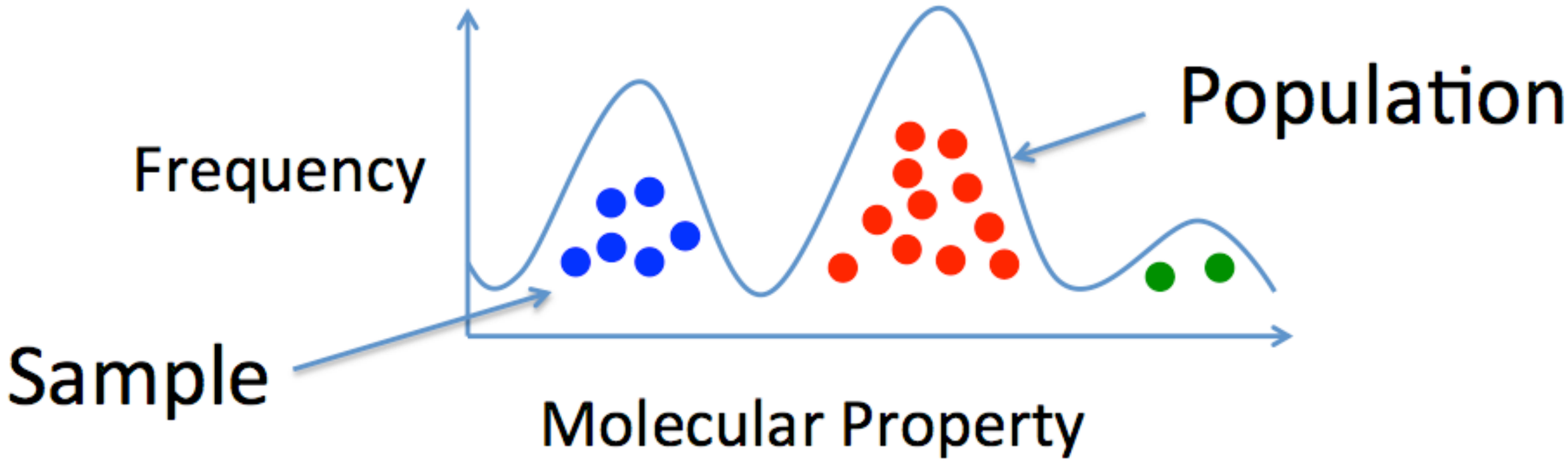
Today: Physical properties and other more empirical models

(QSAR models, sort of)

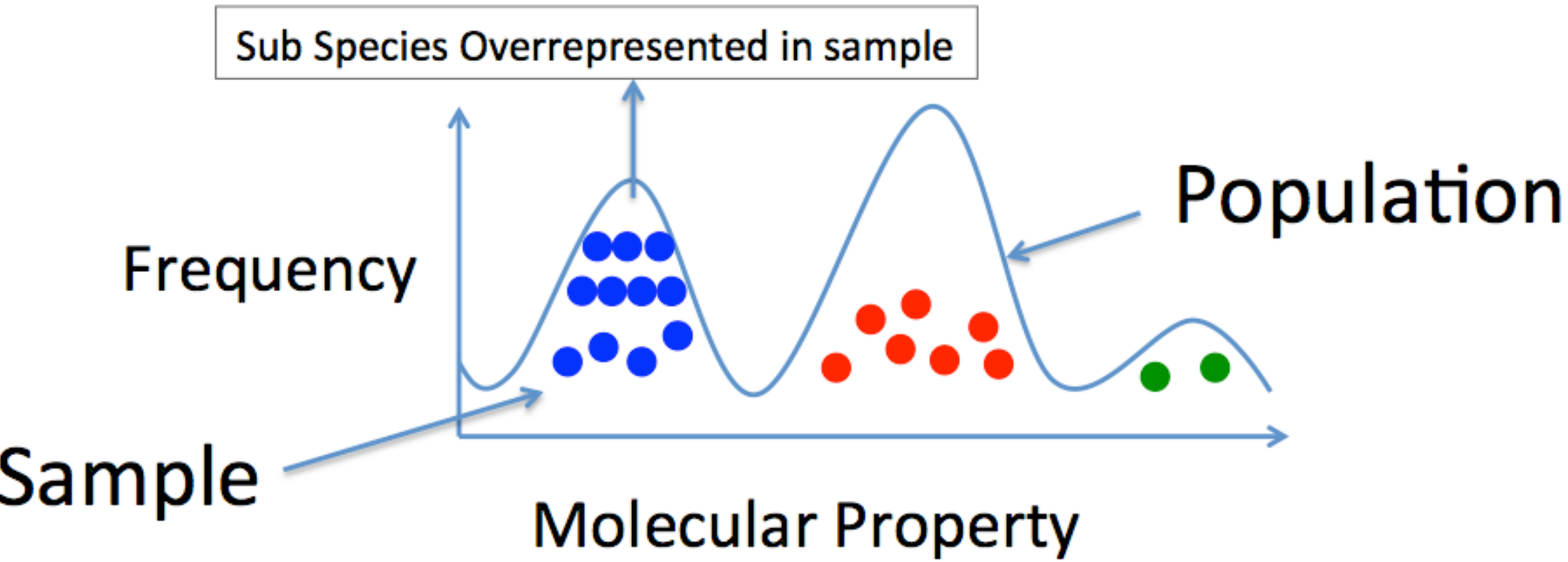
Validation depends on sampling quality

Cross-validation, bootstrapping depend on
sample distribution ≈ population distribution



Validation is *not* straightforward if sampling is *uneven*

Cross-validation, bootstrapping depend on sample distribution ≈ population distribution



Solubility Challenge: Can You Predict Solubilities of 32 Molecules Using a Database of 100 Reliable Measurements?

Antonio Llinàs,* Robert C. Glen, and Jonathan M. Goodman*

Pfizer Institute for Pharmaceutical Materials Science & Unilever Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

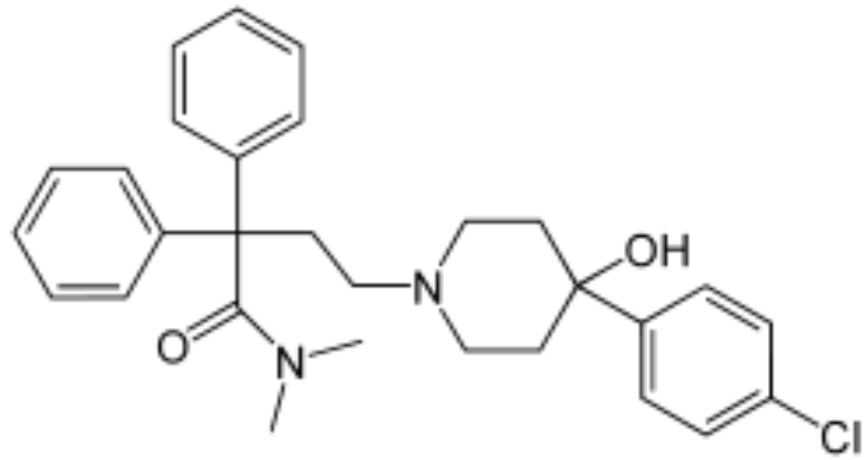
Received February 15, 2008

Solubility is a key physicochemical property of molecules. Serious deficiencies exist in the consistency and reliability of solubility data in the literature. The accurate prediction of solubility would be very useful. However, systematic errors and lack of metadata associated with measurements greatly reduce the confidence in current models. To address this, we are accurately measuring intrinsic solubility values, and here we report results for a diverse set of 100 druglike molecules at 25 °C and an ionic strength of 0.15 M using the CheqSol approach. This is a highly reproducible potentiometric technique that ensures the thermodynamic equilibrium is reached rapidly. Results with a coefficient of variation higher than 4% were rejected. In addition, the Potentiometric Cycling for Polymorph Creation method, $[PC]^2$, was used to obtain multiple polymorph forms from aqueous solution. We now challenge researchers to predict the intrinsic solubility of 32 other druglike molecules that have been measured but are yet to be published.

Up to 40% of drug discovery programs abandoned due to PK problems

- HTS screens have in general increased MW and lipophilicity of leads
- Easy way to gain affinity: More hydrophobicity
- Solubility is a common problem
- Cosolvents (like DMSO) often used in testing, and eliminating them in formulations can lead to unforseen problems
- Designing for solubility really not possible at present aside from rules of thumb

Example: loperamide

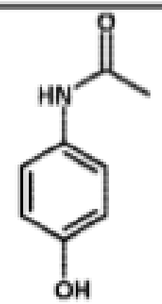
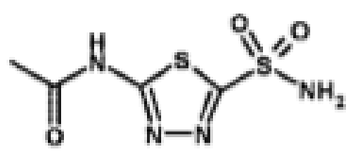
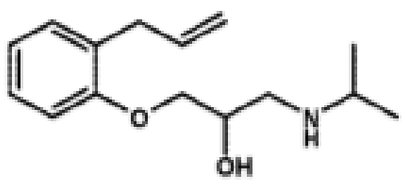


- log S = -4.38 for HCl salt
- log S = -7.07 intrinsic
- So can precipitate when injected in some circumstances

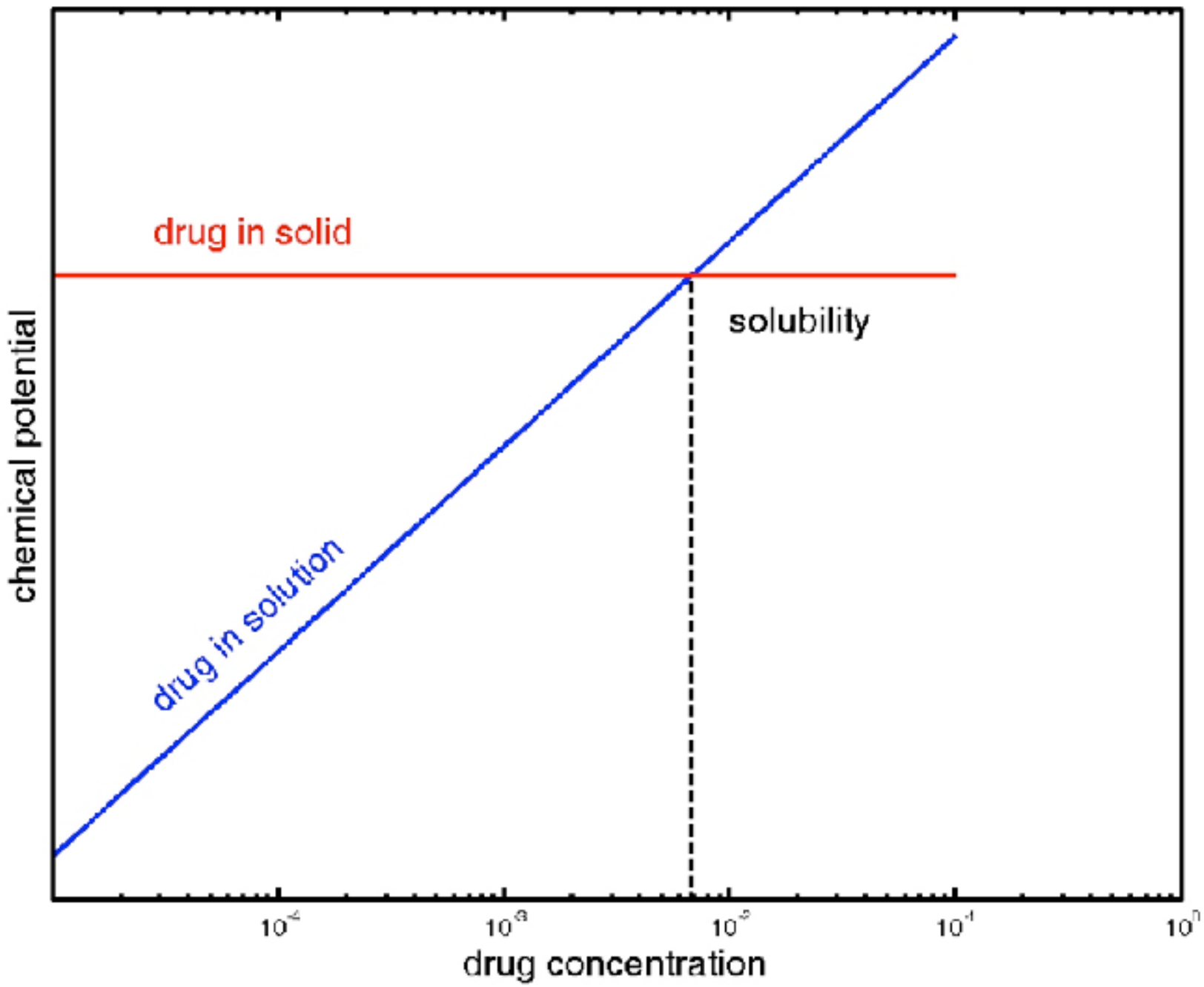
Llinas warns of overfitting

- “Any aqueous solubility prediction method producing standard errors of estimate 0.4-0.5 log unit for a diverse data set is probably overfitting the data.”

Examples from the Llinas dataset

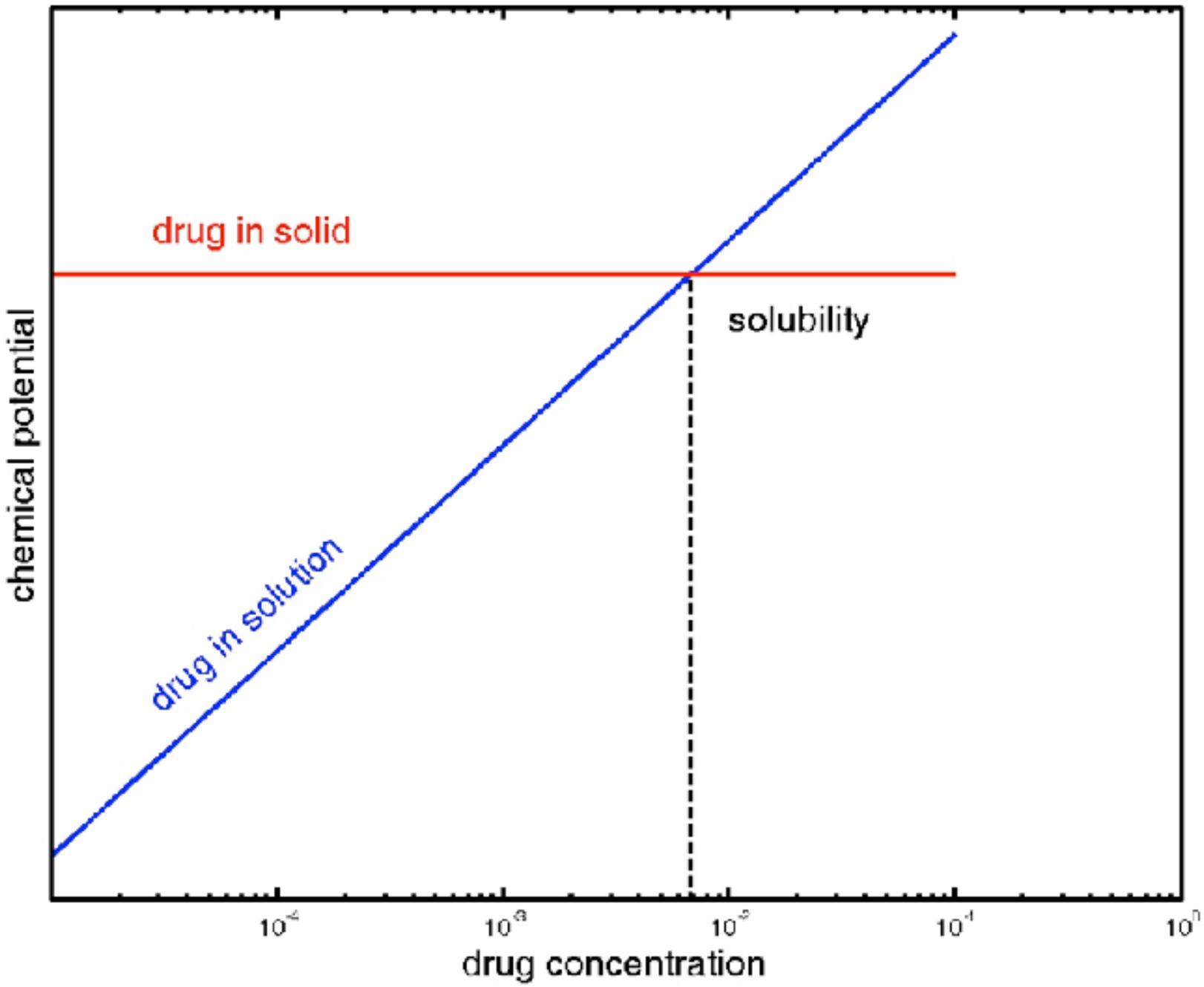
Structure	Name	MW Neutral Form	pKa	Kinetic Solubility µM µg/ml	Intrinsic Solubility µM µg/ml
	Acetaminophen	151.17	9.52 ± 0.01	161700 ± 7000 µM 24400 ± 1060 µg/ml	86300 ± 7000 µM 13000 ± 1060 µg/ml
	Acetazolamide	222.25	8.75 ± 0.02 7.31 ± 0.04	6100 ± 3840 µM 1360 ± 850 µg/ml	3670 ± 80 µM 816 ± 18 µg/ml
	Alprenolol	249.36	9.47 ± 0.01	5080 ± 50 µM 1266 ± 12 µg/ml	2320 ± 40 µM 580 ± 10 µg/ml
	Amantadine	151.25	10.48 ± 0.01	17300 ± 3960 µM 2620 ± 600 µg/ml	14000 ± 1180 µM 2120 ± 180 µg/ml

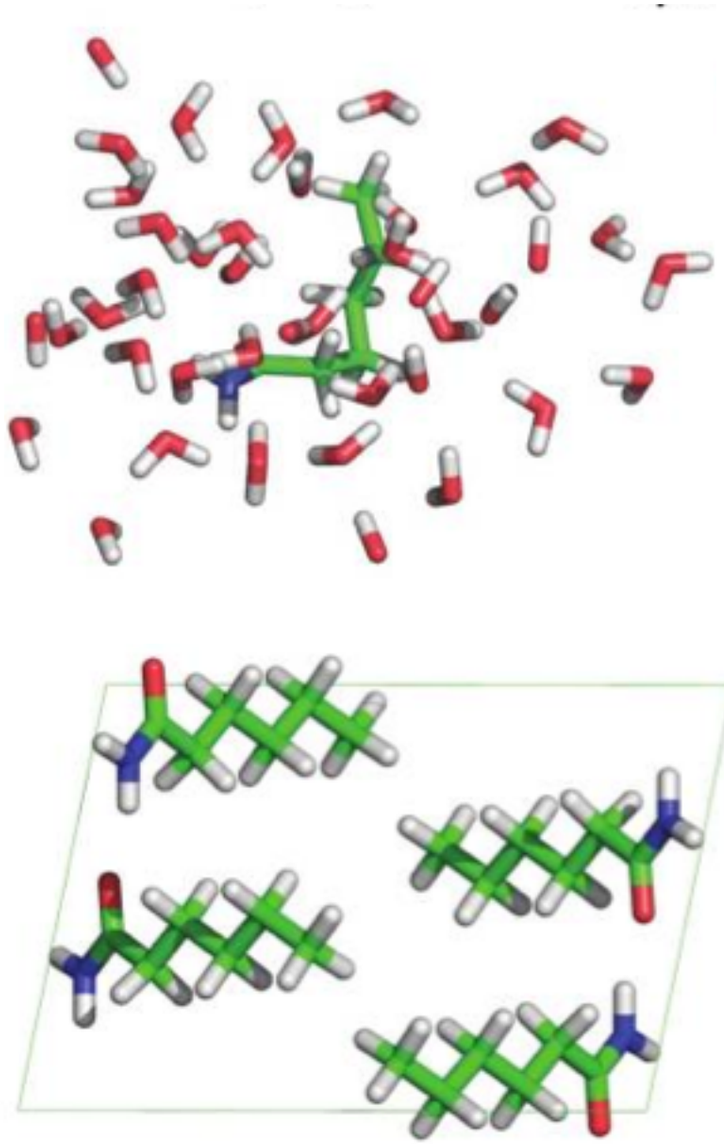
What determines solubility? Relative chemical potential in solution vs solid



(right: Schnieders et al. 2012)

What determines solubility? Relative chemical potential in solution vs solid





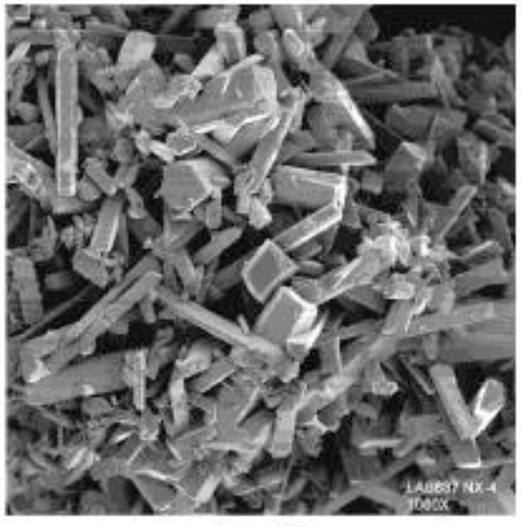
(right: Schnieders et al. 2012)

The crystal structure itself can be particularly problematic and challenging

- Famous example: HIV inhibitor ritonavir
- Similarly:

The crystal structure itself can be particularly problematic and challenging

- Famous example: HIV inhibitor ritonavir
- Similarly:

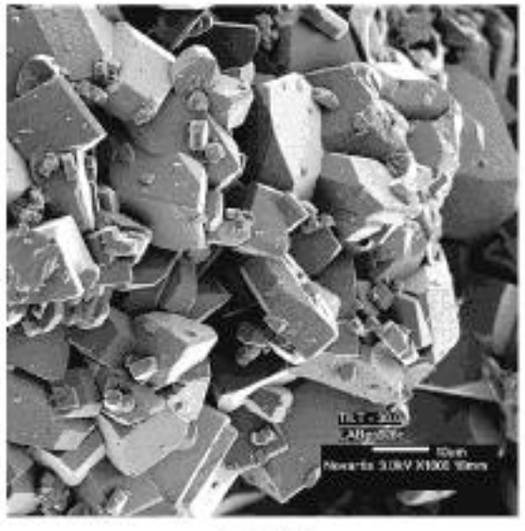


Form B

Was initial material

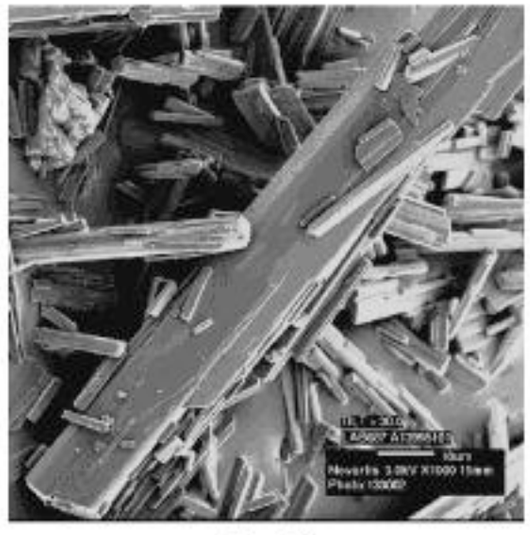
Form A

Next, discovered new polymorph: "Form A". Embarked on polymorph investigation...



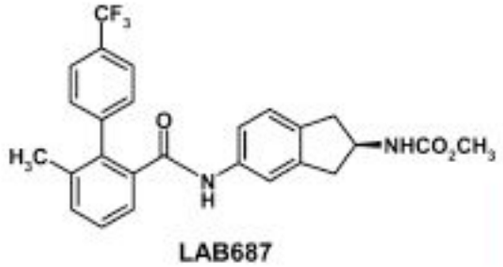
Form C

...and discovered "Form C"; now Form B cannot be made in laboratories



Form D

Selected and scaled up "Form C" in pilot plant, where "Form D" was discovered and began to predominate. Aiiieeeee!



LAB687

Pictures and structure from Prashad et al., Org. Process Res. Dev. 2010, ASAP.
DOI: 10.1021/op100115u

Neglect of polymorphs/solid state probably limits accuracy most empirical/ group-based methods can achieve

- Llinas et al. suggest polymorph differences usually a factor of 7 or less, some suggest less than a factor of 2.

LETTER

Findings of the Challenge To Predict Aqueous Solubility

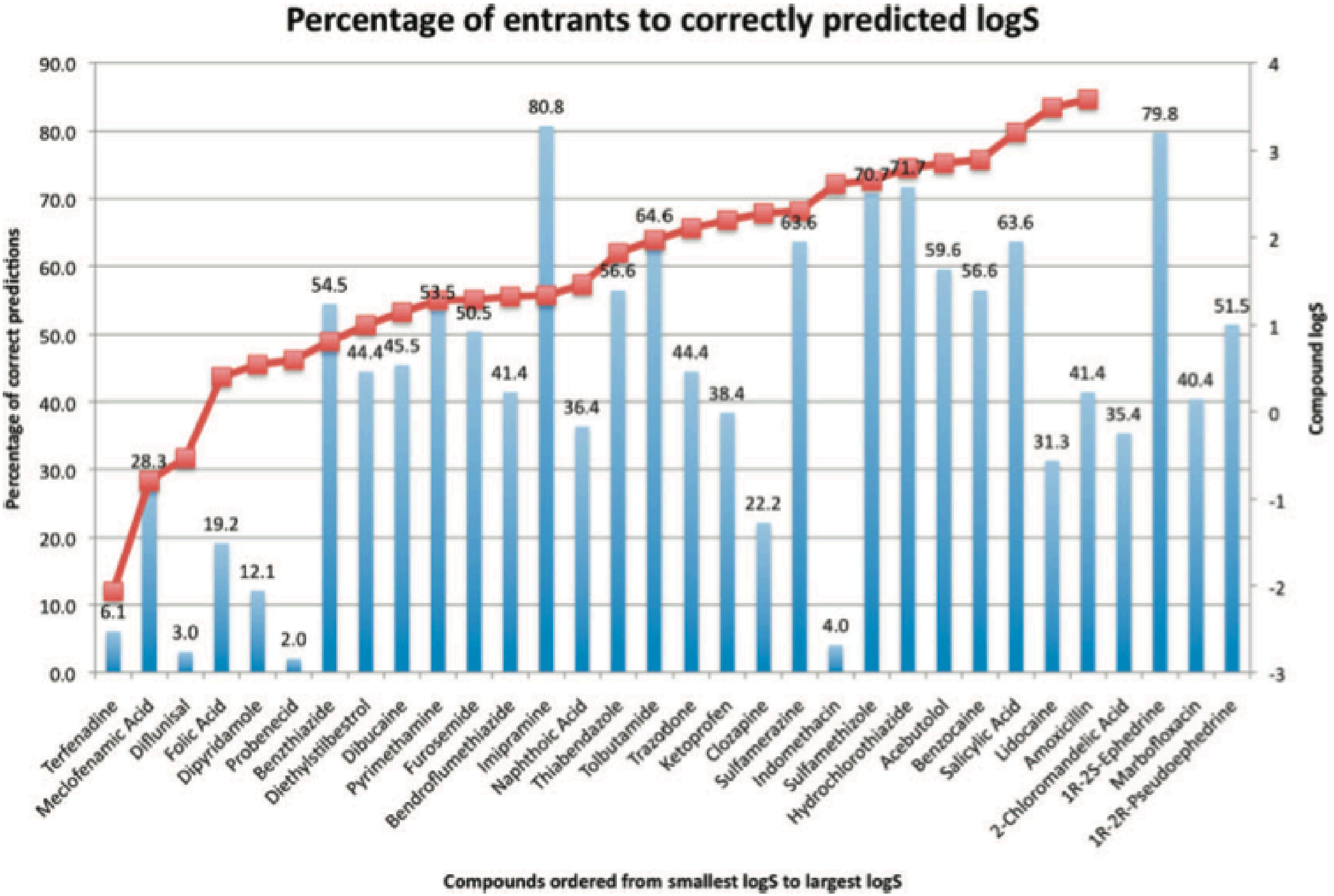
Anton J. Hopfinger,*[,†,‡] Emilio Xavier Esposito,[‡] A. Llinàs,[||,§] R. C. Glen,[§] and J. M. Goodman[§]

College of Pharmacy, University of New Mexico, MSC09 5360, Albuquerque, New Mexico 87131-0001, The Chem21 Group, Inc., 1780 Wilson Drive, Lake Forest, Illinois 60045, and Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

Received December 1, 2008

Table 1. Prediction Set Compounds and Their Measured S Values with Error Limits

compound	measured solubility (μg/mL)	compound	measured solubility (μg/mL)
1 acebutolol	711 ± 140	17 indomethacin	410 ± 5
2 amoxicillin	3900 ± 200	18 ketoprofen	157 ± 4
3 bendroflumethiazide	21.2 ± 2.6	19 lidocaine	3130 ± 100
4 benzocaine	780 ± 150	20 marbofloxacin	too soluble to measure
5 benzthiazide	6.4 ± 0.1	21 meclofenamic acid	0.16 ± 0.01
6 2-chloromandelic acid	too soluble to measure	22 naphthoic acid	28.96 ± 0.01
7 clozapine	188.9 ± 0.7	23 1R-2R-pseudoephedrine	too soluble to measure
8 dibucaine	14 ± 2	24 probenecid	3.9 ± 0.1
9 diethylstilbestrol	10 ± 3	25 pyrimethamine	19.4 ± 2.7
10 diflunisal	four forms:	26 salicylic acid	1620 ± 40


Percentage of entrants to correctly predicted logS
Percentage of correct predictions
Compound logS
Compounds ordered from smallest logS to largest logS
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
4
3
2
1
0
-1
-2
-3
28.3
6.1
3.0
19.2
12.1
2.0
54.5
44.4
45.5
53.5
50.5
41.4
80.8
36.4
56.6
64.6
44.4
38.4
22.2
63.6
4.0
70.7
71.7
59.6
56.6
63.6
31.3
41.4
35.4
79.8
40.4
51.5
Terfenadine
Meclofenamic Acid
Diflunisal
Folic Acid
Dipyridamole
Probenecid
Benzthiazide
Diethylstilbestrol
Dibucaine
Pyrimethamine
Furosemide
Bendroflumethiazide
Imipramine
Naphthoic Acid
Thiabendazole
Tolbutamide
Trazodone
Ketoprofen
Clozapine
Sulfamerazine
Indomethacin
Sulfamethizole
Hydrochlorothiazide
Acebutolol
Benzocaine
Salicylic Acid
Lidocaine
Amoxicillin
2-Chloromandelic Acid
1R-2S-Ephedrine
Marbofloxacin
1R-2R-Pseudoephedrine

Major findings they state

- Percent correct on S (+/-10%) overall: 0%-18%
- Percent correct on logS (0.5 logs): 10-60%
- R-squared range for S: 0 to 0.64:
- No polymorph predictions
- "Normal" solubility compounds (logS=0.5 to 3) were easiest
- Worst two compounds 2% and 4% correct
- Many methods, in general empirical
- No clear winner - a long ways to go

In Silico Prediction of Aqueous Solubility: The Solubility Challenge

M. Hewitt,* M. T. D. Cronin, S. J. Enoch, J. C. Madden, D. W. Roberts, and J. C. Dearden

School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, England

- Most solubility approaches based on some type of quantitative structure-activity relationship (QSAR)
- Compared four types of QSAR models:
  - Multiple linear regression (426 descriptors, genetic algorithm, no more than 5 used at once)
  - Artificial neural network
  - Chemical category formation
  - Various commercial models
- Consensus

The best approach ended up being a simple three-descriptor linear regression

$$\log(\text{intrinsic solubility}) = 7.706 - 0.613 \log P - 0.007\text{BPt} - 5.404\text{R2e} + \qquad (2)$$

$$R^2_{(\text{Train})} = 0.74 \quad R^2_{(\text{Val})} = 0.67 \quad R^2_{(\text{Test})} = 0.51$$

$$\text{RMSE}_{(\text{Test})} = 0.95$$

$$s = 0.77 \quad F = 216.53$$

The best approach ended up being a simple three-descriptor linear regression

log(intrinsic so

$$R^2_{(\text{Train})} = 0.74$$

$$s = 0.77 \quad F =$$

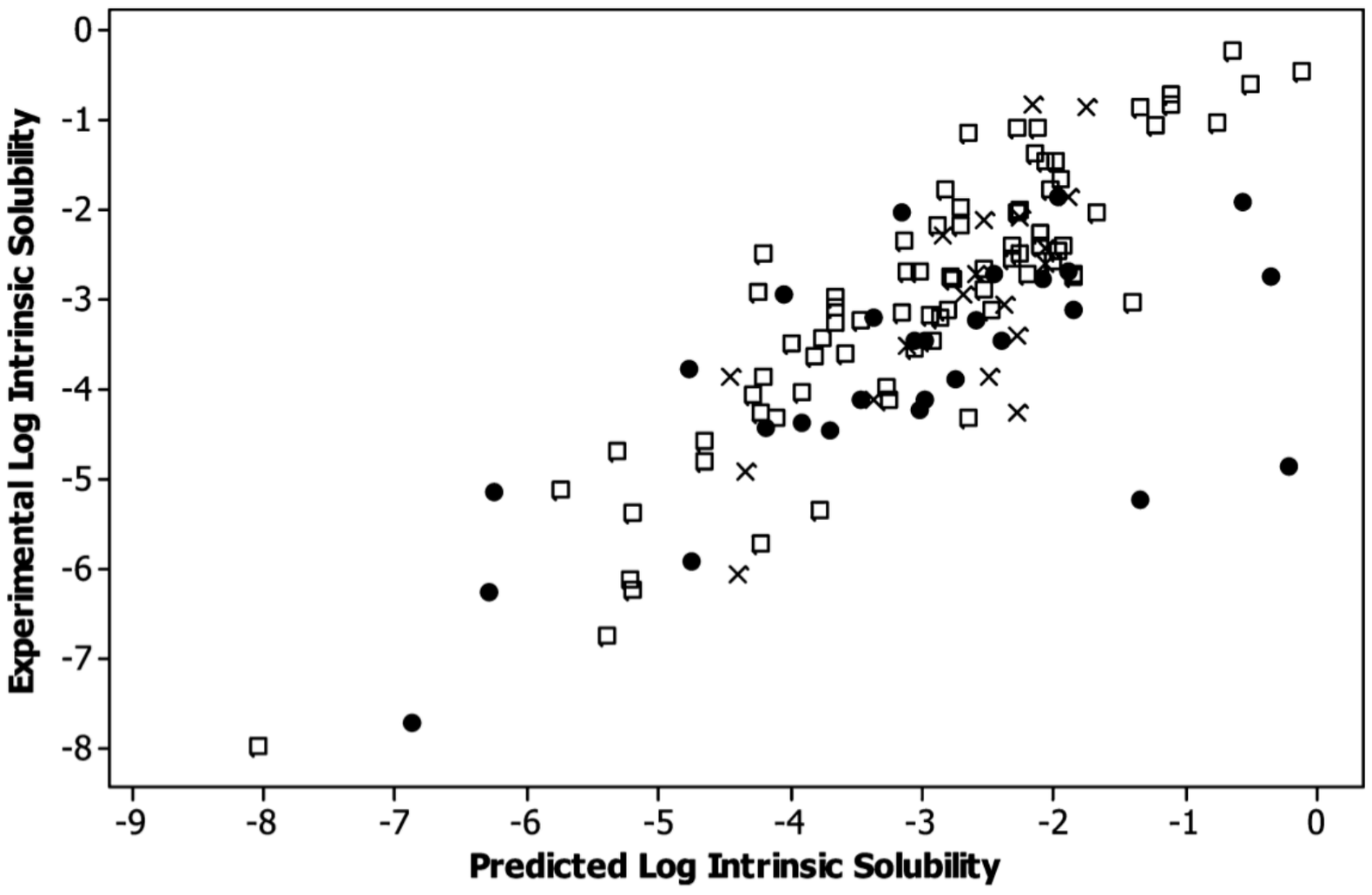


Figure 1. Plot of experimental versus predicted log(intrinsic solubility) using the MLR model (eq 2) for (□) the training set, (×) the validation set, and (●) the test set.

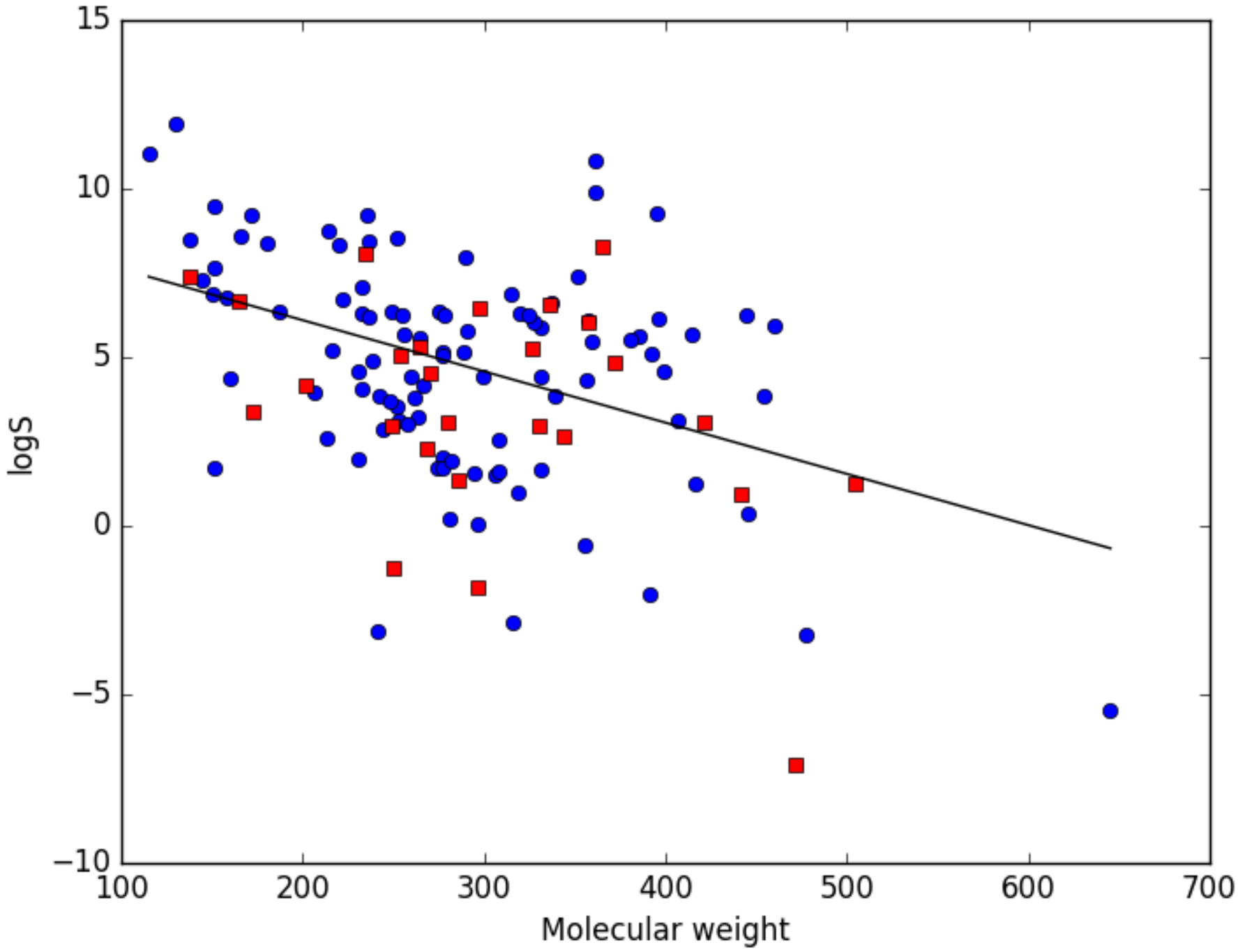
They looked at a wide variety of other approaches as well, but in general none were quite as good

model	$R^2_{(Test)}$	*s*	*F*	RMSE
ChemSilico (CSLogWS)	0.35	1.10	15.56	1.24
Optibrium (StarDrop)	0.28	1.15	11.69	1.38
Pharma Algorithms	0.19	1.22	7.48	1.45
Simulations Plus−SHARK	0.57	0.90	36.59	0.90
in silico consensus	0.38	1.07	17.54	1.14
Simulations Plus−YINAN	0.55	0.92	33.72	0.91
Simulations Plus−UIQBB	0.65	0.81	50.90	0.82
Simulations Plus−LGGAV	0.53	0.93	31.70	0.93
Simulations Plus−A69EM	0.57	0.90	36.04	0.90
Simulations Plus−NSLIC	0.54	0.93	32.60	0.93
Simulations Plus−AM108	0.61	0.85	43.66	0.87
Simulations Plus−OLASM	0.55	0.91	34.51	0.92
SPARC Solubility Model	0.50	0.96	28.17	1.56
New *in silico* consensus	0.60	0.68	41.39	0.90

The physical property/solubility assignment involves tackling the Llinas et al. challenge

- Build simple descriptor-based linear models
- Train on the training set
- Test on the prediction set
- Cross-compare with other methods used in the challenge

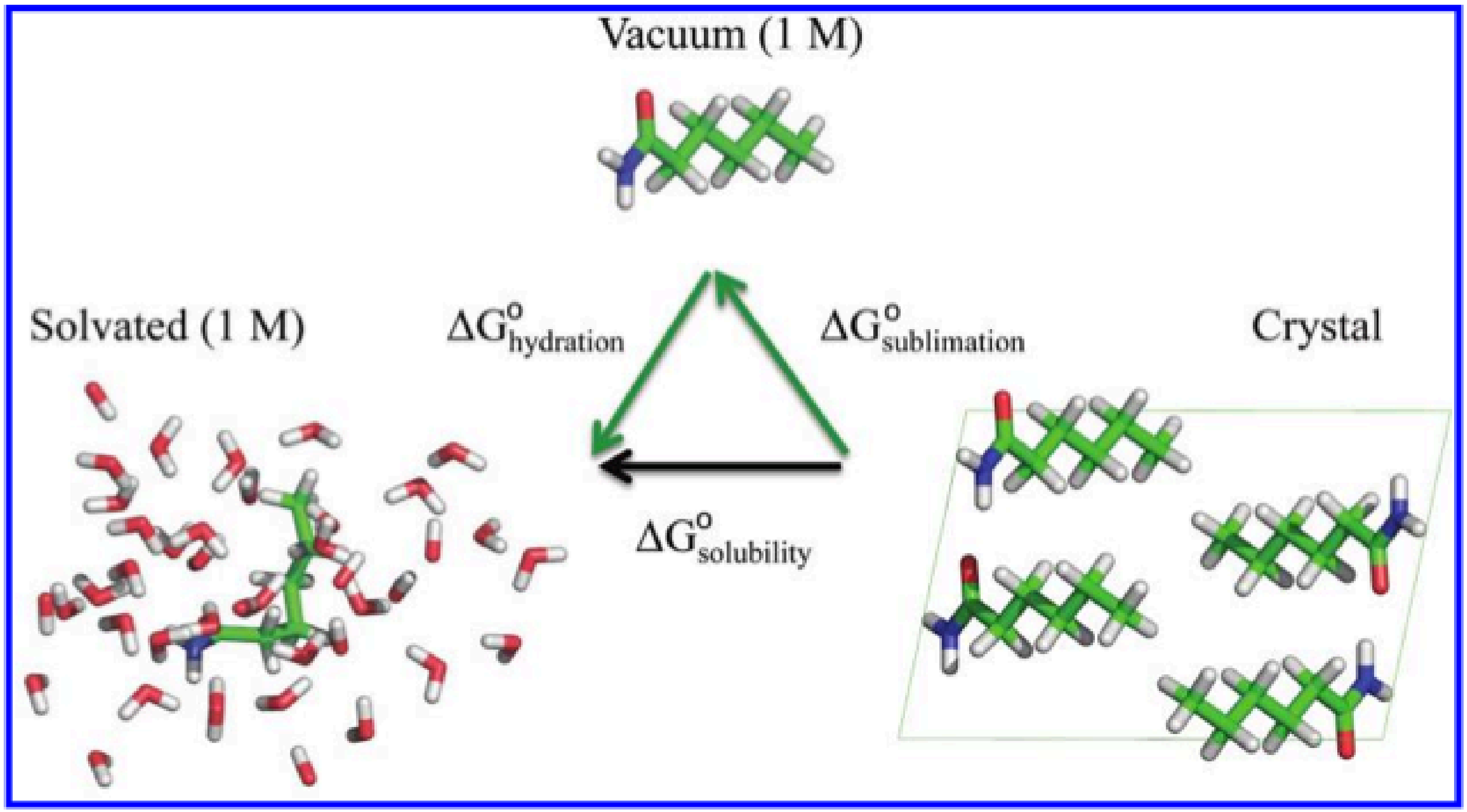
It turns out even molecular weight has some predictive value



One previous student tried all linear four-descriptor models (using the descriptors provided in the assignment):

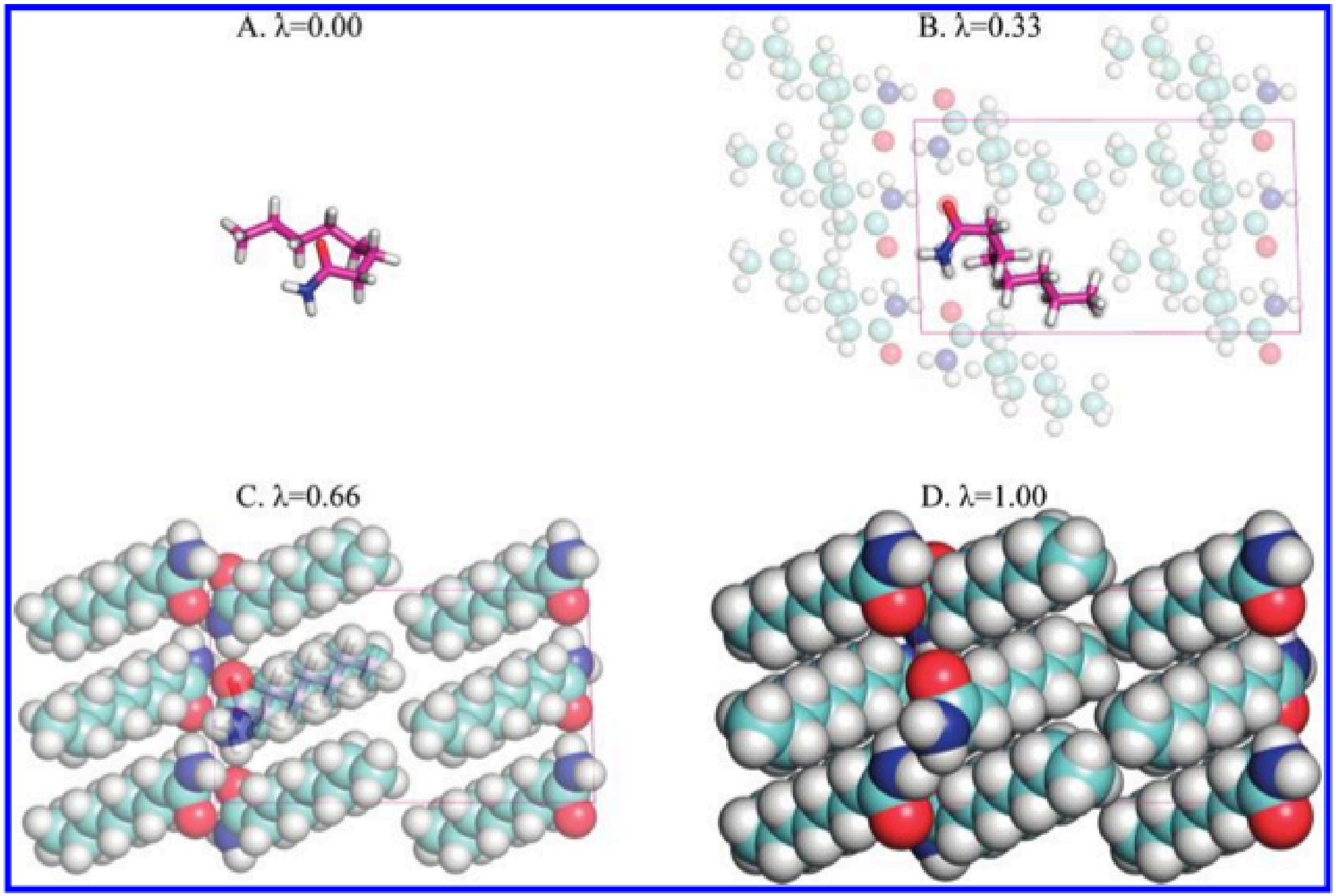
"**In terms of R^2, the best linear four-descriptor model is only out-performed by two models in the Hopfinger paper, out of the 99 entries in the Solubility Challenge** -- AM108 (R^2 = 0.620) and UIQBB (R^2 = 0.650). Linear models are essentially oriented toward minimizing R^2 in particular (least squares fitting is closely related to R^2 maximization), so relatively high R^2s are to to be expected for linear models, especially as the number of descriptors increases. However, many models from the Solubility Challenge outperformed even the best linear four-descriptor model in terms of percent within a half-log of the true solubilities -- this is surprising to me, especially considering how well the models did in terms of R^2. My only guess as to why this could be the case is that **many of the models from the Solubility challenge did not constrain themselves to linear models, so they mad be able to get within a half log of a larger portion of the data set**, sacrificing a bit of precision (lower R^2s) for the advantage of accuracy at a broader range (higher percentage within a half-log)." — *David Wych*

Some efforts are taking solubility prediction in more physical directions



Schnieders et al., JCTC 8:1721-1736 (2012)

Sublimation is calculated via alchemical techniques, as is solvation



Schnieders et al., JCTC 8:1721-1736 (2012)

Results on an initial series appear promising without any empirical tuning

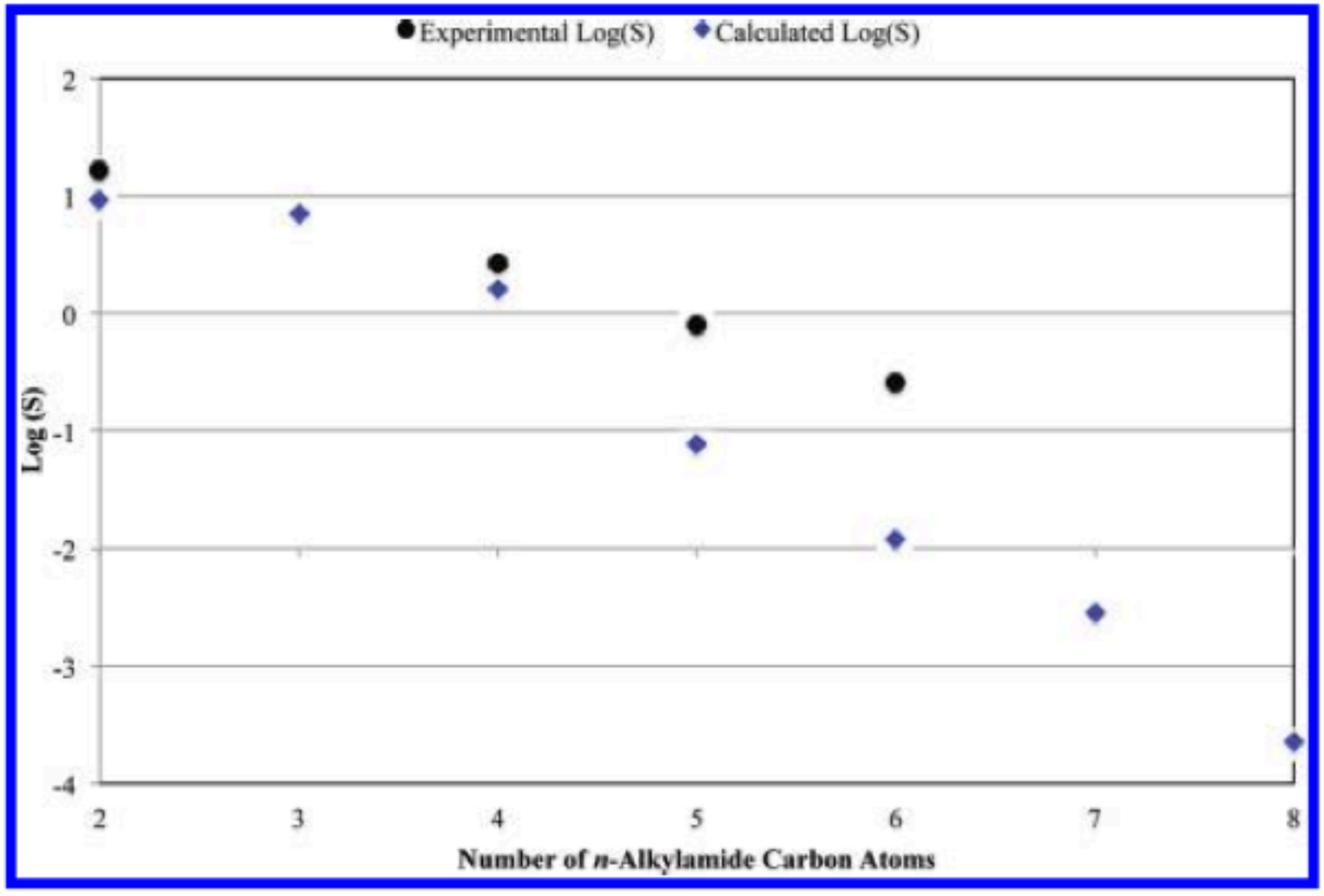
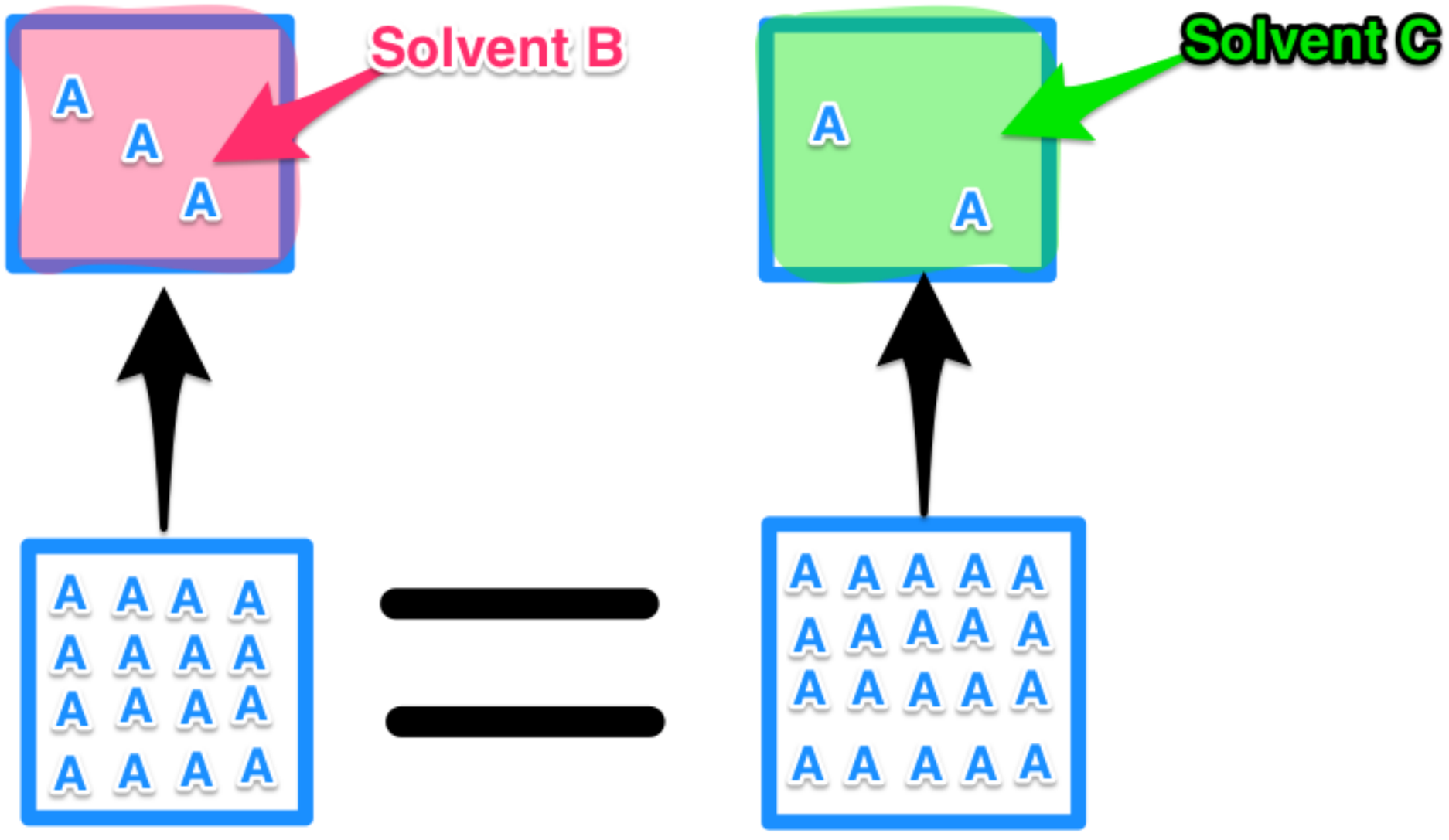


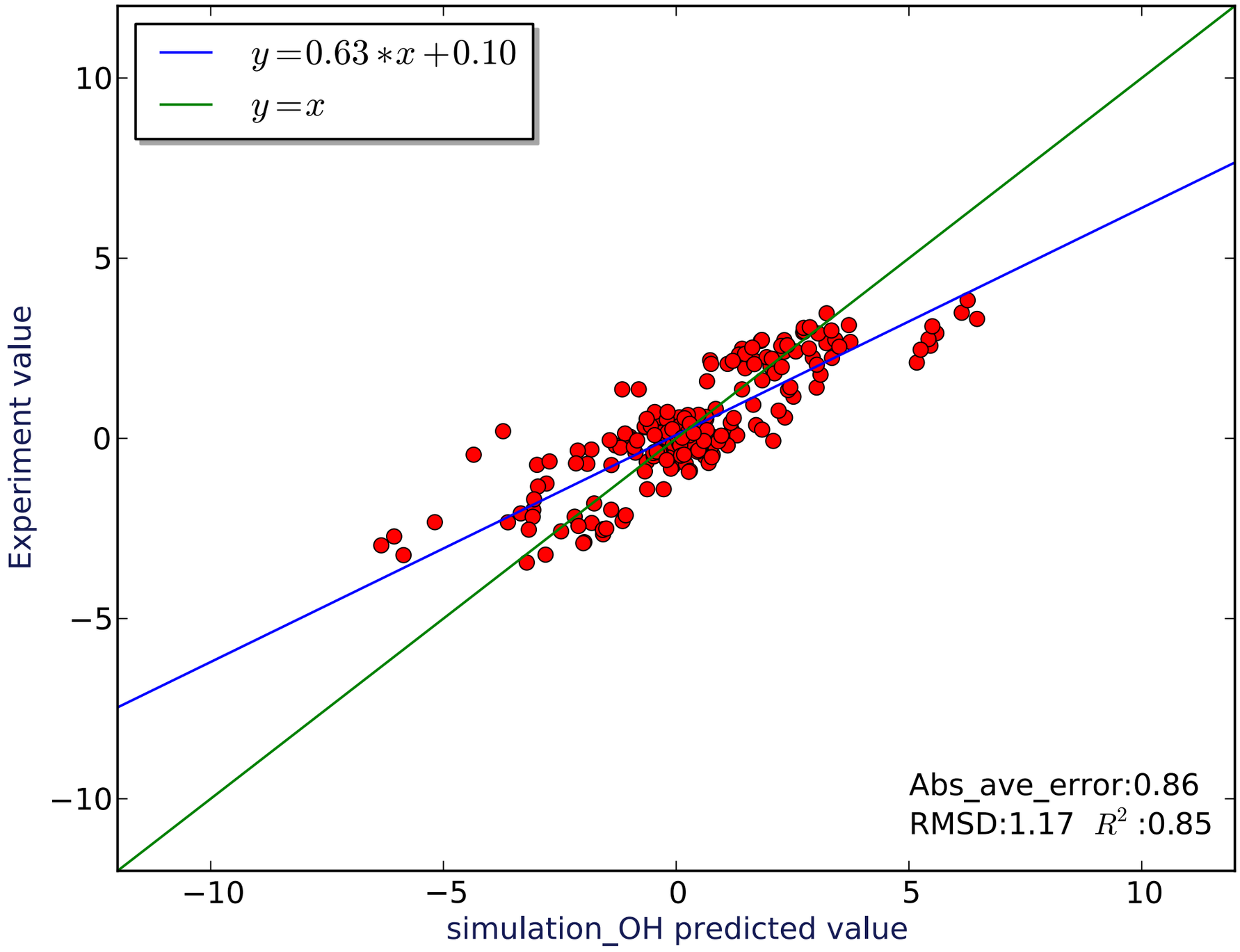
Figure 5. Shown are experimental and calculated log(*S*) values for the *n*-alkylamides (*S* has units of mol/L) from acetamide to octanamide. There is a monotonic trend in both the experimental and calculated values toward lower solubility with each additional CH_2 group due to increasingly favorable deposition and to a lesser extent from unfavorable solvation.

Schnieders et al., JCTC 8:1721-1736 (2012)

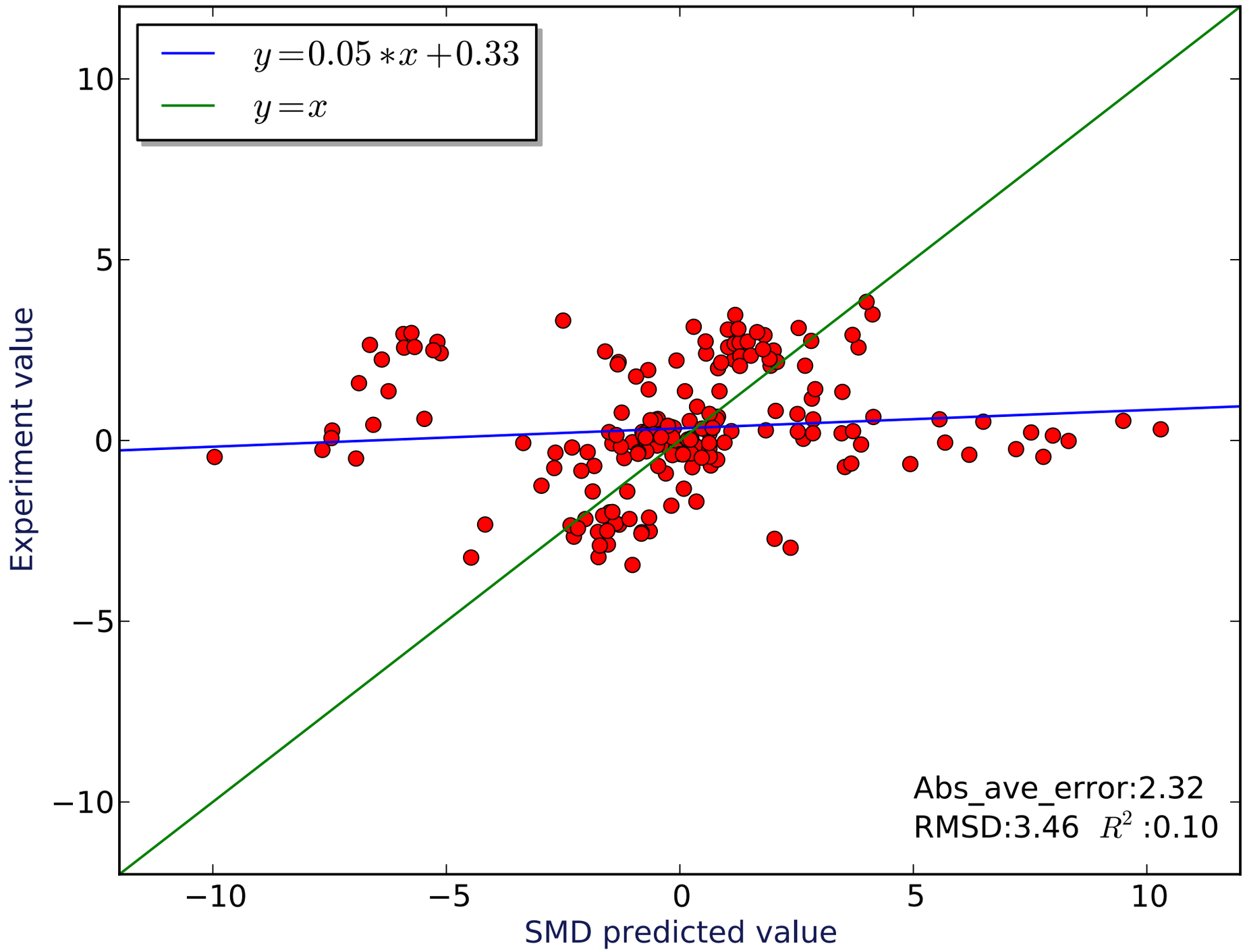
It turns out that relative solubilities can be calculated with much less effort



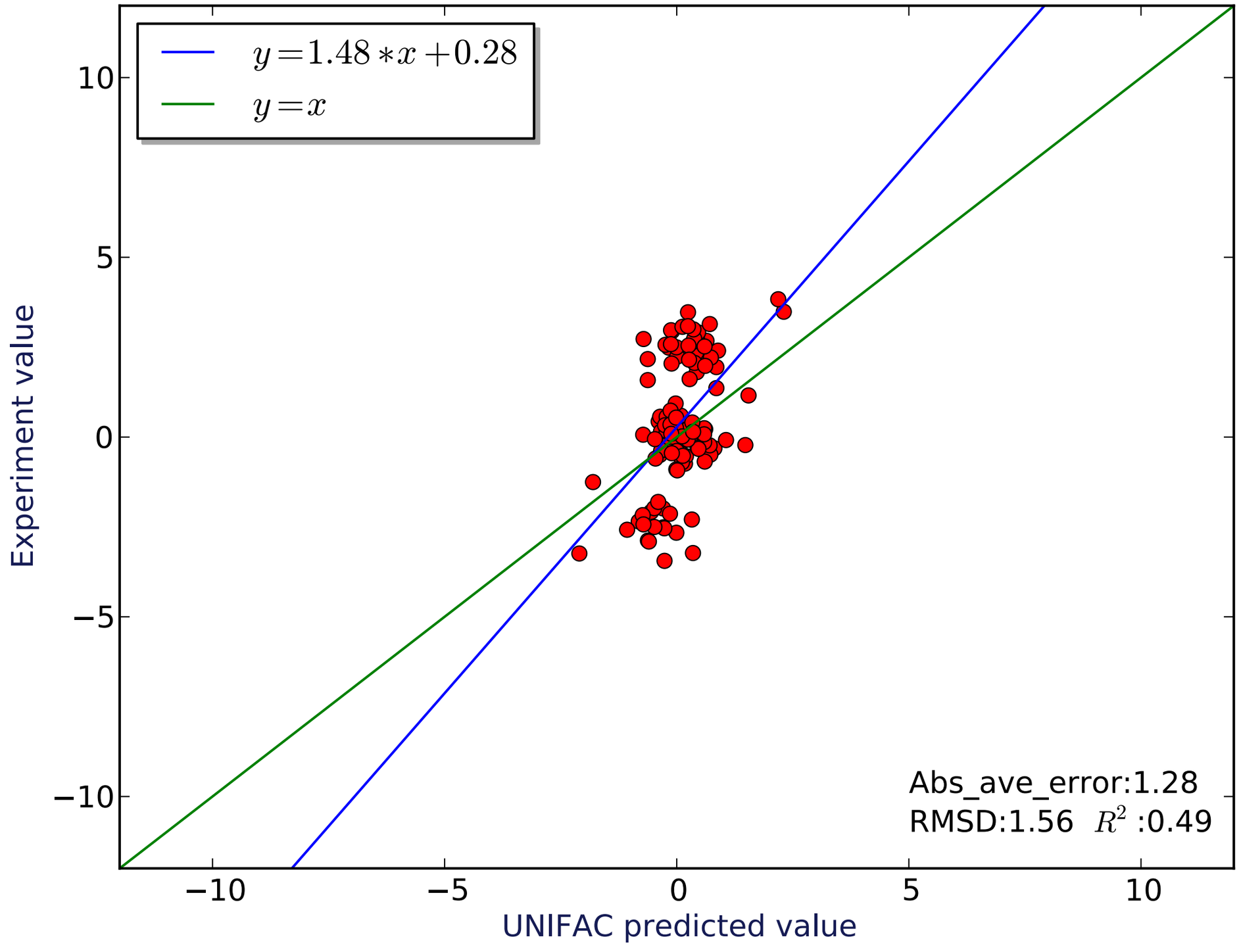
On the whole calculated and experimental values agree rather well



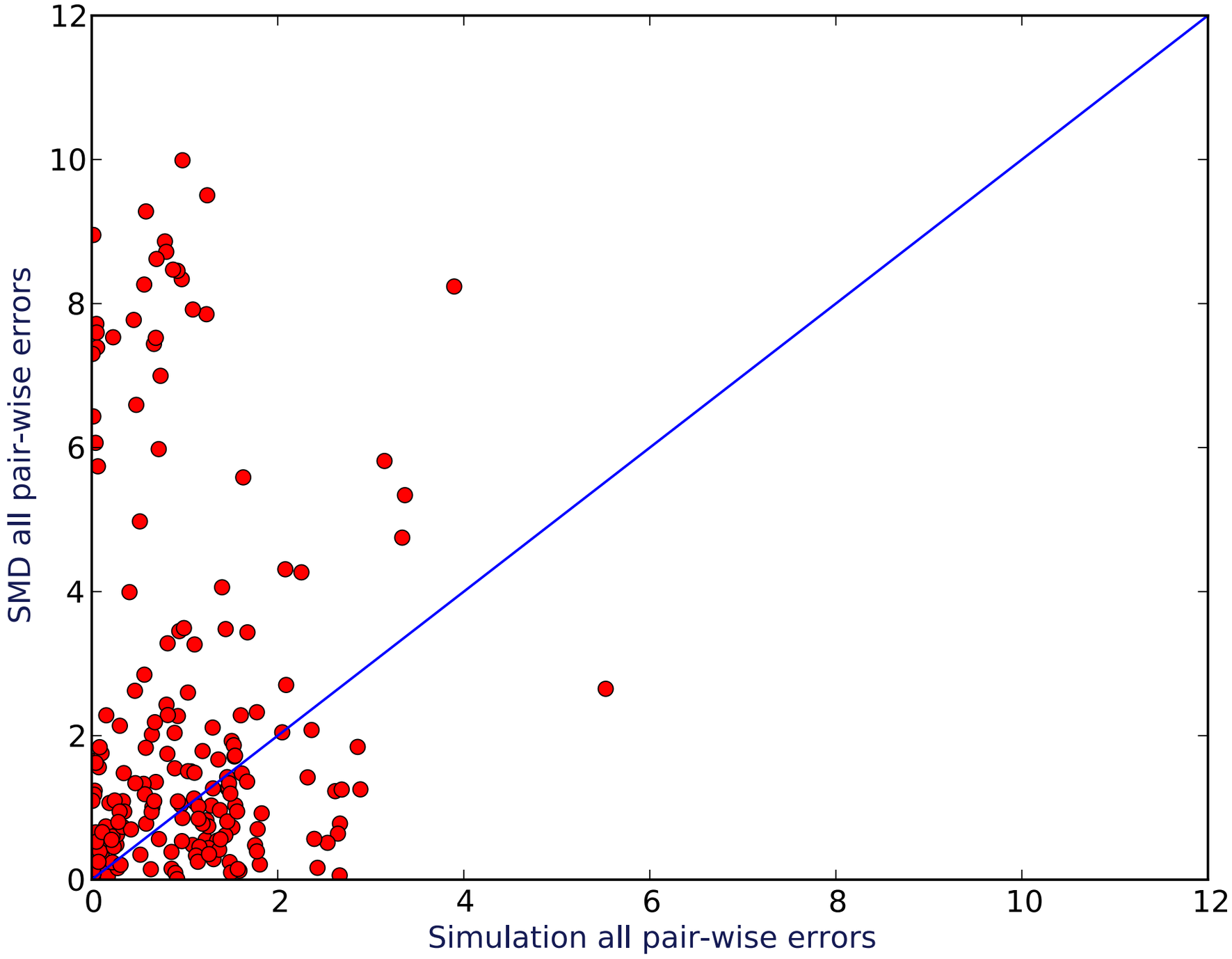
For comparison, SMD is a QM-based solvation model with empirical solvation parameters



UNIFAC does somewhat better but does not cover the whole set



To look at ours vs SMD in another way:



Back to QSAR

- Feature space is key
- The structural features you use in your model have to be adequate to describe what needs to be known to predict the property of interest
- QSAR models have broad application not just in solubility but many other areas
- Including attempts to apply to predict binding (though usually ill-advised there)
- Best when very large data sets are available and properties can reasonably be expected to have some dependence on computable descriptors

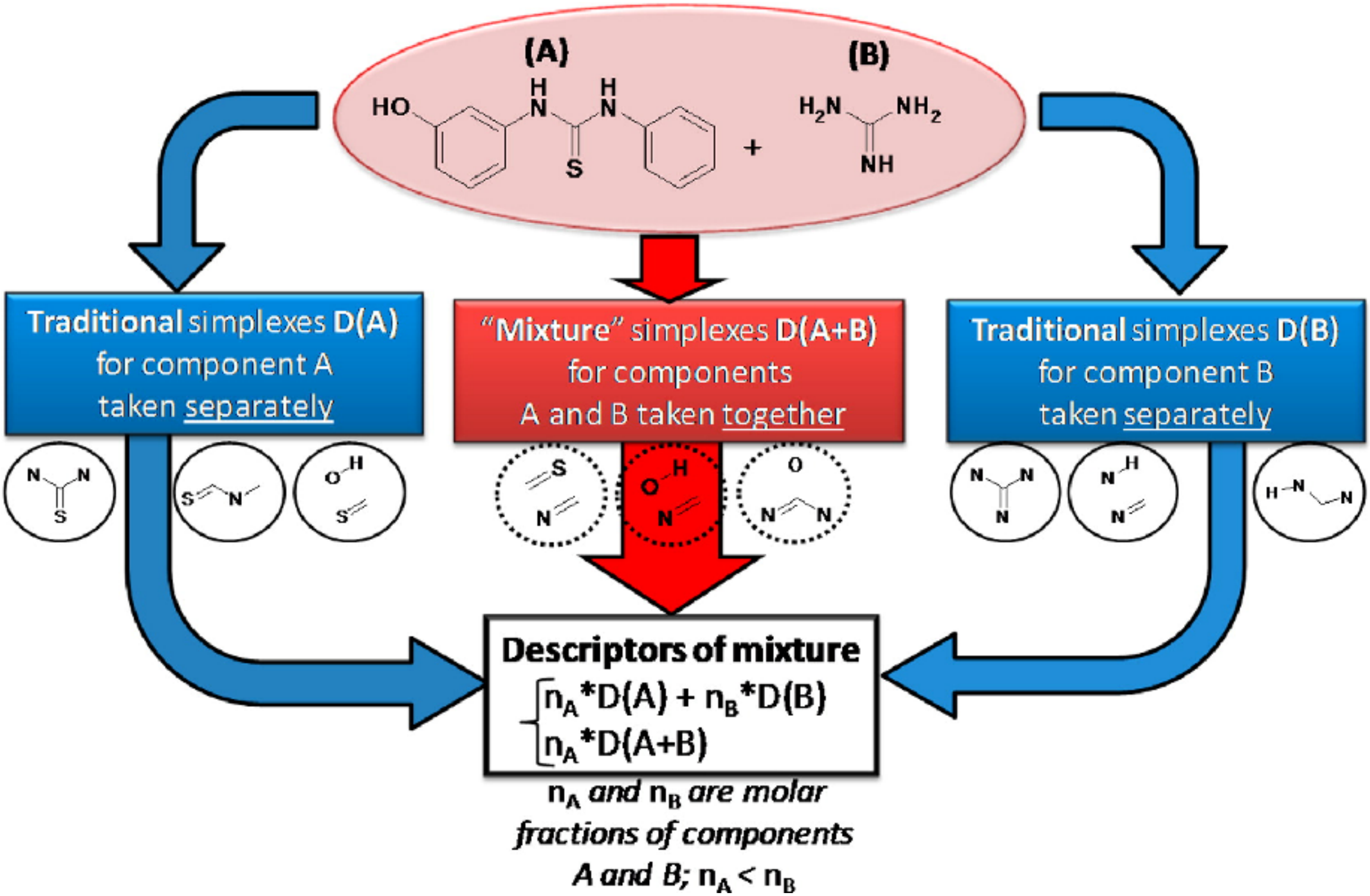
QSAR has a fundamental assumption

- "…[the] fundamental basis of any QSAR investigation: the only possible cause of any difference in biological activity of two structures is their structural difference, regardless of how complex may be the physicochemical and biological interactions that subsequently connect that structural difference to an observed biological effect." (Cherkasov et al., JMC 2013 10.1021/jm4004285)
- Key design criteria: Far more data than descriptors (certainly at least 5:1)

Other targets for QSAR

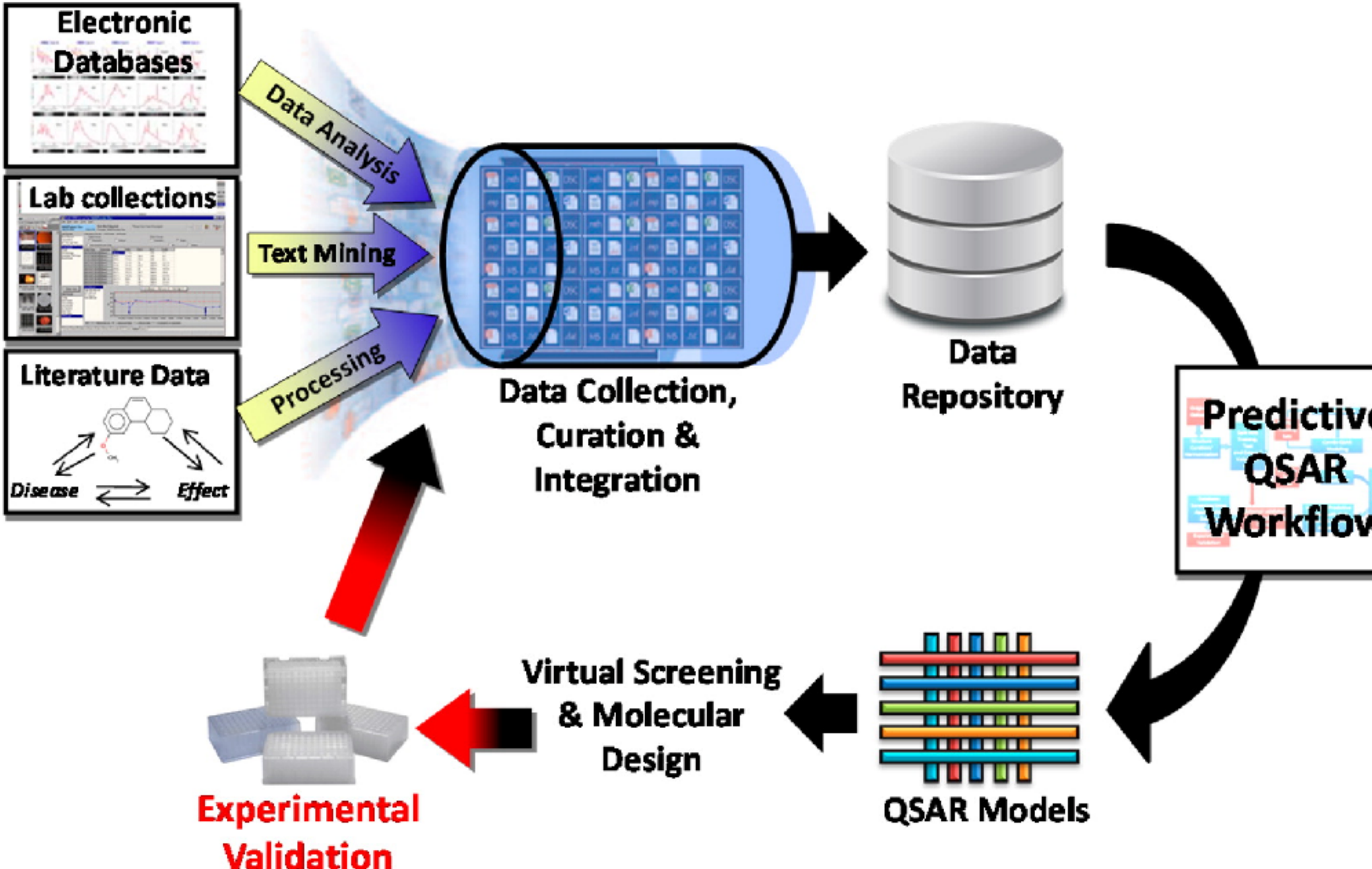
- Toxicology
- Metabolism
- Antimicrobial activity of peptides
- Mixture properties
- …
- (Cherkasov et al., JMC 2013 10.1021/jm4004285)

QSAR modeling can get quite sophisticated

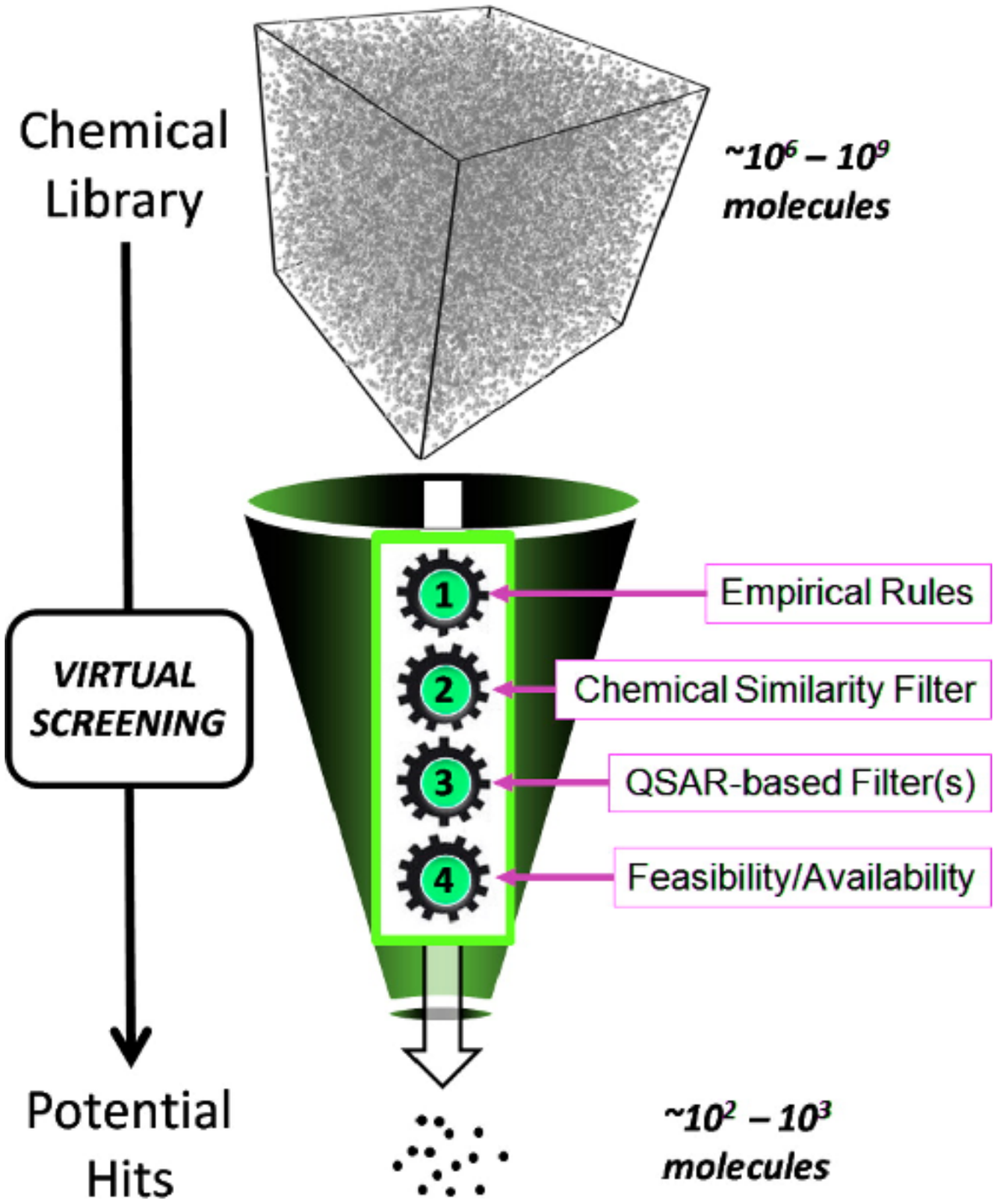


Cherkasov et al., JMC 2013 10.1021/jm4004285

QSAR training is on large datasets



Cherkasov et al., JMC 2014 10.1021/jm4004285



Usually QSAR is used in database filtering

Cherkasov et al., JMC 2013 10.1021/jm4004285