MATLAB searches for potential PDZ ligand motifs

The following MATLAB scripts were used to search for putative PDZ ligand motifs in FASTA amino acid sequences, as reported in Frietze et al.

Consensus PDZ ligand motifs were derived from the following references:

Lenfant N, Polanowska J, Bamps S, Omi S, Borg JP, Reboul J: **A genome-wide study of PDZ-domain interactions in C. elegans reveals a high frequency of non-canonical binding**. *BMC Genomics* 2010, **11**:671.

Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh JH, Reva B, Held HA, Appleton BA, Evangelista M, Wu Y *et al*: **A specificity map for the PDZ domain family**. *PLoS Biol* 2008, **6**(9):e239.

Nourry C, Grant SG, Borg JP: **PDZ domain proteins: plug and play!** *Sci STKE* 2003, **2003**(179):RE7.

Consensus PDZ ligand motifs used were:
Class 1 PDZ ligand: X (S/T) X (V/L)
Class 2 PDZ ligand: X Φ X Φ
Class 3 PDZ ligand: X (D/E) X Φ
Where Φ= Y/F/W/C/M/V/I/L/A

Any sequence to be searched can be entered in the line
sequence = 'YOUR SEQUENCE HERE';
in FASTA format.

Note that because of the degeneracy of these consensus motifs, searches will return large numbers of potential motifs, most of which will not be functionally relevant. Additional bioinformatic, biochemical, and functional methods must be used to validate any potential PDZ ligand motif.

class 1 PDZ ligands:

```
function output = PDZ1(sequence)

sequence = 'MWMRKNNGGNGDDNTAAYQNEREHLSLSPRAESEALGVEAGMKDLPSAPPLVS';

seqshowwords(sequence, {'XSXY', 'XSXF', 'XSXW', 'XSXC', 'XSXM', 'XSXV', 'XSXI', 'XSXL','XSXA','XTXY',
'XTXF', 'XTXW', 'XTXC', 'XTXM', 'XTXV', 'XTXI', 'XTXL', 'XTXA'},'Alphabet', 'AA');

domains = {'XSXY', 'XSXF', 'XSXW', 'XSXC', 'XSXM', 'XSXV', 'XSXI', 'XSXL','XSXA','XTXY', 'XTXF', 'XTXW',
'XTXC', 'XTXM', 'XTXV', 'XTXI', 'XTXL', 'XTXA'};
output = zeros(length(domains),1);
for i = 1:length(domains)
output(i) = seqwordcount(sequence, domains{i},'Alphabet', 'AA');
end



end
```

class 2 PDZ ligands:

```
function output = PDZ2 (sequence)

sequence = 'MWMRKNNGGNGDDNTAAYQNEREHLSLSPRAESEALGVEAGMKDLPSAPPLVS';

seqshowwords(sequence, {'XYXA','XYXV', 'XYXI', 'XYXL', 'XYXM', 'XYXF', 'XYXW', 'XYXC', 'XYXY',
'XAXA','XAXV', 'XAXI', 'XAXL', 'XAXM', 'XAXF', 'XAXW', 'XAXC', 'XAXY', 'XVXV', 'XVXA', 'XVXY', 'XVXI', 'XVXL',
'XVXM', 'XVXF', 'XVXW', 'XVXC', 'XIXV', 'XIXY', 'XIXA', 'XIXI','XIXL', 'XIXM', 'XIXF', 'XIXW', 'XIXC', 'XLXY',
'XLXA', 'XLXV', 'XLXI', 'XLXL', 'XLXM', 'XLXF', 'XLXW', 'XLXC', 'XMXY', 'XMXA', 'XMXV', 'XMXI', 'XMXL',
'XMXM', 'XMXF', 'XMXW', 'XMXC', 'XFXY', 'XFXA', 'XFXV', 'XFXI', 'XFXL', 'XFXM',' XFXF', 'XFXW', 'XFXC',
'XWXY', 'XWXA', 'XWXV', 'XFXI', 'XWXL', 'XWXM', 'XWXF', 'XWXW', 'XWXC', 'XCXY', 'XCXA', 'XCXV', 'XCXI',
'XCXL', 'XCXM', 'XCXF', 'XCXW', 'XCXC'},'Alphabet', 'AA');

domains = {'XYXA','XYXV', 'XYXI', 'XYXL', 'XYXM', 'XYXF', 'XYXW', 'XYXC', 'XYXY', 'XAXA','XAXV', 'XAXI',
'XAXL', 'XAXM', 'XAXF', 'XAXW', 'XAXC', 'XAXY', 'XVXV', 'XVXA', 'XVXY', 'XVXI', 'XVXL', 'XVXM', 'XVXF',
'XVXW', 'XVXC', 'XIXV', 'XIXY', 'XIXA', 'XIXI','XIXL', 'XIXM', 'XIXF', 'XIXW', 'XIXC', 'XLXY', 'XLXA', 'XLXV',
'XLXI', 'XLXL', 'XLXM', 'XLXF', 'XLXW', 'XLXC', 'XMXY', 'XMXA', 'XMXV', 'XMXI', 'XMXL', 'XMXM', 'XMXF',
'XMXW', 'XMXC', 'XFXY', 'XFXA', 'XFXV', 'XFXI', 'XFXL', 'XFXM',' XFXF', 'XFXW', 'XFXC', 'XWXY', 'XWXA',
'XWXV', 'XFXI', 'XWXL', 'XWXM', 'XWXF', 'XWXW', 'XWXC', 'XCXY', 'XCXA', 'XCXV', 'XCXI', 'XCXL', 'XCXM',
'XCXF', 'XCXW', 'XCXC'};
output = zeros(length(domains),1);
for i = 1:length(domains)
output(i) = seqwordcount(sequence, domains{i},'Alphabet', 'AA');
end



end
```

class 3 PDZ ligands:

```
function output = PDZ3(sequence)

sequence = 'MWMRKNNGGNGDDNTAAYQNEREHLSLSPRAESEALGVEAGMKDLPSAPPLVS';

seqshowwords(sequence, {'XDXY', 'XDXF', 'XDXW', 'XDXC', 'XDXM', 'XDXV', 'XDXI', 'XDXL','XDXA','XEXY',
'XEXF', 'XEXW', 'XEXC', 'XEXM', 'XEXV', 'XEXI', 'XEXL', 'XEXA'},'Alphabet', 'AA');

domains = {'XDXY', 'XDXF', 'XDXW', 'XDXC', 'XDXM', 'XDXV', 'XDXI', 'XDXL','XDXA','XEXY', 'XEXF', 'XEXW',
'XEXC', 'XEXM', 'XEXV', 'XEXI', 'XEXL', 'XEXA'};
output = zeros(length(domains),1);
for i = 1:length(domains)
output(i) = seqwordcount(sequence, domains{i},'Alphabet', 'AA');
end



end
```