Volume 13 Number 21 1985 Nucleic Acids Research

T4-induced α- and β-glucosyltransferase: cloning of the genes and a comparison of their products based on sequencing data

Jörg Tomaschewski, Hermann Gram, John W.Crabb[1]* and Wolfgang Rüger+

Arbeitsgruppe Molekulare Genetik, Lehrstuhl Biologie der Mikroorganismen, and [1]Protein Sequenzierungslaboratorium, Institut für Physiologische Chemie I, Ruhr-Universität Bochum, D-4630 Bochum 1, FRG

Received 23 September 1985; Accepted 11 October 1985

ABSTRACT

Bacteriophage T4 α- and ß-glucosyltransferases link glucosyl units to the 5-HMdC residues of its DNA. The monoglucosyl group in α-linkage predominates over the one in ß linkage. Having recently reported on the nucleotide sequence of gene αgt (1) we now determined the nucleotide sequence of gene ßgt. The genes were each cloned on a high expression vector under the control of the λp_L promoter. After thermo-induction the proteins were isolated and purified to homogeneity. To verify that the translational starting sites and the proposed reading frames are effective in vivo the sequence of the first 31 amino acid residues from gpαgt and the first 30 amino acid residues from gpßgt were determined by Edman degradation.
The primary structures of the two proteins seem to have only limited structural similarities. The results are discussed comparing secondary structure predictions and homologies with other proteins from the protein sequence database of the Protein Identification Resource.

INTRODUCTION

Bacteriophages T2, T4 and T6 are known to carry in their DNA the pyrimidine base 5-hydroxymethylcytosine (5-HMC) in place of cytosine (2). In T4 DNA all 5-HMC residues are glucosylated, 70% in α-glycosidic and 30% in ß-glycosidic linkage (3). The two enzymes catalyzing the transfer of the sugar from UDPG to the hydroxymethyl group of 5-HMC are α- and ß-glucosyltransferase, respectively. The basis for the differential glucosylation is attributed in part to nearest neighbour conditions around the 5-HMC residues in a DNA molecule (4,5), but intracellular Mg^{2+} concentration and physiological partitioning (6) might be additional factors. The glucosyl-residues linked to 5-HMC prevent the in vivo restriction of the infecting viral DNA by means of the rgl system of the host, which recognizes and cleaves 5-HMC-containing DNA (for a review see ref. 7). The biological function of the glucosylation is not limited to protection of the infecting DNA, but it also influences transcription in vivo and in vitro (8, 9, 10). Josse and Kornberg (11) had previously purified both

enzymes and gave evidence that the general reaction mediated by the two polypeptides is

UDP-glucose + HMdC-DNA = UDP + glucosyl-HMdC-DNA

In their study T4 ß-glucosyltransferase (gpßgt) differed from α-glucosyltransferase (gpαgt) in its requirements for sulfhydryl, ionic environment, and in the extent of glucosylation of different 5-HMC-containing deoxyribonucleic acids. However, to date the enzymes have not been thoroughly characterized.

Gene αgt is located between genes 47 and 55 on the genetic map of bacteriophage T4 and gene ßgt is located between genes 41 and 42 (12, 13, 14). In vivo transcription of the two genes seems to take place at early and quasi late times of the viral growth cycle, and the amount of enzyme generated in the cells is relatively limited.

Recently we published the nucleotide sequence of gene αgt (1), which codes for a polypeptide of 400 amino acid residues (Mol. wt. of 46,651 daltons). We now report the nucleotide sequence of gene ßgt and the purification and partial characterization of the of α- and ß-glucosyltransferases from overproducing bacteria.

MATERIALS AND METHODS

Agarose was purchased from Sigma Biochemical Company and acrylamide and N,N'bisacrylamide from Serva, Heidelberg. Biodyne nylon membranes (BNNG2) used for hybridization were obtained from Pall, Dreieich, unlabelled deoxyribonucleoside triphosphates from Boehringer, Mannheim, dideoxyribonucleoside triphosphates from PL Biochemicals, α-^{32}P-dATP (400 Ci/mmol) and uridine diphospho-D- 6-^{3}H glucose (4.7 Ci/mmol) from Amersham Buchler. The ion exchange and gel filtration media were DE52 (Whatman), Bio-Rex70 (BioRad), and Sephacryl S200 (Pharmacia).

Enzymes: T4 DNA ligase was isolated by a combination of methods described in the literature from a ligase-overproducing E. coli strain. (15, 16, 17, 18). Restriction endonucleases were either isolated as described (19) or they were purchased from Boehringer or Bethesda Research Laboratories. E. coli DNA polymerase I, nuclease SI and alkaline phosphatase were obtained from Boehringer, Mannheim.

Bacteria and Phages: T4 dC-DNA was prepared by growing the pentuple mutant T4 42^{-} (amN55x5)-56^{-}(amE51)-denA^{-}(nd28)-denB^{-}(D2a23)-alc^{-}(alc10) on E. coli B834 gal^{-}(U56)(met^{-}, gal^{-}U, su^{-}). Due to several point mutations in genes of the nucleotide pathway, this strain replaces the HMC in its DNA by cy-

tosine when grown on a permissive host. Cytosine-containing T4 DNA is cleaved by restriction endonucleases, T4 wild type DNA, with few exceptions, is not. Marker rescue experiments were carried out with the amber double mutant T4 αgt^-(am8), βgt^-(am10) grown on E. coli CR63. The phages were kindly supplied by Dr.J.S.Wiberg and E. coli B834 by Dr.E.M.Kutter. M13 mp8 and mp9 phages were grown on E. coli JM103 [Δ(lac,pro),supE, thi, strA, endA, sbcB15, hsdR4,(F'traD36, proAB, $lacI^qZ\Delta M15$)]. The M13 phages and E. coli JM103 were obtained through Dr.R.Eichenlaub. E. coli K12 1100 (λNM 989)3, the ligase-overproducing strain, was kindly donated by Dr.K.Murray. E. coli W4597, which has a deficiency in the pathway involving UDPG-phosphorylase and which upon infection yields wild type phage progeny carrying non-glucosylated DNA (20, 21), as well as bacteriophages T2 and T6, were taken from this institute's collection.

Plasmids: The vector for the construction of a T4 clone library was pBR322 (22) which was grown in E. coli JA221, a derivative of E. coli C600. The genetic markers are recA1, leuBG, trp ES, $hsdR^-$, $hsdM^+$, lacy. For the expression of the enzymes plasmid vector pPLc2833 (23) was transformed first into E. coli MC1061 [araD139, Δ(ara,leu)7697, ΔlacX74, $galU^-$, $galK^-$, hsr^-, hsm^-, strA] for selection and amplification and then for expression intoK12ΔH1Δtrp [Sm^r, lacZam, Δbio-uvrB, ΔtrpEA2 (λNam7,Nam53,cI857,ΔH1)]. The expression vector and both host strains were kindly supplied by Dr.E.Remaut.

DNA Sequencing: The subclones necessary for the sequencing procedure were constructed as already described (1). Restriction fragments carrying gene βgt were either subjected to further digestion with restriction endonucleases Taq1, Sau3A, HinfI, HpaII, and AluI, which cleaves T4 DNA frequently, or they were broken by sonic shear (24). The subfragments were cloned into M13mp8 and sequenced at random. Sequencing by chain termination (25) and preparation of buffer gradient gels was carried out as described for α^{32}P-dATP by Biggin et al. (26). Single- and double-stranded M13 DNA was prepared by the method of Messing et al. (27). For all transfections we followed the method of Dagert and Ehrlich (28) and in later experiments that of Hanahan (29). As the sequencing progressed, M13 clones which carry DNA fragments spanning gaps in the sequence were identified by dot and plaque hybridization (30, 31, 32).The DNA sequencing data were compiled and analyzed using the computer programs of Staden (33).

Marker Rescue Experiments: E. coli JM103 were infected with phage M13 carrying defined DNA inserts. The M13-infected bacteria were grown to

$5x10^8$ cells per ml. Cells were then superinfected with T4 αgt^-(am8), βgt^-(am10) with an m.o.i. of 5. After two hours of incubation at 37^oC the bacteria were lysed by the addition of chloroform. Phage progeny were titered on E. coli K803 or on E. coli JM103 for total phage count and on E. coli BA for wildtype recombinants.

Purification of Enzymes: All operations were carried out at $0-4^oC$ unless otherwise stated. Cells were grown in 1 liter of a tryptone-broth-yeast medium (TBY) containing 100 mg of ampicillin. For efficient aeration the culture was grown in a 3-liter Erlenmayer flask on a New Brunswick shaker at 30^oC. When the cells reached a density of about $5x10^8$ per ml, they were induced by the addition of 340 ml of TBY at 90^oC. This manipulation brought the temperature of the medium at once to about 42^oC. Cells were allowed to continue growth at this temperature for another 2h. Induced cells were harvested by centrifugation for 10 min at 5 000xg yielding 3-5 g of cell paste.

Isolation of α-Glucosyltransferase: 5 g of induced cells were resuspended in 40 ml of buffer A containing 50 mM Tris-HCl, pH8.0, 10 mM EDTA, 10 mM dithioerythritol (DTT) and 0.02 mM phenylmethylsulfonyl fluoride. Resuspended cells were sonicated with a Branson sonifier for 6 x 30 sec at maximum energy output. Cell debris was sedimented by centrifugation for 30 min at 48 000xg. To 38 ml of the centrifugation supernatant 62 ml of a solution of ammonium sulfate, saturated at 4^oC, were added. After 30 min of stirring the precipitate was collected by centrifugation for 20 min at 27 000xg. The protein pellet was resuspended in 10 ml of buffer B containing 20 mM Tris-HCl, pH 7.4, 1 mM EDTA and 5mM ß-mercaptoethanol (ß-ME). After dialysis against buffer B the protein was applied to a column (18x2 cm) of BioRex70 eqilibrated with buffer B. The column was washed with two volumes of buffer B (EDTA omitted) and developed with a linear gradient from 0.1-0.8 M NaCl in buffer B. α-Glucosyltransferase elutes at about 0.35 M NaCl. The peak fractions were pooled and dialysed against buffer C containing 20 mM Tris-HCl, pH 7.8, 50 mM NaCl, and 5mM ß-ME. The protein then was applied to a DEAE column (10x1 cm), equilibrated with buffer C. The column was washed with two volumes of buffer C and eluted with a linear gradient from 0.05 M-0.3 M NaCl in buffer C. The enzyme elutes at about 0.12 M salt. The peak fractions were concentrated by dialysis against buffer D containing 50 mM Tris-HCl, pH 8.0, 1 mM DTT and 50% glycerol and stored at -20^oC. No appreciable loss of activity was observed during 2-3 month.

Isolation of ß-Glucosyltransferase: 5g of induced cells were resuspended in 10 mM potassium phosphate, pH 8.0, containing 1 mM DTT (buffer A). Cells were sonicated as described for gpαgt and cell debries were removed by 20 min centrifugation at 14 000xg, followed by 2h of centrifugation at 100 000xg. The supernatant was applied to a column of BioRex70 (20x1 cm) equilibrated in buffer A. The column was washed and thereafter developed with a linear gradient from 0-1 M NaCl. The enzyme eluted at about 0.35 M salt. The peak fractions (16 ml) were collected and the protein was precipitated by the addition of 9g of crystalline ammonium sulfate (final concentration about 80%). The precipitate was collected by centrifugation for 15 min at 27 000xg, resuspended in buffer A, dialyzed, and transferred to a Sephacryl S200 column (30x1.8 cm). The column was developed with buffer A containing 100 mM NaCl. The enzyme fractions were pooled and concentrated by dialysis against storage buffer which contained 10 mM potassium phosphate, pH7.0, 0.25 mM DTT, and 50% glycerol. The enzyme was stored at -20°C. The activity of the enzyme decreased to 50% during a storage through 3 months.

Enzyme Activity Assay: The glucosylation of 5'HMC residues was measured by incorporation of ^{3}H-glucose into an HMC-containing DNA acceptor after incubation with ^{3}H-UDPG in the presence of α- or ß-glucosyltransferase. The mixture for αgt contained in a 100 ul volume 100 mM Tris-Hcl, pH 8, 80 mM ß-Me, 20 mM ammonium sulfate, 3.4 ug T4* DNA (isolated from T4 phage progeny grown in E. coli 4597), 0.5 uCi ^{3}H UDPG (spec. act. 4.7 Ci/mmol), 0.1 mM of unlabeled UDPG, and variable amounts of gpαgt. Incubation was at 30°C for 15 min.

The assay mixture for ßgt contained in a 100 ul volume 100 mM potassium phosphate, pH 7.6, 25 mM $MgCl_2$, 3.4 ug of T4*DNA, 0.5 uCi of ^{3}H-UDPG (spec.act. as given above), 0.1 mM of unlabeled UDPG and variable amounts of gpßgt. After 15 min of incubation at 30°C the DNA was precipitated with 7% perchloric acid (PCA). The precipitate was collected by centrifugation, washed with 5% PCA, then with 80% ethanol. The sediment was taken up in 100 ul of 2N NH_4OH, pipetted onto Whatman 540 filters, dried and counted in a toluene scintillation cocktail.

Protein Structural Characterization: Automated Edman degradations were performed on the intact, unmodified proteins (380-2500 pmol) with an Applied Biposystems model 470A gas-phase sequenator (34). Polybrene was prepared essentially as described by Lai (35) and precycled once in the sequenator. Phenylthiohydantoin (PTH) amino acids were identified by HPLC on an IBM Cyano column (36) using a modified 20 min program, a column

buffer composed of 5% tetrahydrofuran in 20 mM Na-acetate, pH 5.2, and an acetonitrile gradient. Amino acid analysis was performed as described by Tarr (37) using HPLC on a Beckman ultrasphere ODS column to separate and quantify phenylthiocarbamyl (PTC) derivatized amino acids.
Search for Homologous Sequences: The Protein Identification Resource data bank of the National Biomedical Research Foundation (3061 sequences, February 1985 revision) was searched for homologous sequences, using a 25 residue window probe on a VAX/VMS computer according to Dayhoff (38). Other computer homology search systems were also utilized as indicated under Acknowledgements.

RESULTS

Determination of the DNA Sequence of gßgt

Using cytosine-containing T4 DNA isolated from purified phage particles of strain T4 JSW800 and plasmid pBR322 as a cloning vector a T4xHindIII clone library was constructed. The ligated DNA was transformed into E. coli JA221. Amp^r, tet^s colonies were selected. The amplified plasmid DNA of 160 clones was isolated (39) and hybridized against patterns of restriction fragments obtained by digestion of T4dC DNA with restriction endonucleases SalI, XhoI and EcoRI (40). Since the physical map of T4 is well established and aligned with the genetic map (19, 41,42), five clones could be chosen which according to their hybridization to distinct restriction fragments probably carried T4 gene ßgt. The T4 inserts of these five clones were separated from the pBR322 DNA by HindIII cleavage and analyzed with regard to the pattern of other restriction enzyme cleavage sites predicted to be situated in this area of the physical map of T4. One of the clones (pHG 205) was chosen, the 2.2 kb T4 insert was purified on and recovered from preparative agarose gels. The DNA insert was fragmented to pieces smaller than 500 bases (43, 44,1). These subfragments were cloned into phage M13mp8 and finally sequenced by the chain termination method. A set of about 75 M13 clones carrying mutually overlapping T4 DNA inserts covered the region between 23.24 kb and 25.46 kb on the map of T4. The nucleotide sequence of the entire region was composed by means of a computer program (33) and analyzed for open reading frames. In Fig.1 we show the position of 45 gel readings which resulted in the sequence of gßgt and the position of gßgt reading frame within the 2.2 kb sequence. The complete DNA sequence of gßgt and the derived amino acid sequence are presented in Fig.2.
Gene ßgt was identified by marker rescue experiments. To this end E. coli

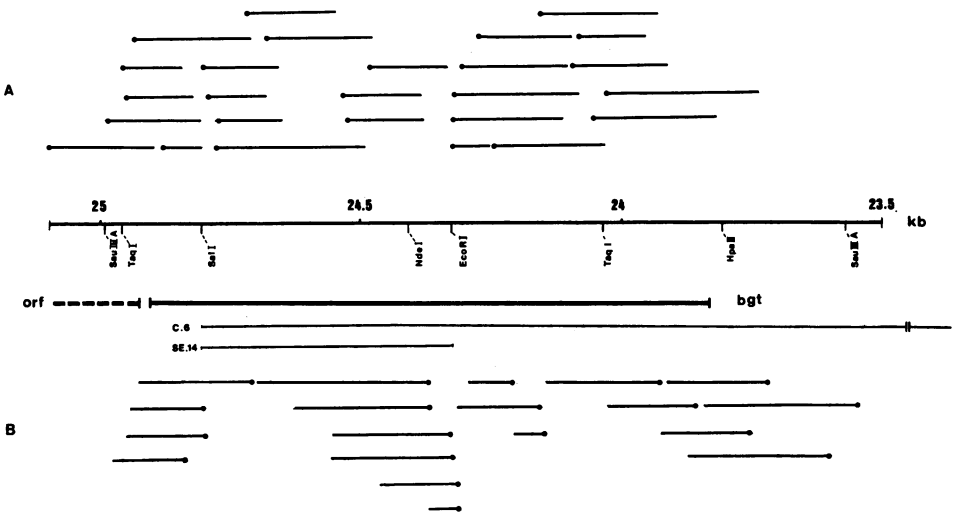

Fig. 1: The Position of gßgt within the Physical Map of Bacteriophage T4 and the Sequencing Strategy. The schematic outline represents the genomic region between 23.0 and 25.0 kb on the physical map of T4 (42, 55). Shown are the positions of the restriction enzyme cleavage sites in gßgt and its flanking regions. Phage M13 derivatives carrying the T4 inserts C.6 and SE.14 responded positively in the rescue experiments for the genetic marker ßgt. A and B show the positions of 45 gel readings from which the nucleotide sequence of gßgt was composed. Points represent the 5' end of each individual sequence, bars and points together indicate the extent of each reading.

JM103 was infected with selected recombinant M13 phages utilized in the sequencing procedure. These cells subsequently were infected separately with the T4 αgt^-,βgt^-, amber mutant. Due to recombinational events between mutant alleles on the T4 genome and the T4 wild type alleles inserted into the phage M13 DNA, wild type progeny are generated which can be selected for on E. coli su^- strains (45). The results of the marker rescue tests are summarized in table 1.

The product of gene ßgt is composed of 351 amino acid residues which add up to a mol. wt. of 40 613 daltons.The polypeptide carries 15.4% of positively charged and 12.3% of negatively charged amino acid residues. A polarity index of 48.7 was calculated according to Capaldi and Vanderkooi (46). For a comparison the corresponding values for gpαgt are 400 amino acid residues adding up to a mol. wt. of 46 651 daltons, 15.5% of positively charged and 14.3% of negatively charged amino acid residues and a polarity index of 51.3. Secondary structure predictions according to Chou and Fasman (47) indicate 29% of α-helix conformation, 36% of ß-pleated

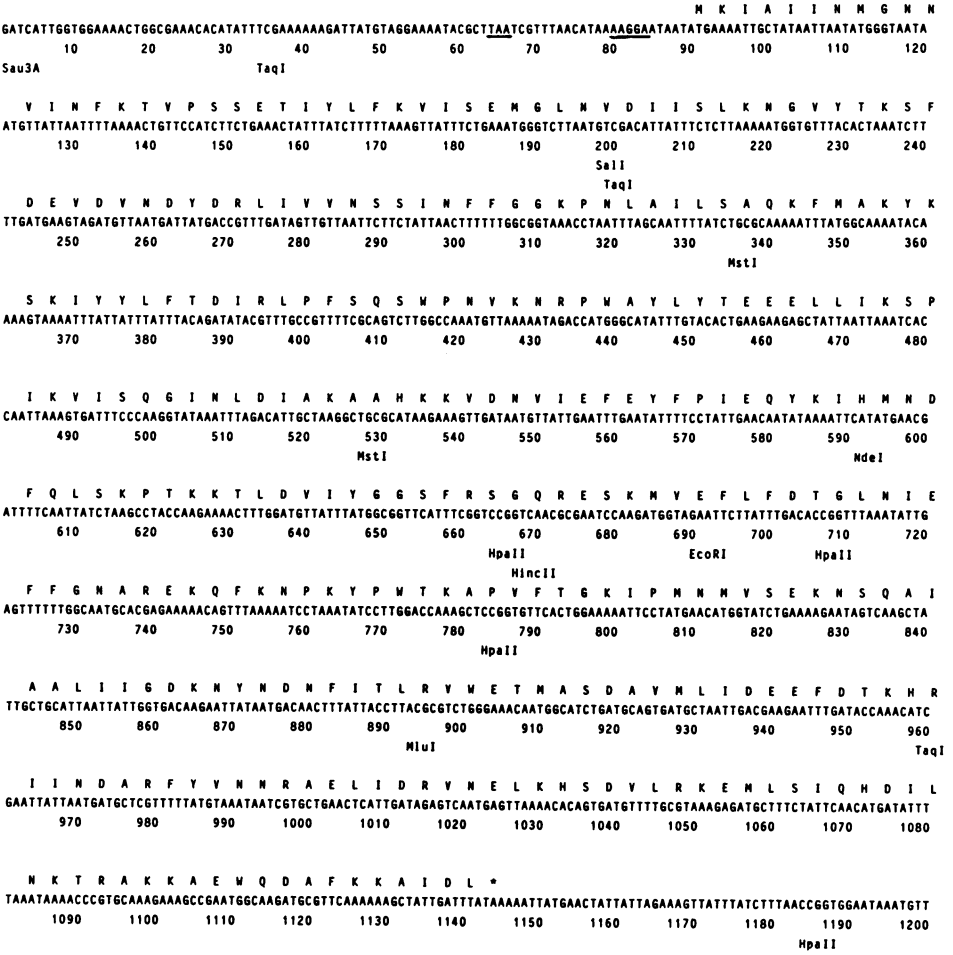

Fig. 2: The Nucleotide Sequence of gßgt and the Deduced Primary Structure of the Gene Product. The sequence of 1200 base pairs shown here contains an open reading frame coding for 351 amino acid residues. The primary structure of the protein is given in the one letter amino acid symbols. The stop codon of the preceding orf and the hypothetic ribosome binding site upstream from gßgt are underscored.

sheet and 25% of reverse turns for gpßgt. For gpαgt these values were calculated earlier to be 18% α-helix conformation, 44% of ß-pleated sheet and 27% of reverse turns (1). Interestingly, in contrast to gpαgt which contains 6 Cys residues, gpßgt contains no Cys and, therefore, its tertiary structure is not influenced by -S-S- bonds. All 6 Cys residues found in the amino acid sequence of α-glucosyltransferase of bacteriophage T4

Table 1: The Rescue Tests for Marker ßgt⁻ (am 10).

Plasmid	pHG205	C6	SE14	pBR322	mp8
Ratio	8.0×10^{-3}	3.3×10^{-3}	1.8×10^{-3}	4.6×10^{-7}	2.3×10^{-8}

Shown are the ratios of wild type progeny (titered on E. coli BA) to total progeny (titered on E. coli JM103 or K803). Cells infected with M13mp8-T4 hybrids and superinfected with the T4 amber mutant served as a control. The αgt⁻,ßgt⁻ double mutant was chosen since its growth is restricted in E. coli BA due to the rgl system (68). For experimental details see Materials and Methods.

are conserved in the corresponding enzyme coded by bacteriophage T6, and in the T2-coded enzyme 5 Cys residues are conserved and one is replaced (48).

Codon Usage in gαgt and gßgt.

The codon usage in these two T4 genes was compared to the one in strongly expressed genes of E. coli (49). The results are summarized in table 2. Generally, tripletts with C or G in the third position are used less frequently than those with A or U. Certain codons that are rarely used in strongly expressed E. coli genes appear more frequently in gαgt and gßgt. Among these are the ones recognized by T4 encoded tRNAs (under-

Table 2: Codon Usage in Genes αgt and ßgt of Bacteriophage T4.

1.		2. U				2. C				2. A				2. G				3.
		A.A.	αgt	ßgt	E.coli	A.A	αgt	ßgt	E.coli	A.A.	αgt	ßgt	E.coli	A.A.	αgt	ßgt	E.coli	
		PHE	82.6	86.4	26	SER	35.5	59.1	36	TYR	77.7	78.6	26	CYS	66.7	0.0	36	U
	U	PHE	17.4	13.6	74	SER	9.6	13.6	34	TYR	22.3	21.4	74	CYS	33.3	0.0	64	C
		LEU	19.3	52.0	3	SER	29.0	9.1	2	---	---	---	--	---	---	---	--	A
		LEU	19.3	20.0	4	SER	3.2	4.5	5	---	---	---	--	TRP	100	100	100	G
		LEU	25.8	16.0	6	PRO	50.0	50.0	10	HIS	90.0	80.0	20	ARG	61.9	46.3	67	U
	C	LEU	12.9	4.0	8	PRO	0.0	0.0	1	HIS	10.0	20.0	80	ARG	4.7	15.4	30	C
		LEU	9.6	8.0	1	PRO	14.3	33.3	12	GLN	46.1	80.0	18	ARG	29.0	15.4	1	A
		LEU	12.9	0.0	79	PRO	35.7	16.7	77	GLN	53.9	20.0	82	ARG	4.7	7.7	0.3	G
		ILE	89.6	88.6	20	THR	52.3	42.9	36	ASN	81.0	85.7	8	SER	12.9	13.6	4	U
	A	IIE	10.4	0.0	79	THR	4.7	42.9	48	ASN	19.0	14.3	92	SER	9.6	0.0	19	C
		ILE	0.0	11.4	1	THR	33.3	14.3	5	LYS	90.3	77.8	71	ARG	0.0	15.4	1	A
		MET	100	100	100	THR	9.4	0.0	10	LYS	9.7	22.2	29	ARG	0.0	0.0	0.3	G
		VAL	63.6	54.4	44	ALA	56.2	42.9	37	ASP	84.6	68.2	36	GLY	72.2	69.2	54	U
	G	VAL	0.0	13.6	9	ALA	0.0	4.8	10	ASP	15.4	31.8	64	GLY	11.1	23.1	42	C
		VAL	31.8	18.2	27	ALA	37.5	38.1	25	GLU	87.1	81.0	76	GLY	16.6	7.7	1	A
		VAL	4.7	13.6	19	ALA	6.3	14.3	28	GLU	12.9	19.0	24	GLY	0.0	0.0	3	G

The values for strongly expressed genes are as shown by Grosjean and Fiers (49). All values are expressed in per cent to take into account the different amino acid compositions of the gene products. The codons underscored are those that are complementary to the anticodons of seven T4 encoded tRNAs (69). The eigth T4 encoded tRNA, $tRNA^{ILE}$, carries the anticodon NAU (70) and codes preferentially for AUA but also for AUC and AUU.

lined) but also AAU, CGA and CGG. The codons CCC, GGG and AGG do not appear in either one of the genes.

Cloning Genes αgt and ßgt into a High Expression Vector.

Both genes were cloned into a plasmid expression vector to obtain substantial amounts of the enzymes. Remaut et al. (23, 50) recently constructed a pBR322 derivative which allows the insertion of genes downstream from the leftward promoter (p_L) of bacteriophage λ. The activity of p_L is fully repressed at 28°C by a thermolabile repressor (cI875). The genetic information for this repressor is encoded in a nonexcisable λ prophage which resides in the genome of the bacterial host strain. The prophage at the same time is deleted for the cro gene and does not express a functional N gene product. At 42°C the repressor is fully inactivated resulting in the transcription of the inserted DNA.
Gene ßgt was isolated from plasmid pHG205 as a coherent Sau3A fragment and cloned directly into the unique BamHI site of pPLc28. The resulting plasmid carrying gßgt was designated pPLcßgt. Gene αgt had to be fused by ligation from two subfragments taking advantage of a SstI/SacI cleavage site found in its sequence (Fig.3). The major portion of the gene was isolated from plasmid pHG28, a derivative of pBR322 carrying a 2694 bp T4 insert which contains about 91% of gαgt. The missing 9% were taken from the M13-T4 hybrid S18 which reacts positively in marker rescue experiments for gαgt (1). The EcoRI-SalI fragment resulting from this ligation starts about 160 bases upstream and ends about 340 bases downstream of gαgt. It was cloned into the EcoRI and SalI sites of plasmid pPLc2833, yielding pPLcαgt. Since the EcoRI site is situated 5' of the SalI site we could expect that gαgt was inserted in the proper orientation with respect to the λp_L promoter. Since in pPLc2833 there is no system selective for DNA inserted at the multi-linker site, a number of ampicillin resistant colonies were screened on polyacrylamide gels for protein over-production at 42°C (Fig.4). It can be seen that in the case of clone pPLcαgt an over-produced protein appears at approximately 46 000 dal and that with pPLcßgt a strong protein band appears at about 40 000 dal. These molecular weights agree with the values predicted from the nucleotide sequences of the genes.

Purification of the α- and ßgt Enzymes.

The α- and ß-glucosyltransferases were purified from bacteria carrying the T4 genes αgt and ßgt on expression vectors by shorter and higher yield methods than those originally described by Josse and Kornberg (11). For the preparation of both gene products induced cells harbouring pPLcαgt or pPLcßgt

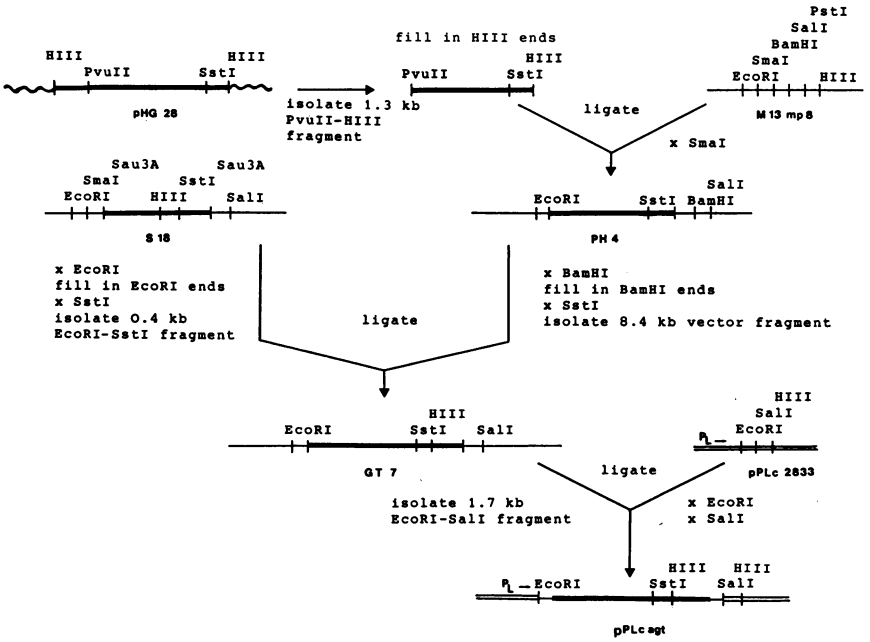

Fig. 3: The Cloning of T4 Gene αgt into a high Expression Vector. The cloning experiment was performed in three steps. First the larger portion of gαgt was isolated as a 1.3 kb PvuII-HindIII fragment from pHG28 (1). This part of gαgt then was cloned by blunt end ligation into the SmaI site of phage M13mp8 resulting in the recombinant phage PH4. The double-stranded DNA of PH4 was cleaved by restriction endonuclease BamHI and the generated ends were converted into blunt ends using the Klenow fragment of DNA polymerase I. The linearized PH4 was then cleaved with restriction endonuclease SstI. The resulting 8.4 kb vector fragment was isolated and the short SstI-BamHI fragment was discarded.
In the second step the smaller portion of of gαgt was isolated from the recombinant M13 phage S.18 (1) as a 0.4 kb EcoRI-SstI fragment. To this end M13 S.18 was first cleaved with EcoRI. To obtain blunt ends at this position the EcoRI sites were filled in using the Klenow fragment of DNA polymeraseI. A successive cleavage with SstI generates a 0.4 kb fragment which was recovered from an agarose gel and then ligated into the 8.4 kb vector fragment gained previously from PH4. The fragment was forced into the proper direction taking advantage of the SstI sites on the one hand and the filled in EcoRI and BamHI sites on the other. The resulting M13-T4 hybrid was designated GT7.
In the third step GT7 was cleaved successively with EcoRI and SalI; the resulting EcoRI-SalI fragment carries gαgt. It was cloned into the EcoRI-SalI site of pPLc2833. Note that the EcoRI sites cleaved in M13 S.18 and M13 GT7 are not identical.
In this context it should be pointed out that S18 but not PH4 reacts positively in marker rescue experiments for gαgt. Hence, the portion of the COOH-terminus of α-glucosyltransferase coded by the last 12 triplets is necessary for the activity of gpαgt. The only codon in register which can be mutated in one step into an amber codon is at position 1190 and codes for Gln.

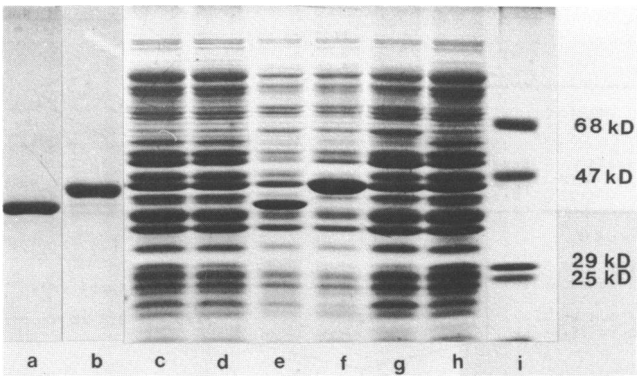

Fig. 4: SDS-Polyacrylamide Electrophoresis of gpαgt and gpßgt.
Polyacrylamide gel electrophoresis in the presence of SDS was performed essentially according to (67). Channel a: 25 ug of gpßgt; channel b: 25 ug of gpαgt; channel c: crude extract of cells grown at 30°C carrying pPLcßgt; channel d: crude extract from cells grown at 30°C and carrying pPLcαgt; channel e: crude extract from cells carrying pPLcßgt after heat induction; channel f: crude extract from cells carrying pPLcαgt after heat induction; channels g and h: crude extracts from cells carrying pPLc2833, grown at 42°C and at 30°C, respectively. channel i: molecular weight markers, of 68 kD (bovine serum albumin), 47 kD (chicken serum albumin), 29 kD (carbonic anhydrase) and 25 kD (chymotrypsinogen).

were harvested by centrifugation and resuspended in a low salt buffer. The cellular suspension was sonicated and cellular debris were removed by low speed centrifugation (15 000xg).

Gpαgt then was subjected to an ammonium sulfate precipitation up to 60%. The precipitated protein was harvested by centrifugation, resuspended and dialyzed against low salt buffer followed by two ion exchange chromatographies, first on BioRex70 then on DE52.

Gpßgt was purified by high speed centrifugation (100 000xg), the supernatant was fractionated by an ion exchange chromatography on BioRex70, followed by an ammonium sulfate precipitation of those portions containing the enzymic activity, and a fractionation on Sephacryl S200. The final enzyme preparations appeared to be homogeneous as judged by SDS-polyacrylamide electrophoresis and Edman degradation.

Both proteins catalyzed the reaction expected, i.e. the glucosylation of 5-HMC-containing T4 DNA. A summary of the purification of the α- and ß-glucosyltransferases is given in Table 3. We also assayed the extent of glucosyl transfer by the purified gene products to T2, T4*, and T6 DNA. In these experiments the HMC residues in T4*DNA were glucosylated

Table 3: Purification Scheme for α- and ß-Glucosyltransferase.

	protein mg	spec. activity µmoles/µg	yield %
Purification of gpαgt			
crude extract	304	14.4	100
Ammoniumsulfate precipitate	152	23.0	79.9
Bio-Rex 70	28	77.4	49.5
DE 52	12	117.1	38.2
Purification of gpßgt			
crude extract	109	19.7	100
100.000 g supernatant	85	12.4	48.7
Bio-Rex 70	15	20.2	13.5
Sephacryl G200	4	38.6	7.0

to 62.5% and 92.6%, in T2 DNA to 15.0% and 27.7% and in T6 DNA to 1.5% and 15.0% by gp gt and gpßgt, respectively. These results are in reasonable agreement with those obtained by Josse and Kornberg (11). In gel sieve chromatographies and on sucrose gradients gpßgt is fractionated into two peaks of enzymic activity, one having approximately twice the mol. wt. of the other. Analysis of the two enzyme fractions on polyacrylamide gels in the presence of SDS shows identical, single protein bands suggesting that this protein tends to form dimers (data not shown).

NH_2-Terminal Sequence Analysis of the α- and ß-Glucosyltransferases.

The NH_2-terminal amino acid sequence of the unmodified T4 α- and ß-glucosyltransferases were determined directly by automated Edman degradation of the purified enzymes. The sequence of the first 31 residues of gpαgt, determined from 2500 pmoles protein (repetitive yield, 94.8%), is Met-Arg-Ile-?-Ile-Phe-Met-Ala-Arg-Gly-Leu-Glu-Gly-?-Gly-Val-Thr-Lys-Phe-Ser-Leu-Glu-Gln-Arg-Asp-Trp-Phe-Ile-Lys-Asn-Gly. The amino acids unidentified by Edman degradation at positions 4 and 14 are cysteine residues as deduced from the αgt gene sequence. The sequence of the first 30 residues

of gpßgt, determined from 560 pmoles protein, (repetitive yield, 97.8%) is Met-Lys-Ile-Ala-Ile-Ile-Asn-Met-Gly-Asn-Asn-Val-Ile-Asn-Phe-Lys-Thr-Val-Pro-Ser-Ser-Glu-Thr-Ile-Tyr-Leu-Phe-Lys-Val-Ile. Both the αgt and ßgt N-terminal sequences determined directly from the proteins correspond with their respective translated DNA sequences.

DISCUSSION

Knowing the position of a large number of restriction endonuclease cleavage sites on the physical map of bacteriophage T4, we identified, cloned and sequenced a T4 DNA fragment encoding gßgt. The nucleotide sequence upstream from this gene exhibits no homology with either the -10 or -35 conserved sequences of promoters serving other phage T4 early or middle mode genes (51, 52). There is, however, a non-coding region of 22 nucleotides in length between the AUG start codon of gßgt and the UAA stop codon of an open reading frame preceding gßgt. Within this leader region there is the pentanucleotide -A-A-G-G-A-, which is complementary to the 16S rRNA and which could serve as a ribosome binding site (53). Stem and loop structures characterizing transcriptional termination sites (54) are found neither proximal nor distal to gßgt. There is also no overlap between the stop codon UAA of gßgt and the AUG initiation codon of the following gene. These findings suggest that gene ßgt may not have a promoter to its own and therefore may be transcribed into a polycistronic message originating from a promoter further upstream. A strong early promoter located previously between genes 43 and 42 (55) might be a possible candidate. Translation of the ßgt message seems to be independent of the translation of the flanking genes (56).

Upon transformation of *E. coli* with pPLcαgt and pPLcßgt and subsequent induction of transcription, α- and ß-glucosyltransferases are produced at about 15% of the soluble protein fraction as judged from polyacrylamide gels. Crude extracts obtained by sonication of these cells are capable of glucosylating 5-HMC-containing DNA. Since McNicol and Goldberg (6) proposed a physiological partitioning of α- and ß-glucosyltransferase to explain the differential glucosylation catalyzed *in vivo* by the two proteins it is of interest that both T4 enzymes were isolated from the soluble portion of the cell extracts. Virtually no enzyme activity was detected in the membrane fractions. Neither the hydrophobicity indices close to 50 for both proteins (46) nor the hydropathy profiles

(57) of the two enzymes, which do not show membrane spanning regions, give evidence that one of the two enzymes might reside in the membrane. The fact that the ß-glucosyltransferase is capable of totally glucosylatting the DNA seems to imply that the DNA is initially exposed only to the α-glucosylating enzyme. Sequential glucosylation might be controlled by a multienzyme complex (58) of the infected cell.

A comparison of the primary structures of these two proteins by the means of computer programs revealed one sequence identity of 4 amino acids common to both proteins and in another region an identity of 6 amino acids within a sequence of 9. The tetrapeptide -Gln-Leu-Ser-Lys- appears in gpαgt at residues 292-295 and in gpßgt at residues 173-176. The other region is -Val-Asn-Asp-Cys-Asp-Ile-Leu-Ile-Ile- in gpαgt (residues 72-80) and -Val--Asn-Asp-Tyr-Asp-Arg-Leu-Ile-Val- in gpßgt (residues 56-64). Homology searches performed with the programs 'Align' and 'Relate' (38) revealed no similarities clearly above statistical margins. Computed secondary predictions (59, 60, 61, 63) indicate no obvious conformational relationship and the supersecondary structure ßαßaß, which has been correlated with a nucleotide fold in a number of proteins (64,52) is not evident in either gpαgt or gpßgt. If at all, the two proteins seem to be only distantly related and more experimentation will be necessary to elucidate whether the sequence homologies mentioned above are of functional significance.

Since the amino acid sequences of a number of proteins which bind to DNA and/or ribonucleotides are known, the gpαgt and gpßgt sequences were compared with the protein sequence database of the Protein Identification Resource. Some similarities were detected with proteins such as RNA polymerase ß chain, T7 DNA polymerase, the probable DNA polymerase of hepatitis B virus, EcoRI endonuclease, the D-galactose binding protein (gpßgt) and T4 DNA ligase, EcoRI and EcoRV endonuclease, EcoRI and EcoRV methylase, T7 RNA polymerase, the probable DNA polymerase of the hepatitis B virus and the D-galactose binding protein (gpαgt). The most extended similarity observed was between gpßgt and proteins of the leucocyte interferon family of the database (residues 241-338 of gpßgt, 24 amino acid identities and a number of homologous exchanges within 98 residues). Indepth statistical analyses could perhaps point out significant structural relationships, however, the extent of similarity does not at present appear to be promising. In addition, ambiguities still associated with conformational predictions (65, 66) limit such comparisons.

ACKNOWLEDGMENT
The authors wish to thank all persons who provided us with bacterial and phage strains. We are grateful to Dr.Staden (MRC Laboratory, Cambridge, UK) for allowing us to use his sequence analysis program, to Drs M. Parker (Seattle), C. Sander (Heidelberg), and E. Spicer (New Haven) for performing the computer homology searches in the protein sequence database, to Dr. C.Block (Dortmund) for allowing us to use his version of protein secondary structure predictions, to M.F.J. Rolofs for adapting the computer programs, to Drs C.K.Mathews (Corvallis) and V.Tanyashin (Pushino) for discussion. W.R. thanks Mrs.U.Aschke for expert technical assistance. The financial support of this study by the Minister für Wissenschaft und Forschung, Nordrhein-Westfalen, is gratefully acknowledged.
Additional data i.e. the hydropathy profiles of the enzymes, their computed secondary predictions, homology searches as well as the complete αgt nucleotide and amino acid sequences are available from the authors upon request.

Abbreviations: bp = base pair, g = gene, gp = gene product, kb = kilobase pairs, orf = open reading frame.

*Present address: W. Alton Jones Cell Science Center
Lake Placid, New York, N.Y.12946, USA

+To whom correspondence should be addressed

REFERENCES

1. Gram,H. and Rüger,W. (1985) EMBO J. 4, 257-264.
2. Wyatt,G.R., and Cohen,S.S. (1953) Biochem. J. 55, 774-782.
3. Lehman,I.R. and Pratt,E.A. (1960) J. Biol Chem. 235, 3254-3258.
4. de Waard,A., Ubbink,T.E.C.M., and Beukman,W. (1967) Eur. J. Biochem. 2, 303-308.
5. Lunt,M.R., and Newton,E.A. (1965) Biochem. J. 95, 717-723.
6. McNicol,L.A., and Goldberg,E.B. (1973) J. Mol. Biol. 76, 285-301.
7. Revel,H.R. (1983) In "Bacteriophage T4" Mathews,C.K., Kutter,E.M., Mosig,G. and Berget,P.B. (eds) pp 156-165, ASM Publications.
8. Cox,G.S. and Conway,T.W. (1973) J. Virol. 12, 1279-1287.
9. Dharmalingam,K. and Goldberg,E.B. (1979) Virology 96, 393-403.
10. Rüger,W. (1978) Eur. J. Biochem. 88, 109-117.
11. Josse,J. and Kornberg, A. (1962) J. Biol. Chem. 237, 1968-1975.
12. Georgopoulos,C.P. (1968) Virology, 34, 364-366.
13. Russell,R.L. (1974) Genetics 78, 967-988.
14. Wood,W.B., and Revel,H.R.(1976) Bacteriol. Rev., 40, 847-868.
15. Wilson,G.G., and Murray,N.E. (1979) J. Mol. Biol. 132, 471-491.
16. Murray,N.E., Bruce,S.A., and Murray,K. (1979) J. Mol. Biol. 132, 493-505.
17. Knopf,K.W. (1977) Eur. J. Biochem. 73, 33-38.
18. Tait,R.C., Rodriguez,L., and West,R.W. (1980) J. Biol. Chem. 255, 813-815.
19. Niggemann,E.,Green,I.,Meyer,H.P., and Rüger,W. (1981) Mol. Gen. Genet.,184, 289 - 299.
20. Hattman,S. and Fukasawa,T. (1963) Proc. Natl. Acad. Sci. USA 50, 297-300.
21. Shedlovsky,A., and Brenner,S. (1963) Proc. Natl. Acad. Sci. USA 50, 300-305.
22. Sutcliffe,J.G. (1978) Nucleic Acids Res. 5, 2721-2728.
23. Remaut,E., Tsao,H., and Fiers,W. (1983) Gene 22, 103-113.
24. Deininger,P.L. (1983) Anal. Biochem., 129, 216-223.

25. Sanger,F., Nicklen,S., and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463-5467.
26. Biggin,M.D., Gibson,T.J., and Hong,G.F. (1983) Proc. Natl. Acad. Sci. USA, 80, 3963-3965.
27. Messing,J.,Crea,R., and Seeburg,P.H.(1981) Nucleic Acids Res.,9, 309-321.
28. Dagert,M. and Ehrlich,S.D. (1979) Gene 6, 23-28.
29. Hanahan,D.(1983) J. Mol. Biol., 166,557-580.
30. Brown,D.M., Frampton,J., Goelet,P., and Karn,J. (1982) Gene, 20, 139-144.
31. Benton,W.D. and Davies,R.W. (1977) Science 196, 180-192.
32. Hu,N. and Messing,J.(1982) Gene, 17,271-277.
33. Staden,R. (1980) Nucl. Acid. Res.9, 3673 - 3694.
34. Herrick,R.M., Hunkapiller,M.W., Hood,L.E. and Dreyer,W.I.(1981) J. Biol. Chem. 256, 7990-7997.
35. Lai,P.-H. (1985) Analytica Chemica Acta, in press
36. Hunkapiller,M.W. and Hood,L.E.(1983) Methods Enzymol. 91, 485-493.
37. Tarr,G.E. (1985) In Microcharacterization of Polypeptides: A Manual, Shirely,J.E. ed., Humana Press, New Jersey, in press.
38. Dayhoff,M.O., Barker,W.C., and Hunt,L.T. (1983) Methods Enzymol. 91, 524-545.
39. Birnboim,H.C. and Doly,J. (1979) Nucl. Acids Res. 7, 1513-1523.
40. Southern,E.M. (1975) J. Mol. Biol. 98, 503-517.
41. Kutter,E. and Rüger,W. (1983) In "Bacteriophage T4", Mathews, C.K., Kutter,E.M., Mosig,G., Berget,P.B. (eds) pp 277 - 290, ASM Publications.
42. Rüger,W., and Kutter,E. (1984) In "Genetic Maps/1984" Vol 3, O'Brian,S.J. (ed) Laboratory of Viral Carcinogenesis, Natl. Cancer Institute, pp 28-34, Cold Spring Harbor.
43. Fuhrman,S.A., Deininger,P.L., LaPorte,P., Friedmann,T., and Geiduschek,E.P. (1981) Nucleic Acid Res. 9, 6439-6456.
44. Rand,K.N., and Gait,M.J. (1984) EMBO J. 3, 397-402.
45. Mattson,T., van Houwe,G., Bolle,A., Selzer,G., and Epstein,R. (1977) Mol. Gen. Genet. 154, 319-326.
46. Capaldi,R.A. and Vanderkooi,G. (1972) Proc. Natl Acad. Sci. USA 69, 930-932.
47. Chou,P.Y. and Fasman,G.D. (1978) Adv. Enzymol. 47, 45-148.
48. Gram and Rüger, in preparation
49. Grosjean,H. and Fiers,W. (1982) Gene 18, 199-209.
50. Remaut,E., Stanssens,P., amd Fiers,W. (1981) Gene 15, 81-93.
51. Brody,E., Rabussay,D., and Hall,D.H. (1983) In "Bacteriophage T4". Mathews,C.K., Kutter,E.M., Mosig,G. and Berget,P.B.(eds) pp 174-183, ASM Publications.
52. Spicer,E.K., Nossal,N.G., and Williams,K.R. (1984) J. Biol. Chem. 259, 15425-15432.
53. Shine,J., and Dalgarno,L.(1974) Proc. Natl. Acad. Sci. USA, 71, 1342-1346.
54. Holmes,W.M., Platt,T., and Rosenberg,M. (1983) Cell 32, 1029-1032.
55. Gram,H., Liebig,H.D., Hack,A., Niggemann,E., and Rüger,W. (1984) Mol. Gen. Genet., 194, 232 - 240.
56. Baugham,G. and Nomura,M. (1983) Cell 34, 979-988.
57. Kyte,J. and Doolittle,R.F. (1982) J. Mol. Biol., 157, 105-132.
58. Allen,J.R., Lasser,G.W., Goldman,D.A., Booth,J.W., and Mathews,C.K. (1983) J. Biol. Chem. 258, 5746-5753.
59. Burgess,A.W., Ponnuswamy,P.K., and Sheraga,H.A. (1974) Israel J. Chem. 12, 239-286.
60. Chou,P.Y. and Fasman,G.D. (1979) Biopphys. J. 26, 367-384.
61. Nagano,K. (1977) J. Mol. Biol. 109, 251-274.
62. Robson,B. and Suzuki,E. (1976) J. Mol. Biol. 107, 327-356.

63. Rawlings,N., Ashman,K., and Wittmann-Liebold,B. (1983) Int. J. Pept. Prot. Res. 22, 515-524.
64. Rao,S.T., and Rossmann, M.G. (1973) J. Mol. Biol. 76, 241-256.
65. Kabsch,W. and Sander,C. (1983) Febs Lett. 155, 179-181.
66. Kabsch,W. and Sander,C. (1984) Proc. Natl. Acad. Sci. USA 81, 1075-1078.
67. Laemmli,U.K. (1970) Nature 227, 680-685.
68. Fleischman,R.A., Campbell,J.L., and Richardson,C.C. (1976) J. Biol. Chem. 251, 1561-1570.
69. Fukada,K. and Abelson,J. (1980) J. Mol. Biol. 139, 377-391.
70. Guthrie,C. and McClain,W.H. (1979) Biochem. 18, 3786-3795.