Validating protein biomarkers with mass spectrometry and molecular biology: a revolution

IAN A. BLAIR, PHD

Penn SRP Center

Center of Excellence in Environmental Toxicology
University of Pennsylvania
Philadelphia, PA 19104
Biomarker discovery: the problem in 2011

- Estimated number of papers claiming a biomarker: 150,000
- Estimated number of biomarkers in routine clinical use: ~100

No significant change since 2011

Biomarkers needed for early detection of cancer
The danger molecule - high mobility group box 1 (HMGB1) as a case study
Danger associated molecular pattern (DAMP) molecules

- Mono- and poly-saccharides
- DNA, RNA, and purine metabolites
- S-100 proteins
- Amyloid-β peptides

**DAMPs**

- HMGB1
- Monocytes
- Macrophages
- Platelets
- Lymphocytes

**Adaptive immune response**

- TIM3
- RAGE
- TLRs
- CXCR3

**Inflammation**

- Diseases:
  - Autoimmune
  - Cancer
  - Cardiovascular
  - Immune
  - Metabolic
  - Neurodegenerative
Asbestos-induced release of HMGB1 from macrophages

Asbestos → 150 nm → Cytosol → HMGB1

ROS → 5 µm → Cytosol → 8-oxo-dGuo → NLRP3 → ASC → NLRP3 Inflammasome → IL18, IL1β

NF-κB

Nucleus

Dostert et al. Science. 2008;320:674
## Serum HMGB1 concentrations in cancer cases measured by ELISA

<table>
<thead>
<tr>
<th>Study</th>
<th>Disease Type</th>
<th>Control mean (ng/mL)</th>
<th>Case mean (ng/mL)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Leukemia and lymphoma</td>
<td>0.9</td>
<td>3.1</td>
<td>Inoue 2013</td>
</tr>
<tr>
<td>2</td>
<td>Pancreatic</td>
<td>1.2</td>
<td>2.0</td>
<td>Wittwer 2013</td>
</tr>
<tr>
<td>3</td>
<td>Head and neck squamous cell</td>
<td>1.5</td>
<td>4.0</td>
<td>Wild 2012</td>
</tr>
<tr>
<td>4</td>
<td>Breast</td>
<td>2.0</td>
<td>4.5</td>
<td>Sun 2011</td>
</tr>
<tr>
<td>5</td>
<td>NSCLC</td>
<td>3.0</td>
<td>7.1</td>
<td>Niki 2017</td>
</tr>
<tr>
<td>6</td>
<td>Gastrointestinal</td>
<td>3.9</td>
<td>16.5</td>
<td>Qiu 2014</td>
</tr>
<tr>
<td>7</td>
<td>Mesothelioma (pleural)</td>
<td>5.4</td>
<td>27.0</td>
<td>Tabata 2013</td>
</tr>
<tr>
<td>8</td>
<td>Hepatocellular carcinoma</td>
<td>7.0</td>
<td>84.2</td>
<td>Cheng 2008</td>
</tr>
<tr>
<td>9</td>
<td>Cervical</td>
<td>7.6</td>
<td>50.8</td>
<td>Sheng 2009</td>
</tr>
<tr>
<td>10</td>
<td>Colorectal</td>
<td>39.7</td>
<td>58.8</td>
<td>Lee 2012</td>
</tr>
</tbody>
</table>
# Analysis of serum HMGB1: the problems

<table>
<thead>
<tr>
<th>#</th>
<th>Issue</th>
<th>Consequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Serum HMGB1 autoantibodies are present in normal serum and from diseased individuals</td>
<td>Serum autoantibodies compete with ELISA antibodies for HMGB1 leading to incorrect values for serum HMGB1</td>
<td>Sims 2010 Adulahad 2011 Hwang 2013</td>
</tr>
<tr>
<td>2</td>
<td>HMGB1 binds to haptoglobin, a high abundance serum protein (30-200 mg/dL)</td>
<td>HMGB1 binding to serum haptoglobin competes with ELISA antibodies leading to incorrect values for serum HMGB1</td>
<td>Yang 2015</td>
</tr>
<tr>
<td>3</td>
<td>Amino acid sequences of HMGB1 and HMGB2 are 74% similar,</td>
<td>ELISA antibodies to HMGB1 can cross-react with HMGB2, which would give incorrect values for serum HMGB1</td>
<td>Yamada 2006</td>
</tr>
<tr>
<td>4</td>
<td>22 lysine residues in HMGB1 are reported to be acetylated</td>
<td>Acetylation of HMGB1 could inhibit binding to ELISA antibodies and give incorrect values for serum HMGB1 concentrations</td>
<td>Sterner 1979 Bonaldi 2003 Rabadi 2004 Pasheva 2015</td>
</tr>
<tr>
<td>5</td>
<td>Platelets contain HMGB1, which is released when they aggregate during blood clotting</td>
<td>Normal platelets contain approximately 3 ng of HMGB1/mL of whole blood that will contribute to concentrations of serum HMGB1</td>
<td>Vogel 2016</td>
</tr>
<tr>
<td>6</td>
<td>Monocytes and lymphocytes contain HMGB1, which is released when blood clots</td>
<td>Monocyte and lymphocyte HMGB1 will contribute to concentrations of serum HMGB1</td>
<td>Maganelli 2010 Palmblad 2015 Cerwenka 2016</td>
</tr>
</tbody>
</table>
### Analysis of serum HMGB1: the solutions

<table>
<thead>
<tr>
<th>#</th>
<th>Issue</th>
<th>Potential solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Serum HMGB1 autoantibodies are present in normal serum and from diseased individuals</td>
<td>Synthesize labeled HMGB1 and allow to equilibrate with endogenous HMGB1 bound to autoantibodies</td>
</tr>
<tr>
<td>2</td>
<td>HMGB1 binds to haptoglobin, a high abundance serum protein (30-200 mg/dL)</td>
<td>Synthesize labeled HMGB1 and allow to equilibrate with endogenous HMGB1 bound to haptoglobin</td>
</tr>
<tr>
<td>3</td>
<td>Amino acid sequences of HMGB1 and HMGB2 are 74% similar,</td>
<td>Conduct stable isotope dilution LC-MS analysis of proteolytic peptides that are specific for HMGB1</td>
</tr>
<tr>
<td>4</td>
<td>22 lysine residues in HMGB1 are reported to be acetylated</td>
<td>Convert serum HMGB1 to a single molecular form by acetylation with labeled acetic anhydride</td>
</tr>
<tr>
<td>5</td>
<td>Platelets contain HMGB1, which is released when they aggregate during blood clotting</td>
<td>Compare HMGB1 in plasma containing anti-coagulant with serum HMGB1</td>
</tr>
<tr>
<td>6</td>
<td>Monocytes and lymphocytes contain HMGB1, which is released when blood clots</td>
<td>Compare HMGB1 in plasma containing anti-coagulant with serum HMGB1</td>
</tr>
</tbody>
</table>
Stable isotope labeling by amino acids in cell culture (SILAC) for labeled HMGB1

- Plasmid containing HMGB1
- mRNA
- Primers
- PCR amplification
- GST-HMGB1 fragment
- Restriction sites
- Restriction enzyme digestion
- Digested fragment containing GST-HMGB1
- pRK5 vector
- Assembled DNA
- HMGB1 pRK5 (Amp^r)
- Ligase
- Transformation and plating
- Ampicillin plate for bacterial selection
- SILAC-HMGB1 cell lysates
- HEK 293 cells
- SILAC medium
- \([^{13}\text{C}_6,^{15}\text{N}_2]\)-lysine
- \([^{13}\text{C}_9,^{15}\text{N}_1]\)-tyrosine
- Cell Transfection
- Bacterial expansion of recombinant DNA
Glu-C digestion of HMGB1

<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>MGKG</td>
<td>DPKKPR</td>
<td>GKMS</td>
<td>SYAFFV</td>
<td>QTCREEHKKK</td>
</tr>
<tr>
<td>60</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td>TMSAK</td>
<td>EKGKF</td>
<td>EDM</td>
<td>MAKADKAR</td>
<td>YEREMKTYIP</td>
</tr>
<tr>
<td>110</td>
<td>120</td>
<td>130</td>
<td>140</td>
<td>150</td>
</tr>
<tr>
<td>AFFLFCSEYR</td>
<td>PKIKG</td>
<td>EHPGL</td>
<td>SIG</td>
<td>DVAKKLG</td>
</tr>
<tr>
<td>160</td>
<td>170</td>
<td>180</td>
<td>190</td>
<td>200</td>
</tr>
<tr>
<td>LKEKYE</td>
<td>DIA</td>
<td>AYRA</td>
<td>KGKPD</td>
<td>AKKGVV</td>
</tr>
<tr>
<td>210</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDEEDEEEEEE</td>
<td>DDDDE</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NLS1:** E^{25}HKKKH {D^{33}} H^{26}KKKH {D^{33}}

K^{56}GKF {E^{60}} A^{170}AKKGVV {KAE^{179}}

**NLS2:** K^{180}SKKKKE^{186}

Five Glu-C peptides contain eleven of the important acetylation sites. Peptides too polar for consistent LC-MS analysis.
**Acetylation with labeled acetic anhydride and Glu-C digestion of HMGB1**

<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>MGKG</td>
<td>DPKN</td>
<td>GKMS</td>
<td>SYA FFV</td>
<td>QTCR</td>
</tr>
<tr>
<td>60</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td>TMSA</td>
<td>KEKG</td>
<td>F</td>
<td>EDM</td>
<td>ADA KAR</td>
</tr>
<tr>
<td>110</td>
<td>120</td>
<td>130</td>
<td>140</td>
<td>150</td>
</tr>
<tr>
<td>AFFL</td>
<td>FCSE</td>
<td>YR</td>
<td>PKIKG</td>
<td>EHPGL</td>
</tr>
<tr>
<td>160</td>
<td>170</td>
<td>180</td>
<td>190</td>
<td>200</td>
</tr>
<tr>
<td>LKEKY</td>
<td>EK</td>
<td>DIA</td>
<td>AYRAKGK</td>
<td>PD</td>
</tr>
<tr>
<td>210</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**EDEDEDEEEE DDDDE**

**[\textsuperscript{2}H_6\textsuperscript{]}-acetic anhydride**

**Glu-C**

**NLS1:** $E^{25}$HK(Ac\textsuperscript{*})K(Ac\textsuperscript{*})K(Ac\textsuperscript{*})HPD\textsuperscript{33} $H^{26}$K(Ac\textsuperscript{*})K(Ac\textsuperscript{*})K(Ac\textsuperscript{*})HPD\textsuperscript{33} $K^{56}$GK(Ac\textsuperscript{*})FE\textsuperscript{60} $A^{170}$AK(Ac\textsuperscript{*})K(Ac\textsuperscript{*})GVVK(Ac\textsuperscript{*})AE\textsuperscript{179}  

**NLS2:** K(Ac\textsuperscript{*})\textsuperscript{180}SK(Ac\textsuperscript{*})K(Ac\textsuperscript{*})K(Ac\textsuperscript{*})K(Ac\textsuperscript{*})E\textsuperscript{186}  

Ac\textsuperscript{*} = C[\textsuperscript{2}H_3]CO
UHPLC-MS analysis of the SILAC-HMGB1 standard coupled with acetylation and Glu-C digestion
Challenges in the analysis of low abundance serum proteins

- **Albumin**: 35-50 mg/mL
- **IgG**: 3-5 mg/mL
- **Transferrin**: 2-4 mg/mL

HMGB1 concentrations in healthy control serum are approximately 2.2 ng/mL or $22 \times 10^6$ lower than albumin.

Low abundance Proteins (10%)
Final scheme for analysis of HMGB1 in plasma and serum

1. Serum or plasma
2. SILAC-HMGB1
3. HMGB1 pAb magnetic beads
4. SDS-PAGE
5. HMGB1
6. [²H₆]-acetic anhydride
7. acetyl-[²H₃]-HMGB1
8. SILAC-HMGB1
9. HMGB1-[²H₃]-acetyl
10. In-gel Glu-C digestion
11. acetyl-[²H₃]-[²H₃]-acetyl
12. 2D-nano-UHPLC-PRM/MS
Analysis of HMGB1 in citrated plasma

- $E^{26}HK(Ac^*)K(Ac^*)K(Ac^*)HPDAS\text{FSE}^3+$: NS: 0, 630.0 to 753.9041, ratio < 0.001
- $E^{26}HK^*(Ac^*)K^*(Ac^*)K^*(Ac^*)HPDAS$ $\text{VNFS}^3+$: NS: 8.84E5, 638.0 to 765.9254, ratio < 0.001
- $K(Ac^*)^{57}GK(Ac^*)FE^{2+}$: NS: 3.23E3, 349.7 to 404.2785, ratio = 0.001
- $K^*(Ac^*)^{57}GK^*(Ac^*)FE^{2+}$: NS: 2.71E6, 357.7 to 420.3059
- $K(Ac^*)^{146}K(Ac^*)AAK(Ac^*)LK(Ac^*)E^{2+}$: NS: 1.86E3, 548.4 to 321.1848, ratio < 0.001
- $K^*(Ac^*)^{146}K^*(Ac^*)AAK^*(Ac^*)LK^*(Ac^*)E^{2+}$: NS: 4.44E5, 564.4 to 329.1990
- $K(Ac^*)^{180}SK(Ac^*)K(Ac^*)K(Ac^*)K(Ac^*)E^{2+}$: NS: 0, 550.9 to 780.5367, ratio < 0.001
- $K^*(Ac^*)^{180}SK^*(Ac^*)K^*(Ac^*)K^*(Ac^*)E^{2+}$: NS: 5.31E5, 570.9 to 812.5935
Standard curves for HMGB1 in citrated plasma

- **E$^{26}$HKK$^{30}$ (NLS-1)**
  - Equation: $y = 0.0055x + 0.0039$
  - $r = 0.9920$

- **K$^{57}$FGFE$^{61}$**
  - Equation: $y = 0.0053x + 0.0008$
  - $r^2 = 0.9854$

- **K$^{146}$KAALKE$^{152}$**
  - Equation: $y = 0.0056x + 0.0006$
  - $r^2 = 0.9947$

- **K$^{180}$SKKKKE$^{186}$ (NLS-2)**
  - Equation: $y = 0.0052x + 0.0003$
  - $r^2 = 0.9959$

- **Mean four peptides**
  - Equation: $y = 0.0054x + 0.0014$
  - $r^2 = 0.9931$
Analysis of serum HMGB1 by UPLC-HRMS
Levels of serum HMGB1 determined by stable isotope dilution UPLC-HRMS (n=24)

- Citrated plasma LC-MS: < 1.0 ng/mL
- EDTA plasma LC-MS: < 1.0 ng/mL
- Control serum LC-MS: 6.0 ng/mL
Big picture

- **Combining methodologies can accelerate biomarker development and molecular understanding of disease**
  - **Immunopurification** permits efficient isolation of very low abundance proteins from biofluids
  - **Mass spectrometry** provides specificity, precision, accuracy, and ultra-high sensitivity for biomarker analysis
  - **Molecular biology** provides protein internal standards and structural confirmation of protein biomarkers

- **HMGB1** is not present in plasma from healthy volunteers

- **HMGB1 is secreted** when blood is allowed to clot
  - **HMGB1** is not a circulating biomarker
  - **Serum HMGB1** studies (> 850) will have to be re-evaluated
  - **Plasma or whole blood** should be used for biomarker studies not serum
  - **Acetylation**, known to occur on 29 lysine residues, should be monitored
Acknowledgements

Anil Vachani
Wei-Ting Hwang
Kevin Qingqing

the BLAIR lab
University of Pennsylvania

Liwei

Anil Vachani
Trevor Penning
Wei-Ting Hwang

T32ES019851
P42ES023720
P30ES013508