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

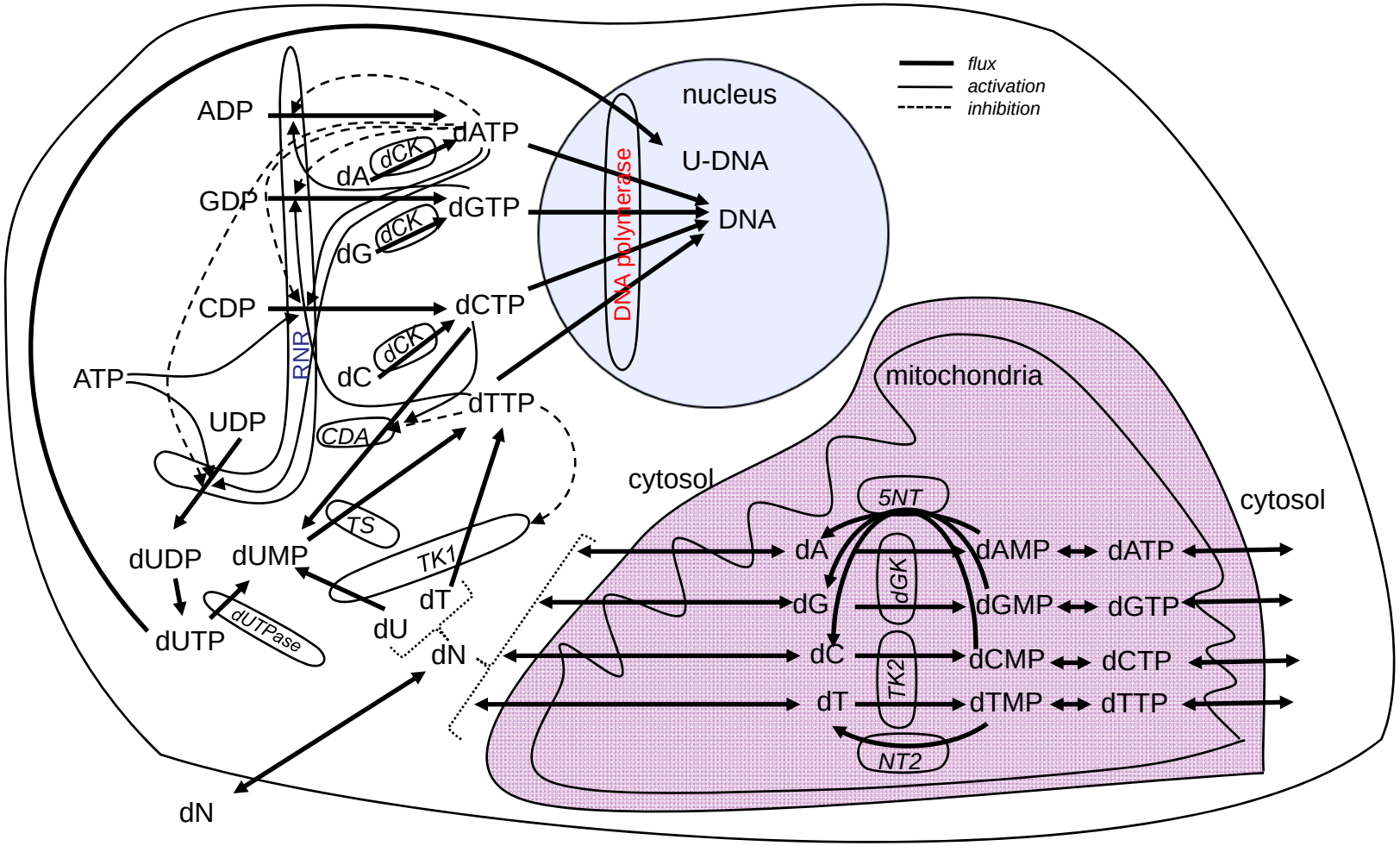


Figure 1. dNTP supply. Many anticancer agents act on or through this system to kill cells. The most central enzyme of this system is RNR.

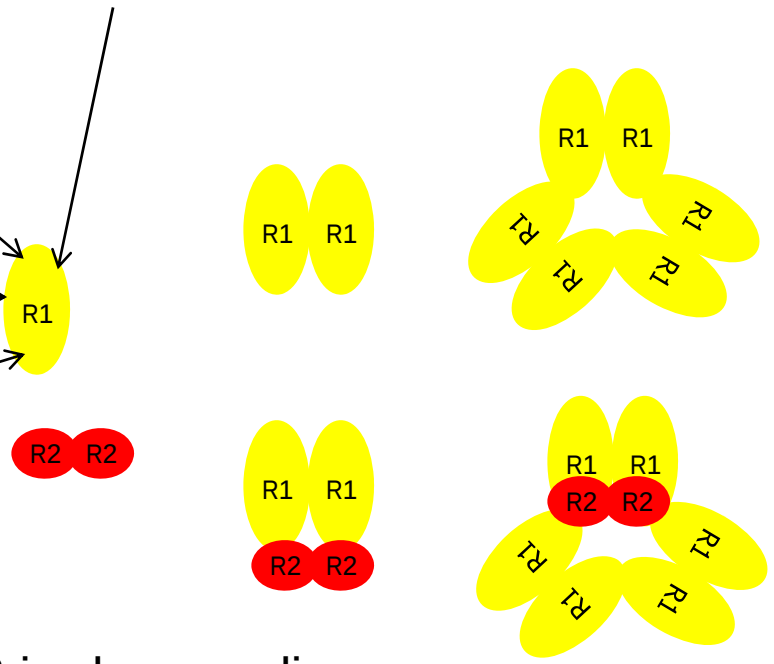
RNR Literature

ATP activates at hexamerization site??

dATP inhibits at activity site, ATP activates at activity site?

dTTP, dGTP, dATP, ATP bind to selectivity site

UDP, CDP, GDP, ADP bind to catalytic site



Selectivity site binding promotes R1 dimers. R2 is always a dimer.

ATP drives hexamer. Controversy: dATP drives inactive tetramer vs. inactive hexamer

Controversy: Hexamer binds one $R2_2$ vs. three $R2_2$

Total concentrations of R1, $R2_2$, dTTP, dGTP, dATP, ATP and NDPs control the distribution of R1-R2 complexes and this changes in S, G_1-G_2 and G_0

Michaelis-Menten Model

$$E + S \rightleftarrows ES$$

$$V_{\max}\left(\frac{[S]}{[S]+K_m}\right) = k_{cat}E_0\left(\frac{[S]/K_m}{[S]/K_m+1}\right) + 0E_0\left(\frac{1}{[S]/K_m+1}\right)$$

$$= k_{cat}E_0\left(\frac{[ES]}{[ES]+[E]}\right) + 0E_0\left(\frac{[E]}{[ES]+[E]}\right)$$

$$= k_{cat}E_0P(ES) \qquad + 0E_0P(E)$$

With RNR: no NDP and no R2 dimer => k_{cat} of complex is zero. Otherwise, many different R1-R2-NDP complexes can have many different k_{cat} values.

Michaelis-Menten Model
[S] vs. [S_T]

$$0 = [E_T] - [E] - \frac{[E][S]}{K_{d\_E\_S}} \quad (1)$$

$$0 = [S_T] - [S] - \frac{[E][S]}{K_{d\_E\_S}} \quad (2)$$

$$\frac{d[E]}{d\tau} = [E_T] - [E] - \frac{[E][S]}{K_{d\_E\_S}}$$

$$\frac{d[S]}{d\tau} = [S_T] - [S] - \frac{[E][S]}{K_{d\_E\_S}}$$

$$[E](0) = 0, \quad [S](0) = 0$$

Substitute this in here to get a quadratic in [S] which solves as

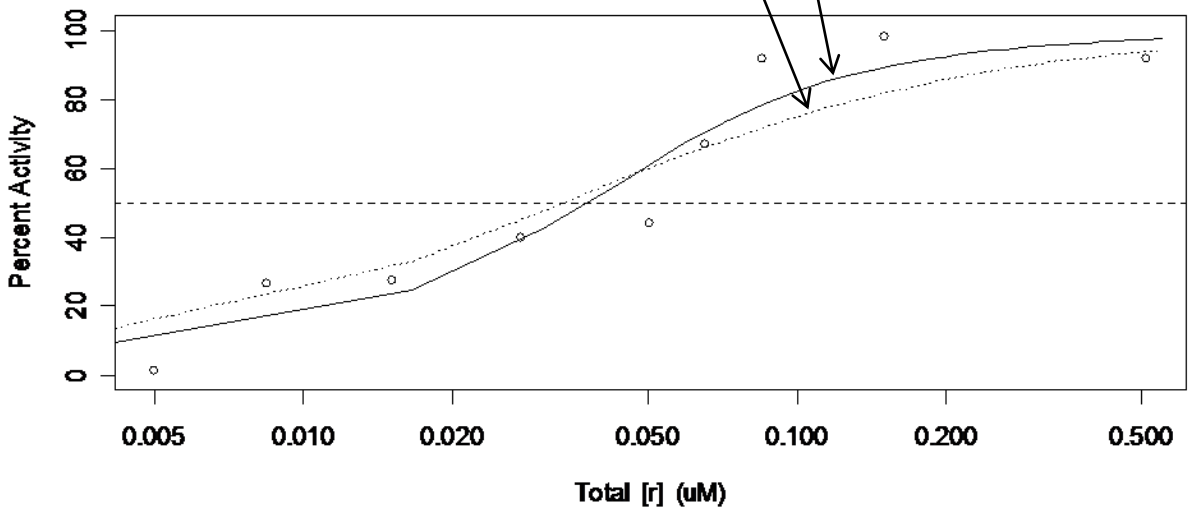
$$[ES] = [S_T] + .5\left((K_{d\_E\_S} - [S_T] + [E_T]\right) \pm \sqrt{\left((K_{d\_E\_S} - [S_T] + [E_T]\right)^2 + 4K_{d\_E\_S}[S_T]}$$

Bigger systems of higher polynomials cannot be solved algebraically => use ODEs (above)

$$[E_T] = [E]\left(1 + \frac{[S]}{K_{d\_E\_S}}\right) \Rightarrow [E] = [E_T]\frac{1}{\left(1 + \frac{[S]}{K_{d\_E\_S}}\right)} \Rightarrow [ES] = [E_T]\frac{\frac{[S]}{K_{d\_E\_S}}}{\left(1 + \frac{[S]}{K_{d\_E\_S}}\right)} \quad \textit{versus} \quad [ES] = [E_T]\frac{\frac{[S_T]}{K_{d\_E\_S}}}{\left(1 + \frac{[S_T]}{K_{d\_E\_S}}\right)} \quad (3)$$

Model	Parameter	Initial Value	Optimal Value	Confidence Interval
RRGGttr1.1.0	RRGGtt_r	0.020	0.012	(0.007, 0.024)
	SSE	1070.252	823.793	
	AIC	45.006	42.650	
MM	Kd	0.020	0.033	(0.022, 0.049)
	SSE	2016.335	1143.682	
	AIC	50.706	45.603	

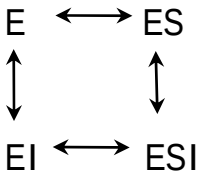
R=R1 G=GDP r=$R2_2$ t=dTTP

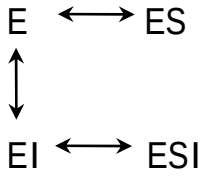


solid line = Eqs. (1-2)
dotted = Eq. (3)

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

Enzyme, Substrate and Inhibitor

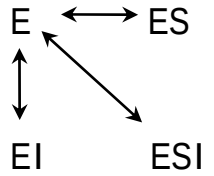




$$0=[E_T]-[E]-\frac{[E][S]}{K_{d\_E\_S}}-\frac{[E][I]}{K_{d\_E\_I}}-\frac{[E][I][S]}{K_{d\_E\_I}K_{d\_EI\_S}}$$

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

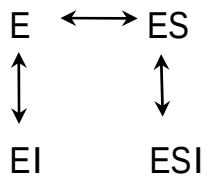
$$0=[I_T]-[I]-\frac{[E][I]}{K_{d\_E\_I}}-\frac{[E][I][S]}{K_{d\_E\_I}K_{d\_EI\_S}}$$



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

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

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

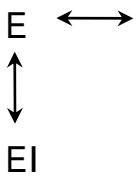


$$0=[E_T]-[E]-\frac{[E][S]}{K_{d\_E\_S}}-\frac{[E][I]}{K_{d\_E\_I}}-\frac{[E][S][I]}{K_{d\_E\_S}K_{d\_ES\_I}}$$

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

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

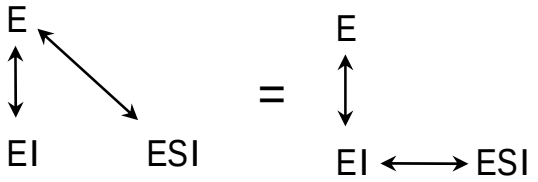
Competitive inhibition



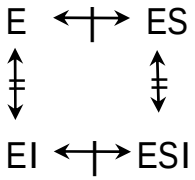
$$0=[E_T]-[E]-\frac{[E][S]}{K_{ES}}-\frac{[E][I]}{K_{EI}}$$

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

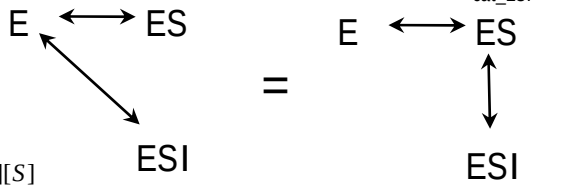
$$0=[I_T]-[I]-\frac{[E][I]}{K_{EI}}$$

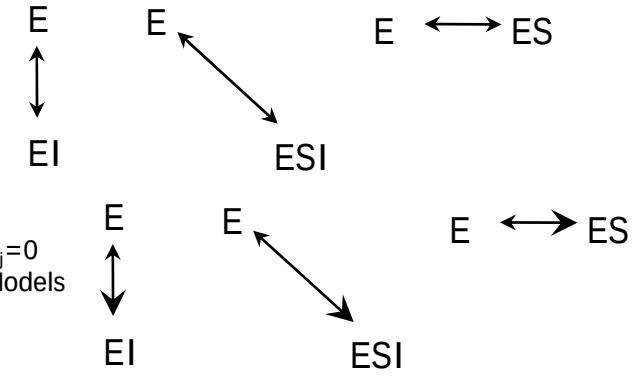


noncompetitive inhibition
Example of $K_d=K_d'$ Model



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





E

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

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

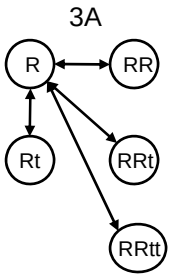
$$0=[I_T]-[I]$$

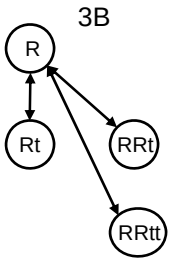
as K_{ES} approaches 0

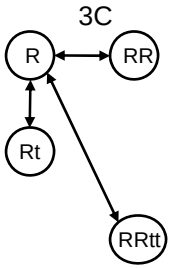
Let p be the probability that an E molecule is undamaged. Then in each model $[E_T]$ can be replace with $p[E_T]$ to double the number of models to $2*(2^3+3+1)=24$.

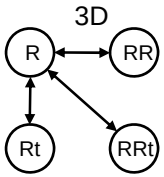
if $([E_T]>[S_T])$ $[ES]=[S_T]$, $[E]=[E_T]-[S_T]$ and $[S]=0$, else $[ES]=[E_T]$, $[S]=[S_T]-[E_T]$ and $[E]=0$

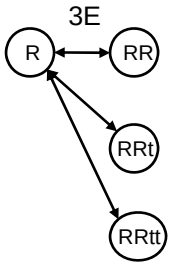
Rt Spur Graph Models

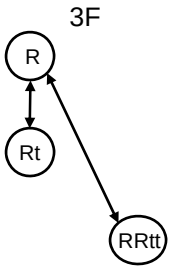


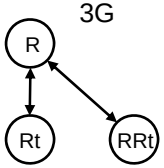


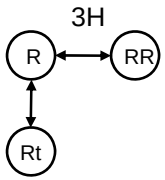


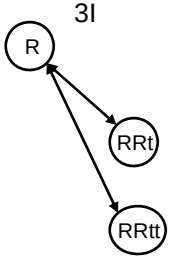


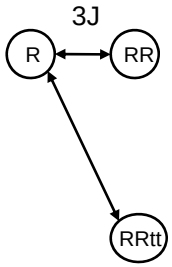


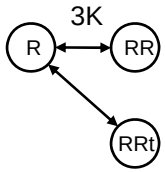


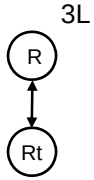


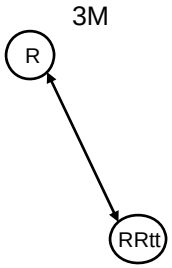


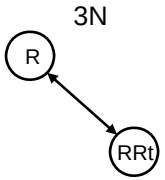


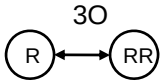


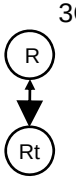


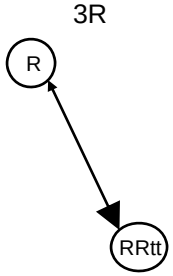


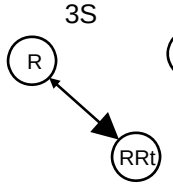


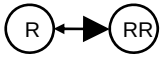












$$0 = p[R_T] - [R] - \frac{[R][t]}{K_{Rt}} - 2\frac{[R]^2}{K_{RR}} - 2\frac{[R]^2[t]}{K_{RRt}} - 2\frac{[R]^2[t]^2}{K_{RRtt}}$$

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

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

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

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

Total number of spur graph models is 16+4=20

Radivoyevitch, (2008) BMC Systems Biology 2:15

Rt Grid Graph Models

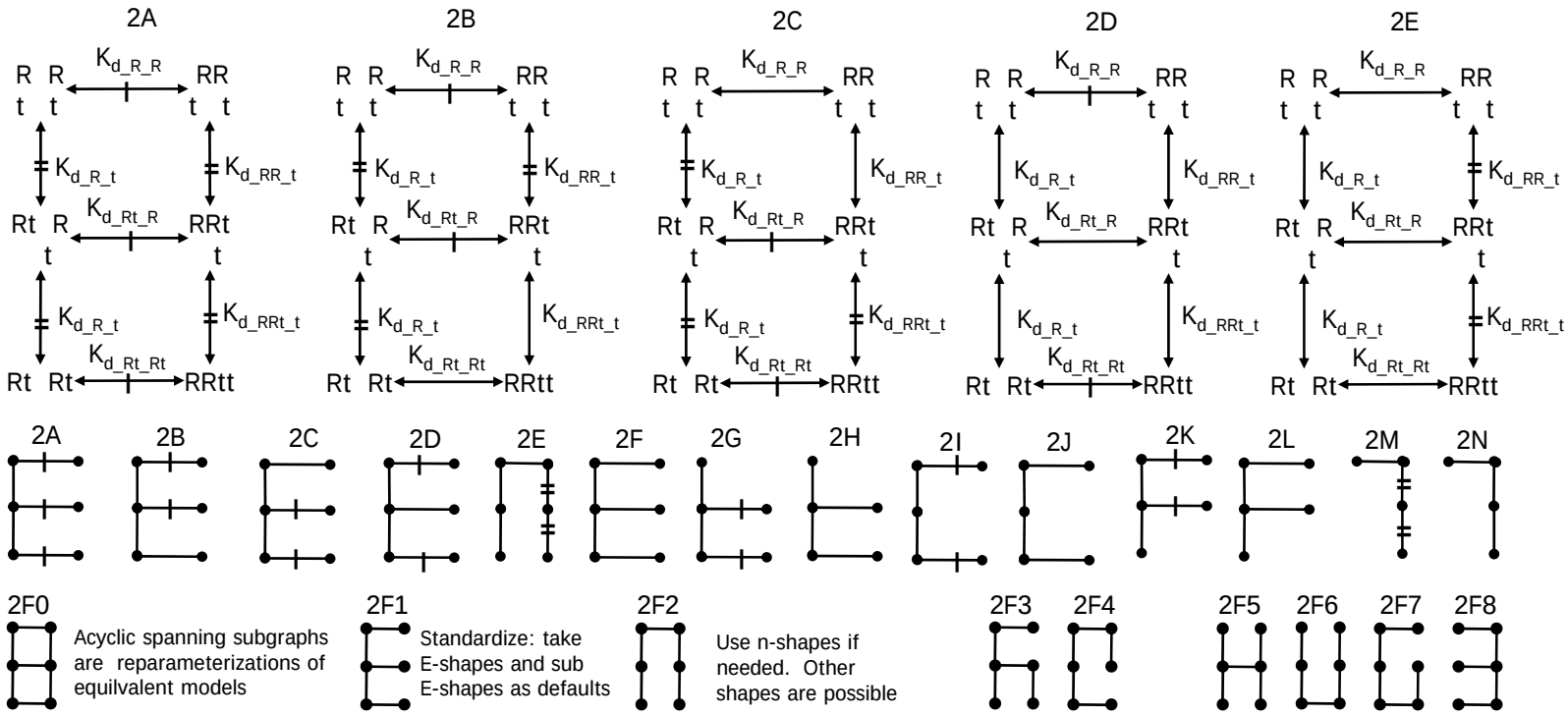


Figure 2. Grid graph models.

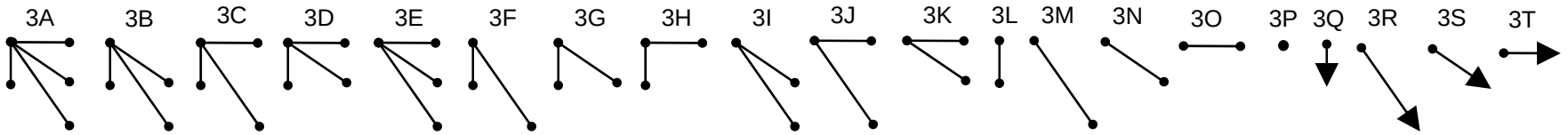


Figure 3. Spur graph models. The following models are equivalent: 3A=2F, 3B=2H, 3C=2J, 3D=2L, 3E=2N

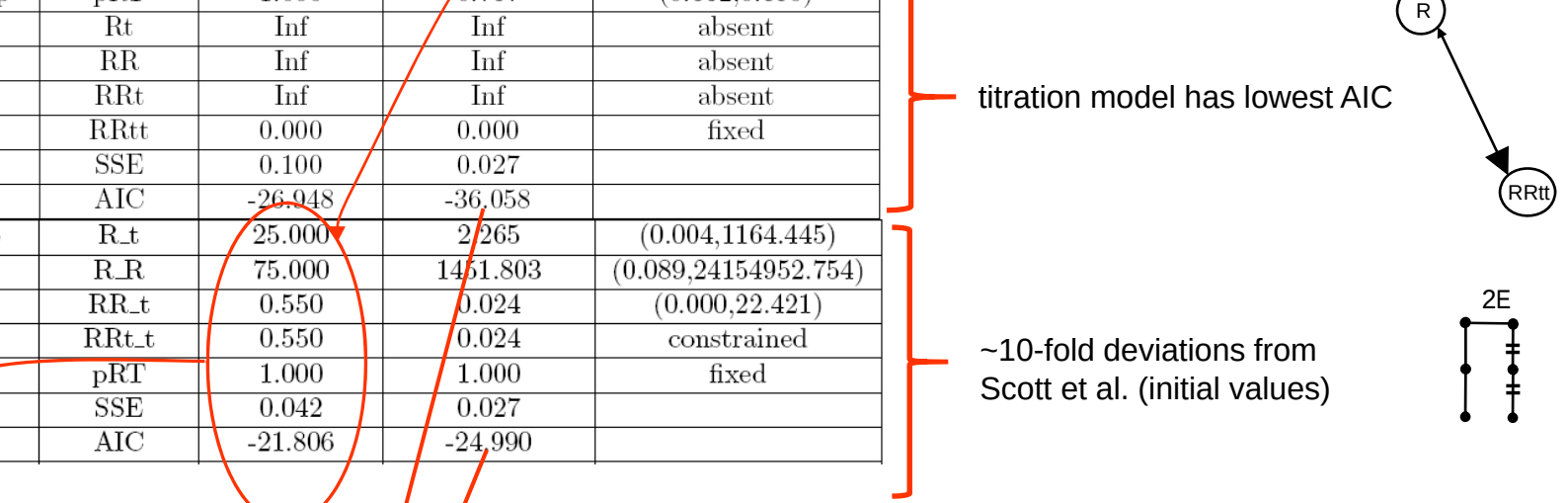
Application to Data

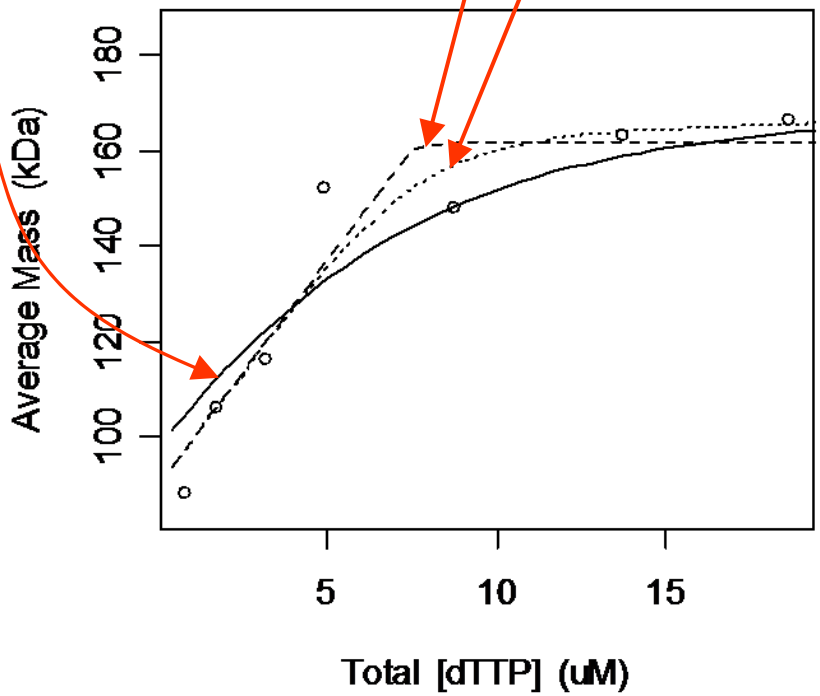
Data and fit from Scott, C. P., Kashlan, O. B., Lear, J. D., and Cooperman, B. S. (2001) *Biochemistry* **40**(6), 1651-166

Model	Parameter	Initial Value	Optimal Value	Confidence Interval
3Rp	pRT	1.000	0.767	(0.662,0.890)
	Rt	Inf	Inf	absent
	RR	Inf	Inf	absent
	RRt	Inf	Inf	absent
	RRtt	0.000	0.000	fixed
	SSE	0.100	0.027	
	AIC	-26.948	-36.058	
2E	R_t	25.000	2.265	(0.004,1164.445)
	R_R	75.000	1451.803	(0.089,24154952.754)
	RR_t	0.550	0.024	(0.000,22.421)
	RRt_t	0.550	0.024	constrained
	pRT	1.000	1.000	fixed
	SSE	0.042	0.027	
	AIC	-21.806	-24.990	

titration model has lowest AIC

~10-fold deviations from Scott et al. (initial values)





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

Infinitely tight binding situation wherein free molecule annihilation (the initial linear ramp) continues in a one-to-one fashion with increasing $[dTTP]_T$ until $[dTTP]_T$ equals $[R1]_T$ =7.6 μM, the plateau point where R exists solely as RRtt.

Experiment becomes a titration scan of $[t_T]$ to estimate $[R_T]$, but $[R_T]$=7.6 μM was already known.

$$M_a = 90\frac{[R]+[R_T](1-p)}{[R_T]}+180\frac{2[RR]+2[RRt]+2[RRtt]}{[R_T]}$$

Radivoyevitch, (2008) BMC Systems Biology 2:15

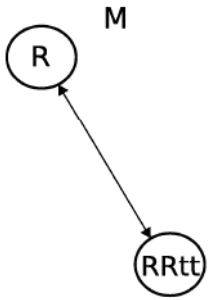
Model Space Fit with New Data

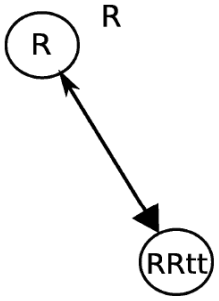
Table 3 – Rofougaran's R1 dimerization data

R_T	t_T	Dimer	Monomer	Average Mass
2.700	100	18100	910	175.692
0.135	100	693	98	168.850
2.700	0	935	19766	94.065

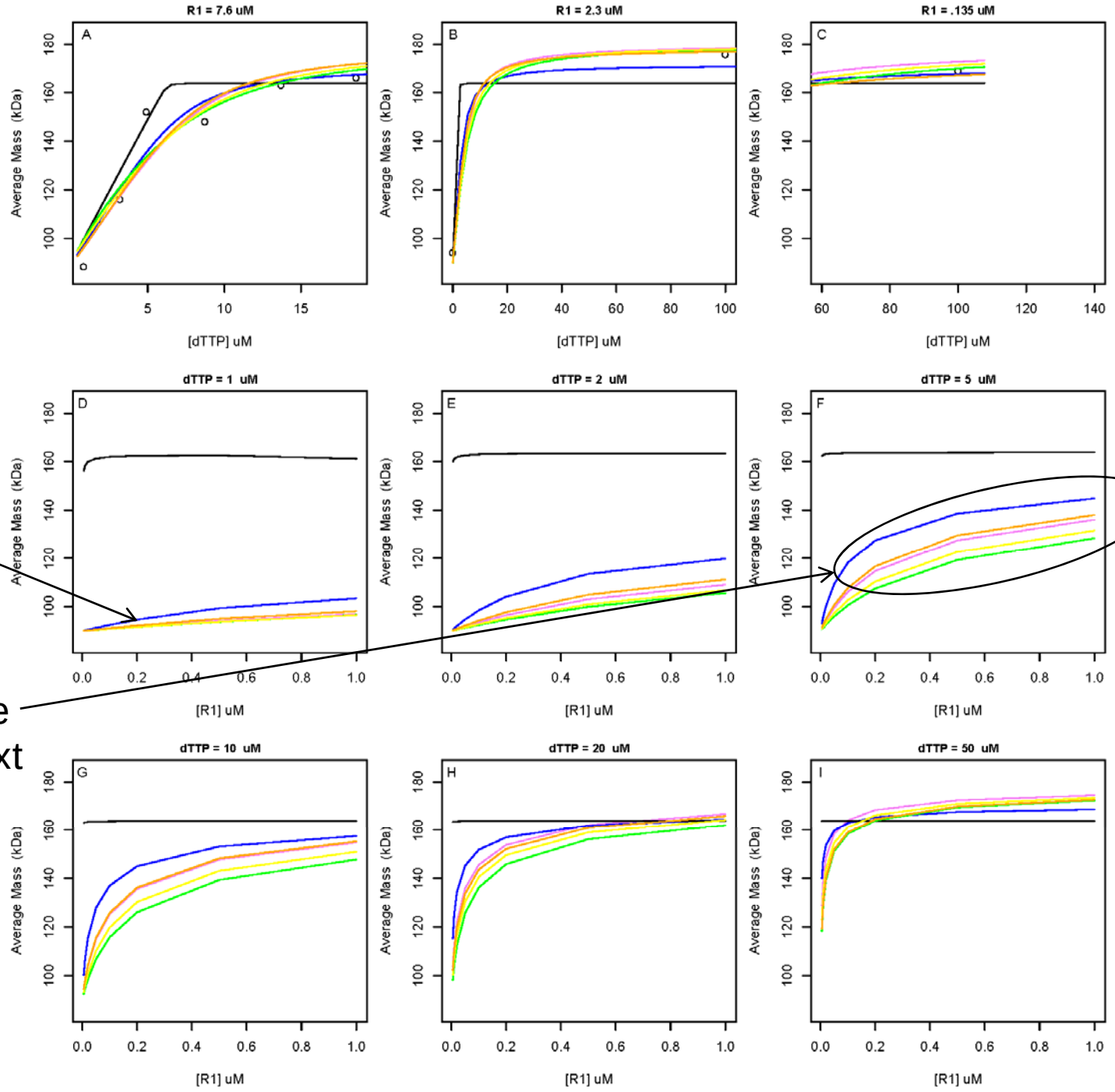
Table 4 – Joint Data Analysis

Model	Parameter	Initial Value	Optimal Value	Confidence Interval
3M	RRtt	1.000	18.697	(4.807,72.966)
	Rt	Inf	Inf	absent
	RR	Inf	Inf	absent
	RRt	Inf	Inf	absent
	pRT	1.000	1.000	fixed
	SSE	0.064	0.034	
	AIC	-48.066	-54.448	
	cpu	0.000	0.445	fit succeeded
3Mp	RRtt	1.000	5.558	(0.370,83.931)
	pRT	1.000	0.907	(0.787,1.044)
	Rt	Inf	Inf	absent
	RR	Inf	Inf	absent
	RRt	Inf	Inf	absent
	SSE	0.064	0.027	
	AIC	-44.852	-53.308	
	cpu	0.000	0.199	fit succeeded
3Rp	pRT	1.000	0.822	(0.736,0.918)
	Rt	Inf	Inf	absent
	RR	Inf	Inf	absent
	RRt	Inf	Inf	absent
	RRtt	0.000	0.000	fixed
	SSE	0.106	0.041	
	AIC	-42.954	-52.590	
	cpu	0.000	0.104	fit succeeded





Radivoyevitch, (2008) BMC Systems Biology 2:15

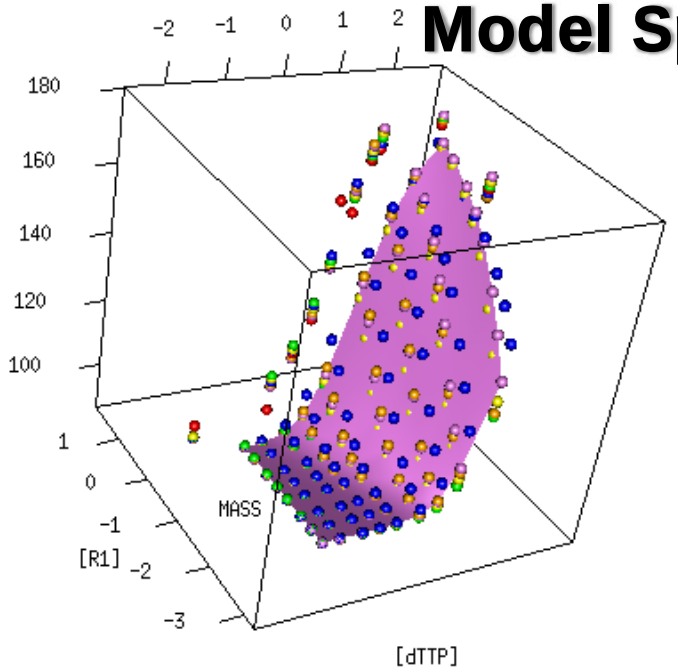


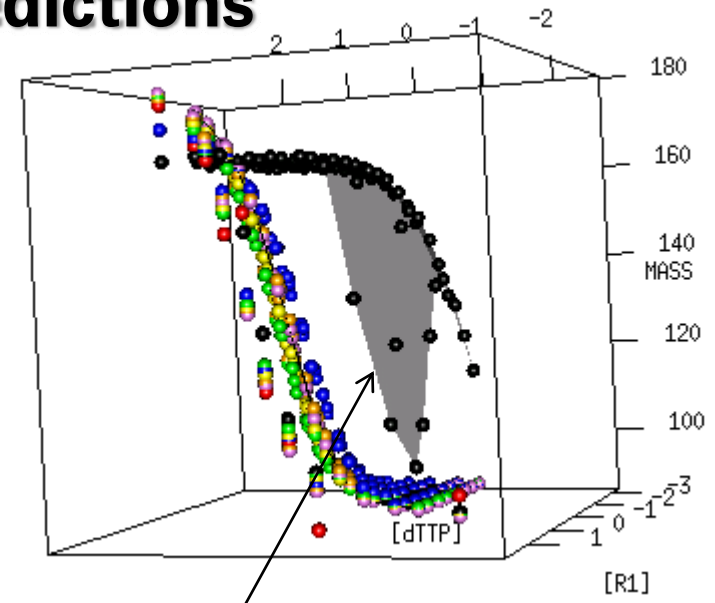
One additional data point here would reject 3Rp

If so, new data here would be logical next

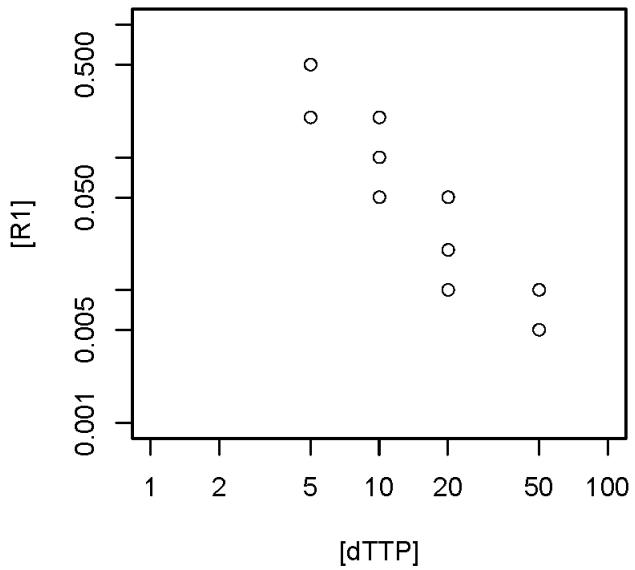
No need to constrain data collection to such profiles

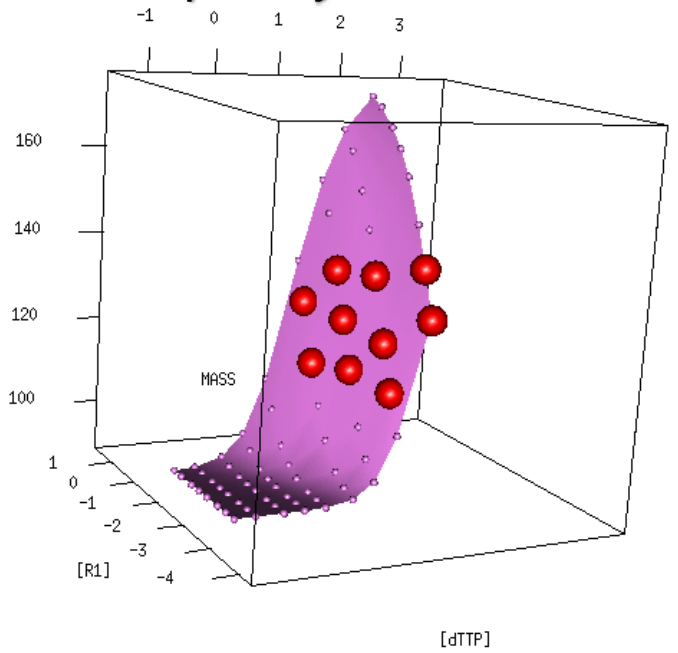
Model Space Predictions





Best next 10 measurements if 3Rp is rejected





Final Remarks

- Fast Total Concentration Constraint (TCC; i.e. *g*=0) solvers are critical to model estimation/selection. TCC ODEs (#ODEs = #reactants) solve TCCs faster than k_{on} =1 and k_{off} = K_d systems (#ODEs = #species = high # in combinatorially complex situations)

- Semi-exhaustive approach = fit all models with same number of parameters as parallel batch, then fit next batch only if current shows AIC improvement over previous batch. This reduces Rt model space fitting times by a factor of 5.

- The best of a best-guess lot of ~10 models may be adequate in many cases

Acknowledgements

- Case Comprehensive Cancer Center
- NIH (K25 CA104791)
- Anders Hofer (Umea)
- Thank you