Carbohydrates and proteins

Risto Renkonen
Haartman institute

Medicel
RESEARCH REDEFINED

Sokereita biologiassa

Nukleiinihapot, proteiinit ja sokerit ovat biopolymeereja

Sokereiden erityispiirre on monimuotoisuus:

- DNA:n neljästä nukleotidistä voidaan järjestää 64:llä eri tavalla kolmen nukleotidin ketju
- Neljästä monosakkaridista voidaan saada yli 10.000 erilaista kolmikkoa

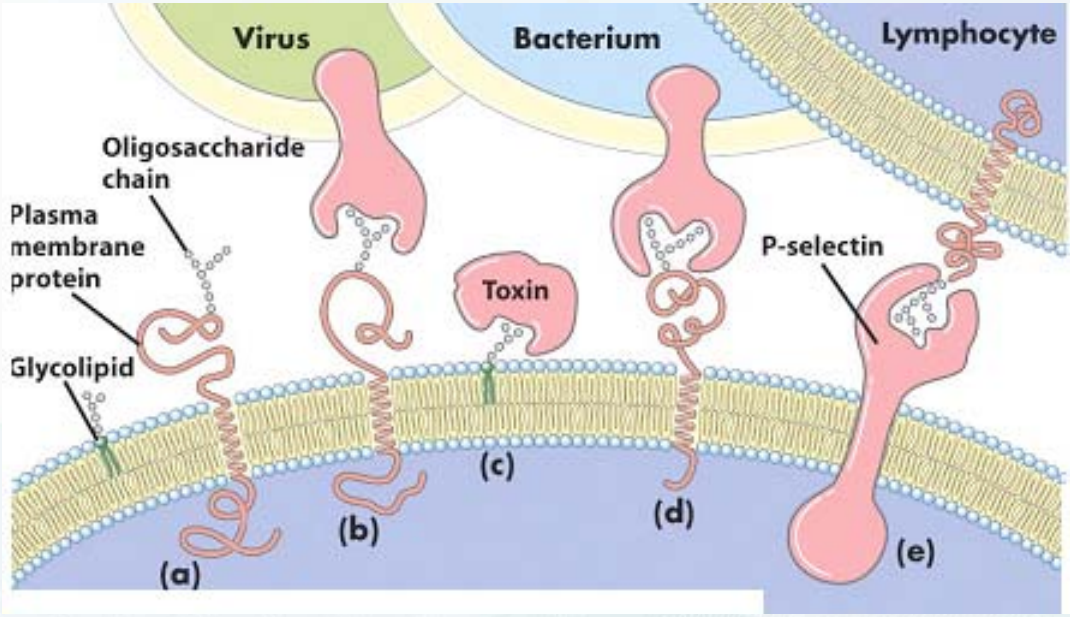
Lektiinit

- 1948 HY:ssä löydettiin lektiinit
  - 'lectin' latinan sanasta *legere*, 'valikoida'
  - proteiineja, jotka tunnistavat sokereita
  - ABO-veriryhmille spesifit lektiinit

- 1990 nisäkäslektiinit
- Selektiinit - mukana tulehdusreaktioissa
- Galectins, C-type lectins, Siglecs jne.

Vicia cracca
Hiirenvirna

Role of glycans in medicine



a = sokerit-lektiinit
b = flunssa virus tarttuu epiteelin glykoproteiineihin
c = pertussis toksiini tarttuu glykolipideihin
d = *Helicobacter pylori* voi tarttua moniin sokerirakenteisiin
e = selektiinit välittää valko- ja syöpäsolujen tarttumista

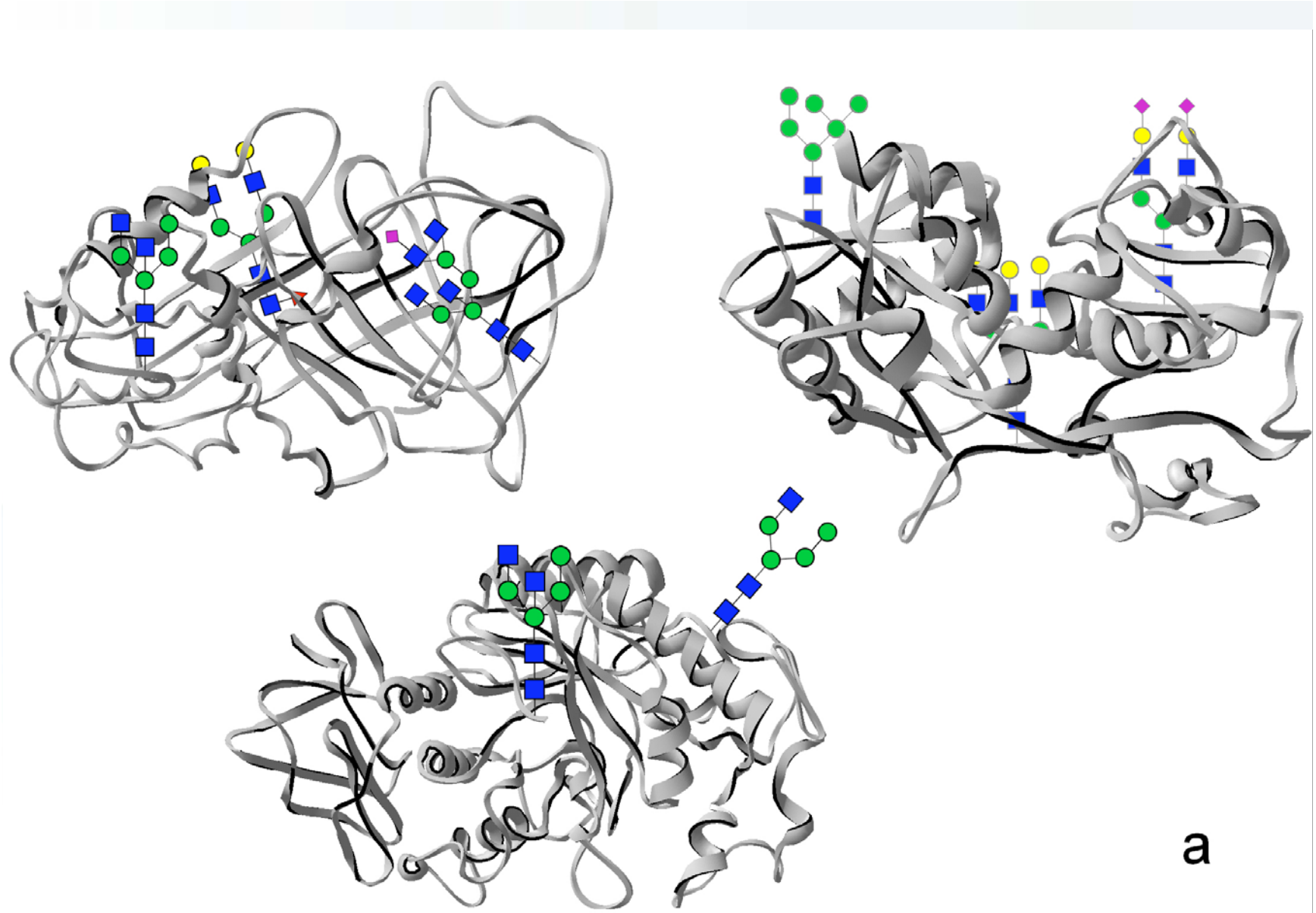
Medicel
RESEARCH REDEFINED

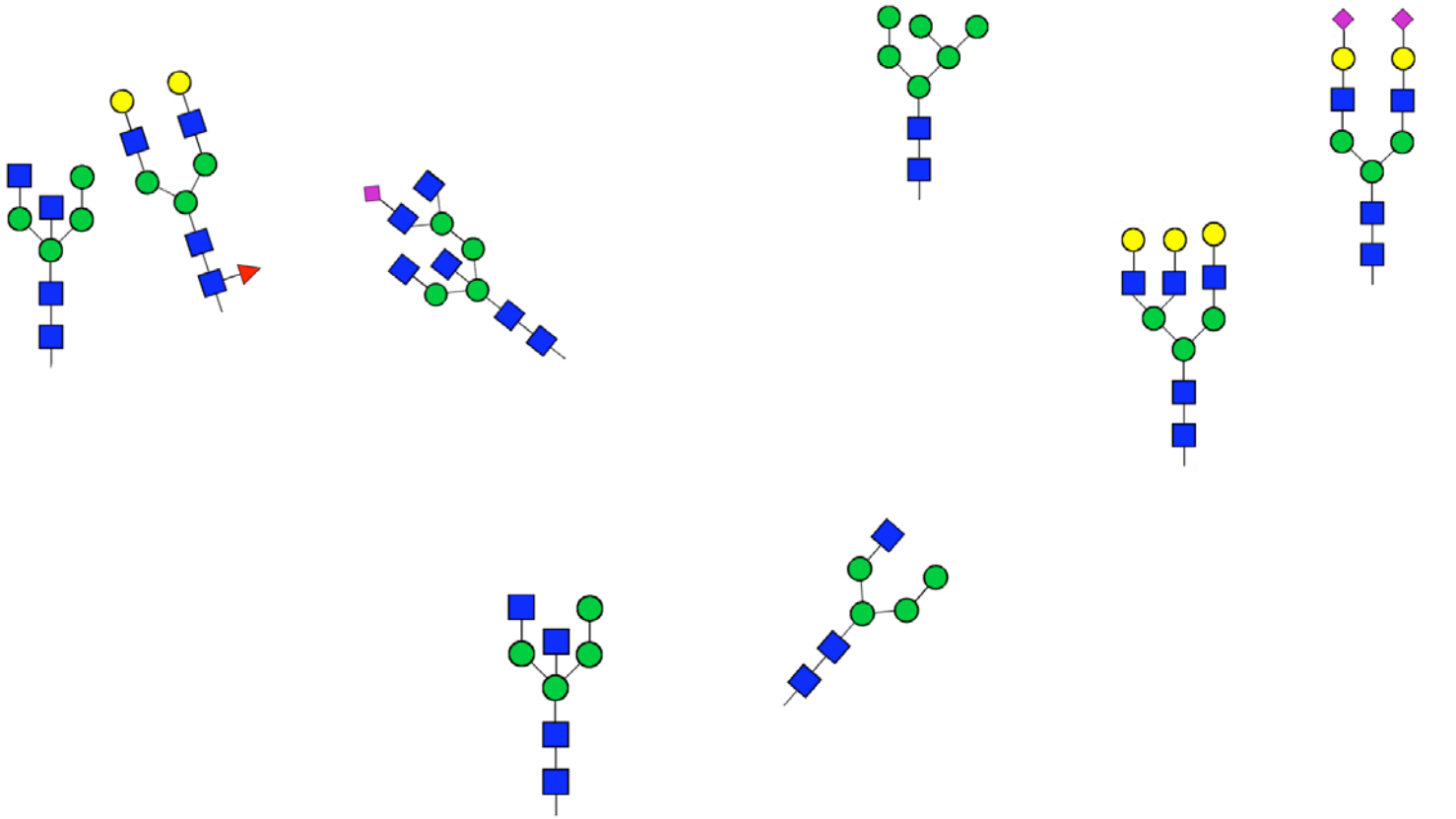
N-Glycopeptide Analysis

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Part 1 – General concepts

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

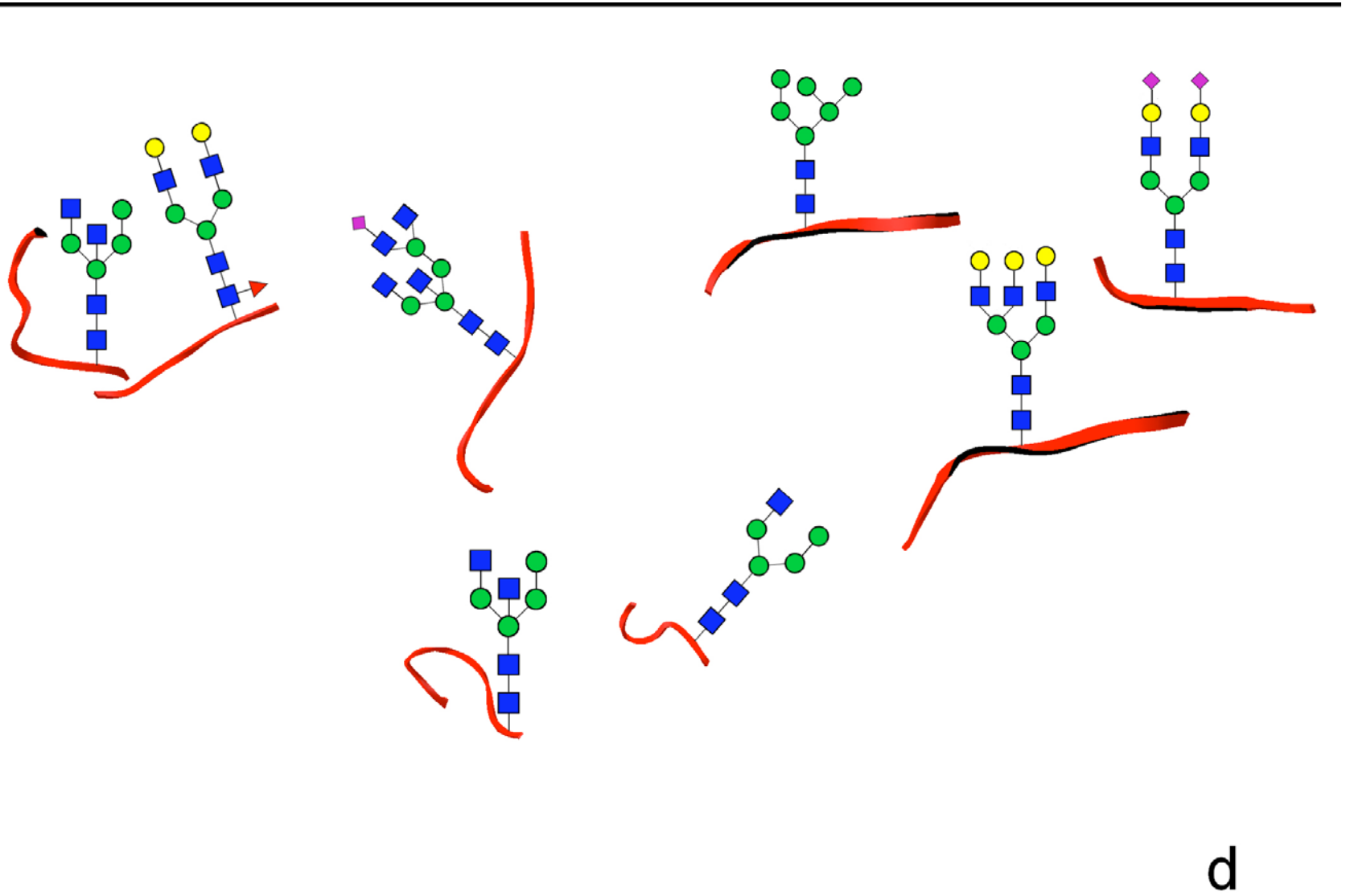

a

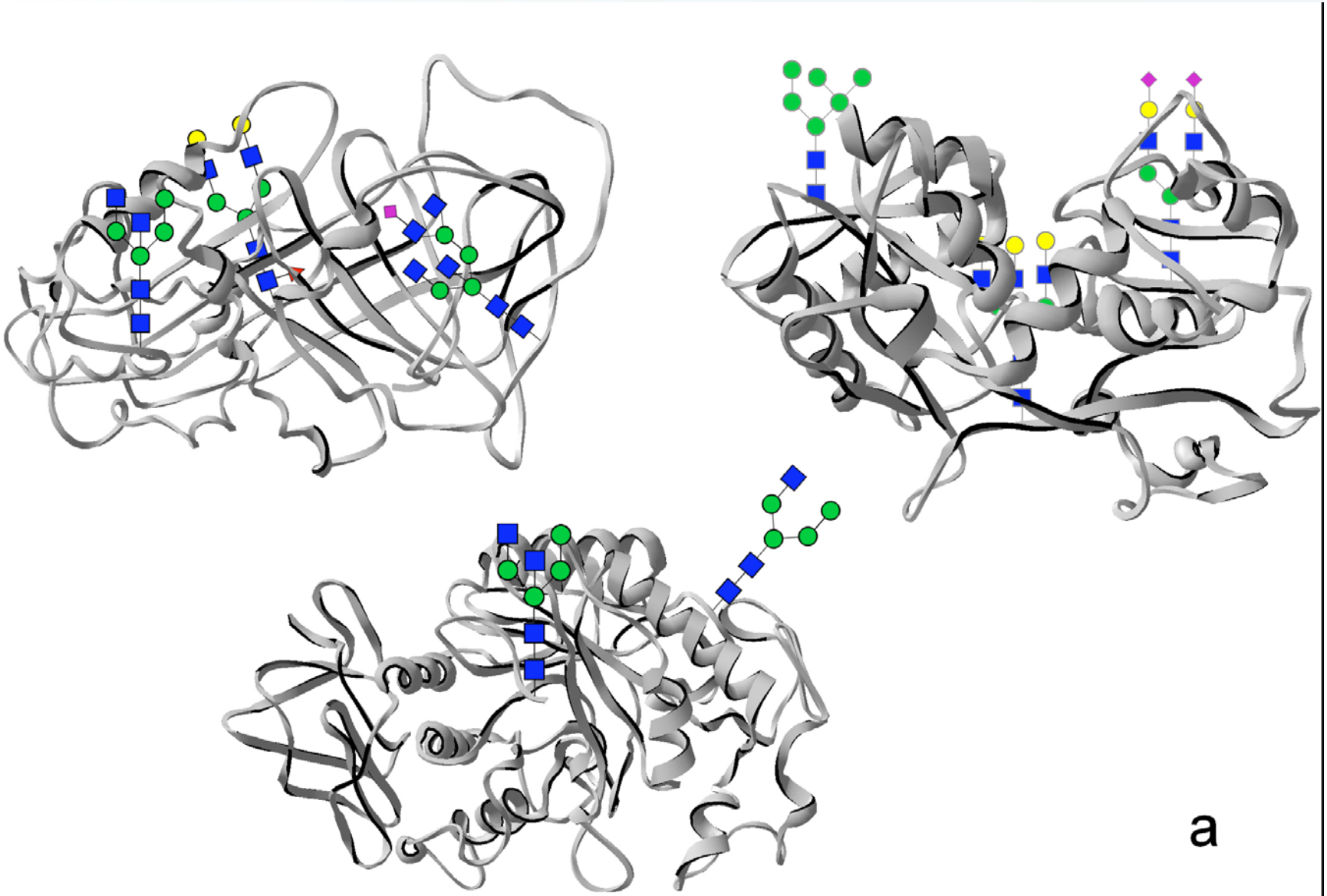


C

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com


d


a

Part 2 – Enrichment of the N-glycopeptides

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Enrichment of the glycopeptides is essential 1/2

- Glycans have a low proton affinity in the MS compared to the peptides
- Glycopeptide microheterogenity – glycovariants
  - Same glycosylation site can have a multitude of glycans
  - molar ratios of individual glycopeptide species can be extremely low compared to the unmodified peptides
- The precursor glycopeptide ion may be for instance
  - protonated, sodiated or an adduct composed of mixture of protons and sodium ions further reducing the ion abundancies
- Many potential methods: Lectins, HILIC, SEC...

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

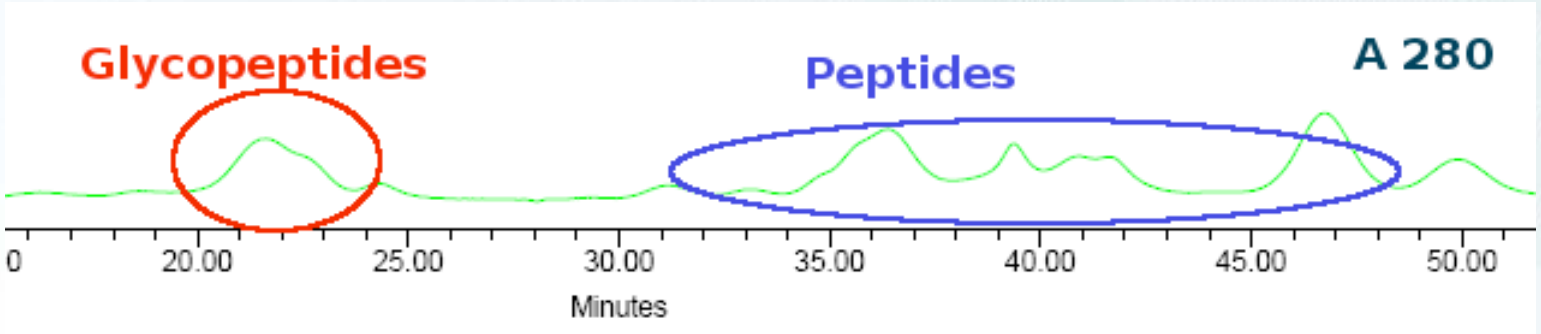
Enrichment of the glycopeptides is essential 2 /2

- SEC is our choise:
  - Glycopeptides elute first due the size compared to the unmodified peptides
  - No preference of particular glycan

Journal of Proteome Research, 2006

Tools for Glycoproteomic Analysis: Size Exclusion Chromatography Facilitates Identification of Tryptic Glycopeptides with N-linked Glycosylation Sites

Gerardo Alvarez-Manilla, James Atwood III, Yan Guo, Nicole Lynn Warren, Ron Orlando and Michael Pierce*



Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Medicel
RESEARCH REDEFINED

Part 3 – MS of N-Glycopeptides

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Medicel
RESEARCH REDEFINED

Mass Spectrometry of the glycopeptides

I. Positive ion mode LC – MS / MS
II. Recognition of the glycopeptides
    I. Glycan fragments in the CID (m/z: 366 (HexHexNAc), 292 (NeuAc))
III. Acquisition with LC-MS / MS
    I. Ramping of the collision energy
        I. Peptide fragmentation: 25 - 60 eV
        II. Glycan fragmentation: 10 - 30 eV
    II. Stopflow acquisition
        I. Essential to collect high quality data
        II. Ion intensities will increase at least 6 – 10 x
IV. Deconvolution of the spectra (MaxEnt3)
    I. Multiply charged CID daughter ions deconvoluted to singly charged species

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Medicel
RESEARCH REDEFINED

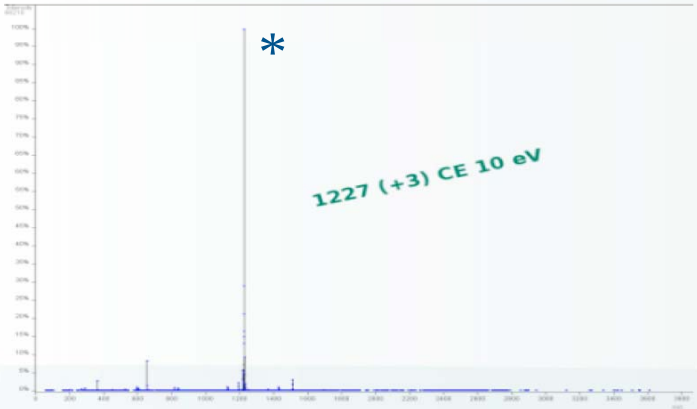
Mass Spectrometry

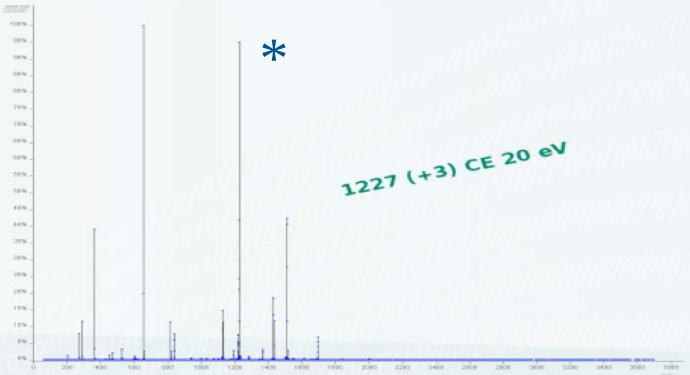
I. CapLC:
   I. Analytical reversed-phase column (RPC): 75 um x 250 mm, LC Packings PepMap C18
   II. Reversed-phase trapping column (RPC): 0.32x5 mm, LC Packings PepMap
   III. Mobile phase A: 0.1% formic acid in 5% MeCN
   IV. Mobile phase B: 0.1% formic acid in 95% MeCN
   V. Loading buffer: 0.1% formic acid in water
   VI. Flow rate: 300 – 450nl/ min

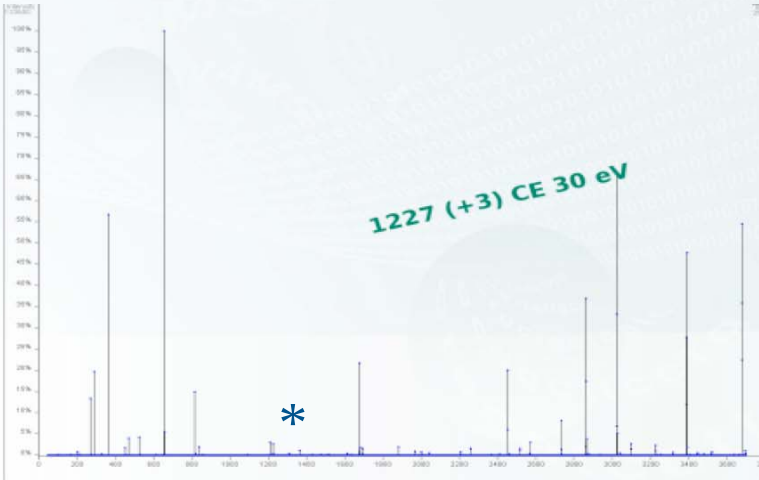
I. Micromass QTOF Ultima Global
   I. Positive ion
   II. V-mode
   III. Nano- ESI
   IV. Stopflow

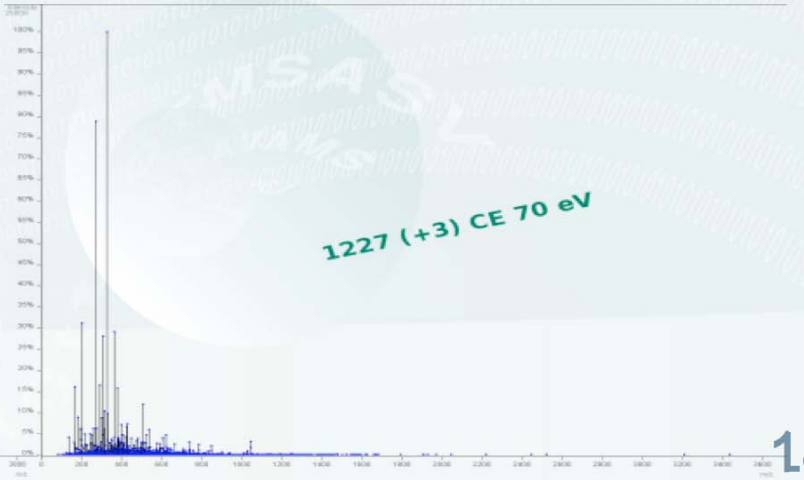
Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Effect of CE to the glycopeptide CID



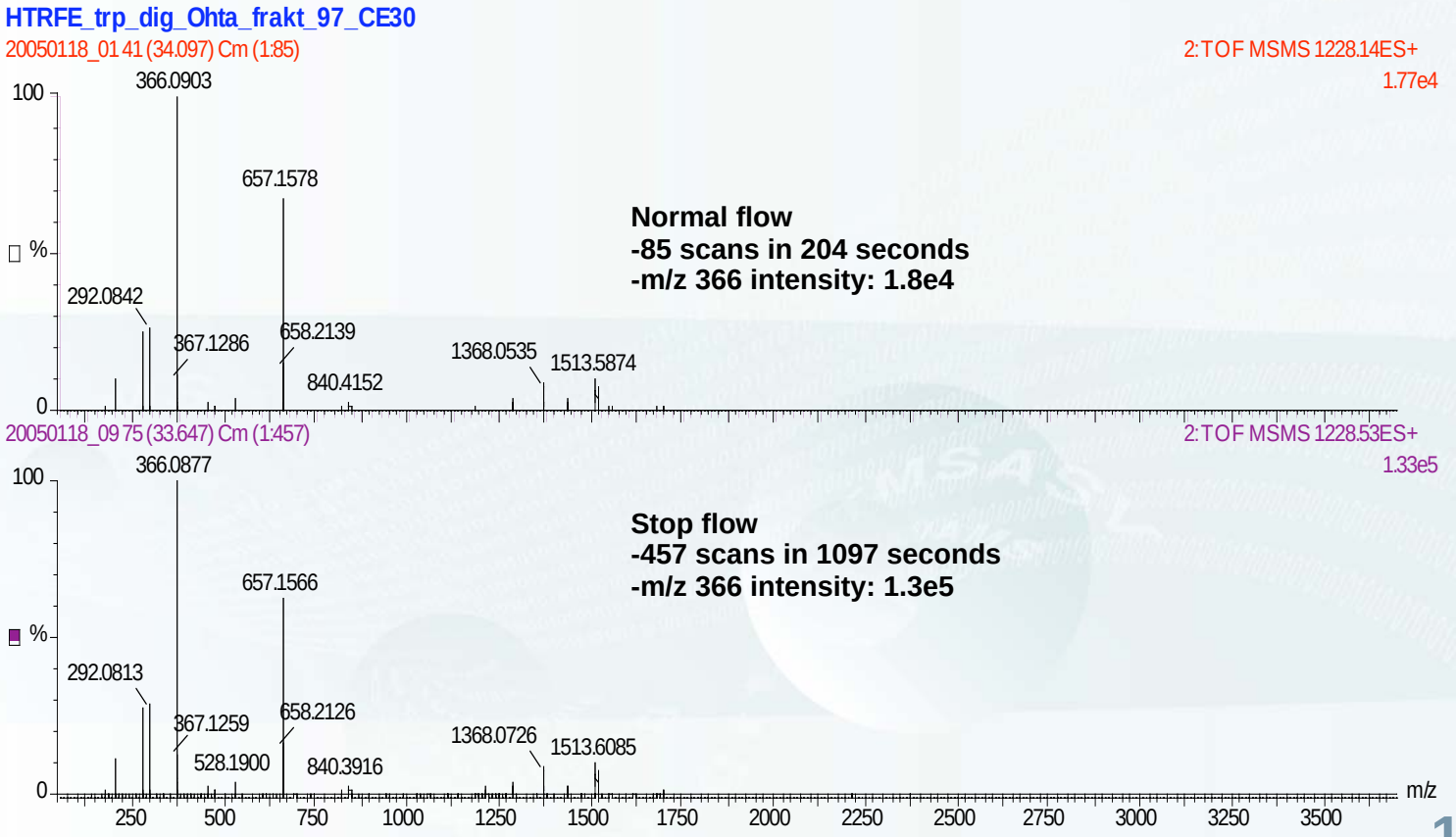






Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

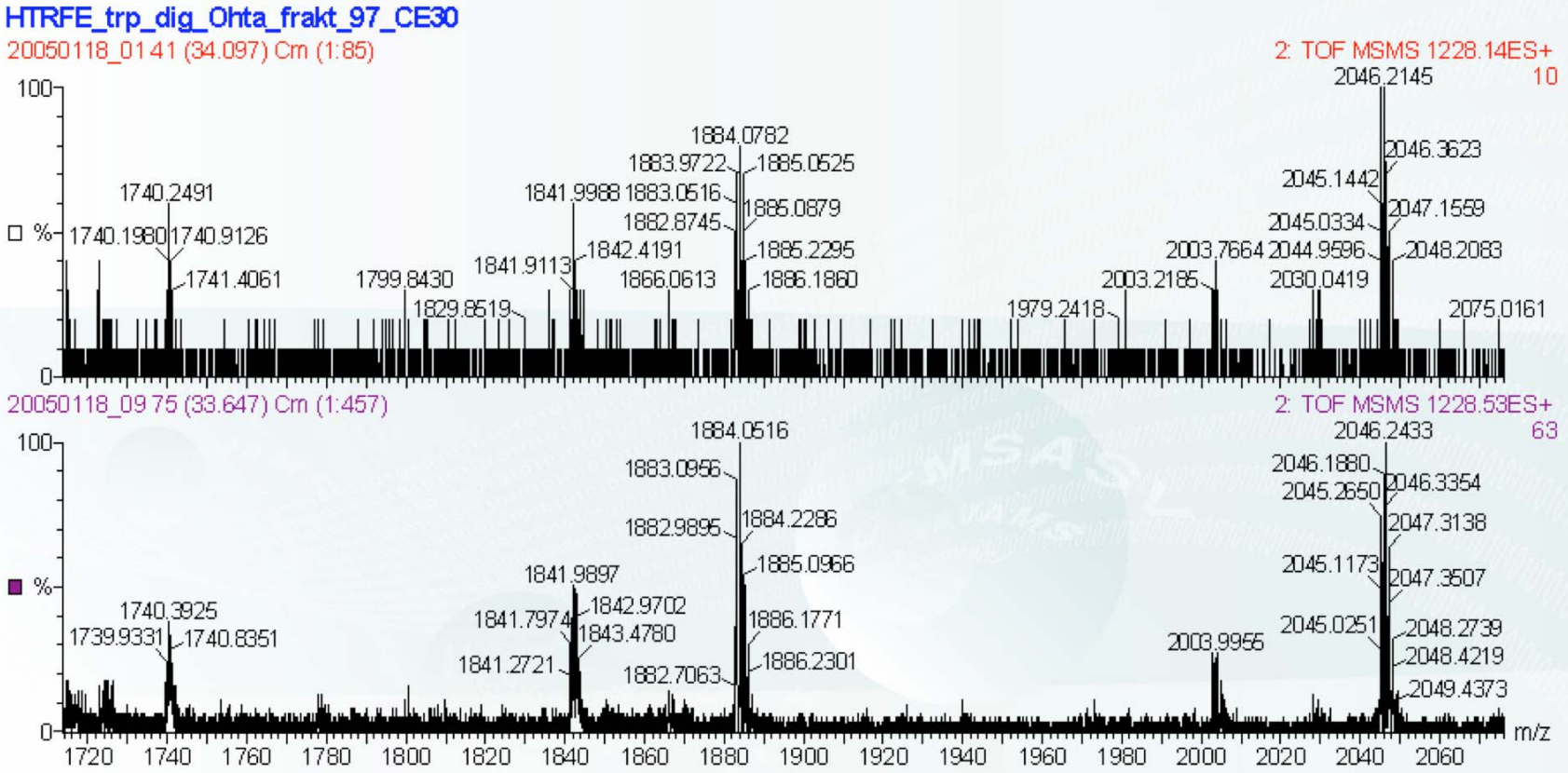
Effect of stopflow to the glycopeptide acquisition 1/2



Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Medicel
RESEARCH REDEFINED

Effect of stopflow to the glycopeptide acquisition 2/2

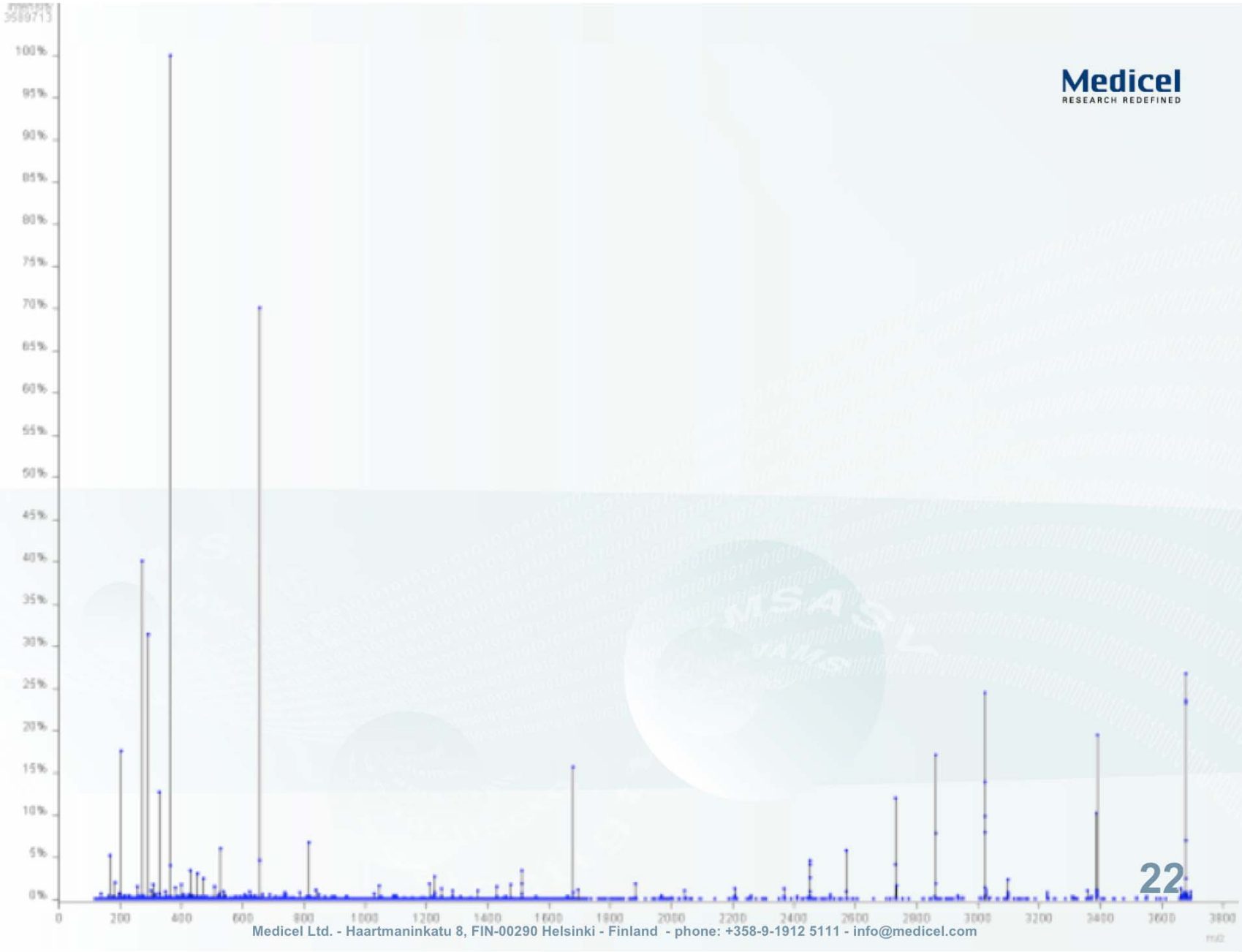


Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Medicel
RESEARCH REDEFINED

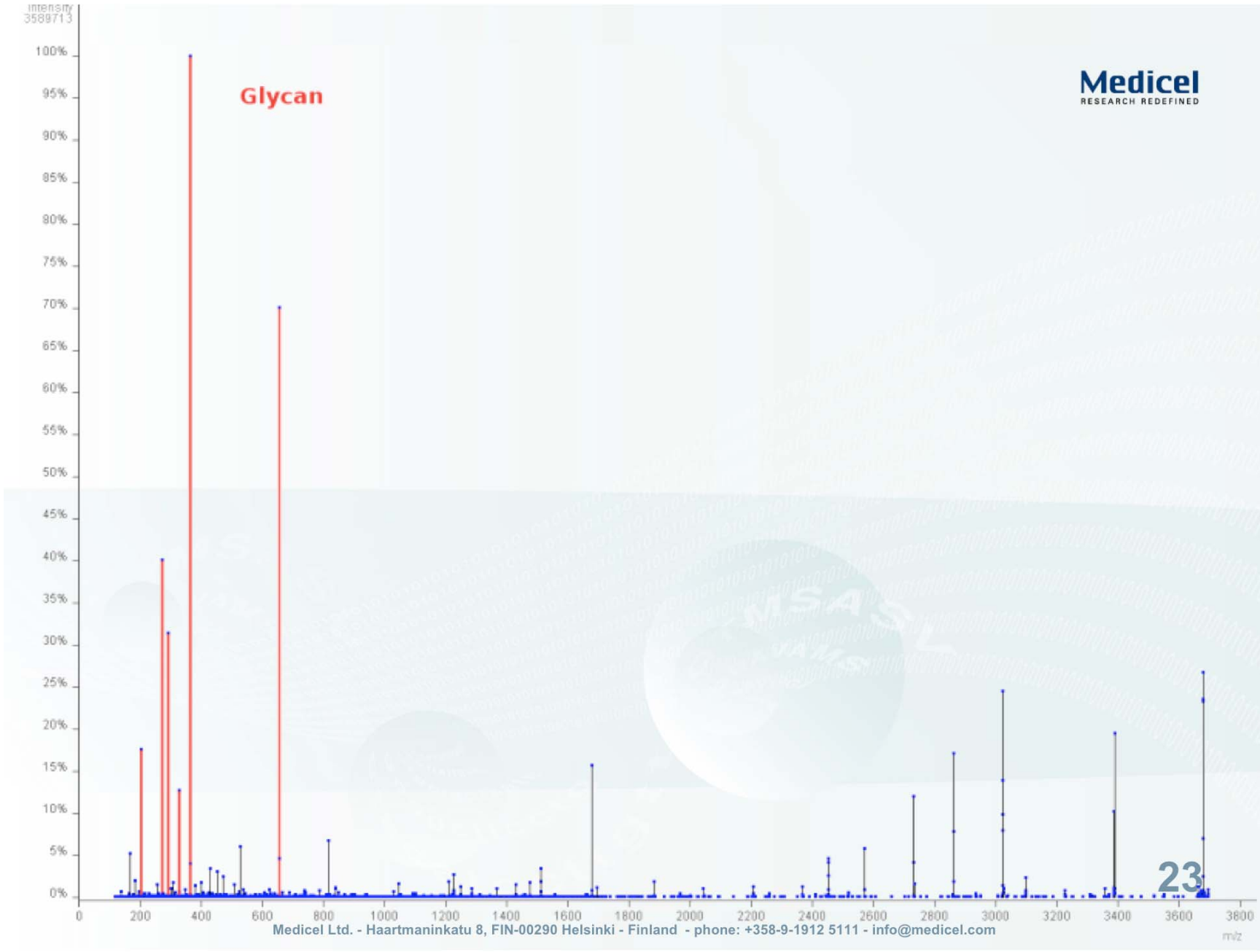
Part 3 – Analysis of Glycopeptide spectra

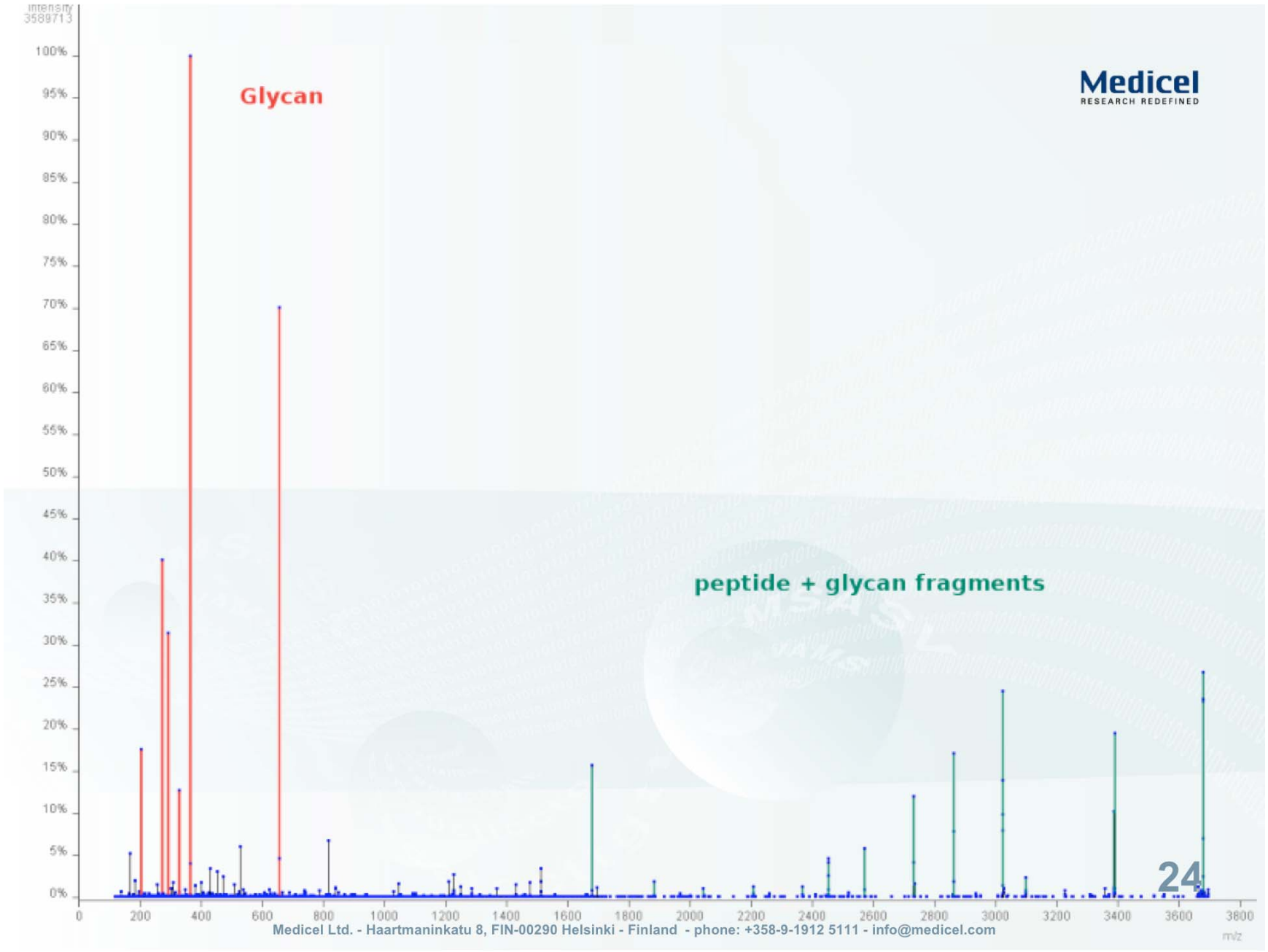
Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

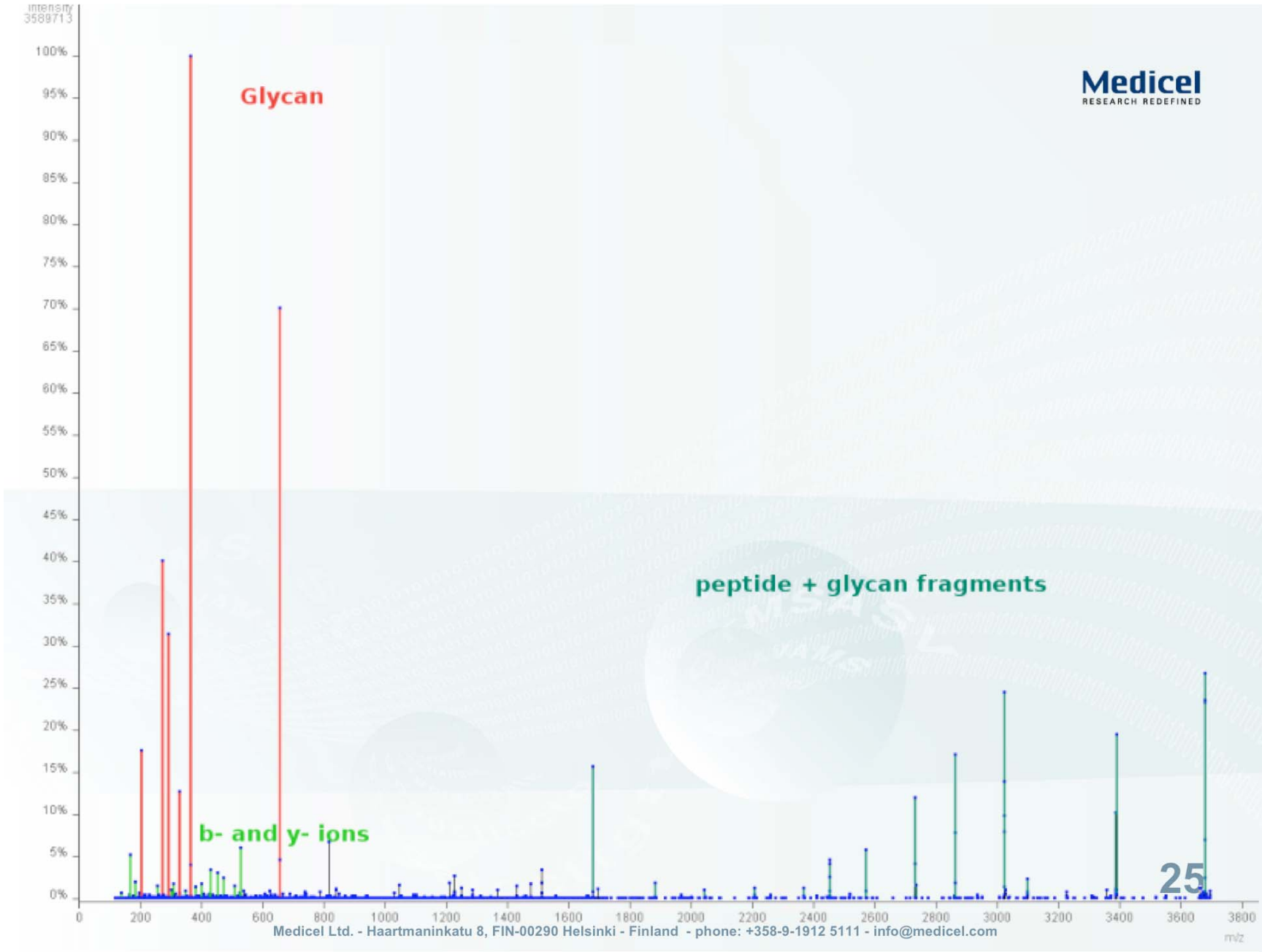


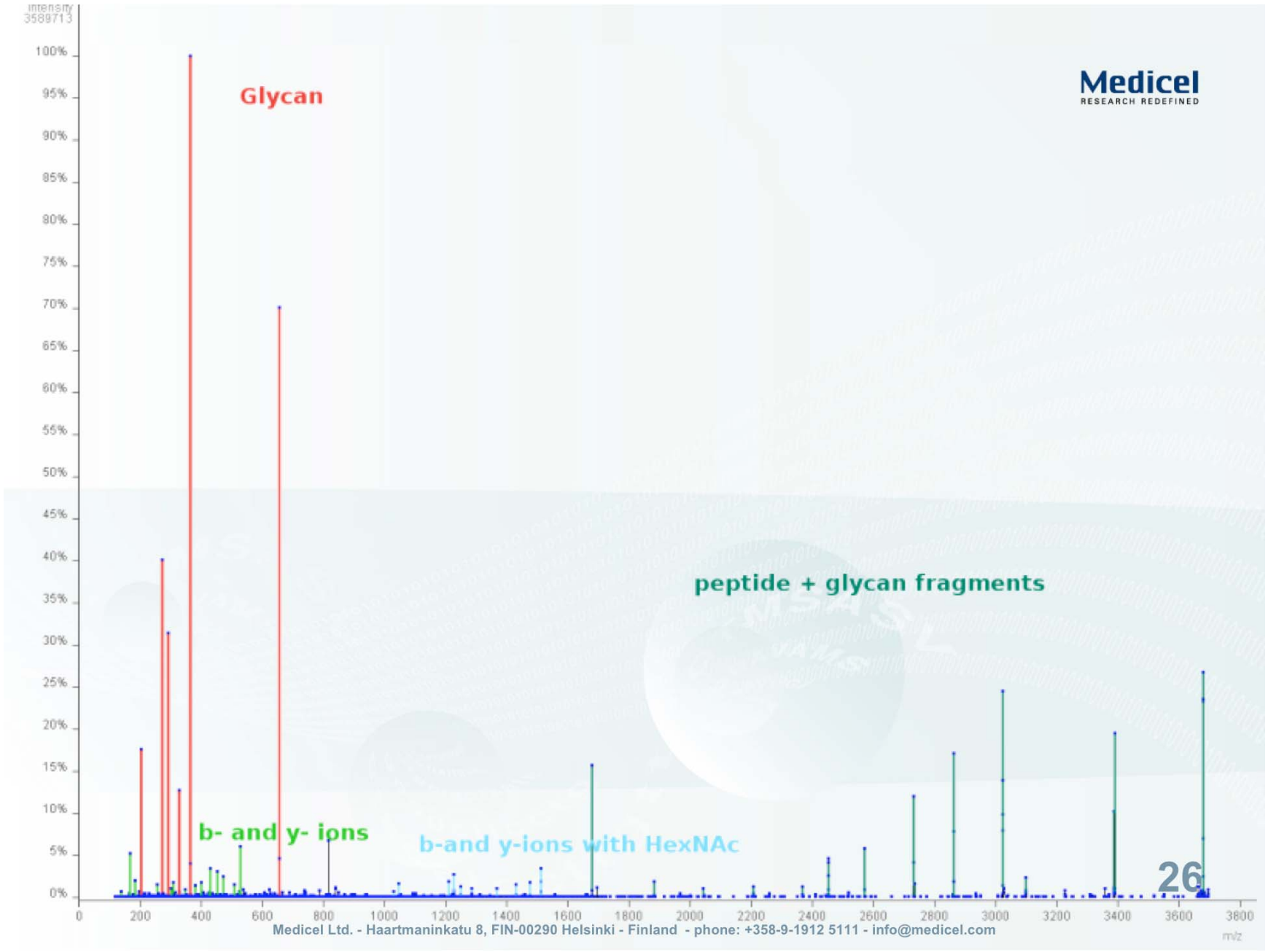
Medicel
RESEARCH REDEFINED

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com


Intensity
3589713
100%
95%
90%
85%
80%
75%
70%
65%
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
Glycan
Medicel
RESEARCH REDEFINED
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600
2800
3000
3200
3400
3600
3800
m/z
23
Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com


Intensity
3589713
100%
95%
90%
85%
80%
75%
70%
65%
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
Glycan
Medicel
RESEARCH REDEFINED
peptide + glycan fragments
24
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600
2800
3000
3200
3400
3600
3800
m/z
Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com


Intensity
3589713
100%
95%
90%
85%
80%
75%
70%
65%
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
Glycan
Medicel
RESEARCH REDEFINED
peptide + glycan fragments
b- and y- ions
25
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600
2800
3000
3200
3400
3600
3800
m/z
Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com


Intensity
3589713
100%
95%
90%
85%
80%
75%
70%
65%
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
Glycan
Medicel
RESEARCH REDEFINED
peptide + glycan fragments
b- and y- ions
b-and y-ions with HexNAc
26
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600
2800
3000
3200
3400
3600
3800
m/z
Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

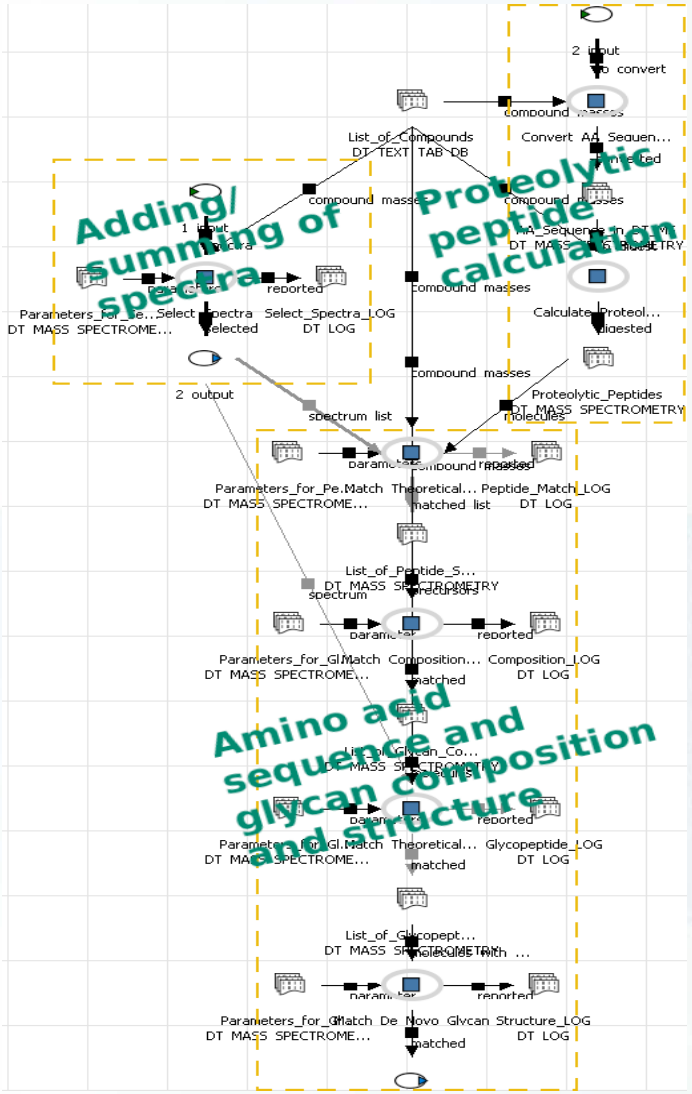
Medicel
RESEARCH REDEFINED

Protein N-glycopeptide Analysis

- **N-Glycopeptide**
  - Amino Acid Sequence Analysis
  - Glycan Composition
  - Glycan Structure
  - Glycolysation Site

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Workflow overview



N-glycopeptide analysis workflow

1) **Proteolytic N-glycopeptide database generated**
2) **Glycopeptide mass spectra added**
3) **Amino acid sequences are searched from the spectra against our N-glycopeptide database**
4) **Compositions of a N-glycopeptide determined**
5) **Putative amino acid sequences together with the glycan compositions are fitted to the spectra**

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Medicel
RESEARCH REDEFINED

Amino acid sequence determination

I. Proteolytic peptides are calculated
II. N-Glycopeptide database generated
    I. Peptide has to have N-X-S / T / C, X not a P sequence
III. N-Glycopeptide database *in silico* fragmented
IV. Empirical spectra imported
V. Matching of the empirical CID spectra to the N-glycopeptide database
    I. Peptide+HexNAc peak required
        I. The amide bond between peptide and glycan is stronger compared to the following glycosidic linkages
    II. y- and b- ion series matched
VI. True positive rate calculated against inverted database (target -decoy)

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

N-Glycopeptide database

I. FASTA sequences of given species retrieved from the database

II. *In silico* proteolytic cleavage (0, 1, or 2 misscleavages)

III. Potential glycopeptide sequences are filtered with the N-glycosylation consensus sequence N-X-S / T / C, X not a proline

	Misscleavages		
	0	1	2
All human proteins	1.2×10^6	5×10^6	10×10^6
N-Glycopeptide library	78 362	229 209	441 772

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Plasma: Amino acid sequence of a glycopeptide (m/z 1194, +4)

Peptide Score	Δm to glycopeptide mass	Protein	Peptide sequence
13.14	2860.79	A1AG_HUMAN	QDQCIYNTTYLNVQR
5.25	494.51	PPL2_HUMAN	YHCPVLFTVFTNNTHIVAVRTTGN VYAYEAVEQLNIK
5.25	494.51	PPIL2_HUMAN_ISOFORM1/02752949	YHCPVLFTVFTNNTHIVAVRTTGN VYAYEAVEQLNIK
5.25	494.51	PPIL2_HUMAN_ISOFORM2/02755796	YHCPVLFTVFTNNTHIVAVRTTGN VYAYEAVEQLNIK
3.51	2657.58	HI14_HUMAN	QPDKENVTLLHWAAINNR
3.51	2657.58	ZDH17_HUMAN_ISOFORM1/02767848	QPDKENVTLLHWAAINNR
3.51	2657.58	ZDH17_HUMAN_ISOFORM2/02762647	QPDKENVTLLHWAAINNR
3.51	2657.58	ZDH17_HUMAN_ISOFORM3/02760815	QPDKENVTLLHWAAINNR

spectrum 437
Number of molecules matched:1747

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

The best scoring peptide - QDQCIYNTTYLNVQR

Δm	Ion
-0.01	y2
-0.01	b3
-0.01	y3
-0.01	y4
-0.02	b4
-0.01	y5
-0.01	b5
-0.01	y6
-0.04	y8
-0.05	peptide

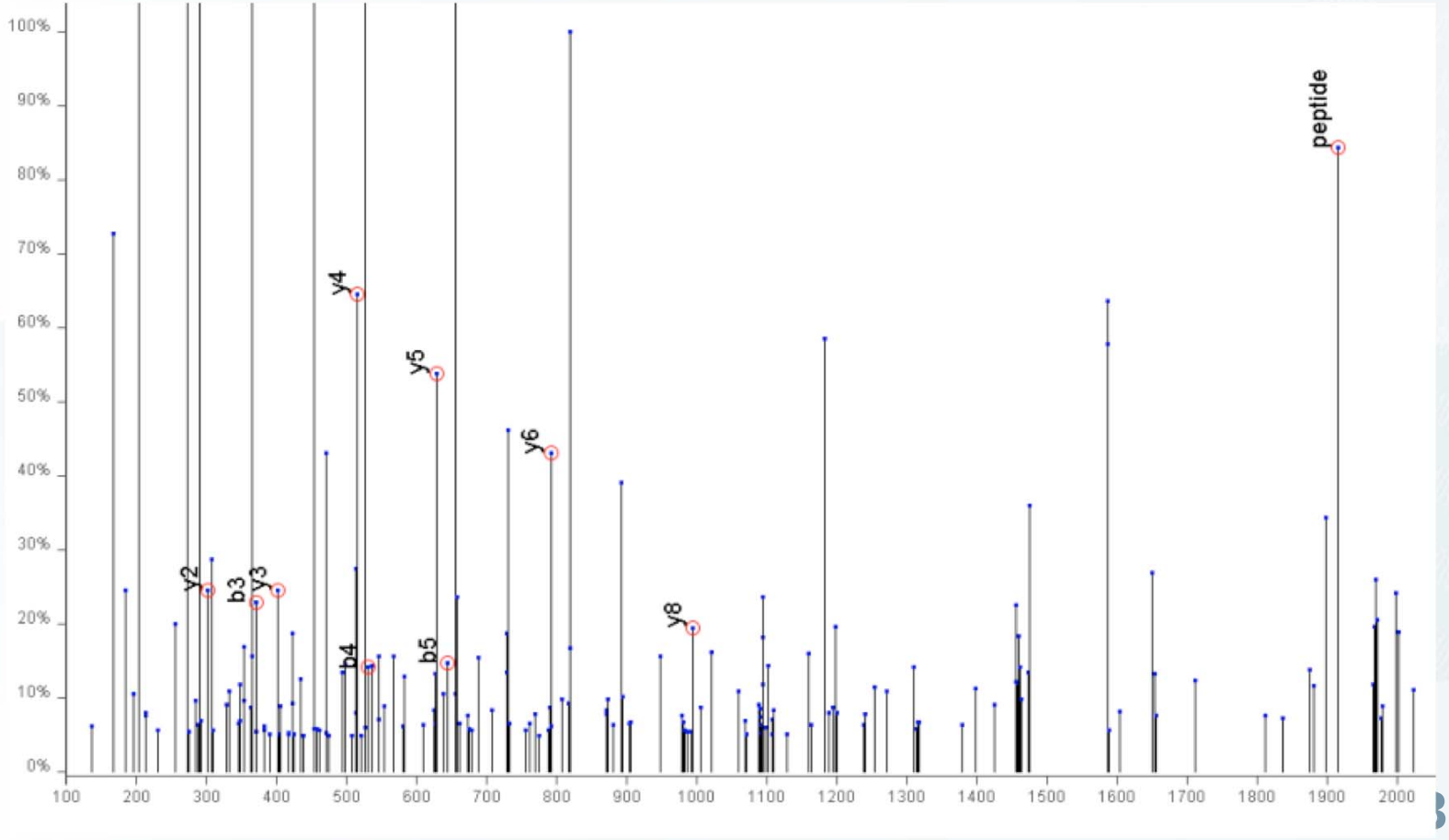
n number of potential sequences are taken to the following step

typically 20 – 50 best matching sequences

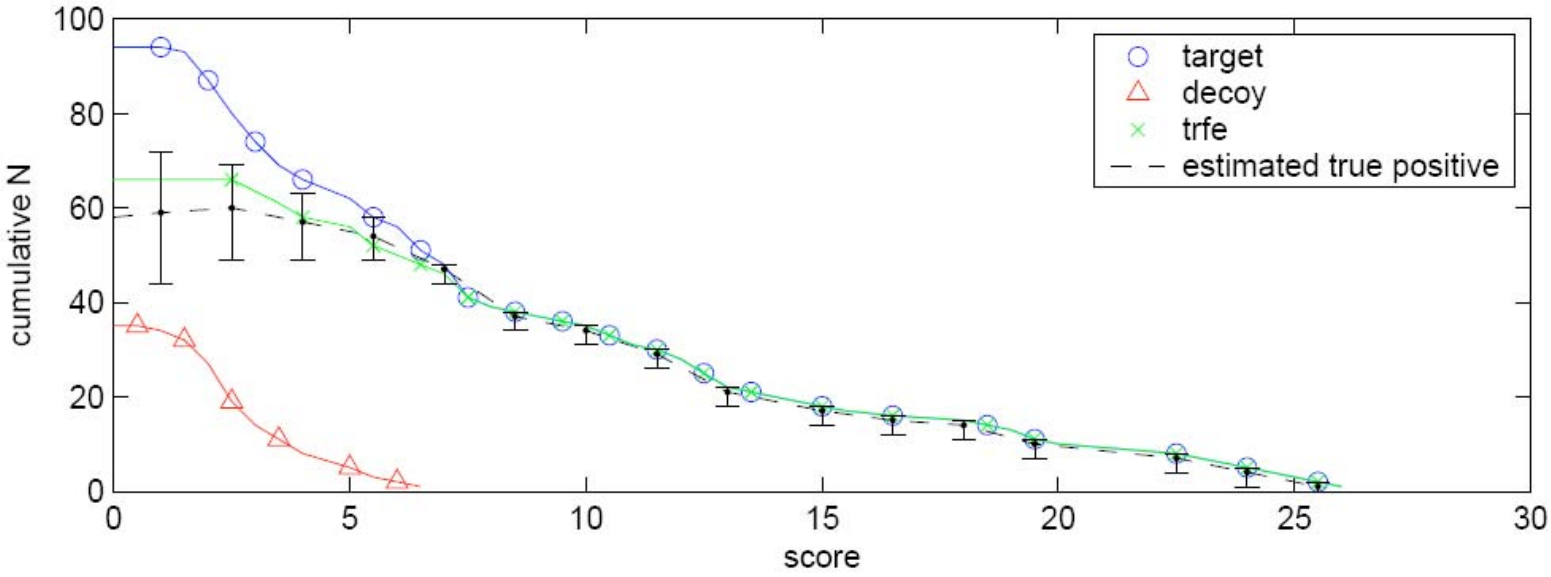
Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Medicel
RESEARCH REDEFINED

The best scoring peptide - QDQCIYNTTYLNVQR



Target – decoy database approach



Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Medicel
RESEARCH REDEFINED

N-Glycan composition determination

I. Peptide mass is calculated from previously identified aa sequences
II. Potential glycan masses are calculated from peptide and total experimental glycopeptide precursor mass
III. Glycan components are given (Hex, HexNAc, deoxyHex, NeuAc...)
IV. Rules used for N-glycan composition analysis (not necessary)
    I. N-Core is required [HexNAc]2 [Hex]3
    II. The number of deoxyhexoses plus 1 must be less than or equal to the sum of the number of hexoses and N-acetylhexosamines (GlycoMod)
    III. If there are no N-acetylhexosamines except in N-Glycan core, then the number of sialic acids is zero (GlycoMod)
V. All glycan compositions are calculated
VI. Order by the smallest Δm

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Medicel
RESEARCH REDEFINED

Plasma: N-Glycan composition determination (m/z 1194, +4)

Peptide Score	Δm	Protein	Sequence	Glycan composition
13.14	-0.16	A1AG_HUMAN	QDQCIYNTTYLNVQR	13 Hex 2 HexNAc 2 DeoxyHex
13.14	-0.21	A1AG_HUMAN	QDQCIYNTTYLNVQR	6 Hex 5 HexNAc 3 NeuAc
3.51	-0.34	HI14_HUMAN	QPDKENVTLLHWAAINNR	6 Hex 4 HexNAc 3 NeuAc
3.51	-0.34	ZDH17_HUMAN_ISOFORM1/02767848	QPDKENVTLLHWAAINNR	6 Hex 4 HexNAc 3 NeuAc
3.51	-0.34	ZDH17_HUMAN_ISOFORM2/02762647	QPDKENVTLLHWAAINNR	6 Hex 4 HexNAc 3 NeuAc
3.51	-0.34	ZDH17_HUMAN_ISOFORM3/02760815	QPDKENVTLLHWAAINNR	6 Hex 4 HexNAc 3 NeuAc
3.51	0.69	HI14_HUMAN	QPDKENVTLLHWAAINNR	6 Hex 3 HexNAc 3 NeuAc 1 DeoxyHex
3.51	0.69	ZDH17_HUMAN_ISOFORM1/02767848	QPDKENVTLLHWAAINNR	6 Hex 3 HexNAc 3 NeuAc 1 DeoxyHex
3.51	0.69	ZDH17_HUMAN_ISOFORM2/02762647	QPDKENVTLLHWAAINNR	6 Hex 3 HexNAc 3 NeuAc 1 DeoxyHex
3.51	0.69	ZDH17_HUMAN_ISOFORM3/02760815	QPDKENVTLLHWAAINNR	6 Hex 3 HexNAc 3 NeuAc 1 DeoxyHex

All aa sequences and their respective glycan compositions are taken to the next step

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Medicel
RESEARCH REDEFINED

Matching proposed N-Glycopeptides to spectra

I. Glycosidic linkages theoretically fragmented from putative N-Glycan compositions
II. Glycan fragments attached to intact peptide are matched to the spectra
    I. N-Core assumed
    II. NeuAc cannot be attached to the N-Core branching hexose
    III. NeuAc and deoxyHex cannot be linked to each other
III. Score to a theoretical glycopeptide is calculated as a negative logarithm of a probability that a random set of fragments would have as many or more shared peaks with measured spectrum as the ranked glycopeptide.
IV. The probability that the random spectrum have more or equal shared peaks than the glycopeptide spectrum is calculated using binomial distribution.

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Plasma: Matching proposed N-Glycopeptides to spectra (m/z 1194, +4)

Score: peptide + comp. fit	Δm	Protein	Sequence	Glycan composition
50.1	-0.21	A1AG_HUMAN	QDQCIYNTTYLNVQR	6 Hex 5 HexNAc 3 NeuAc
32.01	-0.34	HI14_HUMAN	QPDKENVTLLHWAAINNR	6 Hex 4 HexNAc 3 NeuAc
32.01	-0.34	ZDH17_HUMAN_ISOFORM1/02767848	QPDKENVTLLHWAAINNR	6 Hex 4 HexNAc 3 NeuAc
32.01	-0.34	ZDH17_HUMAN_ISOFORM2/02762647	QPDKENVTLLHWAAINNR	6 Hex 4 HexNAc 3 NeuAc
32.01	-0.34	ZDH17_HUMAN_ISOFORM3/02760815	QPDKENVTLLHWAAINNR	6 Hex 4 HexNAc 3 NeuAc
23.63	-0.16	A1AG_HUMAN	QDQCIYNTTYLNVQR	13 Hex 2 HexNAc 2 DeoxyHex
21.31	0.69	HI14_HUMAN	QPDKENVTLLHWAAINNR	6 Hex 3 HexNAc 3 NeuAc 1 DeoxyHex
21.31	0.69	ZDH17_HUMAN_ISOFORM1/02767848	QPDKENVTLLHWAAINNR	6 Hex 3 HexNAc 3 NeuAc 1 DeoxyHex
21.31	0.69	ZDH17_HUMAN_ISOFORM2/02762647	QPDKENVTLLHWAAINNR	6 Hex 3 HexNAc 3 NeuAc 1 DeoxyHex
21.31	0.69	ZDH17_HUMAN_ISOFORM3/02760815	QPDKENVTLLHWAAINNR	6 Hex 3 HexNAc 3 NeuAc 1 DeoxyHex

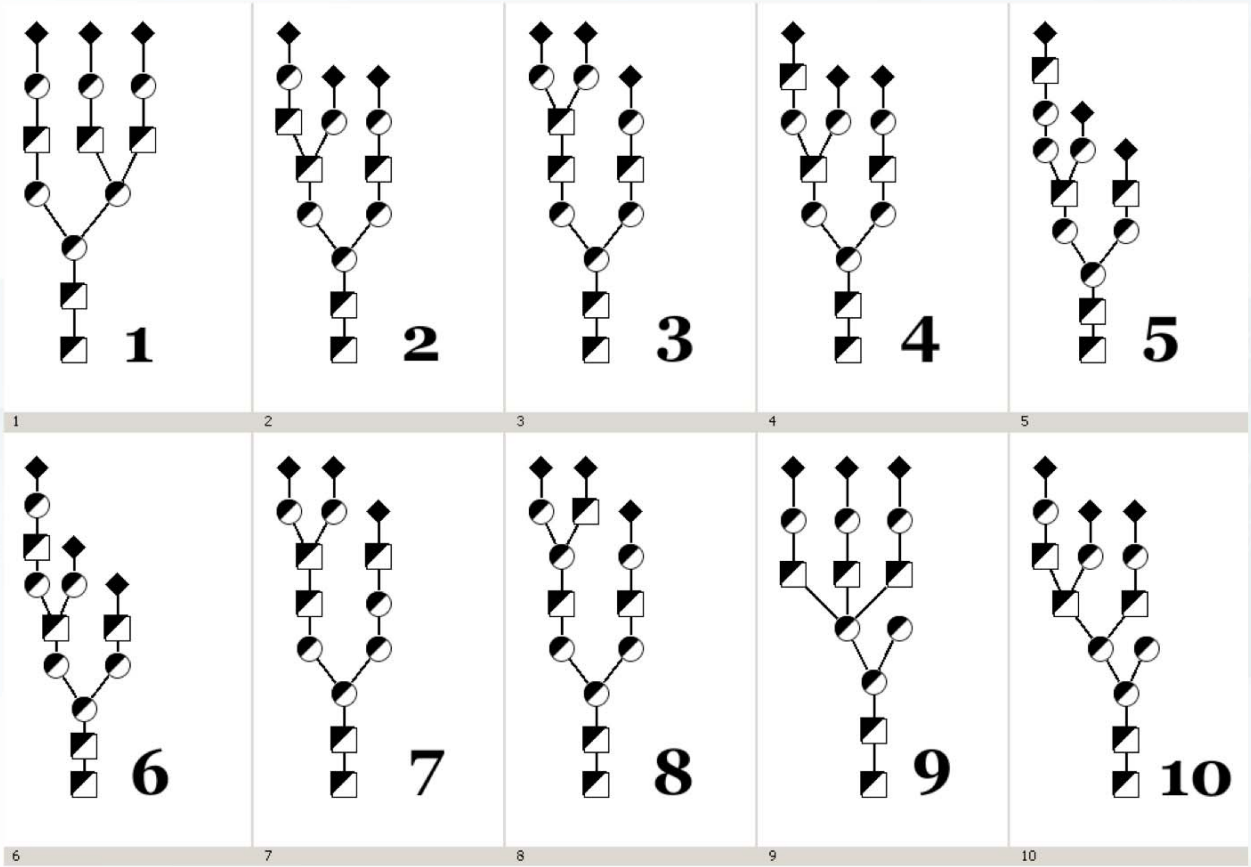
Medicel
RESEARCH REDEFINED

Building glycan structures to glycopeptides

I. *De novo* glycan structures, which have minimal difference with measured spectrum were searched with *Branch and Bound* type algorithm
II. A list of possible glycan compositions and matched fragments with measured spectrum are given as input.
III. Search is started from a given glycan core structure, new carbohydrate residues are added iteratively and a population of glycan structures is generated
IV. Process is continued until the structures have the target composition

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

Plasma: Building glycan structures to glycopeptides
m/z 1194, +4
A1AG_HUMAN, QDQCIYNTTYLNVQR



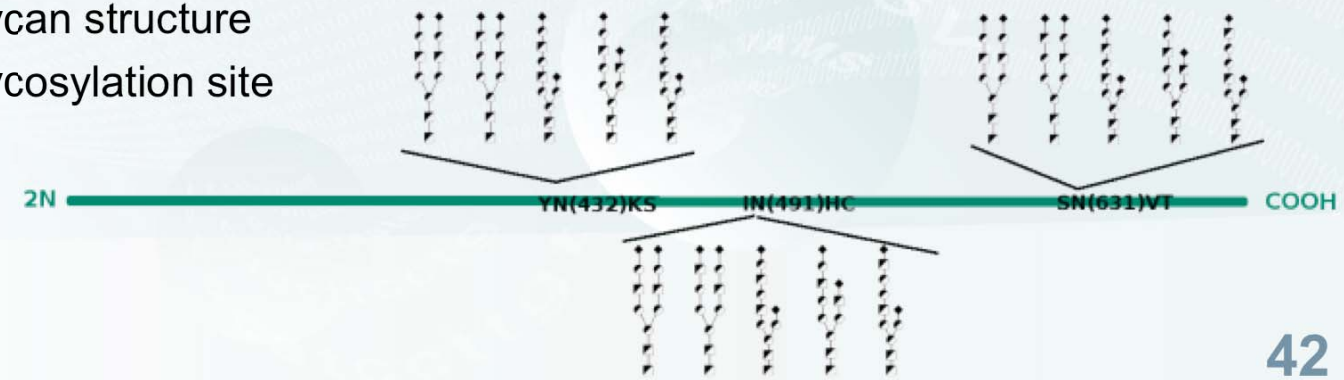
Medicel Ltd. - Keilaranta 12, FIN-02150 Espoo - Finland - phone: +358-9-386-7092 - info@medicel.com

Part 4 – Summary of N-glycopeptide analysis

Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

N-Glycosylation Analysis

I. Medicel Integrator
    I. Integrated database: define one or more species
        I. (Uniprot (SWISS-Prot and Trembl), ENSEMBLE and NCBI refseq)
II. Template workflow library
    I. N-glycopeptide database
    II. Amino acid sequence
    III. Glycan composition
    IV. Glycan structure
    V. Glycosylation site



Medicel Ltd. - Haartmaninkatu 8, FIN-00290 Helsinki - Finland - phone: +358-9-1912 5111 - info@medicel.com

References

CFG nature functionalglycomicsgateway

SEARCH go

Research Highlights

N-glycomics: Of apicomplexa and algorithms

Functional Glycomics (13 March 2008)

GLYCOBIOLOGY

Sakari Joenväärä, Ilja Ritamo, Hannu Peltoniemi, and Risto Renkonen

N-Glycoproteomics – an automated workflow approach

Glycobiology Advance Access published on February 13, 2008. OPEN ACCESS ARTICLE