

MĚŘENÍ A INTERPRETACE SPEKTER BIOMOLEKUL

Miloslav Šanda

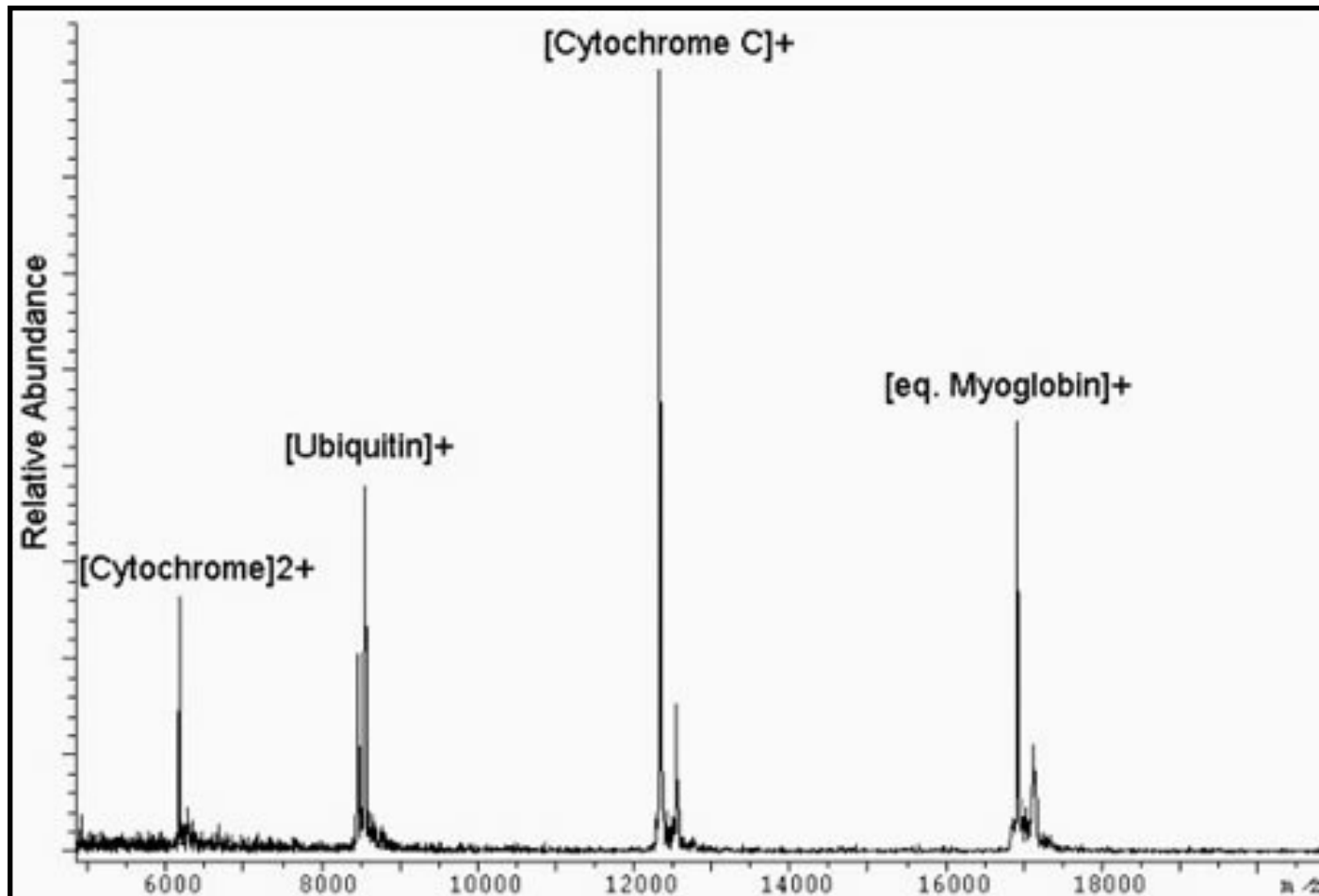
Ionizační techniky využívané k analýze biomolekul (biopolymerů)

- MALDI : proteiny, peptidy, oligonukleotidy, sacharidy ...
- ESI : proteiny, peptidy, oligonukleotidy, sacharidy ...
- APCI: syntetické polymery, nepolární analyty jako triacylglyceroly

MALDI intaktních proteinů

- Tvorba 1x nabitých iontů molekulárních aduktů
- V minoritním množství tvorba 2x nabitých iontů a tvorba dimerů
- Při prosté MALDI-TOF analýze nejsou zpravidla patrné fragmentační ionty
- Používané matrice – kyselina sinapová, v menší míře α -hydroxyskořicová kyselina a DHB

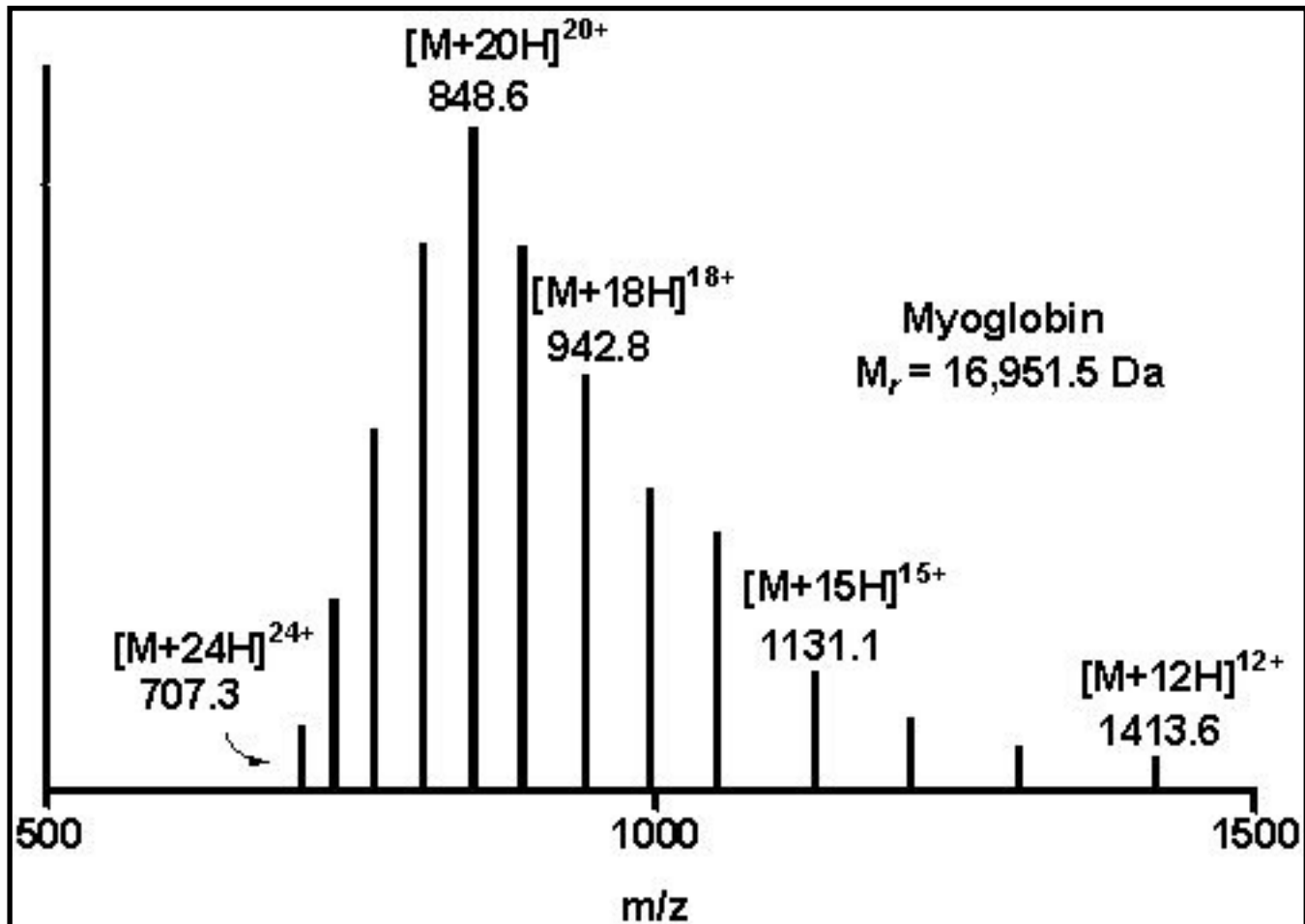
MALDI/MS spektrum proteinů



ESI/MS intaktních proteinů

- Tvorba vícenásobně nabitých iontů – možná analýza velkých proteinů spektrometry s nižším rozsahem hmot
- Při použití analyzátoru s vyšším rozlišením možnost stanovení MW proteinu s přesností $Da - 0,1Da$ v závislosti na rozlišení instrumentu

ESI/MS intaktného proteínu



Výpočet molekulové hmotnosti proteinu z m/z sousedních píků

- M – MW proteinu, z – nábojový stav
- Y – pík s nižší m/z , X – pík s vyšší m/z
- $X = (M+z)/z$, $Y = (M+z+1)/(z+1)$
- nábojový stav píku X $z_X = (Y-1)/(X-Y)$
- $MW = (X * z_X) - z_X = (Y * z_Y) - z_Y$

MALDI-MS peptidů

- Tvorba 1x nabitých iontů
- V případě zasolení vzorků možná tvorba aduktů se složkami solí
- Vhodné matrice – α -hydroxyskořicová kyselina, DHB
- Při volbě vysoké intensity laseru prolínání a potlačení nižších hmot ionty matrice

Využití MALDI/MS peptidů

- V analýze molekulové hmotnosti peptidu
- Identifikace proteinu po naštěpení pomocí specifického enzymu (peptide mass fingerprinting) – srovnání naměřeného spektra s databází aminokyselinových sekvencí proteinů a počítačově vytvořených teoretických digestů (směsí teoretických peptidů)

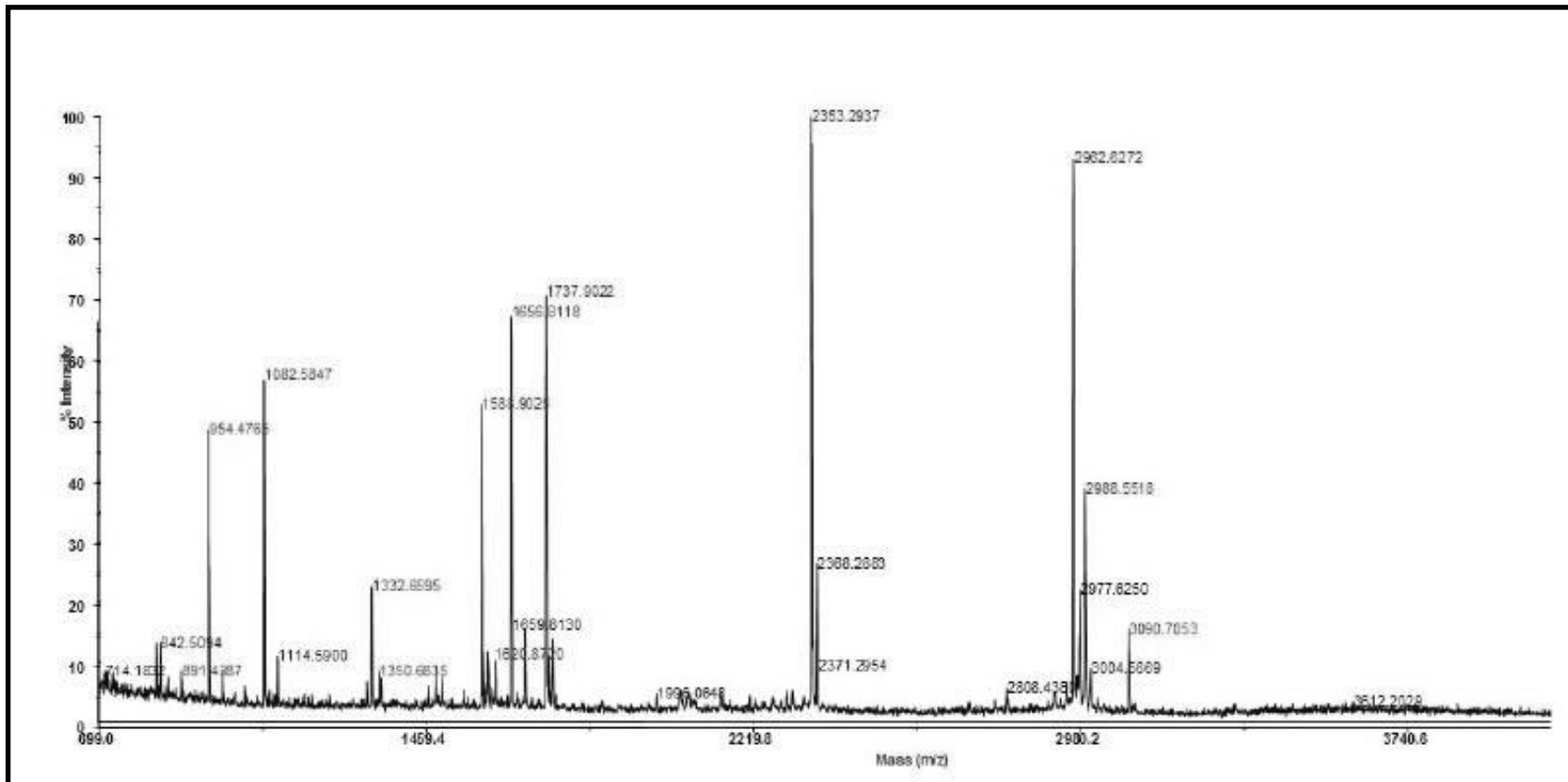
Používané specifické štěpení

Enzyme and chemical cleavage reagents	Cleavage sites, comments
trypsin	C-terminal to R and K (except R-P, K-P bond); R-K, K-K, R-R, K-R cleave slower
endoproteinase Lys-C	C-terminal to K (rarely K-P bond)
endoproteinase Arg-C	C-terminal to R (except R-P bond)
endoproteinase Glu-C (V8 protease)	C-terminal to E and D (except C-P, D-P bond)
chymotrypsin	C-terminal to F, Y, W, L, I, V, M (except X-P bond)
elastase	C-terminal to G, A, S, V, L, I (not very specific)
pepsin	C-terminal to F, L, E (pH 2-4 active range)
Asp-N	N-terminal to D (sometimes E)
thermolysin	N-terminal to L, I, V, F (and others to a lesser extent)
carboxypeptidase A/Y	cleaves C-terminal residues
mild acid	cleaves D-P bond
cyanogen bromide (CNBr)	C-terminus of M; M-T and M-S cleave slower; oxidized M do not cleave
BNPS-skatole	cleaves W-X bond
PNGase F	cleaves N-linked (Asn-glyco) glycoproteins; leaves entire carbohydrate portion intact and converts N to D
alkaline phosphatase	dephosphorylation of phosphoproteins

Výhody a nevýhody MALDI/MS peptidů

- + relativně snadná interpretace spekter, jednoduchost experimentu a instrumentace
- - vysoké nároky na přesnost změřených spekter
- - problémy interpretace spekter při směsi proteinů nebo při identifikaci proteinů s možnou mutací nebo nedostatečně známou sekvencí

MALDI/MS digestu proteinu



Vyhledávací program MASCOT

The screenshot shows the MASCOT Peptide Mass Fingerprint search interface. The browser window title is "Matrix Science - Mascot - Peptide Mass Fingerprint - Windows Internet Explorer". The address bar shows the URL "http://www.matrixscience.com/cgi/search_form.pl?FORMER-S&SEARCH-PMF". The page header includes the Matrix Science logo and navigation links: HOME | WHAT'S NEW | MASCOT | HELP | PRODUCTS | SUPPORT | TRAINING | CONTACT. A search bar is located in the top right corner.

The main heading is "MASCOT Peptide Mass Fingerprint". Below it is a form with the following fields and options:

- Your name:** sands
- Email:** sands@uochb.cas.cz
- Search title:** protein 1
- Database:** MSDB
- Taxonomy:** Homo sapiens (human)
- Enzyme:** Trypsin
- Allow up to:** 1 missed cleavages
- Fixed modifications:** Acetyl (K), Acetyl (N-term), Acetyl (Protein N-term), Amidated (C-term), Amidated (Protein C-term)
- Variable modifications:** Amidated (C-term), Amidated (Protein C-term), Ammonia-loss (N-term C), Biotin (K), Biotin (N-term)
- Protein mass:** kDa
- Peptide tol. ±:** 100 ppm
- Mass values:** NH⁺, M_r, H-H
- Monoisotopic:** Average
- Data file:** Prohibit...
- Query:** 1234.67, 1452.62, 1558.14, 1596.32, 1788.69

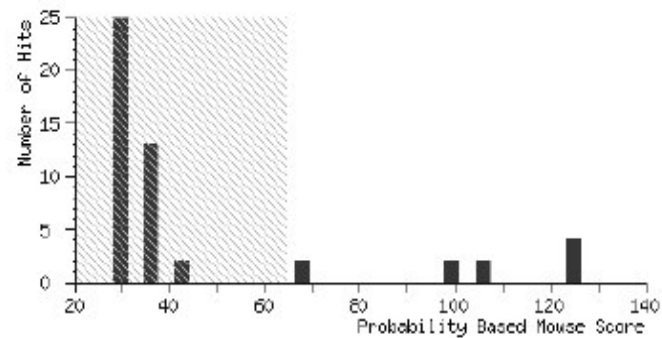
At the bottom of the browser window, the status bar shows "Internet" and "100%".

{MATRIX}
{SCIENCE} **Mascot Search Results**

User : ubik
 Email : ubik@seznam.cz
 Search title : EMG3_4_2
 Database : MSDB 20050929 (2344227 sequences; 779380795 residues)
 Taxonomy : Homo sapiens (human) (141910 sequences)
 Timestamp : 10 Apr 2006 at 05:22:44 GMT
 Top score : 125 for CLIC1_HUMAN, Chloride intracellular channel protein 1 (Nuc)

Probability Based Mowse Score

Protein score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event.
 Protein scores greater than 64 are significant ($p < 0.05$).



Protein Summary Report

Format As [Help](#)
 Significance threshold $p <$ Max. number of hits

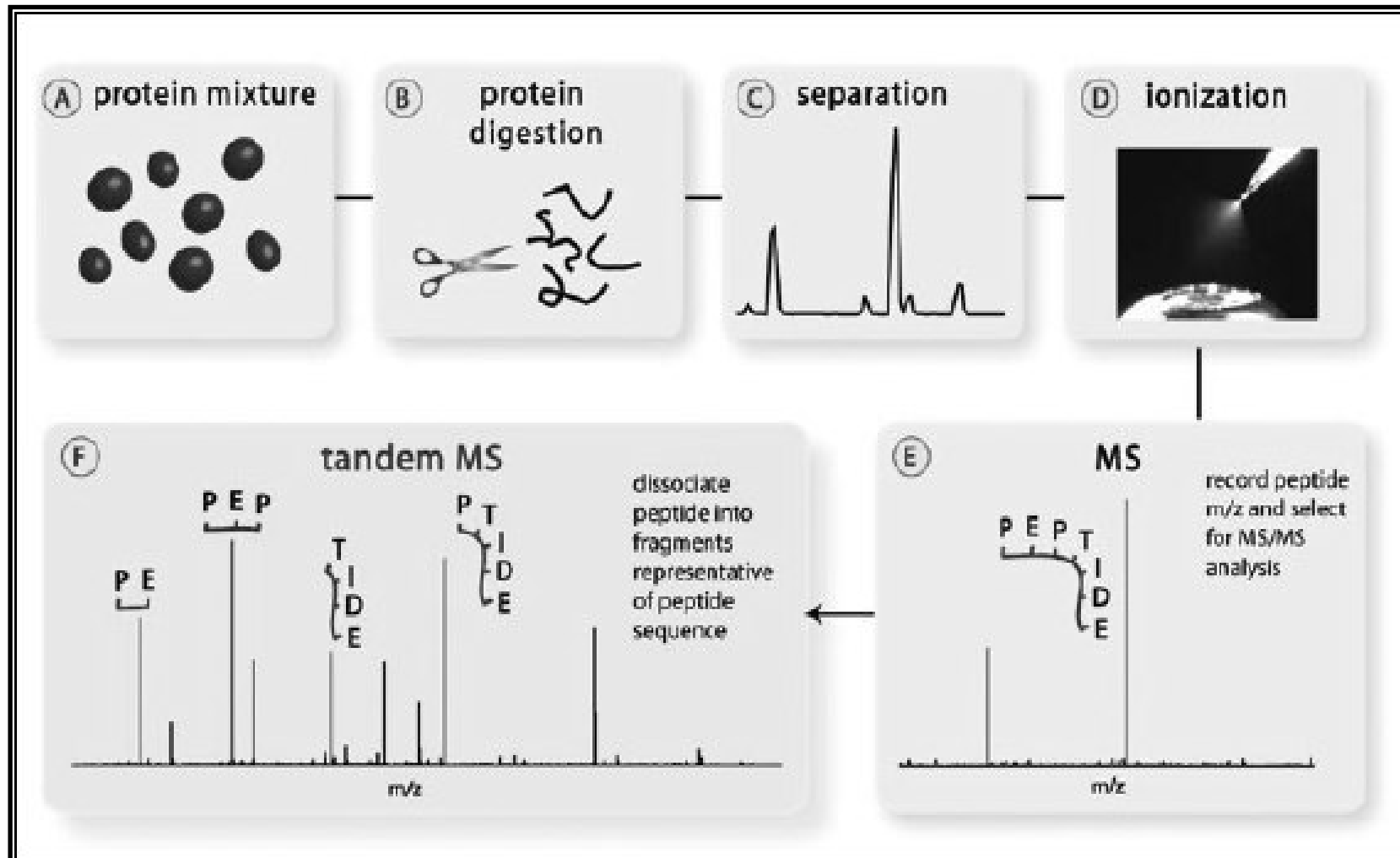
Overview Table

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

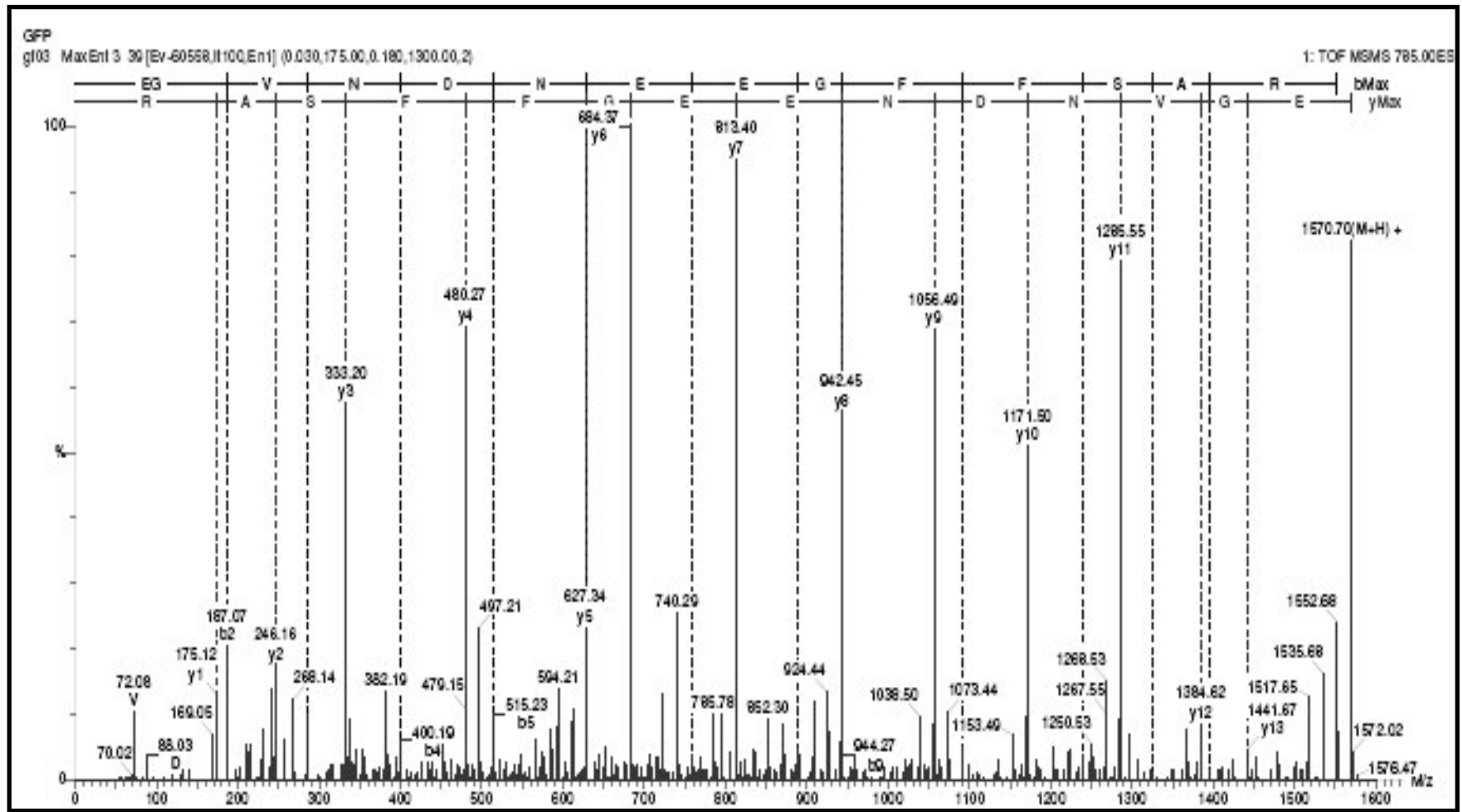
LC-MS/MS (MALDI/MS/MS) peptidů

- Potvrzení přesnější identifikace pomocí hmotnosti peptidu a fragmentačního spektra peptidu
- Možnost popisu sekvence peptidu z fragmentačního spektra (DE-NOVO sekvenování)
- Možnost analýzy digestů celých organel nebo částí organismů popřípadě směsí proteinů
- Možná analýza posttranslačních modifikací pomocí MS/MS

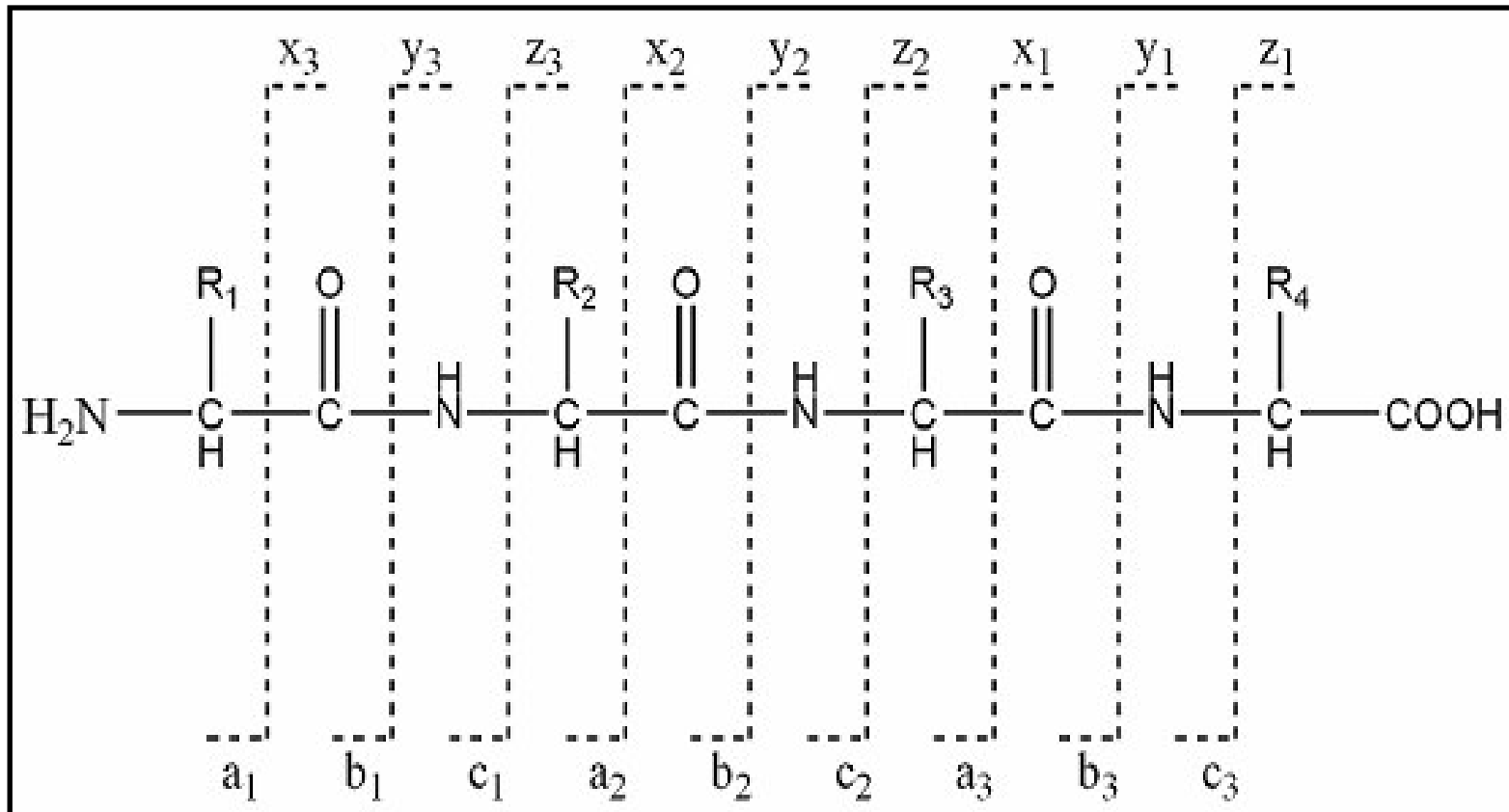
Postup identifikace proteinů pomocí ESI-MS/MS



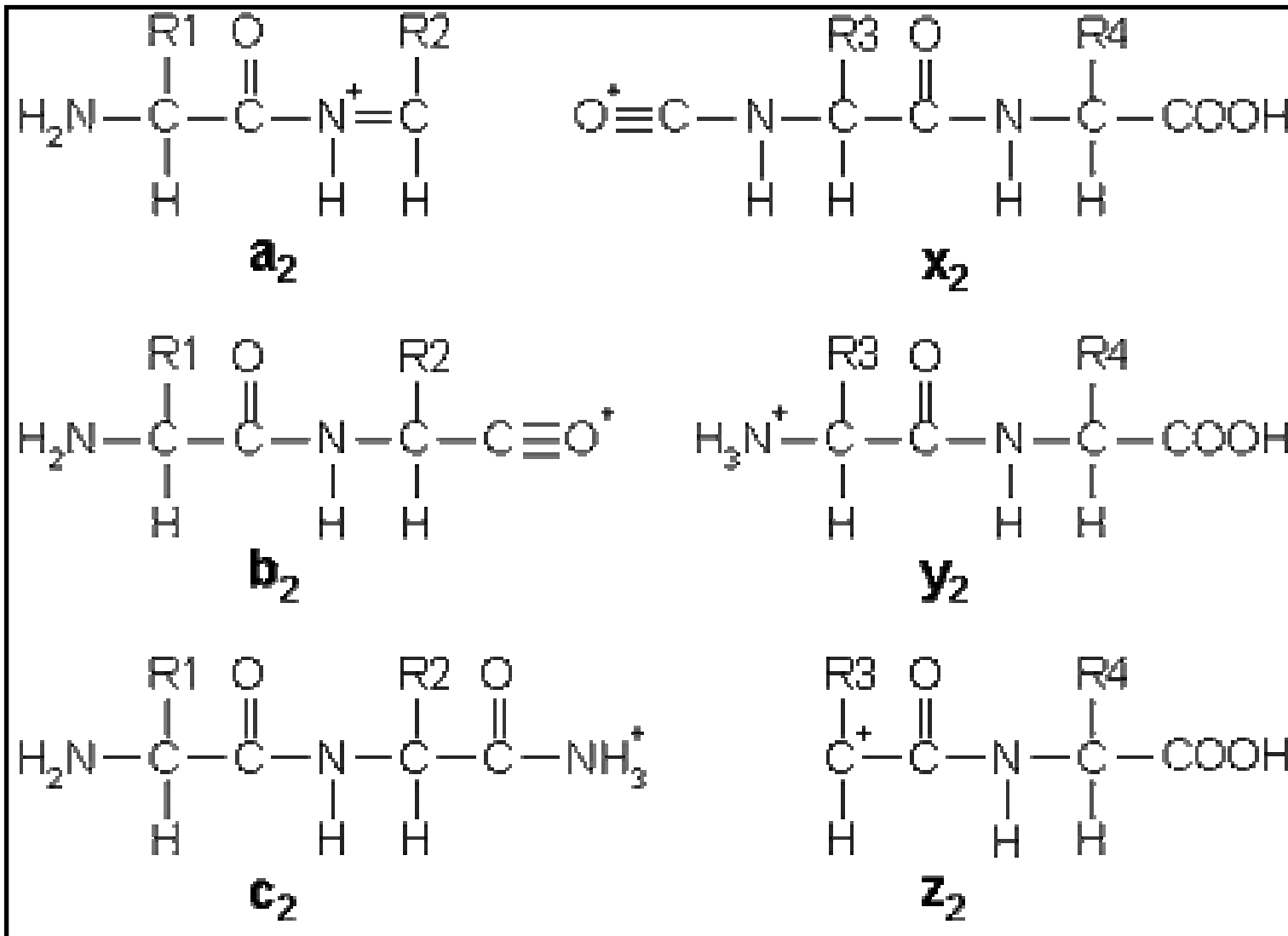
ESI-MS/MS spektrum peptidu



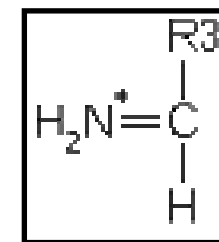
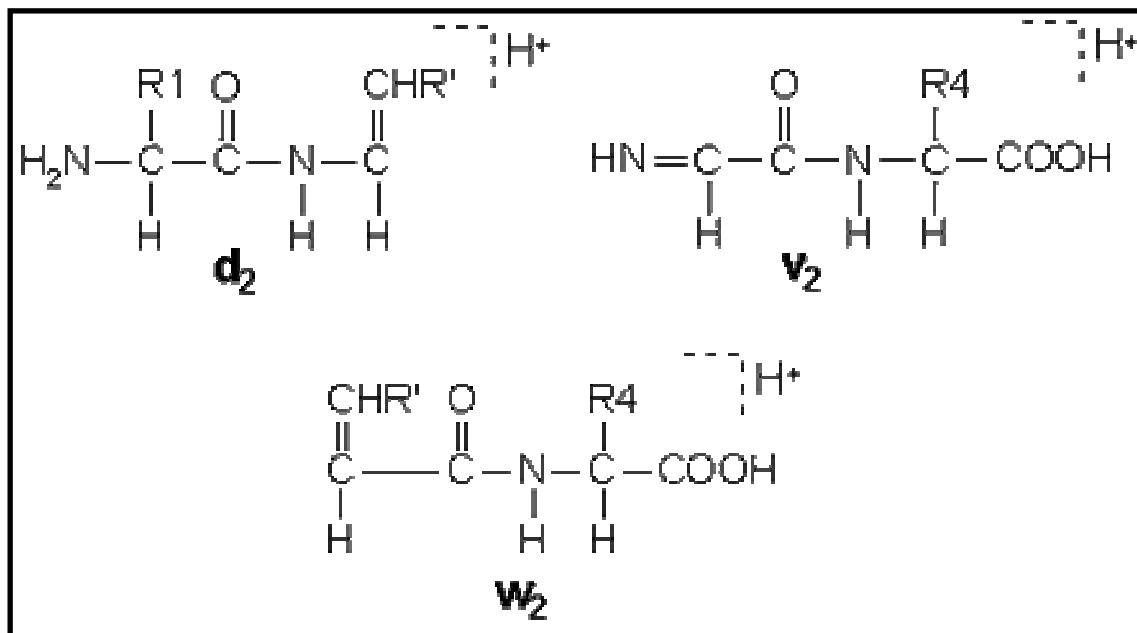
Fragmentace peptidu



Majoritní fragmenty peptidů



Další pozorované fragmenty



immoniový
iont

Hodnoty hmot příspěvků jednotlivých aminokyselin a jejich immoniových iontů

AA	Mass	Side chain	Immonium ions
G Gly	57.02	1	30
A Ala	71.08	15	44
S Ser	87.03	31	60
P Pro	97.05	41	70
V Val	99.07	43	72
T Thr	101.05	45	74
C Cys	103.01	47	76
L Leu	113.08	57	86(72)
I Ile	113.08	57	86(72)
N Asn	114.04	58	87(70)
D Asp	115.03	59	88
Q Gln	128.06	72	101(84, 129)
K Lys	128.09	72	101(129, 112, 84, 70)
E Glu	129.04	73	102
M Met	131.04	75	104(61)
H His	137.06	81	110(166, 138, 123, 121, 82)
F Phe	147.07	91	120(91)
R Arg	156.10	100	129(112, 100, 87, 73, 70, 59)
Y Tyr	163.06	107	136
W Trp	186.08	130	159

MS/MS peptidů

- Fragmentační spektrum pomocí CID obsahuje převážně y a b ionty a v menší míře potom ostatní fragmentační ionty
- Při analýze posttranslačních modifikací lze kromě neutrálních ztrát pozorovat i ionty modifikací
- Volba různých kolizních energií pro různě nabitě ionty peptidů

Instrumenty používané pro CID fragmentaci v proteomických laboratořích

- ESI-IT , ESI-LIT
- ESI-QQQ
- ESI-Q-TOF, ESI-Q-TRAP
- MALDI-Q-TOF
- MALDI-TOF/TOF
- Méně rozšířené jsou ESI/MALDI-FT-ICR, ESI-ORBITRAP

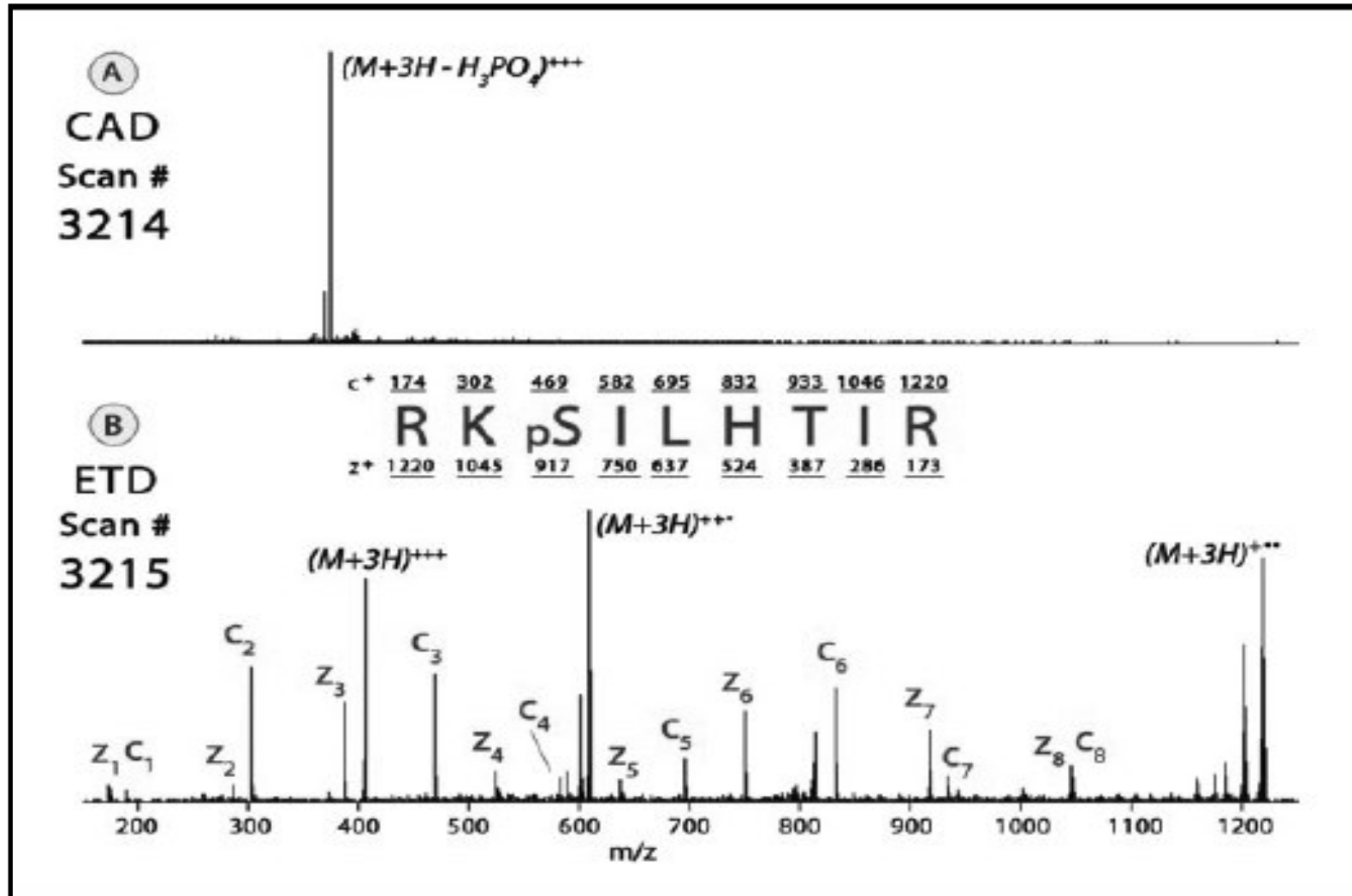
Nízkoenergetické fragmentace ECD,ETD

- Disociace záchytem elektronu ECD nebo přenosem elektronu ETD
- Při použití nízkoenergetické ECD nebo ETD fragmentace disociují převážně peptidové vazby za tvorby c a z iontů

MS/MS peptidů CID vs. ECD,ETD

- Při analýze posttranslačních modifikací v případě více modifikačních míst v jednom peptidu nelze pro zjištění místa modifikace použít CID fragmentaci (vysoká disociační energie, disociace peptidové vazby a současně vazby modifikace)
- V případě ETD, ECD vznikají jednodušší spektra obsahující pouze dva druhy iontů c a z

CID vs ECD, ETD



Některé proteinové databáze

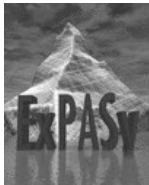
Table 2 Protein databases available on the internet

NCBI nr	A nonredundant database compiled by the NCBI by combining most of the public domain databases (ESTs not included).
Swiss Prot	A curated protein sequence database which strives to provide a high level of annotation, such as the description of the function of a protein, its domain's structure, post-translational modifications, variants, etc. This database offers a minimal level of redundancy and high level of integration with other databases.
OWL	A nonredundant composite of four publicly available primary sources: SWISSPROT, PIR, (1-3), GenBank (translation) and NRL-3D. SWISSPROT is the highest priority source, all others being compared against it to eliminate identical and trivially different sequences.
Genpept	Protein translation of Genbank (ESTs not included).
Unknome	A theoretical database used in de novo MS/MS spectral interpretation that is created on-the-fly and contains all amino acid sequence permutations consistent with the parent mass and amino acid composition information contained in an MS/MS spectrum.

NCBI, National Center for Biotechnology Information; EST, expressed sequence tag; MS/MS, tandem mass spectrometry (2nd series).

Programy pro zpracování proteomických dat a vyhledávání v databázích

- MASCOT (Matrixscience)
- Proteinlynx (Waters)
- Protein prospektor (UC SF)
- a mnoho dalších k nalezení například na:



(<http://www.expasy.org/tools/>)

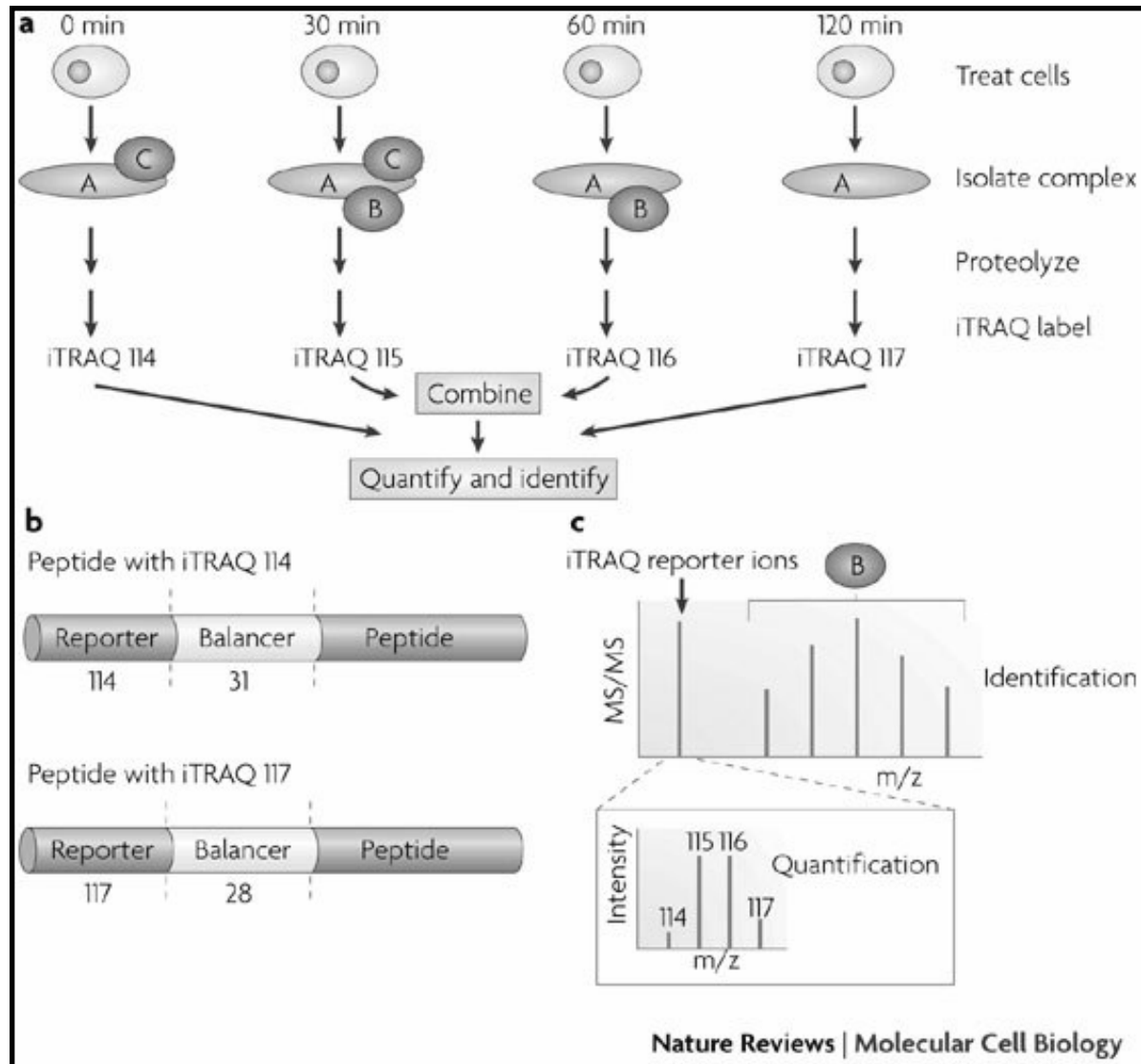
TOP-DOWN Proteomika

- Analýza na úrovni intaktních proteinů
- Fragmentace CID, ETD, ECD
- Neztrácíme některé informace o které můžeme přijít při chemickém (enzymatickém) zpracování proteinu
- Nutnost náročnější MS instrumentace, než u analýzy digestů

Kvantifikace změn koncentrace proteinů

- A, absolutní kvantifikace pomocí isotopově značeného peptidu (AQUA ...)
- B, pomocí dalších technik se kvantifikují změny koncentrací proteinů pomocí isotopických značek (ICAT, ITRAQ ..) nebo isotopicky značených aminokyselin (SILAC...), popřípadě značení při digestu (voda – O^{18})
- V poslední době se stále více uplatňuje kvantifikace bez nutnosti značení

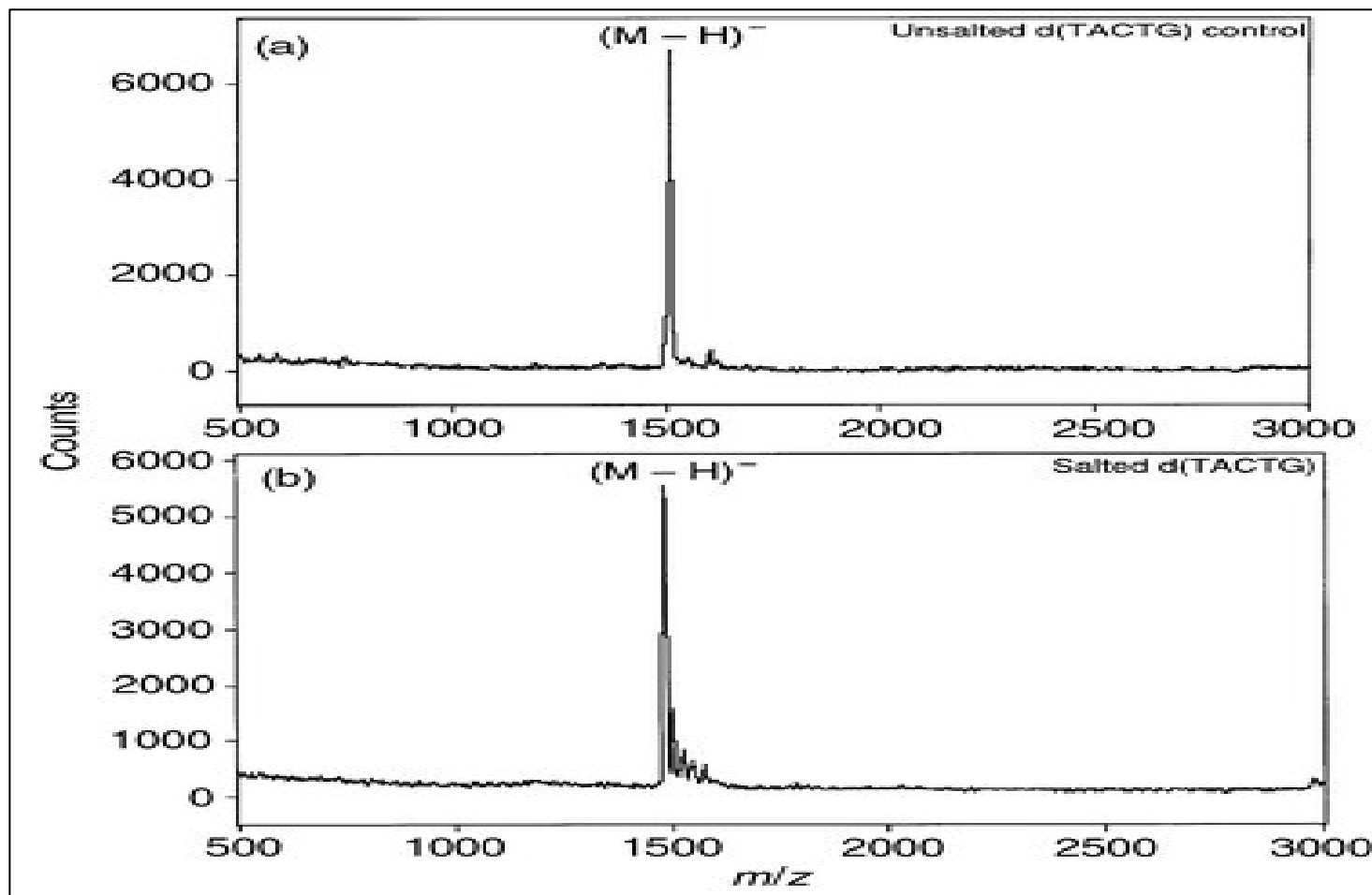
iTRAQ metoda



Stanovení oligonukleotidů pomocí hmotnostní spektrometrie

- MALDI/MS
- Nejlepší výtěžnost ionizace v negativním módu
- Obvyklé matrice PA, 3-HPA, 2,3,6-trihydroxyacetophenon
- Mnohem menší výtěžnost ionizace oproti peptidům
- Spektra jsou komplikovaná tvorbou fragmentů
- Nutné dostatečné odsolení, velmi náchylné k tvorbě aduktů s ionty alkalických kovů

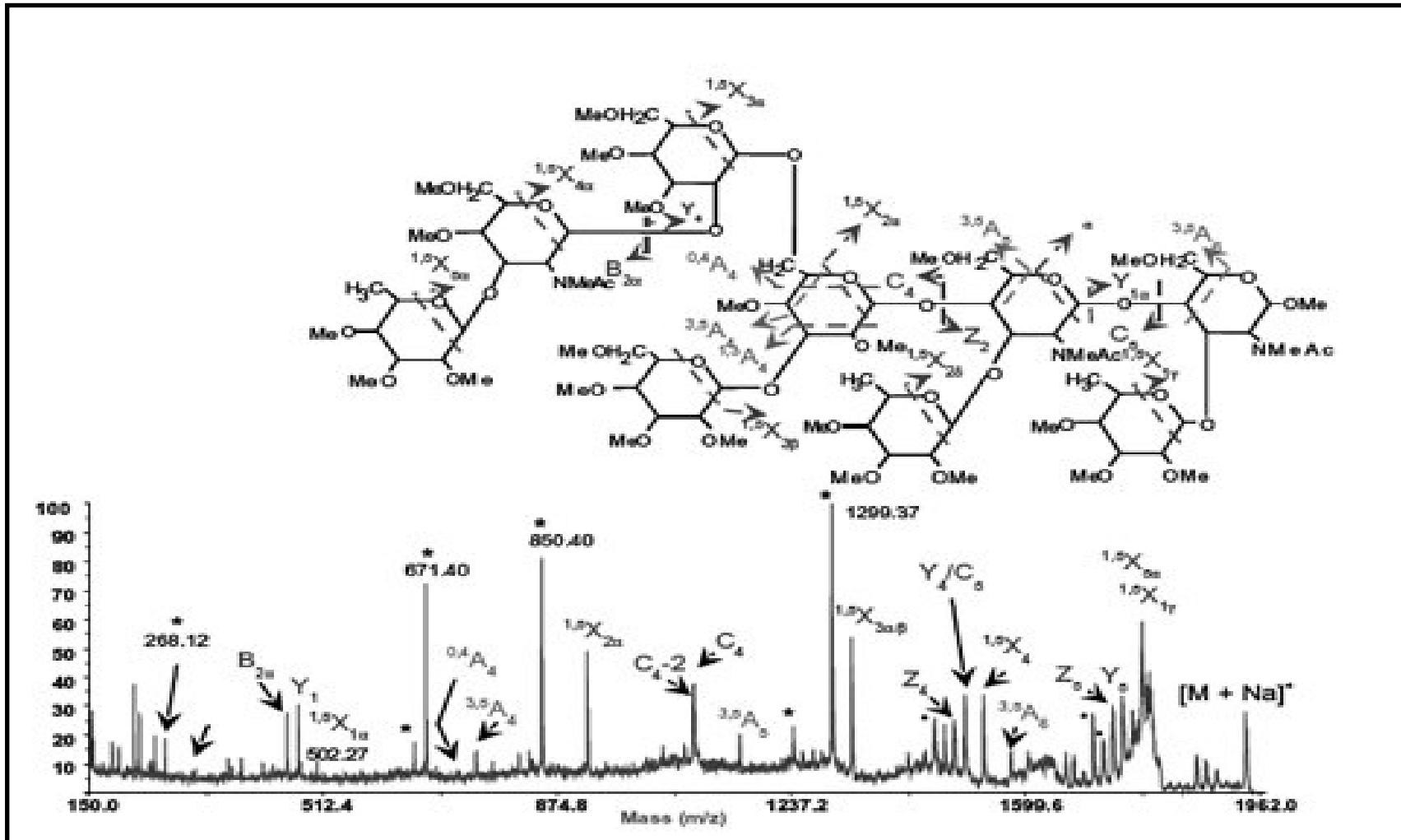
Vliv solí na stanovení oligonukleotidů pomocí MALDI/MS



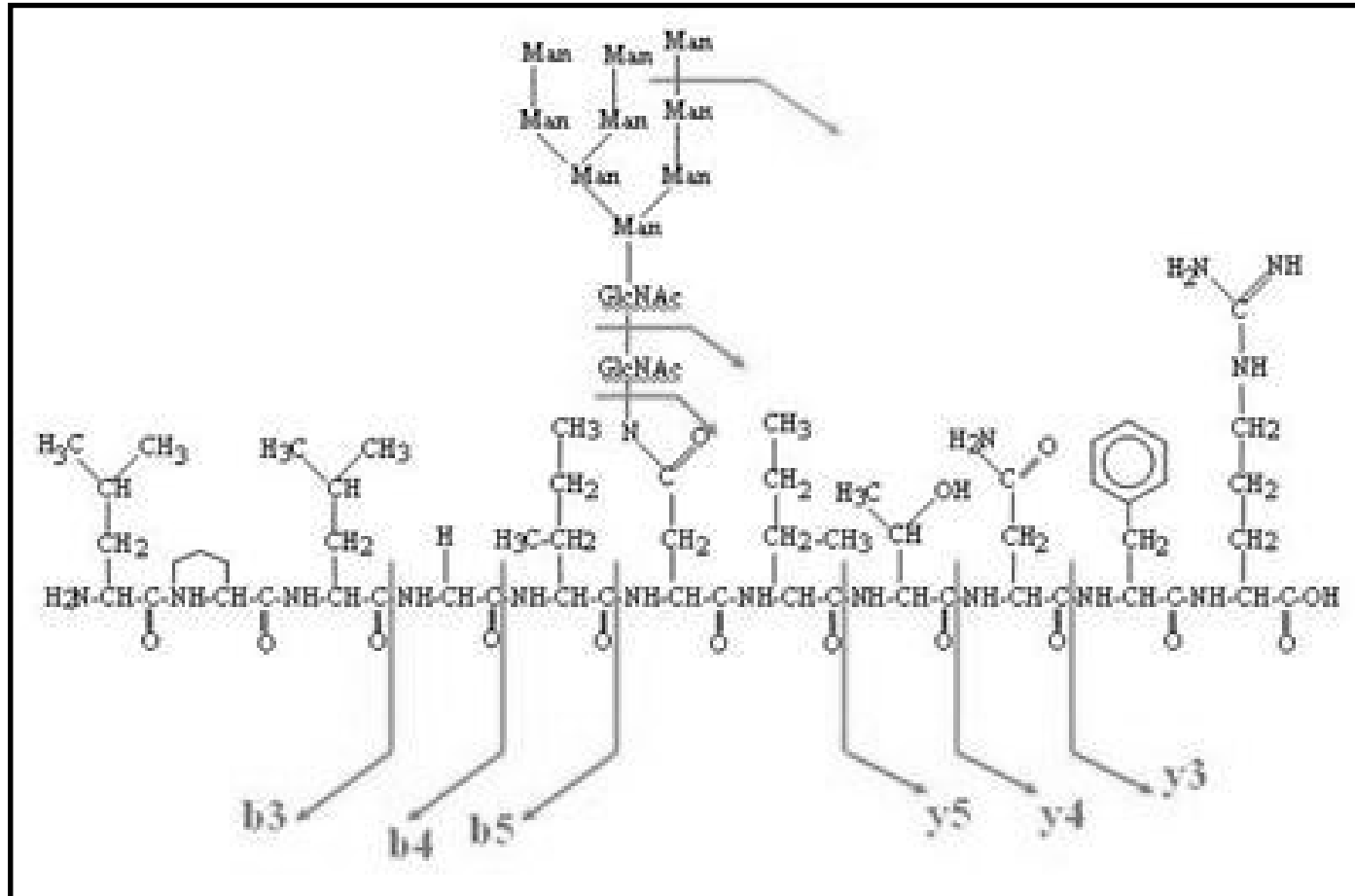
Stanovení sacharidů pomocí hmotnostní spektrometrie

- Používají se především měkké ionizační techniky pro stanovení sekvence a struktury (ESI, MALDI) ve spojení s MSⁿ
- Stanovení polysacharidů je mnohem složitější než stanovení proteinů, peptidů nebo nukleových kyselin díky isomerické povaze základních stavebních kamenů a jejich možnému větvení

Fragmetace glykopeptidu



Fragmentace glykopeptidu



Děkuji za pozornost

