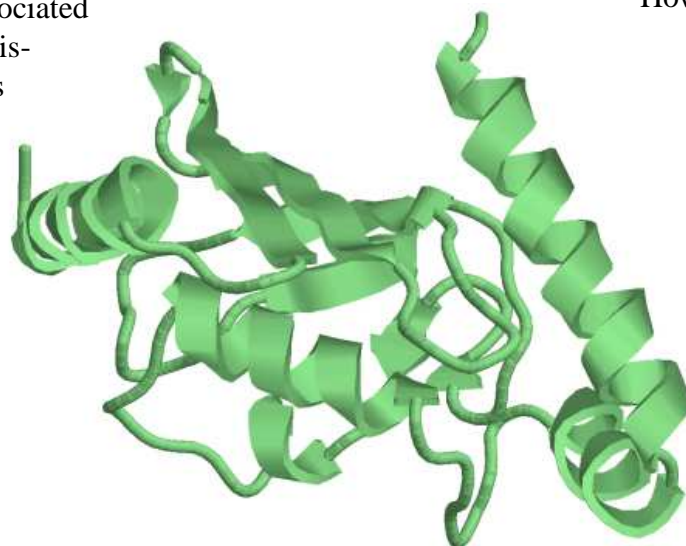


Simple sequences are rare in the Protein Data Bank

Simple sequence protein repeats are associated with several human neurodegenerative diseases, and essential developmental genes in other species. Indeed they are the most commonly shared peptide segments among eukaryotic proteins. However very little is known about what types of structures these simple sequence protein repeats may form. Their functional roles also remain elusive.

A protein repeat can be defined as a segment of protein composed primarily of one of a limited number of amino acids.

The figure shows the crystal structure of the yeast ubiquitin-conjugating enzyme Rad6, sub-chain A. The repeat sequence (highlighted in yellow) is composed primarily of aspartic acid residues (D).



```
MSTPARRRLMRDFKRMKEDAPPGVSASPLPDNVMVWNAMI IGPADTPYED  
GTFRLLEFDEEYPNKPPHVKFLSEMFHPNVYANGEICLDILQNRWTPTY  
DVASILTSIQSLFNDPNPASPANVEAATLFKDHKSQYVKRVKETVEKSWE  
DDMD DDDDDDDDDDDDEAD
```

However there is no structural information for that region of the protein.

The Protein Data Bank (PDB) stores all known 3D structures of proteins but it contains significantly less protein repeat sequence than a randomly constructed protein database matched for taxa and sequence length.

It has been suggested that simple sequence protein repeats do not form stable tertiary structures. This would inhibit their structural characterization, and explain their rarity within the PDB. We must therefore look outside the current structure-function paradigm to elucidate the function of these abundant repetitive peptide sequences.

Melanie A. Huntley and G. Brian Golding, 2002, Proteins 48:134-140.