Histone genes of *Volvox carteri*: DNA sequence and organization of two H3-H4 gene loci

Kurt Müller and Rüdiger Schmitt*

Lehrstuhl für Genetik, Universität Regensburg, D-8400 Regensburg, FRG

Received November 22, 1987; Revised and Accepted March 23, 1988 Accession nos:H3=X06963, H4=X06964

ABSTRACT

Two *Volvox* genomic clones each containing a pair of histone H3-H4 genes were sequenced. In both loci the H3 and H4 genes show outwardly divergent polarity, their coding regions being separated by short intercistronic sequences containing TATA boxes and a conserved 14-bp element. The 3'untranslated regions contain a characteristic motif with hyphenated dyad symmetry otherwise only found associated with animal histone genes. Derived amino acid sequences of histones H3 and H4 are highly conserved and identical between the two sets. The *Volvox* H3 genes both contain one intron whose relative position is shifted by one basepair. Sequence comparisons led to a new interpretation of intron sliding. The *Volvox* H3 gene structure combines the exon-intron organization of fungal H3 and vertebrate H3.3 genes with a termination signal typical for animal H3.1 genes. These features are discussed in view of histone gene evolution.

INTRODUCTION

The classical, highly expressed animal histone genes share the following features: (i) lack of introns, (ii) non-polyadenylated mRNAs and (iii) a typical hyphenated dyad symmetry element in the 3'untranslated region (3'UTR) essential for the 3'end processing of histone mRNA (1,2). In addition to these replication-type histones whose expression is strictly coupled to the S-phase, recent studies have revealed the existence of replacement-type histones, minor variants that are expressed at low levels throughout the cell cycle (3,4,5,6). The corresponding genes frequently contain introns, are transcribed into polyadenylated mRNAs and lack the 3'palindromes. This is similarly true for the histone genes of fungi (7) and protists (8). The latter groups evidently represent a more ancient type of histone genes from which the classical type has evolved, thereby undergoing dramatic

changes in gene structure, while the histone protein structures remained highly conserved.

This recommends the histone gene family as an interesting system to study the evolution of eukaryotic gene structure and to obtain evolutionary markers for the analysis of phylogenetic relationships. As has been shown in a recent comparison of H3 genes (9), differences in gene structure between replication type (H3.1 and H3.2) and replacement-type (H3.3) gene families are paralleled by specific differences in amino acid sequences, codon usage and mode of expression. Based on these data a heterodox evolutionary model in which the intron-bearing, primitive H3.3 genes are the ancestors of the intronless evolved H3.1 genes has been proposed (9). In apparent conflict with this model is the recent observation (10) that maize H3 genes are transcribed into poly-adenylated mRNAs, a feature formerly only attributed to fungal H3 and vertebrate H3.3 genes. On the other hand, the amino acid sequence relates maize histone H3 to the evolved animal H3.1 proteins. These initial data suggest that the evolution of higher plant histone genes may have taken an unusual route and that information on intermediate forms would be useful for clarifying the picture. We have, therefore, sequenced two pairs of histone H3 and H4 genes of *Volvox carteri*, a primitive green alga considered a relative of ancestral land plants.

The results show that *Volvox* histone genes exhibit a combination of structural features not previously described. Specifically, in both sequenced *Volvox* H3 genes an intron (characteristic of fungal H3 and vertebrate H3.3-like genes) is present in combination with a 3' palindromic motif similar to those associated with H3.1-like genes of animals. This suggests that green algae may provide a "missing link" of histone gene evolution. In addition, the exon-intron structure of the *Volvox* H3 genes documents an example of intron sliding within one organism and provides a new interpretation of intron mobility during evolution.

MATERIALS AND METHODS

Strains

Volvox carteri f. *nagariensis* female strain HK10 was kindly supplied by R. Starr (University of Texas, Austin, USA). Synchro-

nous cultures were grown at 28°C in standard *Volvox* medium (11) supplemented with 10mM NH₄Cl under an 8h dark/16h light regime.

Screening of a *Volvox* genomic library

A *Volvox* genomic library in phage vector λ EMBL3 (12) was screened for recombinant clones containing histone H4 genes by *in situ* plaque hybridization (13) using Schleicher & Schuell BA85 nitrocellulose membranes. A 0.4-kb *Bam*HI fragment containing *Xenopus* histone H4 cDNA (from the recombinant plasmid pcX1H4WI; 14) was used as a probe. Five resulting *Volvox* clones positive for histone H4 were probed for the presence of other genes by four DNA fragments containing the *Strongylocentrotus purpuratus* histone genes H1 (1.9-kb *Eco*RI/*Xho*I), H2A (0.5-kb *Hpa*II), H2B (0.9-kb *Bam*HI/*Eco*RI) and H3 (1.2-kb *Bam*HI/*Hpa*II) derived from the recombinant plasmids pSp102 and pSp117 (15). The probes were labelled with ³²P-dCTP (3000 Ci/mmol; Amersham) by nick translation (16). Plasmids pcX1H4WI, pSp102 and pSp117 were kindly provided by H.R. Woodland (University of Warwick, UK).

DNA-Sequencing

Complete DNA sequences of both strands of CopI and CopII were determined by cloning overlapping sets of subfragments of recombinant phage DNA into vectors pUC8 (17), M13mp8 or M13mp9 (18) and by subsequent analysis of double-stranded and single-stranded inserts (19). The dideoxynucleotide chain termination method (20) was applied using synthetic primers and a gradient acrylamide gel as described by Heinrich (21). DNA sequences were processed by the UWGCG program (22) on a VAX-VT100 computer. The sequences reported here will appear in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases under the accession numbers X06963 and X06964.

RESULTS

DNA sequences of two non-allelic H3-H4 gene sets

The *Volvox carter*i genome contains between 10 and 20 histone H4 genes as assessed by Southern hybridization, dot blot hybridization and two independently performed library screenings using *Xenopus* H4 cDNA (14) as a probe (data not shown). Four H4-positive clones isolated from a λ EMBL3 genomic library (12) also contained an H3-positive region linked to the H4 locus suggesting that the two genes may be regularly linked on the *Volvox* genome. These non-allelic H3-H4 gene sets may have originated by duplica-

tion. By comparing the sequences of two such H3-H4 loci we intended to identify conserved signal structures essential for transcriptional control, elucidate the overall gene organization and detect features relevant to their evolution.

The two genomic clones λ VH443 and λ VH352 contained the H3-H4-positive regions on a 1.9-kb *Bam*HI/*Sma*I and a 1.5-kb *Eco*RI/*Rsa*I fragment, respectively, which were ligated into pUC8 (17) to yield the subclones CopI and CopII (Figure 1). Complete DNA sequences of both strands of each insert spanning one H3-H4 gene pair were determined. The two sequences are presented together in Figure 2, aligned with respect to the congruent reading frames of the H3 and H4 genes. Both loci revealed essentially the same overall organization (Figure 1). The H3 and H4 genes are divergently arranged, their coding regions being separated by common 5'UTRs of almost identical lengths (263 bp in CopI, 269 bp in CopII). Both H3 genes contain one intron in positions that differ by one bp and their 3'UTRs exhibit characteristic sequences with hyphenated dyad symmetry.

Coding regions

DNA sequences of the nonallelic gene sets show 51 (H3) and 43 (H4) nucleotide differences, respectively, but encode identical H3 and H4 proteins. The codon usage is clearly biased against A in the third codon position and shows a strong overall preference for G or C, as has been shown for other highly expressed *Volvox* genes (12). It may be noted that the codon UUU for Phe101 of CopII histone H4 has not been found in any H4 gene previously sequenced. A comparison of the *Volvox* H4 predicted amino acid sequence to other H4 proteins (23) revealed three (pea; wheat; maize, 24), four (calf thymus), seven (yeast), ten (*Neurospora*) and 19 (*Tetrahymena*) exchanges, respectively. All replacements occurred at variable positions within the otherwise strictly conserved protein sequence (23). The H4 proteins of calf thymus and higher plants differ only by two conservative exchanges, namely, Val60 for Ile and Lys77 for Arg. *Volvox* H4 protein shares the Arg77 with higher plants, but has Asn in position 60. The other two differences pertain to Thr56 for Gly and Ser69 for Ala. Thus, the primary structure of *Volvox* histone H4 shows significant divergence from higher plant and animal H4 proteins. It is,

	10	20	30	40	50	60	70
Bovine H3.1	ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALRREIRRYQKSTELLIRKLPFQRL						
Bovine H3.2							
Sea urchin H3							
Human H3.3	S.....						
Yeast H3	S.....S.....K.....F.....						
Neurospora H3	S.....S.....K.....						
Wheat/Pea H3	F.....K.....						
Maize H3	F.....K.....						
Tetrahymena H3.1	A.....S.....I.....F.....K.....D.....						
Tetrahymena H3.2	V.....VS.....KF.....K.....T.D.....						
Volvox H3	TA.....K.....						
	80	90	100	110	120	130	
Bovine H3.1	VREIAQDFKTDLRFQSSAVMALQEAASEAYLVGLFEDTNLCIAHAKRVTIMPKDIQLARRIRGERA						
Bovine H3.2	C.....						
Sea urchin H3	E.....						
Human H3.3	A.IG.....						
Yeast H3	IG...SV...S.....A.....QK...K...L...S						
Neurospora H3	IGL...SV.S...S.....QS.....L...N						
Wheat/Pea H3	S...A.....						
Maize H3	A...A.....						
Tetrahymena H3.1	..D..HE..AE.....L...A.....R....T..M.....F						
Tetrahymena H3.2	..D..MEM.N.I...Q..IL...A.....R....T..M.....F						
Volvox H3	Q..L...A.....						

Figure 1. Restriction maps of two H3-H4 genomic clones (1VH443 and 1VH352) and of expanded subclones (CopI and CopII) containing the coding regions. Nucleotide sequences of both strands of CopI (1898bp) and CopII (1504bp) were determined (Figure 2). The aligned coding regions (hatched boxes) and introns (open boxes) assigned to the H3 and H4 genes of CopI and CopII, their divergent polarities (arrows), sizes of intercistronic regions and the 27-bp conserved sequences with hyphenated dyad symmetry (arrows) 25 to 39bp downstream of the translation stops are shown. Numbering above boxed regions refers to codon positions. Relevant restriction sites are marked: B = *Bam*HI, E = *Eco*RI, H = *Hind*III, R = *Rsa*I, S = *Sal*I, Sm = *Sma*I.

however, more closely related to the latter than to the H4 proteins of fungi and ciliates (23).

Comparisons of histone H3 proteins are more illuminating owing to (i) their higher variability and (ii) our knowledge of the

I	3'	CCTAGGGGOTT GAGGGGTACG ATCTGGTGGC AGTCCCGCGA CGGGGAGCGG GACGGAGGTT TTGCTACTTT GTCCATCTTA CAGGGAAGG AGAAGCTATA	100
I	3'	AGGGAGCGGC ATGCTTAGC CTGTACGGAA CGATCGTGAA GTGAAGCGTC TTGTTACCG CTTTATATTT TTTTGGCTCC TCGTGCCTT TTCTTACTTT	200
I	3'	AAACTTACTT GCACACCAAA CAAATTTTGT TGGCCCAAAA TACCACCGCA ACCTGTTGGG TAGCTGGCGT GGAGGTGG	278
II	3'	CAT GCGATGTAGG TTGCTACCA CCAATTTCTT TGTGGCTAT CAACTTCTAT TACTTCAGCA ATACTGCT	71
I	3'	AAT TGG TGG CTT CGG CAT CTC CCA CGC CGG GAC CGC GAA GTC TCG TAT GTG CTG CAG GTA CCG CCA GTG TCA GAA	353
II	3'	G C T T G T A C A A C	146
104		End Gly Gly Phe Gly Tyr Leu Thr Arg Gly Gln Arg Lys Leu Ala Tyr Val Val Asp Met Ala Thr Val Thr Lys	
I	3'	CGC CGC TCG CAC GAG CCA CAT GCA CTG GCT CAG TGC CTA CTG TAA GAG TTC CTT CAA GAA GTC CTG TCA CGC CCA	428
II	3'	T T T T A T C G C	221
79		Arg Arg Ala His Glu Thr Tyr Thr Val Ser Asp Arg Ile Val Asn Glu Leu Phe Asn Lys Leu Val Thr Arg Thr	
I	3'	GAG GAG CAT CTA CTC CGG CGA CTA TGC GAA GTG TGG CGG TGC TGC ACG GTC TGC TGC CTA TCG TCC GAA GCA CTA	503
II	3'	A T G T A C T A G T C	296
54		Glu Glu Tyr Ile Leu Gly Ser Ile Arg Lys Val Gly Gly Arg Ala Leu Arg Arg Ile Ala Pro Lys Thr Ile	
I	3'	TGG GAC TTA CAA CAG CGC TTC GTG GAA CGC TAC CGC GAA TCG AGG AGG GAA CGG GTT CGG GAA AGG CGG GAA CGG	578
II	3'	C A C T A T C C T	371
29		Gly Gln Ile Asn Asp Arg Leu Val Lys Arg His Arg Lys Ala Gly Gly Lys Gly Leu Gly Lys Gly Lys Gly	
I	3'	TGC AGG GCT GTA	590
II	3'	C T	383
4		Arg Gly Ser Met ←H4	
I	5'	GGTACTTAA TGTACTTGG ATAGCAAGG TGAAGTTGG GCAAGATTAG CAGCATGTT GTTCTGTAA TTTGGGAAGT CGCGGGCCTT ATTGGACGTA	690
II	5'	CCAAATGATT ACACTGAACC ATCTGCTTCC ACTTTCAACC GCTTCTAATC GTGCTACAAA CAAAGACATT AAACGGTTCA CGCGCGGAA TAACTGCGAT	690
I	5'	GGTAGTGGT ATATAACAT GTGAGGACAC AAGAGCAGAA CACGGTGGAA AGTGTTCGT TTATAGTGT ATCCAAGCTA TTCGTATTTA TAGCTTTGGT	483
II	5'	CCATCACCAC TATATTGTTA CACTCTGTG TTCTGCTTT GTCCAGCTT TCACAACGCA AATATCACAA TAGGTTGAT AAGCATAAAT ATCGAAACCA	483
I	5'	ATTGTGTGGA CGAGCGTACG CGGACCGCGG TGAGCGGACC TCGAAGCTAT TGCAAATTTG AGAAGCATGC CCATGATATA AACTCCCGCG TCCAAATGAT	790
II	5'	TAAACAACCT GCTCGCATCC CGCTGGGGCC ACTCGCTGGG AGCTTCGATA ACGTTTAAAC TCTTGCTACG GGTACTATAT TTGAGGGGGG AGGTTCACTA	790
I	5'	TTGGACATTA CCACTATGTT CGAGTTTGGT TTGGATCGCG TCAGCGGGGG AACCGAAGTC GTACAAACAC CTCGCGAARA TTCAATAACAC ACCCTAAAT	583
II	5'	AACTGTAAT GTGATACAAA GCTCAACAAA AACTATCGCG AGTGGCGCGG TTGGCTTCAG CATGTTGGTG GAGGCGTTTT AAGTATTGTG TGGGATTTTA	583
I	5'	CTTATACATA GTGATAACAA AGACAAAGCT CTGATATAC TTCTCGCTTA TCAGATTTGT GAA	853
II	5'	GAATGTGAAA ATTGATTGT TCTGTTTGA GAATATATG AAGAGCGAAT AGTCTAACAA CTT	853
I	5'	CACGCTGCAA GATCTTAAC TCGAAGAGCT CTAGAGTGC ACTTGATTT TACTTAACCT TGATACAGA	652
II	5'	CTGGCAGCTT CTAAGATTGA CCGTCTCGCA GATCTCCAGC TGAAGTAAGA ATGAATTGAA ACTATGTCA	652
I	5'	H3→Met Ala Arg Thr Lys Gln Thr Ala Arg Lys Ser Thr Gly Gly Lys Ala Pro Arg Lys Gln Leu Ala Thr	23
II	5'	ATG GCC CGC ACC AAG CAA ACT GGT CGC AAG GCG CCC CGC AAG CAG CTT GCC ACC	922
		T T G G C	721
I	5'	Lys Ala Ala Arg Lys Thr Pro Ala Thr Gly Gly Val Lys Lys Pro His Arg Tyr Arg Pro Gly Thr Val Ala Leu	48
II	5'	AAG GCA GCT CGC AAG ACT CCG GCC ACA GGT GGC GTG AAG AAG CCG CAC CGC TAC CGC CTT GGC ACT GTG GCT CTC	1250
		C A C T C C C G A C	897
		IVS H3-I 181bp IVS H3-II 233bp	
I	5'	Arg Glu Ile Arg Lys Tyr Gln Lys Ser Thr Glu Leu Leu Ile Arg Lys Leu Pro Phe Gln Arg Leu Val Ala Leu	73
II	5'	CGC GAG ATT CGT AAG TAC CAG AAG AGT ACT GAG CTG CTT ATT CGC AAG CTT CTT CCA CGC GTG GTG CGC GAG	1325
		A C C C G C C C C	972
I	5'	Ile Ala Gln Asp Phe Lys Thr Asp Leu Arg Phe Gln Ser Gln Ala Val Leu Ala Leu Gln Glu Ala Ala Glu Ala	98
II	5'	ATT CGC CAG GAC TTC AAG AAG GAT CTG CGC TTC CAG AGC CAG GCC GTG CTG CGC CTG CAG GAG GCT GCC GAG GCC	1400
		C T C C G	1047
I	5'	Tyr Leu Val Gly Leu Phe Glu Asp Thr Asn Leu Cys Ala Ile His Ala Lys Arg Val Thr Ile Met Pro Lys Asp	123
II	5'	TAC CTG GTG GGC CTG TTC GAG GAT ACC AAC CTG TGC GCC ATC CAC GCC AAG CGC GTC ACC ATC ATG CCC AAG GAC	1475
		T G C T	1122
I	5'	Ile Gln Leu Ala Arg Arg Ile Arg Gly Glu Arg Ala End	136
II	5'	ATC CAG CTG GCG CGC CGC ATT CGT GGC GAG CGG GCT TAA	1514
		C T C C G	1161
I	5'	AGCTCGGGCT GCGGTGTGCG CTTCGCCCAA AAATCGGTGT TTTTAAACAC CACCATATTG AAACGCAAGG ACCTTGTGAG TTTGAAGGTT GCTTTCGGCA	1614
II	5'	GAGGTCCAGG ACAGGCCCTCA AAGCAAAATC CGGTGTITTT CAACACCAAC GTGAATTTGA GGATTTTGCT CTAGGCGCGG GGGACATGCT ATTGGTGTGG	1261
I	5'	CCATGGTGTG TCGAGCTGTG ATCATTTTT CTAGATATTT GGGTTTCGCA TGGATCTCTA CAGCGGGCTG CATCGCAAA TGAGCAAGC CTCGGCAACA	1714
II	5'	CAACTTCTTA GTGCGCAATG TTGTGGATTG TATCTACCGG TTGTGTTGCTG AAGGTATGGG GCGCATTCGA TGCCAGGCAT TCGTAACCTT GGAAGGTGAT	1361
I	5'	TATCAAGGCG TTAAGCGAGG CGATCGAAT ACTACCTAAA ATGTACACTG GCACGGCTAC GACACTTTGC CCGTCTGTA CCGCGAATA CCACACAGTG	1814
II	5'	CGAGCTTACC AAATATTGA ATTAGGCTTT CATTACATTT CCATCGCGGT GTTACTATAC CGGTTTGTAGT CCGTCTCTG TCGAAAATGC	1461
I	5'	CCTACCATCC CTAAGTTAAC GTTACATGG GGCAGGAT TGGTATATTT TGGACCTCCC CGTCCATGAC CCGCGTCCCC CGGG	1898
II	5'	CGCTAGCAAA AGCGTTTGA TTTCACGCT TTGCTATGAA TTC	1504

Figure 2. Nucleotide sequences of CopI(=I) and CopII(=II) aligned with respect to their divergently transcribed H3 and H4 genes. Only the coding strands are shown, except in the intercistronic region. Nucleotides in the coding regions of CopI and CopII are identical, except where indicated and the predicted amino acid sequences of the encoded H3 and H4 are given (codon numbering by dots). The shifted positions of single introns (IVS) in H3-I and H3-II are indicated by arrows and their lengths are marked (intron sequences are shown in Figure 6). Numbers at the margin run sequentially and separately for CopI and CopII.

Figure 3. Histone H3 amino acid sequence comparison. Sequences aligned to bovine histone H3.1 are identical (dots), except where indicated (one letter code; Δ = deletion). The upper seven sequences were taken from Wells (23), the maize sequence comes from Chaubet *et al.* (25), the *Tetrahymena* sequences from Hayashi *et al.* (26), the derived *Volvox* sequence is from Figure 2. Note that numbering starts with an Ala following the initiating Met, which is absent from all sequenced H3 proteins (23).

primitive H3.3-like animal and fungal histones. We have, therefore, compared the deduced *Volvox* H3 amino acid sequence to ten other sequences (Figure 3). Residues Ala31, Ser87, Val89 and Met90 distinguish H3.1-like histones from H3.3-type proteins containing Ser31, Ala87, Ile89 and Gly90. Fungal H3 proteins representing the more divergent members of the H3.3 group comprise the diagnostic Ser31, Ile89 and Gly90, whereas higher plant-H3 proteins strongly resemble the H3.1 class in containing Ala31, Ser87 and Val89 (9). Residues Ala31 and Val89 of the *Volvox* histone H3 relate this protein to the H3.1 group; Lys53

and Ala96 are shared by the *Volvox*, higher plant and *Tetrahymena* H3 proteins. On the other hand, there are five differences with reference to higher plant H3, six with reference to bovine H3.1, 18/19 with reference to fungal H3 and 17/20 with reference to ciliate H3 proteins. Another remarkable feature of the *Volvox* H3 sequence is the deletion of Ala29 present in all other sequenced H3 proteins. We, therefore, consider the *Volvox* histone H3 an H3.1 variant that is slightly more similar to the H3 histones of higher plants than to the animal H3.1 class and about as much diverged from the fungal and ciliate H3 as both the higher plant and animal H3 proteins.

Analysis of 5'UTR sequences

The striking similarity between the CopI and CopII gene pairs includes extremely short, overlapping 5'UTRs of 263 bp and 269 bp, respectively, with both regions showing the same low GC content of 41%. In Figure 4 these sequences were aligned to demonstrate overall homology, which required some adjustment (indicated by dots). Screening for local homologies revealed three types of highly conserved sequence elements upstream of each H3 and H4 gene, respectively. (i) Presumptive TATA-boxes precede H3 (TATAAA in CopI at -87 bp; TAAAAAT in CopII at -75 bp) and H4 (AATAA in CopI at -93 bp; TATAAA in CopII at -92 bp). (ii) Following the H3 TATA-box are two conserved motifs, TC/GCAAG at -74/-64 bp and TACTT at -26/-19 bp, the latter resembling the "TCAPyTT" element located between TATA-box and start codon of other H3 genes (23). Flanking the H4 TATA-box are two conserved elements, CAACA at -61/-57 bp and GTCCAA at -111/-106 bp analogous to the "PuPyCA" and "PuTCC" sequences found in comparable positions upstream of other H4 genes (23). (iii) Centered between the intermittent stretches of sequence homology assigned to either the H3 or the H4 genes is the highly conserved 14-bp AT-rich sequence GCAAAATTG/CAG/TAAC thought to be involved in transcription control of both the H3 and H4 genes. Similar elements were found in the 5'UTRs of the divergently transcribed H2A-H2B and H3-H4 genes of yeast (27,28,29).

Analysis of 3'UTR sequences

The 3'UTRs of the sequenced *Volvox* histone genes (Figure 2) contain a 27-bp conserved element starting 25 to 39 bp downstream of the respective stop codons. This sequence closely resembles a

Figure 4. Aligned intergenic regions between the H3 and H4 genes of CopI and CopII. Only one strand is shown. Gaps (dotted) were introduced in both sequences for optimal alignment. Strongly conserved regions are boxed: TATA (---), central 14-bp sequence (....) and gene specific upstream elements (----). Numbers and arrowheads indicate distances to the H3 (▶) or H4 (◀) gene translation starts, numbering at the margin runs sequentially.

conserved 24-bp downstream motif obligatorily present in animal H3.1-type genes (23), but absent from all higher plant, fungal and *Tetrahymena* histone genes analyzed. The consensus deduced from the four *Volvox* sequences not only exhibits substantial homology to the animal consensus sequence, but also shares a characteristic palindromic structure with hyphenated dyad symmetry centered on several T's (Figure 5). This downstream element has been proven essential for efficient 3'processing of sea urchin pre-mRNA (30). Beyond this conserved palindromic region no substantial homologies are present in the 3'UTRs of the four *Volvox* histone genes.

Introns

A comparison of the histone H3 primary structure derived from nucleotide sequences of the *Volvox* H3 genes to the H3 consensus (23) revealed the presence of an intervening sequence in both *Volvox* genes (Figure 2). The positions of the introns were unambiguously determined by (i) restoration of the conserved amino acid sequence and (ii) identity to 5' and 3'splice site consensus sequences. Both introns follow the GT-AT rule and show extensive homology to both plant and animal 5' and 3'consensus sequences (31).

The H3-I gene contains a 253-bp intron splitting codon 46 (Val) in phase I, whereas H3-II contains a 101-bp intron separating codons 45 (Thr) and 46 (Val). Thus, the H3-I intron is shifted one bp downstream to the position of the H3-II intron. To our

```

H3-I      A A A A A T C G G T G T T T T T A A C A C C A C C A
H3-II     A A A A T C C G G T G T T T T T C A A C A C C A C C G
H4-I      A A A A C C C G G T G T T T T T A A A C A C C A C C A
H4-II     A C T A T C C G G T G T T T C T T A A C A C C A C C A

V-CON     AAAATCCGGTGT TTTT TAACACCACCA
          :::::  :::  :::::  :::  :::::  :::
A-CON     AAAA...GGCTCTTTTTCAGAGCCACCA
          C      C      G      G      C
          <----->

```

Figure 5. Top: Conserved sequences with hyphenated dyad symmetry downstream of the *Volvox* H3 and H4 genes (suffices I and II refer to CopI and CopII, respectively). **Bottom:** Derived *Volvox* consensus sequence, V-CON, compared to the animal consensus (A-CON;23). Arrows indicate dyad symmetry.

knowledge this is the first case of intron sliding found in a member of a multigene family within one organism. By direct comparison the two introns exhibit 38% homology, but insertion of appropriate gaps increases the homology to 51%. As shown in Figure 6, sequence similarities are non-randomly distributed. The homology of the longer H3-I intron to the entire H3-II intron is confined to the 3' half of the former suggesting that the shorter H3-II intron may have been derived from a copy of the H3-I intron by a deletion during divergent evolution of the two histone gene loci. A model accounting for the observed variations in position and nucleotide sequence of analogous introns will be discussed below.

DISCUSSION

Gene organization

The sequences of two analyzed *Volvox* histone loci revealed an identical overall organization. H3 and H4 genes show outwardly divergent transcription, the coding regions varying only in silent nucleotide changes and being separated by intercistronic regions of almost identical lengths. Both H3 genes are split by one intron whose relative position is shifted over one bp. These common features strongly support our notion that both H3-H4 pairs arose from one ancient locus by gene duplication. Sequence comparisons between the putative promoter regions of both loci revealed structures homologous in sequence and position (Figure 4)

that have been conserved during evolution probably due to their functional importance for the control of gene expression. We have assigned TATA-boxes to all four genes. Two additional short stretches of sequence homologies resembling conserved H4-specific elements precede the *Volvox* H4 genes. Another two elements are found upstream of the H3-start codons, one showing homology to a similarly positioned conserved H3-specific sequence (23). All four genes lack a CAAT-box; but as a recent compilation analysis of eukaryotic promoter sequences has revealed, this element cannot be considered a universal eukaryotic promoter signal (32).

A 14-bp AT-rich structure located centrally in the overlapping upstream region of each H3-H4 gene set exhibits striking homology to an upstream activating sequence (UAS) found in yeast histone genes, where it promotes cell cycle-dependent activation of transcription. Histone expression in *Volvox* appears to be strictly regulated during life cycle (unpublished observation); thus the *Volvox* AT-rich element may perform an activating function analogous to yeast UAS elements.

Intron sliding

An interesting feature of the sequenced *Volvox* H3 genes is their possession of an intron whose relative position is shifted over one bp. Comparison of exon-intron structures of genes belonging to the same multigene family but present in different organisms have frequently been performed to obtain clues on the origin and the role of introns in gene evolution. An extensive analysis of the serine protease superfamily has lead Rogers (33) to the view that differently positioned introns may have originated by independent intron insertions rather than by classical mutation of pre-existing introns. On the other hand, exon-intron patterns of triosephosphate isomerase genes from maize and chicken have been interpreted in terms of two or more compensating mutational events (34). But until now, no compelling evidence for one or the other pathway of intron sliding during the evolution of a particular gene has been provided.

Similarities between the two sequenced *Volvox* H3 genes strongly suggest that they originated by a gene duplication event. As introns evidently are a primitive mark of H3 genes (see below) and in view of the substantial sequence homology between these

introns (Figure 6), we conclude that the divergence of intron position occurred subsequent to duplication of the locus and thus likely constitutes an example of intron sliding by deletion and insertion of nucleotides. To accomplish a one-bp shift of intron position, at least two events are necessary to restore the coding sequence: one changing the 5'splice site and one producing a compensating shift around the 3'splice junction. A comparison of both intron sequences (Figure 6) shows that homology between the H3-I and H3-II introns starts within a sequence CTGGTTAC strongly resembling the 5'splice junction CTGGTAAG that interrupts the H3-I coding region. This may be accounted for by a deletion event removing the 5'terminal 132 bp of the H3-I intron -- including the 3'terminal TG of exon 1 -- that shifts the intron one bp in the 5'direction and generates a new 5'splice site (31). As a compensating event at the 3'splice junction, an insertion of one G would be sufficient to restore the protein sequence without changing the 3'splicing position. Alternatively, a secondary deletion of the nonamer CATT TTTTA preceding the 3'splice junction of H3-I plus a transversion from T to A100 in H3-II will also generate the observed intron shift. Clearly, such secondary mutation events, though rare, would be favored by selective pressure. On the other hand, the transient inactivation of one H3 gene by the proposed deletion event would not have been particularly deleterious in view of the 10 to 20 gene copies present in the *Volvox* genome.

Figure 6. Intron sequences of H3-I and H3-II genes aligned by best fit. Only one strand is shown as in Figure 2. Note that homologies are restricted to the 3'portion and start within a sequence CTGGTTAC (boxed) strongly resembling the 5'splice junction CTGGTAAG of the H3-I intron.

Although we can only compare the available intron sequences, it is obvious that the *Volvox* H3 genes provide excellent examples for elucidating mechanisms of intron sliding during gene evolution. It will be interesting to obtain sequences of the remaining *Volvox* H3 genes and examine their intron structures in search for additional clues concerning the pathway of intron evolution in the gene family.

Evolution of histone H3 genes

In classification systems based on morphological, anatomical, cytological and biochemical approaches (35,36,37), the green algae are generally regarded as contemporary relatives of organisms ancestral to land plants. Sequence comparisons of 5S rRNA (38) and 18S rRNA loci (H.Rausch, N.Larsen and R.Schmitt, to be published) support such a view. The primitive green alga *Volvox carteri* was therefore considered appropriate for elucidating the phylogeny of higher plant histone genes.

The presence of an intron in the *Volvox* H3 genes is a property shared by fungal H3 and vertebrate H3.3 genes but not by the replication-type H3 genes of animals (9). Nevertheless, the following features indicate that the *Volvox* H3 genes are closely related to the latter class.

First, the deduced amino acid sequence diverges in several places from all previously known H3 proteins, but in the diagnostic positions it clearly exhibits closer identity to H3.1- than to H3.3-type proteins (Figure 3).

Second, the codon usage pattern (12) resembles that previously reported for H3.1-type genes (being strongly biased towards G or C, and disfavoring A in the degenerate third base position) as distinguished from the H3.3 pattern, where third position A's and T's are used with statistical frequency (9).

Third, our preliminary studies indicate that the *Volvox* histone genes show the replication-dependent mode of expression characteristic of H3.1-type genes (2,6).

Fourth, all analyzed *Volvox* histone genes contain the hypenated dyad symmetry element characteristically found in the 3'UTRs of animal H3.1 genes (whose mRNAs are not poly-adenylated), but absent from ciliate and fungal histone genes and the animal H3.3 genes (whose mRNAs do contain a polyA tail). Yet, it

has to be shown experimentally that polyadenylation is not involved in *Volvox* histone gene expression. Interestingly, higher plant histone genes generally lack a 3'palindrome and, as recently reported for maize H3 and H4 genes, are transcribed into polyadenylated mRNAs (10).

On first sight, these differences between *Volvox* and higher plant histone genes are disconcerting, since the anticipated intermediate in plant histone gene evolution does not become obvious. Rather, the 3'palindrome relates *Volvox* histone genes to animal H3.1 genes, whereas poly-adenylation relates the maize histone genes to fungal H3 and to vertebrate H3.3 genes. Conversely, the split-gene organization associates *Volvox* H3 genes with fungal and vertebrate H3.3-like genes, whereas the lack of introns in higher plant histone H3 genes classifies them as evolved H3.1-like genes.

Two alternative pathways of plant histone gene evolution emerge from the available, though unsufficient, data.

(i) *Volvox* marks a true phylogenetic step on the route to higher plants. In this case, the 3'palindrome of *Volvox* histone genes was exchanged for a new termination signal (39,40) that triggers poly-adenylation of transcripts, a feature reminiscent of fungal H3 and vertebrate H3.3 genes. The polyA signal may have been readapted from the non-histone genes operative in plants.

(ii) *Volvox* marks an evolutionary line that very early separated from the route leading to higher plants. This alternative assumes that histone genes representative of the two modes of transcription termination were present in a common ancestor of plants and animals (a situation that has been conserved in vertebrates). Early divergence of *Volvocales* and higher plants was accompanied by the loss of one or the other gene variant, an event, which led to higher plant histone genes with a polyA signal and *Volvox* histone genes with a 3'palindrome.

A decision between these alternatives may well be possible once the structure of more plant histone genes, especially those of other green algae, are known.

ACKNOWLEDGMENTS

We thank Richard Starr for the provision of *Volvox* strain HK10 and Hugh R.Woodland for the gift of DNA probes, Mike Salbaum for

stimulating discussions that initiated this work, and David Kirk for critical review of the manuscript. This investigation was supported by the Deutsche Forschungsgemeinschaft (SFB 43).

*To whom correspondence should be addressed

REFERENCES

1. Hentschel, C.C. and Birnstiel, M.L. (1981) *Cell* 25, 301-313.
2. Old, R.W. and Woodland, H.R. (1984) *Cell* 38, 624-626.
3. Engel, J.D., Sugarman, B.J., Dodgson, J.B. (1982) *Nature* 297, 434-436.
4. Harvey, R.P., Whiting, J.A., Coles, L.S., Krieg, P.A., Wells, J.R.E. (1983) *Proc. Natl. Acad. Sci. USA* 80, 2819-2823.
5. Wells, D. and Kedes, L. (1985) *Proc. Natl. Acad. Sci. USA* 82, 2834-2838.
6. Zweidler, A. (1984) In Stein, G.S., Stein, J.L. and Marzluff, W.F. (eds.), *Histone Genes*, pp. 339-371, John Wiley & Sons Inc., New York.
7. Woudt, L.P., Pastnik, A., Kempers-Veenstra, A.E., Jansen, A.E.M., Mager, W.H., Planta, R.J. (1983) *Nucl. Acids Res.* 11, 5347-5360.
8. Wu, M., Allis, C.D., Richman, R., Cook, R.G., Gorovsky, M.A. (1986) *Proc. Natl. Acad. Sci. USA* 83, 8674-8678.
9. Wells, D., Bains, W., Kedes, L. (1986) *J. Mol. Evol.* 23, 224-241.
10. Chaubet, N., Chaboute, M.-E., Clement, B., Ehling, M., Phillips, G., Gigot, C. (1988) *Nucl. Acids Res.* 16, 1295-1304.
11. Starr, R.C. (1969) *Arch. Protistenk.* 111, 204-222.
12. Mages, W., Salbaum, M.J., Harper, J.F. and Schmitt, R. (1988) *Mol. Gen. Genet.*, in press.
13. Benton, W.D. and Davis, R.S. (1977) *Science* 196, 180-182.
14. Turner, P.C. and Woodland, H.R. (1982) *Nucl. Acids Res.* 10, 3769-3780.
15. Sures, I., Maxam, A., Cohn, R.H., Kedes, L.H. (1976) *Cell* 9, 495-502.
16. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning, a Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
17. Vieira, J. and Messing, J. (1982) *Gene* 19, 259-268.
18. Messing, J. and Vieira, J. (1982) *Gene* 19, 269-276.
19. Sanger, F., Coulson, A.R., Barrell, B.G., Smits, A.J.H. and Roe, B.A. (1980) *J. Mol. Biol.* 143, 161-178.
20. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
21. Heinrich, P. (1986) *Guidelines for Quick and Simple Plasmid Sequencing*. Boehringer Monography, Mannheim.
22. Devereux, J., Haeberli, P., Smithies, O. (1984) *Nucl. Acids Res.* 12, 387-395.
23. Wells, D.E. (1986) *Nucl. Acids Res. Suppl.* 14, r119-149.
24. Phillips, G., Chaubet, N., Chaboute, M.-E., Ehling, M., Gigot, C. (1986), *Gene* 42, 225-229.
25. Chaubet, N., Philipps, G., Chaboute, M.E., Ehling, M. and Gigot, C. (1986) *Plant Mol. Biol.* 6, 253-263.
26. Hayashi, T., Hayashi, H., Fusauchi, Y., Iwai, K. (1984) *J. Biochem.* 95, 1741-1749.

27. Hereford, L., Fahrner, K., Woolford, J. Jr., Rosbash, M., Kaback, D. B. (1979) *Cell* 18, 1261-1271.
28. Smith, M. M. and Murray, K. (1983) *J. Mol. Biol.* 169, 641-661.
29. Osley, M. A., Gould, J., Kim, S., Kane, M., Hereford, L. (1986) *Cell* 45, 537-544.
30. Georgiev, O. and Birnstiel, M. L. (1985) *EMBO J.* 4, 481-489.
31. Brown, J. W. S. (1986) *Nucl. Acids Res.* 14, 9549-9559.
32. Bucher, P. and Trifonov, E. N. (1986) *Nucl. Acids Res.* 14, 10009-10026.
33. Rogers, J. (1985) *Nature* 315, 458-459.
34. Marchionni, M. and Gilbert, W. (1986) *Cell* 46, 133-141.
35. Scagel, R. F., Bandoni, R. J., Rouse, G. E., Schofield, W. B., Stein, J. R., Taylor, T. M. C. (1969) *Plant Diversity: An Evolutionary Approach*, Wadsworth Publishing Company Inc., Belmont CA.
36. Stefureac, T. I. (1972) *Symp. Biol. Hung.* 12, 213-225.
37. Whittacker, R. H. (1977) In Kreier, J. P. (ed.), *Parasitic Protozoa*, Vol. 1, pp. 1-34, Academic Press, New York.
38. Hori, H., Lim, B.-L., Osawa, S. (1985) *Proc. Natl. Acad. Sci. USA* 82, 820-823.
39. Tabata, T. and Iwabuchi, M. (1986) *Mol. Gen. Genet.* 196, 397-400.
40. Tabata, T., Terayama, C., Mikami, K., Uchiyama, H., Iwabuchi, M. (1987) *Plant Cell Physiol.* 28, 73-82.