six lectures on systems biology

jeremy gunawardena
department of systems biology
harvard medical school

lecture 2
31 march 2011

part 2 seminar room, department of genetics
a rather provisional syllabus

0. why mathematical models?

1. post-translational modification of proteins
2. microscopic cybernetics
3. development and evolution
1. post-translational modification (PTM)
the cartoon view of PTMs

but under the hood

many potential phospho-forms
phosphorylation is reversible and dynamic

- ATP
- Protein kinase
- Phosphoprotein phosphatase
- Released ADP
- Released inorganic phosphate
- Metabolic recharging process
- Amino acid in protein
- Phosphorylated amino acid
- $H_2O$
phospho-form distribution

relative proportion of each phospho-form in the molecular population
many kinds of reversible PTMs

<table>
<thead>
<tr>
<th>Modification</th>
<th>Product</th>
<th>Reagent</th>
<th>Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>phosphorylation</td>
<td>PO$_3^{2-}$</td>
<td>ATP</td>
<td>S, T, Y, H, D</td>
</tr>
<tr>
<td>sulfation</td>
<td>SO$_3^{-}$</td>
<td>PAPS</td>
<td>Y$^\dagger$ (extracellular)</td>
</tr>
<tr>
<td>methylation</td>
<td>CH$_3$</td>
<td>SAM</td>
<td>E, K(1-3), R(1-2)$^\dagger$</td>
</tr>
<tr>
<td>acetylation</td>
<td>CH$_3$CO</td>
<td>AcCoA</td>
<td>K</td>
</tr>
<tr>
<td>GlcNAcylation</td>
<td>C$<em>8$H$</em>{15}$NO$_6$</td>
<td>UDP-GlcNAc</td>
<td>S, T</td>
</tr>
<tr>
<td>ubiquitin-like</td>
<td>Ub, SUMO, Nedd</td>
<td></td>
<td>K (linear/branched polymers)</td>
</tr>
</tbody>
</table>

$^\dagger$ = reverse enzymes not known

which interact with each other

histone H3 N-terminal tail

C-terminal domain of p53

in many different ways

Binary switches and modification cassettes in histone biology and beyond


The Age of Crosstalk: Phosphorylation, Ubiquitination, and Beyond

distributions are biologically relevant - example 1

Kv2.1 voltage-gated potassium channel

16 S/T phosphorylation sites
7 are dephosphorylated by calcineurin in response to elevation of intracellular Ca\(^{2+}\)

Mutagenesis studies reveal that each of the seven calcineurin-modulated sites imparts a unique and incremental change in voltage-dependent gating.

Park, Mohapatra, Misonou, Trimmer, “Graded regulation of the Kv2.1 potassium channel by variable phosphorylation”, Science 313:976-9 2006
Distributions are biologically relevant – example 2

Steroid receptor co-activator 3 (SRC-3)

6 S/T phosphorylation sites

We showed that distinct combinations of phosphorylation sites were responsible for the function of different transcription factors and identified multiple cellular kinsases involved in the site-specific phosphorylations. Finally, we showed that different combinations of phosphorylation sites were required for different physiological functions.

Wu, Qin, Yi, Wong, Tsai, Tsai, O'Malley, “Selective phosphorylations of the SRC-3/AIB1 coactivator integrate genomic responses to multiple cellular signaling pathways”, Mol Cell 15:937-49 2004
distributions are biologically relevant – example 3

FRQ circadian clock component from Neurospora crassa

75 S/T phosphorylation sites on 63 tryptic peptides

networks of PTM

Phosphorylation pathways, comprised of kinases, phosphatases, and their substrates, are frequently studied as linear entities in isolation from their surrounding cellular context (Chen and Thomer, 2007; Fiedler et al., 2009). Although this simplistic treatment has identified thousands of kinase and phosphatase substrates, many of which display tissue specificity (Old et al., 2009), regulatory modifications are more realistically viewed as a network in which individual signaling cascades are interconnected by common substrates and interdependent regulation.

Klein, Dioum, Cobb, “Exposing contingency plans for kinase networks”, Cell 143:867-9 2010
another enchanting loom?

“Swiftly the head mass becomes an enchanted loom where millions of flashing shuttles weave a dissolving pattern, always a meaningful pattern though never an abiding one; a shifting harmony of subpatterns.”

Charles Sherrington, *Man on his Nature*, CUP 1942

1. how can we measure mod-form distributions?

2. how do PTM networks regulate the distributions?
measuring phospho-form distributions

1. phospho-specific antibodies – site-specific information

2. mass spectrometry – excellent for small proteins like histone tails


measuring phospho-form distributions

there is no oracle to tell us what is “really” there, so we take a comparative approach –

1. mass spectrometry (MS)
2. nuclear magnetic resonance spectroscopy (NMR)
3. phospho-specific antibodies

using differentially phosphorylated samples of Erk2 42kD MAP kinase, doubly phosphorylatable on TEY
4 phospho-forms – TY, pTY, TpY, pTpY

four samples

1. Transfect
   - Xenopus
   - His6-Erk2

2. Grow
   - E coli

3. Extract
   - 15N-labelled minimal media
   - Ni-NTA beads

4. Purify
   - doubly phosphorylate
   - Erk-TpY
   - Erk-pTY
   - activated GST-Mek + ATP

5. Dephosphorylate
   - PP2A
   - Erk-TpY
   - Erk-pTY
mass spectrometry with peptides

peptide-based LC/MS (pepMS) with internal standards
nuclear magnetic resonance

protein

→ digest

peptides

→ NMR

Bruker Avance 600MHz
but this elephant has a tail ...

at least two additional S/T phosphorylations on the Erk-pTpY sample
mass spectrometry with proteins

protein-based LC/MS (proMS)
4-site phospho-form distribution

Erk-pTpY

Proportion of total

50% < \(a_1\) < 65%

20% < \(a_{14}\) + \(a_{15}\) < 30%

\(a_8\) + \(a_{10}\) + \(a_{13}\) = 10%

\(a_1 = a_2 = a_3 = a_4 = a_5 = a_6 = a_7 = 0\)
\(a_9 = 0\)
\(a_{12} = 0\)

Phospho-forms

<table>
<thead>
<tr>
<th>(a_1)</th>
<th>(a_2)</th>
<th>(a_3)</th>
<th>(a_4)</th>
<th>(a_5)</th>
<th>(a_6)</th>
<th>(a_7)</th>
<th>(a_8)</th>
<th>(a_9)</th>
<th>(a_{10})</th>
<th>(a_{11})</th>
<th>(a_{12})</th>
<th>(a_{13})</th>
<th>(a_{14})</th>
<th>(a_{15})</th>
<th>(a_{16})</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>pS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>pS</td>
<td>pS</td>
<td>pS</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>pS</td>
<td>pS</td>
<td>S</td>
<td>pS</td>
<td>S</td>
</tr>
<tr>
<td>S</td>
<td>S</td>
<td>pS</td>
<td>S</td>
<td>S</td>
<td>pS</td>
<td>S</td>
<td>pS</td>
<td>S</td>
<td>S</td>
<td>pS</td>
<td>S</td>
<td>pS</td>
<td>S</td>
<td>pS</td>
<td>S</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>pT</td>
<td>T</td>
<td>pT</td>
<td>T</td>
<td>pT</td>
<td>T</td>
<td>pT</td>
<td>T</td>
<td>pT</td>
<td>T</td>
<td>pT</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>pY</td>
<td>Y</td>
<td>pY</td>
<td>Y</td>
<td>pY</td>
<td>Y</td>
<td>pY</td>
<td>Y</td>
<td>pY</td>
<td>Y</td>
<td>pY</td>
<td>pY</td>
<td></td>
</tr>
</tbody>
</table>

1 MAHHHHHHHAS AAAAASSNPG GGPMEVRGBQA FDVGPRTNL SYIGEGAYGM
51 VCSAHCELLNVR VRAAIKISR FEHQTCQRT LREQKILLRF KHEINIGIND
101 IIRAPTIEQM KDYIVQDLM EDTLKKLKT QHLSNDHICY FLYQILRLGK
151 YHISANVLHR DKLPSNLLNN TCDLHLICDF GLARVADPDH DHTGFLEXY
201 ATRWRPMBEI MLNSKGYRTS IDISVGCIL AEMLSNPRIF PGKHYLDQLN
251 HILGILGSPS QEDLNCIINL KARNYLLSLP HKNKVPWNRL FPNADPKALD
301 LLDKMHTFNP HKRIEVEAAAL AHPYLEQYYD PSDEPVAEAP FKFEMELDDL
351 PKETLKEFLF EETARFQPGY
summing up

1. *mod-form* distributions carry the most information about PTM state

2. downstream biological function depends on the distribution

3. biophysical measurements (*pepMS, proMS, NMR*) agree to within 10%

4. a hybrid strategy of *pepMS* + *proMS* can uncover phospho-form distributions for small numbers of sites (*n < 10?*)

5. *phospho-specific* antibodies are a biological readout (see point 2) – beware!