mixture method ***(44)***, calculates weights in each position of the matrix, and then scans a sequence database iteratively in order to identify sequence segments that have a score with the matrix above a chosen cutoff. The cutoff is defined as the ratio of the expected number of selected segments to the actually observed number. A typical MoST command line is **most nr aa.mot r0.01 i80% >aa.mres**, where **aa.mot** is a multiple alignment block file used to calculate the position-dependent weight matrix, **r0.01** indicates the expected/observed ratio cutoff, and i80% indicates the weighting scheme, under which each sequence segment in a group of *n* sequences with greater than 80% (or other user-specified percentage) identity is a given a weight of 1/*n*.

When the most conserved C-terminal motif from the alignment of the rod-shaped plant virus capsid proteins (**Fig. 6**) was used as the input for MoST to search the entire NR database, a unique set of proteins was retrieved from the NR database that included, in addition to the capsid protein sequences from which the initial alignment block was derived, segments of nonstructural polyproteins encoded in RNA-2 of a bymovirus, barley yellow mosaic virus (**Fig. 7A**). Subsequent analysis using MACAW showed that the proteins of two bymoviruses also contained the other two motifs typical of the capsid proteins (**Fig. 7B**). This unexpected similarity between the capsid proteins and apparently nonstructural virus proteins has been noticed previously in the course of a comprehensive analysis of the (+)-strand RNA virus protein sequences ***(6)***. Here we show that this similarity is statistically significant and unique in the entire protein sequence database. In retrospect, the similarity with the protein of another bymovirus, barley mild mosaic virus, is detectable in the BLAST outputs for some of the capsid proteins (e.g., **Fig. 4A**). However, it is not distinguishable from spurious hits with unrelated cellular and viral proteins without additional analysis.

It appears most likely that the sequence coding for the capsid protein homolog has been introduced into an ancestor bymovirus genome by recombination between distantly related viruses replicating in the same host. The function of this protein in the bymovirus replication remains to be studied.

The above example shows how motif analysis may help reveal nontrivial relationships between protein sequences. Similarly, when the NR database was searched with the most conserved alignment block from capsid proteins of small spherical plant viruses ***(45)***, a segment of the hepatitis E virus (HEY) capsid protein has been detected (**Fig. 8**). This may suggest that the capsid protein of HEV belongs to the large class of spherical capsid protein with the jellyroll structural fold ***(45–47)***.

Database searches with the most conserved block derived from the alignment of the capsid proteins of plant DNA viruses, geminiviruses, allowed us to confirm statistical significance of a previously noticed local similar-