Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
MOST ANIMAL SPERM are designed to be hydrodynamically efficient, motile cells. During spermatogenesis, precursors resembling typical somatic cells undergo complex and extensive biochemical and morphological transformations. These include flagellar formation, elimination of cytoplasm, complete repression of gene expression, cessation of replication and mitosis and highly compact packaging of nuclear DNA. Unlike nuclei of terminally differentiated cells, however, the nucleus of the fertilizing sperm must be reactivated. This occurs upon transfer to egg cytoplasm at fertilization, as it is transformed into a male pronucleus and eventually combines with the egg nucleus.

There is little information about the global packaging and unpackaging of the sperm genome, since few organisms have lent themselves to direct biochemical analyses of nuclear transitions in both spermatogenesis and fertilization. In addition, sperm DNA is packaged by very different types of proteins in different organisms, and a seemingly bewildering variety of special proteins has evolved to this end.

Two species of histones in sea urchin sperm (Sp H1 and Sp H2B) are chimeric molecules whose highly basic amino-terminal domains are dephosphorylated at the last stage of sperm cell differentiation, and rephosphorylated immediately following fertilization. The phosphorylated regions consist largely of repeating tetrapeptides with two basic residues flanking Ser-Pro residues ('SPKK' motifs) and are predicted to have β-turn secondary structures. Alteration of the charge and structure of the SPKK sites may play a role in the unusually dense DNA packaging of the mature sperm chromatin. The motif resembles the target site of cell-cycle-associated cdc2 kinases and is found in several other proteins whose nucleic acid affinities may be altered during the cell cycle.

Packaging and unpackaging the sea urchin sperm genome

Dominic L. Poccia and G. R. Green

Two species of histones in sea urchin sperm (Sp H1 and Sp H2B) are chimeric molecules whose highly basic amino-terminal domains are dephosphorylated at the last stage of sperm cell differentiation, and rephosphorylated immediately following fertilization. The phosphorylated regions consist largely of repeating tetrapeptides with two basic residues flanking Ser-Pro residues ('SPKK' motifs) and are predicted to have β-turn secondary structures. Alteration of the charge and structure of the SPKK sites may play a role in the unusually dense DNA packaging of the mature sperm chromatin. The motif resembles the target site of cell-cycle-associated cdc2 kinases and is found in several other proteins whose nucleic acid affinities may be altered during the cell cycle.

Elements in the amino-terminal regions of two male-germ-line-specific histones (Sp H1 and Sp H2B)2,3. The basic charge of these clustered sites in the termini of the histones is essentially neutralized upon multiple phosphorylation. This is believed to drastically lower the affinity of these regions for DNA, altering chromatin physical properties and packaging.

Transformations of sperm chromatin during spermatogenesis and fertilization

The mature sea urchin sperm nucleus appears to contain typical 'beads-on-a-string' chromatin containing core nucleosomes, each of which includes 146 base pairs of core DNA and an octamer of core histone proteins (H2A-H2B-H3-H4). However, the chromatin differs from somatic chromatin in several significant ways2,3. With 100–110 base pairs of linker DNA between cores, it has the highest nucleosomal repeat length ever reported. Two of the five histone classes are restricted to the male germ line. These Sp histones (Sp H1 and Sp H2B) are larger than their somatic counterparts, which are completely absent from sperm. The sperm chromatin contains little or no non-histone protein or RNA and is completely inactive in RNA and DNA synthesis. It is exceptionally stable to thermal or ionic denaturation of its DNA, is more resistant to linker digestion by micrococcal nuclease than somatic chromatin and is more densely packed than mitotic chromosomes.

Clues about the relationship between the Sp histones and the unusual physical properties of sperm chromatin emerged from studying transformations occurring during spermatogenesis and pronuclear formation. The genes for the
Sp histones are expressed exclusively in the early replicating cells of the male germ line (spermatogonia and premeiotic spermatocytes). The genes are similar in structure and expression to typical cell-cycle-regulated somatic histone genes. Indeed, Sp histones are immediately incorporated into chromatin of spermatogonia and are the only species of their class present throughout spermatogenesis. Thus, these large cell-type-specific histone variants must be compatible with normal nuclear activities of spermatogonia, spermatocytes and spermatids of the male germ line, including mitosis, meiosis, transcription and replication.

The only change in histone composition observed in Sp histones during sea urchin spermatogenesis occurs in the terminal stages. As the highly condensed late spermatid nucleus is transformed to the mature sperm nucleus, the Sp histones, which were highly phosphorylated throughout spermatogenesis, become dephosphorylated (Fig. 1). Following fertilization, the same molecules are immediately and specifically rephosphorylated. This process occurs before decondensation of the compact chromatin, or reactivation of replication and transcription. Within hours of fertilization, the Sp histones brought by the sperm to the zygote are replaced or diluted by a programmed succession of embryo histone subtypes (Fig. 1). Virtually all phosphorylation of Sp histones during spermatogenesis occurs in the amino-terminal regions of the molecules, exclusively on Ser residues. Following fertilization, the same regions are rephosphorylated and, in addition, the carboxy-terminal region of Sp H1 is highly phosphorylated. The kinases responsible for phosphorylation of Sp histones are unknown. No histone kinase activity has been found in the mature sperm, but candidates have been found in egg cytoplasm.

Major changes in the physical properties of the chromatin correlate with the changes in phosphorylation state of the Sp histones. Spermatid (or male pronuclear) chromatin that contains phosphorylated Sp H1 and Sp H2B have (1) shorter repeat lengths (less linker DNA) typical of somatic chromatin, (2) linkers more easily digested by micrococcal nuclease (less well protected by histone), and (3) DNA that is more easily heat denatured (less shielded charges on the DNA-phosphate backbone) than sperm chromatin. These differences are at least as great or greater than in chromatins from different cell types having different histone variant compositions. In the case of the male germ line, the variants are identical. Therefore, the unusual physical transitions are most simply attributed to histone secondary modifications.

**Phosphorylation of clustered SPKK sites**

The parallel developmental appearance and phosphorylation patterns of Sp H1 and Sp H2B molecules suggested

---

**Table I. Repeating tetrapeptides in Sp H1 and Sp H2B**

| Histone | Species | Amino-terminal sequence |
|---------|---------|-------------------------|
| Sp H1   | (S. purpuratus) | (X1)1A SPKR SPKK SPRK SPRK AGA SPKR AGA SPKR RS |
| Sp H1   | (P. angulosus)  | (X1)1A SPKR SPKK SPRK SPRK AGA SPKR KA |
| Sp H1   | (E. crassale)   | (X1)1A SPKR SPKK SPRK SPRK K? SPKK KA |
| Sp H1   | (S. granulosus) | (X1)1A SPKR SPRK SPRK SPRK | |
| Sp H1   | (S. nudus)      | (X1)1A SPKR SPRK SPRK SPRK K | |
| Sp H1   | (P. miliaria)   | (X1)1A SPKR SPRK SPRK AGA SPKR AGA SPRK K | |
| Sp H1   | (E. esculentus) | (X1)1A SPKR SPRK SPRK AGA SPRK RS |
| Sp H2B (1) | (S. purpuratus) | PRSPKT SPKR T SPKR Q |
| Sp H2B (1) | (P. miliaria)  | PRSPKT SPKR T SPKR Q |
| Sp H2B (2) | (E. esculentus) | PRSPKT SPKR T SPKR Q |
| Sp H2B (2) | (P. angulosus)  | PRSPKT SPKR T SPKR Q |
| Sp H2B (2) | (L. pictus)     | PRSPKT SPKR T SPKR Q |
| Sp H2B (2) | (P. angulosus)  | PRSPKT SPKR T SPKR Q |
| Sp H2B (2) | (E. esculentus) | PRSPKT SPKR T SPKR Q |
| Sp H2B (3) | (P. angulosus)  | PRSPKT SPKR T SPKR Q |
| Sp H2B (3) | (E. esculentus) | PRSPKT SPKR T SPKR Q |

*\( x = X_1X_2X_3SPQKR, \) where \( X_1 = P \) or \( A; \) \( X_2 = G \) or \( A. \) SPKK sequences are in bold (see Ref. 21 for references).*
that they might share common structural features, even though one is a linker and the other a core histone. Both molecules were known to have amino-terminal regions not found in their somatic counterparts, which accounted for much of their increased size. These extensions are highly basic, comprised largely of repeating tetrapeptides (Sp H1) or pentapeptides (Sp H2B)\(^1\). Assuming the repeating sequences might be recognized by a common kinase, a simple consensus sequence, a tetrapeptide of Ser-Pro adjacent to two basic amino acids (Lys and/or Arg), was proposed\(^2\). Such sites are present in all Sp H1 and Sp H2B histones from all sea urchin species examined so far (Table 1). In Sp H2B, they are arranged slightly differently in the two or three variants found in a given species, and are separated by an intervening amino acid which yields the pentapeptide repeat.

Several features and consequences of this tetrapeptide, subsequently named 'SPKK' (Ref. 14), were pointed out\(^2\). The SPKK motif is similar to known phosphorylation sites in the carboxy-terminal regions of many H1 histones, although in these proteins it is not tandemly repeated; Ser-Pro may be flanked on both sides by two basic residues and Thr can substitute for Ser.

(2) Such H1 sites are phosphorylated by what was known as the growth-associated kinase, was later called the M-kinase and is now known as the p34\(^{cd/cd/20}\) kinase, part of the maturation promoting factor (MPF) complex, which is maximally active at mitosis. (3) In each tetrapeptide, the two positive charges of the basic amino acids are expected to form strong ionic links with the DNA backbone. (4) The phosphorylation of the single Ser (or Thr) would provide two negative charges to effectively neutralize the charge of each segment. (5) The predicted secondary structure associated with the clustered tetrapeptide regions is predominantly \(\beta\)-turn. Proline residues are known to prevent formation of secondary structures such as \(\alpha\)-helices or \(\beta\)-sheets.

The nature of the potential \(\beta\)-turns of Sp histones was later clarified\(^15\). A turn may be formed by Ser-Pro and any two subsequent amino acids in a structure stabilized by two hydrogen bonds between the OH and CO groups of Ser and the amide NH groups of the third and fourth amino acids, respectively. While spectral evidence for \(\beta\)-turns is harder to obtain than for other forms, circular dichroic and Fourier-transformed infrared spectra support such a structure in a synthetic 18-amino-acid peptide containing two SPKK sequences\(^8\). The potential \(\beta\)-turn can be considered an independent structural cassette\(^17\) that requires only four amino acids and may function in a variety of backgrounds. An additional effect of Ser phosphorylation would be to disrupt the stabilizing hydrogen bonds of the cassette. A review of the \(\beta\)-turn structural motif has recently appeared in TIBS\(^9\).

The location of known phosphorylation sites of Sp H1 and Sp H2B histones and their cognates are shown in Fig. 2. Alignment of the molecules by their conserved regions emphasizes the chimeric nature of the Sp histones\(^4\). Among the H2B histones, only Sp types have amino-terminal extensions containing SPKK phosphorylation sites. Sp H1 has a similar amino-terminal extension, and also a smaller carboxy-terminal extension containing two additional sites beyond the three found in H1 \(\alpha\), typical of most somatic H1 molecules. Interestingly, two embryonic or adult urchin H1 histones (\(\beta\) and \(\gamma\)) lack such sites and are not phosphorylated during development\(^18\). In fact, all known sea urchin H1 molecules can be approximated as rather repetitive, variable length Lys/Arg-Ala polymers (with small amounts of Pro) surrounding a conserved globular core, punctuated by variable numbers of SPKK phosphorylation sites. The known sites of phosphorylation of Sp histones in spermatids and pronuclei\(^2,3,7,12,20\) (indicated by asterisks in Fig. 2) confirm that dephosphorylation would result in a massive increase in net positive charge (28–32 per nucleosomal repeat) concentrated in the limited amino-terminal predicted \(\beta\)-turn regions of the Sp histones. Additional charge alterations would also occur in the carboxyl terminus of Sp H1.

**Models of Sp histone function**

Demonstration of phosphorylation and dephosphorylation of the tetrapeptide-containing regions of Sp histones...
Table II. Proteins with multiple SPKK sequences*

| Accession Numbers | Protein | Source (and function) |
|-------------------|---------|------------------------|
| PIR (Genbank/EMBL) | Histones | Most H1 histones |
|                   |         | α and H5s Wheat       |
| A02689, X03473, | H2A     |                        |
| X13171, X00169, |         |                        |
| K26212            |         |                        |
| A02602, S00623    | Nucleo-acid-binding proteins (spermiogenesis) | |
|                   | Sp H1 and Sp H2B Sea urchin |             |
| A02662            | Galline | Chicken                |
|                   | Φ-3 proteins Mussel |                     |
|                   |          |                        |
| A28885, A03714, | Core antigen | Hepatitis viruses |
| A0423             |         |                        |
| A26793            | Nucleocapsid protein | Coronavirus |
| B28136            | Protease | HTLV-1                  |
|                   | Nucleo-acid-binding proteins (virions) | |
| C209659, B28887   | DNA polymerases | Hepatitis, simian immunodeficiency |
| B28133            | DNA glycosylase | Herpes simplex 1 |
| A31844            | Replication protein | Cytomegalovirus |
| JV0049            | Topoisomerase III and IV | E. coli |
|                   | Nucleic acid enzymes |          |
|                   | DNA polymerases | Hepatitis, simian immunodeficiency |
|                   | DNA glycosylase | Heroes simplex 1 |
|                   | Replication protein | Cytomegalovirus |
|                   | Topoisomerase III and IV | E. coli |
|                   | Nucleic acid regulators |         |
| A27345            | Gene 62 | Varicella-zoster (transcription regulator) |
| JS0331            | Vpr | HIV (putative splicing or transcription regulator) |
|                   | A31638 | tr-2 | Drosophila (putative splicing) |
|                   | S06028 | su(w-a) | Drosophila (putative splicing suppressor) |
|                   | A03152 | Retinoiblastoma | Human (tumour suppressor) |
|                   | A26793 | RCC1 | Human (negative regulator of chromosome condensation) |
|                   | SWIS | | Yeast (transcription regulator of HO gene) |

*Criterion for inclusion was at least two different sites per molecule containing Ser/Thr-Pro adjacent to two basic amino acids (Lys and/or Arg).

**PIR, Protein Identification Resource database.**

tones in vivo has led to working models to explain how altered histone-DNA interactions might account for some properties of chromatin in male germ line nuclei and pronuclei.

The chimeric Sp histones are assumed to function during spermatogenesis through their conserved portions, while the extensions are kept from strong interactions with DNA due to their highly phosphorylated, neutralized state. Most of the condensation of spermatid DNA occurs without alteration of the phosphorylation state. In the last stages of spermiogenesis, as the spherical, late spermatid nucleus converts to a conical mature form, the Sp histones are dephosphorylated, unmasking the positive charge of the SPKK repeats, which can then interact with the additional linker DNA to form a highly stable, long-repeat-length sperm chromatin. In this sense, the unusually long linker provides binding sites for the increased charge associated with the amino- and carboxy-terminal extensions. The strong interaction of the extensions with linker DNA would account for the increased thermal stability and resistance to digestion of linker DNA. A possible function of the long linker may be to form three equivalent super-helical turns per repeat, which may optimize packing efficiency.

An altered interaction leading to formation of crosslinks might serve to complete the last steps of condensation, stabilize the condensed state, or assist in the shape change in sperm nucleus formation.

Mature sperm chromatin may be stabilized by crosslinks between adjacent 30 nm chromatin fibers, although there is as yet no direct evidence for crosslinks in vivo. However, in vitro, unphosphorylated Sp H1 is a much more effective crosslinker than phosphorylated Sp H1, inducing salt-dependent folding of long oligonucleosomes, forming "pseudo-higher-order" aggregates of oligonucleosomes, and promoting self-association of DNA filaments. The calculated distances between chromatin fibers in the mature sperm nucleus and the lengths of the Sp histone extensions indicate that there may be extensive interstrand crosslinking in vivo. If crosslinks exist, they could counteract the tendency of the sperm chromatin to swell.

Following fertilization, the amino-terminal regions are repolymerized, which should lower their DNA affinities, and could permit decondensation or the moving apart of chromatin fibers. The phosphorylation events may be permissive for decondensation, but are not sufficient, since they may proceed even though decondensation is blocked. Under these conditions, linker DNA cannot be fully neutralized by the Sp histones, suggesting that swelling does not simply follow from hydration due to the increase in linker charge, but involves a second step. This second transition may require molecules like nuclease, recently implicated in decondensation of frog sperm nuclei in egg lysates.

Both Sp H1 and Sp H2B molecules may interact with linker DNA, which would account for the full neutralization of the additional linker DNA in mature sperm compared to embryo chromatin. H2B histones are usually assumed to be core histones and not associated with linker DNA. However, recent experiments using DNA-histone crosslinking or reductive methylation of histone Lys residues exposed upon dissociation from linker DNA, support the view that the amino termini of Sp H2B do bind to linker DNA.

Although evidence for a β-turn structure for the SPKK motifs and its potential disruption by phosphorylation is still limited, several studies demonstrate that phosphorylation lowers the affinity of Sp histones for DNA. Sp H1 phosphorylated in vivo or in vitro binds less tightly to DNA affinity columns than the unphosphorylated form. Most of the effect appears to be due to the amino-terminal region.

Intermolecular migration of Sp H1 between DNA molecules at moderate salt concentrations is undetectable unless the H1 is phosphorylated. The binding constant of the amino terminus of Sp H1 for DNA is high, estimated by competitive inhibition of Hoechst dye binding to the minor groove, but its alteration upon phosphorylation has not been reported. However, a single phosphorylation of a
13-residue fragment of the protein HMGI, containing one SPKK site and six basic residues altogether, lowers its affinity for DNA in a similar assay by 60%29. This suggests that multiple clustered phosphorylations in the amino-terminal extensions of Sp histones should virtually abolish DNA binding.

**SPKK in non-histone proteins**

SPKK motifs resemble targets for cdc2 kinases30,31. Phosphorylation of H1 histones at mitosis is thought to play a role in mitotic chromosome condensation, but several correlations with interphase condensation have been noted (see Ref. 31 for review). Many cdc2-type kinases seem to be regulated by association with cyclin molecules that act at various phases of the cell cycle. The identification of SPKK sites suggested mechanisms used to regulate histone-DNA interactions might be more widely used to mediate reversible nucleic-acid-binding interactions in other proteins, perhaps in a cell-cycle-regulated fashion. We originally identified a small number of non-histone proteins containing multiple SPKK sites in a protein sequence database by searching for the 16 tetrapeptides containing Ser/Thr-Pro adjacent to two basic residues (Lys or Arg), and have expanded this list more recently (Table II)32. Virtually all proteins containing more than one SPKK were nucleic-acid-binding proteins. They include sperm basic proteins of chicken and mussel, several proteins found in mature virions and a number of nucleic acid enzymes and regulators (such as transcription factors, splicing regulators and topoisomerases). Sperm and viral proteins are involved in highly compact nucleic acid packages, suggesting that dephosphorylation may be used to stabilize tight nucleic acid-protein interactions. Other molecules such as Rb (retinoblastoma protein implicated in cell cycle progression at G1), RCC1 (controlling progression of cells into M phase in human cells) and SWI5 (a G1 activated regulator of transcription of yeast mating type genes) may use the motif for cell-cycle-regulated activities. All three are targets for cdc2 kinases and are located in the nucleus in a cell-cycle-dependent manner.

Alternating the search criteria (to only one tetrapeptide or to closely related sequences) generates many more candidates containing potential cdc2 kinase sites30. These have been suggested to occur with unusual frequency in reg-

**References**

Due to TIBS policy of short reference lists, the number of references cited in this article has been limited. Much of the united work can be found in the reviews cited1,4,18,21,23.

1 Poccia, D. (1986) Int. Rev. Cytol. 105, 1–65
2 Green, G. R. and Poccia, D. (1985) Dev. Biol. 108, 235–245
3 Hill, C. S., Packman, L. C. and Thomas, J. O. (1990) EMBO J. 9, 805–813
4 Poccia, D. (1987) in Molecular Regulation of Nuclear Events in Mitosis and Meiosis (Schlegel, R. A., Halleck, M. S. and Rao, P. N., eds), pp. 149–177, Academic Press
5 Poccia, D. L. (1989) in The Molecular Biology of Reproduction (Schatten, H. and Schatten, G., eds), pp. 115–139, Academic Press
6 Poccia, D., Leiber, T. and Chids, G. (1989) Mol. Reprod. Dev. 1, 219–229
7 Poccia, D. L., Simpson, M. V. and Green, G. R. (1987) Dev. Biol. 121, 445–453
8 Porter, D. C., Moy, G. W. and Vacquier, V. D. (1989) Comp. Biochem. Physiol. 92B, 301–384
9 Suzuki, M. et al. (1990) J. Biochem. 108, 356–364
10 Savic, A., Richman, P., Williamson, P. and Poccia, D. (1981) Proc. Natl Acad. Sci. USA 78, 3706–3710
11 Green, G. R. and Poccia, D. L. (1988) Biochemistry 27, 619–626
12 Hill, C. S. et al. (1991) EMBO J. 10, 1939–1948
13 von Holt, C., deGroot, P., Schwager, S. and Brandt, W. F. (1984) in Histone Genes (Stein, G. S., Stein, J. L. and Marzluff, W. F., eds), pp. 65–116, John Wiley
14 Suzuki, M. (1989) EMBO J. 8, 799–804
15 Suzuki, M. (1989) J. Mol. Biol. 207, 61–84
16 Erard, M., Laknndar-Ghalaz, F. and Amaric, F. (1990) Eur. J. Biochem. 191, 19–26
17 Hyres, R. R. et al. (1989) Nature 339, 73–76
18 Churchill, M. E. A. and Travers, A. A. (1991) Trends Biochem. Sci. 16, 92–97
19 Cohen, L. H., Mfadden, G. and Nardi, R. (1979) J. Cell Biol. 83, 171a
20 Porter, D. C. and Vacquier, V. D. (1988) Biochem. Biophys. Res. Commun. 151, 1200–1204
21 Poccia, D. (1992) in Comparative Spermatology: 20 Years After (Baccetti, B., ed.), pp. 61–96, Raven Press
22 Hill, C. S. and Thomas, J. D. (1990) Eur. J. Biochem. 187, 145–163
23 Poccia, D., Pavan, W. and Green, G. R. (1990) Exp. Cell Res. 188, 226–234
24 Philippot, A., Leno, G. H. and Laskey, R. A. (1991) Cell 65, 569–578
25 Ghaiemi, K. and Kastagiri, C. (1991) Dev. Biol. 148, 295–305
26 Zalenskaya, I. A., Zalensky, A. O., Zalenskaya, E. O. and Vorobjov, V. I. (1986) FEBS Lett. 128, 40–42
27 Bayskin, G. S., Usachenko, S. I., Zalensky, A. O. and Mirzabekov, A. D. (1990) J. Mol. Biol. 212, 495–513
28 Churchill, M. E. A. and Suzuki, M. (1989) EMBO J. 8, 4189–4195
29 Reeves, R., Langan, T. A. and Nissen, M. S. (1991) Proc. Natl Acad. Sci. USA 88, 1671–1675
30 Moreno, S. and Nurse, P. (1990) Cell 61, 549–551
31 Roth, S. Y. and Allis, C. D. (1992) Trends Biochem. Sci. 17, 93–98
32 Meijer, L. and Pordaven, P. (1988) Exp. Cell. Res. 174, 116–129
33 Devereux, J. M., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–396