Sequence of a rat MHC class II-associated invariant chain cDNA clone containing a 64 amino acid thyroglobulin-like domain

Andrew J.McKnight, Don W.Mason and A.Neil Barclay

MRC Cellular Immunology Unit, Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK
Submitted April 11, 1989
EMBL accession no. X14254

In the course of studies aimed at cloning rat Interleukin 2 (IL-2) we fortuitously isolated several cDNA clones encoding the rat MHC class II-associated invariant chain (1). The complete nucleotide sequence of one of these clones (pRinv.B9) is shown in figure 1. A cDNA library, constructed from RNA extracted from the spleen of a rat undergoing graft-versus-host disease, had been screened with a mouse IL-2 cDNA probe (2). The rat invariant chain cDNA clones isolated were presumably recognised due to a repeating CAG sequence found within the open reading frame of mouse IL-2 and in the 5' non-coding region of the rat invariant chain (McKnight et al. submitted).

Fig. 1: Nucleotide sequence of cDNA clone pRinv.B9. The initiation site codon, termination codon and polyadenylation signal are shown in upper case.

The sequence of a rat invariant chain cDNA clone has recently been published (3), but this clone lacked a 64 amino acid segment identified in some alternative forms of human and mouse invariant chain (4-6). The complete protein sequence of the rat invariant chain is shown in figure 2, with the alternatively-spliced 64 amino acid domain in bold type. This 64 amino acid region shows considerably higher cross-species homology than the rest of the molecule (Fig. 3) and contains six cysteine residues plus two potential N-linked glycosylation sites, in addition to a degree of homology to thyroglobulin based around a Cys-Trp-Cys-Val motif which is found in most of the type I repetitive units of human thyroglobulin (7). Our nucleotide sequence, apart from the extra segment, is similar to that published by Henkes et al. apart from a discrepancy over the number of C and G residues between nucleotides 997-1004 in the 3' untranslated region, and the

```
MDDQRDLISNHEQLPILGQRARAPESNCNRGVLYTSVSVLVALLLAGQATTAYFLYQQQG 60
RLDKLTVTSQNLQLENLRMKLPKSAKPVSPMRMATPLLMRPLSMDNMLQAPVKNVTKYGN 120
MTQDHVMHLLTKSGPVNYPQLKGSFPENLKHLKNSMNGLDWKVFESWMKQWLLFEMSKNS 180
LEEKQPTQTPPKVLTKCQEEVSH1PDVHPGAFRPKCDENGNYMPLQCHGSTGYCWCVFPN 240
GTEVPHTKSRGRHNCSEPLDMEDPSSGLGVTKQDMGQMFL 280
```

Fig. 2: Complete protein sequence of rat invariant chain deduced from clone pRinv.B9. The additional 64 amino acid segment, not seen in a previous rat clone, is shown in bold type.

© IRL Press 3983

```
Rat inv (193-256) V L T K C Q E E V S H I P D V H P G A F R P K C D E N G N Mou inv (192-255) V L T K C Q E E V S H I P A V Y P G A F R P K C D E N G N Hum inv (193-256) V L T K C Q E E V S H I P A V H P G S F R P K C D E N G N Hum Tg (97-160) Q L Q K Q Q I L L S G Y I N S T D T S Y L P Q C Q D S G D Y M P L Q C H G S T G Y C W C V F P N G T E V P H T K S R G R H N C S Y L P L Q C H G S T G Y C W C V F P N G T E V P H T K S R G R H N C S Y L P L Q C Y G S I G Y C W C V F P N G T E V P N T R S R G H H N C S Y L P L Q C Y G S I G Y C W C V F P N G T E V P N T R S R G H H N C S Y L P L Q C Y G S I G Y C W C V F P N G T E V P N T R S R G H H N C S Y L P L Q C Y G S I G Y C W C V F P N G T E V P N T R S R G H H N C S Y L P L Q C Y G S I G Y C W C V F P N G T E V P N T R S R G H H N C S Y L P L Q C Y G S I G Y C W C V F P N G T E V P N T R S R G H H N C S Y L P L Q C Y G S I G Y C W C V F P N G T E V P N T R S R G H H N C S Y L P L Q C Y G S I G Y C W C V F P N G T E V P N T R S R G H H N C S Y L P L Q C Y G S I G Y C W C V F P N G T E V P N T R S R G H H N C S Y L P L Q C Y G S I G Y C W C V F P N G T E V P N T R S R G H H N C S Y L P L Q C Y G S I G Y C W C V F P N G T E V P N T R S R G H H N C S Y L P L Q C Y G S I G Y C W C V F P N G T E V P N T R S R G H H N C S Y L P L Q C Y G S I G Y C W C V F P N G T E V P N T R S R G H H N C S Y L P L Q C Y G S I G Y C W C V F P N G T R V T G T R Q L G R P K R C
```

Fig. 3: Alignment of the amino acid sequence encoded by the additional segment in rat invariant chain cDNA with the sequence encoded by exon 6b of mouse and human invariant chain genes and the sequence preceding and including the second type I repetitive unit of human thyroglobulin. Amino acids identical in 2 or more sequences are boxed. The amino acid homology between rat and human invariant chain proteins in this region is 88% compared with 75% for the rest of the protein.

presence of eight extra nucleotides at the extreme 5' end of our clone. Also, we have noted four differences between the protein sequences at positions 16, 17, 136 and 147 where Henkes *et al.* have incorrectly translated the nucleotide sequence.

<u>Acknowledgements</u>: The mouse IL-2 cDNA clone pcD-IL-2 was kindly provided by Tim Mosmann. A. J. McK. was supported by the Arthritis and Rheumatism Council, UK.

## References

- (1) Jones, P. P., Murphy, D. B., Hewgill, D. and McDevitt, H. O. (1978) *Immunochem.* 16: 51-60.
- (2) Yokota, T., Arai, N., Lee, F., Rennick, D., Mosmann, T. and Arai, K-I. (1985) *Proc. Natl. Acad. Sci. U.S.A.* 82: 68-72.
- (3) Henkes, W., Syha, J. and Reske, K. (1988) Nucleic Acids Res. 16: 11822.
- (4) Strubin, M., Berte, C. and Mach, B. (1986) EMBO J. 5: 3483-3488.
- (5) Koch, N., Lauer, W., Habicht, J. and Dobberstein, B. (1987) *EMBO J*. 6: 1677-1683
- (6) Zhu, L. and Jones, P. P. (1989) Nucleic Acids Res. 17: 447-448.
- (7) Malthiéry, Y. and Lissitzky, S. (1987) Eur. J. Biochem. 155: 491-498.