domains, sites, stubs...

graphical representation of pY-mediated recruitment (Fig. 2). These diagrams provide a system-level view of the ErbB receptors, showing biophysical interactions between signalling proteins and known sites of tyrosine phosphorylation. Which proteins are actually recruited in a given cell will depend on many factors, including the effective concentrations of both the activated receptors and the signalling proteins. These diagrams should therefore be viewed as quantitative maps of the receptors, rather than a depiction of protein recruitment in any specific cell type or state.

To evaluate how well our microarray experiments recapitulate known interactions, we compiled a list of previously reported interactions between SH2/PTB-containing proteins and the ErbB receptors (Supplementary Table 5). For interactions with EGFR and ErbB2, we relied on hand-curated databases (ref. 10 and http://proteome.incyte.com/); for ErbB3 and ErbB4, we surveyed the literature ourselves. Overall, our arrays detected 43 of the 65 previously reported interactions. For example, we observed that peptides derived from EGFR were able to bind strongly (Kd, 2mM) to the SH2/PTB domains of Crk, Grb2, Nck1, PI3Ka (also known as PIK3R1), PI3Kb (also known as PIK3R2), PLC-g1 (also known as PLCG1), PLC-g2 (also known as PLCG2), Shp2 (also known as PTPN11), RasGAP (also known as RASA1), Shc1, Shc3, Syk and Vav1, and weakly to the SH2 domains of Grb10, Grb7, Nck2, Shp1 (also known as PTPN6), Nsp1 (also known as SH2D3A), Socs1, Stat1, Stat3, Vav2 and Vav3. Many of the known interactions that were not detected were members of the STAT and SOCS families of binding takes center stage.
domains (cont’)

Kim and Gerstein papers/hubs party-, date- ...

what are the forces driving this mass action graph rewriting?
  – randomness
  – causalities/conditional binding
  – geometric hindrance/structural biology
agents and rules

Protein $A$ is activated by kinase $K$ on site $x$.

Phosphorylated site $y$ prevents kinase to activate $x$.

lhs=subgraph (don’t care don’t write)
actions

Site modification

Binding

Unbinding

Synthesis

Degradation
event:=rule application
a simple mag with dim=$\infty$

One ab binding rule with low dissociation $T$ (viscous)
snapshot of the CTMC (away from t=0)

50 val2 vs 100 val3

(2*s_val2<<3*s_val3)
100 of each

\((2 \times s_{val2} < 3 \times s_{val3})\)
giant cluster! phase transition
criticality happens easily and early (not just at $\infty$!)
sensitive to $T$ structure $=\,$ dissociation rate
sensitive to $E$ structure $:=\,$ local stub correlations
liquidity index
A random graph with sites consists of the following data:
- $n$ the set of nodes
- $K$ the (finite) set of colours
- $Z$ the node random variable with values in $\mathbb{N}^K$
- for each $a, b \in K$ a dissociation constant $\Gamma_{ab} \in [0, \infty]$
the dRG model II

- [binding] two free sites $x$, $y$ of respective colours $a$, $b$ bind each other with a probability proportional to $\gamma_{ab}^+$;
- [unbinding] two sites $x$, $y$ of respective colours $a$, $b$, and already bound together, unbind with a probability proportional to $\gamma_{ab}^-$. 
A random graph with sites consists of the following data:
- \( n \) the set of nodes
- \( K \) the (finite) set of colours
- \( Z \) the node random variable with values in \( \mathbb{N}^K \)
- for each \( a, b \in K \) a dissociation constant \( \Gamma_{ab} \in [0, \infty] \)

\[
n_a = n_a^f + \sum_b e_{ab}
\]

\[
\Gamma_{ab} e_{ab} = n_a^f n_b^f
\]

\[
\Gamma_{ab} \cdot e_{ab} = (n_a - \sum_c e_{ac})(n_b - \sum_d e_{db})
\]

\[
K_{ab} e_{ab} = (\langle m_a \rangle - \sum_c \epsilon_{ac})(\langle m_b \rangle - \sum_d \epsilon_{bd})
\]

- \( n \) the set of nodes
- \( K \) the set of colours together with \( * \) a special value not in \( K \)
- \( Z \) the node random variable with values in \( \mathbb{N}^K \)
- for each \( a \in K \), \( Y_a \) the edge random variable with values in \( K + \{ * \} \)
from dRG to sRG

limit distribution of dRG gives $Y_a/T_{ab}$ as a function of $e_{ab}$

$p(Y_a = b) := \epsilon_{ab}/\langle m_a \rangle$

$T_{ab} := \epsilon_{ab}/\langle m_a \rangle \langle m_b \rangle$
size and gf-ology

\[ T_{ab} m_b p_m \]

\[ S_p^a(z) := \sum_n p(S_p^a = n) z^n \]

\[ Z(x_c; c \in K) := \sum_{m \in \mathbb{N}^K} p(Z = m) \prod_{c \in K} x_c^{m_c} \]
\[ S_p^a(z) - p(Y^a = *) \]
\[ = z \sum_{n>0} p(S_p^a = n)z^{n-1} \]
\[ = z \sum_{n>0,m} \sum_{b \in K} T_{ab} m_b p_m p(\sum_{c \in m-b} S_{p-1}^c = n-1)z^{n-1} \]
\[ = z \sum_{b \in K} T_{ab} \sum_m m_b p_m (\sum_{n>0} p(\sum_{c \in m-b} S_{p-1}^c = n-1)z^{n-1}) \]
\[ = z \sum_{b \in K} T_{ab} \sum_m m_b p_m \prod_{c \in m-b} S_{p-1}^c(z) \]
\[ = z \sum_{b \in K} T_{ab} \partial_b Z(S_{p-1}^c(z); c \in K) \]
liquidity index

\[ \psi_a(x_b; b \in K) := p(Y^a = *) + \sum_{b \in K} T_{ab} \partial_b Z(x_b; b \in K) \]

\[ \partial_c \psi_a(x_b; b \in K) = \sum_{b \in K} T_{ab} \partial_c \partial_b Z(x_b; b \in K) \]

\[ \partial_c \psi_a(\mathbf{1}) = \sum_{b \in K} T_{ab} E_{bc} = (TE)_{ac} \]

liquidity index = highest modulus eigenvalue of TE

\[ \lambda_1(TE) \]
bicolor case

\[ N := \frac{\langle m_a m_b \rangle + \sqrt{\langle m_a (m_a - 1) \rangle \langle m_b (m_b - 1) \rangle}}{\langle m_a \rangle \langle m_b \rangle} \]

\[ \epsilon := \frac{\langle m_a \rangle + \langle m_b \rangle + K - \sqrt{(\langle m_a \rangle + \langle m_b \rangle + K)^2 - 4\langle m_a \rangle \langle m_b \rangle}}{2} \]
4.4 Subcritical bicolor systems

A particular and particularly simple case of bicolor systems is when one has a single agent type bearing one stub of each colour $a$ and particularly simple case of bicolor systems is when one has a single agent type bearing one stub of each colour $a_b$. Connected components are chains $\langle m_a \rangle = \langle m_b \rangle = vq$ and $\lambda = \epsilon \leq v$ which is only critical if $K = us$. Clearly the probability that a given chain has length $k$ will vary as $\epsilon^k$ and decrease rapidly with $k$. This is in fact a more general phenomenon.

If all nodes contain exactly one stub of type $a_q$ then the underlying system is subcritical runless $K = us$. Indeed, the assumption forces $\langle m_a m_m m_a - v \rangle = u$ and $N = vq$ so:

$$\lambda = \epsilon = \frac{vp}{\langle m_b \rangle p} K - \sqrt{3p(1 - p)} \leq v$$

One sees that the noise term $N$ plays a key role in criticality. Intriguingly, this suggests that large scale polymers made of divalent monomers because they cannot use too low a $K$ that would lead to irreversible behaviours. Of course this must be taken with a pinch of salt because biological polymers usually grow in a directed way and therefore should be idealised by conditional rules which our analytic approach cannot cope with at the moment and because we are dealing with an idealisation in the first places.

Note that this does not apply to our original $a_q b_x$ example $m$ $\sqrt{y}$ $s v_n q$ and indeed, by choosing carefully the parameters $Z_q$ and $K_q$ it is possible to obtain critical behaviours; the liquidity index is given by:

$$\lambda(p, K) := \frac{2 + p + K - \sqrt{(2 + p + K)^2 - 24p(1 - p)}}{2\sqrt{3p(1 - p)}}$$

with $p := p m Z = x b n$ the ratio of $x_b$ agents. Plotting $\lambda Figs y_n$ shows where critical behaviour happens.
simulations
n=200, \( K_{ab} = 1/4, \)

\[ \epsilon_{ab} = \frac{3}{4}, \quad \epsilon_{ba} = \frac{1}{2} \]

\[ TE = \frac{1}{2} \begin{pmatrix} 0 & 3 \\ 1 & 0 \end{pmatrix} \]
n=1000 size distribution
\[ n = 0.2 \times 10^6 \ - \text{max size} \]

simplx --sim 2a_3b.ka --time 0.03 --no-measure --no-maps --output-final-state --rescale 1000

\[
\begin{align*}
'a1-b1' & \quad 3b(b1),2a(a1) \leftrightarrow 3b(b1!1),2a(a1!1) @ 1.0,50 \\
'a1-b2' & \quad 3b(b2),2a(a1) \leftrightarrow 3b(b2!1),2a(a1!1) @ 1.0,50 \\
'a1-b3' & \quad 3b(b3),2a(a1) \leftrightarrow 3b(b3!1),2a(a1!1) @ 1.0,50 \\
'a2-b1' & \quad 3b(b1),2a(a2) \leftrightarrow 3b(b1!1),2a(a2!1) @ 1.0,50 \\
'a2-b2' & \quad 3b(b2),2a(a2) \leftrightarrow 3b(b2!1),2a(a2!1) @ 1.0,50 \\
'a2-b3' & \quad 3b(b3),2a(a2) \leftrightarrow 3b(b3!1),2a(a2!1) @ 1.0,50
\end{align*}
\]

\%init: 100 * (2a(a1,a2))
\%init: 100 * (3b(b3,b1,b2))

\%obs: 2a(a1!_)
\%obs: 3b(b1!_)
\%obs: 3b(b2!_)
\%obs: 3b(b3!_)
\%obs: 2a(a2!_)

the above takes \( n = 0.2 \times 10^6 \):
- Initialization: 27.5 sec. CPU
- Simulation: 213.5 sec. CPU the final state is 2MB - takes ages to write in a file! - max size is 168, of relative size < 0.1%. 
contact map

a + 2b vs a + b

multisets are enumerated

a + 2b allows branching

Gamma = 50, Z = 90 * ab + 10 * a2b
conclusions

• TE have non monotonic effects in the case of conflicting contact maps

• how good is liquidity a proxy for the size distribution

• what is the influence of other forces (extend to view-local systems)

• compute liquidity of yeast!

• where is information (more later)