Volume 231, number 2, 291–298 FEB 05726 April 1988

Hypothesis

Generation of high specificity of effect through low-specificity binding of proteins to DNA

Emile Zuckerkandl and Ruxton Villet[+]

Linus Pauling Institute of Science and Medicine, 440 Page Mill Road, Palo Alto, CA 94306 and [+] *USDA, Agricultural Research Service/NPA, Bldg 055, BARC-West, Beltsville, MD 20705, USA*

Received 25 January 1988

It is proposed that proteins can bind with relatively low-affinity and specificity to multiple sites, defined as sequence motifs, on polynucleotide chains, and that such binding can collectively be turned into high-affinity, high-specificity binding through cooperative effects, especially when the sequence motifs recur periodically. The selection of individual nucleotides has in general been thought to be the condition of the existence and conservation of function in most of the noncoding sequences. This condition seems unnecessary. Calculations are presented as a step in the direction of giving credibility to a model of stable gene repression.

Gene regulation; Noncoding sequence; Sequence motif; Cooperative binding; Distributive specificity; *c*-Value paradox; Genetic load; Determination

1. INTRODUCTION

Most DNA of higher organisms is neither coding nor involved in highly specific protein binding. This is one reason why most eukaryote DNA has been considered to be 'junk' [1]. This reason would be weakened and investigations into possible functional roles of 'excess' DNA, or of part of it, would be encouraged if low-affinity binding of individual protein molecules to polynucleotide, with binding constants not far above the value for nonspecific binding, could result in specific protein-polynucleotide complexes, which would then be available for exercising specific regulatory effects. Calculations presented here, using a classical model for cooperative binding, support the hypothesis that relatively low-affinity binding of proteins to individual polynucleotide receptor sites can, through cooperativity, lead to a high saturation of binding sites with ligand. This implies that, given a proper distribution of binding sites, low-affinity binding is compatible with precisely positioned high order structures, such as would be required if relatively long sectors of noncoding DNA or RNA sequences filled certain regulatory roles.

It has been asserted that a low specificity of binding to DNA by chromosomal proteins is indeed likely to be compatible with precise regulatory effects, provided that the number of binding sites is relatively large [2]. Such multi-site binding, referred to as mass binding [2], could as well involve long primary RNA transcripts, determine their conformation, and affect processing and perhaps other functions of the transcripts. In the case of DNA, documented examples of mass-binding proteins include histone H1 [3,4], α-protein [5], and high mobility group proteins 1, 2, 14 and 17 [6].

Correspondence address: E. Zuckerkandl, Linus Pauling Institute of Science and Medicine, 440 Page Mill Road, Palo Alto, CA 94306, USA

Published by Elsevier Science Publishers B.V. (Biomedical Division)

2. EXPERIMENTAL EVIDENCE FOR DIFFERENTIAL EFFECTS OF MASS BINDING

Experimental encouragement of this concept is indirect. The existence of Bernardi et al.'s [7] base-compositional isochores in vertebrates is compatible, in different parts of the genome, with different modes of the interaction of DNA with proteins that are bound with moderate affinity. The modes in question are expected to relate (i) to preferred DNA-binding protein variants which, in a given compositional sector of DNA, will displace most molecules of other variants that have a lesser average affinity for the sector; (ii) to the average affinity constants of these higher-affinity protein variants binding to the compositional sector; and (iii) to the average distance between binding sites belonging to a subpopulation of sites that most strongly bind protein variants with affinities in the upper range of moderate-affinity multi-site binding.

Isochores have been found in warm-blooded vertebrates. They may exist also in Diptera (*Sciara*), even though GC content in different sectors of *Sciara* euchromatin does not seem to vary considerably [8]. Nonetheless different polytene chromosome bands do not bind anti-adenine and anti-cytosine antibodies in the same ratios [8], suggesting different average base compositions for different bands. Furthermore in polytene chromosomes of *Drosophila*, many bands display individual reactivities to ionically altered environments [9,10]. Such observations point to the possibility that different representatives of a family of multi-site regulator proteins bind preferentially to different sectors of the genome of the size of a polytene chromosome band or, in higher organisms, of the larger G-bands or R-bands [11], and that the same proteins could bind with different average affinities to different bands characterized by different predominant sequence motifs. A sequence motif [2] is a short run of nucleotides, a certain fraction of which is free to vary from repeat to repeat, while another fraction, not always in the same positions, remains constant.

It has been shown in transfection experiments that genes contained in a segment of about 20 kb, integrated into the genome of a foreign host cell, can be co-regulated. Three genes, which had been brought together from different sources and ligated, were found, according to the state of the cell, to be either collectively transcribed or collectively repressed [12]. By analogy with similar work, it could be inferred [12] that this switch was associated with a change in DNase I sensitivity and therefore (e.g. [13]) with a change in chromatin structure. Changes in the mass binding of proteins may be presumed to be responsible for these changes in chromatin structure and gene expression.

It should be noted, however, that the inserted 20 or so kb fragment was compositionally heterogeneous, since the genic regions that were fused before transfection are known from another source [7,14] to have different GC contents. A uniform behavior with respect to transcription of the fragment as a whole is not expected, if different GC contents of adjacent sequence sectors do lead, in a given cell, to different stabilities of the highest order deoxyribonucleoprotein structures. There are several possible reasons why a 20 kb-long, compositionally heterogeneous higher-order structure could be destabilized in toto and therefore its DNA made available for transcription. For instance, a destabilization of part of the higher order structure may be transmitted to other parts, when these are only a few kilobase pairs long. Also, regulatory specificity relationships may have been peculiar in the experiment under discussion, since the sources of the genes were human and herpesvirus, whereas the host cells were from mouse. Furthermore, one must remember that changes in mass binding may be at times brought about by a local change in punctate binding, one that results in a phase shift of binding opportunities over a sector of DNA and in a switch in the mode of mass binding. Punctate binding [2] has been defined as the binding with high affinity of a regulator protein to one or a few often contiguous or at any rate not distant (e.g. [15]) polynucleotide sites. If a phase shift in mass binding occurred through punctate binding, the position of the protein/polynucleotide complexes would not strongly depend on base composition and sequence over a certain sequence sector [16].

Though punctate regulation also involves multi-site protein binding [17], it is distinct from mass binding in that sites for punctate binding occur in

small clusters. With the exception of the binding of protein by enhancers, which again represent small clusters of protein-DNA complexes [17], the clusters occur in the almost immediate neighborhood of the coding sequences that they control. Mass binding involves much larger sectors of DNA and is hypothesized and, in the case of some proteins (histones), known to be periodical. It is important to recall, however, that at the level of *primary* structure, this periodicity can be inexact, in that the length of the linker sequence, the sequence that links the nucleosomal cores to one another, may vary between adjacent nucleosomes [20,21]. (Such variation is apparently not reflected in the next higher order structure, the nucleosome solenoid [22].) The variation may however occur primarily by increments of about 10 bp and may not frequently exceed ± 10% of the locally predominant length of DNA per nucleosome [21]. Thus linear periodicity is qualified rather than eliminated.

One may again tentatively attribute to mass binding the observation that genes, when transferred into cells in which they are not active, can be expressed in these cells, whereas the corresponding endogenous cellular gene remains repressed [23]. The expression of the transfected gene must be attributable to at least one cellular protein, one that presumably acts through punctate binding. The activity of the transfected gene implies that this regulator protein is present in the cell. Therefore the endogenous gene, or merely its associated punctate receptor sites, are apparently inaccessible to the regulatory protein that has access to the exogenous gene. This state of inaccessibility of the endogenous gene and its lasting repression can be attributed to a higher order regional chromatin structure that is formed and maintained through the interaction between the DNA and certain mass-binding proteins. The same interpretation can tentatively be given of the observation that at gastrulation the oocyte 5 S RNA genes of *Xenopus* cease being expressed even in the presence of an excess of transcription factor TFIIIA. Up to that stage the presence of excess transcription factor indeed leads to an excess in transcription [24]. Later it becomes unable to promote transcription, as if at gastrulation the transcription factor's binding site became inaccessible.

Further indirect encouragement for the concept of a regulatory function of mass-binding comes from the finding that supranucleosome structures encompassing at least 20–40 kb, and thus approaching, or representing, the size of chromatin domains (loops) (e.g. [25]) exist in two structurally distinct types, of which one contains transcriptionally active, the other transcriptionally repressed genes [26]. The large size alone of these chromatin fragments suggests the possibility of structures of an order higher than that of the nucleosome solenoid.

Some mass-binding proteins whose presence locally in chromatin correlates with the release of repression of a gene have been implicated, notably HMG, high mobility group, proteins 14 and 17 [27], though their role is not clear. These HMG proteins do not appear to bind cooperatively to the DNA-protein complex of chromatin [6]. They may however interfere with the cooperative binding of certain 'compaction proteins', of which histone H1 is an example [3,4]. The repressed complex in which genes are incapable of responding to punctate regulators of transcription, a complex that we presume dependent on cooperativity, is stable and transmissible by cells to a number of cell generations, yet reversible under certain circumstances [28].

States of chromatin defined by other criteria share these characteristics. One is the state corresponding to the so-called intermediate sensitivity to DNase I. It is also intrinsically stable and transmissible over a number of cell generations [29], though subject to modulation [30,31]. In regions of intermediate sensitivity to DNase I, mass binding of proteins may again be presumed to be implicated, since the structural state or states involved may extend to sectors of DNA of 100 kb or more. The stabilization of any particular structure over so long a segment of DNA is, in principle, best accounted for by multiple protein/protein/DNA interactions.

Paigen's group [32] considered mass binding of regulator protein molecules as the most likely explanation both of the lag period in transcriptional activation of the β-glucuronidase gene in mice and, following the lag period, of the progressively increasing rate of transcription of this gene.

All in all, mass binding appears to be an indispensable determinant, or participant in the determination, of certain structural modes of

chromatin, representing stages of structural preparedness for functional action or, on the contrary, stable repression.

3. THE ALLOSTERIC BASIS FOR MULTI-SITE BINDING IN GENE REGULATION

The positioning along chromatin of certain DNA-protein complexes is expected frequently to depend on the number and distribution of certain sequence motifs [2]. Individual protein molecules of the mass-binding type might bind to such motifs with an equilibrium constant not very much higher than 10^{-6} M^{-1} [5,33], a value in the range of those characteristic of nonspecific protein-DNA complexes, as against, for instance a value of 10^{-12} M^{-1} in specific binding [34]. The stability of the higher order structure formed as a consequence of the occupancy of most of the relatively low-affinity binding sites in a given region of chromatin would of course be considerably increased by cooperative effects. That there is cooperativity in oligonucleosome-histone H1 interaction has been demonstrated [3]. The mass binding of the HMG-like α-protein to satellite DNA is probably also cooperative [5].

An example both of the existence of a cooperative binding effect and of periodically recurring sequence motifs on DNA (very short ones in this case) is, in a sense, provided by the formation of the nucleosomes themselves. To bind to DNA with a high affinity, the four core histones must, first, form the histone octamer. The octamer then is able to bind preferentially to the periodically recurring dinucleotides AA or TT [35,36].

Consider a Monod-Wyman-Changeux type model for cooperative binding [37]. For simplification, assume that the ligand binds well-nigh exclusively to the R conformational state for n binding sites:

$$\bar{Y}_F = \frac{\alpha(1+\alpha)^{n-1}}{(1+\alpha)^n + L}$$

where $\bar{Y}_F$, the equilibrium fractional saturation with respect to ligand F (the chromosomal regulator protein); α, $[F]/K_R$ where K_R is the dissociation constant (reciprocal of affinity) for ligand binding to the R state of the receptor DNA molecule; L, the conformational equilibrium, T/R. If the receptor begins chiefly in the T state, then L is very large.

It is instructive to calculate values of $\bar{Y}_F$ as a function of α, using various values of n and L. Consider the effect of varying n at a constant and large value of $L = 10000$. In fig.1 results of calculations are shown for the following values of n: (a) 4; (b) 6; (c) 10.

For a given value of α (for example, $\alpha = 5.0$), as the number of binding sites is increased from 4 to 6 and then to 10, the equilibrium fractional saturation increases from 0.10 to 0.67 and then to 0.83, respectively. Moreover, a particular fractional saturation is achieved with successively lower values of α, as the number of sites is increased: for example, half-saturation ($\bar{Y}_F = 0.5$) is achieved at $\alpha = 9.6$, 4.0 and 1.9 for $n = 4$, 6, and 10, respectively. Since $\alpha = [F]/K_R$, low-affinity binding (that is, with a large value for K_R) nevertheless can give substantial values for $\bar{Y}_F$, provided n is increased.

In fig.2 the effect of L is shown at a constant number of binding sites, $n = 4$. Clearly, for the same number of sites the fractional saturation increases markedly as L is decreased. This is because of the preferential binding to the R state; less 'pulling' of protomers from the T to the R state is needed at low L.

To relate these values to a real physiological situation it is instructive to examine measured values for ligand concentration and dissociation constant. In preliminary experiments by Strauss and Varchavsky [5], for nuclear proteins in green monkey cells [F] = 8 nM and K_R = 10 nM, giving a value for α of 0.8. This is within the range of values for α considered in this study.

The following question arises: in order to achieve the same fractional saturation as when ligand is bound at a small number of sites, but tightly, what reduction in affinity could be tolerated by using a larger number of binding sites? Compare, as an example, binding with one site against that with 10 sites. For $n = 1$, $L = 10000$ and $\alpha = 50000$ a fractional saturation of 0.83 is calculated, using the Monod-Wyman-Changeux model for fractional saturation. From fig.1, curve c ($n = 10$; $L = 10000$) a value of $\bar{Y}_F = 0.83$ is achieved with $\alpha = 5.0$. Thus, at a constant ligand concentration with 10 sites the affinity per site can be lowered by four orders of magnitude and yet

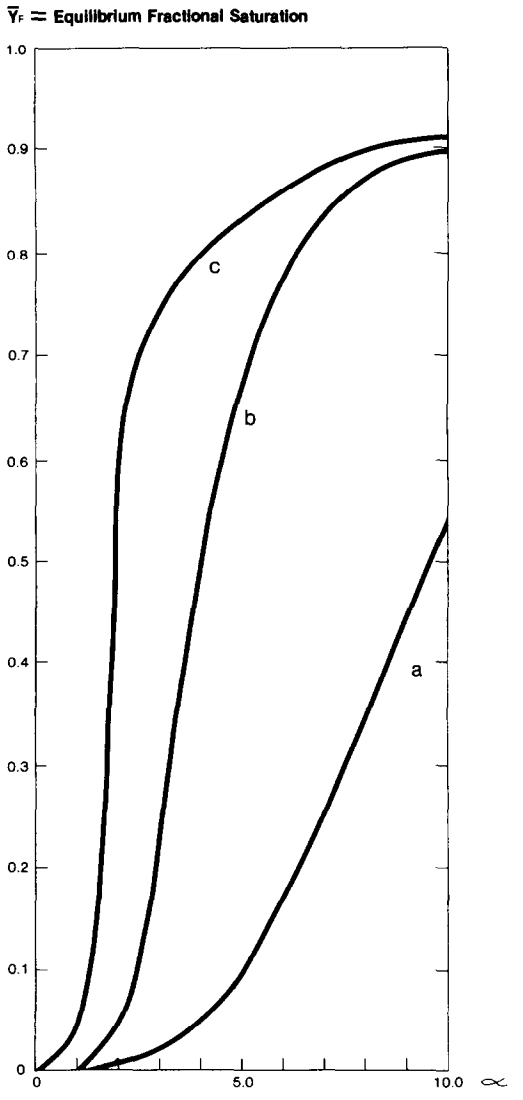

Fig.1. Effect of n, number of binding sites, on equilibrium fractional saturation. Plots of $\bar{Y}_F$, equilibrium fractional saturation of regulatory proteins bound to DNA, as a function of $\alpha = [F]/K_R$, where [F] is the concentration of proteins and K_R the dissociation constant. The curves represent calculations using the Monod-Wyman-Changeux model [39] for cooperative interaction. L, the conformation equilibrium, is constant at 10000. n is varied as follows: (a) 4; (b) 6; (c) 10.

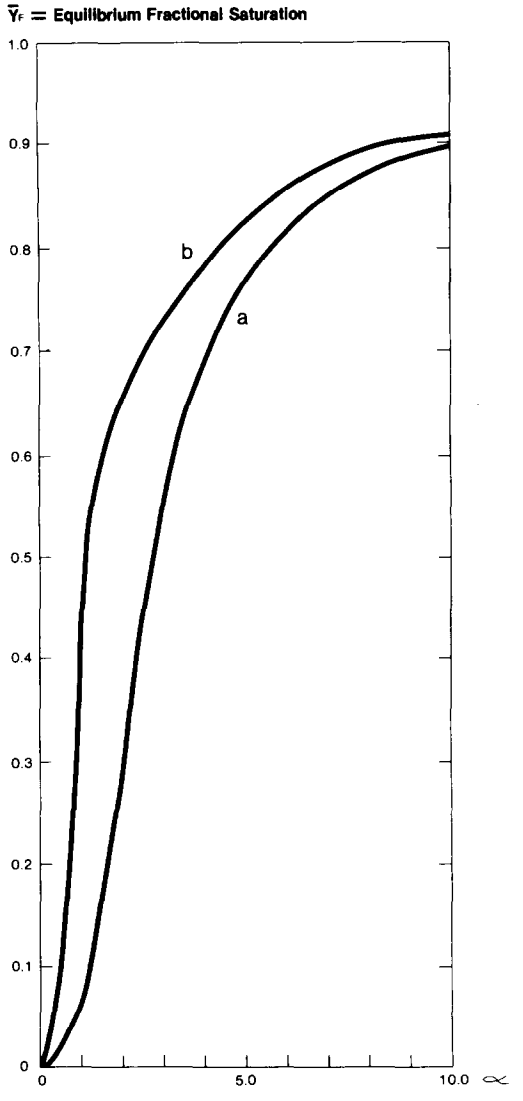

Fig.2. Effect of L, conformational equilibrium, on fractional saturation. From the results of calculations $\bar{Y}_F$ is plotted for $n = 4$ and for the following values of L: (a) 100; (b) 1.

give the same fractional saturation as with a single tightly bound site. This lends credence to the notion of a role of multi-site binding in gene regulatory processes involving noncoding as well as coding DNA.

4. GENERAL DISCUSSION

One of us [2] has proposed that part of noncoding DNA sequences in higher organisms fill a function, namely that of providing for a certain family of regulatory proteins multiple periodically recurring binding sites, in the form of sequence motifs whose sequence requirements are rather relaxed and with a tolerance for the skipping of

some periods (e.g. in exons). Such units of regulation might encompass about 10 to 100 kb, or more, with 10 kb possibly representing a critical size in regard to the stability of a high order structure [38]. Protein/protein/DNA interactions in such mass binding would permit the formation or the abolition of certain high order structures and thereby make sites respectively inaccessible or available to local (punctate) molecular signals. The calculations presented here show such a concept to be compatible with the cellular parameters considered.

The specificity and stability of a mass binding complex requires a quasi-periodic distribution of binding sequence motifs between two boundaries, on each side of which these sequence motifs are no longer present or are distributed too sparsely or not with appropriate periodicity. Periodically recurring sequence motifs, in addition to overlapping and interperiodic motifs, are indeed expected to be found provided that the sequence requirements be sufficiently relaxed and in accord with a regional average base composition. Under such conditions, the presence of a periodically recurring sequence motif is statistically probable and therefore not considered as 'significant'. One may, however, focus on a different concept of significance, because there is no strong reason to anticipate that only improbable sequence motifs are potentially functional. One may consider that statistically probable periodic recurrences of sequence motifs, probable on the basis of the regional base composition, could be at the same time statistically nonsignificant and functionally significant. (Obviously the statistical investigations of sequences could not demonstrate nor even suggest the functional significance of a statistically nonsignificant recurrence of motifs; only experimental determinations could.) For certain regulatory functions, the genome would use probable, rather than improbable sequences. Through the control of sectorial composition, rather than of precise sequence, mechanisms for rigorous regional gene control, in particular repression, could thus be provided at low selective cost, i.e. with little increase in genetic load. This mode of gene control could be based on primary structural features that are largely random, because *highly relaxed sequence requirements are probably compatible with stringently determined higher order structures, thanks to the amplifying effect on overall intermolecular affinity obtained through cooperative binding.*

The specificity of the effect would involve a relatively large segment of DNA or RNA, of the order of a number of kilobases. It would be of a type that has been referred to as distributive specificity [39]. Namely, it would depend on the distribution of sequences rather than on their precise nature, with boundaries set to the sector of cooperative interaction by a discontinuation, on both sides of the interaction zone, of the recurrence of the sequence motif at sufficiently short periodic frequencies. In this fashion the control of gene expression could in part be accomplished through the approximate conservation of sectorially distinct base compositions and through switches to different average base compositions in the flanking sectors. (Cooperative binding of protein (e.g. a phage repressor) to noncontiguous sites can also occur through the formation of DNA loops [40]. This kind of cooperative binding is unlikely to involve more than a small number of interacting DNA sites and may have a functional potential different from the one offered by periodically recurring binding sites. There is no apparent reason for anticipating that such loops are of equal size, i.e., that the recurrence of these interacting sites is periodical. Presumptive sites for this type of interaction therefore could not be discovered by the study of DNA sequences, when the sequence specificity is low. Postulating periodicities (with a moderate tolerated variance) is heuristically more promising.)

The presence of a few 'favorable' bases at appropriate positions along a DNA sector may be thought locally to increase the average equilibrium constant of single protein regulator/polynucleotide receptor pairs or protein/protein/polynucleotide receptor complexes over the equilibrium constant for 'nonspecific binding' and to do so to a moderate, yet sufficient extent for initiating the zipper actions of cooperative binding. In this fashion a specific function of DNA could be compatible with sequences of very low specificity [41].

One such function can be tentatively proposed. Cellular commitment or determination (not distinguished here) is in part characterized by the permanent repression of certain genes. In higher organisms this repression can be strikingly effec-

tive. In the case of the growth hormone gene in rat, Ivarie et al. [42] have found that in tissue in which this gene is repressed (liver) there are at most only about four growth hormone molecules per cell, corresponding to about 0.01 to 0.001 mRNA molecules per cell. Repression in prokaryotes appears to be less effective by five to six orders of magnitude. We propose here that this difference in effectiveness of repression is in large measure accounted for by the difference between punctate binding and mass binding, either at the level of primary gene transcripts [43,44] or at the level of transcription, or at both. The case of primary RNA transcripts, in which mass binding of proteins also occurs [45,46], has not been addressed here.

In DNA, the mechanism of quasi-periodic mass binding might play a regulatory, as distinguished from a merely packaging, role. This may apply to only a fraction of the genes, in particular, though not necessarily exclusively, to genes expressed in terminal differentiation, whose expression is repressed in most tissues. Heat shock genes might be another case in point, since in *Drosophila* the chromatin region defined by the hsp 70 sequences becomes DNase I- and micrococcal nuclease-sensitive upon heat shock [20,47], the sensitivity to DNase I being of the 'intermediate' type. Such a process implies a structural change which in turn must involve a change in protein binding. Extraction experiments involving mass-binding proteins have led to the conclusion that a deoxyribonucleoprotein structure higher than that of the 30 nm fiber could be involved in the appearance of intermediate DNase I-sensitivity in the avian β-globin gene complex [48]. Such structures and, correlatively, sequences of DNA of the order of 50 or more kilobases as well as mass binding of proteins have thereby been implied to function, as predicted [2], in cellular determination, namely in the determination of avian red cells.

Such a function of a fraction of the noncoding sequences in eukaryotes would obviously represent a partial solution to the *c* value paradox, presumably as a complement to the solution proposed by Cavalier-Smith [49], who points out correlations between *c* value and certain general cellular and organismal parameters.

Sites that bind a regulator protein with high affinity can be presumed to be homologous. Shared high affinity and specificity indeed suggest extensive sequence similarity. In mass binding, when the protein binds with a much lower affinity and specificity to the *individual* polynucleotide receptor sites, even though the sites are expected to share certain similarities, their chances of being homologous are smaller.

Prokaryotes and eukaryotes share punctate binding. On the other hand, cooperative mass binding, even though perhaps not 'invented' by eukaryotes, may be thought to have been greatly developed in the line leading to the eukaryotes or during their own evolution.

In summary, available data as well as the calculations presented here are compatible with the view that the *c* value paradox can be partly accounted for on grounds of molecular function, that genomic niches for 'junk' DNA may well be less extensive than has been claimed, and that cooperative mass binding of proteins to polynucleotide provides a reasonable, if hypothetical, general molecular basis for mechanisms of cellular determination.

Acknowledgements: The authors express their appreciation to Drs E.N. Trifonov and T. Boulikas for their critical reading of the manuscript and their suggestions. This work was in part supported by the Japan Shipbuilding Industry Foundation.

REFERENCES

[1] Ohno, S. (1972) in: Evolution of Genetic Systems (Smith, H.H. ed.) pp.366–370, Gordon-Breach, New York.
[2] Zuckerkandl, E. (1981) Mol. Biol. Rep. 149, 158.
[3] Renz, M., Nehls, P. and Hozier, J. (1977) Proc. Natl. Acad. Sci. USA 74, 1879–1883.
[4] Felsenfeld, G. (1978) Nature 271, 115–117.
[5] Strauss, F. and Varshavsky, A. (1984) Cell 37, 889–901.
[6] Reeck, G.R. and Teller, D.C. (1985) in: Progress in Nonhistone Protein Research, vol.II, pp.1–21, CRC Press, Boca Raton, FL.
[7] Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. and Rodier, F. (1985) Science 228, 953–958.
[8] Eastman, E.M., Goodman, R.M., Erlanger, B.F. and Miller, O.J. (1980) Chromosoma (Berlin) 79, 293–314.
[9] Lezzi, M. and Robert, M. (1972) in: Developmental Studies on Giant Chromosomes (Beermann, W. ed.) pp.35–57, Springer, New York.
[10] Kroeger, H. and Muller, G. (1973) Exp. Cell Res. 82, 89–94.
[11] Holmquist, G. (1988) in: Chromosomes and Chromatin, vol.II, CRC Press, Boca Raton, FL, in press.

[12] Roginski, R.S., Skoultchi, A.I., Henthorn, P., Smithies, O., Hsiung, N. and Kucherlapati, R. (1982) Cell 35, 149–155.
[13] Weisbrod, S. (1982) Nature 297, 289–295.
[14] Bernardi, G. and Bernardi, G. (1986) J. Mol. Evol. 24, 1–11.
[15] Myers, R.M., Rio, D.C., Robbins, A.K. and Tjian, R. (1981) Cell 25, 373–384.
[16] Kornberg, A. (1981) Nature 292, 579–580.
[17] Echols, H. (1986) Science 233, 1050–1056.
[18] Serfling, E., Jasin, M. and Schaffner, W. (1985) Trends Genet. 224–230.
[19] Herr, W. and Clarke, J. (1986) Cell 45, 461–470.
[20] Wu, C. (1980) Nature 286, 854–860.
[21] Samal, B. and Worcel, A. (1981) Cell 23, 401–409.
[22] Widon, J., Finch, J.T. and Thomas, J.O. (1985) EMBO J. 4, 3189–3194.
[23] Robins, D.M., Paek, I., Seeburg, P.H. and Axel, R. (1982) Cell 623, 623–631.
[24] Andrews, M.T. and Brown, D.D. (1987) Cell 51, 445–453.
[25] Hancock, R. and Boulikas, T. (1982) in: International Review of Cytology, vol.79, pp.165–215, Academic Press, New York.
[26] Weintraub, H. (1984) Cell 38, 17–27.
[27] Weisbrod, S., Groudine, M. and Weintraub, H. (1980) Cell 19, 289–301.
[28] Chiu, C.P. and Blau, H.M. (1984) Cell 37, 879–887.
[29] Zaret, K.S. and Yamamoto, K.R. (1984) Cell 37, 29–38.
[30] Stalder, J., Groudine, M., Dodgson, J.B., Engel, J.D. and Weintraub, H. (1980) Cell 19, 973–980.
[31] Weintraub, H., Larsen, A. and Groudine, M. (1981) Cell 24, 333–344.
[32] Watson, G., Davey, R., Labarca, C. and Paigen, K. (1981) J. Biol. Chem. 256, 3005–3011.
[33] Robert Tjian, personal communication.
[34] Wu, C., Wilson, S., Walker, B., Dawid, I., Paisley, T., Zimarino, V. and Ueda, H. (1987) Science 238, 1247–1252.
[35] Trifonov, E.N. (1980) Nucleic Acids Res. 8, 4041–4053.
[36] Mengeritsky, G. and Trifonov, E.N. (1984) Cell Biophys. 6, 1–8.
[37] Monod, J., Wyman, J. and Changeux, J.P. (1965) J. Mol. Biol. 12, 88–105.
[38] Butler, P.J.G. and Thomas, J.O. (1980) J. Mol. Biol. 144, 89–93.
[39] Zuckerkandl, E. (1974) Biochimie 56, 937–954.
[40] Griffith, J., Hochschild, A. and Ptashne, M. (1986) Nature 322, 750–752.
[41] Zuckerkandl, E. (1986) J. Mol. Evol. 24, 12–27.
[42] Ivarie, R.D., Schacter, B.S. and O'Farrell, P.H. (1983) Mol. Cell. Biol. 3, 1460–1467.
[43] Wold, B.J., Klein, W.H., Hough-Evans, B.R., Britten, R.J. and Davidson, E.H. (1978) Cell 14, 941–950.
[44] Davidson, E.H. and Britten, R.J. (1979) Science 204, 1052–1059.
[45] Van Eekelen, C., Ohlsson, R., Philipson, L., Mariman, E., Van Beek, R. and Van Venrooij, W. (1980) Nucleic Acids Res. 10, 7115–7131.
[46] Thomas, J.O., Glowacka, S.K. and Szer, W. (1983) J. Mol. Biol. 171, 439–455.
[47] Levy, A. and Noll, M. (1981) Nature 289, 198–203.
[48] Goodwin, G.H., Nicola, R.H., Cockerill, P.N., Zavou, S. and Wright, C.A. (1985) Nucleic Acids Res. 13, 3561–3579.
[49] Cavalier-Smith, T. (1985) in: The Evolution of Genome Size (Cavalier-Smith, T. ed.) pp.105–184, John Wiley and Sons, New York.