

MASARYKOVA UNIVERZITA  
INNOVATION LECTURES (INNOIEC) www.muni.cz

**Binding and Kinetics for Experimental Biologists**  
Lecture 8  
**Optimal design of experiments**

Petr Kuzmič, Ph.D.  
BioKin, Ltd.  
WATERTOWN, MASSACHUSETTS, U.S.A.

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.




INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



**Lecture outline**

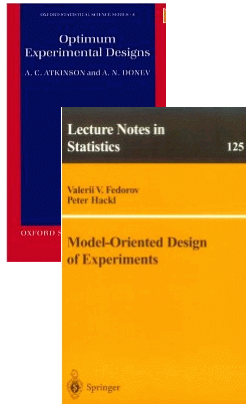
---

- **The problem:**  
How should we **plan an experiment** such we learn the most from it?
- **The solution:**  
Use the **Optimal Design Theory** of statistics
- **An implementation:**  
Software **DynaFit**
- **An example:**  
Kinetics of clathrin cage disassembly

 BKEB Lec 8: Optimal design 2

## Optimal Experimental Design: Where to find basic information

DOZENS OF **BOOKS**



- *Fedorov, V.V. (1972)*  
"Theory of Optimal Experiments"
- *Fedorov, V.V. & Hackl, P. (1997)*  
"Model-Oriented **Design** of Experiments"
- *Atkinson, A.C & Donev, A.N. (1992)*  
"Optimum Experimental **Designs**"
- *Endrenyi, L., Ed. (1981)*  
"**Design** and Analysis of **Enzyme** and Pharmacokinetics Experiments"

## Theory of D-optimal design

MAXIMIZE THE **DETERMINANT ("D")** OF **FISHER INFORMATION MATRIX**

$$y_i = f(x_i, \mathbf{p})$$

$f$	algebraic fitting function
$x_i$	independent variable, $i$ th data point ( $i = 1, 2, \dots, N$ )
$y_i$	dependent variable, $i$ th data point
$\mathbf{p}$	vector of $M$ model parameters

$$s_{i,j} = \frac{\partial f(x_i, \mathbf{p})}{\partial p_j}$$

$s_{i,j}$	sensitivity of $f$ with respect to $j$ th parameter, $i$ th data point
-----------	--

$$F_{j,k} = \sum_{i=1}^N s_{i,j} s_{i,k}$$

$F_{i,j}$	$(i,j)$ th element of the <b>Fisher information matrix</b> $\mathbf{F} = \begin{pmatrix} F_{1,1} & F_{1,2} & \dots & F_{1,M} \\ F_{2,1} & F_{2,2} & \dots & F_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ F_{M,1} & F_{M,2} & \dots & F_{M,M} \end{pmatrix}$
-----------	--

**D-Optimal Design:**

$$\max_{x_1, x_2, \dots, x_N} |\det \mathbf{F}|$$

Choose the independent variable  $x_1, \dots, x_N$  (e.g., total or initial concentrations of reagents) such that the determinant of  $\mathbf{F}$  is maximized.

## D-Optimal design example: Michaelis-Menten equation

RONALD DUGGLEBY – UNIVERSITY OF QUEENSLAND, AUSTRALIA (1979)

*J. Theor. Biol.* **81**, 671-684 (1979)

$$v_i = V \frac{S_i}{S_i + K}$$

$v_i$  initial rate of enzyme reaction,  $i$ th data point  
 $S_i$  substrate concentration,  $i$ th data point ( $i = 1, 2, \dots, N$ )  
 $V, K$  vector of model parameters  
 $K$  ... Michaelis constant,  $V$  ... maximum rate

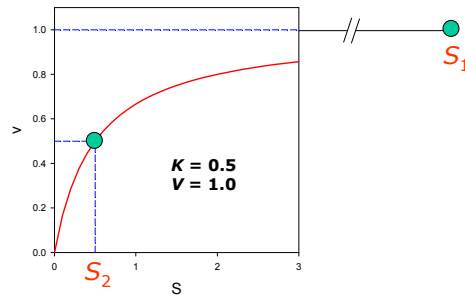
$$s_{i,V} \equiv \frac{\partial v_i}{\partial V} = \frac{S_i}{S_i + K}$$

$$s_{i,K} \equiv \frac{\partial v_i}{\partial K} = -V \frac{S_i}{(S_i + K)^2} \quad \text{sensitivity functions}$$

Box-Lucas **two-point** design:

$$\max_{S_1, S_2} \det \begin{pmatrix} s_{1,V} & s_{1,K} \\ s_{2,V} & s_{2,K} \end{pmatrix}$$

→  $S_1 \rightarrow \infty$   
 $S_2 = K$



BKEB Lec 8: Optimal design

5

## Realistic design for the Michaelis-Menten equation

"INFINITE" SUBSTRATE CONCENTRATION (TO GET  $V_{\max}$ ) IS IMPOSSIBLE TO ACHIEVE

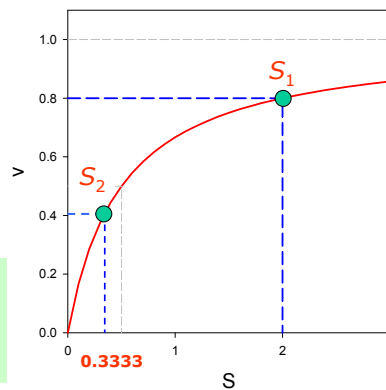
Box-Lucas **two-point** design with one point ( $S_{\max}$ ) already given: **assume  $K = 0.5$ ,  $S_{\max} = 2.0$**

$$\max_{S_2} \det \begin{pmatrix} S_{\max,V} & S_{\max,K} \\ S_{2,V} & S_{2,K} \end{pmatrix}$$

→  $S_1 = S_{\max}$   
 $S_2 = \frac{S_{\max} K}{S_{\max} + 2K}$

### RECIPE

In the determination of  $K_M$ , always **include a substrate concentration** that corresponds to a reaction rate approximately **one half of maximum achievable rate.**

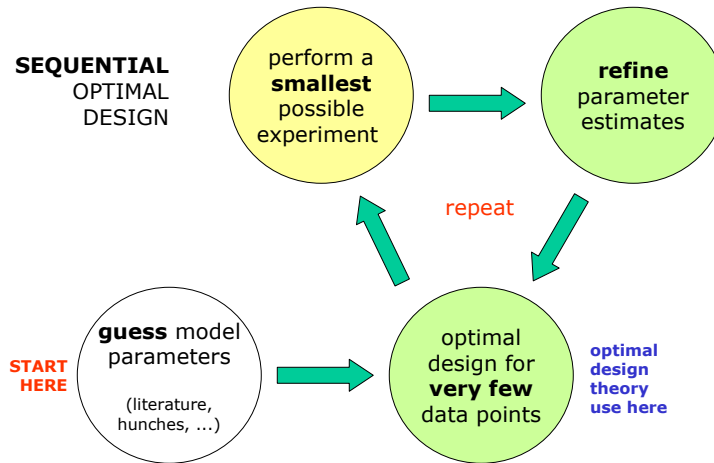


BKEB Lec 8: Optimal design

6

## Theory of D-optimal design: The chicken-and-egg problem

**PROBLEM:** TO DESIGN AN EXPERIMENT WE MUST FIRST GUESS THE FINAL ANSWER!



*Special case:* Time-course experiments with fixed time-points

## A typical kinetics experiment: Fixed mesh of time-points

**PROBLEM:** WE DO NOT ALWAYS HAVE A CHOICE OF INDEPENDENT VARIABLE VALUES

Rothnie, Kuzmic, *et al.* (2011) *Proc. Natl. Acad. Sci. USA* **108**, 6927-35

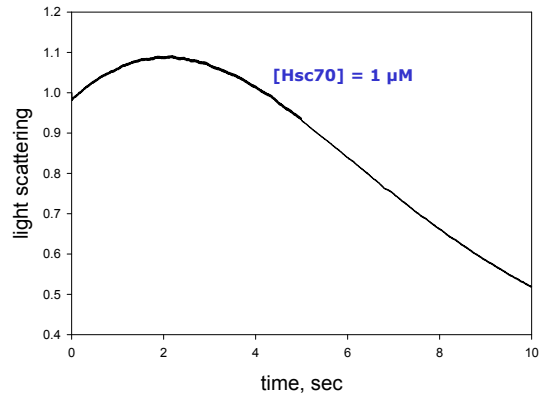
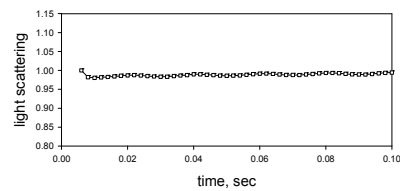


Fig. 4. Stopped-flow measurements of light scattering to examine the early stages of clathrin cage disassembly. Clathrin cages (0.09 μM triskelia) pre-mixed with 0.1 μM auxilin were mixed with Hsc70 (concentrations in μM shown on graph) and 500 μM ATP, and perpendicular light scattering was measured using stopped-flow techniques.

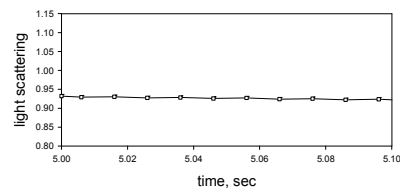
## Stopped flow experimentation: Fixed step size

**AT BEST, WE CAN CHANGE THE (FIXED) STEP SIZE FOR A GIVEN INTERVAL**

Rothnie, Kuzmic, *et al.* (2011) *Proc. Natl. Acad. Sci. USA* **108**, 6927-35, Fig. 4



$t = 0$  to 5 sec:  
step size 20 msec



$t = 5$  to 10 sec:  
step size 100 msec

## Variation of the design problem: Optimize initial conditions

IF WE CAN'T CHOOSE OBSERVATION TIME, AT LEAST WE CAN CHOOSE **INITIAL CONCENTRATIONS**

### BASIC PRINCIPLE:

- In *kinetic* studies
  - the **independent variable** is **time**;
  - **initial concentrations** are considered "parameters" of the model.
- D-Optimal design theory is concerned with optimal choice of **independent variable**
  - in this case the observation time.
- Unfortunately, in the real-world we cannot choose particular observation times:
  - usual instruments are offering us only a fixed mesh of output points.
- But we can turn things around and
  - **treat initial concentrations as "independent variables"**.
  - Then we can optimize the choice of initial concentrations, using the usual formalism of the D-Optimal Design theory.

## Theory of D-optimal design: Initial conditions in ODE systems

MAXIMIZE THE **DETERMINANT ("D") OF FISHER INFORMATION MATRIX**

$$\begin{aligned} dc_i / dt &= f_i(\mathbf{c}, \mathbf{k}) \\ t = 0 &: \mathbf{c} = \mathbf{c}_0 \end{aligned}$$

**initial value problem** (first-order ordinary differential equations)

$\mathbf{c}$  ... vector of concentrations  
 $\mathbf{k}$  ... vector of rate constants  
 $\mathbf{c}_0$  ... concentrations at time zero

$$y_i = g(\mathbf{c}(t_i), \mathbf{r})$$

$y_i$  experimental signal at  $i$ th data point (time  $t_i$ )  
 $\mathbf{c}$  concentrations at time  $t_i$   
 $\mathbf{r}$  vector of molar responses and/or offset on signal axis

$$s_{i,j} = \frac{\partial g(\mathbf{c}(\mathbf{k}, t_i), \mathbf{r})}{\partial p_j}$$

$s_{i,j}$  sensitivity of  $f$  with respect to  $j$ th parameter,  $i$ th data point  
 $\mathbf{p}$  ... model parameters: vectors  $\mathbf{k}$  and  $\mathbf{r}$  combined

$$F_{j,k} = \sum_{i=1}^N s_{i,j} s_{i,k}$$

**D-Optimal Design:**

$$\max |\det \mathbf{F}|$$

## Optimize initial conditions: DynaFit notation

THE SOFTWARE TAKES CARE OF ALL THE MATH

```
[task]
  data = progress
  task = design

[mechanism]
...

[data]
  set ...
  concentration X = 1 ?? (0.01 .. 100)
```

syntax otherwise used for **confidence intervals**

lower and upper bounds must be given

this value is ignored  
(present for syntactical reasons only)

## Optimize initial conditions: Algorithm and DynaFit settings

THE DIFFERENTIAL EVOLUTION ALGORITHM REQUIRES SPECIAL SETTINGS

```
[task]
  data = progress
  task = design

[mechanism]
...

[settings]

(DifferentialEvolution)

  PopulationSizeFixed = 300
  MaximumEvolutions = 1
  MinimumEvolutions = 1
  TestParameterRange = n
  TestParameterRangeAll = n
  TestParameterRangeFull = n
  StopParameterRange = 0.1
  TestCostFunctionRange = y
  StopCostFunctionRange = 0.01
  TestCostFunctionChange = y
  StopCostFunctionChange = 0.00001
  TestCostFunctionChangeCount = 5
```

copy these settings from one of the distributed example problems

population **not too large**

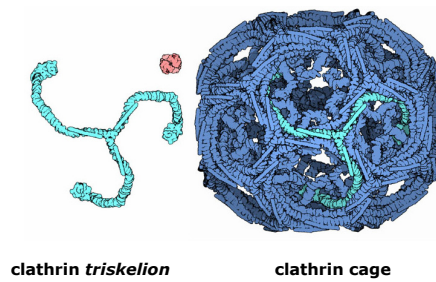
perform the optimization **only once**

relatively weak **convergence criteria**

*Case study:* Kinetics of clathrin cage disassembly

## Clathrin structure: triskelions and cages

CLATHRIN CAGES ARE LARGE ENOUGH TO BE VISIBLE IN **MICROSCOPY** AND **LIGHT SCATTERING**

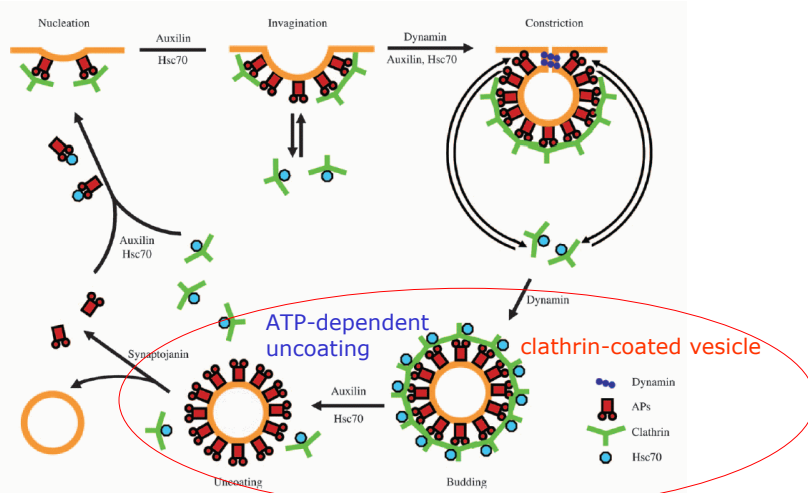




## Clathrin biology: Role in endocytosis

### CLATHRIN IS INVOLVED IN INTRACELLULAR TRAFFICKING

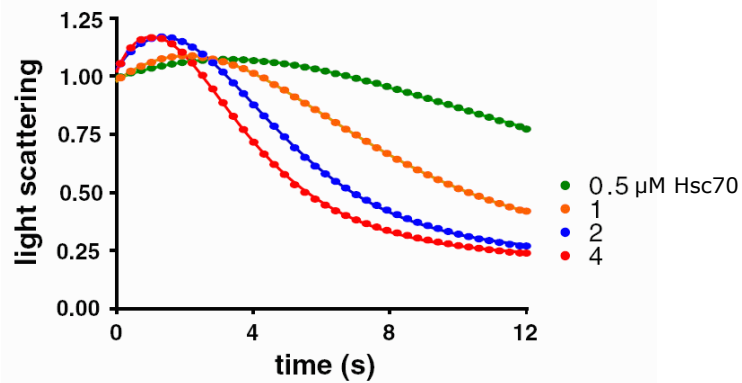
Eisenberg *et al.* (2007) *Traffic* **8**, 640-646



## *In vitro* kinetics of clathrin disassembly: Experimental data

### WATCHING CLATHRIN CAGES TO FALL APART BY PERPENDICULAR LIGHT SCATTERING

Rothnie, Kuzmic, *et al.* (2011) *Proc. Natl. Acad. Sci. USA* **108**, 6927-35, Fig. 4



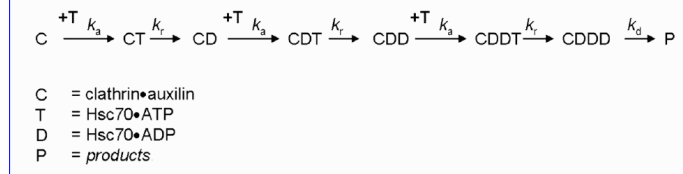
"Clathrin cages (0.09 μM triskelia) premixed with 0.1 μM auxilin were mixed with Hsc70 (concentrations in μM shown on graph) and 500 μM ATP, and perpendicular light scattering was measured using stopped-flow..."

## In vitro kinetics of clathrin disassembly: Theoretical models

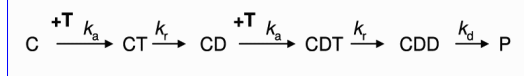
### MODEL SELECTION USING THE AKAIKE INFORMATION CRITERION (DYNAFIT)

Rothnie, Kuzmic, *et al.* (2011) *Proc. Natl. Acad. Sci. USA* **108**, 6927-35

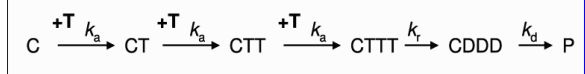
#### THREE-STEP SEQUENTIAL:



#### TWO-STEP SEQUENTIAL:



#### THREE-STEP CONCERTED:



Etc. In total *five* different models were evaluated.



BKEB Lec 8: Optimal design

19

## In vitro kinetics of clathrin disassembly: DynaFit notation

### THE MOST PLAUSIBLE MODEL: THREE STEP SEQUENTIAL

Rothnie, Kuzmic, *et al.* (2011) *Proc. Natl. Acad. Sci. USA* **108**, 6927-35

#### DYNAFIT INPUT:

```
[task]
task = fit
data = progress
model = AHAHAH ?
[mechanism]
CA + T -> CAT : ka
CAT -> CAD + Pi : kr
CAD + T -> CADT : ka
CADT -> CADD + Pi : kr
CADD + T -> CADDT : ka
CADDT -> CADD + Pi : kr
CADD -> Prods : kd
...
[task]
task = fit
data = progress
model = AHAH ?
...
```

#### AUTOMATICALLY GENERATED MATH MODEL:

```
d[CA]/dt = - ka[CA][T]
d[T]/dt = - ka[CA][T] - ka[CAD][T] - ka[CADD][T]
d[CAT]/dt = + ka[CA][T] - kr[CAT]
d[CAD]/dt = + kr[CAT] - ka[CAD][T]
d[Pi]/dt = + kr[CAT] + kr[CADT] + kr[CADDT]
d[CADT]/dt = + ka[CAD][T] - kr[CADT]
d[CADD]/dt = + kr[CADT] - ka[CADD][T]
d[CADDT]/dt = + ka[CADD][T] - kr[CADDT]
d[CADD + Pi]/dt = + kr[CADDT] - kd[CADD + Pi]
d[Prods]/dt = + kd[CADD + Pi]
```



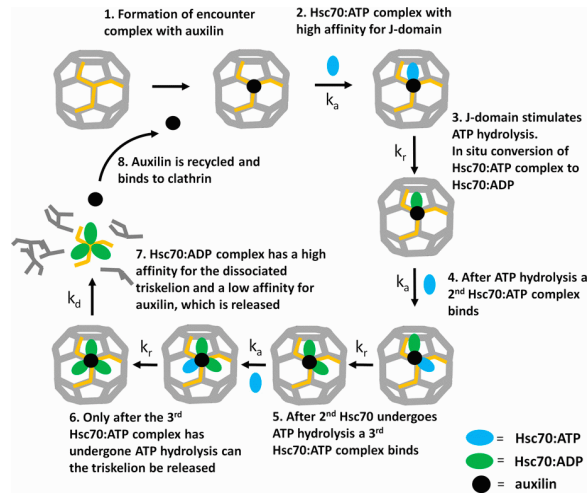
BKEB Lec 8: Optimal design

20

## In vitro kinetics of clathrin disassembly: Preferred mechanism

CONCLUSIONS: **THREE ATP MOLECULES MUST BE HYDROLYZED BEFORE THE CAGE FALLS APART**

Rothnie, Kuzmic, *et al.* (2011) *Proc. Natl. Acad. Sci. USA* **108**, 6927-35

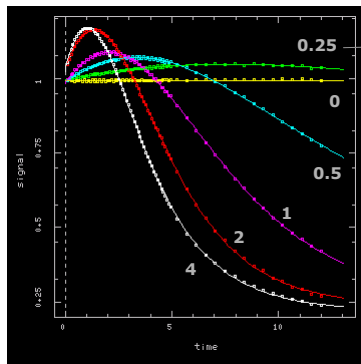


## In vitro kinetics of clathrin disassembly: Raw data

THIS WAS A **VERY EXPENSIVE EXPERIMENT TO PERFORM**

Rothnie, Kuzmic, *et al.* (2011) *Proc. Natl. Acad. Sci. USA* **108**, 6927-35

ACTUAL EXPERIMENTAL DATA:



[Hsc70]

- six assays / experiment
- 90 nM clathrin
- 100 nM auxilin
- up to 4  $\mu$ M Hsc70

a lot of material

expensive and/or  
time consuming  
to obtain

## How many assays are actually needed?

### D-OPTIMAL DESIGN IN DYNAFIT

#### [task]

```
task = design
data = progress
```

#### [mechanism]

```
CA + T -> CAT      : ka
CAT -> CAD + Pi    : kr
CAD + T -> CADT    : ka
CADT -> CADD + Pi  : kr
CADD + T -> CADDT  : ka
CADDT -> CADDD + Pi : kr
CADDD -> Prods     : kd
```

#### [constants]

```
ka = 0.69 ?
kr = 6.51 ?
kd = 0.38 ?
```

"Choose eight initial concentration of **T** such that the rate constants  $k_a$ ,  $k_r$ ,  $k_d$  are determined most precisely."

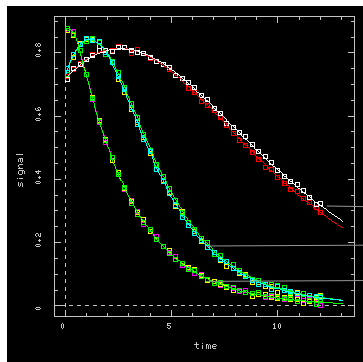
#### [data]

```
file run01 | concentration CA = 0.1 T = 1 ?? (0.001 .. 100)
file run02 | concentration CA = 0.1 T = 1 ?? (0.001 .. 100)
file run03 | concentration CA = 0.1 T = 1 ?? (0.001 .. 100)
file run04 | concentration CA = 0.1 T = 1 ?? (0.001 .. 100)
file run05 | concentration CA = 0.1 T = 1 ?? (0.001 .. 100)
file run06 | concentration CA = 0.1 T = 1 ?? (0.001 .. 100)
file run07 | concentration CA = 0.1 T = 1 ?? (0.001 .. 100)
file run08 | concentration CA = 0.1 T = 1 ?? (0.001 .. 100)
```

## Optimal Experimental Design: DynaFit results

SURPRISE: WE DID TOO MUCH WORK FOR THE INFORMATION GAINED

### SIMULATED DATA – OPTIMAL EXPERIMENT:



#### D-Optimal initial concentrations:

```
[T] = 0.70 μM, 0.73 μM
[T] = 2.4 μM, 2.5 μM, 2.5 μM
[T] = 76 μM, 81 μM, 90 μM
```

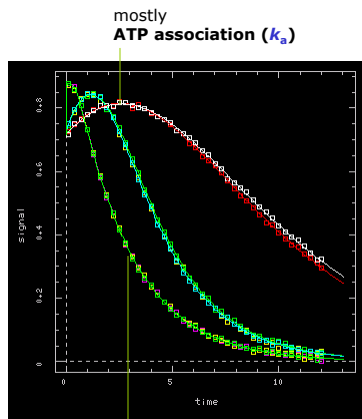
"maximum feasible concentration"  
upswing phase no longer seen

Just **three** experiments would be sufficient for follow-up!

**One half of the material** compared to the original experiment.

## Optimal Experimental Design: DynaFit results - discussion

EACH OF THE **THREE** UNIQUE ASSAYS TELLS A DIFFERENT "STORY"



```
[mechanism]
CA + T -> CAT      : ka
CAT -> CAD + Pi    : kr
CAD + T -> CADT    : ka
CADT -> CADD + Pi  : kr
CADD + T -> CADDT  : ka
CADDT -> CADD + Pi : kr
CADD -> Prods     : kd
```

association ("upswing") is  
no longer visible

## Optimal Experimental Design in DynaFit: Summary

**NOT A SILVER BULLET !**

- Useful for **follow-up (verification)** experiments only
  - Mechanistic model must be known already
  - Parameter estimates must also be known
- Takes a **very long time** to compute
  - Constrained global optimization: "Differential Evolution" algorithm
  - Clathrin design took 30-90 minutes
  - Many design problems take multiple hours of computation
- **Critically** depends on assumptions about **variance**
  - Usually we assume **constant variance** ("noise") of the signal
  - Must verify this by plotting **residuals against signal** (not the usual way)