The Integrons In0, In2, and In5 Are Defective Transposon Derivatives

HEIDI J. BROWN,^{1,2} H. W. STOKES,¹ AND RUTH M. HALL^{2*}

School of Biological Sciences, Macquarie University, New South Wales 2109,¹ and Sydney Laboratory, CSIRO Division of Biomolecular Engineering, North Ryde, New South Wales 2113,² Australia

Received 5 February 1996/Accepted 24 May 1996

The class 1 integrons In0, In2, and In5, found in different locations in pVS1, Tn21, and pSCH884, have closely related structures. All three integrons contain an insertion sequence, IS1326, that is a new member of the IS21 family. IS1326 has caused deletions of adjacent 3'-conserved segment and transposition module sequences, and all three integrons retain a complete copy of only one of four genes required for transposition of related transposons and are thus defective transposon derivatives. In2 contains an additional insertion sequence, IS1353, located within IS1326. IS1353 is a member of the IS3 family and appears to have been acquired after the integron was inserted into an ancestral mercury resistance transposon to create the ancestor of Tn21 and several other transposons that are close relatives of Tn21.

Many of the antibiotic resistance genes found in clinical isolates of gram-negative bacteria are known to be located in integrons (reviewed in references 12, 14, 15, and 31). The resistance genes are contained in discrete mobile elements known as gene cassettes and are integrated at a specific site, *attI*, in the integron. The integron encodes a site-specific recombinase that belongs to the lambda integrase family, is responsible for the insertion of cassettes at *attI*, and also provides the promoter responsible for expression of the cassette-encoded antibiotic resistance genes.

Three classes of integrons have been identified (see reference 31). However, the integrons most commonly isolated from antibiotic-resistant clinical isolates of members of the family Enterobacteriaceae and pseudomonads belong to class 1 and most of them contain the sull sulfonamide resistance gene. These integrons appear to be mobile elements, as they are found in many distinct locations, in plasmids such as R46 and R388 and within transposons such as Tn21 and Tn1696 (16, 43). However, early studies aimed at identifying the sequences common to all integrons failed to identify features consistent with the classification of integrons as transposons (43). Two conserved regions located on either side of the integrated gene cassettes were identified. The 5'-conserved segment (5'-CS) includes a gene, intI1, encoding the integrase; a gene attI, encoding the cassette integration site; and the promoter, Pant, that is responsible for expression of the cassette genes (13, 16, 32, 43). The 3'-CS defined by Stokes and Hall (43) included the sull sulfonamide resistance determinant (43, 44) and a further two open reading frames (ORFs), one of which has recently been identified as a defective version of the quaternary ammonium compound resistance determinant qacE and designated $qacE\Delta 1$ (26).

Subsequently, a more extensive study of the sequences that were common to six integrons from independent locations revealed that all contained the same 5'-CS but the 3'-CS extended different distances beyond the *sulI* gene (13). It was concluded that a variety of events, such as deletions and insertions, that removed part of or interrupted the 3'-CS had occurred to give rise to the integrons studied. The previous 3'-CS endpoint (43) was found to represent one endpoint of a deletion in one of the two integrons (In1 and In2) used initially to define the 3'-CS, and the sequences assigned to the 3'-CS were extended to include a further 359 bp (13). In three cases, In2 from Tn21, In0 from pVS1, and In5 from pSCH884, two further common regions were found. The first directly abuts the 3'-CS endpoint and extends at least 0.7 kb. The second is located 4 to 6 kb away, and an outer end that terminates with a 25-bp sequence that is an inverted repeat of the sequence found at the outer end of the 5'-CS was identified (13). This 25-bp inverted repeat had previously been found at the boundaries of the 11.2-kb insertion in Tn21, identified by comparison of the sequences of Tn501 and Tn21 (2), and this insertion corresponds to In2. This finding is consistent with the identification of integrons as transposons; however, previous attempts to demonstrate movement of In2 were unsuccessful (see reference 11), suggesting that In2 is a defective transposon derivative.

Recently, the sequence of Tn402 (Tn5090), which is a class 1 integron that does not include the sull gene, has been reported (30). Tn402 was first identified as a transposon present in IncP plasmid R751 (39). Tn402 is bounded by the same 25-bp inverted repeats and includes a 5'-CS identical to that found in the sull-containing integrons and two integrated cassettes, dfrB3 (dhfrIIc) and orfD. Tn402 also includes the complete qacE gene (26) and a 4-kb region that contains four ORFs, three of which were identified as encoding potential transposition functions, on the basis of similarities to known transposition proteins (30). Close relatives of these four genes are also found in the transposition module of mercury resistance transposon Tn5053, and it has been shown that three of the genes tniA, tniB, and tniQ, are essential for transposition and the fourth, *tniR*, is required for cointegrate resolution (18). Partial sequencing of In2 revealed that a 2.7-kb segment of the Tn402 transposition module is also present in In2 from Tn21, but only one complete gene (tniA) and a truncated form of a second gene (tniB) are present in this fragment (30). Thus, it appears that In2 may be a defective transposon that does not contain a complete set of transposition functions.

In In2, the 3'-CS sequences (13, 43) and the transposition module segment (30) are separated by a region of about 4 kb, and short inverted repeats (27 bp) have been identified at or close to the boundaries of this segment in In2 (45). The se-

^{*} Corresponding author. Mailing address: CSIRO Division of Biomolecular Engineering, Sydney Laboratory, P.O. box 184, North Ryde NSW 2113, Australia. Phone: 61-2-8864961. Fax: 61-2-8864805. Electronic mail address: hall@pelican.dbe.csiro.au.

Plasmid	Description	Relevant phenotype(s) ^a	Reference	
pVS1	Contains In0 and Tn501	Su ^r Hg ^r	42	
pUB2401	pACYC184::Tn21, contains In2	Cm ^r Sm ^r Sp ^r Su ^r Hg ^r	8	
pSCH884	Contains In5	Gm ^r Tb ^r Su ^r	40	
pRMH467	2.06-kp EcoRI fragment of pVS1 in pUC19	Ap ^r	This work	
pRMH473	6.8-kb HindIII fragment of pVS1 in pUC19	Apr	This work	
pRMH451	1,643-bp EcoRI fragment of Tn21 in pUC19	Apr	This work	
pRMH456	1,847-bp EcoRI fragment of Tn21 in pUC19	Apr	This work	
pRMH458	661-bp SphI fragment of Tn21 in pUC19	Apr	This work	
pRMH474	4.2-kb EcoRI fragment of Tn21 in pUC19	Apr	This work	
pRMH475	1,573-bp SphI fragment of Tn21 in pUC19	Apr	This work	
pRMH477	4.6-kb <i>Eco</i> RI fragment of pSCH884 in pUC19	Apr	This work	

TABLE 1.	Plasmids	used in	this study
----------	----------	---------	------------

^a Ap, ampicillin; Cm, chloramphenicol; Gm, gentamicin; Hg, mercury; Sm, streptomycin; Sp, spectinomycin; Su, sulfonamide; Tb, tobramycin.

quence of the left-hand end of this region has been determined (13), and identical sequences were found in In0 and In5, suggesting that In2, In0, and In5 share a common lineage. However, in In0 and In2, a deletion of 359 bases of the 3'-CS that are present in In5 has occurred (13). Together, these findings suggest that a further transposon or insertion sequence may be present in In2, In0, and In5.

To clarify the relationships between the structures found in different integrons, we examined the sequence of the region between the 3'-CS and transposition module regions in In2, In0, and In5. In all three integrons, an insertion sequence, IS1326, was found to be present, and in In2, a further insertion sequence, IS1353, inserted just inside one of the terminal inverted repeats of IS1326 was found. The regions adjacent to IS1326 were identical in In2 and In0, indicating that these two integrons have a recent common origin. In5 includes extra 3'-CS sequences to the left of IS1326 and has 439 bp less of the transposition module to the right, indicating that IS1326 has caused deletions of adjacent sequences. Possible routes for the evolution of the In2, In0, and In5 group from an integron with a backbone structure similar to that of Tn402 are discussed, and a coherent tree for the evolution of the Tn21 transposon family is presented.

MATERIALS AND METHODS

Bacterial strains and plasmids. Escherichia coli JM109 [($\Delta lac-proAB$) supE thi F' (traD36 proAB⁺ lacI^q lacZ Δ M15)] was used to propagate plasmid DNA. The plasmids used are listed in Table 1. The pRMH series of plasmids listed were constructed by randomly cloning fragments from *EcoRI*, *Hind*III, and *SphI* digests of the relevant parental plasmid (pVS1, Tn21, and pSCH884) into pUC19 (52) by using standard procedures (37). Plasmids containing the appropriate fragment were identified by restriction mapping and by sequencing of the fragment ends with a universal primer. Cells were grown in L broth (37) at 37°C if plasmid free and in L broth supplemented with ampicillin (Sigma Chemical Co., St. Louis, Mo.) at 100 µg/ml if plasmids were present.

DNA isolation and restriction mapping. Plasmid DNA was recovered from 2-ml samples of overnight cultures by the alkaline lysis method (37). Plasmid DNA for sequencing was purified by using the Magic Minipreps DNA Purification System (Promega Corp., Madison, Wis.). Restriction enzymes were used in accordance with the manufacturers' instructions. Fragments were separated by electrophoresis on 1% (wt/vol) agarose gels and visualized by staining with ethidium bromide. The size standard used was an *Eco*RI digest of bacteriophage SPP1 (Bresatec, Adelaide, South Australia, Australia).

DNA sequencing. Nucleotide sequencing was performed with a Sequenase 2.0 system (46) as recommended by the manufacturer (U.S. Biochemicals, Cleveland, Ohio) by using dITP reaction mixtures followed by a 30-min incubation with a 1 mM deoxynucleoside triphosphate mixture and terminal deoxynucleotidyl transferase (Boehringer GmbH, Mannheim, Germany). The specific primers used were 17-mers synthesized with an Applied Biosystems 380A DNA synthesizer.

Sequence analysis. Sequence data were collated, assembled, and interpreted by using Staden programs BAP and NIP (7). Database searches (GenBank, February 1996 version) for sequence similarity were performed by using the Match and Match-Translate programs (50). The sequences reported were compiled with previously published contiguous sequences.

Nucleotide sequence accession numbers. The GenBank accession numbers of

the sequences determined in this study are as follows: In0, U49101; In2, U42226; In5, U38230; IS1326, U38187; IS1353, U40482.

RESULTS

Maps of the region between the 3'-CS and the integron boundary differ. In a previous study, examination of the sizes of PCR products revealed significant differences in the length of the region between the *Eco*RI site that marks the end of available sequences (Fig. 1) and the right-hand outer boundary of three integrons, In0, In2, and In5 (13). To resolve the structures of these three integrons, restriction fragments from this region, generated by digestion with HindIII, EcoRI, or SphI, were cloned and mapped. By using these data, together with available maps of In0 (17) and In2 (25, 45), detailed maps were generated (Fig. 1). All three integrons include similarly located HindIII and SphI sites immediately to the right of the EcoRI site at the end of the sequenced region, and thereafter the maps diverge. Although in In0 and In2, further EcoRI sites are present (one in In0 and three in In2), a single 4.6-kb EcoRI fragment from pSCH884 includes the remainder of In5.

Insertion sequence IS1326 has caused adjacent deletions in In0 and In5. The regions shown in Fig. 1 were sequenced, and the sequence of In0 extending from the last BamHI site in the 3'-CS is shown in Fig. 2. The sequence of In5 (accession no. U38230) is identical to this sequence, except that a 439-bp region (boxed in Fig. 2) is not present in In5 and a further region of 359 bp is present in In5 (13) at the position indicated in Fig. 2. A simple explanation for these findings is that In0 and In5 have arisen from a common ancestor by deletion of sequences to the left or right of bases 152 and 2625, respectively (Fig. 2). Careful examination of the region between these bases revealed features consistent with its identification as an insertion sequence, which was designated IS1326. IS1326 is 2,470 bp long, is bounded by inverted repeats of 26 bp with 22 of the 26 bp identical, and includes two ORFs (see below for details). The apparent deletion in In0 relative to In5 originates precisely at the left boundary of IS1326, and the apparent deletion in In5 relative to In0 originates precisely at the right boundary of IS1326 (Fig. 3). This suggests that In0 and In5 have arisen from a common ancestor by IS1326-mediated deletion of sequences adjacent to the IS1326 insertion in that ancestor.

The sequence to the right of IS1326 in In0 (bases 2626 to 3110 in Fig. 2) is identical to bases 2891 to 2406 in Tn402 (30), which is also a class 1 integron, and the sequence identity extends to and includes the 25-bp repeat at the right-hand terminus of In0 (13) and Tn402. This 2.7-kb region lies within the transposition module of Tn402, which contains the ORFs encoding transposition functions (18, 30). Only 2.26 kb of the transposition module is present in In5. As these regions in-



FIG. 1. Restriction maps of *sull*-type integrons In0, In2, and In5. Arrowheads indicate the 25-bp inverted repeats at the outer boundaries, and the vertical arrows indicate the point at which the In0 and In2 sequences diverge from that of In5. Solid lines below the maps indicate the extent of published sequences collated from the following sources: 1, Bissonnette and Roy (1); 2, Hall et al. (13); 3, Diver et al. (9); 4, Stokes and Hall (43); 5, Sundström et al. (44); 6, Brown et al. (2); 7, Shaw et al. (40). Dashed lines show the extent of sequences determined in this study. Restriction enzyme sites: B, *Bam*HI; Bg, *Bg*II; E, *Eco*RI; H, *Hind*III; Ps, *Pst*I; S, *SaII*; Sm, *SmaI*; Sp, *SphI*.

clude only the complete *tniA* gene and part of the *tniB* gene, it seems reasonable to conclude that the ancestor of In0 and In5 contained the complete transposition module, part of which was subsequently lost as the result of one or more IS1326-mediated deletion events.

Insertion sequence IS1326 is a member of the IS21 family. IS1326 includes two long ORFs that overlap by 13 bp (Fig. 2). The polypeptides predicted by these ORFs were found to be related to the two polypeptides encoded by the *istA* and *istB* genes of IS21 (34) and to the corresponding polypeptides in three further known relatives of IS21, namely, IS232 (23), IS5376 (51), and IS100 (27). On the basis of the similarity of the IS1326 ORFs to IstA and IstB from IS21, which have been shown to be required for transposition (33), the IS1326 genes are also designated istA and istB. A detailed description of the features common to members of the IS21 family, including IS1326, will be presented elsewhere (2a). Briefly, IstA includes an unusual D, D(45), E transposase motif and a helix-turnhelix motif and IstB has a distinctive set of NTP-binding domains. Inverted repeats of 27 bp are present at the boundaries of IS1326 in In0, In2, and In5, as has been noted previously for In2 (45). However, as many of the mobile elements that encode a transposase with a D, D(35), E motif are bounded by 5'-TG and CA-3' (10, 28), we have assumed that IS1326 is bounded by the same dinucleotides and that the inverted repeats are most likely to be 26 bp long.

In2 is closely related to In0 but contains IS1353 inserted within IS1326. The restriction map of In2 includes an extra segment not found in In0 and In5 (see Fig. 1). As this region has previously been shown to be absent from transposons such as Tn2411, which are otherwise very close relatives of Tn21(20), the possibility that this extra segment corresponds to an insertion sequence was examined. The sequence of In2 from the *Eco*RI site in IS1326 to the outer, right-hand end of In2 was determined (GenBank accession number U42226), and this completes the sequences of In2 and Tn21. In2 contains a complete copy of IS1326, but a further new insertion sequence, designated IS1353, has been inserted just within the one end (right-hand end as shown in Fig. 3) of IS1326. The part of the In2 sequence containing IS1353 is shown in Fig. 4. IS1353 is 1,613 bp long and is bounded by inverted repeats of 12 and 13 bp (boxed in Fig. 4). Comparison of the In2 sequence with that of In0 (Fig. 2) revealed that IS1353 is flanked by a 2-bp direct duplication of the target sequence (indicated in Fig. 2). IS1353 is a member of the IS3 family (see below). Beyond the end of IS1326, the In2 sequence is identical to that of In0, and identity to the sequence of part of the transposition module of Tn402 extends up to the right-hand end of In2, as suggested by Råd-ström et al. (30), who sequenced several parts of this region. Thus, In2 is identical to In0, except that it includes an integrated aadA1 gene cassette, that determines resistance to streptomycin and spectinomycin, and IS1353.

IS1353 is a member of the IS3 family. Analysis of the IS1353 ORFs longer than 50 amino acids revealed that on one strand there were two ORFs, orfA of 222 codons and orfB of 304 codons, spanning almost the entire insertion sequence (Fig. 4). Several ORFs were also identified on the other strand. Only ORFB was found to be related to known sequences. ORFB shares 20 to 37% identity with the ORFB products of IS3, IS911, and the other 32 known members of the IS3 family (10, 21, 28, 49), and several conserved regions found in the transposases of IS3 family members are shared by ORFB. In particular, the acidic residue motif D, D(35), E, which was originally identified in the IS3 family (10), is present in ORFB. This motif is also present in other bacterial transposases, including those of Tn552 (36) and the IS4 family (35), as well as in the integrases of many retrotransposons and retroviruses (10, 22), and is believed to be part of the catalytic domain of these transposases and integrases (see reference 38).

IS1353 also exhibits an organization similar to that of members of the IS3 family. Most members of the IS3 family are bounded by 5'-TG and CA-3' (10, 28), and IS1353 shares this feature. Members of this group generally code for two ORFs, one short ORF (orfA) that overlaps a larger downstream ORF (orfB), which is in phase -1 with respect to ORFA. Expression studies with IS911 (29) and IS150 (48) have shown that these

· ·	• •		· ·	
GGATCCAACCTTCATCGCAGAACC	SCCTAGAACTCGAAATCTAA	CGTCCGTTCGGGCATCGAGGTC	CATGTCGGGGTGGGACGGGCCCGT	GCTTCAAGATCACTTGCAGTCCGACCGCGA
	EcoRV	Smal.	. 200	· · · · · ·
TGTCTTGGTTGCGCGAGAGGTTG	ICGATATCTGPTGATTTGCA	CLCAAATTTGACCCGGGATTTG	CATTGAATTTTGACCCACCCCTTG	T E I K L K
250	1	300	R2	350
230 . CGGGTCTGTTTTTCCTCCTGCTT/	ATTCTGAGTTGAACTGTGTT	JUU TGAAGCGGTAACTTTCATTGCC	GG TT TCCAGGATGTGGCAGTGGTG	350 GGTTAGTCGGTCCAACAACGCTGTTGTCATC
RTQKEEQK	NQTSSH	KFRYSENG	тецінснн	TLRDLLATTM
	. 400		45	
TTTTCATCGCCAAACACTCGGCTC	CATTCCGAGAAGCTCAAGI	TGGTGGTCAGTATCACGCTGGT	TTTTTCGTACAGCTTTGAGAGCAG	TGAAACAGCAGTGCCCCACCGGTTTGGCTA
KEDGFVK5	w 6 3 F 3 U	N I I D I V S I		R F L L A G G T Q S
. 500 AAAGGCAAATATCCCAGCTCATCC	AGAATCACCAAATCGCAT	ACAACAGACGGTTTCCGATTTC	550 . TCCCTGACGCCCAGATCATTTCTC	
FPLYGLED	LIVLDA	YLLRNAIQ	GQRGSSKE	Q E L A N V L D V T
		650		700
GAGAAGAAACGCACCCGTCGGTTC	CAAGTGCATCACTGCTTGTG	TACCAATGGCTGTGGCCAGGTG	AGTCTTGCCTGTGCCTGGCCCACC	ATCAGCACCACGTTCTGGGCTTGTTCCATG
SFFRVRRN	LHMVAQ	тдіатасн	ткстсрсс	ILVVNQAQEM
	750 .		. 800	
AAGTCGCACCGGTGCAATTGTTTC	JACCGTGGCCTCATTAACCA	GGCTTTGACTGAAGTCAAAGCC	CACCAAGTCCCGATACACGGGGAA	TTGGCCACCCGCAATTGATAGTTCACCGAA KAVRLOYNVS
850 . CGTACTTCACGCTCTGCCACTTCA	AGCTTTAATCAAGCTGTCCA	. 900 GCATGGGCAAGGCTTGATTAAA	. ECORI . TGCTGGTGAATTCTGATTGCCCAA	950 TCCTCAATGGCGTGTGCCATGCCAAAGAGT
RVEREAVE	AKILSD	LMPLAQNF	A P S N Q N G L	EEIAHAMGFL
	. 1000	.rbs	:	
TTCAAGGATTTGAGGATTCTCAC	ATGGCCTTCATGCTGCATCA	TGGGCTCTCCTTAAACTGTCAT	AGCGGTTCACGTTGGCCTGTGGTT	CAATGTCAGCCTTAACCCCTTGGGAATTGG
K L S K L I R V	* A A D H G E H Q M	HARRLSD' MIstB	ΥΚΝΥΝΑΟΡΙ	S L T L R L G K P I P
•			1150	
. 1100 AATCGGTTTGGGTGGAGGTTCTTC	 IGGTCAAACGTCCCAACAGA	TTAAGCACATGCTCCTTCGATG	LLDU GCTTGCCACACTCCAATGCCAATTC	1200 CCACAGCACTGAGTACCGCACCTTCATCGTG
IPKPPPEE	ETLRGLL	NLVHEKS	PKGCELALI	EVASLVAGEDH
		1250 .	HindIII .	1300
GTGCAATACAAGGGCCAGAATTTC	CACCATGTCACGGTCACCG	CCGGGGCGTTGCAGCAAGATGG	ATTGAAGCTTCTTGAACGCGGGTG	CAATTCAGCAAATGGCGCACCATTGCGCAA
HLVLALIE	s v M D R D G	GPKQLLI	5 Q L К К F А Р I	' LEAFPAGNRL
	1350 .		. 1400	
A G P K K O L V	LAGACAAGTAATGGTGCCAG V S L Y H H W	D Y Q T H G R I	GAGCGTGGCCACTGCCAAACAATC R A H G S G F L I	TIGGATGCTUGGCAATGTGTTGGCCTTCGGC R P H E A I H Q G E A
AGCCATCACCAGCTTGTCTGCATA A M V L K D A)	AAATCCGAAGGCTGATGGCC Y I R L S I A . 1600	CTGTTGGCGTAACTGGCAGGAA RNAYSAP	CGCTGTAGCGATTGCCCTCGTGGT V S Y R N G E H I	BACAAGGCAGGTTGAAGTGACTCGCTTGGT H V L C T S T V R K T
TTGCTCCACGAATGCATCAAAGGC Q E V F A D F A	CATTGGGTAGCGCCATCAAC A N P L A M L	TCGCCTTGTTCATCGGCAAAGG	CCTCTTGCACGGTTTGGTCCAATT A E Q V T Q D L J	CGGGGTGGCGCAGCTCAGACCACAGCGCTTT E P H R L E S W L A K
TTGCTCCACGAATGCATCAAAGGC Q E V F A D F A	CATTGGGTAGCGCCATCAAC A N P L A M L	TCGCCTTGTTCATCGGCAAAGG E G Q E D A F	CCTCTTGCACGGTTTGGTCCAATT A E Q V T Q D L J	GGGGTGGCGCAGCTCAGACCACAGCGCTTT E P H R L E S W L A K
TTGCTCCACGAATGCATCAAAGGC Q E V F A D F Z . 1700 GCAGCGATGCTCAAGCCACACAT	CATTGGGTAGCGCCATCAAC A N P L A M L	TCGCCTTGTTCATCGGCAAAGG E G Q E D A F AAGTCTGGTGCCCCTTGCCACA	CCTCTTGCACGGTTTGGTCCAATT A E Q V T Q D L J 1750 . GGCGTTGGCGGGAATCCTGCACGT	COGGGTGGCCAGCTCAGACCACAGCGCTTT PHRLESWLAK 1800
TTGCTCCACGAATGCATCAAAGGC Q E V F A D F 2 . 1700 GCAGCGATGCTCAAGGCCACACATT C R H E L W V N	CATTGGGTAGCGCCATCAAC A N P L A M L	TCGCCTTGTTCATCGGCAAAGG E G Q E D A F : AAGTCTGGTGCCCCTTGCCACA F D P A G Q W :	CCTCTTGCACGGTTTGGTCCAATT A E Q V T Q D L J 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V 1	CEGGETGGCGCAGCTCAGACCACAGCGCTTT P H R L E S W L A K 1800 ICTTCTCAATCTGGCCTTTCTCCCAACCCGA N K E I Q G K E W G S
TTGCTCCACGAATGCATCAAAGGC Q E V F A D F 2 1700 GCAGCGATGCTCAAGCCACACATT C R H E L W V N	CATTOGGTAGCGCCATCAAC A N P L A M L	TEGECTTOTTCATEGGCAAAGG E G Q E D A F . AAGTETGGTGCECCETTGCEACA F D P A G Q W 1 1850	CCTCTTGCACGGTTTGGTCCAATT A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V 1	CGGGGTGGCGCAGCTCAGACCACAGCGCTTT P H R L E S W L A K 1800 ICTTCTCAATCTGGCCTTTCTCCCAACCCGA N K E I Q G K E W G S 1900
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P 2 1700 GCAGCGATGCTCAAGCCACACAT C R H E L W V N TGCTGGATTACAGAACTGCGCATC	CATTOGGTAGCGCCATCAAC A N P L A M L ICAAATCAGCAAGGCTITGA N L D A L S Q CAAACAGGTAGTGGCTGGCCGCCAC	TEGECTTOTTCATEGGCAAAGG E G Q E D A F . AAGTETGGTGCECECTTGECACA F D P A G Q W 1 1850 ATGGCAGTGAACEGGTGATTGA	CCTCTTGCACGGTTTGGTCCAATT A E Q V T Q D L J 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V I CCCTGCGCTCTTTGCCACGCCCCA	CGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P 7 1700 GCAGCGATGCTCAAGCCCACACAT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I	CATTOGGTAGCGCCATCAAC A N P L A M L CCAAATCAGCAAGGCTTTGA N L D A L S Q CAAACAGGTAGTGGCTGACC D P L Y H S V	TEGECTTGTTCATCGGCAAAGG E G Q E D A F AAGTCTGGTGCCCCTTGCCACA F D P A G Q W 1 1850 ATGGCATGAACCGCTGATTGA M A T F R Q N	CCTCTTGCACGGTTTGGTCCAATT A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V N CCCTGCGCTCTTTGCCACGCCCCA V R R E X G R G V	CGGGGTGGCGCAGCTCAGACCACAGCGCTTT P H R L E S W L A K 1800 ICTTCTCAATCTGGCCTTTCTCCCAACCCGA N K E I Q G K E W G S 1900 ICGAATCCACAGCGGTCTTCATGTTGTCGTA 7 S D V A T K M N D Y
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P P 1700 GCAGCGATGCTCAAGCCACACAT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I	CATTOGGTAGCGCCATCAAC A N P L A M L CAAACAGCAAGGCTTTGA L D A L S Q CAAACAGGTAGTGGCTGACC D P L Y H S V 1950 SPICE A AGGCTGGCTGACC	TEGECTTGTTCATCGGCAAAGG E G Q E D A F AAGTETGGTGGCCCCTTGCCACA F D P A G Q W 1 1850 ATGGCAGTGAACCGCTGATTGA M A T F R Q N	CCTCTTGCACGGTTTGGTCCAATT A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAC V R R E K G R G V 2000 CATGTTGTTGCTGCTGTTGTCACGCCCCCAC	COGGGTGGCGCAGCTCAGACCACAGCGCTTT P H R L E S W L A K 1800 TCTTCTCAATCTGGCCTTTCTCCCAACCCGA K E I Q G K E W G S 1900 CCGAATCCACAGCGGTCTTCATGTTGTCGTA 7 S D V A T K M N D Y +
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P P 1700 GCAGCGATGCTCAAGCCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C	CATTOGGTAGCGCCATCAAC A N P L A M L CCAAACAGCAAGGCTTTGA N L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 Spi CGAAGATTGAAAAG <u>CCATC</u> G F I Q F A H	TEGECTTGTTCATEGGCAAAGG E G Q E D A F AAGTETGGTGGCCCCTTGCCACA F D P A G Q W 1 1850 ATGGCAGTGAACCGCTGATTGA M A T F R Q N CAGTGGGCATCAAACAGCATTT W H A D F L M 1	CCTCTTGCACGGTTTGGTCCAATY A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAC V R R E K G R G V 2000 CATGTTTTTGCTGGTAGTAAGCCCC E H K Q Q Y Y A D	CGGGGTGGCGCAGCTCAGACCACAGCGCTTT E P H R L E S W L A K 1800 ICTTCTCCAATCTGGCCTTTCCCCAACCGA N K E I Q G K E W G S 1900 CGAATCCACAGCGGTCTTCATGTTGTCGTA 7 S D V A T K M N D Y + SAAGCACAAAGGCCCGGCTGTGGGCCAACTT 8 L V F A R S H A L K
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P P 1700 GCAGCGATGCTCAAGCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C	CATTOGGTAGCGCCATCAAC A N P L A M L CCAAACAGGCAAGGCTTTGA N L D A L S Q CCAACAGGTAGTGGCTGACC D F L Y H S V 1950 SphI CGAAGATTGGAAG <u>GCATGC</u> S F I Q F A H	TEGECTTOTTCATEGGCAAAGG E G Q E D A F AAGTETGGTGECECETTGECEACA F D P A G Q W 1 1850 ATGGCAGTGAACEGETGATTGA M A T F R Q N CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100	CCTCTTGCACGGTTTGGTCCAATN A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAA V R R E K G R G V 2000 CATGTTTTTGCTGGTAGTAAGCCCC E H K Q Q Y Y A D	CEGGGETGGCGCAGCTCAGACCACAGCGCTTT E P H R L E S W L A K 1800 ICTTCTCAATCEGGCCTTTCTCCCAACCCGA N K E I Q G K E W G S 1900 CEGAATCCACAGCGGCTCTCATGTTGTCGTA 7 S D V A T K M N D Y 4 SAAGCACAAAGGCCCGGCTGTGGGGCCAACTT 8 L V F A R S H A L K 2150
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P P I700 GCAGCGATGCTCAAGCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCATCTGAAGTTTGAC	CATTOGGTAGCGCCATCAAC A N P L A M L CAAACAGGCAAGGCTTTGA N L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 SphI CGAAGATTTGAAAG <u>GCATGC</u> 3 F I Q P A H	TEGECTTOTTCATEGGCAAAGG E G Q E D A F AAGTETGGTGCCCCTTGCCACAA F D P A G Q W 1 1850 ATGGCAGGGAGCGCTGATTGAA M A T F R Q N CAGTGGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGGCCTCCACTCCAATGA	CCTCTTGCACGGTTTGGTCCAATT A E Q V T Q D L J 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V I CCCTGCGCTCTTTGCCACGCCCCA V R R E K G R G V 2000 CATGTTTTTCCTGGTAGTAAGCCCC E H K Q Q Y Y A J ATTGGAAGGCTTCGCCACAAGCAAJ	CGGGGTGGCGCAGCTCAGACCACAGCGCTTT E P H R L E S W L A K 1800 ICTTCTCAATCTGGCCTTTCTCCCAACCCGA K E I Q G K E W G S 1900 ICGAATCCACAGCGGCTCTTCATGTTGTCGTA 7 S D V A T K M N D Y AAGGACCAAAGGCCCGGCTGTGGGCCAACTT R L V F A R S H A L K 2150 IGCGCAAGGGGATGAAACAACCTTGCCCGA
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P P I700 GCAGCGATGCTCAAGCCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCATCTGAAGTTTGAC F Q A I Q L K V	CATTOGGTAGCGCCATCAAC A N P L A M L CAAACAGGTAGCGCTATGA N L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 SphI CGAAGATTTGAAAG <u>GCATGC</u> 3 F I Q F A H CCTGTTTGCCCGCTATGCGG V Q K G A I R	TEGECTTOTTCATEGGCAAAGG E G Q E D A F AAGTETGGTGCCCCTTGCCCACA F D P A G Q W 1 1850 ATGGCAGTGAACCGCTGATTGA M A T F R Q N CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCCACTCCAATGA	CCTCTTGCACGGTTTGGTCCAATT A E Q V T Q D L J 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V I CCCTGCGCTCTTTGCCACGCCCCAA V R R E K G R G V 2000 CATGTTTTTCCTGGTAGTAAGCCCC E H K Q Q Y Y A J ATTGGAAGGCTTCGCCACAAGCAAJ F Q F A E G C A J	CGGGGTGGCGCAGCTCAGACCACAGCGCTTT E P H R L E S W L A K 1800 CCTCTCCCAATCTGGCCTTTCTCCCAACCCGA K E I Q G K E W G S 1900 CGAATCCACAGCGGTCTTCATGTTGTCGTA 7 S D V A T K M N D Y AAGGACCAAAGGCCCGGCTGTGGGCCAACTT R L V F A R S H A L K 2150 GGGCGAAGGGGATGAAACAACCTTGCCCGA F R L P I F C G K G S
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P Z I700 GCAGCGATGCTCAAGCCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCATCTGAAGTTTGAC F Q A I Q L K V	CATTOGGTAGCGCCATCAAC A N P L A M L CCAAACAGGCAAGGCTTTGA N L D A L S Q CCAAACAGGTAGTGGCTGACC D F L Y H S V 1950 SphI GGAGGATTTGAAGGCATGC G F I Q P A H CCTGTTTGCCCGCTATGCGG V Q K G A I R 2200	TEGECTTOTTCATEGGCAAAGG E G Q E D A F AAGTETGGTGCCCCTTGCCACA F D P A G Q W 1 1850 ATGGCAGTGAACCGCTGATTGA M A T F R Q N CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCACTCCAATCGA A F D E S W D ϕ	CCTCTTGCACGGTTTGGTCCAATT A E Q V T Q D L J 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V I CCCTGCGCTCTTTGCCACGCCCCAA V R R E K G R G V 2000 CATGTTTTTCCTGGTAGTAAGCCCC E H K Q Q Y Y A J ATTGGAAGGCTTCGCCACAAGCAAA F Q F A E G C A J 2250	CGGGGTGGCGCAGCTCAGACCACAGCGCTTT E P H R L E S W L A K 1800 CCTCTCCCAATCTGGCCTTTCTCCCAACCCGA N K E I Q G K E W G S 1900 CGAATCCACAGCGGTCTTCATGTTGTCGTA 7 S D V A T K M N D Y AGGCCAAGGGCCGGCCTGTGGGCCAACTT R L V F A R S H A L K 2150 GGGCGAAGGGCATGAACACCTTGCCCGA F R L P I F C G K G S
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P Z I700 GCAGCGATGCTCAAGCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCATCTGAAGTTTGAC F Q A I Q L K X	CATTOGGTAGCGCCATCAAC A N P L A M L ICAAATCAGCAAGGCTTTGA N L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 SphI GAAGATTTGAAAGGCATCC G F I Q F A H CCTOTTTGCCCGCTATCCGG V Q K G A I R 2200 AATCTTCCACTGTCGGGCA	TEGECTTOTTCATEGGCAAAGG E G Q E D A F AAGTETGGTGCCCCCTTGCCACA F D P A G Q W 1 1850 ATGGCATGGACCGCTGATTGA M A T F R Q N 1 CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCACTCCAATCGA A F D E S W D $\stackrel{\circ}{}$	CCTCTTGCACGGTTTGGTCCAATT A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAA V R R E K G R G N 2000 CATGTTTTTGCTGGTAGTAAGCCCC E H K Q Q Y Y A D ATTGGAAGGCTTCGCCACAAGCAA F Q F A E G C A D 2250 ACCCGGTAAAGCCCAAAGCACAAGCACAA	CGGGGTGGCGCAGCTCAGACCACAGCGCTTT E P H R L E S W L A K 1800 CCTCTCCCAATCTGGCCTTTCTCCCAACCGA N K E I Q G K E W G S 1900 CGAATCCACAGCGGTCTTCATGTTGTCGTA / S D V A T K M N D Y AAGCACAAAGGCCCGGCTGTGGGCCAACTT R L V F A R S H A L K 2150 GCGCCAAGGGGATGAACAACCCTTGCCCGA F R L P I F C G K G S AATCCCGGTACATGCTGCGCAGGTTTCTGCG
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P Z I700 GCAGCGATGCTCAAGCACACATT C R H E L W V V TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCTTGAACTGTTCGGZ T Q A K F Q E S	CATTOGGTAGCGCCATCAAC A N P L A M L ICCAAATCAGCAAGGCTTTGA N L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 SphI CGAAGATTTGAAAGGCATCC G F I Q F A H CCTOTTTGCCCCGCTATCCGG V Q K G A I R 2200 AATCTTTCCACTGTCGGGGCA S D K W Q R A	TEGECTTOTTCATEGGCAAAGG E G Q E D A F AAGTETGGTGGCCCCTTGCCACA F D P A G Q W 1 1850 ATGGCATGGACCGCTGATGAA M A T F R Q N 1 CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCACTCCTCAATGCA A F D E S W D $\stackrel{\circ}{}$ AAGGCACACACTCGGTCATAAG F A C V R D Y $\stackrel{\circ}{}$	CCTCTTGCACGGTTTGGTCCAATT A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAA V R R E X G R G V 2000 CATGTTTTTGCTGGTAGTAAGCCCC E H X Q Q Y Y A D ATTGGAAGGCTTCGCCACAAGCAA F Q F A E G C A D 2251 ACCCGGTAAAGCCCAAAGCACCAA S G T F G L A V D	CGGGGTGGCGCAGCTCAGACCACAGCGCTTT E P H R L E S W L A K 1800 CCTCTCCCAATCTGGCCTTTCTCCCAACCGA N K E I Q G K E W G S 1900 CGAATCCACAGCGGTCTTCATGTTGTCGTA X S D V A T K M N D Y AAGCACAAAGGCCCGGCTGTGGGCCAACTT R L V F A R S H A L K 2150 CGGCAAGGGATGAACAACCTTGCCCGA F R L P I F C G K G S AACCCGGTACATGCTGCGCAGGTTCTGCGG L D R Y M S R L N R R
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P Z I700 GCAGCGATGCTCAAGCACACATT C R H E L W V V TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCTTGAACTGTTCGGZ T Q A K F Q E S 2300	CATTOGGTAGCGCCATCAAC A N P L A M L ICAAATCAGCAAGGCTTTGA N L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 SphI GAAGATTTGAAAGGCATCC G F I Q F A H COTOTTTGCCCGCTATCGGG V Q K G A I R 2200 AATCTTTCCACTGTCGGGGCA S D K W Q R A	TEGECTTOTTCATEGGCAAAGG E G Q E D A F AAGTETGGTGGCCCCTTGCCACA F D P A G Q W 1 1850 ATGGCATGAACCGCTGATGAA M A T F R Q N 1 CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCAACCACCCCAATGCA A F D E S W D 1 AAGGCACACACTCGGTCAATAG F A C V R D Y 1 200	CCTCTTGCACGGTTTGGTCCAATTA A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAA V R R E X G R G V 2000 CATGTTTTGCTGGTAGTAAGCCCCA E H K Q Q Y Y A D ATTGGAAGGCTTCGCCACAAGCAA F Q F A E G C A D 2250 ACCCGGTAAAGCCCAAGCGACCAA S G T F G L A V D 2350	CGGGGTGGGGCAGCTCAGACCACAGCGCTTT E P H R L E S W L A K 1800 CCTCTCCCAATCTGGCCTTTCTCCCAACCGA N K E I Q G K E W G S 1900 CGAATCCACAGCGGTCTTCATGTTGTCGTA X S D V A T K M N D Y AAGCACAAAGGCCCGGCTGTGGGCCAACTT R L V F A R S H A L K 2150 GGCGCAAGGGATGAAACACCTTGCCGA F R L P I F C G K G S AATCCCGGTACATGCTGCGCAGGTTCTGCG L D R Y M S R L N R R 2400
TTGCTCCACGAATGCATCAAAGGC Q E V F A D F Z I700 GCAGCGATGCTCAAGCACACATT C R H E L W V V TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCTTGAACTGTTCGGA T Q A K F Q E S 2300 CAGCTTCTTTTGTCTTTTGTGTGCT	CATTOGGTAGCGCCATCAAC A N P L A M L ICCAAATCAGCAAGGCTTTGA N L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 SphI CGAAGATTTGAAAGGCATGC G F I Q F A H CTOTTTGCCCCGTATGCGG V Q K G A I R 2200 AATCTTTCCACTGCGGGCA S D K W Q R A CGGTGGAGAGGCCACTGCCTT B T S L W Q R	TEGECTTOTTCATEGGCAAAGG E G Q E D A F AAGTETGGTGCCCCTTGCCACA F D P A G Q W 1 1850 ATGCCATGAACCGCTGATGAA M A T F R Q N 1 CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCACAACAGCATTAA A F D E S W D 1 AAGGCACACACCTGGTCATAAG F A C V R D Y 1 AAGGCACACACTGGGCTCAAAAGGACTTAA L K P E F P S 1	CCTCTTGCACGGTTTGGTCCAATTA A E Q V T Q D L 1 1750 GGCGTTGGCGGGAATCCTGCACGTT L R Q R S D Q V 1 CCCTGCGCTCTTTGCCACGCCCCAA V R R E X G R G V 2000 CATGTTTTTGCTGGTAGTAAGCCCC E H K Q Q Y Y A 1 ATTGGAAGGCTTGCCCCAAGCCAA F Q F A E G C A 1 2250 ACTGCCAAGCCGTGTCGCCGCTGCC L K G V S D R A 1	CGGGGTGGGGCAGCTCAGACCACAGCGCTTT P H R L E S W L A K 1800 CCTCTCCCAATCTGGCCTTTCTCCCAACCGA W K E I Q G K E W G S 1900 CGAATCCACAGCGGTCTTCATGTTGTCGTA X S D V A T K M N D Y AAGCACAAAGGCCCGGCTGTGGGCCAACTT R L V F A R S H A L K 2150 SGGGATGGAGCATGCAGCCTTGCCGCA F R L P I F C G K G S ACCCCGGTACAGCGTGCCGCAGGTTTCTGCG D R Y M S R L N R R 2400 GGTACTGCGGTCAACCCTGGCTTGCGCTAAC
TTGCTCCACGAATGCATCAAAGGC Q E V F A D F F 	CATTOGGTAGCCCATCAAC A N P L A M L CRAAATCAGCAAGGCTTTGG N L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 Sphi CAAGAGTTGAAAGGCATC G F I Q F A H CCTOTTTGCCCGCTATCCGG ATCTTTCCCCGCTATCCGG ATCTTTCCCTGTGGGCGAC S D K W Q R A CGGTGGAGAGCCACTGCCTT E T S L W Q R	TCGCCTTGTTCATCGGCAAAGG E G Q E D A F AAGTCTGGTGCCCCTTGCCACA F D P A G Q W 1 1850 ATGCCATGAACCGCTGATTGA M A T F R Q N 1 CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCACTCCAATCGA A F D E S W D 1 AAGGCACACACTCGGTCATAAG F A C V R D Y 1 AACTTGGGCTCAAAAGGACTTA L K P E F P S 1 2450	CCTCTTGCACGGTTTGGTCCAATTA A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCA V R R E X G R G V 2000 CATGTTTTGCTGGTAGTAAGCCCCA E H K Q Q Y Y A D ATTGGAAGGCTTCGCCCAAAGCAAJ F Q F A E G C A D 2250 ACCCGGTAAAGCCCAAAGCCAAGCAAJ S G T F G L A V D 2350 ACTTGCCAACGCTGTCTGGCGCTGC	CGGGGTGGGCAGCTCAGACCACAGCGCTTT C P H R L E S W L A K 1800 CCTCTCCCAATCTGGCCTTTCTCCCAACCGA N K E I Q G K E W G S 1900 CCAATCCACGCGGTCTTCATGTTGTCGTA 7 S D V A T K M N D Y HAAGCACAAAGGCCCGGCTGTGGGCCAACTT R L V F A R S H A L K 2150 CGGGCAAGGGGATGAAACAACCACCTGCCCGA F R L P I F C G K G S ATCCCCGGTACATGCTGCGCAGGTTTCTGCG L D R Y M S R L N R R 2400 GGTACTGCGGTTCAACCACCTTGCCTTGCCAA P Y Q P E V V K S Q L
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P P 1700 GCAGCGATGCTCAAGCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCTTGAACTGTTCGGJ T Q A K F Q E S 2300 CAGCTTCTTTGTCTTTTTGTCCTC L K K T K K H E	CATTOGGTAGCCCCTTCAAC A N P L A M L CAAATCAGCAAGGCTTTCAA N L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 Sphi CAAGGTTGAAGGCATCGGC J F I Q F A H CCTOTTTGCCCGCTATCGGG ATCTTTCCCCGCTATCGGG S D K W Q R A CGGTGGAGAGGCCACTGCCTT E T S L W Q R	TCGCCTTGTTCATCGGCAAAGG E G Q E D A F AAGTCTGGTGCCCCTTGCCACA F D P A G Q W 1 1850 ATGCCATGAACCGCTGATTGA M A T F R Q N 1 CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCACTCCATCGATCGA A F D E S W D 1 AAGGCACACACTCGGTCATAAG F A C V R D Y 1 AACTTGGGCTCAAAAGGACTTA L K P E F P S 1 2450 XTTCCCGAATCGACGCACCAT	CCTCTTGCACGGTTTGGTCCAATTA A E Q V T Q D L 1 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V 1 CCCTGCGCTCTTTGCCACGCCCCA V R R E K G R G V 2000 CATGTTTTGCTGGTAGTAAGCCCA E H K Q Q Y Y A 1 ATTGGAAGGCTTCGCCCAAAGCAAI F Q F A E G C A 1 2250 ACTTGCCAACGCTGTCTCGCGCCGC ACTTGCCAACGCTGTCTCGCGCCGC L K G V S D R A 1 CGCGAAAATGCCAGCGTCGAATGG	CGGGGTGGGCAGCTCAGACCACAGCGCTTT C P H R L E S W L A K 1800 CCTCTCCCAATCTGGCCTTTCTCCCAACCGA N K E I Q G K E W G S 1900 CCAATCCACGCGGTCTTCATGTTGTCGTA 7 S D V A T K M N D Y AGAGCACAAAGGCCCGGCTGTGGGCCAACTT R L V F A R S H A L K 2150 CGGGCAAGGGGATGAACAACCCTTGCCGA F R L P I F C G K G S ATCCCGGTACAGCGGTGGGCGAGGTTCTCGCG L D R Y M S R L N R R 2400 GGTACTGCGGTCAACCACCTTGCTTTGCAA P Y Q P E V V K S Q L 2500 2500 CTCCCTTGCCTTGCCTTGCCTTGCAA
TIGCITCCACGAATGCATCAAAGGC Q E V P A D P P 1700 GCAGCGATGCTCAAGCACCATC C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCTTGAACTGTCCGJJ T Q A K P Q E S 2300 CAGCTTCTTTGCTTTTTGTCTC L K K T K K H F ATACTTGCGAACGGTGTTCCTGJJ Y K R V T N R S	CATTOGGTAGCGCCATCAAC A N P L A M L CAAACAGGTAGCGCATGGCA N L D A L S Q CAAACAGGTAGTGGCTAACC D P L Y H S V 1950 Sphi CAAGATTGAAAGGCATCG G F I Q P A H COTOTTTGCCCGCTATCGGG A Q K G A I R 2200 AATCTTTCCCTGTGGGCA S D K W Q R A CGGTGGAGAGCCACTGCCTT E T S L G S R R A	TCGCCTTGTTCATCGCCAAAGG E G Q E D A F AAGTCTGGTGCCCCTTGCCACA F D P A G Q W 1 1850 ATGGCAGTGAACCGCTGATTGA M A T F R Q N 1 CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCACTCCAATCGA A F D E S W D 1 + AAGGCACACACTCGATCATAAG F A C V R D Y 1 AACTTGGGCTCAAAAGGACTTA L K P E F P S 2450 ATTTCCCGAATCGACGCACCAT I E R I S A G	CCTCTTGCACGGTTTGGTCCAATTA A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAC V R R E K G R G V 2000 CATGTTTTGCTGGTAGTAAGCCCC E H K Q Q Y Y A D ATTGGAAGGCTTCGCCACAAGCAA F Q F A E G C A D 2250 ACCTGCCAACGCTGTCTCGCGCTGG ACTTGCCAACGCTGTCTCGCGCTGG L K G V S D R A D CGCGAAAATGCCAGGCGTCGAATTGC D R F H W R R I J	COGGGTGGCGCAGCTCAGACCACAGCGCTTT E P H R L E S W L A K 1800 RCTACCAACCGGCCTTTCTCCCAACCCGA K E E I Q G K E W G S 1900 CCGAATCCACAGCGGCTCTTCATGTTGTCGTA 7 S D V A T K M N D Y SAAGCACAAAGGCCCGGCTGTGGGCCAACTT R L V F A R S H A L K 2150 SCGCAAGGGGATGAAACACCCTTGCCGG 7 R L P I F C G K G S SAAGCCCGGTTCAACCACCTTGCCGG 4 D R Y M S R L N R R 2400 SGTAACTGCGGTTCAACCACCTTGCTTTGCAA P Y Q P E V V K S Q L 2500 200 2
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P P 1700 GCAGCGATGCTCAAGCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCTTGAACTGTTCGGA T Q A K F Q E S 2300 CAGCTTCTTTGTCTTTTTGTCTT L K K T K K H F ATACTTGCGAACGGTGTTCCTGGJ Y K R V T N R S <u>rbs</u> .	CATTOGGTAGCGCCATCAACA A N P L A M L CAAAACAGGTAGTGGCTAGCG D P L Y H S V 1950 Sphi CGAAGATTGAAGGCATGGG G F I Q F A H CCTGTTTGCCCGCTATGCGG 7 Q K G A I R 2200 AATCTTTCCACTGTCGGGCA S D K W Q R A CGGTGGAGAGCCACTGCCTT E T S L W Q R ACAGGCCGCTTCGTCGGGGCT S L G S R R A 2550	TCCCCATTGGCCTCALAAGGACTTA L K P L R I S A G C C C C C C C C C C C C C C C C C C	CCTCTTGCACGGTTTGGTCCAATTA A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAC V R R E K G R G V 2000 CATGTTTTTGCTGGTAGTAAGCCCCA E H K Q Q Y Y A D ATTGGAAGGCTTGCCACAAGCAAJ F Q F A E G C A D 2250 ACTTGCCAACGCTGTCTCGCGCTGG ACTTGCCAACGCTGCTCTCGCGCTGG CGCGGAAAATGCCAGCGGTCGAATTGG D R F H W R R I D -10 1 2600 L3	COGGGTGGGCGAGCTCAGACCACAGCGCTTT E P H R L E S W L A K 1800 RCTACCAACCGGCCTTCTCCCCAACCCGA K E I Q G K E W G S 1900 CCGAATCCACAGCGGCTCTCATGTTGTCGTA 7 S D V A T K M N D Y 4 SAAGCACAAAGGCCCGGCTGTGGGCCAACTT 8 L V F A R S H A L K 2150 AGGCCAAGGGATGAAACAACCCTTGCCGA 7 R L P I F C G K G S ACCCGGTACATGCTGCGGAGGTTTCTGCGG 1 D R Y M S R L N R R 2400 MCTACTGCGGTTCAACCACCTTGCTTTGCAA P Y Q P E V V K S Q L 2500 2
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P P 1700 GCAGCGATGCTCAAGCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCTTGAACTGTTCCGGA T Q A K F Q E S 2300 CAGCTTCTTTGGAACGGTGTTCCTGGI Y K R V T N R S <u>LES</u> .	CATTOGGTAGCGCCATCAACA A N P L A M L CAAACAGGTAGCGCAAGGCTTTGA L D A L S Q CAAACAGGTAGTGGCTGACC D P L Y H S V 1950 Sphi CGAAGATTTGAAGGCATGC G F I Q F A H CCTGTTTGCCCCGCTATGCGG V Q K G A I R 2200 AATCTTTCCACTGCCGGCA S D K W Q R A CGGTGGAGAGCCACTGCCTT S L G S R R A 2550 GGAAACAGTGCATACGTGG	TEGECTTOTTCATEGGCAAAGG E G Q E D A F AAGTETGGTGGCCCCTTGCCACAA F D P A G Q W 1 1850 ATGGCAGTGAACCGCTGATTGA M A T F R Q N $^{\prime}$ CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCACTCCAATGAA A F D E S W D $^{\prime}$ AAGGCACACACTCGGTCATAAG F A C V R D Y $^{\prime}$ AACTTGGGCTCAAAAGGACTTA L K P E F P S $^{\circ}$ 2450 MITTCCGGAATGGACGCACATT I E R I S A G	CCTCTTGCACGGTTTGGTCCAATTA A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAA V R R E K G R G V 2000 CATGTTTTGCTGGTAGTAAGCCCCA E H K Q Q Y Y A D ATTGGAAGGCTTGCCCAAAGCAAJ F Q F A E G C A D 2250 ACTTGCCAAGCCTGTCTCGCGCTGAA L K G V S D R A D 2350 ACTTGCCAAGCGTCGCCGCGAATTGG L K G V S D R A D 2350 ACTTGCCAAGCGCCGGTCGAATTGG CGCGGAAAATGCCAGCGGTCGAATTGG D R F H W R R I D 2400 TACCCTAAGTGGGTCAATTTTAAA	COGGGTGGCGCAGCTCAGACCACAGCGCTTT E P H R L E S W L A K 1800 RCTACCAATCAGCCTTCTCCCCAACCGA N K E I Q G K E W G S 1900 CCGAATCCACAGCGGCTCTCATGTTGTCGTA 7 S D V A T K N N D Y A CCGAATCCACAGCGGCCGGCTGTGGGCCAACTT SAAGCACAAAGGCCCGGCTGTGGGCCAACTT SAAGCACAAAGGCCCGGCTGTGGGCCAACTT C L V F A R S H A L K 2150 AGCGCAAGGGGATGAAACAACCCTTGCCGA F R L P I F C G K G S ACTCCCGGTACATGCTGGCGCAGGTTTCTGCGG C D R Y M S R L N R R 2400 GGTACTGCGGTTCAACCACCTTGCCTTGCAA P Y Q P E V V K S Q L 2500 CCCAACTCAACGGCCATGCTTGACGTTGCCGA C S L I A V N I M ISTA -35 NGCAACTCAACGGGCCATGCTGAGTGTGCCA
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P Z 1700 GCAGCGATGCTCAAGCCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCATGAACTGTTCCGGJ T Q A K F Q E S 2300 CAGCTTCTTTGCTCTTTTGTGTCC L K K T K K H F ATACTTGCGAACGGTGTTCCTGG/ Y K R V T N R S <u>EDS</u>	CATTOGGTAGCGCCATCAACA A N P L A M L CAAAATCAGCAAGGCTTTGA V L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 Sphi CGAAGATTTGAAGGCATGC G F I Q F A H CCTGTTTGCCCCGCTATGCGG V Q K G A I R 2200 AATCTTTCCACTGCGGCAT S D K W Q R A CGGTGGAGAGCCACTGCCTT E T S L W Q R ACAGGCCGCTTCGTCGGGCT S L G S R R A 2550 GGAAACAGTGTCATACGTGC	TCGCCTTGTTCATCGCCAAAGG E G Q E D A F AAGTCTGGTGCCCCTTGCCACAA F D P A G Q W 1850 ATGGCAGTGAACCGCTGATTGA M A T F R Q N CAGTGGGCATCAAACAGCATTT W H A D F L M 2100 GCAAAGTCCTCATCCAAACAGCATTT W H A D F L M 2100 GCAAAGTCCTCATCCAAACAGCATTA A F D E S W D AAGGCACACACTCGGTCATAGA F A C V R D Y AACTTGGGCTCAAAAGGACTTA L K P E F P S 2450 ATTTCCCGAATCGACGCACATT I E R I S A G GTCAAATTTCGACGCACAATCTT L2	CCTCTTGCACGGTTTGGTCCAATT A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAA V R R E K G R G V 2000 CATGTTTTTGCTGGTAGTAGTAGCCCCAA CATGGAAGGCTTCGCCCCAAGGCAAJ F Q F A E G C A D 1225 ACTCGCGAAAGGCCTGCCCAAGGCAAJGCGACCAAJ CCCGGGAAAATGCCAAGGCGTCGAATTGG L K G V S D R A D 2350 ACTTGCCAACGCTGTCTCGCGCGCGC L K G V S D R A D 2450 CGCGGAAAATGCCAGCGGTCGAATTGG D R F H W R R I Z -10 2600 TACCCTAAGTGGGGCAATTGTACA	COGGGTGGCGCAGCTCAGACCACAGCGCTTT E P H R L E S W L A K 1800 RCTCCCAATCTGGCCTTTCTCCCAACCCGA N K E I Q G K E W G S 1900 CCGAATCCACAGCGGCTCTCATGTTGTCGTA 7 S D V A T K M N D Y CGGAATCACAGCGGCCGGCTGTGGGCCAACTT R L V F A R S H A L K 2150 AGCGCAAGGGATGAAACACCCTTGCCGA C L P I F C G K G S AGCGCAAGGGGATGCAAACCACCTTGCCGA C L P I F C G K G S ACTCCCGGTACATGCTGGCGAGGTTCTGCCGA C D R Y M S R L N R R 2400 SGTACTGCGGTTCAACCACCTTGCCTTGCAA P Y Q P E V V K S Q L 2500 C C CAATCCAACGCCAGCTTGTCCCTTG C C S V N I M ISTA -35 NCCACTCAACAGGCCATGCTGAGTGTGCGA A M S L T R
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P Z 1700 GCAGCGATGCTCAAGCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCATGAACTGTTCCGGJ T Q A K F Q E S 2300 CAGCTTCTTTGCGAACGGTGTTCCTGG/ Y K R V T N R S <u>LES</u> ATTCTCCCCGCCATATCCAGACGC	CATTOGGTAGCGCCATCAACA A N P L A M L CAAAATCAGCAAGGCTTTGA V L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 Sphi CGAAGATTTGAAGGCATGC G F I Q F A H CCTGTTTGCCCCGCTATGCGG V Q K G A I R 2200 AATCTTTCCACTGCGGGCA S D K W Q R A CGGTGGAGAGCCACTGCCTT E T S L W Q R ACAGGCCGCTTCGTCGGGCT S L G S R R A 2550 GGAAACAGTGTCATACGTGC	TCGCCTTGTTCATCGCCAAAGG E G Q E D A F AAGTCTGGTGCCCCTTGCCACAA F D P A G Q W 1850 ATGGCAGTGAACCGCTGATTGA M A T F R Q N CAGTGGGCATCAAACAGCATTT W H A D F L M 2100 GCAAAGTCCTCATCCAACAGCATTT W H A D F L M 2100 GCAAAGTCCTCACTCCATCGA A F D E S W D AAGGCACACACTCGGTCATAAG F A C V R D Y AACTTGGGCTCAAACAGCACTTA L K P E F P S 2450 ATTTCCCGAATCGACGCACACTT I E R I S A G GTCAAATTTCGACGCACATCTT L2 2700	CCTCTTGCACGGTTTGGTCCAATTA A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAA V R R E K G R G V 2000 CATGTTTTTGCTGGTAGTAGGACGAC E H K Q Q Y Y A D ATTGGAAGGCTTCGCCCCAAGCAAJ F Q F A E G C A D 12250 ACTGGCAAAGCCCAAAGCGACGAA ACCGGGAAAATGCCAGCGGTCGAATTGG L K G V S D R A D 2350 ACTTGCCAACGCTGTCTCGCGCGCGC L K G V S D R A D 2350 ACTGCCAAAGCGGTCGAATTGG D R F H W R R I D 2600 TACCCTAAGTGGGTCAATTGTACA	COGGGTGGGGCAGCTCAGACCACAGCGCTTT E P H R L E S W L A K 1800 RCTCCCAATCTGGCCTTTCTCCCAACCGA N K E I Q G K E W G S 1900 CGAATCCACAGCGGCTCTCATGTTGTCGTA 7 S D V A T K M N D Y A CGGAATCCACAGCGGCCTGGGGCCAACTT R L V F A R S H A L K 2150 AGCGCAAGGGATGAAACACCCTTGCCGA R L P I F C G K G S ACCCCGGTACATGCTGGGCAGGTTCTGCG C D R Y M S R L N R R 2400 SGTACTGCGGTTCAACCACCTTGCTTTGCAA P Y Q P E V V K S Q L 2500 CGCCAATATCGCCAGCTTACCCTGCTTGCAA P Y Q P E V V K S Q L 2500 CGCCAATATCGCCCAGCTTACCCTGCCTTGCAA P Y Q P E V V K S Q L 2500 CGCCAATATCGCCCAGCTTACCCTGCTTGCAA P Y Q P E V V K S Q L 2500 CGCCAATATCGCCACGTTACCCTGCCTTGCCAA P Y Q P E V V K S Q L 2500 CGCCAATATCGCCACGTTACCCTGCTTGCCAA P Y Q P E V V K S Q L 2500 CGCCAATATCGCCCCCGCTGCGAGTGTGCGA A S L I A V N I M ISLA
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P P 1700 GCAGCGATGCTCAAGCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCTTGAACTGTTCGGJ T Q A K F Q E S 2300 CAGCTTCTTTGTCTTTTTGTGCTC L K K T K K H H ATACTTGCGAACGGTGTTCCTGGJ Y K R V T N R S <u>EbS</u> ATTTCTCCCGCCATATCCAGACGC	CATTOGGTAGCGCCATCAAC A N P L A M L CAAATCAGCAAGGCTTTGA N L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 Sphi CGAAGATTTGAAGGCTATGCGG V Q K G A I R 2200 AATCTTTCCACTGTCGGGCA S D K W Q R A CGGTGGAGAGCGCTATGCGGCT S T S L W Q R ACAGGCCGCTTCGTCGGGCT S L G S R R A 2550 CGAAGACGGCGACGACGCCTATACGTGG C V A V I A	TCGCCTTGTTCATCGGCAAAGG E G Q E D A F AAGTCTGGTGCCCCTTGCCACAA F D P A G Q W 1 1850 ATGGCAGTGAACCGCTGATTGAA M A T F R Q N 1 CAGTGGGCATCAAACAGCGTTGA M A A D F L M 1 2100 GCAAAGTCCTCACTCCAATCGA A F D E S W D 1 AAGGCACACACTCGGTCATAAG F A C V R D Y 1 AACTTGGGCTCAAAAGGACTTA L K P E F P S 1 2450 ATTTCCCGAATCGACGCACAT I E R I S A G 1 GCCAAATTCGACGCACAAATCTT L2 2700 CCGCCATCAGCACACGCCGCCAG	CCTCTTGCACGGTTTGGTCCAATTG A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGTT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAA V R R E K G R G V 2000 CATGTTTTTGCTGGTAGTAGTAAGCCCCC E H K Q Q Y Y A D ATTGGAAGGCTTGCCACAAGCAAJ F Q F A E G C A D 2250 ACTGGCAAAGCCCTGCCAAAGCGACCAA S G T F G L A V D 2350 ACTTGCCAAAGCGGTCGAATTGG L K G V S D R A D 2350 ACTGCCAAAGTGCCACGGTCGAATTGG D R F H W R R I D 2600 TACCCTAAGTGGGTCAATTGTAGT CGCCGAAAATGCCAGCGTCGAATTGG D R F H W R R I D 2600 TACCCTAAGTGGGTCAATTGTAGTTTTAGA	CGGGGTGGGGCAGCCCGGCGGCGCAGCCAGCCAGCCAGCC
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P P 1700 GCAGCGATGCTCAAGCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCTTGAACTGTTCGGJ T Q A K F Q E S 2300 CAGCTTCTTTGTCTTTTGTCCTCGJ L K K T K K H E ATACTTGCGAACGGTGTTCCTGGJ Y K R V T N R S <u>L'DS</u> ATTCCCCGCCATATCCAGACGC 2650 TGGTTGATCGCTTCCTCGCCGCTC H N I A E E G S	CATTOGGTAGCGCCATCAAC A N P L A M L CAAACAGCAAGGCTTTGA N L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 SphI GAAGATTTGAAAGGCACTGC G F I Q F A H COTOTTTGCCCGCTATGCGG V Q K G A I R 2200 AATCTTTCCACTGTCGGGCA S D K W Q R A CGGTGGAGAGCCACTGCCTT E T S L W Q R ACAGGCCGCTTCGTCGGGGCT S L G S R R A 2550 GAAACAGGTGTCATACGTGG CTCCACGGCGACGATGGCCATGCCGT E V A V I A	TCGCCTTGTTCATCGGCAAAGG E G Q E D A F AAGTCTGGTGCCCCTTGCCACAA F D P A G Q W 1 1850 ATGGCAGTGAACCGCTGATTGAA M A T F R Q N 1 CAGTGGGCATCAAACAGCGTTT W H A D F L M 2 100 GCAAAGTCCTCACTCCAATCGA A F D E S W D 1 AAGGCACACACTCGGTCATAAG F A C V R D Y 1 AACTTGGGCTCAAAAGGACTTA L K P E F P S 2 2450 ATTTCCCGAATCGACGACACAT I E R I S A G C GCCAAATTCGACGCAAAATCTT L2 2700 CCGCCCATCAGCAGTGCGCCAG A M L L H A L	CCTCTTGCACGGTTTGGTCCAATTG A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAA V R R E K G R G V 2000 CATGTTTTTCCTGGTAGTAGACGCCCA E H K Q Q Y Y A D ATTGGAAGGCTTCGCCACAAGCAAJ F Q F A E G C A D 2250 ACCTGCCAAAGTGCCCAAAGCGACCAA S G T F G L A V D 2350 ACCTGCCAAAGTGCCCCGCTGCAATTGG L K G V S D R A D 2350 L K G V S D R A D 2350 L CCCGAAAATGCCAGCGTCCGAATTGG L R F H W R R I Z -10 2600 L3 TACCCCTAAGTGGGTCAATTGTACA	CGGGCTGGGCGCAGCCCGGCGCGCGCGCGCGCGCGCGCGC
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P P I700 GCAGCGATGCTCAAGCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAACTGTTCGGJ T Q A I Q L K V GGTTTGCGCCTTGAACTGTTCGGJ T Q A K F Q E S 2300 CAGCTTCTTTGTCTTTTGTCCTCGJ L K K T K K H F ATACTTGCGAACGGTGTTCCTGGJ Y K R V T N R S <u>L'DS</u> ATTTCCCCCGCCATATCCAGACGC GGTTGATCGCTTCCTCGCCGCTC H N I A E E G S GCAATTGGGGAAGGCCGCCGCGCGCAGC	CATTOGGTAGCGCCATCAAC A N P L A M L CAAACAGGTAGCGCATGAG CAAACAGGTAGTGGCTAGCG D F L Y H S V 1950 SphI GAAGATTTGAAAGGCAGCGCA C F L Q F A H COTOTTTGCCCGCTATGCGG V Q K G A I R 2200 AATCTTTCCACTGTCGGGCA S D K W Q R A CGGTGGAGAGGCCACGCCTTG E T S L W Q R ACAGGCCGCTTCGTCGGGGCT S L G S R R A 2550 3GAAACAGGCGACGACGCCGCG E V A V I A 2800 CGGGAGGGAGAGGCGACGGCGCACG	TCGCCTGTGTCATCGGCAAAGG E G Q E D A F AAGTCTGGTGCCCCTTGCCACAA F D P A G Q W 1 1850 ATGGCAGGGATGAACCGCTGATGAA M A T F R Q N \sim CAGTGGGCATCAAACAGCATTGA W H A D F L M \sim 2100 GCAAAGTCCTCAACAGCATTAA A F D E S W D \sim AAGGCACACACTCGGTCATAAG F A C V R D Y \sim AACTTGGGCTCAAAAGGACTTA L K P E F P S \sim 2450 ATTTCCCCAATCGACGACGACCAT I E R I S A G \sim GCGCCATCAGCAGTGGCGCCAG A M L L H A L TGGCCAGCAGTGAGCGACGACATC	CCTCTTGCACGGTTTGGTCCAATTG A E Q V T Q D L J 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V I CCCCGCGCTCTTTGCCACGCCCCAA V R R E K G R G V 2000 CATGTTTTTCCTGGTAGTAAGCCCC E H K Q Q Y Y A J ATTGGAAGGCTTCGCCACAAGCAAJ F Q F A E G C A J 2250 ACCCGGTAAAGCCCAAAGCGACCAA S G T F G L A V J 2350 ACCTGCCAAGGGTGCTCTCGCCGCTGC L K G V S D R A J 2350 CACCTGCCAAGCGGTGCAATTGTAGA CGCGAAAATGCCAGCGGTGCAATTGTAGA CGCGAAAATGCCAGCGGCGGCGAATTGT D R F H W R R I Z 10 \downarrow 2600 L3 TACCCTAAGTGGGTCAATTGTAGA	CGGGCTAGGCGCAGGCCTGGCCAGGCTTCTCGCG CTTCTCCAATCTGGCCTTTCTCCCAACCGA K E I Q G K E W G S 1900 CGGATCCACAGCGGTCTTCATGTTGTCGTA X S D V A T K M N D Y AGGCCAAGGGCCGGCCTGTGGCCCAACTT R L V F A R S H A L K 2150 CGGCAAGGGGATGAAACAACCTTGCCGA R L P I F C G K G S CGCCAAGGGGATGAAACAACCTTGCTTGCGA P Y Q P E V V K S Q L 2500 CGCCAATCAACGGCCAGGTTATCCACTCGGA CGCCAACTCAACTGGCCAGGTTATCCACTCGGA CGCCAACTCAACAGGCCATGCCGAGGTTATCCACTCCTTG A S L I A V N I M ISTA -35 MCCAACTCAACAGGGCATGCGGCACTTCCCAGCGTG T L L Y R A M D L T CACCGGCACGTCGAACGGATTTCCAACTGG
TTGCTCCACGAATGCATCAAAGG Q E V F A D F Z 1700 GCAGCGATGCTCAAGCACACATT C R H E L W V N TGCTGGATTACAGAACTGCGCACATT A P N C F Q A I GATGCCGCGCCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCTTGAACTGTTCGGZ T Q A K F Q E S 2300 CAGCTTCTTTGTCTTGTCTTTTGTGCTC L K K T K K H E ATACTTGCGAACGGTGTTCCTGGZ Y K R V T N R S <u>LEBS</u> ATTTCTCCCGCCATATCCAGACGC TGGTTGATCGCTTCCTCGCGCCCCC H N I A E E G S GCAATTGGGGAAGGCCGCCCCCCC	CATTOGGTAGCGCCATCAAC A N P L A M L CAAACAGGTAGCGCATCAAC A N L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 SphI GAAGATTTGAAAGGCACAC C F L Q F A H CTOTTTGCCCGCTATGCGG V Q K G A I R 2200 AATCTTTCCACTGTCGGGCA S D K W Q R A COGTGGAGAGGCCACCGCCTT E T S L W Q R ACAGGCCGCTTCGTCGGGGCT S L G S R R A 2550 GGAAGAGGCGCATCGCCGGGCC E V A V I A 2800 CGGGGAGGGACGACGCCCC P L S A A F	TCGCCTQTTCATCGGCAAAGG E G Q E D A F A AAGTCTGGTGCCCCTTGCCACA F D P A G Q W 1 1850 ATGGCAGGGCATCAAACGGCTGATTGA M A T F R Q N 1 CAGTGGGCATCAAACGGCATTGA M A T F R Q N 1 CAGTGGGCATCAAACGGCATTGA M H A D F L M 1 2100 GCAAAGTCCTCACTCCAATCGA A F D E S W D 1 AAGGCACACACTCGGTCATAGG F A C V R D Y 1 AACTTGGGCTCAAAAGGACTTA L K P E F P S 1 2450 ATTTCCCCAATCGACGCACACAT I E R I S A G 1 GCAAAGTTCGACGCCAATCGACGCACAT I E R I S A G 1 GCACACACACGCAATCGACGCACAT I E R I S A G 1 GCACACACACGCACACAGCCCACAT I E R I S A G 1 GCCAAATTTCGACGCAAAACGCCACAT I L L S C C D	CCTCTTGCACGGTTTGGTCCAATTG A E Q V T Q D L J 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V I CCCTGCGCTCTTTGCCACGCCCCCA V R R E K G R G V 2000 CATGTTTTTCCTGGTAGTAAGCCCC E H K Q Q Y Y A J ATTGGAAGGCTTCGCCACAAGCAAJ F Q F A E G C A J 2250 ACCTGCCAAGGTCAAAGCGACCAA S G T F G L A V J 2350 ACCTGCCAAAGCCCAAGCGAATTGG L K G V S D R A J 2350 ACCTGCCAAAGCGGTCGAATTGG L K G V S D R A J CCCCAAAATGCCAGCGTCGAATTGG L K G V S D R A J 2350 CACGCGAAAATGCCAGCGTCGAATTGG L K G V S D R A J 2350 CCCCGAAAATGCCAGCGTCGAATTGG L K G V S D R A J CCCCAAAGTGGGTCGAATTGGCGCTGCAATTGGTGGGTCAATTTTAGA L1 TTCCCCTAAGTGGGTCCAATTGCGCGCGG E G I T G E S R D N A E W V P L	CGGGCTAGGGCAGGCAGGCAGGCAGGCAGGCAGGCAGGCA
TTGCTCCACGAATGCATCAAAGG Q E V F A D F F 1700 GCAGCGATGCTCAAGCACACATG C R H E L W V N TGCTGGATTACAGAACTGCGCACATT A P N C F Q A I GATGCCGCGCCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCTTGAACTGTTCGGJ T Q A K F Q E S 2300 CAGCTTCTTTGTCTTTTTGTCCTCG L K K T K K H E ATACTTGCGAACGGTGTTCCTGGA Y K R V T N R S <u>EbS</u> ATTTCTCCCGCCATATCCAGACGC TGGTTGATCGCTCCTCGCGCCTC H N I A E E G S GCAATTGGGAAGGCCGGCCGCCGC	CATTOGGTAGCCCCTATACAACAAAAAAAAAAAAAAAAAA	TCGCCTQTTCATCGGCAAAGG E G Q E D A F AAGTCTGGTGCCCCTTGCCACA F D P A G Q W 1 1850 ATGGCAGGGCATCAAACGGCTGATTGA M A T F R Q N CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCACTCCAATCGA A F D E S W D \cdot AAGGCACACACTCGGTCATAGG F A C V R D Y \cdot AAGGCACACACTCGGTCATAGGACTTA L K P E F P S \cdot 2450 ATTTCCGCAATCGACGACAATCTT L2 2700 CCGCCATCAGCAGGCACAATCTT L2 2700 CCGCCACACGAGTGAGCGCACATC A M L L H A L TGGCCAGCAGTGAGCGCACATC R A L L S C C D	CCTCTTGCACGGTTTGGTCCAATTG A E Q V T Q D L I 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V I CCCTGCGCTCTTTGCCACGCCCCAA V R R E K G R G V 2000 CATGTTTTTGCTGGTAGTAAGCCCC E H K Q Q Y Y A I ATTGGAAGGCTTCGCCCAAGCAAA F Q F A E G C A I 2250 ACCCGGTAAAGCCCAAGCAAAGGACCAA S G T F G L A V I 2350 ACCTGCCAAGCGTGTCTCGCGCGCGAATTGG L K G V S D R A I CGCGAAAATGCCACGGGTCGAATTGG L K G V S D R A I 2350 ACCTCGCTAAGTGGGTCGAATTGG CGCGAAAATGCCACGCGGCGAATTGG L K G V S D R A I 2350 ACCTCGCTAAGTGGGTCGAATTGG L L LI TTCCCCTAAGTGGGTCGCACCGCGCGGCG E G I T G E S R 2850 D N A E W V P L 2950	CGGGGTGGGGCAGGCAGGCTGGGCCAAGCGCTTT E P H R L E S W L A K 1800 CCTCTCCCAATCTGGCCTTTCTCCCAACCGA K E I Q G K E W G S 1900 CGAATCCACAGGGGTCTTCATGTTGTCGTA X S D V A T K M N D Y AAGGCAAAAGGCCCGGCTGTGGGCCAACTT R L V F A R S H A L K 2150 CGGCAAGGGATGAAACACCTTGCTGGCAA F R L P I F C G K G S ACCCGTACATGCTGGCCAGGTTCTGGG CGTCAATCGGGTCAACCGCGTTATCCACTTGGA P Y Q P E V V K S Q L 2500 CGCCAATCCCAGGGTCGCACGGTTATCCACTTGGA P Y Q P E V V K S Q L 2500 CGTCCAATCGCGGTCAACCACCTTGCTTGGCA P Y Q P E V V K S Q L 2500 CGCCAATATCGCCAGGTTATCCACTCTGG A S L I A V N I M ISTA -35 CGCCAATCCAACGGGAGGACATGTCCAGGGTTCTCGGG T L L Y R A M D L T CATCAGGCGTGAAGGGATTTTCCAACTGG M M P E P R N E L Q ECORI . 3000
TTGCTCCACGAATGCATCAAAGG Q E V P A D P P 1700 GCAGCGATGCTCAAGCACCACATG C R H E L W V N TGCTGGATTACAGAACTGCGCACT A P N C F Q A I GATGCCGCGCCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCTTGAACTGTTCGGJ T Q A K F Q E S 2300 CAGCTTCTTTGTCTTGAACTGTTCGGJ Y K R V T N R S <u>LDS</u> ATTTCTCCCGCACGATGCTCCTGGACGCC A I P S P R R L 2900 TCATCGGAGCGATGGCTAGGTAGCTAGGTA	CATTOGGTAGCGCCATCAAC A N P L A M L CCAAATCAGCAAGGCTTTGA N L D A L S Q CAAACAGGTAGTGGCTGACC D F L Y H S V 1950 SphI CGAAGATTGAAAGGCTATCGG 7 Q K G A I R CCTOTTTGCCCGCTATCGGG ATCTTTCCCCGCTATCGGG ATCTTTCCCCGCTATCGGGCT S D K W Q R A CGGTGGAGAGGCCACTGCCTT E T S L W Q R A 2550 SGAACAGTGTCATACGTGG CTCCCACGGCGACGACGGCGAACC P L S A A F SGCGTCCCGCGTGCCTACCC	TEGECTTGTTCATCGGCAAAGG E G Q E D A F AAGTCTGGTGGCCCCTTGCCACA F D P A G Q W 1 1850 ATGGCAGTGAACCGCTGATTGA M A T F R Q N CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCACTCCAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCACTCCAATCGAC A F D E S W D 1 4 AAGGCACACACTCGATCGATGACAATCTT I E R I S A G AC V R D Y 1 AACTTGGGCTCAAAAGGACTTA L K P E F P S 2700 ATTTCCCGAATCGACGCAGCAATCTT I E R I S A G A M L L H A L TGGCCAGCAGCGAGTGAGCAGCAATC R A L L S C C D CCAACCAACGGATCGCGCAGTC	CCTCTTGCACGGTTGGTCCAATT A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAC V R R E K G R G V 2000 CATGTTTTGCTGGTAGTAAGCCCC E H K Q Q Y Y A D ATTGGAAGGCTTCGCCACAAGCAA F Q F A E G C A D 2250 ACCTGCCAAGTGCCCCACGCGCGCGCGCGCGCGCGCGCGC	Considering the set of the set o
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P P 1700 GCAGCGATGCTCAAGCACCAACATT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCATCTGAACTGTTGAC F Q A I Q L K V GGTTTGCGCCATGAACTGTTCGAP T Q A K P Q E S 2300 CAGCTTCTTTGTCTTGAACTGTTCGAP T Q A K P Q E S 2300 CAGCTTCTTTGTCTTTTTGTCTC L K K T K K H F ATACTTGCGAACGGTGTTCCTGGA Y K R V T N R S <u>L'DS</u> ATTTCTCCCGCCATATCCAGACGCC TGGTTGATCGCTTCCTCGCCCCTCG H N I A E E G S GCAATTGGGAAGGCCGGCCAGC A I P S P R R L 2900 TCATCGGAGCGGATGGTAGGTAGGTAGGTAGGTAGGTAGG	CATTOGGTAGCGCCATCAAC A N P L A M L CAAATCAGCAAGGCTTTGA N L D A L S Q CAAACAGGTAGTGGCTGACC D P L Y H S V 1950 SphI CGAAGATTTGAAAGGCATGC G F I Q P A H CCTOTTTGCCCGCTATGCGG 7 Q K G A I R 2200 AATCTTTCCCTGTGGGCGATGCCTT 5 L G S R R A 2550 GGAAGCGCCGCTGCCTTCGTCGGGCT S L G S R R A 2550 GGAAGCGCGCGCGACGCGAACC P L S A A F GGCGGCGCGCGCCCTACCC A D R T G V	TCGCCTTGTTCATCGCCAAAGG E G Q E D A F AAGTCTGGTGGCCCCTTGCCACA F D P A G Q W 1 1850 ATGGCAGGAACGCGCGATTGA M A T F R Q N CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCACTCCATCGA A F D E S W D 1 AAGGCACACACTCGGTCATAAG F A C V R D Y 1 AAGGCACACACTCGATCGACGCACTTA L K P E F P S 1 2450 ATTTCCCGAATCGACGCACAATCTT I E R I S A G 1 GTCAAATTTCGACGCAAATCTT L2 2700 CCGCCATCAGCAGGAGCAGCAATCT R A L L S C C D CAACCAACGGGATGCGCAGTTC G V L P I R L E	CCCCTGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	Considering the second
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P P 1700 GCAGCGATGCTCAAGCACACACAT C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C F Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAAGTTTGAC F Q A I Q L K V GGTTTGCGCCATGAACTGTTCGAA T Q A K F Q E S 2300 CAGCTTCTTTGTGCTTGTTTTTGTGCTT L K K T K K H I ATACTTGCGAACGGTGTTTCCTGGA Y K R V T N R S 2550 TGGTTGTCCCCGCCTATCCAGACGCC 4 I P S P R R L 2900 TCATCGGAACGCGATGGTAGGTAGGTAG	CATTOGGTAGCGCATCAACA A N P L A M L CAAATCAGCAAGGCTTTGA V L D A L S Q CAAACAGGTAGTGGCTGACC D P L Y H S V 1950 SphI CGAAGATTTGAAGGCATGC G F I Q P A H CCTGTTTGCCCGCTATCCGG 7 Q K G A I R 2200 AATCTTTCCACGGCGAGAGCCACTGCCTT S L G S R R A 2550 CGGTGGAGAGGCCAATGCCGG E V A V I A 2800 CGGGAGCGAAGGCGAAGCCCAACC P L S A A F SGCTGCGCCGCGCGCCACCC A D R T G V . SphI .	TCGCCTTGTTCATCGCCAAAGG E G Q E D A F AAGTCTGGTGCCCCTTGCCACA F D P A G Q W 1 1850 ATGGCAGTGAACCGCTGATTGA M A T F R Q N CAGTGGGCATCAAACAGCATTT W H A D F L M 2100 GCAAAGTCCTCACTCCAATCGAC A F D E S W D AAGGCACACACTCGGTCAAAAGGACTTA A F D E S W D AAGGCACACACTCGGTCAAAAGGACTTA L K P E F P S 1 2450 ATTTCCGCAATCGACGCACATTT L R I S A G GTCAAATTTCGACGCAAATCTT L2 2700 CCGCCATCAGCAAGTGAGCAGCAATCT R A L L S C C D CAGTGGCAACACGGGATGCGCAGTTC G V L P I R L E 3050	CCTCTTGCACGGTTTGGTCCAATTG A E Q V T Q D L D 1750 GGCTTTGGCGGGGAATCCTGCACGT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCAC V R R E K G R G V 2000 CATGTTTTGCTGGTAGTAAGCCCC E H K Q Q Y Y A D ATTGGAAGGCTTCGCCACAAGCAAJ F Q F A E G C A D 2250 ACTTGCCAAGCTGTCTCGCGCAAGCAAJ S G T F G L A V D 2350 ACTTGCCAAGCTGTCTCGCGCTGG L K G V S D R A D 2350 ACTTGCCAAGCTGTCTCGCGCTGG L K G V S D R A D 2350 ACTTGCCAAGTGGGTGAATTGTAGA CGCGGAAAATGCCAGCGTGGAATTGT CGCGAAAATGCCAGCGTGCAATTGG D R F H W R R I D 10 1 2600 TACCCTAAGTGGGTGCAATTGTAGA CL1 TTCCCCTAAGTGGGTGCAATTGTAGA CL1 TTCCCCTAAGTGGGTCCAATCGGCAGCAG D N A E W V P L 2950 GTTGCCGAGGAAGCGCAGCAGGCAGGGTTT N G L F R L L N	Considering the set of the set o
TTGCTCCACGAATGCATCAAAGGC Q E V P A D P P 1700 GCAGCGATGCTCAAGCACACAATGC C R H E L W V N TGCTGGATTACAGAACTGCGCATC A P N C P Q A I GATGCCGCGCTTGGGAATGCCACC I G R K P I G C 2050 AAACTGGGCAATCTGAACTGTTCGAA F Q A I Q L K V GGTTTGCGCCTTGAACTGTTCGAG T Q A K P Q E S 2300 CAGCTTCTTTGGCTTTGTGTTTTTGTGCTT L K K T K K H F ATACTTGCGAACGGTGTTCCTGGA Y K R V T N R S <u>KDS</u> ATTCTCCCCGCCATATCCAGACGCC 2650 TGGTTGATGCGTCTCCTCGCCGCTC A I P S P R R L 2900 TCATCGGAAGGCAGGCGCGCGCGCGCGCAGG A I P S P R R L 2900 TCATCGGACGTTGTGCAGCTCGTCG	CATTOGGTAGCGCCATCAAC A N P L A M L CAAACAGGTAGCGCATGGC CAAACAGGTAGTGGCTAGCG D P L Y H S V 1950 SphI CGAAGATTTGAAGGCTATGCG 7 Q K G A I R . 2200 AACTOTTTGCCCCGCTATGCGGC 7 Q K G A I R . 2200 AACTOTTGCCCGCGCATGCGGC A D R R A 2550 CTCCCACGGCGACGATGCCCTG C V A V I A . 2800 CGGGAGCGAAGGCCAACCGC P L S A A F . 2800 CGGGAGCGAAGGCCAACCGCCAACC P L S A A F . 2800 CGGGAGCGAAGGCCAACCGCCAACC P L S A A F . 2800 CGGGAGCGAAGCCCCACCCCACCC A D R T G V . SphI . SphI	TEGECTTOTTCATEGGCAAAGG E G Q E D A F AAGTETGGTGGCCCCTTGGCCACAA F D P A G Q W 1 1850 ATGGCAGTGAACCGCTGATTGA M A T F R Q N \sim CAGTGGGCATCAAACAGCATTT W H A D F L M 1 2100 GCAAAGTCCTCACTCCAATGAA A F D E S W D \sim AAGGCACACACTCGGTCATAGG F A C V R D Y \sim AACTTGGGCTCAAAAGGACTTA L K P E F P S \sim 2450 MITTCCCGGAATGGACGCACTTA I E R I S A G GTCAAATTTCGACGCAAATCTT L2 2700 CCGCCCTCACGCAGTGGCGCAATCTT L2 2700 CCGCCCTGCGAGTGGCGCAGTTCGACGCACTTC A L L S C C D CCGACCTTGCGCCAGTGCCGAGTGCCAGTGCGCAGTGCCG A A M L L H A L TGGCCAGCGGCAGTGCGCAGTTCCG G V L P I R L E 3050 CCGACCTTGCGCCAGTGCGAGTGCGAGTGCGAGTGCGAGTGCGGC	CCTCTTGCACGGTTTGGTCCAATTG A E Q V T Q D L D 1750 GGCGTTGGCGGGAATCCTGCACGTT L R Q R S D Q V D CCCTGCGCTCTTTGCCACGCCCCAA V R R E K G R G V 2000 CATGTTTTGCTGGTAGTAAGCCCC E H K Q Q Y Y A D ATTGGAAGGCTTGCCCAAGCAAGCAAGCAA F Q F A E G C A D 2250 ACTTGCAAGCCAAGCCAAGCGAATTGG L K G V S D R A D 2350 ACTTGCCAAGCGCTGCTCGCGCGCGAATTGG CGCGGAAAATGCCACGGTCGAATTGG D R F H W R R I D 2350 ACTTGCCAAGCGTGTCTCGCGCGCGA L K G V S D R A D 2350 ACTTGCCAAGCGCGCAGCGCGCGAATTGG D R F H W R R I D 2350 TACCCTAAGTGGGCCCAATCGCGCGCG L K G V S D R A D 2350 TACCCTAAGTGGCCCCCATACCGGCAGC D R F H W R R I D 2350 TACCCTAAGTGGCCCCCCATACCGGCAGC D N A E W V P L 2950 GTTGCCGAGGAAGCGCCGCGCAGCAGGTTT N G L F R L L N ACCCGGTTGCTCCCTTTCCGGCCCAA	Considering the set of the set o

FIG. 2. Nucleotide sequence of In0, including the sequence of IS1326. The sequence shown extends from the *Bam*HI site commencing at base 1876 of the 3'-CS (13). The sequence of the corresponding fragment of pSCH884 is identical, except that pSCH884 contains an extra 359 bp located at the position indicated by the arrow between bases 151 and 152 (13) and a deletion of 439 bp (boxed). The terminal 26-bp inverted repeats of IS1326 are shaded, and repeats of 22 bp found near the ends of the insertion sequence are indicated by arrows and designated R1, R2, L1, L2, and L3. Potential ribosome-binding sites (rbs) and promoter -10 and -35 regions are indicated. Diamonds indicate the conserved acidic residues of the D, D(45), E motif. The vertical arrow at bases 2594 and 2595 marks the position of insertion of IS1353 in In2 (see Fig. 3 and 4).

elements produce three polypeptides, ORFA, ORFB, and fusion protein ORFAB. ORFAB, which appears to be the active transposase for IS911 (29), is a transframe product, resulting from a translational -1 frameshift that occurs in the region where orfA and orfB overlap. Specific signals induce the -1frameshift; a heptanucleotide motif, AAAAAAG, in the overlap region; and a downstream sequence capable of forming a strong secondary structure, such as a stem-loop structure (29, 48). These signals are also present in many other members of the IS3 family (3, 21, 49), and in the region of overlap between ORFA and ORFB in IS1353, the potential frameshift motif A₆G is present and a pair of inverted repeats that have the potential to form a stem-loop structure are found downstream (Fig. 4). This suggests that IS1353, in common with other members of the IS3 family, also utilizes translational frameshifting. It is highly likely that in IS911, translation of ORFB is initiated at an ATT codon that lies immediately upstream of the A_6G motif (29). In IS1353, there is also an ATT codon 3 bp upstream of the A_6G motif; however, an ATG codon is also present six codons downstream, and which codon is in fact used for initiation of ORFB needs to be determined experimentally.

DISCUSSION

The sequence data presented here show that integrons In0, In2, and In5 all contain IS1326 and are also otherwise closely related (Fig. 3). The two differences between In0 and In2 are readily explained by known mechanisms for DNA insertion.



1 kb

FIG. 3. Structures of In0, In2, In5, and Tn402. The 5'-CS is shown as an open box, the 3'-CS is a solid box, and the *tni* module is a bold hatched box. The region of the Tn402 *qacE* gene that is not present in the 3'-CS is stippled. Vertical arrows indicate the positions of the integrated cassettes (in parentheses). IS1326 is represented by a narrow hatched box, and IS1353 is a solid line. The regions shown are contiguous, and the gaps were introduced only to align identical sequences in the four integrons shown. The extents of IS1326-mediated deletions in In0, In2, and In5 are indicated below the maps. Arrowheads indicate the terminal inverted repeats of the integrons (designated 1) and IS1326 (designated 2). Abbreviations for restriction sites are the same as in Fig. 1.

ATCACTCCTTGATTTCTCCCGCCATATCCAGACGGGAAACAGTGTCATACGTGGGTCAAATTTCGACGCAAATCTTTACCCTAAGTTGGGGTGCGGACAAAATCTTGGACTACTTTAGGA -35 rbs	120
orfa M Y S Y E D R L R A V R L Y L K L G R R M S A T L R Q L G Y P T K N S L K A <u>GTAGTTCATGTATGCGATGAGAGATCGCCTTCGAGCCGTGAGGTTGTACCTGAAGGC</u> <u>-10</u> <i>Hin</i> dIII	240
W L À E F E R N Q D L R R G Y Q R I K R Q Y T D E Q K Q R A V D H Y I E Q G Y C CTGGTTGGCA <u>GAATTC</u> GAACGGAATCAGGATCTTCGCCGAGGGCTATCAACGGATAAAACGGCAGTGAGCAAAGGCAAAGGCAACGGGCAGTAGATCACTATATCGAACAAGGCTACTG <i>Eco</i> RI	360
LSHTIRSLGYPSREALRAWIRDLRLNSLGRSSAAALPQSP CCTGAGTCACACAATCCGAAGCCTGGGCTACCCCAAGCCGCGGGGGCCTGGGGGCGGGGGGGG	480
A L A S R S K Q A V I A L N L R V G S A K D V A D T V G V S R P T L Y N W Q H R CGCTCTCGCCTCGAGAAGCAAGCAAGCCGTCATTGCACTGAACCTGCGCGTAGGTTCGGCAAAGGATGTGGCCGACACTGTCGGTGTATCGCGACCAACGTTGTATAACTGGCAGCATCG	600
LLGKVPLKPMTKKKGDTSLEQRHEALLRELAELESQNQRL ATTACTTGGCAAAGTGCCCCTAAAACCCATGACAAAGAAGAAGGTGACACCTGGCCGGACCGGGCATGAGGCACTAGCGGGCACTGGGGGACCGGGGCGGGACCGGGCGGC	720
OTÍBIDKKDMGINPLELTSREKTKVVDALRVT RMENAILEKASELIKKTWASTPSN* TCGCATGGAGAATGCAATTCTGGAGAAGGCGAGTGAATTGATAAAAAAGACATGGGCATCAACCCCTCGAACTGACAAGCCGAGAAAAAACGAAGGTGGTGATGACCCCCCAGAGTCACC	840
FPLANLLCGLKLARSTYFYQRLRQTRPDKYTQVREVIRTI TTTCCATTAGCCAATCTGTTGTGCGGCCTGAAGCTGGCGCGCAGCACATACTTCTATCAACGCCTGCGGCAGACGCGCCCGACAAGTACACGCAGGTGCGTGAGGTCATTCGGACTATC	960
FEDNYRCYGYRRIDSALRLGGMRVSEKVVRRLMAQERLVV TTCGAGGACAACTACCGCTGCTATGGCCATGGACGCATTGATAGTGCCTTGGCGCCTTGGTGGCGTGCGT	1080
R T P R P N L L N R D F H A S A P N L L N R D L A A A A A T D L A A A A A T D L A A A A A A T D L A	1200
TEIHIPAGKVYVSPIVDCFDGLVVAWNIGTSPDANLVNTM ACGGAAATACACATTCCGGCAGGGAAGGTCTACGTCTGCGATCGTCGATGGCTGGTGGTGGTGGTGGCCTGGAATATCGGCACCAGCCCGGATGCGAACCTGGTCAATACCATG	1320
L D H A V R T L R P G E H P V I H S D R G S H Y R W P A W I R R T E N A Q L T R CTGGATCACGCGGTACGGACACTGCGACCCGGTGAGCATCCGGTTATCCATTCGGACAGGGGGCTCGCATTATCGCTGGCCTGCGTCGCCGCACTGAAAATGCCCAATTAACGCGG BamhI	1440
SMSKKGCSPDNAACEGFFGRLKTELIYPRNWQHVTLKDLM TCGATGTCCAAAAAGGGCTGCTCGCCAGACAATGCT <u>GCATGC</u> GAGGGGCTTTTTCGGACGATGAAGACCGAACGAACGAACG	1560
T R I D A Y I H W Y N E R R I K V S L G G R S P I E Y R H A V G L M S V * ACGCGAATCGATGCCTATATCCACTGGTACAACGAGCGCCGCATCAAAGTGTCGCTTGGCGGGCG	1680
GAAATCGTCCGCACCCCCAGTGGGTCAATTTTAGATGCAACTCAACAGGCCATGCTGAGTGTGCGATGGTTGATCGCTTCCTCGCCGCTCTCCACGGCGACGATGGCCGCCGCCATCAGC	1800
AAGTGCGCCAGTTCCCCTATGGTGCCCTCGCT	1832

FIG. 4. Nucleotide sequence of the In2 region containing IS1353. The IS1353 inverted repeats are boxed. The potential heptanucleotide frameshift signal A_6G is underlined, and the inverted repeats of the downstream potential stem-loop structure are indicated by arrows. Diamonds indicate the conserved acidic residues of the D, D(35), E motif. The terminal inverted repeat from IS1326 is underlined with an arrow.

Whereas In0 contains no integrated cassettes (1), In2 contains the *aadA1* cassette (44) that has been integrated at *attI* by the action of the integron integrase IntI1 (4). The further extra segment in In2 corresponds to an acquired insertion sequence, IS1353. There are only three differences between In0 and In5, one of which is the presence of an *aacA* cassette integrated at *attI* in In5 (40). The remaining differences are both deletions with one endpoint at an IS1326 boundary and can be explained by invoking deletion events originating at the boundaries of IS1326. These events are presumably mediated by the action of the insertion sequence-encoded transposition functions in a manner analogous to the formation of adjacent deletions by other insertion sequences and transposons (19).

Recent studies on Tn5053, which has a transposition module

closely related to that of Tn402, have demonstrated that three genes, tniA, tniB, and tniQ, are essential for transposition of Tn5053 and that a further gene, tniR, that encodes a resolvase is required for efficient resolution of cointegrate intermediates (18). Because the complete tni module found in Tn402 is not present in any of the three integrons examined here, it seems reasonable to conclude that these integrons are all defective derivatives of an ancestral transposon and are unable to direct their own movement. However, the fact that In2 and In0, which contain identical 3'-CS and tni module segments, are found in independent locations, in Tn21 and in pVS1, respectively (see reference 13), suggests that movement can occur. As movement of Tn5053 derivatives with defects in any or all of the tniA, tniB, and tniQ genes could be restored by supplying



FIG. 5. Evolution of In0, In2, and In5 from an ancestral, transpositionproficient integron similar to Tn402. The 5'-CS, 3'-CS, transposition modules, and IS*I326* are indicated as in Fig. 3. Only Tn402 is shown as containing integrated gene cassettes, as any cassette configuration at the position indicated by a vertical arrow can be readily achieved by IntI-mediated recombination. InX and InY are hypothetical intermediates.

the defective function(s) in *trans* (18), it is likely that integrons such as In0, In2, and In5, which retain intact ends, could be mobilized when a transposon with a complete *tni* module, such as Tn402 or Tn5053, is present in the same cell. Indeed, movement of In0 and In2 in the presence of mercury resistance transposon Tn502 (41) has recently been demonstrated (40a).

A model for the evolution of In0, In2, and In5 from an ancestral transposition-proficient class 1 integron similar to Tn402 is shown in Fig. 5. Acquisition and loss of gene cassettes are well documented (4–6, 14, 31) and can occur at any time. For simplicity, cassette exchange events have been assumed to occur last. The ancestor, represented by Tn402, consists of four distinct domains: the 5'-CS encoding the integrase responsible for the movement of the gene cassettes, the variable region which contains the integrated gene cassettes (*dfrB3* and orfD in Tn402), the *qacE* gene region, and the transposition module which includes the *tniA*, *tniB*, *tniQ* (orf6), and *tniR* (*tniC*) genes (18, 30).

The proposed transposition-proficient ancestor of the *sull* family of class 1 integrons contains the complete *tni* module and is derived from the Tn402 ancestral configuration by integration of the *sull* gene and other 3'-CS sequences downstream of *qacE* and deletion of the end of the *qacE* gene to generate $qacE\Delta I$, which is found in the 3'-CS. The exact mechanism of the evolution of the 3'-CS module remains a matter for spec-

ulation, but the 3'-CS may represent a particular series of gene cassettes that have become fixed. We have previously suggested that the *sulI* gene may have originally been part of a cassette which was integrated at *attI* and subsequently fixed by loss of the functional 59-base element recombination site (43). The same scenario has recently been suggested for the origin of the $qacE\Delta I$ gene (30), and the fact that the qacE gene of Tn402 has since been shown to be part of a functional mobile gene cassette (31a) supports this conclusion.

Insertion of IS1326 into the ancestral *sull*-containing integron is assumed to have occurred either within a further, unidentified, region of the 3'-CS (dotted region in Fig. 5) or within the *tni* module but between the positions shown by the horizontal bar in Fig. 5. Thereafter, deletion events originating at the ends of IS1326 and leading to loss of known 3'-CS sequences from In0 and In2 and to loss of part of the *tni* module in In0, In2, and In5 have occurred. To explain the different extents of 3'-CS and *tni* modules retained in In5 and In0 (In2), it is necessary to invoke at least two insertion sequence-mediated events in the evolution of one of them. The acquisition of IS1353 by In2 is presumed to be a recent event (see below).

Evolution of the Tn21 transposon family. By using information from this study, together with data from previously published restriction maps of Tn21 and related transposons and heteroduplex analyses involving pairs of transposons from this family (11, 20, 24, 25, 38, 47), it is possible to construct a coherent model for the evolution of members of this family (Fig. 6). Only close relatives of Tn21 identified by the presence of the characteristic motif of EcoRI and HindIII sites found in the Tn21 transposition gene (tnpA, tnpR, and res) region are considered here as members of the Tn21 family. This group includes a little over half of the transposons listed as belonging to the Tn21 subgroup of the Tn3 transposon family in a recent review (see Table 1 in reference 11). Tn1401, Tn1406, and Tn1409 were not considered because the most recently published maps of these transposons (24) are not sufficiently accurate to permit analysis. One further transposon, Tn5086, that has subsequently been identified as a member of this group (45) was considered.

In the simplest evolutionary tree consistent with all of the published data, the first step in the evolution of this family is the insertion of an integron identical to In0 (or equivalent to In0 but containing one or more integrated gene cassettes) into an ancestral mercury resistance transposon, possibly Tn2613 (47). As most members of this group include the *aadA1* cassette, for simplicity it has been assumed that this cassette was present in the incoming integron. Tn2411, rather than Tn21, is assumed to be the true ancestor of the integron-containing members of this family because the acquisition of IS1353 by In2 (in Tn21) appears to be a recent event that occurred after the movement of the integron to its present location. The region corresponding to IS1353 is not present in transposons that are otherwise very closely related to Tn21, such as Tn4 and Tn2411 (20, 25, 47), and it has been suggested that this region was acquired in Japan, as it is generally present in Tn21-like transposons isolated in Japan (11).

Most of the differences between Tn2411 and the known transposons in this family can be explained by site-specific integration of further cassettes into the integron or insertion of further transposons (e.g., Tn3 in Tn4 and IS1353 in Tn21) at various positions. Two further differences can be explained by IS1326-mediated events. The large deletion in Tn1831 (25) originates close to one boundary of IS1326, extends beyond the right-hand end of the integron into the *mer* region, and is likely to be due to a further IS1326-mediated deletion event. Trans-



FIG. 6. Evolution of the Tn21 transposon family. The relationships of all of the transposons in this group (see reference 11 for references) can be explained by invoking the insertion of an integron with the structure shown followed by subsequent events including integration, excision, or exchange of gene cassettes, insertion of transposons, or insertion sequences, insertion sequence-mediated deletions, or insertion sequence excision.

poson Tn5086 contains all of the bases to the right of IS1326 in In2 and all but the last four bases to the left (TATC), which are replaced by two bases (CA) (45). The simplest explanation for this configuration is excision of IS1326 with imprecise rejoining and repair of the ends generated. A similar configuration is found in Tn2608 (47), which may also have arisen by this route.

Other members of the Tn21 subgroup are also known to include integrons, the best known being Tn1696, which contains the integron In4 (13, 43). However, Tn1696 is not a close relative of Tn2411 and the integron is located at a different position (38) and has a different structure, which includes IS6100 (13). Thus, it appears that Tn1696 has a different evolutionary history and that integrons have been inserted into ancestral transposons containing *tnp* and *mer* modules on more

ACKNOWLEDGMENTS

We thank Vilma Stanisich for communicating data prior to publication. Access to DNA databases and sequence analysis programs and assistance in their use were provided by the Australian National Genomic Information Service (A.N.G.I.S.).

This work was supported by a grant from the National Health and Medical Research Council.

REFERENCES

- Bissonnette, L., and P. H. Roy. 1992. Characterization of In0 of *Pseudomonas aeruginosa* plasmid pVS1, an ancestor of integrons of multiresistance plasmids and transposons of gram-negative bacteria. J. Bacteriol. 174:1248–1257.
- Brown, N. L., T. K. Misra, J. N. Winnie, A. Schmidt, M. Seiff, and S. Silver. 1986. The nucleotide sequence of the mercuric resistance operons of plasmid R100 and transposon Tn501: further evidence for mer genes which enhance the activity of the mercuric ion detoxification system. Mol. Gen. Genet. 202:143–151.
- 2a.Brown, H. J., and R. M. Hall. Unpublished data.
- Chandler, M., and O. Fayet. 1993. Translational frameshifting in the control of transposition in bacteria. Mol. Microbiol. 7:497–503.
- Collis, C. M., G. Grammaticopoulos, J. Briton, H. W. Stokes, and R. M. Hall. 1993. Site-specific insertion of gene cassettes into integrons. Mol. Microbiol. 9:41–52.
- Collis, C. M., and R. M. Hall. 1992. Site-specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. J. Bacteriol. 174:1574–1585.
- Collis, C. M., and R. M. Hall. 1992. Gene cassettes from the insert region of integrons are excised as covalently closed circles. Mol. Microbiol. 6:2875– 2885.
- Dear, S., and R. Staden. 1991. A sequence assembly and editing program for efficient management of large projects. Nucleic Acids Res. 19:3907–3911.
- de la Cruz, F., and J. Grinsted. 1982. Genetic and molecular characterization of Tn21, a multiple resistance transposon from R100.1. J. Bacteriol. 151: 222-228.
- Diver, W. P., J. Grinsted, D. C. Fritzinger, N. L. Brown, J. Altenbuchner, P. Rogowsky, and R. Schmitt. 1983. DNA sequences of and complementation by the *tnpR* genes of Tn21, Tn501 and Tn1721. Mol. Gen. Genet. 191:189– 193.
- Fayet, O., P. Ramond, P. Polard, M. F. Prère, and M. Chandler. 1990. Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences? Mol. Microbiol. 4:1771–1777.
- Grinsted, J., F. de la Cruz, and R. Schmitt. 1990. The Tn21 subgroup of bacterial transposable elements. Plasmid 24:163–189.
- Hall, R. M., D. E. Brookes, and H. W. Stokes. 1991. Site-specific insertion of genes into integrons: role of the 59-base element and determination of the recombination crossover point. Mol. Microbiol. 5:1941–1959.
- Hall, R. M., H. J. Brown, D. E. Brookes, and H. W. Stokes. 1994. Integrons found in different locations have identical 5' ends but variable 3' ends. J. Bacteriol. 176:6286–6294.
- Hall, R. M., and C. M. Collis. 1995. Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. Mol. Microbiol. 15:593–600.
- Hall, R. M., and H. W. Stokes. 1993. Integrons: novel DNA elements which capture genes by site-specific recombination. Genetica 90:115–132.
- Hall, R. M., and C. Vockler. 1987. The region of the IncN plasmid R46 coding for resistance to β-lactam antibiotics, streptomycin/spectinomycin and sulphonamides is closely related to antibiotic resistance segments found in IncW plasmids and in Tn21-like transposons. Nucleic Acids Res. 15:7491– 7501.
- Itoh, Y., J. M. Watson, D. Haas, and T. Leisinger. 1984. Genetic and molecular characterization of the *Pseudomonas* plasmid pVS1. Plasmid 11:206– 220.
- Kholodii, G. Y., S. Z. Mindlin, I. A. Bass, O. V. Yurieva, S. V. Minakhina, and V. G. Nikiforov. 1995. Four genes, two ends, and a *res* region are involved in transposition of Tn5053: a paradigm for a novel family of transposons carrying either a *mer* operon or an integron. Mol. Microbiol. 17: 1189–1200.
- 19. Kleckner, N. 1977. Translocatable elements in procaryotes. Cell 11:11-23.
- Kratz, J., F. Schmidt, and B. Wiedemann. 1983. Characterization of Tn2411 and Tn2410, two transposons derived from R-plasmid R1767 and related to Tn2603 and Tn21. J. Bacteriol. 155:1333–1342.
- Kropinski, A. M., M. A. Farinha, and I. Jansons. 1994. Nucleotide sequence of the *Pseudomonas aeruginosa* insertion sequence IS222: another member of the IS3 family. Plasmid 31:222–228.
- 22. Kulkosky, J., K. S. Jones, R. A. Katz, J. P. G. Mack, and A. M. Skalka. 1992.

Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol. Cell. Biol. **12**:2331–2338.

- Menou, G., J. Mahillon, M.-M. Lacadet, and D. Lereclus. 1990. Structural and genetic organization of IS232, a new insertion sequence of *Bacillus thuringiensis*. J. Bacteriol. 172:6689–6696.
- Mercier, J., J. Lachapelle, F. Couture, M. Lafond, G. Vézina, M. Boissinot, and R. C. Levesque. 1990. Structural and functional characterization of *Inpl*, a recombinase locus in Tn21 and related β-lactamase transposons. J. Bacteriol. 172:3745–3757.
- Meyer, J. F., B. A. Nies, J. Kratz, and B. Wiedemann. 1985. Evolution of Tn21-related transposons: isolation of Tn2425, which harbours IS161. J. Gen. Microbiol. 131:1123–1130.
- Paulsen, I. T., T. G. Littlejohn, P. Rådström, L. Sundström, O. Sköld, G. Swedberg, and R. A. Skurray. 1993. The 3'-conserved segment of integrons contains a gene associated with multidrug resistance to antiseptics and disinfectants. Antimicrob. Agents Chemother. 37:761–768.
- Podladchikova, O. N., G. G. Dikhanov, A. V. Rakin, and J. Heesemann. 1994. Nucleotide sequence and structural organization of *Yersinia pestis* insertion sequence IS100. FEMS Microbiol. Lett. 121:269–274.
- Polard, P., and M. Chandler. 1995. Bacterial transposases and retroviral integrases. Mol. Microbiol. 15:13–23.
- Polard, P., M. F. Prère, M. Chandler, and O. Fayet. 1991. Programmed translational frameshifting and initiation at an AUU codon in gene expression of bacterial insertion sequence IS911. J. Mol. Biol. 222:465–477.
- Rådström, P., O. Sköld, G. Swedberg, J. Flensburg, P. H. Roy, and L. Sundström. 1994. Transposon Tn5090 of plasmid R751, which carries an integron, is related to Tn7, Mu, and the retroelements. J. Bacteriol. 176: 3257–3268.
- Recchia, G. D., and R. M. Hall. 1995. Gene cassettes: a new class of mobile element. Microbiology 141:3015–3027.
- 31a.Recchia, G. D., and R. M. Hall. Unpublished data.
- Recchia, G. D., H. W. Stokes, and R. M. Hall. 1994. Characterisation of specific and secondary recombination sites recognised by the integron DNA integrase. Nucleic Acids Res. 22:2071–2078.
- Reimmann, C., and D. Haas. 1990. The *istA* gene of insertion sequence IS21 is essential for cleavage at the inner 3' ends of tandemly repeated IS21 elements *in vitro*. EMBO J. 9:4055–4063.
- Reimmann, C., R. Moore, S. Little, A. Savioz, N. S. Willetts, and D. Haas. 1989. Genetic structure, function and regulation of the transposable element IS21. Mol. Gen. Genet. 215:416–424.
- Rezsöhazy, R., B. Hallet, J. Delcour, and J. Mahillon. 1993. The IS4 family of insertion sequences: evidence for a conserved transposase motif. Mol. Microbiol. 9:1283–1295.
- Rowland, S.-J., and K. G. H. Dyke. 1990. Tn552, a novel transposable element from *Staphylococcus aureus*. Mol. Microbiol. 4:961–975.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

- Schmidt, F., and I. Klopfer-Kaul. 1984. Evolutionary relationship between Tn21-like elements and pBP201, a plasmid from *Klebsiella pneumoniae* mediating resistance to gentamicin and eight other drugs. Mol. Gen. Genet. 197:109–119.
- Shapiro, J. A., and P. Sporn. 1977. Tn402: a new transposable element determining trimethoprim resistance that inserts in bacteriophage lambda. J. Bacteriol. 129:1632–1635.
- Shaw, K. J., C. A. Cramer, M. Rizzo, R. Mierzwa, K. Gewain, G. H. Miller, and R. S. Hare. 1989. Isolation, characterization, and DNA sequence analysis of an AAC(6')-II gene from *Pseudomonas aeruginosa*. Antimicrob. Agents Chemother. 33:2052–2062.
- 40a.Stanisich, V. A. Personal communication.
- Stanisich, V. A., R. Arwas, P. M. Bennett, and F. de la Cruz. 1989. Characterization of *Pseudomonas* mercury-resistance transposon Tn502, which has a preferred insertion site in RP1. J. Gen. Microbiol. 135:2909–2915.
- Stanisich, V. A., P. M. Bennett, and M. H. Richmond. 1977. Characterization of a translocation unit encoding resistance to mercuric ions that occurs on a nonconjugative plasmid in *Pseudomonas aeruginosa*. J. Bacteriol. 129:1227– 1233.
- Stokes, H. W., and R. M. Hall. 1989. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. Mol. Microbiol. 3:1669–1683.
- 44. Sundström, L., P. Rådström, G. Swedberg, and O. Sköld. 1988. Site-specific recombination promotes linkage between trimethoprim and sulfonamide resistance genes. Sequence characterization of *dhfrV* and *sulI* and a recombination active locus of Tn21. Mol. Gen. Genet. 213:191–201.
- Sundström, L., G. Swedberg, and O. Sköld. 1993. Characterization of transposon Tn5086, carrying the site-specifically inserted gene *dhfrVII* mediating trimethoprim resistance. J. Bacteriol. 175:1796–1805.
- Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84:4767–4771.
- Tanaka, M., T. Yamamoto, and T. Sawai. 1983. Evolution of complex resistance transposons from an ancestral mercury transposon. J. Bacteriol. 153: 1432–1438.
- Vögele, K., E. Schwartz, C. Welz, E. Schiltz, and B. Rak. 1991. High-level ribosomal frameshifting directs the synthesis of IS150 gene products. Nucleic Acids Res. 19:4377–4385.
- Walker, D. C., and T. R. Klaenhammer. 1994. Isolation of a novel IS3 group insertion element and construction of an integration vector for *Lactobacillus* spp. J. Bacteriol. 176:5330–5340.
- Wilbur, W. J., and D. J. Lipman. 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA 80:726–730.
- Xu, K., Z. Q. He, Y. M. Mao, R. Q. Sheng, and Z. J. Sheng. 1993. On two transposable elements from *Bacillus stearothermophilus*. Plasmid 29:1–9.
- Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119.