Identification and characterization of the *Drosophila* histone H4 replacement gene

Anna Akhmanova, Koos Miedema, Wolfgang Hennig*

Department of Molecular and Developmental Genetics, Faculty of Sciences, Catholic University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands

Received 2 May 1996

Abstract Replacement variant genes for different histones have been described in most higher eukaryotes. However, so far no such gene has been found for histone H4. We have isolated from both *Drosophila melanogaster* and *D. hydei* a novel histone H4 encoding gene, H4r, which displays all the properties of a histone replacement variant gene: it contains introns, generates polyadenylated mRNA, represents the predominant H4 transcript in non-dividing tissues and is present in the genome as a single copy. The encoded polypeptide is identical to the *Drosophila* cell-cycle regulated histone H4. The fact that it is a single copy gene makes it prone to genetic analysis and it might be a useful tool for studying nucleosome structure and function.

Key words: Histone H4; Histone replacement gene; Evolution; *Drosophila melanogaster*; *Drosophila hydei*

1. Introduction

Histones in higher eukaryotes are encoded by two types of genes. Genes for replication-dependent histones are active predominantly in the S phase of the cell cycle, they are usually present in multiple copies, contain no introns and their transcripts are not polyadenylated [1]. Replacement histone genes are usually single copy, contain introns and produce polyadenylated mRNAs [1]. All histone replacement genes characterized so far encode proteins which are at least in a few positions different from their replication-dependent counterparts. Although cell-cycle regulated genes, encoding all five histones (H1, H2A, H2B, H3 and H4), have been isolated from many organisms, in no species has a full complement of replacement genes been found. For example, replacement variants of histone H3 were described in different animal species, plants and protozoa [2]. Since newly synthesized histones H3 and H4 form a complex before they are assembled into chromatin [3], one would expect to find histone H4 replacement genes as well. Until now no such gene has been described.

In this study, we report the identification and characterization of histone H4 replacement genes from *Drosophila melanogaster* and *D. hydei*. The implications of our finding for understanding of histone gene evolution are discussed.

2. Materials and methods

2.1. Reverse transcriptase-polymerase chain reaction (RT-PCR)

Total RNA was isolated from flies of both *Drosophila* species by the guanidinium-thiocyanate method [4] and reverse transcribed in the presence of oligo(dT) with the aid of Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase (Gibco BRL). PCR was carried out on the resulting cDNA with oligo(dT)-adapter primer with the sequence 5′-GCGGATCCGAATTCATCG$(T)_{17}$ and a degenerate histone H4 primer mh4-I with the sequence 5′-AAAGATCTAYGAR-GARACNCGNGGNGT (see Fig. 1A). For the amplification of the 5′ part of the *D. melanogaster* cDNA 10 μg of total RNA were reverse transcribed in the presence of 0.5 pmol of the oligonucleotide m4-I with the sequence 5′-TTTGATCAAGTATTTTGAATGTAGT-TAAG. The resulting RT mix was diluted 1:50 and oligo(dC) tails were added with the terminal transferase (Boehringer Mannheim) in the presence of 1 mM dCTP. PCR was performed with oligo(dG)-adapter primer with the sequence 5′-AGCTCAGAGCGGGCC-GCAAGCTT$(G)_{12}$ and primer m4-II 5′-CAAGATCTTGCGCT-TGGCGTGCTCG. For the amplification of the 5′ part of the *D. hydei* cDNA, a pool of total cDNA was C-tailed and PCR was performed with oligo(dG)-adapter primer and primer h4-I with the sequence 5′-GAGAATTCTTGGCTGCTGCTTCAG. PCR products were subcloned and several independent subclones were sequenced.

2.2. Screening of the genomic library

A *D. melanogaster* cDNA fragment, obtained by PCR with oligo(dT) and primer mh4-I was labeled with [α-^{32}P]dCTP by PCR with the same primers and used as a probe to screen the cosmid genomic library of *D. melanogaster* Canton S strain in Lawrist4 [5]. Two high-density filters with arrayed cosmid clones of this library were kindy provided by Dr. J. Hoheisel. The filters were hybridised in 0.5 M Na-phosphate buffer (pH 7.2), 7% SDS, 1 mM EDTA and 1% BSA at 60°C and washed in 0.1 M Na-phosphate buffer, 0.1% SDS at 25°C. After exposure to the X-ray film, the filters were stripped and rescreened with a *D. melanogaster* histone H2B probe, derived from the cluster of the cell-cycle regulated histone genes ([6], positions 677–968). Clones which hybridized to the histone H4, but not the H2B probe, were analysed further.

2.3. DNA sequencing and sequence analysis

DNA sequence was determined on both strands with the help of the ALFexpress automatic sequencer (Pharmacia). Sequence analysis was performed using the programs of the GCG Sequence Analysis package [7]. The number of synonymous substitutions per site (K_s) was calculated according to Nei and Gojobory [8]. Standard errors (S.E.) were calculated according to Nei and Jin [9].

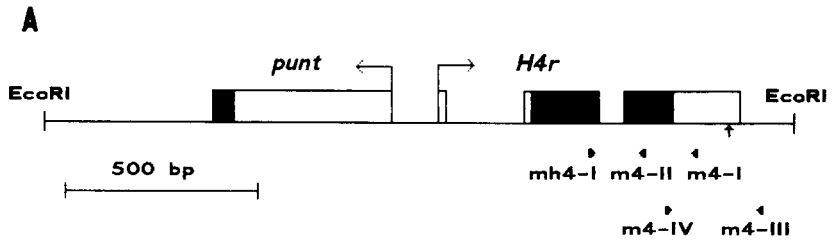
2.4. Northern blotting

Total RNA (10 μg per lane) was separated on 1.2% agarose-formaldehyde gel and transferred to Hybond nylon membrane (Amersham). Hybridization was performed under the same conditions as described for library screening and the filters were washed in 0.3 M Na-phosphate buffer, 0.1% SDS at 60°C. All probes were labeled by PCR. The *D. melanogaster* H4r 3′UTR probe was amplified with primers m4-III 5′-GAGGATCCTAAGCGAGTTTGTTGGACCT and m4-IV 5′-TTACGGATTTGGCGGTTAAG (Fig. 1A).

*Corresponding author. Fax: (31) (24) 3653390.
E-mail: kunmy1@caos.kun.nl

The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank and DDBJ nucleotide sequence databases with the following accession numbers: *D. melanogaster* mRNA for histone H4r — X97438, *D. melanogaster* histone H4r gene — X97437, *D. hydei* mRNA for histone H4r — X97436.

PII S0014-5793(96)00551-0



B

```
          punt gene            <
-148 AAATCTTTTTTGTCGCAACGGGGCGGCGATAgcgatggccgataggcgataactgaccagggctggtagtgagacaatag
                                                                  > H4r gene
 -68 aaatactagatacgcgcattttgcgaaaggtattttgaggtgtgcgagatctggtcacgctcacaagcAATTTTCTGGAG
  13 ATTTCTTGAAgtgagttttgttgctgagaaatttaagtaaataatagaagaatcggtcaacaaagttttttttgtgcataa
  93 acataaaaaaaaataattggagaccattattttgtatgaaaacagacttattaacaacagttttcacctgtcaaatgaacg
 173 tttacctttcaccaaacatggcgtcgctttacggtttattgtgtcttctcttcttttcagCCAACTTTACGAAGCAACTG

       M  T  G  R  G  K  G  G  K  G  L  G  K  G  G  A  K  R  H  R  K  V  L  R  D  N   26
 253 AGAAATGACTGGTCGTGGAAAGGGTGGCAAAGGATTGGGAAAGGGGGGCGCCAAGCGTCATCGTAAGGTGCTTCGTGATA

       I  Q  G  I  T  K  P  A  I  R  R  L  A  R  R  G  G  V  K  R  I  S  G  L  I  Y    52
 333 ACATCCAGGGTATCACCAAGCCTGCTATTCGCCGTTTGGCTCGTCGCGGCGGTGTTAAGCGTATCTCTGGCTTGATTTAC

       E  E  T  R  G  V  L  K                                                          60
 413 GAGGAAACTCGCGGTGTGCTAAAGgtacgttctggcgaaacgaagggaaagtggtgaaaaactgattaatttctcggaaa

             V  F  L  E  N  V  I  R  D  A  V  T  Y  T  E  H  A  K  R  K  T  V  T  A    84
 493 attcacagGTATTCCTTGAGAACGTTATCCGTGACGCTGTCACCTACACCGAGCACGCCAAGCGCAAGACCGTGACCGCC

       M  D  V  V  Y  A  L  K  R  Q  G  R  T  L  Y  G  F  G  G  *                   103
 573 ATGGACGTGGTCTATGCCCTCAAGCGCCAGGGACGCACCCTTTACGGATTTGGCGGTTAAGCACGCCCTCTAACCTCCAC
 653 TTGATAAATAACCATGGAGTTAACTTAACTACATTCAAAATACTTGATCCAATTGAATCTTAATCAGTAGTATTCAACGC
 733 AAACTTAACCGAACGCAAATATTTTATAATTTCACAATATACGGACTAAATAAATCAAAAGATATCACATAGAGAACTCA
 813 ATATCATGCATATGTGCAttttcacttggggcggttaatcatccgaaaataaggtccaacaaactcgcttagggagcaca
```

Fig. 1. (**A**) Structure of the *D. melanogaster* H4r gene. Transcription start sites are indicated by horizontal arrows. Black boxes represent translated parts and open boxes represent untranslated parts of the exons of the H4r and the *punt* gene. The vertical arrow indicates the polyadenylation signal in the H4r gene. Horizontal arrowheads indicate the positions of the oligonucleotides used for PCR. (**B**) Nucleotide sequence and the encoded amino acid sequence of the *D. melanogaster* histone H4r gene. The nucleotides of the *D. melanogaster* sequence are numbered at the left starting from the transcription initiation site, which is indicated in bold and italic. Amino acids are numbered at the right and the stop codon is indicated by an asterisk. Introns and the intergenic region between the H4r and the *punt* genes are shown in lower case letters. The polyadenylation signal is underlined and the experimentally determined polyadenylation sites are shown in bold.

3. Results and discussion

3.1. Isolation of the H4r cDNA and genomic clones

We have undertaken a systematic search for histone H4 replacement variants in *Drosophila*. It was assumed that, due to the extremely high evolutionary conservation of histone H4, a replacement gene would encode a protein quite similar to cell-cycle regulated H4 and that its mRNA would be polyadenylated. A degenerate oligonucleotide primer was designed from the histone H4 sequence and used in combination with oligo(dT) for RT-PCR on total RNA from flies. One major fragment was obtained. The 5′ part of the fragment encoded a polypeptide, identical to the C-terminus of the *Drosophila* histone H4. The 3′ part of the fragment was novel in its sequence and had no similarity to the 3′UTR of the cell-cycle regulated histone H4. It contained no inverted repeat sequence, characteristic for the replication-dependent histones. Rather, it contained a polyadenylation signal near its 3′ end, as was expected for a replacement histone gene (Fig. 1). We called this gene histone H4r (H4 replacement gene). The sequence of this fragment was used to design two new primers to isolate the 5′ part of the cDNA using the rapid amplification of cDNA ends procedure [10]. One PCR product was obtained. It was cloned and several independent subclones were sequenced. The fragment overlapped with the 3′ cDNA fragment and was identical to it in the overlapping region. The 5′ ends of the different subclones were the same, so we concluded that the isolated cDNA was complete in its 5′ part. The same approach was used to isolate the H4r cDNA from *D. hydei*. The 5′UTR of this cDNA is probably not complete, because we used a total pool of oligo(dT)-primed *D. hydei* cDNA to isolate it.

The obtained H4r cDNA fragment was used to screen a cosmid genomic DNA library of *D. melanogaster*. The filters with arrays of cosmid clones, representing together four genomic equivalents, were hybridized to the probe under non-stringent conditions. Approximately 60 clones gave a positive signal. Most of these signals were the result of a cross-reaction of our probe with the histone H4 gene of the histone repeat cluster (which is present in the genome in ~100 copies [11]). We assumed that histone H4r gene is localised outside of the histone cluster, similar to the replacement genes

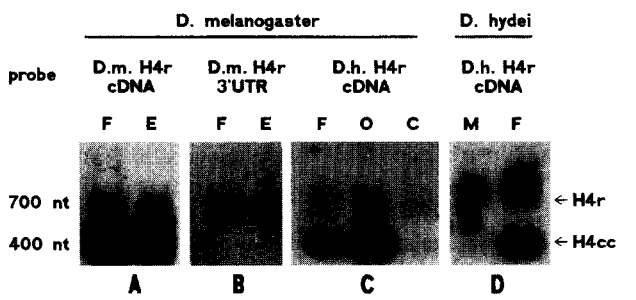


Fig. 2. Northern blot analysis of the H4r expression. Total RNA (10 μg) was prepared from: F — female flies, M — male flies, E — embryos, C — fly carcasses, O — ovaries. Length of the transcripts in nucleotides (nt) is indicated on the left.

H2AvD [12] and H3.3A and H3.3B [13]. To eliminate the cosmids containing the histone cluster copies, we rescreened the library with the cell-cycle regulated histone H2B probe and analysed those clones which did not react with it. Three of these clones, JHD1:72G2, JHD4:19G11 and JHD5:58C4, hybridized strongly to the probe. They proved to be overlapping cosmids, which contained the same 2 kb *Eco*RI fragment encompassing the entire H4r gene (Fig. 1). We also analysed several weaker hybridizing clones. In none of them was an H4-like gene found.

Taken together, the results of the PCR experiments on the cDNA pool and the genomic library screen suggest that the isolated gene, H4r, is the only replacement histone H4 gene in the *D. melanogaster* genome. Also genomic Southern blotting (results not shown) and Northern blotting (see below) provided no indication for the presence of additional copies of the H4-encoding replacement histone gene in this species. The fact that the H4r is a single copy gene makes it susceptible to genetic analysis. It provides a good opportunity to perform genetic experiments with a histone H4 in a higher eukaryote.

3.2. *Genomic organization of the* D. melanogaster *histone H4r*

Comparison of the cDNA and genomic sequence of the histone H4r showed that the gene contains two introns; one of ~200 bp interrupts the 5'UTR while the second one is located in the ORF of the gene (Fig. 1). This structure is quite similar to that of the *Drosophila* H3.3B genes except that the first intron of H4r is much shorter than that of histone H3.3B [13]. The small size of the H4r 3'UTR makes this gene more similar to the H3.3A than to the H3.3B gene.

Analysis of the sequence flanking the histone H4r gene showed that there is another gene, *punt*, which is localized immediately upstream from the H4r gene and is transcribed from the opposite DNA strand (Fig. 1A). *punt* encodes a type II *dpp* receptor and is expressed at all developmental stages [14,15]. The transcription start sites of the two genes are separated by only ~120 bp. This does not leave much space for the promoters of the two genes and suggests that they might be to a certain extent coregulated. Similarly close linkage to a housekeeping gene, encoding the oligosaccharyltransferase 48 kDa subunit, was also observed for the *D. melanogaster* histone H3.3B [13].

Analysis of the promoter region revealed no presence of a TATA-box around position −25, which is again similar to the promoter of H3.3B gene. The GAGA-factor binding sites [16], present in the *Drosophila* cell-cycle regulated H4 promoter, were not found either. Further, we could not identify any promoter elements, common for the *Drosophila* H2AvD, H3.3B and H4r genes. So even if these histone replacement genes are coregulated, it is not apparent from the comparison of their promoter sequences.

The H4r gene was mapped by hybridizing the 2 kb *Eco*RI genomic fragment to the polytene chromosomes (not shown) and was found in the locus 88C of chromosome 3. This is in agreement with the localization of the *punt* gene [14,15].

3.3. *Analysis of the histone H4r gene expression by Northern blotting*

Hybridization of the *D. melanogaster* H4r cDNA probe to Northern blots with total fly and embryonic RNA revealed two bands of ~400 nt and ~700 nt (Fig. 2A). The ~400 nt band corresponds to the cell-cycle regulated histone H4 (H4cc) mRNA, while the ~700 nt band corresponds to the H4r transcript. This was confirmed by probing the same blot with the H4r 3'UTR probe (Fig. 2B). The cell-cycle regulated H4 mRNA is more abundant both in embryos and in females. The latter is caused solely by the storage of the histone cell-cycle regulated messengers in the oocytes. To show that, we hybridized a Northern blot with *D. melanogaster* RNA from ovaries and fly carcasses (remainder of the fly after removing the ovaries) to the *D. hydei* H4r cDNA. This cDNA is almost equally diverged (~17%) from *D. melanogaster* H4 genes of both types in the ORF region and has no similarity to either of the genes in the untranslated regions. As one can see from Fig. 2C, H4cc mRNA is very abundant in the ovary, while the H4r is the predominant transcript in the fly carcass (somatic, mainly non-dividing tissues). This result is exactly as expected for a replacement histone and is very similar to that obtained for the H3.3 genes [13]. A single H4r transcript of ~800 bp was observed in *D. hydei* (Fig. 2D). This size is in agreement with a longer 3'UTR of the H4r cDNA in this species.

3.4. *Sequence comparison of the* Drosophila *histone H4 genes*

The *D. melanogaster* and *D. hydei* H4r cDNAs encode a protein, identical to the *Drosophila* cell-cycle regulated H4. This is the first case when a replication-dependent and a replacement histone gene of a particular species encode exactly the same polypeptide. The function of the replacement-type histone genes is not yet clear, but one can imagine two different possibilities. On the one hand, incorporation of the replacement variant histone proteins might be necessary to alter the structure of certain chromatin domains in the absence of replication. If this is true, one would expect to find functionally important differences between replacement and cell-cycle regulated histone proteins. On the other hand, it is possible that low-level synthesis of histones per se is necessary for chromatin maintenance in non-dividing cells. In the latter case the proteins encoded by the genes of the two types might be identical or differ by neutral amino acid substitutions. Our data on replacement gene of H4 support the second hypothesis. However, one should bear in mind that histones H3 and H4 are assembled into the chromatin as an $(H3–H4)_2$ tetramer. It is possible, therefore, that in order to incorporate into chromatin the replacement variant histone H3.3, replication-independent expression of histone H4 is required and that the H4r gene fulfils this function.

On the nucleotide level the four *Drosophila* H4 ORF sequences are quite divergent, because of the presence of multi-

ple synonymous substitutions. As expected, phylogenetic analysis by different methods [7] indicates that the cell-cycle regulated and replacement genes of two species are more related to each other than the two types H4 genes within one species. The number of synonymous substitutions per site between the H4r genes ($K_s = 1.30$, S.E. = 0.16) is somewhat higher than that between the cell-cycle regulated H4 genes ($K_s = 0.94$, S.E. = 0.10). This is in contrast to the values obtained for the *Drosophila* H3 histones, whose variants evolve slower than replication-dependent genes [13]. The base composition at synonymous positions and the codon usage bias are quite similar between the two types of H4 genes. This indicates that the type of genomic organization (single copy versus tandem repeats) and the mode of regulation do not have much influence on the evolution of these histone-encoding sequences in *Drosophila*.

The untranslated regions of the *D. melanogaster* and *D. hydei* cDNAs display very little similarity. The same lack of conservation was observed for the *Drosophila* H3.3 genes [13]. This indicates that the evolutionary constraints on these regions of histone replacement genes are quite low in *Drosophila*, in contrast to the vertebrate genes [17,18].

3.5. Evolution of the H4 replacement gene

Replacement variants of core histones are thought to be evolutionarily ancient. The histone H2A variant probably diverged from its cell-cycle regulated counterpart before the separation of fungi and ciliates from the rest of the eukaryotic phyla [19]. The histone H3.3 variant probably arose very early in animal evolution, before the split between invertebrates and vertebrates [2,13]. One would expect a similar evolutionary history for a histone H4 replacement gene.

To establish the evolutionary relationships between genes, comparison of the intron positions is very useful. Histone H4 genes were sequenced in many different species, but most of them do not contain introns. The exceptions are the H4 genes of certain fungi (from *Neurospora*, *Aspergillus*, *Physarum*). The intron positions in all these organisms are different from that of the *Drosophila* H4r. This correlates with their very early separation from the rest of the eukaryotes. Recently, two nematode intron-containing H4 sequences were deposited in the data bases. One of them, from *C. elegans* cosmid C50F4, which was sequenced as a part of the *C. elegans* Genome Sequencing Project [20], contains an intron at a position different from the H4r gene. The other, an incomplete genomic sequence from *Ascaris lumbricoides* (EMBL accession numbers Z69289 and Z69290, P. Duda et al., unpublished), contains an intron at a position identical to that of the *Drosophila* H4r. The H4 gene from *Ascaris* displays other properties of a replication-independent histone gene — the absence of the 3′ terminal palindrome and the presence of the polyadenylation signal. Assuming that this gene is expressed, one can conclude that it is the nematode homologue of the *Drosophila* H4r gene. This indicates that the histone H4 replacement and cell-cycle regulated genes separated from each other early in animal evolution. It is noteworthy that the polypeptide, encoded by the *Ascaris* gene, is at four positions different from the *Drosophila* sequence, while the two *Drosophila* histone H4 sequences are identical. However, judging from the intron position and the 3′ regulatory sequences, the replacement H4 genes from *Ascaris* and *Drosophila* are evolutionarily closer to each other than the replication-dependent and the replacement H4 genes of *Drosophila*. This would mean that histone H4 genes undergo coevolution within a species and that establishing the phylogenetic relationships between core histone genes based on protein alignments may be difficult. This conclusion is strengthened by the fact that the intron-containing H4 gene from the *C. elegans* cosmid C50F4 also encodes a protein identical to the cell-cycle regulated H4 of this species. The fact that its intron is positioned differently from that in *Drosophila* and *Ascaris* H4 genes complicates the picture of the evolution of the replacement histone H4. Characterization of the replacement histone H4 genes from other higher eukaryotes will hopefully clarify this picture.

Acknowledgements: We are very grateful to Dr. J. Hoheisel for providing the cosmid library filters and the clones, to Denise Vlooswijk for her contribution to the analysis of the genomic clones and to Dr. G. Korge for help with the polytene chromosome hybridization. We thank Petra Bindels, Wenli Gu and Harry Harhangi for help in conducting the experiments and for valuable discussions.

References

[1] Schümperli, D. (1986) Cell 45, 471–472.
[2] Thatcher, T.H., MacGaffey, J., Bowen, J., Horowitz, S., Shapiro, D.L. and Gorovsky, M.A. (1994) Nucleic Acids Res. 22, 180–186.
[3] Perry, C.A., Dadd, C.A., Allis, C.D. and Annunziato, A.T. (1993) Biochemistry 32, 13605–13614.
[4] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[5] Hoheisel, J.D., Lennon, G.G., Zehetner, G. and Lehrach, H. (1991) J. Mol. Biol. 220, 903–914.
[6] Matsuo, Y. and Yamazaki, T. (1989) Nucleic Acids Res. 17, 225–238.
[7] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[8] Nei, M. and Gojobory, T. (1986) Mol. Biol. Evol. 3, 418–426.
[9] Nei, M. and Jin, L. (1989) Mol. Biol. Evol. 6, 290–300.
[10] Frohman, M.A. (1990) in: PCR Protocols: a Guide to Methods and Applications (Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J., eds) Academic Press, San Diego, pp. 28–38.
[11] Lifton, R.P., Goldberg, M.L., Karp, R.W. and Hogness, D.S. (1977) Cold Spring Harbor Symp. Quant. Biol. 42, 1047–1051.
[12] van Daal, A., White, E.M., Gorovsky, M.A. and Elgin, S.C. (1988) Nucleic Acids Res. 16, 7487–7497.
[13] Akhmanova, A.S., Bindels, P.C.T., Xu, J., Miedema, K., Kremer, H. and Hennig, W. (1995) Genome 38, 586–600.
[14] Childs, S.R., Wrana, J.L., Arora, K., Attisano, L., O'Connor, M.B. and Massague, J. (1993) Proc. Natl. Acad. Sci. USA 90, 9475–9479.
[15] Ruberte, E., Marty, T., Nellen, D., Affolter, M. and Basler, K. (1995) Cell 80, 889–897.
[16] Gilmour, D.S., Thomas, G.H. and Elgin, S.C.R. (1989) Science 245, 1487–1490.
[17] Wells, D., Hoffman, D. and Kedes, L. (1987) Nucleic Acids Res. 15, 2871–2889.
[18] Chalmers, M. and Wells, D. (1990) Nucleic Acids Res. 18, 3075.
[19] van Daal, A., White, E.M., Elgin, S.C.R. and Gorovsky M.A. (1990) J. Mol. Evol. 30, 449–455.
[20] Wilson, R., Ainscough, R., Anderson, K., Berks, C., Bonfield, J., Barton, J., Connel, M., Copsey, T., Cooper, J. et al. (1994) Nature 368, 32–38.