



THE UNIVERSITY
of ADELAIDE

**School of Biological Sciences
Department of Molecular and Biomedical Science**

Structural Characterization of Lysogeny-promoting Transcription Factors of Bacteriophage 186

Thesis submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy

Jia Quyen Truong

16th December 2020

Supervisors: Keith Shearwin, Ian Dodd & John Bruning

DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Jia Truong

15th December 2020

ACKNOWLEDGEMENTS

This thesis is dedicated to my parents. I extend my deepest gratitude to my parents who have supported my every endeavour. You are the sunshine of my life and I am so blessed to have been your son.

During my time at the University of Adelaide, I was fortunate enough to meet many people that have molded me into the scientist I am today. I am deeply indebted to my supervisors Dr. Keith Shearwin, Dr. Ian Dodd and Dr. John Bruning for all their guidance, mentorship and patience. My successes, including the completion of this thesis, would not have been possible without their support and encouragement. My primary supervisor, Dr. Shearwin, has always encouraged me to take part in and funded my travels to conferences and interstate training. He has given me much support to learn and blossom as a scientist. He is an excellent supervisor and I am lucky to have chosen him. I thank him for bringing me into your laboratories. I have thoroughly enjoyed my stay. To all the past and present members of the Shearwin/Dodd and Bruning labs, I wish to express my gratitude for contributing to the positive and supportive lab environment. In particular, I wish to thank Dr. Andrew Hao, Dr. Iain Murchland and Dr. Andrew Marshall for teaching me in my undergraduate placements and starting me off on my scientific career.

I want to thank my partner, Stephanie Nguyen, who I met during my PhD. Thank you for all the support and I appreciate all the snacks that magically appear on my tabletop. You are an amazing scientist and human being, who inspires the people around you, including me.

I also had the great pleasure of working with many other talented scientists at the University of Adelaide, the Australian Synchrotron and the University of Wollongong.

ABSTRACT

Bacteriophage 186, like bacteriophage λ , is a UV-inducible temperate bacteriophage of *Escherichia coli*. At the DNA sequence level, these two phages show little similarity. The arrangements of promoters, genes and regulatory elements in the genome of the two phages are different, yet show similarity in the regulatory network topology. In fact, proteins with equivalent functions can be pinpointed and previous studies have shown that these proteins act as key effectors in the lytic-lysogenic decision. Studies described in this thesis were aimed at investigating these bacteriophage 186 proteins to understand their structure-function relationship. These proteins regulate transcriptional behavior and may be useful as parts for the construction of synthetic circuits.

In 186, the CII protein (analogous to λ CII) is responsible for establishment of lysogeny. It is a transcriptional activator that upon binding to a pair of half sites at the 186 p_E promoter, activates it to generate the initial levels of immunity repressor to allow the phage to enter lysogeny. Its unusual property of binding to half sites two DNA turns apart and its potent ability to activate p_E over 400-fold, suggests it may possess a novel mechanism of promoter activation.

The X-ray crystal structure of the 186 CII protein was solved. It revealed the protein adopts a tetrameric quaternary structure, consisting of rigid dimeric subunits, which further dimerize to form a tetramer. Through mass-spectrometry measurements and mutagenesis studies, it was demonstrated that this tetrameric state is necessary for its function. Molecular modelling with the crystal structure provides insight into how CII may bind DNA and interact with the host RNA polymerase to activate its promoter.

The role of CII is to generate initial pools of 186 CI (analogous to λ CI). 186 CI is responsible for repressing the lytic functions of 186 to maintain lysogeny. Previous studies have shown that CI dimers multimerize into a higher order structure that interacts with DNA. High resolution structures of a CI dimer and higher order CI CTD oligomer, but no high-resolution structure of the full complex, are available. We used a combination of X-ray crystallography, small angle X-ray scattering and mass-spectrometry to demonstrate that 186 CI forms a wheel-like dodecameric structure that could act as a scaffold for DNA wrapping, giving insight into the mechanism of transcriptional regulation provided by 186 CI.

Following the structural characterization of 186 CI, a temperature sensitive mutant of 186 CI was characterized to show that it can be used for temperature-based induction of gene expression. In addition, a temperature-sensitive chimeric repressor containing domains from 186 CI and λ CI was developed. This success illustrates the ability of bacteriophage regulatory proteins to be used as a source of biological parts.

X-ray crystallography was extensively used in this PhD project to structurally characterize proteins. When solving novel protein structures, the generation of diffraction quality crystals and the derivatization of protein crystals with heavy atoms for experimental phasing are two commonly encountered bottlenecks. In the final section of this thesis, a technique to simultaneously optimize crystallization and derivatization of a protein is presented. Crystalline precipitate of a protein was crushed up to form a microseed stock, which was used to promote crystallization by adding it to a sparse matrix crystallization screen of the protein. An iodinated compound I3C was added to the crystallization screen, to allow incorporation of I3C into the crystal during the crystallization process. This technique was employed to solve the crystal structures of Hen-Egg White Lysozyme and a putative lysin domain Orf11 from bacteriophage P68 and should be applicable to many other crystallization targets.

TABLE OF CONTENTS

1	Introduction	1
1.1	Bacteriophage λ	2
1.1.1	Lytic gene cascade of bacteriophage λ	2
1.1.2	Establishment of lysogeny in bacteriophage λ by λ CII	2
1.1.3	Maintenance of lysogeny via λ CI	5
1.1.4	Summary	9
1.2	Bacteriophage 186	9
1.2.1	Lytic gene cascade in bacteriophage 186	9
1.2.2	Diverting towards the lysogeny cycle in bacteriophage 186	10
1.2.3	The 186 CII protein	11
1.2.4	Maintenance of lysogeny with 186 CI	13
1.2.5	DNA recognition by 186 CI	14
1.2.6	Binding to multiple operators	15
1.2.7	Summary	17
1.3	References	17
2	Structural and biochemical characterization of the CII protein	21
3	Structural studies of 186 CI using X-ray crystallography, small angle X-ray scattering and mass spectrometry	82
3.1	Introduction	82
3.2	Expression of the 186 CI repressor	83
3.3	Purification	84
3.4	Crystallization	85
3.5	X-ray data collection and processing	85
3.6	Attempts to solve the phase problem	87
3.6.1	Principles of molecular replacement	87
3.6.2	Attempt to solve phase problem	87
3.6.3	Reasons that may contribute molecular replacement failing	88
3.7	SAXS	88
3.8	Preparation of 186 CI protein for SAXS	89
3.9	SAXS data acquisition and processing	89
3.10	SAXS results	91

3.10.1	Kratky analysis	91
3.10.2	Guinier analysis	93
3.10.3	Pair distance-distribution function P(r).....	97
3.10.4	Ensemble optimization method to examine flexibility in CI ^{E146K}	101
3.11	Discussion of SAXS data	104
3.12	Ion-mobility mass spectrometry	105
3.13	Mass spectrometry experimental procedures	108
3.13.1	Preparation of samples for mass spectrometry	108
3.13.2	Mass-spectrometry examination of 186 CI-CTD and 186 CI	108
3.14	Oligomerization of 186 CI	113
3.14.1	Previous studies of 186 CI oligomeric state	113
3.14.2	An alternative crystallography approach to study 186 CI-CTD oligomerization....	115
3.15	References.....	116
4	Expression, purification, crystallization and structure solution of a hybrid CI protein.....	121
4.1	Introduction	121
4.2	Materials and Methods	123
4.2.1	Overexpression of the hybrid protein	123
4.2.2	Preparation for protein-DNA complex for crystallography	125
4.2.3	X-ray data collection, processing and structure solution	126
4.2.4	Analysis of structure	128
4.2.5	Mass spectrometry characterisation of CI-hybrid protein	128
4.3	Results	129
4.3.1	Structural analysis of the CI-hybrid protein	129
4.3.2	Mass spectrometry analysis of the CI-hybrid protein	133
4.4	Discussion	136
4.4.1	Alternative wheels configurations mediated by 186 CI CTD	136
4.4.2	Dimer-dimer interface analysis	138
4.4.3	Model of transcription regulation using a dodecamer	139
4.5	Conclusion	144
4.6	References.....	144
5	Construction of temperature sensitive expression systems	148
6	Novel method to derivatize crystals during initial crystallography screening trials	179
7	Conclusions and future directions.....	215

7.1	Bacteriophage 186 as a model for a genetic switch	215
7.1.1	Understanding the function of CII in establishment of lysogeny	215
7.1.2	Maintenance of lysogeny by CI	217
7.2	Temperature sensitive repressors from phage components	218
7.3	Random microseed matrix screening method to derivatize protein crystals.....	220
7.4	References	221
Appendix A	Review on lysogeny	225
Appendix B	Experimental determination of 186 CI extinction coefficient	238

1 INTRODUCTION

In this post-genomic era, one of the challenges to tackle in molecular biology is explaining the complex regulation of large genomic networks. Examples include understanding cell differentiation decisions in mammalian development and how states of dysregulation can cause diseases such as cancer. The study of these complicated systems could be aided by studying a simpler biological system with well characterized components to better understand the underlying principles. A temperate bacteriophage is one such system. Although they are one of the simplest biological systems, they can make seemingly rational decisions between alternate developmental regimes based on environmental signals.

Bacteriophage λ , the most extensively characterized bacteriophage, has provided many valuable insights into gene regulation. After infecting its host, it can make a decision between the lytic and lysogenic modes of reproduction, based on environmental cues. In the lytic mode of infection, the bacteriophage coaxes the bacterial host to produce new phage particles, ultimately to release them via cell lysis. Alternatively, it can enter the lysogenic state, where the λ genome is integrated into the host chromosome and replicates along with host replication. The lytic functions of phage λ are repressed in this non-bactericidal state, which is stably maintained through bacterial generations. At a later point, bacteriophage λ can make a second decision to exit the lysogenic state and reenter lytic development. This process often occurs in response to DNA damage (Ptashne, 2004; Oppenheim *et al.*, 2005). For a more in-depth review of lysogeny, see Shearwin and Truong (Shearwin and Truong, 2019) (Appendix A).

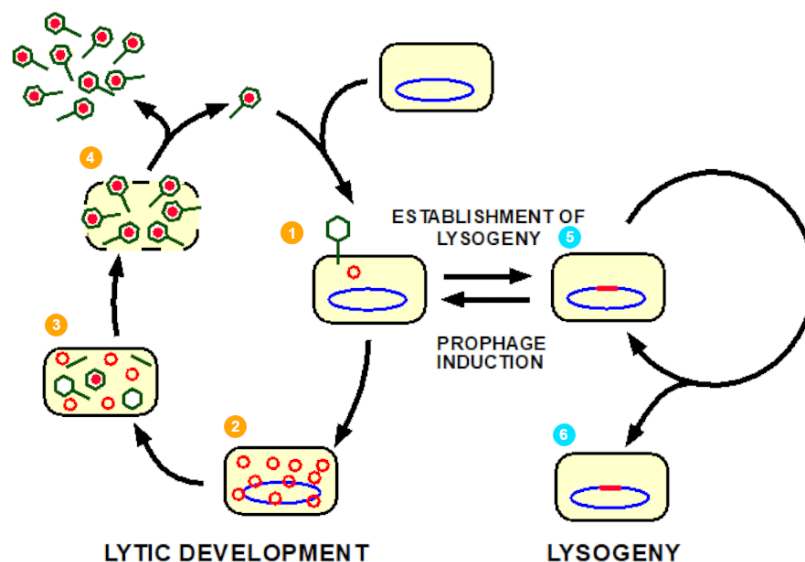


Figure 1-1 - A temperate bacteriophage can choose between the lytic and lysogenic developmental regimes. A bacteriophage infects a host bacterium by injecting its DNA into the host cell (step 1). In the lytic lifecycle, the host cellular machinery is hijacked to replicate the phage genome (step 2), produce new virions (step 3), lysing the cell to release it (step 4). In the lysogenic state, depicted with blue numbering, the phage DNA integrates into the host chromosome (step 5). This state is stable, with prophage DNA maintained in daughter cells over generation (step 6). The phage can exit lysogeny to reenter the lytic phase, at a later point in time.

Bacteriophage λ and the lambdoid family of phages to which it belongs, which include HK022, 434, P22 and HK97, broadly present one solution to the requirements of a temperate bacteriophage. Comparing how evolutionarily divergent bacteriophages tackle these problems could provide more general principles that could be applied to larger systems.

Temperate bacteriophage 186, belonging to the P2-like family of bacteriophages, also infects *E. coli* and is UV-inducible, like bacteriophage λ . Despite sharing no sequence similarity at the nucleotide level, similarities can be identified in the network signaling architecture (Trusina *et al.*, 2005). Studying bacteriophage 186 has revealed it as an excellent counterpoint model to bacteriophage λ .

1.1 Bacteriophage λ

1.1.1 Lytic gene cascade of bacteriophage λ

The genome of bacteriophage λ contains a key regulatory region involved in the lytic-lysogeny decision (Figure 1-2). After injection of the phage DNA into the cell, λ undergoes a temporally controlled cascade of gene expression, with the default cascade leading into lytic development.

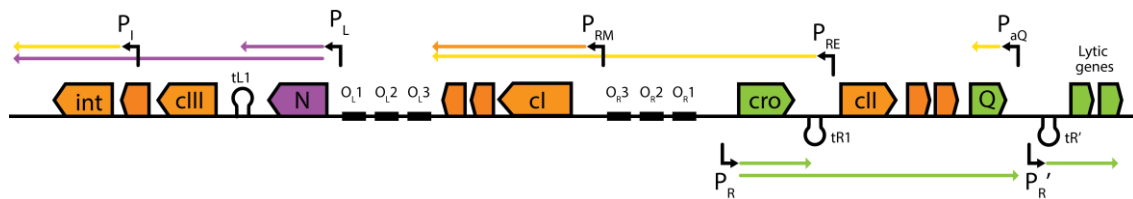


Figure 1-2 – Simplified genetic map of the regulatory region of bacteriophage λ . Promoter regions are denoted with bent arrows. Transcripts are shown as coloured arrows extending from the promoters. Purple arrows correspond to lytic transcripts. Yellow arrows correspond to early lysogenic transcripts. Orange arrows correspond to late lysogenic transcripts. Genes and open reading frames are shown by coloured boxes. Transcripts from promoters activated by CII are shown in yellow. Lytic and lysogenic genes are coloured green and orange, respectively.

Early transcription from the p_L and p_R promoters drives the expression of the *cro* and *N* regulators. *Cro* is a repressor that promotes the lytic cycle. It binds to O_L and O_R operators. At O_R , the affinity of *Cro* to the operator sites is in the order of $O_{R3} > O_{R2} > O_{R1}$. This preferential binding allows *Cro*, in lytic development and in prophage induction, to occupy O_{R3} to repress the expression of the *cl* immunity repressor (Johnson, Pabo and Sauer, 1980; Schubert *et al.*, 2007). These early transcripts from p_L and p_R terminate at the $tL1$ and $tR1$ terminators, respectively. The *N* antitermination factor and host Nus proteins complex with RNA polymerase to overcome these terminators to allow expression of delayed early genes (Franklin, 1974; Mishra *et al.*, 2013). The delayed early genes include lysogeny-promoting regulators *CII* and *CIII*, and lytic late gene regulator *Q*.

Here, the cascade can commit to the lytic cycle or divert to the lysogenic cycle. In the default lytic cascade, *Q* alters RNA polymerase to overcome the tR' terminator downstream of the constitutive late lytic promoter pR' (Roberts *et al.*, 1998). The result is the expression of late genes required for virion production and release. An alternative lysogenic cascade is described in the next section.

1.1.2 Establishment of lysogeny in bacteriophage λ by λ CII

In order for λ to form a lysogen, the bacteriophage λ *CI* repressor and the *Int* (integrase) protein need to be expressed. The *CI* repressor is required to repress the lytic promoters, pR and p_L , to establish and maintain the lysogenic transcriptional state. *Int* is needed for the integration of the phage genome into the host chromosome. The *CII* transcriptional activator and *CIII* protease inhibitor are required to generate the initial pool of *CI* and *Int*. λ plaques are normally turbid due to the presence of lysogenic cells. Bacteriophage λ mutants with

mutations in CI, CII or CIII give rise to clear plaques (Kaiser, 1957) and have lost the ability to enter the lysogenic state.

The crucial event in this switch to lysogeny is the accumulation of the CII transcriptional activator above a threshold level. CII actively inhibits Q activity to pre-empt expression of late lytic functions (Kobiler *et al.*, 2005). It also activates the expression of the Int protein and CI repressor. CII achieves these three objectives through activation of the paQ (Hoopes and McClure, 1985), pI (Schmeissner *et al.*, 1981) and pRE promoters (Schmeissner *et al.*, 1980), respectively. After this initial pool of CI exists, CI represses lytic functions of the phage (including cII) and autoregulates its own expression in the absence of CII.

The level of CII is central to the lytic-lysogenic decision and is regulated *in vivo* at many levels. It has been reviewed more extensively elsewhere (Oppenheim *et al.*, 2005; Casjens and Hendrix, 2015). Transcription of the cII gene is repressed by CI and Cro binding to the O_R operators. It is activated by the N antitermination factor binding to the nutR site to allow RNA polymerase to overcome the tR1 terminator. The stability of CII mRNA is regulated through the OOP RNA. OOP is an antisense RNA that can form a double stranded RNA complex with cII mRNA, which is recognized and cleaved by RNase III (Krinke, Mahoney and Wulff, 1991). The OOP promoter is repressed by the host LexA repressor. Translation initiation of CII requires integration host factor (IHF) (Mahajna *et al.*, 1986), which is also required for the phage genome integration reaction. Finally, CII is also regulated at the protein level. The transcriptional activator protein is rapidly degraded by the host HlfB (FtsH) protease (Kobiler *et al.*, 2002). CIII, which is co-expressed with CII, acts as an inhibitor of FtsH to increase CII levels (Kobiler, Rokney and Oppenheim, 2007).

The structure and function of λ CII

CII functions as a tetrameric activator, binding to TTGC direct repeats, present in all its cognate promoters. Its crystal structure as a tetramer and in a complex with DNA have been solved (Figure 1-3) (Datta *et al.*, 2005; Jain *et al.*, 2005). It is a single domain, helix-turn-helix (HTH) containing protein. It consists of four α-helices, with the helix-4 from four monomers interacting to form a four-helix bundle to mediate tetramerization. This helix can move relative to the remainder of the protein due to conformation flexibility afforded by the linker between helix 3 and 4. This flexibility allows it to adopt the asymmetric tetramer bundle quaternary structure required to bind to direct repeat half sites (Jain *et al.*, 2005). The protein ends with an unstructured and flexible tail, around residues 85-97, which acts as a recognition site for the protease FtsH (Kobiler *et al.*, 2002).

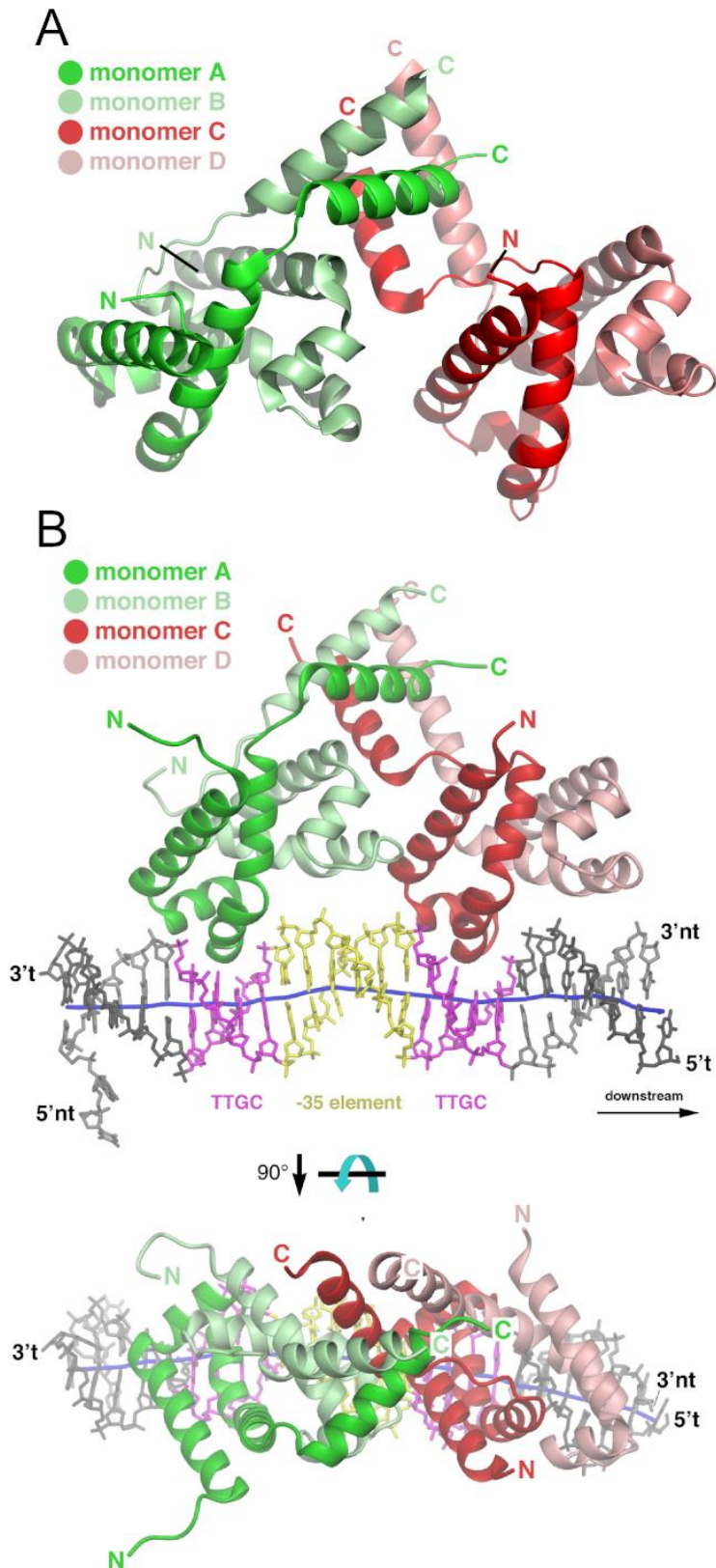


Figure 1-3 – Structure of λ CII with and without DNA. A) Structure of λ CII as a tetramer (PDB ID 1XWR). Each protein chain is displayed as a different colour. Figure reproduced from Datta *et al.* (2005) B) Structure of λ CII in complex with synthetic 27-mer DNA (PDB ID 1ZS4). -35 annotation based on p_{RE} promoter. CII half sites (TTGC) are shown in magenta. Figure reproduced from Jain *et al.* (2005).

λ CII is functionally unique in several aspects. Many activators bind as a dimer rather than a tetramer. It recognizes direct repeat half-sites instead of inverted repeats more frequently observed in prokaryotic activators. In addition, the CII half sites in p_I , p_{aQ} and p_{RE} straddle the -35 hexamer of the promoter, making CII a class II transcriptional activator. These

activators normally make direct contacts with domain 4 (σ_4) of σ subunit of RNA polymerase. However, λ cII binds on the opposite face of DNA relative to the σ subunit and contacts the CTD of the α subunit of RNA polymerase. α -CTD contacts DNA at -41 and in turn contacts the σ^{70} subunit (Figure 1-4) (Kedzierska *et al.*, 2004; Jain *et al.*, 2005).

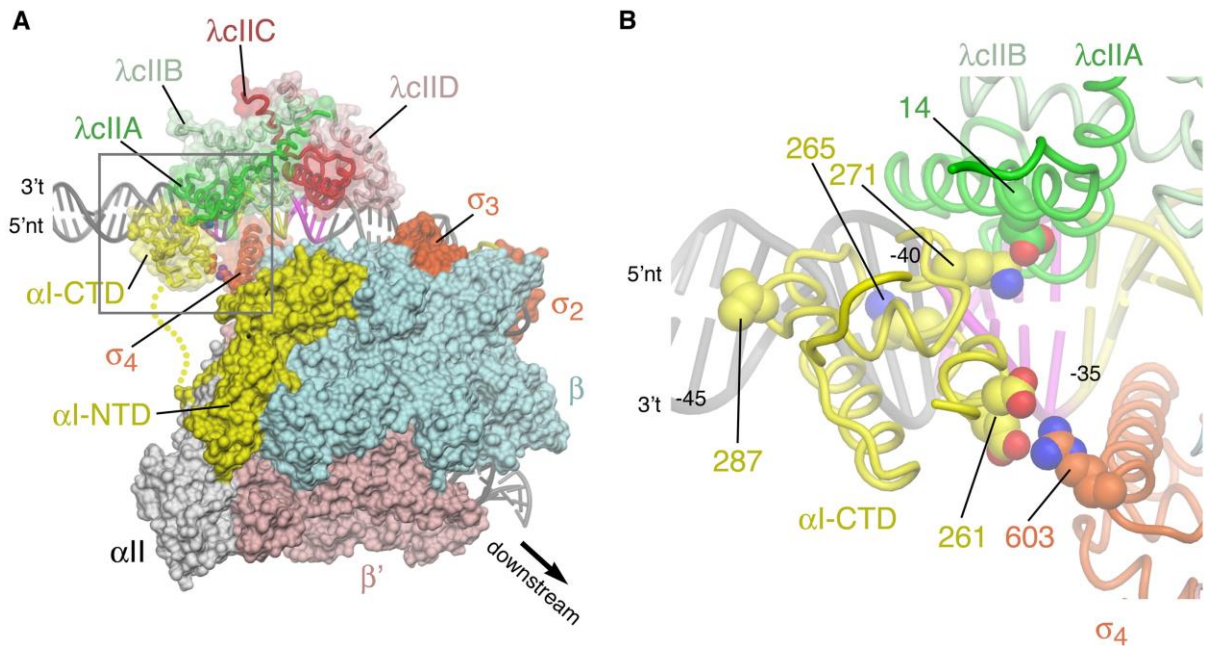


Figure 1-4 – Structural model for λ CII recruitment of RNA polymerase to its cognate promoters. Model was reproduced from Jain *et al.* (Jain *et al.*, 2005) A) Model of λ CII-RNA polymerase holoenzyme complex, constructed through superimposition the *Thermus aquaticus* RNA polymerase open promoter complex (Murakami *et al.*, 2002) and *E. coli* α -CTD (Benoff *et al.*, 2002) onto the structure of structure of the λ CII-DNA complex. B) Magnified region of (A) showing contacts between RNA polymerase and λ CII. λ CII E14 contacts RNA polymerase α -CTD (predicted to be α -CTD residue K271). α -CTD contacts DNA via its 265 determinant. α -CTD contacts σ_4 via its 261 determinant (D259 and E261).

1.1.3 Maintenance of lysogeny via λ CI

λ CI activates the expression of λ CI. If the threshold level of λ CI is reached, a sufficient initial pool of CI is established to prevent the lytic transcriptional cascade by repressing the p_R and p_L promoters. CI also positively and negatively autoregulates its own expression from the p_{RM} promoter. The p_R and p_{RM} promoters are arranged back to back, with the other lytic promoter p_L positioned 2.3 kb away (Figure 1-5A). The mechanism employed by λ CI to efficiently repress its cognate promoters is shown in Figure 1-5B.

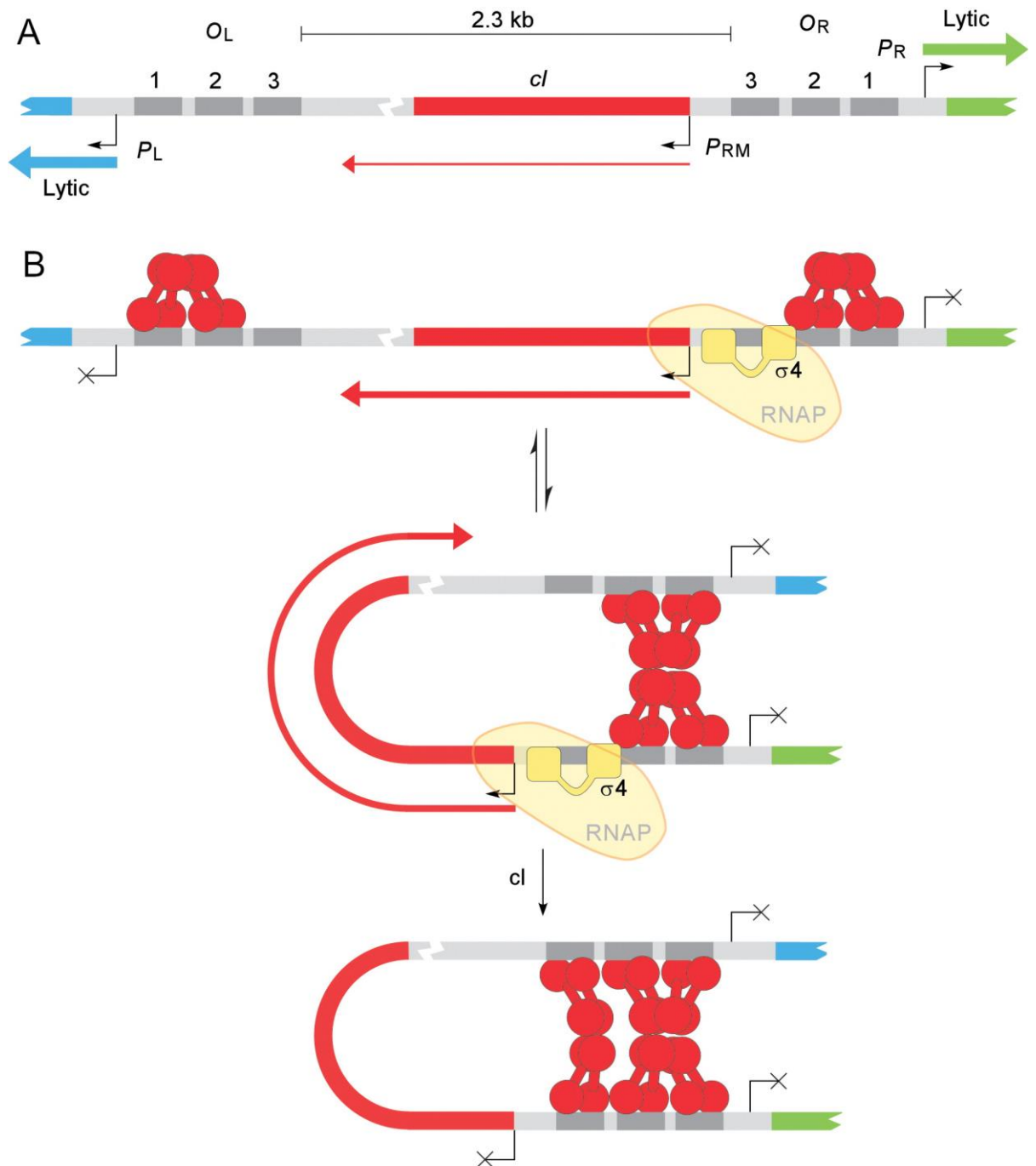


Figure 1-5 – Regulation of the p_R , p_L and p_{RM} promoters of bacteriophage λ CI is mediated through multiple cooperative interactions between λ CI dimers and long range DNA looping. A) Arrangement of p_R , p_L and p_{RM} promoters in bacteriophage λ . Promoters are depicted with bent arrows. RNA transcripts are depicted as coloured arrows. B) Regulation of the p_R , p_L and p_{RM} promoters by λ CI, shown as red dumbbells. Top - λ CI binds to the O_R and O_L operators to block transcription. Adjacent dimers can form tetramers through oligomerization mediated by the λ CI C-terminal domain. The N-terminal domain of λ CI at O_R2 contacts RNA polymerase (depicted in yellow) to activate p_{RM} to drive expression of lysogenic genes including λ CI. Middle – Long range DNA looping mediated by λ CI tetramers bound at O_R1 and O_R2 at O_R and O_L1 and O_L2 at O_L further improves the repression of p_R and p_L . Bottom – DNA looping aligns the O_R3 and O_L3 sites. At high levels of CI, λ CI bound at these two operator sites can interact cooperatively to form a tetramer to repress p_{RM} . Figure adapted from Dodd, Shearwin and Egan (2005).

The λ CI protein consists of two domains connected by a flexible linker. The N-terminal domain contains an HTH motif and is responsible for contacting DNA, while the C-terminal domain mediates oligomerization (Bell and Lewis, 2001; Nickels *et al.*, 2002). The relative affinities of λ CI to its operators follow the order of $O_{R1} > O_{R2} > O_{R3}$ and $O_{L1} > O_{L2} > O_{L3}$. λ CI

dimers preferentially bind to O_{L1} and O_{L2} operators at O_L and O_{R1} and O_{R2} operators at O_R over the O_{L3} and O_{R3} operators. These dimers cooperatively form tetramers at O_{R1} - O_{R2} and O_{L1} - O_{L2} . This is the mechanism of p_R and p_L repression by λ CI. Another level of cooperativity is introduced by interaction of the two tetramers to form an octamer, mediated by 2.3 kb DNA loop, to further improve the repression of these promoters (Ptashne, 2004). λ CI at O_{R2} activates λ p_{RM} . By occupying this operator, the N-terminal domain of λ CI is positioned to contact σ_4 subunit of RNA polymerase that binds the -35 promoter element (Nickels *et al.*, 2002; Jain *et al.*, 2004). This allows λ CI to positively regulate its own expression. The octameric λ CI aligns the O_{L3} and O_{R3} such that in the presence of high λ CI concentrations, λ CI can bind and form a tetramer across the O_{L3} and O_{R3} operators (Dodd *et al.*, 2004). This binding represses the p_{RM} promoter to reduce λ CI levels and is the mechanism of negative autoregulation in λ CI. The positive and negative autoregulation of λ CI creates a stable lysogenic state, but also prevents λ CI levels to escalate to the point where re-entering the lytic cycle (prophage induction) is difficult (Dodd, Shearwin and Egan, 2005).

Structure and function of λ CI

Several λ CI protein structures in complex with DNA have been solved (Figure 1-6) (Beamer and Pabo, 1992; Jain *et al.*, 2004; Stayrook *et al.*, 2008). In these structures, a protein dimer is complexed with O_{L1} DNA, with each monomer contacting an O_{L1} half site. The N-terminal domain responsible for contacting DNA is made up of five alpha helices. Helices 1-4 forms a compact globular domain from which helix 5 extends. Helix 5 forms interactions with helix 5 from another monomer to mediate weak dimerization (Pabo and Lewis, 1982; Beamer and Pabo, 1992). Helix 2 and 3 constitute a HTH motif. Helix 2 makes contacts with the DNA backbone and helix 3 (recognition helix) is positioned in the major groove, with side chains of the helix contacting the DNA base-pairs. The N-terminal arm (residues 1-6) also forms sequence specific contacts with the DNA. In the consensus O_{L1} half site, the N-terminus wraps around the DNA along the major groove (Beamer and Pabo, 1992).

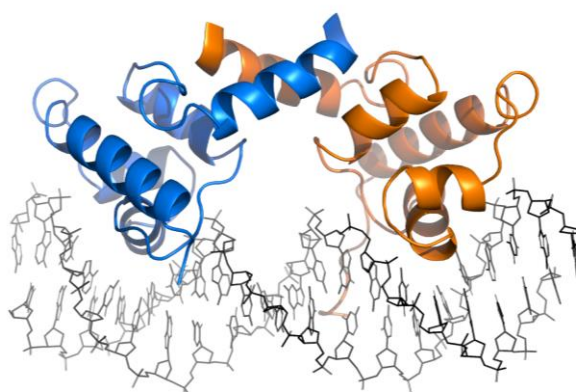


Figure 1-6 – Structure of λ CI NTD dimer in complex with O_{L1} DN (PDB ID 1LMB). The recognition helix 3 is positioned in the major groove of the half sites.

As described in section 1.1.3 and Figure 1-5, λ CI dimers oligomerise to form tetramers and octamers to achieve cooperative transcriptional regulation. The C-terminal domain mediates strong dimerization and subsequent higher order oligomerization. Several crystal structures containing λ CI CTD as an oligomer have been solved. These structures include the structure of full length λ CI with O_{L1} DNA, λ CI-CTD as a tetramer and λ CI-CTD as an octamer (Figure 1-7). By superimposing the structure of the λ CI dimer bound to DNA onto the λ CI CTD octamer, Stayrook *et al.* (2008) presented a structural model for the λ CI octamer bound to

adjacent operators. Cui *et al.* (2013) demonstrated that the Stayrook model could not physically accommodate the DNA linking the operators. Cui *et al.* (2013) refined the model by modelling conformational flexibility in the linker between the NTD and CTD to generate a plausible structural model for a DNA loop mediated by an λ CI octamer in complex with RNA polymerase at P_{RM} (Figure 1-8B). This provides structural basis to understand cooperative regulation by bacteriophage λ CI.

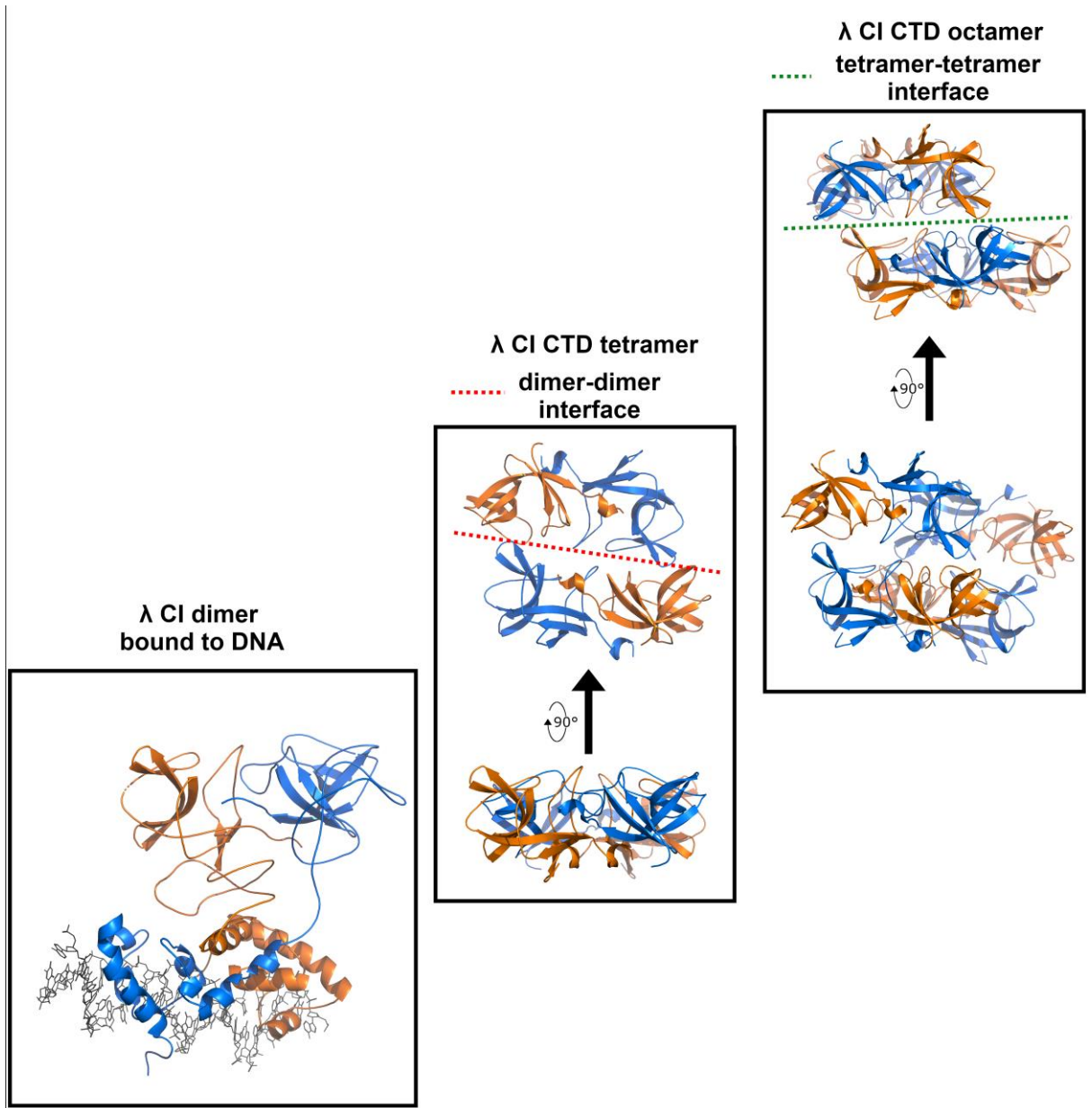


Figure 1-7 – Structure of full length λ CI with O_L1 DNA (PDB ID 3BDN), λ CI-CTD as a tetramer (PDB ID 1F39) and λ CI-CTD as an octamer (PDB ID 1KCA). Dimer-dimer interface and tetramer-tetramer interface have been indicated using a dashed red and dashed green line, respectively.

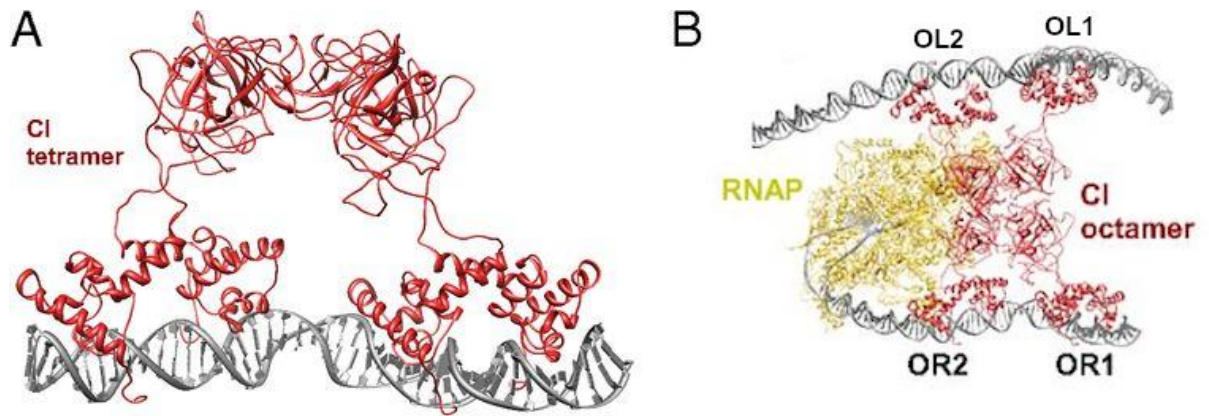


Figure 1-8 – Structural modelling was used to investigate how λ CI could bind DNA and RNA polymerase *in vivo*. A) Model of CI tetramer binding OR1-OR2 DNA. B) Structural model of the λ CI octamer in complex with OL1-OL2 and OR1-OR2 DNA and RNA polymerase. Figures were adapted from Cui et al. (2013).

1.1.4 Summary

The decision making process between the lytic and lysogenic lifecycles in temperate bacteriophage λ , a simple genetic switch relative to complex eukaryotic systems, involves a small set of regulatory proteins with temporally regulated expression. Two key transcriptional regulators λ CI and λ CII have been well structurally characterised. Combining this information with other genetic studies has enriched our understanding of mechanisms employed by transcriptional regulators to activate and/or repress transcription and autoregulate their expression, which involve macromolecular interactions between these proteins, DNA and host transcriptional machinery.

Bacteriophage λ has enriched our understanding of fundamental principles in gene regulation. The study of an evolutionary divergent temperate bacteriophage may provide novel regulatory mechanisms to the lytic-lysogenic process, to further our knowledge on this subject.

1.2 Bacteriophage 186

Bacteriophage 186, originally isolated from the sewers of Paris (Jacob and Wollman, 1961), has been used as a counterpoint model of a genetic switch to bacteriophage λ . 186 and λ are SOS-inducible temperate bacteriophages that infect *Escherichia coli* and can both make a developmental decision between the lytic and lysogenic pathways (Woods and Egan, 1974). These two phages are unrelated at the DNA level (Portelli *et al.*, 1998). Therefore, comparing them should reveal important features of a genetic switch, as they represent independently evolved solutions to the lytic-lysogenic problem (Shearwin, Dodd and Egan, 2002).

1.2.1 Lytic gene cascade in bacteriophage 186

Similarly to bacteriophage λ , bacteriophage 186 displays a temporally controlled pattern of gene expression in lytic development. This can be divided up into three transcriptional phases – early, middle and late (Finnegan and Egan, 1981). A schematic diagram of the 186 genome is shown in Figure 1-9.

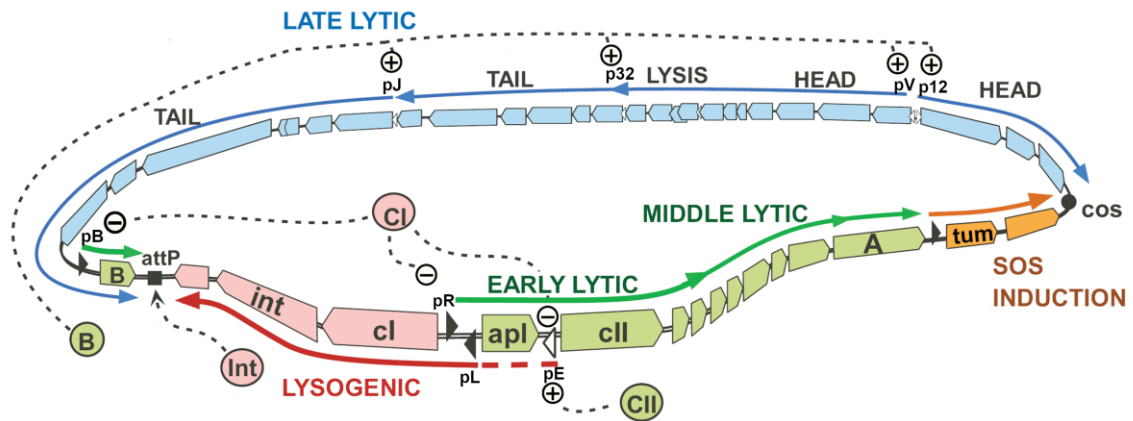


Figure 1-9 – Circular map of the bacteriophage 186 genome. Genes are shown as boxes. Promoters are shown as arrows pointing in the direction of transcription. Proteins are shown as coloured circles. Early and middle lytic functions are driven by the p_B and p_R promoters. Late lytic functions are driven by the p_J , p_V and p_{12} promoters. Lysogenic functions are expressed from the p_L promoter. The CI transcriptional activator represses the lytic promoters p_B , p_R and early lysogenic promoter p_E . CII protein activates the early lysogenic promoter p_E . B activator activates the late lytic promoters p_J , p_{32} , p_V and p_{12} . Figure adapted from Dodd (unpublished).

After infection, the early lytic promoter p_R is active (Finnegan and Egan, 1981). As a result of the convergent arrangement of p_R and p_L , transcription from the stronger p_R promoter interferes transcriptional activity from the weaker p_L promoter (Callen, Shearwin and Egan, 2004). The early lytic transcript from p_R terminates at the t_R cluster of terminators and includes the *apl*, *cll*, *fil* and *dhr* genes. Apl is a repressor of p_R and p_L and also behaves as an excisionase during prophage induction (Dodd, Reed and Egan, 1993; Reed *et al.*, 1997). The Fil and Dhr proteins are inhibitors of host cell division and host DNA replication, respectively (Richardson and Egan, 1989). The 186 CII protein, similarly to the λ CII protein, is a transcriptional activator required to divert 186 off the lytic pathway towards lysogeny. This protein will be covered later in more detail. The replicase protein A is expressed from the middle lytic region as part of the p_R transcript that has extended beyond t_R due to an antitermination event (Richardson, Puspurs and Egan, 1989). The A protein initiates phage genome replication (Sivaprasad *et al.*, 1990).

Also during the early lytic phase, the B activator is expressed from the p_B promoter. This transcription activator activates the p_V , p_{32} , p_J and p_{12} promoters to express late lytic genes required for cell lysis, DNA packaging and bacteriophage capsid proteins (Dibbens and Egan, 1992; Portelli *et al.*, 1998).

1.2.2 Diverting towards the lysogeny cycle in bacteriophage 186

Diversion from the lytic cascade and establishment of lysogeny requires repression of the lytic promoters, p_B and p_R , and expression of the immunity repressor CI and Int from the lysogenic transcript (Figure 1-10). One of the proteins expressed under p_R is CII, which activates the alternative lysogenic operon promoter, p_E (Figure 1-11). Analogous to bacteriophage λ , the initial decision to switch into lysogeny is dependent on the concentration of CII. The default lytic cascade occurs if CII doesn't accumulate to sufficient levels. However, high CII levels lead to high CI levels, which repress lytic promoters p_R and p_B , as well as expression of the Int integrase protein, which catalyses the integration of phage

DNA into the bacterial chromosome. CI repression of p_R activates p_L by removal of transcription interference to sustain its own expression to maintain the lysogenic state.

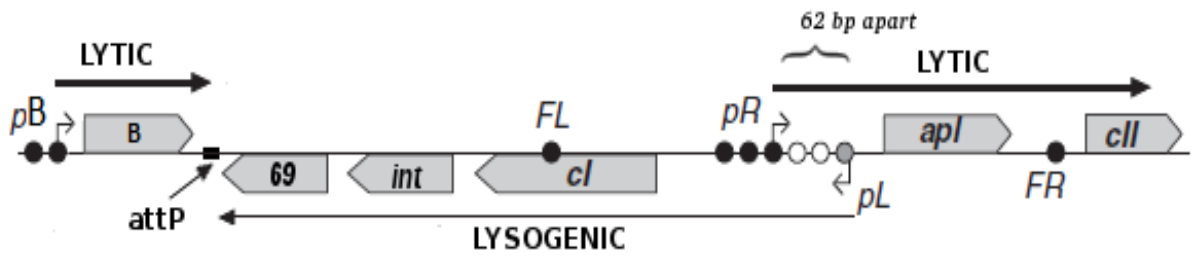


Figure 1-10 – A schematic of the “switch” region in the bacteriophage 186 genome involved in deciding between the lytic or lysogenic cycle. Repressing the lytic promoters, p_B and p_R , and expression from the lysogenic promoter p_L is required for the maintenance of lysogeny. 186 immunity repressor CI sites are shown in circles (unfilled circles are poorly defined weak sites). Promoter regions are labelled with arrows in the direction in which transcription occurs. Adapted from Wang et al (2013).

The action of CII in 186 is thus similar to lambda. However, the co-location of *int* with *cl* in the lysogenic operon allows CII to express both these genes from a single promoter, in contrast to the two promoters used in lambda. An equivalent to the lambda p_aQ promoter has not been found in 186.

1.2.3 The 186 CII protein

CII is a 169 residue protein initially expressed under the control of p_R . CII binds to the 186 p_E promoter to activate it and is capable of upregulating promoter activity about 400 fold (Neufing *et al.*, 1996) (Figure 1-11). This activation only requires CII and RNA polymerase and allows for the rapid production of immunity repressor CI in sufficient levels to block lytic development (Shearwin and Egan, 2000). This process is essential to the establishment but not maintenance of lysogeny, as rare lysogens formed by *cII* mutants are stable.

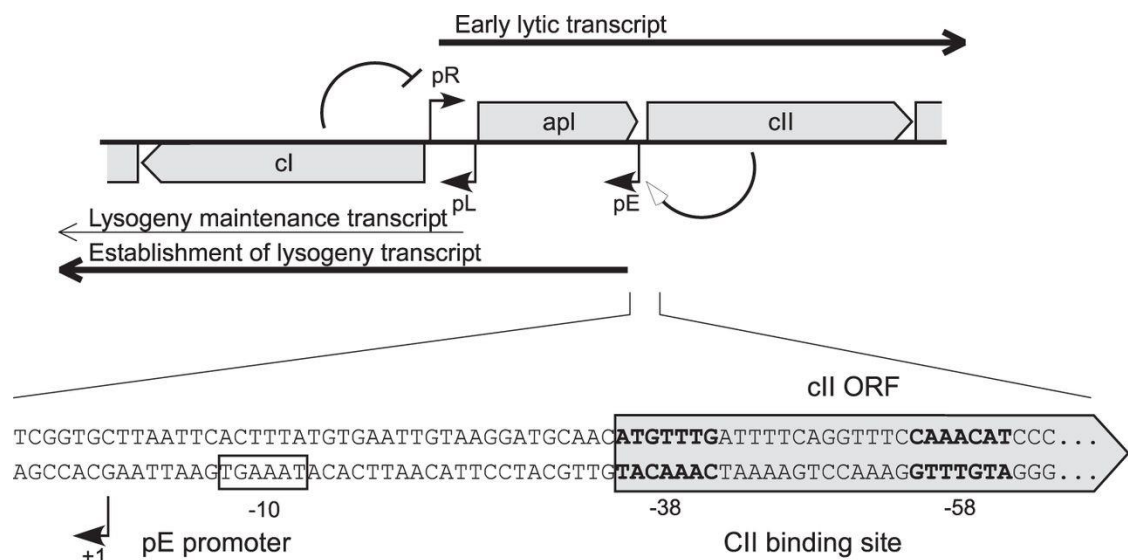


Figure 1-11 – CII’s role in promoting the establishment of lysogeny in bacteriophage 186. After initial infection, an early lytic transcript from the p_R promoter is made. One product made is CII, which activates an early lysogenic promoter p_E . Transcription from this promoter generates the initial pool of CI required to enter lysogeny. CI represses the lytic promoter p_R to maintain lysogeny. CI levels are maintained by the p_L promoter. CII half sites (ATGTTTG) are centered around -38 and -58 and are emphasized in bold. Figure reproduced from Murchland et al. (2014).

186 CII binds to two inverted repeat 7-mer half sites (ATGTTTG) centering at -38 and -58 of the p_E promoter (Neufing *et al.*, 1996) (Figure 1-11). The -38 half site lies just upstream of where the σ^{70} subunit of RNA polymerase would be expected to contact DNA. Thus, 186 CII,

like λ CII, is a class II activator. However, unlike λ CII which binds direct repeats, 186 CII binds to inverted repeat half sites spaced two turns of the DNA helix apart. The spacing is important for CII binding as changes to the spacing, even an insertion or removal of a single base pair, significantly reduced affinity of the sites for CII binding (Shearwin and Egan, 2000). This wider spacing between the CII half sites, compared to the single DNA helix spacing normally observed, is unique among prokaryotic transcriptional activators, which suggests that CII employs a novel mechanism to activate p_E .

Similarly to λ CII, 186 CII is also rapidly degraded *in vivo* due to a degradation tag at the C-terminus, which recruits proteases to cleave the protein, initially at residue 135. Truncating the protein at residue 145 (to give CII145) eliminates this degradation tag, prolongs the *in vivo* half-life around 14-fold and increases the steady state levels of the protein inside the cell. As a result, CII145 shows significantly increased activation of p_E (Murchland *et al.*, 2014).

There has yet to be a high resolution structure of 186 CII solved, but other biophysical and genetic experiments have provided insight into its structure and function. CII is predicted to be a two domain protein, consisting of an N-terminal DNA binding domain and a C-terminal oligomerization domain (Murchland *et al.*, 2014). The DNA binding domain is predicted to contain a helix-turn-helix motif which is used to contact DNA (Neufing *et al.*, 1996). The full length protein can be expressed in *E. coli*, yet the predicted N-terminal domain and C-terminal domains could not be purified separately (Murchland *et al.*, 2014). Thus, it is unknown if the domains are truly structurally independent.

CII can self-associate into dimers and tetramers. Sedimentation equilibrium experiments of purified CII suggested it exists in a monomer-dimer-tetramer equilibrium in solution (Neufing *et al.*, 1996). The tetrameric species of CII could also be detected via Size Exclusion Chromatography - Multi-Angle Laser Light Scattering (SEC-MALS) (Murchland *et al.*, 2014). Sedimentation equilibrium experiments of purified CII with DNA suggested CII binds DNA as a preformed dimer (one monomer binding a half site), with weak binding as a tetramer at higher concentrations. Upon DNA binding, CII generates a $41^\circ \pm 5$ bend in the DNA (Shearwin and Egan, 2000). The degradation product of CII, CII135, self-associates into dimers but not tetramers. Although CII is thought to bind and activate as a dimer, the CII135 protein was unable to activate p_E *in vivo*. Either CII135 has lost a critical activation epitope or CII has two independent dimerization interfaces that yield two distinct dimers, only one of which is an active dimer (Murchland *et al.*, 2014).

A genetic screen to identify CII mutants that are activation deficient but still capable of binding DNA identified CII residues that contact RNA polymerase. Three residue positions R17, E46 and R115 were found to be important. R17 and E46 are thought to cluster on the DNA binding domain and are part of an epitope on the N-terminal domain that contacts σ^{70} K593, which was validated by charge reversal mutagenesis experiments with K593 and E46. R115 is hypothesized to contact the α -CTD of RNA polymerase. The model of p_E promoter activation by CII proposed by Murchland *et al.* is summarized in Figure 1-12. This model proposes that the unusual arrangement of CII half sites two DNA turns apart enables one or both α -CTD to bind DNA between the half sites, stabilized by contacts with CII.

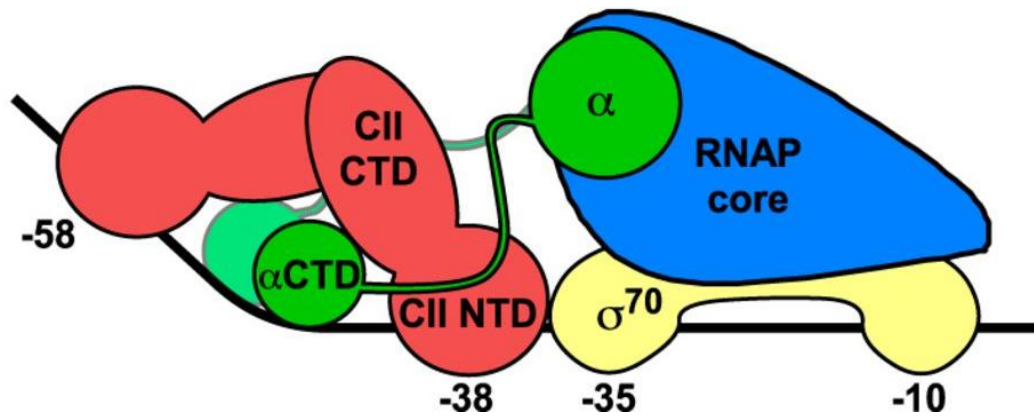


Figure 1-12 – Current model of 186 CII-pE promoter activation. CII binds to the pE promoter at -38 and -58 and bends the DNA. The CII NTD makes contacts with RNA polymerase σ^{70} subunit, which can bind at -35. The CII CTD can contact α -CTD of RNA polymerase, which can bind to the bent AT rich DNA between the half sites. These contacts are used to recruit RNA polymerase to the pE promoter. Figure adapted from Murchland et al. (2014)

A high-resolution structure of CII would be immensely valuable in understanding CII function and would test many of the assumptions in the current model. Analysis of this structure should inform about CII's contacts with RNA polymerase and DNA (via cocrystallization, cryoEM or docking). It should test whether R17 and E46 are clustered together into an epitope and whether CII contains an HTH motif. In a protein crystal, protein molecules are arranged in a lattice, which are revealed in a crystal structure. The arrangement of protein molecules within a crystal can often reflect their biological oligomeric arrangement. A crystal structure of CII could provide insight into functional oligomeric requirements of CII (dimer or tetramer). Crystallography is unable to definitively determine the oligomeric state, so complimentary techniques such as mass spectrometry would be required to validate predictions from the crystal structure. This information would be key to testing the proposed novel arrangement of CII and RNA polymerase. Finally, previous CII mutational data could be explained by mapping onto a structure. This data includes promoter/binding site mutants (Shearwin and Egan, 2000). RNAP contact mutants (Murchland *et al.*, 2014) and protease resistant mutants (Keith Shearwin, unpublished data).

1.2.4 Maintenance of lysogeny with 186 CI

CII establishes the initial pool of the immunity repressor CI, which is responsible for the maintenance of the lysogenic state (Dodd and Egan, 1996). The immunity repressor rapidly shuts off lytic functions by repressing the early lytic promoters, p_R and p_B in the switch region (Figure 1-10). The lytic promoter p_R and lysogenic promoter p_L are arranged face to face, with 62 base pairs between their transcriptional start points. Transcription from the stronger p_R promoter (approximately 10 fold stronger than p_L) inhibits transcription from the weaker p_L promoter by 10 to 20-fold (Dodd, Kalionis and Egan, 1990). Thus, as CI represses p_R , it indirectly stimulates its own expression by activating the p_L promoter through the relief of transcriptional interference. This mechanism of positive autoregulation by 186 CI (Dodd and Egan, 2002) is distinct from the λ CI mechanism of using direct λ CI-RNA polymerase contacts to stimulation of its own expression (Kuldell and Hochschild, 1994). Similarly to λ CI, 186 CI also displays negative autoregulation at high concentrations (Dodd, Kalionis and Egan, 1990)

Structure and function of 186 CI

Pinkett *et al.* (2006) solved the crystal structures of two different variants of the 186 CI protein. One structure was of a cooperativity mutant of the full length repressor CI^{E146K} that could dimerize but could not form higher order oligomers (Figure 1-13a). The repressor has a dual domain topology consisting of an N-terminal helix-turn-helix DNA binding domain with a C-terminal oligomerization domain. The NTD contains a single HTH motif which is responsible for DNA binding (Figure 1-13b) (Shearwin, Dodd and Egan, 2002; Pinkett *et al.*, 2006).

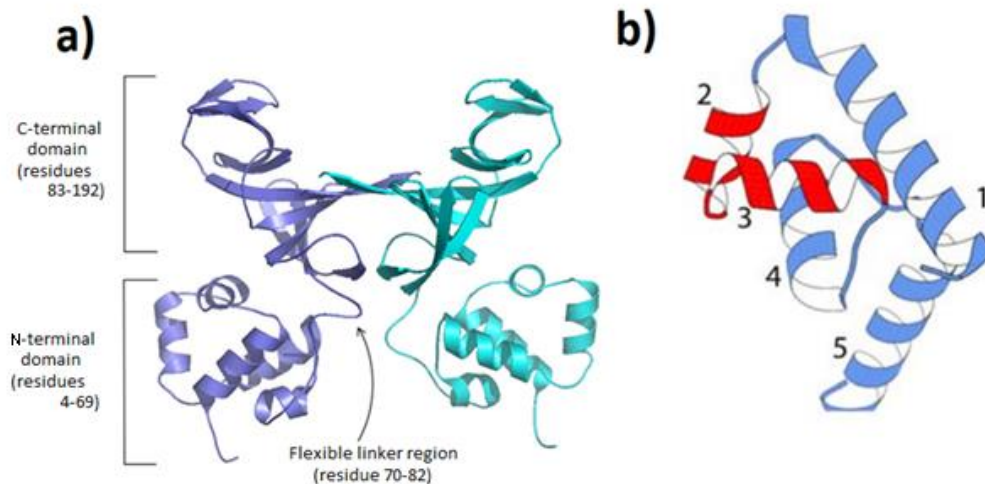


Figure 1-13 – Structure of 186 CI^{E146K}. A) Crystal structure of CI^{E146K} forming a dimer. The E146K substitution prevents formation of higher order structures. B) The CI NTD contains a helix-turn helix motif consisting of helices 2 and 3, shown in red. Adapted from Pinkett *et al.* (2006).

1.2.5 DNA recognition by 186 CI

There are five regions in the 186 genome where CI can bind – p_B , p_R , p_L , F_R and F_L (Figure 1-10). Dodd and Egan (1996) categorized CI operators into three types — A, A' and B sites. A and A' sites consist of inverted repeats with the same consensus sequence separated by 5 or 4 bp AT rich spaces, respectively. B sites also have inverted repeats but with half sites with a different consensus to the A sites (Figure 1-14). Interestingly, one HTH motif is responsible for binding both recognition sequences with variable spacing between the operators (Shearwin, Dodd and Egan, 2002). It is speculated that flexibility of the protein may allow this variable binding. In the crystal structure of the dimer, the HTH motifs are positioned such that they would not be able to contact adjacent half sites. Thus, the linker between the NTD and CTD is flexible, as previously suggested by Shearwin and Egan (Shearwin, Dodd and Egan, 2002), which may allow different DNA complexes with variable half site spacing to form. In addition, the HTH motif itself may be flexible due to a lack of structural restraint from helix 5, allowing the adoption of different conformations to bind distinct sequences. It may be a combination of these flexibilities that allows recognition of two different sequences. Dodd and Egan (1996) observed in DNaseI footprinting that CI makes extended contacts with DNA adjacent to the primary CI operators, indicating that CI binds to DNA as a multimer. Thus, characterization of the flexibility and high order oligomerization of CI is required to fully understand its DNA binding behavior.

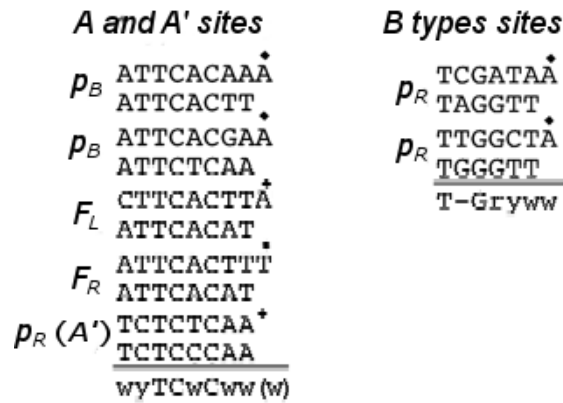


Figure 1-14 - The sequences of CI binding operators of the p_B , p_R , F_L and F_R sites. F_L and F_R sites each have one A site. p_B contains two A sites. p_R has an A' site and two B type sites, arranged in a B-A'-B fashion. For the consensus sequences mentioned above, R = A or G, y = C or T and w = T or A. The w of the A type consensus sequence in parentheses reflects the different spacing between the inverted repeats in A and A' type sites. Adapted from Shearwin et al. (Shearwin, Dodd and Egan, 2002).

1.2.6 Binding to multiple operators

The crystal structure of the oligomerization domain was also solved (Pinkett *et al.*, 2006) and revealed a heptamer of dimers in a ring arrangement (Figure 1-15B), with the arrangement consistent with protein mutations that disrupt CI cooperativity *in vivo*. Combining this CTD arrangement with the structure of CI as a dimer suggests CI forms 14mer wheel-like structure, with DNA binding domains situated on the outside of the wheel (Figure 1-15B) (Pinkett *et al.*, 2006).

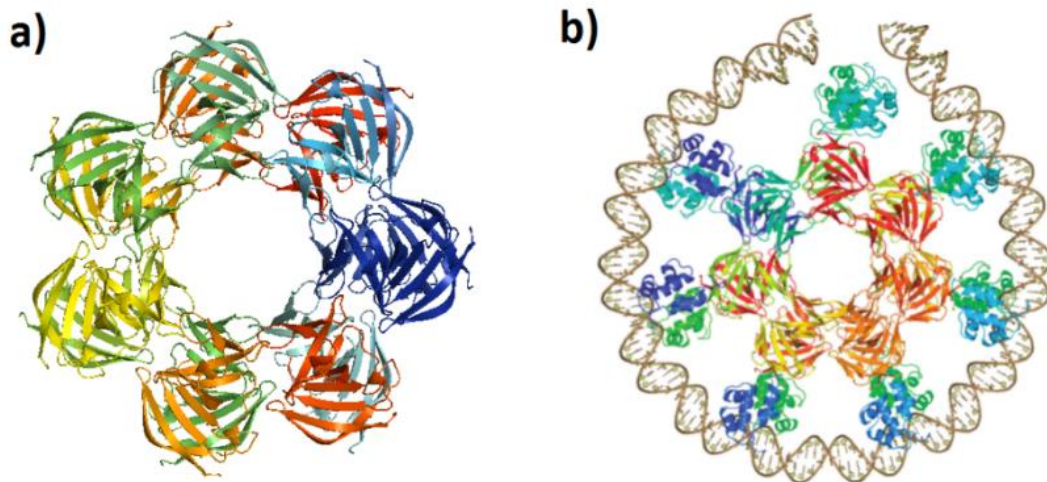


Figure 1-15- 186 CI is proposed to form a 14mer wheel. A) X-ray crystal structure of CI CTD forming a 14mer wheel. B) A model of DNA wrapping around a CI wheel, based on the crystal structures of the dimer and C-terminal domain 14mer. Reproduced from Wang et al. (2013)

The current model of 186 CI transcriptional regulation represents a very different mode of controlling lytic and lysogenic promoters than that seen in lambda or other phages. The model is based on the CI 14mer wheel acting as a scaffold for DNA to wrap on and off. The availability of DNA binding domains all around the wheel allows the formation of different DNA looped species to regulate the lytic and lysogenic promoters (Figure 1-16). The operators around p_R and p_L DNA can wrap onto the CI wheel to give species A, to repress both the p_R and p_L promoters. The operators that flank p_R and p_L , F_L and F_R , are proposed to displace the more weakly bound p_L , to create species C. In this arrangement, p_L is derepressed whilst maintaining repression of p_R , rationalizing the ability of these flanking

sites to stimulate pL activity *in vivo* even in the absence of active pR (Dodd and Egan, 2002). At high levels of CI, the F_L and F_R flanking sites are sequestered by further CI wheels (e.g. species D), to eliminate derepression of p_L via displacement by F_L and F_R. This negative autoregulation places a limit on CI levels *in vivo*, preventing the excessive accumulation of CI which would hinder prophage induction. This model was supported by the observation of CI discs of the correct dimensions by atomic force microscopy (AFM) interacting with DNA to give complexes that resemble the proposed species (Wang *et al.*, 2013). In addition, SEC-MALS measurements of the CI protein estimates the molecular weight of the oligomeric complex at 280 ± 8 kDa (12.1 ± 0.4 monomers). This observation supports 186 CI adopting a high order oligomeric state, although the estimated complex is slightly smaller than the expected 14mer.

Although the wheel structure is consistent with current observations, there is currently no high-resolution confirmation of the proposed wheel structure with or without DNA. Determining the quaternary structure from a crystal structure (such as in this case with the CI-CTD crystal structure) can sometimes result in erroneous conclusions, as the crystal lattice can favour the formation of non-physiologically relevant oligomers (Bahadur *et al.*, 2004). An experimental crystallographic or SAXS structure of the whole complex would validate this model of an alternative genetic switch, making it a useful counterpoint model to λ CI regulation. In addition, a high resolution structure of the full complex would extend the model by showing the arrangements of DNA binding domains. This information would provide insight into the 186 CI wheel-DNA interactions such as the ideal length of DNA and phasing of recognition sites in order for DNA to effectively wrap around the wheel.

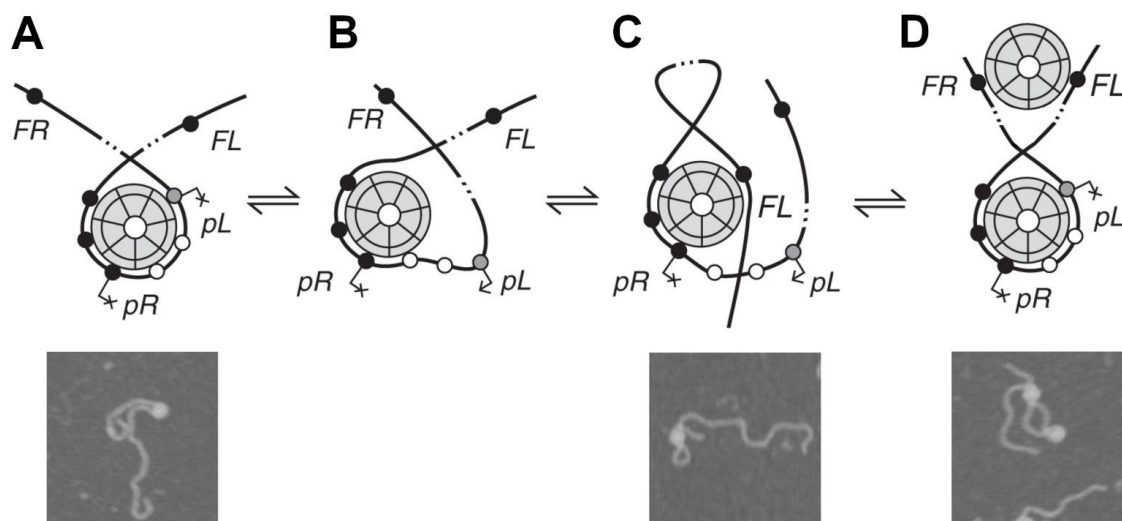


Figure 1-16 — Proposed structures formed when 186 switch region binds onto the 186 CI wheel. A corresponding CI-DNA complex observed with AFM with consistent dimension is shown below the proposed species. The HTH domains of three adjacent CI dimers can bind three major grooves of adjacent p_R operators. Weaker binding sites of p_L can then wrap onto the wheel, to give species a, where p_R and p_L are repressed. This binding behavior is consistent with CI repressing p_L and p_R in the absence of F_L and F_R sites (Dodd and Egan 1996). The weakly binding p_L sites may unbind (species b) and F_L or F_R sites may bind to the wheel to give species c. p_L is active in this state. This agrees with previous data demonstrating the presence of F_L and F_R sites relieves CI repression of p_L whilst p_R is still repressed (Dodd and Egan 1996). High concentrations of CI even in the presence of F_L/F_R still repress p_L and p_R and can be explained by the formation of species d. Figure adapted from Wang *et al.* (2013).

1.2.7 Summary

Bacteriophage 186, although being evolutionarily divergent to bacteriophage λ , functions very similarly. Both are temperate, UV-inducible, phages that infect *E. coli*. Both phages use a transcriptional activator to establish the initial pool of immunity repressor. The immunity repressor represses the lytic functions of the phage and activates the lysogenic promoter. However, the details of the mechanism of regulation by these proteins differ significantly between the two phages. More structural data on the 186 CII and CI proteins would further highlight these differences, and would position bacteriophage 186 as a more valuable counterpoint model to λ .

1.3 References

- Bahadur, R. P. *et al.* (2004) 'A Dissection of Specific and Non-specific Protein-Protein Interfaces', *Journal of Molecular Biology*, 336(4), pp. 943–955. doi: 10.1016/j.jmb.2003.12.073.
- Beamer, L. J. and Pabo, C. O. (1992) 'Refined 1.8 Å crystal structure of the λ repressor-operator complex', *Journal of Molecular Biology*, 227(1), pp. 177–196. doi: 10.1016/0022-2836(92)90690-L.
- Bell, C. E. and Lewis, M. (2001) 'Crystal structure of the λ repressor C-terminal domain octamer', *Journal of Molecular Biology*, 314(5), pp. 1127–1136. doi: 10.1006/jmbi.2000.5196.
- Benoff, B. *et al.* (2002) 'Structural basis of transcription activation: The CAP- α CTD-DNA complex', *Science*, 297(5586), pp. 1562–1566. doi: 10.1126/science.1076376.
- Callen, B. P., Shearwin, K. E. and Egan, J. B. (2004) 'Transcriptional interference between convergent promoters caused by elongation over the promoter', *Molecular Cell*, 14(5), pp. 647–656.
- Casjens, S. R. and Hendrix, R. W. (2015) 'Bacteriophage lambda: Early pioneer and still relevant', *Virology*, 479–480, pp. 310–330. doi: 10.1016/j.virol.2015.02.010.
- Cui, L. *et al.* (2013) 'Enhancer-like long-range transcriptional activation by λ CI-mediated DNA looping', *Proceedings of the National Academy of Sciences of the United States of America*, 110(8), pp. 2922–2927. doi: 10.1073/pnas.1221322110.
- Datta, A. B. *et al.* (2005) 'Structure of λ CII: Implications for recognition of direct-repeat DNA by an unusual tetrameric organization', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0504535102.
- Dibbens, J. A. and Egan, J. B. (1992) 'Control of gene expression in the temperate coliphage 186. IX. B is the sole phage function needed to activate transcription of the phage late genes', *Molecular Microbiology*, 6(18), pp. 2629–2642. doi: 10.1111/j.1365-2958.1992.tb01440.x.
- Dodd, I. B. *et al.* (2004) 'Cooperativity in long-range gene regulation by the λ CI repressor', *Genes and Development*, 18(3), pp. 344–354. doi: 10.1101/gad.1167904.
- Dodd, I. B. and Egan, J. B. (1996) 'DNA Binding by the Coliphage 186 Repressor Protein CI', *Journal of Biological Chemistry*, 271(19), pp. 11532–11540. doi: 10.1074/jbc.271.19.11532.
- Dodd, I. B. and Egan, J. B. (2002) 'Action at a distance in CI repressor regulation of the bacteriophage 186 genetic switch', *Molecular Microbiology*, 45(3), pp. 697–710.

- Dodd, I. B., Kalionis, B. and Egan, J. B. (1990) 'Control of gene expression in the temperate coliphage 186: VIII. Control of lysis and lysogeny by a transcriptional switch involving face-to-face promoters', *Journal of Molecular Biology*, 214(1), pp. 27–37.
- Dodd, I. B., Reed, M. R. and Egan, J. B. (1993) 'The Cro-like Apl repressor of coliphage 186 is required for prophage excision and binds near the phage attachment site', *Molecular Microbiology*, 10(5), pp. 1139–1150. doi: 10.1111/j.1365-2958.1993.tb00983.x.
- Dodd, I. B., Shearwin, K. E. and Egan, J. B. (2005) 'Revisited gene regulation in bacteriophage λ ', *Current Opinion in Genetics and Development*, 15(2), pp. 145–152. doi: 10.1016/j.gde.2005.02.001.
- Finnegan, J. and Egan, J. B. (1981) 'In vivo transcription studies of coliphage 186', *Journal of Virology*, 38(3), pp. 987–995. doi: 10.1128/jvi.38.3.987-995.1981.
- Franklin, N. C. (1974) 'Altered reading of genetic signals fused to the N operon of bacteriophage λ : Genetic evidence for modification of polymerase by the protein product of the N gene', *Journal of Molecular Biology*, 89(1), pp. 33–48. doi: 10.1016/0022-2836(74)90161-2.
- Hoopes, B. C. and McClure, W. R. (1985) 'A cII-dependent promoter is located within the Q gene of bacteriophage λ ', *Proceedings of the National Academy of Sciences of the United States of America*, 82(10), pp. 3134–3138. doi: 10.1073/pnas.82.10.3134.
- Jacob, F. and Wollman, E. L. (1961) *Sexuality and the Genetics of Bacteria*. Academic Press, London, UK.
- Jain, D. *et al.* (2004) 'Structure of a Ternary Transcription Activation Complex', *Molecular Cell*, 13(1), pp. 45–53. doi: 10.1016/S1097-2765(03)00483-0.
- Jain, D. *et al.* (2005) 'Crystal structure of bacteriophage λ cII and its DNA complex', *Molecular Cell*, 19(2), pp. 259–269. doi: 10.1016/j.molcel.2005.06.006.
- Johnson, A. D., Pabo, C. O. and Sauer, R. T. (1980) '[76] Bacteriophage λ Repressor and cro Protein: Interactions with Operator DNA', *Methods in Enzymology*, 65(C), pp. 839–856. doi: 10.1016/S0076-6879(80)65078-2.
- Kaiser, A. D. (1957) 'Mutations in a temperate bacteriophage affecting its ability to lysogenize Escherichia coli', *Virology*. doi: 10.1016/0042-6822(57)90022-3.
- Kedzierska, B. *et al.* (2004) 'Role of the RNA polymerase α subunits in CII-dependent activation of the bacteriophage λ pE promoter: Identification of important residues and positioning of the α C-terminal domains', *Nucleic Acids Research*, 32(2), pp. 834–841. doi: 10.1093/nar/gkh230.
- Kobiler, O. *et al.* (2002) 'The phage λ CII transcriptional activator carries a C-terminal domain signaling for rapid proteolysis', *Proceedings of the National Academy of Sciences of the United States of America*, 99(23), pp. 14964–14969. doi: 10.1073/pnas.222172499.
- Kobiler, O. *et al.* (2005) 'Quantitative kinetic analysis of the bacteriophage λ genetic network', *Proceedings of the National Academy of Sciences of the United States of America*, 102(12), pp. 4470–4475. doi: 10.1073/pnas.0500670102.
- Kobiler, O., Rokney, A. and Oppenheim, A. B. (2007) 'Phage lambda CIII: A protease inhibitor regulating the lysis-lysogeny decision', *PLoS ONE*, 2(4). doi: 10.1371/journal.pone.0000363.
- Krinke, L., Mahoney, M. and Wulff, D. L. (1991) 'The role of the OOP antisense RNA in

- coliphage λ development', *Molecular Microbiology*, 5(5), pp. 1265–1272. doi: 10.1111/j.1365-2958.1991.tb01900.x.
- Kuldell, N. and Hochschild, A. (1994) 'Amino acid substitutions in the -35 recognition motif of σ 70 that result in defects in phage λ repressor-stimulated transcription', *Journal of Bacteriology*, 176(10), pp. 2991–2998. doi: 10.1128/jb.176.10.2991-2998.1994.
- Mahajna, J. *et al.* (1986) 'Translation initiation of bacteriophage lambda gene cII requires integration host factor.', *Journal of Bacteriology*, 165(1), pp. 167–174. doi: 10.1128/jb.165.1.167-174.1986.
- Mishra, S. *et al.* (2013) 'The interaction surface of a bacterial transcription elongation factor required for complex formation with an antiterminator during transcription antitermination', *Journal of Biological Chemistry*, 288(39), pp. 28089–28103. doi: 10.1074/jbc.M113.472209.
- Murakami, K. S. *et al.* (2002) 'Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex', *Science*, 296(5571), pp. 1285–1290. doi: 10.1126/science.1069595.
- Murchland, I. *et al.* (2014) 'Promoter activation by CII, a potent transcriptional activator from bacteriophage 186', *Journal of Biological Chemistry*, 289(46), pp. 32094–32108.
- Neufing, P. J. *et al.* (1996) 'The CII protein of bacteriophage 186 establishes lysogeny by activating a promoter upstream of the lysogenic promoter', *Molecular microbiology*, 21(4), pp. 751–761.
- Nickels, B. E. *et al.* (2002) 'Protein-protein and protein-DNA interactions of σ 70 region 4 involved in transcription activation by λ cl', *Journal of Molecular Biology*, 324(1), pp. 17–34. doi: 10.1016/S0022-2836(02)01043-4.
- Oppenheim, A. B. *et al.* (2005) 'Switches in bacteriophage lambda development', *Annual Review of Genetics*, 39(1), pp. 409–429. doi: 10.1146/annurev.genet.39.073003.113656.
- Pabo, C. O. and Lewis, M. (1982) 'The operator-binding domain of λ repressor: Structure and DNA recognition', *Nature*, 298(5873), pp. 443–447. doi: 10.1038/298443a0.
- Pinkett, H. W. *et al.* (2006) 'The structural basis of cooperative regulation at an alternate genetic switch', *Molecular cell*, 21(5), pp. 605–615.
- Portelli, R. *et al.* (1998) 'The late-expressed region of the temperate coliphage 186 genome', *Virology*, 248(1), pp. 117–130.
- Ptashne, M. (2004) *A genetic switch 3rd edition*, CSHL Press. doi: 10.1038/nrc1424.
- Reed, M. R. *et al.* (1997) 'The dual role of Apl in prophage induction of coliphage 186', *Molecular Microbiology*, 23(4), pp. 669–681. doi: 10.1046/j.1365-2958.1997.2521620.x.
- Richardson, H. and Egan, J. B. (1989) 'DNA replication studies with coliphage 186. II. Depression of host replication by a 186 gene', *Journal of Molecular Biology*, 206(1), pp. 59–68. doi: 10.1016/0022-2836(89)90523-8.
- Richardson, H., Puspurs, A. and Egan, J. B. (1989) 'Control of gene expression in the P2-related temperate coliphage 186. VI. Sequence analysis of the early lytic region', *Journal of Molecular Biology*, 206(1), pp. 251–255. doi: 10.1016/0022-2836(89)90539-1.
- Roberts, J. W. *et al.* (1998) 'Antitermination by bacteriophage λ Q protein', in *Cold Spring*

Harbor Symposia on Quantitative Biology. doi: 10.1101/sqb.1998.63.319.

Schmeissner, U. *et al.* (1980) 'Promotor for the establishment of repressor synthesis in bacteriophage λ ', *Proceedings of the National Academy of Sciences of the United States of America*, 77(6 I), pp. 3191–3195. doi: 10.1073/pnas.77.6.3191.

Schmeissner, U. *et al.* (1981) 'Positively activated transcription of λ integrase gene initiates with UTP in vivo', *Nature*, 292(5819), pp. 173–175. doi: 10.1038/292173a0.

Schubert, R. A. *et al.* (2007) 'Cro's role in the CI-Cro bistable switch is critical for λ 's transition from lysogeny to lytic development', *Genes and Development*, 21(19), pp. 2461–2472. doi: 10.1101/gad.1584907.

Shearwin, K. E., Dodd, I. B. and Egan, B. J. (2002) 'The helix-turn-helix motif of the coliphage 186 immunity repressor binds to two distinct recognition sequences', *Journal of Biological Chemistry*, 277(5), pp. 3186–3194. doi: 10.1074/jbc.M107740200.

Shearwin, K. E. and Egan, J. B. (2000) 'Establishment of Lysogeny in Bacteriophage 186 DNA BINDING AND TRANSCRIPTIONAL ACTIVATION BY THE CII PROTEIN', *Journal of Biological Chemistry*, 275(37), pp. 29113–29122.

Shearwin, K. E. and Truong, J. Q. (2019) 'Lysogeny', *Reference Module in Life Sciences*. doi: 10.1016/b978-0-12-809633-8.20963-1.

Sivaprasad, A. V. *et al.* (1990) 'DNA replication studies with coliphage 186. III. A single phage gene is required for phage 186 replication', *Journal of Molecular Biology*, 213(3), pp. 449–463. doi: 10.1016/S0022-2836(05)80207-4.

Stayrook, S. *et al.* (2008) 'Crystal structure of the λ repressor and a model for pairwise cooperative operator binding', *Nature*, 452(7190), pp. 1022–1025. doi: 10.1038/nature06831.

Trusina, A. *et al.* (2005) 'Functional alignment of regulatory networks: a study of temperate phages', *PLoS Computational Biology*, 1(7), p. e74.

Wang, H. *et al.* (2013) 'Single molecule analysis of DNA wrapping and looping by a circular 14mer wheel of the bacteriophage 186 CI repressor', *Nucleic Acids Research*, 41(11), pp. 5746–5756.

Woods, W. H. and Egan, J. B. (1974) 'Prophage induction of noninducible coliphage 186', *Journal of Virology*, 14(6), pp. 1349–1356.

2 STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF THE CII PROTEIN

The following chapter is an unpublished manuscript detailing our structural and biochemical characterization of the 186 CII transcriptional activator. In this manuscript, we presented several X-ray crystal structures of the CII protein and investigated the mechanism of action of this protein *in vitro* and *in vivo*. This manuscript was written with the intention to submit to Nucleic Acids Research. I performed all the experiments in this manuscript, with the exception of the collection of mass spectrometry data on the CII and CII-DNA samples. I also performed data analysis, prepared figures for publication, deposited crystal structures in the Protein Data Bank and wrote the manuscript. This manuscript has not yet been submitted.

Statement of Authorship

Title of Paper	Crystal structure of 186 CII: a tetrameric transcriptional activator that employs Class I and Class II activation mechanisms
Publication Status	<input type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input checked="" type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	This draft manuscript presents the first X-ray crystal structures of 186 CII and our biochemical characterization of the protein <i>in vitro</i> and <i>in vivo</i> . We also present a structural model on the mechanism of action of this transcriptional activator.

Principal Author

Name of Principal Author (Candidate)	Jia Quyen Truong		
Contribution to the Paper	The principal author performed all the experiments in this manuscript, with the exception of the collection of mass spectra on the CII and CII-DNA samples that they prepared. They also performed data analysis, prepared figures, deposited crystal structures in the Protein Data Bank and wrote the manuscript.		
Overall percentage (%)	60%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	16/12/2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- the candidate's stated contribution to the publication is accurate (as detailed above);
- permission is granted for the candidate to include the publication in the thesis; and
- the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Tara Pukala		
Contribution to the Paper	Collected mass spectrometry data on CII		
Overall percentage (%)	5		
Signature		Date	16/12/2020

Name of Co-Author	Santosh Panjikar		
Contribution to the Paper	Assisted in the structure solution on one of the CII structures (condition 1)		
Overall percentage (%)	5		
Signature		Date	16/12/2020

Name of Co-Author	John Bruning		
Contribution to the Paper	Provided guidance and contributed reagents for protein purification and crystallography experiments; aided in the preparation of the manuscript		
Overall percentage (%)	10		
Signature		Date	17/12/2020

Name of Co-Author	Ian Dodd		
Contribution to the Paper	Assisted in the design of experiments, contributed reagents and analytical tools, analysis of data and the development of the structural model to explain CII activity		
Overall percentage (%)	10		
Signature		Date	16/12/2020

Name of Co-Author	Keith Shearwin		
Contribution to the Paper	Assisted in the design of experiments, contributed reagents and analytical tools, analysis of data, the development of the structural model to explain CII activity and drafting of the manuscript		
Overall percentage (%)	10		
Signature		Date	16/12/2020

CRYSTAL STRUCTURE OF BACTERIOPHAGE 186 CII: A TETRAMERIC TRANSCRIPTIONAL ACTIVATOR THAT EMPLOYS CLASS I AND CLASS II ACTIVATION MECHANISMS

Jia Truong¹, Tara Pukala², Santosh Panjekar³, John Bruning¹, Ian Dodd¹, Keith Shearwin^{1*}

¹Department of Molecular and Cellular Biology, School of Biological Science, University of Adelaide, Adelaide, South Australia 5005, Australia

²Department of Chemistry, School of Physical Sciences, University of Adelaide, Adelaide, South Australia 5005, Australia

³Australian Synchrotron, Clayton, Victoria 3168, Australia

*Corresponding email: keith.shearwin@adelaide.edu.au

Abstract

The temperate bacteriophage 186, after infecting its host bacterium *Escherichia coli*, can follow either the lytic or the lysogenic developmental pathways. Crucial to this developmental decision is the lysogeny promoting CII protein. This potent transcriptional activator activates the early lysogenic promoter p_E at least 400-fold, to build up sufficient immunity repressor levels so that a significant fraction of infections commit to lysogeny. Its potency and its unusual property of binding to operator sites separated by 20 base pairs, center-to-center, suggest it may activate the p_E promoter by a novel mechanism. Three crystal structures of the CII protein were solved to 2-3 Å resolution. The structures reveal a tetrameric arrangement of CII, shown to be necessary for DNA binding by mutational analysis and native mass-spectrometry. CII is degraded *in vivo* into a specific transcriptionally inactive product. The crystal structures reveal a reduced self-association of the degradation product and explain the loss of activity. Modelling of the CII-RNA polymerase complex at the p_E promoter suggests that CII activates p_E via recruitment of RNA polymerase through both σ^{70} and α -CTD contacts, in an arrangement different to previously studied transcriptional activators.

Introduction

Bacteriophage λ has long been used as a model system to study many aspects of prokaryotic gene regulation (Ptashne, 1986). As a temperate bacteriophage, λ makes a developmental decision between the lytic and lysogenic regimes. In λ , the decision between the lytic and lysogenic states is dependent on a transcriptional regulator CII. λ CII is produced soon after λ infection, under the control of the lytic promoter λp_R . λ CII activates three promoters p_{RE} , p_I and p_{aQ} , which drive the expression of the immunity repressor (λ CI), integrase (λ Int) for site specific integration of the genome, and an antisense RNA (anti-Q) to reduce late gene

expression (Oppenheim *et al.*, 2005) (Figure 1A). All three products promote lysogeny in λ (Herskowitz and Hagen, 1980). If insufficient λ CII produced, the default lytic cycle is followed.

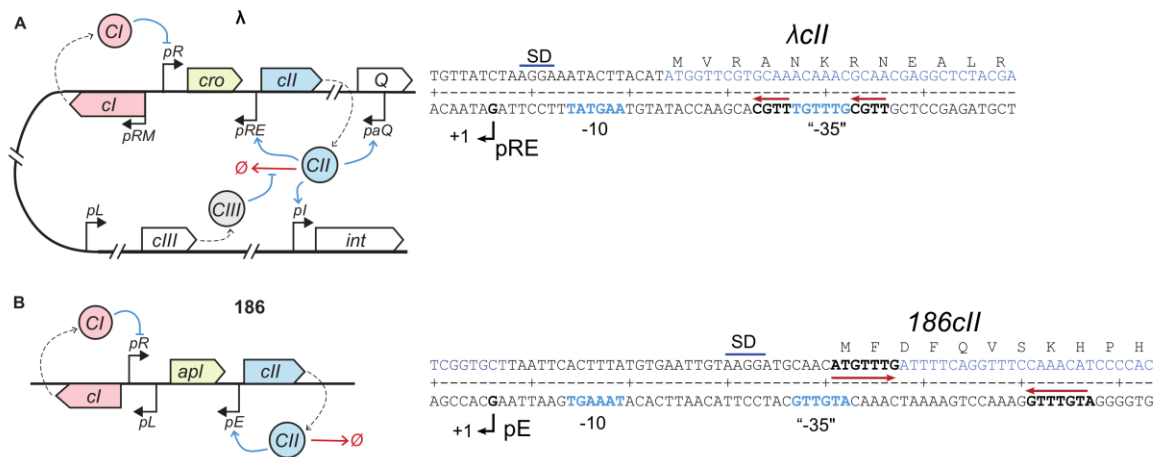


Figure 1 – Schematic diagram illustrating the role of CII in the developmental switch in bacteriophage λ and 186. (A) The λ p_R promoter drives the expression λ CII. λ CII activates the p_{RE} , p_{aQ} and p_I promoters to promote lysogeny. The sequence of p_{RE} promoter is shown, with -10 and “-35” hexamers in blue, and the direct repeat 4-mer λ CII half sites indicated with red arrows. The -35 promoter element is poorly defined. (B) In 186, the p_R promoter drives the expression 186 CII. 186 CII activates the p_E promoter to promote lysogeny. The sequence of p_E promoter is shown, with -10 and “-35” hexamers in blue, and the inverted repeat 7-mer CII half sites indicated with a red arrow. Figure adapted from Murchland *et al.* (2021).

Although λ has made many seminal contributions to field of gene regulation, it represents just one example of bacteriophage decision making. Bacteriophage 186 is an evolutionarily distinct UV inducible temperate phage belonging to the P2 family, that shares a similar network topology to λ (Trusina *et al.*, 2005). By studying bacteriophage 186 as a counterpoint to λ , more general principles in temperate bacteriophage regulation may be derived.

186 also depends on a transcriptional activator protein named CII (186CII and λ CII will be used for clarity) to establish lysogeny, in a concentration dependent manner (Lamont *et al.*, 1993; Neufing *et al.*, 1996; Neufing, Shearwin and Egan, 2001; Murchland *et al.*, 2014, 2021). 186 CII is produced during early infection from the early lytic promoter 186 p_R . 186 CII transiently activates the p_E promoter to generate initial pools of the immunity repressor and the integrase protein (Figure 1B). The immunity repressor represses the lytic promoters, whilst the integrase gene catalyzes the integration of the phage genome into the bacterial chromosome (Neufing *et al.*, 1996).

186 CII recruits RNA polymerase (RNAP) to activate transcriptional activity. The mechanism by which 186 CII activates the p_E promoter is unknown but unusually, mutagenesis studies suggest contacts with both RNAP σ^{70} and the α -CTD (Murchland *et al.*, 2014). 186 CII binds to inverted half sites (ATGTTTG) separated by two turns of the DNA helix, bends the DNA and is quite intolerant to even small changes to spacing (Neufing *et al.*, 1996; Shearwin and Egan, 2000). This spacing is very uncommon among prokaryotic activators and when considered with its ability to activate its cognate p_E promoter approximately 400-fold, suggests a possible novel mechanism of RNAP recruitment.

In this study, we present several crystal structures of the CII protein to show that CII is tetrameric and this oligomeric arrangement, in turn, dictates the half site spacing and the protein's intolerance to spacing changes. Native mass spectrometry and *in vivo* p_E-lacZ experiments are used to validate a model of tetrameric CII-DNA binding. Residue specific interactions between CII and RNAP were probed via alanine scanning mutagenesis. Finally, molecular modelling was used integrate the various structural and biochemical information into a model of the promoter-CII-RNA polymerase complex. This model proposes that the RNAP is recruited through its σ^{70} subunit by CII and that additional stabilization contacts are made with the α -CTD. Specifically, the α -CTD is proposed to contact promoter DNA between the two half sites, positioned into a binding site formed by the CII protein and bent DNA.

This mechanism of anchoring the α -CTD within a protein-DNA cavity is distinct from other well-characterized prokaryotic transcriptional activators.

Materials and Methods

Oligonucleotides

Oligonucleotides used in this study are summarized in Table S 1.

Plasmids and bacterial strains

Plasmids and bacterial strains used in this study are summarised in Table 1 and Table 2, respectively. A number of variants of 186 CII have been used in this study. The nomenclature CII_x, refers to CII truncated at the x-th amino acid. Amino acid substitutions of CII are denoted as superscripts (CII^{Q93A} refers to CII where glutamine 93 has been mutated to alanine. CII^{Q93A}145 refers to CII^{Q93A} truncated at the 145th amino acid).

For pE-lacZ reporter assays, CII derivatives were expressed from a pZS45 plasmid (Murchland *et al.*, 2014), which follows the modular design and nomenclature of the pZ system (Lutz and Bujard, 1997).

For lacZ assays investigating the interactions between CII and the RNAP α subunit, RNAP α subunit mutant variants were expressed in IM50 (IM26 reporter strain, harbouring the pZS45 CII^{E46K} expression vector) from pHTf1-rpoA and pREII α -rpoA plasmids, from a previously described scanning alanine mutagenesis library (Kainz and Gourse, 1998). Alanine substitution mutants at residues 255-271 and 273-329 were expressed from the pHTf1-rpoA and pREII α -rpoA plasmids, respectively.

Table 1 – Plasmids used in this study

Plasmid name	Brief description	Features	Reference
pET15b-CII135 pET15b-CII145 pET15b-CII163	Vectors to overexpress different truncations of CII. Protein expression driven from an IPTG-inducible T7 promoter.	ColE1 ori; Ampicillin resistance. N-terminal His ₆ tag.	Murchland <i>et al.</i> (2014)
pZS45-CII pZS45-CII _{xxx} pZS45-CII ^{Q93A} pZS45-CII ^{V96A} pZS45-CII ^{M97A} pZS45-CII ^{M97C} pZS45-CII ^{M100A} pZS45-CII ^{M100C} pZS45-CII ^{V111A} pZS45-CII ^{S112A}	Low copy vector to allow IPTG-inducible expression of various CII and CII-derivatives from the p _{lac} promoter. 'xxx' denotes the amino acid at which CII is truncated (ranges between 141 and 169 in this study). Amino acid mutations are written in superscript	pZ modular plasmid; SC101 origin; spectinomycin resistance	Murchland <i>et al.</i> (2014) and this work
pZS45	Low copy vector used as parental control in lacZ assays	pZ modular plasmid; SC101 origin; spectinomycin resistance	Murchland <i>et al.</i> (2014) and Lutz and Bujard (1997)
pUHA-1	Used to constitutively supply lacI from p _{lac} promoter in lacZ assays	p15a ori; kanamycin resistance; wild-type lacI ORF and promoter	H. Bujard, Heidelberg University, Germany
pIT3CL-pE-LacZ	p _E lacZ reporter construct ; p _E (-120 to +115) is upstream of	R6K-γ ori; chloramphenicol resistance; can be integrated into λ attB	Murchland <i>et al.</i> (2014)

	a RNaseIII cleavage site and a lacO2- lacZ reporter gene		
pIT3CH- pCIIR-LacZ	pCIIR lacZ reporter construct	R6K- γ ori; chloramphenicol resistance; can be integrated into HK022 attB	Murchland <i>et al.</i> (2014)
pHTf1- rpoA pREII-rpoA	Series of alanine scanning mutagenesis vectors to express single amino acid substitutions of RNA polymerase α subunit. Expression is IPTG-inducible.	Ampicillin resistance	Kainz and Gourse (1998)

Table 2 – Bacterial strains used in this study

Strain	Genotype	Notes	Supplier/ Reference
HMS174(λ DE3)	F- <i>recA1 hsdR</i> ($r_{K12^-} m_{K12^+}$) (DE3) (Rif ^R)	Strain used to overexpress 186 CII and derivatives for protein purification	Merck
T7 crystal express	<i>fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--Tet^S)2 [dcm] R(zgb-210::Tn10--Tet^S) endA1 metB1 Δ(mcrC-mrr)114::IS10</i>	Strain used to overexpress selenomethionine labelled CII145	New England Biolabs
BW25113	(F-(araD-araB) 567 lacZ4787(::rrnB-3)-rph-1 (rhaD-rhaB)568 hsdR514	Reporter gene modules were integrated into the bacterial chromosome of this strain	Datsenko and Wanner (2000)
IM26	BW25113 [pIT3CL-pE-lacZ] λ + pUHA-1	Strain used to measure in vivo activation of pE. pE:lacZ reporter integrated at the λ attachment site	Murchland et al. (2014)
IM13	BW25113 [pIT3CH-pCIIR-lacZ] _{HK022} + pUHA-1	Strain to assay CII binding to DNA in vivo The pCIIR:lacZ reporter was integrated at the HK022 attachment site. pCIIR is an artificial promoter repressed by CII binding	Murchland et al. (2014)
IM50	IM26 + pZS45-CII ^{E146K}	Strain used to measure the effect of RNA polymerase alpha CTD mutants on activation of the p _E promoter	Murchland et al. (2014)

Expression and purification of CII135 and CII145

CII135 and CII145 were overexpressed in HMS174(λ DE3) cells carrying the pLysS plasmid as described by Murchland *et al.* (2014). Selenomethionine labelled CII145 was overexpressed in T7 Express Crystal cells. T7 crystal express cells harbouring the pET15b CII145 plasmid was grown overnight in 5 mL of LB supplemented with 100 $\mu\text{g mL}^{-1}$ ampicillin. The cells were diluted 1 in 200 into M9 media (1 x M9 salts, 0.4% glucose, 1 mM MgSO_4 , 0.1 mM CaCl_2 , 0.0002% ferric ammonium sulfate) supplemented with 100 $\mu\text{g mL}^{-1}$ ampicillin and 50 $\mu\text{g mL}^{-1}$ L-methionine. Cultures were grown at 37°C with shaking until OD_{600} 0.8. Cells were pelleted at 5000 x g for 20 mins and resuspended in an equal volume of culture media without methionine supplementation. The culture was incubated for 2.5 h to deplete methionine from the cells. Solid DL-selenomethionine was added to the culture to give a final concentration of 100 $\mu\text{g mL}^{-1}$ and cultures grown for a further 30 mins. IPTG (isopropyl β -D-1-thiogalactopyranoside) was added to a final concentration of 0.2 mM to induce labelled protein expression and the culture was grown overnight at 16°C. Cells from protein overexpression were harvested by centrifugation at 5000 x g for 20 mins. Cell pellets were stored at -80°C.

For purification of CII variants from cell pellets, cell pellets were thawed and resuspended in START buffer (1 x TBS, 350 mM NaCl, 20 mM imidazole, 5 mM β -mercaptoethanol, pH 7.2). Cell resuspensions were lysed by sonication on ice (5 x 30s, 50% duty cycle, Sonifier Cell Disruptor B-30). Lysate was clarified by centrifugation at 40 000 x g for 1 hour and subsequent filtration through a 0.2 μm syringe filter (Sartorius Minisart) using a disposable syringe. Supernatant was subjected to Ni^{2+} -IMAC chromatography on a 5 mL HisTrap FF column (GE Healthcare) using NGC chromatography FPLC system (BioRad). The column was washed with 10 column volumes of START buffer. The protein was eluted from the column with START buffer with a gradient of imidazole from 70 mM to 220 mM over 8 column volumes, followed by a gradient of imidazole from 220 mM to 500 mM over two column volumes. Fractions showing UV absorbance A_{280} above baseline were tested for purity using SDS-PAGE. Fractions containing the purified protein were pooled, dialysed against 1 x TBS, 50 mM EDTA pH 8, 10% glycerol, 5 mM β -mercaptoethanol. In the case of CII145, the protein was further purified using size exclusion chromatography on the Enrich 650 size exclusion column (Biorad) into TEG150 buffer (1 x TBS, 50 mM EDTA, 10% glycerol, 5 mM BME). For crystallography, CII135 and CII145 proteins were concentrated to 8.5 and 10-24 mg mL^{-1} respectively using the Amicon Ultra-15 centrifugation units (Millipore). For mass spectrometry analysis, proteins were dialysed into 0.3 M ammonium acetate and concentrated to 80-120 μM using the Amicon Ultra-15 centrifugation units (Millipore). Prior to use in crystallography or mass-spectrometry, proteins were flash frozen in liquid nitrogen and stored at -80°C.

Crystallization of CII145, CII135 and selenomethionine derivatized CII145

Crystallization screening for the CII135 and CII145 protein was conducted at 16°C using the sitting drop vapor diffusion method in 96-well Intelliplates (Art Robbins) using conditions from commercial screens. Commercial screens used were PEG/ION HT, Index, Natrix HT, Crystal screen HT (Hampton Research) and the MCSG suite (Microlytic). 1 μL of protein was added to 1 μL of reservoir solution and equilibrated over 75 μL reservoir solution. Crystals were observed in several conditions. Diffraction from several of those crystals were acquired and

used to solve the structure of the CII protein. Details of the crystallization conditions and cryoprotection procedures were summarized in Table 3. A notable attempt to derivatize the CII135 crystal with polyiodide ions was made by adding 0.1 μL 0.47 M I_2 /0.67 M KI solution directly to the crystallization drop before cryoprotection (Evans and Bricogne, 2002). The crystal visually degraded after 30 seconds so the crystal was immediately cryoprotected and frozen in liquid nitrogen.

Table 3 – Crystallization and cryoprotection conditions for CII proteins

Protein	Concentration of protein for crystallization (mg mL^{-1})	Condition	Crystal growth observations	Cryoprotection procedure
CII135	8.5	0.2 M ammonium acetate, 0.1 M HEPES pH 7.5, 25% w/v polyethylene glycol 3350	Crystal observed after 10 months	Exclude water by passing crystal through Parabar 10312 and freezing in liquid nitrogen
CII145 (condition 1)	10	8% Tascimate pH 8.0, 20% PEG 3350	Crystals observed within 3 days	Pass crystal through 17% glycerol in reservoir solution and freezing in liquid nitrogen
Selenomethionine labelled CII145 (condition 1)	9.9	8% Tascimate pH 8.0, 20% PEG 3350	Crystals observed within 3 days and reach maximum size after 14 weeks	Pass crystal through 17% glycerol in reservoir solution and freezing in liquid nitrogen
CII145 (condition 2)	4.8	2 M ammonium sulfate, 0.1 M Bis-Tris: HCl, pH 5.5	Crystal observed after 10 months	Exclude water by passing crystal through Parabar 10312 and freezing in liquid nitrogen

Solving the structure of selenomethionine-labelled CII145

A 3.2 Å resolution MAD data set for selenomethionine CII145 protein was collected at wavelengths of 0.9790 Å (peak), 0.9791 Å (inflection) and 0.9537 Å (high-energy remote) using an ADSC Quantum 210r CCD on the macromolecular beamline MX1 at the Australian Synchrotron. For each dataset, 360 diffraction images with a 1° oscillation width were collected with the crystal-to-detector distance set to 250 mm (Table 4). The diffraction data was processed using XDS (Kabsch, 2010). The initial phases were obtained using 3W-MAD phasing protocol of the Auto-Rickshaw server (Panjikar *et al.*, 2005, 2009). Twenty eight

selenium atoms were placed using SHELXD (Schneider and Sheldrick, 2002) in space group P21. The sites were refined using BP3 (Pannu and Read, 2004), density modification and two-fold NCS averaging was performed on the resulting phases using RESOLVE (Terwilliger, 2004). The density-modified phases were then used for model building in BUCCANEER (Cowtan, 2006). The resulting model was further improved manually in COOT (Emsley *et al.*, 2010). The improved model was then used as a starting model in the MRSAD protocol of Auto-Rickshaw (Panjikar *et al.*, 2009) against the PEAK dataset for further iterative phase improvement, model building and model refinement. After 5 MRSAD iterative cycles including SAD refinement, the model resulted with 90% of complete model with R/R_{free} 23.4%/29.0%.

Solving the structure of CII145 (condition 1)

A 2.1 Å resolution data for native CII145 protein was collected at 0.95 Å wavelength using an ADSC Quantum 315 CCD on the micro-focus beamline MX2 at the Australian Synchrotron. 358 diffraction images were collected with a 1° oscillation width, with crystal-to-detector distance set to 300 mm. Diffraction data was processed and scaled using iMosflm and Scala (Battye *et al.*, 2011; Evans and Murshudov, 2013). The resolution limit was chosen to ensure CC_{1/2} > 0.3. The data processing statistics of the diffraction from this crystal are listed in Table 4. The crystal structure was solved by molecular replacement using Phaser MR (McCoy *et al.*, 2007), with the model built for the selenomethionine-labelled protein used as a search model. Each asymmetric unit contained 4 monomers. The model was iteratively refined and rebuilt using phenix.refine (Afonine *et al.*, 2012) and COOT (Emsley *et al.*, 2010), respectively, until R factors converged. Molprobity was used for model validation (Chen *et al.*, 2010). Final refinement statistics are shown in Table 5.

Structure solution of CII145 (condition 2)

A 2.1 Å resolution data for native CII145 protein in condition 2 was collected at 1.4586 Å wavelength using an ADSC Quantum 315 CCD on the micro-focus beamline MX2 at the Australian Synchrotron. 360 diffraction images were collected at 1° oscillation width, with crystal-to-detector distance set to 200 mm. Diffraction images were processed as described for CII145 condition 1, with processing statistics summarized in Table 4. The structure was solved using the MR-SAD module of Auto-Rickshaw (Panjikar *et al.*, 2009). The SAD procedure of the module was not used. Molecular replacement was performed using MOLREP (Vagin and Teplyakov, 1997), with chains A and D of the CII145 (condition 1) structure used as a search model. CNS (Brünger *et al.*, 1998), REFMAC5 (Murshudov *et al.*, 2011) and phenix.refine (Afonine *et al.*, 2012) were used to refine to structure to a model with R_{work}/R_{free} 0.2627/0.3135. An attempt to identify heavy atom sites was performed with Phaser (McCoy *et al.*, 2007), with heavy atom refinement performed with MLPHARE (Collaborative Computational Project Number 4, 1994), SHARP (De La Fortelle and Bricogne, 1997) and BP3 (Pannu and Read, 2004), without successfully identifying any heavy atom sites. Iterations of model rebuilding and refinement were completed using COOT (Emsley *et al.*, 2010) and phenix.refine (Afonine *et al.*, 2012), respectively, until R factors converged. Molprobity was used for model validation (Chen *et al.*, 2010). Final refinement statistics are shown in Table 5.

Structure solution of CII135

A 2.75 Å resolution native data for native CII135 protein was collected at 0.95 Å wavelength using an ADSC Quantum 210r CCD detector on the MX1 beamline at the Australian

Synchrotron. 241 diffraction images were collected at 1° oscillation width, with crystal-to-detector distance set to 250 mm. Diffraction images were processed as described for CII145 condition 1, with processing statistics summarized in Table 4. The structure was solved using the MR module of Auto-Rickshaw (Panjikar *et al.*, 2009). The search model for molecular replacement used was CII145 structure in condition 1, with residues after E106 removed. Molecular replacement was performed using MOLREP (Vagin and Teplyakov, 1997). CNS (Brünger *et al.*, 1998), REFMAC5 (Murshudov *et al.*, 2011) and phenix.refine (Afonine *et al.*, 2012) were used to refine to structure to a model with $R_{\text{work}}/R_{\text{free}}$ 0.2939/0.2925. Iterations of model rebuilding and refinement were completed using COOT (Emsley *et al.*, 2010) and phenix.refine (Afonine *et al.*, 2012), respectively, until R factors converged. Molprobit was used for model validation (Chen *et al.*, 2010). Final refinement statistics are shown in Table 5.

Table 4 – Data processing statistics for crystallography datasets. Statistics for highest resolution shell are shown in parentheses.

	Selenomethionine-labelled CII145					
	(condition 1)				(condition 2)	
	CII145 (condition 1)	PEAK	INFLECTION	REMOTE	CII145 (condition 2)	CII135
Wavelength (Å)	0.95	0.9790	0.9791	0.9537	0.95	0.95
Total images	358	360	360	360	360	241
Oscillation (°)	1.00	1.00	1.00	1.00	1.00	1.00
Space group	P2 ₁	P2 ₁	P2 ₁	P2 ₁	C2	P3 ₂ 21
Unit cell	a=55.47 b=53.51 c=106.35 α=90.00 β=100.88 γ=90.00	a=54.31 b=58.62 c=106.92 α=90.00 β=97.15 γ=90.00	a=54.31 b=58.60 c=106.89 α=90.00 β=97.15 γ=90.00	a=54.35 b=58.65 c=106.99 α=90.00 β=97.14 γ=90.00	a=167.86 b=46.14 c=38.31 α=90.00 β=99.42 γ=90.00	a=59.86 b=59.86 c=112.97 α=90.00 β=90.00 γ=120.00
Resolution (Å)	2.10	3.20	3.20	3.20	2.21	2.75
Mosaicity (°)	1.61	0.160	0.161	0.161	0.830	0.820
Total reflections	233656 (19505)	82503 (13032)	82766 (13014)	82793 (13015)	99947 (8094)	80395 (11988)
Unique reflections	35935 (2958)	21360 (3383)	21358 (3374)	21389 (3357)	14298 (1197)	6505 (939)
Redundancy	6.5 (6.6)	3.86 (3.85)	3.88 (3.86)	3.87 (3.88)	7.0 (6.8)	12.4 (12.8)
Completeness (%)	99.6 (99.6)	99.2 (99.1)	99.2 (98.9)	99.1 (98.3)	97.8 (95.5)	99.9 (100)
Average I/σ(I)	5.0 (1.3)	11.50 (3.91)	10.41 (3.33)	8.75 (2.71)	11.2 (3.4)	4.9 (2.5)
CC _{1/2} ^c	98.6 (52.8)	99.4 (93.5)	99.2 (89.8)	99.0 (86.9)	99.5 (77.5)	94.8(75.3)
CC _{1/2} anom	-	54 (20)	19 (10)	29 (6)	-	-
Rmerge ^a (%)	0.233 (136.9)	11.6 (35.4)	12.9 (42.1)	16.0 (52.1)	14.4 (85.4)	46.8 (126.7)
R-meas ^b (%)	27/162.2	13.5 (41.1)	15.0 (48.8)	18.6 (60.5)	15.5 (92.6)	50.1 (136.6)
Wilson-B (Å ²)	22.51	41.33	43.19	43.00	18.601	20.363

$${}^a R_{merge} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)} \text{ (Arndt, Crowther and Mallett, 1968)}$$

$${}^b R_{meas} = \frac{\sum_{hkl} \{N(hkl) / [N(hkl) - 1]\}^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)} \text{ (Diederichs and Karplus, 1997)}$$

$${}^c CC_{1/2} = \frac{\sum (x - \langle x \rangle)(y - \langle y \rangle)}{[\sum (x - \langle x \rangle)^2 \sum (y - \langle y \rangle)^2]^{1/2}} \text{ (Karplus and Diederichs, 2012)}$$

Table 5 – Refinement statistics of CII structures. Statistics for the highest-resolution shell are shown in parentheses.

	CII145 (condition 1)	CII145 (condition 2)	CII135
Molecular mass (kDa)	18.33	18.33	17.33
Resolution (Å)	2.1	2.21	2.75
Reflections used in refinement	35898 (3562)	14296 (1395)	6481 (642)
Reflections used for R-free	1995 (194)	1421 (139)	1092 (118)
R-work ^a	0.2209 (0.2944)	0.2280 (0.2656)	0.2122 (0.2771)
R-free ^b	0.2606 (0.3489)	0.2829 (0.3243)	0.2651 (0.2867)
Number of non-hydrogen atoms			
total	4899	2340	1402
macromolecules	4596	2191	1338
ligands	16	6	
solvent	287	143	64
Protein residues	594	285	167
RMSD bonds (Å)	0.002	0.005	0.004
RMSD angles (°)	0.49	0.66	0.94
Ramachandran favored (%)	97.61	96.80	97.55
Ramachandran allowed (%)	2.05	2.49	2.45
Ramachandran outliers (%)	0.34	0.71	0.00
Rotamer outliers (%)	0.00	0.42	0.00
Clashscore	4.16	9.69	4.98
Average B-factor (Å ²)			
overall	37.71	43.21	33.59
macromolecules	37.50	43.59	33.92
ligands	64.78	56.76	
solvent	39.66	36.79	26.69
Number of TLS groups	31	10	9

^a $R_{work} = \sum |F_o - F_c| / \sum |F_o|$ for all data with $F_o > 2\sigma(F_o)$, excluding data to calculate R_{free}

^b $R_{free} = \sum |F_o - F_c| / \sum |F_o|$ for all data with $F_o > 2\sigma(F_o)$, calculated with 5% of reflections that are randomly chosen (Brünger, 1992).

PDB codes

The coordinates and structure factors for CII145 (condition 1), CII145 (condition 2) and CII135 have been deposited in the Protein Data Bank as entries 6VLI, 6VPE and 6VHM, respectively.

Molecular modelling

Intramolecular residue-residue contacts between different CII monomers were investigated with the Protein Interaction Calculator webserver (Tina, Bhadra and Srinivasan, 2007). The crystal structure of CII145 (condition 1) was modified to remove residues from the N-terminus of the protein corresponding to the hexahistidine tag and thrombin cleavage sites, not present in the wild-type protein. This altered crystal structure was used for subsequent modelling. The

pE promoter -67 to -29 DNA was modelled using 3D-DART (van Dijk and Bonvin, 2009). The protein was docked onto the DNA using HADDOCK webserver (De Vries, Van Dijk and Bonvin, 2010; van Dijk and Bonvin, 2010). Full details on docking parameters used are outlined in the Supplemental Materials. The same segment of pE DNA was modelled in its native conformation using the bend.it server (Vlahoviček, Kaján and Pongor, 2003).

Thermus aquaticus RNAP polymerase was initially modelled onto the docked CII-DNA complex. The -35 DNA promoter element from a *T. aquaticus* σ_4 /-35 DNA complex (PDB ID 1KU7) (Campbell *et al.*, 2002) was superimposed onto the -35 promoter element of the CII-DNA complex. The positioned σ_4 /-35 structure was then used to position the structure of the *E. coli* RNAP holoenzyme open promoter complex (PDB ID 4YLN) (Zuo and Steitz, 2015) by superimposing the σ_4 subunit of both structures.

Hotspot residues in the putative tetramerization interface of CII were predicted using the KFC (Darnell, LeGault and Mitchell, 2008) and Robetta alanine scanning (Kim, Chivian and Baker, 2004) webserver. *In silico* mutagenesis of CII was carried out through ICM-Pro (Molsoft).

Molecular visualization

UCSF Chimera (Pettersen *et al.*, 2004) and PyMOL (DeLano, 2002) were used to analyse and visualize the crystal structures, CII-DNA and CII-DNA-RNAP structural models.

Preparation of DNA for mass-spectrometry

Single stranded oligonucleotides were chemically synthesized (Integrated DNA Technologies). Oligonucleotides were resuspended in IDTE buffer (10 mM Tris pH 8, 0.1 mM EDTA) to a final concentration of 3 mM. Complementary nucleotides were then mixed in a 1:1 ratio, heated to 90-95°C and allowed to slowly cool to room temperature. The annealed oligonucleotides were buffer exchanged into 0.3 M ammonium acetate using Illustra MicroSpin G-25 columns (GE Healthcare).

Mass-spectrometry examination of CII

Mass-spectrometry measurements were performed on a Synapt HDMS system (Waters, Manchester, UK). Electrospray from in-house prepared gold coated borosilicate capillaries was used to introduce the samples (Hernández and Robinson, 2007). The instrument parameters were optimized to remove adducts while preserving noncovalent interactions and were typically as follows: capillary voltage, 1.5 kV; cone voltage, 50 V; trap collision energy, 10 V; source temperature, 20°C; and backing pressure, 6 mBar.

Complementation assays

The functional activity of CII was tested by complementation. cII473 is a bacteriophage 186 clear-plaque mutant, with a gene encoding for defective cII (Huddleston, 1970). As expected, cII473 formed clear plaques when IM26 was used as an indicator. When IM26 carried the pZS45-CII plasmid, encoding functional CII, cII473 plaques were turbid, showing that plasmid-derived CII can complement the lack of active CII on the phage. Activity of various CII mutants was tested by plating cII473 onto IM26 strains expressing the mutant CII from pZS45 plasmids. Strains were grown on solid LB agar media supplemented with 100 μ M IPTG to induce CII expression. The activity of CII was qualitatively assessed from the turbidity of the plaques.

Kinetic lacZ assays

Kinetic lacZ assays were performed in 96 well microtitre plates, as previously described (Dodd *et al.*, 2001), with modifications described here. Details of reporter strains and growth media supplementation with antibiotics and IPTG for different lacZ experiments are summarized Table S 2 and Table S 3, respectively. All bacterial culture incubations were carried out at 37°C. Reporter strains were restreaked from glycerol stocks onto selective LB agar plates and incubated overnight. Fresh colonies from the selective plates were used to inoculate 100 µL of selective media in a well of the 96 well microtiter plate. Plates were sealed and incubated overnight with shaking. After overnight incubation, cells were diluted to OD₆₀₀ 0.003 in 100 µL fresh selective media, supplemented with a defined concentration of IPTG, in a microtitre plate well. Plates were sealed and incubated at 37°C until culture OD₆₀₀ reached 0.6-0.9. Cultures were assayed for lacZ activity as previously described (Dodd *et al.*, 2001). Two different lysis-assay reagents were used in this study. The original method for lysing cells involved adding 20 µL of culture to well of a microtitre plate containing 30 µL LB media, 150 µL of TZ8 (100 mM Tris-HCl, pH 8.0, 10 mM KCl, 1 mM MgSO₄), 40 µL of o-nitrophenyl-β-D-galactopyranoside (ONPG) (4 mg mL⁻¹ in TZ8), 1.9 µL of 2-mercaptoethanol and 0.95 µL of polymyxin B (20 mg mL⁻¹). To more accurately measure high lacZ units, some assays used an improved lysis reagent, where each well contained 20 µL LB media, 88 µL of TZ8, 60 µL of ONPG (4 mg mL⁻¹ in TZ8), 2 µL of hen egg white lysozyme (10 mg mL⁻¹ in TZ8), 4 µL of polymyxin B (20 mg mL⁻¹) and 6 µL MQ water.

Western blot assays

IM26 cells with pZS45 plasmids, containing variants of the CII gene under the control of the p_{lac} promoter (Table 1), were grown in LB supplemented with 50 µg mL⁻¹ spectinomycin, 50 µg mL⁻¹ kanamycin and 100 µM IPTG. Cells were harvested by centrifugation (7000 *x g*, 3 min) and pellets were stored at -20°C. Cells were incubated at 37°C to mid-log phase. Cells were lysed with B-PER (Pierce) and Benzonase (0.25 unit µL⁻¹) (Sigma Aldrich) and lysate was analyzed on 4-15% Tris-glycine gels (Bio-Rad) in Tris-glycine running buffer. After electrophoresis, a semi-dry transfer to a PVDF membrane was performed using Trans-Blot Turbo Transfer System (Bio-Rad) with PVDF transfer pack (Bio-Rad). The membranes were blocked with 5% BSA. 186 CII primary detection used anti-CII rabbit antisera (IMVS Veterinary Services). Secondary detection used Cy5-labelled ECL Plex Goat anti-rabbit IgG antibody (GE Healthcare). The membranes were scanned on a Chemidoc imaging system (Bio-Rad) using the system's Cy5 channel (Epi-red illumination at 625-650 nm with 695/55 emission filter) and images analysed by ImageLab (Bio-Rad).

Results

Truncating CII results in higher *in vivo* activity

In order to study the structure-function relationship of 186 CII, one of the aims of this study was to obtain an atomic resolution structure of 186 CII. 186 CII has a natural degradation tag at its C-terminus, which if truncated, extends the half-life of the protein leading to higher expression levels and higher transcriptional activity *in vivo* (Murchland *et al.*, 2014). Thus, the optimal construct to trial for crystallography is likely to be the CII truncation that displays the highest *in vivo* activity.

The 169 amino acid long 186 CII protein was systematically truncated, residue by residue, to give truncated proteins of between 141 and 168 residues long. The activities of these truncated proteins were assayed using an *in vivo* p_E lacZ reporter gene assay (Figure 2). Truncating CII even by one residue resulted in a statistically significant 38% increase in p_E activation. A further statistically significant increase of 32% was observed for CII166 to CII165, suggesting that the degradation tag at the C-terminus can be gradually removed. Further gradual increases in the activity of CII were observed, up to a maximal activity corresponding to CII145 and CII146. Truncating the protein to residue 141 resulted in a sharp reduction in activity. No statistically significant activity of CII141 was seen, relative to the empty control (Figure 2).

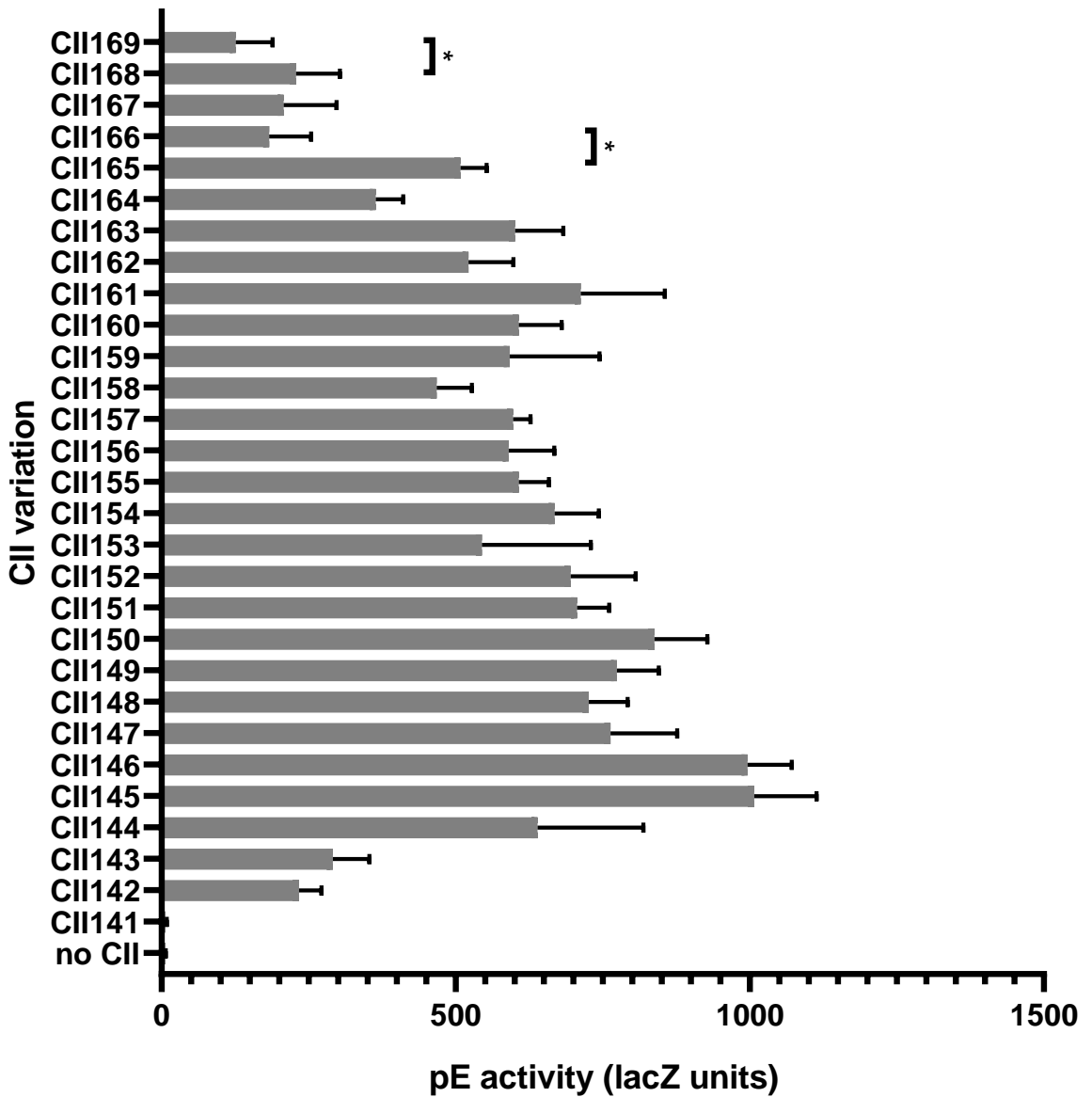


Figure 2 – Truncation of CII results in increased *in vivo* activation of the p_E promoter. Activity of p_E was measured using an *E. coli in vivo* p_E lacZ reporter system using lysis buffer containing hen egg white lysozyme (HEWL) to lyse cells. CII expression was induced by growing cells in 80 μM IPTG. Bars represent the mean. Error bars represent the 95% CI (n ≥ 8). * denotes p > 0.05. ** denotes p > 0.01.

CII forms a dimer of dimers

CII145 was chosen as a target for crystallization on the basis that it displayed the highest *in vivo* activity. 186 CII145 was crystallized and solved in two independent conditions, conditions 1 and 2 (Table 3) (see Material and Methods for full description of the conditions), with the structure within the asymmetric unit shown in Figure 3A and B, respectively. The crystals have the symmetry of the spacegroups in the space group $P2_1$ and $C2$, respectively. The structure of CII contained alpha helices and loops, with two separate domains joined by a linker. The N-terminal domain, containing 5 helices, is the DNA binding domain, containing a helix-turn-helix motif, made by helices 2 and 3 (Figure 3C). A typical section of the electron density map for both structures is shown in Figure S 1.

Previous studies have attempted to define the oligomeric state of the 186 CII protein. The best fit for data from sedimentation equilibrium experiments of the CII protein was a monomer-tetramer equilibrium (Neufing *et al.*, 1996). Subsequent SEC-MALS experiments with CII145 and CII169 detected tetramers in solution (Murchland *et al.*, 2014). Our native mass spectrometry measurements of CII reaffirm this tetrameric oligomeric state (Figure 4B). With this information, we attempted to define the CII biological assembly from the crystal structures using the PISA webserver (Krissinel and Henrick, 2007).

PISA analysis of CII145 structure from condition 1 and condition 2 predicts three potential quaternary complexes being stable in solution, two dimer complexes and a tetramer complex. In both cases, the tetrameric complex was predicted to be the most probable state. The tetramer predicted for the crystal structures in condition 1 and condition 2, corresponded to the structure in Figure 3A and Figure 3D, respectively.

In discriminating between a crystal contact and a true oligomeric interface, an interface area (defined as half the difference in total surface area and interfacing surface area) cutoff of 856 \AA^2 has been used to differentiate between dimer interfaces and artifacts of crystallization (Ponstingl, Henrick and Thornton, 2000). In the proposed tetrameric structures, two pairs of monomers (A/B and C/D in condition 1 and A/B in condition 2) have large interface areas between 2548 to 2949 \AA^2 , strongly suggesting an oligomeric interface. The interface area between monomers across these pairs is significantly lower than the cutoff, ranging from 175.8 \AA^2 – 422.5 \AA^2 . Thus, it is possible that a tetramer interface is only established after the initial dimerization of the pair of monomers with the larger interface areas.

Our evidence supports the CII protein oligomerizing into dimers and tetramers (as a dimer of dimers), with tetramerization mediated by a methionine zipper forming a four-helix bundle. There are eight methionine residues at the core of this tetramerization interface. Each monomer provides two methionine residues, M97 and M100 (Figure 4A). One notable aspect of the CII145 structure in condition 2 is the presence of two intramolecular disulfide bonds in each dimer. C78 forms a disulfide bond with C81 from the other protein chain in the dimer. However, these disulfide bonds are absent in the crystal structure of CII145 in condition 1. This suggests that when the CII proteins are purified from *E. coli*, disulfide bonds are not present. The disulfide bonds observed in condition 2 are likely to be artifacts of oxidation of protein during crystallization.

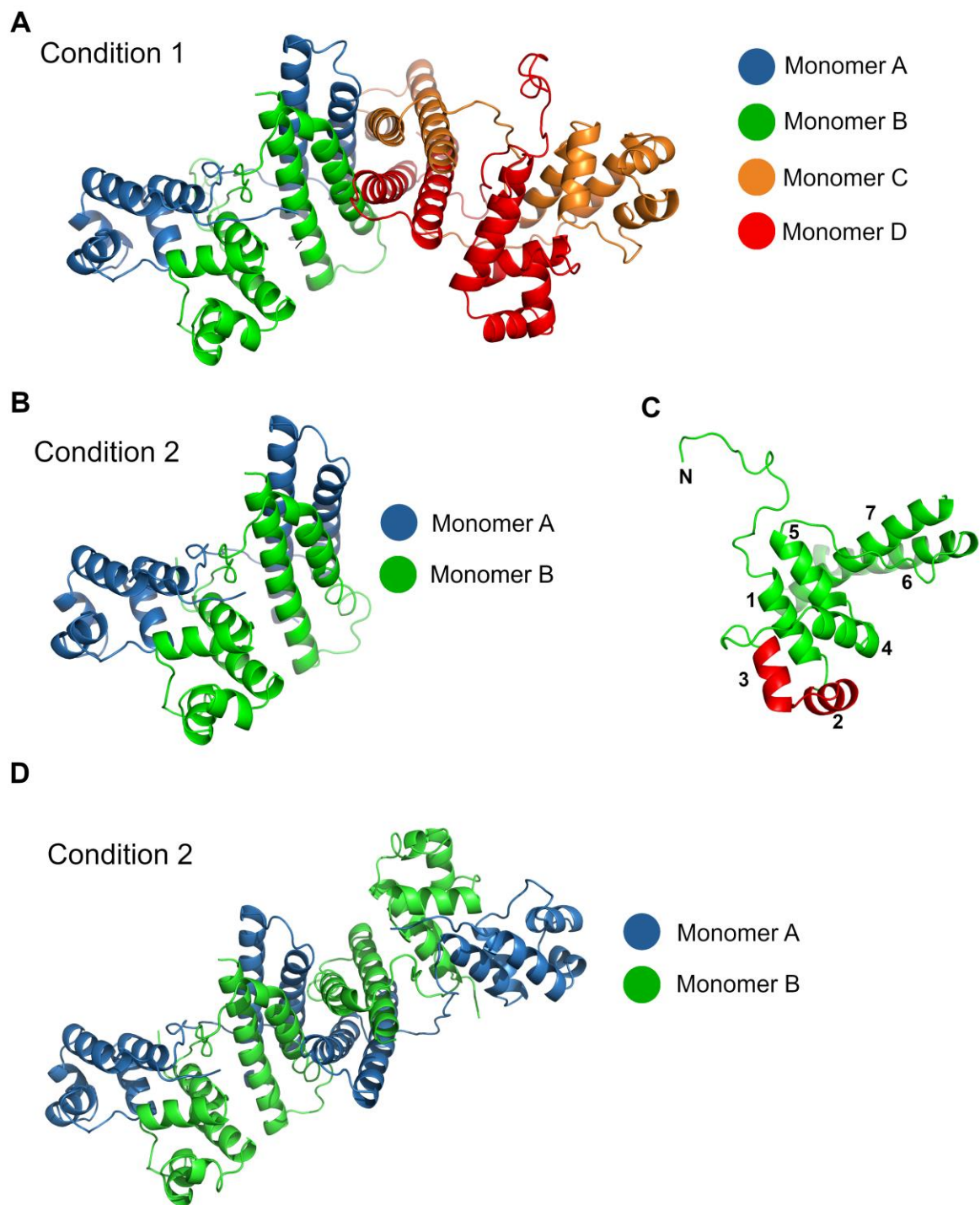


Figure 3 – Two crystal structures of 186 CII145. (A) Crystal structure of CII145 in condition 1 (8% Tascimate pH 8.0, 20% PEG 3350). (B) Crystal structure of CII145 in condition 2 (2 M ammonium sulfate, 0.1 M Bis-Tris: HCl, pH 5.5). (C) Ribbon representation of a subunit of CII with helix-turn-helix motif in red. The helix-turn-helix motif is predicted using the Dodd and Egan weight matrix (Dodd and Egan, 1990). (D) Tetrameric quaternary structure of CII145 (condition 2) proposed by PISA to be stable in solution (Krissinel and Henrick, 2007).

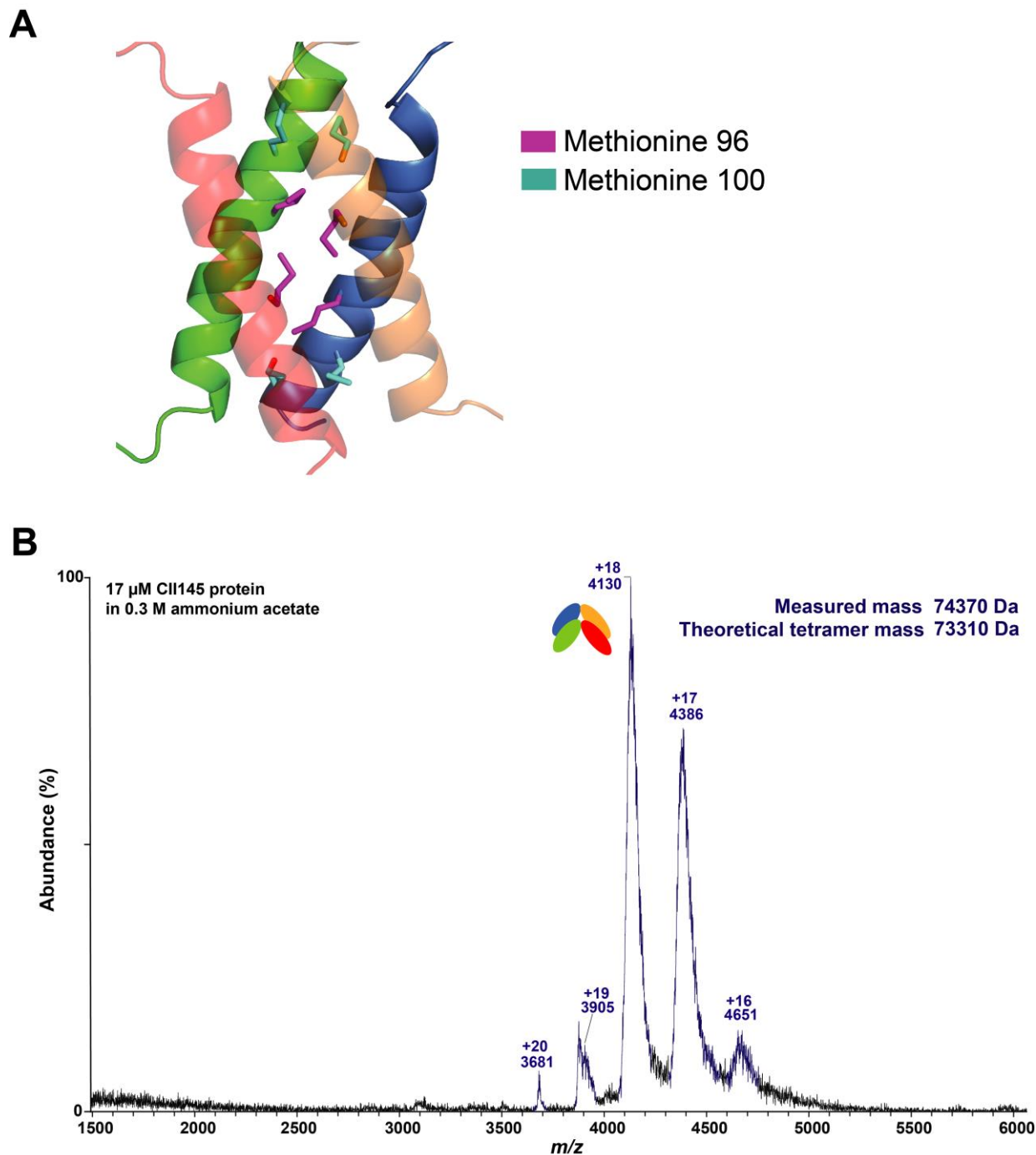


Figure 4 – CII145 forms a tetramer. A) Tetramerization interface in the CII145 (condition 1) crystal structure. CII145 tetramerization is mediated by a methionine zipper. Each monomer contributes two methionine, M96 and M100, which interact to form a four-helix bundle. B) Native ESI-MS of the CII145 protein (17 μ M) in 300 mM ammonium acetate detects CII145 forming a tetramer.

Docking suggests that CII binds to DNA as a tetramer

The CII145 crystal structure revealed a tetrameric protein, mediated by a methionine zipper interface. Previous studies have defined two DNA half sites that CII recognises at the p_E promoter (Shearwin and Egan, 2000). We hypothesized that the CII145 tetramer could be the DNA binding species *in vivo* that mediates transcriptional activation. To investigate this, HADDOCK (De Vries, Van Dijk and Bonvin, 2010) was used to dock the CII145 structure in condition 1 onto *in silico* models of p_E DNA bent at various angles, modelled using 3D-DART (van Dijk and Bonvin, 2009), with the optimal predicted complex shown in Figure 5. In this model, each constituent dimer in the CII tetramer supplies a single helix-turn-helix (HTH) motif that contacts a p_E half site. HTH motifs within a constituent dimer are too close to one another to contact half sites with the spacing observed in p_E , without extreme distortions to the DNA. To accommodate the HTH motifs, the DNA is bent at 40 degrees, consistent with the experimentally measured $41 \pm 5^\circ$ (Shearwin and Egan, 2000). The model shows CII helix 3, the recognition helix of the HTH motif, slotting into the major groove of the p_E binding site. This model is consistent with previous DNaseI footprinting data (Neufing *et al.*, 1996; Shearwin and Egan, 2000), and with data showing that mutagenesis of key HTH residues (CII169^{V36E/Q37S}) gives a protein which cannot bind or activate the p_E promoter (Murchland *et al.*, 2014). These residues are within the recognition helix of CII and likely involved in key DNA binding interactions.

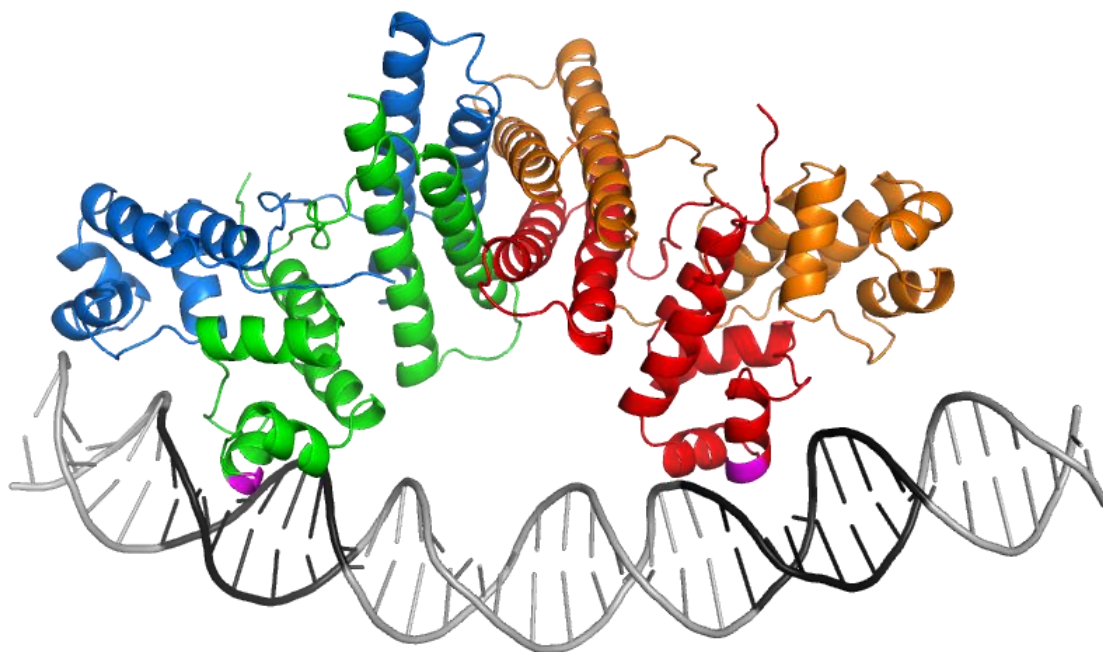


Figure 5 - Model of 186 CII binding to the 186 p_E promoter. CII tetramer structure (CII145 structure 1) docked onto p_E DNA modelled with 3D-DART (van Dijk and Bonvin, 2009). Docking was carried out on the HADDOCK webserver (De Vries, Van Dijk and Bonvin, 2010). One helix-turn-helix motif from each CII dimer contacts a p_E half site. DNA is shown in grey with CII half sites are coloured in black. Residues V36 and Q37 (shown in magenta) if mutated to glutamate and serine abolish DNA binding (Murchland *et al.*, 2014).

Native mass spectrometry was used to investigate DNA binding by CII *in vitro*. Similar approaches have successfully been employed to characterize protein-DNA interactions (Craig *et al.*, 1999; Satiaputra *et al.*, 2019). CII145 was mixed with annealed oligonucleotides containing the CII binding site. The minimum DNA fragment length required for binding is unknown, so two lengths of 31 bp and 35 bp were tested (subsequently referred to as pE31 and pE35) (Figure 6A). In each case, two DNA bound species could be detected, with one species apparently much more abundant than the other (Figure 6B and Figure 6C). The most abundant species corresponds to the mass of a CII145 tetramer bound to one single piece of DNA, consistent with the model of binding proposed in Figure 5. It should be noted that this result does not differentiate between a tetramer binding to one or two p_E sites. The low-abundance species corresponds to the mass of a CII tetramer bound to two separate DNA molecules. This binding may occur through one dimer within the tetramer interacting with one DNA fragment and the second dimer with a different DNA fragment. This mode of binding would be facilitated by the very short lengths of the DNA fragments used here. The caveat to judging abundance from mass spectra is that the ease of ionization between the species may be different and thus relative ratios cannot reliably be determined.

A CII tetramer binding to DNA may introduce significant strain into the DNA due to the 40° bend. The native conformation of pE DNA was modelled using the bend.it server (Vlahoviček, Kaján and Pongor, 2003). The predicted conformation showed a 21° bend in the DNA in the same direction as in the predicted CII-DNA complex. If pE DNA possessed an intrinsic bend that was complementary to the CII tetramer, it would maximise the probability of one CII tetramer occupying both half sites. The extra strain required to bend pE to 40° for CII to occupy both half sites may reduce this probability, and hence contribute to the observation of pE DNA bound by 2 CII tetramers..

To confirm that the CII binding seen here is specific, CII145 was mixed with a 35 bp DNA fragment from the FtsK gene (FtsK35), a sequence unrelated to the p_E promoter sequence. No binding to this non-specific DNA sequence was detected via mass spectrometry, with only free DNA and free CII145 detected (Figure 7).

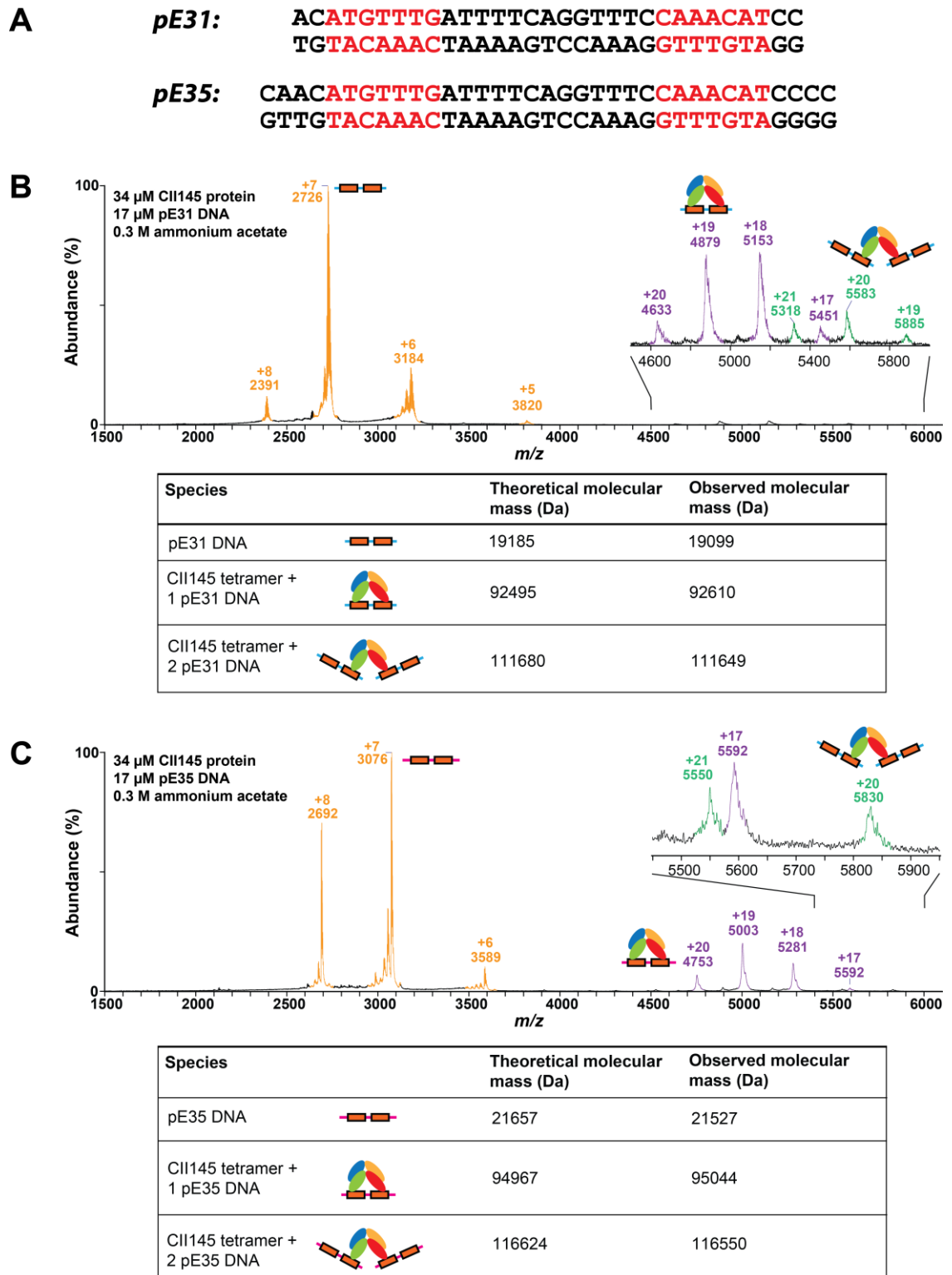
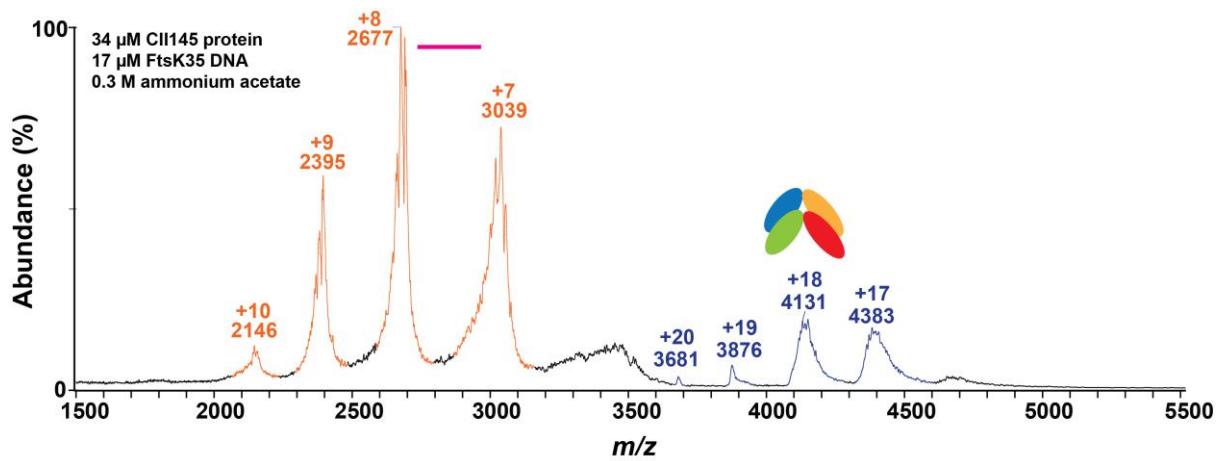


Figure 6 –Mass spectra of native CII145-DNA complexes. (A) The DNA sequences of the oligonucleotides used in the mass spectrometry experiments below with the CII operator sites highlighted in red. (B) ESI-MS spectra of CII145 (34 μ M) as a mixture with 31 bp p_E DNA (17 μ M). (C) ESI-MS spectra of CII145 (34 μ M) as a mixture with 35 bp p_E DNA (17 μ M). Two protein-DNA complexes were observed for both pE31 and pE35 DNA.





Species		Theoretical molecular mass (Da)	Observed molecular mass (Da)
FtsK35 DNA		21503	21478
CII145 tetramer		73310	74385

Figure 7 – No non-specific binding of CII to DNA could be detected via mass spectrometry. ESI-MS spectra of CII145 (34 μ M) as a mixture with 35 bp FtsK DNA (17 μ M). Free DNA and a CII145 tetramer were the only species detected.

***In vivo* proteolysis disrupts CII's ability to form tetramers**

Murchland *et al.* (2021) showed recently that CII has a sequence at the C-terminal region that recruits the RseP protease, resulting in cleavage of the protein at residue 135. The degradation product (referred to as CII135) is further degraded by the FtsH protease. Based on the CII145 structure, this cleavage point lies within helix 7. In the CII tetramer, the α -helical methionine zipper bundle that mediates tetramerization is surrounded by four helix 7, one supplied by each monomer. Thus, cleavage within helix 7 was hypothesized to disrupt the methionine zipper tetramerization interface of CII, preventing tetramerization. In reporter gene assays, CII135 was shown to display minimal transcriptional activation, although there was a statistical significance slightly above basal promoter levels (Murchland *et al.*, 2014). To investigate the structural basis of this post-translational regulation, the crystal structure of CII135 was determined.

The structure of CII135 was solved. The crystals have the symmetry of space group P3₂21. The asymmetric unit contained a dimer of CII135 (Figure 8A). A typical section of the electron density map for the CII135 structure is shown in Figure S 1. In the structure of CII135, electron density for residues after E86 was not observed. The truncation may result in residues after E86 no longer adopting a defined secondary structure and generating extremely flexible regions in the protein that cannot be seen in the crystal structure. To check that the protein had not become further truncated during overexpression or purification, the molecular mass of the CII135 protein used for crystallization was determined by mass spectrometry under denaturing conditions (Figure 9A). The observed mass of the denatured protein was only slightly smaller than the theoretical molecular mass of His-tagged CII135, suggesting that the initiator methionine is cleaved *in vivo*, but that no further degradation had occurred. Examination of the crystal packing of CII135 shows that the protein does not adopt the same

tetrameric oligomerization in the crystal lattice as observed for CII145. The next protein molecule in the crystal lattice is present in the region where the methionine-zipper helices would be expected (Figure 8B and C). This change in oligomeric state to a dimer was confirmed *in vitro* using native mass spectrometry (Figure 9B).

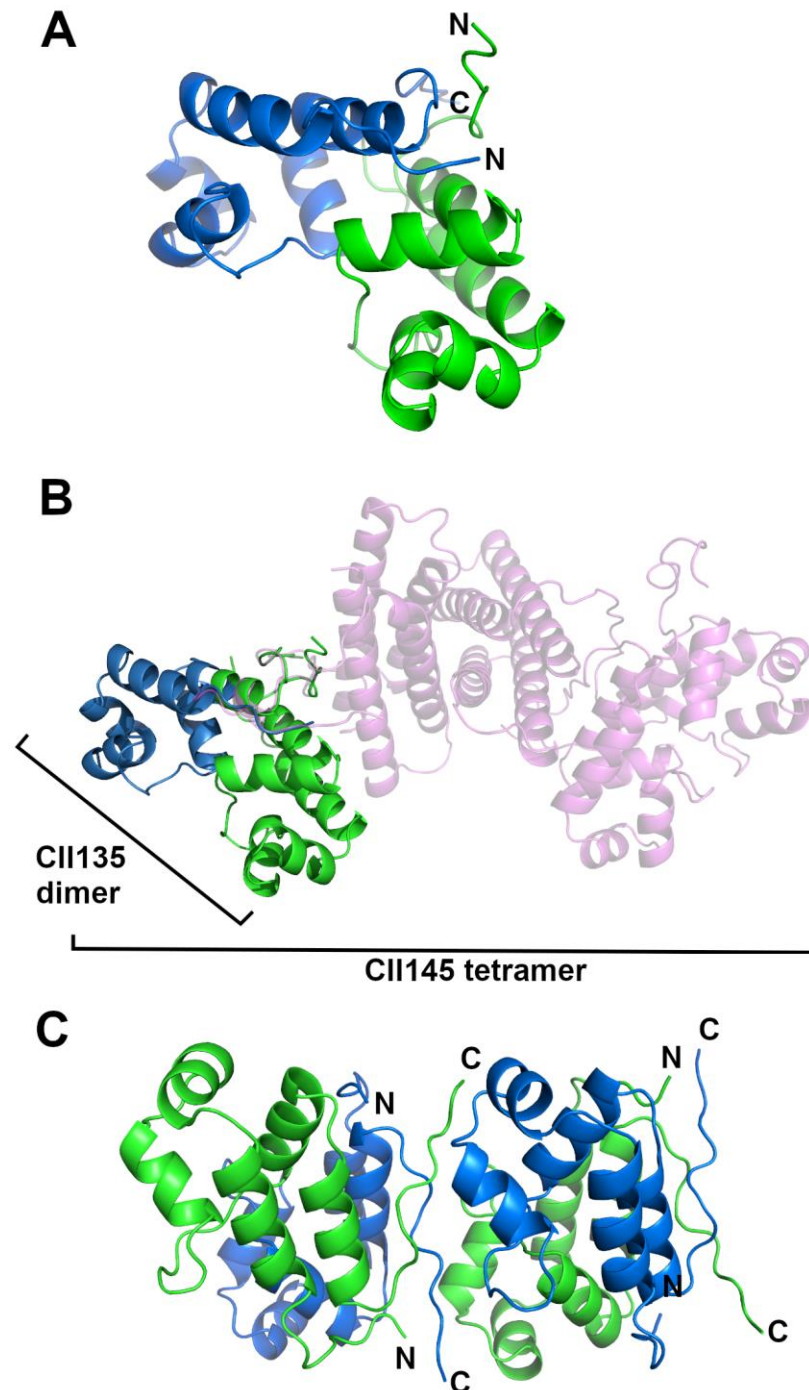


Figure 8 – Crystal structure of CII135. (A) Crystal structure of CII135 was solved as a dimer. Electron density for residues after E86 was not observed. Residues after this point could not be modelled into the structure. (B) Crystal structure of CII135 superimposed onto the structure of CII145 (condition 1). CII135 chains are shown in green and blue. CII145 structure is shown in magenta. (C) CII135 does not adopt the same tetrameric quaternary state as CII145 in the crystal lattice. The packing of CII135 in the lattice excludes the possibility of a CII tetramer. A CII135 dimer is found where an intact methionine zipper interface would be present.

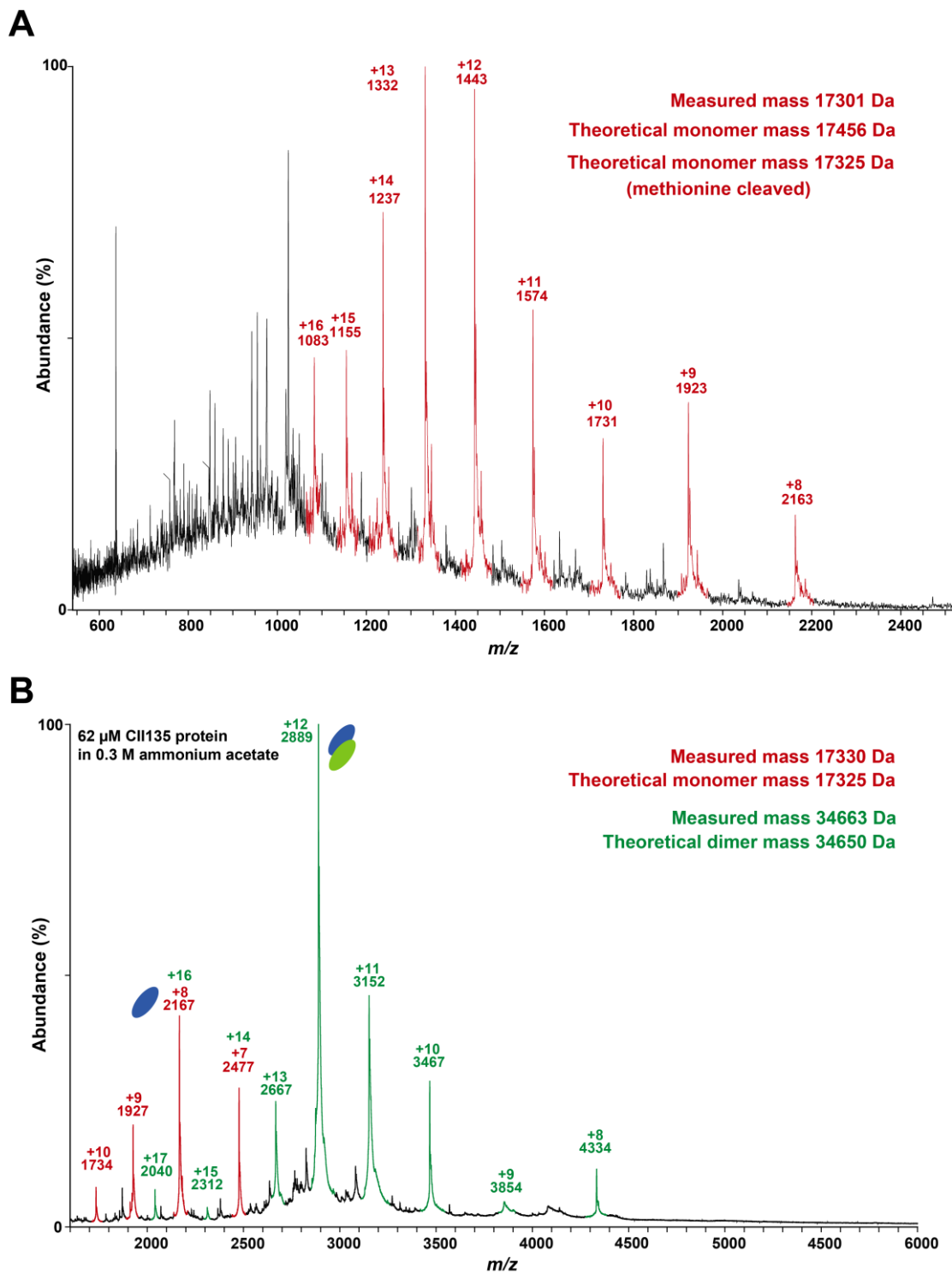


Figure 9 – CII135 forms dimers. (A) Mass spectra of denatured CII135 suggests that initiator methionine of CII is removed. The spectra also confirms that CII135 was not degraded prior to crystallization trials. Protein was denatured by mixing protein solution with a methanol/acetic acid solution in a 1:1 ratio. (B) Native ESI-MS spectrum of CII135 shows the presence of CII135 monomers and dimers. No CII135 tetramers were observed.

Superposition of the CII dimer units from all three CII crystal structures onto each other reveals that they adopt the same conformation (Figure 10A). As described earlier, residues after residue E86 were not observed in the crystal structure of CII135 and the methionine zipper oligomeric interface is disrupted in this degradation product. However, even in the absence of a functional methionine zipper segment, the remainder of the CII135 protein can adopt the same dimeric conformation. This result suggests the possibility of the NTD and/or the linker region play a role in CII dimerization. To investigate the oligomerization, possible residue-residue interactions between different monomers were computationally defined using the crystal structures. Interactions including ionic bonds, hydrogen bonds, aromatic interactions (such as π - π stacking) and hydrophobic interactions were observed throughout the structure and not isolated to just the CTD (Figure 10B). Thus, interactions in the NTD and linker region may assist in dimerization. There is a salt bridge between E14 in the NTD and H143 in the CTD of the other protein chain, which would significantly restrict the movements of one NTD relative to the tetramerization interface. This interaction, in addition to other intramolecular interactions between the NTDs and between the linker regions, could contribute to a CII dimer acting as a rigid body.

The crystal structure of CII135 contained the same two intramolecular disulfide bonds as the structure of CII145 in condition 2. As explained for the structure of CII145 in condition 2, these disulfide bonds are likely due to oxidation of the protein after purification.

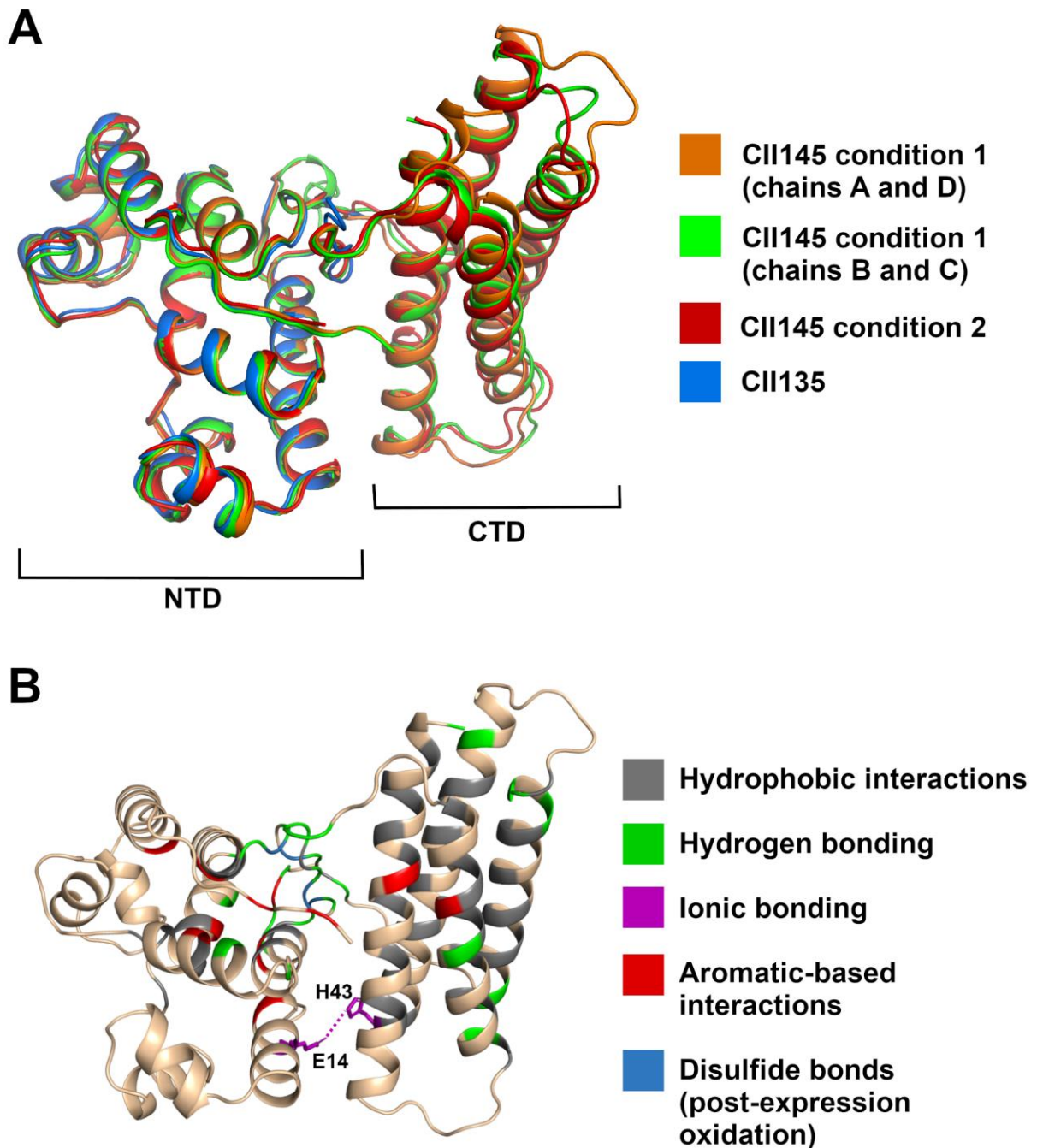


Figure 10 - A CII dimer adopts the same conformation in different crystal conditions. (A) The structures of the CII dimers found in CII145 condition 1, CII145 condition 2 and CII135 were superimposed onto each other. (B) Plausible intramolecular contacts, defined using the Protein Interaction Calculator (Tina, Bhadra and Srinivasan, 2007), exist throughout the protein dimer. In the dimer, there are intramolecular contacts between the two CTDs, the NTDs and between the linkers. There is also a salt bridge between an NTD residue and the CTD of the other monomer. Residues involved in forming disulfide bonds, which are hypothesized to be the result of oxidation of cysteine residues after protein purification, are also shown in blue.

Tetrameric CII is required for efficient DNA binding and transcriptional activation

CII145, shown here to adopt a tetrameric oligomeric state, is a functional truncation of CII that retains the same mode of activation as the wild-type protein. CII135 is the natural degradation

product of CII and displays minimal activity (Murchland *et al.*, 2014). Our experiments show the cleavage of CII at residue 135 allows dimerization but prevents subsequent oligomerization into a tetramer. An attempt to investigate DNA binding by CII135 via mass-spectrometry was made. However, the resulting spectra was too noisy to interpret confidently (see Figure S 3 for details). To examine if a change in oligomeric state affects the ability of the protein to activate p_E , mutagenesis was carried out to disrupt the tetrameric interface of CII. The key residues in the tetramerization interface were computationally predicted with KFC (Darnell, LeGault and Mitchell, 2008) and Robetta scanning alanine (Kim, Chivian and Baker, 2004) webserver. A combined total of six key residues were predicted- Q93, V96, M97, M100, V111, S112. Site directed mutagenesis was used to substitute these residues with alanine. The methionine residues were also mutated to cysteine.

As judged qualitatively by western blot, the mutants V96A, M97A, M100A, V111A and S112 are solubly expressed *in vivo*, at levels similar to the wild-type protein (Figure 11A). The activity of these mutants was first qualitatively assessed using phage complementation assays (Table 6). The V96A, M97A and M100A mutants of CII were unable to complement a CII-deficient phage. M97C and M100C could produce turbid plaques with visually reduced turbidity compared to wild-type CII. The inability of V96A, M97A and M100A mutants to complement a CII-deficient phage arises from their loss of transcriptional activation of p_E (Figure 11B). These three CII mutants showed vastly reduced levels of p_E activation (0-1.4% of wild-type CII). Two residues, M97 and M100 are proposed to be key residues in the putative methionine zipper interface depicted in Figure 4A. Mutating these residues to alanine likely prevents a tetramer formation. These results support the hypothesis that CII binds to DNA and activates p_E as a tetramer. All other mutants Q93A, M97C, M100C, V111A, and S112A display intermediate p_E activation activity between wild-type CII protein and the no CII control, retaining enough activity to form turbid plaques in complementation assays. The functional tetramerization interface is likely retained in these mutants, possibly with reduced affinity for self-association.

Table 6 – Phage complementation assays to measure the effect of tetramerisation interface mutants on CII activity in vivo. Mutagenesis of some residues within the putative tetramerization interface of 186 CII results in loss of transcriptional activation activity of the protein. CII-deficient 186 phage (cII473) was plated onto lawns of IM26 expressing either wild-type or variants of CII from a pZS45 plasmid. Bacteria were grown on LB agar media supplemented with 100 μ M IPTG to induce CII expression. Active CII would complement the CII-deficient phage and give turbid plaque morphology, or no plaques if all infections result in lysogeny. Inactive CII results in clear plaque morphology

CII variant expressed	Morphology of cII473 plaque
Wild-type	Turbid
no CII	Clear
Q93A	Turbid
V96A	Clear
M97A	Clear
M97C	Weakly turbid
M100A	Clear
M100C	Weakly turbid
V111A	Turbid
S112A	Turbid

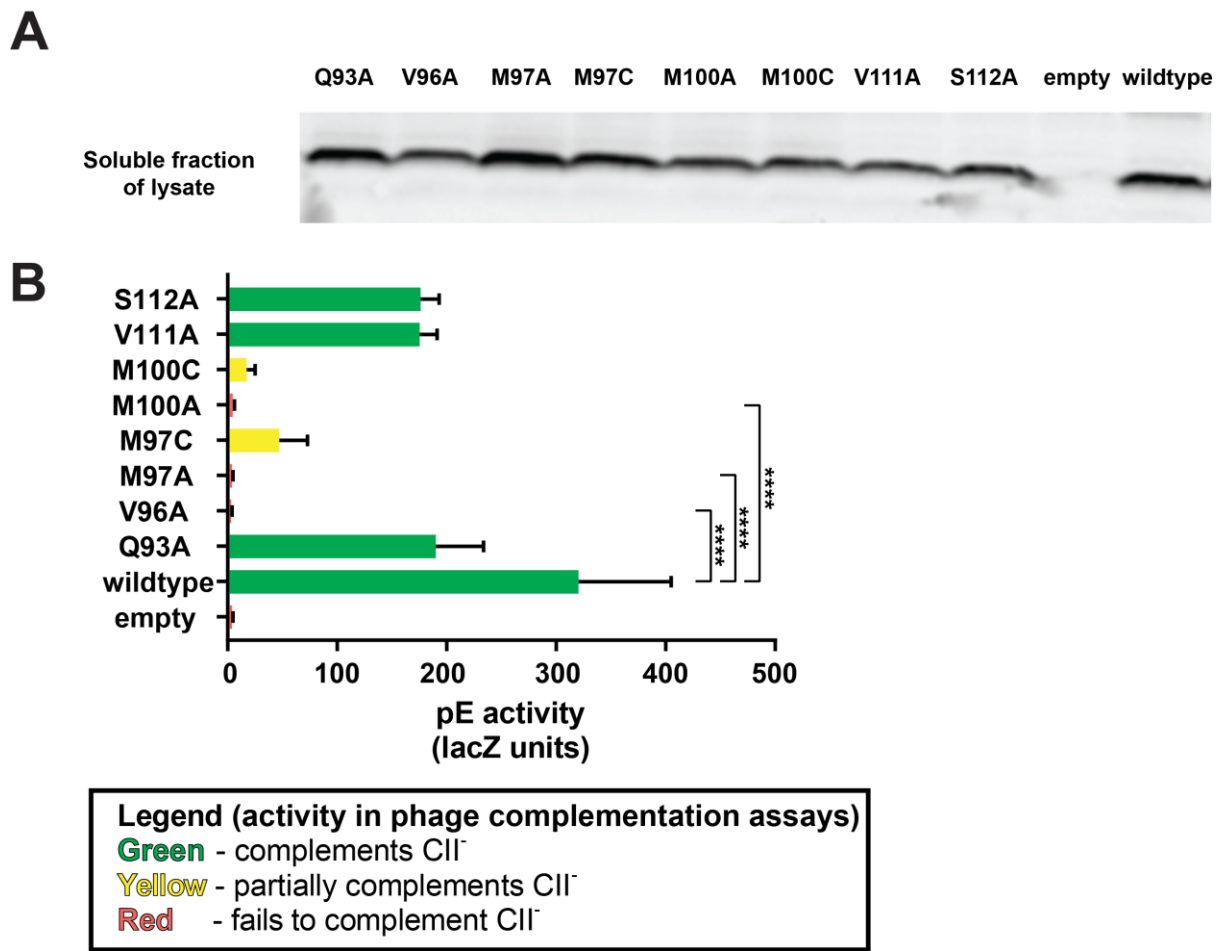


Figure 11 – Measuring p_E activation and DNA binding of CII with amino acid substitutions in the putative tetramerization interface *in vivo*. A) Mutants of CII are solubly expressed by p_E lacZ reporter strains. Western blot analysis of CII expression in p_E-lacZ reporter strains grown in media containing 100 μM IPTG, show qualitatively that the mutants of CII are expressed and that the protein is soluble. The soluble fraction of cell lysate was obtained by centrifugation of whole cell extract (21,130 g, 10 min) to pellet insoluble material. B) p_E-lacZ reporter assays reveal that mutagenesis of V96, M97 and M100 reduces the activation ability of the protein. Bar represents the mean. Error bars represent the 95% CI. n=9. **** denotes p > 0.001. Statistical significance was determined using a paired two-tail t-test.

CII interacts with the RNAP σ⁷⁰ subunit and α-CTD

A common mechanism employed by prokaryotic transcription factors to activate a promoter is to bind to DNA and "recruit" (RNAP) to the promoter by making direct protein-protein contacts (Lee, Minchin and Busby, 2012). This is the basic mechanism used by CII to activate p_E, however the unusual 20bp operator spacing and set of existing CII activation-deficient mutants suggest an unusual set of contacts may be involved (Murchland *et al.*, 2014).

To model how a CII tetramer bound at p_E may interact with RNAP to activate the promoter, the crystal structure of *E. coli* RNAP was modelled onto the CII-DNA docked complex from Figure 5 (Figure 12A). Murchland *et al.* (2014) have previously proposed that CII residues E46 and R17 interact with K593 of the *E. coli* σ⁷⁰ subunit, based on CII mutants isolated from genetic screens. The modelling presented here provides further evidence of an interaction between these three residues. In this docked complex, it is found that one of the CII E46

residues is spatially positioned such that it can interact with σ^{70} residue K593 (Figure 12B). R17 contacts E46 and may have a role in positioning it correctly for the K593 interaction.

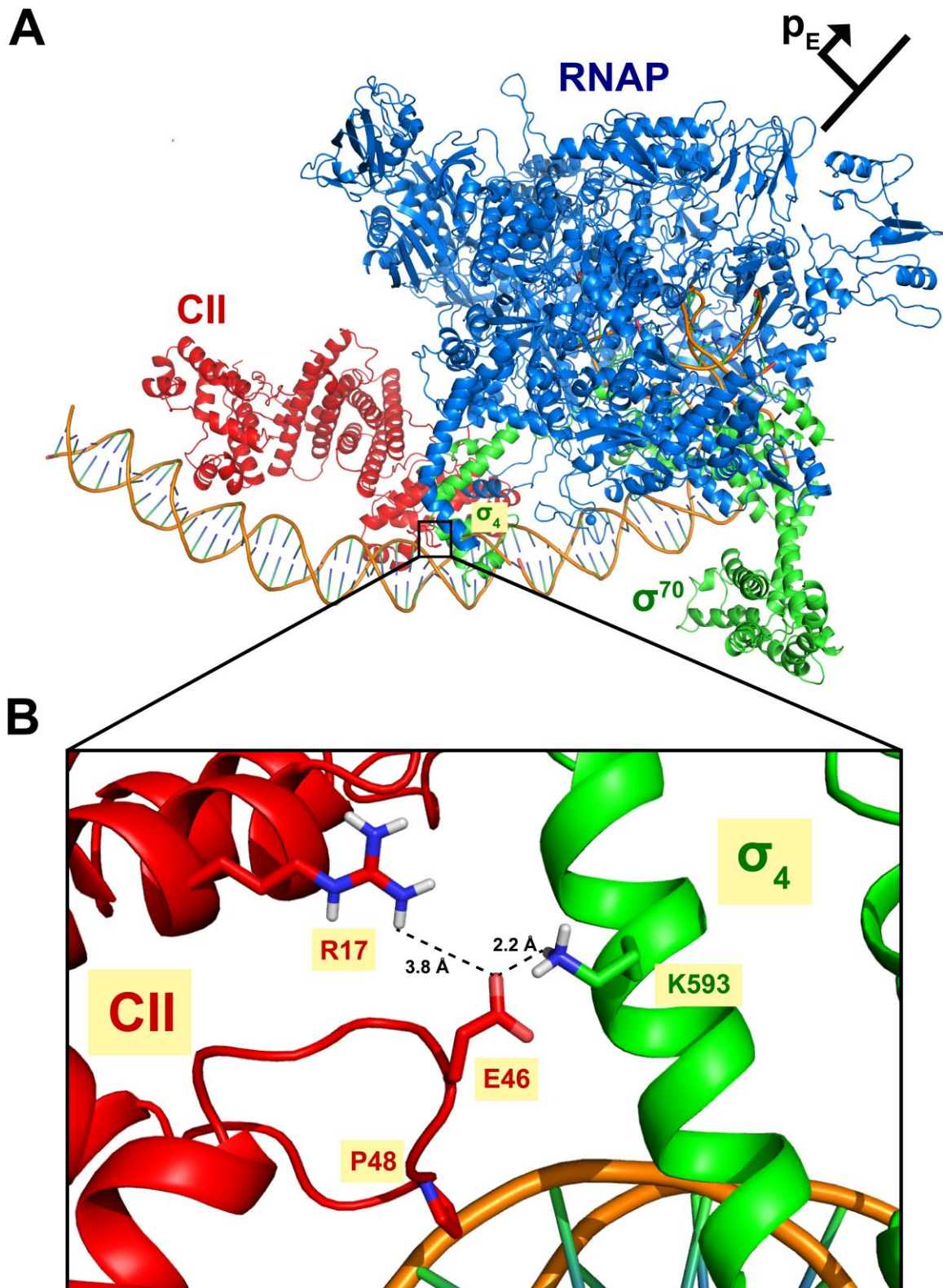


Figure 12 - Model of 186 CII binding and activation of the 186 p_E promoter. (A) Model of the arrangement of CII and RNAP at the 186 p_E promoter. RNAP model used is *E. coli* RNAP crystal structure (PDB ID 4YLN), positioned by placing σ^{70} over the -35 sequence of p_E. The σ^{70} subunit of RNAP is in green, and CII in red. (B) CII residues E46 and R17 are positioned in the model close enough to plausibly contact RNA polymerase σ^{70} K593.

To further validate the E46 contact, this residue was mutated to the other 19 natural amino acids and the transcriptional activation ability was assayed via lacZ reporter assays (Figure 13A). Mutating CII at this position generated mutants with varying levels of activity. An acidic residue in the 46th position, either glutamate (the wild-type residue) or aspartate is necessary for strong transcriptional activation. Interestingly, the aspartate mutation resulted in 13% higher activity than wild-type protein. The structural model indeed predicts that the E46D mutant would retain activity, as D46 should maintain the salt bridge with K593. Placing a non-acidic residue at position 46 resulted in a significant reduction in activation ability by over 65%. These mutants still retained significant activity, suggesting that E46-K593 is not the only stabilizing contact that CII provides to RNAP at the p_E promoter. These lower activity mutants, such as E46G which does not have a side chain to form extra interactions, may have lost the stabilizing interaction from the 46th residue but still provide RNAP recruitment through other contacts. The most significant loss of function occurred when E46 was mutated to a basic residue, arginine or lysine. A basic residue at this position is predicted to create an antagonistic interaction with the key σ^{70} residue. The effect of the E46K mutation on the protein complex was modelled using ICM-Pro (Molsoft). *In silico* energy minimization of the mutant and wild-type complexes show a 2 Å shift of K593 away from the CII 46K residue. In summary, the E46K mutation repels the RNAP σ_4 helix, opposing recruitment at p_E (Figure 13B).

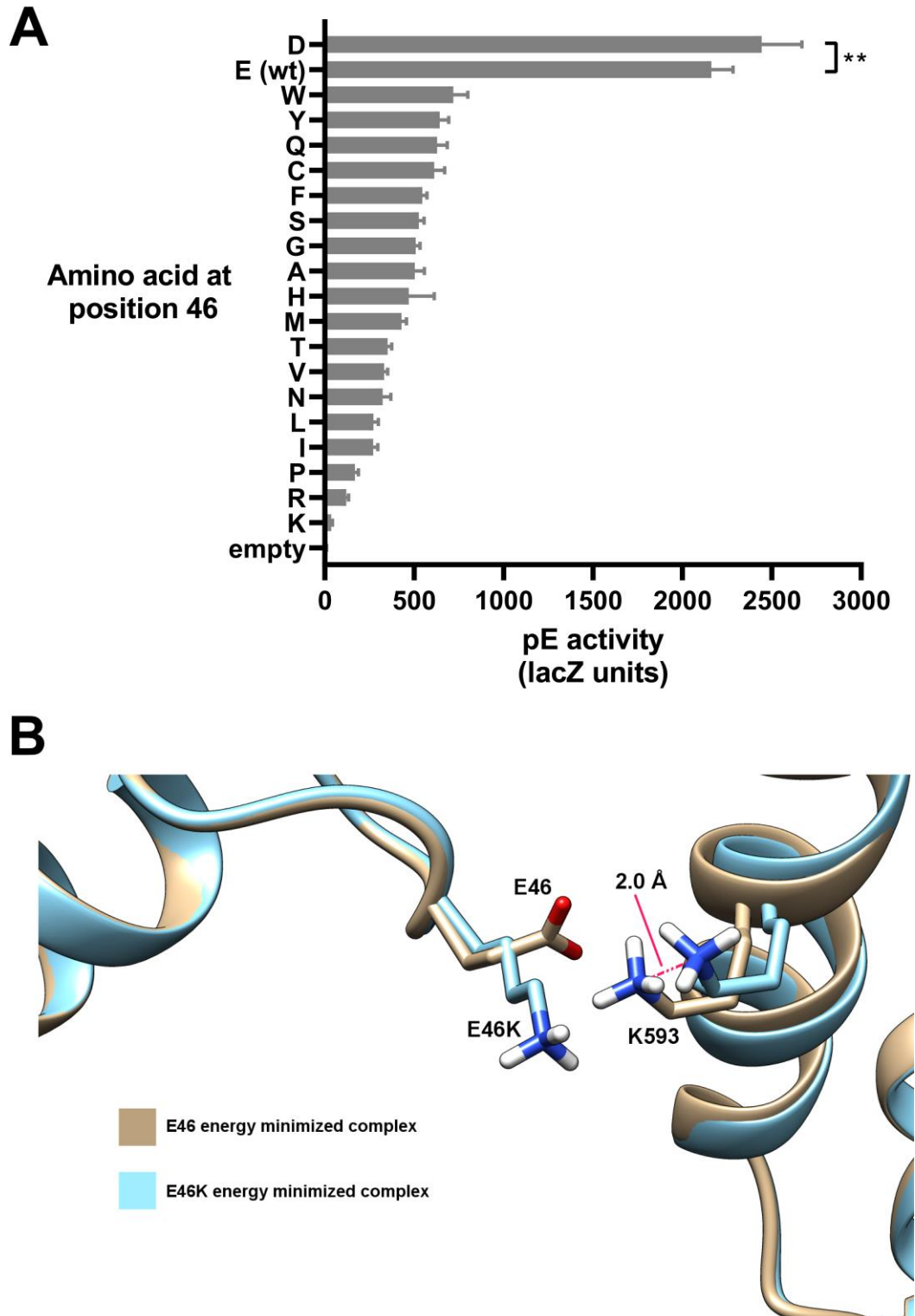


Figure 13 – CII recruits RNA polymerase via an activation epitope at E46. (A) CII requires an acidic residue at the 46th position for effective transcriptional activation. Basic residues in the 46th position resulted in maximal loss of function as measured by p_E-lacZ activity. CII expression was induced by growing cells in 200 μM IPTG. Error bars represent the 95% CI (n ≥ 8). ** denotes p > 0.01. (B) *In silico* simulations show the E46K mutation creates an antagonistic interaction that pushes K593 away by 2.0 Å. The energy minimization of the RNAP-DNA complexes with mutant and wild-type CII was carried out using ICM-Pro (Molsoft). The wild-type and E46K complex are depicted in tan and blue, respectively.

Even with the loss of an important contact between CII residue 46 and σ K593, a significant amount (~20%) of transcriptional activity remains, suggesting another contact between CII and RNAP. Combined with the observation that CII loss-of-transcriptional activity mutants map to different regions of CII (Murchland *et al.*, 2014), additional contacts with the alpha subunit of RNAP was proposed. The alanine scanning mutagenesis library constructed over the studies of Tang *et al.* (1994), Gaal *et al.* (1996), Kainz and Gourse (1998) and Savery *et al.* (1998), were used in this study to investigate possible contacts the α subunit may have with 186 CII, using the p_E lacZ reporter system, and expressing CII^{E46K} (Figure 14). The E46K mutant of CII was selected to measure the effect of α -CTD mutations, as σ^{70} dependent promoter activation masks the contribution of α -CTD to promoter activation in lacZ assays with wild-type CII (Murchland *et al.*, 2014).

The results show that alanine substitutions at residue positions 265, 268, 269, 295, 296, 298, 299 and 302 resulted in a 46-61% loss in activity. These residues lie within the well-known 265 DNA-binding determinant of α -CTD (Gaal *et al.*, 1996; Murakami, Fujita and Ishihama, 1996), indicating that DNA binding by α -CTD is important in p_E activation. Alanine substitutions at positions 262, 281, 289, 303 and 307 result in >40% loss in activity and have sidechains directed towards the core of the protein. These effects are likely to be indirect and not a result of loss of contacts with CII. Alanine substitutions at residues 262 and 303 in a previous screen have been shown to result in a moderate loss of activity and were proposed to have indirect effects on DNA binding (Ross *et al.*, 2003). Other mutations at residues 281, 289 and 303 may alter the position of α -CTD helix 2, which contains D280. As will be explained later, D280 is predicted to interact with CII. Alanine substitutions at positions with surface exposed side chains (residues 273, 280, 287, 291, 293, 314 and 320) also resulted in >40% loss of activity. A subset of these residues may be forming stabilizing contacts with CII to stabilize RNAP at the p_E promoter. Interestingly, alanine substitutions at position 264 and 271 gave an increase in p_E activity. V264 is positioned next to the main DNA binding residue R265 and could plausibly contact DNA. Thus, V264 mutations could have improved DNA binding. This improvement was not observed in other screens using this library and may be specific to the p_E promoter sequence (Gaal *et al.*, 1996; Savery *et al.*, 1998, 2002; Ross *et al.*, 2003). The K271A mutation is likely to be generating a new contact or removing an antagonistic interaction to CII.

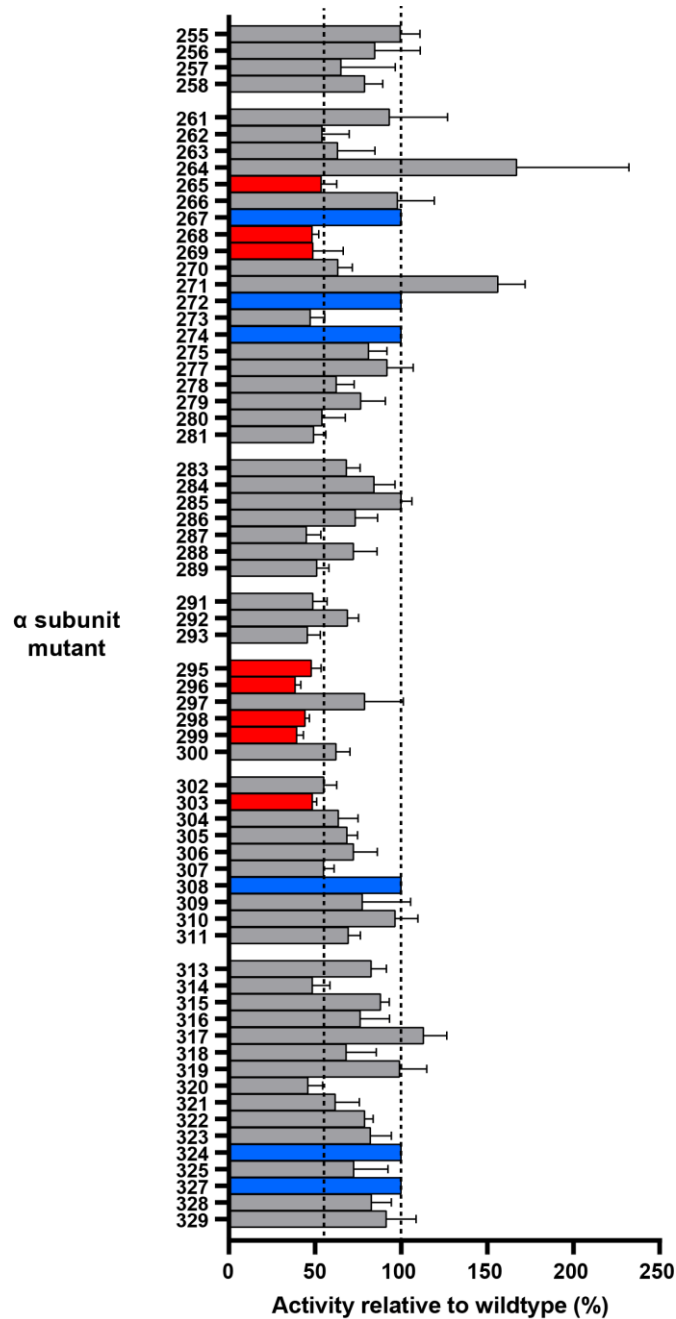


Figure 14 - Alanine scanning mutagenesis to identify RNA polymerase α -CTD residues involved in CII-dependent activation of p_E . (A) Activity of p_E was measured using an *E. coli in vivo* p_E lacZ reporter system expressing CII^{E46K} (IM50) and RNA polymerase α subunit mutants where each residue was replaced with alanine, encoded by pHTf1 and pREII α plasmids. Mutants 255-271 were expressed from the pHTf1 vector system. Mutants 272-329 were expressed from the pREII α vector system. All mutant activities are normalized to the wild-type activity of their respective plasmid system. The activities presented are the average of four independent experiments. The error bars represent the 95% confidence interval (CI). A gap in the mutant axis indicates a residue untested because of poor growth of that strain. CII and RNA polymerase α subunit expression was induced by growing cells in the presence of 200 μ M IPTG. Blue bars indicate where alanine naturally occurs (same as wild-type α). The red bars correspond to residues within the DNA-binding determinant (Gaal *et al.*, 1996; Murakami, Fujita and Ishihama, 1996)

The structural model of CII activation of the p_E promoter was extended to include the α -CTD, with the positioning of the subunit driven by the α -CTD mutagenesis data and known DNA binding contacts of the α -CTD. The α -CTD model was derived from a crystal structure of a catabolite activator protein (CAP), α -CTD and DNA complex (Benoff *et al.*, 2002) (PDB ID 1LB2). The α -CTD was positioned such that DNA binding determinants, most importantly R265, were contacting the A-T rich minor groove of the DNA, between the p_E half sites. R115 is an epitope used by CII to contact α -CTD (Murchland *et al.*, 2014). Different α -CTD positions were tested to evaluate possible contacts with CII R115, with the final model shown in Figure 15A. Due to the rotational symmetry of the CII-DNA complex, CII can contact two α -CTDs, with each dimer providing a R115 contact epitope.

In the proposed model, α -CTD D280 forms a salt bridge with R115 of CII. Thus, mutation of α -CTD D280 to alanine (this work) or CII R115E (Murchland *et al.*, 2014) abolishes this salt bridge and reduces CII activity. In the model, α -CTD K271 also forms a repulsive interaction with K46 of the CII^{E46K} mutant (Figure 15B). This prediction is consistent with the K271A mutant of α providing increased p_E activation in lacZ assays, as the CTD mutant would abolish this repulsive interaction. Remarkably, K271 is predicted to form a stabilising salt bridge with the crucial E46 residue in wild-type CII. The E46 residue from one dimer of a CII tetramer is involved in the interaction with σ^{70} K593 and may be unable to interact with α -CTD. However, the CII dimer subunit distal to the σ^{70} has a free E46 to interact with α -CTD.

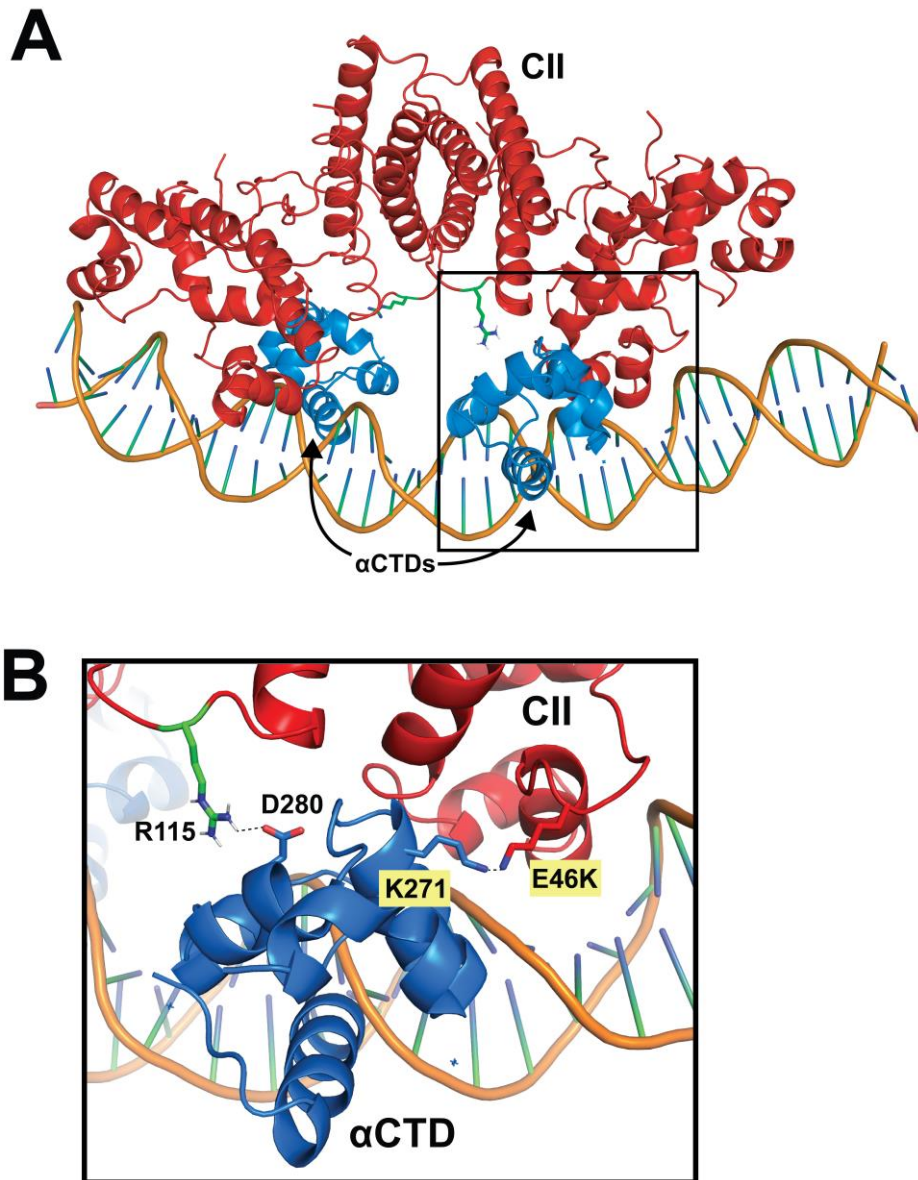


Figure 15 – Proposed structural model of 186 CII anchoring α -CTD at the p_E promoter. (A) α -CTD is positioned within a binding site formed between 186 CII and the bent p_E DNA. The *E. coli* α -CTD of RNAP (PDB ID 1LB2) was modelled into the CII-DNA complex using PyMOL (DeLano, 2002). The R115 epitope of CII that contacts α -CTD is shown in green. Region of interest where α -CTD interacts with CII, denoted in a black box, is magnified in (B). The model proposes that CII R115 contacts D280 of the α -subunit. The K46 residue in the CII^{E46K} mutant used in alanine scanning lacZ assays is proposed to form an electrostatic repulsive interaction with K271, which would be abolished in the K271A mutant of α . In the wild-type situation α 271 is proposed to form a stabilizing electrostatic interaction with CII E46.

Discussion

Rationale for structurally characterized CII truncations

The 186 CII protein is 169 residues long containing two functional domains – an N-terminal DNA binding domain and a C-terminal oligomerization domain. The C-terminal region of 186 CII contains a degradation signal that leads to CII being degraded rapidly *in vivo* to give a specific degradation product, CII135. Rapid degradation presents technical difficulties in overexpressing wild-type CII, both in terms of obtaining sufficient quantities for structural studies, and in purifying it away from its degradation product. Murchland *et al.* (2014) demonstrated that CII145 functionally activates the p_E promoter in the same manner as the wild-type protein. In lacZ reporter assays from the current study (Figure 2), CII145 was shown to be one of the most active truncations and thus potentially most amenable to crystallization. The CII145 truncation was used throughout this study as it was straightforward to purify, and its structure should be reflective of the full-length protein. The degradation product CII135 was also structurally characterized to understand the structural changes that occur as a result of site-specific degradation.

Currently, no structural data on CII residues after residue 145 is available. A secondary structure prediction suggests an α -helix spanning residues 151-160 (Murchland *et al.*, 2014). Sequential single amino acid truncations back to residue 145 increased *in vivo* activity. Thus the structural elements after residue 145 are likely to be involved in the degradation tag recognition by proteases. These structural elements would be shortened or disrupted as the C-terminus is progressively truncated. Attempts to crystallize full-length CII were unsuccessful, as were attempts to co-crystallise CII with DNA. Future studies could employ alternative techniques such as NMR to characterize this C-terminal sequence in isolation.

The active species of 186 CII is a tetramer

186 CII forms a dimer of dimers. The crystal structures suggest that 186 CII initially dimerizes into a rigid body. This dimerization is mediated by residues in both domains of the protein and can occur in the absence of a structured CTD. The presence of the salt bridge in the protein between the NTD and CTD indicates the domains may be structurally dependent and could explain how the NTD and CTD fail to overexpress, each expressed on their own (Murchland *et al.*, 2014). Half of a methionine zipper interface is created in the dimeric protein to mediate subsequent dimerization into a tetramer (Figure 4A). This order of assembly from monomer to dimer to tetramer was previously shown for CII (Shearwin and Egan, 2000) and is almost universally the pathway that homotetrameric proteins use to assemble (Powers and Powers, 2003). A tetrameric CII oligomeric state has previously been observed via analytical ultracentrifugation (Shearwin and Egan, 2000) and SEC-MALS (Murchland *et al.*, 2014). In this study, native mass-spectrometry further validates that CII forms a tetrameric species.

In the tetramer, an HTH motif from each dimer is positioned to bind to an inverted repeat DNA half site, spaced two turns of the DNA apart (Figure 6). Computational modelling of DNA showed that a DNA bend of 40° gave the optimal spatial arrangement for DNA binding by a CII145 tetramer (Figure 5). This angle is consistent with the experimentally determined CII induced DNA bend of $41^\circ \pm 5$ (Shearwin and Egan, 2000). The rigidity of the dimers within the tetramer places physical constraints on the positions of the HTH motif, explaining the intolerance of CII to changes in half site spacing.

Native mass-spectrometry analysis of CII145 with p_E DNA identified a complex of a tetramer of CII with DNA as the predominant species (Figure 6) consistent with the complex predicted by protein-DNA docking (Figure 5). To investigate the role of tetramerization in CII function, residues predicted to mediate tetramerization were mutated. Mutagenesis of residues in the methionine zipper (dimer-dimer) interface abolished the transcriptional activation ability of CII, and the ability to complement a CII deficient phage, indicating that the tetramer is needed to function effectively as an activator.

λ CII, like 186 CII, also exists as a dimer of dimers. The unique arrangements of binding sites in λ CII and 186 CII (Figure 1) are the result of the positioning of the λ CII and 186 CII dimers in their tetramers. Despite both λ CII and 186 CII binding DNA through a helix-turn-helix motif, the overall topology of the proteins differs (Figure 16B). λ CII multimerizes into a dimer of dimers, to form an asymmetric tetramer bundle mediated by a single helix (helix 4) from each monomer. The dimers are arranged such that the HTH motifs that contact the major groove of DNA face the same direction, resulting in λ CII binding direct DNA repeats (Jain *et al.*, 2005) (Figure 16A). 186 CII forms a tetrameric species, mediated by two helices (helices 6 and 7) from each monomer. In this species, the HTH motifs are pointed in opposite directions, resulting in the protein binding inverted repeats (Figure 5).

A commonality between the mechanism of λ CII and 186 CII is that they act through the stabilization of both the σ^{70} subunit and α -CTD of RNAP. 186 CII makes direct contacts to both σ^{70} (CII E46 forms an ionic interaction with σ^{70} K593) and α -CTD (CII E46 and R115 is predicted to form ionic interactions with α residues K271 and D280, respectively). λ CII, on the other hand, binds to the opposite side of DNA relative to RNAP and only contacts an α -CTD subunit of RNAP (Figure 16C). λ CII is predicted to contact the same residue (α K271) as 186 CII. α -CTD helps to stabilize the σ^{70} subunit (Figure 16D). Thus, λ CII indirectly exerts its effect on the σ^{70} subunit. 186 CII presents an alternative solution to efficient activation of a promoter.

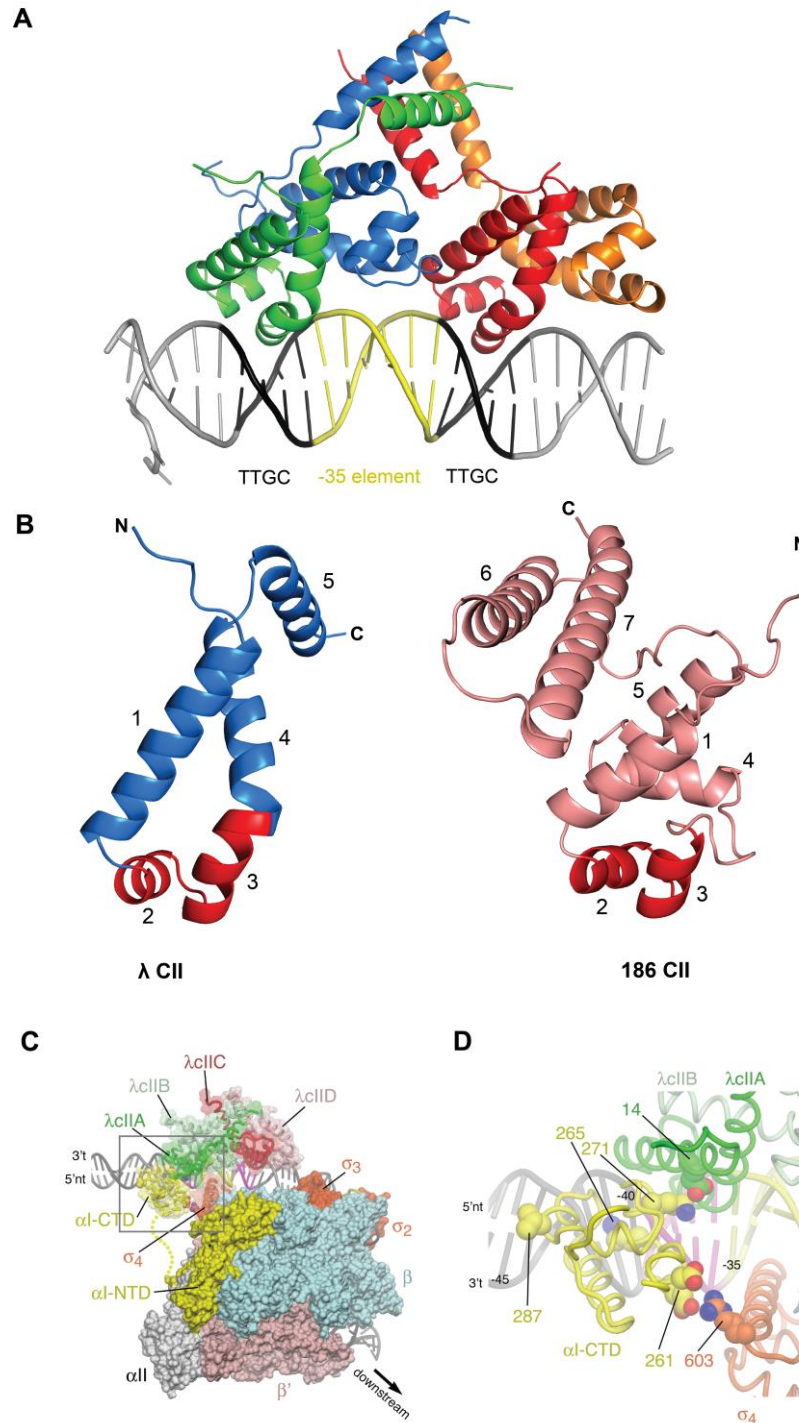


Figure 16 – Mechanism that λ CII uses to bind DNA and recruit RNA polymerase. (A) Structure of λ CII tetramer bound to DNA (PDB ID 1ZS4) (Jain *et al.*, 2005). λ CII binds to TTGC direct repeat half site that flank the -35 promoter element. (B) Structure of a monomer of λ CII (PDB ID 1ZS4; chain A), shown in blue, and a monomer of 186 CII (condition 1; chain A), shown in pink. The helix-turn-helix (HTH) motifs in the proteins are shown in red and are positioned in the same orientation. (C) Structural model of RNA polymerase recruitment by λ CII. λ CII binds DNA and contacts the α -CTD subunit of RNA polymerase on the opposite face of the DNA to recruit it to the promoter. (D) Magnified region from (C) showing a predicted contact between RNA polymerase λ CII residue E14 and α residue K271. α -CTD contacts σ_4 via its 261 determinant (D259 and E261). Models shown in (C) and (D) were reproduced from Jain *et al.* (2005).

A tetrameric activator may present useful regulatory features to the lytic-lysogenic decision

The role of the CII proteins in bacteriophage λ and 186 is to influence the decision between the lytic and lysogenic cycles. If the effective concentration of the activator protein does not reach a threshold concentration, the phage commits to the lytic cycle. In this context, CII providing an ultrasensitive response may be beneficial, where at high CII levels, it decisively begins the lysogenic cascade. Evidence from our laboratory indicate that 186 CII can provide an ultrasensitive response (Crooks, 2006; Murchland *et al.*, 2014). A tetrameric arrangement may provide increased ultrasensitivity compared to a dimer. The degree of ultrasensitivity due to cooperative binding is less than or equal to the stoichiometry of the reaction. This property is described by Hill coefficient, n_H , which have maximal values of 2 and 4 for a dimer and tetramer, respectively.

The stability of 186 CII and λ CII have important roles during the lytic-lysogenic decision stage of infection. Both proteins have a short *in vivo* half-life of around 2 minutes (Shotland *et al.*, 2000; Murchland *et al.*, 2014). λ CII is quickly degraded by the host protease FtsH (HflB) (Banuett *et al.*, 1986; Shotland *et al.*, 2000). It has been hypothesized that FtsH, which is upregulated during times of increased nutrition, allows λ CII to sense nutrient levels and skew development toward the lytic development in times of nutrient abundance (Bandyopadhyay *et al.*, 2011). Mutations in FtsH which increase λ CII half-life, have been shown to promote lysogeny (Herman *et al.*, 1993).

186 CII, like λ CII, is also regulated at the proteolysis level. It is degraded by a site-specific protease RseP recruited by a degradation signal present at the C-terminus of the protein (Murchland *et al.*, 2014; 2020), and also by FtsH. The rapid clearance of CII is required for the lytic-lysogenic switch to work efficiently. A modified 186 phage where the wild-type CII was replaced with the stabilized CII145 always entered lysogeny after initial infection. The ability of this prophage reenter into the lytic cycle (prophage induction) after UV irradiation was also severely impaired. Stable CII145, in initial infection and prophage induction, will build up to levels to drive enough CI expression to force the phage almost permanently into lysogeny (Murchland *et al.* 2020). The structural data explains how proteolysis inactivates the CII protein.

In the crystal structure of the 186 CII135 degradation product, the CTD helices that form the tetramerization interface is not observed and are likely to be unstructured. It however can still dimerize through contacts found in the NTD and linker region. Truncating the protein at residue 135 disrupts this interface, preventing the CII135 dimers from forming the active tetramer. This is consistent with SEC-MALS data showing CII135 forms a dimer (Murchland *et al.*, 2014). Any truncation of CII past residue 141 results in a reduction in activation activity by over 90%, likely due to the inability to form the methionine zipper helix for tetramerization.

It is possible that the degradation product itself plays a minor role in shutting down CII activity. If CII135 exists as a monomer, it may be able to sequester a CII169 monomer into an inactive dimer. This would further reduce levels of active CII in the cell.

Site-specific proteolysis of 186 CII provides a rapid way for the phage to inactivate CII. Similar to λ CII, it is likely the proteases involved are regulated by environmental signals and that protease activity is an input in the lytic/lysogenic decision. Bacteriophage λ produces a

protease inhibitor CIII that inhibits the FtsH protease involved in degradation of λ CII (Kobiler, Rokney and Oppenheim, 2007), whereas 186 does not appear to have a CIII equivalent. Future studies could investigate whether 186 has any direct regulatory effects on the proteases that degrade 186 CII.

186 CII presents a novel mechanism of contacting RNA polymerase

In this study, experimentally determined structures of 186 CII and the RNAP holoenzyme, including the σ^{70} and α CTD subunits was used generate a structural model of CII activation of the p_E promoter (Figure 12 and Figure 15). Computational modelling and docking were driven by data from mutagenesis studies. Bacterial activators that activate σ^{70} dependent promoters often do so via direct contact with RNAP as it engages with the promoter. The two major targets of this activation are domain 4 of the σ^{70} subunit and CTD of the α subunit of RNAP (Lee, Minchin and Busby, 2012).

Class II activators tend to bind around the -41 element and activate a promoter by recruitment of RNAP through direct contacts with domain 4 of σ^{70} subunit (Browning and Busby, 2004). This contact is the primary mode of activation of p_E by 186 CII. CII uses the E46 epitope to contact σ^{70} K593 on the face of the α -helix adjacent to the interface between domain 4 of σ^{70} and the p_E -35 element. The crystal structure of CII shows that R17 also lies in this epitope and plays a role in positioning E46 to make contacts (Figure 12B). Interactions with this face of the α -helix are conserved in many other class II activators including λ cI in the activation of λ p_{RM} promoter, bacteriophage Mu Mor protein and AraC (Lonetto *et al.*, 1998; Dove, Darst and Hochschild, 2003; Lee, Minchin and Busby, 2012).

Class I transcriptional activators bind upstream of the -10 and -35 sites and achieve activation through a direct contact to the α -CTD of RNAP (Browning and Busby, 2004). α -CTD binds to promoter UP elements and class I activators to stabilize this interaction. Examples of class I transcriptional activators include the *E. coli* catabolite activator protein (CAP) (Busby and Ebright, 1999) and λ CII (Jain *et al.*, 2005). 186 CII, in addition to making contacts with domain 4 of σ^{70} , also contacts the α -CTD using its R115 epitope (Murchland *et al.*, 2014). In this study, we proposed that the α -CTD binds the p_E promoter around the -40 and/or -56 position between the p_E half sites, with CII residue R115 contacting D280 residue on the α -CTD. Another contact may exist between K271 and E46. It would be difficult to measure the contribution of E46 to class I transcriptional activation *in vivo*, as it is also a key epitope in contacting the σ^{70} subunit. The unusual spacing of p_E half sites of two turns of the DNA helix apart may be selected to allow CII to contact both α -CTDs simultaneously. α -CTD contacts promoter UP elements which tend to be AT rich (Estrem *et al.*, 1998), which is the case for the DNA between the p_E half sites. CII binding to this DNA also generates a DNA bend (Shearwin and Egan, 2000) and it has been noted that the curve conformation of DNA upstream of a promoter may affect its interactions with RNAP α -CTD (Gaal *et al.*, 1994; Katayama *et al.*, 1999; Asayama and Ohyama, 2007). α -CTD binds the minor groove of DNA (Gourse, Ross and Gaal, 2000), of which there are two between the p_E half sites on the same face that RNAP polymerase binds. Above each minor groove, there is an accessible R115 activation epitope. Thus, CII is likely to contact both α -CTD of RNAP simultaneously.

How does the structure of CII beget its function as an effective decision point in the lytic lysogenic cascade? A single CII tetramer upon binding the p_E promoter could potentially simultaneously stabilize three subunits of RNAP, the σ^{70} subunit and both α -CTDs, to provide

potent activation of the p_E promoter. To stabilize the same three subunits in a class I and class II transcriptional activator, such as CAP, would require two activators binding to two tandem binding sites (Barnard, Wolfe and Busby, 2004). The use of a tetramer as the active species could potentially provide ultrasensitivity, to concisely commit to lysogeny once the threshold CII concentration has been reached. Proteolytic degradation of CII disrupts the tetramerization interface of CII, providing 186 a means to rapidly remove CII and the ability to default to the lytic cycle or reenter the lytic cycle, in the case of prophage induction.

Acknowledgments

This work was supported by funding from the Australian Research Council (DP150103009 and DP160101450). We gratefully acknowledge Prof. Richard Gourse for providing the RNAP alpha-CTD alanine scanning library. We also acknowledge University of Adelaide Biochemistry III practical classes from 2014-2019 for preliminary experiments that have supported this work.

References

- Afonine, P. V. *et al.* (2012) 'Towards automated crystallographic structure refinement with phenix.refine', *Acta Crystallographica Section D: Biological Crystallography*, 68(4), pp. 352–367. doi: 10.1107/S0907444912001308.
- Arndt, U. W., Crowther, R. A. and Mallett, J. F. W. (1968) 'A computer-linked cathode-ray tube microdensitometer for X-ray crystallography', *Journal of Physics E: Scientific Instruments*, 1(5), pp. 510–516. doi: 10.1088/0022-3735/1/5/303.
- Asayama, M. and Ohyama, T. (2007) 'Curved DNA and Prokaryotic Promoters', in *DNA Conformation and Transcription*, pp. 37–51. doi: 10.1007/0-387-29148-2_3.
- Bandyopadhyay, K. *et al.* (2011) 'Studies on Escherichia coli HflKC suggest the presence of an unidentified factor that influences the lysis-lysogeny switch', *BMC Microbiology*, 11. doi: 10.1186/1471-2180-11-34.
- Banuet, F. *et al.* (1986) 'hflB, a new Escherichia coli locus regulating lysogeny and the level of bacteriophage lambda cII protein', *Journal of Molecular Biology*, 187(2), pp. 213–224. doi: 10.1016/0022-2836(86)90229-9.
- Barnard, A., Wolfe, A. and Busby, S. (2004) 'Regulation at complex bacterial promoters: How bacteria use different promoter organizations to produce different regulatory outcomes', *Current Opinion in Microbiology*, 7(2), pp. 102–108. doi: 10.1016/j.mib.2004.02.011.
- Battye, T. G. G. *et al.* (2011) 'iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM', *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp. 271–281. doi: 10.1107/S0907444910048675.
- Benoff, B. *et al.* (2002) 'Structural basis of transcription activation: The CAP- α CTD-DNA complex', *Science*, 297(5586), pp. 1562–1566. doi: 10.1126/science.1076376.
- Browning, D. F. and Busby, S. J. W. (2004) 'The regulation of bacterial transcription initiation', *Nature Reviews Microbiology*, 2(1), pp. 57–65. doi: 10.1038/nrmicro787.
- Brünger, A. T. (1992) 'Free R value: A novel statistical quantity for assessing the accuracy of crystal structures', *Nature*, 355(6359), pp. 472–475. doi: 10.1038/355472a0.
- Brünger, A. T. *et al.* (1998) 'Crystallography & NMR system: A new software suite for macromolecular structure determination', *Acta Crystallographica Section D: Biological Crystallography*, 54(5), pp. 905–921. doi: 10.1107/S0907444998003254.
- Busby, S. and Ebright, R. H. (1999) 'Transcription activation by catabolite activator protein (CAP)', *Journal of Molecular Biology*, 293(2), pp. 199–213. doi: 10.1006/jmbi.1999.3161.
- Campbell, E. A. *et al.* (2002) 'Structure of the bacterial RNA polymerase promoter specificity σ subunit', *Molecular Cell*, 9(3), pp. 527–539. doi: 10.1016/S1097-2765(02)00470-7.
- Chen, V. B. *et al.* (2010) 'MolProbity: All-atom structure validation for macromolecular crystallography', *Acta Crystallographica Section D: Biological Crystallography*, 66(1), pp. 12–21. doi: 10.1107/S0907444909042073.
- Collaborative Computational Project Number 4 (1994) 'The CCP4 suite: Programs for protein crystallography', *Acta Crystallographica Section D: Biological Crystallography*, 50(5), pp. 760–763. doi: 10.1107/S0907444994003112.
- Cowtan, K. (2006) 'The Buccaneer software for automated model building. 1. Tracing protein

- chains', *Acta Crystallographica Section D: Biological Crystallography*, 62(9), pp. 1002–1011. doi: 10.1107/S0907444906022116.
- Craig, T. A. *et al.* (1999) 'Analysis of transcription complexes and effects of ligands by microelectrospray ionization mass spectrometry', *Nature Biotechnology*, 17(12), pp. 1214–1218. doi: 10.1038/70767.
- Crooks, M. (2006) *Transcriptional Interference between the Switch Promoters of 186*. University of Adelaide.
- Darnell, S. J., LeGault, L. and Mitchell, J. C. (2008) 'KFC Server: interactive forecasting of protein interaction hot spots.', *Nucleic acids research*, 36(Issue suppl_2), pp. W265–W269. doi: 10.1093/nar/gkn346.
- Datsenko, K. A. and Wanner, B. L. (2000) 'One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products', *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), pp. 6640–6645. doi: 10.1073/pnas.120163297.
- DeLano, W. L. (2002) 'The PyMOL molecular graphics system'.
- Diederichs, K. and Karplus, P. A. (1997) 'Improved R-factors for diffraction data analysis in macromolecular crystallography', *Nature Structural Biology*, 4(4), pp. 269–275. doi: 10.1038/nsb0497-269.
- van Dijk, M. and Bonvin, A. M. J. J. (2009) '3D-DART: A DNA structure modelling server', *Nucleic Acids Research*, 37(SUPPL. 2). doi: 10.1093/nar/gkp287.
- van Dijk, M. and Bonvin, A. M. J. J. (2010) 'Pushing the limits of what is achievable in protein-DNA docking: Benchmarking HADDOCK's performance', *Nucleic Acids Research*, 38(17), pp. 5634–5647. doi: 10.1093/nar/gkq222.
- Dodd, I. B. *et al.* (2001) 'Octamerization of λ CI repressor is needed for effective repression of PRM and efficient switching from lysogeny', *Genes and Development*, 15(22), pp. 3013–3022. doi: 10.1101/gad.937301.
- Dodd, I. B. and Egan, J. B. (1990) 'Improved detection of helix-turn-helix DNA-binding motifs in protein sequences', *Nucleic Acids Research*, 18(17), pp. 5019–5026. doi: 10.1093/nar/18.17.5019.
- Dove, S. L., Darst, S. A. and Hochschild, A. (2003) 'Region 4 of σ as a target for transcription regulation', *Molecular Microbiology*, 48(4), pp. 863–874. doi: 10.1046/j.1365-2958.2003.03467.x.
- Emsley, P. *et al.* (2010) 'Features and development of Coot', *Acta Crystallographica Section D: Biological Crystallography*, 66(4), pp. 486–501. doi: 10.1107/S0907444910007493.
- Estrem, S. T. *et al.* (1998) 'Identification of an UP element consensus sequence for bacterial promoters', *Proceedings of the National Academy of Sciences of the United States of America*, 95(17), pp. 9761–9766. doi: 10.1073/pnas.95.17.9761.
- Evans, G. and Bricogne, G. (2002) 'Triiodide derivatization and combinatorial counter-ion replacement: Two methods for enhancing phasing signal using laboratory Cu K α X-ray equipment', *Acta Crystallographica Section D: Biological Crystallography*, 58(6 II), pp. 976–991. doi: 10.1107/S0907444902005486.
- Evans, P. R. and Murshudov, G. N. (2013) 'How good are my data and what is the resolution?'

- Acta Crystallographica Section D: Biological Crystallography*, 69(7), pp. 1204–1214. doi: 10.1107/S0907444913000061.
- Gaal, T. *et al.* (1994) 'Localization of the intrinsically bent DNA region upstream of the E.coli *rrnB* P1 promoter', *Nucleic Acids Research*, 22(12), pp. 2344–2350. doi: 10.1093/nar/22.12.2344.
- Gaal, T. *et al.* (1996) 'DNA-binding determinants of the α subunit of RNA polymerase: Novel DNA-binding domain architecture', *Genes and Development*, 10(1), pp. 16–26. doi: 10.1101/gad.10.1.16.
- Gourse, R. L., Ross, W. and Gaal, T. (2000) 'UPs and downs in bacterial transcription initiation: The role of the alpha subunit of RNA polymerase in promoter recognition', *Molecular Microbiology*, pp. 687–695. doi: 10.1046/j.1365-2958.2000.01972.x.
- Herman, C. *et al.* (1993) 'Cell growth and λ phage development controlled by the same essential Escherichia coli gene, *ftsH/hflB*', *Proceedings of the National Academy of Sciences of the United States of America*, 90(22), pp. 10861–10865. doi: 10.1073/pnas.90.22.10861.
- Hernández, H. and Robinson, C. V. (2007) 'Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry', *Nature Protocols*, 2(3), pp. 715–726. doi: 10.1038/nprot.2007.73.
- Herskowitz, I. and Hagen, D. (1980) 'The Lysis-Lysogeny Decision of Phage lambda: Explicit Programming and Responsiveness', *Annual Review of Genetics*, 14(1), pp. 399–445. doi: 10.1146/annurev.ge.14.120180.002151.
- Huddleston, V. (1970) *Coliphage 186 control of the lysogenic response*. University of Adelaide, Adelaide, Australia.
- Jain, D. *et al.* (2005) 'Crystal structure of bacteriophage λ CI and its DNA complex', *Molecular Cell*, 19(2), pp. 259–269. doi: 10.1016/j.molcel.2005.06.006.
- Kabsch, W. (2010) 'XDS', *Acta Crystallographica Section D Biological Crystallography*, 66(2), pp. 125–132. doi: 10.1107/S0907444909047337.
- Kainz, M. and Gourse, R. L. (1998) 'The C-terminal domain of the alpha subunit of Escherichia coli RNA polymerase is required for efficient rho-dependent transcription termination', *Journal of Molecular Biology*, 284(5), pp. 1379–1390. doi: 10.1006/jmbi.1998.2272.
- Karplus, P. A. and Diederichs, K. (2012) 'Linking crystallographic model and data quality', *Science*, 336(6084), pp. 1030–1033. doi: 10.1126/science.1218231.
- Katayama, S. *et al.* (1999) 'Promoter upstream bent DNA activates the transcription of the *Clostridium perfringens* phospholipase C gene in a low temperature-dependent manner', *EMBO Journal*, 18(12), pp. 3442–3450. doi: 10.1093/emboj/18.12.3442.
- Kim, D. E., Chivian, D. and Baker, D. (2004) 'Protein structure prediction and analysis using the Robetta server', *Nucleic Acids Research*, 32(WEB SERVER ISS.), pp. 526–531. doi: 10.1093/nar/gkh468.
- Kobiler, O., Rokney, A. and Oppenheim, A. B. (2007) 'Phage lambda CIII: A protease inhibitor regulating the lysis-lysogeny decision', *PLoS ONE*, 2(4). doi: 10.1371/journal.pone.0000363.
- Krissinel, E. and Henrick, K. (2007) 'Inference of Macromolecular Assemblies from Crystalline State', *Journal of Molecular Biology*, 372(3), pp. 774–797. doi: 10.1016/j.jmb.2007.05.022.

- De La Fortelle, E. and Bricogne, G. (1997) 'Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods', in *Methods in Enzymology*, pp. 472–494. doi: 10.1016/S0076-6879(97)76073-7.
- Lamont, I. *et al.* (1993) 'Genes for the establishment and maintenance of lysogeny by the temperate coliphage 186.', *Journal of bacteriology*, 175(16), pp. 5286–5288.
- Lee, D. J., Minchin, S. D. and Busby, S. J. W. (2012) 'Activating transcription in bacteria', *Annual Review of Microbiology*, 66(1), pp. 125–152. doi: 10.1146/annurev-micro-092611-150012.
- Lonetto, M. A. *et al.* (1998) 'Identification of a contact site for different transcription activators in region 4 of the Escherichia coli RNA polymerase σ 70 subunit', *Journal of Molecular Biology*, 284(5), pp. 1353–1365. doi: 10.1006/jmbi.1998.2268.
- Lutz, R. and Bujard, H. (1997) 'Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements', *Nucleic Acids Research*, 25(6), pp. 1203–1210. doi: 10.1093/nar/25.6.1203.
- McCoy, A. J. *et al.* (2007) 'Phaser crystallographic software', *Journal of Applied Crystallography*, 40(4), pp. 658–674. doi: 10.1107/S0021889807021206.
- Murakami, K., Fujita, N. and Ishihama, A. (1996) 'Transcription factor recognition surface on the RNA polymerase α subunit is involved in contact with the DNA enhancer element', *EMBO Journal*, 15(16), pp. 4358–4367. doi: 10.1002/j.1460-2075.1996.tb00809.x.
- Murchland, I. *et al.* (2014) 'Promoter activation by CII, a potent transcriptional activator from bacteriophage 186', *Journal of Biological Chemistry*, 289(46), pp. 32094–32108.
- Murchland, I. *et al.* (2021) 'Instability of CII is needed for efficient switching between lytic and lysogenic development in bacteriophage 186', *Nucleic Acids Research*, 48(21), pp. 12030–12041. doi: 10.1093/nar/gkaa1065.
- Murshudov, G. N. *et al.* (2011) 'REFMAC5 for the refinement of macromolecular crystal structures', *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp. 355–367. doi: 10.1107/S0907444911001314.
- Neufing, P. J. *et al.* (1996) 'The CII protein of bacteriophage 186 establishes lysogeny by activating a promoter upstream of the lysogenic promoter', *Molecular microbiology*, 21(4), pp. 751–761.
- Neufing, P. J., Shearwin, K. E. and Egan, J. B. (2001) 'Establishing Lysogenic Transcription in the Temperate Coliphage 186', *Journal of Bacteriology*, 183(7), pp. 2376–2379. doi: 10.1128/JB.183.7.2376-2379.2001.
- Oppenheim, A. B. *et al.* (2005) 'Switches in bacteriophage lambda development', *Annual Review of Genetics*, 39(1), pp. 409–429. doi: 10.1146/annurev.genet.39.073003.113656.
- Panjikar, S. *et al.* (2005) 'Auto-Rickshaw: An automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment', *Acta Crystallographica Section D: Biological Crystallography*, 61(4), pp. 449–457. doi: 10.1107/S0907444905001307.
- Panjikar, S. *et al.* (2009) 'On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination', *Acta Crystallographica Section D: Biological Crystallography*, 65(10), pp. 1089–1097. doi:

10.1107/S0907444909029643.

Pannu, N. S. and Read, R. J. (2004) 'The application of multivariate statistical techniques improves single-wavelength anomalous diffraction phasing', *Acta Crystallographica Section D: Biological Crystallography*, 60(1), pp. 22–27. doi: 10.1107/S0907444903020808.

Pettersen, E. F. *et al.* (2004) 'UCSF Chimera—a visualization system for exploratory research and analysis', *Journal of computational chemistry*, 25(13), pp. 1605–1612.

Ponstingl, H., Henrick, K. and Thornton, J. M. (2000) 'Discriminating between homodimeric and monomeric proteins in the crystalline state', *Proteins: Structure, Function, and Genetics*. doi: 10.1002/1097-0134(20001001)41:1<47::aid-prot80>3.3.co;2-#.

Powers, E. T. and Powers, D. L. (2003) 'A Perspective on Mechanisms of Protein Tetramer Formation', *Biophysical Journal*, 85(6), pp. 3587–3599. doi: 10.1016/S0006-3495(03)74777-8.

Ptashne, M. (1986) 'A genetic switch: Gene control and phage. lambda', *Blackwell Scientific Publications*.

Ross, W. *et al.* (2003) 'An intersubunit contact stimulating transcription initiation by E. coli RNA polymerase: Interaction of the α C-terminal domain and σ region 4', *Genes and Development*, 17(10), pp. 1293–1307. doi: 10.1101/gad.1079403.

Satiaputra, J. *et al.* (2019) 'Native mass spectrometry identifies an alternative DNA-binding pathway for BirA from *Staphylococcus aureus*', *Scientific Reports*, 9(1). doi: 10.1038/s41598-019-39398-6.

Savery, N. J. *et al.* (1998) 'Transcription activation at class II CRP-dependent promoters: Identification of determinants in the C-terminal domain of the RNA polymerase α subunit', *EMBO Journal*, 17(12), pp. 3439–3447. doi: 10.1093/emboj/17.12.3439.

Savery, N. J. *et al.* (2002) 'Determinants of the C-terminal domain of the *Escherichia coli* RNA polymerase α subunit important for transcription at class I cyclic AMP receptor protein-dependent promoters', *Journal of Bacteriology*, 184(8), pp. 2273–2280. doi: 10.1128/JB.184.8.2273-2280.2002.

Schneider, T. R. and Sheldrick, G. M. (2002) 'Substructure solution with SHELXD', *Acta Crystallographica Section D: Biological Crystallography*, 58(10 I), pp. 1772–1779. doi: 10.1107/S0907444902011678.

Shearwin, K. E. and Egan, J. B. (2000) 'Establishment of Lysogeny in Bacteriophage 186 DNA BINDING AND TRANSCRIPTIONAL ACTIVATION BY THE CII PROTEIN', *Journal of Biological Chemistry*, 275(37), pp. 29113–29122.

Shotland, Y. *et al.* (2000) 'Proteolysis of bacteriophage λ CII by *Escherichia coli* FtsH (HflB)', *Journal of Bacteriology*, 182(11), pp. 3111–3116. doi: 10.1128/JB.182.11.3111-3116.2000.

Tang, H. *et al.* (1994) 'Location, structure, and function of the target of a transcriptional activator protein', *Genes and Development*, 8(24), pp. 3058–3067. doi: 10.1101/gad.8.24.3058.

Terwilliger, T. (2004) 'SOLVE and RESOLVE: Automated structure solution, density modification, and model building', *Journal of Synchrotron Radiation*, 11(1), pp. 49–52. doi: 10.1107/S0909049503023938.

Tina, K. G., Bhadra, R. and Srinivasan, N. (2007) 'PIC: Protein Interactions Calculator', *Nucleic*

Acids Research, 35(SUPPL.2). doi: 10.1093/nar/gkm423.

Trusina, A. *et al.* (2005) 'Functional alignment of regulatory networks: a study of temperate phages', *PLoS Computational Biology*, 1(7), p. e74.

Vagin, A. and Teplyakov, A. (1997) 'MOLREP: An Automated Program for Molecular Replacement', *Journal of Applied Crystallography*, 30(6), pp. 1022–1025. doi: 10.1107/S0021889897006766.

Vlahoviček, K., Kaján, L. and Pongor, S. (2003) 'DNA analysis servers: Plot.it, bend.it, model.it and IS', *Nucleic Acids Research*, 31(13), pp. 3686–3687. doi: 10.1093/nar/gkg559.

De Vries, S. J., Van Dijk, M. and Bonvin, A. M. J. J. (2010) 'The HADDOCK web server for data-driven biomolecular docking', *Nature Protocols*, 5(5), pp. 883–897. doi: 10.1038/nprot.2010.32.

Zuo, Y. and Steitz, T. A. (2015) 'Crystal structures of the e.coli transcription initiation complexes with a complete bubble', *Molecular Cell*, 58(3), pp. 534–540. doi: 10.1016/j.molcel.2015.03.010.

Supplementary materials

Table S 1 - Oligonucleotides used in this study

Name	Sequence
1241	GTAATGACCTCAGAACTCCATCTGG
1242	GCTTGGATTCTACCAATAAAAAACG
1756 CII-146-stop	TCATTGTCGGCTCTGGCGCTGCATGCACGTCTGtgaACTAATCCCGCTATGTCTCGA G
1757 CII-146-rev	CAGACGTGCATGCAGCGCCAGAGCCGACAATGAC
1758 CII-147-stop	TCGGCTCTGGCGCTGCATGCACGTctgCAGtgaAATCCCGCTATGTCTCGAGCGTG
1759 CII-147-rev	CTGcagACGTGCATGCAGCGCCAGAGCCG
1760 CII-148-stop	GCTCTGGCGCTGCATGCACGTCTGCAGACTtgaCCCGCTATGTCTCGAGCGTGGTC
1761 CII-148-rev	AGTCTGCAGACGTGCATGCAGCGCCAGAGC
1762 CII-149-stop	CTGGCGCTGCATGCACGTCTGCAGACTAATtgaGCTATGTCTCGAGCGTGGTCCG
1763 CII-149-rev	ATTAGTCTGCAGACGTGCATGCAGCGCCAG
1764 CII-150-stop	GCGCTGCATGCACGTCTGCAGACTAATCCctgaATGTCTCGAGCGTGGTTCGATAC
1765 CII-150-rev	GGGATTAGTCTGCAGACGTGCATGCAGCGC
1766 CII-151-stop	CTGCATGCACGTCTGCAGACTAATCCCGCTtgaTCGAGCGTGGTTCGATACCATG
1767 CII-151-rev	AGCGGGATTAGTCTGCAGACGTGCATGCAG
1768 CII-152-stop	CATGCACGTCTGCAGACTAATCCCGCTATGtgaAGCGTGGTTCGATACCATGAG
1769 CII-152-rev	CATAGCGGGATTAGTCTGCAGACGTGCATG
1770 CII-153-stop	GCACGTCTGCAGACTAATCCCGCTATGTCTGtgaGTGGTTCGATACCATGAGCGG
1771 CII-153-rev	CGACATAGCGGGATTAGTCTGCAGACGTGC
1772 CII-154-stop	CGTCTGCAGACTAATCCCGCTATGTCTGAGctgaGTCGATACCATGAGCGGTATT G
1773 CII-154-rev	GCTCGACATAGCGGGATTAGTCTGCAGACG
1774 CII-155-stop	CTGCAGACTAATCCCGCTATGTCTGAGCGTtgaGATACCATGAGCGGTATTGG
1775 CII-155-rev	CACGCTCGACATAGCGGGATTAGTCTGCAG
1776 CII-156-stop	CAGACTAATCCCGCTATGTCTGAGCGTGGTctgaACCATGAGCGGTATTGGCGC

1777 CII-156- rev	GACCACGCTCGACATAGCGGGATTAGTCTG
1778 CII-157- stop	ACTAATCCCGCTATGTCGAGCGTGGTCGATtgaATGAGCGGTATTGGCGCATC
1779 CII-157- rev	ATCGACCACGCTCGACATAGCGGGATTAGTC
1780 CII-158- stop	AATCCCGCTATGTCGAGCGTGGTCGATACCtgaAGCGGTATTGGCGCATCGTTT G
1781 CII-158- rev	GGTATCGACCACGCTCGACATAGCGGGATTAG
1782 CII-159- stop	CCCGCTATGTCGAGCGTGGTCGATACCATGtgaGGTATTGGCGCATCGTTTGG
1783 CII-159- rev	CATGGTATCGACCACGCTCGACATAGCGGG
1784 CII-160- stop	GCTATGTCGAGCGTGGTCGATACCATGAGCtgaATTGGCGCATCGTTTGGTCT
1785 CII-160- rev	GCTCATGGTATCGACCACGCTCGACATAGC
1786 CII-161- stop	ATGTCGAGCGTGGTCGATACCATGAGCGGTtgaGGCGCATCGTTTGGTCTGAT
1787 CII-161- rev	ACCGCTCATGGTATCGACCACGCTCGACATAG
1788 CII-162- stop	TCGAGCGTGGTCGATACCATGAGCGGTATTtgaGCATCGTTTGGTCTGATTTG
1789 CII-162- rev	AATACCGCTCATGGTATCGACCACGCTCGAC
1790 CII-163- stop	AGCGTGGTCGATACCATGAGCGGTATTGGCtgaTCGTTTGGTCTGATTTGAgg
1791 CII-163- rev	GCCAATACCGCTCATGGTATCGACCACGCTC
1792 CII-164- stop	GTGGTCGATACCATGAGCGGTATTGGCGCAtgaTTTGGTCTGATTTGAggatc
1793 CII-164- rev	TGCGCCAATACCGCTCATGGTATCGACCAC
1794 CII-165- stop	GTCGATACCATGAGCGGTATTGGCGCATCGtgaGGTCTGATTTGAggatccTA
1795 CII-165- rev	CGATGCGCCAATACCGCTCATGGTATCGAC
1796 CII-166- stop	GATACCATGAGCGGTATTGGCGCATCGTTTtgaCTGATTTGAggatccTAGAG
1797 CII-166- rev	AAACGATGCGCCAATACCGCTCATGGTATC
1870 attB1TEV_CII s tart	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCGAAAACCTGTATTTTCAGGGAA TGTTTGATTTTCAGGTTTCCAA
1903 attB2CII145 st op	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAACGTGCATGCAGCG

1904 CII-167-stop	ACCATGAGCGGTATTGGCGCATCGTTTGGTtgaATTTGAGGTggatccTAGAGGCATC
1905 CII-167-rev	ACCAAACGATGCGCCAATACCGCTCATGGT
1906 CII-168-stop	ATGAGCGGTATTGGCGCATCGTTTGGTCTGtgaGGTggatccTAGAGGCATC
1907 CII-168-rev	CAGACCAAACGATGCGCCAATACCGCTCAT
1908 CII-144-stop	TTGTCATTGTCGGCTCTGGCGCTGCATGCAtgaCTGCAGACTAATCCCGCTATG
1909 CII-144-rev	TGCATGCAGCGCCAGAGCCGACAATGACAA
1910 CII-143-stop	ATGTTGTCATTGTCGGCTCTGGCGCTGCATtgaCGTCTGCAGACTAATCCCGC
1911 CII-143-rev	ATGCAGCGCCAGAGCCGACAATGACAACAT
1912 CII-142-stop	CGCATGTTGTCATTGTCGGCTCTGGCGCTGtgaGCACGTCTGCAGACTAATCC
1913 CII-142-rev	CAGCGCCAGAGCCGACAATGACAACATGCG
1914 CII-141-stop	ATTCGCATGTTGTCATTGTCGGCTCTGGCGtgaCATGCActgCTGCAGACTAATC
1915 CII-141-rev	CGCCAGAGCCGACAATGACAACATGCGAAT
2128 CII_Q93A_fw d	CCGGTTAATGAGCTGGCTAAAGATAAATTGGCGTCTTACGTCATGCGCGCAATGAGTGAATC
2129 CII_Q93A_rev	CAATTTATCTTTAGCCAGCTCATTAACCGG
2130 CII_V96A_fw d	GAGCTGGCTAAAGATAAATTGCAGTCTTACGCGATGCGCGCAATGAGTGAATC
2131 CII_V96A_rev	GTAAGACTGCAATTTATCTTTAGCCAGCTC
2132 CII_M97A_fw d	CTGGCTAAAGATAAATTGCAGTCTTACGTCGCGCGCAATGAGTGAATC
2133 CII_M97C_fw d	CTGGCTAAAGATAAATTGCAGTCTTACGTCGCGCGCAATGAGTGAATC
2134 CII_M97_rev	GACGTAAGACTGCAATTTATCTTTAGCCAG
2135 CII_M100A_fw d	GATAAATTGCAGTCTTACGTCATGCGCGCAGCGAGTGAATC
2136 CII_M100C_fw d	GATAAATTGCAGTCTTACGTCATGCGCGCATGCAGTGAATC

2137 CII_M100_rev	TGCGCGCATGACGTAAGACTGCAATTTATC
2140 CII_V111A_fw wd	AGTGAAGCTCGGTGAACTGGCGAGCGGTGCGGCGTCTGATGAGCGTCTGACCAC TGCCCGTAAG
2141 CII_V111A_rev	CGCACCGCTCGCCAGTTCACCGAG
2142 CII_S112A_fw d	GAACTCGGTGAACTGGCGAGCGGTGCGGTAGCGGATGAGCGTCTGACCACTG CCCGTAAGCAC
2143 CII_S112A_rev	TACCGCACCGCTCGCCAGTTCAC
pE27 top	ATGTTTGATTTTCAGGTTTCCAAACAT
pE27 bottom	ATGTTTGAAACCTGAAAATCAAACAT
pE31 top	ACATGTTTGATTTTCAGGTTTCCAAACATCC
pE31 bottom	GGATGTTTGAAACCTGAAAATCAAACATGT
pE35 top	CAACATGTTTGATTTTCAGGTTTCCAAACATCCCC
pE35 bottom	GGGGATGTTTGAAACCTGAAAATCAAACATGTTG

Table S 2 - Strains for lacZ assay

Reporter assay	Strains used	Notes
Measuring activity of CII truncations	IM26 + pZS45 CII	CII variants and empty control expressed from pZS45 plasmid
Measuring activity of CII tetramerization mutants	IM26 + pZS45 CII	CII variants and empty control expressed from pZS45 plasmid
Alanine scanning mutagenesis of RNA polymerase α -CTD	IM50 + pHTf1 rpoA IM50 + pREII α rpoA	RNA polymerase α subunit mutants expressed from pHTf1 or pREII α plasmids CII ^{E46K} expressed from pZS45 plasmid

Table S 3– Growth media for lacZ assays

Step	Media		
	Measuring activity of CII truncations	Measuring activity of CII tetramerization mutants	Scanning alanine mutagenesis of RNA polymerase α -CTD
Restreak from glycerol stock	50 $\mu\text{g mL}^{-1}$ spectinomycin and 50 $\mu\text{g mL}^{-1}$ kanamycin	50 $\mu\text{g mL}^{-1}$ spectinomycin and 50 $\mu\text{g mL}^{-1}$ kanamycin	50 $\mu\text{g mL}^{-1}$ spectinomycin, 50 $\mu\text{g mL}^{-1}$ kanamycin and 100 $\mu\text{g mL}^{-1}$ ampicillin
Inoculation of cultures from restreak plate	50 $\mu\text{g mL}^{-1}$ spectinomycin and 50 $\mu\text{g mL}^{-1}$ kanamycin	50 $\mu\text{g mL}^{-1}$ spectinomycin and 50 $\mu\text{g mL}^{-1}$ kanamycin	50 $\mu\text{g mL}^{-1}$ spectinomycin, 50 $\mu\text{g mL}^{-1}$ kanamycin and 100 $\mu\text{g mL}^{-1}$ ampicillin
Inoculation of cultures for lacZ measurements	50 $\mu\text{g mL}^{-1}$ spectinomycin, 50 $\mu\text{g mL}^{-1}$ kanamycin and 80 μM IPTG	50 $\mu\text{g mL}^{-1}$ spectinomycin, 50 $\mu\text{g mL}^{-1}$ kanamycin and 100 μM IPTG	50 $\mu\text{g mL}^{-1}$ spectinomycin, 50 $\mu\text{g mL}^{-1}$ kanamycin, 100 $\mu\text{g mL}^{-1}$ ampicillin and 200 μM IPTG

CII-DNA docking parameters used for HADDOCK

DNA construction

```
1  TGCAACATGTTTGATTTTCAGGTTTCCAAACATCCCCAC  39
40 GTGGGGATGTTTGGAAACCTGAAAATCAAACATGTTGCA  78
```

DNA half sites are underlined. A global 40° angle bend was adopted by the DNA with 1° steps. Various DNA binding angles between 35°-45° were tested, with 40° found to be the optimal fit for CII145 structure 1.

Docking parameters

DNA active residues (DNA half sites): A7 T8 G9 T10 T11 T12 G13 C27 A28 A29 A30 C31 A32 T33 A46 T47 G48 T49 T50 T51 G52 C66 A67 A68 A69 C70 A71 T72

CII active residues: N32 Q37 T38 N41 K42

CII passive residues: M25 A29 A32 G33 M34 V36 R40 L43 N44 P45 Q47 T52 P54 L58

DNA segments between the half sites, center-to-center, were defined to be semi-flexible. This corresponds to the segments T7-A30 and A49-C69.

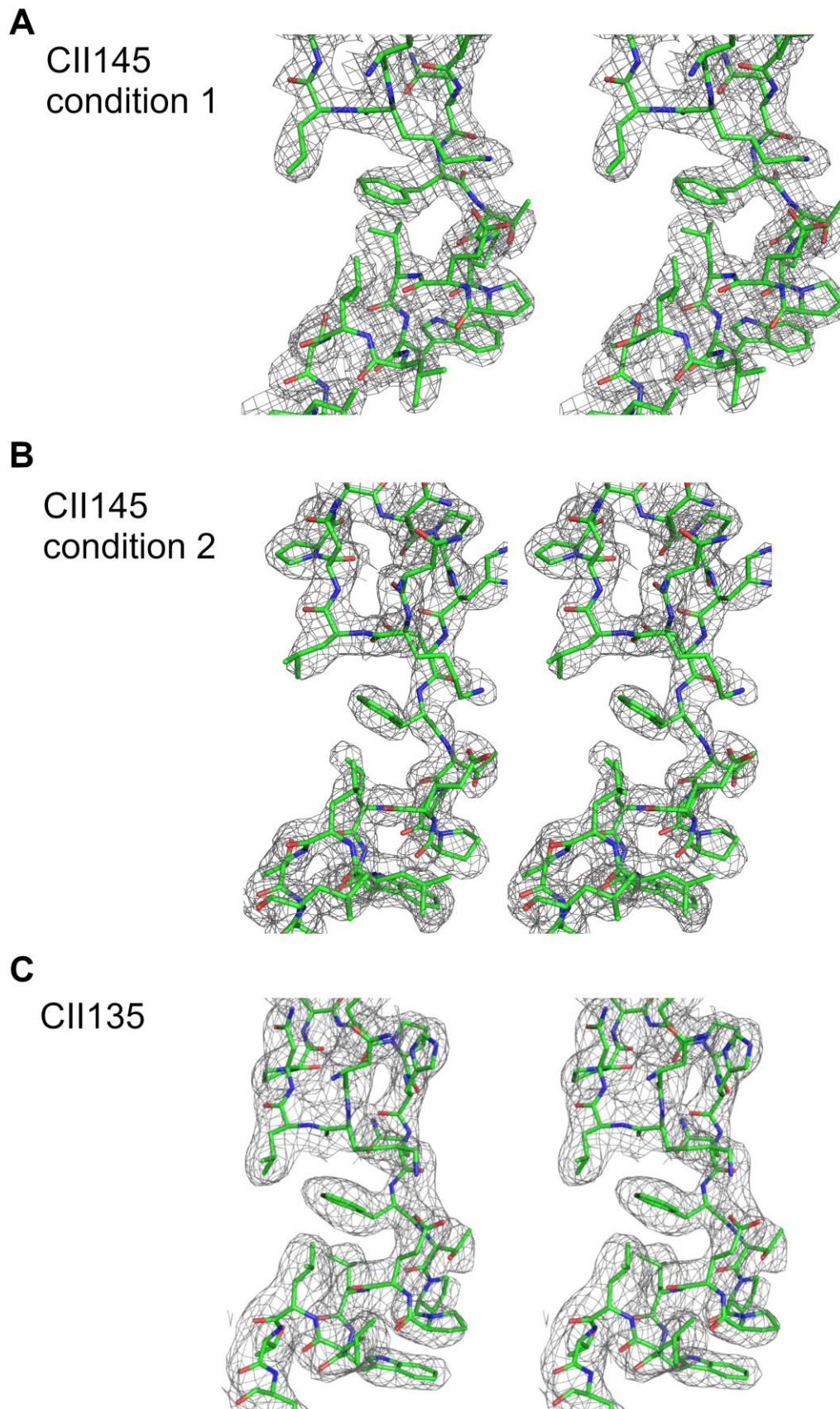


Figure S 1 - Stereoviews of composite omit 2mFo-dFc electron density map of CII145 (condition 1), CII145 (condition 2) and CII135 are shown in (A), (B) and (C), respectively. The contour level was set at 1σ .

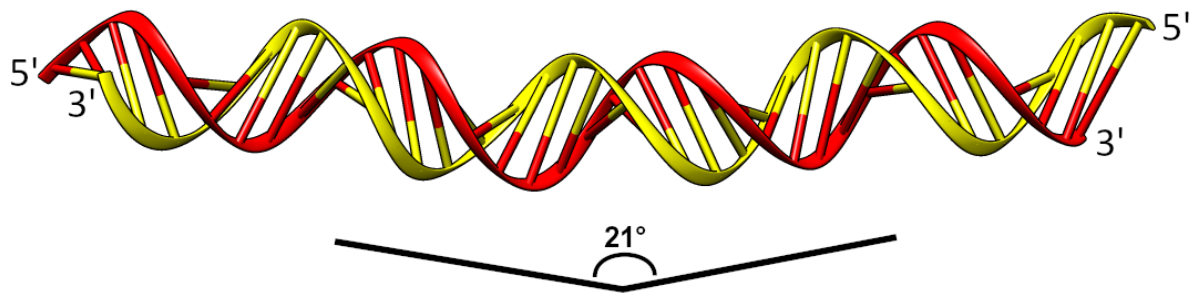


Figure S 2 – pE DNA (-67 to -29) was predicted to have an intrinsic bend of 21°. DNA was modelled using the bend.it server (Vlahoviček, Kaján and Pongor, 2003). DNA strand in red and yellow corresponds to 5'-TGCAACATGTTTGATTTTCAGGTTTCCAAACATCCCCAC-3' and 5'-GTGGGGATGTTTGAAACCTGAAAATCAAACATGTTGCA-3', respectively.

CII135 can bind a single pE half site

Native mass spectrometry was used to investigate DNA binding by CII135 *in vitro*. CII135 was mixed with a single p_E half site and analysed by electrospray ionization mass-spectrometry (ESI-MS). However, the resulting spectra are noisy with many unassigned peaks, making it difficult to draw definitive conclusions. A low abundance peaks corresponding to CII135 in complex with one p_E half site DNA could be identified (Figure S 3). This result provides an initial indication that a CII135 dimer can bind a single p_E half site. A CII135 dimer being able to bind an oligonucleotide via a single p_E half site would allow a CII145 tetramer to bind to two oligonucleotides via two p_E half sites. This species was detected as a low abundance species in the mass-spectra of CII145 in the presence of p_E DNA (Figure 6).

The ability of CII135 to bind DNA is preserved *in vivo*. It displayed significantly reduced DNA binding activity relative to wild-type CII and CII145 in pCIIR lacZ assays (Murchland *et al.*, 2014). However, the activity is greater than a no-CII control. CII135 exhibits 97% less transcriptional activation than wild-type, but the activity is detectably measured above inactivated CII (HTH inactivation mutant) and no-CII controls. The loss of activation function is CII135 is likely due to reduced DNA binding and loss of activation epitopes for RNAP. CII135 binding p_E operators *in vivo* also presents potential regulatory effects, where inactive dimers could bind to p_E half sites and occlude active CII from the promoter. It is uncertain how substantial this effect is, as CII tetramers have higher affinity to p_E and should outcompete the inactive protein.

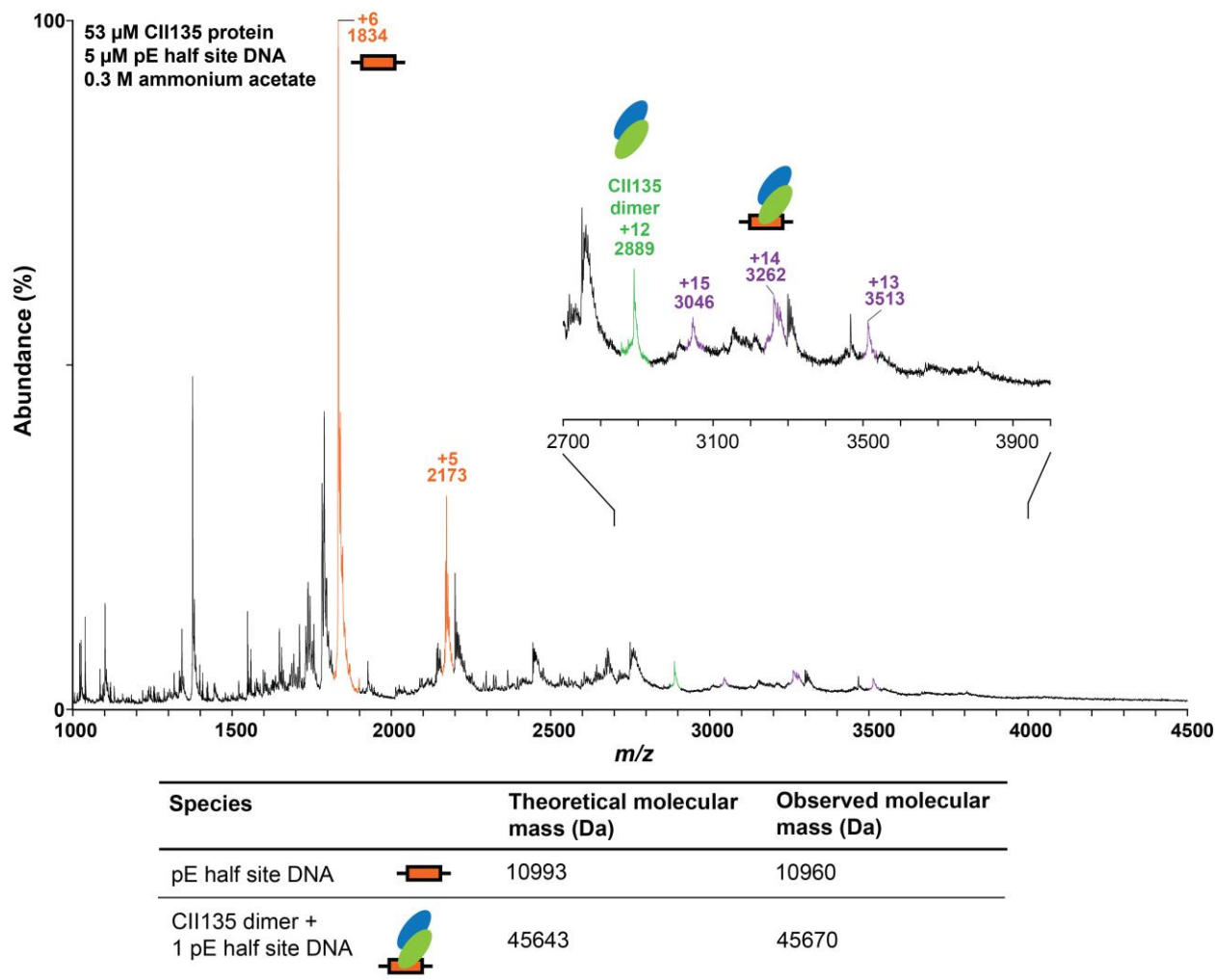


Figure S 3 - Native ESI-MS spectra of CII135 (53 μM) as a mixture with 10 bp p_E half site DNA (5 μM). A CII135 dimer was observed to form a complex with 1 piece of p_E half site DNA. The 2889 m/z peak was assigned as CII135 based on m/z assignments from an ESI-MS spectra of CII135 (Figure 9B).

3 STRUCTURAL STUDIES OF 186 CI USING X-RAY CRYSTALLOGRAPHY, SMALL ANGLE X-RAY SCATTERING AND MASS SPECTROMETRY

3.1 Introduction

In order for a phage to establish lysogeny, it must promptly shut down lytic gene expression after initial infection of the host bacterium and establish lysogenic gene expression, which is then maintained through subsequent cell generations. Chapter 1 reviewed the lytic-lysogenic decision in bacteriophage λ and 186, where the respective immunity repressors were highlighted as central to the maintenance of lysogeny. It is evident that cooperative interactions between the functional dimeric binding units of these repressors are essential to achieve efficient repression of lytic functions and exert positive and negative autoregulation on immunity repressor levels. This enables the lysogenic state to be stable, yet poised for escape. Higher order oligomerization of immunity repressor dimers has also been observed in other phages such as phage 434 in the presence of DNA (Ciubotaru and Koudelka, 2003), TP901-1 (Pedersen et al., 2008) and phage P22 (De Anda et al., 1983). The key points from Chapter 1 will be summarized below.

The immunity repressor of bacteriophage λ , λ CI, uses cooperativity at multiple levels to maintain lysogeny. λ CI binds to operators at the O_L and O_R sites of the λ regulatory region. The relative intrinsic affinity of the operators favours the occupation by λ CI dimers of O_{L1} and O_{L2} at O_L and O_{R1} and O_{R2} at O_R , over the weaker affinity O_{L3} and O_{R3} operators. This binding represses the lytic p_R and p_L promoters. λ CI dimers occupying these adjacent operators can form tetramers. λ CI tetramers can subsequently loop the intervening DNA between O_L and O_R to form an octameric complex, to further enhance the repression of the lytic promoters (Dodd et al., 2001; Ptashne, 2004). λ CI also regulates its own expression. At lower λ CI levels, the N-terminal domain of the λ CI protein occupying O_{R2} contacts the σ^{70} subunit of RNA polymerase, to activate its own expression along with expression of other lysogenic genes from p_{RM} (Jain et al., 2004; Nickels et al., 2002). However, at high levels of λ CI, the octameric complex that bridges O_L and O_R align O_{L3} and O_{R3} , allowing a λ CI tetramer to occupy them. Occupation of O_{L3} and O_{R3} represses λ p_{RM} to limit levels of λ CI in the cell (Dodd et al., 2004). This mechanism aids the responsiveness of prophage induction by allowing RecA to more rapidly reduce active λ CI levels following activation of the host SOS response.

Like bacteriophage λ , a supramolecular immunity repressor-DNA complex is involved in transcriptional regulation in bacteriophage 186. The 186 CI repressor is thought to oligomerize

into a 14mer wheel-like quaternary structure (Pinkett et al., 2006; Wang et al., 2013). This current model of 186 regulation states that the 186 CI 14mer acts as a scaffold, where 186 regulatory DNA can wrap on and off to regulate the lytic and lysogenic states (Pinkett et al., 2006) (Figure 1-15). DNA encompassing the 186 lytic p_R and lysogenic p_L promoters can wrap onto the wheel to repress both promoters. This repression is mediated by contacts with CI operator sites between p_R and p_L . The operators at p_R have stronger affinity and are better defined, relative to operators at p_L . The p_L - p_R region is flanked (~ 300 bp away) by CI operators F_L and F_R . These flanking sites can loop onto the 186 CI wheel to compete with and push off the p_L DNA, without disrupting the 186 CI- p_R interactions. This wrapping and looping mechanism allows for 186 CI to repress 186 p_R , whilst allowing lysogenic genes to be transcribed by p_L . The p_L promoter is further activated by relief of the transcriptional interference generated by the strong p_R promoter. At high levels of 186 CI, the flanking sites F_L and F_R are thought to be sequestered by additional 186 CI wheels, to abolish p_L derepression by F_L/F_R . This autoregulation limits 186 CI levels *in vivo* (Dodd and Egan, 2002). This is analogous to λ CI negative autoregulation, to prevent CI levels from becoming excessive and impeding prophage induction.

Structural characterisation has been conducted on 186 CI. Pinkett *et al.* (2006) solved the crystal structures of two different variants of the 186 CI protein, a full length dimer (E146K mutant) and the C-terminal oligomerization domain, which was solved as a 14mer. Wang *et al.* (2013) observed 186 CI interacting with DNA as low resolution discs via atomic force microscopy. The 186 CI 14mer complex is central to the current model of lytic-lysogenic transcriptional regulation in bacteriophage 186, yet no high resolution structure of the complex has been obtained. If we wish to use 186 as a counterpoint model to λ , it would be extremely useful to have an atomic resolution structure of the full-length wild type 186 CI wheel. Firstly, such a structure would validate our model of cooperative regulation in this phage system. Secondly, it would demonstrate how the N-terminal DNA binding domains are arranged when the protein exists as a wheel. This allows further refinement of model by showing how the operators are bound by the wheel and, in turn, explain the arrangement of operators on the prophage DNA.

3.2 Expression of the 186 CI repressor

The full length 186 CI protein was encoded under the control of the T7 promoter of the pET3aHis₆ vector (Shearwin et al., 2002). The expressed protein contains a C-terminal thrombin cleavage sequence followed by a hexahistidine tag. This same affinity tag was previously used to purify 186 CI CTD. 186 CI-His₆ was previously shown to be active *in vitro* (Shearwin et al., 2002). *E. coli* BL21(λ DE3) pLysS was used as the host for the expression of the CI protein. Cells were grown at 37°C to an OD₆₀₀ of 0.6 in 2 x 1 L LB medium containing 100 μ g mL⁻¹ ampicillin and 30 μ g mL⁻¹ chloramphenicol. 186 CI production was induced upon addition of 200 μ M IPTG and

cultivation overnight at 25°C. Cells were harvested by centrifugation, suspended in approximately 40 mL of lysis buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, 5 mM 2-mercaptoethanol, pH 7.4) and stored at -20°C.

3.3 Purification

All purification steps were performed at 4°C on a Bio-Rad NGC Chromatography system. Frozen *E. coli* cells were thawed and lysed using high pressure disruption (Microfluidics cell disruptor). The lysate was clarified using centrifugation at 40 000g for 1 hr at 4°C. Protein for crystallization trials was produced as follows. The cell lysate was applied to 5 mL HisTrap FF column (GE Healthcare) pre-equilibrated with lysis buffer. Bound proteins were eluted using an imidazole gradient (20 mM – 500 mM). Fractions showing above baseline UV absorbance at 280 nm were tested for purity using SDS-PAGE (Figure 3-1). Fractions containing the purified protein were pooled and dialysed against 50 mM Tris-Cl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.1 mM DTT. The protein was concentrated to 7.7 mg mL⁻¹ by ultracentrifugation on Amicon Ultra-15 centrifugation units (Millipore) prior to crystallization trials.

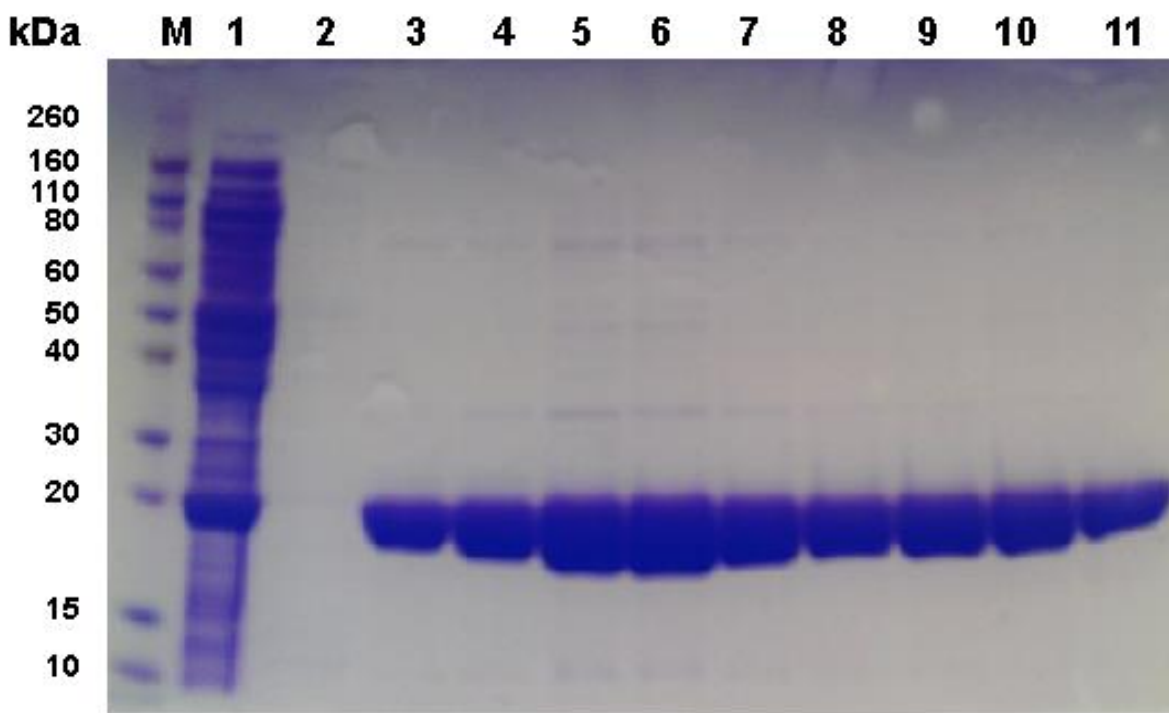


Figure 3-1 – SDS-PAGE analysis of fractions from Ni-NTA chromatography to purify 186 CI. Samples were run on 4-12%NuPAGE BisTris Protein Gels (ThermoFisher Scientific) with NuPAGE MOPS buffer. Lane M: Novex Sharp Pre-Stained Protein Standard (ThermoFisher Scientific, 2 µL), Lane 1: Crude *E. coli* lysate prior to purification. Lane 2-11: Various elution fractions that show UV absorbance above baseline.

The 186 CI repressor protein could be purified to high purity in a single step using Ni²⁺ IMAC chromatography. SDS-PAGE analysis revealed a dominant band in the Ni²⁺ IMAC purification fractions around 20 kDa, consistent with the theoretical molecular weight of CI-His₆ of 22.6 kDa.

Several minor higher molecular weight impurities were also observed but did not prevent subsequent crystallization of the protein.

3.4 Crystallization

Crystallization screening was conducted via sitting-drop vapour diffusion in 96 well Intelliplates (Art-Robbins) using a Phoenix robot (Art Robbins). The commercial screens used to screen for crystallization conditions were the PEG/ION, Natrix, Crystal screens from Hampton Research and the MCSG suite from Microlytic (<https://production.moleculardimensions.com/products/microlytic-screens>). 1 μ L of protein was mixed with an equal volume of reservoir solution and equilibrated over 75 μ L reservoir solution. Three crystals were observed in the G10 condition of the MCSG1 suite. This corresponded to reservoir solution containing 15% (w/v) PEG 3350 and 0.1 M magnesium formate. Crystals were cryoprotected prior to data collection by passing the crystal through mother liquor containing 13% (v/v) glycerol. Subsequent attempts to reproduce the crystals were not successful.

3.5 X-ray data collection and processing

Only one of the three crystals, shown in Figure 3-2, from the initial screen gave diffraction greater than 4 \AA resolution. Data was collected from the single crystal at 100 K at the Australian Synchrotron on the MX1 beamline (Aragão et al., 2018; McPhillips et al., 2002). Diffraction data was processed and scaled using iMosflm and Scala (Battye et al., 2011; Evans and Murshudov, 2013). The data processing statistics of the diffraction from this crystal are listed in Table 3-1.

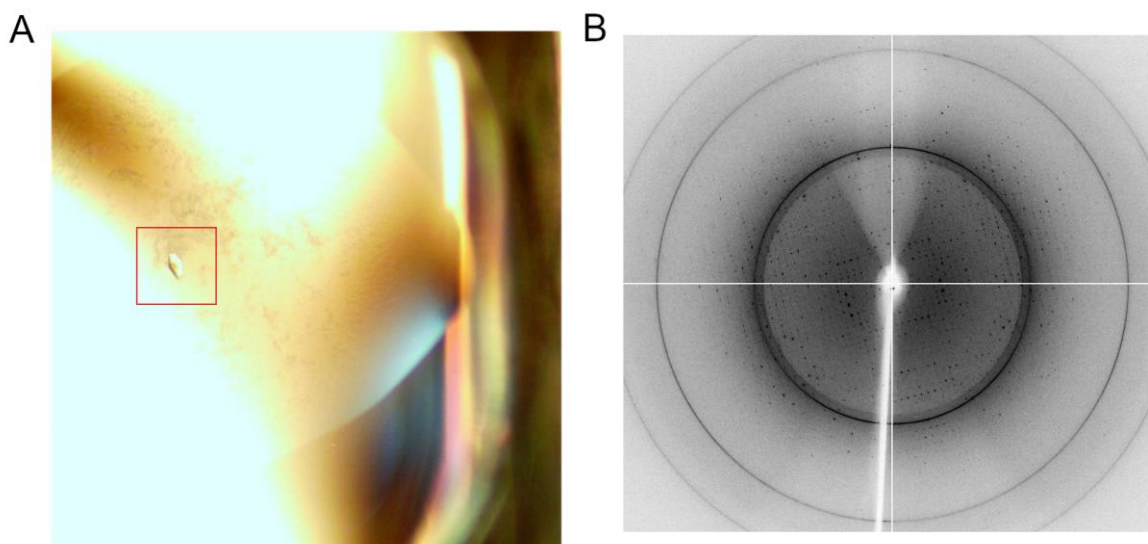


Figure 3-2 – 186 CI was successfully crystallized and diffraction data was collected from the crystal. A) Crystal of 186 CI grown in MCSG1 G10. Crystal position is indicated by the red box. B) Representative image of the diffraction acquired from the crystal shown in A.

Table 3-1 - Data collection statistics for the native data. Values in parentheses are for the highest resolution shell.

Synchrotron radiation source	MX1, Australian Synchrotron, Melbourne
Detector	Quantum 210r (ADSC)
Wavelength (Å)	0.9537
Data collection temperature (K)	100
Space-group	P2 ₁
Unit-cell parameters (Å, °)	a=73.40 b=81.08 c=113.50 α =90.00 β =90.03 γ =90.00
Matthews coefficient V _M	2.54 assuming 6 molecules in the ASU 2.17 assuming 7 molecules in the ASU
Solvent content (%)	51% assuming 6 molecules in the ASU 43% assuming 7 molecules in the ASU
Resolution range (Å)	54.41-2.30
Total reflections	424697 (32888)
Unique reflections	59330 (4575)
Completeness (%)	99.9 (100)
Redundancy	7.2 (7.2)
$\langle I/\sigma(I) \rangle$	16.5 (7.1)
CC _{1/2} (%)	99.8 (98.3)
R _{merge} (%)	8.5 (24.5)
R _{pim} (%)	3.7 (10.6)

The Matthews coefficient was used to estimate the number of monomers in the asymmetric unit. The Matthews coefficient, defined as

$$V_M = \frac{\text{volume of unit cell}}{ZX \times \text{molecular weight of macromolecule}}$$

Z is the number of asymmetric units in the unit cell. X is the number of molecules in the asymmetric unit and is initially not known. Empirically it has been found that V_M typically falls between 2 and 3 Å³/Da. This suggests that there are six or seven monomers in the asymmetric unit, which give a V_M of 2.54 Å³ Da⁻¹ and 2.17 Å³ Da⁻¹, respectively.

3.6 Attempts to solve the phase problem

3.6.1 Principles of molecular replacement

To solve the 3-dimensional structure from the X-ray diffraction of its protein crystal, two parameters need to be determined for the structure factors; amplitude and phase. The amplitude can be measured directly from the diffraction image, but the phase cannot. One of the methods to obtain the phases is molecular replacement (Rossmann and Blow, 1962). In molecular replacement, the phases come from a previously solved or predicted model, called a probe or search model.

For the molecular replacement method to provide phases, the probe needs to be positioned in the unit cell of the target crystal such that the theoretical diffraction from the probe resembles the experimental diffraction. Six spatial parameters are fitted to each probe (3 translational and 3 rotational) to position it in the unit cell.

Different molecular replacement algorithms are available to arrive at these six parameters. The molecular replacement program Phaser (McCoy et al., 2007), which employs a maximum likelihood algorithm, breaks up the search for the 6 parameters into two 3-dimensional searches. First, the three rotational parameters are fitted, followed by the three translational parameters. Some programs search for all six parameters simultaneously. The EPMR program performs a six-dimensional search by using evolutionary search algorithm to optimize the rotation and translation of the probe simultaneously (Kissinger et al., 1999).

3.6.2 Attempt to solve phase problem

Attempts to solve the phase problem using molecular replacement, using the programs Phaser (McCoy et al., 2007) and EPMR (Kissinger et al., 1999), failed to produce a correct solution. The search models used for molecular replacement were structures of previous 186 CI crystal structures solved by Pinkett (2006). Molecular replacement was attempted with the full CI^{E146K} dimer, a CI^{E146K} monomer, the CI N-terminal domain and the CI C-terminal domain, as a monomer and an oligomer with arrangements guided by the existing crystal structure.

3.6.3 Reasons that may contribute to molecular replacement failing

The search models used for molecular replacement were either domains of CI or a single amino acid substitution mutant. However, even with these high similarity search models, molecular replacement was unable to solve the phase problem.

Molecular replacement using a high similarity probe can fail when there have been conformational changes in the protein, such that the probe is no longer sufficiently structurally similar to the true crystal structure. The CI NTD and CI CTD are connected via a flexible linker and a movement of the CI NTD relative to the CTD may make the CI^{E146K} search model an unsuitable probe. One possible way to tackle this is to use the NTD and CTD as separate search models, but this strategy has its own caveats.

One factor that makes molecular replacement more difficult in this case is the presence of an estimated 6-7 molecules per asymmetric unit of the crystal. As the number of components in the asymmetric unit increases, the proportion of scattering that each component contributes to the total decreases. Thus, the low signal of each isolated component becomes harder for MR algorithms to find. Assuming six molecules in the asymmetric unit, the NTD and CTD contribute roughly 6% and 9%, respectively, of total protein scattering and may be missed by molecular replacement programs.

With the roadblocks encountered with solving the crystal structure of wildtype 186 CI, the alternate biophysical techniques small angle X-ray scattering (SAXS) and electrospray ionization mass spectrometry (ESI-MS) were used to study the structure and oligomerization of 186 CI.

3.7 SAXS

SAXS is a solution-based technique. A collimated X-ray beam with defined wavelength is passed through a sample in solution and the intensity of scattered X-rays is measured on a detector (Figure 3-3). In solution, the sample can adopt all orientations. Solution scattering from all the orientations is averaged to give rise to radially symmetrical scattering collected at small angles to determine particle shape. In SAXS, the intensity of detected X-rays (I) is measured as a function of scattering angle. In most cases, scattering angle is expressed as momentum transfer $q = (4\pi\sin\theta)/\lambda$, where λ is the wavelength of incident X-ray and 2θ is the angle of the scattered X-ray relative to the direct beam.

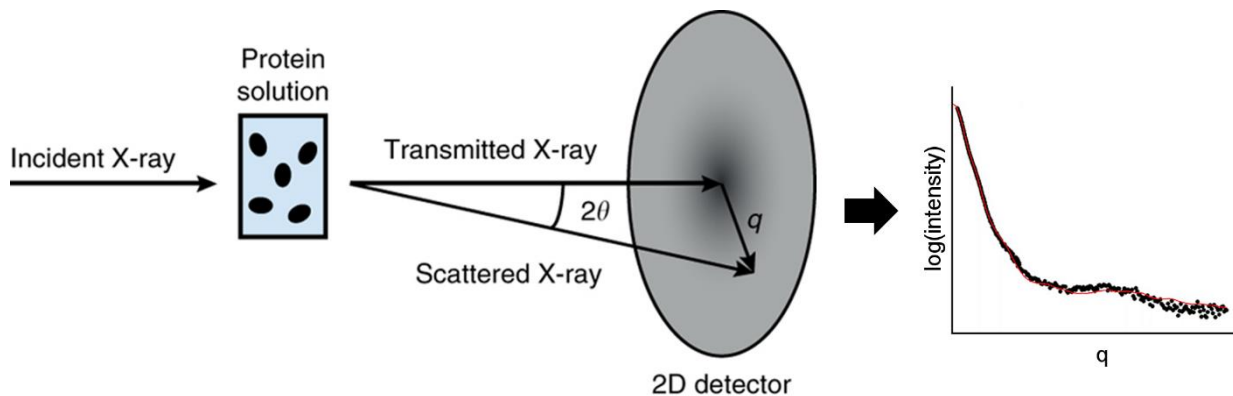


Figure 3-3. The basic setup for a SAXS experiment. A protein sample in solution is illuminated with a collimated, monochromatic X-ray beam. The solution scattering by the protein is measured using an X-ray detector. The intensity of scattered X-rays is measured as a one-dimensional function of momentum transfer q . Figure adapted from Skou et al. (Skou et al., 2014) and SAXIER (SAXIER, 2018).

3.8 Preparation of 186 CI protein for SAXS

Two variants of the 186 CI protein were overexpressed and purified for SAXS analysis. They were the wildtype 186 CI and the 186 CI^{E146K} proteins, each with a C-terminal hexahistidine tag and intervening thrombin cleavage site. 186 CI^{E146K} is a 186 CI cooperativity mutant that can form dimers, but is unable to form higher order oligomers (Pinkett et al., 2006). Both proteins were overexpressed and purified as described in Section 3.3, with modifications described here. After Ni-IMAC purification, both proteins were dialysed into 50 mM Tris-HCl, 150 mM NaCl, 10 mM EDTA, 5% glycerol, 5 mM TCEP, pH 7.5. The proteins were concentrated by ultracentrifugation using an Amicon Ultra-15 centrifugation unit (Millipore). The concentrated protein solutions were dialysed again into the same dialysis buffer. The final concentrations of 186 CI and 186 CI^{E146K} proteins as determined by UV absorbance were 8.7 mg/mL and 14.7 mg/mL, respectively. The extinction coefficient of the proteins at 280 nm was calculated using ExPASy ProtParam webserver (Gasteiger et al., 2005) to be 23950 M⁻¹ cm⁻¹.

3.9 SAXS data acquisition and processing

SAXS data for 186 CI and 186 CI^{E146K} was collected at the P12 EMBL BioSAXS beamline (Deutsches Elektronen-Synchrotron, Hamburg, Germany) on a Pilatus 2M photon-counting pixel detector, in collaboration with Dr. Haydyn Mertens. The sample-to-detector distance was set at 3 m. The temperature of the cell was 15°C. The protein concentrations of the samples used for data collection were 7.4, 3.7, 1.85, 0.93 and 0.46 mg/mL for CI^{E146K} and 8.7, 4.4, 2.2, 1.1 and 0.5 mg/mL for CI. The protein was diluted with dialysis buffer. Dialysis buffer was also collected as a blank reading for subtraction of background scattering. Background subtracted scattering profiles are shown in Figure 3-4.

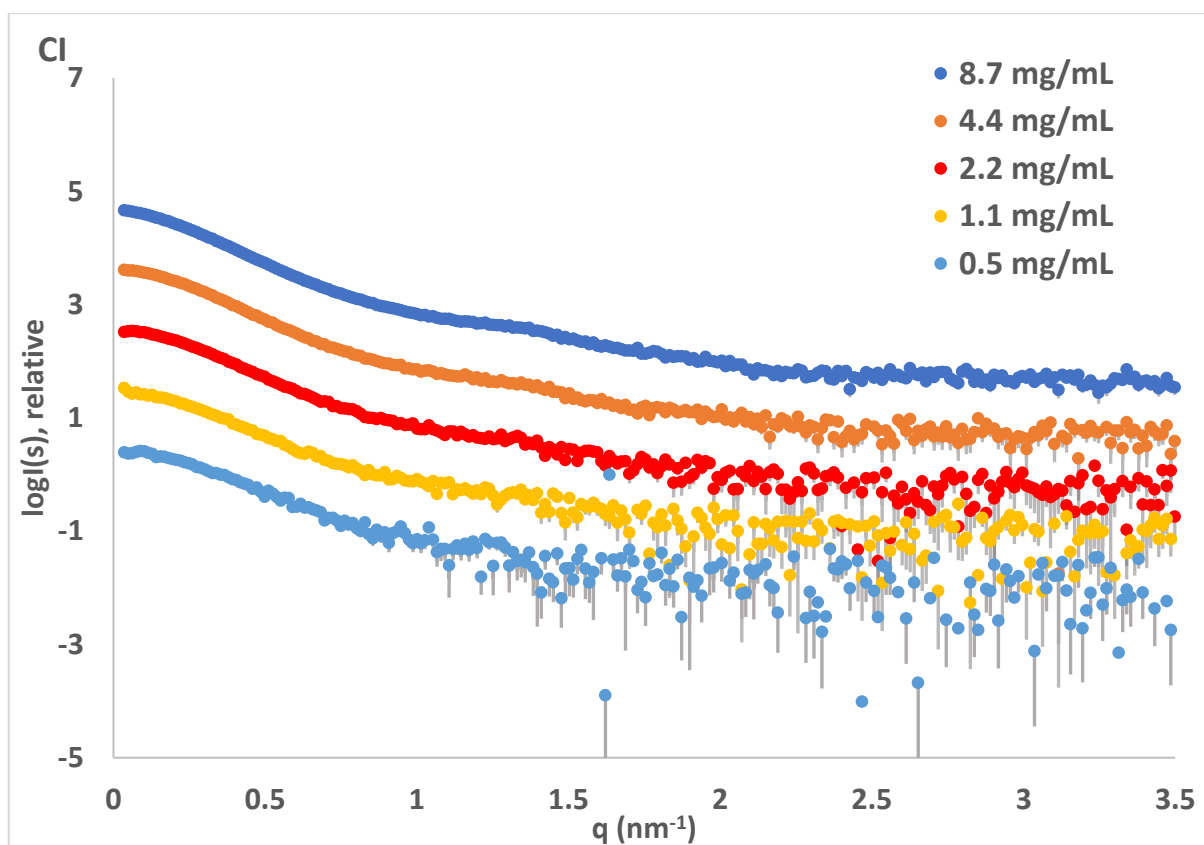
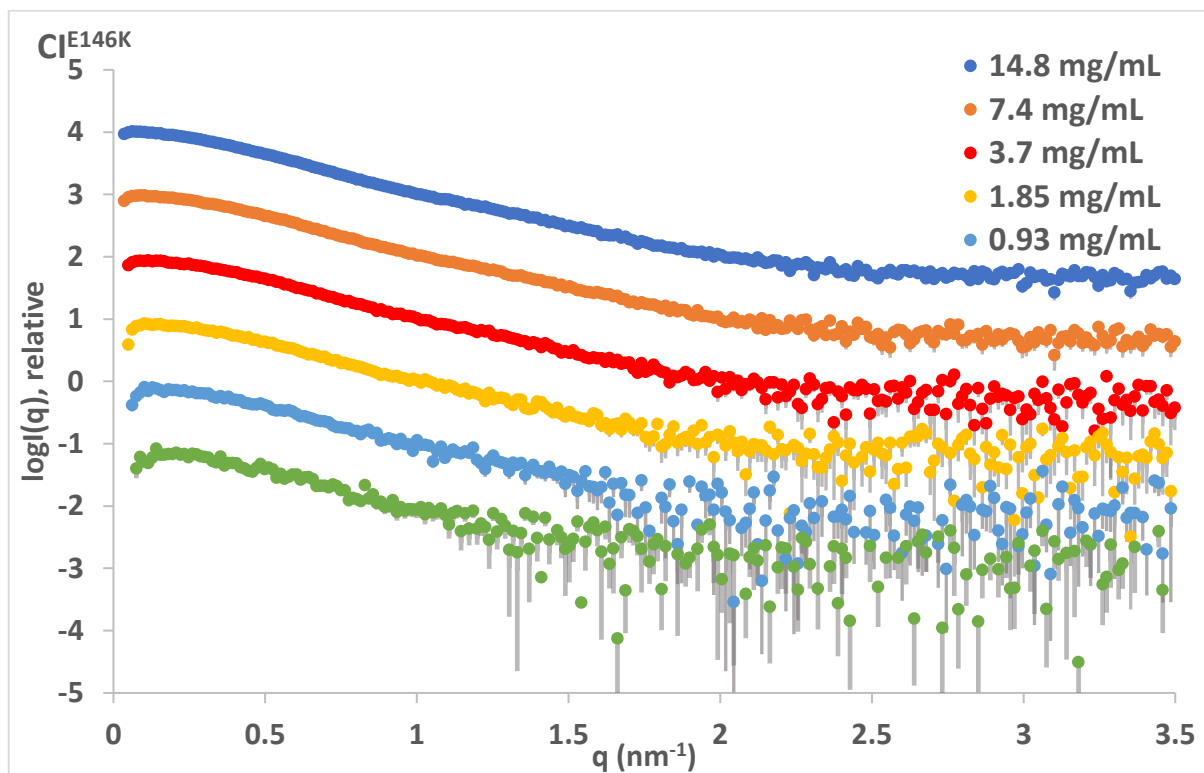


Figure 3-4 – Normalized scattering profiles of 186 Cl and 186 Cl^{E146K}. Scattering profiles are plotted as the logarithm of intensity of scattering against the momentum transfer q . Each plot has had background scattering subtracted and normalized to the protein concentration. Plots have been offset on the log scale. Grey lines represent the error bars on the measurements.

All programs used to analyse and visualize SAXS data are available in the ATSAS package (Franke et al., 2017). SAXS data analysis was conducted with the help of Dr. Haydyn Mertens and Dr. Fiona Whelan. Kratky plots for CI and CI^{E146K} at 8.4 mg/mL and 14.8 mg/mL, respectively, were drawn in SASplot (Franke et al., 2017). Guinier analysis was performed using the AutoRg function of Primus for each dataset (Konarev et al., 2003). The upper limit for data points selected for Guinier analysis was determined such that $qR_g \leq 1.3$. The lower limit was automatically selected by Primus. Pair distance-distribution functions $P(r)$ were calculated by GNOM (Svergun, 1992). The intensity of zero angle scattering $I(0)$ and radius of gyration R_g were extracted from Guinier and $P(r)$ plots. The maximum linear dimension of the scattering molecules was also determined using $P(r)$ plots.

Several approaches were used to fit structural models to the scattering data of CI^{E146K}. The simulated scattering curve of CI^{E146K} crystal structures (PDB ID 2FJR) and CI-hybrid dimer (Chapter 4) was evaluated using CRY SOL (Svergun et al., 1995). Rigid body modelling of CI^{E146K} was performed using CORAL (Petoukhov et al., 2012). CI^{E146K} was treated as a multi-domain protein containing rigid domains, linked by dummy residues where the NTD and CTD domains could move relative to each other. The Advanced Ensemble Optimization Method (Tria et al., 2015) was used to fit the CI^{E146K} SAXS scattering curve with an ensemble of CI^{E146K} conformers, based on the crystal structure.

3.10 SAXS results

3.10.1 Kratky analysis

SAXS can be used as method for investigating unfolded and flexible proteins. A qualitative assessment of flexibility in a protein studied by SAXS can be made using a Kratky plot (Glatter and Kratky, 1982). A Kratky plot is a plot of $q^2 I(q)$ against q and has distinctive characteristics depending on the degree of protein folding (Figure 3-5) (Kikhney and Svergun, 2015). Well folded globular proteins display a peak at low q and return to near zero as q increases. In contrast, Kratky plots of completely unfolded proteins characteristically rise to high $q^2 I(q)$ at low q , but do not return to zero but feature a rising baseline at high q . Kratky plots for CI and CI^{E146K} indicate the protein is partially folded. Partially folded proteins show a peak at low q , with an intermediate baseline. The X-ray crystal structure of CI^{E146K} comprises two folded globular domains connected by a 15 residue linker. The Kratky plot indicates that in solution this linker is likely disordered. In addition, the NTD of 186 CI has previously been suggested to require some conformational flexibility to allow it to bind to two distinct DNA half-sites. Comparison of λ CI and 186 CI^{E146K} structures revealed that the NTD of 186 had a shorter helix 5, which may contribute to NTD flexibility by reducing structural constraints (Pinkett et al., 2006).

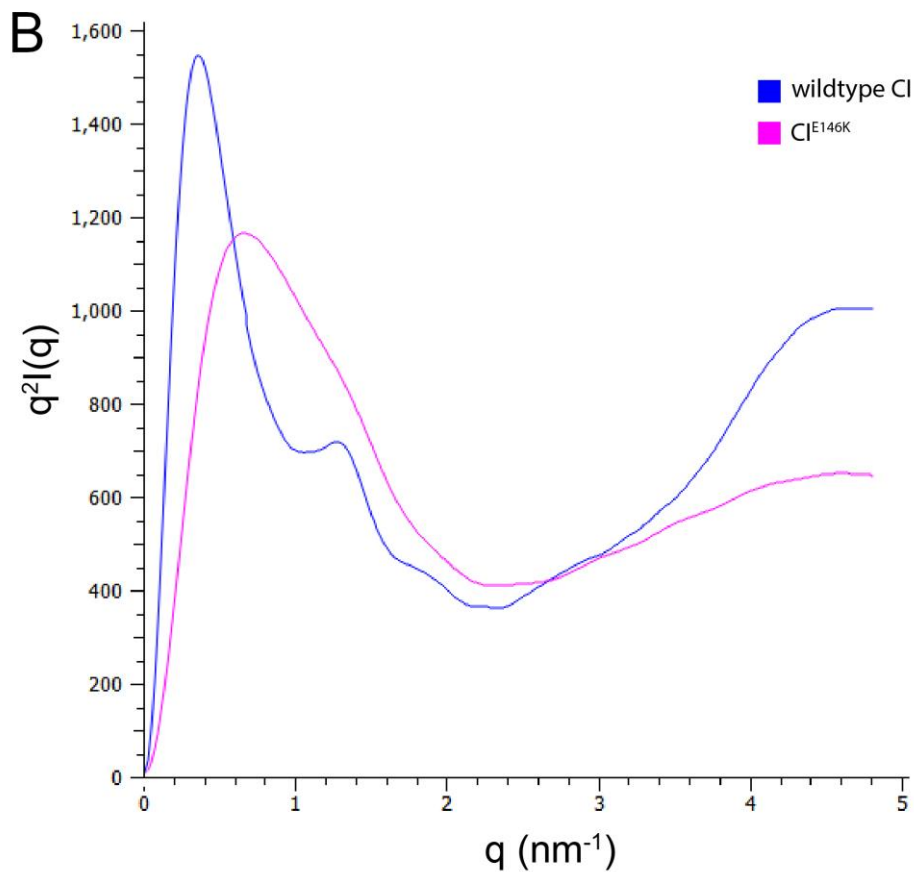
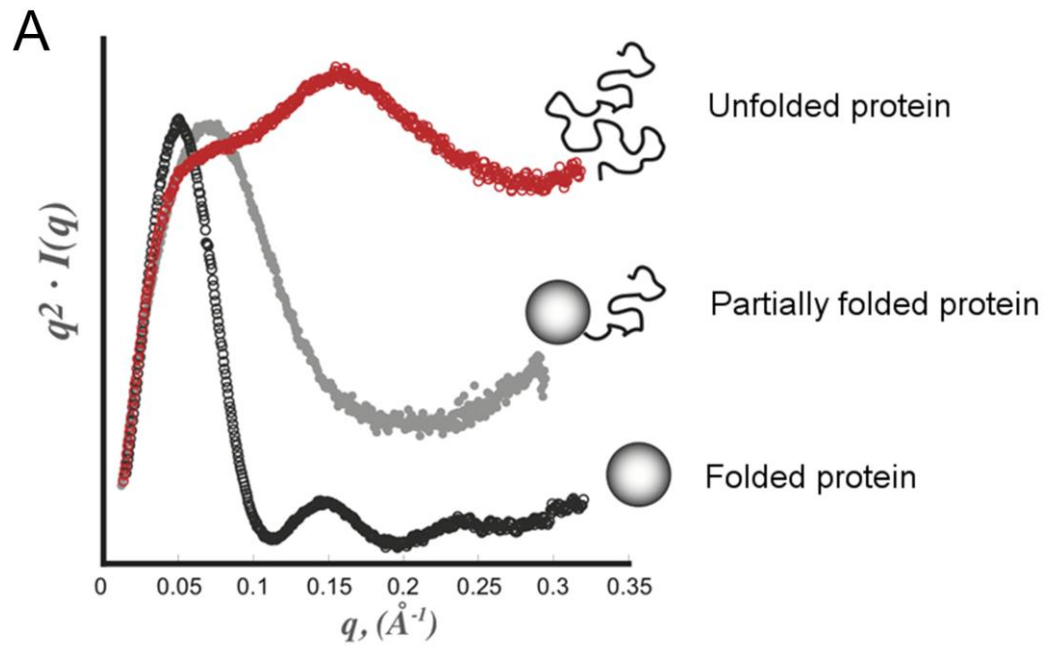


Figure 3-5 – Kratky plot for CI and CI^{E146K} suggests that the protein is partially folded. (A) Characteristic Kratky plots ($q^2 \cdot I(q)$ vs q) for proteins displaying different proportions of foldedness. Well folded globular proteins display a bell-shaped curve in a Kratky plot, returning to near zero values at higher q . Unfolded proteins and peptides have increasing $q^2 \cdot I(q)$ values that do not return to baseline at high q . Figure adapted from Rambo and Tainer (2011). (B) Kratky plots for CI and CI^{E146K} show a peak at low q and an intermediate decrease in the mid to high q range, characteristic of a partially folded protein.

3.10.2 Guinier analysis

SAXS scattering at low resolutions can be described by the Guinier approximation

$$I(q) = I(0)e^{-\frac{R_g^2 q^2}{3}} \quad (\text{Equation 1})$$

where $I(0)$ is the scattering intensity at an angle of zero and R_g , the radius of gyration, is the root mean square distance of each contributing scatterer from the centre of the particle. For globular particles, the Guinier approximation works over a range where $qR_g < 1.3$ (Guinier, 1939). Rearranging Equation 1 gives Equation 2.

$$\ln I(q) = -\frac{R_g^2}{3} q^2 + \ln I(0) \quad (\text{Equation 2})$$

Thus, a Guinier plot ($\ln I(q)$ vs. q^2) has a y -intercept of $\ln I(0)$ and a slope of $-R_g^2/3$. Thus a Guinier approximation allows one to estimate $I(0)$ and R_g , which provides structural parameters of the protein. Assuming a monodisperse solution, $I(0)$ is directly proportional to the molecular mass of the complex in solution. R_g is dependent on the conformation of the protein, with more compact proteins giving smaller R_g values than extended conformations, when comparing two proteins of the same molecular weight (Kikhney and Svergun, 2015).

Guinier plots were made for each dataset and used to determine $I(0)$ and R_g for each of the proteins at different concentrations (Figure 3-6 and Figure 3-7). Visual assessment of the Guinier plots show linearity in this region for both proteins, which is indicative of monodispersity (Figure 3-6). However, subsequent analysis of the $I(0)$ and R_g fits of the SAXS demonstrated polydispersity, which is possible if aggregation (such as oligomerization) and interparticle interference are subtle (Grant et al., 2015). In the low q data beyond the lower limit set for Guinier analysis, there was a slight deviation upwards from linear fit, also suggesting potential aggregation.

186 CIE^{E146K} is a cooperativity mutant which forms dimers, but is unable to form higher order oligomers (Pinkett et al., 2006). This is reflected in its R_g and $I(0)$ values, which are significantly smaller than for wildtype 186 CI. The slight increase in R_g and $I(0)$ likely reflects the increase in dimeric species as concentration increases. The larger R_g and $I(0)$ observed for wildtype CI is consistent with the higher order oligomeric species being formed by this protein. As protein concentration increases for the wild type species, R_g and $I(0)$ also increase, reflecting a shift in the concentration-dependent equilibrium towards larger oligomeric species (Figure 3-7).

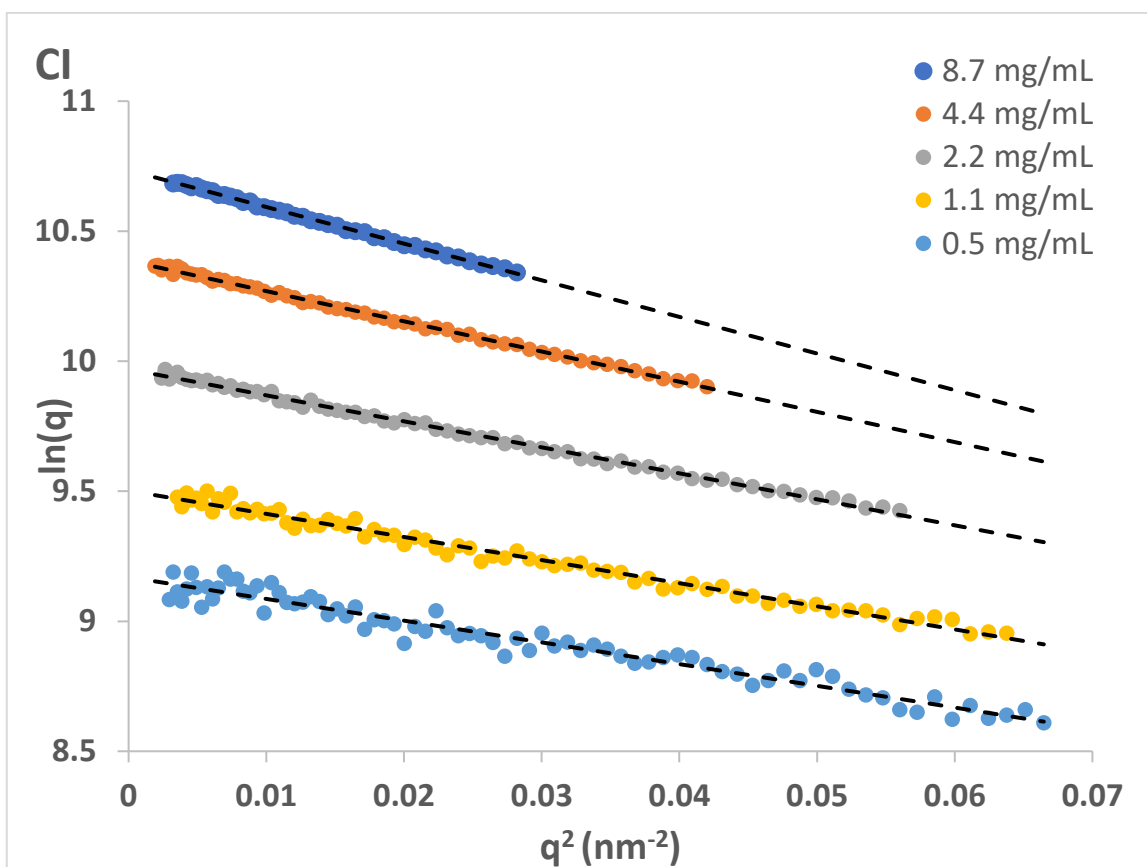
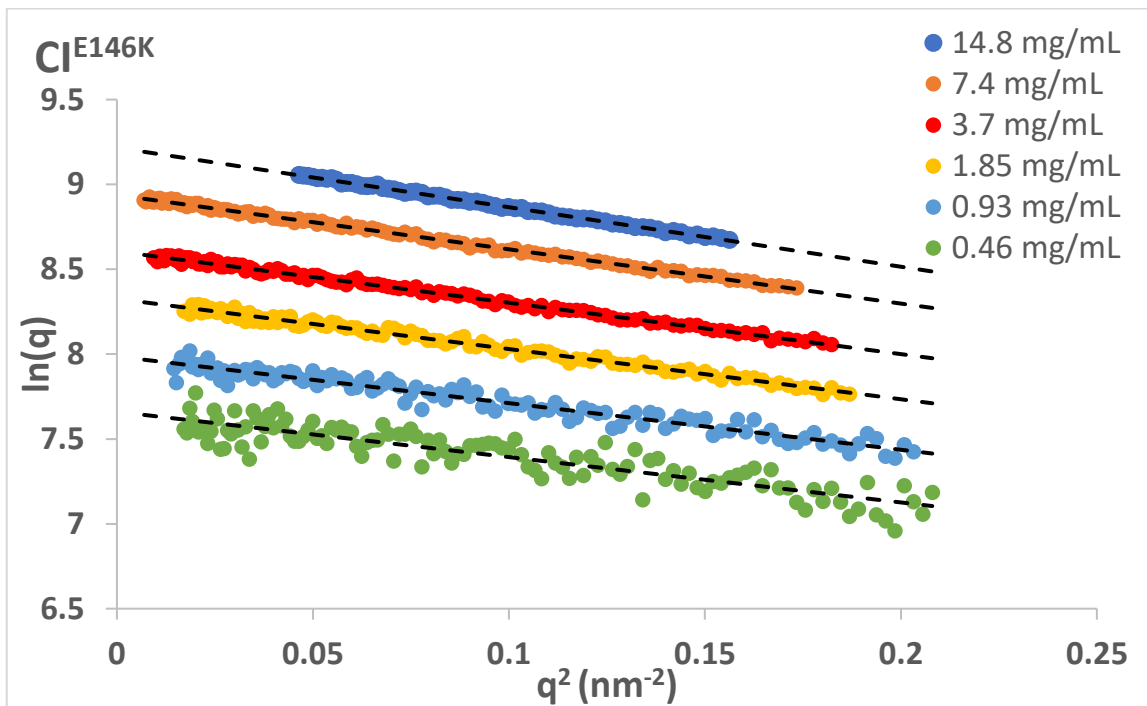


Figure 3-6 – Guinier plots from SAXS scattering by 186 Cl^{E146K} (upper) and 186 Cl (lower). Plots are offset on the log scale.

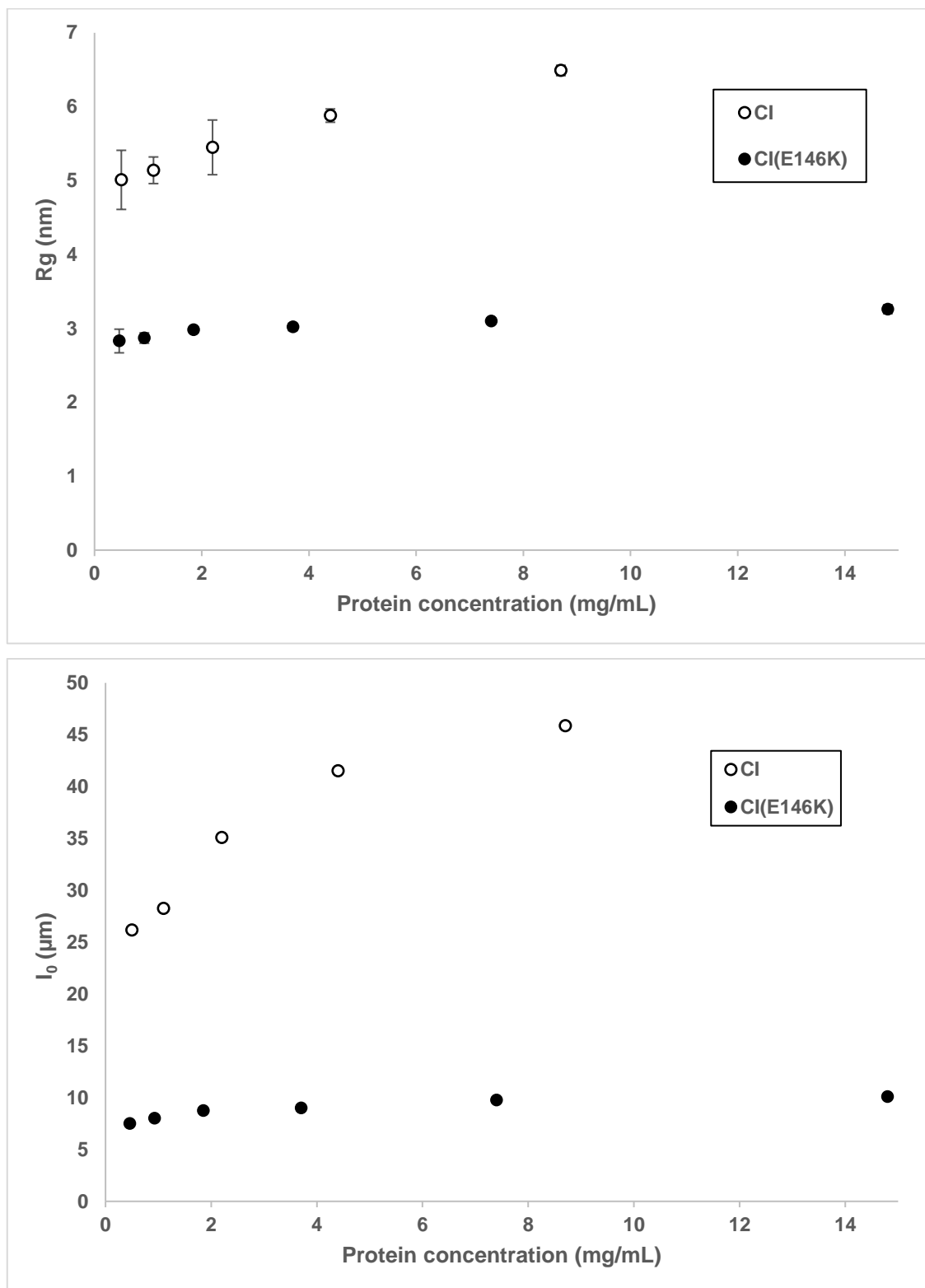


Figure 3-7 – Guiner plots were used to extract the radius of gyration R_g (upper) and intensity of zero angle scattering $I(0)$ (lower) for 186 CI and 186 CI^{E46K} at different concentrations. The concentration dependent formation of higher order wild type CI oligomers is reflected in increasing $I(0)$ and R_g .

As zero angle scattering intensity $I(0)$ is directly proportional to the molecular weight of the complex, $I(0)$ can be used to measure the molecular weight of the complex. The absolute intensity of scattering is not directly measured, thus to convert to absolute units, secondary standards are used (Mylonas and Svergun, 2007). The standard used in this experiment was water, which has a known scattering of $1.637 \times 10^{-2} \text{ cm}^{-1}$ at 15°C (Orthaber et al., 2000). The $I(0)$

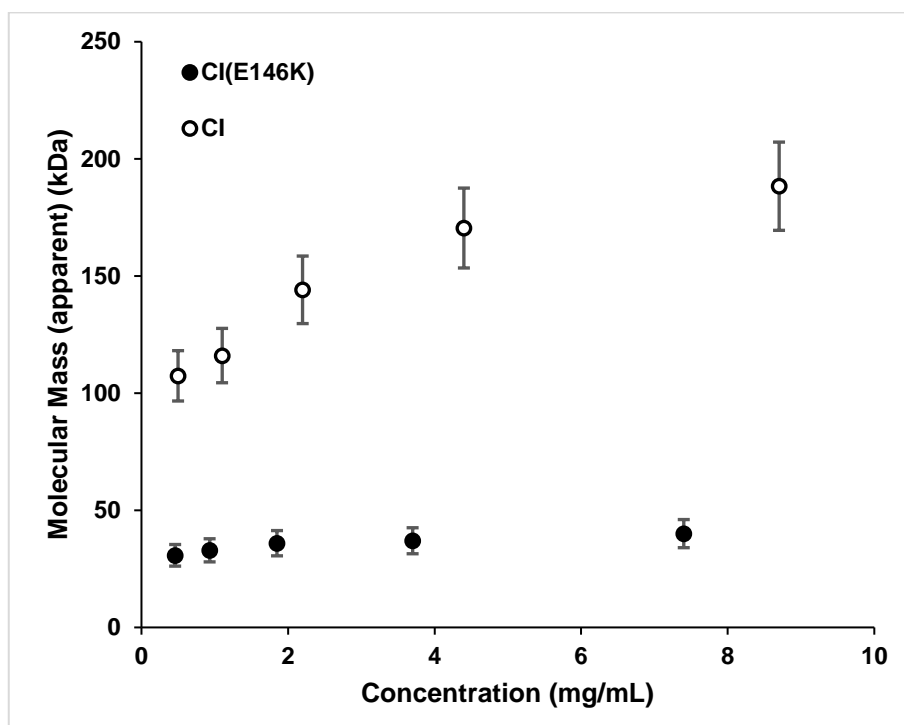
of water was determined to be $5879 \times 10^{-2} \text{ cm}^{-1}$. To scale the $I(0)$ of each protein sample into absolute units, equation 3 was used.

$$I(0)_{Absolute} = \frac{I(0)_{Relative}}{5879} \times 1.637 \times 10^{-2} \text{ cm}^{-1} \text{ (Equation 3)}$$

where $I(0)_{Absolute}$ is the scattering in absolute units and $I(0)_{Relative}$ is the experimentally measured scattering. The molecular mass of the scattering complexes can be calculated from $I(0)_{Absolute}$ using equation 4 (Mylonas and Svergun, 2007).

$$MM = \frac{N_A I(0)_{Absolute}}{\Delta\rho_M^2} \text{ (Equation 4)}$$

where $N_A = 6.023 \times 10^{23} \text{ mol}^{-1}$ is Avogadro's number and $\Delta\rho_M$ is the scattering contrasting of the sample. $\Delta\rho_M$ is the product of the scattering length difference of the protein and the partial specific volume. The scattering length difference of the protein was calculated using the MULCh Contrast Module webserver (Whitten et al., 2008) to be $2.72 \times 10^{10} \text{ cm}^{-2}$ for both 186 CI and 186 CI^{E146K}. The partial specific volume was calculated using the NucProt webserver (Voss and Gerstein, 2005) to be 0.7432 mL/g. $(\Delta\rho_M)^2$ was calculated to be $4.08 \times 10^{20} \text{ g}^{-2}\text{cm}^2$. The estimated molecular masses for CI and CI^{E146K} are shown in Figure 3-8. The molecular mass estimated for CI^{E146K} and CI at the highest concentration are 40 kDa and 188 kDa, respectively. This was smaller than the expected molecular mass of 45.2 kDa for CI^{E146K} and 316 kDa (14mer) for CI.



	Concentration (mg/mL)	Relative I(0)	Absolute I(0) (cm ⁻¹)	Estimated MW from I(0) (Da)	Equivalent monomer units
CI(E146K)	7.4	9752.6	0.027156	40047.61	1.77
	3.7	9011.41	0.025092	37004.02	1.64
	1.85	8756.36	0.024382	35956.69	1.59
	0.93	8015.75	0.02232	32915.49	1.46
	0.46	7497.2	0.020876	30786.14	1.36
CI	8.7	45865.2	0.127711	188338.6	8.34
	4.4	41519.5	0.115611	170493.7	7.55
	2.2	35096.4	0.097725	144118.2	6.37
	1.1	28260.8	0.078692	116048.8	5.13
	0.5	26155.5	0.07283	107403.7	4.75

Figure 3-8 – Molecular weight determination of CI and CI^{E146K} using Guinier analysis. Guinier plots were made for CI and CI^{E146K} to estimate $I(0)$ at each of the concentrations. $I(0)$ was converted into absolute units using X-ray scattering by water as a calibrant. Absolute $I(0)$ was used to calculate the molecular mass of the protein using the formula $MM = \frac{N_A I(0)_{Absolute}}{\Delta\rho_M^2}$ where N_A is Avogadro's number and $\Delta\rho_M$ is the scattering length difference. Errors in molecular mass were estimated at 10%, based on (2007). A table of calculated values during various steps of MW estimation using $I(0)$ is provided.

3.10.3 Pair distance-distribution function $P(r)$

The pair distance-distribution function, also called the pair-density distribution function, pair distribution function or $P(r)$ function, is the distribution of paired-sets of distances between all points within a scattering object weighted by their respective electron densities. In SAXS, the $P(r)$ function describes the paired set of distances between all the electrons in a macromolecule or complex. The $P(r)$ function is generated via a Fourier transform of the SAXS scattering profile $I(q)$ (Putnam et al., 2007; Rambo, 2018). The $P(r)$ distribution can provide structural details of the protein. In theory, the $P(r)$ distribution is zero at $r=0$ and $r \geq D_{max}$, where D_{max} is the largest

linear dimension in the scattering particle. In addition, the shape of the distribution is dependent on the structure of the macromolecule (Figure 3-9) and can be sometimes be used as an initial indication of overall shape.

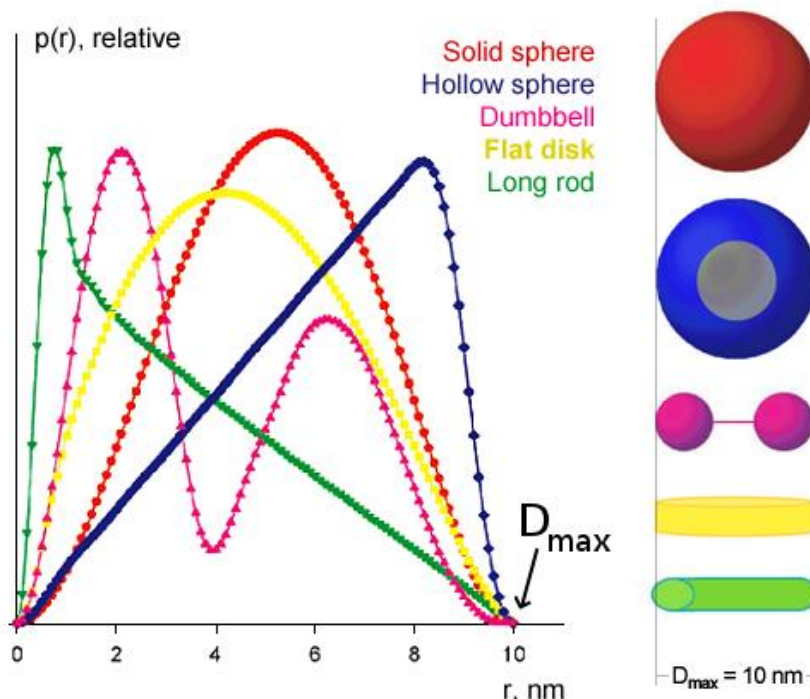


Figure 3-9 – Pair distance-distribution functions $P(r)$ of different geometric bodies. Figure adapted from Svergun and Koch (2003).

In addition, $P(r)$ distributions can be used to calculate R_g and $I(0)$ values for SAXS datasets with data that lies outside the Guinier approximation and can be used to cross-check R_g and $I(0)$ values obtained from a Guinier plot. In practice, an indirect Fourier transform is used to generate $P(r)$ based on a user-entered D_{max} value. $P(r)$ plots are generated through trial-and-error, where the user selects different D_{max} values and visually assesses the $P(r)$ distribution, and the fit to the scattering data, to determine if the chosen D_{max} is appropriate. $P(r)$ distributions were generated for both CI and CI^{E146K} using the GNOM program and the ATSAS package (Franke et al., 2017), and R_g and $I(0)$ were extracted from the plots (Figure 3-10). R_g and $I(0)$ values obtained from $P(r)$ distributions and Guinier analysis are very close, with a deviation of less than 2% in all cases (Figure 3-11). This cross-check shows that the Guinier approximation used was appropriate.

The D_{max} of CI^{E146K} at the concentrations tested were between 9 and 11 nm. However, the maximum linear dimension of the CI^{E146K} dimer in the crystal structure is only 7 nm (Figure 3-10B). This suggests that the dimer conformation of CI in solution may be less compact than the crystal structure due to the flexible linker between the NTD and CTD of CI allowing the NTD to

extend. D_{\max} , R_g and $I(0)$ for CI increase with concentration, consistent with the idea that increasing higher order oligomers form with increasing concentration.

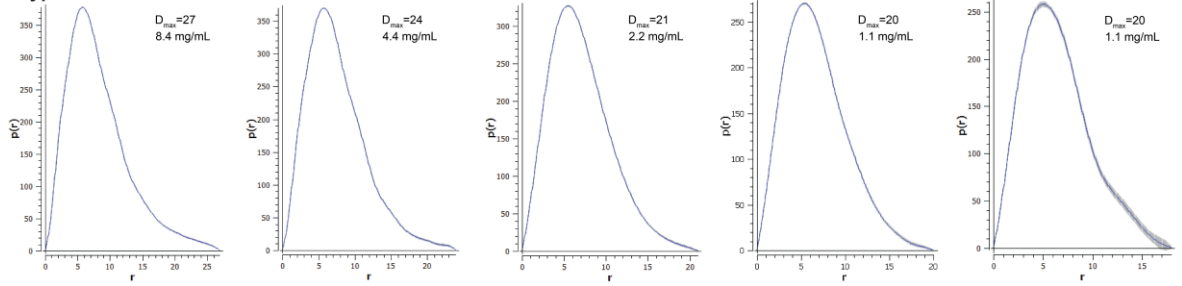
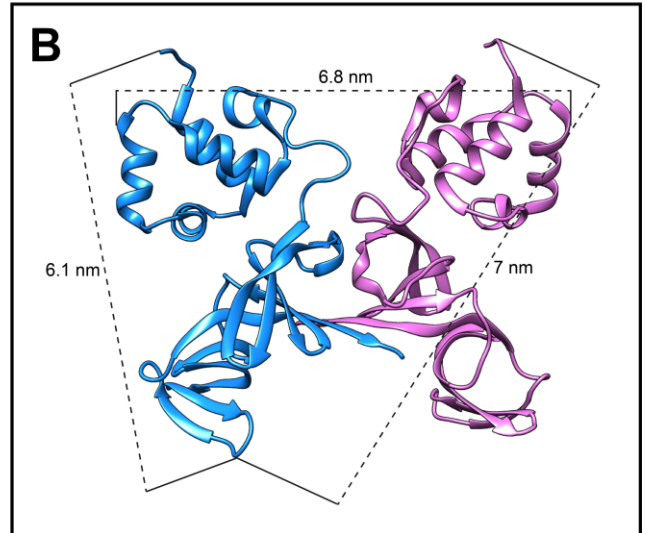
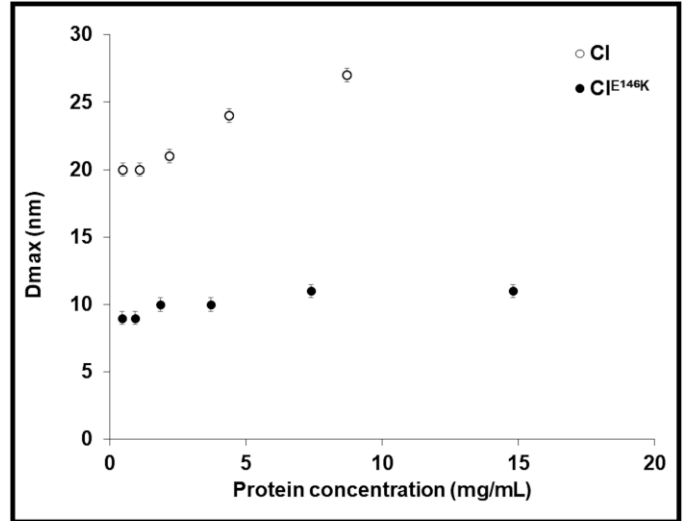
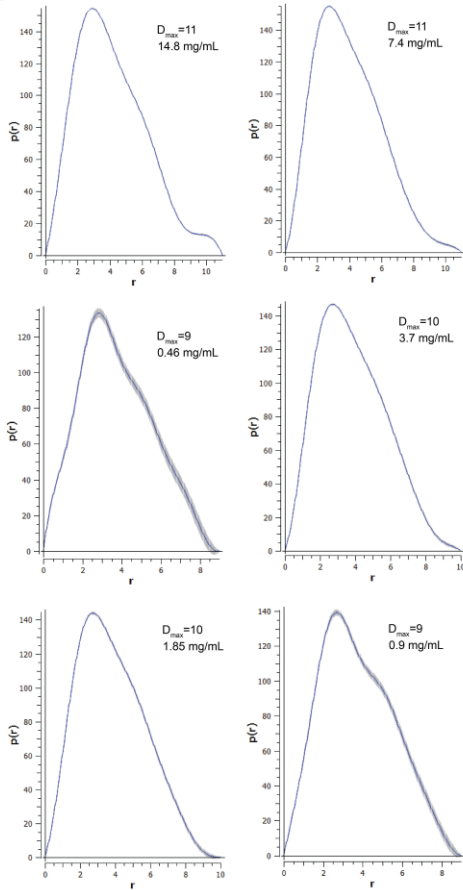
A**Wildtype CI****CI^{E146K}**

Figure 3-10 – Pair distance distribution functions of CI and CI^{E146K}. (A) $P(r)$ distributions for CI and CI^{E146K}. Different D_{max} values were used to generate $P(r)$ distributions for CI and CI^{E146K} using the GNOM program of the ATSAS package (Franke et al., 2017). The $P(r)$ distributions that visually fit the data and the corresponding D_{max} value are shown. (B) The dimensions of the CI^{E146K} protein in the crystal structure are smaller than the D_{max} values determined from $P(r)$ distributions, with the largest atom to atom distance at 7 nm.

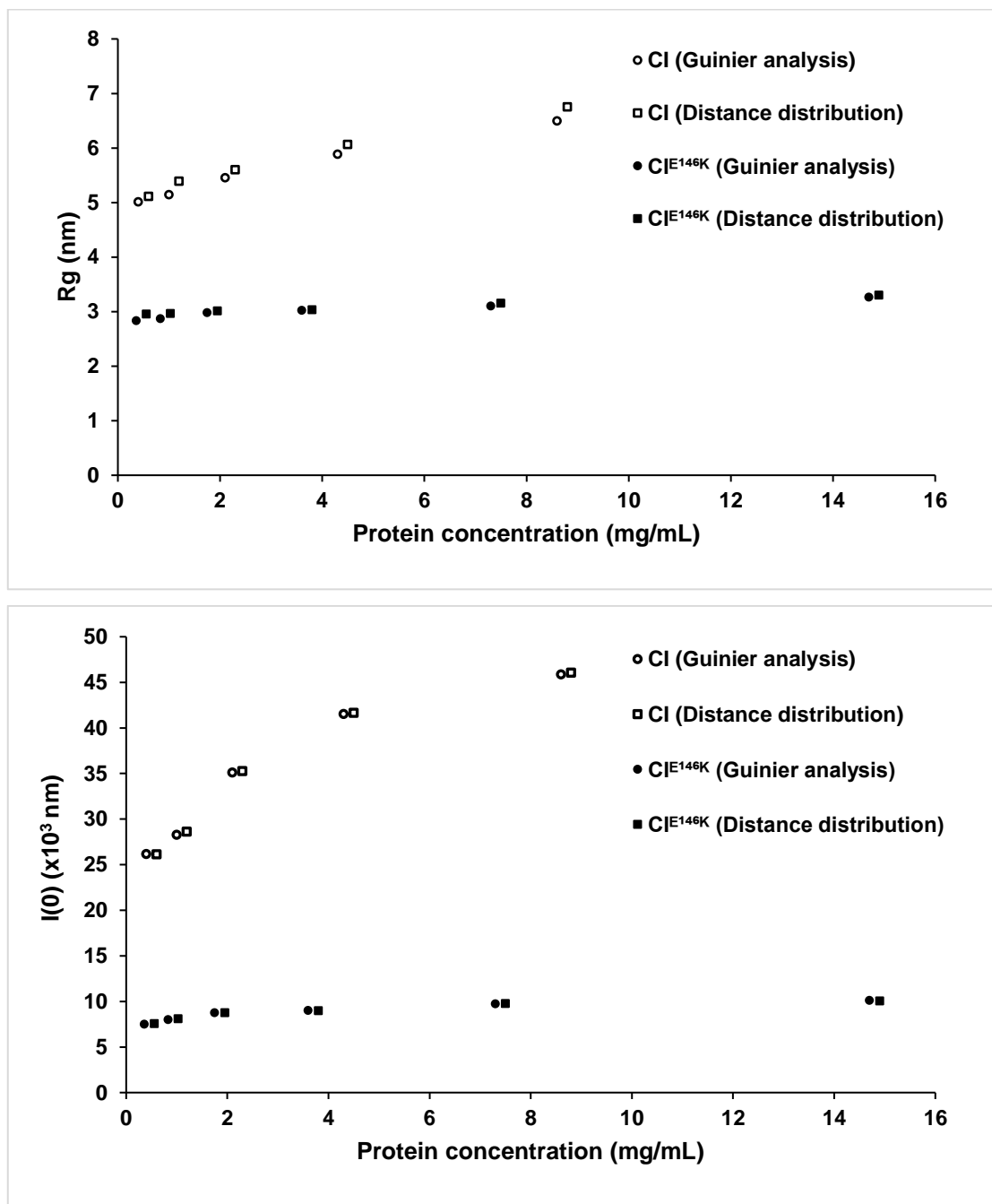


Figure 3-11 Guinier analysis and $P(r)$ distribution analysis for 186 CI and 186 CIE^{146K} give radius of gyration R_g and intensity of zero angle scattering $I(0)$ values that are consistent with each other at all concentrations tested.

3.10.4 Ensemble optimization method to examine flexibility in CIE^{146K}

Given a structural model, it is possible to generate the theoretical scattering pattern. This theoretical scattering pattern would account for orientation averaging effects in solution. However, for flexible systems, the theoretical scattering pattern would also need to account for conformational averaging. Each new conformation in the sample adds to the scattering pattern, such that the final observed scattering is a linear combination of scattering from its constituent conformations. Given K distinct conformations, the SAXS intensity is given by $I(q) = \sum_{k=1}^K v_k I_k(s)$, where v_k and I_k are the fractional proportion and scattering intensity of the k^{th} conformation (Tria et al., 2015).

One method to account for this conformational flexibility is known as the Ensemble Optimization Method (EOM) (Bernadó et al., 2007; Tria et al., 2015). In EOM, a large pool of random conformers is generated to approximate conformational space. The scattering profile for each conformer is calculated and a genetic algorithm is used to select the conformations that best fit the experimental scattering. The algorithm aims to minimize discrepancy χ^2 (Bernadó et al., 2007; Tria et al., 2015), given by

$$\chi^2 = \frac{1}{K-1} \sum_{j=1}^K \left[\frac{\mu I(q_j) - I_{exp}(q_j)}{\sigma(q_j)} \right]^2 \text{ Equation 5}$$

where K is the number of conformations, $I_{exp}(q)$ is the experimental scattering, μ is a scaling factor and σ are the standard deviations. Discrepancy (χ^2) can also be used as a goodness of fit metric to compare the fits between different modelling methods.

Several structural models were fitted to the CI^{E146K} SAXS scattering curve (Figure 3-12). The theoretical scattering profiles for the CI^{E146K} crystal structure (PDB ID 2FJR) and CI-hybrid crystal structure (Chapter 4) were generated and compared to the observed scattering curve of CI^{E146K}. A significantly better fit was obtained by fitting a rigid body model of CI^{E146K}, made by translating the NTD and CTD from the CI^{E146K} crystal structure relative to each other, with linkers represented as interconnected dummy residues. The model that better fits the data, as judged by χ^2 , has a reoriented NTD (Figure 3-12A). In all these cases, the scattering profile assumes only one scattering species. EOM 2.0 was also employed to generate an ensemble of scattering conformers of CI^{E146K} to fit the data. The final ensemble selected by the algorithm consists of five conformers (Figure 3-12B). The conformers differed in the position of the NTDs relative to the CTD dimer, demonstrating the flexibility of the linker.

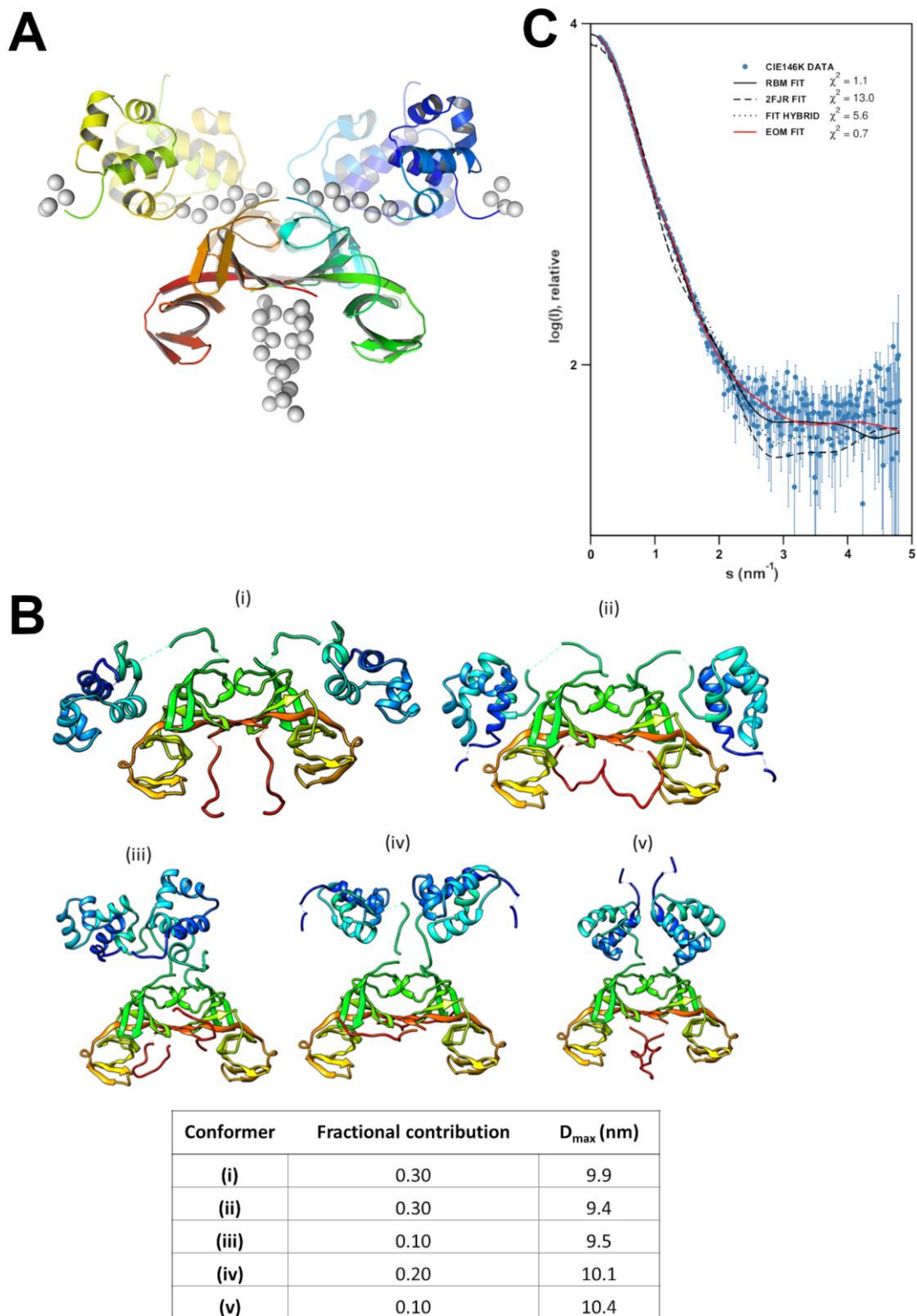


Figure 3-12 – Fitting structural models to CIE^{E146K} SAXS data. (A) A rigid body model of CIE^{E146K}, modelled to fit the CIE^{E146K} SAXS data. The flexible termini and linkers between the domains are represented by dummy atoms. The CIE^{E146K} crystal structure that was used as a template for rigid body modelling is superimposed and shown as a slightly transparent structure. (B) Ensemble of CIE^{E146K} dimer conformers fitted to the SAXS scattering profile using the EOM 2.0 algorithm. Each conformer is assigned a fractional contribution, which corresponds to the predicted fraction of CIE^{E146K} dimers that occupy this conformation at any point in time. (C) The theoretical scattering patterns of the structural models compared to the experimental scattering profile. χ^2 is used as a goodness of fit metric, with lower values indicating a closer fit. The EOM structural models, followed by the rigid body model, provided the closest fit to the data.

3.11 Discussion of SAXS data

SAXS was a useful tool for studying 186 CI. Crystallographic studies of the CI protein described in Pinkett et al. (2006) have provided atomic level structural detail on the domains of the protein and have suggested a possible 14-mer assembly that could be adopted by the protein. However, non-biologically relevant conformations and oligomers can form due to packing of proteins into a solid crystal. SAXS provides a complementary source of structural information on the CI protein in the solution state, avoiding the effects of crystal packing on the protein quaternary structure. SAXS is a lower resolution technique which can be used to derive global shape and conformation, which when combined with high resolution structural data can produce detailed structural models (Putnam et al., 2007).

The adoption of a higher-order quaternary structure by 186 CI dimers is thought to be crucial to its ability to transcriptionally regulate lytic and lysogenic gene expression in the phage lifecycle (Dodd and Egan, 2002; Wang et al., 2013). The E146K mutation in CI abolishes higher order oligomer formation, as the X-ray structure of the protein revealed a dimeric structure. This is supported by SAXS data which shows that the apparent molecular weight and shape of CI^{E146K} as indicated by the Guinier R_g and maximum linear dimension D_{max} parameters, remain unchanged with increasing concentration. This reflects that CI^{E146K} forms a dimer, which does not associate further. The estimated molecular mass was 40 ± 4 kDa, very close to the sequence calculated value of 45.2 kDa. In contrast, wild type CI shows an overall increase in global parameters correlated with increasing concentration, which may reflect the equilibrium shifting towards higher order oligomers as concentration increases.

The molecular weight for CI at the highest concentration tested was estimated to be 188 ± 18 kDa. This is smaller than the estimated molecular mass for the 14-mer complex at 316 kDa. As the high order oligomerization of CI is concentration dependent, it is possible that the concentrations tested were too low to observe the 14mer complex. There are several inherent uncertainties in the measurements and calculations that could also have caused this discrepancy. The molecular weight calculation relies on protein concentration, partial specific volume derived from the primary structure and the experimentally determined $I(0)$ value. Errors in these values would propagate into the mass determination. The concentration of protein was calculated from near UV-absorbance using native protein based on an extinction coefficient calculated from the primary sequence. This extinction coefficient implicitly assumes that the protein is denatured and its extinction coefficient is the linear combination of the major chromophores phenylalanine, tyrosine and tryptophan. In Appendix B, we experimentally determine the native extinction coefficient to be $3 \pm 5\%$ larger than the denatured extinction coefficient. The estimated molecular weight adjusted for concentration effects is 194 ± 30 kDa,

which still is not consistent with a 14mer. Mylonas and Svergun (2007) estimated the molecular mass of 13 proteins using $I(0)$ and compared them to their true values. For molecular masses that were calculated using the partial specific volume derived from the primary sequence, the average deviation from the true value was 16.5%, with 5 of the 13 proteins having a deviation of greater than 20%. It was pointed out by the authors that particle specific volumes calculated from the primary sequence, as used in our calculations, are often significantly smaller than their experimentally measured values. Additionally, if we assume that wild type CI can form an oligomeric assembly and does so dynamically, as implied by the progressive increase in particle shape and size parameters, it is possible that the scattering particles present a weighted averaged MW for a range of different oligomeric states (Graziano et al., 2006). That the estimated molecular mass of the CI complex isn't consistent with the 14mer is therefore unsurprising.

The SAXS data has reaffirmed the flexibility of the CI protein. Kratky analysis of the data qualitatively indicates flexibility in both the CI^{E146K} and wild type CI. EOM modelling of CI^{E146K} can be used to model this flexibility, by modelling it as an ensemble of proteins in different flexed states. This model contained five constituent states with the NTD in various positions relative to the CTD. The NTDs of the CI^{E146K} protein in the crystal structure are positioned in the same plane as the CI wheel. For DNA to wrap around a CI wheel, each dimer's NTDs need to be positioned roughly orthogonal to the wheel plane. In support of the CI wheel model, conformations iii—v of the EOM model position the NTDs in this CTD orthogonal position such that DNA would indeed be able to wrap around. Hypothetically, the mobility of the NTD defined by the ensemble may describe dynamic regulation of DNA binding, such that the predominant dimeric species have the NTD DNA binding interface 'hidden' by the CTD oligomeric domain, and wheel formation may switch the NTD to a DNA binding accessible state, as reflected in the relative percentage contribution (60%) of scattering by the 'hidden' NTD states i-ii in the CI^{E146K} dimer (note that the dimer mutation is in the oligomeric interface of the CTD, not at the NTD/CTD interface).

Neither SAXS nor the previous techniques employed to interrogate the CI complex (AUC; (Shearwin et al., 2002); AFM, SEC-MALLS; (Wang et al., 2013)) are of sufficient resolution to definitively demonstrate the presence of a 14mer oligomeric complex that is central to the current model of 186 CI transcriptional regulation. To tackle this problem, native ion-mobility mass-spectrometry was used to directly measure the mass of the 186 CI complex with high resolution, to conclusively determine the true oligomeric state *in vitro*.

3.12 Ion-mobility mass spectrometry

Ion-mobility mass spectrometry (IM-MS) couples ion mobility separation (IM) with mass spectrometry (MS) to allow the masses and topology of protein complexes to be simultaneously

determined. An IM-MS instrument performs four functions – sample ionisation, IM separation, mass-to-charge separation and ion detection. To study 186 CI oligomeric state, IM-MS measurements on the protein were performed on the Synapt HDMS system (Waters, Manchester, UK) (Figure 3-13). The instrumentation has been described in detail elsewhere (Pringle et al., 2007), and its main functions will be described below.

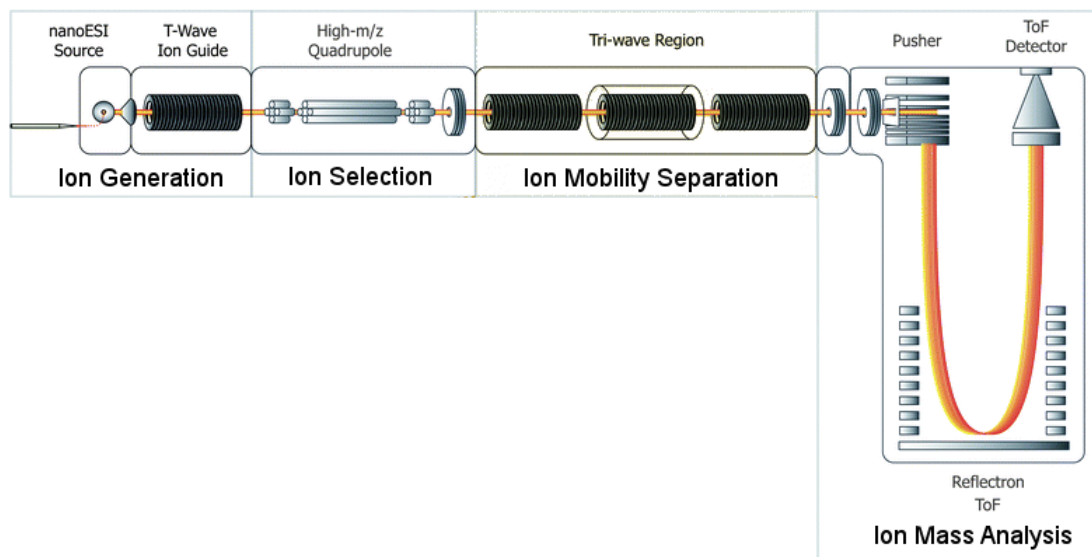


Figure 3-13 – Schematic diagram of the Waters Synapt HDMS system, displaying the sequential order of equipment required for ionization, ion selection, separation based on ion mobility and mass detection. Figure adapted from (Zhong et al., 2011).

Sample ionisation

The Synapt HDMS instruments ionises protein samples by electrospray ionisation (ESI) (Dole et al., 1968; Fenn et al., 1989). In this technique, the liquid sample is passed through a needle. An applied electric field results in the needle tip charging the surface of the liquid as it exits. Coulomb repulsion causes the liquid to disperse and a drying gas is applied to encourage evaporation. The net result is the production of gas phase ions (Fenn et al., 1989). By using volatile aqueous buffers (most commonly ammonium acetate) and physiological pHs, large non-volatile biomolecules can be ionized with their native-like structures maintained.

Ion-mobility separation

Ion mobility separation involves passing ions through a chamber of neutral carrier gas at a controlled pressure in the presence of an external field. The mobility of the ion through the gas depends on the collisional cross-sectional area (CCS) and charge of the ion. Less compact ions collide more frequently with neutral gas molecules and hence take longer to travel through the chamber. Thus, ion mobility measurements provide information about protein shape. The drift time, the time it takes for the ions to traverse the chamber, is measured and can be converted to

CCS area using a standard curve of CCS area against drift time, made using calibrants of known CCS area (Ruotolo et al., 2008).

Mass-to-charge separation

The Synapt HDMS system uses the time of flight (ToF) analyser. When samples enter the mass analyser compartment, they are pushed by an electric potential from the pusher. The time t it takes for a sample ion to traverse the time-of-flight tube is given by $t = d \sqrt{\frac{m}{2Vz}}$, where m is the mass of the ion, z is the charge of the ion, d is the distance traversed and V is the electric potential of the pusher. With both distance and electric potential constant, the time of flight of an ion to reach the detector can be used to determine its mass-to-charge ratio (Lane, 2005). For proteins and protein complexes, a series of multiply charged peaks is often observed for each species, allowing for deconvolution of the spectra to determine the molecular weights of the species.

Computational calculation of CCS

The CCS of a known or modelled structure can be computationally calculated. In our studies, CCS was computationally calculated from structural models using the Leeds method (Smith et al., 2009). This method employs the Projection Approximation Algorithm. The ions cross-section is replaced with its projection (or shadow) which is averaged over all the possible orientations of the ion (Knapman et al., 2010). This approximates the ion as a sphere with this average projection. The collisions of this model sphere with buffer gas is simulated *in silico* to calculate the theoretical CCS.

Mass selection and collision induced dissociation

The Synapt HDMS instruments contains a quadrupole analyser and a collision cell. The quadrupole consists of four parallel rods. To select for ions with specific m/z , alternating radiofrequency (RF) potentials and fixed direct current (DC) are applied to these rods, creating an electric field, to oscillate the ions. With a defined RF potential and DC, only ions with one defined m/z will have a stable trajectory and pass through the quadrupole. All the other ions will be ejected from the quadrupole. In MS mode, no ion-selection is employed, and the quadrupole acts as an ion guide. In MS/MS mode, the quadrupole selects for ions of a particular m/z ratio. These ions can be activated in the collision cell. In the collision cell, the ions collide with an inert gas to excite the ions and cause collision induced dissociation (CID).

3.13 Mass spectrometry experimental procedures

3.13.1 Preparation of samples for mass spectrometry

The 186 CI was overexpressed in *E. coli* and purified as outlined in sections 3.2 and 3.3, respectively. The 186 CI C-terminal domain was overexpressed and purified as described by Pinkett et al. (2006). All proteins were exhaustively dialysed into 0.3 M ammonium acetate at 4°C, flash frozen into liquid nitrogen and stored at -80°C.

3.13.2 Mass-spectrometry examination of 186 CI-CTD and 186 CI

Mass-spectrometry experiments were performed on a Synapt HDMS system (Waters, Manchester, UK). Ions were produced by electrospray ionisation (ESI) from sample introduced from gold-coated borosilicate capillaries prepared in-house (Hernández and Robinson, 2007). The instrument parameters were optimized to remove adducts while preserving noncovalent interactions and were typically as follows: capillary voltage, 1.5 kV; cone voltage, 50 V; trap collision energy, 10 V; source temperature, 20°C; and backing pressure, 6 mBar.

Native mass spectrometry was performed on 186 CI-CTD, using exactly the same protein construct used for crystallography by Pinkett et al. (2006). The concentrations 13 μ M and 34 μ M were used and the spectra are shown in Figure 3-14. At both concentrations, the major species detected were 10mer and 12mer species. There was no evidence for a 14mer species. To unequivocally assign these high molecular weight species, CID experiments were performed to fragment the 10mer and 12mer species. The mass of the fragmented ions can be determined from the MS/MS spectra and will sum up to give the mass of the parent ion. CID of the 5689 m/z 10mer species caused it monomer to be ejected from the ion to leave a 9mer, both of which were detected on the MS/MS spectra (Figure 3-15A). CID of the 5963 and 6075 m/z 12mer species fragmented the ion into a 11mer and monomer, both of which were clearly observed on the MS/MS spectra (Figure 3-15B and C).

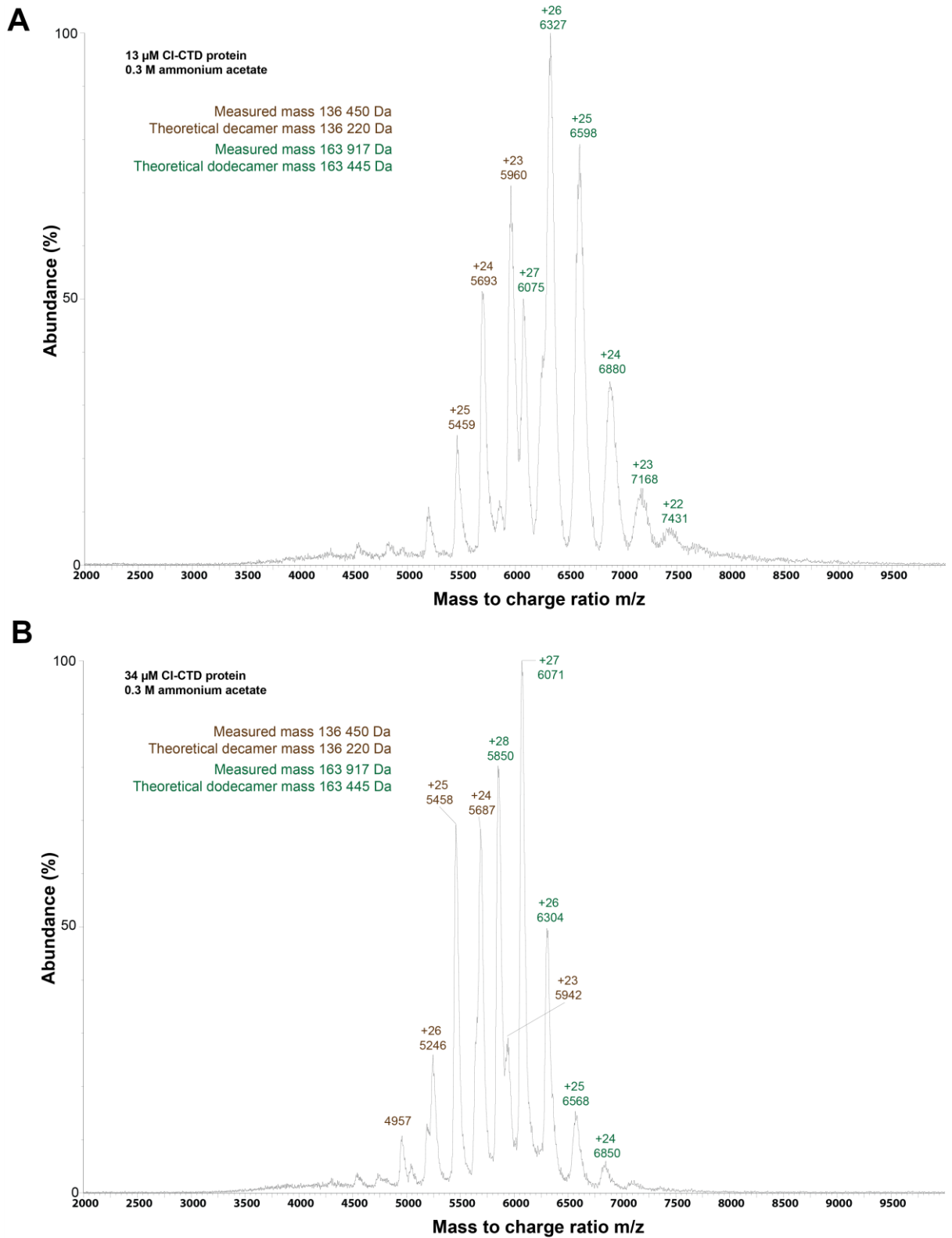


Figure 3-14 - ESI-TOF spectra of 186 CI-CTD at (A) 13 μM and (B) 34 μM was measured in aqueous ammonium acetate (0.3 M). Decamers (10mer) and dodecamers (12mer) of 186 CI-CTD were detected.

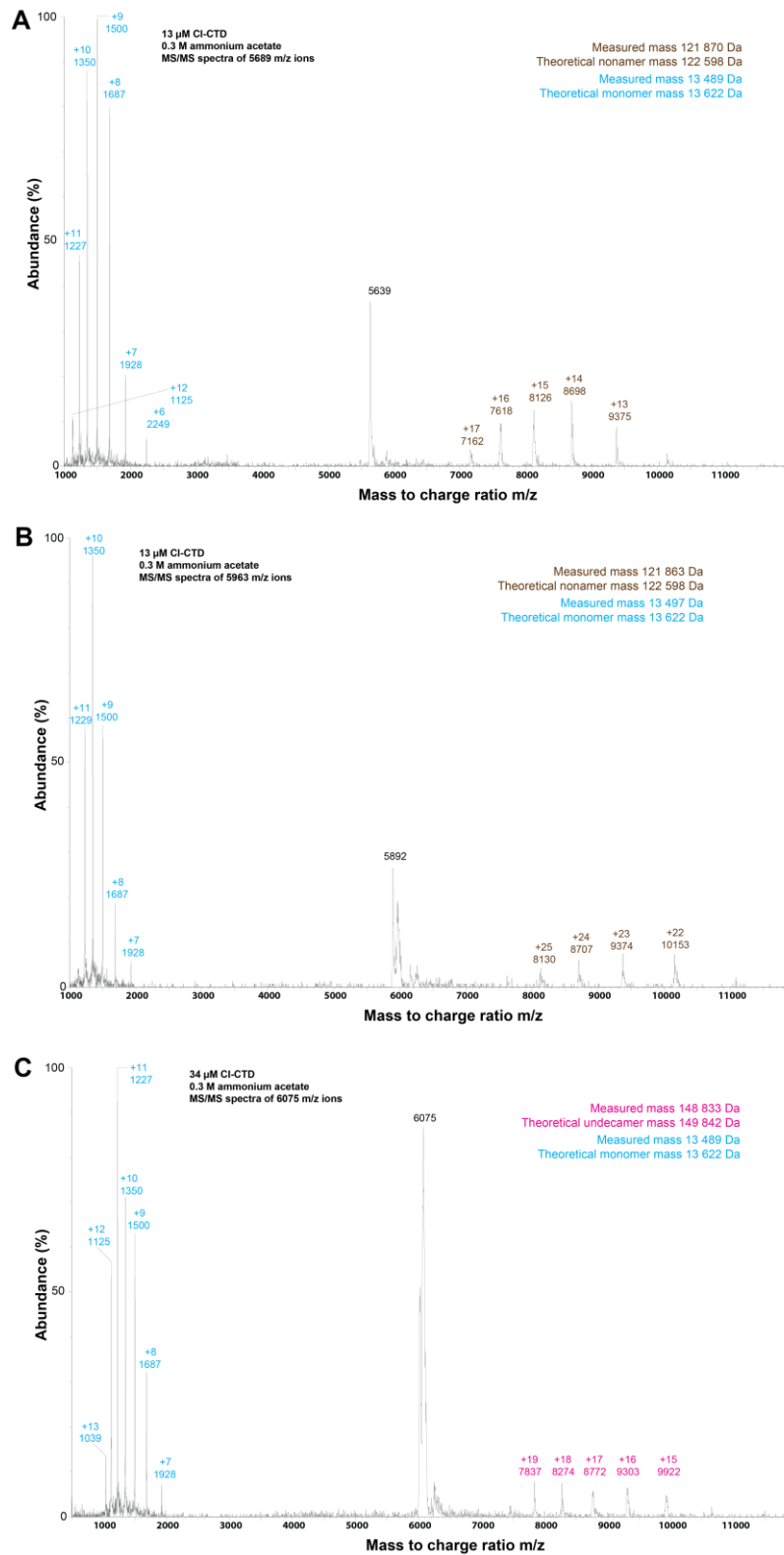


Figure 3-15 – Ionized species corresponding to decamers (10mer) and dodecamers (12mer) of 186 CI-CTD, detected via ESI-MS in Figure 3-14, were subjected to collision induced dissociation (CID) to confirm their oligomeric assignment. The MS/MS spectra of the CID of the 5689 and 5963 m/z ions are shown in (A) and (B), respectively. Both showed the presence of a nonamer and monomer, confirming the presence of a decamer in the native MS spectra. The MS/MS spectra of the collision induced dissociation of the 6075 m/z ions (C) showed the presence of an undecamer (11mer) and monomer, confirming the presence of a dodecamer in the native MS spectra.

Given the unexpected result of the dodecamer as the highest molecular weight oligomeric species observed for CI-CTD, rather than the 14-mer seen in the crystal structure, mass spectrometry was also used to investigate the oligomeric state of full length 186 CI. Mass spectrometry was conducted at 10 μM and 30 μM , with mass spectra shown in Figure 3-16. Dimer, tetramer, hexamer, octamer and 10mer species could be assigned at 10 μM . A very obvious 12mer species was detected at 30 μM . However, there is no evidence of a species above a 12mer in either spectrum. To validate this result, CID experiments were performed on the 7632 m/z ion that was assigned to the 12mer (Figure 3-17). This ion fragmented into a 11mer and monomer to confirm the assignment was correct.

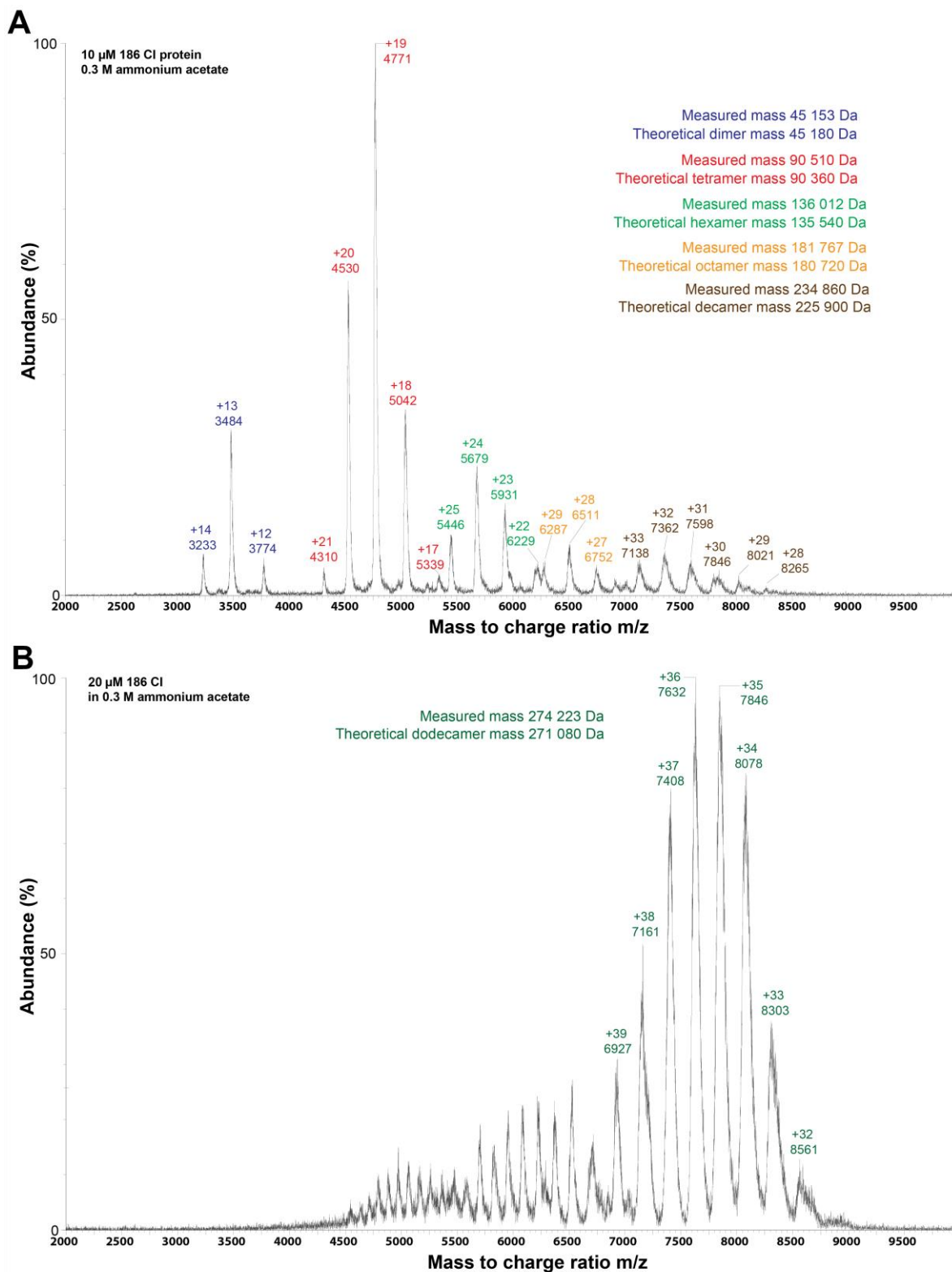


Figure 3-16 - ESI-TOF spectra of 186 Cl at (A) 10 μM and (B) 20 μM was measured in aqueous ammonium acetate (0.3 M). 186 Cl forms dimers that sequentially oligomerize into a dodecamer (12mer) species. Dimers, tetramers, hexamers, octamers, decamers and dodecamers are detected. The proportion of dodecamers increases at the higher protein concentration.

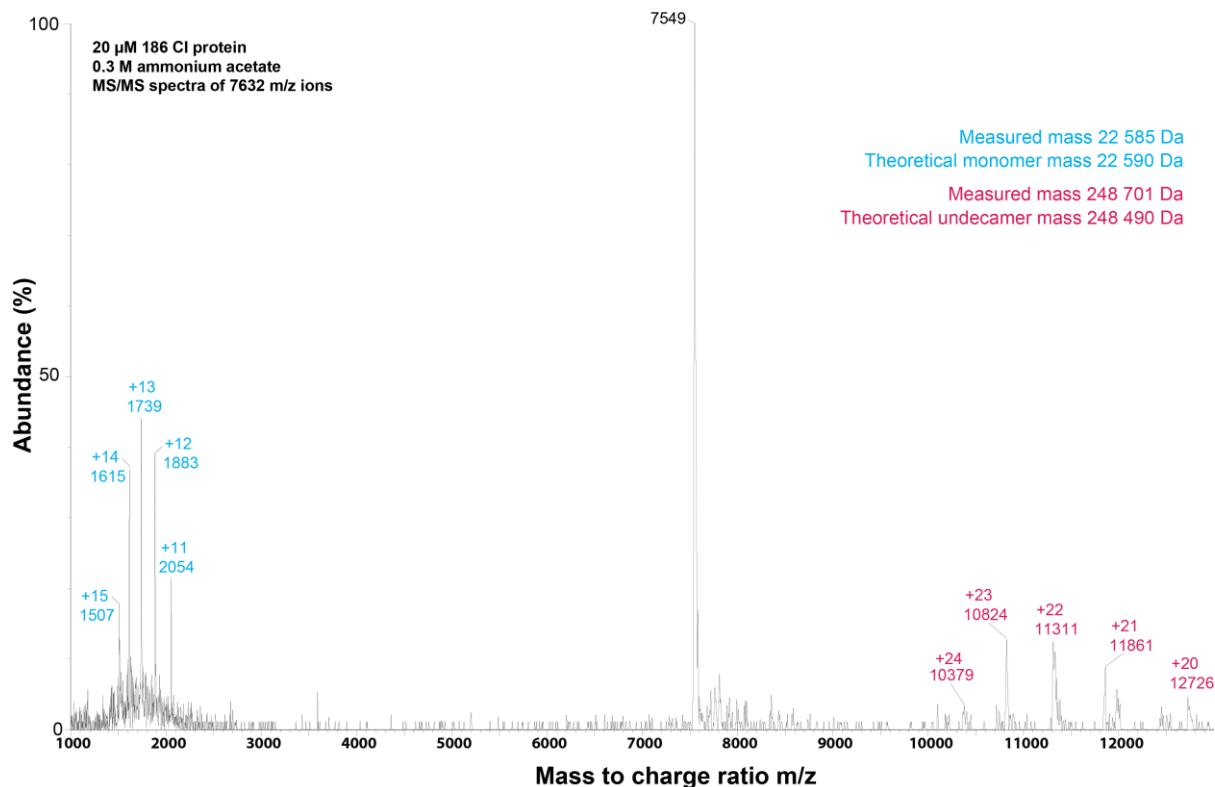


Figure 3-17 – Ionized species corresponding to dodecamers (12mer) of 186 CI, detected via ESI-MS in Figure 3-16B, were subjected to collision induced dissociation (CID) to confirm their oligomeric assignment. The MS/MS spectra of the CID of the 7632 m/z ions showed the presence of undecamer (11mer) and monomer, confirming the presence of a dodecamer in the native MS spectra.

3.14 Oligomerization of 186 CI

Measuring the oligomerization state of 186 CI shows that the protein can form dimers, tetramers, hexamers, octamers, decamers and dodecamers. This oligomerization pattern suggests that the protein forms dimers that sequentially oligomerise until it reaches a final dodecamer state. This higher order oligomerization is mediated via the CTD, as the isolated CTD also forms decamers and dodecamers. Taken together, it is likely that the true final oligomeric state of 186 CI, mediated by the 186 CI-CTD, is a dodecamer. In all the spectra, no 14mer species was observed. The 14mer complex observed in the crystal structure of the CI-CTD could be a crystal artefact and not reflective of its true quaternary structure. It could also be argued that CI concentrations are too low to observe a 14mer. However, the highest concentrations of CI (20 μM) tested exceed the concentration found in a 186 lysogen, which is estimated around 1.1 μM (Dodd and Egan, 2002).

3.14.1 Previous studies of 186 CI oligomeric state

There have been several past experiments into CI oligomerization aside from X-ray crystallography. CI has been visualized directly through atomic force microscopy (AFM) experiments. The oligomeric state of CI has also been inferred from size exclusion chromatography combined with multi-angle laser light scattering (SEC-MALLS) and

sedimentation equilibrium experiments. In this section, this data is examined to see if it supports the 12-mer or 14-mer arrangement.

Wang *et al.* (2013) performed AFM on 186 CI in the presence and absence of DNA. The resolution of the images was sufficient to observe 186 DNA interacting with CI oligomeric particles to form specific CI-DNA looped species. This supported the hypothesis that 186 CI exerts its transcriptional regulatory activity by wrapping of 186 DNA onto it. AFM also allowed the measurement of particle volume, which was used to infer the molecular weight of the particle using a calibration curve. The observed volume of $632 \pm 152 \text{ nm}^3$ encompasses both the 12mer and 14mer arrangement (Figure 3-18). The 14mer deviates less from the standard curve and seems to fit the data better. However, it's also difficult to judge the reliability of extrapolating the standard curve over 150 kDa past its largest calibrant. In the same publication, SEC-MALS was performed on 186 CI and 186 CI HTH⁻, measuring weighted average molecular weight peaks $280 \pm 8 \text{ kDa}$ and $274 \pm 5 \text{ kDa}$, respectively. This corresponds to 12.4 ± 0.4 monomers and 12.1 ± 0.2 monomers for 186 CI and 186 CI HTH⁻, respectively. Thus, the SEC-MALS data supports 186 CI forming a 12mer over a 14mer. SAXS measurements of 186 CI obtained a highest average MW of 188 kDa, equivalent to 8.34 monomers, which is significantly smaller than both the 12mer (271 kDa) and 14mer (316 kDa). This discrepancy could be due to inaccuracies in the partial specific volume parameter used in the calculations, as explained in section 3.11.

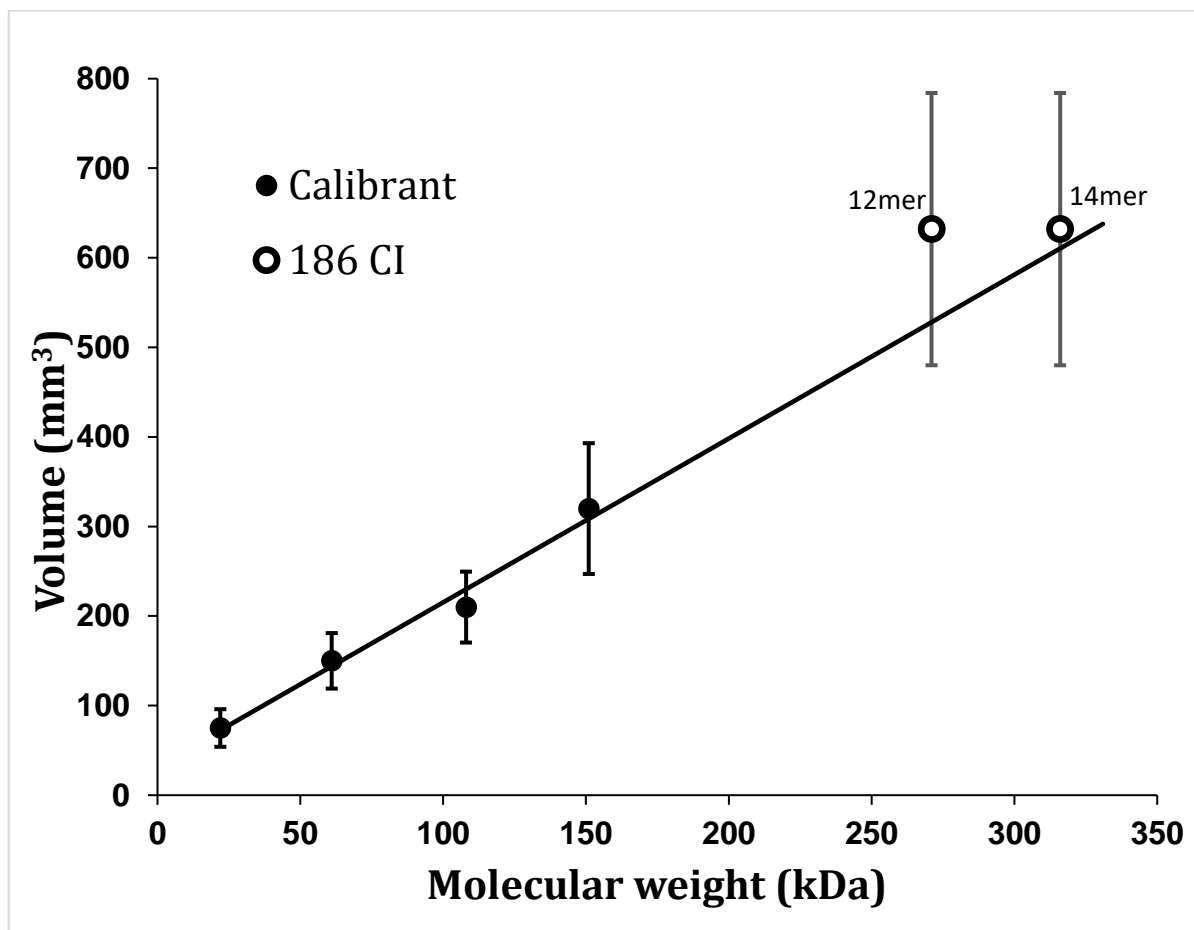


Figure 3-18 – Volume calibration curve for AFM. The molecular weight of 186 CI can be inferred from its volume using a volume vs. molecular weight standard curve. Calibrant proteins (solid points) used from left to right are λ CI monomer (25 kDa), λ CI dimer (50 kDa), histone octamer (108 kDa) and the lac repressor tetramer (150 kDa). 186 CI (unfilled points) were measured to have a volume of $632 \pm 152 \text{ nm}^3$ and plotted with the theoretical molecular weight of a 12mer and 14mer. The chart was based on Figure S1 of Wang *et al.* (Wang *et al.*, 2013).

Sedimentation equilibrium was conducted on his-tagged 186 CI NTD (residues 1-82) and his-tagged CI CTD (residues 83-192) (Pinkett, 2004; Shearwin *et al.*, 2002). Both experiments suggested that 186 CI-CTD and full length 186 CI self-associated into a species 7 to 8-fold larger than the monomer. The self-association in solution was proposed to be through an equilibrium of monomers, dimers, tetramers and octamers (Shearwin *et al.*, 2002). As pointed out by Pinkett *et al.* (Pinkett *et al.*, 2006), the sedimentation equilibrium data could be fitted to higher order species but would require data to be collected at higher concentrations to justify their inclusion.

In summary, native mass spectrometry, with its high mass resolution, has determined with high confidence that the 12mer is the largest oligomeric species formed by 186 CI. Results of previous experiments examining oligomerization of 186 CI, with the exception of the crystal structure of 186 CI CTD, did not provide strong support for the 14mer over the 12mer complex.

3.14.2 An alternative crystallography approach to study 186 CI-CTD oligomerization

The possibility of a 186 CI 12mer complex presents some new questions. What could this 186 CI 12mer look like and how might it achieve regulation? To investigate this, it would be useful to solve the crystal structure showing 186 CI-CTD higher order oligomerization in the presence of

an N-terminal DNA binding domain. The N-terminal domain may provide steric limitations to alter the oligomerization of the protein. However, previous endeavours (Pinkett, 2004) and our attempts to solve the structure of the full CI complex have failed (section 3.6). One of the possible issues with crystallizing the full length 186 CI as a complex is that the protein is very flexible, as revealed by Kratky analysis of SAXS data (section 3.10). This flexibility, likely arising from the flexible linker and the inherent flexibility in the NTD (Pinkett et al., 2006; Shearwin et al., 2002), would reduce crystallization propensity of the protein. One possible approach to improving crystallizability is to remove the long linker between the domains and replace the semi-flexible 186 CI NTD with a more rigid domain, which is attempted in the next chapter.

3.15 References

- Aragão, D., Aishima, J., Cherukuvada, H., Clarken, R., Clift, M., Cowieson, N.P., Ericsson, D.J., Gee, C.L., Macedo, S., Mudie, N., Panjekar, S., Price, J.R., Riboldi-Tunncliffe, A., Rostan, R., Williamson, R., Caradoc-Davies, T.T., 2018. MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. *J. Synchrotron Radiat.* 25, 885–891.
- Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., Leslie, A.G.W., 2011. iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 271–281.
- Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., Svergun, D.I., 2007. Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* 129, 5656–5664.
- Ciubotaru, M., Koudelka, G.B., 2003. DNA-stimulated assembly of oligomeric bacteriophage 434 repressor: Evidence for cooperative binding by recruitment. *Biochemistry* 42, 4253–4264.
- De Anda, J., Poteete, A.R., Sauer, R.T., 1983. P22 c2 repressor. Domain structure and function. *J. Biol. Chem.* 258, 10536–10542.
- Dodd, I.B., Egan, J.B., 2002. Action at a distance in CI repressor regulation of the bacteriophage 186 genetic switch. *Mol. Microbiol.* 45, 697–710.
- Dodd, I.B., Perkins, A.J., Tsemitsidis, D., Egan, J.B., 2001. Octamerization of λ CI repressor is needed for effective repression of PRM and efficient switching from lysogeny. *Genes Dev.* 15, 3013–3022.
- Dodd, I.B., Shearwin, K.E., Perkins, A.J., Burr, T., Hochschild, A., Egan, J.B., 2004. Cooperativity in long-range gene regulation by the λ CI repressor. *Genes Dev.* 18, 344–354.

- Dole, M., Mack, L.L., Hines, R.L., Chemistry, D.O., Mobley, R.C., Ferguson, L.D., Alice, M.B., 1968. Molecular beams of macroions. *J. Chem. Phys.* 49, 2240–2249.
- Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* 69, 1204–1214.
- Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M., 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science (80-.)*. 246, 64–71.
- Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., Svergun, D.I., 2017. ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* 50, 1212–1225.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb.* 571–607.
- Glatter, O., Kratky, O., 1982. *Small angle X-ray scattering*. Academic press.
- Grant, T.D., Luft, J.R., Carter, L.G., Matsui, T., Weiss, T.M., Martel, A., Snell, E.H., 2015. The accurate assessment of small-angle X-ray scattering data. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 71, 45–56.
- Graziano, V., McGrath, W.J., Yang, L., Mangel, W.F., 2006. SARS CoV main proteinase: The monomer-dimer equilibrium dissociation constant. *Biochemistry* 45, 14632–14641.
- Guinier, A., 1939. La diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques. *Ann. Phys. (Paris)*.
- Hernández, H., Robinson, C. V., 2007. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc.* 2, 715–726.
- Jain, D., Nickels, B.E., Sun, L., Hochschild, A., Darst, S.A., 2004. Structure of a Ternary Transcription Activation Complex. *Mol. Cell* 13, 45–53.
- Kikhney, A.G., Svergun, D.I., 2015. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* 589, 2570–2577.
- Kissinger, C.R., Gehlhaar, D.K., Fogel, D.B., 1999. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 55, 484–491.
- Knapman, T.W., Berryman, J.T., Campuzano, I., Harris, S.A., Ashcroft, A.E., 2010. Considerations in experimental and theoretical collision cross-section measurements of small molecules using travelling wave ion mobility spectrometry-mass spectrometry. *Int. J. Mass Spectrom.* 298, 17–23.

- Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M.H.J., Svergun, D.I., 2003. PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* 36, 1277–1282.
- Lane, C.S., 2005. Mass spectrometry-based proteomics in the life sciences. *Cell. Mol. Life Sci.*
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674.
- McPhillips, T.M., McPhillips, S.E., Chiu, H.J., Cohen, A.E., Deacon, A.M., Ellis, P.J., Garman, E., Gonzalez, A., Sauter, N.K., Phizackerley, R.P., Soltis, S.M., Kuhn, P., 2002. Blu-Ice and the distributed control system: Software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron Radiat.* 9, 401–406.
- Mylonas, E., Svergun, D.I., 2007. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J. Appl. Crystallogr.* 40, 245–249.
- Nickels, B.E., Dove, S.L., Murakami, K.S., Darst, S.A., Hochschild, A., 2002. Protein-protein and protein-DNA interactions of $\sigma 70$ region 4 involved in transcription activation by λ cI. *J. Mol. Biol.* 324, 17–34.
- Orthaber, D., Bergmann, A., Glatter, O., 2000. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *J. Appl. Crystallogr.* 33, 218–225.
- Pedersen, M., Lo Leggio, L., Grossmann, J.G., Larsen, S., Hammer, K., 2008. Identification of Quaternary Structure and Functional Domains of the CI Repressor from Bacteriophage TP901-1. *J. Mol. Biol.* 376, 983–996.
- Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P. V., Svergun, D.I., 2012. New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* 45, 342–350.
- Pinkett, H.W., 2004. Structural Determination of the Phage 186 Repressor Protein. University of Pennsylvania.
- Pinkett, H.W., Shearwin, K.E., Stayrook, S., Dodd, I.B., Burr, T., Hochschild, A., Egan, J.B., Lewis, M., 2006. The structural basis of cooperative regulation at an alternate genetic switch. *Mol. Cell* 21, 605–615.
- Pringle, S.D., Giles, K., Wildgoose, J.L., Williams, J.P., Slade, S.E., Thalassinou, K., Bateman, R.H., Bowers, M.T., Scrivens, J.H., 2007. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass Spectrom.* 261, 1–12.

- Ptashne, M., 2004. A genetic switch 3rd edition, CSHL Press.
- Putnam, C.D., Hammel, M., Hura, G.L., Tainer, J.A., 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: Defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.*
- Rambo, R.P., 2018. The $P(r)$ distribution [WWW Document]. Bioisis. URL <http://www.bioisis.net/tutorial/5> (accessed 2.5.18).
- Rambo, R.P., Tainer, J.A., 2011. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* 95, 559–571.
- Rossmann, M.G., Blow, D.M., 1962. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.* 15, 24–31.
- Ruotolo, B.T., Benesch, J.L.P., Sandercock, A.M., Hyung, S.J., Robinson, C. V., 2008. Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protoc.* 3, 1139–1152.
- SAXIER, 2018. What is Synchrotron Small-angle X-ray scattering? [WWW Document]. URL <https://www.saxier.org/aboutus/saxs.shtml> (accessed 1.19.18).
- Shearwin, K.E., Dodd, I.B., Egan, B.J., 2002. The helix-turn-helix motif of the coliphage 186 immunity repressor binds to two distinct recognition sequences. *J. Biol. Chem.* 277, 3186–3194.
- Skou, S., Gillilan, R.E., Ando, N., 2014. Synchrotron-based small-angle X-ray scattering of proteins in solution. *Nat. Protoc.* 9, 1727–1739.
- Smith, D.P., Knapman, T.W., Campuzano, L., Malham, R.W., Berryman, J.T., Radford, S.E., Ashcroft, A.E., 2009. Deciphering drift time measurements from travelling wave ion mobility spectrometry- Mass spectrometry studies. *Eur. J. Mass Spectrom.* 15, 113–130.
- Svergun, D., Barberato, C., Koch, M.H., 1995. CRY SOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* 28, 768–773.
- Svergun, D.I., 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* 25, 495–503.
- Svergun, D.I., Koch, M.H.J., 2003. Small-angle scattering studies of biological macromolecules in solution. *Reports Prog. Phys.* 66, 1735–1782.
- Tria, G., Mertens, H.D.T., Kachala, M., Svergun, D.I., 2015. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* 2, 207–217.
- Voss, N.R., Gerstein, M., 2005. Calculation of standard atomic volumes for RNA and comparison

- with proteins: RNA is packed more tightly. *J. Mol. Biol.* 346, 477–492.
- Wang, H., Dodd, I.B., Dunlap, D.D., Shearwin, K.E., Finzi, L., 2013. Single molecule analysis of DNA wrapping and looping by a circular 14mer wheel of the bacteriophage 186 CI repressor. *Nucleic Acids Res.* 41, 5746–5756.
- Whitten, A.E., Cai, S., Trehwella, J., 2008. MULCh: Modules for the analysis of small-angle neutron contrast variation data from biomolecular assemblies. *J. Appl. Crystallogr.* 41, 222–226.
- Zhong, Y., Hyung, S.J., Ruotolo, B.T., 2011. Characterizing the resolution and accuracy of a second-generation traveling-wave ion mobility separator for biomolecular ions. *Analyst* 136, 3534–3541.

4 EXPRESSION, PURIFICATION, CRYSTALLIZATION AND STRUCTURE SOLUTION OF A HYBRID CI PROTEIN

4.1 Introduction

In the previous chapter, native mass spectrometry analysis of the full length 186 CI suggested that the protein forms a 12mer oligomeric complex, mediated by the C-terminal domain. This stoichiometry is inconsistent with the observation, from X-ray crystallography, that the C-terminal domain alone forms a 14mer complex. It would be ideal to structurally characterize the full-length CI complex in order to address this puzzle, and also to understand how DNA can interact with the CI to achieve its sophisticated transcriptional regulation. However, as described in Chapter 3, attempts to solve the crystal structure of the full-length 186 CI protein have been unsuccessful.

In the absence of a full length 186 CI structure, a hybrid λ -186 CI repressor (henceforth referred to as the CI-hybrid or the hybrid protein) was investigated as a model to probe the oligomeric state and supramolecular assembly of 186 CI. The Shearwin lab constructed this hybrid protein (Pinkett et al., 2006), which is a fusion of the λ CI NTD (residues 1-92) with the 186 CI CTD (residues 84-192). One potential advantage, from the structural biology perspective, of the hybrid protein is that its design reduces the length of the flexible linker found in wild type 186 CI (Shearwin et al., 2002). In addition, the use of the well characterised λ CI NTD, which has been successfully crystallised bound to lambda operator DNA (Beamer and Pabo, 1992), may also reduce the flexibility of the protein. Movements of NTD within the oligomeric protein would be further decreased by the dimerization of the λ CI NTD domain (Pabo and Lewis, 1982), a phenomenon not observed for 186 CI NTD (Shearwin et al., 2002). These reductions in flexibility may in turn aid protein crystallization (Derewenda, 2004).

A case can also be made for the hybrid protein being a suitable model to investigate 186 CI. Both λ CI and 186 CI are two domain proteins, consisting of an N-terminal DNA binding domain and a C-terminal oligomerization domain. The N-terminal domains from λ CI and 186 CI share significant structural similarity with both domains consisting of 5 alpha helices, with helices 2 and 3 forming a helix-turn-helix motif (Pinkett et al., 2006). Helix 5 is significantly shorter in 186 CI than λ CI, which is thought to provide some additional flexibility to 186 CI NTD. As the N-terminal DNA binding domains are structured similarly and are comparable in size, it is likely that the 186 CI CTD would be able to mediate the hybrid protein into the same overall oligomeric state. This hypothesis is further supported by size exclusion chromatography - multi-angle laser light scattering (SEC-MALLS) data

from our laboratory which measures the hybrid protein forming an oligomer with a molecular weight equivalent to approximately 12.9 monomers (Cui Lun, unpublished data).

In their study, Pinkett *et al.* (2006) measured the activity of this hybrid protein *in vivo* against chromosomally integrated, synthetic reporter constructs (Figure 4-1A). The constructs utilized the lambda O_L and O_R binding sites and a lacZ reporter gene placed under the control of the p_R or p_{RM} promoters.

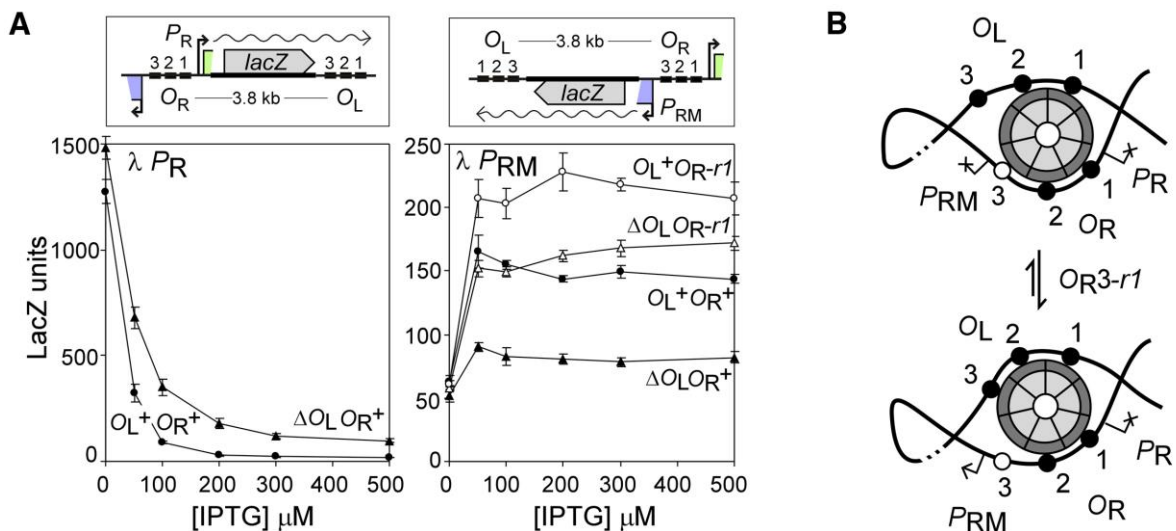


Figure 4-1 - The λ -186 CI-hybrid protein, consisting of λ CI NTD (residues 1-92) fused to the 186 CI CTD (residues 84-192), functions as a repressor. (A) The λ -186 CI-hybrid protein is under the control of an IPTG-inducible plac promoter, with pUHA-1 supplying lac repressor. The λ -186 CI-hybrid protein can repress the p_R promoter and activate the p_{RM} promoter in the p_R:lacZ and p_{RM}:lacZ constructs shown. Error bars represent 95% confidence intervals. (B) Model proposing a mechanism for ability of the O_L operators to increase p_{RM} activation. The original model assumes the hybrid protein adopts a 14mer wheel, as a heptamer of dimers. O_R1, O_R2 and O_R3 operators can bind the wheel to repress p_R and p_{RM}. The three O_L operators can also bind the wheel. Figures are reproduced from Pinkett *et al.* (2006).

In these constructs, the CI-hybrid protein can repress the λ p_R promoter. This repression is aided by the presence of the O_L binding sites, indicating that hybrid protein can simultaneously contact O_L and O_R (Figure 4-1A left construct). The regulation of λ p_{RM} by the hybrid protein is different from λ CI. With three intact O_R operators and in the absence of O_L, the hybrid protein only slightly activates the p_{RM} promoter (Figure 4-1A right construct $\Delta O_L O_R^+$). The 14mer wheel-centric model suggests that the hybrid protein can occupy all three O_R operators. Similar to the mechanism with wild type lambda CI, when the hybrid protein occupies O_R1-O_R2, it can activate the p_{RM} promoter by directly contacting RNAP bound at p_{RM} (Jain *et al.*, 2004; Nickels *et al.*, 2002). However, the hybrid protein can more efficiently occupy O_R3 than is the case in the lambda switch, and so can more efficiently compete with RNAP binding and thus reduce the level of p_{RM} activation.

Reducing the affinity of the hybrid protein for the O_R3 operator, without affecting p_{RM} strength, by introducing the O_R3-r1 mutation (Dodd *et al.*, 2001) increases the activation of p_{RM} by over two fold

(Figure 4-1A right construct $\Delta O_L O_{R-r1}$). This activation can be explained by the reduced occupation of O_{R3} by the CI-hybrid protein and thus reduced competition with RNAP binding at p_{RM} . In contrast, λ CI activates p_{RM} significantly more (regardless of the strength of O_{R3}) due to weaker occupation of O_{R3} (Dodd et al., 2001).

Introducing O_L into the constructs improved the activation of p_{RM} by the hybrid protein (Figure 4-1A right construct $O_L^+ O_{R-r1}$ and $O_L^+ O_{R-r1}$). Thus, O_L appears to reduce the occupation of O_{R3} . The proposed mechanism is that O_L can compete with O_{R3} for binding sites on the hybrid CI wheel, to relieve repression on p_{RM} . This activity is analogous to DNA looping to the 186 flanking F_L/F_R sites to relieve repression of p_L in 186 CI model (Chapter 1). Thus, the presumed wheel formed by the CI-hybrid can show similar DNA binding and looping behaviour to 186 CI, making it a suitable alternative model to the 186 CI.

In addition, the hybrid protein has successfully been used in a genetic screen, based on that originally used for lambda CI by Hochschild and Ptashne (1988), to select for cooperativity mutants in the 186 CI-CTD (Pinkett et al., 2006). Many of these cooperativity mutants were positioned on the dimer-dimer interface when mapped onto the crystal structure of the 186 CI-CTD wheel, which is perhaps the strongest evidence that the hybrid protein uses a similar oligomerisation mechanism to 186 CI.

In this chapter, we successfully solved the X-ray crystal structure of the CI-hybrid protein bound to single DNA operators, and characterized its oligomerization using mass-spectrometry.

4.2 Materials and Methods

4.2.1 Overexpression of the hybrid protein

The expression strain for the hybrid protein, CL1950, was provided by Cui Lun (Shearwin lab). CL1950 is BL21(DE3) pLysS cells harbouring the hybrid protein expression vector pET3a λ -186 CI-hybrid His₆. The vector contains the T7 promoter controlling the expression of the hybrid protein (λ CI residues 1-92 fused to 186 CI residues 84-192) followed by a thrombin cleavage consensus sequence (LVPRGS) and a hexahistidine affinity tag.

2 x 5 mL LB supplemented with 100 $\mu\text{g mL}^{-1}$ ampicillin and 30 $\mu\text{g mL}^{-1}$ chloramphenicol were inoculated with CL1950 glycerol stocks and grown overnight at 37°C. 2 mL of the overnight culture was used to inoculate 2 x 500 mL of the same selective LB media and cultured at 37°C to an OD₆₀₀ of 0.52. The cultures were induced with 200 μM IPTG and grown overnight at room temperature. Cells were harvested by centrifugation at 5000 g for 10 min and stored at -80°C until purification. Cell pellets were thawed and resuspended in 40 mL of lysis buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, 5 mM 2-mercaptoethanol, pH 7.4) and sonicated on ice (4 x 30 s, 1 min rest between sonication cycles). The cell lysate was clarified by centrifugation (40 000 g, 40 min) and supernatant was sequentially passed through a 0.8 μm , 0.45 μm and 0.2 μm syringe filters to remove

remaining suspended solids. The clarified lysate was applied to a 5 mL HisTrap FF column (GE Healthcare) preequilibrated with lysis buffer. The column was washed with 10 CV of lysis buffer, followed by 2 CV of lysis buffer supplemented with 150 mM imidazole. Protein was eluted from the column with 2 CV of lysis buffer supplemented with 500 mM imidazole. Fractions showing UV absorbance $A_{280\text{ nm}}$ above baseline were pooled and tested for purity using SDS-PAGE (Figure 4-2). Fractions containing purified protein were dialysed against TEG150 (50 mM Tris-Cl, 0.1 mM EDTA, 150 mM NaCl, 10% glycerol, pH 7.5, 0.5 mM DTT). The protein was concentrated to 8 mg/mL by centrifugation with Amicon Ultra-15 centrifugation units (Millipore). Protein was further purified by size-exclusion chromatography on an Enrich SEC 650 column (Bio-Rad), using TEG150 as the chromatography buffer. Protein was then concentrated, as before, to a concentration of 8.6 mg/mL.

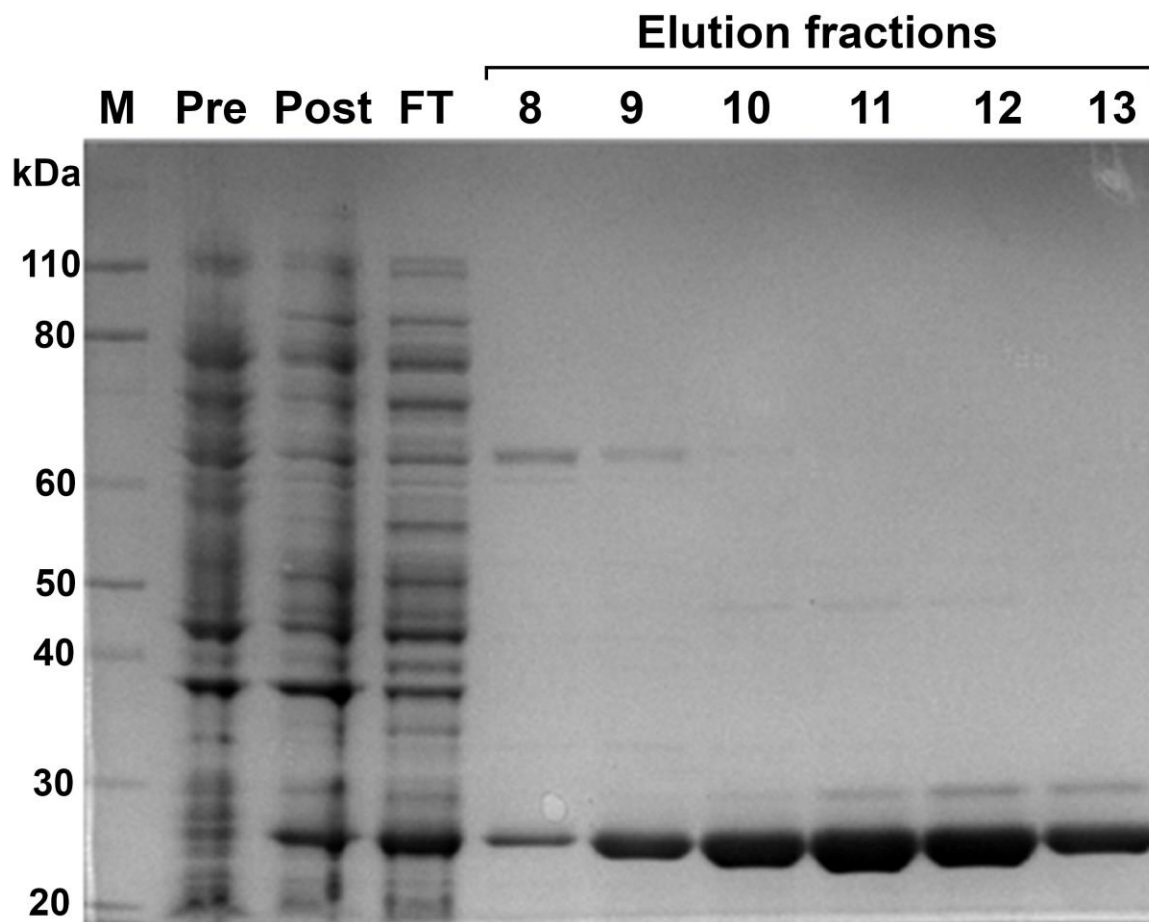


Figure 4-2 - SDS-PAGE analysis of fractions from Ni-NTA chromatography to purify CI-hybrid protein. Samples were run on 4-12% NuPAGE BisTris Protein Gels (ThermoFisher Scientific) with NuPAGE MOPS buffer. Lane M: Novex Sharp Pre-Stained Protein Standard (ThermoFisher Scientific, 2 μ L), Pre: Pre-induction sample. Post: Post-induction sample. FT: Flowthrough fraction. Elution fractions that show UV absorbance above baseline, corresponding to lanes 5-10, have been labelled on the gel image.

4.2.2 Preparation for protein-DNA complex for crystallography

Some of the published X-ray crystal structures of λ CI have the protein in complex with DNA (Beamer and Pabo, 1992; Jain et al., 2004; Stayrook et al., 2008). An attempt was made to crystallize the hybrid protein in complex with the same DNA (OL1) as described in Stayrook et al (2008). The oligonucleotide sequences are listed in Table 4-1. DNA oligonucleotides were commercially synthesized and purified with ion-exchange HPLC purification (Integrated DNA Technologies). Each oligonucleotide was dissolved in TE buffer (10 mM Tris-Cl, 0.1 mM EDTA, pH 8) to a concentration of 3 mM. 50 μ L of the two oligonucleotides were mixed, heated to 95°C and slowly cooled to room temperature to anneal the strands together. 100 μ L of 8.6 mg/mL CI-hybrid protein was mixed with 16.5 μ L of 1.5 mM OL1 DNA and incubated on ice for 30 minutes. Crystallization screening was conducted via sitting drop vapour diffusion in 96 well Intelliplates (Art Robbins), set up using a Phoenix robot (Art Robbins). 1 μ L of CI-hybrid/OL1 DNA mixture was mixed with 1 μ L of reservoir solution and equilibrated over 75 μ L reservoir solution at 16°C. Commercial screens were used to

provide crystallization conditions. The commercial screens used were NatrixHT, Index HT, Crystal Screen HT and PEG/ION HT (Hampton Research) (<https://hamptonresearch.com/cat-1.html>). A single crystal was observed in the NatrixHT screen condition F2 (0.08 M sodium chloride, 0.012 M potassium chloride, 0.04 sodium cacodylate trihydrate pH 8.0, 30% w/v (+/-)-2-methyl-2,4-pentadiol, 0.012 M spermine tetrahydrochloride) after 3 months. The crystal was cryoprotected by passing it through a cryoprotection solution (7.5 μ L mother liquor mixed with 1.5 μ L of 80% glycerol) and flash freezing in liquid nitrogen.

Table 4-1 - Oligonucleotide sequences for crystallography studies

Oligonucleotide name	Sequence
OL1 top	TATATCACCGCCAGTGGTAT
OL1 bottom	AATACCACTGCCGGTGATAT

4.2.3 X-ray data collection, processing and structure solution

Data was collected from a single crystal at 100 K at the Australian Synchrotron MX1 beamline using the ADSC Quantum 210r CCD detector (Aragão et al., 2018; McPhillips et al., 2002). 240 diffraction images were collected with 1° oscillation width, with crystal-to-detector distance set to 275 mm. Diffraction images were processed and scaled using XDS and Scala (Evans and Murshudov, 2013; Kabsch, 2010). The resolution limit was chosen such that the overall $CC_{1/2} > 0.3$. The data processing statistics are summarised in Table 4-2. The structure was solved by molecular replacement using Phaser MR (McCoy et al., 2007), positioning two search models in the asymmetric unit sequentially. The search models were a 186 CI-CTD dimer (PDB ID 2FKD chains A and B) and the λ CI in complex with OL1 DNA with the CTD removed from the model (PDB ID 3BDN with residues 93 to 236 removed from chains A and B). Each asymmetric unit contained a protein dimer in complex with one double stranded OL1 DNA. The model was iteratively refined and rebuilt using phenix.refine (Afonine et al., 2012) and COOT (Emsley et al., 2010), respectively. Model validation was performed using Molprobity (Chen et al., 2010). Refinement statistics for the final model are summarized in Table 4-3. The final structure can be accessed from the Protein Data Bank with PDB accession ID 7JVT.

Table 4-2. Data processing statistics of CI-hybrid crystallography dataset. Statistics for the highest resolution shell are shown in parentheses.

Wavelength (Å)	0.9537
Total images	240
Oscillation (°)	1.00
Space group	H3 ₂
Unit cell	a=177.34 b=177.34 c=108.61 α=90.00 β=90.00 γ=120.00
Resolution (Å)	44.34-3.16
Mosaicity (°)	0.32
Total reflections	165170 (29299)
Unique reflections	11372 (2045)
Redundancy	14.5 (14.3)
Completeness (%)	99.8 (99.4)
Average I/σ(I)	14.6 (3.1)
R _{merge} ^a (%)	13.5 (73.8)
R _{meas} ^b (%)	14.0 (76.5)
R _{pim} ^c	3.7 (0.200)
CC _{1/2} ^d	99.9 (91.0)

$$^a R_{merge} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)} \text{ (Arndt et al., 1968)}$$

$$^b R_{meas} = \frac{\sum_{hkl} \{N(hkl)/[N(hkl) - 1]\}^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)} \text{ (Diederichs and Karplus, 1997)}$$

$$^c R_{pim} = \frac{\sum_{hkl} \{1/[N(hkl) - 1]\}^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)} \text{ (Weiss and Hilgenfeld, 1997)}$$

$$^d CC_{1/2} = \frac{\sum (x - \langle x \rangle)(y - \langle y \rangle)}{[\sum (x - \langle x \rangle)^2 (y - \langle y \rangle)^2]^{1/2}} \text{ (Karplus and Diederichs, 2012)}$$

Table 4-3 - Refinement statistics of CI-hybrid structure

Reflections used in refinement	11352 (1121)
Reflections used for R-free	1142 (115)
R-work ^a	0.2271 (0.2917)
R-free ^b	0.2814 (0.3521)
CC(work)	0.916 (0.588)
CC(free)	0.933 (0.307)
Number of non-hydrogen atoms	3882
macromolecules	3831
solvent	51
Protein residues	397
RMS(bonds)	0.002
RMS(angles)	0.44
Ramachandran favored (%)	96.44
Ramachandran allowed (%)	3.31
Ramachandran outliers (%)	0.25
Rotamer outliers (%)	0.00
Clashscore	6.92
Average B-factor	84.99
macromolecules	85.37
solvent	56.47
Number of TLS groups	11

^a $R_{work} = \sum |F_o - F_c| / \sum |F_o|$ for all data with $F_o > 2\sigma(F_o)$, excluding data to calculate R_{free}

^b $R_{free} = \sum |F_o - F_c| / \sum |F_o|$ for all data with $F_o > 2\sigma(F_o)$, calculated with 5% of reflections that are randomly chosen (Brünger, 1992).

4.2.4 Analysis of structure

Structures were visualized in UCSF Chimera (Pettersen et al., 2004) and PyMOL (DeLano, 2002). Surface area calculations were performed on the PDBEPIA webserver (Krissinel and Henrick, 2007). Intramolecular residue-residue contacts between different protein chains were investigated with the Protein Interaction Calculator webserver (Tina et al., 2007), using the default distance constraints of the server. Protein-DNA contacts were analysed using NUCPLOT (Luscombe et al., 1997).

4.2.5 Mass spectrometry characterisation of CI-hybrid protein

The CI-hybrid protein was exhaustively dialysed into 0.3 M ammonium acetate at 4°C, flash frozen into liquid nitrogen and stored at -80°C. Mass-spectrometry experiments were performed on a Synapt HDMS system (Waters, Manchester, UK). Ions were produced by electrospray ionisation (ESI) from sample introduced from gold-coated borosilicate capillaries prepared in-house (Hernández and Robinson, 2007). The instrument parameters were optimized to remove adducts while preserving noncovalent interactions and were typically as follows: capillary voltage, 1.5 kV; cone voltage, 50 V; trap collision energy, 10 V; source temperature, 20°C; and backing pressure, 6 mBar.

4.3 Results

4.3.1 Structural analysis of the CI-hybrid protein

The asymmetric unit contains a dimer of CI-hybrid in complex with one double stranded OL1 DNA (Figure 4-3A). Despite the different oligomerization C-terminal domains, the λ -derived CI NTDs in the hybrid structure contact the OL1 half sites in the same manner as λ CI in the crystal structure (3BDN) that was used as a search model (Figure 4-3B). Structural alignment of λ CI NTDs in complex with DNA from this structure to previously solved structures showed very little structural deviation as measured by all-atom RMSD, which lies between 0.56Å – 1.28Å (Figure 4-4A). The DNA provides strong steric constraints on the relative positions on the NTDs, which would be further restricted by the weak dimerization interface between helix 5 of the monomers (Pabo and Lewis, 1982; Sauer et al., 1990). The DNA present in the crystal structure of the hybrid is identical in sequence to previously solved structures of the λ CI dimer in complex with DNA (PDB ID: 3BDN) and λ CI NTD dimer in complex with DNA (PDB ID: 1LMB). An attempt was made to compare DNA-protein interactions across these structures (Figure 4-4B). However, the ability to accurately determine interactions between the CI-hybrid protein and DNA is hampered by the resolution (3.16 Å) of the model. At this resolution, many flexible side chains cannot be modelled from the electron density, which translates into many of the interactions between side chains and DNA being undetected. For this reason, there are several interactions detected in the highest resolution 1.8Å published structure (1LMB) that were not detected in the CI-hybrid structure. All of the DNA-protein interactions that were detected in the CI-hybrid structure were also found in the highest resolution model, with the exception of two interactions Asn58 and Asn61 of NTD 2 contacting the phosphates of guanine 13 and guanine 14, respectively. The Asn58 contact was also observed in the structure of the λ CI dimer complex with DNA (3BDN). In the 1LMB model, N58 instead contacts the phosphate of the next base guanine 14. Thus, there is a local rearrangement of side chains in this area of the protein between the structures.

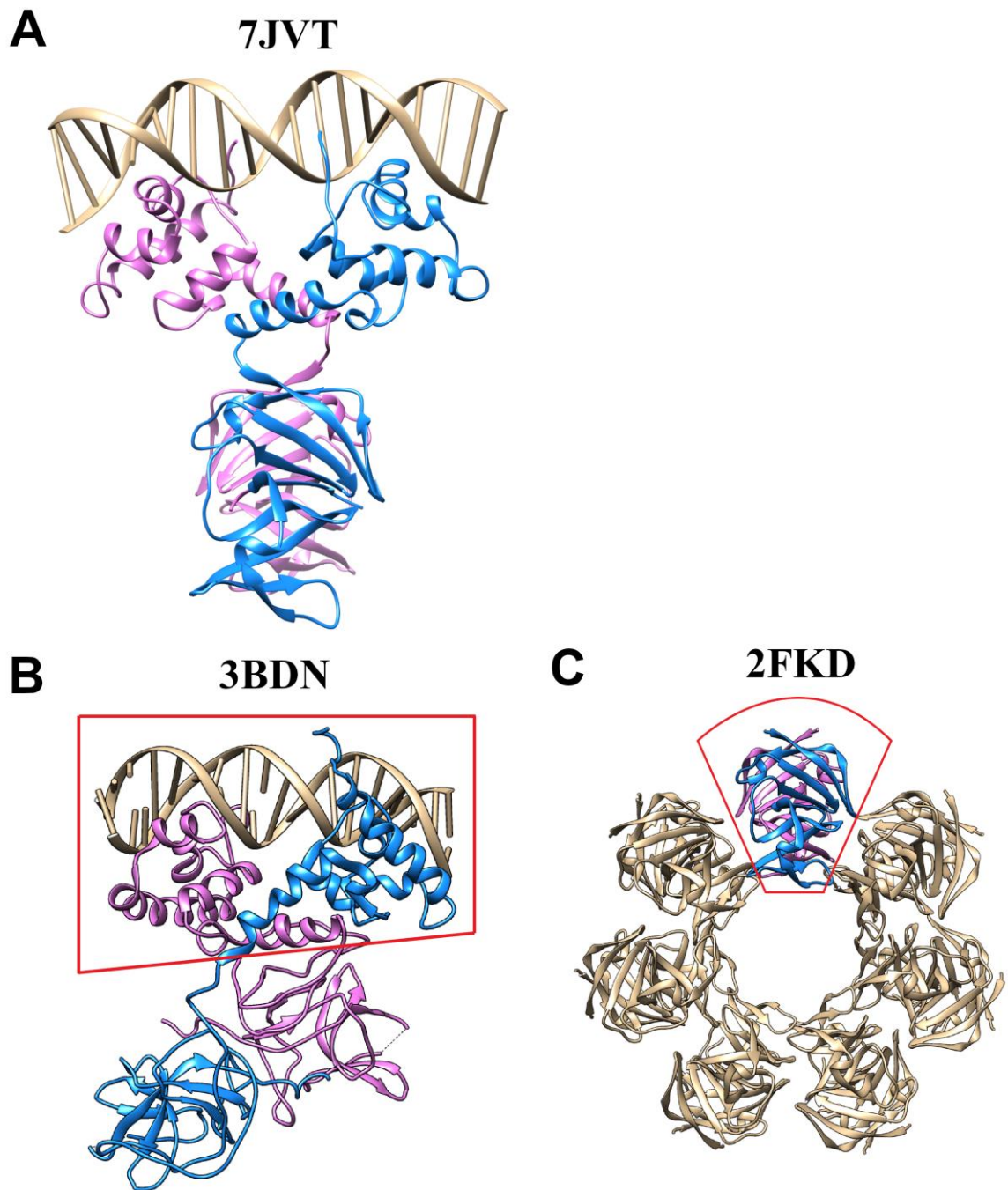


Figure 4-3. The CI-hybrid structure (PDB ID 7JVT) shows a high degree of structural similarity to the search models containing individual domains. (A) Asymmetric unit of the solved structure of the CI-hybrid protein contains a dimer of CI-hybrid bound to OL1 DNA. The DNA binding and dimerization are mediated by the λ CI NTD and 186 CI CTD portions, respectively. The structure shares structural similarity to the previously solved structures of the domains in (B) λ CI (PDB ID: 3BDN) and (C) 186 CI CTD (PDB ID: 2FKD). The search models used in molecular replacement to solve the CI-hybrid structure have been approximated in red in (B) and (C).

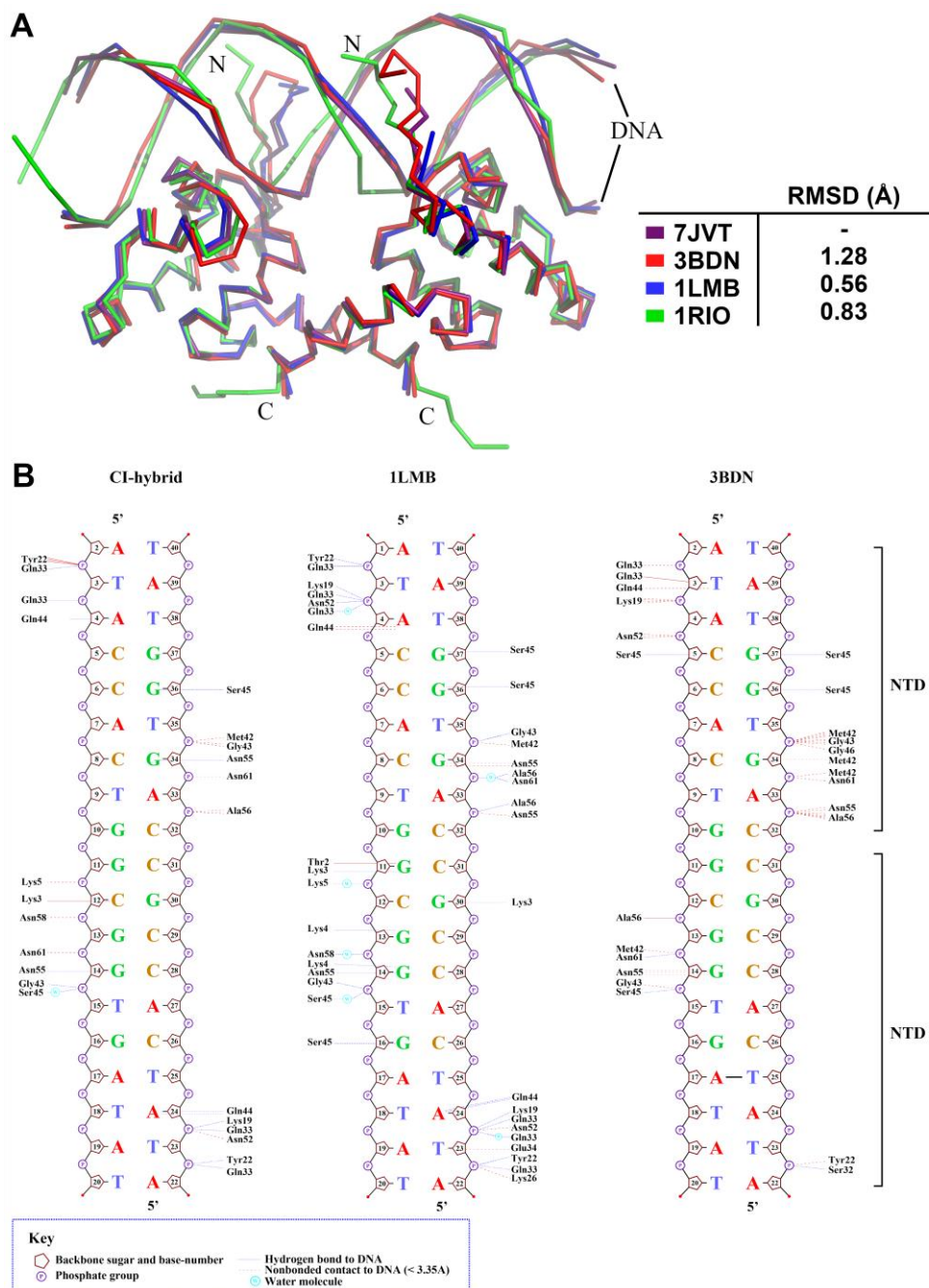


Figure 4-4 – The CI-hybrid protein and λ CI share the same DNA-binding conformation. A) Structural superposition of λ CI NTDs (residues 1-92) in complex with DNA from the CI-hybrid protein (PDB ID 7JVT) with previously solved structures of λ CI NTDs contacting DNA (PDB ID 1LMB, 1RIO) and λ CI dimer binding DNA (PDB ID 3BDN). Minimal structural deviations were observed as indicated by the low all-atom RMSD values measured. B) Nucleotide-residue contact map summarising the interactions between λ CI NTDs and DNA in the CI-hybrid and previously solved structures 1LMB and 3BDN reveals similar binding interactions. The protein chains to which the interacting residues belong are indicated by “NTD 1” and “NTD 2” on the right. Maps were generated using NUCPLOT (Luscombe et al., 1997).

The C-terminal domain derived from 186 CI mediated the dimerization of the protein in a manner similar to the 186 CI-CTD 14mer structure (Figure 4-3C). Structural superposition of these C-terminal domains of the CI-hybrid protein against existing structures of the CI-CTD 14mer structure and the 186 CI^{E146K} dimer also shows minimal structural deviation, with all-atom RMSD measuring 0.537 Å and 0.525 Å, respectively.

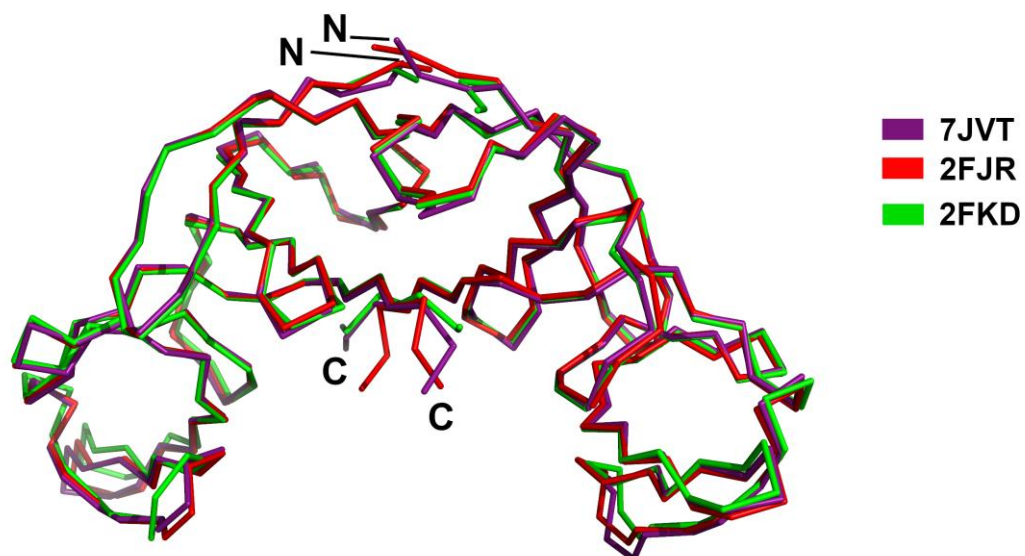


Figure 4-5 – Superposition of a dimer of 186 CI CTD domains from the CI-hybrid protein (PDB ID 7JVT), 186 CI^{E146K} dimer (PDB ID 2FJR) and 186 CI-CTD 14mer wheel (PDB ID 2FKD). The 186 CI CTD in the CI-hybrid protein dimerises in the same manner as the CI-CTD and 186 CI^{E146K} crystal structures.

Having established that the dimers within CTD of the hybrid are essentially identical to the dimers within the isolated 186 CTD, the key question of the overall stoichiometry of the wheel assembly could be addressed. The quaternary structure of the protein in the lattice is a dodecamer wheel, arranged as a hexamer of dimers (Figure 4-6). This is consistent with mass spectrometry experiments of the 186 CI-CTD and 186 CI proteins described in Chapter 3, which determined that the final higher order oligomer that formed from oligomerization of 186 CI-CTD dimers is a dodecamer. The arrangement of the DNA binding domains on the outside of the wheel would still allow DNA to wrap on and off the wheel. The protein binds OL1 DNA in the same plane as the wheel, as expected from this DNA wrapping model. Interestingly, the 186 CI^{E146K} crystal structure has the DNA binding domains positioned relative to the CTD in a way where DNA would bind orthogonal to the plane of the wheel. However, it is clear from the structure of the hybrid protein that wild type 186 CI can bind DNA in a wheel-like manner, as the domains are connected by a flexible region which allows the DNA binding domains to rearrange into a conformation that facilitates wrapping of DNA onto the protein.

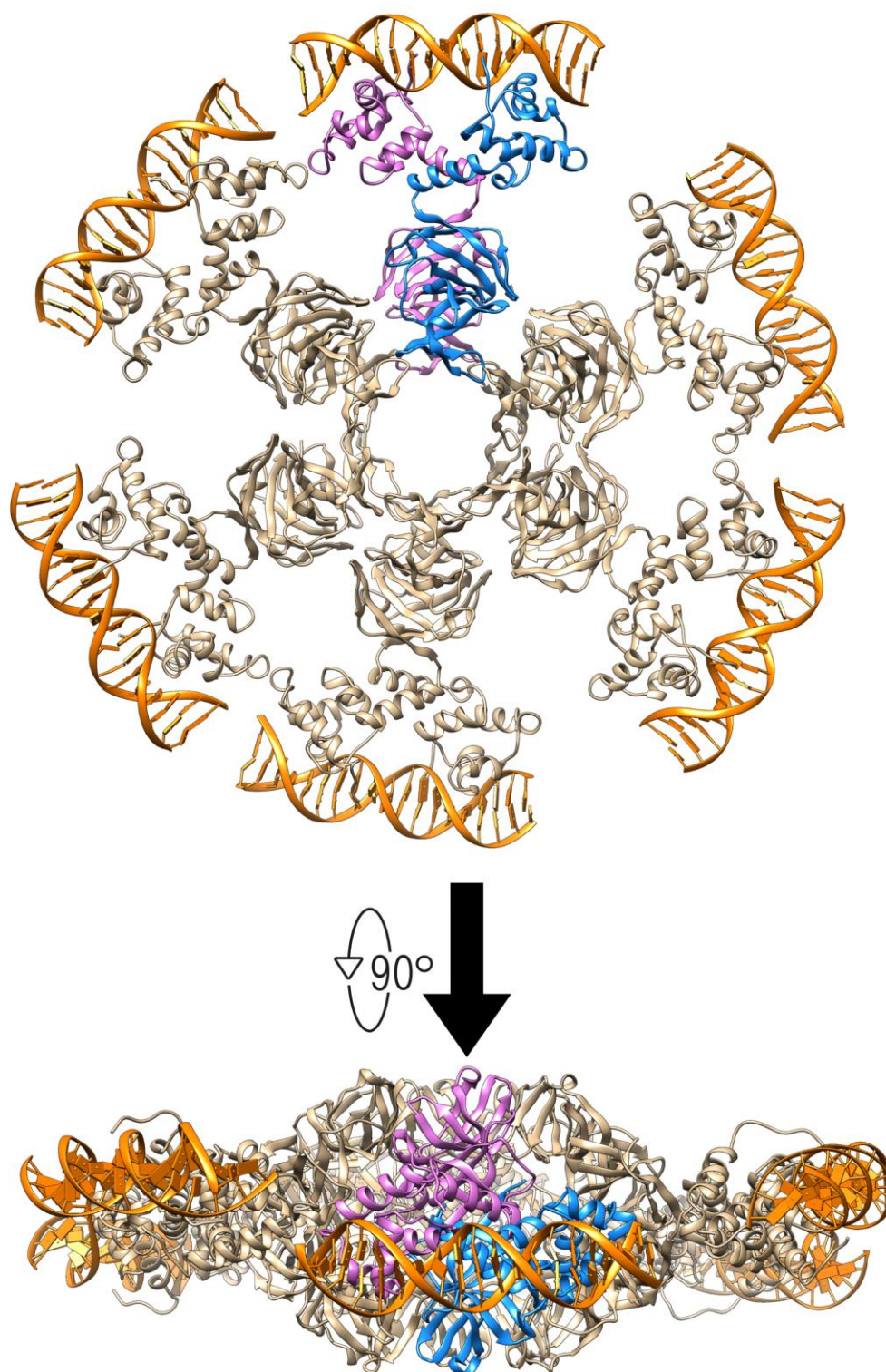


Figure 4-6 - CI-hybrid protein forms a dodecamer quaternary structure in the crystal lattice. Dimers of CI-hybrid oligomerize through the C-terminal domain to form a hexamer of dimers in a circular arrangement.

4.3.2 Mass spectrometry analysis of the CI-hybrid protein

To investigate the oligomeric state of the CI-hybrid protein outside a crystalline system, native mass spectrometry experiments were carried out on the CI-hybrid protein in the absence of DNA (Figure 4-7). Dimers, tetramers, hexamers, octamers, decamers and dodecamers of the CI-hybrid protein were identified in the mass spectra of the CI-hybrid protein at concentrations 10 μM and 30 μM . The

same oligomeric states were also observed in the mass spectra of 186 Cl (Chapter 3), suggesting that the same oligomerization pathway is followed by both proteins. The dimer and tetramer species were more abundant at 10 μM , compared to 30 μM . In contrast, the dodecamer species is more abundant at 30 μM . This is consistent with higher order oligomerisation being a concentration dependent process.

Finally, collision induced dissociation (CID) was performed to provide additional evidence for the stoichiometry. CID of the 7950 m/z ions, assigned as a 12mer species, produced both 11mer and monomer fragment ions (Figure 4-8). This result provides a high level of confidence that the highest molecular weight species was accurately assigned as a 12mer species.

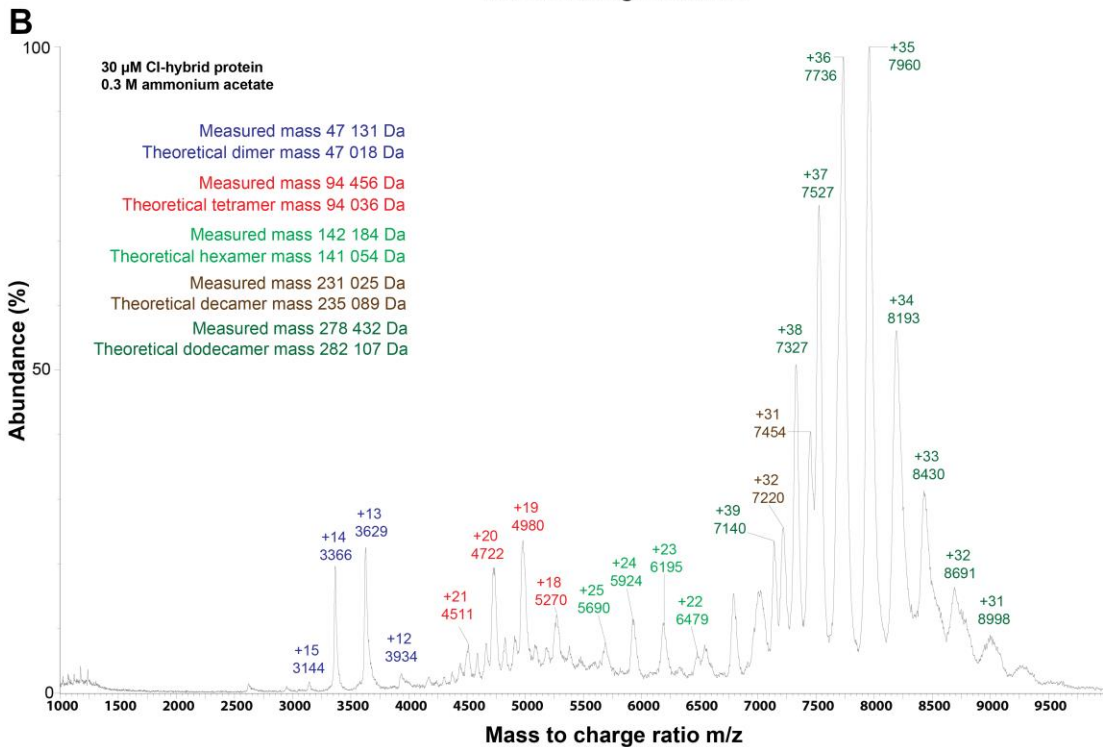
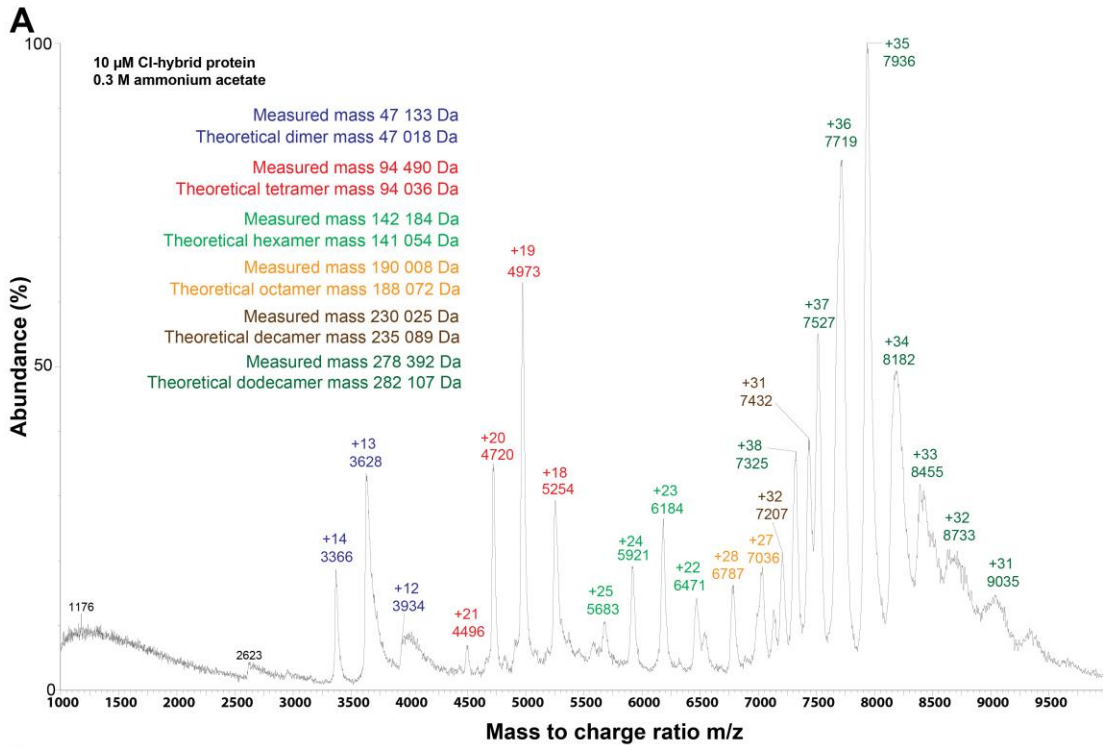


Figure 4-7 - ESI-TOF spectra of the CI-hybrid protein at (A) 10 μ M and (B) 30 μ M were measured in aqueous ammonium acetate (0.3 M). Dimers, tetramers, hexamers, octamers, decamers and dodecamers (12mer) are detected.

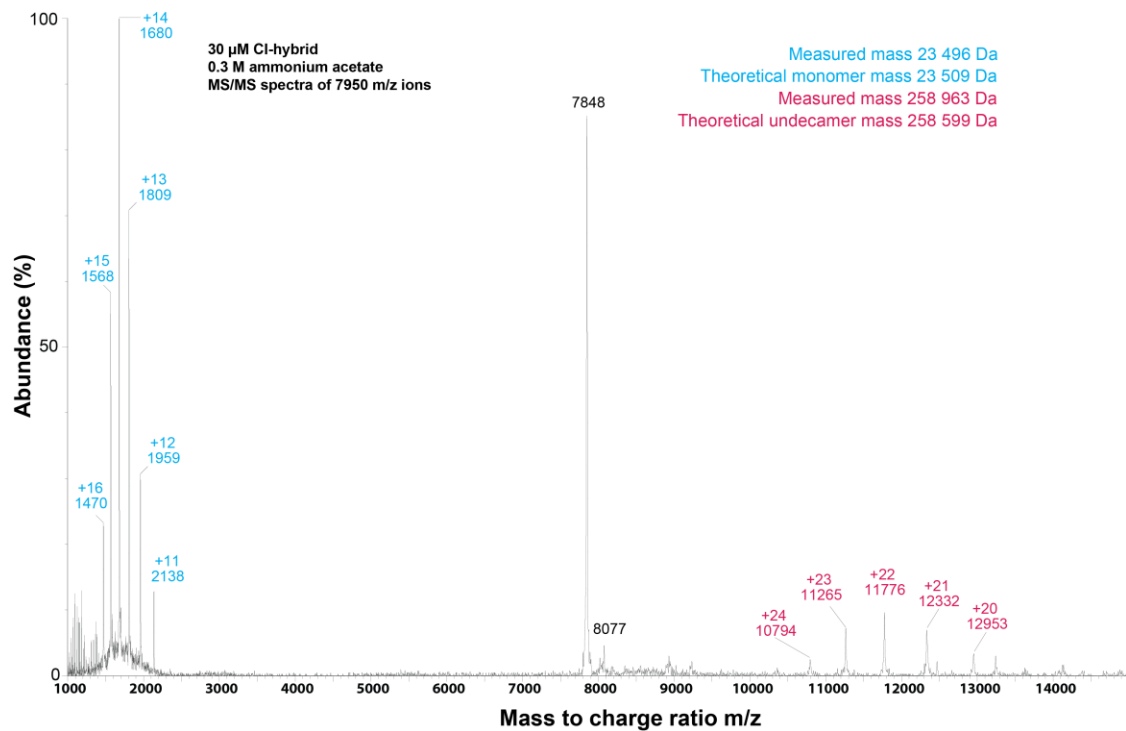


Figure 4-8 - The MS/MS spectra of the collision induced dissociation of the 7950 m/z species, assigned as a 12mer in Figure 4-7B, detected a undecamer (11mer) and monomer.

4.4 Discussion

4.4.1 Alternative wheels configurations mediated by 186 CI CTD

The dodecamer wheel arrangement of the CI-hybrid crystal structure differs from the 14mer wheel observed in the 186 CI-CTD crystal structure, which was composed of a heptamer of dimers (Figure 4-3C). Circular protein assemblies, like 186 CI, present a straightforward mechanism for forming alternative oligomeric states. A subunit can be added and subtracted from the circular ring, provided the physical constraints, such as the distance between protein-protein interfaces, are satisfied (Figure 4-9). For example, bacteriophage portal proteins exclusively form 12-mer rings within intact phage particles. However, isolated phage portal proteins have been observed to form other ring oligomers when purified following overexpression from plasmids. These oligomers range from 11mer to 14mer structures (Cardarelli et al., 2010; Cerritelli and Studier, 1996; Dube et al., 1993; Orlova et al., 2003). Other circular assemblies with alternative oligomeric states include the bacterial pneumolysin toxin (Tilley et al., 2005), the *Salmonella typhimurium* flagella motor (Young et al., 2003) and components of the *S. typhimurium* type III secretion system needle complex (Marlovits et al., 2004). A similar mechanism may be explain the observation that the 186 CI and CI-hybrid crystal structures have adopted different oligomeric states.

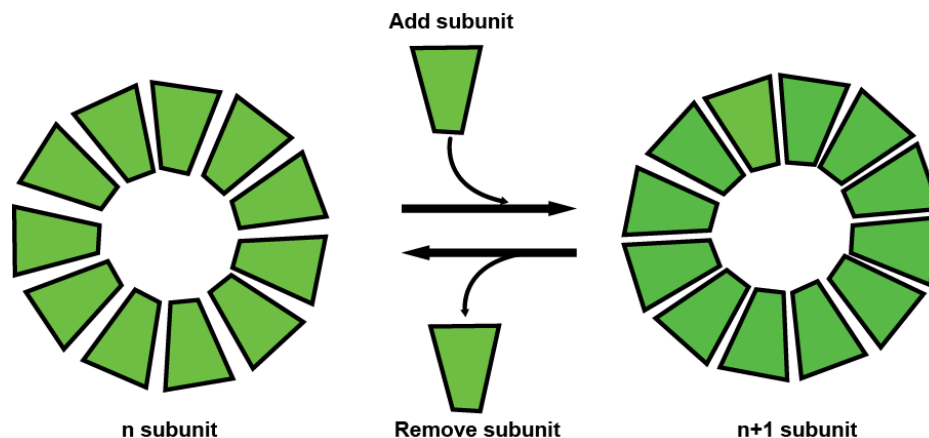


Figure 4-9 – Mechanism of forming alternative ring assemblies. Proteins that oligomerize into rings can often form larger or smaller rings by adding and removing subunits from the circular assembly. The same protein-protein interfaces are utilized in the n and $n+1$ oligomers.

Proteins have been found to form oligomeric structures in crystals that differ from their true state in solution. Crystallography uses protein concentrations many-fold more concentrated than found in nature, which may cause non-physiologically relevant oligomers to form. Conditions in crystallography are also optimised to promote protein-protein interactions to form a crystal lattice. The crystal lattice may stabilise non-physiologically relevant oligomers and complexes through crystal packing. Bahadur et al. (2004) studied structures present in the Protein Data Bank and found over 100 non-physiologically relevant dimers, where proteins exist as monomers in solution. In addition, some oligomeric assemblies can prefer to crystallize in certain spacegroups as their oligomer symmetry can be incorporated into the symmetry of the crystal (Chruszcz et al., 2008). Proteins that form rings and wheels are often radially symmetric and the crystallization process may preferentially select for certain oligomeric rings.

Another possibility is that the oligomeric state of 186 CI in the cellular context could be dependent upon whether it is bound to DNA. For example, if not interacting with DNA, 186 CI could in principle form dodecamers solely via CTD mediated oligomerisation. If bound to DNA, CI might favour a different oligomeric state through the influence of the DNA mediated protein-protein interactions between adjacent dimers. This would be somewhat similar to bacteriophage λ , where DNA looping mediates the formation of the λ CI octamer (Dodd et al., 2005), which is otherwise less favourable in the absence of DNA (Senear et al., 1993). In the lambda situation, DNA binding and looping simply assist a pre-existing oligomerisation, rather than drive the formation of a distinct oligomer (Cui et al., 2013).

Thus, further experiments that follow from this study of 186 CI should aim for structural characterisation of a 186 CI-DNA complex. While attempts at crystallisation of CI in the presence of short fragments of DNA have been unsuccessful, it may be possible with DNA designed to be long enough to wrap around the whole wheel. Alternatively, cryo-electron microscopy may be a more amenable approach for observation of the 186 CI protein-DNA complex than X-ray crystallography.

4.4.2 Dimer-dimer interface analysis

The dimer-dimer interface of the 186 CI-CTD and CI-hybrid buries approximately 970 Å² and 885 Å² of solvent accessible surface area, respectively. This higher buried surface area is reflective of the closer packing of the domain in the 14mer wheel.

Pinkett *et al.* (2006) conducted a genetic screen for cooperativity mutants in the 186 CI C-terminal domain using the CI-hybrid protein. The genetic screen was designed to isolate mutants that could still bind DNA but showed reduced cooperativity between the operator sites. The screen isolated 15 mutations at 13 positions that reduced cooperativity between dimers. These mutations can be mapped onto both the 12mer and 14mer wheel structures. As expected, the mutations both cluster around the dimer-dimer interface in both cases (Figure 4-10).

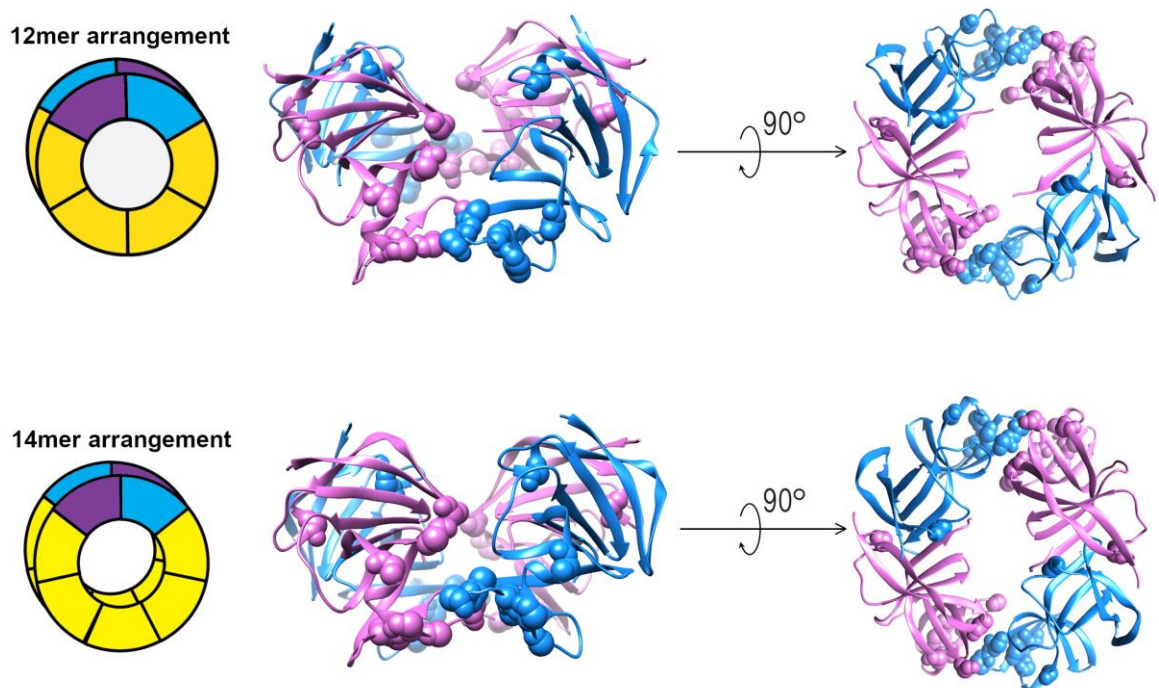


Figure 4-10 – The 186 C-terminal domain uses the same interface to form the dodecamer and 14mer. The CI-CTD cooperativity mutants isolated by Pinkett *et al.* (Pinkett *et al.*, 2006) cluster in this oligomerization interface in both arrangements. The 14mer arrangement has the dimers packed closer together to accommodate an extra dimeric subunit.

The cooperativity mutants were investigated for possible residue-residue interactions between adjacent dimers in a wheel configuration. The distances of these interactions in the 12-mer and 14-mer configuration in the crystal structures were quantified (Table 4-4). It is important to note that the interaction distances of the 14mer were based upon 13 or more values, but the interaction distances of the 12mer were only based on 2 values. This is because the crystal structure of the 186 CI CTD contains 14 subunits in the asymmetric unit, and each dimer interacts with an adjacent dimer slightly differently. However, in the case of the CI-hybrid structure, the asymmetric unit contained a CI-hybrid dimer. The 12mer arrangement is generated through symmetry operations based on the

spacegroup of the crystal. This results in only 2 unique bond interaction distances in the arrangement. Each residue interaction distance was observed 6 times in the arrangement.

Out of the residue positions where 186 CI-CTD dimer cooperativity mutants were found, four residues were found to form plausible ionic interactions with residues from adjacent dimers. K156 from one dimer can form an ionic interaction with either E96 or E146 on an adjacent dimer. Interaction with one glutamate side chain positions the lysine side away from interacting with the other. In the 14mer crystal structure, K156 preferentially interacts with E96. In the 12mer crystal structure, K156 was observed to interact with either E96 or E146, at the detriment of the other interaction. This choice between ionic interactions possibly contributes to seeing both 12mer and 14mer oligomeric states in the crystal structures. The other ionic interaction between K168 and E153 is closer in the 12-mer than 14-mer.

Table 4-4 – Residue-residue interaction distances of 186 CI cooperativity mutants in the 12mer and 14mer configuration based on crystal structures.

Mutation	Possible intramolecular residue contact	Type of interaction	Contact distance in 12mer (Å)	Contact distance in 14mer (Å) *
E96G	K156	Ionic	10.1 2.4	2.9±0.1
F110L	none identified			
E146K	K156	Ionic	4.1 8.1	8.5±0.3
K156I	E96	Ionic	10.1 2.4	2.9±0.1
	E146		4.1 8.1	8.5±0.3
L157P	V169	Hydrophobic	4.3 4.7	4.0±0.1
	F171		4.0 3.8	4.4±0.1
P158S	none identified			
L162P	none identified			
H163R	none identified			
K168E	E153	Ionic	2.3 4.0	5.3±0.3
F171L F171S F171V	L157	Hydrophobic	4.0 3.8	4.4±0.1
	P158		5.0 3.7	4.6±0.1
C173R	none identified			
I178T	none identified			

*Average of all the residue interaction distances in the crystal structure (n≥13). Error stated is the standard error.

4.4.3 Model of transcription regulation using a dodecamer

The previous model of transcriptional regulation of p_L and p_R by CI involved DNA from these promoters wrapping around a 14mer wheel (Pinkett et al., 2006). It is important to consider whether this model of transcriptional regulation by DNA looping remains valid with a 12mer wheel. A

schematic diagram of the regulation model with a 12mer and 14mer is shown in Figure 4-11. One of the assumptions of the 14mer model is that DNA strands in close proximity can sterically clash to occlude and diminish binding. F_L and F_R are proposed to use this steric clash property to push p_L operators off the wheel to activate p_L (Figure 4-11 species c). However, in the structure of the hybrid protein (Figure 4-6 bottom), the DNA does not lie strictly in the plane of the wheel due to flex in the positions of the NTDs. It may be possible for DNA strands to enter and exit the 186 CI wheel from above and below the plane, to mitigate clashing effects. Thus, it is important to discuss the possible roles of clashing in the 12mer and 14mer models.

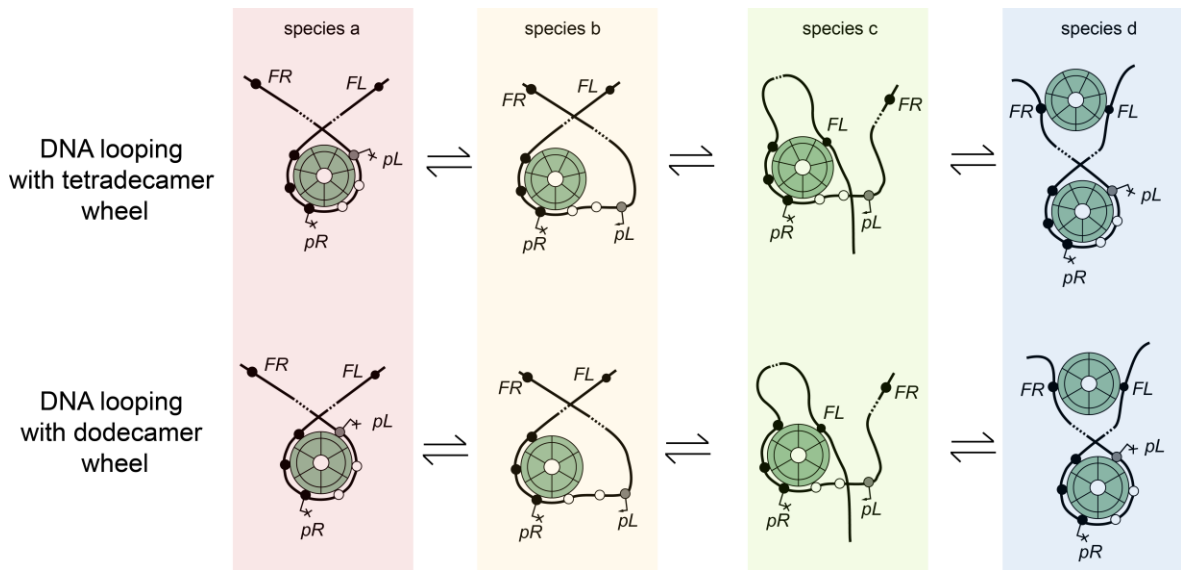
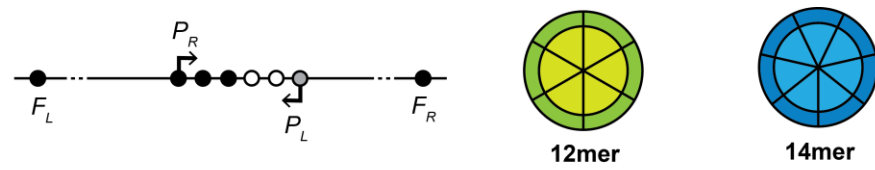


Figure 4-11 – Comparison of the dodecamer and tetradecamer wheel models of transcriptional regulation. 186 CI forms a wheel with the DNA binding HTH-containing domains facing outwards. DNA around the p_R and p_L promoters can wrap around the wheel to repress both the p_R and p_L promoters (species a). The p_L promoter with weaker affinity operators can loop off the wheel scaffold to be replaced by a flanking site (species b and c). At higher concentrations, both F_L and F_R are both sequestered by another CI wheel (species d).

There are two regulatory features of the 186 CI wheel model where DNA clashes can be considered, which are summarized in Figure 4-12. They are (i) corepression of the p_L and p_R promoters and (ii) p_L derepression by competition off the wheel by F_L and F_R binding. 186 CI can corepress p_R and p_L (Dodd and Egan, 2002), which is proposed to be driven by six contiguous operators (including two poorly defined operators) between p_R and p_L all bound to the wheel (Figure 4-11 species a). The dodecamer complex has six dimers of CI, and thus should still be able to complex with the six CI operator sites between the p_L and p_R promoters. However, a strong steric clashing effect between DNA entering and exiting the wheel would prevent the operator p_L from binding, preventing Figure 4-11 species a from forming and would not be consistent with p_R - p_L corepression. Thus, strong steric clashing effects are inconsistent with the 12mer model, but not the 14mer model.

The flanking sites F_L and F_R can improve p_L derepression by CI. This is proposed to be an effect of F_L and F_R binding to the wheel and sterically pushing off the operator at p_L . In both the 12mer and 14mer models, the presence of steric clashes would allow the p_L DNA to be pushed off by flanking

sites. In a 12mer model with no steric clashes, p_L can be activated by an F_L or F_R site binding to the wheel. As there are only 6 available binding positions on the wheel, the operator at p_L may free due to a lack of binding positions. Thus, F_L/F_R assisted derepression of P_L can be observed with a 12mer wheel even in the absence of DNA clashes. In a 14mer model with no steric clash effect, there are enough binding positions on the wheel to bind to all 6 operators between p_L and p_R and whilst allowing an F_L or F_R operator to bind. Thus, this model would require an extra binding site (or significant non-specific binding of DNA) adjacent to F_L or F_R to push off p_L . In summary, the 14mer wheel model requires DNA clashes to explain the regulatory effects of F_L and F_R .



	No clash	Clash
Corepression of p_R and p_L		
F_L/F_R assisted derepression of P_L (via pushing P_L off the wheel)		

Figure 4-12- The effect of DNA clashes on transcriptional regulation by a 12mer and 14mer CI wheel. A 12mer wheel would be consistent with the CI transcriptional regulation model in the absence of strong steric clashes between DNA entering and exiting the wheel. Strong DNA clashes would not allow both p_R and p_L to simultaneously occupy the wheel and is inconsistent with observed p_R - p_L corepression. The regulation model based around the 14mer would be inconsistent in the absence of strong steric clashes as p_R , p_L and F_L (or F_R) could simultaneously occupy the wheel. This simultaneous occupation would result in F_L/F_R not efficiently derepressing p_L .

When examining the CI-hybrid protein regulating p_R and p_{RM} , it was observed that O_L reduces repression of p_{RM} (Figure 4-1). This was attributed to a clashing effect whereby an O_L operator pushes off p_{RM} , which would be consistent in both a 12mer and 14mer model. However, without the

presence of clashes, both the 12mer and 14mer CI-hybrid wheels would have enough binding positions to allow binding. It is possible that the extent of DNA clashes differs between the 186 CI wheel and CI-hybrid wheel. The linker between the DNA binding domain and CTD of the 186 CI protein is significantly longer than the CI-hybrid protein. As will be explained below, this longer linker may allow for the DNA to sit even higher or lower than the plane of the wheel compared to the CI-hybrid protein. This arrangement may eliminate DNA clashes that still exist in the CI-hybrid wheel.

In 186 CI, if the NTDs sat above and below the plane of the wheel, the DNA could follow an “oscillatory” trajectory around the wheel, allowing even longer pieces of DNA wrapping around the wheel. This arrangement may be the method 186 CI uses to alleviate steric clashes between the entry and exit points of the DNA into the well (Figure 4-14). 186 CI can bind two distinct half sites and bind to operators with variable spacing, as explained in Chapter 1. The flexible positioning of the DNA binding domains in terms of distance from the centre of the wheel and its position above or below the plane of the wheel may contribute to its ability to bind different arrangements of operators.

Changing the number of subunits in the wheel will alter the internal cavity diameter, the symmetry and curvature of the wheel. Changing the curvature of the wheel would affect the length of DNA that can wrap around the wheel. To estimate the amount of DNA that will wrap around a single wheel (Figure 4-11 species a), a structural model of the CI wheel was created based on the hybrid wheel structure and the crystal structure of the 186 CI dimer. A DNA molecule was traced around the wheel to approximately match the diameter (Figure 4-13A). This modelled DNA was 164 bp. The linker connecting the DNA binding domain to the oligomerization domain is flexible, thus the true diameter of the wheel could be marginally larger or smaller than the model. Wang *et al.* (Wang et al., 2013) measured by AFM an average of 180 bp wrapping around the wheel. This could be the result of the NTD domains extending slightly further away from the wheel centre. The DNA modelled to wrap around the wheel lies fully in the plane of the wheel. This would create the aforementioned steric clashes when DNA enters and exits the wheel (Figure 4-13B). Modelling an oscillatory DNA trajectory, as depicted in Figure 4-14, to avoid DNA clashes would predict a longer piece of DNA would be required to wrap the wheel.

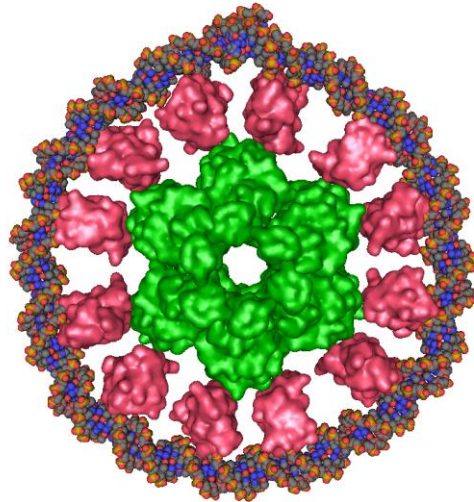
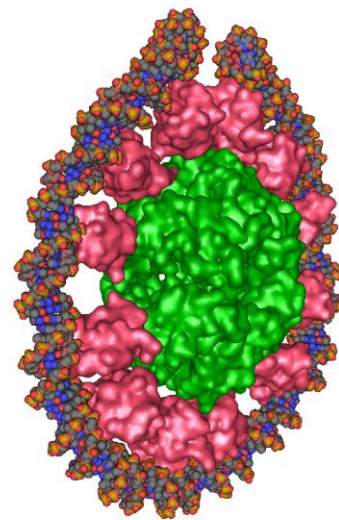
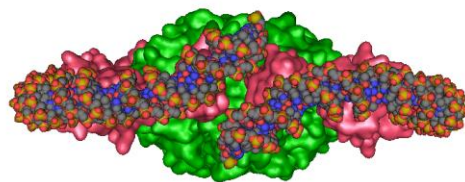
A**Front view****B****DNA entry and exit from wheel**

Figure 4-13 – Model of DNA wrapping around a 186 CI wheel. (A) Model of dodecameric CI wheel constructed through superimposing the CI-NTD (PDB 2FJR) onto the hybrid wheel structure. 164 bp DNA was modelled to wrap around the wheel using GraphiteLiteExplorer (Hornus et al., 2013). (B) DNA was modelled in the plane of the wheel. The model predicts that binding in this manner would have steric effects on DNA entering and exiting the wheel.

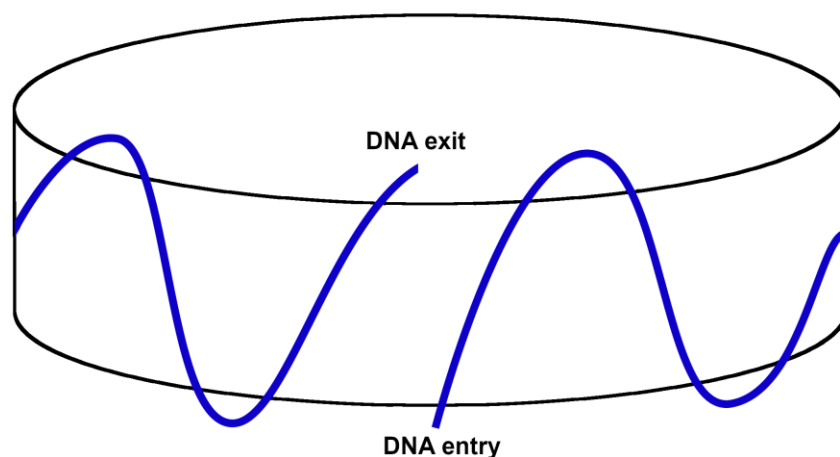


Figure 4-14 – DNA wrapping around the CI wheel could adopt an oscillatory conformation, resulting in more DNA wrapped than if the DNA wrapping was fully in the plane of the wheel. The conformation could also circumvent steric clashes between the DNA entry and exit points. DNA and CI wheel are depicted as a blue line and cylinder, respectively.

4.5 Conclusions

The immunity repressor CI from bacteriophage 186 presents a new mechanism of regulating lytic and lysogenic transcription in a temperate bacteriophage that shows differences to the conventionally studied bacteriophage λ . This mechanism utilises DNA looping onto a wheel scaffold and the relief of transcriptional interference to activate the lysogenic promoter. To validate this model, structural characterization of a CI-hybrid protein was conducted and strongly suggests that 186 CI adopts a dodecamer wheel *in vitro*, rather than the previously proposed tetradecamer wheel. This oligomerization complex does not contradict the DNA looping model of regulation proposed by Pinkett et al. (2006) and Wang et al. (2013). Further studies are required to determine the quaternary structure of 186 CI in the presence of DNA.

4.6 References

- Afonine, P. V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D., 2012. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 68, 352–367.
- Aragão, D., Aishima, J., Cherukuvada, H., Clarken, R., Clift, M., Cowieson, N.P., Ericsson, D.J., Gee, C.L., Macedo, S., Mudie, N., Panjekar, S., Price, J.R., Riboldi-Tunncliffe, A., Rostan, R., Williamson, R., Caradoc-Davies, T.T., 2018. MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. *J. Synchrotron Radiat.* 25, 885–891.
- Arndt, U.W., Crowther, R.A., Mallett, J.F.W., 1968. A computer-linked cathode-ray tube

- microdensitometer for X-ray crystallography. *J. Phys. E Sci. Instruments* 1, 510–516.
- Bahadur, R.P., Chakrabarti, P., Rodier, F., Janin, J., 2004. A Dissection of Specific and Non-specific Protein-Protein Interfaces. *J. Mol. Biol.* 336, 943–955.
- Beamer, L.J., Pabo, C.O., 1992. Refined 1.8 Å crystal structure of the λ repressor-operator complex. *J. Mol. Biol.* 227, 177–196.
- Brünger, A.T., 1992. Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472–475.
- Cardarelli, L., Lam, R., Tuite, A., Baker, L.A., Sadowski, P.D., Radford, D.R., Rubinstein, J.L., Battaile, K.P., Chirgadze, N., Maxwell, K.L., Davidson, A.R., 2010. The Crystal Structure of Bacteriophage HK97 gp6: Defining a Large Family of Head-Tail Connector Proteins. *J. Mol. Biol.* 395, 754–768.
- Cerritelli, M.E., Studier, F.W., 1996. Purification and characterization of T7 head-tail connectors expressed from the cloned gene. *J. Mol. Biol.* 258, 299–307.
- Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, 12–21.
- Chruszcz, M., Potrzebowski, W., Zimmerman, M.D., Grabowski, M., Zheng, H., Lasota, P., Minor, W., 2008. Analysis of solvent content and oligomeric states in protein crystals-does symmetry matter? *Protein Sci.* 17, 623–632.
- Cui, L., Murchland, I., Shearwin, K.E., Dodd, I.B., 2013. Enhancer-like long-range transcriptional activation by λ CI-mediated DNA looping. *Proc. Natl. Acad. Sci. U. S. A.* 110, 2922–2927.
- DeLano, W.L., 2002. The PyMOL molecular graphics system.
- Derewenda, Z.S., 2004. The use of recombinant methods and molecular engineering in protein crystallization. *Methods* 34, 354–363.
- Diederichs, K., Karplus, P.A., 1997. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* 4, 269–275.
- Dodd, I.B., Egan, J.B., 2002. Action at a distance in CI repressor regulation of the bacteriophage 186 genetic switch. *Mol. Microbiol.* 45, 697–710.
- Dodd, I.B., Perkins, A.J., Tsemitsidis, D., Egan, J.B., 2001. Octamerization of λ CI repressor is needed for effective repression of PRM and efficient switching from lysogeny. *Genes Dev.* 15, 3013–3022.
- Dodd, I.B., Shearwin, K.E., Egan, J.B., 2005. Revisited gene regulation in bacteriophage λ . *Curr. Opin. Genet. Dev.* 15, 145–152.

- Dube, P., Tavares, P., Lurz, R., van Heel, M., 1993. The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *EMBO J.* 12, 1303–1309.
- Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, 486–501.
- Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* 69, 1204–1214.
- Hernández, H., Robinson, C. V., 2007. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc.* 2, 715–726.
- Hochschild, A., Ptashne, M., 1988. Interaction at a distance between lambda repressors disrupts gene activation. *Nature* 336, 353–357.
- Hornus, S., Lévy, B., Larivière, D., Fourmentin, E., 2013. Easy DNA Modeling and More with GraphiteLifeExplorer. *PLoS One* 8.
- Jain, D., Nickels, B.E., Sun, L., Hochschild, A., Darst, S.A., 2004. Structure of a Ternary Transcription Activation Complex. *Mol. Cell* 13, 45–53.
- Kabsch, W., 2010. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, 125–132.
- Karplus, P.A., Diederichs, K., 2012. Linking crystallographic model and data quality. *Science* (80-.). 336, 1030–1033.
- Krissinel, E., Henrick, K., 2007. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* 372, 774–797.
- Luscombe, N.M., Laskowski, R.A., Thornton, J.M., 1997. NUCPLOT: A program to generate schematic diagrams of protein-nucleic acid interactions. *Nucleic Acids Res.*
- Marlovits, T.C., Kubori, T., Sukhan, A., Thomas, D.R., Galán, J.E., Unger, V.M., 2004. Structural insights into the assembly of the type III secretion needle complex. *Science* (80-.). 306, 1040–1042.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674.
- McPhillips, T.M., McPhillips, S.E., Chiu, H.J., Cohen, A.E., Deacon, A.M., Ellis, P.J., Garman, E., Gonzalez, A., Sauter, N.K., Phizackerley, R.P., Soltis, S.M., Kuhn, P., 2002. Blu-Ice and the distributed control system: Software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron Radiat.* 9, 401–406.
- Nickels, B.E., Dove, S.L., Murakami, K.S., Darst, S.A., Hochschild, A., 2002. Protein-protein and protein-DNA interactions of $\sigma 70$ region 4 involved in transcription activation by λ cl. *J. Mol. Biol.* 324, 17–34.

- Orlova, E. V., Gowen, B., Dröge, A., Stiege, A., Weise, F., Lurz, R., Van Heel, M., Tavares, P., 2003. Structure of a viral DNA gatekeeper at 10 Å resolution by cryo-electron microscopy. *EMBO J.* 22, 1255–1262.
- Pabo, C.O., Lewis, M., 1982. The operator-binding domain of λ repressor: Structure and DNA recognition. *Nature* 298, 443–447.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Pinkett, H.W., Shearwin, K.E., Stayrook, S., Dodd, I.B., Burr, T., Hochschild, A., Egan, J.B., Lewis, M., 2006. The structural basis of cooperative regulation at an alternate genetic switch. *Mol. Cell* 21, 605–615.
- Sauer, R.T., Jordan, S.R., Pabo, C.O., 1990. λ Repressor: A model system for understanding protein–DNA interactions and protein stability. *Adv. Protein Chem.* 40, 1–61.
- Seneor, D.F., Laue, T.M., Eaton, S., Ross, J.B.A., Waxman, E., Rusinova, E., 1993. The Primary Self-Assembly Reaction of Bacteriophage λ cI Repressor Dimers Is to Octamer. *Biochemistry* 32, 6179–6189.
- Shearwin, K.E., Dodd, I.B., Egan, B.J., 2002. The helix-turn-helix motif of the coliphage 186 immunity repressor binds to two distinct recognition sequences. *J. Biol. Chem.* 277, 3186–3194.
- Stayrook, S., Jaru-Ampornpan, P., Ni, J., Hochschild, A., Lewis, M., 2008. Crystal structure of the λ repressor and a model for pairwise cooperative operator binding. *Nature* 452, 1022–1025.
- Tilley, S.J., Orlova, E. V., Gilbert, R.J.C., Andrew, P.W., Saibil, H.R., 2005. Structural basis of pore formation by the bacterial toxin pneumolysin. *Cell* 121, 247–256.
- Tina, K.G., Bhadra, R., Srinivasan, N., 2007. PIC: Protein Interactions Calculator. *Nucleic Acids Res.* 35.
- Wang, H., Dodd, I.B., Dunlap, D.D., Shearwin, K.E., Finzi, L., 2013. Single molecule analysis of DNA wrapping and looping by a circular 14mer wheel of the bacteriophage 186 CI repressor. *Nucleic Acids Res.* 41, 5746–5756.
- Weiss, M.S., Hilgenfeld, R., 1997. On the use of the merging R factor as a quality indicator for X-ray data. *J. Appl. Crystallogr.* 30, 203–205.
- Young, H.S., Dang, H., Lai, Y., DeRosier, D.J., Khan, S., 2003. Variable symmetry in *Salmonella typhimurium* flagellar motors. *Biophys. J.* 84, 571–577.

5 CONSTRUCTION OF TEMPERATURE SENSITIVE EXPRESSION SYSTEMS

The following chapter describes the development of new temperature sensitive expression systems for *E. coli*. In this manuscript, we characterize temperature sensitive mutants of the 186 CI immunity repressor and CI-hybrid repressor. These proteins were structurally characterized in Chapter 3 and Chapter 4, respectively. Two new temperature sensitive expression systems were developed based around these repressors and temperature sensitive expression was successfully demonstrated. This manuscript was written with the intention to submit to Journal of Biological Engineering or ACS Synthetic Biology. I performed all the experiments described, analysed the data, prepared figures for publication and wrote the manuscript. This manuscript has not yet been submitted.

Statement of Authorship

Title of Paper	Engineering temperature sensitive expression modules with bacteriophage repressors
Publication Status	<input type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input checked="" type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	This draft manuscript characterizes two temperature sensitive transcriptional repressors. A temperature sensitive module was constructed using these transcriptional repressors and temperature sensitive gene induction was demonstrated.

Principal Author

Name of Principal Author (Candidate)	Jia Quyen Truong		
Contribution to the Paper	The principal author performed all the experiments in this manuscript. They also performed data analysis, prepared figures, and drafted the manuscript.		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	16/12/2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Ian Dodd		
Contribution to the Paper	Assisted in the design of experiments, contributed reagents and analytical tools, analysis of data and the preparation of the manuscript		
Overall percentage (%)	10		
Signature		Date	16/12/2020

Name of Co-Author	Keith Shearwin		
Contribution to the Paper	Assisted in the design of experiments, contributed reagents and analytical tools, analysis of data and the preparation of the manuscript		
Overall percentage (%)	10		
Signature		Date	16/12/2020

ENGINEERING TEMPERATURE SENSITIVE EXPRESSION MODULES WITH BACTERIOPHAGE REPRESSORS

Jia Truong¹, Ian Dodd¹, Keith Shearwin¹

¹Department of Molecular and Cellular Biology, School of Biological Sciences, University of Adelaide, Adelaide, South Australia 5005, Australia

*Corresponding email: keith.shearwin@adelaide.edu.au

Abstract

The λ cI857- p_{RM} - p_R system provides temperature-induced gene expression, relying on a temperature-sensitive mutant of the λ cI protein. Two alternative temperature-sensitive transcriptional repressors are characterized in this study. The CI immunity repressor from bacteriophage 186 also has a thermolabile mutant CI^{G22R}. A lacZ *in vivo* reporter gene assay showed that the protein tightly repressed its cognate promoters at 30°C, but lost activity at temperatures 37°C-40°C. A temperature-sensitive chimeric repressor containing domains from both the λ and 186 CI proteins is also described. This protein, the hybrid^{ts}, allows for temperature-controlled repression of the λ p_R promoter and responds more rapidly to elevated temperatures than cI857. Two self-contained expression systems 186 CI^{G22R}- p_R - p_L and hybrid^{ts}- p_{RM} - p_R , analogous to the cI857 system, showed comparable temperature-induced gene expression to the cI857 system. At 37°C, both the CI^{G22R} and hybrid^{ts} systems display high promoter activity compared to λ cI857, which may be desirable for applications that wish to avoid initiation of the *E. coli* heat shock response.

Introduction

Inducible promoters have been extensively employed for regulated gene expression in bacteria for applications such as protein overexpression, synthetic gene circuits, and controlling metabolic pathways. Several temperature-sensitive expression systems have been developed that use elevated temperatures as the signal to induce gene expression (Hasan and Szybalski, 1995; Terpe, 2006). These expression systems avoid several of the major pitfalls of chemically inducible systems, which use an added ligand as the induction signal. Inducers such as IPTG (used in the lac repressor system (Gronenborn, 1976; Simons et al., 1987)) can be expensive which limits their application for large scale production. For the industrial production of therapeutic products, chemical inducers such as IPTG and antibiotics also present a potential biohazard and must be removed from the final product (Andrews et al., 1996; Lee and Keasling, 2008). In addition, thermo-regulated systems are scalable to

large fermenters as culture temperatures are externally adjustable, with an associated reduced risk of contamination from the addition of inducer solution.

Many temperature inducible systems employ a thermolabile mutant of the λ immunity repressor *ci* called *ci857* (Lieb, 1966; Valdez-Cruz et al., 2010). Applications of *ci857*-controlled gene expression include protein overexpression (Cheng et al., 1984; Yoakum et al., 1982), toggle switches (Gardner et al., 2000) and chromosomal integration of DNA (Haldimann and Wanner, 2001). The *ci857* protein is active and effectively represses the λ p_L and p_R promoters at permissive temperatures around 28-30°C. Raising the temperature to 40-42°C inactivates the protein, allowing p_R and p_L to exhibit strong promoter activity (Sussman and Jacob, 1962; Villaverde et al., 1993).

A common implementation of λ *ci857* in a inducible expression system is the λ *ci857*- p_{RM} - p_R expression system (Jechlinger et al., 2005), which utilizes the natural bacteriophage λ genomic architecture. The gene of interest is placed under the control of the λ p_R promoter. The divergently orientated λ p_{RM} promoter drives the expression of *ci857*. Between these promoters lie the O_{R1} , O_{R2} and O_{R3} operators, which regulate the activities of these two promoters (Meyer and Ptashne, 1980; Ptashne, 1986). A λ *ci* repressor tetramer binding to O_{R1} and O_{R2} represses p_R and activates p_{RM} through stabilizing contacts with RNA polymerase positioned at that promoter. When *ci* levels are high, the repressor binds to the low affinity O_{R3} to reduce repressor production (Meyer and Ptashne, 1980; Ptashne et al., 1980). In this setup, *ci857* positively and negatively autoregulates its own levels and is inactivated at high temperatures to allow expression of the gene of interest.

λ *ci857* systems has one main drawback for production-based applications. Effective inactivation of λ *ci857* requires temperatures above 37°C which places undesirable physiological stresses on *Escherichia coli*-based systems. The heat stress can exacerbate protein misfolding during expression, leading to increased inclusion body formation and proteolysis (Georgiou et al., 1994; Schein, 1991). The bacterial heat shock response is also triggered, which can also negatively impact the quantity, quality and stability of protein produced. This response, under the control of the alternative signal factor σ_{32} (Grossman et al., 1984), imposes a metabolic burden by diverting resources towards the synthesis of molecular chaperones and heat shock proteins (Hoffmann and Rinas, 2001). For example, transferring cultures from 30°C to 42°C for 5 minutes resulted in the upregulation of five representative heat shock proteins (such as GroEL) to the point they constitute about 20% of intracellular protein (Yamamori et al., 1978). Upregulated protein synthesis under heat shock can deplete intracellular amino acid pool, upregulate protease activity, and trigger the stringent response. Depending on the protein, the upregulated proteases could be detrimental to obtaining an intact product. In the stringent response, translational machinery is downregulated to limit protein production capacity. Thus, it would be useful to have inducible expression systems that respond to lower temperatures, where the promoters are active at 37°C.

Additionally, many artificial gene circuits employ the same set of transcriptional regulators, namely, tetR, lacI, λ cl and araC (Brödel et al., 2016; Elowitz and Leibler, 2000; Gardner et al., 2000; Singh, 2020; Stricker et al., 2008). cl857 cannot be utilized in such systems already expressing λ cl, as the non-temperature sensitive repressors would repress the promoters at the restrictive temperatures. Alternative temperature inducible systems that do not cross-react with these commonly used transcriptional regulators would expand the toolkit available for building increasingly complex gene circuits.

Thus, a thermo-regulated inducible system, unrelated to the common set of transcriptional regulators, that can achieve high induction of gene expression without invoking heat stress responses would be valuable. We hypothesized that 186 CI^{G22R} may be a suitable candidate repressor to use in such a system. Originally isolated by Baldwin et al (1966), 186 CI^{G22R} is a temperature sensitive mutant of the bacteriophage 186 immunity repressor, with a glycine 22 to arginine substitution, that represses its cognate 186 p_R promoter at permissive temperatures (Dodd et al., 1990; Lamont et al., 1993). It is further hypothesized that due to the highly cooperative nature with which 186 CI binds its operator DNA (Dodd and Egan, 1996), it may be more rapidly inactivated by temperature upshifts.

λ cl and 186 CI both consist of two structurally independent domains, a N-terminal DNA binding helix-turn-helix (HTH) containing domain and a C-terminal oligomerization domain. Pinkett et al. (2006) has previously created a hybrid repressor protein by replacing the oligomerization domain of λ CI, which can form octamers, with the oligomerization domain of 186 CI, which can form even higher order structures. This chimeric repressor protein is proposed to cooperatively oligomerize like 186 CI and yet remains able to repress the lambda p_R promoter. Thus, replacing the oligomerization domain of λ cl857 with the 186 CI oligomerization domain may create a highly cooperative temperature sensitive repressor with a lower restrictive temperature, than λ cl857.

In this study, we constructed two thermo-regulated systems based on these two repressors and demonstrate that they are practical alternatives to the λ cl857-p_{RM}-p_R system. These systems have different temperature response profiles, which could be beneficial when optimizing applications that utilize the λ cl857-p_{RM}-p_R inducible expression system.

Results

186 CI^{G22R} gradually loses activity from 30°C to 40°C

To characterize the activity of 186 CI^{G22R} at different temperatures, a series of lacZ reporter assays were conducted. The CI repressor and its temperature sensitive variant were expressed from an IPTG inducible p_{lac} promoter on a low copy number plasmid. The ability of the 186CI protein variants to repress the 186 p_R promoter were tested in the reporter strain KS980 (Figure 5-1A). This strain

contains the lacZ reporter gene under the control of the 186 pR promoter, without the F_L and F_R flanking sites that normally enhance the efficiency of CI repression on 186 pR (Dodd and Egan, 2002).

A range of IPTG concentrations between 0-200 μM was used to produce a range of repressor levels *in vivo*. As expected, wildtype 186 CI repressor could efficiently repress 186 pR, reducing pR activity to 1-2% of unrepressed promoter activity levels at the 200 μM IPTG at each of the tested temperatures. In contrast, a gradual loss of repression of pR by 186 CI^{G22R} was observed as the temperature increased (Figure 5-1B). The slightly reduced promoter activity in the absence of IPTG is consistent with incomplete LacI repression of the plasmid p_{lac} (Penumetcha et al., 2010). At 30°C, CI^{G22R} shows comparable activity to 186 CI, repressing pR to 14 lacZ units at 200 μM IPTG. At 37°C and 40°C, there is a substantial drop in the observed pR repression by CI^{G22R} to 150 and 290 lacZ units, respectively, at 200 μM IPTG. While these results confirm the temperature sensitive nature of the G22R mutation, some activity remains at 40°C.

The effect of the G22R mutation on a C-terminally His₆-tagged version of the 186 CI was also tested. This protein has a C-terminal thrombin cleavage tag followed by a hexahistidine tag and was previously studied by Pinkett et al (2006), where they solved the crystal structure of the protein. The G22R mutation also confers thermolability on the tagged protein (Figure 5-2B). The untagged protein displays increased repression activity over the tagged version, even at 30°C, suggesting the His₆ tag partly impairs protein function or reduces protein levels *in vivo*.

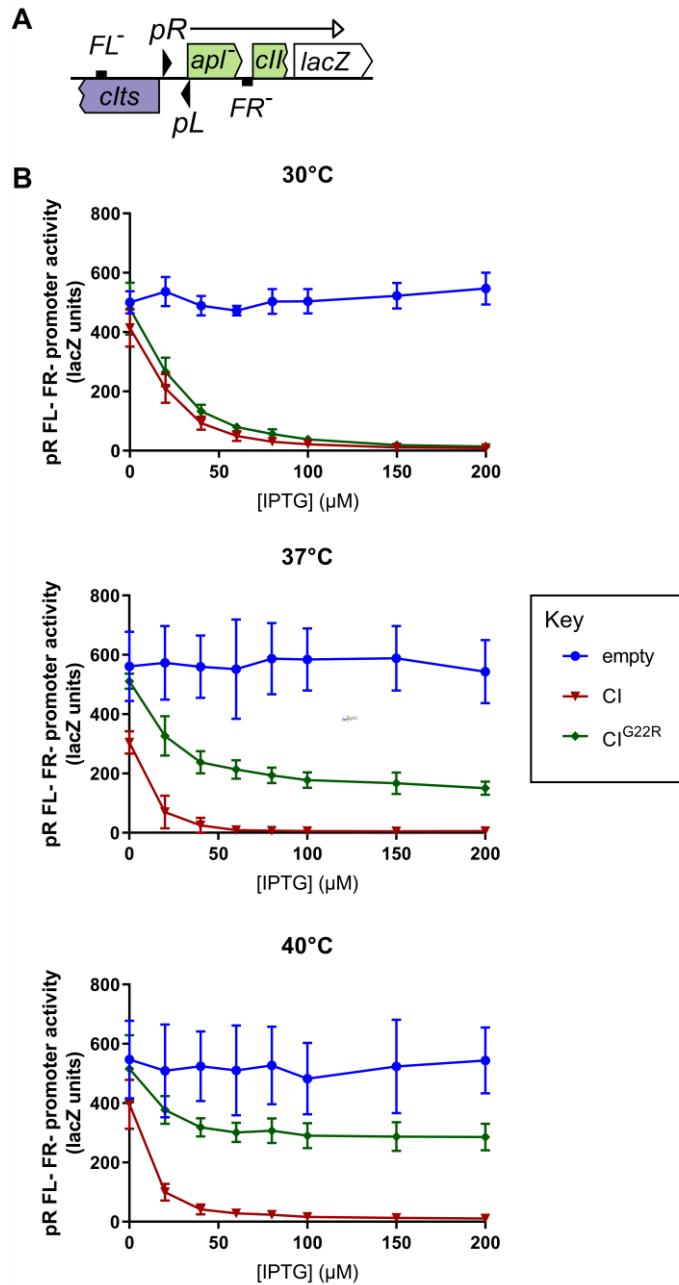


Figure 5-1 – *In vivo* lacZ reporter assays demonstrate that 186 CI^{G22R} is a thermolabile protein that gradually loses the ability to repress its cognate promoter 186 pR as temperature increases from 30°C to 40°C. A) pR:lacZ reporter strain used to measure activity of 186 CI protein variants, KS980 = NK7049 (λ RS45lacZ Δ YA-pBC2-HincII-SnaBI FL⁻pR⁺pL⁺FR⁻) carries a chromosomally integrated single copy of the pR:lacZ fusion (Dodd and Egan, 2002). Diagram based on figure by Pinkett et al (2006). B) At 30°C, 37°C and 40°C, 186 CI represses pR FL- FR⁻ efficiently. At 30°C (top), 186 CI^{G22R} shows similar pR repression as wildtype CI. At 37°C (middle) and 40°C (bottom), 186 CI^{G22R} markedly reduced repression of pR compared to wildtype CI, with greater loss of function observed at 40°C. The sample sizes for empty vectors and CI variants were $n \geq 6$ and $n \geq 8$, respectively.

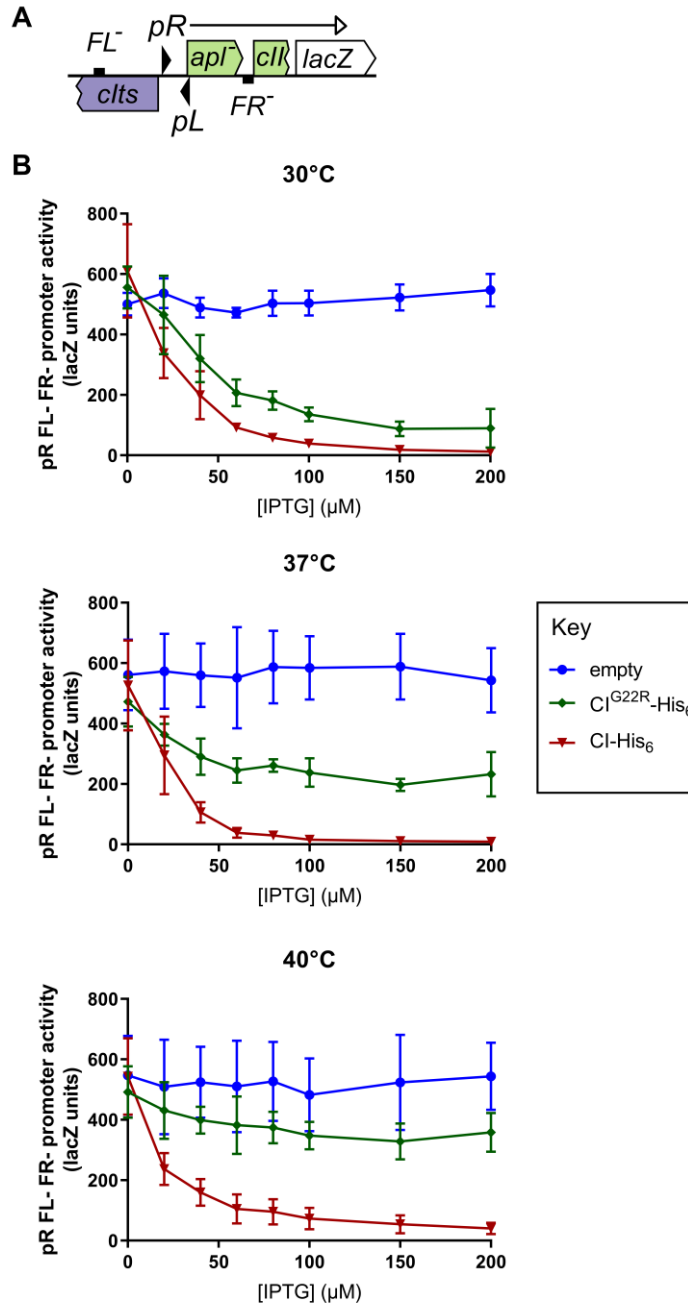


Figure 5-2 - The G22R amino acid substitution makes a His₆-tagged 186 Cl thermolabile, with reduced pR repression as temperature increases from 30°C to 40°C. A) pR:lacZ reporter strain used to measure activity of 186 Cl protein variants, KS980 = NK7049 (λ RS45lacZΔYA-pBC2-HincII-SnaBI FL⁻pR⁺pL⁺FR⁻) carries a chromosomally integrated single copy of the pR:lacZ fusion (Dodd and Egan, 2002). Diagram based on figure by Pinkett (2006). B) 186 Cl^{G22R}-His₆ loses activity in the order of 30°C (top), 37°C (middle) and 40°C (Bottom). Data error bars are 95% confidence intervals. The sample sizes for empty vectors and Cl variants were $n \geq 6$ and $n \geq 8$, respectively.

The observed loss of function of Cl^{G22R} *in vivo* may arise from changes in Cl^{G22R} production or degradation in the cell in response to elevated temperatures. A reduction in the steady state levels of the Cl^{G22R} protein in the cell would lead to pR derepression. To investigate this possibility, the steady state levels of thermolabile and wildtype Cl were quantified *in vivo* via western blotting (Figure 5-3). At all temperatures tested between 30°C-40°C, 186 Cl and Cl^{G22R} can be observed on the blot. Cl^{G22R}

did not show reduced levels relative to wildtype as temperature increased. This result suggests that CI^{G22R} inactivation is not based around proteolytic degradation.

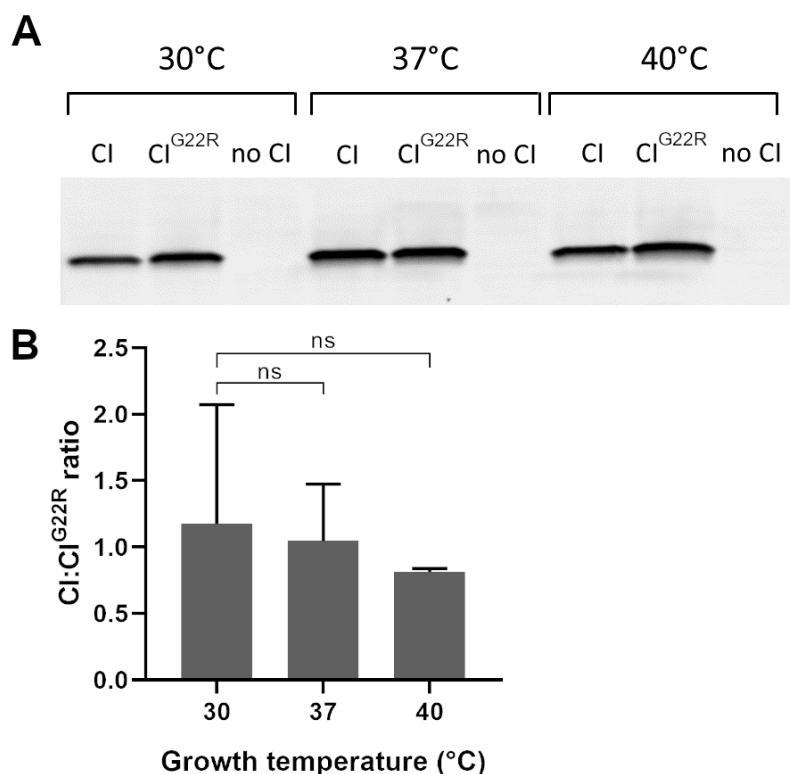


Figure 5-3 –Western blot analysis of 186 CI and 186 CI^{G22R} steady state levels in pR FL- FR- reporter strains. A) Representative western blot image showing the level of expression of CI and CI^{G22R} . B) Densitometric analysis showing the $CI:CI^{G22R}$ ratios at 30°C, 37°C and 40°C. CI^{G22R} levels do not preferentially decrease relative to wildtype at higher temperatures. Sample size was $n=3$ and error bars represent the 95% confidence interval. Changes in $CI:CI^{G22R}$ ratios between the temperatures were found to be not statistically significant. Significance was evaluated by a paired one-tailed Student's t -test on the \log_{10} transformed data.

186 CI^{G22R} - p_R - p_L system provides comparable temperature-inducible gene expression to λ $cl857$ - p_{RM} - p_R system

The natural genomic architecture of 186 was used as a model to design an inducible expression system equivalent to the λ $cl857$ - p_{RM} - p_R system. In both bacteriophage λ and 186, the immunity repressor is expressed from the lysogenic promoter λ p_{RM} and 186 p_L , respectively. Both repressors act to repress transcription from a lytic promoter that drives transcription in the opposite direction as the lysogenic promoter. These promoters, λ p_L and 186 p_R , respectively, are situated close to the lysogenic promoter. A key difference between the two bacteriophages is that 186 uses a convergent promoter arrangement whilst λ uses divergent promoters.

In this 186 CI^{G22R} - p_R - p_L system, the expression of 186 CI^{G22R} is driven by the 186 p_L promoter. Raising the temperature of the system should inactivate 186 CI^{G22R} and allow expression of the gene of interest from 186 p_R . Furthermore, promoter activity from p_R would reduce 186 p_L activity due to transcriptional interference (Callen et al., 2004), further reducing 186 CI^{G22R} gene activity. To test this proposed system, a genome-integrated reporter was constructed (Figure 5-4A). In this reporter, the

natural ci - p_R - p_L - apl layout of the 186 p_R and p_L promoters is retained, including the F_L/F_R flanking sites. F_L and F_R are CI operators ~300 bp either side of the 186 p_R and p_L promoters that aid repression of p_R by CI , and increase the activity of p_L . In this arrangement, 186 CI^{G22R} is under the control of p_L and a $lacZ$ reporter gene is placed downstream of p_R , following the apl gene mutated to inactivate its DNA binding function (apl^{HTH-}), and avoid Apl repression of 186 p_R and p_L (Cutts et al., 2020; Dodd et al., 1993). A control reporter carrying a ci^{HTH-} mutant was also constructed.

An analogous reporter construct was made to test λ $ci857$ in the same system (Figure 5-4B). The thermolability of λ $ci857$ arises from a single amino acid substitution A66T in the N-terminal domain of λ ci (Lieb, 1981). In the λ reporter construct, the natural λ ci - p_{RM} - p_R - cro layout is retained, with a $lacZ$ reporter gene placed downstream of cro and thus under the control of the p_R promoter (the λ $tR1$ terminator is removed). The cro gene, cro^{HTH-} , encodes a mutant λ cro protein defective in binding DNA, to prevent it from repressing p_R and p_{RM} (Schubert et al., 2007). The $lacZ$ reporter gene has the same ribosome binding site as the $lacZ$ reporter gene in the 186 CI^{G22R} construct to keep the constructs comparable. λ $ci857$ repressor is under the control of p_{RM} . The O_L operators, normally 2.3 kb upstream of p_R , were included to allow for λ CI positive and negative autoregulation via DNA looping (Dodd et al., 2004, 2001; Lewis et al., 2011).

Both the 186 CI^{G22R} - p_R - p_L and λ $ci857$ - p_{RM} - p_R systems showed tight repression of the 186 and λ p_R promoters at 30°C, and strong temperature induction in the *in vivo* reporter gene assays (Figure 5-4A and B). Maximal induction occurred at the highest temperature tested (40°C), with all constructs measuring >88% unrepressed promoter activity (Figure 5-4C). $lacZ$ expression from λ p_R was stronger than 186 p_R at all the temperatures tested, possibly reflective of intrinsic differences in λ p_R and 186 p_R promoter strengths. Both unrepressed promoter activities were observed to increase in activity with temperature. This effect was more pronounced with the 186 p_R promoter, which increased from 62 $lacZ$ units at 30°C to 124 units at 40°C.

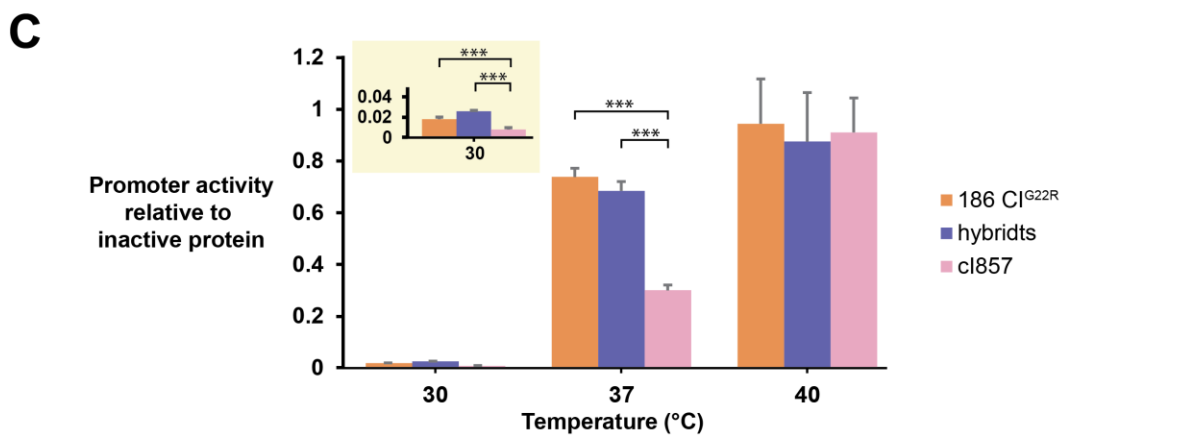
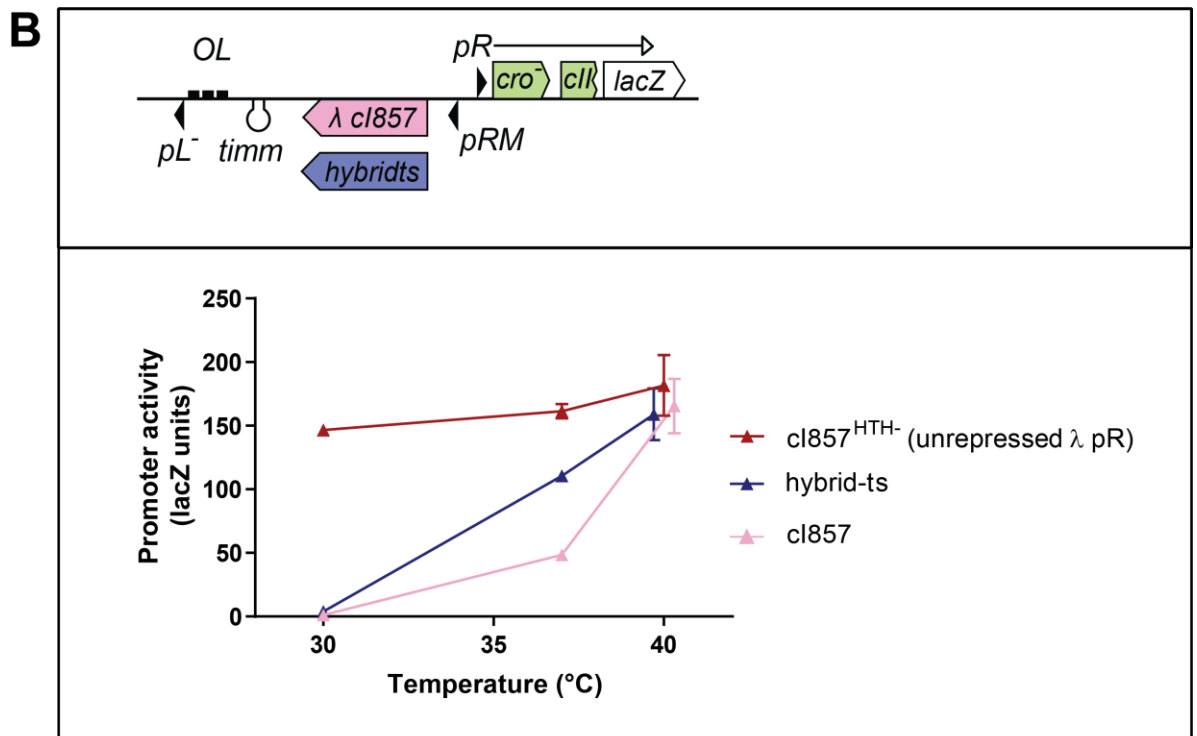
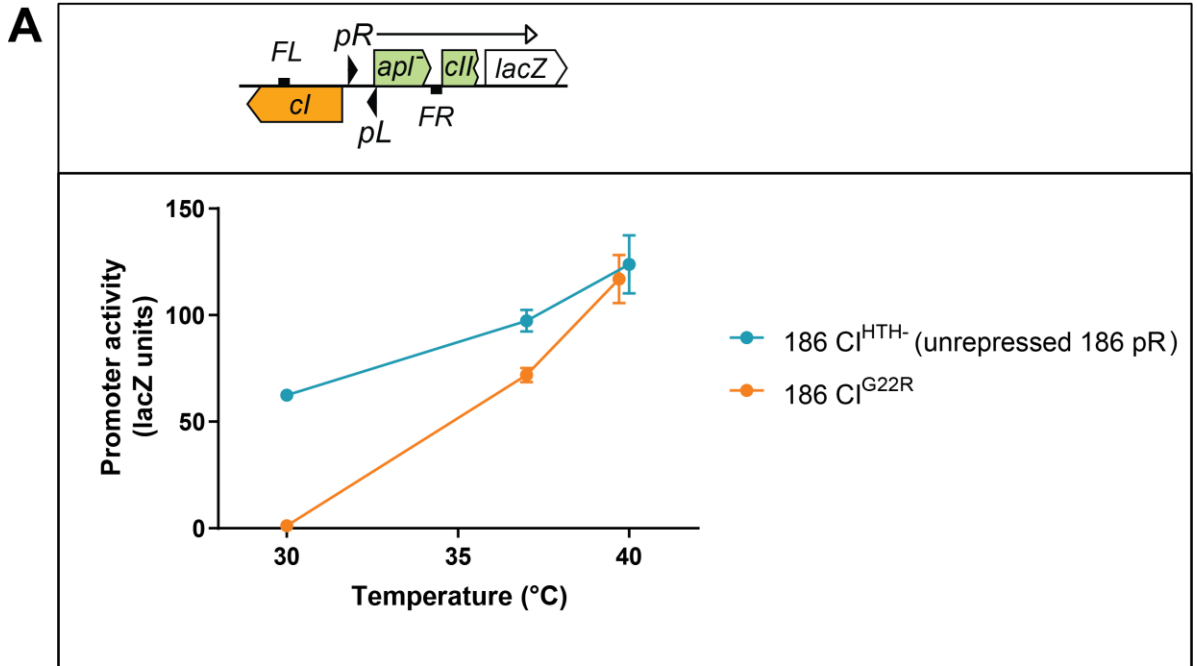


Figure 5-4 –186 CI^{G22R} - p_R - p_L , λ cI857- p_{RM} - p_L , and hybrid^{ts}- p_{RM} - p_R constructs display temperature inducible promoter expression as temperature increases from 30°C to 40°C.

A) Top: 186 p_R :lacZ reporter constructs was created to measure the temperature inducibility of a 186 CI^{G22R} - p_R - p_L system. The natural arrangement of 186 p_R - p_L is retained, with CI^{G22R} under the control of 186 p_L . 186 CI^{HTH} - (186 $CI^{S37R/S38E}$) contains mutations in the HTH motif to abolish DNA binding, to allow measurement of unrepressed 186 p_R activity. Bottom: 186 p_R in the 186 CI^{G22R} construct displays tight repression at 30°C (1.1 lacZ units), with promoter induction temperature increases to 40°C.

B) Top: λ p_R :lacZ reporter constructs were created to measure the of temperature inducibility of λ cI857- p_{RM} - p_L and hybrid^{ts}- p_{RM} - p_R systems. These constructs use the natural arrangement of p_{RM} - p_L from λ . λ cI857^{HTH}-(λ cI857^{Q34R/S34E}) contains mutations in the HTH motif that abolish DNA binding, allowing measurement of unrepressed p_R activity. Bottom: λ p_R in the cI857 and hybrid^{ts} construct shows temperature dependent derepression as temperature increases from 30°C to 40°C. At 30°C, cI857 and hybrid^{ts} constructs measured 1.2 and 3.8 lacZ units, respectively. The hybrid^{ts} shows 2.3 fold greater derepression at 37°C than λ cI857.

Error bars in (A) and (B) represent 95% confidence interval ($n \geq 6$ for unrepressed promoter measurements and $n \geq 8$ for the λ 186 CI^{G22R} , λ cI857 and hybrid^{ts} measurements). Data points at 40°C with overlapping error bars have been staggered to improve readability.

C) Ratio of promoter expression by CI^{G22R} , λ cI857 and hybrid^{ts} constructs to inactive repressor constructs. All three constructs are well repressed at 30°C and significantly derepressed at 40°C. At 37°C, the 186 CI^{G22R} and hybrid^{ts} constructs display substantially more derepression than the cI857 construct. Error bars represent 95% confidence intervals, calculated through the error propagation from measurements from (A) and (B). Statistical significance was evaluated by a paired one-tailed Student's t-test on the \log_{10} transformed data (***) denotes $p < 0.001$.

Importantly, the 186 CI^{G22R} constructs produced significantly greater promoter activity at 37°C compared to the cI857 construct. This was true in terms of absolute units and relative to unrepressed promoter activity (Figure 5-4A and B).

Hybrid^{ts} protein provides temperature-controlled expression in p_{RM} - p_R systems

The λ -186 CI-hybrid repressor described by Pinkett et al. (2006), is a fusion between the N-terminal DNA-binding domain of λ CI (amino acids 1-92) to the C-terminal oligomerization domain of 186 CI (amino acids 83-192). This protein was shown to repress λ p_R and activate λ p_{RM} , with both processes enhanced by O_L (Pinkett et al., 2006). It was hypothesized that adding the cI857 amino acid substitution to the λ CI domain would confer temperature sensitivity to the CI-hybrid repressor. To test if this protein, named the hybrid^{ts} repressor, could be used to provide temperature induced protein expression, the λ p_R reporters were modified by replacing the λ cI857 gene with the hybrid^{ts} gene (Figure 5-4B).

Surprisingly, the hybrid^{ts} repressor reporter behaved more similarly to the 186 CI^{G22R} reporter in its temperature response than the λ cI857 repressor, whose mutation it carries, being more than 60% induced at 37°C (Figure 5-4C). Its fractional induction at 40°C was similar to the other repressors, while repression at 30°C was not as tight either of the other repressors (Figure 5-4C).

Discussion

New temperature inducible systems have practical applications

The use of λ cI857 for temperature inducible expression has two main advantages. The first advantage is induction can easily be controlled. Increasing the temperature from 30°C to 42°C in an incubator or fermenter is sufficient to achieve maximal induction. This makes temperature induction scalable for large scale commercial applications. One caveat is that in large scale bioreactors, the rate of temperature change in the culture can be throttled by heat transfer limitations, as heating rate decreases as the fermenter size increases (Caspeta et al., 2009). Temperature inducible expression is particularly relevant for industrial and pharmaceutical applications as it does away with expensive and/or toxic inducer compounds. Furthermore, when fully derepressed, p_R and p_L act as strong promoters. The second advantage is that λ cI857 works in a wide variety of bacterial systems. Besides *E. coli*, λ cI857 has been tested and shown to work in many bacterial species including *Bacillus subtilis*, several species of *Pseudomonas*, *Salmonella enterica*, *Klebsiella pneumoniae* and *Serratia rubidae* (Breitling et al., 1990; Jawale et al., 2012; Winstanley et al., 1989).

Different applications have different temperature requirements. This is an important consideration when deploying a temperature inducible system. As an example, bacterial ghost vaccine candidates have been generated via λ cI857-based temperature inducible expression of the phiX174 lysin gene E in bacteria (Jawale et al., 2012; Jechlinger et al., 2005). However, this approach, when using the traditional λ cI857- p_{RM} - p_R requires growing cells below the permissible temperature of 30°C which may be suboptimal or not possible for organisms such as *Helicobacter pylori*. To address the needs of this application, Jechlinger et al. (2005) developed an altered p_{RM} - p_R promoters that could be extremely well repressed up to 37°C and 39°C. However, some applications have the opposite temperature requirement. Overexpression of protein in *E. coli* can often result in proteins sequestered into insoluble inclusion bodies. Higher induction temperatures can favor aggregation, with lower induction temperatures often promoting soluble protein expression (Kaur et al., 2018; Sørensen and Mortensen, 2005). Physiological responses from heat shock can also further compromise yield (Valdez-Cruz et al., 2010). Biosynthetic fermentations often benefit from growth at lower temperatures, as enzymes involved in the fermentation can have a shorter half-life at higher temperatures, causing decreased yields (Jiang et al., 2013). Thus, in these cases, a higher degree of de-repression at lower temperatures and avoiding heat shock would be advantageous. To answer the needs of applications requiring lower restrictive temperatures, two alternative temperature sensitive repressors that may be inactivated at lower temperatures were investigated. These repressors were a mutant of the 186 immunity repressor, 186 Cl^{G22R}, and a hybrid^{ts} repressor.

The activity of 186 CI^{G22R} was measured *in vivo* via a lacZ reporter assay. The repressor expressed from a plasmid under the control of an IPTG inducible promoter. This protein is active at 30°C and gradually loses activity as temperature increases to 40°C. The dynamic range of the p_R FL- FR-promoter (ratio of induced to uninduced promoter activity) in this setup is 20.3 ± 5.6.

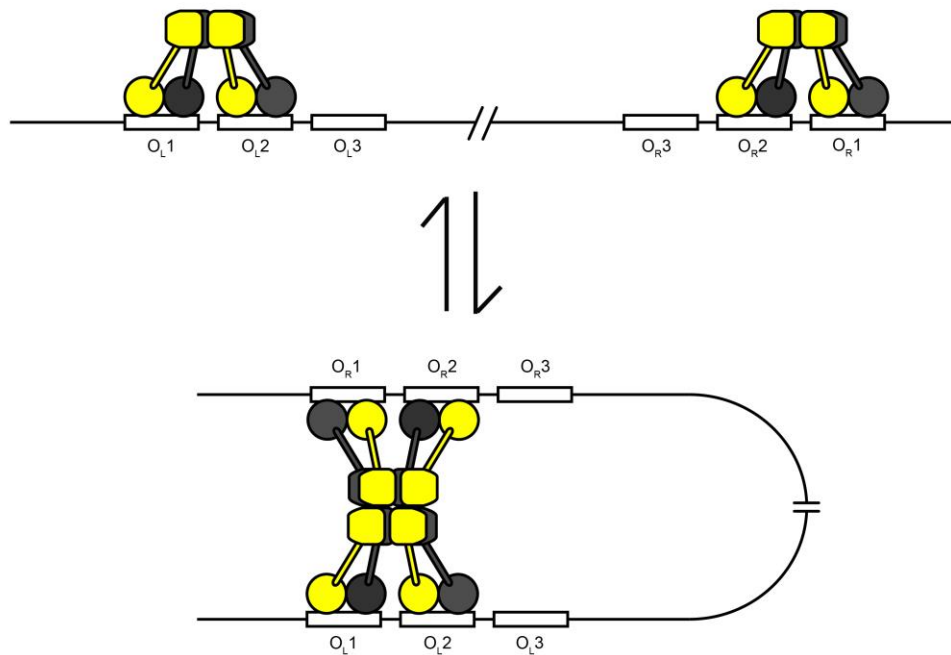
The activities of 186 CI^{G22R}, hybrid^{ts} and λ cI857 were tested in single integrant chromosomal reporter constructs. An advantage of these systems is that all the elements required for heat inducible expression were present on a single DNA construct, simplifying the molecular cloning required to apply these to applications. The λ cI857 and hybrid^{ts} constructs have identical λ p_{RM}-p_R promoter layouts and both proteins contain the λ cI NTD with the same temperature sensitive mutation, yet their temperature response differs. At 30°C, λ cI857 and the hybrid^{ts} repressor keep the λ p_R promoter well repressed at 0.8% and 2.5% of unrepressed p_R levels, respectively. At 37°C, the hybrid^{ts} repressor loses significantly more activity relative to λ cI857, at 68% unrepressed p_R activity compared to 30% unrepressed p_R activity, respectively. At 40°C, the p_R promoter in the λ cI857 and hybrid^{ts} constructs are close to maximum induction levels at 91% and 87% of unrepressed p_R activity.

The hybrid^{ts} protein represses the same p_R promoter and thus can substitute for λ cI857 in many λ p_R-based systems. This can potentially be utilized for the optimization of new and existing temperature inducible systems based around λ cI857. The higher induction levels at 37°C could be useful in systems described earlier that would benefit from avoiding heat shock stress.

186 CI^{G22R} also demonstrated characteristics of a practical temperature sensitive repressor. At its permissive temperature of 30°C, it keeps its cognate promoter 186 p_R well repressed, to levels similar to wildtype protein (Figure 5-1A). It loses activity gradually as the temperature approaches 40°C. When implemented into a single DNA construct, the 186 CI^{G22R}-p_R-p_L system displayed temperature inducible gene expression. Expression from 186 p_R at 30°C is well repressed at 1.1 ± 0.4 lacZ units, comparable to λ cI857-p_{RM}-p_R at 1.2 ± 0.3 lacZ units. At 37°C, the 186 p_R promoter has largely been derepressed and displays 1.5 fold more promoter activity than the λ cI857 system. At 40°C, the 186 p_R promoter is almost completely unrepressed.

186 CI^{G22R} presents a practical alternative to λ cI857 systems. One advantage is that it acts on 186 lytic promoters and can be used in many of the biological systems that already employ λ cI repressor with the λ p_L, p_R and p_{RM} promoters, where λ cI857 could not be used. The 186 CI^{G22R}-p_R-p_L system, like the hybrid^{ts} system, has the same improvement over the λ cI857 system where it can achieve higher levels of induction at 37°C. Both 186 CI^{G22R} and the hybrid^{ts} proteins contain the 186 CI oligomerization domain that mediates the formation of a 12mer wheel, as a hexamer of dimers (unpublished data). This increased rate of loss of activity with increasing temperature in 186 CI^{G22R} and the hybrid^{ts} repressor may be due to differences in oligomerization of the protein (Figure 5-5), where cooperativity is higher for 186 CI-CTD mediated oligomerization.

A λ cI oligomerization



B 186 CI oligomerization λ -186 CI-hybrid oligomerization

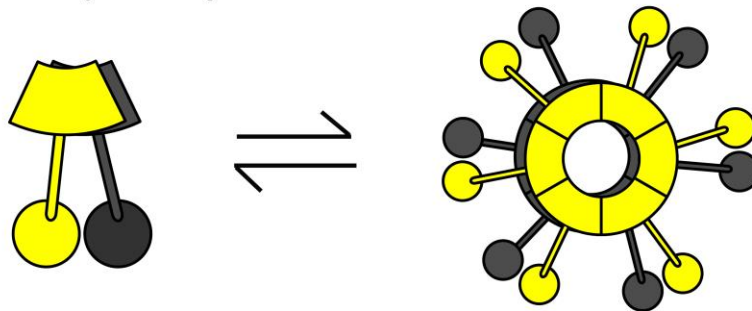


Figure 5-5 – The oligomeric organization of λ cI, 186 CI and λ -186 CI-hybrid repressor. A) λ cI dimers binds to the stronger O_{L1} , O_{L2} , O_{R1} , O_{R2} sites on the DNA. Tetramers form at these sites, and can further form an octamer, mediated by DNA looping. B) 186 CI and the λ -186 CI-hybrid oligomerize via the 186 CI oligomerization domain. These proteins form dimers, followed by subsequent oligomerization into a wheel-like arrangement, proposed to be a hexamer of dimers.

Insights into the mechanisms of temperature sensitivity

This study also provides some insight into the mechanism of thermolabile DNA binding repressor proteins. The N-terminal domain (NTD) and C-terminal domain (CTD) of λ cI thermally unfold independently from each other, with melting temperatures at 54°C and around 73°C, respectively (Pabo et al., 1979; Sauer et al., 1990). The λ cI NTD consists of 5 helices, with helices 2 and 3 forming the DNA-binding HTH motif. The A66T mutation in the N-terminal domain helix 4 (Figure 5-6) lowers the melting temperature of the NTD to 42°C (Hecht et al., 1984). It is believed that the mutation partially disrupts the hydrophobic core of the NTD to increase the thermolability of the domain (Hecht et al., 1984; Sauer et al., 1990). The A66T mutation sits at the end of helix 4 in the DNA binding domain, close to helix 5. Helix 5-helix 5 contacts are involved in the weak dimerization of the NTD, with the CTD mediating strong

dimerization and subsequent high order oligomerization of the protein (Sauer et al., 1990). Thus, conformational changes of the N-terminal domain with the A66T mutation at high temperatures may also impair proper alignment of the HTH motifs for operator binding.

The temperature sensitivity of both the λ cI857 and the hybrid-ts repressors is most consistent with the N-terminal domain unfolding or a loss of correct NTD alignment for binding the λ operators. It seems unlikely that the same mutation acts by causing dissociation of the λ CI CTD and the 186 CI CTD. One possible explanation for the hybrid-ts protein losing activity at lower temperatures than λ cI857 is that λ cI multimerization to tetramers and octamers is facilitated by DNA binding (Dodd et al., 2001; Lewis et al., 2011; Senear et al., 1993). The DNA should select for active dimers from the pool of λ cI857, and may thus effectively select for active DNA-bound tetramers and octamers. Therefore, repression of pR may be able to remain strong in the presence of inactive CI monomers. In contrast, atomic force microscopy of 186 CI suggests that wheel oligomers may form prior to binding DNA (Wang et al., 2013). Thus, it could be possible that incorporation of a smaller proportion of monomers with inactive DNA binding domains would compromise the highly cooperative DNA binding ability of a large proportion of the wheels.

The 186 CI G22 residue lies just after the first helix of the NTD. It is not known why the G22R mutation causes temperature sensitivity, though a reduced melting temperature or impaired NTD dimerization seem likely. Dimerization of the NTD appears weak, as the NTDs were not dimerized in the crystal structure of the full length CI dimer (Pinkett et al., 2006), and NTD association constant was measured to be at least 10^4 weaker than the full length protein (Shearwin et al., 2002). The 186 CI NTD shares structural similarity to the NTD of λ cI. The 186 CI NTD also consists of five helices. Helices 2 and 3, like λ cI, form the HTH motif (Pinkett et al., 2006). Interestingly, the λ cI protein also has a temperature sensitive mutation T22H (Hecht et al., 1984), near the end of the first helix like G22R from the temperature-sensitive mutant of 186 CI (Figure 5-6). These mutations to positively charged amino acids in this area of the domain may moderately disrupt the hydrophobic core of the domain to decrease melting temperatures. If the CI^{G22R} NTD unfolds at 40°C, it does not change the *in vivo* steady state levels of CI, suggesting that the NTD unfolding does not encourage proteolysis.

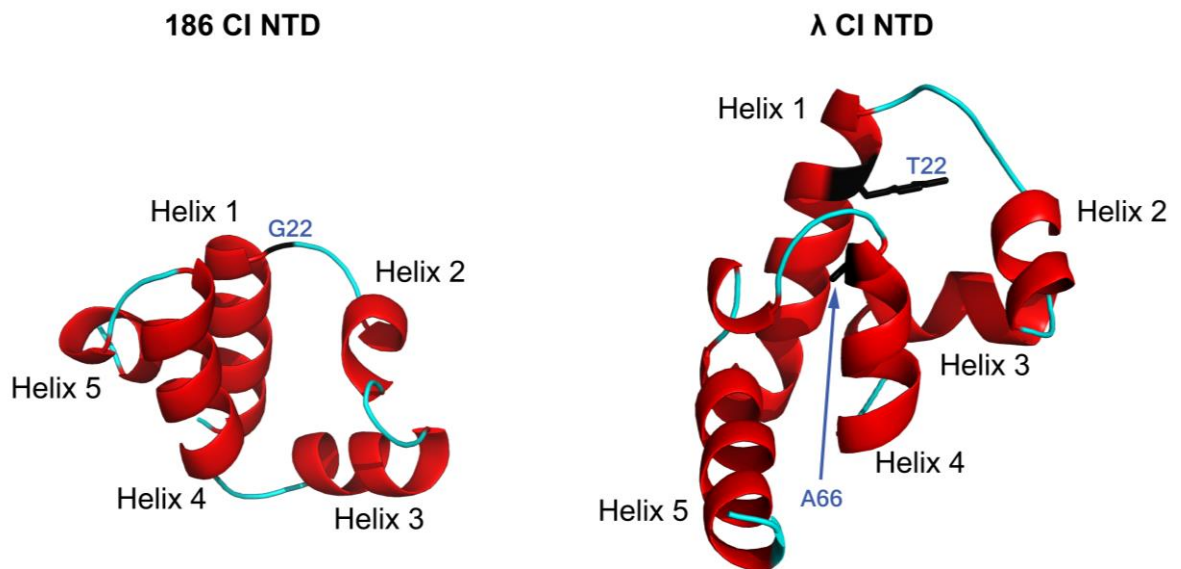


Figure 5-6 – Temperature sensitive mutations mapped onto the structure of the 186 CI NTD (PDB ID: 2FJR; (Pinkett et al., 2006)) and λ CI NTD (PDB ID: 3BDN; (Stayrook et al., 2008)) Mutation sites are shown in black. Structures were visualized In PyMOL (Schrödinger, LLC, 2010).

Despite the knowledge that thermolability can arise from amino acid substitutions that disrupt the hydrophobic core of a protein (Varadarajan et al., 1996), rational design of thermolabile proteins is not always straightforward. TSpred is an online server that uses hydrophobic moment and residue depth into the protein core to predict residues that when mutated may confer temperature sensitivity (Tan et al., 2014). In this study, two buried hydrophobic residues (V148 and V185) in the CTD of 186 CI were chosen by TSpred and reporters carrying mutations at these positions created. However, the resulting 186 CI repressors were functional and were not temperature sensitive (data not shown).

Concluding remarks

Many applications such as overexpression of protein, reprogramming of metabolic pathways and synthetic gene circuits require controllable promoters and the optimization of growth conditions such as temperature. This work has expanded the tunable promoter expression systems available for biologists, allowing for greater fine tuning of promoter activity across the 30-42°C temperature range.

Methods and Materials

Plasmids and bacterial strains

Bacterial strains used in this study are summarized in Table 5-1. Plasmids used in this study are summarized in Table 5-2.

Table 5-1 – *E. coli* strains used in this study.

Strain	Genotype	Notes	Reference
E4300	NK7049 (Δ lacIZYA) \times 74 galOP308 Str ^R Su ⁻	Strain used for lacZ assays with IPTG inducible repressors	Simons et al. (1987)
KS980	E4300(λ Δ YA pBC2 HincII-SnaB1) FL- FR- pR lacZ pUHA-1	Reporter strain with lacZ reporter gene under the control of pR without 186 flanking sites. pUHA-1 plasmid constitutively provides lac repressor for inducible expression of repressor proteins	Dodd and Egan (2002)
E4643	MG1655 rph ⁺ Δ lacIZYA	Self-contained temperature-sensitive modules were integrated into the bacterial chromosome of this strain	Cui et al. (2013)
JT341	E4643 [pIT3_CL-pL- OL_UP_ λ pR pRM_ λ CI857_cro-_lacZ]	Strain to test temperature-sensitive λ ci857-p _R -p _{RM} system Self-contained temperature sensitive module is integrated into the lambda attachment site	This work
JT343	E4643 [pIT3_CL-pL- OL_UP_lambdapRp RM_hybridCIts_cro- _lacZ]	Strain to test temperature-sensitive hybrid ^{ts} -p _R -p _{RM} system Self-contained temperature sensitive module is integrated into the lambda attachment site	This work
JT358	E4643 [pIT3_CL-pL- OL_UP_lambdapRpRM_ λ CI(HTH-)_cro-_lacZ]	Control strain with inactive λ ci857 λ ci857 gene has mutations in the HTH motif to inactivate DNA binding Control module is integrated into the lambda attachment site	This work
JT172	E4643 [pIT3_CL_186_Ci ^{G22R} pR pL apl- lacZ]	Strain to test temperature-sensitive Ci ^{G22R} -p _R -p _L system Self-contained temperature sensitive module is integrated into the lambda attachment site	This work
JT317	E4643 [pIT3_CL_186CI(HTH-)]	Control strain with inactive 186 ci 186 ci gene has mutations in the HTH	This work

	pR pL apl- lacZ]	motif to inactivate DNA binding Control module is integrated into the lambda attachment site	
--	------------------	--	--

Table 5-2 – Plasmids used in this study

Plasmid name	Brief description	Features	Reference
pZS35-186 CI pZS35-186 CI ^{G22R} pZS35-186 CI-His ₆ pZS35-186 CI ^{G22R} -His ₆	Low copy vector to allow IPTG-inducible expression of various CII and CII-derivatives from the p _{lac} promoter Amino acid mutations are written in superscript His ₆ indicates a C-terminal hexahistidine tag was present on the protein	pZ modular plasmid; SC101 origin; chloramphenicol resistance	Lutz and Bujard (1997) and this work
pZS35-empty	Low copy vector used as no CI control in lacZ assays	pZ modular plasmid; SC101 origin; chloramphenicol resistance	Lutz and Bujard (1997)
pUHA-1	Used to constitutively supply lacI from p _{lac} promoter in lacZ assays	p15a ori; kanamycin resistance; wildtype lacI ORF and promoter	H. Bujard, Heidelberg University, Germany
pIT3_CL-pL-OL_UP_λpR pRM_λCI857_cro- _lacZ pIT3_CL-pL-OL_UP_λpRp RM_hybridCIts_cro- _lacZ pIT3_CL-pL-	Self-contained temperature sensitive modules λ CI(HTH-) and 186 CI(HTH-) are repressors that have mutations in the DNA binding motif that render them functionally inactive	R6K-γ ori; chloramphenicol resistance; can be integrated into λ attB	This work

OL_UP_λpRpRM_λCI(HTH-)_cro-_lacZ			
pIT3_CL_186_CI ^{G22R}			
pR pL apl- lacZ			
pIT3_CL_186CI(HTH-)			
pR pL apl- lacZ			

Characterisation of 186 CI^{G22R}

A series of lacZ assays were performed to determine how the repressor activity of 186 CI^{G22R} changes with temperature. *Escherichia coli* strain used for lacZ assays was KS980 (Table 5-1). This reporter strain provided a p_R FL- FR- promoter to test the activity of the 186 CI^{G22R} protein at different temperatures. The CI and 186CI^{G22R} expression plasmids were constructed using the modular pZ system (Lutz and Bujard, 1997). Strains not expressing CI possessed an empty vector. Vectors pZS*35 186 CI, pZS*35 186CI^{G22R} and pZS*35 empty were used to allow inducible expression of 186 CI, 186 CI-His₆, 186 CI^{G22R}, 186 CI^{G22R}-His₆ and no CI under the control of the IPTG inducible p_{lac} promoter.

Western blots of CI^{G22R}

KS981 cells harbouring pZS*35 empty, pZS*35 CI-His₆, pZS*35 CI^{G22R}-His₆ were grown in LB supplemented with 30 µg/mL chloramphenicol, 50 µg/mL kanamycin and 100 µM IPTG at 30°C, 37°C and 40°C to mid-log phase. Cells were lysed using BPER (ThermoScientific) and Benzonase (2.5 units/µL) (Novagen) and cell extracts run on a Bio-Rad Any kDa Criterion TGX gel in Tris-glycine running buffer. Protein bands were transferred to a PVDF membrane using a TransBlot Turbo Transfer system (Bio-Rad) and TransBlot Turbo Transfer Pack PVDF (Bio-Rad). Membranes were blocked using 5% BSA. 186 CI primary detection was with rabbit polyclonal antibody raised against purified 186 CI. Secondary antibody used to probe for 186 CI antibody was goat anti-rabbit IgG Cy5-labelled ECL plex secondary antibody (GE Healthcare). Dried membranes were scanned using the Chemidoc imaging system (Bio-Rad) and images analysed by ImageJ. (Breitling et al., 1990)

Single copy integrants for self-contained temperature based expression system

All constructs were integrated in single copy in the chromosome of E4643 (MG1655 rph⁺ ΔlacIZYA) (Cui et al., 2013), at the λ attB site using the method described in Priest *et al* (2014).

The self-contained 186 CI^{G22R} temperature-based expression module was a pIT3_CL_186_CI^{G22R} pR pL apl- lacZ reporter. The plasmid contained the 186 ci-pR-pL-apl-cII region with a lacZ reporter gene as a translational fusion to the first amino acid of CII. The *apl* gene has been mutated to introduce the E28R, R29E, R33Q mutations into the helix-turn-helix region of Apl to inactivate its DNA binding activity. To measure the unrepressed pR promoter activity of the self-contained module at different

temperatures, 186 CI was substituted with an inactive mutant of 186 CI. This CI mutant CI^{S37R/S38E} had amino acid substitutions in the helix-turn-helix DNA binding motif .

An equivalent temperature-based module was constructed for λ cI857. This was constructed by Gibson assembly of two synthesized G-blocks (IDT) into a BamHI/XmaI-digested pIT3_CL_lacZtrimfuse. One fragment contains the _OL_UP_timm_cI857 sequence, with the natural λ architecture. The other fragment contains cro_pR_pRM_cI-NTD. Additional details are provided in Supplementary Methods.

A temperature sensitive chimeric repressor was designed based on the λ -186 chimeric repressor described by Pinkett et al (2006). The temperature sensitive chimeric repressor is a translational fusion between the λ cI857 NTD (amino acids 1-92) and the 186 CI CTD (amino acids 83-192). The amino acid substitutions responsible for the temperature sensitive characteristic of cI857, A66T, is also present in this chimeric repressor. For the measurement of the unrepresed λ pR promoter in these constructs, cI857 was substituted with cI857^{Q34R/S35E}, which has amino acid substitutions in the helix-turn-helix DNA binding motif.

Kinetic lacZ assays

Kinetic lacZ assays were performed in 96 well microtitre plates as previously described (Dodd et al., 2001), with modifications described here. Growth media conditions for different lacZ experiments are summarized in Table 5-3. The reporter strains were restreaked from glycerol stocks onto selective LB agar media and incubated overnight. Fresh colonies from the restreak plates were used to inoculate 100 μ L of selective LB broth in a well of a 96 well microtitre plate and incubated overnight at 30°C with shaking. After overnight incubation, cells were diluted to OD₆₀₀ 0.003 in 100 μ L fresh selective LB media in a microtitre plate. Plates were sealed and grown at one of three temperatures - 30°C, 37°C and 40°C. When culture OD₆₀₀ reached 0.6-0.9, cultures were assayed for lacZ activity as described by Dodd *et al* (2001).

Table 5-3 – Growth media and growth conditions for lacZ assays

Step	Media
LacZ assays for pZS*35 CI system	
Restreak from glycerol stock	LB agar supplemented with 30 μ g/mL chloramphenicol and 50 μ g/mL kanamycin
Inoculation of cultures from restreak plate	LB broth supplemented with 30 μ g/mL chloramphenicol and 50 μ g/mL kanamycin
Inoculation of cultures for lacZ measurements	LB broth supplemented with 30 μ g/mL chloramphenicol, 50 μ g/mL kanamycin and 0-200 μ M IPTG
LacZ assays for pIT3 integrant constructs	

Restreak from glycerol stock	LB agar supplemented with 30 µg/mL chloramphenicol
Inoculation of cultures from restreak plate	LB broth supplemented with 20 µg/mL chloramphenicol
Inoculation of cultures for lacZ measurements	LB broth supplemented with 20 µg/mL chloramphenicol

Construction of pZS*35 CI, pZS*35 CI^{G22R} and pZS*35 empty

The His₆-tag was removed from the CI-His₆ gene in via a 2-fragment Gibson assembly reaction. Two PCR reactions using pZS*35 186 CI-His₆ as a template were performed using two different pairs of primers 1399 & 1342 and 1400 & 1331 (Table S 1). Primer 1399 was designed to anneal to the CI protein sequence and the vector to loop out the his-tag DNA. Primer 1400 was designed to overlap with 1399 to allow for Gibson assembly. Purified PCR products were assembled into a circular plasmid using Gibson isothermal assembly to create pZS*35 186CI.

pZS*35 186 CI^{G22R} and pZS*35 186 CI^{G22R}-His₆ was made by a two-fragment Gibson assembly reaction. pZS*35 186 CI^{G22R} and pZS*35 186 CI^{G22R}-His₆ minipreps was used as PCR template, respectively. Two PCR reactions using the template were performed using two different pairs of primers 1340 & 1342 and 1329 & 1331 (Table S 1). Primer 1329 contained the mutation to generate the G22R amino acid substitution. Purified PCR products were assembled into a circular plasmids using Gibson isothermal assembly.

References

- Andrews, B., Adari, H., Hannig, G., Lahue, E., Gosselin, M., Martin, S., Ahmed, A., Ford, P.J., Hayman, E.G., Makrides, S.C., 1996. A tightly regulated high level expression vector that utilizes a thermosensitive lac repressor: Production of the human T cell receptor Vβ5.3 in Escherichia coli. *Gene* 182, 101–109.
- Baldwin, R.L., Barrand, P., Fritsch, A., Goldthwait, D.A., Jacob, F., 1966. Cohesive sites on the deoxyribonucleic acids from several temperate coliphages. *J. Mol. Biol.* 17, 343–357.
- Breitling, R., Sorokin, A. V., Behnke, D., 1990. Temperature-inducible gene expression in *Bacillus subtilis* mediated by the c1857-encoded repressor of bacteriophage lambda. *Gene* 93, 35–40.
- Brödel, A.K., Jaramillo, A., Isalan, M., 2016. Engineering orthogonal dual transcription factors for multi-input synthetic promoters. *Nat. Commun.* 7, 13858.
- Callen, B.P., Shearwin, K.E., Egan, J.B., 2004. Transcriptional interference between convergent promoters caused by elongation over the promoter. *Mol. Cell* 14, 647–656.
- Caspeta, L., Flores, N., Pérez, N.O., Bolívar, F., Ramírez, O.T., 2009. The effect of heating rate on escherichia coli metabolism, physiological stress, transcriptional response, and production of

- temperature-induced recombinant protein: A scale-down study. *Biotechnol. Bioeng.* 102, 468–482.
- Cheng, S.C., Kim, R., King, K., 1984. Isolation of gram quantities of EcoRI restriction and modification enzymes from an overproducing strain. *J. Biol. Chem.* 259, 11571–11575.
- Cui, L., Murchland, I., Shearwin, K.E., Dodd, I.B., 2013. Enhancer-like long-range transcriptional activation by λ CI-mediated DNA looping. *Proc. Natl. Acad. Sci. U. S. A.* 110, 2922–2927.
- Cutts, E.E., Barry Egan, J., Dodd, I.B., Shearwin, K.E., 2020. A quantitative binding model for the Apl protein, the dual purpose recombination-directionality factor and lysis-lysogeny regulator of bacteriophage 186. *Nucleic Acids Res.* 48, 8914–8926.
- Dodd, I.B., Egan, J.B., 1996. DNA Binding by the Coliphage 186 Repressor Protein CI. *J. Biol. Chem.* 271, 11532–11540.
- Dodd, I.B., Egan, J.B., 2002. Action at a distance in CI repressor regulation of the bacteriophage 186 genetic switch. *Mol. Microbiol.* 45, 697–710.
- Dodd, I.B., Kalionis, B., Egan, J.B., 1990. Control of gene expression in the temperate coliphage 186: VIII. Control of lysis and lysogeny by a transcriptional switch involving face-to-face promoters. *J. Mol. Biol.* 214, 27–37.
- Dodd, I.B., Perkins, A.J., Tsemitsidis, D., Egan, J.B., 2001. Octamerization of λ CI repressor is needed for effective repression of PRM and efficient switching from lysogeny. *Genes Dev.* 15, 3013–3022.
- Dodd, I.B., Reed, M.R., Egan, J.B., 1993. The Cro-like Apl repressor of coliphage 186 is required for prophage excision and binds near the phage attachment site. *Mol. Microbiol.* 10, 1139–1150.
- Dodd, I.B., Shearwin, K.E., Perkins, A.J., Burr, T., Hochschild, A., Egan, J.B., 2004. Cooperativity in long-range gene regulation by the λ CI repressor. *Genes Dev.* 18, 344–354.
- Elowitz, M.B., Leibier, S., 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338.
- Gardner, T.S., Cantor, C.R., Collins, J.J., 2000. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339–342.
- Georgiou, G., Valax, P., Ostermeier, M., Horowitz, P.M., 1994. Folding and aggregation of TEM β -lactamase: Analogies with the formation of inclusion bodies in *Escherichia coli*. *Protein Sci.* 3, 1953–1960.
- Gronenborn, B., 1976. Overproduction of phage Lambda repressor under control of the lac promoter of *Escherichia coli*. *MGG Mol. Gen. Genet.* 148, 243–250.

- Grossman, A.D., Erickson, J.W., Gross, C.A., 1984. The htpR gene product of *E. coli* is a sigma factor for heat-shock promoters. *Cell* 38, 383–390.
- Haldimann, A., Wanner, B.L., 2001. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J. Bacteriol.* 183, 6384–6393.
- Hasan, N., Szybalski, W., 1995. Construction of lacI_{ts} and lacI_{qts} expression plasmids and evaluation of the thermosensitive lac repressor. *Gene* 163, 35–40.
- Hecht, M.H., Sturtevant, J.M., Sauer, R.T., 1984. Effect of single amino acid replacements on the thermal stability of the NH₂-terminal domain of phage λ repressor. *Proc. Natl. Acad. Sci. U. S. A.* 81, 5685–5689.
- Hoffmann, F., Rinas, U., 2001. On-line estimation of the metabolic burden resulting from the synthesis of plasmid-encoded and heat-shock proteins by monitoring respiratory energy generation. *Biotechnol. Bioeng.* 76, 333–340.
- Jawale, C. V., Chaudhari, A.A., Jeon, B.W., Nandre, R.M., Lee, J.H., 2012. Characterization of a novel inactivated *Salmonella enterica* serovar enteritidis vaccine candidate generated using a modified cI857/ λ PR/gene E expression system. *Infect. Immun.* 80, 1502–1509.
- Jechlinger, W., Glocker, J., Haidinger, W., Matis, A., Szostak, M.P., Lubitz, W., 2005. Modulation of gene expression by promoter mutants of the λ cI857/pRM/pR system. *J. Biotechnol.* 116, 11–20.
- Jiang, X., Zhang, H., Yang, J., Liu, M., Feng, H., Liu, X., Cao, Y., Feng, D., Xian, M., 2013. Induction of gene expression in bacteria at optimal growth temperatures. *Appl. Microbiol. Biotechnol.* 97, 5423–5431.
- Kaur, Jashandeep, Kumar, A., Kaur, Jagdeep, 2018. Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. *Int. J. Biol. Macromol.* 106, 803–822.
- Lamont, I., Richardson, H., Carter, D.R., Egan, J.B., 1993. Genes for the establishment and maintenance of lysogeny by the temperate coliphage 186. *J. Bacteriol.* 175, 5286–5288.
- Lee, S.K., Keasling, J.D., 2008. Heterologous protein production in *Escherichia coli* using the propionate-inducible pPro system by conventional and auto-induction methods. *Protein Expr. Purif.* 61, 197–203.
- Lewis, D., Le, P., Zurla, C., Finzi, L., Adhya, S., 2011. Multilevel autoregulation of λ repressor protein CI by DNA looping in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 108, 14807–14812.

- Lieb, M., 1966. Studies of heat-inducible lambda bacteriophage: I. Order of genetic sites and properties of mutant prophages. *J. Mol. Biol.* 17, 309.
- Lieb, M., 1981. A fine structure map of spontaneous and induced mutations in the lambda repressor gene, including insertions of IS elements. *MGG Mol. Gen. Genet.* 184, 364–371.
- Lutz, R., Bujard, H., 1997. Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* 25, 1203–1210.
- Meyer, B.J., Ptashne, M., 1980. Gene regulation at the right operator (OR) of bacteriophage λ . III. λ Repressor directly activates gene transcription. *J. Mol. Biol.* 139, 195–205.
- Pabo, C.O., Sauer, R.T., Sturtevant, J.M., Ptashne, M., 1979. The λ repressor contains two domains. *Proc. Natl. Acad. Sci. U. S. A.* 76, 1608–1612.
- Penumetcha, P., Lau, K., Zhu, X., Davis, K., Eckdahl, T.T., Campbell, A.M., 2010. Improving the Lac System for Synthetic Biology. *Bios* 81, 7–15.
- Pinkett, H.W., Shearwin, K.E., Stayrook, S., Dodd, I.B., Burr, T., Hochschild, A., Egan, J.B., Lewis, M., 2006. The structural basis of cooperative regulation at an alternate genetic switch. *Mol. Cell* 21, 605–615.
- Priest, D.G., Cui, L., Kumar, S., Dunlap, D.D., Dodd, I.B., Shearwin, K.E., 2014. Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and λ repressors. *Proc. Natl. Acad. Sci. U. S. A.* 111, 349–354.
- Ptashne, M., 1986. *A genetic switch: Gene control and phage. lambda.* Blackwell Sci. Publ.
- Ptashne, M., Jeffrey, A., Johnson, A.D., Maurer, R., Meyer, B.J., Pabo, C.O., Roberts, T.M., Sauer, R.T., 1980. How the λ repressor and cro work. *Cell* 19, 1–11.
- Sauer, R.T., Jordan, S.R., Pabo, C.O., 1990. λ Repressor: A model system for understanding protein–DNA interactions and protein stability. *Adv. Protein Chem.* 40, 1–61.
- Schein, C.H., 1991. Optimizing protein folding to the native state in bacteria. *Curr. Opin. Biotechnol.* 2, 746–750.
- Schrödinger, LLC, 2010. The {PyMOL} Molecular Graphics System, Version~1.3r1.
- Schubert, R.A., Dodd, I.B., Egan, J.B., Shearwin, K.E., 2007. Cro's role in the CI-Cro bistable switch is critical for λ 's transition from lysogeny to lytic development. *Genes Dev.* 21, 2461–2472.
- Senear, D.F., Laue, T.M., Eaton, S., Ross, J.B.A., Waxman, E., Rusinova, E., 1993. The Primary Self-Assembly Reaction of Bacteriophage λ ci Repressor Dimers Is to Octamer. *Biochemistry* 32, 6179–6189.

- Shearwin, K.E., Dodd, I.B., Egan, B.J., 2002. The helix-turn-helix motif of the coliphage 186 immunity repressor binds to two distinct recognition sequences. *J. Biol. Chem.* 277, 3186–3194.
- Simons, R.W., Houman, F., Kleckner, N., 1987. Improved single and multicopy lac-based cloning vectors for protein and operon fusions. *Gene* 53, 85–96.
- Singh, V., 2020. Introduction to Synthetic Biology. In: *Advances in Synthetic Biology*. pp. 1–15.
- Sørensen, H.P., Mortensen, K.K., 2005. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact.* 4.
- Stayrook, S., Jaru-Ampornpan, P., Ni, J., Hochschild, A., Lewis, M., 2008. Crystal structure of the λ repressor and a model for pairwise cooperative operator binding. *Nature* 452, 1022–1025.
- Stricker, J., Cookson, S., Bennett, M.R., Mather, W.H., Tsimring, L.S., Hasty, J., 2008. A fast, robust and tunable synthetic gene oscillator. *Nature* 456, 516–9.
- Sussman, R., Jacob, F., 1962. On a thermosensitive repression system in the *Escherichia coli* lambda bacteriophage. *C. R. Hebd. Seances Acad. Sci.* 254, 1517–1519.
- Tan, K.P., Khare, S., Varadarajan, R., Madhusudhan, M.S., 2014. TSpred: A web server for the rational design of temperature-sensitive mutants. *Nucleic Acids Res.* 42, W277–W284.
- Terpe, K., 2006. Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 72, 211–222.
- Valdez-Cruz, N.A., Caspeta, L., Pérez, N.O., Ramírez, O.T., Trujillo-Roldán, M.A., 2010. Production of recombinant proteins in *E. coli* by the heat inducible expression system based on the phage lambda pL and/or pR promoters. *Microb. Cell Fact.* 9, 18.
- Varadarajan, R., Nagarajaram, H.A., Ramakrishnan, C., 1996. A procedure for the prediction of temperature-sensitive mutants of a globular protein based solely on the amino acid sequence. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13908–13913.
- Villaverde, A., Benito, A., Viaplana, E., Cubarsi, R., 1993. Fine regulation of cI857-controlled gene expression in continuous culture of recombinant *Escherichia coli* by temperature. *Appl. Environ. Microbiol.* 59, 3485–3487.
- Wang, H., Dodd, I.B., Dunlap, D.D., Shearwin, K.E., Finzi, L., 2013. Single molecule analysis of DNA wrapping and looping by a circular 14mer wheel of the bacteriophage 186 CI repressor. *Nucleic Acids Res.* 41, 5746–5756.
- Winstanley, C., Morgan, J.A., Pickup, R.W., Jones, J.G., Saunders, J.R., 1989. Differential regulation of lambda pL and pR promoters by a cI repressor in a broad-host-range thermoregulated plasmid

marker system. Appl. Environ. Microbiol. 55, 771–777.

Yamamori, T., Ito, K., Nakamura, Y., Yura, T., 1978. Transient regulation of protein synthesis in *Escherichia coli* upon shift-up of growth temperature. J. Bacteriol. 134, 1133–1140.

Yoakum, G.H., Yeung, A.T., Mattes, W.B., Grossman, L., 1982. Amplification of the *uvrA* gene product of *Escherichia coli* to 7% of cellular protein by linkage to the p(L) promoter pKC30. Proc. Natl. Acad. Sci. U. S. A. 79, 1766–1770.

Supplementary materials

Table S 1 Oligonucleotides used in this study

Primer	Sequence
1329	GTACGCCTCGCAGATGCGGTCAAGTAC
1331	CACGGCTCTGGCAGTGAATGGG
1340	GCTAATTGAATTTTCTGCGAAAAGCGGTACGCCTCGCAGATGCGGTCAAGT
1342	CCTTTTATGGATTCATGCAAGGAAAC
1399	GGTGGGTGTATACAGCGAGGTTAACTAAGGATCCTAGAGGCATC
1400	gttAACCTCGCTGTATACACCCACC



Figure S 2 – Plasmid map of pIT3-CL-pL-OL-UP-λpRλpRM-hybridCIts cro-lacZ plasmid. (A) Temperature sensitive module was incorporated into the pIT3-CL-lacZtrimfuse plasmid (Priest et al., 2014). The DNA that makes up the temperature sensitive model, defined by the grey box, is annotated in (B). The temperature sensitive mutation in hybrid^{ts} is shown in red. HTH mutations that abolish DNA binding by cro is shown in green. Mutations that abolish pL promoter activity are shown in pink.



Figure S 3 – Plasmid map of pIT3-CL-pL-OL-UP-λpRλpRM-λCl857 cro- lacZ plasmid. (A) Temperature sensitive module was incorporated into the pIT3-CL-lacZtrimfuse plasmid (Priest et al., 2014). The DNA that makes up the temperature sensitive model, defined by the grey box, is annotated in (B). The temperature sensitive mutation in cl857 is shown in red. The HTH-mutations present in the pIT3-CL-pL-OL-UP-λpRλpRM-λCl(HTH-) cro- lacZ control plasmid is shown in purple. HTH mutations that abolish DNA binding by cro is shown in green. Mutations that abolish pL promoter activity are shown in pink.

6 NOVEL METHOD TO DERIVATIZE CRYSTALS DURING INITIAL CRYSTALLOGRAPHY SCREENING TRIALS

The following chapter presents two publications detailing a novel method to efficiently generate derivatized protein crystals by combining random microseeding matrix screening with derivatization with a heavy atom molecule I3C (5-amino-2,4,6-triiodoisophthalic acid). We demonstrate the efficacy of this technique with two test cases. Two protein crystal structures, the Orf11 N-terminal domain from bacteriophage P68 and hen egg white lysozyme, were successfully solved using this technique, solely using the anomalous signal from I3C. I performed all the experiments detailed in these publications, analysed the data, prepared figures for publication, deposited data into the Protein Data Bank and wrote the manuscripts.

Statement of Authorship

Title of Paper	Combining random microseed matrix screening and the magic triangle for the efficient structure solution of a potential lysin from bacteriophage P68
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	This publication describes a new crystallization protocol to produce derivatized crystals by performing a random microseed matrix screen in the presence of a phasing molecule I3C.

Principal Author

Name of Principal Author (Candidate)	Jia Quyen Truong		
Contribution to the Paper	The principal author performed all the experiments described in this manuscript, performed data analysis, prepared figures for publication and contributed to writing the manuscript.		
Overall percentage (%)	70%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	16/12/2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- the candidate's stated contribution to the publication is accurate (as detailed above);
- permission is granted for the candidate to include the publication in the thesis; and
- the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Santosh Panjekar		
Contribution to the Paper	Provided guidance in experimental design and manuscript preparation		
Overall percentage (%)	5		
Signature		Date	16/12/2020

Name of Co-Author	Linda Shearwin-Whyatt		
Contribution to the Paper	Provided bacteria strains for protein purification experiments and aided in manuscript preparation		
Overall percentage (%)	5		
Signature		Date	16/12/2020

Name of Co-Author	John Bruning		
Contribution to the Paper	Provided guidance and contributed reagents for protein purification and crystallography experiments		
Overall percentage (%)	10		
Signature		Date	17/12/2020

Name of Co-Author	Keith Shearwin		
Contribution to the Paper	Assisted in the design of experiments, contributed reagents and analytical tools, analysis of data and preparation of the manuscript		
Overall percentage (%)	10		
Signature		Date	16/12/2020

Statement of Authorship

Title of Paper	Derivatization of Protein Crystals with I3C using Random Microseed Matrix Screening
Publication Status	<input type="checkbox"/> Published <input checked="" type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	This manuscript describes a new crystallization protocol to produce derivatized crystals by performing a random microseed matrix screen in the presence of a phasing molecule I3C.

Principal Author

Name of Principal Author (Candidate)	Jia Quyen Truong		
Contribution to the Paper	The principal author performed all the experiments in this manuscript. They prepared figures, wrote the manuscript and also assisted with the production of the associated video that will be published alongside this manuscript.		
Overall percentage (%)	40%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	16/12/2020

Name of Principal Author	Stephanie Nguyen		
Contribution to the Paper	Contributed to the production of a video protocol that will be published alongside the manuscript and to the preparation of the manuscript		
Overall percentage (%)	40%		
Signature		Date	16/12/2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Keith Shearwin		
Contribution to the Paper	Contributed to the preparation of the manuscript		
Overall percentage (%)	10		
Signature		Date	16/12/2020

Name of Co-Author	John Bruning		
Contribution to the Paper	Contributed to the preparation of the manuscript		
Overall percentage (%)	10		
Signature		Date	17/12/2020



Combining random microseed matrix screening and the magic triangle for the efficient structure solution of a potential lysin from bacteriophage P68

Jia Quyen Truong,^a Santosh Panjikar,^b Linda Shearwin-Whyatt,^a John B. Bruning^c and Keith E. Shearwin^{a*}

Received 1 March 2019

Accepted 24 June 2019

Edited by B. Kobe, University of Queensland, Australia

Keywords: crystallography; seeding; random microseed matrix screening; I3C; magic triangle; phasing; lysin.

PDB references: lysin protein from *Staphylococcus* phage P68, 6o43; hen egg-white lysozyme, 6pbb

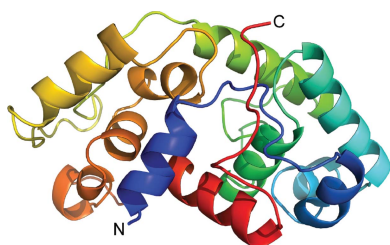
Supporting information: this article has supporting information at journals.iucr.org/d

^aSchool of Biological Sciences, The University of Adelaide, North Terrace, Adelaide, South Australia 5005, Australia, ^bMX, Australian Synchrotron, 800 Blackburn Road Clayton, Melbourne, VIC 3168, Australia, and ^cInstitute of Photonics and Advanced Sensing, School of Biological Sciences, The University of Adelaide, North Terrace, Adelaide, South Australia 5005, Australia. *Correspondence e-mail: keith.shearwin@adelaide.edu.au

Two commonly encountered bottlenecks in the structure determination of a protein by X-ray crystallography are screening for conditions that give high-quality crystals and, in the case of novel structures, finding derivatization conditions for experimental phasing. In this study, the phasing molecule 5-amino-2,4,6-triiodoisophthalic acid (I3C) was added to a random microseed matrix screen to generate high-quality crystals derivatized with I3C in a single optimization experiment. I3C, often referred to as the magic triangle, contains an aromatic ring scaffold with three bound I atoms. This approach was applied to efficiently phase the structures of hen egg-white lysozyme and the N-terminal domain of the Orf11 protein from *Staphylococcus* phage P68 (Orf11 NTD) using SAD phasing. The structure of Orf11 NTD suggests that it may play a role as a virion-associated lysin or endolysin.

1. Introduction

X-ray crystallography is commonly used to determine the three-dimensional structures of proteins, with over 148 000 structures deposited in the Protein Data Bank (PDB) as of February 2019 (Burley *et al.*, 2019). However, there are many bottlenecks in successful structure determination. For a protein for which no X-ray crystal structure exists, the conditions that will give rise to crystals are not known *a priori*. Many crystallization conditions are screened to find one that provides diffracting crystals. However, the large number of factors that affect crystallization, including protein concentration, precipitant, buffer, salt concentrations and temperature, make an exhaustive screening expensive and laborious (Jancarik & Kim, 1991). Sparse-matrix screens are often used to efficiently find crystallization conditions (Jancarik & Kim, 1991). Sparse matrix is a data-mining approach that chooses crystallization conditions based on known and published crystallization conditions. It can be generalized or tailored to the macromolecule, for example membrane proteins or DNA-binding proteins. Many sparse-matrix screens are readily available commercially. Ireton & Stoddard (2004) pioneered the microseed matrix screening (MMS) technique of transferring crystal seeds from existing crystals to new crystallization conditions to create new conditions in which crystals can grow. D'Arcy *et al.* (2007) further developed the technique for use with random crystallization screens, such as the sparse-matrix screens already described. This approach was named random microseed matrix screening (rMMS) and was found



OPEN ACCESS

to be an effective method for increasing the number of hit crystallization conditions. In many cases, improved diffraction-quality crystals could be obtained.

A protein structure can only be obtained using X-ray crystallography if the phase problem can be solved. This presents another bottleneck. If a suitable search model is available through a homologous structure, the phase problem can be solved by molecular replacement (MR; Rossmann, 1990; McCoy *et al.*, 2007; Bibby *et al.*, 2012). However, many protein targets being studied do not have a suitable template and require experimental phasing.

Several experimental phasing techniques have been developed. The traditional phasing method is isomorphous replacement, which involves soaking crystals in solutions containing heavy atoms such as lead, mercury or uranium. Another phasing technique is anomalous dispersion at a single wavelength (SAD; Wang, 1985) or multiple wavelengths (MAD; Hendrickson, 1991). The anomalous dispersion technique requires atoms to be incorporated into the crystal lattice that scatter anomalously at the wavelengths available at the X-ray source. These can be incorporated through soaks or co-crystallization with heavy-atom solutions, chemical modification (for example selenium modification of nucleic acids) or labeling during protein expression (selenomethionine and selenocysteine). The theories of isomorphous replacement and anomalous dispersion are described in Taylor (2010). Suitable compounds and conditions for heavy-atom derivatization are found empirically by screening. This process can be very laborious, although some rational approaches have been devised to guide the screening process (Lu & Sun, 2014).

Beck *et al.* (2008, 2010) synthesized the 'magic triangle' (I3C) and 'MAD triangle' (B3C) compounds specifically for phasing, showing that they could phase several model proteins (Fig. 1). Currently, there are 19 structures in the Protein Data Bank that have been solved using I3C. The triangular arrangement of anomalous scatterers can easily be identified in the substructure. The two carboxylate groups and one amino group can interact with the protein through hydrogen bonding to either the protein backbone or to side chains to facilitate more specific binding.

In this study, we incorporated I3C into an rMMS screen to increase the chance of obtaining diffraction-quality crystals while simultaneously incorporating heavy atoms into the crystal for experimental phase determination. This method aims to overcome the two identified bottlenecks in structure

determination of novel targets in one step. We demonstrated this technique using hen egg-white lysozyme (HEWL) and the N-terminal domain of Orf11 from *Staphylococcus* bacteriophage P68 (Orf11 NTD). Orf11 NTD is a good example of a target that does not have a suitable template in the PDB to use for MR and that did not produce sufficient hit conditions in an initial screen. Our devised method very quickly gave diffraction-quality crystals that were already derivatized with a heavy-atom compound and allowed the structure to be solved.

2. Methods

2.1. Crystallization of HEWL

HEWL was commercially acquired from Sigma–Aldrich (catalog No. L6876) as a lyophilized powder. The powder was dissolved in TBS (50 mM Tris–HCl pH 7.6, 150 mM NaCl) to a final concentration of 30 mg ml⁻¹ as determined using UV absorbance at 280 nm.

Lysozyme crystals were grown via hanging-drop vapor diffusion. 1 µl protein solution was mixed with 1 µl reservoir solution [0.2 M ammonium tartrate dibasic pH 7.0, 20% (w/v) polyethylene glycol 3350] and equilibrated against 500 µl reservoir solution. The crystals obtained were crushed to generate an rMMS seed stock as described by D'Arcy *et al.* (2007).

Sitting-drop vapor-diffusion crystallization screening of HEWL was performed in 96-well Intelli-Plates (Art Robbins) using the commercial Index HT screen (Hampton Research). Four crystallization screens were carried out corresponding to lysozyme without seeds or I3C, lysozyme with seeds, lysozyme with I3C, and lysozyme with seeds and I3C.

For screens containing I3C, 1 M I3C stock was directly added to the protein to give a final concentration of 20 mM I3C. Sitting-drop vapor-diffusion trays were set up with the I3C-containing protein stock. 1 µl protein solution was added to 1 µl reservoir solution and 0.1 µl seed stock and equilibrated over 75 µl reservoir solution.

A seeded crystal from condition C6 of Index HT grown in the presence of I3C showed suitable diffraction. These crystals appeared after three days and reached their maximum size within two weeks. The size of the crystals was estimated to be 150 µm. Crystals were mounted on cryoloops (Hampton Research), passed through Paratone (Hampton Research) for cryoprotection and flash-cooled in liquid nitrogen for data collection (Teng, 1990).

2.2. Data collection, structure solution and refinement of HEWL

A 1.87 Å resolution SAD data set was collected at a wavelength of 1.459 Å using an ADSC Quantum 210r CCD detector on the macromolecular beamline MX1 at the Australian Synchrotron (McPhillips *et al.*, 2002). This wavelength allows iodine to have a large anomalous signal ($f'' = 6.3$ e and $f' = -0.13$ e). 360 diffraction images with 1° oscillation width were collected at a crystal-to-detector distance of 120 mm. The diffraction data from two crystals were processed

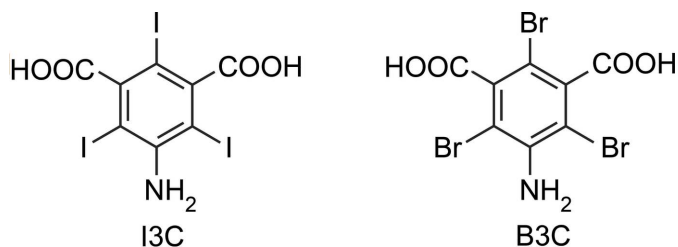


Figure 1
Chemical structures of I3C (5-amino-2,4,6-triiodoisophthalic acid) and B3C (5-amino-2,4,6-tribromoisophthalic acid).

Table 1

Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Protein	HEWL	Orf11 NTD
Diffraction source	MX1, Australian Synchrotron	MX1, Australian Synchrotron
Wavelength (Å)	1.459	1.459
Temperature (K)	100	100
Detector	ADSC Quantum 210r CCD detector	ADSC Quantum 210r CCD detector
Rotation range per image (°)	1	1
Total rotation range (°)	360	360
Space group	$P4_3212$	$P3_121$
a, b, c (Å)	77.16, 77.16, 38.20	61.397, 61.397, 101.606
α, β, γ (°)	90, 90, 90	90, 90, 120
Resolution range (Å)	38.20–1.87 (1.92–1.87)	36.73–2.08 (2.13–2.08)
Total No. of reflections	428577 (7041)	267451 (14604)
No. of unique reflections	9810 (475)	13689 (853)
Completeness (%)	98.4 (76.4)	98.3 (80.0)
Multiplicity	43.7 (14.8)	19.5 (17.1)
$\langle I/\sigma(I) \rangle$	23.0 (1.8)	22.2 (3.5)
Wilson B factor (Å ²)	17.501	42.71
$R_{p.i.m.}$	0.019 (0.198)	0.013 (0.134)
R_{meas}	0.127 (0.850)	0.060 (0.578)
R_{anom}	0.0777 (0.171)	0.0528 (0.160)
CC_{anom}	0.452 (−0.120)	0.291 (−0.038)
$CC_{1/2}$	0.999 (0.938)	1 (0.983)

using *XDS* (Kabsch, 2010) and combined and scaled using *AIMLESS* (Evans & Murshudov, 2013) (Table 1).

The structure was solved using the SAD protocol of *Auto-Rickshaw*, the EMBL Hamburg automated crystal structure-determination platform (Panjikar *et al.*, 2005). The diffraction data processed using *AIMLESS* (Evans & Murshudov, 2013) were used as input. F_A values were calculated using *SHELXC* (Sheldrick, 2015). Based on an initial analysis of the data, the maximum resolution for substructure determination and initial phase calculation was set to 1.8 Å. 17 potential heavy-atom sites out of the maximum number of 20 heavy atoms requested were found using *SHELXD* (Schneider & Sheldrick, 2002). The correct hand for the substructure was determined using *ABS* (Hao, 2004) and *SHELXE* (Sheldrick, 2002). Initial phases were calculated after density modification using *SHELXE* (Sheldrick, 2002). 89.48% of the model was built using *ARP/wARP* (Perrakis *et al.*, 1999; Morris *et al.*, 2004).

The resulting structure was used as a starting model in the MRSAD module (Panjikar *et al.*, 2009) of *Auto-Rickshaw* in space group $P4_3212$ for further phase improvement, model completion and refinement. The molecular-replacement step of the pipeline was skipped. Refinement of the structure was carried out using *CNS* (Brünger *et al.*, 1998), *REFMAC5* (Murshudov *et al.*, 2011) and *phenix.refine* (Afonine *et al.*, 2012) within *Auto-Rickshaw*. A search for and refinement of heavy atoms were conducted using *Phaser* (McCoy *et al.*, 2007) and *MLPHARE* (Winn *et al.*, 2011), which identified 13 sites. Nine sites were subsequently determined to correspond to the three I atoms from each of three I3C ligands, and four sites corresponded to the S atoms of two methionine and two cysteine residues in the structure. Density modification was performed in *Pirate* (Cowtan, 2000). Model building was

Table 2

Structure-solution statistics.

Values in parentheses are for the highest resolution shell.

Protein	HEWL	Orf11 NTD
Molecular mass (kDa)	14.3	22.8
Reflections used in refinement	9547 (921)	13296 (1219)
Reflections used for R_{free}	913 (97)	893 (94)
R_{work}	0.1967 (0.3442)	0.1792 (0.2498)
R_{free}	0.2404 (0.3874)	0.2245 (0.3030)
CC_{work}	0.948 (0.887)	0.966 (0.910)
CC_{free}	0.923 (0.778)	0.938 (0.866)
No. of non-H atoms		
Macromolecules	1142	1675
I3C	1004	1535
Solvent	64	16
No. of protein residues	73	124
R.m.s.d., bonds (Å)	129	199
R.m.s.d., angles (°)	0.003	0.003
Ramachandran favored (%)	0.54	0.50
Ramachandran allowed (%)	99.21	96.95
Ramachandran outliers (%)	0.79	3.05
Rotamer outliers (%)	0.00	0.00
Clashscore	0.00	2.00
Average B factor (Å ²)	5.39	3.32
Overall	36.01	54.11
Macromolecules	33.76	53.53
Ligands	67.83	90.66
Solvent	38.66	56.48
No. of TLS groups	5	5

conducted using *SHELXE* (Sheldrick, 2002), *RESOLVE* (Terwilliger, 1999, 2000) and *Buccaneer* (Cowtan, 2006), which collectively built 128 out of 129 residues that were correctly docked in the electron density. *REFMAC5* and *phenix.refine* were used to further refine the structure, which resulted in an R and R_{free} of 31.48% and 34.53%, respectively. The structure was then iteratively rebuilt and refined using *Coot* (Emsley & Cowtan, 2004) and *phenix.refine* to an R and R_{free} of 19.67% and 24.04%, respectively. Structure-solution statistics are summarized in Table 2.

2.3. Expression of Orf11 NTD

Staphylococcus phage P68 Orf11 N-terminal domain (residues 2–200) was expressed in *Escherichia coli* BL21(λ DE3) cells carrying the pLysS and pET-15b Orf11 NTD plasmids (strain JT438). JT438 cells were grown overnight in 5 ml LB medium supplemented with 100 μ g ml^{−1} ampicillin. The cells were diluted 1:200 into fresh prewarmed LB medium supplemented with 100 μ g ml^{−1} ampicillin. The culture was grown at 37°C with shaking until an OD₆₀₀ of 0.6 was reached and were then induced overnight at 16°C using isopropyl β -D-1-thiogalactopyranoside (IPTG) at a final concentration of 200 μ M. The cells were harvested by centrifugation at 5000g for 20 min. The cell pellets were stored at −80°C.

2.4. Purification of Orf11 NTD

The cell pellets were thawed and resuspended in START buffer [20 mM Tris pH 7.6, 500 mM NaCl, 20 mM imidazole, 5 mM β -mercaptoethanol (BME) pH 7.2]. Cell resuspensions were lysed by sonication on ice (5 \times 30 s, 50% duty cycle,

Sonifier Cell Disruptor B-30). The lysate was clarified by centrifugation at 40 000g for 1 h and subsequent filtration through 0.45 and 0.2 μm syringe filters (Sartorius Minisart) using a disposable syringe. The supernatant was subjected to Ni^{2+} -IMAC chromatography on a 5 ml HisTrap FF column (GE Healthcare) using an NGC FPLC system (Bio-Rad). The column was washed with ten column volumes of START buffer. The protein was eluted from the column with START buffer with an imidazole gradient from 70 to 220 mM over eight column volumes, followed by an imidazole gradient from 220 to 500 mM over two column volumes. Fractions showing UV absorbance A_{280} above baseline were tested for purity using SDS-PAGE. Fractions containing the purified protein were pooled and dialysed against 25 mM Tris pH 7.5, 150 mM NaCl, 10% glycerol, 5 mM BME. The protein was concentrated to 50 mg ml⁻¹ using Amicon Ultra-15 centrifugation units (Millipore) and was stored at -80°C .

2.5. Crystallization of Orf11 NTD using the rMMS protocol in the presence of I3C

Crystallization screening for the Orf11 NTD protein was conducted at 16°C by the sitting-drop vapor-diffusion method in 96-well Intelli-Plates (Art Robbins) using the commercial PEG/Ion HT screen (Hampton Research). The stored protein was diluted to 30 mg ml⁻¹ in 25 mM Tris pH 7.5, 150 mM NaCl, 10% glycerol, 5 mM BME for crystallization. 1 μl protein solution was added to 1 μl reservoir solution and equilibrated over 75 μl reservoir solution. Crystals from condition G12 of PEG/Ion HT were used to generate an rMMS seed stock, as described by D'Arcy *et al.* (2007). A 1 M stock of lithium I3C solution was prepared as described in Beck *et al.* (2008). The protein stock was diluted to 30 mg ml⁻¹ using heavy-atom buffer (25 mM Tris pH 7.5, 150 mM NaCl, 10% glycerol, 5 mM BME, 50 mM I3C). The protein stock contained I3C at a final concentration of 20 mM. Sitting-drop vapor-diffusion trays were set up with the I3C-containing protein stock. 1 μl protein solution was added to 1 μl reservoir solution and 0.1 μl seed stock and was equilibrated over 75 μl reservoir solution. For crystallization screens without I3C, I3C was omitted from the buffer used to dilute the protein. The sizes of the crystals used for data acquisition were estimated to be between 50 and 75 μm . Crystals appeared after one day and reached their maximum size within a week. If too many crystals were observed in the drop, the seed stock was diluted with PEG/Ion HT condition G12 reservoir solution in subsequent optimization steps.

Crystals were mounted on cryoloops (Hampton Research), passed through Paratone (Hampton Research) for cryoprotection and flash-cooled in liquid nitrogen for data collection (Teng, 1990).

2.6. Data collection, structure solution and refinement of Orf11 NTD

A seeded crystal from condition H3 of PEG/Ion HT grown in the presence of I3C showed suitable diffraction. A 2.0 \AA resolution SAD data set was collected at a wavelength of

1.459 \AA using an ADSC Quantum 210r CCD detector on the macromolecular beamline MX1 at the Australian Synchrotron (McPhillips *et al.*, 2002). At this wavelength, iodine has an f'' and f' of 6.3 and -0.13 e, respectively. 360 diffraction images with 1° oscillation width were collected at a crystal-to-detector distance of 120 mm. The diffraction data were processed using *XDS* (Kabsch, 2010) and scaled with *AIMLESS* (Evans & Murshudov, 2013) (Table 1).

The structure was solved using the SAD protocol of *Auto-Rickshaw* (Panjekar *et al.*, 2005). The input diffraction data were prepared and converted for use in *Auto-Rickshaw* using programs from the *CCP4* suite (Winn *et al.*, 2011). F_A values were calculated using *SHELXC* (Sheldrick, 2015). Based on an initial analysis of the data, the maximum resolution for substructure determination and initial phase calculation was set to 3.4 \AA . Five heavy atoms out of the maximum number of nine heavy atoms requested were found using *SHELXD* (Schneider & Sheldrick, 2002). The correct hand for the substructure was determined using *ABS* (Hao, 2004) and *SHELXE* (Sheldrick, 2002). Based on the analysis of the hand of the substructure, the space group of the data was changed from $P3_221$ to $P3_121$. Initial phases were calculated after density modification using *SHELXE* (Sheldrick, 2002). 84.54% of the model was built using *ARP/wARP* (Perrakis *et al.*, 1999; Morris *et al.*, 2004).

The resulting structure was used as a starting model in the MRSAD module (Panjekar *et al.*, 2009) of *Auto-Rickshaw* in space group $P3_121$ for further phase improvement, model completion and refinement. Refinement of the structure was carried out using *CNS* (Brünger *et al.*, 1998), *REFMAC5* (Murshudov *et al.*, 2011) and *phenix.refine* (Afonine *et al.*, 2012) within *Auto-Rickshaw*. A search for and refinement of heavy atoms were conducted using *Phaser* (McCoy *et al.*, 2007) and *MLPHARE* (Winn *et al.*, 2011), which identified four sites. Three sites were subsequently determined to correspond to the three I atoms from a single I3C, and one site corresponded to the S atom of a methionine. Density modification was performed in *Pirate* (Cowtan, 2000). Model building was conducted using *SHELXE* (Sheldrick, 2002), *RESOLVE* (Terwilliger, 1999, 2000) and *Buccaneer* (Cowtan, 2006), which collectively built 194 out of 207 residues that were correctly docked in the electron density. *REFMAC5* and *phenix.refine* were used to further refine the structure, which resulted in an R and R_{free} of 26.75% and 32.27%, respectively. The structure was then iteratively rebuilt and refined using *Coot* (Emsley & Cowtan, 2004) and *phenix.refine* to an R and R_{free} of 17.34% and 22.64%, respectively. Structure-solution statistics are summarized in Table 2.

2.7. PDB codes

The coordinates and structure factors for Orf11 NTD and HEWL have been deposited in the Protein Data Bank as entries 6o43 and 6pbb, respectively.

3. Results

3.1. I3C can be added to a random microseed matrix screen and still give rise to more crystals than unseeded protein

The effect of adding I3C to an rMMS optimization screen was tested using HEWL as a model crystallization protein. rMMS optimization has previously been successfully applied to lysozyme to increase the number of conditions supporting

crystal growth in the JCSG, PACT and Morpheus screens (Till *et al.*, 2013). Consistent with previous studies, we observe that rMMS increases the number of conditions in which lysozyme can crystallize in the Index HT screen (Hampton Research). The rMMS screen generated 35 more conditions that supported crystal growth [Fig. 2(a)]. Adding I3C without seeding to the screen did not notably increase the number of crystallization conditions found. Eight conditions no longer

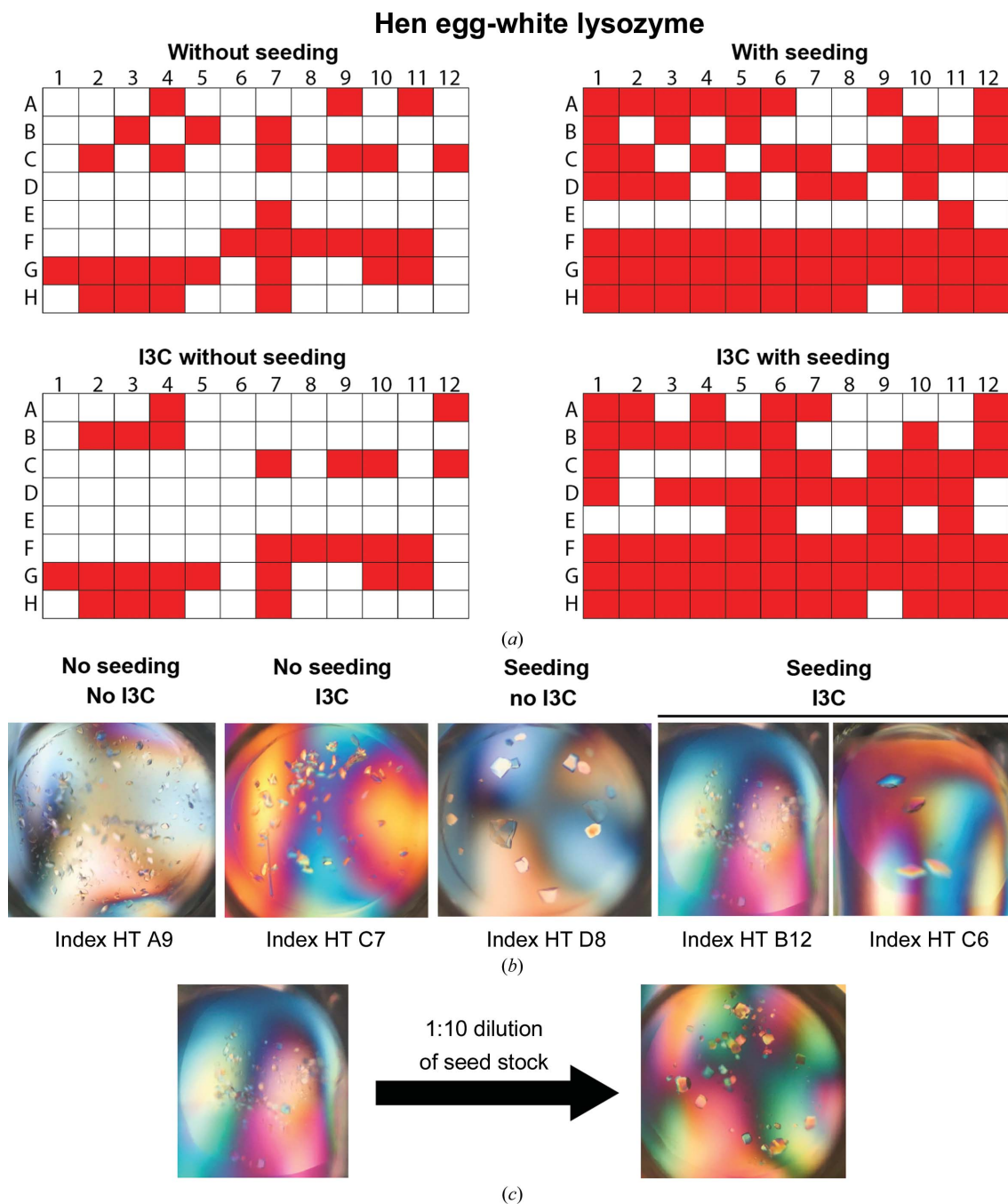


Figure 2 rMMS can generate new conditions that support the crystal growth of HEWL in the presence of I3C. (a) Crystallization conditions supporting the crystal growth of HEWL in the Index HT screen are shaded in red. Seed stock was made from lysozyme crystals grown in 0.2 M ammonium tartrate dibasic pH 7.0, 20% (w/v) polyethylene glycol 3350. (b) Representative photographs showing crystals that formed with and without seeding both in the presence and absence of I3C. (c) Dilution of seed stock was used to reduce the number of seeds to influence the crystal size. Seed stock was diluted 1:10 to give fewer and larger crystals in condition B12 of Index HT.

supported crystal growth and three new conditions were found in the presence of I3C. Adding I3C to the lysozyme rMMS optimization screen generated substantially more conditions than screens without seeding with and without I3C [Fig. 2(a)]. Having used HEWL as a test case, we tested this technique with a protein with significantly lower crystallizability, Orf11 NTD.

Only one hit condition was obtained from the initial screen with Orf11 NTD using the PEG/Ion HT screen (Hampton Research). This condition generated a shower of microcrystals that were too small to mount for X-ray diffraction. Adding

I3C to the screen resulted in the same hit condition but gave no new conditions. The crystalline material from the unseeded screen was used to make a seed stock for an rMMS screen using PEG/Ion HT with and without I3C added. Five and six new crystallization hits were identified in these optimization screens, respectively [Fig. 3(a)].

The new hit condition H3 from PEG/Ion found by adding undiluted seed stock produced too many crystals and was optimized for crystal number by the dilution of seed stock as detailed in Fig. 2. The growth of fewer crystals resulted in larger crystals [Fig. 3(c)]. These two screens using HEWL and

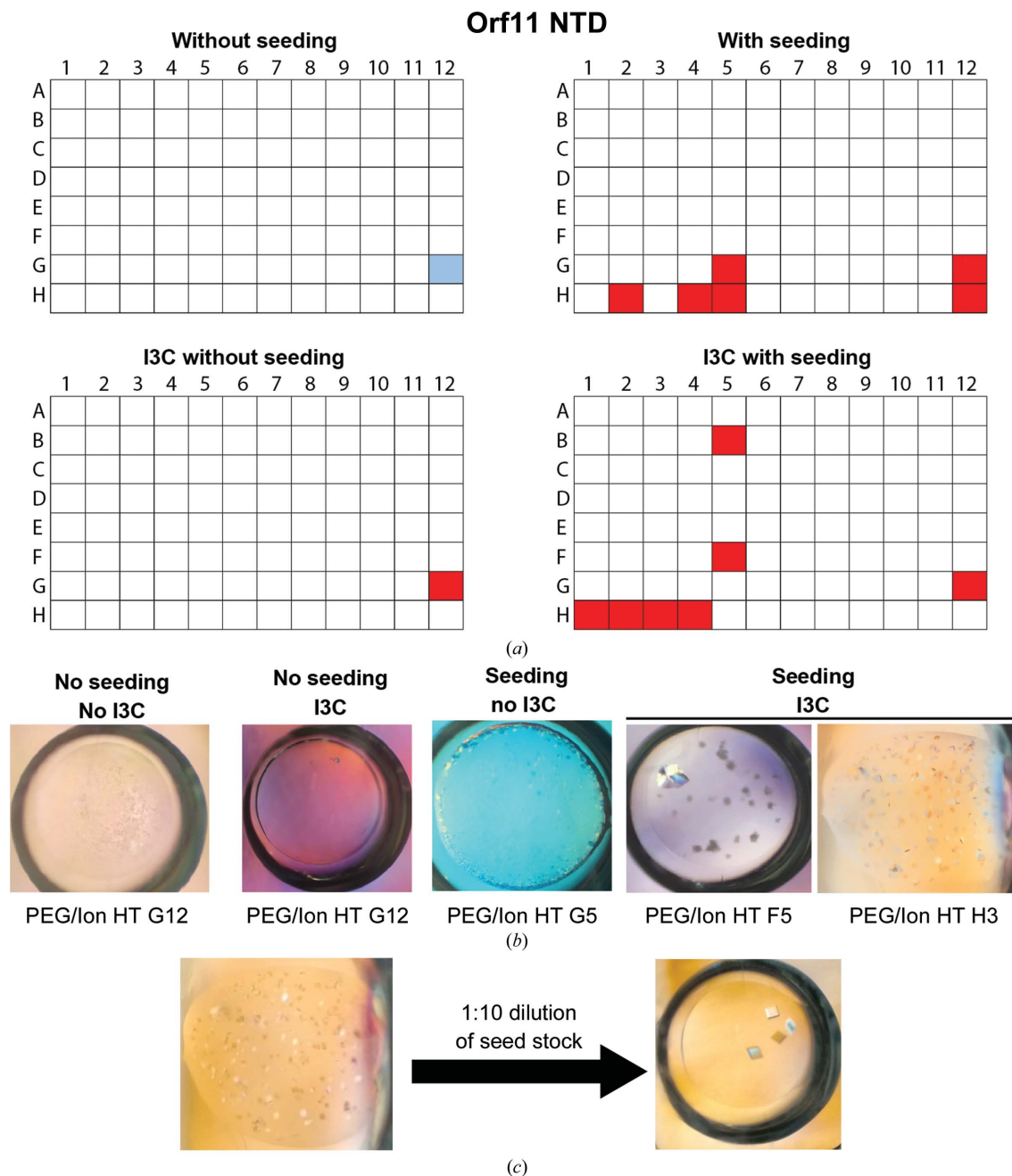


Figure 3

rMMS was used to find new hit conditions that support the crystal growth of Orf11 NTD in the presence of I3C. (a) Crystallization conditions supporting crystal growth of Orf11 NTD in the PEG/Ion HT screen are colored. The crystallization condition used for seed stock is shown in blue. (b) Representative photographs showing Orf11 NTD crystals that formed without and with seeding. (c) Dilution of seed stock can reduce excess nucleation to give larger crystals. Seed stock was diluted 1:10 to give fewer but larger crystals in condition H3 of PEG/Ion HT.

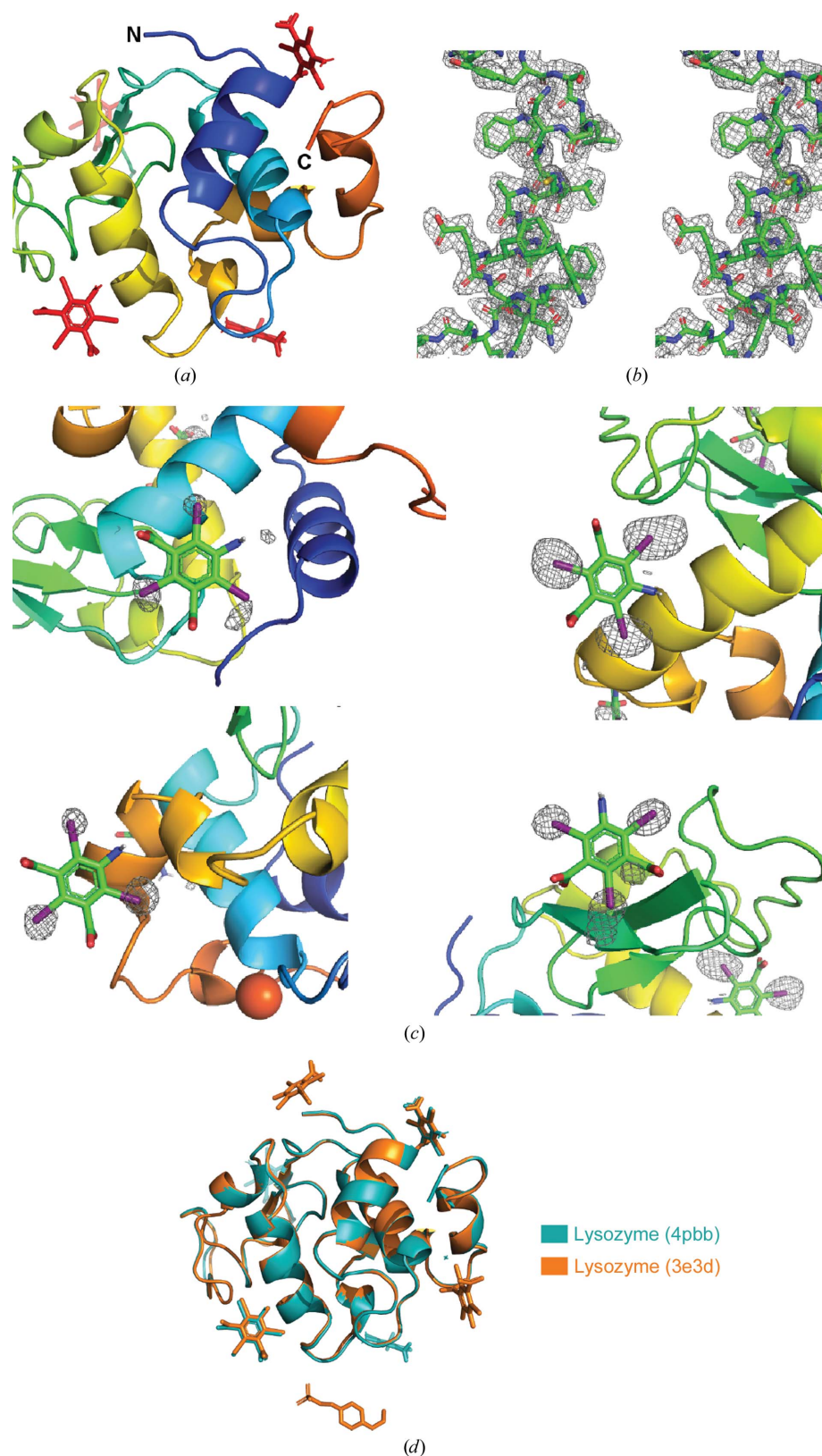


Figure 4
 Structure solution of HEWL. (a) The crystal structure of HEWL. (b) Stereoview of a composite OMIT $2mF_o - DF_c$ electron-density map of HEWL. The contour level was set to 1σ . (c) Substructure density of each I3C molecule. The anomalous difference map is contoured at 5σ . The map was generated using *phenix.maps*. (d) The crystal structure of HEWL (PDB entry 6pbp) was superimposed onto a previously solved structure of lysozyme soaked with I3C (PDB entry 3e3d) and shows that two I3C molecules are bound in the same position.

Orf11 NTD provide a proof of concept that I3C-derivatized crystals can be grown and optimized using seeding from a nonderivatized crystal.

To confirm that I3C was incorporated into the crystals, crystals from the I3C rMMS screen were harvested and the crystal structures of the two proteins were solved using the anomalous signal without any external template information.

3.2. Structure of HEWL

The structure of HEWL was solved by SAD phasing using diffraction data from a crystal grown using condition C6 of Index HT [Fig. 4(a)]. The structure showed four I3C molecules bound to a single HEWL monomer in the asymmetric unit, with occupancies of 57, 54, 32 and 26% [Fig. 4(c)]. A structural superposition of the structure with a previously solved structure of HEWL from a crystal co-crystallized with I3C (PDB entry 3e3d; Beck *et al.*, 2008) showed that two of the four I3C molecules bind to the same positions in the protein [Fig. 4(d)].

3.3. Structure of Orf11 NTD

The structure of Orf11 NTD was obtained by SAD phasing using diffraction data from a crystal grown using condition H3 of PEG/Ion HT [Fig. 3(c)] in an rMMS screen in the presence of I3C. This structure is shown in Fig. 5(a), with representative electron density shown in Fig. 5(b). One I3C molecule was bound to the protein with an occupancy of 53% and provided sufficient signal to phase the structure. An intrinsic S atom from Met138 provided a fourth anomalous scattering atom. All scattering atoms displayed clear density in the anomalous difference map [Figs. 5(c) and 5(d)].

4. Discussion

4.1. The I3C-rMMS method shows promise

Molecular-replacement (MR) phasing is the most popular method for solving protein structures; however, in many cases the protein target lacks a suitable homology or *ab initio* model for

successful MR phasing. In such cases, experimental phasing is the method of choice for solving the phase problem. For proteins that do not intrinsically possess scattering atoms, external anomalous scatterers can be incorporated into the crystal by soaking or co-crystallizing the protein with heavy atoms. If heavy-atom ions or small molecules are incorporated into a crystal via co-crystallization, it usually results in higher occupancy. Another benefit of co-crystallization over soaking is that it reduces manual crystal-manipulation steps that may damage the crystal.

However, soaking is often preferred over co-crystallization to derivatize crystals. Co-crystallization requires the set up of additional screens. In addition, attempting to co-crystallize a heavy-atom ligand in the same condition that yields

underivatized crystals can fail to yield crystals. This observation is unsurprising as the ligand can change the equilibrium of the crystallization condition or the crystal contacts (Garman & Murray, 2003; McPherson & Cudney, 2014). In this study, we present an efficient method for screening for crystallization conditions that co-crystallize a phasing compound into a protein crystal.

rMMS often substantially increases the number of hit conditions and produces improved diffraction-quality crystals in new crystallization conditions (D'Arcy *et al.*, 2007; Obmolova *et al.*, 2010). I3C has previously been demonstrated to bind to proteins within a protein crystal, often via multiple functional groups. The amino group and carboxyl groups of I3C can form hydrogen bonds directly to the protein backbone

and side chains or via water bridges to the protein. The benzene ring and iodine allow π - π interactions and halogen bonds, respectively. The ability to form multiple interactions results in I3C binding with high specificity and occupancy in a protein crystal (Beck *et al.*, 2008).

It is possible that I3C can also stabilize a lattice and/or generate new crystal contacts, resulting in a protein crystal with improved mechanical and diffraction properties. In the HEWL crystal, three of the four I3C molecules interact with three different protein monomers [Fig. 6(a)]. In the Orf11 NTD crystal, I3C was found at the interface between three Orf11 NTD molecules and makes interactions with all three protein molecules [Fig. 6(b)]. Such contacts between the protein molecules bridged by I3C may explain why this crystallization condition was only found when rMMS was combined with I3C and not with rMMS alone. We are aware of one published case in which protein crystals would only form in the presence of I3C and would fail without it (Leverrier *et al.*, 2011). Thus, it seems likely that I3C can increase the number of hits found by rMMS by creating new crystal contacts. In this study, a small increase in the number of hits was found by adding I3C to rMMS over a standard rMMS screen. However, this is not a clear positive effect as some hit conditions were lost upon adding I3C.

By adding I3C to the protein solution and using rMMS, we are efficiently searching for crystallization conditions in which I3C can bind the protein. rMMS increases the likelihood of obtaining crystals, and any such protein

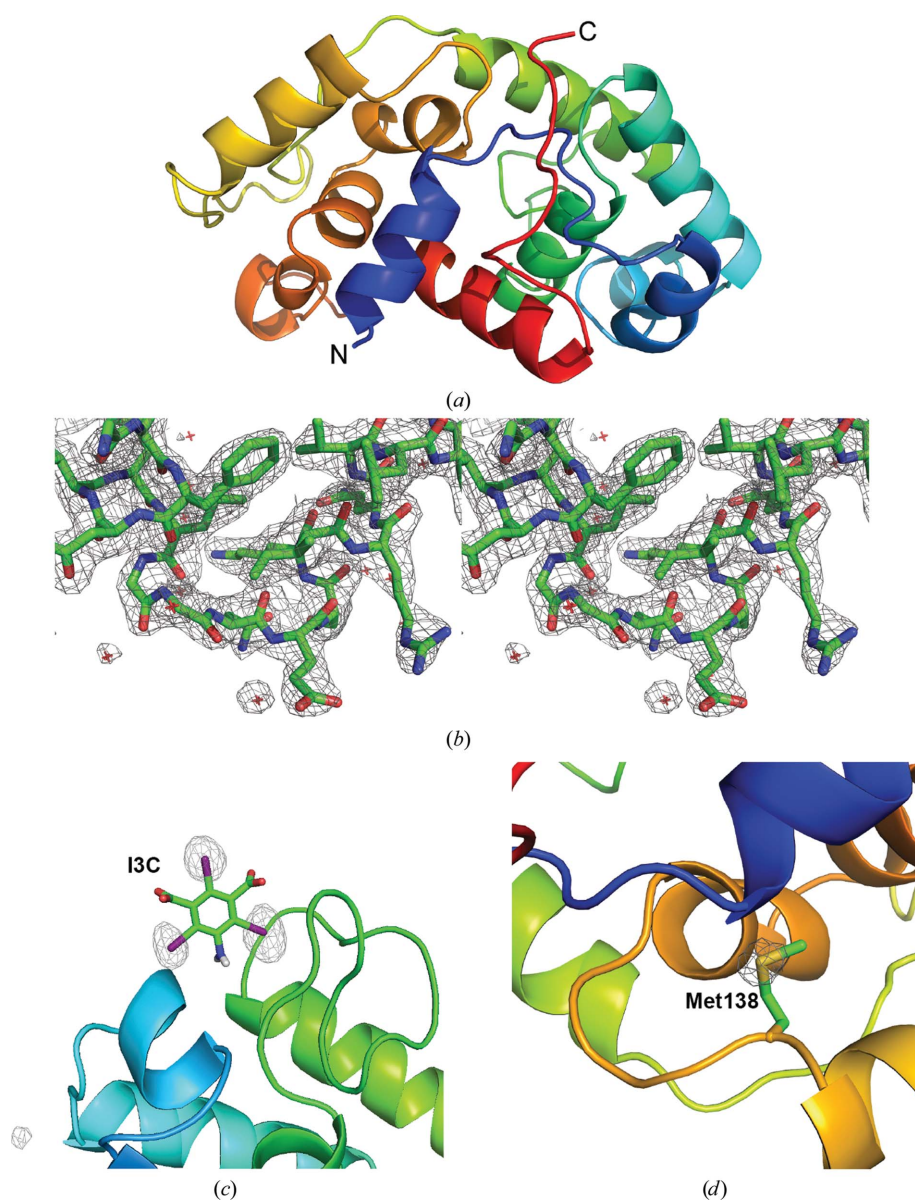
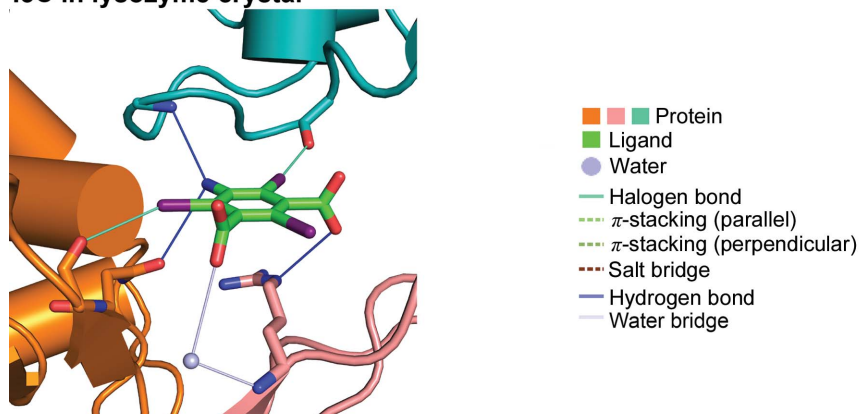


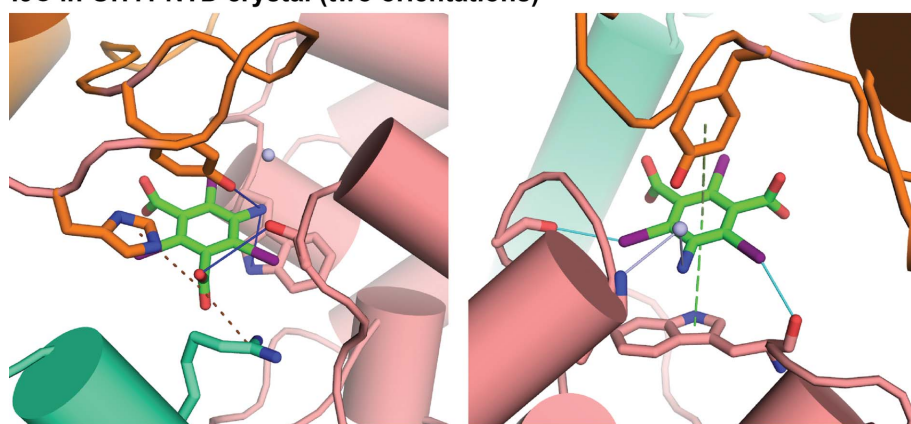
Figure 5
Structure solution of Orf11 NTD. (a) The crystal structure of Orf11 NTD. (b) Stereoview of a composite OMIT $2mF_o - DF_c$ electron-density map of Orf11 NTD. The contour level was set to 1σ . Substructure density of I3C and the intrinsic S atom are shown in (c) and (d), respectively. The anomalous difference map is contoured at 5σ . The map was generated using *phenix.maps*.

I3C in lysozyme crystal



(a)

I3C in Orf11 NTD crystal (two orientations)



(b)

Figure 6

I3C can mediate contacts between protein molecules in a crystal. One I3C molecule from (a) the HEWL crystal and (b) the Orf11 NTD crystal are shown. The same I3C molecule from the Orf11 NTD crystal is displayed in two different orientations for clarity. I3C–protein interactions include hydrogen bonding, π – π stacking interactions, salt bridges, water bridges and halogen bonding. In both cases, each I3C molecule forms interactions with three different protein molecules (each shown in a different color), which could assist in lattice packing. Protein–ligand interaction analysis was conducted using the *PLIP* web server (Salentin *et al.*, 2015).

crystals obtained are likely to have I3C incorporated, ready for SAD phasing. This approach has successfully been applied to both the HEWL and Orf11 NTD proteins. In both cases, an increase in the number of hit conditions was identified in the rMMS with I3C screens compared with the unseeded screens with and without I3C. For both HEWL and Orf11 NTD, one of these new conditions was confirmed to produce I3C-derivatized crystals that allowed solution of the protein structure.

There are several benefits to the use of I3C with rMMS screening compared with many of the compounds that are commonly used for derivatization. Firstly, I3C is inexpensive and readily available. Secondly, many of the commonly used heavy-metal salts such as uranyl acetate and platinum potassium chloride are incompatible with our rMMS approach. Many of these metal ions will form insoluble compounds when mixed with sulfate and phosphate buffers. Citrate and acetate in certain crystallization conditions can also chelate divalent metal ions to reduce their effective concentration (Pike *et al.*, 2016). Sulfate, phosphate, citrate and acetate are all commonly used in crystallization conditions and would preclude many heavy metals from our rMMS approach. Finally, some heavy atoms, but not I3C, react with HEPES and Tris and/or bind to DTT and β -mercaptoethanol, which are used in many protein preparations (Pike *et al.*, 2016).

I3C intrinsically provides two benefits when solving structures. Each I3C molecule provides three heavy scattering atoms, providing a significant anomalous signal for phasing. The heavy atoms are arranged in an equilateral triangle with sides of 6 Å. This arrangement allows the presence of I3C to be confirmed in the substructure-determination stage. If a triangle with these dimensions is found, it indicates a correct substructure. The specificity of I3C binding also makes it preferable to halide or alkali metal ions, which bind to many positions on the protein. This lack of specificity can result in many poorly occupied sites, making substructure determination difficult.

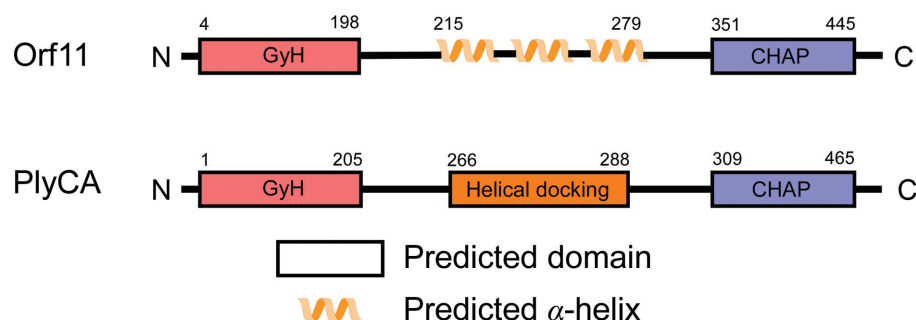


Figure 7

Orf11 is predicted to have a GyH domain and a CHAP domain, connected via a linker containing three α -helices. This predicted domain architecture of Orf11 is similar to that of the endolysin protein PlyCA from *Streptococcus* phage C1. The PlyCA domain architecture was defined using the crystal structure of the protein (PDB entry 4f88; McGowan *et al.*, 2012). The domain architecture of Orf11 was assigned using multiple bioinformatics tools. The GyH domain was identified using the *FFAS-3D* homology-detection server (Xu *et al.*, 2014). The CHAP domain was determined using the Pfam protein-domain database (Finn *et al.*, 2016). The α -helical content within the linker region was defined using the *JPred4* secondary-structure prediction server (Drozdetskiy *et al.*, 2015). The residue numbers for the domain boundaries are annotated above the diagrams.

Another benefit of using iodine is that it provides a large anomalous signal ($f'' = 6.9 e$) at the Cu $K\alpha$ wavelength used by many home-source X-ray generators (and can also be used on home sources for isomorphous replacement). This property would allow the collection of anomalous data for phasing without the need to wait for synchrotron beamtime.

This screening approach can be applied to identify new co-crystallization conditions using other heavy-atom molecules. One possibility is the MAD triangle (Beck *et al.*, 2010), which has a similar molecular structure with iodine substituted by bromine. This phasing molecule has the benefit of allowing MAD phasing as the Br K edge is accessible at many

Orf11_NTD	MND...QEKIDKFTHSYINDDFGLTIDQLVPKVKGYGRFN...VWLGGNESK.IRQVILKA	53
PlyCA	MSKKYTQQQYEKYLAPANTFGLSPQOVADWFMGQAGARPVINSYGVNASNLVSTYILPK	60
Orf11_NTD	VKEIGVSPTLFAVYEKNECFSSGLGWLNHTSARGDYLTDAKFIARKLVSQSKQAGQPSWY	113
PlyCA	MQEYGVSYTLFLMYTVFEGGGAG.NWINHYM.....YDTGSNGLECLEHDLQYIHGVWE	113
Orf11_NTD	D...AGNIVHFVPODVQRKGNAD.FAKNMKAGTIGRAYIPLTAAA...TWAYYPIGLK	165
PlyCA	TYFPPALSAPECYPAEDNAGALDRFYOSLPGRTWGDVMIPTMAGNAWVWAYNYCVNNO	173
Orf11_NTD	ASYNKVNQYGNPFLDGANITLAWGGKLDGKGGSPSSDSSGSSGDSGSSLALAKQAMQE	225
PlyCA	GAAP.LVYFGNPDSDSLIAMDADP.FTGGSIITGDGKNPVGTCGNATVSAASEANRE	230
Orf11_NTD	LKK.IQDALQWDVHSIGSDKFFSNDYFTLEKTFNNTYHIKMTIGLIDSLKKLIDSVOVD	284
PlyCA	KLKA.LTDLFNNNLEHL.SGEFYGNQVLNAMK.YGTILKCDLTDGDLNLAITLADVNLQ	288
Orf11_NTD	SGSSSSNPPTDDDDGDH.KP.....ISGKSVKPNKSGRVIGGNWTYAQLPEKYKKAIGVP	337
PlyCA	TNPNPDKPTVQSPGQNDLGSQSDRVAANLANAQAQVQKYIGDGGCYAWVGVWSARVCGYS	348
Orf11_NTD	LFKKEYLYKPGN.IFFQTCNA..GQCTELITWAYMSQLHGKRQPTDDGOITNGQRVWYVYK	394
PlyCA	I.....SYSTCDPMLPLITDGMNAHSIHLGW.....DWSIANTGITVNYYPVG	389
Orf11_NTD	KLGAKTTHNP...TVGYGFSKPPYLQATAYGI.GHTGVVVAVFEDGSFLVANYVPP.YV	450
PlyCA	TVGRKEDLRVGAIWCATA.FSGAPFYTG...QYGHTGIIIE.SWSDTTVTLEQNLILGSPV	444
Orf11_NTD	APSRVLYTLINGVPPNAGDNIVFVSGIA.....	450
PlyCA	I.....RSTYDLNTEFLSTLTGLITFK	444

KEY

SIMILAR
STRICT IDENTITY

(a)

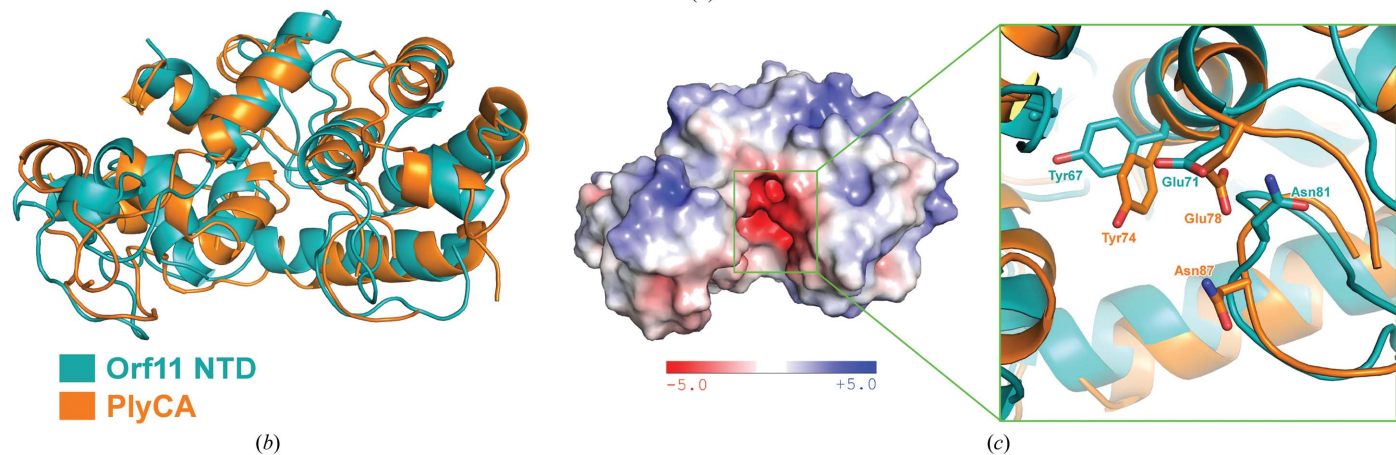


Figure 8 Orf11 shows limited sequence similarity and high structural similarity to the PlyCA protein from *Streptococcus* phage C1. (a) Sequences of the PlyCA GyH domain (residues 1–213) and the Orf11 NTD domain (residues 1–201) were aligned using *T-Coffee* (Armougom *et al.*, 2006), with an identity of 22% and a sequence similarity of 40%. Alignments were displayed using *ESPrpt* (Gouet *et al.*, 1999) and similarity was calculated using the Risler matrix (Risler *et al.*, 1988). (b) A superposition of the PlyCA GyH domain and Orf11 NTD structures shows that the proteins adopt similar tertiary folds. The two proteins were superimposed with *PyMOL* (DeLano, 2002) using C^α positions, with an r.m.s.d. of 3.5 Å over 127 atoms. (c) The conserved catalytic residues of the glucosaminidase domain of GyH are Glu78, Tyr74 and Asn87. The equivalent residues Glu71, Tyr67 and Asn81 in Orf11 NTD appear in similar positions. The residues appear in an electronegative cleft in the protein. The electrostatic potentials of Orf11 NTD were calculated using the *APBS* plugin in *PyMOL*.

synchrotron beamlines. It would also be of interest to test the tantalum bromide cluster (Knäblein *et al.*, 1997), which provides an extremely large amount of phasing power, especially for large proteins or protein–protein/DNA complexes.

4.2. Orf11 is likely to be a lysin

Bacteriophage P68 Orf11 was identified as a putative endolysin using a bioinformatics approach. Endolysins are enzymes that degrade the peptidoglycan structure in bacterial cell walls and, along with holins that form pores in the inner membrane, form the escape system for bacteriophages to exit their host via lysis (Young, 2014). A homology search using the FFAS (Xu *et al.*, 2014) and Pfam (Finn *et al.*, 2016) web servers identified two putative domains in the Orf11 protein: an N-terminal glycosyl hydrolase (GyH) domain and a C-terminal cysteine, histidine-dependent aminohydrolase/peptidase (CHAP) domain (Fig. 7). Both domains commonly appear in endolysin proteins (Oliveira *et al.*, 2013). Another possibility is that Orf11 corresponds to a virion-associated lysin protein (VAL). Some phages, particularly those that infect Gram-positive bacteria, use glycosyl hydrolases associated with the virion to degrade the peptidoglycan layer (Fernandes & São-José, 2018). This process clears a path to enable the phage to find the cell membrane and allows the phage machinery to inject DNA into the cell.

The structure of Orf11 NTD is very similar to that of the GyH domain of PlyC. PlyC is an endolysin protein from *Streptococcus* phage C1 and is the most potent endolysin protein discovered to date (McGowan *et al.*, 2012; Riley *et al.*, 2015). PlyC consists of one PlyCA polypeptide mounted on a PlyCB octamer. A structural superposition of the GyH domain of PlyCA (PDB entry 4f88; McGowan *et al.*, 2012) with Orf11 NTD reveals similar overall folds, despite a low sequence identity between the domains of 22% [Figs. 8(a) and 8(b)]. However, attempts to solve the data set via molecular replacement using the PlyC GyH domain as a model failed to yield a solution. Small deviations from the true structure in the template may have prevented the algorithm from finding an initial solution. The glycosyl hydrolase domain of PlyCA has three key catalytic residues, Glu78, Tyr74 and Asn87, thought to form the catalytic center. In the structure of Orf11 NTD these same residues appeared at similar spatial positions within an electronegative cleft in the protein [Fig. 8(c)]. From this, it is possible that the Orf11 NTD is a glycosyl hydrolase.

A secondary-structure prediction of Orf11 shows that a region containing three α -helices lies between the two domains (Fig. 7). This topology is similar to that of PlyCA, which has a helical docking domain consisting of three α -helices that anchors it to the cell-wall-binding protein PlyCB. All of this, taken together, suggests that Orf11 is a lysin protein that perhaps requires an additional protein to direct it to its bacterial target.

Acknowledgements

This research was undertaken in part using the MX1 beamline at the Australian Synchrotron, which is part of ANSTO.

Author contributions are as follows. JQT performed the crystallographic experiments and structural modeling and prepared the manuscript. JQT and LS performed the bioinformatics analysis. SP provided support with crystallographic data processing. KES and JBB contributed to manuscript preparation.

Funding information

The following funding is acknowledged: Australian Research Council (grant Nos. DP150103009 and DP160101450 to Keith Shearwin); University of Adelaide (Adelaide Postgraduate Scholarship to Jia Quyen Truong).

References

- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.
- Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V. & Notredame, C. (2006). *Nucleic Acids Res.* **34**, W604–W608.
- Beck, T., Gruene, T. & Sheldrick, G. M. (2010). *Acta Cryst.* **D66**, 374–380.
- Beck, T., Krasauskas, A., Gruene, T. & Sheldrick, G. M. (2008). *Acta Cryst.* **D64**, 1179–1182.
- Bibby, J., Keegan, R. M., Mayans, O., Winn, M. D. & Rigden, D. J. (2012). *Acta Cryst.* **D68**, 1622–1631.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* **D54**, 905–921.
- Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Costanzo, L., Christie, C., Duarte, J. M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D. S., Green, R. K., Guranovic, V., Guzenko, D., Hudson, B. P., Liang, Y., Lowe, R., Peisach, E., Periskova, I., Randle, C., Rose, A., Sekharan, M., Shao, C., Tao, Y., Valasatava, Y., Voigt, M., Westbrook, J., Young, J., Zardecki, C., Zhuravleva, M., Kurisu, G., Nakamura, H., Kengaku, Y., Cho, H., Sato, J., Kim, J. Y., Ikegawa, Y., Nakagawa, A., Yamashita, R., Kudou, T., Bekker, G., Suzuki, H., Iwata, T., Yokochi, M., Kobayashi, N., Fujiwara, T., Velankar, S., Kleywegt, G. J., Anyango, S., Armstrong, D. R., Berrisford, J. M., Conroy, M. J., Dana, J. M., Deshpande, M., Gane, P., Gáborová, R., Gupta, D., Gutmanas, A., Koča, J., Mak, L., Mir, S., Mukhopadhyay, A., Nadzirin, N., Nair, S., Patwardhan, A., Paysan-Lafosse, T., Pravda, L., Salih, O., Sehnal, D., Varadi, M., Vařeková, R., Markley, J. L., Hoch, J. C., Romero, P. R., Baskaran, K., Maziuk, D., Ulrich, E. L., Wedell, J. R., Yao, H., Livny, M. & Ioannidis, Y. E. (2019). *Nucleic Acids Res.* **47**, D520–D528.
- Cowtan, K. (2000). *Acta Cryst.* **D56**, 1612–1621.
- Cowtan, K. (2006). *Acta Cryst.* **D62**, 1002–1011.
- D’Arcy, A., Villard, F. & Marsh, M. (2007). *Acta Cryst.* **D63**, 550–554.
- DeLano, W. (2002). *CCP4 Newsl. Protein Crystallogr.* **40**, 82–92.
- Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. (2015). *Nucleic Acids Res.* **43**, W389–W394.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* **D69**, 1204–1214.
- Fernandes, S. & São-José, C. (2018). *Viruses*, **10**, 396.
- Finn, R. D., Cogill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J. & Bateman, A. (2016). *Nucleic Acids Res.* **44**, D279–D285.
- Garman, E. & Murray, J. W. (2003). *Acta Cryst.* **D59**, 1903–1913.
- Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.

- Hao, Q. (2004). *J. Appl. Cryst.* **37**, 498–499.
- Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
- Ireton, G. C. & Stoddard, B. L. (2004). *Acta Cryst. D***60**, 601–605.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- Kabsch, W. (2010). *Acta Cryst. D***66**, 125–132.
- Knäblein, J., Neufeind, T., Schneider, F., Bergner, A., Messerschmidt, A., Löwe, J., Steipe, B. & Huber, R. (1997). *J. Mol. Biol.* **270**, 1–7.
- Leverrier, P., Declercq, J.-P., Denoncin, K., Vertommen, D., Hiniker, A., Cho, S.-H. & Collet, J.-F. (2011). *J. Biol. Chem.* **286**, 16734–16742.
- Lu, J. & Sun, P. D. (2014). *FEBS J.* **281**, 4021–4028.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- McGowan, S., Buckle, A. M., Mitchell, M. S., Hoopes, J. T., Gallagher, D. T., Heselpoth, R. D., Shen, Y., Reboul, C. F., Law, R. H. P., Fischetti, V. A., Whisstock, J. C. & Nelson, D. C. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 12752–12757.
- McPherson, A. & Cudney, B. (2014). *Acta Cryst. F***70**, 1445–1467.
- McPhillips, T. M., McPhillips, S. E., Chiu, H.-J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M. & Kuhn, P. (2002). *J. Synchrotron Rad.* **9**, 401–406.
- Morris, R. J., Zwart, P. H., Cohen, S., Fernandez, F. J., Kakaris, M., Kirillova, O., Vonrhein, C., Perrakis, A. & Lamzin, V. S. (2004). *J. Synchrotron Rad.* **11**, 56–59.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst. D***67**, 355–367.
- Obmolova, G., Malia, T. J., Teplyakov, A., Sweet, R. & Gilliland, G. L. (2010). *Acta Cryst. D***66**, 927–933.
- Oliveira, H., Melo, L. D. R., Santos, S. B., Nóbrega, F. L., Ferreira, E. C., Cerca, N., Azeredo, J. & Kluskens, L. D. (2013). *J. Virol.* **87**, 4558–4570.
- Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Acta Cryst. D***61**, 449–457.
- Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2009). *Acta Cryst. D***65**, 1089–1097.
- Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
- Pike, A. C. W., Garman, E. F., Krojer, T., von Delft, F. & Carpenter, E. P. (2016). *Acta Cryst. D***72**, 303–318.
- Riley, B. T., Broendum, S. S., Reboul, C. F., Cowieson, N. P., Costa, M. G. S., Kass, I., Jackson, C., Perahia, D., Buckle, A. M. & McGowan, S. (2015). *PLoS One*, **10**, e0140219.
- Risler, J. L., Delorme, M. O., Delacroix, H. & Henaut, A. (1988). *J. Mol. Biol.* **204**, 1019–1029.
- Rossmann, M. G. (1990). *Acta Cryst. A***46**, 73–82.
- Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F. & Schroeder, M. (2015). *Nucleic Acids Res.* **43**, W443–W447.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst. D***58**, 1772–1779.
- Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
- Sheldrick, G. M. (2015). *Acta Cryst. C***71**, 3–8.
- Taylor, G. L. (2010). *Acta Cryst. D***66**, 325–338.
- Teng, T.-Y. (1990). *J. Appl. Cryst.* **23**, 387–391.
- Terwilliger, T. C. (1999). *Acta Cryst. D***55**, 1863–1871.
- Terwilliger, T. C. (2000). *Acta Cryst. D***56**, 965–972.
- Till, M., Robson, A., Byrne, M. J., Nair, A. V., Kolek, S. A., Shaw Stewart, P. D. & Race, P. R. (2013). *J. Vis. Exp.*, 50548.
- Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst. D***67**, 235–242.
- Xu, D., Jaroszewski, L., Li, Z. & Godzik, A. (2014). *Bioinformatics*, **30**, 660–667.
- Young, R. (2014). *J. Microbiol.* **52**, 243–258.

Derivatization of Protein Crystals with I3C using Random Microseed Matrix Screening

Jia Quyen Truong^{*1}, Stephanie Nguyen^{*1,2}, John B. Bruning^{1,2}, Keith E. Shearwin¹

¹ School of Biological Sciences, The University of Adelaide ² Institute of Photonics and Advanced Sensing (IPAS), School of Biological Sciences, The University of Adelaide

*These authors contributed equally

Corresponding Author

Keith E. Shearwin

keith.shearwin@adelaide.edu.au

Citation

Truong, J.Q., Nguyen, S., Bruning, J.B., Shearwin, K.E. Derivatization of Protein Crystals with I3C using Random Microseed Matrix Screening. *J. Vis. Exp.* (), e61894, doi:10.3791/61894 (2020).

Date Published

December 15, 2020

DOI

10.3791/61894

URL

jove.com/t/61894

Abstract

Protein structure elucidation using X-ray crystallography requires both high quality diffracting crystals and computational solution of the diffraction phase problem. Novel structures that lack a suitable homology model are often derivatized with heavy atoms to provide experimental phase information. The presented protocol efficiently generates derivatized protein crystals by combining random microseeding matrix screening with derivatization with a heavy atom molecule I3C (5-amino-2,4,6-triiodoisophthalic acid). By incorporating I3C into the crystal lattice, the diffraction phase problem can be efficiently solved using single wavelength anomalous dispersion (SAD) phasing. The equilateral triangle arrangement of iodine atoms in I3C allows for rapid validation of a correct anomalous substructure. This protocol will be useful to structural biologists who solve macromolecular structures using crystallography-based techniques with interest in experimental phasing.

Introduction

In the field of structural biology, X-ray crystallography is regarded as the gold standard technique to determine the atomic-resolution structures of macromolecules. It has been utilized extensively to understand the molecular basis of diseases, guide rational drug design projects and elucidate the catalytic mechanism of enzymes^{1,2}. Although structural data provides a wealth of knowledge, the process of protein expression and purification, crystallization and structure determination can be extremely laborious. Several bottlenecks are commonly encountered that hinder the

progress of these projects and this must be addressed to efficiently streamline the crystal structure determination pipeline.

Following recombinant expression and purification, preliminary conditions that are conducive to crystallization must be identified which is often an arduous and time-consuming aspect of X-ray crystallography. Commercial sparse matrix screens that consolidate known and published conditions have been developed to ease this bottleneck^{3,4}.

However, it is common to generate few hits from these initial screens despite using highly pure and concentrated protein samples. Observing clear drops indicates that the protein may not be reaching the levels of supersaturation required to nucleate a crystal. To encourage crystal nucleation and growth, seeds produced from pre-existing crystals can be added to the conditions and this allows for increased sampling of the crystallization space. Ireton and Stoddard first introduced the microseed matrix screening method⁵. Poor quality crystals were crushed to make a seed stock and then added systematically to crystallization conditions containing different salts to generate new diffraction-quality crystals that would not have otherwise formed. This technique was further improved by D'Arcy et al. who developed random microseed matrix screening (rMMS) in which seeds were introduced into a sparse matrix crystallization screen^{6,7}. This improved the quality of crystals and increased the number of crystallization hits on average by a factor of 7.

After crystals are successfully produced and an X-ray diffraction pattern is obtained, another bottleneck in the form of solving the 'phase problem' is encountered. During the data acquisition process, the intensity of diffraction (proportional to the square of the amplitude) is recorded but the phase information is lost, giving rise to the phase problem that halts immediate structure determination⁸. If the target protein shares high sequence identity to a protein with a previously determined structure, molecular replacement can be used to estimate the phase information^{9,10,11,12}. Although this method is fast and inexpensive, model structures may not be available or suitable. The success of the homology model-based molecular replacement method drops significantly as sequence identity falls below 35%¹³. In the absence of a suitable homology model, *ab initio* methods, such as ARCIMBOLDO^{14,15} and AMPLE¹⁶, can be tested. These

methods use computationally predicted models or fragments as starting points for molecular replacement. AMPLE, which uses predicted decoy models as starting points, struggles to solve structures of large (>100 residues) proteins and proteins containing predominately β -sheets. ARCIMBOLDO, which attempts to fit small fragments to extend into a larger structure, is limited to high resolution data (≤ 2 Å) and by the ability of algorithms to expand the fragments into a full structure.

If molecular replacement methods fails, direct methods such as isomorphous replacement^{17,18} and anomalous scattering at a single wavelength (SAD¹⁹) or multiple wavelengths (MAD²⁰) must be used. This is often the case for truly novel structures, where the crystal must be formed or derivatized with a heavy atom. This can be achieved by soaking or co-crystallizing with a heavy atom compound, chemical modification (such as 5-bromouracil incorporation in RNA) or labelled protein expression (such as incorporating selenomethionine or selenocysteine amino acids into the primary structure)^{21,22}. This further complicates the crystallization process and requires additional screening and optimization.

A new class of phasing compounds, including I3C (5-amino-2,4,6-triiodoisophthalic acid) and B3C (5-amino-2,4,6-tribromoisophthalic acid), offer exciting advantages over pre-existing phasing compounds^{23,24,25}. Both I3C and B3C feature an aromatic ring scaffold with an alternating arrangement of anomalous scatters required for direct phasing methods and amino or carboxylate functional groups that interact specifically with the protein and provide binding site specificity. The subsequent equilateral triangular arrangement of heavy metal groups allows for simplified validation of the phasing substructure. At the time of writing,

there are 26 I3C-bound structures in the Protein Data Bank (PDB), of which 20 were solved using SAD phasing²⁶.

This protocol improves the efficacy of the structure determination pipeline by combining the methods of heavy metal derivatization and rMMS screening to simultaneously increase the number of crystallization hits and simplify the crystal derivatization process. We demonstrated this method was extremely effective with hen egg white lysozyme and a domain of a novel lysin protein from bacteriophage P68²⁷. Structure solution using the highly automated Auto-Rickshaw structure determination pipeline is described, specifically tailored for the I3C phasing compound. There exists other automated pipelines that can be used such as AutoSol²⁸, ELVES²⁹ and CRANK2³⁰. Non-fully automated packages such as SHELXC/D/E can also be used^{31,32,33}. This method is particularly beneficial to researchers who are studying proteins lacking homologous models in the PDB, by significantly reducing the number of screening and optimization steps. A prerequisite for this method is protein crystals or a crystalline precipitate of the target protein, obtained from previous crystallization trials.

Protocol

1. Experimental planning and considerations

1. Use pre-existing crystals of the protein of interest, preferably generated through vapor diffusion crystallization. For a generalized protocol of vapor diffusion crystallization, see Benvenuti and Mangani³⁴. Other methods of crystallization such as microbatch under oil and free interface diffusion will require harvesting the crystals prior to crushing to generate microseeds.

2. In the preparation of a seed stock, use the highest quality crystals that can be sacrificed. The highest quality crystal can be judged visually based on morphology or the best diffracting crystal can be selected, if such data is available. It is very likely that even better quality crystals are obtained after optimization through seeding. In the case where no crystals are available, crystalline precipitate such as spherulites and needles can be used.
3. Identify salt crystals. Salt crystals can grow in crystallization screens and can look like protein crystals. Using salt crystals in rMMS will provide no benefit and will waste precious sample, so it is important to eliminate salt false positives.
 1. Salt crystals are loud when they are crushed. Crystals must be crushed to generate a seed stock, so this strategy is particularly relevant. If an audible crack sound is heard when crushing up the crystals, the crystal is likely to be salt.
 2. If the protein contains tryptophan and tyrosine residues, use ultraviolet fluorescence microscopy to identify protein crystals which fluoresce under these lighting conditions.
 3. Use IZIT dye (methylene blue) to stain protein crystals to differentiate them to salt crystals which remain relatively unstained. However, this procedure is more destructive and is only recommended if one has crystals to spare from replicates of the same drop.

NOTE: Although the aforementioned tests may give promising results, salt crystals may still be mistaken for protein crystals. In this case, diffraction experiments can be used to definitively discern between a protein and salt crystal.

2. Preparation of lithium I3C stock

1. Measure out 120 mg of I3C (5-amino-2,4,6-triiodoisophthalic acid) into a 1.5 mL microcentrifuge tube.
2. Dissolve I3C in 200 μ L of 2 M lithium hydroxide. The solution can be gently heated using a heat block at 40-60 $^{\circ}$ C to encourage dissolution. The resulting lithium I3C solution should be brown and has a concentration of 1 M. CAUTION: Lithium hydroxide is corrosive. Safety glasses, gloves and a lab coat should be worn.
3. Measure the pH of the solution. If necessary, add small amounts of 1 M hydrochloric acid or 2 M lithium hydroxide to adjust the pH to between 7 and 8. Add milliQ water to make the final solution volume to 400 μ L. The concentration of the I3C stock solution is 0.5 M.

NOTE: Step 2.3 is optional. The pH of the solution should be between pH 7-8 prior to any pH adjustment. This step should be performed if the protein of interest is strongly affected by pH. The protocol can be paused here. Lithium I3C can be kept in the dark at 4 $^{\circ}$ C for at least two weeks³⁵.

3. Addition of I3C to the protein stock

1. Method 1
 1. Add stock lithium I3C to a 150 μ L aliquot of the target protein. The final concentration should be between 5-40 mM lithium I3C.
2. Method 2 (gentler method)
 1. Prepare a protein dilution buffer that matches the buffer of the target protein. To this dilution buffer, add stock lithium I3C to give a concentration of lithium I3C between 10-80 mM.

2. Dilute the protein 1:1 with protein dilution buffer to give a final concentration of lithium I3C between 5-40 mM.

NOTE: Some proteins will precipitate upon coming into contact with high concentrations of lithium I3C in method 1, while other proteins can tolerate it. Method 2 reduces the likelihood of precipitation. However, this method halves the protein concentration. For proteins that do not have an established crystallization protocol, a protein concentration of 10 mg/mL is generally recommended for initial crystallization screening. An initial molar ratio of I3C to protein of 8 is recommended. Protein concentration and molar ratio of I3C to protein can be optimized after the initial screen.

4. Making a seed stock

1. Make a rounded probe for crushing crystals.
 1. With a Bunsen burner on the blue flame, heat a Pasteur pipette towards its middle. Using a tweezer, pull the end of the Pasteur pipette to draw it out into a thin diameter of less than 0.3 mm.
 2. Once the midsection is thin enough, hold that segment in the flame to separate the pipette at this point and round the end of the pipette to finish the glass probe.

NOTE: Rounded probe crystal crushers are sold by third party vendors. These are an alternative to making rounded probes.
2. Place five 1.5 mL microcentrifuge tubes on ice.
3. Under a light microscope, examine the crystallization tray for a suitable condition to generate microcrystals. Ideally,

good morphology large crystals are selected. However, this technique also works with poor morphology crystals, needles, plates, microcrystals and spherulites.

4. Open up the crystallization tray well. For 96 well crystallization trays sealed with plastic, use a scalpel to cut the plastic sealing the well. For hanging drop trays sealed with grease, the coverslip can be removed using tweezers and inverted onto an even surface.
5. Transfer 70 μL of reservoir solution to a microcentrifuge tube and chill it on ice. To the other microcentrifuge tubes, add 90 μL of reservoir solution and return to ice to chill.
NOTE: If the reservoir does not have enough volume or does not exist (in the case of microbatch under oil), create crystallization reservoir by mixing the appropriate reagents.
6. Agitate the crystal in the drop using the crystal probe to thoroughly crush it up. The crystal needs to be completely crushed up which can be monitored under the microscope.
7. Remove all the liquid from the drop and transfer it to the microcentrifuge tube with the reservoir solution. Mix and subsequently take 2 μL of mixture from the microcentrifuge tube and add it back to the well. Rinse the well with the solution and transfer it to the microcentrifuge tube. Repeat this rinse step once more. From this point on, keep the microcentrifuge tube cold to avoid melting the microseeds in the mixture.
8. Vortex the tube at maximum speed at 4 $^{\circ}\text{C}$ for 3 min, stopping regularly to chill the tube on ice to prevent overheating.

NOTE: Some microseeding protocols add a polytetrafluoroethylene seed bead to the microcentrifuge tube to aid crystal crushing^{7,36}. We have employed the

technique without the use of a seed bead with success, but see no problems with utilizing a seed bead to crush up crystals.

9. Make a 1 in 10 serial dilution of the seed stock by sequentially transferring 10 μL between the chilled reservoir solutions.
10. Store seed stocks that will not be used immediately at -80°C .

5. Setting up an rMMS screen

1. Setting up a 96 well screening plate using a liquid dispensing robot. In the absence of a robot, a multichannel pipette may also be used.
 1. Transfer 75 μL from a deep well block to a 96 well crystallization tray. Add 1 μL to the crystallization drop and 74 μL to the reservoir.
 2. Transfer 1 μL of protein supplemented with lithium I3C, made in step 2, to the crystallization drop.
 3. Transfer 0.1 μL of seed stock to the crystallization drop.
 4. Seal the plate with clear sealing tape and incubate the plate at a constant temperature to allow crystal growth.
2. Setting up a hanging drop screens
 1. Grease the edges of the hanging drop wells (hanging drop crystallization trays can be found in 24 and 48 well formats).
 2. Transfer 500 μL crystallization solution into reservoir.
 3. Near the center of a glass cover slide, place a 1 μL drop of protein supplemented with lithium I3C, made in step 2.

4. Add 1 μL of the crystallization solution to the drop.
5. Transfer 0.1 μL of seed stock to the crystallization drop.
6. Invert the cover slide and seal the crystallization well by pushing the cover slide into the grease.
7. Incubate the plate at a constant temperature to allow crystal growth.

NOTE: With new and untested seed stocks, it is recommended to use the most concentrated seed stock to maximize the chances of getting crystallization hits. Subsequent conditions can be set up with reduced seed concentration to optimize the number of crystals.

3. Inspect crystal trays under a microscope regularly for crystal growth. If crystals are of sufficient quality, they can be harvested for data collection. Crystals can also be used to generate new seed stocks and new rMMS screens to allow for iterative optimization.

6. Data collection

1. Harvest crystals using cryoloops, cryoprotect the crystals and flash cool them in liquid nitrogen. For additional information on flash cooling crystals, refer to Teng³⁷ and Garman and Mitchell³⁸.
 1. During the cryoprotection stage, if the crystal is passed through a new aqueous solution, I3C can be lost from the crystal due to it leeching into the cryoprotection solution. Use lithium I3C in the cryoprotection solution at a concentration that matches the crystallization condition to mitigate this.

2. Crystals grown using this protocol have successfully been cryoprotected using Parabar 10312 oil based cryoprotectant (Hampton Research).

NOTE: The protocol can be paused here while crystals are stored in liquid nitrogen.

CAUTION: Liquid nitrogen can cause cold burns. Liquid nitrogen can also cause asphyxiation if used in enclosed spaces.

2. Mount the crystal on the X-ray source goniometer and collect diffraction data using the protocol specific for the X-ray source.
3. This technique relies on anomalous signal from iodine atoms in I3C. Thus, select the energy of the X-ray to maximize this signal.
 1. Set synchrotron X-ray sources with tunable energies as low as possible. For many macromolecular crystallography beamlines, the lowest configurable energy is 8000 to 8500 eV.
 2. Rotating anode X-ray sources cannot be tuned. Commonly used anode sources with copper have the $K\alpha$ edge at 8046 eV, which provides a good anomalous signal for iodine ($f'' = 6.9$ e). Anode sources with chromium have a $K\alpha$ edge at 5415 eV, which provides a large anomalous signal for iodine ($f'' = 12.6$ e).
4. Radiation damage is a significant problem during data collection as it will degrade the anomalous signal³⁹. Select the exposure time and attenuation of the beam to achieve the best diffraction while minimizing radiation dose.

NOTE: In a similar phasing compound with the iodine atoms replaced with bromine atoms, radiation damage has been shown to cause the radiolysis of the carbon

bromine bond and a reduction in the occupancy of the bromine atoms²⁴.

1. Use inverse beam SAD data collection as a collection strategy. The data is collected in wedges, with opposite wedges collected after each other. This allows Friedel pairs to be collected with an equivalent dose, resulting in an improved measurement of anomalous signal less affected by radiation damage. For example, an eight wedge strategy to collect 360° would involve collecting the data in the order of wedge 1 (0°-45°), wedge 2 (180°-225°), wedge 3 (46°-90°), wedge 4 (225°-270°), wedge 5 (90°-135°), wedge 6 (270°-315°), wedge 7 (135°-180°) and wedge 8 (315°-360°).

NOTE: Continuous rotation is an alternative collection strategy to that of inverse beam data collection. For a recent comparison of the collection strategies, see Garcie-Bonte & Katona⁴⁰.

7. Data processing and structure solution

1. Perform data reduction on the diffraction data using XDS⁴¹, with the aim of maximizing the anomalous signal. Data reduction input parameters are specific to the dataset and may require some trial and error. Here are some recommendations to start.
 1. Set FRIEDEL'S LAW=FALSE. Execute CORRECT twice, setting STRICT_ABSORPTION_CORRECT = TRUE and STRICT_ABSORPTION_CORRECT = FALSE. One run can have a higher anomalous signal than the other. Compare the anomalous signals between the runs using the 'Anomal Corr'

and 'SigAno' disciplines in the output. This provides an indicator of data quality.

2. Run SHELXC on the XDS_ASCII.HKL file for a more accurate indication of anomalous signal. The 'Ranom' discipline will give an indication of the anomalous signal at different resolutions.
2. Run POINTLESS⁴² and AIMLESS⁴³ to scale the data. In AIMLESS, set the parameter ANOMALOUS ON. If the GUI is used, select the option **Separate anomalous pairs for outlier rejection and merging statistics**. Testing different resolution cutoffs may be required to maximize anomalous signal.
3. Solve the protein structure using Auto-Rickshaw automated crystal structure determination pipeline⁴⁴. Auto-Rickshaw will attempt to solve the phase problem and build the crystal structure of the protein automatically with protein modelling and refinement software.
 1. For proteins without a homology model template, run the SAD protocol of Auto-Rickshaw in Advanced Mode. Enter the required parameters.
 1. Select PROTEIN as the molecule type.
 2. Enter the data collection wavelength in angstroms (Å).
 3. Select "I" as substructure element to indicate iodine atoms was used.
 4. Select "i3c" as substructure type to indicate I3C was the phasing molecule.
 5. Select "sub_direct" as the substructure determination method. This method employs SHELXD³² to search for the substructure.
 6. Select "3" as the number of expected substructure per monomer.

7. Enter "1" as the resolution cutoff of substructure search. This allows Auto-Rickshaw to automatically determine a suitable resolution cutoff.
 8. Enter the number of residues in a single monomer, spacegroup of the dataset, and number of molecules in the asymmetric unit based on the Matthews coefficient.
 9. Select the appropriate dissemination level of X-ray data that suits the needs. Selecting "AutoRickshaw developers" will allow Auto-Rickshaw developers to troubleshoot the run if problems arise.
 10. Input the anomalous data as an mtz file.
 11. Input the protein sequence as a seq, pir or txt file. A seq file can be generated in a text editor (such as Notepad++⁹ on Windows or nano in Linux). Create a new file, enter the primary sequence of the protein as one long line or separated by line breaks. Save the file with the .seq file extension.
 12. Enter an institutional email address.
4. Results are delivered via a web-link sent to the email address provided.

NOTE: AutoRickshaw is an automated pipeline that invokes various crystallography software packages to solve an X-ray crystal structure^{32,33,45,46,47,48,49,50,51,52,53,54,55,56,57,58}.

If the Auto-Rickshaw run fails to solve the structure, other Auto-Rickshaw settings can be tested. The structure determination method can be changed to "sub_phassade" to use Phaser⁵⁹ instead of SHELXD³².

The number of expected substructure per monomer can be also increased or decreased.

5. During the experimental phasing of the crystal structure, Auto-Rickshaw will attempt to position heavy atoms in the unit cell, creating a substructure. The equilateral triangle arrangement of iodine atoms in I3C presents an efficient way of validating the substructure. If step 6.3 fails, validating the substructure could aid in troubleshooting structure solution.
 1. Download the list of heavy atom sites from the Auto-Rickshaw results page. It is a hyperlink called "heavy atom sites". This will download a text file with the heavy atom sites.
 2. Change the file extension of the file from .txt to .pdb.
 3. Open the PDB file in Coot⁶⁰. Turn on symmetry to see other heavy atoms from neighboring asymmetric units.
 4. Measure the distances between the heavy atoms, including across asymmetric units. I3C will appear as an equilateral triangle with a side length of 6 angstroms. The presence of a triangle with these dimensions indicates the placements of those heavy atoms are correct.

Representative Results

Incorporating I3C into rMMS can generate new conditions supporting derivatized crystal growth

The efficacy of simultaneous rMMS screening and I3C derivatization was demonstrated in two proteins, hen egg white lysozyme (HEWL, obtained as a lyophilized powder) and the putative Orf11 lysin N-terminal domain (Orf11 NTD) from bacteriophage P68. Each protein was screened against PEG/ION HT under four different conditions including: unseeded, seeded, unseeded with I3C and seeded with I3C (**Figure 1**). For both proteins, the sole addition of I3C did not increase the number of conditions conducive to crystallization. In the case of Orf11 NTD, only one suitable condition was identified with and without I3C (**Figure 1B**). When I3C was added to the HEWL screens, the number of hits was reduced from 31 to 26, highlighting the added complexities of crystallisation when introducing phasing compounds (**Figure 1A**). Consistent with other studies, adding seed to commercial sparse matrix screens to generate an rMMS screen significantly increased the number of possible crystallization conditions for both proteins, resulting in a 2.1 and 6 fold increase for HEWL and Orf11 NTD, respectively^{6,61} (**Figure 1**). Most importantly, simultaneous addition of I3C and seed increased the number of hits relative

to an unseeded screen, demonstrating a 2.3 and 7 fold increase for HEWL and Orf11 NTD, respectively. Many of the crystals from rMMS in the presence of I3C show excellent crystal morphology (**Figure 2**).

Seeding allows careful control of crystal number in I3C rMMS screens

In microseeding experiments, the number of seeds introduced into a crystallization trial can be controlled by dilution of the seed stock and this allows for precise control of nucleation in the drop^{7,36}. This often allows larger crystals to form since there is reduced competition of protein molecules at nucleation sites. This advantage also extends to the I3C-rMMS method and has been demonstrated successfully in both HEWL and Orf11 NTD. Recreation of a crystallization condition identified from the I3C-rMMS screen with a diluted seed stock yielded fewer but larger crystals (**Figure 3**).

SAD phasing can be used to solve the structures from crystals derived from rMMS I3C screen

Crystals grown using the diluted seed stock shown in **Figure 3** were used to solve the structure of the proteins using SAD phasing using diffraction data from a single crystal (**Figure 4**). Data was collected on the Australian Synchrotron MX1 beamline⁶². Detailed data collection and structure solution details are described elsewhere²⁷.

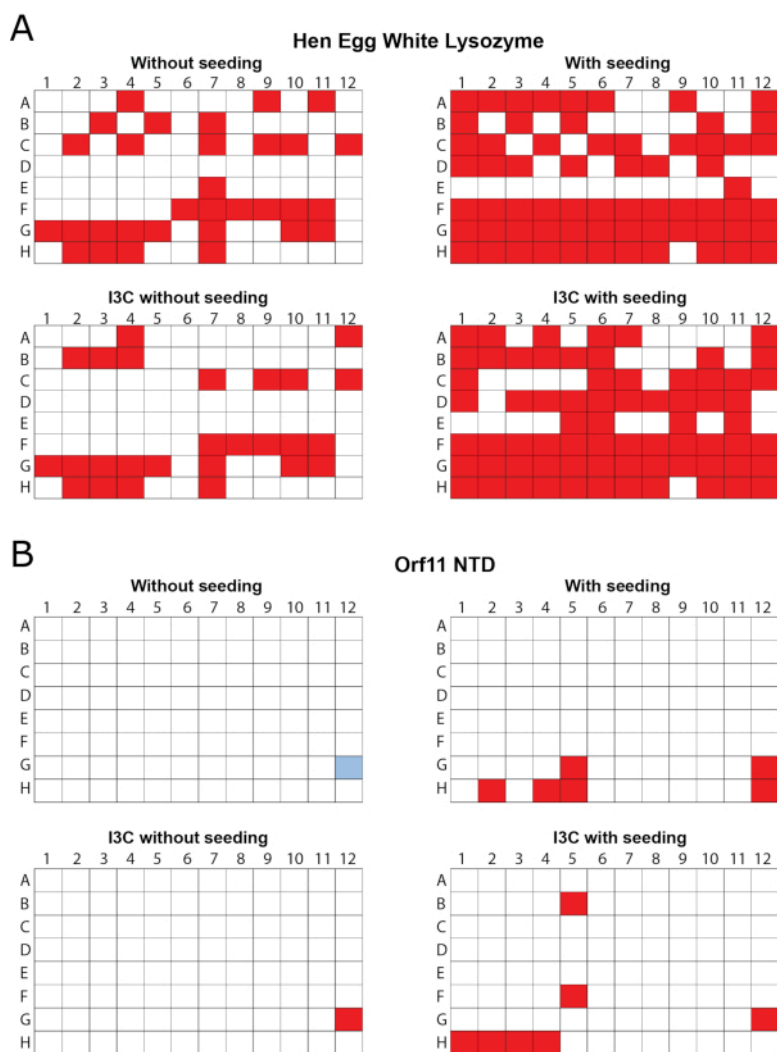


Figure 1 - rMMS was used to generate new conditions for crystal growth in the presence of I3C for two test proteins. 96 well vapor diffusion crystallization screens were carried out using commercial sparse matrix screens. **(A)** Hen egg white lysozyme was tested with the Index HT screen. Trays were seeded with HEWL crystals grown in 0.2 M ammonium tartrate dibasic pH 7.0, 20% (w/v) polyethylene glycol 3350. **(B)** Orf11 NTD from bacteriophage P68 was tested with the PEG/ION screen. Orf11 NTD trays were seeded from crystals from condition G12 from the unseeded screen, shown in blue. Conditions supporting crystal growth are shown in red. rMMS seeding in the presence and absence of I3C both gave significantly more crystal hits than unseeded trays. Figure adapted from Truong et al.²⁷. [Please click here to view a larger version of this figure.](#)

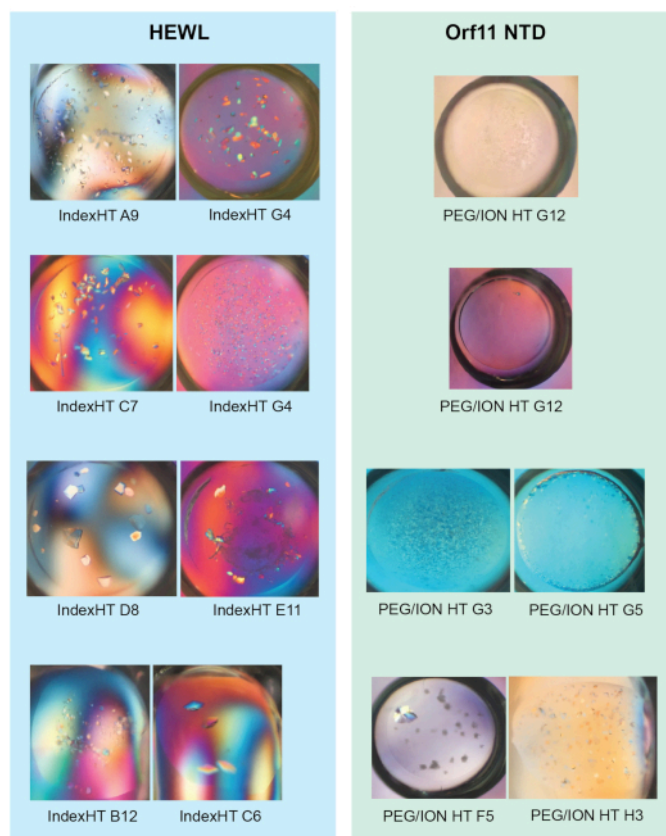


Figure 2 - Representative images of crystals grown from the vapor diffusion trials shown in Figure 1 (a) and (b).

Figure adapted from Truong et al.²⁷. [Please click here to view a larger version of this figure.](#)

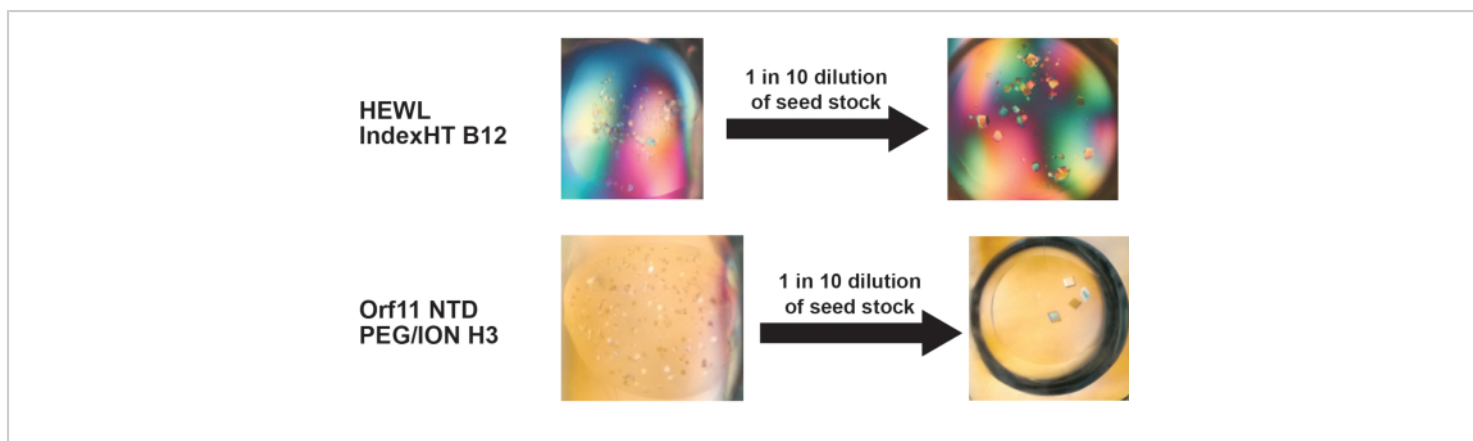


Figure 3 - Dilution of the seed stock is an effective way to reduce nucleation in a crystallization condition found using the I3C-rMMS method, to control the number of crystals that form. Reducing nucleation within a drop often results in crystals growing to larger dimensions. Figure adapted from Truong et al.²⁷. [Please click here to view a larger version of this figure.](#)

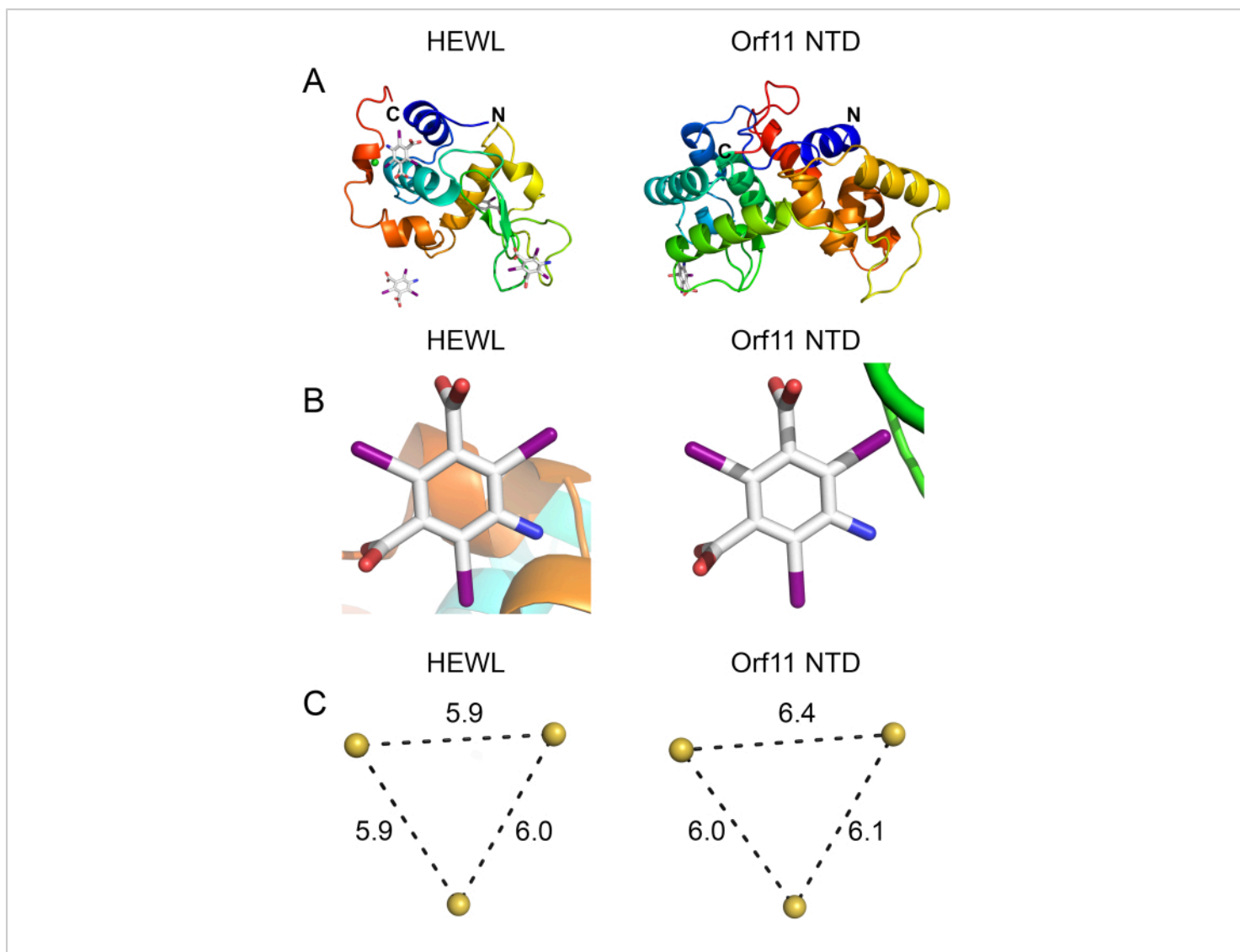


Figure 4 - Orf11 NTD (PDB ID 6O43) and HEWL (PDB ID 6PBB) were crystallized using the I3C-rMMS method and solved using Auto-Rickshaw SAD phasing. (A) Ribbon structures of HEWL and Orf11 NTD solved through experimental phasing. **(B)** I3C molecule bound to HEWL and Orf11 NTD. **(C)** Anomalous iodine atoms in I3C are arranged in an equilateral triangle of 6 Å. Thus the presence of this triangle in the phasing substructure indicates that there is an I3C molecule in that position. [Please click here to view a larger version of this figure.](#)

Discussion

Structure determination of a novel protein in the absence of a suitable homology model for molecular replacement requires experimental phasing. These methods require incorporation of heavy atoms into the protein crystal which adds a level

of complexity to the structure determination pipeline and can introduce numerous obstacles that must be addressed. Heavy atoms can be incorporated directly into the protein through labelled expression using selenomethionine and selenocysteine. As this method is costly, laborious and can result in lower protein yields, labelled protein is often

expressed after crystallization conditions has been found and optimized with unlabeled protein. Alternatively, crystals can be derivatized by soaking in a solution containing heavy atoms^{22,63,64}. This method often uses high quality crystals and is therefore performed after a robust crystallization method has already been developed. Successfully obtaining a derivatized crystal using this method requires further optimization of soaking procedures and screening of different phasing compounds, therefore adding more time to an already laborious process.

Co-crystallization of the protein with the heavy atom can be performed at the screening stage, thus efficiently streamlining the process and reducing crystal manipulation steps that can cause damage. However, there still exists the potential scenario of obtaining few initial crystallization hits and the problem of choosing a compatible heavy atom compound. Many currently available phasing compounds are incompatible with precipitants, buffers and additives commonly found in crystallization conditions. They may be insoluble in sulphate and phosphate buffers, chelate to citrate and acetate, react unfavorably with HEPES and Tris buffers or become sequestered by DTT and β -mercaptoethanol²¹. As the I3C phasing compound does not suffer from these incompatibilities, it is a robust phasing compound that could be amenable to many different conditions.

In this study, a streamlined method of producing derivatized crystals ready for SAD phasing through simultaneous co-crystallization of the I3C phasing compound and rMMS is presented. The combination of both techniques increases the number of crystallization hits, with many of the conditions having improved morphology and diffraction characteristics. In both Orf11 NTD and HEWL test cases, new conditions in the I3C-rMMS screen were identified that were absent

when I3C was not present. Potentially, I3C may bind favorably to the protein, facilitating the formation and stabilization of crystal contacts²⁷. In turn, this may induce crystallization and possibly improve diffraction characteristics. Besides being a compound compatible with sparse matrix screens, I3C is also an attractive phasing compound due to its intrinsic properties. The functional groups that alternate with iodine on the aromatic ring scaffold allow specific binding to proteins. This leads to greater occupancy and potentially reduces background signal²³. Furthermore, the arrangement of anomalous scatterers in an equilateral triangle is obvious in the substructure and can be used to rapidly validate binding of I3C (**Figure 4B** and **4C**). Finally, it can produce an anomalous signal with tunable synchrotron radiation as well as chromium and copper rotating anode X-ray sources. Thus, it can be applied to many different workflows. As I3C is widely available and inexpensive to purchase, this approach is within reach for most structural biology laboratories.

There are several experimental considerations that must be addressed when using the I3C-rMMS method. This method cannot be applied if initial crystalline material of the protein cannot be obtained. In difficult cases, crystalline material from a homologous protein can also be used to generate seed stock. This cross-seeding approach to rMMS has shown some promising results⁷. Optimizing crystal number through dilution of the seed stock is a crucial step, which should not be overlooked, to maximize the chance of producing high quality large crystals and acquiring suitable diffraction data. If there are few I3C sites identified in the asymmetric unit, conditions conducive to crystallization should be further optimized with an increased concentration of I3C. This may increase the occupancy of I3C to maximize the anomalous signal and aid crystal derivatization.

There can be cases where this technique may not be the optimal method to derivatize protein crystals. As the size of a protein or protein-complex increases, the limited number of I3C sites on the protein surface may not provide sufficient phasing power to solve the structure. In these scenarios where protein size is suspected to be impeding phasing, selenomethionine labelling of the protein may be a more viable approach to phasing the protein. If the protein has adequate numbers of methionine residues in the protein (recommended having at least one methionine per 100 residues⁶⁵) and high efficiency selenomethionine incorporation into a protein can be achieved (such as in bacterial expression systems⁶⁶), multiple high occupancy selenium atoms will be present in the crystals to phase the structure.

In addition, some proteins may inherently be unsuited for derivatization with I3C. I3C binding sites on proteins are dependent on protein structure. There may exist proteins that naturally have few exposed patches compatible with I3C binding. Thus, it is not unforeseeable that there may be difficulties in co-crystallizing some target proteins with I3C.

Disclosures

The authors have nothing to disclose.

Acknowledgments

This research was undertaken on the MX1 beamline at the Australian Synchrotron, part of ANSTO. The authors would like to acknowledge members of the Shearwin and Bruning laboratories for discussions on this work. The authors would also like to acknowledge Dr. Santosh Panjekar and Dr. Linda Whyatt-Shearwin who contributed to the original work that pioneered this protocol.

The following funding is acknowledged: Australian Research Council (grant Nos. DP150103009 and DP160101450 to Keith E. Shearwin); University of Adelaide (Australian Government Research Training Program stipend scholarship to Jia Quyen Truong and Stephanie Nguyen).

References

1. Zheng, H., Hou, J., Zimmerman, M.D., Wlodawer, A., Minor, W. The future of crystallography in drug discovery. *Expert Opinion on Drug Discovery*. **9** (2), 125-137 (2014).
2. Oakley, A.J., Wilce, M.C.J. Macromolecular crystallography as a tool for investigating drug, enzyme and receptor interactions. *Clinical and Experimental Pharmacology and Physiology*. **27** (3), 145-151 (2000).
3. Jancarik, J., Kim, S.H. Sparse matrix sampling. A screening method for crystallization of proteins. *Journal of Applied Crystallography*. **24** (pt 4), 409-411 (1991).
4. Newman, J. et al. Towards rationalization of crystallization screening for small- To medium-sized academic laboratories: The PACT/JCSG+ strategy. *Acta Crystallographica Section D: Biological Crystallography*. **61** (10), 1426-1431 (2005).
5. Ireton, G.C., Stoddard, B.L. Microseed matrix screening to improve crystals of yeast cytosine deaminase. *Acta Crystallographica Section D: Biological Crystallography*. **60** (3), 601-605 (2004).
6. D'Arcy, A., Villard, F., Marsh, M. An automated microseed matrix-screening method for protein crystallization. *Acta Crystallographica Section D: Biological Crystallography*. **63** (4), 550-554 (2007).
7. D'Arcy, A., Bergfors, T., Cowan-Jacob, S.W., Marsh, M. Microseed matrix screening for optimization in protein crystallization: What have we learned?

- Acta Crystallographica Section:F Structural Biology Communications*. **70** (9), 1117-1126 (2014).
8. Taylor, G. The phase problem. *Acta Crystallographica - Section D Biological Crystallography*. **59** (11), 1881-1890 (2003).
 9. Rossmann, M.G. The molecular replacement method. *Acta Crystallographica Section A*. **46** (2), 73-82 (1990).
 10. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J. Phaser crystallographic software. *Journal of Applied Crystallography*. **40** (4), 658-674 (2007).
 11. Millán, C., Jiménez, E., Schuster, A., Diederichs, K., Usón, I. ALIXE: a phase-combination tool for fragment-based molecular replacement. *Acta Crystallographica Section D*. **76** (3), 209-220 (2020).
 12. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology*. **75**, 861-877 (2019).
 13. Abergel, C. Molecular replacement: Tricks and treats. *Acta Crystallographica Section D: Biological Crystallography*. **69** (11), 2167-2173 (2013).
 14. Pröpper, K. et al. Structure solution of DNA-binding proteins and complexes with ARCIMBOLDO libraries. *Acta Crystallographica Section D: Biological Crystallography*. **70** (6), 1743-1757 (2014).
 15. Rodríguez, D.D. et al. Crystallographic ab initio protein structure solution below atomic resolution. *Nature Methods*. **6** (9), 651-653 (2009).
 16. Bibby, J., Keegan, R.M., Mayans, O., Winn, M.D., Rigden, D.J. AMPLE: A cluster-and-truncate approach to solve the crystal structures of small proteins using rapidly computed ab initio models. *Acta Crystallographica Section D: Biological Crystallography*. **68** (12), 1622-1631 (2012).
 17. Green, D.W., Ingram, V.M., Perutz, M.F. The structure of haemoglobin - IV. Sign determination by the isomorphous replacement method. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*. **225** (1162), 287-307 (1954).
 18. Blow, D.M., Rossmann, M.G. The single isomorphous replacement method. *Acta Crystallographica*. **14** (11), 1195-1202 (1961).
 19. Wang, B.C. Resolution of phase ambiguity in macromolecular crystallography. *Methods in Enzymology*. **115** (C), 90-112 (1985).
 20. Hendrickson, W.A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*. **254** (5028), 51-58 (1991).
 21. Pike, A.C.W., Garman, E.F., Krojer, T., Von Delft, F., Carpenter, E.P. An overview of heavy-atom derivatization of protein crystals. *Acta Crystallographica Section D: Structural Biology*. **72** (3), 303-318 (2016).
 22. Dauter, Z., Dauter, M., Rajashankar, K.R. Novel approach to phasing proteins: Derivatization by short cryo-soaking with halides. *Acta Crystallographica Section D: Biological Crystallography*. **56** (2), 232-237 (2000).
 23. Beck, T., Krasauskas, A., Gruene, T., Sheldrick, G.M. A magic triangle for experimental phasing of macromolecules. *Acta Crystallographica Section D: Biological Crystallography*. **64** (11), 1179-1182 (2008).
 24. Beck, T., Gruene, T., Sheldrick, G.M. The magic triangle goes MAD: Experimental phasing with a bromine

- derivative. *Acta Crystallographica Section D: Biological Crystallography*. **66** (4), 374-380 (2010).
25. Beck, T., Da Cunha, C.E., Sheldrick, G.M. How to get the magic triangle and the MAD triangle into your protein crystal. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*. **65** (10), 1068-1070 (2009).
 26. Berman, H.M. et al. The Protein Data Bank. *Nucleic Acids Research*. (2000).
 27. Truong, J.Q., Panjikar, S., Shearwin-Whyatt, L., Bruning, J.B., Shearwin, K.E. Combining random microseed matrix screening and the magic triangle for the efficient structure solution of a potential lysin from bacteriophage P68. *Acta Crystallographica Section D: Structural Biology*. **75** (7), 670-681 (2019).
 28. Terwilliger, T.C. et al. Decision-making in structure solution using Bayesian estimates of map quality: The PHENIX AutoSol wizard. *Acta Crystallographica Section D: Biological Crystallography*. **65** (6), 582-601 (2009).
 29. Holton, J., Alber, T. Automated protein crystal structure determination using ELVES. *Proceedings of the National Academy of Sciences of the United States of America*. **101** (6), 1537-1542 (2004).
 30. Skubák, P., Pannu, N.S. Automatic protein structure solution from weak X-ray data. *Nature Communications*. **4** (2013).
 31. Sheldrick, G.M. Crystal structure refinement with SHELXL. *Acta Crystallographica Section C: Structural Chemistry*. **71**, 3-8 (2015).
 32. Schneider, T.R., Sheldrick, G.M. Substructure solution with SHELXD. *Acta Crystallographica Section D: Biological Crystallography*. **58** (10 I), 1772-1779 (2002).
 33. Sheldrick, G.M. Macromolecular phasing with SHELXE. *Zeitschrift für Kristallographie*. **217** (12), 644-650 (2002).
 34. Benvenuti, M., Mangani, S. Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nature Protocols*. (2007).
 35. Beck, T. *Sticky triangles: New tools for experimental phasing of biological macromolecules*. (2010).
 36. Luft, J.R., DeTitta, G.T. A method to produce microseed stock for use in the crystallization of biological macromolecules. *Acta Crystallographica Section D: Biological Crystallography*. **55** (5), 988-993 (1999).
 37. Teng, T.-Y. Mounting of crystals for macromolecular crystallography in a free-standing thin film. *Journal of Applied Crystallography*. **23** (5), 387-391 (1990).
 38. Garman, E.F., Mitchell, E.P. Glycerol concentrations required for cryoprotection of 50 typical protein crystallization solutions. *Journal of Applied Crystallography*. **29**, 584-587 (1996).
 39. Garman, E.F., Weik, M. X-ray radiation damage to biological samples: recent progress. *Journal of Synchrotron Radiation*. **26** (4), 907-911 (2019).
 40. Garcia-Bonete, M.J., Katona, G. Bayesian machine learning improves single-wavelength anomalous diffraction phasing. *Acta Crystallographica Section A: Foundations and Advances*. **75**, 851-860 (2019).
 41. Kabsch, W. XDS. *Acta Crystallographica Section D: Biological Crystallography*. **66** (2), 125-132 (2010).
 42. Evans, P. Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography*. **62** (1), 72-82 (2006).
 43. Evans, P.R., Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallographica*

- Section D: *Biological Crystallography*. **69** (7), 1204-1214 (2013).
44. Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S., Tucker, P.A. Auto-Rickshaw: An automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallographica Section D: Biological Crystallography*. **61** (4), 449-457 (2005).
 45. Jones, T.A., Thirup, S. Using known substructures in protein model building and crystallography. *The EMBO journal*. **5** (4), 819-822 (1986).
 46. Kleywegt, G.J., Jones, T.A. Template convolution to enhance or detect structural features in macromolecular electron-density maps. *Acta Crystallographica Section D: Biological Crystallography*. **53** (2), 179-185 (1997).
 47. Perrakis, A., Morris, R., Lamzin, V.S. Automated protein model building combined with iterative structure refinement. *Nature Structural Biology*. **6** (5), 458-463 (1999).
 48. Morris, R.J. et al. Breaking good resolutions with ARP/wARP. *Journal of Synchrotron Radiation*. **11** (1), 56-59 (2004).
 49. Yao, D.Q. et al. SAD phasing by OASIS-2004: Case studies of dual-space fragment extension. *Acta Crystallographica Section D: Biological Crystallography*. **62** (8), 883-890 (2006).
 50. Hao, Q. ABS: A program to determine absolute configuration and evaluate anomalous scatterer substructure. *Journal of Applied Crystallography*. **37** (3), 498-499 (2004).
 51. Collaborative Computational Project Number 4 The CCP4 suite: Programs for protein crystallography. *Acta Crystallographica Section D: Biological Crystallography*. **50** (5), 760-763 (1994).
 52. Sheldrick, G.M., Hauptman, H.A., Weeks, C.M., Miller, R., Usón, I. Ab initio phasing. *International Tables for Crystallography*. 333-345 (2006).
 53. Smith, G.D. Matching selenium-atom peak positions with a different hand or origin. *Journal of Applied Crystallography*. **35** (3), 368-370 (2002).
 54. Pannu, N.S., McCoy, A.J., Read, R.J. Application of the complex multivariate normal distribution to crystallographic methods with insights into multiple isomorphous replacement phasing. *Acta Crystallographica - Section D Biological Crystallography*. **59** (10), 1801-1808 (2003).
 55. Pannu, N.S., Read, R.J. The application of multivariate statistical techniques improves single-wavelength anomalous diffraction phasing. *Acta Crystallographica Section D: Biological Crystallography*. **60** (1), 22-27 (2004).
 56. De La Fortelle, E., Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods in Enzymology*. **276**, 472-494 (1997).
 57. Cowtan, K. Joint CCP4 and ESF-EACBM Newsletter on Protein. *Crystallography*. **31**, 34-38, at <<https://ci.nii.ac.jp/naid/10010645386/en/>> (1994).
 58. Terwilliger, T.C. Maximum-likelihood density modification. *Acta Crystallographica Section D: Biological Crystallography*. **56** (8), 965-972 (2000).

59. Read, R.J., McCoy, A.J. Maximum-likelihood determination of anomalous substructures. *Acta Crystallographica Section D: Structural Biology*. (2018).
60. Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*. **66** (4), 486-501 (2010).
61. Till, M. et al. Improving the success rate of protein crystallization by random microseed matrix screening. *Journal of visualized experiments : JoVE*. (78), e50548 (2013).
62. McPhillips, T.M. et al. Blu-Ice and the distributed control system: Software for data acquisition and instrument control at macromolecular crystallography beamlines. *Journal of Synchrotron Radiation*. **9** (6), 401-406 (2002).
63. Nagem, R.A.P., Polikarpov, I., Dauter, Z. Phasing on Rapidly Soaked Ions. *Methods in Enzymology*. **374**, 120-137 (2003).
64. Taylor, G.L. Introduction to phasing. *Acta Crystallographica Section D: Biological Crystallography*. **66** (4), 325-338 (2010).
65. Hendrickson, W.A., Ogata, C.M. Phase determination from multiwavelength anomalous diffraction measurements. *Methods in Enzymology*. **276**, 494-523 (1997).
66. Doublé, S. Production of Selenomethionyl Proteins in Prokaryotic and Eukaryotic Expression Systems. *Macromolecular Crystallography Protocols. Methods in Molecular Biology*. 91-108 (2007).

7 CONCLUSIONS AND FUTURE DIRECTIONS

7.1. Bacteriophage 186 as a model for a genetic switch

Bacteriophage present a unique opportunity to study gene regulatory mechanisms in the context of a whole genome. Temperate bacteriophage have been considered to contain a genetic “switch” as they choose between the production of new phage virions and the dormant persistence of the phage genome within a host bacterium. To understand prokaryotic gene regulation, the mechanisms that temperate phages employ to lysogenize and lyse their host, and decide between these two fates have been extensively investigated over many decades. The most thoroughly studied temperate phage, phage λ , has provided many seminal contributions to the fields of prokaryotic transcriptional regulation and biological decision making (Ptashne, 2004). However, bacteriophage λ is just a single example amongst many of nature’s solutions to creating a genetic switch. Through the study of other temperate bacteriophage, alternate gene regulation mechanisms to achieve a bistable switch can be uncovered.

Bacteriophage 186 has used as a useful counterpoint model to bacteriophage λ (Dodd and Egan, 2002; Pinkett *et al.*, 2006; Murchland *et al.*, 2014). Prior to this work, work by the Shearwin/Dodd/Egan laboratory had provided a detailed picture on the lysis-lysogenic system of 186. The specific functions of many gene products involved in decision making, such as transcriptional activators and repressors, are known. The transcriptional cascade observed in the establishment of the lytic and lysogenic cycles after initial infection had been mapped. Refer to Section 1.2.1 for more details. To divert phage development away from the default lytic cascade, two key transcriptional regulators, CII and CI, are required. CII establishes the initial pool of the immunity repressor CI. After the initial pool of CI is generated, CI activates its own transcription and represses the lytic promoter to generate a stable state where lytic functions of the phage are shut off (Neufing *et al.*, 1996). The work described in this thesis has expanded our knowledge of gene regulation in 186 by providing a structural basis for understanding the function of these transcriptional regulators. As Stephen A. Wainwright once stated, “Structure without function is a corpse; function without structure is a ghost”. Form and function are never separate, even within the biochemical world. The behaviour of transcriptional regulators must make sense when considering the structures they adopt.

7.1.1 Understanding the function of CII in establishment of lysogeny

CII is a transcriptional activator that is essential for the efficient establishment of lysogeny in 186. Prior to this work, it was understood that CII activates p_E via the recruitment of RNA polymerase, with limited structural information to explain its activity. The protein oligomerizes into dimers and tetramers, and bind to 7mer half sites at p_E , with an unusual wider spacing of two turns of the DNA helix apart (Shearwin and Egan, 2000). CII was shown to activate p_E via interactions with RNA polymerase σ^{70} and the CTD of the α subunit via two epitopes around E46 and R115, respectively (Murchland *et al.*, 2014).

In Chapter 2, I presented several atomic resolution crystal structures of the CII protein. At the time of writing, these structures were novel and there were no known structural homologues to 186 CII in the Protein Data Bank. 186 CII homologues are found in genomes across many bacterial species, possibly within prophage DNA or remnants of prophage DNA. Thus, the

structure of CII could potentially be used as a model to understand a wider family of CII-like proteins. Through molecular docking approaches, guided by data from mutagenesis and mass spectrometry experiments, a structural model of CII stabilizing RNA polymerase at the p_E promoter was proposed. In this model, a CII tetramer is positioned at the p_E promoter and interacts with RNA polymerase via two epitopes. The CII epitope containing E46 and R17 contacts K593 of the σ^{70} subunit of RNA polymerase domain 4 positioned around the -35 site of p_E . E46 forms an ionic interaction with K593. The CII epitope containing R115 is proposed to contact residue D280 on the CTD of the α -subunit. α -CTD binds to the minor groove of DNA between the p_E half sites, with potentially two α -CTDs stabilized by a single CII tetramer. The stabilization of multiple subunits of RNA polymerase (two or potentially three subunits) may be the key to its potent activation ability. The structure also presents a simple structural explanation of the rapid inactivation of CII by proteolysis. The protease RseP is recruited by the C-terminal degradation tag (Murchland *et al.*, 2021). Cleavage of the protein disrupts the tetramerization interface of the protein and prevents the formation of an active CII tetramer to abolish its activity. Recent work from Murchland *et al.* (2021) suggests that CII's role in the switch is to rapidly equilibrate CI to lysogenic levels and that it is achieved with a potent activator with a short half-life. Our work provides a structural basis to understand how CII satisfies both of these functions.

Several aspects of this model require further validation. The contact between α -CTD residue D280 and CII residue R115 was inferred from scanning alanine mutagenesis experiments (Chapter 2) and an a previous genetic screen of CII (Murchland *et al.*, 2014). However, reduction in activity from mutating an α -CTD residue could be indirect, due to destabilization of α -CTD, which has been observed in residues in other screens using the same scanning alanine mutant library. The possibility of D280 interacting with other residues that are not R115 cannot be excluded. As these residues carry the opposite charge, an interaction between D280 and R115 is likely to be an ionic interaction. This interaction could be validated using a charge swap experiment, where a D280R and D280K mutation would be predicted to restore lost activity observed in the R115E mutant of CII (Murchland *et al.*, 2014).

CII simultaneously stabilizing both α subunits was predicted from the model, as there were no physical constraints identified that prevents such a complex from forming. One potential experiment that can test this prediction is to determine the region of the promoter that α -CTD contacts by hydroxyl radical-based DNA cleavage. In these experiments, with methods described by Murakami *et al.* (1997), (*p*-bromoacetamidobenzyl)-EDTA·Fe is conjugated onto the RNA polymerase α subunit and incorporated into a RNA polymerase holoenzyme. If α -CTD is stabilized in the positions proposed in the model, DNA cleavage would be observed at two sites, centering at around -40 and -56 positions at the p_E promoter.

The most direct way to confirm this structural model is to experimentally determine the structure of a CII interacting with DNA and/or RNA polymerase. Approaches to solving these structures could include X-ray crystallography, small angle X-ray scattering and protein cryo-electron microscopy. Complexes that potentially could be structurally characterized include

- CII145 in complex with $pE31$ or $pE35$ DNA
- CII145 in complex with pE DNA and the σ^{70} subunit of RNA polymerase
- CII145 in complex with pE DNA and α -CTD
- CII145 in complex with pE DNA and the RNA polymerase holoenzyme

7.1.2 Maintenance of lysogeny by CI

Pinkett et al. (2006) proposed a structural model to explain the maintenance of lysogeny by 186 CI. The crystal structure of the 186 CI dimer revealed that it was a two-domain protein, with the CTD mediating oligomerization. The crystal structure of the CI-CTD revealed it adopted a 14mer ring (heptamer of dimers) in the crystal lattice. This led to the model that CI forms a 14mer wheel (Figure 1-15). The positioning of DNA binding domains on the outside of the wheel allows phage DNA to wrap on and off the wheel to repress the lytic promoters, whilst activating the lysogenic promoter p_L (Figure 1-16). Wang et al. (2013) employed atomic force microscopy (AFM) to demonstrate 186 p_R - p_L DNA can wrap around a CI oligomeric disc. The formation of a 14mer CI wheel is central to this model of cooperative regulation by 186 CI. In this thesis, I describe experiments to structurally characterize the intact CI complex to validate this model.

In Chapter 3, CI oligomerization was analyzed by native mass spectrometry. The experiments determined that CI forms dimers that sequentially oligomerize to form dimers, tetramers, hexamers, octamers, 10mers and 12mers. However, a 14mer species was not identified by mass spectrometry and was the first piece of evidence to cast doubt on the 14mer wheel model. Chapter 4 describes the structural characterization of the CI-hybrid protein. This protein is a translational fusion of the λ CI NTD (residues 1-92) with the 186 CI CTD (residues 84-192) and was postulated to oligomerize in the same manner as 186 CI. Mass spectrometry also revealed that the CI-hybrid protein forms a 12mer oligomer. The CI-hybrid protein adopted a 12mer (hexamer of dimers) circular ring assembly in the crystal structure, that is very similar to the previously postulated 14mer wheel assembly of CI, but with one less dimeric subunit. These results taken together suggest that 186 CI also oligomerizes into a 12mer ring.

In Chapter 4, I also discuss the ramifications of updating the model of transcriptional regulation with a 12mer CI wheel. p_R - p_L DNA wrapping on and off a 12mer wheel still adequately explains the observed behavior of 186 CI, provided that DNA entering and exiting the wheel does not sterically clash to prevent binding of operators to DNA. In the original model, DNA entry and exit into and out of the wheel on adjacent dimers of the wheel was assumed to be sterically unfavorable in order to adequately explain the transcriptional regulation.

The studies outlined in Chapters 3 and 4 did not yield a structure of a CI wheel nor did they investigate interactions between CI and DNA. Solving a crystal structure of full-length CI as a higher-order oligomer has proven difficult, with previous attempts described in Chapter 3 and Pinkett (2006). An alternate technique to solve the structure of the CI wheel is protein cryo-electron microscopy (cryoEM). This technique is most amenable for larger protein complexes, although there has been some success with sub-100 kDa proteins and complexes (Merk *et al.*, 2016; Herzik, Wu and Lander, 2019). A 12mer CI wheel has a molecular mass of 248 kDa and is a suitable size for cryo-EM.

As discussed in Chapter 3, 186 CI contains two domains connected via a flexible linker. This linker allows the NTD to be move relative to the CTD domains. Flexibility in a protein or complex may make solving the structure of the 12mer CI complex via cryo-EM and X-ray crystallography prohibitively difficult. Flexible regions of proteins in crystals often impede protein crystallization by introducing heterogeneity into the system (Dale, Oefner and D'Arcy,

2003). Conformational heterogeneity due to flexibility can adversely affect various cryo-EM image processing steps including the alignment, classification and determination of the Euler angle. Additionally, flexible domains have lower densities as different species in the ensemble are conformationally averaged (Spahn and Penczek, 2009). Restricting the movement of these domains could aid in structure determination via cryo-EM and crystallography. This restriction could be achieved by complexing 186 DNA (such as p_R - p_L DNA) with CI to reduce the motions of the DNA binding domains. Furthermore, characterizing a CI-DNA complex naturally provides structural information on how CI can interact with DNA. This information will help us more comprehensively model the 186 genome interacting with the CI wheel to paint a detailed picture as to how regulation is achieved at this alternate genetic switch.

7.2. Temperature sensitive repressors from phage components

In Chapter 5, I characterized temperature sensitive mutants of the 186 CI and CI-hybrid proteins, which were named 186 CI^{G22R} and hybrid^{ts} proteins, respectively. These repressors allowed for the temperature-controlled expression of genes and are viable alternatives to the commonly used λ cI857- p_{RM} - p_R temperature-controlled expression systems. This work aimed to address two limitations of the λ cI857-based systems. Firstly, efficient induction of the promoter in λ cI857 systems require heating the system to 40-42°C (Villaverde *et al.*, 1993). These temperatures are often suboptimal for production-based applications, as high temperatures can lead to protein misfolding and trigger the heat-shock and stringent stress responses, leading to reduced yields. Secondly, many synthetic gene circuits also employ λ cI as a transcriptional regulator, making those systems incompatible with λ cI857.

The temperature-controlled induction of gene expression was measured in three systems, λ cI857- p_{RM} - p_R , hybrid^{ts}- p_{RM} - p_R and 186 CI^{G22R}- p_R - p_L . In all three systems, the target promoter was tightly repressed at 30°C. At 37°C, the hybrid^{ts} and 186 CI^{G22R} systems displayed significantly more derepression of their target promoters than λ cI857. Thus, these systems could be used for induction of gene expression at 37°C, which may reduce the heat stresses placed on these systems. In addition, 186 CI^{G22R} acts on the 186 p_R promoter, distinct from λ cI regulated promoters, allowing it to be used in systems utilizing λ cI. These expression systems can also be used at 40°C, where they show close to maximal induction.

The hybrid^{ts} repressor could be used to optimize existing λ cI857- p_{RM} - p_R system. Our work has shown that substituting λ cI857 with hybrid^{ts} in these systems allows for increased gene expression from λ p_R at 37°C. However, λ cI857 has been implemented into other temperature sensitive expression modules outside the described λ cI857- p_{RM} - p_R layout, including some that utilize the λp_L as the promoter for inducible expression (Table 7-1). It would be valuable to know if hybrid^{ts} also works in these systems and whether it provides any regulatory benefits, such as higher induction at lower temperatures.

Table 7-1 – Examples of λ cI857 based temperature sensitive expression systems reported in the literature

System	Protein expressed	Notes	Reference
λ pL, λ cI857	β -galactosidase	Both λ cI857 and λ pL elements are on the same plasmid	Villaverde et al. (1993)
λ pL, λ cI857	IHF α , β	IHF α , β expressed from two different plasmids in the same cell. Expression of both proteins driven by λ pL promoter. λ cI857 expressed from an integrated prophage	Nash et al. (1987)
λ pRpL, λ cI857	TNF- α	λ pR and λ pL promoters are placed in tandem to increase promoter strength. λ pR, λ pL and λ cI857 are all present on the same plasmid.	Menart et al. (2003)
λ cI857- p_{RM} - p_R	phiX174 lysin gene E	λ cI857, λp_{RM} , λp_R elements were present on the same plasmid	Witte and Lubitz (1989)

λ cI857 has been shown to work in other bacterial systems such as *Salmonella enterica*, *Bacillus subtilis*, several species of *Pseudomonas* and *Klebsiella pneumoniae* (Winstanley *et al.*, 1989; Breitling, Sorokin and Behnke, 1990; Jawale *et al.*, 2012). One application that has been explored is the generation of bacterial ghost vaccines. In the target bacteria, the phiX174 lysin gene E is placed under the control of a promoter repressed by λ cI857. Elevating the temperature induces the lysin gene, leading to cell lysis. The resulting empty bacterial envelopes are investigated as potential vaccines against that bacteria or as an adjuvant in a vaccine preparation (Jechlinger *et al.*, 2005; Fu *et al.*, 2021). Future work could investigate the feasibility of substituting λ cI857 with hybrids and 186 CI^{G22R} in these systems. The use of hybrids and 186 CI^{G22R} could lower the temperatures at which the bacteria lyse, to reduce heat induced alterations of surface envelope antigens in bacterial ghost preparations.

It is impossible to anticipate all the potential use cases of temperature sensitive expression systems in advance. These applications with differing temperature requirements could require further optimization of existing systems or the development of new systems. The hybrids and 186 CI^{G22R} systems could be optimized at the promoter level. The two proteins act on different promoters, λp_L and 186 p_B , respectively. Substituting with these promoters could be used to adjust the promoter strength and potentially shift the induction temperatures required. Modified promoters such the tandem λ pRpL promoter, described by Menart et al. (2003) could be also tested. Additional tweaking of promoter strength and sensitivity to temperature induction could be achieved via the mutagenesis of the promoter and repressor binding sites. Strategies to carry this out have been described in Jechlinger et al. (1999, 2005) and Rohlhill, Sandoval and Papoutsakis (2017).

Bacteriophage have been a fruitful source of biological components for synthetic biology. Examples include λ CI (Elowitz and Leibier, 2000), T7 RNA polymerase with the T7 promoter (Noireaux, Bar-Ziv and Libchaber, 2003; Kim, White and Winfree, 2006), and Bxb1 and Cre

recombinases (Friedland *et al.*, 2009; Bonnet, Subsoontorn and Endy, 2012). Our work has reiterated this point and used phage components to expand the toolkit of available tunable promoter expression systems. Future work in this field could focus on developing new temperature-controlled expression systems to cover a wider variety of use cases.

Temperate bacteriophage themselves present a straightforward resource for generating new temperature sensitive transcriptional regulators. Temperate phages plated onto a lawn of their host bacteria generate turbid plaques. Many temperate phages, such as λ , 186, TP901-1 and P2, use an immunity repressor to repress the lytic promoters of the phage (Ptashne *et al.*, 1980; Lundqvist and Bertani, 1984; Dodd and Egan, 1996; Pedersen, Ligowska and Hammer, 2010). Inactivation of this immunity repressor results in clear plaques. Mutant phages with a temperature sensitive immunity repressor will give rise to clear and turbid plaques when grown at the restrictive and permissible temperatures, respectively. λ phage with λ cI857 was isolated in this manner (Lieb, 1966). When interpreting results from these plaque assays, mutations should be mapped onto the genome as temperature sensitive mutations in other lysogeny promoting factors, such as the integrase and CII (in the case of λ and 186) could give rise to the same plaque behaviour. A temperature sensitive 186 CII activator could be valuable for certain production-based applications.

In many cases, the soluble expression of protein is favoured by lower culture temperatures (Schein and Noteborn, 1988; San-Miguel, Pérez-Bermúdez and Gavidia, 2013). Thus, one of the strategies for expressing recalcitrant proteins in bacteria is to lower the culture temperature after induction. As it is currently used, in λ cI857, 186 CII^{G22R} and hybridts systems, induction of target gene expression requires high induction temperatures. A temperature sensitive activator protein, which loses activity at higher temperatures, could be used to induce expression of genes at lower temperatures. The minimal basal activity of 186 p_E in the absence of functional CII in conjunction with the potent activation provided by 186 CII makes a temperature sensitive 186 CII/p_E system an attractive prospect.

7.3. Random microseed matrix screening method to derivatize protein- crystals

X-ray crystallography was used in Chapter 2, 3 and 4 to structurally characterize proteins and their complexes. During these experiments, two commonly encountered technical challenges were identified. These are the generation of diffraction quality crystals and solving the phase problem.

Firstly, identifying compatible crystallization conditions for a protein is a trial and error process that requires laborious screening. Secondly, for a protein without a suitable model for molecular replacement, which often is the case when the closest structural homologue available shares less than 35% sequence identity to the target protein (Abergel, 2013), experimental phasing must be used instead. Experimental phasing requires derivatizing a protein crystal with heavy atoms. Finding a compatible heavy atom to incorporate into the crystal and suitable derivatization conditions adds another time-consuming screening step to the crystallization project.

In Chapter 6, I present a new approach to simultaneously optimizing crystal growth and derivatizing crystals with a heavy atom phasing molecule, tackling both crystallography bottlenecks in one step. In this study, two proteins were derivatized at the screening stage

using co-crystallization with the heavy atom ligand I3C. Random matrix microseeding screening (rMMS) was used to increase the number of hits in the presence of the phasing ligand, where many of the resulting crystals showed improved size and morphology. For the two tested proteins, a 2.3 and 7-fold increase in the number of hits were observed when I3C and microseed was introduced into the screen, relative to the unseeded screen. Most importantly, a crystal from each of these screens was shown to be derivatized with I3C and was used successfully in experimental phasing to solve the structure of the protein.

Often in a structure determination project with a novel protein, it is known in advance that derivatization of the crystal is required as molecular replacement is likely to fail. In these cases, derivatizing the protein crystal while screening for crystallization conditions could streamline the process to obtain a structure by reducing the number of screening steps. The significance of this work lies in that it provides a general and straightforward screening method for crystals that simultaneously derivatizes them, with initial empirical evidence demonstrating that it gives a higher number of crystal hits and derivatized crystals contain sufficient phasing power to solve the structures. This technique can also be applied to target proteins with a potential molecular replacement template to reduce model bias from molecular replacement, and as a backup phasing option if molecular replacement fails.

One shortcoming of this study was that it was only tested on two relatively small proteins with molecular masses 14.3 kDa and 22.8 kDa. Future work could verify that the I3C-rMMS method functions with other proteins. There may also exist an upper size limitation on this technique, which should be explored. As I3C derivatization is applied to larger proteins and protein complexes, the limited number of I3C sites on the protein surface may not provide enough anomalous signal to phase the protein. One potential solution is to test rMMS with other phasing ligands. Ligands that bind to more sites on the average protein surface could increase the available phasing power. However, these ligands also tend to be less specific in binding, leading to low occupancy binding sites, increasing the anomalous noise in the data. Thus, there is a balance of promiscuity and binding occupancy that needs to be maintained. Ligands that have increased phasing power per molecule could also be tested, such as the tantalum bromide cluster and B3C. The tantalum bromide cluster contains 6 tantalum metal atoms and 12 bromide atoms, providing an extremely large phasing power per molecule (Knäblein *et al.*, 1997). The B3C phasing ligand has the same structure as I3C, except with the iodine atoms replaced with bromine atoms (Beck, Gruene and Sheldrick, 2010). B3C can provide an increased anomalous phasing signal, as it can be used with the multiwavelength anomalous dispersion (MAD) phasing technique. MAD provides additional phasing information when compared to the single wavelength anomalous dispersion (SAD) phasing technique which I3C is limited to (Rice, Earnest and Brunger, 2000). Tobias Beck, the pioneer of the I3C and B3C phasing molecules, in his doctoral dissertation also proposed three other phasing molecules that can be tested with rMMS (Beck, 2010). Potentially, a cocktail of phasing ligands could be tested. Multiple ligands would increase the collective number of potential binding sites on the protein without resorting to low specificity ligands that adds undesirable noise.

7.4. References

Abergel, C. (2013) 'Molecular replacement: Tricks and treats', *Acta Crystallographica Section D: Biological Crystallography*, 69(11), pp. 2167–2173. doi: 10.1107/S0907444913015291.

- Beck, T. (2010) *Sticky triangles : New tools for experimental phasing of biological macromolecules*.
- Beck, T., Gruene, T. and Sheldrick, G. M. (2010) 'The magic triangle goes MAD: Experimental phasing with a bromine derivative', *Acta Crystallographica Section D: Biological Crystallography*, 66(4), pp. 374–380. doi: 10.1107/S0907444909051609.
- Bonnet, J., Subsoontorn, P. and Endy, D. (2012) 'Rewritable digital data storage in live cells via engineered control of recombination directionality', *Proceedings of the National Academy of Sciences of the United States of America*, 109(23), pp. 8884–8889. doi: 10.1073/pnas.1202344109.
- Breitling, R., Sorokin, A. V. and Behnke, D. (1990) 'Temperature-inducible gene expression in *Bacillus subtilis* mediated by the c1857-encoded repressor of bacteriophage lambda', *Gene*, 93(1), pp. 35–40. doi: 10.1016/0378-1119(90)90132-B.
- Dale, G. E., Oefner, C. and D'Arcy, A. (2003) 'The protein as a variable in protein crystallization', *Journal of Structural Biology*, 142(1), pp. 88–97. doi: 10.1016/S1047-8477(03)00041-8.
- Dodd, I. B. and Egan, J. B. (1996) 'DNA Binding by the Coliphage 186 Repressor Protein CI', *Journal of Biological Chemistry*, 271(19), pp. 11532–11540. doi: 10.1074/jbc.271.19.11532.
- Dodd, I. B. and Egan, J. B. (2002) 'Action at a distance in CI repressor regulation of the bacteriophage 186 genetic switch', *Molecular Microbiology*, 45(3), pp. 697–710.
- Elowitz, M. B. and Leibler, S. (2000) 'A synthetic oscillatory network of transcriptional regulators', *Nature*, 403(6767), pp. 335–338. doi: 10.1038/35002125.
- Friedland, A. E. *et al.* (2009) 'Synthetic gene networks that count', *Science*, 324(5931), pp. 1199–1202. doi: 10.1126/science.1172005.
- Fu, L. X. *et al.* (2021) 'Controlled expression of lysis gene E by a mutant of the promoter pL of the thermo-inducible λ c1857-pL system', *Journal of Applied Microbiology*, 130(6), pp. 2008–2017. doi: 10.1111/jam.14690.
- Herzik, M. A., Wu, M. and Lander, G. C. (2019) 'High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM', *Nature Communications*, 10(1), p. 1032. doi: 10.1038/s41467-019-08991-8.
- Jawale, C. V. *et al.* (2012) 'Characterization of a novel inactivated *Salmonella enterica* serovar enteritidis vaccine candidate generated using a modified c1857/ λ PR/gene E expression system', *Infection and Immunity*, 80(4), pp. 1502–1509. doi: 10.1128/IAI.06264-11.
- Jechlinger, W. *et al.* (1999) 'Altered temperature induction sensitivity of the lambda p(R)/c1857 system for controlled gene E expression in *Escherichia coli*', *FEMS Microbiology Letters*, 173(2), pp. 347–352. doi: 10.1016/S0378-1097(99)00094-4.
- Jechlinger, W. *et al.* (2005) 'Modulation of gene expression by promoter mutants of the λ c1857/pRM/pR system', *Journal of Biotechnology*, 116(1), pp. 11–20. doi: 10.1016/j.jbiotec.2004.10.002.
- Kim, J., White, K. S. and Winfree, E. (2006) 'Construction of an in vitro bistable circuit from synthetic transcriptional switches', *Molecular Systems Biology*, 2. doi: 10.1038/msb4100099.
- Knäblein, J. *et al.* (1997) 'Ta6Br2+12, a tool for phase determination of large biological

- assemblies by X-ray crystallography', *Journal of Molecular Biology*, 270(1), pp. 1–7. doi: 10.1006/jmbi.1997.1074.
- Lieb, M. (1966) 'Studies of heat-inducible lambda bacteriophage: I. Order of genetic sites and properties of mutant prophages', *Journal of Molecular Biology*, 17(1), p. 309. doi: 10.1016/S0022-2836(66)80115-8.
- Lundqvist, B. and Bertani, G. (1984) 'Immunity repressor of bacteriophage P2. Identification and DNA-binding activity', *Journal of Molecular Biology*, 178(3), pp. 629–651. doi: 10.1016/0022-2836(84)90242-0.
- Menart, V. *et al.* (2003) 'Constitutive versus thermoinducible expression of heterologous proteins in Escherichia coli based on strong PR, PL promoters from phage lambda', *Biotechnology and Bioengineering*, 83(2), pp. 181–190. doi: 10.1002/bit.10660.
- Merk, A. *et al.* (2016) 'Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery', *Cell*, 165(7), pp. 1698–1707. doi: 10.1016/j.cell.2016.05.040.
- Murakami, K. *et al.* (1997) 'Positioning of two alpha subunit carboxy-terminal domains of RNA polymerase at promoters by two transcription factors', *Proceedings of the National Academy of Sciences of the United States of America*, 94(21), pp. 11274–11278. doi: 10.1073/pnas.94.21.11274.
- Murchland, I. *et al.* (2014) 'Promoter activation by CII, a potent transcriptional activator from bacteriophage 186', *Journal of Biological Chemistry*, 289(46), pp. 32094–32108.
- Murchland, I. M. *et al.* (2021) 'Instability of CII is needed for efficient switching between lytic and lysogenic development in bacteriophage 186', *Nucleic Acids Research*, 48(21), pp. 12030–12041. doi: 10.1093/nar/gkaa1065.
- Nash, H. A. *et al.* (1987) 'Overproduction of Escherichia coli integration host factor, a protein with nonidentical subunits.', *Journal of bacteriology*, 169(9), pp. 4124–4127. doi: 10.1128/jb.169.9.4124-4127.1987.
- Neufing, P. J. *et al.* (1996) 'The CII protein of bacteriophage 186 establishes lysogeny by activating a promoter upstream of the lysogenic promoter', *Molecular microbiology*, 21(4), pp. 751–761.
- Noireaux, V., Bar-Ziv, R. and Libchaber, A. (2003) 'Principles of cell-free genetic circuit assembly', *Proceedings of the National Academy of Sciences of the United States of America*, 100(22), pp. 12672–12677. doi: 10.1073/pnas.2135496100.
- Pedersen, M., Ligowska, M. and Hammer, K. (2010) 'Characterization of the CI repressor protein encoded by the temperate lactococcal phage TP901-1', *Journal of Bacteriology*, 192(8), pp. 2102–2110. doi: 10.1128/JB.01387-09.
- Pinkett, H. W. *et al.* (2006) 'The structural basis of cooperative regulation at an alternate genetic switch', *Molecular cell*, 21(5), pp. 605–615.
- Ptashne, M. *et al.* (1980) 'How the λ repressor and cro work', *Cell*, 19(1), pp. 1–11. doi: 10.1016/0092-8674(80)90383-9.
- Ptashne, M. (2004) *A genetic switch 3rd edition*, CSHL Press. doi: 10.1038/nrc1424.
- Rice, L. M., Earnest, T. N. and Brunger, A. T. (2000) 'Single-wavelength anomalous diffraction phasing revisited', *Acta Crystallographica Section D: Biological Crystallography*, 56(11), pp. 1413–1420. doi: 10.1107/S0907444900010039.

- Rohlfhill, J., Sandoval, N. R. and Papoutsakis, E. T. (2017) 'Sort-Seq Approach to Engineering a Formaldehyde-Inducible Promoter for Dynamically Regulated Escherichia coli Growth on Methanol', *ACS Synthetic Biology*, 6(8), pp. 1584–1595. doi: 10.1021/acssynbio.7b00114.
- San-Miguel, T., Pérez-Bermúdez, P. and Gavidia, I. (2013) 'Production of soluble eukaryotic recombinant proteins in E. coli is favoured in early log-phase cultures induced at low temperature', *SpringerPlus*, 2(1), pp. 1–4. doi: 10.1186/2193-1801-2-89.
- Schein, C. H. and Noteborn, M. H. M. (1988) 'Formation of soluble recombinant proteins in escherichia coli is favored by lower growth temperature', *Bio/Technology*, 6(3), pp. 291–294. doi: 10.1038/nbt0388-291.
- Shearwin, K. E. and Egan, J. B. (2000) 'Establishment of Lysogeny in Bacteriophage 186 DNA BINDING AND TRANSCRIPTIONAL ACTIVATION BY THE CII PROTEIN', *Journal of Biological Chemistry*, 275(37), pp. 29113–29122.
- Spahn, C. M. and Penczek, P. A. (2009) 'Exploring conformational modes of macromolecular assemblies by multiparticle cryo-EM', *Current Opinion in Structural Biology*, 19(5), pp. 623–631. doi: 10.1016/j.sbi.2009.08.001.
- Villaverde, A. *et al.* (1993) 'Fine regulation of cl857-controlled gene expression in continuous culture of recombinant Escherichia coli by temperature', *Applied and Environmental Microbiology*, 59(10), pp. 3485–3487. doi: 10.1128/aem.59.10.3485-3487.1993.
- Wang, H. *et al.* (2013) 'Single molecule analysis of DNA wrapping and looping by a circular 14mer wheel of the bacteriophage 186 CI repressor', *Nucleic Acids Research*, 41(11), pp. 5746–5756.
- Winstanley, C. *et al.* (1989) 'Differential regulation of lambda pL and pR promoters by a cl repressor in a broad-host-range thermoregulated plasmid marker system.', *Applied and environmental microbiology*, 55(4), pp. 771–777. doi: 10.1128/aem.55.4.771-777.1989.
- Witte, A. and Lubitz, W. (1989) 'Biochemical characterization of ϕ X174-protein-E-mediated lysis of Escherichia coli', *European Journal of Biochemistry*, 180(2), pp. 393–398. doi: 10.1111/j.1432-1033.1989.tb14661.x.

APPENDIX A: REVIEW ON LYSOGENY

The following chapter is an encyclopedia entry on lysogeny co-authored by my principal supervisor, Dr. Keith Shearwin, and I. It was published in the *Reference Module in Life Sciences*. I contributed to the publication by reviewing the literature, preparing figures and writing the manuscript.

Statement of Authorship

Title of Paper	Lysogeny
Publication Status	<input type="checkbox"/> Published <input checked="" type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	This publication is an encyclopedia entry that reviews lysogeny in bacteriophage.

Principal Author

Name of Principal Author (Candidate)	Jia Quyen Truong		
Contribution to the Paper	Reviewed the literature and contributed to the manuscript		
Overall percentage (%)	50%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	16/12/2020

Name of Principal Author	Keith Shearwin		
Contribution to the Paper	Reviewed the literature and contributed to the manuscript		
Overall percentage (%)	50%		
Signature		Date	16/12/2020

Lysogeny

Keith E Shearwin and Jia Q Truong, The University of Adelaide, Adelaide, SA, Australia

© 2019 Elsevier Inc. All rights reserved.

Glossary

Lysogen A bacterial cell carrying one or more bacteriophage genomes, either integrated into the host chromosome or existing as independently replicating extra-chromosomal elements. A true lysogen is able to exit lysogeny and re-enter lytic development.

Lysogenic conversion The situation where a bacterial host acquires a new trait as a direct result of the expression of a gene or genes encoded by a prophage.

Prophage The lysogenic form of a bacteriophage.

Prophage induction The process of a bacteriophage exiting lysogeny and entering lytic development, resulting in lysis of the host cell.

Temperate bacteriophage A bacteriophage capable of entering either lytic development or lysogeny upon host cell infection.

Terminally redundant DNA DNA that contains repeated sequences at each end called terminal repeats. These ends are used to join the ends of the linear DNA to form circular DNA.

Transcriptional interference The suppressive influence of one transcriptional process, directly and *in cis* on a second transcriptional process.

Introduction

Bacteriophages (phages) are obligate bacterial parasites. Many phages exhibit a purely lytic lifecycle, where following infection of a susceptible host, the phage DNA is replicated and the phage hijacks the bacterium's cellular machinery to produce new virion particles, which are released upon lysis of the host cell. Other phages, the so-called temperate phages, are able to make a developmental 'decision' between two developmental regimes, the lytic and lysogenic cycles. In lysogeny, the phage persists indefinitely inside the bacterial host as a prophage, with the phage genome either integrated into the bacterial chromosome or existing extra-chromosomally (Fig. 1). Lysogeny is a non-bacteriocidal state where the bacteriophage genome replicates without virion production. In this state, most of the phage genome is transcriptionally inactive, as the expression of lytic functions would be lethal to the host cell. A bacterial host cell lysogenic for a particular phage is also immune to further infection by the same phage.

Filamentous phage, which have ssDNA genomes packaged into filament-like virions, are able to replicate without killing the host. Among the filamentous phage, some integrate in the host chromosome, while non-integrative filamentous phage replicate exclusively as extrachromosomal elements or episomes. Both classes of filamentous phages continually shed viral particles without host cell death, even while inserted into the bacterial genome as a prophage. Thus, bacteria infected permanently with filamentous phage represent a form of lysogeny, but don't meet the definition of true temperate phage, in that there is no stage which brings about host cell lysis.

In this article, we discuss specific examples of how true temperate phage establish and maintain lysogeny. The phenotypic effects of lysogeny on the host bacterium and the evolutionary impacts of lysogeny are also discussed.

Why Lysogeny?

For a phage to successfully propagate, it must coax its host into manufacturing new virions. These new virions are released through host cell lysis, and need to infect another host bacterium in order for the phage to continue to propagate. From the phage perspective, during times of abundant resources and the presence of many host cells, lytic development allows for large numbers of phages to be produced, to subsequently infect other host cells and thus continue the cycle. However, when resources and/or susceptible host cells are scarce, the phage can benefit from entering lysogeny until conditions improve. The process of making a choice between the lytic and lysogenic pathways following host cell infection, is covered in more detail in another article (Golding). While there must be some (small) fitness cost in replicating extra genomic material, many phages have evolved such that lysogeny can provide a selective advantage to the host cell in a number of ways. Such benefits may include carrying genes which confer some growth advantage to the host, or by providing protection from subsequent infection, and potential lysis, by phage from the same or different families.

Persistence of DNA

Integration Into the Chromosome

The long term persistence of phage DNA inside their bacterial host cell is the key feature of a lysogen. Phage have evolved many mechanisms to achieve this long term persistence. In many cases, such as the paradigm bacteriophage λ , the phage DNA is

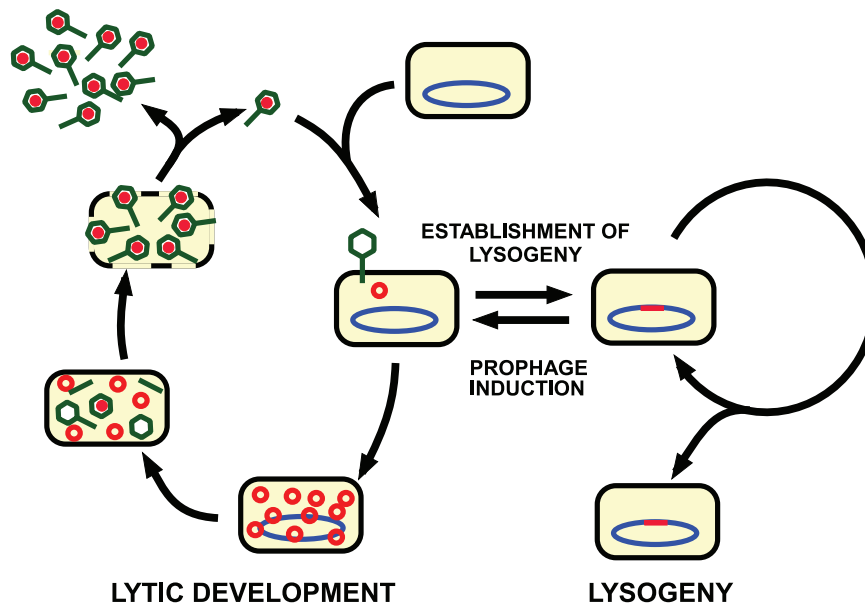


Fig. 1 Lytic and lysogenic developmental cycles of a temperate bacteriophage. A bacteriophage infects a bacterial host by injecting its DNA (red) into the host cell. In the lytic lifecycle, phages produce new virions, lysing the cell to release them. The lysogenic lifecycle has the phage DNA integrated into the bacterial chromosome or persisting extra-chromosomally. This state is stable, with prophage DNA maintained in daughter cells over subsequent generations. Most temperate phage can efficiently exit lysogeny to reenter the lytic phase in response to specific environmental cues.

integrated into the bacterial chromosome via site-specific recombination (**Fig. 2**). The phage-encoded integrase protein is a site-specific recombinase that recognizes attachment (*att*) sites on both the phage genome (*attP*) and the bacterial chromosome (*attB*). The integrase catalyses recombination between the *attP* and *attB* sites. This process generates two new *att* sites, *attL* and *attR* that flank the phage genome. When the phage exits lysogeny, the excision machinery, consisting of at least the integrase and a phage-encoded recombination directionality factor (RDF), and often with additional host factors, recognizes the *attL* and *attR* sequences to catalyse the reverse reaction to excise the prophage genome.

In general, integrase proteins belong to one of two families, the serine integrases or the tyrosine integrases. These proteins use a catalytic serine or tyrosine, respectively, to carry out the recombination reaction. Serine integrases create a double stranded break in both the phage and bacterial chromosome, and mediate strand rearrangement to bring them into the recombinant configuration, followed by strand ligation. Tyrosine recombinases, on the other hand, mediate recombination via a Holliday junction intermediate. Single stranded cuts are made in DNA and rejoined to single stranded DNA from the recombinant partner. Both types of integrases have been used extensively for biotechnology and molecular biology applications.

Some temperate phage integrate their DNA less specifically into the bacterial genome. Upon infection of a host cell, phage Mu initially integrates at random into the host chromosome. Only after integration is a decision on lytic or lysogenic development made. If the lytic pathway is followed, the phage genome undergoes successive rounds of replicative transposition. In each round, Mu DNA is duplicated, and a range of host genome rearrangements (deletions, inversions, translocations) are introduced. Viral genomes, along with variably sized segments of the host genome, are packaged directly from the integrated copies dispersed around the scrambled host genome. Similar to other temperate phages, the Mu lysogenic pathway (termed the latent pathway in Mu) is followed if the Repc repressor protein accumulates to a sufficiently high concentration. Repc competes for overlapping operators with DDE-recombinase A, which binds to an internal activation sequence (IAS), to initiate the lytic pathway.

In Mu, the initial integration process occurs by random transposition (**Fig. 3**). This process utilizes transposition machinery consisting of phage-encoded and host-supplied components to cleave the Mu DNA and insert it into target sites in the host genome with little or no target specificity. When Mu injects its DNA into a new host cell, its DNA is linear and flanked by DNA that was acquired from the previous host from which it was assembled. The phage supplied N protein, injected alongside DNA, non-covalently circularizes the DNA. To integrate Mu DNA into the chromosome, the phage DNA is nicked by MuA transposase to generate two 3' hydroxyl (-OH) ends on both strands at the Mu-Host DNA junctions. The 3'-OH ends attack two phosphodiester bonds spaced 5 bps apart, which join the 3'-OH to the 5' phosphate of the host DNA at the attack site. The flanking segments of DNA from the previous host are removed and the DNA gap is repaired.

A natural consequence of phage DNA integration into the host chromosome is that host DNA replication will also replicate the prophage DNA. Thus, all daughter cells of a lysogen will inherit the prophage. This form of lysogeny can be extremely stable, with very low rates of 'spontaneous' induction.

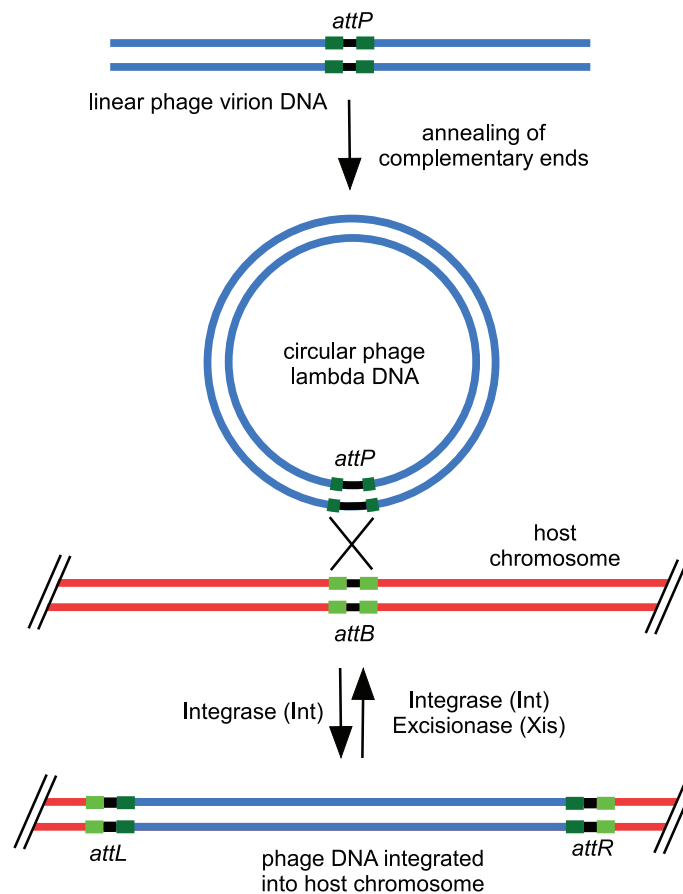


Fig. 2 Prophage DNA can be maintained by site-specific integration into the bacterial chromosome. This reaction is mediated by a phage-encoded integrase protein (Int). Initially upon injection into the host, the phage DNA is linear. Circularization of the viral DNA occurs through annealing of complementary ends, followed by ligation. Integrases recognize the attachment sites *attP* on the viral DNA (dark green) and *attB* on the host chromosome (light green) to catalyse the site-specific recombination. This generates two new *att* sites, *attL* and *attR* which flank the viral DNA in the chromosome. The attachment sites have a conserved core, shown in black. To return to the lytic cycle, the viral DNA must be excised from the chromosome through a reversal of the site-specific recombination between *attL* and *attR*, mediated by the excision machinery, which include the Int and excisionase (or recombination directionality factor RDF) proteins.

Extrachromosomal Persistence of DNA

Some prophages can also maintain their DNA separate from the host chromosome as a plasmid, either circular or linear. A well-characterized example of a circular extrachromosomal lysogen is bacteriophage P1. Infectious particles of P1 consist of a large (93 kb) linear double stranded genome with terminal redundancy of 10–15 kb. Upon being injected into the P1 host *E. coli*, the ends of the DNA undergo homologous recombination to circularize into an autonomous plasmid. In lysogeny, the prophage/plasmid copy number is maintained to be the same as the number of bacterial chromosomes. This is achieved through the expression of P1 proteins involved in origin-specific initiation and partitioning. The phage also uses a toxin-antitoxin system to kill daughter cells that have failed to receive a P1 molecule upon cell division. The antitoxin is degraded rapidly and needs to be continually expressed to counteract the toxin's activity. Loss of the P1 molecule in daughter cells results in a rapid depletion of antitoxin to a level where the more stable toxin will kill the host cell.

Some prophages, such as *E. coli* bacteriophage N15 maintain their DNA as a linear plasmid. Similar to linear eukaryotic chromosomes, this presents a problem as DNA polymerase fails to replicate all the way to the ends of the DNA. Bacteriophage N15 belongs to a group of phages that express a protelomerase to create covalently closed hairpins. After entering the cell, N15 DNA circularizes through its cohesive termini. The phage encoded protelomerase cuts an inverted sequence in the phage genome and ligates the phosphodiester bonds of each strand to form two closed hairpin ends (Fig. 4).

The replication of such as linear plasmid also presents mechanistic challenges and requires prophage encoded proteins. In the case of bacteriophage N15, the phage *repA* protein is necessary for replication. This multi-domain protein contains domains that resemble prokaryotic primases and helicases required for DNA replication.

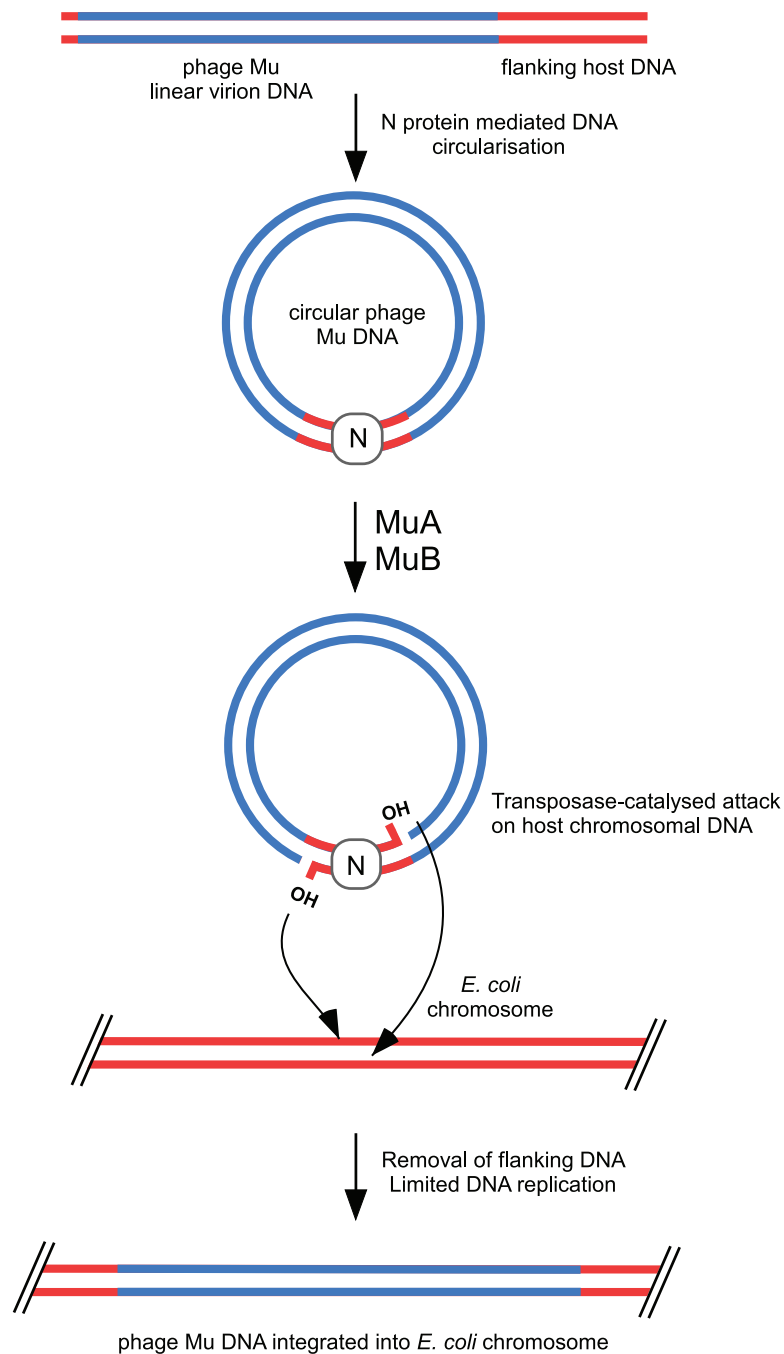


Fig. 3 Bacteriophage Mu integrates into the *E. coli* chromosome via random transposition. Linear Mu DNA is non-covalently circularized via the N protein. The Mu B transposase generates single stranded nicks at the junctions of the Mu (blue)-prior host DNA (red) sequence, revealing two 3'-OH ends. These ends attack a target site on the *E. coli* chromosome to join the 3'-OH of Mu DNA to the 5' phosphate of the host DNA. Removal of the prior host DNA and subsequent repair of the DNA gap at the host-Mu DNA junction via limited replication results in successful integration of Mu into the chromosome.

Maintenance of the Lysogenic State

Temperate phages often contain natural examples of genetic switches which enable the developmental choice between the lytic or lysogenic lifecycles. The biochemical basis of this decision-making process will be covered in detail in a separate chapter of the encyclopedia. After entering lysogeny, true lysogens are able to exit back into the lytic cycle, either spontaneously or in response to

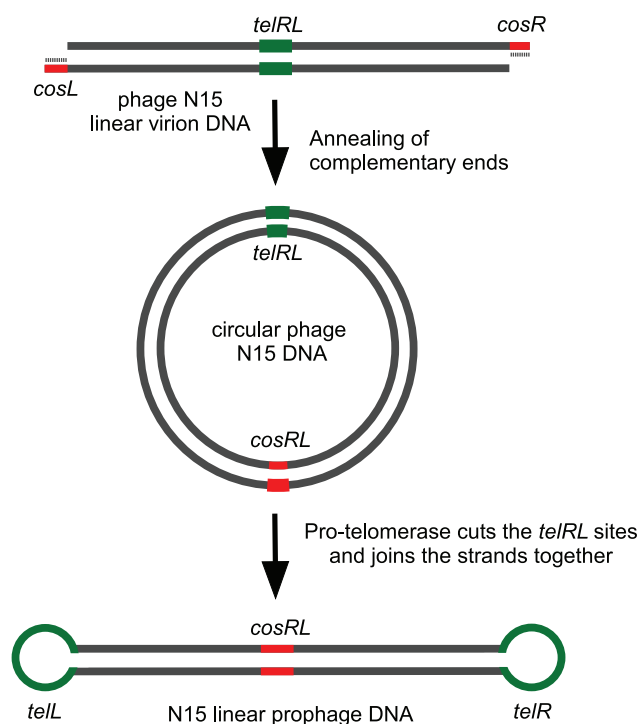


Fig. 4 A prophage of bacteriophage N15 maintains its DNA as a linear plasmid with closed hairpin loops. Upon injecting its DNA into *E. coli*, the phage initially circularizes through the annealing of complementary ends. A phage encoded protelomerase binds and cleaves an inverted repeat sequence (*telRL*) within the genome. The phosphodiester backbone of single strands are then ligated together to form two hairpin loops, creating one covalently closed DNA molecule. Linear plasmid replication requires a phage encoded protein, *repA*.

environmental cues. One strategy to generate such a bistable decision making circuit is through mutually repressing transcription factors. The most well characterized bacteriophage employing this strategy is bacteriophage λ .

Bacteriophage λ

The bistable regulatory circuit in λ is generated by the *CI* and *Cro* repressors, which block each other's transcription. To maintain lysogeny, the promoters that drive lytic functions must be repressed, whilst maintaining production of the *CI* repressor. In general, the repressor in temperate phages responsible for turning off lytic function is termed the immunity repressor, as it prevents secondary infections by phage of the same family, by blocking the lytic functions of incoming phage.

λ has a lysogenic promoter, P_{RM} , positioned back to back with the adjacent lytic promoter P_R , with a second lytic promoter, P_L located 2.3 kb away (Fig. 5). The transcriptional cascade in λ is such that repression of the early lytic promoters P_L and P_R will stop all of the lytic genes from being expressed, pushing the phage into lysogeny. This repression is achieved by the λ *CI* protein, which binds to operators at *OR* and *OL* operators that overlap the P_R and P_L promoters, respectively. In the lysogenic state, *CI* simultaneously represses transcription from P_L and P_R , while activating its own expression from P_{RM} .

The lambda *CI* protein is a two domain protein consisting of an N-terminal DNA binding domain connected by a cleavable linker to an oligomerization domain. The repressor homodimerizes and these dimers can bind to individual operators. However, repressor binding is cooperative such that pairs of dimers bound to adjacent operators (primarily *OR*1-*OR*2 and *OL*1-*OL*2) interact to improve overall binding affinity. In this state, P_R and P_L are repressed, whilst P_{RM} is activated by a contact between the *CI* dimer bound at *OR*2 and the σ subunit of RNA polymerase. In this manner, λ *CI* produces positive feedback to increase its own expression. A further level of cooperativity also occurs, when a pair of dimers occupying two operators at *OR* interact with a pair of dimers bound at *OL* to form a *CI* octamer, looping out the intervening ~ 2.3 kb of DNA (Fig. 5). Octamer formation in turn positions the *OL*3 and *OR*3 operators to allow λ *CI* dimers bound at each of these sites to interact and repress P_{RM} at high levels of λ *CI*. This negative autoregulation is thought to limit the repressor levels to allow for efficient prophage induction back into lytic cycle (Fig. 5(D)).

Cro is the second transcriptional regulator within the lambda genetic switch. The *cro* gene is the first of the early genes expressed from P_R following phage infection, and is required to enforce prophage induction – the decision point when the phage switches from lysogeny towards lysis. Upon activation of the host SOS response, the *CI* repressor is degraded by a mechanism involving activated *RecA* improving an intrinsic but normally weak *CI* self-cleavage activity. Thus the lytic repressors are at least partially de-repressed and *Cro* is made. *Cro* binds as a dimer to the same *OL* and *OR* operators as *CI*, but with different affinity, binding most

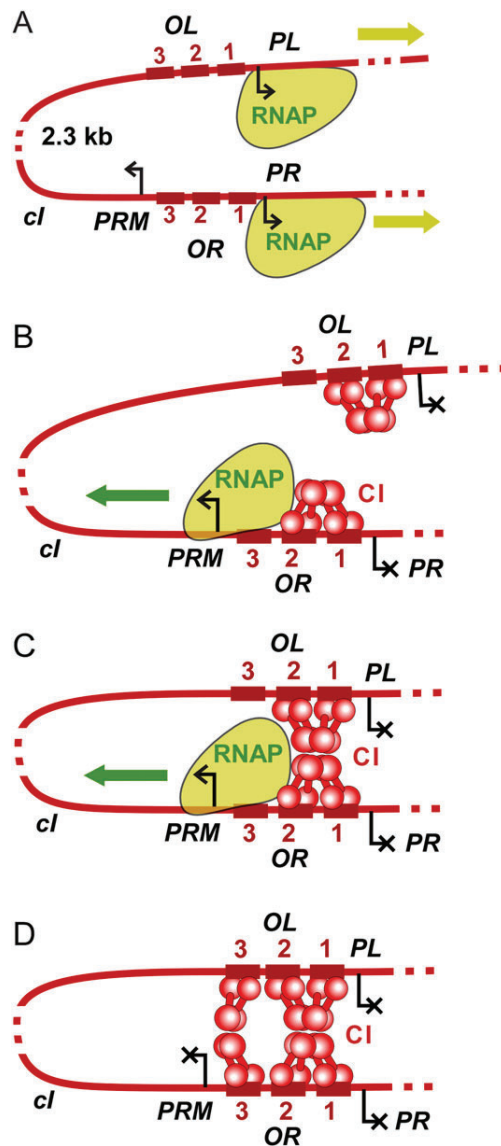


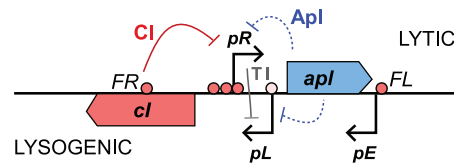
Fig. 5 Regulation of the λ lytic and lysogenic promoters by λ CI involves cooperative interactions at multiple levels. A. The arrangement of early lytic and lysogenic promoters in bacteriophage λ . The regulation of these promoters is crucial in the lytic-lysogeny decision. Promoters are depicted as bent arrows. RNA transcripts are depicted as arrows. Early lytic transcripts from P_L and P_R are shown in yellow. OL and OR operators are depicted as red bars on the DNA. B. λ CI (depicted as red dumbbells) binds to operators at OL and OR. λ CI dimers at adjacent operators can further interact to form tetramers, mediated through the CI C-terminal domain. Binding of λ CI blocks transcription from P_R and P_L , by blocking RNA polymerase (RNAP) binding. Contacts between the N-terminal domain of CI at O_{R2} and the sigma subunit of RNAP activates the P_{RM} promoter to drive expression of λ CI and other lysogenic functions. The lysogenic transcript from P_{RM} is depicted in green. C. The repression of P_R and P_L is further enhanced by interactions between CI tetramers at OR and OL, assisted by DNA looping. D. Formation of the CI octamer allows CI dimers bound at OL3 and OR3 interact cooperatively, to repress P_{RM} and downregulate CI's own expression.

strongly to OR3 at the OR operator. This preferential and non-cooperative binding to OR3 dampens CI expression from the P_{RM} promoter, while leaving the P_L and P_R promoters active. Thus, Cro repression of P_{RM} is required to prevent re-synthesis of new CI during prophage induction, which would otherwise re-establish the prophage state.

Bacteriophage 186

Bacteriophage 186 is a temperate bacteriophage of the P2 related family, a group of phages evolutionarily distinct from the lambdoid phages. 186 and lambda infect the same host (*E. coli*) and have very similar temperate lifecycles, yet the mechanisms by which these two phages enter and maintain lysogeny are quite different. In 186, early lytic pR and lysogenic promoters pL are in a

A. Bacteriophage 186 switch region



B. DNA wrapping and looping by 186 CI

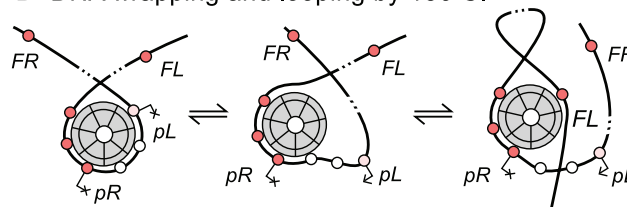


Fig. 6 Relief of transcriptional interference is used to activate transcription of lysogenic promoters in bacteriophage 186. A. The early lytic promoter pR and the lysogenic promoter pL are arranged face-to-face. CI is the 186 immunity repressor. Apl acts as both a repressor of the pR and pL promoters, and also as the excisionase for phage 186. The 186 CI repressor binding sites are drawn as circles – there are three strong sites (red) over the pR promoter, a weak operator located over pL (pink) and two distant sites (FL and FR), located ~300 bp from pR. B. Model of transcriptional activation of 186 pL by 186 CI. CI is shown as a wheel-like structure where DNA wraps around the circumference of the wheel. CI operators are shown as circles. The weak operator at pL is sometimes wrapped to the wheel (left) thus repressing pL, in an equilibrium with a form (middle) where the pL operator is not bound, allowing pL to be active. The flanking sites FL and FR can compete with pL for binding to the CI wheel (right), fine tuning the response of the system to CI levels.

convergent arrangement (Fig. 6), rather than divergent as is the case for the λ P_R and P_{RM} promoters. A consequence of this convergent layout is that RNA polymerases either bound at, or elongating from, the switch promoters can potentially collide with each other. In the absence of the immunity repressor (CI), transcription from the stronger lytic pR promoter represses transcription from the weaker lysogenic promoter pL by this mechanism of transcriptional interference (TI), favouring the lytic lifecycle. Conversely, in the presence of CI, pR is efficiently repressed, relieving transcriptional interference, and providing a form of positive feedback on CI expression.

The 186 CI protein is the immunity repressor responsible for maintenance of lysogeny. Similarly to λ CI, it contains an N-terminal DNA binding domain and a C-terminal oligomerization domain and binds to its operators as a dimer. Crystal structures suggest 186 CI forms a wheel consisting of a heptamer of dimers. The repressor's DNA binding domains face outwards from this wheel, allowing DNA to wrap around the protein (Fig. 6(B)) to regulate transcription of pR and pL. The CI wheel binds cooperatively to the three strong operators to repress pR and relieve transcriptional interference on pL. Binding of the weaker CI operator at pL to the CI wheel is in an equilibrium between wrapped (repressed) and unwrapped (active) forms. There is further fine tuning of pL promoter activity by the two flanking CI operator sites (FL and FR, located ~300 bp away) which can loop to the wheel and displace pL from the wheel. When the CI operator at pL is displaced by either of the stronger flanking site operators, pL is transcriptionally active, maintaining CI levels. Interestingly, in phage 186, integrase is also expressed on the lysogenic transcript, such that integrase is continually present in a lysogen. When prophage induction is triggered, excision of the phage genome requires only expression of the excisionase (Apl), the first gene of the lytic transcript.

Thus, phage 186 CI repressor maintains an active lysogenic promoter by relieving transcriptional interference on pL, to drive its own expression and to maintain a stable, self-correcting lysogenic level of CI. The lytic pR promoter is tightly repressed in this state, leading to low levels of spontaneous induction.

Integration-Dependent Bacteriophage Immunity

A class of temperate mycobacteriophage have been described which use site-specific integration as the key decision point of the genetic switch. In these phage, the immunity repressor and integrase are expressed from the same promoter P_{rep} . To the right of the repressor gene, under the control of a separate divergent promoter, P_R , is a Cro-like protein. Both the immunity repressor and Cro-like protein are predicted to have a DNA-binding helix-turn-helix (HTH) motif. Crucially, the bacteriophage attachment site *attP* is located within the open reading frame of the immunity repressor (Fig. 7).

Unlike bacteriophages λ and 186, the promoter driving the immunity repressor function is not regulated by the immunity repressor. Rather, the key step needed to establish lysogeny is the integration of the phage DNA into the host chromosome. Integration of the phage genome through recombination of *attP* and *attB* sites results in a truncated immunity repressor without a short destabilizing C-terminal tag. This truncated repressor supplies sufficient repression activity to establish lysogeny. The tag, thought to target the viral repressor for degradation by host proteases, reduces its effective activity below the level required for lysogeny.

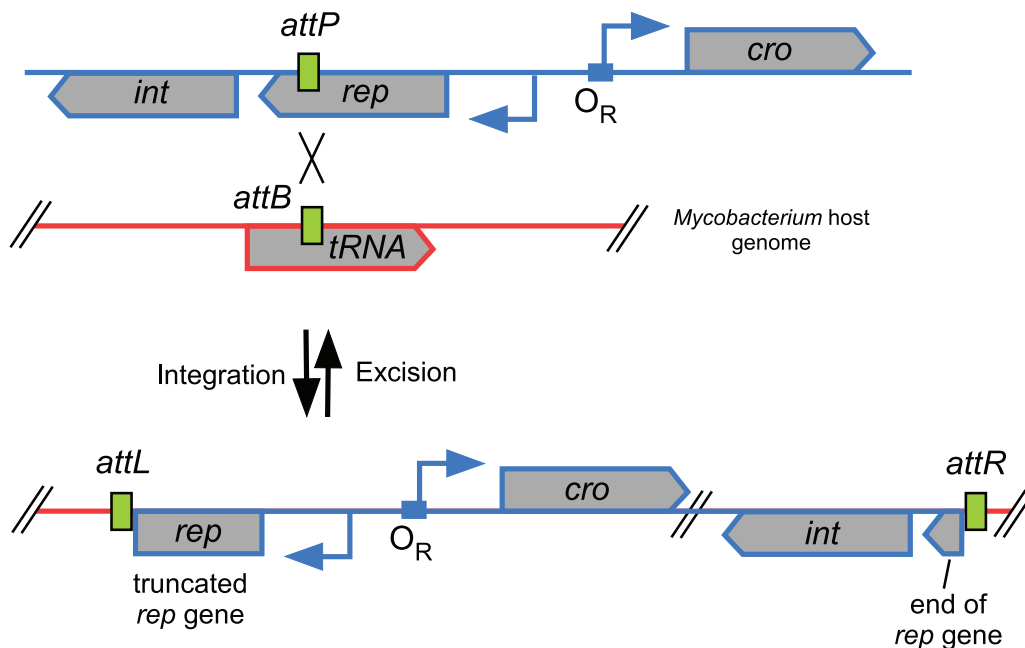


Fig. 7 The organization of integrase (*int*), immunity repressor (*rep*) and *cro*-like protein (*cro*) genes in integration-dependent immunity systems. Promoters are shown as bent arrows. Site-specific attachment sites are shown as green boxes. The *attP* attachment site lies within the *rep* open reading frame. Site specific recombination of *attP* with *attB* results in a truncated repressor gene within the bacterial chromosome. This truncated gene encodes for a more stable immunity repressor, driving the phage towards lysogeny.

The determining factor of the phage's developmental fate is the stability of the integrase protein. If conditions favor high integrase activity, such as high multiplicity of infection (MOI) or when host protease levels are low, integration of the phage genome is favored. Integration subsequently establishes and maintains lysogeny via immunity repressor stabilization.

Use of a Host Protein as an Immunity Repressor

In most phages, a phage-encoded immunity repressor binds to promoters that drive lytic functions to repress them. GIL01, a phage that infects *Bacillus thuringiensis*, is an interesting example of a phage that does not have encode an immunity repressor at all. It instead uses the *B. thuringiensis* host LexA repressor as its immunity repressor.

It has been shown that the DNA binding activity of LexA in *B. thuringiensis* is regulated by GIL01 phage encoded proteins. When complexed with specific phage proteins, LexA represses lytic promoters by binding to a conserved LexA site within the promoter region. If these additional phage proteins are not expressed, LexA binding alone is too weak to repress the lytic promoter and GIL01 is unable to form stable lysogens.

RNA to Maintain Lysogeny

The mechanisms of maintenance of lysogeny that have been described thus far have been based on proteins, responsible for mediating either repression or integration. However, there are also examples of phage which use RNA-based mechanisms to maintain a stable lysogenic state.

Lactobacillus casei phage A2 has a similar genetic arrangement to that of the *O_R* region of lambda, with back-to-back promoters driving expression of *CI* and *Cro* repressors. However, in phage A2, the *CI* and *Cro* repressors have very similar affinities for their operators, leading to the expectation that the much stronger lytic pR promoter should dominate, leading to expression of *Cro*, and favoring the lytic cycle. However, A2 is able to effectively establish and maintain stable lysogens. It has recently been shown that the second gene of the A2 lytic transcript encodes an RNA binding protein (*gp25*) that is able to specifically bind to a region between the ribosome binding site and *cro* start codon on the lytic transcript. RNA binding by *gp25* thus prevents translation of *Cro*, allowing establishment and maintenance of stable lysogens.

P4 is a so-called satellite phage that relies on an unrelated helper phage, such as P2 or 186, to supply the structural genes needed for its propagation. P4 has the ability to redirect the genetic network of the helper phage in order to redirect the capsid assembly process for its own purposes. The P4 replicon can exist as a plasmid or can be stably integrated into the host chromosome as a lysogen. The integrated lysogenic form is maintained by an RNA-based mechanism. In a P4 lysogen, transcription from a constitutive early promoter generates a ~300 nucleotide transcript that is

processed to form the CI RNA. The CI RNA interacts with two specific regions in the untranslated leader sequence of the nascent transcript to mediate transcriptional termination, prevent expression of the P4 replication functions and thus, maintain lysogeny.

Prophage Induction

Established lysogens are generally stable with low probabilities of reverting back to the lytic state. Certain environmental stresses can stimulate the prophage to revert to the lytic state. This process is called prophage induction. The immunity repressor must be degraded or otherwise inactivated during prophage induction in order to establish lytic gene expression.

Many temperate phages are inducible by DNA damage, such as through mitomycin C treatment or UV irradiation. Such DNA damage results in activation of the host SOS response, a global cascade of gene expression centered on DNA repair. As explained previously, the stress response genes are under the control of the host LexA protein. LexA contains a labile linker region that can be cleaved by a C-terminal protease domain within LexA itself. In the SOS response, the activation of co-protease RecA as a result of the accumulation of single stranded DNA, promotes the auto-proteolysis of the host LexA repressor. Phage GIL01, which relies on LexA to directly repress lytic genes, will undergo prophage induction simply due to the removal of LexA. Some phage repressors, such as λ CI repressor, have structural similarity to LexA, and undergo similar RecA-stimulated self-cleavage. Inactivation of CI allows the expression of lytic genes, including the genes for excision and replication.

Other phages, such as bacteriophage 186, express specific anti-repressor proteins, which complex with their cognate repressor to relieve repression of the lytic genes. For example, bacteriophage 186 expresses the anti-repressor Tum, which is in turn under the control of a LexA repressible promoter. Upon activation of the SOS response, RecA mediated proteolysis of LexA results in expression of the Tum anti-repressor, which inactivates 186 CI, to drive prophage induction. The related coliphage P2 does not code for an anti-repressor, and does not induce in response to DNA damage. However, it has a somewhat higher level of spontaneous induction than 186, which must be sufficient for it to sustain a viable number of free phage.

The satellite phage P4, which uses an RNA based transcriptional termination mechanism for maintaining lysogeny, is induced by transcription from an alternative promoter activated by a protein from the helper phage. Translation of nested genes within this longer transcript leads to suppression of the CI RNA mediated termination, and expression of the P4 genes needed for replication.

Evolutionary and Phenotypical Effects of Lysogeny

It is estimated that the number of phage outnumber bacteria in the biosphere by at least an order of magnitude. Thus, it is unsurprising that phages have a large effect on the evolution of their hosts and play a significant role in shaping bacterial communities in almost all ecosystems. Here, we discuss how lysogeny has altered the phenotype and evolution of its hosts.

Lysogenic Conversion

Lysogenic conversion describes the situation where a bacterial host acquires a new trait as a direct result of the expression a gene encoded by a lysogen. Clinically relevant examples include the acquisition of virulence factors by bacterial pathogens. Many of the toxins contributing to virulence in bacteria such as *Staphylococcus aureus*, *Shigella* spp., *Salmonella enterica* and pathogenic *E. coli* have been acquired through lysogenic conversion. *E. coli* O157:H7, a causative agent of food poisoning, harbors Sp5 and Sp15 – two Shiga toxin expressing prophages. *Vibrio cholera*, some strains of which cause Asiatic cholera, acquired the cholera toxin from its CTX ϕ prophage.

Besides toxins, other phage-encoded virulence factors include adhesion factors, superantigens and factors that aid immune system invasion. Various examples of phage-encoded virulence factors are summarized in [Table 1](#).

In many cases, the virulence genes are arranged as a cassette flanked by a σ^{70} promoter on one side and a transcription terminator on the other. Such a configuration reduces interference with the prophage genes adjacent to the cassette. It has been postulated that virulence genes confer some fitness advantage to their host, either offensive (e.g., toxin production, invasins), defensive (e.g., detoxification) or survival (such as improved nutrient uptake).

Very recently, a type of temperate filamentous bacteriophage that infects and integrates into *Pseudomonas aeruginosa* (*Pa*) was found to be associated with chronic human wound infections. In mice, phage-infected *Pseudomonas aeruginosa* led to more severe and longer-lasting wounds compared to wounds colonized by *Pa* alone. The authors showed that uptake of phage-infected *Pseudomonas aeruginosa* by cells of the immune system resulted in phage RNA production, inappropriate antiviral immune responses and suppression of bacterial clearance.

Gene Disruption

Prophage integration into the host chromosome can alter bacterial phenotype by disruption of an open reading frame. The integration of phage may also disrupt sequences with regulatory roles within an intergenic region, altering how downstream genes are regulated. Prominent loci for *attB* phage integration sites are within tRNA open reading frames. The loss of a functional tRNA could potentially lead to a reduction in host fitness, so many phage have evolved such that integration will reconstitute a functional tRNA gene. This holds true for many other *attB* sites within important genes.

Table 1 Non-exhaustive list of phage-encoded virulence factors

<i>Protein</i>	<i>Phage</i>	<i>Bacterial host</i>	<i>Gene</i>
Extracellular toxins			
Diphtheria toxin	β-Phage	<i>C. diphtheriae</i>	<i>tox</i>
Neurotoxin	Phage C1	<i>C. botulinum</i>	<i>C1</i>
Shiga toxins	H-19B	<i>E. coli</i>	<i>stx1, stx2</i>
Enterohaemolysin	φFC3208	<i>E. coli</i>	<i>hly2</i>
Cytotoxin	φCTX	<i>P. aeruginosa</i>	<i>ctx</i>
Enterotoxin	NA	<i>S. aureus</i>	<i>see, sel</i>
Enterotoxin P	φN315	<i>S. aureus</i>	<i>sep</i>
Enterotoxin A	φ13	<i>S. aureus</i>	<i>entA</i>
Enterotoxin A	φMu50A	<i>S. aureus</i>	<i>sea</i>
Exfoliative toxin A	φETA	<i>S. aureus</i>	<i>eta</i>
Toxin type A	T12	<i>S. pyogenes</i>	<i>speA</i>
Toxin type C	GS112	<i>S. pyogenes</i>	<i>speC</i>
Cholera toxin	CTXφ	<i>V. cholerae</i>	<i>ctxAB</i>
Enterotoxin	CTXφ	<i>V. cholerae</i>	<i>ace, zot</i>
Leukocidin	fPVL	<i>S. aureus</i>	<i>pvl</i>
Superantigens	8232.1	<i>S. pyogenes</i>	<i>speA1, speA3, speC, speI, speH, speM, speL, speK, ssa</i>
Cytolethal distending toxin	Unnamed	<i>E. coli</i>	<i>cdt</i>
Proteins altering antigenicity			
Membrane proteins	Pnm1	<i>N. meningitidis</i>	<i>Mu-like</i>
Glucosylation	ε ³⁴	<i>S. enterica</i>	<i>rfb</i>
Glucosylation	P22	<i>S. enterica</i>	<i>gtr</i>
O-antigen acetylase	Sf6	<i>S. flexneri</i>	<i>oac</i>
Glucosyl transferase	SfII, SfV, SfX	<i>S. flexneri</i>	<i>gtrII</i>
Effector proteins involved in invasion			
Type III effector	SopEφ	<i>S. enterica</i>	<i>sopE</i>
Type III effector	GIFSY-2	<i>S. enterica</i>	<i>ssel (gtgB)</i>
Type III effector	GIFSY-3	<i>S. enterica</i>	<i>sspH1</i>
Enzymes			
Superoxide dismutase	Sp4, 10	<i>E. coli O157</i>	<i>sodC</i>
Superoxide dismutase	GIFSY-2	<i>S. enterica</i>	<i>sodC-I</i>
Superoxide dismutase	Fels-1	<i>S. enterica</i>	<i>sodC-III</i>
Neuraminidase	Fels-1	<i>S. enterica</i>	<i>nanH</i>
Hyaluronidase	H4489A	<i>S. pyogenes</i>	<i>hylP</i>
Leukocidin	φPVL	<i>S. aureus</i>	<i>pvl</i>
Staphylokinase	φ13	<i>S. aureus</i>	<i>sak</i>
Phospholipase	315.4	<i>S. pyogenes</i>	<i>sla</i>
DNase/streptodornase	315.6, 8232.5	<i>S. pyogenes</i>	<i>sdn, sda</i>
Serum resistance			
OMP	λ	<i>E. coli</i>	<i>bor</i>
OMP	λ-like	<i>E. coli</i>	<i>eib</i>
Adhesions for bacterial host attachment			
Vir	MAV1	<i>M. arthritidis</i>	<i>vir</i>
Phage coat proteins	SM1	<i>S. mitis</i>	<i>pblA, pblB</i>
Others			
Mitogenic factors	370.1, 370.3, 315.3	<i>S. pyogenes</i>	<i>mf2, mf3, mf4</i>
Mitogenic factor	Unnamed	<i>P. multocida</i>	<i>toxA</i>
Mitogenic factor	phisc 1	<i>S. canis</i>	<i>Unnamed</i>
Virulence	GIFSY-2	<i>S. enterica</i>	<i>gtgE</i>
Antivirulence	GIFSY-2, Fels-1	<i>S. enterica</i>	<i>grvA</i>

Note: Table adapted and extended from Brüssow, H., Canchaya, C., Hardt, W., 2004. Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews* 68 (3), 560–602.

There are cases where prophage integration does not reconstitute a functional gene and leads to loss of protein function. For example, the integration of L54a and ϕ 13 phages into the *S. aureus* chromosome leads to a loss of a functional lipase and β -toxin, respectively.

Genomic Rearrangement

Bacteria harboring multiple prophages with similar DNA sequences can give rise to homologous recombination events to cause large genomic rearrangements. The location of rearrangements in the chromosome are often correlated with the loci of prophages. For example, from a sequence alignment, it was observed that the genomic differences between an American and Japanese *S. pyrogenes* M3 isolate could be largely explained by two sequential inversions. One of these inversions occurred between the lysis modules of two prophages.

Genomic rearrangements may also aid phage evolution. In the example of *S. pyrogenes*, the lysogenic conversion genes were outside of the inversion region. The inversion event shuffled virulence genes between the prophages. This could lead to new prophages with altered host specificity and behaviors.

Conclusion

Lysogeny can change the phenotype, fitness and evolution of the bacterial host cell. In the case of bacterial pathogens, prophages allow the acquisition of new traits such as virulence factors which in turn may improve the host's fitness. This is emphasized by the observation that prophages have played a vital role in the emergence of several clinically relevant human pathogens, such as the food-borne pathogen *E. coli* O157:H7 and cholera causing *V. cholerae*. Prophages also drive bacterial evolution through the introduction of new genes, the disruption of existing genes and mediating large genomic rearrangements within the host.

In the laboratory, temperate bacteriophages have provided numerous tools and reagents for molecular biology, and have been used as models of regulatory networks and biological switches. The study of bacteriophages λ , P2 and its relatives, Mu, P1, N15 amongst others have provided many seminal contributions to these fields. This article has touched on several of these mechanisms and simultaneously highlighted the diversity in means for achieving the persistence of DNA, maintaining the prophage in a quiescent state and for exiting lysogeny. As the number of bacterial genomes sequenced continues to grow, new prophages using novel mechanisms to achieve these lysogenic functions will no doubt be discovered.

Acknowledgments

The Shearwin lab is supported by grants from the Australian Research Council (DP160101450) and the National Health and Medical Research Council (APP1100653).

Further Reading

- Brüssow, H., Canchaya, C., Hardt, W., 2004. Phages and the evolution of bacterial pathogens : From genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews* 68 (3), 560–602.
- Carrasco, B., *et al.*, 2016. Modulation of *Lactobacillus casei* bacteriophage A2 lytic/lysogenic cycles by binding of Gp25 to the early lytic mRNA. *Molecular Microbiology* 99 (2), 328–337.
- Chandra, B., Ramisetty, M., Sudhakari, P.A., 2019. Bacterial 'grounded' prophages: Hotspots for genetic renovation and innovation. *Frontiers in Genetics* 10 (February), 1–17.
- Christie, G.E., Calendar, R., 2016. Bacteriophage P2. *Bacteriophage* 7081 (November), e1145782.
- Christie, G.E., Dokland, T., 2012. Pirates of the Caudovirales. *Virology* 434 (2), 210–221.
- Dodd, I.B., Shearwin, K.E., Egan, J.B., 2005. Revisited gene regulation in bacteriophage lambda. *Current Opinion in Genetics and Development* 15 (2), 145–152.
- Golding, I., 2016. Single-cell studies of phage λ : Hidden treasures under occam's rug. *Annual Review of Virology* 3, 7.1–7.20.
- Harshey, R.M., 2014. Transposable phage Mu. *Microbiology Spectrum* 2 (5), 1–22.
- Łobocka, M.B., Rose, D.J., Plunkett, G., *et al.*, 2004. Genome of bacteriophage P1. *Journal of Bacteriology* 186 (21), 7032–7068.
- Mai-Prochnow, A., Gee, J., Hui, K., *et al.*, 2015. Big things in small packages : The genetics of filamentous phage and effects on fitness of their host. *FEMS Microbiology Reviews* 39 (February), 465–487.
- Merrick, C.A., Zhao, J., Rosser, S.J., 2018. Serine integrases: Advancing synthetic biology. *ACS Synthetic Biology* 7, 299–310.
- Ptashne, M., 2004. *A Genetic Switch – Phage Lambda Revisited*. New York: Cold Spring Harbor Laboratory Press.
- Ravin, N.V., 2015. Replication and maintenance of linear phage-plasmid N15. *Microbiology Spectrum* 3 (1), 1–12.
- Schubert, R.A., Dodd, I.B., Egan, J.B., Shearwin, K.E., 2007. Cro's role in the CI-Cro bistable switch is critical for lambda's transition from lysogeny to lytic development. *Genes and Development* 21 (19), 2461–2472.
- Shao, Q., Trinh, J.T., Zeng, L., 2019. High-resolution studies of lysis – Lysogeny decision-making in bacteriophage lambda. *Journal of Biological Chemistry* 294 (10), 3343–3349.

Relevant Websites

- <https://ecocyc.org/>
EcoCyc.
- <https://viralzone.expasy.org/>
ViralZone root – ExPASy.

APPENDIX B: EXPERIMENTAL DETERMINATION OF 186 CI EXTINCTION COEFFICIENT

B.1 Introduction

Accurate knowledge of protein concentration is important when analysing small angle X-ray scattering (SAXS) data. Errors in protein concentration would result in proportionate errors in the estimation of molecular weight of the protein (Chapter 3).

Knowledge of the extinction coefficient of a protein allows its concentration in solution to be determined via UV spectroscopy. The Beer-Lambert law (equation 6) is used to related UV absorbance of the protein to its concentration

$$A_{280\text{ nm}} = \epsilon_{280\text{ nm}} \times c \times l \text{ (Equation 6)}$$

where $A_{280\text{ nm}}$ is the absorbance of the protein at 280 nm, $\epsilon_{280\text{ nm}}$ is the extinction coefficient of the protein at 280 nm, c is the concentration of the protein and l is the pathlength of protein solution the light travels through.

The extinction coefficient of 186 CI-His₆ was estimated theoretically to be 23950 M⁻¹cm⁻¹ (assuming all cysteine residues are reduced) from its primary sequence using the Edelhoch method (Edelhoch, 1967), implemented into the ProtParam webserver (Gasteiger et al., 2005). This calculates the extinction coefficient of a denatured protein as a linear combination of the extinction coefficients of major chromophores phenylalanine, tyrosine, tryptophan and cysteine. This is summarised in equation 7

$$\epsilon_{M,Gdn.HCl} = a\epsilon_{\text{phenylalanine}} + b\epsilon_{\text{tyrosine}} + c\epsilon_{\text{tryptophan}} + d\epsilon_{\text{cysteine}} \text{ (Equation 7)}$$

where a , b , c and d are the number of phenylalanine, tyrosine, tryptophan and cysteine, respectively. However, this implicitly assumes that the protein is in a denatured state where all the aromatic groups are exposed to solvent. A spectral shift could be observed in the chromophores within the native structure depending on if the chromophore is buried or present on the surface of the protein.

The extinction coefficient of the native 186 CI protein was determined by UV-spectroscopy.

B.2 Methods

The sample of 186 CI-His₆ used for SAXS analysis (Chapter 3) was also used to determine the extinction coefficient of native 186 CI-His₆. The estimated protein concentration measured via UV-spectroscopy, using the denatured protein extinction coefficient on native protein, was 8.7

mg/mL. An aliquot of 186 CI and the corresponding matched buffer (from the final dialysis of the protein) was thawed in an ice bath. The protein was denatured by adding the protein solution to concentrated solutions of guanidine hydrochloride.

The protein and buffer solutions made are summarised in Table B- 1.

Table B- 1: Protein and buffer solutions.

Sample ID	Solution constituents
Sample 1 (blank for sample 2)	10 μ L matched buffer + 90 μ L of 7.5 M guanidine hydrochloride
Sample 2	10 μ L CI + 90 μ L of 7.5 M guanidine hydrochloride
Sample 3	10 μ L CI + 90 μ L matched buffer
Sample 4 (blank for sample 5)	18 μ L matched buffer + 82 μ L of 7.5 M guanidine hydrochloride
Sample 5	9 μ L CI + 9 μ L matched buffer + 82 μ L of 7.5 M guanidine hydrochloride
Sample 6	9 μ L CI + 91 μ L of matched buffer

The UV absorbance for the samples was measured using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific) with the UV wavelength set to 280 nm (Table B- 3). Each reading was made 4 times with independently prepared samples. The blanking solutions used for each sample is described in Table B- 2.

Table B- 2: The buffer used to blank the spectrophotometer for different samples.

Sample measured	Blanking buffer
Sample 2	Sample 1
Sample 3	Matched buffer
Sample 5	Sample 4
Sample 6	Matched buffer

B.3 Results

Samples 2 and 3, along with samples 5 and 6 have the same concentration of 186 CI protein. The difference is that one sample contains high concentrations of guanidine hydrochloride which denatures the protein. Thus, the difference in UV absorbance between denatured and native sample can be used to calculate the extinction coefficient of native protein (Table B-3).

Table B- 3: Measured absorbance of native and denatured samples of 186 CI at 280 nm. Errors represent standard error in the mean (n=4)

Sample measured	Absorbance at 280 nm
Sample 2 (denatured)	0.751±0.01
Sample 3 (native)	0.729±0.02
Sample 5 (denatured)	0.654±0.003
Sample 6 (native)	0.618±0.01

The absorbance of native and denatured protein at the same concentration are related by equation 8.

$$\varepsilon_{native} = \frac{A_{native} \times \varepsilon_{denatured}}{A_{denatured}} \text{ (Equation 8)}$$

A_{native} and $A_{denatured}$ denote the absorbance of native and denatured protein, respectively. ε_{native} and $\varepsilon_{denatured}$ are the extinction coefficient of native and denatured protein, respectively. $\varepsilon_{denatured}$ was estimated to be 23950 M⁻¹cm⁻¹ using the Edelhoch method, as described in Section B.1.

Using results from Table B- 3, the extinction coefficient of 186 CI-His₆ was calculated to be 23000 ± 1000 M⁻¹cm⁻¹ (using samples 2 and 3) and 22900 ± 600 M⁻¹cm⁻¹ (using samples 5 and 6).

B.4 References

- Edelhoch, H., 1967. Spectroscopic Determination of Tryptophan and Tyrosine in Proteins. *Biochemistry* 6, 1948–1954.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb.* 571–607.