

# Nucleotide sequence of an HLA-A1 gene

J.Girdlestone

Laboratory of Molecular Biology, MRC Centre, Hills Road, Cambridge CB2 2QH, UK

Submitted October 12, 1990

EMBL accession no. X55710

I report here the complete genomic sequence of the HLA-A1 gene from the human thymoma line MOLT4 (1), which has been serologically typed as A1, A3, Bw57, B18 (2). A  $\lambda$  2001 library constructed with BamHI-digested DNA from the MOLT4 derivative NH (kindly provided by Dr. N. Migone) was screened at low stringency with nick-translated insert of the HLA-B7 clone p001 (kindly provided by Dr. S. Weissman). Positive clones were re-screened at high stringency ( $0.1 \times \text{SSC}$ ,  $65^\circ$ ) with prime-cut probes derived from M13 cDNA clones representing the 3' untranslated regions of HLA-A and -B alleles from MOLT4 (2). A clone which hybridized with only the HLA-A probe contained a complete gene on a 4.7 kb HindIII fragment, which is diagnostic for the HLA-A1 allele (3). This fragment was sub-cloned into pUC9, and further shotgun-cloned (4) into Mp18 for sequencing on both strands using the Sanger dideoxy method (5). The

sequence encodes a functional HLA class I gene, and the exons show perfect identity at the nucleotide level with a genomic clone for HLA-A1 derived from the cell line LCL 721 (6).

## REFERENCES

1. Minowada, J., Ohnuma, T. and Moore, G.E. (1972) *J. Natl. Canc. Inst.* **49**, 891.
2. Girdlestone, J. and Milstein, C. (1988) *Eur. J. Immunol.* **18**, 139–143.
3. Orr, H.T. and DeMars, R. (1983) *Immunogenetics* **18**, 489–502.
4. Bankier, A.T. and Barrell, B.G. (1983) In Flavell, R.A. (ed.), *Techniques in the Life Sciences*. Elsevier Scientific Publishers, Ireland, Vol. **B5**, 1–34.
5. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5476.
6. Parham, P., Lawlor, D.A., Lomen, C.E. and Ennis, P.D. (1989) *J. Immunol.* **142**, 3937.

```

AAGCTACTCTGACCAAACTCCATGGGGTATTTTTCTTAGAAGAGTCCAGGTGGACAGGTAAGGAGTGGGAGTCAGGGATCCAGTTCAGGGACAGAGATAATGGGATGAAAA 120
GTGAAAGGAGAGGGACGGGGCCATGCCGAGGGTTTCTCCCTGTTTCTCAGACAGCTCCTGGGCCAAGACTCAGGGAGACATTGAGACAGAGCGCTTCGCACAGGACAGAGGGGTCTAG 240
GGCGAAGTCCCAGGGCCCAAGCGTSGCTCTCAGGGTCTCAGGGCCCGAAGGCGGTGTATGATTGGGGAGTCCCAGCCTTGGGGATTCCCCAACTCCCGCAGTTTCTTTCTCCCTCCTCC 360
CAACCTACGTAGGGTCTTCATCCTGATACTCAGCAGCGGGACCCAGTCTCACTCCCATTTGGGTGTCCGGTTTCCAGAGAAGCCAAATCAGTGTCTGCGGGTCTGCTTCTTAAAGTCC 480
      M A V M A P R T L L L L L S G A L A L T Q T W A
GCACGCCACCACGGGACTCAGATTCTCCCCAGACGCCGAGGATGGCCGTCATGGCCGCCGAAACCTCTCCTGCTACTCTCGGGGGCCCTGGCCCTGACCCAGACCTGGGGCGGTGAG 600
TGCGGGTCCGGGAGGAAACCGCCTCTGCGGGGAGAAGCAAGGGGCCCTCTGGCGGGGGCGCAGGACCGGGGAGCCGCCGGGAGGGGTCCGGCAGTCTCAGCCACTGCTCGCC 720
      G S H S M R Y F F T S V S R P G R G E P R F I A V G Y V D D T Q F V R F D S D
CCCAGGCTCCCACTCAGGATTTCTTTCACATCCGTGTCGCCGGCCGGCCGGGGAGCCCGCTTCATCGCCGTGGGCTACGTGGACGACACGCACTTCGTGGCGGTTCGACAGCGA 840
A A S Q K M E P R A P W I E Q E G P E Y W D Q E T R N M K A H S Q T D R A N L G
CGCCCGAGCCAGAAGTGGAGCCGGGGCCCGTGGATAGAGCAGGAGGGCCGGAGTATGGGACCAGGAGACACGGAATATGAAGGCCCACTCACAGACTGACCGAGCGAAACCTGGG 960
      T L R G Y Y N Q S E D
GACCTGCGCGGTACTACAACAGAGCAGGACGGTGAAGTACCCCGGCCGGGGCGCAGGTACAGACCCCTCATCCCCACGGACGGGGCCAGGTCCGCCACAGTCTCCGGGTCCGAGAT 1080
CCACCCGAAGCCGGGACTCCGAGACCCCTTGTCCCGGAGAGGCCAAGCCGCTTACCCTGTTTCACTTTTCACTTTAGGCCAAAAATCCCCCGGGTGGTCCGGGGCGGGGGCGGG 1200
      G S H T I Q I M Y G C D V G P D G R F L R G Y R Q D A Y D
TCGGGGACTGGGCTGACCGGGGTCCGGGCCAGGTTCTCACACCTCAGATAATGATAGCTGCGACGTGGGGCCGGACGGGGCGTCTCTCCCGGGTACCGGCAGGACCGCTACGA 1320
G K D Y I A L N E D L R S W T A A D M A T G Q I T K R K W E A V H A E Q R R V Y
CGGCAAGGATTACATCCCTGACAGGACCTGCGCTTGGACCGCGCGGACATGGCAGTTCAGATCACCAAGCGCAAGTGGGAGGGTCCATGGCGGGAGCAGCGGAGGTCTA 1440
L E G R C V D G L R Y L E N G K E T L Q R T
CTGGAGGGCCGGTTCGGTGGACGGGCTCCGACATACCTGGAGAACGGGAAGGACGCTGCAGCCGCGGTACCGGGGCCCGGGCGCCTCCCTGATCGCCTATAGATCTCCCGGG 1560
CTGGCCTCCCAAGGAGGGGAGCAATTTGGGACCAACTAGAATATCACCTCCCTCTGGTCTGAGGGAGAGGAATCCTCCTGGGTTTCCAGATCTGTACCAGAGAGTGACTCTGA 1680
GGTCCCGCCTGCTCTCAGCAAAATTAAGGGATAAAATCTCTGAAGGAGTACGGGAAGCAGATCCCTCGAATCTGATGAGTGGTCCCTTTGACACCGGCAGCAGCTTGGGCCCGT 1800
GACTTTCCCTCAGGCCCTGTCTCTGCTCACACTCAATGTGTGGGGTCTGAGTCCAGCACTCTGAGTCTCTCAGCCTCCACTCAGGTCAGGACAGAAAGTGCCTGTTCCTCTC 1920
TCAGGGAATAGAAGATTATCCAGGTGCCTGTGTCCAGGCTGGTGTCTGGTCTGTCTCTTCCCATCCCGGGTGTCTGTCCATTCTCAAGATGGCCACATGCGTGTGGTGGAG 2040
      D P P K T H M T H H P I S D H E A T L R C W A L
TGTCCTGACAGATGCAAAATGCCTGAATTTTCTGACTCTTCCCGTCAGACCCCCCAAGACATATGACCCACCCACCTCTGACCATGAGGCGACCCCTGAGGTGCTGGGCCCT 2160
G F Y P A E I T L T W Q R D G E D Q T Q D T E L V E T R P A G D G T F Q K W A A
GGGCTTACCTCGGAGATCACACTGACCTGGCAGCGGATGGGGAGGACAGCCAGGACCGGAGCTCGTGAGACAGGCCCTGACGGGATGGAACTTCCAGAAAGTGGGGCGG 2280
V V V P S G E E Q R Y T C H V Q H E G L P K P L T L R W
TGTGTGTGCTCTGAGAGGAGCAGAGATACACTGCCATGTGACAGTGAAGGTTGCCCCAAGCCCTCACCTGAGTGGGTAAGGAGGAGATGGGGGTGTCTGTCTCTTAG 2400
      E L S S Q P T I P I V G I I A G L V
GGAAAGCAGGACCTCTCTGGAGCCTTAGCAGGGTCAAGGCCCTCACCTTCCCTCTTTCCAGAGCTTCTCCAGAGCTTCTCCAGCCACCTCCCATCTGCTGCGCTGTG 2520
L L G A V I T G A V V A A V M W R R K S S
TCTCCTGGAGTGTGATCACTGGAGTGTGGTGCCTGTGAGGAGGAGGAGCTCAGGTGGAGAAGGGTGAAGGGTGGGGTCTGAGATTTCTGTCTCACTGAGGGTTCCA 2640
AGCCCCAGCTAGAAATGTCCCTGTCTCACTTACTGGGAAGCACCTCCACAATCATGGGCCGACCCAGCCCTGGGCCCTGTGTGCCAGCACTACTCTTTTGAAGCACCTGTTAAATG 2760
AAGGACAGATTATACCTTGTATACGGCGGTGATGGGACCTGATCCAGCAGTCAAAAGTCAAGGGAAGGTCCTGAGGACAGACCTCAGGAGGGCTATTGGTCCAGGACCCACCC 2880
TGCTTTCTCATGTTTCTGATCCCGCCTGGGTCTGCAGTCAACATTTCTGAAACTCTCTGGGGTCAAGACTAGGAGGTTCTCTAGGACCTTAAAGCCCTGGCTCCTTCTGGT 3000
      D R K G G S Y T Q A A
ATCTCACAGGACATTTCTTCCACAGATAGAAAAGGAGGGAGTTACACTCAGGCTGCAAGTAAATGATGAAGGAGGCTGATGCCTGAGGTCCTGGGATGTTGGGAGCCCATGG 3120
      S S D S A Q G S D V S L T
GGGAGCTCACCCACCCACAATCTCTCTAGCCACATCTTCTGTGGATCTGACAGGTTCTGTTTTTGTCTACCCAGGCAGTACAGTGCACAGGGCTCTGATGTGTCTCACA 3240
A C K
GCTTGAAGGTGAGAGCTTGGAGGGCTGATGTGTGTGGTGTGGGTGGAAACAGTGGACACAGCTGTGCTATGGGGTTCTTTGCGTGGATGATTAGCATGCGATGGGCTGTT 3360
      V
AAGGTGTACCCCTCACTGTGATGGATGAATTTGTTTCATGATATTTTTCTTAGTGTGAGACAGCTGCCTGTGTGGACTGAGAGGCAAGAGTGTTCCTGCCCTCCCTTTGT 3480
GACTTGAAGAACCTGACTTTGTTCTGCAAAAGGCACCTGCAATGTGTCTGTGTTGTTAGGCAATAATGTAGGAGTGGGGAGGACACCCCAATGTCACCATGACCTCTTCC 3600
CACGCTGACCTGTGCTCCCTCTCAATCATCTTCTGTTCAGAGAGTGGGGTGAAGTGTCTCCATCTCTGCTCAACTTCATGGTGCACGTGAGCTGTAACCTCTTCTTCCCTT 3720
AAAATTAGAACCTGAGTATAAAATTTACTTCTCAAAATCTTCCATGAGAGGTTGATGAGTAAATTAAGGAGAAGATTCCTAAATTTGAGAGACAAAATTAATGAACGCATGAGAAC 3840
CCAAT, TATA, and poly-adenylation signals are underlined

```