



8th Meeting from Genome to Proteome, Siena, Sep 2, 2008

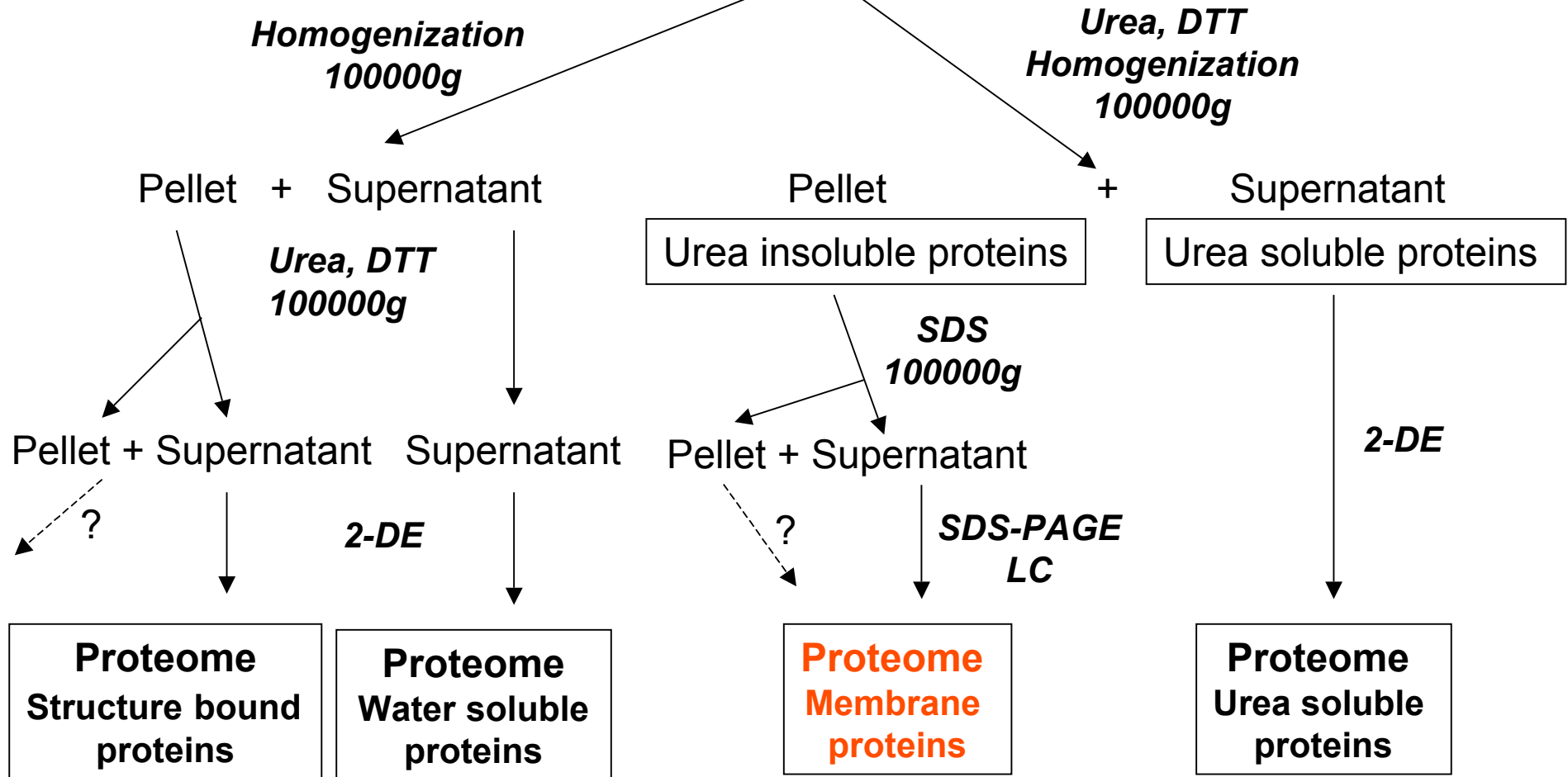
**Data Mining of 2-DE/MS and 1-DE-LC/MS
Helicobacter pylori Proteome Data: Fate of
membrane proteins**

**Peter R. Jungblut, Robert Stein, Klaus-Peter Pleißner,
Thomas F. Meyer**

**Max Planck Institute for Infection Biology,
Core Facility for Protein Analysis,
Berlin**

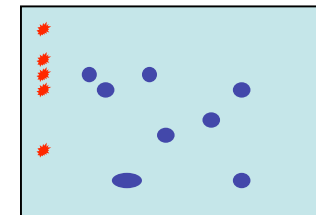
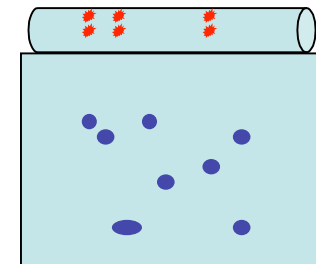
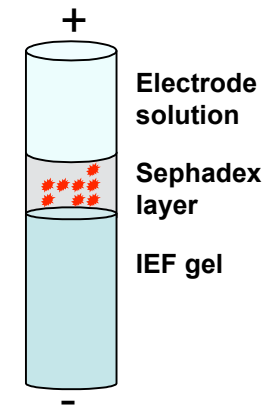
Hypothesis

Cells



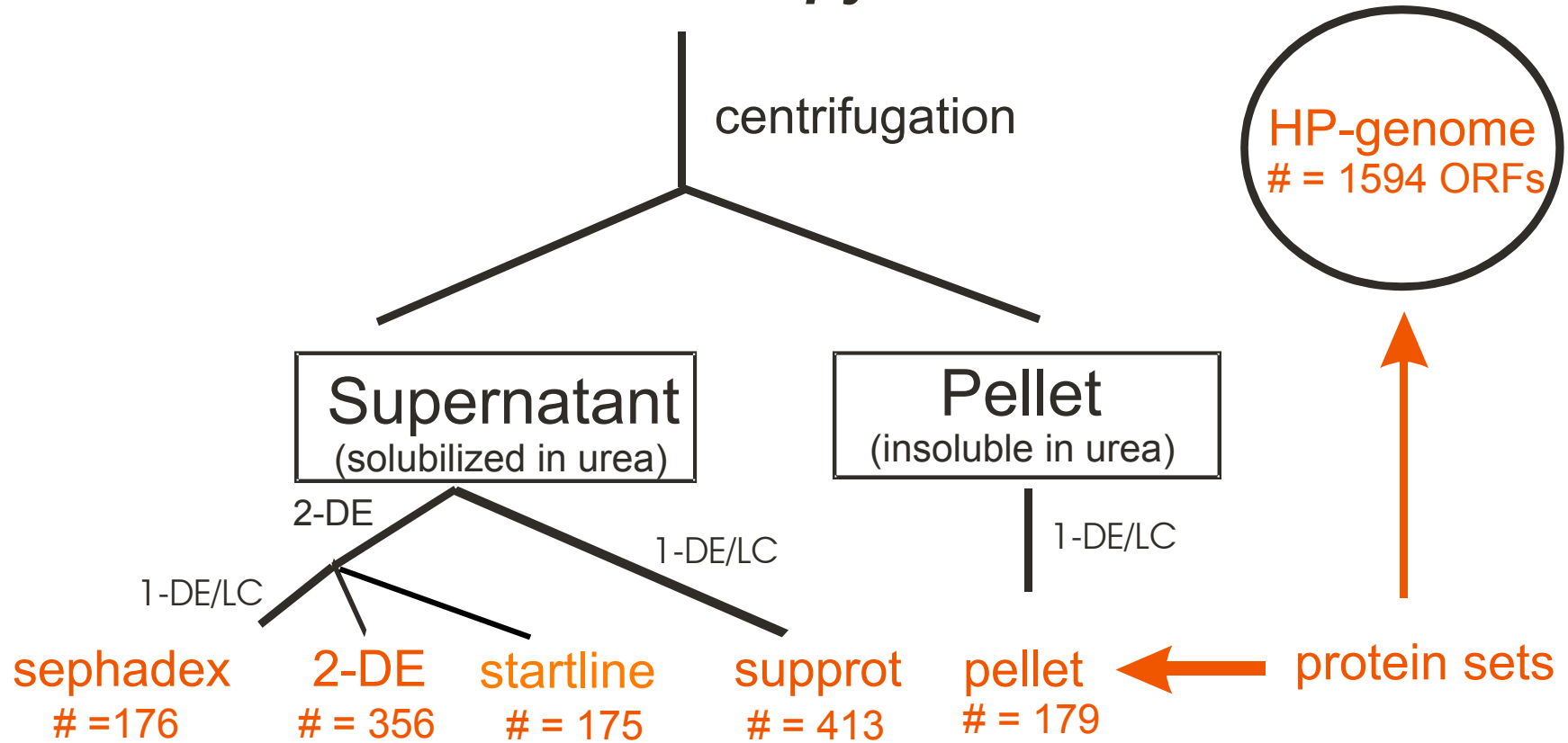
Where could the membrane proteins be lost?

- Membrane proteins are not solubilized in urea.
- Precipitation at the application site of IEF: in the case of NEPHGE within the Sephadex layer
- Membrane proteins are rare proteins and below our analysis window.
- Membrane proteins are precipitated at their pI and do not move out of the IEF gel into the SDS-PAGE gel.
- Membrane proteins are lost at surfaces.
- Membrane proteins are precipitated at the start region of IEF gel and appear as start line within the 2-DE gel.

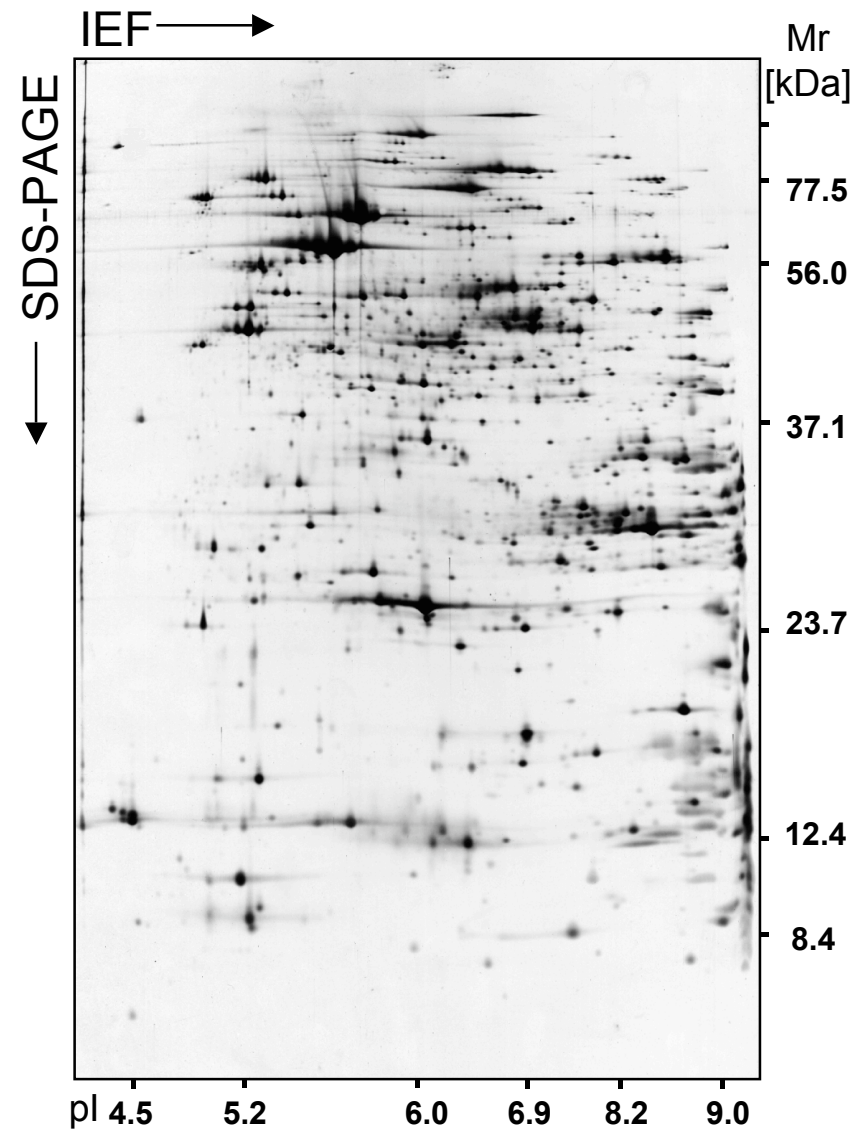


Protein preparations investigated

Helicobacter pylori



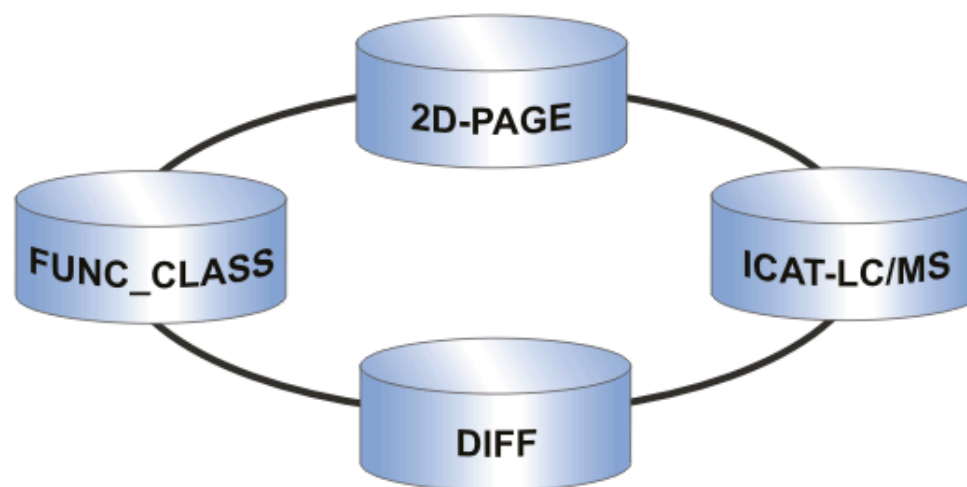
2-DE *H. pylori* 26695





<http://www.mpiib-berlin.mpg.de/2D-PAGE>

Proteome Database System for Microbial Research at Max Planck Institute for Infection Biology



in cooperation with
 BROAD
INSTITUTE



PROMPT

Protein Mapping and Comparison Tool

Menu:

- Home
- Screenshots
- Case Studies
- Start PROMPT
- Downloads
- Mailinglist
- Help + Contact

Imprint



open source

Free for academic,
commercial users please
request a [free evaluation
license](#).

PROMPT: *Protein Mapping and Comparison Tool*

PROMPT is a platform independent system for retrieval, analysis, mapping and comparison of protein sets. It allows easy mapping of different types of sequence identifiers, automatical data retrieval and integration, a multitude of analysis and comparison algorithms and a full-featured easy to use graphical user interface (GUI) application with an integrated help-system. Moreover exhaustive statistical tests are conducted in appropriate cases automatically or could be performed manually. Furthermore all analysis results can be viewed or visualized with one mouse-click and exported in various formats e.g. to Microsoft Excel. Additionally all methods can be used easily in your own Java code or with powerful beanshell scripting in own scripts, a pipeline or grid systems.

[Start PROMPT with Java Web Start](#)

[Download Standalone](#)

[Screenshots](#)

[Case Studies](#)

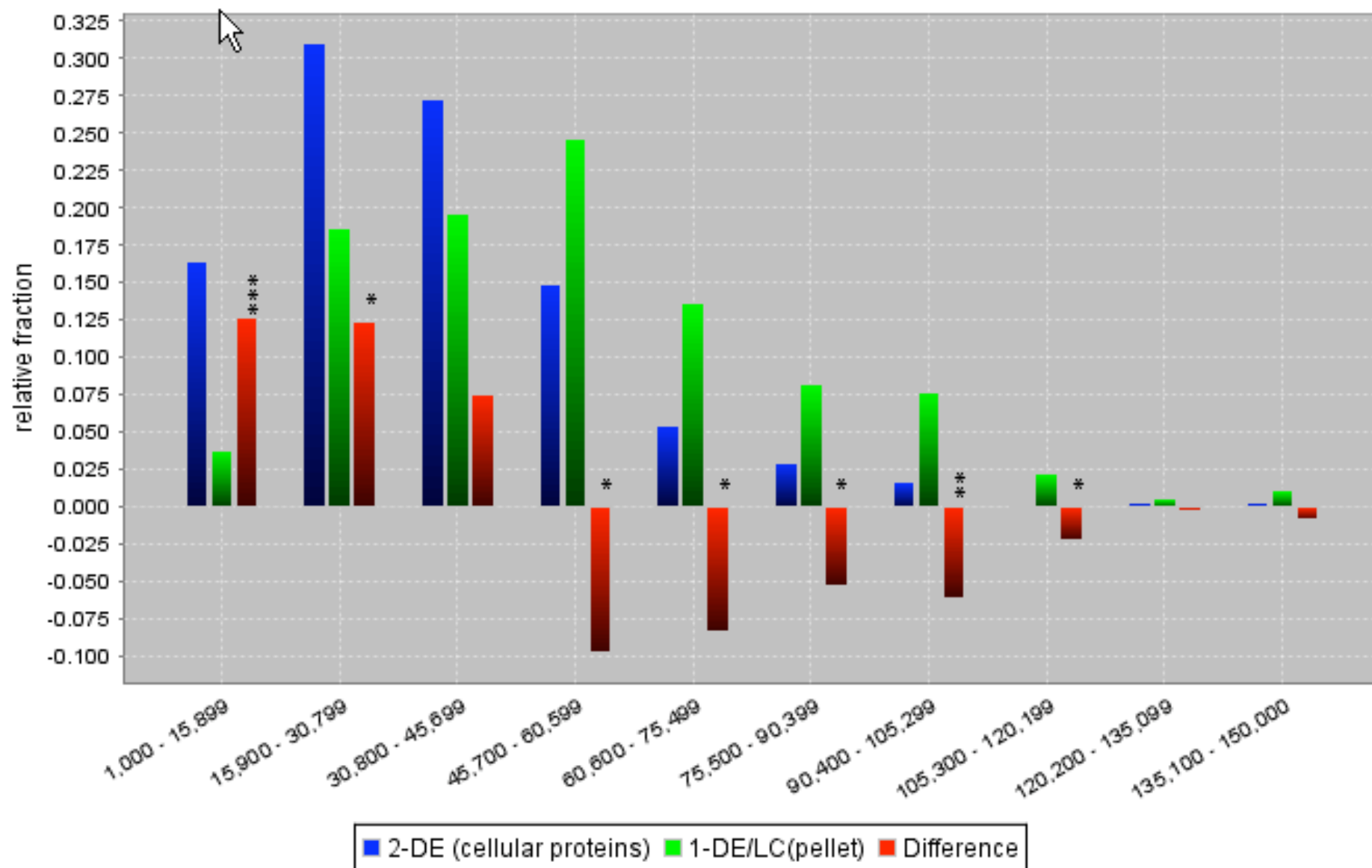
New: [Web based Fast Mapping to large databases](#) by F. Hager

<http://webclu.bio.wzw.tum.de/prompt/>

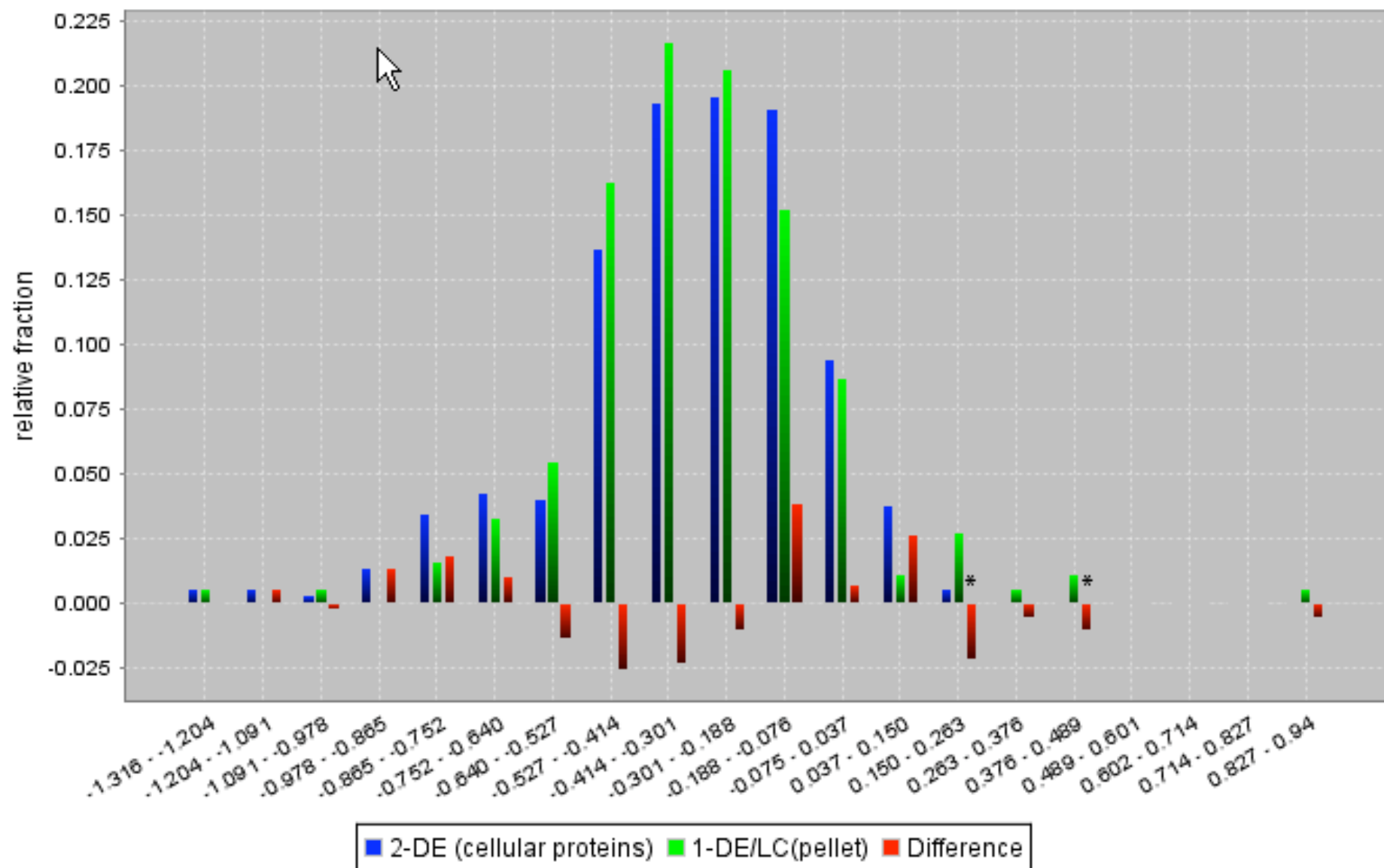
Select a graphic or data by navigation

- [-] PROMPT Results (graphics)
 - [+] pellet_vs_Hpylori_without_pellet
 - [+] pellet_vs_sephadex
 - [+] pellet_vs_supprot
 - [+] pellet_vs_startline
 - [+] pellet_vs_startline_without_pellet
 - [+] pellet+2DE+sephadex+supprot_vs_startline_without
 - [+] sephadex_vs_Hpylori_without_sephadex
 - [+] sephadex_vs_supprot
 - [+] sephadex_vs_startline
 - [+] sephadex_vs_startline_without_sephadex
 - [+] (sephadex+pellet)_vs_Hpylori_without_(sephadex+p
 - [+] supprot_vs_(pellet+sephadex)
 - [+] supprot_vs_2-DE
 - [+] supprot_vs_Hpylori_without_supprot
 - [+] (se+pe+su)_vs_Hpylori_without_(se+pe+su)
 - [+] startline_vs_sephadex_without_startline
 - [+] startline_vs_Hpylori_without_startline
 - [-] 2-DE_vs_pellet(pe)
 - [+] mass
 - [+] pi
 - [+] hydrophobicity
 - [+] Keywords
 - [+] Keywords(large)
 - [+] Features
 - [+] Features(large)
 - [+] 2-DE_vs_sephadex (se)
 - [+] 2-DE_vs_(pellet+sephadex)
 - [+] 2-DE_vs_(pellet+sephadex+supprot)
 - [+] 2-DE_vs_Hpylori_without_2-DE
 - [+] 2-DE_vs_supprot(su)
 - [+] 2-DE_vs_startline
 - [+] 2-DE_vs_startline_without_2-DE
 - [+] all_experiments(pellet+sephadex+supprot+2DE+star
 - [+] Transmembrane_vs_Hpylori_without_transmembran
- [+] Func. classification of protein sets (graphics)
- [+] Data files

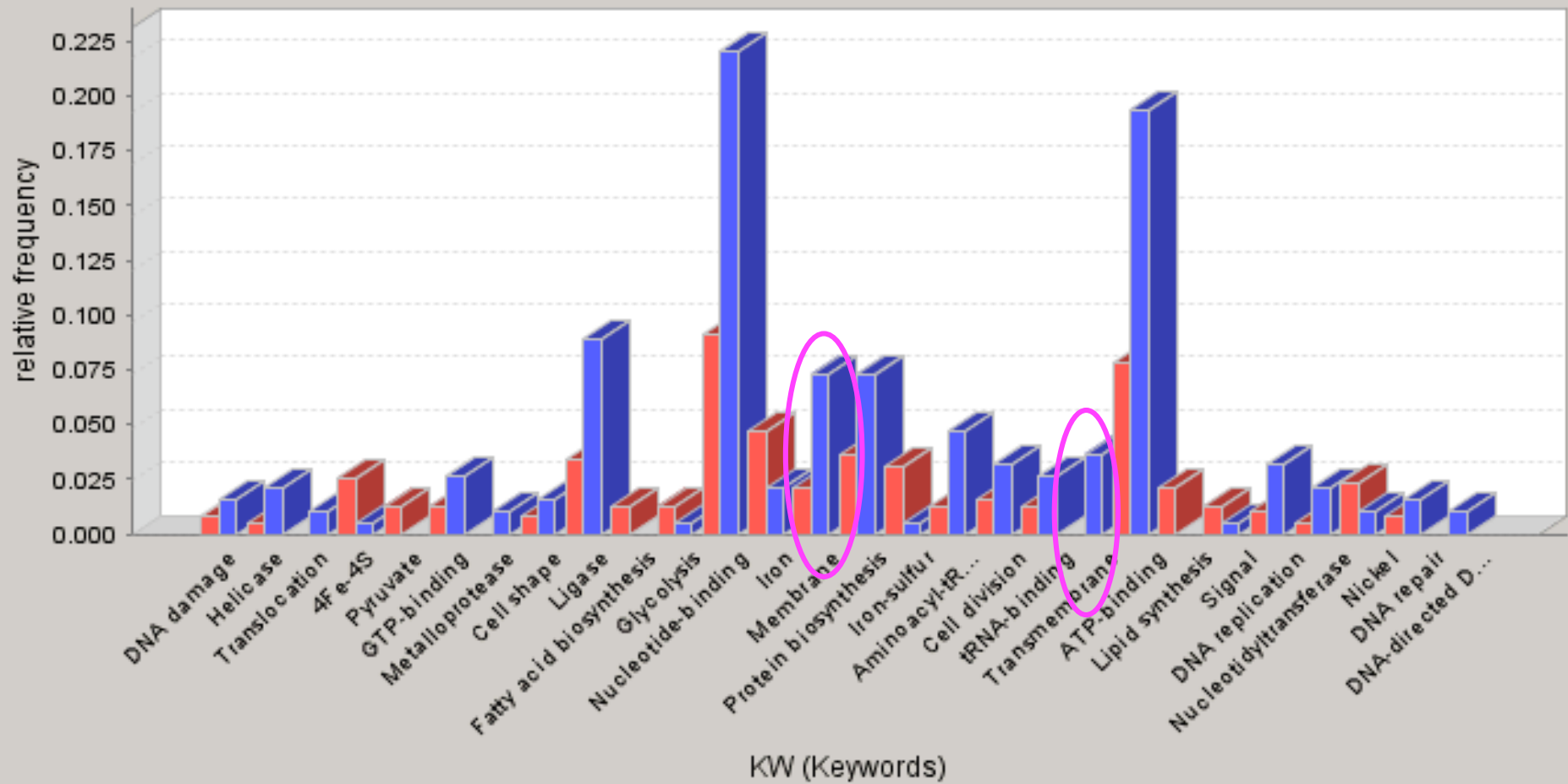
Mr



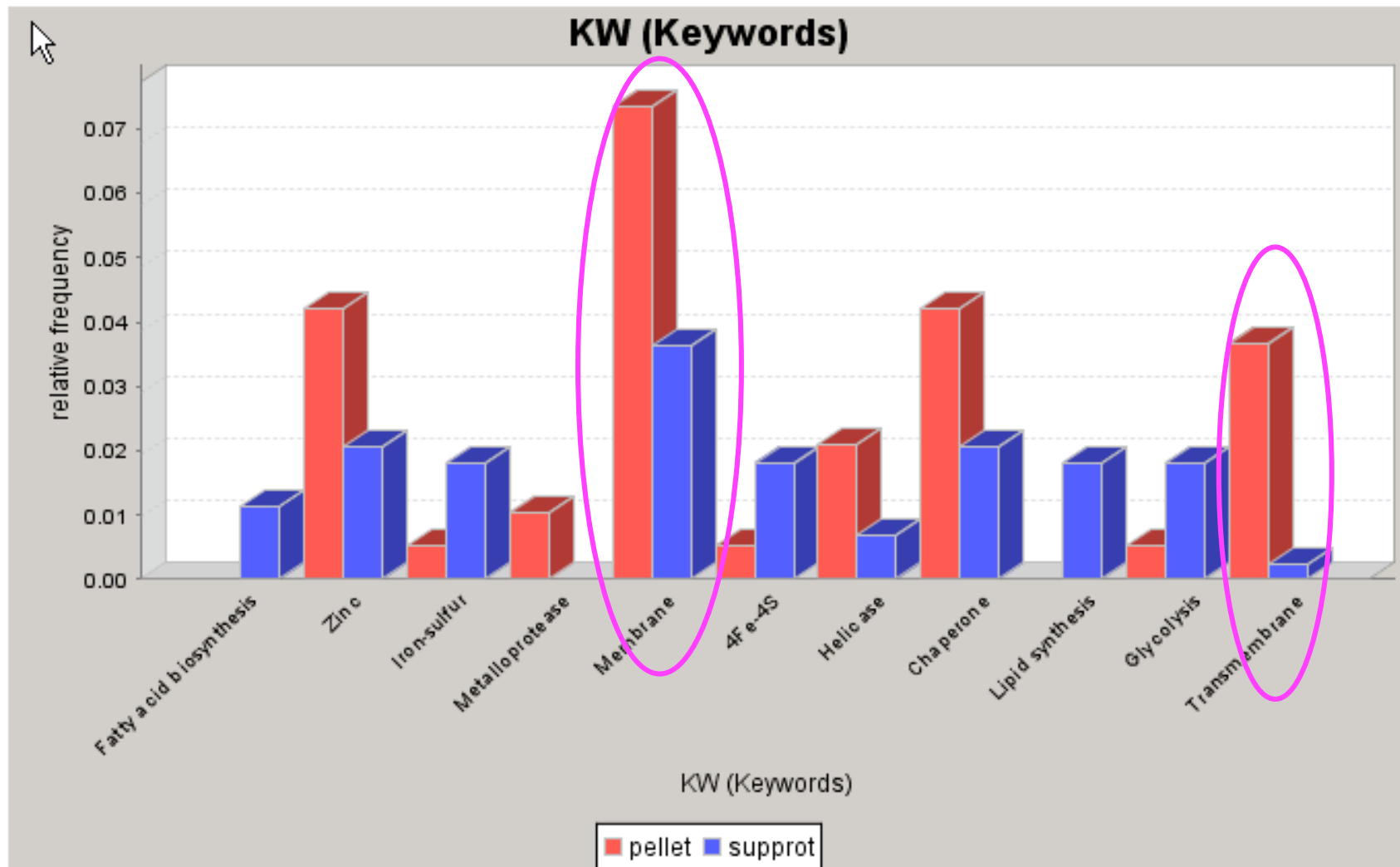
Hydrophobicity





























KW (Keywords)



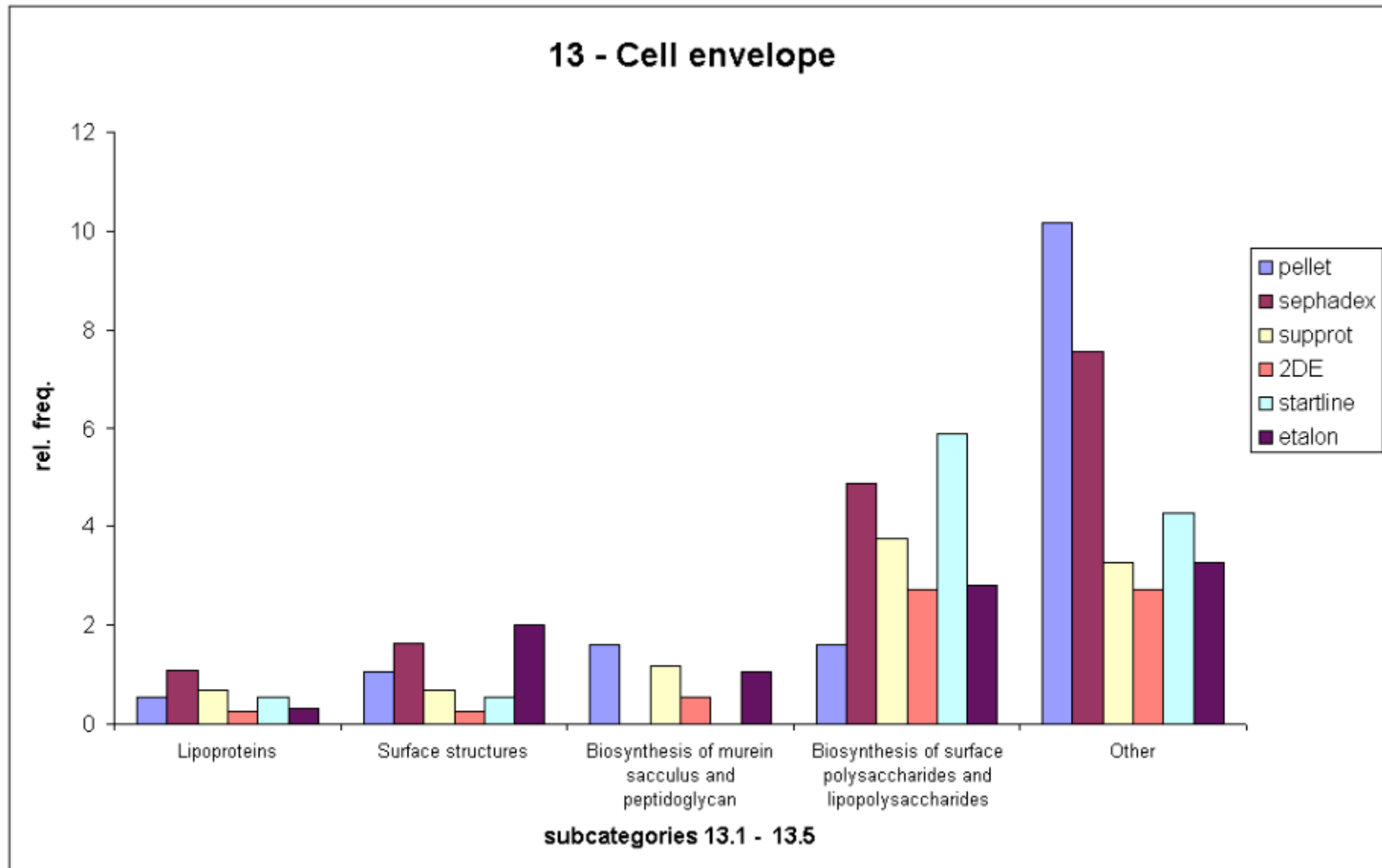
■ 2-DE (cellular proteins) ■ 1-DE/LC (pellet)



-  Select a graphic or data by navigation
-  PROMPT Results (graphics)
-  Func. classification of protein sets (graphics)
 -  Main categories (all sets)
 -  1-Amino acid biosynthesis
 -  2-Purine, pyrimidines, nucleosides
 -  3-Fatty acid, phospholipid metabolism
 -  4-Biosynthesis
 -  5-Central intermediary metabolism
 -  6-Energy metabolism
 -  7-Transport and binding proteins
 -  8-DNA metabolism
 -  9-Transcription
 -  10-Protein synthesis
 -  11-Protein fate
 -  12-Regulatory functions
 -  13-Cell envelope
 -  14-Cellular processes
 -  15-Other categories
 -  16-Unknown
 -  17-Hypothetical
 -  Main categories (all vs. HP without all)
 -  Subcategories (individual sets)
 -  Subcategories (combined sets)
 -  Combined subcategories (individual sets)
-  Data files

Functional classification

13 - Cell envelope



Select a graphic or data by navigation

- [-] PROMPT Results (graphics)
- [-] Func. classification of protein sets (graphics)
- [-] Data files
 - [+] Functional Classification Codes
 - [+] Helicobacter pylori 26955
 - [+] Helicobacter pylori 26955 (EXCEL)
 - [+] 5 experimentally determined protein sets
 - [+] Combined protein sets
 - [+] experimental & theoretical data (EXCEL)
 - [+] rel. frequencies with graphics (EXCEL)
 - [+] Data for functional classification (EXCEL)
- [-] Venn diagrams of protein sets (graphics)

- [+] Open all directories
- [+] Close all directories

#	pellet		sephadex		supprot		2DE		Startline	
	HP	UniProt-ID	HP	UniProt-ID	HP	UniProt-ID	HP	UniProt-ID	HP	UniProt-ID
1	HP0010	P42383	HP0003	P56060	HP0002	O24854	HP0002	O24854	HP1198	O25806
2	HP0011	P0A0R3	HP0010	P42383	HP0003	P56060	HP0003	P56060	HP0696	O25403
3	HP0013	O24858	HP0011	P0A0R3	HP0006	P56061	HP0009		HP0695	O25402
4	HP0014	O24859	HP0014	O24859	HP0010	P42383	HP0010	P42383	HP1195	P56002
5	HP0025	O24870	HP0025	O24870	HP0011	P0A0R3	HP0011	P0A0R3	HP1266	O25856
6	HP0056	O24897	HP0026	P56062	HP0014	O24859	HP0013	O24858	HP1512	O26042
7	HP0068	Q09066	HP0027	P56063	HP0019	O24864	HP0014	O24859	HP0680	P55982
8	HP0082	O24911	HP0044	O24885	HP0025	O24870	HP0025	O24870	HP0779	P56418
9	HP0083	P66637	HP0045	O24886	HP0026	P56062	HP0026	P56062	HP0264	P71404
10	HP0086	O24913	HP0056	O24897	HP0027	P56063	HP0027	P56063	HP1422	P56456
11	HP0088	P55993	HP0069	Q09065	HP0043	O24884	HP0030	O24873	HP0786	O25475
12	HP0097	O24923	HP0072	P69996	HP0045	O24886	HP0049	O24890	HP0109	P55994
13	HP0098	O24924	HP0073	P14916	HP0047	O24888	HP0056	O24897	HP0392	O25153
14	HP0099	O24925	HP0083	P66637	HP0056	O24897	HP0067	Q09067	HP0025	O24870
15	HP0100	O24926	HP0086	O24913	HP0060	O24900	HP0068	Q09066	HP1241	P56452
16	HP0103	O24929	HP0097	O24923	HP0067	Q09067	HP0069	Q09065	HP1205	P56003
17	HP0121	P56070	HP0098	O24924	HP0068	Q09066	HP0070	Q09064	HP1045	O25686
18	HP0145	O24956	HP0106	P56069	HP0069	Q09065	HP0072	P69996	HP1177	O25791
19	HP0153	P42445	HP0109	P55994	HP0072	P69996	HP0073	P14916	HP0056	O24897
20	HP0163	P56074	HP0121	P56070	HP0073	P14916	HP0075	P25177	HP0402	P56145
21	HP0166	O24973	HP0123	P56071	HP0075	P25177	HP0083	P66637	HP0547	P55980
22	HP0182	P56126	HP0127	O24941	HP0083	P66637	HP0084	P56038	HP0501	P55992
23	HP0193	O06912	HP0137	O24949	HP0086	O24913	HP0086	O24913	HP1379	P55995
24	HP0197	P56460	HP0153	P42445	HP0088	P55993	HP0090	O24916	HP0121	P56070
25	HP0210	P56116	HP0154	P48285	HP0090	O24916	HP0096	O24922	HP1430	P56185
26	HP0213	P56138	HP0166	O24973	HP0096	O24922	HP0104	O24930	HP0919	O25577
27	HP0220	O25008	HP0170	O24976	HP0097	O24923	HP0105	O24931	HP0317	O25086
28	HP0227	O34523	HP0175	P56112	HP0098	O24924	HP0106	P56069	HP0396	O25157
29	HP0229	O25015	HP0176	P56109	HP0099	O24925	HP0107	P56067	HP0192	O06913
30	HP0233	O25019	HP0182	P56126	HP0104	O24930	HP0109	P55994	HP0632	O25349
31	HP0243	P43313	HP0183	P56089	HP0106	P56069	HP0110	P55970	HP1213	O25812
32	HP0247	O25029	HP0191	O06914	HP0107	P56067	HP0116	P55991	HP1547	P56457
33	HP0248	O25030	HP0192	O06913	HP0109	P55994	HP0123	P56071	HP0322	O25089
34	HP0250	O25032	HP0201	O24993	HP0110	P55970	HP0129	O24943	HP0480	O25225
35	HP0252	O25034	HP0207	O24999	HP0112	O24934	HP0130	O24944	HP0210	P56116
36	HP0254	O25036	HP0210	P56116	HP0121	P56070	HP0134	O24947	HP1460	P56157

Detected ORFs with transmembrane regions

Supprot HP	Transmembrane regions (Pedant)	Pellet HP	Transmembrane regions (Pedant)	Pellet HP	Transmembrane regions (Pedant)	2-DE HP	Transmembrane regions (Pedant)
HP0097	1	HP0082	2	HP0791	6	HP0160	1
HP0099	2	HP0097	1	HP0920	7	HP0410	1
HP0160	1	HP0099	2	HP0977	1	HP0490	2
HP0248	1	HP0103	2	HP0983	3	HP0506	1
HP0410	1	HP0145	1	HP1069	2	HP0977*	1
HP0473	1	HP0193	5	HP1088	1	HP1038	1
HP0490	2	HP0248	1	HP1136	2	HP1118	1
HP0506	1	HP0284	3	HP1349	3	HP1203	1
HP0887	1	HP0351	2	HP1450	6	HP1540	1
HP0977	1	HP0358	1	HP1479	1		
HP1118	1	HP0597	1	HP1550	5		
HP1488	1	HP0607	12	HP1556	1		

Total number of identified proteins:

Supprot: 413

Pellet: 179

2-DE: 356

Absolute and relative numbers of ORFs with transmembrane regions

Fraction	ORFs	ORFs with keyword transmembrane in Uniprot ¹		ORFs with transmembrane regions in Pedant ²	
		No	%	No	%
	Total No				
2-DE	355	0	0	9	2.5
Supprot	413	1	0.2	12	2.9
Pellet	179	7	3.9	24	13.4
Sephadex	176	1	0.6	5	2.8
Startline	175	0	0	0	0

¹<http://www.ebi.ac.uk/ebisearch/search.ebi?t=Helicobacter+pylori+26695&db=proteinSequences&refine=transmembrane>.

This database contains 103 proteins with at least 1 transmembrane region in *H. pylori* 26695.

²http://pedant.gsf.de:3345/pedant3htmlview/pedant3view?Method=start_method&Db=p3_p233_Hel_pylor.

This database contains 331 proteins with at least 1 transmembrane region in *H. pylori* 26695.

ORF No	Transmembrane Region* No	2-DE	Supprot	Pellet	Sephadex	Startline
HP0160	1	x	x	-	-	-
HP0410	1	x	x	-	x	-
HP0490	2	x	x	-	-	-
HP0506	1	x	x	-	-	-
HP0977	1	x	x	x	-	-
HP1038	1	x	-	-	-	-
HP1118	1	x	x	-	x	-
HP1203	1	x	-	-	-	-
HP1540	1	x	-	-	-	-
HP0473	1	-	x	-	-	-
HP0887	1	-	x	-	x	-
HP1488	1	-	x	-	x	-
HP0082	2	-	-	x	-	-
HP0097	1	-	x	x	x	-
HP0099	2	-	x	x	-	-
HP0103	2	-	-	x	-	-
HP0145	1	-	-	x	-	-
HP0193	5	-	-	x	-	-
HP0248	1	-	x	x	-	-
HP0284	3	-	-	x	-	-

Examples of questions for the database

- With which technology can a special protein class best be analysed?**
- With which technology can a special protein or even protein species optimally be analysed?**
- With which combinations of technologies proteins of energy metabolism can optimally be analysed ?**
- Check of the gene prediction: Are the unknown proteins by all of the separation techniques underrepresented?**
- Which protein class is overrepresented in the startline or within the Sephadex layer?**
- With which method most of the low Mr proteins are accessible?**
- With which method basic proteins are optimally accessible?**

Conclusion

- **Hypothesis confirmed: Membrane proteins are enriched in the urea-insoluble protein fraction (pellet-fraction).**
- **LC does not per se have an advantage to resolve membrane proteins as compared with 2-DE (supprot has the same low relative number of membrane proteins as 2-DE). The 100000g centrifugation step precipitates and therefore removes the membrane proteins out of the sample.**
- **After opening the data in the www, the scientific community will be able to do with the accessible data their own data mining at three different levels:**
 1. **Original data: 2-DE gels, MS spectra and peak lists**
 2. **Identification lists**
 3. **Prompt and functional classification data**



**Franziska Schiele
Alexander Krah
Wolfgang Höhenwarter
Frank Schmidt
Ursula Zimny-Arndt
Monika Schmid
Renate Ackermann
Hajar Habibi
Anja Freiwald
Robert Stein
Klaus-P. Pleissner
Thomas F. Meyer**

**European Union EBP (QLK2CT200001536) and XTB, SFB577
Bundesministerium für Bildung und Forschung, Germany (031U107A/207A)**



MAX-PLANCK-GESellschaft