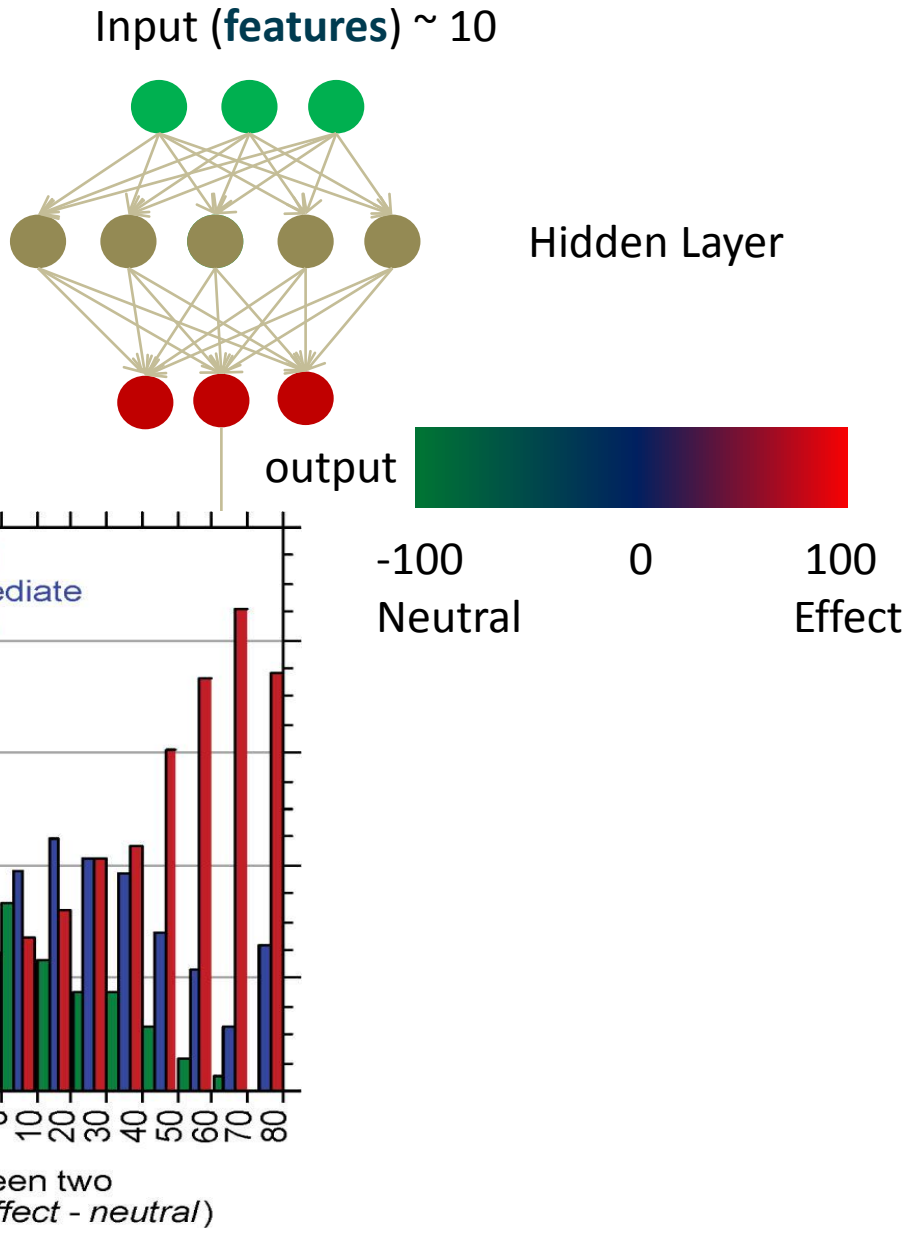
SNAP

Predict effect of non-synonymous polymorphisms on function

What is SNAP?

Features:

- Bio-chemical properties
- Sequence information
- Transition frequencies
- PSI-BLAST profiles
- SWISS-PROT annotation
- …



http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1920242/

Protein features for SNPs predictions in HUMAN (**strong high** and **neutral high**)

No differences:

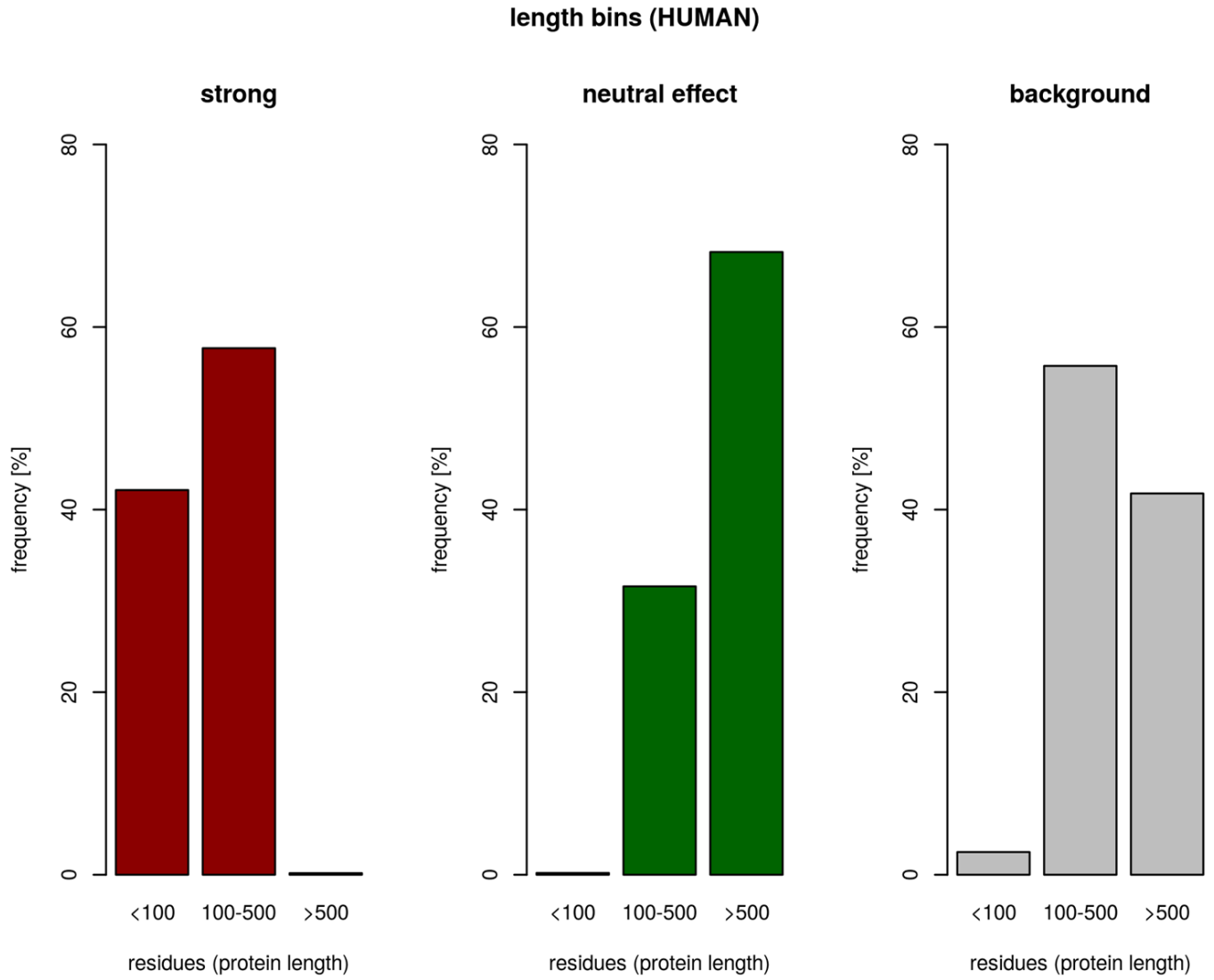
- #TM helices (mostly 0)
- MetaDisorder (predicted as disordered)
- coiled coil (motif is present)

Small differences:

- RNA binding sites
  - **Strong**: 60% with 0 binding sites (up to 25 sites)
  - **Neutral**: 90% with 0 binding sites (up to 9 sites)
- DNA binding sites
  - **Strong**: 40% with 0 binding sites (up to 24 sites)
  - **Neutral**: 90% with 0 binding sites (up to 14 sites)

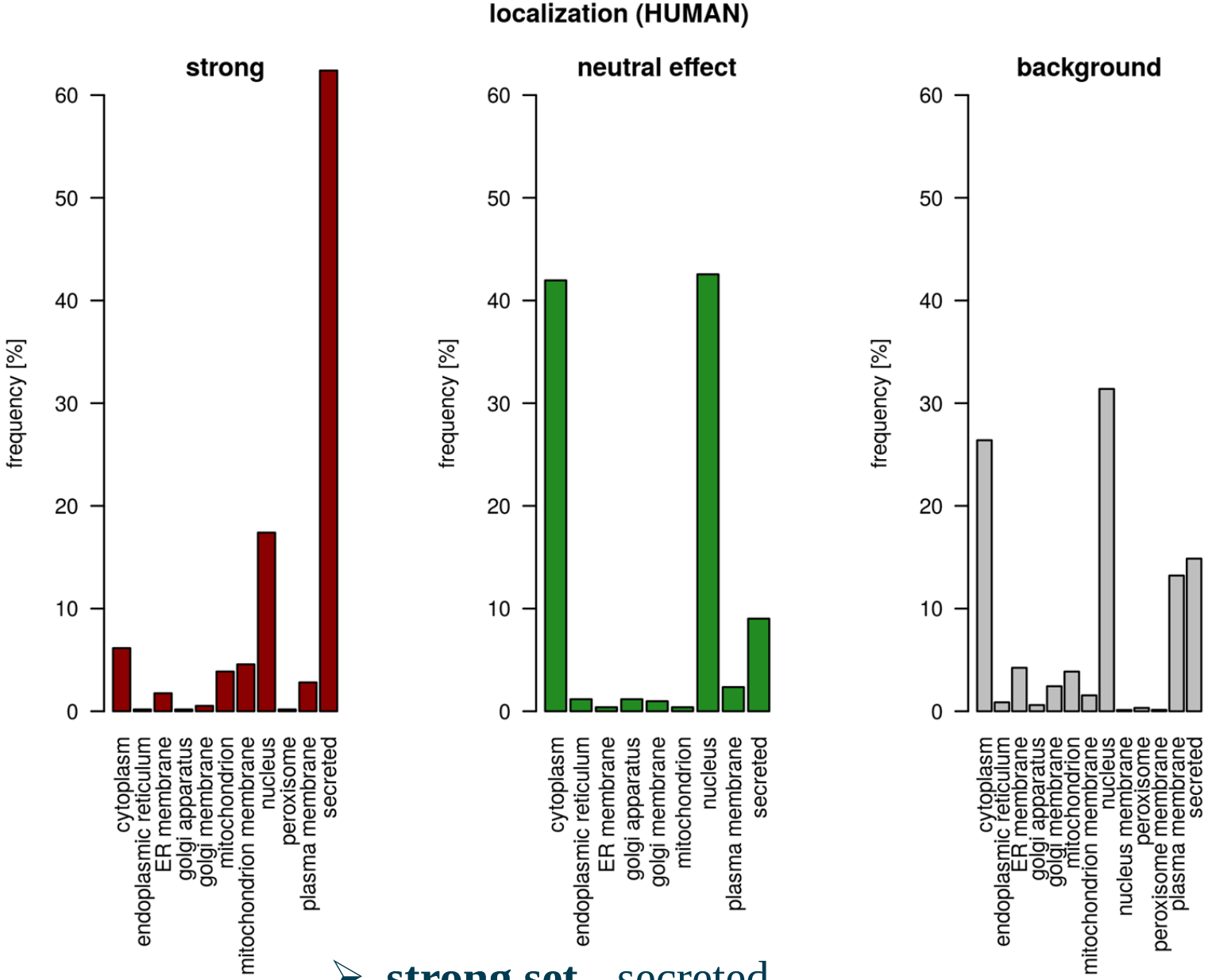
➢ more binding sites for **strong** although "smaller proteins"

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1920242/



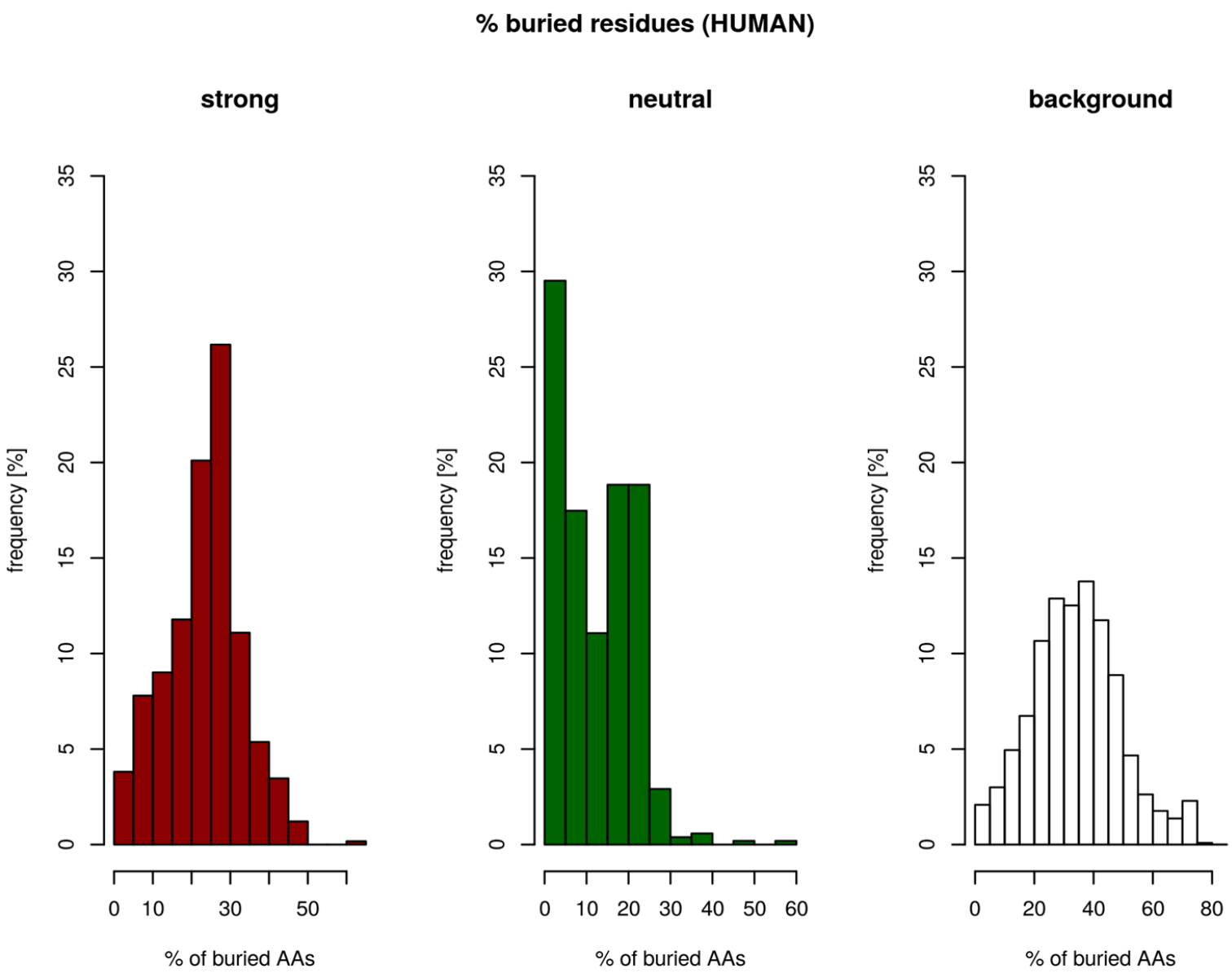
- **strong set** - more short proteins (11-232 residues long)
- **neutral set** - more long proteins (125-5654 residues long)

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1920242/



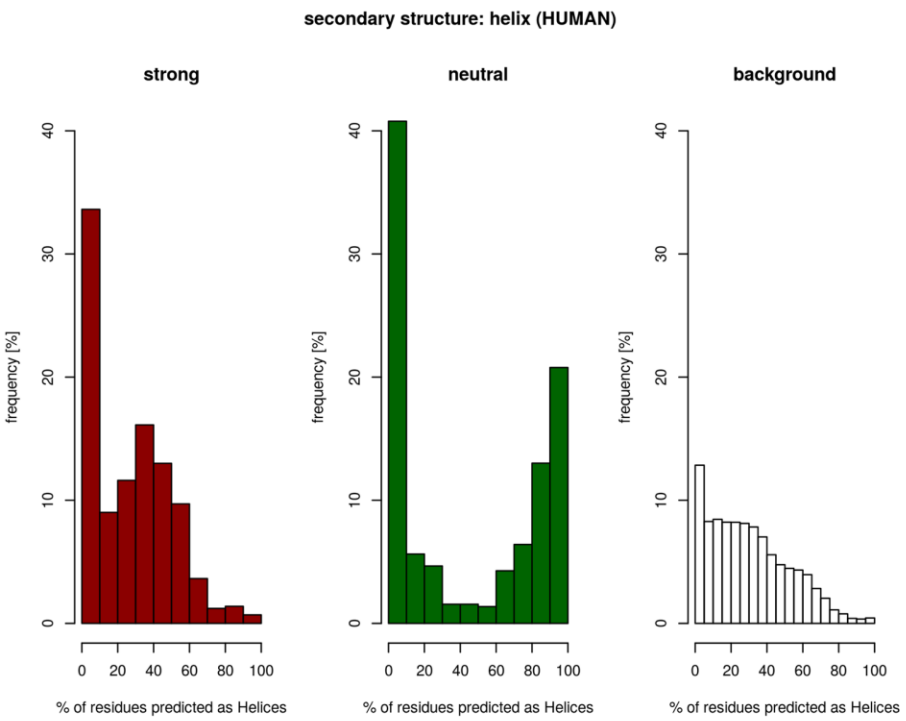
- **strong set -** secreted
- **neutral set -** cytoplasm, nucleus

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1920242/

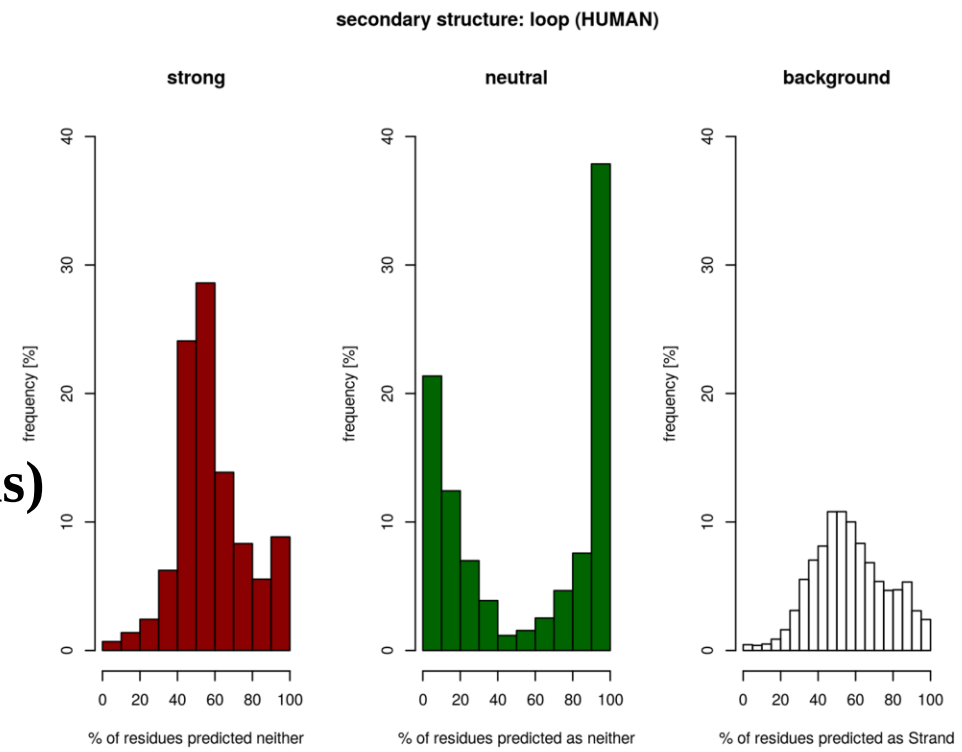


- **strong set** - more buried (10-40%)
- **neutral set** - less buried (0-25%)

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1920242/



e.g. 40% of proteins from neutral set are predicted as 0% helix



Strong

- 40% without strands
- versatile structure (mix of loops, helices and strands)

Neutral

- 70% without strands
- "all-or-nothing" loop and helices

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1920242/

Keywords for SNPs predictions

Human Keywords					
Strong effect		Strong neutral		Background	
Polymorphism	4,08%	Polymorphism	7,98%	Polymorphism	5,47%
Signal	3,47%	Phosphoprotein	6,55%	Phosphoprotein	3,41%
Membrane	2,91%	Cytoplasm	3,85%	Membrane	3,22%
Transmembrane	2,61%	Nucleus	3,19%	3D-structure	2,56%
Phosphoprotein	2,22%	Cytoskeleton	2,16%	Transmembrane	2,38%
Secreted	2,05%	Repeat	2,02%	Nucleus	2,33%
Acetylation	2,03%	Membrane	1,62%	Repeat	2,25%
3D-structure	1,94%	Transcription	1,08%	Cytoplasm	2,14%
Nucleus	1,80%	Metal-binding	0,85%	Glycoprotein	2,07%
Repeat	1,22%	Zinc	0,70%	Signal	1,67%

A. Thaliana

Secreted
Antimicrobial
Fungicide

DNA-binding
Transport
Zinc
Phosphoprotein

Hydrolase
Nucleotide-binding
ATP-binding

E. Coli

Ribonucleoprotein
DNA-binding
Toxin
Transcription
tRNA-binding

Periplasm

Metal-binding
Transferase

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1920242/

Thank you

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1920242/