

## Hypothesis

*cI* and *lexA* repressors consist of three *cro*-like domains

O.B. Ptitsyn, A.V. Finkelstein, M.P. Kirpichnikov\* and K.G. Skryabin\*

*Institute of Protein Research, USSR Academy of Sciences, 142292 Poustchino, Moscow Region and \*Institute of Molecular Biology, USSR Academy of Sciences, Moscow, USSR*

Received 27 July 1982

<i>cI</i> repressor	<i>lexA</i> repressor	<i>cro</i> repressor	<i>cro</i> -like domain	Operator region, $o_R$
	Secondary structure prediction		DNA binding	

## 1. INTRODUCTION

The switching of phages  $\lambda$  and 434 from the lysogenic state to the lytic growth or vice versa depends mainly on 2 repressors: *cro* and *cI*. This switching can be induced by external agents such as UV light. On the other hand, various DNA-damaging agents (including UV light) induce the 'SOS-system' of *Escherichia coli* [1] which is normally inhibited by *lexA* repressor. Among the stages of the 'SOS-system' work one includes *recA* protein activation which cleaves proteolytically *lexA* repressor [2,3]; the same protein cleaves also *cI* repressor [4,5].

The choice between lytic and lysogenic types of development of  $\lambda$ -like phages is due to the interaction of *cro* and *cI* repressors with operator regions of phage genome. The regulation takes place mainly on the operator region  $o_R$  which lies between genes *cI* and *cro* and controls 2 promoters,  $p_R$  and  $p_{RM}$ . The  $o_R$  region consists of 3 sites of specific binding ( $o_{R1}$ ,  $o_{R2}$  and  $o_{R3}$ ) which have similar primary structures with an approximate 2-fold internal symmetry [6–11]. Both *cro* and *cI* repressors interact specifically with  $o_{R1}$ ,  $o_{R2}$ ,  $o_{R3}$  having, however, different binding constants with different regions [12,13]. The *cI* repressor can bind cooperatively (due to protein–protein interactions) with 2 adjacent binding sites  $o_{R1}$  and  $o_{R2}$  (or  $o_{R2}$  and  $o_{R3}$ ). The binding of *cI* repressor to  $o_{R1}$  turns off the transcription (from  $p_R$  promoter) of early phage genes including the *cro* gene which are necessary for the lytic growth. On the other hand, the

binding of *cI* to  $o_{R2}$  (which is cooperative with binding to  $o_{R1}$ ) turns on (and, moreover, stimulates) the transcription of the *cI* gene from  $p_{RM}$  promoter. At the same time *cro* repressor does not bind cooperatively with the sites  $o_{R1}$ ,  $o_{R2}$ ,  $o_{R3}$  and its binding to  $o_{R3}$  turns off the transcription of *cI* gene from  $p_{RM}$  promoter [10,11]. *lexA* protein binds with 2 DNA regions (each with an approximate 2-fold internal symmetry) before the *lexA* gene and with one DNA region before the *recA* gene turning off the transcription of those genes [14,15].

*cI* repressors of  $\lambda$  and related phages exist in solution in an equilibrium between monomers and dimers but bind tightly to the operator only in a dimeric form [11,16] while *cro*  $\lambda$  probably exists in solution in a dimeric [17] or even tetrameric [18] form.

The three-dimensional structure of *cro*  $\lambda$  tetramers have been recently obtained from X-ray analysis [18]. Each monomer (66 residues) consists of a 3-stranded  $\beta$ -sheet ( $\beta$ -strands 2–6, 39–45 and 48–55) covered by a helical hairpin ( $\alpha$ -helices 7–14 and 27–36) with a third  $\alpha$ -helix [15–23] in the hairpin loop. A similar structure has been predicted (O.B.P., A.V.F., unpublished) from the *cro*  $\lambda$  amino acid sequence [19,20] before X-ray data became available. As has been shown by the high resolution NMR method and by optical methods, *cro*  $\lambda$  repressor in solution has an expressed tertiary structure, the secondary structure being similar to the crystal one (M.P.K., unpublished).

Below we predict the secondary structures of *cI*

repressor of phages  $\lambda$  and 434 and of *lexA* repressor of *E. coli* (as well as that of *cro* 434 repressor) and compare them with the secondary structure of *cro*  $\lambda$  predicted by the same method and known from X-ray data [18]. The results suggest that despite the large difference between the primary structures of these proteins, the secondary structure of *cro* 434 is similar to that of *cro*  $\lambda$ , and repressors *cI* and *lexA* consist of 3 domains each having a secondary structure (and thus, probably, also a tertiary one) similar to that of *cro*  $\lambda$ . The relative positions of these 3 domains will also be discussed.

## 2. METHOD

The usual method for searching the similarities between different proteins with unknown three-dimensional structures is the alignment of their primary structures. However, this method suffers from the mutation noise which can alter protein primary structures leaving unaltered their three-dimensional ones. Here, we use the alignment of the predicted secondary structures of the studied proteins rather than the alignment of their primary structures, considering that the similarity between the secondary structures of different proteins often means that their three-dimensional structures are similar also.

We have predicted secondary structures of 5 repressors from their reported primary structures: *cro*  $\lambda$  [19,20]; *cro* 434 [21,22]; *cI*  $\lambda$  [23]; *cI* 434 (R. Rogers Yocum, personal communication); and *lexA* [3]. The predictions have been made using a new computer algorithm (A.V.F., unpublished) based on the molecular theory of secondary structure formation [24]. According to this theory,  $\alpha$ -helices and  $\beta$ -strands are formed in the protein regions enriched by non-polar residues. The choice between an  $\alpha$ -helix and  $\beta$ -strand in each of these regions is determined mainly by helix-coil and  $\beta$ -coil equilibrium constants of its residues while the lengths of  $\alpha$ -helices and  $\beta$ -strands are determined by the lengths of their continuous hydrophobic surfaces [25]. The algorithm reliably distinguishes the irregular regions from the  $\alpha$ -helices and  $\beta$ -strands. The choice between the  $\alpha$ -helix and  $\beta$ -strand is somewhat less precise. Therefore we shall use the alternative predictions ( $\alpha$ -helix or  $\beta$ -strand) in all cases when the prediction is not completely unambiguous.

## 3. RESULTS

### 3.1. Inter- and intramolecular homologies of secondary structure

Fig.1 shows the alignment of predicted secondary structures of 5 repressors. Large repressors (*cI* and *lexA*) were each divided into 3 parts: the N-terminal (N), the middle (M) and the C-terminal (C) with lengths approximately equal to those of *cro*-repressors. One can see that the predicted secondary structures of these parts of all the 3 large repressors (*cI* 434, *cI*  $\lambda$  and *lexA*) are similar to each other and to the predicted secondary structures of both *cro*  $\lambda$  and *cro* 434. Moreover, the predicted secondary structures of all these 11 sequences are similar to the secondary structure of *cro*  $\lambda$  known from X-ray data [18] and shown at the top of fig.1: 3  $\alpha$ -helices ( $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$ ) and the  $\beta$ -hairpin ( $\beta$ -strands  $\beta_1$  and  $\beta_2$ ) of *cro*  $\lambda$  are predicted as the first choice in 8–10 sequences (of the 11) and as the second choice in almost all the remaining ones.

The relative positions of  $\alpha$ -helices and  $\beta$ -strands in these 11 sequences are also very similar: the alignment shown in fig.1 contains no deletions. The only degree of freedom being the shift of sequences as a whole.

### 3.2. Similarity of tertiary structures

The predicted secondary structures of all these 11 sequences do not have much freedom for the formation of their tertiary structures: the  $\alpha$ -helices of their N-terminal parts must form a helical complex and the two  $\beta$ -strands ( $\beta_1$  and  $\beta_2$ ) of their C-terminal parts must form an antiparallel  $\beta$ -hairpin which can also adjoin the N-terminal and/or the C-terminal  $\beta$ -strands referred to as  $\beta_0$  and  $\beta_3$ . The helical complex and the  $\beta$ -hairpin (or  $\beta$ -sheet) must join by their hydrophobic side groups forming a three-dimensional structure similar to the structure of the *cro*  $\lambda$  monomer known from the X-ray data [18].

We have come to the conclusion that *cro* 434 and each of the 3 parts of large repressors (*cI* and *lexA*) must have a tertiary structure similar (though probably not identical) to the structure of *cro*  $\lambda$  monomer. It is worthwhile to stress that *cro*-like structures are not only in the N-terminal parts of large repressors (which probably bind operators [5,11]) but also in their middle and C-terminal parts which do not bind operators but promote the

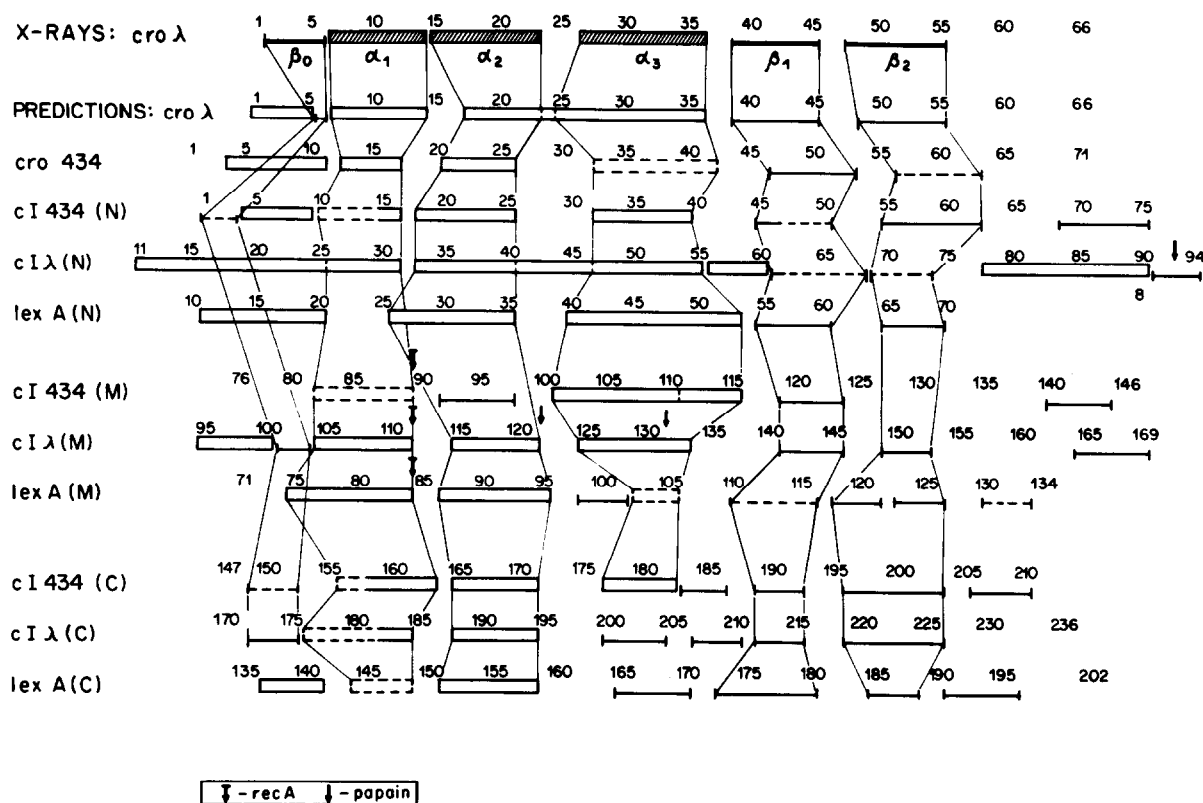


Fig.1. Alignment of the predicted secondary structures of *cro* and the domains of *cI* and *lexA*. Rectangles stand for  $\alpha$ -helices and solid lines for  $\beta$ -strands. The first choice of predictions is shown by solid lines and the second choice by dashed ones. The 'second choice' means that the alternative secondary structure is predicted as a more probable one for that chain region (e.g., for region 60-75 of *cI* λ the first choice is  $\alpha$ -helix and  $\beta$ -structure is only the second one). The X-ray secondary structure of *cro* λ [18] is shown at the top. Light lines connect the analogous elements of the secondary structures in the considered sequences. Arrows show the cleavage points by *recA* [2,3] and *papain* [4,5] (see text).

aggregation of subunits [5,11]. This rather surprising result can be explained if we remember that the *cro* λ protein has both these functions: it binds the operator and aggregates into dimers or tetramers [18].

The similarity between the structures of repressors *cI* and *lexA* is of special interest. The protein *recA* which switches off all these large repressors by cleaving their Ala-Gly bonds [2-5,11] cleaves all of them at exactly the same points of their secondary structures (just after the end of the M-domain  $\alpha_2$ -helix) despite the different positions of this bond in the primary structures of these proteins (89-90 in *cI* 434, 111-112 in *cI* λ and 84-85 in *lexA*).

### 3.3. Homology of primary structures

In the frames of the register for  $\alpha$ -helices and  $\beta$ -strands of these 11 sequences an additional fine alignment can be made to obtain the best homologies of their primary structures. Good homologies between different proteins can be obtained with only few deletions (fig.2). These homologies include all those in [11,22] as well as some new ones. The most striking homology is between *cro* 434 and the N-part of *cI* 434 [11]. However, there are also remarkable homologies between *cro* repressors of phages λ and 434 as well as between *cI* repressors of these phages. There is also a considerable homology between *E. coli* *lexA* repressor and *cI* phage repressors especially in their M- and

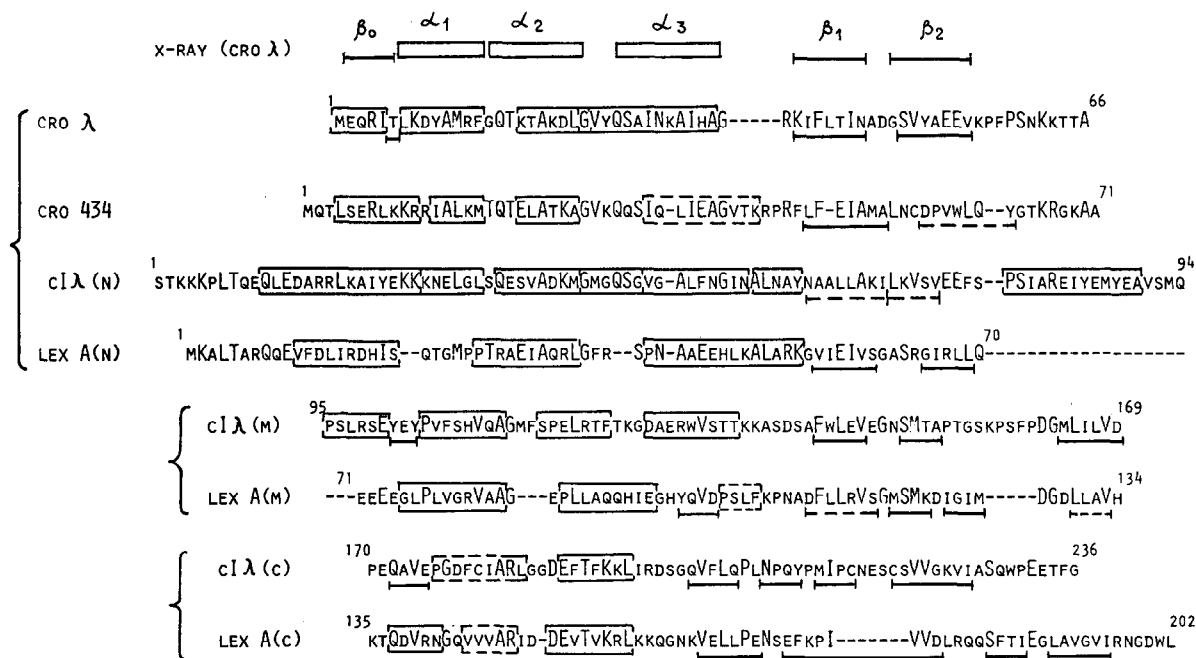


Fig.2. Alignment of the primary structures of 8 sequences in the frames of the alignment of their secondary structures. Amino acid residues common for at least 2 sequences in each of the 3 groups (*cro* and N-domains, M-domains and C-domains) are shown by large letters. All other notations are as in fig.1.

C-parts. The point is that the method of establishing these homologies (from the similarity of secondary structures) suggests that they are not occasional but important for the conservative three-dimensional structure of the proteins studied. We have not found any significant intramolecular homologies between different domains of large repressors as well as between *cro* repressors and M- or C-domains of the large ones.

One of the possible explanations of the observed homologies between primary structures is the existence of their common ancestor; however, there is no direct evidence in favour of this suggestion.

#### 4. DISCUSSION

The data on papain splitting of *cI* λ and the temperature melting of *cI* λ and its papain fragments [5] lead to the conclusion [5] that *cI* λ includes 2 domains, 1–92 and 132–236. However, the experimental data in [5] are more consistent with our 3-domain model. Moreover, in conjunction with the X-ray data on the tertiary structure of *cro* λ tetramers [18] these data permit us to speculate on the relative positions of 3 *cro*-like domains in *cI* dimers

and, probably, also in *lexA*.

The first splitting point by papain (92–93) is almost exactly at the end of our N-terminal domain. Both fragments 1–92 and 93–236 are relatively stable to further papain splitting and have a cooperative temperature of melting. The fragment 132–236 (which also melts cooperatively) appears after 1 h digestion and remains the only stable

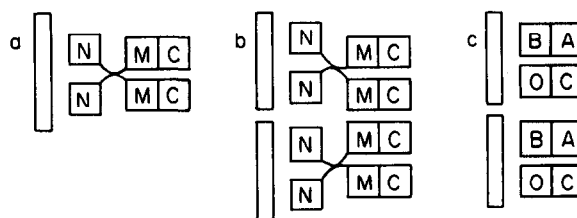


Fig.3. Schematic presentation of the structure of the *cI* dimer (a) and of the cooperative binding of 2 dimers to 2 adjacent operator sites (b). The non-cooperative binding of 2 *cro* λ tetramers [18] is shown for comparison (c). Operator sites are shown as rectangles and the domains of *cI* λ and subunits of *cro* λ (O, A, B and C [18]) as squares.

fragment after 10–22 h (together with traces of fragment 1–92). It is clear, therefore, that we are dealing with 3 (rather than 2) stable fragments: 1–92 (our N-domain); 93–236 (the complex of our M- and C-domains); and 132–236 (which is our C-domain together with the C-terminal  $\beta$ -region of the M-domain). This suggests that the C-terminal  $\beta$ -strands of the M-domain can form a common  $\beta$ -sheet with the  $\beta$ -strands of the C-domain (a similar common  $\beta$ -sheet has been found between molecules O and C or B and A of *cro*  $\lambda$  tetramer [18]). In [5] fragments 1–92 of *cI*  $\lambda$  bound specifically to DNA and fragments 93–236 and 132–236 oligomerized.

Given these data we can propose the principle scheme of *cI* dimer structure in fig.3a. This scheme implies that M- and C-domains are tightly bound both intra- and intermolecularly and that N-domains are practically not bound to M- or C-domains. Fig.3b shows how this scheme of *cI* dimer structure can explain their cooperative binding to 2 adjacent operators in contrast with the non-cooperative binding of *cro* tetramers or dimers (fig.3c). Our alignments (fig.1, 2) imply that the N-terminal 17 residues of *cI*  $\lambda$  and 14 residues of *lexA* have no equivalent in *cro*  $\lambda$  and therefore could play some special role in DNA binding.

## ACKNOWLEDGEMENTS

The authors express their gratitude to Professor A.A. Bayev for his continuous interest to the work and useful discussions. The authors thank Drs A.G. Murzin and T.A. Erokhina for valuable help. The authors are very grateful to Dr R. Rogers Yocum for the presentation to Dr K.G. Skryabin the unpublished primary structure of *cI* 434 repressor.

## REFERENCES

- [1] Gottesman, S. (1981) *Cell* 23, 1–2.
- [2] Little, J.W., Edmiston, S.H., Pacelli, L.Z. and Mount, D.W. (1980) *Proc. Natl. Acad. Sci. USA* 77, 3225–3229.
- [3] Markhom, B.E., Little, J.W. and Mount, D.W. (1981) *Nucleic Acids Res.* 9, 4149–4161.
- [4] Roberts, J.W., Roberts, C.W. and Craig, N.L. (1978) *Proc. Natl. Acad. Sci. USA* 75, 4714–4718.
- [5] Pabo, C.O., Sauer, R.T., Sturtevant, J.M. and Ptashne, M. (1979) *Proc. Natl. Acad. Sci. USA* 76, 1608–1612.
- [6] Ptashne, M., Backman, K., Humayun, M.Z., Jeffrey, A., Maurer, R., Meyer, B. and Sauer, R.T. (1976) *Science* 194, 156–161.
- [7] Maurer, R., Meyer, B.J. and Ptashne, M. (1980) *J. Mol. Biol.* 139, 147–161.
- [8] Meyer, B.J., Maurer, R. and Ptashne, M. (1980) *J. Mol. Biol.* 139, 163–194.
- [9] Meyer, B.J. and Ptashne, M. (1980) *J. Mol. Biol.* 139, 195–205.
- [10] Ptashne, M., Jeffrey, A., Johnson, A.D., Maurer, R., Meyer, B.J., Pabo, C.O., Roberts, T.M. and Sauer, R.T. (1980) *Cell* 19, 1–11.
- [11] Johnson, A.D., Poteete, A.R., Lauer, G., Sauer, R.T., Ackers, G.K. and Ptashne, M. (1981) *Nature* 294, 217–223.
- [12] Johnson, A.D., Meyer, B.J. and Ptashne, M. (1978) *Proc. Natl. Acad. Sci. USA* 75, 1783–1787.
- [13] Takeda, Y. (1979) *J. Mol. Biol.* 127, 177–191.
- [14] Little, J.W., Mount, D.W. and Yanisch-Perron, C.R. (1981) *Proc. Natl. Acad. Sci. USA* 78, 4199–4203.
- [15] Brent, R. and Ptashne, M. (1981) *Proc. Natl. Acad. Sci. USA* 78, 4204–4208.
- [16] Chadwick, P., Pirrotta, V., Steinberg, R., Hopkins, N. and Ptashne, M. (1970) *Cold Spring Harb. Symp. Quant. Biol.* 35, 283–294.
- [17] Takeda, Y., Folkmanis, A. and Echols, H. (1977) *J. Biol. Chem.* 252, 6177–6183.
- [18] Anderson, W.F., Ohlendorf, D.H., Takeda, Y. and Matthews, B.W. (1981) *Nature* 290, 754–758.
- [19] Hsiang, N.E., Cole, R.D., Takeda, Y. and Echols, H. (1977) *Nature* 270, 275–278.
- [20] Shimatake, H., Brady, C. and Roseberg, M. (1977) *Nature* 270, 274–275.
- [21] Ovchinnikov, Yu.A., Guryev, S.O., Krayev, A.S., Monastyrskaya, G.S., Skryabin, K.G., Sverdlov, E.D., Zakharyev, V.M. and Bayev, A.A. (1979) *Gene* 6, 235–249.
- [22] Grosscheld, R. and Schwarz, E. (1979) *Nucleic Acids Res.* 6, 867–881.
- [23] Sauer, R.T. and Andereg, R. (1978) *Biochemistry* 17, 1092–1100.
- [24] Ptitsyn, O.B. and Finkelstein, A.V. (1979) *Int. J. Quant. Chem.* 16, 407–418.
- [25] Ptitsyn, O.B. and Finkelstein, A.V. (1980) *Quart. Rev. Biophys.* 13, 339–386.