```
SAPS.  Version of March 24, 1993.
Date run: Fri May  5 10:25:55 1995

File: plrv

SWISS-PROT ANNOTATION:
ID     COAT_PLRVW

number of residues:  208

      1  MSTVVVKGNV NGGVQQPRRR RRQSLRRRAN RVQPVVMVTA PGQPRRRRRR RGGNRRSRRT
     61  GVPRGRGSSE TFVFTKDNLM GNSQGSFTFG PSLSDCPAFK DGILKAYHEY KITSILLQFV
    121  SEASSTSSGS IAYELDPHCK VSSLQSYVNK FQIPQGGAKT YQARMINGVE WHDSSEDQCR
    181  ILWKGNGKSS DTAGSFRVTI RVALQNPK
------------------------------------------------------------------------
COMPOSITIONAL ANALYSIS (extremes relative to: swp23s)

A  : 10( 4.8%); C  :  3( 1.4%); D  :  7( 3.4%); E  :  6( 2.9%); F  :  8( 3.8%)
G  : 20( 9.6%); H  :  3( 1.4%); I  :  8( 3.8%); K  : 11( 5.3%); L- : 10( 4.8%)
M  :  4( 1.9%); N  : 10( 4.8%); P  : 10( 4.8%); Q  : 13( 6.2%); R++: 26(12.5%)
S+ : 24(11.5%); T  : 11( 5.3%); V  : 17( 8.2%); W  :  2( 1.0%); Y  :  5( 2.4%)


KR    + :   37 ( 17.8%);    ED     :   13 (  6.2%);    AGP    :   40 ( 19.2%);
KRED    :   50 ( 24.0%);    KR-ED ++:   24 ( 11.5%);   FIKMNY :   46 ( 22.1%);
LVIFM   :   47 ( 22.6%);    ST     :   35 ( 16.8%).


------------------------------------------------------------------------
CHARGE DISTRIBUTIONAL ANALYSIS

      1  000000+000 0000000+++ ++000+++00 +000000000 0000++++++ +000++0++0
     61  000+0+000- 00000+-000 0000000000 0000-0000+ -000+000-0 +000000000
    121  0-00000000 000-0-000+ 000000000+ 00000000+0 000+00000- 00-00--00+
    181  000+000+00 -00000+000 +000000+

CHARGE CLUSTERS.

Positive charge clusters (cmin = 13/30 or 18/45 or 22/60):
  1) From    45 to   66:    RRRRRRRGGNRRSRRTGVPRGR
                            +++++++000++0++0000+0+
     quartile: 1; size: 22, +count: 13, -count:  0, 0count:  9; t-value:  5.07 *
     G:  4 (18.2%); R: 13 (59.1%);
Negative charge clusters (cmin =  7/30 or  9/45 or 11/60):  none
Mixed charge clusters (cmin = 16/30 or 22/45 or 27/60):  none

C. CHARGE RUNS AND PATTERNS.

  (+)   7(0,0,0); at   45- 51:    RRRRRRR
        (1. quartile)            +++++++
```

Fig. 1. Statistical analysis of a protein sequence. An excerpt of the SAPS output for the potato leaf roll virus (PLRV) capsid protein. The + value shows statistical significance of residue-clustering expressed as the number of standard deviations above the mean.

2. Methods and Examples

2.1. Intrinsic Properties of Protein Sequences

2.1.1. Statistical Analysis of Amino Acid Sequences

Exploring simple statistical properties, such as the presence of clusters of certain types of residues (e.g., charged or hydrophobic) or short repeats, is a good starting point in analyzing an amino acid sequence. The straightforward