The EMBO Journal vol.3 no.7 pp.1587–1593, 1984

Molecular cloning, sequencing, and expression of the *crr* gene: the structural gene for III^{Glc} of the bacterial PEP:glucose phosphotransferase system

Stephen O. Nelson[1], Anja R.J. Schuitema[1], Rob Benne[2], Lex H.T. van der Ploeg[3], Johan S.Plijter[1], Frank Aan[1] and Pieter W. Postma[1]

[1]Section of Biomembranes, Laboratory of Biochemistry, BCP Jansen Institute, University of Amsterdam, PO Box 20151, 1000 HD Amsterdam, [2]Section for Medical Enzymology and Molecular Biology, Laboratory of Biochemistry, Jan Swammerdam Institute, PO Box 60000, 1005 GA Amsterdam, and [3]Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Communicated by L.A.Grivell

The phosphoenolpyruvate:glucose phosphotransferase system (PTS) of *Salmonella typhimurium* is involved both in glucose transport and in the regulation and synthesis of adenylate cyclase and several transport systems. The *crr* gene has been implicated in this regulating mechanism. A 9.6-kb segment of the *S. typhimurium* chromosome containing the *crr* gene was cloned in pAT153. The cloned fragment also complemented *cysA* mutations but did not contain a functional *pts* operon which is closely linked to the *crr* gene and codes for two enzymes of the PTS. Although *cysA* and *crr* have been reported to be located on opposite sides of *ptsHI*, our results suggest that the correct gene order is *cysK-ptsHI-crr-cysA*. Expression of *crr* plasmids in a maxicell system yielded two proteins which reacted with specific anti-serum against III^{Glc}. The apparent mol. wts. in SDS-polyacrylamide gels were 20 000 and 21 000, the former corresponding to the major band of purified III^{Glc}. Both forms were also observed in bacterial extracts and purified III^{Glc}. The *crr* gene was localized on a 1-kb *Eco*RI-*Eco*RV fragment of the 9.6-kb insert and sequenced. It codes for a single protein (18 556 D) containing 169 amino acid residues and identified as III^{Glc}.

Key words: phosphotransferase system/*crr* gene/III^{Glc}/regulatory protein/*Salmonella typhimurium*

Introduction

The phosphoenolpyruvate:sugar phosphotransferase system (PTS) catalyses the transport and concomitant phosphorylation of a number of sugars in *Salmonella typhimurium* and *Escherichia coli*. In addition, the PTS is thought to be involved in the regulation of the synthesis and/or activity of a number of non-PTS transport systems and adenylate cylase (for review, see Postma and Roseman, 1976; Saier, 1977). A central role in this regulation is attributed to III^{Glc}, a protein which together with the membrane-bound II^{Glc} is also involved in the transport of glucose and methyl α-glucoside via the PTS. Factor III^{Glc} can exist in two forms: phosphorylated and non-phosphorylated. Phosphorylation of III^{Glc} by phosphoenolpyruvate (PEP) is catalysed by two proteins of the PTS, enzyme I and HPr. In its non-phosphorylated form, III^{Glc} is an inhibitor of a number of non-PTS uptake systems, including those for maltose, melibiose, lactose and glycerol. Phosphorylated III^{Glc}, $P{\sim}III^{Glc}$, is supposed to activate adenylate cyclase. Thus when the cells have a low steady-state level of $P{\sim}III^{Glc}$ (for instance when PTS sugars are present or the activity of enzyme I/HPr is limited due to mutation), the intracellular cAMP level will be low and the transport of substrates (inducers) will be inhibited. This mechanism may explain the preference of *S. typhimurium* and *E. coli* for glucose (and other PTS sugars) compared to certain non-PTS compounds. Two classes of carbon sources are known to be affected by an alteration of the level of III^{Glc} phosphorylation: (i) growth on compounds such as maltose, melibiose, lactose and glycerol (class I compounds, Postma, 1982) is dependent on cAMP, which is required for transcription initiation of the respective operons specifying the transport systems. In addition, the uptake of these substrates is inhibited by III^{Glc}; (ii) compounds such as citrate, succinate and xylose (class II compounds) also require cAMP for transcription of the respective operons but their transport systems are not sensitive to inhibition by non-phosphorylated III^{Glc}. According to this hypothesis, III^{Glc} plays a central role in the regulation of cell metabolism.

Previously we have described the purification of III^{Glc} from *S. typhimurium* (Scholte *et al.*, 1981), the production of antibodies against III^{Glc} and the characterisation of a number of *crr* mutants which lack soluble III^{Glc} or contain a modified III^{Glc} (Scholte *et al.*, 1982). From these studies it appeared that the *crr* gene is most likely the structural gene for III^{Glc} and that soluble III^{Glc} may occur in several forms. It has been shown that purified III^{Glc} interacts with the purified lactose carrier in a stoichiometric fashion. Upon binding of III^{Glc} to the lactose carrier, the carrier becomes inactive (Nelson *et al.*, 1983). Recently, III^{Glc} has also been shown to inhibit glycerol kinase (Postma *et al.*, 1984). As a first step in a more detailed study of the interaction between III^{Glc} and its various target proteins, we report here the cloning of the *crr* gene, its expression in maxicells, and the determination of its base sequence.

Results

Selection of a λgt4 hybrid phage that carries the crr gene

A *crr ptsM* double mutant of *E. coli* was used to select a λgt4 hybrid phage that carries the *Salmonella crr* gene. This *E. coli* strain, LM1 (Lengeler *et al.*, 1981), does not grow on succinate due to the *crr* mutation (Scholte and Postma, 1980) and does not grow on glucose because both PTS glucose transport systems, II^{Man} (*ptsM*) and III^{Glc}/II^{Glc} (*crr/ptsG*) are defective. Thus complementation of *crr ptsM* strains by a hybrid phage can yield two types of clones: (i) complementation of *crr* will result in Glucose$^+$ Succinate$^+$ cells; (ii) complementation of *ptsM* will yield Glucose$^+$ Succinate$^-$ cells. *crr* Revertants could be discriminated from real lysogens because λgt4 contains a temperature-sensitive repressor and at 42°C the cell is lysed (Panasenko *et al.*, 1979).

Using LM1, we obtained a number of Glucose$^+$ lysogens. The majority were also Succinate$^+$ and temperature-sensitive and represent most likely Crr$^+$ lysogens. Some were Glucose$^+$, Succinate$^-$ and temperature-sensitive and are pre-

© IRL Press Limited, Oxford, England.

Table I. Production of III^{Glc} in *S. typhimurium* and *E. coli* plasmid strains

Strain	Relevant genotype	$III^{Glc a}$		
		no plasmid	pBCP1	pBCP2
E. coli				
LR2-167	–	2	–	50
LM1	*crr*	0	6	26
HB101	–	3	4	22
S. typhimurium				
SB3507	–	3.5	7	55
ST422	*recA*	4	9	40
PP1461	*recA*	4	–	32
PP692	*cya*	3	–	51
PP642	*ptsHI*△	2	–	77
PP1407	(*ptsI-crr*)△	0	–	80
PP994	*crr*	0	–	45

[a]Cells were grown to the stationary phase in LB medium containing 50 μg ampicillin/ml if a plasmid was present. III^{Glc} was determined as described in Materials and methods and expressed as μg III^{Glc} per mg protein. – not determined.

sumably PtsM$^+$ lysogens. Phages were isolated from a number of the Crr$^+$ lysogens and used to lysogenise LM1 again. One such lysogen, LM1-C1, was studied further.

While LM1 lacks III^{Glc} completely (Table I), LM1-C1 grown at 31°C has about wild-type levels. Upon heat induction III^{Glc} was overproduced in the lysogens containing the *crr*$^+$ gene at least 8–10-fold compared to the *E. coli crr*$^+$ strain LR2-167 (data not shown). The level of III^{Glc} increased for at least 40 min after heat induction and cell lysis started after ~50 min. None of the lysogens tested contained the complete *pts* operon which is tightly linked to *crr*. We could not detect any significant increase of enzyme I levels (as measured with antibodies against enzyme I) above the level already present in LM1 (see below). We have also been unable to complement an *E. coli ptsI* strain directly with the same λgt4 bank.

Amplification of the crr gene

The λgt4 clone contained a 9.6-kb *Eco*RI fragment of *Salmonella* DNA (data not shown). This fragment was inserted into the multicopy plasmid pAT153 and LM1 was transformed with the recombinant plasmid by selecting for tetracycline resistance and growth on glucose and succinate. Two different types of Glucose$^+$, Succinate$^+$ transformants were isolated which differed in the level of III^{Glc} produced. While the *crr* strain, LM1, contains no III^{Glc}, one transformant (LM1/pBCP1) had about twice the wild-type levels and the other (LM1/pBCP2) had 13 times the wild-type levels of III^{Glc}. This difference in production of III^{Glc} was seen in other *E. coli* and *Salmonella* strains containing these plasmids (Table I).

III^{Glc} synthesized by plasmid-containing cells is in an active form since the protein isolated from LM1/pBCP2 has the same specific activity in a phosphorylation assay *in vitro* as III^{Glc} from a wild-type *Salmonella* strain (Scholte *et al.*, 1981). A similar specific activity has been reported by Meadow *et al.* (1982b). Overproduction of III^{Glc} does not seem to be harmful to the cell. A *S. typhimurium recA* strain (PP1461) containing pBCP2 grows normally on all carbon sources tested, including PTS sugars, class I compounds (maltose, melibiose, and glycerol) and class II compounds (xylose, citrate and succinate).

Complementation of S. typhimurium crr and pts mutants by pBCP2

Since pBCP2 was isolated originally by complementation of an *E. coli crr* strain, it was important to show that *S. typhimurium crr* mutations are also complemented by the plasmid. We have introduced pBCP2 by transformation or by P22 transduction (using PP1461/pBCP2 as donor) in various *S. typhimurium crr* mutants, including *crr-303*, *crr-306* and *crr-307*::Tn*10* (Scholte *et al.*, 1982). In all cases tested, the strains obtained had a Crr$^+$ phenotype, i.e., had regained growth on succinate, citrate and xylose and produced III^{Glc}. By selecting for growth on mannitol, which requires an intact *pts* operon, we found that pBCP2 does not complement a *ptsHI* or *ptsI-crr* deletion strain. Thus, although *pts* and *crr* are closely linked in *S. typhimurium* [>95% co-transduction between *ptsI17* (SB1476) and *crr-307*::Tn*10* (PP994)], the plasmid does not contain a functional *pts* operon.

Surprisingly, both plasmids pBCP1 and pBCP2 did complement *cysA* mutations of both *E. coli* and *S. typhimurium*. Since *cysA* has been reported to be located in *S. typhimurium* on the opposite side of *pts* from *crr* (Cordaro and Roseman, 1972), from the presence of *cysA* and *crr* on one chromosomal fragment we also expected the presence of the *pts* genes. Two other results suggest that the correct order is *cysK-ptsHI-crr-cysA* rather than *cysA-cysK-ptsHI-crr* as published (Cordaro and Roseman, 1972). First, we were unable to transduce a *ptsI* mutation into a *cysA*::Tn*10 crr* strain by selecting for *cys*$^+$ and retaining *crr* at the same time. This suggests that *crr* is located between *cysA* and *ptsI*. On the contrary, we can transduce *crr* into a *cysA*::Tn*10 ptsI* strain by selecting for *cys*$^+$ and retaining *ptsI*. This would be unlikely if the published order were correct. Secondly, we have constructed *ptsP*::Tn*10* insertions that have ~30% of the normal levels of enzyme I and HPr. Excision of Tn*10* using selection on fusaric acid (Bochner *et al.*, 1980) yielded strains that became *cysA crr ptsI* lacking all III^{Glc} and enzyme I. This result is difficult to explain with the published order but fits the reverse order, which is similar to the one published for *E. coli* (Bachmann, 1983).

Identification of products synthesized by pBCP2

Although overproduction of III^{Glc} both by temperature-induced lysogens and pBCP2-containing cells indicates that the structural gene for III^{Glc} is contained in the cloned DNA fragment, more conclusive evidence was sought by investigating protein synthesis directed by plasmid DNA in the absence of chromosomal protein synthesis. To this end, the 'maxicell' system described by Sancar *et al.* (1981) was used. *E. coli* CSR603 (*recA uvrA phr*) was transformed with pBCP2. Figure 1 shows that after u.v. treatment no proteins are synthesized in CSR603. When pBR322 is present, the major labelled band visible corresponds to β-lactamase (cells containing the plasmids are grown in the presence of ampicillin). pBCP2 codes for a number of additional proteins, two of which have an apparent mol. wt. close to 21 000, the mol. wt. of III^{Glc} (Scholte *et al.*, 1981). The following experiments show that both these proteins are closely related forms of III^{Glc}.

Figure 2 shows crossed immunoelectrophoresis of an extract of permeabilized, ^{35}S-labelled CSR603/pBCP2 cells with anti-serum against III^{Glc}. A continuous precipitation line is obtained with two peaks. The Coomassie-stained pattern (Figure 2A) coincides with that obtained after autoradiogra-

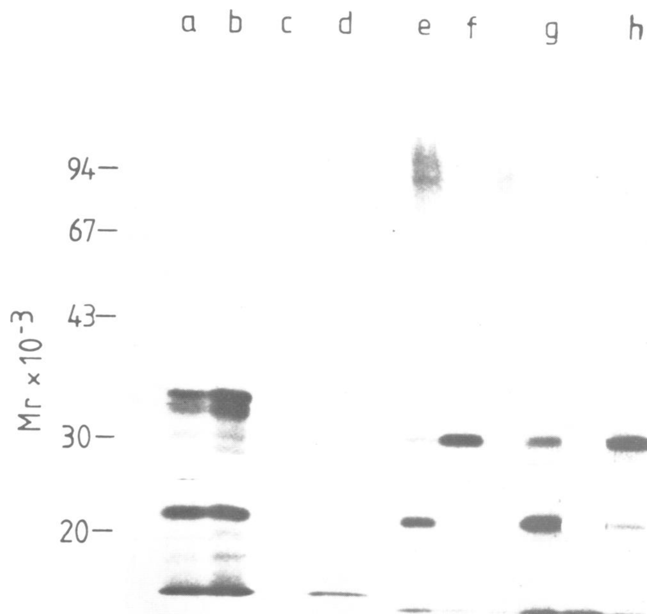

Fig. 1. Protein synthesis directed by pBCP2. CSR603 cells with or without plasmid were u.v. treated and labeled with [^{35}S]methionine. After SDS-polyacrylamide gel electrophoresis, proteins were visualised by autoradiography; **a,b,** CSR603/pBCP2; **c,** CSR603; **d,** CSR603/pBR322; **e,** CSR603/pBCP7; **f,** CSR603/pAT153; **g,** CSR603/pBCP18; **h,** CSR603/pBCP20.

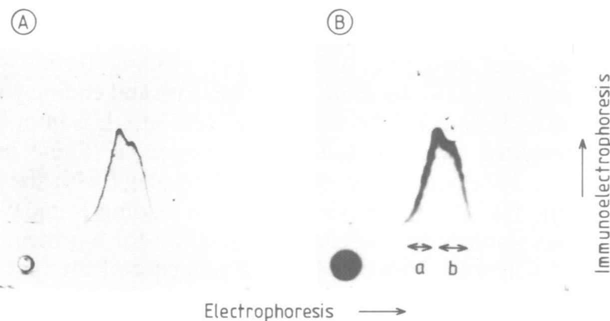

Fig. 2. Crossed immunoelectrophoresis of CSR603/pBCP2 cells using anti-serum against III^{Glc}. After treatment of [^{35}S]methionine-labeled cells with toluene/chloroform/Triton X-100, proteins were electrophoresed for 2 h (5 V/cm) in the first dimension and for 18 h (5 V/cm) against anti-III^{Glc} anti-serum in the second dimension. About 10^5 c.p.m. were applied in the slot. **(A)** Coomassie staining; **(B)** autoradiography.

phy (Figure 2B). The former represents the III^{Glc} synthesized before u.v. treatment and present in large amounts, whereas the latter represents III^{Glc} synthesis directed by the plasmid during a 1 h period in the absence of chromosomal protein synthesis.

Thus, pBCP2 directs the synthesis of III^{Glc}. The continuous precipitation line shown in Figure 2 indicates that the two proteins are very closely related. To characterise both peaks, the precipitated III^{Glc} anti-serum complex was cut out of the agarose gel and re-electrophoresed in SDS-polyacrylamide. Figure 3A shows that the two peaks consist of proteins that have different apparent mobilities. The peak with the highest mobility in agarose (non-denaturing conditions, Figure 2B) represents a protein that has an apparent mol. wt. of ~21 000. The other peak contains this protein and a much

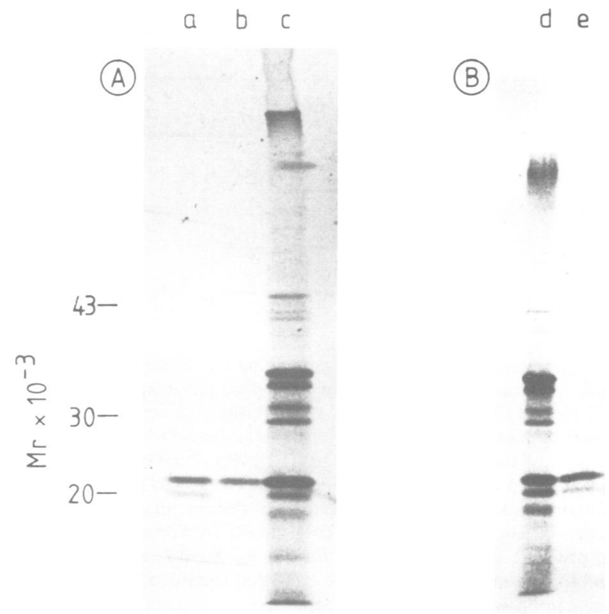

Fig. 3. Characterization of ^{35}S-labeled proteins with anti-serum against III^{Glc}. **(A)** After crossed immunoelectrophoresis of CSR603/pBCP2 cells described in the legend to Figure 2 two agarose fractions, indicated as (a) and (b) in Figure 2B, containing the precipitated complex were cut out. After re-electrophoresis in SDS-polyacrylamide gels, the labeled proteins were visualised by autoradiography. **a,** fraction a; **b,** fraction b; **c,** CSR603/pBCP2. **(B)** III^{Glc} was precipitated from [^{35}]methionine-labeled cells with anti-III^{Glc} and protein A-Sepharose and visualised by autoradiography. **d,** CSR603/pBCP2; **e,** immunoprecipitate.

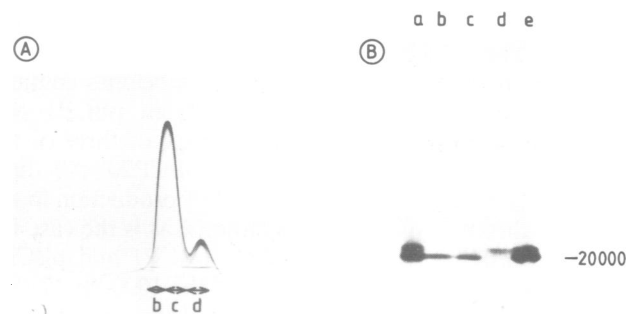

Fig. 4. Crossed immunoelectrophoresis and SDS-polyacrylamide gel electrophoresis of purified III^{Glc}. After electrophoresis of purified III^{Glc} (1 μg) against anti-serum, agarose fractions containing the precipitation lines, as shown in **(A)** were cut out and re-electrophoresed in SDS-polyacrylamide gel. III^{Glc} was visualised by protein blotting on nitrocellulose, labeling with anti-serum and [^{125}I]protein A and autoradiography. **(A)** crossed immunoelectrophoresis; **(B)** SDS-polyacrylamide gel electrophoresis; **a** and **e**, purified III^{Glc}; **b**–**d**, fractions b, c and d of Figure 4A, respectively.

more weakly labelled protein with an apparent mol. wt. of 20 000. Similar results were obtained by direct immunoprecipitation of the ^{35}S-labelled proteins using anti-serum against III^{Glc} and protein A-Sepharose beads (Figure 3B). Two bands were obtained after SDS-polyacrylamide gel electrophoresis.

Moreover, after crossed immunoelectrophoresis of both purified III^{Glc} and bacterial extract a double peak is found, the ratio of which may vary (Figure 4A, see also Scholte *et al.*, 1981, 1982). Analysis of both peaks from purified III^{Glc} by SDS-polyacrylamide gel electrophoresis, followed by protein blotting, treatment with anti-serum and [^{125}I]protein A

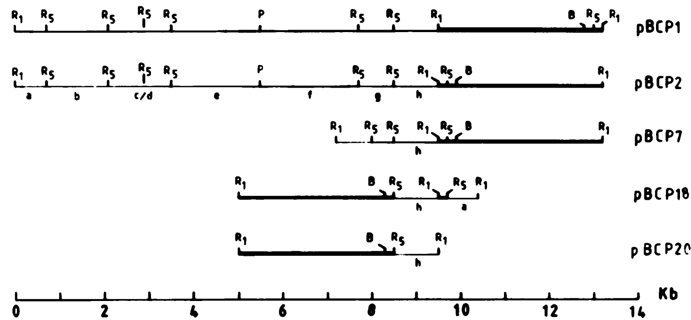

Fig. 5. Restriction maps of plasmids containing *crr*. Subcloning of the 9.6-kb *S. typhimurium* DNA insert of the hybrid phage λgt-4-C1 in pAT153 yielded pBCP1 and pBCP2. pBCP7 was obtained as a deletion (generated by partial *AluI* digestion) of material between the outermost *Eco*RV sites of the insert in pBCP2. pBCP18 was obtained by re-arrangement and deletion of the DNA of pBCP7. It was produced by digestion with *Eco*RV followed by ligation of the resulting mix of fragments. Deletion of the smaller *Eco*RI-*Eco*RI fragment of pBCP18 yielded pBCP20. Abbreviations: R_1, *Eco*RI; R_5, *Eco*RV; P, *Pvu*II; B, *Bam*HI. The numbers indicate the estimated lengths of the restriction fragments in bp with respect to standard DNA fragments.

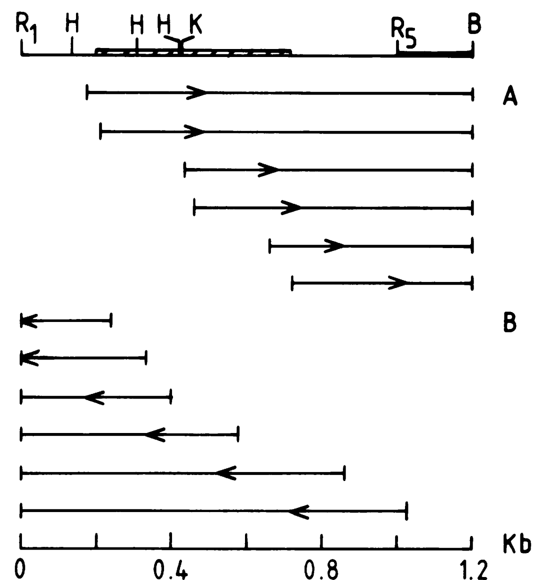

Fig. 6. Sequence strategy. The map was generated by restriction analysis of pBCP20 and includes the 1000-bp *Eco*RI-*Eco*RV *S. typhimurium* insert plus the adjacent *Eco*RV-*Bam*HI vector fragment. Phage M13 clones containing *Eco*RI-*Bam*HI fragment-derived clones in two orientations were constructed and submitted to sequence analysis as described. **(A)** clones derived from procedure A. **(B)** clones derived from procedure B. The arrows indicate the regions for which sequences were determined with the clones. Abbreviations: R_1, *Eco*RI; H, *Hind*II; K, *Kpn*I; B, *Bam*HI.

and autoradiography shows again two bands with slightly different mobilities (apparent mol. wts. 21 000 and 20 000), the lower mobility in agarose corresponding with the higher mobility in SDS-polyacrylamide (Figure 4B). Direct blotting of purified III^{Glc} yields the same results although the upper band is much weaker after Coomassie staining and also reacts more weakly with anti-serum.

Restriction map of the 9.6-kb insert and localisation of the crr gene

Restriction maps of pBCP1 and pBCP2 were determined (Figure 5). The 9.6-kb *S. typhimurium* insert has opposite orientations in the two plasmids. Several subclones containing the *crr* gene have been produced from pBCP1 and pBCP2. Figure 5 shows the restriction maps of three of the most interesting, pBCP7, pBCP18 and pBCP20. All three complement *crr* mutations and direct III^{Glc} production in the absence of chromosomal protein synthesis, as is the case for pBCP1 and pBCP2 (see below). Only pBCP1 and pBCP2 complement *cysA*. The *crr* is located in the 1000-bp *Eco*RI-*Eco*R5 fragment, h. This is the only insert fragment common to all five plasmids and is the only one present in pBCP20 (Figure 5). Although the three smaller plasmids are all derived from pBCP2, only pBCP7 and pBCP18 overproduce III^{Glc}. In both plasmids, a 185-bp *Eco*RI-*Eco*R5 vector fragment is situated before the *crr* fragment, h, as is the case with pBCP2.

The *bla* promoter of pAT153 (pAT153 is a derivative of pBR322) is located in this vector fragment. Transcription from this promoter is directed toward the *Eco*RI site (Brosius *et al.*, 1982) and therefore towards the *crr* insert. In pBCP20, which does not overproduce III^{Glc}, this *bla* promoter is deleted. In pBCP1, the *crr* gene is inserted in the other orientation (at ~10 kb distance from the second *bla* promoter). We conclude that overproduction could be due to transcription from the *bla* rather than the *crr* promoter. The low level of expression of *crr* from its own promoter is not easily explained because pAT153 is a multicopy plasmid that should give overproduction. Two possibilities are that the *crr* gene is autoregulated when expression occurs via its own promoter or that *crr* has been subcloned with a damaged promoter.

Nucleotide sequence analysis of the crr gene

The nucleotide sequence of the 1000-bp *Eco*RI-*Eco*RV fragment, h (Figure 5), was determined using a non-random M13 sequencing method (Poncz *et al.*, 1982; Benne *et al.*, 1983). The sequence strategy is shown in Figure 6 and is described in Materials and methods. A large open reading frame was found beginning 197 bp from the *Eco*RI site and ending with two stop codons at 712 bp from the *Eco*RI site. It is immediately preceded by a Shine-Dalgarno sequence (Shine and Dalgarno, 1974) exhibiting very high homology with the 3′ end of the 16s rRNA. Figure 7 shows this reading frame plus the region immediately preceding it. It codes for a protein of 169 amino acid residues of which the sequence of the first 29 (minus the methionine) is in perfect agreement with the partial amino-terminal sequence recently determined by Edman degradation of pure III^{Glc} from *S. typhimurium* (Meadow and Roseman, 1982).

Discussion

Regulation of cellular metabolism by the PEP:sugar phosphotransferase system in *S. typhimurium* and *E. coli* is a complex process. One of the PTS phosphoproteins, III^{Glc}, which is involved in transport and phosphorylation of glucose and methyl α-glucoside, has been proposed as the central regulatory protein (for reviews, see Postma and Roseman, 1976; Saier, 1977). According to this hypothesis, in its phosphorylated form III^{Glc} activates adenylate cyclase, whereas it inhibits certain non-PTS transport systems in its non-phosphorylated form.

Here we report a first analysis of a chromosomal fragment of *S. typhimurium* containing the *crr* locus. Fragments of the *Salmonella* chromosome (varying in length from 1 to 9.6-kb) complementing *crr* mutations were subcloned in pAT153. Both phenotypic characteristics of the *crr* mutation, namely absence of or lowered transport of glucose and methyl α-glucoside, and inability to grow on Krebs-cycle intermediates such as succinate, were restored by our *crr* plasmids. The

```
  1 GAATTCTCTATGAGCGCCATTTCTATCCCGCGCATTAAGAAGATTATCCGTAACACGAAC
 61 TTCGAAGAAGCGAAGGTGTTAGCAGAGCAGGCTCTTGCTCAACCGACAACGGACGAGTTA
121 ATGACGCTGGTTAACAAGTTCATTGAAGAAAAAACAAACTGCTAATCCACGAGACGCGGC
                                                            MET   GLY
181  C C A A T T T A C T G C T T A G G A G A A G A T C A T G G G
        LEU   PHE   ASP   LYS   LEU   LYS   SER   LEU   VAL   SER
211  T T T G T T C G A T A A A C T A A A A T C T C T G G T T T C
        ASP   ASP   LYS   LYS   ASP   THR   GLY   THR   ILE   GLU
241  T G A T G A T A A G A A A G A C A C C G G A A C T A T T G A
        ILE   VAL   ALA   PRO   LEU   SER   GLY   GLU   ILE   VAL
271  G A T T G T T G C C C C G C T C T C T G G C G A G A T C G T
        ASN   ILE   GLU   ASP   VAL   PRO   ASP   VAL   VAL   PHE
301  C A A C A T C G A A G A C G T G C C G G A T G T A G T T T T
        ALA   GLU   LYS   ILE   VAL   GLY   ASP   GLY   ILE   ALA
331  T G C T G A A A A A A T C G T T G G T G A T G G C A T C G C
        ILE   LYS   PRO   THR   GLY   ASN   LYS   MET   VAL   ALA
361  T A T C A A A C C A A C C G G T A A C A A A A T G G T C G C
        PRO   VAL   ASP   GLY   THR   ILE   GLY   LYS   ILE   PHE
391  C C C T G T T G A C G G T A C C A T C G G C A A A A T C T T
        GLU   THR   ASN   HIS   ALA   PHE   SER   ILE   GLU   SER
421  T G A A A C C A A C C A T G C G T T C T C T A T C G A A T C
        ASP   SER   GLY   ILE   GLU   LEU   PHE   VAL   HIS   PHE
451  C G A T A G C G G C A T T G A G C T G T T C G T T C A C T T
        GLY   ILE   ASP   THR   VAL   GLU   LEU   LYS   GLY   GLU
481  C G G T A T C G A C A C C G T T G A G C T G A A A G G C G A
        GLY   PHE   LYS   ARG   ILE   ALA   GLU   GLU   GLY   GLN
511  A G G C T T C A A G C G T A T T G C T G A A G A A G G T C A
        ARG   VAL   LYS   VAL   GLY   ASP   PRO   VAL   ILE   GLU
541  G C G C G T G A A A G T T G G C G A T C C G G T C A T C G A
        PHE   ASP   LEU   PRO   LEU   LEU   GLU   GLU   LYS   ALA
571  G T T C G A T C T G C C G T T G C T G G A A G A A A A A G C
        LYS   SER   THR   LEU   THR   PRO   VAL   VAL   ILE   SER
601  C A A G T C T A C C C T G A C T C C G G T T G T T A T C T C
        ASN   MET   ASP   GLU   ILE   LYS   GLU   LEU   ILE   LYS
631  C A A C A T G G A T G A A A T C A A A G A A C T G A T C A A
        LEU   SER   GLY   SER   VAL   THR   VAL   GLY   GLU   THR
661  A C T G T C C G G T A G C G T G A C C G T G G G T G A A A C
        PRO   VAL   ILE   ARG   ILE   LYS   LYS   ***
691  T C C G G T T A T C C G C A T T A A G A A G T A A T T C T T
721 GCCGCAGTGAAGA
```

Fig. 7. The nucleotide sequences of the *crr* region of the R_1-R_5 fragment. The sequence of the coding strand is shown for nucleotides 1 – 733 beginning by the *Eco*RI site. The sequence was assembled using the computer programs of Staden (1982). The Shine-Dalgarno (1974) consensus sequence is underlined and neighbouring bases identical to the corresponding bases in the 16S rRNA 3′ end are individually underlined. The stop codons are marked with stars.

results can be summarised as follows. (i) pBCP2 overproduces III^{Glc} 10- to 20-fold compared with wild-type cells. Overproduction depends somewhat on growth conditions and genetic background (Table I) and is the result of transcription of *crr* via the second *bla* promoter of the vector. The high level of (phosphorylated) III^{Glc} has no apparent harmful effects, since strains containing the plasmid grow well on PTS compounds and on both class I and class II non-PTS compounds. We have subcloned pBCP2 and obtained a plasmid (pBCP20) in which the chromosomal fragment is 1 kb and still complements *crr*. (ii) The five *crr* complementing plasmids described here direct the synthesis of III^{Glc} in the absence of chromosomal directed protein synthesis. In the case of pBCP20, this is the only protein product other than β-lactamase (Figure 1). (iii) *In vivo*, two proteins are synthesized which react with anti-serum against III^{Glc} and have slightly different mobilities in SDS-polyacrylamide gels. From crossed immunoelectrophoresis (Figure 2) it can be deduced that both proteins are closely related since one continuous precipitation line is obtained. Similar peaks have been observed in different preparations from various strains without plasmids, including crude bacterial extracts and purified III^{Glc}, although the proportion of both peaks may differ, as judged from the peak areas in crossed immunoelectrophoresis and the intensity after autoradiography. The protein with lower mobility in SDS-polyacrylamide gels moves faster during electrophoresis under non-denaturing conditions (Figure 4).

These results strengthen the conclusion (Scholte *et al.*, 1981, 1982; Meadow and Roseman, 1982) that III^{Glc} can exist in different forms. These forms might have different functions. Temperature determines (at least *in vitro*) whether III^{Glc} exists as a monomer or associates to dimers, trimers and hexamers. At low temperatures, the monomeric state dominates. Mutant III^{Glc} molecules have been isolated in which the association-dissociation equilibrium is displaced (Scholte *et al.*, 1982). In addition, III^{Glc} can exist in forms with different apparent mol. wt. Both forms are absent from *crr* deletion mutants and certain point mutants. Both forms are synthesized from pBCP2. Indeed, recent experiments (Nelson *et al.*, 1983) have shown that III^{Glc} isolated from overproducing cells and consisting mainly (>95%) of the 20 000 mol. wt. form is able to bind to the *E. coli* lactose carrier and to inhibit β-galactoside counter transport, efflux and active transport in liposomes reconstituted with the purified lactose carrier.

Meadow and Roseman (1982) also describe two forms of III^{Glc}, the lower mol. wt. form (III^{Glc}_{fast}) being inactive in *in vitro* methyl α-glucoside phosphorylation. They reported that their III^{Glc}_{fast} was an artifact produced by proteolysis during isolation. Our III^{Glc} is isolated in the presence of protease inhibitors (Scholte *et al.*, 1981); it consists predominantly of a lower mol. wt. form and is active. It cannot yet be decided whether both forms have different functions in transport and regulation. (iv) The structural gene for III^{Glc} has been identified. It is located on a 1000-bp *S. typhimurium* chromosomal fragment that directs the synthesis of only one protein product, III^{Glc}, and complements *crr* mutations. There is conclusive evidence that *crr* mutations can result not only in loss of III^{Glc} but also, in some cases, in the production of a changed III^{Glc} (Scholte *et al.*, 1982; Meadow *et al.*, 1982a). This led to the conclusion that the structural gene for III^{Glc} and *crr* must be one and the same. There remained little doubt, therefore, that the structural gene for III^{Glc} should be located on this fragment, and we found an open reading frame coding for a protein containing an amino-terminal sequence consistent with the incomplete sequence of pure III^{Glc} from *S. typhimurium* recently determined by Edman degradation (Meadow and Roseman, 1982). From the sequence we have calculated the mol. wt. to be 18556 if the initial methionine is included. The amino terminal analysis of Meadow and Roseman (1982) indicates that this residue may not be present in isolated III^{Glc}. The mol. wt. calculated here is in fairly good agreement with published values for the apparent mol. wt., which vary from 19000 to 21000 (Scholte *et al.*, 1981; Meadow and Roseman, 1982).

Recently Meadow *et al.* (1982b) identified the III^{Glc} produced by the cloned *crr* gene of *E. coli* by immunological methods whereas Britton *et al.* (1982) identified it on the basis of apparent mol. wt. only. To our knowledge the sequence of the *E. coli crr* gene has not yet been published.

We have compared the amino acid sequence of III^{Glc} with the sequence of three other proteins which we thought might

be evolutionarily related to III^{Glc}. No significant homology was found with II^{Mtl} (also a PTS protein which interacts with HPr and catalyses mannitol transport) or with calmodulin (an eucaryotic protein with multiple regulatory functions). A strong homology was found between the last nine residues of III^{Glc} (carboxy terminus: gly-glu-thr-pro-val-ile-arg-ile-lys-lys) and a block of eight residues near the carboxy terminus of the cAMP-binding protein, CRP (residues 182–189: glu-thr-val-gly-arg-ile-leu-lys (Cossart and Gicquel-Sanzey, 1982)). Both CRP and III^{Glc} appear to be involved in the regulation of adenylate cyclase activity (for reviews, see Saier, 1977; Ullmann and Danchin, 1983). The significance of homology between two such small blocks of amino acids is not clear. These sequences may define areas of interactions between both proteins and adenylate cyclase or with an as yet undefined area of the cytosolic membrane. It should also be noted that residues 180–190 of CRP and residues 27–37 of the CRO protein have been reported to be structurally very closely related. These segments of both proteins have also been proposed to interact with the major groove of B DNA (Steitz *et al.*, 1982). No such function for III^{Glc} has been proposed to date.

III^{Glc} is a multifunctional protein which is able to interact with a large number of other proteins. It is a PTS component which transfers a high energy phosphate from HPr to glucose via interactions with a membrane-bound component of the PTS, II^{Glc}. It serves to regulate the activity of several non-PTS transport systems via stoichiometric interactions (Nelson *et al.*, 1982,1983). III^{Glc} is also involved in the regulation of adenylate cyclase activity although the nature of this last regulatory effect is still not known. The construction of multicopy plasmids containing the structural gene for III^{Glc} together with the sequencing of this gene makes it possible to produce defined mutations and deletions of this gene. Analysis of the properties and activities of the resulting gene products should yield much information on the functional domains of this interesting protein and on the nature of its interactions with other proteins.

Materials and methods

Bacteria and Salmonella genomic phage bank

The *Salmonella* gene pool using λgt4 as a vector (Panasenko *et al.*, 1977) was constructed by A. DeFranco and T. Ingolia (DeFranco and Koshland, 1980) and generously donated by Dr D.E. Koshland Jr. Infection of the recipient *E. coli* strain LM1 (maltose grown) with the genomic phage bank was done by mixing 2 x 10^8 cells and 10^7 phage. The infected cells were plated and incubated at 30°C to obtain lysogens.

E. coli LM1 (*ptsM crr*) was a gift from Dr J. Lengeler. It is defective in both II^{Man} and III^{Glc} of the PTS.

Transduction with phage P22 was performed as described by Ely *et al.* (1974).

Media and growth of cells

Cells were grown in minimal salts medium A (Scholte and Postma, 1981) supplemented with the required nutrients (amino acids, vitamins) and 0.2% of a carbon source. LB medium (1% tryptone, 0.5% yeast extract and 0.5% NaCl) was used when growth in a rich medium was required. Eosine Methylene Blue plates (EMB plates) containing 1% sugar were used to monitor sugar fermentation.

Isolation of phage and phage DNA

Hybrid phage was isolated from a fully lysed culture which was grown in LB medium at 30°C, kept at 43°C for 15 min and grown for another 2 h at 37°C. The culture was centrifuged for 30 min at 50 000 *g* to remove cell debris, followed by 12 h at 50 000 *g* to pellet the phage. The pellet was suspended in 10 ml 5 M CsCl, containing 10 mM $MgSO_4$ and 0.1 mM EDTA and centrifuged for 24 h at 135 000 *g* at 4°C. The phage band was removed from the gradient with a syringe and contained ~10^{11} phage/ml. Phage titer was determined using *E. coli* C600.

After treatment of 10^{11} phage with pancreatic DNase I (Sigma), proteinase K (Merck) was added to a concentration of 10 μg/ml in a medium containing 25 mM EDTA, 1% sarkosyl and 10 mM Tris-HCl, pH 8. DNA was phenol-extracted and enthanol-precipitated twice. DNA was stored in 10 mM Tris-HCl, pH 8, at –20°C.

Subcloning of the crr gene in pAT153

λgt4-C1 DNA and the vector pAT153 (Twigg and Sheratt, 1980) were digested with *Eco*RI. After phosphatase treatment, the pAT153 vector was ligated with the λgt4-C1 *Eco*RI-digested DNA as described by Grosveld *et al.* (1981). The ligated DNA was used to transform LM1 cells as described by Mandel and Higa (1970). The mixture was plated on EMB glucose plates containing 100 μg ampicillin/ml and resulting fermenting colonies were analysed.

Isolation of plasmid DNA

Plasmid DNA was isolated according to the procedure of Birnboim and Doly (1979).

Synthesis of plasmid-encoded proteins in vivo

The proteins encoded by pBCP2 were labeled with [^{35}S]methionine *in vivo* using the procedure of Sancar *et al.* (1981).

Electrophoretic methods

SDS-polyacrylamide gel electrophoresis, rocket electrophoresis and crossed immunoelectrophoresis were performed as described elsewhere (Scholte *et al.*, 1982). Protein blotting (Towbin *et al.*, 1979), detection of polypeptides with anti-serum against III^{Glc} and labeling with [^{125}I]protein A have been described (Scholte *et al.*, 1982).

Immunoprecipitation of labeled proteins

^{35}S-labeled CSR603/pBCP2 cells (5 x 10^5 c.p.m.) were treated with toluene/chloroform/Triton X-100 (Scholte *et al.*, 1981) and incubated with anti-serum against III^{Glc} for 1 h at 37°C in 25 mM Tris-HCl, pH 7.5, containing 0.5 M NaCl and 0.5% Triton X-100. After addition of protein A-Sepharose beads (Pharmacia, 50 mg) the mixture was incubated for 16 h at 4°C with gentle shaking. The immunoprecipitate was washed twice with 50 mM Tris-HCl, pH 7, containing 1 M NaCl and 1% Triton X-100, twice with 0.5 M LiCl containing 1% Triton X-100 and 0.1% SDS and twice with 10 mM Tris-HCl, pH 7.5. The pellet was sedimented each time by centrifugation for 30 s in an Eppendorf microfuge. The pellet was analysed by SDS-polyacrylamide gel electrophoresis, using as marker proteins those supplied by Pharmacia.

Enzymes and reagents

Restriction endonucleases were from New England Biolabs or Boehringer Mannheim; DNA polymerase (large fragment), calf intestine phosphatase and T_4 DNA ligase from Boehringer; exonuclease Bal-31 from New England Biolabs; low-melting agarose from Bethesda Research Laboratories.

Assays

Restriction enzyme digestions, agarose gel electrophoresis, blot analysis of DNA fragments, nick-translation and hybridization were performed as described by Hoeijmakers *et al.* (1981). Bal-31 digestion was performed at 30°C for varying periods of time; 0.1 U of Bal-31 was used per μg of DNA. Incubations were stopped by the addition of phenol.

Cloning in bacteriophage M13 and sequence analysis

The plasmid pBCP20 containing the 1000-bp *Eco*RI-*Eco*RV insert fragment complementing *crr* was used as starting material for the sequence determination. M13 clones containing a varying part of the *Eco*RI-*Bam*HI fragment of pBCP20 (including ~200 bp of vector DNA) in two orientations were generated with BAL-31 using non-random cloning procedures (Poncz *et al.*, 1982) except that the gel isolation step of Bal-31 treated fragments was omitted. The procedure has been described in detail (Benne *et al.*, 1983). The resulting recombinant DNA probes either all have the *Bam*HI site in common and are progressively shortened by Bal-31 on the *Eco*RI side (procedure A), or all have the *Eco*RI site in common and are progressively shortened by Bal-31 on the *Bam*HI site (procedure B). In all cases, the side shortened by Bal-31 is oriented towards the vector priming site. The DNA sequence analysis was carried out by the dideoxynucleotide chain termination technique (Sanger *et al.*, 1977). The procedure yielded the sequences of overlapping segments of the *Eco*RI-*Eco*RV fragment in both orientations as shown in Figure 6.

Acknowledgements

We thank Dr D.E. Koshland for the generous gift of the lambda bank, Dr J. Lengeler and Dr J. Roth for the donation of strains, Dr L. Hensgens for the donation of M13mp8 phage, Dr M. Meyer for the use of restriction enzymes, and Dr A. Danchin for introducing us to the maxicell system.

References

Bachmann,B.J. (1983) *Microbiol. Rev.*, **47**, 180-230.

Benne,R., De Vries,B.F., Van den Burg,J. and Klaver,B. (1983) *Nucleic*

Acids Res., **11**, 6925-6941.
Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513-1523.
Bochner,B.R., Huang,H.-C., Schienen,G. and Ames,B.N. (1980) *J. Bacteriol.*, **143**, 926-933.
Britton,P., Murfitt,D., Parra,F., Jones-Mortimer,M.C. and Kornberg,H.L. (1982) *The EMBO J.*, **1**, 907-911.
Brosius,J., Cate,R.L. and Perlmutter,A.P., (1982) *J. Biol. Chem.*, **257**, 9205-9210.
Cordaro,J.C. and Roseman,S. (1972) *J. Bacteriol.*, **112**, 17-29.
Cordaro,J.C., Melton,T., Stratis,J.P., Atagün,M., Gladding,C., Hartman, P.E. and Roseman,S. (1976) *J. Bacteriol.*, **128**, 785-793.
Cossart,P. and Gicquel-Sanzey,B. (1982) *Nucleic Acids Res.*, **10**, 1363-1378.
DeFranco,A.L. and Koshland,D.E., Jr. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 2429-2433.
Ely,B., Weppelman,R.M., Massey,H.C. and Hartman,P.E. (1974) *Genetics*, **76**, 625-631.
Grosveld,F.G., Dahl,H.H.M., de Boer,E. and Flavell,R.A. (1981) *Gene*, **13**, 227-237.
Hoeijmakers,J.H.J., Snijders,A., Jansen,J.W.G. and Borst,P. (1981) *Plasmid*, **5**, 329-350.
Lengeler,J., Auburger,A.M., Mayer,R. and Pecher,A. (1981) *Mol. Gen. Genet.*, **183**, 163-170.
Mandel,M. and Higa,A. (1970) *J. Mol. Biol.*, **53**, 159-162.
Meadow,N.D. and Roseman,S. (1982) *J. Biol. Chem.*, **257**, 14526-14537.
Meadow,N.D., Rosenberg,J.M., Pinkert,H.M. and Roseman,S. (1982a) *J. Biol. Chem.*, **257**, 14538-14542.
Meadow,N.D., Saffen,D.W., Dottin,R.P. and Roseman,S. (1982b) *Proc. Natl. Acad. Sci. USA*, **79**, 2528-2532.
Nelson,S.O., Scholte,B.J. and Postma,P.W. (1982) *J. Bacteriol.*, **150**, 604-615.
Nelson,S.O., Wright,J.K. and Postma,P.W. (1983) *The EMBO J.*, **2**, 715-720.
Panasenko,S.M., Cameron,J.R., Davis,R.W. and Lehman,I. (1977) *Science (Wash.)*, **196**, 188-189.
Poncz,M., Solowiekczyk,E., Ballantine,M., Schwartz,E. and Surrey,S. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4298-4302.
Postma,P.W. (1982) *Ann. Microbiol. Inst. Pasteur*, **133A**, 261-267.
Postma,P.W. and Roseman,S. (1976) *Biochim. Biophys. Acta*, **457**, 213-257.
Postma,P.W., Epstein,W., Schuitema,A.R.J. and Nelson,S.O. (1984) *J. Bacteriol.*, **158**, 351-353.
Saier,M.H., Jr. (1977) *Bacteriol. Rev.*, **41**, 856-871.
Sancar,A., Wharton,R.P., Selzer,S., Kacinski,B.M., Clark,M.D. and Rupp, W.D. (1981) *J. Mol. Biol.*, **148**, 45-62.
Sanger,F., Nicklen,S. and Coulsen,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Scholte,B.J. and Postma,P.W. (1980) *J. Bacteriol.*, **141**, 751-757.
Scholte,B.J. and Postma,P.W. (1981) *Eur. J. Biochem.*, **114**, 51-58.
Scholte,B.J., Schuitema,A.R. and Postma,P.W. (1981) *J. Bacteriol.*, **148**, 257-264.
Scholte,B.J., Schuitema,A.R. and Postma,P.W. (1982) *J. Bacteriol.*, **149**, 576-586.
Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342-1346.
Staden,R. (1982) *Nucleic Acids Res.*, **10**, 4731-4751.
Steitz,T.A., Ollendorf,D.H., McKay,D.B., Anderson,W.F. and Mathews, B.W. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 3097-3100.
Towbin,H., Staehelin,T. and Gordon,J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4350-4354.
Twigg,A.J. and Sheratt,D. (1980) *Nature*, **283**, 216-218.
Ullmann,A. and Danchin,A. (1983) *Adv. Cyclic Nucleotide Res.*, **15**, 1-53.

Received on 2 April 1984