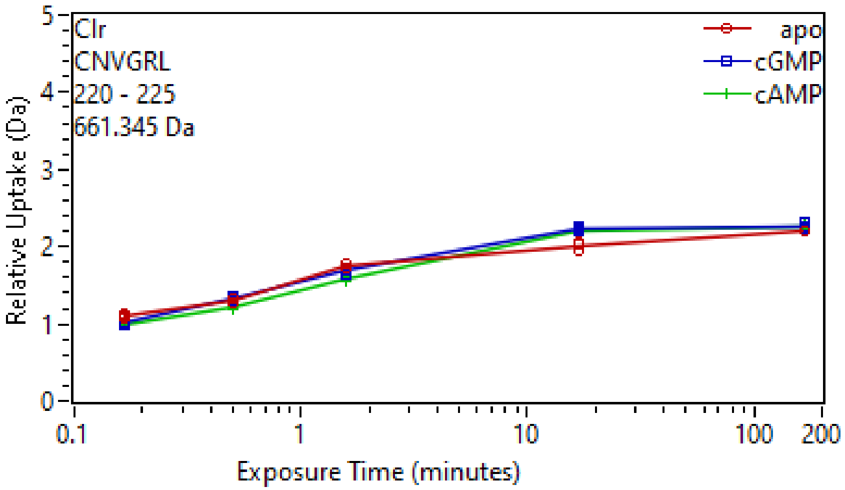

ClR
GASRPKVNRAILSL
192 - 205
1481.891 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
0.1
1
10
100
200
Exposure Time (minutes)

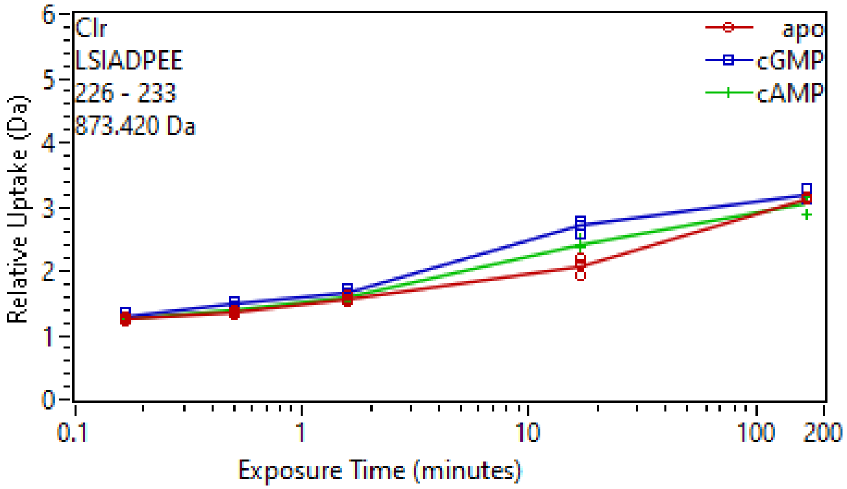

ClR
SLEESGA
204 - 210
692.310 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
0.1
1
10
100
200
Exposure Time (minutes)

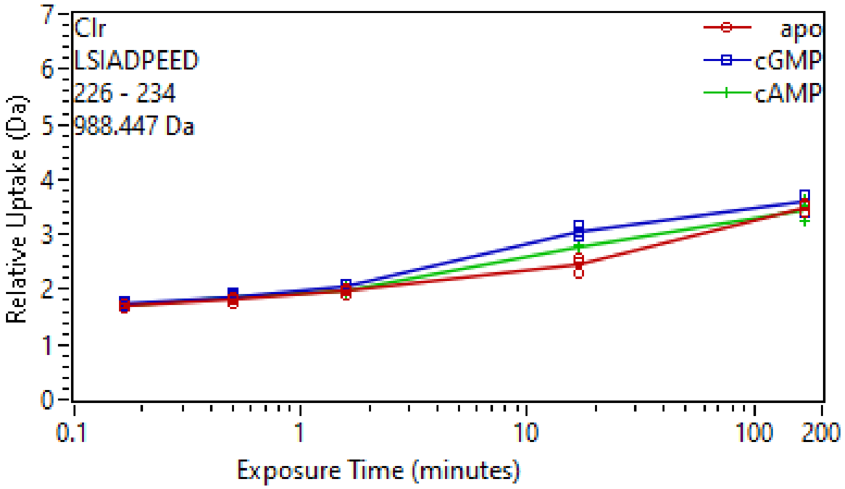

ClR
SLEESGAIKRADGIICC
204 - 220
1764.862 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
14
16
0.1
1
10
100
200
Exposure Time (minutes)

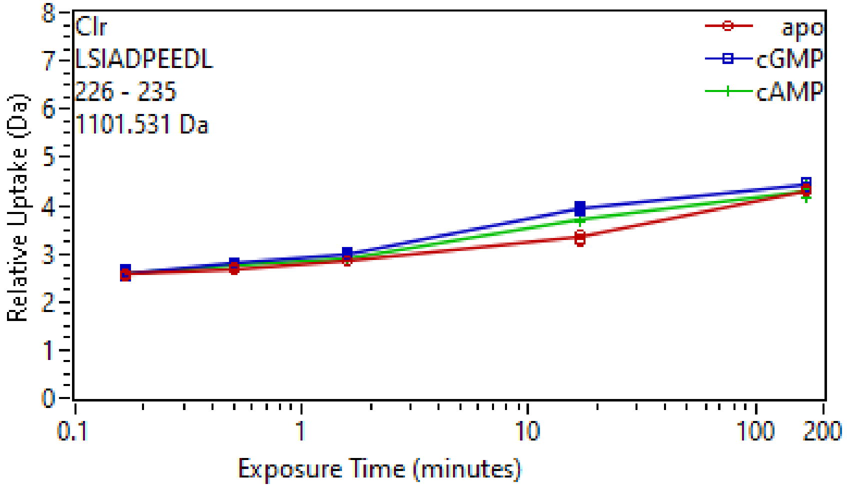

ClR
EESGAIKRADGIIC
206 - 219
1461.737 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
13
0.1
1
10
100
200
Exposure Time (minutes)

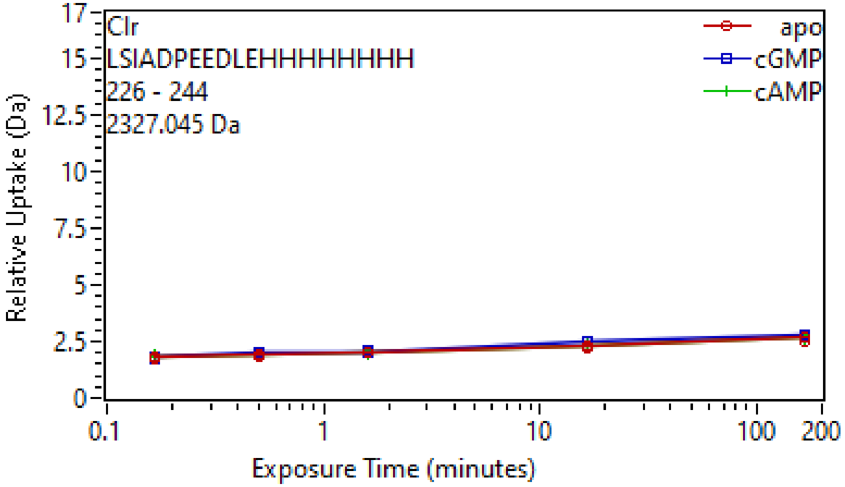

Clr
EESGAIKRADGIICC
206 - 220
1564.746 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
14
0.1
1
10
100
200
Exposure Time (minutes)

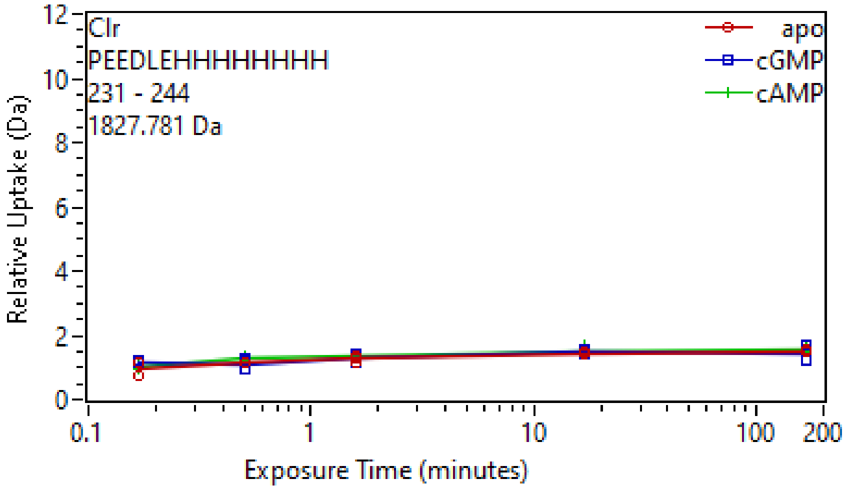

Clr
CNVGRL
220 - 225
661.345 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
0.1
1
10
100
200
Exposure Time (minutes)

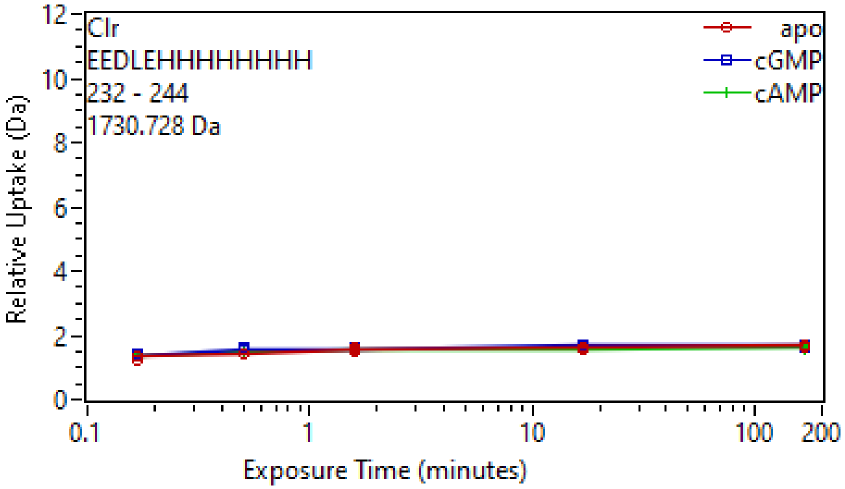

Clr
LSIADPEE
226 - 233
873.420 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
0.1
1
10
100
200
Exposure Time (minutes)

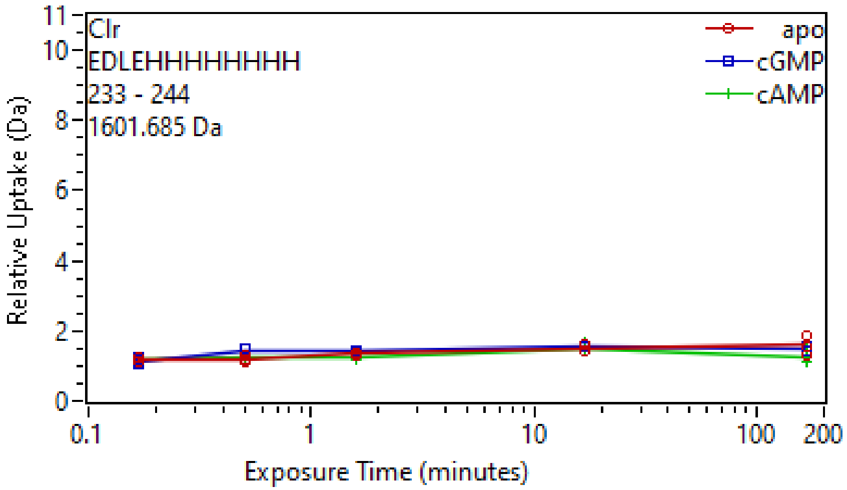

Clr
LSIADPEED
226 - 234
988.447 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
0.1
1
10
100
200
Exposure Time (minutes)


Clr
LSIADPEEDL
226 - 235
1101.531 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
1
2
3
4
5
6
7
8
0.1
1
10
100
200
Exposure Time (minutes)


Clr
LSIADPEEDLEHHHHHHHHH
226 - 244
2327.045 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2.5
5
7.5
10
12.5
15
17
0.1
1
10
100
200
Exposure Time (minutes)


Clr
PEEDLEHHHHHHHHH
231 - 244
1827.781 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
0.1
1
10
100
200
Exposure Time (minutes)


Clr
EEDLEHHHHHHHHH
232 - 244
1730.728 Da
apo
cGMP
cAMP
Relative Uptake (Da)
0
2
4
6
8
10
12
0.1
1
10
100
200
Exposure Time (minutes)


Clr
EDLEHHHHHHHH
233 - 244
1601.685 Da
apo
cGMP
cAMP
Relative Uptake (Da)
11
10
8
6
4
2
0
0.1
1
10
100
200
Exposure Time (minutes)

8.4 NMR analysis of potential *Sm*CpdA substrates

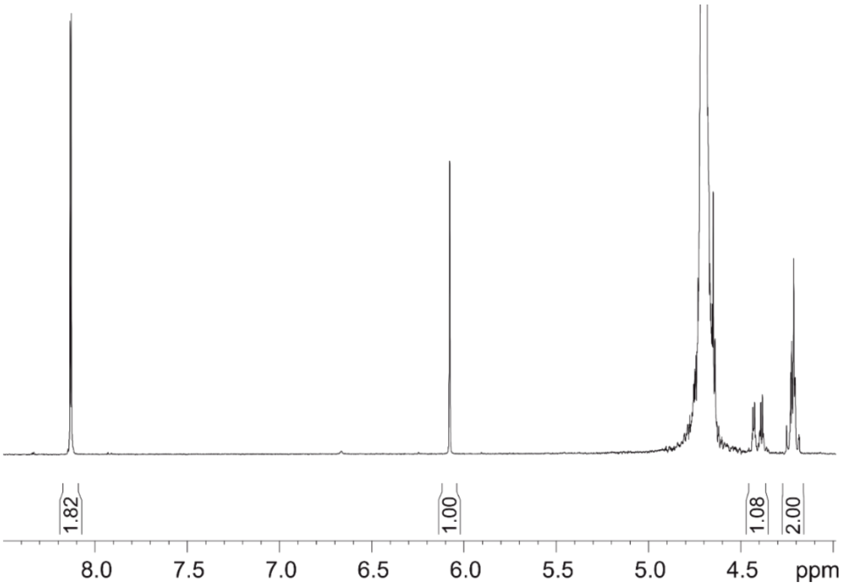


Figure 75: ^{1}H spectrum of 3’, 5’-cAMP 5.8 mM in D_2O at room temperature. The two singlets between 8.1 and 8.2 ppm correspond to the 2’H and 8’H of the adenine moiety, the single peak at 6.08 ppm is caused by the 1’H of the ribose. 2’H and 3’H are severely overlapped by the H_2O peak illustrating the need for a water free sample. The ddd at 4.4 ppm corresponds to 5’H and the following overlapping dd signals at 4.3 ppm are caused by the 5’’H atom and the 4’H. Additional 2D spectra used to verify the peak assignment are shown below.

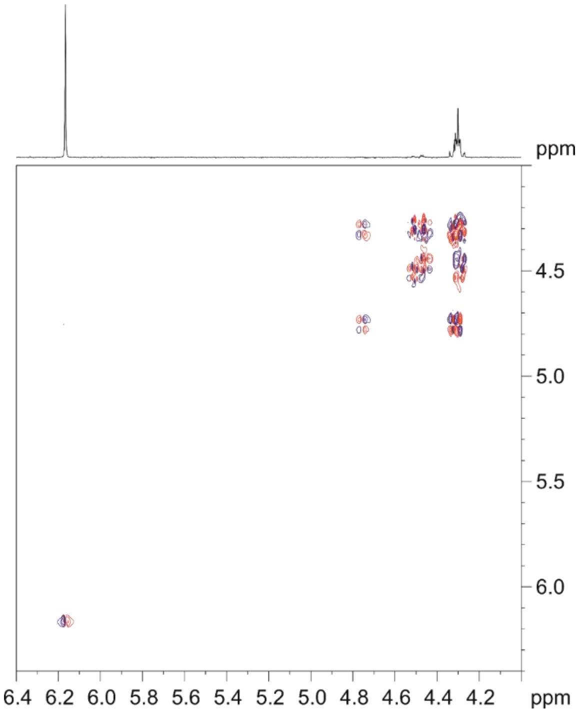


Figure 76: ^1H-^1H DQF-COSY spectrum of 3',5'cAMP. The H_2O signal was suppressed by Dr. Xiulan Xie using excitation sculpting. The spectrum reveals a strong interaction of the 5’H and 5’’H with each other. Additionally, 5’’H interact with 4’H and weakly with 3’H.

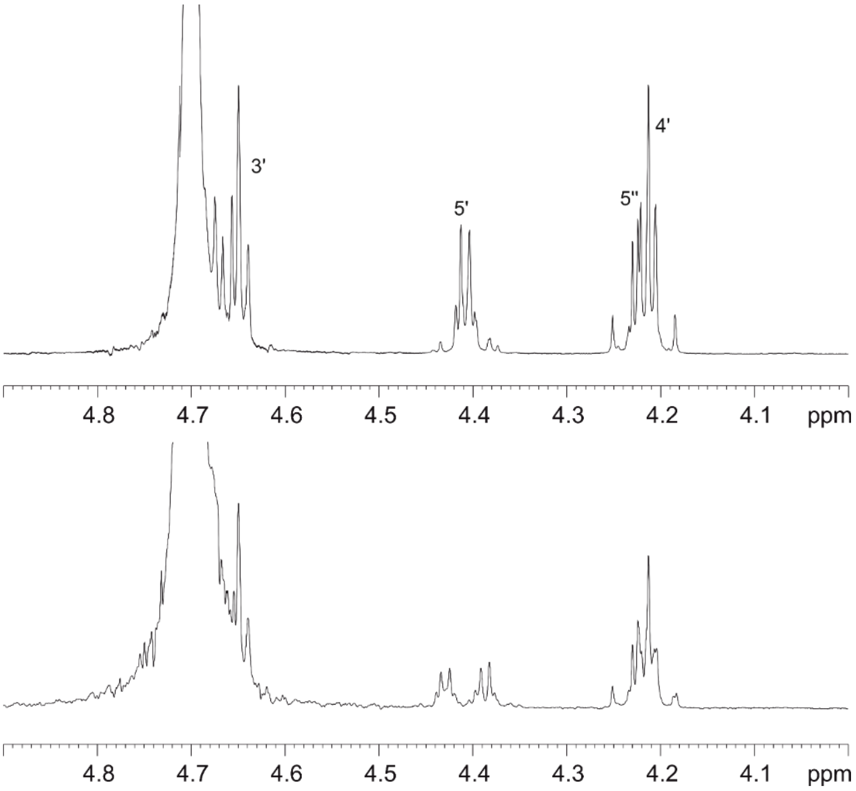


Figure 77: Comparison of ^1H spectra of 3’, 5’-cAMP with (upper trace) and without (lower trace) ^{31}P decoupling in D_2O at room temperature. Labels stand for signal assignment. The signal of the 5’H is reduced from a ddd to a dd by the decoupling.

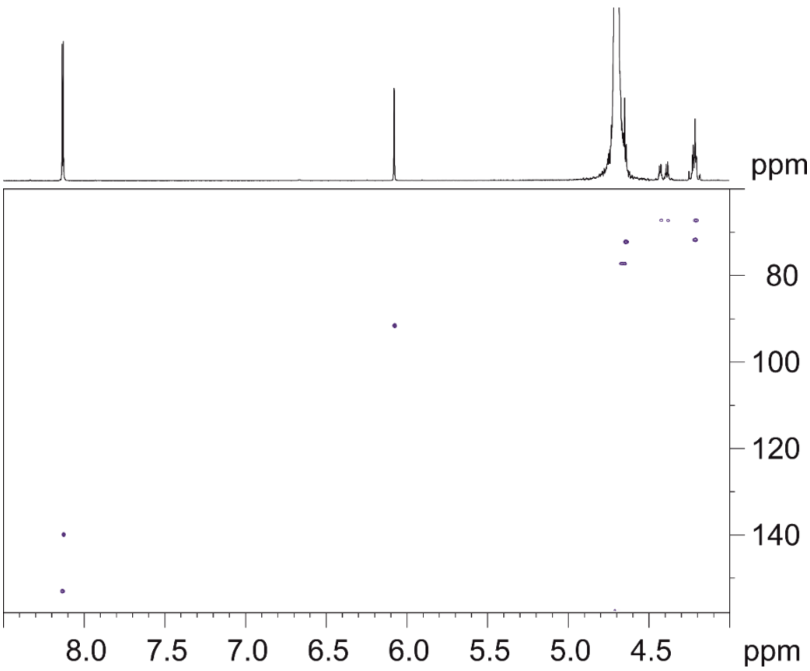


Figure 78. ^{1}H-^{13}C HSQC spectrum of 3’, 5’-cAMP in D_2O at room temperature. The spectrum can be used to assign the ^{13}C signals, as it reflects the coupling between both nuclei. It also clearly reveals the overlapping of one 5’’H signal with 4’H and the presence of the 2’H and 3’H within the water peak.

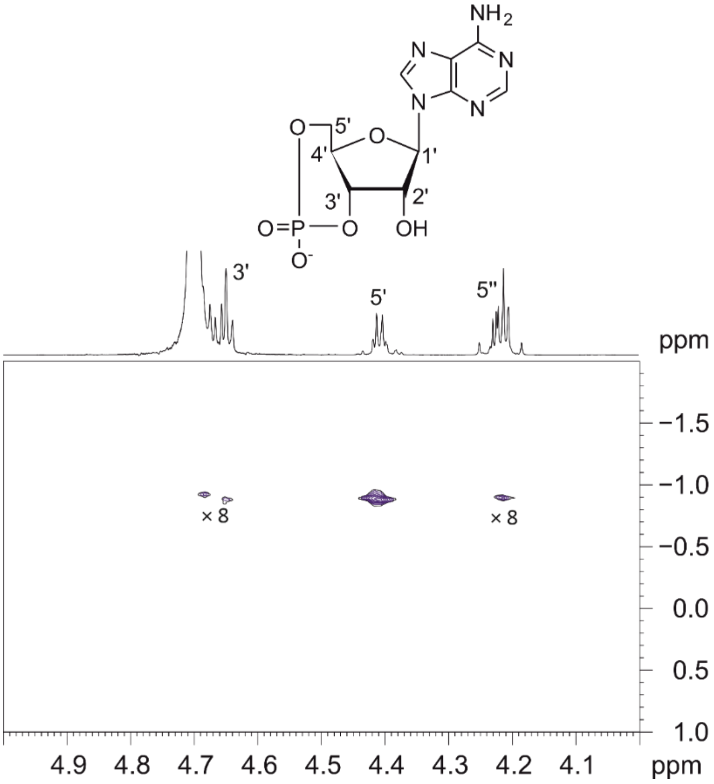


Figure 79: ^{1}H-^{31}P HMBC spectrum of 3’, 5’-cAMP in D_2O at room temperature. Long-range correlations were observed between P-atom and 5’H (strong), 5’’H (weak) and 3’H (weak). These are experimental evidence of the ring closure and the substitution positions at 3’ and 5’, as well.

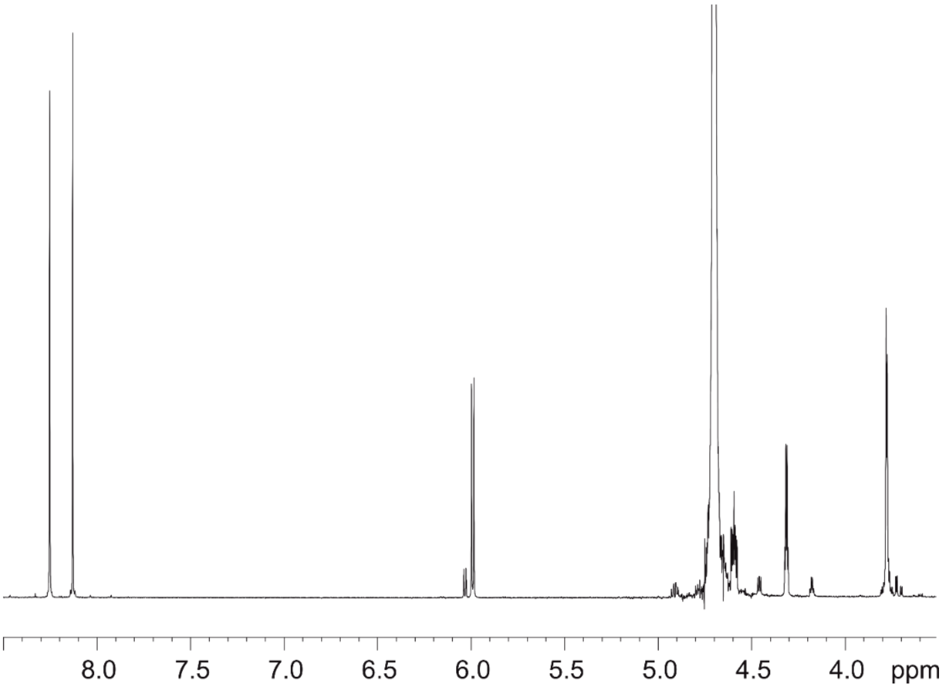


Figure 80: ^{1}H spectrum of a mixture of 2'- and 3'-AMP in D_2O at room temperature. 3'AMP is the major compound present in the sample and therefore elicits the stronger peaks. The peaks at 8.3 and 8.1 ppm belong to the aromatic hydrogens again and overlap for 2'- and 3'-AMP, the peaks around 6 ppm belong to the 1'H of the ribose. The remaining signals are overlapping the water peak in some cases and were assigned using the 2 D spectra shown below.

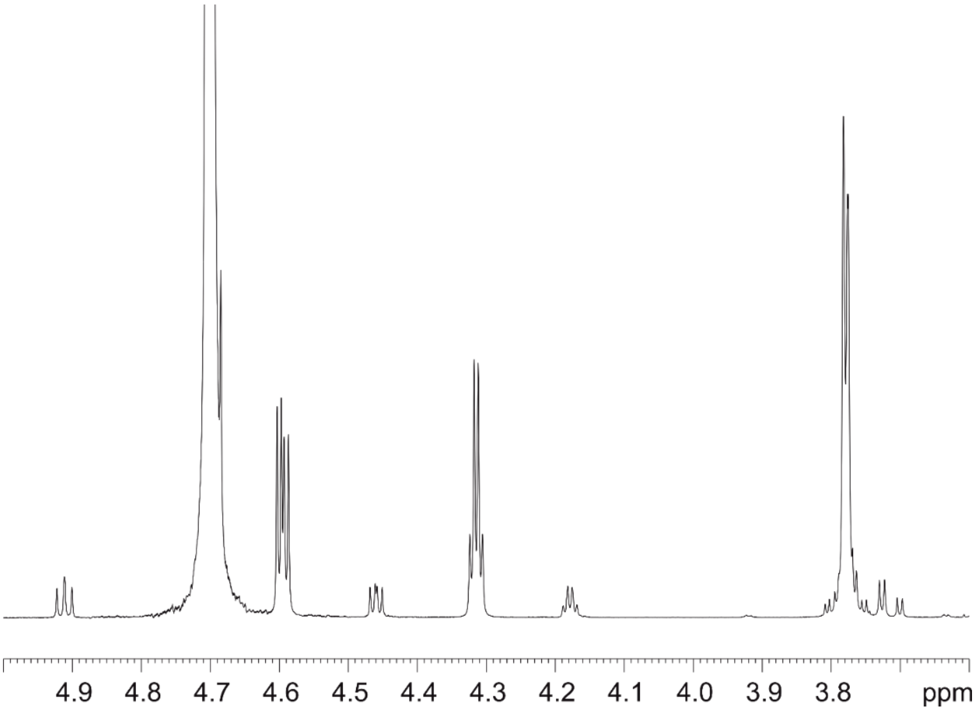


Figure 81: ^{1}H spectrum in the region 3.6 - 5.0 ppm of with ^{31}P decoupling of a mixture of 2'- and 3'-AMP in D_2O at room temperature. The minor peaks belong to 2'- and the major peaks belong to 3'-AMP.

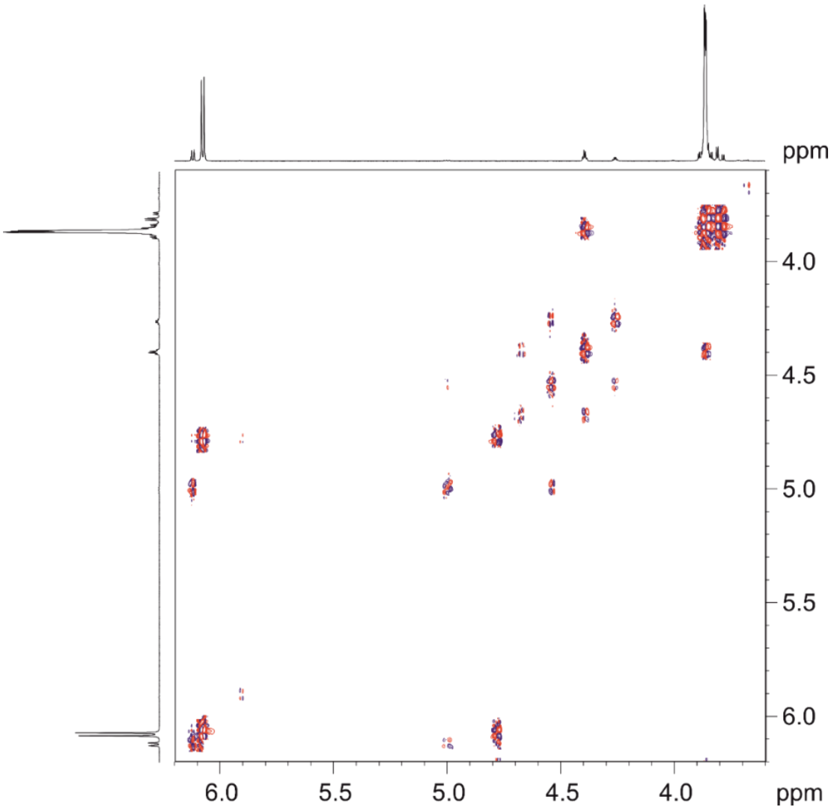


Figure 82: ^{1}H-^{1}H DQF-COSY spectrum in the region 3.6 - 6.2 ppm of a mixture of 2'- and 3'-AMP in D_2O at room temperature. The H_2O signal was suppression by Dr. Xiulan Xie using an excitation sculpting technique. The 1'H at 6.1 ppm interacts with the neighbouring 2'H at 4.8-5 ppm, 3'H is located around 4.6 ppm, and 4'H at 4.3-4.4 ppm couples with the 5'H located at 3.9 ppm.

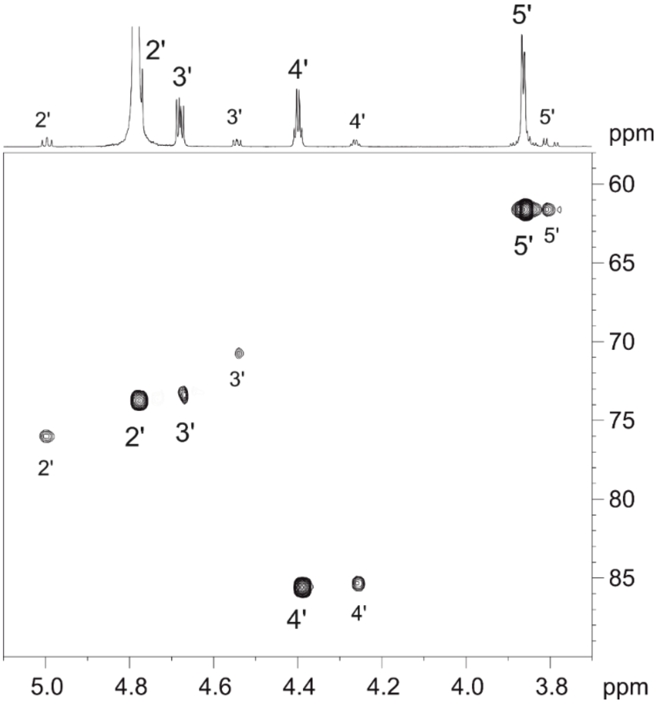


Figure 83: ^{1}H-^{13}C HSQC spectrum of a mixture of 2'- and 3'-AMP in D_2O at room temperature. The ^{13}C signals of the nucleotide are mostly overlapping. The exception are the 2'/3'H, which are shifted due to the proximity to the phosphate

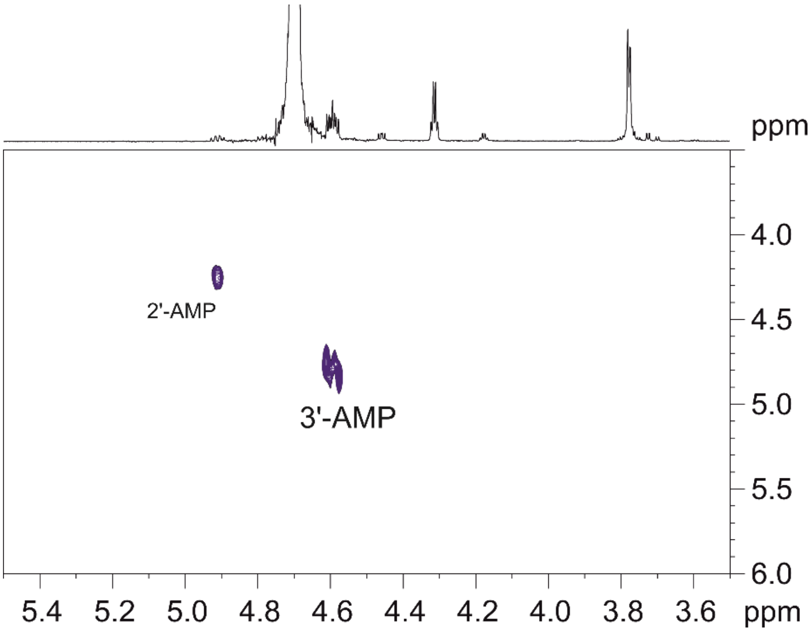


Figure 84: ^{1}H-^{31}P HMBC spectrum of a mixture of 2'-AMP and 3'-AMP in D_2O at room temperature. The major peak belongs to 3'-AMP and the minor peak belongs to 2'-AMP. The respective other hydrogen shows no ^{1}H-^{31}P coupling.

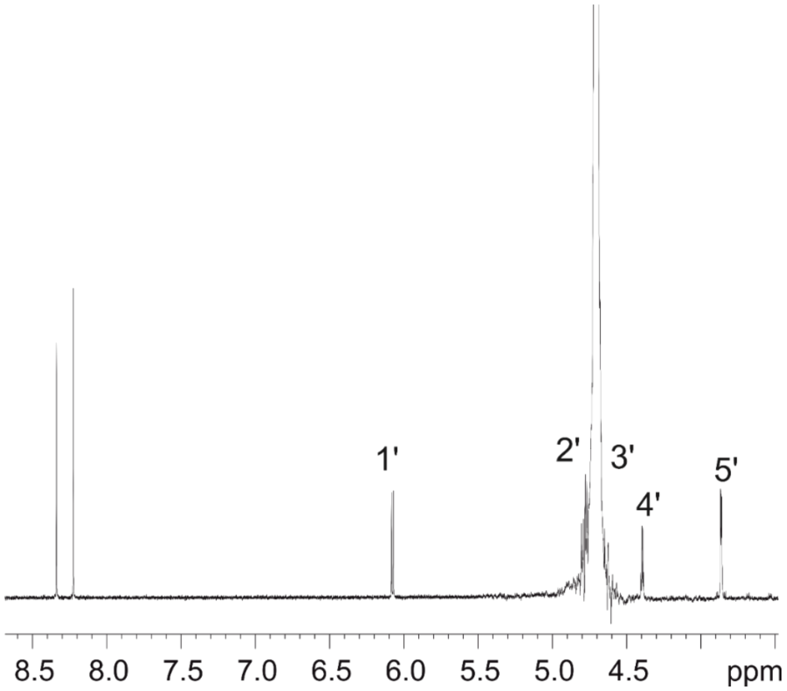


Figure 85: ^{1}H spectrum of 3'-AMP in D_2O at room temperature. The signals could be assigned based on the mixed spectra shown above. 2'H and 3'H severely overlap with the H_2O signal.

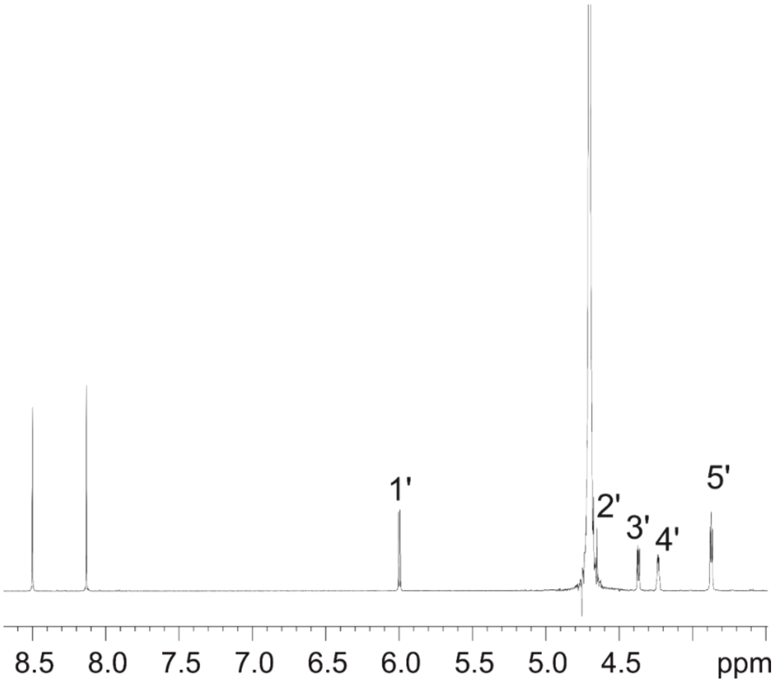


Figure 86: ^{1}H spectrum of 5'-AMP in D_2O at room temperature. The signals of the 2'H, 3'H and 3'H are shifted to lower ppm similarly to the 3'5'cAMP spectrum in comparison to the 2'- and 3'-AMP spectra.

8.5 CpdA kinetics in ITC measurements

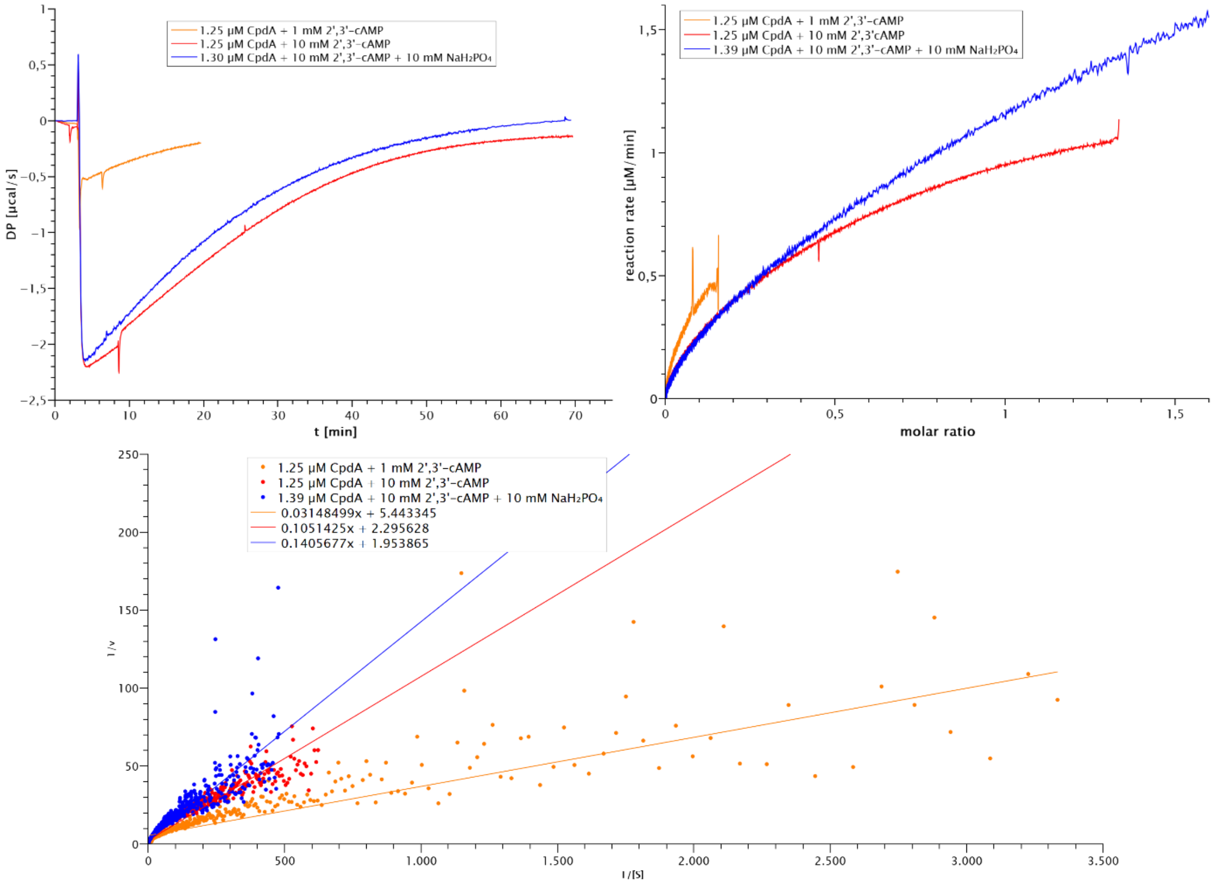


Figure 87: Raw thermograms for the determination of *Sm*CpdA kinetics using a single injection method. The curves never reach a defined baseline, which is reflected in the nonlinearity of the corresponding Lineweaver-Plot at the bottom. The Michealis-Menten Plot on the right similarly has no apparent V_{max}.

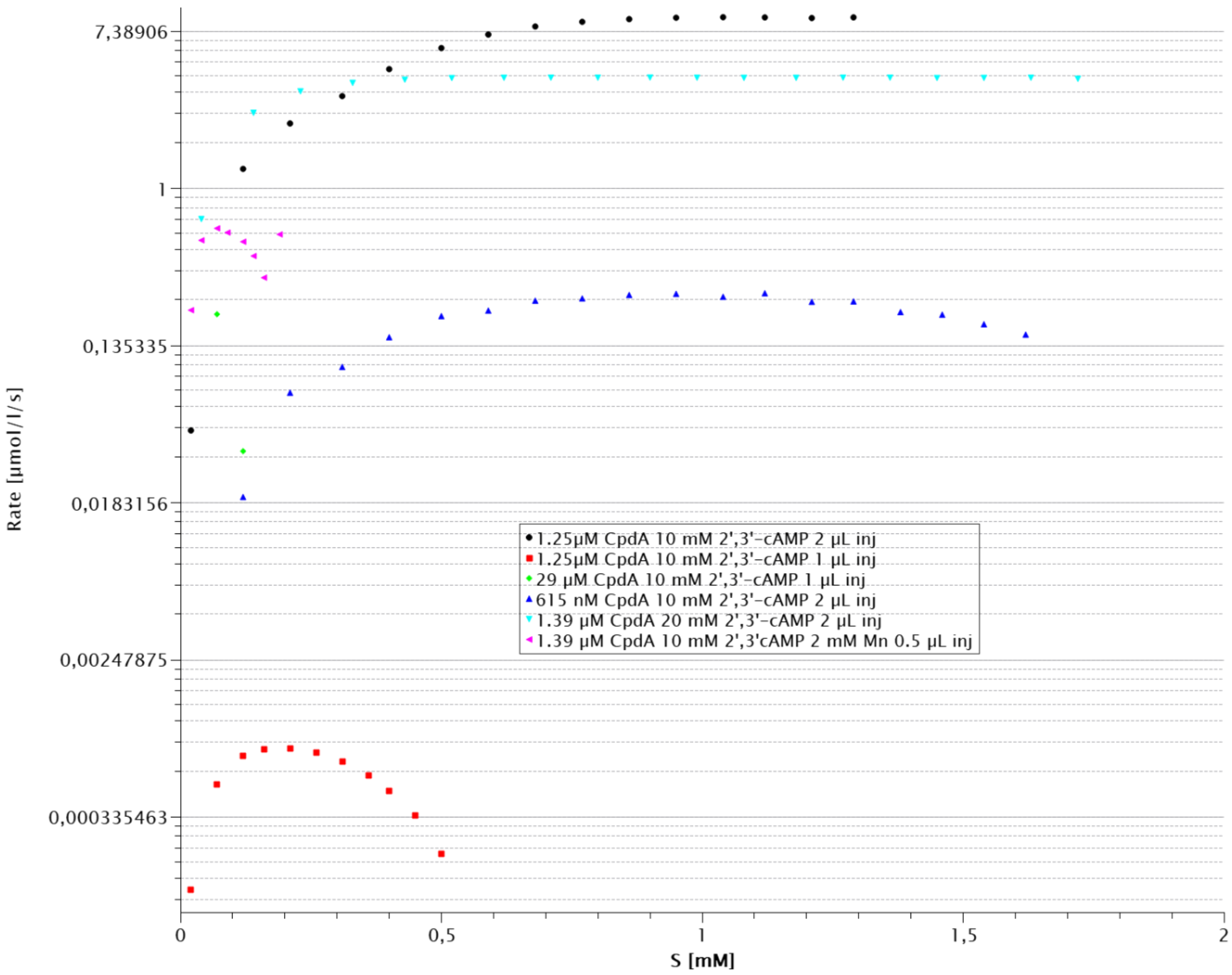


Figure 88: Raw thermograms for the determination of *Sm*CpdA kinetics using a multiple injection method. The curves deviate very strongly from each other, so that a logarithmic view had to be used to even fit them in the same graph. Additionally, the rate drops down after a few injections, so that no meaningful interpretation is possible.

8.5 CHASE2 homologs

Organism	UniProt-ID	Protein function	truncation	Length [AA]	MW [Da]	% unpolar AA
Sinorhizobium meliloti	Q92SD3	cyclase	45-270	225	25140	60.2

MGSSHHHHHHSQDPNSASPRSPDIIVVDVDRKAYEAHGGSWDRAATADLITRLAAAGPKTIAVDFVFSSACDPASPANAALAAAIGRAPVALGFLAADRVLARPRPVPPLALRRQLAVPAQWFIEGAETSCPAFMDRAKAATAAFLVGDEDARVRRVQAYAILGNDAYPALGIEAGRLAADSSTPVLGGEPAWLRLDHAIIPLDESGNLRFPASSGEAIAARTISAADILAGRLEKSRLAG

Organism	UniProt-ID	Protein function	truncation	length [AA]	MW [Da]	% unpolar AA
Thermosynechococcus elongatus	Q8DIP1	Serine/threonine protein kinase	84-376	293	32340	54.1

LELRVYDQWLRWRSTPATSQRLLIVEITEADIQTLKQYPVPDEVLIQAVNELQEYQPRVIGIDIFRDFPVP
DRFKLPTNGLPSLGRVMMTQPNTVIVCKAGSEGDPGIAPPAGLLNDQVGFADIPIDDDSVVRRAILATQP
EANDRCSTPQSFALALARLFLGVNPQAVTENRLELGTARFQSLTRNWGGYNNLDAAGFQILINYARPTQ
PYETVTLSEVLRGEVLPSKVRDRAVLIGLTGSSSNDKFLIPITLPEYTNRLTPGVVVQAAILEDLLAAALD
HRSPMGTWPQEA

gctagcctggaactgcgcgtttatgatcagtggctgcgttggcgcagtaccccggcaaccagtcagcgcctgctgattgttgaaattaccgaagccgatattcagac
cctgaaacagtatccggtgccggatgaagttctgattcaggcagtgaatgaactgcaggaatatcagccgcgtgtgattggtattgatatttttcgtgatttcccggttcc
ggatcgctttaaactgccgaccaatggtctgccgagtctgggccgtgttatgatgacccagccgaataccgttattgtttgcaaagcaggtagtgaaggtgacccgg
gcattgccccgccggcaggtttactgaatgatcaggtgggctttgccgatattccgattgatgatgatagtgtggttcgtcgtgccattctggcaacccagccggaag
ccaatgatcgttgcagtaccccgcagagttttgccctggcactggcccgcctgtttctgggcgttaatccgcaggccgtgaccgaaaatcgtctggaactgggcacc
gcccgctttcagagcctgacccgtaattggggtggctataataatctggatgccgcaggttttcagattctgattaattatgcacgtccgacccagccgtatgaaaccgt
taccctgagcgaagttctgcgtggtgaagtgctgccgagcaaagttcgtgatcgcgcagttctgattggcctgaccggcagtagtagcaatgataaatttctgattcc
gattaccctgccggaatataccaatcgtctgaccccgggtgttgtggtgcaggcagcaattctggaagatctgctggccgccgccctggatcatcgcagtccgatgg
gcacctggccgcaggaagcataactcgag

Organism	UniProt-ID	Protein function	truncation	length [AA]	MW [Da]	% unpolar AA
Myxococcus xanthus	Q1D4J2	Adenylate/guanylate cyclase	48-365	317	33495	63.1

ERPLYDQAVSRLLPAVPPSADLVLVEVDDRALATLSERWPLSRTTWARAFHALAAQRPAAVAVDVVF
DQPGPRDALELGEDILEALRRSGLTEQPAGEALAADLEARLLARDGDARLAEAISEGNGVVLGAAALT
DDVPVMAPLQDGALGTPLALPAHALRLQAREVAGSIAPLRMAARGSGTLNMLVDGDGVIRRYPYAVG
VSGQAWPSLALATALHLTPERAEPLRELAAWDHGAPLMRLPAPNWLPRVSLADLLHADPRSVGLDLA
LRGKTVFVGVTATGLHGQSTLPGQVAVPGVEIHAFALDNLRSGRLMRSS

gctagcatggaacgtccgctgtatgatcaggcagttagtcgcctgctgccggcagttccgccgagtgctgatctggtgctggttgaagtggatgatcgtgccctggc
caccctgagtgaacgttggccgctgagtcgcaccacctgggcccgtgcttttcatgccctggccgcccagcgtccggcagcagtggcagttgatgttgtgtttgatc
agccgggtccgcgcgatgccctggaactgggtgaagatattctggaagccctgcgtcgtagtggtctgaccgaacagccggcaggtgaagcactggccgcagat
ctggaagcacgcctgctggcccgtgatggtgacgcacgtctggccgaagccattagcgaaggtaatggcgttgtgctgggcgccgcagcactgaccgatgatgt
gccggttatggcaccgctgcaggatggtgcactgggtaccccgctggcactgccggcacatgccctgcgtctgcaggcacgcgaagttgccggcagtattgcac
cgctgcgtatggcagcacgcggtagcggtaccctgaatatgctggtggatggcgatggcgtgattcgtcgctatccgtatgccgttggtgtgagcggccaggcctg
gcctagcctggcactggctaccgcactgcatctgaccccggaacgtgccgaaccgctgcgtgaactggccgcatgggatcatggcgcaccgctgatgcgtctgc
cggcacctaattggctgccgcgtgtgagtctggccgatctgctgcatgcagatccgcgtagcgtgggtctggatctggcactgcgcggcaaaaccgtttttgttggc
gttaccgcaaccggtctgcatggccagagcaccctgccgggccaggtggcagtgcctggtgttgaaattcatgcctttgcactggataatctgcgcagtggtcgtct
gatgcgcagtagttaactcgag

Organism	UniProt-ID	Protein function	truncation	length [AA]	MW [Da]	% unpolar AA
Synechococcus elongatus PCC 6301	A0A0H3K3Y5	Adenylate/guanylate cyclase	36-303	267	29472	56.3

TDRALQQQLLLLQGTRPAPNDLLIVKIDDASFDQAAFFRDDPRAPAWAQDLEGFPWPRAIYAQIIDRLF
AAGARAIAIDLLFVTPSAFGPADDAALAQRLQRYGDRISLATEYREPDGSGGGLSLGSLTETLGPNLRVG
YINTLPADNGQILDHPARYRETVVEPLGLPLLPSLPETLQTQQSPARALRYYGPRGHLPQISAWQVLDDA
EWQRIQGRVRDRLILIGPTAASLQDQHATPWGVQAGVEILATALANDRDRSGLRFWPQS

gctagcatgaccgatcgcgcactgcagcagcagctgctgctgctgcagggcacccgccctgctcctaatgatctgctgattgttaaaattgatgacgcaagctttgatcaggccgcctttttccgcgatgatccgcgcgcaccggcatgggcacaggatctggaaggttttccgtggccgcgtgccatctatgcacagattattgatcgtctgtttgcagcaggcgcacgtgcaattgcaattgatctgctgtttgttaccccgagcgcatttggcccggccgatgatgcagccctggcccaacgcctgcagcgttatggtgaccgtattagtctggcaaccgaatatcgtgaaccggatggtagtggcggtggcctgagtctgggtagcctgaccgaaaccctgggtccgaatctgcgcgttggctatattaataccctgccggcagataatggccagattctggatcatccggcacgctatcgtgaaaccgtggttgaaccgctgggcctgccgctgctgccgtcattaccggaaaccctgcagacccagcagagtccggcacgtgcactgcgctattatggtccgcgtggccatctgccgcagattagcgcatggcaggtgctggatgatgcagaatggcagcgcattcagggccgcgttcgcgatcgcctgattctgattggcccgaccgcagcaagtctgcaggatcagcatgccaccccgtggggcgtgcaggctggtgttgaaattctggccaccgcactggccaatgatcgtgatcgtagcggcctgcgcttttggccgcagagttaactcgag

Organism	UniProt-ID	Protein function	truncation	length [AA]	MW [Da]	% unpolar AA
Deinococcus radiodurans	Q9RTW9	GGDEF family protein	38-272	234	24974	58.7

HNRALENSLNRALPSRFDERLLLVGIDDATLGDYGPLPQWPRSLYLRAFDTLEQAGAQVVGLDVVLGRSSDPDPAVQAAFDRPHLVLATAPGDPVLDWLSARQTVTGVSALNSGSGPLREVQTAYPVGSGPATVLIPSFARQVAAQAGQPLPLDTQPRLLRYSPPAQVAAQTLSFRDLVGGNVRYADLQGRVVLIGQNASGVPGLNLPDIDGRLTPGLEWQARAVSSLLSAPLRR

gctagcatgcataaccgcgcactggaaaatagtctgaatcgtgcactgccgagtcgttttgatgaacgtctgctgctggttggcattgatgatgcaaccctgggtgactatggtccgctgccgcagtggccgcgtagtctgtatctgcgtgcctttgataccctggaacaggccggcgcacaggttgtgggtctggatgtggttctgggccgtagtagtgatccggatccggccgttcaggcagcatttgatcgtccgcatctggtgctggcaaccgcaccgggtgacccggttctggattggctgagtgcacgtcagaccgttaccggtgtgagcgcactgaatagcggcagcggcccgctgcgtgaagttcagaccgcatatccggttggcagtggcccggccaccgtgttaattccgagttttgcacgtcaggttgccgcacaggcaggccagccgctgcctttagatacccagccgcgtctgctgcgctatagtccgccggcacaggtggccgcacagaccttaagttttcgtgatctggtgggtggtaatgttcgttatgcagatctgcagggtcgcgttgttctgattggccagaatgcaagcggtgtgccgggtctgaatctgccggatattgatggccgcctgaccccgggtctggaatggcaggcacgtgcagttagcagtctgctgagtgcaccgctgcgccgctaactcgag