

A quantitative binding model for the Apl protein, the dual purpose recombination-directionality factor and lysis-lysogeny regulator of bacteriophage 186

Erin E. Cutts, J. Barry Egan, Ian B. Dodd and Keith E. Shearwin ^{*}

Department of Molecular and Biomedical Science, The University of Adelaide, Adelaide 5005, Australia

Received April 17, 2020; Revised July 21, 2020; Editorial Decision July 24, 2020; Accepted July 28, 2020

ABSTRACT

The Apl protein of bacteriophage 186 functions both as an excisionase and as a transcriptional regulator; binding to the phage attachment site (*att*), and also between the major early phage promoters (pR-pL). Like other recombination directionality factors (RDFs), Apl binding sites are direct repeats spaced one DNA helix turn apart. Here, we use *in vitro* binding studies with purified Apl and pR-pL DNA to show that Apl binds to multiple sites with high cooperativity, bends the DNA and spreads from specific binding sites into adjacent non-specific DNA; features that are shared with other RDFs. By analysing Apl's repression of pR and pL, and the effect of operator mutants *in vivo* with a simple mathematical model, we were able to extract estimates of binding energies for single specific and non-specific sites and for Apl cooperativity, revealing that Apl monomers bind to DNA with low sequence specificity but with strong cooperativity between immediate neighbours. This model fit was then independently validated with *in vitro* data. The model we employed here is a simple but powerful tool that enabled better understanding of the balance between binding affinity and cooperativity required for RDF function. A modelling approach such as this is broadly applicable to other systems.

INTRODUCTION

Apl is one of the large family of recombination directionality factors (RDFs) (1), that modulate the directionality of site-specific recombination reactions catalysed by the tyrosine integrase/recombinase proteins (2). RDFs are small proteins that bind to the DNA flanking the recombination site and, by altering the DNA architecture or by interacting with the integrase protein, facilitate the assembly of the integrase complex to promote excision of the prophage.

RDFs appear to share a mode of DNA binding in which protomers bind with high cooperativity in a head-to-tail manner to tandem DNA repeats spaced one DNA turn apart, shown for the archetypal RDF, Xis from λ (3), as well as for Gifsy-1 Xis (4) and Pukovnik Xis (5). The crystal structure of P2 Cox has been solved in the absence of DNA, revealing an extensive interaction with neighbouring Cox protomers (i+1) and also interactions with i+2 (6). Where examined, RDFs have been shown to cause large bends in attachment site (*att*) DNA (λ Xis (7–9), P2 Cox (10), L5 Xis (11), P4 Vis (12), W ϕ Cox (10), P22 Xis (13) and Pukovnik Xis (5)). A cryo-EM structure of the Holliday junction intermediate of the λ excisive complex (9) revealed three key roles for λ Xis in formation of the complex, including promoting integrase (Int) binding, mediating an Xis/Int interface, and bending of *att* DNA to position the DNA for cooperative Int binding. A crystal structure of λ Xis, showed three Xis monomers bound to the X1-X1.5-X2 sites in *attR* causing a 72° non-planar bend in the DNA, leading to the hypothesis that a twisted microfilament forms (14), a hypothesis supported by DNA compaction studies on P2 Cox (15). Another apparent common feature of RDFs is relaxed DNA specificity, with binding at non-canonical DNA sites seen *in vitro* at higher RDF concentrations. The crystal structure of lambda Xis–DNA complex also showed fewer sequence-specific contacts are made to the X1.5 site, compared with the X1 and X2 sites (14). Based on amino acid sequence, Apl is an outlier in the RDF family (1) but appears to fit this DNA binding pattern, being monomeric in solution (16) and binding to the 186 *att* site, along with the integrase protein (Int) and the integration host factor (IHF), to five direct repeats sequences with 10–11 bp spacing (17) (Figure 1).

A subset of RDFs also function as transcriptional regulators. In the KlpE (18) and P4 (12) prophages, the promoter for the integrase gene lies near the *att* site and is repressed by binding of the RDF to its sites within *att*. Such regulation is potentially widespread given the common proximity of *att* sites and *int* genes. Apl and other RDFs from P2-like bacteriophages and P4 also regulate transcription

^{*}To whom correspondence should be addressed. Tel: +61 8313 5361; Fax: +61 8313 4348; Email: keith.shearwin@adelaide.edu.au
Present address: Erin Cutts, Division of Structural Biology, Institute of Cancer Research, London, UK.

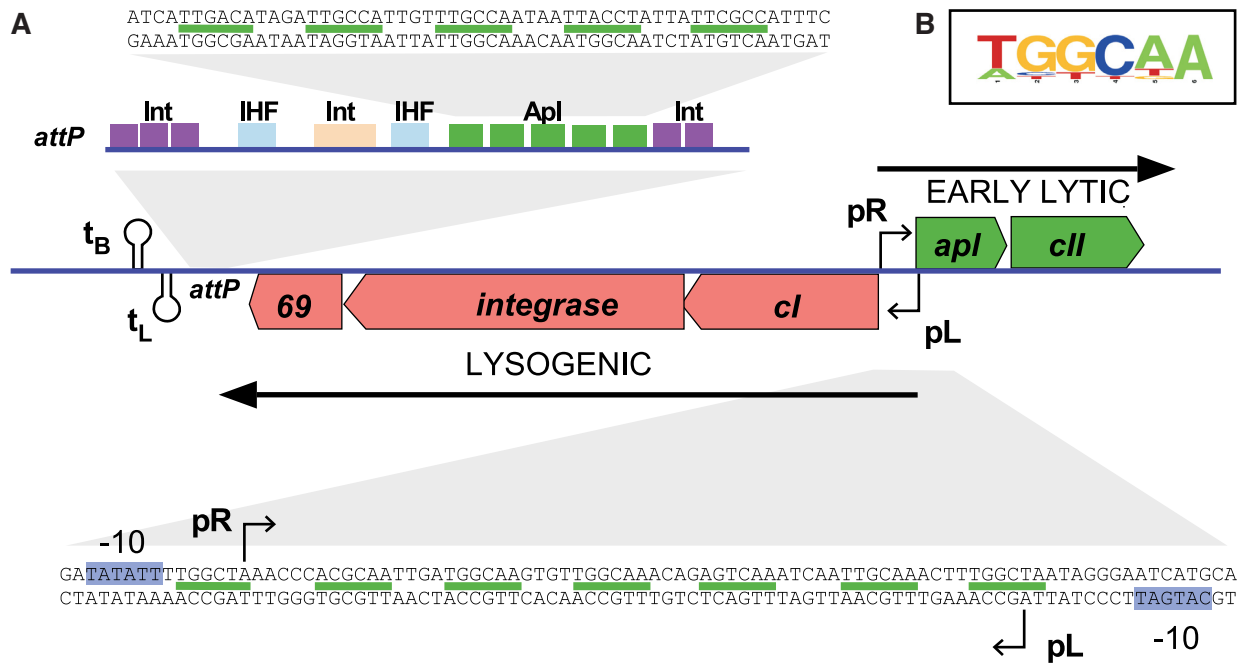


Figure 1. (A) Map indicating the location of Apl binding sites in bacteriophage 186. The centre of the diagram shows the switch region of 186, where the lytic pR promoter and the lysogenic pL promoter are arranged in a convergent orientation. There are seven Apl operators (indicated by green lines) located between the pR and pL promoters, and five Apl operators located at *attP*. Individual Apl operator sequences are six base pairs long, present in a direct repeat arrangement. At *attP*, the IHF sites are shown in blue and Integrase binding sites shown in tan (core site) and purple (arm sites). (B) Logo plot (47) representing the consensus of all Apl operator sequences from coliphage 186.

at locations well away from their attachment sites. These HTH-motif proteins are each encoded by the first gene of the phage early lytic operon and regulate the balance between lytic and lysogenic transcription, using recognition sequences of similar sequence and arrangement to the *att* sequences (19–21). In the P2-like phages, these RDF binding sequences typically lie between and overlapping the lytic and lysogenic promoters, which are arranged face-to-face and separated by 40–60 bp (22,23). In 186, the region between the pR lytic promoter and the pL lysogenic promoter contains seven Apl recognition sequences that, like the sites at *attP*, are direct repeats with 10–11 bp spacing (Figure 1). Apl binding represses both promoters (17); however, whilst Apl has a clear function at the *att* site as an excisionase (19), the function of its repressive activity at pR and pL is not well understood.

To better understand the mechanism of action and function of Apl, particularly with regard to its regulation of lytic and lysogenic transcription, we further investigated its mode of DNA binding. We show that purified Apl binds at pR-pL with high cooperativity, bends the DNA and spreads from specific binding sites into adjacent non-specific DNA. Although we were unable to detect Apl binding to a single site, we were able to use a simple mathematical model to extract estimates of binding energies for specific and non-specific sites and cooperativity by measuring Apl binding *in vitro* and *in vivo* to different numbers and arrangements of DNA sites. Each Apl monomer binds to DNA with low sequence specificity but with strong cooperativity between immediate neighbouring monomers.

MATERIALS AND METHODS

Assay and expression strains

NK7049 (Δ lacIZYA) X74 galOP308 Str^R Su⁻ from R. Simons (24) was the host strain for all LacZ assays. DH5 α and XLI-blue were hosts for recombinant DNA work. Strains were grown at 37°C in lysogeny broth (LB), with the addition of ampicillin (100 μ g ml⁻¹ for pZE15 based plasmids) and kanamycin (50 μ g ml ml⁻¹ for pUHA1) where necessary.

pZE15-P_{lac}-LacZ was constructed by inserting the *lacZ* gene into the BamHI and Hind-III sites of the ampicillin resistant, *colE1* based plasmid pZE15 (25). Lac repressor was supplied by pUHA-1, a p15A based plasmid encoding kanamycin resistance and carrying the wild-type *lacI* gene and promoter, obtained from H. Bujard (Heidelberg University, Germany).

Chromosomally integrated LacZ reporters were NK7049 (λ RS45 Δ YA pBC2-based or pMMR9-R) based 186 pR- or 186 pL-lacZ reporters. The pR- and pL-lacZ reporter plasmids were created as described in (26). These plasmid-based lacZ fusions were then transferred to the lacZ reporter phage λ RS45 Δ YA for insertion into the *Escherichia* colichromosome. Plasmid-containing strains were infected with λ RS45 Δ YA, and blue plaque-forming phage amongst the progeny were identified and purified on NK7049, on plates containing X-gal (24). Lysogenising NK4079 with the reporter phage ensures that the reporters are all located at an identical position (*att* lambda) in the chromosome. Chromosomal integrants were checked for monolysogens by polymerase chain reaction (PCR) (27).

Apl was supplied to the reporter strains by pZE15Apl, a colE1 based plasmid (25), where Apl expression was under control of the plac promoter. The parental pZE15 plasmid was used as an Apl⁻ control. Reporter strains also carried the pUHA-1 plasmid, as a source of lac repressor. Thus, Apl expression was controlled by addition of isopropyl-β-D-thiogalactoside (IPTG) to the growing culture, and promoter activities assayed in a microtiter plate format, according to Palmer *et al.* (28).

In vitro DNA binding assays

Gel mobility shift assays. For gel shift assays, double stranded DNA fragments with one strand ³²P end-labelled were generated by PCR in which one of the primers had been ³²P end-labelled using polynucleotide kinase. The double stranded ³²P-labelled PCR product was purified by polyacrylamide gel electrophoresis, the DNA eluted from the gel slice overnight at 37°C, ethanol precipitated and resuspended in binding buffer before use.

The DNA sequences of the oligonucleotides used in binding assays are given in Supplementary Figure S1.

Binding reactions (10 μl) were prepared by addition of DNA (~300 cpm), Apl (exhaustively dialysed against 50 mM Tris-HCl (pH 7.5), 0.1 mM ethylenediaminetetraacetic acid (EDTA), 10% (v/v) glycerol, 150 mM NaCl (TEG 150)) and binding buffer (TEG150). Reactions were left on ice for at least 30 min to allow attainment of equilibrium, and 6 μl loaded onto running polyacrylamide (0.5 × TBE) gels containing 10% glycerol. For binding to short DNA fragments, 15% gels were used, whilst 8% gels were used for DNA bending assays. Gels were electrophoresed at 4°C at constant current (20 mA) for ~2 h. Upon completion of electrophoresis, gels were dried, exposed to a phosphorimager screen and quantitated using the volume integration feature of Imagequant (Molecular Dynamics) or Imagelab (BioRad) software.

The fraction of DNA bound in each lane was calculated as (counts for the retarded band)/(counts for the whole lane), and corrected for a small degree of protein independent smearing using a no protein control lane. The DNA concentration was sufficiently low that total protein concentration could be substituted for free protein concentration.

Bending assay. DNA fragments containing 3, 4, 5, 6 or 7 Apl binding sites (Supplementary Figure S1) were prepared by annealing complementary oligonucleotides and ligating into the blunt *Hpa* I site of pBend 5 (29). The region of this plasmid containing 17 circularly permuted restriction sites and the Apl binding sites were then amplified by PCR using primers pBend SK (TAGTGGATCCCCGGGCTGCA) and pBend KS (CGACGGTATCGATAAGCTTGG). This fragment was labelled by inclusion of ³²P α-ATP (10 μCi) in the PCR reaction. The PCR product was purified by polyacrylamide gel electrophoresis and an aliquot digested in 10 μl reactions with MluI, EcoRV or BamHI. Digestion produced three fragments, including a fragment containing the Apl binding site located at either the left end, centre or right end within the fragment. A total of 2 μl of this digest was used in binding reactions for determination by gel shift assay of the electrophoretic mobility of the

protein–DNA complex. Binding reactions were performed in TEG 50 buffer and contained 3.2 μM Apl. Samples were loaded on 8% polyacrylamide, 0.5 × TBE gels containing 10% glycerol and run at 20 mA constant current and 4°C. Loading dye was run in a separate lane. Following electrophoresis, gels were dried, exposed to a phosphor imager screen and DNA mobility quantitated using Imagequant software (Molecular Dynamics). The apparent bend angles were quantitated according to Equation (1):

$$\frac{\mu_M}{\mu_E} = \cos\left(\frac{\alpha}{2}\right) \quad (1)$$

where, α is the bend angle and μ_M and μ_E are the relative mobility of DNA fragments containing the binding site at the middle and at the end of the fragment, respectively (7,29). Apparent bend angles were calculated from the mean of four independent experiments.

DNase I footprinting. Experiments were performed essentially according to (30), with modifications described by (31). This method uses magnetic beads to facilitate sample preparation. Double stranded DNA fragments for footprinting were prepared by PCR using a ³²P end-labelled primer and a biotinylated unlabelled primer (biotin-RSP). The PCR reaction (20 μl) was passed over a PCR purification spin column (Geneworks, Adelaide) to remove any unincorporated biotinylated primer which would compete with full length product for binding to the beads. The eluate from the spin column (60 μl) was added to 75 μl of streptavidin-coated magnetic beads (Dynabeads, Dynal), prepared according to the manufactures recommendations and incubated for 1 h at room temperature to allow the biotinylated, radiolabelled PCR product to bind. The beads were then washed several times, resuspended in 50–100 μl binding buffer and stored at 0°C for up to 1 week. Bead DNA (5 μl, ~6000 cpm) was added to binding buffer containing appropriate Apl concentrations, in a total volume of 40 μl. The footprint binding buffer consisted of 50 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 10% (v/v) glycerol, 75 mM NaCl, 10 mM MgCl₂ 1.5 mM CaCl₂ 1 μM bovine serum albumin (BSA). These binding reactions were incubated at 37°C for at least 30 min to allow attainment of equilibrium, prior to addition of DNase 1 (0.5 ng). The DNase 1 reaction was allowed to proceed for exactly 10 min at 37°C before being stopped with 50 μl of stop solution (4M NaCl, 100 mM EDTA). The beads were washed once with 100 μl of 2M NaCl, 20 mM EDTA, once with 100 μl of 10 mM Tris-Cl, 1 mM EDTA, pH 8.0 and resuspended in 6 μl of loading buffer (90% formamide, 10 mM EDTA). The reactions were heated to 90°C for 3 min and 5 μl loaded immediately onto a 6% denaturing polyacrylamide gel. Electrophoresis was at 1500 V (constant voltage) for ~2 h. Following electrophoresis, gels were dried onto filter paper, exposed overnight to a phosphor imager screen and viewed using Imagequant. Apl concentrations used in the footprints were: 3000, 2000, 1000, 794, 631, 500, 400, 319, 100 and 10 nM.

In vivo Apl expression system. Apl expression from pZE15Apl was controlled from the pLac promoter by Lac repressor supplied by pUHA-1. Relative expression of

Apl from pZE15Apl in NK7049 λ RS45 Δ YA-pMRR9R-MMpR⁺pL⁺.lacZ from the pLac promoter at 0, 3.6, 5.4, 7.2, 9, 13.5, 18, 20, 27, 40, 60 and 100 μ M IPTG was determined by comparison with lacZ expression from pZE15-plac-LacZ in NK7049 (Supplementary Figure S2). Data was pooled from assays performed on two to three different days, each with four biological replicates. For each dataset, the Apl-containing lacZ reporter value was divided by the mean parental (no Apl) plasmid value for that IPTG concentration, and relative repression pooled. The 95% confidence intervals of the pooled data were calculated, and relative repression curves plotted with *y*-axis error bars being 95% confidence intervals in relative repression value and *x*-axis error bars being 95% confidence intervals in pLac promoter activity.

Statistical mechanical modelling

We have taken a statistical mechanical approach to modelling both *in vivo* and *in vitro* Apl binding data, where the relative probability of each possible species in the proposed binding model is explicitly considered (32–35).

In vivo modelling. Initial modelling aimed to fit the *in vivo* LacZ reporter data, using the seven specific Apl sites between the pR and pL promoters and two non-specific flanking sites on either side of the specific sites, making a total of 11 sites. All possible Apl bound states could then be described as an 11 digit binary number, with a 1 if Apl is bound that site or a 0 if Apl is not, resulting in 2^{11} (= 2048) different Apl bound states.

The weight for each of these states was stored in an array 2^{11} long. Each state is given the initial weight of 1, for each non-specific site occupied each weight multiplied by cU , where U is the non-specific binding parameter, and c is the concentration, and multiplied by $c(B+U)$, where B is the specific binding parameter. For all states, where any site, i , and neighbouring site, $i+1$, are bound by Apl, the weight is multiplied by the cooperatively parameter, F .

Binding at the first two non-specific and the first two specific sites (i.e. sites 1–4) was considered to compete with RNA polymerase (RNAP) at the pR promoter, and the last two non-specific and last specific sites (i.e. sites 9–11) considered to compete at the pL promoter, where RNAP covers at least -45 to $+10$ of a promoter region. Additional states were defined where RNAP could bind to pR when sites 1–4 were not bound by Apl and multiplied by a pR binding parameter, R . The same was done with pL with a pL binding parameter, L . This makes the reasonable assumption that RNAP is at a fixed cellular concentration. Transcription, hence LacZ gene expression from either the pR or pL promoter is considered to occur from all states where RNAP can bind the promoter, hence relative expression is proportional to the probability of RNAP being at the promoter. The probability of RNAP occupying the promoter was determined as the sum of all states where RNAP is at the promoter, divided by the sum of all possible states. To relate this to relative expression Lac Z data, this is normalized by a constant that is related to pR and pL, such that at an Apl concentration of 0, relative expression is 1.

Fitting in vivo repression data. Four constructs were assayed for Apl binding, each with a different number of specific sites from the seven present in wild-type. The arrangements were:

pR 0011111100 pL
pR 0011010100 pL
pR 0011000100 pL
pL 0011111100 (-pR),

where 0 is a non-specific site, 1 is a specific site, pR represents a pR promoter, pL represents a pL promoter, (-pR) represents a mutant inactive pR promoter and the promoter on the left indicates the promoter from which the LacZ reporter is expressed.

Each different arrangement of Apl sites was modelled as described above, and the relative expression curves were globally fit.

Data fitting. The model was fitted to experimental data derived from LacZ assays. In these assays, a LacZ reporter gene was expressed from either a pR or a pL promoter integrated into the bacterial chromosome. Apl was supplied from an IPTG inducible promoter on a separate plasmid, pZE15Apl. Each LacZ assay was repeated on 8–12 biological replicates, and relative expression was determined for each IPTG concentration by dividing by the value at 0 μ M IPTG for each replicate. Values for all replicates were used in data fitting, with the error between the model and data for each dataset divided by the total number of points in that dataset.

The minimum error was found using a combined random Monte Carlo and linear search method. Random parameter guesses were made, and the best guesses were used as starting guesses for linear optimization. The advantage of this method is that it avoids local minima, whilst searching a wide range of parameter values. Initially, 1000 random guesses were taken, and the error minimized with the *fmincon* function of MATLAB. This was performed 100 times. The error as a function of each parameter value was examined, the bounds of the random guesses updated and the minimization was performed an additional 200 times. Values for both rounds were pooled together, and each parameter plotted against the error to assess convergence. Plotting each parameter value against the error for the 300 rounds of minimization clearly showed convergence to the lowest error (Supplementary Figure S3). The twenty fits with the lowest errors were then averaged and the standard deviation determined, resulting in $B = 4.55 (\pm 0.53) \times 10^{-5}$ Apl expression units, $U = 0.99 (\pm 0.09) \times 10^{-5}$ Apl expression units, $F = 50.7 (\pm 7.9)$, $pR = 6.88 (\pm 1.09)$, $pL = 0.29 (\pm 0.15)$, with an error of 0.37.

The parameter values with the lowest 100 errors were also plotted to investigate parameter correlations (Supplementary Figure S4). The effect of parameter variation was also examined, where for clarity the species distribution a three site operator sequence was simulated, varying one parameter at a time by ± 1 standard deviation.

Fitting of in vitro data. Fitting of *in vitro* binding data was used as an independent test of the binding model.

Firstly, this was done using gel shift data. Gel shift data was quantified with Imagequant software. The total volume of each lane and the volume of the free DNA of each lane was quantified separately, corresponding to total loaded DNA and unbound DNA. The fraction of free DNA was determined as the amount of unbound DNA divided by the total loaded DNA. The amount of free DNA decreased with increasing Apl concentration, down to a limit which corresponds to a small proportion of DNA that is not annealed correctly and hence is unable to bind Apl. The fraction of free DNA was then normalized to 1 when the Apl concentration was 0, and normalized to 0 when free DNA plateaus. This was done for the seven site gel shift data, for concentrations 0, 50, 100, 200, 400, 800 and 1600 nM of Apl.

As gel shift data was used as an independent measure of the predictive power of the model, the mean of the parameter values for B, U, and F from the *in vivo* fitting, given above, were used. The promoter parameters pR and pL were not used as RNA polymerase is not present in the gel shift assays. To convert B and U into nM units from Lac units, these were multiplied by a calibration factor, *d*, which was determined by fitting the seven site gel shift data for free DNA using linear optimization, resulting in a fit value of $d = 1.75$. The resultant fit to the seven site data was excellent, considering only the concentration was rescaled resulting in $B = 79\,700\text{ M}^{-1}$ and $U = 17\,300\text{ M}^{-1}$.

These same parameter values were used to calculate the distributions of bound species, as a function of Apl concentration, for all fragments used in gel shift assays.

RESULTS

Three adjacent operators are required for efficient Apl binding *in vitro*

Purified, refolded Apl protein (16) was used in gel shift assays to determine the number and arrangement of recognition sequences needed for efficient binding to the pR-pL region *in vitro* (Figure 2). Binding was detected only when three or more operators were present, suggesting that Apl binding is highly cooperative. The apparent K_D decreased as the number of operators was increased. When six or seven operators were present, multiple shifted species were observed, demonstrating multiple, distinct relatively stable Apl–DNA complexes.

To examine whether cooperation between multiple operators is adjacent or longer range, binding was tested using DNAs with combinations of intact and mutated Apl recognition sequences. These were designated 101, 1101 and 10101, where 1 indicates an intact operator and 0 indicates a scrambled site. No binding was detected to any of these fragments (Figure 2G–I). The binding to the 111 fragment, but lack of binding to the fragment containing scrambled sites, indicates that three adjacent operators are needed for efficient binding *in vitro*. A gel shift experiment using the five operators at *attP* (Figure 2J) showed a similar binding pattern to that seen using the central five operators at pR-pL, indicating a similar mechanism of binding.

Apl bends DNA upon binding

As DNA bending appears to be a conserved property of the tyrosine integrase family of RDF proteins, the ability of Apl to bend DNA was tested more directly using the ‘circularly permuted gel shift’ technique (Figure 3) (29,36–37). Fragments containing three, four, five, six or seven adjacent operators from pR-pL were tested. Retardation of the fragments differed markedly depending on the position of the set of binding sites within the fragment, suggestive of bending. The calculated bend angle (from Equation 1) for the 3-operator segment was $87 \pm 1^\circ$ (Figure 3). This value is similar to the 72° seen in the crystal structure of λ Xis bound to three adjacent sites (14), though smaller than the 120° estimate using a similar gel shift technique (7). Increased apparent bend angles were obtained for the four-, five- and six-operator segments, with a possible slight decrease for the seven-operator segment (Figure 3). These bends are of similar magnitude to those seen using the same technique with P22 Xis (13), P2 Cox (10) and Puckovnik Xis (5) at their attachment sites. As the angle estimate is based on an assumption of a planar bend, which may not be the case for DNA with multiple Apl operators, these changes in apparent bending indicate only that the architecture of the DNA changes with increasing bound Apl.

Apl binding spreads into adjacent non-specific sites

Whilst Apl DNase I footprinting performed by (17) revealed protections and enhancements that extended beyond the specific binding sites at pR-pL and at *attP*, these assays were performed with crude cell extracts and with native 186 sequences flanking the operator DNA. Hence, the observed DNA alterations may have been due to proteins other than Apl or to Apl binding to cryptic operators within the 186 sequence. To test for spreading of Apl binding into adjacent non-specific sequences, DNase I footprinting was repeated using purified Apl and with either three (Figure 4A) or five (Figure 4B) adjacent Apl operators embedded in non-186 DNA. In both cases, the periodic protections and enhancements previously observed within the Apl operators (17) were apparent. Furthermore, protections continued on either side of the specific sequences, showing Apl spreading to non-specific sites in the vector DNA, with the extent of spreading increasing with Apl concentration.

Interestingly, there is no sharp transition between occupation of the specific operator sites and occupation of the adjacent non-specific sites. That is, protection of the flanking DNA begins at Apl concentrations at which the operator sites are not fully occupied. This suggests that Apl’s affinity for its operators is not substantially greater than for non-specific sites. If these affinities were very different, then one would have expected that some Apl concentrations would give strong protection of the operator sites with very little spreading.

These results show that, like other RDFs, Apl binding at specific sites can seed spreading of Apl into adjacent non-specific sites, and indicate strong binding cooperativity and weak discrimination between the operator sites and non-specific DNA.

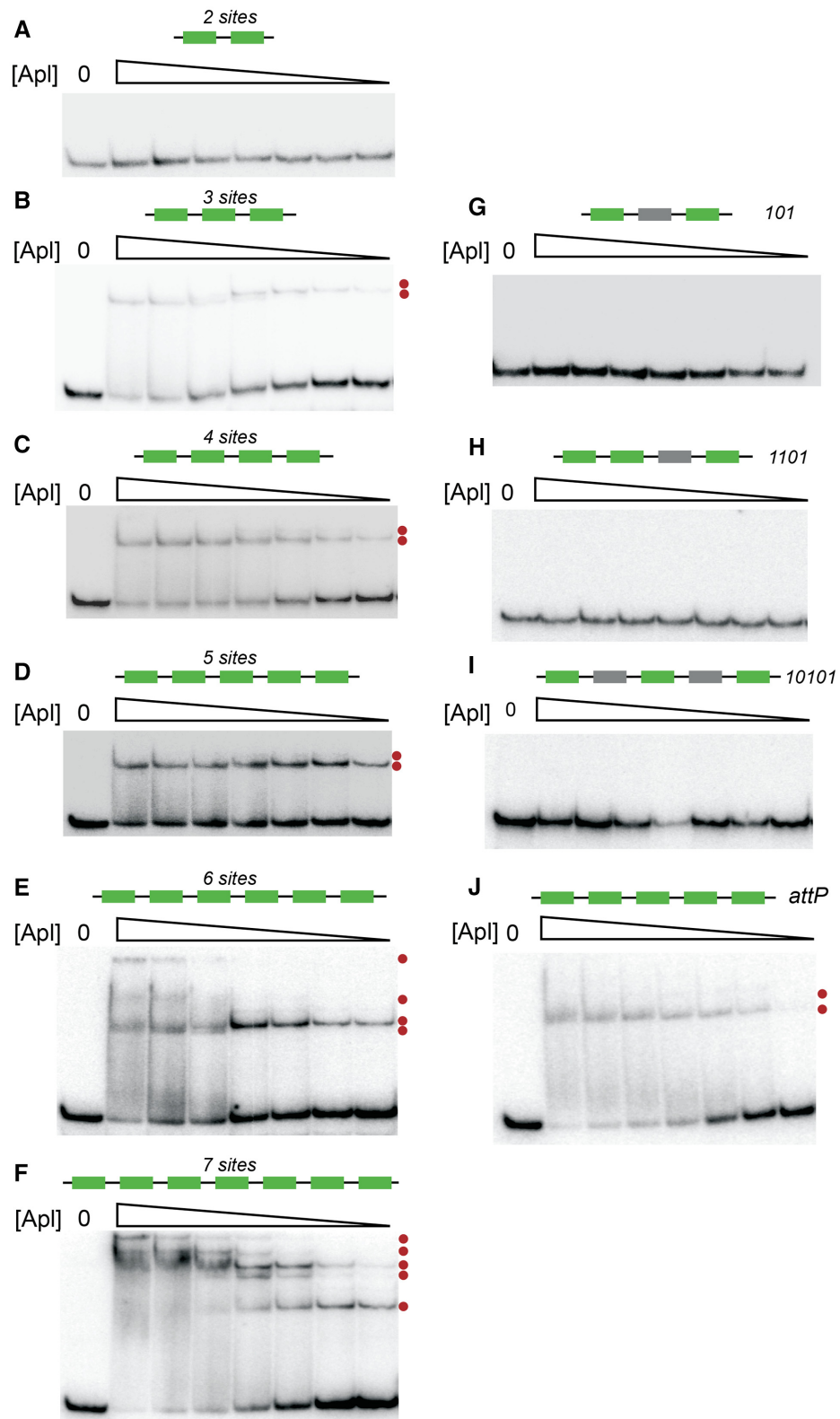


Figure 2. Apl DNA binding to DNA. DNA gel shift assays using DNA oligomers containing three to seven Apl (A–F) binding sites. Binding was assayed to DNA fragments with various combinations of specific and non-specific sites; 101 (G), 1101 (H) and 10101 (I), where 1 represents a specific Apl operator and 0 represents a scrambled operator. An assay was also performed on a fragment containing the five specific operators from the *attP* site (J). Apl concentrations were 6400, 3200, 1600, 800, 400, 200 and 100 nM. Dots next to a gel indicate the presence of a different protein–DNA species.

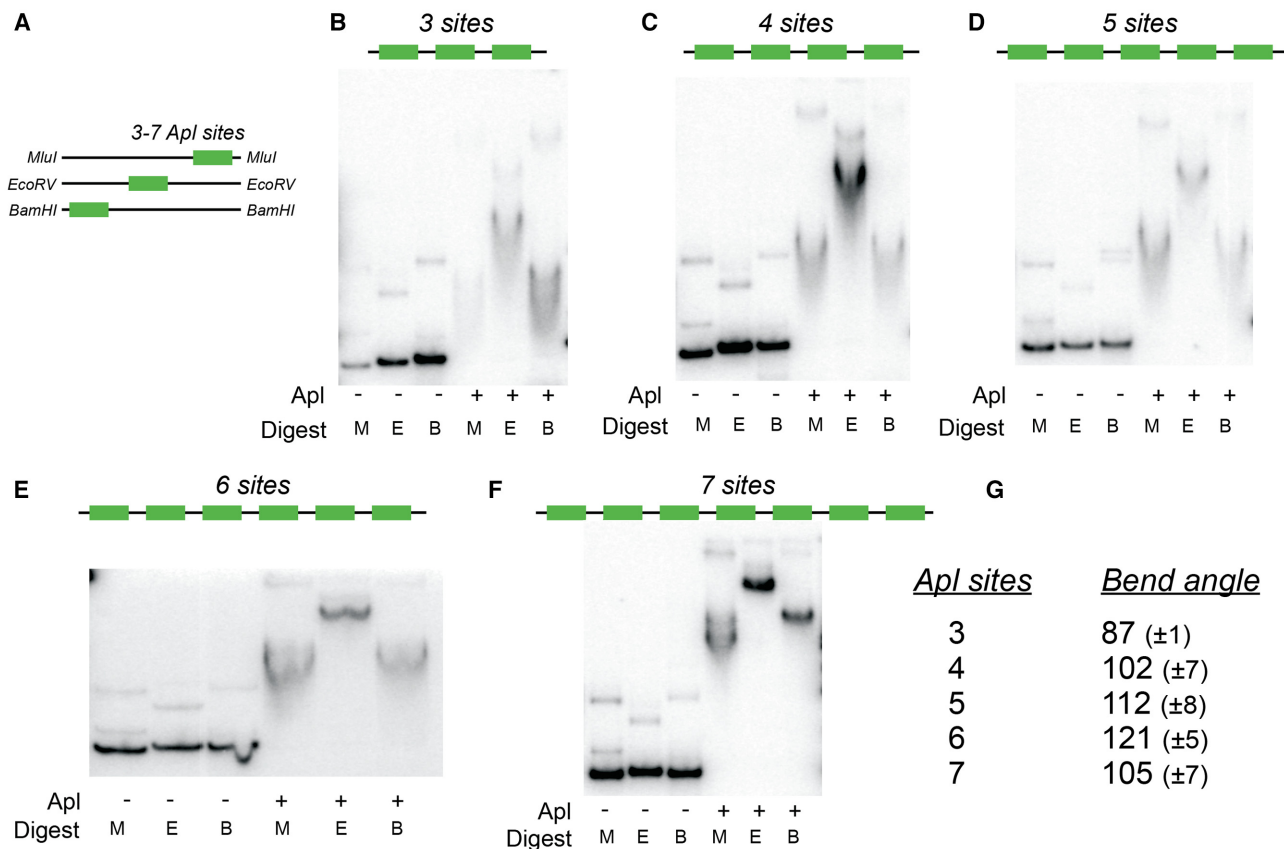


Figure 3. DNA bending assay. (A) The arrangement of Apl binding sites at the ends or centre of the DNA fragment allows estimation of DNA bending angle from gel shift assays. (B–F) shows the gel shifts in the absence and presence of 3.2 μM Apl for fragments containing 3–7 Apl operators. M, E and B indicate the fragment corresponding to digestion with MluI, EcoRV or BamHI, as in panel A. (G) Summary of apparent bend angles derived from the bending assays. Errors are confidence limits based on four independent experiments.

Effect of operator mutations on Apl repression of the pR and pL promoters *in vivo*

To examine the role of cooperative binding of Apl on its activity at pR-pL, we used lacZ reporter constructs carrying mutations at the central three Apl operators (Figure 5). Three different fragments were tested: 111111 (wild-type (WT), bearing operators 1–7), 1101011 (operators 3 and 5 scrambled) and 1100011 (operators 3, 4 and 5 scrambled). Fragments oriented to report either pR activity or pL activity were fused to lacZ in a lambda prophage, as described in Materials and Methods. The pL reporter carried mutations inactivating pR, in order to remove pR's inhibition of pL by transcriptional interference (38). There is no reciprocal interference of pL on pR (38). Apl was supplied from a multicopy plasmid (pZE15Apl) under plac/IPTG control. We have not measured absolute Apl concentrations resulting from this expression system, however relative Apl expression was quantitated by constructing an equivalent pZE15-plac.lacZ plasmid and assaying LacZ activity in response to IPTG (Supplementary Figure S2). Thus, relative Apl concentrations are given in terms of expression units.

Induction of Apl expression with IPTG resulted in repression of both pR and pL (Figure 5A). Promoter expression is expressed relative to the Apl minus control plasmid (~800 units for pR, ~100 units for pL(pR⁻)). Wild-

type pR was repressed by Apl to ~0.4 of unrepressed and wild-type pL(pR⁻) to ~0.2, a difference in sensitivity consistent with previous observations (19). Testing the effect of the 11001011 and 1100011 operator mutations on pR repression showed that in both mutants repression was much weaker but was still detectable at the highest Apl concentrations (Figure 5A).

Modelling of Apl repression of pR and pL

To achieve a more quantitative understanding of the mechanism of Apl regulation of pR and pL, we tested whether a simple statistical mechanical model of Apl DNA binding could explain the reporter data and quantify cooperativity and DNA binding affinity. The pR-pL region was modelled as comprising seven specific Apl operators (O₁-O₇), two non-specific Apl binding sites on each side (N₋₁, N₀, N₈ and N₉) and two RNAP binding sites (Figure 5B). Apl binding at operator site O₁ and O₂ was assumed to compete with RNAP binding at pR, since the conserved sequences lies at pR -1 to -6, and +5 to +10, respectively, as was binding to the two non-specific sites N₋₁ and N₀ adjacent to site 1. The conserved 6 bp O₇ sequence lies at pL +3 to +8 and Apl binding to this site or the adjacent non-specific N₈ and N₉ sites was also assumed to compete with RNAP binding to pL (Figure 1).

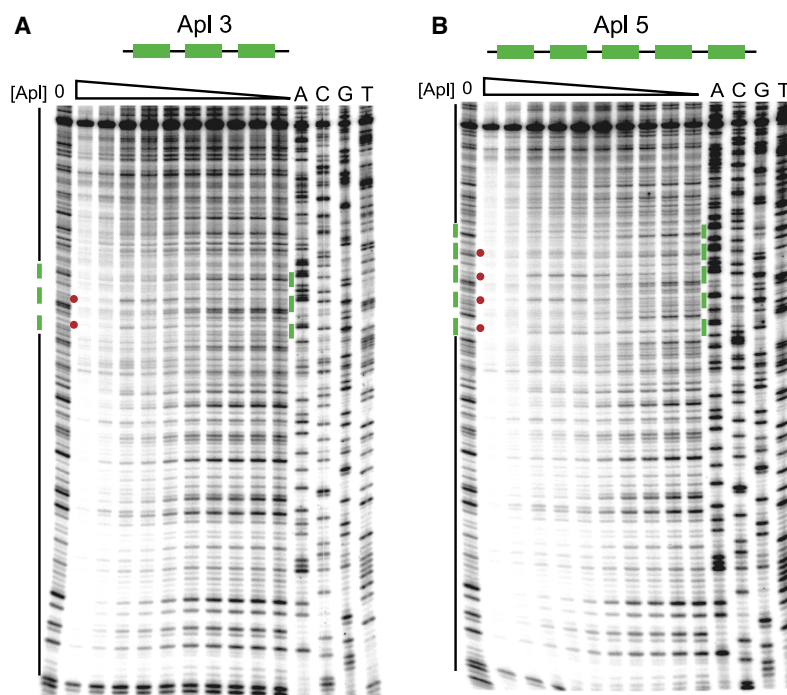


Figure 4. DNase I footprinting shows spreading of Apl binding into surrounding DNA of unrelated sequence. DNA containing either (A) three or (B) five Apl operator sequences was embedded into unrelated DNA sequence (pBluescript plasmid). DNaseI footprint reactions, examining the equivalent of the 186 top strand, were performed as described in ‘Materials and Methods’ section. The leftmost lane in each gel contained no Apl. Apl concentrations were 3000, 2000, 1000, 794, 631, 500, 400, 319, 100 and 10 nM (left to right). The four lanes on the right hand side of each gel contain dideoxy sequencing reactions as indicated. At the side of each gel the rectangles indicate the region of DNA corresponding to Apl operator sequences, and the black lines indicate sequence corresponding to plasmid DNA. The dots represent the characteristic enhancements of DNase I cleavage seen with Apl binding, indicative of distortion of the DNA (17).

All possible states were then defined and given a statistical weight based on the interactions present (Figure 5B). Binding of an Apl monomer to an operator was given a statistical weight $[Apl] \cdot B$, where B is the specific association constant, and $[Apl]$ is a scaled Apl concentration. Non-specific Apl binding of an Apl monomer was given a weight $[Apl] \cdot U$, where U is the non-specific association constant. A cooperation parameter, F , was applied for two adjacent bound Apl monomers. RNAP binding to the pR and pL promoters was given weights R and L , respectively, combining the unknown but constant cellular concentration of RNAP and unknown RNAP binding constants to the promoters. Based on these parameters, a weight for each of the 2448 possible states can be calculated. The probability of any state occurring is the weight of that state divided by the sum of the weights for all states. The activity of each promoter was assumed to be proportional to the sum of the probabilities of all the states where the RNAP is bound to the promoter. Thus, possible specific effects of Apl on transcription initiation, promoter clearance or elongation were ignored. A scaling factor for each promoter was applied to set its activity in the absence of Apl to 1.

The model parameters were adjusted to optimize the fit to the complete set of *in vivo* repression data, using a combined Monte-Carlo/linear optimization approach (Supplementary Figures S3 and 4). The repression of the WT pL(pR⁻) and WT pR reporters by Apl are reproduced well at higher $[Apl]$ concentrations, however, the model predicts

a greater difference between the 1101011 and 1100011 reporters than is seen in the data (Figure 5A). The obtained Apl binding constants were $B = 4.55 (\pm 0.53) \times 10^{-5}$ and $U = 0.99 (\pm 0.09) \times 10^{-5}$ (Apl expression units)⁻¹. Thus, the K_D for monomer binding to a single operator is $\sim 22\,000$ Apl expression units ($1/B$). Since the maximum *in vivo* Apl concentration, produced from pLac on a multicopy plasmid, was ~ 15 -fold less than this (1500 expression units; Supplementary Figure S2), it is clear that binding to a single operator is weak. The RNAP binding values were fit as $R = 6.88 (\pm 1.09)$, and $L = 0.29 (\pm 0.15)$ and although these values are harder to interpret as they are used to normalize the binding curves, the fact R is larger than L is consistent with pR being a stronger promoter.

Remarkably, non-specific binding is predicted to be only ~ 4.5 -fold weaker than specific binding. The fitted value for cooperativity, $F = 50.7 (\pm 7.9)$, equivalent to a free energy ($\Delta G = -RT \ln F$) (32,39) of -2.4 kcal/mol, reflects a large contribution to Apl binding from cooperativity between adjacent monomers. The apparent K_D for cooperative binding to two operators becomes ~ 3100 expression units, only twice the maximum expression level, whilst for three operators, the apparent K_D is ~ 1600 expression units. For seven consecutive operators, the apparent K_D falls to ~ 750 expression units.

We explored more complex versions of the model, such as having different site binding strength according to sequence variation and adding in an additional cooperation

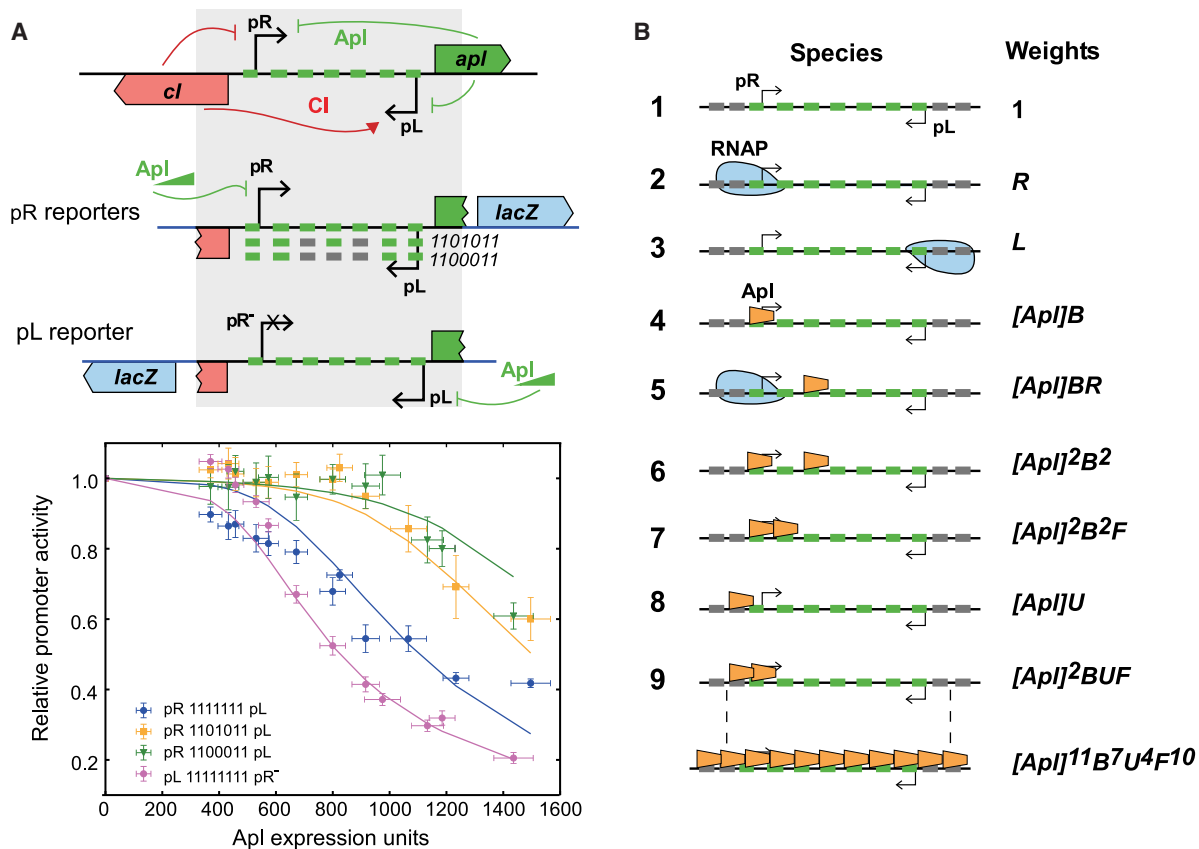


Figure 5. Mechanistic modelling of Apl DNA binding. (A) *Upper panel:* LacZ reporters measure Apl's function as a repressor of the switch region, where it represses both pR and pL. In the pR reporter, pR drives expression of a LacZ reporter gene. The pR construct was also made with operators 3 and 5, and 3, 4 and 5 scrambled. In the pL reporter, pL drives the expression of LacZ, with the pR promoter mutated to prevent transcriptional interference. *Lower panel:* LacZ reporter data, with the model fit shown as solid lines. Error bars indicate 95% confidence intervals. (B) Explanation of model parameters, R and L for RNAP binding to pR and pL respectively, B and U, for specific and non-specific DNA binding respectively, scaled by Apl concentration. When two Apl monomers are bound adjacently, this is further multiplied by the cooperativity parameter, F. More complex states are described by multiplication of these parameters.

term as the P2 Cox structure suggests an *i*+2 contact (6), but these models did not converge to one unique solution and did not significantly improve the data fit. Although the parameter values obtained with the more complex models were often different, the basic observations of high cooperativity and low discrimination were robust. Thus, a simple model of Apl binding applied to the *in vivo* repression data was able to confirm the qualitative conclusions from the gel shift and footprinting experiments that Apl binds with high cooperativity and with low discrimination between specific and non-specific sites.

Modelling *in vitro* Apl-DNA binding

To test whether the Apl binding model is consistent with the *in vitro* Apl binding data, we tested whether it could reproduce key features of both the gel shift and DNaseI footprinting data.

To obtain binding constants in units of Apl concentration, rather than Apl expression units, the value of B was fitted to the 7-operator gel shift binding data, holding B/U and F fixed ('Materials and Methods' section). The fraction of DNA remaining unbound in the Apl 7 gel shift (Figure 2) was quantified, the number of Apl sites in the model was

set to 7, and the model modified by removing all species involving RNAP. This gave association constants $B = 79,700 \text{ M}^{-1}$ and $U = 17,300 \text{ M}^{-1}$ and a good match to the Apl 7 data (Figure 6A).

These *in vitro* derived parameters were then used to predict binding patterns as a function of Apl concentration for each of the other fragments used in gel shifts (Figure 2), each time adjusting the model for the numbers of operators and the presence (where applicable) of scrambled sites. The results, over a range of Apl concentrations, are plotted in Figure 6A. The first point to note is that, as expected, the apparent K_D , defined as the Apl concentration which gives 50% of DNA unbound, decreases as the number of specific sites increases. This is mirrored in the gel shifts with intact operators, where the apparent K_D decreases from ~800 nM for the three operator fragment to ~275 nM for the seven operator fragment. Although the model does predict some weak binding to a two site fragment, we did not detect this experimentally in the gel shift. It is likely that there is rapid dissociation of weakly bound Apl during the course of the *in vitro* gel shift experiment. In contrast, in the *in vivo* promoter repression experiments, molecular crowding (40) will tend to favour association of the protein-DNA interactions.

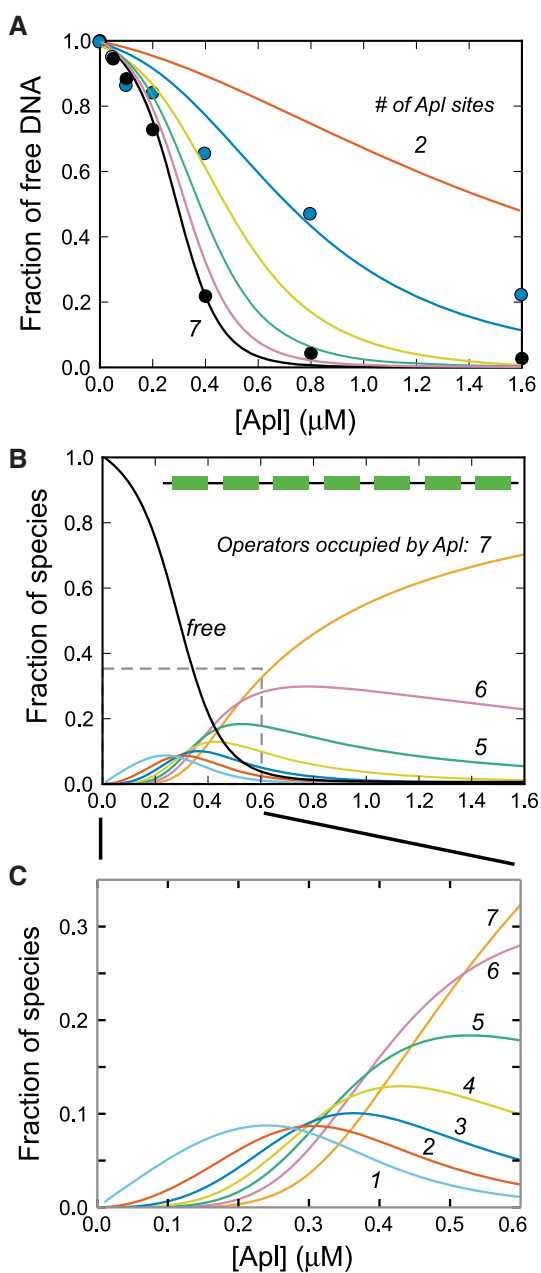


Figure 6. Model prediction of Apl binding. The model was used to predict the proportion of free DNA as a function of Apl concentration with different numbers of Apl operators, shown in solid coloured lines, for models with two to seven specific sites (A). The seven site gel shift data that was used to calibrate the Apl concentration range is shown with black circles. Data from the three site gel shift experiment are shown with blue circles. In panel B, the predicted distributions of binding stoichiometry to the full seven operators are shown, whilst panel C shows an enlargement of the Apl concentration regime for the Apl7 model, where several species of different stoichiometries are predicted to co-exist.

The model allows us to calculate the relative proportion of all the possible states, and hence calculate the proportion of each Apl binding stoichiometry (Figures 6 and 7; Supplementary Figures S5 and 6). Both the experiments (Figure 2) and the simulations indicate that Apl binding is much stronger when there are adjacent specific sites. Although the

101, 1101 and 10101 fragments are predicted by the model to bind Apl at high Apl concentrations, the overall affinity was very much weakened. Whilst our model assumes for simplicity a constant cooperativity value (F) between adjacently bound Apl monomers, it is possible that cooperativity between a specifically bound Apl and an Apl bound to a scrambled site is weaker. However, the species distributions are not particularly sensitive to variations in the fitted values of B , U and F , illustrated by recalculating the distributions when holding two of the parameters fixed, and varying the third, up or down, by one standard deviation (Supplementary Figure S7).

Thus, the model reproduced the binding affinity changes observed in the *in vitro* gel shift data reasonably well. In addition, the fitting enables the affinity of Apl to a single site to be calculated. The binding of Apl to a single operator is very weak, with an estimated K_D of 12.5 μM. As a result of the moderate level of cooperativity between bound Apl monomers, the modelling also predicts that at intermediate Apl concentrations (~0.2–0.7 μM), where all operators are not occupied, significant fractions of species with different numbers of bound Apl monomers should be present (Figure 6B and C). Indeed, consistent with the predicted distribution of species, two retarded bands were seen for fragments containing three, four or five operators, and more than three bands were apparent with the six- and seven-site fragments. Uncertainty in how the various Apl–DNA complexes migrate in the gel means it is not possible to assign bands to specific species.

We also used the model to simulate Apl spreading in the DNase I footprints. Two non-specific sites were placed on each side of three and five specific operators, and the probability for each of the sites to be occupied was calculated over a range of Apl concentrations. The result is depicted in ‘footprint form’, showing the expected spreading and the lack of a clear boundary between the specific and non-specific sites (Figure 7), comparable to the corresponding experimental footprint data (Figure 4).

DISCUSSION

Apl’s DNA binding mode

DNase I footprints of Apl at its binding sites at pR–pL and at *attP* are indicative of Apl binding on the inside face of bent DNA (17). Our results and previous studies establish a number of features of Apl DNA binding, many or all of which are shared with other well-studied RDFs (1–2,5,9,12–14,20,41): (i) the operators are arranged as direct repeats, spaced roughly one turn of the DNA helix apart, presumably with one monomer bound per operator; (ii) binding to a single specific operator is weak; (iii) binding to adjacent operators is highly cooperative; (iv) binding causes DNA bending; and (v) the difference in affinity for specific and non-specific sites is small, presumably reflecting flexible sequence recognition. The presence of these features in Apl supports the idea that this basic mode of DNA binding is universal in this group of proteins (13,14).

Taking advantage of Apl’s activity as a transcriptional repressor and by using operator mutants, we were able to generate *in vivo* data that enabled model-based extraction of estimates for the basic biochemical parameters for Apl bind-

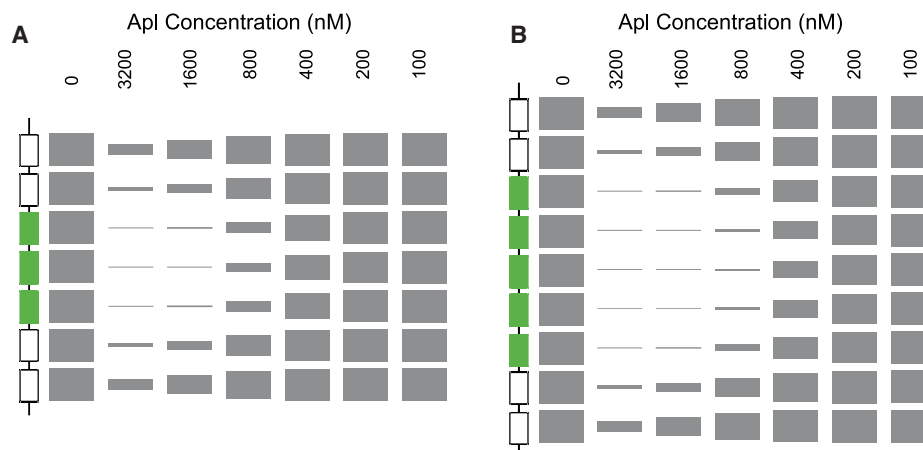


Figure 7. Simulation of DNase footprinting data for (A) three and (B) five operators, respectively, reproduces the observed spreading of Apl binding into flanking DNA sequences.

ing. A simple model that uses two DNA binding affinities for Apl (specific and non-specific) and a single parameter for cooperation between adjacent monomers was able to reproduce the *in vivo* and *in vitro* binding data reasonably well. Monomer binding to a single site was estimated to have K_D s of $\sim 12.5 \mu\text{M}$ (ΔG of -6.9 kcal/mol) for a specific operator, and $\sim 58 \mu\text{M}$ (ΔG of -6.0 kcal/mol) for a non-specific site. The fitted value for cooperativity, $F \sim 50$ is equivalent to a ΔG of -2.4 kcal/mol. Although these values should be regarded with caution, they compare reasonably with estimates for other DNA-binding proteins. In gel-shift experiments, a K_D of $1 \mu\text{M}$ was estimated for P22 Xis binding to a DNA with a single operator (13), and weak but detectable binding to single operators at nM concentrations was seen for λ Xis (42), HPI Cox (43) and Gifsy-1 Xis (4) suggesting single site affinities that are substantially higher than Apl. Thus, it seems that Apl's affinity for its specific sites is relatively low, hence more binding sites may be needed in order to compensate. The $12.5 \mu\text{M}$ K_D is an average over all the Apl operators at pR-pL, but as the consensus is not strictly conserved, some operators may have higher affinity than others.

The relative levels of intermediate species seen in gel shifts with multiple operator DNA, and the coordinate occupation of multiple adjacent sites in DNaseI footprinting experiments reflects a balance between binding affinity and cooperativity. Our model allowed us to derive a quantitative measure of cooperativity, a parameter which is not available for other RDFs. Apl's cooperativity factor of ~ 50 is comparable to $F = 60\text{--}130$ between λ CI dimers ($= -2.5$ to -3 kcal/mol; (39), but less than the ~ 2000 (-4.7 kcal/mol) for HK022 repressor dimers (44).

Hence, the derived values for both binding affinity and cooperativity are comparable to other regulators of expression, and both affinity and cooperativity can be tuned for the desired regulatory outcomes.

The DNA binding mode and RDF function

The conservation of DNA binding mode amongst RDFs suggests that it is particularly suited to allow RDFs to foster a specific spatial arrangement of the DNA flanking their

binding sites in order to create appropriate DNA substrates for binding of the integrase and other recombination proteins.

Strong cooperativity between adjacent RDFs is likely to be needed to impart a static bend and a stiffening of the DNA to fix an optimal recombination structure. Each additional binding site provides an 'architectural increment' to this structure. Thus, a monomer/one-DNA-turn binding unit for an RDF may be an advantage over a typical dimer/two-DNA-turn binding unit because it allows for smaller architectural increments in the evolutionary construction of *att* sites.

The ability to spread into adjacent sites provides an 'RDF concentration window' for the creation of a particular DNA structure, since one less or more RDF in the chain is likely to significantly change the overall DNA arrangement. How the length of the RDF chain responds to RDF concentration can be readily tuned by alterations in the DNA sequence to foster or hinder the addition of the next monomer in the chain or by alteration in the cooperativity between RDFs. Spreading may also position the RDF where it can make favourable or competitive contacts with other recombination proteins, providing further concentration-dependent regulation of recombination. Mattis *et al.* (13) proposed that occupation of non-specific sites that overlap *Int* binding sites may cause high concentrations of P22 Xis to inhibit reintegration after excision.

The DNA binding mode and transcriptional regulation

Although the DNA binding properties of Apl and other RDFs seem well suited to their recombinational role, two features can also be used to provide effective and unusual transcriptional regulation. The first is highly cooperative binding, a feature that is common amongst transcription factors and which is used to generate sharp transitions between promoter activity and inactivity in response to small changes in regulator concentration (35). The second is spreading. Spreading from specific sites into non-specific sites is not often used in transcriptional control. One example is the ParB family of proteins, which mediate chromosome partitioning in various replicons. These are dimeric

HTH proteins that bind to specific ‘centromere’ sites but also are capable of spreading kilobase distances into adjacent DNA (45). This spreading is able to silence adjacent genes and may play a role in the partitioning process. Similar spreading has been observed for the *Drosophila* transcription factor Yan and its human homologue TEL/ETV6 (35).

The spreading of an RDF from its specific operators at *att* could in theory allow repression of the promoter for the recombinase gene when this is located adjacent to *att* but somewhat distant from the primary RDF binding sites. However, in P4 and KplE1, specific RDF sites at *att* overlap the integrase promoter (12,46,18), so spreading is not required, at least in these cases.

A spreading mechanism seems to be used in control of the lytic and lysogenic promoters of many P2 related bacteriophages. In these phages, the lytic and lysogenic promoters are arranged face-to-face and the primary operators for the immunity repressor lie over the lytic promoter, whilst the Apl/Cox operators are distinct from these and tend to lie over the lysogenic promoter or between the transcriptional start sites (23). In all tested cases, the Apl/Cox proteins repress the lysogenic promoter (41,19,21), and in some cases also the lytic promoter. It is not clear to what degree the ability of Apl to spread into non-specific sites is important for its regulation of pL and pR, since specific sites overlap both promoters. If binding to non-specific sites is removed in the model, then repression is weakened, but only slightly (repression can be restored by small increases in B or F). However, in P2, Cox repression of the *pe* early lytic promoter at high concentrations may be due to its spreading from its specific sites over the lysogenic promoter into non-specific sequences at *pe*. The face-to-face arrangement of the lytic and lysogenic promoters in P2-like phages provides sequence between the promoters that can be used to specify Apl/Cox binding and to set the required balance of repression by adjustment of spreading, without compromising the sequences for RNAP recognition or for immunity repressor binding. In contrast, in the lambdoid phages, the arrangement is more compact, with the lytic and lysogenic promoters back-to-back and the immunity repressor and Cro protein sharing the same operators. Thus, those features that make Apl and the other Cox proteins able to function as RDFs also adapt them well for their roles as lytic regulators of the face-to-face lytic and lysogenic promoters.

DATA AVAILABILITY

Matlab scripts are available at <https://doi.org/10.25909/5f17973d89b52> and <https://doi.org/10.25909/5f1798607cd90>.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Australian Research Council [DP110100824, DP150103009 to K.E.S]. Funding for open access charge: University of Adelaide Biochemistry grant [75125608].

Conflict of interest statement. None declared.

REFERENCES

- Lewis, J.A. and Hatfull, G.F. (2001) Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res.*, **29**, 2205–2216.
- Esposito, D. and Sococa, J.J. (1997) The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res.*, **25**, 3605–3614.
- Sam, M.D., Papagiannis, C.V., Connolly, K.M., Corselli, L., Iwahara, J., Lee, J., Phillips, M., Wojciak, J.M., Johnson, R.C. and Clubb, R.T. (2002) Regulation of directionality in bacteriophage λ site-specific recombination: Structure of the Xis protein. *J. Mol. Biol.*, **324**, 791–805.
- Flanigan, A. and Gardner, J.F. (2007) Interaction of the Gifsy-1 Xis protein with the Gifsy-1 attP sequence. *J. Bacteriol.*, **189**, 6303–6311.
- Singh, S., Plaks, J.G., Homa, N.J., Amrich, C.G., Héroux, A., Hatfull, G.F. and Vandemark, A.P. (2014) The structure of Xis reveals the basis for filament formation and insight into DNA bending within a mycobacteriophage intasome. *J. Mol. Biol.*, **426**, 412–422.
- Berntsson, R., Odegrip, R., Sehlén, W., Skaar, K., Svensson, L.M., Massad, T., Högbom, M., Haggård-Ljungquist, E. and Stenmark, P. (2014) Structural insight into DNA binding and oligomerization of the multifunctional Cox protein of bacteriophage P2. *Nucleic Acids Res.*, **42**, 2725–2735.
- Thompson, J.F. and Landy, A. (1988) Empirical estimation of protein-induced DNA bending angles: applications to λ site-specific recombination complexes. *Nucleic Acids Res.*, **16**, 9687–9705.
- Cho, E.H., Gumport, R.I. and Gardner, J.F. (2002) Interactions between integrase and excisionase in the phage lambda excisive nucleoprotein complex. *J. Bacteriol.*, **184**, 5200–5203.
- Laxmikanthan, G., Xu, C., Brilot, A.F., Warren, D., Steele, L., Seah, N., Tong, W., Grigorieff, N., Landy, A. and Van Duyne, G.D. (2016) Structure of a holliday junction complex reveals mechanisms governing a highly regulated DNA transaction. *Elife*, **5**, e14313.
- Ahlgren-Berg, A., Cardoso-Palacios, C., Eriksson, J.M., Mandali, S., Sehlén, W., Sylwan, L. and Haggård-Ljungquist, E. (2009) A comparative analysis of the bifunctional Cox proteins of two heteroimmune P2-like phages with different host integration sites. *Virology*, **385**, 303–312.
- Lewis, J.A. and Hatfull, G.F. (2003) Control of directionality in L5 integrase-mediated site-specific recombination. *J. Mol. Biol.*, **326**, 805–821.
- Calì, S., Spoldi, E., Piazzolla, D., Dodd, I.B., Forti, F., Dehò, G. and Ghisotti, D. (2004) Bacteriophage P4 Vis protein is needed for prophage excision. *Virology*, **322**, 82–92.
- Mattis, A.N., Gumport, R.I. and Gardner, J.F. (2008) Purification and characterization of bacteriophage P22 Xis protein. *J. Bacteriol.*, **190**, 5781–5796.
- Abbani, M.A., Papagiannis, C.V., Sam, M.D., Cascio, D., Johnson, R.C. and Clubb, R.T. (2007) Structure of the cooperative Xis-DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 2109–2114.
- Frykholm, K., Berntsson, R.P.-A., Claesson, M., de Battice, L., Odegrip, R., Stenmark, P. and Westerlund, F. (2016) DNA compaction by the bacteriophage protein Cox studied on the single DNA molecule level using nanofluidic channels. *Nucleic Acids Res.*, **44**, 7219–7227.
- Shearwin, K.E. and Egan, J.B. (1996) Purification and self-association equilibria of the lysis-lysogeny switch proteins of coliphage 186. *J. Biol. Chem.*, **271**, 11525–11531.
- Dodd, I.B., Reed, M.R. and Egan, J.B. (1993) The Cro-like Apl repressor of coliphage 186 is required for prophage excision and binds near the phage attachment site. *Mol. Microbiol.*, **10**, 1139–1150.
- Panis, G., Méjean, V. and Ansaldi, M. (2007) Control and regulation of KplE1 prophage site-specific recombination: a new recombination module analyzed. *J. Biol. Chem.*, **282**, 21798–21809.
- Reed, M.R.R., Shearwin, K.E.E., Pell, L.M.M. and Egan, J.B.B. (1997) The dual role of Apl in prophage induction of coliphage 186. *Mol. Microbiol.*, **23**, 669–681.

20. Yu, A. and Haggard-Ljungquist, E. (1993) The Cox protein is a modulator of directionality in bacteriophage P2 site-specific recombination. *J. Bacteriol.*, **175**, 7848–7855.
21. Esposito, D., Wilson, J.C.E. and Scocca, J.J. (1997) Reciprocal regulation of the early promoter region of bacteriophage HP1 by the cox and CI proteins. *Virology*, **234**, 267–276.
22. Karlsson, J.L., Cardoso-Palacios, C., Nilsson, A.S. and Haggard-Ljungquist, E. (2006) Evolution of immunity and host chromosome integration site of P2-like coliphages. *J. Bacteriol.*, **188**, 3923–3935.
23. Nilsson, H., Cardoso-Palacios, C., Haggård-Ljungquist, E. and Nilsson, A.S. (2011) Phylogenetic structure and evolution of regulatory genes and integrases of P2-like phages. *Bacteriophage*, **1**, 207–218.
24. Simons, R.W., Houman, F. and Kleckner, N. (1987) Improved single and multicopy lac-based cloning vectors for protein and operon fusions. *Gene*, **53**, 85–96.
25. Dodd, I.B., Perkins, A.J., Tsemitsidis, D. and Egan, J.B. (2001) Octamerization of lambda CI repressor is needed for effective repression of PRM and efficient switching from lysogeny. *Genes Dev.*, **15**, 3013–3022.
26. Dodd, I.B. and Egan, J.B. (2002) Action at a distance in CI repressor regulation of the bacteriophage 186 genetic switch. *Mol. Microbiol.*, **45**, 697–710.
27. St-Pierre, F., Cui, L., Priest, D.G., Endy, D., Dodd, I.B. and Shearwin, K.E. (2013) One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol.*, **2**, 537–541.
28. Palmer, A.C., Ahlgren-Berg, A., Egan, J.B., Dodd, I.B. and Shearwin, K.E. (2009) Potent transcriptional interference by pausing of RNA polymerases over a downstream promoter. *Mol. Cell*, **34**, 545–555.
29. Zwieb, C. and Adhya, S. (2009) Plasmid Vectors for the Analysis of Protein-Induced DNA Bending. In: Moss, T. and Leblanc, B. (eds). *Methods in Molecular Biology. DNA-Protein Interactions*. Humana Press, NY, Vol. **543**, pp. 547–562.
30. Sandaltzopoulos, R. and Becker, P.B. (1994) Solid phase DNase I footprinting: quick and versatile. *Nucleic Acids Res.*, **22**, 1511–1512.
31. Shearwin, K.E. and Egan, J.B. (2000) Establishment of lysogeny in bacteriophage 186. DNA binding and transcriptional activation by the CII protein. *J. Biol. Chem.*, **275**, 29113–29122.
32. Ackers, G.K., Johnson, A.D. and Shea, M.A. (1982) Quantitative model for gene regulation by lambda phage repressor. *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 1129–1133.
33. Dodd, I.B., Shearwin, K.E., Perkins, A.J., Burr, T., Hochschild, A. and Egan, J.B. (2004) Cooperativity in long-range gene regulation by the lambda CI repressor. *Genes Dev.*, **18**, 344–354.
34. Cui, L., Murchland, I., Shearwin, K.E. and Dodd, I.B. (2013) Enhancer-like long-range transcriptional activation by lambda CI-mediated DNA looping. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 2922–2927.
35. Hope, C.M., Rebay, I. and Reinitz, J. (2017) DNA occupancy of polymerizing transcription factors: a chemical model of the ETS family factor yan. *Biophys. J.*, **112**, 180–192.
36. Wu, H.M. and Crothers, D.M. (1984) The locus of sequence-directed and protein-induced DNA bending. *Nature*, **308**, 509–513.
37. Kim, J., Zwieb, C., Wu, C. and Adhya, S. (1989) Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. *Gene*, **85**, 15–23.
38. Callen, B.P., Shearwin, K.E. and Egan, J.B. (2004) Transcriptional interference between convergent promoters caused by elongation over the promoter. *Mol. Cell*, **14**, 647–656.
39. Johnson, A.D., Meyer, B.J. and Ptashne, M. (1979) Interactions between DNA-bound repressors govern regulation by the lambda phage repressor. *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 5061–5065.
40. Zimmerman, S.B. and Trach, S.O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J. Mol. Biol.*, **222**, 599–620.
41. Saha, S., Haggard-Ljungquist, E. and Nordstrom, K. (1989) Activation of prophage P4 by the P2 Cox protein and the sites of action of the Cox protein on the two phage genomes. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 3973–3977.
42. Yin, S., Bushman, W. and Landy, A. (1985) Interaction of the lambda site-specific recombination protein Xis with attachment site DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 1040–1044.
43. Esposito, D. and Scocca, J.J. (1997) Purification and characterization of HP1 Cox and definition of its role in controlling the direction of site-specific recombination. *J. Biol. Chem.*, **272**, 8660–8670.
44. Carlson, N.G. and Little, J.W. (1993) Highly cooperative DNA binding by the coliphage HK022 repressor. *J. Mol. Biol.*, **230**, 1108–1130.
45. Sanchez, A., Cattoni, D.I., Walter, J.C., Rech, J., Parmeggiani, A., Nollmann, M. and Bouet, J.Y. (2015) Stochastic self-assembly of ParB proteins builds the bacterial DNA segregation apparatus. *Cell Syst.*, **1**, 163–173.
46. Piazzola, D., Cali, S., Spoldi, E., Forti, F., Sala, C., Magnoni, F., Dehò, G. and Ghisotti, D. (2006) Expression of phage P4 integrase is regulated negatively by both Int and Vis. *J. Gen. Virol.*, **87**, 2423–2431.
47. Crooks, G., Hon, G., Chandonia, J. and Brenner, S. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.