

Bioinformatics Challenges for Protein Identification from Tandem Mass Spectra

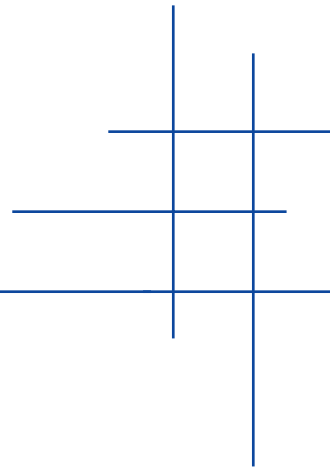
Nathan Edwards



CELERA

AB Applied
Biosystems

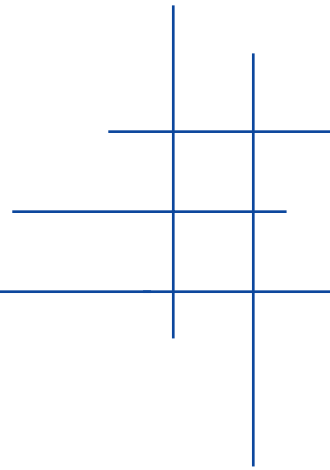
Proteomics is...



...the high-throughput analysis of the expressed protein complement of a cell.



Proteomics is...

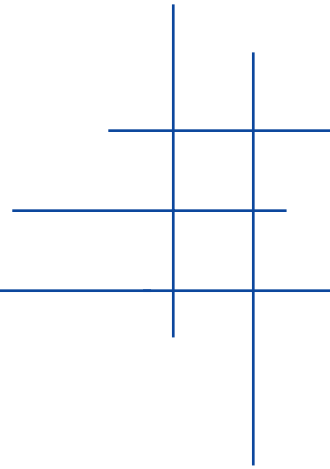


...one of the pillars of **systems biology**

- **Genotyping**
 - SNP arrays
 - (RT-)PCR, Multiplexed PCR
 - Resequencing
- **Gene expression**
 - Expression Arrays
 - RT-PCR.



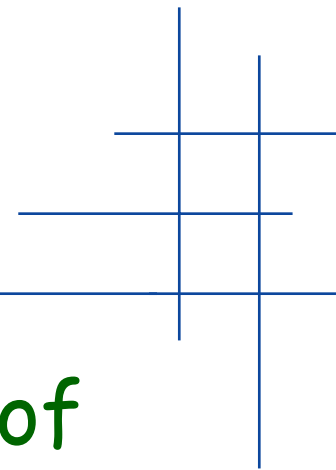
Proteomics is...



	Gene / Transcript / Protein	
Measurement	Predetermined	Unknown
Discrete	Genotyping	
Continuous	Gene Expression	Proteomics



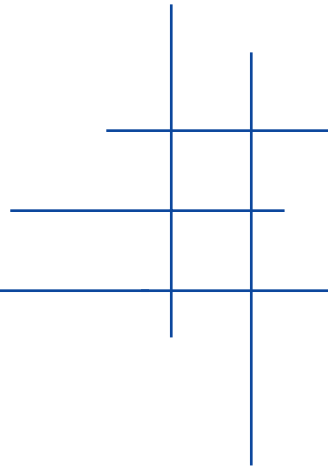
Celera Proteomics



The high-throughput identification of
diagnostics markers &
therapeutics targets
by
differential protein abundance
in
healthy and diseased samples.



Celera Proteomics

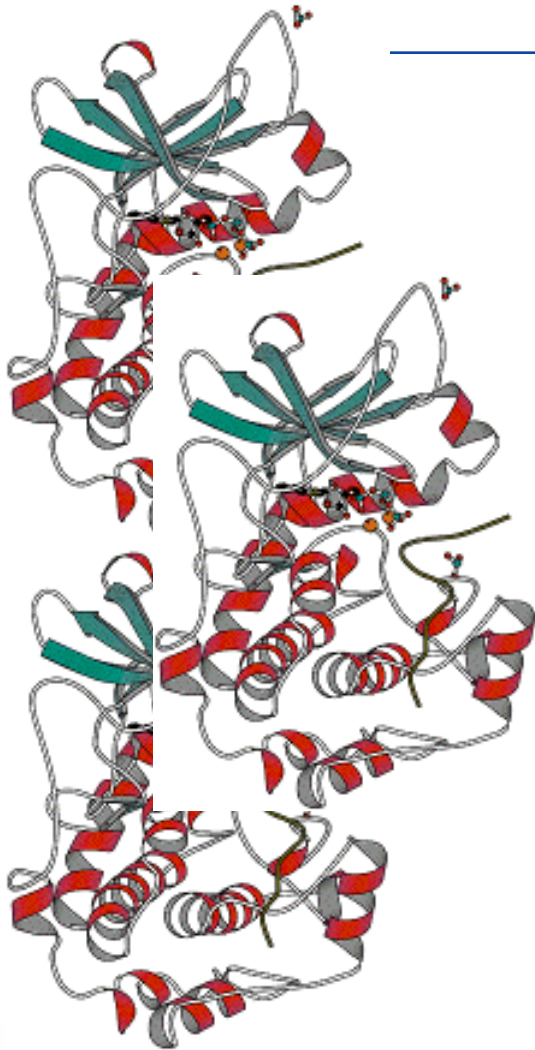


Key technologies

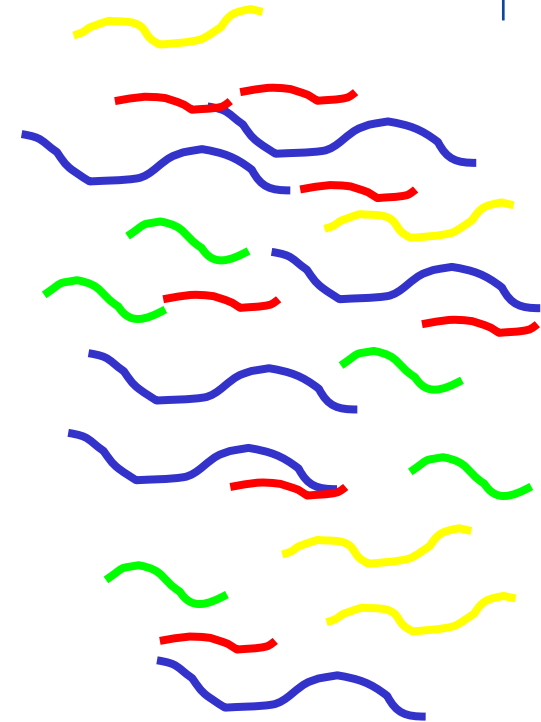
- Cell biology
- Liquid chromatography (LC)
- Mass spectrometry (MS)
- Isotope-coded affinity tags (ICAT)
- Tandem Mass Spectrometry (MS/MS)
- Automation, Pipelines, Data-Volume
- Bioinformatics



LC/MS for Protein Expression



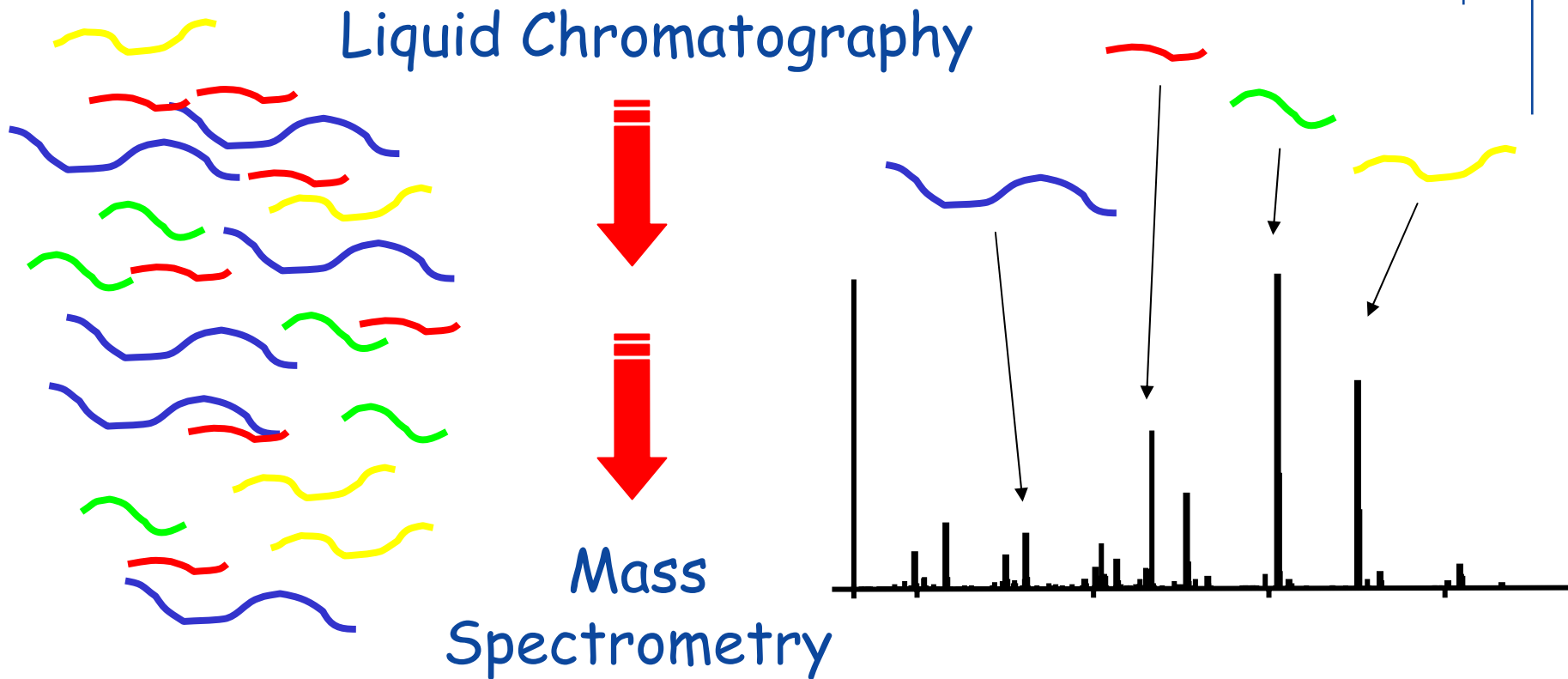
Enzymatic Digest
and
Fractionation



CELERA

AB Applied
Biosystems

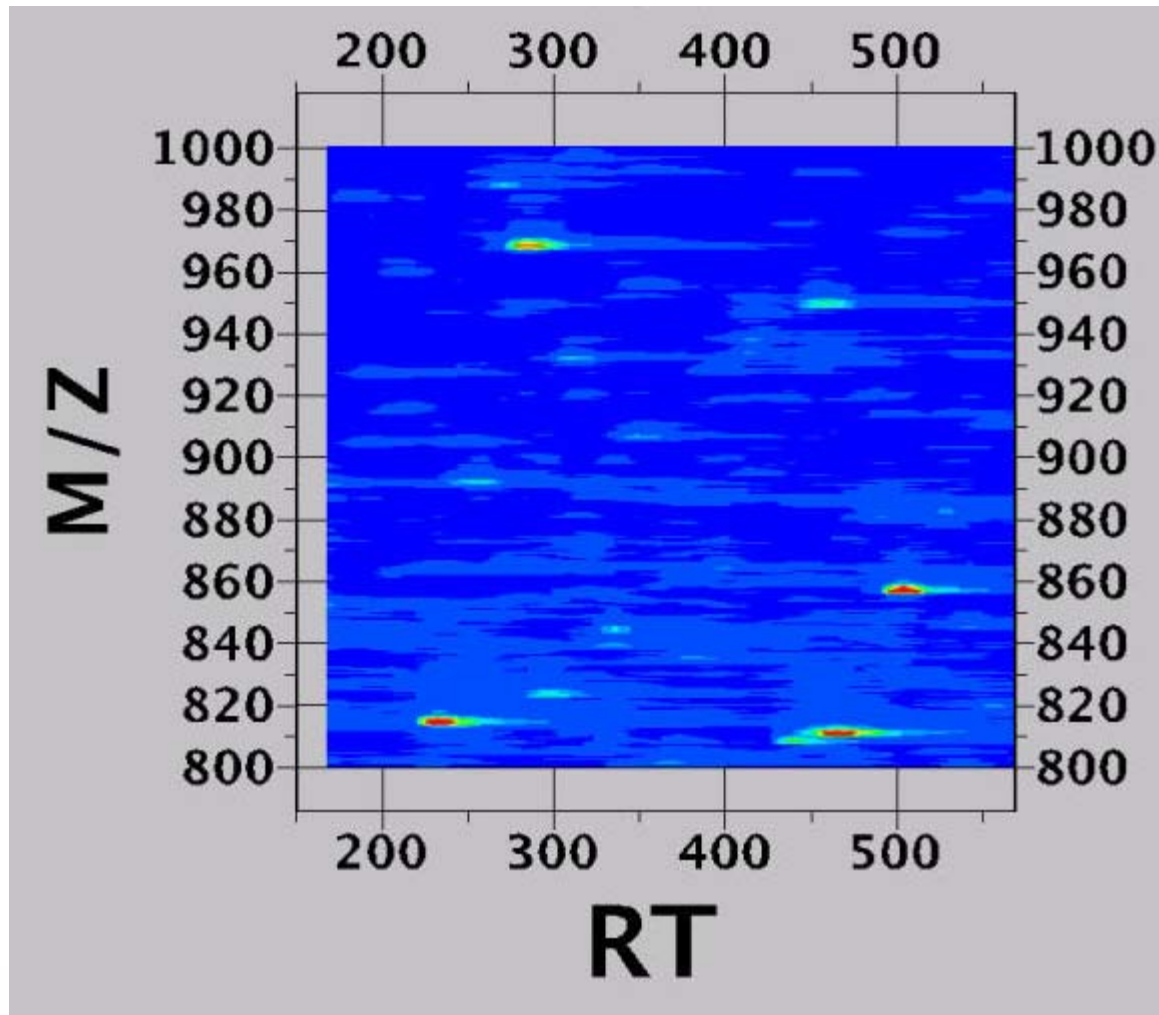
LC/MS for Protein Expression



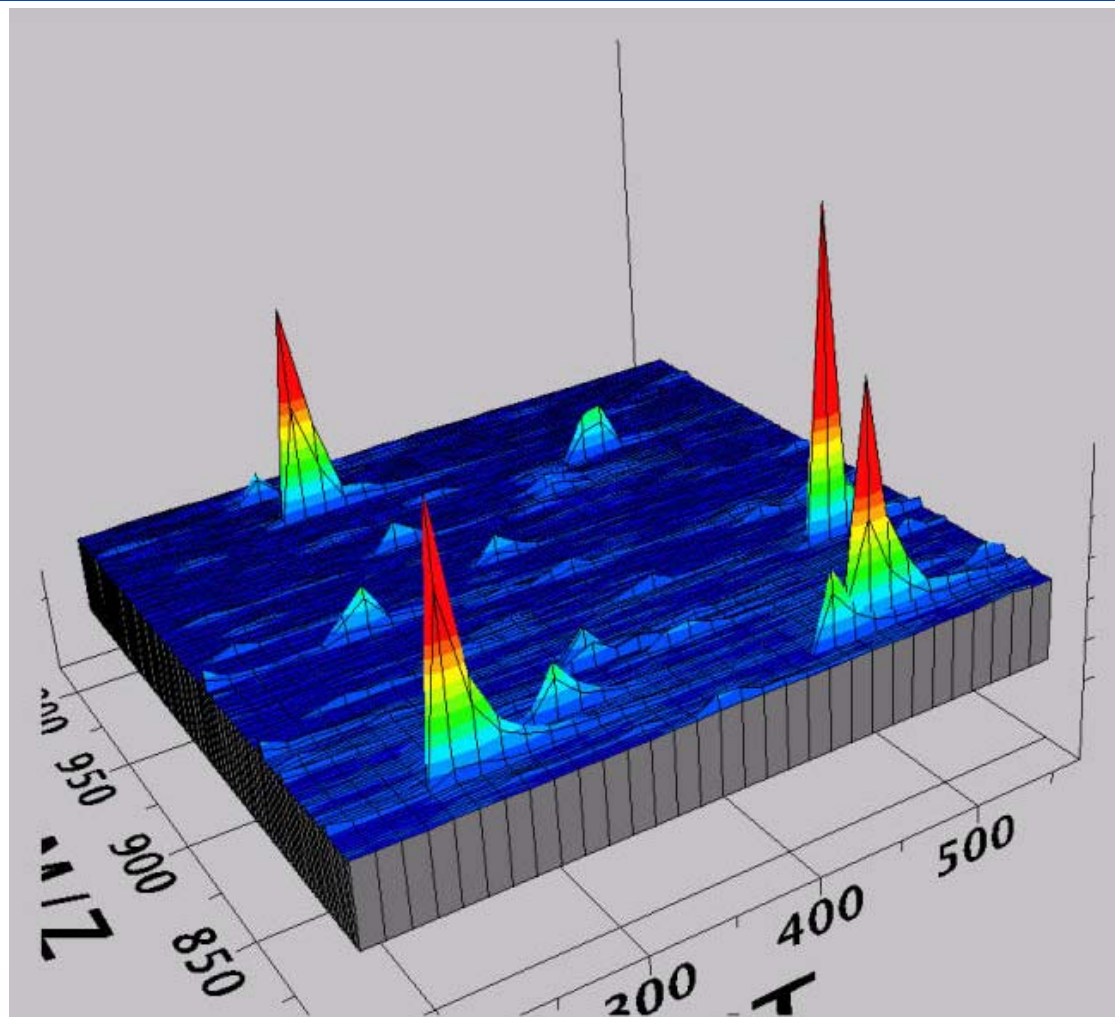
LC/MS: 1 MS spectrum every 1-2 seconds



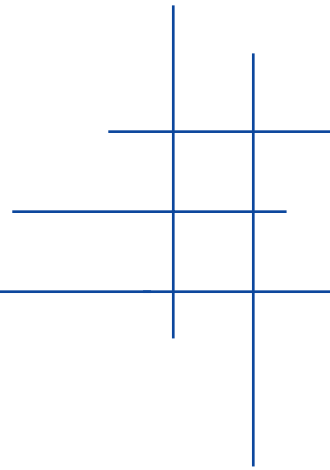
LC/MS for Protein Expression



LC/MS for Protein Expression



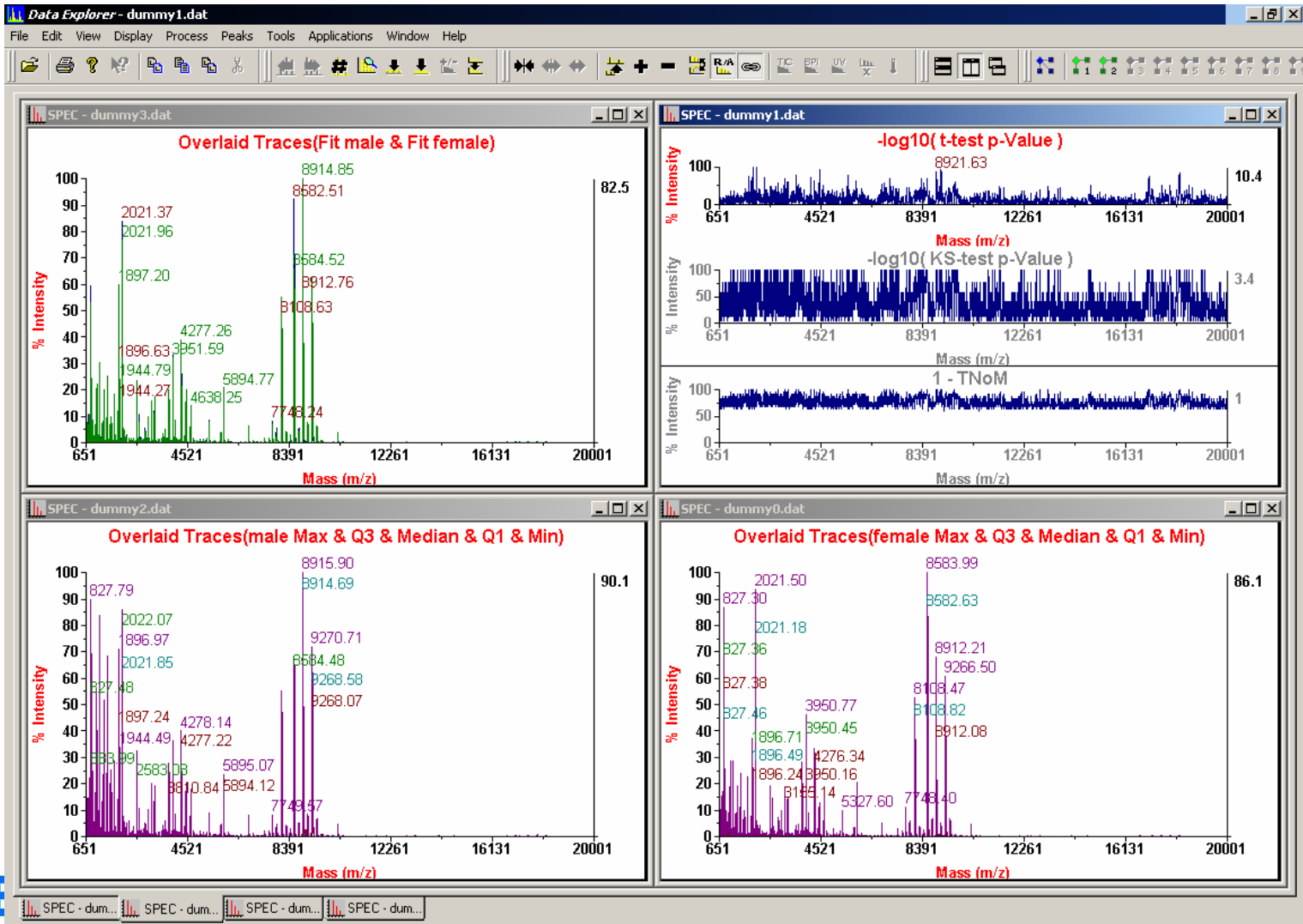
NCI-FDA Proteomics



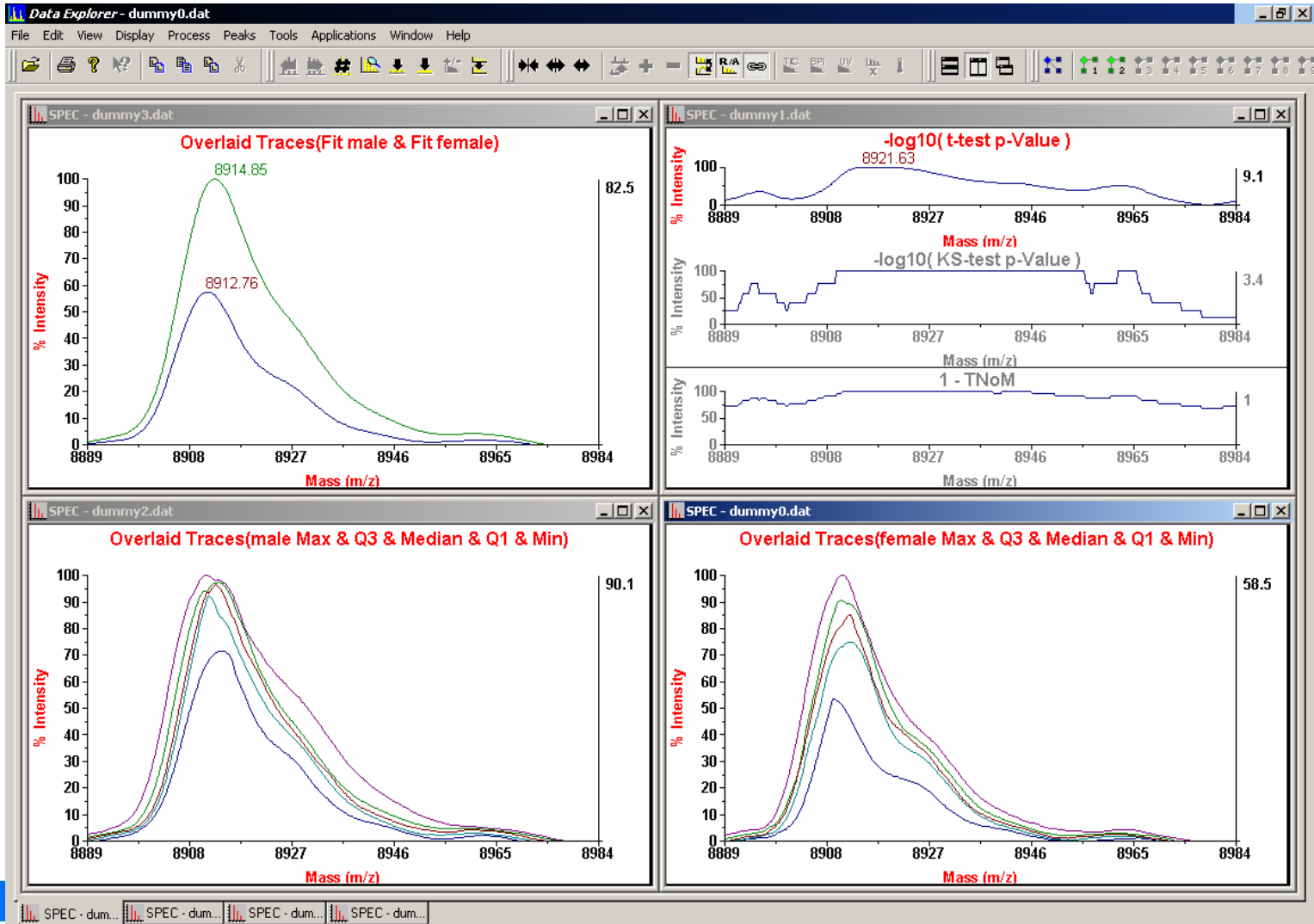
The identification of
diagnostic patterns
by
statistical data-mining
of spectra from
healthy and diseased
samples.



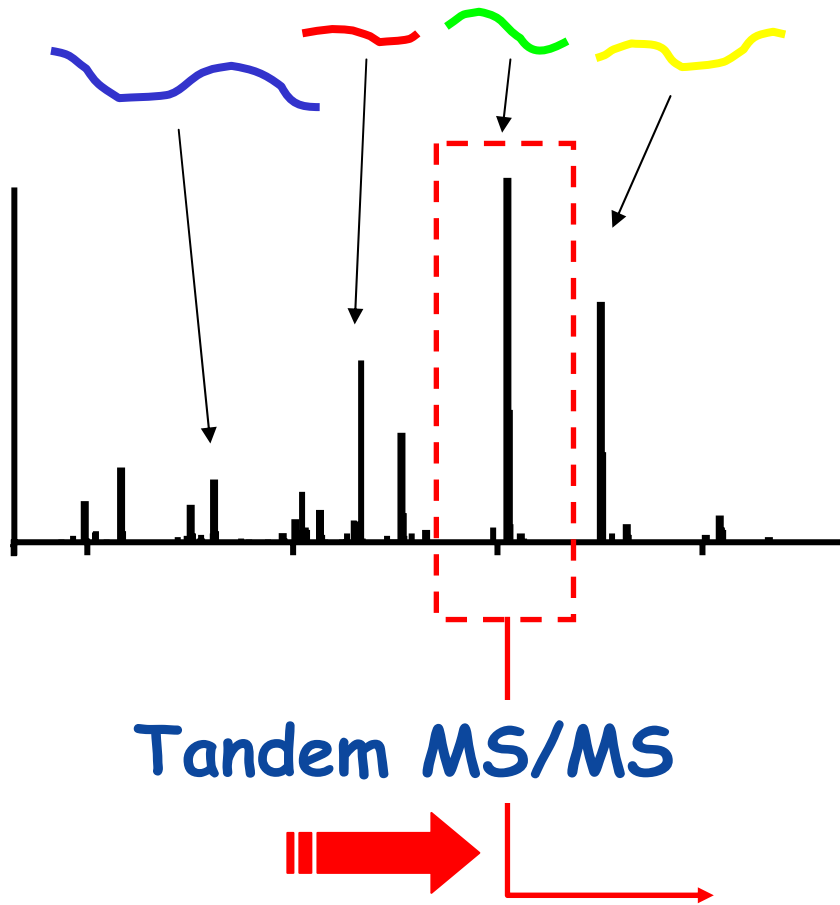
Protein Profiling Toolbox



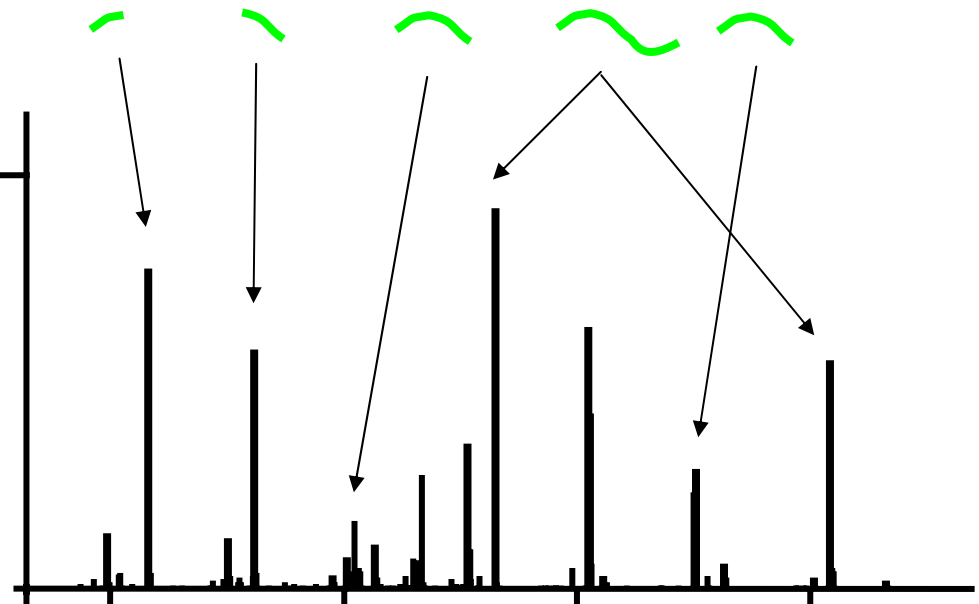
Protein Profiling Toolbox



LC/MS/MS for Protein Id



LC/MS/MS: 1 MS spectrum followed by 2-5 Tandem MS/MS spectra every 5-10 sec.



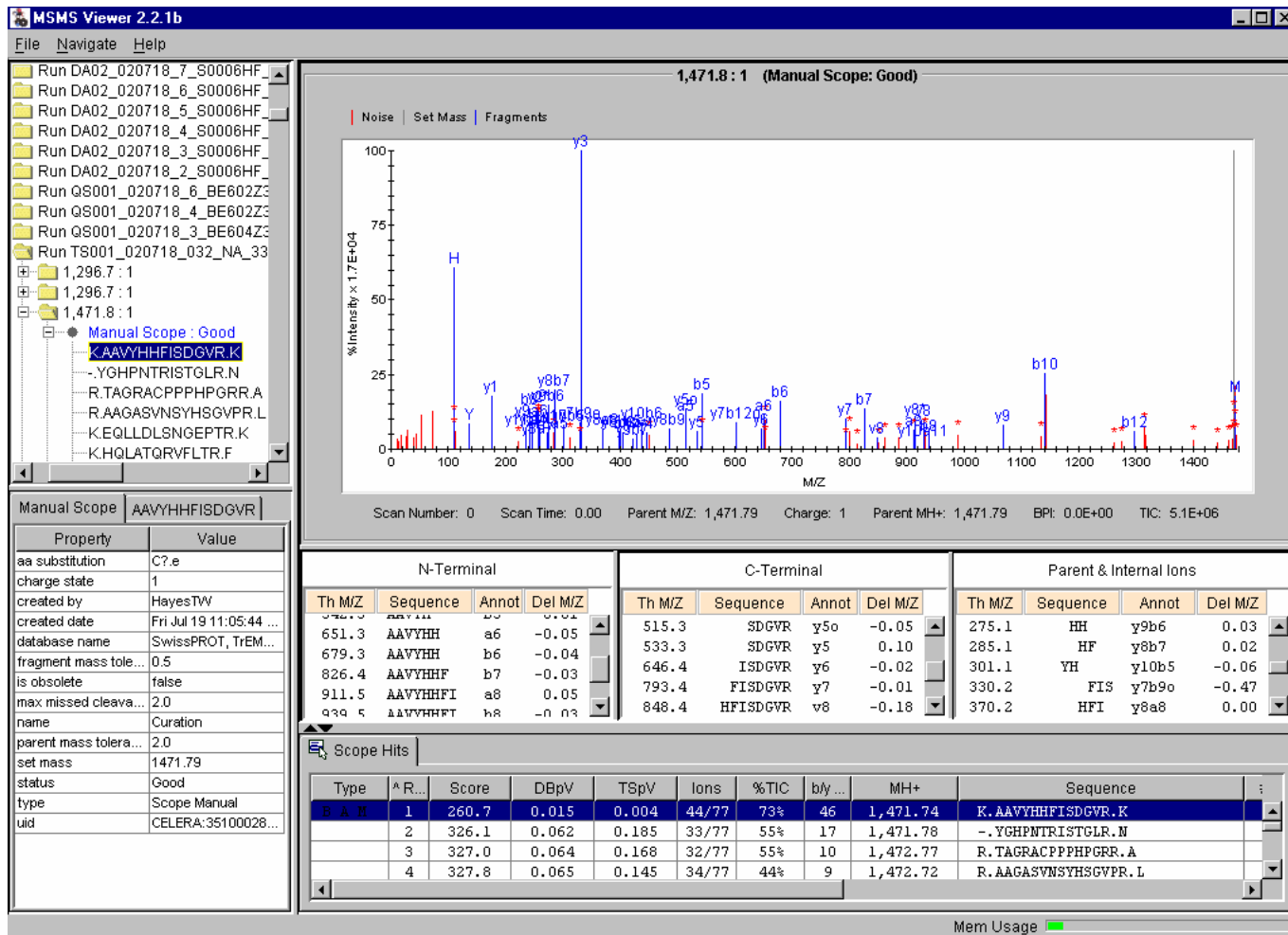
LC/MS/MS for Protein Id



- 1 experiment produces 1000's of MS/MS spectra
- Suitable for complex mixtures
- 100's of proteins identified from a single experiment
- **High-throughput protein identification!**



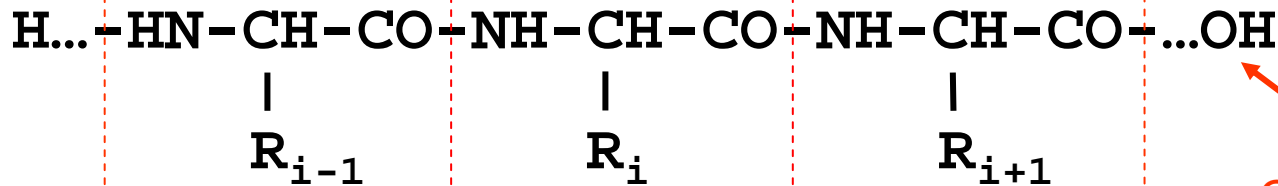
Tandem Mass Spectrometry



Peptide Fragmentation

Peptides consist of amino-acids arranged in a linear backbone.

N-terminus



AA residue_{*i-1*}

AA residue_{*i*}

AA residue_{*i+1*}

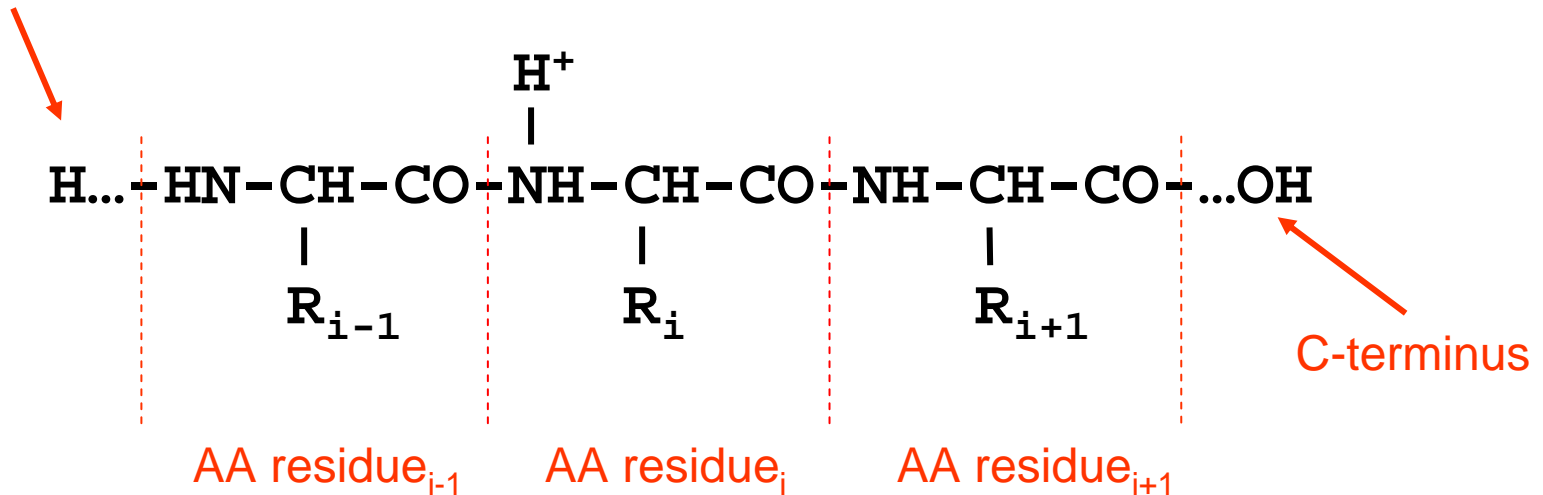
C-terminus



Peptide Fragmentation

Peptides consist of amino-acids arranged in a linear backbone.

N-terminus



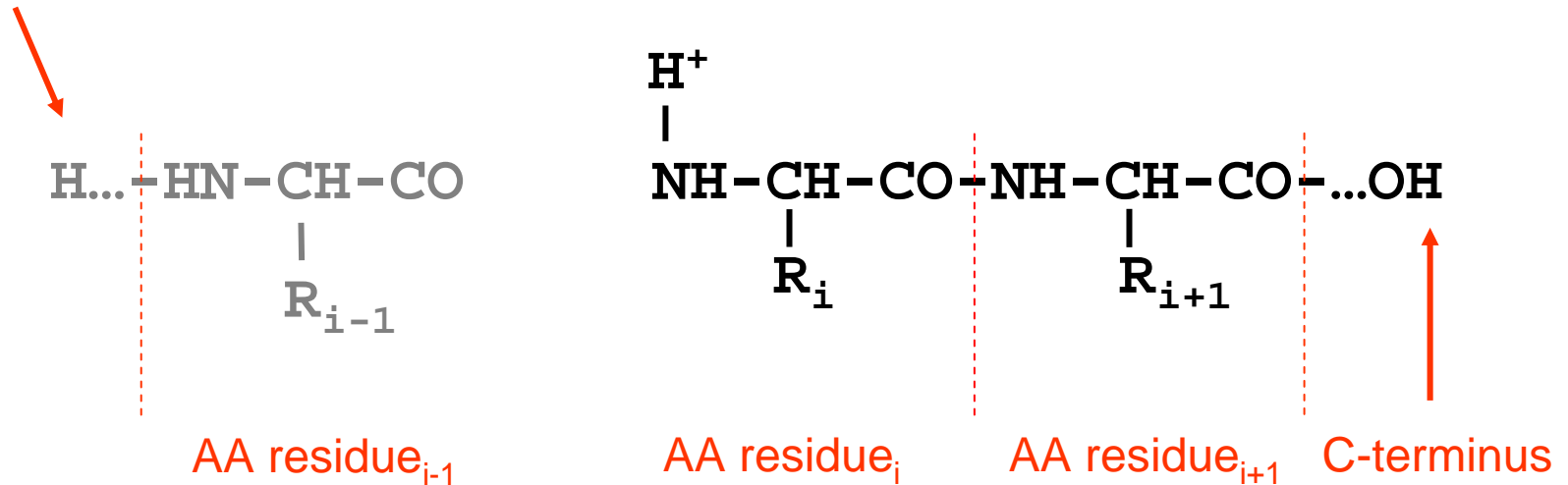
Ionized peptide (addition of a proton)



Peptide Fragmentation

Peptides consist of amino-acids arranged in a linear backbone.

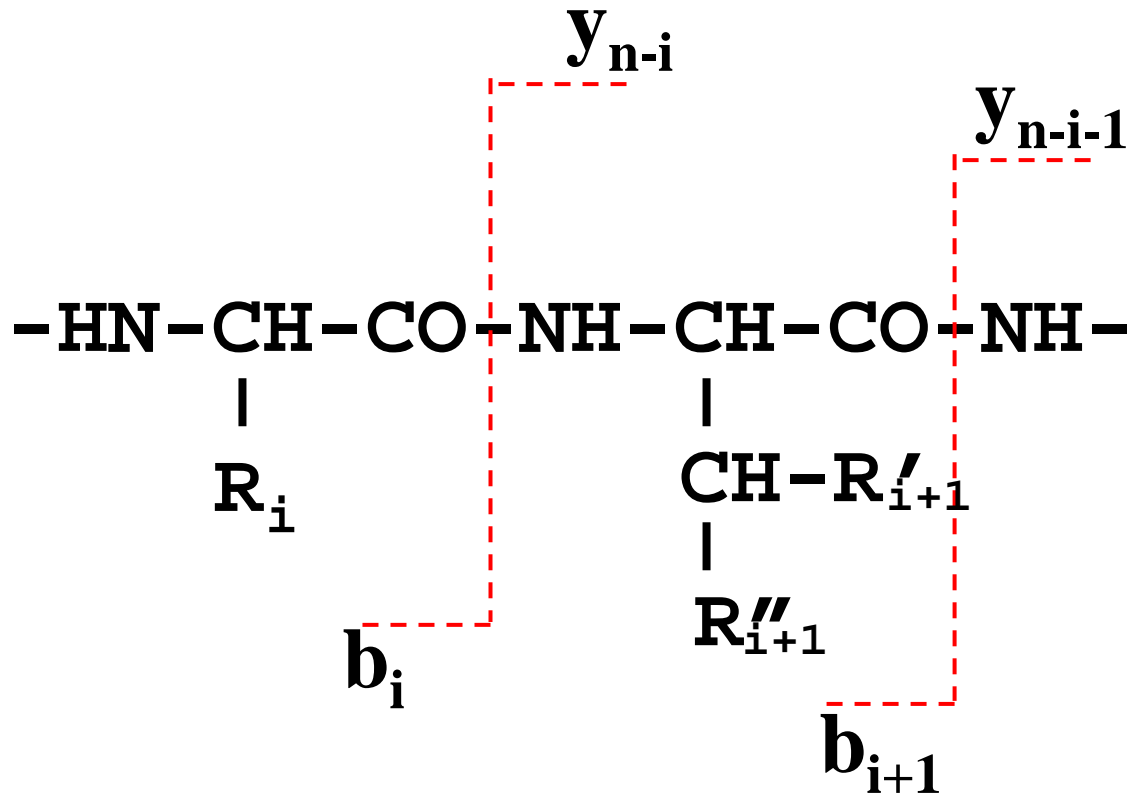
N-terminus



Fragmented peptide
C-terminus fragment observed



Peptide Fragmentation



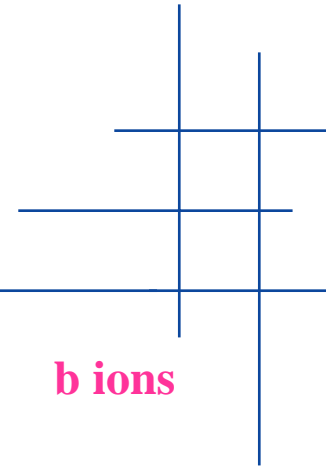
Peptide Fragmentation

Peptide: S-G-F-L-E-E-D-E-L-K

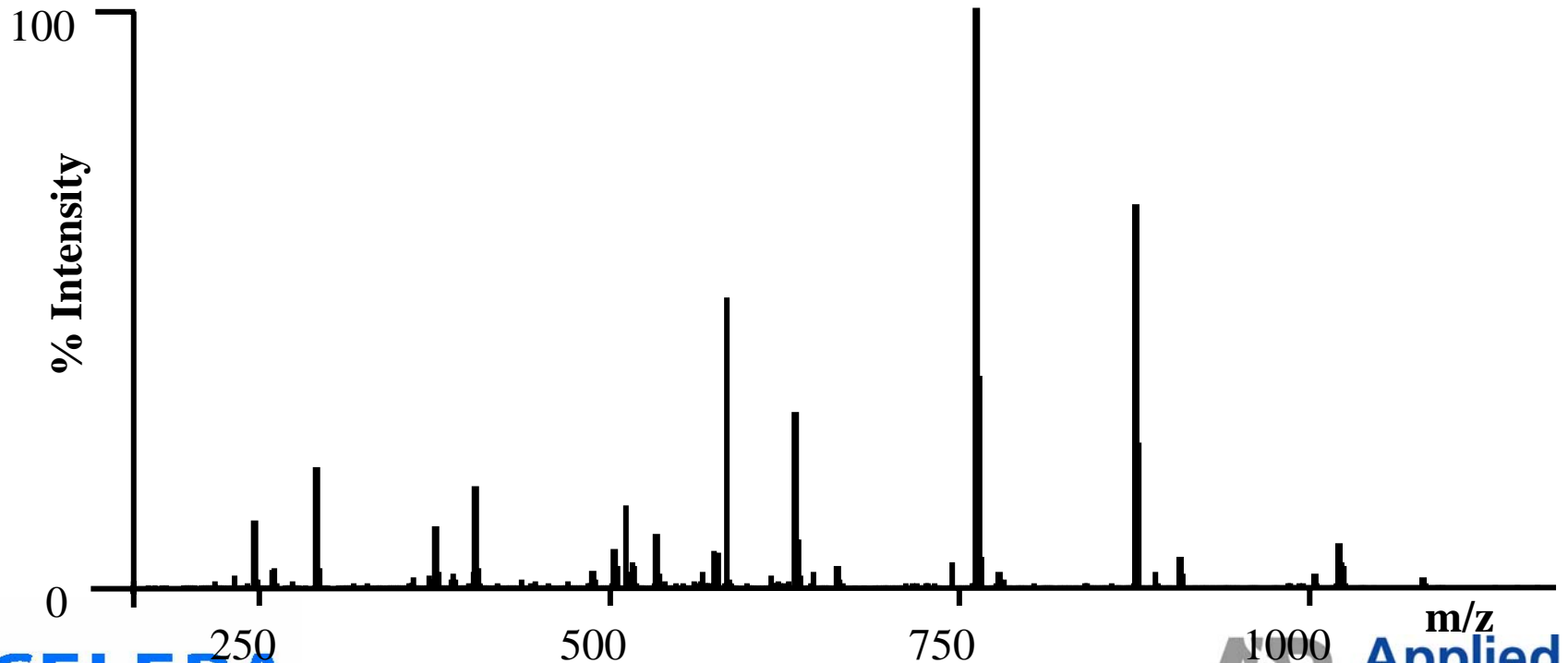
MW	ion			ion	MW
88	b ₁	S	GFLEEDELK	y ₉	1080
145	b ₂	SG	FLEEDELK	y ₈	1022
292	b ₃	SGF	LEEDELK	y ₇	875
405	b ₄	SGFL	EEDELK	y ₆	762
534	b ₅	SGFLE	EDELK	y ₅	633
663	b ₆	SGFLEE	DELK	y ₄	504
778	b ₇	SGFLEED	ELK	y ₃	389
907	b ₈	SGFLEEDE	LK	y ₂	260
1020	b ₉	SGFLEEDEL	K	y ₁	147



Peptide Fragmentation

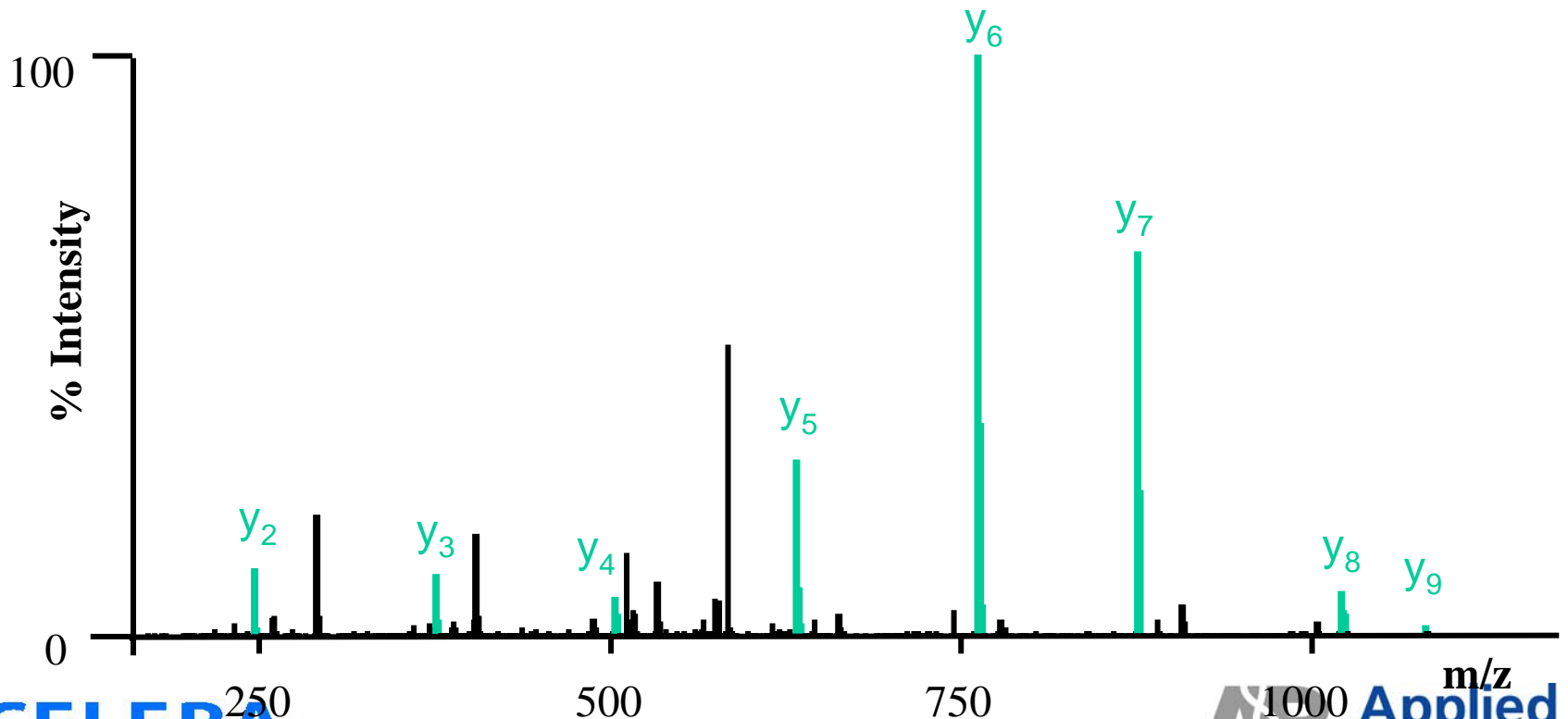


<u>88</u>	<u>145</u>	<u>292</u>	<u>405</u>	<u>534</u>	<u>663</u>	<u>778</u>	<u>907</u>	<u>1020</u>	1166	b ions
S	G	F	L	E	E	D	E	L	K	
1166	<u>1080</u>	<u>1022</u>	<u>875</u>	<u>762</u>	<u>633</u>	<u>504</u>	<u>389</u>	<u>260</u>	147	y ions



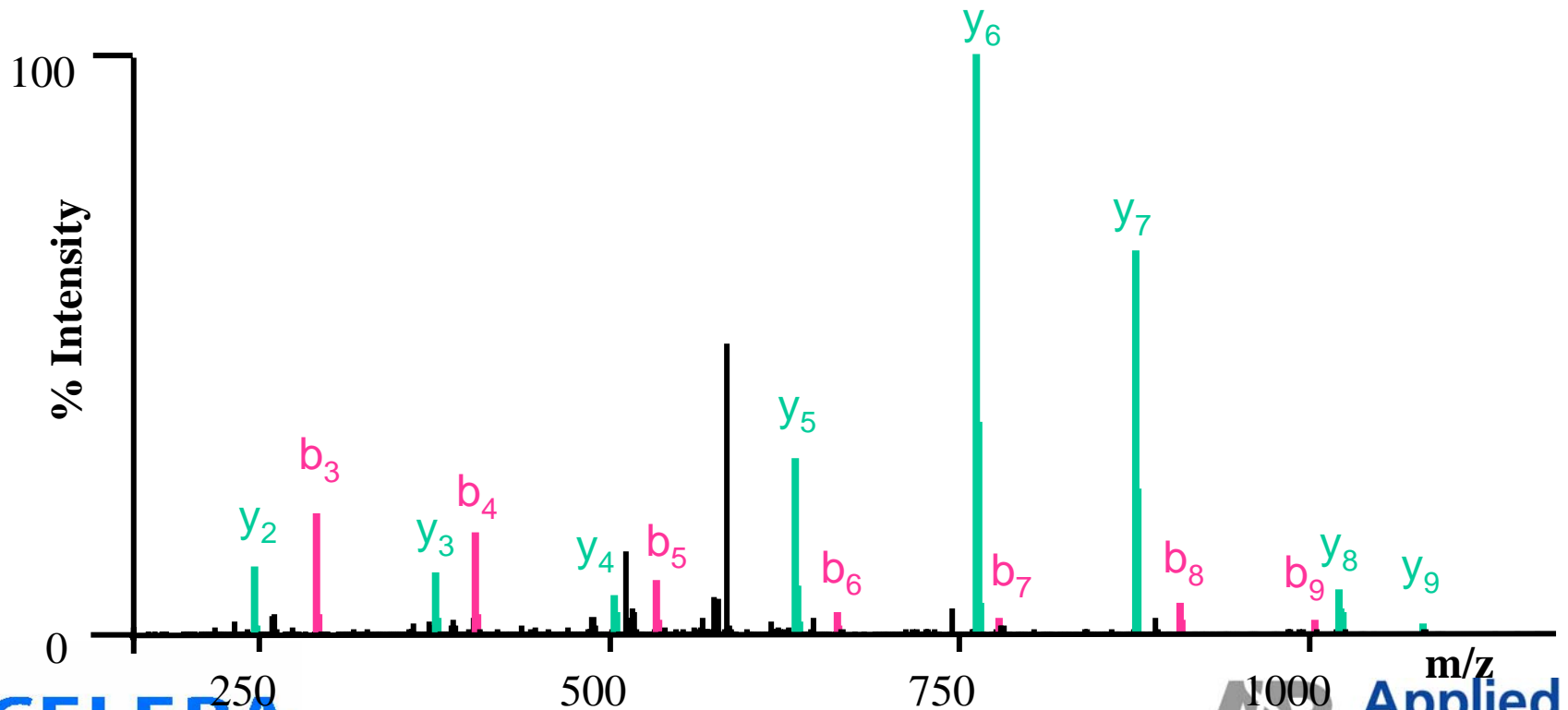
Peptide Fragmentation

<u>88</u>	<u>145</u>	<u>292</u>	<u>405</u>	<u>534</u>	<u>663</u>	<u>778</u>	<u>907</u>	<u>1020</u>	1166	b ions
S	G	F	L	E	E	D	E	L	K	
1166	<u>1080</u>	<u>1022</u>	<u>875</u>	<u>762</u>	<u>633</u>	<u>504</u>	<u>389</u>	<u>260</u>	147	y ions

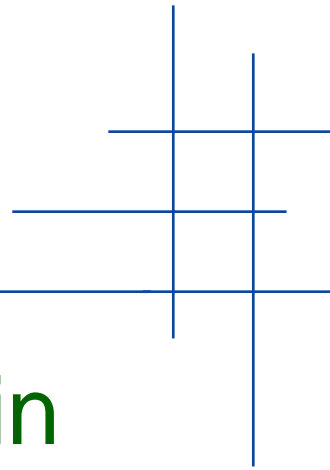


Peptide Fragmentation

<u>88</u>	<u>145</u>	<u>292</u>	<u>405</u>	<u>534</u>	<u>663</u>	<u>778</u>	<u>907</u>	<u>1020</u>	<u>1166</u>	b ions
S	G	F	L	E	E	D	E	L	K	
<u>1166</u>	<u>1080</u>	<u>1022</u>	<u>875</u>	<u>762</u>	<u>633</u>	<u>504</u>	<u>389</u>	<u>260</u>	<u>147</u>	y ions



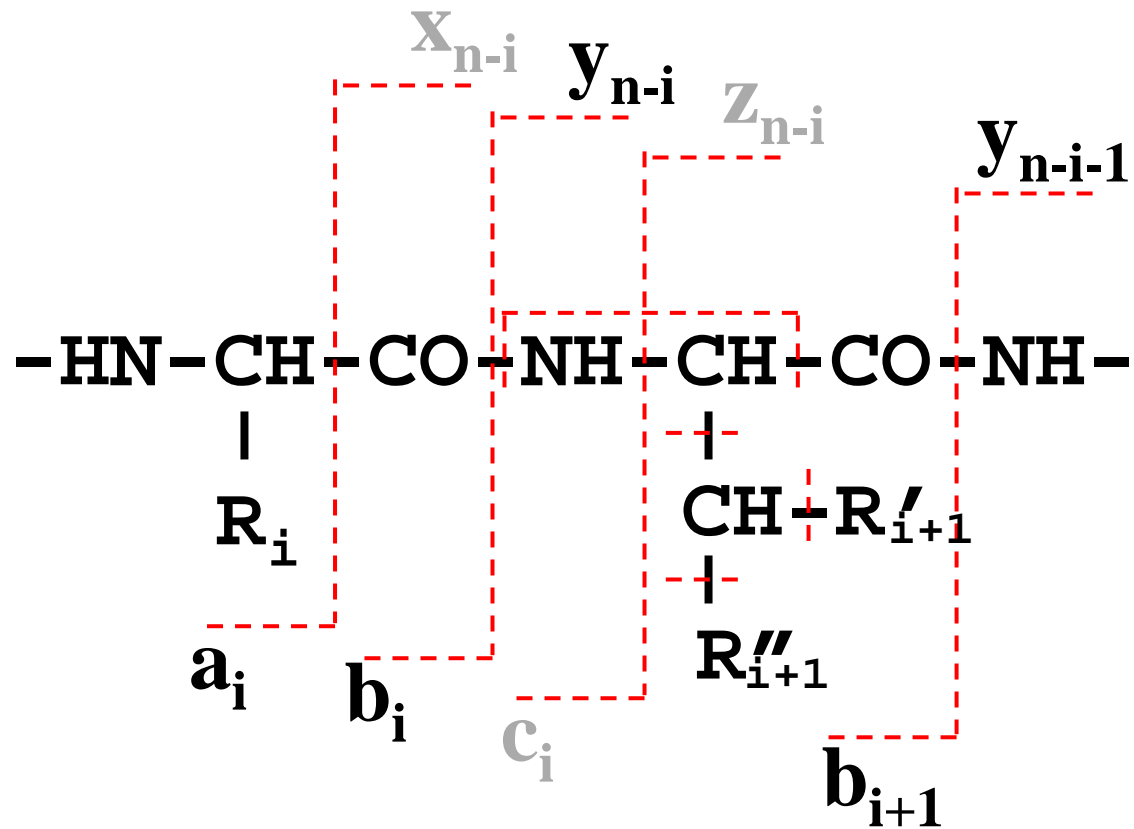
Sequence Database Search



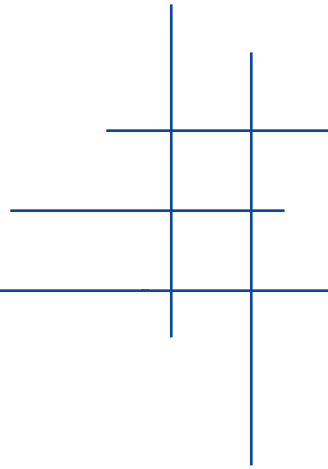
- Compares peptides from a protein sequence database with spectra
- Filter peptide candidates by
 - Parent mass
 - Digest motif
- Score each peptide against spectrum
 - Generate all possible peptide fragments
 - Match putative fragments with peaks
 - Score and rank



Peptide Fragmentation



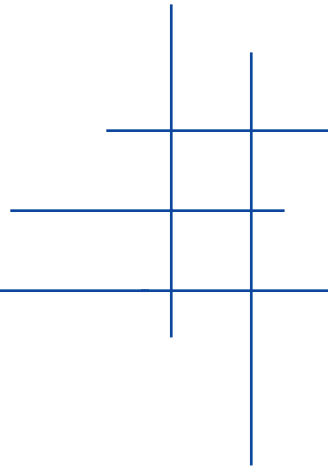
Peptide score statistics



- Peptide fragmentation not well understood
- Current scoring models are like blast e-values without PAM matrices
- Results often difficult to interpret



Sequence Databases

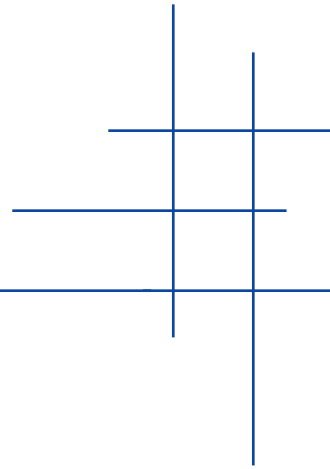


Issues:

- Search engines need exact sequence
- Sequence databases contain redundant sequence
- Protein annotation is poor



Simple Linear Scan



Query Mass = 2018.07

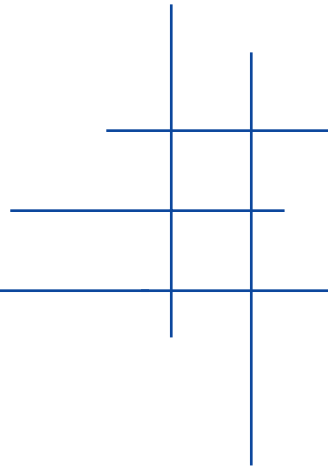
MKWVTFISLLFLFSSAYSRGV...

↑↑ ~~P~~ ~~D~~ ~~S~~ ~~L~~ ~~L~~ ~~F~~ ~~L~~ ~~F~~ ~~S~~ ~~S~~ ~~A~~ ~~Y~~ ~~S~~ ~~R~~ ~~G~~ ~~V~~ ... ↑↑ ~~1~~ ~~8~~ ~~2~~ ~~0~~ ~~2~~ ~~2~~ ~~1~~ ~~7~~ ~~0~~ ~~9~~

Output: WVTFISLLFLFSSAYSR



Search engines need exact sequence



- Enrich sequence database content
 - Protein isoforms
 - Post-translational modifications
 - Alternate splice forms
 - SNPs and polymorphism
- Link proteomics to genomics!
 - Phenotypic or clinical annotations



Swiss-Prot

NiceProt View of Swiss-Prot: P13746 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Folders Links

Address http://us.expasy.org/cgi-bin/niceprot.pl?1A11_HUMAN Go

Google Search Web Search Site Search Groups Search Directory News Page Info 822 blocked Options Up High

[ExPASy Home page](#) [Site Map](#) [Search ExPASy](#) [Contact us](#) [Swiss-Prot](#)

Hosted by [NCSC US](#) Mirror sites: [Australia](#) [Bolivia](#) [Canada](#) [China](#) [Korea](#) [Switzerland](#) [Taiwan](#)

Search for

NiceProt View of Swiss-Prot: P13746

[\[Entry info\]](#) [\[Name and origin\]](#) [\[References\]](#) [\[Comments\]](#) [\[Cross-references\]](#) [\[Keywords\]](#) [\[Features\]](#) [\[Sequence\]](#) [\[Tools\]](#)

Note: most headings are clickable, even if they don't appear as links. They link to the [user manual](#) or [other documents](#).

Entry information

Entry name	1A11_HUMAN
Primary accession number	P13746
Secondary accession numbers	O19605 O19606 Q29747 Q29835 Q9BCN0 Q9MYI5 Q9TQE9 Q9TQP6 Q9TQP7
Entered in Swiss-Prot in	Release 13, January 1990
Sequence was last modified in	Release 13, January 1990
Annotations were last modified in	Release 42, October 2003

Name and origin of the protein

Protein name	HLA class I histocompatibility antigen, A-11 alpha chain [Precursor]
Synonym	MHC class I antigen A*11
Gene name	HLA-A or HLAA
From	Homo sapiens (Human) [TaxID: 9606]
Taxonomy	Eukaryota ; Metazoa ; Chordata ; Craniata ; Vertebrata ; Euteleostomi ; Mammalia ; Eutheria ; Primates ; Catarrhini ; Hominidae ; Homo .



Swiss-Prot Variant Annotations

NiceProt View of Swiss-Prot: P13746 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://us.expasy.org/cgi-bin/niceprot.pl?1A11_HUMAN

Google SwissPROT Search Web Search Site Search Groups Search Directory News Page Info 822 blocked Options Up High

Comments

- **FUNCTION:** Involved in the presentation of foreign antigens to the immune system.
- **SUBUNIT:** Heterodimer of an alpha chain and a beta chain (beta-2-microglobulin).
- **SUBCELLULAR LOCATION:** Type I membrane protein.
- **ALTERNATIVE PRODUCTS:**
 - Alternative splicing [2 named forms] [Display all isoform sequences in Fasta format](#)

Name	1
Isoform ID	P13746-1
This is the isoform sequence displayed in this entry .	

Name	2
Synonyms	Long
Isoform ID	P13746-2
<i>Note:</i> Only produced by allele A*1103.	
Features which should be applied to build the isoform sequence: VSP_008099 .	

- **POLYMORPHISM:** The following alleles of A-11 are known: A*1101 (A-11E), A*1102 (A-11K), A*1103, A*1104, A*1105 and A*1107. The sequence shown is that of A*1101.

Copyright

This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

Cross-references

X13111; CAA31503.1; -	[EMBL / GenBank / DDBJ] [CoDingSequence]
X13112; CAA31504.1; -	[EMBL / GenBank / DDBJ] [CoDingSequence]
D16841; BAA04117.1; -	[EMBL / GenBank / DDBJ] [CoDingSequence]
D16842; BAA04118.1; -	[EMBL / GenBank / DDBJ] [CoDingSequence]
M16010; AAA65449.1; -	[EMBL / GenBank / DDBJ] [CoDingSequence]
M16007; AAA65449.1; JOINED	[EMBL / GenBank / DDBJ] [CoDingSequence]

Internet



Swiss-Prot Variant Annotations

NiceProt View of Swiss-Prot: P13746 - Microsoft Internet Explorer

Address: http://us.expasy.org/cgi-bin/niceprot.pl?P1A11_HUMAN

Features

[Feature table viewer](#) [Feature aligner](#)

Key	From	To	Length	Description	FTId
SIGNAL	1	24	24		
CHAIN	25	365	341	HLA class I histocompatibility antigen, A-11 alpha chain.	
DOMAIN	25	114	90	Extracellular alpha-1.	
DOMAIN	115	206	92	Extracellular alpha-2.	
DOMAIN	207	298	92	Extracellular alpha-3.	
DOMAIN	299	308	10	Connecting peptide.	
TRANSMEM	309	332	24		
DOMAIN	333	365	33	Cytoplasmic tail.	
CARBOHYD	110	110		N-linked (GlcNAc...) (By similarity).	
DISULFID	125	188		By similarity.	
DISULFID	227	283		By similarity.	
VARSPPLIC	337	337		S -> SGGEGVK (in isoform 2).	VSP_008099
VARIANT	43	43	*	E -> K (in allele A*1102).	VAR_004353
VARIANT	133	133	*	F -> L (in allele A*1107).	VAR_016731
VARIANT	168	168	*	K -> E (in allele A*1105).	VAR_016732
VARIANT	175	175	*	H -> R (in allele A*1103).	VAR_016733
VARIANT	176	176	*	A -> E (in allele A*1103).	VAR_016734
VARIANT	187	187	*	R -> T (in allele A*1104).	VAR_016735
VARIANT	345	345	*	T -> S (in allele A*1105).	VAR_016736

Sequence information

Length: 365 AA [This is the length of the unprocessed precursor]	Molecular weight: 40937 Da [This is the MW of the unprocessed precursor]	CRC64: FE449CE2D4BF6CC5 [This is a checksum on the sequence]
--	--	--

10 20 30 40 50 60

MANARPTL...L...G...A...L...T...G...H...G...M...R...E...Y...T...S...I...D...R...C...D...E...R...E...L...A...N...G...W...I...D...T...O...S...U...R...E...



Swiss-Prot Sequence

NiceProt View of Swiss-Prot: P13746 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Folders

Address http://us.expasy.org/cgi-bin/niceprot.pl?1A11_HUMAN Go

Google Search Web Search Site Search Groups Search Directory News Page Info 822 blocked Options Up High

Sequence information

Length: 365 AA [This is the length of the unprocessed precursor] Molecular weight: 40937 Da [This is the MW of the unprocessed precursor] CRC64: FE449CE2D4BF6CC5 [This is a checksum on the sequence]

10	20	30	40	50	60
MAVMAPRTLL	LLLSGALALT	QTWAGSHSMR	YFYTSVSRPG	RGEPRFIAVG	YVDDTQFVRF
70	80	90	100	110	120
DSDAASQRME	PRAPWIEQEG	PEYWDQETRN	VKAQSQTRV	DLGTLRGYYN	QSEDEGSHTIQ
130	140	150	160	170	180
IMYGCDVGPD	GRFLRGYRQD	AYDGKDYIAL	NEDLRSWTA	DMAAQITKRK	WEAAHAAEQQ
190	200	210	220	230	240
RAYLEGRCVE	WLRRYLENGK	ETLQRTDPPK	THMTHHPISD	HEATLRCWAL	GFYPAEITLT
250	260	270	280	290	300
WQRDGEDQTQ	DTELVETRPA	GDGTFQKWA	VVVPSEGEQR	YTCHVQHEGL	PKPLTLRWEL
310	320	330	340	350	360
SSQPTPIVIG	IIAGLVLLGA	VITGAVVA	MWRKSSDRK	GGSYTQAASS	DSAQGSVSL

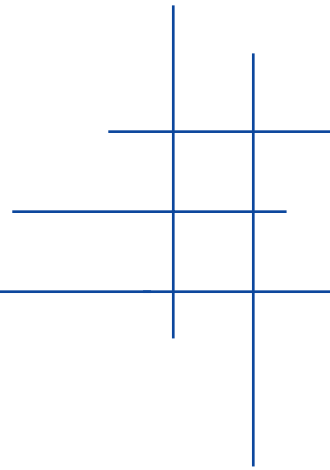
TACKV

P13746 in [FASTA format](#)

[View entry in original Swiss-Prot format](#)
[View entry in raw text format \(no links\)](#)



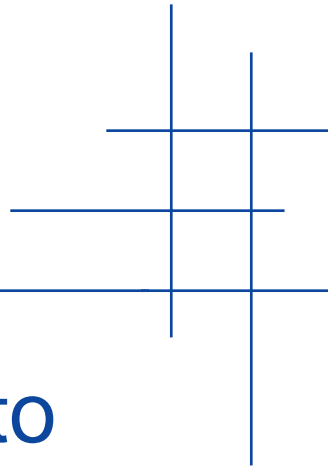
Swiss-Prot



- VarSplic enumerates all variants, conflicts, isoforms
- Swiss-Prot sequence size:
 - 56 Mb
- VarSplic sequence size:
 - 90 Mb
- How many more peptide candidates?



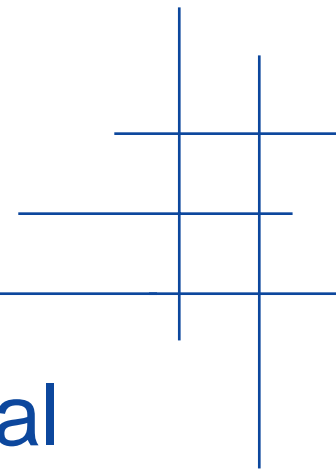
Alternate Splice Forms and SNPs



- Map all types of sequence evidence to genome
- Enumerate sequence appropriately & encode evidence
- Peptide to evidence decoding



Sequence Databases are Redundant

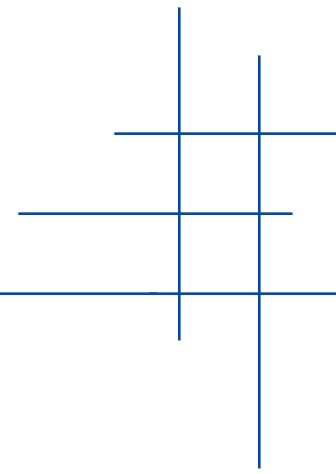


- Search times are typically proportional to database size
- Protein isoform enumeration makes for *very* redundant databases

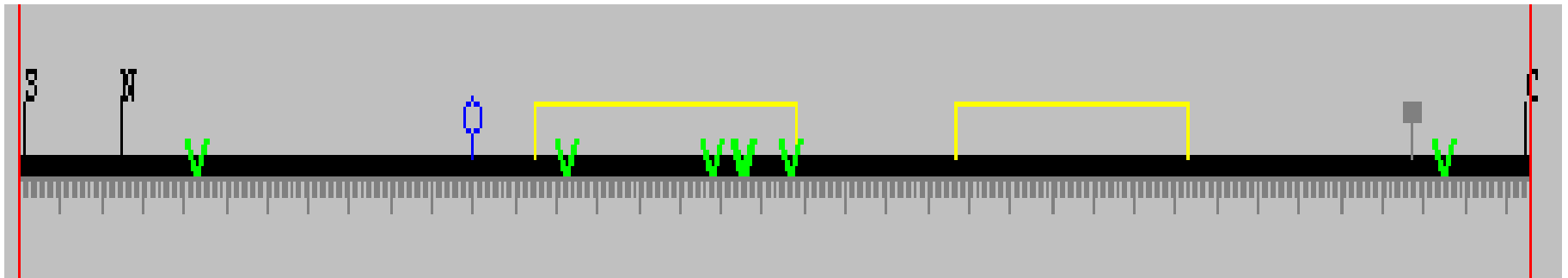
Sequence Database	Swiss-Prot	Swiss-Prot All Isoforms
Size	56 Mb	90 Mb
30-mers	44 Mb	45 Mb
Overhead	27%	97%



Swiss-Prot Variant Annotations



Feature viewer



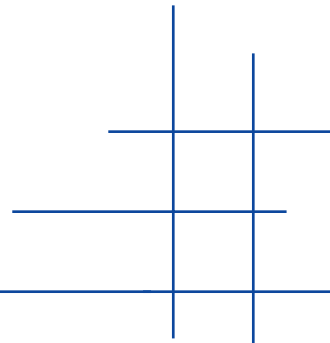
Variants



CELERA

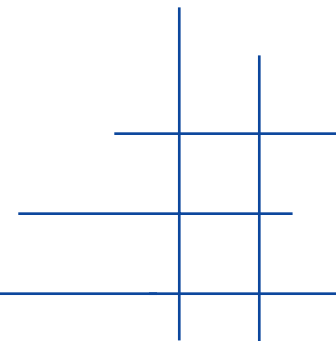
AB Applied Biosystems

Swiss-Prot VarSplic Output



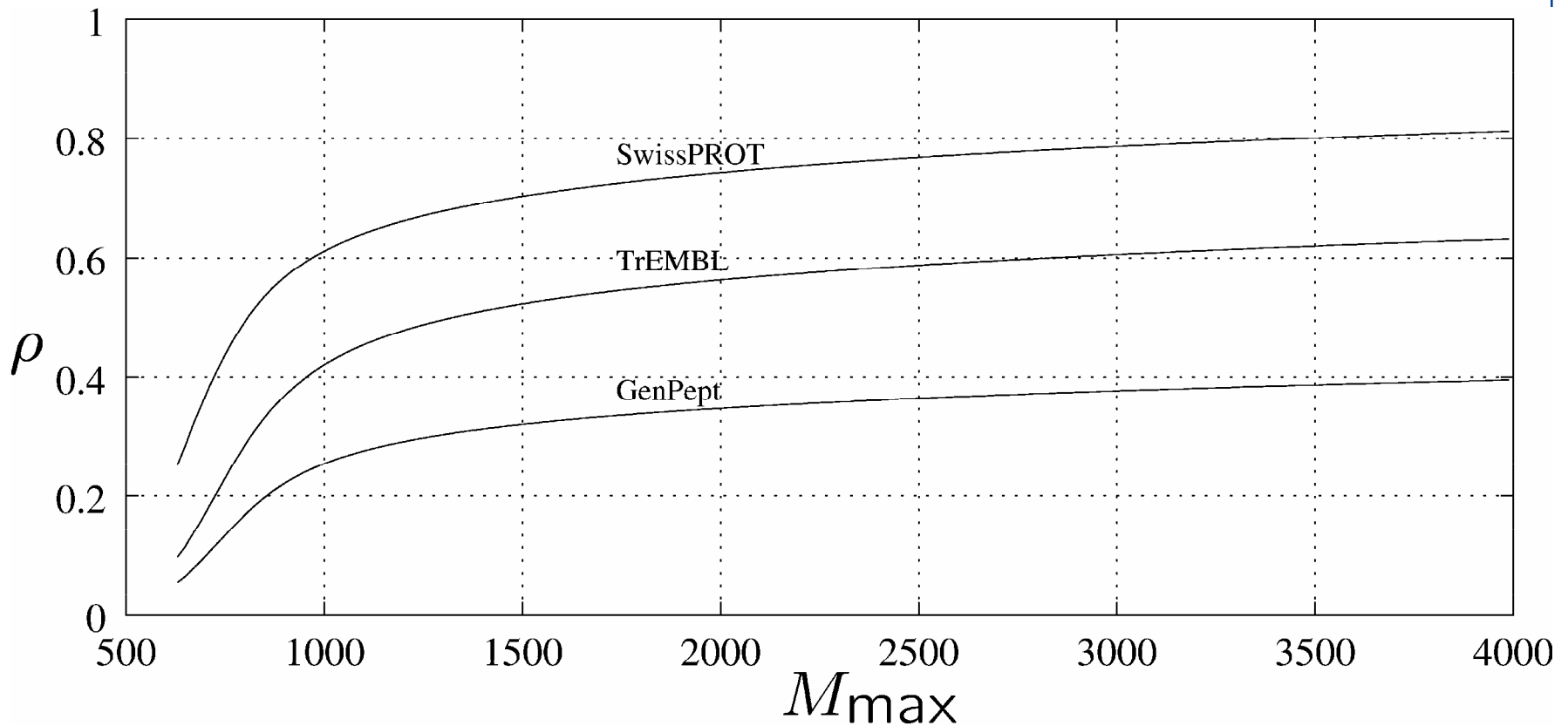
```
P13746-00-01-00      MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F Y T S V S R P G R G E P R F I A V G Y V D D T Q F V R F
P13746-01-01-00      MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F Y T S V S R P G R G E P R F I A V G Y V D D T Q F V R F
P13746-00-00-00      MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F Y T S V S R P G R G E P R F I A V G Y V D D T Q F V R F
P13746-00-03-00      MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F Y T S V S R P G R G E P R F I A V G Y V D D T Q F V R F
P13746-01-03-00      MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F Y T S V S R P G R G E P R F I A V G Y V D D T Q F V R F
P13746-00-04-00      MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F Y T S V S R P G R G K P R F I A V G Y V D D T Q F V R F
P13746-01-04-00      MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F Y T S V S R P G R G K P R F I A V G Y V D D T Q F V R F
P13746-00-05-00      MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F Y T S V S R P G R G E P R F I A V G Y V D D T Q F V R F
P13746-01-05-00      MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F Y T S V S R P G R G E P R F I A V G Y V D D T Q F V R F
P13746-01-00-00      MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F Y T S V S R P G R G E P R F I A V G Y V D D T Q F V R F
P13746-00-02-00      MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F Y T S V S R P G R G E P R F I A V G Y V D D T Q F V R F
P13746-01-02-00      MAVMAPRTL L L L L L S G A L A L T Q T W A G S H S M R Y F Y T S V S R P G R G E P R F I A V G Y V D D T Q F V R F
***** : *****
```

Swiss-Prot VarSplic Output

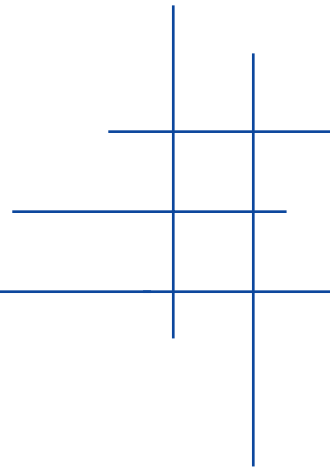


```
P13746-00-01-00      SSQPTIPIVGIIAGLVLLGAVITGAVVAAVMWRKSS-----DRKGGSYTQAASSDSAQ
P13746-01-01-00      SSQPTIPIVGIIAGLVLLGAVITGAVVAAVMWRKSSGGEGVKDRKGGSYTQAASSDSAQ
P13746-00-00-00      SSQPTIPIVGIIAGLVLLGAVITGAVVAAVMWRKSS-----DRKGGSYTQAASSDSAQ
P13746-00-03-00      SSQPTIPIVGIIAGLVLLGAVITGAVVAAVMWRKSS-----DRKGGSYTQAASSDSAQ
P13746-01-03-00      SSQPTIPIVGIIAGLVLLGAVITGAVVAAVMWRKSSGGEGVKDRKGGSYTQAASSDSAQ
P13746-00-04-00      SSQPTIPIVGIIAGLVLLGAVITGAVVAAVMWRKSS-----DRKGGSYTQAASSDSAQ
P13746-01-04-00      SSQPTIPIVGIIAGLVLLGAVITGAVVAAVMWRKSSGGEGVKDRKGGSYTQAASSDSAQ
P13746-00-05-00      SSQPTIPIVGIIAGLVLLGAVITGAVVAAVMWRKSS-----DRKGGSYTQAASSDSAQ
P13746-01-05-00      SSQPTIPIVGIIAGLVLLGAVITGAVVAAVMWRKSSGGEGVKDRKGGSYTQAASSDSAQ
P13746-01-00-00      SSQPTIPIVGIIAGLVLLGAVITGAVVAAVMWRKSSGGEGVKDRKGGSYTQAASSDSAQ
P13746-00-02-00      SSQPTIPIVGIIAGLVLLGAVITGAVVAAVMWRKSS-----DRKGGSYSQAASSDSAQ
P13746-01-02-00      SSQPTIPIVGIIAGLVLLGAVITGAVVAAVMWRKSSGGEGVKDRKGGSYSQAASSDSAQ
*****                               *****:*****
```


Substring Density (ρ)



Peptide Candidates



- Parent ion
 - Typically < 3000 Da
- Tryptic Peptides
 - Cut at K or R
- Search engines
 - Don't handle > 4+ well
 - Long peptides don't fragment well
- # of distinct 30-mers upper bounds total peptide content



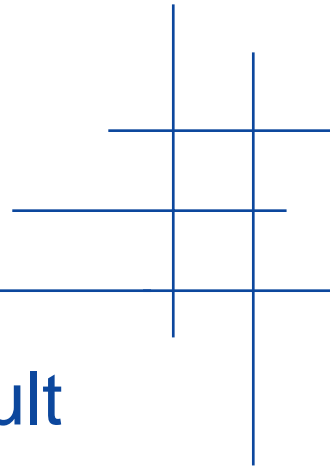
Sequence Databases are Redundant

- Compressed sequence is guaranteed
 - Complete, Correct, Compact

Sequence Database	Swiss-Prot	Swiss-Prot All Isoforms
Original Size	56 Mb	90 Mb
Distinct 30-mers	44 Mb	45 Mb
Overhead	27%	97%
SBH-Graph Comp. Size	53 Mb	55 Mb
SBH-graph Compression	93%	61%



Protein Annotation is Poor



- FASTA file deflines are inconsistent, difficult to parse, and lack controlled vocabularies
- Proteomics workflows interrogate a whole lot more than mass!
- Need to provide slice and dice, by
 - Organism
 - Genomic region
 - Cellular location
 - Molecular function



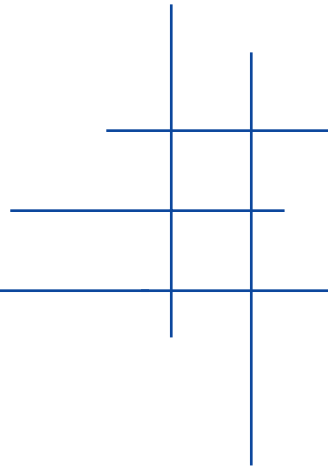
NCBI - NRAA

```
>gi|4504255|ref|NP_002097.1| H2A histone family, member Z; H2AZ histone [Homo sapiens]
>gi|7949045|ref|NP_058030.1| H2A histone family, member Z; histone H2A.Z [Mus musculus]
>gi|12083611|ref|NP_073165.1| H2A histone family, member Z [Rattus norvegicus]
>gi|27807373|ref|NP_777234.1| H2A histone family, member Z [Bos taurus]
>gi|121994|sp|P17317|H2AZ_HUMAN Histone H2A.z (H2A/z)
>gi|89608|pir||S03642 histone H2A.Z - bovine
>gi|92380|pir||S03644 histone H2A.Z - rat
>gi|106267|pir||A35881 histone H2A.Z - human
>gi|11513399|pdb|1F66|C Chain C, 2.6 A Crystal Structure Of A Nucleosome Core Particle
  Containing The Variant Histone H2a.Z
>gi|11513403|pdb|1F66|G Chain G, 2.6 A Crystal Structure Of A Nucleosome Core Particle
  Containing The Variant Histone H2a.Z
>gi|410|emb|CAA36554.1| histone H2A.Z (AA 1-127) [Bos taurus]
>gi|31975|emb|CAA36553.1| histone H2A.Z (AA 1-127) [Homo sapiens]
>gi|57808|emb|CAA36552.1| histone H2A.Z (AA 1-127) [Rattus sp.]
>gi|163150|gb|AAA30566.1| histone (H2A.Z)
>gi|184060|gb|AAA35984.1| histone (H2A.Z)
>gi|204599|gb|AAA41329.1| histone (H2A.Z)
>gi|1575713|gb|AAB09578.1| histone H2A.Z [Mus musculus]
>gi|3649600|gb|AAC61625.1| histone [Homo sapiens]
>gi|17389988|gb|AAH18002.1|AAH18002 H2A histone family, member Z [Homo sapiens]
>gi|18089228|gb|AAH20936.1|AAH20936 H2A histone family, member Z [Homo sapiens]
>gi|18535649|gb|AAL71863.1| histone H2A.Z [Ovis aries]
>gi|18535651|gb|AAL71864.1| histone H2A.Z [Mus musculus]
>gi|26353770|dbj|BAC40515.1| unnamed protein product [Mus musculus]
>gi|26389789|dbj|BAC25791.1| unnamed protein product [Mus musculus]
```

```
MAGGKAGKDSGKAKTKAVSRSQRAGLQFPVGR IHRHLKSRTTSHGRVGAT
AAVYSAAILLEYLTAEVLELAGNASKDLKVKRITPRHLQLAIRGDEELDSL
IKATIAGGGVIPHIHKS LIGKKGQOKTV
```

CELERA

Mascot & Organism

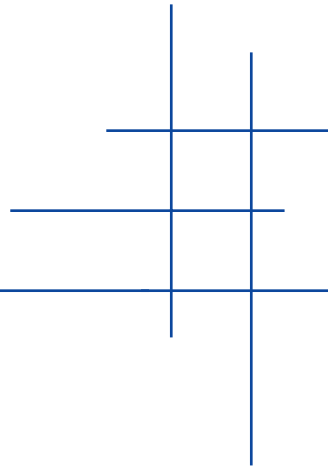


Mascot slices sequence databases by organism at run-time!

- Organism name from defline
- NCBI controlled vocabulary & synonyms
- NCBI taxonomy tree
 - Specify mammals, rodents, humans.



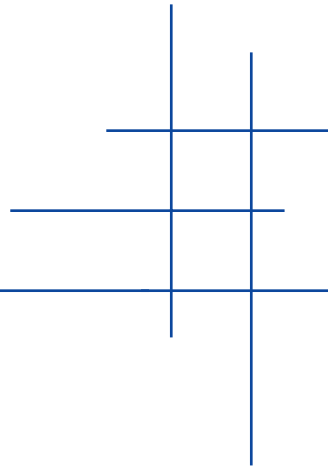
Other Annotations



- Panther
 - Biological Process
 - Molecular Function
- Gene Ontology
- Chromosome Cytobands
 - 12p4 – 12p6
- Biochemical Properties
 - Hydrophobicity
 - Iso-Electric Point



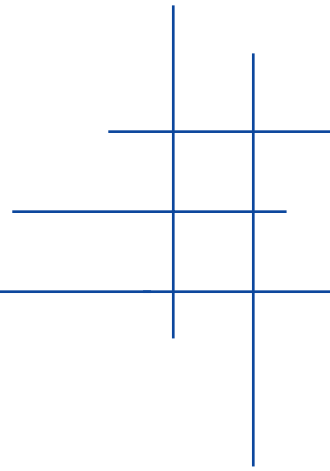
Final words



- Many unsolved bioinformatics challenges in protein identification
- Disconnect between sequence databases and their use in proteomics
- Good quality protein annotation is important



Questions?



CELERA

AB Applied
Biosystems