

$T_1$  oligonucleotide on the 3'-end of the DNA sequence was not recovered, nor was a seven nucleotide-long  $T_1$  oligonucleotide on the 5'-end of the DNA sequence. Both of these  $T_1$  oligonucleotides contain extensive polypyrimidine tracts which might explain the poor recovery of these oligonucleotides in the conditions of pancreatic RNase digestion used here to isolate ribosome protected fragments.

We have carefully compared the sequences of these four ribosome binding sites with the six potential ribosome binding sites described by Barrell *et al.*<sup>12</sup> but are as yet unable to explain why only four are recovered in this *in vitro* system. Those which are recovered allow us to define, in conjunction with the accompanying paper and earlier studies<sup>4-6,12</sup>, additional functional regions of the  $\Phi X174$  DNA sequence.

We thank Bart Barrell, Nigel Brown and Mike Smith for generously providing DNA sequencing data in advance of publication. We further thank Bart Barrell for helpful suggestions, and George Pieczenik and Norton Zinder for encouragement, advice and suggestions. We thank Edward Pelle and Carolyn McGill for technical assistance. This research was supported

in part by grants from the US NFS and the US National Institutes of Health.

Received November 25; accepted December 24, 1976.

- 1 Denhardt, D. T. *CRC Critical Rev. Microbiol.* **4**, 161-223 (1975).
- 2 Burgess, A. B. and Denhardt, D. T. *J. molec. Biol.* **44**, 377-386.
- 3 Godson, G. N. *J. molec. Biol.* **57**, 541-553 (1971).
- 4 Robertson, H. D., Barrell, B. G., Weith, H. L., Donelson, J. E., *Nature new Biol.* **241**, 38-40 (1973).
- 5 Robertson, H. D. *J. molec. Biol.* **92**, 363-375 (1975).
- 6 Barrell, B. G., Weith, H. L., Donelson, J. E. & Robertson, H. D. *J. molec. Biol.* **92**, 377-393 (1975).
- 7 Peiczenik, G., Model, P. & Robertson, H. D. *J. molec. Biol.* **90**, 191-214 (1974).
- 8 Kondo, M., Eggerston, G., Eisenstadt, J. & Lengyel, P. *Nature* **220**, 368-371 (1968).
- 9 Steitz, J. A. *Nature* **224**, 957-964 (1969).
- 10 Hindley, J. & Staples, D. H. *Nature* **224**, 964-967 (1969).
- 11 Brown, N. & Smith, M. *Nature* **265**, 695-698 (1977).
- 12 Barrell, B. G., Air, G. H. & Hutchinson, C. A. III *Nature* **264**, 34-41 (1976).
- 13 Brownlee, G. G. & Sanger, F. *Eur. J. Biochem.* **21**, 395-399 (1969).
- 14 Air, G. M. *et al.* *J. molec. Biol.* (in the press).
- 15 Air, G. M. & Bridgen, J. *Nature new Biol.* **241**, 40-41 (1973).
- 16 Gelfand, D. H. & Hayashi, M. *Proc. natn. Acad. Sci. U.S.A.* **63**, 135-137 (1969).
- 17 Shine, J. & Dalgarno, L. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1342-1346 (1974).
- 18 Steitz, J. A. & Jakes, K. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4734-4738 (1975).
- 19 Ravetch, J. V. & Zinder, N. D. *Nature* **262**, 150-153 (1976).
- 20 Platt, T. & Yanofsky, C. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2399-2403 (1975).
- 21 Model, P. & Zinder, N. D. *J. molec. Biol.* **83**, 231-251 (1974).
- 22 Barrell, B. G. in *Procedures in Nucleic Acid Research* (eds Cantoni, G. L. & Davies, D. R.) 2, 751-779 (Harper and Row, New York).

## DNA sequence at the C termini of the overlapping genes *A* and *B* in bacteriophage $\Phi X174$

M. Smith, N. L. Brown, G. M. Air, B. G. Barrell, A. R. Coulson, C. A. Hutchison III & F. Sanger

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

*Sequence determination of mutants in genes A and B confirms that these constitute another pair of genes in  $\Phi X174$  which are translated from the same DNA sequence using different reading frames.*

BACTERIOPHAGE  $\Phi X174$  is known to code for nine proteins in its DNA (ref. 1) which is about 5,400 nucleotides in length (Fig. 1). The locations of the boundaries of six of the genes—*D*, *E*, *J*, *F*, *G* and *H*—have been defined by sequence determination (F.S. *et al.*, in preparation). Two of these genes (*D* and *E*) overlap on the DNA sequence<sup>2</sup>. To define the boundaries for genes *A*, *B*

and *C* which lie in the remaining region between gene *H* and gene *D*, the sequence of  $\Phi X174$  am3 DNA in the region between the *HaeIII* 6a/6b and the *TaqI* 2/7 sites (Fig. 1) has been determined (N.L.B. and M.S., in preparation). A newly recognised ribosome binding site<sup>3</sup> has been located in the sequence<sup>4</sup>. The sequence determination of a *B* gene mutant—*am16*—strongly suggested that the *A* and *B* genes are translated from the same DNA using different reading frames (N.L.B. and M.S., in preparation). We now report the 200 nucleotide sequence of  $\Phi X174$  am3 DNA from the *Taq* 2/7 to the *Hha* 8b/4 sites (Fig. 1). In addition, the sequences of three mutants from this region—*am18* and *am35* (gene *A*) and a revertant of *am18*, *ts116* (gene *B*)—have been determined. The mutants *am18* and *am35* have an

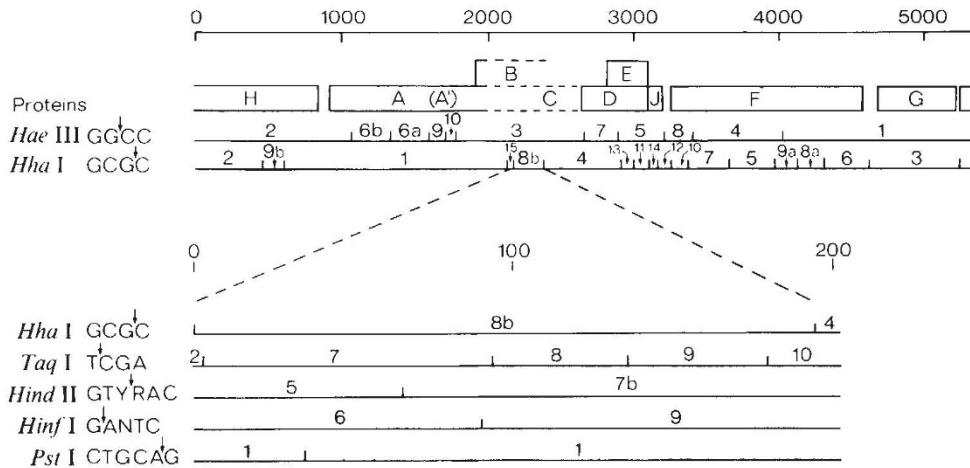


Fig. 1 Map of the  $\Phi X174$  genome showing the *HaeIII* and *HhaI* fragments and the protein-coding sequences and details of other restriction cleavages within *Hha* fragment 8b.

amber codon at the same site. The gene *B* mutant, *ts116*, results from a single base change in the termination codon of *am18/35*. These results confirm the overlap of genes *A* and *B* and establish the C-termini of the two genes.

### Nucleotide sequence analysis

The cleavage sites of five restriction enzymes which cleave in the region are shown in Fig. 1. The sequence of the DNA was obtained by a slight modification of the 'plus and minus' method (ref. 5 and N.L.B. and M.S., in preparation) by priming from the *Hinf*I 6/9 site, on both viral and complementary templates and from the *Pst*I site on the complementary template (Fig. 2). The sequence was confirmed by depurination analysis<sup>6</sup> of *Taq*I fragments 7, 8, 9 and 10 labelled by 'nick translation'<sup>7</sup> of  $\Phi X 174$  *am3* RFI DNA.

### Nucleotide sequence of *A* and *B* mutants

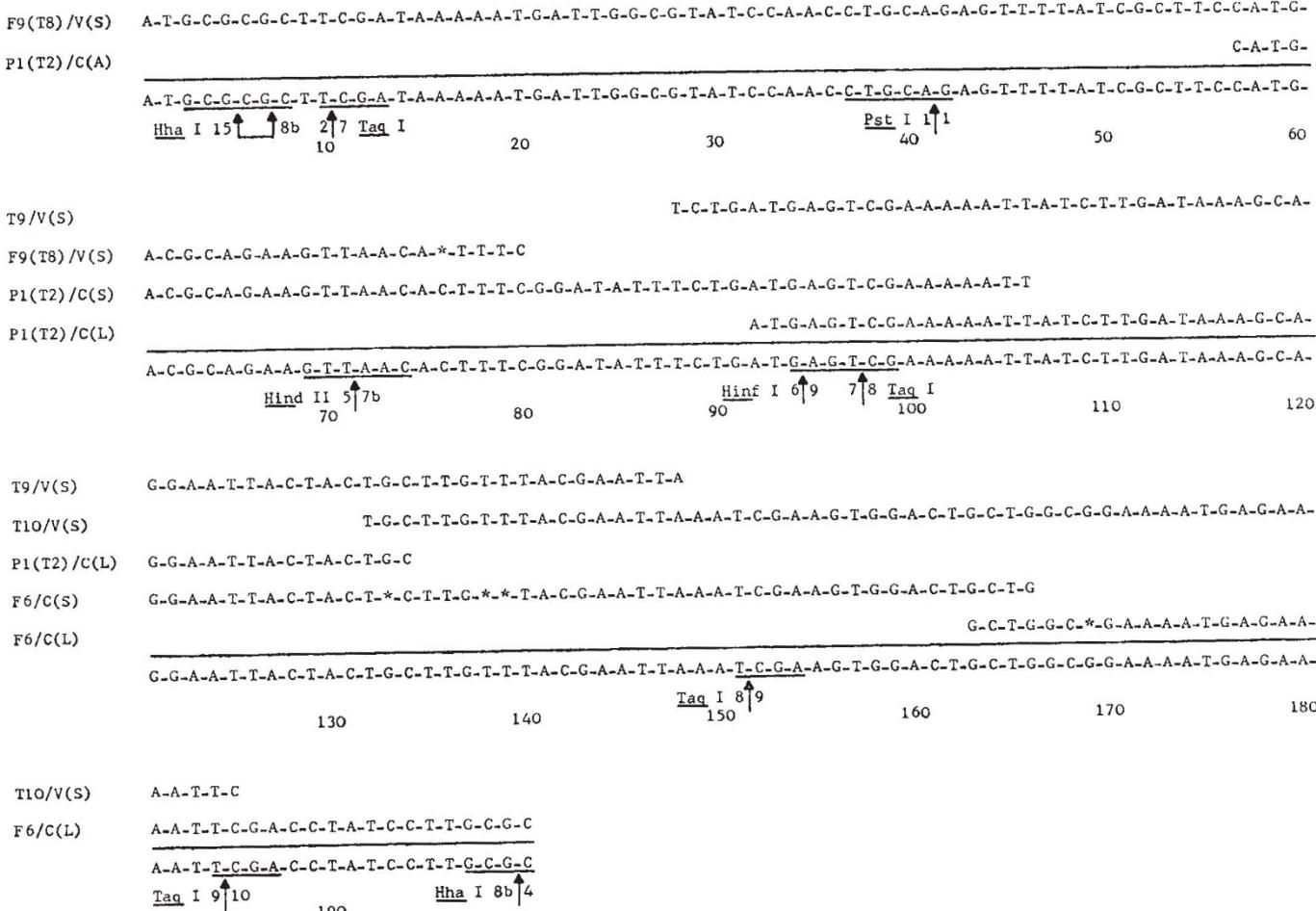
Gene *A* mutants (*am18*, *am35*) have been mapped<sup>8</sup> between the left end of *Hha*I fragment 8b (fragment 7b in ref. 8) and the right end of *Hind*II fragment 5. The mutant *am18* and its revertant *ts116* do not recombine to produce *wt* progeny at a frequency above background. These experiments set an upper limit of approximately three nucleotides on the distance separating these mutants<sup>9</sup>. The *B* gene mutant (*ts116*) maps close to, or in, the *Hind*II 5/7b site<sup>10</sup>. Consequently, the mutant sequences can be determined very easily using the plus and minus method by priming from the *Hinf*I 6/9 site on the mutant viral strand template. The results of these determinations for *am3* (*E* gene

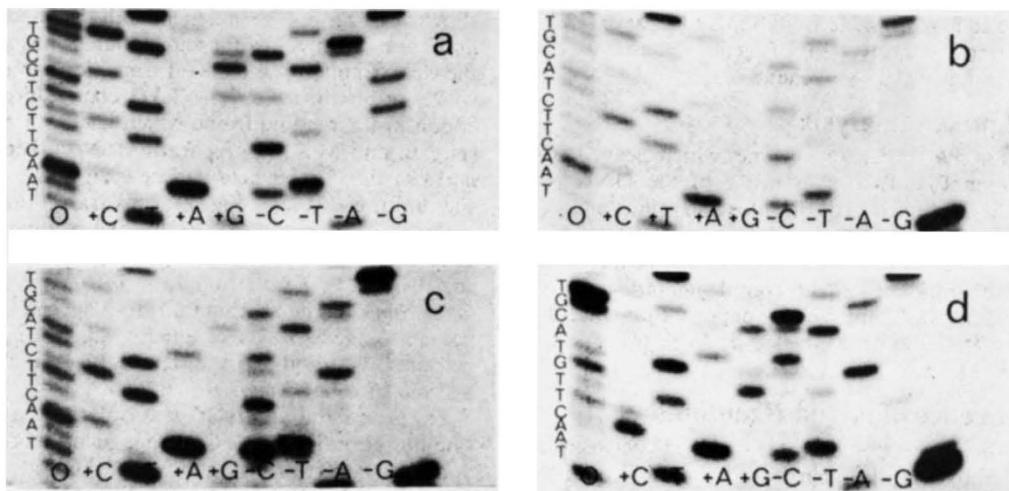
mutant; wild type in this region), *am18*, *am35* and *ts116* are shown in Fig. 3. The independently isolated *A* gene mutants, *am18* and *am35*, show the same altered nucleotide sequence. In the viral strand, this is a C → T transition at nucleotide 64 (Fig. 2) which results in a CAG → TAG change producing an amber codon in the reading frame which has been assigned to gene *A* (N.L.B. and M.S., in preparation). No other alterations in the *am18* or *am35* sequences can be seen between the *Hind*II 5/7b site and the *Pst*I site (Fig. 2). This transition also results in an amino acid change (Ala → Val) in the reading frame assigned to gene *B* on the basis of the phasing of the ribosome binding site<sup>3,4</sup> and the gene *B* mutant, *am16*. This assignment of the *am18*, *am35* (gene *A*) mutation to a specific site to the right of the *am16* (gene *B*) as originally suggested by recombination mapping<sup>1,9,11</sup>, and to a different reading frame, verifies the overlap of genes *A* and *B*.

The gene *B* mutant, *ts116*, was isolated as a revertant of *am18*. The mutation is a G → C change at nucleotide 66, which produces a TAG → TAC (amber → Tyr) codon change in the gene *A* reading frame. In gene *B*, the mutation results in a GAA → CAA (Glu → Gln) codon change. The resultant change in charge may be responsible for the *ts* phenotype.

The original stocks of these *am* mutants were obtained after nitrous acid treatment of viral DNA. The mutations characterised so far include G → A (*am3,34*; *amN11,27*) (ref. 2), C → T (*am18,35*) and G → T (*am16*). This variety of mutations illustrates the caution which should be exercised in predicting the consequences of chemical mutagenesis<sup>12</sup>.

**Fig. 2** Summary of the plus and minus sequence determination of the *Hha*I fragment 8b region. At the left of each determination is indicated the fragment used as the origin of the plus and minus priming. (The single letter designations used for restriction endonucleases are: *Taq*I (T); *Hinf*I (F); and *Pst*I, (P).) Fragments indicated in parentheses are the primers used when these differed from the one corresponding to the origin (for example, F9(T8) indicates that T8 was the primer with the origin at the adjacent *Hinf*I site). The letters V and C indicate that the template was viral or complementary strand DNA, respectively. The designations (S) or (L) indicate the use of a short (4 h) or long (8 h) gel electrophoresis run.





**Fig. 3** Sequences of the *am18*, *am35* and *ts116* mutations. The *am3* (wild type in this region) sequence is also shown. *TagI* fragment 8 was used as primer with *HinfI* as the datum restriction endonuclease with either (a) *am3*, (b) *am18*, (c) *am35* or (d) *ts116* viral DNA as template. The same portion of the autoradiogram of each plus and minus determination is shown with the sequence of the complementary strand deduced. The viral strand sequences, positions 60–74 (Fig. 2), are therefore:

<i>am3</i> ('wt')	G-A-C-G-C-A-G-A-A-G-T-T-A-A-C . . , . , . , . , .
<i>am18</i> (A)	G-A-C-G-T-A-G-A-A-G-T-T-A-A-C . . , . , . , . , .
<i>am35</i> (A)	G-A-C-G-T-A-G-A-A-G-T-T-A-A-C . . , . , . , . , .
<i>ts116</i> (B)	G-A-C-G-T-A-C-A-A-G-T-T-A-A-C . . , . , . , . , .

Dots indicate the reading frame for gene *A* and commas that for gene *B*.

B protein	- Tyr - Ala - Arg - Phe - Asp - Lys - Asn - Asp - Trp - Arg - Ile - Gln - Pro - Ala - Glu - Phe - Tyr - Arg - Phe - His -
A protein	- Met - Arg - Ala - Ser - Ile - Lys - Met - Ile - Gly - Val - Ser - Asn - Leu - Gln - Ser - Phe - Ile - Ala - Ser - Met -
	A-T-G-C-G-C-G-C-T-T-C-G-A-T-A-A-A-A-T-G-A-T-T-G-G-C-G-T-A-T-C-C-A-A-C-C-T-G-C-A-G-A-G-T-T-T-A-T-C-G-C-T-T-C-C-A-T-G-
	Hha I/15      /8b    2/7    Tag I
	10                20                30                40                50                60
	Pst I 1/1

B protein      Asp - Ala - Glu - Val - Asn - Thr - Phe - Gly - Tyr - Phe

↑                    ↑

(Val)    Gln

The diagram shows the amino acid sequence of the A protein. Key features include:

- Termination Codons:** Tyr (at position 1), Ter (at position 18), B stop (at position 90), and T am6 (at position 100).
- Restriction Enzyme Sites:**
  - Hind II:** Cleavage site at position 70.
  - Hinf I:** Cleavage site at position 90.
  - Tag I:** Cleavage site at position 100.
- Start Codon:** TGA (at position 1).
- Other Labels:** am18, am35, 5/7b, C ts116, and positions 70, 80, 90, 100, 110, 120.

A protein sequence diagram showing positions 130 to 180. The sequence is: Gly - Ile - Thr - Thr - Ala - Cys - Leu - Arg - Ile - Lys - Ser - Lys - Trp - Thr - Ala - Gly - Gly - Lys. Below the sequence, the nucleotide sequence is given: G-G-A-A-T-T-A-C-T-A-C-T-G-C-T-T-G-T-T-A-C-G-A-A-T-T-A-A-A-A-T-C-G-A-A-G-T-G-G-A-C-T-G-C-T-G-G-C-G-A-A-A-A-T-G-A-G-A-A-. Vertical arrows indicate positions 130, 140, 150, 160, 170, and 180. A double-headed arrow between positions 150 and 160 is labeled "8/9 Taq I". An upward arrow from position 159 points to the word "Ter". A downward arrow from position 159 points to the label "A am6". Another downward arrow from position 159 points to the label "A stop".

**Fig. 4** The sequence of  $\Phi X 174$  viral DNA in the region of gene *B* and the C terminus of gene *A*. The amino acid sequences of the gene *A* and gene *B* proteins predicted from the DNA sequence are shown together with the changes resulting from the mutations *am18* and the subsequent *ts116*, *am35* and *am6*.

The changes induced in the second gene by mutation in its overlapping companion are not necessarily restrictive, as evidenced by the Leu → Phe change in gene *A* which accompanies the *am16* mutation in gene *B*, the Ala → Val change in gene *B* accompanying the *am18,35* mutation in gene *A*, and the reversion of *am18*, which by two steps induces a Gln → Tyr in gene *A*. There is no evidence for any impairment in protein function due to these changes, whereas in *ts116* the second amino acid change (Glu → Gln) in gene *B* which follows the initial change (Ala → Val) induced by the gene *A* amber mutation evinced a phenotypic response (*ts*).

### C termini of A and B proteins

The sequence determination and establishment of the *A* gene reading frame taken together with the *B* gene reading frame allow the prediction of the C termini of the two proteins (Fig. 4). The *B* protein is 120 amino acids long with a molecular weight of 13,845. The *A* gene extends for 85 nucleotides (28 codons) beyond the C terminus of gene *B*. If, as has been suggested (N.L.B. and M.S., in preparation), the gene *A*/gene *B* overlap arose by read-through of a previously shorter gene *A*, then a question arises about the original role of the DNA beyond the C terminus of gene *B*. The mutant *am6* is a double amber mutant mapping in fragments F9 (gene *A*) and Z7 (gene *E*) (P. Weisbeek, personal communication). In gene *E*

the mutation is identical to *am3*. The gene *A* mutation has also been sequenced and is a G → A change at nucleotide 158. This gives an amber (TAG) codon, and would result in a gene *A* product six amino acids shorter than in the wild type. In addition there is a C → T change at nucleotide 98 which would result in a missense Ser → Leu in the gene *A* protein, and preliminary results suggest yet another change in this region. The resulting gene product is impaired in function, indicating that there is some requirement for the region of the *A* protein coded by the DNA beyond the end of gene *B*.

We thank R. Kamen, R. J. Roberts and S. Sato for gifts of some of the enzymes used in this study. The *am18* and *am35* stock were kindly provided by H. S. Jansz.

Received November 25; accepted December 24, 1976.

- 1 Benbow, R. M., Mayal, R. F., Picchi, J. C. & Sinsheimer, R. L. *J. Virol.* **10**, 99–114 (1972).
- 2 Barrell, B. G., Air, G. M. & Hutchison, C. A. III *Nature* **264**, 34–41 (1976).
- 3 Ravetch, J., Model, P. & Robertson, H. D. *Nature* **265**, 698–702 (1977).
- 4 Brown, N. L. & Smith, M. *Nature* **265**, 695–698 (1977).
- 5 Sanger, F. & Coulson, A. R. *J. molec. Biol.* **94**, 441–448 (1975).
- 6 Ling, V. *J. molec. Biol.* **64**, 87–102 (1972).
- 7 Maniatis, T., Jeffrey, A. & Kleid, D. G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1184–1188 (1975).
- 8 Baas, P. D., van Heusden, G. P. H., Vereijken, J. M., Weisbeek, P. J. & Jansz, H. S. *Nucl. Acids Res.* **3**, 1947–1960 (1976).
- 9 Hutchison, C. A. III thesis, California Institute of Technology (1969).
- 10 Weisbeek, P. J., Vereijken, J. M., Baas, P. D., Jansz, H. S. & van Arkel, G. G. *Virology* **72**, 61–71 (1976).
- 11 Sinsheimer, R. L. *Prog. Nucl. Acid Res. molec. Biol.* **8**, 115–169 (1968).
- 12 Vanderbilt, A. S. & Tessman, I. *Genetics* **66**, 1–10 (1970).

# Photochemical attachment of cyclic AMP binding protein(s) to the nuclear genome

John Kallos\*

Department of Biochemistry, College of Medicine and Dentistry of New Jersey, New Jersey Medical School, Newark, New Jersey 07103

*The nature of the cyclic AMP-receptor-nucleus interactions was examined by a novel combination of two photoreactions. A photosensitive derivative of cyclic AMP, N<sup>6</sup>-butyryl cyclic AMP, was covalently attached to its cytoplasmic receptor by photo-affinity labelling and this receptor complex was photo-crosslinked by the DNA in the rat liver nuclei. The photolytic reactions seemed to be specific since stable links were formed only when substantial noncovalent binding occurred.*

CLOSE similarity between two modes of signalling to the nucleus in mammalian cell regulation, one by steroid hormones<sup>1–5</sup> and the other by cyclic nucleotides<sup>6–10</sup>, has been suggested. Both steroid hormones and cyclic AMP combine specifically with a cytoplasmic receptor protein, and then the resulting ligand receptor complex enters into the nucleus, where it binds to the genome. Little is known of the nature of these hormone-receptor-genome interactions. A requirement for the solution of this problem is the identification of the residues involved in the protein-DNA interactions in the nucleus.

I intend to study the chemistry of the ligand-receptor and receptor genome interactions and to freeze the *in situ* interacting residues by the covalent attachment of the ligand to the receptor and the receptor complex to the genome in the intact nucleus, so that these residues can be

identified. I report here the combined use of two well-known photolabelling techniques. In one, photo-affinity labelling<sup>11–13</sup>, the ligand derivative is covalently attached to the receptor protein to identify the ligand binding site. In the other, the photo-induced crosslinking reaction<sup>11,14–19</sup>, the protein and nucleic acid are linked together to locate the segments of the macromolecules that are in close contact. This method of 'photo-affinity crosslinking' is based on the ability of the mobile cytoplasmic receptor to bind its ligand specifically and then to migrate to the nuclear site, thus delivering a photosensitive probe to form a ternary nuclear complex.

Photo-affinity crosslinking was used to study the covalent attachment of cyclic AMP to its 'receptor' and the cross-linking of the receptor complex to DNA in the intact nucleus. (In this paper the term 'receptor' is used to signify a high affinity cyclic AMP binding site.) The previously described cyclic-AMP-receptor-nucleus *in vitro* system from rat liver was used<sup>10</sup>. The N<sup>6</sup>-butyryl derivative of cyclic AMP (NBcAMP)<sup>20–21</sup> was chosen for photo-affinity labelling because the carbonyl group on the acyl side chain can be electronically excited by irradiation to a chemically reactive species<sup>22</sup>. This excitation can produce a variety of free radicals from the rupture of the bond in the  $\alpha$  position to the carbonyl ( $\alpha$  cleavage) or a 1,5 hydrogen shift of the  $\gamma$ -hydrogen (hydrogen abstraction) and leads to the covalent bond formation between the ligand and the receptor<sup>11,22</sup>.

Preliminary photo-affinity labelling studies of the cytoplasmic receptor with NBcAMP encouraged us to explore the applicability of receptor site labelling in a more complex system, the cell nucleus. It was expected that NBcAMP would bind specifically to its cytoplasmic receptor proteins,

\*Present address: Research Institute, Hospital for Joint Diseases, 1919 Madison Avenue, New York, N.Y. 10035