Automated SPR Data Processing and Sensogram Classification

Constantine Kreatsoulas
Chemistry Modeling & Informatics – West Point, PA
Merck & Co., Inc.
constantine_kreatsoulas@merck.com

Where patients come first MERCK

Public disclosed

Abstract

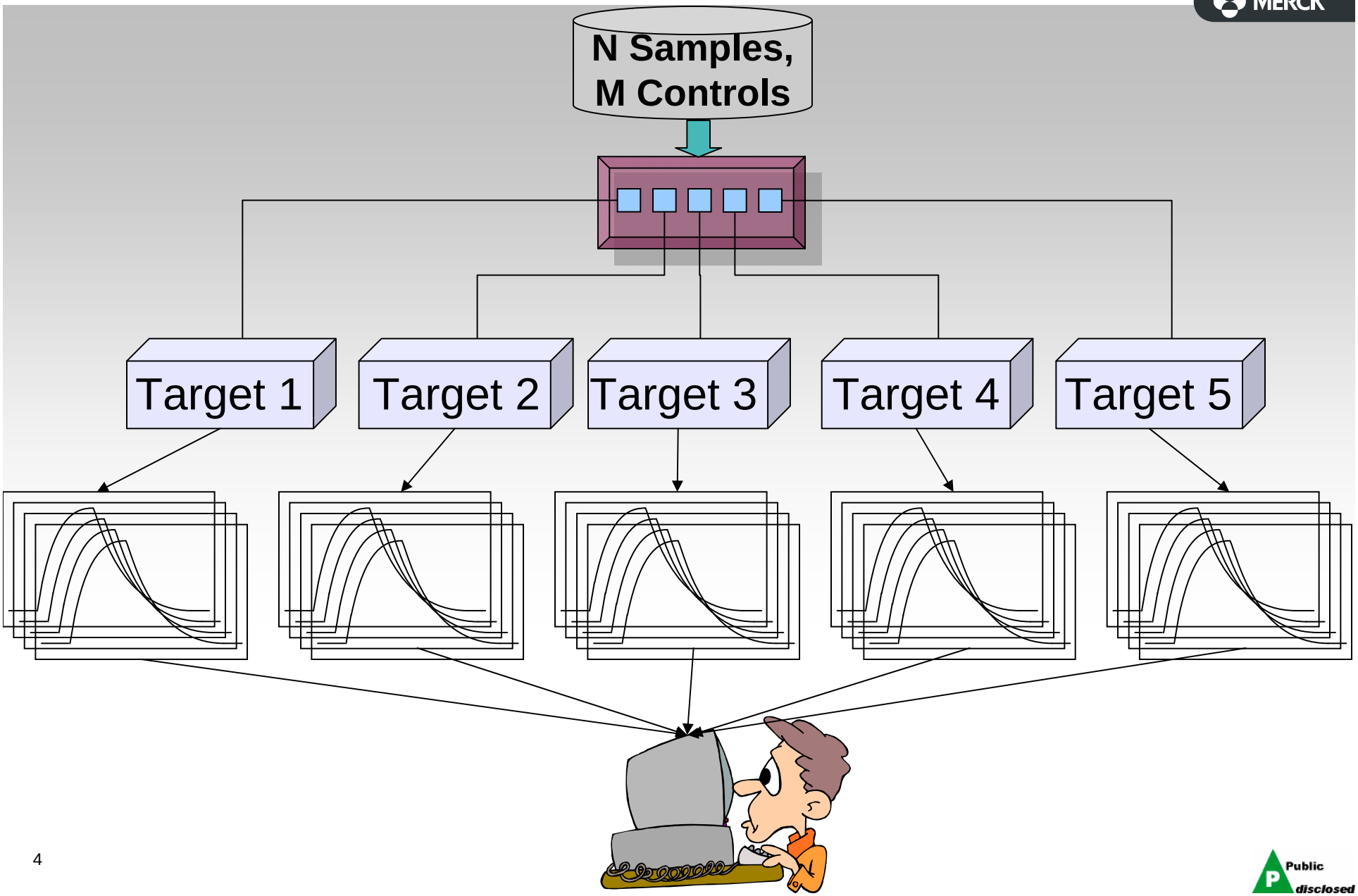
Fragment-based approaches have added to the arsenal of tools used to identify novel small molecule leads with high ligand efficiencies. A variety of label-free technologies have been developed and implemented throughout the industry for fragment screening. Using surface plasmon resonance (SPR) as a fragment screening platform is in its infancy. The miniaturization and automation of this technology has led to the associated problem of dealing with the large volume of raw data generated. The resources needed for the analysis, integration and prioritization of compounds screened makes it challenging to take the results of an SPR screen into the workflow of project teams engaged in the discovery process in a timely fashion. As such, several sets of equations were derived and implemented on Merck's intranet to score single sensograms to distinguish stable binders from weak or anomalous binders. This set of equations was optimized and validated on simulated data to both capture "fragment-like" behavior from SPR experiments and to filter out much of the anomalous behavior commonly observed. This talk will cover the creation of the theoretical data sets, the resultant equations, and the performance metrics used to evaluate them.

Discovery Project Needs Drove Collaboration

MERCK

- Fragment Screening in the Discovery Process
  - Lead ID – time is critical!
  - Need rapid turnaround of results
    - Assay development takes time for novel targets
    - Biacore A100 assay takes little time (*N* fragments + *M* controls on *5* spots)
    - Visual analysis of each of the $5 \times (N + M)$ curves
      - ~1 week (Bottleneck)
- Selection criteria for follow up
  - Response relative to theoretical R_{max}
  - Appropriate curve shape (box) for fragment-like kinetics
  - Selectivity over other targets

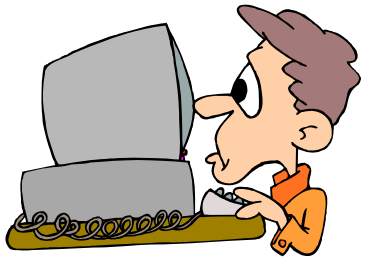
The Problem in a Nutshell



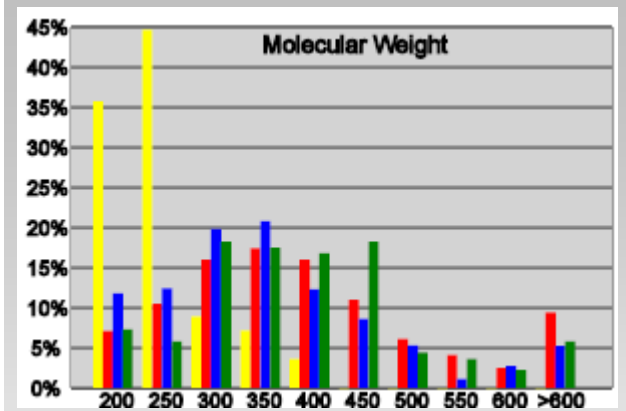
The Problem in a Nutshell

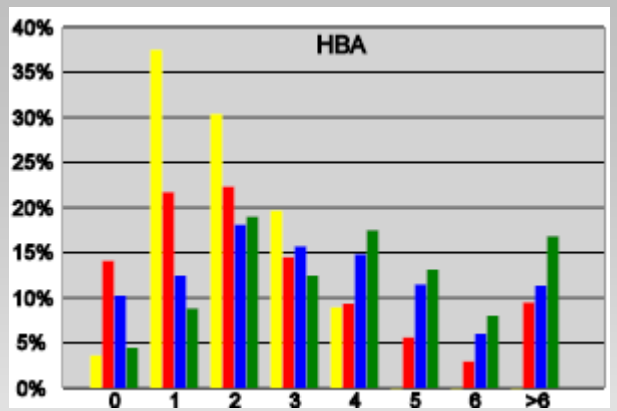
MERCK

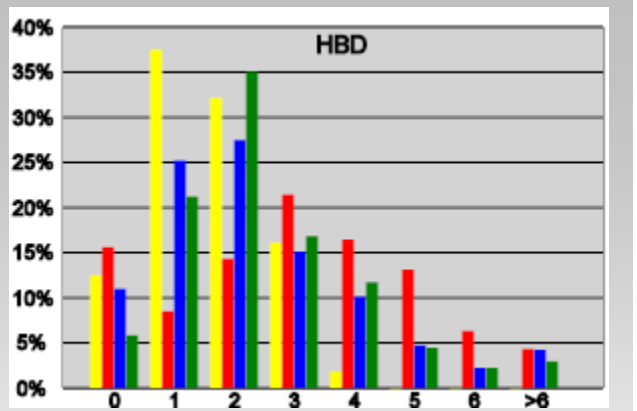
How can we speed this up to have an immediate project impact?

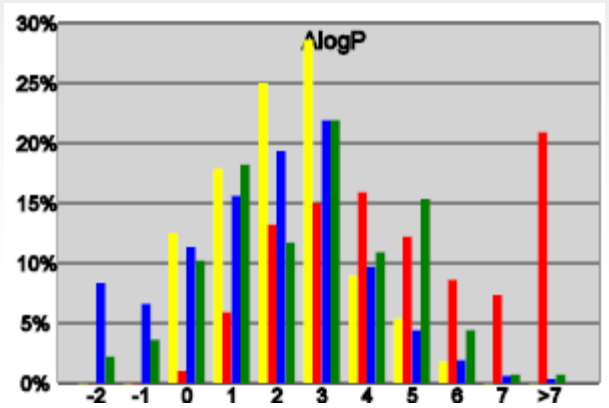


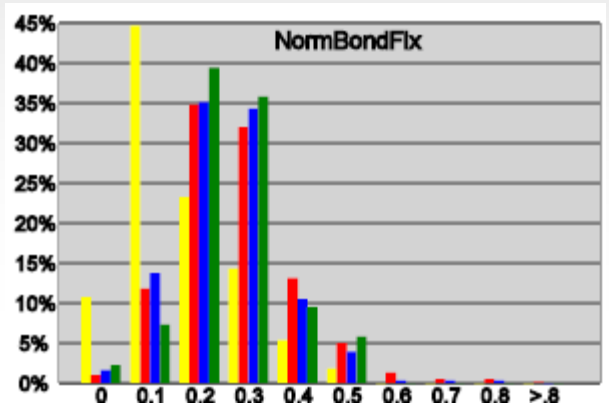
Aside: Summary of Merck Fragment Library Properties

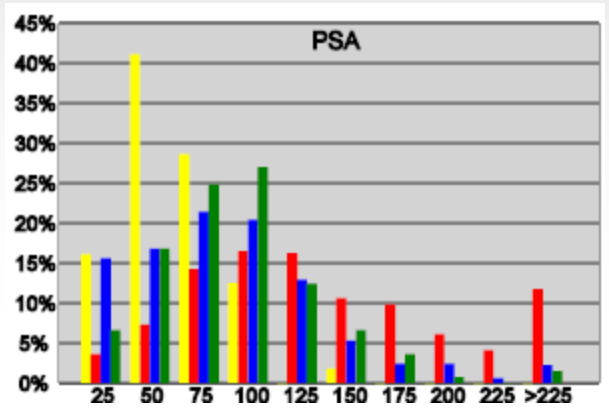






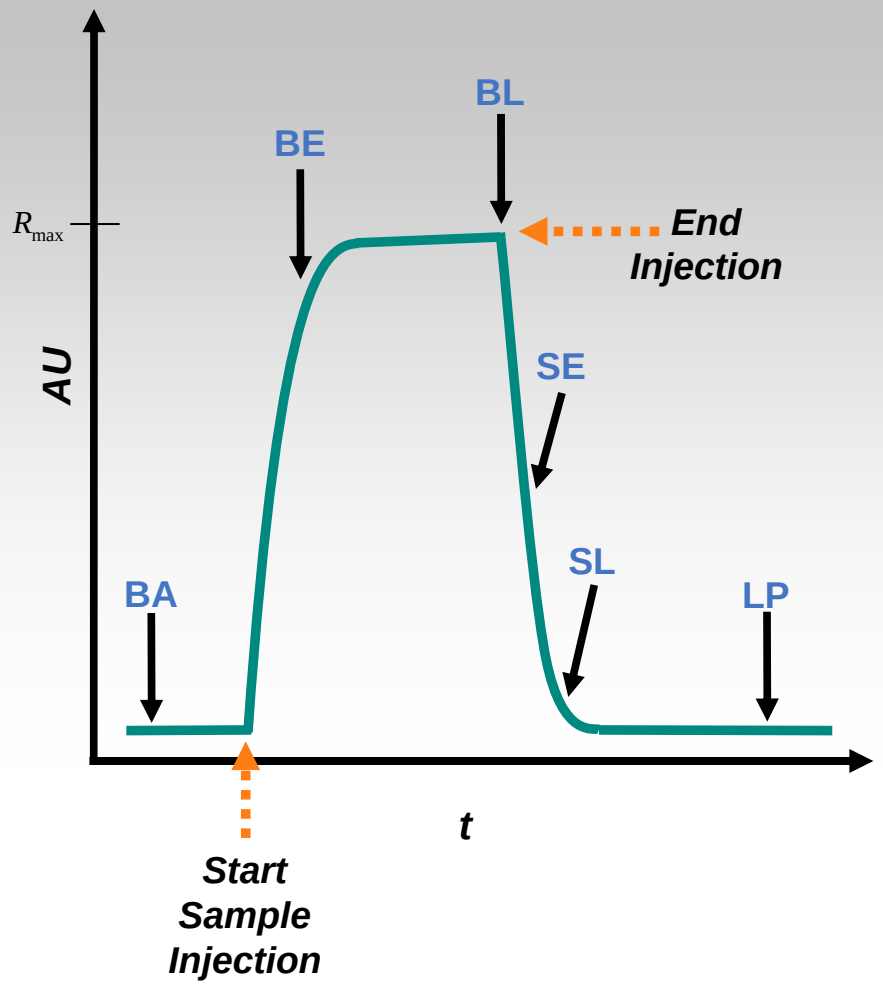






Yellow: Fragments Assayed
Red: All Launched Drugs as of 2006 (MDDR)
Blue: Orally Delivered Drugs
Green: Top 200 Small Molecule Drugs of 2008

Another Aside: A Single Sensogram



Single-point experiment

- *Not* titrated
- > 1500 compounds/screen

$$R_{\max} = \frac{n \cdot R_L \cdot MW_{analyte}}{MW_{ligand}}$$

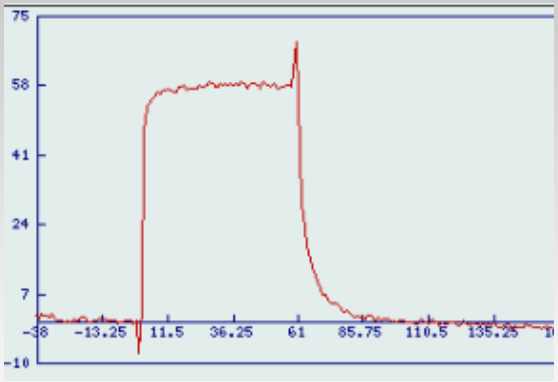
Reporter Points

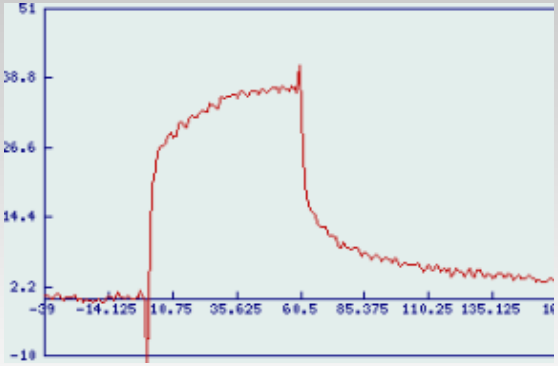
- BA: baseline (15 sec window)
- BE: binding early (10 sec)
- BL: binding late (10 sec)
- SE: stability early (10 sec)
- SL: stability late (10 sec)
- LP: late points (10 sec)

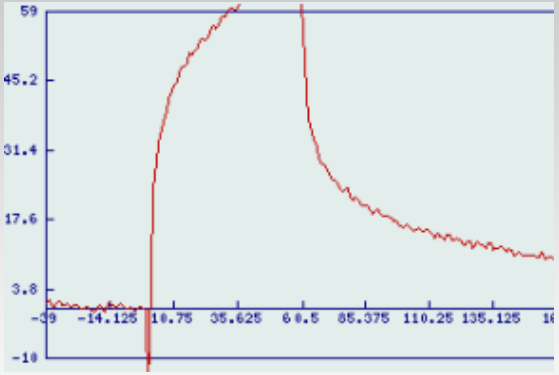
Theory vs. Experiment

Many curve types, need to account for them all:

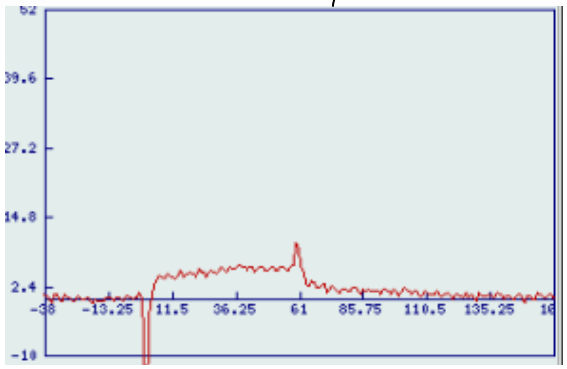
The Good

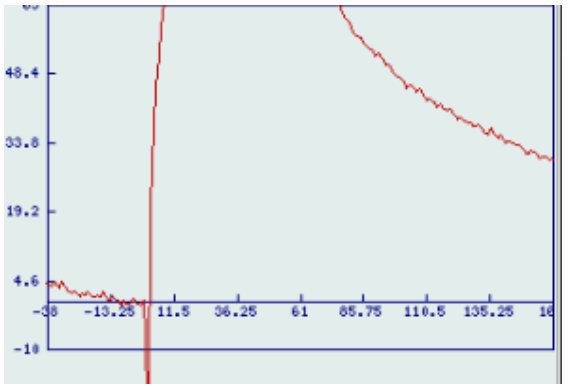


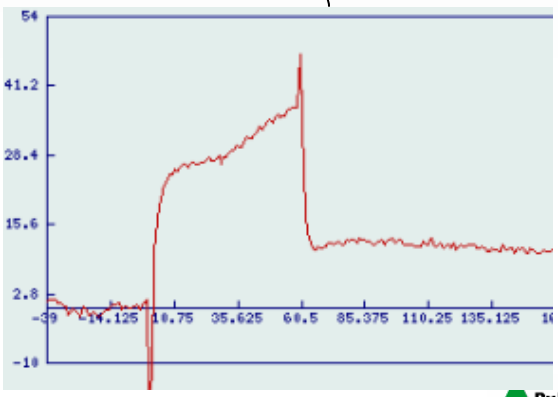




The Bad and Ugly





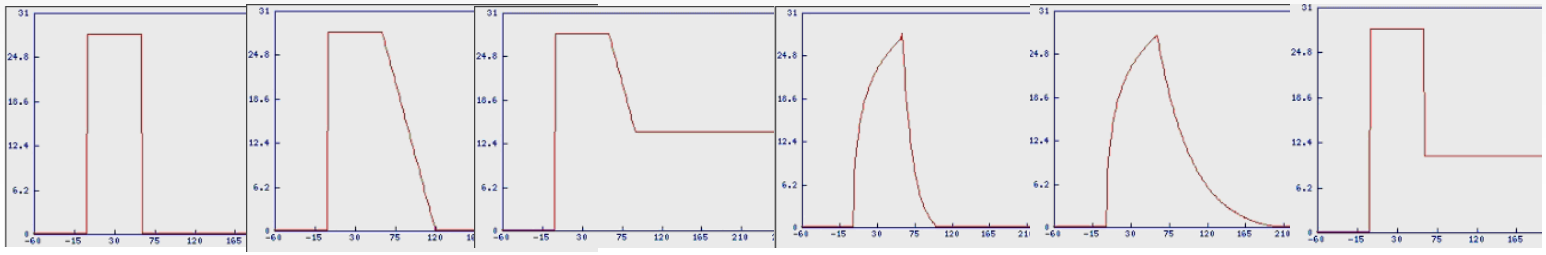


Theoretical Data Sets

MERCK

Used to mimic experimental data and benchmark ranking equations

Data Set	*IS1*	*IS2*	*IS3*	*IS4*	*IS5*	*IS6*
Design Criteria	*Fast-on, fast-off kinetics (fragment-like)*	*Fast-on, slow-off kinetics*	*Fast-on, slow-off; incomplete dissociation*	*Slow-on, moderately fast-off (high affinity)*	*Slow-on, slow-off kinetics (drug-like)*	*Fast-on,fast-off kinetics; incomplete dissociation*



<u>24 total “training” sets:</u>

- **Curves modulated from 100%, 75%, 50% and 25% of R_{max}.**
- **Two modulated to 150% of R_{max}.**

How to Evaluate the Resultant Rankings?

MERCK

Mathematically:

AUC = Area Under Accumulation Curve

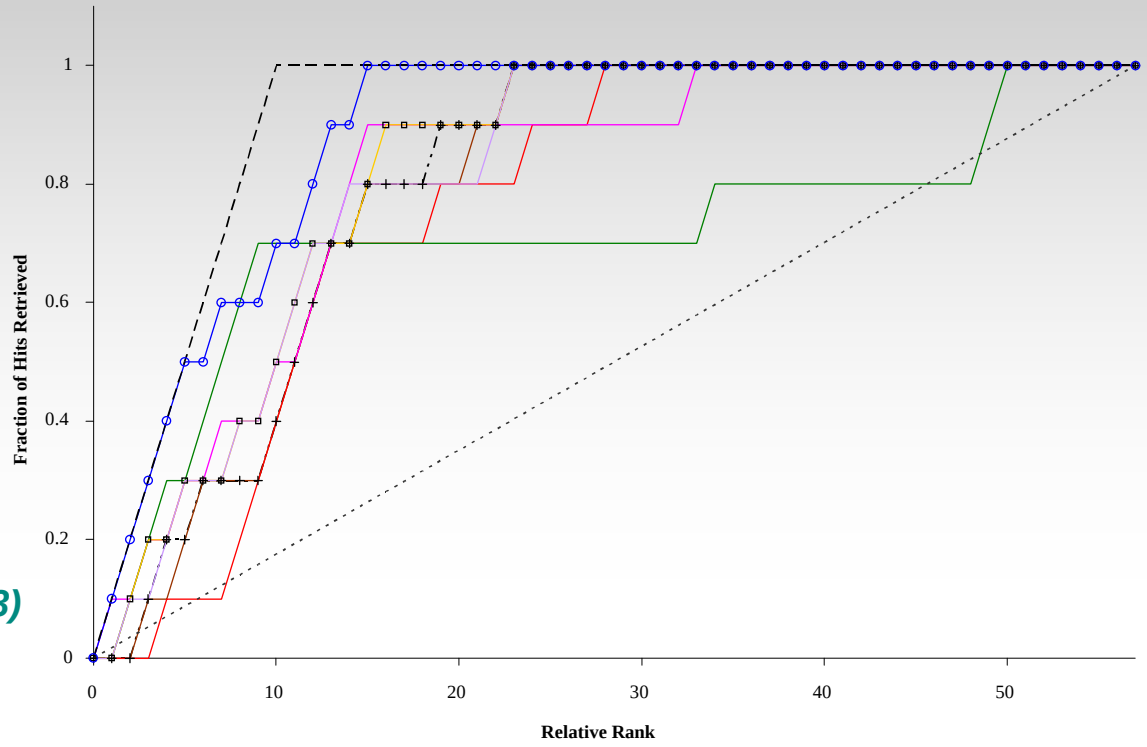
$$ROC = \frac{2AUC - R_a}{2R_i} \quad (1)$$

$$RIE = \frac{\sum_{i=1}^{hit} e^{-\alpha(r_i/N)}}{R_a\left(\frac{1-e^{-\alpha}}{e^{\alpha/N}-1}\right)} \quad (2)$$

$$BEDROC = \frac{RIE - \left(\frac{1-e^{aR_a}}{R_a\left(1-e^{\alpha}\right)}\right)}{\left(\frac{1-e^{-aR_a}}{R_a\left(1-e^{-\alpha}\right)}\right) - \left(\frac{1-e^{aR_a}}{R_a\left(1-e^{\alpha}\right)}\right)} \quad (3)$$

Graphically:

Accumulation Curve



Possible Ranking Metrics

Simple, widely used:

$$S = \frac{B_L - B_A}{R_{\max}} \quad (4) \qquad S_{Occ} = 100 \cdot \frac{B_L}{R_{\max}} \quad (5)$$

Penalize slow on-rate:

$$S = \frac{2B_L - (B_E + B_A)}{R_{\max}} \quad (6)$$

Penalize slow off-rate:

$$S = \frac{B_L - S_E}{R_{Max}} \quad (7) \qquad S = \frac{B_L - (2S_E - S_L)}{R_{\max}} \quad (8)$$

Penalize aggregators:

$$S = \frac{B_L - L_P}{R_{Max}} \quad (9) \qquad S = \frac{B_L - (2S_L - L_P)}{R_{\max}} \quad (10)$$

or combine them…

Combined Ranking Metric (aka *Eq. 17)*

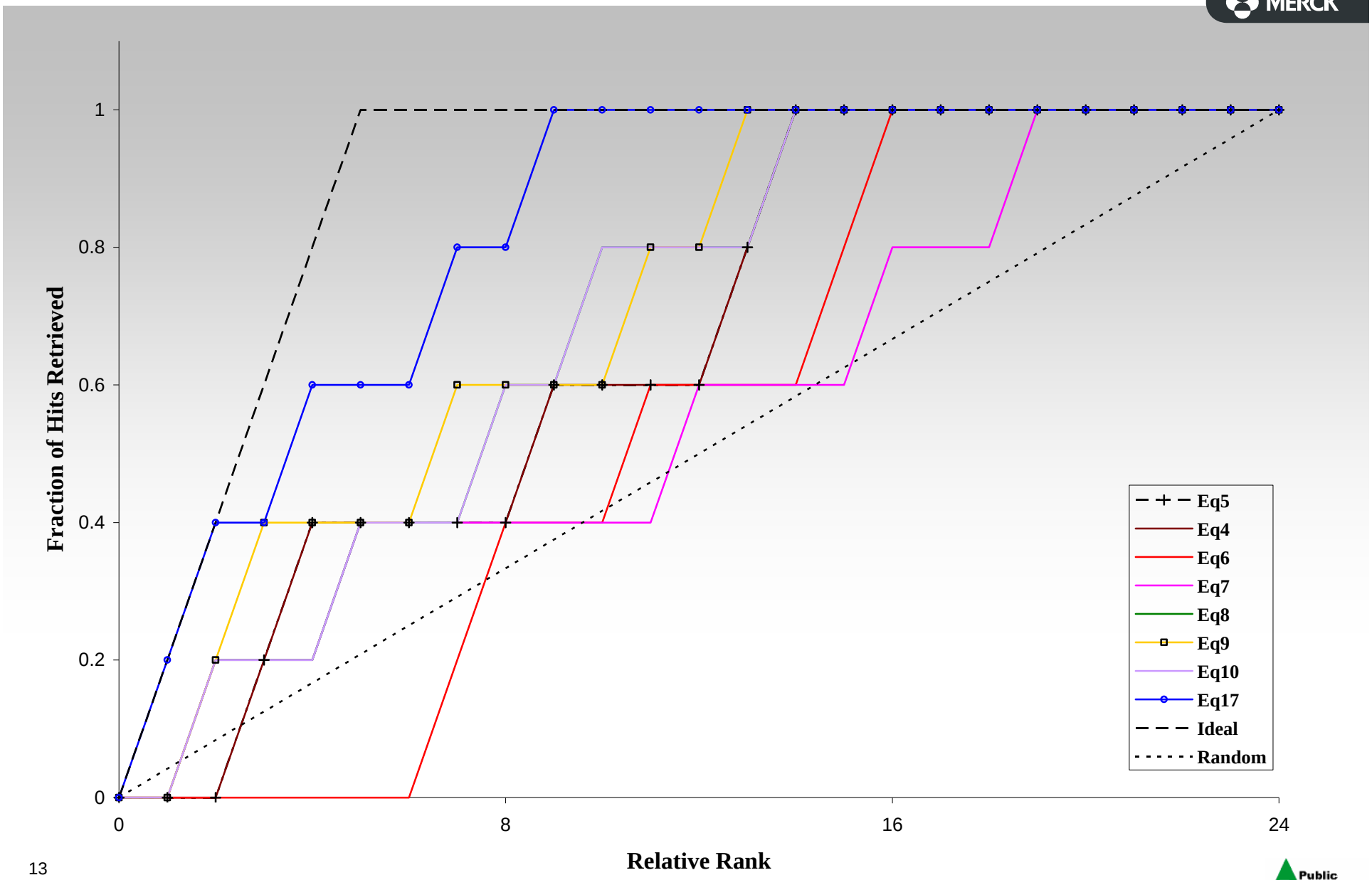
MERCK

Fast On & Occupancy | **Fast Off** | **Drug-Like** | **Irreversible**

$$S = \frac{w_1\left\{1-\left|1-\frac{\|\overline{(BE,BL)}-BA\|}{R_{\max}}\right|\right\} + w_2\left\{\left[\frac{\|BL-SE\|}{\|BL-BA\|}\right]^3\right\} + w_3\left\{\left[\frac{\|BL-SL\|}{\|BL-BA\|}\right]^3\right\} + w_4\left\{\left[1-\frac{\|LP-BA\|}{R_{\max}}\right]^3\right\}}{\sum_{i=1}^{4} w_i}$$

Weights, w_i, allow for a single equation to be used for fragment and small molecule ranking.

Performance on Theoretical Data Set

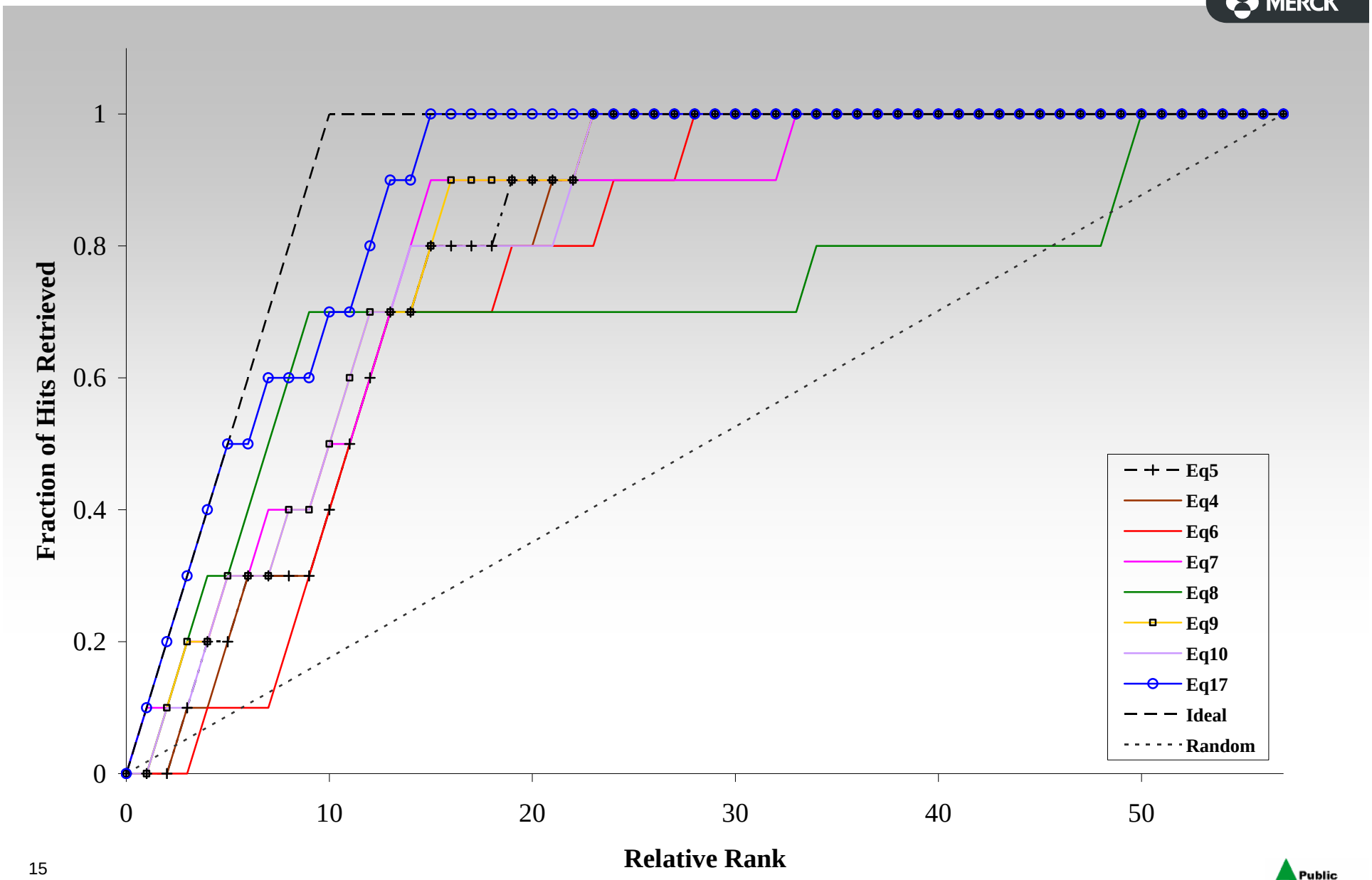


Performance on Theoretical Data Set

MERCK

	AUC	*ROC*	*RIE*	*BEDROC*
Eq. (4)	0.65	0.71	2.34	0.49
Eq. (5)	0.65	0.71	2.34	0.49
Eq. (6)	0.53	0.56	1.15	0.24
Eq. (7)	0.55	0.59	1.93	0.41
Eq. (8)	0.68	0.76	2.36	0.50
Eq. (9)	0.71	0.79	2.76	0.58
Eq. (10)	0.68	0.76	2.36	0.50
Eq. (17)	0.82	0.93	3.55	0.75
Ideal	0.89	1.02	4.48	0.95
Random	0.49	0.52	1.76	0.37

Performance on a Fragment Screen



Performance on a Fragment Screen

MERCK

	AUC	*ROC*	*RIE*	*BEDROC*
Eq. (4)	0.80	0.87	1.91	0.35
Eq. (5)	0.81	0.87	1.92	0.35
Eq. (6)	0.78	0.83	1.13	0.20
Eq. (7)	0.81	0.88	3.31	0.60
Eq. (8)	0.81	0.88	2.51	0.45
Eq. (9)	0.82	0.89	2.76	0.50
Eq. (10)	0.81	0.88	2.51	0.45
Eq. (17)	0.89	0.97	5.18	0.94
Ideal	0.91	1.00	5.53	1.00
Random	0.49	0.49	1.85	0.33

Deployed via Intranet: Modify on the Fly

MERCK

MERCK | myMerck portal: home | help

MolSys >

Sensogram File Upload

Tab Separated file to upload: Browse..

Ligand Mwt File Upload

Comma or tab separated file to upload: Browse..

Set Parameters

Analysis Control:

R_max only used when there is no Ligand mwt file given.

○ Drug-like ○ Fragment-like ○ User-defined: W1: 1 W2: 1 W3: 1.25 W4: 0.75

R_L(RU): 5000 R_max(RU): 28 mwt Prot(D): 70000

BA start: -55	BE start: 6	BL start: 38
BA stop: -20	BE stop: 28	BL stop: 58
SE start: 62	SL start: 169	LP start: 300
SE stop: 86	SL stop: 181	LP stop: 380

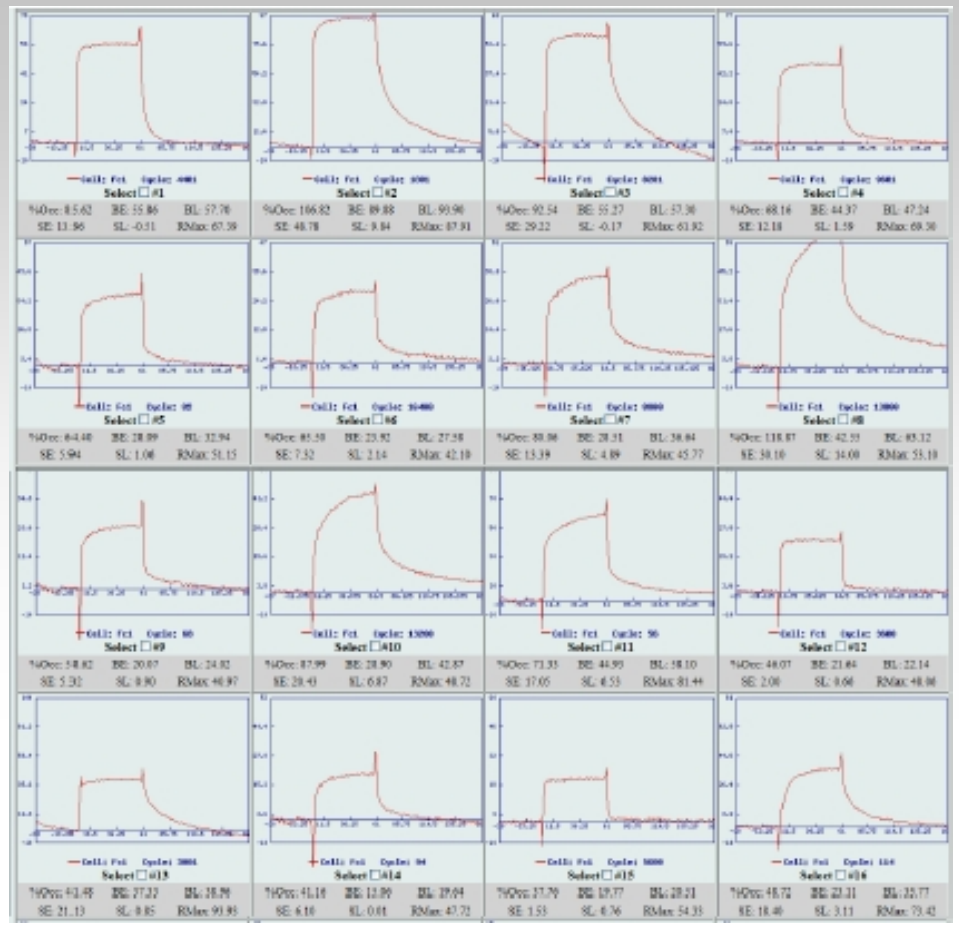
Plot Control:

Leave as "Calc" to let the software calculate the values from the input.

Sort by:	Y Min:	Y Max:	X Min:	X Max:
Scoring Eq'n Eq17	Calc	Calc	Calc	Calc

Process Data Reset

If nothing happens, press the button again.



Comments

MERCK

Impact

- Shortened SPR screen analysis time > 50x
- Rapid selection of true positives
- Outperforms usual methods of fragment selection
- Can use for fragments or fully elaborated molecules

Next Steps

- Use of machine learning to select best w_i from a series of screens and users' rankings
- Optimization of w_i for drug-like molecules
- Would love to hear how this works in others' hands
- Not meant to replace titrations!

Acknowledgements

Automated Biotechnology

- ***Kartik Narayan***
- Delphine Collin
- Krista Getty

Chemistry Modeling & Informatics

- Chris Culberson
- John Sanders

Thank you!

Reference:

Kreatsoulas & Narayan, *Analytical Biochemistry,* 402 (2010) 179-184.