Structure of allelic variants of subtype 5 of histone H1 in pea *Pisum sativum* L.

VS Bogdanova[1], DR Lester[2], VA Berdnikov[1] and I Andersson[3]

[1]*Institute of Cytology and Genetics, Siberian Division of the Russian Academy of Sciences, Novosibirsk 630090, Russia;* [2]*Institute of Cell and Molecular Biology, Uppsala University, PO Box 596, S-751 24 Uppsala, Sweden;* [3]*Department of Molecular Biology, Swedish University of Agricultural Sciences, PO Box 590, S-751 24 Uppsala, Sweden*

The pea genome contains seven histone H1 genes encoding different subtypes. Previously, the DNA sequence of only one gene, *His1*, coding for the subtype H1-1, had been identified. We isolated a histone H1 allele from a pea genomic DNA library. Data from the electrophoretic mobility of the pea H1 subtypes and their *N*-bromosuccinimide cleavage products indicated that the newly isolated gene corresponded to the H1-5 subtype encoded by *His5*. We confirmed this result by sequencing the gene from three pea lines with H1-5 allelic variants of altered electrophoretic mobility. The allele of the slow H1-5 variant differed from the standard allele by a nucleotide substitution that caused the replacement of the positively charged lysine with asparagine in the DNA-interacting domain of the histone molecule. A temperature-related occurrence had previously been demonstrated for this H1-5 variant in a study on a worldwide collection of pea germplasm. The variant tended to occur at higher frequencies in geographic regions with a cold climate. The fast allelic variant of H1-5 displayed a deletion resulting in the loss of a duplicated pentapeptide in the C-terminal domain.
Heredity (2005) **94,** 582–588. doi:10.1038/sj.hdy.6800650
Published online 23 February 2005

Keywords: pea; *Pisum sativum*; histone H1; allelic variants

Introduction

The DNA of the eukaryotic cell is packaged into chromatin by histones and nonhistone proteins. The fundamental chromatin unit, the nucleosome, is an octameric complex of the four conserved core histones with 146-bp DNA wrapped around it (van Holde, 1988). Adjacent nucleosomes are separated by a 20-bp stretch of linker DNA with a single less conserved linker histone, or H1, associated. A histone H1 has a characteristic three-domain structure made up of a central globular domain, flanked by unstructured N- and C-terminal tails (Wolffe *et al*, 1997). Most multicellular organisms have several subtypes of linker histones which sometimes display developmentally regulated or tissue-specific expression (Stein *et al*, 1984; Tanaka *et al*, 1999). However, the functional significance of histone H1 heterogeneity is obscure.

Since H1 subtypes can differ in their strength of interaction with DNA (Liao and Cole, 1981; De Lucia *et al*, 1994) or intranuclear distribution (Schulze *et al*, 1993), it is tempting to suppose that patterns of genomic distribution of histone H1 variants create differential 'molecular environments' for gene transcription. Another type of histone H1 variability is allelic polymorphism, which occurs most frequently in plant families. In Fabaceae, allelic variants of histone H1 subtypes are common (Berdnikov *et al*, 1992, 2003; Belyaev and Berdnikov, 1985). Variation in the structure of allelic variants might influence gene expression over the genome. For example, allelic substitutions in histone H1 loci of legumes were associated with small statistically significant differences in a number of quantitative traits (Bogdanova *et al*, 1994; Berdnikov *et al*, 1999, 2003).

Protein electrophoretic studies on a worldwide collection of 883 accessions of the garden pea (*Pisum sativum* L.) germplasm have revealed several allelic variants within each of seven H1 subtypes (Berdnikov *et al*, 1993b). All allelic variants within subtypes 3, 4 and 7 showed random territorial distribution, while the distribution of certain allelic variants within subtypes 1, 6 and 5 correlated with the geographic origin of the accession. One allele of subtype 5 was especially interesting. It was most abundant in regions where the mean vegetational temperature was low. The distribution could not be explained in historical terms and implied that the allele had physiological significance (Berdnikov *et al*, 1993b). This finding was consistent with the hypothesis that histone H1 participates in the adaptation to new environments. This concept had emerged from a detailed electrophoretic study on insects, in which it was found that the intra-order variation of H1 molecule length correlated with the number of species in an order but not with its evolutionary age (Berdnikov *et al*, 1993a).

The genomic organisation of histone H1 gene clusters might reflect their functional differentiation (Trieschmann *et al*, 1997). In garden pea, seven genes, *His1–His7*, encoding the various subtypes of histones H1 lie in three regions of the genetic map, with *His2–His6* tightly linked (Trusov *et al*, 1994; Kosterin *et al*, 1994). Only *His1* has been coupled with its DNA sequence

(Berdnikov et al, 2003). Gantt and Key (1987) sequenced cDNA coding for a H1 subtype, and Woo et al (1995) described the sequence of a putative histone H1; however, the relation of these two sequences to His1–His7 is unclear.

In this study, we use electrophoretic data together with available and newly acquired DNA sequences to further characterise the histones H1 of pea. We identify the DNA sequence of three alleles of His5 and gain insight into the structural background of adaptive changes in histone H1 construction.

Materials and methods

Plant material
The pea cultivar Torsdag provided DNA for genomic DNA library construction. The pea accessions: VIR-3971 (Kirov), VIR-4362 (Vologda), VIR-6560 (Tadjikistan) provided DNA for sequencing of the H1-5 alleles, H1-5-1, H1-5-2 and H1-5-3, respectively. Lathyrus sativus, Lens culinaris, and Vicia faba leaves were used to isolate and analyse histone H1 electrophoretic mobility.

Genomic DNA extraction
Genomic DNA was isolated from pea lines according to the method of Ellis (1994).

Histone H1 gene isolation and sequencing
A pea (cv. Torsdag) genomic DNA library was constructed with a Universal Genome Walker kit (BD Sciences), using methods recommended by the manufacturer. Degenerate PCR primers were designed according to conserved regions of known pea histones H1. The Universal Genome Walker strategy was employed together with these primers to obtain the 3′ end of a histone H1-like gene. In the first round of PCR, the primer 5′-CARTAYGCIATIRCIAARTTIATYGARGAR AA-3′ was used versus the Genome Walker adaptor 1 (AP1) primer. In the second round, the primer 5′-AARAARWIIGTIGCIWSIGGIAARCTIRT-3′ was used versus the Genome Walker Adaptor 2 (AP2) primer. Reactions were held at 95°C for 60 s and then cycled 35 times between 94°C for 59 s, 46°C for 59 s and 72°C for 60 s. PCR products were ligated into pGEM-T (Promega) and sequenced using Big Dye Terminator methods (Applied Biosystems). Two primers were designed according to the sequence of a 750 bp product. These two primers, 5′-GCTTTTGGCTTAGCAACTGTG-3′ and 5′-TTCATACAAGCTTTCAGCAGCGGCC-3′, were used in successive rounds of nested PCR versus the AP1 and AP2 primers, respectively to isolate the 5′ end of the histone H1-like gene from the Genome Walker library. PCR reactions were held at 95°C for 60 s and then cycled 35 times between 94°C for 59 s, 58°C for 59 s and 72°C and 60 s. The resulting 650 bp product was sequenced and found to overlap with the 3′ fragment. The combined sequence of these two fragments enabled the design of the PCR primers 5′-CCACACTCATTTCACTATTTAA ACC-3′ and 5′-AGCATTGTAAAGATGTTTTGTGTTC-3′. These primers (7.5 pmol) were used on a template of genomic DNA of the VIR-4362 line (about 5 ng) with the use of Advantage cDNA polymerase mix (CLONTECH). Conditions for PCR were 95°C for 60 s and then 35 cycles of 94°C for 59 s, 58°C for 59 s, and 72°C for 60 s. A PCR product of 1025 bp was obtained and sequenced.

Sequence analysis
Multiple sequence alignments were made with Clustal X, version 1.8 (Thompson et al, 1997). The alignments were further adjusted by eye to minimise insertion/deletion events. Synonymous divergence between DNA sequences was estimated by the Kumar method that is included in MEGA package version 2.1 (Kumar et al, 2001). Corresponding standard errors were obtained after 1000 bootstrap trials. Estimation of the divergence time was based on synonymous distances, d_S, and mean rate of synonymous substitutions accepted to be 6.96×10^{-9} per site per year, as estimated for actin genes of dicot plants (Moniz de Sa and Drouin, 1996).

Histone H1 isolation and electrophoresis
Histone H1 proteins were isolated by an express method (Rozov et al, 1986; Kosterin et al, 1994) based on the Johns (1964) method. The preparations were subjected to electrophoresis in slabs of 15% polyacrylamide/0.5% N,N′- methylenbisacrylamide gel containing 6.25 M urea and 0.9 M acetic acid following a modification (Berdnikov and Gorel, 1975) of Panyim and Chalkley's (1969) method. After electrophoresis, gels were stained in 0.01% (w/v) Coomassie R-250 in 0.9 M acetic acid and destained by diffusion in 0.9 M acetic acid.

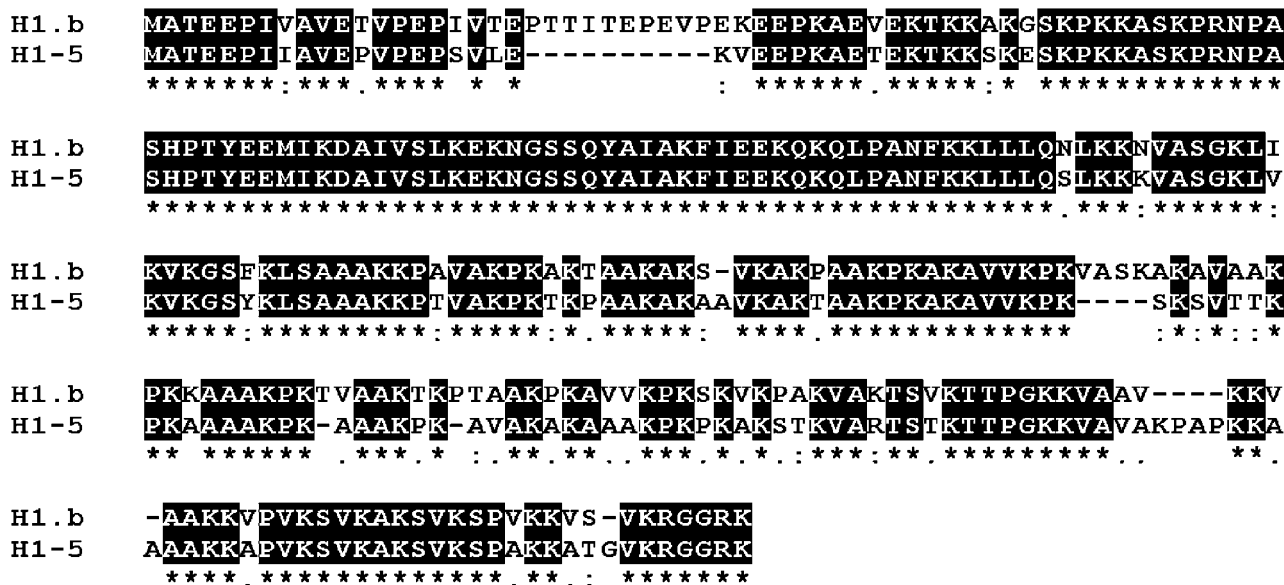## Results

Isolation of a new pea histone H1 gene
We isolated a gene from the pea Genome Walker library (submitted to GenBank as AY231151) that coded for a protein with the typical characteristics of histone H1. The length of the predicted protein is 255 amino acids and it displays a central hydrophobic domain containing aromatic amino-acid residues flanked by positively charged N- and C-tails.

The derived amino-acid sequence of the novel histone H1 most closely resembles the previously identified subtype H1.b (Gantt and Key, 1987); they share 82 % identity (Figure 1). The highly conserved globular domains of the proteins differ by four amino-acid replacements, one of which, Lys→Asn, is associated with a change of electric charge. The newly isolated histone H1 subtype is notable in that, unlike most plant H1 histones, it does not contain the tripeptide FKL at the C-terminal border of its globular domain. Instead, the sequence YKL is present. The N-terminal domain of the novel H1 protein is shorter than that of H1.b due to a deletion of 10 amino acids (Figure 1). Divergence between the novel histone H1 sequence and H1.b is seen mainly in the C-terminal domain; identity in this part of the molecule is only 76%. The features of C-terminal region of the newly isolated histone H1 are described below.

Identification of the novel H1 subtype in electrophoregram
Histone H1 of pea leaves contains up to seven subtypes resolved by gel electrophoresis in the presence of urea and acetic acid (Kosterin et al, 1994). Multiple results from electrophoretic studies on histones H1 s hed light on the identity of the newly isolated histone H1 gene.

An analysis of products formed after cleavage of pea H1 subtypes with N-bromosuccinimide allowed

```
H1.b  MATEEPIVAVETVPEPIVTEPTTITEPEVPEKEEPKAEVEKTKKAKGSKPKKASKPRNPA
H1-5  MATEEPIIAVEPVPEPSVLE----------KVEEPKAETEKTKKSKESKPKKASKPRNPA
      ******:***.**** * *           :  *****.*****:* ************

H1.b  SHPTYEEMIKDAIVSLKEKNGSSQYAIAKFIEEKQKQLPANFKKLLLQNLKKNVASGKLI
H1-5  SHPTYEEMIKDAIVSLKEKNGSSQYAIAKFIEEKQKQLPANFKKLLLQSLKKKVASGKLV
      ***********************************************.***:******:

H1.b  KVKGSFKLSAAAKKPAVAKPKAKTAAKAKS-VKAKPAAKPKAKAVVKPKVASKAKAVAAK
H1-5  KVKGSYKLSAAAKKPTVAKPKTKPAAKAKAAVKAKTAAKPKAKAVVKPK----SKSVTTK
      *****:*********:*****.*:.***** ****.************    :*:*::*

H1.b  PKKAAAKPKTVAAKTKPTAAKPKAVVKPKSKVKPAKVAKTSVKTTPGKKVAAV----KKV
H1-5  PKAAAAKPK-AAAKPK-AVAKAKAAAKPKPKAKSTKVARTSTKTTPGKKVAVAKPAPKKA
      ** ******  .***.*  :.**.**..***.*.*:***:**.*********..   **.

H1.b  -AAKKVPVKSVKAKSVKSPVKKVS-VKRGGRK
H1-5  AAAKKAPVKSVKAKSVKSPAKKATGVKRGGRK
       ****.**************.**.: *******
```

Figure 1 An alignment of pea histone H1 subtypes H1.b (Gantt and Key, 1987) and novel subtype (H1-5). Sequence identities are shaded in black.

Kosterin et al (1994) to locate the positions of tyrosine residues in the molecule. It was shown that H1-1 and H1-5 contained a residue susceptible to N-bromosuccinimide cleavage at the border of the G- and C-domains, while H1-2, H1-3, H1-4 and H1-6 did not. This implied that H1-1 and H1-5 contained a tyrosine at the border site. This result was confirmed for H1-1 from the sequence of its gene (Berdnikov et al, 2003). The derived amino-acid sequences of other H1 genes, H1.b (Gantt and Key, 1987) and H1.41 (Woo et al, 1995), display phenylalanine at the corresponding position. Considering that phenylalanine is generally conserved in plant histones H1, the lack of N-bromosuccinimide cleavage in H1-2, H1-3, H1-4 and H1-6 is probably a result of the presence of the tripeptide FKL at the border of G-domain. The presence of the tripeptide YKL corresponding to the newly isolated histone H1 gene suggested that it encoded subtype H1-5.

The electrophoretic mobility, V, of a protein in PAGE in the presence of 0.9 M acetic acid and urea is a function of its length, L (the number of amino-acid residues) and positive charge, Z (Berdnikov et al, 1993a). Under the electrophoresis conditions used (pH 3.2), Z is equal to the sum of lysine, arginine and histidine residues. To determine the character of function relating V with L and Z we measured the mobilities of Vicieae H1 subtypes of known primary structure, their fragments produced by N-bromosuccinimide cleavage and the pea histone H4 (Table 1). It turned out that the equation $V = \alpha Z L^{\beta}$ perfectly described the data, where the constants α and β were estimated to be 583.6 and −1.89, respectively. After logarithmic transformation, the relationship was linearised (Figure 2). The theoretical electrophoretic mobility of the sequenced subtype could then be estimated as 1.21, a value most closely corresponding to the electrophoretic mobility of H1-5 (Table 1). Thus, the hypothesis that the protein encoded by the newly sequenced gene was H1-5 was supported.

Table 1 The electrophoretic mobility (V), length (L) and positive charge (Z) of some histones and their fragments

Number	Name	L	Z	V^{a}	$V_T{}^{a}$
1	Pea H1-1-3	290	79	1.01	1.01
2	Pea H1-1-2	296	81	1.00	1.00
3	Pea H1-1-1	301	82	0.99	0.98
4	Pea H1-1-3(T_3C^b)	170	56	1.98	1.97
5	Pea H1-1-2(T_3C^b)	176	58	1.91	1.91
6	Pea H1-1-1(T_3C^b)	181	60	1.88	1.88
7	Pea H1-1-3(T_2C^b)	211	67	1.57	1.57
8	Pea H1-1-2(T_2C^b)	217	69	1.52	1.53
9	Pea H1-1-1(T_2C^b)	222	70	1.49	1.47
10	Lat. sativus H1-1-1	306	82	0.94	0.95
11	Lat. sativus H1-1-2	295	82	1.01	1.02
12	Lentil H1-1-1	293	79	1.00	0.99
13	Lentil H1-1-2	281	77	1.06	1.05
14	V. faba H1-1	278	78	1.05	1.08
15	Pea H4	102	27	2.49	2.50
16	H1.b (Gantt and Key, 1987)	264	74		1.13
17	H1.41 (Woo et al, 1995)	184	52		1.58
18	Putative pea H1-5	255	74		1.21
19	Pea H1-2[c]			1.10	
20	Pea H1-3[c]			1.11	
21	Pea H1-4[c]			1.15	
22	Pea H1-5[c]			1.19	
23	Pea H1-6[c]			1.44	
24	Pea H1-7[c]			1.60	

L = number of amino-acid residues; Z = number of lysines, arginines and hystidines; V_T = calculated mobility with the use of the approximation formula $V = 583.6 Z L^{-1.89}$.
[a]Mobility of the pea standard variant H1-1-2 accepted to be 1.0.
[b]T_2C and T_3C are fragments of the molecule from the second and third tyrosine residues to the C-terminus.
[c]Mobility of the most common allelic variants is given.

Analysis of H1-5 allelic variants

An electrophoretic analysis of a pea worldwide germplasm collection (Berdnikov et al, 1993b) revealed three alleles of the His5 gene (Figure 3). We sequenced the
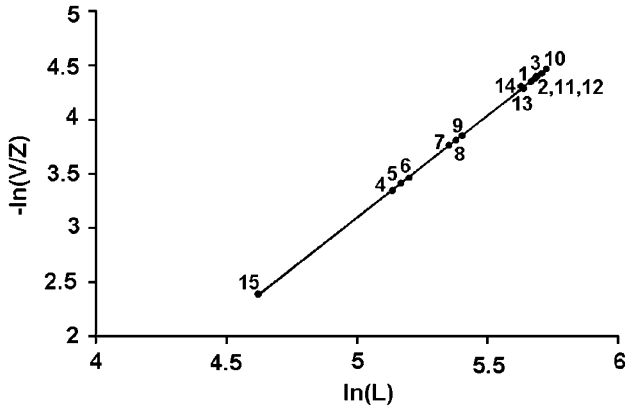
Figure 2 The relationship of electrophoretic mobility (*V*), molecular length (*L*) and net positive charge (*Z*) of a protein. Calculations were made using Microsoft Excel 2000. Numbers in the plot correspond to those of Table 1.

presumptive *His5* gene from the pea germplasm accessions known to contain slow (VIR-3971), standard (VIR-4362) and fast (VIR-6560) allelic variants.

The slow allele from the accession VIR-3971 differed from the standard one by a single nucleotide substitution G→T. This change results in the substitution of a positively charged lysine 178 with the neutral asparagine in the middle of the C-tail (Figure 4). Such a reduction in charge should cause a decrease in electrophoretic mobility that would register under the conditions employed (Figure 3).

The fast allele from the accession VIR-6560 differed from the standard allele by a deletion of 15 nucleotides in the C-tail and five nucleotide substitutions, two of which cause amino-acid changes (Figure 4). Alanine is replaced by glycine in the globular domain and in the C-tail glycine is replaced by arginine. The variation in protein length and charge displayed by the allele are expected to cause an increase of electrophoretic mobility.

The calculated mobilities of the three sequenced allelic variants are 1.19, 1.21 and 1.24, where mobility of the most abundant H1-1 subtype is adopted as unity. These values fit well with the observed ones; the mobilities of H1-5 variants shown in Figure 3 are related as 1.19 : 1.21 : 1.23.

The correlation between the electrophoretic mobility of H1-5 variants and the structure of the predicted proteins of the sequenced alleles allows us to identify the newly described histone H1 gene as *His5* coding for subtype H1-5.

C-domain structure of subtype H1-5

The C-domain of H1-5 has an amino-acid composition typical for plant H1 histones with more than 70% represented by the sum of lysine (32%), alanine (29%) and proline (10%). Lysine makes up a third of the domain, accounting for 44 of its 138 amino acids.

If the lysines were distributed evenly, the mean spacing between them would be two amino acids. However, there are only five spacers of this length (of 45 possible), four of which are in the terminal half of the C-domain. About 10% of the lysine residues (4–5) are
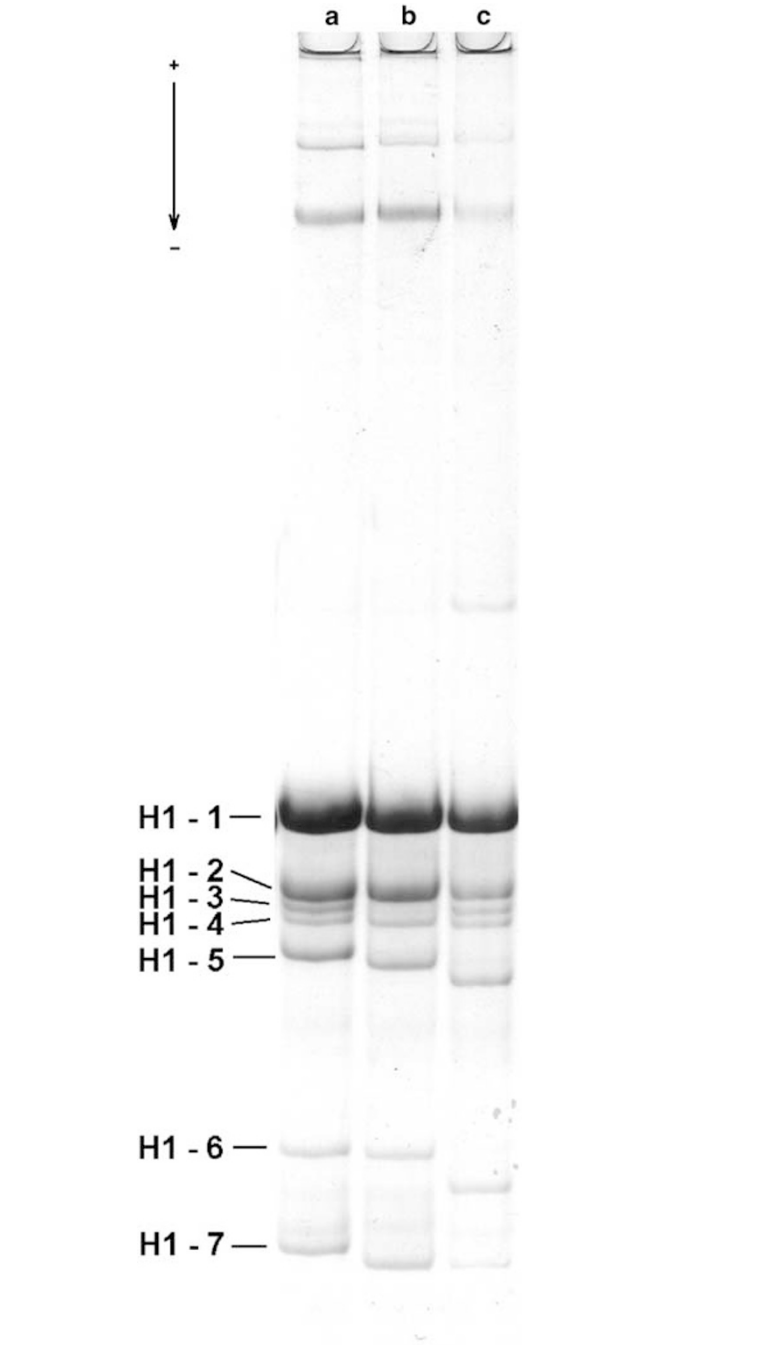
Figure 3 An electrophoregram of pea histone H1. Subtypes are numbered according to the increase of electrophoretic mobility. The following pea accessions and EMBL accession numbers (in parentheses) are presented: (**a**) – VIR-3971 (AJ635199); (**b**) – VIR-4362 (AJ543403); (**c**) – VIR-6560 (AJ635198).

expected to be paired over the C-domain. One KK doublet is close to the globular domain and four others are concentrated in the terminal half. Taking into account the presence of KR and RK pairs, this part of the molecule is enriched in doublets of positively charged amino acids. In the middle of the C-domain, lysine residues tend to be separated alternately by one residue or three residues.

```
H1-5-1  MATEEPIIAVEPVPEPSVLEKVEEPKAETEKTKKSKESKPKKASKPRNPASHPTYEEMIK
H1-5-2  MATEEPIIAVEPVPEPSVLEKVEEPKAETEKTKKSKESKPKKASKPRNPASHPTYEEMIK
H1-5-3  MATEEPIIAVEPVPEPSVLEKVEEPKAETEKTKKSKESKPKKASKPRNPASHPTYEEMIK
        ************************************************************

H1-5-1  DAIVSLKEKNGSSQYAIAKFIEEKQKQLPANFKKLLLQSLKKKVASGKLVKVKGSYKLSA
H1-5-2  DAIVSLKEKNGSSQYAIAKFIEEKQKQLPANFKKLLLQSLKKKVASGKLVKVKGSYKLSA
H1-5-3  DAIVSLKEKNGSSQYAIAKFIEEKQKQLPGNFKKLLLQSLKKKVASGKLVKVKGSYKLSA
        *****************************.******************************

H1-5-1  AAKKPTVAKPKTKPAAKAKAAVKAKTAAKPKAKAVVKPKSKSVTTKPKAAAAKPKAAANP
H1-5-2  AAKKPTVAKPKTKPAAKAKAAVKAKTAAKPKAKAVVKPKSKSVTTKPKAAAAKPKAAAKP
H1-5-3  AAKKPTVAKPKTKPAAKAKAAVKAKTAAKPKAKAVVKPKSKSVTTKPKAAAAKPKAAAKP
        *********************************************************:.*

H1-5-1  KAVAKAKAAAKPKPKAKSTKVARTSTKTTPGKKVAVAKPAPKKAAAAKKAPVKSVKAKSV
H1-5-2  KAVAKAKAAAKPKPKAKSTKVARTSTKTTPGKKVAVAKPAPKKAAAAKKAPVKSVKAKSV
H1-5-3  KAVAKAKAAAKPKPKAKSTKVARTSTKTTPRKKVAVAKPAPKKAAAAKKAPVKSVK----
        ******************************  ************************

H1-5-1  KSPAKKATGVKRGGRK
H1-5-2  KSPAKKATGVKRGGRK
H1-5-3  -SPAKKATGVKRGGRK
        ***************
```

Figure 4 An alignment of three allelic variants of subtype H1-5. Sequence identities are shaded in black.

The alanine residues of the C-domain tend to be distributed in blocks. Although the alanine content is only slightly lower than that of lysine, only 17 out of 40 alanines occur singly. Proline residues are often found between the lysines. Of 10 prolines, eight are present as a member of the KPK tripeptide.

These biased patterns of amino-acid distribution are most conspicuous in the middle part of the C-domain. It harbours the AAAKPK hexapeptide motif tandemly repeated four times: AAAKPK AAAKPK AVAKAK AAAKPK. The deviating hexapeptide AVAKAK differs from the adjacent AAAKPK by only two synonymous nucleotide substitutions, which suggests that the four repeats have a common origin. The slow allelic variant H1-5-1 differs from the other two by one amino-acid replacement, which gives rise to the hexapeptide AAANPK instead of the second AAAKPK.

The last third of the C-domain contains a 10-amino-acid-sequence VKSVK AKSVK consisting of two pentapeptides that differ by only one amino acid. The second pentapeptide AKSVK is deleted in the fast allelic variant H1-5-3.

Discussion

Divergence of pea H1 subtypes
This study characterises the second pea histone H1 gene that can be related to a specific protein electrophoretic band and identified as *His5*. The previously characterised *His1* gene maps to another chromosome, while *His5* is tightly linked to the genes coding for other subtypes. We are now in a position to consider the remaining two DNA sequences available that are presumed to encode H1 histones, H1.b (Gantt and Key, 1987) and H1.41 (Woo *et al*, 1995).

The electrophoretic mobility of the predicted protein of H1.41 is 1.58 (Table 1). This corresponds well to H1-7 encoded by *His7*, which is on the same chromosome as *His5* at a distance of 30 cM (Kosterin *et al*, 1994). The predicted mobility of H1.b, the sequence displaying closest homology with H1-5, is 1.13, which can be equally attributed to H1-3 or H1-4 encoded by *His3* and *His4*, respectively (Table 1). These two genes map to the gene cluster *His(2–6)* containing *His5* (Trusov *et al*, 1994).

Considering the tight linkage between *His3* and *His5*, it is of interest to estimate the time of evolutionary divergence between H1-5 and H1.b. Since the rate of nucleotide substitutions in synonymous sites is nearly constant (Kimura, 1983), the synonymous differences between the two sequences reflect the time that passed after their separation from a common ancestor. The synonymous distance, d_S, between the sequences coding for the globular domains of H1-5 and H1.b is 0.187 ± 0.067. If we accept the mean rate of synonymous substitutions to be 6.96×10^{-9} per site per year as estimated for actin genes of dicot plants (Moniz de Sa and Drouin, 1996), the gene duplication with following divergence of subtypes H1-5 and H1.b occurred 13.4 ± 4.8 million years ago.

H1.b and H1-5 are highly homologous (Figure 1), even in the most variable C-domain. The C-domains of H1-5 and H1.b are similar not only in their length (137 and 138 amino-acids) but also in their amino-acid content. This applies particularly for lysine, proline, serine and threonine. The distribution of lysine residues is highly conserved. Lysine doublets are concentrated in the

terminal part of C-domain; in the central part the alternating pattern xxxKxK is more common, with the short spacer often proline. A similar pattern of lysine distribution was observed in the H1-1 subtype as a result of tandem repeats of the motif PAAKAK (Berdnikov et al, 2003).

At the same time, subtypes H1.b and H1-5 have differences in their structure, that might reflect their functional divergence. First of all, the N-domains contain a large deletion/insertion of a decapeptide that is about 15% of the domain. The C-domains of both subtypes are about the same length, however, they differ by 31 amino-acid replacements (23.7%) and 13 residues covered by indels. Their amino-acid content is similar, except that the C-domain of H1.b contains much more valine (16.1%) than that of H1-5 (9.4%).

Divergence of *His5* alleles

Previously, we have analysed allelic variants of the subtype H1-1 and found that they differed by the number of repeats of the AKPAAK motif in the C-domain (Berdnikov et al, 2003). The most frequently occurring variant has 12 repeats, the slow variant – 13, and the fast variant – 11 repeats. The repeats occur in an extended zone comprising half of the C-domain. In lentil, where allelic polymorphism of H1-1 was also associated with changes in the number of repeated AKPAAK units, we found a very rare fast variant in which two units were lost. In the corresponding part of the C-domain in H1-5, there is a repeated zone of four AAAKPK motifs. However, in pea germplasm, we did not find variants that differed in the number of these repeats.

The main difference in the fast H1-5 variant from the two other variants is the deletion of the pentapeptide AKSVK situated adjacent to a similar pentapeptide VKSVK (Figure 4). The corresponding DNA sequences differ by three nucleotide substitutions. This pair of pentapeptides seems to be an ancient duplication. Strikingly, both pentapeptides are present in the same place in H1.b. That is, the duplication occurred before the separation of subtypes H1-5 and H1.b. It suggests some important function performed by these repeats, however, the loss of one repeated unit in the fast variant H1-5-3 does not appear to cause a drastic reduction of fitness. This allele is frequent in Central Asia, occurring in pea landraces that are considered to belong to the subspecies *Pisum sativum asiaticum* (Berdnikov et al, 1993b).

The slow variant H1-5-1 arose from the standard one due to a single nucleotide substitution resulting in amino-acid replacement Lys→Asn. This substitution occurred in the region of the C-tail harbouring four repeats of the AAAKPK hexapeptide motif. Repetition of this hexapeptide gives rise to alternating KPK and AAA tripeptides. The former carries a strong positive charge, and the latter may be considered as a neutral spacer. The KPK tripeptide is typically present in the C-terminal domains of histone H1 (Ponte et al, 2003).

In spite of an approximately equal percentage of alanine (29%) and lysine (32%) in the C-terminal domain of H1-5, eight prolines occur between lysine residues and none between alanine residues. It is natural to suppose that KPK plays an important role in interaction with linker DNA. Replacement of a positively charged lysine in a region saturated with AAAKPK motifs by a neutral

asparagine might change substantially the strength of histone–DNA interaction. A single amino-acid replacement in a functionally important motif TPKK→EPKK affected dramatically the binding ability of human H1.1 (Hendzel et al, 2004).

Earlier, we revealed a strong negative correlation between H1-5-1 frequency and the sum of vegetational temperatures (Berdnikov et al, 1993b). While comparing near-isogenic lines, we registered a slight increase in the total seed number and the mean number of seeds per pod in the line carrying this allelic variant (Bogdanova et al, 1994). However, the precise mechanism relating the charge of the histone H1 molecule and adaptation to low temperature is obscure.

References

Belyaev AI, Berdnikov VA (1985). Intraspecific polymorphism of H1 histone of the wild Fabaceae species. *Genetika (USSR)* **21**: 605–613 (in Russian).
Berdnikov VA, Gorel FL (1975). A study of ratios between histone fractions. *Mol Biol (USSR)* **9**: 699–705 (in Russian).
Berdnikov VA, Bogdanova VS, Gorel FL, Rozov SM (1992). Territorial distribution of histone H1 alleles in a population of *Vicia unijuga* A. Br. formed after urbanization of natural habitat. *Can J Bot* **70**: 1591–1595.
Berdnikov VA, Rozov SM, Temnykh SV, Gorel FL, Kosterin OE (1993a). Adaptive nature of interspecies variation of histone H1 in insects. *J Mol Evol* **36**: 497–507.
Berdnikov VA, Bogdanova VS, Rozov SM, Kosterin OE (1993b). The geographic patterns of histone H1 allelic frequencies formed in the course of pea (*Pisum sativum* L.) cultivation. *Heredity* **71**: 199–209.
Berdnikov VA, Gorel FL, Bogdanova VS, Kosterin OE, Trusov YA, Rozov SM (1999). Effect of a substitution of a short chromosome segment carrying a histone H1 locus on expression of the homeotic gene *Tl* in heterozygote in the garden pea *Pisum sativum* L. *Genet Res* **73**: 93–109.
Berdnikov VA, Bogdanova VS, Gorel FL, Kosterin OE, Trusov YA (2003). Large changes in the structure of the major histone H1 subtype result in small effects on quantitative traits in legumes. *Genetica* **119**: 167–182.
Bogdanova VS, Rozov SM, Trusov YA, Berdnikov VA (1994). Phenotypic effect of substitutions of short chromosomal segments containing different alleles of histone H1 genes in garden pea (*Pisum sativum* L.). *Genet Res* **64**: 35–41.
De Lucia F, Faraone-Mennella MR, D'Erme M, Quesada P, Caiafa P, Farina B (1994). Histone-induced condensation of rat testis chromatin: testis-specific H1t versus somatic H1 variants. *Biochem Biophys Res Commun* **198**: 32–39.
Ellis THN (1994). Approaches to the genetic mapping of pea. In: Linskens H-F, Jackson JF (eds) *Modern Methods of Plant Analysis. Vegetables and Vegetable Products*. Springer-Verlag: Berlin Vol 16, pp 117–160.
Gantt JS, Key JL (1987). Molecular cloning of a pea H1 histone cDNA. *Eur J Biochem* **166**: 119–125.
Hendzel MJ, Lever MA, Crawford E, Th'ng JPH (2004). The C-terminal domain is the primary determinant of histone H1 binding to chromatin *in vivo*. *J Biol Chem* **279**: 20028–20034.
Johns EW (1964). Studies of histones. 7. Preparative methods for histone fractions from calf thymus. *Biochem J* **92**: 55–59.

Kimura M (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press: Cambridge.

Kosterin OE, Bogdanova VS, Gorel FL, Rozov SM, Trusov YA, Berdnikov VA (1994). Histone H1 of the garden pea (*Pisum sativum* L.); composition, developmental changes, allelic polymorphism and inheritance. *Plant Sci* **101**: 189–202.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001). *MEGA2: Molecular Evolutionary Genetics Analysis Software*. Arizona State University: Tempe, AZ, USA.

Liao LW, Cole RD (1981). Differences among subfractions of H1 histone in their interactions with linear and superhelical DNA. Circular dichroism. *J Biol Chem* **256**: 6751–6755.

Moniz de Sa M, Drouin G (1996). Phylogeny and substitution rates of angiosperm actin genes. *Mol Biol Evol* **13**: 1198–1212.

Panyim S, Chalkley R (1969). High resolution acrylamide gel electrophoresis of histones. *Arch Biochem Biophys* **130**: 337–346.

Ponte I, Vila R, Suau P (2003). Sequence complexity of histone H1 subtypes. *Mol Biol Evol* **20**: 371–380.

Rozov SM, Bogdanova VS, Berdnikov VA (1986). Different chromosomal localizations of genes coding for *Pisum* histone H1 fractions. *Genetika (USSR)* **22**: 2159–2166 (in Russian).

Schulze E, Trieschmann L, Schulze B, Schmidt ER, Pitzel S, Zechel K *et al* (1993). Structural and functional differences between histone H1 sequence variants with differential intranuclear distribution. *Proc Natl Acad Sci USA* **90**: 2 481–2485.

Stein GS, Stein JL, Marzluff WF (eds) (1984). *Histone genes: Structure, Organization, and Regulation*. John Wiley & Sons: New York.

Tanaka I, Akahori Y, Gomi K, Suzuki T, Ueda K (1999). A novel histone variant localized in nucleoli of higher plant cells. *Chromosoma* **108**: 190–199.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl Acids Res* **24**: 4876–4882.

Trieschmann L, Schulze E, Schulze B, Grossbach U (1997). The histone H1 genes of the dipteran insect, *Chironomus thummi*, fall under two divergent classes and encode proteins with distinct intranuclear distribution and potentially different functions. *Eur J Biochem* **250**: 184–196.

Trusov YA, Bogdanova VS, Berdnikov VA (1994). Recombination within the complex locus His(2-6) containing genes for five histone H1 subtypes in pea. *Pisum Genet* **26**: 34–35.

van Holde KE (1988). *Chromatin*. Springer Verlag: New York.

Wolffe AP, Khochbin S, Dimitrov S (1997). What do linker histones do in chromatin? *BioEssays* **19**: 249–255.

Woo HH, Brigham LA, Hawes MC (1995). Molecular cloning and expression of mRNAs encoding H1 histone and an H1 histone-like sequences in root tips of pea (*Pisum sativum* L.). *Plant Mol Biol* **28**: 1143–1147.