Appl Microbiol Biotechnol (2011) 92:997–1008
DOI 10.1007/s00253-011-3582-y

GENOMICS, TRANSCRIPTOMICS, PROTEOMICS

Comparative structural bioinformatics analysis of *Bacillus amyloliquefaciens* chemotaxis proteins within *Bacillus subtilis* group

Anna Yssel · Oleg Reva · Ozlem Tastan Bishop

Received: 24 March 2011 / Revised: 17 August 2011 / Accepted: 15 September 2011 / Published online: 13 October 2011

Abstract Chemotaxis is a process in which bacteria sense their chemical environment and move towards more favorable conditions. Since plant colonization by bacteria is a multifaceted process which requires a response to the complex chemical environment, a finely tuned and sensitive chemotaxis system is needed. Members of the *Bacillus subtilis* group including *Bacillus amyloliquefaciens* are industrially important, for example, as bio-pesticides. The group exhibits plant growth-promoting characteristics, with different specificity towards certain host plants. Therefore, we hypothesize that while the principal molecular mechanisms of bacterial chemotaxis may be conserved, the bacterial chemotaxis system may need an evolutionary tweaking to adapt it to specific requirements, particularly in the process of evolution of free-living soil organisms, towards plant colonization behaviour. To date, almost nothing is known about what parts of the chemotaxis proteins are subjected to positive amino acid substitutions, involved in adjusting the chemotaxis system of bacteria during speciation. In this novel study, positively selected and purified sites of chemotaxis proteins were calculated, and these residues were mapped onto homology models that were built for the chemotaxis proteins, in an attempt to understand the spatial evolution of the chemotaxis proteins. Various positively selected amino acids were identified in semi-conserved regions of the proteins away from the known active sites.

Keywords Homology modeling · Chemotaxis receptor · Positive selection · Purifying selection

Introduction

Bacterial movement in an aqueous environment is controlled by alternating the tumble and swim phases of clockwise (CW) and counter-clockwise (CCW) rotation of flagella. A coordinated interaction of chemotaxis proteins and signal transduction from chemoreceptors to the flagellar motor apparatus enable bacteria to bias their motion towards a more favorable chemical environment (Garrity and Ordal 1995; Rao et al. 2004). Thus, this complex system of chemotaxis is absolutely important for survival of microorganisms in their habitats. For example, chemotaxis has been shown to be critical for plant colonization of both pathogenic (Yao and Allen 2006) and symbiotic plant-associated microbes (Van de Broek et al. 1998). Plant colonization is a multifaceted process which requires bacteria to respond to the complex chemical environment of the plant rhizosphere, in particular to detect and resist plant defense systems, as well as the ability to sense attractants and initiate growth on the plant surface. Therefore, a finely tuned and sensitive chemotaxis system is required.

Of interest to this paper is *Bacillus amyloliquefaciens*. It forms a closely related taxonomic unit with *Bacillus subtilis*, *Bacillus licheniformis*, and *Bacillus pumilus* (Fritze 2004). They are commonly referred to as the *B. subtilis*

Electronic supplementary material The online version of this article (doi:10.1007/s00253-011-3582-y) contains supplementary material, which is available to authorized users.

A. Yssel · O. Tastan Bishop (✉)
Department of Biochemistry, Microbiology and Biotechnology, Rhodes University Bioinformatics (RUBi), Rhodes University, Grahamstown 6140, South Africa
e-mail: O.TastanBishop@ru.ac.za

O. Reva
Department of Biochemistry, Bioinformatics and Computational Biology Unit, University of Pretoria, Pretoria 0002, South Africa

group. Members of this group are free-living soil microorganisms but several strains of *B. amyloliquefaciens* and *B. subtilis* are plant growth-promoting rhizobacteria (Reva et al. 2004). It has been shown that different strains of members of the *B. subtilis* group exhibit different specificity towards certain host plants (Reva et al. 2004). A recent paper argued that the chemotactic-competent bacteria present in the rhizosphere of wheat and cowpea are different from, and less diverse than, those in the bulk soil, indicating the development of specialized microbial communities (Buchan et al. 2010). Therefore, it might be expected that the chemotaxis proteins of the members of the *B. subtilis* group may be adapted to their specific habitats. This then raises the question as to the precise nature, at the molecular level, of the adaptation, or tweaking, mechanism of the chemotaxis proteins in related group of organisms living in different environments. To our knowledge, there has been no prior study of this question, and such an investigation is the primary objective of this paper.

B. subtilis group has an important role in various industrial applications. *B. amyloliquefaciens* produces a large number of anti-fungal and anti-bacterial substances with pharmacological and agricultural value (Chen et al. 2009). In this study, we used five *B. amyloliquefaciens*, as described below, with agricultural significance. Two more distant organisms, *B. licheniformis* and *B. pumilus*, were used for comparison, both of which also have industrial applications (Choudhary and Johri 2009).

The core proteins of all known chemotaxis systems is comprised of the methyl accepting chemotaxis proteins (MCPs), sensor histidine kinase CheA, receptor coupling protein CheW, response regulator CheY, methyltransferase and methylesterase proteins CheR and CheB, and a number of other proteins specific for different taxa (Garrity and Ordal 1995; Hamer et al. 2010; Rao et al. 2004). The chemotaxis mechanism of *Escherichia coli* is no doubt the best understood and well-studied model for chemotaxis; however, more recent studies on *B. subtilis, Thermatoga maritima*, and *Rhodobacter sphaeroides*, amongst others, have shown that there are more complex mechanisms of chemotaxis than in *E. coli* (Hamer et al. 2010). Six chemotaxis receptors in *E. coli* and ten in *B. subtilis* have been discovered (Manson et al. 1998; Mowbray and Sandgren 1998). All of them are transmembrane proteins, except for HemAT (Zhang and Phillips 2003). The McpA, McpB, and McpC receptors are well studied in *B. subtilis*. In addition to the core proteins, *B. subtilis* also contains CheV which is functionally similar to CheW (Rosario et al. 1994), CheC—a member of a similarly named family of phosphatases, and also the regulatory protein CheD (Rosario et al. 1995; Rao et al. 2008). It has been suggested that the more complex system of *B. subtilis* may be representative of that employed by the ancestral organism from which the archaea and bacteria arose (Garrity and Ordal 1995).

As mentioned above, the key proteins of the chemotaxis system are shared by organisms belonging to different bacterial classes of proteobacteria and archaea. For instance, it was shown in experiments that in *trans*-complementation of *E. coli* cells with knocked out chemotaxis proteins CheA and CheB with their counterparts from *B. subtilis* rescued the mutation (Garrity and Ordal 1995). It was shown in a previous study that CheA is quite variable even on the level of subspecies of bacteria of *B. subtilis* group (Reva et al. 2004). Thus, we hypothesize that while the principal molecular mechanisms of bacterial chemotaxis are conserved, the bacterial chemotaxis system may need an evolutionary tweaking to meet the unique requirements of the ecological niche where the species exists. The molecular mechanisms of self-adaptation of the chemical gradient-sensing system of *Bacillus* to environmental changes were studied in detail (Rao et al. 2008).

To date, almost nothing is known about which of the chemotaxis proteins are subjected to positive amino acid substitutions, involved in adjusting the chemotaxis system of bacteria during speciation. In this paper, homology models of chemotaxis proteins were built, and positively selected and purified sites of chemotaxis proteins were calculated. Next, these residues were mapped onto the 3-dimensional (3D) models of these proteins in an attempt to understand the spatial evolution of the chemotaxis system. The plant promoting strain *B. amyloliquefaciens* FZB42 (Chen et al. 2007) was selected as a reference organism. Newly sequenced *B. amyloliquefaciens* strains DSM7^T, B946, B9601Y2, and GaoB3 (Borriss R., personal communication) were also used for comparative analysis. The availability of complete genome sequences of closely related organisms exhibiting different capacities of plant colonization, allowed studying the micro-evolution of the chemotaxis system at the subspecies level. The obtained knowledge will allow better understanding of the plant-colonizing ability of industrial strains of *B. amyloliquefaciens* used in bio-pesticides and will aid in better bio-preparation design.

Materials and methods

Data retrieval

The nucleotide and amino acid sequences for each chemotaxis protein and chemoreceptor in *B. amyloliquefaciens* ssp. *plantarum* FZB42^T (NC_009725), *B. subtilis* ssp. *subtilis* 168 (NC_000964), *B. subtilis* SMY (ABQN01000001-ABQN01000009), *B. subtilis* JH642 (ABQM01000001-ABQM01000009), *B. subtilis* NCBI 3610 (ABQL01000001-

ABQL01000005), *B. subtilis* ssp. *spizizenii* (NC_014479), *B. amyloliquefaciens* ssp. *amyloliquefaciens* DSM7^T (NC_014551), *B. licheniformis* str. ATCC 14580 (NC_006270), and *B. pumilus* SAFR-032 (NC_009848) were retrieved from the RefSeq database; and newly sequenced *B. amyloliquefaciens* B946, *B. amyloliquefaciens* B9601Y2 and *B. amyloliquefaciens* GaoB3 were provided by Prof. R. Borriss (Humboldt University, Berlin). These newly sequenced genomes were used as drafts in a work published by Borriss et al. (2010). Orthologs were confirmed by a protein BLAST (Altschul et al. 1990) search against the reference strain *B. amyloliquefaciens* ssp. *plantarum* FZB42.

Sequence alignment and sequence identity calculations

Full length sequences of the proteins CheA, CheB, CheC, CheD, CheB, CheR, CheV, CheW, CheY, McpA, McpB, and McpC from the representative members of the *B. subtilis* group were aligned with known structures (if available) using PROMALS3D (Pei et al. 2008). The positions and residue compositions of known active and functionally important sites were determined. Additionally, prior to submission to Selecton, whole nucleotide sequences were codon aligned by MUSCLE with the maximum number of iterations set to 16 (Edgar 2004). Then, the phylogenetic tree and hypothetical ancestral states of the sequences were inferred using the maximal likelihood approach with the program PHYLIP dnaml (Felsenstein and Churchill 1996; Hernández-Sánchez et al. 2008). In the phylogenetic analysis, *B. pumilus* SAFR-032 was used as an outgroup. The sequence identities for the homologous proteins of the different organisms used in this study were calculated with BioEdit (Hall 1999).

Determining positive and purifying selection

The Selecton server (http://selecton.tau.ac.il/) (Doron-Faigenboim et al. 2005) was used to calculate the ratio of non-synonymous to synonymous substitutions, known as the K_a/K_s ratio. The program identifies regions in protein sequences with a K_a/K_s ratio significantly greater than 1 as positively selected and those with the K_a/K_s ratio significantly smaller than 1 are considered as areas of purifying selection, and the randomly mutated regions have K_a/K_s ratio close to 1 (Doron-Faigenboim et al. 2005). Codon-aligned DNA sequences of the chemotaxis gene and predicted ancestral sequences were analyzed by Selecton. The ancestral sequences, predicted by the maximal likelihood algorithm, were added to the alignment to improve the statistical reliability of the analysis. The mechanistical empirical model (MEC) was set for the Selecton program run. The MEC model differs from the other methods provided by Selecton in that it takes into account the differences between various amino acid replacement probabilities (Doron-Faigenboim and Pupko 2007). Fourteen categories of discreet approximations of amino acid substitution likelihoods were set for the program run, which is the maximum allowed by Selecton and gives the most accurate results. Statistical analysis was performed by comparing Akaike Information Content scores between the MEC and M8a models provided by Selecton server (Table S1). Statistical confidence was controlled by calculated p values that depend on the number of sequences in the sample.

Homology modeling and validation

Template selection was done using the HHpred server (Soding et al. 2005). After template selection, a target-template sequence alignment was built with HHpred's built-in alignment function. Additionally, the target-template alignments were compared with the multiple sequence alignments produced with PROMALS3D to verify overall accuracy, and no further editing was performed. The homology models of the chemotaxis proteins, CheB, CheC, CheD, CheR, CheW, CheY of *B. amyloliquefaciens* FZB42 were calculated using available crystal structures of their homologs, obtained from the Protein Data Bank (PDB). The respective templates used were: 1A2O (Djordjevic et al. 1998), 1XKR (Park et al. 2004a), 2F9Z (Chao et al. 2006), 1AF7 (Djordjevic and Stock 1997), 2QDL (Yao et al. 2007), 1TMY (Usher et al. 1998), respectively. For further information on target-template matches, see Table 1. The target-template alignments were used as inputs for the program MODELLER-9v7 for homology modeling (Sali and Blundell 1993). When needed, problematic loops were refined using MODELLER. The models were evaluated based on their discrete optimized protein energy (DOPE Z) scores and MetaMQAPII results (Pawlowski et al. 2008).

Homology modeling is a promising new approach for predicting protein complex structures (Tastan Bishop and Kroon 2011). Protein complexes of CheAp2–CheY, CheAp4p5–CheW, CheC–CheD, and CheY in complex with the second phosphorylation center of CheC were modeled using the following crystal structure complexes as templates: 1U0S (Park et al. 2004b); 2CH4 (Park et al. 2006), 2F9Z (Chao et al. 2006), and 3HZH (Pazy et al. 2010). More information on templates can be found in Table 1. Target-template alignments were constructed using PROMALS3D. Model quality assessment of complexes was performed by ProQ (Wallner and Elofsson 2003) and Verify3D (Eisenberg et al. 1997). Problematic loops were refined when necessary. Structures were visualized and analyzed using PyMOL (DeLano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org). The residues under positive and purifying selection were mapped on the 3D structures of proteins using the Selecton color code.

Table 1 Target-template information and model quality assessment results

Protein	CheB	CheC	CheD	CheR	CheW	CheY	CheAP2 CheY	CheAP4P5 CheW	CheC CheD	CheY CheC
Templates used										
PDB id	1A2O_A	1XKR_A	2F9Z_C	1AF7_A	2QDL_A	1TMY_A	1U0S_A	2CH4_A	2F9Z_A	3HZH_A
							1U0S_Y	2CH4_W	2F9Z_C	3HZH_B
Sequence identity	39%	30%	42%	29%	30%	71%	34%	47%	31%	43%
							71%	27%	42%	22%
Residue range[a]	2–354	4–207	11–158	2–254	9–156	1–119	163–247	349–669	7–207	1–120
							1–120	7–148	1–165	111–136
E-value[b]	0	1.6e-42	0	1.4e-45	2.7e-32	3.6e-29	3e-12	0	1.6e-420	1.1e-25
							3.6e-29	09.9e-32	0	2.9e-16
Before loop refinement										
Dope Z	−0.832	−1.085	−1.002	−1.150	−1.347	−1.836	−1.552	−0.677	−1.002	−1.034
GDT-TS	72.958	78.469	68.289	69.238	65.287	82.917	–	–	–	–
Verify3D	–	–	–	–	–	–	89%	89%	77%	61%
ProQ LG	–	–	–	–	–	–	4.932	5.328	5.504	2.562
After loop refinement										
Dope Z	−1.079	−1.490	−1.281	−1.470	−1.668	−1.88	–	−1.002	−0.776	–
GDT-ts	80.493	83.732	76.342	78.613	73.885	84.792	–	–	–	–
Verify3D	–	–	–	–	–	–	–	96%	83%	–
ProQ LG	–	–	–	–	–	–	–	6.451	5.998	–

A sound model typically has a GDT-TS score of >75 and Dope *Z* score <−0.5 Å. A ProQ LG score >5 indicates a very good model. A model with a Verify3D score of more than 80% indicates a structure of experimental quality

[a] Residue range of target that aligns with template

[b] The E-value is a measure of reliability; it gives the average number of false positives with a score better than the one for the template when scanning the database. E-values near 0 signify a very reliable hit

Results

Sequence similarity comparison

The similarity of protein sequences of CheA, CheB, CheC, CheD, CheR, CheV, CheW, CheY, McpA, McpB, and McpC of organisms related to *B. subtilis* were compared to the reference organism *B. amyloliquefaciens* FZB42. They showed variability on the level of 55% to 99% identity, as detailed in Fig. 1. *B. amyloliquefaciens* chemotaxis proteins were most similar to their counterparts from *B. subtilis* 168, followed by *B. licheniformis* ATCC14580, then *B. pumilus* SAFR032. The level of similarity between *B. subtilis* and *B. amyloliquefaciens* was above 65% for all homologous proteins, indicating a high level of conservation. However, recognizable variations in proteins were discovered between subspecies *B. amyloliquefaciens* and *B. subtilis*. The frequency of amino acid substitutions in chemotaxis proteins on average was the same as or a little bit higher than in the gyrase A (GyrA) subunit (data not shown) broadly used for phylogenetic inferences (Chun and Bae 2000; Reva et al. 2004). A general belief is that GyrA accumulates amino acid substitutions in a random molecular clock fashion without any significant evolutionary pressure.

Homology models of chemotaxis proteins

This study aimed to determine the selective forces acting upon the proteins of the chemotaxis system within the *B. subtilis* group; particularly to elucidate the adaptive spatial evolution of the chemotaxis proteins of plant-associated *B. amyloliquefaciens*. The homology models were assessed using DOPE *Z* score (Shen and Sali 2006) and model evaluation algorithms provided by the MetaMQAPII server (Pawlowski et al. 2008). Generally, models with a DOPE *Z* score lower than −1 are likely to be native like. All final models of individual proteins have very good DOPE *Z* score, GDT-TS (according to the MetaMQAPII results) (Table 1). In a similar approach, the protein complexes were also modeled, based on solved complex structures, in order to analyze the spatial interactions between chemotaxis proteins. Model quality assessment results of complexes can be found in Table 1.

No structure has been solved for any of the MCPs from *Bacillus*, but the overall topology was predicted based on a

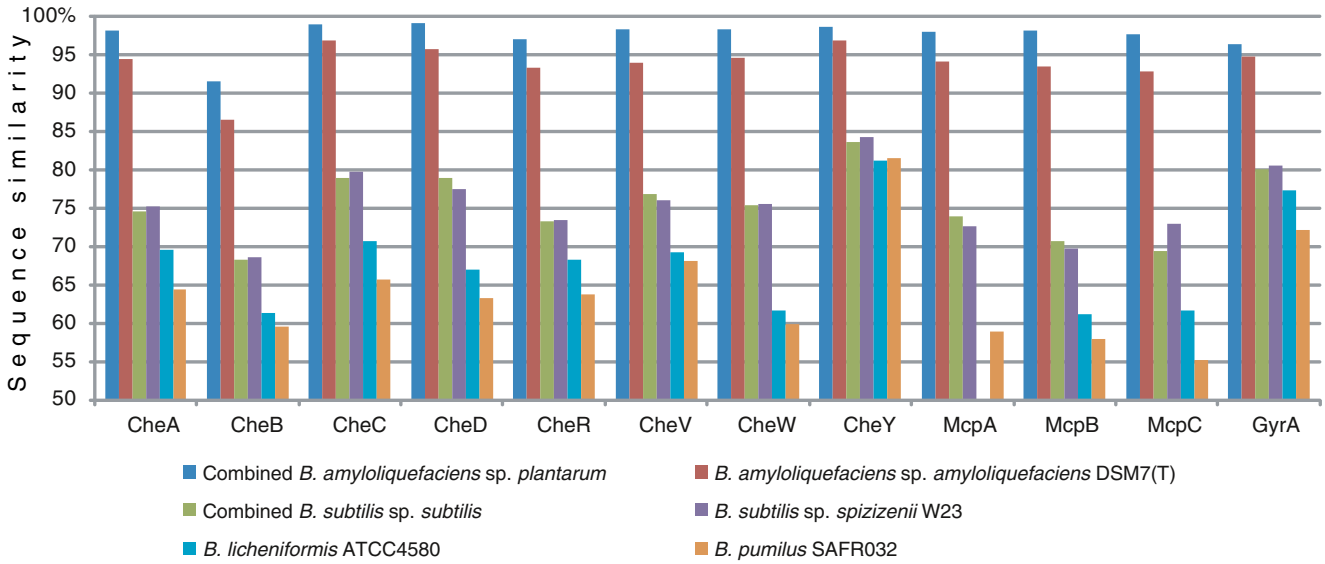


Fig. 1 Sequence similarity histogram. Sequence similarity of *B. amyloliquefaciens* FZB42 chemotaxis proteins compared to their counterparts in genomes of other organisms of the *B. subtilis* group

modular architecture as revealed by comparative sequence analysis and homology to partially solved structures of these receptors from other organisms (Bunn and Ordal 2003; Kristich et al. 2003; Zimmer et al. 2000).

Positive and purifying selection—mapping to 3D structures

The homologous sequences were aligned, and the rates of substitutions were determined for every residue position in each alignment. Purified and positively selected sites were identified and assessed by the Selecton server (Doron-Faigenboim et al. 2005). Thereafter, the position-specific substitution scores were checked for every predicted site. Finally, 34 residues in seven of the chemotaxis proteins and the three chemoreceptors were considered as positively selected. A position-specific replacement-score table and an additive tree graph were constructed, based on concatenated strings of positively selected residues (Fig. 2). Please note that the figure shows only 32 residues, and 6 proteins. CheV (site 26) and McpB (site 2) were omitted. McpB (site 2) falls within the N-terminal loop region of the protein. Usually the N-terminal and C-terminal regions of proteins are highly disordered, and amino acids are normally highly variable at these positions, so the result of finding positive selection is not unexpected. CheV was omitted because this protein was not discussed in detail in this study, and no structure for this protein has been solved. The organisms belonging to the same species were grouped together, showing that the variation at these sites was not random. Remarkably, the chemotaxis proteins of *B. subtilis* strains are much more conserved than those of *B. amyloliquefaciens* strains, which comprise soil dwelling and plant-associated organisms.

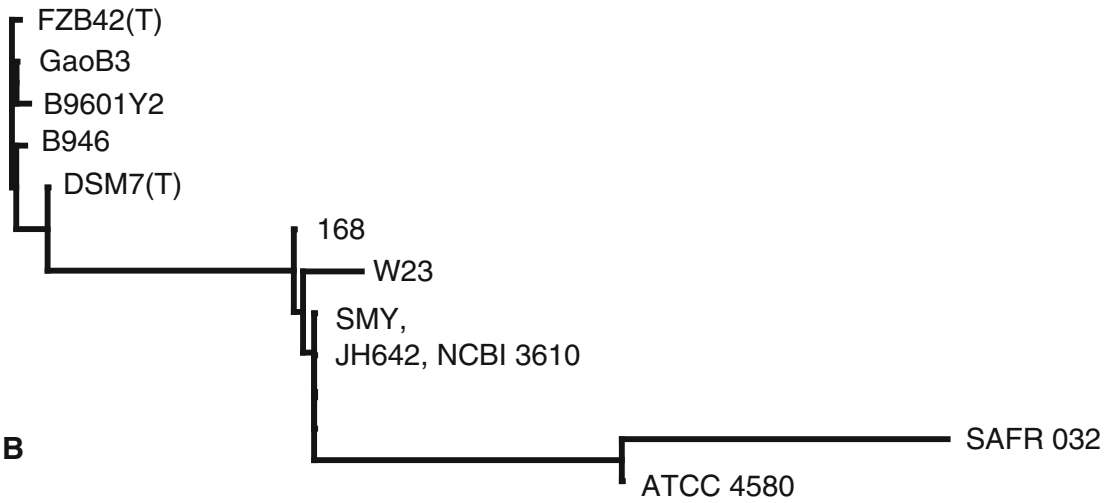
The sites under positive Darwinian selection may confer adaptability of the chemotaxis system of microorganisms to their specific habitats. Amino acid sites under significant positive selective pressure were found in methylesterase CheB, phosphatase CheC, coupling protein CheW, and the chemoreceptors McpA, McpB, McpC. No positively selected sites were identified in the GyrA protein, used as a hallmark of the neutral evolution (Fig. S1). Sites identified as being under evolutionary pressure (Fig. 3A and Table S2) were mapped to the predicted 3D structures of proteins and colored according to the Selecton score (Fig. 3B). Many structural elements of the proteins are under strong purifying selection, as expected considering the high level of conservation of these proteins (Fig. 1). The highest level of conservation was detected in active sites of the proteins highlighted in Fig. 3. Other conserved residues were found at predicted binding sites of the proteins (Fig. 4).

Discussion

In this paper, we investigated the chemotaxis system of industrially important bacteria *B. amyloliquefaciens*, a member of the *B. subtilis* group. We combined the methods of protein structural modeling with the prediction of positively selected residues. This represents a novel attempt to identify possible sites of evolutionary tweaking, which can play a role in the adaptation of the bacterial chemotaxis system to specific requirements imposed on it by the environment, particularly in the process of evolution of free-living soil organisms towards plant-colonizing behaviour. Positively selected sites were defined as those which

	CheB				CheC	CheW	McpA							McpB							McpC									
	25	90	277	279	86	18	95	99	112	234	235	244	289	92	99	288	376	408	412	478	52	71	123	169	171	242	276	322	354	536	647	651
B. amyloliquefaciens sp. *plantarum* FZB42(T)	V	R	R	S	D	G	T	T	Y	S	S	Y	L	S	G	A	S	V	H	H	G	G	D	Q	S	Q	T	R	A	S	I	S
B. amyloliquefaciens sp. *plantarum* B946	V	R	R	S	H	G	T	T	Y	T	S	Y	L	S	G	A	S	A	H	H	S	D	D	Q	S	Q	T	R	A	S	I	S
B. amyloliquefaciens sp. *plantarum* B9601Y2	V	R	R	S	D	G	T	T	F	T	S	Y	L	S	G	S	S	A	H	H	G	D	E	Q	S	Q	T	R	A	S	I	S
B. amyloliquefaciens sp. *plantarum* GaoB3	V	R	R	S	D	G	T	T	F	T	S	Y	L	S	G	A	S	A	H	H	G	G	D	Q	S	Q	T	R	A	S	I	S
B. amyloliquefaciens sp. *amyloliquefaciens* DSM7(T)	G	R	R	S	H	G	T	S	Y	S	S	Y	L	T	G	A	S	A	H	H	G	G	D	Q	N	R	T	R	T	S	I	S
B. subtilis SMY	K	C	K	T	F	G	A	S	Y	S	A	M	A	E	S	I	G	H	L	S	E	N	S	R	-	N	R	R	L	L	M	A
B. subtilis JH642	K	C	K	T	F	G	A	S	Y	S	A	M	A	E	S	I	G	H	L	S	E	N	S	R	-	N	R	R	L	L	M	A
B. subtilis NCBI 3610	K	C	K	T	F	G	A	S	Y	S	A	M	A	E	S	I	G	H	L	S	E	N	S	R	-	N	R	R	L	L	M	A
B. subtilis sp. *subtilis* 168	K	C	K	T	F	G	R	S	Y	S	A	M	A	E	S	I	G	H	L	S	E	N	D	R	-	N	R	R	L	L	I	R
B. subtilis sp. *spizizenii* W23	K	C	K	A	Y	G	K	S	Y	N	D	L	T	E	S	I	G	H	L	S	E	N	D	L	G	E	R	H	L	L	M	T
B. licheniformis ATCC 4580	H	I	S	R	S	H	V	E	Y	G	S	F	A	D	E	G	D	Q	-	A	K	D	T	N	-	-	K	Q	S	M	L	N
B. pumilus SAFT 032	Q	L	H	K	E	K	Q	E	F	S	Q	F	G	S	Q	S	G	E	L	Q	-	T	E	K	-	-	Q	S	G	D	R	D
Norm. Ka/Ks:	1.9	1.8	2.1	1.9	1.6	1.6	1.7	2.3	1.6	1.6	1.7	2.4	1.9	1.9	1.6	2.1	2.2	1.6	2.0	1.6	2.6	1.8	2.2	2.4	2.2	2.2	1.9	1.9	1.9	1.7		

A

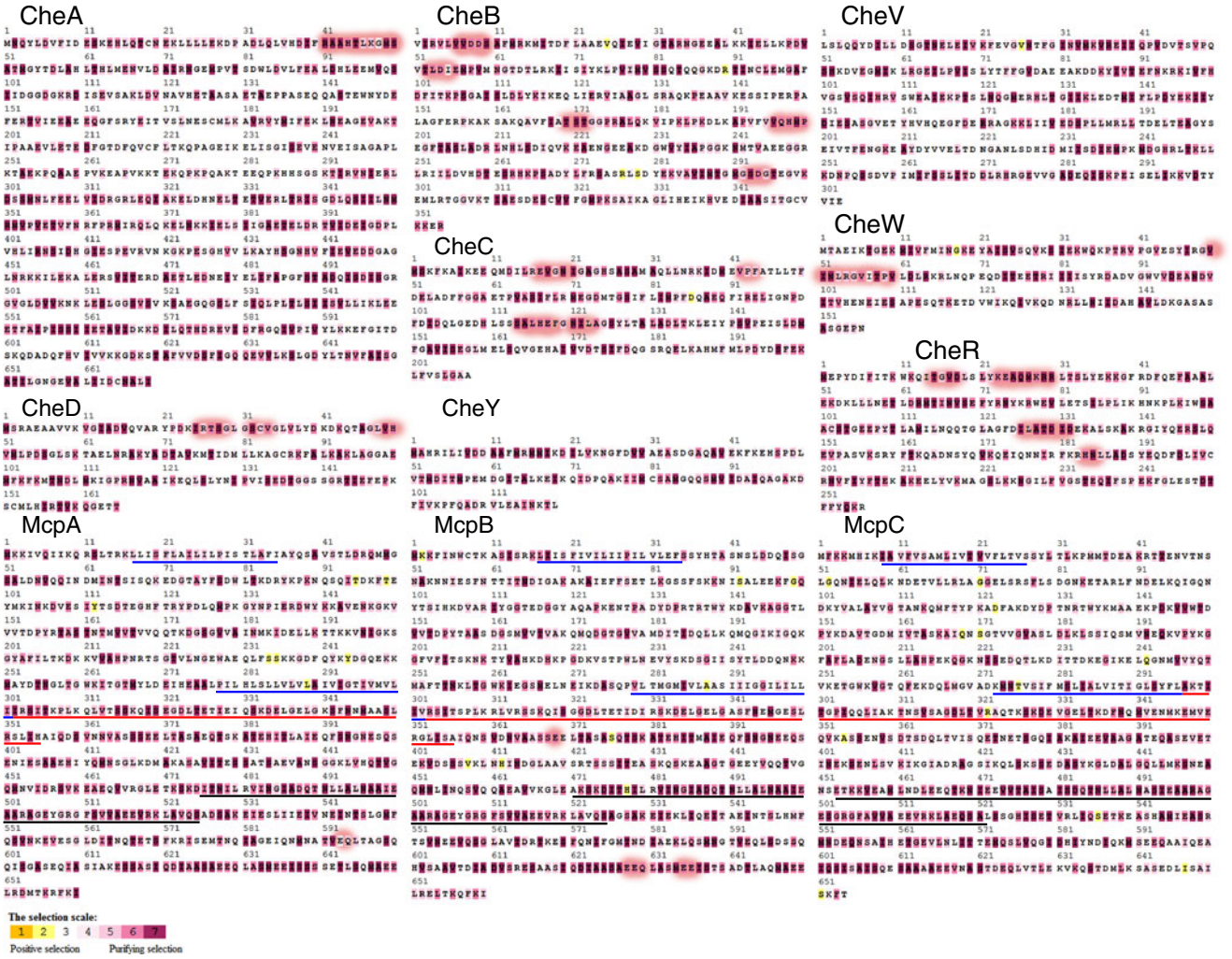


B

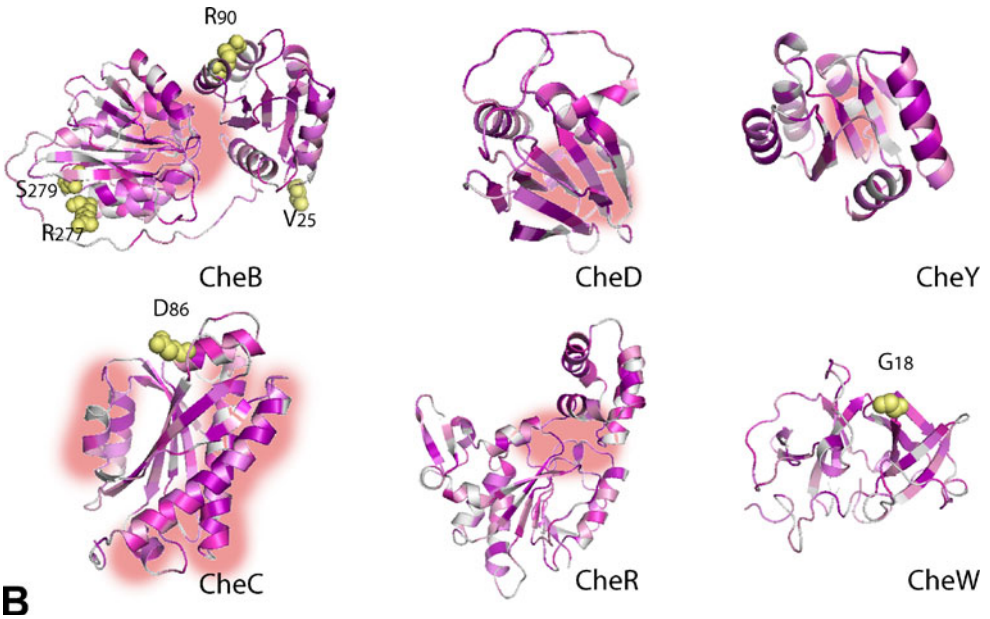
Fig. 2 Positively selected residues and additive tree graph based on these residues. **A** Position-specific substitution table of amino acid residues under positive Darwinian selection. Gaps in the sequence alignment are denoted by "-". **B** An additive tree graph calculated by the maximum likelihood algorithm based on concatenated strings of these residues. Note: *B. licheniformis* ATCC 4580 has two paralogous versions of McpA (A and B). McpAA was used for alignment as it is more similar to McpA of FZB42. However, McpAA was excluded from the Selecton input alignment to avoid ambiguity

are variable between strains of closely related organisms for which the relative rate of non-synonymous nucleotide substitutions at the corresponding codons is higher than expected for randomly mutated sites. Hence, a selective evolutionary pressure may be supposed. Positively selected amino acids were identified in semi-conserved accessory regions of the proteins away from the known enzymatic and binding sites (Fig. 3). Substitution of these amino acids may result in conformational changes of the proteins that will eventually affect the enzymatic activity of the protein and the timing of the response to the signal; however, this hypothesis requires further investigation. Adaptation of the chemotaxis system of free-living and plant/rhizosphere-associated bacteria most likely involves adjustment of the sensitivity to different attractants and repellents by modifying corresponding N-terminal extracellular and C-terminal intracellular domains of corresponding MCPs; and by

Fig. 3 Color code presentation of Selecton results for the selective forces. **A** Sequences of chemotaxis proteins of *B. amyloliquefaciens* FZB42 are highlighted in accordance with the Selecton color code. The selection color scale ranges from positively selected residues (*orange* and *yellow*) to purified residues (*dark pink*) through randomly mutated residues (*white*). Known active and binding sites of the proteins are depicted by pink halo. In sequences of chemoreceptors, the functional domains are *blue* (*TM*), *red* (*HAMP*) and *black* (*SH*) *underlined*. The areas between HAMP and SH domains and from SH domain to the end of the sequence are MH1 and MH2 domains, respectively. **B** Homology models of six *B. amyloliquefaciens* FZB42 proteins are colored according to the Selecton color code using a PyMOL script. Labelled residues represented as *yellow spheres* are subject to positive selection. Known active and binding sites of the proteins are depicted by *pink halo*


CheA
CheB
CheV
CheC
CheW
CheD
CheY
CheR
McpA
McpB
McpC
The selection scale:
1 2 3 4 5 6 7
Positive selection
Purifying selection

A


R90
S279
R277
V25
CheB
CheD
CheY
D86
CheC
CheR
G18
CheW

B

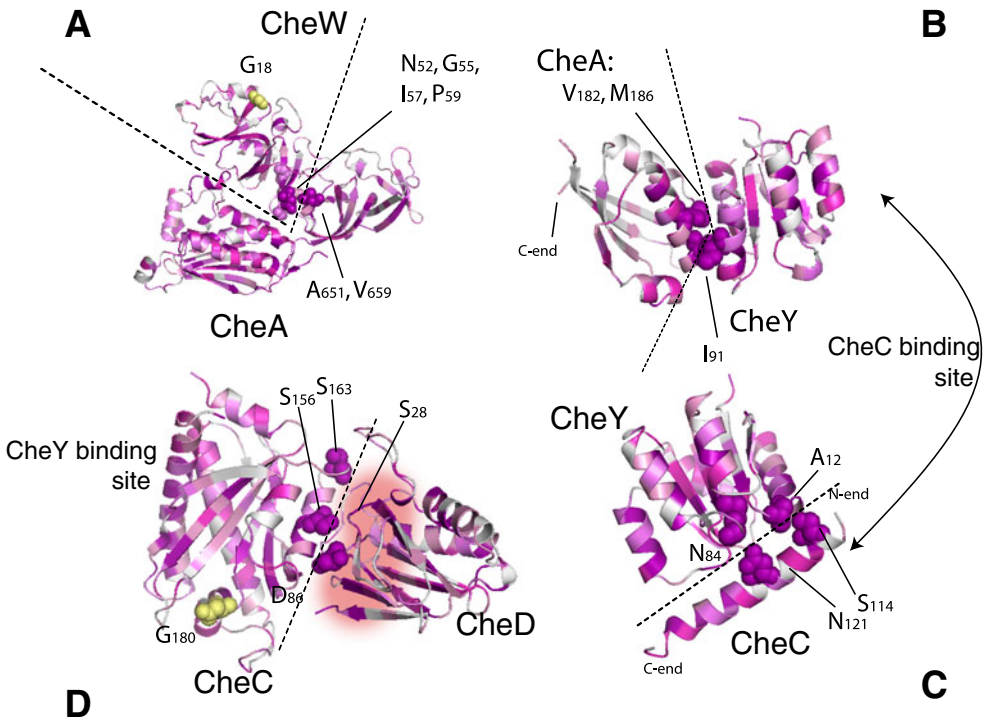


Fig. 4 Color code presentation of Selecton results for the selective forces mapped to protein complexes. Homology models of interacting proteins from *B. amyloliquefaciens* FZB42 are colored according to the Selecton color code. Labelled yellow spheres are residues under positive selection. Labelled purple and violet spheres are residues that play a role in interaction between the two interaction partners. *Dashed lines* separate domains of different proteins. **A** CheW binds CheAp4 and CheAp5 domains to the left and right from the *dashed line*, respectively; **B** CheAp2–CheY complex. The CheC binding site is on the opposite end of where CheA binds to CheY; **C** CheY–CheC complex. N-terminal and C-terminal residues of the CheC fragment are L111 and K136, respectively; **D** CheC–CheD complex. CheY and CheD bind at opposite ends of the CheC molecule. The *shaded pink area* denotes the enzymatic active region of CheD that is blocked when bound to CheC

adjustment of the positive and negative response timing through modifying several specific loci of the regulatory proteins CheB and CheC.

Relationships between chemotaxis proteins, known from the literature, are summarized in the scheme in Fig. 5. It has to be noted that several important elements most likely are missing from this scheme.

Discussion about MCPs

The MCPs that are responsible for detecting stimuli in the environment and transmitting the signal to CheA have a modular structure that is comprised of an N-terminal transmembrane domain (TM1), variable extracellular domain, second transmembrane region (TM2), HAMP domain and two pairs of C-terminal methylation (MH), and conserved SH domains (Bunn and Ordal 2003; Kristich et al. 2003; Zimmer et al. 2000).

Overall, our results showed that the positively selected sites mostly were found in the extracellular sensing domains. In each chemoreceptor, one selective residue was identified in or close to the transmembrane domain; McpB and McpC have common variable sites between HAMP and MH1 domains. Positively selected amino acid residues were identified also in the methylation helix and the transmembrane helix domains of McpB and McpC.

The McpC protein is the most variable between strains of *B. amyloliquefaciens* (Fig. 2). It contains 12 positively selected sites. These sites are at positions: 52, 71, 123, 169, 171, and 242 in the sensing domain; 276th in TM1 domain, 322th in HAMP domain, 354th between HAMP and MH1 domains and the residues 536, 647, and 651 in the C-terminal MH2 domain. The residue 536 is located close to the intracellular signaling domain-sensing carbohydrates of the phosphoenolpyruvate-dependent phosphotransferase system (PTS). Plant-colonizing behavior may be associated with carbohydrate sensing. It is known that in both *E. coli* and *B. subtilis*, the PTS of the transmembrane transport of carbohydrates is also involved in chemotaxis regulation towards the PTS carbohydrates (glucose, fructose, and mannitol); however, due to the inverted character of the chemotaxis system in *E. coli* and *B. subtilis*, the molecular mechanisms of this signaling pathway has to be different. It was hypothesized that in *E. coli* the PTS transport induces dephosphorylation of CheY that causes bacterial cells to swim smoothly

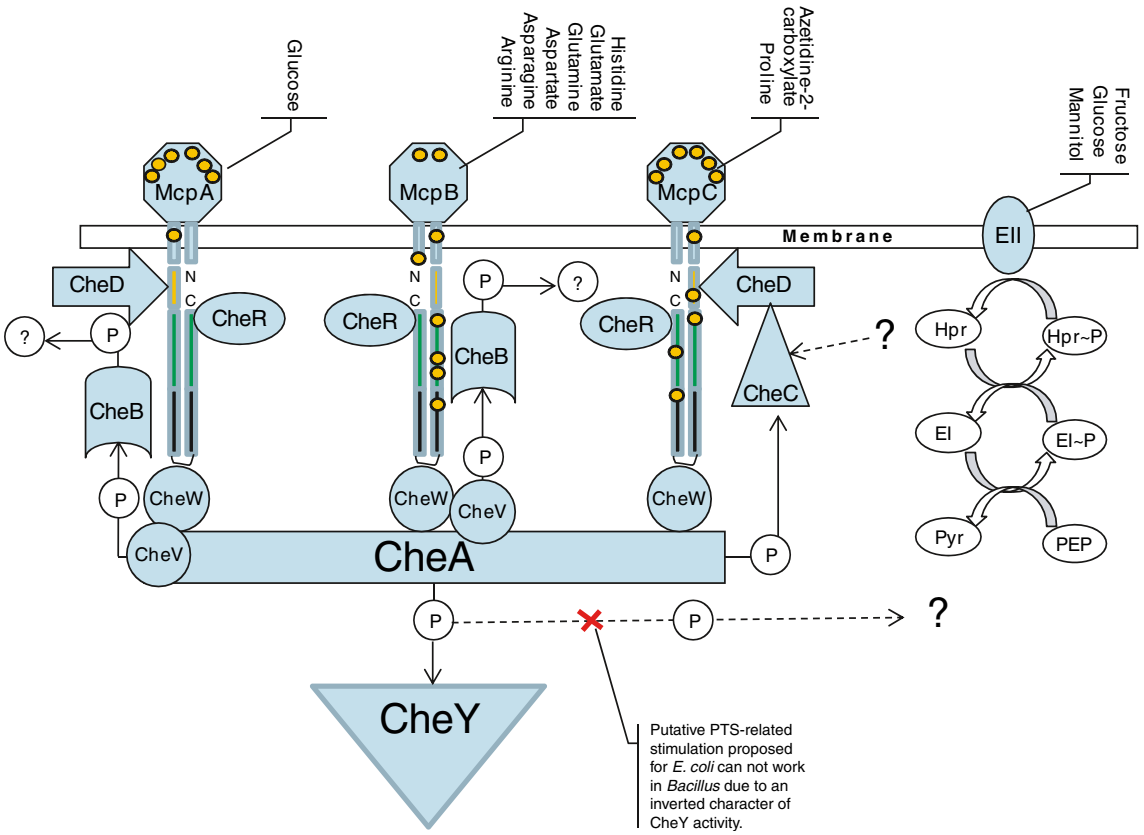


Fig. 5 The schema of interaction of chemotaxis proteins and chemoreceptors. In chemoreceptors, TM domains are shown as *blue bars*; HAMP domains—*yellow bars*; SH-domain—*black bars*; MH domains—*green bars*; and extracellular domains are shown as *polygons*. Positions of positively selected residues (Fig. 3A) are depicted by *orange-filled spheres*. Phosphate flow is shown by *plain arrows*. The phosphoenolpyruvate-dependent phosphotransferase system (PTS) is depicted by core elements of the PTS phosphorylation cascade: enzyme I (EI), enzyme II complex (EII), HPr, pyruvate (Pyr), and phosphoenolpyruvate (PEP). Possible links between PTS and chemotaxis systems in *B. subtilis* and *E. coli* are indicated by *dashed arrows*

towards carbohydrate attractants (Lux et al. 1999). This scheme is not applicable for *B. subtilis* as the dephosphorylation of CheY would lead to tumbling. In *B. subtilis*, the cytoplasmic domain of McpC is responsible for sensing PTS carbohydrates (Garrity et al. 1998; Kristich et al. 2003), while the N-terminal extracellular domain of McpC binds proline and senses this amino acid and several other compounds. There are two additional interacting chemotaxis proteins in *B. subtilis* CheD and CheC that are absent in *E. coli*. Knocking-out of the CheD protein impairs sensing of both proline and PTS carbohydrates. It was shown that CheD interacts specifically with the HAMP domain of McpC (Kristich and Ordal 2004); catalyzes amide hydrolysis of specific glutaminyl side chains of McpA (Kristich and Ordal 2002) and probably does not interact at all with McpB. It may be hypothesized that in *B. subtilis*, CheD may be a key element of the signal transduction flow from the PTS system to CheY and the flagellar motors, through the intracellular domain of McpC (Fig. 5).

Little is known about the receptor McpA except that it may be involved in sensing D-glucose. Seven positively selected residues were predicted in extracellular and TM2 domains of this protein at positions 95, 99, 112, 234, 235, 244, and 289.

The transmembrane organisation of the *B. subtilis* chemoreceptor McpB has been deduced by cysteine disulfide cross-linking (Bunn and Ordal 2003). The transmembrane segments TM1 and TM2 stretch from residues 17–34 and 280–302, respectively (Bunn and Ordal 2003). These transmembrane positions correspond to 17–34 and 279–301 in *B. amyloliquefaciens* FZB42. The transmembrane domain provides a structural link between the extracellular sensing domain and the cytoplasmic signaling domain; hence, any conformational changes induced by ligand binding to the sensing domain must be communicated through the TM2 domain (Bunn and Ordal 2003; Falke and Hazelbauer 2001). It is conceivable that a mutation under positive selection in this domain may confer an adaptational advantage to the organism, in terms of the efficiency in which a message from the receptors is communicated to the rest of the chemotactic machinery.

Three methylation sites have been identified in *B. subtilis* (Zimmer et al. 2000), which correspond to the residues E370,

Q629, and E636 in *B. amyloliquefaciens*. It is interesting that the first methylation site, which is a glutamine in *B. subtilis* and then later converted to a glutamate via deamidation; is encoded as glutamate in *B. amyloliquefaciens* and *B. pumilis*; therefore, it cannot undergo deamidation (Figs. 2 and 3). The residues at positions 376, 408, 412, and 478 of McpB fall in the MH domains and may interfere with the methylation and demethylation processes. In *E. coli*, the demethylation of receptors (all belong to class I) takes place when ligands are being released from the extracellular domains creating a negative stimulus. *B. subtilis* McpB receptor can sense both positive and negative stimuli (Zimmer et al. 2002).

Discussion about chemotaxis proteins

Loci of positive selection were also found in CheB (Fig. 2). CheB methylesterase has a multidomain architecture comprised of an N-terminal regulatory domain and a C-terminal effector domain which are joined by a linker region (Fig. 3). CheB accepts a phosphor group from CheA (Lupas and Stock 1989), thereby competing for phosphor residues with CheY that leads to increased probability of tumbling of the bacterial cell when its movement is directed downstream of the attractant gradient (Kirby et al. 2000). When CheB is activated by phosphorylation, it demethylates certain amino acid residues of McpB. This demethylation is important for adaptation to persistent stimuli (Goldman and Ordal 1984; Rao et al. 2008; Zimmer et al. 2000). In dephosphorylated CheB, the N-terminal domain packs tightly against the active site of the C-terminal domain, thereby restricting access to the active site causing inhibition of methylesterase activity (Djordjevic et al. 1998). Two positively selected sites were predicted in the N-terminal domains at positions 25 and 90, and another two residues in the C-terminal effector domain at positions 277 and 279 (Fig. 3). None of these sites is in close spatial proximity to the N-terminal phosphor binding site or the catalytic triad that consists of S171, H200, and D294 (Djordjevic et al. 1998). However, these amino acids are quite conserved in *B. subtilis* and *B. amyloliquefaciens* but vary between species. Alterations of these amino acids may modulate biding of CheB to CheA and McpB, and in this way, tune compensatory response of bacteria to attractant removal.

The coupling protein CheW has one positively selected site at position 18. This residue is located on the surface of CheW opposite to the area that interacts with CheA domains P4 and P5. No amino acid residues under positive selection were identified in CheA, CheR, and CheV proteins.

Binding of CheC to CheD was proved experimentally (Chao et al. 2006) (see Fig. 4D). On the other hand, CheC has two similar α-helixes that has been proven to bind to CheY to dephosphorylate it, with the second active site showing higher activity than the first one (Chao et al. 2006; Muff and Ordal 2007; Park et al. 2004a, b). A stronger phosphatase activity was predicted for the C-proximal helix. It contains more negatively selected amino acid residues and, probably, is a more likely candidate for binding to the corresponding CheY domain (Fig. 4C). CheY and CheD binding sites are opposite on the CheC molecule. The affinity that CheC has for CheD is enhanced in the presence of CheYp and at the same time the phosphatase activity of CheC increases when bound to CheD (Muff and Ordal 2007; Rao et al. 2008). Binding of CheD to CheC recruits CheD away from the receptors and obstructs the CheD's active site (Chao et al. 2006). It was hypothesized that CheC may serve as a depot of CheD that may be released by a signal coming from CheY to CheC (Rao et al. 2008).When CheD is released from the MCPs, the conformation changes in such a way that CheA kinase activity is inhibited, and this feedback loop causes levels of CheYp to decrease (Rao et al. 2008). The presence of the additional α-helixes in CheC that remains unused in this scheme suggests that this circuit may involve other yet unknown components that may be PTS-related chemicals. The loss of CheC impacts chemotaxis globally by a considerable reduction of switching frequency between CCW and CW rotation that makes the chemotaxis process less sensitive (Saulmon et al. 2004). However, several bacterial genomes are known where CheD is present but CheC is absent that suggests CheC is not an obligatory part of *Bacillus* chemotaxis but an important regulatory element (Kirby et al. 2001). The residue at position 86 is under positive selection and variable in *B. amyloliquefaciens* strains (Fig. 2). This position is not part of the interfaces that interact with CheD or CheY and may provide the protein with a novel property.

In conclusion, future reconstruction of the spatial organization of the whole chemotaxis system of plant-colonizing *Bacillus* and self-regulatory mechanisms of bacterial behavior, will aid in better understanding of the growth-promoting activity of industrially important *Bacillus* group.

Acknowledgements AY thanks Rhodes University and National Research Foundation (NRF) of South Africa for financial support, OR's work was funded by the NRF grant 71261. We thank Prof. R. Borriss (Humboldt University, Berlin) for providing us with newly sequenced *B. amyloliquefaciens* genomes. We also thank Prof. N. Bishop and Dr. T. Bray for kindly proofreading the manuscript. AY and OTB started this work at the Bioinformatics and Computational Biology Unit, University of Pretoria; then continued at Rhodes University, Grahamstown.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Borriss R, Chen X, Rueckert C, Blom J, Becker A, Baumgarth B, Fan B, Pukall R, Schumann P, Spröer C, Junge H, Vater J, Pühler A, Klenk HP (2010) Relationship of *Bacillus amyloliquefaciens* clades associated with strains DSM7T and FZB42: a proposal for *Bacillus amyloliquefaciens* subsp. *amyloliquefaciens* subsp. nov. and *Bacillus amyloliquefaciens* subsp. *plantarum* subsp. nov. based on their discriminating complete genome sequences. Int J Syst Evol Microbiol. doi:10.1099/ijs.0.023267-0

Buchan A, Crombie B, Alexandre GM (2010) Temporal dynamics and genetic diversity of chemotactic-competent microbial populations in the rhizosphere. Environ Microbiol 12(12):3171–3184

Bunn MW, Ordal GW (2003) Transmembrane organization of the *Bacillus subtilis* chemoreceptor McpB deduced by cysteine disulfide crosslinking. J Mol Biol 331:941–949

Chao X, Muff TJ, Park S, Zhang S, Pollard AM, Ordal GW, Bilwes AM, Crane BR (2006) A receptor-modifying deamidase in complex with a signaling phosphatase reveals reciprocal regulation. Cell 124:561–571

Chen XH, Koumoutsi A, Scholz R, Eisenreich A, Schneider K, Heinemeyer I, Morgenstern B, Voss B, Hess WR, Reva O, Junge H, Voigt B, Jungblut PR, Vater J, Süssmuth R, Liesegang H, Strittmatter A, Gottschalk G, Borriss R (2007) Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. Nat Biotech 25:1007–1014

Chen XH, Koumoutsi A, Scholz R, Schneider K, Vater J, Süssmuth R, Piel J, Borriss R (2009) Genome analysis of *Bacillus amyloliquefaciens* FZB42 reveals its potential for biocontrol of plant pathogens. J Biotechnol 140:27–37

Choudhary DK, Johri BN (2009) Interactions of *Bacillus* spp. and plants—with special reference to induced. Microbiol Res 164 (5):493–513

Chun J, Bae KS (2000) Phylogenetic analysis of *Bacillus subtilis* and related taxa based on partial *gyrA* gene sequences. Antonie van Leeuwenhoek 78:123–127

Djordjevic S, Stock AM (1997) Crystal structure of the chemotaxis receptor methyltransferase CheR suggests a conserved structural motif for binding *S*-adenosylmethionine. Structure 5:545–558

Djordjevic S, Goudreau PN, Xu Q, Stock AM, West AH (1998) Structural basis for methylesterase CheB regulation by a phosphorylation-activated domain. Proc Natl Acad Sci USA 95:1381–1386

Doron-Faigenboim A, Pupko T (2007) A combined empirical and mechanistic codon model. Mol Biol Evol 24:388–397

Doron-Faigenboim A, Stern A, Mayrose I, Bacharach E, Pupko T (2005) Selecton: a server for detecting evolutionary forces at a single amino-acid site. Bioinformatics 21:2101–2103

Edgar R (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113

Eisenberg D, Lüthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396–404

Falke JJ, Hazelbauer GL (2001) Transmembrane signaling in bacterial chemoreceptors. Trends Biochem Sci 26:257–265

Felsenstein J, Churchill GA (1996) A Hidden Markov Model approach to variation among sites in rate of evolution. Mol Biol Evol 13:93–104

Fritze D (2004) Taxonomy of the genus *Bacillus* and related genera: the aerobic endospore-forming bacteria. Phytopathology 94:1245–1248

Garrity LF, Ordal GW (1995) Chemotaxis in *Bacillus subtilis*: how bacteria monitor environmental signals. Pharmacol Ther 68:87–104

Garrity LF, Schiel SL, Merrill R, Reizer J, Saier MH Jr, Ordal GW (1998) Unique regulation of carbohydrate chemotaxis in *Bacillus subtilis* by the phosphoenolpyruvate-dependent phosphotransferase system and the methyl-accepting chemotaxis protein McpC. J Bacteriol 180:4475–4480

Goldman DJ, Ordal GW (1984) In vitro methylation and demethylation of methyl-accepting chemotaxis proteins in *Bacillus subtilis*. Biochemistry 23:2600–2606

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hamer R, Chen P, Armitage J, Reinert G, Deane C (2010) Deciphering chemotaxis pathways using cross species comparisons. BMC Systems Biology 4:3

Hernández-Sánchez C, Mansilla A, de Pablo F, Zardoya R (2008) Evolution of the insulin receptor family and receptor isoform expression in vertebrates. Mol Biol Evol 25:1043–1053

Kirby JR, Niewold TB, Maloy S, Ordal GW (2000) CheB is required for behavioural responses to negative stimuli during chemotaxis in *Bacillus subtilis*. Mol Microbiol 35:44–57

Kirby JR, Kristich CJ, Saulmon M, Zimmer MA, Garrity L, Zhulin IB, Ordal G (2001) CheC is related to the family of flagellar switch proteins and acts independently from CheD to control chemotaxis in *Bacillus subtilis*. Mol Microbiol 42:573–585

Kristich CJ, Ordal GW (2002) *Bacillus subtilis* CheD is a chemoreceptor modification enzyme required for chemotaxis. J Biol Chem 277:25356–25362

Kristich CJ, Ordal GW (2004) Analysis of chimeric chemoreceptors in *Bacillus subtilis* reveals a role for CheD in the function of the McpC HAMP domain. J Bacteriol 186:5950–5955

Kristich CJ, Glekas GD, Ordal GW (2003) The conserved cytoplasmic module of the transmembrane chemoreceptor McpC mediates carbohydrate chemotaxis in *Bacillus subtilis*. Mol Microbiol 47:1353–1366

Lupas A, Stock J (1989) Phosphorylation of an N-terminal regulatory domain activates the CheB methylesterase in bacterial chemotaxis. J Biol Chem 264:17337–17342

Lux R, Munasinghe VRN, Castellano F, Lengeler JW, Corrie JET, Khan S (1999) Elucidation of a PTS-carbohydrate chemotactic signal pathway in *Escherichia coli* using a time-resolved behavioral assay. Mol Biol Cell 10:1133–1146

Manson MD, Armitage JP, Hoch JA, Macnab RM (1998) Bacterial locomotion and signal transduction. J Bacteriol 180:1009–1022

Mowbray SL, Sandgren MOJ (1998) Chemotaxis receptors: a progress report on structure and function. J Struct Biol 124:257–275

Muff TJ, Ordal GW (2007) The CheC phosphatase regulates chemotactic adaptation through CheD. J Biol Chem 282:34120–34128

Park S, Chao X, Gonzalez-Bonet G, Beel BD, Bilwes AM, Crane BR (2004a) Structure and function of an unusual family of protein phosphatases: the bacterial chemotaxis proteins CheC and CheX. Mol Cell 16:563–574

Park S, Beel BD, Simon MI, Bilwes AM, Crane BR (2004b) In different organisms, the mode of interaction between two signaling proteins is not necessarily conserved. Proc Natl Acad Sci USA 101:11646–11651

Park S, Borbat PP, Gonzalez-Bonet G, Bhatnagar J, Pollard AM, Freed JH, Bilwes AM, Crane BR (2006) Reconstruction of the chemotaxis receptor-kinase assembly. Nat Struct Mol Biol 13:400–407

Pawlowski M, Gajda M, Matlak R, Bujnicki J (2008) MetaMQAP: a meta-server for the quality assessment of protein models. BMC Bioinformatics 9:403

Pazy Y, Motaleb MA, Guarnieri MT, Charon NW, Zhao R, Silversmith RE (2010) Identical phosphatase mechanisms achieved through distinct modes of binding phosphoprotein substrate. Proc Natl Acad Sci USA 107:1924–1929

Pei J, Kim B, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucl Acids Res 36:2295–2300

Rao CV, Kirby JR, Arkin AP (2004) Design and diversity in bacterial chemotaxis: a comparative study in *Escherichia coli* and *Bacillus subtilis*. PLoS Biol 2:e49

Rao CV, Glekas GD, Ordal GW (2008) The three adaptation systems of *Bacillus subtilis* chemotaxis. Trends Microbiol 16:480–487

Reva ON, Dixelius C, Meijer J, Priest FG (2004) Taxonomic characterization and plant colonizing abilities of some bacteria related to *Bacillus amyloliquefaciens* and *Bacillus subtilis*. FEMS Microbiol Ecol 48:249–259

Rosario MM, Fredrick KL, Ordal GW, Helmann JD (1994) Chemotaxis in *Bacillus subtilis* requires either of two functionally redundant CheW homologs. J Bacteriol 176:2736–2739

Rosario MML, Kirby JR, Bochar DA, Ordal GW (1995) Chemotactic methylation and behavior in *Bacillus subtilis*: role of two unique proteins, CheC and CheD. Biochemistry 34:3823–3831

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815

Saulmon MM, Karatan E, Ordal GW (2004) Effect of loss of CheC and other adaptational proteins on chemotactic behaviour in *Bacillus subtilis*. Microbiology 150:581–589

Shen M, Sali A (2006) Statistical potential for assessment and prediction of protein structures. Protein Sci 15:2507–2524

Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucl Acids Res 33:W244–W248

Tastan Bishop O, Kroon M (2011) Study of protein complexes via homology modeling, applied to cysteine proteases and their protein inhibitors. J Mol Model. doi:10.1007/s00894-011-0990-y

Usher KC, De La Cruz AFA, Dahlquist FW, James Remington S, Swanson RV, Simon MI (1998) Crystal structures of CheY from *Thermotoga maritima* do not support conventional explanations for the structural basis of enhanced thermostability. Protein Sci 7:403–412

Van de Broek A, Lambrecht M, Vanderleyden J (1998) Bacterial chemotactic motility is important for the initiation of wheat root colonization by *Azospirillum brasilense*. Microbiology 144:2599–2606

Wallner B, Elofsson A (2003) Can correct protein models be identified? Protein Sci 12:1073–1086

Yao J, Allen C (2006) Chemotaxis is required for virulence and competitive fitness of the bacterial wilt pathogen *Ralstonia solanacearum*. J Bacteriol 188:3697–3708

Yao W, Shi L, Liang D (2007) Crystal structure of scaffolding protein CheW from *Thermoanaerobacter tengcongensis*. Biochem Biophys Res Commun 361:1027–1032

Zhang W, Phillips GN (2003) Structure of the oxygen sensor in *Bacillus subtilis*: signal transduction of chemotaxis by control of symmetry. Structure 11:1097–1110

Zimmer MA, Tiu J, Collins MA, Ordal GW (2000) Selective methylation changes on the *Bacillus subtilis* chemotaxis receptor McpB promote adaptation. J Biol Chem 275:24264–24272

Zimmer MA, Szurmant H, Saulmon MM, Collins MA, Bant JS, Ordal GW (2002) The role of heterologous receptors in McpB-mediated signalling in *Bacillus subtilis* chemotaxis. Mol Microbiol 45:555–568