Eur. J. Biochem. *161*, 727–731 (1986)
© FEBS 1986

Cooperative DNA binding by lambda integration protein – a key component of specificity

Stephen J. MINTER [1], G. Marius CLORE [2], Angela M. GRONENBORN [2] and R. Wayne DAVIES [1,3]

[1] Department of Biochemistry and Applied Molecular Biology, University of Manchester Institute of Science and Technology
[2] Max-Planck Institut für Biochemie, Martinsried
[3] Allelix Inc., Missisauga, Ontario

(Received June 19, 1986) – EJB 86 0630

Quantitative analysis of nitrocellulose filter binding data by the method of Clore, Gronenborn and Davies [(1982) *J. Mol. Biol. 155*, 447–466] has been used to show that lambda integration protein (Int) exhibits cooperativity in binding to specific recognition sites within the attachment site region (λ*att*P) of bacteriophage lambda DNA. Optimal values of the equilibrium constant obtained were $3.0(\pm 1.0) \times 10^{10}$ M^{-1} for the P′ site using a model of three sites with equal affinity and $1.9(\pm 0.4) \times 10^{10}$ M^{-1} for the P1 site on a two-site model. The value of the cooperativity parameter α is 172^{+106}_{-66} in all cases. The occurrence of a consensus recognition sequence is necessary but not sufficient for strong binding; cooperative interaction between Int molecules binding to adjacent members of an array of binding sites is also essential. The occurrence of binding site arrays distinguishes λ*att*P very clearly from other DNA sequences containing single recognition sites by chance.

Specific recognition of regions of DNA by proteins underlies many important biological phenomena, particularly the regulation of gene expression at the transcriptional level. All proteins that interact with specific regions of DNA also show some degree of affinity for random DNA sequences ('non-specific' binding). Cooperative binding of the regulatory protein to an array of sites can raise the affinity for the correct region of DNA well above that for even a perfect recognition sequence that occurs elsewhere in the genome by chance, and the probability of the occurrence of arrays of recognition sequences by chance is very low. A clear example of the occurrence of multiple recognition sites within a functionally important region of DNA is that of the phage attachment site of bacteriophage lambda (λ*att*P) which contains multiple recognition sites for three proteins: the lambda integration protein (Int), the lambda excision protein, and the integration host factor. The lambda integration protein has been shown by a variety of experiments [1, 2] to be both the specificity determinant for recognition of λ*att*P and *att*B, and to carry out strand exchange in the site-specific recombination reaction by means of which the circularised genome of phage lambda is inserted into the chromosome of its host *Escherichia coli.* We have used the binding of Int to λ*att*P to investigate whether the occurrence of arrays of binding sites is correlated with cooperative interactions between molecules of the DNA-binding protein being important.

The possible importance of cooperative interactions between proteins in DNA-protein interaction has been recognized, but has not been formally demonstrated, and quantitative estimates of the importance of cooperativity have not been made. In the case of Int, it is generally accepted that Int binds to itself and other proteins, because Int has been shown to condense the whole of λ*att*P into a 14-nm-diameter particle [3]. It is assumed that these interactions occur with a strong cooperative component, but this has never actually been demonstrated. We have therefore applied the expressions derived by Clore et al. [4] to equilibrium binding curves obtained by nitrocellulose filter binding assays [5, 6] in which the binding of Int to λ*att*P, and to fragments of λ*att*P containing particular single recognition sites as defined by protection experiments [7] was measured. The results show that Int does exhibit cooperativity when binding to DNA regions containing closely spaced recognition sites, and that cooperativity contributes significantly to specific recognition of λ*att*P by Int.

MATERIALS AND METHODS

Preparation and radioactive labelling of DNA fragments

The 493-bp fragment containing the whole of λ*att*P was obtained by digestion of CsCl-gradient-purified pac129 [8] with *Hin*dIII and *Bam*HI. All DNA fragments were end-labelled by filling in recessed 3′ ends using the Klenow fragment of DNA polymerase I (Boehringer Corp.) as in [9]. After end-labelling, the fragments from the digestion were separated by electrophoresis in thin 8% polyacrylamide gels [10], the positions of the desired fragments determined by comparison with the positions of labelled fragments of known size from digests of pBR322 or M13mp7 RF DNA, and eluted from the gel by soaking for more than 12 h at 37 °C in 0.1 M NaCl,

Correspondence to G. M. Clore, Max-Planck Institut für Biochemie, Am Klopferspitz 18a, D-8033 Martinsried bei München, Federal Republic of Germany

Abbreviations. bp, base pairs; Int, lambda integration protein; λ*att*P, the attachment site of bacteriophage lambda that occurs within the bacteriophage DNA; P, left arm of λ*att*P; P, right arm of λ*att*P.

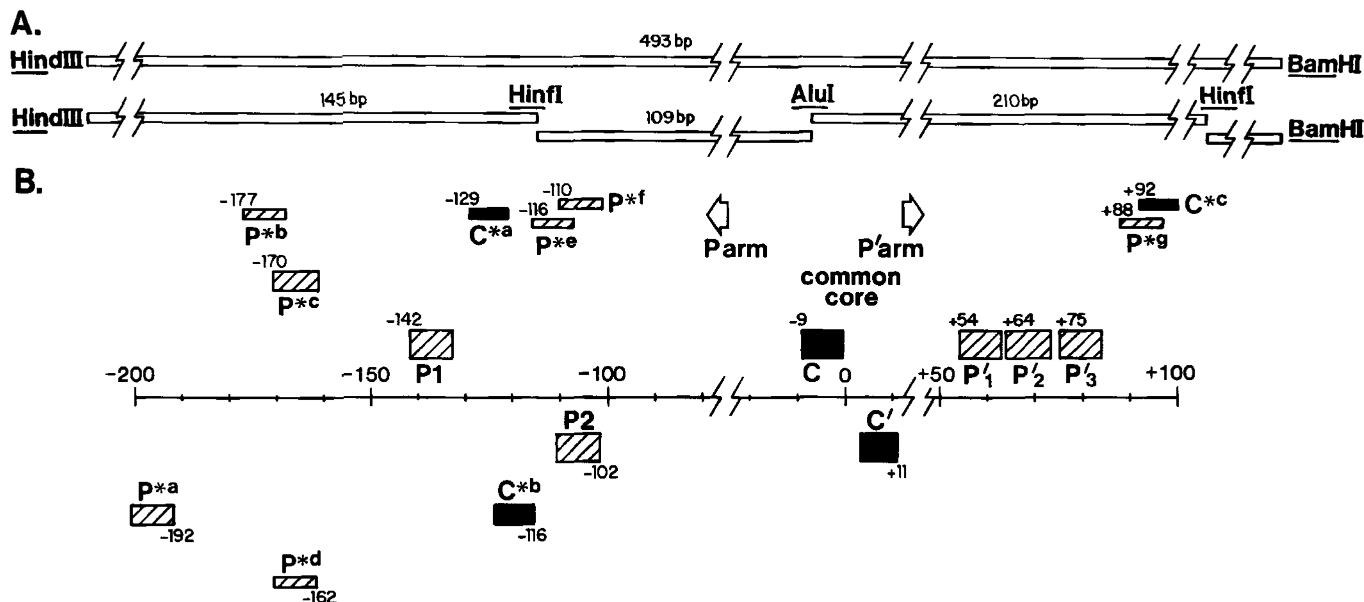

Fig. 1. *(A) Restriction fragments used in the nitrocellulose filter binding assay. (B) Location of known arm-type and core-type recognition sites for Int, and the location of sequences closely related to binding sites for Int, within λattP.* C* indicates a sequence resembling closely a core-type site, P* a sequence closely resembling an arm-type site. C, C′, P1, P2, P′1, P′2 and P′3 are previously known sites [17, 18] within the λ*att*P DNA sequence [23, 24]. Blocks indicating the previously known sites are shown closest to the centre line. The potential weak recognition sites (P^{Ea}, P^{*b} and C^{*b}) that resemble the known binding sites most closely are shown nearer to the centre line than the remainder; the position of the blocks with respect to the centre line is thus a rough indication of the likelihood that Int may bind to them. A base pair between C and C′ in the centre of the common core region is designed 0 by convention. The first 5′ end bases of all sequences are numbered according to this convention, − to the left (P arm), + to the right (P′ arm). The minimum functional λ*att*P includes P1 on the left and all of P′ on the right [7, 25]. Potential secondary sites were chosen after (a) a search with the FITCONSENSUS program using consensus tables assembled for core-type and arm-type sites by the CONSENSUS program, (b) a search for zero or one mismatch to the first (most conserved) 6 bp of either *P′*1, P′2, P′2, or P1 (or to C or C′) using the MAP program, (c) searches for matches at various levels of stringency to the set of sequences CAGTCA AAGTCA, AGGTCA, CGGTCA, TAGTCA, NATCANNAT, NANNCANNAT using the MAP program. Detailed descriptions of the programs and their output are available with the University of Wisconsin Genetics Computer Group sequence analysis package. A Digital Equipment Corporation VAX 750 computer was used

10 mM Tris/HCl pH 8.0, 1 mM EDTA. All fragments were further purified by binding to a small (0.2-ml-volume) DEAE-cellulose column, washing extensively with gel-elution buffer, collecting the DNA in a 1-ml block of 1 M NaCl, 10 mM Tris/HCl pH 8.0, 1 mM EDTA, and precipitating with ethanol. A 145-bp *Hin*dIII-*Hin*fI fragment containing the P1 site, a 109-bp *Hin*fI-*Alu*I fragment containing the P2 site and a 210-bp *Alu*I-*Hin*fI fragment containing the P′ and common core C′ sites were purified from digests of the 493-bp fragment with both *Hin*fI and *Alu*I (Fig. 1).

Purification of Int protein

Int protein was purified from extracts of *E. coli* strain HN442, which harbours a plasmid that contains an *Eco*RI fragment of lambda DNA carrying the *att-int* regions from an *int*^{c} strain of lambda [11]. The purification procedure of Kikuchi and Nash [12] was followed with minor modifications. Int was more than 95% pure as judged by polyacrylamide gel electrophoresis. The amount of protein in Int preparations was measured by the dye binding assay of Bradford [13].

Nitrocellulose filter binding assay

Aliquots of end-labelled DNA fragments were incubated with different amounts of pure Int protein in 250 μl of a solution containing 25 mM Tris/HCl pH 7.5, 1 mM $MgCl_2$, 0.1 mM dithiothreitol, 0.02 mM EDTA and 70 mM KCl for 5 min at 4 °C before being passed slowly (1 ml/min) through a nitrocellulose filter (Millipore HAWP 02500) that had been soaked in the above solution for at least 1 h on ice. Filters were washed with 0.5 ml of the above buffer, then with 0.5 ml of this buffer containing 500 mM NaCl and finally with 0.5 ml of the incubation buffer again. Radioactivity bound to the filters was determined by scintillation counting following dissolution of the filter in Scintacount (Beckman).

Curve fitting

The equilibrium nitrocellulose filter binding data were fitted to approximate expressions derived on the basis of the theoretical framework developed by Clore et al. [4] using the FACSIMILE program [14] which employs Powell's [15] method of non-linear optimization as described by Clore [16].

RESULTS

Fig. 1 shows the regions of DNA within λ*att*P for which Int has been shown by protection experiments [7] to possess high affinity. The collection of binding sites is made up of two sets based on two distinct consensus sequences. Ross and Landy [17] distinguished 'core-type' sites at the borders of the 'common core' region that is found in both bacteriophage and bacterial *att* sites, and 'arm-type' sites that occur to the

Table 1. *Equilibrium binding parameters for the specific DNA binding of Int to λattP*
All the experimental data shown in Fig. 2 were fitted simultaneously using Eqns (1–4). K_1 and K_2 are the specific equilibrium binding constants for the sites on the 145-bp fragment, i.e. essentially the P1 arm-type site region. K_3 and K_4 are the specific equilibrium binding constants for the sites on the 210-bp fragment, i.e. essentially the P′ arm-type site. It should be noted that if the 210-bp fragment contains three specific sites of equal affinity, then the value of the equilibrium constant will be reduce from $\approx 5\times10^{10}$ to $\approx 3\times10^{10}$ M^{-1}

Parameter	Cooperative binding $\alpha = 0$	Non-cooperative binding ($\alpha \neq 1$)
$K_1 = K_2$ (M^{-1})	$1.9\,(\pm0.4)\times10^{10}$	$9.2\,(\pm0.8)\times10^{10}$
$K_3 = K_4$ (M^{-1})	$5.0\,(\pm1.0)\times10^{10}$	$2.7\,(\pm0.2)\times10^{11}$
α	172^{+106}_{-66}	
Overall SD of fit (%)	4.5	9.1
Distribution of residuals	random	systematic errors

left and right of the common core (C and C′ respectively [18]). Footprinting defined as 'arm-type' sites a near-consensus (P1) and a consensus (P2; in the opposite orientation to P1) sequence to the left of common core (Fig. 1), and a region to the right of common core (P′ site; Fig. 1) consisting of three tandem consensus or near-consensus sequences. The sequences of these sites and of derived consensus sequences are given in Table 1. The P′ arm-type site is still protected by Int in the presence of heparin, while all other sites are not [7, 19], showing that Int has the highest affinity for the P′ site.

The DNA fragments used in filter binding experiments are also given in Fig. 1. The 493-bp *Hin*dIII-*Bam*HI fragment contains the whole of λ*att*P. The 210-bp *Alu*I-*Hin*fI fragment contains the P′ arm-type site, the C′ core-type site at the right core-arm junction, and a portion of the left core-arm junction site C; this latter site is presumably non-functional. Since the P′ triple site is the site for which Int has by far the highest affinity, the filter-binding experiments with the 210-bp fragment essentially measure binding to the P′ site. The 145-bp *Hin*dIII-*Hin*fI fragment contains P1 only. The 109-bp *Hin*fI-*Alu*I fragment contains the P2 site and the left portion (13 bp) of the C site which is destroyed by the *Alu*I cut. The P2 site is close to the *Hin*fI site, but the whole of the sequence related to the arm-type consensus sequence and the whole of the region clearly protected by Int against DNase I and neocarzinostatin [7] is within the fragment. The radioactively labeled fragments detected by filter binding have filled-in *Hin*fI sites, so that the first base pair of the arm-type binding site is the sixth base pair from the end. Therefore the 109-bp *Hin*fI-*Alu*I fragment should contain the intact P2 site.

The nitrocellulose filter binding assay, as described in Materials and Methods, was used to investigate the binding of Int to λ*att*P and the various DNA fragments containing portions of λ*att*P. The equilibrium filter binding curves for the binding of Int to all the DNA fragments are given in Fig. 2. Visual inspection of the equilibrium curves shows that the affinity of Int for the 210-bp P′(+C′) fragment (L_{50} $\approx$1.4 pM) is almost identical to its affinity for λ*att*P (L_{50} $\approx$1.2 pM), that Int has an approximately threefold lower affinity for the isolated P1 site compared to P′ (L_{50} $\approx$3.8 pM), and at least a 100-fold lower affinity for the isolated P2 site.

To obtain further insight into the mode of Int binding, we analysed the data quantitatively. The quantity measured in

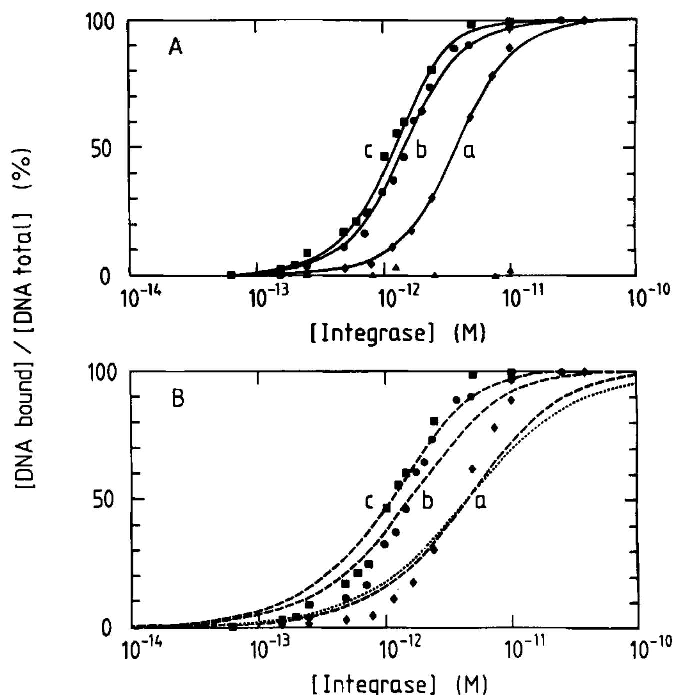

Fig. 2. *Comparison of the experimental equilibrium nitrocellulose filter binding curves for the binding of Int to the 493-bp (■), 210-bp (●), 145-bp (◆) and 109-bp (▲) fragments (Fig. 1) with the best fit theoretical curves for (A) cooperative (——) and (B) non-cooperative (– – – –) DNA binding.* The theoretical curves were calculated using Eqns (1–4) and the best-fit parameters are given in Table 1. A rectangular hyperbola given by $kL_F/(1 + KL_F)$ is shown for comparison in B as a dotted line (·······) with $K = 2.29\times10^{11}$ M^{-1}. The experimental conditions are: 70 mM KCl, 25 mM Tris/HCl; pH 7.5; 1 mM $MgCl_2$. The symbols are the means of the actual data points obtained from three separate experiments

the experiment is the fraction $\theta(L_F)$ of DNA to which at least one molecule of protein is bound as a function of the free protein concentration L_F [4]. $\theta(L_F)$ is given by

$$\theta(L_F) = [DNA]_{bound}/[DNA]_{total} = (Z-1)/Z \qquad (1)$$

where Z is the binding polynomial. For a single site, Z is given by $1 + kL_F$, and $\theta(L_F)$ is a simple rectangular hyperbola. In considering expressions of Z to fit the binding data, the contribution from non-specific DNA binding was neglected, as the experiment with the 109-bp *Hin*fI-*Alu*I fragment indicated that at concentrations of Int below 10 pM, this was insignificant.

Examination of Fig. 2B indicates that the binding curves for DNA fragments containing the whole of λ*att*P, the P′ + C′ sites, and the P1 site all deviate significantly from a rectangular hyperbola. Thus, even for the 145-bp fragment within which the only previously identified site is P1, a two-site binding model has to be considered. The simplest model that can, therefore, be put forward, is one in which the binding curves to the 145-bp and 210-bp fragments are each described by a two-site binding model with the binding curve for the 493-bp fragment described by the appropriate combination of the binding polynomials for the 145-bp and 210-bp fragments with no interaction between the sites on the 145-bp fragment on the one hand and the 210-bp fragment on the other. Thus, for such a model, Z is given by

$$Z = 1 + (K_1 + K_2)L_F + \alpha K_1 K_2 L_F^2 \qquad (2)$$

for the 145-bp fragment, by

$$Z = 1 + (K_3 + K_4)L_F + \alpha K_3 K_4 L_F^2 \quad (3)$$

for the 210-bp fragment, and by

$$\begin{aligned} Z = 1 &+ (K_1 + K_2 + K_3 + K_4)L_F \\ &+ (\alpha K_1 K_2 + \beta K_3 K_4 + K_1 K_3 + K_1 K_4 + K_2 K_3 + K_2 K_4)L_F^2 \\ &+ [\alpha K_1 K_2 (K_3 + K_4) + \beta K_3 K_4 (K_1 + K_2)] L_F^3 \\ &+ \alpha\beta K_1 K_2 K_3 K_4 L_F^4 \end{aligned} \quad (4)$$

for the 493-bp fragment, where K_1 and K_2 are the two specific equilibrium binding constants for the 145-bp fragment, K_3 and K_4 are the two specific equilibrium binding constants for the 210-bp fragment, and α and β are the cooperativity constants between the sites on the 145-bp and 210-bp fragments respectively. To simplify the model further, we take $K_1 = K_2$, $K_3 = K_4$ and $\alpha = \beta$. The binding curves for the 145-bp, 210-bp and 493-bp fragments were then fitted simultaneously to two models, namely cooperative ($\alpha \neq 1$) and non-cooperative ($\alpha = 1$) binding. The best fit curves for the cooperative model are shown in Fig. 2A and for the non-cooperative model in Fig. 2B. The SD of the fit for the cooperative model is 4.5%, which is within the estimated standard error of the data ($\approx 5\%$) and the distribution of residuals is random; in contrast, the SD of the fit for the non-cooperative model is approximately 9% and there are clear systematic differences between the experimental data and the computed curves. It should also be noted that the fit of the non-cooperative model cannot be improved by valuing $K_1 \neq K_2$ and $K_3 \neq K_4$, as this can only reduce the steepness of the curves. The optimized values of the parameters are given in Table 1.

DISCUSSION

The equilibrium filter binding curves for Int binding to λattP or the P′ recognition site clearly fit to a cooperative DNA-binding model, and not to a non-cooperative model. Thus we provide the first experimental evidence that λ Int does indeed, as long supposed by those in the field, bind to DNA in a cooperative fashion. The value of the cooperativity parameter α is 172^{+106}_{-66}, showing that cooperativity is quantitatively important for optimal binding of Int to its specific recognition sequence. This correlates with the fact that the P′ site is actually an array of three recognition sites in tandem. Davies et al. [20] showed that exonuclease III digestion stopped preferentially at three major sites within the P′ region when Int was present, and Hsu et al. [7] showed that deletions leaving only P_1' (Fig. 1B) intact are completely defective in recombination. Therefore the occurrence of the single perfect recognition site P_1' is insufficient to allow recombination to occur at normal intracellular Int concentrations, but two or three in tandem is dramatically better. Indeed, the complete triple P′ site is the strongest Int binding site, being the only site which is still detectable by footprinting in the presence of heparin [7, 19], and being protected against DNase I and methylation at lower concentrations of Int than all other sites [17]. Fig. 2 shows that the equilibrium filter binding curve for P′(+C′) is very close to that for complete λattP. The importance of cooperative interactions between Int molecules binding at P′ under physiological conditions is clear.

The curves were fitted to a two-site cooperative model. If P′ is taken as being functionally three recognition sites in tandem, then the value of the equilibrium constant of $\approx 5 \times 10^{10}$ M^{-1} for the two-site model is reduced to $\approx 3.0(\pm 1.0) \times 10^{10}$ M^{-1} for a three-site model. The value of the cooperativity parameter α would not be altered.

The equilibrium filter binding data obtained for the fragments defined as containing the Int recognition sites P1 or P2 require some explanation. Binding of Int to the fragment containing P1 is clearly cooperative. Since a single protein molecule bound to a DNA fragment results in its retention on a nitrocellulose filter, the cooperativity observed must be at the level of DNA binding, as for the P′ site. No sites strong enough to show up in protection experiments have been found in this fragment other than P1, but is is possible that weak binding of Int to poor recognition sites close to P1 might be sufficient to shift the equilibrium in favour of the bound state via cooperative interactions. Indeed, something of this sort must occur to give these data. Possible locations of weak binding sites found by extensive computer analysis are shown in Fig. 1B. The single P2 site, although it is a perfect recognition site, shows no significant Int binding in the concentration range used. This may reflect the equilibrium state in the complete absence of cooperativity, which would be consistent with the failure of P′ alone to provide enough Int binding for recombination. There are no weak Int recognition sites at all near P2 in this fragment. However, we cannot rule out the possibility that P2 is too near the end of the restriction fragment used to be recognised.

Analysis of the equilibrium filter binding data for λattP and P′(+C′) shows unequivocally the occurrence and the importance of cooperativity in the binding of Int to its specific recognition sequences. These data, when combined with the results of Hsu et al. [7] showing the insufficiency of a single site, demonstrate conclusively the importance of cooperative interactions between Int molecules binding to adjacent members of an array of recognition sites. When cooperative interactions contribute significantly, it is clear that the occurrence of recognition sites in arrays leads to very great discrimination between correct binding sites and other DNA sequences containing single recognition sites by chance. This principle may apply to the recognition of multiple binding sites by proteins controlling eukaryotic transcription, where arrays of recognition sites are often found [21, 22].

REFERENCES

1. Nash, H. A. (1981) *Annu. Rev. Genet. 15*, 143–167.
2. Craig, N. L. & Nash, H. A. (1983) *Cell 35*, 795–803.
3. Better, M., Lu, C., Williams, R. C. & Echols, H. (1982) *Proc. Natl Acad. Sci. USA 77*, 3220–3224.
4. Clore, G. M., Gronenborn, A. M. & Davies, R. W. (1982) *J. Mol. Biol. 155*, 447–466.
5. Riggs, A. D., Suzuki, H. & Bourgeois, S. (1970) *J. Mol. Biol. 48*, 67–83.
6. Hinkle, D. C. & Chamberlain, M. J. (1972) *J. Mol. Biol. 70*, 157–185.
7. Hsu, P.-L., Ross, W. & Landy, A. (1980) *Nature (Lond.) 285*, 85–91.
8. Brenner, S., Cesarini, G. & Korn, J. (1982) *Gene 17*, 27–44.
9. Davies, R. W. (1982) in *Gel electrophoresis of nucleic acids: a practical approach* (Rickwood, D. & Hames, B. D., eds) pp. 117–172, IRL press, Oxford and Washington DC.
10. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett. 7*, 107–110.
11. Honigman, A., Hu, S.-L. & Szybalski, W. (1979) *Virology 92*, 542–556.
12. Kikuchi, Y. & Nash, H. A. (1978) *J. Biol. Chem. 253*, 7149–7157.
13. Bradford, M. (1976) *Anal. Biochem. 72*, 248–254.

14. Chance, E. M., Curtis, A. R., Jones, I. P. & Kirby, C. R. (1979) *At. Energy Res. Estab. (G. Brit.) Rep.* R8775.
15. Powell, M. J. D. (1972) *At. Energy Res. Estab. (G. Brit.) Rep.* TP492.
16. Clore, G. M. (1983) in *Computing in biological science* (Geisow, M. J. & Barret, A. N., eds) pp. 313–348, Elsevier/North-Holland, New York.
17. Ross, W. & Landy, A. (1982) *Proc. Natl Acad. Sci. USA 79*, 7724–7728.
18. Ross, W. & Landy, A. (1983) *Cell 33*, 261–272.
19. Ross, W., Landy, A., Kikuchi, Y. & Nash, H. A. (1979) *Cell 18*, 297–307.
20. Davies, R. W., Schreier, P. H., Kotewicz, M. L. & Echols, H. (1979) *Nucleic Acids Res. 7*, 2255–2273.
21. Jones, K. A. & Tjian, R. (1984) *Cell 36*, 155–162.
22. Dynan, W. S. & Tjian, R. (1985) *Nature (Lond.) 316*, 774–778.
23. Davies, R. W., Schreier, P. H. & Büchel, D. E. (1977) *Nature (Lond.) 270*, 757–760.
24. Landy, A. & Ross, W. (1977) *Science (Wash. DC) 197*, 1147–1160.
25. Mizuuchi, M. & Mizuuchi, K. (1980) *Proc. Natl Acad. Sci. USA 77*, 3220–3224.