Automated model generation and selection methods for combinatorially complex biochemical equilibriums

Tom Radivoyevitch[1,2,*]

1. Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio 44106 USA
2. Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, Ohio 44106 USA

* Contact author: txr24@case.edu

Keywords: Enzymes, Rate Laws, Systems Biology, Model Selection

Background: Biochemical equilibrium models can be generated from a full model via hypotheses that some dissociation constants *K* are infinite and/or that two or more *K* are equal. For example, in enzyme-substrate-inhibitor (ESI) equilibriums, competitive inhibition models hypothesize that *K* for ESI is infinite and non-competitive inhibition models hypothesize that *K* for E_S equals *K* for EI_S (Fig. 1). In combinatorially complex systems, the number of plausible protein complexes is large relative to the number of reactants, and far more *K* infinity and equality hypotheses arise than can be specified by hand. Automated model generation and selection methods are needed for these situations.

Results: Biochemical equilibrium models of ATP-induced ribonucleotide reductase R1 hexamerization were generated via *K* infinity and equality hypotheses from a full model that included three (*s*, *a* and *h*) ATP binding sites on R1. Assuming, based on the crystal structure of yeast R1 dimers [*PNAS* 2006, **103**, 4022-4027], that the *s*-site is created at the R1 dimer interface, it is reasonable to assume that R1 oligomer *s*-sites are always fully occupied (i.e. that oligomers cannot form without full *s*-site occupancy) and that R1 monomer *s*-sites are always unoccupied (i.e. that the *s*-site does not exist in R1 monomers). With ATP and R1 denoted by X and R, respectively, the full spur graph system equations are then

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

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

where the T denotes totals and a lack thereof denotes free concentrations. The number of complexes represented is thus 2 + 5 + 9 + 13 = 29 and this implies 2^{29} = ~500 million spur models. Not all of these models need to be fitted, however, as one can first fit the 29 single edge models, then the $\binom{29}{2}=406$ two edge models, then the $\binom{29}{3}=3654$ three parameter models, etc., stopping once the lowest AIC of the current batch is greater than the lowest AIC of the previous batch. Using this approach to analyze recent dynamic light scattering data [*Biochemistry* 2002, **41,** 462-474], assuming *h*-sites are filled only after all of the *a*-sites are filled (and that, in oligomers, these are filled only after all of the *s*-sites are filled), Figure 2 shows that the best models (those with the lowest Akaike Information Criterion) do not support the existence of an *h*-site.

Conclusions: Automated model space generation and analysis methods for combinatorially complex biochemical equilibriums in which the number of models is too large to enumerate by hand, can be realized. Such methods let data speak. They are important because they can lead to inferences that might otherwise be missed.

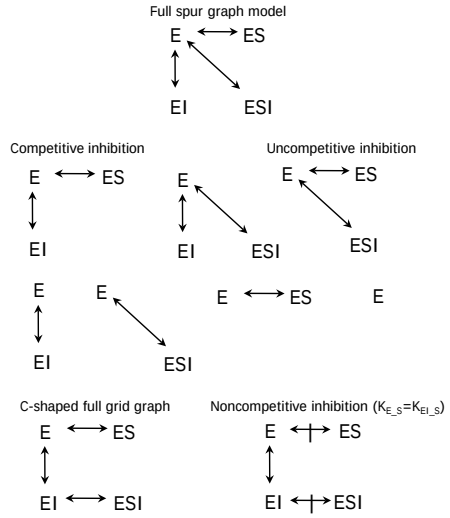

Figure 1. ESI models/graphs. The full spur graph at the top generates the seven models/graphs below it via hypotheses taken one at a time, two at a time, etc, that dissociation constants are infinite. The C-shaped grid graph is a data-fitting equivalent of the full spur graph. It is important because it generates the non-competitive inhibition model where parallel edges (K_d's) are equal.

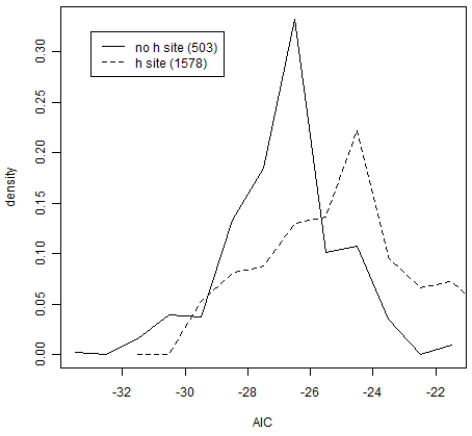

Figure 2. Normalized densities of models with SSEs less than twice the minimum SSE (legend indicates model numbers). Though occupied *h*-site models outnumber unoccupied *h*-site models 3 to 1, the latter make up 28 of the top 30 models and all of the top 5models.

References

Radivoyevitch (2008). Equilibrium model selection: dTTP induced R1 dimerization. *BMC Systems Biology*, **2**, 15

Radivoyevitch (2008). *Equilibrium Model Selection*, User!2008 (Dortmund, Germany), July 2008, p 151

Radivoyevitch (2009). *Combinatorially Complex Equilibrium Model Selection*, http://epbi-radivot.cwru.edu/ccems/overview.html