Molecular Engineering of Solar-Powered Windows

Jacqueline Cole

Cavendish Laboratory, University of Cambridge, UK

ISIS Neutron and Muon Source, Rutherford Appleton Laboratory, UK

Argonne National Laboratory, USA

Molecular Engineering of Solar-powered Windows

Jacqueline M. Cole

Cavendish Laboratory, University of Cambridge, UK
ISIS Neutron Facility, Rutherford Appleton Laboratory, UK
Argonne National Laboratory, USA

Synopsis

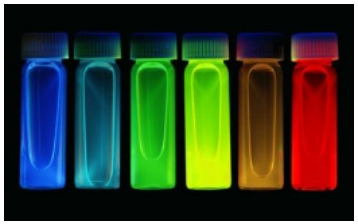
- Scientific background and motivation
- Materials discovery using existing databases: pre-ADSP
- Auto-generating Custom Databases: ChemDataExtractor
- Materials discovery "Take #2" – The ADSP project
- Next Steps

Background and Motivation

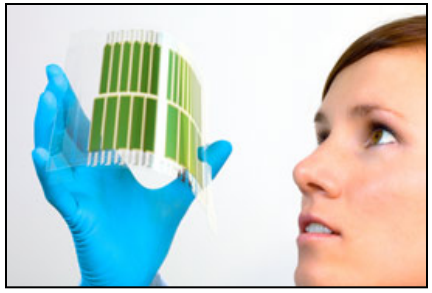
Types of Solar Cells

Silicon

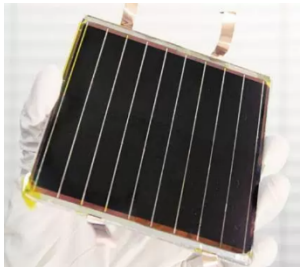
Thin-film Si



Quantum Dot



Organic PV



Perovskite PV

Dye-Sensitized
Solar Cell (DSC)

Solar-powered Windows

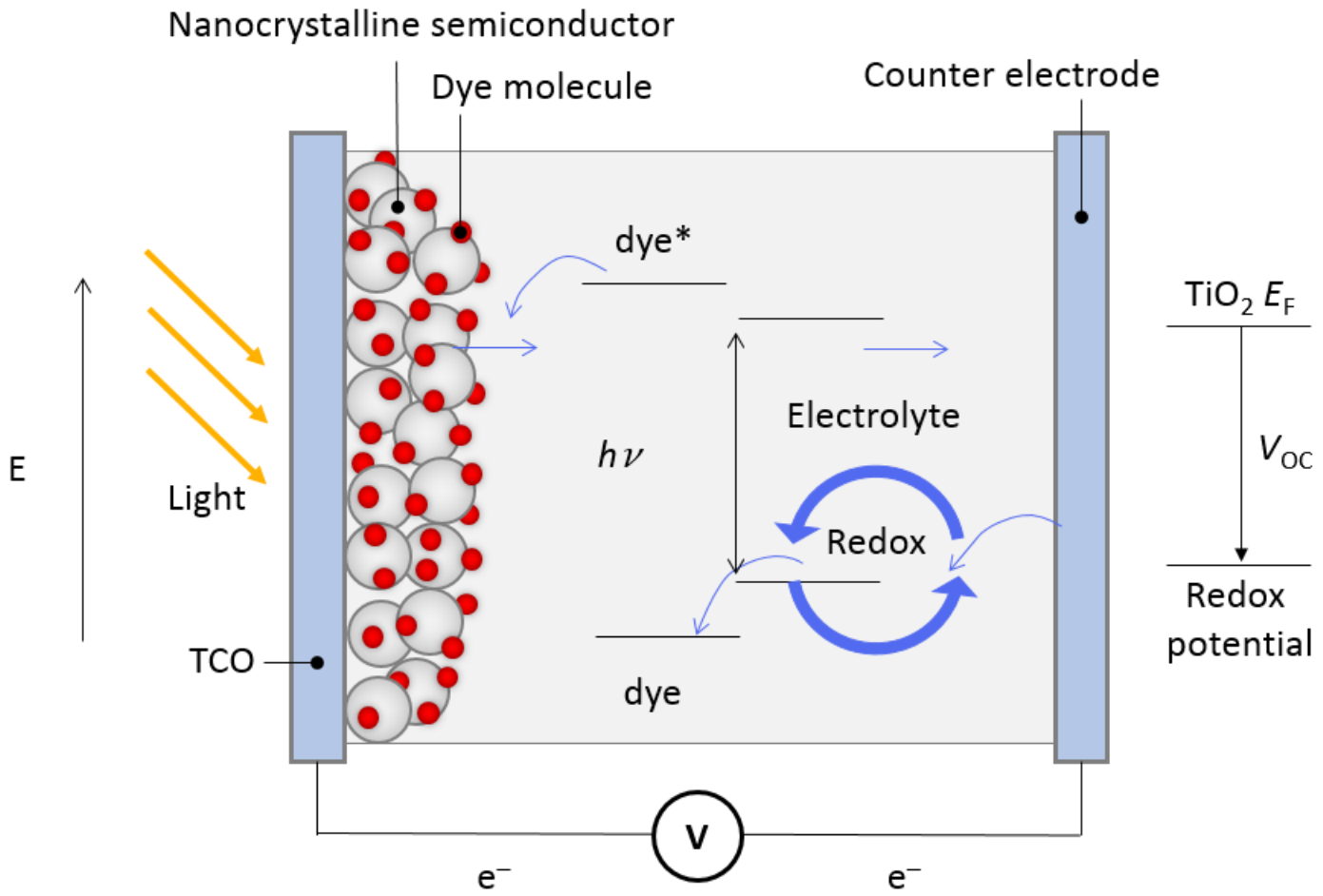
40% Total energy drain in the USA comes from buildings

DSCs in real-world action …

Graz Science Tower (Sep 17)

Swiss Tech Convention Centre

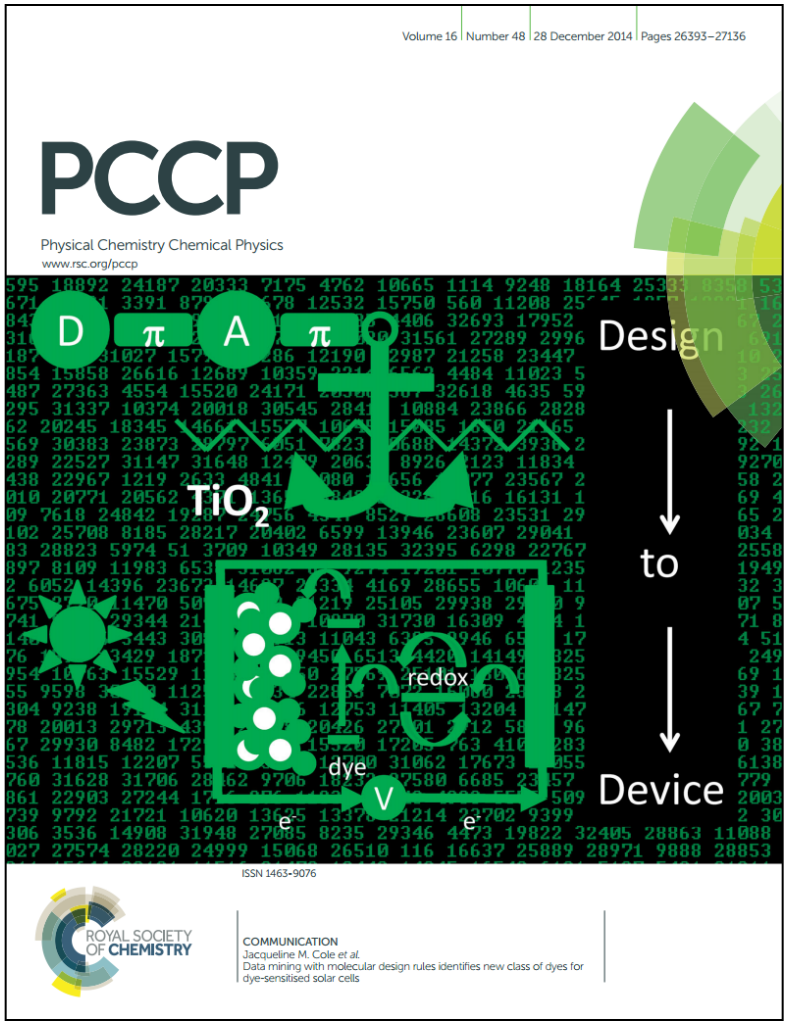
Dye-sensitized solar cell

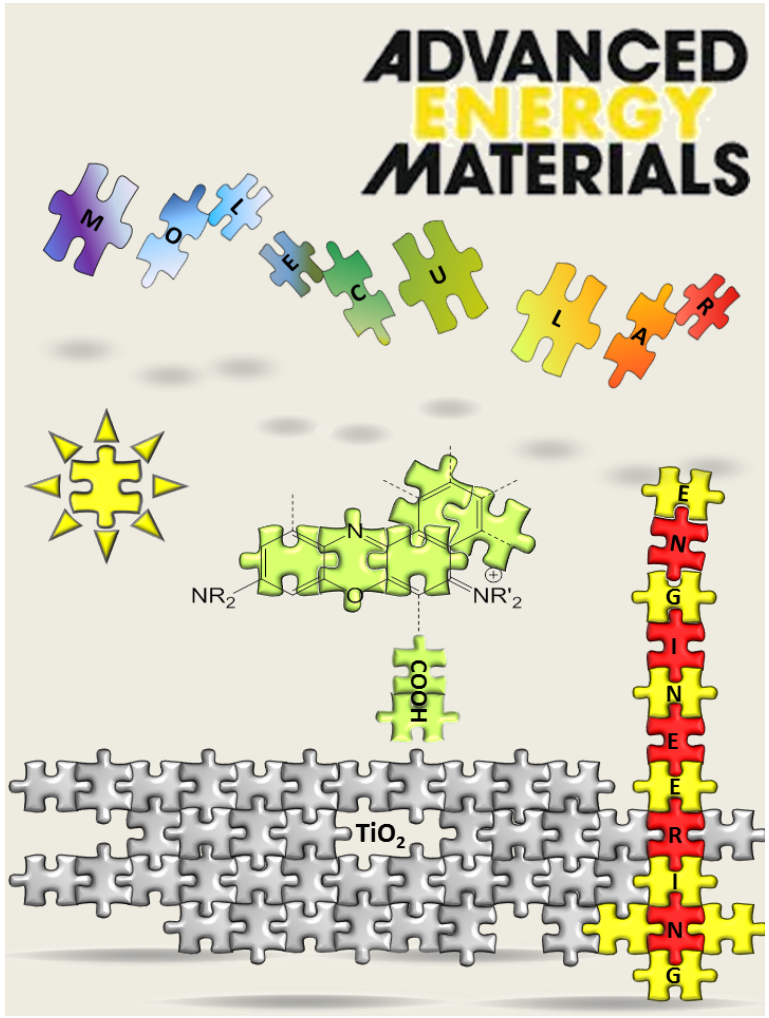


Economics: Price-to-Performance Ratio (and color options)

Materials Discovery pre-ADSP

Materials Discovery of DSC dyes

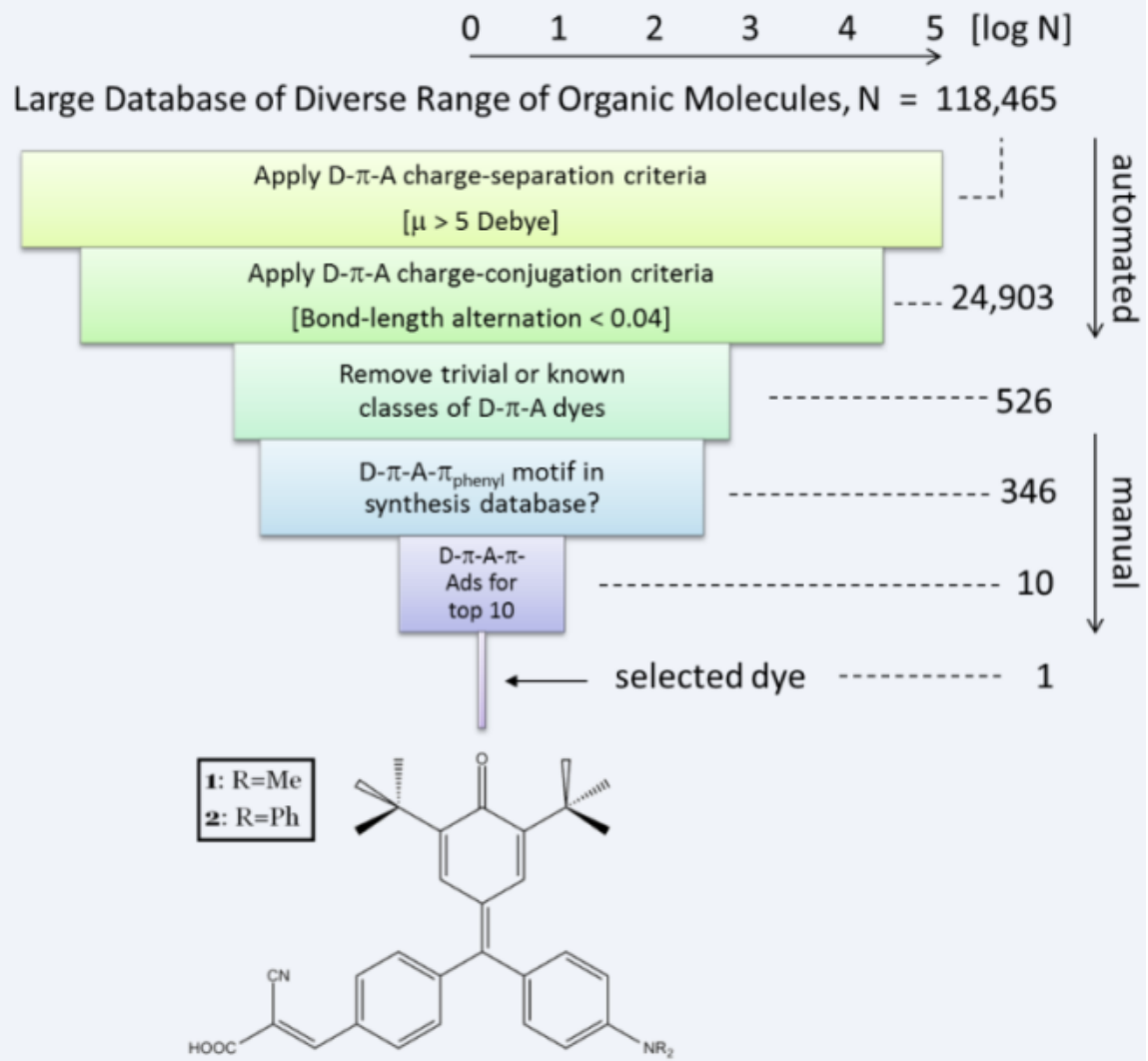




'Top-down' approach: Cole et al, PCCP, 16 (2014) 26684

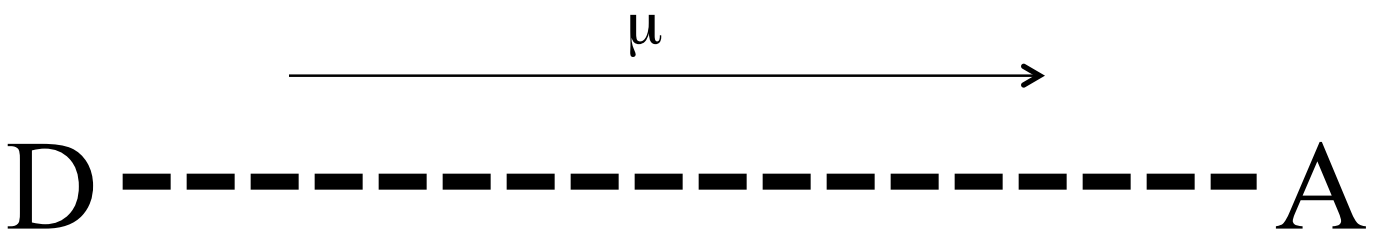
'Bottom-up' approach: Cole et al, AEM 5 (2015) 1401728

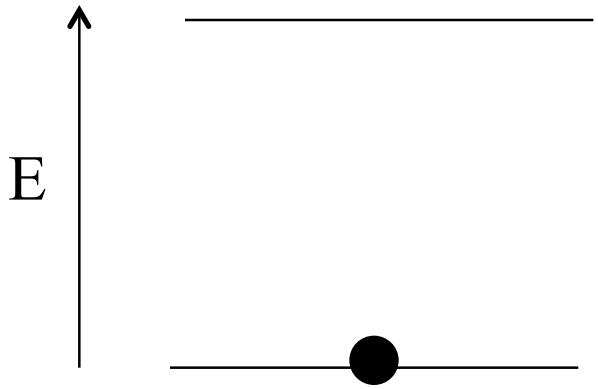
Overarching Data-Mining Strategy



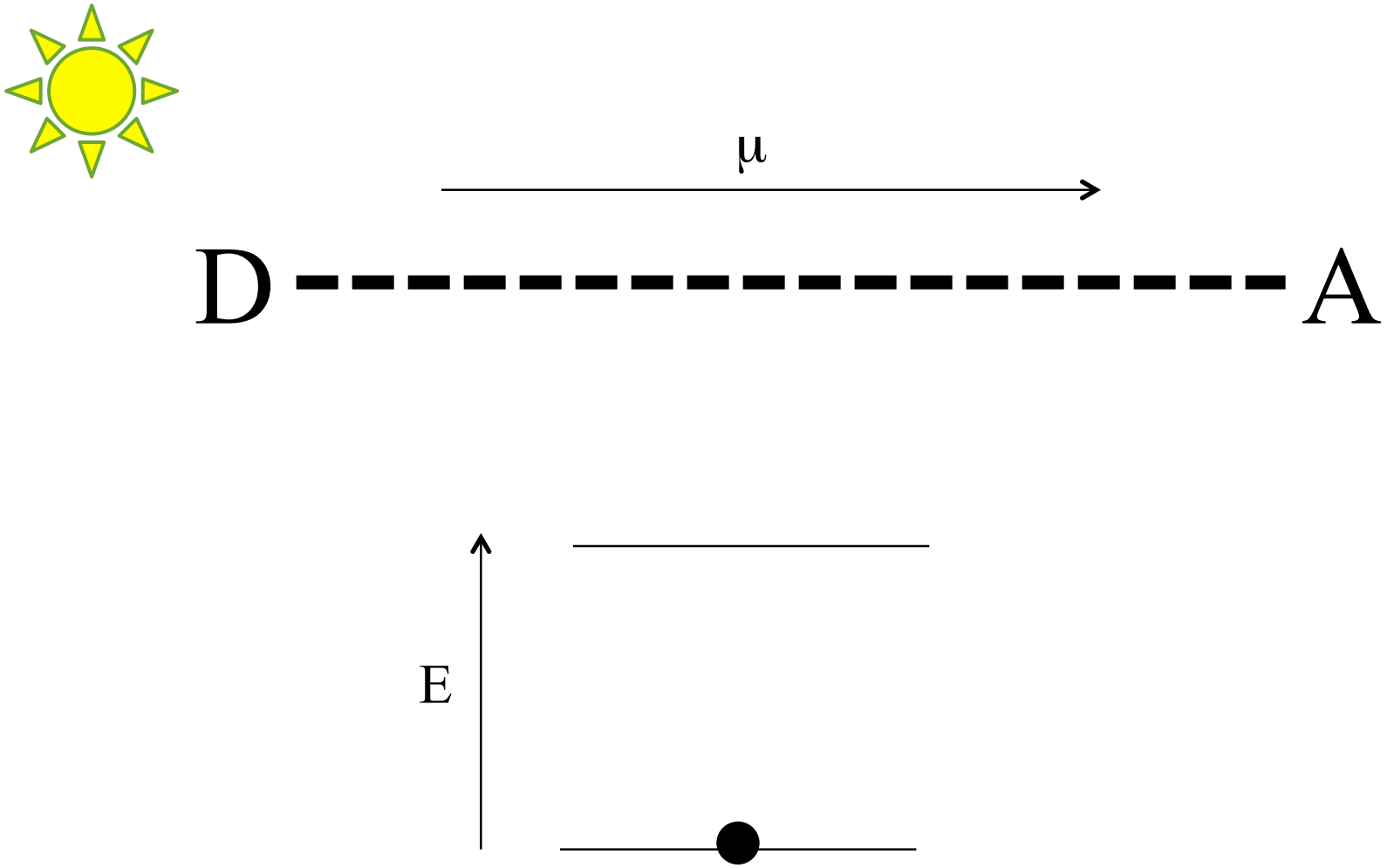
Creating molecular design rules

Molecular Design Rules

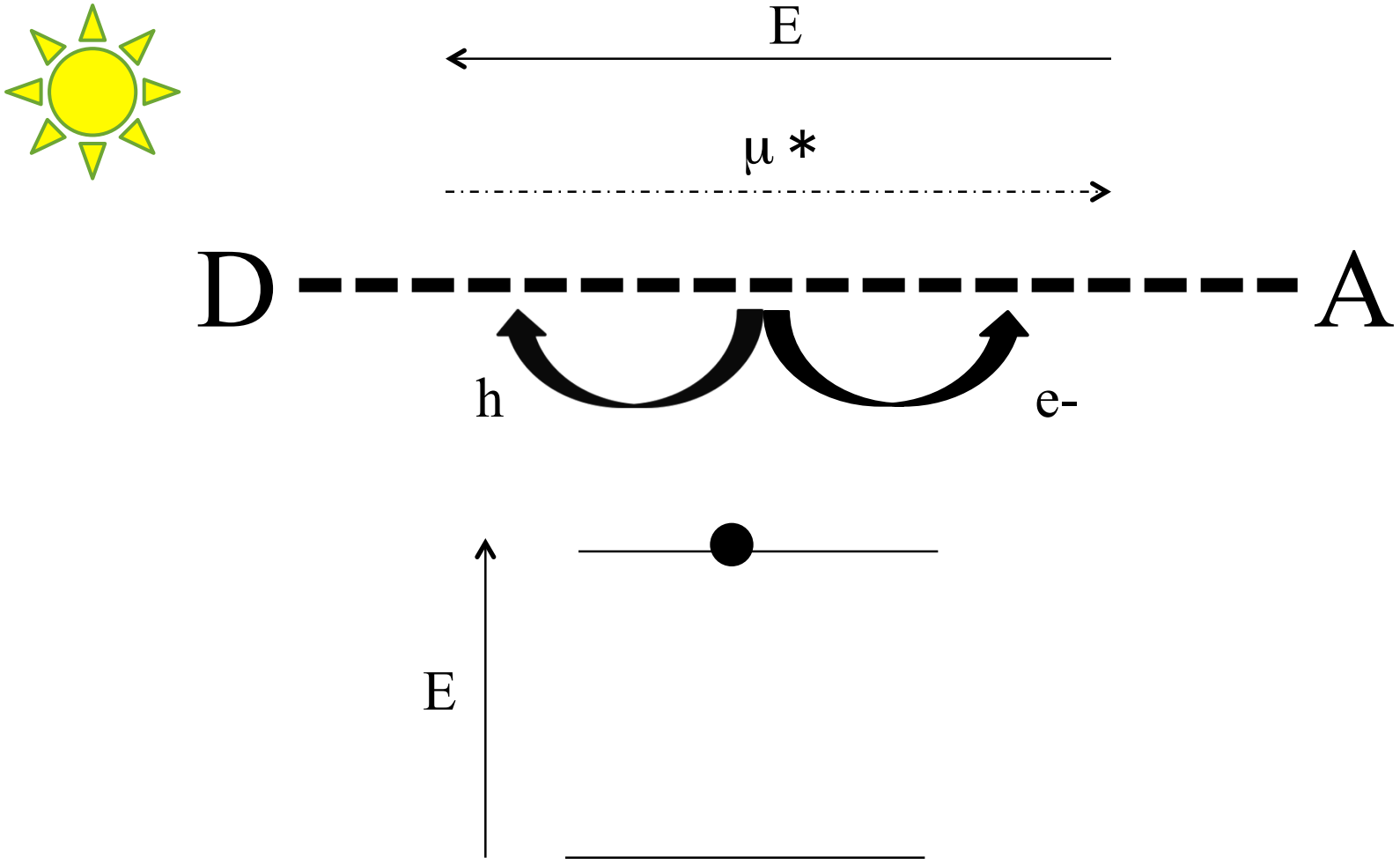




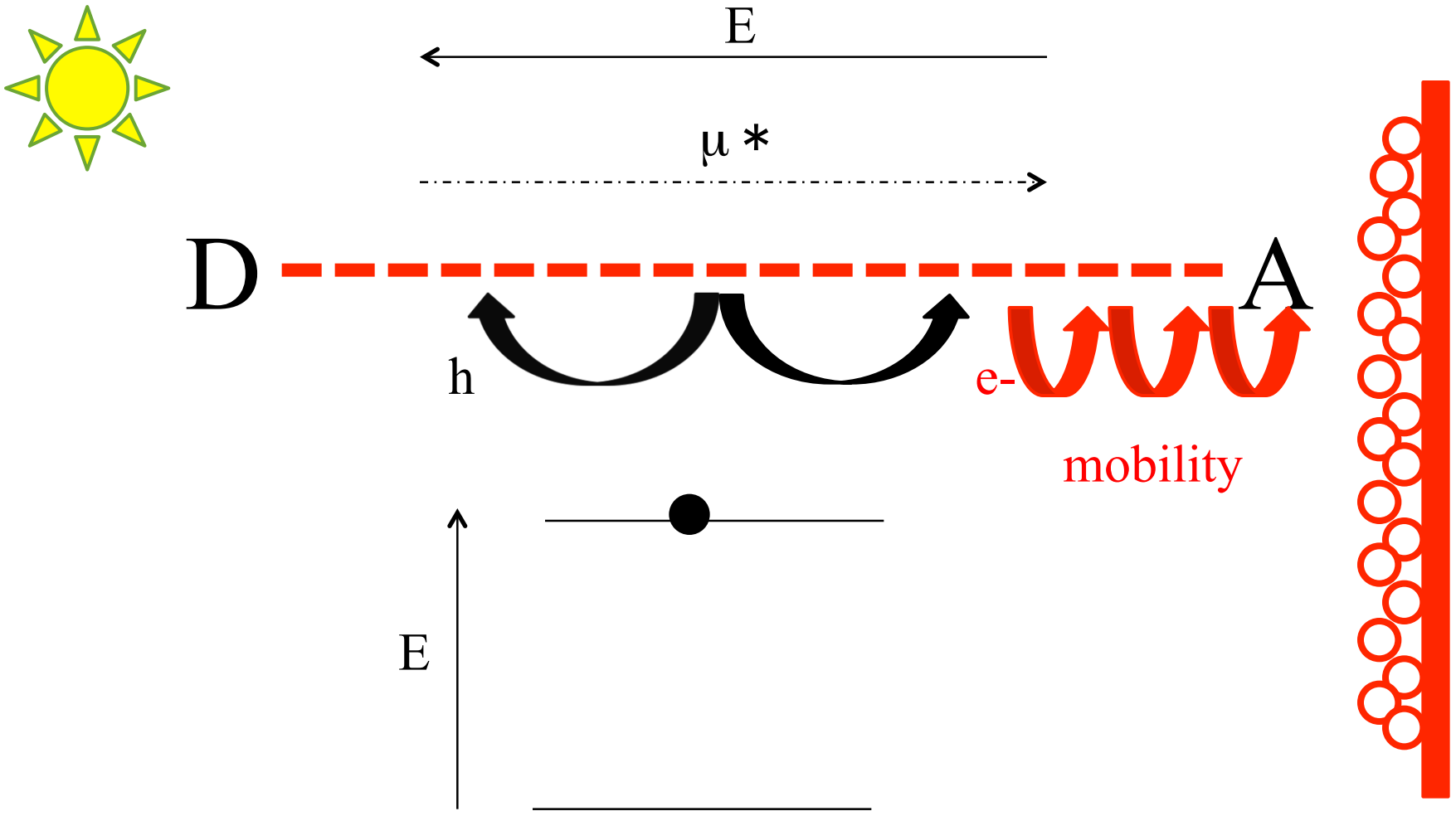
Molecular Design Rules



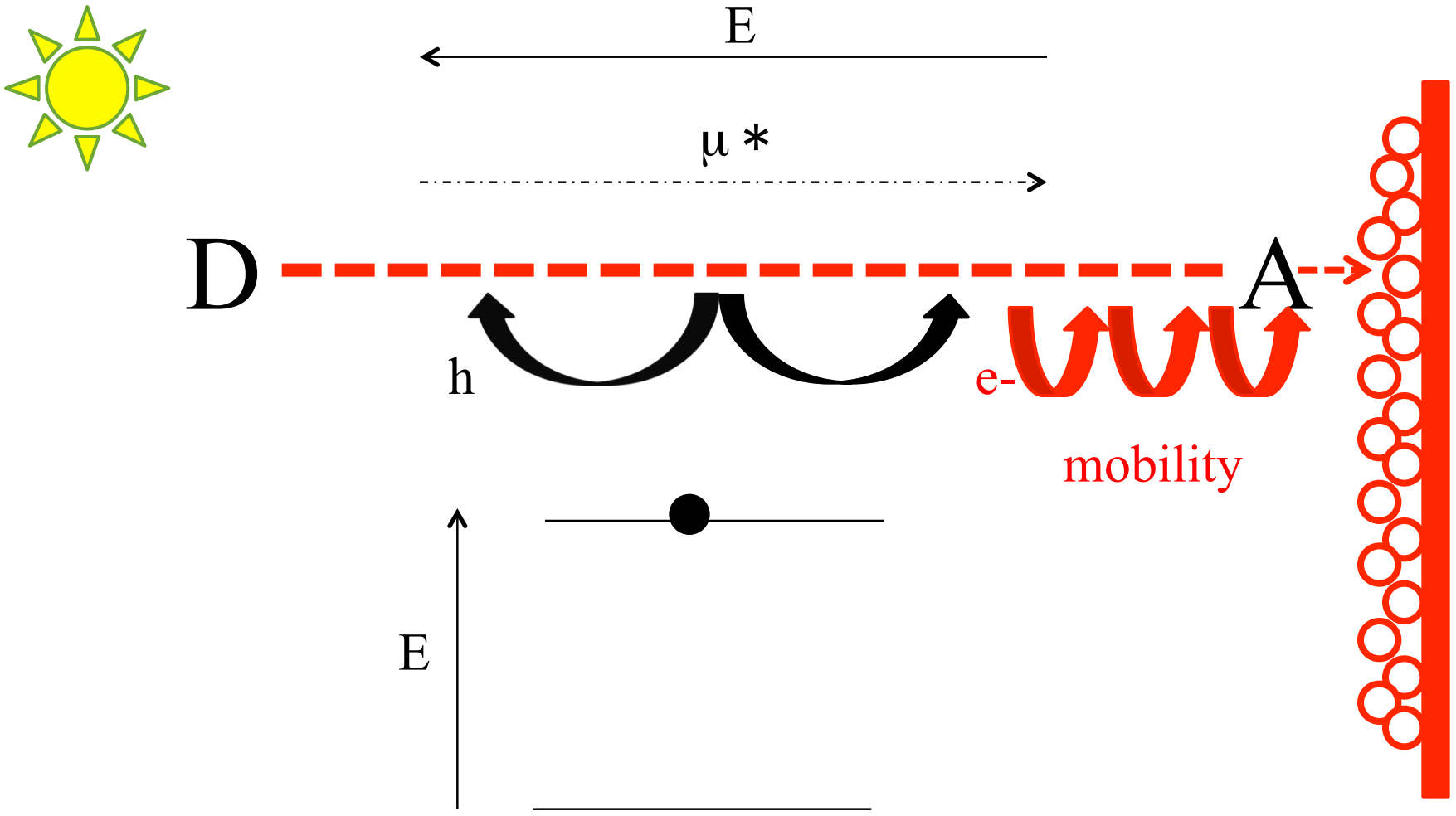
Molecular Design Rules



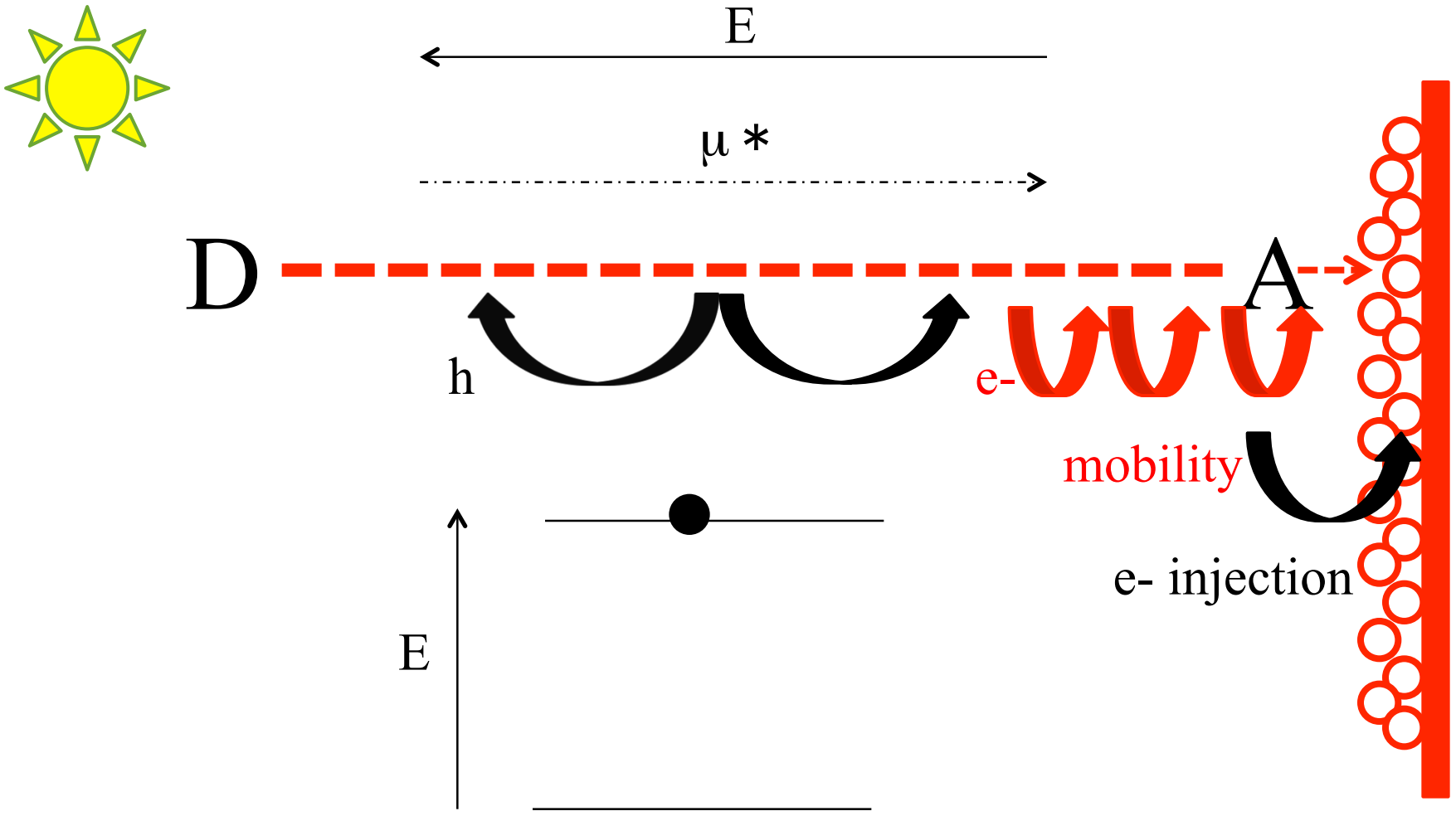
Molecular Design Rules



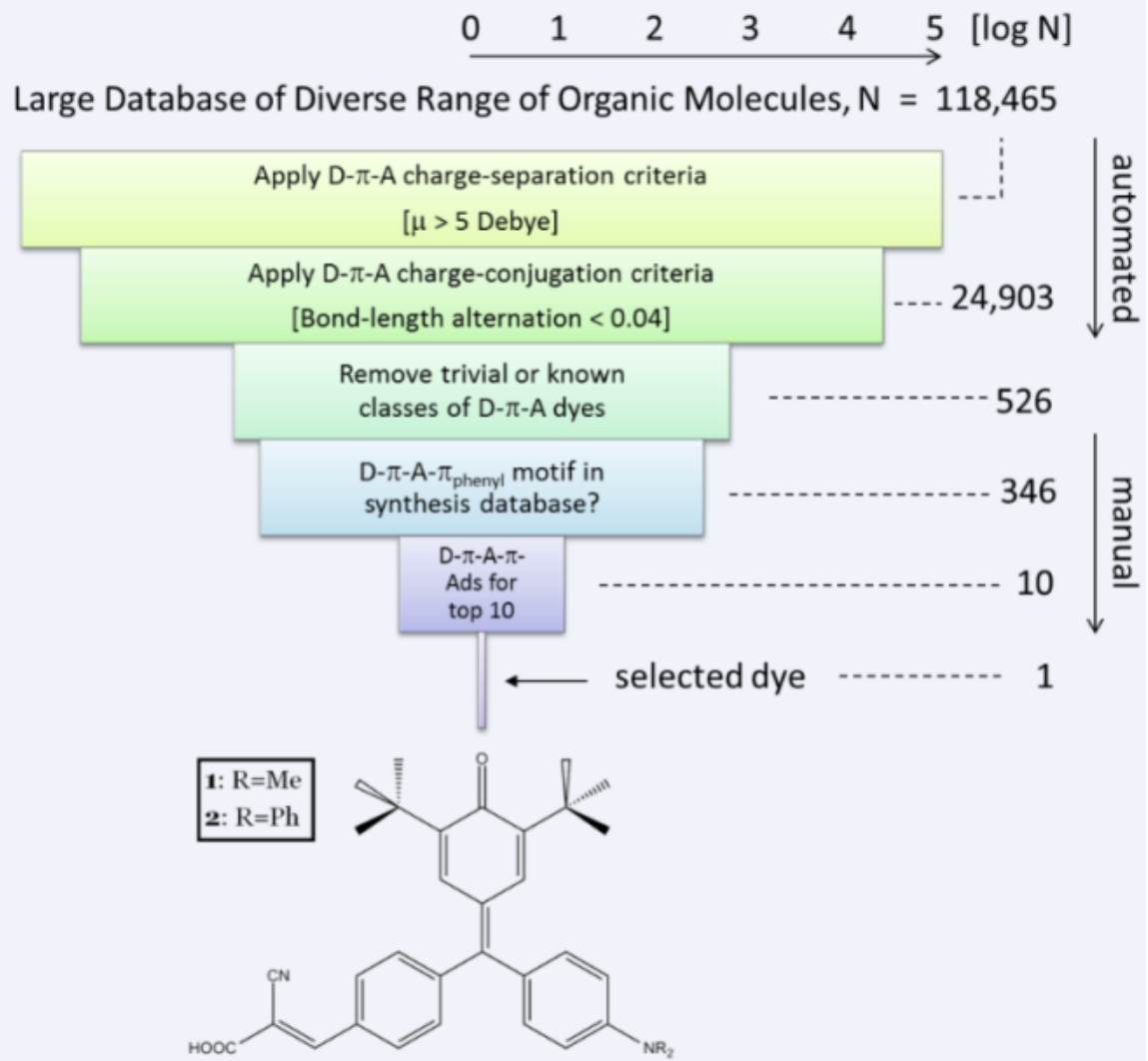
Molecular Design Rules



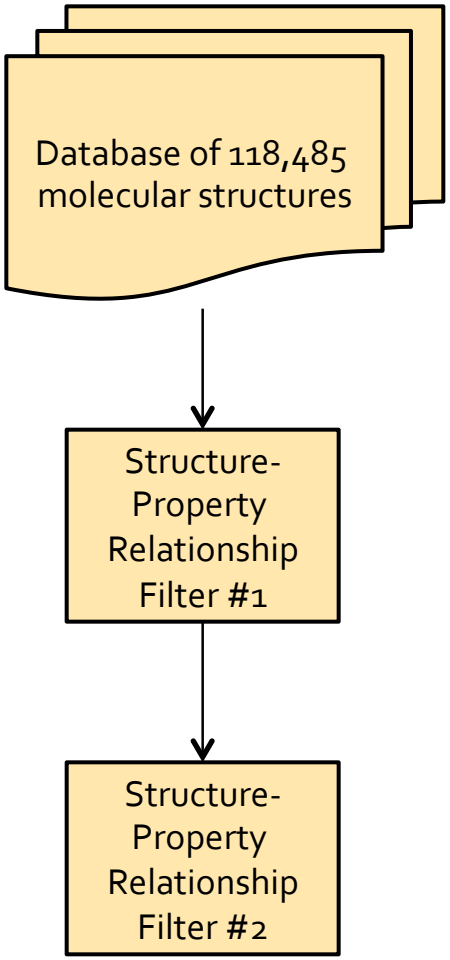
Molecular Design Rules



Overarching Data-Mining Strategy



Auto-sorting a set of chemical space



A representative set of organic chemical space

#1 Apply D-π-A charge-separation criterion

Calculate μ for all 118,465 molecules

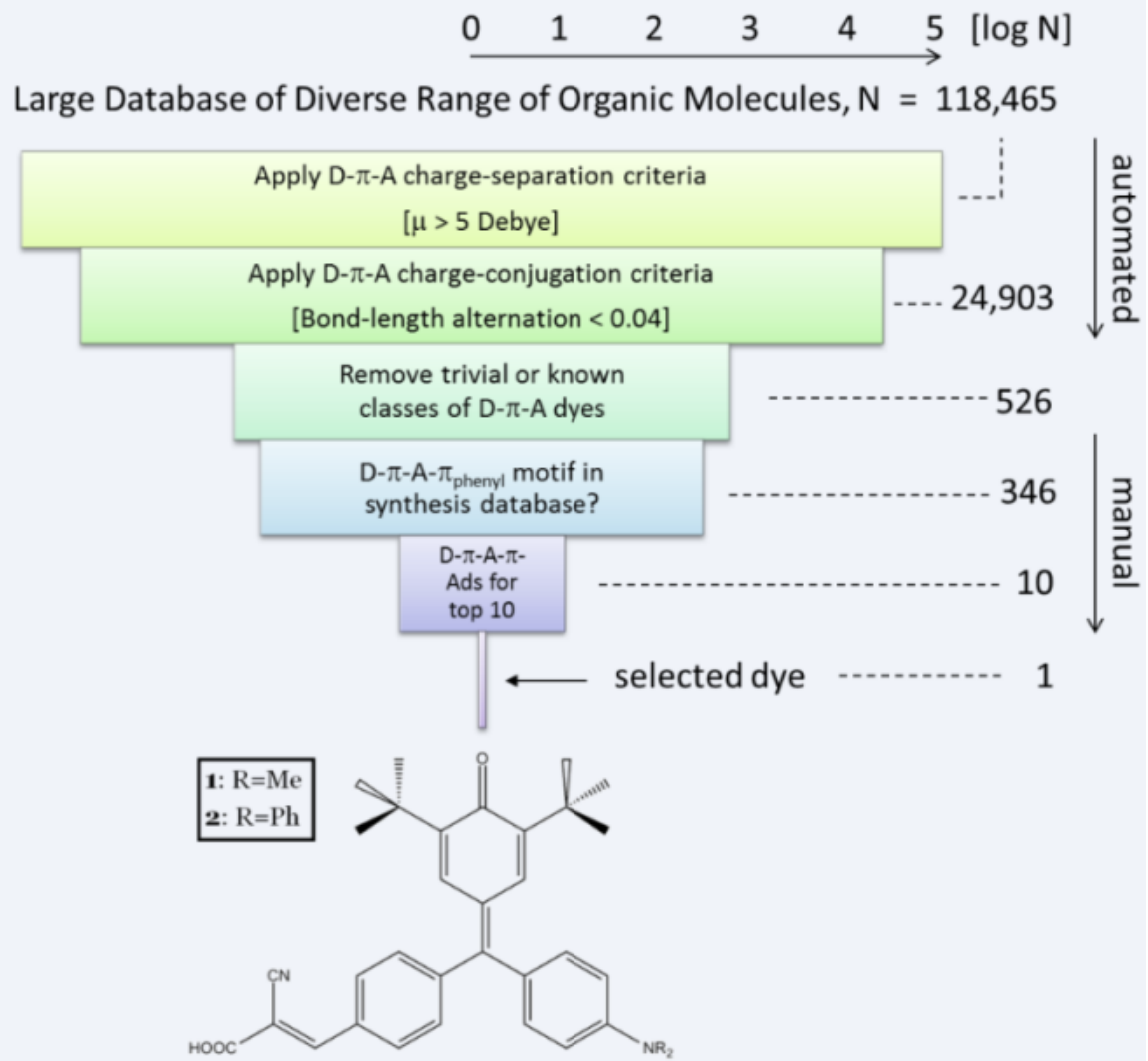
Set threshold: $\mu > 5$ D

Semi-empirical PM7 in MOPAC

#2 Apply D-π-A charge-conjugation criterion

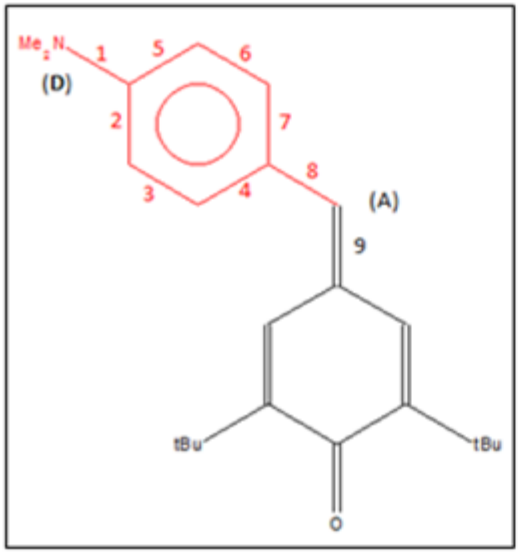
Encode bond-length Alternation calculation

Overarching Data-Mining Strategy



Encoding D-π-A Bond-length Alternation

Example Compound



BLA for each bond emanating from D (nitrogen)

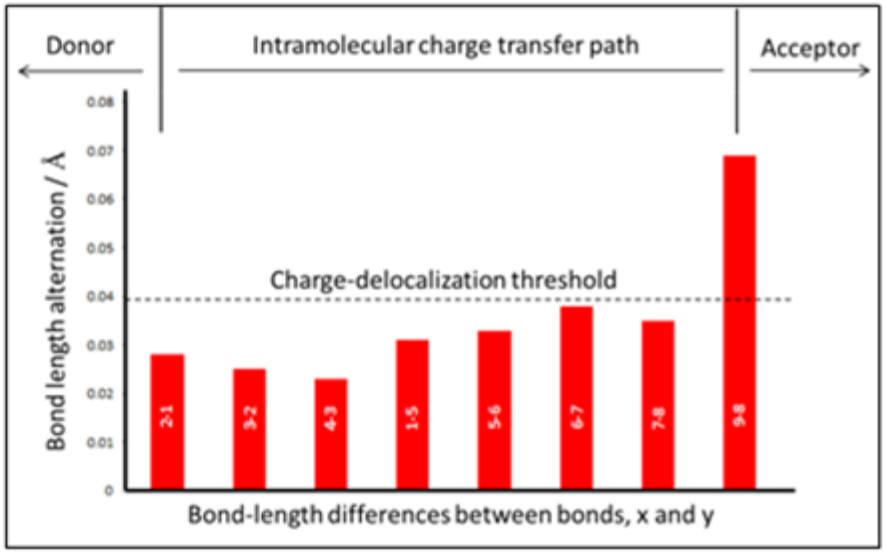


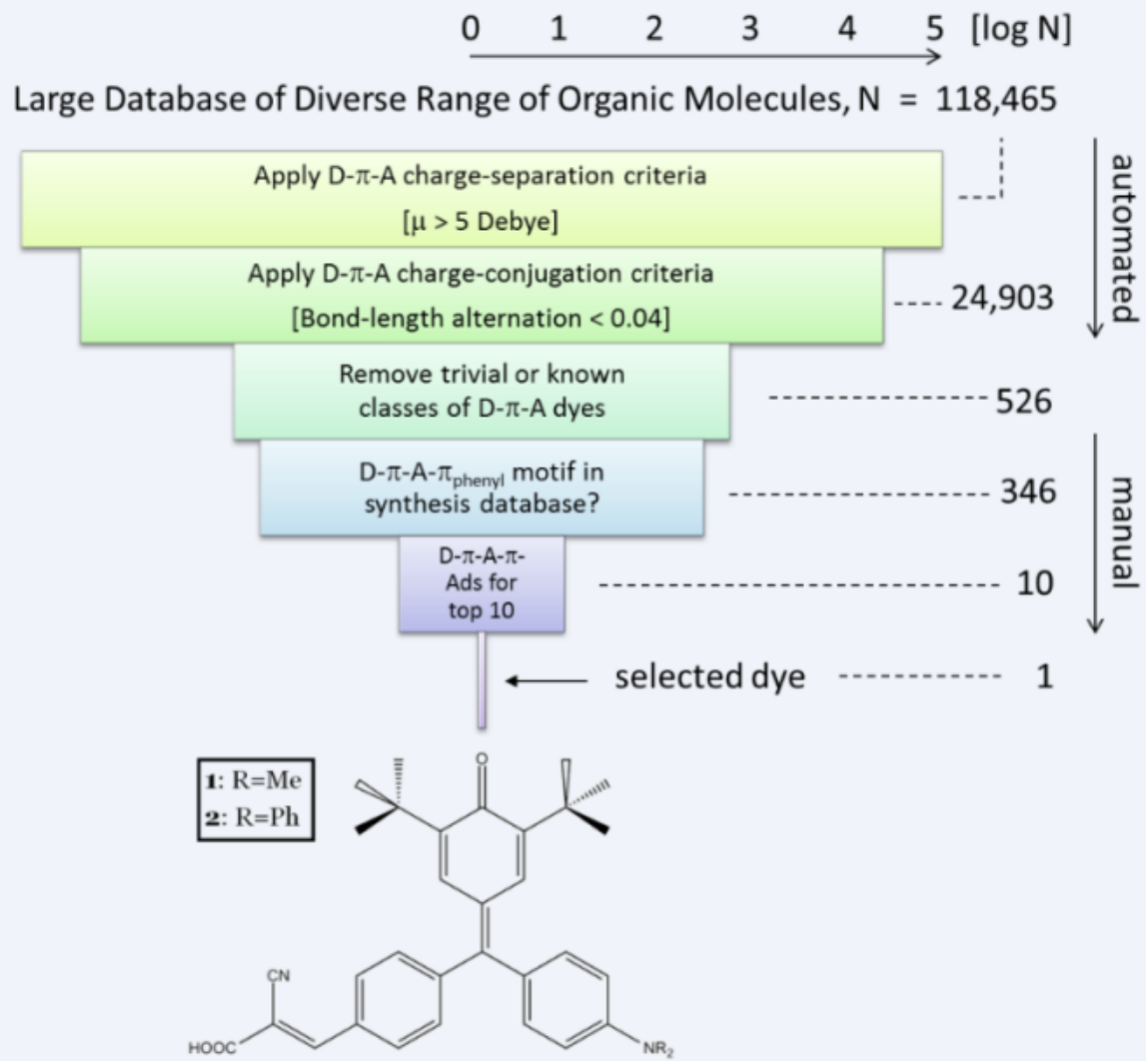
Illustration of the BLA calculation of all bonds, n, in the red fragment of the example compound.

The donor, D, is pre-selected from Hammett constants; here, D is the nitrogen on the N-Me_2 substituent.

The electron acceptor group begins at A, as identified by the individual BLA that first exceeds the BLA threshold, 0.04 Å, as bonds emanate from D. In this molecule, this BLA (0.069 Å) lies between bond 8 and 9.

The overall BLA for the D-π-A molecular fragment is 0.0308 Å, *i.e.* significantly less than the threshold, BLA < 0.04 Å. Thus, this molecule passes the charge-delocalization criteria in the filtering algorithm.

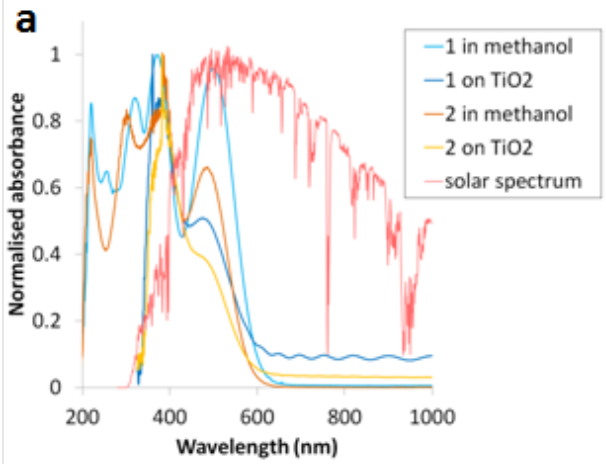
Overarching Data-Mining Strategy

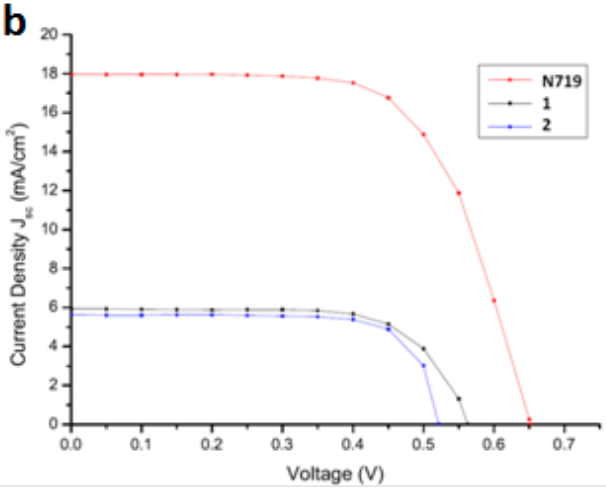


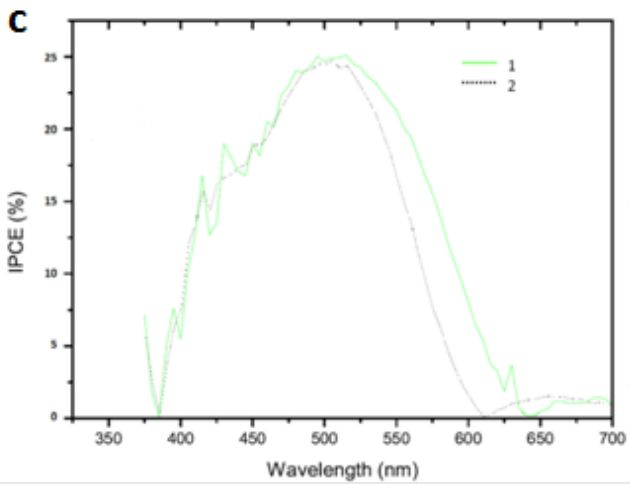
Experimental Validation

DSC Device Testing

Dye	V_{oc} (V)	J_{sc} (mA/cm^2)	FF	η (%)	η_{dye}:η_{N719} (%)
1	0.56	5.92	0.69	2.32	30.8
2	0.52	5.62	0.75	2.20	29.2
N719	0.65	17.96	0.64	7.54	100

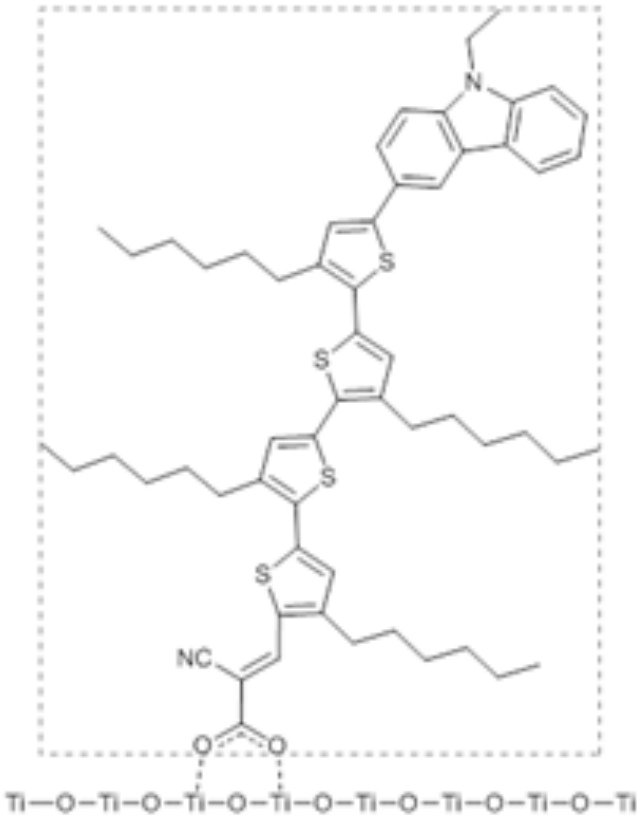




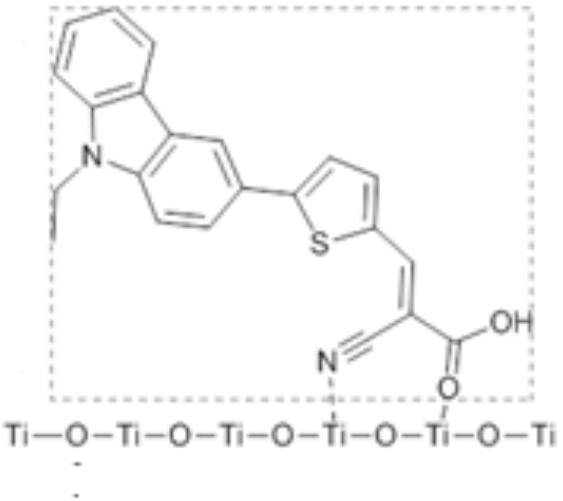


Molecular Building Blocks of DSC Dyes

e.g.



MK-2 [PV eff: 7.7%]



MK-44 [PV eff: 1.8%]

Materials discovery of DSC dyes

Molecular building blocks of predicted new classes of DSC dyes
→1/3 world's best performance DSC dye in experimental validation

→ Far from a null result - Prediction strategy works!
(and should improve with further refinement of data-mining method)

No significant dye aggregation effects observed in 1 or 2

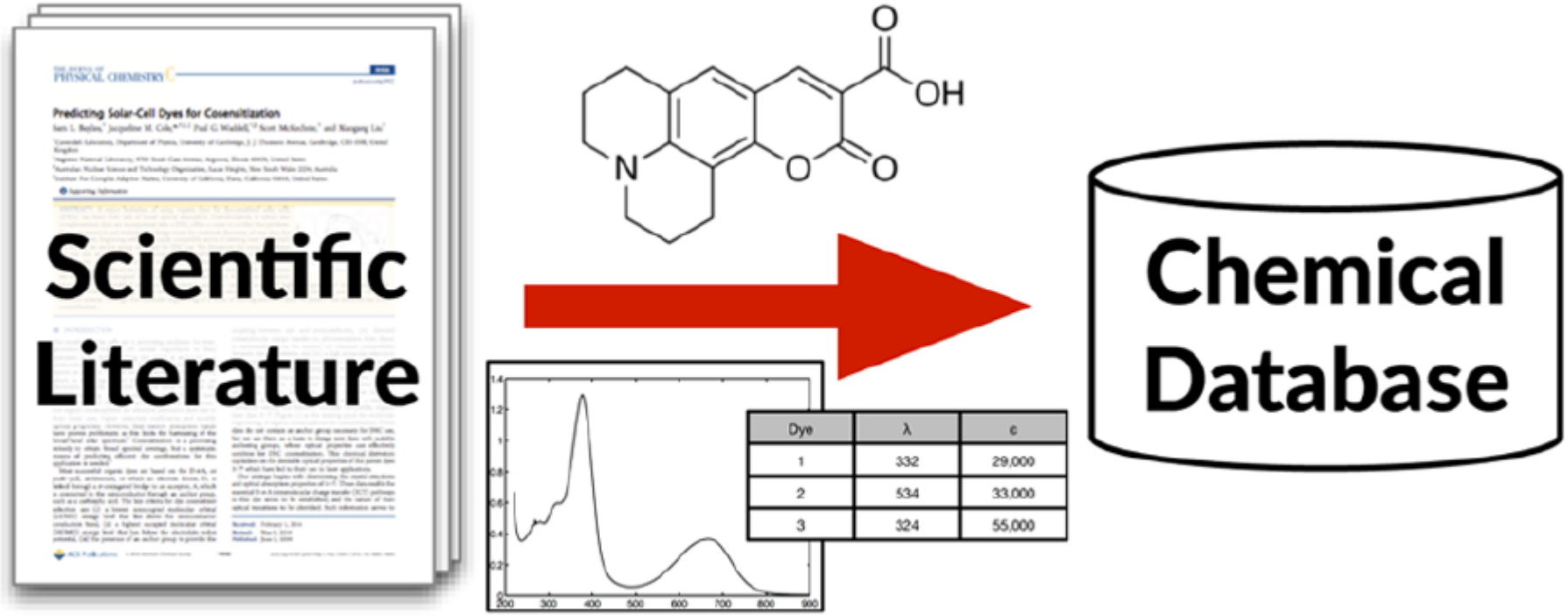
1 and 2 exhibit unusually high melting points

What next?

Use custom-made material databases?

Database Auto-generation tools

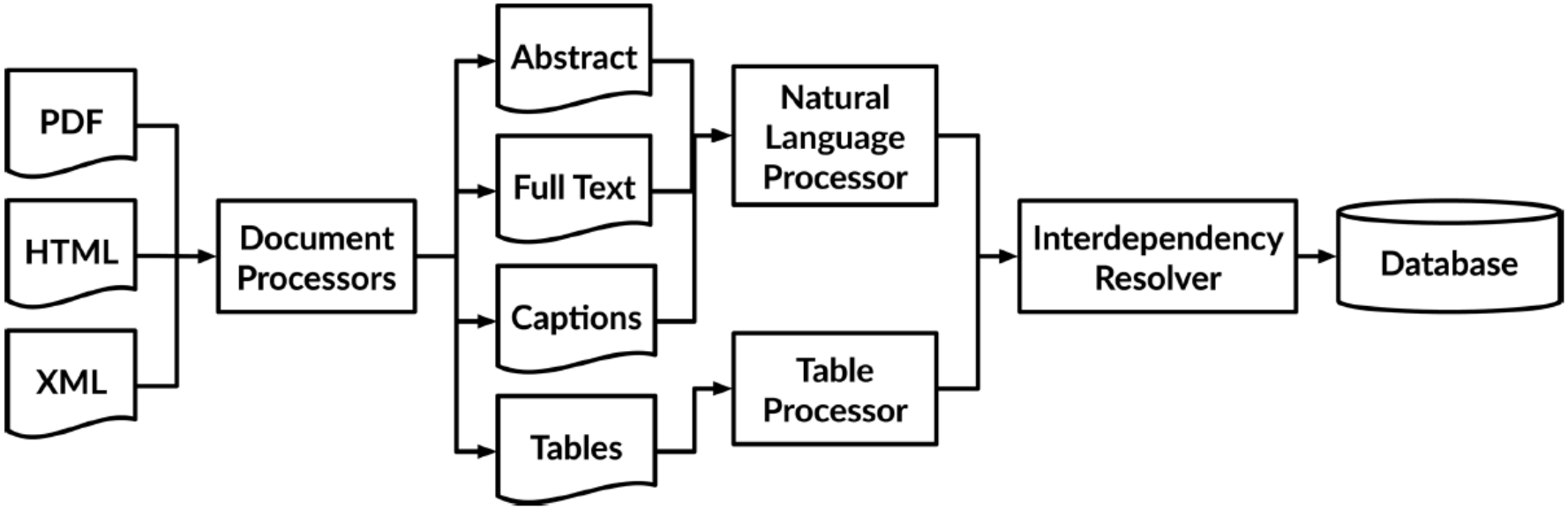
ChemDataExtractor



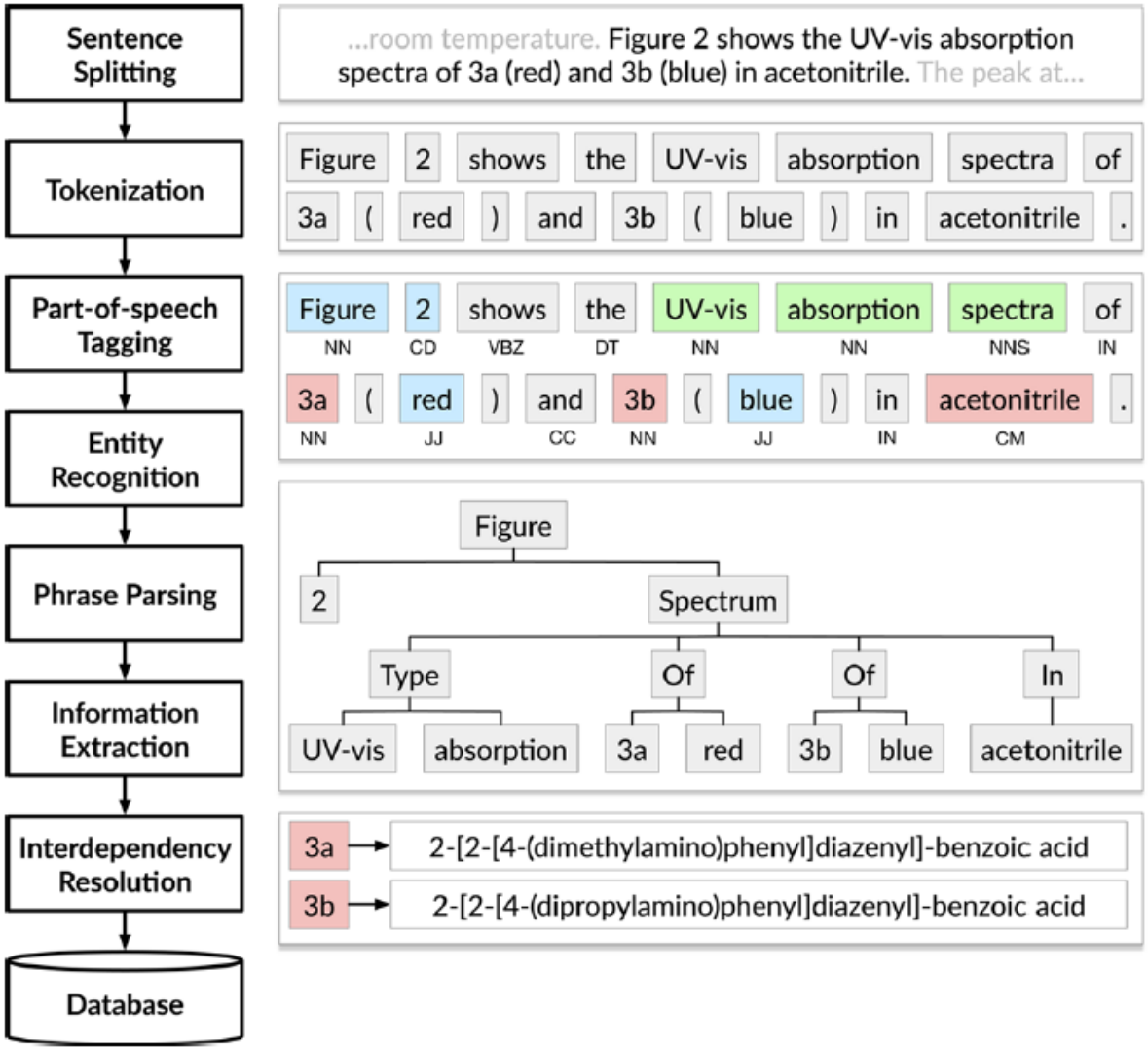
www.chemdataextractor.org

M. C. Swain, J. M. Cole *J. Chem. Inf. Model*. 56 (2016) 1894-1904

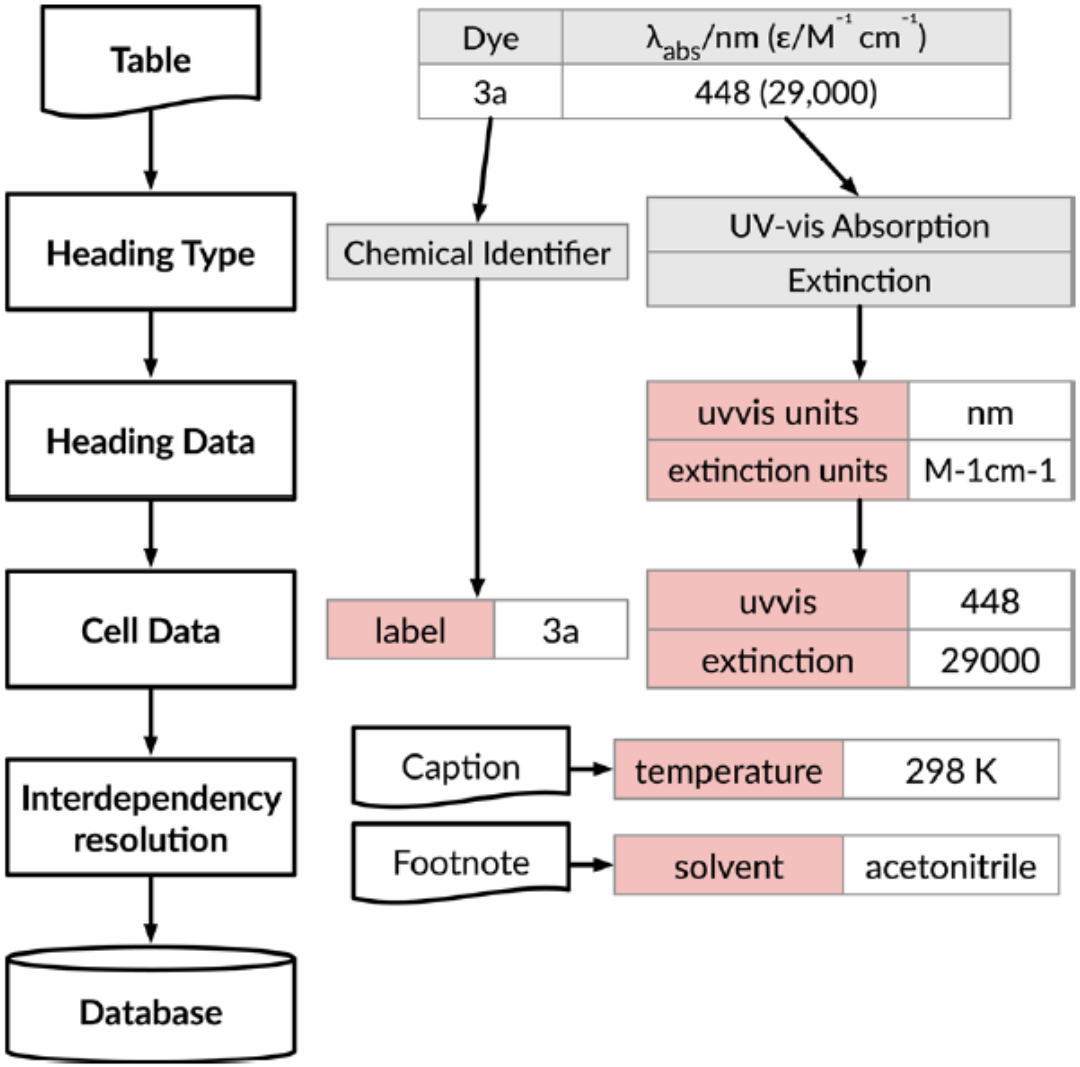
ChemDataExtractor



ChemDataExtractor



ChemDataExtractor



ChemDataExtractor

The dye 2-[2-[4-(dimethylamino)phenyl]diazenyl]-benzoic acid (**3a**) was added…

UV-vis spectra were recorded using an Agilent8453 diode array spectrophotometer.


```
{
  “name”: “2-[2-[4-(dimethylamino)phenyl]diazenyl]-benzoic acid”,
  “label”: “3a”,
  “uvvis”: [ {
    “solvent”: “acetonitrile”,
    “apparatus”: “Agilent8453 diode array spectrophotometer”,
    “peaks”: [ {“wavelength”: “448”, “extinction”: “29,000” } ],
  } ],
}
```


Figure 2: UV-vis absorption spectra of 3a in acetonitrile.

Dye	λ_{max}/nm (ε/M^{-1} cm^{-1})
3a	448 (29,000)
3b	415 (48,000)

ChemDataExtractor

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$\text{F-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

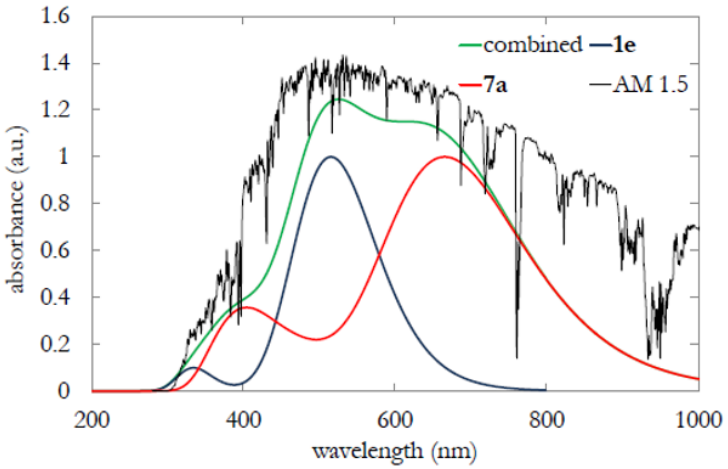
	precision	recall	F-score
chemical identifier records	94.1%	92.7%	93.4%
spectrum records	88.3%	85.4%	86.8%
chemical property records	93.5%	89.6%	91.5%

	precision	recall	F-score
chemical names	97.4%	96.3%	96.8%
alphanumeric labels	99.3%	97.3%	98.3%
full chemical identifier records	**94.1%**	**92.7%**	**93.4%**

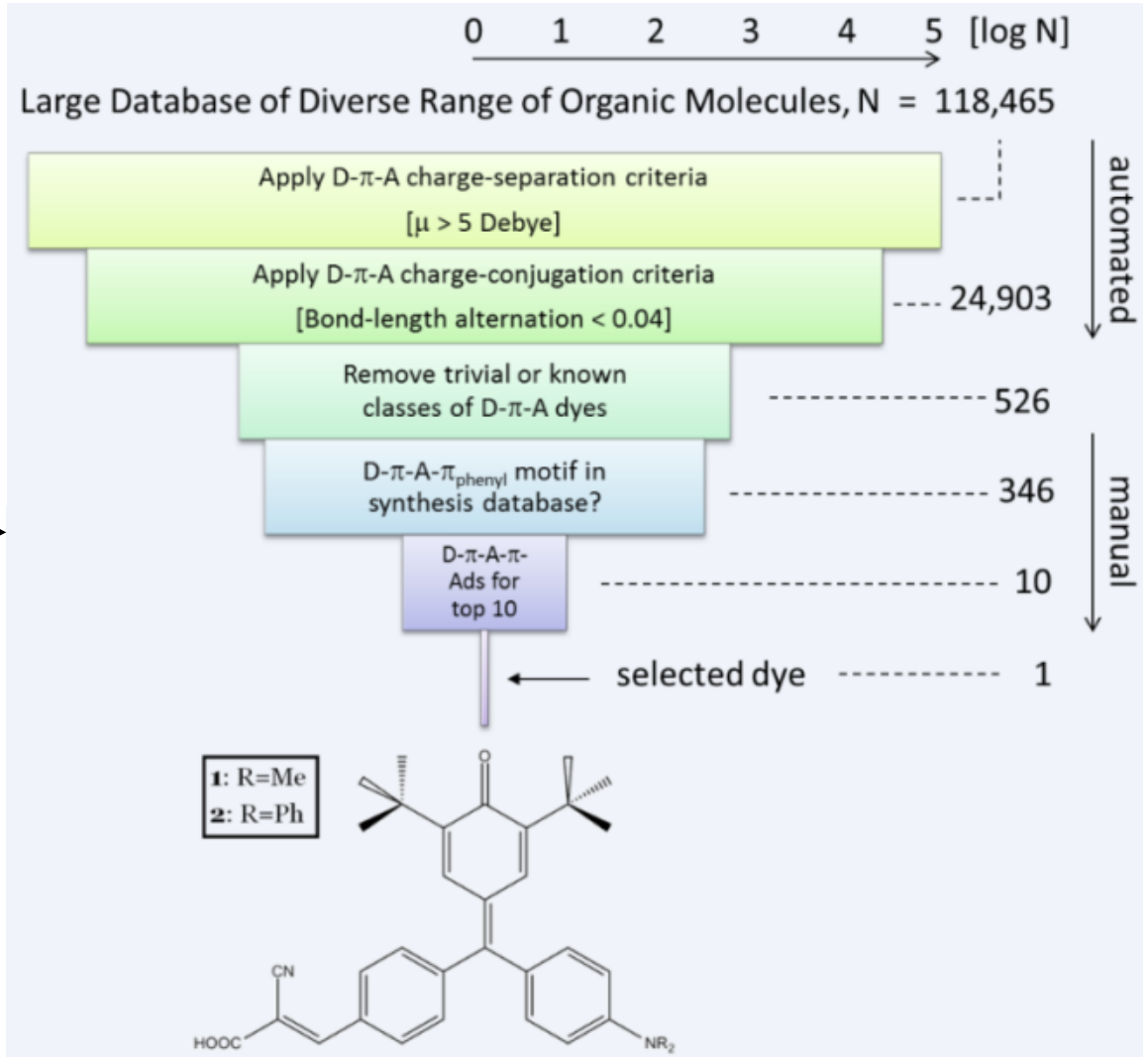
	precision	recall	F-score
spectrum type	99.9%	96.7%	98.4%
chemical subject	93.4%	90.3%	91.8%
peak values	98.6%	95.4%	96.9%
solvent	99.5%	96.7%	98.1%
temperature	100%	87.5%	93.3%
apparatus	96.9%	91.0%	93.8%
full spectrum records	**88.3%**	**85.4%**	**86.8%**

	precision	recall	F-score
property value	100%	95.9%	97.9%
property units	100%	94.8%	97.4%
chemical subject	93.5%	89.6%	91.5%
solvent	100%	94.4%	97.1%
temperature	100%	88.9%	94.1%
apparatus	100%	87.5%	93.3%
full property records	**93.5%**	**89.6%**	**91.5%**

Materials discovery of new DSC dyes



Large database of UV/vis absorption spectra



Materials Discovery 'Take #2' - The ADSP project

Overview of Project: 4 Key Steps

Academic Literature

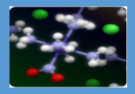
Data Sources

Apply/Create Software Tools for Data Extraction:

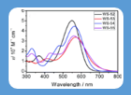
ChemDataExtractor

Stage 1

Data Extraction



Material structure



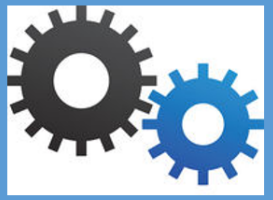
Physical properties

Databases

Stage 2

Data Enrichment

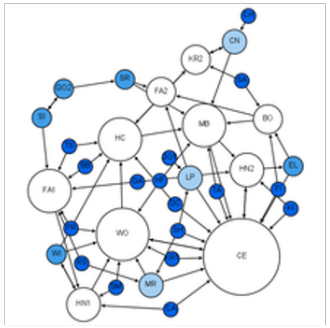
- Machine learning
- Quantum-chemical calculations (φ)



Enriched Data

Stage 3

Create algorithms for materials prediction



Stage 4

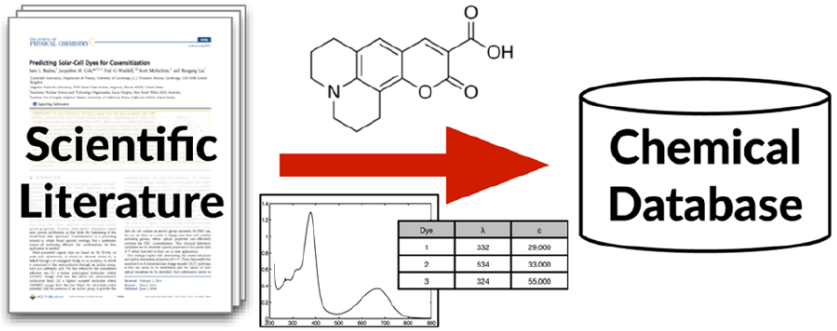
Prediction & Validation

Feedback

- Materials Prediction
- Experimental Validation

Step 1: Data Extraction

- Negotiating publisher agreements
  - Royal Society of Chemistry
  - Elsevier
  - Wiley
  - Springer



- Parallelization of ChemDataExtractor code for Mira and Theta
- MongoDB for storing data

Stage 2: Enriching Data with Machine Learning and Quantum-Chemical Calculations

Chemical (SMILES)	property(1)	property(2)	Molecular descriptors
------	----	----	10110110…
------	----	----	10111110…
------	----		11110101…
------	----	---	11011010…

Missing Data

Training Data

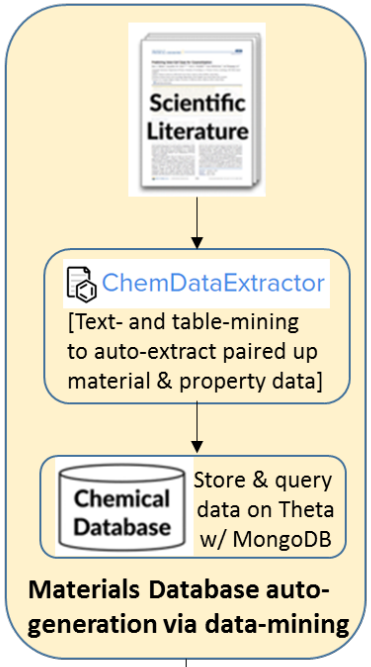
Machine learning:

Neural network, Random forest…

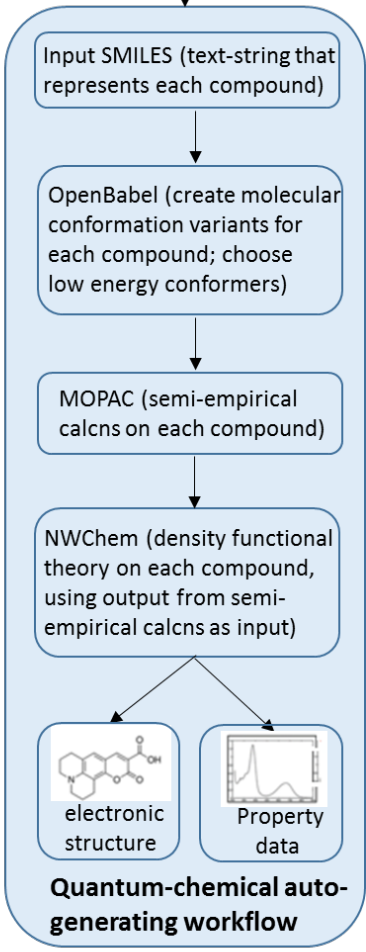
....
....
....
....

Feature vector

Predict Missing Data



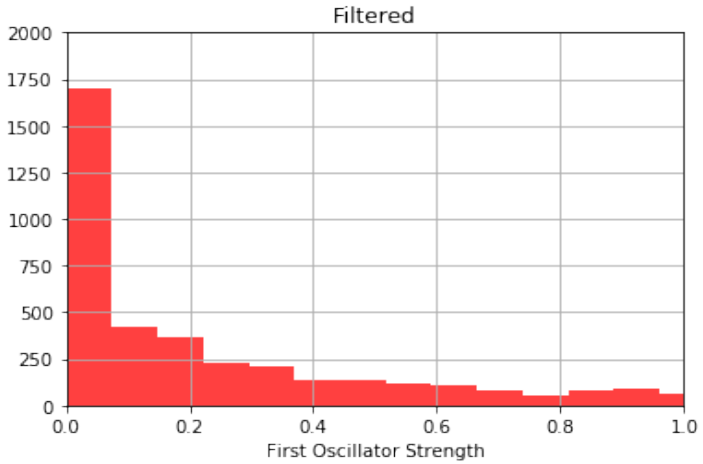
Augmenting Data with Quantum-Chemical Calculations
(We have developed a high-throughput workflow that pairs experimental measurements on a chemical with its molecular wave function: for **property prediction**)

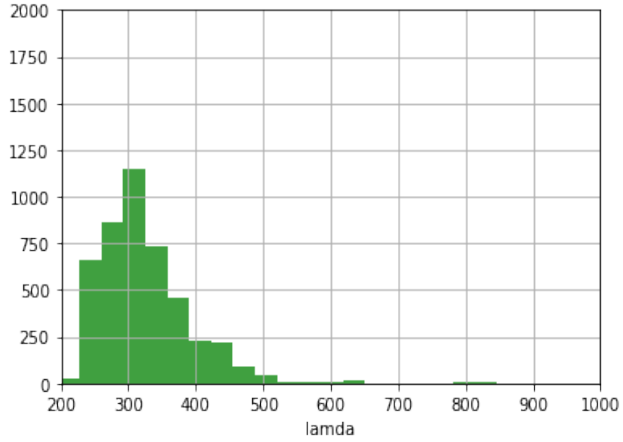


High-throughput Computation

So far…

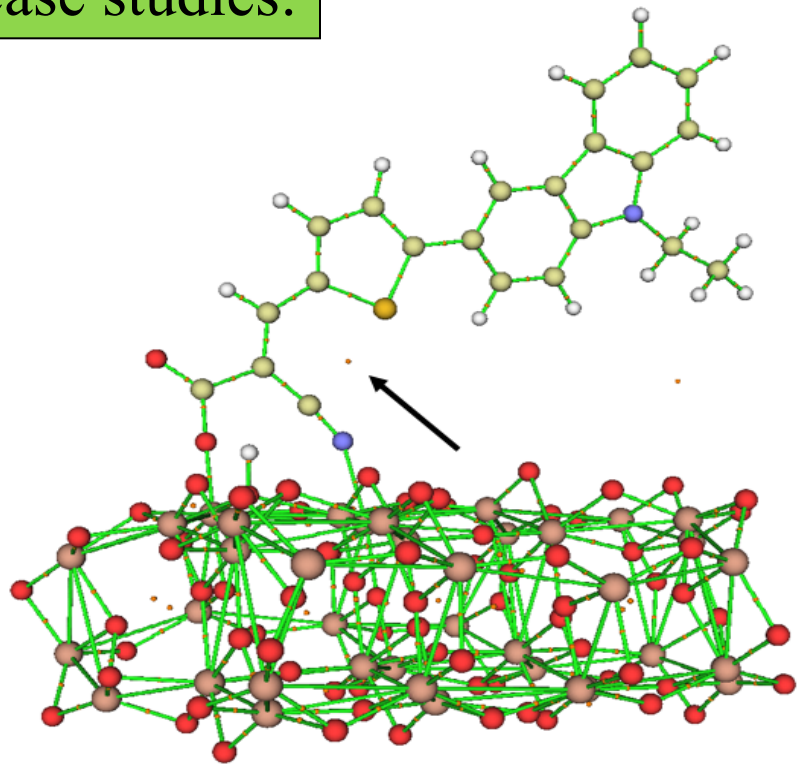
- 18,309 lambda max values from experiments
- 9,094 compounds.
- 4,519 computed with DFT and sTDA/sTDDFT .
- (sTDXX is an approximation to TDDFT)
- 917 possible candidates to be computed with Linear Response Coupled Cluster

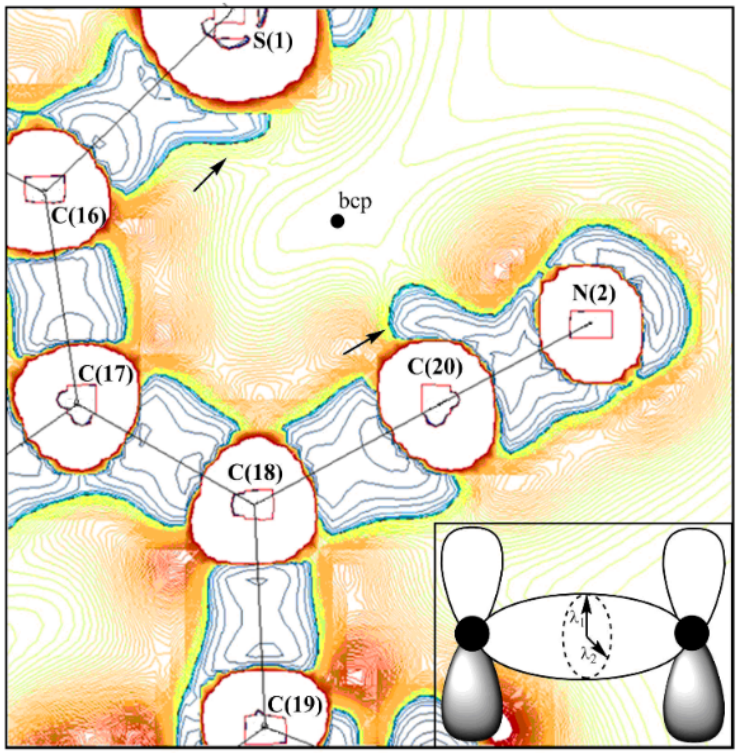




Step 3: Design materials prediction

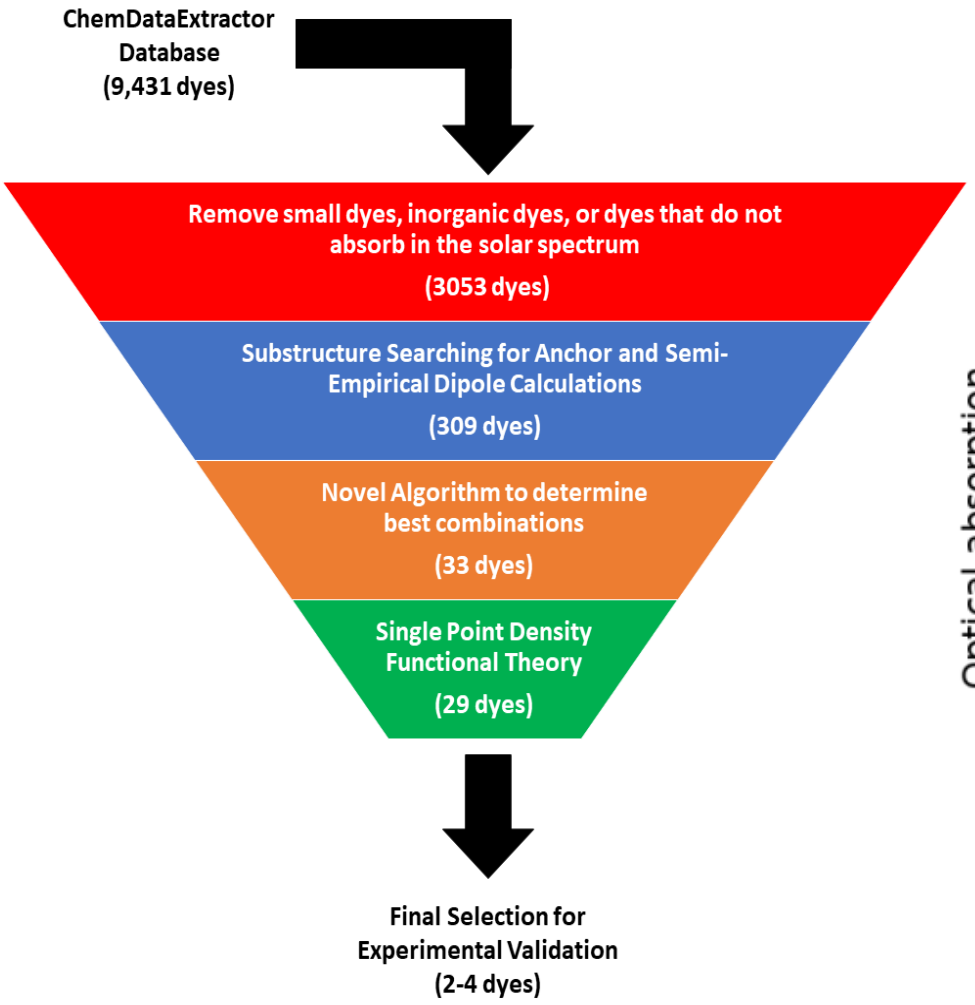
Case studies:

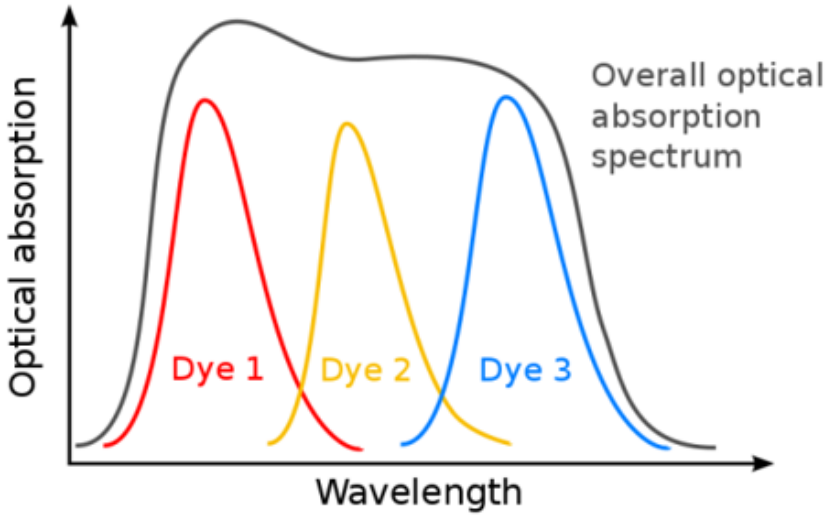




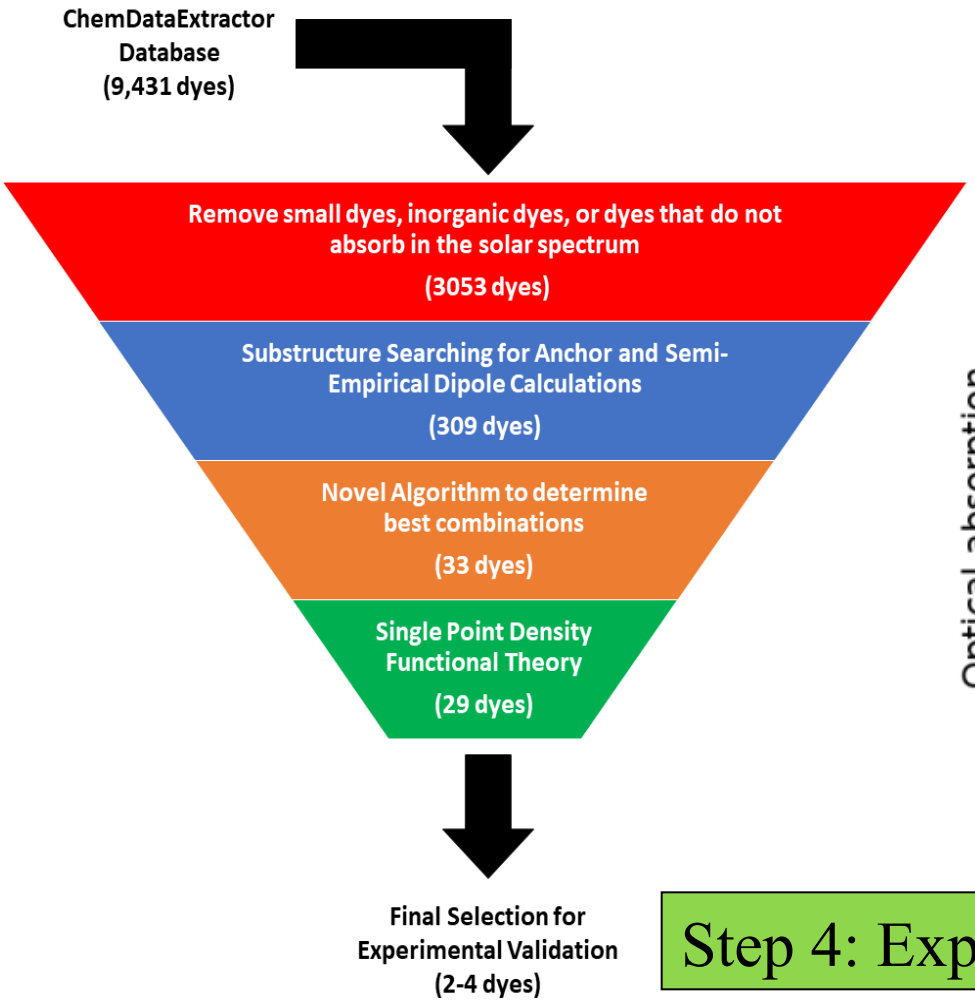
Cole, Vázquez-Mayagoitia et al, *ACS Appl. Mater. & Interfaces*, 9 (2017) 25952-25961.

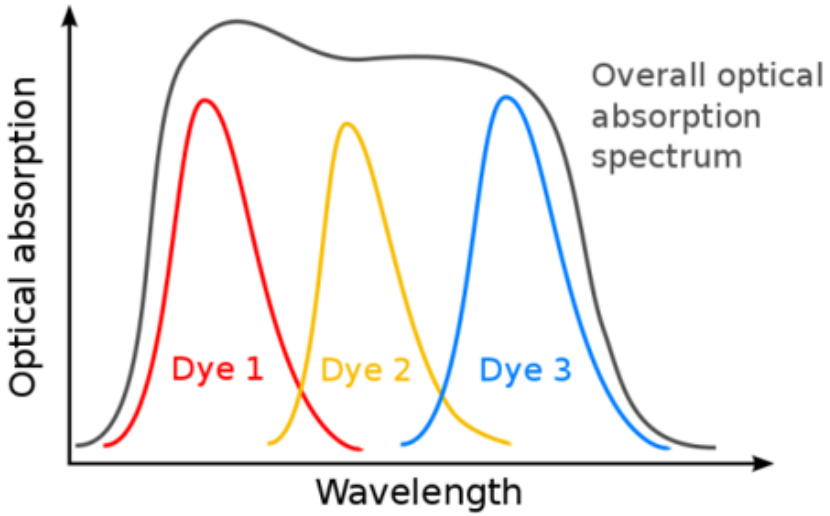
Step 3: Prediction workflows





Step 3: Prediction workflows





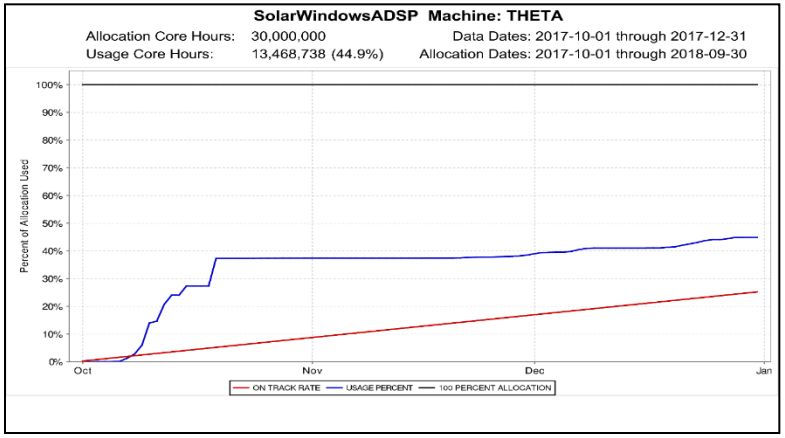
Step 4: Experimental Validation (NOW)

Next Steps

- More Data: New Data Extraction capabilities:

FigureDataExtractor

- More Theta/Mira calculations



- Computation versus Experiment
- Data enrichment: Machine Learning
- Algorithm Design for Structure-Function relationships:
  - Case studies using DFT and MD
- Experimental Validation (CNM)

Acknowledgements

Alvaro Vázquez-Mayagoitia, Ganesh Sivaraman, Samuel Flender, Yasi Ghadar, Elise Jennings,
Venkat Vishwanath, Paul Coffman, Pragneshkumar Balvantbhai Patel, Misha Salim, User Support (ALCF)
Petra Rudolf, Giota Stathi (Zernike Institute of Advanced Materials, Netherlands)
Hiroaki Ozoe, K. Fuji, C.Kitamura, Takeshi Kawase (Uni. of Hyogo, Japan)
Hiroyuki Kurata (University of Osaka, Japan)
Chris Cooper, Kian Sing Low, Martin Blood-Forsythe, Yun Gong, Ed Beard, Matt Swain, Hao Chen
(Cambridge University)

References:

Data-mining led DSC dye-discovery:
Cole et al, *PCCP*, 16 (2014) 26684

Text- and table mining:
Swain & Cole *J. Chem. Inf. Model*. 56 (2016) 1894-1904

ChemDataExtractor
www.chemdataextractor.org