J. Mol. Biol. (1995) **249**, 654–664

JMB

Investigating the Structural Determinants of the p21-like Triphosphate and Mg^{2+} Binding Site

Philippe Cronet[1]*, Lluís Bellsolell[2,3]*, Chris Sander[1], Miquel Coll[2,3] and Luis Serrano[1]

[1]*European Molecular Biology Laboratory, Meyerhofstrasse 1 6900 Heidelberg, Germany*

[2]*Departament de Biologia Molecular i Cellular Centre d'Investigació i Desenvolupament-CSIC, Jordi Girona 18-26, Barcelona 08034, Spain*

[3]*Departament d'Enginyeria Química, Universitat Politècnica de Catalunya Diagonal 647, 08028 Barcelona, Spain*

**Corresponding author*

Amongst the superfamily of nucleotide binding proteins, the classical mononucleotide binding fold (CMBF), is the one that has been best characterized structurally. The common denominator of all the members is the triphosphate / Mg^{2+} binding site, whose signature has been recognized as two structurally conserved stretches of residues: the Kinase 1 and 2 motifs that participate in triphosphate and Mg^{2+} binding, respectively. The Kinase 1 motif is borne by a loop (the P-loop), whose structure is conserved throughout the whole CMBF family. The low sequence similarity between the different members raises questions about which interactions are responsible for the active structure of the P-loop. What are the minimal requirements for the active structure of the P-loop? Why is the P-loop structure conserved despite the diverse environments in which it is found? To address this question, we have engineered the Kinase 1 and 2 motifs into a protein that has the CMBF and no nucleotide binding activity, the chemotactic protein from *Escherichia coli*, CheY. The mutant does not exhibit any triphosphate / Mg^{2+} binding activity. The crystal structure of the mutant reveals that the engineered P-loop is in a different conformation than that found in the CMBF. This demonstrates that the native structure of the P-loop requires external interactions with the rest of the protein. On the basis of an analysis of the conserved tertiary contacts of the P-loop in the mononucleotide binding superfamily, we propose a set of residues that could play an important role in the acquisition of the active structure of the P-loop.

Keywords: protein design; CheY; crystal structure; P-loop; α-helix

Introduction

The mononucleotide triphosphate binding site is a frequently occurring feature in proteins. Amongst the mononucleotide triphosphate binding proteins for which a structure has been determined, only a few distinct classes have been recognized (reviewed by Schulz, 1992): the adenylate kinase fold, or ''classical mononucleotide binding fold'' (Schulz *et al*., 1974), the cAMP-dependent kinase fold (Knighton *et al*., 1991; Zheng *et al*., 1993), the actin fold (Kabsch *et al*., 1990; Flaherty *et al*., 1991), the type II tRNA synthetase (reviewed by Moras, 1992), the glutathione synthetase (Yamagushi *et al*., 1993) and very recently the nucleotide diphosphate kinase (Cherfils *et al*., 1994).

The classical mononucleotide binding fold (CMBF) is one of the best characterized structural families (Figure 1). About 15 members have been crystallized and described so far, with or without cofactors, so a fairly complete picture is available (Traut, 1994). Well-known members of the family are p21ras, the product of the Ha-*ras* oncogene (Pai *et al*., 1990; De Vos *et al*., 1988), adenylate kinase (ADK) (Dreusicke *et al*., 1988) and elongation factor Tu (EFTu) (La Cour *et al*., 1985). More recently three new structures have been determined, Rec A (Story & Steitz, 1992), dethiobiotin synthetase (Huang *et al*., 1994) and the Gα subunit of the heterotrimeric G protein transducin (Noel *et al*., 1993; Lambright *et al*., 1994). The members of the family show a high degree of structural similarity (Artymiuk *et al*., 1990; Holm & Sander, 1993) although the sequence identity between the members is very low (5 to 6%). Earlier studies have shown that two conserved peptide units, the Walker A and B motifs were highly

Abbreviations used: CMBF, the classical mononucleotide binding fold; ADK, adenylate kinase; EFTu, elongation factor Tu; p2lras, product of the Ha-*ras* oncogene; CheY, chemotactic protein Y from *E. coli*; MNTP, mononucleotide triphosphate; cAMP, cyclic adenosine monophosphate; P-loop, phosphate-binding loop; r.m.s.d., root-mean-square deviation; BSA, bovine serum albumin.

0022–2836 / 95 / 230654–11 $08.00 / 0

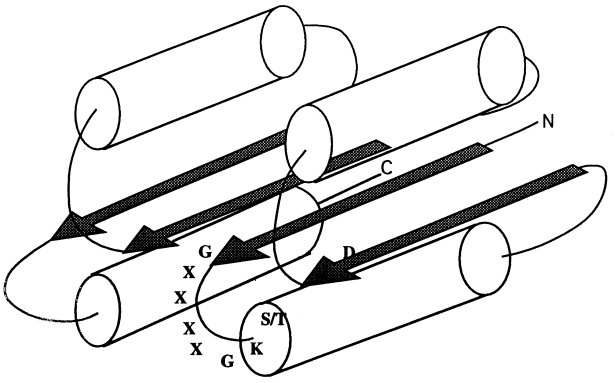

Figure 1. Schematic representation of the classical mononucleotide binding fold. The sequence of the Kinase 1 (GXXXXGKS/T) and 2 (D) motifs are inserted into their respective positions. N and C are the N- and C-terminal ends of the protein chain.

conserved in several members of the group (Walker *et al*., 1982). These units were proposed to be responsible for triphosphate and Mg^{2+} binding, respectively. As more known structures joined the group, these motifs have been reconsidered and updated, especially the Walker B motif, which does not participate in the Mg^{2+} binding as originally proposed. On the basis of further work (Dever *et al*., 1987; Traut, 1994) the Walker A consensus motif has been renamed the Kinase 1 motif and has the following fingerprint sequence: G(X)XXXXGKS/T/G (the X in parenthesis has been added to account for an insertion found in the sequence of the dethiobiotin synthetase). The insertion in the dethiobiotin synthetase is the first exception to the usual GXXXXGKS/T/G pattern, demonstrating the relatively weak predictive value of the sequence alignment-derived motifs. This motif in this family is always found on a loop: the phosphate-binding loop or ''P-loop'' (Saraste *et al*., 1990), between the first β-strand and the first α-helix of the protein. The structure of the P-loop is extremely well conserved in the CMBF family although the environment around it is not. Even in dethiobiotin synthetase (Huang *et al*., 1994), where an insertion is present in the P-loop, the structural features, i.e. the dihedral angles of the conserved residues of the motif, are identical to those of the other members. It is responsible for the binding of the phosphate moiety of the nucleotide by forming a ''giant anion hole'' (Dreusicke & Schulz, 1986), in which the α- and β-phosphates interact with the helix dipole (Wierenga *et al*., 1985). The last serine residue, or threonine residue, of the Kinase 1 motif is directly or indirectly involved in the coordination of a Mg^{2+} associated with the nucleotide, together with a conserved aspartic acid residue, the Kinase 2 motif (Dever *et al*., 1987; Traut, 1994). This aspartate residue is usually found at the C-terminal extremity of a β-strand in the immediate vicinity of the Kinase 1 motif. The relative positions of the motifs are well conserved in the available structures, probably reflecting the precise adjustment needed for the proper function of the binding site.

What are the structural determinants of the P-loop? Is it a self-determined structure or does it require external factors to maintain its active conformation? To address this question, we have designed several modifications of a protein that shows high structural similarity with the members of the CMBF, but does not bind nucleotides and does not have the Kinase 1 and 2 motifs: i.e. the chemotactic protein from *Escherichia coli*, CheY (Stock *et al*., 1989; Artymiuk *et al*., 1990; Chen *et al*., 1990). CheY is a small α/β parallel protein whose fold (Stock *et al*., 1989; Volz & Matsumura, 1991; Bruix *et al*., 1993; Bellsolell *et al*., 1994) is the same as that of the CMBF family although it does not have any significant sequence similarity with any of the members (Artymiuk *et al*., 1990). CheY is involved in regulating the chemotactic movement of the bacteria in response to several chemical stimuli in the environment (Parkinson, 1983). In response to the right stimulus the protein is phosphorylated at Asp57 (Sanders *et al*., 1989). In order to be phosphorylated or dephosphorylated, the protein requires a Mg^{2+} to be bound in its active site (Hess *et al*., 1988).

On the basis of the high structural similarity between CheY and the CMBF, we identified the regions of CheY that occupy equivalent positions to the triphosphate/Mg^{2+} binding site in the CMBF family. The first loop of CheY was mutated to a typical P-loop sequence and the Kinase 2 motif was introduced simultaneously in the appropriate position. The mutant was purified and crystallized. Here, we report and discuss the structure of the triphosphate/Mg^{2+} binding site of the mutant. On the basis of the comparison with the usual active structure of the binding site, we propose a set of important tertiary contacts for the native structure of this widely occurring triphosphate binding motif.

Results

Structural comparison of CheY and the CMBF binding proteins

To determine the regions of CheY suitable for introduction of the Kinase 1 and 2 motifs, CheY was superimposed on the members of the CMBF family. In the next paragraph, we present a detailed comparison of the P-loop, and of the Kinase 2 motif, with the corresponding regions in CheY. The regions including the Kinase 1 and 2 motifs, as well as the position of the phosphate group of the nucleotide are virtually identical in the CMBF family (Story & Steitz, 1992). Therefore, for simplicity, we choose p21ras as our main template model, unless specified otherwise. The residue numbers will be that of CheY and the corresponding ones from p21ras will appear in parentheses.

The first β/α unit of CheY is very similar to the first β/α unit of p21ras and of any member of the CMBF

(a)

P21:	G_{10}	A	G	G	V	-	G	K	S_{17}	///	D_{57}	T	A	G_{60}
EFTU:	G	H	V	D	H	-	G	K	T	///	D	C	P	G
ADK:	G	G	P	G	S	-	G	K	G	///	D	G	Y	P
DETHIO:	G	T	D	T	E	*V	G	K	T	///	E	G	A	G

CheY:	D_{12}	D	F	S	T	-	M	R	R_{19}	///	E_{35}	A	E	D_{38}

(b)

Figure 2.(a) Superimposition of p21ras and CheY Kinase 1 (residues G10/D12 to S17/R19) and 2 (residue D57/E35) regions. The residue positions are given as p21ras/CheY. The main chain (thin lines) and ribbon (thick lines) are colored in green/yellow for p21ras/CheY. The GTP of the p21ras binding site is colored by atom type (C, green; O, red; N, blue). There is a very good general agreement apart from the one extra turn in the first helix of CheY. (b) Structurally derived sequence alignment of p21ras (PDB entry 5P21; Pai *et al*., 1990), elongation factor Tu (PDB entry 1ETU; La Cour *et al*., 1985), adenylate kinase (PDB entry 3ADK; Dreusicke *et al*., 1988) and dethiobiotin synthetase (Huang *et al*., 1994; D. Andreev, personal communication) in the region spanning the P-loop (Kinase 1 motif) and the Kinase 2 motif. The sequences of the CheY segments superimposed on the Kinase motifs are included. The residue numbers for p21ras and CheY appear in indices. This gives an idea of the sequence diversity found in this part of the structures. *The P-loop of dethiobiotin synthetase has an insertion in the P-loop after the second ''X'' of the motif and the D of the Kinase 2 motif is an E.

family containing the P-loop (Gly10 to Lys16; Figure 2(a) and (b)). Superimposition on the C^{α} trace of residues 6 to 23 from CheY on residue 4 to 21 of p21ras gives an r.m.s.d. of 1.9 Å. The same superimposition excluding the loop residues 14 to 17 from CheY and 12 to 15 from p21ras gave an r.m.s.d. of 1.0 Å. The first β-strand of CheY matches the first β-strand of p21ras up to Phe14 (Gly12). Ser15 (Gly13) shows the main differences with p21ras. This residue assumes unusual φ angles in p21ras (and related structures), whereas those in CheY fit the most populated area of the (φ, ψ) map (data not shown). Glycine is therefore the residue most frequently occupying this position in the CMBF family, although different residues could be found (i.e. an aspartate in EFTu and a serine residue in RecA; Schulz, 1992). The last part of the P-loop consists of the transition between the loop *per se* and the first helix of p21ras. The helix in p21ras starts at Lys16, with Gly15 being the N-capping residue (Richardson & Richardson, 1988). This transition region constitutes the triphosphate-binding interface where the α and β phosphate groups of the nucleotide interact with the free amide groups of the first turn of the helix. In CheY, the first helix is longer by one turn and starts at the level of Phe14. The last glycine residue of the G(X)XXXXGKS/T/G motif corresponds to Met17 (Gly15), while Arg18 is the equivalent in the sequence alignment to Lys16 in p21ras. It is interesting to note that the α and β oxygen atoms of the triphosphate moiety are perfectly superimposable on the carbonyl main-chain oxygens of Phe14 and Ser15. This suggests that these phosphate groups could play the role of an extra turn of the helix in this type of binding site (Figure 2(a)).

Concerning the Kinase 2 motif, a glutamic acid residue (Glu35), is found in CheY at a similar position as Asp57 of p21ras (Figure 2(b)). When we started this study, Glu had never been observed in place of the conserved Asp interacting with Mg^{2+}. In the crystal structure of dethiobiotin synthetase (Huang *et al*., 1994; D. Andreev, personal communication), a glutamic acid residue (Glu115) is found at this position. However, the role of this Glu deserves careful interpretation, since the structure represents the apo-form of the enzyme.

Another important difference between CheY and p21ras is the location of the active site. If one draws a plane containing the axis of the first β-strand and of the first α-helix, CheY's Mg^{2+} binding site is on the opposite side of the plane compared to the mononucleotide triphosphate (MNTP) binding site of p21ras and related structures. This has an important consequence. Whereas the Mg^{2+} binding site of CheY, which involves Asp12, -13 and -57 (Bellsolell *et al*., 1994; Protein Data Bank (PDB) entry 1CHN), is clearly hydrophilic and exposed to the solvent, the equivalent residue to Asp57 in p21ras is a hydrophobic residue (Val81) that is shielded from the solvent by the P-loop.

Design of the mutant

In order to introduce the Kinase 1 and 2 motifs, the following mutations were introduced into CheY: Asp12 to Gly, Asp13 to Asn, Phe14 to Gly, Ser15 to Gly, Met17 to Gly, Arg18 to Lys, Arg19 to Ser, Ile20 to Thr and Glu35 to Asp (Figure 4(a)). Apart from the residues directly involved in the consensus motifs (Asp12 to Gly, Met17 to Gly, Arg18 to Lys, Arg19 to Ser and Glu35 to Asp), other modifications were introduced into the original CheY sequence for specific purposes. Phe14 is the equivalent residue to Gly12 in p21ras. This residue was mutated to a Gly, to compensate for the possible destabilizing effect of deleting the buried hydrophobic side-chain of Met17. Mutation of this hyperexposed residue to a hydrophilic one results in a large increase in stability of CheY by the inverse hydrophobic effect (Muñoz *et al*., 1994). Asp13, which is part of CheY active site and is involved in the binding of Mg^{2+} (Bellsolell *et al*., 1994), was mutated to Asn to avoid any interference in the original CheY Mg^{2+} binding activity. Ile20 is equivalent to Ala18 in p21ras. Though there is an Ala

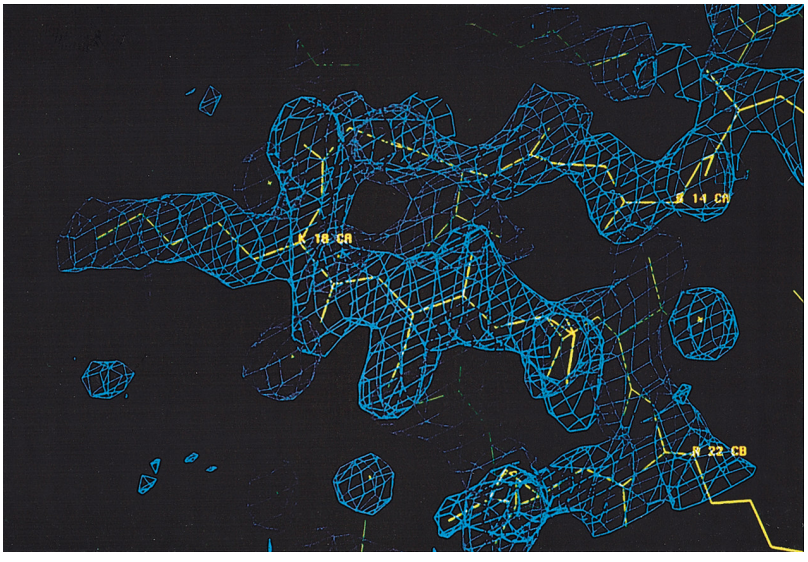

Figure 3. $2F_o - F_c$ electron density map (blue) of the P-loop area calculated at 2.24 Å resolution and contoured at 1σ and the final crystallographic model (yellow) fitted into the electron density.

Consensus	G	X	X	X	X	G	K	S/T	A/T	---//--	D
P21	G_{10}	A	G	G	V	G	K	S	A_{18}	---//---	D_{57}
CheY	D_{12}	D	F	S	T	M	R	R	I_{20}	---//---	E_{35}
Mutant 2	G_{12}	N	G	G	T	G	K	S	T_{20}	---//---	D_{35}

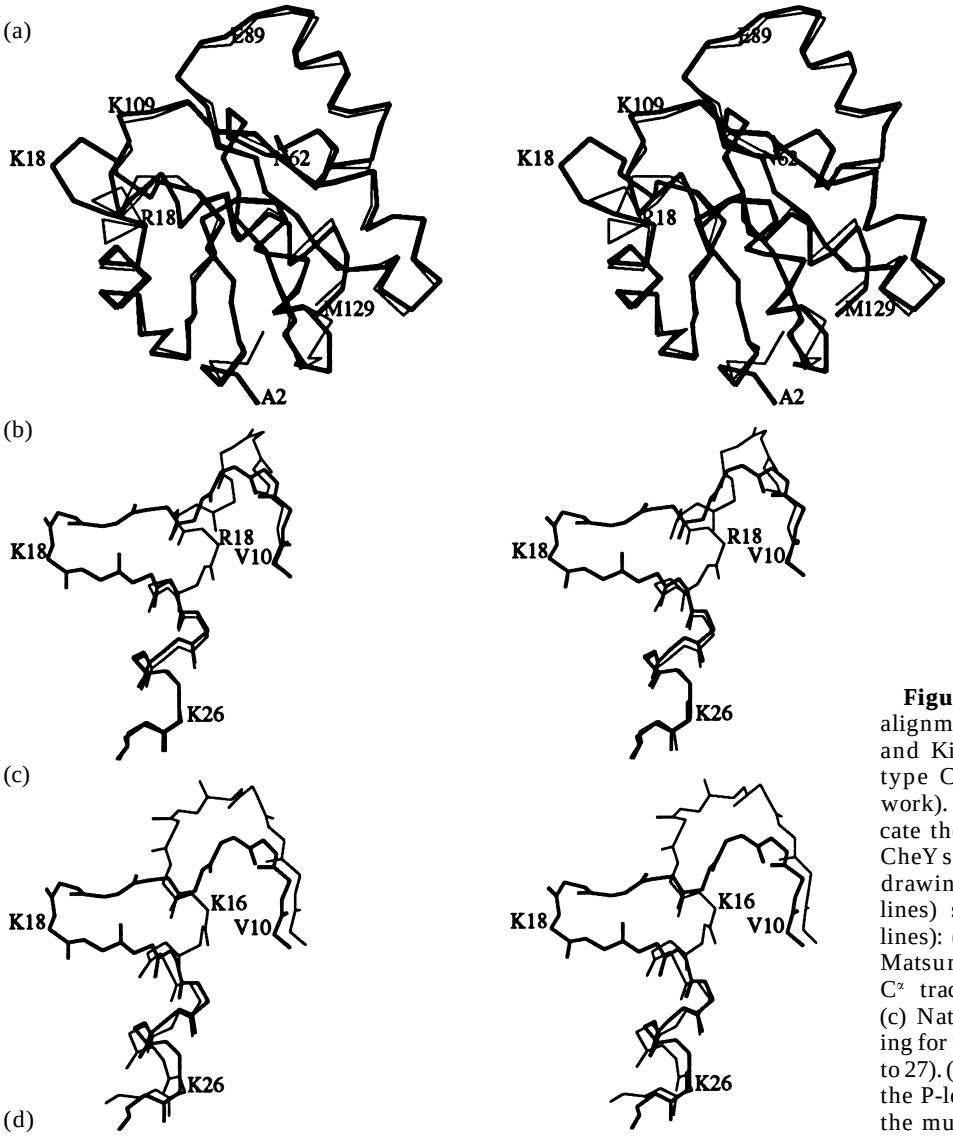

Figure 4. (a) Structurally derived alignment of the consensus Kinase 1 and Kinase 2 motifs. p21ras, wild-type CheY and mutant CheY (this work). The characters in bold indicate the mutations of the wild-type CheY sequence. (b), (c) and (d) Stereo drawings of the mutant (in thick lines) superimposed onto (in thin lines): (b) Native apo-CheY (Volz & Matsumura, 1991; PDB entry 3CHY), C^{α} tracing for the whole protein. (c) Native apo-CheY backbone tracing for the first loop area (residues 10 to 27). (d) p21ras backbone tracing for the P-loop area (residues 10 to 27 of the mutant, 8 to 25 of p21ras).

here in p21ras, Ser or Thr are often found at this position in many proteins of the family, their side-chains making a hydrogen bond to the guanosine ring of the nucleotide. We mutated Ile20 to a Thr with the idea of using it in the future to introduce the full nucleotide binding site in CheY.

Modeling of the mutant

The purpose of this model was to get a rough idea of the new environment created in CheY upon insertion of the loop. No clashes were observed between the inserted loop and the rest of CheY. Apart from the hole created by removal of the Met17 side-chain, no drastic perturbation of the structure could be forecast on the basis of this model (data not shown).

Nucleotide binding properties of the mutant

Filter-binding assays with radioactive GTP and ATP did not show any detectable binding of the nucleotide onto the protein (data not shown).

Structural analysis of the mutant

The global structure of the mutant CheY (excluding the new P-loop) is similar to the apo-CheY structure. Figure 4(b) shows a superimposition of the C^{α} tracings of both structures. The r.m.s.d. in the C^{α}

atomic positions is only 0.58 Å if one excludes the engineered P-loop. Table 1 provides a list of those residues that have moved by more than 3 Å with respect to their original position in CheY. This new loop is drastically different from the equivalent loop in the apo-CheY structure (Figure 3 and 4(b), (c)). The first helix starts at Thr20, which acts as the N-cap residue. The hydroxyl side-chain group of this residue is hydrogen bonded to two of the free main-chain amides of the first helical turn (Richardson & Richardson, 1988). This helix starts at Phe14 in wild-type CheY with (Bellsolell *et al*., 1994; Stock *et al*., 1994) and without Mg^{2+} (Stock *et al*., 1989; Volz & Matsumura, 1991; Bruix *et al*., 1993). The loop is extended by six residues in the mutant when compared with the wild-type structures and has shifted towards the loop made by residues 108 to 112. The distal segment of the first helix has not undergone any structural modification compared to the wild-type. The region around Asp57 is now a little more exposed to the solvent than it is in the wild-type.

As a consequence, the P-loop is rather different from the active P-loop structure of p21ras (Figure 4(d)). As an example, the C^{α} position of Lys18 has moved by 11 Å with respect to its position as expected from the model (Table 1). The loop now forms a short irregular ribbon with two main chain-main chain hydrogen bonds: O(Gly14)–N(Val21) and N(Thr16)–O(Ser19). There is a crystallographic 2-fold axis close to the new P-loop. The P-loops of symmetry-related molecules in the crystal lattice interact with each other. The interaction includes two hydrogen bonds from N(G17) to O(G17*) and O(G17) to N(G17*), where the asterisk indicates a symmetry-related molecule. This could account for the ''distorted'' conformation of the loop. However, the NMR structure of the same mutant clearly shows that this region, although rather mobile, adopts a very similar conformation in solution to the structure observed in the crystal (Bruix, M., Cronet, P., Santoro, J., Sander, C., Serrano, L. & Rico. M., unpublished results).

Table 1

Structural differences observed between CheY and the mutant

Residue	C^{α} displacement (Å)
Asn13	3.5
Gly14	6.0
Gly15	6.3
Thr16	7.5
Gly17	7.0
Lys18	11.0
Ser19	5.9

Atomic displacements (in Å) observed in the CheY mutant structure compared with the wild-type crystallographic structure. The values were measured after least squares superimposition of apo-CheY (Volz & Matsumura, 1991) and the mutant using the program MOLFIT (written by S. J. Remington).

Table 2

Conserved contacts with the P-loop in the CMBF family

	P21	EFTu	ADK
	N116	**N135**	**R128**
G	**V81**	**V104**	**V118**
	K117		
	D57	D80	
	T58		I192
K	A59	Y87	N193
	G60		
	V81	**V104**	**V118**
	T35		
S	**D57**	**D80**	**D93**

Tertiary contacts observed between the GKS/T/G tripeptide of the Kinase 1 motif and the rest of the molecule in three representative members of the CMBF. The residues appearing on the same line and in bold characters make conserved contacts with the corresponding residue of the P-loop throughout the whole CMBF family.

The next region of interest in this mutant is the Kinase 2 motif where Glu35 has been mutated to Asp. The new Asp35 does not adopt the same rotamer as the equivalent aspartic acid residue in p21ras (Asp57). We think that this could be due to the presence of the side-chain of Arg22 interacting with the carboxyl group of Asp35.

Helical propensity of the first helix in MNTP-binding proteins

The algorithm AGADIR (Muñoz & Serrano, 1994a,b), that evaluates accurately the intrinsic helical propensity of α-helices in solution at the residue level, suggests that the α-helix in the different members of the CMBF family should start after the conserved Ser/Thr residue of the Kinase 1 motif (data not shown). The reason for this is that Ser and Thr are very good N-capping residues (Richardson & Richardson, 1988; Fersht & Serrano, 1993). Moreover, Lys is a very unfavorable residue when close to the N terminus, due to the repulsion between its positive charge and the N terminus helix dipole (Muñoz & Serrano, 1994b). Therefore the helix is predicted to be shorter by one turn in solution than it is in the CMBF family. Confirming this hypothesis, a protein fragment corresponding to the first 45 residues of ADK shows that the first helix starts after the conserved Thr23 residue of the Kinase 1 motif (Fry *et al*., 1988).

In the crystal structure of our CheY mutant the helix starts at the level of Thr20 (two positions after the conserved Lys of the Kinase 1 motif), supporting the hypothesis that specific tertiary contacts are needed to induce the active conformation of the P-loop.

Analysis of the tertiary contacts with the P-loop

Table 2 shows the summary of the tertiary contacts with the GKS/T motif of p21ras, EFTu and ADK. There is a conserved hydrogen bond between the side-chain of the conserved Asp of the Kinase 2 motif

and the side-chain of the conserved Ser/Thr of the Kinase 1 motif. More interestingly, there is always a hydrophobic residue (mainly a Val), on a β-strand adjacent to the P-loop (Val81 in p21ras, which is equivalent to Asp57 in CheY). This residue makes strong van der Waals' contacts with the conserved Gly-Lys of the Kinase 1 motif. In the GTP-binding subfamily, there is an asparagine residue (i.e. Asn116 in p21ras), whose side-chain makes a hydrogen bond to the main-chain carbonyl of the last Gly of the Kinase 1 motif. In the ATP-binding subfamily, this Asn is not present, but similar van der Waals' contacts with the Gly are made by other residues (i.e. Arg128 in ADK).

Discussion

Comparison of Kinase 1 and 2 motifs (Walker *et al*., 1982), in the CMBF family (Schulz, 1974), reveals their very high degree of structural conservation, with a low degree of sequence conservation. This paradox raised two questions in our minds. What are the minimal requirements for the active structure of the P-loop? Why is the P-loop structure conserved despite the diverse environments in which it is found?

We have addressed these two questions by introducing the complete P-loop sequence (Kinase 1 motif) and the Asp of the Kinase 2 motif into a protein that has a high level of structural similarity to the CMBF family, but which does not bind a mononucleotide and does not have the Walker motifs: the chemotactic protein from *E. coli*, CheY (Stock *et al*., 1989; Volz & Matsumura, 1991).

Three possible situations could have been observed, depending on the structural autonomy of the two motifs. (1) The P-loop could contain enough information to constrain its structure into the active form independently of the rest of the protein. (2) The P-loop could contain the information but steric hindrances with the original CheY structure force it to adopt a different conformation. (3) Tertiary contacts could be responsible for the conformation of the P-loop in the CMBF family and, therefore, since these are different in CheY, the structure of the loop should adopt a different conformation.

Our results show that the loop adopts a very different conformation from the one expected. Modeling the P-loop into CheY indicates that there are no steric clashes or very unfavorable interactions that could prevent the new loop adopting the same conformation as in p21ras. Another possibility is that the structure of the loop is influenced by the contacts with the neighboring molecule in the crystal. However, NMR analysis of this mutant in solution shows that the P-loop adopts a very similar conformation to the one in the crystal, proving that the X-ray structure is not determined by the packing forces of the crystal lattice (Bruix, M., Cronet, P., Santoro, J., Sander, C., Serrano, L. & Rico, M., unpublished data). This suggests that the structure of the P-loop is strongly influenced by the protein environment (i.e. tertiary contacts, binding of Mg^{2+} or binding of the nucleotide–Mg^{2+}). As observed by Tramontano *et al*. (1989), the conformations of medium-sized loops are often determined by the packing of a particular residue or several residues against the rest of the molecule and by a network of hydrogen bonds. Some of the residues in the core of the antibody framework were found, as in our results, to be important for stabilizing the loops and increasing the affinity for the antigen (Foote & Winter, 1992).

In all the CMBF binding sites we analyzed, residues GKS/T of the Kinase 1 motif are always found in a helical conformation, whereas in our mutant the helix starts three residues after the Gly of the Kinase 1 motif. As indicated by AGADIR and

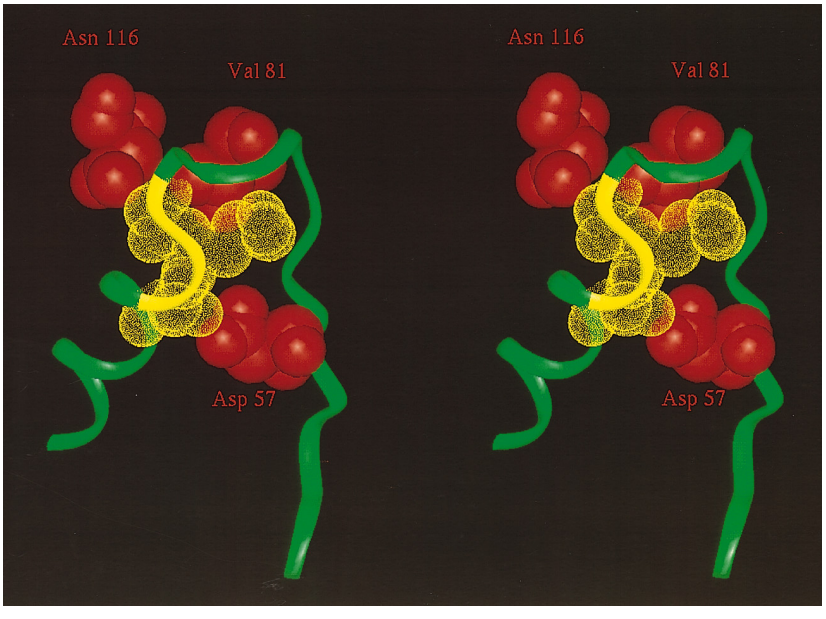

Figure 5. Stereo plot of the tertiary contacts with the first turn of the first helix in p21ras. The backbone of the P-loop is presented as a green ribbon. The 3 residues forming the first turn of the helix (yellow ribbon) are in yellow van der Waals' surface. The residues making conserved tertiary contacts are represented as red spheres; Asp57, Val81 and Asn116.

shown in the NMR analysis of an ADK fragment comprising residues 1 to 45 (Fry *et al*., 1988), this reveals the natural tendency of the sequence in the absence of tertiary contacts with the rest of the protein, or the substrate.

Two external factors could account for the longer helical conformation in the CMBF family. The first candidate is the NTP–Mg^{2+} complex interacting with the amides of the GKS/T sequence as well as with the S/T side-chain of the Kinase 1 motif. This creates favorable interactions that can help the helix to by-pass the capping residue (Ser or Thr), until the next suitable N-cap (Gly17 in our mutant and Gly15 in p21ras). This possibility can be ruled out, since it has been shown that the P-loop of the beef heart mitochondrial adenylate kinase with no ATP in the active site (Diederichs & Schulz, 1991) is almost identical to the one from yeast containing the NTP analog Ap5A. Also, the work of Fry *et al*. (1988), indicates that NTP–Mg^{2+} does not modify the helical content of the first helix of an ADK fragment in solution. The second possibility is the existence of specific tertiary interactions with other elements of the structure. Three main contacts are always present in all the structures we have examined (Table 2). (1) van der Waals' contacts between the side-chain of a buried hydrophobic residue on the β-strand adjacent to the P-loop (Val81 in p21ras) and the backbone of the three residues before the S/T of the Kinase 1 motif. The P-loop covers the hydrophobic residue and shields it from the solvent. In our mutant, Asp57 occupies the equivalent position (to that in wild-type CheY) and consequently conditions for its burial by the P-loop will not be favorable. (2) A hydrogen bond between the S/T of the Kinase 1 motif and the Asp of the Kinase 2 motif. Our mutant contains the Asp of the Kinase 2 motif, although it is in a different conformer than in p21ras. Therefore, it cannot constrain the helix in its active conformation. (3) In the first Gly of the GKS/T triplet, a hydrogen bond is found between the main-chain carbonyl group and a side-chain amide group of an Asn residue in p21ras and EFTu, or a van der Waals' contact with an Arg from an adjacent loop in ADK. In our mutant, Lys109 is in a suitable position to provide this hydrogen bond but it is salt-bridged to Asp57 and therefore unavailable for other interactions.

On the basis of these points we suggest that the hydrophobic cluster made by the packing of the conserved hydrophobic residue (mainly a Val) with the P-loop could play a central role in determining its structure. The other two interactions could help in a synergistic way. A consensus structural model of the nucleotide binding site is proposed in Figure 5.

In order to create the hydrophobic cluster as well as the complete hydrogen bonding network with the P-loop we suggest mutating the Asp57 of CheY to a Val. This will also leave Lys109 free for interacting with Gly17. Asp35 is not in a suitable conformation to bind Mg^{2+} in the crystal structure and this is mainly due to interactions with Arg22. Mutation of Arg22 to a Glu should force the Asp35 to adopt a different rotamer by electrostatic repulsion.

Conclusion

The structure of the P-loop is very conserved amongst the different members of the CMBF. This observation is paradoxical, since the sequence similarity between the different members is low. The results obtained here clearly show that the conformation of the P-loop is very dependent on the tertiary interactions. We propose that the P-loop is anchored in its active conformation through a network of contacts with the rest of the structure, two hydrogen bonds and two van der Waals' contacts with a hydrophobic core on the back of the P-loop. A number of mutations have been proposed in order to help the P-loop to adopt an active conformation. This aspect of loop stability is not limited to this particular case, since similar conclusions were reached about the importance of intra-core residues for the stability of hypervariable loops in antibodies. This principle of protein structure therefore deserves careful consideration in any protein engineering study.

Materials and Methods

Structural comparison of α/β proteins and mutant design

Structural alignments were made using the program DALI (Holm & Sander, 1993). Graphical inspection and visualization were done using the software INSIGHT II (Biosym) and WHATIF (Vriend, 1990). Energy minimizations of the model were carried out using DISCOVER (Biosym). The mutations were examined on the model using the rotamer search option of WHATIF.

Site-directed mutagenesis and protein preparation

The mutants were obtained by site-directed mutagenesis using a method based on the polymerase chain reaction (Landt *et al*., 1990), as previously described (Filimonov *et al*., 1993). The cloning was done using a vector derived from ptz18U (Pharmacia). Detailed procedures for the cloning, expression and purification of the mutant have been described elsewhere (Bruix *et al*., 1993).

Nucleotide binding assay

Blot experiments were carried out with [α-^{32}P]GTP and ATP (Amersham). A 10 μl sample of 100 μM protein solution was put onto a nitrocellulose disk (4 mm diameter) and dried at room temperature. The nitrocellulose disk was saturated with 0.5% (w/v) bovine serum albumin (BSA) in 1 ml 5 mM Tris-HCl (pH 7), 2 mM Mg^{2+}. The buffer was removed and replaced by 1 ml of the same buffer without BSA but with 2 μCi/ml of the appropriate nucleotide (0.003 μM) and incubated for one-hour. Controls were carried out with Rab5 (positive control) and SH3 domain (negative control) at a concentration equal to that of the mutant. The disks were then quickly washed in 1 ml of 5 mM Tris-HCl (pH 7), 2 mM Mg^{2+} and exposed to a phosphoimager screen (Molecular Dynamics) for one hour.

Modeling of the P-loop into CheY

A model of the complete P-loop on the CheY structure was built in order to get a better idea of how the

phosphate/Mg^{2+} binding site could be accommodated in the CheY structure. The model was constructed according to the following procedure. CheY and p21ras were superimposed to get a perfect match of residues CheY-Asp12/p21ras-Gly10 and CheY-Arg18/p21ras-Lys16. The P-loop from p21ras was cut from its original structure and inserted into CheY. Glu35 was mutated into an Asp. The model was then energy minimized for 1000 steps with a steepest descent algorithm fixing all the atoms but the P-loop and the residues at the junctions of the loop with CheY. A further 1000 steps were then applied to the whole structure using the conjugate gradient algorithm.

Analysis of the tertiary contacts with the P-loop

We have searched for all the conserved tertiary contacts made by the residues in the first turn of the first helix (comprising residues GKS/T of the Kinase 1 motif), in all the members of the CMBF family. For this purpose, the option CONTAC of WHATIF was used, defining a maximum distance of 4.5 Å between the van der Waals' surfaces of the residues.

Crystallization, structure determination and refinement

The mutant protein was crystallized by the hanging drop method at 4°C. Drops were prepared by mixing 2 µl of protein solution, containing 40 mg/ml of protein in 5 mM Tris-HCl (pH 7.5), with 2 µl of precipitant solution, containing 30% (w/v) PEG 4000 and 0.2 M Li_2SO_4 in 0.1 M Tris-HCl (pH 8.5). Crystals, with the shape of elongated hexagonal pyramids, appeared after one week. They belong to space group $P6_522$ and have cell dimensions $a = b = 60.5$ Å and $c = 129.6$ Å. X-ray diffraction data were collected from a crystal measuring 0.1 mm × 0.1 mm × 0.4 mm using synchroton radiation at the EMBL Oustation at DESY (Hamburg) on beam line X31 at 4°C. The data were processed with the XDS program (Kabsch, 1988) giving an R_{merge} of 11.1%. From a total number of 74,759 measurements, 6707 independent reflections were observed at the $2\sigma(F_o)$ level. At this level the data set are 90.7% complete (84.5% complete in the highest 2.36 to 2.24 Å resolution shell). The structure was solved by the molecular replacement method using the structure of the CheY *mutant 2* (Bellsolell, L., Cronet, P., Majolero, M., Sander, C., Serrano, L. & Coll, M., unpublished results) as a starting model. CheY *mutant 2* is an intermediate mutant in the present P-loop design carrying the mutations Phe14 to Gly, Ser15 to Gly and Met17 to Gly. Rotation and translation functions were calculated with the program AMoRe (Navaza, 1994) using 15 to 4 Å data. The correct enantiomorph $P6_522$ was established according to the translation function solution. The routine fitting of AMoRe was used for rigid body refinement giving an *R*-factor of 45.3% and a correlation factor of 0.48. Refinement of the structure continued with the program XPLOR (Brünger, 1992), including cycles of simulated annealing, and model building with TURBO (Rousell & Cambillau, 1989). The initial 3 Å resolution Fourier maps calculated with coefficients $2F_o - F_c$ and $F_o - F_c$ clearly indicated that the P-loop region had to be retraced. These residues were omitted from the structure factor calculation and the structure was further refined. New maps calculated at 2.8 Å permitted a new tracing of the P-loop. The final *R*-factor is 20.0% for 6494 reflections $>2\sigma(F_o)$ and 22.4% for all reflections between 8 and 2.24 Å resolution. The final model includes 83 solvent molecules and shows r.m.s.d. from target values of 0.015 Å for bond lengths and 1.80° for bond angles. The atomic coordinates have been deposited with the Brookhaven Protein Data Bank and are available directly from the authors on request until they have been processed and released.

Acknowledgements

This work was supported by grants from the E.C. BRIDGE program (C. S.) from DGICYT (PB92-0117) and the CESCA (M.C.). L.B. is the recipient of a doctoral fellowship from the CSIC.

We thank Montserrat Majolero for doing crystallization experiments, and Alfonso Valencia and Ernest Feytmans for their helpful and critical comments. We are also very grateful to Marta Bruix for sharing the preliminary NMR results prior to publication.

References

Artymiuk, P. J., Rice, D. W., Mitchell, E. M. & Willet, P. (1990). Structural resemblance between the families of bacterial signal-transduction proteins and of G proteins revealed by graph theoretical techniques. *Protein Eng.* **4**, 39–43.

Bellsolell, L., Prieto, J., Serano, L. & Coll. M. (1994). Magnesium binding to the bacterial chemotaxis protein CheY results in large conformational changes involving its functional surface. *J. Mol. Biol.* **238**, 489–495.

Brünger, A. T. (1992). *XPLOR Version 3.1*, Yale University Press, New Haven & London.

Bruix, M., Pascual, J., Santoro, J., Prieto, J. & Serrano, L. (1993). ^{1}H- and ^{15}N-NMR assignment and solution structure of the chemotactic *E. coli* CheY protein. *Eur. J. Biophys.* **215**, 573–585.

Chen, J. M., Lee, G., Murphy, R., Brandt-Rauft, P. & Pincus, M. R. (1990). Comparison between the three-dimensional structures of the chemotactic protein CheY and the normal G12-p21ras protein. *Int. J. Pept. Protein Res.* **36**, 1–6.

Cherfils, J., Moréra, S., Lascu, I., Véron, M. & Nanin, J. (1994). X-ray structure of nucleoside diphosphate kinase complexed with thymidine diphosphate and Mg^{2+} at 2 Å resolution *Biochemistry*, **33**, 9062–9069.

Dever, T. E., Glynias, M. J. & Merrick, C. (1987). GTP-binding domain: three consensus sequence elements with distinct spacing. *Proc. Natl Acad. Sci., U.S.A.* **84**, 1814–1818.

DeVos, A. M., Tong, L., Milburn, M. V., Matias, P. M., Jancarik, J., Noguchi, S., Nishimura, S., Muira, K., Othsuka, E. & Kim, S.-H. (1988). Three-dimensional structure of an oncogene protein: catalytic domain of human c-H-*ras* p21. *Science*, **239**, 888–893.

Diederichs, K. & Schulz, G. E. (1991). The refined structure of the complex between adenylate kinase from beef heart mitochondrial matrix and its substrate AMP at 1.85 Å resolution. *J. Mol. Biol.* **217**, 541–549.

Dreusicke, D. & Schulz, G. E. (1986). The glycine-rich loop of adenylate kinase forms a giant anion hole. *FEBS Letters*, **208**, 301–304.

Dreusicke, D., Karplus, P. A. & Schulz, G. E. (1988). The refined structure of the procine cytosolic adenylate kinase at 2.1 Å resolution. *J. Mol. Biol.* **199**, 359–371.

Fersht, A. & Serrano, L. (1993). Principles of protein

stability derived from protein engineering experiments. *Curr. Opin. Struct. Biol.* **3**, 75–83.

Filimonov, V. V., Prieto, J., Martinez, J. C., Bruix, M., Mateo, P. & Serano, L. (1993). Thermodynamic analysis of the chemotactic protein from *Escherichia coli*, CheY. *Biochemistry,* **32**, 12906–12920.

Flaherty, K. M., McKay, D. B., Kabsch, W. & Holmes, K. C. (1991). Similarity of the three-dimensional structure of actin and the ATPase fragment of a 70 kDa heat shock cognate protein. *Proc. Natl Acad. Sci., U.S.A.* **88**, 5041–5045.

Foote, J. & Winter, G. (1992). Antibody framework residues affecting the conformation of the hypervariable loops. *J. Mol. Biol.* **224**, 487–499.

Fry, D. C., Byler, M., Susi, H., Brown, E. M., Kuby, S. A. & Mildvan, A. S. (1988). Solution structure of the 45-residue Mg-ATP binding peptide of adenylate kinase as examined by 2-D NMR, FTIR and CD spectroscopy. *Biochemistry,* **27**, 3588–3598.

Hess, J. F., Bourret, R. B., Oosawa, K., Matsumura, P. & Simon, M. I. (1988). Protein phosphorylation and bacterial chemotaxis. *Cold Spring Harbor Symp. Quant. Biol.* **53**, 41–47.

Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.

Huang, W., Linqvist, Y., Schneider, G., Gibson, K. J., Flint, D. & Lorimer, G. (1994). Crystal structure of an ATP-dependent carboxylase, dethiobiotin synthetase, at 1.65 Å resolution. *Structure,* **2**, 407–414.

Kabsch, W. (1988). Evaluation of single crystal X-ray diffraction from a position-sensitive detector. *J. Appl. Crystallog.* **21**, 916–924.

Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. (1990). Atomic structure of the actin: DNase-I complex. *Nature,* **347**, 37–44.

Knighton, D. R., Zheng, J., Ten Eyck, L. F., Ashord, V. A., Xuong, N. H., Taylor, S. S. & Sowadsky, J. M. (1991). Crystal structure of the catalytic subunit of the cyclic adenosine monophosphate-dependent kinase. *Science,* **253**, 407–414.

La Cour, T. F. M., Nyborg, J., Thirup, S. & Clark, B. F. C. (1985). Structural details of the binding of guanosine diphosphate to elongation factor Tu from *E. coli* as studied by X-ray crystallography. *EMBO J.* **4**, 2385–2388.

Lambright, D. G., Noel, J. P., Hamm, H. E. & Sigler, P. B. (1994). Structural determinants for activation of the α-subunit of a heterotrimeric G protein. *Nature,* **369**, 621–628.

Landt, O., Grunert, H. P. & Hahn, U. (1990). A general method for rapid site-directed mutagenesis using the polymerase chain reaction. *Gene,* **96**, 125–128.

Moras, D. (1992). Structural and functional relationships between the aminoacyl-tRNA synthetases. *Trends Biochem. Sci.* **17**, 159–164.

Muñoz, V. & Serrano, L. (1994). Elucidating the folding problem of helical peptides using empirical parameters. *Nature Struct. Biol.* **1**, 399–409.

Muñoz, V. & Serrano, L. (1995). Elucidating the folding problem of α-helical peptides using empirical parameters, II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol.* **245**, 275–296.

Muñoz, V., Lopez, E. M., Jager, M. & Serrano, L. (1994). Kinetic characterisation of the chemotactic protein from *Escherichia coli*, CheY. Kinetic analysis of the inverse hydrophobic effect. *Biochemistry,* **33**, 5858–5866.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect., A***50**, 157–163.

Noel, J. P., Hamm, H. E. & Sigler, P. B. (1993). The 2.2 Å crystal structure of transducin-α complexed with GTPγS. *Nature,* **366**, 654–663.

Pai, E. F., Krengel, U., Petsko, G. A., Goody, R. S., Kabsch, W. & Wittinghofer, A. (1990). Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 Å resolution: implications for the mechanism of GTP hydrolysis. *EMBO J.* **9**, 2351–2359.

Parkinson, J. S. (1983). Complementation analysis and deletion mapping of *Escherichia coli* mutants defective in chemotaxis. *J. Bacteriol.* **155**, 265–274.

Richardson, J. S. & Richardson, D. C. (1988). Amino acid preferences for specific locations at the ends of α helices. *Science,* **240**, 1648–1652.

Rousell, A. & Cambillau, C. (1989). TURBO-FRODO. In *Silicon Graphics Geometry Partners Directory* (Fall 1989), pp. 77–78, Silicon Graphics, Mountain View, CA.

Sanders, D. A., Gillece-Castro, B. L., Stock, A. M., Burlingame, A. L. & Koshland, D. E. (1989). Identification of the phosphorylation site of the chemotaxis response regulator protein, CheY. *J. Biol. Chem.* **264**, 21770–21778.

Saraste, M., Sibbald, P. R. & Wittinghofer, A. (1990). The P-loop, a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430–434.

Schulz, G. E. (1992). Binding of nucleotides by proteins. *Curr. Opin. Struct. Biol.* **2**, 61–67.

Schulz, G. E., Elzinga, M., Marx, M. & Schrimer, R. H. (1974). Three-dimensional structure of adenylate kinase. *Nature,* **250**, 120–123.

Stock, A. M., Mottonen, J. M., Stock, J. B. & Schutt, C. E. (1989). Three-dimensional structure of CheY, the response regulator of bacterial chemotaxis. *Nature,* **377**, 745–749.

Stock, A. M., Martinez-Hackert, E., Rasmussen, B. F., West, A. H., Stock, J. B., Ringe, D. & Petsko, G. A. (1994). Structure of the Mg^{2+}-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis. *Biochemistry,* **49**, 13375–13380.

Story, R. M. & Steitz, T. A. (1992). Structure of the recA protein-ADP complex. *Nature,* **355**, 374–376.

Tramontano, A., Chothia, C. & Lesk, A. M. (1989). Structural determinants of the conformations of medium-sized loops in proteins. *Proteins: Struct. Funct. Genet.* **6**, 382–394.

Traut, T. W. (1994). The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. *Eur. J. Biochem.* **222**, 9–19.

Volz, K. & Matsumura, P. (1991). Crystal structure of *Escherichia coli* CheY refined at 1.7 Å resolution. *J. Biol. Chem.* **266**, 15511–15519.

Vriend, G. (1990). WHATIF: A molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52–55.

Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982). Distantly related sequence in the α- and β-subunit of ATP synthetase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945–951.

Wierenga, R. K., De Maeyer, M. C. H. & Hol, W. G. J. (1985). Interaction of the phosphate moieties with α-helices in dinucleotide binding proteins. *Biochemistry,* **24**, 1346–1357.

Yamagushi, H., Kato, H., Hata, Y., Nishioka, T., Kimura, A., Oda, J. & Katsube, Y. (1993). Three-dimensional structure of the glutathione synthetase from

Escherichia coli B at 2 Å resolution. *J. Mol. Biol.* **229**, 1083–1100.

Zheng, J., Knighton, D. R., Ten Eyck, L. F., Karlsson, R., Xuong, N.-H., Taylor, S. S. & Sowadsky, J. M. (1993). Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. *Biochemistry*, **32**, 2154–2161.

Edited by F. E. Cohen

(Received 4 November 1994; accepted 28 February 1995)