Sequence and organization of the mouse poly (ADP-ribose) polymerase gene

Konrad Huppi, Kishor Bhatia[1], David Siwarski, Denis Klinman[2], Barry Cherney[1] and Mark Smulson[1]

Laboratory of Genetics, NCI/NIH, Bldg 37, Rm 2B-21 Bethesda, MD 20892, [1]Department of Biochemistry, Georgetown University School of Mediciine, Washington, DC 20007 and [2]Laboratory of Molecular and Viral Pathogenesis, NINDS/NIH Bethesda, MD, USA

Received February 17, 1989; Revised and Accepted March 29, 1989 EMBL accession no. X14206

ABSTRACT

Using a human cDNA probe, we have isolated murine genomic and cDNA clones corresponding to the nuclear enzyme poly (ADP-ribose) polymerase (ADPRP). Northern analysis with the mouse cDNA clones reveals transcripts of 3.7−3.8kb corresponding in size to the human ADPRP transcript. DNA sequence comparisons between mouse and human clones reveals extensive amino acid sequence conservation within regions harboring DNA binding, NAD^+ binding or automodification domains. A survey among mouse inbred strains for restriction fragment length polymorphism (RFLP) reveals at least three distinct ADPRP alleles. The segregation of alleles among mouse genetic recombinants positions *ADPRP* on mouse chromosome 1 between the complement receptor-related gene *At-3* and the Fc receptor locus *FcR*. Furthermore, *ADPRP* is closely associated with the autoimmune locus *gld* (*g*eneralized *l*ympha*d*enopathy).

INTRODUCTION

The nuclear enzyme poly (ADP-ribose) polymerase (ADPRP) utilizes the coenyzme NAD^+ as a substrate for the covalent transfer of multiple ADP-ribose units to various nuclear acceptors including nucleosomal core histones, histone H1, HMG proteins, topoisomerase I and II and SV40 large T antigen. This complex enzyme contains functional domains specific in binding NAD^+, and DNA as well as an automodification domain. The activity of ADPRP is stimulated by DNA strand breaks suggesting ADPRP may function alone or with other enzymes in DNA repair, replication and differentiation (1,2). High levels of ADPRP have been detected in patients with rheumatic disorders suggesting aberrant or excessive ADPRP activity in lymphocytes may also be associated with certain autoimmune disorders (3). Indeed, many autoreactive human diseases such as Ataxia telangietasia (AT), Fanconi's Anemia (FA), Bloom's syndrome and Cockayne Syndrome are characteristically associated with DNA repair defects (4).

Recent cDNA cloning of the full-length human ADPRP (huADPRP) gene has established the sequence (5−7) and chromosomal location of three closely-related ADPRP genes (6). We have determined that the expressed gene for huADPRP is located on human chromosome 1 and related genes or pseudogenes are found on chromosomes 13 and 14 (6). Studies from this laboratory and others have also identified potential regions in the mouse which harbor genes involved in effective DNA repair and/or reduced susceptibility to tumorigenesis. (8,9) We have now used the huADPRP cDNA probe to obtain genomic and cDNA clones corresponding to the mouse expressed gene for ADPRP. Identity in size of the human and mouse transcripts as well as direct sequence comparisons confirm the mouse gene is indeed the direct homologue of huADPRP.

Preliminary analysis of genomic phage clones of mouse ADPRP (moADPRP) indicate

the mouse locus must be larger than 22 kilobases (Kb) and consists of approximately 15−20 exons; each separated by more than 1 Kb of intervening sequence. Furthermore, Southern hybridization studies using the mouse cDNA probe have established restriction fragment length polymorphisms (RFLPs) among Hind III or Eco RV digested genomic DNA from inbred strains of mice. As a result, we are able to assign inbred strain haplotypes and to position the expressed moADPRP gene on chromosome 1 close to the autoimmune locus *gld* (*generalized lymphoproliferative disorder*) (10). In addition, cDNA clones homologous to ADPRP are also presumed to encode genes related to ADPRP functional domains and map to additional chromosomal loci (i.e., mouse chromosome 4).

MATERIALS AND METHODS

DNA and RNA resources-

DNA was prepared as described previously (11) from spleen or liver tissue of inbred strains, recombinant inbred strains or congenic strains of mice: AKR/N, BALB/cAn, C3H/Fg, C57BL/10N, DBA/2N, NZB/B1NJ, SJL/JAn, BALB/cAn. DBA/2N (C.D2) C.D2 Fv-$1^{n/n}$, C.D2 Pep3/Akp-1 (Hazleton Laboratories, Rockville MD)., AU/SsJ, BSVS, C3H/HeJ, C57BL/6J, C57L/J, C57/BR, C58/J, CBA/J, CBA/N, Dwf/Dwf, HRS/J, I, MA/MyJ, MRL, NH, N2W, PL/J, RIIIs/J, RF/J, SM/J, ST/bJ, SWR/J, 129/J, C57BL/6J X DBA/2N (BXD) recombinant inbred (RI) mice (The Jackson Labs), 020/A (J. Hilgers, Amsterdam).

RNA was prepared from tissue or tumors as described (12). The following mouse tumors were obtained from Hazleton laboratories: TEPC 2027, TEPC 1017, TEPC 1033, ABPC18, ABPC33, MOPC 104E. The human HeLa tumor sample is maintained in culture at Georgetown University.

Genomic and cDNA libraries-

A partial EcoR1 genomic library was made by size fractionation of partially digested high molecular weight DNA from BALB/cAn liver. The fractions (12−23 kb) were pooled and ligated to EcoRI digested Charon 30 arms and subsequently packaged in vitro.

A cDNA library was generated by synthesis of cDNA from poly A+ enriched splenic mRNA of the autoimmune mouse BXSB. The inserts carrying EcoRI linkers were ligated to a λ gT10 vector.

DNA probes and hybridization conditions-

A full-length 3.8kb cDNA probe for human ADPRP, pcD, was originally cloned and sequenced as described (6). The conditions for digestion of genomic DNA, agarose gel electrophoresis, southern transfer and hybridization have been described (11). For Northern hybridization analysis, poly A+ selected mRNA was prepared (12) and electrophoresed on 4mM methylmercury hydroxide agarose gels (1.0%) in E buffer (50mM Boric Acid, 5mM sodium borate, 10mM sodium sulfate). Transfer was performed by blotting in 20×SSC (SSC=0.15 M sodium chloride, 0.015 M sodium citrate) to Nytran filters followed by a 2 hour bake at 80°C. Hybridization was carried out at 42°C in 50% formamide, 5×SSC, 0.1% sodium dodecyl sulfate (SDS), 5×Denhardt's solution (Denhardt's = 0.02% each of bovine serum albumin, Ficoll, and polyvinylpyrrolidone) and 10% dextran sulfate. All final wash conditions, unless otherwise mentioned are 0.2×SSC, 0.1% SDS and 5mM EDTA at 65°C.

DNA Sequencing-

All DNA sequencing reactions were performed on Sau3A, RsaI, EcoR1, BamH1-EcoR1,

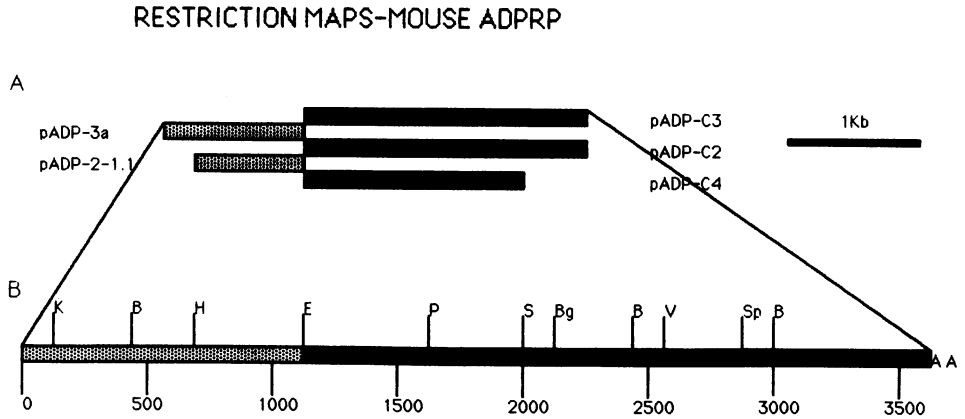


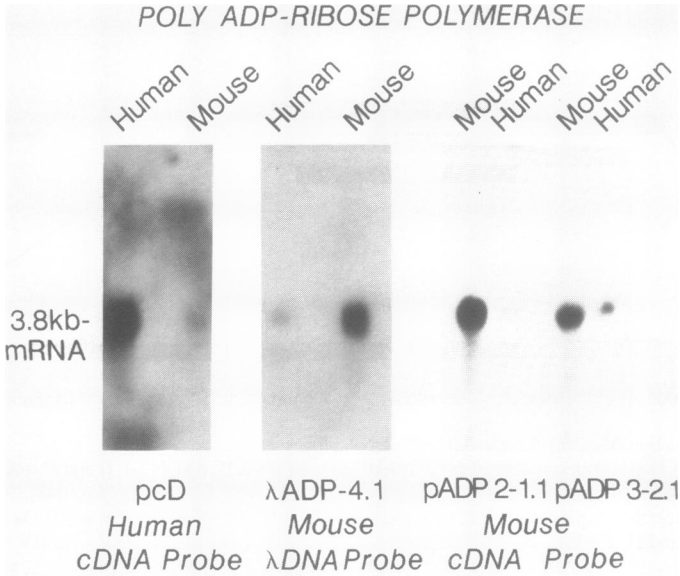
Figure 1—Restriction Maps of Mouse ADPRP cDNA clones.
A). BXSB cDNA clones corresponding to the 5′ (pADP-3a (pADP3-2.1), pADP-2-1.1) and 3′ (pADP-C3, pADP-C2, pADP-C4) halves of mouse ADPRP with respect to the internal EcoR1 site (res. 340).
B). Detailed restriction map of entire coding region of mouse ADPRP (3.8kb) from pADP-3a and pADP-C3. K=KpnI, H=HindIII, P=PstI, Bg=BglII, Sp=SphI, E=EcoR1, B=BamH1, RV=EcoRV

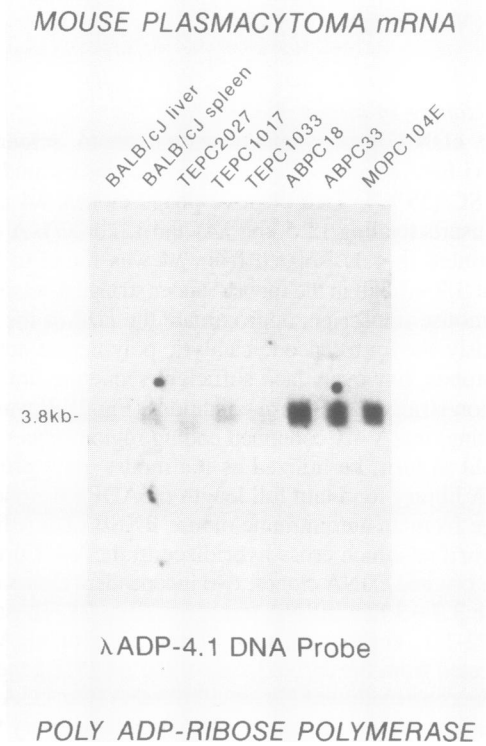
SphI-EcoR1 or HindIII-EcoR1 M13 subclones by the dideoxysequencing protocol described by the manufacturer (N.E. Biolabs or BRL).

RESULTS

Genomic and cDNA cloning of mouse ADPRP

The human ADPRP cDNA probe, pCD, was used to screen approximately 10^6 recombinants of a BALB/cAn partial EcoR1 genomic library under relaxed stringency conditions of 0.2 ×SSC (55°C). Two positive phage clones, λ4 and λ7, were purified and found to contain inserts totaling 12.5 kb (λ4) and 12kb(λ7). A genomic 4kb HindIII-EcoRI restriction fragment (λ 4.1) isolated from λ4 was found to hybridize specifically to an mRNA doublet of 3.7−3.8kb in the mouse (under stringent wash conditions 0.2×SSC, 65°C) (Fig. 2). The mouse transcripts approximate the size of the huADPRP transcript which we had previously shown to encode catalytic polymerase activity (5). The human and mouse ADPRP probes, however, lack sufficient sequence similarity that they cross-hybridize only under non-stringent Northern conditions (Fig.2). It was assumed (and shown later by direct sequencing) that λ 4.1 contained coding region sequences of mouse ADPRP (moADPRP) and could in turn, be utilized as the mouse gene probe for more efficient screening of the cDNA library to obtain full length moADPRP clones. Among more than 10^6 recombinant phage from an autoimmune mouse BXSB cDNA library, 8 independent clones were initially purified which cross-hybridized to the λ 4.1 probe under the relaxed stringency. Among the original cDNA clones, two independent clones continued to hybridize specifically to a mouse 3.7−3.8kb mRNA transcript (Fig. 2). These clones, pADP 3-2.1 (pADP3a) and pADP2-1.1, correspond to different sizes of moADPRP NH_2-terminal sequence but are truncated from the carboxy-end of the ADPRP gene by an internal EcoRI site (used initially in the construction of the λ gt10 library) (Fig 1). A differential screening process was subsequently used to rescreen the BXSB cDNA library to obtain carboxy-end


POLY ADP-RIBOSE POLYMERASE
Human
Mouse
Human
Mouse
Mouse
Human
Mouse
Human
3.8kb-
mRNA
pcD
Human
cDNA Probe
λADP-4.1
Mouse
λDNA Probe
pADP 2-1.1 pADP 3-2.1
Mouse
cDNA Probe


MOUSE PLASMACYTOMA mRNA
BALB/cJ liver
BALB/cJ spleen
TEPC2027
TEPC1017
TEPC1033
ABPC18
ABPC33
MOPC104E
3.8kb-
λADP-4.1 DNA Probe
POLY ADP-RIBOSE POLYMERASE

sequences of moADPRP. Three cDNA clones (pADP-C2,C3,C4) where obtained, among which pADP-C2 and pADP-C3 corresponded to mRNA from the internal EcoRI site in the 3′ direction to the polyadenylation signal. The combination of cDNA inserts encompassing the NH_2-terminal (1.2 Kb) pADP-3a and carboxy terminal (2.6 Kb) pADP-C3 or pADP-C2 sequences yields a complete mRNA for moADPRP. DNA sequence comparison with the full length huADPRP cDNA sequence reveals that, indeed, the combined mouse sequences of 5′ pADP-3a and 3′ pADP-C3 extend uninterrupted −84 from base pairs (bp)5′ of residue[1] (MET) to more than 400bp 3′ of the termination signal. Furthermore, these sequences are contiguous and nonoverlapping (Fig.3). In fact only a single amino acid residue (Leu) at position 505 relative to the human sequence appears to be missing in the BXSB mouse sequence (Fig.4). There are no other deletions between mouse and human cDNA sequences in the 1014 residues encoding huADPRP.

Human and Mouse Sequence Comparisons

The 5′ untranslated region of the mouse contains very little sequence similarity with the corresponding region of huADPRP. In contrast, the 3′ untranslated region of moADPRP is very strongly conserved with respect to the human sequence with the exception of an extra 120bp insertion just prior to the poly-adenylation site. This 120 bp sequence may be the murine counterpart to the human alternative splicing previously recognized (7) and may result in the observation of the doublet observed in Fig 2b.

There is approximately 88% nucleotide sequence similarity within the 3042bp encoding the ADPRP structural gene between human and mouse. The amino acid similarity between mouse and human is much higher (92%) and specific regions reveal greater amino acid sequence conservation than others (Fig. 4). For example, the region between mouse residues 727−812 contains nearly the same frequency of nucleotide changes as adjacent regions without a single amino acid substitution. Whereas nearly 50% of the amino acids are substituted between mouse residues 361−386. This area, in fact, lies between the DNA binding and automodification domains of the protein and therefore, may be less conserved.

Comparison of the huADPRP sequence similarities originally demonstrated several conserved sequence motifs. One motif identified corresponds to a consensus nucleotide binding fold characteristic of ATP requiring enzymes (5,6). As shown in Table I, this fold consists of an 'A' site at mouse residues 888−914 and a 'B' site at mouse residues 940−957. The mouse and human amino acid sequences have complete identity within these regions. The so-called Zn^{++} finger motif (13) has also been identified within the human ADPRP protein sequence (6). This conclusion was based upon a specific arrangement of His and Cys residues as well as a precise duplication of sequence between fingers (Fig.5). The mouse sequence (res 2−res 198) now reveals considerable conservation among amino acid residues deemed important in the Zn^{++} finger structure, despite the fact that there

Figure 2(top)—Northern Hybridization Analysis of Human and Mouse with ADPRP. Poly A+ mRNA from Human (HeLa) or BALB/cJ mouse (spleen) was hybridized to a Human cDNA probe pCD (Left), mouse λ4.1 insert (center) or mouse cDNA probes pADP-3-2.1 (pADP-3a), pADP-2-1.1 (right) under stringent wash conditions (0.2×SSC, 65°C).

(bottom)—Northern Analysis of mouse plasmacytomas with ADPRP λ4.1 probe. The message levels from plasmacytoma ABPC18, ABPC33 and MOPC 104E are much more intense than from a equivalent amount of mRNA in liver, spleen, or TePC 2027, TePC 1017 and TePC 1033. The ADPRP message appears as a doublet of 3.7−3.8kb and is more intense in some tumors such as ABPC18, ABPC33, or MOPC 104E.

```
1    /    1                                   31   /   11
ATG GCG GAG GCC TCG GAG AGG CTT TAT CGA GTG CAG TAC GCG AAG AGC GGG CGC GCC TCT
met ala glu ala ser glu arg leu tyr arg val gln tyr ala lys ser gly arg ala ser
61   /   21                                  91   /   31
TGC AAG AAA TGC AGC GAG AGT ATT CCC AAG GAC TCC CTC CGC ATG GCC ATC ATG GTG CAG
cys lys lys cys ser glu ser ile pro lys asp ser leu arg met ala ile met val gln
121  /   41                                  151  /   51
TCA CCC ATG TTC GAT GGG AAA GTC CCA CAC TGG TAC CAC TTC TCC TGC TTC TGG AAG GTG
ser pro met phe asp gly lys val pro his trp tyr his phe ser cys phe trp lys val
181  /   61                                  211  /   71
GGC CAG TCC ATC CGG CAC CCT GAT GTT GAG GTG GAT GGC TTC TCT GAG CTG CGC TGG GAT
gly gln ser ile arg his pro asp val glu val asp gly phe ser glu leu arg trp asp
241  /   81                                  271  /   91
GAT CAG CAG AAG GTC AAG AAG ACG GCC GAG GCT GGA GGC GTG GCA GGC AAA GGC CAG GAT
asp gln gln lys val lys lys thr ala glu ala gly gly val ala gly lys gly gln asp
301  /  101                                 331  /  111
GGA AGT GGG GGC AAG GCG GAG AAG ACA TTG GGT GAC TTT GCA GCA GAG TAT GCC AAG TCC
gly ser gly gly lys ala glu lys thr leu gly asp phe ala ala glu tyr ala lys ser
361  /  121                                 391  /  131
AAC AGG AGC ATG TGC AAG GGC TGC CTG GAG AAG ATA GAG AAG GGC CAG ATG CGC CTG TCC
asn arg ser met cys lys gly cys leu glu lys ile glu lys gly gln met arg leu ser
421  /  141                                 451  /  151
AAG AAG ATG GTG GAT CCA GAG AAG CCA CAG CTG GGT ATG ATT GAC CGC TGG TAC CAT CCA
lys lys met val asp pro glu lys pro gln leu gly met ile asp arg trp tyr his pro
481  /  161                                 511  /  171
ACT TGC TTT GTC AAG AAG CGG GAC GAG CTG GGC TTC CGG CCC GAG TAC AGT GCC AGT CAG
thr cys phe val lys lys arg asp glu leu gly phe arg pro glu tyr ser ala ser gln
541  /  181                                 571  /  191
CTC AAG GGC TTT AGC CTC CTC TCT GCA GAA GAC AAA GAA GCT CTG AAG AAG CAG CTC CCG
leu lys gly phe ser leu leu ser ala glu asp lys glu ala leu lys lys gln leu pro
601  /  201                                 631  /  211
GCC ATC AAG AAT GAA GGA AAG AGA AAA GGC GAC GAG GTG GAT GGA ACA GAT GAA GTG GCC
ala ile lys asn glu gly lys arg lys gly asp glu val asp gly thr asp glu val ala
661  /  221                                 691  /  231
AAA AAG AAA TCT AGA AAG GAG ACA GAT AAG TAT AGT AAG CTT GAG AAG GCC CTA AAG GCT
lys lys lys ser arg lys glu thr asp lys tyr ser lys leu glu lys ala leu lys ala
721  /  241                                 751  /  251
CAG AAT GAG CTG ATC TGG AAT ATC AAA GAC GAG CTG AAG AAA GCG TGT TCC ACC AAT GAC
gln asn glu leu ile trp asn ile lys asp glu leu lys lys ala cys ser thr asn asp
781  /  261                                 811  /  271
CTG AAG GAG CTG CTC ATC TTC AAC CAG CAG CAA GTG CCG TCA GGA GAG TCA GCG ATC TTG
leu lys glu leu leu ile phe asn gln gln gln val pro ser gly glu ser ala ile leu
841  /  281                                 871  /  291
GAC AGA GTT GCT GAT GGC ATG GCG TTT GGG GCC CTT CTG CCC TGC AAG GAG TGT TCA GGC
asp arg val ala asp gly met ala phe gly ala leu leu pro cys lys glu cys ser gly
901  /  301                                 931  /  311
CAG CTG GTC TTT AAG AGC GAC GCT TAT TAC TGT ACT GGG GAT GTC ACT GCC TGG ACC AAG
gln leu val phe lys ser asp ala tyr tyr cys thr gly asp val thr ala trp thr lys
961  /  321                                 991  /  331
TGC ATG GTC AAG ACA CAG AAT CCT AGC CGA AAG GAA TGG GTA ACT CCA AAG GAA TTC CGA
cys met val lys thr gln asn pro ser arg lys glu trp val thr pro lys glu phe arg
1021 /  341                                 1051 /  351
GAA ATA TCC TAC CTC AAG AAG TTA AAG GTC AAA AAA CAG GAC CGA ATA TTC CCT CCA GAA
glu ile ser tyr leu lys lys leu lys val lys lys gln asp arg ile phe pro pro glu
1081 /  361                                 1111 /  371
AGC AGC GCC CCC ATC ACG GTG CAC TGG CCG CTC TCT GTC ACC TCA GCA CCC ACA GCT GTG
ser ser ala pro ile thr val his trp pro leu ser val thr ser ala pro thr ala val
1141 /  381                                 1171 /  391
AAC TCC TCT GCT CCA GCA GAT AAG CCC CTG TCT AAC ATG AAG ATC CTG ACT CTT GGG AAG
asn ser ser ala pro ala asp lys pro leu ser asn met lys ile leu thr leu gly lys
1201 /  401                                 1231 /  411
CTC TCC CAG AAC AAG GAC GAA GCA AAA GCT GTG ATT GAG AAA CTG GGA GGC AAG TTG ACA
leu ser gln asn lys asp glu ala lys ala val ile glu lys leu gly gly lys leu thr
1261 /  421                                 1291 /  431
GGA TCT GCC AAC AAG GCC TCC TTG TGT ATC AGC ATC AAA AAG GAG GTG GAA AAG ATG AAT
gly ser ala asn lys ala ser leu cys ile ser ile lys lys glu val glu lys met asn
1321 /  441                                 1351 /  451
AAG AAG ATG GAG GAA GTA AAG GAA GCC AAC ATC CGA GTT GTG TCT GAG GAC TTC CTC CAG
lys lys met glu glu val lys glu ala asn ile arg val val ser glu asp phe leu gln
1381 /  461                                 1411 /  471
GAC GTC TCC GCC TCC ACC AAG AGC CTT CAG GAC TTG CTC TCG GCC CAC AGC TTG TCC CCT
asp val ser ala ser thr lys ser leu gln asp leu leu ser ala his ser leu ser pro
1441 /  481                                 1471 /  491
TGG GGG GCT GAG GTG AAG GCA GAG CCT GGT GAA GTG GTG GCC CCA AGA GGG AAG TCA GCT
trp gly ala glu val lys ala glu pro gly glu val val ala pro arg gly lys ser ala
1501 /  501                                 1531 /  511
GCA CCC TCC AAG AAG AGC AAG GGC TGC TTC AAG GAG GAA GGT GTC AAC AAA TCT GAA AAG
ala pro ser lys lys ser lys gly cys phe lys glu glu gly val asn lys ser glu lys
1561 /  521                                 1591 /  531
AGG ATG AAA TTA ACT CTG AAG GGA GGA GCA GCC GTT GAT CCT GAC TCT GGT CTG GAA CAC
arg met lys leu thr leu lys gly gly ala ala val asp pro asp ser gly leu glu his
```

```
1621 /  541                              1651 /  551
TCT GCA CAC GTC CTG GAG AAA GGT GGG AAG GTG TTC AGC GCC ACA CTT GGC CTG GTG GAC
ser ala his val leu glu lys gly gly lys val phe ser ala thr leu gly leu val asp
1681 /  561                              1711 /  571
ATT GTG AAA GGG ACG AAC TCC TAT TAC AAA CTG CAG CTT CTG GAG GAC GAC AAG GAG AGC
ile val lys gly thr asn ser tyr tyr lys leu gln leu leu glu asp asp lys glu ser
1741 /  581                              1771 /  591
AGG TAC TGG ATC TTC CGG TCC TGG GGC CGG CTG GGC ACA GTT ATC GGC AGT AAC AAA CTT
arg tyr trp ile phe arg ser trp gly arg leu gly thr val ile gly ser asn lys leu
1801 /  601                              1831 /  611
GAG CAG ATG CCC TCC AAA GAG GAA GCC GTT GAG CAA TTC ATG AAG CTG TAT GAA GAG AAG
glu gln met pro ser lys glu glu ala val glu gln phe met lys leu tyr glu glu lys
1861 /  621                              1891 /  631
ACT GGG AAT GCC TGG CAC TCG AAA AAC TTC ACA AAA TAT CCC AAG AAG TTC TAC CCT CTG
thr gly asn ala trp his ser lys asn phe thr lys tyr pro lys lys phe tyr pro leu
1921 /  641                              1951 /  651
GAG ATT GAC TAT GGC CAG GAC GAA GAG GCA GTA AAG AAG CTG ACG GTG AAG CCT GGC ACC
glu ile asp tyr gly gln asp glu glu ala val lys lys leu thr val lys pro gly thr
1981 /  661                              2011 /  671
AAG TCG AAG CTG CCG AAG CCA GTG CAG GAG CTC GTG GGG ATG ATC TTC GAC GTG GAC AGC
lys ser lys leu pro lys pro val gln glu leu val gly met ile phe asp val asp ser
2041 /  681                              2071 /  691
ATG AAA AAG GCC TTG GTG GAG TAC GAG ATT GAC CTC CAG AAG ATG CCC TTG GGG AAG CTG
met lys lys ala leu val glu tyr glu ile asp leu gln lys met pro leu gly lys leu
2101 /  701                              2131 /  711
AGC AGA AGG CAG ATC CAG GCC GCC TAC TCT ATC CTC AGC GAG GTC CAG CAG CCA GTG TCT
ser arg arg gln ile gln ala ala tyr ser ile leu ser glu val gln gln pro val ser
2161 /  721                              2191 /  731
CAA GGC AGC AGT GAA TCC CAG ATC CTA GAT CTC TCC AAT CGC TTC TAC ACT CTG ATC CCC
gln gly ser ser glu ser gln ile leu asp leu ser asn arg phe tyr thr leu ile pro
2221 /  741                              2251 /  751
CAT GAC TTT GGA ATG AAG AAG CCC CCA CTC CTG AAC AAC GCA GAC AGC GTG CAG GCC AAG
his asp phe gly met lys lys pro pro leu leu asn asn ala asp ser val gln ala lys
2281 /  761                              2311 /  771
GTG GAG ATG CTA GAC AAC CTC CTG GAC ATC GAG GTG GCC TAT AGT CTT CTC AGG GGT GGC
val glu met leu asp asn leu leu asp ile glu val ala tyr ser leu leu arg gly gly
2341 /  781                              2371 /  791
TCT GAC GAC AGC AGC AAG GAT CCC ATC GAC GTC AAC TAC GAG AAA CTC AAA ACT GAC ATT
ser asp asp ser ser lys asp pro ile asp val asn tyr glu lys leu lys thr asp ile
2401 /  801                              2431 /  811
AAG GTG GTT GAC AGA GAT TCT GAA GAG GCC GAG GTC ATC AGG AAG TAC GTG AAG AAC ACT
lys val val asp arg asp ser glu glu ala glu val ile arg lys tyr val lys asn thr
2461 /  821                              2491 /  831
CAT GCT ACC ACG CAC AAC GCC TAT GAC CTG GAA GTG ATC GAT ATC TTC AAG ATA GAG CGC
his ala thr thr his asn ala tyr asp leu glu val ile asp ile phe lys ile glu arg
2521 /  841                              2551 /  851
GAG GGG GAG AGC CAG CGC TAC AAG CCC TTC AGG CAG CTT CAC AAC CGG AGG CTG CTG TGG
glu gly glu ser gln arg tyr lys pro phe arg gln leu his asn arg arg leu leu trp
2581 /  861                              2611 /  871
CAC GGC TCC AGG ACC ACC AAC TTT GCT GGC ATC CTG TCG CAG GGT CTG CGG ATA GCC CCA
his gly ser arg thr thr asn phe ala gly ile leu ser gln gly leu arg ile ala pro
2641 /  881                              2671 /  891
CCT GAA GCG CCC GTG ACA GGC TAC ATG TTT GGG AAA GGG ATC TAC TTT GCC GAC ATG GTG
pro glu ala pro val thr gly tyr met phe gly lys gly ile tyr phe ala asp met val
2701 /  901                              2731 /  911
TCC AAA AGT GCA AAC TAC TGC CAC ACA TCT CAG GGA GAC CCG ATT GGC TTA ATA ATG CTG
ser lys ser ala asn tyr cys his thr ser gln gly asp pro ile gly leu ile met leu
2761 /  921                              2791 /  931
GGA GAG GTT GCC CTT GGA AAT ATG TAT GAA CTC AAG CAT GCT TCA CAT ATC AGC AAG TTA
gly glu val ala leu gly asn met tyr glu leu lys his ala ser his ile ser lys leu
2821 /  941                              2851 /  951
CCC AAG GGC AAG CAC AGT GTC AAA GGT TTG GGA AAA ACC ACC CCT GAC CCT TCG GCC AGC
pro lys gly lys his ser val lys gly leu gly lys thr thr pro asp pro ser ala ser
2881 /  961                              2911 /  971
ATC ACC CTG GAG GGT GTA GAG GTT CCA CTG GGA ACA GGG ATC CCA TCT GGT GTC AAC GAC
ile thr leu glu gly val glu val pro leu gly thr gly ile pro ser gly val asn asp
2941 /  981                              2971 /  991
ACT GCC CTG CTG TAT AAT GAG TAC ATT GTC TAC GAC ATT GCT CAG GTG AAT CTC AAA TAC
thr ala leu leu tyr asn glu tyr ile val tyr asp ile ala gln val asn leu lys tyr
3001 / 1001                              3031 / 1011
CTG CTG AAA CTC AAG TTC AAT TTT AAG ACA TCC CTG TGG
leu leu lys leu lys phe asn phe lys thr ser leu trp
```

Figure 3—Nucleotide Sequence of BXSB coding region—The BXSB cDNA sequence of pADP-3a begins at coding residue 1 and continues in an open reading frame to the internal EcoRI site at mouse residue 1012. The sequence extending 3′ to the poly A^{++} site is from pADP-C3. Only the coding region is shown for clarity. Sequences from −85bp 5′ of residue 1 and 3′ of residues 1012 can be found in EMBL/Genbank accession number X14206.

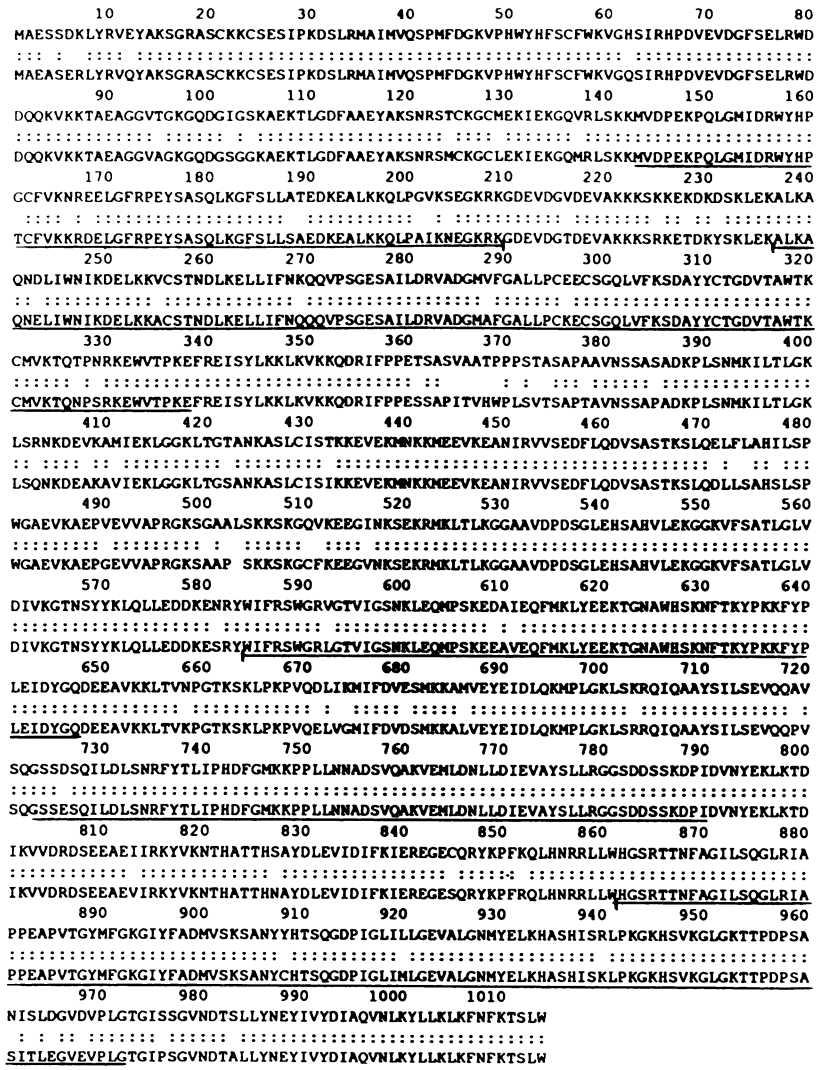


Figure 4—Derived Amino Acid Sequence Comparison of human (top) versus mouse (bottom) ADPRP cDNA and genomic clones. Direct sequence analysis between the human sequence derived from pCD (6) is scored as a match with the mouse sequences (:). Additional comparisons was made with genomic sequences from BALB/c genomic clones λ4 and λ7 and those regions are underlined Exon−intron boundaries identified by a genomic sequencing are indicated by a '+'. Only one amino acid residue was found to differ between BXSB and BALB/c at residue 724 (ser Gly). There 5 additional substitutions between BXSB and BALB/c which at the third base position of codons 174, 259, 271, 285, 972 and therefore are silent changes.

are 9% amino acid substitutions within this domain (Fig 5). These observations further support the role this region may play as recognition sites for DNA binding. Interestingly, a region between the DNA-binding fingers containing an unusual amount of glycine residues (6/12 residues) which may allow a considerable amount of flexibility between fingers was

Table I. Consensus nucleotide binding fold

"A"	Gly	Xaa_4	Gly	Lys	Gly	Xaa_6	Val	Xaa_{12-14}	Asp
HuADPRP (888-914)	Gly	TyrMetPhe	Gly	Lys	Gly	IleTyrPheAlaAspMet	Val		Asp
MoADPRP	Gly	TyrMetPHe	Gly	Lys	Gly	IleTyrPheAlaAspMet	Val		Asp
"B"	Arg	Xaa2-3	Gly	Xaa3 (hydrophobic) 4-6					Asp
HuADPRP (940-957)	Arg	LeuProLys	Gly	LysHisSerValLysGlyLeuGlyLysThrThrPro					Asp
MOADPRP	Lys	LeuProLys	Gly	LysHisSerValLysGlyLeuGlyLysThrThrPro					Asp

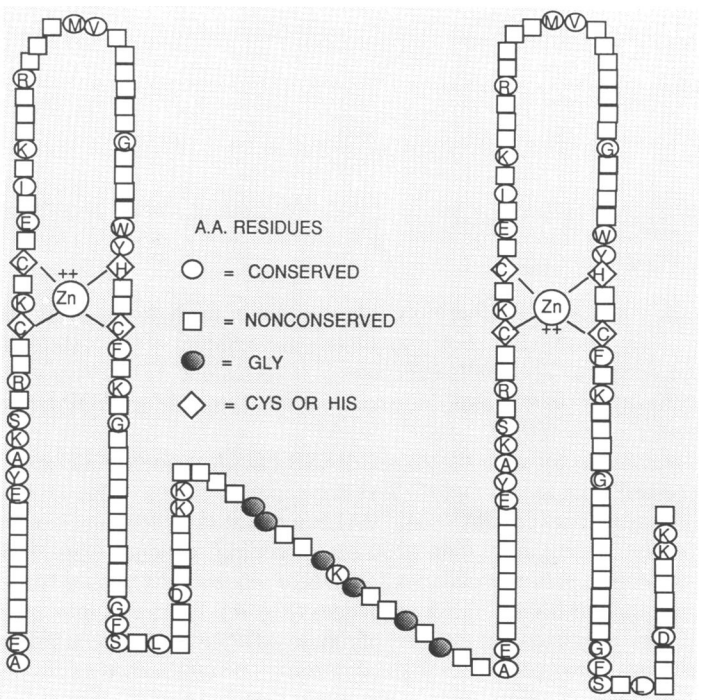


Figure 5—Schematic comparison of human and mouse amino acid residues in the putative Zn^{++} binding region of ADPRP. A computer generated ADPRP molecule shows the location of the Zn^{++} binding regions starting at residue 2 through highlights conserved or nonconserved amino acid substitutions by designating the mouse residue in the indicated box or circle. Glycine, Cysteine and Histidine residues which are conserved are also designated.

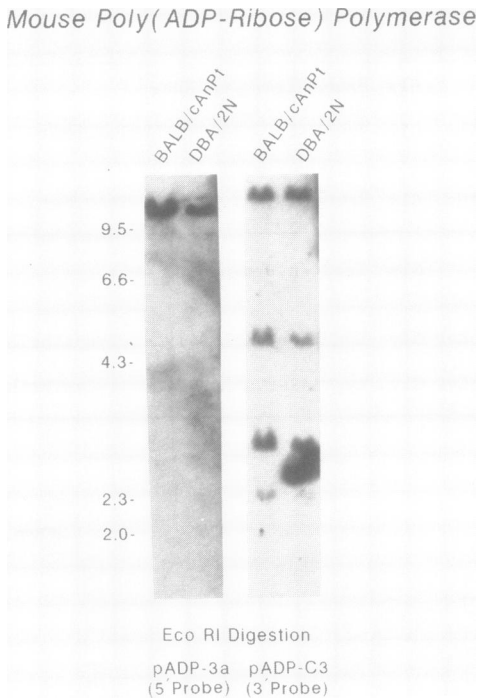


Figure 6—EcoR1 Southern hybridization analysis of mouse DNA with moADPRP. EcoRI restricted genomic DNA from BALB/cAnPt or DBA/2N was hybridized to 5′ (pADP-3a) or 3′ (pAPP-C3) molecular probes and washed under stringent wash conditions.

also found to be conserved. Furthermore, comparison of the human and mouse sequences in the region of the automodification domain (mouse residues 400−550) identified 15/16 conserved Glu residues. The single exception is (Glu residue 471 to Asp which is buried in a very strong hydrophobic area, not necessarily accessible to ADP-ribosylation.

Exon Boundaries of moADPRP

Comparison was made between the mouse BXSB cDNA sequence and exon nucleotide sequences obtained from the λ4 and λ7 BALB/c genomic clones. (Fig. 4). Coding region sequence comparisons between BXSB and BALB/c revealed 5 silent mutations at the codon third base position among more than 1200 bp. of coding sequence (Fig. 4) We are able to establish some intron−exon boundaries at points of sequence divergence between BALB/c genomic clones and the BXSB cDNA sequences (Fig. 4). From this information, we can determine a preliminary exon composition of mouse ADPRP as being comprised of perhaps 10−15 exons (each separated by 1−2 kb of intron). The established exons do not appear to interrupt functional domains such as the nucleotide binding folds at the A or B sites nor any of the Zn^{++} finger motifs. Thus, exon structure of ADPRP closely approximates the domain structure.

Chromosomal Localization of moADPRP

The human loci corresponding to ADPRP are polymorphic and genes associated with huADPRP map to three distinct chromosomes (6). Therefore, it is interesting to compare

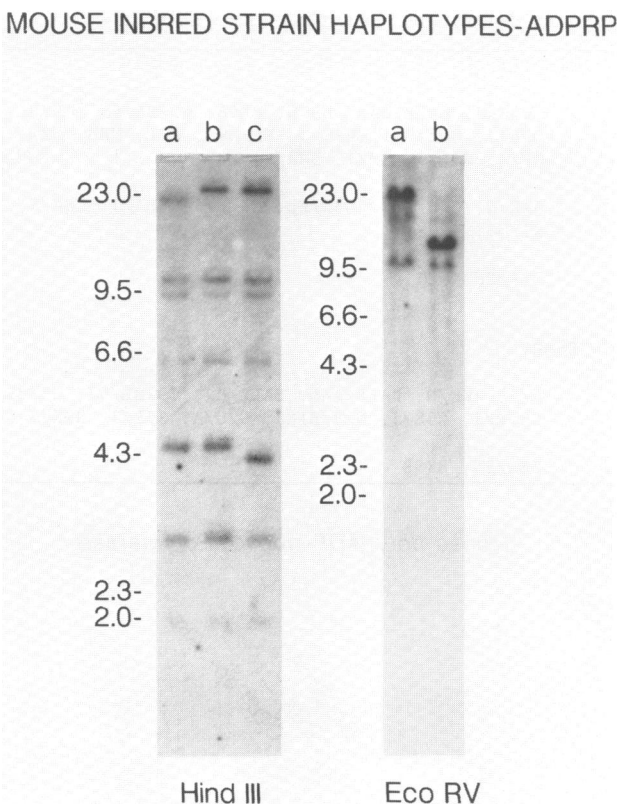


Figure 7—Mouse Inbred strain genotypes for ADPRP A HindIII (Left panel) or EcoRV (right panel) RFLP segregates most mouse inbred strains. Three distinct HindIII haplotyes and two EcoRV haplotyes are distinguishable with the probe pADP-C3 as listed in table II.

the abundance and corresponding chromosomal positions in the mouse. Hybridization of the pADP-3a 5′ probe to EcoR1 restricted genomic DNA from the BALB/c inbred mouse exhibits a single 10 kb RF band (Fig 6). The 3′ probe, pADP-C3, hybridizes to four EcoRI RF bands of 2.3 Kb, 3.0 Kb, 4.3 Kb and 12 kb in BALB/c (Fig. 6). Since the genomic clones λ4 and λ7 can account for the 10 kb and 12 kb bands, we suspect that the 3′ most exons are probably contained within the small EcoRI fragments. Direct comparison of two inbred mouse strains, DBA/2N and BALB/c, by hybridization with pADP3a or pADP-C3 revealed that among multiple enzymes only HindIII, or EcoRV are polymorphic. In an expanded survey of inbred mice, three distinct HindIII RF patterns and two distinct EcoRV patterns are found (Fig. 7 and Table II).

Originally, we utilized the Hind III RFLP between the inbred strains DBA/2N, BALB/cAn or C57BL/6J for the purpose of genetic mapping of moADPRP. Two Hind III RF bands of approximately 4.3kb and 23.0 kb differ between the DBA/2N haplotype (RF^c) and the BALB/c or C57BL/6J haplotype (RF^a). An extensive panel of BALB/c by DBA/2N (C.D2) congenics mice have been maintained in our laboratory by which the DBA/2N genes have been introgressively backcrossed onto the BALB/cAn background and selected

Table II. Mouse Inbred Genotypes - ADPRP

ADPRP alleles - HindIII	
a.	AU/SsJ, BALB/cAn, BSVS, BXSB, C57BL/6J, C57BL/10N, C57/BR, C58/J, Dwf/Dwf, HRS/J, MA/MyJ, MRL, NZB, NZW, PL/J, RIII/sJ, RF/J, SJL/J, SM/J, SWR/J
b.	AKR/N, C3H/HeJ, C3H/Fg, C57L/J, CBA/J, CBA/N, I, NH, 020/A, ST/bJ, 129/J
c.	DBA/2N
ADPRP alleles - EcoRV	
a.	A/J, AKR/N, BALB/cAn, BALB/cJ, C57BL/6J, C57BL/10N, C57L/J, C58/J, NZB/BlNJ, 020/A, SJL/J, SWR/J
b.	DBA/2N, BXSB

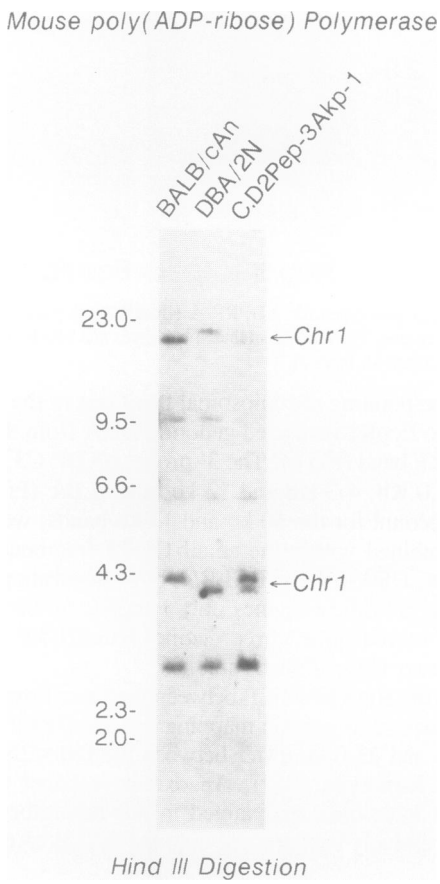


Table III. Segregation of ADPRP on Mouse Chromosome 1

A. BXD Recombinant Inbreds*+

	1	2	5	6	9	11	12	13	14	15	16	19	21	22	24	25	27	28	29	30	32
Pep-3	B	B	D	D	B	D	B	D	B	D	B	B	B	B	D	B	D	D	D	D	D
Lamb-2	B	B	D	-	D	D	B	D	B	B	B	D	B	D	D	B	D	D	D	D	B
At-3	B	B	D	-	D	D	B	D	B	B	B	B	B	D	D	D	D	D	D	D	D
ADPRP	B	B	D	B	D	D	B	B	B	B	B	B	B	D	D	D	D	D	D	D	D
FcR	B	D	D	-	D	D	B	B	B	B	B	B	B	D	D	D	D	D	B	D	D

*Data for Pep-3, Lamb-2, FcR and At-3 are from (16).
+B refers to C57BL/6J alleles, whereas D refers to DNA/2J alleles.

B. C.D2 Congenics

	BALB/c	DBA/2N	C.D2 Pep3 Akp-1
Pep-3	a	b	b
ADPRP	a	c	c
FcR	a	c	c

*FcR refers to Fcε RIα (14)

for markers on various mouse chromosomes. We screened this panel of congenics with the pADP-C3 probe looking for specific retention of the HindIII RF^c haplotype. A single congenic, C.D2 Pep3-Akp-1 (N6), specific to a region of mouse chromosome 1 between Pepsinogen (*Pep-3*) and *Alkaline Phosphatase-1* (*Akp-1*), was found to contain the Hind III RF^c component (Fig. 8). Since the C.D2 Pep-3/Akp-1 (N6) congenic has been determined to be restricted in DBA/2N genes exclusively to this region of mouse chromosome 1 including the genetic loci for Fc receptors of IgE and IgG (14, and unpublished, KH), we concluded that moADPRP is contained within this region.

An EcoRV RFLP between BALB/c and DBA/2 enabled us to reexamine the C.D2 Pep-3/Akp-1 congenic for the presence of the DBA/2 EcoRV RF^b pattern. Since there are only two EcoRV RF bands, and one Eco RV site falls within the moADPRP coding region, it is likely that the entire moADPRP gene is contained within both bands. As predicted, the C.D2 Pep-3 Akp-1 congenic displayed an EcoRV RF^b pattern (data not shown). Therefore, it is likely that the chromosome 1 structural gene is also the expressed moADPRP.

To confirm the positioning of moADPRP on chromosome 1, we tested a series of 26 recombinant inbred (RI) mice developed between C57BL/6J (RF^a) and DBA/2J (RF^b)

Figure 8—Mouse ADPRP maps to chromosome 1. The segregation of a DBA/2N RRF^c) allele in a C.D2 Pep-3/Akp-1 congenic reveals that the moADPRP 23.0 kb and 4.0 kb HindIII RF bands map to mouse chromosome 1. For the purpose of ongoing breeding, the C.D2 Pep-3/Akp-1 congenic is a heterozygote and carries one BALB/c and one DBA/2 chromosome 1 complement, hence, the congenic appears heterozygous for moADPRP. The 9.0 kb HindIII RF band is absent in comparison to Fig. 6. The reason is unclear but may be related to slight variation in wash stringency (see Discussion).

(BXD) (15) by following the genetic segregation distribution with an EcoRV RFLP (Table III). The BXD distribution of moADPRP is 100% concordant with the chromosome 1 loci Epoxide hydratase-1 (*Eph-1*) and *Mtv-7*, the integration site of mouse mammary tumor virus. Further examination of the segregation patterns of ADPRP indicates close linkage to the complement receptor-related gene *At-3* (19/20 concordancies) and the *Ly-17* locus (18/20 concordancies) (16). Using accepted statistical parameters for linkage analysis among RI lines, moADPRP lies within 1.3cM (.995) of *At-3* and 2.9cM (.972) of *Ly-17* (17). Examination of the BXD segregation patterns from Table III and attempting to minimize double crossover frequency suggests moADPRP is between these loci and may be close to the autoimmune locus *gld* (16).

DISCUSSION

ADPRP is activated by DNA strand breaks and thereby becomes critical in DNA repair, differentiation and regulation of transcription. Although seemingly multifaceted in its interactions, ADPRP probably serves a much more uniform role in cellular metabolism. Most likely, this role is a stabilizing effect utilizing two functional domains of ADPRP, DNA binding and NAD binding. ADPRP has been found in such divergent organisms as fish, birds and snails. Comparisons between human and mouse ADPRP support strict sequences conservation across species boundaries. Furthermore, we have been able to predict some regions of greater conservation between mouse and human which may ultimately aid in the definition of the various functional domains of ADPRP.

The positioning of the moADPRP locus to chromosome 1 suggests: i). that the expressed ADPRP gene lies within a syntenic unit in the region of human 1q23-q25(16) and ii). that the moADPRP and huADPRP are homologues. Furthermore, the establishment of inbred mouse haplotypes for ADPRP allows for more detailed examination of the segregation among inbred strains or between genetic crosses of inbred strains. This has permitted us to precisely position *moADPRP* between *At-3* and *FcR*, and near the mouse autoimmune locus *gld* (10). The close proximity of *ADPRP* and *gld* is provocative in view of the finding that excessive stimulation of the immune response which occurs in autoimmune reactions leads to a concomitant increase of ADPRP activity (18). Excessive ADPribosylation causes hyperutilization of NAD and consumes ATP, which in turn leads to energy poor cells. Disruption of the normal interaction of ADPRP with its target has been suggested as a causative agent leading to transformation and/or autoimmunity. In fact, several forms of lymphoid neoplasia display defective DNA repair as well as immunological abnormalities (i.e. autoimmunity). Moreover, the involvement of ADP-ribosylation in the integration of foreign DNA into Eucaryotic cells suggests ADPRP is associated with DNA recombination events (19). The genetic distribution of moADPRP haplotypes shows an interesting segregation with autoimmune strains such as BXSB, C3H/HeJ, AKR, CBA/N, C57L/J and 129/J (20). We are presently investigating whether different ADPRP alleles encode structurally distinct proteins or whether regulatory differences in expression of ADPRP may result in changes of expression patterns of ADPRP. More directly, expression of ADPRP in MRL/gld, C57BL/6J/gld or C3H/HeJ gld mice carrying the autosomal recessive trait needs to be examined in order to unambiguously derive any relationship between expressed ADPRP and gld. Preliminary evidence from the additional 6 cDNA clones isolated in this laboratory suggest ADPRP-like domains may also be associated with other related genes. For example ADPRP-like sequences lie on human chromosomes 13

and 14; in addition to the expressed gene on human chromosome 1. However, it has not been determined whether these sequences represent pseudogenes or fragments of genes relating to DNA or NAD binding domains. Preliminary evidence reveals that a 9.0 kb HindIII RF band (see Fig. 8 legend) identified by pADP-C3 may map to the distal end of mouse chromosome 4 (data not shown). The distal end of mouse chromosome 4 encodes both a tumor suppressor locus(21) as well as DNA repair genes (8,9), which could conceivably encode domains related to DNA binding. This region may be analogous to a region on human chromosome 13 which carries chromosomal deletions of associated regions of ADPRP-like sequences frequently associated with some hematopoietic malignancies such as Burkitt Lymphoma (K.B. manuscript in preparation).

ACKNOWLEDGEMENT

We wish to acknowledge the expert secretarial assistance of V. Rogers, M. Millison and J. Coopersmith in the preparation of this manuscript. We appreciate the efforts of J. Owens in the computer setup and analysis.

REFERENCES

1. Smulson, M.E. and Sugimura, T. eds. (1980) Novel ADP-Ribosylations of Regulatory Enzymes and Proteins (Elsevier, New York).
2. Benjamin, R.C. and Gill, D.M. (1980) J. Biol. Chem. 255, 10502–10508.
3. Carson, D.A., Seto, S. and Wasson, D.B. (1986) J. Exp. Med. 163:746–751.
4. Schweiger, M., Awer, B., Burtscher, H.J., Hirsch-Kaufmann, M., Klocker, H., and R. Schneider (1986) Biol. Chem. Hoppe-Seyler 367, 1185–1195.
5. Cherney, B.W., McBride, O.W., Chen, D., Alkhatib, H., Bhatia, K., Hensley, P. and M.E. Smulson (1987) Proc. Natl. Acad. Sci. USA. 84; 8370–8374.
6. Kurosaki, T., Ushiro, H., Mitsuuchi, N. Kangawa, K., Matsuo, H., Hirose, T., Inayama, S., and Shizuta, Y. (1987) J. Biol. Chem. 262: 15990–15997.
7. Uchida, K., Morita, T., Sato, T., Ogura, T., Yamashita, R., Noguchi, S., Suzuki, H., Nyunaya, H. Miwa, M. and Sugimura, T. (1987). Biochem, Biochem, Biophys Res. Commun. 148 617–622.
8. Lin, P.F. and Ruddle, F.H. (1981) Nature (London) 289;191–194.
9. Potter, M., Sanford, K.K., Parshad, R., Tarone, R.E., Price, F.M., Mock, B. and Huppi, K. (1988) Genomics 2; 257–262.
10. Roths, J.B., Murphy, E.D. and Eicher, E.M. (1984) J. Exp. Med. 159; 1–20.
11. Huppi, K., Duncan, R., and Potter, M. (1988) Immunogenetics 27:215–219.
12. Mushinski, J.F., Potter, M., Bauer, S.R., and Reddy, E.P. (1983) Science 220:795–798.
13. Berg, J.M. (1988) Proc. Natl. Acad. Sci. (USA). 85;99–102.
14. Huppi, K., Mock, B.A., Hilgers, J., Kochan, J. and Kinet, J-P. (1988) J. Immunol. 141;2807–2810.
15. Taylor, B.A. (1980) Origins of Inbred Strains, H.C. Morse (ed.), pp. 423–428, Academic Press, New York.
16. Seldin, M.F. Morse, H.C., Leboeuf, R.C., and Steinberg, A.D. (1988) Genomics 2, 48–56.
17. Silver, J., and Buckler, C.E. (1986) Proc. Natl. Acad. Sci. (USA) 83:1423–1427.
18. Yamamaka, H., Penning, C.A., Willis, E.H., Wasson, D.B. and Carson, D.A. (1988) J. Biol. Chem. 263, 3879–3883.
19. Farzaneh, F., Panayotou, G.N., Bowler, L.D., Hardas, B.D., Broom, T., Walther, C., and Shall, S. (1988) Nucl. Acids. Res. 16; 11319–11326.
20. Shultz, L.D. and Sidman, C.L. (1987) Ann Rev. Immunol. 367–403.
21. Fenyo, E.M., Klein, G., Povey, S., Jeremiah, S.J., Spira, J., Wiener,F., and Harris, H. (1980) Euro. J. Cancer 16; 357–368.
22. Hansen, M.F. and Cavenee, W.K. (1988) Trends in Genetics 4, 125–128.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.