Volume 13 Number 24 1985 Nucleic Acids Research

# A processed pseudogene in an intron of the HLA-DPβ1 chain gene is a member of the ribosomal protein L32 gene family

John A.T.Young and John Trowsdale

Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK

Received 24 October 1985; Accepted 21 November 1985

## ABSTRACT

A sequence in an intron of the human HLA-DPβ1 gene was identified by its homology to the gene encoding ribosomal protein L32 (rpL32). It lacked introns indicating that it was derived from a processed rpL32 mRNA transcript. A human cDNA clone encoding rpL32 was isolated and compared to this human pseudogene and to several related mouse sequences, one of which is contained in an intron of the murine dihydrofolate reductase gene. Comparison of these sequences revealed that they were more related within species than between, suggesting that they became inserted in the genome after man and mouse diverged.

## INTRODUCTION

There are several examples of processed pseudogenes derived from genes transcribed by RNA polymerase II. These include those derived from so called "housekeeping" genes which are presumably expressed in the germ cell, ie dihydrofolate reductase (1,2), cytochrome C (3), β-tubulin (4), α-tubulin (5), metallothionein (6), and ribosomal protein genes L30 (7), L7 (8), L18 (9), and L32 (10). Others appear to be derived from aberrant transcripts of genes normally expressed only in differentiated cell types, ie α-globin (11, 12) and immunoglobin λ light chain (13).

Several features of these elements indicate that processed mRNA molecules are involved in their formation. Namely, they lack the intervening sequences of the parent gene, and their homology to the parent gene is usually confined to the precise limits of the mature transcript. Presumably a reverse transcriptase activity acted upon these processed mRNA transcripts to produce complementary double stranded DNA molecules which then integrated into the genome at a site distinct from that occupied by the parent gene (13). These elements are generally flanked by direct repeat DNA sequences - a hallmark of some integrated DNA elements.

We previously described a processed pseudogene in an intron of the human

© IRL Press Limited, Oxford, England.

HLA-DPβ1 gene and showed that it was highly related to a human cDNA clone (14). However, the 5' end of this clone was completely unrelated to the corresponding region of the pseudogene, implying that it did not represent the progenitor of this sequence. To identify this pseudogene we isolated other related cDNA clones, and characterised one encoding ribosomal protein L32 (rpL32) which appears to represent the parent gene from which the pseudogene is derived.

## MATERIALS AND METHODS

DNA probes - 100 ng samples of DNA were nick translated in 5μg/ml BSA, 2μM dATP, dTTP, dGTP, 20μCi $\alpha^{32}$PdCTP (>3000 Ci/mM, Amersham), 50mM Tris.Cl pH7.5, 5mM MgCl2, and 1mM β-mercaptoethanol (15). 5 units of DNA polymerase I (Kornberg, Boehringer Mannheim) was added, incubated at 14°C for 75 mins, and labelled nucleic acid was separated from unincorporated nucleotides on Sepharose G50.

Other probes, subcloned into bacteriophage M13mp8 (BRL), were labelled by performing the primed sequencing reactions in the presence of 10μCi $\alpha^{32}$PdCTP (>3000Ci/mM, Amersham), and 400 μm dATP, dGTP, dTTP, in the absence of dideoxynucleotides. The labelled recombinant DNA was digested with the enzymes EcoRI, HindIII, and PstI, which all cut in the polylinker sequence flanking, but not within, the specific insert. The labelled insert was then separated from M13 DNA by electrophoresis on a 1.4% agarose gel, cut from the gel, homogenised in 10mM Tris pH7.4, 1mM EDTA, and denatured by boiling for 6 minutes before adding to hybridisation buffer.

DNA sequencing - subclones of DNA fragments were inserted into bacteriophage M13mp8 and were sequenced in both directions by the chain termination method (16, 17).

Southern blotting - DNA blots were prepared as described by Southern (18). Hybridisation was performed at 68°C for 16 hours in 6XSSC, 10X Denhardts solution, 10% Dextran Sulphate, 0.1% SDS, and 100μg/ml denatured salmon sperm DNA, using a probe concentration of 1-5 ng/ml. Filters were washed 3 x 45 minutes in 0.1 x SSC, 0.1% SDS at 68°C, and exposed to Kodak XAR-5 film for 6 days at -70°C.

Cell lines - Human Bu cells were a generous gift from Dr S Povey. Mouse 1R cells and mouse/human somatic cell hybrid HorlI cells have been previously described (19, 20).

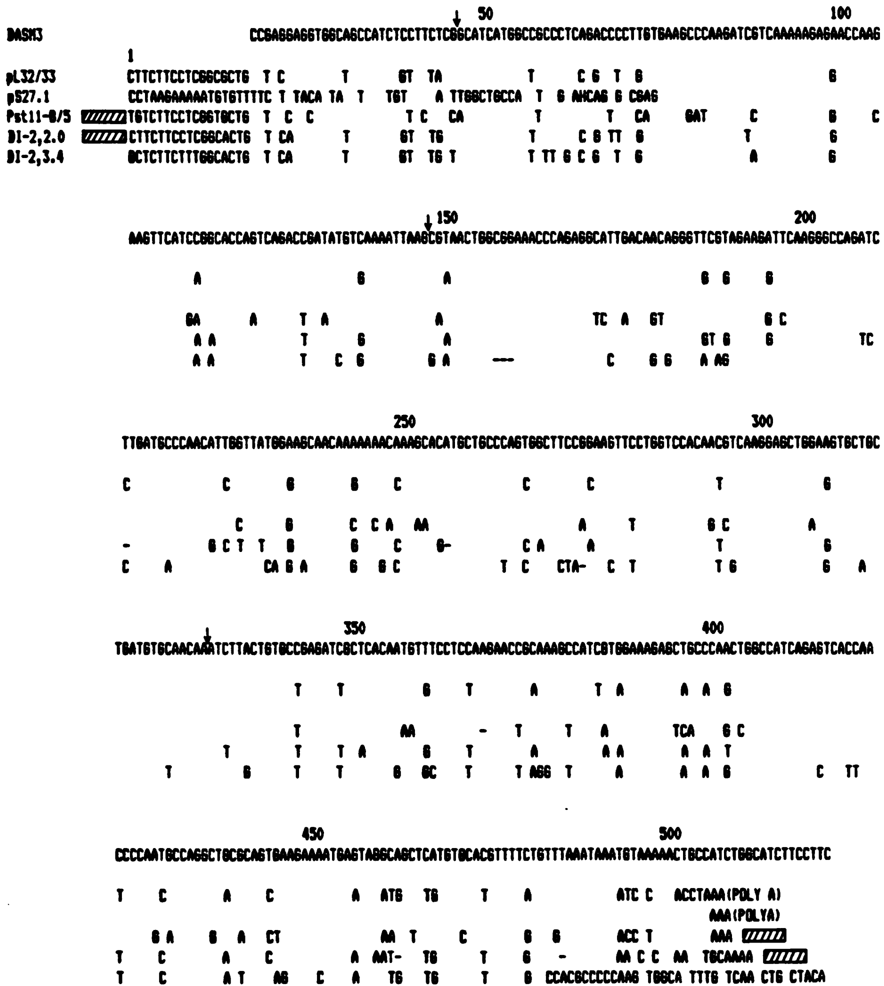


Figure 1
Comparison of human and mouse rpL32 gene family sequences. The sequence of the human cDNA clone described here, DASM3, is written in full and aligned with those of two published human sequences; cDNA clone pS27.1, and pseudogene Pstll-8/5 (14). The sequence of clone pS27.1 is truncated at the 5' end for the purpose of comparison. Three murine sequences; pseudogenes DI-2,2.0 and DI-2,3.4 (21) and rpL32 cDNA clone pL32/33 (10), are also aligned with the clone DASM3. Several small insertions have been omitted from the pseudogene sequences to allow optimal alignment. These pseudogene sequences are published in full elsewhere (14). Dashes in the pseudogene sequences represent bases deleted. Hatched boxes represent 17 base pair direct repeats which flank the elements Pst11-8/5 and DI-2,2.0. Gaps in each sequence indicate sequence identity with clone DASM3. Bases are numbered from the transcription start site (indicated by "1") of the murine rpL32 gene (10). Arrows represent positions of intron removal from the primary murine rpL32 transcript (10).

## RESULTS

### A human cDNA clone encoding ribosomal protein L32

We previously reported two related human sequences: a processed pseudogene (Pst 11-8/5) in an intron of a DPβ gene, and a cDNA clone called pS27.1 (14). Using nick translated pS27.1 DNA as a probe, further related cDNA clones were isolated from a Burkitt's lymphoma cell line Daudi cDNA library the longest of which, DASM3, was sequenced. The sequence of the clone DASM3 was identical to that of clone pS27.1 (bases 75-505, Figure 1), except at the 5' end (bases 18-74, Figure 1).

A search of the literature identified this family of sequences. Both human clones pS27.1 and DASM3 were highly related to a murine cDNA clone pL32/33 (10), which encodes rpL32 (Figure 1). The region where the two human cDNA clones pS27.1 and DASM3 were identical (bases 75-505, Figure 1) was 90% related to the corresponding region of the murine rpL32 cDNA clone pL32/33. However, the 5' end of the human clone DASM3 (bases 18-74, Figure 1) was much more highly related (78%) to the corresponding region of the murine clone pL32/33, than was the same region of the other human clone pS27.1 (32%). This suggested that the clone DASM3 is the human equivalent of the murine

```
rpL32                                                       Met Ala Ala Leu Arg Pro Leu Val Lys Pro Lys Ile Val Lys Lys Arg
DASM3                     CCGAGGAGGTGGCAGCCATCTCCTTCTCGGCATC ATG GCC GCC CTC AGA CCC CTT GTG AAG CCC AAG ATC GTC AAA AAG AGA
pL32/33    CTTCTTCCTCGGCGCTG--T-C--------T-------GT--TA------- --- --T --- --- C-G --T --G --- --- --- --- --- --- --- --- --G

Thr Lys Lys Phe Ile Arg His Gln Ser Asp Arg Tyr Val Lys Ile Lys Arg Asn Trp Arg Lys Pro Arg Gly Ile Asp Asn Arg Val Arg Arg
ACC AAG AAG TTC ATC CGG CAC CAG TCA GAC CGA TAT GTC AAA ATT AAG CGT AAC TGG CGG AAA CCC AGA GGC ATT GAC AAC AGG GTT CGT AGA
--- --- --- --- --- A-- --- --- --- --- --- --- --G --- --- --- --A --- --- --- --- --- --- --- --- --- --- --- --G --G ---

Arg Phe Lys Gly Gln Ile Leu Met Pro Asn Ile Gly Tyr Gly Ser Asn Lys Lys Thr Lys His Met Leu Pro Ser Gly Phe Arg Lys Phe Leu
AGA TTC AAG GGC CAG ATC TTG ATG CCC AAC ATT GGT TAT GGA AGC AAC AAA AAA ACA AAG CAC ATG CTG CCC AGT GGC TTC CGG AAG TTC CTG
--G --- --- --- --- --- C-- --- --- --- --C --- --- --G --- --- --G --- --C --- --- --- --- --- --C --- --- --C --- --- ---

Val His Asn Val Lys Glu Leu Glu Val Leu Leu Met Cys Asn Lys Ser Tyr Cys Ala Glu Ile Ala His Asn Val Ser Ser Lys Asn Arg Lys
GTC CAC AAC GTC AAG GAG CTG GAA GTG CTG CTG ATG TGC AAC AAA TCT TAC TGT GCC GAG ATC GCT CAC AAT GTT TCC TCC AAG AAC CGC AAA
--- --- --T --- --- --- --- --G --- --- --- --- --- --- --- --- --- --- --T --- --T --- --- --- --G --- --T --- --- --A ---

Ala Ile Val Glu Arg Ala Ala Gln Leu Ala Ile Arg Val Thr Asn Pro Asn Ala Arg Leu Arg Ser Glu Glu Asn Glu
GCC ATC GTG GAA AGA GCT GCC CAA CTG GCC ATC AGA GTC ACC AAC CCC AAT GCC AGG CTG CGC AGT GAA GAA AAT GAG TAG  GCAGCTCATGTGC
--- --T --A --- --- --A --A --G --- --- --- --- --- --- --T --- --C --- --- --A --- --C --- --- --- --A ---  ATG---TG-----

ACGTTTTCTGTTTAAATAAATGTAAAAACTGCCATCTGGCATCTTCCTTC
-T-----A-----------ATC-C---ACCTAAAAAAAAAAAAAAAAAA
```

Figure 2
Comparison of human and murine rpL32 cDNA clones. The nucleotide sequences of the human clone DASM3 and murine rpL32 cDNA clone pL32/33 are compared. Dashes indicate identity between the two sequences.

rpL32 cDNA clone pL32/33. Indeed, the human cDNA clone DASM3 encodes a polypeptide identical to murine rpL32 (Figure 2).

Comparing the translated regions of this human clone and the murine rpL32 cDNA clone pL32/33, there are 36 silent nucleotide substitutions: 33 are third base pair and 3 are first base pair changes (Figure 2).

Assuming that the human and murine rpL32 genes are similarly organised, the human rpL32 cDNA clone DASM3 is full length except for 17 base pairs of 5' terminal sequence (the transcription start site identified in the mouse (10) is indicated by "1" in Figure 1).

The other human clone, pS27.1 (14), appears to encode a 163 amino acid polypeptide which is highly related to rpL32 but with a different amino terminus (not shown).

rpL32 processed pseudogenes are found in introns of the human HLA-DPβ1 gene and murine dihydrofolate reductase gene

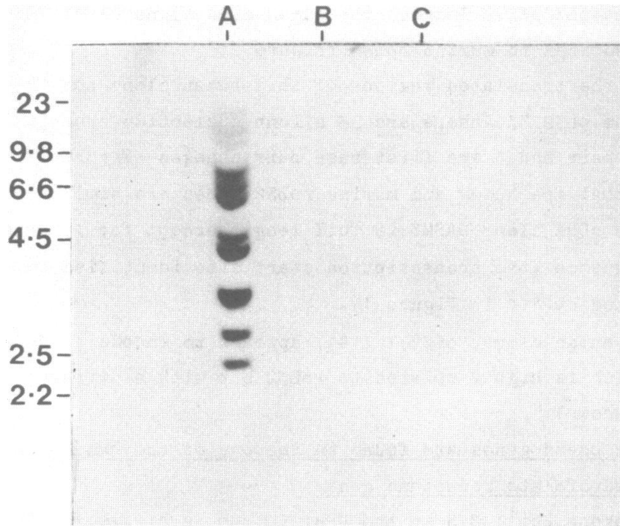
The pseudogene Pst11-8/5 in the DPβ1 intron is highly related to the human and murine rpL32 cDNA clones DASM3 and pL32/33, and bears features which suggest that it is derived from a correctly processed rpL32 mRNA transcript. For instance, the homology between this pseudogene and the rpL32 gene is confined to the precise limits of the mature rpL32 transcript, with the introns precisely removed [arrows represent positions of intron removal from the murine rpL32 primary transcript (10), Figure 1]. Furthermore, this pseudogene is immediately flanked by 17 base pair direct repeat DNA sequences (represented by hatched boxes, Figure 1), which are presumably generated from the integration target site (see Discussion).

Murine sequences DI-2,2.0 and DI-2,3.4 (21) have similar features and are presumably also rpL32 processed pseudogenes (Figure 1). The element DI-2,2.0 is contained in the second intervening sequence of the murine dihydrofolate reductase gene (21).

The rpL32 processed pseudogenes probably arose after the divergence of mouse and man

The human rpL32 processed pseudogene Pst 11-8/5 is more closely related to the human rpL32 cDNA clone DASM3 (86% homology) than it is to the corresponding murine clone pL32/33 (79% homology). Similarly, the murine rpL32 processed pseudogenes DI-2,2.0 and DI-2,3.4 are more closely related to the murine rpL32 clone (92% and 83% homology respectively) than they are to the human rpL32 clone (84% and 77% homology respectively). These pseudogenes are therefore more highly related to the rpL32 gene within their own species.

This was demonstrated directly by probing Southern blots of DNA from




<u>Figure 3</u>
<u>Southern blot of human, mouse and human/mouse somatic cell hybrid genomic DNA.</u>
DNA samples digested with EcoRI (10μg/track) were loaded onto a 0.8% agarose gel, and electrophorosed at 20V for 20 hours, then the DNA was denatured and transferred to a nitrocellulose filter. The filter was hybridised to a DASM3 specific probe prepared from an EcoRI/Sau3A fragment (bases 18-83, Figure 1), subcloned into bacteriophage M13mp8, as described in Materials and Methods. Lane 1, Bu human DNA; Lane 2, IR mouse DNA; Lane 3, HorlI DNA. HorlI is a human/mouse somatic cell hybrid which contains only human chromosome 15 (20).

both human and mouse origin using a probe specific for the human rpL32 cDNA clone DASM3. A specific EcoRI-Sau3A fragment of human clone DASM3 (base pairs 18-83, Figure 1) was subcloned into EcoRI-BamHI cut bacteriophage M13mp8 DNA, and used as a probe for hybridisation to a Southern blot of digested human and mouse genomic DNA. This radioactively labelled probe hybridised to multiple human bands (Figure 3, track A), the majority of which are presumably rpL32 processed pseudogenes. However, under the conditions of high stringency washing used, the same human probe did not hybridise to any of the murine rpL32 gene family (Figure 3, track B).

<u>The two human cDNA clones are encoded by unlinked genes</u>

As the human cDNA clones DASM3 and pS27.1 are identical except for their 5' ends they may have been derived from alternatively spliced transcripts of a single gene. Since we had previously shown that clone pS27.1 maps to human chromosome 15 (14), we tested whether clone DASM3 also maps to this chromosome.

Using the probe specific for the 5' end of clone DASM3 described above, we found that none of the multiple human rpL32 bands (Figure 3, track A) were present in the digested genomic DNA of a mouse/human somatic cell hybrid HorlI (Figure 3, track C), which contains only human chromosome 15 (20). The gene encoding rpL32 therefore does not appear to map to the same human chromosome as that respresented by clone pS27.1. In fact, other experiments indicate that rpL32 genes are dispersed into several different human chromosomes (N Spurr, personal communication).

## DISCUSSION

We have isolated and identified a human cDNA clone encoding ribosomal protein L32, and have shown that the element Pst 11-8/5 (14) in the intron of the human HLA-DPβ1 gene is a rpL32 processed pseudogene. To our knowledge this is one of the first examples of a sequence of this type to be found in an intron of an expressed gene.

Although there is 9% nucleotide sequence divergence (36 silent nucleotide substitutions) between the coding regions of the human rpL32 cDNA clone and its murine counterpart they encode identical proteins. This indicates that rpL32 plays a highly important role in the function of the eukaryotic ribosome.

The other human cDNA clone pS27.1 was homologous to clone DASM3 except at the extreme 5' end (bases 18-74, figure 1). This clone may encode a novel ribosomal protein although we cannot rule out the possibility that it may represent transcription of a pseudogene or an aberrant cDNA clone. This is currently being investigated.

The element Pst 11-8/5 (14) bears all the features of a processed pseudogene as it lacks intervening sequences and its homology to the parent gene is confined precisely to the limits of the mature rpL32 transcript. It is also flanked by 17 base pair direct repeat DNA sequences which presumably are derived from DNA at the target site for integration; the HLA-DPβ2 gene which is related to the DPβ1 gene by a relatively recent gene duplication event contains a single copy of this 17 base pair sequence and does not contain this pseudogene (14). These direct repeats therefore most likely formed by staggered cleavage at this 17bp sequence in the DPβ1 intron prior to integration of the pseudogene, followed by repair on either side of the integrated element.

It will be interesting to see whether the equivalents of the DPβ1 gene in other species, eg murine Aβ3 (22), contain this pseudogene, in order to

aid our understanding of the evolution of the major histocompatibility complex genes.

In view of their strong sequence homology to DASM3, the two published murine sequences are identified as rpL32 processed pseudogenes (21). All of these pseudogenes described most likely arose following the divergence of man and mouse as they are more highly related to the rpL32 gene within, than between species. Alternatively, they could have evolved before man and mouse diverged and subsequent gene conversion-like events with the rpL32 gene may have homogenised them independently in each species. Such events are proposed to operate in several eukaryotic multi-gene families including globin (23), immunoglobin (24), MHC classI (25, 26, 27) and MHC classII (28, 29, J Young, unpublished observations).

It is intriguing that several of these pseudogenes are found in introns of other genes since such elements might be expected to integrate preferably into regions of the genome which are most accessible in the germ cell, eg those being transcribed and replicated. As the murine DHFR gene is transcribed in the germ cell, it is possible that a rpL32 processed pseudogene integrated into its second intron while this region was open for transcription (21). However, there is no evidence to suggest that the HLA-DPβ1 gene is expressed in he germ cell. It is therefore tempting to speculate that the pseudogene in the DPβ1 intron became integrated when this region of the DNA was accessible in the absence of transcription eg during a gene conversion event. If so, we anticipate that other such elements will be found not only in and around those genes which are transcribed in the germ cell, but also in other genes which participate in genetic events associated with the generation of the extremely high degree of polymorphism which is characteristic of HLA genes.

ACKNOWLEDGEMENTS

We are grateful to Drs R L Friedman and M McMahon for allowing us to screen their cDNA library derived from the Burkitt's lymphoma cell line Daudi. We also thank Adrian Kelly for help with sequencing, and W F bodmer and J Rogers for helpful discussions. We also thank Dr N Spurr for providing human and mouse DNA samples, Dr S Povey for providing Bu cells and Drs P Goodfellow and E Solomon for providing somatic cell hybrid DNA. We are very grateful to Susan Simmons for typing this manuscript.

REFERENCES

1. Masters J N, Yang J K, Cellini A and Attardi G (1983). J Mol Biol 167, 23-36.
2. Chen M, Shimada T, Moulton A D, Harrison M and Nienhuis A W (1982). Proc Natl Acad Sci USA 79, 7435-7439.

3. Scarpulla R C (1984). Mol Cell Biol, 4, 2279-2288.
4. Lee M G, Lewis S A, Wilde C D and Cowan N J (1983). Cell 33, 477-487.
5. Lemischka I and Sharp P A (1982). Nature 300, 330-335.
6. Karin M and Richards R I (1982). Nature 299, 797-802.
7. Wiedemann L M and Perry R P (1984). Mol Cell Biol 4, 2518-2528.
8. Klein A and Meyuhas O (1984). Nucleic Acids Res 12, 3763-3776.
9. Peled-Yalif E, Cohen-Binder I and Meyuhas O (1984). Gene 29, 157-166.
10. Dudov K P and Perry R P (1984). Cell 37, 457-468.
11. Nishioka Y, Leder A and Leder P (1980). Proc Natl Acad Sci USA 77, 2806-2809.
12. Vanin E F, Goldberg G I, Tucker P W and Smithies O (1980). Nature 286, 222-226.
13. Hollis G F, Hieter P A, McBride O W, Swan D and Leder P (1982). Nature 296, 321-325.
14. Trowsdale J, Kelly A, Lee J, Carson S, Austin P and Travers P (1984). Cell 38, 241-249.
15. Rigby P W J, Diekmann M, Rhodes C and Berg P (1977). J Mol Biol 113, 237-251.
16. Sanger F, Coulson A R, Barrell B G, Smith A J H and Roe B A (1980). J Mol Biol 143, 161-178.
17. Biggin M D, Gibson T J and Hong G F (1983). Proc Natl Acad Sci USA 80, 3963-3965.
18. Southern E (1975). J Mol Biol 98, 503-517.
19. Nabholz M, Miggiano V and Bodmer W (1969). Nature 223, 358-363.
20. Goodfellow P and Solomon E (1982). In Monoclonal Antibodies in Clinical Medicine, A J McMichael and J W Fabre, eds. (New York: Academic Press) pp 365-393.
21. Feagin J E, Setzer D R and Schimke R T (1983). J Biol Chem 258, 2480-2487.
22. Widera G and Flavell R A (1985). Proc Natl Acad Sci USA 82, 5500-5504.
23. Slightom J L, Blechl A E and Smithies O (1980). Cell 21, 627-638.
24. Bothwell A, Paskind M, Reth M, Imanishi-Kari T, Rajewsky K and Baltimore D (1981). Cell 24, 625-637.
25. Mellor A L, Weiss E H, Ramachandran K and Flavell R A (1983). Nature 306, 792-795.
26. Weiss E, Golden L, Zakut R, Mellor A, Fahrner K, Kvist S and Flavell R A (1983). EMBO J 2, 453-462.
27. Weiss E H, Mellor A, Golden L, Fahrner K, Simpson E, Hurst J and Flavell R A (1983). Nature 301, 671-674.
28. Widera G and Flavell R A (1984). EMBO J 3, 1221-1225.
29. Auffray C, Lillie J W, Arnot D, Grossberger D, Kappes D and Strominger J L (1984). Nature 308, 327-333.