

So fragments..

An out-of-depth assessment

Spot the connection

- Arthroscopic knee surgery ~ 300,000 per year
- Coronary Angioplasty > 1,000,000 per year
- QSAR > 1000s? Per year

A Problem of Inference

Prob(Event given the data) \neq Prob(The data given the Event)

A Problem of Inference

Prob(Event given the data) \neq Prob(The data given the Event)

Prob(Ant is Ranting given he's giving a Talk)= 0.5 = $P(\text{Rant} | \text{Talk})$

Prob(Ant is giving a Talk given he's Ranting)= 0.02 = $P(\text{Talk} | \text{Rant})$

Bayes' Rule

$$P(\text{Event} \mid \text{data}) = B * P(\text{data} \mid \text{Event})$$

$$B = \frac{P(\text{Event})}{P(\text{data})}$$

$$P(\text{Rant} \mid \text{Talk}) = P(\text{Talk} \mid \text{Rant}) * \frac{P(\text{Rant})}{P(\text{Talk})}$$

Prosecutor's Fallacy

- Data = blood sample match
- $\text{Prob}(\text{Data} \mid \text{Innocent}) \sim 1/1000$
- $\text{Prob}(\text{Innocent} \mid \text{Data}) = B * \text{Prob}(\text{Data} \mid \text{Innocent})$
- **Fallacy:** $\text{Prob}(\text{Innocent} \mid \text{Data}) \cong \text{Prob}(\text{Data} \mid \text{Innocent})$
- True *only* if $\text{Prob}(\text{Innocent}) / \text{Prob}(\text{Data}) = 1$

Modeler's Fallacy: Flexible Docking

$P(\text{Flexible Docks} \mid \text{Active}) > P(\text{Inflexible Docks} \mid \text{Active})$

Therefore:

$P(\text{Active} \mid \text{Flexible Docks}) > P(\text{Active} \mid \text{Inflexible Docks})!$

Inflexible Docking vs Flexible Docking

$$P(\text{Active} | \text{Docks}) = P(\text{Docks} | \text{Active}) * \frac{P(\text{Active})}{P(\text{Docks})}$$

$$P(\text{Docks}) = P(\text{Docks} | \text{Active}) + P(\text{Docks} | \text{Inactive})$$

Correct Question:

$$\frac{P(\text{Flexible Docks} | \text{Active})}{P(\text{Flexible Docks})} > \frac{P(\text{Inflexible docks} | \text{Active})}{P(\text{Inflexible Docks})} \quad ???$$

Fallacious Modeling

Flexible proteins & docking

Flexible proteins & posing

Multiple active site ionization states & posing

Flexible shape fitting = better ROCS

Discrete waters = better docking

More conformers = better ROCS

More tautomers = better FRED/ EON

The Drug Designers Fallacy

- 1) Prosecutor's Fallacy + Anecdotal Reasoning
 - Concentrate only on successes of X
 - $P(X | \text{Find Drug}) \cong 1.0$

Fallacy: $\text{Prob}(\text{Find Drug} | X) \cong \text{Prob}(X | \text{Find Drug}) \cong 1.0$

Think I'm kidding?

90% of JMC pre-reviews	
We did X, and only X	So $\text{Prob}(X) = 1$
It worked	Therefore $\text{Prob}(X \text{Success}) = 1$
Publish my damn paper	<i>Clearly</i> $\text{Prob}(\text{Success} X) = 1$

Rigorous Assessment of Drug Design Approach X

$$P(\text{Find Drug} | X) = P(X | \text{Find Drug}) * \frac{P(\text{Find Drug})}{P(X)}$$

$P(\text{Find Drug})$ = % of successful projects

$P(X | \text{Find Drug})$ = % of successful projects using X not Y

$P(X)$ = % of projects that used X not Y

$P(\text{Find Drug} | X) > P(\text{Find Drug})$ *iff* X is better than Y

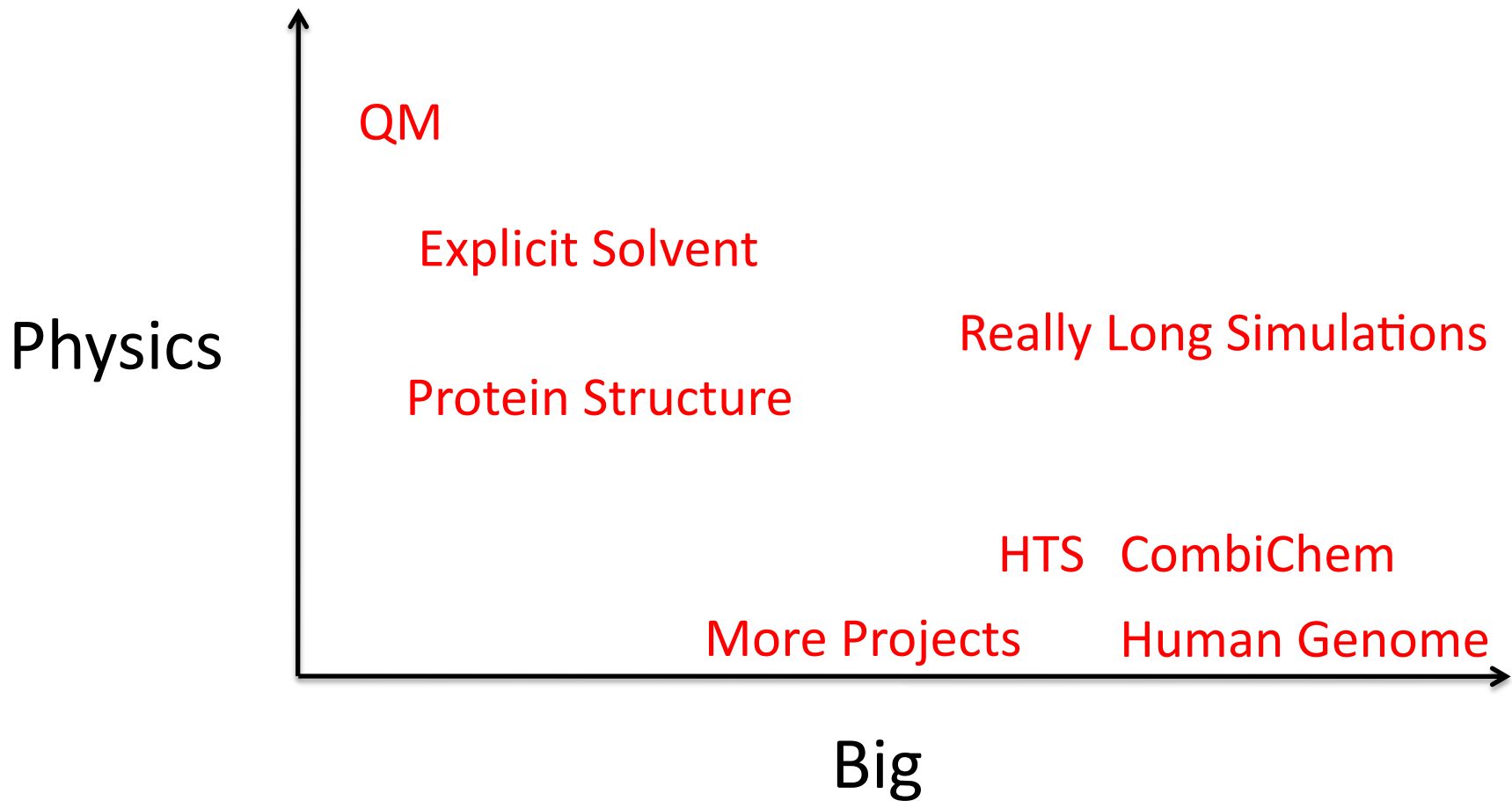
Intuitive Assessment of Drug Design Approach X

$$P(\text{Find Drug} | X) = \frac{P(X | \text{Find Drug})}{P(X)} * P(\text{Find Drug})$$

$$\frac{P(X | \text{Find Drug})}{P(X)} > 1, \text{ Surely}$$

The {Surely, Clearly, Obviously} Justification

{Surely, Clearly, Obviously} Justifications



Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery

Michael M. Hann, Andrew R. Leach, and Gavin Harper

J. Chem. Inf. Comput. Sci., **2001**, 41 (3), 856-864 • DOI: 10.1021/ci000403i • Publication Date (Web): 27 April 2001

Downloaded from <http://pubs.acs.org> on April 1, 2009

Hypothesis	Evidence
Larger molecules bind less often	None
Drugs are larger than leads	Sneader DB (<2 methyls>)
Can't make large libraries without big molecules	None
It is better to start with a small, weak binder and build	None



{Surely, Clearly, Obviously}

Other {Surely, Clearly, Obviously} of FBDD

- Fragment affinity transfers to large molecules
- Molecules containing that fragment bind similarly
- More information when a fragment binds
- Two good fragments link to give one great one
- It's easier/faster/cheaper to screen a few fragments

Simple Experiments: Pose

- Consider large ligands of known binding pose
- Break up into N fragments
- $P(B)$ = probability a fragment binds
- $P(R|B)$ = probability it binds within R RMSD

Simple Experiments: Affinity

Select a screening library

Screen some fragments (virtually or by NMR/ Xray) $\rightarrow p(\text{Frag binds})$

Run a full screen, determine hits $\rightarrow p(\text{Active})$

The fraction of hits contain the fragments that bind $\rightarrow p(F | A)$

Calculate $p(A | F)$

Simple Experiments: Information

- Screen some fragments (virtually or by NMR/ Xray) $\rightarrow p(F)$
- Choose N larger molecules that contain binding fragments
- Choose M molecules by some other means Y
- Run the screen and count $p(F|A)$, $p(Y|A)$, $P(A)$
- $P(A|F) > 1$ means F is 'better' than Y

What I do accept:

Stephanie Someless: “It is useful when nothing else works”

Bobby Cupboard: “It has a really low entry cost to drug discovery”

What I fear:

“I'll Tip My Hat To The New Constitution
Take A Bow For The New Revolution
Smile And Grin At The Change All Around
Pick Up My Mac Air and Code
Just Like Days of Old
Then I'll Get On My Knees And Pray
We Don't Get Fooled Again”