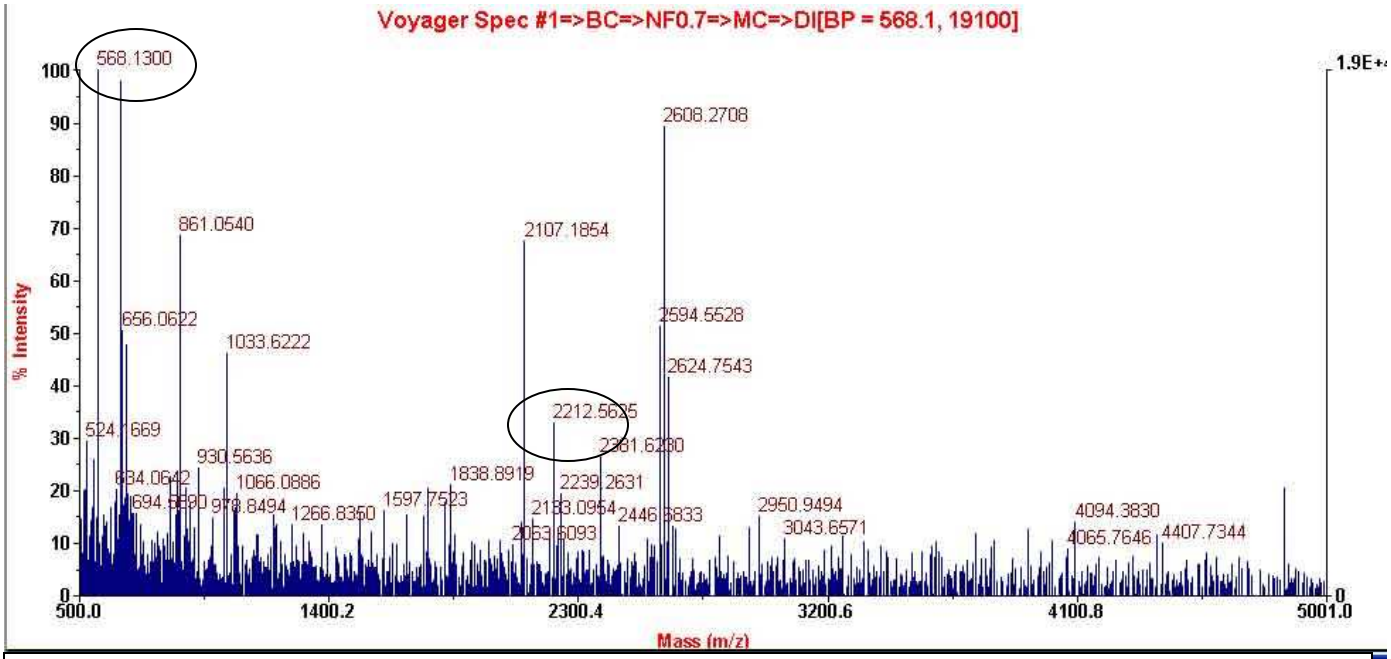
Bioinformatic analysis for protein identification based on Peptide mass fingerprinting and Mascot search

For generating the peptide mass fingerprint of a given unknown protein, the MALDI-TOF instrument was calibrated using insulin oxide β, ACTH fragment, angiotensin I, and bradykinin as external markers. The digested protein sample was then scanned and the raw data collected. The mass fingerprint data was processed and low intensity peaks were removed. The data was then calibrated using internal markers namely the trypsin autodigestion peaks and the matrix peaks. The mass peak list was then obtained from the calibrated mass spectrum for further analysis using the Mascot search engine.



Supplementary Figure 1: A typical mass spectrum showing the internal markers [the matrix fragment peak (m/z 568) and trypsin autodigestion peak (m/z 2212)] used for calibration.

The Mass peak list analysis with the help of Mascot search engine is exemplified by the identification of the mycobacterial GroEL protein below.

Mass peak list of the identified Gro EL protein of *Mycobacterium immunogenum*
568.130000,644.027962, 650.037983, 702.389178, 719.369343,855.060043, 971.553400, 1060.094248, 1067.540281, 1124.670212, ,1223.659512,1264.663920,1267.759728,1319.703215, 1348.865491, 1383.841602, 1413.658774,1477.781354,1478.057436,1482.123235, 1529.883423, 1545.880128, 1553.839224, 1567.970404,
1611.920965,1616.946832,1628.948559,1650.93001,1687.005462
1787.062278, 1804.077961,1817.965843,1826.088307, 1839.969713,1967.166654, 2175.314314
2277.428885, 2411.379643, 2591.438100, 2613.372461, 2658.607830, 2777.524738, 2780.429246, 2807.410604,
2827.601567, 2828.469083, 2838.588839, 2960.752111, 2983.709902, 3705.153086
3710.019362

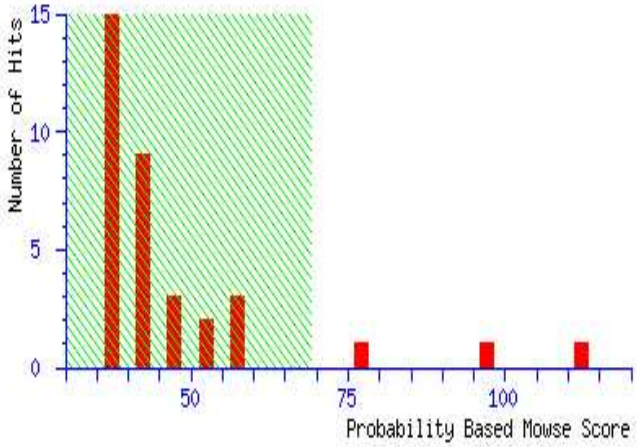
The protein mass peak list was used as the input in the Mascot search engine using the following criteria: Possible modification as a result of alkylation of cysteine (i.e. carbamidomethylation) and oxidation of methionine; mass tolerance of 100 ppm; one mis-cleavage allowed.

Mascot Search Results

```
User            : manish
Email           : guptam2@ucmail.uc.edu
Search title    :
Database        : SwissProt 56.6 (405506 sequences; 146166984 residues)
Timestamp       : 25 Dec 2008 at 06:41:24 GMT
Top Score       : 112 for CH602_MYCUA, 60 kDa chaperonin 2 OS=Mycobacterium ulcerans (strain Agy99) GN=groL2 PE=3 SV=1
```

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 69 are significant (p<0.05).



Concise Protein Summary Report

Format As | Concise Protein Summary | Help

Significance threshold p< 0.05 | Max. number of hits 5

Supplementary Figure 2: Mascot search results on the GroEL protein identification

Concise Protein Summary Report

1. CH602_MYCUA **Mass:** 56524 **Score: 112** **Expect:** 2.6e-06 **Queries matched:** 14
   60 kDa chaperonin 2 OS=Mycobacterium ulcerans (strain Agy99) GN=groL2 PE=3 SV=1
   CH602_MYCA1 **Mass:** 56580 **Score:** 56 **Expect:** 0.93 **Queries matched:** 10
   60 kDa chaperonin 2 OS=Mycobacterium avium (strain 104) GN=groL2 PE=3 SV=1
   CH602_MYCPA **Mass:** 56608 **Score:** 56 **Expect:** 0.93 **Queries matched:** 10
   60 kDa chaperonin 2 OS=Mycobacterium paratuberculosis GN=groL2 PE=3 SV=4
   CH601_MYCBP **Mass:** 56692 **Score:** 46 **Expect:** 10 **Queries matched:** 9
   60 kDa chaperonin 1 OS=Mycobacterium bovis (strain BCG / Pasteur 1173P2) GN=groL1 PE=3 SV=1
   CH601_MYCTA **Mass:** 56692 **Score:** 46 **Expect:** 10 **Queries matched:** 9
   60 kDa chaperonin 1 OS=Mycobacterium tuberculosis (strain ATCC 25177 / H37Ra) GN=groL1 PE=3 SV=1
   CH602_MYCBO **Mass:** 56692 **Score:** 46 **Expect:** 10 **Queries matched:** 9
   60 kDa chaperonin 2 OS=Mycobacterium bovis GN=groL2 PE=3 SV=2
   CH602_MYCTU **Mass:** 56692 **Score:** 46 **Expect:** 10 **Queries matched:** 9
   60 kDa chaperonin 2 OS=Mycobacterium tuberculosis GN=groL2 PE=1 SV=2

---

2. CH601_MYCSJ **Mass:** 56563 **Score: 98** **Expect:** 6.9e-05 **Queries matched:** 13
   60 kDa chaperonin 1 OS=Mycobacterium sp. (strain JLS) GN=groL1 PE=3 SV=1
   CH601_MYCSK **Mass:** 56563 **Score: 98** **Expect:** 6.9e-05 **Queries matched:** 13
   60 kDa chaperonin 1 OS=Mycobacterium sp. (strain KMS) GN=groL1 PE=3 SV=1
   CH601_MYCSS **Mass:** 56563 **Score: 98** **Expect:** 6.9e-05 **Queries matched:** 13
   60 kDa chaperonin 1 OS=Mycobacterium sp. (strain MCS) GN=groL1 PE=3 SV=2
   CH601_MYCGI **Mass:** 56308 **Score: 89** **Expect:** 0.0005 **Queries matched:** 12
   60 kDa chaperonin 1 OS=Mycobacterium gilvum (strain PYR-GCK) GN=groL1 PE=3 SV=1
   CH601_MYCVP **Mass:** 56479 **Score: 80** **Expect:** 0.0043 **Queries matched:** 11
   60 kDa chaperonin 1 OS=Mycobacterium vanbaalenii (strain DSM 7251 / PYR-1) GN=groL1 PE=3 SV=1
   CH601_MYCS2 **Mass:** 56453 **Score: 75** **Expect:** 0.012 **Queries matched:** 11

60 kDa chaperonin 1 OS=Mycobacterium smegmatis GN=groL1 PE=3 SV=1

CH602_NOCFA **Mass:** 56373 **Score:** 42 **Expect:** 29 **Queries matched:** 8

60 kDa chaperonin 2 OS=Nocardia farcinica GN=groL2 PE=3 SV=2

CH60_COREQ **Mass:** 56506 **Score:** 38 **Expect:** 61 **Queries matched:** 7

60 kDa chaperonin OS=Corynebacterium equii GN=groL PE=3 SV=1

---

3. CH602_MYCLE **Mass:** 56857 **Score:** **78** **Expect:** 0.0069 **Queries matched:** 11

60 kDa chaperonin 2 OS=Mycobacterium leprae GN=groL2 PE=3 SV=3

---

4. RADC_CITK8 **Mass:** 24757 **Score:** 58 **Expect:** 0.59 **Queries matched:** 7

DNA repair protein radC homolog OS=Citrobacter koseri (strain ATCC BAA-895 / CDC 4225-83 / SGSC4696) GN=radC PE=3 SV=1

---

5. CH601_RHOSR **Mass:** 56617 **Score:** 57 **Expect:** 0.79 **Queries matched:** 9

60 kDa chaperonin 1 OS=Rhodococcus sp. (strain RHA1) GN=groL1 PE=3 SV=1

As shown in the figure on Mascot search results, there were three significant hits (scores at >69 cut off value). This led to a positive identification of the protein as GroEL protein.

Protein view for top most hit

Match to: **CH602_MYCUA** Score: **112** Expect: **2.6e-06**
60 kDa chaperonin 2 OS=Mycobacterium ulcerans (strain Agy99) GN=groL2 PE=3 SV=1

Nominal mass (M_r): **56524;** Calculated pI value: **4.89**
NCBI BLAST search of CH602_MYCUA against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Mycobacterium ulcerans

Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: **51**
Number of mass values matched: **14**
Sequence Coverage: **32%**

Matched peptides shown in **Bold Red**

```
  1 MAKTIAYDEE ARRGLERGLN ALADAVKVTL GPKGRNVVLE KKWGAPTITN
 51 DGVSIAKEIE LEDPYEKIGA ELVKEVAKKT DDVAGDGTTT ATVLAQALVK
101 EGLRNVAAGA NPLGLKRGIE KAVEKVTETL LKSAKEVETK EQIAATAAIS
151 AGDQSIGDLI AEAMDKVGNE GVITVEESNT FGLQLELTEG MRFDKGYISG
```

```
201 YFVTDAERQE AVLEDPYILL VSSKVSTVKD LLPLLEKVIQ GGKPLLIIAE
251 DVEGEALSTL VVNKIRGTFK SVAVKAPGFG DRRKAMLQDM AILTGGQVIS
301 EEVGLSLETA DISLLGKARK VVITKDETTL VEGAGDSDAI AGRVAQIRAE
351 IENSDSDYDR EKLQERLAKL AGGVAVIKAG AATEVELKER KHRIEDAVRN
401 AKAAVEEGIV AGGGVALLHA VPSLDELKLS GDEATGANIV RVALEAPLKQ
451 IAFNGGLEPG VVAEKVRNSP VGTGLNAATG VYEDLLKAGV ADPVKVTRSA
501 LQNAASIAGL FLTTEAVVAD KPEKAAAPAG DPTGGMGGMD F
```

Show predicted peptides also

Sort Peptides By Residue Number Increasing Mass Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Sequence
4 - 12	1067.5403	1066.5330	1066.4931	37	0	K.TIAYDEEAR.R
4 - 13	1223.6595	1222.6522	1222.5942	47	1	K.TIAYDEEARR.G
18 - 27	971.5534	970.5461	970.5447	1	0	R.GLNALADAVK.V
43 - 57	1529.8834	1528.8761	1528.7886	57	0	K.WGAPTITNDGVSIAK.E
58 - 67	1264.6639	1263.6566	1263.5870	55	0	K.EIELEDPYEK.I
79 - 100	2175.3143	2174.3070	2174.1431	75	1	K.KTDDVAGDGTTTATV LAQALVK.E
105 - 116	1124.6702	1123.6629	1123.6349	25	0	R.NVAAGANPLGLK.R
196 - 208	1477.7814	1476.7741	1476.6885	58	0	K.GYISGYFVTDAER.Q
209 - 224	1804.0780	1803.0707	1802.9666	58	0	R.QEAVLEDPYILLVSSK.V
276 - 282	719.3693	718.3621	718.3398	31	0	K.APGFGDR.R
349 - 360	1413.6588	1412.6515	1412.5692	58	0	R.AEIENSDSDYDR.E
394 - 399	702.3892	701.3819	701.3708	16	0	R.IEDAVR.N
400 - 428	2828.4691	2827.4618	2827.5443	-29	1	R.NAKAAVEEGIVAGGGVALL HAVPSLDELK.L
450 - 465	1628.9486	1627.9413	1627.8570	52	0	K.QIAFNGGLEPGVVAEK.V