```
>COAT_TMV

                                      1-68    SYSITTPSQFVFLSSAWADPIELINLCTNA
                                              LGNQFQTQQARTVVQRQFSEVWKPSPQVTV
                                              RFPDSDFK
vyrynavldplvtallgafdtrnriieven        69-158
qanpttaetldatrrvddatvairsainnl
ivelirgtgsynrssfesssglvwtsgpat


>COAT_TBSVB

amttrnnnnvlavskkqlgvlaasaavgal        1-174
rnyigesspallqsavglgkkalnkvrnrr
kqgnqqiithvggvggsimapvavsrqlvg
skpkftgrtsgsvtvtghreyltqvnnssg
fvvnggivgnslqlnpsngtlfswlpalas
       nfdqysfnsvvldyvplcgttevg
                                      175-232  RVALYFDKDSQDPEPADRVELANFGVLKET
                                               APWAEAMLRIPTDKVKRYCNDSATVDQK
lidlgqlgiatyggagadavgelflarsvt        233-387
lyfpqptntllsskrldltgsladatgpgy
lvltrtptvlthtfratgtfnlsgglrclt
sltlgatgavvindilaidnvgtasdyfln
ctvsslpatvtftvsgvaagillvgraran
                            vvnll
```

Fig. 2. Amino acid sequence complexity. The SEG output for capsid proteins of tobacco mosaic tobamovirus (TMV) and tomato bushy stunt tombusvirus (TBSV). Low-complexity regions are shown in the upper case. Conserved motifs shared by the capsid proteins of the viruses with rod-shaped virions (TMV), or with small isometric particles (TBSV), are shown in boldface.

statistical theory for this type of analysis has been described *(7,8)* and implemented in the Statistical Analysis of Protein Sequences (SAPS) program *(9)*. **Figure 1** shows excerpts of the SAPS output for the sequence of potato leaf roll luteovirus capsid protein. A notable feature highlighted by this analysis is the positive charge cluster in the N-terminal portion of the protein, which results in the statistically significant overrepresentation of arginine in the entire protein.

2.1.2. Amino Acid Sequence Complexity

Many proteins contain stretches of nonrandom amino acid composition. These compositionally biased, or low-complexity, segments, built of a mosaic of a few amino acid residues, are abundant in large cellular proteins, where they have been shown to fulfill important regulatory functions (reviewed in **ref. 10**). We used the SEG program *(11,12)* to evaluate the sequence complexity in some plant virus capsid proteins. Although virus capsid proteins are relatively compact and are thought to be well-structured, many of them contain considerable proportion of the low-complexity regions (**Fig. 2**; and unpublished observations). At the same time, regions of the best conservation between dis-