Combinatorially Complex Equilibrium Model Selection

Tom Radivoyevitch
Assistant Professor
Epidemiology and Biostatistics
Case Western Reserve University

Email: txr24@case.edu
Website: http://epbi-radivot.cwru.edu/

dNTP Supply System

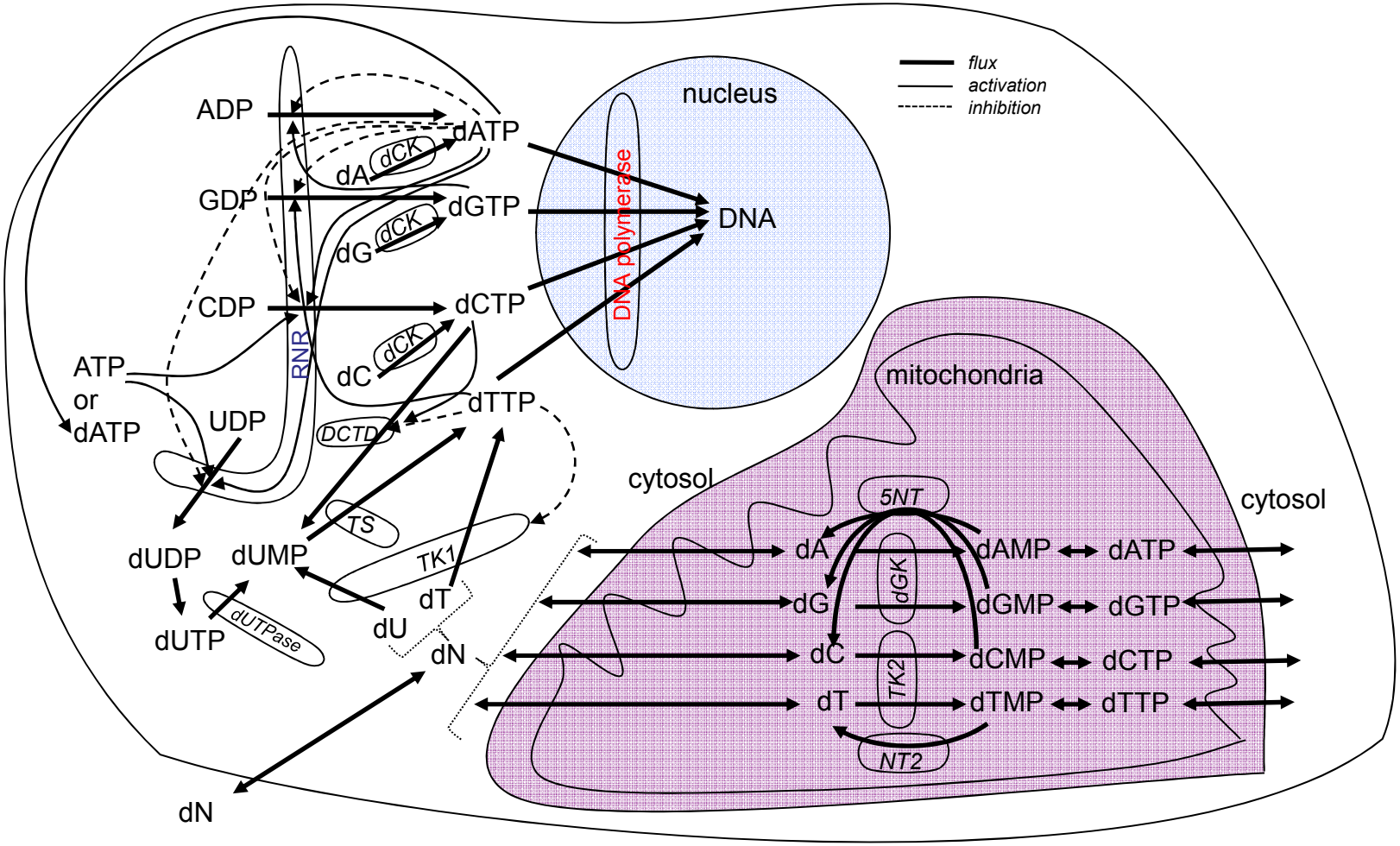
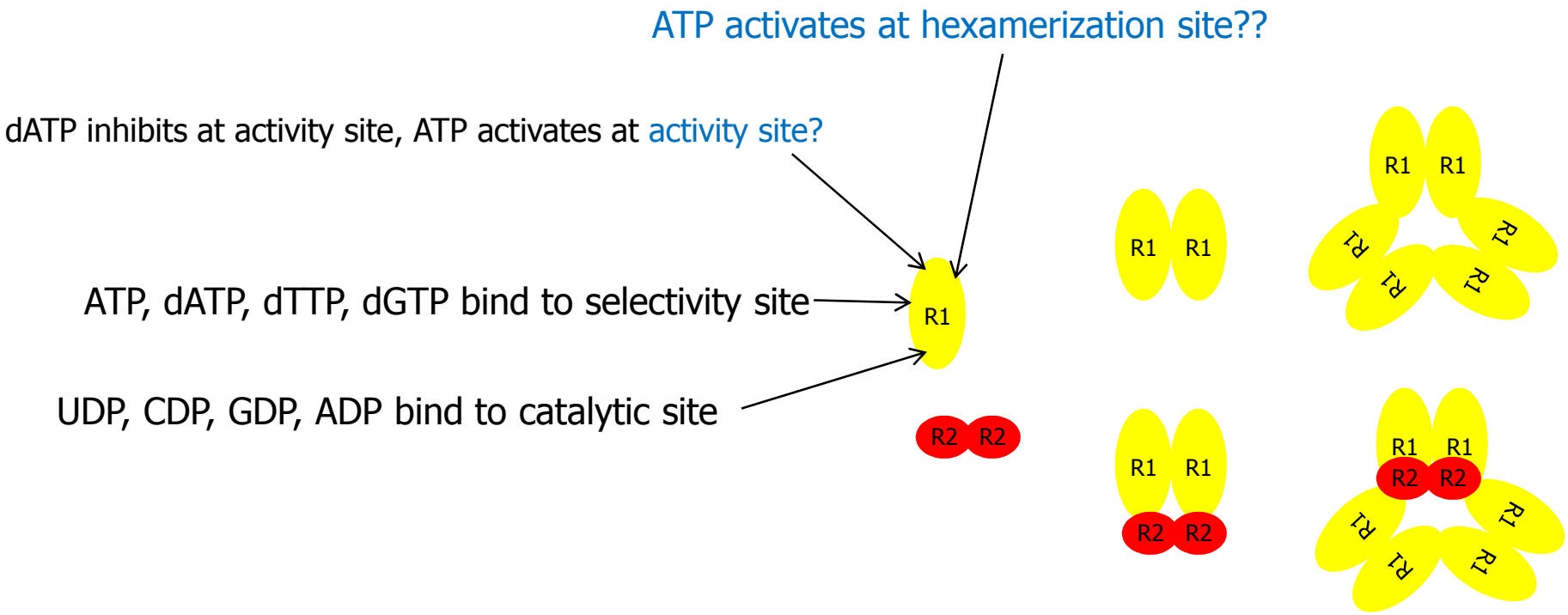


Figure 1. dNTP supply. Many anticancer agents act on or through this system to kill cells. The most central enzyme of this system is RNR.

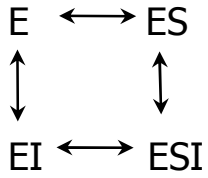
Ribonucleotide Reductase (RNR)

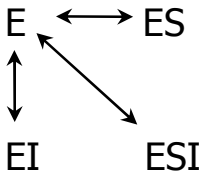


5 catalytic site states x 5 *s*-site states x 3 *a*-site states x 2 *h*-site states = 150 states
⇒ $(150)^6$ [=1.1×10^{13}] different hexamer complexes
⇒ $2^{(150)^6}$ models ~ 1 followed by a trillion zeros

RNR is Combinatorially Complex

Enzyme, Substrate and Inhibitor

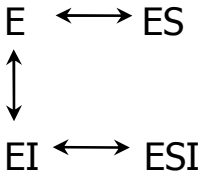




$$0 = [E_T] - [E] - \frac{[E][S]}{K_{ES}} - \frac{[E][I]}{K_{EI}} - \frac{[E][S][I]}{K_{ESI}}$$

$$0 = [S_T] - [S] - \frac{[E][S]}{K_{ES}} - \frac{[E][S][I]}{K_{ESI}}$$

$$0 = [I_T] - [I] - \frac{[E][I]}{K_{EI}} - \frac{[E][S][I]}{K_{ESI}}$$

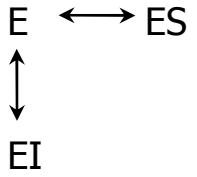


$$0 = [E_T] - [E] - \frac{[E][S]}{K_{E\_S}} - \frac{[E][I]}{K_{E\_I}} - \frac{[E][I][S]}{K_{E\_I}K_{EI\_S}}$$

$$0 = [S_T] - [S] - \frac{[E][S]}{K_{E\_S}} - \frac{[E][I][S]}{K_{E\_I}K_{EI\_S}}$$

$$0 = [I_T] - [I] - \frac{[E][I]}{K_{E\_I}} - \frac{[E][I][S]}{K_{E\_I}K_{EI\_S}}$$

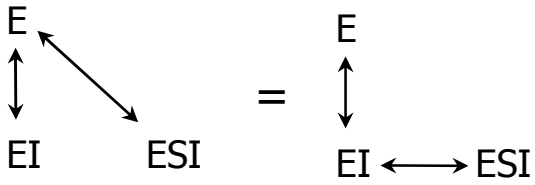
Competitive inhibition



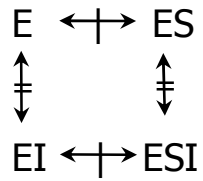
$$0 = [E_T] - [E] - \frac{[E][S]}{K_{ES}} - \frac{[E][I]}{K_{EI}}$$

$$0 = [S_T] - [S] - \frac{[E][S]}{K_{ES}}$$

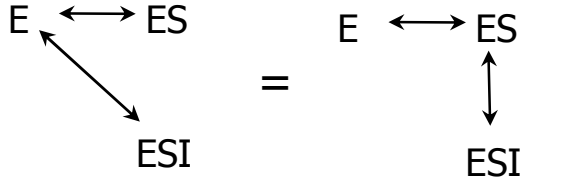
$$0 = [I_T] - [I] - \frac{[E][I]}{K_{EI}}$$

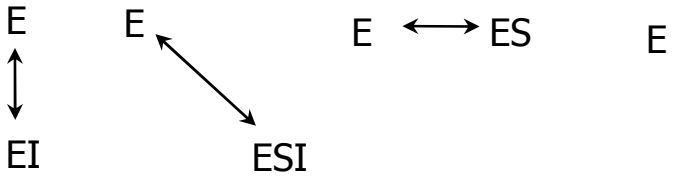


noncompetitive inhibition
Example of K=K' Model



uncompetitive inhibition if $k_{cat\_ESI}=0$





ATP-induced R1 Hexamerization

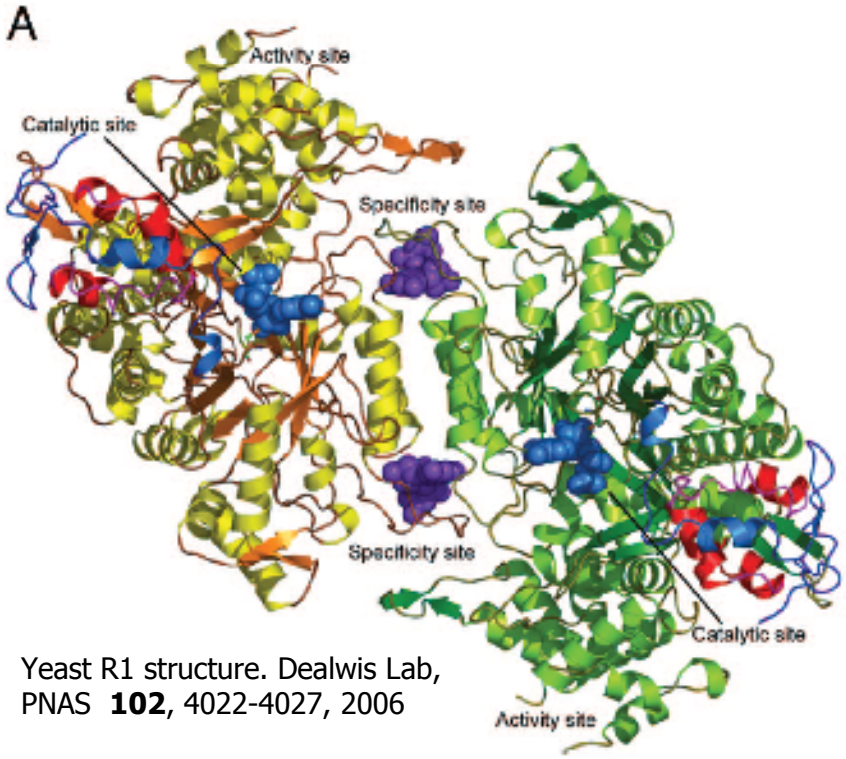
$$0 = [R_T] - [R] - \sum_{i=1}^{2} \frac{[R][X]^i}{K_{RX^i}} - 2\left(\sum_{i=2}^{6} \frac{[R]^2[X]^i}{K_{R^2X^i}}\right) - 4\left(\sum_{i=4}^{12} \frac{[R]^4[X]^i}{K_{R^4X^i}}\right) - 6\left(\sum_{i=6}^{18} \frac{[R]^6[X]^i}{K_{R^6X^i}}\right)$$

$$0 = [X_T] - [X] - \left(\sum_{i=1}^{2} i\frac{[R][X]^i}{K_{RX^i}}\right) - \left(\sum_{i=2}^{6} i\frac{[R]^2[X]^i}{K_{R^2X^i}}\right) - \left(\sum_{i=4}^{12} i\frac{[R]^4[X]^i}{K_{R^4X^i}}\right) - \left(\sum_{i=6}^{18} i\frac{[R]^6[X]^i}{K_{R^6X^i}}\right)$$

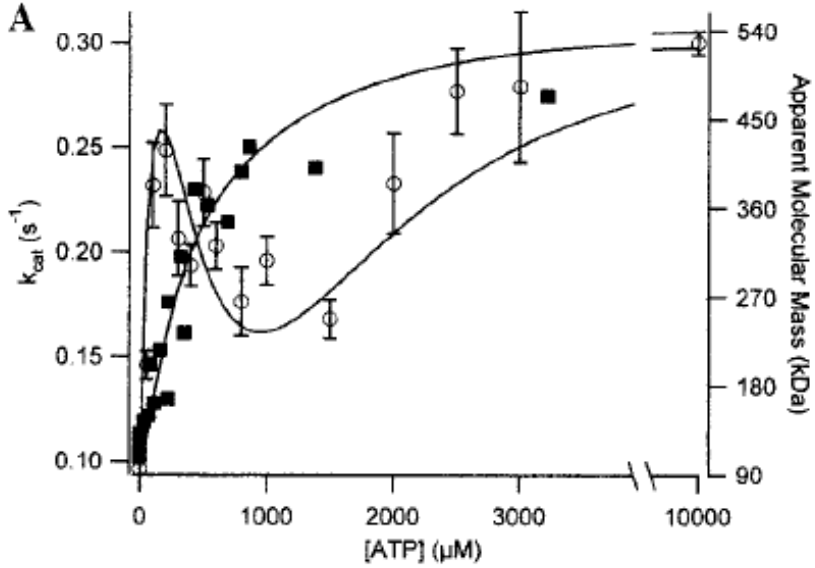
2+5+9+13 = 29 parameters => $2^{29}=5\text{x}10^8$ spur graph models via $K_j=\infty$ hypotheses

29 models with 1 parameter, 408 models with 2, 3654 models with 3, 23751 with 4

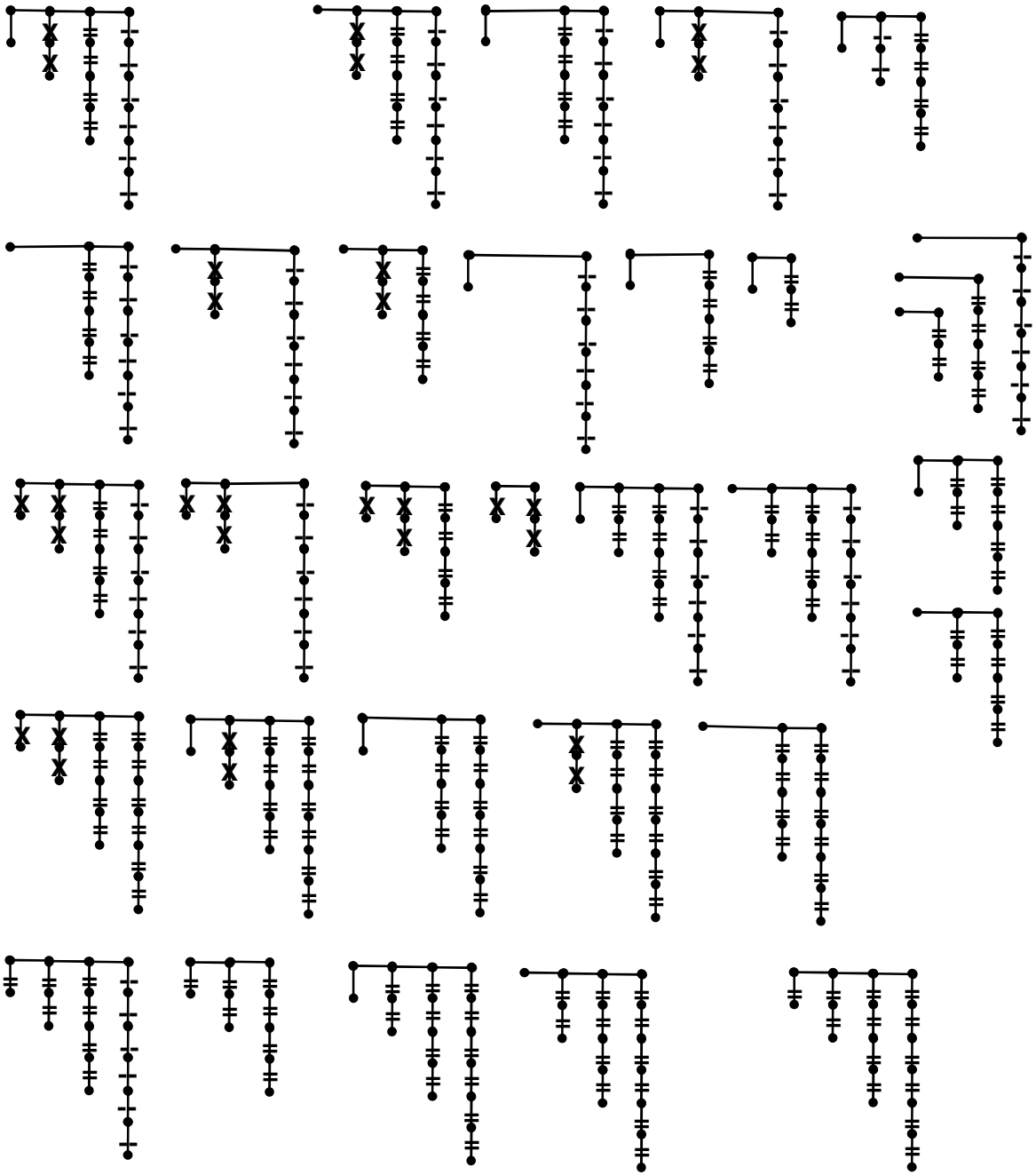
R = R1
X = ATP



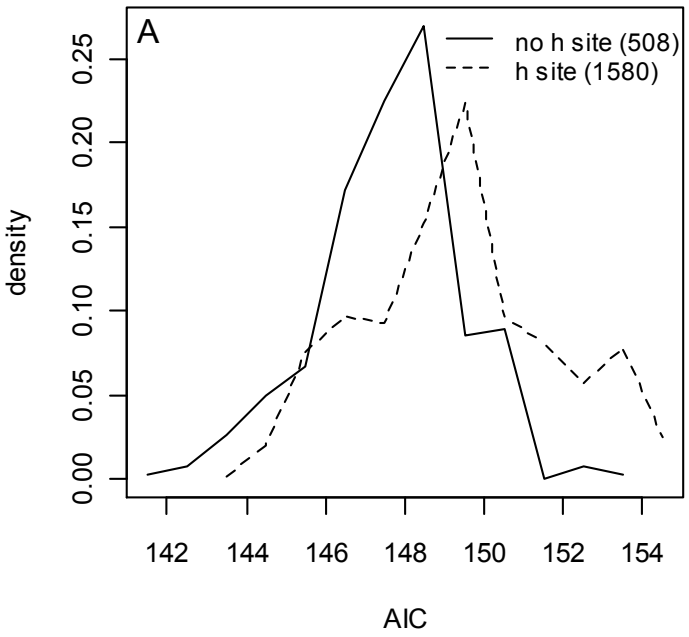
Yeast R1 structure. Dealwis Lab, PNAS **102**, 4022-4027, 2006

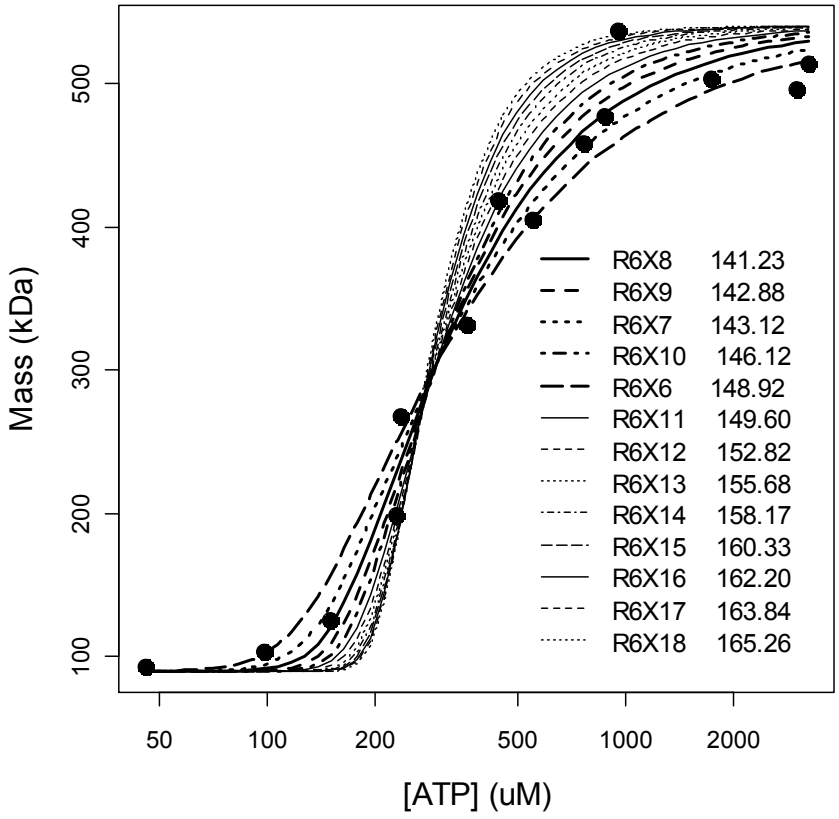


Kashlan et al. Biochemistry 2002 **41**:462



2088 Models with SSE < 2 min (SSE)





$AIC_c = N*\log(SSE/N)+2P+2P(P+1)/(N-P-1)$

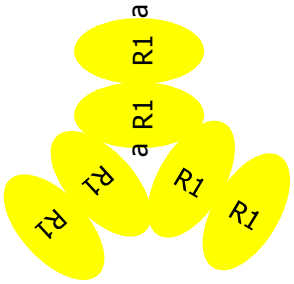
Kolomogorov-Smirnov Test $p < 10^{-16}$
28 of top 30 did not include an *h*-site term; $28/30 \neq 503/2081$ with $p < 10^{-16}$
This suggests no h-site.

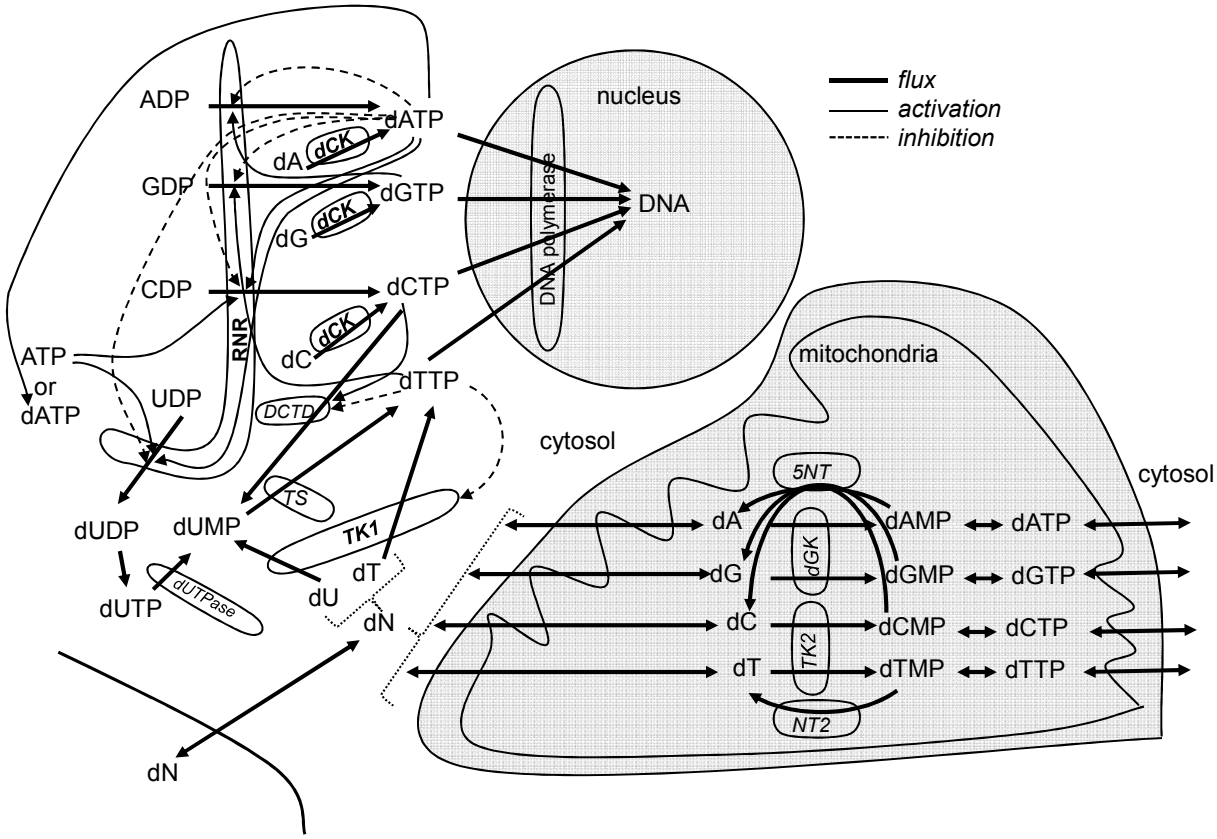
Top 13 included R6X8 or R6X9, save one, single edge model R6X7
This suggests that less than 3 a-sites are occupied in hexamer.

Conclusions (so far)

1. The dataset does not support the existence of an h-site
2. The dataset suggests that ~1/2 of the a-site are not occupied by ATP

hypothesis



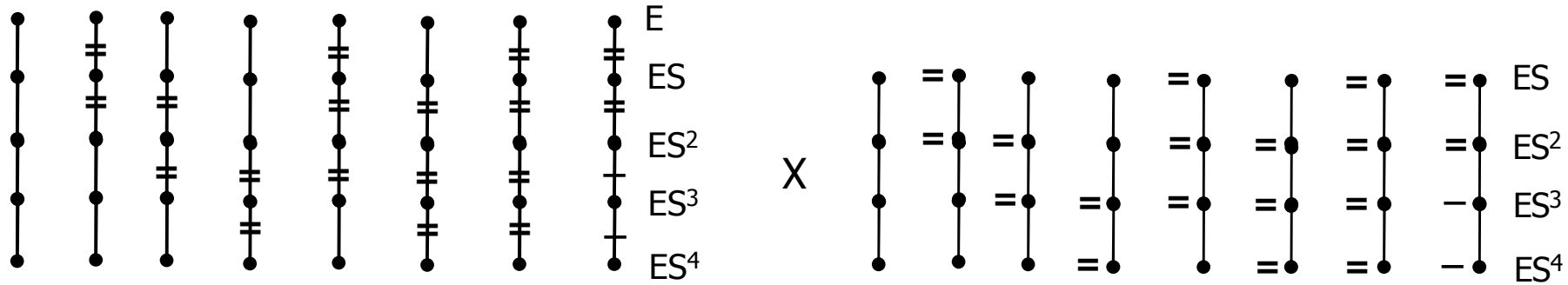


[ATP]=~1000[dATP]

Conjecture: system prefers to have 3 a-sites empty and ready for dATP

Conjecture: Inhibition versus activation is partly due to differences in pockets

Tetrameric Enzyme Models (e.g. Thymidine Kinase 1)



8 K Models x 8 k Models => 64 Models

$$0 = [E_T] - [E] - \frac{[E][S]}{K_{ES^1}} - \frac{[E][S]^2}{K_{ES^2}} - \frac{[E][S]^3}{K_{ES^3}} - \frac{[E][S]^4}{K_{ES^4}}$$

$$0 = [S_T] - [S] - \frac{[E][S]}{K_{ES^1}} - 2\frac{[E][S]^2}{K_{ES^2}} - 3\frac{[E][S]^3}{K_{ES^3}} - 4\frac{[E][S]^4}{K_{ES^4}}$$

$$[ES^i] = \frac{[E][S]^i}{K_{ES^i}} \qquad i = \{1, 2, 3, 4\}$$

$$k = \frac{\sum_{i=1}^{4} i k_i [ES^i]}{4[E_T]}$$

$$k = \frac{\sum_{i=1}^{4} i k_i [ES^i]}{4[E_T]} = \frac{\sum_{i=1}^{4} i k_i \frac{[E][S_T]^i}{K_{ES^i}}}{4[E]\left(1 + \sum_{i=1}^{4} \frac{[S_T]^i}{K_{ES^i}}\right)} = \frac{\sum_{i=1}^{4} i k_i \frac{[S_T]^i}{K_{ES^i}}}{4\left(1 + \sum_{i=1}^{4} \frac{[S_T]^i}{K_{ES^i}}\right)}$$

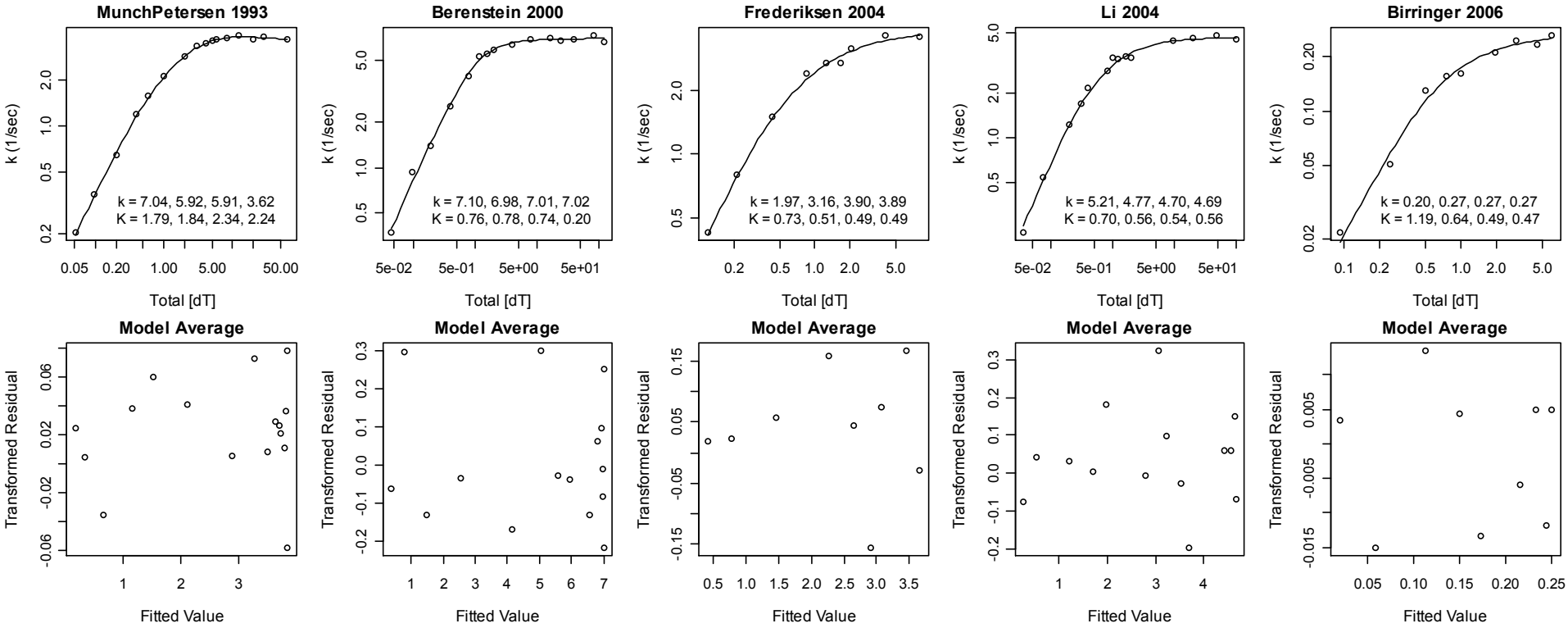
$$k_{off}[ES] = 4k_{on}[E][S]$$

$$\Rightarrow \quad K_{E\_S} \equiv \frac{k_{off}}{k_{on}} = \frac{4[E][S]}{[ES]} = 4K_{ES}$$

$$\Rightarrow \quad K_{ES} = \frac{K_{E\_S}}{4}$$

$$K_{E\_S}K_{ES\_S} = \frac{4[E][S]}{[ES]}\frac{3[ES][S]}{2[ES^2]} = \frac{6[E][S]^2}{[ES^2]} = 6K_{ES^2}$$

$$\Rightarrow \quad K_{ES^2} = \frac{K_{E\_S}K_{ES\_S}}{6}$$



K = (0.85, 0.69, 0.65, 0.51) μM
k = (3.3, 3.9, 4.1, 4.1) sec^{-1}

$$k = \frac{k_{\max}\left(\frac{[S_T]}{S_{50}}\right)^h}{1+\left(\frac{[S_T]}{S_{50}}\right)^h}$$

k_{max} = 4.1/sec
S_{50} = 0.6 μM
h = 1.1

Model Averages

Probability Model is true = $e^{\Delta AIC}/\sum e^{\Delta AIC}$

$AIC_c = N*\log(SSE/N)+2P+2P(P+1)/(N-P-1)$

Burnham, K. P., and Anderson, D. R. (2002) *Model Selection and Multimodel Inference: A Practical-Theoretic Approach*, Springer-Verlag

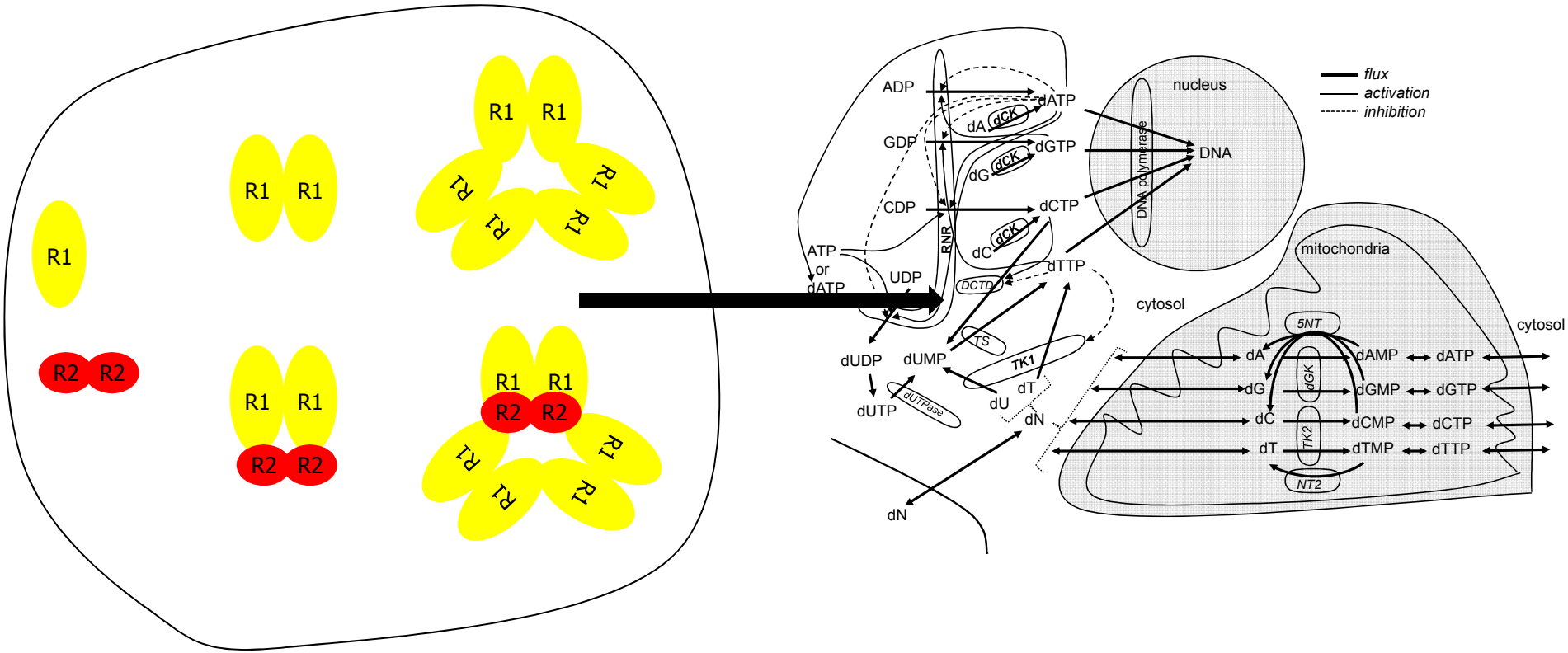
SUMMARY

Combinatorially Complex Equilibrium Model Selection (CCEMS, CRAN 2009)

Model individual enzymes

Systems Biology Markup Language interface to R (SBMLR, BIOC 2004)

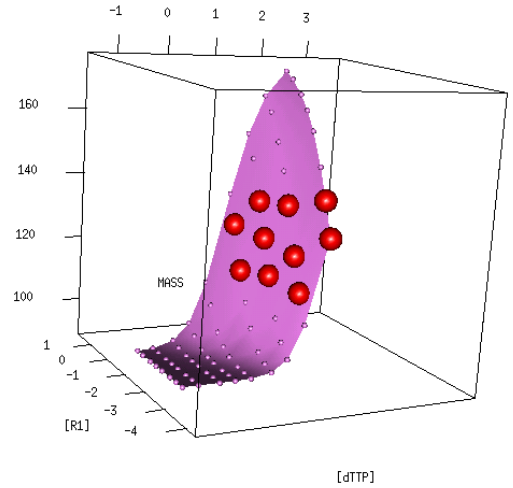
Model networks of enzymes

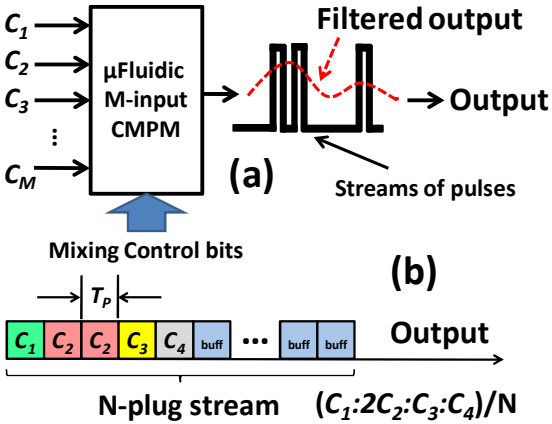


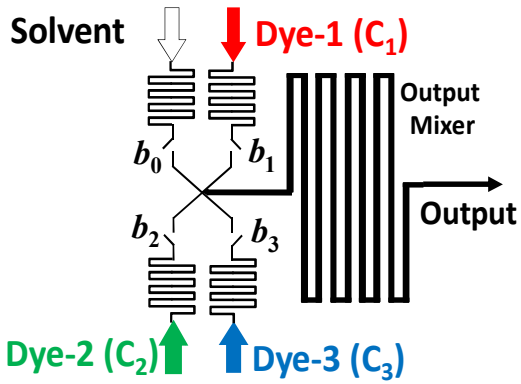
Adaptive Experimental Designs

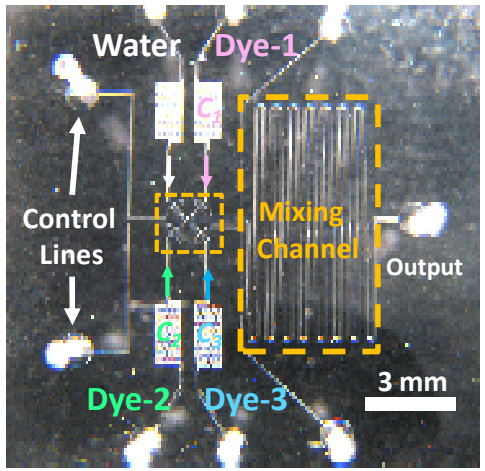
Find best next 10 measurement conditions given models of data collected.

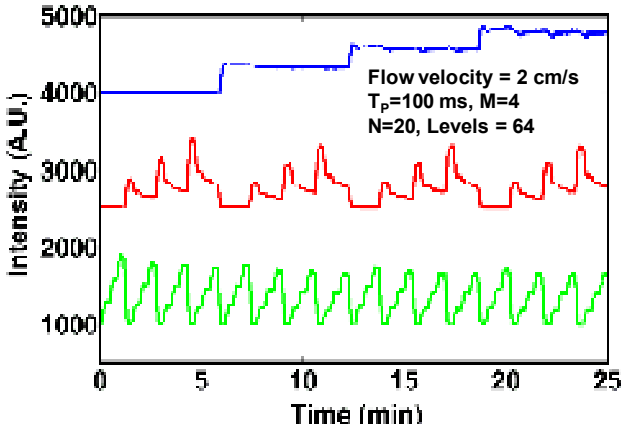
Need automated analyses in feedback loop of automatic controls of microfluidic chips

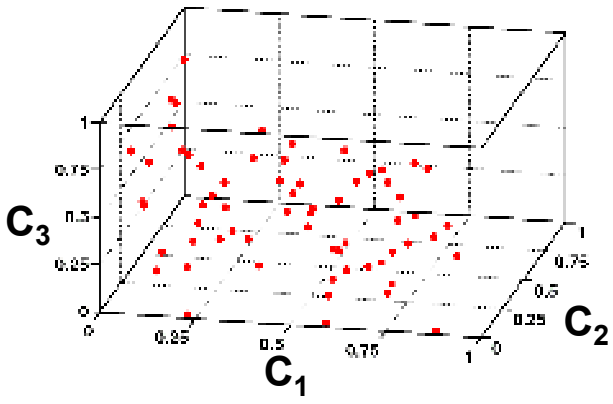












Acknowledgements

- Case Comprehensive Cancer Center
- NIH (K25 CA104791)
- Birgitte Munch-Petersen (Roskilde)
- Chris Dealwis (CWRU Pharmacology)
- Anders Hofer (Umea)
- Thank you

```
library(ccems) #     Thymidine Kinase Example
topology = list(
    heads=c("E1S0"), # E1S0 = substrate free E
    sites=list(
        c=list(    # c for catalytic site
            t=c("E1S1","E1S2","E1S3","E1S4")
        ) # t for tetramer
    )
) # TK1 is 25kDa = 25mg/umole, so 1 mg = .04 umoles
g = mkg(topology, activity=T,TCC=F)
dd=subset(TK1,(year==2000),select=c(E,S,v))
dd=transform(dd, ET=ET/4,v=ET*v/(.04*60))# now uM/sec
tops=ems(dd,g,maxTotalPs=8,kIC=10,topN=96)#9 min on 1 cpu
```

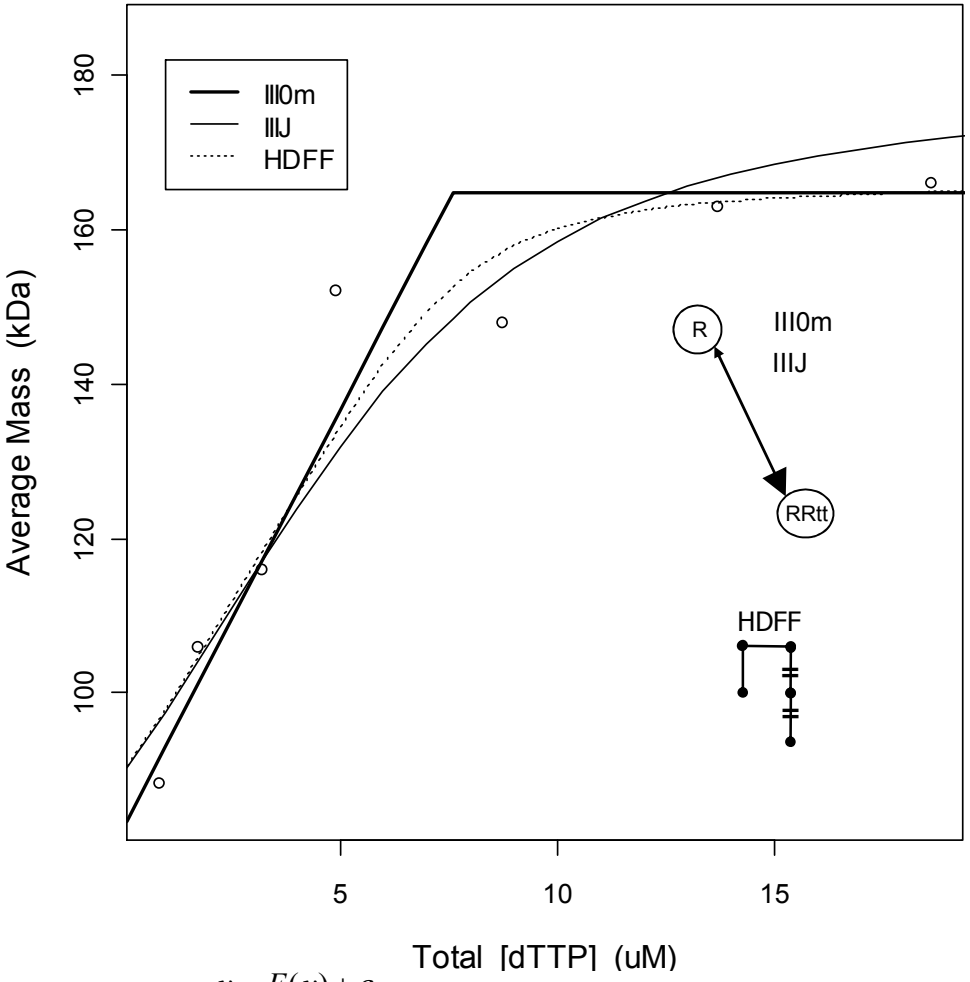
$K = e^{\Delta G/RT}$ => $K = (0, \infty)$ maps to $\Delta G = (-\infty, \infty)$

Model weights $e^{\Delta AIC}/\Sigma e^{\Delta AIC}$

```
library(ccems)  # Ribonucleotide Reductase Example
topology <- list(
    heads=c("R1X0","R2X2","R4X4","R6X6"),
    sites=list(                # s-sites are already filled only in (j>1)-mers
        a=list(  #a-site                                                    thread
            m=c("R1X1"),                                          # monomer   1
            d=c("R2X3","R2X4"),                                   # dimer     2
            t=c("R4X5","R4X6","R4X7","R4X8"),                     # tetramer  3
            h=c("R6X7","R6X8","R6X9","R6X10", "R6X11", "R6X12")   # hexamer   4
        ), # tails of a-site threads are heads of h-site threads
        h=list(   # h-site
            m=c("R1X2"),                                          # monomer   5
            d=c("R2X5", "R2X6"),                                  # dimer     6
            t=c("R4X9", "R4X10","R4X11", "R4X12"),                # tetramer  7
            h=c("R6X13", "R6X14", "R6X15","R6X16", "R6X17", "R6X18")# hexamer   8
        )
    )
)
g=mkg(topology,TCC=TRUE)
dd=subset(RNR,(year==2002)&(fg==1)&(X>0),select=c(R,X,m,year))
cpusPerHost=c("localhost" = 4,"compute-0-0"=4,"compute-0-1"=4,"compute-0-2"=4)
top10=ems(dd,g,cpusPerHost=cpusPerHost, maxTotalPs=3, ptype="SOCK",KIC=100)
```

Application to Data

Scott, C. P., Kashlan, O. B., Lear, J. D., and Cooperman, B. S. (2001) *Biochemistry* **40**(6), 1651-166



Model	Parameter	Initial Value	Optimal Value	Confidence Interval
1 III0m	m1	90.000	82.368	(79.838, 84.775)
	SSE	4397.550	525.178	
	AIC	71.965	57.090	
2 IIIJ	R2t2	1.000^3	2.725^3	(2.014^3, 3.682^3)
	SSE	2290.516	557.797	
	AIC	67.399	57.512	
27 HDFF	R2t0	1.000	12369.79	(0, 1308627507869)
	R1t0_t	1.000	1.744	(0.003, 1187.969)
	R2t0_t	1.000	0.010	(0.000, 403.429)
	SSE	25768.23	477.484	
	AIC	105.342	77.423	

$$y = E(y) + \varepsilon$$

$$E(y) = M_1 \frac{[R] + [R_T](1-p)}{[R_T]} + 2M_1 \frac{2[RR] + 2[RRt] + 2[RRtt]}{[R_T]}$$

$$0 = p[R_T] - [R] - \frac{[R][t]}{K_{Rt}} - 2\frac{[R]^2}{K_{RR}} - 2\frac{[R]^2[t]}{K_{RRt}} - 2\frac{[R]^2[t]^2}{K_{RRtt}}$$

$$0 = [t_T] - [t] - \frac{[R][t]}{K_{Rt}} - \frac{[R]^2[t]}{K_{RRt}} - 2\frac{[R]^2[t]^2}{K_{RRtt}}$$

$$\frac{d[R]}{d\tau} = p[R_T] - [R] - \frac{[R][t]}{K_{Rt}} - 2\frac{[R]^2}{K_{RR}} - 2\frac{[R]^2[t]}{K_{RRt}} - 2\frac{[R]^2[t]^2}{K_{RRtt}}$$

$$\frac{d[t]}{d\tau} = [t_T] - [t] - \frac{[R][t]}{K_{Rt}} - \frac{[R]^2[t]}{K_{RRt}} - 2\frac{[R]^2[t]^2}{K_{RRtt}}$$

$$[R](0) = 0; \quad [t](0) = 0.$$

$$\left[R^i t^j\right] = \frac{[R]^i[t]^j}{K_{R^i t^j}}$$

AIC_c = N*log(SSE/N)+2P+2P(P+1)/(N-P-1)

Radivoyevitch, (2008) BMC Systems Biology 2:15