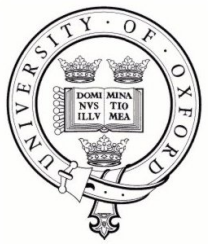
ROLES OF THE TWO CHEMOTAXIS CLUSTERS IN *RHODOBACTER SPHAEROIDES*

Jennifer Anne de Beyer

A thesis submitted in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy at the University of Oxford



MERTON COLLEGE
UNIVERSITY OF OXFORD
TRINITY TERM 2013

Abstract

ROLES OF THE TWO CHEMOTAXIS CLUSTERS IN *RHODOBACTER SPHAEROIDES*

Jennifer Anne de Beyer

Merton College, University of Oxford

Submitted for the degree of Doctor of Philosophy, Trinity Term 2013

Bacteria swim towards improving conditions by controlling flagellar activity via signals (CheY) sent from chemosensory protein clusters, which respond to changing stimuli. The best studied chemotactic bacterium, *E. coli*, has one transmembrane chemosensory protein cluster controlling flagellar behaviour. *R. sphaeroides* has two clusters, one transmembrane and one cytoplasmic. The roles of the two clusters in regulating swimming and chemosensory behaviour are explored here.

Newly-developed software was used to measure the effect of deleting or mutating each chemotaxis protein on unstimulated swimming and on the chemosensory response to dynamic change. New behaviours were identified by using much larger sample sizes than previous studies. *R. sphaeroides* chemotaxis mutants were classified as (i) stoppy unresponsive; (ii) smooth unresponsive or (iii) stoppy inhibited compared to wildtype swimming and chemosensory behaviour. The data showed that the ability to stop during free-swimming is not necessarily connected to the ability to respond to a chemotaxis challenge. The data suggested a new model of connectivity between the two chemosensory pathways.

$CheY_3$ and $CheY_4$ are phosphorylated by the transmembrane polar cluster in response to external chemoeffector concentrations. $CheY_6$-P produced by the cytoplasmic cluster is a requirement for chemotaxis, whether or not the polar cluster is able to produce $CheY_6$-P. $CheY_6$-P stops the motor, whereas $CheY_{3,4}$-P allow smooth swimming. When chemoeffector levels fall, the signals through $CheY_{3,4}$ fall, allowing $CheY_6$-P to bind and stop the motor. As the polar cluster adapts to the fall by the action of the adaptation proteins $CheB_1$ and $CheR_2$, the concentration of $CheY_{3,4}$-P increases again, to compete with $CheY_6$-P and allow periods of smooth swimming. Under aerobic conditions, the cytoplasmic cluster controls the basal stopping frequency and does not appear to respond to external chemoeffector changes.

The role of the adaptation proteins in resetting the signalling state in *R. sphaeroides* is unclear, particularly the roles of the proteins associated with the cytoplasmic cluster, $CheB_2$ and $CheR_3$. Tandem mass spectrometry was used to identify glutamate and glutamine (**EQ**) sites on the cytoplasmic *R. sphaeroides* chemoreceptor TlpT that are deamidated and methylated by the *R. sphaeroides* adaptation homologues. In *E. coli*, adaptation sites are usually **EQ/EQ** pairs. However the sites reported in TlpT vary at the first residue in the pair. Mutation of the putative **EQ** adaptation sites caused changes in adaptation, suggesting that $CheY_6$-P levels are controlled and reset by $CheB_2$ and $CheR_3$ controlling the adaptation state of TlpT.

Declaration

The work in this thesis was undertaken at the University of Oxford; in the Department of Biochemistry. Work was performed from October 2009 to October 2013 under the supervision of Prof. J.P. Armitage. All the work in this thesis is my own unless otherwise stated and has not been submitted for a degree at this or any other university.

Acknowledgements

Firstly, thank you to my supervisor, Prof Judy Armitage, for giving me the freedom to satisfy my curiosity, and for her unfailing enthusiasm for new data. Thank you to Dr Kathryn Scott and Dr Mark Roberts, for their encouragement, guidance and teaching.

The lovely people of the Armitage and Wadhams labs, past and present, for their generosity with their time and knowledge. Thank you for all the morning coffee sessions, the Friday pub trips and the general hilarity. Its been a privilege to work with you all.

The work by Elaine Byles and Qin Qi in making the TlpT mutant strains and corresponding $CheW_4$-CFP tagged strains used in Chapter 5 is gratefully acknowledged.

The Clarendon Fund, the Commonwealth Scholarship Commission and Merton College, many thanks for their generous financial assistance.

Friends far and near, new and old. In particular, Denise, Martin and Philippa, thank you for always being just a phonecall or email away. The Oxford Aikido dojo, for keeping me (mostly) on my feet. Mireille, John, Emma, Jacqui, Andrew and Daniel, thank you for all the laughter, your enthusiastic reception of my kitchen experiments, and your support and hugs when things got tricky.

Finally, and with all my heart, thank you to my family. The Shropshire lot, for being a home away from home for five years. The Cape Town lot, Mom and Dad, Paul and Leigh, thank you for the visits and the kospakkies, the cheerleading and the support and the love. Its easy to risk and dream when you have such strength behind you.

Publications

Alice C. Ind, Steven L. Porter, Mostyn T. Brown, Elaine D. Byles, **Jennifer A. de Beyer**, Scott A. Godfrey and Judith P. Armitage (2009) Inducible-expression plasmid for *Rhodobacter sphaeroides* and *Paracoccus denitrificans* and its use to investigate chemotaxis. *Appied and Environmental Microbiology* 75(20): 6613–6615.

Gabriel Rosser, Alexander G. Fletcher, David A. Wilkinson, **Jennifer A. de Beyer**, Christian A. Yates, Judith P. Armitage, Philip K. Maini and Ruth E. Baker (2013) Novel methods for analysing bacterial tracks reveal persistence in *Rhodobacter sphaeroides*. *PLoS Computational Biology* 9(10): e1003276.

Contents

List of Figures

List of Tables

Abbreviations

ANOVA	analysis of variance
ATP	adenosinetriphosphate
bp	base pairs
BSA	bovine serine albumin
CFP	cyan fluorescent protein
Che	chemotaxis
$CheB_c$	CheB methylesterase domain
cheOp	chemotaxis operon
CCW	counter clockwise
CD	cytoplasmic domain
CW	clockwise
Δ	Delta, the deletion of a protein
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside 5-triphosphate
EDTA	ethylene diamine tetra acetic acid
E^{met}	methylated glutamate
EIC	extracted ion chromatography
EMS	effective mean speed
gDNA	genomic DNA
HAMP	histidine kinase, adenyl cyclase, methyl binding protein and phosphatase domain
HCD	highly conserved domain
HEPES	sodium N-2-hydroxyethylpiperazine-N-2-ethanesulphuric acid
HMM	hidden Markov model
HRP	horse radish peroxidase
IPTG	isopropyl-β-D-thiogalactopyranoside
Kan	kanamycin
kb	kilobase

kDa	kilo Dalton
lac	lactose
LB	Luria-Bertani
LBA	Luria-Bertani agar
MAC	mean angle change
MBR	minimum bounding radius
MCP	methyl-accepting chemotaxis protein
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MSA	multiple sequence alignment
MilliQ	ultrapure water
Nal	naladixic acid
NC	Non-chemotactic
NEMS	Normalised effective mean speed
Ni-NTA	nickel nitrilotriacetic acid
NM	Nonmotile
NMR	nuclear magnetic resonance
OD_{600nm}	optical density at 600 nm
-P	phosphorylated
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PVDF	polyvinylidene fluoride
RNA	ribonucleic acid
SAM	S-adenosyl-L-methionine
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
Sux	succinate medium
TBE	Tris-borate-EDTA buffer
TFB	transformation buffer
Tlp	transducer-like protein
T_m	melting temperature
Tris	2-amino-2(hydroxymethyl)-1,3-propandiol
TM	transmembrane
UV	ultraviolet
X	any amino acid
YFP	yellow fluorescent protein

Standard one letter abbreviations used for amino acids

Chapter 1

Introduction

Chemotaxis, the ability to alter movement in response to chemical stimuli in the environment, is a trait that many bacterial species use in their interactions with their environments, which includes us. The formation of symbiotic relationships between nitrogen-fixing bacteria and legumes relies on chemotaxis and motility (for example, see Doyle and Luckow (2003); Miller *et al.* (2007)). Pathogens such as *Campylobacter* and *Helicobacter* use chemotaxis for infection, linking virulence and chemotaxis (reviewed in Lertsethtakarn *et al.* (2011)). Biofilms foul up areas great (industrial processing plants) and small (catheters and those hard to reach places between teeth), and again chemotaxis is implicated (reviewed in Verstraeten *et al.* (2008)).

Our strategy for countering the less desirable of these interactions with microbes is to understand and disrupt. Knocking out the chemotaxis pathway in pathogens greatly decreases the ability to infect (for example, McGee *et al.* (2005)), while disrupting biofilm formation leaves bacteria vulnerable to traditional antibiotics (for example, Rogers *et al.* (2010)). However, a finer-grained understanding of the network is required if effective drug targets against chemotaxis in pathogens are to be produced. Any understanding gained of the chemotaxis network can also be applied to the understanding of other signalling pathways within microbes.

We have therefore expended much energy in understanding chemotaxis within the

model organism *Escherichia coli*, chosen for the ease with which it can be cultured and manipulated. However, as will be explained below, the *E. coli* chemotaxis pathway is the simplest known and is not necessarily representative of the pathways used by the pathogens and biofilm-formers that are our targets. By unravelling and understanding these more complex pathways to the same extent as the *E. coli* pathway, we increase our understanding of within-cell signalling in general, and bring the development of anti-biofilm and chemotaxis-targeting drugs one step closer.

Rhodobacter sphaeroides is one species with a complex chemotaxis pathway. It has multiple homologues of the core chemotaxis proteins, which localise to two distinct regions in the cell. Although it is the only species known to have two clusters, many species also have multiple homologues of the core chemotaxis proteins, such as *Vibrio cholera*, and the search for additional clusters is ongoing. Understanding the complex network connectivity in *R. sphaeroides* chemotaxis therefore has practical implications in future drug target development.

This thesis focuses on the requirement for the two clusters of chemotaxis proteins for *R. sphaeroides* chemotaxis to function correctly. Here, pertinent previous literature is presented. Flagellate bacterial motility is introduced and the core set of proteins used to effect chemotaxis is explained. What is already known of the *R. sphaeroides* chemotaxis network is described and compared to that found in other model organisms, namely *E. coli* and *Bacillus subtilis*.

1.1 Bacterial motility

Microbes have evolved motility systems to deal with the challenge of movement in a low Reynolds environment, in which inertial forces are overridden by viscous forces, and at a size where Brownian motion must be overcome (Purcell, 1977). A number of strategies exist for moving across a solid surface. Twitching motility has the microbe pulling itself along the surface by extending and retracting pili (re-

viewed in Burrows (2012)), adventurous gliding involves proteins arranged in a helical pattern engaging and disengaging the surface, propelling the cell body forward by rotation (Mignot *et al.*, 2007), and *Mycoplasma* species use a series of protein 'legs' attaching and releasing to pull the cell body along a surface (reviewed in Miyata (2010)). Free swimming in a liquid medium, however, is usually accomplished by flagellate motility and is the subject of this thesis.

1.1.1 Flagellate motility

Flagellate bacteria propel themselves forward through liquid by rotating a propeller, formed of one or more flagella embedded in the cell membrane by a rotating motor (Berg and Anderson, 1973), powered by a proton or cation gradient across the membrane .

The flagellum is a helical, semi-rigid protein filament, made up of polymerised strands of the protein FliC. The filament attaches into the flagellar motor by a flexible hook (Figure 1.3). Flagellar arrangements are diverse (see Figure). Some species have a single flagellum, situated at the pole (*Vibrio cholerae*) or at random along the cell body (*R. sphaeroides*), called a monotrichous arrangement. Some species have a tuft of flagella at one (*Helicobacter pylori*) or both poles (*Pseudomonas spp*), lophotrichous and amphitrichous respectively. Some species have flagella arranged at random around the cell body, called a peritrichous arrangement (*E. coli*).

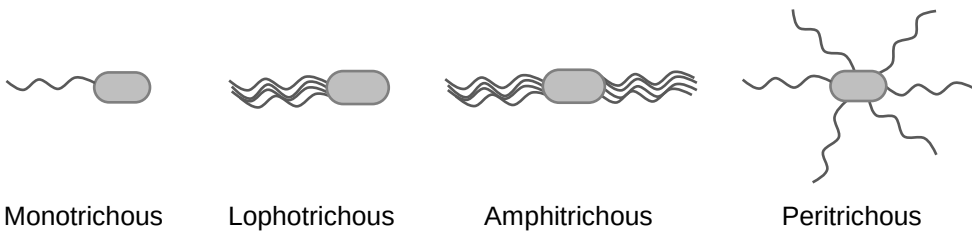


Figure 1.1: Flagellar arrangements

The majority of flagellar motors are able to rotate in both directions. When switching directions, the shape of the flagellum helix and the interactions of the filaments changes, causing periods of forward motion (runs) and periods of reorientation. When the flagellar motor of a peritrichious bacteria rotates counter-clockwise (CCW), the flagella come together to form a flagellar bundle and drive forward motion. Switching to clockwise (CW) rotation causes the bundle to fly apart and the cell to actively tumble on the spot. Lophotrichous, amphitrichous and monotrichous bacteria usually move forward by CCW rotation and reorientate by reversing briefly during CW rotation, Brownian motion introducing a new direction when forward motion is restored (Taylor and Koshland, 1974). *Vibrio alginolyticus* adds an additional movement, using a run-reverse-flick motion to actively introduce direction change rather than relying on Brownian motion (Xie *et al.*, 2011).

R. sphaeroides has an unusual motility pattern. The flagellar motor is unidirectional (Armitage and Macnab, 1987). Forward motion occurs during CCW motor rotation. When the motor stops, the cell is passively reorientated by Brownian motion (Pilizota *et al.*, 2009).

All forms of motility allow the cell to explore space by a 'random walk', as there is no control over each run's direction. However, the cell can integrate information from the environment to bias the frequency of runs and tumbles/reverses/stops, to bias movement into a particular direction. This is referred to as a biased random walk (Figure 1.2). Bacteria are able to use external stimuli such as light (phototaxis), oxygen (aerotaxis) and chemoeffector concentrations (chemotaxis) in this way. Chemotaxis is discussed in more detail in Section 1.2.

1.1.2 The flagellar motor

The flagellar motor (reviewed in Berg (2003); Brown *et al.* (2011)) is an electrical motor, powered by a gradient of H^+ or Na^+ across the inner membrane, depending

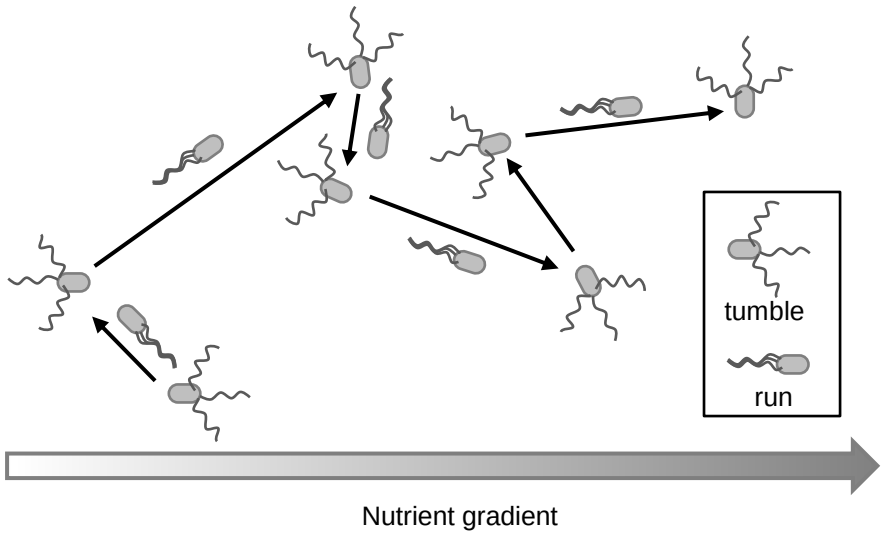


Figure 1.2: Motile chemotactic flagellate bacteria explore the environment using a random directed walk. Increases in the chemoeffector concentration biases motion towards runs and decreases in the chemoeffector concentration increases the frequency of tumbles.

on the species. The motor requires approximately 50 gene products for synthesis, 25 of which are found in the final structure. As mentioned above, the complete flagellum is formed of three parts: the filament, the flexible hook and the motor. The motor is itself formed of the basal body, which is the rotor, and the stationary stators. A great deal is known about the structure of the motor, thanks to biochemical and structural studies and successful electron cryomicroscopic imagery.

The core of the rotor structure is highly conserved. It is formed of a central rod, connected to the flexible hook, and a series of rings of protein subunits embedded in the layers of membrane and cell wall, able to rotate. These are the MS (inner membrane) and C (cytoplasmic) rings. In Gram-negative bacteria, two additional rings are stationary around the rotating rod, the L (lipopolysaccharide outer membrane) and P (peptidoglycan cell wall) rings. The rod acts as the drive-shaft in the middle of the rotor.

Rotational torque is generated by the stators. Stators are stable complexes formed of four MotA and two MotB subunits (Kojima and Blair, 2003). Eleven inner-membrane-

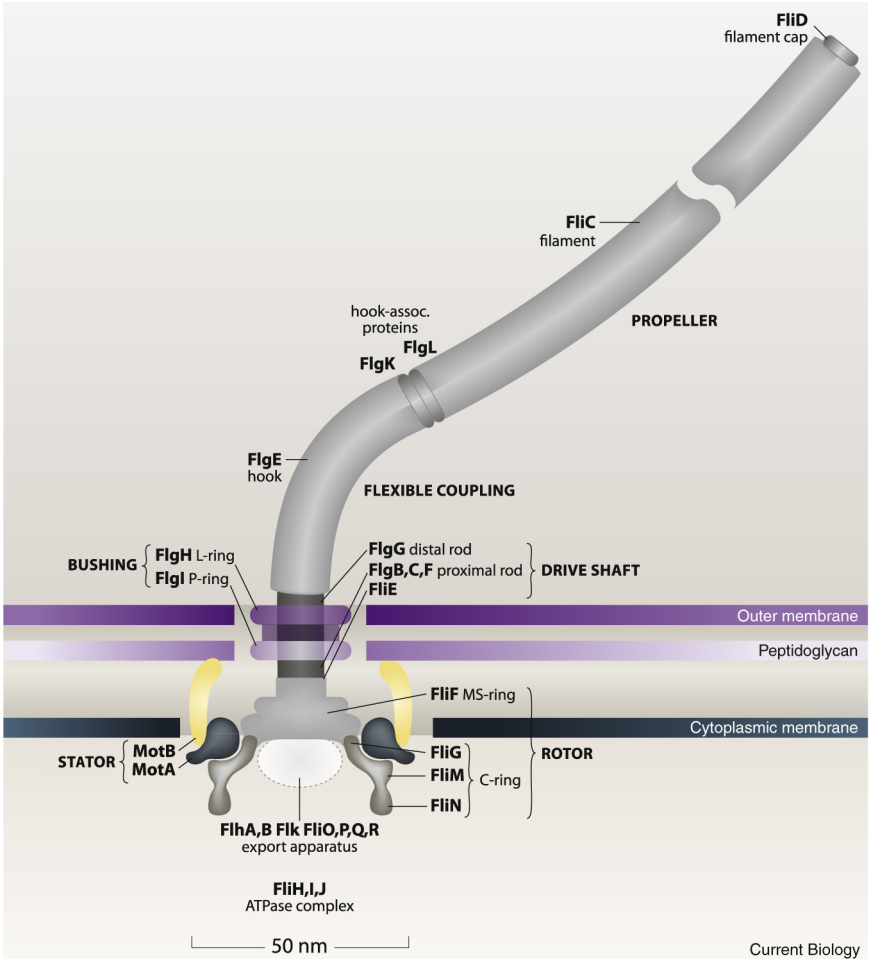


Figure 1.3: The flagellar motor. The flagellar filament (FliC) connects to the basal body of the flagellar motor by a flexible hook. The basal body is embedded in the outer liposaccharide membrane (L-ring) and peptidoglycan cell wall (P-ring). Ions flow through channels in the piston, the stator complex, to generate rotation (MotA and MotB). Binding of CheY-P to the switching complex (the C-ring: FliG, FliM and FliN) determines the direction of rotation. Reproduced from Berg (2008) with permission from Elsevier, license number 3692660107125.

bound stators bind to the peptidoglycan layer and dynamically associate with the motor at a time (Reid *et al.*, 2006). How torque is generated is not yet fully understood, but protons are able to bind to the stator complex at the periplasmic interface, then move through the complex channel to the cytoplasmic interface. The conformational changes produced in the stator complex by the binding and release of protons to a conserved aspartate residue (Asp32) drives rotation of the C-ring and thus rotation of the rest of the rotor and, in turn, filament. The stators must hold tightly to their inner membrane and peptidoglycan layer docking sites in order to have suffi-

cient leverage to rotate the full flagellum. Individual stators remain associated to the membrane for as little as 30 s at a time, exchanging with the pool of stators within the inner membrane (Leake *et al.*, 2006).

The C-ring, or switching complex, interacts with the stators and determines the direction of rotation. It is formed of three rings of proteins, FliM, FliN and FliG. The precise conformation of these dynamic proteins is not yet known. Torque generation is driven by FliG-stator interaction. Rotational direction is driven by FliM binding to the signalling protein CheY-P. In *E. coli*, when FliM is unbound, CCW rotation occurs. As more FliM subunits bind CheY-P, the probability of the motor switching rotation to CW increases. The opposite occurs in *B. subtilis*. In *R. sphaeroides*, binding of CheY-P produces a stop.

Rotation of the motor can reach 100 Hz in proton-driven motors and 1300 Hz in sodium-driven motors. Typical mean swimming speeds are in the range of 15-50 μm/s in different species (for a recent range of speeds, see Youle *et al.* (2012)), with speeds of 600 μm/s and 1 mm/s recorded in *Thiovulum majus* (Garcia-Pichel, 1989) and *Candidatus* Ovobacter propellens (Fenchel and Thar, 2004) respectively.

1.1.3 Tracking swimming bacteria

Swimming behaviour must be analysed to compare different species and to characterise the effects of biochemical changes to pathway components. Swimming behaviour under low load conditions can be analysed either one cell at a time or on a population scale. Single cell data can be obtained by holding a cell within an optical trap, which uses a narrow beam of highly focused light to hold a dielectric object stable in three-dimensions, while still allowing movement like flagellar rotation (Ashkin *et al.*, 1986). The flagellum can be truncated and stuck to a bead, known as an optical bead, through antibody interactions. The regularly shaped bead is easily recorded as it rotates under the power of the flagellar motor within the optical trap

(Ryu *et al.*, 2000). This gives clean clear data, but is time consuming and suffers from cell to cell variability. The majority of studies have focused on whole populations, averaging out the variability seen in individual cells across the population.

Populations of bacterial cells can be imaged under a microscope and then tracked. Individual cells can be followed as they swim, by physically moving the microscope stage. This gives long tracks of around 20 s, but introduces a selection bias, as the experimentalist chooses which cells to follow (Berg and Brown, 1972). Alternatively, cells can be captured as they swim in and out of a field of view. This removes any selection bias for which cells to image, but decreases track length and introduces many 'junk' tracks, such as dead cells.

Once cells have been imaged, the images must be converted into tracks. This has been done by active selection of a cell in an image and confirmation of that cell's identity in successive images (Poole *et al.*, 1988). More commonly, automated image segmentation using differences in pixel intensity has been used to identity cells, and tracks have been joined up using for example nearest neighbour algorithms (for example, Sager *et al.* (1988); Alon *et al.* (1998)). A challenge for these kinds of algorithms is that, due to their size and their low Reynolds number environment, bacteria are constantly buffeted by the environment, introducing stochasticity into what should be a straight track (Purcell, 1977). Often runs will appear as curved tracks, rather than straight lines.

The resulting tracks are then summarised by mean swimming speed (Alon *et al.*, 1998) and mean angle changes (Berg and Brown, 1972) frame by frame or mean curvature across a track, using these measures as a shorthand for swimming behaviour as a cell that tumbles more frequently is likely to have a lower mean speed and a higher curvature.

Tracks can also be classified into runs and tumbles. This is either done manually (Sager *et al.*, 1988) or using heuristics, such as frame to frame speed and angle change

cut-offs to define a tumble (Berg and Brown, 1972; Alon *et al.*, 1998). Both methods are subjective. In manual classification, only stopping events that can be seen by eye can be identified and the definition of a stop is subjective, as there is no 'gold-standard' stop for comparison. In heuristic classification, the measures used to define a stop require knowing a great deal about swimming behaviour before analysis and prior knowledge about what a particular strain's swimming looks like may colour the parameters selected. Of the studies mentioned, only Alon *et al.* (1998) considered how robust their classification scheme is to the selection of heuristic parameters.

1.2 Bacterial chemotaxis

Chemotaxis, or chemically directed movement, is how bacteria are able to bias their movement to move up concentration gradients of certain chemicals, known as chemoeffectors, and down concentration gradients of toxins. Bacteria use two-component signalling to respond to drops or increases in the chemoeffector concentration (reviewed in Falke *et al.* (1997), Stock *et al.* (2000) and West and Stock (2001)).

In one-component signalling, a stimulus directly produces a response within a cell. The enzyme or protein that senses the stimulus change also causes the response within the cell. In two-component signalling, the sensing protein and effector protein are different. A histidine kinase enzyme is activated by the stimulus. It in turn activates its cognate response regulator by either phosphorylating or dephosphorylating it. As the cell uses the histidine kinase-response regulator motif in many pathways, such as nitrogen sensing and sporulation, within the cell, these pairs are highly specified to one another.

The *E. coli* chemotaxis system (Figure 1.4) is not only the paradigm for our understanding of chemotaxis, but also for the study of two-component signal transduction (reviewed in Wadhams and Armitage (2004)). A transmembrane protein, the

methyl-accepting chemotaxis protein (MCP), binds chemoeffectors in the environment (Hazelbauer *et al.*, 2008). When chemoeffector levels drop, the MCPs change conformation, causing a conformational change in the associating cytoplasmic kinase CheA. CheA autophosphorylates, then phosphorylates the response regulator CheY (Borkovich *et al.*, 1989).

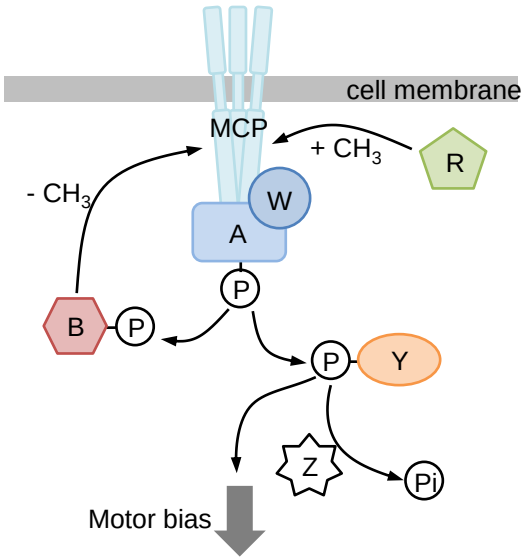


Figure 1.4: The *E. coli* chemotaxis network. Decreased ligand binding to the MCPs causes autophosphorylation of CheA, which then phosphorylates CheY. CheY-P binding to the flagellar motor increases the probability of a tumble. CheY-P is dephosphorylated by CheZ. CheA also phosphorylates CheB. CheB-P desensitises the MCPs by demethylating them, ending the signal. CheR remethylates the MCPs, increasing sensitivity. Adapted from Scott *et al.* (2012) with permission from the authors and Caister Academic Press.

Phosphorylated CheY (CheY-P) binds to FliM in the motor, biasing its rotation to result in tumbling, allowing a change in direction (Dyer *et al.*, 2009; Sarkar *et al.*, 2010). The system must be reset through the adaptation pathway before the running motion can resume. CheY-P is dephosphorylated through both an autodephosphorylation and an enzymatic route using CheZ.

The kinase CheA also phosphorylates the response regulator CheB, a methylesterase. This phosphorylation reaction occurs at a slower rate than to CheY, so that CheB-P's

effects occur after CheY-P has had a chance to cause a stop(Hess *et al.*, 1988). The MCP can be methylated on certain glutamate residues (Terwilliger and Koshland, 1984). Methylation induces a conformational change in the MCP, changing the dissociation constant between the MCP and its chemoeffector. The more methylated the MCP, the more likely it is to change conformations and activate CheA at a particular chemoeffector concentration and thus the more sensitive it is to a small drop in chemoeffector. Activated CheB-P rapidly demethylates the MCP, increasing MCP-chemoeffector binding, thus lowering the MCP's sensitivity to the current chemoeffector concentration and making it more likely to return CheA to the inactive conformation (Kehry and Dahlquist, 1982; Kehry *et al.*, 1985). CheY-P production stops and the proportion of unphosphorylated CheY rises. This returns the flagellar motor to the running motion.

In preparation for the next signalling event, the methylation state of the MCP is gradually increased through the action of the constitutively active methyltransferase CheR (Springer and Koshland, 1977).

To complicate the above story, MCPs are found in large complexes (Maddock and Shapiro, 1993). Within a cluster of MCPs, different methylation states and thus sensitivities will be found, giving the cell sensitivity to changes in attractant across a range of background concentrations of five to six orders of magnitude. This also leads to population variability at the cell level in how cells respond to environmental changes.

Each of these steps and proteins will be explored in greater detail below.

1.2.1 Sensing: the chemoreceptors

In chemotaxis, changes in the concentration of various chemoeffectors are sensed by the chemoreceptors. The number of chemoreceptors varies greatly between species, from single receptors in species such as *Methanococcoides burtonii* and *Mesorhizobium*

loti, to those with dozens, such as *Magnetospirillum magnetotacticum*'s 65 putative receptors (Alexander and Zhulin, 2007). *E. coli* has four chemoreceptors that probably arose from gene duplication given their genetic similarity, each responding to a different ligand or group of ligands: Tar (aspartate), Tsr (serine), Tgr (galactose and ribose) and Tap (dipeptide receptor). A fifth receptor, Aer, associates with the four chemoreceptors in the cluster. It responds to oxygen concentration and is responsible for aerotaxis. Within a species, there is also significant disparity in the expression levels of the different chemoreceptors: *E. coli*'s Tar and Tsr are highly abundant with several thousand copies of each (Clarke and Koshland, 1979; Stock and Koshland, 1981), whereas Trp and Tap are low-abundance with ten-fold less (Hazelbauer *et al.*, 1981).

The majority of chemoreceptors are methyl-accepting chemotaxis proteins (MCPs) and share a common conserved structure (Figure 1.5). Chemoreceptors that share conserved features but do not have a transmembrane region are classified as transducer-like proteins (Tlps) and will be discussed later.

MCPs are transmembrane proteins, approximately 350 Å in length. Structural studies of the extracellular (Chi *et al.*, 1997) and cytoplasmic (Kim *et al.*, 1999) regions of the MCP show that the majority of the protein is composed of α-helices. The extracellular ligand binding region is flanked by two transmembrane (TM) regions. A HAMP domain connects TM2 to the cytoplasmic region, known as the highly conserved domain (HCD). The HCD is further divided into two signalling domains and two adaptation domains. The C-terminus tail of the MCP ends in a conserved recognition site for CheR binding, known as the pentapeptide tail.

Ligand-binding region

The periplasmic domain of the MCP can be highly variable, as each chemoreceptor has affinity for different ligands and stimuli. In cases such as *E. coli*, where the four

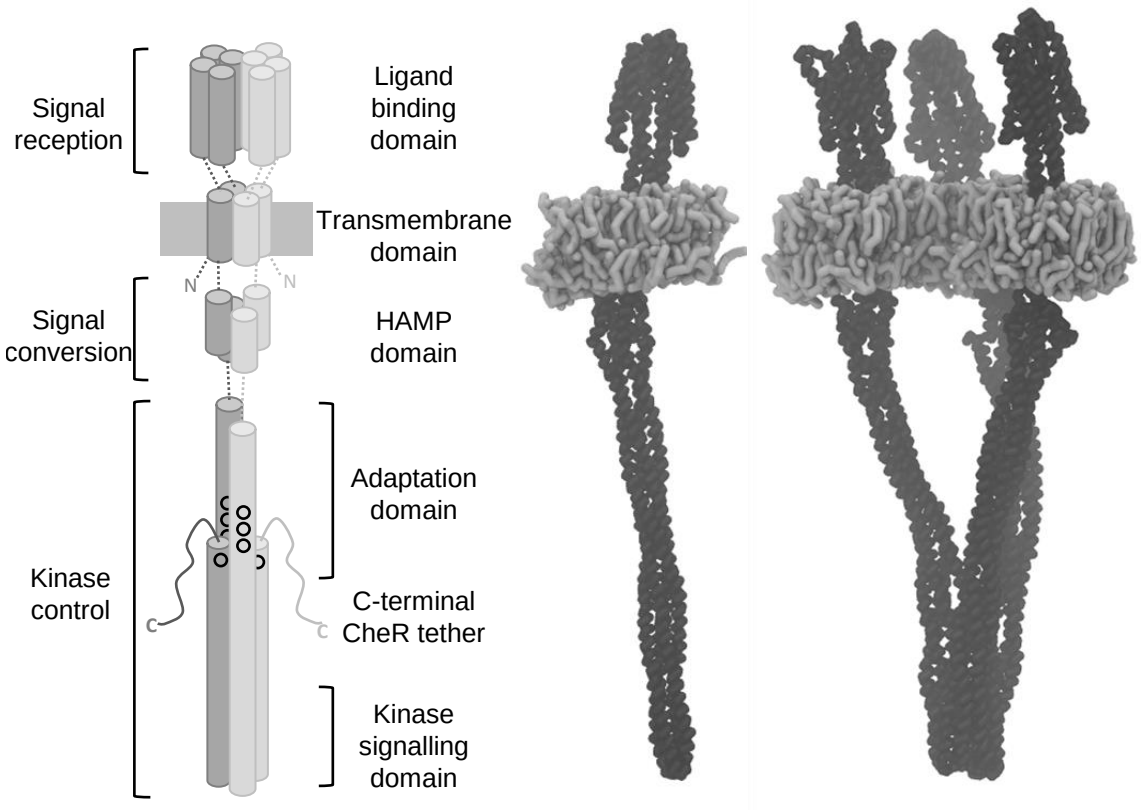


Figure 1.5: Conserved structure of the MCP, in cartoon form (left) and as a molecular dynamics simulation, as a dimer (centre) and trimer of dimers (right). Taken from Scott *et al.* (2012) with permission from the authors and Caister Academic Press, © Caister Academic Press.

receptors likely arose from gene duplication, the receptors are very similar. The periplasmic domain of Tar has been crystallised bound to its ligand, aspartate (Chi *et al.*, 1997).

Transmembrane domain

E. coli MCPs have two α-helix transmembrane domains: TM1 is N-terminus of the ligand binding domain and TM2 links the ligand binding and cytoplasmic domains. Changes in binding rates of ligand in the periplasmic space are transmitted into the cell, to the HAMP domain, via TM2. The precise mechanism by which the signal is transducted is as yet not understood and a number of possible mechanisms have

been proposed (Hulko *et al.*, 2006; Kitanovic *et al.*, 2011). What is known is that attractant binding to *E. coli*'s Tar results in a small, 1-3 Å displacement of one helix in TM2 (Draheim *et al.*, 2006). Shifting TM2 up and down within the membrane and altering its length alters signalling through TM2 and how readily the adaptation enzymes modify the receptor (Adase *et al.*, 2012; Wright *et al.*, 2011).

HAMP domain

Bridging the transmembrane and HCD regions in the majority of MCPs is a histidine kinase, adenylate cyclase, methyl-accepting chemotaxis protein and phosphatase (HAMP) domain. The conformation of the HAMP domain determines how a signal from the ligand binding domain will be propagated. Different conformations of the HAMP domain can convert ligand binding into either activation or deactivation of the associating cytoplasmic kinase (Airola *et al.*, 2013). For a recent review of HAMP domains in chemoreceptors, see Parkinson (2010).

Highly conserved domain (HCD)

The cytoplasmic region of the MCPs is highly conserved across all species and can be used to classify the MCPs, suggesting the evolutionary path of MCP development. Initial studies (Moual and Koshland, 1996) with a limited database of MCPs showed a conserved motif running throughout the length of the HCD of seven-residue repeats **a-b-c-d-e-f-g**, corresponding to two turns of the α-helix, where the first (**a**) and fourth (**d**) position are hydrophobic. These initial sequence alignments suggested three MCP classes based on the presence of eight of these repeats. The repeats in question are found as four sets of 14 amino acids each, positioned as two pairs, each pair symmetrical about the central glycine turn of the HCD. The repeats were shown to be lost in sets of four, as an entire pair of 14 amino-acids was lost concurrently.

Class III MCPs contain both pairs of 14 amino acids and are likely to be the original receptors. Class II (one loss) and Class I (two losses) MCPs may have formed through either direct loss or gene transfer.

This analysis was repeated a decade later, using a database of over 2,000 potential MCPs (Alexander and Zhulin, 2007). The existence of the seven-residue repeat was confirmed across this much larger data set. Sequence conservation and the number of symmetrical indels (symmetrical about the central glycine turn) were used to define seven major classes of MCPs, of which three were the original classes. The seven classes and the positions of their heptad repeats are indicated in Figure 1.6.

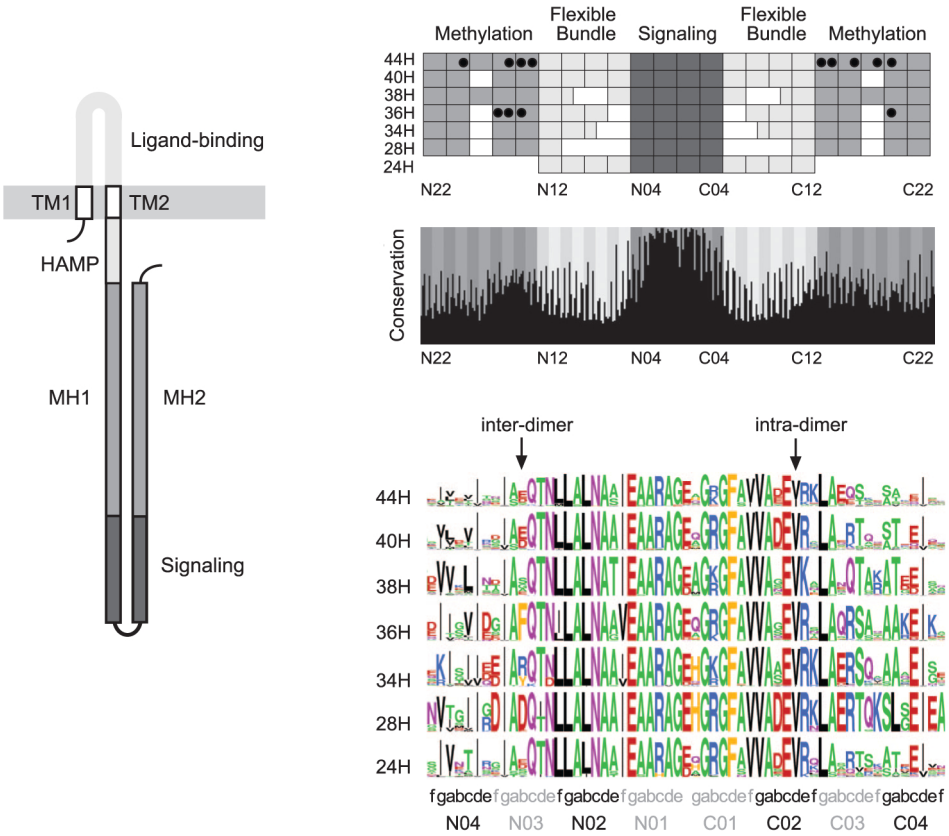


Figure 1.6: Classification of MCPs by the length of the HCD. Left: Cartoon showing the regions of an MCP. The methylation domains (MH1 and MH2) and the signalling domain form the HCD. Right, top: Heptad insertions and deletions in the major MCP classes. Right, bottom: Sequence homology and heptad positions in the MCP signalling domain. Taken from Alexander and Zhulin (2007) with permission, © 2007 National Academy of Sciences.

Class 44H corresponds to Class III in the original system and is typified by *Bacillus subtilis* and *Thermatoga maritima*. Class 36H corresponds to the Class I MCPs and is typified by *E. coli* and *Salmonella*, extensively studied in the model *Salmonella enterica* subsp. *enterica* serovar Typhimurium.

This analysis also clearly delineated the HCD into three domains, each comprising two regions lying symmetrically to one another on either side of the glycine turn: the methylation region, a flexible bundle and the signalling domain. The methylation region will be described in more detail in section 1.2.4.

The flexible bundle region shows the least conservation of residues in the HCD.

C-terminal flexible tail

The C-terminal region of the MCP is an unstructured, flexible linker region (Bartelli and Hazelbauer, 2011). In *E. coli*, it was shown that the final pentapeptide (NWETF) is the sole site of CheR binding to the MCP (Wu *et al.*, 1996). The high abundance *E. coli* and *Salmonella* Typhimurium MCPs Tar and Tsr end in this motif and the *Salmonella* Typhimurium citrate receptor Tcr ends in the highly similar NWESF. Loss of this pentapeptide tail from these receptors greatly reduces methylation and thus chemotaxis in these organisms. The two low-abundance receptors (Tap and Tgr) do not have the NWETF motif and, when expressed without the presence of one of the high-abundance receptors, are methylated poorly. This suggests that they are usually methylated by CheR bound to a high-abundance MCP.

Structural studies (Djordjevic and Stock, 1998) and multiple sequence alignment analyses were used to suggest a general consensus sequence **X-[HFWY]-X-X-[HFWY]** for this pentapeptide motif (Alexander and Zhulin, 2007; Perez and Stock, 2007). Both groups found this consensus sequence in approximately 10% of MCPs analysed, and only in 34H and 36H class MCPs. As yet, this consensus sequence has not been confirmed experimentally beyond the initial *E. coli* and *Salmonella* Typhimurium

sequences.

MCPs from classes other than 34H and 36H possess the unstructured flexible tail as above, but do not appear to have a specific binding site for their adaptation enzymes as in the proteobacteria. The *T. maritima* MCPs share a similar C-terminal region, but removal of this does not reduce methylation of these receptors *in vitro* by CheR (Perez and Stock, 2007).

MCP dimerisation and packing

MCPs form stable homodimers in the cytoplasmic membrane, then form the functional unit, a trimer-of-dimers (Liu *et al.*, 1997). Trimers-of-dimers can be mixed, including different chemoreceptors, but it is believed that only receptors of the same HCD length can form a cluster together. Trimer-of-dimers pack together into clusters in a conserved 12 nm hexagonal lattice arrangement (Briegel *et al.*, 2009, 2012) (Figure 1.7), although *T. maritima* may have an alternative hedgerow of dimers arrangement (Park *et al.*, 2006).

1.2.2 Signalling

When MCPs change conformation due to a decrease in ligand binding, associating CheA in the cytoplasm autophosphorylates, then phosphorylates CheY. CheY-P then diffuses to the motor. It takes approximately 100 ms for the signal to travel from the receptor cluster to the motor (Sourjik and Berg, 2002).

CheA

CheA is a histidine kinase protein, functioning in a two-component signalling system with CheY and CheB. CheA is made up of five functional domains (Figure 1.8). P1 domain contains the conserved histidine residue (H48) that is phosphorylated.

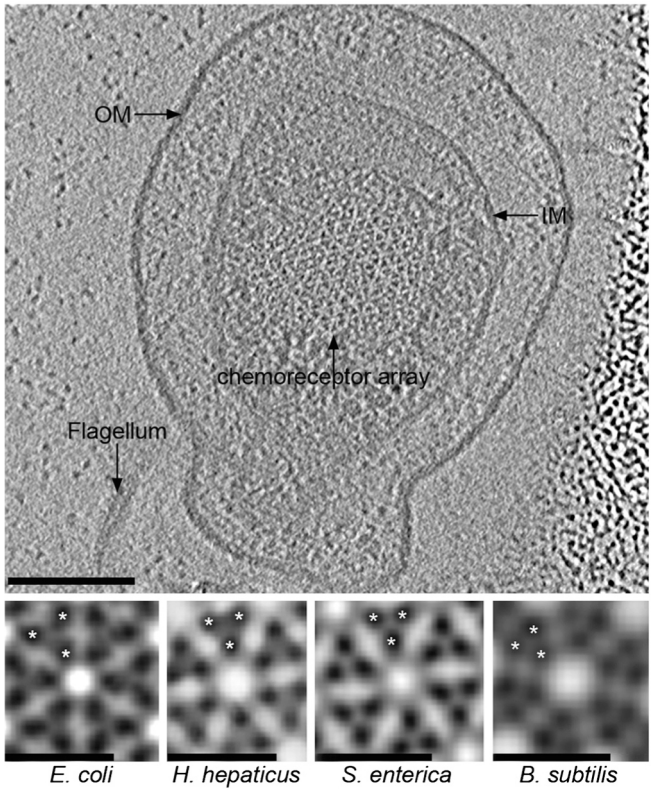


Figure 1.7: Architecture of native chemoreceptor arrays as seen by electron cryotomography. (Upper) Tomographic slice through the top of a *S. enterica* minicell. OM, outer membrane; IM, inner membrane. (Scale bar: 100 nm.) (Lower) Subtomogram averages of (from left to right) *E. coli*, *H. hepaticus*, *S. enterica*, and *B. subtilis* chemoreceptor arrays, showing individual receptor dimers (asterisks). Trimers-of-dimers are oriented so that one receptor dimer points toward (scale bars: 12 nm.) Taken from Briegel *et al.* (2012) with permission, © 2012 National Academy of Sciences.

P2 is the binding domain for the response regulators CheB and CheY (Stewart *et al.*, 2000; Jahreis *et al.*, 2004). As CheY has greater affinity for the P2 domain than CheB, signal is transmitted to the flagellar motor before feedback shuts the signal down. The P3 domain is the dimerisation domain, as CheA functions as a homodimer. The P4 domain is the kinase domain, splitting an ATP molecule and transferring the γ-phosphoryl group to the P1 domain on the other half of the dimer (Swanson *et al.*, 1993). The final domain, P5, controls localisation of CheA to the cluster and couples chemoreceptor signalling to the rate of CheA autophosphorylation (Bourret *et al.*, 1993; Kim *et al.*, 1999).

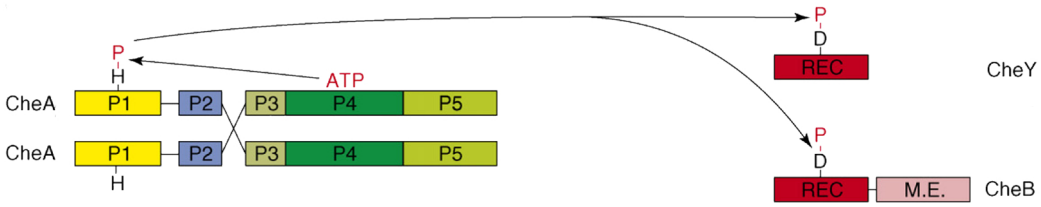


Figure 1.8: Structure of a typical CheA histidine kinase, indicating domains (P) and its response regulator partners. REC=response regulator domain, ME=methylesterase domain in CheB proteins. Adapted from Porter *et al.* (2008b), with permission from Elsevier, license number 3692710333670.

CheY

CheY is a small protein (14 kDa in *E. coli*) and the only *E. coli* chemotaxis protein known to diffuse in the cell. It is the response regulator for CheA. After binding to CheA's P2 domain, CheY is phosphorylated on a conserved aspartate residue (D57). CheY-P has a lower affinity for CheA than CheY, so dissociates and diffuses to the flagellar motor. Binding of CheY-P to FliM increases the probability of motor switching to CW rotation (Dyer and Dahlquist, 2006). The relationship between CheY-P and the motor bias is steeply sigmoidal, so that a small change in CheY-P concentration results in a rapid change in the motor bias (Cluzel *et al.*, 2000). The CheY-P concentration is therefore a crucial step in signal amplification from receptor to motor.

CheY is phosphorylated on the residue D57. Mutations of this residue in *E. coli* result in non-chemotactic strains. The majority of mutations, including D57A, result in a CheY that cannot be phosphorylated, a motor that only supports CCW rotation and a smooth swimming cell. However, the mutation D57N allows flagellar switching in the absence of CheZ, through an additional phosphorylation event on S56. If this additional site is also mutated, then CheY cannot be phosphorylated or bind effectively to the motor and a smooth swimming phenotype is seen (Appleby and Bourret, 1999).

1.2.3 Signal termination

CheY-P can be dephosphorylated by autodephosphorylation or by the action of the phosphatase CheZ. In *E. coli*, the half life of CheY-P in the absence of CheZ is 14 s (Lukat *et al.*, 1991; Appleby and Bourret, 1998), whereas CheZ decreases this half-life by 100-fold to approximately 0.14 s (Silversmith *et al.*, 2008). Thus, once the MCP switches conformation back to the inactive state, CheY-P is rapidly cleared from the cytoplasm and the system can return to CCW rotation and smooth swimming. *E. coli* is able to respond to a chemoeffector concentration drop by tumbling within 0.2 s of the stimulus (Segall *et al.*, 1982).

1.2.4 Adaptation

There are three known systems by which MCP signalling is halted and MCPs reset to a less sensitive state: methylation-based adaptation, the CheC-CheD system and the CheY-CheV system.

Methylation-based adaptation

Key glutamate residues on the HCD can be methylated by the methyltransferase CheR (Springer *et al.*, 1979). Increased methylation removes negative charges and leads to closer packing, which makes MCPs more likely to activate CheA signalling at a particular chemoeffector concentration. Demethylation by the methylesterase CheB introduces negative charges into the cluster, interfering with packing and decreasing the probability that the MCPs will activate CheA signalling at a particular chemoeffector concentration (Figure 1.9).

Methylation of the MCPs Thus far, methylation has only been observed in chemoreceptors in the heptad classes 36H (*E. coli*) and 44H (*T. maritima*, *B. subtilis*).

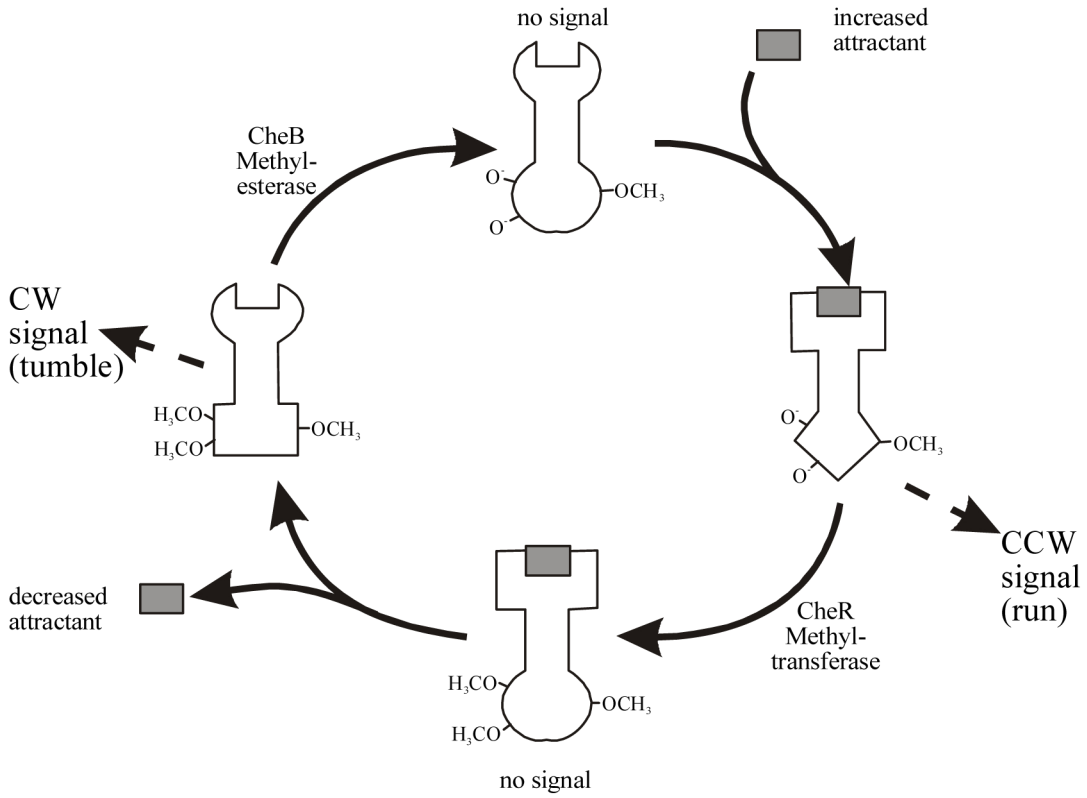


Figure 1.9: The *E. coli* adaptation pathway. Taken from Porter (2002) with permission from the author.

All *E. coli* chemoreceptors have four methylation sites, which on Tsr are Q297, E303, Q311 and E493 (Terwilliger and Koshland, 1984; Kim *et al.*, 1999). Two of the sites are expressed as glutamine residues. These are post-translationally modified to glutamate sites, presumably during the first adaptation response. The receptors are therefore expressed in a state equivalent to half-methylation (two negative charges, two neutral charges). All methylation sites are found in adjacent to another glutamine or glutamate. Methylation is usually observed on the second residue in these **EQ/EQ** pairs. Mutation of these sites to alanine leads to a subtractive loss in chemotaxis ability.

Alexander and Zhulin (2007) predict that methylation can occur in the majority of the heptad classes, in the pattern shown in Figure 1.10.

Preliminary work has been done on *Campylobacter jejuni* an ϵ-proteobacteria with

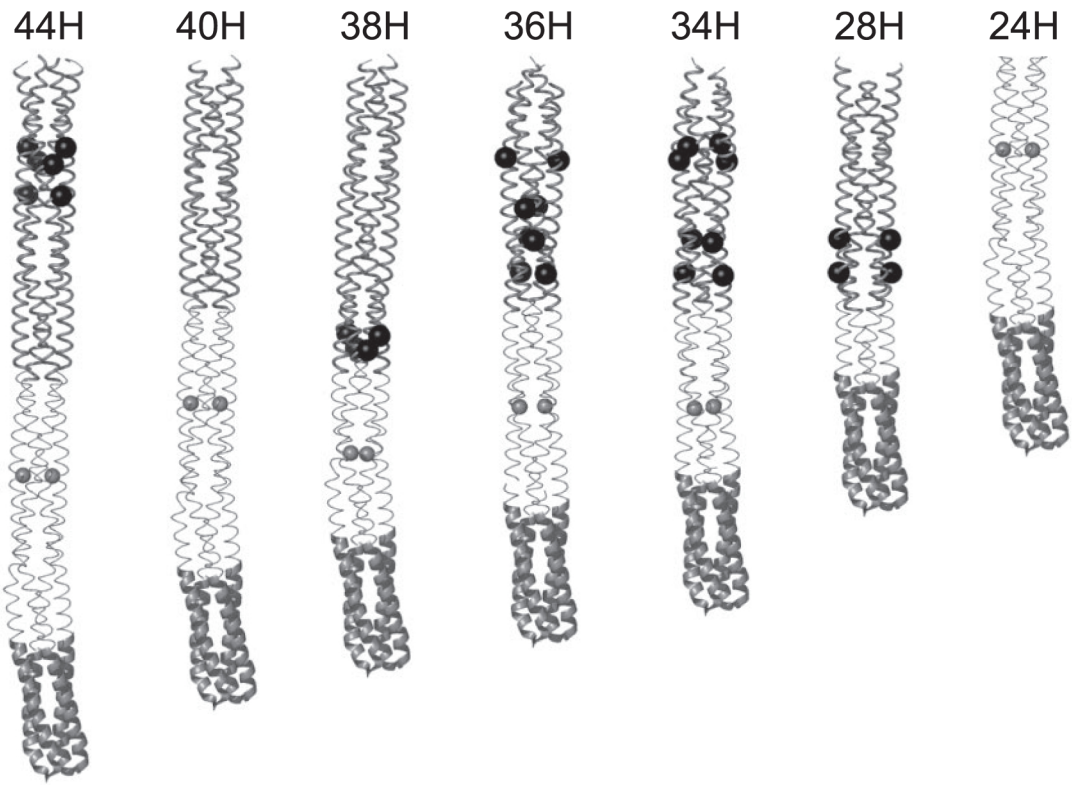


Figure 1.10: Homology models of major MCP CD classes constructed from the *T. maritima* TM1143 structure (class 44H) show positions of the most common methylation sites (dark gray spheres) in each class. The signalling subdomain is shown in dark thick ribbons, the flexible bundle subdomain in light thin ribbons, and the methylation subdomain in dark thin ribbons. The glycine hinge is shown as a light gray sphere. Taken from Alexander and Zhulin (2007) with permission, © 2007 National Academy of Sciences.

ten putative MCPs all classified as 28H by Alexander and Zhulin (2007). Deletion of the putative CheB causes a band shift during electrophoresis of two MCPs studied. It has been suggested that these MCPs are stably methylated in the absence of CheB and presence of CheR (Kanungpean *et al.*, 2011). As yet, the identity of the putative methylation sites has not been discovered, nor is any information available regarding how many sites are expected from experimental work.

CheR CheR is an S-adenosyl-L-methionine : protein-L-glutamate O-methyltransferase. It uses the substrate S-adenosyl-L-methionine (SAM) to methylate glutamate residues on MCP through SN2 nucleophilic substitution (Simms and Subbaramaiah, 1991).

CheR binds to one MCP homodimer at the MCP C-terminal pentapeptide tether, but

methylates adjacent homodimers (Li *et al.*, 1997). The deletion of CheR in *E. coli* results in completely demethylated MCPs. Cells are smooth swimming during steady state. Cells can respond to a repellent stimulus if it is sufficiently large, beginning tumbling, but cannot adapt and recover forward motion again (Goy *et al.*, 1978).

CheB CheB is a response regulator for the histidine kinase CheA, and is itself a methylesterase and deamidase. CheB comprises two domains: an N-terminal regulatory domain that is homologous to CheY, and a C-terminal methylesterase domain (West *et al.*, 1995). Phosphorylation of the regulatory domain causes a conformational change (Anand *et al.*, 2000) that activates CheB, increasing its methylesterase activity (Anand *et al.*, 1998; Anand and Stock, 2002).

CheB is thought to loosely bind to the MCP C-terminus pentapeptide tether in the same manner as CheR. The deletion of CheB in *E. coli* results in overmethylated MCPs. Cells are tumbly during steady state swimming. Cells can respond to an attractant stimulus if it is sufficiently large, beginning smooth swimming, but cannot adapt and recover tumbly movement again (Yonekawa *et al.*, 1983).

The release of methanol from a cell is an indicator for methylation-based adaptation, as methanol is the waste product from CheB methylesterase action.

CheC/CheD adaptation

Concurrent with the methylation-based adaptation system, in some bacteria an alternative adaptation system exists. The CheC-CheD system has primarily been studied in *Bacillus subtilis* although homologues of one or both of the proteins have been found in other species.

CheC is a CheY-P phosphatase, homologous to the CheY-binding domain of FliM (Szurmant *et al.*, 2003, 2004). Deletion of CheC results in poor chemotaxis through a lack of adaptation and does not appear to affect levels of CheY-P (Kirby *et al.*, 2001).

Mutagenesis work has suggested that the deletion phenotype is due to CheC's ability to bind CheY-P, rather than its phosphatase activity (Muff and Ordal, 2007).

CheD is a deamidase, deamidating receptor **Q** residues to **E** for methylation-based adaptation (Kristich and Ordal, 2002; Glekas *et al.*, 2012). It interacts with receptors, increasing the likelihood of receptors activating the associated kinases. CheD also binds CheC, increasing the latter's phosphatase ability. CheD-CheC binding is enhanced by CheC-CheY-P binding. Deletion of CheD results in a tumbly phenotype, due to low levels of CheY-P (Kirby *et al.*, 2001). Mutagenesis work revealed that reducing CheC-CheD binding while maintaining CheC phosphatase activity results in the same lowered chemotaxis ability as the CheC deletion mutant.

The proposed model for CheC-CheD-CheY-P interaction in adaptation is shown in Figure 1.11 (Rao *et al.*, 2008). When attractant binds receptors, CheD interacting with receptors enhances the resulting activation of CheA. The produced CheY-P stimulates a run. Higher levels of CheY-P allow more CheC-CheY-P complexes to form. In turn, these complexes recruit CheD away from the receptors to form CheD-CheC-CheY-P complexes. Without CheD, receptor activation of CheA is decreased, allowing adaptation and a return of tumbling events.

CheV adaptation

CheV is found in a small subset of chemotactic bacteria including some Firmicutes (of which *B. subtilis* is one), β-, δ- and ϵ- (*C. jejuni*) proteobacteria (Alexander *et al.*, 2010). It comprises two domains, a regulatory domain phosphorylatable by CheA and a scaffolding domain homologous to CheW (Rosario *et al.*, 1994). There may be one or many CheV homologues and they may be joined by CheW or alone.

CheV is thought to have both a structural role, taking over from CheW, and an adaptation role. The latter has been studied primarily in *B. subtilis*. Mutation of the regulatory domain so that it is unphosphorylatable renders *B. subtilis* unable to fully

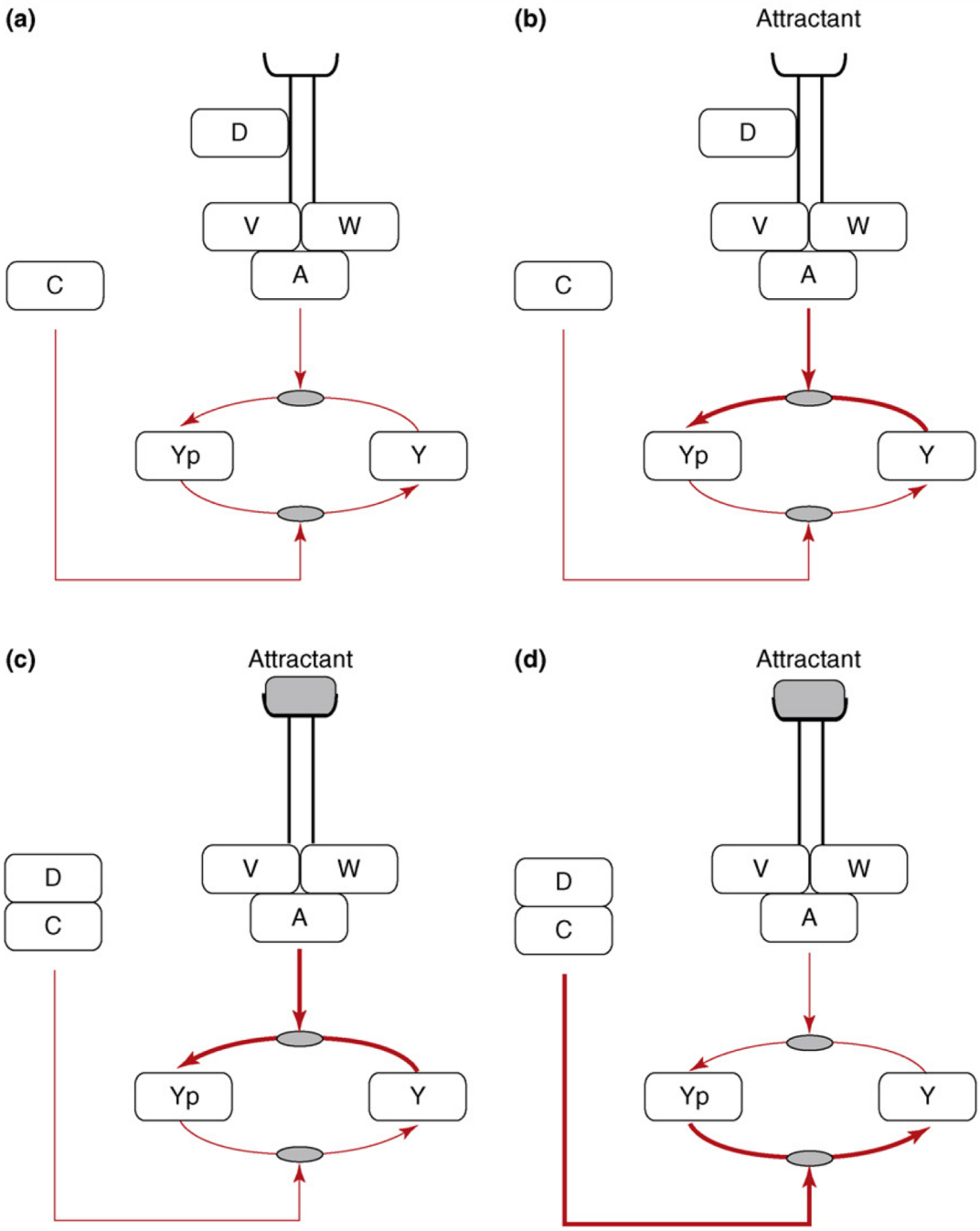


Figure 1.11: Model for the CheC-CheD-CheY-P adaptation system in *B. subtilis*. (a) Prior to attractant binding, CheD is bound to receptors. (b) When attractant binds to receptors, CheA kinase is activated and CheY is phosphorylated. (c) CheC binds CheY-P. CheC-CheY-P complexes attract CheD away from the receptors. (d) Receptors unbound with CheD only weakly activate the CheA kinase, causing less CheY-P to be formed (adaptation). Abbreviations: A, CheA; C, CheC; D, CheD; V, CheV; W, CheW; Y, CheY; Yp, CheY-P. Taken from Rao *et al.* (2008) with permission from Elsevier, license number 3692711391532.

adapt after response to a signal, in a similar manner as to the CheC deletion mutant (Karatan *et al.*, 2001). It is thus thought the CheV forms another negative feedback loop. When CheA is activated, CheV is phosphorylated. CheV-P then affects the structural coupling between CheA and the receptor, reducing CheA activation (Rao *et al.*, 2008)

1.2.5 Complexity in bacterial chemotaxis

The basic chemotaxis system described above, using the full sensing and signalling systems, and only the methylation based adaptation system, is the paradigm of chemotaxis extensively studied in the model organisms of *E. coli* and *Salmonella* Typhimurium. However, some bacteria start with this basic system and complicate it. In many cases, multiple homologues of the chemotaxis proteins are found.

Multiple adaptation systems in concert (*Bacillus subtilis*)

The kinase - response regulator chain of events is reversed in *B. subtilis*. Activation of the kinase is achieved by attractant binding. Resulting CheY-P binds to the motor causing CCW rotation, or runs, while CheY biases the motor towards tumbling.

B. subtilis uses all three distinct adaptation systems: the CheB-CheR methylation-based adaptation system, CheC-CheD system and CheV system (Rao *et al.*, 2008). The systems are to an extent redundant, as adaptation is possible with any two present. Perfect adaptation is however only possible with all three systems. It has been proposed that the CheC-CheD and CheV systems facilitate adaptation at low attractant levels, while the CheB-CheR system facilitates adaptation at high attractant levels.

B. subtilis contains multiple MCPs, the attractants for some of which have been elucidated. The standard receptor for study is McpB, the asparagine receptor. Three

methylation sites have been confirmed on McpB: Q371 (deamidated to E371) on the N-terminus arm, and E630 and E637 on the C-terminus arm. Homology modelling of McpB against the crystal structure of the *T. maritima* chemoreceptor TM1143 suggests that the three sites lie closely together (Figure 1.12).

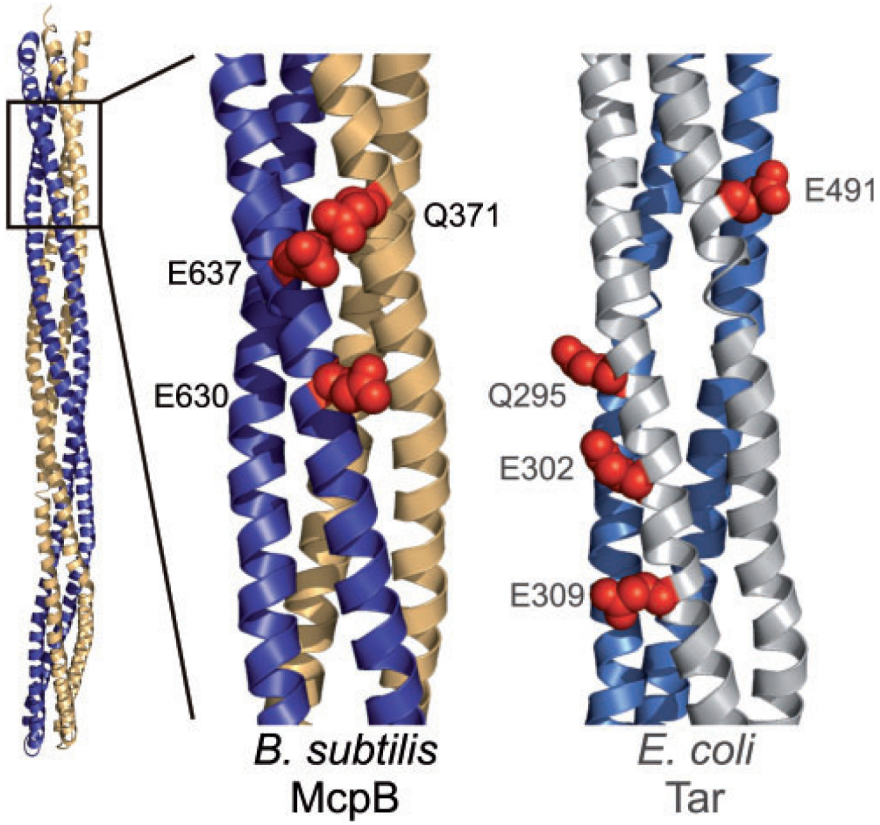


Figure 1.12: Comparison of adaptation sites (red spheres) on *B. subtilis* McpB and *E. coli* Tar dimerised cytoplasmic domains. The McpB homology model is based on the crystal structure of the cytoplasmic domain of the *T. maritima* TM1143 chemoreceptor (Park *et al.*, 2006). McpB adaptation sites, Gln371, Glu630 and Glu637, form a closely spaced, interacting triad on McpB. Tar (Kim *et al.*, 1999) adaptation sites are positioned in a line and are spaced further apart. The adaptation sites are only shown on one face of each receptor. Taken from Glekas *et al.* (2011), with permission from the Society for General Microbiology.

In *E. coli*, methylation of any of the key glutamate residues increases CheA activity. However, the three *B. subtilis* McpB sites are methylated under different conditions and appear to affect McpB's sensitivity towards attractant concentration and the resulting kinase activity differently, although the precise mechanisms are as yet unknown.

Q371 is thought to control receptor sensitivity (Glekas *et al.*, 2011). It is demethylated on attractant addition. The amidated state, analogous to the methylated state, increases McpB's affinity for asparagine as well as increases kinase activity. Thus, methylated E371 allows McpB to bind asparagine and signal through the kinase for a run. E371 is then demethylated to allow adaptation to the new asparagine concentration and allow tumbling to occur again. At lower E371 methylation states, McpB's affinity for asparagine drops. This allows changes in concentration to be sensed at a higher absolute concentration, and allows for relative changes to be sensed over a wide range of concentrations.

E630 and E637 are thought to control kinase activity (Glekas *et al.*, 2011). Addition and removal of attractant causes demethylation of E630, whereas only removal of attractant causes demethylation of E637. The methylation state of these sites does not affect McpB's affinity for asparagine. Amidation as a mimic for methylation at either or both sites results in decreased kinase activity. These sites form a negatively charged patch on McpB. It is possible that neutralising this charge through methylation allows for a conformational change in the McpB packing similar to that seen in *E. coli*, decreasing kinase activity. Thus, when attractant is removed, methylated E630 and E637 amplify the conformational changes leading to complete kinase inactivity. This causes a tumble. E630 and E637 are then demethylated to allow adaptation and a return to run to move to a new area.

If attractant levels rise over a long period, the methylation state of E630 and E637 will rise, causing kinase activity to slowly drop. This works alongside the lowering methylation state of E371 to allow changes in attractant concentration at the high end of the gradient to still be sensed.

A note of caution can be derived from the work on *B. subtilis*. Initial mutagenesis studies mutated sites of interest to aspartate in order to mimic the demethylated state. This has been shown to be valid in work on *E. coli*. However, the results of these early studies were contradicted by later work using the more physiologi-

cally relevant amidation and deamidation states in a *cheBcheR* deletion background. This suggests that where putative methylation sites lie clustered very closely to one another, rather than spaced further apart as in *E. coli* the small change in length between a **D** and **E** sidechain is sufficient to cause conformational changes in the receptor, resulting in non-physiologically relevant results.

Multiple chemotaxis clusters (*R. sphaeroides*)

The arrangement of the chemotaxis genes in operons is instrumental in the robustness of the chemotaxis pathway, buffering the pathway against gene expression noise. Chemotaxis efficiency is sensitive to changes in the expression of a single gene and efficiency dramatically improves when expression of just two genes is linked. Gene order within the operon also affects robustness, with the *E. coli* gene order offering strong robustness to translational noise (Lvdok *et al.*, 2007, 2009).

In a study of 206 bacterial species containing the basic chemotaxis system (CheA, CheB, CheR, CheW and CheY), more than 50% of all sequenced bacterial genomes have multiple homologues of at least one of the chemotaxis proteins and 30% have multiple homologues of the entire set. These multiple homologues tend to be arranged in operons as in *E. coli* and the order of the gene pairs *cheA-cheW* and *cheB-cheR* is strongly conserved (Hamer *et al.*, 2010).

In some cases, these duplicate systems have been co-opted to some other role than chemotaxis. For example, *Rhodospirillum centrum* has three chemosensory pathways. One controls chemotaxis, as in the canonical system, one controls flagellar expression and the third controls cyst formation (Berleman and Bauer, 2005a,b). However, in at least one organism, the duplicate systems are all used in chemotaxis.

R. sphaeroides has three complete operons of chemotaxis genes, two of which are expressed under laboratory conditions. Both of these expressed operons are required for chemotaxis to function and the products of the two operons localise to two dis-

tinct regions in the cell. As *R. sphaeroides* is a model organism for photosynthesis, it is well-represented in the literature and techniques for culturing and genetic manipulation have been developed. It therefore makes a useful model for studying more complex chemotaxis systems. In the next sections, *R. sphaeroides* will be introduced and the specifics of what is known about its chemotaxis system will be presented.

1.3 *Rhodobacter sphaeroides*

Rhodobacter sphaeroides, previously known as *Rhodopseudomonas sphaeroides*, is of the class α-proteobacteria. The proteobacteria are Gram-negative and share a photosynthetic ancestor. *R. sphaeroides* falls within the order *Rhodobacterales* and the family *Rhodobacteraceae*, due to the presence of green and red photosynthetic pigments and the lack of sulphur droplets or use of sulphur-based compounds as a final electron acceptor in photosynthesis (Molisch, 1907).

Originally known as the family *Athiorhodaceae*, the purple non-sulphur bacteria were first fully characterised in van Niel (1944). These bacteria are motile, do not form endospores and are usually found in aqueous environments, most notably "stagnant ponds" (van Niel, 1944).

1.3.1 Strains used

Most of the work on *R. sphaeroides* has focused on the strain 2.4.1, the strain first identified by van Niel (1944). The complete genome for this strain is available (Mackenzie *et al.*, 2001).

However, the work done on motility and chemotaxis has focused on strain WS8N, as this strain is more consistently motile strain. The parent strain WS8, first designated TS/6 (Sistrom, 1977), was originally isolated in Ithica, New York, in 1969 by Clayton and Clayton (1972). WS8N is a spontaneous mutant of the parent species

that displays nalidixic acid resistance (Sockett *et al.*, 1990). The complete genome is now available for strain WS8N in Porter *et al.* (2011).

1.3.2 Genetic structure

The genetic material of *R. sphaeroides* is arranged across two chromosomes (3.14 and 0.97 Mbp) and a number of plasmids, depending on the strain. WS8N contains two large plasmids (200 and 110 kbp), whereas 2.4.1 contains five. There are two theories for the existence of two chromosomes of different sizes. Chromosome II may be the remnant of a plasmid that has accumulated genes from the main chromosome over time. The low density of functional genes and the plasmid-like origin of replication support this theory (Mackenzie *et al.*, 2007). *R. sphaeroides* may have had a single large chromosome, which then split into two unevenly-sized chromosomes. The dispersal of essential genes across both chromosomes and the age of gene duplications found in both chromosomes support this theory (Mackenzie *et al.*, 2001; Choudhary *et al.*, 2004).

Recently, it has been proposed that the second, smaller, permanent DNA structure in some bacterial species (including *R. sphaeroides* Chromosome II) is not a chromosome or a plasmid, but an intermediate structure termed a chromid (Harrison *et al.*, 2010). These structures are thought to be a hallmark of the evolution of a new genus, as they contain essential 'lifestyle' genes, such as metabolism or motility genes. They may arise through either recombination between a chromosome and a plasmid or by recombination between an existing chromid and a plasmid.

1.3.3 Metabolism

The purple non-sulfur photosynthetic bacteria are among the metabolically most diverse species known. *R. sphaeroides* is able to grow in the light or in the dark, in

the presence or absence of oxygen, by photosynthesis, respiration or fermentation (Tabita, 1995).

In the absence of light and the presence of oxygen, *R. sphaeroides* functions as either a chemoheterotroph using reduced carbon compounds as the carbon source and reducing power in aerobic respiration. In the absence of both light and oxygen, fermentation or anaerobic respiration is used for growth. A wide variety of organic compounds can be oxidised as a source of reducing power and carbon in respiration and fermentation, including organic fatty acids, polyols and sugars, as well as carbon dioxide. A number of both inorganic and organic compounds can be reduced as the final electron acceptor in anaerobic respiration, including sulfate, nitrate, toxic metal oxides/oxyanions and thymine.

In the presence of light and absence of oxygen, photosynthesis is used. Again, depending on the availability of different carbon sources, *R. sphaeroides* either functions photoautotrophically, using carbon dioxide as the carbon source and electron acceptor and hydrogen as the electron donor, or photoheterotrophically, using organic compounds such as methanol, acetate, pyruvate or butyrate as the electron donor and/or carbon source and CO_2 as the electron acceptor.

1.3.4 Motility

As with the majority of the purple non-sulphur bacteria, *R. sphaeroides* is motile. It has two sets of flagellar genes, *fla1* and *fla2*. Expressed under laboratory conditions, *fla1* codes for a single, randomly-positioned, unidirectional flagellum. This is typical of the family (van Niel, 1944). The *fla2* operon codes for a tuft of polar flagella. As yet, this operon has only been expressed through mutation work and no laboratory conditions have been found under which endogenous expression takes place (del Campo *et al.*, 2007; Poggio *et al.*, 2007). Only the *fla1* system will be discussed in this thesis.

R. sphaeroides has a different motility pattern from *E. coli* and other species. The flagellar motor alternates between on and off to produce active forward motion and passive reorientation by Brownian motion respectively (Armitage and Macnab, 1987). When the motor is on, the flagellum has a helical conformation similar to that of *E. coli*, but when the motor is off, the flagellum relaxes into a highly coiled form which may enhance reorientation (Armitage *et al.*, 1999).

In a homogenous environment, *R. sphaeroides* reaches averages speeds of approximately 40 μm/s and the flagellar motor a rotational speed of approximately 100 Hz (Armitage and Macnab, 1987; Brown, 2009). Transient stops for direction change occur with a mean frequency of 0.31 ± 0.19 stops per second. Rotation restarts after a mean 0.66 s and a median 0.44 s (range: 0.01-6 s). *R. sphaeroides* has a high run bias, spending 0.80 ± 0.20 of its time in runs during steady state conditions (Brown, 2009). In comparison, *E. coli* has a run bias of 0.6-0.7 (Block *et al.*, 1983).

1.4 *Rhodobacter sphaeroides* chemotaxis

R. sphaeroides demonstrates chemotaxis towards a wide range of chemoeffectors, including its preferred nutrients for growth, propionate and succinate. A drop in one of these attractants from 100 μM to 0 μM results in a responsive stop of approximately 40 s in wildtype cells grown aerobically and 70 s in wildtype cells grown photoheterotrophically (Kojadinovic *et al.*, 2011). The only known repellent is oxygen during anaerobic growth conditions.

R. sphaeroides contains multiple homologues of all the basic chemotaxis proteins, arranged in three operons and multiple orphan genes (Figure 1.13). Deletion of the components of *che*Op$_1$ or the full operon has little effect on chemotaxis (Ward *et al.*, 1995; Hamblin *et al.*, 1997) and it is assumed that these proteins are not expressed under laboratory conditions. When expressed through mutation work, the proteins in *che*Op$_1$ have been shown to control rotation of the *fla2* flagellar system (del Campo

et al., 2007, 2011). The proposed model for this control is a modified version of the *E. coli* paradigm: a polar cluster of MCPs and a single homologue of CheW ($CheW_1$) and CheA ($CheA_1$) activates three CheY homologues. $CheY_2$ interacts directly with the motor to effect stops, $CheY_4$ is thought to act as a phosphate sink, as in *S. meliloti* and the role of $CheY_1$ is as yet unknown. The polar cluster is thought to be subject to methylation-based adaptation, through a CheR homologue ($CheR_1$) and a CheD homologue, as there is no CheB homologue in the cluster. No work on this operon or cluster was done in this thesis, as only endogenously expressed chemotaxis proteins were examined.

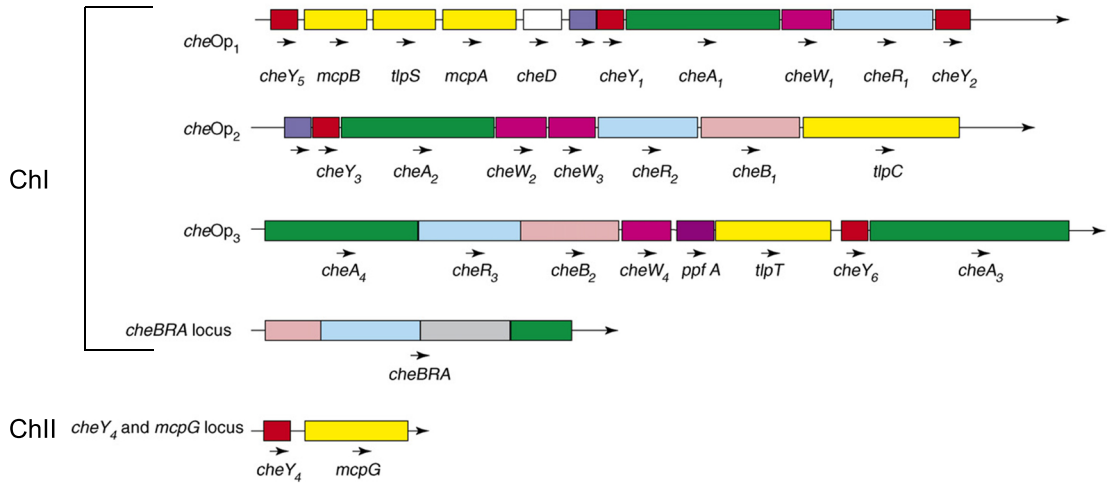


Figure 1.13: The chemotaxis loci of *R. sphaeroides* chemotaxis. Ch=chromosome. Adapted from Porter *et al.* (2008b), with permission from Elsevier, license number 3692710333670.

The deletion of *che*Op_2 (Shah *et al.*, 2000b) or *che*Op_3 (Porter *et al.*, 2002) results in loss of chemotaxis. The rest of this section deals with the components of these two expressed operons.

1.4.1 Receptors

There are thirteen putative chemoreceptors, nine of which appear to be classical MCPs (McpA, McpB, McpE, McpG, McpH, McpJ, McpM, McpR and McpV) and four of which are Tlps without a transmembrane region (TlpC, TlpL, TlpS and TlpT).

Of these, seven of the MCPs and three of the Tlps have been found to be expressed under laboratory conditions. The remaining receptors (McpA, McpB, McpM and TlpS) are assumed to associate with the *che*Op$_1$ proteins.

The receptors do not form a single polar cluster as in *E. coli*. Instead, fluorescent localisation studies on some MCPs indicate that the chemoreceptors cluster in two distinct regions in the cell. The MCPs tested form a classical cluster at one pole, while the Tlps tested form separate clusters within the cytoplasm on the surface of the chromosomes (Wadhams *et al.*, 2000, 2002, 2003).

Unlike in *E. coli*, the chemoreceptors studied do not appear to each sense only one attractant. Instead, deletion studies show that when deletion of a chemoreceptor causes a reduction in chemotaxis, this reduction is seen with a wide range of attractants (Thompson, 2005).

Of the *R. sphaeroides* chemoreceptors that were included in the Alexander and Zhulin (2007) data set, all MCPs were classified as 34H receptors and the single Tlp, TlpT, was classified as a 36H receptor. Classification of the *R. sphaeroides* chemoreceptors is discussed in more detail in Chapter 4.

1.4.2 Clustering of the *R. sphaeroides* chemotaxis proteins

Chemotaxis operons two and three are expressed under laboratory conditions. In general, operon two proteins localise with the MCPs at the pole, while operon three proteins localise with the Tlps in the cytoplasm (Figure 1.14). The exceptions are that the CheB proteins do not localise but are diffuse throughout the cytoplasm, while the CheR proteins are mostly localised but do show some diffusion (Wadhams *et al.*, 2003). Each cluster contains at least one homologue of each of the chemotaxis proteins seen in *E. coli*: A CheA, CheW, CheY, CheR and CheB.

The polar cluster contains two homologues of CheW, CheW$_2$ and CheW$_3$. Both are required for formation of the cluster, as deletion of either results in either abolished

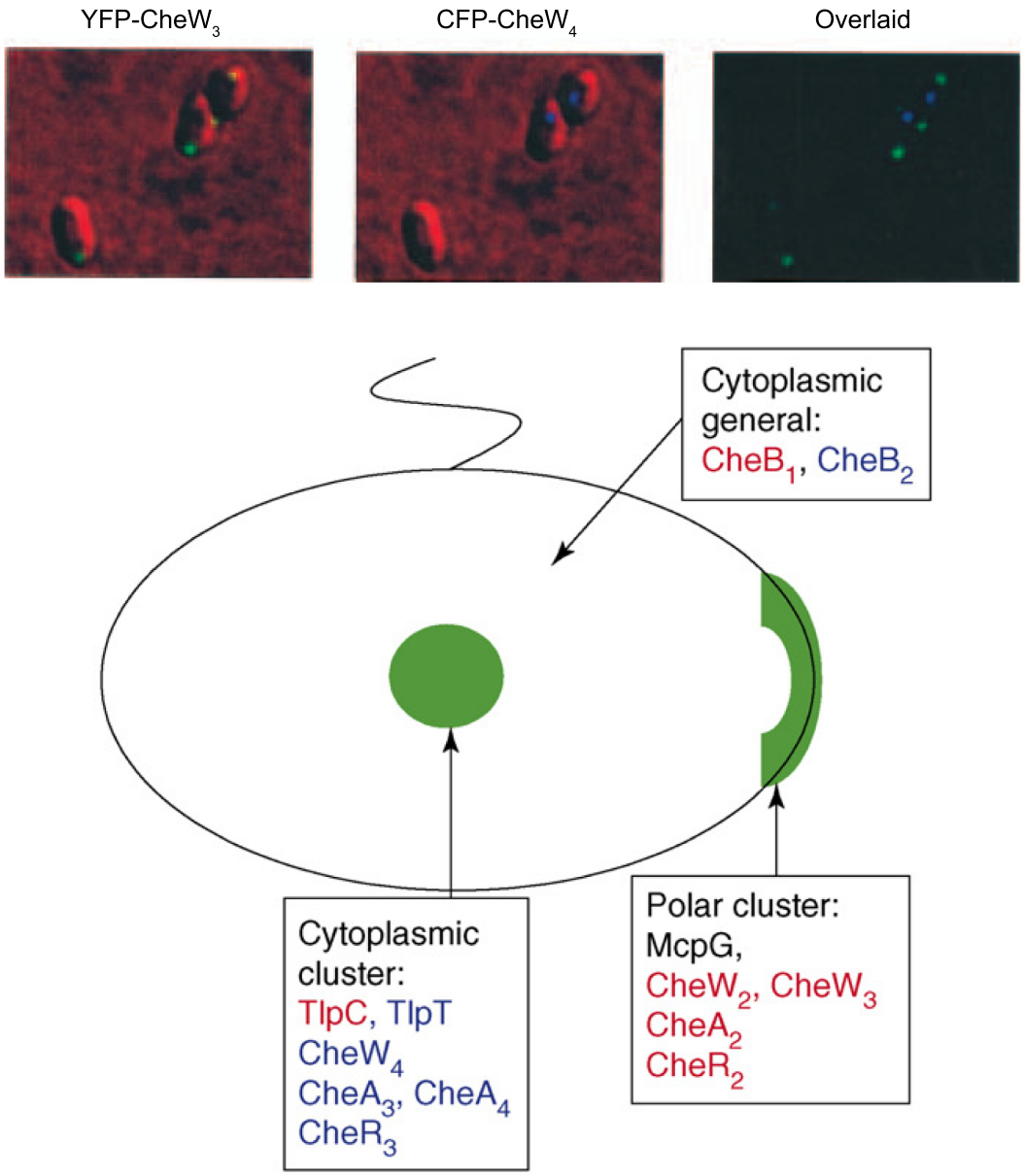


Figure 1.14: Localisation of the *R. sphaeroides* chemotaxis proteins to two distinct regions in the cell. (top) Fluorescent tagging shows localisation of $CheW_3$ to the cell pole whereas $CheW_4$ forms clusters in the cytoplasm. These images are representative of fluorescent tagging of all chemotaxis proteins. (bottom) Summarising *R. sphaeroides* chemotaxis protein localisation based on fluorescent tagging. Taken from (Wadhams *et al.*, 2003), with permission from John Wiley and Sons, license number 3692931072781.

($CheW_2$) or reduced ($CheW_3$) recruitment of $CheA_2$ and the other CheW to the cluster (Wadhams *et al.*, 2005). Why two CheWs are required in this cluster is unknown.

Clustering in the cytoplasm is dependent on the presence of both TlpT and the single CheW in this cluster, $CheW_4$. TlpC is required to a lesser extent for cluster formation, as its deletion leads to less condensed clusters and more clusters. The cytoplasmic cluster contains two truncated homologues of CheA, $CheA_3$ and $CheA_4$. These CheA proteins localise to the cluster independently of one another and the presence of at least one of the pair is required for cluster formation (Wadhams *et al.*, 2005).

1.4.3 Known interactions and signalling

Both clusters are required for normal functioning of chemotaxis and the deletion of either results in the loss of chemotaxis (Porter *et al.*, 2002). The requirement for each chemotaxis protein has been tested through numerous deletion studies and are summarised in Table 1.1. The known interactions between these proteins are summarised in Figure 1.15.

Deletion of either TlpT or TlpC results in both loss of the cytoplasmic cluster and abolition of chemotaxis (Porter *et al.*, 2002; Wadhams *et al.*, 2002, 2005). The latter may be due solely to the former, or may also be due to some function of these chemoreceptors. Similarly, deletion of only McpH results in abolition of chemotaxis, rather than the decrease seen with the other MCPs tested (Thompson, 2005).

The polar cluster is thought to sense chemoeffector concentration drops and signals the flagellar motor via CheY homologues in the same sequence as the classical *E. coli* cluster. $CheA_2$ is able to phosphorylate the response regulators $CheY_3$, $CheY_4$, $CheY_6$, $CheB_1$ and $CheB_2$ (Figure 1.16, top) (Porter and Armitage, 2002).

The cytoplasmic cluster has two truncated CheA homologues (Figure 1.16, bottom). $CheA_4$ is missing the phosphorylatable domain and the response regulator binding site, but is able to dimerise and contains the regulatory and kinase domains required

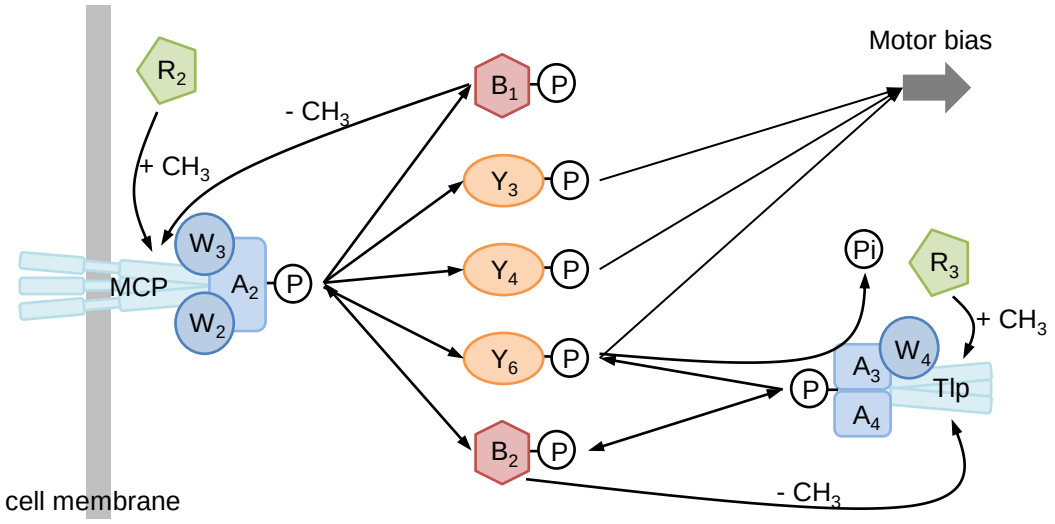


Figure 1.15: Current model for *R. sphaeroides* chemotaxis. **Sensing** The MCPs in the polar cluster sense external chemoeffector concentrations while the Tlps in the cytoplasmic cluster sense the internal energy state of the cell. $CheR_2$ and $CheR_3$ sensitise the MCPs and Tlps respectively through methylation. When chemoeffector levels drop, the chemoreceptors change conformation and activate their associating proteins. **Signalling** $CheA_2$ in the polar cluster autophosphorylates, then phosphorylates $CheY_6$. $CheA_4$ in the cytoplasmic cluster phosphorylates $CheA_3$, which then phosphorylates $CheY_6$. $CheY_6$-P binds the motor and causes a stop. **Adaptation** $CheA_2$-P then phosphorylates the methylesterases $CheB_1$ and $CheB_2$, while $CheA_3$-P phosphorylates just $CheB_2$. The methylesterases desensitise the chemoreceptors, halting their signalling. The specificity here is unknown. $CheA_2$ also phosphorylates $CheY_3$ and $CheY_4$, which either function as a phosphate sink or interfere with $CheY_6$-P binding at the motor. **Signal termination** $CheY_6$-P autodephosphorylates and is dephosphorylated by $CheA_3$'s phosphatase domain. **Reverse phosphorelay** The cytoplasmic cluster can activate the polar cluster through $CheB_2$, as $CheB_2$-P can phosphorylate $CheA_2$. Evidence for this model is included in the text. Adapted from Scott *et al.* (2012) with permission from the authors and Caister Academic Press.

Table 1.1: Effects of deleting expressed *R. sphaeroides* chemotaxis proteins.

Protein	Chemotaxis phenotype when deleted	Study	Complements *E. coli* deletion	Study
Polar cluster				
$CheA_2$	abolished	Hamblin *et al.* (1997)	partial	Shah *et al.* (2000a)
$CheW_2$	abolished	Martin *et al.* (2001a)	no	Shah *et al.* (2000a)
$CheW_3$	abolished	Martin *et al.* (2001a)	partial	Shah *et al.* (2000a)
$CheR_2$	abolished	Martin *et al.* (2001b)	partial	Martin *et al.* (2001b)
Cytoplasmic cluster				
$CheA_3$	abolished	Porter *et al.* (2002)	no	Porter *et al.* (2002)
$CheA_4$	abolished	Porter *et al.* (2002)	no	Porter *et al.* (2002)
$CheW_4$	reduced	Porter *et al.* (2002)	partial	Porter *et al.* (2002)
$CheR_3$	abolished	Porter *et al.* (2002)	no	Porter *et al.* (2002)
Diffuse				
$CheB_1$	abolished	Martin *et al.* (2001b)	yes	Martin *et al.* (2001b)
$CheB_2$	reduced	Porter *et al.* (2002)	partial	Porter *et al.* (2002)
$CheY_3$	none	Porter *et al.* (2006)	no	Shah *et al.* (2000a)
$CheY_4$	none	Porter *et al.* (2006)	no	Shah *et al.* (2000a)
$CheY_6$	abolished	Porter *et al.* (2002)	no	Porter *et al.* (2002)
$CheY_3CheY_4$ double deletion abolishes chemotaxis (Porter *et al.*, 2006) all CheYs complement CheZ deletion in *E. coli* (Shah *et al.*, 2000a)				

for activation and phosphorylation. $CheA_3$ can be phosphorylated, but its dimerisation pattern is unknown. The two CheA proteins function as a single CheA together: $CheA_4$ is activated and transfers a phosphoryl group through $CheA_3$ to a response regulator. This cluster is able to phosphorylate only the response regulators $CheB_2$ and $CheY_6$ (Porter and Armitage, 2002).

$CheY_6$ and one of $CheY_3$ or $CheY_4$ are required for chemotaxis (Porter *et al.*, 2006). All CheYs have been observed binding to FliM *in vitro* and this binding is enhanced by phosphorylation (Ferre *et al.*, 2004), although only $CheY_6$-P binding can stop the motor. The role of $CheY_3$ and $CheY_4$ is as yet unclear. It is possible that $CheY_3$-P or $CheY_4$-P binding to FliM blocks $CheY_6$-P from binding to FliM. The multiple CheYs may also function as phosphate sinks, as in *S. meliloti*. This hypothesis was tested through mathematical modelling, which will be discussed in Section 1.4.5 (Tindall *et al.*, 2010).

In general, deletion mutants and other mutations of the chemotaxis proteins that

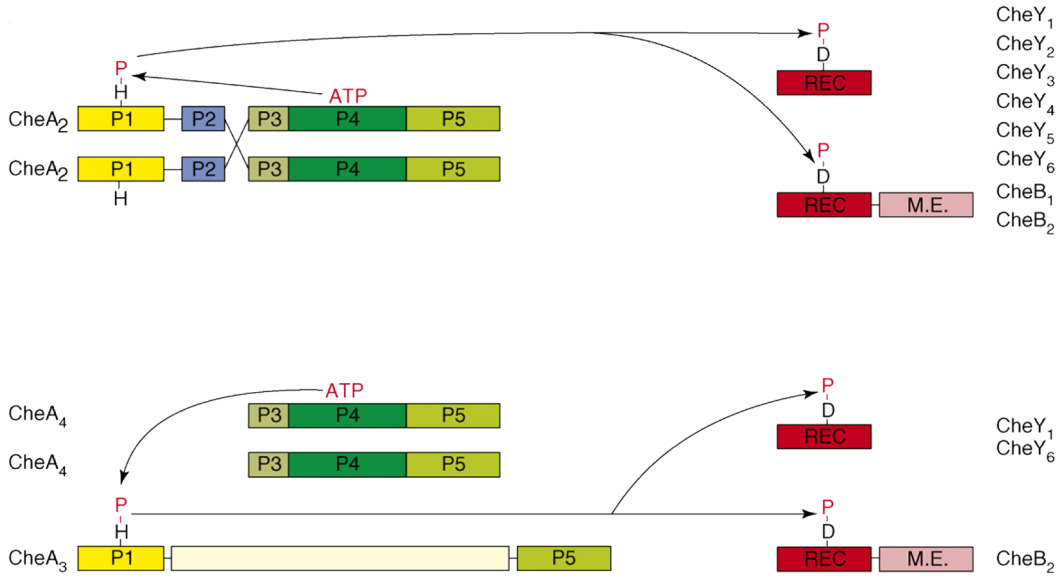


Figure 1.16: Structure of *R. sphaeroides* $CheA_2$ (top) and $CheA_3$and $CheA_4$ (bottom) and their known response regulator partners. P=domains in CheA proteins, REC=response regulator domain, ME=methylesterase domain in CheB proteins. Taken from Porter *et al.* (2008b), with permission from Elsevier, license number 3692710333670.

affect chemotaxis without affecting the flagellar architecture do not affect motility, but only chemotaxis. However, a flagellate nonmotile phenotype can be generated through the D56N mutation of $CheY_6$. As in *E. coli*,(Appleby and Bourret, 1999), $CheA_3A_4$ irreversibly phosphorylates $CheY_6$D56N on an alternative site. $CheY_6$D56N-P binds to FliM to produce a permanently stopped motor (Porter *et al.*, 2006).

It is interesting to note that the mutations in *E. coli* CheY that result in permanently activated CheY result in either wildtype or deletion phenotypes in *R. sphaeroides* CheY proteins (Porter *et al.*, 2006), as CheY is structurally similar to the regulatory domain of CheB.

1.4.4 Signal termination in *R. sphaeroides*

R. sphaeroides does not have a homologue of the *E. coli* CheZ or the *B. subtilis* CheC, which both dephosphorylate CheY in order to end a chemotaxis signal. $CheY_6$-P

autodephosphorylates much faster than *E. coli* CheY-P autodephosphorylation, with a half life of 4 s compared to 14 s. The truncated kinase $CheA_3$ acts as a phosphatase of $CheY_6$-P, although it only accelerates dephosphorylation three-fold, compared to CheZ's 100-fold increase (Porter *et al.*, 2008a).

There are no known phosphatases acting towards $CheY_3$-P and $CheY_4$-P, which have half lives from autodephosphorylation of 36 ± 3 s and 38 ± 3 s respectively (Porter *et al.*, 2008a).

R. sphaeroides responds to a stimulus in approximately 1 s (Berry and Armitage, 2000). Only $CheY_6$-P therefore dephosphorylates quickly enough to likely effect the chemotaxis response.

1.4.5 Adaptation in *R. sphaeroides*

There is as yet no direct evidence of methylated chemoreceptors in *R. sphaeroides*. However, the ability of *R. sphaeroides* cells to respond to drops in the chemoeffector concentration by stopping flagellar motility, then restart movement within a short timescale, shows that some kind of adaptation is occurring. The lack of a CheC, CheD or CheV homologue expressed under laboratory conditions suggests that the alternative adaptation systems used by *B. subtilis* are not used here.

There is evidence suggesting that methylation-based adaptation could be possible in *R. sphaeroides*. Each chemotaxis protein cluster contains the basic requirements for methylation-based adaptation – receptors from MCP class 34H or 36H, a putative methylesterase and a putative methyltransferase – and methanol release has been measured on changes in chemoeffector in the environment.

Putative methyltransferases

$CheR_2$ in the polar cluster is able to complement *E. coli* CheR deletion (Martin *et al.*, 2001b). Deletion of $CheR_2$ results in a loss of chemotaxis, as cells lose the ability to respond to drops in the chemoeffector concentration. As yet, a C-terminal pentapeptide tether has not been identified within the polar cluster receptors. However, as $CheR_2$ localises with this cluster, a tether is likely to exist.

$CheR_3$ localises to the cytoplasmic cluster. Deletion of the final five amino acids in TlpT (GDGDF) results in $CheR_3$ delocalisation (Mark Roberts, unpublished). The deletion of $CheR_3$ results in a loss of chemotaxis as cells lose the ability to respond to chemoeffector changes. However, $CheR_3$ does not complement an *E. coli* CheR deletion (Porter *et al.*, 2002).

Putative methylesterases

As the two *R. sphaeroides* CheB proteins ($CheB_1$ and $CheB_2$) are not localised but are diffuse through the cytoplasm, they are available to both chemotaxis protein clusters. The polar cluster's kinase $CheA_2$ has been shown to phosphorylate both $CheB_1$ and $CheB_2$ (Porter and Armitage, 2002). However, the cytoplasmic cluster's kinase combination $CheA_3CheA_4$ only phosphorylates $CheB_2$.

$CheB_1$ is able to complement an *E. coli* CheB deletion (Martin *et al.*, 2001b). The deletion of $CheB_1$ results in a loss of chemotaxis, as cells lose the ability to respond to drops in the chemoeffector concentration. $CheB_2$ partially complements an *E. coli* CheB deletion (Porter *et al.*, 2002). The deletion of $CheB_2$ results in a decrease in chemotaxis, as cells still respond to drops in the chemoeffector concentration but are very slow to adapt (Porter *et al.*, 2002). These different effects suggest differing roles for the two CheB proteins in adaptation.

An additional role has been proposed for $CheB_2$. Mathematical modelling of the in-

teractions between the histidine kinases and response regulators in the *R. sphaeroides* chemotaxis system found that levels of $CheY_3$-P, $CheY_4$-P and $CheB_1$-P did not fall to zero when $CheA_2$ś ability to autophosphorylate was knocked out, as in Figure 1.17C. This suggested the existence of a reverse phosphorelay from $CheA_3$ through $CheB_2$ to $CheA_2$, following the pathways in Figure 1.17A and B (Tindall *et al.*, 2010). This would allow the cytoplasmic cluster to activate the polar cluster and thus indirectly activate all available response regulators.

It has been shown the $CheB_2$ can phosphorylate $CheA_2$ *in vitro* (Porter and Armitage, 2002), but as yet this has not been shown *in vivo*. However, the model in Tindall *et al.* (2010) only included phosphorylation reactions between the *R. sphaeroides* CheA, CheB and CheY homologues. The sensory and adaptation pathways were not included. The reverse phosphorelay may be an artefact of using this truncated system. The key protein, $CheB_2$, may have a role in adaptation and it is possible that the kinetics of this reaction would make the proposed reverse phosphorelay physiologically irrelevant.

Methanol release in *R. sphaeroides*

In wildtype *R. sphaeroides*, methanol is released when an attractant is added. A much smaller spike of methanol release is seen when some attractants are removed. Methanol release is lost with deletion of $CheB_1$, $CheR_2$, $CheA_2$, $CheW_2$ or $CheW_3$ (Martin *et al.*, 2001b; Mantotta, 2002). The polar cluster must therefore be intact and able to signal for methanol release. That the two adaptation proteins are also required strongly suggests that the polar cluster is involved in methylation-based adaptation.

Only the effect of deleting one Tlp on methanol release has been tested, TlpC. The deletion of TlpC gives a non-chemotactic phenotype, but leaves the methanol release pattern intact (Wadhams *et al.*, 2002). As the cytoplasmic cluster requires TlpC to be

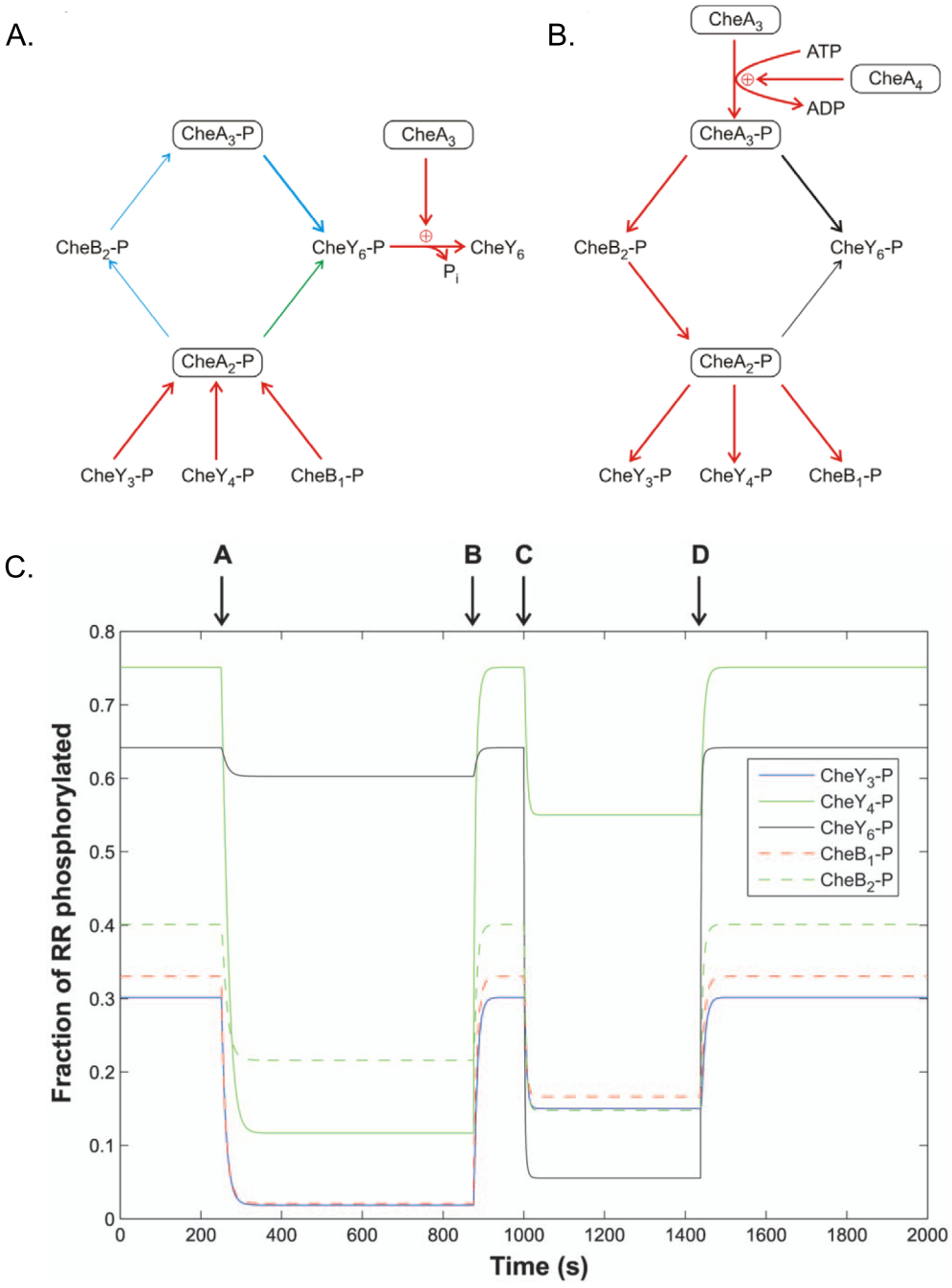


Figure 1.17: Modelling the *R. sphaeroides* chemotaxis pathway predicts the existence of phosphorelays between the two clusters. (A) $CheY_6$ as a phosphate-sink, carrying phosphoryl groups away from the polar cluster. $CheY_6$-P is rapidly dephosphorylated by the cytoplasmic cluster. Arrows show the direction of phosphorylation. (B) Reverse phosphorelay (red reactions) between the cytoplasmic cluster and polar cluster, via $CheB_2$, allows the cytoplasmic cluster to indirectly activate its non-cognate RRs, $CheY_3$, $CheY_4$ and $CheB_1$. (C) The predicted levels of RR-P during a simulated chemotaxis response. When $CheA_2$ autophosphorylation is turned off (A, mimicking chemoeffector drop), the RR-P levels drop. When $CheA_2$ autophosphorylation is turned back on (B, mimicking adaptation), the system returns to its original steady state. Similarly, phosphorylation of $CheA_3$ by $CheA_4$ is turned off at C and turned on again at D. Taken from Tindall *et al.* (2010), published under a CC-BY license.

structurally and functionally stable, this strongly suggests that methylation-based adaptation does not occur here.

Although methanol release is key evidence of methylation-based adaptation, methanol metabolism in *R. sphaeroides* may complicate these results. *R. sphaeroides* can use methanol as a reducing agent and energy source under both aerobic and anaerobic conditions, producing carbon dioxide as a waste product.

Alternative adaptation systems

Although the unexpressed *che*Op$_1$ contains a putative CheD, there is no CheC homologue in the *R. sphaeroides* genome. As this operon does not contain a putative CheB but does contain a putative CheR (CheR$_1$), it is likely this CheD functions more like a CheB than a CheD, if it is ever expressed (del Campo *et al.*, 2011). There is no CheV homologue in the *R. sphaeroides* genome.

1.5 Project aims

This thesis focuses on several broad open questions regarding *R. sphaeroides* chemotaxis.

1. Why does *R. sphaeroides* have two chemotaxis clusters? Do the two clusters have distinct roles and do they both sense and respond to chemoeffectors?

2. Why are two CheYs required for chemotaxis and why do the two clusters phosphorylate different CheYs? These requirements point to different roles for the two clusters.

3. Does canonical methylation-based adaptation occur or some variant? As an adaptation system in 34H receptors (the majority of *R. sphaeroides* receptors) has not yet been recorded, *R. sphaeroides* may have a system similar to *B. sub-*

tilis, to *E. coli*, or a unique system.

4. Are the CheBs involved in signal integration between the clusters, as has been proposed? Is $CheB_2$ required for adaptation, or does it solely act in signal integration?

All of the above research questions require a more complete picture of how the chemotaxis protein network is wired in *R. sphaeroides* than is currently known. The specific aims of this thesis are therefore to:

1. Conduct a systematic investigation of the network connectivity of the chemotaxis system by analysing the motility (under free swimming steady state conditions) and chemotaxis (the response to dynamic changes in a chemoeffector) behaviour of a library of chemotaxis mutants.
   (a) Improve the existing methods for analysing swimming behaviour at steady state, avoiding the used of biased heuristics to define runs and stops.
   (b) Develop a statistical analysis scheme to analyse the population-level tethering (dynamic) data obtainable with the improved image analysis software.
   (c) Use any patterns in the resulting phenotypes to determine the network connectivity of the system.
2. Compare the elements for methylation-based adaptation to the systems in *E. coli* and *B. subtilis*.
   (a) Identify the chemoreceptors likely to be involved in adaptation through bioinformatics. Are they of the same class as those receptors known to be methylated in *B. subtilis* and *E. coli*?
   (b) Develop an experimental protocol for confirming the methylation of a chemoreceptor. Do the methylation sites follow the same pattern as those predicted by Alexander and Zhulin (2007) or match the pattern in *E. coli*

or *B. subtilis*?

(c) Determine the specificity of the diffuse adaptation enzymes.

3. Integrate the above information to develop an improved model of *R. sphaeroides* chemotaxis adaptation.

Chapter 2

Materials and Methods

Recipes for all mentioned media and buffers are included in Appendix B.

2.1 Strains and growth conditions

2.1.1 Strains and plasmids

Rhodobacter sphaeroides and *Escherichia coli* strains were used in this study. Specific strains and plasmids used are listed in Tables 2.1 and 2.2.

Table 2.1: Strains used in this study

Strain	Description	Source
R. sphaeroides strains		
WS8N	Naladixic acid resistant derivative of wildtype WS8	Sockett *et al.* (1990)
JPA425	WS8N Δ$cheY_{3,4}$	Shah *et al.* (2000b)
JPA517	WS8N Δ$cheB_1$	Martin *et al.* (2001b)
JPA565	WS8N Δ$cheR_2$	Martin *et al.* (2001b)
JPA900	WS8N Δ*mcpJ*	Jeevani Mantotta (this laboratory)
JPA915	WS8N $cheY_3$(D57A) Δ$cheY_4$	Mostyn Brown (this laboratory)
JPA917	WS8N $cheY_3$* (inactive) $cheY_4$(D57A)	Mostyn Brown (this laboratory)
JPA919	WS8N $cheY_3$(D57A) $cheY_4$(D57A)	Mostyn Brown (this laboratory)
JPA920	WS8N $cheY_3$(D57N) $cheY_4$(D57N)	Mostyn Brown (this laboratory)
JPA1037	WS8N *tlpT(Q451A)*	Ayse Ozhan and Mark Roberts (this laboratory)

Strain	Description	Source
JPA1038	WS8N *tlpT(Q475A)*	Ayse Ozhan and Mark Roberts (this laboratory)
JPA1039	WS8N *tlpT(Q528A)*	Ayse Ozhan and Mark Roberts (this laboratory)
JPA1204	WS8N $cheA_2$(H48Q)	Porter and Armitage (2004)
JPA1206	WS8N $cheA_2$(G470K)	Porter and Armitage (2004)
JPA1208	WS8N $cheA_2$(H48Q, G470K)	Porter and Armitage (2004)
JPA1210	WS8N $cheA_3$(H48Q)	Porter and Armitage (2004)
JPA1211	WS8N $cheA_4$(G470K)	Porter and Armitage (2004)
JPA1213	WS8N $cheY_6$(D57N) (nonmotile)	Porter *et al.* (2006)
JPA1216	WS8N $cheY_6$(D57A)	Porter *et al.* (2006)
JPA1218	WS8N $cheY_3$(D57N)	Porter *et al.* (2006)
JPA1219	WS8N $cheY_3$(D57N), $\Delta cheY_{1,2,4,5}$	Porter *et al.* (2006)
JPA1220	WS8N $cheY_4$(D57N)	Porter *et al.* (2006)
JPA1221	WS8N $cheY_4$(D57N), $\Delta cheY_{1-3,5}$	Porter *et al.* (2006)
JPA1243	WS8N $cheA_2$(H48Q, G470K) $cheY_6$(D57N) (nonmotile)	Porter *et al.* (2006)
JPA1246	WS8N $cheA_3$(H48Q) $cheY_6$(D57N)	Porter *et al.* (2006)
JPA1260	WS8N $cheA_4$(G470K) $cheY_6$(D57N)	Porter *et al.* (2006)
JPA1320	WS8N $\Delta cheR_3$	Porter *et al.* (2002)
JPA1323	WS8N $\Delta cheB_2$	Porter *et al.* (2002)
JPA1324	WS8N $\Delta cheB_{1,2}$	Steven Porter (this laboratory)
JPA1331	WS8N $\Delta tlpT$	Porter *et al.* (2002)
JPA1336	WS8N $\Delta cheY_6$	Porter *et al.* (2002)
JPA1337	WS8N $\Delta cheY_{1-6}$	Porter *et al.* (2006)
JPA1340	WS8N ΔBRA	Martin *et al.* (2003)
JPA1353	WS8N gutted of chemotaxis genes (non-chemotactic)	Porter *et al.* (2006)
JPA1373	WS8N $\Delta cheR_{1,2,3}$	Steven Porter (this laboratory)
JPA2330	WS8N *tlpT(Q261E)*	this study, Elaine Byles
JPA2331	WS8N *tlpT(Q261A)*	this study, Qin Qi
JPA2333	WS8N *tlpT(Q261D)*	this study, Qin Qi
JPA2334	WS8N *tlpT(E296D)*	this study, Qin Qi
JPA2335	WS8N *tlpT(E296Q)*	this study, Elaine Byles
JPA2337	WS8N *tlpT(Q451D)*	this study, Elaine Byles
JPA2338	WS8N *tlpT(Q451E)*	this study, Elaine Byles
JPA2340	WS8N *tlpT(Q475D)*	this study, Elaine Byles
JPA2341	WS8N *tlpT(Q475E)*	this study, Elaine Byles
JPA2343	WS8N *tlpT(Q528D)*	this study, Elaine Byles
JPA2350	WS8N *cfp-*$cheW_4$*,tlpT(Q261A)*	this study, Elaine Byles
JPA2351	WS8N *cfp-*$cheW_4$*,tlpT(Q261D)*	this study, Elaine Byles
JPA2352	WS8N *cfp-*$cheW_4$*,tlpT(Q261E)*	this study, Elaine Byles
JPA2354	WS8N *cfp-*$cheW_4$*,tlpT(E296D)*	this study, Elaine Byles
JPA2355	WS8N *cfp-*$cheW_4$*,tlpT(E296Q)*	this study, Elaine Byles
JPA2356	WS8N *cfp-*$cheW_4$*,tlpT(Q451A)*	this study, Elaine Byles

Strain	Description	Source
JPA2357	WS8N *cfp-cheW*$_4$,*tlpT(Q451D)*	this study, Elaine Byles
JPA2358	WS8N *cfp-cheW*$_4$,*tlpT(Q451E)*	this study, Elaine Byles
JPA2359	WS8N *cfp-cheW*$_4$,*tlpT(Q475A)*	this study, Elaine Byles
JPA2360	WS8N *cfp-cheW*$_4$,*tlpT(Q475D)*	this study, Elaine Byles
JPA2361	WS8N *cfp-cheW*$_4$,*tlpT(Q475E)*	this study, Elaine Byles
JPA2362	WS8N *cfp-cheW*$_4$,*tlpT(Q528A)*	this study, Elaine Byles
JPA2363	WS8N *cfp-cheW*$_4$,*tlpT(Q528D)*	this study, Elaine Byles
JPA2365	WS8N Δ(*cheB*$_1$, *cheB*$_2$, *cheBRA*)	this study
JPA2365	WS8N Δ(*cheR*$_1$, *cheR*$_2$, *cheR*$_3$, *cheBRA*)	this study
All mutations of *R. sphaeroides* CheAs and CheYs are identified by the residue number of the corresponding mutation in *E. coli* CheA and CheY.		
E. coli strains		
DH5α	Plasmid preparation	Gibco BRL
RP437	Wildtype	Parkinson (1978)
RP1091	RP437 Δ(all chemotaxis genes)	Parkinson and Houts (1982)
RP2867	RP437 Δ(*cheBcheR*)	Parkinson and Houts (1982)
S17-1 λ*pir*	Used for conjugations	Penfold and Pemberton (1992)
XL1-blue	Plasmid preparation	Stratagene

Table 2.2: Plasmids used in this study

Plasmid	Description	Source
pIND$_4$	Protein expression vector for *E. coli* and *R. sphaeroides*, kanamycin resistant, inducible *lac* promoter upstream of various restriction enzyme sites	Ind *et al.* (2009)
pK18*mobsacB*	Allelic exchange suicide vector mobilised by *E. coli* S17 λ*pir*	Schafer *et al.* (1994)
pQE60	Protein expression vector, C-terminus 6xHis tag, inducible *lac* promoter	QIAGEN
pQE80	Protein expression vector, N-terminus 6xHis tag, inducible *lac* promoter and *lacI*	QIAGEN
pREP4	Kanamycin resistant, suppresses leaky IPTG-inducing plasmids with *lacI*	QIAGEN
pUC19	High copy number cloning vector, ampicillen resistant	Pharmacia
pIND4-TlpT	pIND4 containing *R. sphaeroides tlpT*	This study
pQE60-Tsr	pQE60 containing *E. coli tsr*	This study
pQE80-B1	pQE80 containing *R. sphaeroides cheB*$_1$	Steven Porter (this laboratory)
pQE80-B1c	pQE80 containing methylesterase domain of *R. sphaeroides cheB*$_1$	This study
pQE80-B2	pQE80 containing *R. sphaeroides cheB*$_2$	Steven Porter

Plasmid	Description	Source
		(this laboratory)
pQE80-B2c	pQE80 containing methylesterase domain of *R. sphaeroides cheB*$_2$	This study
pQE80-R1	pQE80 containing *R. sphaeroides cheR*$_1$	Steven Porter (this laboratory)
pQE80-R2	pQE80 containing *R. sphaeroides cheR*$_2$	Steven Porter (this laboratory)
pQE80-R3	pQE80 containing *R. sphaeroides cheR*$_3$	Steven Porter (this laboratory)
pQE80-ECCheB	pQE80 containing *E. coli cheB*	This study
pQE80-ECCheBc	pQE80 containing methylesterase domain of *E. coli cheB*	This study
pQE80-ECCheR	pQE80 containing *E. coli cheR*	This study
pQE80-McpJCD	pQE80 containing cytoplasmic domain of *R. sphaeroides McpJ*	Steven Porter (this laboratory)
pQE80-TlpT	pQE80 containing *R. sphaeroides tlpT*	Steven Porter (this laboratory)
pQE80-TsrCD	pQE80 containing cytoplasmic domain of *E. coli tsr*	This study
pK18-delBRA	500 bp 5' upstream of *R. sphaeroides cheBRA* - 500 bp 3' downstream of *cheBRA* in pK18*mobsacB*, for genomic deletion of *cheBRA*	Steven Porter (this laboratory)
pK18-TlpT	*R. sphaeroides tlpT* with 500 bp upstream and downstream in pK18*mobsacB*, for genomic insertion of *tlpT* at wildtype position	Steven Porter (this laboratory)
pK18-TlpT(Q261A)	pK18-TlpT with *tlpT(Q261A)* point mutation	this study, Elaine Byles
pK18-TlpT(Q261D)	pK18-TlpT with *tlpT(Q261D)* point mutation	this study, Elaine Byles
pK18-TlpT(Q261E)	pK18-TlpT with *tlpT(Q261E)* point mutation	this study, Elaine Byles
pK18-TlpT(E296D)	pK18-TlpT with *tlpT(E296D)* point mutation	this study, Elaine Byles
pK18-TlpT(E296Q)	pK18-TlpT with *tlpT(E296Q)* point mutation	this study, Elaine Byles
pK18-TlpT(Q451D)	pK18-TlpT with *tlpT(Q451D)* point mutation	this study, Elaine Byles
pK18-TlpT(Q451E)	pK18-TlpT with *tlpT(Q451E)* point mutation	this study, Elaine Byles
pK18-TlpT(Q475D)	pK18-TlpT with *tlpT(Q475D)* point mutation	this study, Elaine Byles
pK18-TlpT(Q475E)	pK18-TlpT with *tlpT(Q475E)* point mutation	this study, Elaine Byles
pK18-TlpT(Q528D)	pK18-TlpT with *tlpT(Q528D)* point mutation	this study, Elaine Byles
pK18-W4-CFP	last 500 bp of *R. sphaeroides cheW*$_4$ - *cfp* - 500 bp 3' downstream of *cheW*$_4$ in pK18*mobsacB*, for genomic tagging of CheW$_4$ with CFP	George Wadhams (this laboratory)

2.1.2 *E. coli* growth conditions

E. coli liquid cultures were grown aerobically with shaking at 225 rpm in Luria-Bertani (LB) medium at 37^{o}C. *E. coli* frozen stocks or liquid samples requiring selection were streaked out on LB agar plates (1.5% w/v agar) and grown at 37^{o}C. In all cases, 10-16 h growth was required for a stationary culture or developed colonies.

DH5α and XL1-Blue were used for molecular cloning procedures. S17-1 λpir was used for conjugal transfer of plasmids to *R. sphaeroides*.

2.1.3 *R. sphaeroides* growth conditions

R. sphaeroides liquid cultures were grown in succinate medium (Sux) at 30^{o}C, either aerobically in the dark with shaking at 225 rpm, or anaerobically with illumination and without shaking. In both cases, two to three days' growth was required for a stationary culture.

R. sphaeroides frozen stocks or liquid samples requiring selection were streaked out on LB agar plates and grown at 30^{o}C for two to three days.

2.1.4 Antibiotics

Antibiotics were added to medium at the working concentrations given in Table 2.3, to select for successful plasmid uptake or the correct bacterial strain, as indicated in Table 2.1. Antibiotics were made up 1000-fold more concentrated than the working concentrations, filter sterilised and stored at -20^{o}C.

Chloramphenicol (Chl) was used to halt protein production in live cultures (see tethered cell assay, section 2.5.3 and free swimming capillary assay, section 2.5.1).

Table 2.3: Working concentrations of antibiotics used in selection.

Antibiotic	Working concentration	Diluted into
Ampicillen (Amp)	100 μg/ml	water
Chloramphenicol (Chl)	30 μg/ml	ethanol
Kanamycin (Kan)	25 μg/ml	water
Naladixic acid (Nal)	25 μg/ml	water
Tetracycline (Tet)	30 μg/ml	95% ethanol

2.1.5 Storage

Freezer stocks of stationary liquid culture were made for long-term storage at -80^{o}C. 600 μl culture and 400 μl 50% glycerol were mixed in a cryotube and flash frozen in liquid nitrogen.

2.2 Genetic manipulations

All standard genetic techniques were performed as described in Sambrook and Russell (2001).

2.2.1 DNA extraction

Plasmid extraction

Sequencing quality plasmid DNA was extracted from *E. coli* overnight liquid culture using the QIAGEN Plasmid Mini-Kit or Midi-Kit, depending on the culture volume. DNA was eluted in MilliQ and stored at -20^{o}C.

Genomic DNA extraction

Chromosomal DNA was extracted from *R. sphaeroides* and *E. coli* stationary liquid culture using a standard phenol-chloroform extraction (Chomczynski and Sacchi, 1987).

Cells from 1.5 ml culture were collected by centrifugation and flash-frozen in liquid nitrogen. The pellet was resuspended in 0.5 ml 65^{o}C lysis buffer. 100 μg Proteinase K was added and the samples incubated at 45^{o}C for at least 2 h.

The lysed digested cell debris was passed through two rounds of chloroform saturated phenol, sequestering cell debris in the phenol waste and extracting DNA to the aqueous fraction. DNA was precipitated from the aqueous fraction by addition of 1 ml ethanol and incubation at -20^{o}C for at least 30 min. The DNA was pelleted by centrifugation for 15 min at 13,000 g, washed with 1 ml 70% ethanol and centrifuged again.

The final pellet was dried at 37^{o}C overnight, then resuspended in 50 μl MilliQ and stored at -20^{o}C.

2.2.2 Cloning

Polymerase chain reaction (PCR)

All PCRs were performed using *Pfu* DNA polymerase (Promega), according to the manufacturer's directions. All primers were synthesised by Sigma-Genosys.

Unless otherwise mentioned, PCRs included 5-500 ng template DNA, 100 pmol of each primer, 12.5 nmol dNTP mix and 2.5 units polymerase in the supplied buffer, to a total volume of 50 μl. Primers used are listed in Appendix A.

Reactions were run in a Techne thermocycler using the following program. Annealing temperatures were chosen based on the melting temperature (T_m) of the primers.

Step	Temperature (^{o}C)	Time (min)
Initial denaturation	98	10
Repeat the following cycle 25 times		
Denaturation	98	2
Annealing	50-65	2
Extension	72	1 min per 500 bp + 1.5 min
Final extension	72	3 x above extension
Hold sample	4	indefinite

Gel electrophoresis

DNA fragments from PCRs or digest reactions were separated using gel electrophoresis. To make gels, 0.6-2% agarose was dissolved in 0.5x TBE buffer. The concentration of agarose was chosen to optimise separation of the particular size of DNA fragments. DNA samples were mixed with DNA loading dye and loaded onto the gel, alongside 1kb-plus DNA ladder (Invitrogen). The gel was placed in 0.5x TBE buffer and a potential difference of 120 V applied until the loading dye front had travelled the length of the gel.

The gel was stained in ethidium bromide for 15 min and the position of DNA fragments visualised using a UV transilluminator, λ = 360 nm. If required, bands of gel containing the relevant DNA fragment were then excised.

Gel extraction

PCR products and extracted plasmids were purified from agarose gel bands using the GenElute (Sigma) gel extraction kit. Purified DNA was eluted in 50 μl elution buffer, heated to 65^{o}C for plasmids and large fragments.

The gel extraction kit was also used to remove enzymes and replace buffers during sequential restriction enzyme digests, for example.

Restriction enzyme digests

Restriction enzymes were obtained from NEB. Restriction digests were carried out using the supplied buffers. If the enzymes in a double digest required different buffers, then reactions were performed sequentially. A PCR clean-up step using the standard gel extraction recipe (Section 2.2.2) was performed between reactions.

Unless otherwise required, digests were performed for 2 h at 37^{o}C.

Plasmid digests were followed by a 1 h incubation at 37^{o}C with 1 μl calf intestinal phosphatase, dephosphorylating the cut plasmid and thus reducing self-ligation.

Ligations

Digest products were ligated using T4 DNA ligase (NEB). Approximately three times as much insert as vector, total DNA around 1 μg, were incubated together in the appropriate buffer at 16^{o}C overnight.

Sequencing

Sequencing was performed by SourceBioscience Oxford. Sequences were analysed using Clone Manager version 9.0 (Sci Ed Central) and assembled using the Staden software package.

2.2.3 DNA transfer

Transformation of *E. coli*

Completed plasmids were inserted into chemically competent *E. coli* by transformation.

E. coli was made chemically competent using standard transformation buffer (TFB). A liquid overnight stationary culture was diluted 50-fold in LB media and grown aerobically as described until OD_{600nm} = 0.4-0.5. Cells were incubated on ice for 15-30 min, then centrifuged at 1000 x g for 10 min at 4^{o}C. The pellet was resuspended in 16 ml ice-cold TFB I buffer and incubated on ice for 15 min. After incubation, cells were centrifuged at 750 x g for 10 min at 4^{o}C and resuspended in 2 ml ice-cold TFB II buffer. The now competent cells were aliquoted into single use portions and stored at -80^{o}C for up to six months.

To transform the competent cells with DNA, ice-cold DNA was added to thawed cells and incubated on ice for 30 min. Cells were heat-shocked at 42^{o}C for 1.5 min, incubated on ice for 2 min, resuspended in 800 μl LB and incubated at 37^{o}C for 1 h. Cells were then spread onto LBA plates containing appropriate selective antibiotics and incubated overnight at 37^{o}C.

Conjugation into *R. sphaeroides*

As *R. sphaeroides* cannot be made chemically competent, plasmids were transferred into *R. sphaeroides* by conjugation with the *E. coli* strain S17 λpir. Only vectors based on pIND4 and pK18*mobsacB* were used in this way.

Plasmids were transferred into S17 λpir by transformation (Section 2.2.3). A liquid overnight culture of S17 λpir was diluted 50 μl in 5 ml LB and grown to faint cloudiness, or approximately OD_{600nm} = 0.3.

1 ml of this S17 λpir culture and 1 ml of a two-day liquid culture of *R. sphaeroides* were centrifuged separately at 6000 x g for 3 min. The supernatant was discarded and the cell pellets resuspended gently by tapping in 1 ml LB. Cells were pelleted as before, then resuspended in 100 μl LB. 10 μl *E. coli* and 100 μl *R. sphaeroides* were mixed gently and pipetted onto a nitrocellulose filter paper placed on a day-old LBA plate.

After incubating overnight at 30^{o}C, the filter paper was placed in an eppendorf with 800 μl LB and vortexed to resuspend the cell pellet. The LB was spread across four LBA-Nal plates containing the selection antibiotic for the plasmid used, and the plates incubated at 30^{o}C for two to three days.

2.2.4 Genomic changes in *R. sphaeroides*

Overlap-extension PCR

Overlap-extension PCR was used to introduce point mutations in a DNA construct or to splice out a section of DNA and thus to prepare vectors for introducing changes to the genome (Higuchi *et al.*, 1988).

Two pairs of primers were prepared, F1/R1, and F2/R2 (Figure 2.1). F1 binds 500 bp 5′ upstream of the target region to be deleted or mutated. R1 binds on the antisense strand at the target region. If a mutation is made, the 5′ half of R1 contains the mutation; if a deletion is to be made, the 5′ half of R1 contains the first segment of DNA after the gap.

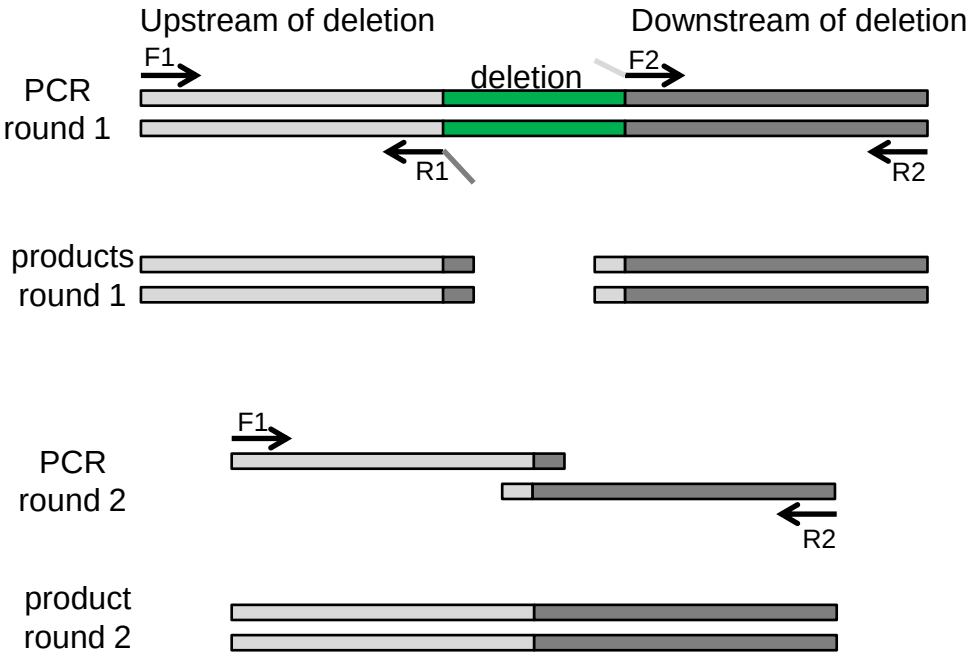


Figure 2.1: Overlap extension PCR for deletion of a region of interest from template DNA.

R2 binds on the antisense strand 500 bp 3′ downstream of the target region. F2 binds on sense strand at the target region. Like R1, the 5′ half of F2 does not match that binding site, but rather contains either the mutation or the first segment of DNA after the gap, in the opposite direction to R1.

The trailing regions of R2 and F1 are thus complementary.

Two rounds of PCR were done. Two PCRs were done in round one, using primer pairs F1/R1 and F2/R2 with wildtype DNA as a template. In round two, both PCR products from round one were used as a template with primer pair F1/R2. The complementary regions in F2 and R1 were introduced into the respective round one products, so that in round two, a single, long product was formed. The final product was either the wildtype DNA with point mutation, or the wildtype DNA with a section spliced out.

The final product was then transferred into a suicide vector for allelic exchange. For allelic exchange to occur, there must be approximately 500 bp DNA complementary to the genome on either side of the region for exchange.

Allelic exchange

The prepared pK18*mobsacB* suicide vector was transferred into the appropriate *R. sphaeroides* strain via conjugation (Section 2.2.3). Only colonies that successfully integrate the plasmid via homologous recombination survive the selective antibiotic Kan (Figure 2.2). A mix of these colonies was then used to inoculate 5 ml Sux-Nal without Kan and incubated for two days aerobically. This allows a second round of recombination to occur, removing the plasmid backbone containing the Kan resistance gene.

Serial dilutions up to 1 in 10,000 of the resulting liquid culture were spread on LBA plates containing 10% sucrose and incubated for three to five days. Sucrose selects for colonies that have successfully lost the plasmid backbone, and thus the *sacB* gene, whose presence inhibits *R. sphaeroides* growth on sucrose. 50 successful colonies were picked and replica plated on LBA-Nal and LBA-Nal-Kan plates and incubated for three to five days.

Only those colonies which grow on LBA-Nal but not on LBA-Nal-Kan have suc-

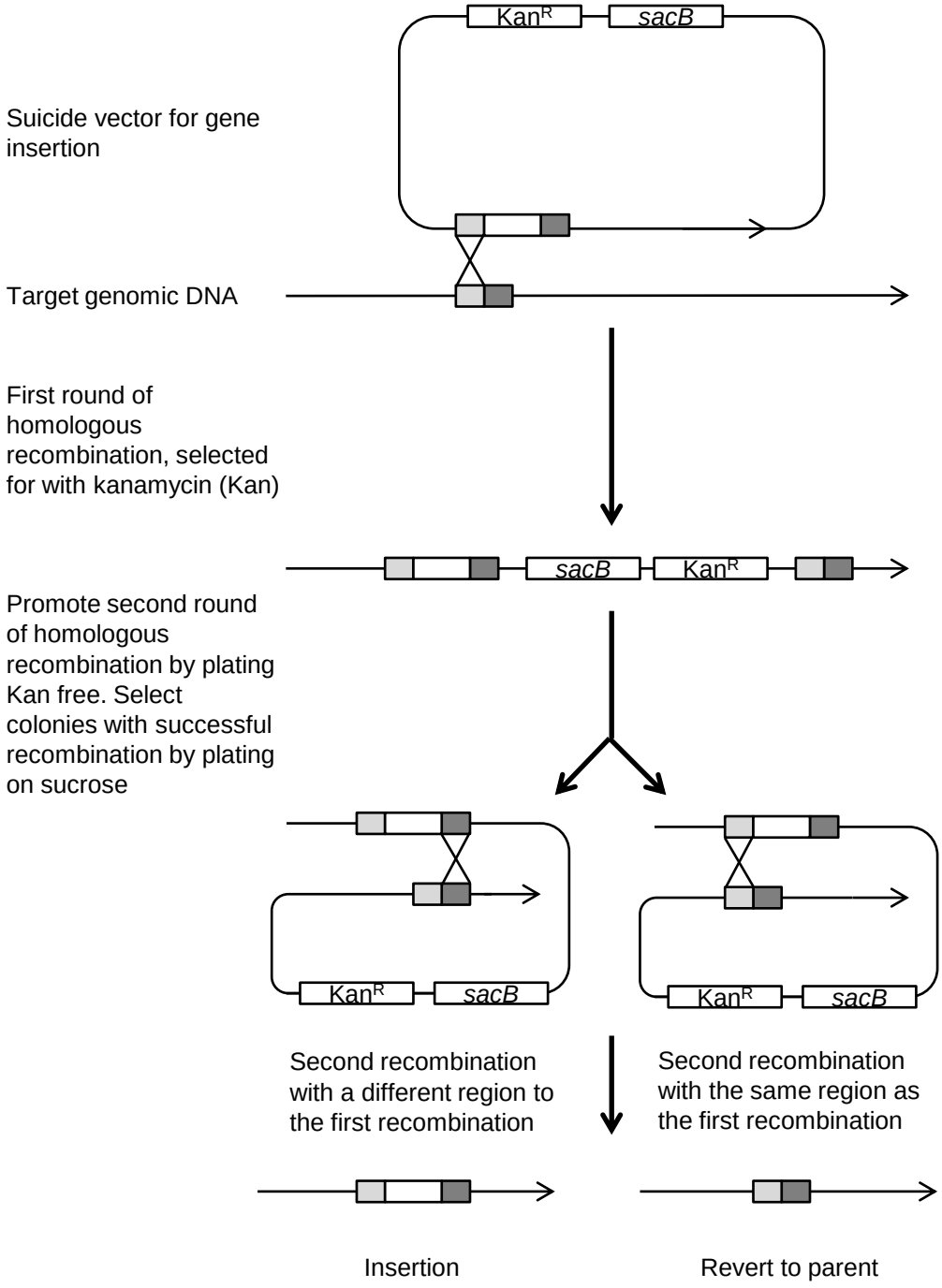


Figure 2.2: Double homologous recombination to insert or delete regions of interest from genomic DNA.

cessfully removed the plasmid backbone. These colonies have either reverted back to wildtype or incorporated the desired change. Southern blotting was used to distinguish between the two.

Southern blotting

Southern blotting allows the visualisation of DNA fragments which contain a sequence complementary to a highly specific probe.

Probe preparation A region of DNA that will have a significant size difference between the wildtype and the allelic exchange strain was chosen. Typically, the pK18*mobsacB* suicide vector contains a fragment that corresponds to this region. An appropriate fragment was cut from the vector and purified. To form a probe that will bind to the complementary region in genomic DNA, the purified DNA was labelled using the DPG DNA Labelling Kit (Roche Diagnostics), following the manufacturer's protocol.

Chromosomal DNA was extracted from the candidate colonies identified on LBA-Nal/LBA-Nal-Kan plates, along with parental strain DNA as a negative control, and a positive control if this particular allelic exchange had already been performed in another background strain. 20 μl resuspended DNA was added to a restriction digest mixture and digests incubated overnight at 37^{o}C. Restriction enzymes were chosen to give fragments of significantly different sizes in the parental strain and strains where the allelic exchange has occurred.

Digest products were separated by gel electrophoresis. The resulting gel was stained in ethidium bromide and photographed alongside a UV ruler.

The gel was soaked in denaturation buffer for 45 min, rinsed in MilliQ, then washed twice in neutralisation buffer, for 30 min and 15 min. The blot was then set up to allow transfer of DNA from the gel to a nitrocellulose membrane by capillary action

through Whatman paper and paper towels, in 20x SSC transfer buffer. Transfer was left overnight.

After transfer, the relative position of the gel wells was marked on the membrane, the membrane washed in 6x SSC, airdried and fixed using a UV auto-crosslinker.

The membrane was placed in a constantly rotating Hybaid tube filled with hybridisation buffer and incubated at 68^{o}C for at least 3 h. The probe was denatured by boiling for 15 min, then added to 25 ml fresh hydridisation buffer preheated to 68^{o}C. The Hybaid tube was emptied, the small volume of buffer-probe added and the membrane incubated with the probe overnight.

The membrane was washed twice with Wash A for 5 min each at room temperature, then twice with Wash B for 15 min each at 68^{o}C. The membrane was then removed from the tube and further washes performed flat at room temperature. The membrane was washed in buffer 1 for 1 min, buffer 2 for 3 min, buffer 2 with antibody conjugate for 30 min, twice in buffer 1 for 15 min each and finally in buffer 3 for 2 min. Detection solution in 20 ml buffer 3 was then poured onto the membrane, which was incubated in a sealed container in the dark until clear bands developed (10 min to two days). The reaction was stopped by 5 min wash in TE buffer, the membrane dried and stored.

2.3 Protein expression and purification

All proteins were expressed using 6xHis-tagging expression vectors, either the pQE system for expression in *E. coli* or pIND4 for expression in *R. sphaeroides*.

2.3.1 Overexpression in *E. coli*

Functional proteins required for *in vitro* assays were overexpressed in *E. coli* and purified under native conditions (Porter *et al.*, 2007).

Expression vectors were transformed into M15 or a relevant deletion strain. An overnight stationary culture was diluted 1 in 50 in 2YT media containing appropriate antibiotics and grown at 37^{o}C to OD_{600nm} = 0.5. Expression was induced with 1 mM IPTG. The culture was then incubated at 30^{o}C for 5 h if a chemosensory protein fragment was being expressed, or overnight at 18^{o}C for all other chemotaxis proteins.

After expression, cells were collected by centrifugation at 6,000 x g, 4^{o}C for 30-40 min and the resulting cell pellet frozen at -80^{o}C.

2.3.2 Low level expression in *E. coli* and *R. sphaeroides*

Full length MCPs and Tlps were expressed at close to endogenous levels in *E. coli* and *R. sphaeroides* strains to test for post-translational modification.

In *E. coli*, the expression vector was transformed into a relevant background strain made chemically competent (Section 2.2.3). An overnight stationary culture was diluted 1 in 50 in 2TY media containing appropriate antibiotics and 2.5 μM IPTG and grown at 37^{o}C to OD_{600nm} = 0.6.

In *R. sphaeroides*, the expression vector pIND4 containing a chemosensory protein was transferred to an appropriate background strain by conjugation via S17 λpir (Section 2.2.3). A two-day stationary culture was diluted 1 in 50 in Sux-Kan and 2.5 μM IPTG and incubated at 30^{o}C to OD_{700nm} = 0.6.

Protein expression was halted by addition of Chl and deamidation ensured by the addition of 100 μM S-methionine. After 30 min incubation, 100 μM attractant (propionate for *R. sphaeroides*, L-serine for *E. coli*) was added and incubation continued

for 20 min.

Cells were collected by centrifugation at 6,000 x g, 4^{o}C for 30-40 min and frozen at -80^{o}C.

2.3.3 Protein purification under native conditions

Ni-NTA chromatography under native conditions

Proteins were purified using nickel affinity chromatography. The frozen cell pellet (Section 2.3.1) was thawed on ice and resuspended in 5 ml/g protein buffer. Cells were lysed by sonication on ice (4 min sonication time, pulsed 5 s on, 15 s off on the Vibracell sonicator). The lysate was centrifuged at 35,000 x g for 30 min and the supernatant filtered through a 0.45 μm filter.

A nickel affinity chromatography column was prepared by pouring 1 ml Ni-NTA Agarose slurry (Bio-Rad) into a chromatography column, allowing to settle for 10 min and equilibrating with 5 ml protein buffer.

Filtered supernatant was applied to the column and run through. The column was then washed with at least twice as much protein buffer as used to resuspend the cell pellet. Protein was then eluted using 5 ml native elution buffer, collected in 1 ml fractions. Protein usually elutes in the first three fractions.

Size exclusion chromatography

Native elution buffer was exchanged for protein buffer and remaining impurities removed from eluted protein fractions by size exclusion chromatography (gel filtration). This was only performed on cytoplasmic proteins.

The protein fraction was thawed on ice then centrifuged at 10,000 x g for 15 min at 10^{o}C. The fraction was injected onto a S75 gel filtration column attached to an Akta

system, using flow 1 ml/min. Protein buffer was used. The void volume was discarded, then fractions of 1 ml collected for approximately 80 ml. For most proteins, UV absorbance could be used to identify which fractions contained significant protein. For proteins with no aromatic residues, the Bradford assay (section 2.3.5) was used to determine which fractions had significant protein.

Fractions containing the protein of interest were identified by subjecting all fractions with significant protein to SDS-PAGE (section 2.3.5). Fractions containing the protein of interest, usually 5-10 ml total, were pooled and concentrated to more than 1 mg/ml using a centrifuge concentrator.

Dialysis

Elution buffer was exchanged for protein buffer in eluted fractions of chemosensory protein fragments by dialysis, as these proteins could not be concentrated using a centrifuge concentrator. Eluted fractions containing chemosensory protein, usually 2 ml total, were placed in dialysis tubing and sealed. The tubing was floated in 2 l fresh protein buffer at 4^oC for 24 h three times.

Protein was then aliquoted into single use aliquots and stored at -80^oC.

2.3.4 Protein purification under denaturing conditions

Proteins were purified using nickel affinity chromatography. The cell pellet was thawed on ice, then resuspended in denaturing protein buffer B (5 ml/g cell pellet) with gentle stirring at room temperature. Cells were lysed by sonication on ice (4 min sonication time, pulsed 5 s on, 15 s off on the Vibracell sonicator). The lysate was centrifuged at 10,000 x g for 30 min at room temperature and the supernatant removed to a clean 50 ml Falcon tube.

1 ml Ni-NTA agarose slurry was added to the supernatant and incubated at room

temperature on an orbital shaker for 1 h. The tube was spun down at 1,000 x g for 10 min at room temperature. Supernatant was discarded, 50 ml buffer C added and the tube incubated on the orbital shaker for 1 h. Buffer was removed as before and a second buffer C wash applied for 1 h. Buffer was removed as before, leaving approximately 5 ml supernatant above the resin.

Resin was gently resuspended in remaining buffer C and the slurry poured into a sealed chromatography column. Once the resin settled, the remaining buffer C was run through. Protein was eluted with 5 ml buffer D, then buffer E, collecting 1 ml fractions. Depending on aggregation, some proteins elute in buffer D and others in buffer E, usually within the first two fractions.

Fractions are then aliquoted into single use aliquots and stored at -80^{o}C.

2.3.5 Confirming protein expression

Bradford assay

The concentration of total protein in a sample was determined using the Bradford assay (Bradford, 1976). Protein produces a colour change from brown to blue in Bradford reagent. There is a linear relationship between protein concentration and the resulting absorbance at OD_{595nm}. For each fresh batch of Bradford reagent, a standard curve was measured. A linear dilution of bovine serum albumin (BSA), concentrations 1 mg/ml to 8 mg/ml in MilliQ was made. 20 μl BSA dilution was added to 1 ml Bradford, incubated for 5 min and the OD_{595nm} recorded. Unknown protein samples were treated in the same way and the protein concentration calculated from the standard curve.

Protein separation by electrophoresis

The identity of the expressed protein in a cell extract or in purified protein was suggested by size comparison after separation using SDS polyacrylamide gel electrophoresis (SDS-PAGE).

5x protein loading dye containing SDS was added to samples to be visualised. Native proteins and cell pellets were boiled at 95^0C for 15 min to denature, whereas denatured proteins (in urea) were not heat treated. Samples were loaded on a precast 12% or 4-20% acrylamide gel, alongside Benchmark Protein Ladder (Invitrogen) or Kaleidoscope Protein Ladder (BioRad), surrounded by SDS-PAGE running buffer. 180 V, 110 mA of current was applied for 1 h.

The gel was then either stained using Invitrogen In-gel His-stain (staining only proteins with a His-tag) or InstaBlue Coomassie blue stain (staining all proteins).

Visualising protein fragments by Western blotting

The identity of a protein could be more conclusively confirmed using Western blotting.

An SDS-PAGE gel was run as before, including Kaleidoscope Protein Ladder (Bio-Rad) as the marker. The gel was layered with sponge, Whatman paper and PVDF membrane (first floated in methanol) in a cassette soaked in blotting buffer, using the standard set-up described in Sambrook and Russell (2001), and the cassette placed in an electrophoresis tanked filled with blotting buffer. Protein was then transferred to the PVDF membrane by electrophoresis using 500 mA current for 1 h.

To visualise protein on the membrane, the membrane was blocked in 5% milk powder in PBS overnight. The membrane was then incubated for at least 1 h in 1% milk powder in PBS with primary antibody (raised against a specific protein or group of proteins) diluted approximately 1 in 2000. The membrane was washed for 10 min

in, successively, PBS, PBS with 0.2% TWEEN20 twice and PBS again. The membrane was then incubated with secondary antibody (HRP conjugated antibody raised in goat against whichever animal was used to produce the primary antibody) diluted approximately 1 in 10,000 for at least 1 h, followed by the same wash protocol.

The Millipore HRP detection kit was used, following the manufacturer's protocol. The membrane was incubated in 10 ml HRP substrate for 5-20 min at room temperature. Excess substrate was removed and the membrane placed between acetate sheets. The luminescence produced by the HRP was collected using the GelDock camera system. Images were taken at 1Mbx1Mb resolution, one image every 1 min for 30 min, adding successive images.

2.4 *In vitro* methylation/deamidation assay

The ability of functional CheB and CheR homologues to respectively deamidate and methylate the cytoplasmic fragment of a chemoreceptor was tested using an *in vitro* assay. This assay uses the principle that the same chemoreceptor at different methylation and deamidation states will travel at different speeds during gel electrophoresis.

A receptor and an adaptation protein were incubated together in a variety of molar ratios, usually 1:1, 5:1 and 1:5. If testing methylation, the CheR substrate SAM was added at 100 μM. If testing deamidation with full-length CheB, the phosphodonor phosphoamidate was added at 100 μM. Final reaction volumes of 100 μl were typically used. Incubation took place at 30^{o}C if an *R. sphaeroides* protein was used, or at 37^{o}C if only *E. coli* proteins were used, for 1 h.

SDS-PAGE as in section 2.3.5 was then used to visualise differing methylation and deamidation states, with some differences. The denatured reaction mix in protein loading dye was loaded alongside Benchmark Protein Ladder (Invitrogen) on a

large (25x25 cm) 12% resolving acrylamide gel topped with a 5% stacking gel. Gel electrophoresis was performed with cooling to 4^{o}C, applying 95 V for 5-10 h. The gel was stained using InVision His-stain, followed by InstaBlue Coomassie stain for comparison.

2.5 Phenotype analysis

2.5.1 Free swimming capillary assay

Motility of *R. sphaeroides* strains at the cell level was tested in a zero-flow microaerobic environment with a glass capillary. 100 μl aerobic mid-log culture (OD_{700nm} = 0.4-0.6) was spun down at 1000 rpm for 1 min, then very gently resuspended in 1 ml 10 mM PIPES (pH 7.2) with Chl by tapping. The resuspension was left to incubate for 20 min on the desk, allowing cells to acclimatise to the new buffer and resume motility.

A 0.2 x 2 mm glass capillary tube was filled with the resuspension and the ends dipped in silicon grease to air seal. This creates a zero-flow environment. Although the culture is grown aerobically, within the capillary oxygen is quickly consumed. The environment is therefore considered microaerobic.

The filled capillary was laid lengthways on a glass slide and visualised under phase contrast microscopy at 20x magnification. The centre of the capillary, or bulk, rather than the surface of the capillary, was visualised. Initial free swimming behaviour under homogenous environmental conditions was noted. Five phenotypes can generally be differentiated by eye: nonmotile, stoppy, wildtype, infrequent stopping and smooth swimming. Videos of 20 ms per frame, 2 min long were recorded.

The image segmentation program Tracker was used to detect cells and form tracks (Wood *et al.*, 2012). In Tracker, cells are detected in each frame by differences in

pixel intensity, defined by the Niblack algorithm including user-defined parameters specific to each video. A moving window size of 15 pixels were used and objects were defined as larger than 2 pixels. For a positive identification, pixel intensity must be higher or lower than the mean pixel intensity, with standard deviation taken into account. Upper and lower bounds of 3.5-4.0 were used for all videos.

After object detection, tracks were formed by identifying an object frame to frame, comparing the probability distributions of where each object in the preceding frame could be, to where each object in the current frame is. *R. sphaeroides* swimming speed and behaviour is used to inform the probability distributions.

Tracks were then censored using a top framewise speed cutoff, minimum bounding radius for the full track, minimum track length and tortuosity. The censoring protocol is discussed in detail in Chapter 3. The remaining tracks were classified into runs and stops using the analysis program RunStopAnalysis (Rosser *et al.*, 2013).

RunStopAnalysis takes as input two sets of tracks, a known nonmotile strain simulating a set of stops, and a known smooth swimming strain, simulating a set of runs. KDE functions for framewise speed distributions are built for both. An exponential function describes the run angle change distribution and a uniform function describes the stop angle distribution. The third input is a set of tracks for analysis. The four functions are first used in a hidden Markov model to determine the probability of any frame transitioning from a run to a stop, and from a stop to a run, for the unknown data set. Following this, the four functions are used to classify each framewise transition in the data set as a run or a stop, on the basis of that transition's speed and angle change. The identity of the preceding transition in the track informs the identity of the current transition, through the use of the calculated switching probabilities.

A post-processing step was applied, restricting each run and each stop to a minimum length of two transitions.

The classified tracks were then be compared using summary statistics, such as track length, fraction of time stopped in a track, length of stop and frequency of switching. The distributions of these summary statistics for tracks are not Gaussian, so non-parametric tests must be used to compare data sets.

2.5.2 Soft agar swim plates

The motility of *R. sphaeroides* strains at the population level was tested using swim plates with 0.25% agar in M22 minimum medium. The attractant propionate was added at 100 μM and is the only organic carbon source.

Plates were inoculated with 5 μl drops of stationary *R. sphaeroides* culture, grown either aerobically with shaking, or anaerobically with illumination at sample sites. Swim plates were incubated at 30^{o}C, either aerobically in the dark or anaerobically with illumination.

As bacteria multiply, a circular colony of culture forms. Motile strains are able to move through the sloppy media, forming a larger spread. As bacteria metabolise the attractant, a concentration gradient forms within the media. Strains with functional chemotaxis pathways move up this concentrations gradient, forming larger spots than motile strains without chemotaxis.

Swim diameters were recorded 48 h after inoculation. Three independent sets of each plate were carried out on separate days using fresh culture, each in triplicate, giving nine repeats.

A one-way ANOVA test was carried out at each condition, with a post-hoc Tukey test. The null hypothesis considers all strains to be a part of the same population. A p-value of below 0.05 is needed to reject the null hypothesis and place two strains in separate groups.

2.5.3 Tethered cell assay

The response of individual *R. sphaeroides* cells to a rise in attractant concentration, then drop in attractant concentration, was tested through tethering.

Cells from 1 ml aerobic or anaeorobic mid-log culture (OD_{700nm} = 0.4-0.6) were collected by centrifugation at 1000 rpm for 1-3 min. The cell pellet was very gently resuspended in 1 ml 10 mM PIPES with Chl by tapping and centrifuged again. This wash step was completed twice. After the third spin-down step, the cell pellet was resuspended in 1 ml 10 mM PIPES (pH 7.2) with Chl.

The edge of one surface of a small circular glass slide (12 mm diameter) was coated in grease, viscosity L (Apiezon Products). The slide was balanced grease side up. 2 μl anti-flagellum antibody, diluted 1 in 1000 from stock preparation, was pipetted onto the centre of the slide. 10 μl resuspended culture was pipetted on top of the antibody droplet. The slide was left to incubate for 20 min, tethering cells by their single flagellum to the glass.

A flow cell was prepared, sealing the small glass slide onto the microfluidic device with the culture droplet facing into the flow cell. 10 mM PIPES with Chl was flowed through the device continuously. The slide was visualised using light microscopy at 40x magnification. A field of view was found with rotating cells lying separate from one another. A recording was made at 10 ms per frame for 13 min. For the first 3 min, 10 mM PIPES with Chl was flowed through the device. From minute 3 to minute 8, 10 mM PIPES with Chl and 100 μM propionate, the attractant, was flowed through. From minute 8 to minute 13, 10 mM PIPES with Chl was flowed through. This drop in attractant causes cells with functional chemotaxis to stop rotating, then recover rotation. If very few or no cells had recovered by minute 13, the recording was continued to minute 18. If a strain in general had not recovered by minute 18, that strain was considered unable to recover.

The image analysis software BRAS and Click&Mean was used to analyse videos of

tethered cells responding to additional and removal of an attractant, described in Kojadinovic *et al.* (2011).

BRAS is an image segmentation program. The user selects which cells to analyse, enclosing each within a moveable window and ensuring that the cell remains within the window over the duration of the video. For this study, any cell which showed rotation was selected at this point. Once each window has been selected, BRAS analyses the video frame by frame. In each frame, the cell in each window is identified by the difference in pixel intensity between the cell and the background and the xy coordinates of the cell's midpoint recorded.

Click&Mean processes and displays the coordinate data recorded by BRAS. The change in x-y coordinates between each frame is converted into a rotational speed. Each window's (and thus cell's) rotation output is processed through a Fourier transformation using a smoothing window of 128 frames.

Cells were classified as either useful or not useful, as in many cases the rotational signal from a cell is not clear enough to be interpreted. The behaviour of each useful cell was recorded as unresponsive (does not stop), responsive (stops and does not recover), adaptive (stops, then recovers) or inhibited (starts on propionate addition, stops before propionate is removed). Cells were further classified as either having a stoppy appearance or a constant rotation signal, and as starting rotation on addition of propionate or rotating in buffer from the start.

The start, end and duration of the stop was recorded for adaptive cells where the stopping event occurs around the point of attractant removal.

If alternative behaviour was seen, such as rotation stopping on attractant addition, this is also recorded.

The duration of an adaptation event was defined as the time taken from the stop, which is instantaneous, to the point of full recovery. Often there is a period of recovery where rotation is not fully restarted, but is jerky and produces an unclear signal.

This was included in the stopped period. Once a period of clear rotation for at least 5 s occurred, the adaptation period was considered over.

For each strain and growth condition, the percentage of unresponsive, responsive and adaptive cells was calculated. The average adaptation time and standard deviation was calculated where relevant. A one-way ANOVA test with a post-hoc Tukey test was used to identify strains which had an adaptation time significantly different to wildtype. A p-value of below 0.05 is needed to reject the null hypothesis that all strains are from the same population and place two strains in separate groups.

2.6 Statistical tests

2.6.1 Box-and-whisker plot

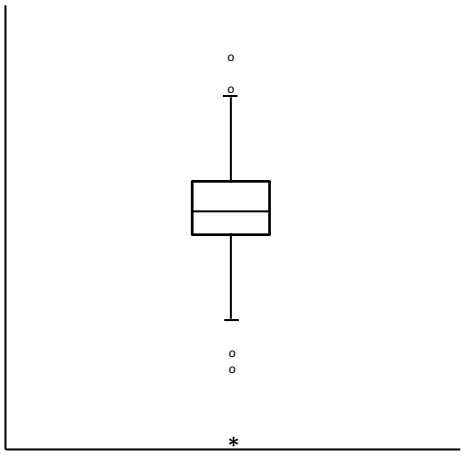


Figure 2.3: Example of a box-and-whisker plot

Box-and-whisker plots were used to graphically represent the distribution of a non-parametric data set. Figure 2.3 shows a representative example of a boxplot. The line in the middle of the box is the median value. Half of the data points are greater than the median and half are lower. The bottom hinge, or the bottom of the box, is the 25th percentile, which is the value below which 25% of the data points are found. The top hinge, or the top the box, is the 75th percentile, which is the value

above which 25% of the data points are found. 50% of the data points therefore lie within the box.

The whiskers, or inner fences, extend 1.5 times the height of the box in either direction. If no data points are that large/small, then the whiskers indicate the minimum and maximum data point values.

Any data points that lie outside the whiskers are considered outliers. Circles represent outliers between 1.5x and 3x the height of the box. Stars represent extreme outliers, which lie beyond 3x the height of the box in either direction.

Chapter 3

Exploring signal integration between the two chemotaxis clusters through *in vivo* methods

In this chapter, two measures for describing chemotaxis and motility are presented.

1. Tethering allows a cell's response to a chemotaxis challenge to be measured. Newly developed software is used to analyse hundreds rather than tens of cells, giving insight into population-level variability.
2. The free swimming behaviour of a cell under homogeneous environmental conditions is explored. A new analysis method for classifying swimming into run and stop events can differential the motility in different strains.

Strains with deletions and mutations of the chemotaxis proteins are analysed using these measures. Novel phenotypes are recorded. Patterns in the motility and chemotaxis behaviour of the different strains suggest how the network of chemotaxis proteins is wired. The results suggest distinct roles for the two chemotaxis clusters.

3.1 Introduction

Unlike other models of bacterial chemotaxis, *R. sphaeroides* has two clusters of chemotaxis proteins, both of which are required for the chemotaxis network to function. Why this is the case and what the role of each cluster is, remains an open question. A standard method for determining the connectivity of a protein network is the deletion of each component and measurement of the resulting changes in phenotype. Although a library of *R. sphaeroides* chemotaxis deletion mutants exists, a systematic analysis of these mutants has not yet been done.

There are two measures which describe the motility of a chemotactic cell. The first is swimming under homogeneous environmental conditions, when the signalling system is at steady state and allows a random exploration of space. Functional motility is a requirement for chemotaxis; the output of a functioning chemotaxis pathway is to alter motility. The second is swimming behaviour in the presence a specific stimulus, such as a sudden drop in the concentration of a chemoeffector, and the resulting dynamic swing away from and return to the steady state. Past analyses of the motility of *R. sphaeroides* deletion mutants have been limited to swim plates and the tethering of a small number of individual cells (for example, Martin *et al.* (2001b); Porter *et al.* (2006)).

Swim plates give an indication of a whole population's steady state and dynamic behaviour. A small amount of culture is dotted onto a sloppy agar plate, made with minimal media and a known attractant. As the culture grows and metabolises, it forms a concentration gradient of that attractant in the agar. The diameter of the resulting colony is dependent on motility, as cells can swim through the sloppy agar, and on efficiency in chemotaxis, as cells can access more nutrients by travelling up the resulting gradient. Strains with different mutations affecting chemotaxis tend have similar swim diameters. Swim plates give an overview of an entire population's motility, but give no information as to how motility has been affected, or

whether it is the steady state or dynamic state that has been affected.

Tethered cells are used to analyse dynamic behaviour. Cells are attached to a glass slide by a single flagellum and subjected to a rise and then a drop in attractant levels. Flagellar motor rotation results in the whole cell body rotating, as the flagellum is held in place. This rotation can be recorded and a stop or change in rotation due to a change in attractant level is easily seen. However, this system is not useful for visualising the swimming behaviour at steady state, as the high load placed on a tethered cell changes the rate of transient stops (Brown, 2009).

These two methods therefore give an indication of general motility and detailed information on dynamic chemotactic behaviour, but do not give the same level of detail regarding steady state behaviour. Analysing the free swimming behaviour of a cell under homogeneous environmental conditions would give this information. In this chapter, the free swimming steady state behaviour and tethered dynamic behaviour of a number of deletion mutants are analysed.

3.2 Free swimming behaviour in *R. sphaeroides*

When within a homogeneous environment, *R. sphaeroides* explores its space using a random walk by switching between runs and stops. Genetic manipulation of the chemotaxis and motility machinery has resulted in a measurable change in motility in only a few cases.

3.2.1 Previous work

Free swimming behaviour in *E. coli* has typically been studied by either manually following a limited number of individual cells swimming freely within a zero-flow environment (e.g., Berg and Brown (1972)) or by recording whole populations of cells swimming freely (e.g., Alon *et al.* (1998)). Tracks are then extracted from record-

ings and tumbling or stopping events classified either by hand based on observable rules (e.g., Sager *et al.* (1988)) or by applying heuristics such as imposing a speed and angle change cut-off, the results of which are checked against by-hand classification (e.g., Alon *et al.* (1998)).

R. sphaeroides free swimming behaviour has been analysed using the whole population and heuristic classification approach (Warren, 2003). In this instance, the recording system used VHS tape with an effective frame rate of 24 frames per second. The analysis took place as video was played back, so a ten minute video clip had a 10 minute analysis time - essentially, a real time analysis with no possibility for optimisation. The software used to generate the data is now obsolete.

In *E. coli*, four clear phenotypes of swimming behaviour in homogeneous conditions have been recorded: nonmotile, tumbly, wildtype (switching between runs and tumbles) and smooth swimming. These behaviours have been linked with specific genetic mutations. Nonmotile behaviour only occurs when the assembly or function of the flagellar motor has been prevented. Tumbly behaviour occurs when there is excessive CheY-P, either through over-activation of CheA, such as when CheB is deleted, or through slow dephosphorylation of CheY-P, such as when CheZ is deleted. Smooth swimming behaviour occurs when CheY-P is not produced, through some failure of the CheA activation process such as deletion of CheR.

Table 3.1: *E. coli* (from Block *et al.* (1983) unless otherwise indicated) and *R. sphaeroides* (Brown, 2009) wildtype swimming parameters.

	E. coli	*R. sphaeroides*
Cell dimension	1 x 2-6 μm	1 x 2 μm
Mean speed	16.5 μm/s (Alon *et al.*, 1998)	40 μm/s
Mean run time	1 s	3 s
Mean tumble or stop time	0.1 s	0.66 (range: 0.01-6) s
Mean tumble or stop frequency	0.23 ± 0.19 s (Weis and Jr, 1990)	0.31 ± 0.19 /s
Mean run bias	0.6-0.7	0.8 ± 0.2

In *R. sphaeroides*, the same diversity of swimming behaviour has not been recorded. We should see a 'stoppy' *R. sphaeroides* phenotype analogous to the *E. coli* tumbly phenotype, and similarly a smooth *R. sphaeroides* swimmer analogous to the *E. coli*

smooth phenotype. Instead, genetic mutations of the individual CheAs, CheBs, CheRs and CheYs have uniformly been reported as having 'wildtype motility', except for some notable exceptions.

When $CheB_2$ is deleted, the resulting swimming behaviour appears to have shorter stops than wildtype. Deletion of $CheR_3$ results in longer stops than wildtype. In both cases, however, no statistical difference between these strains and wildtype was found (Warren, 2003). The mutation $CheY_6$(D57N) results in the only recorded case of statistically more frequent transient stops using the original VHS analysis system, unsurprising as this strain is nonmotile (Porter *et al.*, 2006).

R. sphaeroides's use of stops rather than active tumbles for direction change is possibly why this method has not been as successful as in *E. coli*. The frame-to-frame angle change in a stopped track is not as great as that seen in an actively tumbling track.

R. sphaeroides strain JPA 1353 is gutted for all chemotaxis genes. It has been recorded as smooth swimming and unable to perform transient stops. Rather than using the image segmentation and tracking software available at the time, Brown (2009) used two alternative methods to classify this and other *R. sphaeroides* strains as smoother or stoppier than wildtype.

Brown (2009) used a novel bead assay to explore free swimming. The flagellum was shortened and attached to an optical bead through antibody interactions (Section 1.1.3). Individual flagellum-bead pairs were captured in an optical track and the bead position recorded. Position change was used to calculate rotational velocity at each time point. Stopping events were then defined as events with speeds between 0.4 and 0.1 of the mean speed. Load on the motor is significantly less in this assay than during tethering. Further, the rate of data capture was markedly improved from the VHS recordings of the past.

Although this is an accurate method, clearly showing transient stopping events in

wildtype cells, it is time consuming as a single cell is analysed at a time. Brown (2009) used an alternative screening method to define strains as stoppier or smoother than wildtype. Mid-log phase culture was drawn up into a capillary and laid next to a capillary containing wildtype. In a double-blind fashion, strains were repeatedly classified as smoother or stoppier than the reference capillary. JPA 1353 was repeatedly selected as smoother in this method.

3.2.2 Alternative analysis

The work in Brown (2009) suggests that the lack of reported variability in *R. sphaeroides* swimming behaviour under homogeneous conditions is due to the method of data recording (low frame rate) and the method of analysis (over-smoothing removing significant events). Using current analysis techniques to re-examine existing mutants will indicate if this is true. The bead assay requires too many repeats to be sensible for screening motility of many strains. Any improved method would have the accurate, clean, quantitative data of a bead assay combined with the speed of the double blind assay and would automate as much of the analysis as possible.

As an initial screen, deletion mutants were grown to the mid-log phase of exponential growth (OD_{700nm} = 0.6) and treated with Chl to halt growth and protein synthesis. Capillary tubes were filled with this culture and viewed under a light microscope at 20x magnification (Section 2.5.1). Marked differences were seen in swimming behaviour between deletion mutants, ranging from very smooth to very stoppy. Crucially, when digital recordings of this free swimming were made at 50 Hz, the differences between strains could still be seen in playback.

Variable swimming behaviour had thus been hidden within the noisy video data and analysis by the previous system. It is likely that the loss of detail was due to the use of low frame rate VHS video for data capture and complicated by the fact that stops are not as obviously different from runs as tumbles are. Digital recordings

give the option for higher frame rate capture, as well as freeing analysis from the constraint of 'in real time' image analysis.

The fact that by eye the differences between smooth, stoppy and wildtype are obvious suggests that the free swimming capillary assay has potential as a high throughput method to differentiate swimming behaviour in *R. sphaeroides* as in *E. coli*. Dr G. Rosser (Department of Applied Mathematics, University of Oxford), in collaboration with Dr D. Wilkinson, S. Chacko and this author (Department of Biochemistry, University of Oxford), has developed improved software for analysing free swimming data. As yet, this software has only been used to analyse wildtype behaviour, comparing it to the smooth swimming behaviour of the gutted *R. sphaeroides* strain (Rosser *et al.*, 2013).

The rest of this section explains how this method was extended to allow analysis of strains showing stoppy and smooth swimming.

3.2.3 Data collection and processing

R. sphaeroides cultures were grown and capillary tubes filled as before. Free swimming behaviour was captured at 50 Hz, 20x magnification, for 2 min. Three sets of videos were taken for each strain and the experiment repeated across three days, giving nine repeats across three biological replicates. Three standards were used. WS8N represents wildtype motility. JPA 1353, the strain gutted for all chemotaxis genes, is the known standard for smooth swimming (Brown, 2009). JPA 1246, WS8N with $CheY_6$(D57N) is the known standard for nonmotile, flagellate *R. sphaeroides* (Porter *et al.*, 2006).

Videos were processed in three sections, tracking, clean-up and classification.

Tracking

Tracking software was written by Dr C. Yates and Dr T. Wood (Department of Mathematics, University of Oxford). Each frame is automatically segmented using the Niblack algorithm with user-entered parameters optimised for each video set. The positions for identified objects (areas darker or lighter in average pixel intensity than their surroundings) in each frame of a video are recorded. This information is then converted into tracks, identifying which object in sequential frames is the same object using probability densities (Wood *et al.*, 2012).

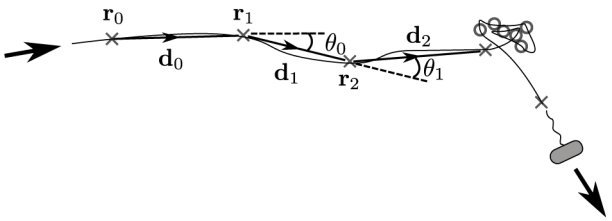


Figure 3.1: Data representation in a free swimming track. The thin black line shows the movement of a *R. sphaeroides* cell. Crosses and circles denote running and stopping phases respectively, and represent the locations $\mathbf{r}_t$ at which the position of the cell is recorded every constant time interval Δt. Dashed black lines and notation illustrate the mathematical representation of the track, in text. Reproduced from Rosser *et al.* (2013), published under a CC-BY license.

From these tracks, the frame-to-frame distance travelled was calculated by

$$\mathbf{d}_t = \mathbf{r}_{t+1} - \mathbf{r}_t, \tag{3.1}$$

where $\mathbf{d}_t$ is the distance travelled between time t and time $t+1$, $\mathbf{r}_t$ is the position at time t and $\mathbf{r}_{t+1}$ is the position at time $t+1$, as shown in Figure 3.1.

The change in direction calculated by

$$\begin{aligned} \theta_t &= tan^{-1}\frac{\mathbf{d}y_t}{\mathbf{d}x_t} - tan^{-1}\frac{\mathbf{d}y_{t\text{-}1}}{\mathbf{d}x_{t\text{-}1}}, \\ \theta &\in [-\pi, \pi], \end{aligned} \tag{3.2}$$

where θ_t is the angle made between the vector of the travel $\mathbf{d}_{t-1}$ and $\mathbf{d}_t$, corrected to fall within the scale $[-\pi, \pi]$, as shown in Figure 3.1.

Data clean-up

Before tracks can be analysed, the data set is cleaned to remove as many invalid tracks as possible. Data censoring was done as described in Rosser *et al.* (2013), with some changes to accommodate differing swimming patterns from wildtype.

All strains are treated in the same way, including the standards for wildtype (strain WS8N) and smooth swimmers (JPA 1353). The nonmotile standard JPA 1243 is treated separately.

Three kinds of unusable tracks are found in the motile data sets: nonmotile, truncated and corkscrews. In all cases, to prevent skewing the data set with extra runs or extra stopping events, these tracks must be identified and removed. A final unusable track is found in all data sets, jumps. Examples of each of these tracks compared with wildtype run-stop-run tracks are shown in Figure 3.2.

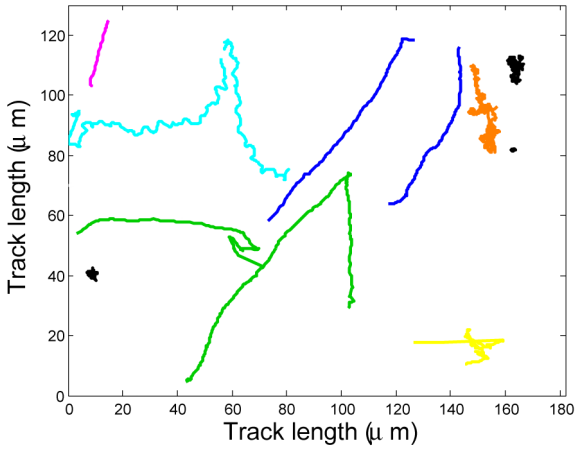


Figure 3.2: Representative free swimming tracks, showing nonmotile (black), drifters (orange), truncated (magenta), corkscrews (cyan), jumps (yellow), smooth tracks (blue) and wildtype run-stop-run (green).

1. Jumps refer to an error in the formation of tracks from the collected centroids. Adjacent tracks can be erroneously joined together. As the mean speed for

R. sphaeroides is approximately 40 μm/s, any single transition that is significantly greater than this speed is probably due to a jump. The top speed cut-off is set to more than twice the average *R. sphaeroides* swimming speed, reflecting the large range in frame-to-frame speeds.

2. Truncated tracks are very short tracks. These result from a cell moving at an angle to the focal plane, so that it is within focus for only a small number of frames. These tracks can skew the data towards runs, as due to the time covered in the track they typically only contain runs. Truncated tracks can be removed either by setting a minimum number of frames for tracks, or by setting a minimum distance that a track must travel.

3. Within a population of motile cells, there is usually a number of nonmotile cells. These are dead or have lost their flagellum through the handling process. The majority of such nonmotile cells remain stationary through the length of a video due to the zero-flow conditions. Drifters are stopped cells that appear to move slightly, due to buffeting by motile cells passing close by and by Brownian motion. These tracks typically do not move very far.

   A track can be described by its minimum bounding radius (MBR) which is the radius of the smallest circle needed to include the entire track. Careful selection of an appropriate MBR can be used to remove nonmotile and drifter cells without removing legitimate motile but stoppy cells.

   The remaining nonmotile and drifter cells can pose a problem for later rounds of analysis. When track sets are classified by eye, these tracks are easily distinguished from motile neighbours. When the RunStopAnalysis system is used, remaining drifting cells have been found to be classified overwhelmingly as permanently stopped in later stages of classification, and thus do not introduce spurious stopping events into the data set (Rosser *et al.*, 2013). If an intermediate classification system is used, relying on summary statistics of tracks'

speeds and angle changes, large numbers of drifters and nonmotile cells can skew the data set.

Removing as many of these tracks as possible via the MBR is also desirable as this allows for removal of the next class of unwanted tracks, the corkscrew.

4. Corkscrew tracks result from a motile cell which rotates during forward motion in such a way as to rock the centre of the cell from side to side. In extreme cases, especially where the moving object is a cell in the process of dividing and thus likely off centre in its movement, the resulting track forms a distinct zigzag which can be mistakenly classified as a series of short runs and stops. Whereas jump tracks are artefacts of the tracking system and nonmotile/drifter tracks are biologically correct but not interesting tracks, corkscrews are valid tracks which may be biologically relevant. However, the analysis system used cannot yet accurately deal with corkscrews (Rosser *et al.*, 2013) and therefore are removed.

   Corkscrew tracks are typically not seen in strains with a stoppier swim than wildtype but are seen in smooth swimming strains. This is likely because with each stop, the flagellum and cell body reorientate with respect to one another, disrupting corkscrew-causing situations. In comparison, if a smooth swimming cell finds itself in a corkscrew orientation, it will remain like this for the duration of the track.

   The angle change in a track, or its tortuosity, can be summarised by the median absolute curvature, where curvature is calculated by

$$\kappa(\mathbf{r}_i) = \frac{\theta_i}{\|\mathbf{d}_{i-1}\| + \|\mathbf{d}_i\|}, \tag{3.3}$$

   where $\kappa(\mathbf{r}_i)$ is the curvature at position $\mathbf{r}_i$, $\mathbf{d}_{i-1}$ and $\mathbf{d}_{i-1}$ are the displacements preceding and following that position and θ_i is the angle between the two dis-

placements, following Figure 3.1.

A highly stoppy or drifting cell will have a higher median curvature than a corkscrew track, but a corkscrew will have a higher median curvature than a wildtype or smooth track. Thus, this measure can be used to remove corkscrews from smooth and wildtype data sets, provided the majority of drifters have already been removed.

All of the strains are filtered using curvature. As per Alon *et al.* (1998), tracks are sorted and the most tortuous are removed. This will remove some legitimate tracks from the stoppy strains, but also remove problematic tracks from the rest of the strains.

The challenge in cleaning data sets for later analysis is to remove sufficient of the artefact (jumps, truncated), uninteresting (nonmotile, drifter) and unusable (corkscrew) tracks to give clear results without losing significant numbers of the usable motile tracks. This is particularly the case for strains whose motility pattern is stoppier than wildtype. Using a high MBR removes more drifters, which ensures that the curvature cut-off percentage removes the corkscrews, rather than the remaining drifters, in wildtype and smooth swimming strains. However, too high an MBR will remove the motile but stoppy tracks. A graphical method for selecting an appropriate MBR is now presented, using the properties of a data set of known nonmotile cells (strain $CheY_6$(D57N)) to filter the nonmotile/drifter cells from a motile population.

A track can be described by two measures (no *et al.*, 2011). The mean absolute frame-wise angle change (MAC) of each track measures how curved a track is. The normalised effective mean speed (NEMS) for each track measures how quickly a cell track reached its destination. The NEMS is calculated by

$$\begin{aligned} \text{NEMS} &= \frac{\text{effective mean speed}}{\text{mean speed}} \\ &= \frac{\mathbf{d}_{final} - \mathbf{d}_0}{\text{mean speed}}, \end{aligned} \tag{3.4}$$

where $\mathbf{d}_0$ is the first displacement measured and $\mathbf{d}_{final}$ is the final displacement measured in a track. The closer to 1 the NEMS is, the more direct the route the cell took to reach its final destination.

The jump tracks, which are due to errors in tracking, are removed from the three standards data sets (wildtype, nonmotile and gutted/smooth). The MAC and NEMS are calculated for each and plotted against one another (Figure 3.3). The wildtype population of cells has two distinct clusters on the MAC-NEMS plot; high MAC low NEMS and low MAC high NEMS. The nonmotile strain has a single cluster – high MAC low NEMS – occupying the same space as the high MAC low NEMS cluster in the wildtype (Figure 3.3A and C). The high MAC low NEMS cluster on the wildtype plot (Figure 3.3A) therefore represents the dead and non-flagellate cells in the population. If an MBR is chosen that would filter out a significant proportion of the nonmotile strain – removing the single cluster on Figure 3.3C – then the same MBR would remove the majority of drifters and nonmotile cells from the wildtype or any other motile strains – removing the high MAC low NEMS cluster in Figure 3.3A.

However, drifters in the nonmotile strain are only generated by Brownian motion. Those in motile strains are also buffeted by motile cells around them. Drifters in a motile strain can move greater distances than in the nonmotile strain. The MBR should be chosen to stringently censor the nonmotile strain, ensuring that the majority of drifters will be removed in motile cells. The MBR is chosen to capture up to 99% of cells in the nonmotile strain CheY_6(D57N).

The few tracks remaining in the nonmotile standard data set after filtering with this MBR are found in the top left corner of the MAC-NEMS plot (Figure 3.3B). Subjected

to the same censoring protocol, the wildtype strain also has a small number of tracks in the same position (Figure 3.3D). Removing all of the truncated tracks of less than 1 s and the top 10% most tortuous tracks using median curvature removes all of the tracks from this position (Figure 3.3E).

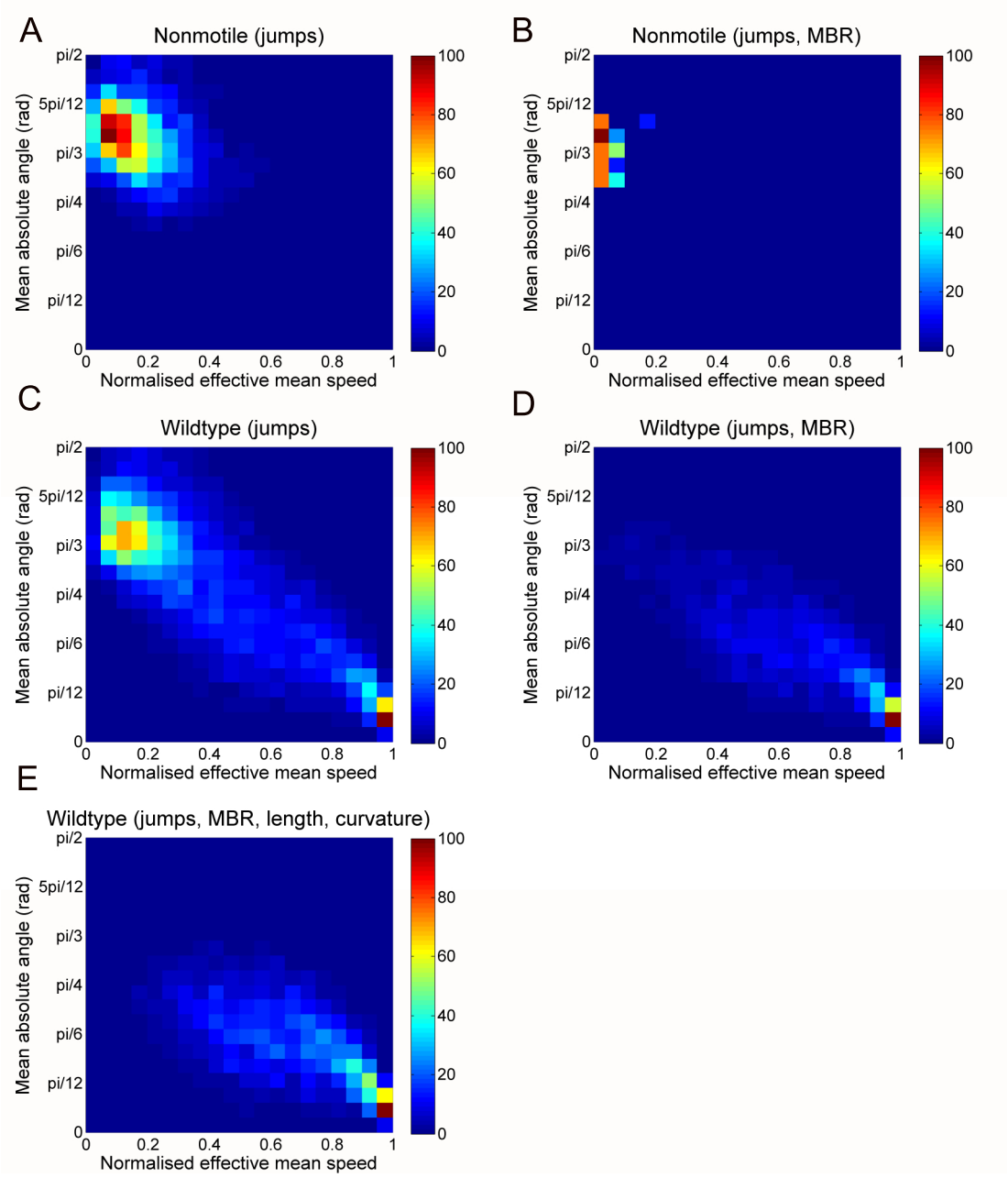


Figure 3.3: 400 bin two-dimensional histograms of the MAC-NEMS space occupied by nonmotile tracks (A) with only jumps removed and (B) after censoring all remaining tracks with MBR less than 9 μm, and of wildtype tracks (C) with only jumps removed, (D) after censoring the remaining tracks with 9 μm MBR and (E) after additional removal of tracks of less than 1 s and the 10% most tortuous tracks.

The final parameters used for censoring all data sets appear in Table 3.2 in the order in which the censoring is applied. The protocol is applied to all motile strains.

Table 3.2: Censoring parameters for sets of tracks from motile strains, before classification of tracks or strains.

Parameter	Cuts which cells	Value
Top speed cut-off	jumpers	90 μm/s
Minimum track length	truncated	50 frames or 1 s
MBR	nonmotile and drifters	9 μm
Percentage most tortuous	corkscrews and remaining drifters	10%

Classification

The cleaned data sets of tracks can now be analysed in a number of ways.

Comparison by hand If strains are sufficiently different from one another, these differences will be clear by eye. Representative tracks can be compared, as well as MAC-NEMS plots which provide a summary of the full data set. Due to the noise inherent in these data sets and polluting tracks remaining after censoring, if a difference cannot be seen by eye, it is unlikely that further statistical tests will discern a difference.

Summary statistics for whole tracks Data sets can be described by summary statistics, the choice of which is driven by what is seen by eye. This method avoids having to define what is a run or a stop, but rather uses how the characteristics of the two states will affect the overall behaviour of a track.

In *E. coli*, significant differences are seen in the mean track speeds of wildtype and smooth swimming cells (Alon *et al.*, 1998) due to the time spent in tumbles in wildtype. It is possible that the same difference exists between *R. sphaeroides* wildtype and smooth swimmers. It is likely that a stoppier strain would have a significantly smaller mean swimming speed than wildtype.

Tortuosity of tracks can be used to compare strains. A smooth strain, with its lack of stops, is likely to show less tortuosity than wildtype, which is turn should be less tortuous than a stoppy strain. Tortuosity can be described by mean absolute curvature.

Distributions of speed and angle change, whether frame by frame or averaged across a track, are non-normal, so all summary statistics and comparisons must use non-parametric tests.

Classify runs and stops Following Rosser *et al.* (2013), data sets of known non-motile and smooth swimming strains can be used to inform the classification of each track into runs and stops. The distributions of frame to frame speeds and angle changes in these two standards are used alongside an unknown strain data set to determine whether each framewise transition is part of a run or a stop.

Processing is done in two steps. The first uses a hidden Markov model to build probability functions for that particular unknown strain, calculating the probability of any particular frame transitioning from one state to the other. The second assigns an identity to each transition, comparing that frame's speed and angle change to the two known distributions and then taking into account the previous frame's state and the HMM's calculated probability of a switch out of that state.

Post processing tweaks these assignments force runs and stops to have a minimum length, usually of two frames each. This more accurately reflects the biological reality.

Once classification is complete, strains can be compared using summary statistics as above. The distributions to be compared are then the duration of runs and stops, the frequency of stops and the total fraction of time spent in runs and stops.

3.2.4 Analysis of free swimming *R. sphaeroides* mutants at steady state

In *E. coli*, the deletion of individual chemotaxis proteins alters the steady state tumbling frequency: deletion or inactivation of CheA, CheR or CheY all lead to smooth swimmers, whereas CheB or CheZ leads to tumbly swimmers. These phenotypes correspond to what is known about the network connectivity in *E. coli* chemotaxis: CheA, CheR and CheY all facilitate signalling from the chemoreceptors to the motor to cause a tumble and a direction change, whereas CheB and CheZ end the signal and allow a return to smooth swimming. In *R. sphaeroides*, however, strains with chemotaxis mutations have responses to dynamic changes that are different from wildtype, but not steady state swimming (see, for example, Porter *et al.* (2002)). Either these previous analyses are incorrect, or in *R. sphaeroides*, motility and chemotaxis are not linked in the same way as in *E. coli*, as changes in the chemotaxis network do not result in changes in the tumble/stop frequency at steady state.

R. sphaeroides chemotaxis mutant strains were selected that, as completely as possible, represented the entire of the *R. sphaeroides* chemotaxis network. Any overlap of phenotypes between the response regulator and methyltransferase deletions will provide clues as to how the system is wired. It is expected that the CheB and CheR homologues will form antagonist pairs, similarly to *E. coli*, whereas the CheY homologues' phenotypes should reflect each CheY's role in chemotaxis.

Single deletions of each CheB and CheR homologue, double deletions of each CheR pair and both CheBs, and a triple deletion of all three CheRs were selected. The three CheYs of importance, $CheY_3$, $CheY_4$ and $CheY_6$, were similarly tested. $CheY_6$ cannot be deleted without affecting transcription of the downstream *che*Op_3 genes. The $CheY_6$(D57A) mutant was used as a substitute (Porter *et al.*, 2006). To confirm the use of this mutation, both deletions of $CheY_{3,4}$ and $CheY_{3,4}$(D57N) mutants were used. The D57A mutation of all three was also used, which should permanently

bind to the motor (Appleby and Bourret, 1999; Porter *et al.*, 2006). Strains with deletions of $CheY_3$, $CheY_4$, and both $CheY_3$ and $CheY_4$, and mutations in each alone and the pair, were chosen, as both are required for chemotaxis.

To test the effects of the proteins' activity, not their structural function in forming a cluster, mutations were used along with deletions. CheA deletions were therefore not tested. CheA mutants with either no autophosphorylation activity (G470K) or no phosphotransferase activity (H48Q) were tested. Separating the two CheA functions allows the existence of a phosphorelay (a CheB phosphorylating a CheA H48 when G470 is not available) to be tested.

Video data of all of the motile strains were recorded and the tracks extracted and censored as described above. The number of tracks removed in each data set at each step and the final number and percentage of tracks surviving is listed in Table 3.3. There is a surprisingly large range of track numbers, considering that 18 min of video was taken of all strains grown to the same optical density range. It is possible that this is related to the motility of each strain. Censoring removes 80-90% of tracks in all strains, the majority of which are through the MBR parameter.

The data sets were initially clustered by eye into smoother and stoppier tracks. Figures 3.4 and 3.5 shows 25 randomly selected tracks from each strain.

Figure 3.4 shows the known smooth swimming strain, gutted for all of the chemotaxis genes. These tracks are clearly different from wildtype (Figure 3.5, top left). Smooth swimming is characterised by many long straight tracks. Some corkscrew tracks remain in the data set. There are a number of coiled tracks, which resemble the swimming pattern of a tethered cell. These are an extreme version of the corkscrew. All other strains with a similar collection of tracks are included in Figure 3.4.

The remaining strains are shown in Figure 3.5 along with wildtype. Some strains are noticeably stoppier than wildtype. These tracks travel shorter effective distances

Table 3.3: Number of tracks removed at each censoring round for each chemotaxis mutant. Numbers in brackets for the nonmotile strain indicate, for unused steps, the number of tracks that would be removed if that censoring step was used.

Strain	Starting tracks	Tracks removed at each stage				Remaining tracks
		Top speed	MBR	Length	Tortuosity	
Wildtype	13841	1989	8387	345	312	2808
Gutted	6672	908	4839	105	82	738
Nonmotile ($CheY_6$(D57N))	5327	706	(4578)	(0)	–	4621 (43)
$CheA_2$(H48Q)	7058	1027	4233	148	165	1485
$CheA_2$(G470K)	8622	1277	6050	90	121	1084
$CheA_2$(H48Q,G470K)	9771	1490	5723	222	234	2102
$CheA_3$(H48Q)	4680	601	3038	132	91	818
$CheA_4$(G470K)	3564	551	2065	120	83	745
$CheA_3$(H48Q), $CheY_6$(D57N)	5111	755	3180	90	109	977
$CheA_4$(G470K), $CheY_6$(D57N)	5782	855	3174	121	164	1468
$\Delta CheB_1$	5194	766	3134	119	118	1057
$\Delta CheB_2$	9833	1394	6103	249	209	1878
$\Delta CheB_1CheB_2$	8413	1198	5473	174	157	1411
$\Delta CheR_2$	10382	1543	7093	134	162	1450
$\Delta CheR_3$	4114	574	2773	90	68	609
$\Delta CheR_2CheR_3$	5070	672	3258	132	101	907
$\Delta CheR_1CheR_3$	5160	785	3429	106	84	756
$\Delta CheR_1CheR_2CheR_3$	7085	1045	4688	180	118	1054
$CheY_3$(D57N)	8170	1223	4901	173	188	1685
$CheY_3$(D57N) $\Delta CheY_{1,2,4,5}$	5930	865	3741	156	117	1051
$CheY_3$(D57A)	4687	650	2436	155	145	1301
$CheY_3$(D57A) $\Delta CheY_4$	8463	1027	6106	95	124	1111
$CheY_4$(D57A)	7795	1234	4339	171	206	1845
$CheY_3$(inactive) $CheY_4$(D57A)	9266	1211	5334	333	239	2149
$CheY_4$(D57N)	3887	518	2591	78	70	630
$CheY_3$(inactive) $CheY_4$(D57N)	3349	375	2072	134	77	691
$CheY_4$(D57N) $\Delta CheY_{1,2,3,5}$	5427	798	3495	130	101	903
$CheY_{3,4}$(D57A)	7677	1181	4746	179	158	1413
$\Delta(CheY_{3,4})$	10243	1684	6377	165	202	1815
$CheY_{3,4}$(D57N)	2657	392	1506	106	66	587
$CheY_6$(D57A)	4889	675	2948	176	109	981
$CheY_6$(D57N) $\Delta CheY_{1-5}$	8134	1022	7050	0	7	55
$\Delta CheY_4$	4811	606	3212	105	89	799
$\Delta CheY_{1-6}$	4663	678	2879	117	99	890
ΔTlpT	7306	934	4964	205	121	1082

and have more direction changes. However, drawing the distinction between wildtype and stoppy is very subjective and not as clear as selecting smooth tracks. The full data set is not taken into account, as it is difficult to visualise more than 25 tracks at a time.

As differences can be seen in the strains by eye, it is worth proceeding with further data analysis. Rather than using a summary statistic to compare the raw tracks data (analysis strategy 2), the tracks were first categorised into runs and stops (analysis strategy 3).

As an aside, the summary statistic often used for unclassified *E. coli* tracks is mean swimming speed of a track. However, the mean swimming speeds for the wildtype and gutted smooth swimming strains are not significantly different from one another. As these strains have significantly different swimming patterns as measured by the bead assay, this testing possibility is rejected.

The full set of censored nonmotile tracks was used to simulate a distribution of stop transitions. The gutted strain tracks were used to simulate a distribution of run transitions. The gutted tracks were further censored before use. Tracks for both sets were broken down into transitions. The nonmotile transitions Gaussian distribution was used to censor the gutted transitions bimodal distribution: a top speed of 22.1 μm/s (1.7 pixels/frame) removed all nonmotile transitions, so a lower speed of 22.1 μm/s was imposed on the gutted transitions.

These distributions direct the classification of each transition in each data set. A post-processing step was imposed, restricting both runs and stops to a minimum length of 2 frames, with stops taking priority (i.e.: short runs were changed first).

The resulting data sets have a number of incorrectly classified tracks which introduce a level of noise into the data. These tracks are typically the most tortuous tracks remaining in the data sets and are often classified as series of runs and stops, rather than one corkscrewed run. These anomalous stops mean that the summary

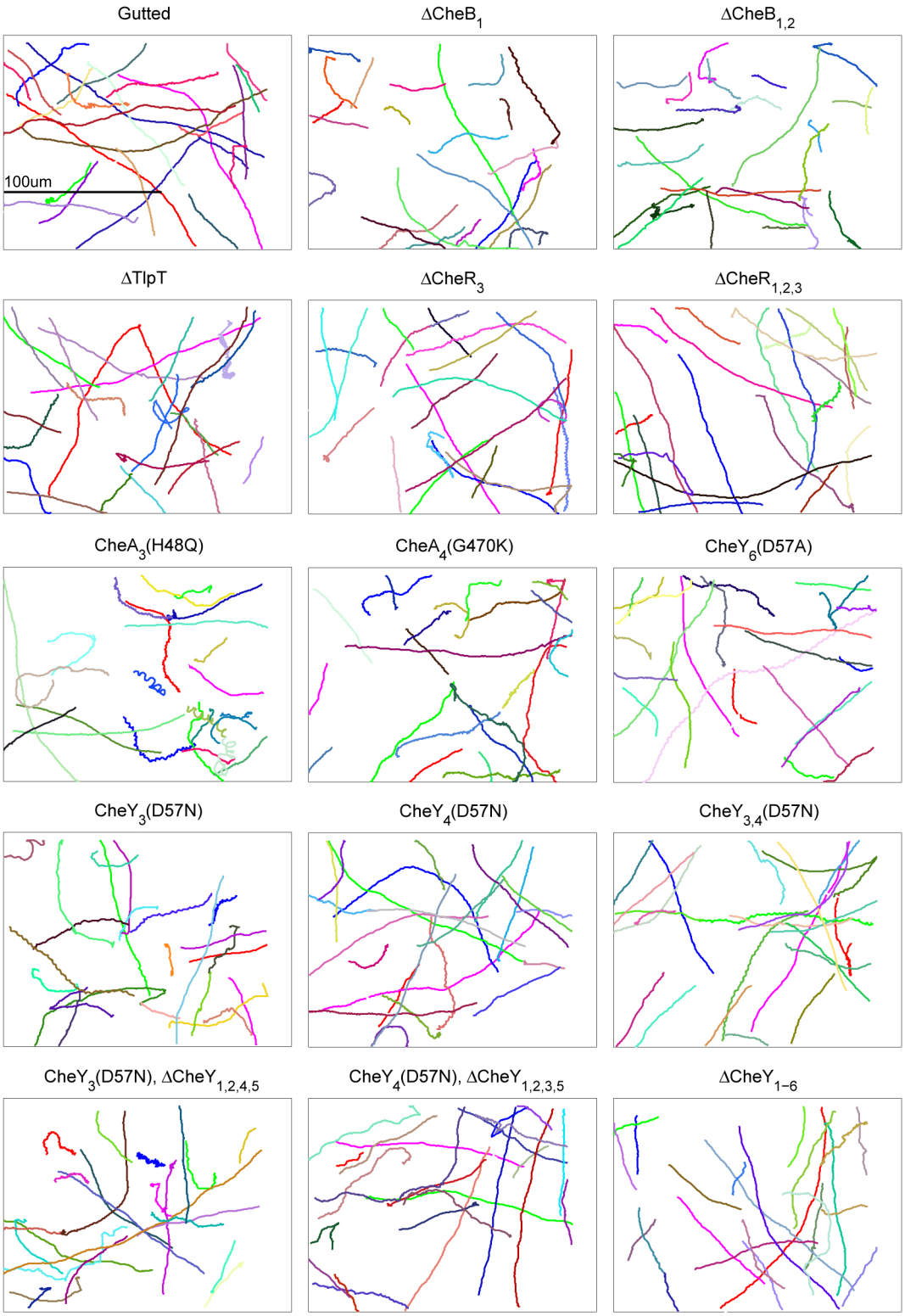


Figure 3.4: Representative tracks from strains which appear similar to the gutted strain. Black bar on gutted indicates 100 μm.

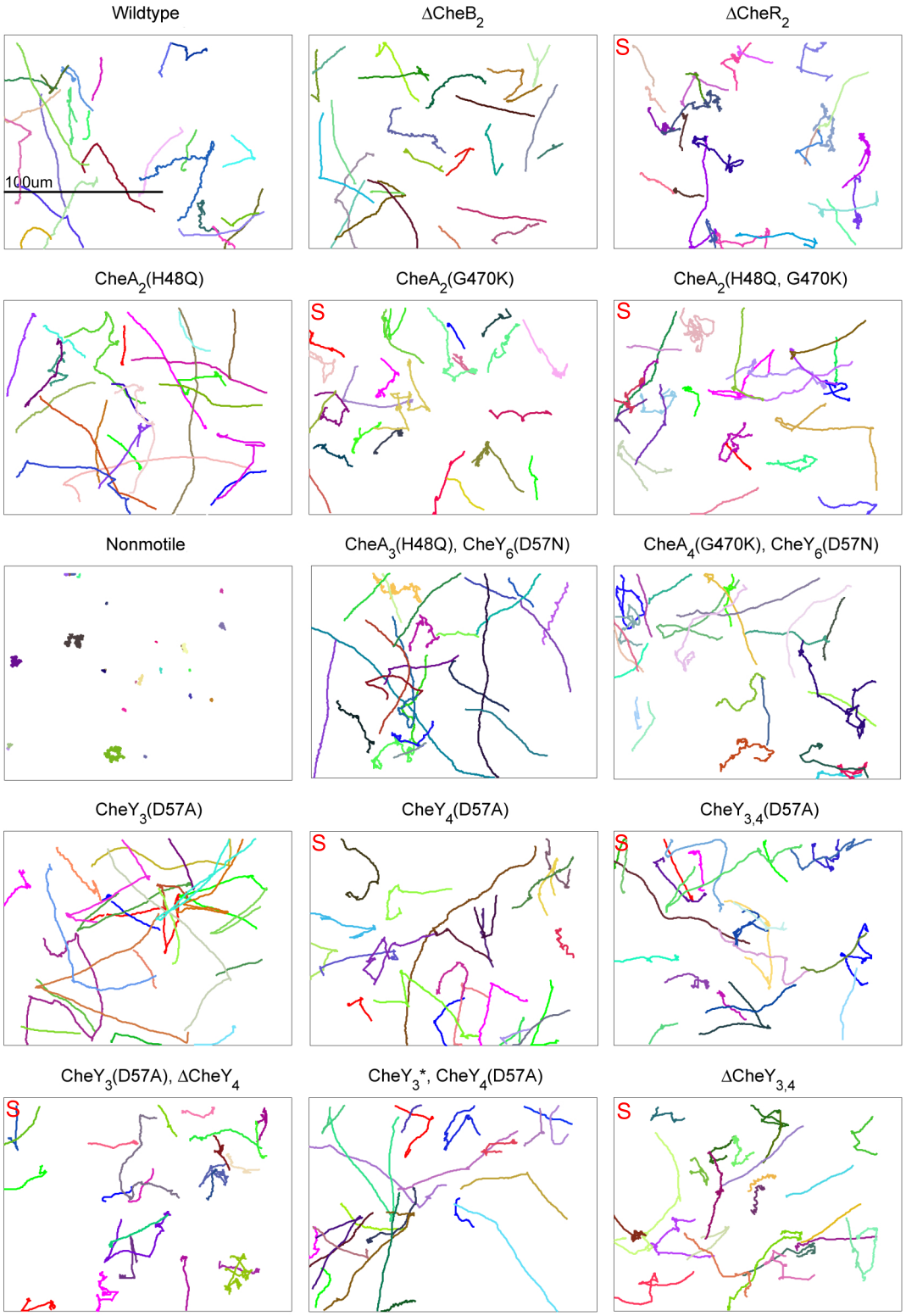


Figure 3.5: Representative tracks from strains which appear similar to or stoppier (red S in top left corner) than wildtype. Black bar on wildtype indicates 100 μm. * = partial deletion, sufficient to inactivate

statistic of 'number of stops transitions in a particular time' is unlikely to give a clear reflection of the differences between strains.

However, anomalous stops are typically very short in duration compared to true stops. Thus, a truly stoppy strain is likely to spend a greater proportion of its time stopped than a wildtype strain or a smooth swimming strain. The distributions of proportion of time stopped in each track do indeed differ (Figure 3.6).

The chemotaxis mutants have similar ranges of values for fraction of time stopped, as all cover almost the entire 0.0-1.0 range. However, there are two clear distribution groups. The strains identified as smooth by eye all have a lower box bound of zero and their medians appear lower than those of the wildtype and stoppy strains. The outlier here is $CheY_3$(D57N). By eye this was classified as smooth, but the distribution appears to match the wildtype/stoppy set.

A one-way Kruskal-Wallis test rejects the hypothesis that these distributions are from the same population ($p<0.001$). Pairwise comparisons are made post-hoc using a Dunn test with a null hypothesis that each pair is from the same population. The resulting probabilities for this hypothesis are shown in Table 3.4. For reasons of space, the set is split into two groups in this table, on the basis that every strain in the top table has $p<0.001$ probability of being in the same population as every strain in the bottom table. The resulting overlapping groups are shown in Table 3.5. The groups are ordered according to the magnitude of the median, as shown in Figure 3.6.

The classification scheme has grouped the strains into approximately six combined groups. Strains with a median time spent stopped higher than wildtype are grouped more clearly than those with median time less than wildtype. This is interesting, as by eye the stoppy tracks were harder to classify, whereas smooth tracks are easy to classify by eye.

Only one strain is not statistically different from wildtype, $CheY_3$(D57A), and this

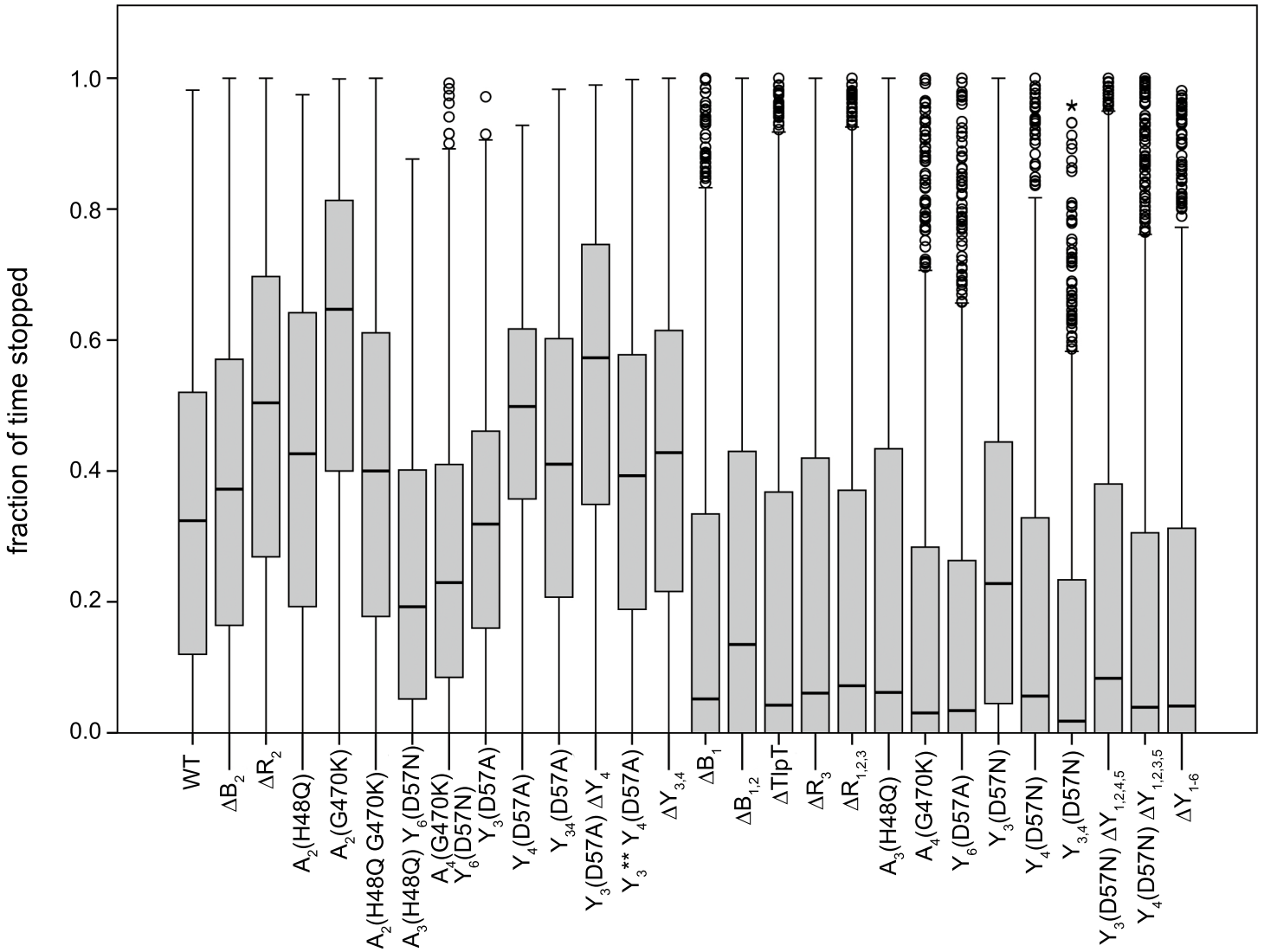


Figure 3.6: Fraction of time each chemotaxis mutant spends stopped. The boxplot edges indicate the first and third percentiles, the box centre indicates the median, the whiskers indicate the range $\pm$ 1.5x the box height, and the circles indicate outliers (Section 2.6.1).

Table 3.4: Pairwise comparisons (Dunn test) between distributions of fraction time stopped for each chemotaxis mutant, giving the probability that a pair is from the same population. The strains are split into two groups to fit the page. Every comparison between a member of the top table and of the bottom table is $p<0.001$. Wildtype is included in the top table to indicate this.

	WT	ΔB_1	$Y_{3,4}$ (D57N)	A_3 (H48Q)	A_4 (G470K)	Y_6 (D57A)	Y_3 (D57N)	Y_3(D57N) $\Delta Y_{1,2,4,5}$	Y_4 (D57N)	Y_4(D57N) $\Delta Y_{1,2,3,5}$	A_3(H48Q) Y_6(D57N)	A_4(G470K) Y_6(D57N)	ΔR_3	$\Delta B_{1,2}$	ΔTlpT	ΔY_{1-6}	ΔR_{1-3}
WT		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ΔB_1	0		0.801	1	1	1	0	1	1	1	0	0	1	0	1	1	1
$Y_{3,4}$(D57N)	0	0.801		0	1	1	0	0	0.179	0.531	0	0	0	0	0.086	1	0.001
A_3(H48Q)	0	1	0		0.072	0	0	1	1	1	0.085	0	1	1	1	1	1
A_4(G470K)	0	1	1	0.072		1	0	0.111	1	1	0	0	0.085	0	1	1	0.427
Y_6(D57A)	0	1	1	0	1		0	0	0.826	1	0	0	0.001	0	0.383	1	0.002
Y_3(D57N)	0	0	0	0	0	0		0	0	0	1	1	0	0.01	0	0	0
Y_3(D57N), $\Delta Y_{1,2,4,5}$	0	1	0	1	0.111	0	0		1	1	0.007	0	1	1	1	1	1
Y_4(D57N)	0	1	0.179	1	1	0.826	0	1		1	0	0	1	0.125	1	1	1
Y_4(D57N), $\Delta Y_{1,2,3,5}$	0	1	0.531	1	1	1	0	1	1		0	0	1	0.001	1	1	1
A_3(H48Q) Y_6(D57N)	0	0	0	0.085	0	0	1	0.007	0	0		1	0.635	1	0	0	0.001
A_4(G470K) Y_6(D57N)	0	0	0	0	0	0	1	0	0	0	1		0	0.002	0	0	0
ΔR_3	0	1	0	1	0.085	0.001	0	1	1	1	0.635	0		1	1	1	1
$\Delta B_{1,2}$	0	0	0	1	0	0	0.01	1	0.125	0.001	1	0.002	1		0.002	0	0.738
ΔTlpT	0	1	0.086	1	1	0.383	0	1	1	1	0	0	1	0.002		1	1
ΔY_{1-6}	0	1	1	1	1	1	0	1	1	1	0	0	1	0	1		1
ΔR_{1-3}	0	1	0.001	1	0.427	0.002	0	1	1	1	0.001	0	1	0.738	1	1	

	WT	$\Delta Y_{3,4}$	ΔR_2	Y_3(D57A) ΔY_4	Y_3* Y_4(D57A)	$Y_{3,4}$ (D57A)	A_2 (H48Q)	A_2 (G470K)	A_2 (H48Q,G470K)	A_4 (G470K)	ΔB_2	Y_3 (D57A)	Y_4 (D57A)
WT		0	0	0	0	0	0	0	0	0	0.009	1	0
$\Delta Y_{3,4}$	0		0	0	0.098	1	1	0	1	0	0	0	0
ΔR_2	0	0		0.02	0	0	0	0	0	0	0	0	1
Y_3(D57A), ΔY_4	0	0	0.02		0	0	0	0.003	0	0	0	0	0.292
Y_3*, Y_4(D57A)	0	0.098	0	0		1	0.249	0	1	0	1	0	0
$Y_{3,4}$(D57A)	0	1	0	0	1		1	0	1	0	0.16	0	0
A_2(H48Q)	0	1	0	0	0.249	1		0	1	0	0	0	0
A_2(G470K)	0	0	0	0.003	0	0	0		0	0	0	0	0
A_2(H48Q, G470K)	0	1	0	0	1	1	1	0		0	1	0	0
A_4(G470K)	0	0	0	0	0	0	0	0	0		0	0	0
ΔB_2	0.009	0	0	0	1	0.16	0	0	1	0		0.003	0
Y_3(D57A)	1	0	0	0	0	0	0	0	0	0	0.003		0
Y_4(D57A)	0	0	1	0.292	0	0	0	0	0	0	0	0	

Table 3.5: The chemotaxis mutants in statistically significant groups, based on Dunn test pairwise comparisons on the proportion of time spent stopped during steady state swimming in a homogeneous, zero-flow environments. The groups are split if $p<0.05$ for one pair. The combined groups merge groups that have overlapping members.

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Combined 1	Combined 2		Combined 3		Combined 4
MOST STOPPY	$CheA_2$(G470K)				
	$\Delta CheR_2$ $CheY_4$(D57A)	$CheY_4$(D57A) $CheY_3$(D57A) $\Delta CheY_4$			
			$CheA_2$(H48Q) $\Delta CheY_{3,4}$ $CheY_3$* $CheY_4$(D57A) $CheY_{3,4}$(D57A) $CheA_2$(H48,G470K)	$CheY_3$* $CheY_4$(D57A) $CheY_{3,4}$(D57A) $CheA_2$(H48Q,G470K) $\Delta CheB_2$	
					Wildtype $CheY_3$(D57A)

Group 7	Group 8	Group 9	Group 10	Group 11	
Combined 5	Combined 6				
$CheY_3$(D57N) $CheA_4$(G470K) $CheY_6$(D57N) $CheA_3$(H48Q) $CheY_6$(D57N)					
	$CheA_3$(H48Q) $CheY_6$(D57N) $\Delta CheB_{1,2}$ $\Delta CheR_3$ $CheA_3$(H48Q)	$\Delta CheB_{1,2}$ $\Delta CheR_3$ $CheA_3$(H48Q) $CheY_3$(D57N) $\Delta CheY_{1,2,4,5}$ $\Delta CheR_{1,2,3}$ $CheY_4$(D57N)	$\Delta CheR_3$ $CheA_3$(H48Q) $CheY_3$(D57N) $\Delta CheY_{1,2,4,5}$ $\Delta CheR_{1,2,3}$ $CheY_4$(D57N) $CheA_4$(G470K) $\Delta CheB_1$ $CheY_4$(D57N) $\Delta CheY_{1,2,3,5}$ $\Delta TlpT$ $\Delta CheY_{1-6}$	$CheY_4$(D57N) $CheA_4$(G470K) $\Delta CheB_1$ $CheY_4$(D57N) $\Delta CheY_{1,2,3,5}$ $\Delta TlpT$ $\Delta CheY_{1-6}$ $CheY_{3,4}$(D57N) $CheY_6$(D57A)	MOST SMOOTH

strain was one of those that by eye could not be distinguished from wildtype.

Smooth vs. stoppy swimmers

There are two groups of strains with median time stopped less than wildtype (Combined groups 5 and 6, Table 3.5). Combined group 6 has median time significantly lower, and close to zero, than combined group 5. This, larger, group of strains show truly smooth swimming and likely cannot stop. Stops seen in this group are likely a result of classification of noisy data or corkscrews, or of damage to flagella. The two groups do have one overlapping strain, the double mutant $CheA_3$(H48Q) $CheY_6$(D57N). However, its matches in combined group 6 are all relatively low ($p<0.01$), barring the strain with the double deletion of the two CheBs.

There are three groups of strains with median time stopped greater than wildtype (Combined groups 1, 2 and 3, Table 3.5), with each median coming closer to 1.0, which would be indicative of a stopped set.

The grouping of the strains' median time stopped can be summarised in terms of their physiology (Table 3.6). The results of each subset of mutants are now discussed in detail.

Table 3.6: Summary of the broad effects of chemotaxis network mutations on swimming behaviour.

Part of network	Smooth	Stoppy
Signalling	Inactivate cytoplasmic kinase	Inactivate polar kinase
	Active $CheY_3$ and $CheY_4$	Inactive or deleted $CheY_3$ and $CheY_4$
	Inactive $CheY_6$	Active $CheY_6$ (stopped)
Adaptation	Insensitive cytoplasmic cluster	Highly sensitive cytoplasmic cluster
	Highly sensitive polar cluster	Insensitive polar cluster

Signalling: The CheAs

Chemotaxis can be disrupted by the deletion of either cluster or the inactivation of the kinases in either cluster. The effects on steady state swimming when altering the

two clusters are opposing (Figure 3.7).

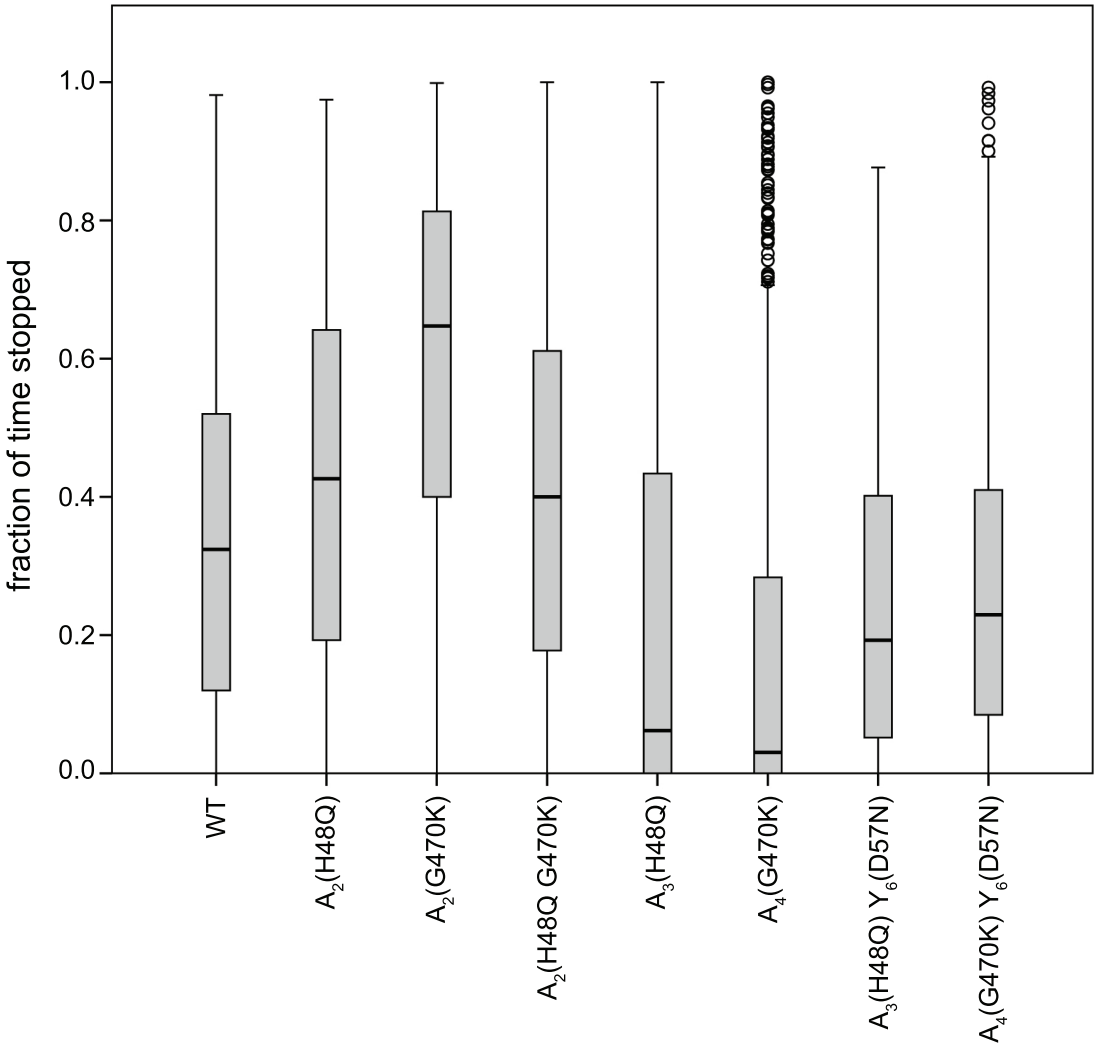


Figure 3.7: Fraction of time that each CheA chemotaxis mutant spends stopped. The boxplot edges indicate the first and third percentiles, the box centre indicates the median, the whiskers indicate the range $\pm$ 1.5x the box height, and the circles indicate outliers (Section 2.6.1).

When the polar cluster kinase, $CheA_2$, is rendered incapable of transferring a signal from the chemoreceptors by phosphorylating its response regulators, steady state swimming becomes significantly stoppier than wildtype ($CheA_2$(H48Q), $CheA_2$(G470K) and $CheA_2$(H48Q, G470K)). Removing the autophosphorylation activity of $CheA_2$ ($CheA_2$ (G470K)) results in a significantly stoppier phenotype than both wildtype and the removal of the transphosphorylation activity ($CheA_2$(H48Q) and $CheA_2$(H48Q, G470K)). It is possible that something else is able to transfer phosphoryl groups to

$CheA_2$'s H48Q, allowing signalling to occur through the polar kinase without any action by the chemoreceptors. Why this activation by another source would result in the same phenotype as the inactivating mutations is not known.

Conversely, when the cytoplasmic kinases, $CheA_3$ and $CheA_4$, are rendered incapable of transferring a signal from the chemoreceptors by phosphorylating their response regulators, steady state swimming becomes significantly smoother than wildtype ($CheA_3$(H48Q) and $CheA_4$(G470K)). There is no significant difference in the swimming of mutants without the ability to autophosphorylate the kinase pair ($CheA_4$(G470K)) or transphosphorylate a response regulator ($CheA_3$(H48Q)). The double mutant $CheA_3$(H48Q), $CheA_4$(G470K) was therefore not tested.

The $CheY_6$(D57N) mutation alone results in a permanently stopped mutation, due to a permanent alternative phosphorylation event. If $CheY_6$ is mutated to $CheY_6$(D57N) in the presence of either of the cytoplasmic kinase mutations, the steady state swimming becomes stoppier. It is still less stoppy than wildtype, but some stops are restored. Thus, the loss of $CheY_6$ phosphorylation via the cytoplasmic cluster is sufficient to restore motility, which agrees with the experimental results of Porter *et al.* (2006). The polar cluster may contribute some phosphorylation of $CheY_6$, as the $CheY_6$(D57N), cytoplasmic kinase double mutations are not smooth swimming. However, the rate of phosphorylation must be very low; even though $CheY_6$(D57N) is permanently, not transiently, phosphorylated, the polar kinase cannot sufficiently phosphorylate it to render a stopped phenotype. It is possible that the few stops seen are not due to polar kinase phosphorylation of $CheY_6$(D57N), but rather weak binding by the unphosphorylated mutant CheY to the flagellar motor.

$CheA_3$'s phosphatase activity is still present in both the cytoplasmic kinase mutations. It can be argued that the $CheA_3$ dephosphorylates $CheY_6$(D57N), allowing the smooth swimming seen in the $CheY_6$(D57N) plus cytoplasmic kinase mutation strains. However, if $CheA_3$ was actually able to dephosphorylate $CheY_6$(D57N), then that mutation should not be able to cause the completed non-motile phenotype

seen. It can therefore be assumed that $CheA_3$ is unable to remove the permanent Ser83-P event from $CheY_6$(D57N).

Signalling: The CheYs

As expected from our current understanding of how stops are produced, smooth swimming is seen in the absence of all of the CheYs (Figure 3.8). Inactivation of just $CheY_6$ ($CheY_6$(D57A)) is sufficient to cause smooth swimming, as when the cytoplasmic kinases are lost. Loss or inactivation of $CheY_3$ and $CheY_4$ (through the D57A mutation) results in a stoppy swimming pattern, as when the polar kinase is lost. This could be indicative of these proteins' role as a phosphate sink (Sourjik and Schmitt, 1998): without the additional route for phosphorylation, $CheY_6$ is over-phosphorylated, changing the bias of stops.

If $CheY_3$ and $CheY_4$ only function as a phosphate sink and have no function at the flagellar motor, any mutation which freezes these proteins into one state should lead to the same stoppy swimming. The D57N mutation should result in permanent phosphorylation at an alternative site. $CheY_3$(D57N) and $CheY_4$(D57N) then should not be able to collect and pass on phosphoryl groups, the dynamic role required for a phosphate sink. Instead, these mutants have a smooth swimming behaviour.

The working model for *R. sphaeroides* chemotaxis and motility assumes that only $CheY_6$-P can bind to the flagellar motor at physiologically relevant rates, initiating a stop. However, the antagonistic effects of the D57A and D57N mutations of $CheY_3$ and $CheY_4$ suggest that these proteins can in fact bind to the motor when phosphorylated, or in some other way block binding of $CheY_6$-P, causing a run. This is in agreement with Ferre *et al.* (2004), who measured *in vitro* binding between FliM and all six *R. sphaeroides* CheY proteins when phosphorylated.

Thus far, the results suggest that the two clusters interact with the flagellum in an antagonistic manner during steady state swimming. The cytoplasmic cluster creates

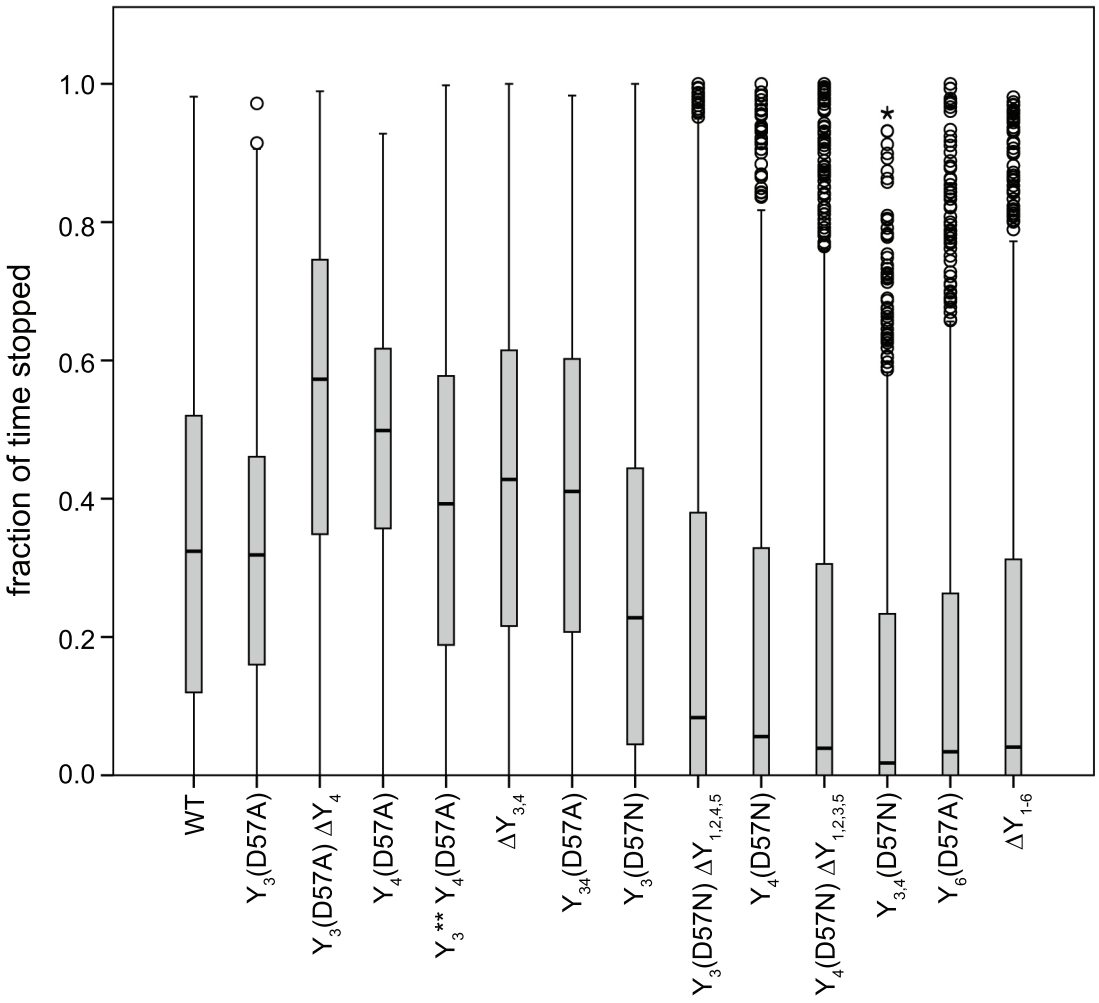


Figure 3.8: Fraction of time that each CheY chemotaxis mutant spends stopped. The boxplot edges indicate the first and third percentiles, the box centre indicates the median, the whiskers indicate the range $\pm$ 1.5x the box height, and the circles indicate outliers (Section 2.6.1).

the stopping events, whereas the polar cluster returns the flagellum to a smooth swimming state. If this is true, then $CheY_3$ and $CheY_4$ copy numbers and $CheY_3$-P and $CheY_4$-P binding rates to FliM must be tuned, along with those of $CheY_6$, so that the $CheY_3$-P and $CheY_4$-P produced on attractant addition is sufficient to overcome the basal levels of $CheY_6$-P.

Adaptation proteins

Again, the effects seen on the adaptation protein pairs connected with the two clusters are opposite (Figure 3.9). The cytoplasmic cluster pair shows the same pattern as in *E. coli*: deletion of $CheB_2$ increases stops, whereas deletion of $CheR_3$ gives a smooth phenotype. The polar cluster has an inverted response: when $CheR_2$ is deleted, stops increase and when $CheB_1$ is deleted, a smooth swimming pattern is seen.

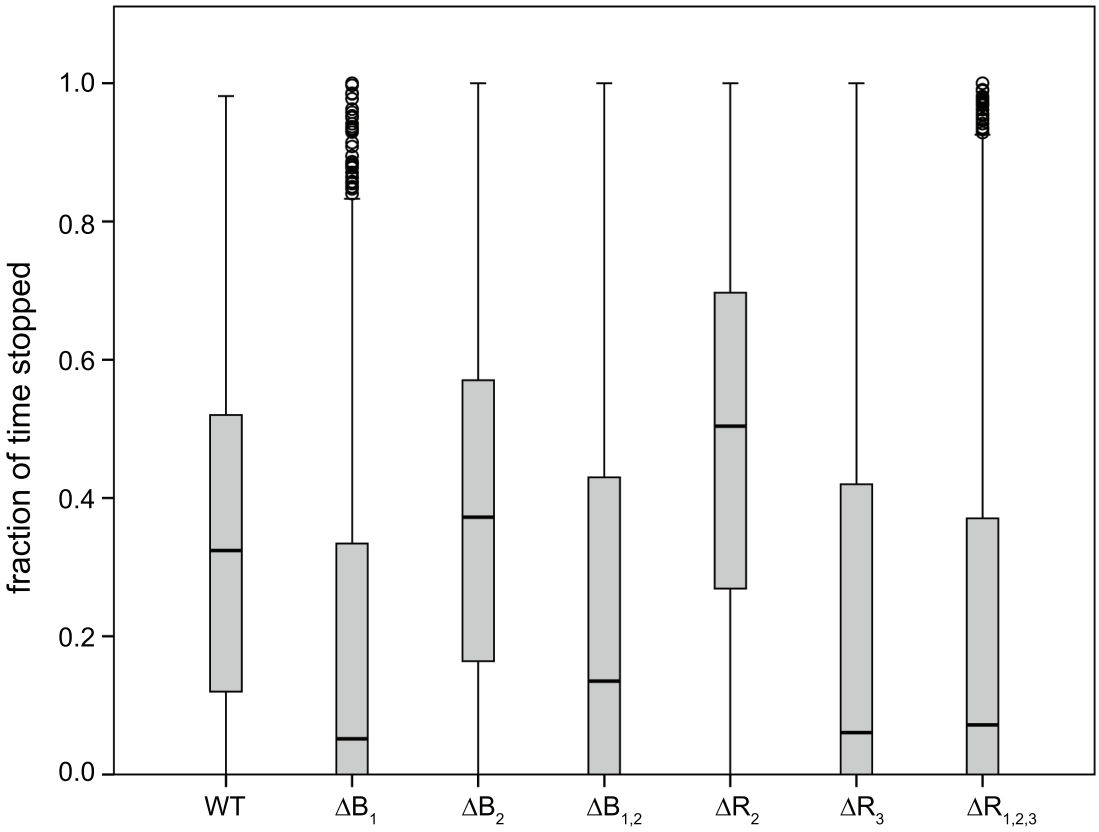


Figure 3.9: Fraction of time that each adaptation chemotaxis mutant spends stopped. The boxplot edges indicate the first and third percentiles, the box centre indicates the median, the whiskers indicate the range $\pm$ 1.5x the box height, and the circles indicate outliers (Section 2.6.1).

If the two pairs of CheBs and CheRs operate as in *E. coli* (see Figure 1.9) to desensitise and sensitise their respective chemoreceptors, then the opposing responses matches the signalling mutant results. The network connectivities of the CheBs are not known. However, the pairing used here is the most likely pairing. In *E. coli* and

other species, deletion of the adaptation protein pair causes antagonistic responses. The localisation of $CheR_2$ and $CheR_3$ are known. Thus, $CheB_1$ must pair with $CheR_2$, as this forms a smooth/stoppy deletion pair and similarly $CheB_2$ with $CheR_3$. This agrees with the gene positions in operons. Additional connectivity between $CheB_2$ and the polar cluster has been suggested and is not ruled out by the above.

$CheR_2$ sensitises the polar cluster, leading to more $CheY_3$-P and $CheY_4$-P and runs. The deletion of $CheR_2$ prevents the polar cluster from signalling, leading to more stops. $CheB_1$ desensitises the polar cluster, stopping $CheY_3$ and $CheY_4$ phosphorylation and allowing steady state stops. Without $CheB_1$, an overproduction of these phosphorylated CheYs leads to smooth swimming.

$CheR_3$ sensitises the cytoplasmic cluster, leading to more $CheY_6$-P and stops. The deletion of $CheR_3$ prevents the cytoplasmic cluster from signalling, leading to smooth swimming under steady state conditions. This reinforces that the polar cluster does not produce $CheY_6$-P at physiologically relevant rates at steady state: $CheR_3$ localises away from the polar cluster, so its deletion cannot affect the polar kinase, and yet when $CheR_3$ is deleted, no stops are seen. $CheB_2$ desensitises the cytoplasmic cluster, stopping $CheY_6$ phosphorylation and allowing steady state stops. Without $CheB_2$, an overproduction of $CheY_6$-P leads to more stops.

Network connectivity

The results have suggested how the two clusters function under steady state conditions (Figure 3.10-3.13). Steady state stops are due entirely to $CheY_6$-P from the cytoplasmic cluster, not the polar cluster. $CheY_3$-P and $CheY_4$-P are needed to modulate the frequency of stops, producing runs. $CheY_3$-P and $CheY_4$-P are able to interact with the motor, possibly by direct binding, and whether the motor biases to a run or a stop is dependent on the interaction of all three phosphorylated proteins.

If it is assumed that the polar cluster is able to sense and respond to the external

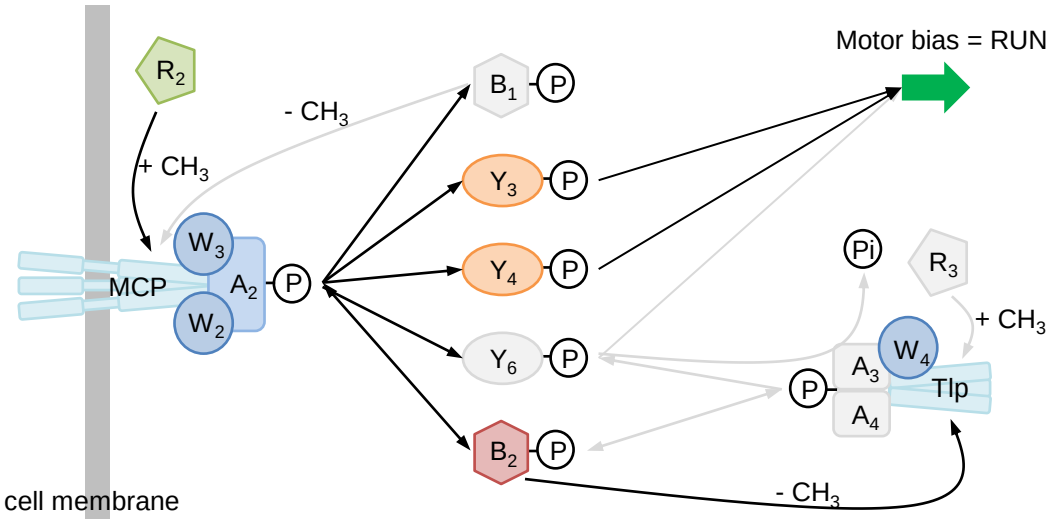


Figure 3.10: Deletion or inactivating mutation of the grey proteins leads to a loss of the greyed out reactions, biasing the motor towards runs.

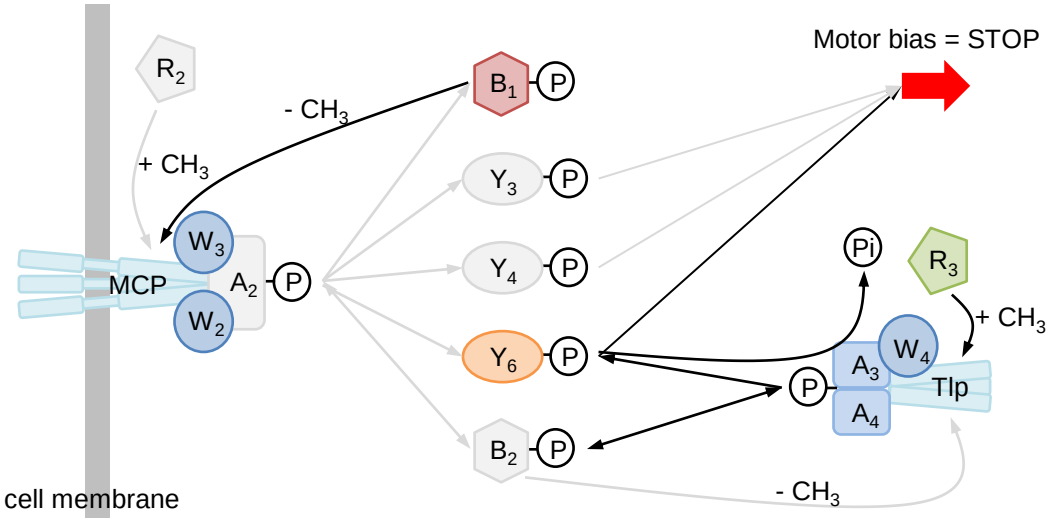


Figure 3.11: Deletion of the grey proteins leads to a loss of the greyed out reactions, biasing the motor towards stops.

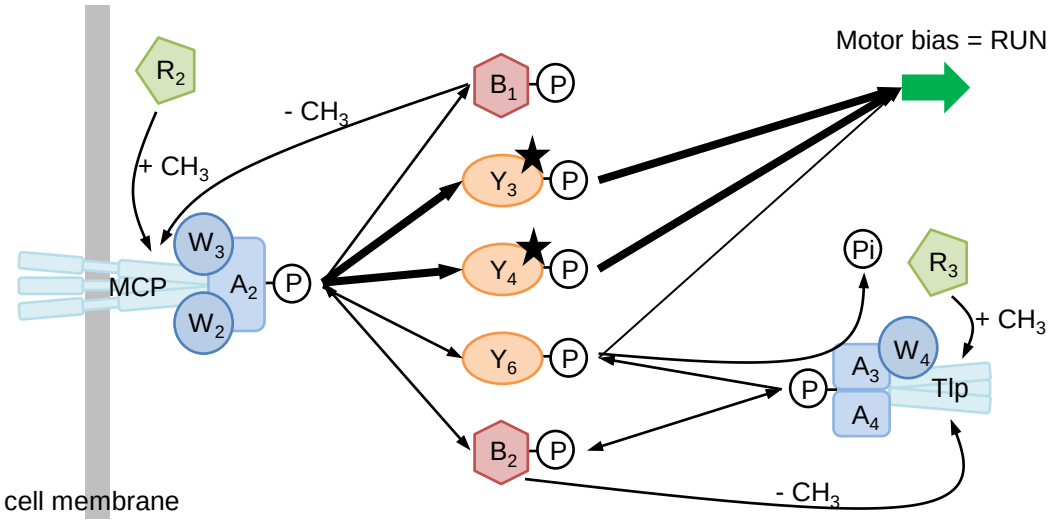


Figure 3.12: Mutations mimicking phosphorylation (black stars) that bias the motor towards runs.

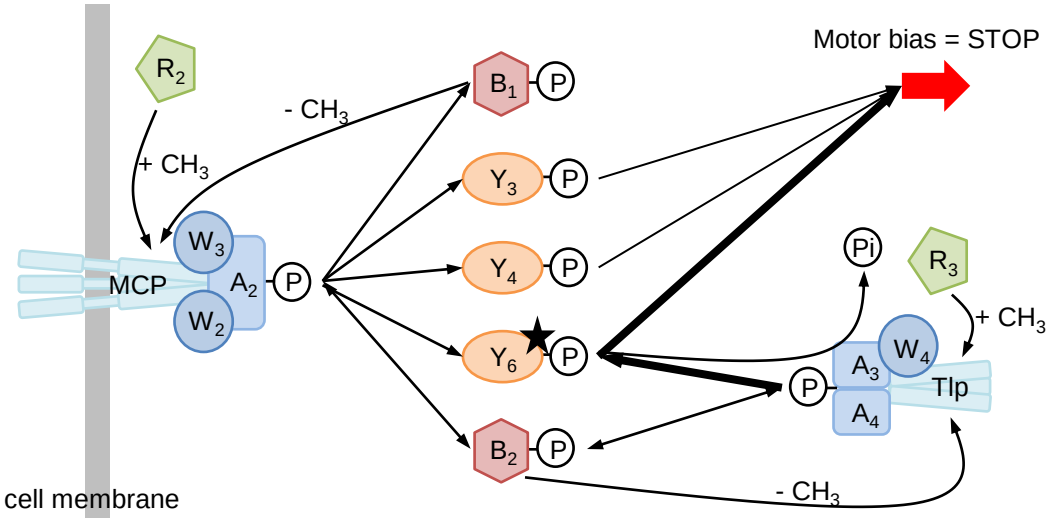


Figure 3.13: Mutations mimicking phosphorylation (black stars) that bias the motor towards stops.

environment, it may also be interacting with the motor solely through $CheY_3$ and $CheY_4$ during dynamic events. The polar cluster may be responding to increasing chemoeffector concentrations by producing $CheY_3$-P and $CheY_4$-P, causing a run. This pattern of interaction is reminiscent of *B. subtilis* where CheY-P binding to the motor results in runs (Rao *et al.*, 2008).

If the cytoplasmic cluster produces steady state stops and the polar cluster operates through $CheY_3$ and $CheY_4$ during dynamic swimming, what is the cytoplasmic cluster's role in dynamic swimming? It could be responsible for setting the levels of $CheY_6$-P, thus influencing the frequency of stops during steady state and the threshold for when the polar cluster's signalling drops sufficiently in response to a chemoeffector drop to produce a complete stop. It may respond to internal metabolite levels, tuning $CheY_6$-P levels in response. It also may be responding to chemoeffector changes itself, either independently or dependent on polar cluster signalling. Previous work has shown that, under photoheterotrophic conditions, the deletion of $CheA_2$ does not prevent a response to a dynamic change in chemoeffector (propionate) (Martin *et al.*, 2001a). Thus, at least under photoheterotrophic conditions, a stop can be produced by the cytoplasmic kinase alone, without the need for $CheY_3$ and $CheY_4$. However, under aerobic conditions, the loss of $CheA_2$ is sufficient to remove the response to dynamic changes in the environment. Only aerobic conditions are considered here.

An in-depth study of the dynamic behaviour of some of the *R. sphaeroides* chemotaxis mutants here will shed further light into the network connectivity of the system. In the discussion that follows, the basic model derived from the phosphorylation and localisation work described in depth in Section 1.4 is assumed. Most importantly, it is assumed that increased methylation increases the probability that a chemoreceptor cluster will change conformation and trigger CheA kinase activity. It is also assumed that $CheR_2$ is a methyltransferase that methylates the polar cluster, increasing the probability of $CheA_2$ autophosphorylation, and that $CheR_3$ is a

methyltransferase that methylates the cytoplasmic cluster, increasing the probability of $CheA_4$ transphosphorylating $CheA_3$.

1. Does the polar cluster respond to drops in chemoeffector by phosphorylating $CheY_6$, producing a stop?

   (a) If so, the deletion of $CheR_3$ should not affect the ability to respond to dynamic changes in chemoeffector concentration despite the lack of $CheY_6$-P under steady state conditions, as the polar cluster should produce its own $CheY_6$-P under dynamic conditions.

2. Does the cytoplasmic cluster, not the polar cluster, produce the $CheY_6$-P required to cause a stop, either by raising the baseline levels of $CheY_6$-P sufficiently for additional $CheY_6$-P from the polar cluster to trigger a stop, or as the sole producer of $CheY_6$-P?

   (a) If so, the deletion of $CheR_3$ should remove the ability to respond to dynamic changes in chemoeffector concentration. The lack of $CheY_6$-P from the cytoplasmic cluster should prevent a stop.

3. Does the polar cluster respond to increases in chemoeffector by phosphorylating $CheY_3$ and $CheY_4$, producing a run?

   (a) If so, then if $CheY_3$ and $CheY_4$ are not available for phosphorylation (D57A), a decrease in chemoeffector concentration should result in an altered adaptation pattern: a stop will be produced as normal from either cluster's action, but the return to run will take longer than normal, as $CheY_6$-P must be dephosphorylated following the stop, rather than $CheY_3$-P or $CheY_4$-P competing to bind and producing the run.

4. Does the cytoplasmic cluster sense and respond to external chemoeffector concentrations independently of the polar cluster under aerobic conditions, as in photoheterotrophic conditions (Martin *et al.*, 2001a)?

   (a) If so, the deletion of $CheR_2$, which prevents the polar cluster from responding to stimuli, retain the ability to respond to dynamic changes, as the cytoplasmic cluster will produce $CheY_6$-P and effect a stop.

   (b) If 4a is true and $CheY_3$-P/$CheY_4$-P do not outcompete $CheY_6$-P, the deletion of $CheB_1$ should not remove the ability to respond to dynamic changes.

5. Is the cytoplasmic cluster required for chemotaxis under aerobic conditions solely to produce a baseline of $CheY_6$-P, while the polar cluster responds to the external chemoeffector concentration and dynamically produces phosphorylated CheYs to effect runs and stops?

   (a) If so, $CheY_6$(D57N) can be substituted to produce the baseline of $CheY_6$-P, as it is permanently phosphorylated and can produce stops in the absence of functional $CheA_3CheA_4$. The removal of functional $CheA_3CheA_4$ along with $CheY_6$(D57N) should not remove the ability to respond to dynamic changes, especially if the polar cluster relies on $CheY_3$-P and $CheY_4$-P to exert its influence, rather than $CheY_6$-P.

6. Does $CheB_2$ function as a methylesterase at the cytoplasmic cluster? If so:

   (a) The deletion of $CheB_2$ will result in high levels of $CheY_6$-P from the cytoplasmic cluster, producing the stoppy steady state swimming phenotype seen here.

   (b) If the polar cluster can respond to the chemoeffector concentration without the action of the cytoplasmic cluster, then a dynamic response should still be seen in a $CheB_2$ deletion. On an increase of chemoeffector, either the polar cluster will cease producing $CheY_6$-P, dropping levels sufficiently to transiently relieve the stoppy phenotype, or the polar cluster will produce $CheY_3$-P and $CheY_4$-P, which will bind to the motor and produce a run, transiently relieving the stoppy phenotype.

The above is by no means an exhaustive list of all of the possibilities for network connectivity. These are merely some of the most obvious possibilities arising from the free swimming data, which can easily be preliminarily tested using the dynamic behaviour of chemotaxis mutant strains with mutations in the four adaptation proteins.

3.3 Dynamic swimming behaviour: tethering analysis of population-sized samples

Swimming behaviour in response to changes in the external environment can be measured by tethering a cell and tracking its rotation through the addition and removal of an attractant. Analysis methods in the past have suffered from high levels of noise in the resulting rotation trace and have restricted the number of cell that can be analysed in one frame at a time.

A novel analysis method allows an entire field of tethered cells to be analysed simultaneously, which can generate data for up to forty cells at a time (Kojadinovic *et al.*, 2011). Noise levels are reduced without losing signal by transforming rotation traces with a Fourier transformation. Thus, detailed cell-level behaviour can be captured using population-sized samples in a time-efficient manner.

Although the method for data capture and software for data extraction have been developed, the method has not yet been used to compare different strains, or used to answer a specific biological question. As the data capture method was used extensively in this study, a data analysis protocol was developed. Areas of future development to extend the method are suggested.

Twelve strains were analysed using this method. WS8N is the baseline wildtype response for comparison with each chemotaxis mutant and $CheY_6$(D57A) is the baseline unresponsive strain. Dynamic responses of the rest (Table 3.7) will invalidate

some network connectivity possibilities, as detailed above.

3.3.1 Tethering data capture and classification protocol

The experimental details of tethering analysis are included in Section 2.5.3. Cells were grown to mid-log phase, then treated with Chl to halt protein production and growth, and transferred to motility buffer. Cells were incubated on a small glass slide with anti-flagellum antibody, sticking cells by their flagella to the slide. The slide was then inserted into a flow chamber. The cells were challenged first by buffer containing attractant, then by the removal of that attractant. Cell behaviour was recorded on high-speed video.

Information on the rotating cells was then extracted using the software BRAS. This requires user input, selecting rotating cells for analysis. In this study, any cell that showed rotation at any point during the experiment was selected for analysis. Cells that were stopped for the duration of the experiment were discarded. Permanently stopped cells may have been damaged during processing, or may represent a biologically relevant phenotype, but are beyond the scope of this study. Figure 3.14 shows a representative field of view.

The raw rotation output is highly variable and noisy. The companion software Click&Mean uses a Fourier transformation to smooth the data, generating clear rotation signals without requiring the data averaging that is often used, losing detail on short events.

A subjective form of user input is then required. The Click&Mean output does not classify cells in any way. Whether the rotational signal is clear enough to use and what phenotype is shown must be decided by the user. The obsolete method used raw rotational signal for a few clear cells and relied on the same subjective decision-making. The difference here is that in allowing all rotating cells to be brought to this stage of analysis, a clearer picture of the variety of phenotypes should be available.

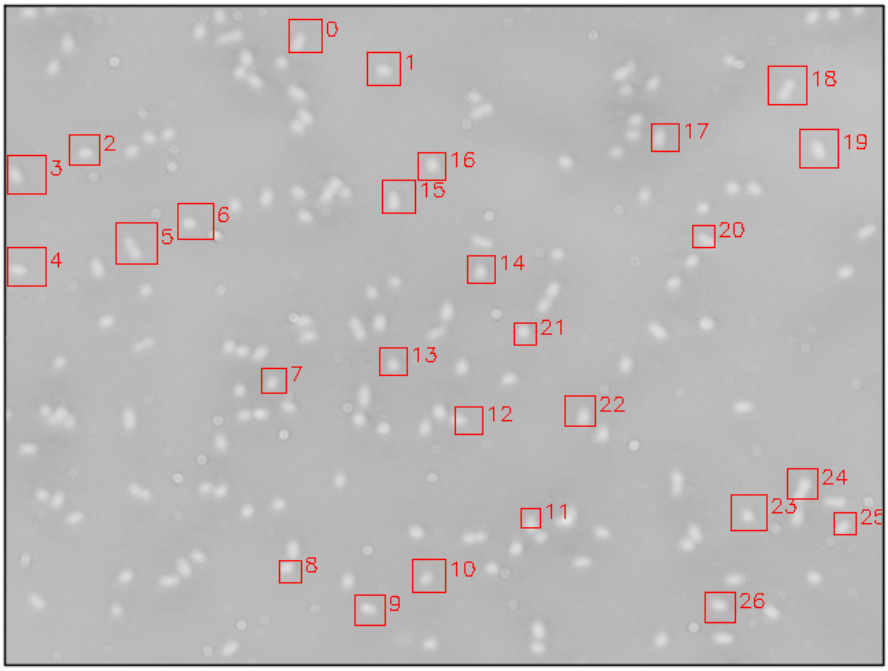


Figure 3.14: Rotating tethered cells in one field of view, selected for BRAS analysis.

Cells without a clear rotational signal were immediately discounted. Cells were then classified into phenotypes on the basis of what is known about chemotaxis in *R. sphaeroides* and other species and, more importantly, upon the variability seen in the Click&Mean outputs obtained. Examples of each phenotype are shown in Figure 3.15.

Tethering phenotypes recorded in previous work studying wildtype and chemotaxis deletion mutants:

1. Does not respond to addition or removal of attractant (unresponsive): cell rotates clearly throughout the experiment (Figure 3.15A).

2. Responds to removal of attractant, does not recover (responsive): cell ceases rotation around 8m00s - 8m45s, and does not restart rotation within the 13m00s or 18m00s experiment (Figure 3.15B).

3. Responds to removal of attractant, then adapts to recover (adaptive): cell ceases rotation around 8m00s - 8m45s, and restarts rotation before the end of the ex-

periment (Figure 3.15C).

Unexpected tethering phenotypes:

4. Generally 'stoppy' motion: as is known from the free swimming analysis, some strains show a stronger bias towards stops than runs in the absence of stimulus (Figure 3.15D).

5. Responds to the addition of attractant (inhibited): cells begin rotation on addition of propionate (around 3m00s - 3m45s). Some of these cells cease rotation before propionate is removed, some do not respond to propionate removal and others are able to adapt (Figure 3.15E).

6. Responds to addition of attractant: cells cease rotation on addition of propionate (Figure 3.15F).

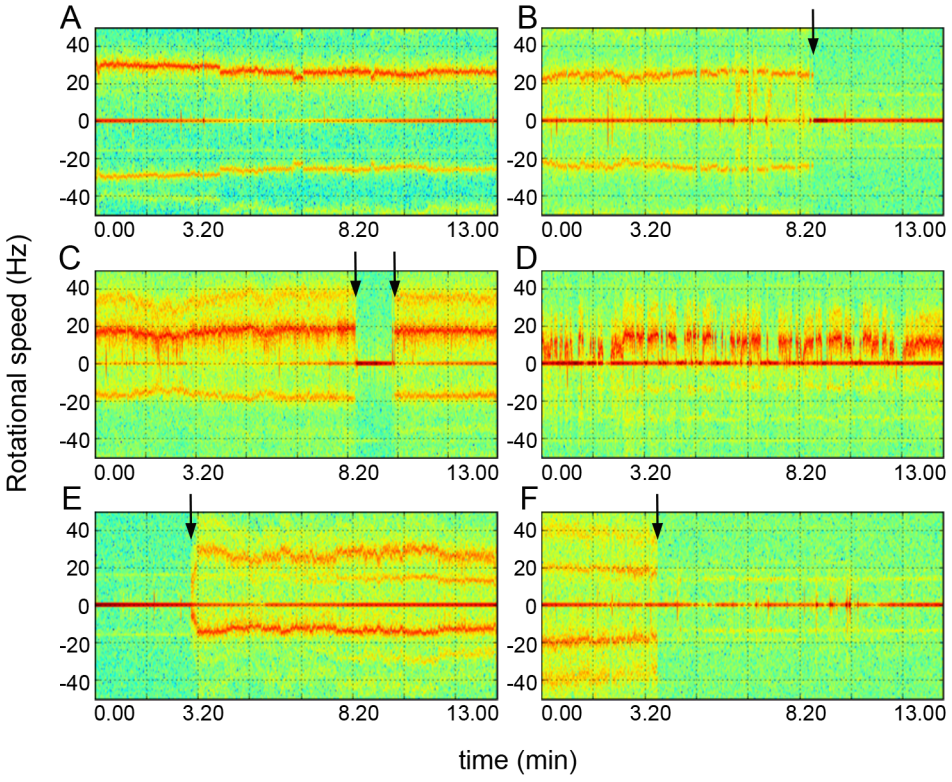


Figure 3.15: Examples of BRAS output for the general phenotypes seen in tethered cells, upon addition (3 min) and removal (8 min) of propionate. Arrows indicate interesting events (stops and starts due to propionate). (A) Unresponsive (B) Responsive (C) Adaptive (D) Stoppy (E) Starts on propionate addition (F) Stops on propionate addition.

The majority of these phenotypes, although subjective, are relatively clear. If a clear rotation signal is obtained, stops over 5 s are easily identified in a heat map of rotational signal. As the attractant is added and removed at specific times, a stop or start around these times may be assumed to be driven by that stimulus.

The most subjective measure is that of stoppiness. This does not describe traces with small gaps due to noise in the signal or an improperly tethered cell. Instead, as can be seen in Figure 3.15D, these traces show numerous clear, long (5-30 s) gaps.

The merit of this classification scheme will be discussed in relation to the results below. It is obvious that the development of software to automatically classify all cells into the categories based solely on the given output would be desirable. However, before this kind of software can be developed, a sense of which phenotypes are possible and what classification is useful statistically is essential.

Once cells had been classified, Click&Mean was used to measure the duration of the stop, called the adaptation period, in cells which stopped rotation on removal of stimulus, then restarted.

As can be seen from Figure 3.16, deciding where a stop begins and ends can be subjective. In many cases, the point where rotation restarts is clear. In others, there are some attempts to start followed by pauses before full rotation begins again. The full adaptation time is defined in Kojadinovic *et al.* (2011) as ending when full rotation begins again. As such, the adaptation time was measured to have ended after these false starts, once a rotation of at least 5 s was seen.

Again, future work should focus on the development of software able to make these classifications and measure stops of interest automatically, using the information gained from this study. However, development of such software is beyond the scope of this project.

Two kinds of data are thus obtained from the tethering analysis, frequencies of cell phenotypes and the durations of adaptation periods.

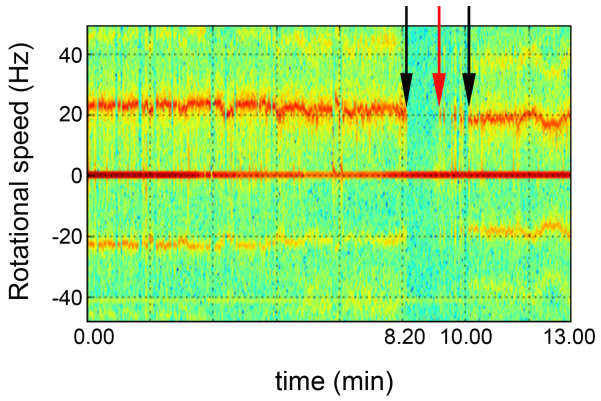


Figure 3.16: In adaptive cells, the adaptation time is defined as the time taken from full cessation of rotation, to the full recovery of rotation (the time between the two black, outer arrows). Stuttering rotation starting from the red, central arrow is not considered full recovery.

3.3.2 Analysis of cell phenotype variability

Due to the small data sets used, strains were previously classified as unresponsive, responsive or adaptive. With a larger population of cells, it became obvious that all strains show variability and include subpopulations of all three basic phenotypes. It would be useful if this variability could be used to indicate how far from wildtype a particular strain is, and in what direction. The appearance of rare events is also interesting, as these are likely the result of stochasticity across the population.

The data set contains categorical nominal data - a set of 'count data' for the phenotypes identified. Pearson's chi-square test for heterogeneity can be used to determine whether a group of strains is significantly different from one another or not based on this data set. This is done by calculating the expected frequency counts for each category (where a category is a strain-phenotype combination), then comparing this to the observed frequencies and estimating the probability of the difference between observed and expected occurring due only to chance.

The limitation of chi-square tests is that no more than 20% of the categories in the resulting contingency table may have an expected count of less than 5, as the estimated probability is no longer accurate. If the contingency table is sparse, this

limitation can be circumvented by using a randomisation test to estimate probabilities more accurately (Mehta and Patel, 2011). An exact test could also be used to directly calculate the probability, but here would be inadvisable due to the numbers of cells collected and the number of strains analysed. Randomisation generates a probability range in which the exact probability falls, with a particular confidence.

The raw frequency data is shown in Table 3.7. As would be expected, Pearson's chi-square test with Monte Carlo randomisation based on 100,000 sampled tables and a randomly selected starting seed gives a two-sided probability of less than 0.001 at 99% confidence that there is no significant statistical difference between these strains.

Although the test shows that at least one strain is statistically different from the others, it does not suggest which strains are similar and which different. Hierarchical clustering analysis can be used to make the initial separation of strains into like groups. This is something that could be done by eye. When comparing a large number of strains over a large number of variables, however, a computational approach is preferable. The approach developed here will ideally be useful for very large databases, for future extensions of the work.

Hierarchical clustering starts with each strain in its own group, then groups the most likely strains together. Complete linkage, or a furthest neighbour approach, is used to decide when to merge two clusters. In the case of count data, the distance between members of a cluster and members of neighbouring clusters is based on the standard chi-square test for equality for two sets of frequencies.

The limitations of chi-square tests based on asymptotic probabilities described above are not a problem here, as clustering is an exploratory technique. Clustering results cannot be used to make a firm decision as to whether the members of a cluster are members of the same population. Rather, clustering gives an initial grouping for further testing.

Table 3.7: Frequency of rotation phenotypes seen in *R. sphaeroides* chemotaxis deletion mutants grown under aerobic conditions, tethered and exposed to the addition and removal of a 100 μM propionate challenge.

strain	rotate from start				rotate on propionate				stoppy		total cells
	unresponsive	responsive	adaptive	stop on propionate	unresponsive	responsive	adaptive	stop early	unresponsive	adaptive	
WS8N	7	1	43	0	2	2	6	2	0	0	63
$\Delta CheB_1$	73	3	2	2	0	0	0	0	0	0	80
$\Delta CheB_2$	3	2	3	0	0	6	0	11	0	0	25
$\Delta CheR_2$	45	2	1	1	0	0	0	0	31	2	82
$\Delta CheR_3$	55	3	0	0	1	0	0	0	0	0	59
$CheY_3$(D57A)	10	3	54	0	0	0	1	1	0	0	69
$CheY_4$(D57A)	6	0	48	0	1	0	0	0	0	0	55
$CheY_{3,4}$(D57A)	30	1	3	0	0	0	0	0	17	2	53
$CheY_{3,4}$(D57N)	68	4	0	0	3	0	0	0	2	0	77
$CheY_6$(D57A)	78	1	1	0	0	0	0	0	0	0	80
$CheA_3$(H48Q), $CheY_6$(D57N)	94	0	1	0	1	0	0	0	1	1	98
$CheA_4$(G470K), $CheY_6$(D57N)	79	1	1	0	1	0	0	0	0	0	82

Table 3.8: Agglomeration schedule for hierarchical clustering of deletion mutants based on phenotype frequencies in tethered cells exposed to changes in attractant. Agglomeration coefficients are scaled in the range [0,1].

Clustering stage	Representative strains from clusters combined in stage		Agglomeration coefficient	Coefficient jump to next stage	Stage each cluster first appears		Next Stage
	Cluster 1	Cluster 2			Cluster 1	Cluster 2	
1	$CheY_6$(D57A)	$CheA_4$(G470K),$CheY_6$(D57N)	0	0.047	0	0	4
2	$CheY_{3,4}$(D57N)	$\Delta CheR_3$	0.047	0.035	0	0	6
3	$\Delta CheR_2$	$CheY_{3,4}$(D57A)	0.082	0.009	0	0	9
4	$CheY_6$(D57A)	$\Delta CheB_1$	0.091	0.048	1	0	6
5	$CheY_3$(D57A)	$CheY_4$(D57A)	0.139	0.059	0	0	8
6	$CheY_{3,4}$(D57N)	$CheY_6$(D57A)	0.198	0.002	2	4	7
7	$\Delta CheB_1$	$CheY_{3,4}$(D57N)	0.2	0.029	0	6	9
8	WS8N	$CheY_3$(D57A)	0.229	0.354	0	5	10
9	$\Delta CheB_1$	$\Delta CheR_2$	0.583	0.059	7	3	11
10	WS8N	$\Delta CheB_2$	0.642	0.358	8	0	11
11	WS8N	$\Delta CheB_1$	1		10	9	0

We can also use clustering to determine whether solutions vary a great deal when a particular (rare) category is excluded or included, to see if this category is useful to record.

The initial exploratory clustering dendrogram is shown in Figure 3.17. An agglomeration coefficient is associated with each group merge event (Table 3.8). A likely clustering arrangement is one before a large jump in the agglomeration coefficient. This makes a final solution of four clusters (precedes a jump of 0.354) or two clusters (precedes a jump of 0.358) most likely.

The two-cluster solution is rejected

Clustering into two clusters is shown in the top panel of Figure 3.17. Splitting up the strains in this way gives the two contingency tables Table 3.9 and 3.10. Both are sparse contingency tables and can again be analysed with Pearson's chi-square test for heterogeneity using Monte Carlo randomisation with 100,000 sample tables to accurately estimate probability.

Both tests give a two-sided probability of less than 0.001 at 99% confidence that there is no statistical difference between strains within the clusters. The clustering solution giving two clusters is rejected, as there is still significant difference between strains within each cluster.

The four-cluster solution is rejected

Clustering into four clusters is shown in the bottom panel of Figure 3.17. Cluster 4 has a membership of one, so further statistics are not required. The remaining three clusters have sparse contingency tables (Tables 3.11, 3.12 and 3.13) and are analysed with Pearson's chi-square test for heterogeneity using Monte Carlo randomisation with 100,000 sample tables to accurately estimate probability.

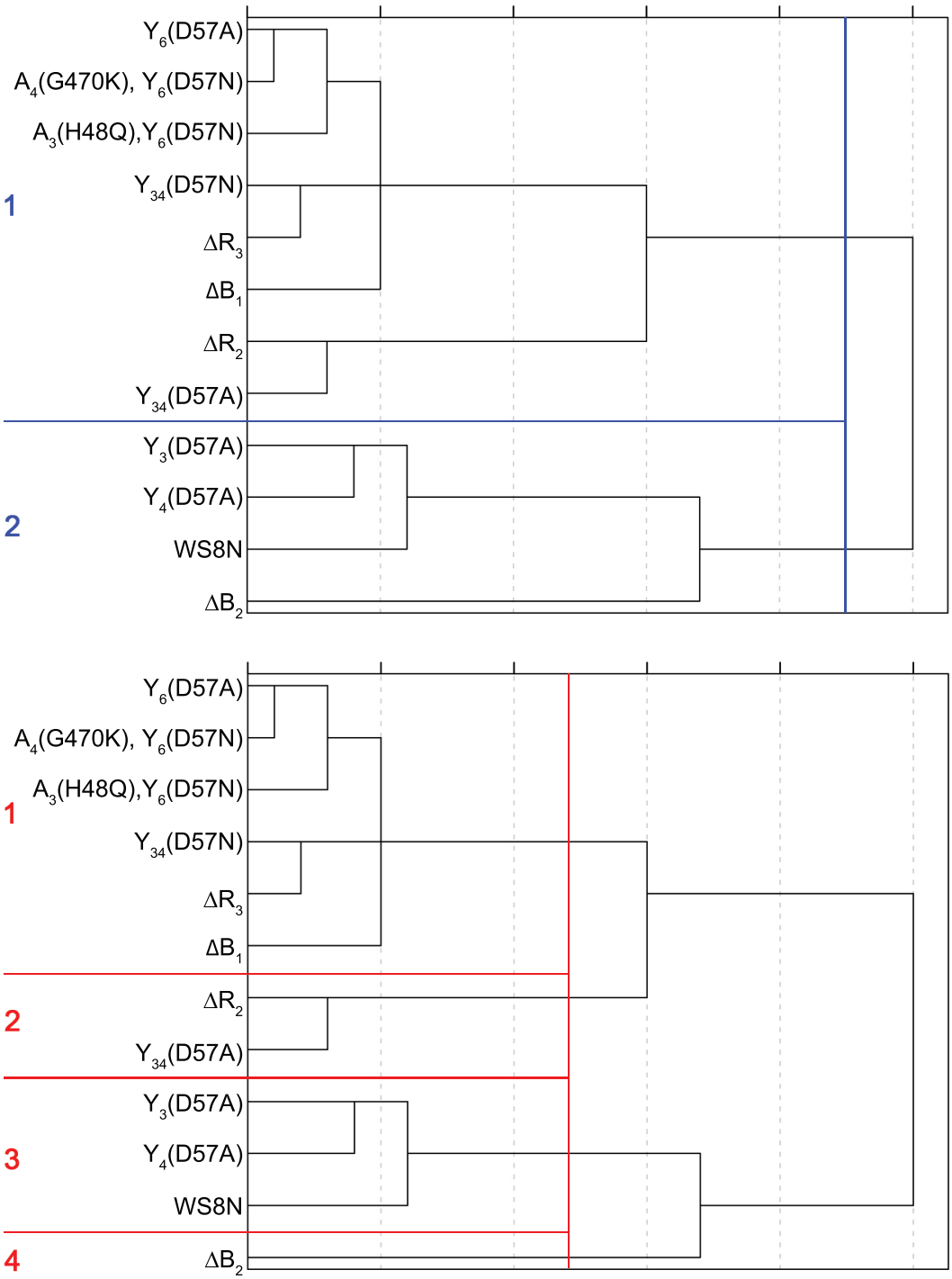


Figure 3.17: Hierarchical clustering of chemotaxis mutant strains based on a complete linkage, chi-square test of tethering phenotype frequencies. The two-cluster solution (top, blue lines) and four-cluster solution (bottom, red lines) are the two most likely clustering solutions. Lines and numbering indicate how the strains are grouped in each of the solutions. Distance between clusters (x-axis) is an arbitrary scale.

Table 3.9: Contingency table of tethering phenotype frequencies for deletion mutants clustered in group 1 of the two-cluster solution. Categories that can be removed without affecting the solution are left blank.

strain	rotate from start				rotate on propionate				stoppy		total cells
	unresponsive	responsive	adaptive	stop on propionate	unresponsive	responsive	adaptive	stop early	unresponsive	adaptive	
$\Delta CheB_1$	73	3	2	2	0				0	0	80
$\Delta CheR_2$	45	2	1	1	0				31	2	82
$\Delta CheR_3$	55	3	0	0	1				0	0	59
$CheY_3$(D57A) $CheY_4$(D57A)	30	1	3	0	0				17	2	53
$CheY_3$(D57N) $CheY_4$(D57N)	68	4	0	0	3				2	0	77
$CheY_6$(D57A)	78	1	1	0	0				0	0	80
$CheA_3$(H48Q), $CheY_6$(D57N)	94	0	1	0	1				1	1	98
$CheA_4$(G470K), $CheY_6$(D57N)	79	1	1	0	1				0	0	82

Table 3.10: Contingency table of tethering phenotype frequencies for deletion mutants clustered in group 2 of the two-cluster solution. Categories that can be removed without affecting the solution are left blank.

strain	rotate from start				rotate on propionate				stoppy		total cells
	unresponsive	responsive	adaptive	stop on propionate	unresponsive	responsive	adaptive	stop early	unresponsive	adaptive	
Wildtype	7	1	43		2	2	6	2			63
$\Delta CheB_2$	3	2	3		0	6	0	11			25
$CheY_3$(D57A)	10	3	54		0	0	1	1			69
$CheY_4$(D57A)	6	0	48		1	0	0	0			55

Table 3.11: Contingency table of tethering phenotype frequencies for deletion mutants clustered in cluster 1 of the four-cluster solution. Categories that can be removed without affecting the solution are left blank.

strain	rotate from start				rotate on propionate				stoppy		total cells
	unresponsive	responsive	adaptive	stop on propionate	unresponsive	responsive	adaptive	stop early	unresponsive	adaptive	
$\Delta CheB_1$	73	3	2	2	0				0	0	80
$\Delta CheR_3$	55	3	0	0	1				0	0	59
$CheY_3$(D57N) $CheY_4$(D57N)	68	4	0	0	3				2	0	77
$CheY_6$(D57A)	78	1	1	0	0				0	0	80
$CheA_3$(H48Q), $CheY_6$(D57N)	94	0	1	0	1				1	1	98
$CheA_4$(G470K), $CheY_6$(D57N)	79	1	1	0	1				0	0	82

Table 3.12: Contingency table of tethering phenotype frequencies for deletion mutants clustered in cluster 2 of the four-cluster solution. Categories that can be removed without affecting the solution are left blank.

strain	rotate from start				rotate on propionate				stoppy		total cells
	unresponsive	responsive	adaptive	stop on propionate	unresponsive	responsive	adaptive	stop early	unresponsive	adaptive	
$\Delta CheR_2$	45	2	1	1					31	2	82
$CheY_3$(D57A) $CheY_4$(D57A)	30	1	3	0					17	2	53

Table 3.13: Contingency table of tethering phenotype frequencies for deletion mutants clustered in cluster 3 of the four-cluster solution. Categories that can be removed without affecting the solution are left blank.

strain	rotate from start				rotate on propionate				stoppy		total cells
	unresponsive	responsive	adaptive	stop on propionate	unresponsive	responsive	adaptive	stop early	unresponsive	adaptive	
Wildtype	7	1	43		2	2	6	2			63
$CheY_3$(D57A)	10	3	54		0	0	1	1			69
$CheY_4$(D57A)	6	0	48		1	0	0	0			55

Tests for clusters one and two give a two-sided probabilities at 99% confidence in the ranges 0.098–0.103 and 0.711–0.718 respectively that there is no statistical difference between strains within the clusters. These clusters are accepted as having only random differences between strain members.

Tests for cluster three, however, give a probability in the range 0.027–0.030. This grouping is thus rejected, as there is still significant difference between member strains. As such, the four-cluster solution cannot be used.

The five-cluster solution is accepted

Three of the four clusters given in the four-cluster solution were shown to have no significant variability between member strains, and the agglomeration coefficient jumps between all other clustering stages are very low. Therefore, rather than jumping to the clustering solution preceding the next highest agglomeration coefficient jump, the next solution in the schedule is taken.

The next valid solution, the five-cluster solution, requires WS8N to be split out of cluster 3 of the four-cluster solution. This gives two clusters. WS8N forms a cluster alone, so no further statistical test are required. The remaining two-member cluster has a probability in the range 0.304–0.311 that the members are from the same population, making this a likely true grouping.

Testing the validity of the phenotype groups used

Either the five-cluster solution is the most able to describe the data set, or the phenotype classification used here is adding in variability where in reality there is none. The latter seems likely, as agglomeration coefficients point more strongly towards the four-cluster solution than the five. Each of the four clusters are explored further and compared to what is known about swimming behaviour to suggest where phenotype binning may be sensible.

Cluster 1 (Unresponsive) The first, largest, cluster comprises the strains with the single mutations $\Delta CheB_1$, $\Delta CheR_3$ and $CheY_6$(D57A) and the double mutations $CheA_4$(G470K)$CheY_6$(D57N), $CheA_3$(H48Q)$CheY_6$(D57N) and $CheY_{3,4}$(D57N). All have a high count for 'unresponsive' and very low counts in all of the other phenotypes (Table 3.11).

All of these strains have smooth swimming behaviour under steady state conditions, which reinforces this grouping.

Cluster 2 (Stoppy unresponsive) The second cluster comprises the strains with the single mutation $\Delta CheR_2$ and the double mutation $CheY_{3,4}$(D57A). Both of these strains are predominantly unresponsive, with these split between normal swimming and stoppy swimming.

The stoppy distinction had not been used before. If this distinction had not proved useful (if, say stoppy cells were very rare events across all of the strains), the resulting phenotypes could easily be removed by binning stoppy-unresponsive and unresponsive cells together, and stoppy-adaptive and adaptive cells together. However, the stoppy phenotype is what creates the distinction between Groups 1 and 2. As both of these strains showed free swimming behaviour significantly stoppier than wildtype under steady state conditions, this split in unresponsive is likely to be based on real differences between the strains, rather than an artefact of by-eye cell classification. This makes the stoppy distinction in initial classification credible.

Cluster 3 (Adaptive) The third cluster of the four-cluster solution includes the wildtype strain WS8N and the strains with the single mutations $CheY_3$(D57A) and $CheY_4$(D57A). These strains are all majority adaptive cells. The variability that drives WS8N from the main cluster is due to the distinction between cells which rotate under zero nutrient/chemoeffector conditions and cells which only begin rotating when a chemoeffector is introduced. These phenotypes are rare events in all

three strains, although less rare in WS8N.

It is possible that binning together cells which rotate from the start and those that start rotating on propionate addition would remove sufficient variability between the strains for cluster 3 to be accepted. The question is whether this is valid or whether it removes an important distinction between the two sets. Unlike the stoppy cells, the split here is not analogous to the grouping seen for steady state swimming. Under steady state, although both of the mutant strains ($CheY_3$(D57A) and $CheY_4$(D57A)) had a median proportion of time stopped greater than wildtype, only the $CheY_4$(D57A) mutant's difference was significant.

Cluster 4 (deletion of$CheB_2$: Inhibited) The final cluster includes only the strain with the $CheB_2$ deletion. This strain is obviously different from the rest of the strains. During data capture, it was difficult to obtain high numbers of rotating cells. The majority of cells that were found rotating needed the addition of propionate to lift inhibition of movement. Of these, the majority stopped rotating before propionate was removed.

This strain has what can be called an inhibited phenotype. Its default motor state is stationary. Chemoeffector binding lifts that inhibition, but adaptation to the stimulus returns cells to the inhibited state.

Interestingly, no cells in this strain were classified as stoppy, although steady state free swimming behaviour was significantly stoppier than wildtype. It is possible that those cells with a stoppy phenotype were so stoppy as to not generate a clear enough rotational signal for Click&Mean to register.

The distinction between rotation starting from the beginning of recording or only on propionate addition seems an important one for defining the behaviour of this strain. However, removing this level of classification would leave intact the inhibited cells - those that started rotating with propionate addition and stopped again

before propionate could be removed. If this phenotype alone is sufficient to define the difference between the $\Delta CheB_2$ mutant strain and the strains in cluster 3, then using binning to remove differences between members of cluster 3 would still keep cluster 3 and 4 distinct from one another.

Repeating hierarchical clustering with alternative phenotype grouping

The data set is binned, so that cells with unresponsive, responsive and adaptive phenotypes are not separated on basis of when rotation starts. The same hierarchical clustering protocol is used on the data set after binning and an identical grouping is obtained as in Figure 3.17. Once again, the two and four cluster solutions are most likely (coefficient jumps of 0.327 and 0.361 respectively) and Pearson's chi-square test for heterogeneity rejects the two cluster solution (probability for either cluster's variability arising from chance less than 0.001).

However, the four-cluster solution is now accepted after Pearson's chi-square test for heterogeneity with Monte Carlo randomisation. Grouping follows the phenotypes unresponsive with normal swimming (Cluster 1, probability 0.113–0.119), unresponsive with stoppy swimming (Cluster 2, probability 0.715–0.722), adaptive (Cluster 3, probability 0.587–0.595) and inhibited (Cluster 4, one member).

Thus, removing the distinction of when cells begin rotating does not affect the unresponsive or inhibited groups. The only cluster affected is the adaptive group. Although in this particular case, counting cells that start rotating on propionate addition separately from those that rotate in attractant-free buffer introduced unnecessary variability in the data, when other strain sets are compared the distinction may prove useful.

Two new phenotypes were introduced here, stoppiness during steady state between challenges and whether rotation was measured from the null buffer or began only on addition of attractant. The second of these proved to be unnecessary, whereas the

first provided the key distinction between two groups of strains. When analysing a new set of strains, it is best to start with as many phenotypes as are needed to accurately describe each cell, regardless of how often each phenotype is used. Only during analysis can phenotypes be discarded, once their usefulness in distinguishing between strains has been tested.

3.3.3 Rare events: responsive cells and propionate stops

The fact that when rotation starts was mostly irrelevant in clustering suggests that events which are always very rare are unlikely to help direct clustering or add much information to a strain. Returning again to the frequencies of all of the phenotypes seen in the full data set (Table 3.7), there are less than 5% responsive cells in all of the strains except the $CheB_2$ deletion mutant. Similarly, 'stop on propionate' is a very rare phenotype, appearing in only two strains and then at less than 5% the total number of cells in that strain.

Removing these two phenotypes completely has no effect on the final result. Clustering suggests the same two or four clusters, both of which fail on Pearson's test for heterogeneity, whereas the five cluster solution is accepted. Again, binning cells which rotate from the beginning and on propionate addition gives an identical clustering solution with an acceptable four cluster solution.

Therefore, when a phenotype appears at less than 5% of total cells in all strains, that phenotype can probably be safely removed from the data set, as it is only a reflection of stochasticity across the population.

However, as with the 'rotate on propionate addition' phenotype above, the phenotypes 'responsive' and 'stop on propionate addition' should still be included in the initial classification and only removed once it is clear that they are rare events in all strains in a set.

3.3.4 Adaptation time variability

From the frequency results above, only the stop lengths of those strains with 10% or more adaptive cells are included for further analysis (Figure 3.18). An unbalanced one-way ANOVA test with a post-hoc Tukey test indicates that the $CheB_2$ deletion mutant is significantly different from the rest of the strains tested. Two-way comparisons indicate a significant difference between the strains with the single mutations $CheY_3$(D57A) and $CheY_4$(D57A) (p=0.004). However, both are not significantly different from either WS8N or the $CheY_{3,4}$(D57A) double mutant strain. The null hypothesis that these four strains are part of the same population is therefore not rejected (p=0.395).

Thus, the difference between inhibited and adaptive cells seen in the phenotype frequencies is carried through in mean adaptation time. Although adaptive and stoppy unresponsive have very different phenotype frequencies, the few adaptive cells within the stoppy unresponsive data set are indistinguishable from wildtype cells.

3.3.5 Comparing steady state and dynamic behaviour

The steady state swimming behaviour of a set of chemotaxis mutants suggested possible network connectivities in *R. sphaeroides* chemotaxis, under aerobic conditions. The same mutants' ability to respond to a dynamic change in one chemoeffector, propionate, under aerobic conditions is now compared to the steady state behaviour recorded in Section 3.2.4. Once again, the basic model derived from the phosphorylation and localisation work described in depth in Section 1.4 is assumed. Most importantly, it is assumed that increased methylation increases the probability that a chemoreceptor cluster will change conformation and trigger CheA kinase activity. It is also assumed that $CheR_2$ is a methyltransferase that methylates the polar cluster, increasing the probability of $CheA_2$ autophosphorylation, and that $CheR_3$ is a

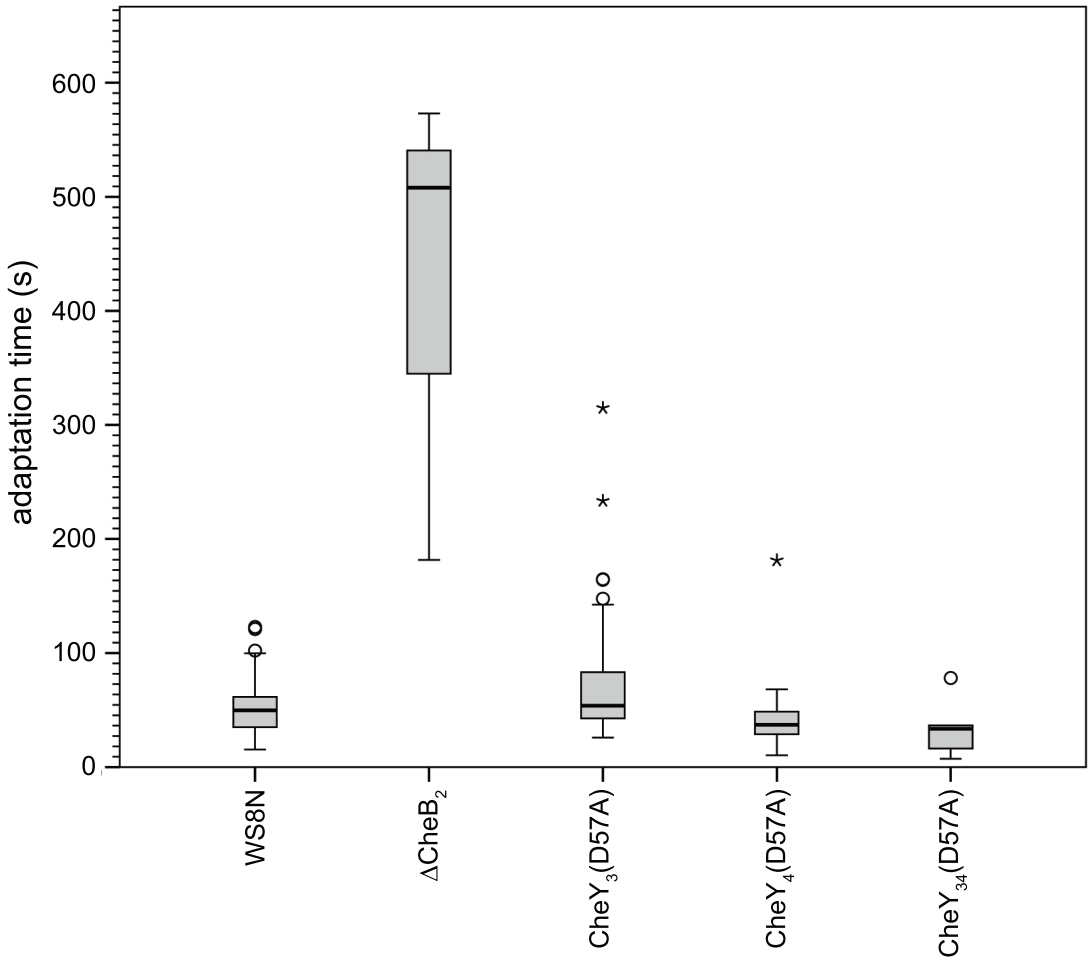


Figure 3.18: Adaptation times in tethered deletion mutants grown aerobically and challenged with a 100 μM drop in propionate concentration. The boxplot edges indicate the first and third percentiles, the box centre indicates the median, the whiskers indicate the range $\pm$ 1.5x the box height, and the circles indicate outliers (Section 2.6.1). The number of data points in each block are 49 (WS8N), 3 (ΔCheB$_2$), 55 (CheY$_3$(D57A)), 48 (CheY$_4$(D57A)) and 5 (CheY$_{3,4}$(D57A)).

methyltransferase that methylates the cytoplasmic cluster, increasing the probability of CheA$_4$ transphosphorylating CheA$_3$.

Deletion of CheR$_3$

Does the polar cluster respond to drops in chemoeffector by phosphorylating CheY$_6$, producing a stop? If so, the deletion of CheR$_3$ should not affect the ability to respond to dynamic changes in chemoeffector concentration despite the lack of CheY$_6$-P under steady state conditions, as the polar cluster should produce its own CheY$_6$-P under dynamic conditions.

Does the cytoplasmic cluster, not the polar cluster, produce the $CheY_6$-P required to produce a stop, either by raising the baseline levels of $CheY_6$-P sufficiently for additional $CheY_6$-P from the polar cluster to trigger a stop, or as the sole producer of $CheY_6$-P? If so, the deletion of $CheR_3$ should remove the ability to respond to dynamic changes in chemoeffector concentration. The lack of $CheY_6$-P from the cytoplasmic cluster should prevent a stop.

The strain with a $\Delta CheR_3$ mutation was classified as unresponsive. It is therefore likely that $CheY_6$-P from the cytoplasmic cluster is required for the cell to respond to a drop in chemoeffector and stop. Either the polar cluster does not respond to a chemoeffector drop by producing $CheY_6$-P (perhaps only signalling via $CheY_3$-P and $CheY_4$-P to effect a run), or the cytoplasmic cluster produces a baseline concentration of $CheY_6$-P and the polar cluster manages the concentration of $CheY_6$-P in a dynamic range capable of effecting a stop.

$CheY_3$(D57A) and $CheY_4$(D57A) single mutant strains

Does the polar cluster respond to increases in chemoeffector by phosphorylating $CheY_3$ and $CheY_4$, producing a run? If so, then if $CheY_3$ and $CheY_4$ are not available for phosphorylation (D57A), a decrease in chemoeffector concentration should result in an altered adaptation pattern: a stop will be produced as normal from either cluster's action, but the return to run will take longer than normal, as $CheY_6$-P must be dephosphorylated following the stop, rather than $CheY_3$-P or $CheY_4$-P competing to bind and producing the run.

$CheY_{3,4}$(D57A) was classified as stoppy unresponsive. The few adaptive cells within the data set were indistinguishable from wildtype (Figure 3.18).

The addition of propionate did not relieve the stoppy phenotype, unlike the addition of propionate to the $\Delta CheB_2$ mutant. The higher baseline of $CheY_6$-P (stoppy steady state swimming) may be from the polar cluster, phosphorylating $CheY_6$ at greater rates when deprived of its usual $CheY_3$ and $CheY_4$. However, if both clusters can produce $CheY_6$-P and can respond to external chemoeffector concentrations, a

complete stop in response to the propionate drop would be expected. Although the CheYs are diffuse in the cytoplasm, some localisation of $CheY_3$ at the both chemotaxis clusters and $CheY_3$ at the cytoplasmic cluster has been recorded (Porter *et al.*, 2006). $CheY_3$-P and $CheY_4$-P may have some action at the cytoplasmic cluster, inhibiting $CheY_6$-P production, and the increased $CheY_6$-P on their loss (stoppy swimming) may be driven by the cytoplasmic cluster

Deletion of $CheB_2$

Does $CheB_2$ function as a methylesterase at the cytoplasmic cluster? If so, the deletion of $CheB_2$ will result in high levels of $CheY_6$-P from the cytoplasmic cluster, producing the observed stoppy steady state swimming phenotype. If the polar cluster can respond to the chemoeffector concentration without the action of the cytoplasmic cluster, then a dynamic response should still be seen in a $CheB_2$ deletion. On an increase of chemoeffector, either the polar cluster will cease producing $CheY_6$-P, dropping levels sufficiently to transiently relieve the stoppy phenotype, or the polar cluster will produce $CheY_3$-P and $CheY_4$-P, which will bind to the motor and effect a run, transiently relieving the stoppy phenotype.

The $CheB_2$ deletion strain was classified as inhibited. The majority of cells were stopped and only began rotating on the addition of chemoeffector. Those cells that were able to rotate stopped on the removal of chemoeffector and either did not recover, or recovered over a much longer adaptation time than wildtype. This result fits the proposed network action above.

Deletion of $CheR_2$

Does the cytoplasmic cluster sense and respond to external chemoeffector concentrations independently of the polar cluster under aerobic conditions, as in photoheterotrophic conditions (Martin et al., 2001a)? If so, the deletion of $CheR_2$, will prevent the polar cluster from responding to stimuli but will not remove the ability to respond to dynamic changes, as the

cytoplasmic cluster will produce CheY$_6$-P and effect a stop.

The strain with a CheR$_2$ deletion is classified as stoppy unresponsive. Thus, it is unlikely that the cytoplasmic cluster can achieve a chemotaxis response without the polar cluster also responding to the external signal, at least under aerobic conditions. The lack of CheY$_3$-P and CheY$_4$-P may affect the cytoplasmic cluster's ability to respond to dynamic changes in some way. The stoppy swimming suggests a high level of baseline CheY$_6$-P: again, it is possible that the cytoplasmic cluster requires the presence of CheY$_3$-P and CheY$_4$-P for tuning.

Deletion of CheB$_1$

If the cytoplasmic cluster senses and responds to external chemoeffector concentrations independently of the polar cluster and CheB$_1$ is the methylesterase of the polar cluster, then the deletion of CheB$_1$ should result in higher levels of CheY$_3$-P and CheY$_4$-P. The deletion mutant may still be able to respond to dynamic changes.

The strain with a CheB$_1$ deletion was classified as unresponsive and smooth swimming. When the polar cluster is permanently active, there is insufficient CheY$_6$-P in the cell to cause a stop. CheY$_3$-P and CheY$_4$-P may flood the motor, preventing any CheY$_6$-P from the cytoplasmic cluster from binding. CheY$_3$-P and CheY$_4$-P may also inhibit the cytoplasmic cluster from forming CheY$_6$-P in some way. The cytoplasmic cluster requires the polar cluster to respond to an external stimulus before it can respond, under aerobic conditions.

Inactive cytoplasmic kinase, CheY$_6$(D57N)

Is the cytoplasmic cluster required for chemotaxis under aerobic conditions solely to produce a baseline of CheY$_6$-P, while the polar cluster responds to the external chemoeffector concentration and dynamically produces phosphorylated CheYs to cause runs and stops? If so, CheY$_6$(D57N) can be substituted to produce the baseline of CheY$_6$-P, as it is permanently

phosphorylated and can cause stops in the absence of functional $CheA_3CheA_4$ (from the free swimming results). The removal of functional $CheA_3CheA_4$ along with $CheY_6$(D57N) should not remove the ability to respond to dynamic changes, especially if the polar cluster relies on $CheY_3$-P and $CheY_4$-P to effect its influence, rather than $CheY_6$-P.

Both the $CheA_3$(H48Q)$CheY_6$(D57N) and $CheA_4$(G470K)$CheY_6$(D57N) double mutants were classified as unresponsive. This would suggest that the cytoplasmic cluster's role is not just to supply $CheY_6$-P, but that some other signalling is required to allow a chemotaxis response.

However, both strains are classified as smoother than wildtype, although not smooth swimmers. This suggests that the $CheY_6$(D57N) mutant has a reduced binding rate to the motor when it is not stably phosphorylated on Ser83. It is possible that even when no $CheY_3$-P or $CheY_4$-P is produced, that $CheY_6$(D57N) alone cannot bind for long enough to cause a dynamic stop.

3.4 Discussion

In this chapter, *R. sphaeroides* chemotaxis mutants were analysed for steady state and dynamic swimming behaviour using two cell-level techniques.

3.4.1 Analysis of large populations of free swimming and tethered cells

Steady state swimming behaviour was characterised by the appearance of tracks and by the distribution of fraction of time stopped in each track in a population. All track data sets were censored using the same parameters, aiming to remove as many anomalous tracks as possible while not sacrificing too many legitimate tracks. Tracks were classified into runs and stops using a hidden Markov model, comparing

each framewise transition to a database of known stops and run.

The resulting data sets are still noisy and additional work in censoring to remove all corkscrews and drifters would improve accuracy. However, despite this noise, clear significant differences between strains could be observed when run-stop data was summarised by fraction of time stopped vs. running.

Three general classes of behaviour were described: stoppier than wildtype, wildtype, and less stoppy (smoother) than wildtype. There is more variability in strains stoppier than wildtype than in strains less stoppy than wildtype, as the majority of strains less stoppy are unable to stop at all. Stoppy strains formed three statistically significant, well-separated groups. Smooth strains formed one statistically significant group, but relaxing of probability for splitting groups allowed two groups: smoother than wildtype and smooth.

Dynamic swimming behaviour was characterised by the frequency of cell phenotypes and the mean adaptation time for a data set. Recorded cells were classified by eye as unresponsive, responsive, adaptive or inhibited. Cells were also separated into those with normal swimming and those with many transient stops, and those which rotate initially in buffer and those that only begin rotation on addition of chemoeffector. The responsive phenotype was rare and could be ignored without affecting results. Distinguishing when cells begin rotation introduced unnecessary variability into the data set and this distinction was removed.

Four statistically different classes of strains were described, unresponsive, stoppy unresponsive, adaptive and inhibited.

All of the smooth steady state swimmers are dynamically unresponsive. Smoother than wildtype steady state swimmers are dynamically adaptive. Stoppy steady state swimmers are either stoppy and dynamically unresponsive, or are dynamically inhibited. The difference between these two is profound, as any stoppy unresponsive cell that does adapt is indistinguishable from wildtype, whereas the dynamic inhib-

ited cell has a significantly longer adaptation time.

Using these two methods of analysis gives significantly more detail about chemotaxis mutants than the perfunctory 'unable to perform chemotaxis and/or normal motility' label that swim plates result in. Both methods show a marked improvement over previous methods, as the conversion from collected videos to descriptive data introduces much less noise and thus finds statistical differences in the same places as by-eye comparisons.

3.4.2 *R. sphaeroides* chemotaxis network connectivity

The steady state swimming results obtained here contradict the majority of published work, as previous studies recorded the majority of *R. sphaeroides* chemotaxis mutants as having normal swimming (for example, Porter *et al.* (2006)). Previous work on steady state swimming using obsolete methods is therefore not discussed further. The tethering results obtained here agree with those obtained in previous work (for example, Martin *et al.* (2001a)), with the exception of the inhibited phenotype obtained for the $CheB_2$ deletion mutant, the complexities of whose behaviour had not fully been recorded.

The most significant result is the existence of a not previously described swimming phenotype, that of the stoppy unresponsive cell. Two different deletion mutants produced a stoppy steady state swimming pattern, while losing the ability to respond to a drop in the chemoeffector concentration by stopping. In *E. coli*, the frequency of tumbling during steady state swimming and the ability to tumble in response to a change in environment are both controlled by the adaptation proteins.

R. sphaeroides has two sets of adaptation proteins and two clusters of chemoreceptors. It is possible that, under aerobic conditions, the two elements of swimming, dynamic and steady state, have been split from one another and are controlled independently by the two clusters. In this model, the polar cluster acts similarly to the

canonical *E. coli* cluster and responds to the chemoeffector concentration.

However, unlike in *E. coli*, activation of the polar cluster switches on smooth swimming and a drop in the chemoeffector concentration inactivates the cluster to allow a stop. Preventing signalling through the polar cluster by permanently making $CheY_3$ and $CheY_4$ unphosphorylatable, by removing the kinase, or by removing the sensitising methyltransferase $CheR_2$, results in a stoppy, unresponsive phenotype. Causing continued signalling, either through the removal of desensitising $CheB_1$ or by the permanent phosphorylation of $CheY_3$ and $CheY_4$ on an alternative site, causes a smooth swimming, unresponsive phenotype. In both cases, the cell is unresponsive, although the steady state swimming behaviour is quite different.

It appears that, under aerobic conditions, the cytoplasmic cluster controls steady state swimming and the ability to stop by controlling the concentration of $CheY_6$-P. The cytoplasmic cluster produces a baseline concentration of $CheY_6$-P, controlling the frequency of stops under steady state conditions. If signalling through the cluster is halted, by making $CheY_6$ unphosphorylatable, by removing the kinase, or by removing the sensitising methyltransferase $CheR_3$, the cell loses the ability to stop and has a smooth swimming phenotype. A side effect of this is a loss of responsiveness under dynamic conditions. Similarly when signalling through the cluster is over-expressed by deleting desensitising $CheB_2$, a stoppy phenotype is generated. This chemotaxis mutant remains responsive and able to adapt, albeit slowly or from the inhibited state. The cell more often defaults to a stopped state and, on nutrient/chemoeffector concentration addition, is able to start swimming again briefly due to the production of $CheY_3$-P and $CheY_4$-P. Adaptation by the polar cluster allows the over-production of $CheY_6$-P from the cytoplasmic cluster to restore the default stop.

However, the inactivation of the cytoplasmic kinase while supplying $CheY_6$(D57N) gives results that disagree with this model. This strain shows smoother swimming than wildtype, indicating that $CheY_6$(D57N) without an additional Ser83 phospho-

rylation binds poorly to the motor. This strain is unresponsive to dynamic changes, which suggests that the model for chemotaxis is missing some interaction between the two clusters. Alternatively, the (D57N) mutation is a poor substitution for phosphorylation of CheY when Ser83 phosphorylation does not follow and any level of $CheY_3$ and $CheY_4$ is sufficient to override any but the most transient of stops initiated by $CheY_6$(D57N) binding. This agrees with Appleby and Bourret (1999), who recorded that *E. coli* CheY(D57N) was able to bind to the flagellar motor and cause a tumble, but a loss of this activity when the additional phosphorylation site Ser86 was mutated to **A**.

All of the results strongly suggest that the polar cluster does not appreciably phosphorylate $CheY_6$-P. This contradicts the work of Tindall *et al.* (2010), who predicted that $CheY_6$-P acts as a phosphate sink for the rest of the polar cluster's response regulators using mathematical modelling. However, that work used a truncated model of chemotaxis that ignored the action of the CheBs and CheRs. It is possible that the phosphate sink activity seen is an artefact of this truncated model.

Whether the polar cluster's $CheY_3$-P and $CheY_4$-P inhibits the cytoplasmic cluster from phosphorylating $CheY_6$, or whether $CheY_3$-P and $CheY_4$-P directly bind to the motor to cause a run is unclear from these results. The working hypothesis of this thesis is that they bind directly to the motor, as this agrees with previous *in vitro* work (Ferre *et al.*, 2004).

3.4.3 Future work

The model of chemotaxis connectivity here presents two potential lines of enquiry.

If the polar cluster does indeed operate like a *B. subtilis* cluster and produces smooth swimming through phosphorylation of $CheY_3$ and $CheY_4$, then the binding of these proteins at the motor can be measured. Competition assays between the three phosphorylated CheYs will indicate whether the proposed model of $CheY_3$-P and $CheY_4$-

P overriding $CheY_6$ is feasible. If sensitivity of the cluster is controlled by $CheB_1$ and $CheR_2$ as proposed, do the chemoreceptors have a pattern of methylation similar to *B. subtilis*, with methylation and demethylation occurring on both attractant addition and removal? If so, does the system use both CheBs to achieve this, or are the multiple CheWs in the cluster involved?

If the cytoplasmic cluster is indeed setting the frequency of stopping under steady state conditions, it is likely from the chemotaxis mutants above that the adaptation machinery localised at this cluster been co-opted for the role. If so, there are adaptation sites on the cytoplasmic chemoreceptors. Do these behave in a similar fashion to other adaptable chemoreceptors, or is altered role a result of alterations of the mechanism?

There is a strong indication that the cytoplasmic cluster has a different role during photoheterotrophic growth, as only under these conditions has a chemotaxis response been measured after removal of signalling through the polar cluster (deletion of $CheA_2$, Martin *et al.* (2001a)). If this set of chemotaxis mutants are grown up photoheterotrophically, will the tethering results differ?

The rest of this thesis focuses on how methylation-based adaptation is used in *R. sphaeroides* to affect the chemotaxis system seen. The proposed model can be invalidated if the two pairs of adaptation proteins do not function as methylesterases and methyltransferases to control sensitivity of the two clusters' chemoreceptors, or achieve this by some other method. The model can also be invalidated if methylation sites within the two clusters cannot be identified.

Chapter 4

The *R. sphaeroides* chemoreceptors: classification and predicting methylation sites

In this chapter, previous work on the thirteen *R. sphaeroides* chemoreceptors is summarised and current gaps highlighted. Receptors are classified according to their CD lengths, as per Alexander and Zhulin (2007), and published consensus sequences are used to identify potential methylation sites. TlpT is identified as the most likely methylation candidate in the cytoplasmic cluster. McpH, McpJ and McpR are identified as the most likely candidates in the polar cluster.

4.1 Introduction

A model for *R. sphaeroides* chemotaxis was proposed in Chapter 3, relying heavily on the ability of the CheB and CheR proteins to sensitise and desensitise the chemoreceptors in both clusters, suggesting that methylation-based adaptation does occur at both clusters. Methylation-based adaptation is likely in *R. sphaeroides*, as methanol release on the increase and decrease of attractant in the environment has been recorded (Martin *et al.*, 2001b; Mantotta, 2002). However, a chemoreceptor involved in adaptation has not yet been identified.

There are two possible methods for identifying chemoreceptors that can be methylated. Potential sites can be mutated *in vivo* and the resulting changes in swimming behaviour measured. Alternatively, specific chemoreceptors can be purified and analysed for the changes in molecular weight associated with methylation and deamidation. The second method has been used successfully in *E. coli*, as all the components required for methylation are known. With so many unknowns in *R. sphaeroides*, it is preferable to keep the system intact and make single changes, as done in the previous chapter.

Either method benefits from prior knowledge of the chemoreceptors and the probability whether one is actually methylated. Alexander and Zhulin (2007) performed a systematic bioinformatics study of probable chemoreceptors in all sequenced bacterial genomes. That work can be used to identify potential methylation sites in the *R. sphaeroides* chemoreceptors.

In this chapter, prior knowledge of the thirteen chemoreceptors is summarised. Possible methylation sites are identified using published consensus sequences for all bacterial genomes (Alexander and Zhulin, 2007) and for specific sequences.

4.2 The *R. sphaeroides* chemoreceptors

Previous work on the thirteen chemoreceptors in *R. sphaeroides* is summarised in Table 4.1.

4.2.1 The MCPs

There are nine MCPs in the *R. sphaeroides* genome. Three are located on Chromosome I and six on Chromosome II.

Two of the Chromosome I MCPs are found within *che*Op$_1$ (McpA and McpB). The

Table 4.1: The *R. sphaeroides* putative chemoreceptors.

Mcp	Chromosome	Chemotaxis operon	Expressed	Localisation	Deletion phenotype	Classification[1]	length (residues)
A	1	*che*Op$_1$	no	not tested	none	not included	792
B	1	*che*Op$_1$	no	not tested	none	34H	560
M	1	separate locus	no	not tested	not tested	not included	687
E	2	separate locus	yes	not tested	none	34H	519
G	2	separate locus	yes	polar	none	34H	534
H	2	separate locus	yes	polar	chemotaxis abolished	34H	561
J	2	separate locus	yes	polar	superswarmer	34H	553
R	2	separate locus	yes	not tested	chemotaxis reduced	34H	572
V	2	separate locus	yes	not tested	none	34H	524
C	1	*che*Op$_2$	yes	cytoplasmic	chemotaxis abolished	not included	580
L	1	separate locus	yes	diffuse	chemotaxis reduced	not included	370
S	1	*che*Op$_1$	no	not tested	none	not included	595
T	1	*che*Op$_3$	yes	cytoplasmic	chemotaxis abolished	36H	567

[1](Alexander and Zhulin, 2007)

third is found on a separate locus not linked to other chemotaxis proteins (McpM). These three are assumed not to be expressed under laboratory conditions (Gould, 2006).

Little is known about the expressed transmembrane chemoreceptors, which are all found on Chromosome II in separate loci. Chromosome II is a chromid (Harrison *et al.*, 2010), formed later in *R. sphaeroides*'s evolution than Chromosome I. All these chemoreceptors fall within the 34H class of MCPs based on the length of the HCD (Section 1.2.1, Alexander and Zhulin (2007)), so presumably can form mixed trimers of dimers and can associate with one another in a cluster. Methylation has not yet been directly observed in 34H receptors, but Alexander and Zhulin (2007) predict that it is possible.

It is unlikely that these receptors are as highly specialised to one particular ligand as the *E. coli* receptors, as a single deletion of most of the expressed chemoreceptors either has no effect on chemotaxis (McpE (Thompson, 2005), McpG (Wadhams *et al.*, 2000), McpV (Thompson, 2005)) or results in decreased chemotaxis towards all attractants (McpR (Thompson, 2005)). Deleting McpJ produces has a unique superswarming phenotype towards certain attractants, suggesting that it is involved in modulating the chemotaxis response, possibly through methylation (Mantotta, 2002). Deleting McpH results in a loss of chemotaxis, suggesting that it is integral to either cluster formation or to signalling through the cluster, perhaps by tethering the CheB and CheR homologues, as in *E. coli*'s Tar and Tsr (Thompson, 2005).

$CheR_2$ localisation

$CheR_2$ is localised to the cluster, but its tethering point has not been identified. Five MCPs share the consensus for C-terminal pentapeptide tails given by Alexander and Zhulin (2007). The matching sequences in McpM (AHAGF) and McpE (LWGDY) are in the middle of the protein rather than at the C-terminal, so are unlikely to func-

tion as CheR tethers. The matching sequences for McpA (GWEDF), McpB (DFRRH) and McpR (EFRRH) are positioned at the tail of the protein. McpH is thus unlikely to be involved in the tethering of the CheBs and CheRs and most likely has some structural function in cluster formation.

As only McpR is expressed under laboratory conditions, it is the most likely candidate for $CheR_2$ tethering. In light of this, the phenotype obtained by deleting McpR is understandable: adaptation still occurs without the pentapeptide tether for $CheR_2$, but $CheR_2$'s efficiency is reduced, hampering the chemotaxis response. Chemotaxis ability has only been measured by swim plate activity, so no further information is available regarding how chemotaxis is affected.

4.2.2 The Tlps

Four transducer-like proteins are located on Chromosome I. TlpS is part of *che*Op_1 and is the only Tlp not expressed under laboratory conditions. Deleting TlpS has no effect on chemotaxis, but deleting TlpS in a *fla2*$^+$ mutant results in reduced chemotaxis (del Campo *et al.*, 2011).

Little is known about TlpL, which is the shortest putative soluble chemoreceptor and the only one not found within a chemotaxis operon. As it is not localised within the cell to a particular cluster, it may associate with either cluster or have a different role entirely. Deletion causes a reduction of chemotaxis ability as measured by swim plates, suggesting that it may associate with one or another of the clusters.

TlpT and TlpC are required for normal clustering. Both are expressed under laboratory conditions from within a chemotaxis operon. Deletion of TlpT causes complete dissociation of the cluster and a loss of chemotaxis (Porter *et al.*, 2002; Wadhams *et al.*, 2005), whereas deletion of TlpC allows for some cluster formation, but no chemotaxis (Wadhams *et al.*, 2002). TlpC is highly soluble (data not shown), whereas TlpT is insoluble (Dr M.A.J. Roberts, unpublished). The two may associate to form a

scaffold for the cytoplasmic cluster, in much the same way as the membrane-bound chemoreceptors do.

TlpT provides the binding site for $CheR_3$, which dissociates from the cluster if the last five residues of TlpT (GFGDF) are deleted (Dr M.A.J. Roberts, unpublished). The N-terminal region of TlpT is implicated in cytoplasmic cluster segregation during cell replication. TlpT is the only chemoreceptor to be classified in the Alexander and Zhulin (2007) database and most resembles the canonical MCP sequence. It is classified as a 36H receptor, one of the only classes where methylation-based adaptation has been directly observed.

4.3 Classification of chemoreceptors

As explained in Section 1.2.1, MCPs are classified according to the number of heptad repeats found in the highly conserved domain of the cytoplasmic domain (Moual and Koshland, 1996; Alexander and Zhulin, 2007), as shown in Figure 4.1. Five of the 13 *R. sphaeroides* chemoreceptors were not included in the Alexander and Zhulin (2007) database and have not yet been classified according to this system. The unclassified chemoreceptors are two of the receptors in *che*Op_1, McpA and TlpS, two of the remaining Tlps, TlpC and TlpL, and finally McpM.

The original database included all of the chemoreceptors identified in all of the sequenced bacterial genomes available at the time, which included *R. sphaeroides* 2.4.1., provided that those chemoreceptors could be classified using the study's automated scheme. The majority of chemoreceptors which were not included were either truncated or were cytoplasmic and did not align with the general chemoreceptor sequence used in that study. Before looking for methylation sites in all of the *R. sphaeroides* chemoreceptors, a classification attempt is made for the five unclassified receptors. If the receptors do not fit the classification scheme, they may have a structure too different from a canonical MCP to be involved in signalling and

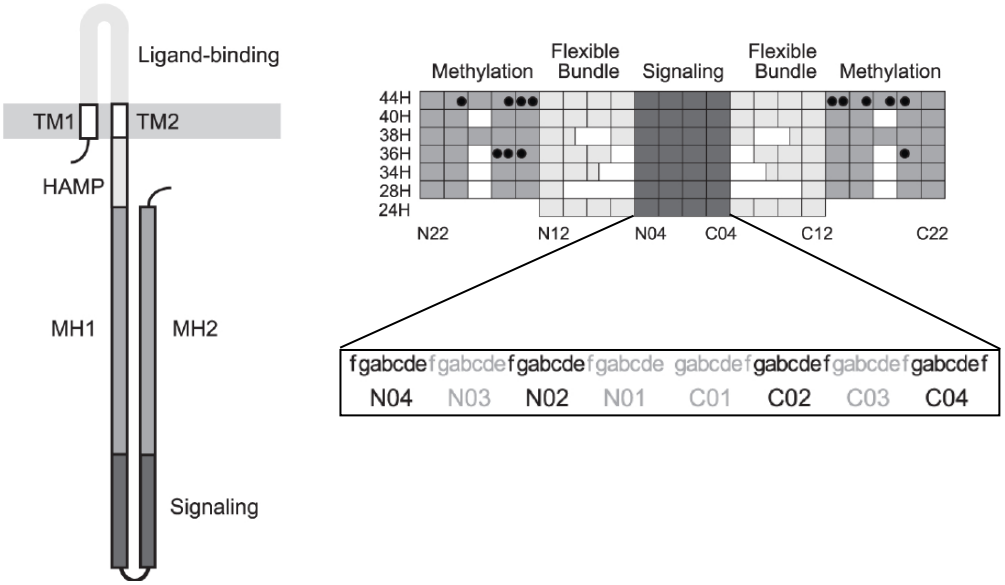


Figure 4.1: MCP classes and heptad positions Left: Cartoon showing the regions of an MCP. The methylation domains (MH1 and MH2) and the signalling domain form the highly conserved domain. Right, top: The 44 heptads and their positions in each class. White squares indicate deletions. Black circles indicate experimentally-determined methylation sites. Right, bottom: The numbering of the residues within each heptad, shown only for the heptads of the signalling domain. Taken from Alexander and Zhulin (2007) with permission, © 2007 National Academy of Sciences.

adaptation as expected.

4.3.1 Method

The protocol used in Alexander and Zhulin (2007) was used to classify the *R. sphaeroides* WS8N chemoreceptors not included in the original study. Three markers were identified in each receptor:

1. Any sensory domains or HAMP domains. The beginning of the HCD is defined as after any such domain.

2. The glycine turn motif connecting the two arms of α-helices, coded as **G-(ENDQ)-X-G-X-G**, where the second amino acid in the motif is the central position in the HCD.

3. A conserved C-terminus LLF motif, coded as **(ILMVQ)-6X-(ILMVQ)-6X-F**, which ends the majority of HCDs.

The HCD is preliminarily defined as running from the glycine turn to the end of the LLF, and starting an equal distance N-terminus to the glycine turn. Dividing this length by seven gives an initial estimate of the number of heptad repeats, and thus the chemoreceptor class.

The initial estimate is confirmed using multiple sequence alignments (MSA). The MSAs of each class from Alexander and Zhulin (2007) are used to generate hidden Markov models (HMM) using HMMR 3.0 (Finn *et al.*, 2011). The unknown chemoreceptor is then scanned against each HMM. A chemoreceptor can be assigned to an HMM if its bit score (Z score) is significantly (50 points) greater than the next highest bit score. The alignment can be used to trim the HCD limits set above, if needed. In most cases, the highest scoring MSA corresponds to the class suggested by the glycine turn to LLF score.

4.3.2 Results of the classification

Table 4.2 shows the location of the critical motifs and resulting heptad repeat classification for the chemoreceptors tested. The protocol was first verified by applying it to the known chemoreceptor TlpT. Both HMM scanning and glycine turn-LLF positioning gave the correct result: 36H.

The protocol was then used on the *R. sphaeroides* chemoreceptors not included in the original data set. The two MCPs were classified by both HMM scanning and positioning to give the same result. McpA was classified as a 36H receptor and McpM as a 38H receptor.

The cytoplasmic chemoreceptors proved more problematic. Both TlpC and TlpS do not include a glycine turn motif, although both include multiple LLF motifs, so cannot be classified using positioning. Neither could be classified using HMM

Table 4.2: HCD classification of the *R. sphaeroides* chemoreceptors that were not included in the Alexander and Zhulin (2007) database.

chemoreceptor	TlpT	McpA	McpM	TlpC	TlpS	TlpL
length (amino acids)	567	792	687	580	595	370
sensory domains (from SMART)	HAMP 110 - 163	PAS 245-311	none	none	HAMP 108 - 211	none
C-terminus cytoplasmic domain (from Phobius)	1 - 567	186 - 792	408 - 687	1 - 580	1 - 595	1 - 370
GXXGXG mid position (use 2nd position in motif)	376	632	549	none	none	267
LLF end positions (use final 17th position in motif)	171 , 502	243 , 758	294 , 682	54 , 295	233, 425, 520	none
HCD domain (directly from GXXGX and LLF)	250 - 502	506 - 758	416 - 682	1 - 580	1-595	164 - 370
HMM database matching	36H	36H	38H	weak to 24H and 28H	none	34H
Final HCD domain (after trimming)	250 - 502	506 - 758	416 - 682	unknown	unknown	176 - 358
Number heptad repeats	36	36	38	unknown	unknown	26 (from 34)

scanning: TlpC gives a low but significant score against 24H and 28H alignments, whereas TlpS has no significant score against any alignment. As both do not include a glycine turn, it is not surprising that they cannot be classified with other chemoreceptors. Without a glycine turn, the structure of these proteins must be quite different from the classical two α-helical arms.

TlpL contains a glycine turn, but no known sensory domain and no LLF motif. This makes positioning unhelpful in defining the HCD. HMM scanning shows the 34H MSA as the most likely match, with a highly significant bit score of 210 units, 71 units greater than the next hit, and an E score (probability that the match is based on random chance) of 1.6 x 10^{-65}. If the glycine turn is used as the only indication of the HCD's position and symmetry around the turn is required, a maximum of 206 amino acids form the HCD. This gives an HCD of less than 30 heptad repeats. This suggests that TlpL is a 34H chemoreceptor with deletion of at least four heptad repeats.

Aligning the full length of TlpL to the MSA for the 34H group shows good homology for the central heptads: N/C1-5b, and N/C8a-16f (Figure 4.2). TlpL shares the 34H deletions of N/C5a - N/C8b. TlpL also appears to have an additional deletion of three to four heptads at its ends.

The edges of the alignments do not initially match up to the boundaries of the heptads found in the 34H sequence, as the alignment indicates matching to C19-c, but N16-f. Upon inspection, however, the C-terminus alignment is highly unlikely for the final few amino acids. The alignment is pushed by an aspartate at C19-c, however this residue is wholly conserved as glutamate in the MSA. Taking this into account, the final amino acids of the TlpL alignment can be discounted, and the alignment taken from C16-f. This then gives equal ends on either side of the TlpL HCD, and gives a full deletion of four heptads from each end.

TlpL is thus possibly a 26H chemoreceptor, generated from the deletion of four pairs


McpE/259-497 EVMDEVSGTARDVAAASGRMATAAGQ.....LKSGTEEQASATVEASAAVEQMVGNITQSE 314
McpG/257-495 DVVGRVTAAVGDVASGSAEVAATSEQ.....LSQGASEQAAATVQASASVEEIAATVRQSA 312
McpH/262-500 EVLGRVWLATDQVAANSQTMAATSEQ.....LSQGSSEQAASTEEASASVEEMAANIRQTA 317
McpJ/257-495 EVVGGVSSVARQVSSGSGEMAATSEQ.....LSQGASEQASATEEASASVEQMAANIKQAA 312
McpR/267-505 EMVGTITGAVNRVSSGSSSMASTSEE.....LSQGAQEQASATAEASASVEQMAANIRQTA 322
McpV/267-505 TVVEAVTASSGALTGLSARVAATSEQ.....MSVNAMEQISATGEAASAVTGIQSSLEQSV 322
TlpL/176-358 .....EVAEEIARALVGASSAMEEVTGTIRHTA 203
N22 N21 N20 N19/17 N16 N15 N14 N13
McpE/259-497 SNAGVTAQIAGSSADDARRCGVAVAEAVRSM.....QEIAERIGVVREIARQTD 363
McpG/257-495 ENSGQTERMARSSAEAARQSGAAVSDAVSAM.....RGIAERIHVVQEIARQTD 361
McpH/262-500 DSAGETERIAAKSAEDARASGDAVREAVAAM.....ASIADRILIVQEIARQTD 366
McpJ/257-495 DNAMQTERIASKAAEDARASGAAVEEAVGAM.....RSIADKIMMVQEIARQTD 361
McpR/267-505 DNAGNTETVARTSADRAQASGSAVNEAVDAM.....QAIAERILVVQEIARQTD 371
McpV/267-505 ERAAACESVAARAAQDARASEKVVGEALRSM.....KKIAEHILVMQEIARQTD 371
TlpL/176-358 GNASETEQLAAQNARAAETGGAAVERSADAM.....RLIADKINVLREIARQTD 252
N12 N11 N10 N09 N08/05 N04 N03
McpE/259-497 LLALNAAVEAARAGEQGRGFAVVAAEVRRLAERSAEAAAEISTLSAG.....TA 412
McpG/257-495 LLALNAAVEAARAGEHGRGFAVVATEVRRLAERSQAAAAEISDLSSA.....TA 410
McpH/262-500 LLALNAAVEAARAGEHGRGFAVVASEVRKLAERSQAAAAEISALSAR.....TS 415
McpJ/257-495 LLALNAAVEAARAGEHGRGFAVVASEVRKLAERSQAAAAEISQLSAG.....TV 410
McpR/267-505 LLALNAAVEAARAGEHGRGFAVVAAEVRKLAERSRGAAEEISALSAR.....TL 420
McpV/267-505 LLALNAAVEAARAGERGLGFAVVASEVRKLAERSQDAAKEVSTLSHS.....TS 420
TlpL/176-358 LLALNAAVEAARAGQHGAGFSVVAAEVRKLAEHAAAASHEIDQLAHT.....TL 301
N02 N01 C01 C02 C03 C04 C05/08

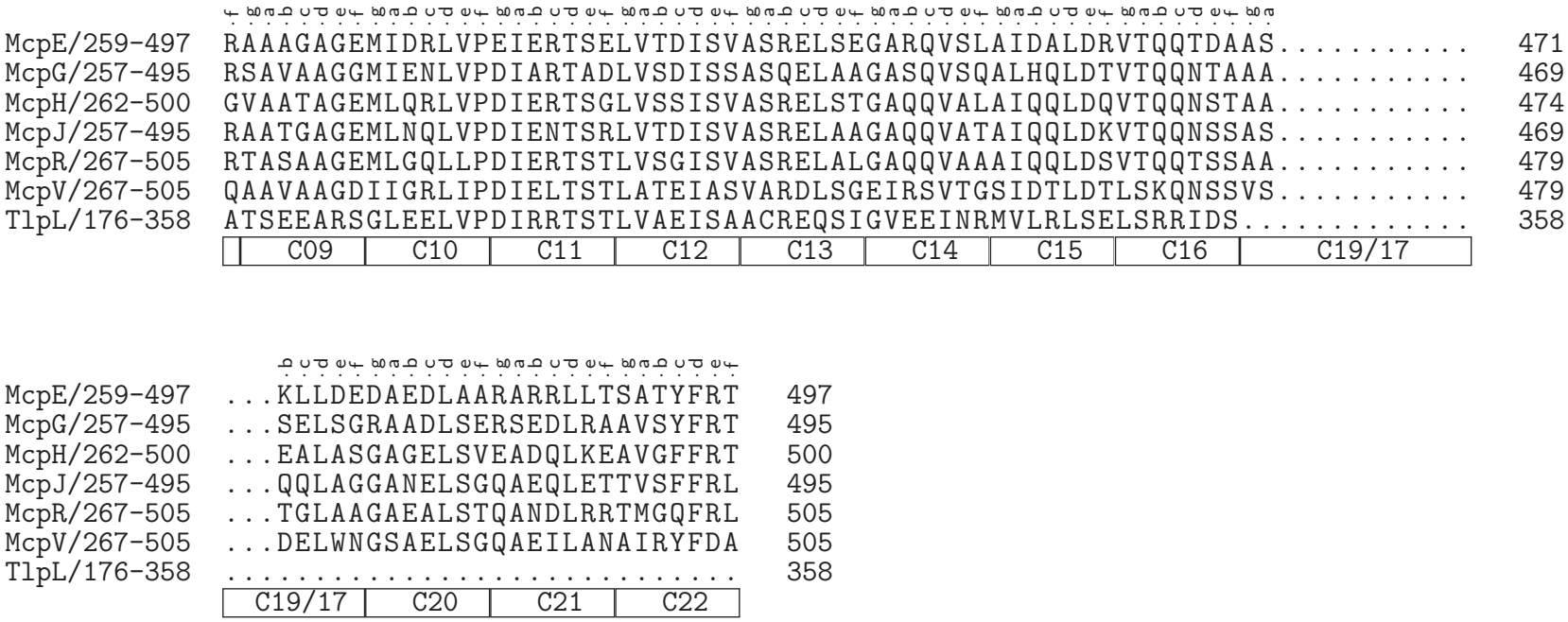


Figure 4.2: Final TlpL alignment with the 34H chemoreceptor MSA. Only the *R. sphaeroides* 34H chemoreceptors are shown. The top ladder shows the position within each heptad and the bottom ladder indicates the heptad number. All of the heptad numbering is according to Alexander and Zhulin (2007).

of heptads from the ends of 34H HCD. The position of this loss explains the lack of LLF motif in TlpL, as this is found in the C20-C22 heptads.

4.4 Putative methylation sites

Methylation has been observed experimentally in chemoreceptors in the classes 44H and 36H (Figure 4.1, top right). Consensus sequences from these experimentally observed sites have been used to identify further sites within the same organisms. The 36H *E. coli* sequence, also valid for *Salmonella* Typhimurium, is **(EQ)-(EQ)-X-X-A-(ST)** (Terwilliger *et al.*, 1986) and the 44H *T. maritima* sequence is **(AS)-(ASTG)-X-(EQ)-E-X-(ASTG)-(AS)** (Perez *et al.*, 2006).

Alexander and Zhulin (2007) postulated that the 34H and 38H receptors are also capable of methylation. They proposed a methylation consensus sequence based on all of the known methylation sites: **(ASTG)-(ASTG)-X-X-(EQ)-(EQ)-X-X-(ASTG)-(ASTG)**. The key feature of this consensus sequence is the requirement for small amino acids within the helical turn upstream and downstream of the **QE/QE** pair. Alexander and Zhulin (2007) applied this consensus sequence to their database of classified chemoreceptors. Figure 4.3 shows all of these chemoreceptors, grouped by heptad class. Every match to the consensus sequence is shown as a black line.

The putative matches form general patterns in each heptad class, except 40H and 24H. Figure 4.4 shows the proposed general pattern in five of the seven major heptad classes. The proposed patterns in the 44H and 36H classes match the positions of the experimentally determined methylation sites in *B. subtilis* (Glekas *et al.*, 2011) and *T. maritima* (44H) (Perez *et al.*, 2006) and in *E. coli* (36H) (Terwilliger and Koshland, 1984; Kim *et al.*, 1999).

Methylation sites have not yet been experimentally determined in the class 34H chemoreceptors, which is the class of the *R. sphaeroides* polar chemoreceptors. There-

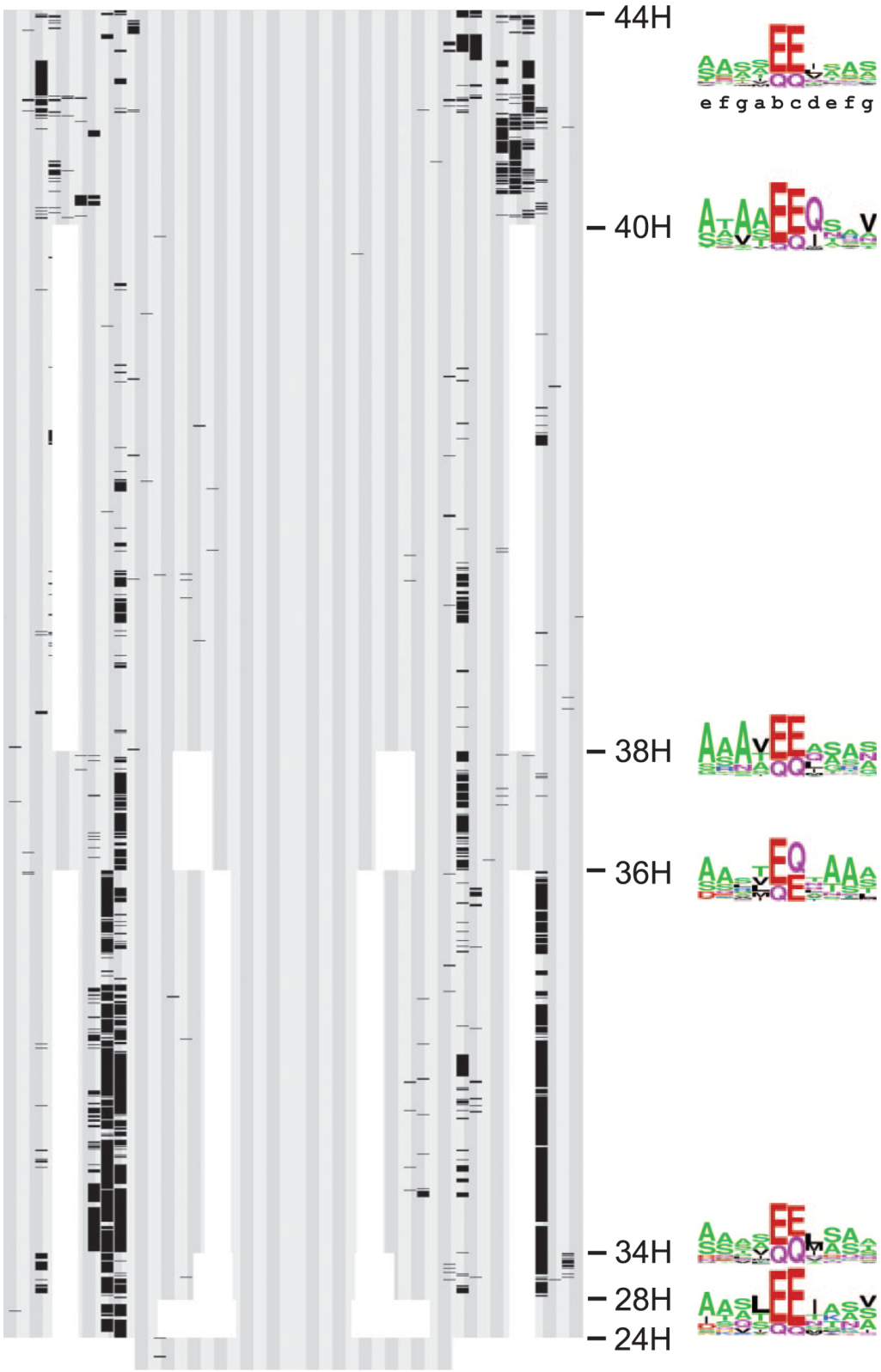


Figure 4.3: Alexander and Zhulin (2007) applied their proposed methylation consensus sequence to 1,656 classified MCPs. The MCPs aligned and grouped according to heptad class. The 44 heptads (N22-C22) are indicated by alternating gray shading and heptad deletions in each class are shown as white squares. Each match to the consensus sequence is shown as a black line. For each class, a sequence logo is shown of the most common residues at each position in the methylation sequence, collated from all of the matches to the consensus sequence. Taken from Alexander and Zhulin (2007) with permission, © 2007 National Academy of Sciences.

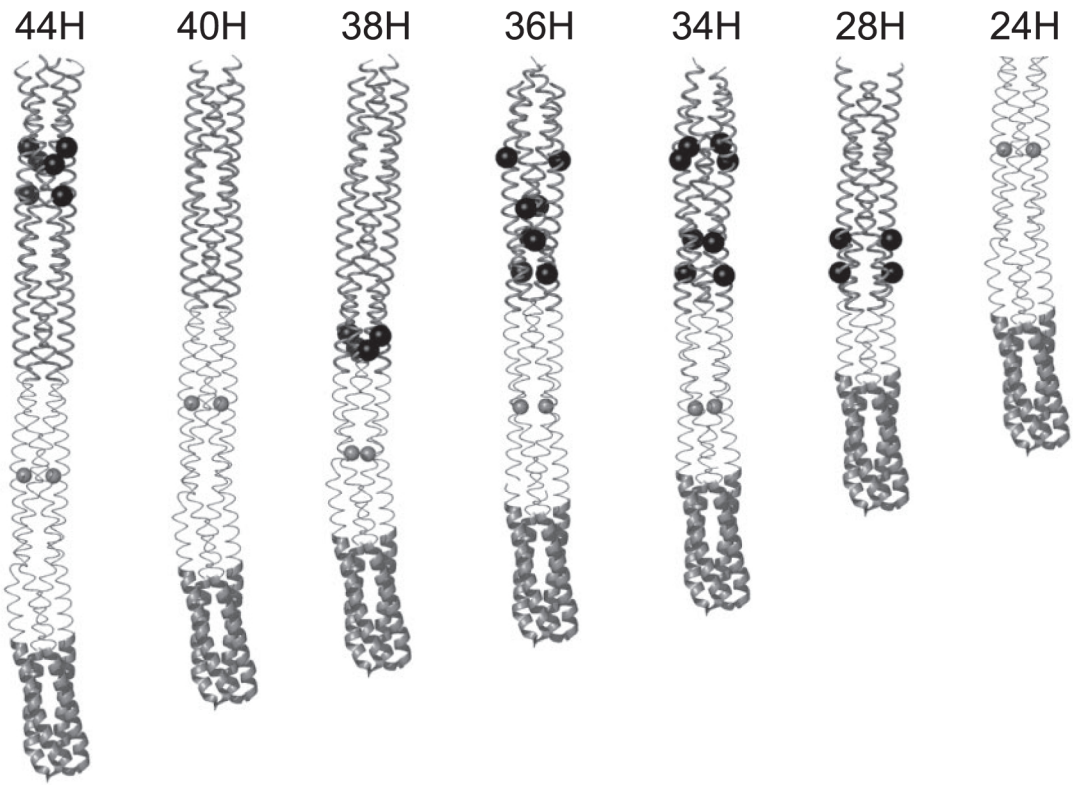


Figure 4.4: Alexander and Zhulin (2007) proposed a methylation pattern in five of the seven major heptad classes, based on the matches in 1,656 classified MCPs to their consensus sequence. The MCP cartoons show, from the bottom, the signalling domain in thick, dark grey ribbon, the flexible linker region thin, light grey and the methylation domain in thin, dark grey. The light grey small circles within the flexible linker are a pair of glycine residues around which the region flexes. The black large circles within the methylation domain are the proposed general positions for methylation in each class. Taken from Alexander and Zhulin (2007) with permission, © 2007 National Academy of Sciences.

fore, the general Alexander and Zhulin (2007) consensus sequence was used to search for possible methylation sites. The *E. coli* consensus sequence was used, as *R. sphaeroides* contains one 36H receptor (TlpT). The *T. maritima* consensus sequence was also used; as nothing is known about 34H receptors, it is worth exploring all possible options.

4.4.1 Putative sites in *R. sphaeroides* receptors matching known consensus sequences

The predicted pattern for 34H receptors is three or four methylation sites, as the clusters of black predicted sites in the 34H section of Figure 4.4 show. Two of the sites lie close to one another on the N-arm of the double helix in heptads 17-15. One or two lie on the C-arm of the helix, in heptads 19-22, although some extend to 13. As helices have regular turns and **QE** residues must be on the open face of the helix, any predicted residues should be separated by multiples of seven residues.

Each *R. sphaeroides* chemoreceptor was scanned for matches to the *E. coli* (Figure 4.5), *T. maritima* (Figure 4.6) and more general Alexander and Zhulin (2007) (Figure 4.7) consensus sequences. The 34H, 36H and 38H receptor HCDs are aligned according to the Alexander and Zhulin (2007) heptad numbering system in these figures.

None of the receptors are likely to be involved in methylation, on the basis of the sites that match each of the consensus sequences. None of the consensus sequences predict three or four sites in the classical 34H or 36H pattern on any of the chemoreceptors. Even if the sites predicted by all three consensus sequences are considered together, the majority of the chemoreceptors only have one or two sites predicted sites. McpJ has three predicted sites according to the *E. coli* consensus sequence, however two of those sites are six residues, not seven, apart, so are highly unlikely to function as methylation sites in reality.

TlpC and TlpS were not classified. They could not be divided into canonical MCP domains. They cannot be aligned with the rest of the receptors: due to the highly repetitive nature of MCP sequences, with the repeating heptad motif, multiple poor alignments are possible. Without an anchoring middle point and a clear indication of which heptads are deleted in these receptors, there is no way to select a correct alignment. These receptors are therefore not shown in Figures 4.5-4.7. Neither of these receptors has any matches to the *E. coli* or Alexander and Zhulin (2007) con-

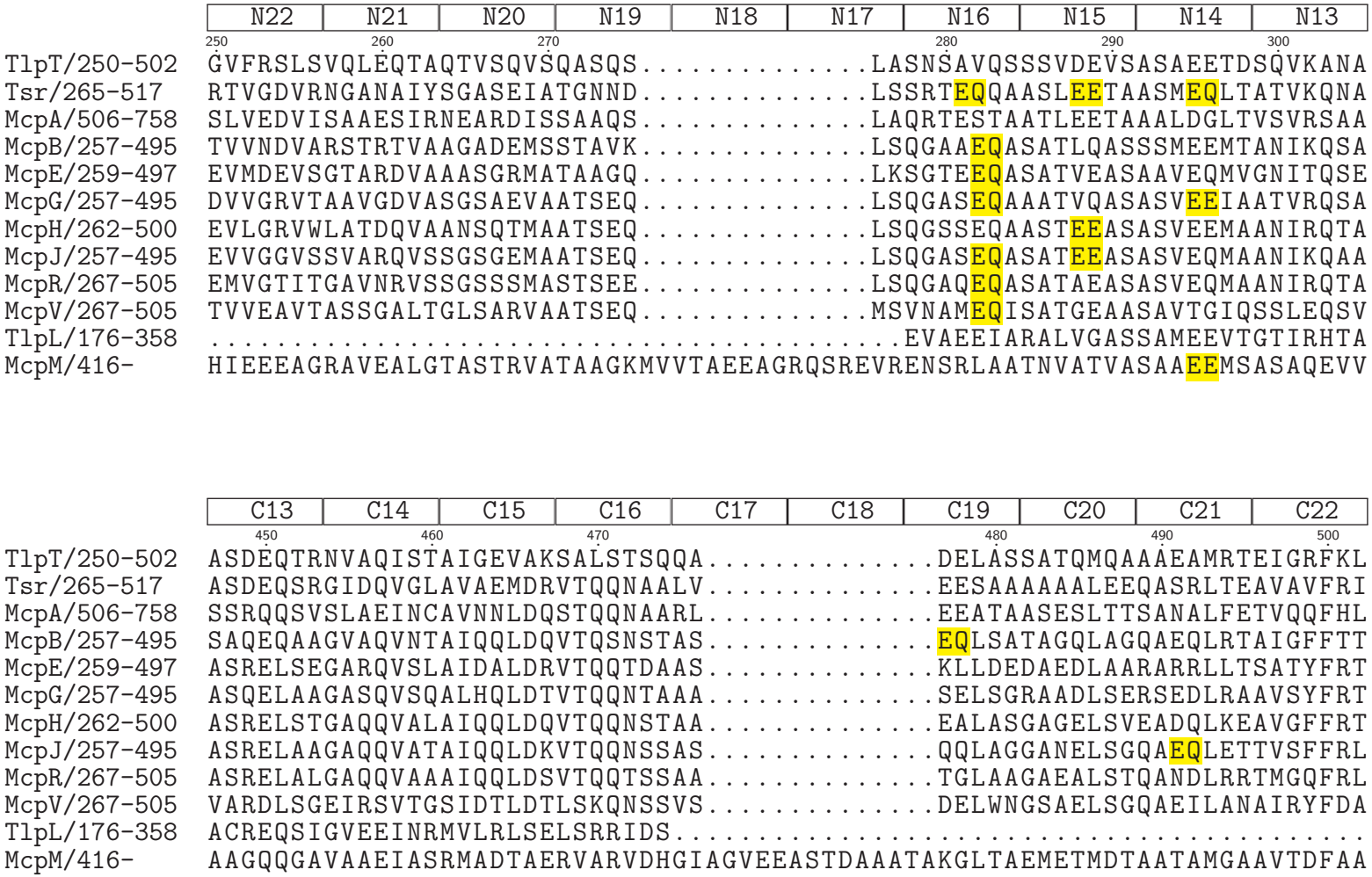


Figure 4.5: The **QE/QE** pairs in the *E. coli* and *R. sphaeroides* chemoreceptor methylation domains (heptads 13-22) that match the *E. coli* consensus sequence **(EQ)-(EQ)-X-X-A-(ST)** are highlighted. The chemoreceptors are aligned according to the heptad numbering used in Alexander and Zhulin (2007). The top ladder residue numbers are relative to TlpT.

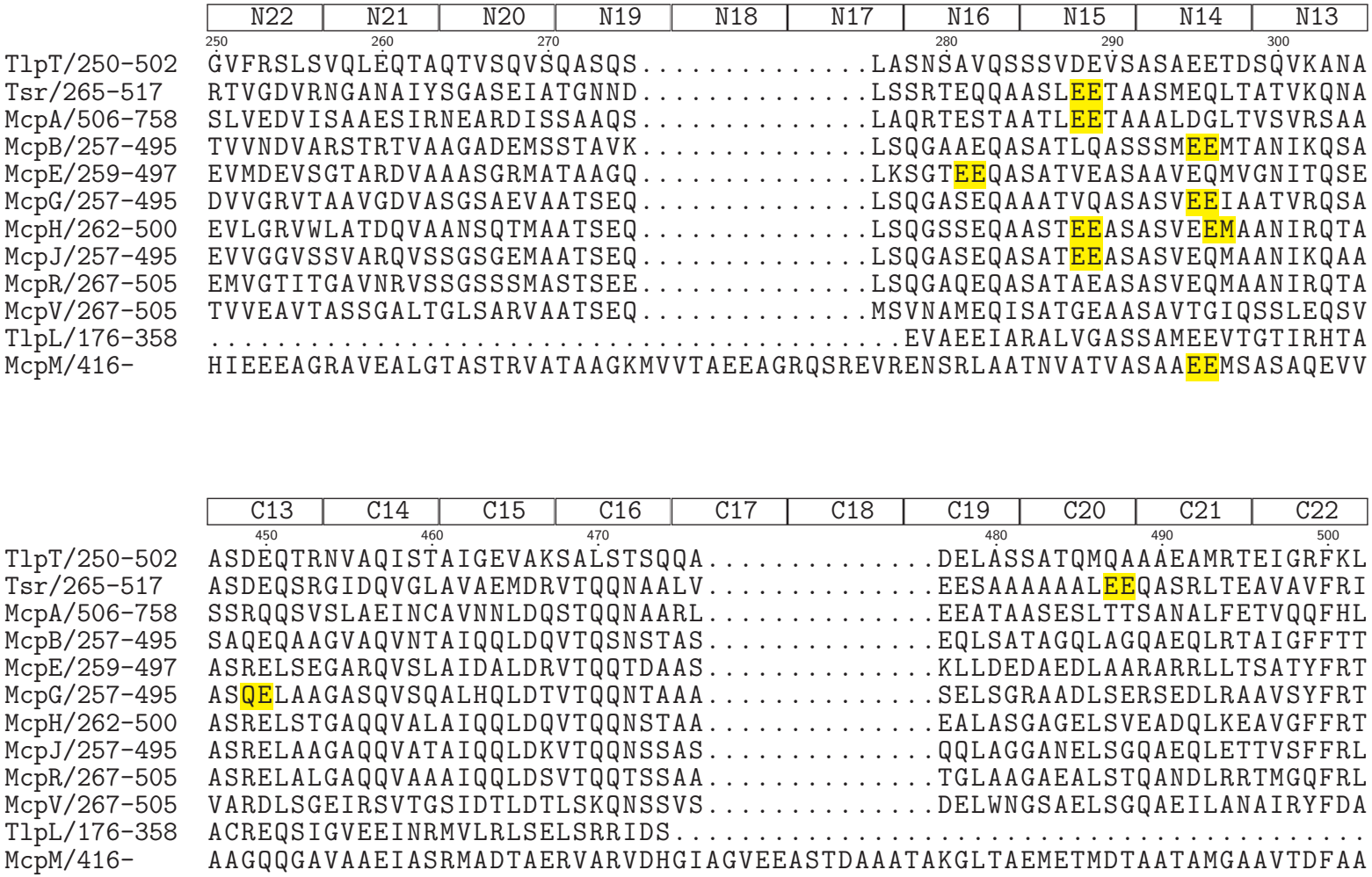


Figure 4.6: The **QE/QE** pairs in the *E. coli* and *R. sphaeroides* chemoreceptor methylation domains (heptads 13-22) that match the *T. maritima* consensus sequence **(AS)-(ASTG)-X-(EQ)-E-X-(ASTG)-(AS)** are highlighted. The chemoreceptors are aligned according to their heptads. The top ladder residue numbers are relative to TlpT.

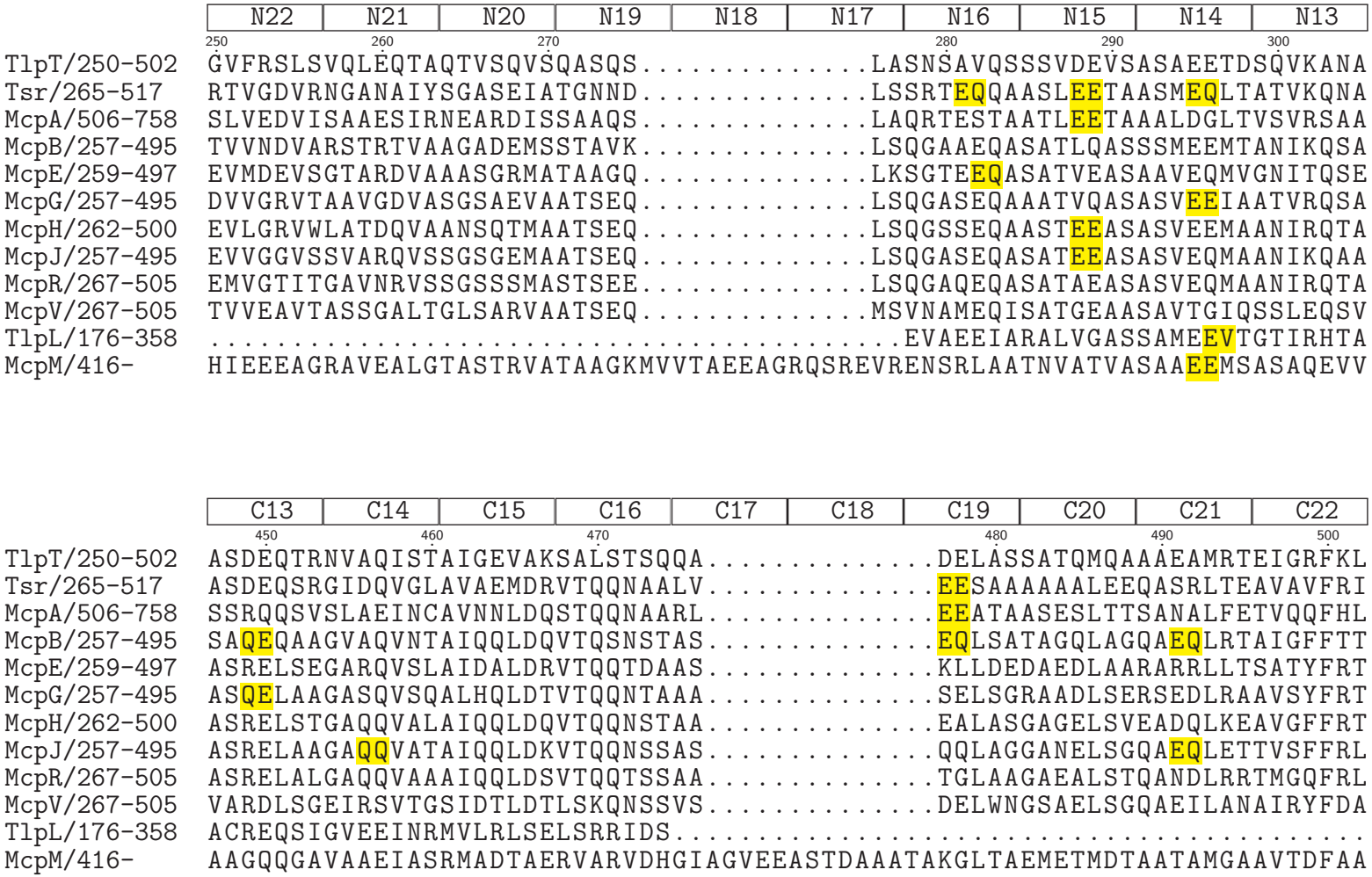


Figure 4.7: The **QE/QE** pairs in the *E. coli* and *R. sphaeroides* chemoreceptor methylation domains (heptads 13-22) that match the Alexander and Zhulin (2007) consensus sequence **(ASTG)-(ASTG)-X-X-(EQ)-(EQ)-X-X-(ASTG)-(ASTG)** are highlighted. The chemoreceptors are aligned according to their heptads. The top ladder residue numbers are relative to TlpT.

sensus sequences. TlpS has one match to the *T. maritima* sequence, N-terminal to the HAMP domain, which is highly unlikely to be a valid methylation site.

It is not necessary to predict all of the methylation sites in a receptor before testing the sites on that receptor by either *in vitro* or *in vivo* methods. However, none of the receptors seem to be a good candidate for further study from these results.

4.4.2 Putative sites in *R. sphaeroides* receptors matching relaxed consensus sequences

It is possible that the *R. sphaeroides* chemoreceptors are not involved in methylation-based adaptation, or that they have fewer sites per receptor than other studied species. It is more likely, however, that the *R. sphaeroides* methyltransferases and methylesterases recognise a slightly different active site. Comparing the known consensus sequences and relaxing some of their constraints will give a useful prediction sequence to identify potential sites for further study. Once a site is positively identified by other means, a consensus sequence for *R. sphaeroides* chemoreceptors can be made.

The consensus sequences all use various groupings of tiny and small amino acids to buttress the **QE/QE** pair. The group **ASTG** is a subset of the tiny and the small amino acids, both hydrophobic and polar (Figure 4.8). This group can be extended to include all tiny amino acids **ACSG** and all small-not-tiny amino acids **DNPTV**.

The three known consensus sequences were compared to one another and two prediction sequences, Up and Down, were built (Table 4.3). The three consensus sequences covered eleven consecutive residues. The least constraining option for each position was selected from across the three consensus sequences, i.e., if the options at a site were **X** or **ASTG**, then that position became **X**. All positions with some subset of small amino acids were relaxed to allow all small amino acids.

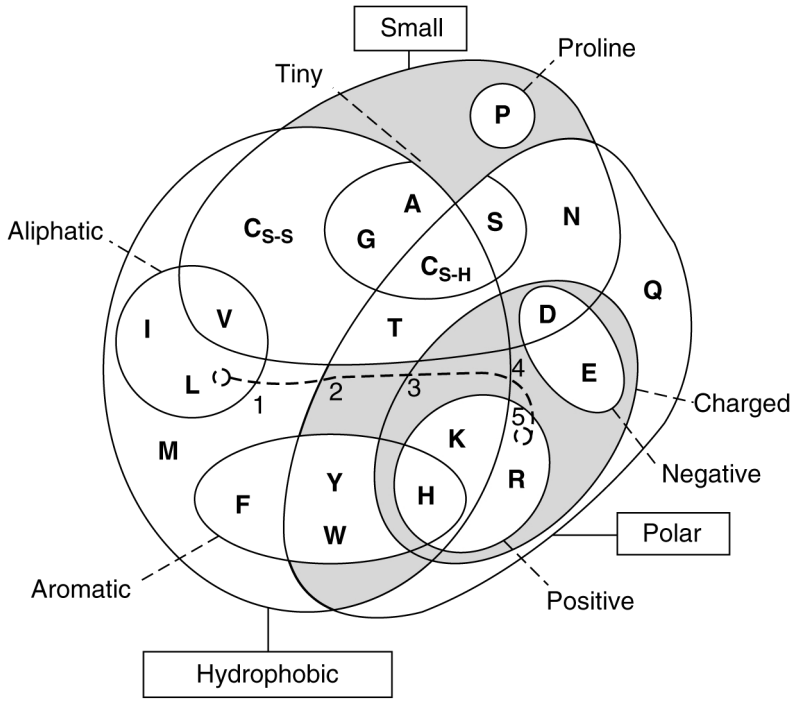


Figure 4.8: Properties of the amino acids. Taken from Betts and Russell (2003) with permission from John Wiley and Sons, license number 3693190627696.

Table 4.3: Generating a 'relaxed' sequence for prediction of methylation sites.

	heptad position									
	e	f	g	a	b	c	d	e	f	g
EC					EQ	EQ	X	X	A	ST
TM		AS	ASTG	X	EQ	EQ	X	ASTG	AS	
AZ	ASTG	ASTG	X	X	EQ	EQ	X	X	ASTG	ASTG
Up	small	small	X	X	EQ	EQ				
Down					EQ	EQ	X	X	small	small

AZ = Alexander and Zhulin (2007), EC = *E. coli*, TM = *T. maritima*

The resulting eleven residue sequence was then cut in half to give prediction sequences Up and Down. It is likely that these prediction sequences are too lax and will identify additional non-methylation sites. Considering that CheR swings from the C-terminal pentapeptide tether to reach sites on both the N- and C-arms of the helix, it is conceivable that the enzyme remains in the same orientation to methylate both arms and thus needs residues buttressing the target **QE/QE** site on either side to be a certain shape. However, these prediction sequences are not to be taken as consensus sequences for *R. sphaeroides* methylation sites, but are meant as an exploratory tool to identify sites most likely to be methylation targets. The *E. coli*

consensus sequence includes all Tsr sites and itself only represents the region downstream of the sites.

All of the residues matching the prediction sequences Up and Down in the *R. sphaeroides* chemoreceptors are highlighted in Figure 4.9. As expected, the prediction sequences generate false positives in Tsr, as the sequences were chosen to prioritise false positives over false negatives.

4.5 Discussion

Swimming behaviour in chemotaxis mutants suggests that both chemotaxis clusters use some form of methylation-based adaptation. The polar cluster is predicted to demethylate on addition of attractant, causing smooth swimming, then remethylate to adapt and return to steady state stopping frequencies. The cytoplasmic cluster is predicted to control steady state stopping frequencies and its activity is regulated by the action of $CheB_2$ and $CheR_3$, through methylation and demethylation. The continued methanol release pattern seen in the non-chemotactic TlpC deletion strain (and thus cytoplasmic cluster-less) strains (Martin *et al.*, 2001b; Mantotta, 2002) argues for methylation-based adaptation independent of the environment in this cluster.

There are a number of possible methods for testing whether methylation-based adaptation occurs. Considering the complexity of the *R. sphaeroides* system, the methods most likely to give clear results require focusing on one chemoreceptor at a time. Chemoreceptors can be purified and tested *in vitro* or mutated *in vivo* and phenotypes studied. As these methods are time-intensive and expensive, only the most likely methylated receptors should be selected for further study.

The *R. sphaeroides* MCPs and Tlps were scanned for potential methylation sites. All the chemoreceptors have at least one site matching at least one of the prediction

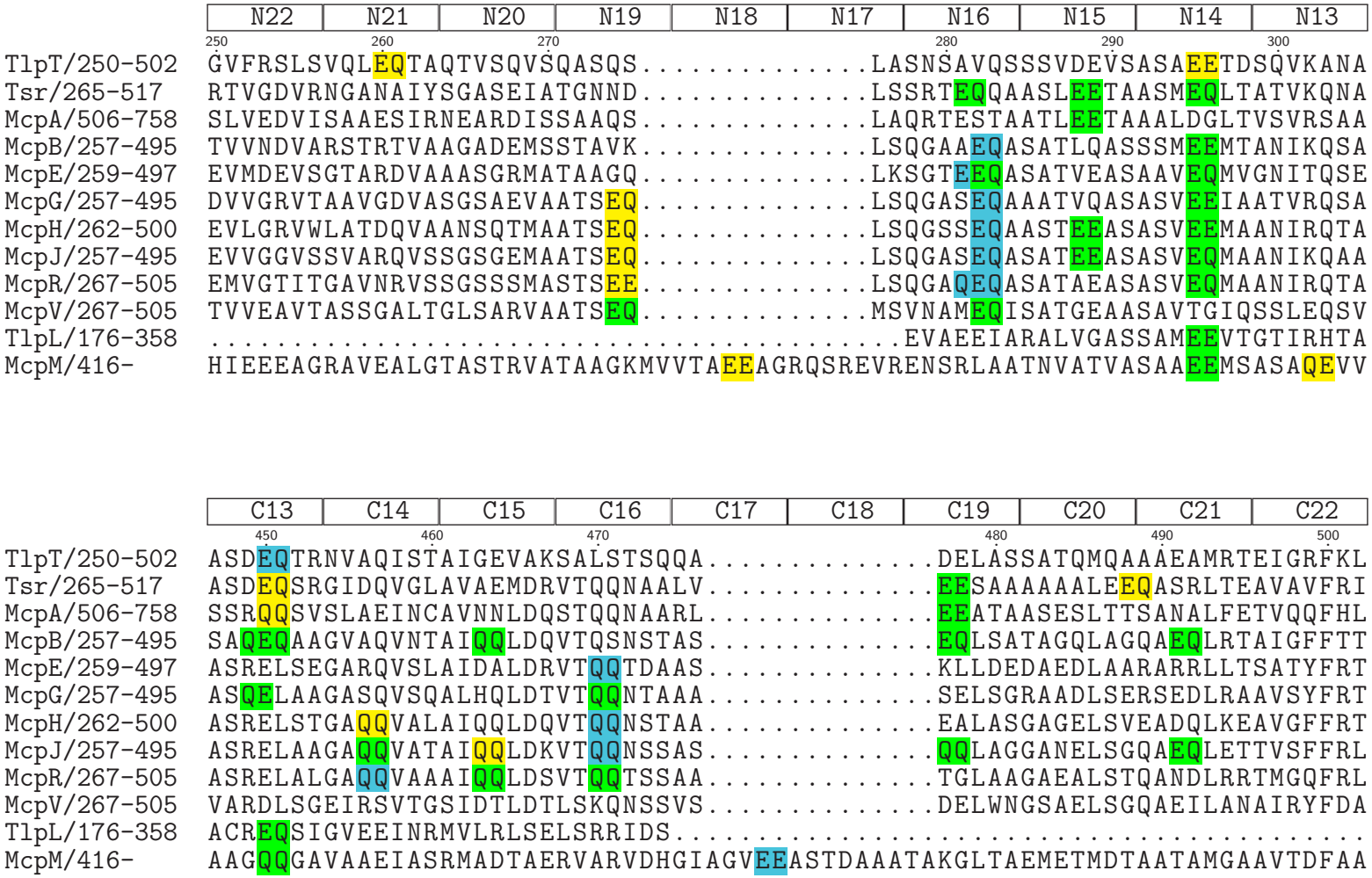


Figure 4.9: The **QE/QE** pairs in the *E. coli* add *R. sphaeroides* chemoreceptor methylation domains (heptads 13-22) that match the Up (yellow) and Down (blue) prediction sequences are highlighted. Sites matching both sequences are highlighted in green. The chemoreceptors are aligned according to their heptads. The top ladder residue numbers are relative to TlpT.

sequences used. None of these sites follow the predicted pattern of methylation for the appropriate heptad class predicted by Alexander and Zhulin (2007), as shown in Figure 4.3. Site location, structural information, deletion phenotypes and other information are used to select the most likely receptors in each cluster.

A deletion phenotype showing an effect on chemotaxis efficiency is not a guarantee that the receptor in question is involved in adaptation. Similarly, receptors may be involved in adaptation and yet not show a deletion phenotype unless all of the adaptation receptors are deleted. However, it seems likely that a receptor showing an adaptation-related phenotype when deleted is involved in adaptation and that these receptors will prove easier to study, as a phenotype change can be tracked.

4.5.1 Polar cluster chemoreceptors

The polar cluster 34H receptors all have two or three predicted methylation sites on the N-terminal arm of the helix, most positioned in multiples of seven residues from neighbouring sites. The often-predicted N16 site is the exception, so is likely not a true methylation site. The C-terminal arm has differing numbers of sites. McpV has none, which strongly suggests that either this receptor is not methylated or that its methylation sites are quite different from known sites in other receptors. The rest of the 34H receptors are equally strong candidates, with multiple sites on both arms. Additional information is thus taken into account.

McpH, McpJ and McpR are the likeliest candidates for methylation. Deletion of any causes disruption to chemotaxis and McpR is predicted to tether $CheR_2$. McpH has sites in N19, N15 and N14 (N16 is out of frame) on the N-arm and C14 and C16 on the C-arm. McpJ has sites in N19, N15 and N14 (N16 is out of frame) on the N-arm and C14, C15, C16, C19 and C21 on the C-arm. McpR has sites in N19, N16 and N14 on the N-arm and C14, C15 and C16 on the C-arm. It is likely that not all these sites are correct, as previous predictions have capped the number of methylation sites on

any one receptor at four.

4.5.2 Cytoplasmic cluster chemoreceptors

In the cytoplasmic cluster, the inability to classify TlpS and TlpC suggests that these receptors may have a structure different from classical MCPs, making methylation-based adaptation as is currently described unlikely. The sites matching Up and Down prediction sequences in TlpC and TlpS are also positioned differently from those in classical MCPs. The methylation sites on classical MCPs are only found in the HCD, are separated from one another by multiples of seven residues and in general follow the pattern predicted by Alexander and Zhulin (2007). The four matching sites in TlpC are spaced out across the protein's full length and are not separated from one another by multiples of seven residues. Three sites are predicted in TlpS, one N-terminal to the HAMP domain and the other two on what would be opposite arms of the helix if it forms like a classical MCP. Thus, both the structure and the position of putative methylation sites in TlpC and TlpS differ sufficiently from classical MCPs that these putative chemoreceptors are unlikely to function similarly to classical MCPs in any way, including methylation-based adaptation.

Two sites are predicted in TlpL, lying in heptads N14 and C13. Little can be said about these sites, as there is nothing known about TlpL's class of receptor.

Three sites are predicted in TlpT, lying in heptads N21, N14 and C13. The two sites on the N-arm lie a multiple of seven residues apart. These sites do not fit the usual 36H receptor methylation pattern. However, it is similar to the 44H *B. subtilis* pattern. Considering what is known now about the possible pattern of methylation in the polar cluster, it is possible that the cytoplasmic cluster too resembles *B. subtilis*. Again, it is possible that there are additional unpredicted sites.

TlpT is known to play a critical role in the structure of the cytoplasmic cluster. Its deletion leads to an unresponsive phenotype, unable to respond to chemoeffector

concentration changes, and a loss of cytoplasmic cluster formation. As the only expressed 36H receptor, it also has the most similarity to the *E. coli* receptors known to be involved in methylation-based adaptation.

Considering the amount of information on TlpT already known and TlpT's role as a tether for $CheR_3$ (Dr M.A.J. Roberts, personal communication), TlpT is the cytoplasmic chemoreceptor of choice for further study.

Chapter 5

In vivo exploration of the predicted TlpT methylation sites

In this chapter, the predicted methylation sites in the cytoplasmic chemoreceptor TlpT were tested using *in vivo* methods. In *E. coli*, mutating methylation sites results in measurable changes in the adaptation response. The TlpT sites were mutated to alanine and to aspartate and the swimming behaviour under steady state and dynamic conditions analysed. The majority of the predicted sites were dismissed as candidates for methylation, although Q475 mutant phenotypes do suggest that legitimate methylation sites exist in that region. Site E296 could not be validated or invalidated as a methylation site.

5.1 Introduction

The swimming behaviour of chemotaxis mutants under dynamic and steady state conditions has suggested that in *R. sphaeroides*, at least under aerobic conditions, the cytoplasmic cluster produces a baseline concentration of $CheY_6$-P, controlling the frequency of runs and stops under steady state conditions. The polar cluster likely responds to increases in chemoeffector concentrations by biasing motion towards runs. Whether the cytoplasmic cluster responds to external chemoeffector concen-

tration drops to bias motion towards stops is as yet unclear. For the cytoplasmic cluster to set the baseline stopping frequency, its activity must be tuned, similarly to the way the *E. coli* polar cytoplasmic cluster is tuned by the adaptation proteins CheB and CheR. It is likely that this tuning is in response to the concentrations of metabolites in the cell that function as chemoeffectors. The candidate proteins for this tuning are the adaptation proteins, one or more of the CheBs, as they are diffuse, and $CheR_3$, which is localised to the cytoplasmic cluster.

If the adaptation homologues in the cytoplasmic cluster are tuning the cluster's activity, it is likely through reversible methylation, as that is how these proteins usually achieve their effect. In Chapter 4, the most likely cytoplasmic chemoreceptor for these effects was identified as TlpT and sites for methylation were predicted. In this chapter, these sites are tested by mutagenesis and phenotyping of the resultant strains.

5.2 Position of predicted methylation sites on TlpT

There is as yet no confirmed structure for TlpT. Repeated attempts in the laboratory to purify full-length TlpT for crystallisation have been unsuccessful (Dr M.A.J. Roberts, unpublished).

There are a number of X-ray crystallography and NMR-generated structures of the periplasmic domain and the CD of Tsr. Hall *et al.* (2012) has produced a molecular dynamics theoretical structure of the full length Tsr, using known structures for HAMP domains in other organisms to join the periplasmic and CD structures. Coarse grain simulations of 200 ns were used to find a stable conformation.

TlpT and Tsr are both classified as 36H receptors by Alexander and Zhulin (2007) and share significant sequence similarity of the CD. A theoretical structure of full length TlpT was generated by threading TlpT against the Tsr theoretical structure

and using molecular dynamics techniques to find the most likely stable structure (James Allen, unpublished data). This structure has not been tested experimentally. The positions of any positively identified methylation sites can be used to comment on the validity of this structure.

Methylation-based adaptation occurs on one of the residues in a **QE/QE** pair on the outer face of the α helix of a receptor dimer. At present, methylation is usually observed on the second residue in these pairs. Although *T. maritima* has been shown to methylate on the first residue (Perez *et al.*, 2006), *E. coli* and other 36H chemoreceptor containing species only show methylation on the second residue. As TlpT is a 36H receptor, it is assumed that it follows the same pattern as Tsr.

Four potential sites were identified on TlpT using the prediction sequences Up and Down in Chapter 4: Q261, E296, Q451 and Q528. Q528 falls outside the defined HCD region and does not have a heptad position number, so was not included in the discussion in Chapter 4 (for the full discussion of the classification of chemoreceptor cytoplasmic domains using heptad repeats, see Sections 1.2.1 and RsChemoreceptorClass). However, as nothing is known about cytoplasmic cluster packing, this site is included in the mutagenesis list. A fifth site, Q475, is included as a negative control site. Q475 is also a **QE/QE** pair, but does not lie within frame to the other mutations, with respect to the usual position of methylation sites within heptads. If mutations at this position show the same results as at the identified positions, then it is likely that the changes we see are not due to methylation, but due to packing changes.

Methylation sites on Tsr are accessible to CheB and CheR (Figure 5.1, left). The side chains are orientated out of the double helix of the monomer and are found on the opposite side of the helix to the interacting face between the two monomers within the dimer.

Mapping the predicted methylation sites onto the theoretical structure for TlpT does

not show the same pattern (Figure 5.1, right). Instead, in the case of Q261, Q451 and Q475, the side chains are orientated into the centre of the dimer. If this structure is correct, then deamidation of these sites is unlikely under physiological conditions, and mutation to **E** is likely to cause disruption of the packing of the dimer. If any of these sites is positively identified as a methylation site, this structure is highly unlikely. E296 alone is found on the surface of the helix and is available for possible methylation. If methylation occurs at this site but not the others, the structure is more likely to be valid.

Q528 is not shown as the flexible C-terminal region of an MCP cannot be resolved through crystal structure and is not included in either the Tsr or TlpT structure. However, no methylation site has yet been identified out of the HCD region and on this flexible region, making this an unlikely candidate for methylation.

5.3 Genomic mutations in TlpT

All except one of the putative methylation sites are **Q** residues. They will require deamidation post-translation to become functional as methylation sites. Mutation of these sites to **E** is a useful positive control, as the **Q** to **E** mutation should be silent in a methylation site, but may cause phenotypic changes if in any other site due to the structural effect of adding a charged residue.

Putative sites were also mutated to **A** and to **D**.

The **Q/E** to **A** mutation sets the site to the permanently neutral charge state, promoting tight packing. This may be analogous to a state that is permanently methylated and sensitive, allowing the receptor to respond to a chemoeffector drop. It is not demethylatable; the receptor should not be able to recover to an insensitive state after a response, increasing the adaptation time (Nowlin *et al.*, 1988).

The **Q/E** to **D** mutation sets the site to the permanently negative charge state, dis-

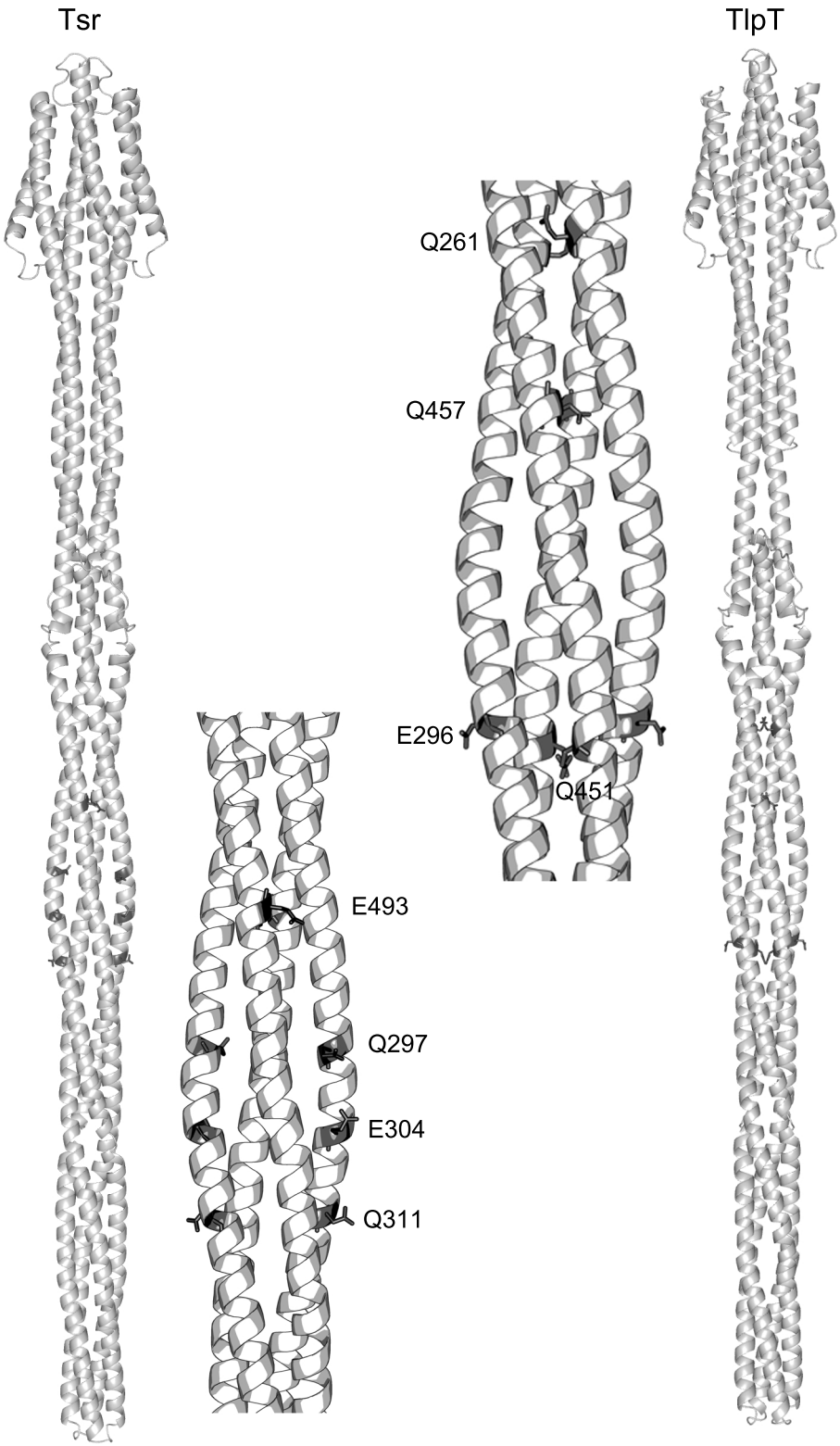


Figure 5.1: Theoretical structures of Tsr (left) and TlpT (right). Known (Tsr) and predicted (TlpT) methylation sites are shown in dark grey on full structure and zoomed in regions.

rupting packing. This is analogous to the permanently demethylated and insensitive state; the receptor should be less able to activate the cytoplasmic kinases in response to a change in chemoeffector concentrations, whether they are internal or external changes. These residues cannot be methylated (Shapiro and Koshland, 1994).

The methylation rate of sites up and downstream of the mutations may also be affected by the mutations (Nowlin *et al.*, 1988; Shapiro *et al.*, 1995).

Strategy for genomic mutations

Point mutations in TlpT were made by inserting mutated TlpT into a TlpT deletion mutant, using variants of the suicide vector pK18-TlpT. This suicide vector contains *tlpT*, flanked by the 500 bp up and downstream of its usual position in the genome. The construct was initially inserted into pK18*mobsacB* with restriction enzymes SphI and EcoRI. Strains TlpT(Q451A), TlpT(Q475A) and TlpT(Q528A) were made by Ayse Ozhan in a previous study, so the making of these is not discussed further. Strains TlpT(Q261A), TlpT(Q261D) and TlpT(Q296A) were made by Qin Qi for this study. All other new strains in this chapter were made by Elaine Byles for this study.

Construction of pK18-TlpT variants Point mutations were made in TlpT using overlap extension PCR (Section 2.2.4) and taking advantage of convenient restriction digest sites. pK18-TlpT was used as the template DNA for the two PCR reactions in round one, using primer pairs F1/R1 and F2/R2. Both products of round one were used as the template for round two, using primer pair F1/R2.

For all mutations in sites Q261 and E296, the same F1 (TlpT-Q261E296-F1) and R2 (TlpT-Q261E296-R2) were used. F1 used the AgeI site starting at position 587 in *tlpT* and R2 used the AatII site starting at position 1203. Specific primers R1 and F2 coded

for each mutation (e.g., TlpT-Q261A-R1) and bound to the region to be mutated. For all mutations in sites Q451, Q475 and Q528, the same F1 (TlpT-Q451Q475Q528-F1) and R2 (TlpT-Q451Q475Q528-R2) were used. F1 used the AatII site starting at position 1203 in *tlpT* and R2 used the Bsu361 site starting 138 bp 3′ from the end of *tlpT*. Specific primers R1 and F2 coded for each mutation (e.g., TlpT-Q451A-R1) and bound to the region to be mutated. All primers are listed in Appendix A.

The PCR products of round two were inserted into pK18-TlpT. The vector and inserts were digested with AgeI and AatI (sites Q261 and E296) or AatI and Bsu361 (Q451, Q475, Q528) before ligation.

Conjugations and recombinations Each TlpT point mutation was inserted into the *R. sphaeroides* TlpT deletion strain using the standard method for allelic exchange with a suicide vector (Section 2.2.4). The two rounds of homologous recombination between the suicide vector and genomic DNA were selected for using Kan for round one, then sucrose in the absence of Kan for round two. Insertion was confirmed using the Southern blot technique (Section 2.2.4).

The Southern blot probe was prepared by digesting pK18-TlpT with AgeI-HF and Bsu36I , producing a 1200 bp band corresponding to the region overlapping the end of TlpT and the region downstream of TlpT. Extracted chromosomal DNA was digested with PvuII and SphI, cutting respectively upstream of the region complementary to pK18-TlpT and within the final downstream 500 bp of the pK18-TlpT insert. Reversion to the parental strain would generate a band of around 800 bp and a successful insertion a band of around 2500 bp.

Thirteen of the fifteen strains were successfully made. The two failed strains, the mutants TlpT(E296A) and TlpT(Q528E), failed repeatedly at the second recombination stage.

5.3.1 Cluster formation after TlpT mutation

Deletion of TlpT results in complete delocalisation of the cytoplasmic chemotaxis cluster proteins. To ensure that the mutant TlpT strains were expressing TlpT and that the mutated TlpT is still able to form clusters, the cytoplasmic cluster protein $CheW_4$ was tagged with CFP. This has been done in wildtype cells with no effect on cluster formation or on functionality.

An existing pK18 construct (pK18-W4-CFP) was used to make the C-terminal CFP tag on $CheW_4$, using the same recombination protocol as for the point mutants.

Cytoplasmic clusters formed in all but one strain (Q475D), suggesting that stable TlpT is produced (data not shown).

5.4 *In vivo* analysis of putative methylation sites

In *E. coli*, loss of each methylation site causes a measurable loss of adaptative ability: the methylation sites' action is additive (Levit and Stock, 2002; Nowlin *et al.*, 1988). Thus, in TlpT, loss of a methylation site may produce a measurable phenotype change distinct from the deletion phenotype. Steady state stopping frequency, ability to respond during the dynamic state and adaptation times after response may all be affected. If the sites mutated are not involved in methylation but are structural, then mutations may result in phenotypic changes closer to that of the deletion phenotype.

The effects of the mutations in TlpT was analysed using the free swimming assay described in Chapter 3. Chemotaxis ability was analysed on a population level using swarm plates and at the individual cell level using the tethered cell assay described in Chapter 3.

In the chemotaxis assays, the analysis was extended to include two growth condi-

tions, aerobic and photoheterotrophic. The stimulus that activates the cytoplasmic cluster is currently uncertain. If it involved the cell's energy state, then it is possible that differences between the two growth states will be seen in chemotaxis. Differences have been seen in the response to external chemoeffector concentrations in aerobic and photoheterotrophic cells (Martin *et al.*, 2001a). The free swimming assay does not allow a clear distinction to be made the effects of an aerobic and a photoheterotrophic environment, as once the capillary is sealed, a micro-aerobic to anaerobic environment will form. A full photoheterotrophic screen is therefore not run. If any strains show a marked difference between the two growth states in the tethering assay, then the free swimming can be extended to include these under photoheterotrophic conditions.

A combined discussion follows the presentation of the three sets of results.

5.4.1 Free swimming analysis of individual cells

18 min of free swimming video footage was collected of each TlpT mutant, the track data extracted and tracks censored using the parameters in Table 3.2 (Section 2.5.1). The number of tracks in each data set before and after censoring again varies widely (Table 5.1). THe wildtype, gutted and nonmotile data from Chapter 3 were reused. Representative tracks from each strain are shown in Figure 5.2.

Once again, representative tracks are easily classified by eye into smooth and wildtype/stoppy, due to the long straight tracks and corkscrews that characterise smooth swimmers, but distinguishing tracks stoppier than wildtype is difficult.

The TlpT(Q528A) mutant shows an unusual motility phenotype. Under shaking aerobic conditions, this strain is nonmotile at all OD_{700nm}. However, after 20 min stationary incubation at room temperature the population becomes motile, albeit a very stopping motion with a high proportion of nonmotile and drifter cells. Further incubation increases the proportion of motile, non-drifter cells, as can be seen

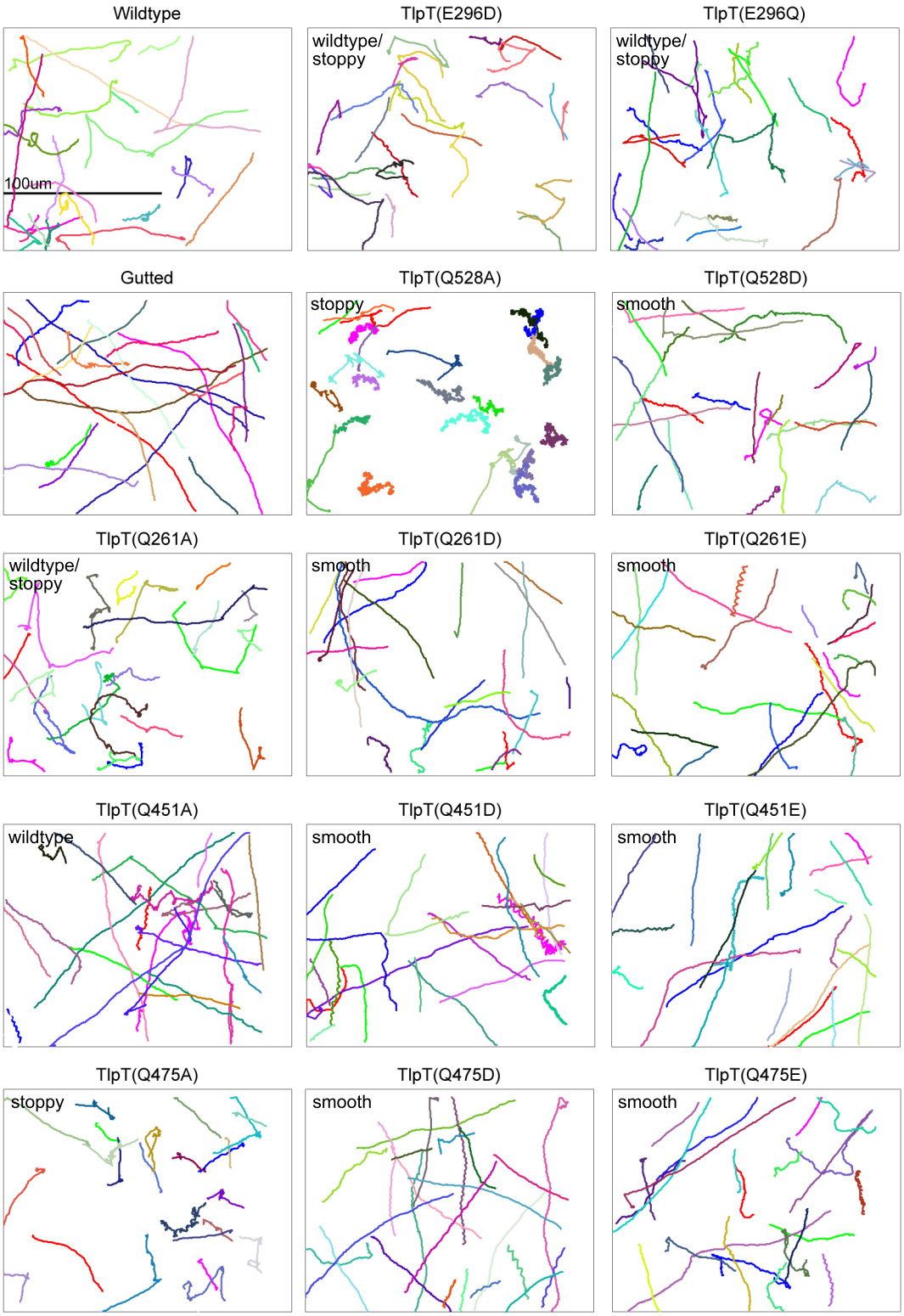


Figure 5.2: Representative free swimming tracks from the TlpT mutants, after censoring. The black bar on wildtype indicates 100 μm.

Table 5.1: Number of free swimming tracks removed at each censoring round for each aerobically grown TlpT mutant. The numbers in brackets indicate the number of tracks that would be removed from the nonmotile data set if that censoring step was used.

	Starting tracks	Tracks removed at each stage				Remaining tracks
		Top speed	MBR	Length	Tortuosity	
Wildtype	13841	1989	8387	345	312	2808
Nonmotile	5327	706	(4578)	(0)	–	4621 (43)
Gutted	6672	908	4839	105	82	738
TlpT(Q261A)	11493	1721	7507	201	207	1857
TlpT(Q261D)	4617	534	3188	107	79	709
TlpT(Q261E)	6297	785	4462	107	95	848
TlpT(E296D)	13112	2108	7984	239	279	2502
TlpT(E296Q)	9704	1452	5659	206	239	2148
TlpT(Q451A)	7177	1087	4215	175	170	1530
TlpT(Q451D)	6923	1014	4422	176	132	1179
TlpT(Q451E)	4683	646	3125	100	82	730
TlpT(Q475A)	10893	1613	8042	77	117	1044
TlpT(Q475D)	5879	733	3923	165	106	952
TlpT(Q475E)	6079	787	3805	185	131	1171
TlpT(Q528A)	7172	972	5991	7	21	181
TlpT(Q528A) after 1 h	6861	861	5778	4	22	196
TlpT(Q528A) after 2 h	7428	938	6089	11	39	351
TlpT(Q528D)	10600	1545	6317	335	241	2162

in Figure 5.3. A much higher percentage of tracks are removed during censoring in these data sets, due to the proportion of nonmotile cells, but the percentage surviving censorship increases with incubation time. Restoration of motility occurs both in the presence and absence of Chl, suggesting that this is does not require protein expression, but instead due to the lifting of some inhibition.

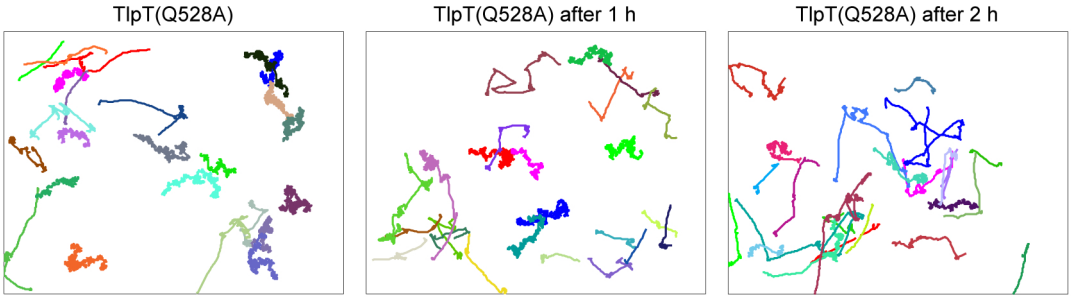


Figure 5.3: Representative free swimming tracks after censoring, from aerobically grown TlpT(Q528A), after 20 min, 1 h and 2 h stationary incubation at room temperature.

In comparison, when the TlpT(Q528A) mutant is grown photoheterotrophically, it swims smoothly with no stops (data not shown). Incubating this strain after it

has been grown photoheterotrophically under aerobic conditions does not show a change in motility pattern, either in the presence or absence of Chl.

Tracks were classified into runs and stops as before. The fraction of time spent stopped is again used to summarise the data set (Figure 5.4). A Kruskal-Wallis test rejects the hypothesis that all data sets are from the same population ($p<0.001$). Mutants are grouped by pairwise comparisons with a Dunn-Sidak test (Table 5.2). Members remain in a group if $p>0.05$ for all members. Groups with overlapping members are combined.

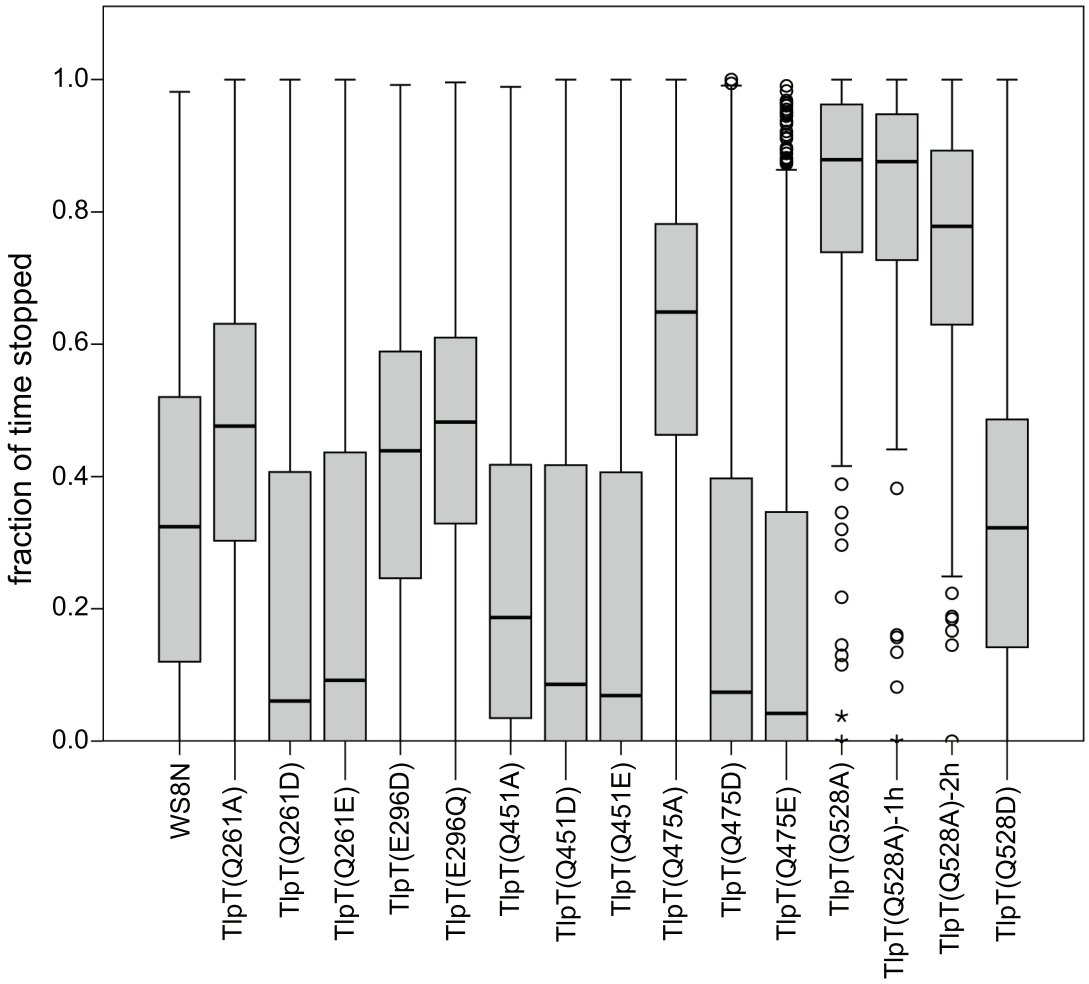


Figure 5.4: Fraction of time stopped TlpT mutant spends stopped during free swimming. The boxplot edges indicate the first and third percentiles, the box centre indicates the median, the whiskers indicate the range $\pm$ 1.5x the box height, and the circles indicate outliers (Section 2.6.1).

As with the chemotaxis mutants, the TlpT mutants are split into two clear groups

Table 5.2: Pairwise comparisons between the distributions of the fraction of time spent stopped during free swimming for each TlpT mutant, giving the probability that a pair is from the same population. Comparisons not shown are $p<0.001$.

	Wildtype	TlpT (Q475A)	TlpT (Q528A)	TlpT (Q528A) 1h	TlpT (Q528A) 2h	TlpT (Q261A)	TlpT (E296D)	TlpT (E296Q)	TlpT (Q451E)	TlpT (Q528D)
Wildtype		0	0	0	0	0	0	0	0	1
TlpT(Q475A)	0		0	0	0	0	0	0	0	0
TlpT(Q528A)	0	0		1	1	0	0	0	0	0
TlpT(Q528A) 1h	0	0	1		1	0	0	0	0	0
TlpT(Q528A) 2h	0	0	1	1		0	0	0	0	0
TlpT(Q261A)	0	0	0	0	0		0.001	1	0	0
TlpT(E296D)	0	0	0	0	0	0.001		0	0	0
TlpT(E296Q)	0	0	0	0	0	1	0		0	0
TlpT(Q451E)	0	0	0	0	0	0	0	0		0
TlpT(Q528D)	1	0	0	0	0	0	0	0	0	

	TlpT (Q451A)	TlpT (Q261E)	TlpT (Q261D)	TlpT (Q451D)	TlpT (Q451E)	TlpT (Q475D)	TlpT (Q475E)
TlpT(Q451A)		1	0.026	0.355	0.356	0.067	0
TlpT(Q261E)	1		1	1	1	1	0.001
TlpT(Q261D)	0.026	1		1	1	1	1
TlpT(Q451D)	0.355	1	1		1	1	0.025
TlpT(Q451E)	0.356	1	1	1		1	0.521
TlpT(Q475D)	0.067	1	1	1	1		0.474
TlpT(Q475E)	0	0.001	1	0.025	0.521	0.474	

(Table 5.3). The mutants with a greater proportion of time spent running than wild-type form one significant group. All of the mutants but one (the TlpT(Q451A) mutant) in this group have boxplots with the third quartile including 0.0 and can be considered smooth swimmers rather than smoother than wildtype. The rest of the strains form a range of small groups or singletons with increasing median time stopped from wildtype to almost nonmotile (the TlpT(Q528A) mutant).

5.4.2 Swim plate analysis of population response

The ability of a strain to swim through a soft agar matrix with a metabolism-generated nutrient/chemoeffector gradient gives a visual indication of the chemotaxis ability of that strain. 5 μl of stationary culture is inoculated onto a set plate of M22 media with 0.25% agar and 100 μl propionate, along with three standards; wildtype, non-chemotactic and nonmotile (Section 2.5.2). As cells grow and divide, attractant is metabolised, creating a concentration gradient. The agar matrix is sufficiently soft for motile strains to swim through it. Motility and chemotaxis affect the diameter of the resulting colony.

The strains showing mutations at the same site were analysed on the same swim plate, so that changes at a single site could be directly compared. The spread of measurements for each strain are shown in box-plot form in Figures 5.5 and 5.6. A one-way balanced ANOVA was performed on each set of site mutations alongside the three standards of wildtype, non-chemotactic and nonmotile.

In all cases, the null hypothesis that all data sets are from the same population was rejected. A post-hoc Tukey-Kramer test was performed to determine which measurement sets did belong to the same population group. This test assumes all sets are taken from the same population and splits the group only if the difference within-group is significant ($p<0.05$). The resulting groupings are shown in Table 5.4. The only mutant with a significant difference between the two growing conditions is the

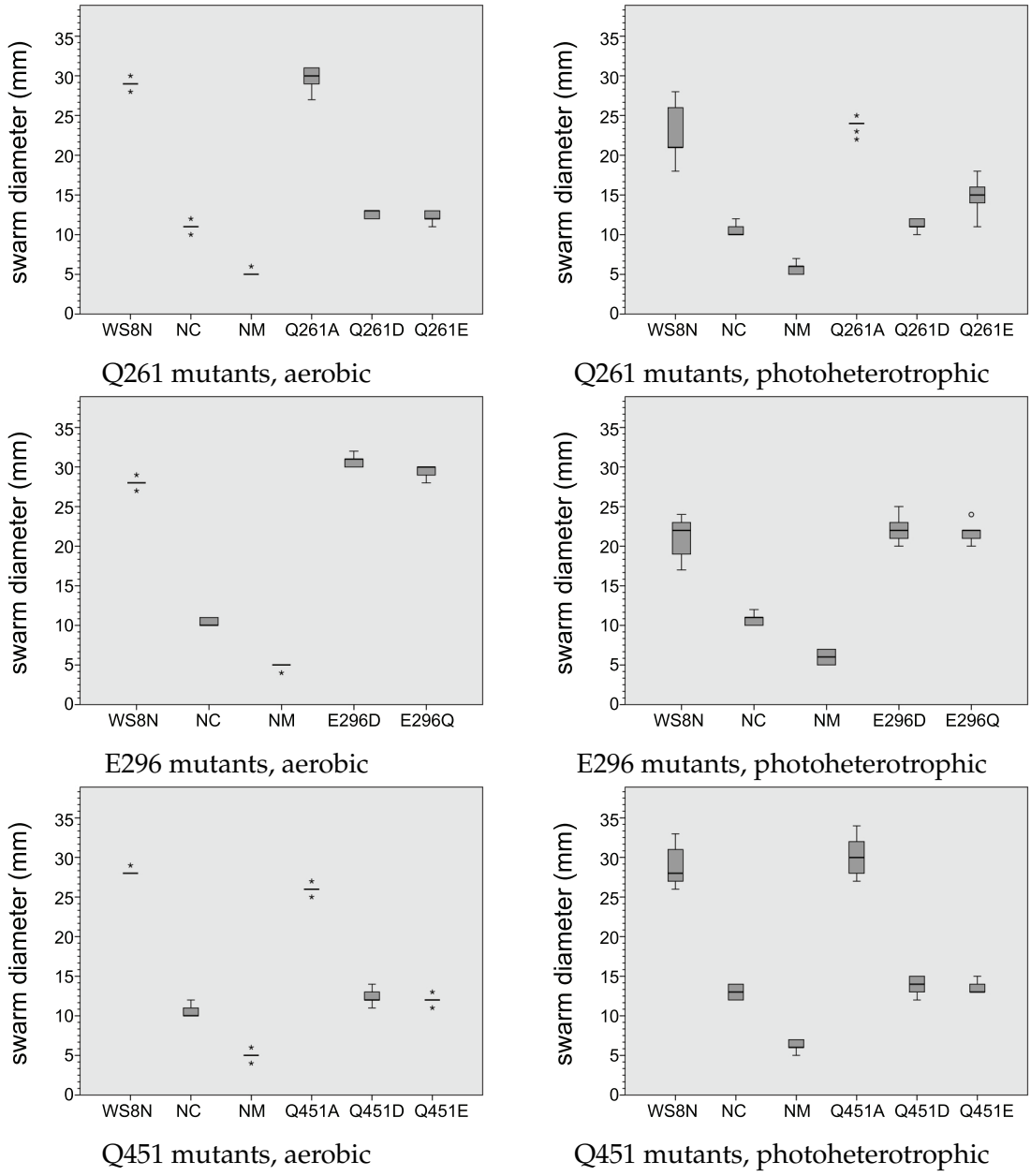


Figure 5.5: Comparison of swim diameters of the TlpT mutants strains under aerobic and photoheterotrophic growing conditions. The boxplot edges indicate the first and third percentiles, the box centre indicates the median, the whiskers indicate the range $\pm$ 1.5x the box height, and the circles indicate outliers (Section 2.6.1). NM = nonmotile, NC = non-chemotactic.

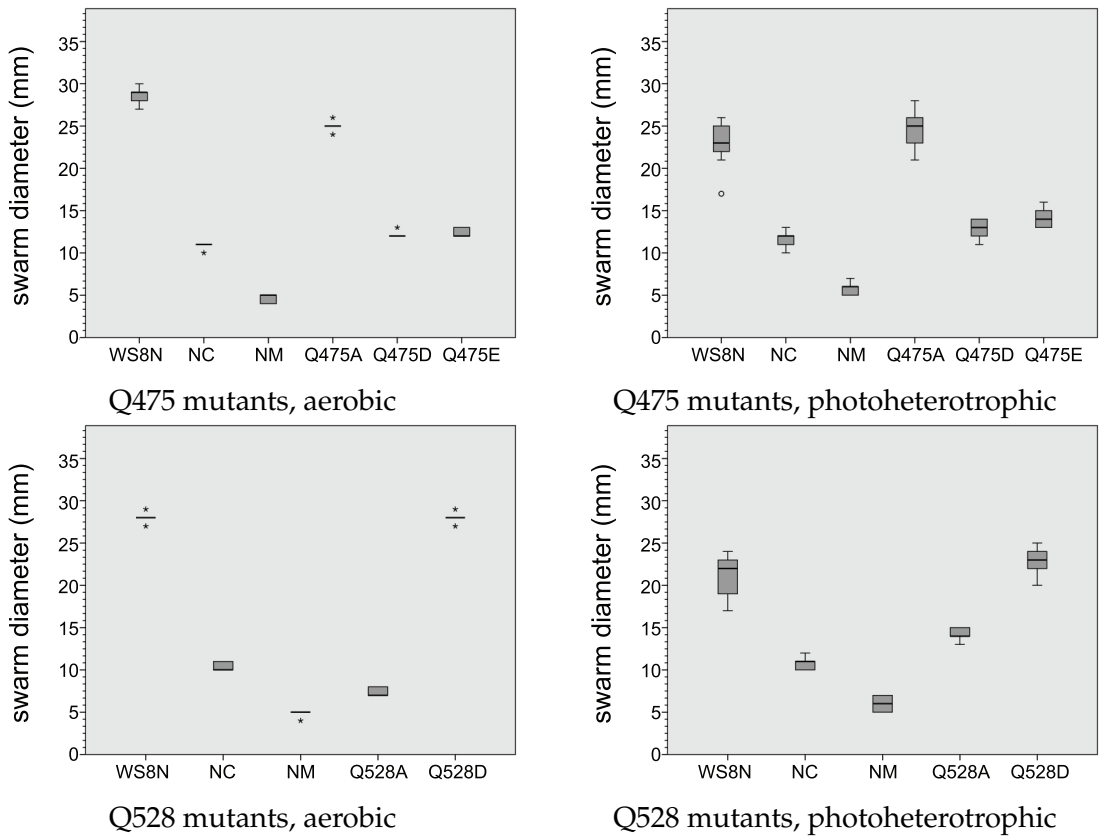


Figure 5.6: Comparison of swim diameters of the TlpT mutants strains under aerobic and photoheterotrophic growing conditions. The boxplot edges indicate the first and third percentiles, the box centre indicates the median, the whiskers indicate the range $\pm$ 1.5x the box height, and the circles indicate outliers (Section 2.6.1). NM = nonmotile, NC = non-chemotactic.

Table 5.3: TlpT mutants in statistically significant groups, on the basis of pairwise comparisons with Dunn tests on the proportion of time spent stopped during free swimming. Groups are split if $p<0.05$ for one pair. Combined groups merge groups that have overlapping members.

	Group 1 Combined 1	Group 2 Combined 2	Group 3 Combined 3	Group 4 Combined 4	Group 5 Combined 5	Group 6 Combined 6	Group 7 Combined 6	Group 8 Combined 6	
MOST STOPPY	TlpT(Q528A) TlpT(Q528A)-1 h TlpT(Q528A)-2 h	TlpT(Q475A)	TlpT(Q261A) TlpT(E296Q)	TlpT(E296D)	Wildtype TlpT(Q528D)	TlpT(Q451A) TlpT(Q261E) TlpT(Q451D) TlpT(Q475D) TlpT(Q451E)	TlpT(Q261E) TlpT(Q451D) TlpT(Q475D) TlpT(Q451E) TlpT(Q261D)	TlpT(Q475D) TlpT(Q451E) TlpT(Q261D) TlpT(Q475E)	MOST SMOOTH

Table 5.4: TlpT mutants grouped by the diameter of swim plate colonies, calculated by a one-way ANOVA and a post-hoc Tukey test.

Mutation	Aerobic phenotype	Photoheterotrophic phenotype
Q261A (JPA 2331)	wildtype	wildtype
Q261D (JPA 2333)	greater than non-chemotactic (with Q261E)	non-chemotactic
Q261E (JPA 2330)	greater than non-chemotactic (with Q261D)	greater than non-chemotactic
E296D (JPA 2334)	greater than wildtype and E296Q	wildtype
E296Q (JPA 2335)	greater than wildtype	wildtype
Q451A (JPA 1037)	less than wildtype	wildtype
Q451D (JPA 2337)	greater than non-chemotactic (with Q451E)	non-chemotactic
Q451E (JPA 2338)	greater than non-chemotactic (with Q451D)	non-chemotactic
Q475A (JPA 1038)	less than wildtype	wildtype
Q475D (JPA 2340)	greater than non-chemotactic (with Q475E)	non-chemotactic
Q475E (JPA 2341)	greater than non-chemotactic (with Q475D)	greater than non-chemotactic
Q528A (JPA 1039)	less than non-chemotactic	greater than non-chemotactic
Q528D (JPA 2343)	wildtype	wildtype

TlpT(Q528A) mutant strain, which agrees with the steady state swimming differences between the two conditions for this strain.

5.4.3 Tethered cell analysis of individual cells

As described in Chapter 3, tethering was used to record the response of individual cells to a step up, followed by a step down, of propionate concentration.

Each strain was analysed on three separate occasions, using fresh culture. On each occasion, at least three slides of data were analysed, each slide made up directly before analysis and subjected to only one challenge to prevent sequential propionate challenges from affecting response times. Each strain was therefore analysed nine times, at the two growth conditions.

Cells were classified using the same phenotypes as in Section 3.3.1 and the resulting frequency table is given in Table 5.5.

Unlike the chemotaxis mutants analysed in Chapter 3, none of the TlpT mutant strains showed a significant number of stoppy cells. Again, there were only isolated cases of stopping on propionate addition. As rare events across all strains were shown not to have an effect on clustering, these categories were excluded by

Table 5.5: Tethering phenotype frequencies for TlpT mutants grown under aerobic and photoheterotrophic conditions, tethered and challenged with an increase and decrease in propionate in the environment.

Growth condition	strain	rotate from start				rotate on propionate				stoppy		total cells
		unresponsive	responsive	adaptive	stop on propionate	unresponsive	responsive	adaptive	stop early	unresponsive	adaptive	
Aerobic												
	WS8N	7	1	43	0	2	2	6	2	0	0	63
	Q261A	6	1	14	0	1	2	2	2	0	0	28
	Q261D	46	2	3	1	0	0	0	0	0	0	51
	Q261E	51	0	0	0	0	0	0	0	2	0	53
	E296D	3	1	29	1	0	0	1	2	0	1	37
	E296Q	6	0	19	1	0	0	0	1	0	4	30
	Q451A	5	1	19	0	0	0	0	0	0	0	25
	Q451D	47	0	0	0	0	0	0	0	0	0	47
	Q451E	50	1	1	1	0	0	0	0	0	0	52
	Q475A	3	0	1	0	0	10	3	12	0	0	29
	Q475D	44	1	0	0	0	0	0	0	1	0	46
	Q475E	32	0	0	0	0	0	0	0	2	0	34
	Q528A	8	3	1	0	0	0	0	0	0	0	12
	Q528D	0	3	25	0	0	1	3	2	0	1	35
Photoheterotrophic												
	WS8N	42	5	47	0	0	0	0	0	0	0	94
	Q261A	14	6	100	0	0	1	0	1	0	0	122
	Q261D	94	0	1	2	3	0	0	0	0	0	98
	Q261E	116	5	26	1	0	0	0	0	0	0	147
	E296D	8	3	96	1	0	0	1	0	0	0	108
	E296Q	5	10	163	0	0	1	2	0	0	0	181
	Q451A	28	4	122	0	0	0	1	0	0	1	156
	Q451D	109	0	4	0	0	0	0	0	0	0	113
	Q451E	106	2	0	0	1	0	1	0	2	0	112
	Q475A	7	4	135	0	0	0	0	0	0	5	151
	Q475D	120	3	6	0	0	0	0	0	0	0	129
	Q475E	64	1	12	0	0	1	0	0	0	0	78
	Q528A	43	13	33	0	0	19	5	13	1	0	127
	Q528D	22	7	128	0	0	0	2	0	0	0	159

removing the cells which stop on propionate addition and binning stoppy cells with normal swimming unresponsive, responsive and adaptive cells.

Hierarchical clustering

Hierarchical clustering based on a complete linkage chi-square test was used to direct grouping of strains based on tethering phenotype frequency data, as described in Chapter 3. The aerobic and photoheterotrophic data sets were first treated separately from one another, then combined.

Hierarchical clustering of the aerobic data set suggests grouping as per the dendrogram in Figure 5.7. Solutions before large jumps in agglomeration coefficients were taken as the starting point for testing for within group heterogeneity with Pearson's chi-square test with Monte Carlo randomisation (100,000 cases). The three-cluster solution precedes a 0.307 jump and was used as the starting point. The first cluster in this solution failed the test for heterogeneity (p=0.003). Splitting this group according the four-cluster solution resulted in two groups that pass the test for heterogeneity, group one (p=0.236) and three (p=0.283) on Figure 5.7, and two singleton groups.

Grouping follows the broad phenotyping of the frequency data. Group one contains mainly unresponsive cells and group three mainly adaptive cells. Both groups two and four show mixed phenotypes, but group four can be considered mainly inhibited, as it contains inhibited and responsive cells.

Similarly, hierarchical clustering applied to the photoheterotrophic data set suggested the three-cluster solution (preceding a 0.264 jump) as the most likely starting point for significant grouping (Figure 5.8). Sequential testing and splitting of each group indicated that the seven-cluster solution gives groups with no significant within group variance using Pearson's chi-square test with Monte Carlo randomisation ($p>0.05$).

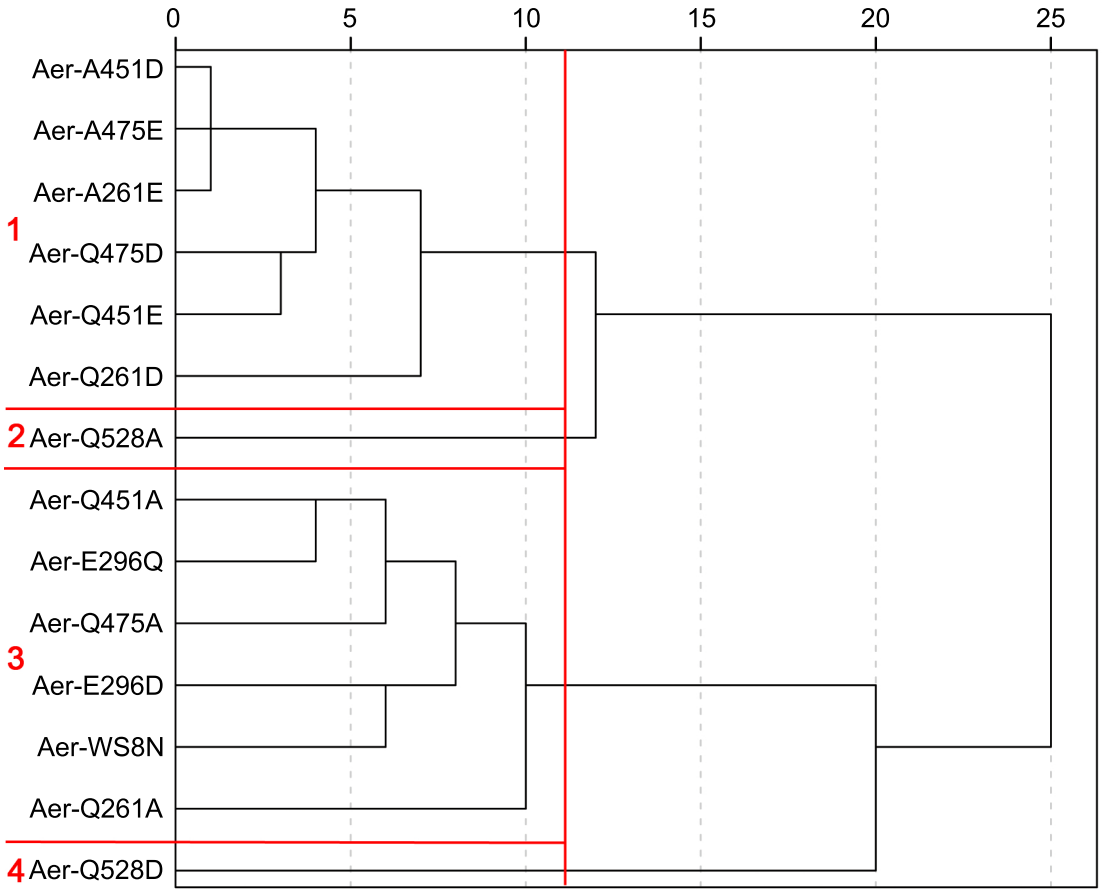


Figure 5.7: Hierarchical clustering of aerobically grown TlpT mutants based on a complete linkage, chi-square test of tethering phenotype frequencies. Red lines and numbering indicate the final accepted clustering solution.

Comparing groups to frequency data, again grouping follows broad phenotyping. Group one (p=0.425) and two (p=0.312) contain strains with mainly adaptive cells, with approximately 80% and 90% adaptive cells respectively. Groups three (p=0.543), four (p=0.298) and five (p=0.132) are contain strains with mainly unresponsive cells, with 15% , 5% and 1% adaptive cells respectively. Groups six and seven are singleton groups. Both contain one strain with a mix of adaptive and unresponsive cells, but group six has more adaptive cells and group seven more unresponsive cells.

Using hierarchical clustering to direct grouping of the complete data set followed by sequential testing and splitting of clusters indicated that an eight-cluster solution best describes the data set (Figure 5.9). The majority of the groupings found in the separate sets are intact here.

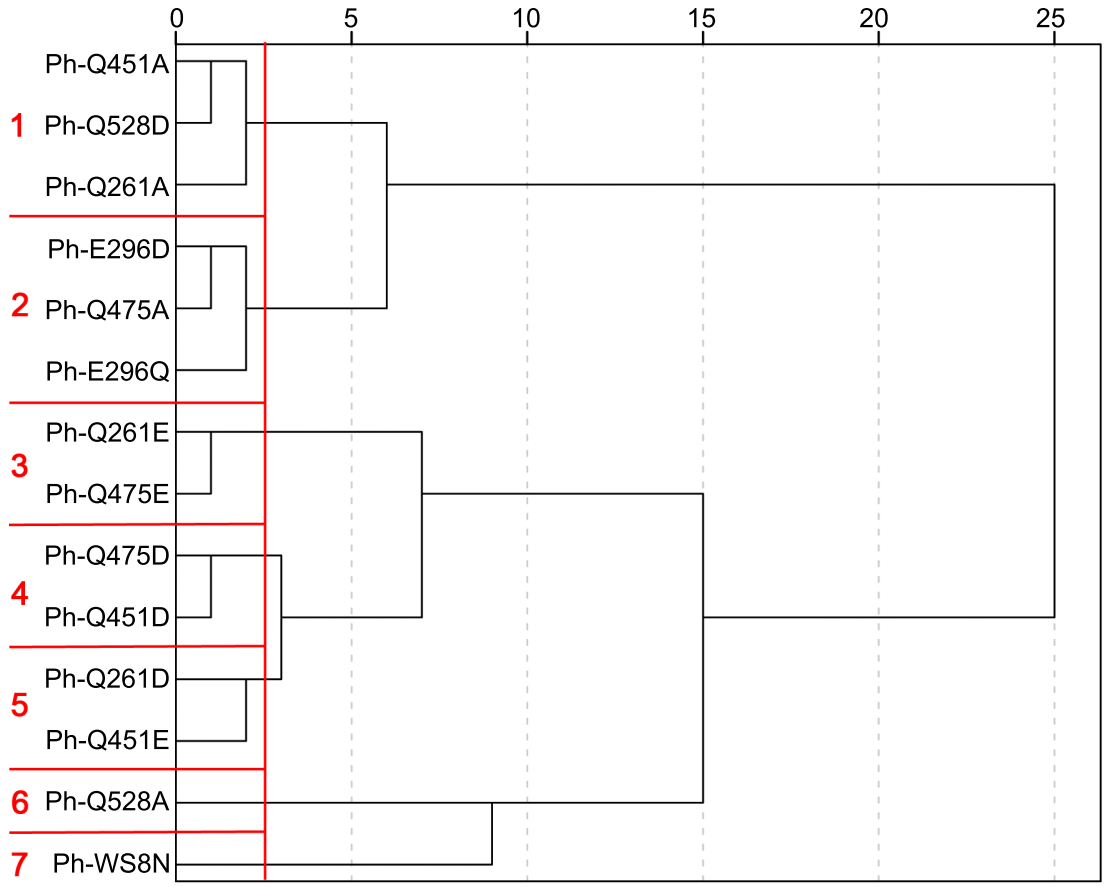


Figure 5.8: Hierarchical clustering of photoheterotrophically grown TlpT mutants based on a complete linkage, chi-square test of tethering phenotype frequencies. Red lines and numbering indicate the final accepted clustering solution.

Group one (p=0.130) and group 2 (p=0.543) contain strains with mainly unresponsive cells. Group one here combines the aerobic unresponsive group (group one), whose strains have 0-6% adaptive cells, and the two photoheterotrophic unresponsive groups, whose strains have insignificant proportions of adaptive cells (groups three and four), forming an unresponsive, <10% adaptive group . Group two here is identical to the photoheterotrophic unresponsive, 15-20% adaptive group.

Groups five (p=0.270), six (p=0.687) and seven (p=0.251) contain strains with mainly adaptive cells. Groups five and six contain strains with 90% and 80% adaptive cells respectively, are both without inhibited cells, and correspond to the photoheterotrophic adaptive groups. Group seven corresponds to the aerobic adaptive group and its strains contain 60-80% adaptive and 3-7% inhibited cells.

The strains in groups three (p=0.077) and eight (singleton) show mixed phenotypes. Group three (combining aerobic group two and photoheterotrophic group seven) strains have more unresponsive cells than adaptive. Group eight (photoheterotrophic group seven) strains have with more adaptive cells than unresponsive. The strains in both of these groups have small proportions of or no inhibited cells.

Group four (aerobic group four) is the inhibited group and its strains have mainly inhibited and responsive cells.

The phenotype groupings for each TlpT mutant under the two growth conditions are shown in Table 5.6. The majority of strains have similar phenotypes under the two conditions. The inhibited phenotype is not seen under photoheterotrophic conditions. Strains that under aerobic conditions are majority adaptive with some inhibited cells have a higher percentage of adaptive cells in photoheterotrophic conditions. The exception to this is wildtype, which has almost equal unresponsive and adaptive cells under photoheterotrophic conditions compared to a majority of adaptive cells under aerobic conditions. Two majority unresponsive strains also see an increase in adaptive, although they remain majority unresponsive.

The most dramatic change is in strain TlpT(Q475A), which converts from an inhibited phenotype to a 90% adaptive with no inhibited cells.

Adaptation times

The adaptation time for each adaptive cell was measured. As described in Chapter 3, the initial stop is easily defined and the point where rotation is considered to have been recovered is once a rotation signal is recorded as sustained for 5 s. Only strains with more than 10% adaptive cells are included in the ANOVA test, to guard against bias from very low cell numbers and to allow for post-hoc tests.

For the aerobic conditions, adaptation times for the six adaptive strains and the

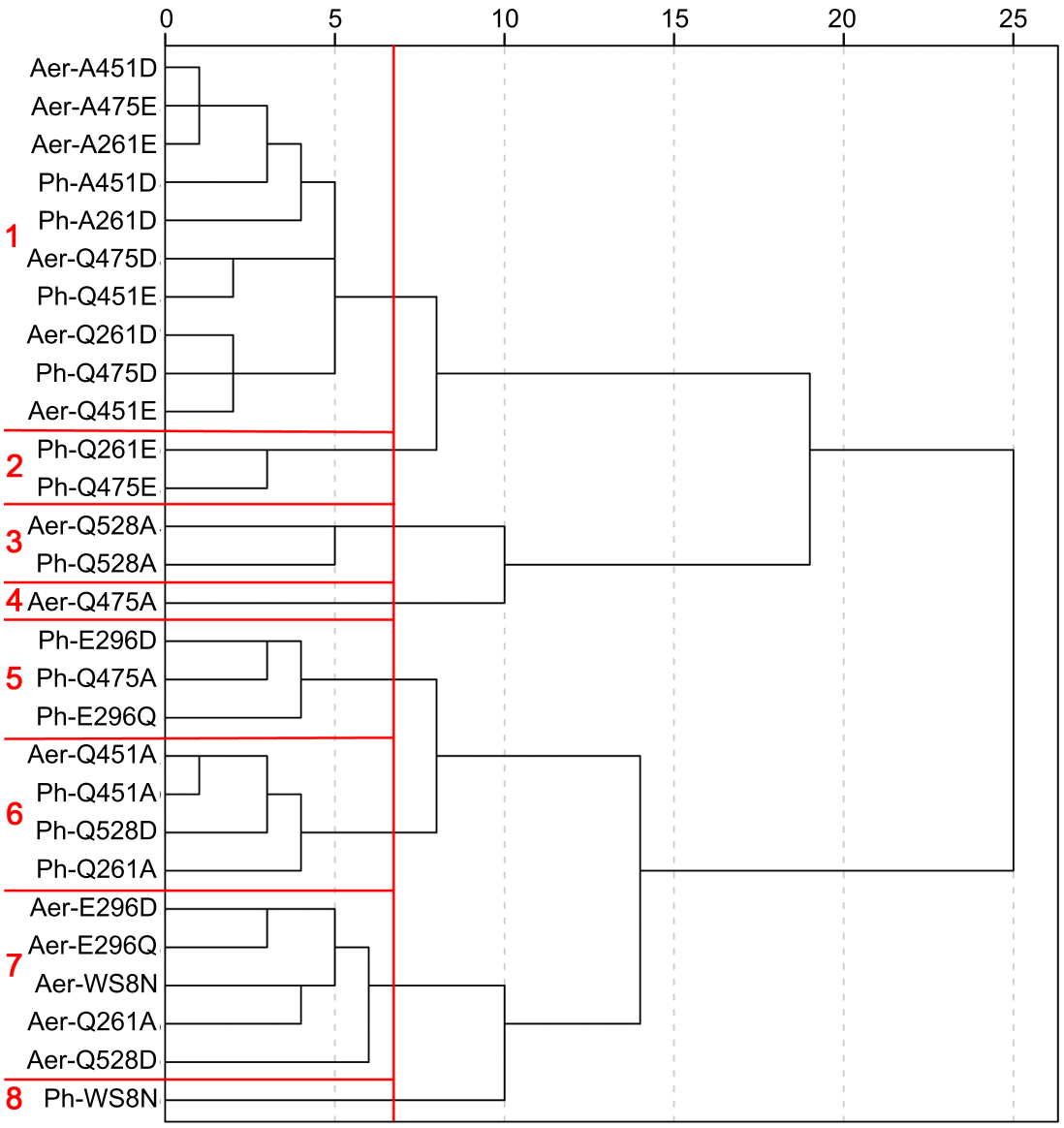


Figure 5.9: Hierarchical clustering of TlpT mutants based on a complete linkage, chi-square test of tethering phenotype frequencies. Both aerobic and photoheterotrophic data is included. Red lines and numbering indicate the final accepted clustering solution.

Table 5.6: Classification of TlpT mutants based on hierarchical clustering of frequencies of phenotypes when tethered and challenged with changes in attractant concentration.

	Aerobic	Photoheterotrophic
Wildtype	60-80% adaptive with inhibited	Mixed, adaptive>unresponsive
TlpT(Q261A)	60-80% adaptive with inhibited	80% adaptive, no inhibited
TlpT(Q261D)	Majority unresponsive, <10% adaptive	Majority unresponsive, <10% adaptive
TlpT(Q261E)	Majority unresponsive, <10% adaptive	Majority unresponsive, 15-20% adaptive
TlpT(E296D)	60-80% adaptive with inhibited	90% adaptive, no inhibited
TlpT(E296Q)	60-80% adaptive with inhibited	90% adaptive, no inhibited
TlpT(Q451A)	80% adaptive, no inhibited	80% adaptive, no inhibited
TlpT(Q451D)	Majority unresponsive, <10% adaptive	Majority unresponsive, <10% adaptive
TlpT(Q451E)	Majority unresponsive, <10% adaptive	Majority unresponsive, <10% adaptive
TlpT(Q475A)	Inhibited	90% adaptive, no inhibited
TlpT(Q475D)	Majority unresponsive, <10% adaptive	Majority unresponsive, <10% adaptive
TlpT(Q475E)	Majority unresponsive, <10% adaptive	Majority unresponsive, 15-20% adaptive
TlpT(Q528A)	Mixed, unresponsive >adaptive	Mixed, unresponsive >adaptive
TlpT(Q528D)	60-80% adaptive with inhibited	80% adaptive, no inhibited

inhibited strain (Q475A) were taken. Boxplots illustrate the population spread for each strain in Figure 5.10. A one-way unbalanced ANOVA test was performed on this data set, confirming that there is a significant difference between the adaptation time for TlpT(Q475A) and the other strains, and that the rest of the strains cluster together.

For the photoheterotrophic conditions, adaptation times for all strains except the two unresponsive groups with very low numbers of adaptive cells. Boxplots indicate the population spread for each strain in Figure 5.11. Once again, a one-way unbalanced ANOVA test separates only one mutant, in this case the mixed phenotype TlpT(Q528A), from the rest. The rest of the data forms four overlapping groups. As before, as groups sizes are different and the ANOVA is thus likely to give Type 1 errors (falsely rejecting the null hypothesis), overlapping groups should be treated as one group.

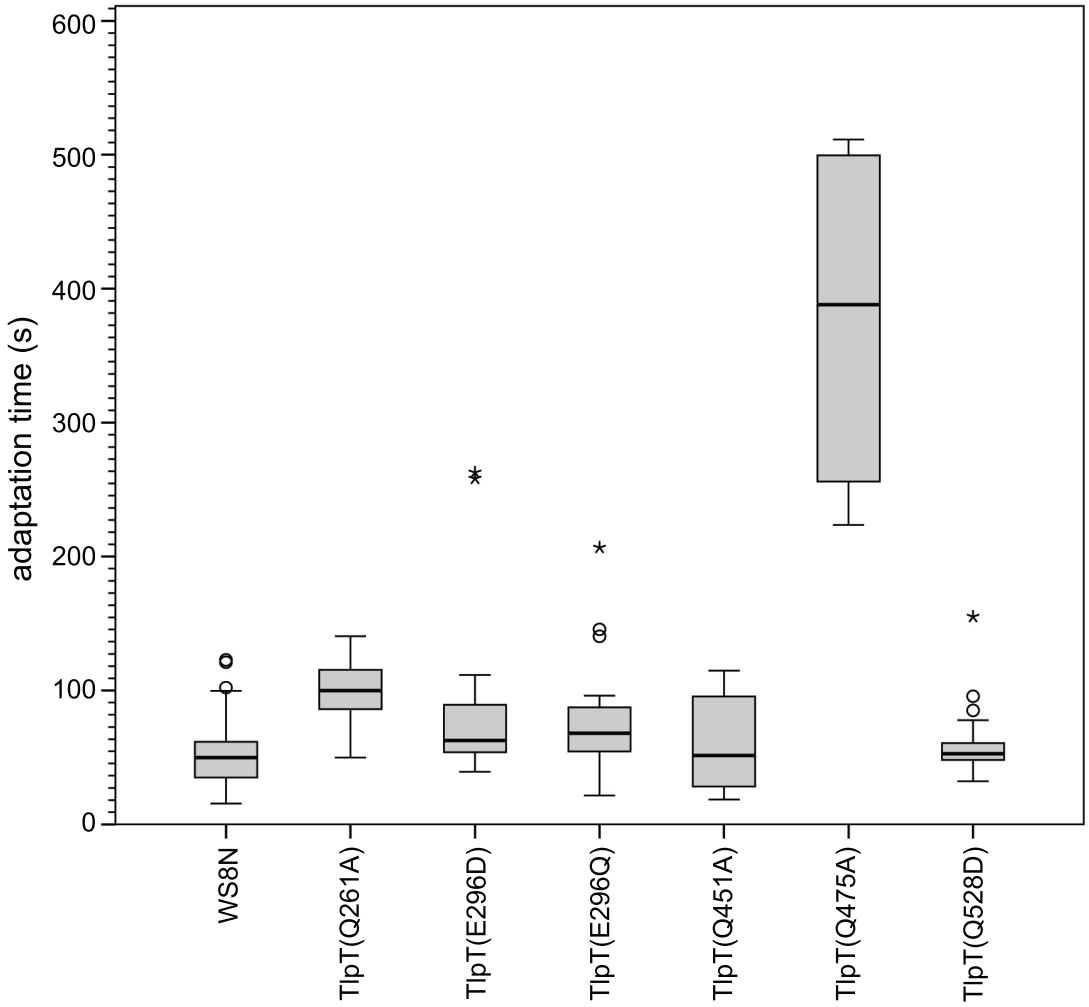


Figure 5.10: Adaptation time in aerobically grown tethered TlpT mutants challenged with a 100 μM drop in propionate concentration. The boxplot edges indicate the first and third percentiles, the box centre indicates the median, the whiskers indicate the range $\pm$ 1.5x the box height, and the circles indicate outliers (Section 2.6.1)

5.5 Discussion

Five potential methylation sites on cytoplasmic chemoreceptor TlpT were mutated to **A**, **D** and **E** on the genome. Thirteen of these mutants were successfully made. The swimming behaviour of the mutants under steady state and dynamic conditions, under aerobic and photoheterotrophic growth conditions, was analysed using swim plates, a free swimming assay and a tethering assay. The phenotypes recorded for each strain for each of these tests are included in Table 5.7.

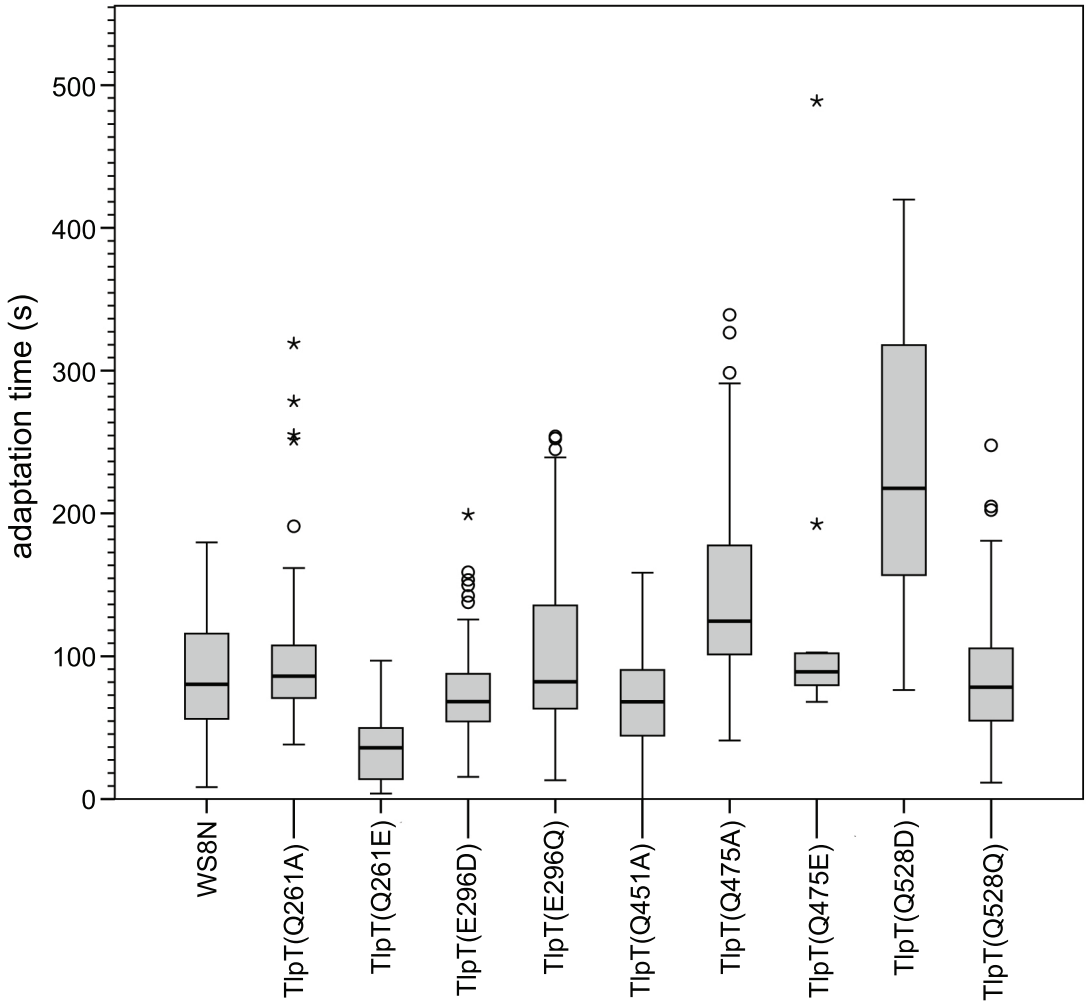


Figure 5.11: Adaptation time in photoheterotrophically grown tethered TlpT mutants challenged with a 100 μM drop in propionate concentration. The boxplot edges indicate the first and third percentiles, the box centre indicates the median, the whiskers indicate the range $\pm$ 1.5x the box height, and the circles indicate outliers (Section 2.6.1)

Table 5.7: All phenotype results for the 13 TlpT mutants. WT = wildtype, NC = non-chemotactic.

	Swarm plate		Free swimming	Tethering	
	Aerobic	Anaerobic		Aerobic	Anaerobic
WS8N	WT	WT	WT	60-80% adaptive with inhibited	Mixed, adaptive>unresponsive
TlpT(Q261A)	WT	WT	stoppier	60-80% adaptive with inhibited	80% adaptive, no inhibited
TlpT(Q261D)	>NC	NC	smooth	Majority unresponsive, <10% adaptive	Majority unresponsive, <10% adaptive
TlpT(Q261E)	>NC	>NC	smooth	Majority unresponsive, <10% adaptive	Majority unresponsive, 15-20% adaptive
TlpT(E296D)	>WT	WT	stoppier	60-80% adaptive with inhibited	90% adaptive, no inhibited
TlpT(E296Q)	>WT	WT	stoppier	60-80% adaptive with inhibited	90% adaptive, no inhibited
TlpT(Q451A)	<WT	WT	smooth	80% adaptive, no inhibited	80% adaptive, no inhibited
TlpT(Q451D)	>NC	NC	smooth	Majority unresponsive, <10% adaptive	Majority unresponsive, <10% adaptive
TlpT(Q451E)	>NC	NC	smooth	Majority unresponsive, <10% adaptive	Majority unresponsive, <10% adaptive
TlpT(Q475A)	<WT	WT	very stoppy	Inhibited	90% adaptive, no inhibited
TlpT(Q475D)	>NC	NC	smooth	Majority unresponsive, <10% adaptive	Majority unresponsive, <10% adaptive
TlpT(Q475E)	>NC	>NC	smooth	Majority unresponsive, <10% adaptive	Majority unresponsive, 15-20% adaptive
TlpT(Q528A)	<NC	>NC	very stoppy	Mixed, unresponsive >adaptive	Mixed, unresponsive >adaptive
TlpT(Q528D)	WT	WT	WT	60-80% adaptive with inhibited	80% adaptive, no inhibited

5.5.1 Wildtype behaviour

WS8N has a different response pattern to dynamic changes under aerobic and photoheterotrophic conditions. In both growth conditions, wildtype does not mean 100% adaptive cells. Instead, the basal levels of $CheY_6$-P and, presumably, the relative levels of sensitivity of chemoreceptors in the polar cluster, are set so there are subpopulations of cells.

Under aerobic conditions, 80% of cells adapt as expected. A very small percentage of the cells are inhibited or responsive, less than 10% combined. These cells likely have a higher basal level of $CheY_6$-P and are primed to respond to increases in chemoeffector, not decreases. The small energy output to keep the cell in the same area, burning ATP to keep $CheY_6$ phosphorylated, is worth the gain of staying in that area. Less than 15% of cells are unresponsive. These cells likely have a lower or zero basal level of $CheY_6$-P and are primed not to respond to drops in chemoeffector. The energy output to prime the cell for active searching of the environment may not be worth the potential gains.

It is unlikely that the environment in which the TlpT mutant strains were grown up in is actually best responded to by switching to the inhibited or unresponsive state. These two small outlier populations are less well adapted to the environment they are currently in, but are more likely to survive if the environment undergoes a large change. By maintaining such a degree of stochasticity in the system that some cells are forced to the outlier conditions, the population may maintain variability for future catastrophes.

Under photoheterotrophic conditions, cells are generally less responsive than under aerobic conditions. This likely reflects the increased metabolic variability and higher expression levels of the chemotaxis proteins under aerobic conditions. Under photoheterotrophic conditions, the population is primed either to move to dark conditions and search for organic materials to ferment or respire (adaptive) or to

move into an anaerobic environment without any organic materials and rely completely on photosynthesis (unresponsive).

5.5.2 Potential methylation sites

If these sites are methylation sites and TlpT functions as predicted in Chapter 3, mutation of a **Q** site to **E** will not have any effect, as this post-translational modification should be present in every receptor in wildtype. The permanent methylation of a site through an **A** mutation should increase the proportion of $CheY_6$-P in the cell, as the cluster is more sensitive and more likely to signal, resulting in a stoppier swim. The permanent demethylation of a site through a **D** mutation could decrease the proportion of $CheY_6$-P in the cell, resulting in a smoother swim. Disrupting one site was not expected to disrupt chemotaxis altogether, based on previous work (Nowlin *et al.*, 1988).

If a site is instead involved in TlpT packing but not methylation, mutations could have a number of effects. Mutations to **A** are likely to be silent, although depending on the environment they are in, they may promote packing. **DE** mutants introduce a negative charge, so are likely to destabilise packing.

Site Q261

Site Q261 is unlikely to be methylation site. Mutation of the **Q** to a **D** or **E** leads to a smooth swimming, unresponsive motility and chemotaxis phenotype, which is the same as that obtained when signalling by the cytoplasmic cluster is lost under aerobic conditions. In the anaerobic state, only 20% of TlpT(Q261E) mutant cells are able to sense and respond to dynamic change. This strongly suggests that if this site holds a negative charge, cluster packing is sufficiently disrupted to completely halt signalling.

TlpT(Q261A) is not a neutral mutation of TlpT, as might be expected from the above, but has a similar profile to an **A** mutated methylation site or a mutation that increases packing and sensitivity. Increased packing leads to stoppier swimming under steady state. Under both growth conditions, the dynamic response profile is apparently skewed towards a greater basal rate of $CheY_6$-P, increasing the proportion of inhibited and responsive cells under aerobic conditions and increasing the proportion of adaptive cells under photoheterotrophic conditions.

Site E296

Although the TlpT(E296D) and TlpT(E296Q) mutant strains both have stoppier swimming at steady state than wildtype, the aerobic response profile is identical to wildtype. The photoheterotrophic response suggests increased signalling through the cluster, as a greater proportion of cells are adaptive. This is the opposite response to that which is expected of a methylation site. Unfortunately, this cannot be compared to a TlpT(E296A) mutant to see if it too has the opposite response.

TlpT(E296Q) is not a silent mutation. If E296 is a methylation site, then deamidation probably does not easily occur on sites that are not normally translated as **Q**.

Site Q451

Site Q451 is highly unlikely to be a methylation site. The **A** mutation is a mostly silent mutation, giving rise to wildtype behaviour, except for a skew towards smoother swimming at steady state. This suggests some disruption to signalling through the cluster (a drop in $CheY_6$-P), but sufficient signalling to allow a stop.

The **D** and **E** mutations both completely disrupt signalling, giving smooth swimming at steady state and an inability to respond to dynamic changes. Very little, if any, $CheY_6$-P is likely produced.

Site Q475

Site Q475 is unlikely to be a methylation site, as both **D** and **E** mutants result in decreased signalling through the cluster. Steady state swimming is smoother and the ability to respond to dynamic changes is lost under aerobic conditions. Under photoheterotrophic conditions there is some signalling through the cluster, as a small percentage of cells can adapt. Thus, these mutations do not completely disrupt the cluster.

However, the **A** mutation gives a different result from what would be expected from a purely structural site. Under aerobic conditions, the cluster appears to be activated, producing additional $CheY_6$-P. Swimming is significantly stoppier than wildtype and the response to dynamic change is similar to that of the $CheB_2$ deletion strain, showing an inhibited profile. This is more dramatic than was expected for a single methylation site.

Under photoheterotrophic conditions, the inhibition lifts completely and the majority of cells are adaptive. This is still a skew away from wildtype, with an apparent higher basal $CheY_6$-P level.

This site thus interferes in some way with cytoplasmic cluster activation and deactivation, promoting sensitivity of the chemoreceptors but not forcing them to be permanently active, as evidenced by the difference in response in the two growth states. It is possible that this site is part of a CheR binding or active site and mutating it promotes or inhibits CheR activity.

Site Q528

Site Q528 was not expected to have much effect on TlpT signalling, as it falls within the flexible tail region of the protein and should not be involved in packing. The phenotype seen for the TlpT(Q528D) mutant agrees with this. The **Q/D** mutation

introduces a negative charge, so if the site is involved in packing it is likely to show some disruption. If the region is part of a flexible linker, **Q** and **D** are similar hydrophilic polar molecules and there is likely to be no effect. This is the case, as the TlpT(Q528D) mutant has wildtype phenotypes at all conditions, barring a slight increase in cluster activity during dynamic behaviour in photoheterotrophic conditions.

However, The TlpT(Q528A) mutant shows a markedly different phenotype. Under aerobic conditions, $CheY_6$-P production is so high that swim diameters are only just larger than nonmotile cells. As with the $CheB_2$ deletion strain, collecting sufficient cells in tethering is challenging as most cells do not rotate. However, those that do rotate are primarily unresponsive and there are no inhibited cells. From free swimming results, it is possible that the skew of $CheY_6$-P is so high that the majority of cells are completely stopped and the few cells that are found to rotate are from the far end of distribution, the unresponsive cells. When there are few rotating cells, it is much easier to find a single currently rotating cell in a field of view than it is to find an inhibited cell that will start rotating on the addition of chemoeffector.

Corroborating this, under photoheterotrophic conditions the $CheY_6$-P levels across the population again appear higher than wildtype. Tethered cells show the same range as wildtype aerobic cells, with a mix of inhibited, unresponsive, responsive and adaptive. This is the only photoheterotrophic strain to include inhibited cells.

The Q528A mutation promotes sensitivity and activation of the cluster. Like Q475, this site may promote CheR activity in the cluster. It is not a methylation site, as there is no strong phenotype for the TlpT(Q528D) mutant, as for the TlpT(Q528A) mutant.

5.5.3 Validity of the TlpT structure

The phenotype differences for the TlpT(Q528A) and TlpT(Q528D) mutants suggest that TlpT does indeed have a flexible C-terminal region like Tsr. The evidence that Q261 and Q451 are not methylation sites and that introduction of a charge disrupts signalling validates the model, as neither of these sites was accessible for deamidation/methylation on the theoretical structure.

The results for site Q475 neither validate nor invalidate the model. This site is not a methylation site, but may be near methylation sites. The profound changes that the usually silent **A** mutant resulted in leave the potential for this site to be actively involved in some way, so its orientation in the theoretical structure cannot be commented on.

5.5.4 Validity of the model

Point mutations in TlpT that do not affect cluster formation are sufficient to disrupt packing and prevent signalling through the cluster, leading to a complete loss of the ability to stop or to respond to dynamic changes (for example, the TlpT(Q451D) mutants). This again strongly suggests that the polar cluster cannot produce sufficient $CheY_6$-P alone to cause a stop. The cytoplasmic cluster is either required to supply a baseline concentration of $CheY_6$-P or to effect a stop itself.

Unlike the chemotaxis mutants in Chapter 3, which are primarily unresponsive or primarily adaptive, the TlpT mutants have a range of tethering phenotypes. The TlpT mutants' adaptation times also have a greater range than with the chemotaxis mutants, although in both sets of mutants the adaptation times form one statistical population. That range covers adaptation times from that of the wildtype strain to less than that of the $CheB_2$ deletion mutant. The spread is much greater under photoheterotrophic conditions than aerobic conditions.

The differences seen between the two conditions does strongly suggest that the cytoplasmic cluster is indeed sensing the internal energy state or redox potential of the cell. Whether the cytoplasmic cluster responds merely by setting the basal $CheY_6$-P level in the cell or also is involved in the dynamic response via the polar cluster cannot be teased apart using these results.

TlpT's role in setting the basal rate of $CheY_6$-P is confirmed by certain mutations causing much stoppier swimming than wildtype at steady state (Q528A and Q475A).

5.5.5 Future work

Targeted mutations were made based on bioinformatics analysis of the *R. sphaeroides* chemoreceptors. Known consensus sequences for methylation sites did not match many sites with the *R. sphaeroides* cytoplasmic cluster, requiring the development of prediction sequences.

None of the sites mutated can conclusively be identified as methylation sites. If a number were possible methylation sites, mutating all of the sites in one receptor would increase the chance of a positive phenotype. However, the majority of the sites can be dismissed as potential methylation sites from these results (Q261, Q451, Q475, Q528), and the remaining site is unconvincing (E296). The results from site Q475 do suggest that there may be unpredicted methylation sites in that region. $CheR_3$ binding and activity is enhanced on Q528 mutation, which again implies that methylation occurs somewhere on TlpT or on one of the other cytoplasmic chemoreceptors.

The prediction sequences used generated too many false positives to be useful. The published consensus sequences did not match many regions in *R. sphaeroides* chemoreceptors, suggesting that *R. sphaeroides* methylation sites sit in a different environment to other chemoreceptors. Rather than extending this method to the next chemoreceptor and testing many false positives, positive identification of methy-

lation sites in TlpT by some other method is needed. This can be used to direct sequential mutation of multiple sites in TlpT and compare results of individual mutations.

Chapter 6

In vitro testing of putative adaptation components

Chemoreceptors in different methylation states travel in distinct bands in gel electrophoresis. In this chapter, this chemoreceptor property is used to test the methylesterase and methyltransferase activity of the *R. sphaeroides* CheBs and CheRs towards the *E. coli* chemoreceptor Tsr. $CheB_1$ was able to modify Tsr, whereas $CheB_2$ had no effect. No conclusive results for the CheRs were obtained. The methylation state of likely methylation candidate McpJ was unchanged by all adaptation proteins tested.

6.1 Introduction

Methylation-based adaptation for chemotaxis is predicted to occur at the *R. sphaeroides* polar cluster, whereas the cytoplasmic cluster is predicted to use the adaptation proteins to control the steady state stopping frequency (Chapter 3). Only one cytoplasmic chemoreceptor matched sufficient criteria to be considered an initial candidate for methylation (Chapter 4), but site-directed mutagenesis of the predicted sites did not give clear results (Chapter 5).

A number of polar chemoreceptors are likely methylation candidates. Three, McpH,

McpJ and McpR, have deletion phenotypes that make them highly likely to be involved in adaptation. If site-directed mutagenesis of the only chemoreceptor thought to be methylated in the cytoplasmic cluster does not give clear phenotypes, then site-directed mutagenesis of one of several adaptable chemoreceptors in the polar cluster is less likely to show clear phenotype changes with the loss of one site on one involved chemoreceptor. The ability of the other involved chemoreceptors to be methylated and demethylated, and thus to change conformation and drive alterations in the packing of the entire cluster, could be sufficient to dampen the effect of losing only one site on one chemoreceptor, especially if that chemoreceptor occurs at low copy numbers.

Using a background strain with only one polar chemoreceptor would remove this issue. However, attempts to sequentially delete the chemoreceptors have resulted in the loss of chemotaxis (Thompson, 2005). Rather than repeating this attempt to make a clean background strain for site-directed mutagenesis, an alternative method for initially screening chemoreceptors is used.

6.1.1 *In vitro* chemoreceptor modification

In gel electrophoresis, different methylation states of the same chemoreceptor travel in distinct separate bands (Boyd and Simon, 1980; Chelsky and Dahlquist, 1981; Engstrom and Hazelbauer, 1980). Why this is the case is not well understood: the single charge difference that methylation gives to a chemoreceptor should not have an effect once the receptor is denatured and bound to negatively-charged SDS (Section 2.3.5).

Although chemotaxis is achieved by a complex network of proteins within the cell, many of the individual reactions can occur *in vitro* in the absence of the majority of the components. The adaptation proteins are able to deamidate, demethylate and methylate receptors *in vitro* (e.g., Anand *et al.* (1998); Anand and Stock (2002)).

These reactions have been shown to occur using either full-length chemoreceptors embedded in nanodiscs or using truncated chemoreceptors, including only the soluble cytoplasmic domain (e.g., Mowbray *et al.* (1985); Perez *et al.* (2006)).

6.1.2 The *R. sphaeroides* adaptation proteins

$CheB_1$ is able to complement a CheB deletion in *E. coli* only at very low levels of induction, restoring chemotaxis and lifting the tumbly swimming bias. At high levels of $CheB_1$ chemotaxis is again inhibited, presumably as bias switches from permanent tumbles to permanent runs (Martin *et al.*, 2001b), mimicking the overexpression of *E. coli* CheB. $CheB_2$ is able to partially complement a CheB deletion in *E. coli*, partially restoring chemotaxis and reducing the tumbly bias (Porter *et al.*, 2002).

$CheR_2$ is able to complement a CheR deletion in *E. coli*, restoring chemotaxis and lifting the smooth swimming bias (Martin *et al.*, 2001b). Neither $CheR_1$ (Martin *et al.*, 2001b) nor $CheR_3$ (Porter *et al.*, 2002) is able to complement a CheR deletion in *E. coli*, having no effect on either chemotaxis or the smooth swimming bias.

The cytoplasmic cluster is likely to have a very different kind of structure to a canonical polar cluster. The difference between the *che*Op_2 and *che*Op_3 adaptation proteins' ability to complement *E. coli* deletions might be due to cluster formation. The differences are unlikely to be driven by an inability of the *che*Op_3 adaptation proteins to localise at the *E. coli* chemotaxis cluster: the pentapeptide tether on *che*Op_3's $CheR_3$ (EFRRH) is almost identical to that of *E. coli*'s CheR (NWETF), whereas the proposed tether on *che*Op_2's $CheR_2$ is less similar (GFGDF). If localisation was the problem, then $CheR_2$ would be less able to complement *E. coli* CheR than $CheR_3$.

If *in vitro* the same differences in the *R. sphaeroides* CheRs ability to methylate Tsr are seen, then those differences are more likely to be due to differences in the active site's recognition of methylation sites, rather than differences in polar and cytoplasmic cluster formation.

6.2 Strategy

An *in vitro* assay was devised to test the ability of a specific CheR and CheB to methylate or deamidate a particular chemoreceptor. An adaptation protein and a chemoreceptor fragment are incubated together with any required substrate (SAM or a phosphodonor) for an hour. After incubation, the reaction mix is separated with gel electrophoresis. The assay is based on the assumptions (discussed in detail above) that the adaptation proteins can act upon the cytoplasmic domain of a chemoreceptor and that the resulting different methylation states will be visible in the form of separate bands after gel electrophoresis. If additional bands are seen, then that adaptation protein is able to modify that chemoreceptor.

The method was first validated using *E. coli* proteins. The cytoplasmic domain of Tsr (Tsr_{CD}) was incubated with purified *E. coli* CheR, CheB and a constitutively active CheB fragment ($CheB_c$). $CheB_c$ is the methylesterase domain of CheB. It is less active than activated full-length CheB but significantly more active than unphorphorylated full-length CheB (Simms *et al.*, 1985; Lupas and Stock, 1989). This construct is included to confirm that phosphoramidate is an effective activator of CheB.

The reaction mixes are separated using gel electrophoresis. Band shifts indicate whether the methylation state of the receptor has changed. The pre-translational deamidation **Q** residues are analogous to methylated **E** residues: both are neutrally charged. Tsr_{CD} is purified in the absence of all adaptation proteins, so is not subject to post-translational modifications before incubation. If Tsr is incubated with only CheB, there are three possible Tsr states: QQEE, QEEE or EEEE. If Tsr is incubated with only CheR, there are three possible states: QQEE, $QQEE^{met}$ (one methylation event) or $QQEE^{metmet}$ (two methylation events).

The method was then extended to test the reactivity of the *R. sphaeroides* adaptation proteins towards Tsr_{CD}. The ability of *E. coli* and *R. sphaeroides* adaptation proteins to alter the methylation state of polar chemoreceptor $McpJ_{CD}$ was then tested.

6.3 Method

6.3.1 Constructs and strains used

The *E. coli* and *R. sphaeroides* adaptation proteins were expressed in the *E. coli* protein expression strain M15. Chemoreceptor fragments was expressed in *E. coli* strain RP2867, which is a wildtype RP437 derivative with Δ*cheBcheR*. This ensured that chemoreceptors were produced without post-translational modification or any methylation.

Existing constructs pQE80-B1, pQE80-B2, pQE80-R1, pQE80-R2 pQE80-R3 and pQE80-McpJc, encoding N-terminal His-tagged proteins $CheB_1$, $CheB_2$, $CheR_1$, $CheR_2$, $CheR_3$ and $McpJ_{CD}$ respectively, were used.

Constructs expressing the *E. coli* proteins and the methylesterase domains of the CheB proteins were made. Domain boundaries were taken from the NCBI conserved domain search (Marchler-Bauer *et al.*, 2011). Genomic DNA (gDNA) extracted from wildtype *E. coli* RP437 (Section 2.2.1) was used as template DNA for *E. coli* constructs. Existing constructs were used as template DNA for *R. sphaeroides* $CheB_c$ constructs.

Primer pairs were used to amplify target genes from the templates, introducing restriction sites BamHI and PstI up and downstream of the gene fragments respectively. These restriction enzymes were then used to insert the gene fragments into the expression vector pQE80. The resulting vectors coded for N-terminally His-tagged proteins as listed in Table 6.1.

6.3.2 Protein expression and purification

Each construct was transformed into the appropriate *E. coli* expression strain using the standard protocol (Section 2.2.3). 1 *l* of each culture was grown with selection

Table 6.1: Components for construction of *E. coli* and *R. sphaeroides* protein expression vectors by PCR and ligation.

Protein	Vector	Forward primer	Reverse primer	Target gene	Template
E. coli constructs					
CheB	pQE80-ECCheB	ECCheB-F	ECCheB-R	*cheB*	RP437 gDNA
$CheB_c$	pQE80-ECCheBc	ECCheBc-F	ECCheB-R	*cheB* MED	RP437 gDNA
CheR	pQE80-ECCheR	ECCheR-F	ECCheR-R	*cheR*	RP437 gDNA
Tsr_{CD}	pQE80-TsrCD	TsrCD-F	TsrCD-R	CD of *tsr*	RP437 gDNA
R. sphaeroides constructs					
$CheB1_c$	pQE80-B1c	B1c-F	B1c-R	$cheB_1$ MED	pQE80-B1
$CheB2_c$	pQE80-B2c	B2c-F	B2c-R	$cheB_2$ MED	pQE80-B2

MED=methylesterase domain

antibiotics until mid-log phase ($OD_{600\ nm}$ = 0.5), then induced with 0.5 mM IPTG. Cultures containing adaptation protein expression vectors were induced overnight at 18^{o}C and those containing chemoreceptor fragments were induced for 5 h at 30^{o}C. After induction, cells were collected by centrifugation and frozen. For more details, see Section 2.3.1.

All proteins were purified from frozen cell pellets under native conditions. Adaptation proteins were purified using Ni-NTA chromatography, followed by size exclusion chromatography to remove imidazole from the final protein buffer. Adaptation proteins were then concentrated using a centrifuge concentrator. Chemoreceptor fragments were purified using Ni-NTA chromatography, followed by three rounds of dialysis to remove imidazole from the final protein buffer. Further details are given in Section 2.3.3.

Presence of the target protein in purification fractions and protein purity in the final sample were analysed by protein separation using electrophoresis under denaturing conditions (SDS-PAGE, Section 2.3.5). Protein concentration in each final sample was calculated using the Bradford assay (Section 2.3.5).

All proteins were successfully purified except for the construct $CheB_{2c}$. The final concentrations of all proteins are given in Table 6.2.

Table 6.2: Purified *R. sphaeroides* and *E. coli* chemotaxis proteins.

Protein	Molecular weight (kDA)	Concentration (uM)
E. coli Tsr_{CD}	32.90	192.56
R. sphaeroides $McpJ_{CD}$	31.36	67.01
E. coli CheB	39.01	31.37
E. coli $CheB_c$	22.24	58.54
E. coli CheR	34.39	29.15
R. sphaeroides $CheB_1$	39.60	43.36
R. sphaeroides $CheB_{1c}$	23.24	68.30
R. sphaeroides $CheB_2$	39.63	42.88
R. sphaeroides $CheB_{2c}$	22.76	N/A
R. sphaeroides $CheR_2$	31.32	50.94
R. sphaeroides $CheR_3$	31.69	7.22

6.3.3 *In vitro* modification

Chemoreceptors and adaptation enzymes were incubated together at 30^oC (*R. sphaeroides* adaptation enzymes) or 37^oC (*E. coli* adaptation enzymes) for 1 h. A range of chemoreceptor to enzyme ratios was used, from 10:1 to 1:1. The ratios used did not affect the final results. The final concentrations of all components in the reaction mix were driven by the concentration of individual components.

An excess of the substrate SAM was added to reactions with CheRs. Chemoreceptor concentrations did not exceed 100 μM and SAM was added at 5 mM. The phosphodonor phosphoramidate was added in excess to reactions with full-length CheBs. CheB concentrations did not exceed 10 μM and phosphoramidate was added at 50 mM.

After incubation, reactions were stopped by the addition of protein loading dye. Samples were normalised by the concentration of the chemoreceptor for ease of comparison. The samples were then diluted to give a final concentration of 10μM chemoreceptor. The samples were denatured and separated by gel electrophoresis. Large polyacrylamide gels were used and low voltage at low temperature was applied over 12 h for clear separation of bands. Gels were stained with a stain specific to the His-tag on each protein (In-Vision His-tag stain from Invitrogen), to avoid staining any contamination in the samples remaining from the purification. Further

details of the method are given in Section 2.4.

6.4 Results

Brightness and contrast are raised on all images. Some images are inverted to aid band detection.

6.4.1 Tsr modification by *E. coli* enzymes

Tsr_{CD} was incubated with *E. coli* CheB, $CheB_c$ and CheR at chemoreceptor to enzyme ratios of 100 μM:10 μM and 50 μM:10 μM. All ratios used produce the same results.

Additional states are identified by comparing bands with those obtained from Tsr_{CD} incubated alone and the run positions of enzymes in test cases. An example of a reaction is shown in Figure 6.1. Purified Tsr_{CD} runs as a single band (lane 2), closer to the 40 kDa marker than its expected weight of 32.9 kDa would suggest, but in the same position as on small gels run for confirmation following purification (data not shown). Incubation with $CheB_c$ (lane 3) results in an additional band running at the position of a protein at least 1 kDa heavier. $CheB_c$ cannot be seen as its concentration is too low for detection by the stain used and at 22.2 kDa, its band lies below the cut-off of this image. An identical result is seen with full length CheB at all ratios tested (not shown). Presumably, the two Tsr_{CD} bands correspond to the QEQE and EEEE states.

Incubation with CheR (lane 4) results in an additional band slower than the Tsr_{CD} band, positioned between the control Tsr_{CD} band and the slower band produced by CheB. There is a definitely shift from unmodified to modified, as the lower band, corresponding to unmodified Tsr, is much lighter in intensity than the upper band. Here it was assumed that CheR was not visible due to its low concentration in the reaction mix (approximately 1 μM following dilution). However, as CheR is close in

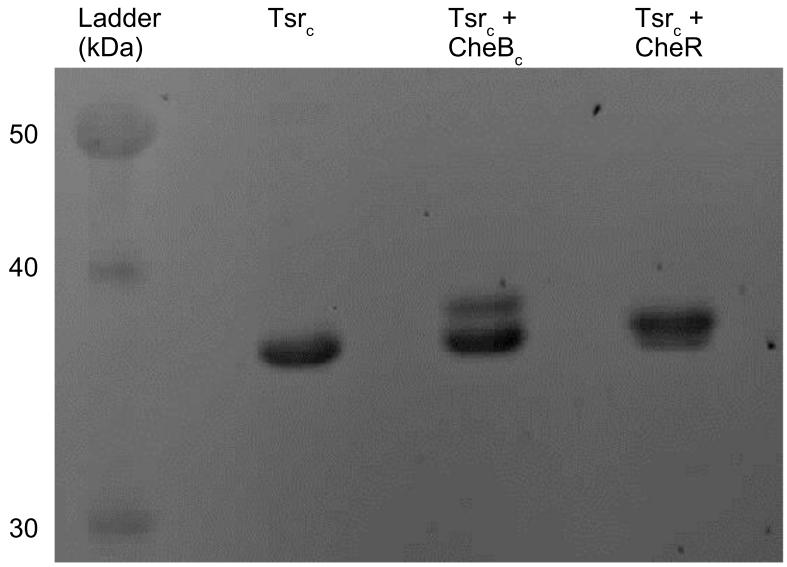


Figure 6.1: Tsr$_{CD}$ after 1 h incubation with *E. coli* adaptation proteins, separated by gel electrophoresis and stained with In-Vision His-tag Stain. The only bands visible are all Tsr$_{CD}$ states.

molecular weight to Tsr$_{CD}$ but slightly larger, there is a chance that the additional band seen is actually CheR, not an additional methylation state. This had not been expected, as CheR separated by gel electrophoresis ran at a lower molecular weight than expected, at approximately 30 kDa, when checked following purification (data not shown).

This result validates the use of this method for identifying chemoreceptors which are modified by adaptation enzymes, provided there is a sufficiently large difference between the molecular weights of the adaptation proteins and the chemoreceptor used.

6.4.2 Tsr modification by *R. sphaeroides* enzymes

Tsr$_{CD}$ was incubated with the *R. sphaeroides* adaptation enzymes at chemoreceptor to enzyme ratios of 100 μM:10 μM, except for CheR$_3$ which was incubated at the ratio of 45 μM:5 μM due to low CheR$_3$ yield. Incubation with the *E. coli* adaptation enzymes was done simultaneously as a control and the band positions were

compared to these lanes and pure Tsr_{CD} (Figure 6.2).

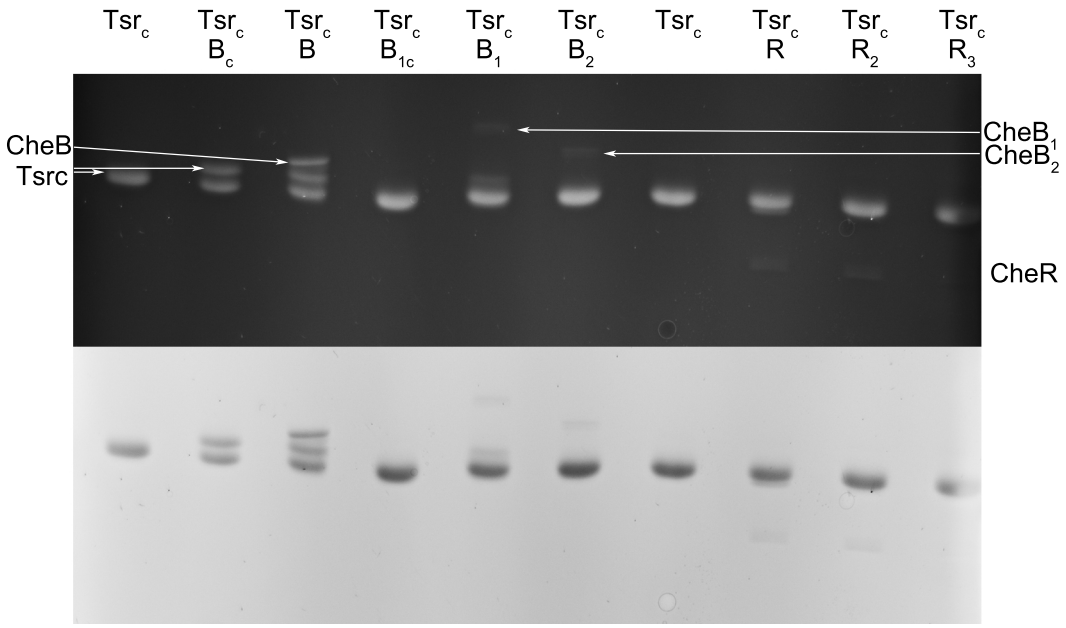


Figure 6.2: Tsr_{CD} after 1 h incubation with *E. coli* and *R. sphaeroides* adaptation proteins, separated by gel electrophoresis and stained with In-Vision His-tag Stain. Band identities are indicated by arrows. *E. coli* $CheB_c$ and $CheR_3$ cannot be seen at this resolution.

Full length $CheB_1$ is able to deamidate Tsr (lane Tsr_c+B_1), as a faint band is seen just above the main Tsr_{CD} band, corresponding to the band seen on incubation with *E. coli* CheB. Neither truncated $CheB_1$ nor $CheB_2$ are able to deamidate Tsr under these conditions (lanes Tsr_c+B_{1c} and Tsr_c+B_2), as both have only the main Tsr_{CD} band. In all cases, the molecular weight of the *R. sphaeroides* CheBs and truncated CheBs are sufficiently different that they do not interfere with the chemoreceptor bands. The full length CheBs can be seen 10 kDa above the chemoreceptor bands, while the truncated CheBs (10 kDa smaller than the chemoreceptors) cannot be seen.

Once again, no conclusions can be drawn regarding methylation. Although a dual band is seen around the chemoreceptor on incubation with *E. coli* CheR, there are no corresponding bands for the *R. sphaeroides* CheRs. Further, although faint bands are seen approximately 5 kDa below the chemoreceptor position in the *E. coli* CheR and *R. sphaeroides* $CheR_2$ lanes, which may correspond to these CheRs, the use of only the

His-tag stain here is inconclusive. If specific antibodies against the chemoreceptor were used, then the adaptation proteins would not be visible on the gel.

6.4.3 McpJ modification by *R. sphaeroides* enzymes

$McpJ_{CD}$ was incubated with $CheB_1$, $CheB_{1c}$, $CheB_2$, $CheR_2$ and $CheR_3$ at chemoreceptor to enzyme ratios of 30 μM:30 μM and 30 μM:10 μM. $McpJ_{CD}$ undergoes spontaneous cleavage, or the sample is contaminated with another protein, so runs as two bands of approximately 30 kDa and 35 kDa (Figure 6.3), compared to the expected weight of 31.4 kDa. This dual band was reproducible and was also seen in the small gels run to confirm protein identity following purification.

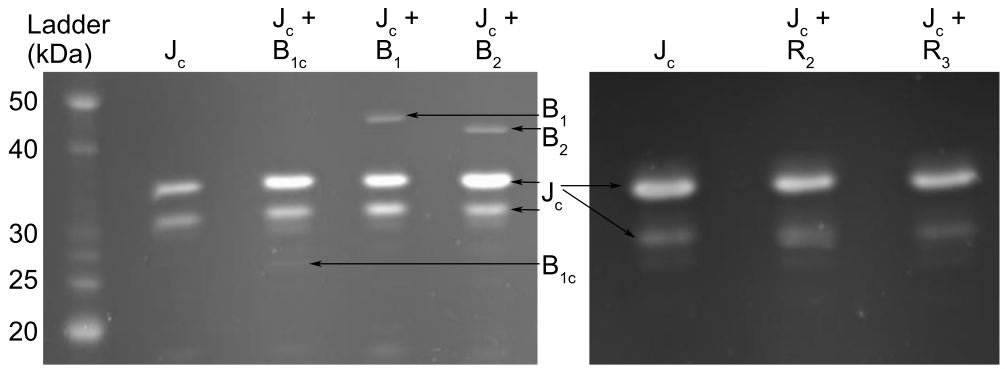


Figure 6.3: $McpJ_{CD}$ after 1 h incubation with *R. sphaeroides* adaptation proteins, separated by gel electrophoresis and stained with In-Vision His-tag Stain. Band identities are indicated with arrows.

No additional bands are seen after incubation with any of the *R. sphaeroides* adaptation enzymes at either ratio: the truncated McpJ was not modified by any *R. sphaeroides* adaptation enzyme under the conditions tested. Only the CheB homologues can be seen on these images, although the truncated $CheB_{1c}$ is very faint.

6.5 Discussion

6.5.1 Method appraisal

This *in vitro* method of screening for chemoreceptors that are modified by adaptation enzymes was able to detect band shifts in a truncated *E. coli* Tsr due to deamidation by CheB and $CheB_c$. Chemoreceptor to enzyme ratios of 5:1 to 10:1 and chemoreceptor concentrations of higher than 50 μM gave optimal results.

However, no clear conclusions could be drawn regarding methylation, as *E. coli* CheR has too similar a molecular weight to the truncated chemoreceptor. Either the His-tag must be removed from CheR before incubation, or a more specific visualisation technique should be used, such as a Western blot with a primary antibody specific to the chemoreceptor.

6.5.2 *R. sphaeroides* adaptation enzymes

R. sphaeroides $CheB_1$ was able to deamidate Tsr_{CD}. However, the constitutively active $CheB_{1c}$ was not able to deamidate Tsr, although *E. coli* $CheB_c$ could. It is possible that the precise domain edges selected for $CheB_1$ were not optimal. Alternatively, truncated $CheB_1$ may not fold correctly. *R. sphaeroides* $CheB_{2c}$ could not be purified, as expression levels were below background noise. Considering that the *R. sphaeroides* CheBs are more similar to one another than they are to *E. coli* CheB, this lends strength to the argument that *R. sphaeroides* CheB structure requires the full protein to fold and function correctly.

The differing ability of the *che*Op_2 and *che*Op_3 adaptation proteins to modify Tsr_{CD} is as would be expected from previous work on *R. sphaeroides* complementation of *E. coli* deletions. The differing abilities suggests that the differences seen in the action of the two sets of enzymes in *R. sphaeroides*, as measured through deletion mutants,

is not just driven by localisation but also by differences in the enzymes' active sites.

6.5.3 *R. sphaeroides* chemoreceptors

McpJ was predicted to be one of the polar cluster chemoreceptors most likely to be involved in methylation-based adaptation. Although the final five residues of McpJ do not correspond to the Alexander and Zhulin (2007) adaptation protein pentapeptide tether consensus sequence, *in vitro* the adaptation proteins should not require localisation before modifying the chemoreceptor. However, none of the *E. coli* or *R. sphaeroides* adaptation enzymes modified $McpJ_{CD}$ under any conditions tested.

It is possible that the conditions of the assay were not conducive to modification. The concentration of McpJ in reaction mixes was less than that of Tsr, as it was expressed at a lower concentration. Unfortunately, it was found that the *E. coli* and *R. sphaeroides* chemoreceptors purified here cannot be concentrated using centrifuge concentrators or size exclusion chromatography, as they are prone to either adhering to the concentrator membrane or aggregating and not eluting in a gel filtration column.

Although conditions were found under which Tsr was modified, it cannot be assumed that these conditions apply to the 34H polar cluster receptors. These receptors may need additional components in the reaction mix, or may require the structure of associating membrane-bound receptors, or may require tethering of the adaptation enzymes through a pentapeptide tail. The ratio of chemoreceptor to enzyme required under *in vitro*, clusterless conditions may be radically different from that of Tsr, or the receptors may need to associate in heterodimers. 34H chemoreceptors may prove not to show methylation band shifts, as seen in the *E. coli* 36H chemoreceptors.

If one receptor was found to be modified through some other means, then it could be used to determine the conditions under which *R. sphaeroides* receptors will be mod-

ified in this assay. There are simply too many variables to test when the chemoreceptor itself is also uncertain.

As in Chapter 5, the method used in this chapter for screening for modified chemoreceptors and the enzymes that modify them is a useful one. However, there are as yet too many unknowns: in the previous chapter's case, lack of a prediction sequence, in this case, lack of a known modified chemoreceptor. Rather than try many iterations of the assay to find the correct chemoreceptor or pair of chemoreceptors needed, the correct ratios of proteins and any additional components needed, Chapter 7 presents a different approach to identify a chemoreceptor involved in adaptation and its modification sites, using the intact dual cluster chemotaxis system to be used to methylate that chemoreceptor.

Chapter 7

Identifying methylation sites in *R. sphaeroides* chemoreceptors using mass spectrometry

In this chapter, *R. sphaeroides* is used as a black-box to methylate and deamidate a tagged chemoreceptor. The receptor is purified and modification sites identified by tandem mass spectrometry following elastase digest. The method is validated using *E. coli* Tsr, then used to identify three adaptation sites on *R. sphaeroides* TlpT. From these sites, a consensus sequence for methylation sites in soluble *R. sphaeroides* chemoreceptors is presented and used to predict an additional site in TlpT.

7.1 Introduction

Is the cytoplasmic cluster involved in methylation-based adaptation? Although the swimming behaviour of chemotaxis mutants, specifically single deletions of the CheRs and CheBs, suggests so (Chapter 3), methanol release is not affected by loss of the cytoplasmic cluster. TlpT was predicted to be the most likely cytoplasmic chemoreceptor to be involved in adaptation (Chapter 4), but mutating the predicted sites did not give the expected phenotype changes (Chapter 5).

An alternative cytoplasmic chemoreceptor could be selected to mutate and test. This is unhelpful if the prediction sequence is the problem. Alternatively, the *in vitro* assay in Chapter 6 could be used to test TlpT for gel electrophoresis shifts after incubation with CheB or CheR. However, if TlpT is actually methylated, it is sufficiently different from *E. coli* that methylation sites could not be predicted from known methylation sites. There is thus no guarantee that band shifts should occur with methylation. Further, as *R. sphaeroides* is a complex system, predicting which components are required to allow *in vitro* methylation is difficult.

TlpT's involvement in methylation-based adaptation can only be invalidated by exposing TlpT to the full *R. sphaeroides* system under conditions that should give rise to adaptation, and directly measuring methylation.

7.1.1 Using *R. sphaeroides* as a black-box-methylator

In this chapter, *R. sphaeroides* is used as a black box containing all the requirements for methylation. By expressing TlpT at low levels within *R. sphaeroides* with all of the other chemotaxis components included, it is certain that the receptor will be methylated/deamidated if it is methylatable/deamidatable. Low expression levels ensure that sufficient of the expressed receptor will integrate into the chemosensory cluster to be modified and will prevent the formation of inclusion bodies.

If the receptor is His-tagged, it can then easily be purified from the cell after methylation. Before purification, it would be advantageous if as much receptor as possible has been modified. This can be achieved by the modification of an existing protocol (Adase *et al.*, 2012). A motile chemotactic strain is grown to the midlog phase, ensuring that it is in its motile state and is expressing the chemotaxis machinery. Chl is added, halting all further protein production. Methionine, the precursor for SAM, the substrate for CheR methylation, is also added. A 30 min incubation allows deamidation of all of the expressed chemoreceptors to occur through normal

transient leakage through the chemotaxis system, transiently activating CheB, and allows the cells to take up methionine and convert it to SAM.

A large stimulus is then given in the form of attractant at 100 μM. During the 20 min incubation, the chemotaxis system is inactivated and CheR methylates the chemoreceptors. Cells are then collected by centrifugation at 4^{o}C and frozen at -80^{o}C, to avoid loss of methylation.

At this point, there are two options. Either the whole cell extract can be analysed, or the expressed chemoreceptor can first be purified and only the purified protein analysed. Considering the low concentration of chemoreceptor in comparison to the contents of the cell and the low rate of methylation seen in other species, the possible loss of some modification during purification is outweighed by the benefit of concentrating the chemoreceptor through purification.

Purification under denaturing conditions prevents demethylation from occurring, either by specific methylesterases or spontaneously. The methylester bond can be broken under strongly alkali conditions. The standard QIAGEN Ni-NTA affinity purification protocol under denaturing conditions uses 8 M urea and buffers in the mildly alkali to acidic range (pH 4.5-8.0).

7.1.2 Identifying methylation sites

Methylation increases a chemoreceptor's molecular mass by 15 Da, whereas deamidation increases the mass by 1 Da. It is possible to identify these mass changes through tandem mass spectrometry (MS/MS). A previous study successfully identified methylation sites in purified *T. maritima* receptor HCDs using MS/MS (Perez *et al.*, 2006). These HCDs were methylated in an *in vitro* assay similar to that used in Chapter 6 and methylation was first confirmed by tracking integration of H^3 into receptors. After purification, full-length chemoreceptors are essentially analogous to this sample.

The advantage of this method is that nothing need be known about the specifics of the adaptation system or of the locations of methylation sites. All that is required is a known chemoreceptor suspected of involved in adaptation, expressed from a plasmid.

7.2 Confirming *E. coli* Tsr sites using MS/MS

The method was first tested on *E. coli* Tsr, to validate that sufficiently high levels of modification remain after processing to be identified, and to ensure that processing does not introduce so many random modifications that the true signal is swamped.

7.2.1 Constructs and strains used

E. coli Tsr was grown up in two background strains. RP437 is a wildtype *E. coli* strain containing all chemotaxis components. RP1091 is an RP437 mutant that is gutted for all chemotaxis genes, including *cheB* and *cheR*.

Primers Tsr-F and Tsr-R were used to amplify the full *tsr* gene from an existing expression vector, introducing restriction sites NcoI and BglII up and downstream of the gene respectively and removing the stop codon. These enzymes were then used to insert the *tsr* gene into the expression vector pQE60. The resulting vector pQE60-Tsr encoded for full length Tsr with a C-terminal His-tag. Tsr thus no longer has a C-terminal pentapeptide tether available for CheR and CheB. However, Adase *et al.* (2012) attached a C-terminal antibody tether to endogenous levels of Tar *in vivo* and still found high levels of deamidation and methylation. It is assumed that the expressed modified chemoreceptor integrates into clusters with unmodified chemoreceptors. The unmodified chemoreceptors are at a sufficiently high concentration in the cluster to recruit and tether the required adaptation proteins, which can then reach all of the chemoreceptors, modified and unmodified, in the vicinity.

7.2.2 Protein collection and analysis

pQE60-Tsr was transformed into the two background strains (Section 2.2.3). 6 l of each strain was grown up and induced using the principles discussed above. Stationary culture was diluted 1 in 50 and induced with a low concentration of IPTG (2.5μl), producing the lowest controllable level of expression from this plasmid possible. At mid-log phase (OD_{600nm} = 0.6) 30 μg/ml Chl and 100 μM L-Met were added and incubation continued for 30 min. 100 μM of the Tsr attractant L-Ser was added and incubation continued for 20 min. Cells were collected by centrifugation and frozen. The full method is detailed in Section 2.3.2.

Tsr-His was purified from frozen cells using Ni-NTA affinity under denaturing conditions (Section 2.3.4). The protein elutes in one of two sequential elution buffers, depending on aggregation, either Buffer D at pH 5.9 or Buffer E at pH 4.5. The fractions containing Tsr were identified by separating a sample of each fraction by SDS-PAGE using a 4-20% acrylamide gel and comparing the sizes of the resulting bands to a molecular ladder.

MS/MS was performed by Dr Roman Fischer, Oxford Centre for Cellular and Molecular Physiology. Samples were digested with elastase, a proteinase without a specific digest site, to give random short peptides. Electrospray-ionisation MS/MS was used to fragment and separate the peptides. Spectra were searched against a database of *E. coli* proteins, generating the most probable peptide identities. Post-translational modifications allowed were **Q** and **E** methylations and **Q** deamidation. Peptides scoring less than 20 were discounted as improbable.

Two biological repeats were done for each background strain.

7.2.3 Results

Total protein in each sample

Good coverage for Tsr under both conditions was obtained, 89% for wildtype and 84% for the chemotaxis gutted strain, with full coverage over the cytoplasmic domain in both cases. All peptide abundances were normalised by comparing the total protein abundance found in each sample, so that the change in abundance for a particular peptide can be directly compared across backgrounds. This is particularly necessary as the mutant gutted strain is less fit than RP437 and generally grows slower, resulting in a smaller total protein yield.

Identified peptides were grouped according to which residue each covered. Every site that was recorded as modified in at least one peptide was selected for further analysis. Table 7.1 shows all sites with a **E** methylation, **Q** deamidation or **Q** deamidation plus methylation. The number of modified and unmodified peptides that include each site is recorded.

Identification of known adaptation sites in Tsr

There are four known methylation sites in Tsr, Q297, E304, Q311 and E493. All four of these sites were positively identified with deamidation and/or methylation here. Site E493 proved difficult to resolve from its **QE/QE** pair partner, E492, and modifications were identified on both sites, which is a common problem in previous work identifying adaptation sites with mass spectrometry. The results from both of these sites were thus combined.

The four sites all showed modification in 10% or more of the identified peptides including that site. Three of the four sites showed modification in 25% or more of the identified peptides. All four sites showed a clear regulatory pattern, with a higher abundance of modified peptides in the wildtype background than in the

Table 7.1: All sites identified with **Q** or **E** modifications in *E. coli* Tsr by MS/MS.

Methylation on **E** sites			
Site	total peptides	methylated peptides	percentage modified
E304	20	2	10
E321	11	2	18.18
E325	8	2	25
E458	14	1	7.14
E479	55	1	1.82
E492	40	15	37.5
E493	40	5	12.5

Deamidation and methylation on **Q** sites				
Site	total peptides	deamidated peptides	deamidated and methylated peptides	percentage modified
Q48	18	1	0	5.56
Q78	46	1	0	2.17
Q154	24	1	0	4.17
Q173	26	1	0	3.85
Q297	37	8	1	24.32
Q311	28	17	1	64.29
Q318	12	2	2	33.33
Q338	19	0	1	5.26
Q408	21	1	0	4.76
Q466	16	0	1	6.25
Q485	54	2	1	5.56
Q486	53	1	2	5.66
Q504	14	1	0	7.14
Q520	23	1	0	4.35

gutted background (Figures 7.1-7.4) and often a corresponding higher abundance of the unmodified peptides in the gutted background.

In every set there are some peptides which run counter to the overall regulatory pattern. A set will often have one modified peptide with equal abundance in both gutted and wildtype backgrounds. However, the most common overall regulatory pattern is clear.

Deamidation of both known **Q** sites (Q297 and Q311) is clear, which is expected as deamidation is a once-off, irreversible reaction. Methylation is transient and reversible. Despite the possibility of methylation loss during processing, methylation was recorded at both **E** sites. However, the adaptation site with the lowest percentage of modified sites is an **E** site (E304) and both **Q** sites showed predominantly deamidated peptides, not methylated peptides. Although this method is efficient

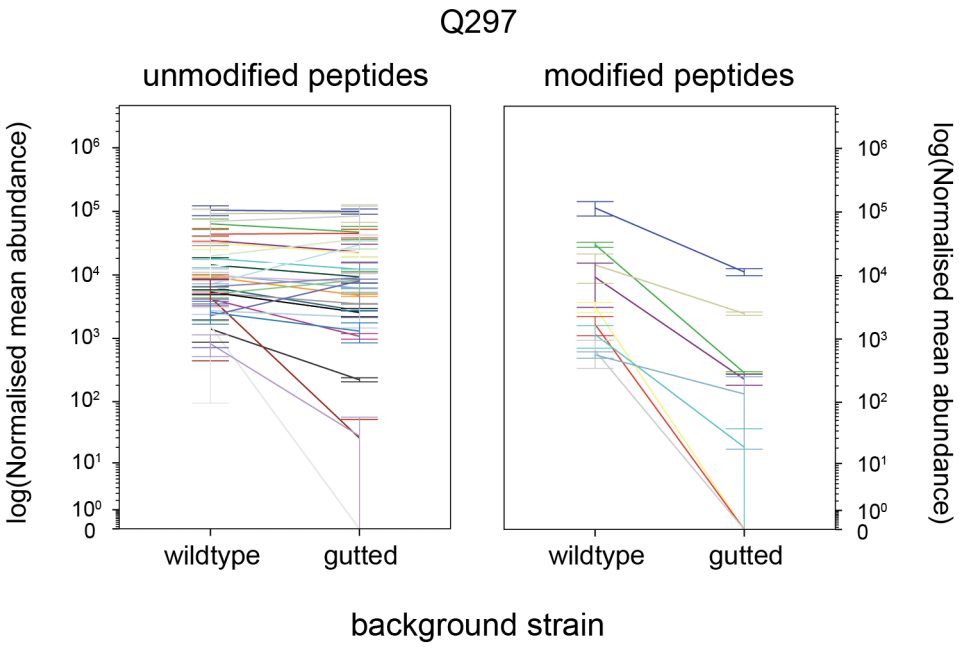


Figure 7.1: Normalised mean abundance of each unmodified (left) and modified (right) Tsr peptide that includes site Q297. Error bars indicate the two actual values recorded.

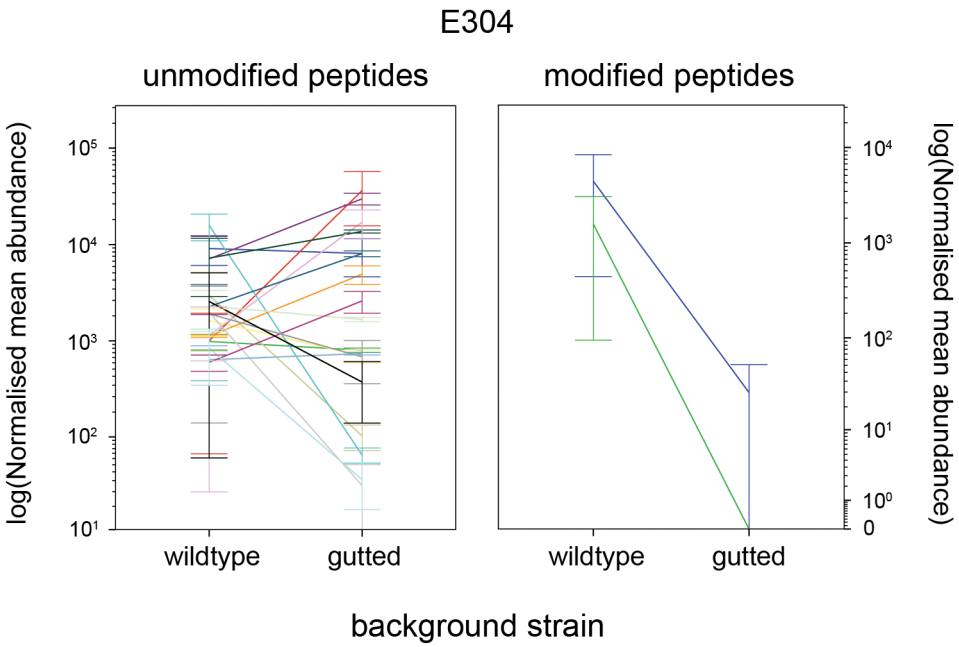


Figure 7.2: Normalised mean abundance of each unmodified (left) and modified (right) Tsr peptide that includes site E304. Error bars indicate the two actual values recorded.

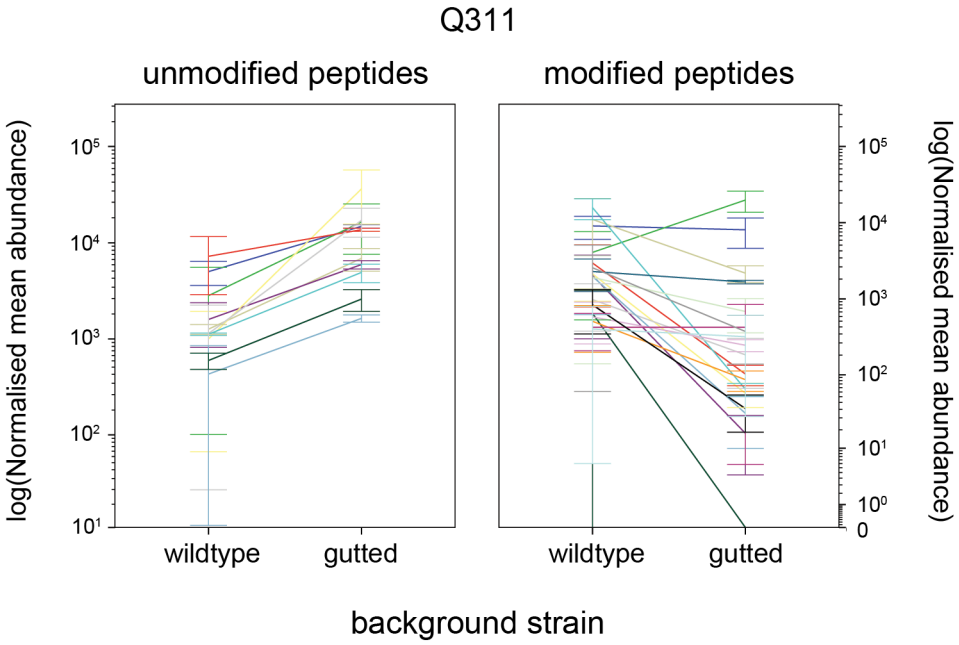


Figure 7.3: Normalised mean abundance of each unmodified (left) and modified (right) Tsr peptide that includes site Q311. Error bars indicate the two actual values recorded.

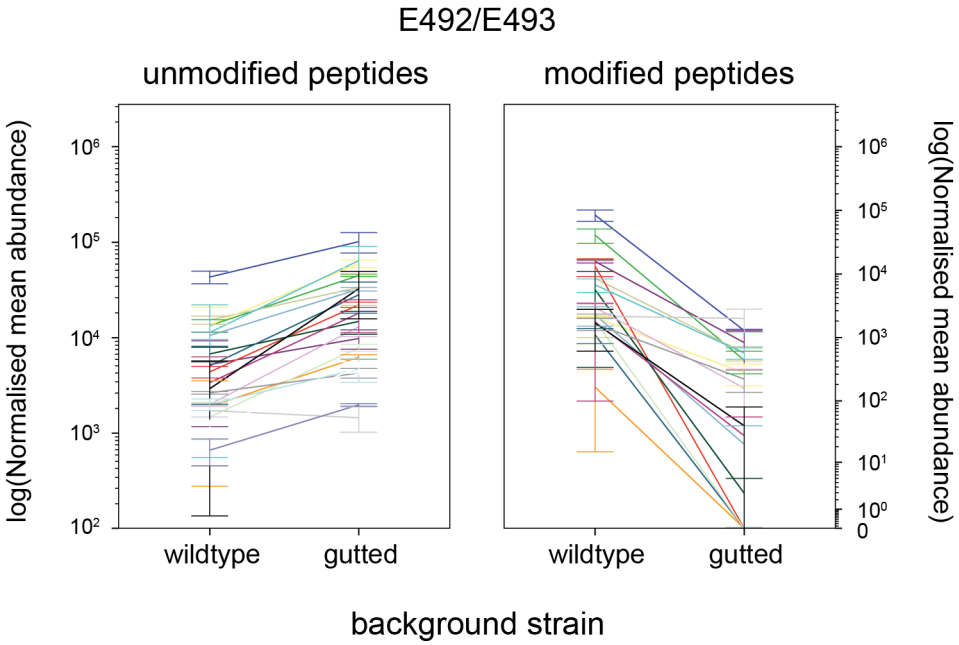


Figure 7.4: Normalised mean abundance of each unmodified (left) and modified (right) Tsr peptide that includes the pair E492/E493. Error bars indicate the two actual values recorded.

at positively identifying **Q** sites through deamidation, the identification of **E** sites through methylation is therefore less efficient. The more reactive an **E** site, the more likely it is to lose methylation during processing, as in E304.

Identification of modification on other sites

A number of sites other than those known to be adaptation sites were identified with modifications. There are two possibilities for these sites.

Firstly, these could be physiologically relevant modifications, made by either the chemotaxis adaptation proteins or by other methyltransferases and deamidases in the cell. This is likely if modifications do not follow the same regulatory pattern as the adaptation sites above.

Secondly, these modifications could be the result of crosstalk within the cell by other methyltransferases from unrelated pathways during the cell collection and lysis process, or the result of spurious deamidation and methylation as a result of the protein purification protocol. This is likely if modifications are seen in a low percentage of peptides covering a particular site or if peptides do not follow any particular abundance pattern.

As the known adaptation sites all show modification in over 10% of peptides, this is taken as the cut-off for modification due to crosstalk. Three sites remain showing modification in a greater percentage of peptides, sites E321, E325 and Q318. All of these show a clearly different regulatory pattern across the two backgrounds compared to the methylation sites (Figures 7.5-7.7). The methylation recorded at these sites is therefore not caused by the chemotaxis adaptation pathway: removal of that pathway in the gutted strain does not result in a decrease in methylation at these sites. Thus, these sites must represent another form of chemoreceptor modification, unrelated to adaptation, that has not previously been recorded.

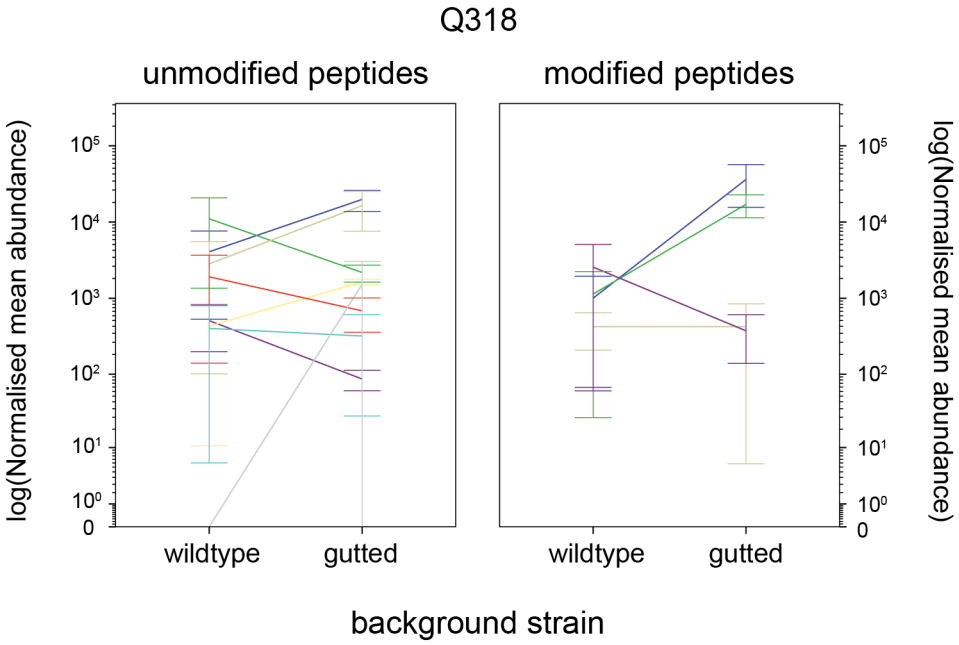


Figure 7.5: Normalised mean abundance of each unmodified (left) and modified (right) Tsr peptide that includes Q318. Error bars indicate the two actual values recorded.

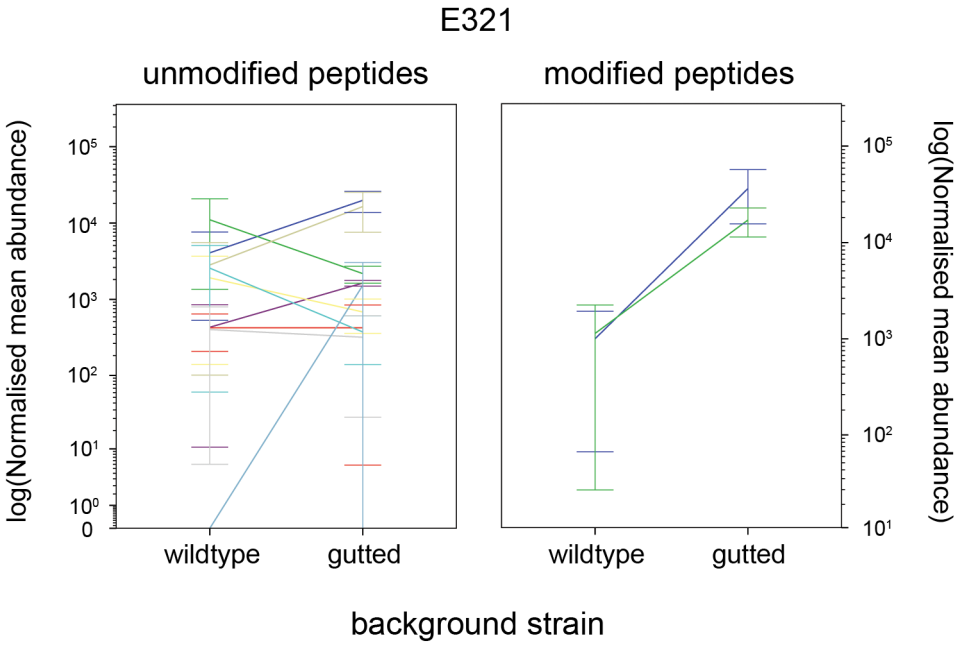


Figure 7.6: Normalised mean abundance of each unmodified (left) and modified (right) Tsr peptide that includes E321. Error bars indicate the two actual values recorded.

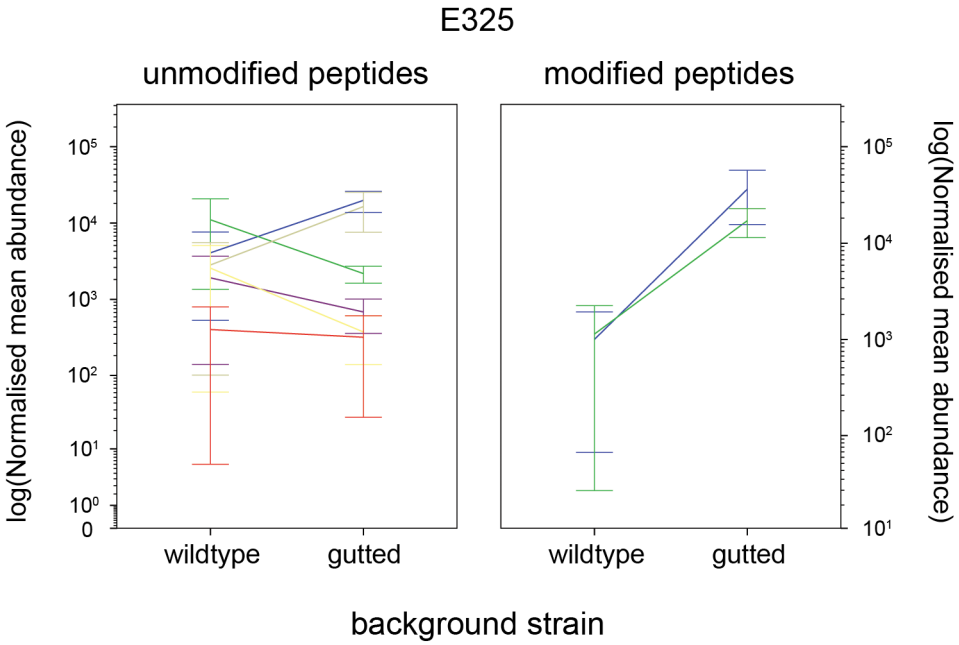


Figure 7.7: Normalised mean abundance of each unmodified (left) and modified (right) Tsr peptide that includes E325. Error bars indicate the two actual values recorded.

7.2.4 Method appraisal

All four known adaptation sites in Tsr were identified using this method. Three of the four sites had modifications in over 25% of identified peptides covering that site. One site, E304, had only 10% modified peptides. Not all **Q** sites were seen to be methylated, although both were seen to be deamidated. Thus, the method suffers from false negatives, as methylation can be lost during processing. The method is more likely to give a false negative for an **E** site than a **Q** site, as deamidation cannot be lost during processing, whereas methylation is a transient modification and can be lost.

False negatives for particular sites are less problematic, as the primary objective of this method is to identify any methylation site rather than exhaustively identify all methylation sites. Provided at least one site is identified as deamidated/methylated, the chemoreceptor can be subjected to more in-depth analysis. Once sufficient sites have been identified across the *R. sphaeroides* chemoreceptors, a consensus sequence

for *R. sphaeroides* can be determined and used to identify more possible sites.

Some **Q/E** sites in the cytoplasmic domain that are known not to be Tsr adaptation sites were found to be methylated. The majority of these were at low levels, which could be assumed to be the result of crosstalk or the effects of processing. Sites with low levels of modification can thus be assumed to be within the background noise of the method. The rest of the sites have levels of modification comparable with those of the known methylation sites. There is a danger of false positives arising from these sites. However, the clear regulatory pattern seen in the known sites – a vastly decreased abundance of modified peptides in the chemotaxis gutted train compared to wildtype – form a filter. None of the unknown sites with a high level of modification share the known sites' regulatory pattern. Therefore, a site must show a level of modification above the determined background and must show a regulatory pattern consistent with adaptation to be accepted as a potential adaptation site. Only the known methylation sites in Tsr matched these criteria.

False positive are therefore unlikely with this system. This is preferable, as the occurrence of false positives when predicting methylation sites (Chapter 4) made it difficult to decide which chemoreceptors are likely to be involved in adaptation and resulted in a set of mutagenesis experiments on structural sites rather than methylation sites (Chapter 5).

7.3 Analysis of *R. sphaeroides* protein TlpT

Once this method for identifying methylation sites had been validated with Tsr, it was extended to include the cytoplasmic chemoreceptor TlpT.

7.3.1 Constructs used

Primers TlpT-F and TlpT-R were used to amplify the full *tlpT* gene from an existing expression vector pQE30-TlpT, introducing restriction site PciI up and downstream of the gene and removing the stop codon. The *tlpT* PCR product was cut with the restriction enzyme PciI and the expression vector pIND4 was cut with NcoI before ligation of the two. The resulting vector pIND4-TlpT encoded for full length TlpT with a C-terminus His-tag.

7.3.2 Background strains

Full length TlpT was to be expressed within a *R. sphaeroides* strain at low levels to allow for modification, then purified and analysed by MS/MS as before. Initially, three existing background strains were chosen to express TlpT in. WS8N, the wild-type *R. sphaeroides* strain, should show methylation and deamidation of any **Q** or **E** site. JPA 1373 is a deletion strain with the three putative methyltransferases $CheR_1$, $CheR_2$ and $CheR_3$ deleted from the genome, so should show no methylation on **Q** or **E** sites, but full deamidation of **Q** sites. JPA 1324 is a deletion strain with the two putative methylesterases $CheB_1$ and $CheB_2$ deleted from the genome, so should show no deamidation (and thus methylation) of **Q** sites, but full methylation on **E** sites. Using two mutant background strains should solve the problem in *E. coli* of losing **E** site methylation during processing.

However, a test case of TlpT purified from each of these backgrounds revealed high levels of deamidation and methylation in both the deletion backgrounds and unclear patterns of regulation on all identified sites (data not shown).

In addition to the known *cheB* and *cheR* homologues, *R. sphaeroides* contains an additional putative methyltransferase-methylesterase fusion gene, *cheBRA*. This fusion also contains a putative redox sensing PAS domain. Deletion of this gene from WS8N, producing strain JPA 1340, had not shown any noticeable changes in phe-

notype in a swarm plate assay, although CheBRA was shown to be expressed under laboratory conditions (Martin *et al.*, 2003). It had therefore been assumed that CheBRA does not play a role in methylation-based adaptation and was not deleted from the methyltransferase and methylesterase free strains.

The above-background levels of deamidation and methylation in the $CheR_{1,2,3}$ deletion strain and the $CheB_{1,2}$ deletion strain suggested, however, that CheBRA is able to function as a replacement for CheB or CheR, although these results may very well be the result of crosstalk rather than any true biological function. In some species, CheBRA homologues have been shown to act as either a methyltransferase or methylesterase for biological function (Wu *et al.*, 2011).

Two new background strains were made, deleting *cheBRA* from both the $CheB_{1,2}$ and the $CheR_{1,2,3}$ strains. The existing construct pK18-delBRA was used in the standard protocol for allelic exchange (Section 2.2.4) and successful deletion of *cheBRA* confirmed using the Southern blot technique (Section 2.2.4).

A Southern blot probe was prepared by digesting pK18-delBRA with XbaI and XhoI, producing a 500 bp band corresponding to the region upstream of *cheBRA*. DNA was extracted from colonies after a successful second round of recombination. Chromosomal DNA was digested with NotI and NcoI, cutting respectively sufficiently upstream and downstream of the gene *cheBRA* so as to cut beyond the regions complementary to pK18-BRA. DNA from colonies that reverted back to the parental strain in the second round of recombination gave a band of approximately 5000 bp containing *cheBRA*. DNA from colonies with a successful deletion in the second round of recombination gave a band of approximately 1600 bp containing sequences up and downstream of *cheBRA*. Both positive and negative chromosomal DNA controls were used, in the form of the BRA deletion strain and WS8N.

50 colonies were checked for each attempted deletion strain. Three positives were obtained for the $CheR_{1,2,3}$ deletion background strain (Figure 7.9), giving the new

strain JPA 2366, and two positives for the $CheB_{1,2}$ deletion background strain, giving new strain JPA 2365 (Figure 7.9).

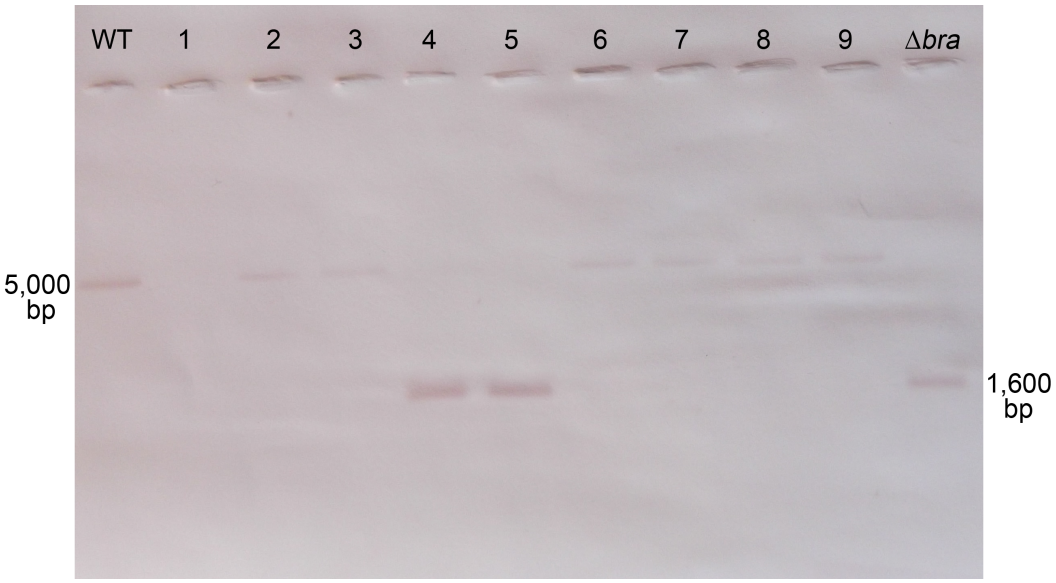


Figure 7.8: Successful deletion of *cheBRA* from the $CheB_{1,2}$ deletion strain (JPA 1324). DNA from colonies after the second recombination round were digested with NotI and NcoI and hybridised with DIG-tagged DNA from the region upstream of *cheBRA*. Bands of 5,000 bp indicate reversion to parental strain, confirmed by comparison to the WS8N DNA digest (lane WT). Bands of 1,600 bp indicate successful deletion of *cheBRA*, confirmed by comparison to the *cheBRA* deletion strain (JPA 1340) DNA digest (lane Δ*cheBRA*). Lanes 4 and 5 show successful deletions.

7.3.3 Protein collection and analysis

The plasmid pIND4-TlpT was conjugated into the three background strains, WS8N, the methylesterase deletion strain (ΔCheB, JPA 2365) and the methyltransferase deletion strain (ΔCheR, JPA 2366), (Section 2.2.3). 6 l of each culture was grown, induced and collected in triplicate on separate days using the same protocol as for *E. coli* Tsr, with the exception of the incubation temperature (30^{o}C) and the attractant (propionate). The tagged chemoreceptor was purified from the cell extract under denaturing conditions.

Samples from the elution fractions were separated by SDS-PAGE (Section 2.3.5), stained with Coomassie protein stain and compared to a molecular weight ladder

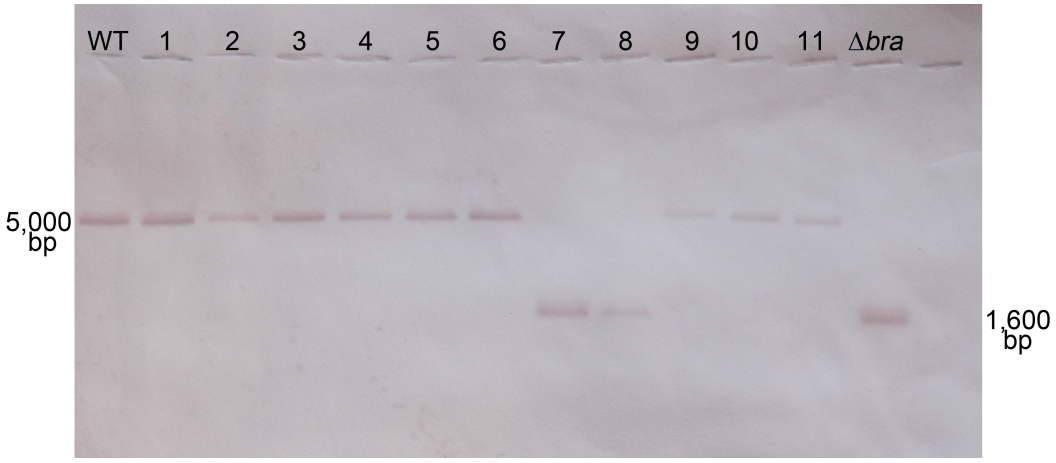


Figure 7.9: Successful deletion of *cheBRA* from the $CheR_{1,2,3}$ deletion strain(JPA 1373). DNA from colonies after the second recombination round were digested with NotI and NcoI and hybridised with DIG-tagged DNA from the region upstream of *cheBRA*. Bands of 5,000 bp indicate reversion to parental strain, confirmed by comparison to the WS8N DNA digest (lane WT). Bands of 1,600 bp indicate successful deletion of *cheBRA*, confirmed by comparison to the *cheBRA* deletion strain (JPA 1340) DNA digest (lane Δ*cheBRA*) Lanes 7 and 8 show successful deletions.

to identify the fractions containing the tagged chemoreceptor. Fractions containing the chemoreceptor were pooled and concentrated approximately ten times using a centrifuge concentrator. An excess of this concentrated sample was then separated using SDS-PAGE and the bands containing TlpT excised (Figure 7.10).

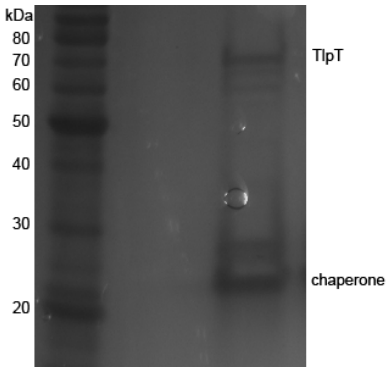


Figure 7.10: TlpT-His is expressed in WS8N and purified by denaturing Ni-NTA chromatography. Elution fractions containing protein are pooled and concentrated, then separated by SDS-PAGE. Two proteins have eluted, TlpT-His and a chaperone (identified through MS/MS, data not shown).

Samples were digested with elastase and separated by MS/MS as for the *E. coli* control. For full details of the method, see Section 2.3.

7.3.4 Results

Quality of TlpT MS/MS results

97% coverage of full-length TlpT was obtained. Total TlpT in each sample is calculated. To allow samples to be directly compared, these total abundances are used to normalise across the samples. Each peptide abundance within each sample is thus normalised.

The three biological replicates for each background strain showed good agreement in terms of abundance for each identified peptide, as shown by a principal component analysis of the nine separate repeats in Figure 7.11. Here, each peptide's normalised abundance is treated as a separate data point relating to its sample. The overall identity and abundance of all the peptides in a repeat determines where in the plot that repeat is represented. Each repeat is represented by a single point, colour-coded as to the known identity of the repeat (what background it represents). In Figure 7.11, the repeats clearly cluster based on their known backgrounds.

The distribution of all normalised abundances recorded does not follow a Gaussian distribution. Triplicates are therefore summarised by median and range (minimum and maximum) throughout, rather than mean and standard deviation.

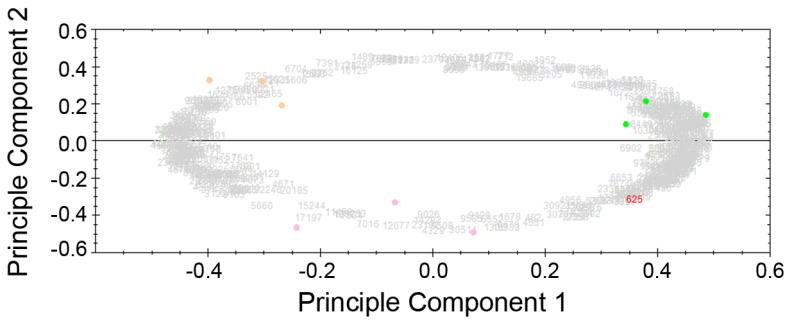


Figure 7.11: Principle component analysis of the MS/MS results for nine TlpT-His samples, grown in three backgrounds. Each peptide is treated as a separate data point. Repeats from the WS8N background are pink, the methylesterase deletion strain (ΔCheB) are orange and the methyltransferase deletion strain (ΔCheR) are green.

As in the *E. coli* case, all modified peptides were examined. Table 7.2 shows all sites with a **E** methylation, **Q** deamidation or **Q** deamidation plus methylation. The number of peptides each site appears in, in the modified and unmodified form, is recorded. In Tsr, sites which only saw modification as a side effect of denaturing protein purification showed less than 10% modified peptides. To avoid prematurely discounting sites in TlpT, only sites which show modification in less than 5% of the peptides (usually only a single peptide) identified are not considered biologically relevant. Discarding these rare events, there are nine sites of interest.

Table 7.2: All sites identified with **Q** or **E** modifications in TlpT by MS/MS, appearing in any of the three background strains

Methylation on **E** sites			
Site	total peptides	methylated peptides	percentage modified
E30	28	3	10.71
E32	28	1	3.57
E96	40	2	5.00
E146	13	3	23.08
E170	44	34	77.27
E296	56	16	28.57
E435	38	1	2.63
E450	37	1	2.70
E478	73	23	31.51
E521	104	1	0.96

Deamidation and methylation on **Q** sites				
Site	total peptides	deamidated peptides	deamidated and methylated peptides	total percentage modified
Q149	34	0	3	8.82
Q183	42	2	1	7.14
Q442	45	1	0	2.22
Q451	37	0	1	2.70
Q457	17	0	2	11.76
Q485	103	42	0	40.78
Q487	91	0	1	1.10
Q525	78	1	0	1.28

Methylation of E residues in TlpT

Sites E30, E146, E170, E296 and E478 show methylation. The abundance for a particular peptide can be directly compared across samples, indicating whether there is a relative change in abundance in different backgrounds. However, the absolute

abundance of different peptides cannot be compared to one another. Whether a particular peptide is recorded at high or low abundance depends on a number of factors, not only the abundance of that peptide in the original sample. Only the relative change in abundance across backgrounds can be directly compared.

Therefore, all modified and unmodified peptides are plotted separately. A logarithmic scale is used for abundance, both to be able to visualise low and high abundance peptides at the same time and to emphasise fold differences between different backgrounds. If a number of similar peptides show a similar relative change in abundance across backgrounds, this can be taken as evidence for regulation.

Due to crosstalk within any system, regulation does not require a modified or unmodified peptide be recorded at zero abundance. All that is required is the relative change across backgrounds.

If an **E** site is methylated by a CheR homologue, the expected regulatory pattern is a high abundance of modified peptides in both the wildtype and CheB deletion backgrounds, and a marked decrease in abundance of each of those peptides in the CheR deletion background. A concurrent increase in unmodified peptides in the CheR deletion background can also be expected. The CheB deletion background and wildtype background are expected to produce similar profiles, as the **E** site does not require deamidation by CheB before it can be methylated. In fact, a small increase in modified peptides may be expected in the CheB deletion background, as methylation will not be lost during processing to active methylesterases.

E30 Site E30 does not show any regulation pattern. Both modified and unmodified peptides (Figure 7.12) have similar abundances across all three backgrounds.

E146 Site E146 shows a consistent pattern of decreased abundance of the unmodified peptides in the two deletion strains, and more commonly in the methylesterase deletion strain (Figure 7.13). In two of the three methylated peptides the same pat-

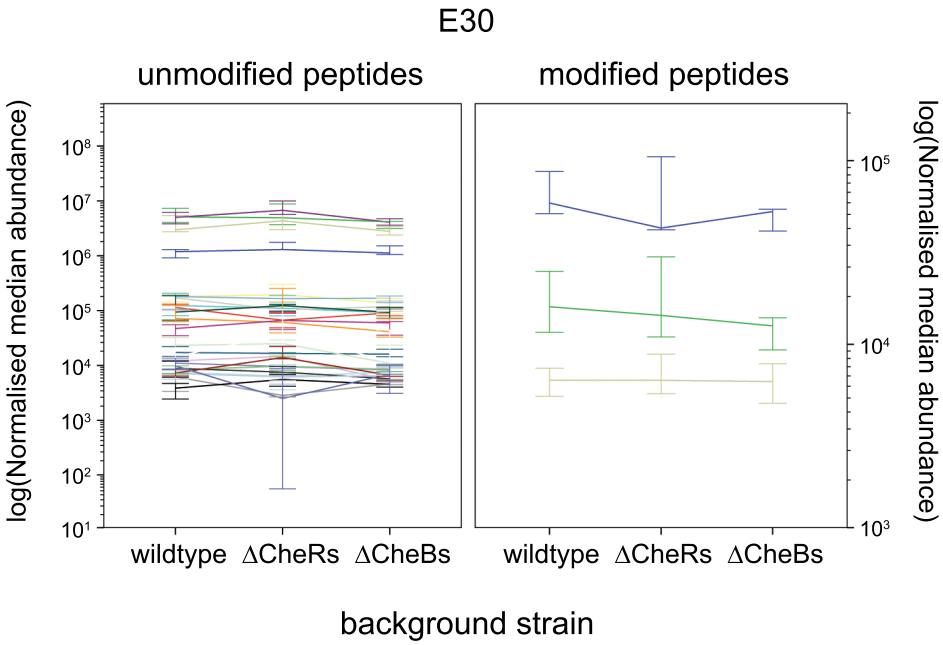


Figure 7.12: Normalised median abundance of each unmodified (left) and methylated (right) TlpT peptide that includes site E30. Error bars indicate the minimum and maximum values in the triplicate.

tern is seen. This pattern of methylation does not fit with methylation via one of the identified methyltransferases.

Site E170 Modified peptides covering site E170 show an increase in abundance in the methyltransferase deletion strain (Figure 7.14). Unmodified peptides show no clear pattern, as some peptides decrease in the methyltransferase deletion strain and others decrease. This site is not methylated by a CheR homologue.

E296 and E478 Both sites within the cytoplasmic domain of TlpT show similar patterns of significant differences between the peptide profiles in the three backgrounds (Figures 7.15 and 7.16 respectively). Levels of modified peptide are ten- to one hundred-fold lower in the CheR deletion strain than in the other two strains for both sites. Zero abundance values were recorded for some peptides in this background. To display these on a logarithmic scale, these values were not transformed to log scale (as they cannot be) and zero was included below the logarithmic scaled

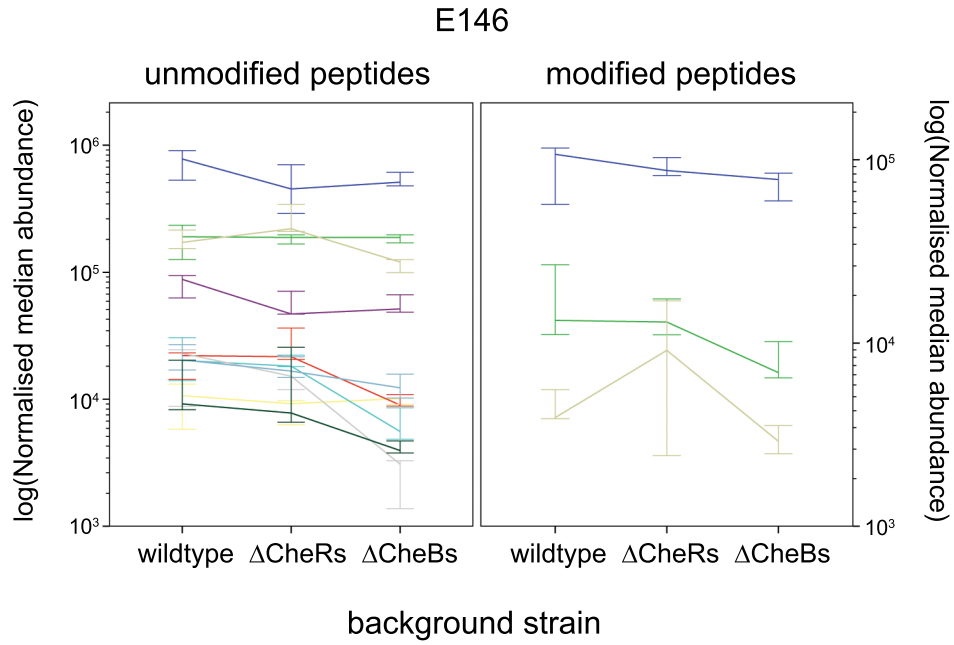


Figure 7.13: Normalised median abundance of each unmodified (left) and methylated (right) TlpT peptide that includes site E146. Error bars indicate the minimum and maximum values in the triplicate.

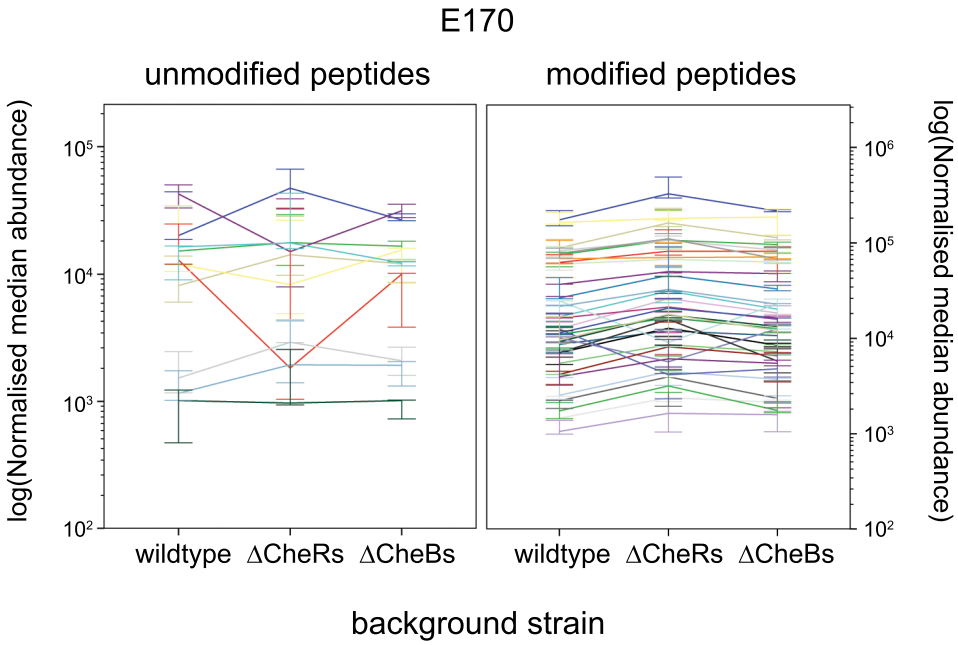


Figure 7.14: Normalised median abundance of each unmodified (left) and methylated (right) TlpT peptide that includes site E170. Error bars indicate the minimum and maximum values in the triplicate.

y-axis.

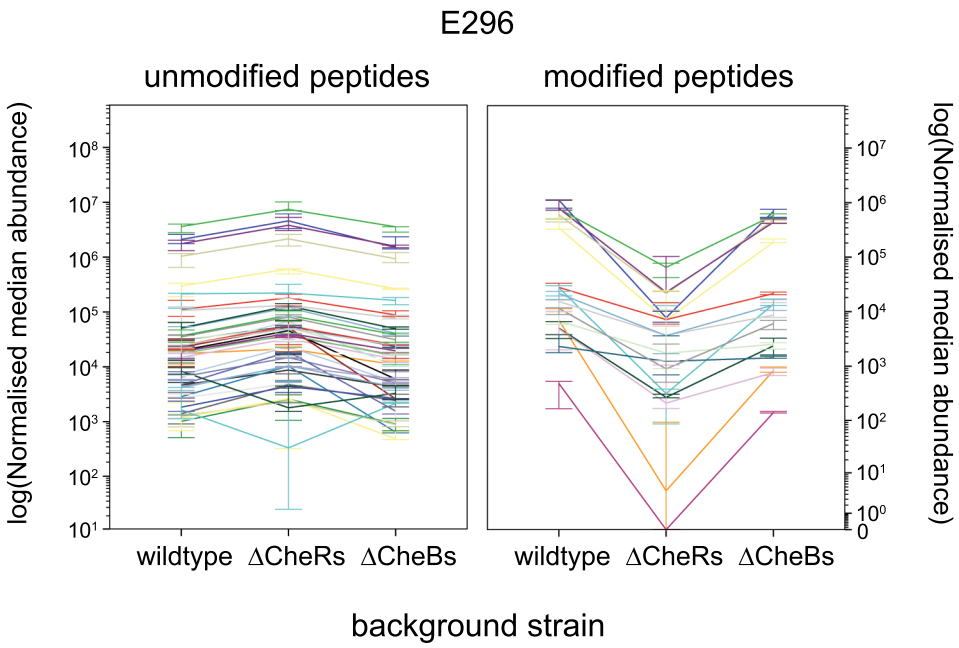


Figure 7.15: Normalised median abundance of each unmodified (left) and methylated (right) TlpT peptide that includes site E296. Error bars indicate the minimum and maximum values in the triplicate. Zero values are indicated below the lowest true value as log(0) cannot be calculated.

Unmodified peptides carrying E296 showed an increase in abundance in the CheR deletion strain. Certain of the both modified and unmodified peptides carrying E478 showed a drop in abundance in the CheB deletion background. These peptides include modification of another site seven residues downstream, Q485, and will be discussed below. The rest of the unmodified peptides show the same pattern as in E296.

Figure 7.17 shows qualitatively the marked difference in profiles between the different backgrounds by focusing on one modified-unmodified peptide pair for each site. These three-dimensional plots show the abundances of peptides normalised to total protein in the sample, separated by their m/z score and the time of elution. Comparing the abundance of a modified and unmodified peptide shows the clear drop in modified peptide in the CheR deletion background, and the often concurrent rise in unmodified peptide in the same.

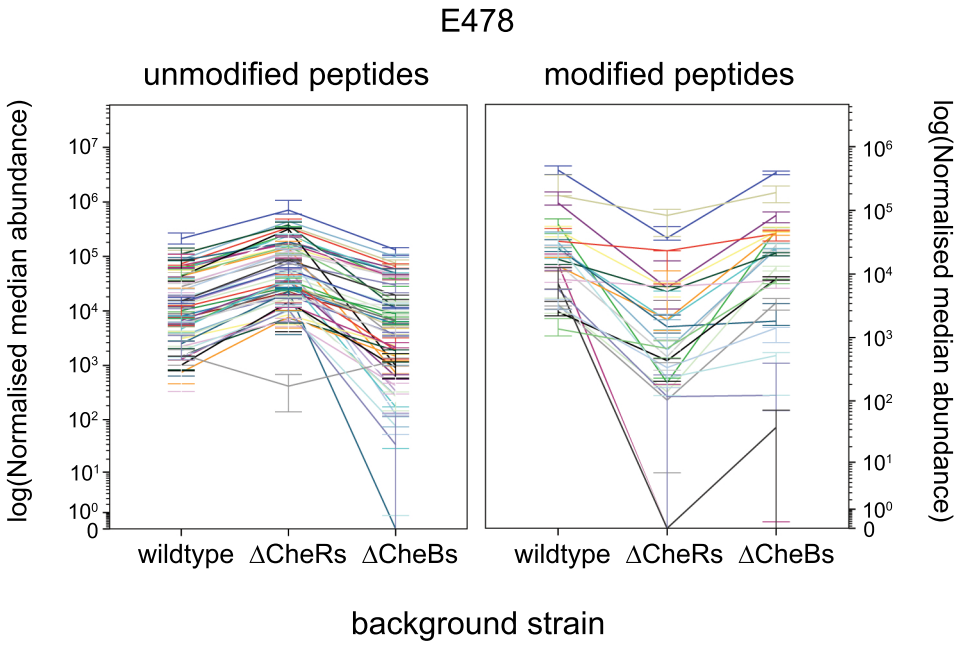


Figure 7.16: Normalised median abundance of each unmodified (left) and methylated (right) TlpT peptide that includes site E478. Error bars indicate the minimum and maximum values in the triplicate. Zero values are indicated below the lowest true value as log(0) cannot be calculated.

The data thus indicates that methylation of both of these sites is regulated by a CheR-homologue methyltransferase.

Deamidation and methylation of Q residues in TlpT

Sites Q149, Q183, Q457 and Q485 show deamidation and, in some cases, methylation. Peptides covering each of these sites were analysed in the same way as for methylated **E** sites. If a **Q** site is an adaptation methylation site, there are two expected regulatory patterns.

A high abundance of deamidated peptides are expected in both the wildtype and CheR deletion backgrounds, and a marked decrease in abundance of each of those peptides in the CheB deletion background. A concurrent increase in unmodified peptides in the CheB deletion background can also be expected. The CheR deletion background and wildtype background are expected to produce similar profiles, as

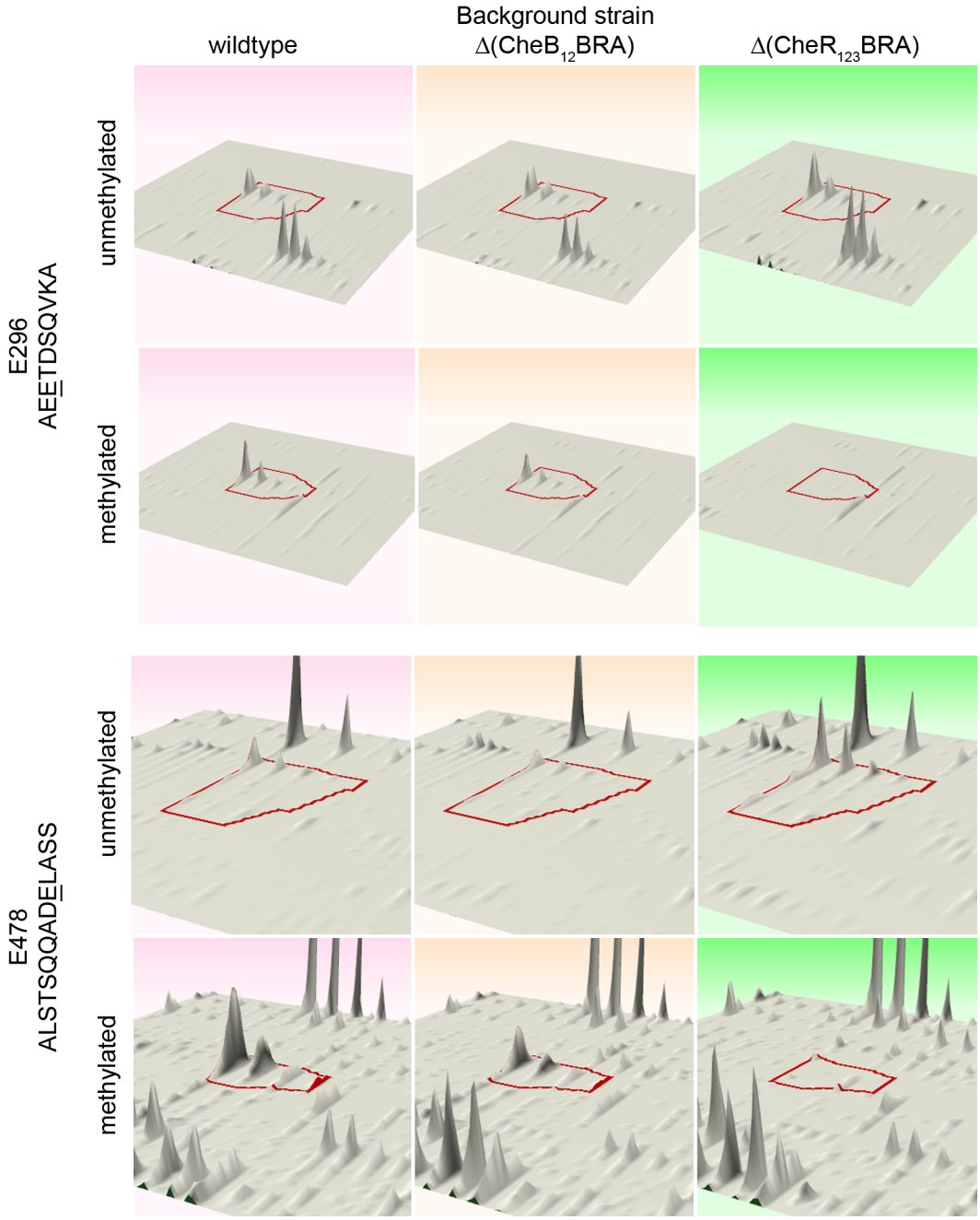


Figure 7.17: A representative peptide for each of the two methylation sites, E296 (top) and E478 (bottom), in TlpT is shown. The normalised abundance (z-axis) of both the unmodified and modified peptide forms is shown for each of the three background strains wildtype, Δ($CheB_1$, $CheB_2$, BRA) and Δ($CheR_1$, $CheR_2$, $CheR_3$, BRA). The red shape encloses the peptide peak of interest. The x and y axes represent m/z and time in this qualitative plot.

the **Q** site does not require methylation by CheR for the deamidation to be recorded

A high abundance of deamidated and methylated peptides are expected in only the wildtype background, with a marked decrease in abundance of each of these peptides in both deletion backgrounds. A concurrent increase in unmodified peptides in both deletion backgrounds is also expected. The deletion backgrounds are expected to produce similar profiles, as the **Q** site must be both deamidated and methylated, requiring both a methylesterase and a methyltransferase, for the modification to be recorded.

Q149 Modification of Q149 is rare and does not follow a fixed pattern (Figure 7.18). In two of the three modified peptides, there is an insignificant drop on abundance in the CheB deletion strain (range bars overlap). In comparison, the modified peptides show similar abundance across all background with an insignificant increase in the CheR deletion strain. This site is unlikely to be modified by the CheB homologues.

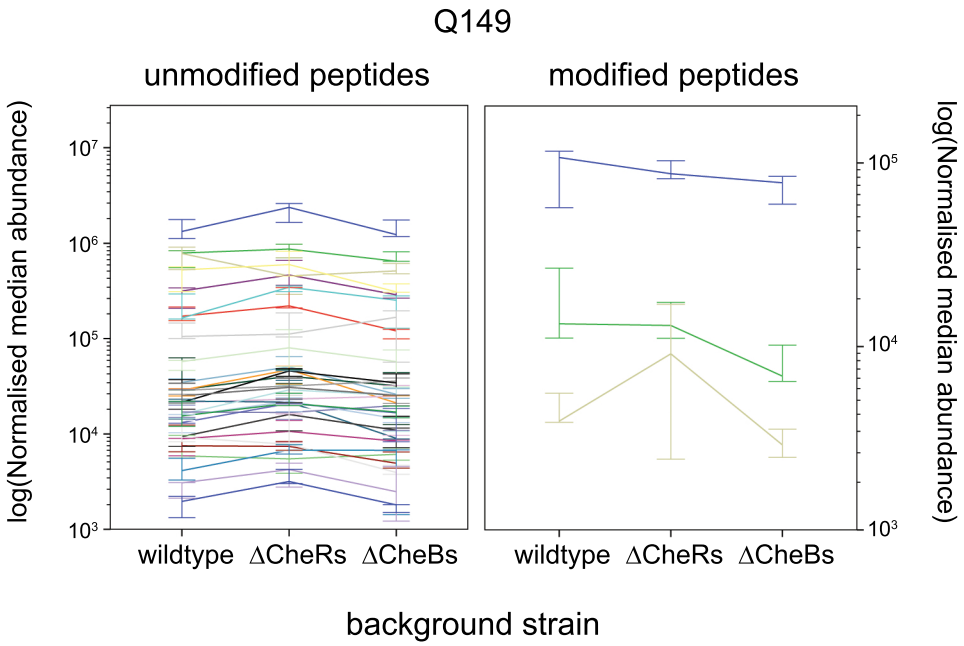


Figure 7.18: Normalised median abundance of each unmodified (left) and methylated (right) TlpT peptide that includes site Q149. Error bars indicate the minimum and maximum values in the triplicate.

Q183 The unmodified peptides that include Q183 have a mixed pattern of abundance changes (Figure 7.19). The abundances in the wildtype and CheB deletion backgrounds are equivalent, whereas the abundances in the CheR deletion background are significantly increased or decreased or also equivalent. The modified peptides all show an increase in abundance in the CheR deletion background, regardless of whether the modification is deamidation or deamidation and methylation. As none of these modified peptides have a decreased abundance in the CheB deletion background, the modification seen is likely not due to the action of the adaptation pathway enzymes.

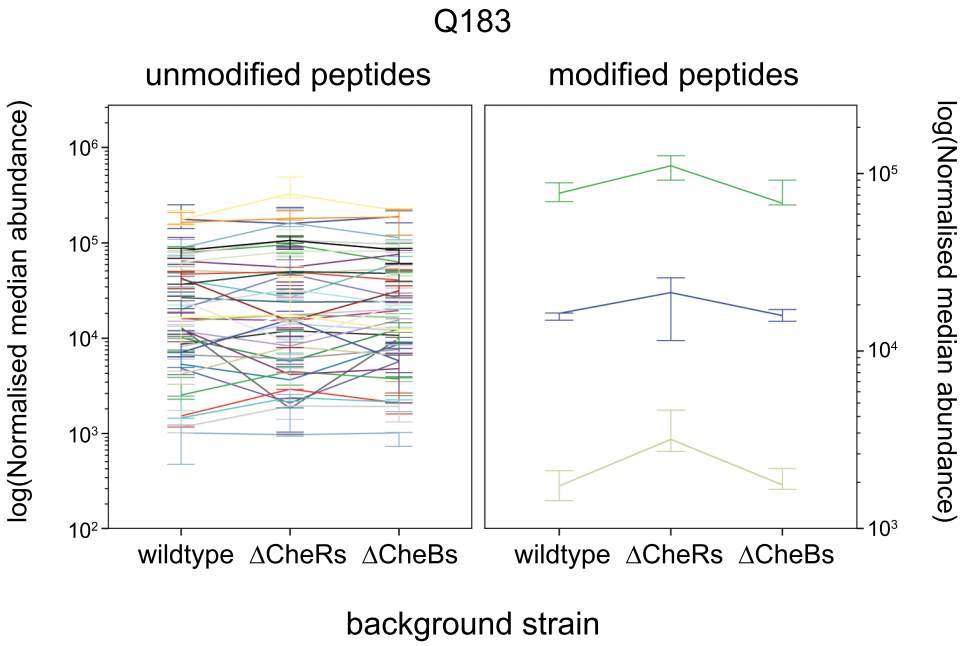


Figure 7.19: Normalised median abundance of each unmodified (left) and methylated (right) TlpT peptide that includes site Q183. Error bars indicate the minimum and maximum values in the triplicate.

Q457 Both modified peptides for site Q457 are deamidated and methylated (Figure 7.20). Neither shows the decrease in abundance in the two deletion backgrounds that this modification would result in if it was due to the action of the adaptation pathway.

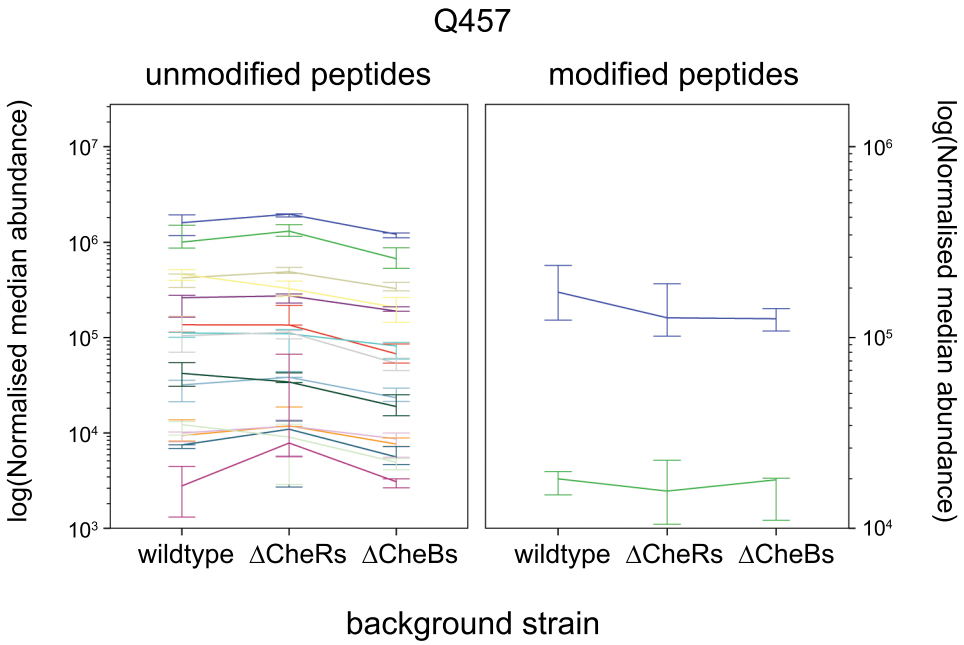


Figure 7.20: Normalised median abundance of each unmodified (left) and methylated (right) TlpT peptide that includes site Q457. Error bars indicate the minimum and maximum values in the triplicate.

Q485 The site Q485 has the highest percentage of peptides showing modification across TlpT. No methylated peptides were seen, only deamidated peptides.

There are significant differences in the abundances of modified and unmodified peptide across the three backgrounds (Figure 7.21). The abundance of modified peptides is significantly decreased in the CheB deletion background compared to wildtype, with some peptides registering zero abundance. This strongly suggests that this site is deamidated by the CheBs.

Like site E478, there appears to be more than one regulatory pattern within both modified and unmodified peptides. In both the modified and unmodified sets, some peptides have equivalent abundances in the wildtype and CheR deletion backgrounds. In others, there is a marked decrease in abundance in the CheR deletion background compared to wildtype. These peptides are those with methylated E478, which is likely methylated by the CheRs.

For a peptide to carry modification on both these sites, both the CheBs and CheRs

must be present. Thus in the modified set, some peptides only show significant abundance in wildtype. In the unmodified set, unmodified only refers to site Q485. Some of these peptides are unmodified on Q485 but methylated on E478. These peptides have low abundance in the CheR deletion background.

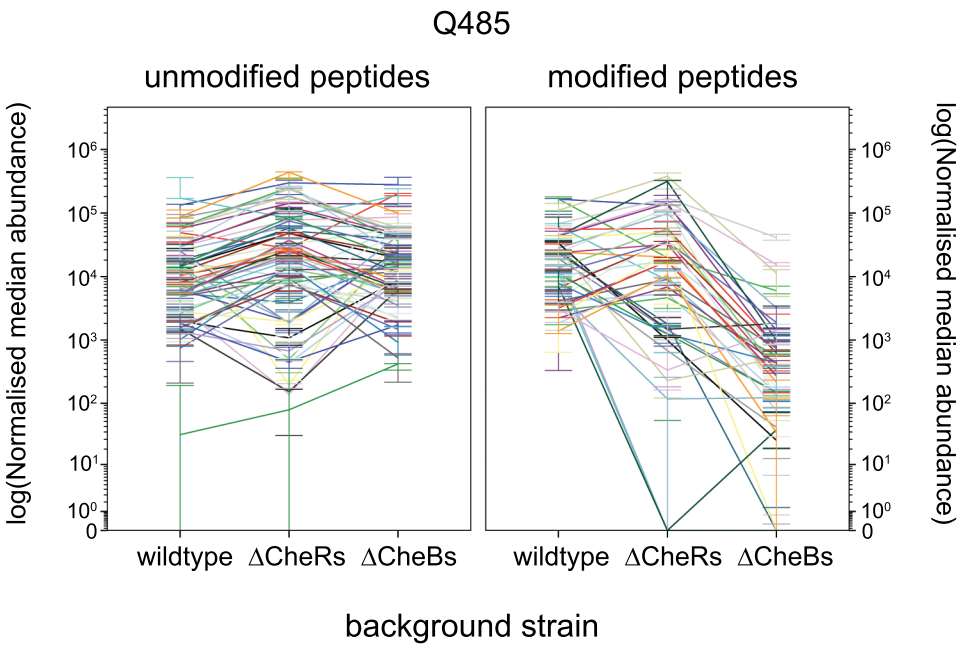


Figure 7.21: Normalised median abundance of each unmodified (left) and methylated (right) TlpT peptide that includes site Q485. Error bars indicate the minimum and maximum values in the triplicate. Zero values are indicated below the lowest true value as log(0) cannot be calculated.

A three dimensional plot of abundance, m/z and time of one modified/unmodified peptide pair qualitatively illustrates the deamidation of Q485 by CheB. (Figure 7.22).

Position of identified sites

The modification sites with a regulatory pattern consistent with adaptation, E296, E478 and Q485, were mapped to the theoretical structure of TlpT obtained through threading and compared to the predicted adaptation sites (Figure 7.23).

Only E296 was predicted to be an adaptation site in Chapter 4. The sidechains for E296 and E478 are in a position likely for methylation, as these sidechains are ori-

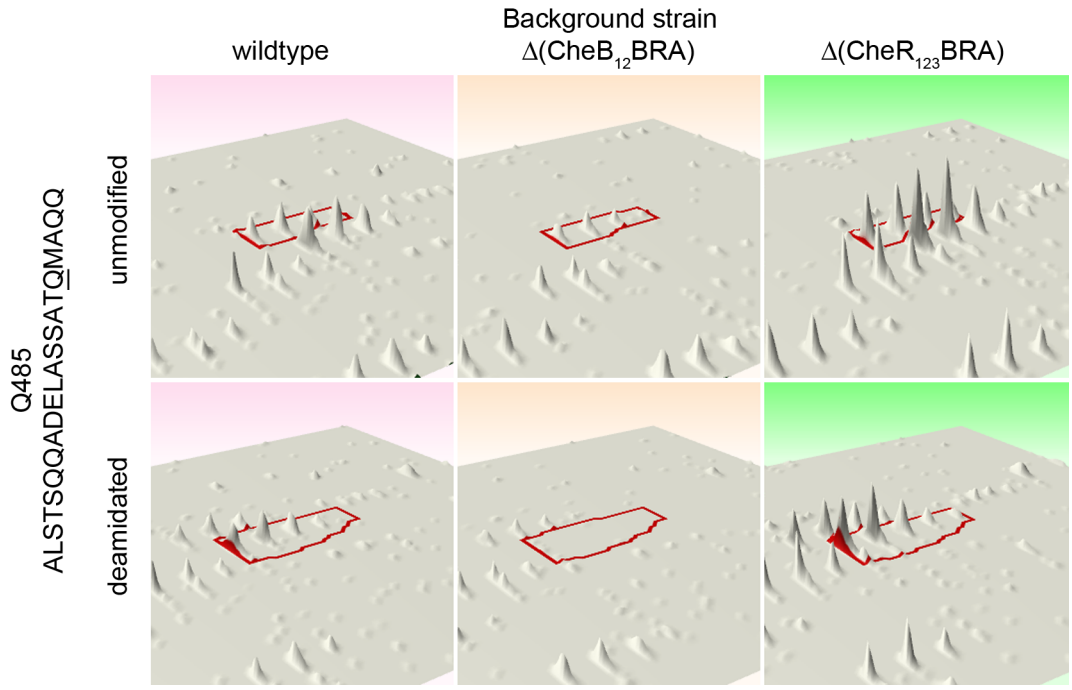


Figure 7.22: A representative peptide for one deamidation site, Q485, in TlpT is shown. The normalised abundance (z-axis) of both the unmodified and modified peptide forms is shown for each of the three background strains wildtype, Δ(CheB$_1$, CheB$_2$, CheBRA) and Δ(CheR$_1$, CheR$_2$, CheR$_3$, CheBRA). The red shape encloses the peptide peak of interest. The x and y axes represent m/z and time in this qualitative plot.

entated out away from the helix in the simulated structure. The sidechain for Q485, however, is orientated into the middle of the helix, a position highly unlikely for deamidation by a CheB.

7.4 Discussion

7.4.1 Using mass spectrometry to identify novel adaptation sites

Using the whole cell as a black box to fully deamidate and methylate a chemoreceptor when little is known about the system is effective, as all four known methylation sites on *E. coli*'s Tsr were identifiable and no false positive identifications were made. Three unknown adaptation sites were identified in TlpT.

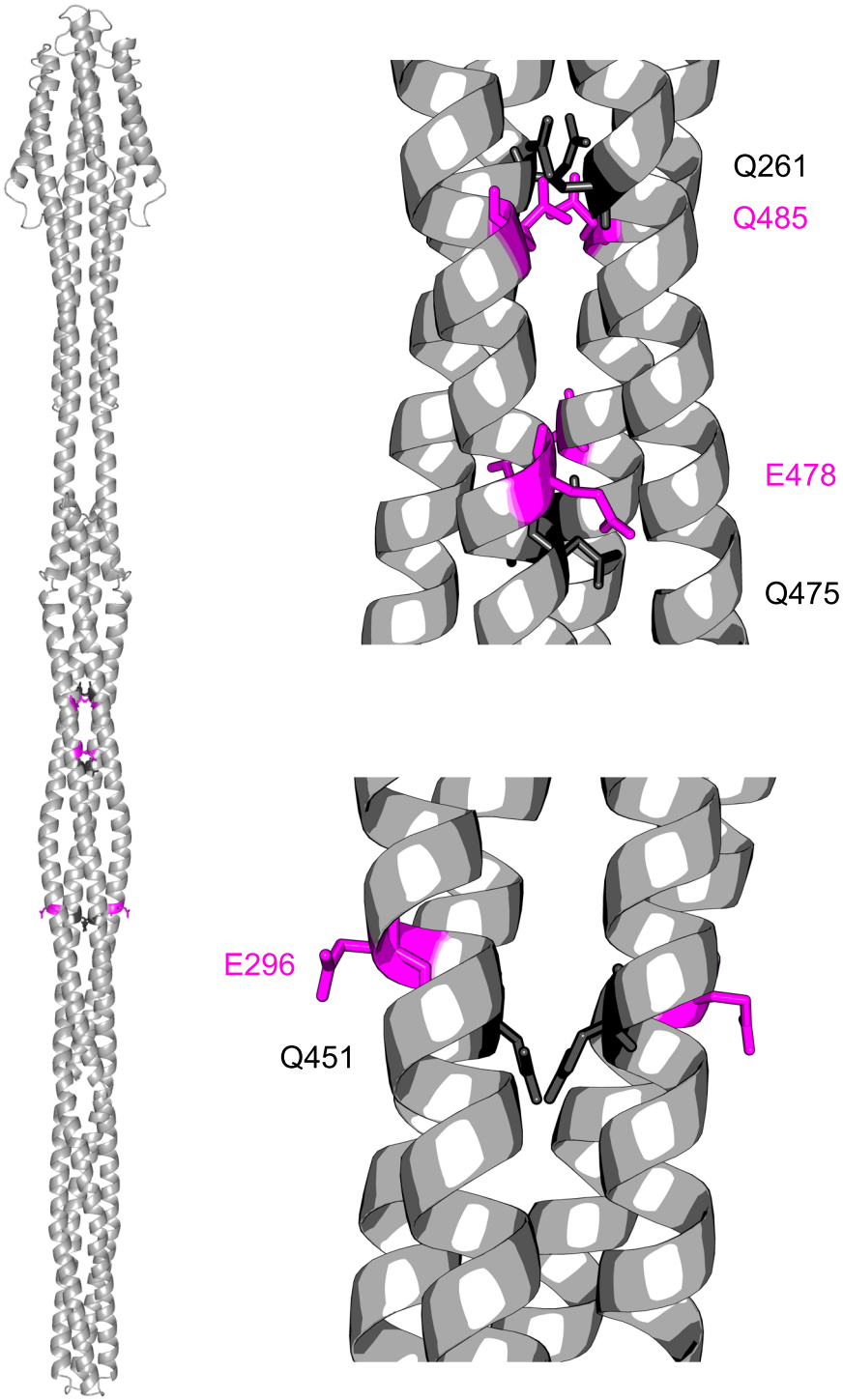


Figure 7.23: Position of predicted (grey) and identified (magenta) **QE** modification sites on the theoretical TlpT structure. Areas of interest are zoomed and rotated to visualise sidechain orientation.

There are two main limitations to this method. Firstly, the method is much more successful at recording deamidation than methylation. Only 10% of the Tsr E304 peptides showed modification. Tsr Q297 and Q311 both recorded high levels of modification, but only one peptide with both deamidation and methylation was recorded in each. In TlpT, no **Q** methylation was recorded. Due to the time taken between challenging cells with a nutrient/chemoeffector increase and spinning and freezing cells, it is possible that some methylation is lost. High reactivity sites are the most likely to be affected by this. When dealing with very small cell volumes, adding ice-cold trichloro-acetic acid to cells before centrifuging would dissolve membranes and halt any further demethylation. However, at the cell volumes used here, the quantities of trichloro-acetic acid required make this impossible.

The results for TlpT show, however, that this limitation does not prevent the method from generating useful information. In particular, sites can be identified without anything being known about the surrounding consensus sequence. If at least a subset of all adaptation sites in a new chemoreceptor or species can be identified, consensus sequences can be adapted from these sites and further potential sites probed using mutation work as in Chapter 5.

If this method is extended to the *R. sphaeroides* polar chemoreceptors, it is likely that higher levels of methylation will survive processing. The propionate challenge given to fully methylate all of the chemoreceptors is an external challenge. If the cytoplasmic cluster responds to internal metabolite levels, perhaps to set the baseline stopping frequency, as was suggested in Chapter 3, then this challenge is unlikely to affect cytoplasmic cluster methylation levels directly. If this is true, then the methylation seen here is a steady state methylation level in TlpT, which may be indirectly caused by metabolising the added propionate.

The second limitation is that only one chemoreceptor is analysed at a time and this receptor must have been identified as such by some other method and be amenable to genetic manipulation. Expressing one tagged chemoreceptor from a plasmid en-

sures that sufficient protein is collected that modifications will survive processing at significant enough levels to be measured. There is also no ambiguity as to which chemoreceptor is modified, a real danger considering the high degree of homology between chemoreceptors.

A possible extension of this method would be to analyse whole cell extracts, allowing all chemoreceptors to be tested at once. This would remove the need to run extensive tests on a chemoreceptor that ultimately proves to not have any modification. A challenge to overcome would be the varying concentrations at which the different chemoreceptors are expressed endogenously, which may make it much more likely to see modifications in some chemoreceptors than in others.

Further, in a whole cell extract, the concentration of a particular chemoreceptor in comparison to the total concentration of the sample is low. Picking up a modification within this, itself only a percentage of the total concentration of that protein, becomes difficult. Some method to enrich for methylation would be required to help improve the concentration of the modification. The cell volumes required may be prohibitive.

7.4.2 Adaptation sites in TlpT

Three modification sites were identified in TlpT, E296, E478 and Q485. Of these, two were **E** sites that were found to be methylated and one a **Q** site that was found to be deamidated. In all cases, the identified sites showed modification on the **c** position in a heptad. However, only one of these sites, E296, is the second residue in a canonical **EQ/EQ** pair, as has been recorded in most studied chemoreceptors. E475 was found in a **DE** pair and Q485 was found in a **TQ** pair. The positions of these sites on the TlpT cytoplasmic domain and their relationships to the known sites in Tsr are shown in the alignment in Figure 7.24.

Of these sites, only one, E296, was predicted in Chapter 4. This site did not match ei-

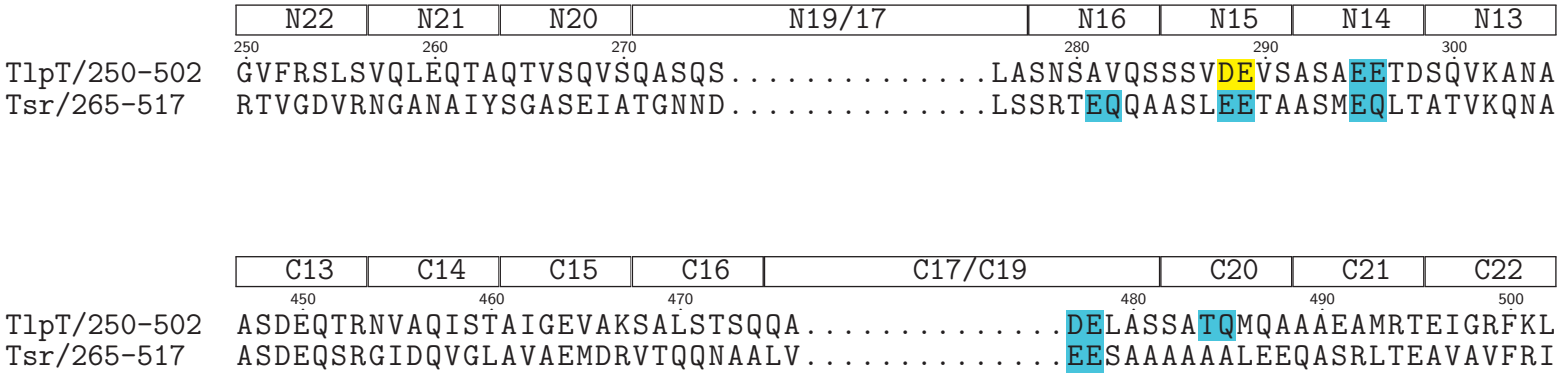


Figure 7.24: The TlpT and Tsr methylation domains (heptads 13-22) are aligned according to the spacing in the 36H chemoreceptor MSA from Alexander and Zhulin (2007). The positions of the methylation and deamidation sites identified using this *in vivo* challenge and MS/MS method are shown in blue. In all cases, the modification was recorded on the second residue in the pair. The identified TlpT sites suggest a consensus sequence, which matches to one additional site in TlpT, highlighted in yellow. The top ruler numbering indicates the TlpT residue numbering.

ther the *E. coli* consensus sequence or the more general Alexander and Zhulin (2007) consensus sequence, as residue Q300 should be a **T** or **ASTG** respectively to match. It was predicted using the prediction sequence Up, which is a relaxed version of the upstream half of the Alexander and Zhulin (2007) consensus sequence, **[ASTG]-[ASTG]-X-X-EQ-EQ**, and confirmed as it is in frame with the Tsr methylation site Q311 on the 36H chemoreceptor MSA.

There are two other methylation sites upstream of E311 in Tsr: Q297 and E304. The in-frame matches to these sites on TlpT may also be methylation sites. However, Tsr Q297 matches TlpT V282 in the alignment, which cannot be methylated. Tsr E304 matches TlpT E289. This site is part of a **DE**, not the usual **EQ/EQ** pair. Unlike canonical methylation sites, the identified TlpT methylation site, E478, is itself part of a **DE** pair and is also in frame with an *E. coli* site, E493. E289 may therefore also be a methylation site, despite its **DE** identity.

No modifications were recorded at E289. However, its in-frame partner on Tsr, E304, was the site which was recorded with the lowest percentage of modified peptides. It is likely that the two sites have similar reactivities and either were not methylated appreciably with the addition of attractant to the cell culture, or were too quickly demethylated during the cell collection process, or were more susceptible to loss of methylation during processing.

Q485 was identified in the deamidated state but not the methylated state. This site is not in frame with any Tsr modification site nor any predicted site in the 36H chemoreceptor MSA. It is in frame with a predicted position on a small number of sequences in the 34H chemoreceptor set (Figure 4.3). It lies seven residues downstream of the identified site E478, correctly in-frame for modification by the adaptation pathway. This site matches the general consensus sequence at all positions bar one: the modified site is in a **TQ** pair rather than an **EQ/EQ** pair.

The three identified sites can be used to build a *R. sphaeroides* cytoplasmic cluster-

specific consensus sequence. Table 7.3 shows the environment surrounding each identified site and how this differs from the Alexander and Zhulin (2007) predicted environment. Two consensus sequences can be built. The first is a specific consensus sequence for TlpT, including only the variability in the three identified sites. The second is a general consensus sequence for all chemoreceptors, including the variability in the TlpT sites in the Alexander and Zhulin (2007) sequence.

Table 7.3: The environment surrounding the identified TlpT modification sites.

	1	2	3	4	5	6	7	8	9	10
	e	f	g	a	b	c	d	e	f	g
E296	S	A	S	A	E	E	T	D	S	Q
E478	S	Q	Q	A	D	E	L	A	S	S
Q485	A	S	S	A	T	Q	M	Q	A	A
TlpT specific	AS	SAQ	SQ	A	DET	EQ	LTM	DAQ	AS	SQA
AZ	ASTG	ASTG	X	X	EQ	EQ	X	X	ASTG	ASTG
AZ with TlpT	ASTG	ASTGQ	X	X	EQDTCN	EQ	X	X	ASTG	ASTGQ

TlpT sites introduce additional variability into the Alexander and Zhulin (2007) consensus at three positions, 2, 5 and 10. Position 5 is the first residue of the **QE/QE** pair and is relaxed to include residues **D** and **T**. As this position can therefore take an uncharged or negatively charged residue, it is possible that additional variability can be allowed at this site, including the polar uncharged or negatively charged residues (**C**, **D** and **N**), but not the aromatic polar uncharged residues as these are likely to disrupt the helix.

Position 2 and 10 is relaxed to include **Q**. These positions occupy the same face of the helix as the modified residues. They are therefore either neutral (**A**, **G**) on the hydrophobicity scale or are hydrophilic (**S**, **T**) and are small so as not to interfere with the modified residue's protruding sidechain. **Q** is polar, but is not small, suggesting a slightly different CheR active site to *E. coli*. Additional variability is not obvious for this site, as the added residue is so different from those in positions 2 and 10 in the original sequence.

The expanded consensus sequence predicts one additional methylation site in the TlpT cytoplasmic domain to the three identified sites: E289. The three identified

sites and additional predicted site together follow the predicted positioning pattern of 34H receptors (Alexander and Zhulin, 2007), two sites on each helix arm.

7.4.3 Comparing results to phenotype data

As only E296 was predicted, it is the only identified site which has been mutated and phenotyped (Chapter 5). Unfortunately, only **D** and **Q** mutants were tested, as attempts to create the **A** mutant were unsuccessful.

E296Q showed wildtype free swimming and tethering phenotypes, suggesting that this site was post-translationally deamidated back to **E**. This suggests that a more specific consensus sequence is not needed for sites that require post-translational deamidation before they can function as methylation sites. CheB specificity for these sites is driven by the same interactions as for later demethylation.

E296D also had wildtype phenotypes. This mutation sets the receptor to 'permanently demethylated and unmethylatable'. From *E. coli* results, it had been predicted that a **D** mutant of a methylation site would be analogous to a decrease in (rather than complete loss of) CheR, giving smoother free swimming, less sensitivity to a drop in chemoeffector, so a higher percentage of unresponsive cells, and a shorter adaptation time for those that do respond. Similarly, an **A** mutant is expected to show stoppier free swimming, greater sensitivity to a drop in chemoeffectors and a longer adaptation time.

It is possible that these predicted effects are very subtle when only one methylation site is mutated. Mutating all known sites to **A** or **D** may give a clearer phenotype.

Within the α-helix, a mutation to **A** is usually a silent mutation, as this does not disrupt structure, unless the site mutated is one involved in some kind of signalling, like the methylation sites. The phenotype results for the other sites mutated follows this, as Q261A and Q451A both gave wildtype phenotypes. Mutating these sites to either **D** or **E** gave a smooth swimming, unresponsive phenotype, suggesting that

these mutations disrupted structure sufficiently to prevent signalling.

However, the mutation Q475A gave a stoppy, slow-adaptive phenotype similar to the phenotype of $\Delta CheB_2$. This site lies directly upstream of E478. Consensus sequences suggest that this mutation should be neutral, as this position in relation to the methylated site should be able to take any residue. However, changing this position may have influenced helix packing sufficiently to change the position of the E478 sidechain. If this is true, then the mutation E478A should give as dramatic a result. It is also possible that the slight shift in helix packing also affected the next methylation site, Q485, which would explain the magnitude of the phenotype change in Q475A.

Chapter 8

General conclusions

R. sphaeroides differs from other bacterial models of chemotaxis as it has multiple copies of the key chemotaxis proteins localised into two distinct clusters. The primary aim of this work was to discover why both the polar and cytoplasmic clusters are required for normal chemotaxis. A systematic analysis of the swimming behaviour of a library of deletion mutants suggests a model for *R. sphaeroides* chemotaxis with distinct roles for the two clusters.

Further exploration of the role of the cytoplasmic cluster through *in vivo* analysis and tandem mass spectrometry of a selected chemoreceptor provides the first direct evidence for chemoreceptor methylation by adaptation proteins in *R. sphaeroides*. The consensus sequence for these methylation sites is contrasted with those seen in other bacterial species.

8.1 *R. sphaeroides* motility under dynamic and steady state conditions

Prior to this study, the majority of the *R. sphaeroides* chemotaxis mutants studied had been recorded with wildtype swimming at steady state in a homogenous environment, but with differences from wildtype during the dynamic changes of chemo-

taxis similar to those seen in *E. coli* (for example, Porter *et al.* (2006)). A systematic investigation of adaptation deletion mutants and related mutant strains revealed that *R. sphaeroides* chemotaxis mutants do indeed exhibit a range of free swimming phenotypes, similar to *E. coli*.

8.1.1 Steady state swimming

Through the use of very simple technology (rather than a more complex assay such as optical traps) applied to whole populations, strains could be classified as smoother or stoppier than wildtype in a homogenous environment.

Strains can be classified by manual inspection of tracks, by summary statistics of the raw tracks, or by first categorising each track into runs and stops and using summary statistics of this data set to compare strains. In this study, smooth swimmers were identified by manual inspection of tracks and by using the proportion of time spent stopped as a summary statistic of the classified data set. Stoppy swimmers were difficult to categorise objectively by eye, but could easily be classified using the proportion of time spent stopped. For the first time, statistical differences between swimming behaviour in wildtype and deletion mutants were recorded.

Tracks were classified into runs and stops using a novel system that does not rely on heuristics. Other systems for whole populations of swimming cells require heuristics such as a cut-off in swimming speed or angle change to define a stop. Usually, these cut-offs are based on what is known about wildtype swimming, which introduces a bias when analysing strains which deviate from wildtype. Prior judgement on what is 'wildtype' will affect how mutated strains are classified. This system uses data sets of known stopped and smooth tracks to determine the probability that a particular framewise transition is a run or a stop. Heuristics based on wildtype swimming or any knowledge of a particular mutant's behaviour are not required.

Wildtype swimming behaviour had been analysed using this system, but it had not

yet been extended to classify tracks from swimming mutants or used to differentiate mutant strains from wildtype. Although the classification of runs and stops is noisy and data sets have a number of obviously anomalous tracks, (alternating between very short runs and stops), the method appears to be robust against this noise. The two strains identified in the literature as differing from wildtype, the $CheB_2$ and $CheR_3$ deletion mutants, were successfully classified as stoppier and smoother than wildtype respectively. Classification of other mutants agreed with by-eye classification. This is a marked improvement on previous attempts to classify free swimming behaviour in *R. sphaeroides*, as all obvious by-eye differences were then lost in analysis noise.

8.1.2 Dynamic swimming behaviour

Cell-to-cell variability under dynamic swimming conditions during the chemotaxis response had largely been ignored. Extending an existing tethering assay to analyse large numbers of cells revealed the extent of population variability, especially under photoheterotrophic growth conditions. Tethered cells' responses to changes in the concentration of a chemoeffector in the external environment were classified by eye. Cells were classified as unresponsive, responsive and adaptive, which are phenotypes that have been used previously. Cells' swimming was also described as stoppy or wildtype, and the additional phenotype 'inhibited' was defined. These new classifications proved crucial for describing the differences between strains.

New phenotypes were described by combining what is seen under steady state and dynamic swimming. The three clear phenotypes are smooth swimming unresponsive (e.g. the $CheB_1$ or $CheR_3$ deletion mutants), stoppy unresponsive (e.g. the $CheR_2$ deletion mutant) and stoppy slowly adaptive/inhibited (the $CheB_2$ deletion mutant). The separation of the unresponsive population into stoppy and smooth swimming subpopulations is a novel one in *R. sphaeroides*.

In *E. coli*, cells with mutations in the adaptation pathway have one of two phenotypes: tumbly responsive (CheB deletion) and smooth swimming responsive (CheR deletion), both of which are unable to adapt. The system which sets the steady state tumbling frequency and the system that allows adaptation to a dynamic change are the same. In *R. sphaeroides*, the control over these two systems must be more complex, as the steady state stopping frequency can be altered in one direction, giving a stoppier swimmer, yet affect the cell's response to dynamic changes either by increased or decreased sensitivity.

Thus far, uncoupling the two elements of swimming behaviour has only been achieved through deletion of a protein from the chemotaxis network, and has disrupted the system completely to one extreme or another (smooth or stoppy). Although the majority of sites mutated in TlpT proved unlikely to be methylation sites, mutation of these sites did cause sufficient structural or functional changes to alter the ratio of unresponsive and adaptive. These mutations also resulted in changes to the adaptation time, although the majority were not significant changes from wildtype.

8.2 Methylation sites in TlpT

Three sites in TlpT that are modified by the adaptation pathway were identified by tandem mass spectrometry, using a method with a low rate of false positives (Chapter 7). Sites E296 and E478 were methylated, whereas Q485 was only recorded in the deamidated state. The environments surrounding these sites suggested that the current consensus sequence used to predict methylation sites (Alexander and Zhulin, 2007) is too restrictive for TlpT. Most notably, two of the identified sites were not part of an **EQ/EQ** pair, but rather a **DETQ/EQ** pair.

Using this information and the relative position of the identified sites compared to the known sites in Tsr, an additional adaptation site, E289, is predicted. To test this site, it should be mutated to **A** and **D** and the resulting dynamic and steady state

swimming behaviour analysed. To confirm that this is a valid approach for testing adaptation sites, all of the identified sites should be similarly mutated, singly and in combination.

An extended bioinformatics analysis of all soluble/cytoplasmic chemoreceptors would also be interesting. Does the TlpT methylation site consensus sequence find many matches in other soluble chemoreceptors, or is it unique?

8.3 Working model for *R. sphaeroides* chemotaxis

The data presented in Chapter 3 suggest that the cytoplasmic cluster produces the majority of the $CheY_6$-P in the cell, as mutations that result in that cluster not activating result in smooth swimming. Even if the polar cluster does produce $CheY_6$-P in response to a drop in chemoeffector, that $CheY_6$-P is insufficient to produce a stop if the cytoplasmic cluster is not functioning, as $CheA_3A_4$ mutants are unresponsive. $CheY_{3,4}$-P and $CheY_6$-P appear to have antagonistic effects on the flagellar motor, either by directly binding or by $CheY_{3,4}$-P preventing the cytoplasmic cluster from signalling in some way.

In Figure 8.1, the working hypothesis for network connectivity during aerobic growth developed in this thesis is presented. $CheY_3$ and $CheY_4$ are assumed to interact directly with the flagellar motor, as this agrees with previous *in vitro* work (Ferre *et al.*, 2004). This is only one interpretation of the data, but does explain the majority of the results.

The proposed model of chemotaxis and signal integration between the two clusters is summarised in Figure 8.2. This figure explores the initial case, assuming that the polar cluster does not produce $CheY_6$-P at physiologically relevant rates. Additional complexities are discussed below.

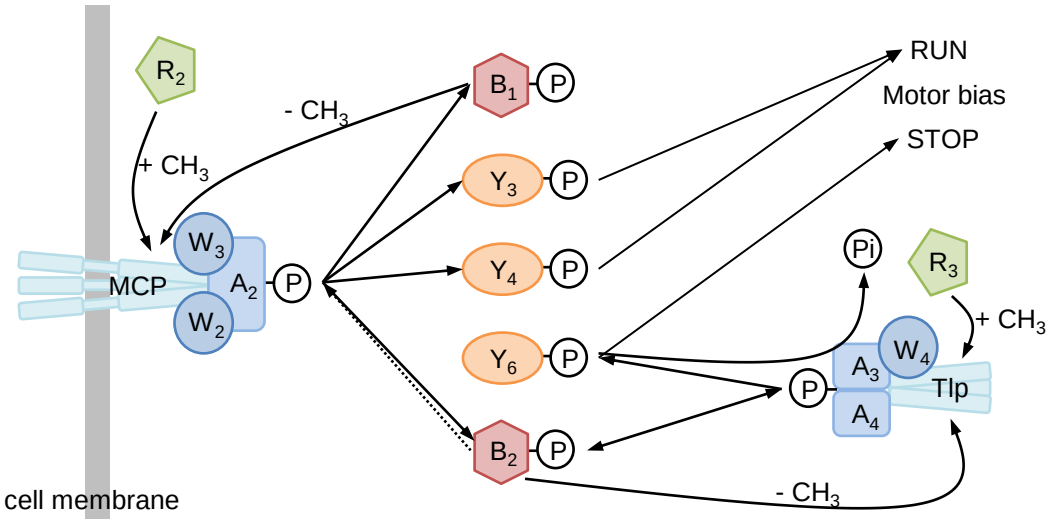


Figure 8.1: Proposed network connectivity in *R. sphaeroides* chemotaxis. In this model, the polar cluster does not phosphorylate $CheY_6$ at physiologically relevant levels. $CheY_{3,4}$-P biases the motor towards runs, whereas $CheY_6$-P biases the motor towards stops.

8.3.1 Role of the cytoplasmic cluster

In the initial working model (Figures 8.1 and 8.2), the cytoplasmic cluster controls the steady state stopping frequency, using a modified version of the canonical *E. coli* chemotaxis system. The cytoplasmic chemoreceptors are made more sensitive to their internal chemoeffectors, which are likely to be metabolites, by the action of the methyltransferase $CheR_3$ and less sensitive by the action of the methylesterase $CheB_2$. When chemoreceptors are activated, the heterodimer $CheA_3CheA_4$ phosphorylates $CheY_6$. $CheY_6$-P binds to the flagellum and initiates a stop. $CheY_6$ autodephosphorylation and $CheA_3$ phosphatase activity dephosphorylates $CheY_6$-P.

In this model, the cytoplasmic cluster does not respond to external chemoeffector changes like the *E. coli* chemotaxis cluster. Drops in external chemoeffector concentration do not cause a flood of $CheY_6$-P to initiate a stop. Instead, the level of $CheY_6$-P determines the threshold of $CheY_{3,4}$-P required to initiate a run or stop (Figure 8.2). The level of $CheY_6$-P is tuned by the antagonistic action of $CheB_2$ and

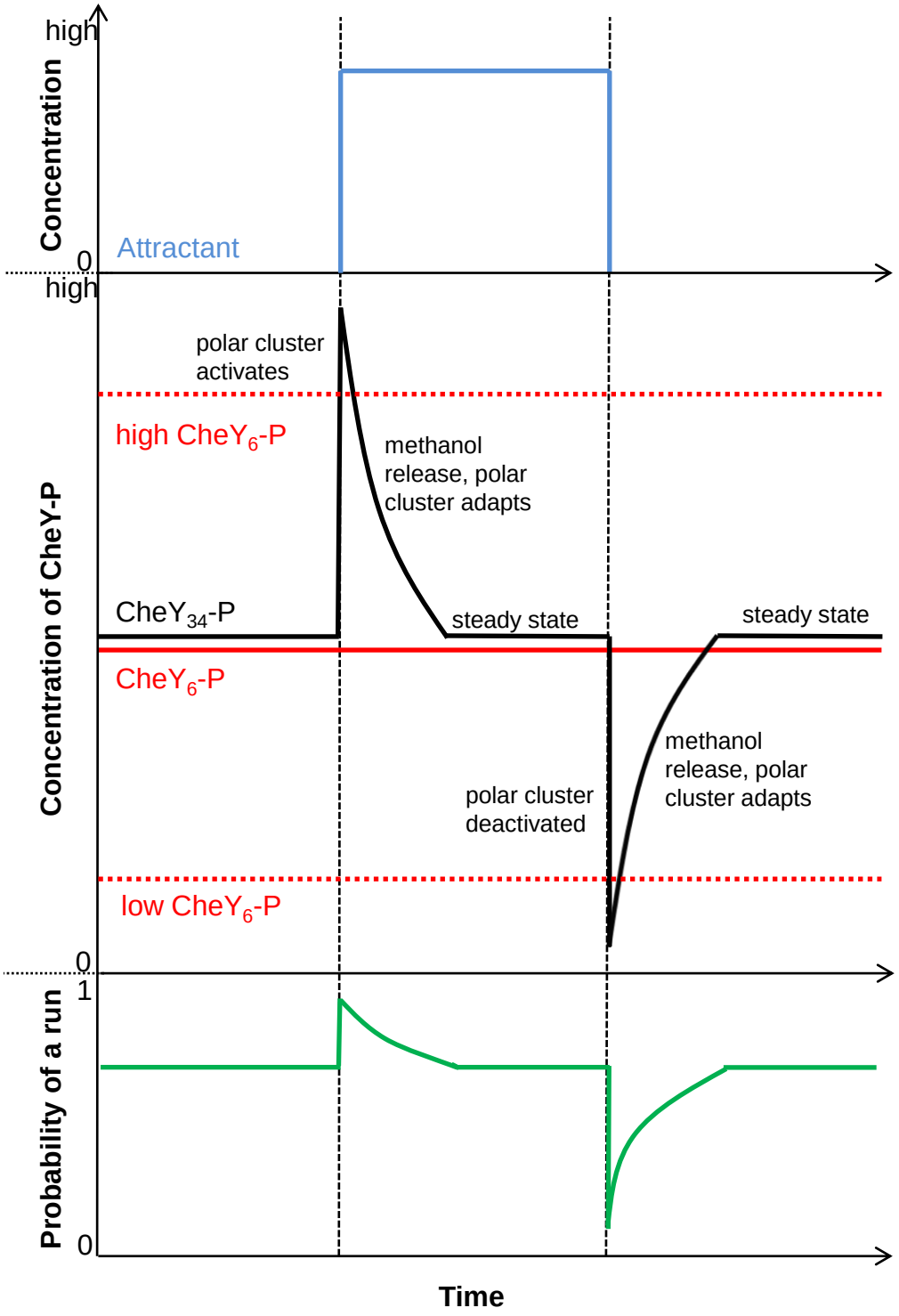


Figure 8.2: Proposed model of chemotaxis in *R. sphaeroides*. At steady state, the concentration of $CheY_{3,4}$-P and $CheY_6$-P (middle panel) is tuned to give the basal probability of a run (bottom panel). An increase in attractant (top panel) activates the polar cluster and increases $CheY_{3,4}$-P, ensuring a run. Adaptation to the stimulus returns the system to steady state. Similarly, a decrease in attractant causes a drop on $CheY_{3,4}$-P levels and a stop. If the concentration of $CheY_6$-P is pushed up (middle panel, top dashed red line), a higher concentration of $CheY_{3,4}$-P is needed to effect a run. If the concentration of $CheY_6$-P is pushed down (middle panel, bottom dashed red line), a lower concentration of $CheY_{3,4}$-P is needed to produce a run.

$CheR_3$, along with the phosphatase action of $CheA_3$. Thus, the deletion of $CheB_2$ results in high $CheY_6$-P, causing the inhibited phenotype (Figure 8.2, middle panel, top dashed red line).

Previous work has shown that under photoheterotrophic conditions, the cytoplasmic cluster can respond and adapt to changes in external chemoeffectors (Martin *et al.*, 2001a). This suggests an additional level of complexity to the model, whereby runs are produced by the polar cluster and stops by the cytoplasmic cluster (Figure 8.3). However, this does not agree with all of the obtained data. As can be seen in the middle panel, if there are high basal levels of $CheY_6$-P, as in the $CheR_2$ deletion mutant, a response to a drop in chemoeffector should still be possible. This was not the case under aerobic conditions (Chapter 3). If this alternative model is correct, it is valid only under photoheterotrophic conditions. An extensive study of the *R. sphaeroides* chemotaxis mutants under photoheterotrophic conditions would give more clarity.

8.3.2 Role of the polar cluster

The working model presented here has the polar cluster controlling dynamic changes in a similar manner to the *B. subtilis* chemotaxis system. The polar chemoreceptors monitor the external environment, including sugar and amino acid concentrations (Thompson, 2005). When chemoeffector levels increase, the chemoreceptors activate $CheA_2$, which autophosphorylates and then phosphorylates $CheY_{3,4}$. $CheY_{3,4}$-P overcomes the threshold of $CheY_6$-P to bias the flagellar motor towards smooth swimming (Figure 8.2). When chemoeffector levels decrease, the polar cluster ceases signalling and $CheY_{3,4}$-P levels drop below steady state levels, allowing $CheY_6$-P to effect a stop of approximately 50 s under aerobic conditions and 80 s under photoheterotrophic conditions in wildtype.

The ability to respond to dynamic changes was lost if the polar cluster was un-

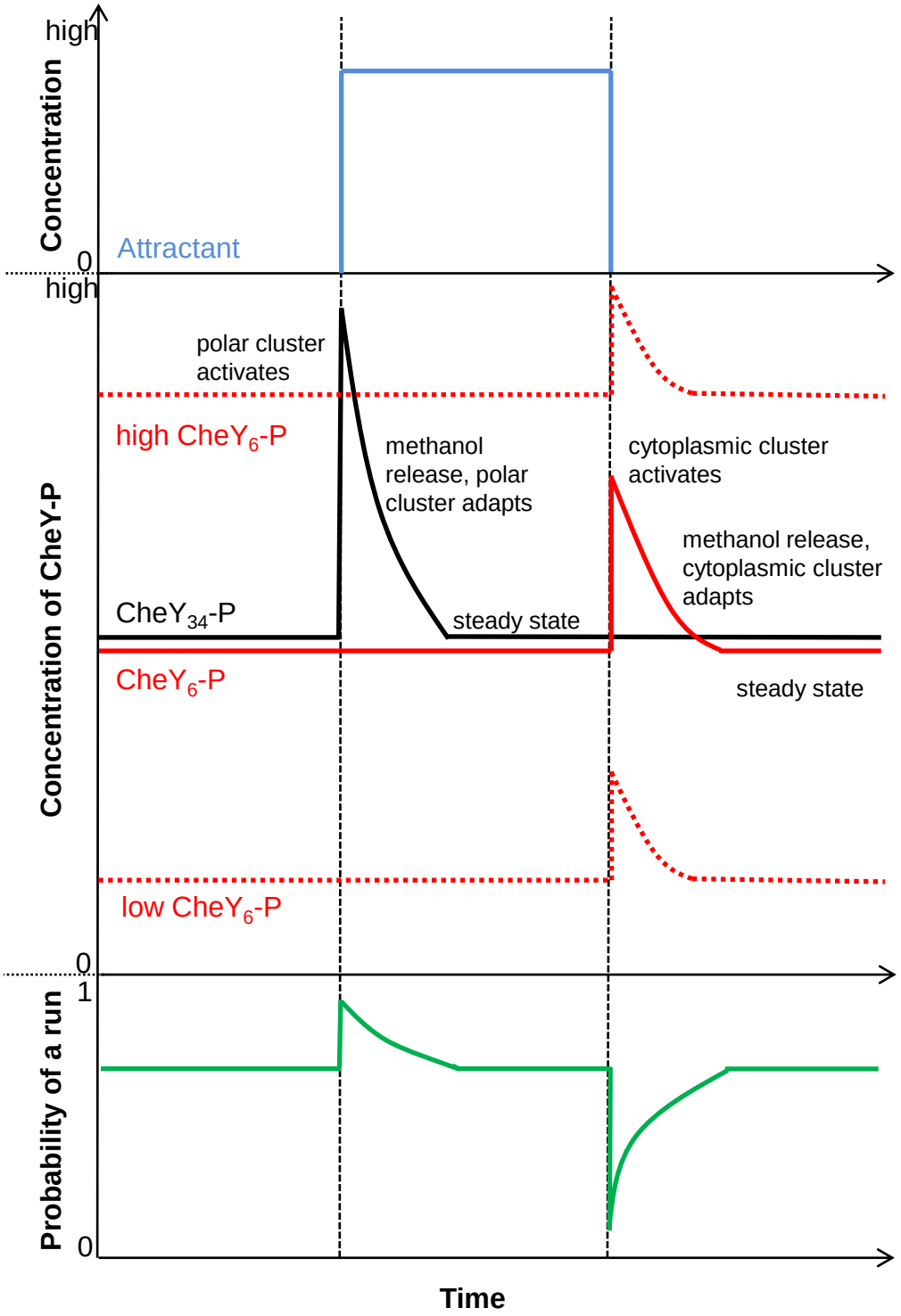


Figure 8.3: Alternative model of chemotaxis in *R. sphaeroides*, allowing the cytoplasmic cluster to respond to a drop in external chemoeffector. At steady state, the concentration of $CheY_{3,4}$-P and $CheY_6$-P (middle panel) is tuned to give the basal probability of a run (bottom panel). An increase in attractant (top panel) activates the polar cluster and increases $CheY_{3,4}$-P, ensuring a run. Adaptation to the stimulus returns the system to steady state. Similarly, a decrease in attractant causes an increase in $CheY_6$-P levels and a stop. If the concentration of $CheY_6$-P is pushed up (middle panel, top dashed red line), a higher concentration of $CheY_{3,4}$-P is needed to effect a run. If the concentration of $CheY_6$-P is pushed down (middle panel, bottom dashed red line), a lower concentration of $CheY_{3,4}$-P is needed to produce a run.

able signal (Chapter 3). For example, inactivation or deletion of $CheY_3$ and $CheY_4$ leads to a stoppy swimmer that cannot respond to dynamic changes, whereas mutations that mimic phosphorylation also result in unresponsive cells, but with smooth swimming. However, even if the polar cluster can respond and adapt to external chemoeffector concentrations, if the cytoplasmic cluster does not also produce $CheY_6$-P, the ability to stop is lost. This explains previous experimental results (deletion of TlpC) that the loss of the cytoplasmic cluster does not affect methanol release on attractant addition or removal, but does abolish chemotaxis (Wadhams *et al.*, 2002): response and adaptation occurs at the polar cluster with concurrent methanol release, but that signal cannot reach the flagellar motor without $CheY_6$-P produced by the cytoplasmic cluster.

The model proposed is a hybrid of the *E. coli* and *B. subtilis* systems. Both addition and removal of attractant results in the release of methanol, suggesting that methylation and demethylation both occur at both steps, as happens in *B. subtilis*. Deletion of the polar cluster adaptation proteins $CheB_1$ and $CheR_2$ results in loss of ability to respond to dynamic changes, which resembles *E. coli* more than *B. subtilis*. The steady state swimming behaviour after these deletions is the reverse of that in *E. coli*, substantiating the idea that the polar cluster operates in the reverse of the *E. coli* cluster, similarly to the *B. subtilis* cluster.

The model requires dephosphorylation of $CheY_{3,4}$ at rates much higher than measured autodephosphorylation. This suggests that a protein with phosphatase activity is associated with or found near to the polar cluster. The identification of such a protein will support the proposed model. The model does not require the existence of the reverse phosphorelay that *in vitro* allows the cytoplasmic cluster kinase to phosphorylate the polar cluster kinase via $CheB_2$.

The proposed model in Figures 8.1 and 8.2 is the current best explanation of the results obtained for *R. sphaeroides* chemotaxis mutants under aerobic conditions. The model does not address the polar cluster's production of $CheY_6$-P, as all of the ob-

tained results suggested that the production of $CheY_{3,4}$-P through this cluster was the physiologically relevant event. Additional work as discussed at the end of Chapter 3 will provide further clarity.

Although the *E. coli* chemotaxis system is well-understood, the very simplicity that made its elucidation possible is its limitation for further application. The majority of chemotactic bacteria have multiple homologues of at least one of the key chemotaxis proteins. Determining network connectivity and whether protein duplications are redundancies or requirements is challenging and requires much more information than can be gleaned from the simple *E. coli* system. With its two chemotaxis clusters and multiple response regulators, the *R. sphaeroides* chemotaxis system is ideal for exploring the complexities of a system with multiple homologues. The systematic analysis of steady state and dynamic swimming behaviour used here has suggested how this complex system functions.

Appendix A

PCR primers used in the study

Restriction enzyme sites in bold

Primers for TlpT mutations (Chapter 5)

TlpT-Q261Q296-F1	GTCATCG**ACCGGT**TCCGC
TlpT-Q261Q296-R2	CAGCTCC**GACGTC**TCGCGC
TlpT-Q261A-R1	CGGTCTGCGCGGTCGCTTCGAGCTGGACCG
TlpT-Q261A-F2	CGGTCCAGCTCGAAGCGACCGCGCAGACCG
TlpT-Q261D-R1	CGGTCTGCGCGGTGTCTTCGAGCTGGACCG
TlpT-Q261D-F2	CGGTCCAGCTCGAAGACACCGCGCAGACCG
TlpT-Q261E-R1	CGGTCTGCGCGGTCTCTTCGATGCGGACCG
TlpT-Q261E-F2	CGGTCCAGCTCGAAGAGACCGCGCAGACCG
TlpT-E296A-R1	CCTGGCTGTCGGTCGCTTCGGCCGAGGCC
TlpT-E296A-F2	GGCCTTGGCCGAAGCGACCGACAGCCAGG
TlpT-E296D-R1	CCTGGCTGTCGGTGTCTTCGGCCGAGGCC
TlpT-E296D-F2	GGCCTCGGCCGAAGACACCGACAGCCAGG
TlpT-E296Q-R1	CCTGGCTGTCGGTCTGTTCGGCCGAGGCC
TlpT-E296Q-F2	GGCCTCGGCCGAACAGACCGACAGCCAGG
TlpT-Q451Q475Q528-F1	GCGCGA**GACGTC**GGAGCTG
TlpT-Q451Q475Q528-R2	CGCAGG**CCTCAGG**CCCGG
TlpT-Q451D-R1	GCCACGTTGCGCGTGTCCTCGTCCGAGGCG
TlpT-Q451D-F2	CGCCTCGGACGAGGACACGCGCAACGTGGC
TlpT-Q451E-R1	GCCACGTTGCGCGTCTCCTCGTCCGAGGCG
TlpT-Q451E-F2	CGCCTCGGACGAGGAGACGCGCAACGTGGC
TlpT-Q475D-R1	CCAGCTCATCCGCGTCCTGGCTGGTCGAAAGC
TlpT-Q475D-F2	GCTTTCGACCAGCCAGGACGCGGATGAGCTGG

TlpT-Q475E-R1	CCAGCTCATCCGCCTCCTGGCTGGTCGAAAGC
TlpT-Q475E-F2	GCTTTCGACCAGCCAGGAGGCGGATGAGCTGG
TlpT-Q528D-R1	GGCCGCCACCATGTCCTGAAGCTGGGCCAGC
TlpT-Q528D-F2	GCTGGCCCAGCTTCAGGACATGGTGGCGGCC
TlpT-Q528E-R1	GGCCGCCACCATCTCCTGAAGCTGGGCCAGC
TlpT-Q528E-F2	GCTGGCCCAGCTTCAGGAGATGGTGGCGGCC

Primers for chemotaxis protein expression (Chapter 6)

B1c-F	AGAACT**GGATCC**CCGCCGCGCCCGGTG
B1c-R	AGTTCT**CTGCAG**TTATCATCGGACGGTCCTTTCCG
B2c-F	AGAACT**GAGCTC**GAGGACTGGATCCAC
B2c-R	AGTTCT**CTGCAG**TTATCAGTCGTTTGACGCGAGGC
ECCheB-F	AGAACT**GGATCC**GGCTCTATGAGCAAAATCAGG
ECCheBc-F	AGAACT**GGATCC**TTGAGTTCTGAAAAAC
ECCheB-R	AGTTCT**CTGCAG**TTATTAAATACGTATCGCCTGTC
ECCheR-F	AGAACT**GGATCC**GGCTCTATGACTTCATCTCTGC
ECCheR-R	AGTTCT**CTGCAG**TTATTAATCCTTACTTAGCGC
TsrCD-F	AGAACT**GGATCC**GGCTCTGAGAGTTTGCGCCATATGC
TsrCD-R	AGTTCT**CTGCAG**TTAGATCAACGACCCACGTAAAG

Primers for full length chemoreceptor expression (Chapter 7)

TlpT-F	GGTTA**CATGTT**AACACGCCGCCCCAAG
TlpT-R	GGTAA**CATGTG**GAAGTCGCCGAAGC
Tsr-F	CGAACG**CCATGG**GAATGTTAAAACGTATCAAAATTGTG
Tsr-R	ATATCG**AGATCT**AAATGTTTCCCAGTTC

Appendix B

Composition of media used in this study

B.1 Growth media

Ingredient	Concentration
Luria-Bertani broth (LB)	
Bacto-tryptone	10 g/l
NaCl	5 g/l
Yeast extract	5 g/l
2YT	
Bacto-tryptone	16 g/l
NaCl	5 g/l
Yeast extract	10 g/l
LB agar	
Agar	20 g/l
Bacto-tryptone	10 g/l
NaCl	5 g/l
Yeast extract	5 g/l
M22 Minimal medium	
Concentrated base	2% (v/v)
Growth factors	0.2% (v/v)

Ingredient	Concentration
NaCl	0.5 g/l
$(NH_4)_2SO_4$	0.5 g/l
Phosphate buffer 1 M pH 7.0	2% (v/v)
Adjust to pH 7.2 with KOH	
Succinate medium (Sux)	
Casamino acids	1 g/l
Concentrated base	2% (v/v)
Growth factors	0.2% (v/v)
NaCl	0.5 g/l
$(NH_4)_2SO_4$	0.5 g/l
Phosphate buffer 1 M pH 7.0	2% (v/v)
Sodium succinate	2.0 g/l
Adjust to pH 7.2 with KOH	
1 M phosphate buffer pH 7.0	
K_2HPO_4	174.18 g/l
KH_2PO_4	136.09 g/l
Adjust to pH 7.0 with KOH	
Concentrated base	
Ammonium molybdate(.$4H_2O$)	4.6 mg/l
$CaCl_2.6H_2O$	2.5 g/l
$FeSO_4.7H_2O$	50 mg/l
Metals 44 solution	2.5% (v/v)
$MgSO_4.7H_2O$	14.5 g/l
Nitrilotriaccetic acid (Na salt)	5.94 g/l
Adjust to pH 6.8	
Metals 44 solution	
EDTA	2.5 g/l
$CoCl_2.6H_2O$	200 mg/l
$CuSO_4.5H_2On$	390 mg/l
$FeSO_4.7H_2O$	5 g/l
H_3BO_3	120 mg/l
$MnSO_4.7H_2O$	2 g/l

Ingredient	Concentration
Sulphuric acid 3 M	0.15% (v/v)
$ZnSO_4.7H_2O$	11 g/l
Growth factors	
Biotin	20 mg/l
$NaHCO_3$	500 mg/l
Niacin (Nicotinic acid)	1 g/l
Thiamine HCl	500 mg/l
PIPES	
Piperazine-N,N′-bis(2-ethanesulfonic acid)	10 mM (3.02 g/l)
Adjust to pH 7.2 with HCl	
PBS	
Phosphate buffered saline (Dulbecco A) tablets	10 in 1l

B.2 Buffers and solutions

Ingredient	Concentration
Lysis buffer (2.2.1)	
EDTA	1 mM
SDS	1%
Tris (pH 8.0)	10 mM
Preheated to 65^oC	
5x DNA loading dye (2.2.2)	
Bromophenol blue	0.25 % (w/v)
Glycerol	70 % (w/v)
TRIS base	30 % (w/v)
Xylene cyanol	0.25 % (w/v)
10x TBE (2.2.2)	
Boric acid	55 $g.l^{-1}$
TRIS base	108 $g.l^{-1}$

Ingredient	Concentration
EDTA	7.4 $g.l^{-1}$
pH 8.3	
TFB I (2.2.3)	
$CaCl_2.2\ H_2O$	10 mM
Glycerol	15 % (v/v)
KCH_3COO	30 mM
$MnCl_2.\ 4\ H_2O$	50 mM
RbCl	100 mM
Adjust to pH 5.8 with acetic acid	
TFB II (2.2.3)	
$CaCl_2.2\ H_2O$	75 mM
Glycerol	15 % (v/v)
PIPES	10 mM
RbCl	10 mM
Adjust to pH 6.8 with HCl	
Denaturation buffer(2.2.4)	
NaCl	0.15 M
NaOH	0.5 M
Neutralisation buffer(2.2.4)	
NaCl	1.5 M
Tris pH 7.5	1 M
20x SSC (2.2.4)	
NaCl	175.3 g/l
Sodium citrate	88.2 g/l
Adjust to pH 7 with HCl	
Hybridisation solution (2.2.4)	
Blocking reagent solution	1 %
N-laurylsarcosine	0.1% (v/v)
SDS	0.02% (w/v)
SSC	5x

Ingredient	Concentration
Blocking reagent supplied in Roche kit	
Wash A (2.2.4)	
SDS	0.1% (v/v)
SSC	2x
Wash B (2.2.4)	
SSC	0.5x
SDS	0.1%
Buffer 1, maleate buffer (2.2.4)	
Maleic acid	100 mM
NaCl	150 mM
Adjust to pH 7.5 with solid NaOH	
Buffer 2, blocking solution (2.2.4)	
Blocking reagent solution	1 %
Made up in buffer 1	
Blocking reagent supplied in Roche kit	
Buffer 3, detection buffer (2.2.4)	
NaCl	100mM
Tris-HCl	100 mM
Adjust to pH 9.5 with NaOH	
Protein buffer (2.3.3)	
DTT	1 mM (if protein has Cys)
Glycerol	10%
Imidazole	10 mM
NaCl	150 mM
Tris-HCl	50 mM
Adjust to pH 8 with NaOH	
Native elution buffer (2.3.3)	
DTT	1 mM (if protein has Cys)
Glycerol	10%

Ingredient	Concentration
Imidazole	500 mM
NaCl	150 mM
Tris-HCl	50 mM
Adjust to pH 8 with HCl	

Denaturing protein buffers B-E (2.3.4)	
Imidazole	10 mM
NaH_2PO_4	100 mM
Tris-HCl	10 mM
Urea	8 M
TWEEN20	2%
Buffer B	adjust to pH 8.0 with NaOH
Buffer C	adjust to pH 6.3 with HCl
Buffer D	adjust to pH 5.9 with HCl
Buffer E	adjust to pH 4.5 with HCl

SDS protein loading dye (2.3.5)	
β-mercaptoethanol	750 μl
Bromophenol blue	75 mg
Glycerol	2.5 ml
MilliQ	1.75 ml
Sodium dodecyl sulphate	750 mg

SDS-PAGE running buffer (2.3.5)	
Glycine	14.41 g/l
SDS	0.1%
Tris-HCl	3.01 g/l

Blotting buffer (2.3.5)	
Glycine	14.41 g/l
Methanol	20% (v/v)
SDS	0.1%
Tris-HCl	3.01 g/l

Bibliography

C.A. Adase, R.R. Draheim, and M.D. Manson. (2012). The residue composition of the aromatic anchor of the second transmembrane helix determines the signaling properties of the aspartate/maltose chemoreceptor Tar of *Escherichia coli*. *Biochemistry*, **51**:1925–1932.

M.V. Airola, N. Sukomon, D. Samanta, P.P. Borbat, J.H. Freed, K.J. Watts, and B.R. Crane. (2013). HAMP domain conformers that propagate opposite signals in bacterial chemoreceptors. *PLoS Biology*, **11**(2):e1001479.

R.P. Alexander and I.B. Zhulin. (2007). Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. *Proceedings of the National Academy of Science USA*, **104**:2885–2890.

R.P. Alexander, A.C. Lowenthal, R.M. Harshey, and K.M. Ottemann. (2010). CheV: CheW-like coupling proteins at the core of the chemotaxis signaling network. *Trends in Microbiology*, **18**(11):494–503.

U. Alon, L. Camarena, M.G. Surette, B. Aguera y Arcas, Y. Liu, S. Leibler, and J.B. Stock. (1998). Response regulator output in bacterial chemotaxis. *The European Molecular Biology Organization Journal*, **17**(15):4238–4248.

G.S. Anand and A.M. Stock. (2002). Kinetic basis for the stimulatory effect of phosphorylation on the methylesterase activity of CheB. *Biochemistry*, **41**(21):6752–6760.

G.S. Anand, P.N. Goudreau, and A.M. Stock. (1998). Activation of methylesterase CheB: Evidence of a dual role for the regulatory domain. *Biochemistry*, **37**:14038–14047.

G.S. Anand, P.N. Goudreau, J.K. Lewis, and A.M. Stock. (2000). Evidence for phosphorylation-dependent conformational changes in methylesterase CheB. *Protein Science*, **9**:898–906.

J.L. Appleby and R.B. Bourret. (1998). Proposed signal transduction role for con-

served CheY residue Thr87, a member of the response regulator active-site quintet. *Journal of Bacteriology*, **180**(14):3563–3569.

J.L. Appleby and R.B. Bourret. (1999). Activation of CheY mutant D57N by phosphorylation at an alternative site, Ser-56. *Molecular Microbiology*, **34**(5):915–925.

J.P. Armitage and R.M. Macnab. (1987). Unidirectional intermittent rotation of the flagellum of *Rhodobacter sphaeroides*. *Journal of Bacteriology*, **169**:514–518.

J.P. Armitage, T.P. Pitta, M.A.S. Vigeant, H.L. Packer, and R.M. Ford. (1999). Transformations in flagellar structure of *Rhodobacter sphaeroides* and possible relationship to changes in swimming speed. *Journal of Bacteriology*, **181**(16):4825–4833.

A. Ashkin, J.M. Dziedzic, J.E. Bjorkholm, and S. Chu. (1986). Observation of a single-beam gradient force optical trap for dielectric particles. *Optics Letters*, **11** (5):288–290.

N.L. Bartelli and G.L. Hazelbauer. (2011). Direct evidence that the carboxyl-terminal sequence of a bacterial chemoreceptor is an unstructured linker and enzyme tether. *Protein Science*, **20**:1856–1866.

H.C. Berg. (2003). The rotary motor of bacterial flagella. *Annual Review of Biochemistry*, **72**:19–54.

H.C. Berg. (2008). Bacterial flagellar motor. *Current Biology*, **18**(16):R689–R691.

H.C. Berg and R.A. Anderson. (1973). Bacteria swim by rotating their flagellar filaments. *Nature*, **245**:380–382.

H.C. Berg and D.A. Brown. (1972). Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. *Nature*, **239**:500–504.

J.E. Berleman and C.E. Bauer. (2005)a. A Che-like signal transduction cascade involved in controlling flagella biosynthesis in *Rhodospirillum centenum*. *Molecular Microbiology*, **55**:1390–1402.

J.E. Berleman and C.E. Bauer. (2005)b. Involvement of a Che-like signal transduction cascade in regulating cyst cell development in *Rhodospirillum centenum*. *Molecular Microbiology*, **56**:1457–1466.

R.M. Berry and J.P. Armitage. (2000). Response kinetics of tethered *Rhodobacter sphaeroides* to changes in light intensity. *Biophysical Journal*, **78**:1207–1215.

M.J. Betts and R.B. Russell. (2003). Amino acid properties and consequences of substitutions. In Michael R. Barnes and Ian C. Gray, editors, *Bioinformatics for Geneticists*, chapter 14, pages 289–316. John Wiley & Sons, Ltd.

S.M. Block, J.M. Segall, and H.C. Berg. (1983). Adaptation kinetics in bacterial chemotaxis. *Journal of Bacteriology*, **154**:312–323.

K.A. Borkovich, N. Kaplan, J.F. Hess, and M.I. Simon. (1989). Transmembrane signal transduction in bacterial chemotaxis involves ligand-dependent activation of phosphate group transfer. *Proceedings of the National Academy of Sciences USA*, **86** (4):1208–1212.

R. B. Bourret, J. Davagnino, and M.I. Simon. (1993). The carboxy-terminal portion of the CheA kinase mediates regulation of autophosphorylation by transducer and CheW. *Journal of Bacteriology*, **175**:2097–2101.

A. Boyd and M.I. Simon. (1980). Multiple electrophoretic forms of methyl-accepting chemotaxis proteins generated by stimulus-elicited methylation in *Escherichia coli*. *Journal of Bacteriology*, **143**(2):809–815.

M.M. Bradford. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, **72**:248–254.

A. Briegel, X. Lib, A.M. Bilwes, K.T. Hughes, G.J. Jensen, and B.R. Crane. (2012). Bacterial chemoreceptor arrays are hexagonally packed trimers of receptor dimers networked by rings of kinase and coupling proteins. *Proceedings of the National Academy of Sciences USA*, **109**(10):3766–3771.

A. Briegel, D.R. Ortega, E.I. Tocheva, K. Wuichet, Z. Li, S. Chen, A. Mueller, C.V. Iancu, G.E. Murphy, M.J. Dobro, I.B. Zhulin, and G.J. Jensen. (2009). Universal architecture of bacterial chemoreceptor arrays. *Proceedings of the National Academy of Science USA*, **106**(40):17181–17186.

M.T. Brown. (2009). *Control of the Unidirectional Motor in Rhodobacter sphaeroides*. PhD thesis, Department of Biochemistry, University of Oxford.

M.T. Brown, N.J. Delalez, and J.P. Armitage. (2011). Protein dynamics and mechanisms controlling the rotational behaviour of the bacterial flagellar motor. *Current Opinion in Microbiology*, **14**:734–740.

L.L. Burrows. (2012). *Pseudomonas aeruginosa* twitching motility: Type IV pili in action. *Annual Review of Microbiology*, **66**:493–520.

D. Chelsky and F.W. Dahlquist. (1981). Multiple sites of methylation in the methyl accepting chemotaxis proteins of *Escherichia coli*. *Progress in Clinical and Biological Research*, **63**:371–381.

Y.-I. Chi, H. Yokota, and S.-H. Kim. (1997). Apo structure of the ligand-binding

domain of aspartate receptor from *Escherichia coli* and its comparison with ligand-bound or pseudoligand-bound structures. *FEBS Letters*, **414**:327–332.

P. Chomczynski and N. Sacchi. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, **162**:156–159.

M. Choudhary, Y.-X. Fu, C. Mackenzie, and S. Kaplan. (2004). DNA sequence duplication in *Rhodobacter sphaeroides* 2.4.1: Evidence of an ancient partnership between chromosomes I and II. *Journal of Bacteriology*, **186**(7):2019–2027.

S. Clarke and D.E. Koshland. (1979). Membrane receptors for aspartate and serine in bacterial chemotaxis. *Journal of Biological Chemistry*, **254**:9695–9702.

R.K. Clayton and B.J. Clayton. (1972). Relations between pigments and proteins in the photosynthetic membranes of *Rhodopseudomonas spheroides*. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **3**:492–504.

P. Cluzel, M. Surette, and S. Leibler. (2000). An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells. *Science*, **287**(5458):1652–1655.

A. Martinez del Campo, T. Ballado, J. de la Mora, S. Poggio, L. Camarena, and G. Dreyfus. (2007). Chemotactic control of the two flagellar systems of *Rhodobacter sphaeroides* is mediated by different sets of CheY and FliM proteins. *Journal of Bacteriology*, **189**(22):8397–8401.

A. Martnez del Campo, T. Ballado, L. Camarena, and G. Dreyfus. (2011). In *Rhodobacter sphaeroides*, chemotactic operon 1 regulates rotation of the flagellar system 2. *Journal of Bacteriology*, **193**(23):6781–6786.

S. Djordjevic and A.M. Stock. (1998). Chemotaxis receptor recognition by protein methyltransferase CheR. *Nature Structural Biology*, **5**:446–450.

J.J. Doyle and M.A. Luckow. (2003). The rest of the iceberg: legume diversity and evolution in a phylogenetic context. *Plant Physiology*, **131**(3):900–910.

R.R. Draheim, A.F. Bormans, R.-Z. Lai, and M.D. Manson. (2006). Tuning a bacterial chemoreceptor with protein-membrane interactions. *Biochemistry*, **45**:14655–14664.

C.M. Dyer and F.W. Dahlquist. (2006). Switched or not?: the structure of unphosphorylated CheY bound to the N terminus of FliM. *Journal of Bacteriology*, **188**: 7354–7363.

C.M. Dyer, A.S. Vartanian, H. Zhou, and F.W. Dahlquist. (2009). A molecular mechanism of bacterial flagellar motor switching. *Journal of Molecular Biology*, **388**:71–84.

P. Engstrom and G.L. Hazelbauer. (1980). Multiple methylation of methyl-accepting chemotaxis proteins during adaptation of *E. coli* to chemical stimuli. *Cell*, **20**:165–171.

J.J. Falke, R.B. Bass, S.L. Butler, S.A. Chervitz, and M.A. Danielson. (1997). The two-component sgnaling pathway of bacterial chemotaxis: A molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annual Review Of Cell Development Biology*, **13**:457–512.

T. Fenchel and R. Thar. (2004). *Candidatus* Ovobacter propellens: a large conspicuous prokaryote with an unusual motility behaviour. *FEMS Microbiology Ecology 48*, **48**: 231–238.

A. Ferre, J. de la Mora, T. Ballado, L. Camarena, and G. Dreyfus. (2004). Biochemical study of multiple CheY response regulators of the chemotactic pathway of *Rhodobacter sphaeroides*. *Journal of Bacteriology*, **186**(15):5172–5177.

R.D. Finn, J. Clements, and S.R. Eddy. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, **39**:W29–W37.

F. Garcia-Pichel. (1989). Rapid bacterial swimming measured in swarming cells of *Thiovulum majus*. *Journal of Bacteriology*, **171**(6):3560–3563.

G.D. Glekas, J.R. Cates, T.M. Cohen, C.V. Rao, and G.W. Ordal. (2011). Site-specific methylation in *Bacillus subtilis* chemotaxis: The effect of covalent modifications to the chemotaxis receptor McpB. *Microbiology*, **157**:56–65.

G.D. Glekas, M.J. Plutz, H.E. Walukiewicz, G.M. Allen, C.V. Rao, and G.W. Ordal. (2012). Elucidation of the multiple roles of CheD in *Bacillus subtilis* chemotaxis. *Molecular Microbiology*, **86**(3):743–756.

M. Gould. (2006). *Chemotaxis gene expression in Rhodobacter sphaeroides WS8N*. PhD thesis, Merton College, University of Oxford.

M.F. Goy, M.S. Springer, and J. Adler. (1978). Failure of sensory adaptation in bacterial mutants that are defective in a protein methylation reaction. *Cell*, **15**:1231–1240.

B.A. Hall, J.P. Armitage, and M.S.P. Sansom. (2012). Mechanism of bacterial signal transduction revealed by molecular dynamics of Tsr dimers and trimers of dimers in lipid vesicles. *PLoS Computational Biology*, **8**(9):e1002685.

P.A. Hamblin, B.A. Maguire, R.N. Grishanin, and J.P. Armitage. (1997). Evidence for two chemosensory pathways in *Rhodobacter sphaeroides*. *Molecular Microbiology*, **26**: 1083–1096.

R. Hamer, P.Y. Che, J.P. Armitage, G. Reinert, and C.M. Deane. (2010). Deciphering chemotaxis pathways using cross species comparisons. *BMC Bioinformatics*, **4**:3.

P.W. Harrison, R.P.J. Lower, N.K.D. Kim, and J.P.W. Young. (2010). Introducing the bacterial chromid: not a chromosome, not a plasmid. *Trends in Microbiology*, **18**(4): 141–148.

G.L. Hazelbauer, P. Engstrom, and S. Harayama. (1981). Methyl-accepting chemotaxis protein-III and transducer gene *trg*. *Journal of Bacteriology*, **145**:43–49.

G.L. Hazelbauer, J.J. Falke, and J.S. Parkinson. (2008). Bacterial chemoreceptors: high-performance signaling in networked arrays. *Trends in Biochemical Sciences*, **33**(1):9 – 19.

J.F. Hess, R.B. Bourret, and M.I. Simon. (1988). Histidine phosphorylation and phosphoryl group transfer in bacterial chemotaxis. *Nature*, **336**:139–143.

R. Higuchi, B. Krummel, and R. Saiki. (1988). A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Research*, **16**:7351–7367.

M. Hulko, F. Berndt, M. Gruber, J.U. Linder, V. Truffault, A. Schultz, J. Martin, J.E. Schultz, A.N. Lupas, and M. Coles. (2006). he hamp domain structure implies helix rotation in transmembrane signaling. *Cell*, **126**:929–940.

A.C. Ind, S.L. Porter, M.T. Brown, E.D. Byles, J.A. de Beyer, S.A. Godfrey, and J.P. Armitage. (2009). Inducible-expression plasmid for *Rhodobacter sphaeroides* and *Paracoccus denitrificans*. *Applied and Environmental Microbiology*, **75**:6613–6615.

K. Jahreis, T.B. Morrison, A. Garzon, and J.S. Parkinson. (2004). Chemotactic signaling by an *Escherichia coli* CheA mutant that lacks the binding domain for phosphoacceptor partners. *Journal of Bacteriology*, **186**:2664–2672.

D. Kanungpean, T. Kakuda, and S. Takai. (2011). Participation of CheR and CheB in the chemosensory response of *Campylobacter jejuni*. *Microbiology*, **157**:1279–1289.

E. Karatan, M.M. Saulmon, M.W. Bunn, and G.W. Ordal. (2001). Phosphorylation of the response regulator CheV is required for adaptation to attractants during *Bacillus subtilis* chemotaxis. *Journal of Biological Chemistry*, **276**(47):43618–43616.

M.R. Kehry and F.W. Dahlquist. (1982). Adaptation in bacterial chemotaxis: CheB-

dependent modification permits additional methylations of sensory transducing proteins. *Cell*, **29**:761–772.

M.R. Kehry, T.G. Doak, and F.W. Dahlquist. (1985). Sensory adaptation in bacterial chemotaxis: regulation of demethylation. *Journal of Bacteriology*, **163**:983–990.

K.K. Kim, H. Yokota, and S.-H. Kim. (1999). Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor. *Nature*, **400**:787–792.

J.R. Kirby, C.J. Kristichm, M.M. Saulmon, M.A. Zimmer, L.F. Garrity, I.B. Zhulin, and G.W. Ordal. (2001). CheC is related to the family of flagellar switch proteins and acts independently from CheD to control chemotaxis in *Bacillus subtilis*. *Molecular Microbiology*, **42**(3):573–585.

S. Kitanovic, P. Ames, and J.S. Parkinson. (2011). Mutational analysis of the control cable that mediates transmembrane signaling in the *Escherichia coli* serine chemoreceptor. *Journal of Bacteriology*, **193**:5062–5072.

M. Kojadinovic, A. Sirinelli, G.H. Wadhams, and J.P. Armitage. (2011). New motion analysis system for characterization of the chemosensory response kinetics of *Rhodobacter sphaeroides* under different growth conditions. *Applied and Experimental Microbiology*, **77**:4082–4088.

S. Kojima and D.F. Blair. (2003). Solubilization and purification of the MotA/MotB complex of *Escherichia coli*. *Biochemistry*, **43**:26–34.

C.J. Kristich and G.W. Ordal. (2002). *Bacillus subtilis* CheD is a chemoreceptor modification enzyme required for chemotaxis. *Journal of Biological Chemistry*, **277**(28): 25356–25362.

M.C. Leake, J.H. Chandler, G.H. Wadhams, F. Bai, R.M. Berry, and J.P. Armitage. (2006). Stoichiometry and turnover in single, functioning membrane protein complexes. *Nature*, **443**:355–358.

P. Lertsethtakarn, K.M. Ottemann, and D.R. Hendrixson. (2011). Motility and chemotaxis in *Campylobacter* and *Helicobacter*. *Annual Reviews in Microbiology*, **65**: 389–410.

M.N. Levit and J.B. Stock. (2002). Receptor methylation controls the magnitude of stimulus-response coupling in bacterial chemotaxis. *Journal of Biological Chemistry*, **277**(39):36760–36765.

J. Li, G. Li, and R.M. Weis. (1997). The serine chemoreceptor from *Escherichia coli* is methylated through an inter-dimer process. *Biochemistry*, **36**:11851–11857.

Y. Liu, M. Levit, R. Lurz, M.G.Surette, and J.B.Stock. (1997). Receptor-mediated protein kinase activation and the mechanism of transmembrane signaling in bacterial chemotaxis. *The EMBO Journal*, **16**(24):7231–7240.

G.S. Lukat, B.H. Lee, J.M. Mottonen, A.M. Stock, and J.B. Stock. (1991). Roles of the highly conserved aspartate and lysine residues in the response regulator of bacterial chemotaxis. *The Journal of Biological Chemistry*, **266**(13):8348–8354.

A. Lupas and J.B. Stock. (1989). Phosphorylation of an N-terminal regulatory domain activates the CheB methylesterase in bacterial chemotaxis. *The Journal of Biological Chemistry*, **264**(29):17337–17342.

L. Lvdok, M. Kollmann, and V. Sourjik. (2007). Co-expression of signaling proteins improves robustness of the bacterial chemotaxis pathway. *Journal of Biotechnology*, **129**(2):173–180.

L. Lvdok, K. Bentele, N. Vladimirov, A. Muller, F.S. Pop, D. Lebiedz, M. Kollmann, and V. Sourjik. (2009). Role of translational coupling in robustness of bacterial chemotaxis pathway. *PLoS Biology*, **7**(8):e1000171.

C. Mackenzie, M. Choudhary, F.W. Larimer, P.F. Predki, S. Stilwagen, J.P. Armitage, R.D. Barber, T.J. Donohue, J.P. Hosler, J.E. Newman, J.P. Shapleigh, R.E. Sockett, J. Zeilstra-Ryalls, and S. Kaplan. (2001). The home stretch, a first analysis of the nearly completed genome of *Rhodobacter sphaeroides* 2.4.1. *Photosynthesis Research*, **70**(1):19–41.

C. Mackenzie, J.M. Eraso, M. Choudhary, J.H. Roh, X. Zeng, P. Bruscella, A. Puskas, and S. Kaplan. (2007). Postgenomic adventures with *Rhodobacter sphaeroides*. *Annual Review of Microbiology*, **61**:283–307.

J.R. Maddock and L. Shapiro. (1993). Polar location of the chemoreceptor complex in the *Escherichia coli* cell. *Science*, **259**(5102):1717–1723.

J.C. Mantotta. (2002). *Investigation of methyl accepting chemotaxis proteins in Rhodobacter sphaeroides*. PhD thesis, Linacre College, University of Oxford.

A. Marchler-Bauer, S. Lu, J.B. Anderson, F. Chitsaz, M.K. Derbyshire, C. DeWeese-Scott, J.H. Fong, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, J.D. Jackson, Z. Ke, C.J. Lanczycki, F. Lu, G.H. Marchler, M. Mullokandov, M.V. Omelchenko, C.L. Robertson, J.S. Song, N. Thanki, R.A. Yamashita, D. Zhang, N. Zhang, C. Zheng, and S.H. Bryant. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*, **39**:225–229.

A.C. Martin, G.H. Wadhams, and J.P. Armitage. (2001)a. The roles of the multiple CheW and CheA homologues in chemotaxis and in chemoreceptor localization in *Rhodobacter sphaeroides*. *Molecular Microbiology*, **40**(6):1261–1272.

A.C. Martin, G.H. Wadhams, D.S.H. Shah, S.L. Porter, J.C. Mantotta, T.J. Craig, P.H. Verdult, H. Jones, and J.P. Armitage. (2001)b. CheR- and CheB-dependent chemosensory adaptation system of *Rhodobacter sphaeroides*. *Journal of Bacteriology*, **183**(24):7135–7144.

A.C. Martin, U. Nair, J.P. Armitage, and J.R. Maddock. (2003). Polar localization of CheA2 in *Rhodobacter sphaeroides* requires specific *che* homologs. *Journal of Bacteriology*, **185**:4667–4671.

D.J. McGee, M.L. Langford, E.L. Watson, J.E. Carter, Y.T Chen, and K.M. Ottemann. (2005). Colonization and inflammation deficiencies in mongolian gerbils infected by *Helicobacter pylori* chemotaxis mutants. *Infectious Immunology*, **73**:1820–1827.

C.R. Mehta and N.R. Patel. (2011). *IBM SPSS Exact Tests*. IBM.

T. Mignot, J.W. Shaevitz, P.L. Hartzell, and D.R. Zusman. (2007). Evidence that focal adhesion complexes power bacterial gliding motility. *Science*, **315**:853–856.

L.D. Miller, C.K. Yost, M.F. Hynes, and G. Alexandre. (2007). The major chemotaxis gene cluster of *Rhizobium leguminosarum* bv. *viciae* is essential for competitive nodulation. *Molecular Microbiology*, **63**(2):348–362.

M. Miyata. (2010). Unique centipede mechanism of *Mycoplasma* gliding. *Annual Review of Microbiology*, **64**:519–537.

Molisch. (1907). *Die Purpurbakterien nach neuen Untersuchungen*. Gustav Fischer, Jena.

H. Le Moual and D.E. Koshland. (1996). Molecular evolution of the C-terminal cytoplasmic domain of a superfamily of bacterial receptors involved in taxis. *Journal of Molecular Biology*, **261**:568–585.

S.L. Mowbray, D.L. Foster, and D.E. Koshland. (1985). Proteolytic fragments identified with domains of the aspartate chemoreceptor. *Journal of Biological Chemistry*, **260**(21):11711–11718.

T.J. Muff and G.W. Ordal. (2007). The CheC phosphatase regulates chemotactic adaptation through CheD. *Journal of Biological Chemistry*, **282**(47):34120–34128.

G. Mi no, T.E. Mallouk, T. Darnige, M. Hoyos, J. Dauchet, J. Dunstan, R. Soto,

Y. Wang, A. Rousselet, and E. Clement. (2011). Enhanced diffusion due to active swimmers at a solid surface. *Physical Review Letters*, **106**(4):048102.

D.M. Nowlin, J. Bollinger, and G.L. Hazelbauer. (1988). Site-directed mutations altering methyl-accepting residues of a sensory transducer protein. *PROTEINS: Structure, Function, and Genetics*, **3**:102–112.

S.Y. Park, P.P. Borbat, G. Gonzalez-Bonet, J. Bhatnagar, A.M. Pollard, J.H. Freed, A.M. Bilwes, and B.R. Crane. (2006). Reconstruction of the chemotaxis-receptor kinase assembly. *Nature Structural and Molecular Biology*, **13**:400–407.

J.S. Parkinson. (1978). Complementation analysis and deletion mapping of *Escherichia coli* mutants defective in chemotaxis. *Journal of Bacteriology*, **135**:45–53.

J.S. Parkinson. (2010). Signaling mechanisms of HAMP domains in chemoreceptors and sensor kinases. *Annual Review of Microbiology*, **64**:101–122.

J.S. Parkinson and S.E. Houts. (1982). Isolation and behaviour of *Escherichia coli* deletion mutants lacking chemotaxis functions. *Journal of Bacteriology*, **151**(1):106–113.

R.J. Penfold and J.M. Pemberton. (1992). An improved suicide vector for construction of chromosomal insertion mutations in bacteria. *Gene*, **118**:145–146.

E. Perez and A.M. Stock. (2007). Characterization of the *Thermotoga maritima* chemotaxis methylation system that lacks pentapeptide-dependent methyltransferase CheR : MCP tethering. *Molecular Microbiology*, **63**:363–378.

E. Perez, H. Zheng, and A.M. Stock. (2006). Identification of methylation sites in *Thermotoga maritima* chemotaxis receptors. *Journal of Bacteriology*, **188**:4093–4100.

T. Pilizota, M.T. Brown, M.C. Leake, R.W. Branch, R.M. Berry, and J.P. Armitage. (2009). A molecular brake, not a clutch, stops the *Rhodobacter sphaeroides* flagellar motor. *Proceedings of the National Academy of Science USA*, **106**(28):11582–11587.

S. Poggio, C. Abreu-Goodger, S. Fabela, A. Osorio, G. Dreyfus, P. Vinuesa, and L. Camarena. (2007). A complete set of flagellar genes acquired by horizontal transfer coexists with the endogenous flagellar system in *Rhodobacter sphaeroides*. *Journal of Bacteriology*, **189**(8):3208–3216.

P.S. Poole, D.R. Sinclair, and J.P. Armitage. (1988). Real time computer tracking of free-swimming and tethered rotating cells. *Analytical Biochemistry*, **175**:52–58.

S.L. Porter. (2002). *The chemotaxis proteins of Rhodobacter sphaeroides*. PhD thesis, University of Oxford.

S.L. Porter and J.P. Armitage. (2002). Phosphotransfer in *Rhodobacter sphaeroides* chemotaxis. *Journal of Molecular Biology*, **324**:35–45.

S.L. Porter and J.P. Armitage. (2004). Chemotaxis in *Rhodobacter sphaeroides* requires an atypical histidine protein kinase. *Journal of Biological Chemistry*, **279**:54573–54580.

S.L. Porter, A.V. Warren, A.C. Martin, and J.P. Armitage. (2002). The third chemotaxis locus of *Rhodobacter sphaeroides* is essential for chemotaxis. *Molecular Microbiology*, **46**(4):1081–1094.

S.L. Porter, G.H. Wadhams, A.C. Martin, E.D. Byles, D.E. Lancater, and J.P. Armitage. (2006). The CheYs of *Rhodobacter sphaeroides*. *Journal of Biological Chemistry*, **281**:32694–32704.

S.L. Porter, G.H. Wadhams, and J.P. Armitage. (2007). *In vivo* and *in vitro* analysis of the *Rhodobacter sphaeroides* chemotaxis signaling complexes. *Methods in Enzymology*, **423**:392–413.

S.L. Porter, M.A.J. Roberts, C.S. Manning, and J.P. Armitage. (2008)a. A bifunctional kinase-phosphatase in bacterial chemotaxis. *Proceedings of the National Academy of Sciences USA*, **105**(47):18531–18536.

S.L. Porter, G.H. Wadhams, and J.P. Armitage. (2008)b. *Rhodobacter sphaeroides*: complexity in chemotactic signalling. *Trends in Microbiology*, **16**(6):251–260.

S.L. Porter, D.A. Wilkinson, E.D. Byles, G.H. Wadhams, S. Taylor, N.J. Saunders, and J.P. Armitage. (2011). Genome sequence of *Rhodobacter sphaeroides* strain WS8N. *Journal of Bacteriology*, **193**:4027–4028.

E.M. Purcell. (1977). Life at low Reynolds Number. *American Journal of Physics*, **45**: 3–11.

C.V. Rao, G.D. Glekas, and G.W. Ordal. (2008). The three adaptation systems of *Bacillus subtilis* chemotaxis. *Trends in Microbiology*, **16**(10):480–487.

S.W. Reid, M.C. Leake, J.H. Chandler, C.J. Lo, J.P. Armitage, and R.M. Berry. (2006). The maximum number of torque-generating units in the fagellar motor of *Escherichia coli* is at least 11. *Proceedings of the National Academy of Sciences USA*, **103**:8066–8071.

S.A. Rogers, R.W. Huigens III, J. Cavanagh, and C. Melander. (2010). Synergistic effects between conventional antibiotics and 2-aminoimidazole-derived antibiofilm agents. *Antimicrobial Agents and Chemotherapy*, **54**(5):2112–2118.

M.M Rosario, K.L Fredrick, G.W Ordal, and J.D Helmann. (1994). Chemotaxis in *Bacillus subtilis* requires either of two functionally redundant CheW homologs. *Journal of Bacteriology*, **176**(9):2736–2739.

G. Rosser, A. G. Fletcher, D. A. Wilkinson, J. A. de Beyer, C. A. Yates, J. P. Armitage, P. K. Maini, and R. E. Baker. (2013). Novel methods for analysing bacterial tracks reveal persistence in *Rhodobacter sphaeroides*. *PLoS Computational Biology*, **9**(10): e1003276.

W.S. Ryu, R.M. Berry, and H.C. Berg. (2000). Torque-generating units of the flagellar motor of *Escherichia coli* have a high duty ratio. *Nature*, **403**:444–447.

B.M. Sager, J.J. Sekelsky, P. Matsumuka, and J. Adler. (1988). Use of a computer to assay motility in bacteria. *Analytical Biochemistry*, **173**:271–277.

J. Sambrook and J.B. Russell. (2001). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

M.K. Sarkar, K. Paul, and D. Blair. (2010). Chemotaxis signaling protein CheY binds to the rotor protein FliN to control the direction of flagellar rotation in *Escherichia coli*. *Proceedings of the National Academy of Sciences USA*, **107**(20):9370–9375.

A. Schafer, A. Tauch, W. Jager, J. Kalinowski, G. Thierbach, and A. Puhler. (1994). Small mobilizable multipurpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19 - selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene*, **145**(1):69–73.

K.A. Scott, E.E. Jefferys, B.A. Hall, M.A.J. Roberts, and J.P. Armitage. (2012). *Bacterial Regulatory Networks*, chapter Bacterial Chemotaxis, pages 223–260. Caister Academic Press.

J.E. Segall, M.D. Manson, and H.C. Berg. (1982). Signal-processing times in bacterial chemotaxis. *Nature*, **296**(5860):855–857.

D.S.H. Shah, S.L. Porter, D.C. Harris, G.H. Wadhams, P.A. Hamblin, and J.P. Armitage. (2000)a. Identification of a fourth *cheY* gene in *Rhodobacter sphaeroides* and interspecies interaction within the bacterial chemotaxis signal transduction pathway. *Molecular Microbiology*, **35**:101–112.

D.S.H. Shah, S.L. Porter, A.C. Martin, P.A. Hamblin, and J.P. Armitage. (2000)b. Fine tuning bacterial chemotaxis: analysis of *Rhodobacter sphaeroides* behaviour under aerobic and anaerobic conditions by mutation of the major chemotaxis operons and *cheY* genes. *EMBO Journal*, **19**:4601–4613.

M.J. Shapiro and D.E. Koshland. (1994). Mutagenic studies of the interaction between the aspartate receptor and methyltransferase from *Escherichia coli*. *Journal of Biological Chemistry*, **269**(15):11054–11059.

M.J. Shapiro, D. Panomitros, and D.E. Koshland Jr. (1995). Interactions between the methylation sites of the *Escherichia coli* aspartate receptor mediated by the methyltransferase. *Journal of Biological Chemistry*, **270**(2):751–755.

R.E. Silversmith, M.D. Levin, E. Schilling, and R.B. Bourret. (2008). Kinetic characterization of catalysis by the chemotaxisvphosphatase CheZ: modulation of activity by the phosphoryated CheY substrate. *The Journal of Biological Chemistry*, **283**(2): 756–765.

S.A. Simms and K. Subbaramaiah. (1991). The kinetic mechanism of S-adenosyl-L-methionine: glutamylmethyltransferase from *Salmonella typhimurium*. *Journal of Biological Chemistry*, **266**(19):12741–12746.

S.A. Simms, M.G. Keane, and J.B. Stock. (1985). Multiple forms of the CheB methylesterase in bacterial chemosensing. *The Journal of Biological Chemistry*, **260** (18):10161–10168.

W.R. Sistrom. (1977). Transfer of chromosomal genes mediated by plasmid R68.45 in *Rhodopseudomonas sphaeroides*. *Journal of Bacteriology*, **131**:526–532.

R.E. Sockett, J.C.A. Foster, and J.P. Armitage. (1990). Molecular biology of the *Rhodobacter sphaeroides* flagellum. *FEMS Symposium*, **53**:473–479.

V. Sourjik and H.C. Berg. (2002). Binding of the *Escherichia coli* response regulator CheY to its target measured in vivo by fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences USA*, **99**(20):12669–12674.

V. Sourjik and R. Schmitt. (1998). Phosphotransfer between CheA, $CheY_1$, and $CheY_2$ in the chemotaxis signal transduction chain of *Rhizobium meliloti*. *Biochemistry*, **37**(8):2327–2335.

M.S. Springer, M.F. Goy, and J. Adler. (1979). Protein methylation in behavioural control mechanisms and signal transduction. *Nature*, **230**:279–284.

W.R. Springer and D.E. Koshland. (1977). Identification of a protein methyltransferase as the *cheR* gene product in the bacterial sensing system. *Proceedings of the National Academy of Sciences USA*, **74**:533–537.

R.C. Stewart, K. Jahreis, and J.S. Parkinson. (2000). Rapid phosphotransfer to CheY from a CheA protein lacking the CheY-binding domain. *Biochemistry*, **39**:13157–13165.

A.M. Stock, V.L. Robinson, and P.N. Goudreau. (2000). Two-component signal transduction. *Annual Review in Biochemistry*, **69**:183–215.

J.B. Stock and D.E. Koshland. (1981). Changing reactivity of receptor carboxyl groups during bacterial sensing. *Journal of Biological Chemistry*, **256**:10826–10833.

R.V. Swanson, S.C. Schuster, and M.I. Simon. (1993). Expression of CheA fragments which define domains encoding kinase, phosphotransfer and CheY binding activities. *Biochemistry*, **32**:7623–7629.

H. Szurmant, M.W. Bunn, V.J. Cannistraro, and G.W. Ordal. (2003). *Bacillus subtilis* hydrolyzes CheY-P at the location of its action, the flagellar switch. *Journal of Biological Chemistry*, **278**(49):48611–48516.

H. Szurmant, T.J. Muff, and G.W. Ordal. (2004). *Bacillus subtilis* CheC and FliY are members of a novel class of CheY-P-hydrolyzing proteins in the chemotactic signal transduction cascade. *Journal of Biological Chemistry*, **279**(21):21787–21792.

F.R. Tabita. (1995). *Anoxygenic Photosynthetic Bacteria*, volume 2 of *Advances in Photosynthesis*, chapter 41, pages 885–914. Kluwer Academic Publishers.

B.L. Taylor and D.E. Koshland. (1974). Reversal of flagellar rotation in monotrichous and peritrichous bacteria: Generation of changes in direction. *Journal of Bacteriology*, **119**(2):640–642.

T.C. Terwilliger and D.E. Koshland. (1984). Sites of methyl esterification and deamination on the aspartate receptor involved in chemotaxis. *The Journal of Biological Chemistry*, **259**(12):7719–7725.

T.C. Terwilliger, J.Y. Wang, and D.E. Koshland Jr. (1986). Kinetics of receptor modification. the multiply methylated aspartate receptors involved in bacterial chemotaxis. *Journal of Biological Chemistry*, **261**:10814–10820.

S.R. Thompson. (2005). *A study of multiple chemosensory gene homologues in Rhodobacter sphaeroides*. PhD thesis, University of Oxford.

M.J. Tindall, S.L. Porter, P.K. Maini, and J.P. Armitage. (2010). Modeling chemotaxis reveals the role of reversed phosphotransfer and a bi-functional kinase-phosphatase. *PLoS Computational Biology*, **6**:e1000896.

C.B. van Niel. (1944). The culture, general physiology, morphology and classification of the non-sulfur purple and brown bacteria. *Bacteriological reviews*, **8**:1–118.

N. Verstraeten, K. Braeken, B. Debkumari, M. Fauvart, J.Fransaer, J. Vermant, and

J. Michiels. (2008). Living on a surface: swarming and biofilm formation. *Trends in Microbiology*, **16**(10):496–506.

G.H. Wadhams and J.P. Armitage. (2004). Making sense of it all: bacterial chemotaxis. *Nature Reviews in Molecular and Cellular Biology*, **5**:1024–1037.

G.H. Wadhams, A.C. Martin, and J.P. Armitage. (2000). Identification and localization of a methyl-accepting chemotaxis protein in *Rhodobacter sphaeroides*. *Molecular Microbiology*, **36**:1222–1233.

G.H. Wadhams, A.C. Martin, S.L. Porter, J.R. Maddock, J.C. Mantotta, H.M. King, and J.P. Armitage. (2002). TlpC, a novel chemotaxis protein in *Rhodobacter sphaeroides*, localizes to a discrete region in the cytoplasm. *Molecular Microbiology*, **46**(5):1211–1221.

G.H. Wadhams, A.V. Warren, A.C. Martin, and J.P. Armitage. (2003). Targeting of two signal transduction pathways to different regions of the bacterial cell. *Molecular Microbiology*, **50**(3):763–770.

G.H. Wadhams, A.C. Martin, A.V. Warren, and J.P. Armitage. (2005). Requirements for chemotaxis protein localization in *Rhodobacter sphaeroides*. *Molecular Microbiology*, **58**:895–902.

M.J. Ward, A.W. Bell, P.A. Hamblin, H.L. Packer, and J.P. Armitage. (1995). Identification of a chemotaxis operon with two *cheY* genes in *Rhodobacter sphaeroides*. *Molecular Microbiology*, **17**:357–366.

A. Warren. (2003). *Analysis of the chemosensory operons of Rhodobacter sphaeroides*. PhD thesis, University of Oxford.

R.M. Weis and D.E. Koshland Jr. (1990). Chemotaxis in *Escherichia coli* proceeds efficiently from different initial tumble frequencies. *Journal of Bacteriology*, **172**(2): 1099–1105.

A.H. West and A.M. Stock. (2001). Histidine kinases and response regulator proteins in two-component signaling systems. *Trends in Biochemical Sciences*, **26**(1):369–376.

A.H. West, E. Martinez-Hackert, and A.M. Stock. (1995). Crystal structure of the catalytic domain of the chemotaxis receptor methylesterase, CheB. *Journal of Molecular Biology*, **250**:276–290.

T. Wood, C.A. Yates, D. Wilkinson, and G. Rosser. (2012). Simplified multitarget tracking using the PHD filter for microscopic video data. In *IEEE Transactions on Circuits and Systems for Video Technology*, volume 22, pages 702–713, (2012).

G.A. Wright, R.L. Crowder, R.R. Draheim, and M.D. Manson. (2011). Mutational analysis of the transmembrane Helix 2-HAMP domain connection in the *Escherichia coli* aspartate chemoreceptor Tar. *Journal of Bacteriology*, **193**:82–90.

J. Wu, J. Li, G. Li, D.G. Long, and R.M. Weis. (1996). The receptor binding site for the methyltransferase of bacterial chemotaxis is distinct from the sites of methylation. *Biochemistry*, **35**:4984–4993.

L. Wu, Y. Cui, Y. Hong, and S. Chen. (2011). A CheR/CheB fusion protein is involved in cyst cell development and chemotaxis in *Azospirillum brasilense* sp7. *Microbiological Research*, **166**(8):606–617.

L. Xie, T. Altindal, S. Chattopadhyay, and X.-L. Wu. (2011). Bacterial flagellum as a propeller and as a rudder for efficient chemotaxis. *Proceedings of the National Academy of Science USA*, **108**(6):2246–2251.

H. Yonekawa, H. Hayashi, and J.S. Parkinson. (1983). Requirement of the *cheB* function for sensory adaptation in *Escherichia coli*. *Journal of Bacteriology*, **156**(3): 1228–1235.

M. Youle, F. Rohwer, A. Stacy, M. Whiteley, B.C. Steel, N.J. Delalez, A.L. Nord, R.M. Berry, J.P. Armitage, S. Kamoun, S. Hogenhout, S.P. Diggle, J. Gurney, E.J.G. Pollitt, A. Boetius, and S.C. Cary. (2012). The Microbial Olympics. *Nature Reviews Microbiology*, **10**:583–588.