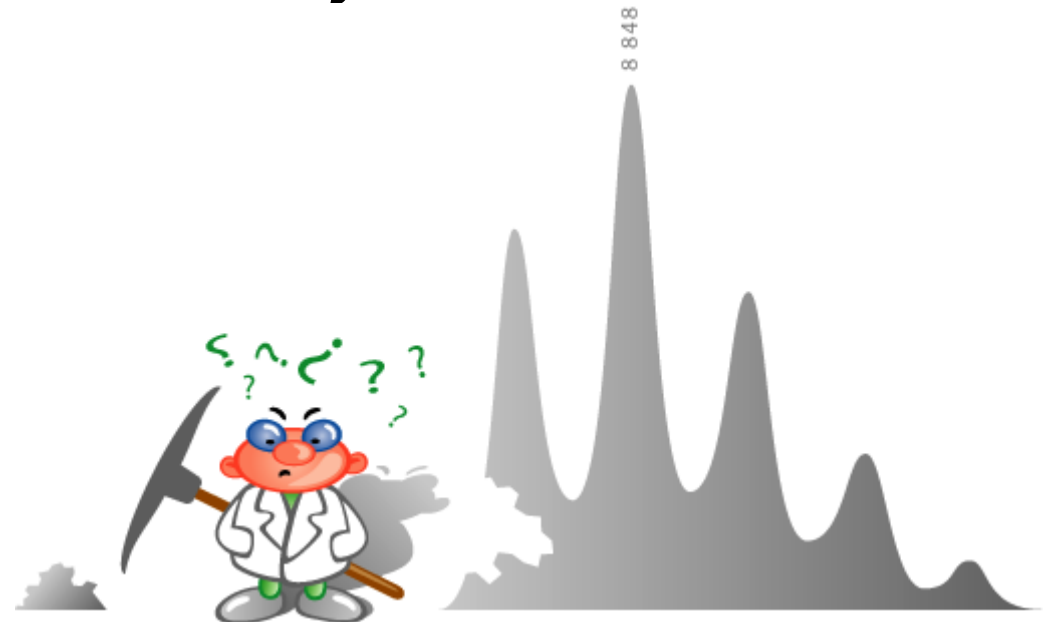


mMass Data Miner - an Open Source Alternative for Mass Spectrometric Data Analysis

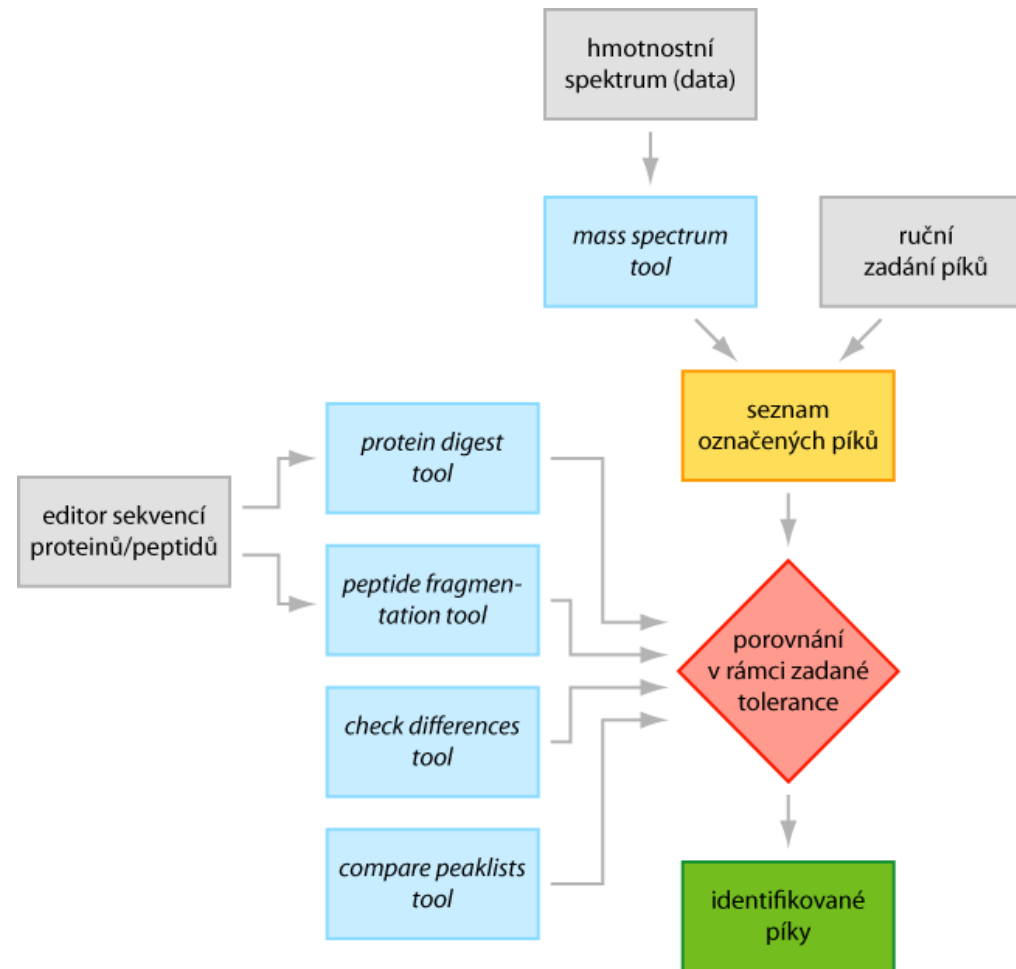
martin strohalm

*Institute of Chemical Technology in Prague
Department of biochemistry and microbiology*



Proč vytvářet další program?

- Přílišné zaměření současných programů na komerční sféru (*high throughput, automatizace*).
- Jiné použití než pro identifikaci proteinů je problematické.
- Příliš drahé licence.
- Absence nástroje dostupného i pro „klienty“ hmotníkářských laboratoří.
- Programy jsou vázány na datový formát výrobce daného instrumentu.



Podporované formáty

- Vlastní formát *.msd*
- Bruker FLEX series
- ASCII XY

m/z	intenzita
500.011169	113.000000
500.030304	112.000000
500.049469	106.000000
500.068604	105.000000
500.087769	125.000000
500.106903	107.000000
500.126068	114.000000
500.145203	91.000000

Formát dokumentu mSD

- Založen na XML
- Vychází z formátu *mzData*

```
<?xml version="1.0" encoding="utf-8" ?>
<mMassDoc version="1.0">

  <description>
    <date value="2002-01-28 15:51:01" />
    <operator value="Martin Strohal" />
    <contact value="martin@biographics.cz" />
    <institution value="ICT Prague, CZ, EU" />
    <instrument value="Bruker Biflex" />
    <notes></notes>
  </description>

  <peaklist>
    <peak mass="993.403163" intens="3620.444052" annots="[WVCNDGR (1x Carbamidomethyl)]" />
    <peak mass="1045.555317" intens="982.448626" annots="(0.01 Da) [117-125/0] [GTDVQAWIR]" />
  </peaklist>

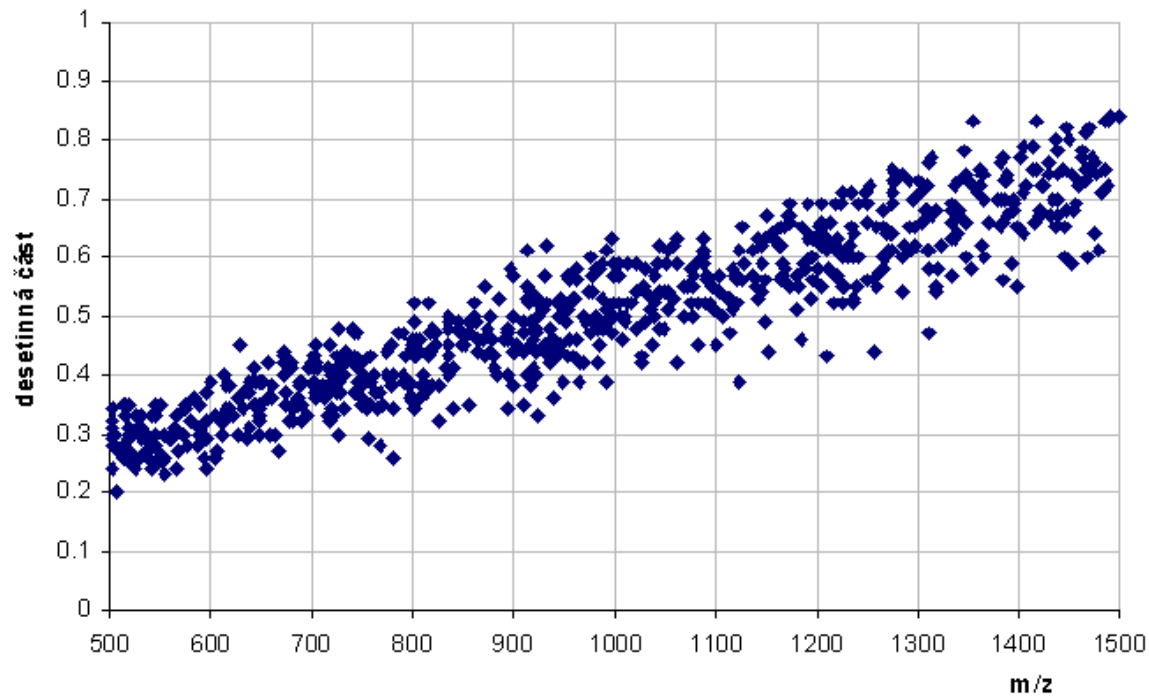
  <sequences>
    <sequence id="seq1">
      <title>In-gel digest experiment - LYSOZYME C - Gallus gallus (Chicken)</title>
      <seq>
        KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINSRWWCN
        DGRTPGSRNLNIPCSALLSSDITASVNCARKIVSDGNGMNAWVWRNRCKGTDVQAWIRGCRLL
      </seq>
      <modifications>
        <modification name="Carbamidomethyl" type="global" amino="C" gain="CH2CONH2" loss="H" />
        <modification name="Phospho" type="residual" amino="Y" position="19" gain="H2PO3" loss="H" />
      </modifications>
    </sequence>
  </sequences>

  <spectrum>
    <points value="148" />
    <startPoint value="993.026001" />
    <endPoint value="996.995605" />
    <mzArray compression="gz">
      eJwNz3loCAAYhvFkm2OE5YgZa2FZcs2xmNae0kxjuaUxW2hLy5H3G01L05o01lZnlabZhiwta2mzrJG1FtkcaxpZsliTjJY13/+3p63fLVIjRP
      Tloh3a0XhOpEQL4ISxNMNIitRLNokepLE3c0iZauYul207RDnd4n43SgWdttEZyio1PFjzRRul/sPShC00VrhttdBjPdHhalR4SOiYXH3cptlt
      uTbrPdnnKb4/a021y3Z3w3z3fz3Z51e84bCrzhgjcUui0S6y+JwCvecNV7i733uvfeEGW3xL7bIqzEv93xb6X+rUwMqxD198SJB2LJQ9FbKSoei
      bQqMf2x+FatimpE4hmXolY01InsehHzTPxqEPcbxYEXIuK16GgS15tFUosY9Uo8fyly3ojYNtH3V1s+F+ntYlaH6PwoijvFls9i7BfR1CVyv4pV
      30R/t6j6LjJ7RFSv6Popbv4SO3+LCX9Ey1+R3y/W/BODA6JmUBwdYswbanQHGCVBRvJwY8pIozXYKBhtxI8xAscZdSFG1nhj8USjd5JRptLICzX
      Cw4z2acbFcGNjhbE8w2icaeREGstnG31RRuUcI2OuETnf+LTAuBZtbFtshCwlmmOMvGVGXKwxsMKoXmn8B1HBBABY=
    </mzArray>
    <intArray compression="gz">
      eJwT0s0rRFEYx/GH8dp4SUIUXbHAUmq8lu49dydL6ykLW/kLJikbC1HWFhaShZGF50UuJOnGJEmykCRJ0pg0SfI9nWfx6dQ95zznd55zRXK+yAM
      KuFUxzcPBdz7kpkPRMqMSNbIbCoUbzqUaCCU9JWRzJyRKMncOmu6UIkmtKMZjeiDhzbU21qowi/1i3jBK95hM73hA39I6NonzSslasx+rGDHiHe
      GLHDHjAlk6yHXC3JSe+8y+2sDdbQ3LOMQdDnxX2+boYc0oJpHS+5SjG4PoRDUefZcp1ho57Z09YwLZQKJWI+NDZOk1S55vK0iJBXV6p1Lsse9Y8
      9ic17jRHuRREbg7F7U/9vunnuNHZRo321/TnGu72j3X6OGuTESYQOLmMEwkoHrc0FrbyHCpr7/DnZ1tP/Ehe9y2/ey75jQdywJ/gH42Xo/
    </intArray>
  </spectrum>

</mMassDoc>
```

Statistická kalibrace

(Ukázka vývoje desetinné části hmotnosti pro set peptidů, získaný *in silico* trypsinovým štěpením proteinů krevní plazmy.)



Použitý jazyk a další knihovny?

- Python (<http://www.python.org/>)
- wxPython (<http://www.wxpython.org/>)
- Numeric (<http://numeric.scipy.org/>)

- PSPad (<http://www.pspad.com/>)

Licence

- GPL – General Public License (<http://www.gnu.org/>)
- Open Source

Další plány

- Testování a odladění pro Linux a OSX.
- Přímé propojení s internetovými nástroji (*Mascot, Proteome Discoverer...*).
- Možnost definice a použití bifunkčních činidel.
- Více podporovaných nativních formátů MS instrumentů.
- Snížení nároků na výkon počítače (zejména překreslování spekter).
- ???

Poděkování

- vývojářům *Python*, *wxPython*, *Numeric*, *PSPad*
- Ing. Martin Hassman
- Ing. Bedřich Košata, Ph.D.
- Prof. RNDr. Milan Kodíček, CSc.
- Ing. Jiří Šantrůček
- Aleš Hnízda
- Martin Svoboda

Czech **FOBIA** Free & Open
BioInformatic Association



<http://mmass.biographics.cz/>