Data Sets 11-30.

Data sets 11-30 are output from the DME program applied to the sequence sets described in the paper. An example of a header line is:

```
> rCCAWkStGy SCORE=50060.67 COUNT=(42,420)[4506.4] INFO/COL=1.34
```

The fields of the header line have the following meanings.

1. "`rCCAWkStGy`" is the degenerate consensus. Upper case letters mean that the corresponding column is closer to the ideal column for that particular degenerate nucleotide.
2. "`SCORE=50060.67`" is the score assigned to the motif during the search.
3. "`COUNT=(42,420)[4506.4]`" indicates that the motif had 42 occurrences in the foreground, 420 in the background, and if the foreground had the same size as the background, the 42 occurrences would translate to 4506.67
4. "`INFO/COL=1.34`" means that the information content per column of the final matrix was 1.34. The "final matrix" is not the matrix discovered during the enumeration if any postprocessing was applied to improve the matrix.

An example matrix is:

```
55      0       33      12
0       95      0       5
0       100     0       0
100     0       0       0
38      0       0       62
0       0       64      36
0       43      57      0
19      0       0       81
0       0       100     0
14      67      0       19
```

The first to fourth columns represent the percentage of bases A, C, G, and T, respectively, in each column.

An example alignment is:

```
   1   94       GCCCCCATATTGCTGTCCCC TCCATTCTGC TGCTCCTCGCGTACAGAACA
   1  532       GCTGCCAACACAGAGGTGCA ACCATGGTGC GTCCGCTGCTGACAAGAACA
   2  354       GCTGCCAACGCGGAGCTGCA ACCATGGTGC GTCTGCGGCTGACAAGACCA
   3  139       CAGCCTGCCACGCCGCTGCC GCCATGCTGA CGCCGAGGACAAGAAGCTCA
   4  124       GGGTGCATCCTGACAGGAGT GCCAAGCAGC CCACGAGGCCAAGGCTGCCA
```

Each line represents a motif occurrence, the first value is the number of the sequence in which the occurrence exists, the second number is the offset of the aligned segment in the sequence, and the sequence is presented with 20 bases of context on either side (the portion between the two spaces is the actual occurrence).

Examples of the presentation for binding site database hits (from TRANSFAC or JASPAR) are as follows:

```
MA0082 1.19701  SQUA
M00392 1.33726  AGL3
```

Each line has the accession of the database motif (accessions beginning with MA are from JASPAR, others are from TRANSFAC). The numerical value is the divergence between the database motif and the one found be dme. The final entry on each line is a list of factor names that are known to bind to the database matrix.