

Different Evolutionary Constraints on Chemotaxis Proteins CheW and CheY Revealed by Heterologous Expression Studies and Protein Sequence Analysis

Gladys Alexandre† and Igor B. Zhulin*

School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332-0230

Received 14 June 2002/Accepted 18 October 2002

CheW and CheY are single-domain proteins from a signal transduction pathway that transmits information from transmembrane receptors to flagellar motors in bacterial chemotaxis. In various bacterial and archaeal species, the *cheW* and *cheY* genes are usually encoded within homologous chemotaxis operons. We examined evolutionary changes in these two proteins from distantly related proteobacterial species, *Escherichia coli* and *Azospirillum brasilense*. We analyzed the functions of divergent CheW and CheY proteins from *A. brasilense* by heterologous expression in *E. coli* wild-type and mutant strains. Both proteins were able to specifically inhibit chemotaxis of a wild-type *E. coli* strain; however, only CheW from *A. brasilense* was able to restore signal transduction in a corresponding mutant of *E. coli*. Detailed protein sequence analysis of CheW and CheY homologs from the two species revealed substantial differences in the types of amino acid substitutions in the two proteins. Multiple, but conservative, substitutions were found in CheW homologs. No severe mismatches were found between the CheW homologs in positions that are known to be structurally or functionally important. Substitutions in CheY homologs were found to be less conservative and occurred in positions that are critical for interactions with other components of the signal transduction pathway. Our findings suggest that proteins from the same cellular pathway encoded by genes from the same operon have different evolutionary constraints on their structures that reflect differences in their functions.

In *Escherichia coli*, the signal transduction pathway for chemotaxis consists of specialized membrane receptors, termed chemotaxis transducers; a CheA-CheY two-component system, which transmits the signal from transducers to flagellar motors; and a docking protein, CheW, which couples the CheA histidine kinase to transducers (for recent reviews, see references 7, 8, and 11). In addition, two proteins, the CheB methyl-esterase and the CheR methyltransferase, comprise an adaptation pathway for chemotaxis. Interaction of CheA with chemotaxis transducers changes the rate of kinase phosphorylation (4). The phosphoryl group on CheA can be transferred to CheY or CheB (18). In a phosphorylated form, CheY binds to the flagellar switch protein FliM and causes a change in the direction of flagellar rotation from counterclockwise to clockwise (43).

Homologous pathways (transducers, CheA, CheW, CheY, CheB, and CheR) govern chemotaxis in all motile bacterial and archaeal species (46). Series of mutational, biochemical, and structural studies have addressed the importance of protein-protein interactions within the signal transduction cascade; for many of these, however, the exact mechanisms remain unclear. Identification of functional residues in interacting proteins is of critical importance for understanding mechanisms of signal transduction. Comparison of evolutionarily distant homologs may be very useful for this purpose.

Proteins are robust to site mutations, and during evolution they may accumulate significant numbers of substitutions with little or no change in structure or function (39). Thus, a distant homolog can be viewed as a result of evolutionarily derived mutagenesis carefully selected for maintaining the function within a particular cellular context. If a component of a signal transduction pathway can be replaced by a homologous element in a distantly related species, that would imply that functionally important residues were preserved during evolution. Therefore, heterologous expression of a “foreign” signal transduction element in *E. coli* may be a useful tool in searching for functional residues in interacting proteins. Like any other approach, it has its limitations. For example, as in laboratory mutagenesis, evolutionary mutations can be compensated for by second-site suppressors acting on the interacting partner. In such a case, heterologous expression of a distant homolog would not result in restoration of function.

Here, we demonstrate that two single-domain proteins from the chemotaxis signal transduction pathway are substantially different with respect to selective pressure on their properties. We used the docking protein CheW and the chemotaxis response regulator CheY from an α -proteobacterium, *Azospirillum brasilense*, to complement corresponding mutants of *E. coli*. Both *A. brasilense* proteins are significantly divergent from their *E. coli* counterparts, but they have essentially the same functions (16). Surprisingly, the more diverged CheW protein allowed the chemotactic signal to occur in the corresponding *E. coli* mutant, whereas we found no such evidence for CheY. Detailed pairwise comparison of protein sequences revealed different patterns and types of amino acid substitutions in the CheW and CheY. The results provide new insights into the

* Corresponding author. Mailing address: School of Biology, Georgia Institute of Technology, 310 Ferst Dr., Atlanta, GA 30332-0230. Phone: (404) 385-2224. Fax: (404) 894-0519. E-mail: igor.zhulin@biology.gatech.edu.

† Present address: Department of Biology, Georgia State University, Atlanta, GA 30303.

TABLE 1. Strains and plasmids used in this study

| Strain or plasmid | Genotype, phenotype, or description | Reference or source |
|--------------------------|--|---------------------|
| Strains | | |
| <i>A. brasilense</i> Sp7 | Wild-type strain | ATCC 29145 |
| <i>E. coli</i> | | |
| DH5 α | General cloning strain | Gibco BRL |
| RP437 | Wild-type chemotaxis strain | J. S. Parkinson |
| RP1616 | Δ <i>cheZ</i> ; tumbly | J. S. Parkinson |
| RP4606 | Δ <i>cheW</i> ; smooth swimming | J. S. Parkinson |
| RP5232 | Δ <i>cheY</i> ; smooth swimming | J. S. Parkinson |
| RP2768 | <i>cheY201</i> ; smooth swimming | J. S. Parkinson |
| Plasmids | | |
| pFAJ451 | pLAFR1 derivative containing 25 kb of <i>A. brasilense</i> DNA, including the chemotaxis operon | 16 |
| pTrc99A | Expression vector; <i>Ptac</i> -based promoter; Ap ^r | Amersham |
| pProEXHTa | Expression vector; <i>Ptac</i> -based promoter; 6-His N-terminal tag to the expressed protein; Ap ^r | Gibco BRL |
| pIZ101 | <i>cheW</i> gene from <i>A. brasilense</i> cloned as an <i>NcoI/HindIII</i> DNA fragment in pTrc99A | This work |
| pIZ103 | <i>cheY</i> gene from <i>A. brasilense</i> cloned as a <i>BamHI/XhoI</i> DNA fragment in pProEXHTa | This work |
| pIZ103D52N | Derivative of pIZ103, carrying a mutant <i>cheY</i> gene which encodes the mutated protein CheYD52N | This work |
| pIZ109 | <i>cheY</i> from <i>A. brasilense</i> cloned as an <i>EcoRI/HindIII</i> DNA fragment in pTrc99A | This work |

mechanism of signal transduction and the molecular evolution of the signal transduction pathway.

MATERIALS AND METHODS

Bacterial strains and plasmids. Bacterial strains and plasmids used in this study are listed in Table 1. *E. coli* strains were grown aerobically at 30°C in Luria-Bertani (LB) medium containing thiamine (1 mM) and supplemented with ampicillin (100 μ g ml⁻¹) for plasmid selection.

Plasmid construction. All standard cloning steps were carried out as described by Sambrook et al. (30). Plasmid DNA was isolated from *E. coli* by using the Qiagen spin column kit. DNA was isolated from agarose gels by using the Qiaquick gel extraction kit (Qiagen). Enzymes were used according to the manufacturers' directions (New England Biolabs, Roche Applied Science, and Stratagene). The *cheW* gene was amplified from cosmid pFAJ451 (16) with primer cheW-For (GGGCCATGGCTAGCAACGCAAGCTGCCCGCC), which includes an *NcoI* site (underlined), and primer cheW-Rev (GGGAAAGCTTTCAGGCCGCTCCATCGTGGT), which includes a *HindIII* site (underlined). The PCR fragment (557 bp) corresponding to the amplified *cheW* gene was gel purified and digested with *NcoI* and *HindIII* followed by ligation into the pTrc99A vector digested with the same enzymes, generating pIZ101. The *cheY* gene was amplified from the pFAJ451 cosmid by using CheY-For (GGAAATTCAAAGTTTGTCTGGTCTCGA), which includes an *EcoRI* site (underlined), and CheY-Rev (CCCAAGCTTTCACAGCAGCCCGACCTGCTC), which includes a *HindIII* site (underlined) and was cloned into the pTrc99A vector digested with *EcoRI* and *HindIII* to generate pIZ109. In addition, the *cheY* gene was amplified from the pFAJ451 cosmid and cloned into the pProEXHTa vector by using the following primers: cheYBam-For (GGGGGATCCGAAGTTTGTGGTCTCGTCTCGA), which includes a *BamHI* site (underlined), and cheYXho-Rev (GGGCTCGAGTCACAGCAGCCCGACTGCTC), which includes an

XhoI site (underlined). The plasmid generated was called pIZ103. All PCR amplifications were performed using the Expand High Fidelity System (Roche Applied Science) under the conditions recommended by the manufacturer. Constructs were transformed into *E. coli* competent cells by standard heat shock procedures (30). All constructs were confirmed by automated DNA sequencing.

Mutagenesis. Mutations were generated by *Pfu* polymerase PCR by using the Quick-Change Site-Directed Mutagenesis kit (Stratagene) according to the manufacturer's recommendations. The putative phosphorylation site of the CheY protein from *A. brasilense* (D52, corresponding to D57 in the *E. coli* protein) was identified by alignment with CheY from *E. coli*. The following primers, replacing the GAC triplet encoding D52 in the CheY protein from *A. brasilense* with the AAC triplet (underlined) encoding asparagine, were used: cheYD52N-For (CGCCATCCCTGCTGAAACTGGAACATGCCG) and cheYD52N-Rev (CGGGCA TGTCCAGTTTCAGCAGGATGGCG). Thermocycling was carried out in a Perkin-Elmer Thermocycler. Site-specific mutations were verified by DNA sequencing using ABI Prism dye terminator cycle sequencing.

Detection of protein expression. Proteins were expressed from the pTrc99A vector in order to allow expression of the native proteins. Induction of the proteins was confirmed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis after Coomassie brilliant blue staining. The wild-type and mutant *cheY* genes were also expressed in the pProEXHTa vector from an isopropylthiogalactoside (IPTG)-inducible *tac* promoter. In this expression vector, proteins are expressed with an in-frame six-histidine N-terminal tag. INDIA HisProbe-HRP (Pierce Chemicals), a nickel-activated derivative of horseradish peroxidase (HRP), was used to directly detect the blotted recombinant polyhistidine-tagged CheY fusion proteins. The HRP was detected by using the SuperSignal West Pico chemiluminescent substrate (Pierce Chemicals) as recommended by the manufacturer. Recombinant proteins were induced according to the manufacturer's directions by using varying amounts of IPTG (Gibco BRL).

Behavioral assays. To analyze *E. coli* behavior, spatial (swarm plates) and temporal gradient assays were used essentially as previously described (2). Because of the smooth-swimming bias of the *cheY* and *cheW* *E. coli* mutants, temporal responses to the addition of the repellent leucine at 10 mM were measured. Upon adaptation, cells resume their smooth-swimming bias.

Protein sequence analysis. Pairwise comparisons of protein sequences were performed using the BLAST 2 Sequences program with default parameters (38). Sequence similarity plots were produced from pairwise alignments based on the PAM250 (10) and BLOSUM62 (17) scoring matrices. Multiple alignments were constructed using the CLUSTAL X program (40). Secondary-structure predictions were carried out using the JPRED2 server (9) and the PHD program (29).

RESULTS

CheW protein from *A. brasilense* restores the chemotactic signal to the *E. coli cheW* mutant. A Δ *cheW* mutant of *E. coli* is incapable of chemotaxis. It has a smooth-swimming phenotype, which remains constant upon addition of any chemoefactor (27). Expression of the CheW protein from a plasmid restores chemotaxis to a Δ *cheW* strain (21, 31). Elevated levels of CheW inhibit chemotaxis; therefore, its expression should be tightly controlled (5, 21, 31). We used the pTrc99a vector (1) to enable IPTG-inducible expression of the *A. brasilense cheW* gene in both the *cheW* null mutant (RP4606) and wild-type (RP437) *E. coli*. We first tested the effect of the heterologous expression in a swarm plate assay. The *A. brasilense* protein restored the swarming ability of the mutant and, as expected, inhibited chemotaxis in the wild type (Fig. 1A). Reduction of the swarm size of the wild type correlated with induction levels and was statistically significant (Fig. 1B), whereas there was no effect on the cell growth rate under these conditions. Restoration of swarming in the mutant was clearly observed, but only under specific conditions of induction (50 to 100 μ M IPTG and 22°C). Restoration of swarming per se does not necessarily mean restoration of the chemotactic response. For example, the increased motility on swarm plates may be due to so-called "pseudotaxis" (44). Several observations indicated that the restoration of swarming in the complemented

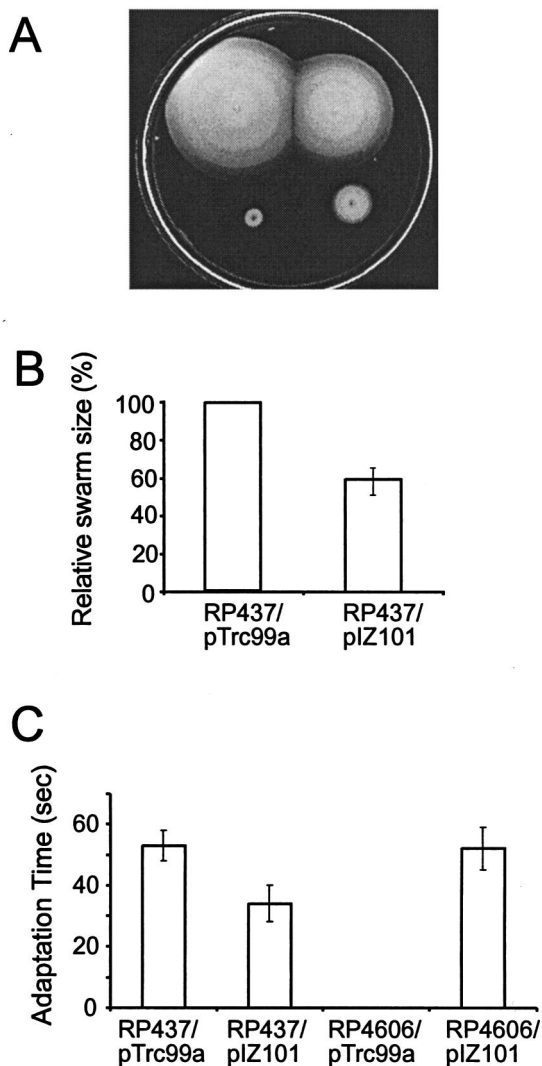


FIG. 1. Heterologous expression of the *A. brasilense* CheW protein in *E. coli*. (A) Swarm plate assay. Conditions were LB medium, 100 μ M IPTG, 22°C, and 12 h of incubation. Strains, clockwise from top left, are RP437/pTrc99a, RP437/pIZ101, RP4606/pIZ101, and RP4606/pTrc99a. (B) Inhibition of chemotaxis of wild-type cells by *A. brasilense* CheW. Relative sizes of swarms are shown. Conditions were LB medium, 500 μ M IPTG, 30°C, and 12 h of incubation. (C) Temporal gradient assay. Cells were induced with 100 μ M IPTG. Times for adaptation to 10 mM leucine are shown.

mutant was the result of true chemotaxis and not pseudotaxis: (i) swarms always had sharp edges (chemotaxis rings), (ii) cells taken from the edge of the swarm had a smooth-swimming bias, whereas in pseudotaxis the ability to swarm is proportional to the percentage of tumbling cells (44), and (iii) cells taken directly from the swarm exhibited chemotaxis in the temporal gradient assay (see below).

In order to conclusively establish that the *A. brasilense* CheW protein can restore the chemotaxis signal from transducers to the flagellar motor in the *cheW* mutant, we carried out a quantitative temporal gradient assay, where changes in behavior of free-swimming cells in response to a chemoeffector are recorded. *E. coli* mutant cells complemented with the *A.*

brasilense cheW gene still exhibited a smooth-swimming bias, which made it difficult to measure chemotaxis to attractants. However, in response to the repellent leucine, which is detected by the Tsr transducer, the cells had a 1-min tumbling response, typical of the wild type (Fig. 1C). Furthermore, when the attractant serine (1 mM) was added to cells that had been exposed to leucine for 15 s and therefore had a 100% tumbling bias, the cells immediately displayed a 100% smooth-swimming bias. Similar responses were observed upon addition of other repellents and attractants (data not shown). These results clearly establish that the CheW protein from *A. brasilense* is able to interact with transducers and the CheA protein of the *E. coli* chemotaxis machinery and to transduce the chemotactic signal, as was suggested for the *E. coli* protein (13).

CheY protein from *A. brasilense* interacts with the chemotaxis machinery of *E. coli* but does not restore chemotaxis to the *cheY* mutant. As in the experiments described above, wild-type *E. coli* (RP437) and the *cheY* null mutant (RP5232) were transformed with plasmids expressing the *cheY* gene from *A. brasilense* under the control of the inducible *Trc* promoter. Because the mutation in strain RP5232 is polar on the downstream *cheZ* gene (J. S. Parkinson, personal communication), experiments were also carried out in parallel with another *cheY* mutant (RP2768). Expression of the CheY protein from *A. brasilense* inhibited chemotaxis in wild-type *E. coli* in both swarm plate (Fig. 2A) and temporal gradient (data not shown) assays. The inhibitory effect was best observed at high (0.5 to 1 mM) IPTG concentrations, which still did not affect cell growth. No inhibition was observed without IPTG, suggesting that the effect was due to the expressed heterologous protein.

The CheY protein from *A. brasilense* did not complement the defect in either of the *cheY* mutants of *E. coli* when tested under a wide range of experimental conditions in either swarm plate or temporal gradient assays (Fig. 2A). There was also no effect on the swimming bias of the *cheY* mutants. In order to verify expression of the recombinant protein, the *cheY* gene from *A. brasilense* was cloned into the pProEx expression vector, which carries a six-His tag. Expression of the CheY protein was monitored with the INDIA His probe, as described in Materials and Methods. The CheY protein was expressed in all strains tested, and its expression levels correlated with the IPTG concentrations (Fig. 2B). All heterologous expression experiments were then repeated with the His-tagged protein and yielded the same results as with the untagged protein: no restoration of chemotaxis in the *cheY* mutant and statistically significant inhibition of chemotaxis in wild-type cells (Fig. 2C and D). We therefore interpreted the lack of complementation as resulting from an insufficient match between the *A. brasilense* CheY protein and the *E. coli* chemotactic machinery. However, the inhibition of wild-type chemotaxis suggested that the CheY protein from *A. brasilense* interacted with the *E. coli* chemotaxis proteins. The *A. brasilense* CheY contains a conserved phosphoacceptor site (Asp52, corresponding to Asp57 in the *E. coli* homolog) (16). In order to test the hypothesis that the *A. brasilense* CheY inhibits chemotaxis in wild-type *E. coli* by competing for a phosphate with its homolog, we replaced the aspartate residue at position 52 with asparagine in the *cheY* gene from *A. brasilense* cloned into the pProEx vector. The plasmid was then transformed into the wild-type *E. coli* strain. Expression levels of the wild-type and mutant CheY proteins

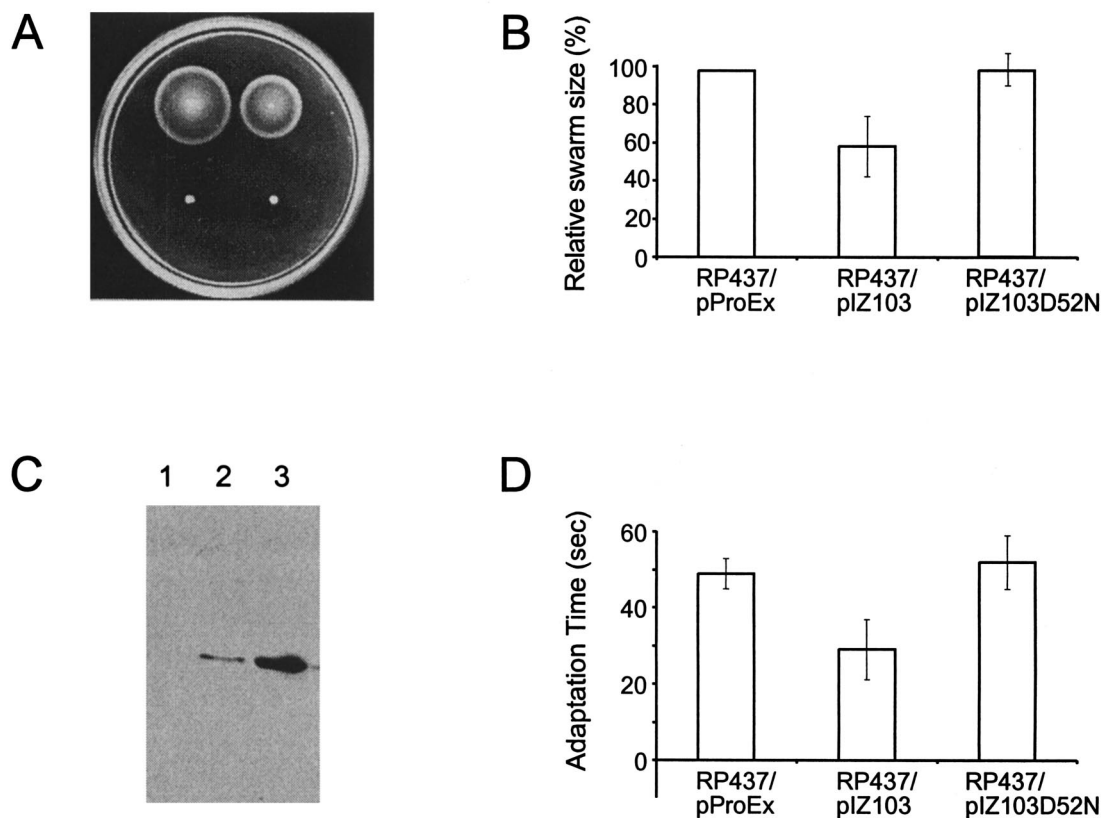


FIG. 2. Heterologous expression of the *A. brasilense* CheY protein in *E. coli*. (A) Swarm plate assay. Conditions were LB medium, 500 μ M IPTG, 30°C, and 12 h of incubation. Strains, clockwise from top left, are RP437/pProEx, RP437/pIZ103, RP5232/pIZ103, and RP5232/pProEx. (B) Expression of *A. brasilense* CheY in *E. coli*. Relative sizes of swarms are shown. (C) Western blot analysis of *A. brasilense* CheY expression in *E. coli* RP5232. IPTG concentrations: 0 μ M (lane 1), 250 μ M (lane 2), and 1 mM (lane 3). (D) Temporal gradient assay. Cells were induced with 500 μ M IPTG. Times for adaptation to 10 mM leucine are shown.

were similar. In contrast to the wild-type protein, the mutant CheY protein had no effect on chemotaxis of wild-type *E. coli* cells at different expression levels in both spatial and temporal gradient assays (Fig. 2C and D). These results clearly indicate that CheY of *A. brasilense*, when expressed in *E. coli*, does interact with its chemotaxis machinery. This serves as a positive control for negative results for complementation of the *E. coli* *cheY* mutant and strengthens our conclusion that the apparent lack of complementation is due to the insufficient match between the *A. brasilense* and *E. coli* proteins. The results can be interpreted further as suggesting that CheA of *E. coli* can phosphorylate CheY of *A. brasilense* in vivo. However, the relatively poor inhibition of wild-type chemotaxis also suggests that the interaction between CheA and CheY in the heterologous system is weak.

Heterologous expression studies with multiple CheY homologs from *Rhodobacter sphaeroides* demonstrated that while these proteins do not restore chemotaxis to the *E. coli* *cheY* mutant, they complement *cheZ* mutants, probably by competing for phosphate with the *E. coli* CheY protein (33). We expressed CheY from *A. brasilense* in the *cheZ* null mutant strain RP1616 from both the pIZ109 and pIZ103 plasmids. Expression of the *A. brasilense* CheY protein was confirmed by using the INDIA-His probe and by Coomassie blue staining. CheY from *A. brasilense* did not complement the defect in

RP1616 over a wide range of IPTG inductions that did not affect cell growth (data not shown).

Protein sequence comparisons of CheW and CheY homologs. Pairwise alignment of CheW protein sequences from *E. coli* and *A. brasilense* revealed 26% identity and 52% similarity based on default parameters of the BLAST 2 Sequences program (38). The same analysis revealed 34% identity and 53% similarity between the CheY proteins. Thus, the *A. brasilense* CheY protein shows a significantly higher degree of identity with its *E. coli* homolog than does CheW, whereas levels of similarity appear to be the same in both cases. We further evaluated the alignments with an emphasis on the types of amino acid substitutions between homologous proteins. The alignments were translated into plots (Fig. 3), where a score was assigned to each residue position based on the amino acid substitution matrix, such as PAM150 (10) or BLOSUM62 (17).

Comparison of the two plots reveals differences between types of amino acid substitutions in CheW and CheY homologs. The diapason of score variation is wider for CheY homologs. Scores of 8 or higher are assigned to five positions in the CheY alignment (none in the CheW alignment), and scores of -6 or lower are assigned to six positions in the CheY alignment (none in the CheW alignment). "Severe" mismatches (scores of -3 or lower) are found only in 12 positions in the CheW alignment (7%), whereas in the CheY alignment

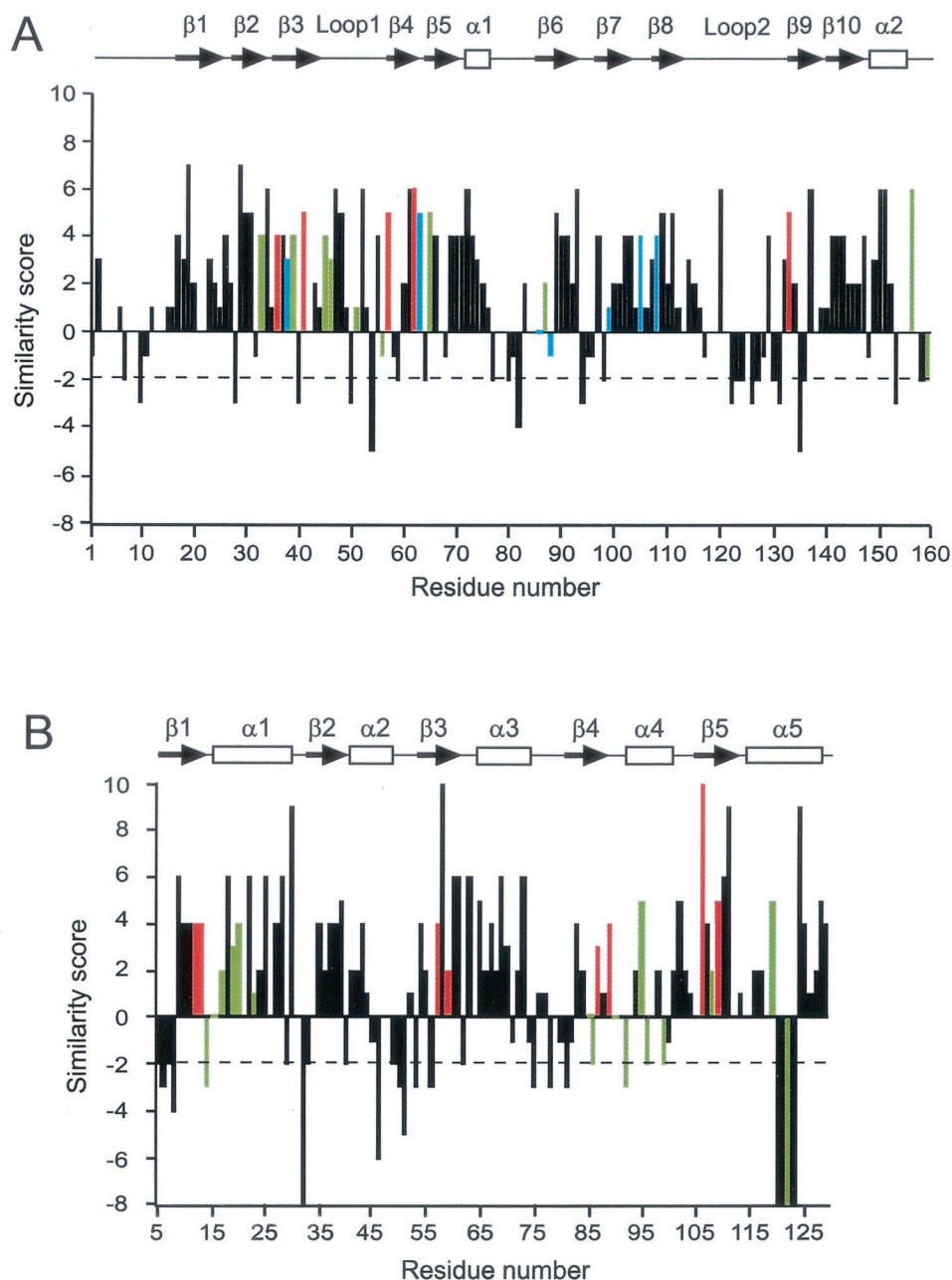


FIG. 3. Plots of sequence similarity between *E. coli* and *A. brasilense* proteins. Positions in the sequence alignment are shown along the x axis; residue numbers given are for *E. coli* sequences. Scores based on the PAM150 amino acid substitution matrix (10) are shown along the y axis; positive scores are given for matches and negative scores are given for mismatches according to chemical groups of amino acids. Dashed line marks the -2 cutoff. (A) CheW proteins. Positions of functionally important residues identified in mutational and biochemical studies are shown in color, as follows: red, null mutations (5); green, mutations that affect CheA and chemoreceptor binding in vitro (6); blue, chemoreceptor-suppressible mutations (21). The secondary structure of CheW shown above the plot is based on the reported structure of the *Thermotoga maritima* protein (15) and predicted two-dimensional structures for *E. coli* and *A. brasilense* proteins. (B) CheY proteins. Positions of functionally important residues identified in mutational, biochemical, and structural studies are shown in color, as follows: red, active-site residues; green, residues proposed to interact with FliM (19, 23), CheZ (23, 45), or CheA (24, 42). The secondary structure of CheY shown above the plot is based on the reported structure of the *E. coli* protein (41). The *A. brasilense* protein contains a 4-amino-acid deletion corresponding to residues 120 to 123 of the *E. coli* homolog. α -Helices and β -strands are indicated.

they are found in 17 positions (13%). Mismatches that score -2 or higher are often found in the same group of amino acids, which is important for structural considerations. For example, Asp and Val are both small amino acids, Ala and Leu are both

hydrophobic, and so forth. Therefore, from the structural point of view, CheW homologs appear to be more conserved (up to 93% similar) than CheY homologs. Furthermore, there is no single insertion or deletion throughout the entire alignment of

TABLE 2. Functionally important amino acid residues in the *E. coli* CheW protein compared to aligned sites in the *A. brasilense* homolog

| Residue no. ^a | Amino acid residue in: | | Phenotype or function ^b | | |
|--------------------------|------------------------|-------------------------|------------------------------------|--------------|-------------|
| | <i>E. coli</i> | <i>A. brasilense</i> | Loss of function | CheA binding | Tar binding |
| 33 | Ile | Val | | | |
| 36 | Val | Val | X | | X |
| 38 | Glu | Asp | X | | X |
| 39 | Ile | Val | | | X |
| 41 | Gly | Gly | | X | X |
| 45 | Val | Ile | | X | |
| 46 | Thr | Thr | | X | |
| 51 | Thr | Ala | | X | |
| 56 | Lys | Ala ^c | | X | |
| 57 | Gly | Gly | X | X | |
| 62 | Arg | Arg | X | | |
| 63 | Gly | Gly | X | | X |
| 65 | Ile | Ile | | X | |
| 86 | Thr | Gly | X | | X |
| 87 | Val | Met | | | X |
| 88 | Val | Ser | X | | X |
| 99 | Gly | Ser | X | | X |
| 105 | Val | Val | X | | X |
| 108 | Val | Val | X | | X |
| 128 | Gly | Gly | | | |
| 133 | Gly | Gly | | X | X |
| 156 | Met | Met | | X | |
| 158 | Leu | Ala | | X | |

^a *E. coli* numbering.

^b Data are from references 5, 6, and 21.

^c Severe substitution, shown in boldface.

CheW proteins, whereas the alignment of CheY proteins contains two insertions or deletions. Finally, an identical secondary structure was predicted for the two CheW homologs. In contrast, a striking deviation was observed between the known structure of the *E. coli* CheY protein and the predicted secondary structure of its homolog from *A. brasilense*. A 4-amino-acid deletion in the C-terminal portion of the *A. brasilense* protein (corresponding to residues 120 to 123 of the *E. coli* homolog) is likely to result in the disruption of α -helix 5 (Fig. 3).

In order to further interpret the results of heterologous expression studies and to obtain additional information about conservation of particular amino acid residues in CheW and CheY sequences, we mapped known structurally and functionally important residues of *E. coli* proteins onto the alignments shown in Fig. 3. These data are shown in a more detailed format in Tables 2 and 3.

DISCUSSION

Conservation of functionally important residues in CheW homologs. Several mutational and biochemical studies identified a number of residues on the *E. coli* CheW protein that are critical for its structure and function. Previously reported invariant (absolutely conserved in all homologs) CheW residues are Gly57, Arg62, Gly62, Gly99, and Gly128 (*E. coli* numbering) (15). We compared a larger set (102 proteins) of CheW homologs from microbial genomes available from more recent versions of major public databases (March 2002) and found that none of the residues is in fact invariant; however, Gly57 and Arg62 are conserved in 95% of sequences, Gly62 is con-

TABLE 3. Functionally important amino acid residues in the *E. coli* CheY protein compared to aligned sites in the *A. brasilense* homolog

| Residue no. ^a | Amino acid residue in: | | Function ^b | | | |
|--------------------------|------------------------|-------------------------|-----------------------|--------------|--------------|--------------|
| | <i>E. coli</i> | <i>A. brasilense</i> | Active site | CheA binding | CheZ binding | FlhM binding |
| 12 | Asp | Asp | X | | X | |
| 13 | Asp | Asp | X | | | |
| 14 | Phe | Ser ^c | | | X | |
| 16 | Thr | Val | | | X | |
| 17 | Met | Val | | | X | |
| 19 | Arg | Lys | | | X | |
| 20 | Ile | Val | | | X | |
| 23 | Asn | Lys | | | X | |
| 27 | Glu | Glu | | | | X |
| 57 | Asp | Asp | X | | X | X |
| 59 | Asn | Asn | X | | X | |
| 85 | Met | Phe | | | X | X |
| 86 | Val | Cys | | | X | X |
| 87 | Thr | Thr | X | | X | X |
| 89 | Glu | Glu | X | | X | |
| 90 | Ala | Asn | | | | X |
| 91 | Lys | Asp | | | X | |
| 92 | Lys | Leu | | X | | X |
| 94 | Asn | His | | | X | |
| 95 | Ile | Ile | | | | X |
| 96 | Ile | Gln | | | X | |
| 98 | Ala | Ala | | | X | |
| 99 | Ala | Leu | | | X | X |
| 100 | Gln | Ser | | X | | |
| 104 | Ser | Asn | | X | X | |
| 106 | Tyr | Tyr | X | X | X | X |
| 107 | Val | Ile | | | X | X |
| 108 | Val | Met | | | X | X |
| 109 | Lys | Lys | X | | X | X |
| 111 | Phe | Phe | | | X | X |
| 112 | Thr | Asn | | | X | X |
| 116 | Leu | Ile | | | | X |
| 119 | Lys | Lys | | X | X | |
| 121 | Asn | Absent | | | | X |
| 122 | Lys | Absent | | X | | X |
| 126 | Lys | Gln | | X | | X |

^a *E. coli* numbering.

^b Data are from references 19, 23, 24, 34, 35, 41, and 45.

^c Severe substitutions are shown in boldface.

served in 90% of sequences, and Gly99 and Gly128 are conserved in 85% of sequences (data not shown). Except for one (G99S), all of these residues are identical in the *E. coli* and *A. brasilense* homologs (Fig. 3A). Recent genetic and biochemical evidence confirmed that Arg62 is an absolutely irreplaceable residue but suggested a role other than interaction with a chemoreceptor or CheA (5). In the same study, Gly57 was shown to be a critical residue, mutation of which leads to a loss-of-function phenotype due to the inability of the mutant protein to interact with the CheA protein. Boukhvalova et al. (5) also identified Val36 as another critical site, mutation of which eliminates CheW binding to a chemoreceptor in vitro and chemotaxis in vivo. Two other residues, Gly41 and Gly133, were identified as "weak titrators" that decrease CheW binding affinities but are tolerated by in vivo signaling systems. Interestingly, not only Val36 but also Gly41 and Gly133 are strictly conserved in the *A. brasilense* protein (Table 2), confirming their important role for interactions within signaling complexes.

Using the yeast two-hybrid system, Boukhvalova et al. identified seven residues in CheW of *E. coli* mutation of which specifically disrupts the interaction with CheA but not with a chemoreceptor, and four residues mutation of which disrupts CheW binding to a chemoreceptor but not to CheA (6). Three out of seven "CheA-specific" residues are identical in the *A. brasilense* protein: Thr46, Ile65, and Met156. Three others can be considered "mild" substitutions: V45I, T51A, and L158A. Interestingly, the T51A substitution in the *A. brasilense* protein is exactly the same as the substitution in the *E. coli* protein, which led only to a small decrease in CheA binding and did not affect chemotaxis in vivo. The K56I substitution in the *E. coli* protein significantly inhibited CheA binding but allowed in vivo chemotaxis. In the *A. brasilense* protein, Lys56 is replaced by alanine, which can be interpreted as a less severe mutation than that of Lys to Ile and also allowed in vivo complementation. None of the four residues implicated in the specific interaction with a chemoreceptor are identical in the *A. brasilense* protein, but all four substitutions are very conservative: I33V, E38D, I39V, and V87M. The E38D substitution in the *A. brasilense* protein is identical to the mutation in the *E. coli* protein that resulted in severe inhibition of chemoreceptor binding in vitro (6). Again, however, our complementation studies are in full agreement with those of Boukhvalova et al., where the E38D mutation did not inhibit in vivo chemotaxis (6).

Earlier genetic evidence (21) implicated eight residues in *E. coli* CheW in interactions with the chemoreceptor in the signaling complex (Fig. 3A). Two of these, Glu38 and Arg62, were also identified in the biochemical studies (5, 6) and have been addressed above. Of the other six residues, three are identical (Gly63, Val105, and Val108) in the *A. brasilense* homolog and another three are conservatively changed (T86G, V88S, and G99S).

Conservation of functionally important residues in CheY homologs. All active-site residues (19, 41), including the phosphoacceptor site Asp57 and Asn59, which is nonconserved but functionally important in *E. coli* (35), are identical in the *E. coli* and *A. brasilense* CheY homologs (Table 3).

The CheY protein controls the direction of flagellar rotation by binding to its target, the FliM protein of the flagellar motor. Residues that might be important for FliM binding were first identified by mutagenesis (34) and nuclear magnetic resonance (NMR) (23) studies. More recently, some of the CheY residues directly involved in FliM binding were determined by solving the structure of activated CheY bound to the N-terminal helix of FliM (19). Eight residues of *E. coli* CheY are critical for FliM binding. Three of these residues (Ile95, Tyr106, and Lys119) are identical in the *E. coli* and *A. brasilense* proteins. Another three are relatively mild substitutions: A90N, A99L, and V108M. However, the important residue Lys92, which forms a hydrogen bond with Ser4 of FliM via its NH group, is replaced with a leucine residue in the *A. brasilense* protein. Furthermore, Lys122, which forms a salt bridge with Asp12 of the FliM protein, is missing from the *A. brasilense* protein (deletion in the α -5 helix). Other residues implicated in FliM binding by the NMR studies (23) include Met85, Val86, and Asn121, for all of which there are severe mismatches in the *A. brasilense* protein (Table 3).

A subset of residues important for CheY binding to the CheA kinase was determined by solving the crystal structure of

CheY with the phosphoacceptor-binding (P2) domain of CheA (14, 24, 42). Seven such residues were identified, and only two of them, Tyr106 (an active-site residue) and Lys119, are identical in the *E. coli* and *A. brasilense* proteins. Four substitutions in the *A. brasilense* protein are K92L, Q100S, S104N, and K126Q. Lys122, which has been suggested to form a hydrogen bond with Ala169 of CheA (24), is missing from the *A. brasilense* protein (deletion in the α -5 helix). It should be stressed that these are only some of the residues involved in CheA binding. It is likely that CheY also contains contact sites for the P1 domain of CheA.

Residues that might be important for CheZ binding were first identified by NMR studies (23) and more recently from the cocrystal structure of CheZ with CheY (45). Twenty-four residues that might participate in CheZ binding were identified, and five of them are replaced in the *A. brasilense* protein with residues that can be classified as severe mismatches (Table 3). One of these mismatches (F14S) is in the position directly implicated in interaction with the side chain of Gln147 of CheZ (45).

Evolutionary and functional considerations. In this study, we investigated evolutionary changes that occurred in two single-domain proteins of the same signal transduction pathway in two distantly related microorganisms. Based on the absolute rates of 16S rRNA divergence (26), we estimated that *A. brasilense* and *E. coli* shared a common ancestor approximately 1 billion years ago. Moreover, no orthologous relationship between the chemotaxis operons in these species is obvious, and the *A. brasilense* chemotaxis operon was likely a subject of lateral-transfer events (16). As a result, the protein sequences of CheW and CheY in the two species are significantly divergent. In both species the *cheW* and *cheY* genes are located in homologous chemotaxis operons; therefore, in a given species the two genes were subjected to similar mutagenic forces throughout the evolutionary process. However, the results demonstrate that the evolutionary changes that occurred in CheW homologs are quite different from those that occurred in CheY homologs.

Numerous, but very moderate, amino acid substitutions were allowed in CheW proteins. As a result, CheW homologs from the two distantly related organisms have identical predicted structures. Identical or very similar residues are maintained in most critical positions that were experimentally shown to be important for interaction with CheA and chemoreceptors. Only in 1 out of 23 such positions can a substitution in the *A. brasilense* protein be classified as a severe mismatch (Table 2). The unique ability of the CheW protein to maintain chemotaxis in vivo even when site mutations significantly diminish its interaction with CheA and chemoreceptors (5) probably reflects the importance of maintaining an overall structure rather than various specific sites. This type of evolutionary change likely reflects the functional constraints. CheW functions to couple the signaling state of chemotaxis transducers to the CheA kinase (4) and is involved in the formation of transducer-kinase complexes (13), but it does not have any catalytic activity on its own.

Very different changes occurred in CheY homologs. The CheY protein is a smaller molecule than CheW, but its functional role appears to be significantly more complex. In *E. coli*, it binds to at least three different partners, namely, CheA,

FliM, and the CheZ phosphatase, and it undergoes CheA-dependent phosphorylation and CheZ-dependent dephosphorylation. The overall number of amino acid substitutions in *E. coli* and *A. brasilense* CheY homologs is comparable to that in CheW homologs. However, there is a clear difference in the pattern of amino acid substitution. More positions are strictly conserved in CheY than in CheW homologs (35 versus 24%), and at the same time more severe changes have occurred in CheY than in CheW homologs (13 versus 7%). The CheY protein has known enzymatic activities (phosphorylation and dephosphorylation) and an active site to carry them out, whereas CheW does not. In CheY, strict conservation is observed in the active-site residues and residues that are critical for protein folding (25). However, many residues that have been implicated in interaction with the CheA, CheZ, and FliM proteins are not conserved in the *A. brasilense* CheY (Table 3).

We interpret our heterologous expression studies as suggesting that the *A. brasilense* protein can weakly interact with and be phosphorylated by *E. coli* CheA. The CheA protein of *A. brasilense* lacks the P2 domain, which serves as the high-affinity binding site for CheY in the *E. coli* protein. However, the weak interaction and phosphorylation of the *A. brasilense* CheY by *E. coli* CheA is in agreement with the experimental data showing that CheA lacking the P2 domain still binds and phosphorylates CheY, although less efficiently (37). CheZ protein is found exclusively in the branch of β/γ -proteobacteria (I. B. Zhulin, unpublished data); therefore, it is not surprising that sites for interaction with CheZ are not conserved in the *A. brasilense* CheY protein. Relaxation of the FliM binding sites and the apparent inability of the *A. brasilense* CheY protein to interact with the *E. coli* FliM suggest that the FliM protein in *A. brasilense* might be significantly divergent from that in *E. coli*. It appears that the CheY protein tolerates dramatic substitutions in the individual sites that are coincident with the presence or absence of interacting partner proteins and domains (Table 3) or reflect potentially significant changes in the interacting surfaces of a partner protein.

Taken together, the results of our analysis fully support the notion that structural and functional constraints on specific positions within a protein sequence are extremely important factors in molecular evolution (22, 28). Quantification of evolutionary constraints on particular amino acid positions has been recently proposed as a new tool for detection of functionally important regions in proteins (36). Recent analysis of evolutionary changes in interacting proteins in eukaryotes demonstrated that such proteins have similar functional constraints and thus have coevolved (12). Our results suggest that despite the facts that CheW and CheY are components of the same signal transduction pathway and the corresponding genes are always located together on microbial DNA, the apparent evolutionary constraints on them are different, likely because they perform different functions within the pathway and do not interact with each other.

It appears that the input elements of the chemotaxis signaling pathway are more conserved than the output elements. All known chemoreceptors from distantly related species of prokaryotes share a highly conserved domain (20, 46). As suggested in this study, the docking protein CheW is also highly conserved, although at a different stringency level. Finally, the CheW-like domain of the CheA histidine kinase, which is pro-

posed to interact with the CheW protein (3), is the most conserved domain of this protein, whereas the output portion of CheA (CheY-interacting domains) are significantly less conserved (G. Alexandre, K. Wuichet, and I. B. Zhulin, unpublished data). Our results also suggest that the CheY protein might be highly variable with respect to critical sites in its C-terminal region involved in CheA, FliM, and CheZ interactions. The flexibility of the output elements of the chemotaxis signaling pathway may reflect different mechanisms for signal termination, such as phosphatase versus a "phosphate sink," and differences in the structure and function of bacterial flagellar motors (32).

ACKNOWLEDGMENTS

We thank Sandy Parkinson for *E. coli* strains, Rick Dahlquist for atomic coordinates of *Thermotoga maritima* CheW, Christophe Mougel for advice on calibrating evolutionary divergence, and Rick Stewart for sharing data prior to publication. We are grateful to Sandy Parkinson and Rick Stewart for helpful discussions and critical reading of the manuscript. We also thank the anonymous referee for many helpful suggestions.

This work was supported by start-up funds from the Georgia Institute of Technology (to I.B.Z.).

REFERENCES

- Amann, E., B. Ochs, and K. J. Abel. 1988. Tightly regulated *tac* promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* **69**:301–315.
- Bespalov, V. A., I. B. Zhulin, and B. L. Taylor. 1996. Behavioral responses of *Escherichia coli* to changes in redox potential. *Proc. Natl. Acad. Sci. USA* **93**:10084–10089.
- Bilwes, A. M., L. A. Alex, B. R. Crane, and M. I. Simon. 1999. Structure of CheA, a signal-transducing histidine kinase. *Cell* **96**:131–141.
- Borkovich, K. A., N. Kaplan, J. F. Hess, and M. I. Simon. 1989. Transmembrane signal transduction in bacterial chemotaxis involves ligand-dependent activation of phosphate group transfer. *Proc. Natl. Acad. Sci. USA* **86**:1208–1212.
- Boukhvalova, M., F. W. Dahlquist, and R. C. Stewart. 2002. CheW binding interactions with CheA and Tar: importance for chemotaxis signaling in *Escherichia coli*. *J. Biol. Chem.* **277**:22251–22259.
- Boukhvalova, M., R. VanBruggen, and R. C. Stewart. 2002. CheA kinase and chemoreceptor interaction surfaces on CheW. *J. Biol. Chem.* **277**:23596–23603.
- Bourret, R. B., and A. M. Stock. 2002. Molecular information processing: lessons from bacterial chemotaxis. *J. Biol. Chem.* **277**:9625–9628.
- Bren, A., and M. Eisenbach. 2000. How signals are heard during bacterial chemotaxis: protein-protein interactions in sensory signal propagation. *J. Bacteriol.* **182**:6865–6873.
- Cuff, J. A., and G. J. Barton. 1999. Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins* **34**:508–519.
- Dayhoff, M. O., R. Schwartz, and B. C. Orcutt. 1978. A model of evolutionary change in proteins, p. 345–352. *In* M. O. Dayhoff, (ed.), *Atlas of protein sequences and structures*, vol. 5, Suppl. 3. National Biomedical Research Foundation, Washington, D.C.
- Falke, J. J., R. B. Bass, S. L. Butler, S. A. Chervitz, and M. A. Danielson. 1997. The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annu. Rev. Cell Dev. Biol.* **13**:457–512.
- Fraser, H. B., A. E. Hirsh, L. M. Steinmetz, C. Scharfe, and M. W. Feldman. 2002. Evolutionary rate in the protein interaction network. *Science* **296**:750–752.
- Gegner, J. A., D. R. Graham, A. F. Roth, and F. W. Dahlquist. 1992. Assembly of an MCP receptor, CheW and kinase CheA complex in the bacterial chemotaxis signal transduction pathway. *Cell* **70**:975–982.
- Gouet, P., N. Chinardet, M. Welch, V. Guillet, S. Cabantous, C. Birck, L. Mourey, and J.-P. Samama. 2000. Further insights into the mechanism of function of the response regulator CheY from crystallographic studies of the CheY-CheA_{124–257} complex. *Acta Crystallogr. D* **57**:44–51.
- Griswold, I. J., H. Zhou, M. Matison, R. V. Swanson, L. P. McIntosh, M. I. Simon, and F. W. Dahlquist. 2002. The solution structure and interactions of CheW from *Thermotoga maritima*. *Nat. Struct. Biol.* **9**:121–125.
- Hauwaerts, D., G. Alexandre, S. K. Das, J. Vanderleyden, and I. B. Zhulin. 2002. A major chemotaxis gene cluster in *Azospirillum brasilense* and relationships between chemotaxis operons in α -proteobacteria. *FEMS Microbiol. Lett.* **208**:61–67.

17. Henikoff, S., and J. G. Henikoff. 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* **89**:10915–10919.
18. Hess, J. F., K. Oosawa, N. Kaplan, and M. I. Simon. 1988. Phosphorylation of three proteins in the signaling pathway of bacterial chemotaxis. *Cell* **53**:79–87.
19. Lee, S. Y., H. S. Cho, J. G. Pelton, D. Yan, R. K. Henderson, D. S. King, L. Huang, S. Kustu, E. A. Berry, and D. E. Wemmer. 2001. Crystal structure of an activated response regulator bound to its target. *Nat. Struct. Biol.* **8**:52–65.
20. Le Moual, H., and D. E. Koshland, Jr. 1996. Molecular evolution of the C-terminal cytoplasmic domain of a superfamily of bacterial receptors involved in taxis. *J. Mol. Biol.* **261**:568–585.
21. Liu, J., and J. S. Parkinson. 1989. Role of CheW protein in coupling membrane receptors to the intracellular signaling system of bacterial chemotaxis. *Proc. Natl. Acad. Sci. USA* **86**:8703–8707.
22. Lopez, P., D. Casane, and H. Philippe. 2002. Heterotachy, an important process in protein evolution. *Mol. Biol. Evol.* **19**:1–7.
23. McEvoy, M. M., A. Bren, M. Eisenbach, and F. W. Dahlquist. 1999. Identification of the binding surfaces on CheY for two of its targets, the phosphatase CheZ and the flagellar switch protein FlhM. *J. Mol. Biol.* **289**:1423–1433.
24. McEvoy, M. M., A. C. Hausrath, G. B. Randolph, S. J. Remington, and F. W. Dahlquist. 1998. Two binding modes reveal flexibility in kinase/response regulator interactions in the bacterial chemotaxis pathway. *Proc. Natl. Acad. Sci. USA* **95**:7333–7338.
25. Mirny, L. A., and E. I. Shakhnovich. 1999. Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. *J. Mol. Biol.* **291**:177–196.
26. Ochman, H., S. Elwyn, and N. A. Moran. 1999. Calibrating bacterial evolution. *Proc. Natl. Acad. Sci. USA* **96**:12638–12643.
27. Parkinson, J. S. 1978. Complementation analysis and deletion mapping of *Escherichia coli* mutants defective in chemotaxis. *J. Bacteriol.* **135**:45–53.
28. Philippe, H., and P. Lopez. 2001. On the conservation of protein sequences in evolution. *Trends Biochem. Sci.* **26**:414–416.
29. Rost, B. 1996. PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol.* **266**:525–539.
30. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
31. Sanders, D. A., B. Mendez, and D. E. Koshland, Jr. 1989. Role of the CheW protein in bacterial chemotaxis: overexpression is equivalent to absence. *J. Bacteriol.* **171**:6271–6278.
32. Scharf, B., and R. Schmitt. 2002. Sensory transduction to the flagellar motor of *Sinorhizobium meliloti*. *J. Mol. Microbiol. Biotechnol.* **4**:183–186.
33. Shah, D. S. H., S. L. Porter, D. C. Harris, G. H. Wadhams, P. A. Hamblin, and J. P. Armitage. 2000. Identification of a fourth *cheY* gene in *Rhodobacter sphaeroides* and interspecies interaction within the bacterial signal transduction pathway. *Mol. Microbiol.* **35**:101–112.
34. Shukla, D., X. Y. Zhu, and P. Matsumura. 1998. Flagellar motor-switch binding face of CheY and the biochemical basis of suppression by CheY mutants that compensate for motor-switch defects in *Escherichia coli*. *J. Biol. Chem.* **273**:23993–23999.
35. Silversmith, R. E., J. G. Smith, G. P. Guanga, J. T. Les, and R. B. Bourret. 2001. Alteration of a nonconserved active site residue in the chemotaxis response regulator CheY affects phosphorylation and interaction with CheZ. *J. Biol. Chem.* **276**:18478–18484.
36. Simon, A. L., E. A. Stone, and A. Sidow. 2002. Inference of functional regions in proteins by quantification of evolutionary constraints. *Proc. Natl. Acad. Sci. USA* **99**:2912–2917.
37. Stewart, R. C., K. Jahreis, and J. S. Parkinson. 2000. Rapid phosphotransfer to CheY from a CheA protein lacking the CheY-binding domain. *Biochemistry* **39**:13157–13165.
38. Tatusova, T. A., and T. L. Madden. 1999. BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174**:247–250.
39. Taverna, D. M., and R. A. Goldstein. 2002. Why are proteins so robust to site mutations? *J. Mol. Biol.* **315**:479–484.
40. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876–4882.
41. Volz, K., and P. Matsumura. 1991. Crystal structure of *Escherichia coli* CheY refined at 1.7-Å resolution. *J. Biol. Chem.* **266**:15511–15519.
42. Welch, M., N. Chinardet, L. Mourey, C. Birck, and J.-P. Samama. 1998. Structure of the CheY-binding domain of histidine kinase CheA in complex with CheY. *Nat. Struct. Biol.* **5**:25–29.
43. Welch, M., K. Oosawa, S. Aizawa, and M. Eisenbach. 1993. Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria. *Proc. Natl. Acad. Sci. USA* **90**:8787–8791.
44. Wolfe, A. J., and H. C. Berg. 1989. Migration of bacteria in semisolid agar. *Proc. Natl. Acad. Sci. USA* **86**:6973–6977.
45. Zhao, R., E. J. Collins, R. B. Bourret, and R. E. Silversmith. 2002. Structure and catalytic mechanism of the *E. coli* chemotaxis phosphatase CheZ. *Nat. Struct. Biol.* **9**:570–575.
46. Zhulin, I. B. 2001. The superfamily of chemotaxis transducers: from physiology to genomics and back. *Adv. Microb. Physiol.* **45**:157–198.