

Complete nucleotide sequence of the cosmid vector pWE15A

Donald Seto, Ben F.Koop, Jason Seto and Leroy Hood

Division of Biology (147-75), California Institute of Technology, Pasadena, CA 91125, USA

Submitted May 26, 1992

EMBL accession no. Z12112

Current efforts to determine the complete nucleotide sequence of the six murine and human T-cell receptor (TCR) loci are limited by the availability of overlapping genomic inserts cloned into a defined and characterized vector. In order to create these mapping and sequencing libraries, our laboratory has modified an existing cosmid vector pWE15 (1, 2). The optimized version, pWE15A (3), contains a polylinker with 15 infrequently cleaved restriction enzyme sites, asymmetrically centered around the BamHI site, which is the cloning site used for the insertion of the genomic DNA described below. The modifications allow recovery of inserted DNA by the use of several restriction enzymes. Each recombinant insert bearing vector may contain sequenceable genomic DNA ranging in size from 35 to 40 kilobases (kb). This complete 8213 nucleotide pWE15A sequence complements the two noncontiguous approximately 2 kb of pWE15A extant in GenBank.

Our laboratory has begun the mapping and sequencing process by making libraries of overlapping human TCR β locus clones (K.Wang, manuscript in preparation), murine TCR β locus clones (C.Boysen, personal communication), human and murine α/δ loci clones (K.Wang, manuscripts in preparation), and murine γ locus clones (B.Vernooij, manuscript in preparation). These libraries are the basis of our current large-scale DNA sequencing endeavor. Initial DNA sequencing projects include a murine V α/δ cosmid clone (D.Seto, manuscript in preparation), murine and human J α clones (B.F.Koop, manuscripts submitted), and human V β clones (L.Rowen, personal communication). These sequencing efforts involve the random 'shotgun' approach (4) which should yield a 5- to 10-fold redundancy of DNA sequences. Due to the sheer number of sequences, which also includes vector sequences, the assembly process may be difficult given current algorithms. The process is simplified when the vector sequences are subtracted from the data set prior to assembly (D.Seto, manuscript submitted).

The pWE15A cosmid vector has been completely sequenced using automated fluorescent techniques as described by our laboratory (6). This 8.2 kb of sequence has been generated from a data set containing an average of approximately 10-fold redundancy. Areas of compression in the sequencing gel have

been resolved by incorporating base analogs such as 7-deazaGTP and 7-deazaITP in place of dGTP, and also by using either T7 DNA polymerase or Taq DNA polymerase. One area (4 nucleotides) remained unresolvable despite these measures; however this area has been sequenced by several other laboratories (example, 7) and its sequence is incorporated into our sequence (positions 1883–1886).

Cosmid vector pWE15A has been completely sequenced and assembled for the first time and is being released into EMBL (accession number Z12112) so that any investigator wishing to use our pWE15A-based libraries may have ready access to the entire complete sequence data. This complete characterization of pWE15A will enhance its role in high-resolution restriction enzyme mapping and rapid chromosomal walking. Importantly, the pWE15A sequences may be used to screen out vector components when 'shotgun' sequencing inserts are cloned into this cosmid or into its predecessor pWE15.

ACKNOWLEDGEMENTS

D.S. is a Lawton Chiles Fellow of the National Institutes of Health, General Medical Sciences (GM 13039). B.F.K. is supported by a DOE Human Genome Distinguished Postdoctoral Fellowship. Additional support came from DOE (FG0391ER61182) and NIH (HG00356). We thank Cecilie Boysen and Bernard Vernooij for critically reading this manuscript.

REFERENCES

1. Wahl, G.M., Lewis, K.A., Ruiz, J.C., Rothenberg, B., Zhao, J. and Evans, G.A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2160–2164.
2. Evans, G.A. and Wahl, G.M. (1987) *Methods Enzymol.* **152**, 604–610.
3. Lai, E., Wang, K., Avdalovic, N. and Hood, L. (1991) *BioTechniques* **11**, 212–217.
4. Deininger, P.L. (1983) *Anal. Biochem.* **129**, 216–223.
5. Staden, R. (1980) *Nucleic Acids Res.* **8**, 3673–3694.
6. Koop, B.F., Wilson, R.K., Chen, C., Halloran, N., Sciamma, R., Hood, L. and Lindelien, J.W. (1990) *BioTechniques* **9**, 32–37.
7. Deng, T., Noel, J.P. and Tsai, M.-D. (1990) *Gene* **93**, 229–234.