

# 11/16/2022

- Quiz 3: Molecular Simulations
- Innovation Day
- Quantitative Structure-Property Relationships
- Solubility Prediction

# FALL 2022 INNOVATION DAY

## OVERVIEW

# Innovation Day Overview

At the Kaplan Institute, we understand that practical, technological breakthroughs only come from people who are not just good at technology, but good at leadership: people who understand that influence, inspiration and human insight are as important as engineering and computer science and clean rooms; people who are looking to drive positive change, both in their communities and in the world, and see technology as the means.

The Innovation Day presentation is in effect the final exam for IPRO projects. The presentations demonstrate the entire development process requiring teams to assemble all work done over the course of the semester and present that work to the partners in person.

# Innovation Day Project Submission

In order for us to be able to develop exhibit materials as well as distribute content with our partners, all teams must submit their project deliverables by **November 30**. The deliverables may include the following:

## **Innovation Day Deliverables**

1. Presentation Slide Deck
2. Research Deck and/ or Written Report
3. Supporting Media/ Documentation

(Poster, Videos/ Animations/ Renderings/ Documentation of Prototypes)

Please submit all relevant project materials using the following the link **below**:

- [Team Submission](#)

# Innovation Day Schedule

**Innovation Day: Friday, Dec. 2nd**

**8:00 - 9:00 AM**  
**Team setup**

**9:30 - 12:30 PM**  
**Doors open/ Judging begins**

**12:30 - 1:30 PM**  
**Judges' deliberation & lunch**

**1:30 PM -2:00 PM**  
**Recognition ceremony**

# Innovation Day Judging Tracks

- Sustainable Cities
- Community and Social Innovation
- Sports and Technology
- Competitions
- Research

# Judging Rubric

<b>COMMITMENT TO POSITIVE COMMUNITY CHANGE</b>	Did the team identify and analyze a contemporary issue and problem? Did the team address and compare different perspectives, both within and across cultures?
<b>CRITICAL AND INNOVATIVE THINKING</b>	Did the team appropriately employ multiple quantitative and qualitative methods of analysis and evaluation? Did the team employ the best available technology to achieve the solution(s)?
<b>PROFESSIONAL AND ETHICAL COLLABORATIONS</b>	Did the team members successfully work with others within disciplines and cultures? Did the team identify and discuss ethical issues?
<b>EFFECTIVE COMMUNICATION</b>	Did the team speak and write appropriately within and across disciplines and cultures? Did the students present their project in clear, concise, and logical form both verbally and through visual/graphic supporting material? Did the team clearly establish an objective and cohesively support it?

[Click here for the evaluation form](#)

# Exhibition Space

Each team will have approximately 2' x 3' of table space with access to power.

Your team needs to be able to stand out.  
**Build prototypes!!!**

Bring your own displays (computer monitors or televisions) for more visuals.

Please have your students fill out [this form](#) for special requests regarding equipment/space needs.

# Poster Guidelines & Best Practices

- Design all graphics for 24" x 36" layout. Do **NOT** scale up from a smaller page size
- Use no less than 24pt font
- Print a black and white full size version to mockup, then adjust image sizes and fonts as needed
- Check all spelling BEFORE you print
- Send files to the Post Office Mail Center at MTCC

# Poster Printing

The IPRO office is offering vouchers for teams to print 24" x 36" posters that can be mounted to foam core.

- Please visit the Kaplan Welcome Desk or the IPRO Help Desk for a voucher to print your poster at the MTCC Post Office (2 vouchers/team).
- Vouchers will become available from **November 9th onwards**.
- Foam core will be available at the Idea Shop from **the week of November 14th**.

# Event Deadlines & Requests

- Teams should fill out [this request form](#) by **Friday, November 25** for any special requests for Innovation Day. We cannot honor any requests made after that date.
- Teams are required to submit their project deliverables by **Wednesday November 30th [here](#)**.
- Teams must have their project space set up by 9 AM on **Friday, December 2nd**.

# Prepping For Innovation Day

## Make A Plan

Envision the experience you wish to deliver, what is the sequence of movements, and set goals for that.

## Visualize

Imagine the exhibition space, then sketch out how you plan to set up your table. How will you display your visuals, prototypes etc.

## Practice

What is the story you want to tell, how do you want the audience to feel and react. Presentations need to be 3-5 minutes.

## Work With Your Professors

Professors can help determine ways to best portray your work

## Divide & Conquer

Everyone on the team needs to contribute, work to each others strengths.

# Using Video & Multimedia

- Video & media is a great medium to tell your story and illicit emotion
- Remember audio will be hard to hear on Innovation Day
- Videos & animations can effectively demonstrate functions of your concept

*\*Videos/ Media is not required for  
Innovation Day*

# Resources

# IPRO Office Resources

## Supplies and Purchasing

1. <https://ideashop.iit.edu/portal/>
2. <https://ipro.iit.edu/class-resources/>

## Software on Campus

1. <https://ots.iit.edu/classrooms-labs/lab-software>
2. [https://wiki.ideashop.iit.edu/index.php?title=Computer\\_Software](https://wiki.ideashop.iit.edu/index.php?title=Computer_Software)

# Visual Assets

## Icons

- The Noun Project: <https://thenounproject.com>
- Font Awesome: <https://fontawesome.com/icons?d=gallery>
- Dryicons: <http://dryicons.com/free-icons/>
- FREEBIESBUG: <http://freebiesbug.com/psd-freebies/icons/>

## Free Imagery - Photography

- UNSPLASH: <https://unsplash.com>
- PixaBay: <https://pixabay.com>
- Free Images: <http://www.freeimages.com>
- PEXELS: <https://www.pexels.com>
- PICJUMBO: <https://picjumbo.com>
- PixaBay: <https://pixabay.com>

# Visual Assets

## Video Clips (Free - select ones on sites)

- <https://www.wedistill.io/>
- <https://coverr.co/>
- <https://www.pexels.com/>
- <https://pixabay.com/>
- <https://www.videvo.net/>
- <https://www.videezy.com/>
- <https://www.vidsplay.com/>
- <https://mazwai.com/>

More generic, aerial & drone shots:

- <https://www.lifeofvids.com/>
- <https://dareful.com/>
- <https://www.splitshire.com/>

# Interaction Design Tools

**Figma:** <https://www.figma.com/>

Figma for beginners: <https://www.youtube.com/watch?v=Saz6S1svYns>

**Sketch:** <https://www.sketch.com/>

A guide to sketch basic: <https://www.youtube.com/watch?v=qywB0JHQeC4>

**Mockplus:** <https://www.mockplus.com/>

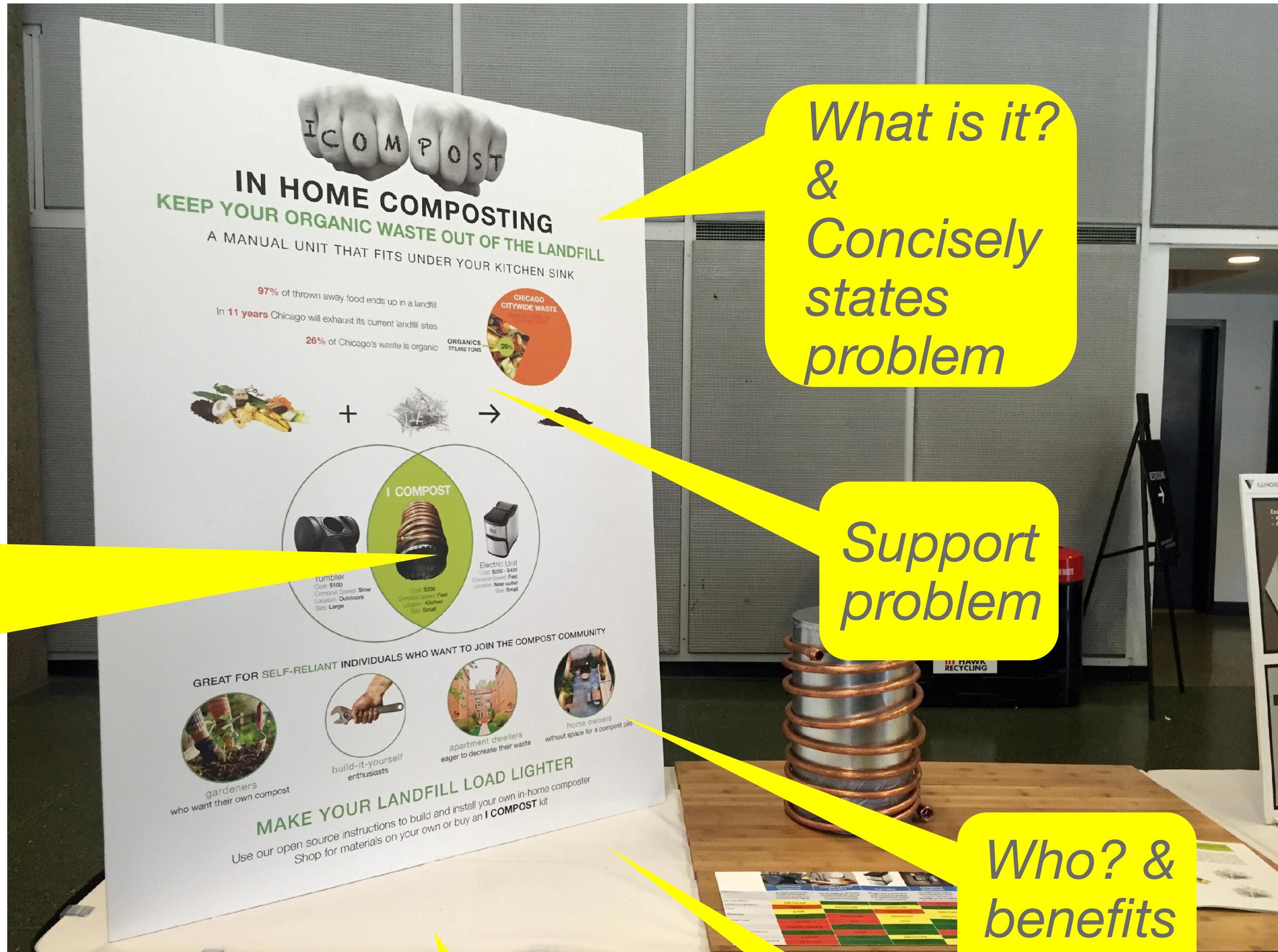
Mockplus Tutorial: <https://www.youtube.com/watch?v=LGZDZ-L0ejo>

**Adobe XD** (can use from Idea Shop at Kaplan: [Ideashop Link](#))

Top ten things to know when getting started with Adobe XD: <https://www.youtube.com/watch?v=JtI6YpmPGI>

# **Examples for Innovation Day**





Strategy  
(this team included cost here bc it was DIY)

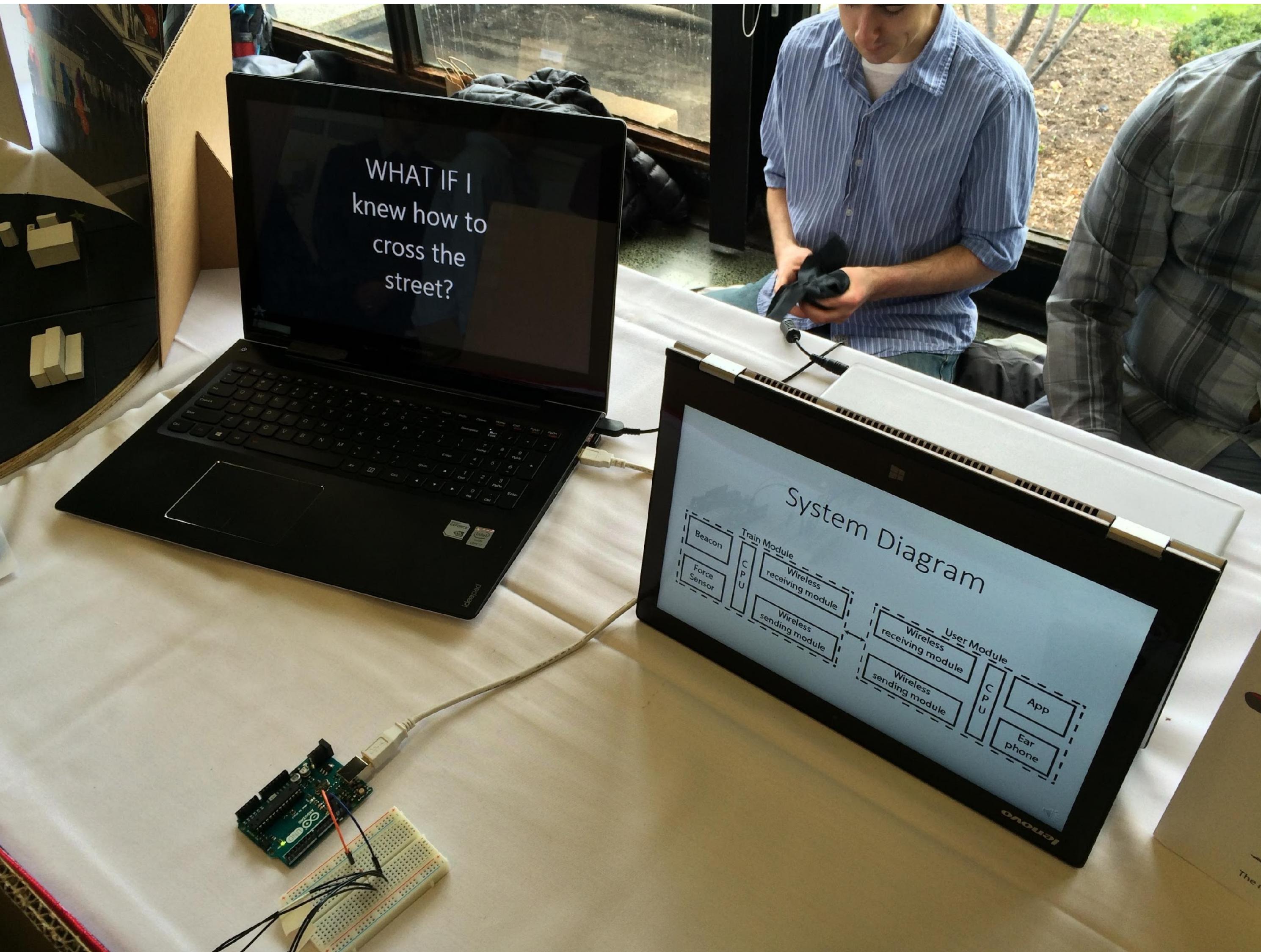
What is it?  
&  
Concisely  
states  
problem

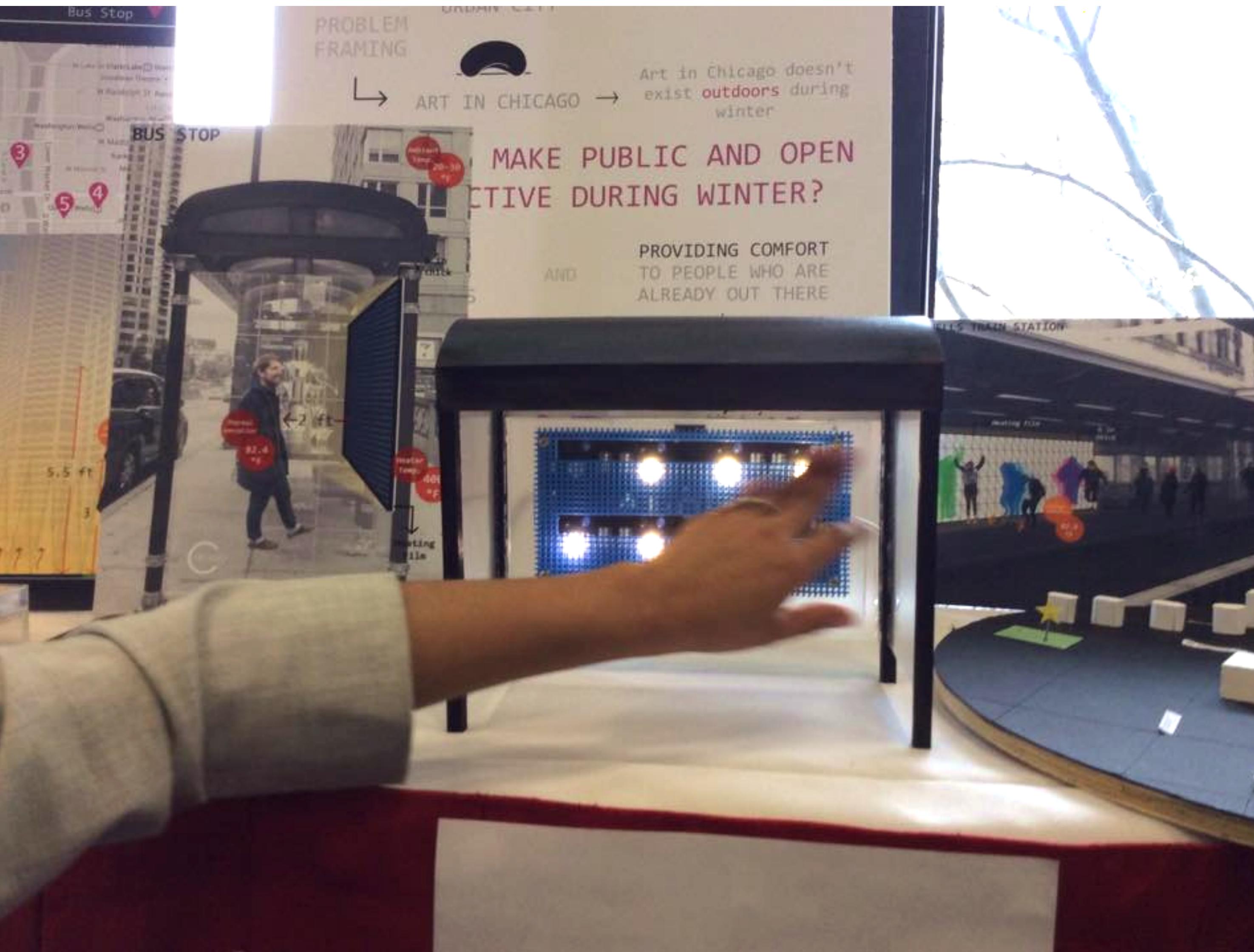
Support  
problem

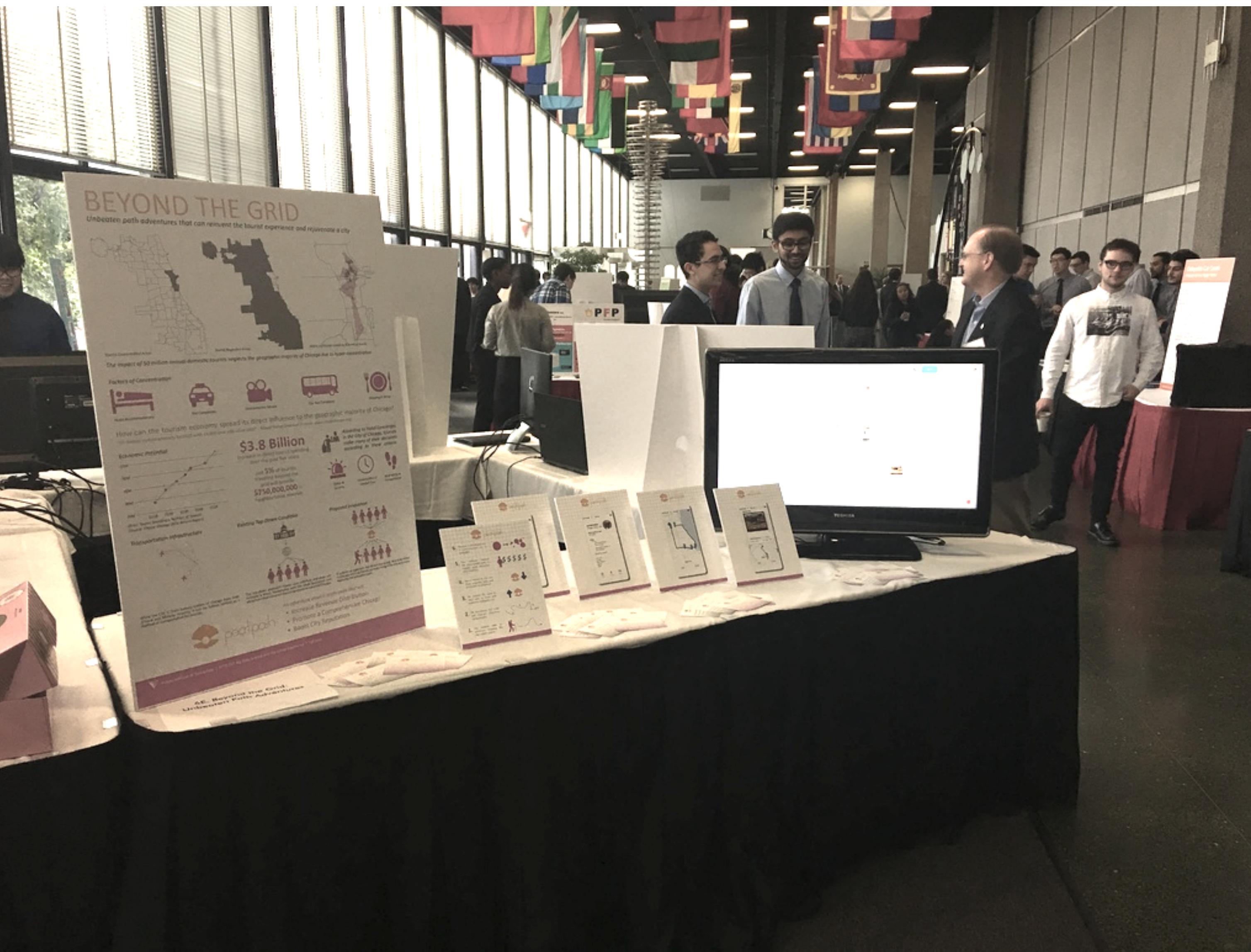
Who? &  
benefits

Why ?

How?



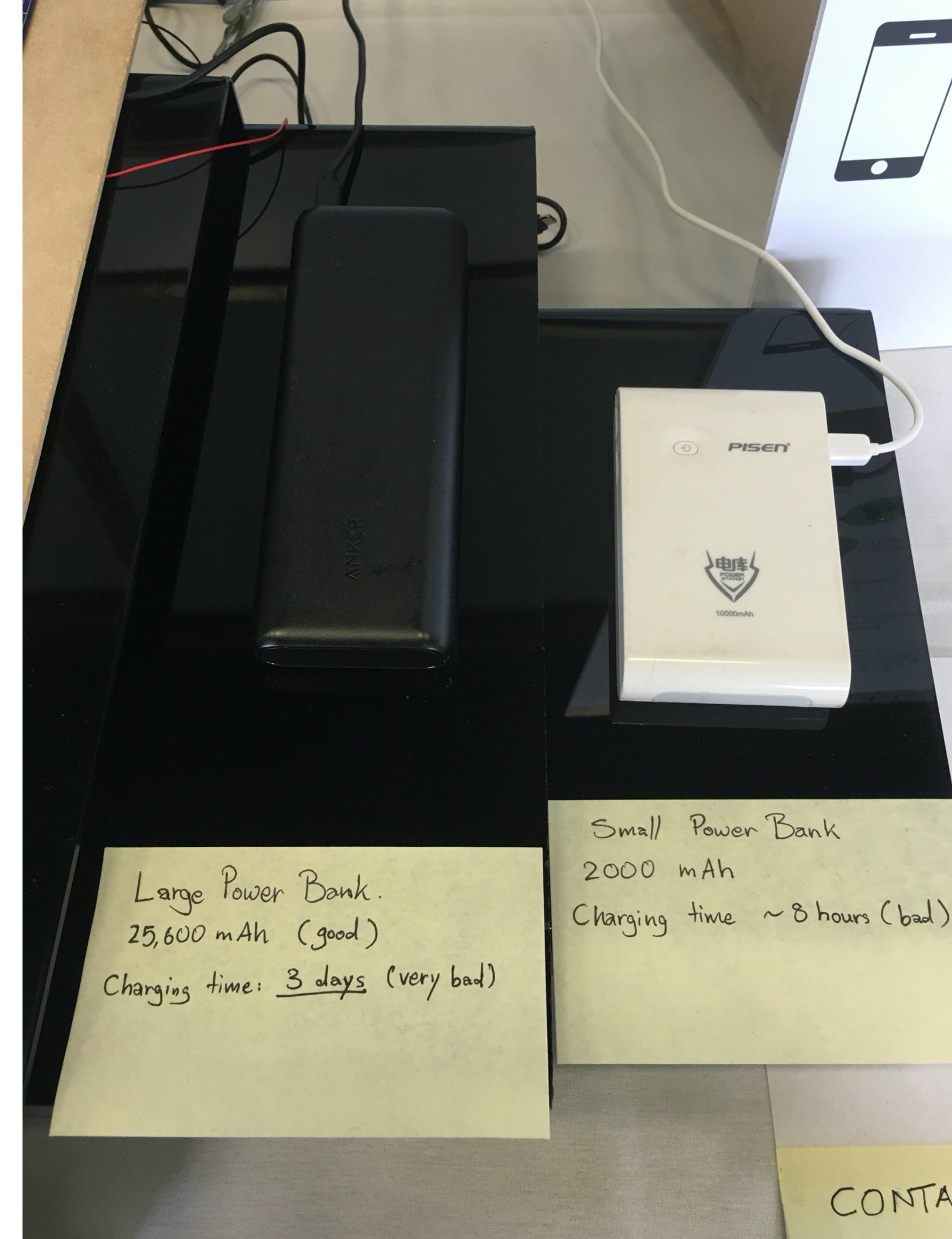
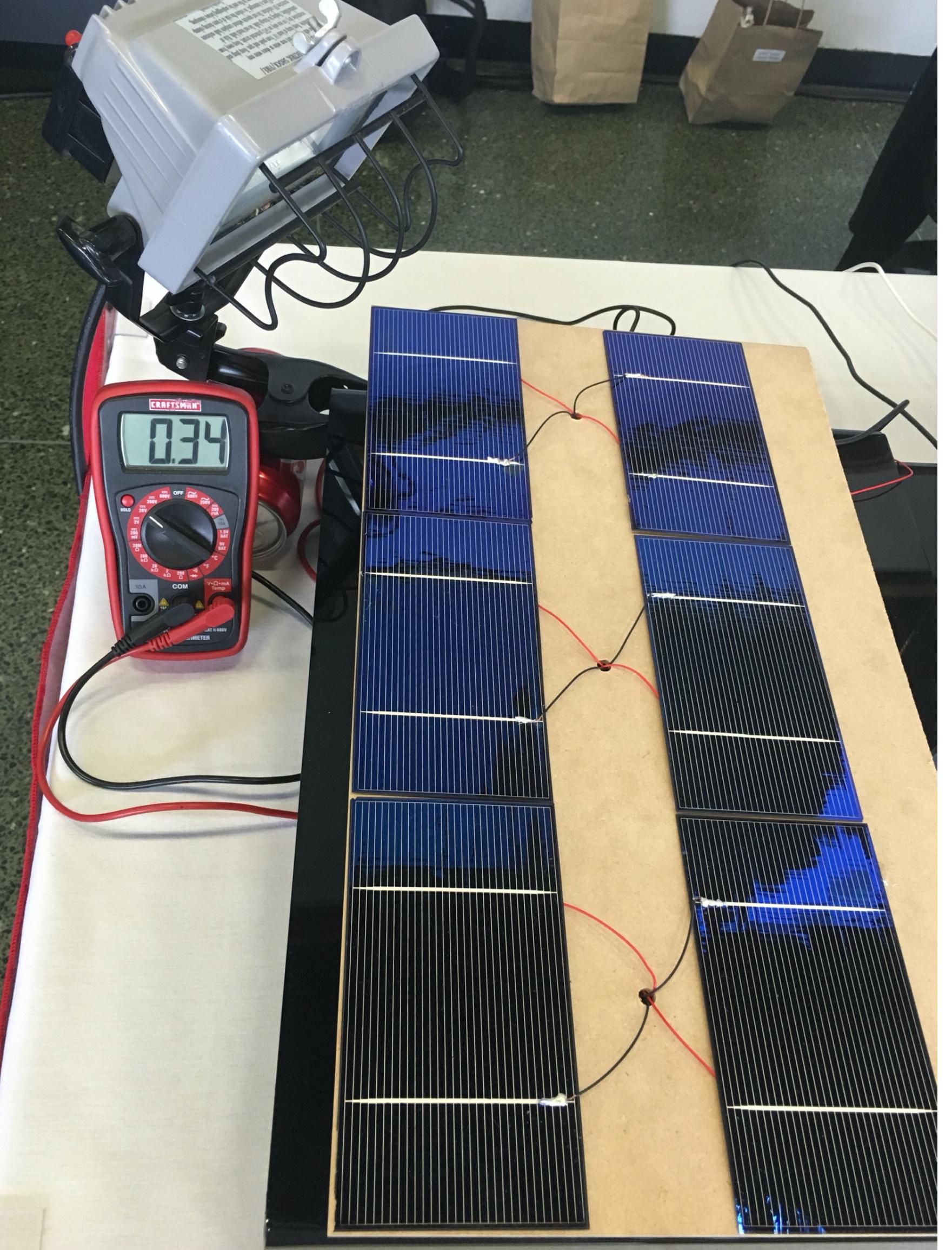
















# Quantitative Structure-Property Relationships

- Today's lecture is intended to help you achieve the following learning objective: Develop a model for a quantitative structure-property relationship. Describe its domain of applicability.
- QSPR Topics
  - Machine learning
  - Features/descriptors
  - Algorithms
  - Validation
  - Limitations

# QSPR is an application of machine learning

- Machine learning [1]
  - “gives computers the ability to learn without being explicitly programmed” - Arthur Samuel (1959)
  - “A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.” -Tom Mitchell (1997)
- In Quantitive Structure-Property Relationships (QSPR)
  - Experience E: chemicals
  - Task T: predicting physical property, e.g.
    - boiling point, chromatography retention times
    - activity against a biological target = QSAR
    - ADMET (absorption, distribution, metabolism, and excretion - toxicity)
  - Performance P: correlation/error in validation set/real-world applications

# Features/Descriptors

- Enables chemicals to be input into machine learning models
- How would you describe a chemical in numbers that can be put into a mathematical formula?
- Features can be
  - from enumerable properties, e.g. number of a certain element, number of H bond donors/acceptors, presence/absence of an element functional group
  - based on 3D structures
    - of a chemical, e.g. surface area, radius of gyration, moments of inertia
    - of a protein-ligand complex
- <http://www.rdkit.org/docs/GettingStartedInPython.html#list-of-available-descriptors>

# Algorithms

- Linear regression
  - $y = mx + b$  for simple linear regression
  - $y = m_1x_1 + m_2x_2 + \dots + b$  for multiple linear regression
  - $x$  is projected onto a new space in partial least squares regression
- Neural networks
  - sets of nodes that transform inputs into an output
  - allows for nonlinear relationships.
- Deep learning
  - neural networks with multiple layers
  - increasingly popular and powerful with faster computers and larger datasets
- Not an exhaustive list!

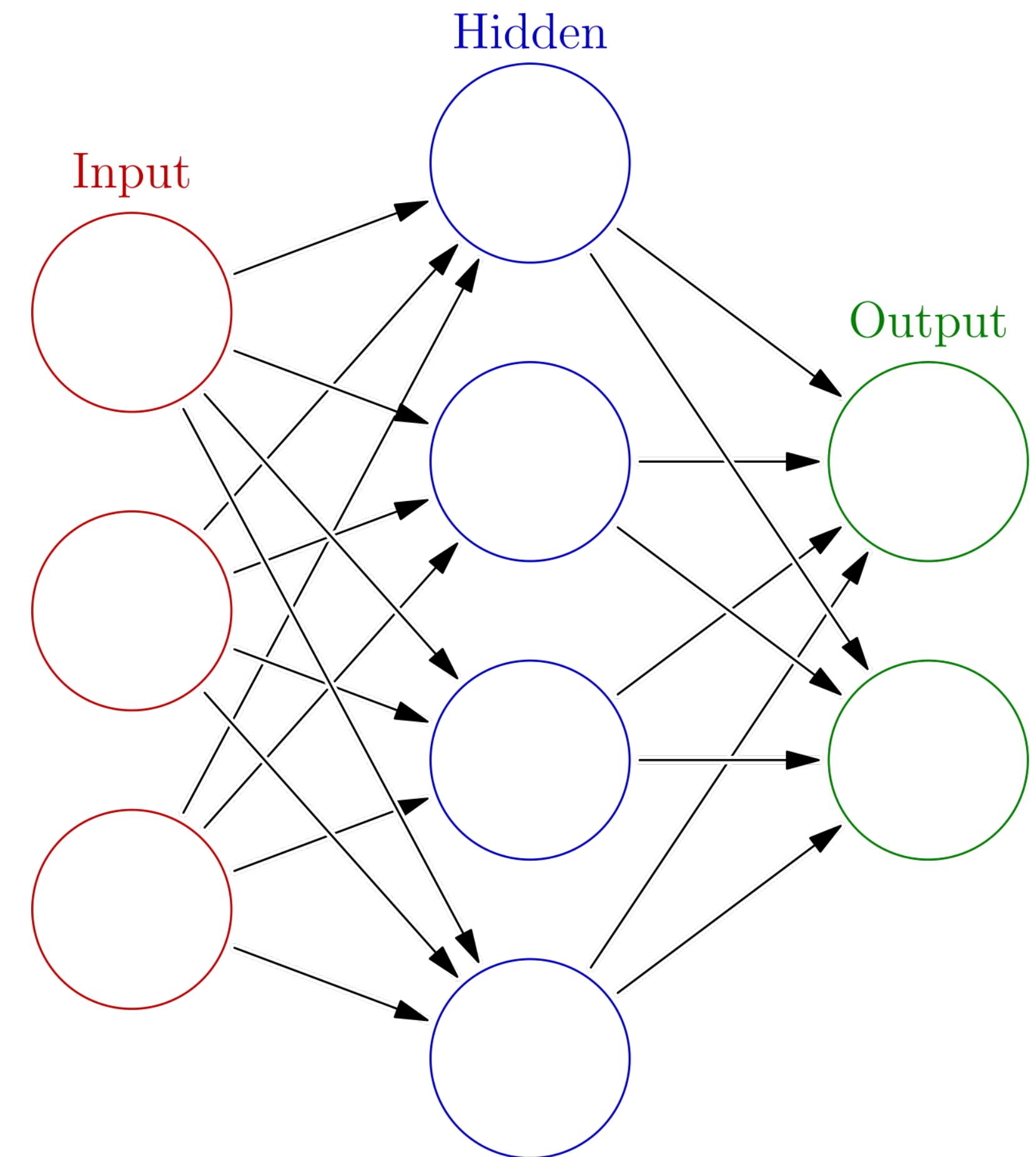


Diagram of an artificial neural network [2]

# Validation

- How do you know if a model is any good?
- Data are split into training and test sets
  - Both sets should represent the same population
  - Model is built using the training set
  - Calculations on test set are compared to data

# Limitations

- What do you think some limitations of QSPPR might be?
- A high-quality training set is necessary
- Models can only be applied to molecules similar to the training set
- Sometimes small changes in structure can lead to large changes in a property
  - adding a methyl can make it so that a ligand no longer fits into a binding pocket
  - in QSAR, known as an “activity cliff”

# References

- [1] <https://www.geeksforgeeks.org/introduction-machine-learning/>
- [2] Downloaded from [https://commons.wikimedia.org/wiki/File:Colored\\_neural\\_network.svg](https://commons.wikimedia.org/wiki/File:Colored_neural_network.svg) and reused under the [CC BY-SA 3.0](#) license

# Solubility Prediction

- The importance of solubility prediction
- The Solubility Challenge (2008)
- QSPR prediction of solubility
- Free energy calculations of solubility

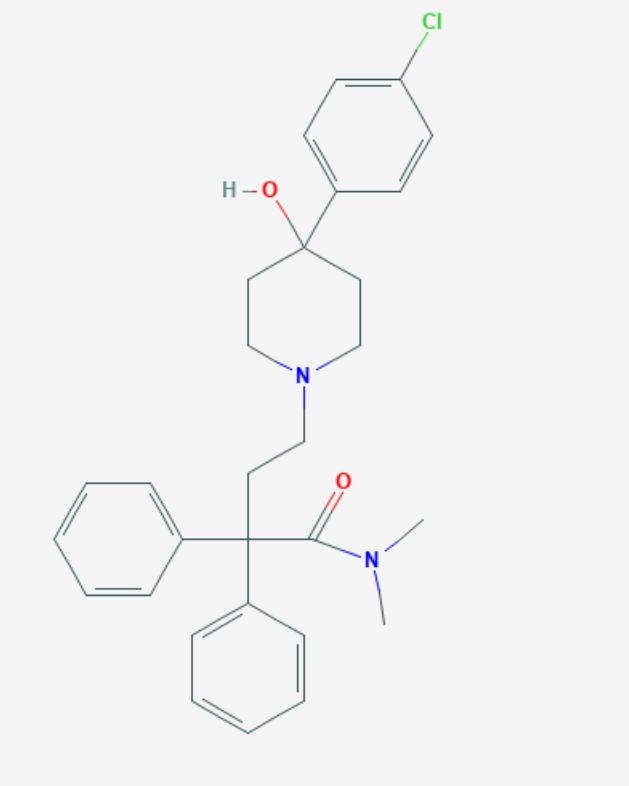
# **The importance of solubility prediction**

# Why predict solubility? [2]

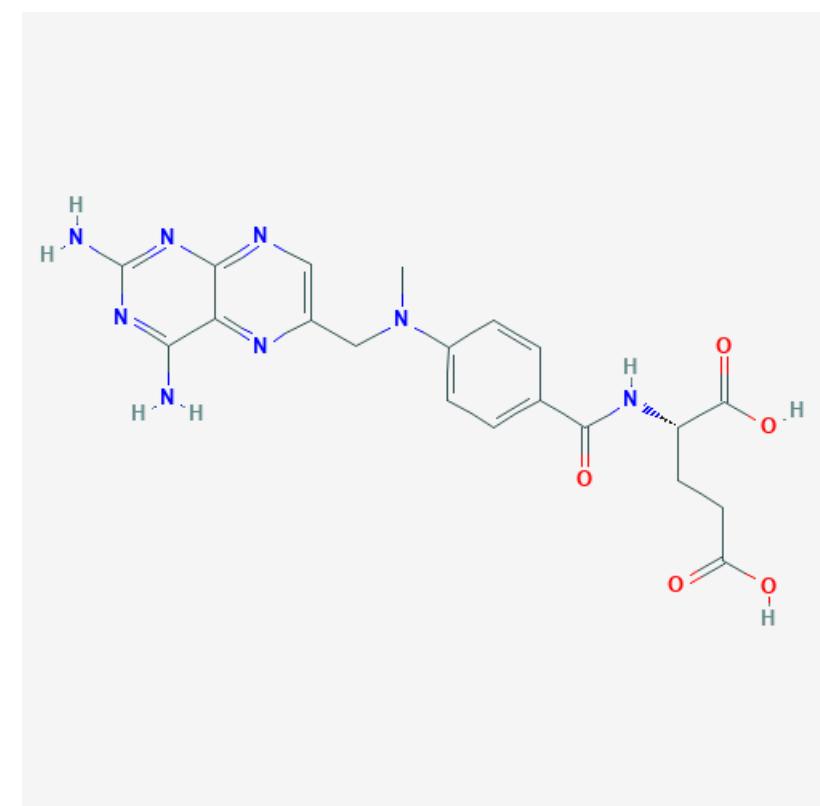
- Drug candidates should have sufficient solubility to be
  - tested in biological assays
  - stored
  - bioavailable
- Rules of thumb are unreliable

# Poor solubility is a major problem [2]

- Pharmacokinetics
  - Up to 40% of drug discovery programs abandoned due to pharmacokinetics - the movement of drugs within the body
  - Low solubility is associated with problems with pharmacokinetics
- Precipitation
  - Loperamide (diarrhea drug) can sometimes precipitate when injected
  - Methotrexate (cancer drug) has kidney toxicity due to precipitation
- Polymorphs can affect solubility and have different bioavailability, e.g. ritonavir (HIV drug) had to be reformulated

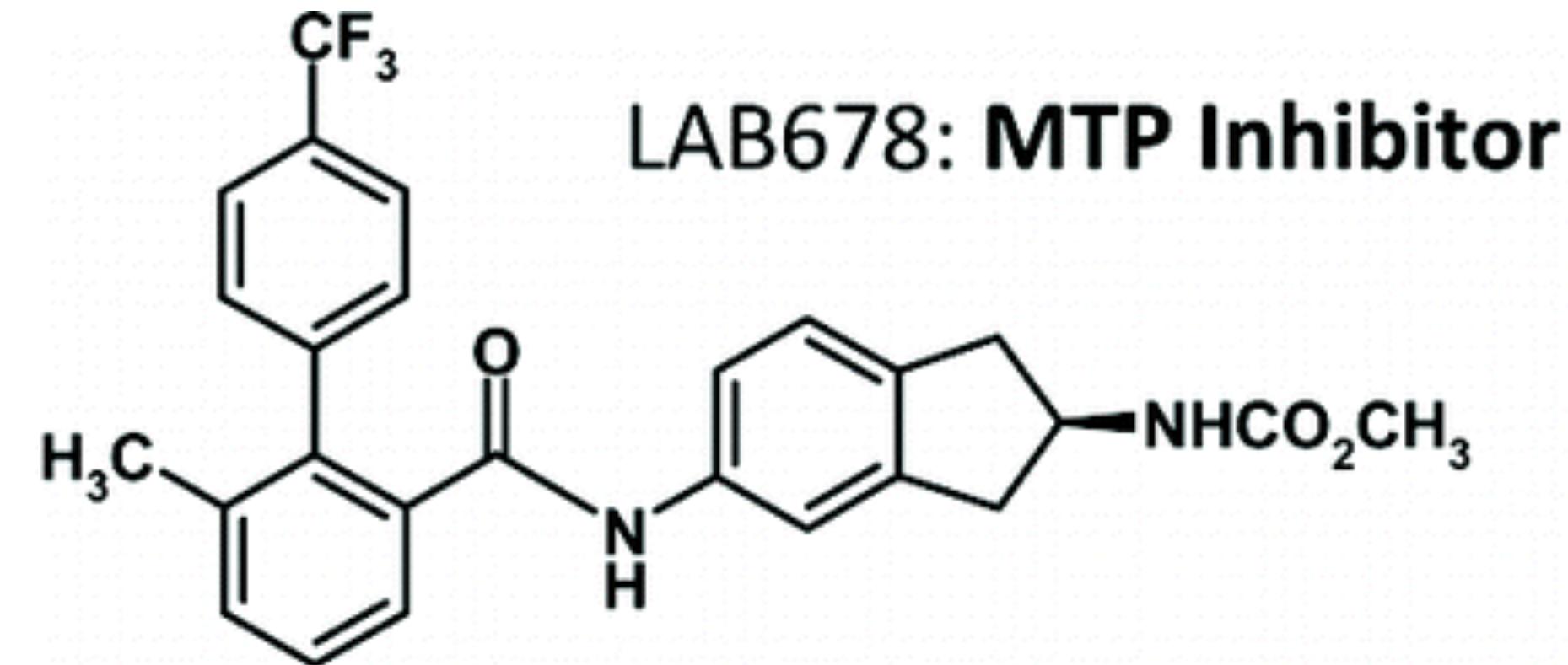


Loperamide

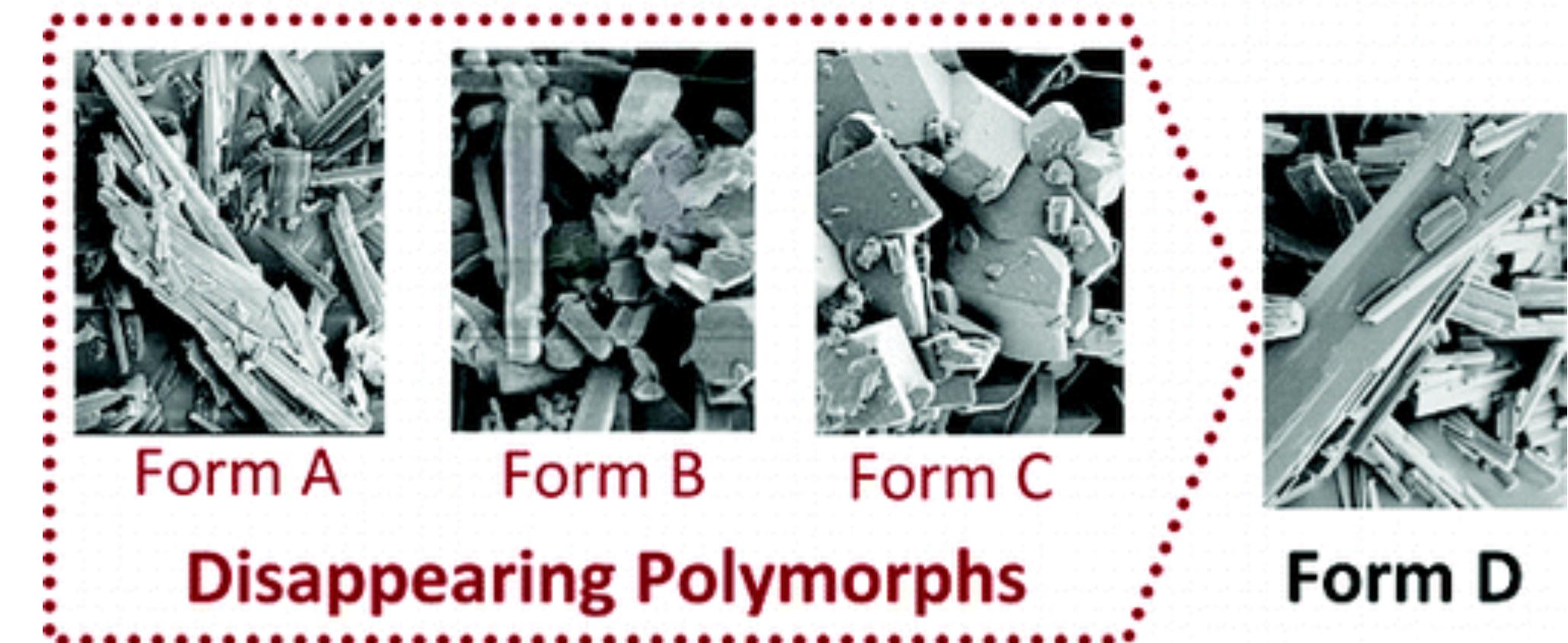


Methotrexate

# Polymorphs are chemical engineering problem



LAB678: MTP Inhibitor



Chemical structure and SEM images from [4]

- LAB687 (lower triglycerides and LDL cholesterol)
- Early on, observed form A and B
- Later, found form C
- Form D observed in scale up, and began to predominate!

# Why work with low-solubility compounds?

- Practical reason
  - increased weight and lipophilicity of test compounds from high-throughput synthesis and screening
- Principled reasons
  - drugs usually need to permeate cell membranes
  - increased binding affinity - the molecule prefers the complex over the solvent
    - often associated with reduced solubility

# How are solubility problems hidden?

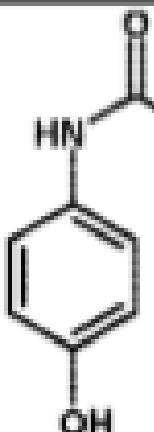
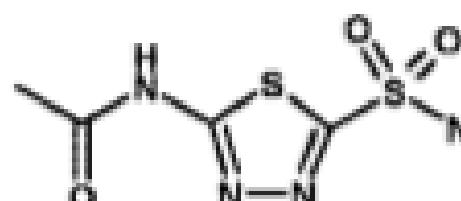
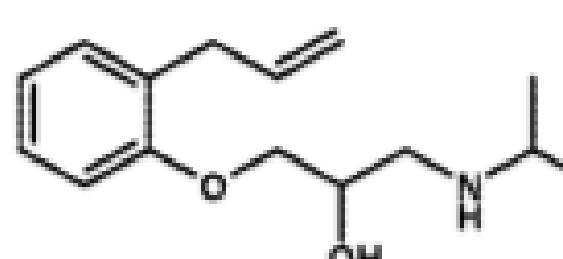
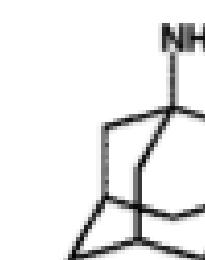
- use of solvents, e.g. DMSO, in early-stage assays
- pH dependence
  - free acid/free base is less soluble than salt
  - intrinsic solubility is solubility of free acid/free base
- kinetic solubility - appearance or disappearance of precipitate at different concentrations than the thermodynamic saturating concentration (sub- or supersaturation)
- polymorphs not always known
- These problems are also reasons why solubility is difficult to measure and predict

# **The Solubility Challenge**

# Solubility Challenge (2008) [2, 3]

- measured
  - intrinsic solubility (free acid/free base)
  - highly reproducible (thermodynamic versus kinetic solubility)
- published 100 solubilities
- challenged researchers to predict 32 more
- received >100 entries between July and September 2008

# Examples from the dataset [2]

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

# Overall performance in the solubility challenge

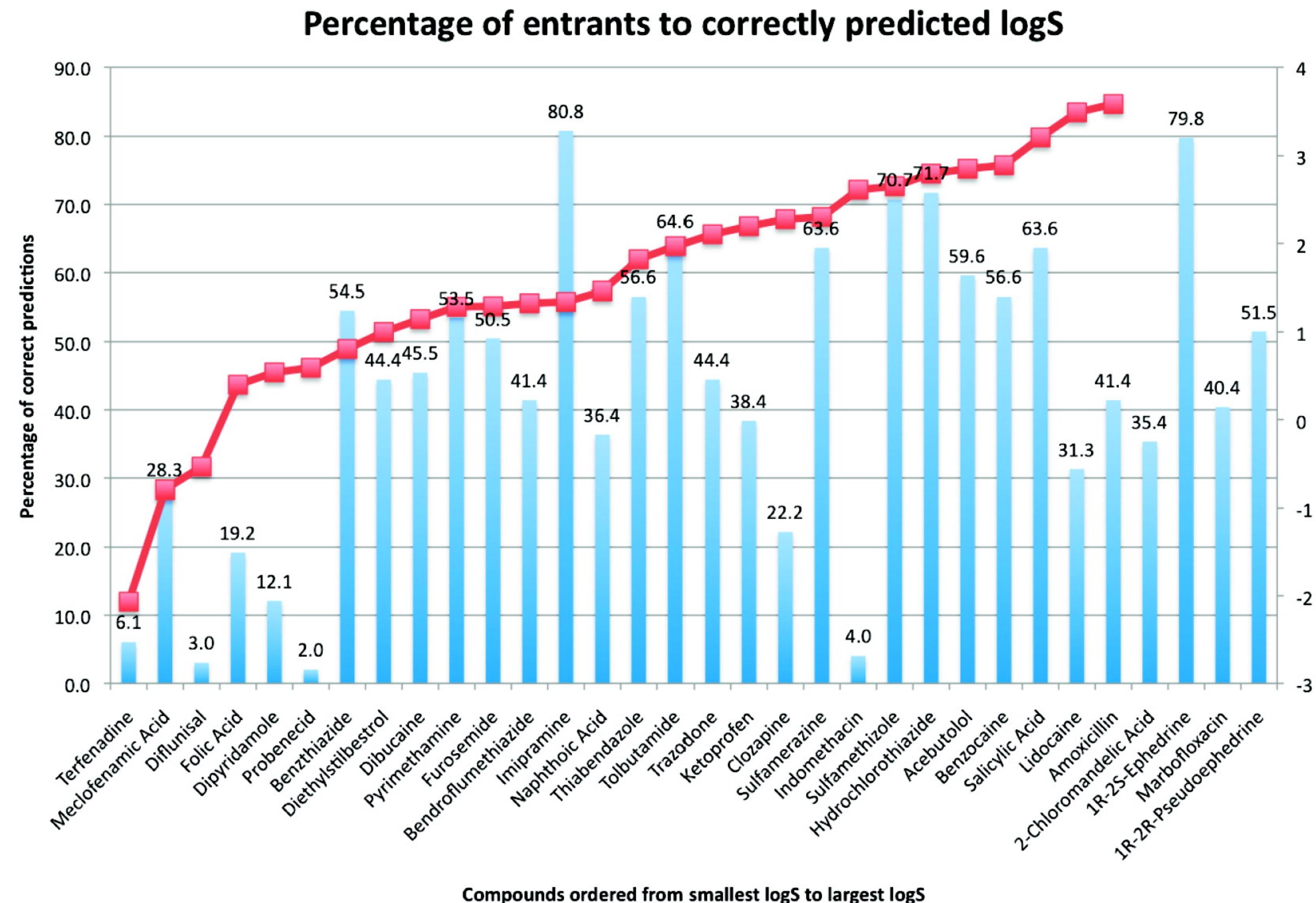


Figure 2 from ref [3]  
“Correct” is within 0.5 logS  
Blue bars are percentage correct  
Red is actual logS

# Stated major findings [3]

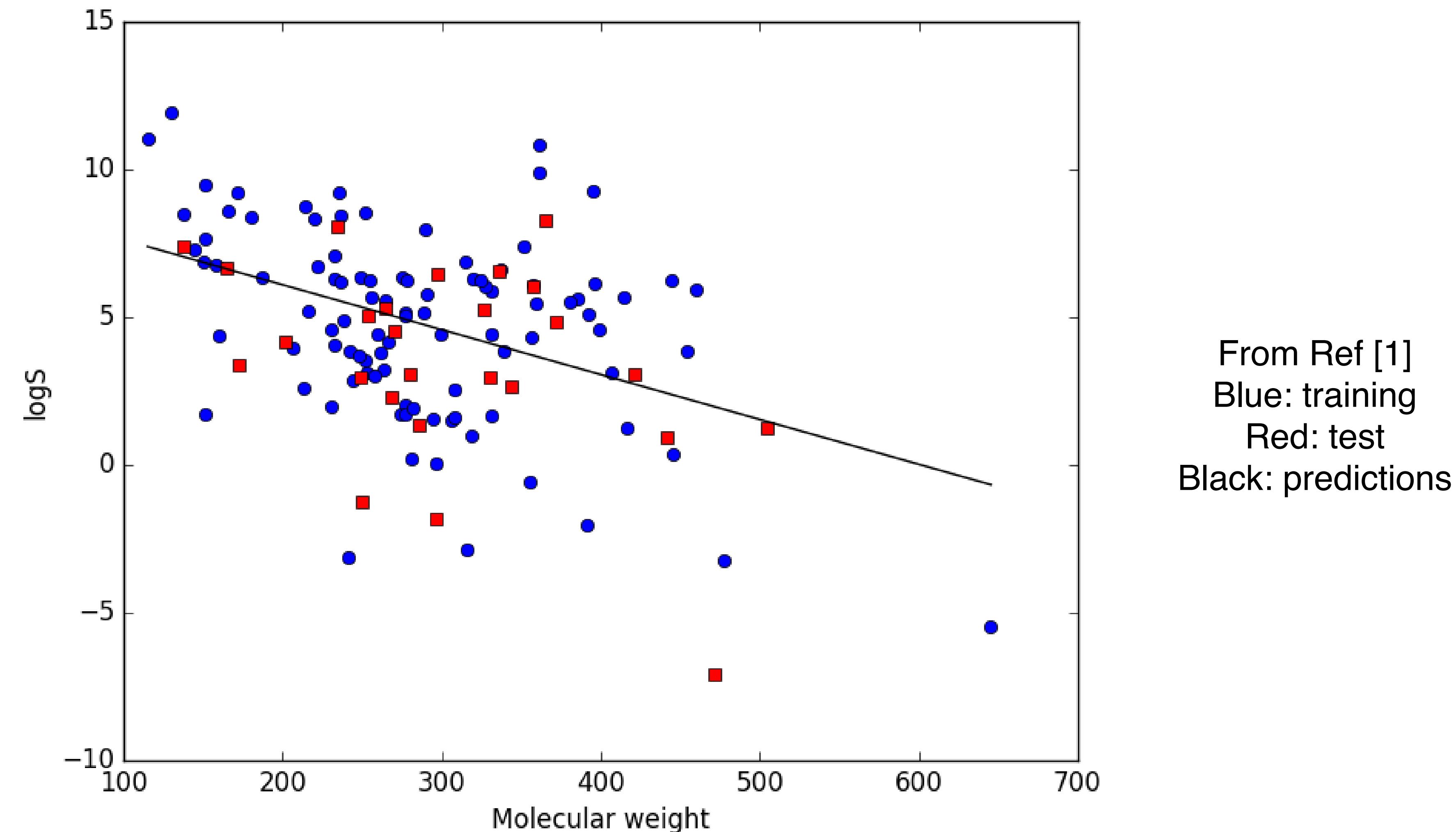
- For measured set of 28 compounds, percent correct on
  - absolute S ( $\pm 10\%$ ) ranges from 0%-18%
  - logS ( $\pm 0.5$  logS) ranges from 10-61%
  - Worst two compounds 2% and 4% correct by logS
- $R^2$  ranges from 0 to 0.65
- No polymorph predictions
- Compounds with “normal” solubility ( $\log S = 0.5$  to 3) were easiest
- Many methods, but none clearly the best

# **QSPR prediction of solubility**

# Historically informative descriptors [5]

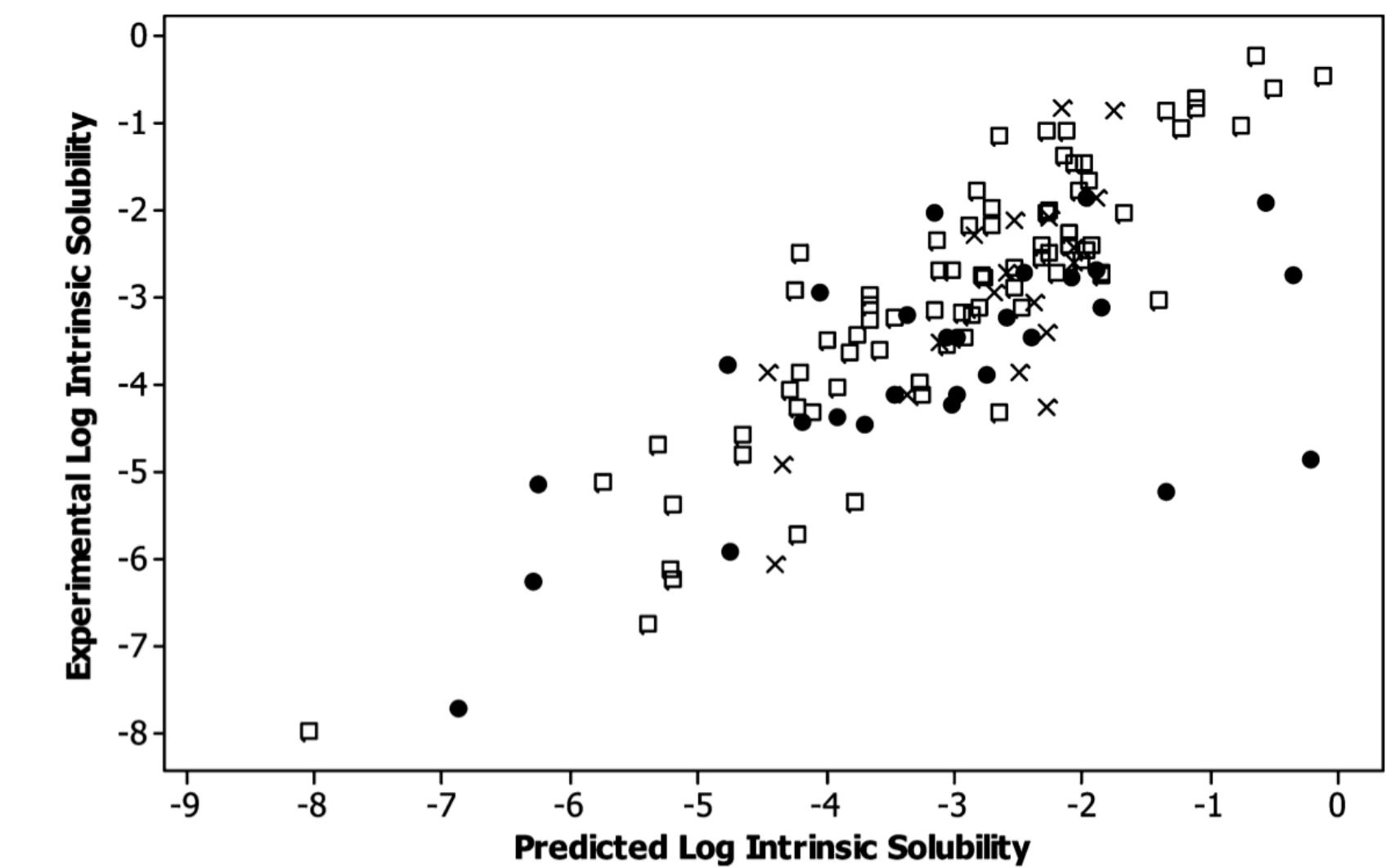
- Molecular size or surface area - cost of forming cavity in water
- Octanol:water partition coefficient ( $P$ ) - solubility relative to a non polar phase
- Melting point - higher means more stable solid
- Hydrogen bonding - favors water solubility
- Atom/group contributions
- Molecular connectivities

# Molecular weight has some predictive power



# QSPR in the solubility challenge

- Hewitt et al tried four QSPR models [5]
  - Multiple linear regression (426 descriptors, genetic algorithm, no more than 5 used at once)
  - Artificial neural network
  - Category-specific models based on H bond ability
  - Various commercial QSPR models
  - Consensus, mean of four models
- The best was a three-descriptor linear regression
  - based on log P, boiling point, and R maximal autocorrelation of lag 2 (related to size and connectivity).
- $R^2_{train} = 0.74$ ,  $R^2_{test} = 0.51$
- “none of the other modeling approaches used in this study was able to improve upon the predictions made by the MLR model”



**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.

# Can deep learning do better? [6]

**Table 5**

Prediction performance and standard deviations using 10- fold cross validation on the Solubility Challenge Dataset (125 molecules)

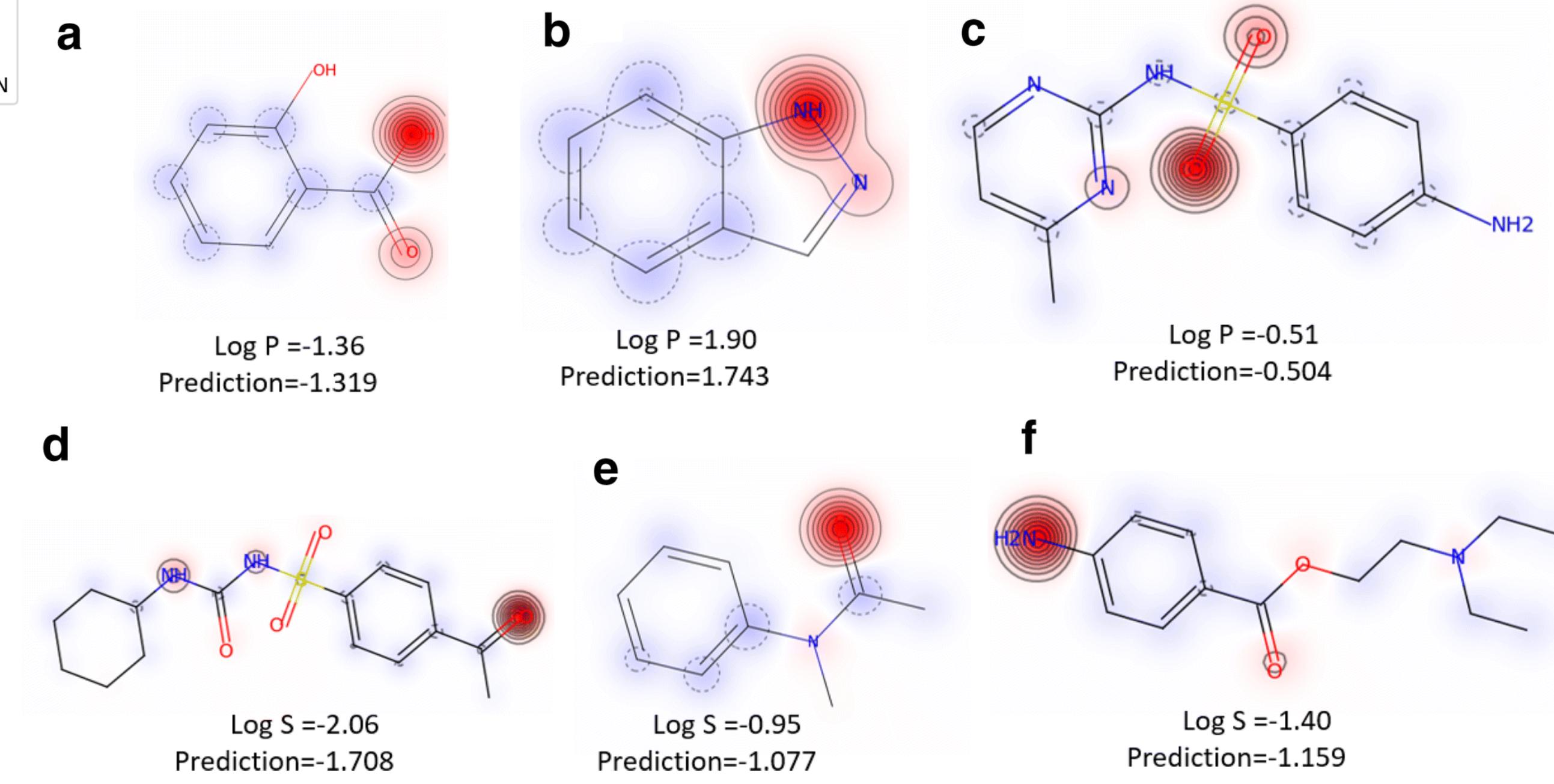
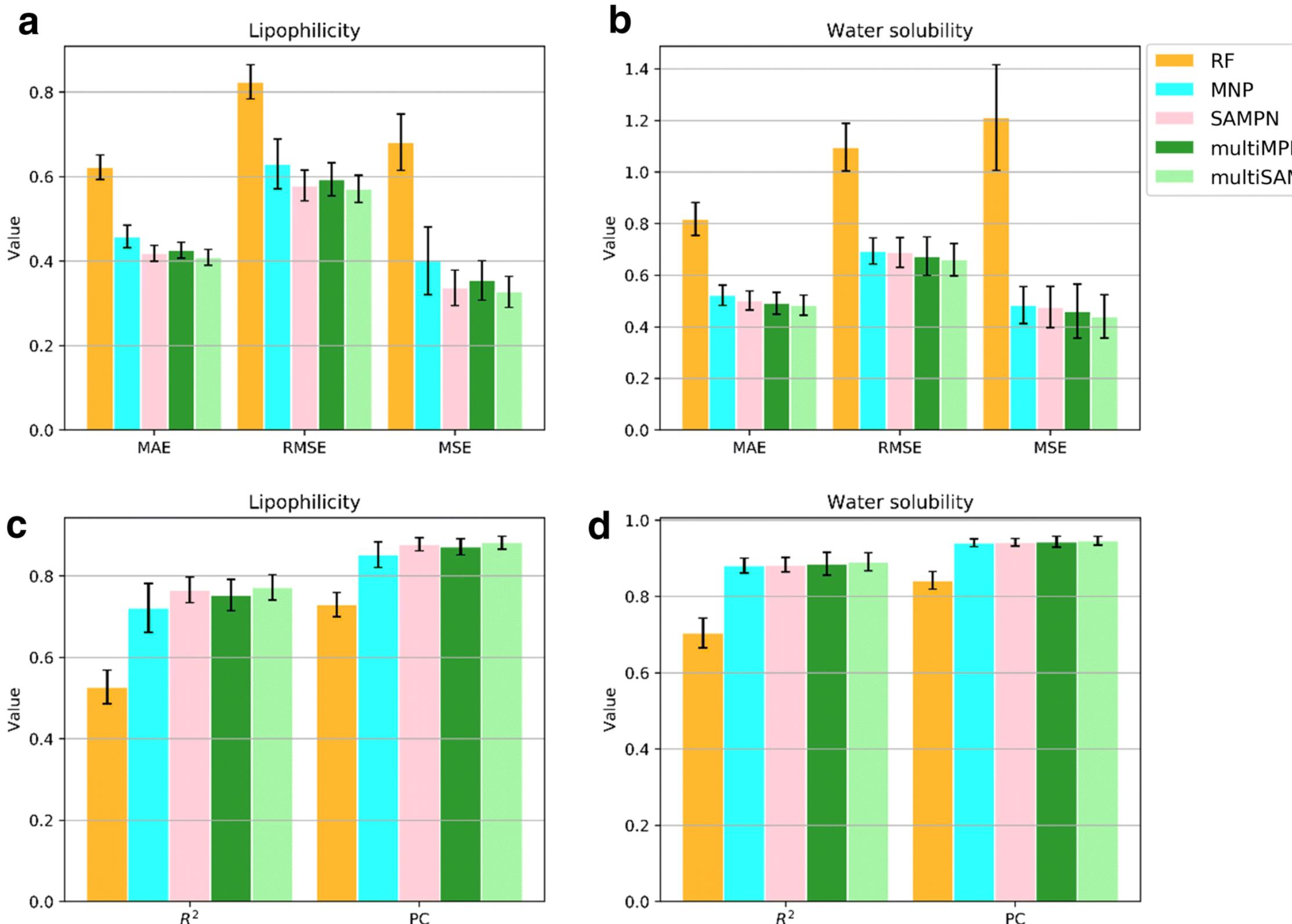
Models	R <sup>2</sup>	std R <sup>2</sup>	RMSE	std RMSE	AAE	std AAE
UG-RNN	0.32	0.03	1.41	0.12	1.08	0.10
UG-RNN-LogP	0.45	0.04	1.27	0.13	1.03	0.11
UG-RNN-CR-LogP	0.44	0.09	1.28	0.18	1.03	0.16
UG-RNN-Huusk	0.43	0.02	1.16	0.03	0.93	0.03
UG-RNN-Huusk-Sub	0.48	0.02	1.11	0.03	0.84	0.01
UG-RNN-LogP-Huusk	0.54	0.02	1.00	0.03	0.82	0.03
UG-RNN-LogP-Huusk-Sub	0.60	0.02	0.94	0.02	0.71	0.02
UG-RNN-CR-LogP-Huusk	0.62	0.03	0.96	0.06	0.83	0.06
UG-RNN-CR-LogP-Huusk-Sub	0.67	0.03	0.90	0.06	0.74	0.05
NN-Sol-Chal <sup>11</sup>	0.40	-	1.51	-	-	-
MLR-Sol-Chal <sup>11</sup>	0.51	-	0.95	-	0.77	-
New <i>in silico</i> consesus <sup>11</sup>	0.60	-	0.90	-	0.68	-

Better than simple neural network, but worse than MLR

Huusk results are based on more data.  
Sub results use different solubility values.

In summary, deep learning can do better if there is more data.

# Opening the black box?

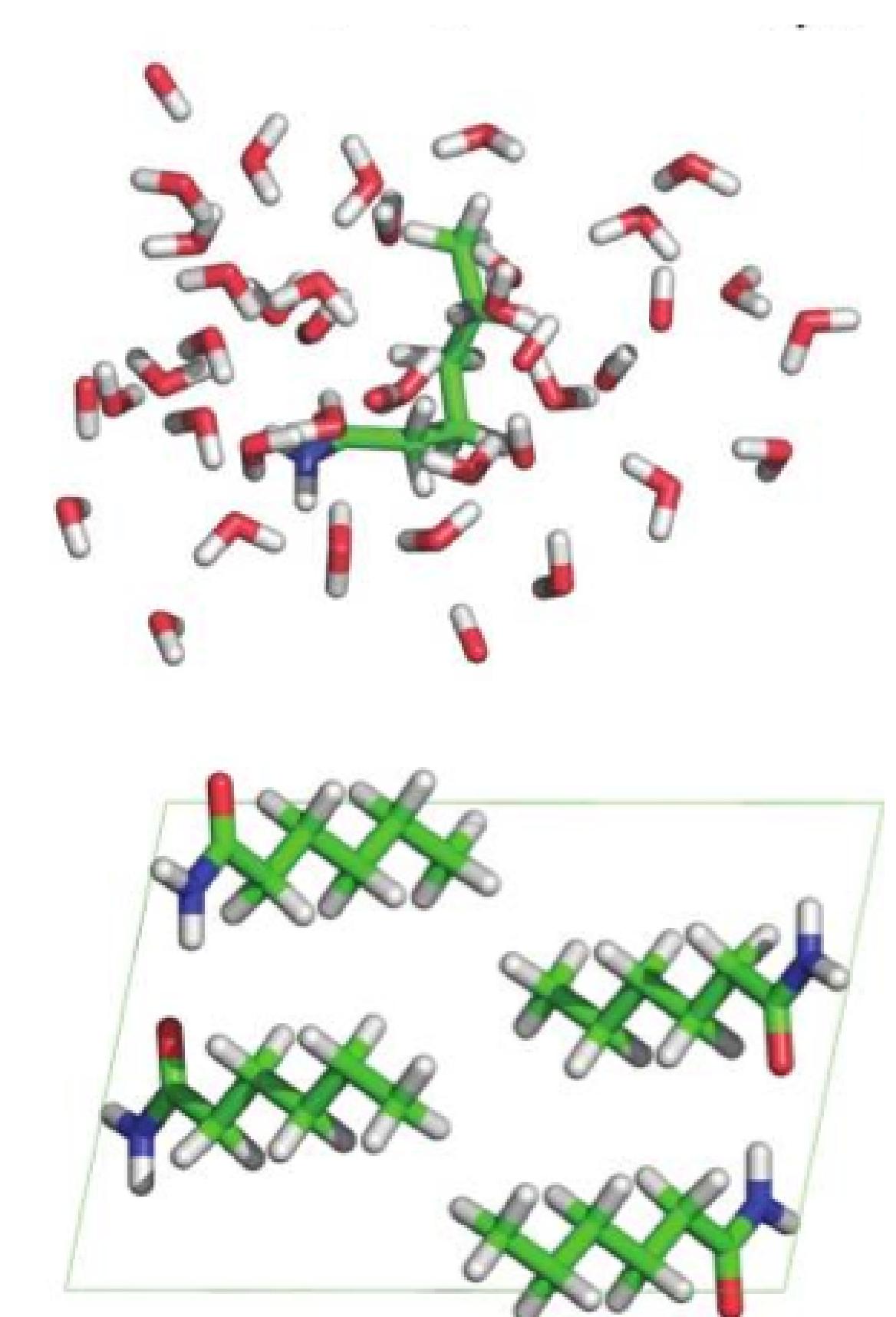
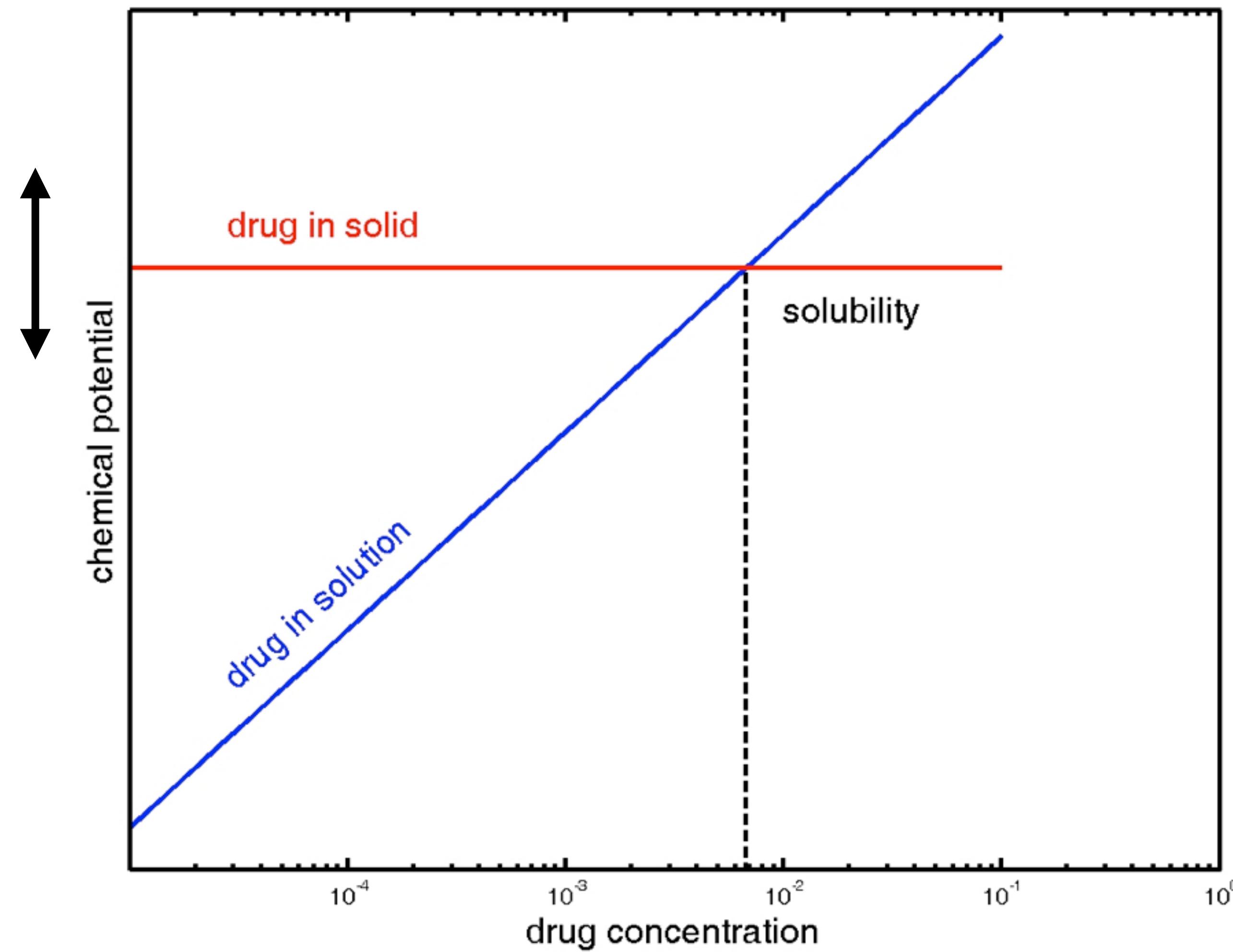


Recent deep learning method seems to perform well in solubility prediction and also highlights moieties that contribute to (red) or detract from (blue) solubility. See Ref. [7].

# **Free energy calculations of solubility**

# Chemical potential determines solubility

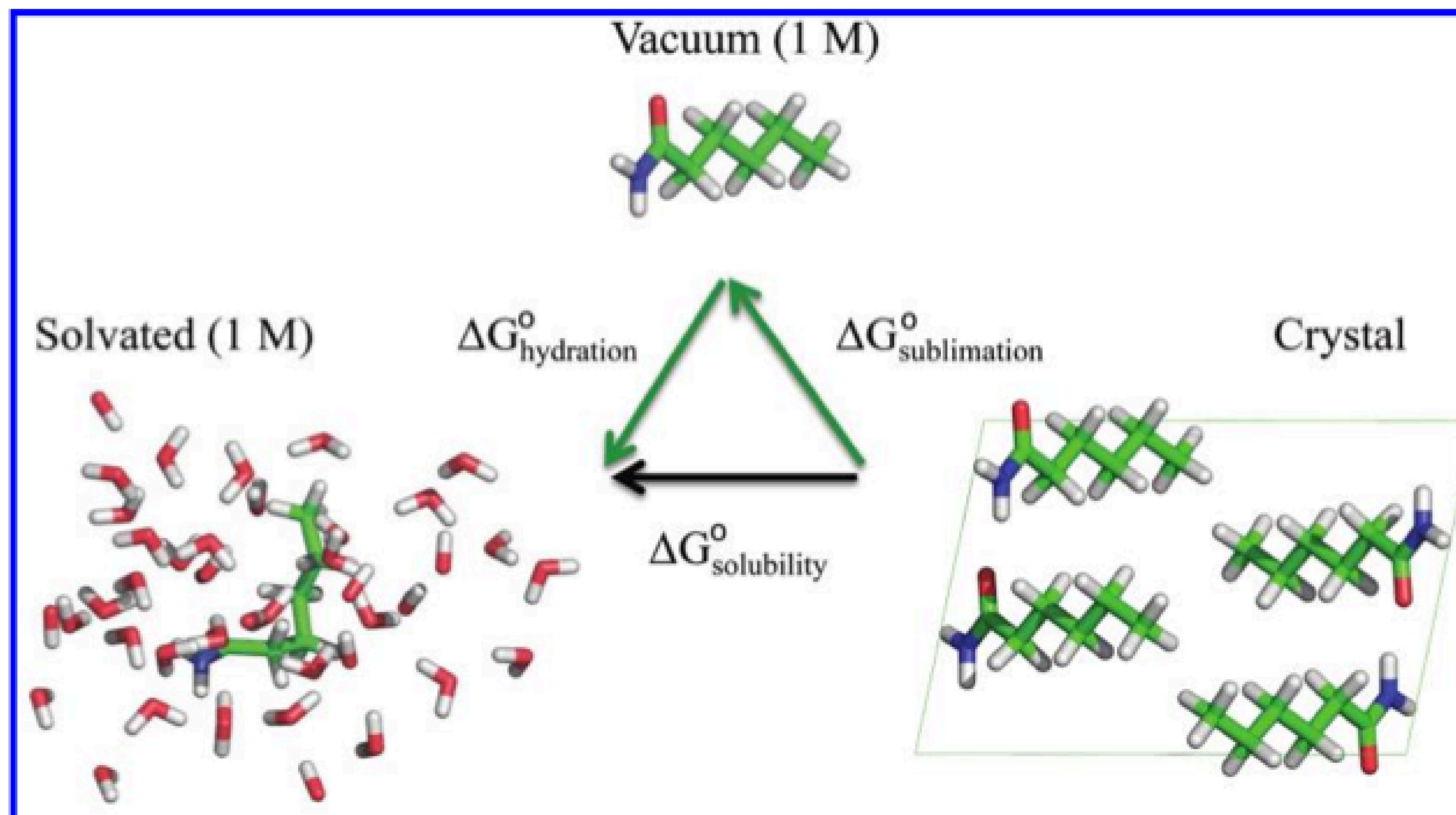
Polymorphs have  
different chemical potential



From Ref [8]

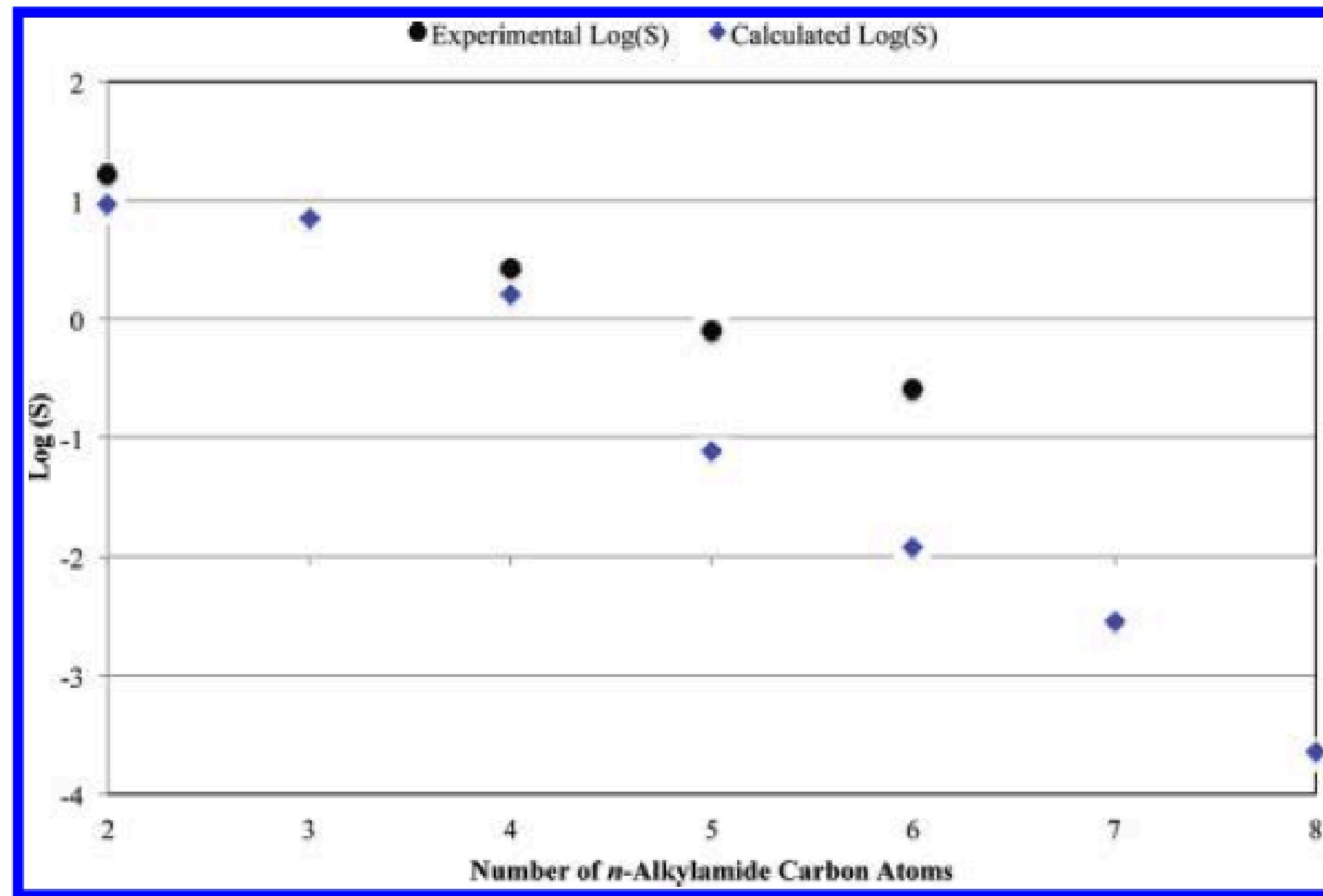
From Ref [1]

# A thermodynamic cycle for solubility



From Ref [8]. Note that solubility  $\neq$  solvation or hydration.  
Hydration + sublimation free energies calculated with alchemical free energy methods.

# Results on an initial series appear promising without any 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<sub>2</sub> group due to increasingly favorable deposition and to a lesser extent from unfavorable solvation.

From Ref [8]

# References

- [1] Many parts of today's lecture were adapted from a lecture by David Mobley ([https://github.com/MobleyLab/drug-computing/tree/master/uci-pharmsci/lectures/free\\_energy\\_basics](https://github.com/MobleyLab/drug-computing/tree/master/uci-pharmsci/lectures/free_energy_basics)) under the [CC BY 4.0 license](#). The lecture is part of the Drug Discovery Computing Techniques course (PharmSci 175/275) at UC Irvine.
- [2] Llinàs, A.; Glen, R. C.; Goodman, J. M. Solubility Challenge: Can You Predict Solubilities of 32 Molecules Using a Database of 100 Reliable Measurements? *J. Chem. Inf. Model.* 2008, 48 (7), 1289–1303. <https://doi.org/10.1021/ci800058v>.
- [3] Hopfinger, A. J.; Esposito, E. X.; Llinàs, A.; Glen, R. C.; Goodman, J. M. Findings of the Challenge To Predict Aqueous Solubility. *J. Chem. Inf. Model.* 2009, 49 (1), 1–5. <https://doi.org/10.1021/ci800436c>.
- [4] Prashad, M.; Sutton, P.; Wu, R.; Hu, B.; Vivelo, J.; Carosi, J.; Kapa, P.; Liang, J. Process Research and Development of a MTP Inhibitor: Another Case of Disappearing Polymorphs upon Scale-Up. *Org. Process Res. Dev.* 2010, 14 (4), 878–882. <https://doi.org/10.1021/op100115u>.

# References

- [5] Hewitt, M.; Cronin, M. T. D.; Enoch, S. J.; Madden, J. C.; Roberts, D. W.; Dearden, J. C. In Silico Prediction of Aqueous Solubility: The Solubility Challenge. *J. Chem. Inf. Model.* 2009, 49 (11), 2572–2587. <https://doi.org/10.1021/ci900286s>.
- [6] Lusci, A.; Pollastri, G.; Baldi, P. Deep Architectures and Deep Learning in Chemoinformatics: The Prediction of Aqueous Solubility for Drug-Like Molecules. *J. Chem. Inf. Model.* 2013, 53 (7), 1563–1575. <https://doi.org/10.1021/ci400187y>.
- [7] Tang, B.; Kramer, S. T.; Fang, M.; Qiu, Y.; Wu, Z.; Xu, D. A Self-Attention Based Message Passing Neural Network for Predicting Molecular Lipophilicity and Aqueous Solubility. *J. Cheminform.* 2020, 12 (1), 15. <https://doi.org/10.1186/s13321-020-0414-z>.  
Figures adapted under the [CC BY 4.0](#) license.
- [8] Schnieders, M. J.; Baltrusaitis, J.; Shi, Y.; Chattree, G.; Zheng, L.; Yang, W.; Ren, P. The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force Field. *J. Chem. Theory Comput.* 2012, 8 (5), 1721–1736. <https://doi.org/10.1021/ct300035u>.