Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 942–946, February 1994
Genetics

A proposed superfamily of transposase genes: Transposon-like elements in ciliated protozoa and a common "D35E" motif

(Tc1/IS*630*/IS*3*/retrovirus/profile)

THOMAS G. DOAK*, F. PAUL DOERDER*†, CAROLYN L. JAHN‡, AND GLENN HERRICK*§

*Department of Cellular, Viral and Molecular Biology, University of Utah School of Medicine, Salt Lake City, UT 84132; and ‡Department of Cell, Molecular, and Structural Biology, Northwestern University Medical School, Chicago, Il 60611

Communicated by Elizabeth H. Blackburn, September 24, 1993 (received for review June 22, 1993)

ABSTRACT **The transposon-like elements TBE1, Tec1, and Tec2 of hypotrichous ciliated protozoa appear to encode a protein that belongs to the IS*630*-Tc1 family of transposases. The *Anabaena* IS*895* transposase also is placed in this family. We note that most family members transpose into the dinucleotide target, TA, and that members with eukaryotic hosts have a tendency for somatic excision that is carried to an extreme by the ciliate elements. Alignments including the additional members, and also *mariner* elements, show that transposases of this family share strongly conserved residues in a large C-terminal portion, including a fully conserved dipeptide, Asp-Glu (DE), and a block consisting of a fully conserved Asp and highly conserved Glu, separated by 34 or 35 residues (D35E). This D35E motif likely is homologous to the previously characterized D35E motif of the family of retroviral-retrotransposon integrases and IS*3*-like transposases. Because it is known that the IS*3*-retroposon D35E region is a critical portion of a domain capable of various *in vitro* transposition-related reactions, the results suggest that the two families share homologous catalytic transposase domains and that members of both families may share a common transposition mechanism.**

Transposons infest a wide variety of organisms and are structurally and functionally diverse (1). Transposon classifications have been based on shared hosts (prokaryotic or eukaryotic), shared structures (inverted terminal repeats or long direct terminal repeats; compound, composite, or complex), shared mechanisms of transposition (conservative or replicative, via RNA or "cut and paste"), and increasingly on shared homologous genes, usually transposases. The trend—as in the field of gene families in general—has been from an initial recognition of many small families to a progressive fusion of families into larger and fewer families. The elements involved in the present study provide a good example of this trend. Families of transposases related to *Caenorhabditis* Tc1, *Drosophila mariner*, and *Shigella* IS*630* grew in isolation (2–5) and were joined later (6, 7). The prokaryotic IS*3* transposase and eukaryotic retrotransposon-retrovirus integrase families were joined upon recognition of a common sequence motif referred to as the "D,D35E" motif. This motif includes a conserved Asp joined—by a variable-length less-conserved segment—to a "D35E" block consisting of invariant Asp and Glu residues separated by a moderately conserved segment usually 35 residues long (8, 9). Here we provide evidence that a related D35E block exists in the IS*630*-Tc1 family.

We uncovered this interfamily connection while searching for sequences similar to open reading frames (ORFs) of transposon-like elements that reside in two hypotrichous ciliated protozoa. Although these elements have not been observed to transpose, they show a variety of features indicating they are transposons: they are repetitive and have inverted terminal repeats flanked by short direct repeats, and in some cases, empty alleles of element-interrupted loci are known (for review, see ref. 10). The sequences of the 5.3-kbp *Euplotes crassus* Tec1 and Tec2 elements were recently determined (11), and we report here the sequencing of a 4.1-kbp *Oxytricha fallax* TBE1 element.¶ Each element contains multiple ORFs, one of which encodes a moderate-sized protein (380, 383, and 354 codons, respectively) that we show belongs to the IS*630*-Tc1 transposase family. The TBE1 and Tec elements are not otherwise obviously related.

The range of hosts for the aggregate IS*630*-Tc1-IS*3*-retron family is extremely broad, including prokaryotes, fungi, plants, invertebrate and vertebrate metazoa (2–9), and now ciliated protozoa.

MATERIALS AND METHODS

TBE1-fa1-1 was subcloned from a clone of *O. fallax* micronuclear DNA (12). Inserts of nested sets of unidirectional deletion clones were constructed and sequenced as described (13). ORFs were identified, translating TAA and TAG as Glu (see ref. 13); the transposase ORF extends from nt 3993 to nt 2932 in the 4073-bp sequence.

The Tec1 and Tec2 transposase ORFs extend from nt 1891 to nt 745 and nt 1908 to nt 760 on the respective element sequences (GenBank accession nos. L03359 and L03360, respectively). Other sequences were obtained from public data bases (see figure legends). Short names have been assigned, with an abbreviation of the host species name, where necessary to avoid ambiguity (e.g., Tc1 and CbTc1).

Sequence Analyses. Data base searches were performed with BLAZE (GenBank–Intelligenetics, Swiss-Prot data base release 22) and BLAST (NCBI data base May 10, 1993; ref. 14). Alignments and construction and use of profiles (15) were performed with various GCG programs (version 7.2), including EXTRACTNAMES, which allows the intersection of two lists to be identified. All alignments were constructed by PILEUP and none was altered by hand.

RESULTS

Additional Members of the IS*630*-Tc1 Family. Database searches with the putative transposase sequences of the ciliate TBE1 and Tec elements indicated their similarities with members of the IS*630*-Tc1 family. For instance in a BLAZE search with the TBE1 sequence, the *Shigella* IS*630* transposase received the top score, and in a BLAST search the *Drosophila bari-1* transposase (16) scored first, fifth, and ninth in searches with Tec1, TBE1, and Tec2, respectively.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: ORF, open reading frame.

†Present address: Department of Biology, Cleveland State University, Cleveland, OH 44115.
§To whom reprint requests should be addressed.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. L23169).

To test the implications of these anecdotal results, we used the PROFILE suite of programs to generate a "profile" or position-specific scoring table (15). In essence, a profile is a set of sequence characteristics that typifies a family of aligned sequences. A database then can in effect be searched for candidates that "fit the profile": a score is calculated for each data base sequence; a score of +1 is 1 SD above the mean of all scores. The sequences of the 5 bacterial IS*630* family transposases (5) were aligned, and a profile was generated. An augmented protein sequence database (supplemented with the TBE1 and Tec sequences) was searched to learn how the ciliate sequences fared relative to the 22 established members of the IS*630*-Tc1 family in the database. The 5 "profiled" sequences got inflated scores (>30) because they are represented in the profile (Fig. 1*A*). Although the Tec1 and Tec2 sequences did not receive impressive scores (2.26 and 2.14, respectively), the TBE1 sequence received the top score (8.71), strongly suggesting family membership (Fig. 1*A*). Similarly, the *Anabaena* IS*895* sequence (17) received an impressive score (8.65), indicating that it too may be a family member. The Tec sequences received much more promising scores when the profile was augmented by the TBE1 and IS*895* sequences (5.79 and 5.02 for Tec1 and Tec2, respectively; Fig. 1*B*). With this second profile, a variety of members of the eukaryotic Tc1 subfamily, including four *mariner* sequences, received scores in the same range as the Tec elements. This result supports Tec element membership, because the *mariner* elements are established relatives of the IS*630* subfamily (see Introduction). The *Drosophila pogo* element (18) also received a high score in this range (3.8). *pogo* has often cropped up in our database searches, but alignments with IS*630*-Tc1 family members are not sufficiently impressive to present. Since *pogo* shares target specificity with IS*630*-Tc1 members (see *Discussion*), we note a potential relationship of *pogo* to the family.

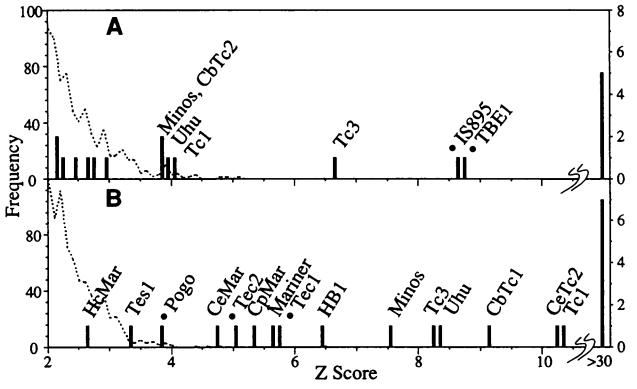

FIG. 1. (*A*) Detection of TBE1 and IS*895* transposases with an IS*630*-family profile. A profile of the protein sequences from IS*630*, StIS*630*, IS*1066*, RSa, and RIATL was used to search an augmented GenPept data base (see below). The frequency of entries with high Z scores (≥2.0) is shown as a curve (left axis). The "hit list" was searched for IS*630*-Tc1 family members and candidates (see list below); the scores of those found are shown with labeled bars (right axis); family candidate names are marked with a large dot; unlabeled bars represent Tec1, *pogo*, Tec2, CeMar, CpMar, CbTc1, and *mariner*. Scores >30 represent the profile constituents. (*B*) Detection of Tec1 and Tec2 transposases. A profile of the five IS*630* family members, plus TBE1 and IS*895*, was used in a search. Of the 22 established IS*630*-Tc1 transposases in the database, only the 5 lacking all or parts of the DE and D35E region got scores <2. Program settings: *A* PILEUP (gap weight, 3.0; gap length, 0.5), PROFILEMAKE (default), and PROFILESEARCH (gap weight, 5.0; gap length, 0.2; minseq, 200); *B* as in *A* except PILEUP (gap length; 0.2). List of IS*630*-Tc1 member and candidate protein sequences, with each corresponding GenBank accession number or locus name in parentheses: TBE1 (L23169), Tec1 (L03359), Tec2 (L03360); IS*630* (SHFIS630), StIS*630* (STYORF), RSa (RHMA9INS1), IS*1066* (PSETNDO), RiATL (RIATL), IS*895* (ANAIS895A), Tc1 (CELTC2), CbTc1 (CELTBCB1), IpTc1 (ICTIGHD), EcTc1 (EURHEMED2), EcTc1a (EURHEMED6), CbTc2 (CELTCB2), Tc3 (CELB0303), Tc6 (CELRTC61), Uhu (DROTEUHU), HB1 (DROFBHB1E), HB2 (DROFBHB2E), Tes1 (EPTVASOB), *minos* (DROMINOS), *pogo* (DROPOGOR11), *bari* (DMBARI1), *mariner* (DROMAR), CeMar (CELZK370), HcMar (HCEMTRMPAB), and CpMar (CHXMARNTN). Augmented GenPept database: besides the ciliate element sequences, the following protein sequences were not in the GenPept data base (release 72) and were obtained from translations of the GenBank DNA sequences, guided by homology to Tc1 or IS*630*, and in cases treated by Henikoff (6), guided by homology blocks identified by him: *mariner*, Tc6, HB1, HB2, EcTc1, EcTc1a, Tes1, IpTc1, HcMar, and RiATL. IS*895* was translated to fuse its two long ORFs at nt 580 (see text).

To further examine whether TBE1, the Tec elements, and IS*895* are members of the IS*630*-Tc1 family, their sequences were aligned with those of a representative set of family members. Regions of high conservation were found in the alignment, as indicated by its associated similarity plot (Fig. 2*A*). While there are hints of conservation in the N-terminal portions, more convincing regions lie beyond position 200, with three or four particularly prominent peaks. Fig. 3 shows the alignment across these regions of high similarity. The first region is centered on a fully conserved DE dipeptide (DE peak, Fig. 2*A*). Two further regions involve a fully conserved Asp (D) and a highly conserved Glu (E), separated in most cases by 34 or 35 residues (35 for all bacterial elements, except 43 for IS*895*; 34 for all eukaryotic elements, except 38 for the hagfish Tes1). We refer to this region as the "D35E" region, as discussed below. Note that the putative family members, TBE1, Tec1, and Tec2, align well with the known family members, convincing us the homologies are real. Even IS*895* aligns well, despite its unusually long D35E region. In addition, the alignment shows the *mariner* elements aligning well with the rest of the family across the full region from DE to beyond D35E, whereas previously only a small portion of the *mariner* D35E region was shown to be homologous to the family (7). The alignment implies that the distal E of D35E has undergone a chemically conservative E→D substitution in the *mariner* elements (both parsimony and distance-method transposase gene trees indicate the *mariner* family E has the derived state; data not shown).

Relatedness of the IS*630*-Tc1 and IS*3*-Retrotransposon-Retrovirus Transposase-Integrase Families. The conservation of a D35E region in IS*630*-Tc1 transposase family members

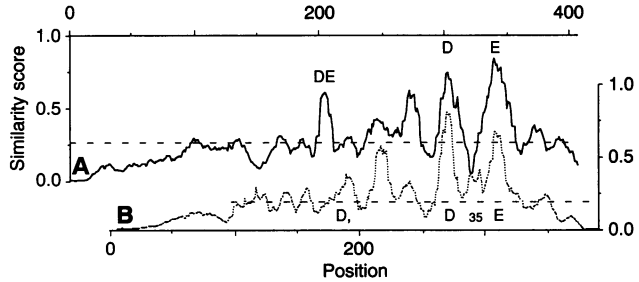

FIG. 2. Similarity plots for the IS*630*-Tc1 and IS*3*-retroposon transposase families. Protein sequences of the two families were assembled into alignments and the associated similarity plots are shown. (*A*) IS*630*-Tc1 family. (*B*) IS*3*-retroposon family. The horizontal dashed line of each plot represents its average similarity value. Similarity peak labels indicate highly conserved residues (see Fig. 3 and text). The IS*630*-Tc1 member and candidate sequences aligned are those shown in Fig. 3. (The N-terminal 60 residues of Tec1 and Tec2 were excluded). The list of 51 phylogenetically diverse IS*3*-retroposon family members aligned is available upon request. Program settings: PLOTSIMILARITY (window, 10) and in *A* PILEUP (gap weight, 1.5; gap length, 0.3), and in *B* PILEUP (gap weight, 1.0; gap length, 0.1).

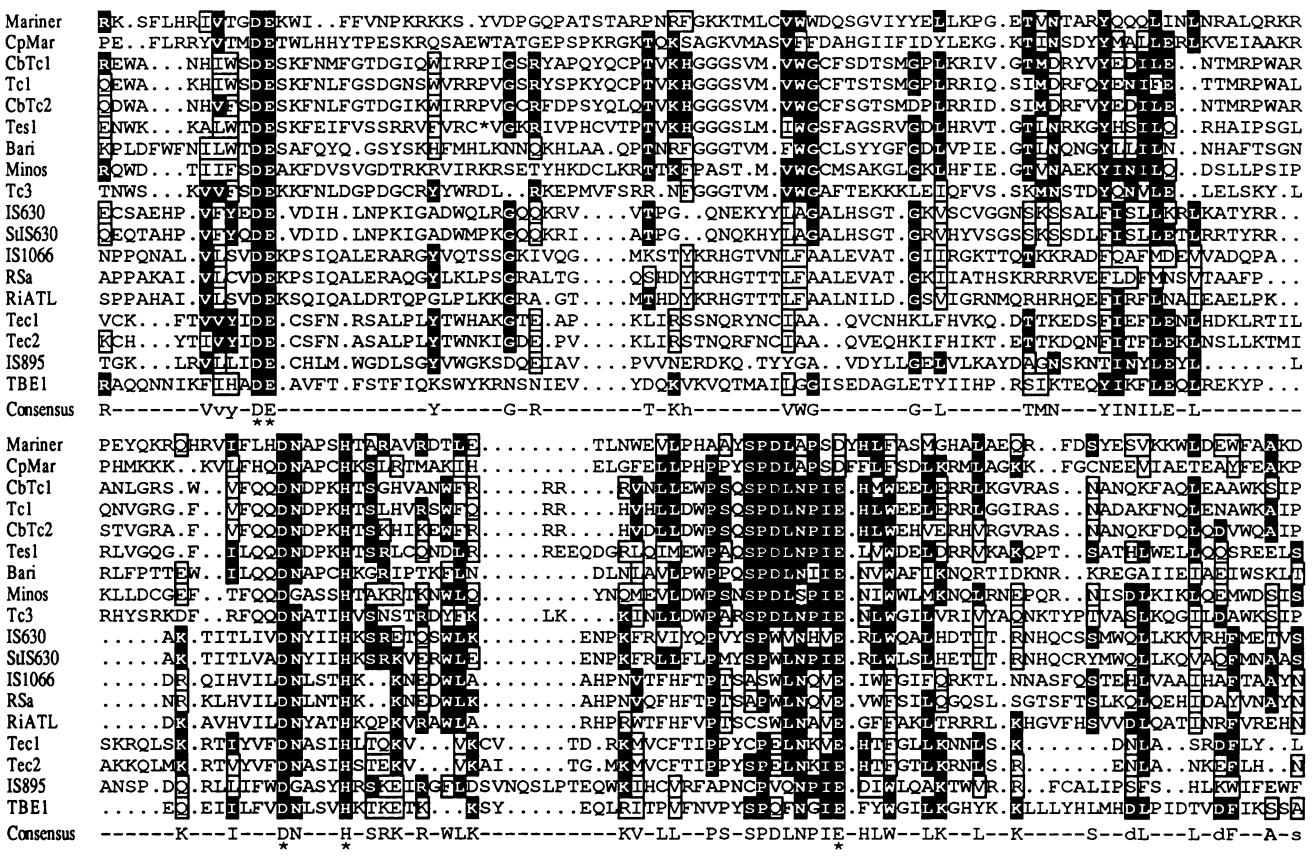

FIG. 3. Alignment of IS*630*-Tc1 family transposases. The figure shows a section of the alignment described in Fig. 2*A* that includes prominent regions of similarity. Asterisks below the consensus indicate completely or prominently conserved residues, including those that identify the three marked peaks in Fig. 2*A* (DE, D, and E). To identify evolutionarily related residues, coalitions and a consensus were calculated for the alignment. Members of the winning coalition are marked: representatives of the coalition leader are indicated by white letters in a black background, and other members are boxed. Where a tie for leader occurred, we assigned a leader, indicated in lowercase type in the consensus sequence. The asterisk in the Tes1 sequence signifies a stop codon. Program setting: PRETTY (plurality, 10; threshold, 1.0) and BLOSUM [62-substitution matrix (7)].

suggests a relationship to a conserved D35E motif in the IS*3*-retroposon family. Several other observations lead us to propose that these families are related. (*i*) They are both families of transposases, and proteins of both families have similar sizes (Fig. 2). (*ii*) IS*895* has two long ORFs that overlap for 25 codons (17). Within this overlap, we noted the potentially "slippery" sequence AAAAAAAAAAG. A −1 frameshift within the string of adenosines would join the two frames to produce the full-length protein sequence we have analyzed. This possible frameshift mechanism is reminiscent of that probably acting to make most IS*3* family transposases (19).

Protein sequence similarity analyses further support the relatedness of the two families. (*i*) Database searches with the TBE1 protein sequence frequently gave high scores to IS*3*-retroposon family members. For instance, in a BLAZE search, the retrotransposon *copia* and the IS*3* relative IS*136* scored second and eighth, respectively; in a BLAST search they scored fifth and fifteenth, respectively. In each case, the matches involved the D35E regions (data not shown). (*ii*) The similarity plots of the two families (Fig. 2) resemble each other in the positions of their D35E regions, and when the D35E regions are lined up, other peaks of similarity coincide. This led us to compare (data not shown) the pattern of conserved residues in the D35E region of our alignments of IS*630*-Tc1 sequences (Fig. 3 *Lower*) with published IS*3*-retroposon alignments (8, 9). Besides the invariant D and E positions and the generally similar spacing between, we saw that the two blocks resemble each other at several other positions, if one accepts the possibility of minor insertion-deletion differences between the families. Although difficult to judge critically, this comparison encouraged us to make a grand alignment including representative sequences from both families (described below). We also compared the conserved DE region of the IS*630*-Tc1 family and the so-called D, region of the IS*3*-retroposon family (see Introduction), but the relationship was difficult to assess and was not pursued.

All members of both families did align across their D35E segments. The D35E section of the alignment is shown in Fig. 4. Aligned residues marked in Fig. 4 belong to "coalitions" of evolutionarily related residues. Most coalitions include residues from members of both families (Fig. 4, above and below the central gap), thus supporting the contention that the two D35E motifs are homologous. For example, at the left end of the block note the invariant Asp (D of D35E), shared by all members of both families, the adjacent N, shared by a large fraction of members of both families, and then the S T and aromatic H F Y coalitions 2 and 4 residues further rightward. Beyond the gapped central region are two nearby consensus prolines (P-YSP), and to the right of the conserved E for each family, an LK dipeptide is shared by a variety of members of both families. The alignment does not, however, show an invariant Glu (E) at the right side of the 35-residue region—the "E" of D35E. Instead it suggests that maybe two nonhomologous Glu residues serve analogous roles for their respective transposases. We are not fully satisfied with the placement of a single gap just to the right of the Glu for most IS*630*-Tc1 members and, if given the license of "by-eye" tinkering, would move it leftward two positions. This would fully align the Glu residues of both families along with the

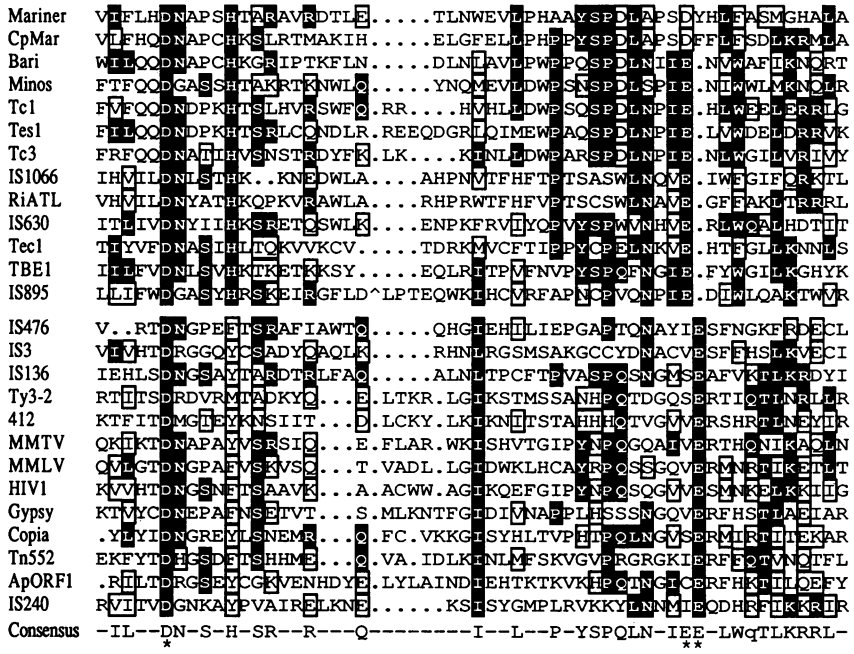

FIG. 4. Alignment to compare the D35E regions of the IS*630*-Tc1 and IS*3*-retroposon transposase families. A subset of 13 IS*630*-Tc1 members and 13 IS*3*-retroposon members was chosen for broad phylogenetic representation, as judged by the dendrogram generated by PILEUP with a large set of members from both families (all Tc1 members represented in Fig. 3 and 51 members of the IS*3* family). From each family of 13 sequences, a region of 120 residues, centered on the D35E region, was chosen and these were aligned. Coalitions and consensus were calculated and residues were marked as described for Fig. 3 [PILEUP (gap weight, 3.0; gap length weight, 0.3) and PRETTY (plurality, 12; threshold, 1.0)]. The caret (^) in the IS*895* sequence represents the sequence SVNQS (removed after alignment). The 13 IS*3*-retroposon sequences chosen were IS*476* (Y161_XANCA), IS*3* (TRA5_ECOLI), IS*136* (ATUIS136), Ty3-2 (YSCTY32A), 412 (GNFF42), mouse mammary tumor virus (MMTV), (MMTPOL), Maloney murine leukemia virus (MMLV) (MLMCG), human immunodeficiency virus 1 (HIV1) (HIVELICG), *gypsy* (DROGYPSY), *copia* (DROCOPIA1), Tn*552* (STATN552), ApORF1 (ACNORF), and IS*240* (BACIS2402).

adjacent residue, which would form a strong V I coalition. Indeed, if the *mariner* sequences are removed, then the calculated alignment confirms our "by-eye" supposition. However, this altered alignment fails to preserve many coalitions left of the gapped region (data not shown). Thus, we feel the potential homology of the namesake Glu residues (E) of the D35E motifs is an open issue.

Because it is difficult to assess whether the similarity between the D35E motifs of the two families represents true homology, we devised a test to determine how unusual the similarity might be. Separate profiles were made representing the D35E motifs of the two families, and each was used to search our augmented protein sequence data base. Fig. 5 shows the results of the search with the IS*630*-Tc1 D35E profile (previously recognized members only). The distribution of all 26,265 scores has an average and SD of 0 and 1 (0.01 ± 1.07), representing a negative control collection of nonsignificant matches with the profile. The excluded IS*630*-Tc1 family members represent a positive control (Fig. 5 *Upper Right*); their scores averaged 5.12 (±0.62) (also note *pogo*). The collection of IS*3*-retroposon sequences in the database received a wide range of scores, with an average, 1.42 (±1.47), well above the negative control; the top of the range overlapped that of the positive control distribution (mouse mammary tumor virus received the top score of 5.40). These results indicate the IS*630*-Tc1 D35E motif matches IS*3*-retroposon sequences far better than sequences in general. The IS*3*-retroposon D35E profile search gave complementary results: IS*630*-Tc1 family member scores ranged up to 5.51 (CbTc2), with an average of 2.33 (±1.50). The result of the two profile searches together strongly support the proposal that the D35E motifs of the two families are homologous. We are, however, unaware of a method to assign statistical significance to this similarity, and homology for us remains a strong hypothesis until a test can be devised.

DISCUSSION

Similarities Within the IS*630*-Tc1 Family. Additional members of the IS*630*-Tc1 family of transposase genes have been identified, including proteins encoded by elements from two ciliated protozoa and a bacterial element, IS*895*. This led us to ask what other properties these elements might share. First, they appear to share the common property of transposition target specificity, and in general use the same target. We surveyed target sequences of all members of the IS*630*-Tc1 family, and in nearly every case the results are consistent with insertion into the same target sequence, the palindromic dinucleotide TA. The evidence varies from irrefutable to simply consistent with the hypothesis. Comparison of sites before and after demonstrated transposition proves TA is the target for Tc1, IS*630*, *mariner*, and IS*895* (17, 20–22) and *pogo* (18). Where element ends can be tentatively identified by comparison to closely related element ends, TA appears adjacent to the ends of a variety of further elements, including the Tec elements (5, 6, 10, 23). The TBE1 target is not TA but ANT (N is any nucleotide; ref. 24).

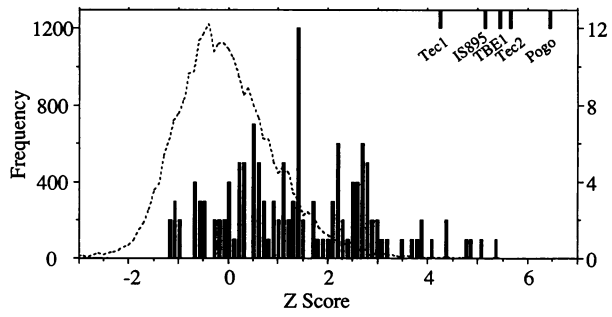

FIG. 5. Detection of IS*3*-retroposon family transposases with an IS*630*-Tc1 profile. Twenty-two IS*630*-Tc1 sequences were aligned. A profile was created from a 59-residue-wide section of the alignment, to include the D35E region. The augmented GenPept data base (see Fig. 1) was searched with the profile, and the Z scores for all database entries are plotted (dotted curve, left axis). A list of 145 members of the IS*3*-retroposon family was prepared and is available upon request. Their Z scores were extracted from the grand list and plotted as standing bars (right axis). Scores of the additional members of the IS*630*-Tc1 family (not represented in the profile) are shown as hanging bars. Z-score distribution means (±SD) were as follows: entire data base, 0.0125 (±1.0717); IS*3*-retroposon family, 1.423 (±1.468); members of profile, 12.728 (±4.721); control members, 5.38 (±0.801). Aligned IS*630*-Tc1 sequences: IS*630*, StIS*630*, RSa, IS*1066*, RiATL; Tc1, CbTc1, IpTc1, EcTc1, EcTc1a, CbTc2, Tc3, Tc6, Uhu, HB1, HB2, Tes1, *minos*; *mariner*, CeMar, HcMar, and CpMar. Program settings: PILEUP (gap weight, 2.0; gap length, 0.2), PROFILEMAKE (default), PROFILESEARCH (gap weight, 20; gap length, 8; minseq, 200).

IS*630* shows target specificity beyond the central TA, preferring the palindromes CTAG or TTAA, and this seems to be true for the other bacterial family members (5, 21). Tc1 also may show preference outside its TA target (20). TBE1s show specificity for CANTG (24).

Because all family members have inverted terminal repeats and the target is centrisymmetric, it is not possible to decide from sequence analyses the exact extent of the element and whether or not the target is duplicated (25). However, Tenzen *et al.* (21) showed experimentally that IS*630* does not end A . . . T and does duplicate the central TA of its CTAG (TTAA) target. This fact at least provides a working model for the other members of the family until experimental tests can be devised in each case.

A second property shared by family members is somatic excision. The ciliate, Tc1, and *mariner* elements all excise from the somatic genomes of their hosts (10, 20, 22). The three ciliate elements are excised during development of the ciliate host somatic nucleus, or macronucleus. Excision is complete, such that no members of the family persist in the mature macronucleus, and excision is precise, such that each transpositional mutation is exactly reverted in the macronucleus. We and others have argued (for review, see ref. 10) that the ability to completely and precisely excise in the soma has allowed these transposon families to create additional members with little regard to the host genome, because the germ line is not expressed. All expression is from the macronucleus, whose genes are not transmitted to sexual progeny. As expected, the copy numbers of these elements in the host micronuclear genomes can be extremely high (1800 for TBE1 and 30,000 each for Tec1 and Tec2).

Tc1 and *mariner* excise only sporadically from the soma, and excision is mostly imprecise (20, 26). Their hosts are metazoans. Unlike the ciliate micronucleus, the metazoan germ-line genome cannot remain silent, because the germ-line genome does not share a common cytoplasm with the somatic genome. As a result, each germ-line nucleus must express genes—housekeeping as well as germ-line-specific genes. These genes therefore are not safe targets for transposition. To take advantage of the benefits of efficient and precise somatic excision, Tc1 and *mariner* would have had to evolve a mechanism to avoid transposition into germ-line-expressed genes. We suggest that the metazoan and ciliate elements evolved from a common ancestor that had some propensity for somatic excision, as Tc1 and *mariner* do today. In ciliates, TBE1 and Tec evolved two additional abilities: excision became fully efficient and precise excision replaced imprecise excision (10, 24).

Significance of the Interrelationship of Transposase Families. The D35E regions of the transposases of the IS*630*-Tc1 and IS*3*-retroposon families are similar. This region of the human immunodeficiency virus integrase resides in a protease-resistant core capable of various transposition-related *in vitro* reactions (27), and its most conserved residues are necessary for these reactions (9, 28–30). Thus it seems possible that the D35E region of the IS*630*-Tc1 transposases catalyzes similar reactions—a testable hypothesis.

A domain N-terminal to the central core appears to contribute to DNA binding for retroviruses and retrotransposons (27-32), the IS*3* family (33), Tc1 (34), and the IS*630* family (5). However, available evidence suggests these N-terminal domains are not homologous, since the retroviral module involves a zinc-binding motif apparently absent from IS*3* family transposases (see ref. 27), which instead show a helix–turn–helix motif (33), as also suggested for IS*630* members (5). The implication is that the transposases of the various subgroups may be homologously related only by a common central catalytic domain.

Mizuuchi (35) has pointed to a common mechanism of transposition for an even wider variety of transposable elements, suggesting that the homology we believe we have detected reaches even deeper than we have been able to detect by sequence analyses. Thus, members of the proposed IS*630*-Tc1/IS*3*-retroposon superfamily are found in an extremely wide spectrum of hosts, now including ciliated protozoa. Given the recently demonstrated horizontal transfer of *mariners* between widely disparate hosts (4), it seems possible that much of the wide distribution of the superfamily of elements may have been horizontal.

We thank L. Okun for continual help with the computing facilities, J. Seger for inciteful discussions, and B. Butler, M. Hogan, and I. Edelman at GCG for generous instruction. This work was supported by National Institutes of Health Grants GM25203 to G.H. and GM37661 to C.L.J.

1. Berg, D. E. & Howe, M. M., eds. (1989) *Mobile DNA* (Am. Soc. Microbiol., Washington, DC).
2. Harris, L. J., Baillie, D. L. & Rose, A. M. (1988) *Nucleic Acids Res.* **16**, 5991–5998.
3. Henikoff S. & Plasterk, R. H. A. (1988) *Nucleic Acids Res.* **16**, 6234.
4. Robertson, H. M. (1993) *Nature (London)* **362**, 241–245.
5. van der Meer, J. R., Zehnder, A. J. B. & de Vos, W. M. (1991) *J. Bacteriol.* **173**, 7077–7083.
6. Henikoff, S. (1992) *New Biol.* **14**, 382–388.
7. Henikoff, S. & Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10915–10919.
8. Fayet, O., Ramond, P., Polard, P., Prere, M. F. & Chandler, M. (1990) *Mol. Microbiol.* **4**, 1771–1777.
9. Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P. G. & Skalka, A. M. (1992) *Mol. Cell. Biol.* **12**, 2331–2338.
10. Klobutcher, L. A. & Jahn, C. L. (1991) *Curr. Op. Genet. Dev.* **1**, 397–403.
11. Jahn, C. L., Doktor, S. Z., Frels, J. S., Jaraczewski, J. W. & Krikau, M. F. (1993) *Gene* **133**, 71–78.
12. Herrick, G., Cartinhour, S., Dawson, D., Ang, D., Sheets, R., Lee, A. & Williams, K. (1985) *Cell* **43**, 759–768.
13. Williams, K. & Herrick, G. (1991) *Nucleic Acids Res.* **19**, 4717–4724.
14. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
15. Gribskov, M., Luthy, R. & Eisenberg, D. (1990) *Methods Enzymol.* **183**, 146–159.
16. Ciazzi, R., Caggese, C. & Pimpinelli, S. (1993) *Genetics* **133**, 335–345.
17. Alam, J., Vrba, J. M., Cai, Y., Martin, J. A., Weislo, L. J. & Curtis, S. E. (1991) *J. Bacteriol.* **173**, 5778–5783.
18. Tudor, M., Lobocka, M., Goodell, M., Pettitt, J. & O'Hare, K. (1992) *Mol. Gen. Genet.* **232**, 126–134.
19. Chandler, M. & Fayet, O. (1993) *Mol. Microbiol.* **7**, 497–503.
20. Moerman, D. G. & Waterston, R. H. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington, DC), pp. 537–556.
21. Tenzen, T., Matsutani, S. & Ohtsubo, E. (1990) *J. Bacteriol.* **172**, 3830–3836.
22. Hartl, D. L. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington, DC), pp. 531–536.
23. Dreyfus, D. H. (1992) *Mol. Immunol.* **29**, 807–810.
24. Williams, K., Doak, T. G. & Herrick, G. (1993) *EMBO J.* **12**, 4593–4601.
25. Rosenzweig, B., Liao, L. W. & Hirsh, D. (1983) *Nucleic Acids Res.* **11**, 7137–7140.
26. Bryan, G., Garza, D. & Hartl, D. (1990) *Genetics* **125**, 103–114.
27. Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P. & Craigie, R. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 3428–3432.
28. Vincent, K. A., Ellison, V., Chow, S. A. & Brown, P. O. (1993) *J. Virol.* **67**, 425–437.
29. Engelman, A. & Craigie, R. (1992) *J. Virol* **66**, 6361–6369.
30. van Gent, D. C., Groeneger, A. A. & Plasterk, R. H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9598–9602.
31. Khan, E., Mack, J. P. G., Katz, R. A., Kulkosky, J. & Skalka, A. M. (1990) *Nucleic Acids Res.* **19**, 851–860.
32. Vink, C., Oude Groeneger, A. A. M. & Plasterk, R. H. A. (1993) *Nucleic Acids Res.* **21**, 1419–1425.
33. Prere, M. F., Chandler, M. & Fayet, O. (1990) *J. Bacteriol.* **172**, 4090–4099.
34. Schukkink, R. F. & Plasterk, R. H. A. (1990) *Nucleic Acids Res.* **18**, 895–900.
35. Mizuuchi, K. (1992) *Annu. Rev. Biochem.* **61**, 1011–1051.