doi:10.1006/jmbi.2001.5286 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2002) **315**, 1099–1110

JMB

Sugar Specificity of Bacterial CMP Kinases as Revealed by Crystal Structures and Mutagenesis of *Escherichia coli* Enzyme

Thomas Bertrand[1], Pierre Briozzo[2]*, Liliane Assairi[3] Augustin Ofiteru[3,4], Nadia Bucurenci[4], Hélène Munier-Lehmann[3] Béatrice Golinelli-Pimpaneau[1], Octavian Bârzu[3] and Anne-Marie Gilles[3]

[1]*Laboratoire d'Enzymologie et de Biochimie Structurales, UPR 9063 du CNRS, 91198 Gif-sur-Yvette Cedex, France*

[2]*Laboratoire de Chimie Biologique, Institut National Agronomique Paris-Grignon 78850, Thiverval-Grignon France*

[3]*Laboratoire de Chimie Structurale des Macromolécules, URA 2185 du CNRS, Institut Pasteur 75724 Paris Cedex 15, France*

[4]*Institutul Cantacuzino, 70100, Bucharest, Romania*

**Corresponding author*

Bacterial cytidine monophosphate (CMP) kinases are characterised by an insert enlarging their CMP binding domain, and by their particular substrate specificity. Thus, both CMP and 2′-deoxy-CMP (dCMP) are good phosphate acceptors for the CMP kinase from *Escherichia coli* (*E. coli* CMPK), whereas eukaryotic UMP/CMP kinases phosphorylate the deoxynucleotides with very low efficiency. Four crystal structures of *E. coli* CMPK complexed with nucleoside monophosphates differing in their sugar moiety were solved. Both structures with CMP or dCMP show interactions with the pentose that were not described so far. These interactions are lost with the poorer substrates AraCMP and 2′,3′-dideoxy-CMP. Comparison of all four structures shows that the pentose hydroxyls are involved in ligand-induced movements of enzyme domains. It also gives a structural basis of the mechanism by which either ribose or deoxyribose can be accommodated. In parallel, for the four nucleotides the kinetic results of the wild-type enzyme and of three structure-based variants are presented. The phosphorylation rate is significantly decreased when either of the two pentose interacting residues is mutated. One of these is an arginine that is highly conserved in all known nucleoside monophosphate kinases. In contrast, the other residue, Asp185, is typical of bacterial CMP kinases. It interacts with Ser101, the only residue conserved in all CMP binding domain inserts. Mutating Ser101 reduces CMP phosphorylation only moderately, but dramatically reduces dCMP phosphorylation. This is the first experimental evidence of a catalytic role involving the characteristic insert of bacterial CMP kinases. Furthermore, this role concerns only dCMP phosphorylation, a feature of this family of enzymes.

Keywords: bacterial CMP kinases; deoxynucleotide phosphorylation; nucleoside monophosphate binding domain insertion; site-directed mutagenesis; X-ray crystallography

Introduction

Nucleoside monophosphate (NMP) kinases are key enzymes in the metabolism of nucleotides. They catalyse the reversible transfer of a γ phosphoryl group from a nucleoside triphosphate to a particular nucleoside monophosphate.[1] Cytidine monophosphate kinase from *Escherichia coli* (*E. coli* CMPK), a monomer of 227 residues, specifically phosphorylates CMP (or dCMP), using ATP as the preferred phosphoryl donor, according to the scheme:

Abbreviations used: AraCMP, cytosine β-D-arabinofuranoside-5′-monophosphate; CDP, cytidine-5′-diphosphate; CMP, cytidine-5′-monophosphate; CMPK, cytidine monophosphate kinase; dCMP, 2′-deoxy-CMP; ddCMP, 2′,3′-dideoxy-CMP; NMP, nucleoside monophosphate; NMPb-insert, the insertion of the NMP binding domain; r.m.s.d., root mean square deviation.

E-mail address of the corresponding author: briozzo@grignon.inra.fr

0022-2836/02/051099–12 $35.00/0

$$ATP \cdot Mg^{2+} + CMP \leftrightarrow ADP \cdot Mg^{2+} + CDP$$

The CDP and 2′-deoxy-CDP produced are further phosphorylated to produce CTP and 2′-deoxy-CTP, precursors of the major biological molecules RNA, DNA, and phospholipids.

Concerning the enzymes involved in the phosphorylation of pyrimidine nucleoside monophosphates, three major differences are observed between bacteria and eukaryotes.

Firstly, in eukaryotes a single enzyme phosphorylates UMP and CMP with similar efficiency. Conversely, bacterial CMP kinases phosphorylate UMP with very low rates, this role being assigned to distinct UMP kinases.

Secondly, *E. coli* CMPK phosphorylates dCMP nearly as well as it does CMP. This is in contrast to eukaryotic UMP/CMP kinases, which phosphorylate the corresponding deoxynucleotides with very low efficiency.[2] Mutants of *E. coli* devoid of CMP kinase activity have a level of CTP comparable to that found in wild-type bacteria, due to the enzyme CTP synthetase which produces CTP through UTP amination. In contrast, the dCTP pool and the rate of DNA replication are reduced in the mutants: ribonucleoside diphosphate reductase, which catalyses the formation of 2′-deoxyribonucleoside diphosphates from the corresponding ribonucleotides, cannot fully compensate for the lack of *E. coli* CMPK.[3] Therefore, phosphorylation of dCMP appears as one of the main roles of CMP kinase in *E. coli*.

Thirdly, the kinetic differences between eukaryotic UMP/CMP kinases and bacterial CMP kinases are echoed by sequence differences. All known NMP kinases have an overall well-conserved fold with three domains:[4,5] the CORE domain, containing a five-stranded parallel β-sheet; the LID domain, which like a lid closes the active site upon binding of the phosphate donor ATP; the NMP binding domain, which closes the active site upon binding of the phosphate acceptor. Classically, NMP kinases are divided into short forms, including eukaryotic UMP/CMP kinases, and long forms. The latter group consists of adenylate kinases with an insertion of around 27 residues into the LID domain. Bacterial CMP kinases represent a third distinct family of NMP kinases, as they possess a short LID domain but have an insertion of around 40 residues into their NMP binding domain (NMPb-insert).

We previously solved the first structure of a bacterial CMP kinase, that of *E. coli*, either alone or in complex with CDP.[6] Binding of CDP induces a movement of the NMP-binding domain which is different from that of the two structurally characterised eukaryotic UMP/CMP kinases: those from yeast[7] and from amoeba.[8] The CDP-CMPK structure explained the specificity of the *E. coli* enzyme for CMP *versus* UMP, but neither its particular ability to phosphorylate dCMP nor the potential role of the NMPb-insert.

To further investigate the phosphorylation of deoxynucleotides by *E. coli* CMPK, we solved its structure in complex with the two natural substrates CMP and dCMP. Then, we looked at the complexes with two nucleotide analogues modified on the pentose moiety, cytosine β-D-arabinofuranoside-5′-monophosphate (AraCMP) and 2′,3′-dideoxy-CMP (ddCMP), which are poorer substrates.

In addition, we tested by site-directed mutagenesis the importance of three residues which, according to crystal structure analysis, were either directly interacting with the pentose in the complexes with CMP or dCMP, or H-bonded to such an interacting residue. The catalytic efficiencies of the three resulting variants for CMP, dCMP, AraCMP and ddCMP are interpreted in relation to crystal structures. The conservation among bacterial CMP kinases of the mutated residues is discussed.

Results

Comparison of *E. coli* CMPK structures complexed either with the substrate CMP or the product CDP

Overall, the CMP-CMPK structure (Figure 1(a)) is the same as that previously published in complex with CDP. The CORE domain, which contains five parallel β strands connected by α helices, and the short LID domain are similar to those of short NMP kinases. The NMP binding domain contains a 40-residue insert (blue trace, Leu63 to Gln102 in *E. coli* CMPK) typical of bacterial CMP kinases: this insert contains three antiparallel β strands and two α helices, the longest one being by far α5 (labelled on Figure 1(b)). When CMP or CDP is bound, the helical part of the NMPb-insert moves closer to the ligand in an induced-fit movement.

A comparison of the active sites with CMP and CDP (Figure 2, top) indicates that the same residues interact with the α phosphate or the cytosine of the two nucleotides. Residues that are H-bonded to the α phosphate (Arg41, Arg131) are conserved in all known NMP kinases, whereas most residues H-bonded to the cytosine (Asp132, Arg110, Arg188) are conserved only in bacterial CMP kinases (see ref.[2] for an alignment of amino acid sequences from different NMP kinases with *E. coli* CMPK).

However, the interactions with the sugar moiety are quite different in the two complexes. With CDP the only direct polar interaction between the *E. coli* enzyme and the ribose is the hydrogen bond involving the 2′-hydroxyl group. Whereas the carboxylate group of Asp185 interacts only with the 2′-OH of CDP, it is H-bonded to the two hydroxyl groups of the CMP ribose. The other important difference with CDP-CMPK concerns Arg181, which is H-bonded to both the 3′-OH and the α phosphate of CMP. In the presence of CDP the side-chain of this arginine points away from the nucleotide.

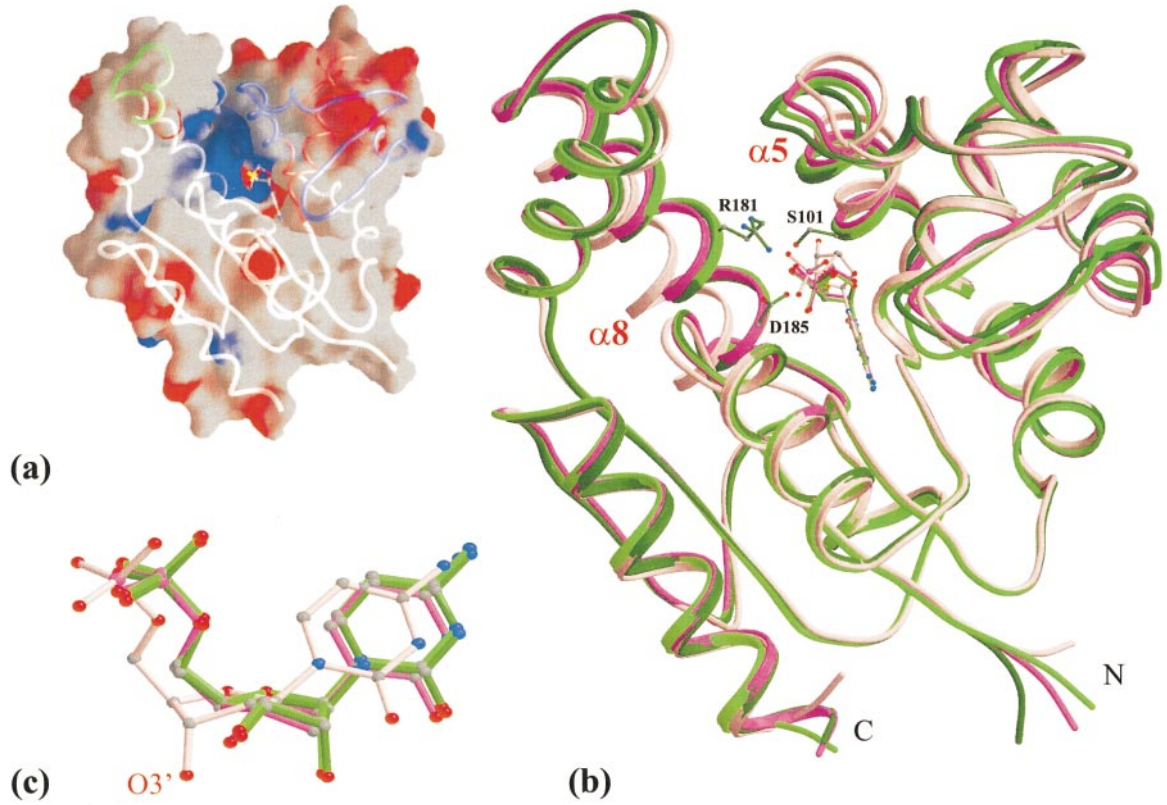

Figure 1. Conformations of *E. coli* CMPK with the natural substrates CMP and dCMP. (a) Molecular surface of CMP-CMPK viewed along the active site, showing the substrate cavity with a blue positive zone around the α phosphate. The backbone is shown in transparency: white for the CORE domain, green for the short LID domain and red and blue (insert) for the NMP binding domain. The ball-and-stick depicts CMP. Drawn with GRASP.[28] (b) Backbone superposition of four structures: the two molecules A and B of an asymmetric unit either from CMP-CMPK or from dCMP-CMPK structures. The orientation is the same as in (a). With CMP molecule A is dark green, with some residues in ball-and-stick, and molecule B light green. With dCMP molecule A is magenta, molecule B is pink. Sticks of the bound substrate are coloured as the corresponding protein. Carbon atoms are grey, oxygen atoms red, nitrogen atoms blue and phosphorus atoms magenta. (c) Magnification of the substrates in a plane parallel to the cytosine rings, which emphasises the conformations of the pentose. Drawn with MOLSCRIPT[29] and Raster3D.[30]

The conformation of the ribose is C3′-*endo* in CMP-CMPK crystals (C3′ for green models is over the sugar ring in the orientation of Figure 1(c)), as opposed to C2′-*endo* in CDP-CMPK. C3′-*endo* and C2′-*endo* conformations, the two principal sugar puckering modes, are in rapid equilibrium in solution.[9] The C3′-*endo* conformation preferred for the bound CMP brings the 3′-OH closer to Arg181 and Asp185. This allows a bidentate interaction of the carboxylate of Asp185 with both hydroxyls of the ribose.

The structure of *E. coli* CMPK complexed with dCMP

The crystal structure with dCMP shows different closures of helices α5 and α8

Both CMP-CMPK and dCMP-CMPK crystals belong to the $P2_12_12_1$ space group, with very similar unit cells (Table 1), and two molecules per asymmetric unit. The two molecules of the asymmetric unit are essentially identical in the structure with CMP (Figure 1(b), green models), in contrast to the structure with dCMP: the calculated root mean square deviation (r.m.s.d.) for C^{α} atoms between the two molecules of an asymmetric unit is 0.89 Å with dCMP, and only 0.45 Å with CMP. With dCMP, most of the backbone remains very similar, but significant variations are observed for helices α8 (Phe171 to Asn187) and to a lesser extent α5 (Arg92 to Ala104). These are the two helices that move towards each other to close the cleft when a nucleotide binds, like two opposite jaws. For the sake of clarity, we will call molecule A of an asymmetric unit that which has the narrower cleft, and molecule B the more open one. The dCMP-CMPK structure residues of α5 and α8 involved in crystal contacts are identical for both molecules of an asymmetric unit. Thus, the differences observed between molecules A and B are not an artefact due to a particular crystal packing in this region. Molecule A from dCMP-CMPK (Figure 1(b), magenta) is very similar to those with CMP. In molecule B (pink), the left jaw α8 moves

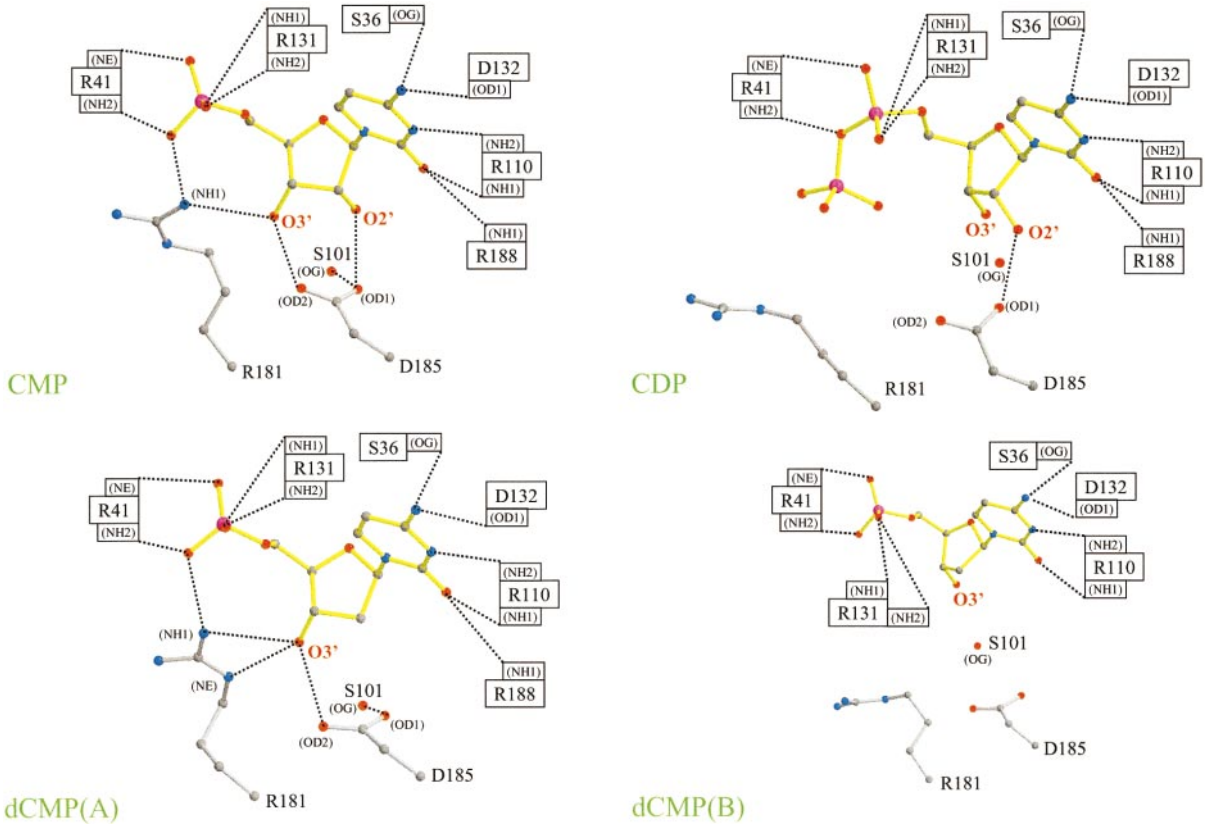

Figure 2. Polar interactions of *E. coli* CMPK with its natural substrates CMP or dCMP, and with its natural product CDP. Top left: with CMP, valid for both molecules of the asymmetric unit. Top right: with CDP (crystals contained one molecule per asymmetric unit). Bottom: with dCMP. For molecule B of the asymmetric unit the magnification is smaller in order to allow visualisation of the distant residues Arg181 and Asp185. Nucleotides have yellow sticks. Enzyme residues H-bonded to the cytosine or the α phosphate of the nucleotides are simply labelled, as they are essentially comparable in all structures. Conversely, side-chains of residues that can be H-bonded to the pentose are shown with grey sticks, and the gamma oxygen of Ser101 as a ball. Dotted lines indicate atoms distant from less than 3.2 Å. CMP-CMPK and dCMP-CMPK structures were superposed on the previously published structure of CDP-CMPK. The orientation, chosen to avoid overlapping of protein residues with bound ligands, differs from those of Figure 1.

further from dCMP (the calculated r.m.s.d. for α8 between molecules A and B is 2.11 Å with dCMP, whereas it is only 0.31 Å with CMP). The right jaw, helix α5 from the NMPb-insert, also moves further from dCMP (r.m.s.d. for α5 is 1.48 Å with dCMP, and 0.49 Å with CMP).

With dCMP, only the molecule A makes polar interactions like those with CMP

The bottom of Figure 2 shows the enzyme interactions with dCMP in molecules A and B. In molecule A, Arg181 interacts with the α phosphate and the 3′-OH as it does with CMP, although the position of the guanidinium group is different. As is the case with CMP, the side-chain of Asp185 interacts with the 3′-OH and with Ser101. Thus, the same residues interact with the ribose and the deoxyribose. The only difference is that the carboxylate of Asp185 no longer has the 2′-OH partner for hydrogen bonding.

In molecule B with dCMP, Arg181 and Asp185 from α8 do not interact anymore with the 3′-OH, and Arg188, a residue adjacent to α8, does not interact with the cytosine. This is consistent with the fact that α8 is further from dCMP than in molecule A. Similarly, the remoteness of α5 from dCMP is probably related to the loss of interaction between the side-chains of Ser101 from α5 and Asp185 from α8. However, most polar interactions of the enzyme with the α phosphate or the cytosine from the nucleotide substrate are conserved in all structures with CDP, CMP and dCMP. Thus, in the relatively more open conformation of molecule B, dCMP remains close to the position observed in molecule A, except that it penetrates less deeply into the binding cleft (see Figure 1(b)).

Another difference between molecules A and B is the sugar conformation (see Figure 1(c)). It is C3′-*endo* in A, as was the case for CMP, whereas it is C2′-*endo* in B (pink model), as with CDP. As a

Table 1. Structural data

Data set	CMP	dCMP	AraCMP	ddCMP
Data collection				
Wavelength (Å)	0.987	1.001	0.968	0.933
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å)				
a	74.09	73.78	73.58	72.65
b	75.74	75.98	73.77	75.31
c	77.99	78.09	77.64	78.04
Resolution (Å)	20-1.9	30-2.3	15-2.25	35-1.95
Observed reflections	336,628	345,577	196,124	503,844
Unique reflections	35,160	20,232	19,192	32,964
Completeness (%)	94.0 (96.4)[a]	96.4 (95.7)	93.6 (94.9)	93.0 (96.1)
$I/\sigma(I)$	12.7 (7.68)	25.5 (7.30)	14.0 (4.25)	9.93 (2.61)
R_{sym}[b] (%)	7.1 (12.1)	4.5 (16.7)	8.3 (27.4)	7.3 (39.6)
Refinement statistics				
R_{cryst}[c] (%)	20.2	21.9	19.9	22.1
R_{free}[d] (%)	24.8	25.5	25.4	25.3
Residues	3-225	3-225	3-223	3-225
Number of water molecules	451	385	80	433
R.m.s.d.				
bond lengths (Å)	0.0043	0.0096	0.0054	0.0059
bond angles (°)	1.08	1.25	1.26	1.19
Average *B* factor (Å^2)				
main chain	16.4	30.8	26.3	28.4
side-chain	18.5	31.6	25.3	31.5
nucleotide	11.9	20.8	28.5	23.3
water molecules	23.0	43.2	29.5	39.6
sulphate S1	12.5	25.6	44.3	26.7
Ramachandran statistics[e] (%)				
most favoured regions	90.8	90.5	88.9	90.8
additional allowed regions	8.8	9.3	10.4	8.8

[a] Numbers in parentheses represent values in the highest resolution shell (last of 10 shells).
[b] $R_{sym} = \Sigma_h\Sigma_i\ |I(h,i) - \langle I(h)\rangle| / \Sigma_h\Sigma_i\ I(h,i)$ where $I(h,i)$ is the intensity value of the ith measurement of h and $\langle I(h)\rangle$ is the corresponding mean value of $I(h)$ for all i measurements.
[c] $R_{cryst} = \Sigma\ \ ||F_{obs}| - |F_{calc}|| / \Sigma\ \ |F_{obs}|$, where $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure factor amplitudes respectively.
[d] R_{free} is the same as R_{cryst} but calculated with a 10 % subset of all reflections that was never used in crystallographic refinement.
[e] As evaluated by PROCHECK.[32]

result, the 3′-OH from the pentose has very different orientations in molecules A and B.

Structures with pentose modified nucleotide analogues

Common features of both structures with AraCMP or ddCMP

To further investigate the role of the pentose hydroxyls, we solved the crystal structures of *E. coli* CMPK with the monophosphates of two nucleoside analogues that have altered affinities for the enzyme (Table 1): aracytidine (Cytarabine), a drug used in cancer chemotherapy, and 2′,3′-dideoxy-cytidine (ddC), used against Human Immunodeficiency Virus.

As was the case with dCMP, the asymmetric unit of both crystals contains two molecules which are rather different (r.m.s.d. between C^α positions: 0.73 Å for AraCMP, 0.78 Å for ddCMP). Besides, with AraCMP and ddCMP the cleft is more widely open than is the case with CMP.

Figure 3 shows the enzyme-nucleotide polar interactions in molecules A and B for both analogues, and the omit maps of the substrates. Despite their high K_m, the electron densities are very clear. At first sight, the two nucleotide analogues have structures that are close to those of the natural substrates. Thus, in molecules A the residues that interact with the α phosphate and with the cytosine of AraCMP or ddCMP also interact with CDP, CMP and dCMP (molecule A). Nevertheless, in the presence of the analogues several interactions with the enzyme are lost: no hydrogen bond connects the α phosphate to Arg181. Then, in molecules B Arg188 does not interact with the bound NMP, as is the case in dCMP (molecule B), but in addition Arg131 is no longer H-bonded to the α phosphate. Compared to the natural substrates CMP and dCMP, the weaker polar interactions of the analogues with the enzyme easily explain their relatively low affinity.

Changing the orientation of the 2′-OH: characteristics of the AraCMP-CMPK structure

AraCMP, an arabinonucleotide, differs from its ribo analogue in the altered configuration at C2′: the 2′-hydroxyl is *cis*-oriented to the glycosyl C1′-N bond. Whereas the absence of the 2′-OH in dCMP is of little consequence on the phoshorylation rate,

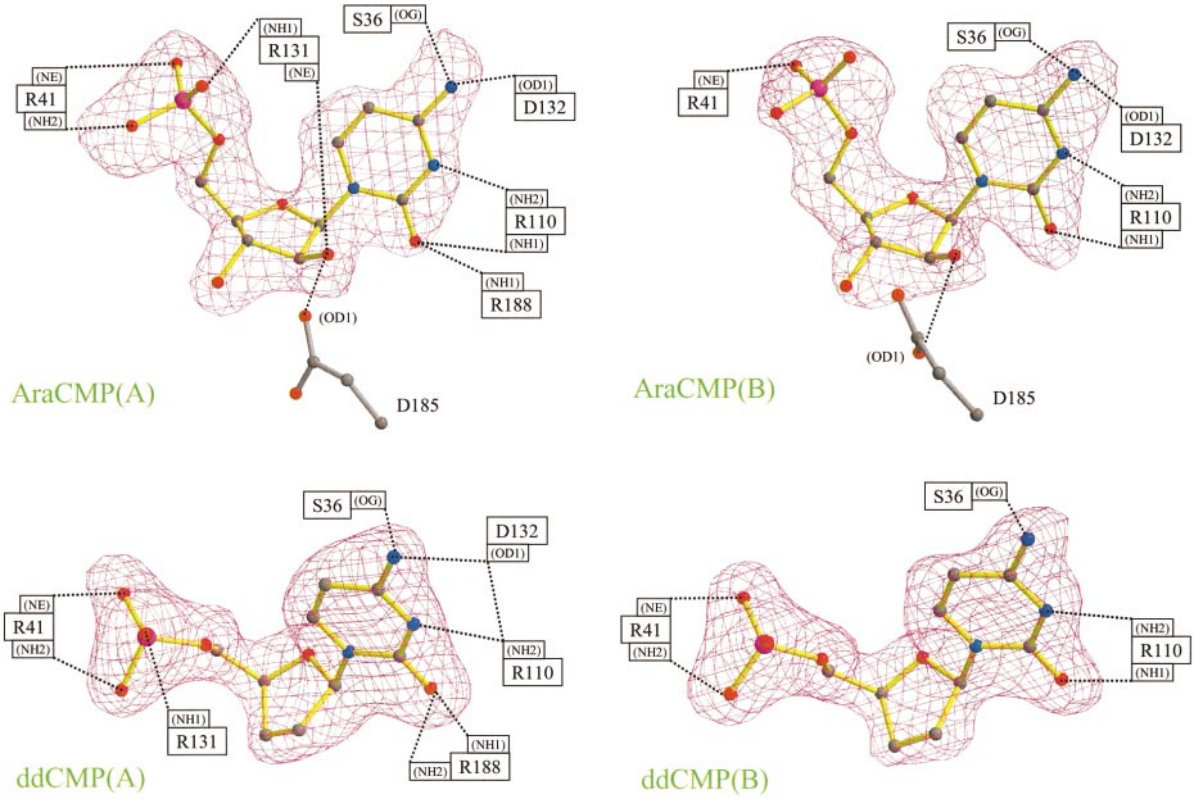

Figure 3. Polar interactions with nucleotide analogues. $F_o - F_c$ omit maps (magenta) were calculated in the absence of the nucleotide model. The contour level, chosen to give comparable volumes of the maps, is 2σ (molecule A) and 1.5σ (molecule B) for AraCMP, and 3.8σ for ddCMP. The orientation is the same as in Figure 2 for ddCMP, and slightly different for AraCMP in order to emphasise the 2′-OH orientation. Drawn with BobScript.[31]

its inversion in AraCMP has a greater effect: k_{cat}/K_m is ten times lower than with dCMP (Table 2). In the crystal structure, for both molecules A and B the 2′-OH typical of AraCMP establishes an hydrogen bond with Asp185. This interaction requires that Asp185 side-chain moves so that it no longer interacts with the 3′-OH and Ser101. Thus the bidentate interaction observed with CMP is lost. However, the 3′-OH orientation remains close to those observed with the natural substrates CMP and dCMP.

Suppressing the 3′-OH: characteristics of the ddCMP-CMPK structure

ddCMP lacks both hydroxyl groups of the pentose: the 2′-OH, like dCMP, and in addition the 3′-OH. Kinetic experiments indicate that it is an

Table 2. Kinetic parameters of *E. coli* CMP kinase with four NMPs

Enzyme	t_m (°C)		CMP	dCMP	AraCMP	ddCMP
Wild-type	52	K_m (mM)	0.035	0.094	0.53	0.46
		k_{cat} (s^{-1})	103	109	56	0.047
		k_{cat}/K_m (s^{-1} mM^{-1})	2940	1160	105	0.102
D185A	49	K_m (mM)	0.47	0.24	1.0	0.15
		k_{cat} (s^{-1})	0.26	0.071	0.085	0.0083
		k_{cat}/K_m (s^{-1} mM^{-1})	0.54	0.30	0.083	0.056
R181M	53	K_m (mM)	0.19	0.24	0.79	0.54
		k_{cat} (s^{-1})	1.38	0.45	1.36	0.12
		k_{cat}/K_m (s^{-1} mM^{-1})	7.4	1.9	1.7	0.22
S101A	52	K_m (mM)	0.08	0.19	0.47	0.65
		k_{cat} (s^{-1})	56	1.2	3.6	0.0033
		k_{cat}/K_m (s^{-1} mM^{-1})	697	6.1	7.5	0.0059

Activity was determined at 30 °C and pH 7.4; ATP was constant (1 mM) whereas the concentration of various NMPs varied between 0.02 and 10 mM. t_m is the temperature at which the denaturation of the protein is half completed. k_{cat} was calculated assuming a molecular mass of *E. coli* CMP kinase of 24.8 kDa.

extremely poor substrate of the wild-type *E. coli* CMPK (Table 2). The k_{cat}/K_m value for ddCMP is at least 10,000 times lower than for dCMP.

Concerning the opening of the α5 and α8 jaws, conformations with ddCMP are close to those with dCMP: for both deoxynucleotides, molecule B is significantly more open than molecule A (not shown). Concerning the enzyme-ligand interactions (Figure 3, bottom), the absence of the 3′-OH results in the loss of the polar interaction between Arg181 and the α phosphate observed with dCMP. For the particular case of molecule B, the interaction with Asp132 is lost.

Site-directed mutagenesis

Substitution of residues directly interacting with the sugar moiety: kinetic parameters of the D185A and R181M variants

As the pentose makes direct interactions with Asp185 in the CDP, CMP, dCMP (molecule A) and AraCMP complexes, and with Arg181 for CMP and dCMP(A), we mutated these two positions. The mid-point of thermal denaturation (t_m) for both variants is close to that of the wild-type enzyme, suggesting that these mutations do not alter the protein structure. Table 2 gives the kinetic parameters of the resulting enzymes. Also shown are those of the wild-type enzyme, which phosphorylates dCMP nearly as well as CMP. For nucleotide analogues, both K_m values of the wild-type enzyme are at least ten times as high as that for CMP, but whereas k_{cat} for AraCMP remains half that with CMP, it is drastically reduced with ddCMP (more than 2000 times lower).

Concerning the natural nucleotides CMP and dCMP, the D185A variant has higher K_m values than the wild-type enzyme, but the difference is at most one order of magnitude. On the other hand, the k_{cat} of this variant is reduced by a factor of 400 with CMP, and 1500 with dCMP. Mutation R181M has the same results concerning K_m values; the results concerning k_{cat} values are similar, although less pronounced, than for mutation D185A. On the whole, both mutations exhibit dramatic consequences on k_{cat}/K_m towards CMP and dCMP (the

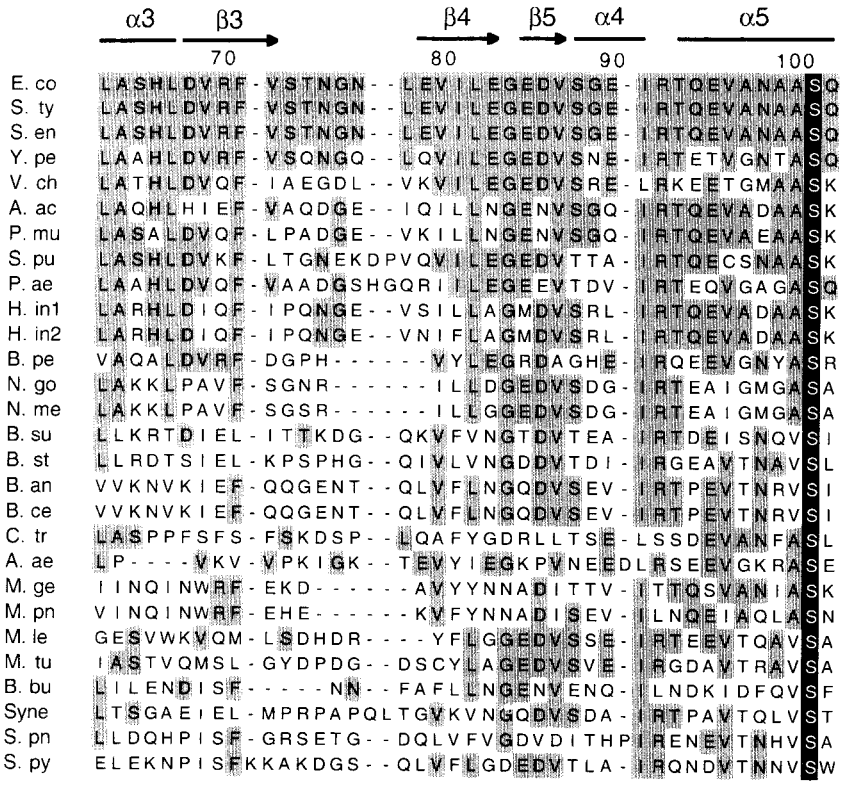

Figure 4. Sequence alignment of the NMP binding domain insert characteristic of bacterial CMP kinases. Alignment made by BLAST.[10] Residues identical to that of *E coli* CMPK have a grey background, and Ser101, the only strictly conserved residue, a black one. The secondary structure assignment and residue numbering for *E. coli* CMPK are shown above. E. co, *Escherichia coli*; S. ty, *Salmonella typhi*; S. en, *Salmonella enteritidis*; Y. pe, *Yersinia pestis*; V. ch, *Vibrio cholerae*; A. ac, *Actinobacillus actinomycetemcomitans*; P. mu, *Pasteurella multocida*; S. pu, *Shewanella putrefaciens*; P. ae, *Pseudomonas aeruginosa*; H. in, *Haemophilus influenzae*; B. pe, *Bordetella pertussis;* N. go, *Neisseria gonorrhoea;* N. me, *Neisseria meningitidis*; B. su, *Bacillus subtilis*; B. st, *Bacillus stearothermophilus*; B. an, *Bacillus anthracis*; B. ce, *Bacillus cereus*; C. tr, *Chlamydia trachomatis*; A. ae, *Aquifex aeolicus*; M. ge, *Mycoplasma genitalium*; M. pn, *Mycoplasma pneumoniae*; M. le, *Mycobacterium leprae*; M. tu, *Mycobacterium tuberculosis*; B. bu, *Borrelia burgdorferi*; Syne, *Synechocystis PCC6803*; S. pn, *Streptococcus pneumoniae*; S. py, *Streptococcus pyogenes*.

decrease is about 5000-fold for D185A, and 500-fold for R181M).

Concerning the nucleotide analogues ddCMP and AraCMP, the K_m values do not vary much when the two mutants and the wild-type enzyme are compared. k_{cat} drops for AraCMP. For ddCMP the low k_{cat} value of the wild-type enzyme does not change much in the mutants.

The S101A mutant: the collapse of dCMP phosphorylation is the first evidence of a catalytic role played by the insert

Asp185, which is essential for the activity of *E. coli* CMPK, is H-bonded to Ser101 in both CMP-CMPK and dCMP-CMPK (molecule A) structures (Figure 2). Moreover, as indicated in Figure 4, Ser101 is the only strictly conserved amino acid residue in the homologous inserts from the 27 different bacterial species identified by BLAST.[10] We substituted this residue with an alanine and examined the catalytic properties of the modified protein. The resulting variant has a t_m close to that of the wild-type enzyme.

The kinetic parameters of this variant differ from those of D185A and R181M mutants. Thus, the S101A variant phosphorylates CMP much like the wild-type enzyme, whereas the three other substrates show significantly decreased reaction rates. The contrast is especially apparent when CMP and dCMP are compared: k_{cat}/K_m for CMP is four times lower than with the wild-type enzyme, whereas it is 200 times lower for dCMP. As was the case for the variant D185A, the decrease is mainly due to a decrease in k_{cat}.

Therefore, Ser101 is involved in dCMP phosphorylation, which is a special feature of bacterial CMP kinases. The effect of the S101A substitution reported here is the first published evidence of an active contribution of the NMPb-insert from bacterial CMP kinases in catalysis.

Discussion

Contribution of the pentose in induced-fit movements of the enzyme

In a previous paper we discussed the domain movements induced by CDP binding to *E. coli* CMPK.[6] Structure comparison of the free and CDP-bound enzymes revealed a movement of the helical part of the NMPb-insert, in particular helix α5, towards the ligand (r.m.s.d. due to CDP binding is 3.17 Å for α5). The same comparison for the free and CMP-bound enzymes indicates a similar, but more pronounced closing movement (r.m.s.d. for α5 is 4.04 Å). If we also take into account molecules A and B of dCMP-CMPK, enzyme conformations from the most open to the most closed might be classified, according to the position of α5, as: free *E. coli* CMPK > dCMP (molecule B) > CDP > dCMP (molecule A) > CMP. A similar classification using the other hydroxyl jaw α8 is less obvious because its C-terminal part (Glu180-Arg188) is disordered in the absence of a bound NMP. Anyway, based on the N-terminal part from α8, which is ordered even in the free enzyme (Phe171-Lys179), with dCMP the molecule B occupies an intermediate position between the open structure of the free enzyme and the closed one induced by CMP, as is the case for the α5-based classification.

Around 40 different structures are known so far for various NMP kinases, and much attention was devoted to the interactions with the phosphate or the base moieties of their NMP substrates. The effect of these interactions on the movements of the NMP binding and LID domains of the enzymes was thoroughly investigated.[11–13] However, the lack of structural data concerning a model enzyme able to phosphorylate various NMPs differing in their sugar moiety hindered discussion about the pentose contribution. The structures described here show that in the case of *E. coli* CMPK the so-called induced-fit movements seen in the presence of ligands are not only due to the base or phosphate moieties of the nucleotides, but also for a large part to the sugar moiety.

Importance of the 3′-OH from the pentose in catalysis

Kinetic parameters indicate that as opposed to dCMP, ddCMP is a very poor substrate. This emphasises the importance of the 3′-OH from the pentose, which is the only chemical difference between these two deoxynucleotides. The structures with the natural substrates CMP and dCMP show two polar interactions with that hydroxyl group, which have no equivalent in other NMP kinases of known structure.

The first residue that is H-bonded to the 3′-OH is Arg181 of α8. In contrast, neither AraCMP nor ddCMP is H-bonded to Arg181, in accordance with a poor K_m. This arginine is highly conserved in all groups of NMP kinases, and it has been identified as a residue involved in catalysis. Indeed, in the structures of UMP/CMP kinase from *Dictyostelium discoideum*, an H-bond connecting the equivalent Arg148 to the 3′-OH of CMP was observed in the ''transition state-like'' complex with CMP, AlF_3 and ADP (pdb code 3ukd), but not in the complex with CMP and ADP (pdb code 2ukd), suggesting that this arginine stabilises the transition state.[8] Except for the complexes with a transition state analogue, the direct hydrogen bond we observe between Arg181 from *E. coli* CMPK and the phosphate acceptor is unprecedented among NMP kinases. As in the phosphorylation reaction CMP-CMPK represents a step preceding the [CMP/γ transferred phosphoryl group/ADP] transition state, this H bond between Arg181 and the 3′-OH is probably productive for catalysis. In the AraCMP-CMPK structure, no hydrogen bond connects Arg181 to the 3′-OH. However, in the

case of the R181M variant the k_{cat} for AraCMP is decreased by a factor of 40. This could be explained by a contribution of Arg181 to the transition state stabilisation.

The second residue that is H-bonded to the 3′-OH of CMP and dCMP is Asp185, again a residue of α8. This aspartate also makes a polar interaction with the 2′-OH from CMP, AraCMP and CDP. The bidentate interaction observed for CMP between the side-chain of a unique residue and both hydroxyls from the ribose has never been described for NMP kinases. Moreover, among known structures of NMP kinases complexed solely with the phosphate acceptor, a direct protein interaction with the 3′-OH of the sugar only occurs with thymidylate kinases like TMP kinase of yeast (pdb coordinates 1tmk[14]), that phosphorylate 2′-deoxynucleotide substrates: the enzyme residue involved is also an aspartate, but one from the phosphate binding loop, that is an N-terminal domain unrelated to α8 in sequence alignments. The D185A variant shows a decrease of k_{cat} for the natural substrates CMP or dCMP. This is also the case for the AraCMP analogue, a result that emphasizes the importance of the H-bond between the 2′-OH and Asp185 for this nucleotide. In all three cases the decrease is even stronger than that observed for R181M. A decrease of k_{cat} when Asp185 is mutated was unexpected from the sequence of the *E. coli* enzyme, as this aspartate has no equivalent in other groups of NMP kinases. These results suggest that Asp185, conserved only in bacterial CMP kinases, also participates in catalysis, possibly through transition-state stabilisation. Such a stabilisation should involve the interaction of Asp185 with the 3′-OH (see CMP and dCMP(A) in Figure 2), or for AraCMP with the 2′-OH (see Figure 3) of the sugar. We propose that this favours a C3′-*endo* conformation, allowing the 3′-OH to interact with Arg181, a residue which can stabilise the α phosphate. An arginine-phosphate H-bond is observed in both crystal structures with CMP and dCMP (molecule A). With AraCMP, we suggest that Arg181 interacts with the α phosphate only in the presence of the transition state.

On the whole, the importance for catalysis of the two residues that are H-bonded to the 3′-OH, Arg181 and Asp185, makes it clear why the catalytic activity of the wild-type enzyme for ddCMP, that does not possess a 3′-OH, is so low.

Understanding the specificity for 2′-deoxynucleotide acceptors: the two conformations observed in crystals with dCMP

With CMP, the two molecules of the crystal asymmetric unit are essentially identical. With dCMP, molecule A has a conformation similar to that with CMP. Compared to CMP, or to dCMP in molecule A, the dCMP in molecule B: (i) penetrates less deeply in the binding cleft, (ii) is less tightly enclosed by the jaws α5 and α8, (iii) lacks the direct polar interactions with Arg188, and the indirect one (through Asp185) with Ser101, these three residues being highly conserved in bacterial CMP kinases, (iv) does not make any direct H-bond between the 3′-OH from its pentose and Arg181 or Asp185 of the enzyme, and (v) has a different sugar conformation (C2′-*endo* instead of C3′-*endo*), so that it has a very different orientation of its 3′-OH. We can hypothesise that the two alternative binding modes seen in molecule A (tight binding) and B (loose binding) coexist in solution. In that case, the conformer of molecule B, in which the 3′-OH is not oriented as in CMP-CMPK, and therefore does not bind the catalytically important residues Asp185 and Arg181, would be a dead-end complex non-productive in catalysis. This hypothesis is in accordance with the observed kinetic parameters: the *E. coli* CMPK activity reflected by k_{cat} remains identical for CMP and dCMP, whereas the K_m of the enzyme for dCMP is about three times as high as that for CMP, indicating a weaker affinity of the deoxynucleotide (Table 2).

The role of Ser101 in dCMP phosphorylation

Bacterial CMP kinases phosphorylate dCMP nearly as well as CMP. This is in contrast to eukaryotic UMP/CMP kinases: for instance, the k_{cat}/K_m ratio for dCMP of UMP/CMP kinase from *D. discoideum* is less than 1% of that for CMP. Similarly, its k_{cat}/K_m for dUMP is around 1% of that for UMP (A.M.G., unpublished data). Site-directed mutagenesis of Ser101 resulted in a variant with a decreased k_{cat} for dCMP. Thus this serine is important for deoxynucleotide catalysis, probably through binding of Asp185. This aspartate binds both the CMP and dCMP substrates. Moreover, crystal structures indicate that it positions these substrates in a productive way for catalysis: all attempts to get crystal complexes of *E. coli* CMPK with the phosphate donor site occupied (either with ADP, or ATP analogues, or with transition state analogues with ADP and aluminium fluoride) failed; nevertheless, in the structures presented here the phosphate from CMP or dCMP points towards a sulphate. In crystal structures of other NMP kinases also crystallised with ammonium sulphate, such a sulphate occupies a position corresponding to that of the β phosphate from the donor ATP.[15] The favourable positioning of the bound substrates is also indicated by the superposition of CMP-CMPK or dCMP-CMPK with structures of other known NMP kinases in which both phosphate donor and acceptor sites are occupied. Thus, these *E. coli* CMPK modelled structures can be considered as structures of the enzyme with both the NMP phosphate acceptor and an equivalent of the β phosphate from ATP.

Why does the substitution of Ser101 only slow down the phosphorylation of dCMP, and not that of CMP? The crystal structures help to understand this effect (Figure 2, left): with dCMP (molecule A), Ser101 interacts with the carboxylate of Asp185 through an oxygen (labelled OD1 on the Figure)

which is not held in position by any other H-bond, as in the CMP complex. We suggest that suppressing this interaction by substituting Ser101 with an alanine results in a more mobile side-chain of Asp185 when the variant binds dCMP. Moreover, the partial negative charge of OD1 is no more compensated through hydrogen bonding. In contrast, the effect of the substitution is minor with CMP as the 2′-OH from the ribose is still H-bonded to OD1. As Ser101 is conserved in all NMP bind inserts of bacterial CMP kinases, its role in dCMP phosphorylation is probably a general feature of these enzymes.

Possible implications for the design of antibiotics

Despite the fact that CMPK is not essential for *E. coli*, a Gram negative bacterium, mutants devoid of CMP kinase activity have a decreased DNA synthesis rate and are cold-sensitive.[3] However, CMP kinase appears essential for the Gram positive bacterium *Bacillus subtilis*.[16] This raises the possibility to design new antibiotics against CMP kinases from bacteria involved in human pathologies. The four crystal structures presented here give a detailed image of the interactions between the *E. coli* enzyme and phosphate acceptor substrates differing in their sugar moiety. Together with site-directed mutagenesis experiments, these structures also show the involvement of the NMPb-insert typical of bacterial CMP kinases in dCMP phosphorylation, that is a main role of *E. coli* CMPK. This gives a wealth of information about enzyme zones specific to bacterial CMP kinases that could be targeted by inhibitors. Such inhibitors would avoid possible side effects related to nucleoside monophosphate kinases of the patients.

Materials and Methods

Chemicals

CMP, dCMP and AraCMP were purchased from Sigma. ddCMP was kindly provided by Dr H.G. Ihlenfeldt from Roche Diagnostics, Penzberg (Germany).

Site-directed mutagenesis

Replacement by site-directed mutagenesis of Arg181 with a methionine and of Ser101 and Asp185 with an alanine was performed on the single-stranded form of plasmid pHSP210[2] in the *ung dut E. coli* strain RZ1032[17] in the presence of the helper phage M13KO7[18] and using respectively the following oligonucleotides: 5′-GTTACGGTCGCGGTCGTC**CAT**TTCTTTGATCTC-3′, 5′-GAATGCCGCGACTTGTGCCGC**TGC**ATTCGCCACTTCCTG-3′ and 5′-TCGGTTACG**AGC**GCGGTCGTCGCG-3′ (modified codons are in boldface). For overexpression, plasmids pLA221, pLA224 and pHL90-5a (harbouring the respective mutated *cmk* genes) were introduced into the *E. coli* strain BL21(DE3)/pDIA17.[19]

Purification of the enzymes and activity assays

The overproduced wild-type and modified variants of *E. coli* CMPK were purified as described previously,[2] and checked by mass spectrometry. Enzyme activity was determined at 30 °C using a spectrophotometric assay.[20] For each mutation, both a single-point mutant and its histidine-tagged (His-tagged) equivalent were produced and purified. The His-tagged variants were purified by Nickel-nitriloacetic acid affinity chromatography[21] using the QIAexpress System. Their kinetic values were essentially the same as those of the non-tagged mutants, indicating that the His-tag did not interfere with catalysis. As the assays with His-tagged variants were not biased by the potential presence of wild-type copurified enzyme, their results are used in Table 2.

Crystallisation

Crystals were grown at 20 °C by the vapour diffusion method, using a 6 μl hanging drop with 0.4 M ammonium sulphate in a 50 mM Tris-HCl buffer (pH 7.4). The wild-type enzyme used for crystallisation was not histidine-tagged. All crystals were obtained by cocrystallization of *E. coli* CMPK (5 mg/ml except for CMP-CMPK, in which case it was 10 mg/ml) with the NMP (20 mM for CMP, 10 mM for dCMP and ddCMP, 40 mM for AraCMP). Concentrations of ammonium sulphate in the pit were 1.7 M with CMP and AraCMP, 1.1 M for dCMP, and 1.3 M for ddCMP. Crystals were allowed to grow for several weeks before data collection. Their typical size was around 0.2 mm × 0.2 mm × 0.2 mm, except with dCMP (0.1 mm × 0.05 mm × 0.05 mm). All crystal systems are orthorhombic, space group $P2_12_12_1$, with two molecules per asymmetric unit.

Data collection

All crystals were frozen at 100 K before data collection, using paraffin oil (Hampton Research) as cryoprotectant. Data were collected on the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, beamline BM-30A (FIP) for CMP-CMPK and dCMP-CMPK crystals, beamline ID14-2 for ddCMP-CMPK crystals, and on the synchrotron of LURE (Orsay, France), beamline DW32, for AraCMP-CMPK crystals. All data were processed and scaled with the DENZO and SCALEPACK programs.[22] Information for all useful data collection and processing is given in the upper half of Table 1.

Phasing

The structures of CMP-, dCMP-, ddCMP-, and AraCMP-CMPK complexes were solved by molecular replacement at 3.5 Å resolution with AMoRe,[23] using only the protein part of the CDP-CMPK model. In all cases a clear density immediately appeared for the NMP.

Model building and structure refinement

Models were built using the O program.[24] CNS[25] version 1.0 was used for refinement, which was monitored using a free *R* factor.[26] With the nucleotide analogues pseudo-merohedral twinning problems occurred: for ddCMP-CMPK, we collected data for several crystals until obtaining an untwinned one; for AraCMP-CMPK, all data sets we collected were twinned with very close *a*

and *b* unit cell parameters, mimicking a tetragonal crystal system. Thus, refinement could only be achieved using SHELX 97-2[27] on the first steps, CNS being used for the final steps with the reflection file corrected by SHELX.

In all cases, simulated annealing was initially used, as well as non-crystallographic constraints. Then the two molecules of the asymmetric unit were refined as different models, using individual *B* factors. In the final steps we placed water molecules in residual density above 2.5 standard deviations. Solvent molecules with *B* factors refining higher than 60 Å^2 were discarded. For each model, side-chains from a few residues (zero to six per molecule) had no clear density and were modelled as alanines. A summary of the refinement results together with model statistics is given in the lower half of Table 1.

Superposition of different structures

Models were superposed with the procedure implemented in O, and the relevant root mean square deviations for C^{α} atoms were calculated using CNS. As residues Arg70 to Val80 have a weak density in all structures, the corresponding region was omitted in r.m.s.d. calculations.

Accession numbers

The pdb codes of the *E. coli* CMPK complexes are 1kdo for the CMP-CMPK structure, 1kdp for dCMP-CMPK, 1kdr for AraCMP-CMPK, and 1kdt for ddCMP-CMPK.

Acknowledgements

We thank Richard Kahn, Michel Roth, Jean-Luc Ferrer (beamline BM-30A, FIP), and Andreas Bracher (beamline ID14-2) from the ESRF for attentive assistance. We are also grateful to Thierry Prangé and Roger Fourme (LURE) for making the DW32 station available to us. We thank Joël Janin for carefully reading the manuscript. This work was supported by grants from the Centre National de la Recherche Scientifique (UPR 9063 and URA 2185), the Institut National de la Recherche Agronomique (UMR 206), and the Institut Pasteur.

References

1. Noda, L. (1973). Adenylate kinase. In *The Enzymes* (Boyer, P. D., ed.), 3rd edit., vol. 8, pp. 279-305, Academic Press, New York.
2. Bucurenci, N., Sakamoto, H., Briozzo, P., Palibroda, N., Serina, L., Sarfati, R. S. *et al*. (1996). CMP kinase from *Escherichia coli* is structurally related to other nucleoside monophosphate kinases. *J. Biol. Chem.* **271**, 2856-2862.
3. Fricke, J., Neuhard, J., Kelln, R. A. & Pedersen, S. (1995). The *cmk* gene encoding cytidine monophosphate kinase is located in the *rspA* operon and is required for normal replication rate in *Escherichia coli*. *J. Bacteriol.* **177**, 517-523.
4. Vonrhein, C., Schlauderer, G. J. & Schulz, G. E. (1995). Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. *Structure,* **3**, 483-490.
5. Yan, H. & Tsai, M. D. (1999). Nucleoside monophosphate kinases: structure, mechanism, and substrate specificity. In *Advances in Enzymology and Related Areas of Molecular Biology* (Purich, D. L., ed.), vol. 73A, pp. 103-134, Wiley, New York.
6. Briozzo, P., Golinelli-Pimpaneau, B., Gilles, A. M., Gaucher, J. F., Burlacu-Miron, S. & Sakamoto, H., *et al*. (1998). Structures of *Escherichia coli* CMP kinase alone and in complex with CDP: a new fold of the nucleoside monophosphate binding domain and insights into cytosine nucleotide specificity. *Structure,* **6**, 1517-1527.
7. Egner, U., Tomasselli, A. G. & Schulz, G. E. (1987). Structure of the complex of yeast adenylate kinase with the inhibitor P^1,P^5-di(adenosine-5′-) pentaphosphate at 2.6 Å resolution. *J. Mol. Biol.* **195**, 649-658.
8. Schlichting, I. & Reinstein, J. (1997). Structures of active conformations of UMP kinase from *Dictyostelium discoideum* suggest phosphoryl transfer is associative. *Biochemistry,* **36**, 9290-9296.
9. Saenger, W. (1984). Structures and conformational properties of bases, furanose sugars, and phosphate groups. In *Principles of Nucleic Acid Structure*, pp. 55-104, Springer-Verlag, New York.
10. Altshul, S. F., Gish, W., Myers, E. W. & Lipmann, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.
11. Müller, C. W., Schlauderer, G. J., Reinstein, J. & Schulz, G. E. (1996). Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure,* **4**, 147-156.
12. Matte, A., Tarie, L. W. & Delbaere, T. J. (1998). How do kinases transfer phosphoryl groups? *Structure,* **6**, 413-419.
13. Ostermann, N., Schlichting, I., Brundiers, R., Konrad, M., Reinstein, J., Veit, T. *et al*. (2000). Insights into the phosphoryl transfer mechanism of human thymidylate kinase gained from crystal structures of enzyme complexes along the reaction coordinate. *Structure,* **8**, 629-642.
14. Lavie, A., Vetter, I. R., Konrad, M., Goody, R. S., Reinstein, J. & Schlichting, I. (1997). Structure of thymidylate kinase reveals the cause behind the limiting step in AZT activation. *Nature Struct. Biol.* **4**, 601-604.
15. 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.
16. Sorokin, A., Serror, P., Pujic, P., Azevedo, V. & Ehrlich, S. D. (1995). The *Bacillus subtilis* chromosome region encoding homologues of the *Escherichia coli mssA* and *rpsA* gene products. *Microbiology,* **141**, 311-319.
17. Kunkel, T. A. (1985). Rapid and efficient site-directed mutagenesis without phenotypic selection. *Proc. Natl Acad. Sci. USA,* **82**, 488-492.
18. Dotto, G. P. & Zinder, N. D. (1984). Reduction of the minimal sequence for initiation of DNA synthesis by qualitative or quantitative changes of an initiator protein. *Nature,* **311**, 279-280.
19. Munier, H., Gilles, A. M., Glaser, P., Krin, E., Danchin, A., Sarfati, R. & Bârzu, O. (1991). Isolation and characterisation of catalytic and calmodulin-binding domains of *Bordetella pertussis* adenylate cyclase. *Eur. J. Biochem.* **196**, 469-474.
20. Blondin, C., Serina, L., Wiesmüller, L., Gilles, A. M. & Bârzu, O. (1994). Improved spectrophotometric

assay of nucleoside monophosphate kinase activity using the pyruvate kinase/lactate dehydrogenase coupling system. *Anal. Biochem.* **220**, 219-221.
21. Crowe, J., Döbeli, H., Gentz, R., Hochulu, E., Stüber, D. & Henco, K. (1994). 6xHis-Ni-NTA chromatography as a superior technique in recombinant expression protein expression/purification. In *Methods in Molecular Biology* (Harwood, A. J., ed.), vol. 31, pp. 371-387, Humana Press, Inc., Totawa.
22. Otwinowski, Z. & Minor, W. (1993). *DENZO, a Film Processing Program for Macromolecular Crystallography*, Yale University Press, New Haven, CT.
23. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A,* **50**, 157-163.
24. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjelgaard, M. (1991). Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A,* **47**, 110-119.
25. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al*. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta crystallog. sect. D,* **54**, 905-921.
26. Brünger, A. T. (1992). Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature,* **355**, 472-475.
27. Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: high resolution refinement. *Methods Enzymol.* **277**, 319-343.
28. Nichols, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association-insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281-296.
29. Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.
30. Merritt, E. A. & Bacon, D. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524.
31. Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. *J. Mol. Graph.* **15**, 132-134.
32. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Edited by K. Morikawa

(Received 5 September 2001; received in revised form 19 November 2001; accepted 20 November 2001)