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

Binding and Kinetics for Experimental Biologists

Lecture 4

Equilibrium Binding: Case Study

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

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.
EVROPSKÁ UNIE esf MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

BioKin L t d.

Lecture outline

- Topics:
  - generalized numerical model for equilibrium binding data
  - *PREVIEW:* model discrimination analysis (Akaike Information Criterion, AIC)
  - representing equilibrium binding mechanisms in *DynaFit*:
    - the "thermodynamic box";
    - exclusive vs. non-exclusive binding;
    - interacting vs. non-interacting binding sites.
- Example:

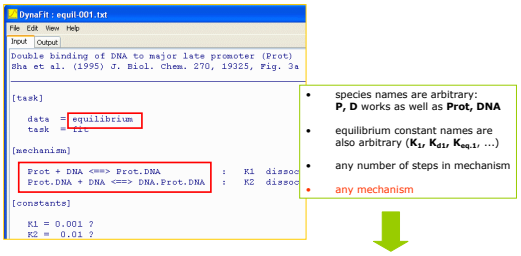
HIV-1 Rev responsible element (RRE) RNA sequence interacting with

(a) a model peptide representing the Rev protein
(b) Neomycin B as a potential Rev competitor

Goal: determine molecular mechanism – "Rev" and "Neo" mutually exclusive?

DynaFit: Analysis of complex equilibria

UNIFORM USER INTERFACE: SYMBOLIC DESCRIPTION OF REACTION MECHANISM



- species names are arbitrary: **P, D** works as well as **Prot, DNA**
- equilibrium constant names are also arbitrary (K_1, K_{d1}, $K_{eq,1}$, ...)
- any number of steps in mechanism
- any mechanism

DynaFit automatically derives the underlying mathematical model

DynaFit: Mathematical model for complex equilibria

"UNDER THE HOOD": A SYSTEM OF SIMULTANEOUS NONLINEAR ALGEBRAIC EQUATIONS

DynaFit uses a modification of algorithm "EQS" by W.R. Smith (1990)

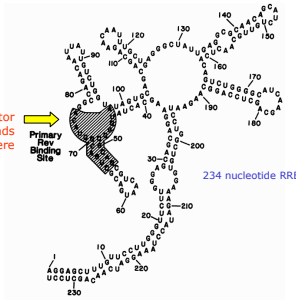
MATHEMATICAL DETAILS:

Royer, C.A.; Smith, W.R.; and Beechem, J.M. (1990)
"Analysis of binding in macromolecular complexes: A generalized numerical approach"
Anal. Biochem., **191**, 287-294.

Royer, C.A. and Beechem, J.M. (1992)
"Numerical analysis of binding data: advantages, practical aspects, and implications"
Methods Enzymol. **210**, 481-505.

Example: HIV-1 Rev response element (RRE)

Rev REGULATES THE TRANSCRIPTION OF HIV-1 REGULATORY PROTEINS



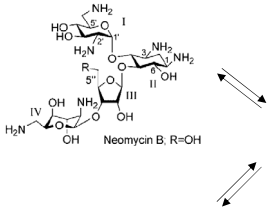
Rev *trans*-activator protein binds near here

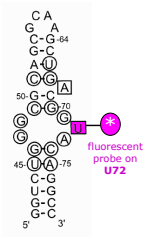
234 nucleotide RRE RNA target sequence

Cullen (1991) *FASEB J.* **5**, 2361-8

HIV-1 RRE / Rev / Neomycin B

NEOMYCIN BINDS TO Rev RESPONSIBLE ELEMENT. COULD IT **DISRUPT THE BINDING OF Rev?**





Rev model peptide:

Suc-TRQARRNRRRRWRERQRAAAAK

Lacourciere *et al.* (2000) *Biochemistry* **39**, 5630-41

HIV-1 RRE / Rev / Neomycin B – study plan

1. Experiment #1: Observe the binding of RRE to Rev
2. Experiment #2: Observe the binding of RRE to Neomycin
3. Experiment #3: Observe the binding of RRE to Rev + Neomycin
4. Compare the observations with two alternate mechanisms:

   a. **Neomycin competes with Rev** peptide ...
   b. **Neomycin does *not* compete with Rev** peptide ...

   ... for binding to the fluorescently labeled RNA fragment
5. Conclude which of the two models is more likely to be true

DynaFit script: Skeleton for fitting equilibrium data

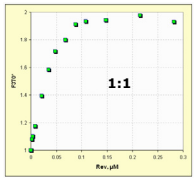
EVERY DYNAFIT SCRIPT HAS TO CONTAIN THESE **SECTIONS**

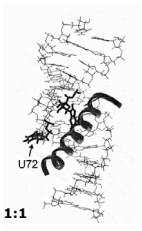
```
[task]
   task = fit
   data = equilibria

[mechanism]          ← definitions of equilibrium constants

[constants]          ← numerical estimates of equilibrium constants

[concentrations]     ← concentrations of reactants applicable to all data sets

[responses]          ← molar response coefficients (e.g., UV/Vis extinction coefficients)

[data]
   variable ...      ← which component is varied in the binding experiment
   set      ...      ← where to find the experimental data (not the data themselves)

[output]
   directory ...

[set:...]            ← experimental data

[end]
```


Experiment #1: DynaFit script - mechanism

NOTHING SPECIAL – JUST SIMPLE 1:1 BINDING

```
[mechanism]

   R72 + Rev <===> R72.Rev     :     K     dissoc
```





Lacourciere *et al.* (2000) *Biochemistry* **39**, 5630-41

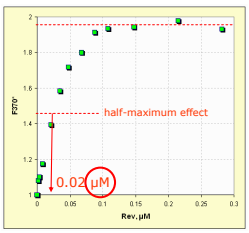
Experiment #1: DynaFit script - constants

LOOK FOR "HALF-MAXIMUM CONCENTRATION" TO ESTIMATE DISSOCIATION CONSTANTS

```
[mechanism]

   R72 + Rev <===> R72.Rev     :     K     dissoc

[constants]

   K = 0.02
```

dissociation constants have the **same dimension as concentrations**

units must be the **same as** those used in the **experimental data!**



Lacourciere *et al.* (2000) *Biochemistry* **39**, 5630-41

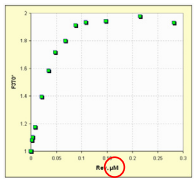
Experiment #1: DynaFit script - concentrations

LIST ONLY **CONSTANT** (NOT VARIABLE) CONCENTRATIONS **IDENTICAL** IN ALL DATA SETS

```
[mechanism]

   R72 + Rev <===> R72.Rev     :     K     dissoc

[constants]

   K = 0.02

[concentrations]

   R72 = 0.03
```

[R72] = 30 nM

units must be the **same as** those used in the **experimental data!**



Lacourciere *et al.* (2000) *Biochemistry* **39**, 5630-41

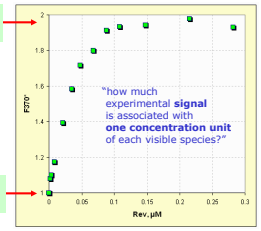
Experiment #1: DynaFit script - responses

LIST **ALL** MOLECULAR SPECIES **"VISIBLE"** IN THE GIVEN EXPERIMENTS

```
[mechanism]

   R72 + Rev <===> R72.Rev     :     K     dissoc

[constants]

   K = 0.02

[concentrations]

   R72 = 0.03

[responses]

   R72     = 33.3
   R72.Rev = 66.6
```

2.0 / **0.03** = 66.6

1.0 / **0.03** = 33.3



Lacourciere *et al.* (2000) *Biochemistry* **39**, 5630-41

Experiment #1: DynaFit script - data

EXPERIMENTAL DATA CAN BE **EMBEDDED** IN THE SCRIPT OR RESIDE IN **SEPARATE FILES**

```
[mechanism]

   R72 + Rev <===> R72.Rev      :      K     dissoc

[constants]

   K = 0.02

[concentrations]

   R72 = 0.03

[responses]

   R72      = 33.3
   R72.Rev = 66.6

[data]

   variable Rev
   set      R72--Rev
```

```
[set:R72--Rev]

Figure 2B in Lacourciere et al. (2000)
Rev,uM   F370*

0.0000   1
0.0020   1.0803
0.0040   1.1005
0.0080   1.1749
0.0213   1.3921
0.0347   1.5824
0.0480   1.7166
0.0680   1.7993
0.0880   1.9123
0.1080   1.9317
0.1480   1.9436
0.2147   1.9781
0.2813   1.9298
```

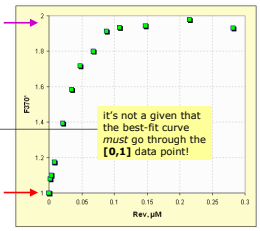
a "comment"

raw data courtesy of Jim Stivers *Johns Hopkins University*

Experiment #1: DynaFit – optimized parameters

WHAT ARE THE "UNKNOWNS" IN THIS EXPERIMENT?

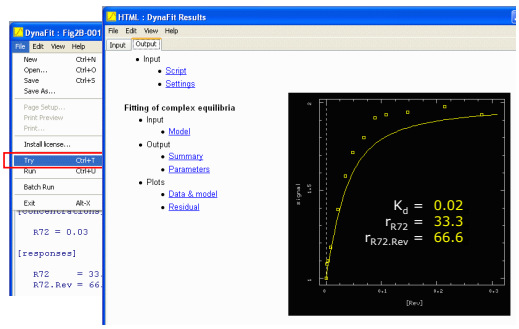
```
[mechanism]

   R72 + Rev <===> R72.Rev      :      K     dissoc

[constants]

   K = 0.02  ?

[concentrations]

   R72 = 0.03

[responses]

   R72      = 33.3 ?
   R72.Rev = 66.6 ?

[data]

   variable Rev
   set      R72--Rev
```



it's not a given that the best-fit curve *must* go through the **[0,1]** data point!

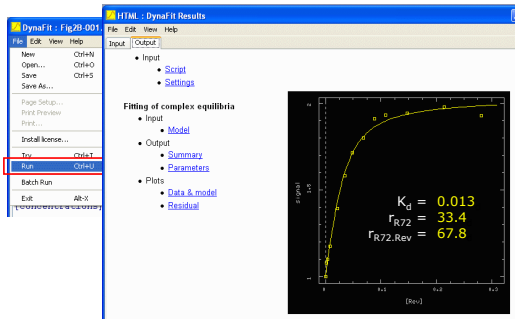
Experiment #1: DynaFit – initial estimate

ALWAYS USE THIS FEATURE TO ASSESS THE QUALITY OF YOUR INITIAL ESTIMATE!



Experiment #1: DynaFit – performing the fit

RUN THE SCRIPT **ONLY** WHEN THE INITIAL ESTIMATE LOOKS **REASONABLY GOOD**!



A devil in the detail: Is our labeled [RNA] correct?

DynaFit output:

Optimized Parameters

No	Par#Set	Initial	Final	Std. Error	CV (%)	Note
#1	K	0.02	0.0126128	0.0021282	16.87	
#2	r(R72)	33.3	33.4412	0.592298	1.77	
#3	r(R72.Rev)	66.6	67.8477	0.868964	1.28	

Special situation: the K_d is *lower* than the (fixed) RNA concentration!

[R72] = **0.030** µM
K_d = **0.013** µM

When the "fixed" concentration is higher than K_d ...

Analytical Biochemistry **286**, 45–50 (2000)

High-Throughput Screening of Enzyme Inhibitors: Simultaneous Determination of Tight-Binding Inhibition Constants and Enzyme Concentration

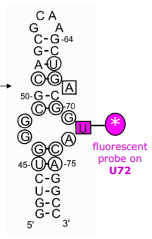
Petr Kuzmič,*,1 Kyle C. Elrod,† Lynne M. Cregar,† Steve Sideris,† Roopa Ra
*BioKin, Ltd., 1652 South Grand Avenue, Suite 337, Pullman, Washington 99163; and †Depa
‡Department of Medicinal Chemistry, Axys Pharmaceuticals, Inc., 180 Kimball Way, South San I

... then it *must* be optimized, along with the K_d!

Experiment #1: Optimized parameters – Take 2

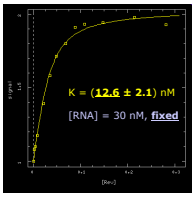
ADD ONE MORE "UNKNOWN" AND SEE WHAT HAPPENS ...

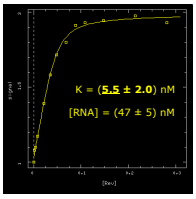
```
[mechanism]

   R72 + Rev <===> R72.Rev      :      K     dissoc

[constants]

   K = 0.02  ?

[concentrations]

   R72 = 0.03  ?

[responses]

   R72      = 33.3  ?
   R72.Rev = 66.6  ?

[data]

   variable Rev
   set      R72--Rev
```



Fixed or optimized [RNA]? Model selection results

AKAIKE INFORMATION CRITERION IS INCONCLUSIVE





Model discrimination analysis

Minimum sum of squares = 0.00497308

Akaike Information Criterion — sum of squares did decrease by a factor of two

	model	n_D	n_P	SS_{min}	AIC_c	ΔAIC_c	weight
[1]	fixed RNA conc.	13	3	1.957	-47.2	3.2	0.171
[2]	optimized RNA conc.	13	4	1.000	-50.4	0.0	0.829

→ this number must be **larger than ~10**

→ "Akaike weight" must be **larger than ~0.95**

however the number of adjustable parameters increased!

Fixed or optimized [RNA]? Confidence intervals

THE **"PLUS OR MINUS"** STANDARD ERRORS ARE ALMOST **ALWAYS WRONG** (TOO SMALL)

```
[task]

   task = fit
   data = equilibria

[mechanism]

   R72 + Rev <===> R72.Rev      :      K     dissoc

[constants]

   K = 0.02  ??

[concentrations]

   R72 = 0.03  ??

[responses]

   R72      = 33.3 ?
   R72.Rev = 66.6 ?
...
```

"PROFILE-T" METHOD

Watts, D. G. (1994)
"Parameter estimation from nonlinear models"
Methods Enzymol. **240**, 24-36.

Bates, D. M., and Watts, D. G. (1988)
Nonlinear Regression Analysis and its Applications
Wiley, New York, pp. 127-130

Confidence intervals: Results

THE NOMINAL [RNA] CONCENTRATION IS *PROBABLY* INCORRECT

DynaFit output:

Optimized Parameters

No.	Par#Set	Initial	Final	Std. Error	CV (%)	Low	Low P (%)	High	High P (%)
#1	K	0.02	**0.0054574**	0.00198625	36.40	0.00219474	95	0.0112241	95
#2	[R72]	0.03	**0.0473544**	0.00484073	10.22	0.0346082	95	0.057239	95
#3	r(R72)	33.3	**21.6544**	2.07801	9.60				
#4	r(R72.Rev)	66.6	**41.9983**	4.55354	10.84				

parameter	best-fit value	formal error, ±	confidence interval (95%)	
K_d, nM	5.5	2.0	2.2 — 11.2	
[R72], nM	47.4	4.8	34.6 — 57.2	... *nominal*: 30.0

reasonable suspicion:
actual **RNA** concentration *might* be **higher by ~60%** than the nominal value

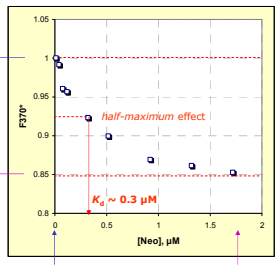
Experiment #2: RRE / Neomycin – raw data

FIXED RRE-72AP CONCENTRATION: **[R72] = 0.1 μM**

INITIAL ESTIMATES:

molar response of **R72**
1.0/0.1 = **10**

molar response of **R72.Neo**
0.85/0.1 = **8.5**



only **R72** (0.1 μM)

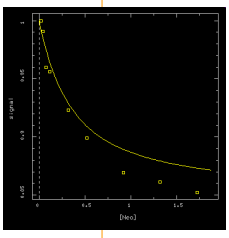
only **R72.Neo** (0.1 μM)

Experiment #2: RRE / Neomycin – script

USING INITIAL ESTIMATES ESTIMATED FROM RAW DATA

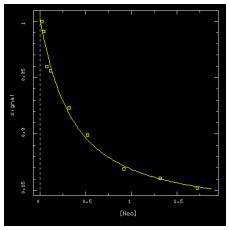
```
[task]

   task = fit
   data = equilibria

[mechanism]

   R72 + Neo <===> R72.Neo      :

[constants]

   K = 0.3 ??

[concentrations]

   R72 = 0.1 ; fixed!

[responses]

   R72      =  10 ?
   R72.Neo = 8.5 ?
...
```

File .. Try



Experiment #2: RRE / Neomycin – results

USING INITIAL ESTIMATES FROM PREVIOUS SLIDE



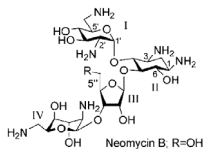
parameter	best-fit value	formal error, ±	confidence interval (95%)
K_d, μM	0.29	0.07	0.15 — 0.56

DynaFit output:

No.	Par#Set	Initial	Final	Std. Error	CV (%)	Low	Low P (%)	High	High P (%)	Note
#1	K	0.3	**0.289936**	0.0728402	25.12	0.15153	95	0.562575	95	
#2	r(R72)	10	**10.0329**	0.0556981	0.56					
#3	r(R72.Neo)	0.5	**0.2671**	0.104526	1.26					

Experiment #1 & #2: Summary

ONLY **BINARY** INTERACTIONS STUDIED SO FAR

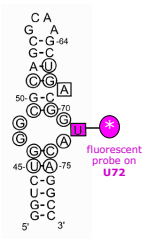


K_d = **290** nM

K_d = **6** nM

Rev model peptide:

Suc-TRQARRNRRRRWRERQRAAAAK



The main question remains unanswered

Could Neomycin *prevent* the Rev peptide from binding to the RNA?

in other words:

Is the binding of Rev and Neomycin *simultaneous* or *exclusive*?

non-competitive competitive

And how do we translate these ideas into *stoichiometric notation*?

DynaFit

Simultaneous vs. exclusive: stoichiometry

IT DEPENDS ON **HOW MANY DIFFERENT COMPLEXES** ARE FORMED

EXCLUSIVE: • ***not necessarily*** *different binding sites*

Neo + RRE + Rev ⟷ Neo·RRE + RRE·Rev

SIMULTANEOUS: • *always at different binding sites*

Neo + RRE + Rev ⟷ Neo·RRE + RRE·Rev + Neo·RRE·Rev

Simultaneous vs. exclusive: DynaFit notation

HOW MANY DIFFERENT COMPLEXES IS NOT THE ONLY QUESTION

```
[mechanism] ; exclusive

        RRE + Rev <===>   RRE.Rev      :  Kr   dissoc
        Neo + RRE <===>   Neo.RRE      :  Kn   dissoc
```

```
[mechanism] ; simultaneous

        RRE + Rev <===>   RRE.Rev      :  Kr   dissoc
        Neo + RRE <===>   Neo.RRE      :  Kn   dissoc

        ??? + ??? <===>   Neo.RRE.Rev  :  ??   dissoc
```

what goes here?

Two new concepts to consider ...

... BEFORE WE CAN FINISH OUR DYNAFIT SCRIPT

1. "thermodynamic box"
2. independent *vs.* interacting sites

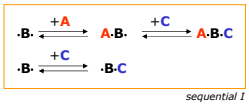
From stoichiometry to molecular mechanism

ONLY **BIMOLECULAR** INTERACTIONS ARE REALISTIC: **THREE** MOLECULES **NEVER** COLLIDE !

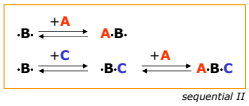
overall stoichiometry:

A + ·B· + C ⇌ A·B· + ·B·C + A·B·C

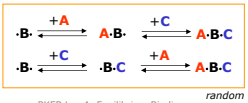
possible molecular mechanisms:



sequential I



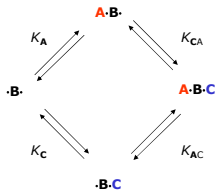
sequential II



random

Thermodynamic box: A very general idea

NO MATTER WHICH PATH WE TAKE, THE **FREE-ENERGY CHANGE** MUST BE THE SAME



all "*K*"s are ***dissociation*** constants

$K_{CA} \times K_A = K_{AC} \times K_C$

dissociation from ABC: first **C** then **A**

dissociation from ABC: first **A** then **C**

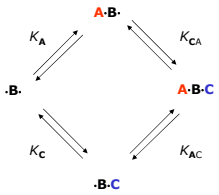
Only **three of four** equilibrium constants can have an **arbitrary** value.

Any one of the *K*'s is ***a priori* defined** in terms of the remaining three.

It does not matter which *K* we select to be dependent on the remaining three.

Thermodynamic box: DynaFit notation

THERE ARE MULTIPLE EQUIVALENT WAYS TO SPECIFY THE "**RANDOM**" MECHANISM IN DYNAFIT



all "*K*"s are ***dissociation*** constants

for example:

```
[mechanism]
   A + B <==> AB   :  Kc   diss
   B + C <==> BC   :  Kc   diss
  AB + C <==> ABC  :  Kca  diss
```

or, equivalently:

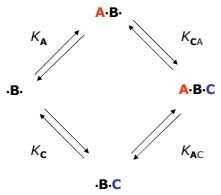
```
[mechanism]
   A + B <==> AB   :  Ka   diss
   B + C <==> BC   :  Kc   diss
  A + BC <==> ABC  :  Kac  diss
```

There must be **only three** steps (**any** three) in the DynaFit notation!

***How many other ways** exist to specify this mechanism in DynaFit ?*

Independent / interacting sites

WHETHER OR NOT PAIRS OF EQUILIBRIUM CONSTANTS IN THE "BOX" ARE THE SAME



all "*K*"s are ***dissociation*** constants

independent sites:

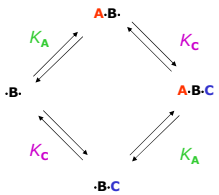
$K_{CA} = K_C$
$K_{AC} = K_A$

interacting sites:

$K_{CA} \neq K_C$
$K_{AC} \neq K_A$

Independent sites: DynaFit notation

THERE ARE MULTIPLE EQUIVALENT WAYS TO SPECIFY THIS, TOO



all "*K*"s are ***dissociation*** constants

for example:

```
[mechanism]
   A + B <==> AB   :  KA   diss
   B + C <==> BC   :  Kc   diss
  AB + C <==> ABC  :  Kc   diss
```

or, equivalently:

```
[mechanism]
   A + B <==> AB   :  Ka   diss
   B + C <==> BC   :  Kc   diss
  A + BC <==> ABC  :  Ka   diss
```

Only **two** distinct dissociation constants.

Simultaneous vs. exclusive: DynaFit notation

FINALLY WE KNOW ENOUGH THEORY TO FINISH THE DYNAFIT SCRIPT

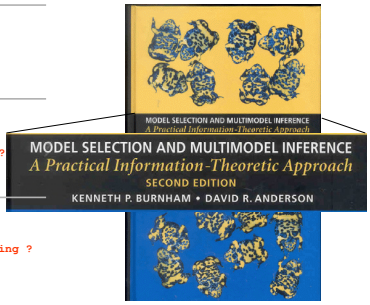
```
[mechanism] ; exclusive

        RRE + Rev <===>   RRE.Rev      :  Kr   dissoc
        Neo + RRE <===>   Neo.RRE      :  Kn   dissoc
```

```
[mechanism] ; simultaneous, non-interacting

        RRE + Rev <===>   RRE.Rev      :  Kr   dissoc
        Neo + RRE <===>   Neo.RRE      :  Kn   dissoc

   Neo.RRE + Rev <===>   Neo.RRE.Rev   :  Kr   dissoc
```

```
[mechanism] ; simultaneous, interacting

        RRE + Rev <===>   RRE.Rev      :  Kr   dissoc
        Neo + RRE <===>   Neo.RRE      :  Kn   dissoc

   Neo.RRE + Rev <===>   Neo.RRE.Rev   :  Krn  dissoc
```

Automatic model selection in DynaFit

```
[task]
   task  = fit
   data  = equilibria
   model = exclusive ?
...
...

[task]
   task  = fit
   data  = equilibria
   model = interacting ?
...
...

[task]
   task  = fit
   data  = equilibria
   model = non-interacting ?
...
...
```

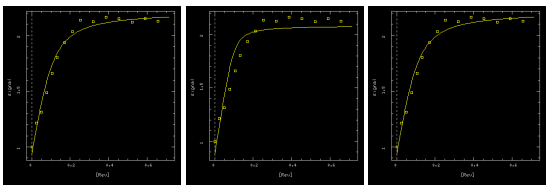


Model selection: round 1 – fixed [RNA]

NEITHER MODEL FITS VERY WELL AT ALL!

experiment #3

labeled [RNA]: 100 nM, constant **[RNA] is under suspicion**
Neomycin B: 990 nM, constant
Rev peptide: 0 – 655 nM, varied



exclusive non-interacting interacting

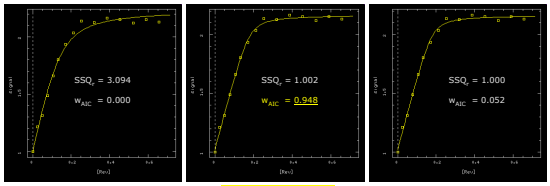
All equilibrium constants were ***fixed*** at values determined in **binary** binding studies.

Model selection: round 2 – optimized [RNA]

GOODNESS-OF-FIT IS MUCH IMPROVED

experiment #3

labeled [RNA]: **178** nM, **optimized** in the fit
Neomycin B: 990 nM, constant
Rev peptide: 0 – 655 nM, varied

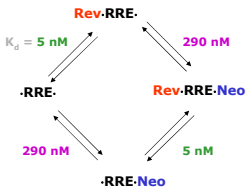


exclusive **non-interacting** interacting

actual **[RNA] 78% higher** than nominal?

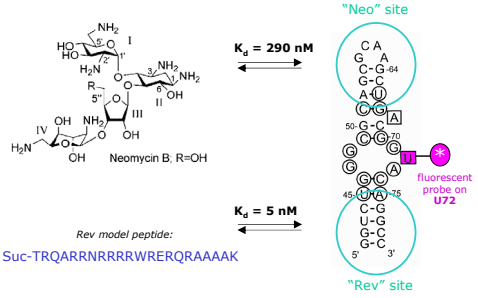
Mechanism for HIV-1 RRE / Neomycin / Rev

NON-EXCLUSIVE BINDING TO TWO DISTINCT, **NON-INTERACTING** SITES



Mechanism for HIV-1 RRE / Neomycin / Rev

STRUCTURAL IMPLICATIONS OF THE BINDING DATA: **SEPARATE BINDING SITES**



Rev model peptide:

Suc-TRQARRNRRRRWRERQRAAAAK

Summary and conclusions

1. Equilibrium binding data are easily handled by numerical models.
   Arbitrary conditions (no "excess of **A** over **B**"); arbitrarily complex mechanisms.
2. Certain restrictions exist on representing reaction mechanisms.
   The "thermodynamic box" rule must always be obeyed.
3. Exclusive vs. non-exclusive binding is expressed
   simply as a **different number of complexes** present in the overall mechanism.
4. Interacting vs. non-interacting sites are expressed
   simply by assigning **identical vs. unique values** to equilibrium constants.
5. Incorrectly specified concentrations have a large impact
   on best-fit values of equilibrium constants *and* on model selection.

   BUT THERE IS SOME RELIEF:

   when the binding is "tight", **actual concentrations can be inferred** from the data;
   when the binding is "loose", systematic concentration errors do not matter (much).
6. DynaFit is not a "silver bullet": You must still use your brain a lot.