Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 11210–11214, November 1995
Biochemistry

Periodicity of strong nucleosome positioning sites around the chicken adult β-globin gene may encode regularly spaced chromatin

COLIN DAVEY, SARI PENNINGS, GEERT MEERSSEMAN, TIM J. WESS*, AND JAMES ALLAN†

Department of Biochemistry, University of Edinburgh, George Square, Edinburgh EH8 9XD, Scotland

Communicated by K. E. van Holde, Oregon State University, Corvallis, OR, July 31, 1995 (received for review February 25, 1995)

ABSTRACT **Positioned nucleosomes contribute to both the structure and the function of the chromatin fiber and can play a decisive role in controlling gene expression. We have mapped, at high resolution, the translational positions adopted by limiting amounts of core histone octamers reconstituted onto 4.4 kb of DNA comprising the entire chicken adult β-globin gene, its enhancer, and flanking sequences. The octamer displays extensive variation in its affinity for different positioning sites, the range exhibited being about 2 orders of magnitude greater than that of the initial binding of the octamer. Strong positioning sites are located 5′ and 3′ of the globin gene and in the second intron but are absent from the coding regions. These sites exhibit a periodicity (≈200 bp) similar to the average spacing of nucleosomes on the inactive β-globin gene *in vivo*, which could indicate their involvement in packaging the gene into higher-order chromatin structure. Overlapping, alternative octamer positioning sites commonly exhibit spacings of 20 and 40 bp, but not of 10 bp. These short-range periodicities could reflect features of the core particle structure contributing to the pronounced sequence-dependent manner in which the core histone octamer interacts with DNA.**

The bulk of DNA in eukaryotic nuclei is packaged into nucleosomes and folded into higher-order chromatin structure. Within a chromosome, this 30-nm fiber appears to be organized into domains (1). The four developmentally regulated chicken β-globin genes, for example, are contained within a domain of ≈30 kb (2). General modifications to this domain accompany gene activation (2–7), but these do not appear to result in substantial, overall disruption of the 30-nm fiber. Distinct structural disruption does occur within the regulatory regions of transcribing genes (3, 8), but evidence suggests that the transcribed sequences themselves retain both core and linker histones (4, 5, 9, 10) and that any unfolded conformation is only transient (10, 11). Thus, gene expression appears to take place in the context of a folded chromatin fiber, a condition which may contribute to the temporal control of developmentally linked genes within a domain. A complete explanation of the mechanisms involved in gene regulation should therefore take into account the long-range organization of the nucleoprotein fiber.

There are indications that the manner in which DNA is packaged into a higher-order chromatin fiber is sensitive to the spatial distribution of nucleosomes (12), which in turn can be determined by the underlying DNA sequence (13, 14). Most studies of sequence-directed nucleosome positioning have analyzed only short regions and have focused on regulatory DNAs. To determine whether there are features in the long-range organization of DNA sequence which could influence the higher-order packaging of chromatin and thereby contribute to regulating gene expression, we have produced a long-range, high-resolution nucleosome positioning map for an entire gene region. Our analysis of the chicken adult β-globin gene, its enhancer, and flanking sequences reveals extensive variation in the strength of octamer positioning and demonstrates that the highest-affinity sites, which are largely contained within the noncoding regions of the sequence, are arranged with a periodicity of about 200 bp.

MATERIALS AND METHODS

Phagemid Construction. A 4.4-kb *Eco*RI–*Bam*HI fragment of chicken β-globin sequence (−1052 to +3369, relative to the transcription start site of the adult gene) was cloned from plasmid pCARB4.4 (15) into pBluescript KS(−) (Stratagene) to generate pCBA4.4. Fourteen overlapping fragments of this β-globin sequence were similarly cloned into pBluescript KS(+) or KS(−) to generate the following phagemids: KS(+) subclones 44 (+201 to +583), 4A (+584 to +1063), 13 (+972 to +1598), 4B (+1599 to +2091), 8 (+2092 to +2608), 64 (+2092 to +2870), and 20 (+2675 to +3177); KS(−) subclones EcSm (−1052 to −110), Le (−406 to +200), SmH5 (−56 to +1063), EgS4 (−8 to +971), GP1 (+1064 to +2091), G3HX (+1471 to +2073), and S6B8 (+1599 to +2608).

Monomer Extension. Phagemid single-stranded DNA was isolated from *Escherichia coli* DH11S (16). Core particle DNA was prepared from pCBA4.4 as described (17); reconstitution was carried out at a core histone/DNA ratio of 0.5:1 (wt/wt), the reconstitute was digested with micrococcal nuclease, and the ≈146-bp core particle DNA fragments were isolated from a 1.8% agarose gel. Monomer extension was performed as described (17): typically, 30 ng of 5′ end-labeled, alkaline-denatured pCBA4.4 core particle DNA was annealed to excess phagemid single-stranded DNA (1 μg) and extended by Klenow DNA polymerase in the absence or presence of appropriate restriction enzymes. Extension products were analyzed by electrophoresis in denaturing 6% polyacrylamide gels.

Monomer Extension Data Analysis. Quantitative densitometer scans were obtained for each extension reaction after PhosphorImager (Molecular Dynamics) analysis of the dried gels. Scans were converted to a linear scale by reference to a sixth-order polynomial generated from the mobilities of 85 marker fragments. Sized extension products were then related to the globin sequence to locate core particle boundaries. For quantitation, scans of extension reactions carried out in the absence of a restriction enzyme were considered to be background. Analyses were normalized with respect to a common octamer positioning site within vector sequence [KS(+) clones] or via shared positioning sites within the globin DNA [KS(−) clones].

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Present address: Department of Biological and Molecular Sciences, University of Stirling, Stirling FK9 4LA, Scotland.
†To whom reprint requests should be addressed.

The entire data set, comprising the intensity amplitudes at each of the 4192 bp analyzed, was subjected to autocorrelation analysis using the equation

$$F(k) = \sum_{t=k+1}^{n} \Big(y(t) - \bar{y}\Big)\cdot\Big(y(t+k) - \bar{y}\Big),$$

where k is the displacement length, n is the total number of data points, y is the pixel intensity at each point $[y(t), y(t+k)]$, and $\bar{y}$ is the average pixel intensity of the data set. This function was calculated for 1-bp increments from 0 bp (self-mapping function) to the entire length of the sequence and was normalized to the value of the self-mapping function. The autocorrelation function was then subjected to Fourier transformation in the range of 20–400 bp by sampling 400 points regularly spread over the reciprocal base-pair separation.

Competitive Reconstitution. pCBA4.4 DNA was sonicated to an average length of 200 bp and then reconstituted at a core histone/DNA ratio of 0.4:1 (wt/wt). The resulting mixture of DNA and mononucleosomes was separated in a 5% polyacrylamide nucleoprotein gel (18). To distinguish the DNA fragments retarded as mononucleosomes, the lane was cut from the gel, soaked in 0.5% SDS in electrophoresis buffer, and electrophoresed in a second-dimension gel containing 0.5% SDS. Mononucleosome DNA was excised from the gel, purified, and radioactively labeled by use of random primers. This DNA was used to probe a filter prepared by electroblotting a set of 12 restriction digests of pCBA4.4 resolved in a 5% polyacrylamide gel. Hybridization to the various restriction fragments was quantitated by PhosphorImager.

RESULTS AND DISCUSSION

Mapping Core Histone Octamer Positioning Sites on the Chicken Adult β-Globin Gene. We have developed a method (monomer extension) to map the precise translational positions adopted by core histone octamers reconstituted onto long DNA sequences (17). The technique maps the boundaries of core particle DNA fragments protected by the histone octamer from micrococcal nuclease digestion. We have applied this approach to produce a long-range nucleosome positioning map for an entire contiguous gene region.

A plasmid containing the chicken adult β-globin gene, its enhancer, and flanking sequences (pCBA4.4; Fig. 1*b*) was reconstituted with limiting amounts of core histones. Under our conditions the formation of nucleosome arrays is not significant and octamer–octamer interactions do not contribute to positioning (17). This chromatin was digested with micrococcal nuclease to produce a population of core particles. Purified core particle DNA molecules (monomers) were 5′ end-labeled and used as a heterogeneous population of primers for extension on each member of a set of 14 single-stranded phagemids containing overlapping sections of the 4.4-kb globin fragment (Fig. 1*b*). Extension reactions were carried out in the presence of a restriction enzyme that cleaves the nascent double-stranded DNA at a unique location just downstream of the globin insert. This produces fragments whose lengths reveal the distances from the restriction site to one of the labeled boundaries of individual core particle DNA molecules (Fig. 1*a*). In this way, the locations at which histone octamers were positioned on the globin DNA were determined at high resolution.

The products of extension reactions were analyzed by electrophoresis in denaturing polyacrylamide gels, and a selection of the results is shown in Fig. 2. In the presence of an appropriate restriction enzyme, extension of the monomer DNA gives rise to a set of discrete, monodisperse bands, the pattern of which is characteristic and reproducible for each section of the gene. In regions where both the upstream and downstream boundaries were mapped, the data were consistent (for example, EgS4 and SmH5; Fig. 1*b*). Control reactions carried out in the absence of restriction enzyme gave rise only to high molecular weight products (Fig. 2). Densitometry traces from the collective analysis of separate regions of the globin gene were quantitatively normalized and combined to produce the map shown in Fig. 3.

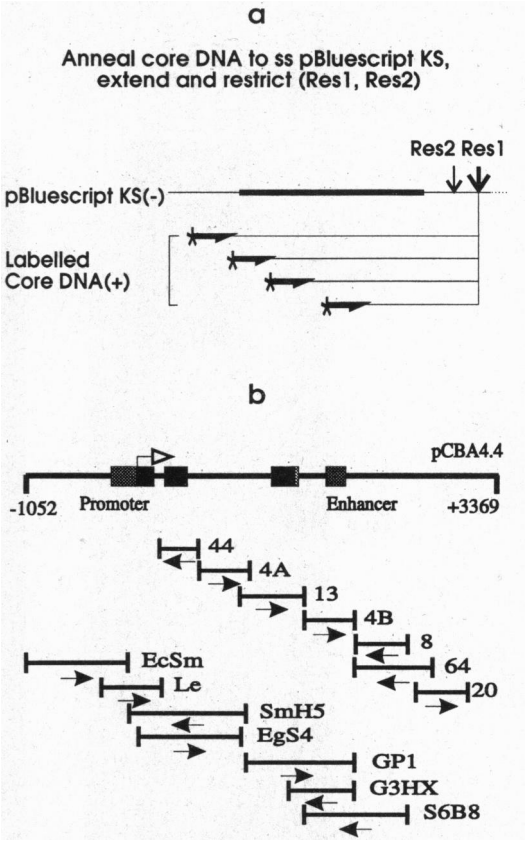

FIG. 1. (*a*) Schematic outline of the monomer extension procedure. ss, Single-stranded. (*b*) The chicken adult β-globin gene region of phagemid pCBA4.4 and subclones. Arrows indicate the direction of monomer extension on single-stranded DNA of each subclone.

Affinity of the Histone Octamer for Positioning Sites. In our analysis, the relative intensity of an extension product is determined by the abundance of a particular core particle DNA in the population, which in turn reflects the relative affinity of core histone octamers for that particular 146-bp stretch of DNA (17). The data in Figs. 2 and 3 therefore provide an indication of the range of affinities that the octamer exhibits for different 146-bp frames of the globin DNA sequence, including those which overlap. Clearly, many of the possible positioning frames on the globin gene are very poorly represented and the signal is essentially at background level. Densitometry (EcSm and 20; Fig. 2) reveals that the integrated intensities of the strongest positioning sites are some 300 times greater than background. This value is an estimate of the range in the relative affinity of core histone octamers for all possible positioning sites.

The relative affinity of the core histone octamer for a variety of DNA fragments has been measured previously by band-mobility shift analysis (19). This approach measures the comparative ability of a DNA fragment to capture octamers during reconstitution but does not necessarily reflect the capacity of the sequence to position octamers (17, 20). To compare the range in binding strength of octamer capture with that of octamer positioning, we have carried out a competitive reconstitution analysis of the β-globin gene region.

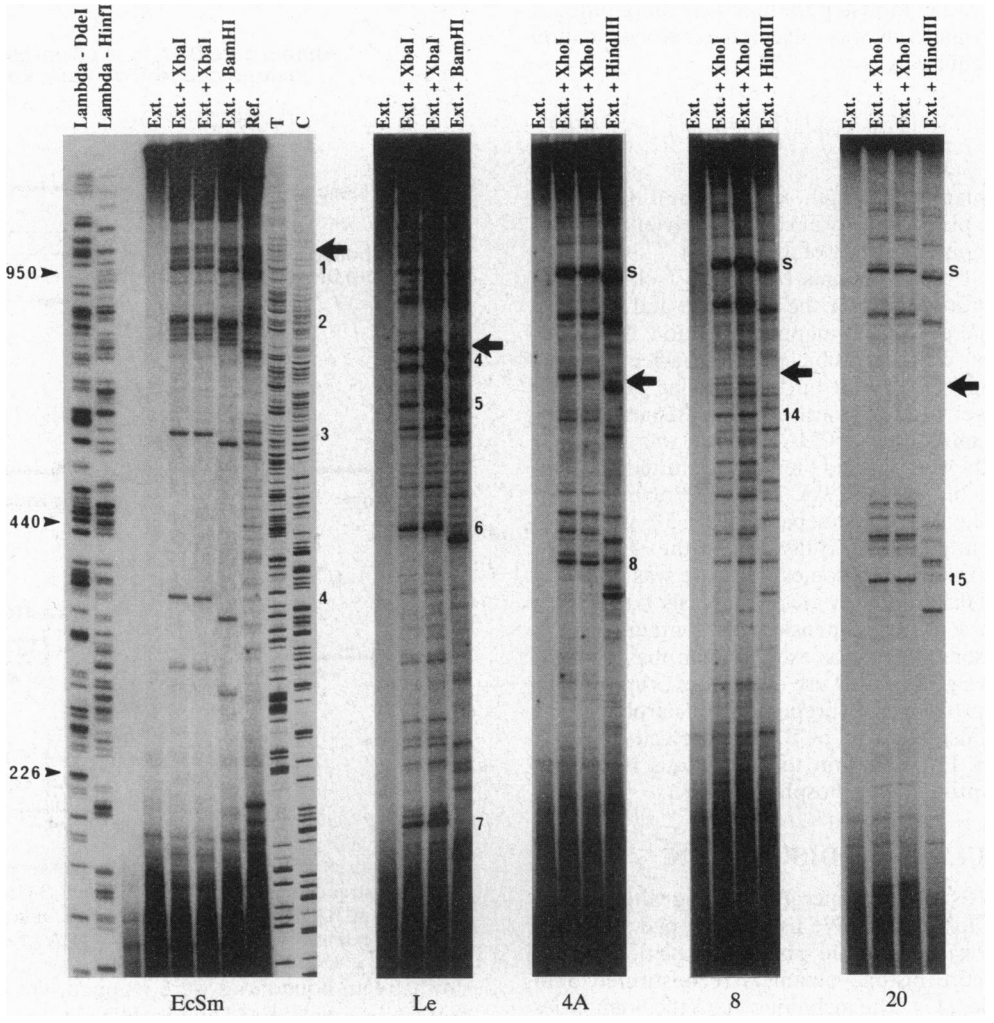

FIG. 2. Denaturing polyacrylamide gel analyses of monomer extension products. Extension reactions were undertaken in the presence of appropriate restriction enzymes (*Xba* I and *Bam*HI, for example), with control extensions in the absence of restriction (Ext.) or in the presence of an enzyme which cuts at the inappropriate end of the globin insert to reveal positioning sites within the vector only (Ref.). Selected products are numbered in accordance with Fig. 3. Arrows denote the limit of extension products derived from the globin inserts; larger products map to the vector. S indicates the prominent vector positioning site employed to quantitatively normalize the KS(+) analyses. Size standards included C and T sequencing reactions of M13mp18 phage DNA and *Hin*fI and *Dde* I digests of phage λ (some sizes, in bases, are shown).

A preparation of short random fragments of pCBA4.4 was reconstituted under competitive conditions, and the DNA recovered from the mononucleosome fraction was identified and quantitated by hybridization to restriction fragments of the globin plasmid. The results reveal the extent to which DNA fragments throughout the globin region compete to capture histone octamers during reconstitution (Fig. 4*a*). Although there is some modulation over the 4.4-kb region, the range in the relative strength of binding is no greater than 4-fold, some 2 orders of magnitude less than the range associated with translational positioning as established by monomer extension (Fig. 3). Thus, the binding of histone octamers to DNA during *in vitro* reconstitution, which occurs in relatively high salt, is, in terms of DNA sequence, a much less discriminating process than that associated with the subsequent positioning of the octamer.

Position and Long-Range Periodicities of Histone Octamer Positioning Sites. The most prominent octamer positioning sites are located 5′ and 3′ to the chicken β-globin gene and in its large second intron (Fig. 3). Strong positioning sites are absent from the coding regions of the gene. Hybridization analysis (Fig. 4*b*) shows that the coding regions are not substantially underrepresented in the core particle DNA population used for monomer extension. Thus, histone octamers do bind to the coding regions of the gene but do not become so discretely positioned. Strong positioning sites in the second intron indicate that such sites are not precluded from transcribed sequences *per se.* Their absence from the exons is likely to reflect, for this gene, an incompatibility between the sequence requirements for coding and core histone positioning.

The strong positioning sites display a notable periodicity along the globin DNA. This feature is most evident at the 5′ end of the gene, where prominent sites are located approximately every 200 bp (sites 1–7; Fig. 3). A less pronounced periodicity of ≈200 bp is also seen in the second intron (sites 8–10) and toward the 3′ end of the gene (sites 11–16). Furthermore, the positioning sites in these latter regions appear to be in register with the periodicity at the 5′ end of the gene.

Autocorrelation analysis (Fig. 5 *a* and *b*) confirms both the nearest-neighbor and long-range periodic spacing of positioning sites. Nearest-neighbor spacings are reflected by the peak centered at ≈200 bp (Fig. 5*b*). This broad maximum contains a number of minor peaks (175, 200, and 230 bp) indicative of discrete subsets of spacings. The regular spacing of peaks between ≈200 and ≈1200 bp and between ≈2800 and ≈3800 bp in the autocorrelation analysis (Fig. 5*a*) indicates a pro-

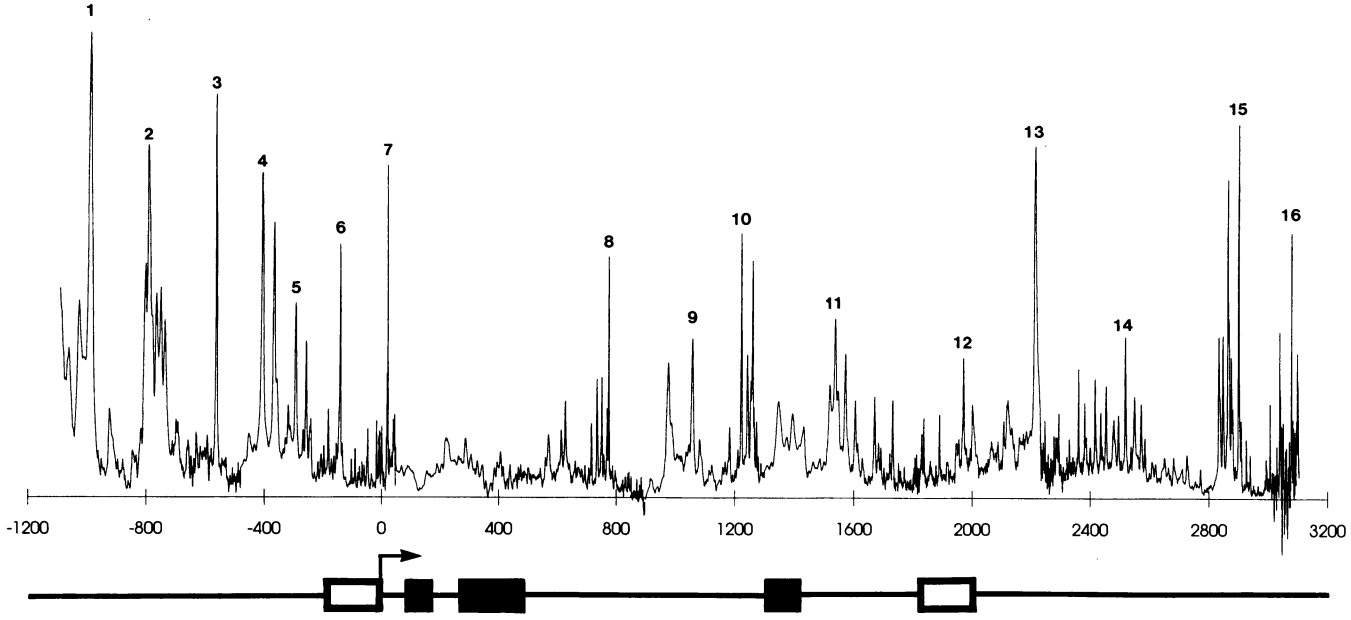

FIG. 3. Map of the histone octamer positioning sites for the chicken adult β-globin gene region. Open boxes represent the promoter and the 3′ enhancer; solid boxes represent the exons. The sequence is numbered with respect to the transcription start site of the gene. By assuming a core particle size of 146 bp, we have arranged the map to depict the centers of positioning sites rather than their boundaries. The positioning sites have been arbitrarily numbered for the purposes of discussion.

nounced long-range periodic function based on about 200 bp. This was confirmed by Fourier transformation of the autocorrelation function (Fig. 5*c*), which reveals a prominent peak centered at 203 bp and extending from approximately 175 to 235 bp.

Core histone octamers assembled onto chicken globin DNA under conditions designed to promote the formation of nucleosome arrays adopt a spacing of ≈180 bp (21). The positioning signals we have detected may well contribute to this spacing, as 180 bp often separates nearest neighbors (Fig. 5*b*).

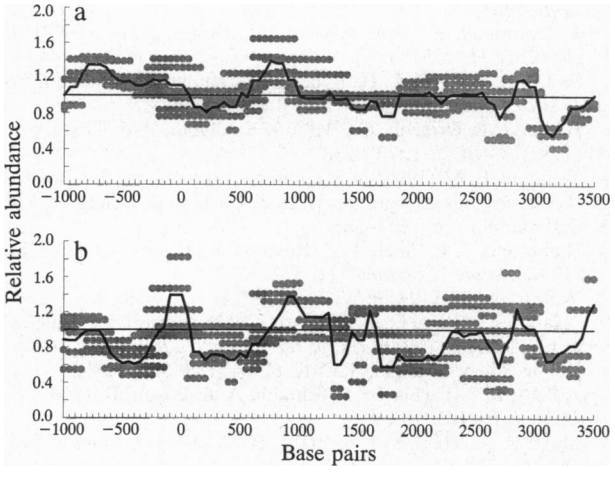

FIG. 4. Hybridization analysis showing the quantitative distribution of octamer binding along the chicken β-globin gene region. A filter containing a panel of restriction digests of pCBA4.4 was probed with nucleosomal DNA formed by competitive reconstitution of histone octamers onto fragmented pCBA4.4 (*a*) and the core particle DNA used for monomer extension (*b*). Blocks of gray symbols depict the length and location of each restriction fragment employed in the hybridization panel (only fragments of <1000 bp are shown). The level of hybridization of nucleosomal DNA to each of these restriction fragments, relative to the level obtained with unreconstituted fragmented DNA, is shown (*y* axis). The solid line is a size-weighted average of the data. The β-globin sequence is numbered with respect to the transcription start site of the gene.

Where direct comparison can be made, octamer positioning 3′ of the globin gene (figure 5 in ref. 21) is consistent with our findings (sites 13–16; Fig. 3).

Estimates of the nucleosome repeat length of inactive chicken β-globin chromatin *in vivo* range from 195 to 200 bp (22–24), close to the 203-bp average spacing we observe. The occurrence and distribution of strong positioning sites over the β-globin gene region could therefore be appropriate to facilitate its packaging into an inactive chromatin structure in nonerythroid tissues. However, it is unclear to what degree these positioning signals would continue to influence chromatin structure when the globin domain adopts an active conformation, as here the average nucleosome spacing is reduced (1, 22) and varies somewhat during development (22). Modification of the chromatin components could affect the capacity of the histones to recognize positioning information in the DNA.

Short-Range Periodicities of Histone Octamer Positioning Sites. Our analysis reveals a number of short-range periodicities relating overlapping, alternative octamer positioning sites. The most notable are 20 and 40 bp (Fig. 5*b*). Numerous examples can be seen in Fig. 3 (20-bp spacing around sites 8, 10, 14, and 15; 40-bp spacing around sites 4, 5, 10, 14, and 15) and in Fig. 2 (with phagemid 8, for example, the *Hin*dIII products are 21 nt shorter than those of *Xho* I, allowing one to easily identify the positioning sites related by 20 bp).

Although the short-range periodicities observed are multiples of 10 bp, a 10-bp periodicity *per se* is infrequent (Fig. 5*b*). Thus, the bendability of DNA, facilitated by a 10-bp periodic distribution of short nucleotide elements (13, 14, 19, 25), is unlikely to be a major determinant of octamer positioning on this gene. The frequent 20- and 40-bp periodicities instead suggest that particular aspects of core particle structure, possibly the sites at which the DNA is distorted (26), are demanding of DNA sequence and consequently of promoting translational positioning. Analysis of the sequences that comprise the histone octamer positioning sites may shed some light upon the sequence elements involved.

Conclusions. Higher eukaryotes have developed an exceptional ability to effect genetic repression (27). This is largely due to the packaging of DNA into higher-order chromatin

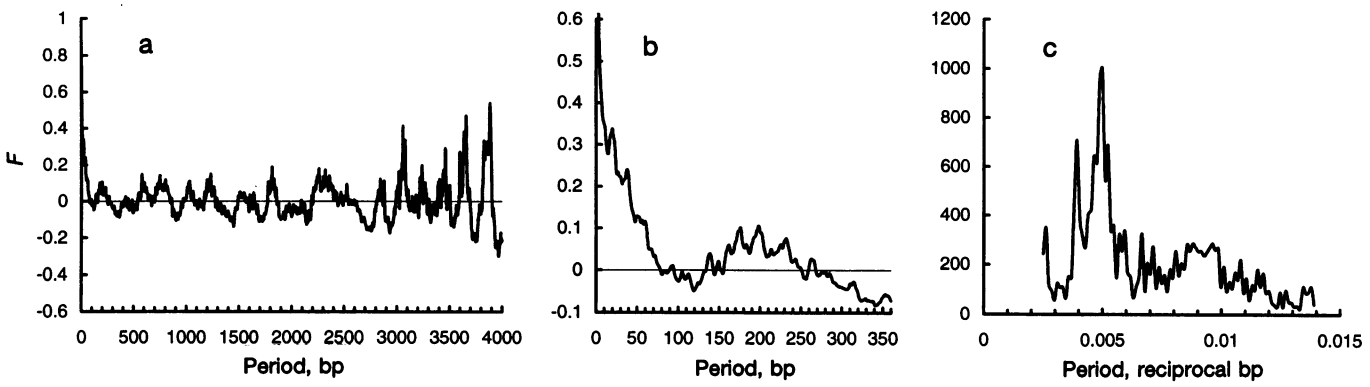

FIG. 5. Analysis of periodicities in positioning sites on the chicken β-globin gene region. (*a*) Autocorrelation analysis of the data set used to generate the map in Fig. 3. (*b*) An expansion in the 0- to 360-bp period range. The variation in amplitude (*F*) is plotted as a function of the period (bp). (*c*) Fourier transformation of the autocorrelation analysis shown in *a*. The variation in amplitude (*F*) is plotted as a function of the reciprocal of the period (1/bp) and was sampled over the range 20–400 bp. Peaks centered at 0.0039 and 0.0049 on the reciprocal scale correspond to periodicities of 256 and 203 bp, respectively.

structure (27), the formation of which is determined by association of the DNA with the core and linker histones (28). We have shown that the propensity of core histones to interact with DNA in a sequence-specific manner, so as to adopt precise translational positions, is prevalent. This property may influence positioning *in vivo* (12–14). In the case of the chicken β-globin gene, the periodic distribution of positioning sites suggests a role in packaging the gene into the higher-order structure responsible for repression of the gene in nonerythroid cells.

Long-range periodicities in sequence or structure have been observed in eukaryotic DNA (29–32). As these are mainly a feature of noncoding sequence, it has been suggested that they could be involved in packaging the DNA into chromatin. This is consistent with our finding that the nucleosome positioning sites with the strongest potential to influence chromatin structure reside in the DNA that flanks and interrupts the coding sequence.

We are grateful to Nick Short for providing certain clones and to Dima Alexeev for assistance with Fourier analysis. This work was supported by a Wellcome Trust project grant to J.A., a Wellcome Trust fellowship to S.P., and a European Molecular Biology Organization fellowship to G.M.

1. van Holde, K. E. (1989) *Chromatin* (Springer, New York).
2. Hebbes, T. R., Clayton, A. L., Thorne, A. W. & Crane-Robinson, C. (1994) *EMBO J.* **13,** 1823–1830.
3. Gross, D. S. & Garrard, W. T. (1987) *Trends Biochem. Sci.* **12,** 293–297.
4. Grunstein, M. (1990) *Annu. Rev. Cell Biol.* **6,** 643–678.
5. Felsenfeld, G. (1992) *Nature (London)* **335,** 219–223.
6. Stadler, J., Larsen, A., Engel, J. D., Dolan, M., Groudine, M. & Weintraub, H. (1980) *Cell* **20,** 451–460.
7. Hebbes, T. R., Thorne, A. W., Clayton, A. L. & Crane-Robinson, C. (1992) *Nucleic Acids Res.* **20,** 1017–1022.
8. Reitman, M. & Felsenfeld, G. (1990) *Mol. Cell. Biol.* **10,** 2774–2786.
9. Postnikov, Y. V., Shick, V. V., Belyavsky, A. V., Khrapko, K. R., Brodolin, K. L., Nikolskaya, T. A. & Mirzabekov, A. D. (1991) *Nucleic Acids Res.* **19,** 717–725.
10. Ericsson, C., Grossbach, U., Björkroth, B. & Daneholt, B. (1990) *Cell* **60,** 73–83.
11. Andersson, K., Mahr, R., Björkroth, B. & Daneholt, B. (1982) *Chromosoma* **87,** 33–48.
12. Woodcock, C. L., Grigoryev, S. A., Horowitz, R. A. & Whitaker, N. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 9021–9025.
13. Simpson, R. T. (1991) *Prog. Nucleic Acid Res. Mol. Biol.* **40,** 143–184.
14. Thoma, F. (1992) *Biochim. Biophys. Acta* **1130,** 1–19.
15. Walmsley, M. E., Buckle, R. S., Allan, J. & Patient, R. K. (1991) *J. Cell Biol.* **114,** 9–19.
16. Lin, J.-J., Smith, M., Jessee, J. & Bloom, F. (1992) *BioTechniques* **12,** 718–721.
17. Yenidunya, A., Davey, C., Clark, D., Felsenfeld, G. & Allan, J. (1994) *J. Mol. Biol.* **237,** 401–414.
18. Meersseman, G., Pennings, S. & Bradbury, E. M. (1992) *EMBO J.* **11,** 2951–2959.
19. Shrader, T. E. & Crothers, D. M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 7418–7422.
20. Drew, H. R. (1991) *J. Mol. Biol.* **219,** 391–392.
21. Liu, K., Lauderdale, J. D. & Stein, A. (1993) *Mol. Cell. Biol.* **13,** 7596–7603.
22. Villeponteau, B., Braeley, J. & Martinson, H. G. (1992) *Biochemistry* **31,** 1554–1563.
23. Sun, Y. L., Xu, Y. Z., Bellard, M. & Chambon, P. (1986) *EMBO J.* **5,** 293–300.
24. Bellard, M., Dretzen, G., Bellard, F., Oudet, P. & Chambon, P. (1982) *EMBO J.* **1,** 223–230.
25. Trifanov, E. N. (1990) in *Theoretical Biochemistry and Molecular Biophysics*, eds. Beveridge, D. L. & Lavery, R. (Adenine, Guilderlane, NY), pp. 377–388.
26. Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D. & Klug, A. (1984) *Nature (London)* **311,** 532–537.
27. Weintraub, H. (1984) *Cell* **38,** 17–27.
28. Allan, J., Cowling, G. J., Harborne, N. R., Cattini, P., Craigie, R. & Gould, H. J. (1981) *J. Cell Biol.* **90,** 279–288.
29. Keene, M. A. & Elgin, S. C. R. (1984) *Cell* **36,** 121–129.
30. Adams, R. L. P., Davis, T., Rinaldi, A. & Eason, R. (1987) *Eur. J. Biochem.* **165,** 107–115.
31. Marx, K. A., Hess, S. T. & Blake, R. D. (1994) *J. Biomol. Struct. Dyn.* **12,** 235–246.
32. Beckmann, J. S. & Trifanov, E. N. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 2380–2383.