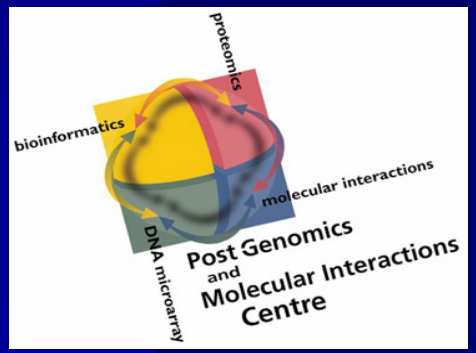
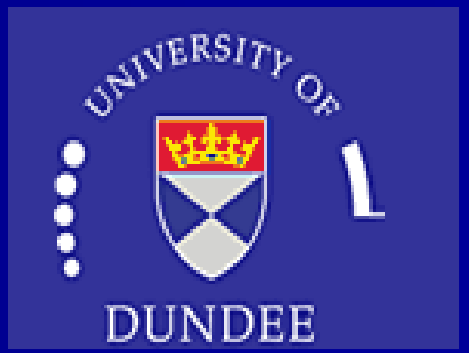
A Normalised Scale To Estimate Protein "Crystallisation Propensity" - The OB-Score

Ian Overton
University of Dundee
ian@compbio.dundee.ac.uk

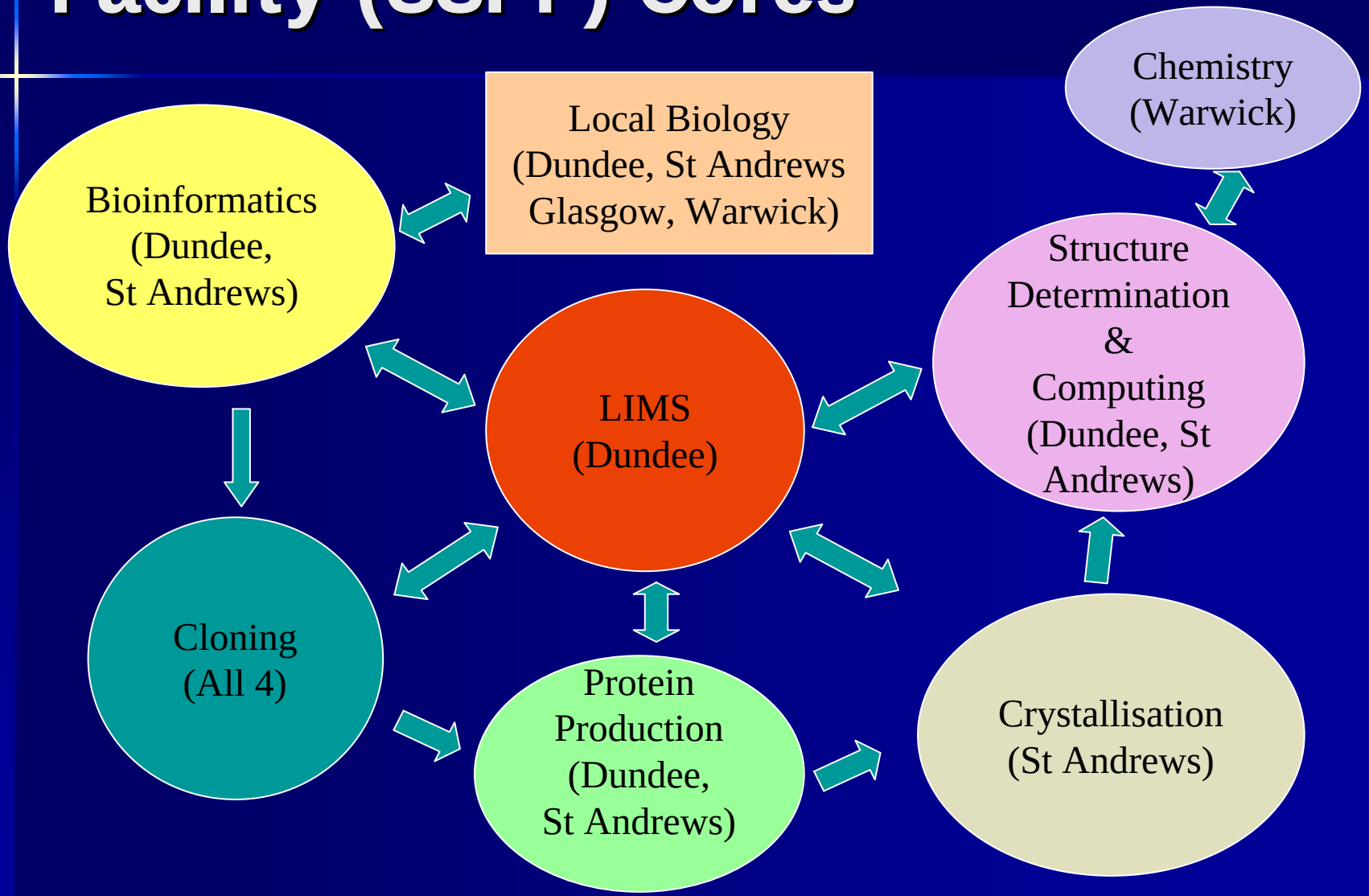




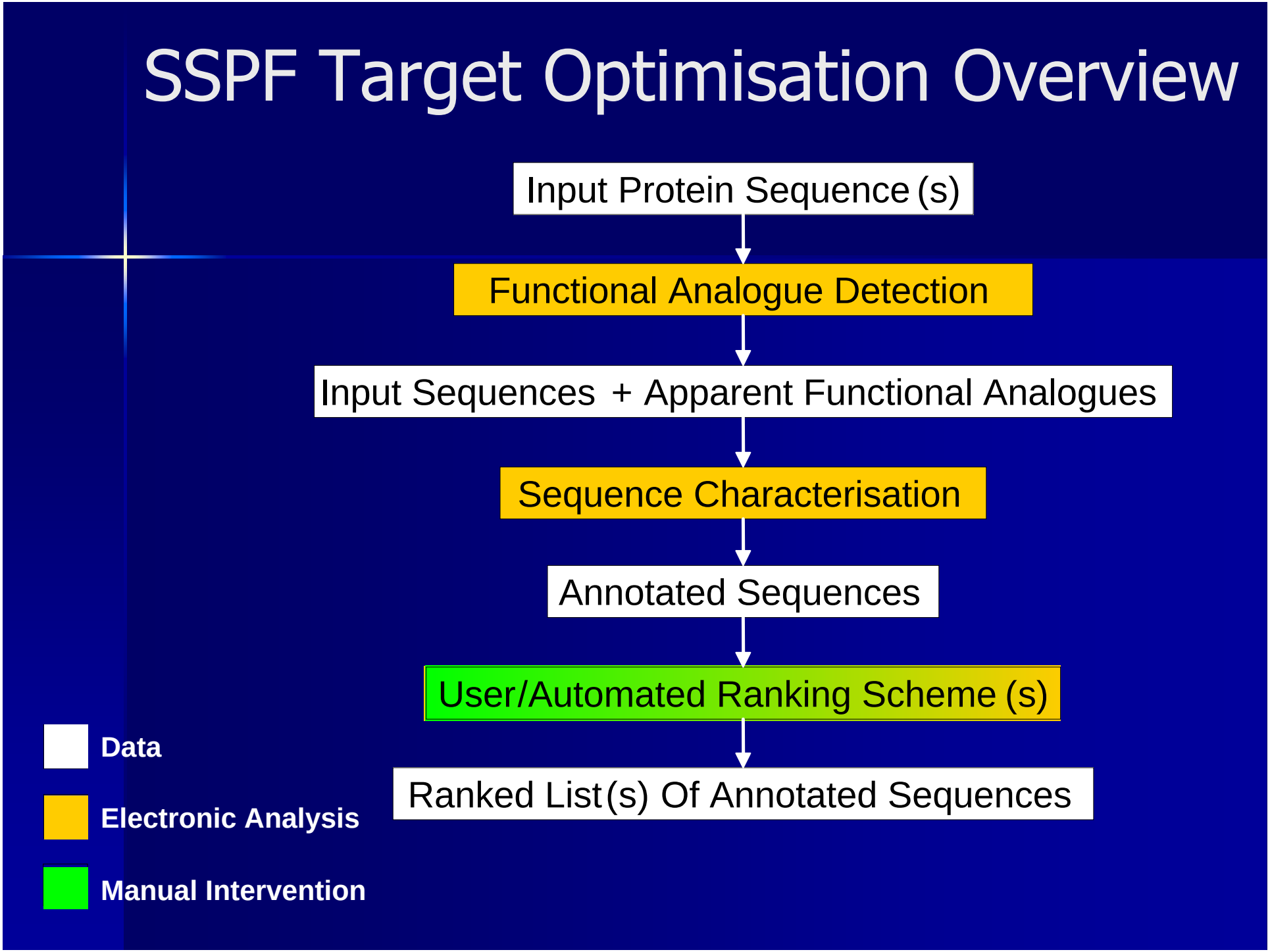
Overview

- Background
- Development of the OB-Score
- Validation
  - TargetDB test datasets
- Application to
  - 241 Proteomes
  - 7,868 Pfam families

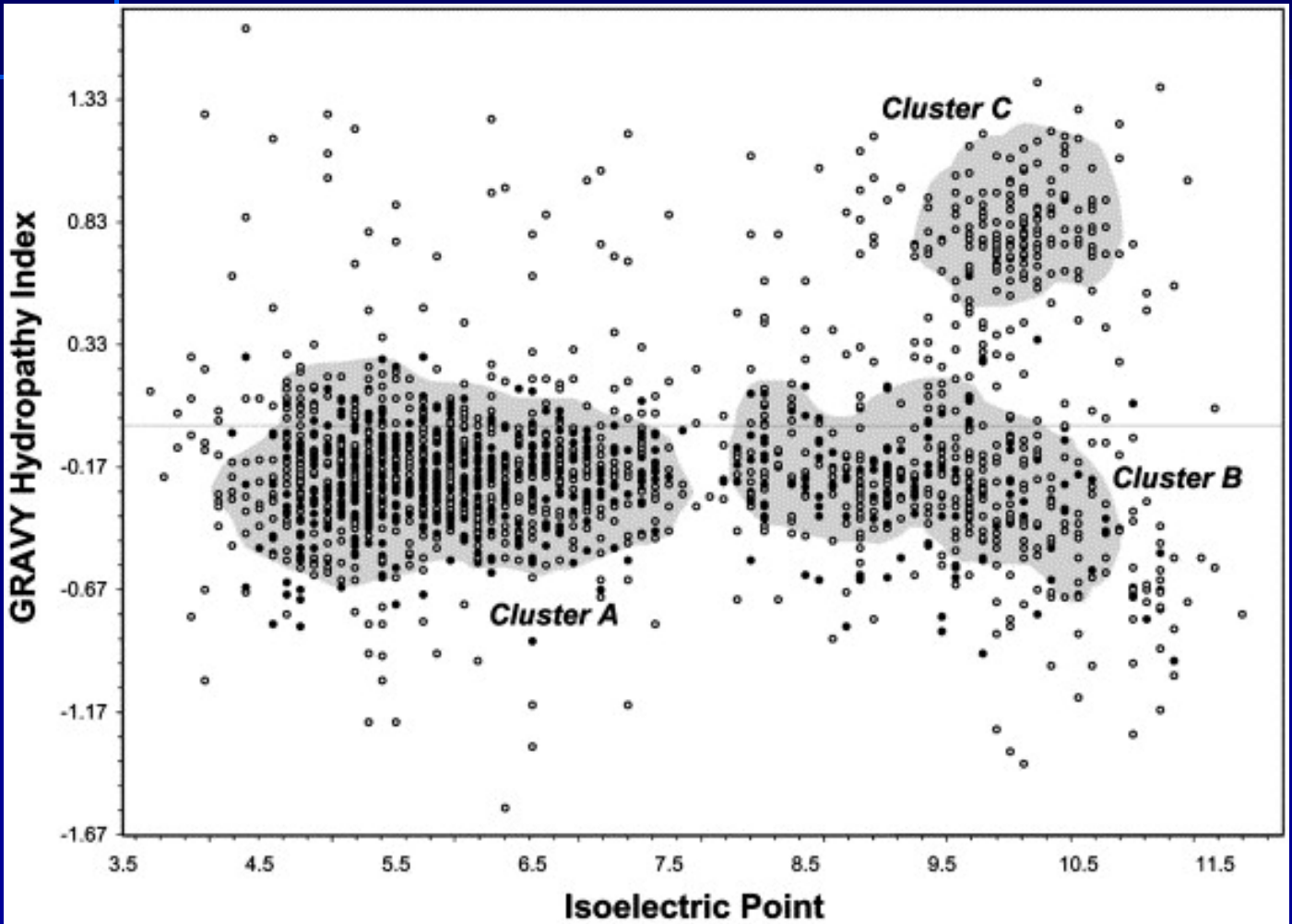
Scottish Structural Proteomics Facility (SSPF) Cores



SSPF Target Optimisation Overview



T. Maritima Correlates of Crystallisation Success

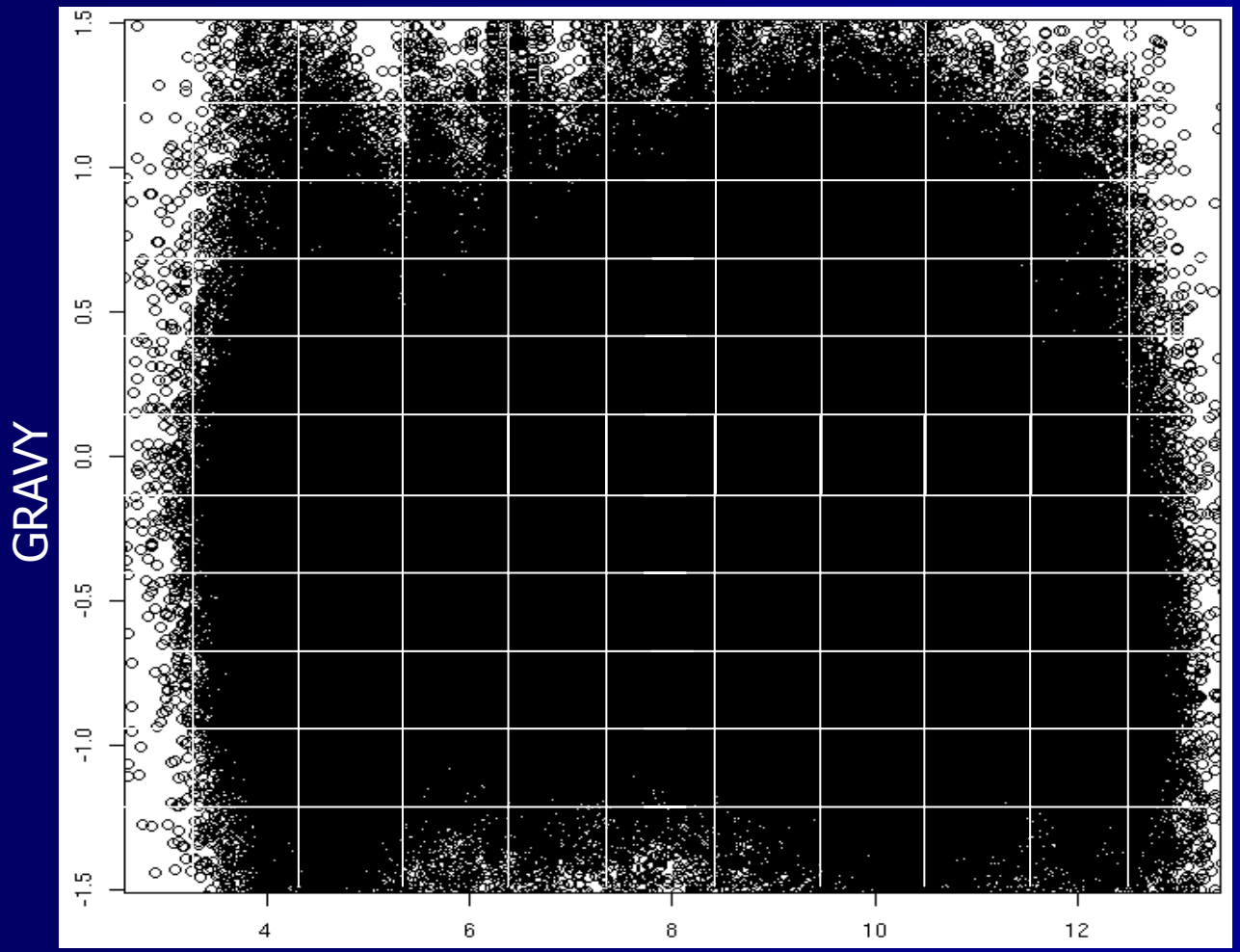


Hydropathy vs pI for *T. maritima* proteome.

Filled circles are proteins which crystallised, open circles are proteins which did not crystallise.

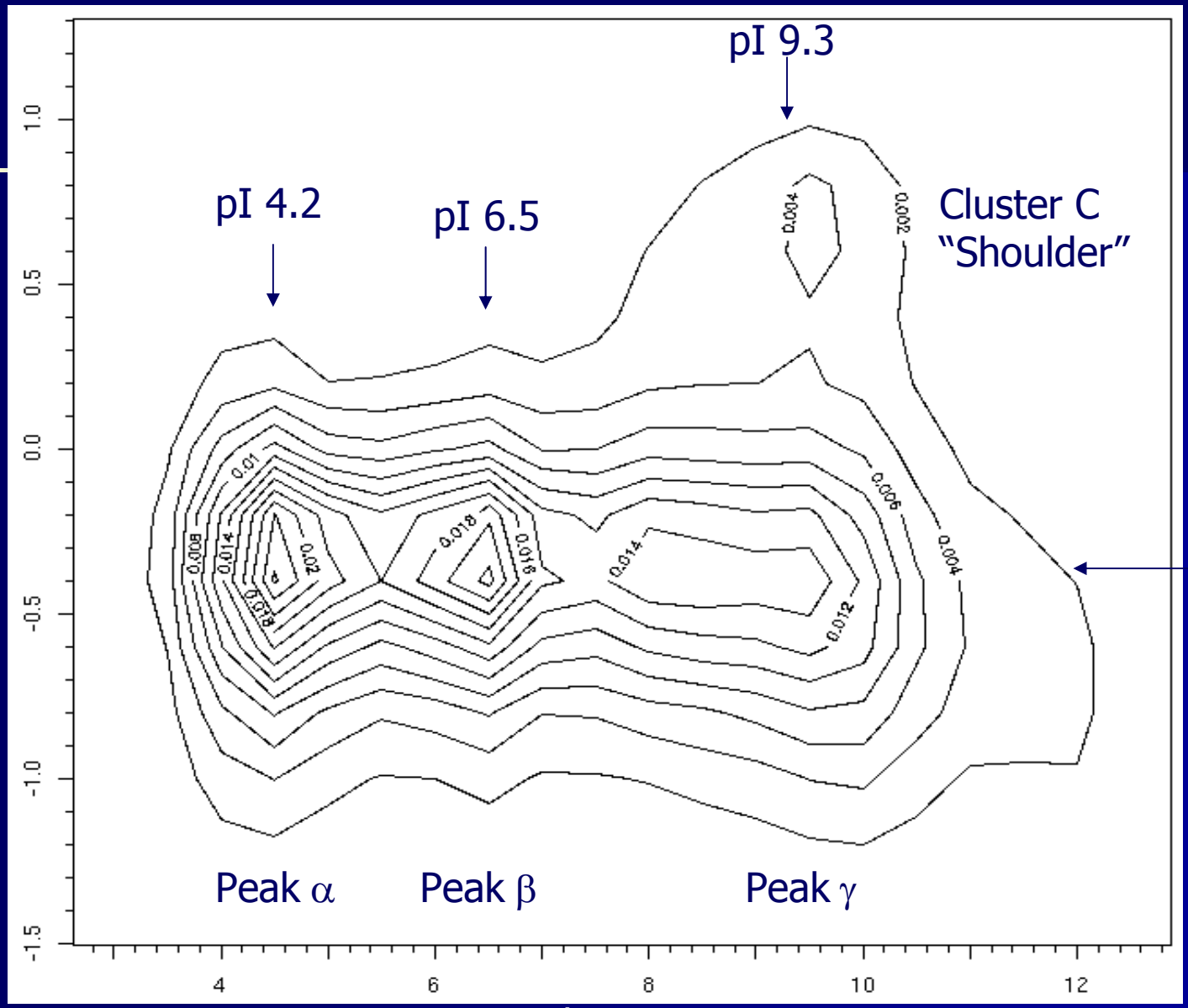
Fig. 6 from: Canaves *et al*. (2004) *J. Mol. Biol*. **344**, 977-991

Development of the OB-Score



UniRef50 GRAVY vs pI scatterplot.

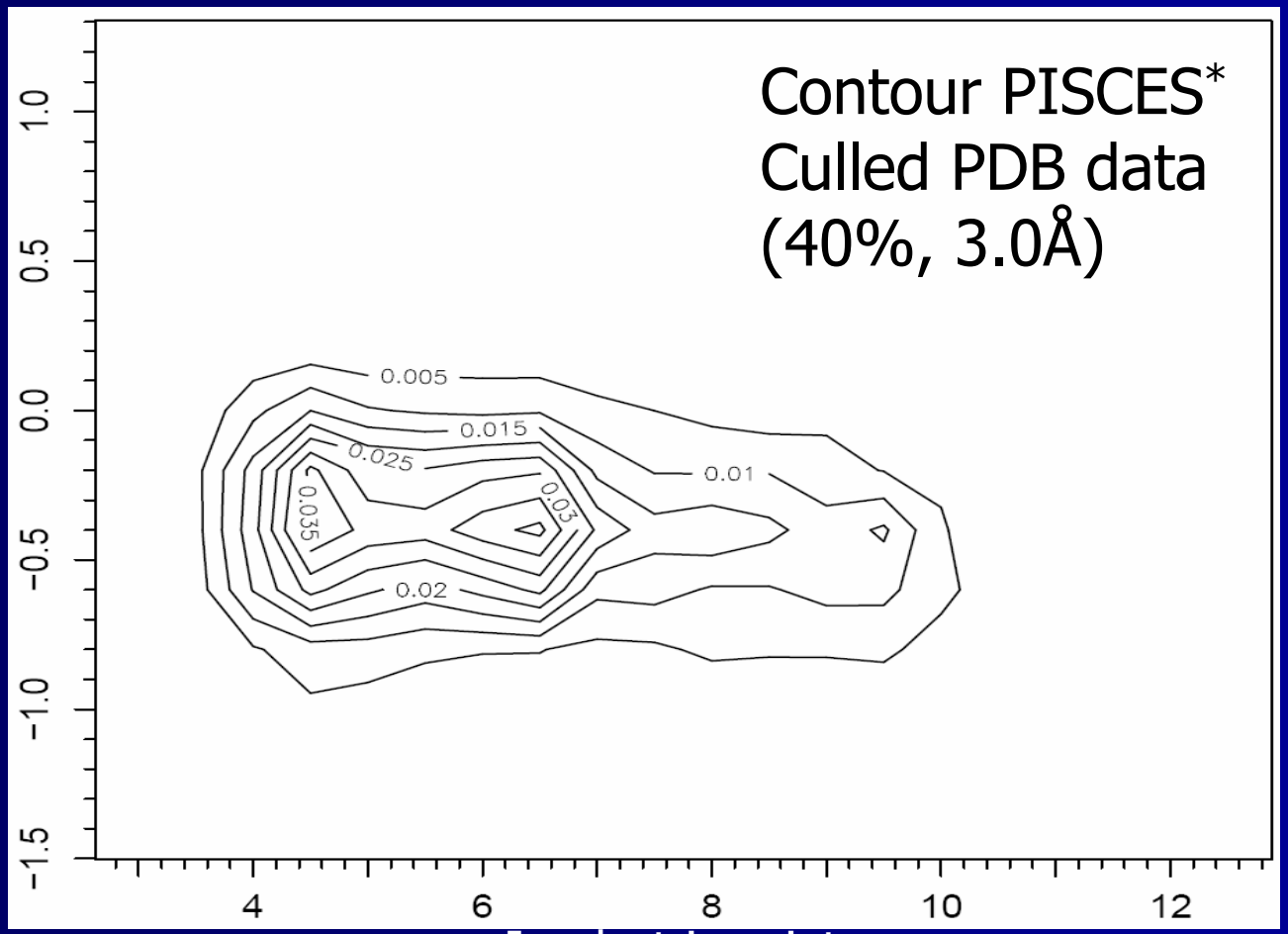
Development of the OB-Score



Contoured UniRef50 Background

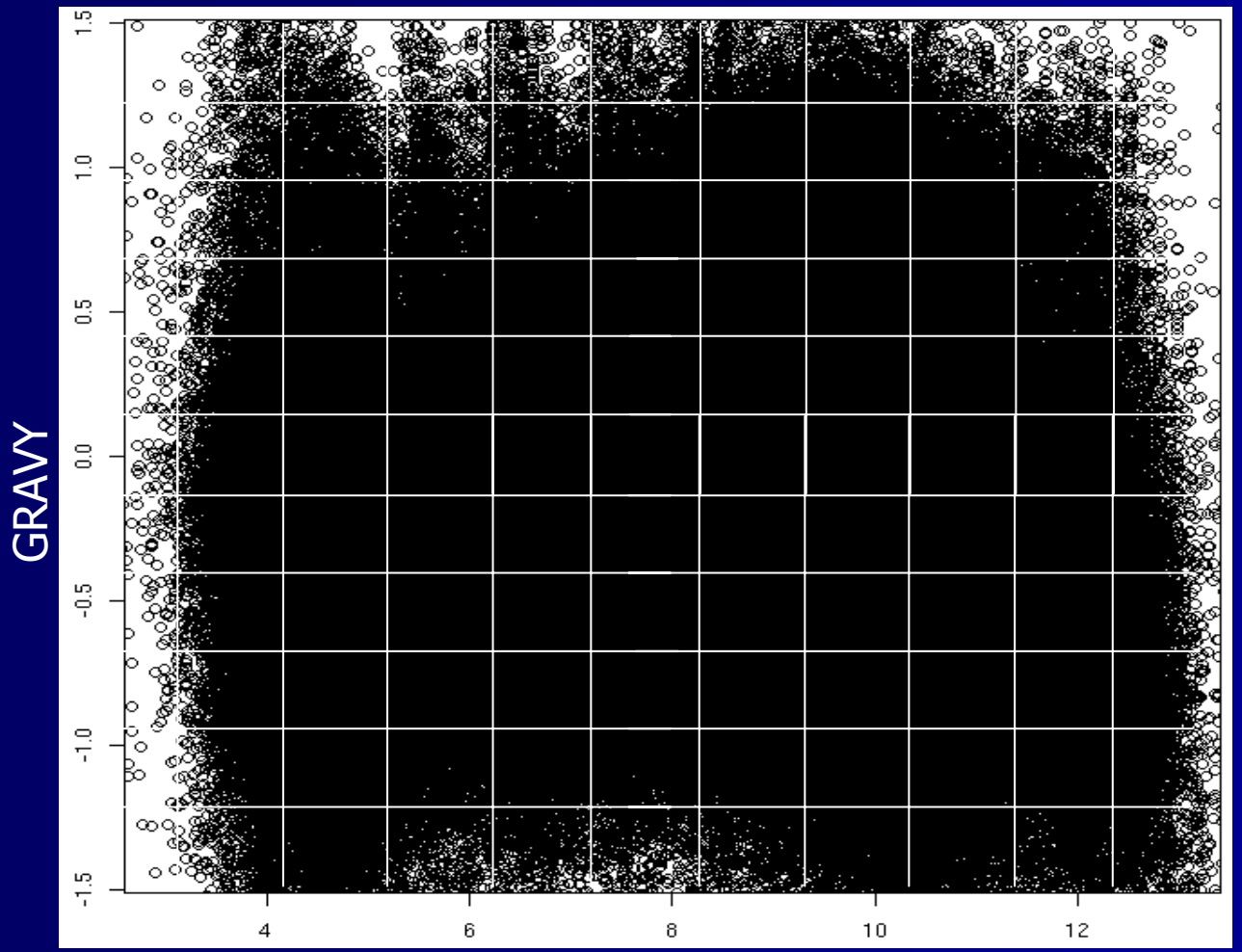
GRAVY -0.4

Development of the OB-Score



*Wang & Dunbrack (2003) *Bioinformatics* **19**, 1589-1591

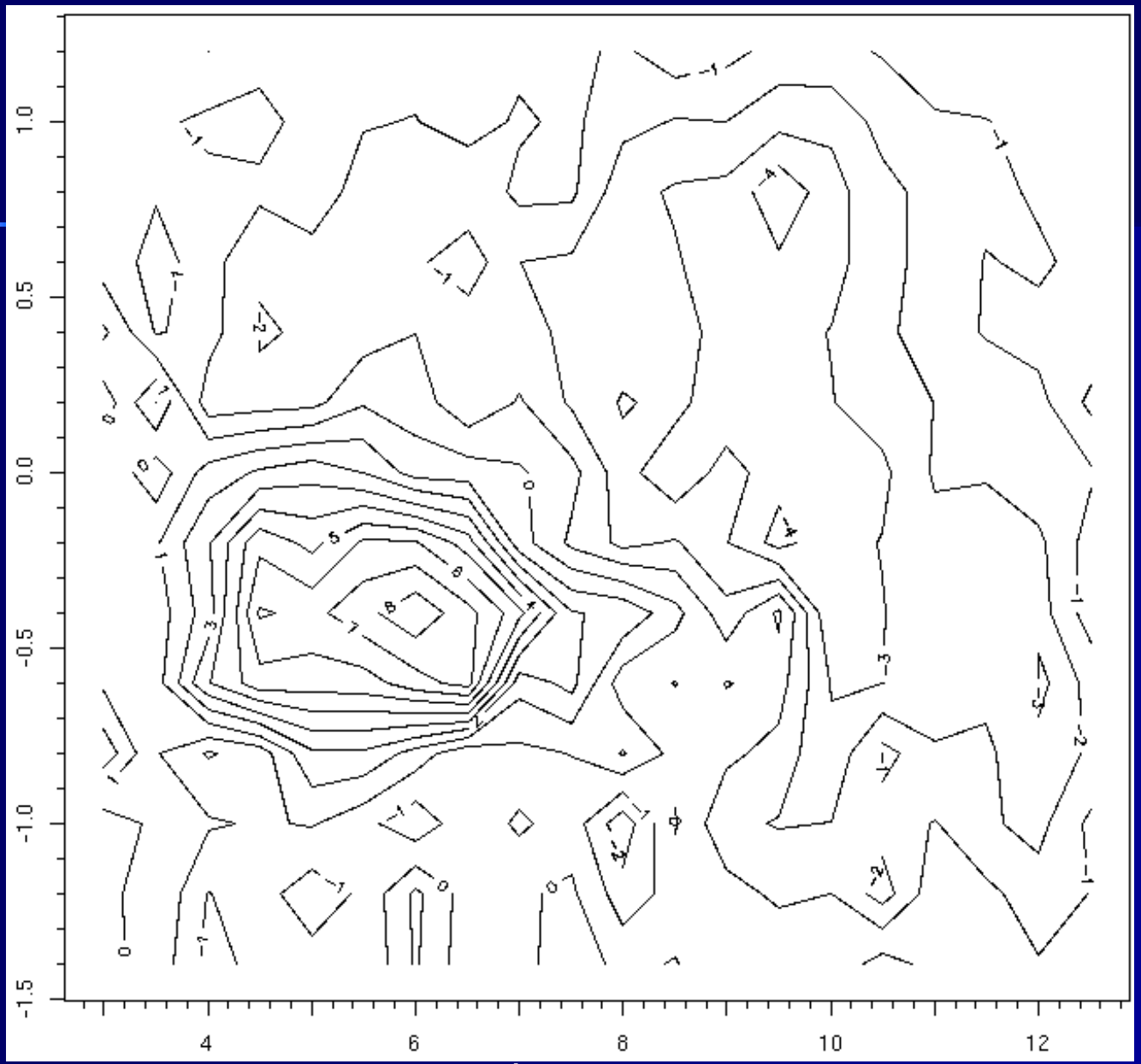
Development of the OB-Score



Sample UniRef50 100 times, 5000 sequences per sample

Calculate mean and standard deviation of the normalised frequencies in each matrix cell

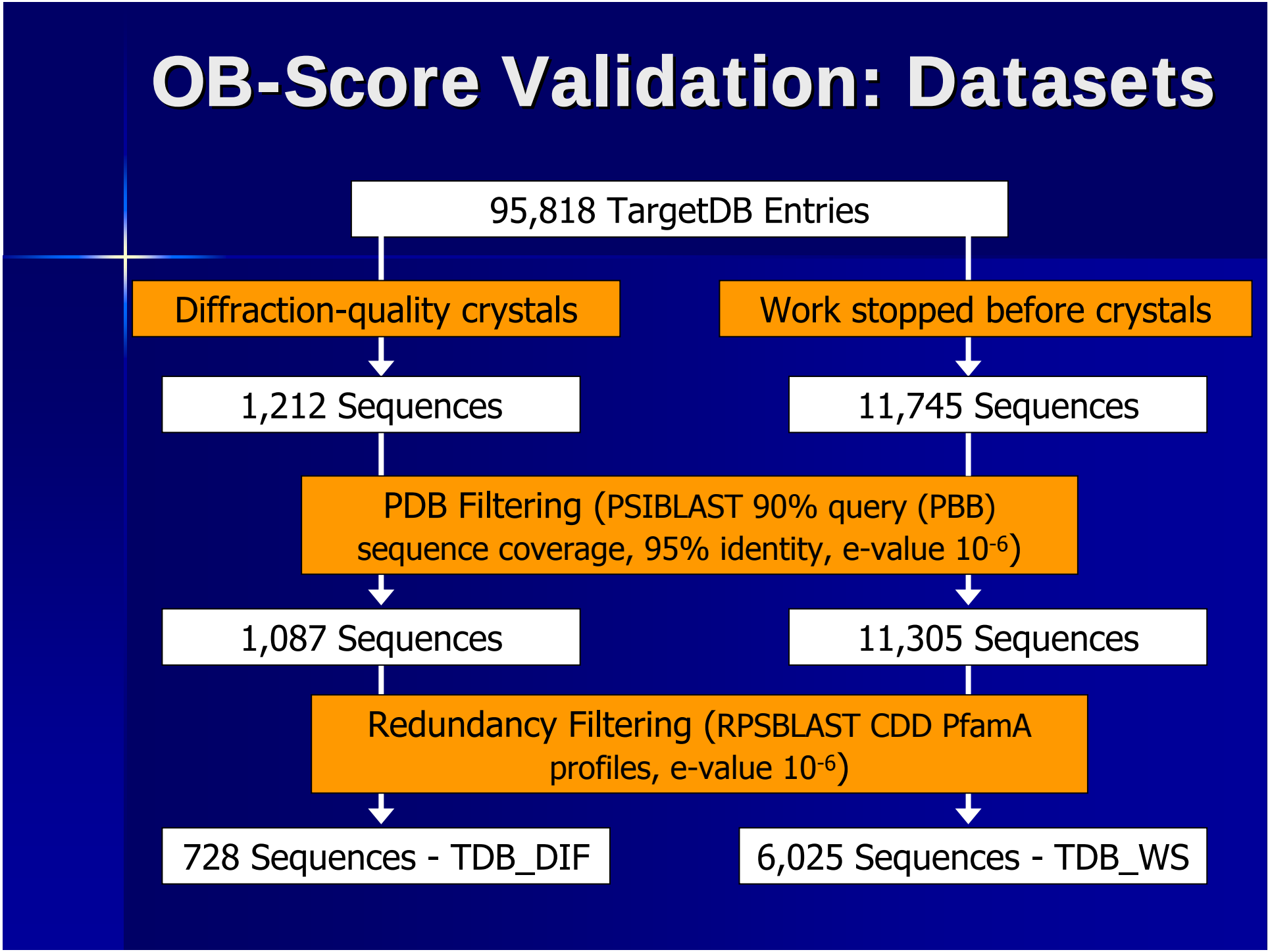
Z-Score Contours for PISCES Culled PDB (40%, 3.0Å)

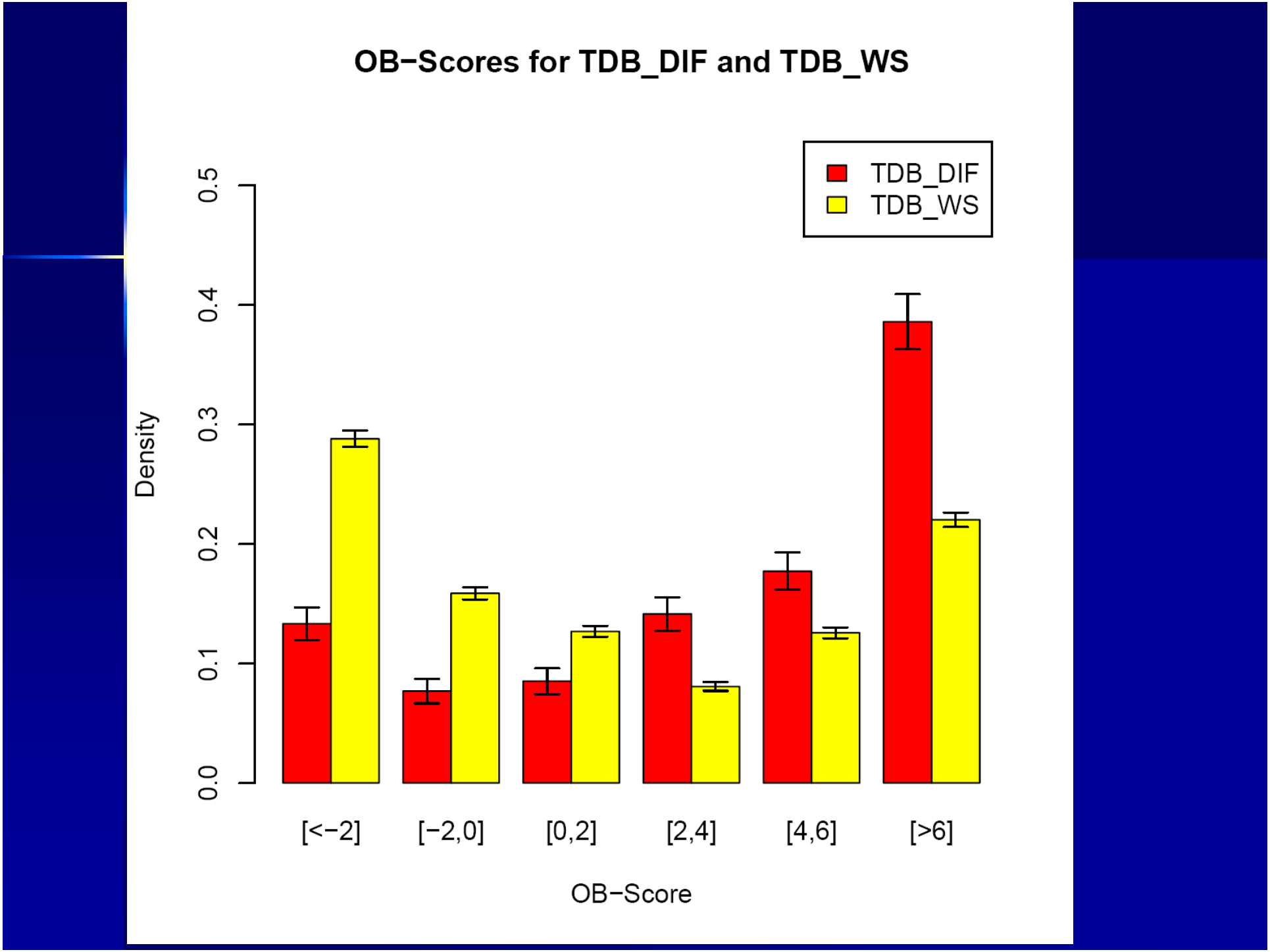


5454 PDB sequences plotted onto pI/GRAVY matrix.

Z-Scores derived from UniRef50 background distribution in each matrix cell

OB-Score Validation: Datasets




OB−Scores for TDB_DIF and TDB_WS
TDB_DIF
TDB_WS
Density
0.0
0.1
0.2
0.3
0.4
0.5
[<−2]
[−2,0]
[0,2]
[2,4]
[4,6]
[>6]
OB−Score

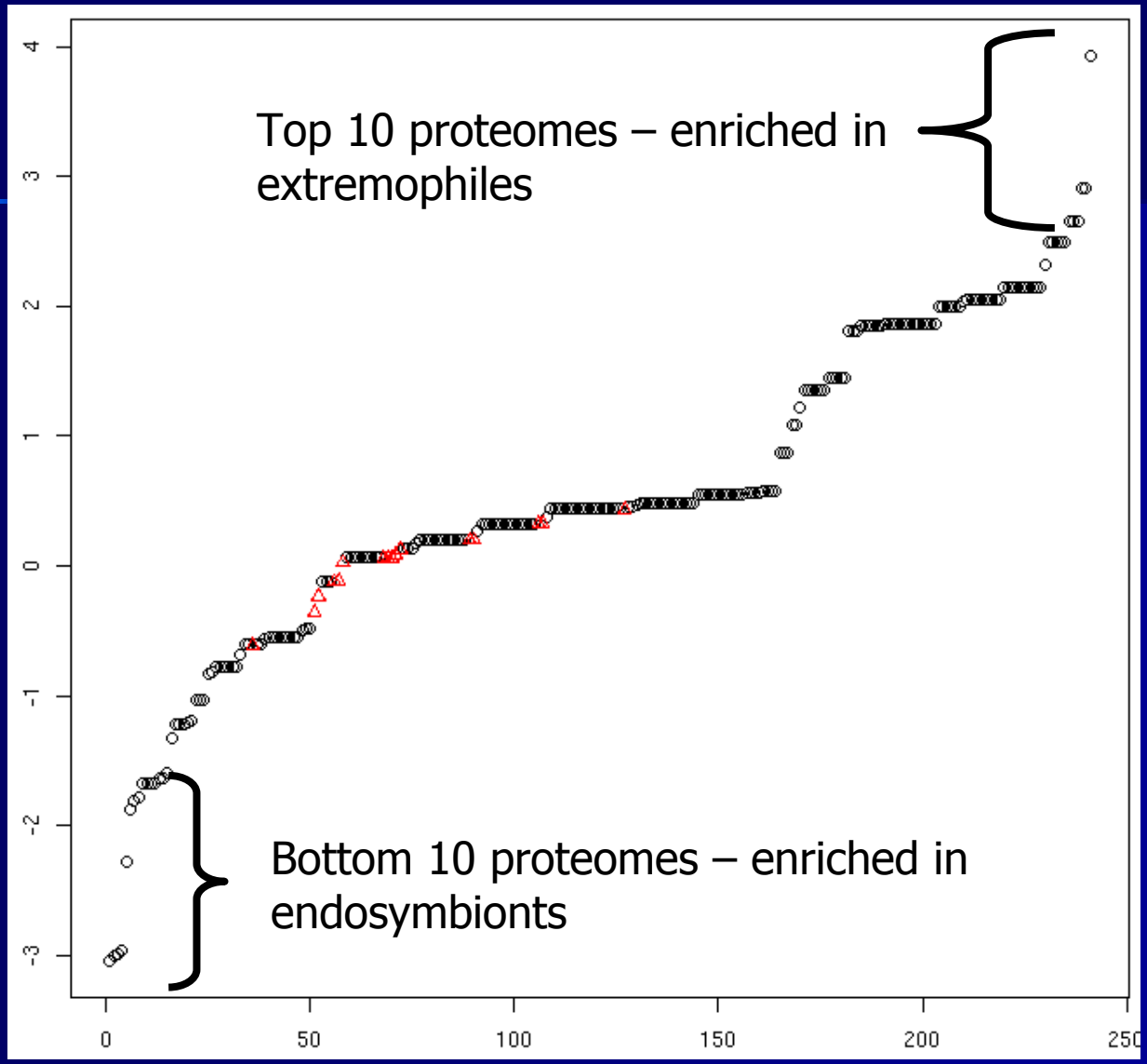
Application to 241 Proteomes & 7868 Pfam Families

- Which organisms are the crystallographer's friend?
- Which Pfam family members are most suitable for structural genomics?

Ranking Protein Families & Proteomes

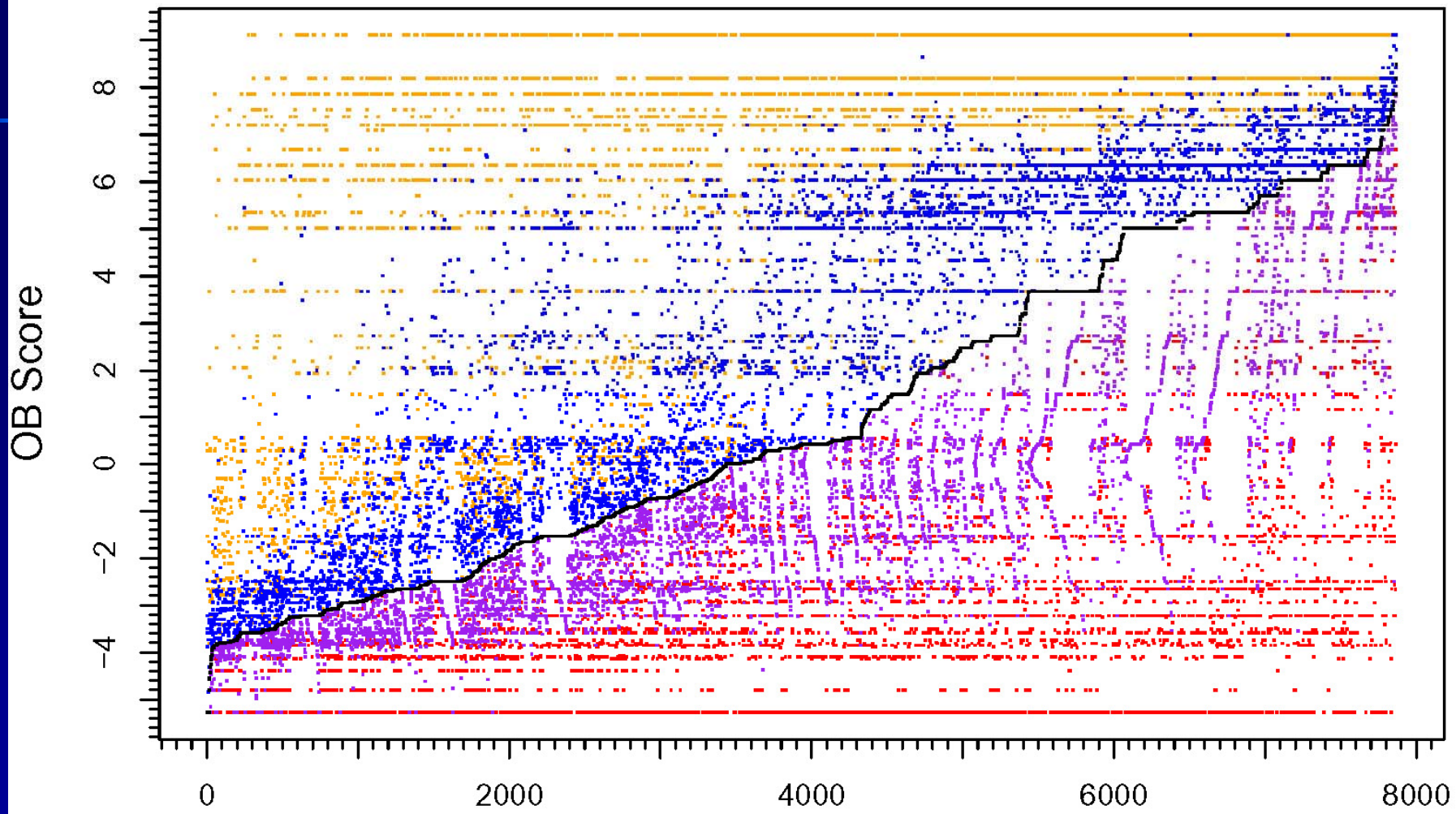
- Look up OB-score for each protein in a proteome/Pfam family
- Sort proteomes/Pfam families on their median OB-score values
- Plot…

Median OB-Scores for 225 Prokaryotic & 16 Eukaryotic Proteomes



Red triangles show Eukaryotic proteomes

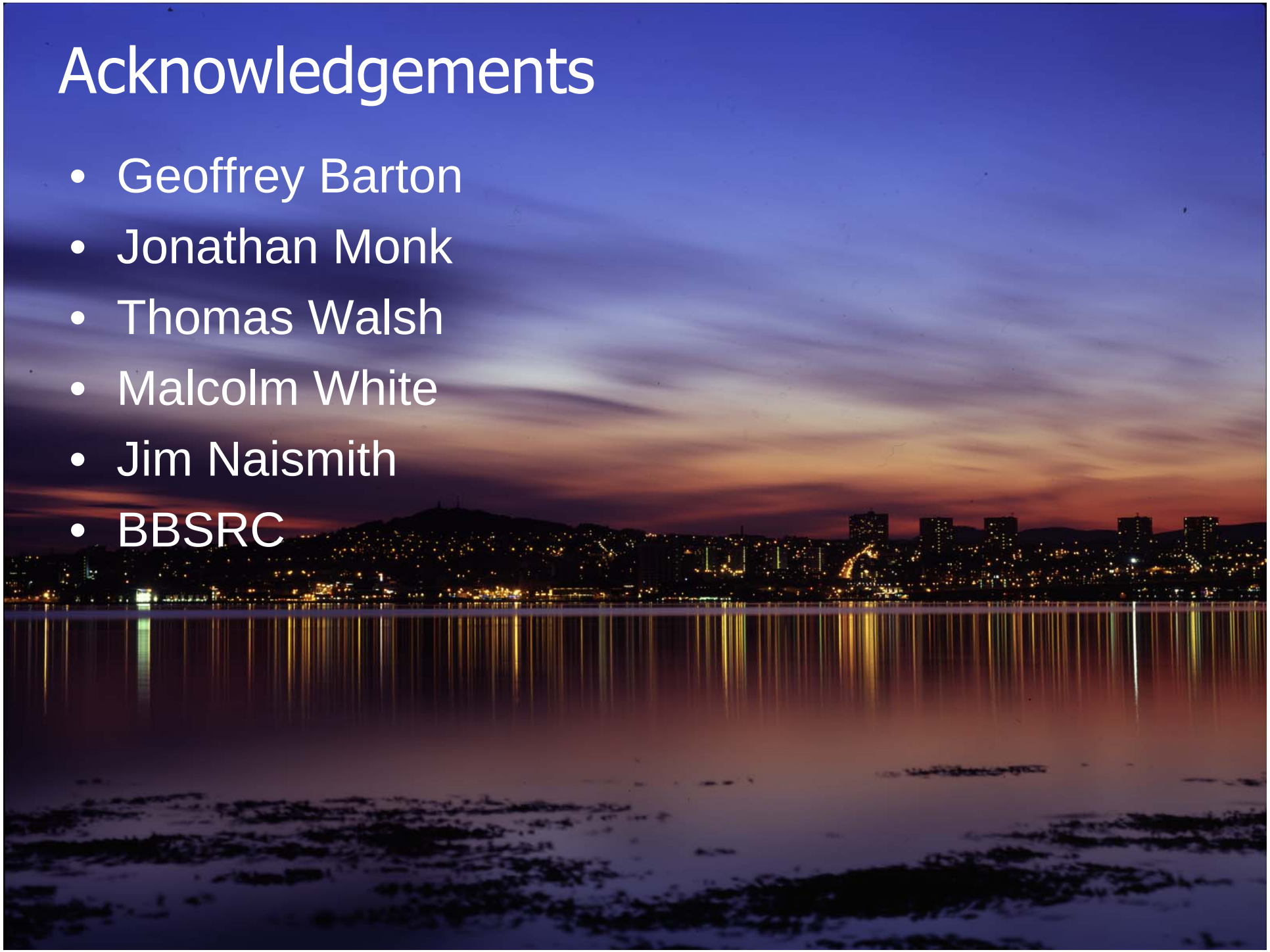
OB Score Statistics Across Pfam



73.4% Families at least
1 member with OB ≥5

Availability

- Data & Software:
  - www.compbio.dundee.ac.uk/obscore
- Also see:
  - Overton & Barton (2006) *FEBS Letters (in press)*



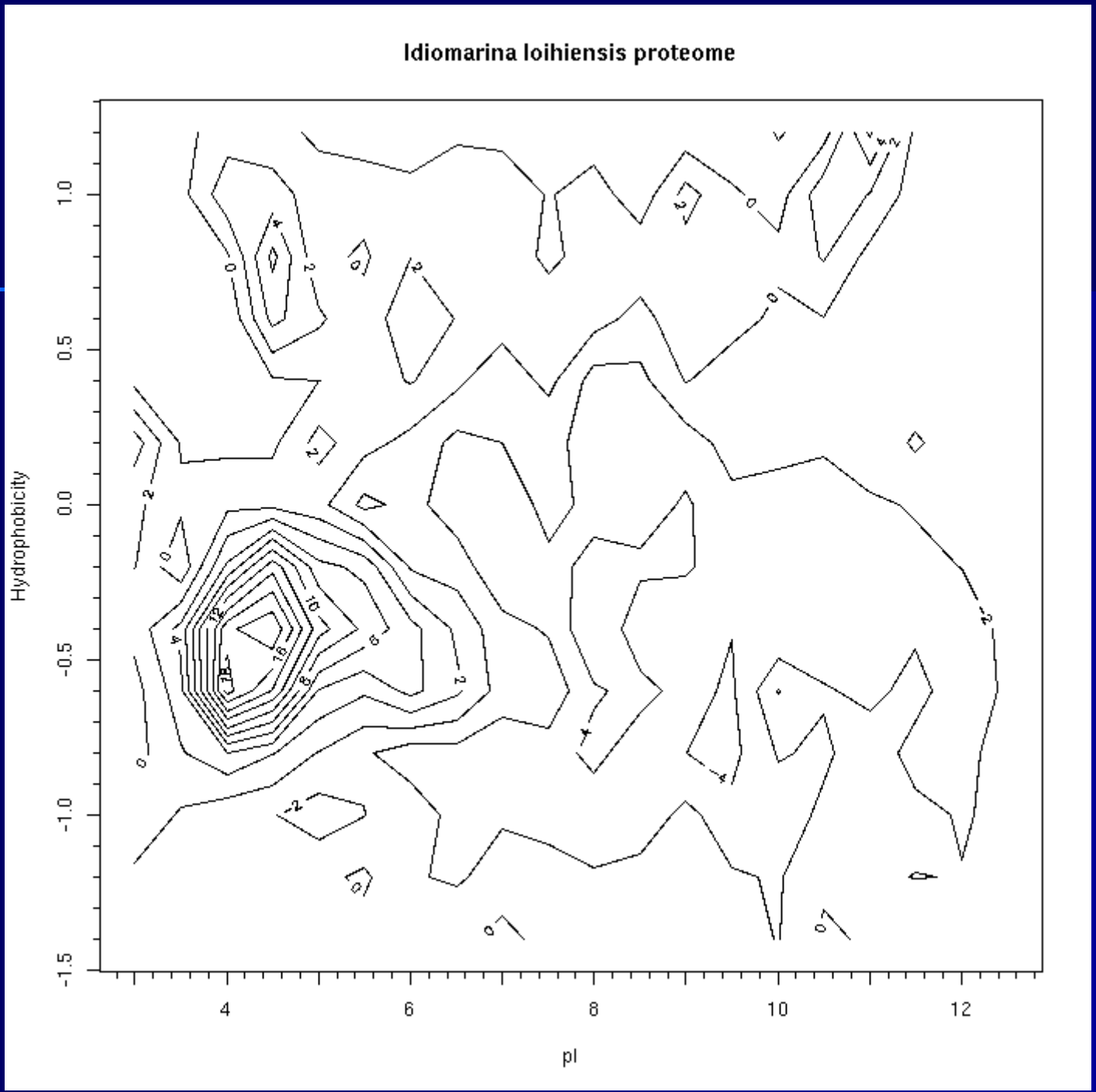
Acknowledgements

- Geoffrey Barton
- Jonathan Monk
- Thomas Walsh
- Malcolm White
- Jim Naismith
- BBSRC

Statistics Across Datasets

Dataset	N	Length				pI		GRAVY	
		Min	Median	Mean	Max	Median	Range	Median	Range
TDB_DIF	728 (1087)	12 (12)	252.5 (262.0)	282.3 (289.1)	1727 (1727)	5.9 (5.9)	8.4 (8.4)	-0.23 (-0.21)	2.4 (2.4)
TDB_WS	6025 (11305)	24 (24)	194.0 (238.0)	242.2 (277.9)	2514 (6048)	7.2 (6.9)	10.7 (10.7)	-0.36 (-0.31)	5.0 (5.0)
Culled_PDB	5454	20	254.4	220.5	1733	6.4	9.9	-0.30	4.9
PfamA	961405	5	150.5	117.0	2799	6.8	11.6	-0.16	6.13
UniRef50	794085	11	360.1	265.0	34350	7.5	12.3	-0.25	7.4

Z-score contours



Z-score contours