Mini-Review

Anionic Regions in Nuclear Proteins

William C. Earnshaw

Department of Cell Biology and Anatomy, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

Recent technological advances have generated a flood of protein sequence data, much of it deduced from the DNA sequences of cloned genes. In this commentary I will discuss an unexpected sequence motif found in a number of proteins that are thought to interact directly or indirectly with DNA. The motif, which I will term an A⁻ region (acidic region), is a local high concentration of acidic residues, which may in extreme cases take the form of monotonous runs of either glutamic or aspartic acid residues. I will list below a number of proteins that have been shown to contain A⁻ regions, and speculate about the possible role of such regions in vivo.

A⁻ Regions in Nuclear Proteins

A⁻ regions have been identified directly in several chromatin and chromosomal proteins by protein sequencing and deduced in a variety of others by DNA sequencing methods (see Table I). In addition, a recent study looking for polyacidic regions (identified by digestion of bulk nuclear non-histone proteins with a mixture of proteolytic enzymes and subsequent isolation of resistant peptides) suggests that such regions are significantly more common in nuclear than cytoplasmic proteins (26). Since it seems unlikely that the A⁻ regions interact directly with DNA, what are they doing in chromatin proteins?

One possibility is that they anchor the proteins to chromatin through electrostatic interactions with the basic histones. However, examination of the solubility properties of a number of proteins for which sequence data are available suggests that this cannot be the whole explanation. Proteins containing A⁻ regions span the entire range of chromatin-binding affinities. For example, nucleoplasmin is highly soluble, showing no appreciable binding to either DNA or chromatin (though it does bind to isolated core histones; references 12 and 27); HMGs 1 and 2 bind to chromatin, but are readily eluted by relatively low ionic strengths (<0.5 M NaCl; reference 18); topoisomerase I is eluted from chromatin at 1 M NaCl (31); and finally, CENP-B and c-myc are tightly associated with the salt-insoluble mitotic chromosome scaffold (13) and nuclear matrix fractions (16), respectively. Therefore chromatin-binding properties cannot be predicted from the presence or absence of A⁻ regions alone.

It has been previously noted that A⁻ regions might be expected to exhibit significant binding to the core histones in vivo. It has been thought that this binding regulates nucleosome assembly and disassembly (28, 41), an idea suggested by the finding that polyglutamic acid assembles nucleosome cores at physiological ionic strengths by organizing histones into octamers and transferring them to DNA (47). There is

evidence to support this contention both in vitro and in vivo. Both nucleoplasmin and HMG-1 can act as nucleosome assembly factors in vitro (6, 27). In addition, it has been shown that part of the stored maternal histone pool in extracts prepared from *Xenopus laevis* oocytes is complexed to nucleoplasmin (23). (However, the presence of A⁻ regions cannot be the sole determinant of histone binding. In these experiments [22, 23], histones were also found to be complexed with N1/N2, but not to HMG-A, the oocyte equivalent of HMG-1. Both proteins have significant A⁻ regions; see Table I.)

Clearly, not all polypeptides containing A⁻ regions promote nucleosome assembly and disassembly. Other roles for this unexpected sequence motif must therefore be considered.

Possible Roles of the A⁻ Regions

The list of plausible roles for the A⁻ regions includes the following.

(a) One function of certain A⁻ regions might be to unfold the condensed higher order chromatin fiber, causing the DNA to become more accessible for scanning by proteins (or by other domains of the proteins containing the A⁻ regions) that recognize specific sequences. This unfolding could result from electrostatic "capture" of the basic domains (21, 53) of the core histones and/or H1. Such capture might be predicted to open up the higher order chromatin fiber, since proteolytic removal of these domains has been shown to prevent formation of the higher order fiber but not to affect nucleosome core assembly or integrity (1, 2, 11, 30, 35, 52, 54). Furthermore, modifications of the basic domains (such as acetylation; reference 43) are correlated with transcriptional activity (reviewed in references 33 and 40). Such modifications could provide a mechanism for restricting the action of freely diffusible proteins containing A⁻ regions to localized chromatin domains. Evidence consistent with this model has been obtained for two yeast proteins with weak A⁻ regions (see next section).

(b) The regions might act through transient weak electrostatic attractions with the histones to enable DNA-binding proteins to track along the chromatin, in a manner similar to that in which repressors have been proposed to find operator sequences in bacteria (3). Such a localized charge effect would greatly increase the effective concentration of the effector near the DNA, and would therefore ultimately increase the rate at which target sites are located.

(c) An A⁻ region on a component of the replication or transcription enzyme complex could act as a "hiking strap"

Table I. Examples of Nuclear Proteins Containing A⁻ Regions

Protein*	Size of A⁻ region	Glu + Asp	Function (if known)	Reference
		%		
CENP-B	61§	79	Centromere protein	14
	31	87		
HMG-1	42	98	? ssDNA binding‖	50
HMG-2	38	92	? ss DNA binding‖	50
Xenopus N1/N2	31	68	? Histone binding in oocytes	24
	12	83		
MCM9 gene product	23	87	Initiation of DNA replication	45
RAD-6	23	87	DNA repair, sporulation	41
Nucleoplasmin	21	81	? Nucleosome assembly in vitro Histone binding in oocytes	10
λ-m6 protein	19	89	? Homeotic analogue (cellular location unknown)	9
DNA polymerase (Herpes Simplex)	18	78	DNA synthesis	17, 37
engrailed region protein	17	76	? Developmental control	36
DNA topoisomerase I	16	75	? Transcription/replication	49
c-myc (chicken)‡	13	77	? Regulation of cell proliferation	51

This table presents only polypeptides either known or assumed to function within nuclei of eukaryotic cells. The list was in part compiled by scanning the Protein Identification Resource (of the National Biomedical Research Foundation) for sequences of 15 contiguous amino acid residues that were ≥75% glu + asp.
* Listed in order of decreasing A⁻ region size.
‡ Similar results were obtained for c-myc from human (38) and mouse (46).
§ Amino acids.
‖ See references 5 and 20.

to which nucleosomes could bind transiently so that they might swing out of the way of the passing enzymes. This might explain results indicating that histones are present in the proximity of transcriptional complexes even though nucleosomes are often not readily observed on genes undergoing intensive transcription.

(*d*) In some cases the A⁻ regions might undergo direct interactions with other chromatin proteins. These could lead to establishment of stable structures or to activation of enzymes (such as has been suggested for GCN4 activation of RNA polymerase; reference 19).

(*e*) One obvious possibility, that A⁻ regions might be involved in targeting of certain classes of nuclear proteins to the nucleus, seems unlikely since the A⁻ regions of nucleoplasmin and GAL4 are experimentally separable from the regions carrying the karyophilic signal sequences (10, 44). The ability of A⁻ regions to act as karyophilic signals could be directly tested by construction of hybrid proteins in *Saccharomyces cerevisiae.*

Specific Examples of Proteins with A⁻ Regions

Nucleoplasmin catalyses nucleosome core assembly in vitro (28), and we have previously proposed that it might function in vivo either in this capacity, or by disassembling nucleosomes to permit the transcription of active genes (27). Alternatively, the first model presented above suggests that nucleoplasmin, the most abundant protein in the *Xenopus* oocyte nucleus (34), might promote transcription in oocytes (which show a tremendous diversity of transcriptional activity) by unfolding higher order chromatin fibers so that sequence-specific transcription factors might function. (Note that N1/N2 or HMG-A might also fulfill such a role in the oocyte nucleus [22, 23].)

In somatic cells a similar role may be taken up by other abundant nonhistone proteins, such as HMG1 and HMG2, since nucleoplasmin is apparently present in much reduced amounts (C. Dingwall, personal communication; see, however, reference 25). In fact, the role of the HMG-1 and -2 proteins in vivo is highly controversial (reviewed in reference 40). Observations suggesting that HMG-1 and -2 play a role in transcription are balanced by others that disagree (40). Several examples follow. (*a*) The polypeptides are, in some cases, associated with regions of transcriptionally active chromatin as identified by increased sensitivity to DNase 1 (numerous references, reviewed in reference 40, p. 372). (*b*) Purified HMGs may stimulate runoff transcription by endogenous polymerase on salt-washed chromatin (reference 48, disputed in reference 8). (*c*) Antibodies to HMG-1 inhibit transcription of lampbrush chromosomes in amphibian oocytes (22), though other antibodies have no apparent effect on transcription in somatic cells (15). Clearly a great deal of work still needs to be done in this area. (Note that HMG-14 and -17 have been excluded from this discussion since they lack A⁻ regions.)

The most suggestive evidence for a role of A⁻ regions in activation of transcription comes from two yeast proteins. GCN4 and GAL4 are transcriptional activator proteins that contain distinct DNA binding and activator domains. The latter work in a sequence-independent fashion when coupled to other DNA binding domains (19, 32). These activator domains, while not as acidic as the A⁻ regions listed in Table I, are highly acidic. In GCN4, the domain contains a region of 60 amino acid residues that is 30% glu + asp (19). In GAL4, two separate regions are present (32). These contain stretches of 29 residues (31% glu + asp) and 20 residues (35% glu + asp; references 32 and 29).

It was postulated that these domains might either open up

the chromatin structure or interact directly with some component of the transcription apparatus (19, 32). Recent results favor the latter interpretation, but do not rule out the possibility that both chromatin unfolding and specific interactions with the transcriptional apparatus are involved (K. Struhl, personal communication). It would be interesting to know whether the polyacidic region of nucleoplasmin could replace these activator regions in vivo.

Proteins with polyacidic regions might also be involved in the recognition of DNA replication origins, regions of DNA homology (in recombination), and DNA damage (in repair). The best examples of the involvement of A⁻ regions in these processes come from two proteins of *Saccharomyces*, both of which have A⁻ regions of identical size and charge located at or near the carboxy terminus. The MCM9 protein appears to influence the replication of a subset of ARS elements (detected as an effect on the stability of certain centric minichromosomes but not others; reference 45), while the RAD6 protein is essential for genetic recombination and induced mutagenesis (which may involve recognition of abnormalities of the DNA helix; reference 41).

Other proteins with A⁻ regions may be involved in the establishment of sequence-dependent chromosomal structures. CENP-B, the 80-kD human centromere autoantigen (14), has the largest A⁻ domain observed to date. It seems fairly certain that this protein serves a structural role at or subjacent to the site of microtubule attachment to the kinetochore. Experiments in *Saccharomyces* have demonstrated clearly that DNA sequence recognition defines kinetochore location (7). Thus, protein recognition of specific centromere DNA sequences must occur. Chromatin unfolding by the A⁻ region of CENP-B could be involved both in rendering centromere sequences accessible for recognition, and in establishment of the altered kinetochore chromatin structure observed by both biochemical (4) and ultrastructural (39, 42) analyses. Alternatively, the A⁻ region of CENP-B might be involved in direct structural interactions with other chromosomal proteins or with the spindle microtubules.

Clearly, many other functions of A⁻ regions are possible, and the list will undoubtedly grow as further proteins bearing the motif are identified.

References

1. Allan, J., N. Harborne, D. C. Rau, and H. Gould. 1982. Participation of core histone "tails" in the stabilization of the chromatin solenoid. *J. Cell Biol.* 93:285–297.
2. Allan, J., T. Mitchell, N. Harborne, L. Bohm, and C. Crane-Robinson. 1986. Roles of H1 domains in determining higher order chromatin structure and H1 Location. *J. Mol. Biol.* 187:591–601.
3. Berg, O. G., R. B. Winter, and P. H. von Hippel. 1982. How do genome-

regulatory proteins locate their DNA target sites? *Trends Biochem. Sci.* 7:52–55.
4. Bloom, K. S., and J. Carbon. 1982. Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromosomes. *Cell.* 29:305–317.
5. Bonne, C., P. Sautiere, M. Duguet, and A.-M. De Recondo. 1982. Identification of a single-stranded DNA binding protein from rat liver with high mobility group protein 1. *J. Biol. Chem.* 257:2722–2725.
6. Bonne-Andrea, C., F. Harper, J. Sobczak, and A.-M. De Recondo. 1984. Rat liver HMG1:a physiological nucleosome assembly factor. *EMBO (Eur. Mol. Biol. Organ.) J.* 3:1193–1199.
7. Clarke, L., and J. Carbon. 1985. The structure and function of yeast centromeres. *Annu. Rev. Genet.* 19:29–56.
8. Cockerill, P. N., G. H. Goodwin, P. D. Cary, C. Turner, and E. W. Johns. 1983. Comparisons of the structures of the chromosomal high mobility group proteins HMG1 and HMG2 prepared under conditions of neutral and acidic pH. *Biochim. Biophys. Acta.* 745:70–81.
9. Colberg-Poley, A. M., S. D. Voss, K. Chowdhury, and P. Gruss. 1985. Structural analysis of murine genes containing homeo box sequences and their expression in embryonal carcinoma cells. *Nature (Lond.).* 314:713–718.
10. Dingwall, C., S. M. Dilworth, S. J. Black, S. E. Kearsey, L. S. Cox, and R. A. Laskey. 1987. Nucleoplasmin cDNA sequence reveals polyglutamic acid tracts and a cluster of sequences homologous to putative nuclear localization signals. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:69–74.
11. Dumuis-Kervabon, A., I. Encontre, G. Etienne, J. Jauregui-Adell, J. Méry, D. Mesnier, and J. Parello. 1986. A chromatin core particle obtained by selective cleavage of histones by clostripain. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:1735–1742.
12. Earnshaw, W. C., B. M. Honda, R. A. Laskey, and J. O. Thomas. 1980. Assembly of nucleosomes: the reaction involving X. laevis nucleoplasmin. *Cell.* 21:373–383.
13. Earnshaw, W. C., N. Halligan, C. Cooke, and N. Rothfield. 1984. The kinetochore is part of the metaphase chromosome scaffold. *J. Cell Biol.* 98: 352–357.
14. Earnshaw, W. C., K. F. Sullivan, P. S. Machlin, C. A. Cooke, D. A. Kaiser, T. D. Pollard, N. F. Rothfield, and D. W. Cleveland. 1987. Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen. *J. Cell Biol.* 104:817–829.
15. Einck, L., and M. Bustin. 1983. Inhibition of transcription in somatic cells by microinjection of antibodies to chromosomal proteins. *Proc. Natl. Acad. Sci. USA.* 80:6735–6739.
16. Eisenman, R. N., C. Y. Tachibana, H. D. Abrams, and S. R. Hann. 1985. v-myc and c-myc-encoded proteins are associated with the nuclear matrix. *Mol. Cell. Biol.* 5:114–126.
17. Gibbs, G. S., H. C. Chiou, J. D. Hall, D. W. Mount, M. J. Retondo, S. K. Weller, and D. M. Coen. 1985. Sequence and mapping analysis of the herpes simplex virus DNA polymerase gene product predict a C-terminal substrate binding domain. *Proc. Natl. Acad. Sci. USA.* 82:7969–7973.
18. Goodwin, G. H., J. M. Walker, and E. W. Johns. 1978. The high mobility group (HMG) nonhistone chromosomal proteins. *In* The Cell Nucleus. Vol. VI. H. Busch, editor. Academic Press, Inc., New York. 181–219.
19. Hope, I. A., and K. Struhl. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell.* 46:885–894.
20. Isakson, P. J., J. L. Fishback, D. L. Bidney, and G. R. Reeck. 1979. Preferential affinity of high molecular weight high mobility group nonhistone chromatin proteins for single stranded DNA. *J. Biol. Chem.* 254:5569–5572.
21. Isenberg, I. 1979. Histones. *Annu. Rev. Biochem.* 48:159–191.
22. Kleinschmidt, J. A., U. Scheer, M.-C. Dabauville, M. Bustin, and W. W. Franke. 1983. High mobility group proteins of amphibian oocytes: a large storage pool of a soluble high mobility group-1-like protein and involvement in transcriptional events. *J. Cell Biol.* 97:838–848.
23. Kleinschmidt, J. A., E. Fortkamp, G. Krohne, H. Zentgraf, and W. W. Franke. 1985. Co-existence of two different types of soluble histone complexes in nuclei of *Xenopus laevis* oocytes. *J. Biol. Chem.* 260:1166–1176.
24. Kleinschmidt, J. A., C. Dingwall, G. Maier, and W. W. Franke. 1985. Molecular characterization of a karyophilic, histone-binding protein: cDNA cloning, amino acid sequence and expression of nuclear protein N1/N2 of *Xenopus laevis. EMBO (Eur. Mol. Biol. Organ.) J.* 5:3547–3552.
25. Krohne, G., and W. W. Franke. 1980. A major soluble acidic protein located in nuclei of diverse vertebrate species. *Exp. Cell Res.* 129:167–189.
26. Kuehl, L., T. J. Childers, and R. M. McCauley. 1986. The occurrence of extended acidic sequences in nonhistone chromosomal proteins. *Arch. Biochem. Biophys.* 248:272–281.
27. Laskey, R. A., and W. C. Earnshaw. 1980. Nucleosome assembly. *Nature (Lond.).* 286:763–767.
28. Laskey, R. A., B. M. Honda, A. D. Mills, and J. T. Finch. Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature (Lond.)* 275:416–420.
29. Laughon, A., and R. Gesteland. 1984. Primary structure of the *Saccharomyces cerevisiae* GAL4 gene. *Mol. Cell. Biol.* 4:260–267.
30. Lilley, D. M. J., and K. Tatchell. 1977. Chromatin core particle unfolding in-

duced by tryptic cleavage of histone. *Nucleic Acids Res.* 4:2039–2055.

31. Liu, L. F., and K. G. Miller. 1981. Eukaryotic DNA topoisomerases: two forms of type I DNA topoisomerases from HeLa cell nuclei. *Proc. Natl. Acad. Sci. USA.* 78:3487–3491.

32. Ma, J., and M. Ptashne. 1987. Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell.* 48:847–853.

33. McGhee, J. D., and G. Felsenfeld. 1980. Nucleosome structure. *Annu. Rev. Biochem.* 49:1115–1156.

34. Mills, A. D., R. A. Laskey, P. Black, and E. M. De Robertis. 1980. An acidic protein which assembles nucleosomes in vitro is the most abundant protein in *Xenopus* oocyte nuclei. *J. Mol. Biol.* 139:561–568.

35. Palter, K. B., and B. M. Alberts. 1979. The use of DNA-cellulose for analysing histone-DNA interactions. Discovery of nucleosome-like histone binding to single-stranded DNA. *J. Biol. Chem.* 254:11160–11169.

36. Poole, S. J., L. M. Kauvar, B. Drees, and T. Kornberg. 1985. The engrailed locus of Drosophila: structural analysis of an embryonic transcript. *Cell.* 40:37–43.

37. Quinn, J. P., and D. J. McGeoch. 1985. DNA sequence of the region in the genome of herpes simplex virus type 1 containing the genes for DNA polymerase and the major DNA binding protein. *Nucleic Acids Res.* 13:8143–8163.

38. Rabbitts, T. H., P. H. Hamlyn, and R. Baer. 1983. Altered nucleotide sequences of a translocated c-myc gene in Burkitt lymphoma. *Nature (Lond.)* 306:760–765.

39. Rattner, J. B. 1986. Organization within the mammalian kinetochore. *Chromosoma (Berl.).* 93:515–520.

40. Reeves, R. 1984. Transcriptionally active chromatin. *Biochim. Biophys. Acta.* 782:343–393.

41. Reynolds, P., S. Weber, and L. Prakash. 1985. RAD6 gene of *Saccharomyces cerevisiae* encodes a protein containing a tract of 12 consecutive aspartates. *Proc. Natl. Acad. Sci. USA.* 82:168–172.

42. Ris, H., and P. L. Witt. 1981. Structure of the mammalian kinetochore. *Chromosoma (Berl.).* 82:153–170.

43. Ruiz-Carrillo, A., L. J. Wangh, and V. G. Allfrey. 1975. Processing of

newly synthesized histone molecules. *Science (Wash. DC).* 190:117–128.

44. Silver, P., L. P. Keegan, and M. Ptashne. 1984. Amino terminus of the yeast GAL4 gene product is sufficient for nuclear localization. *Proc. Natl. Acad. Sci. USA.* 81:5951–5955.

45. Sinha, P., C. Chan, G. Maine, S. Passmore, D. Ren, and B.-K. Tye. 1986. Regulation of DNA replication initiation in yast. *In* Yeast Cell Biology. J. Hicks, editor. Alan R. Liss, Inc., New York. 193–209.

46. Stanton, L. W., P. D. Fahrlander, P. M. Tesser, and K. B. Marcu. 1984. Nucleotide sequence comparison of normal and translocated murine c-myc genes. *Nature (Lond.)* 310:423–425.

47. Stein, A., J. P. Whitlock, Jr., and M. Bina. 1979. Acidic polypeptides can assemble both histones and chromatin *in vitro* at physiological ionic strength. *Proc. Natl. Acad. Sci. USA.* 76:5000–5004.

48. Stoute, J. A., and W. F. Marzluff. 1982. HMG-proteins 1 and 2 are required for transcription of chromatin by endogenous RNA polymerase. *Biochem. Biophys. Res. Commun.* 107:1279–1284.

49. Thrash, C., A. T. Bankier, B. G. Barrell, and R. Sternglanz. 1985. Cloning, characterization, and sequence of the yeast DNA topoisomerase I gene. *Proc. Natl. Acad. Sci. USA.* 82:4374–4378.

50. Walker, J. M. 1982. Primary structures. *In* The HMG Chromosomal Proteins. E. W. Johns, editor. Academic Press, London. 69–87.

51. Watson, D. K., E. P. Reddy, P. H. Duesberg, and T. S. Papas. 1983. Nucleotide sequence analysis of the chicken *c-myc* gene reveals homologous and unique coding regions by comparison with the transforming gene of avian myelocytomatosus virus MC29, Δgag-myc. *Proc. Natl. Acad. Sci. USA.* 80:2146–2150.

52. Weintraub, H., and F. Van Lente. 1974. Dissection of chromosome structure with trypsin and nucleases. *Proc. Natl. Acad. Sci. USA.* 71:4249–4253.

53. Wells, D. E. 1986. Compilation analysis of histones and histone genes. *Nucleic Acids Res.* 14 (Suppl.):r119–r149.

54. Whitlock, J. P., Jr., and A. Stein. 1978. Folding of DNA by histones which lack their NH₂-terminal regions. *J. Biol. Chem.* 253:3857–3861.