Linear Free Energy Relationships:

History and Modern Application in Organic Synthesis

rational design of catalysts

mechanistic understanding

catalyst activity

catalyst properties
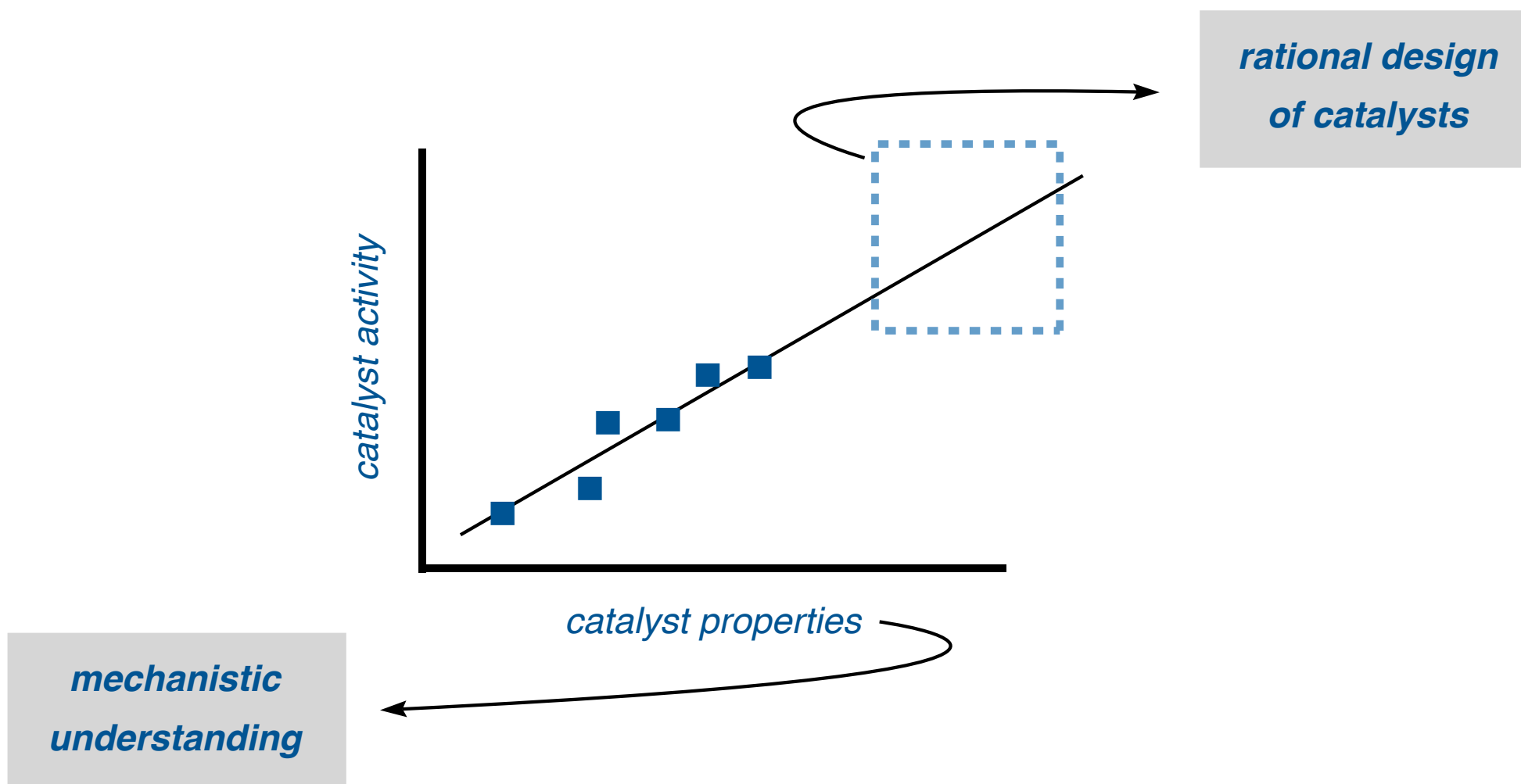
Nick Till

MacMillan Group Meeting

December 12, 2018

Outline for the Presentation

■ **Development of LFER Parameters from Brönsted through 2018**

- origin of the Hammett plot and modifications

- development of steric parameters and modifications

- generalization of LFERs and applications in catalysis

■ **Origin and Adoption of Complex LFERs in Organic Synthesis**

- statistical origins and intuition behind "multivariate linear regressions"

- seminal work in applying MLR to designing enantioselective catalysts

- a study in simultaneous electronic and steric parameterization

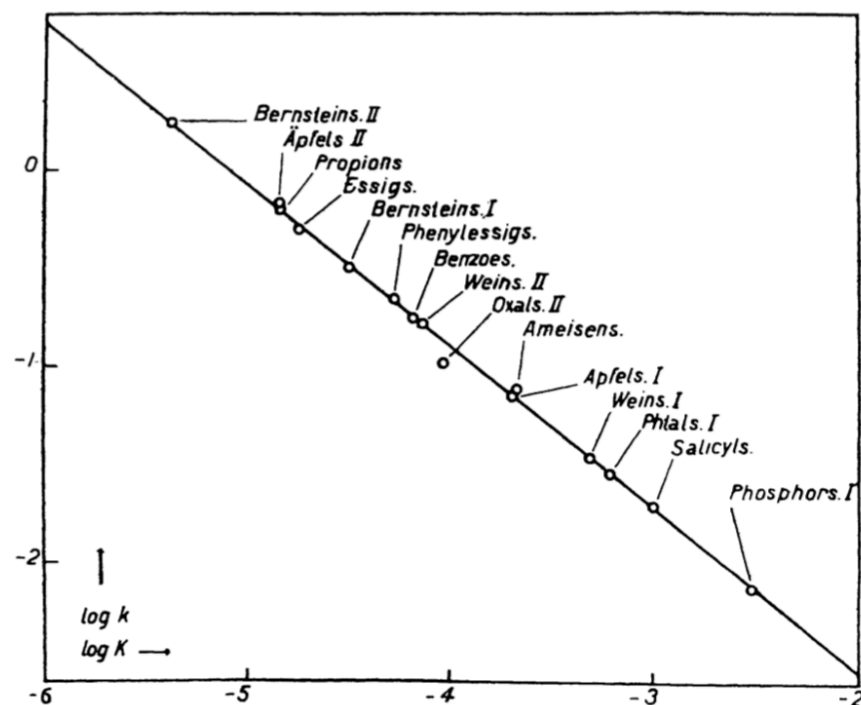■ **Recent Applications of MLR in Catalysis (Sigman-type LFERs)**

- site selectivity predictions in Fe-catalyzed C–H oxidation

- implementing molecular vibration as as a stereoelectronic parameter

- machine learning and the future of LFERs

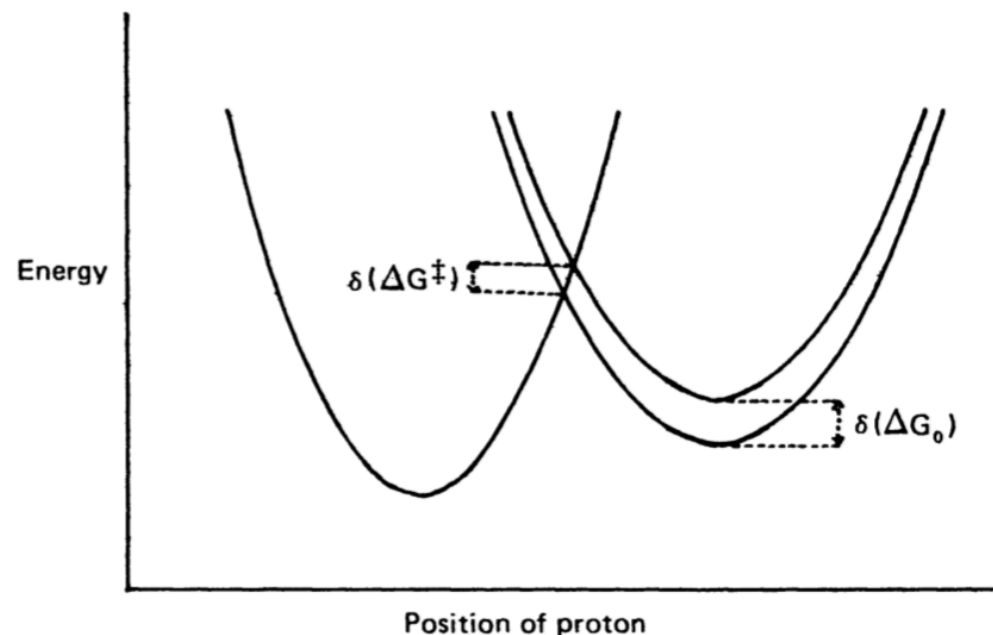# Early Developments in Linear Free Energy Relationships

■ First linear free energy relationship: Brönsted Relation, 1924

$$H_2N-NO_2 \xrightarrow{\ base\ } H_2O \qquad N_2O$$

$$\log_{10}(k) = \alpha \cdot \log_{10}(K_a) + C$$

initially considered purely empirical

Bernsteins. II
Äpfels II
Propions
Essigs.
Bernsteins. I
Phenylessigs.
Benzoes.
Weins. II
Oxals. II
Ameisens.
Apfels. I
Weins. I
Phtals. I
Salicyls.
Phosphors. I

log k
log K ⟶

rate of an acid-catalyzed reaction is a linear function of the acidity of the acid catalyst

Energy

$\delta(\Delta G^{\ddagger})$

$\delta(\Delta G_o)$

Position of proton

Polanyi proposes physical basis for the relation, 1935

J. Horiuti and M. Polanyi, *Acta Physicochim. U.R.S.S.,* **1935**, *2*, 505.

The Hammett Equation: Origin and Definition

■ Hammett relates benzoic acid pK_a values to electronic effects, 1937
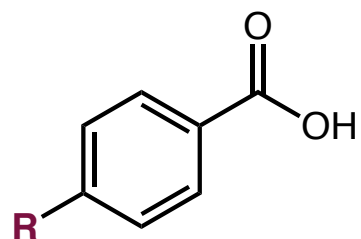
– y-axis –

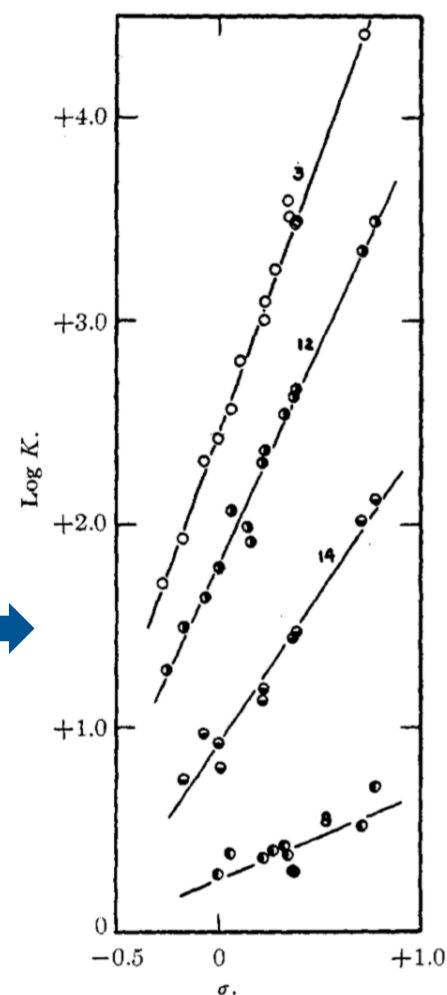$$\sigma = \log_{10}\left(\frac{K_a(\mathbf{R})}{K_a(\mathbf{R} = H)}\right) \qquad \log_{10}\left(\frac{K_{reaction}(\mathbf{R})}{K_{reaction}(\mathbf{R} = H)}\right) \quad or \quad \log_{10}\left(\frac{k_{reaction}(\mathbf{R})}{k_{reaction}(\mathbf{R} = H)}\right)$$

Fundamental hypothesis: acidity, $K_a(\mathbf{R})$ is directly proportional to electron-withdrawing ability of R

Hammett found that 33/38 reactions analyzed followed the proposed linear relationship

different sensitivies to electronic effects
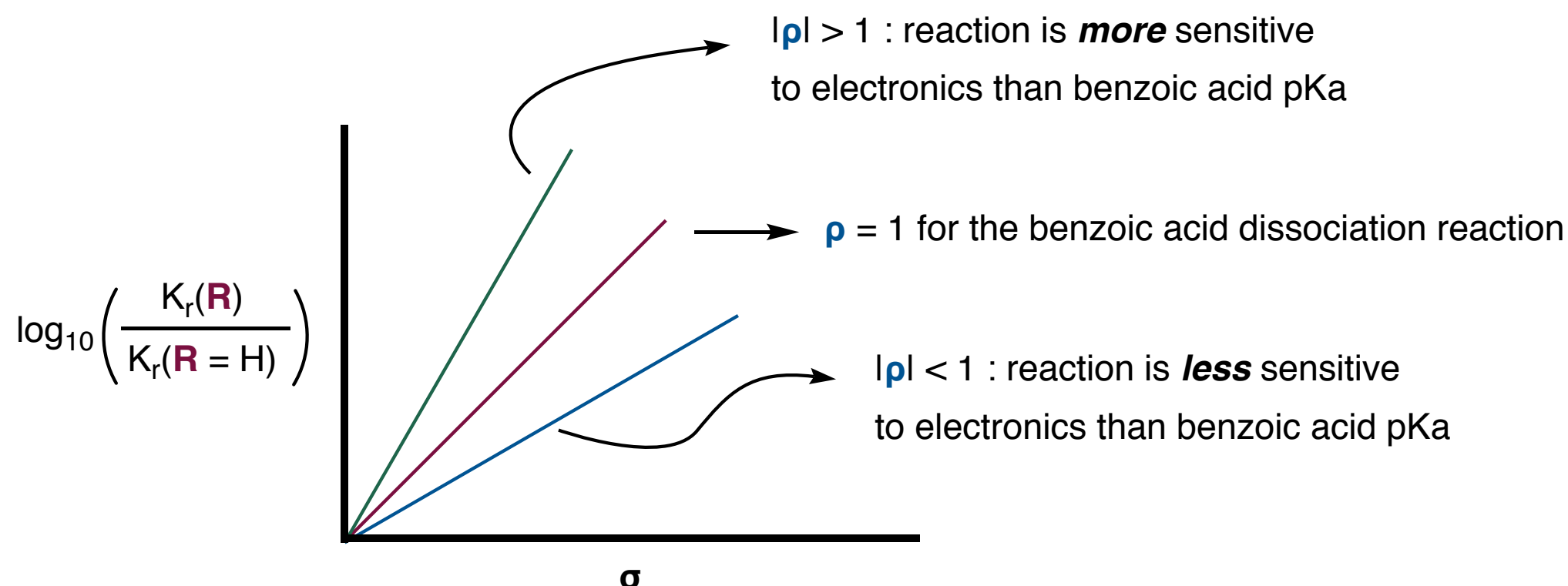
different slopes ($\boldsymbol{\rho}$)

"linear" free energy relationship: $\Delta G° = -2.303 \cdot RT \cdot \log_{10}(\mathbf{K})$

Hammett, L. P. *J. Am. Chem. Soc.* **1937**, *59*, 96–103.

The Hammett Equation: Interpreting Hammet Plots

- Hammett plots can reveal magnitude *and* nature of electronic influences

|ρ| > 1 : reaction is ***more*** sensitive
to electronics than benzoic acid pKa

ρ = 1 for the benzoic acid dissociation reaction

$$\log_{10}\left(\frac{K_r(\mathbf{R})}{K_r(\mathbf{R} = H)}\right)$$

|ρ| < 1 : reaction is ***less*** sensitive
to electronics than benzoic acid pKa

σ

$$\log_{10}\left(\frac{K_{reaction}(\mathbf{R})}{K_{reaction}(\mathbf{R} = H)}\right) = \sigma\boldsymbol{\rho} + C$$

slope

ρ < 0 : rate or equilibrium is inversely related
to electron-withdrawing ability of group

- sign of **ρ** reveals nature of electronic influence on reaction of interest
- absolute value of **ρ** reveals magnitude of electronic influence on reaction

Hammett, L. P. *J. Am. Chem. Soc.* **1937**, *59*, 96–103.

Early Anomalies with the Hammett Parameters

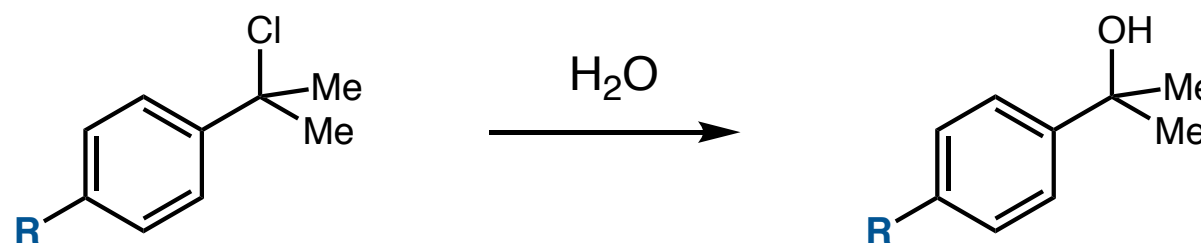■ One set of Hammett parameters does not cover all reactions - new parameters developed, 1958

Br$_2$, AcOH

newly defined σ$^+$ values correlate
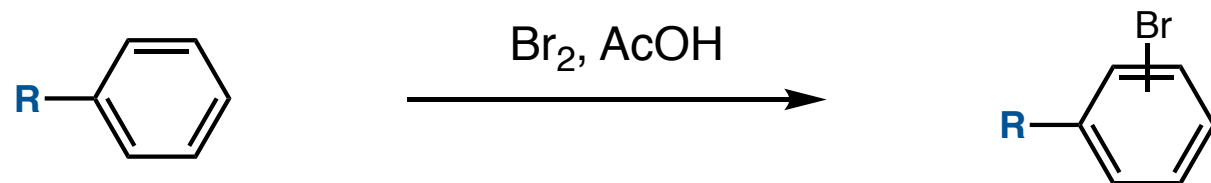
traditional σ values fail to correlate
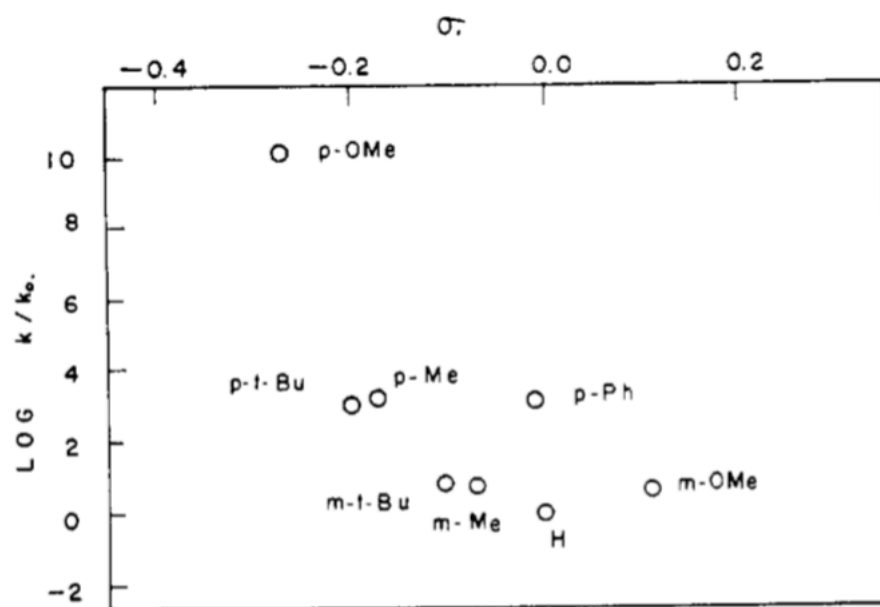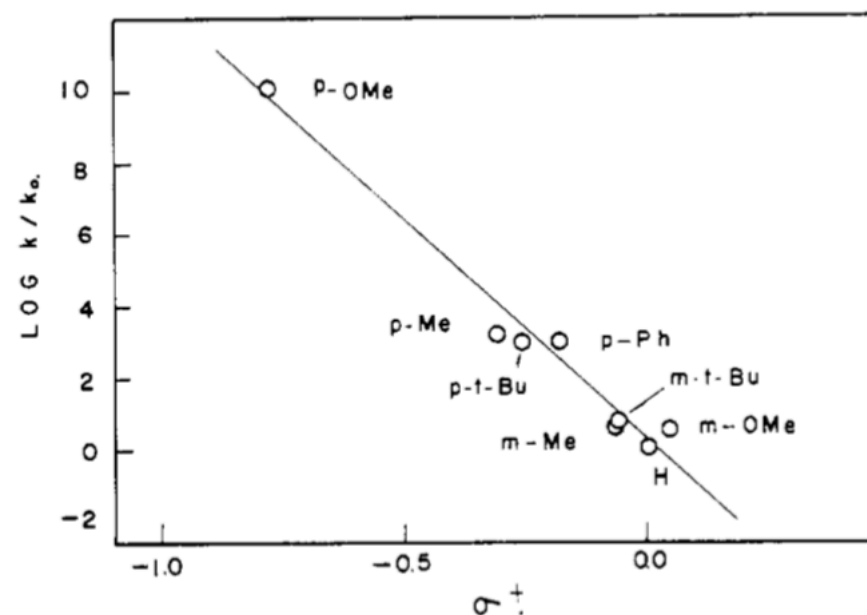
σ$^+$ values origin: *t-cumyl chloride solvolysis rates*

H$_2$O

Brown, H. C.; Okamoto, Y. *J. Am. Chem. Soc.* **1958**, *80*, 4979–4987.

Development of Additional Electronic Parameters

■ More electronic parameters are continually developed, *through 2018*

σ^+ *(hydrolysis rate)*

Brown, 1958

σ^- *(pK$_a$ values)*

Brown, 1963

σ' *(pK$_a$ values)*

Swain, 1968

σ_I *(pK$_a$ values)*

Grob, 1983
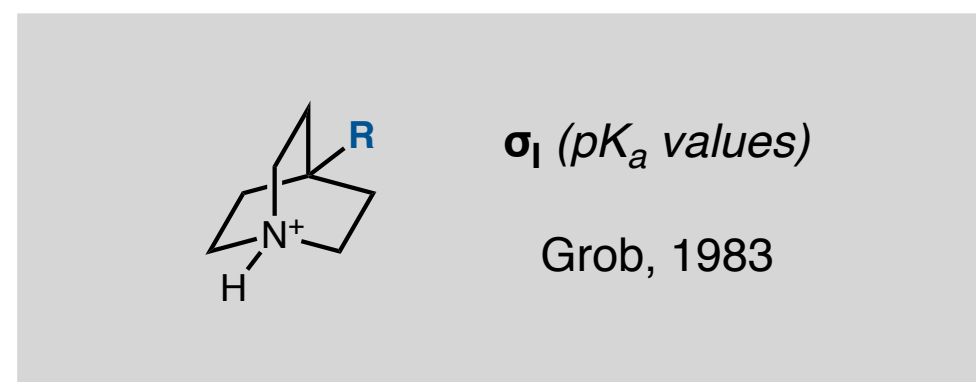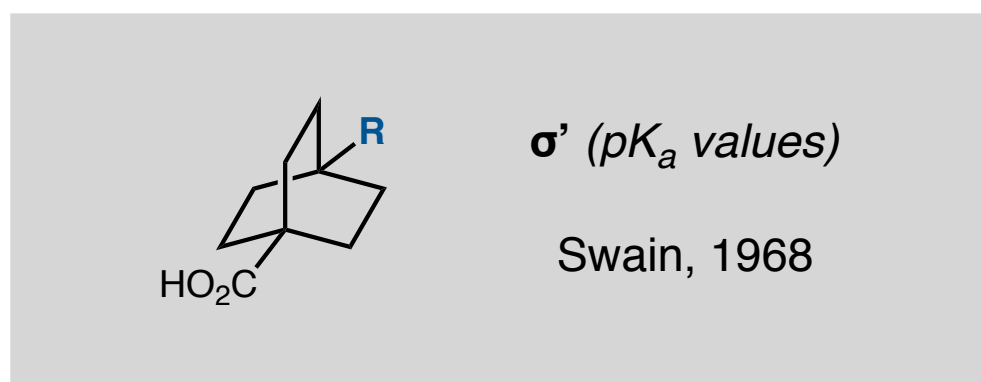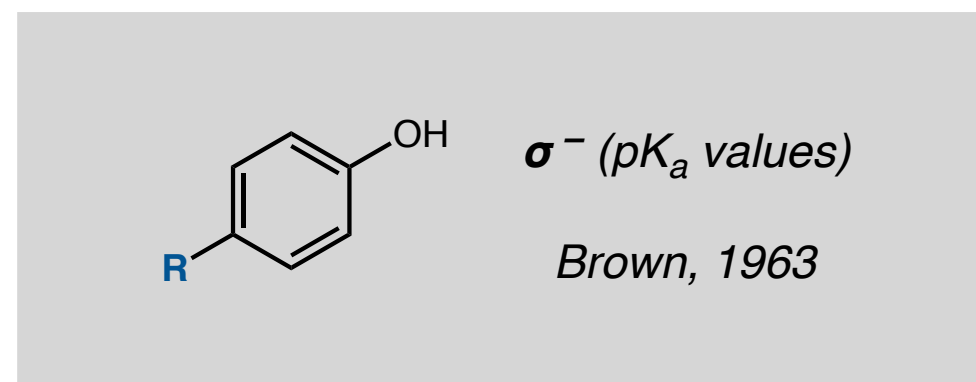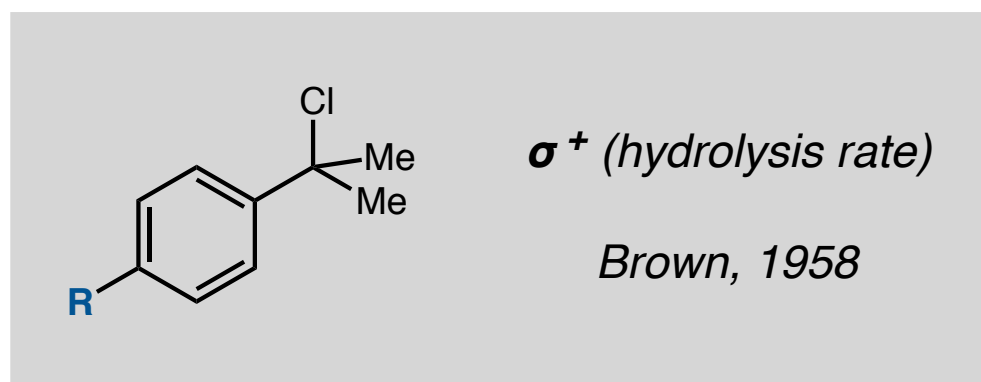
■ Lupton and Swain split **σ** into a linear combination of resonance and polar effects, 1968

$$\sigma = f \cdot \mathscr{F} + r \cdot \mathscr{R} \longrightarrow$$ *requires fitting binomial*

■ DFT is heavily utilized in determining electronic effects today, *continually developed*

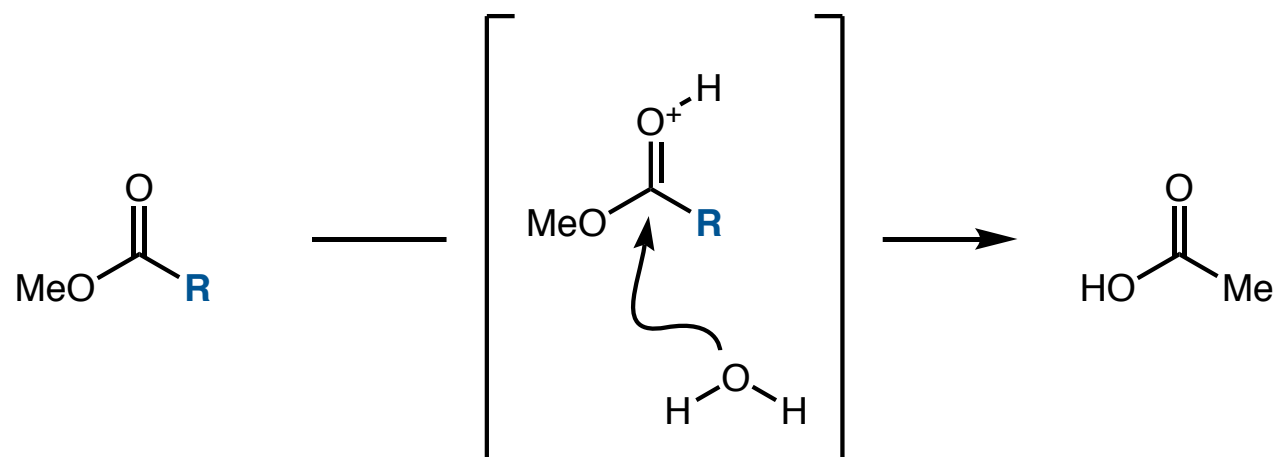Brown, H. C.; Okamoto, Y. *J. Am. Chem. Soc.* **1958**, *80*, 4979–4987.

Brown, H. C.; Stock, L. M. *Adv. Phys. Org. Chem.* **1963**, *1*, 35

Swain, C. G.; Lupton, E. C. *J. Am. Chem. Soc.* **1968**, *90*, 4328–4337.

Grob, C. A. *Acc. Chem. Res.* **1983**, *16*, 426–431.

Taft Parameters and Subsequent Advancements

■ Taft proposes ester hydrolysis rate measurements to assess steric size, 1952

■ Charton calls into question Taft's σ*, replaces E_s with ν, 1975

Taft parameters

$$E_s = \ln \left(\frac{k_{hydrolysis}(R)}{k_{hydrolysis}(R = H)} \right)$$

(steric parameter)

$$\sigma^* = \left(\frac{E_{s(acidic)} - E_{s(basic)}}{2.48} \right)$$

(electronic parameter)

Taft's **σ*** is rarely used today, but E_s remains

Charton's ν is commonly used as a steric parameter

Taft, R. W. *J. Am. Chem. Soc.* **1952**, *74*, 2729–2732.

Charton, M. *J. Am. Chem. Soc.* **1975**, *97*, 1552–1556.

Taft Parameters and Subsequent Advancements

■ Winstein and Holness establish A-values, 1955

■ Tolman defines the ligand cone angle (θ) to explain anomalous metal binding results, 1970

■ Verloop defines computationally derived sterimol parameters for medicinal chemistry, 1976

Sterimol Parameters

L: length of longest bond axis

B1: minimum perpendicular radius

B5: maximum perpendicular radius

commonly used in both drug design and catalytic applications
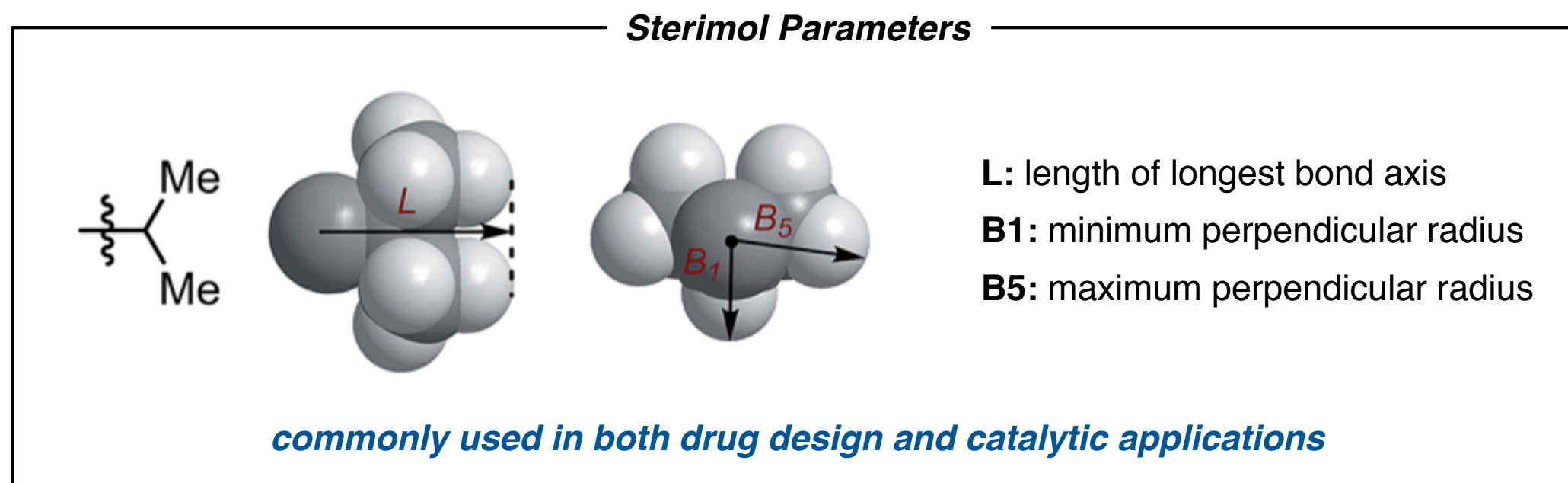
■ Computational methods to account for conformational flexibility are still being developed

A. Verloop, in Drug Design, Academic Press, New York, **1976**.

C. A. Tolman, Chem. Rev., **1977**, 77, 313.

Winstein, S.; Holness, N. J. *J. Am. Chem. Soc.* **1955**, *77*, 5562–5578.

Beyond Electronic and Steric Parameters: Generalization of LFERs

reaction outcome = **function** of measurable (*or computable*) **molecular properties**

Y-axis

- % yield
- rate
- % ee
- d.r.
- r.r.

functions

only limited by

physical meaning

- linear (y = mx + b)
- polynomial (X^n)
- logarithmic (ln(x))
- exponential (e^x)
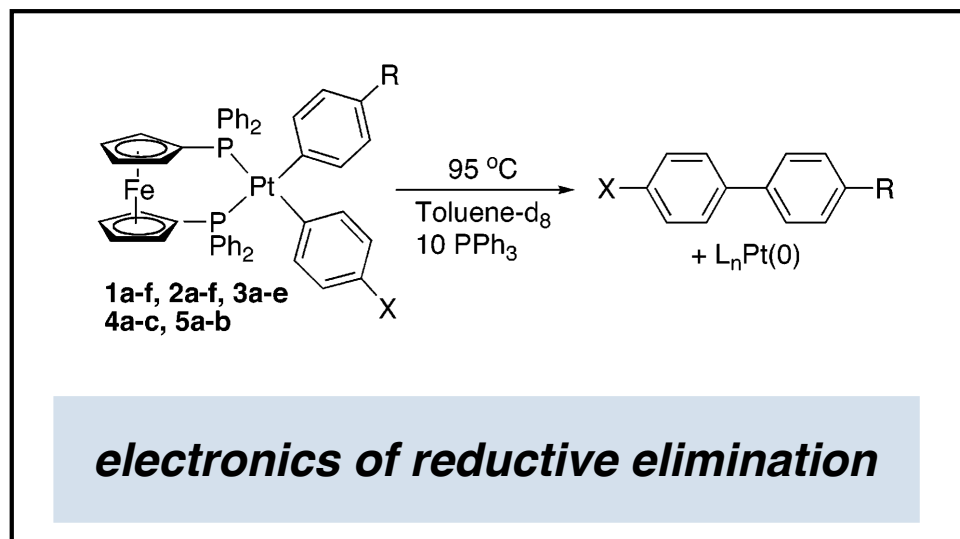
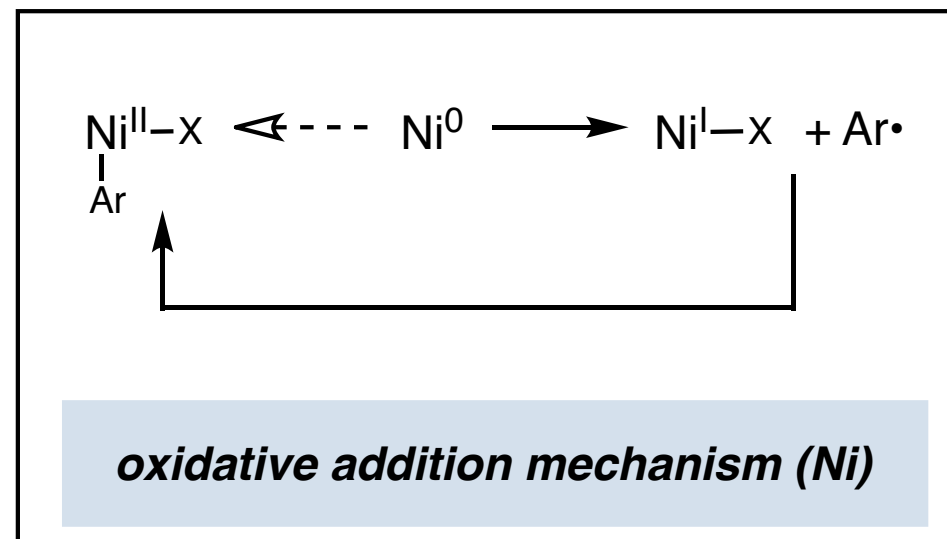any linear combination

of the above functions

and more…

X-axis

- steric parameters
- electronic parameters
- calculated HOMO
- calculated LUMO
- NMR shift values
- vibrational frequency
- ligand bite angle
- reduction potential
- oxidation potential
- triplet energy
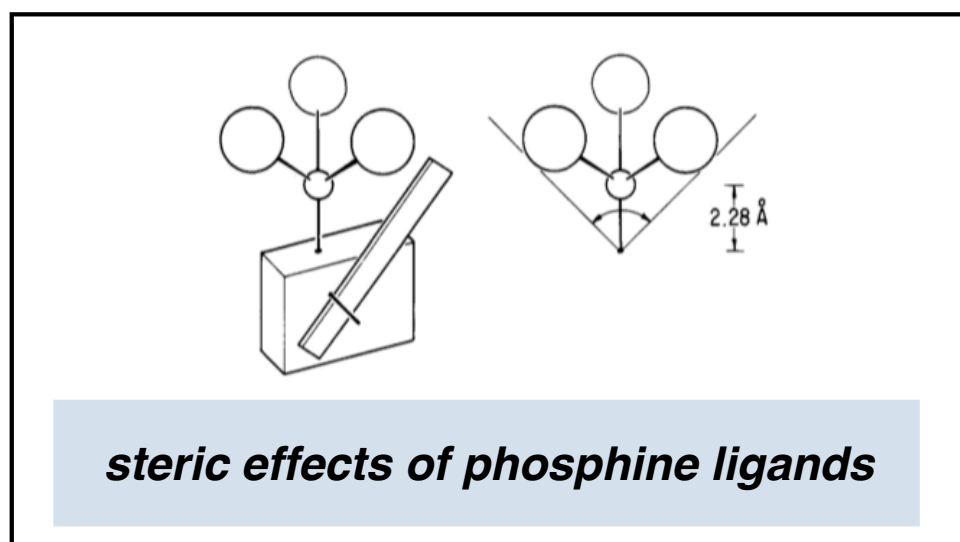- bond energy
- pK_a values
- ring strain

B. Santiago, C.; Guo, J.-Y.; S. Sigman, M. *Chem. Sci.* **2018**, *9*, 2398–2412.

Utility of single-parameter LFERs in Organometallic Catalysis

electronics of reductive elimination

Shekhar, S.; Hartwig, J. F. *J. Am. Chem. Soc.* **2004**, *126*, 13016–13027.

oxidative addition mechanism (Ni)

Tsou, T. T.; Kochi, J. K. *J. Am. Chem. Soc.* **1979**, *101*, 6319–6332.

steric effects of phosphine ligands

Tolman, C. A. *Chem. Rev.* **1977**, *77*, 313–348.

asymmetric epoxidation mechanism
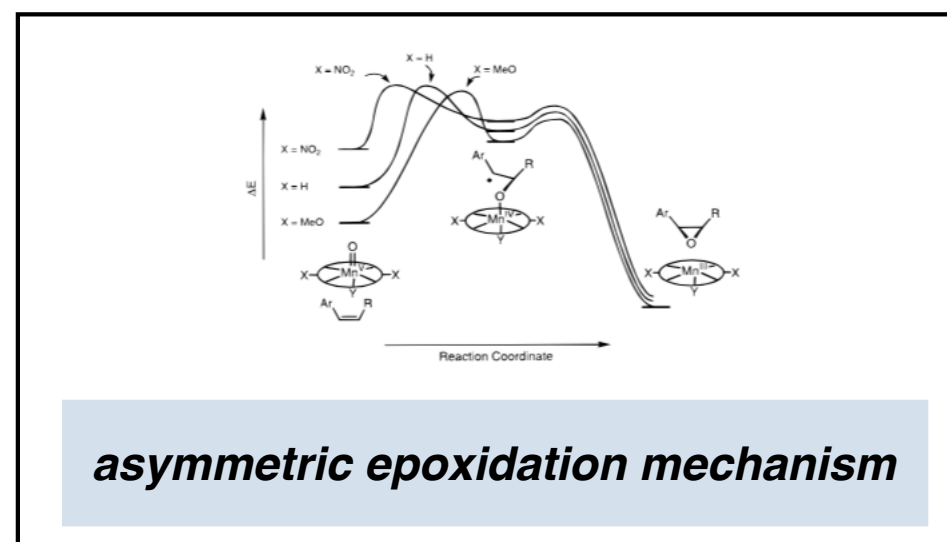
Jacobsen, E. N. and coworkers *J. Am. Chem. Soc.* **1998**, *120*, 948–954.

1937 — *Hammett values*

1962 — Hansch analysis

2010 — *Sigman parameterization*
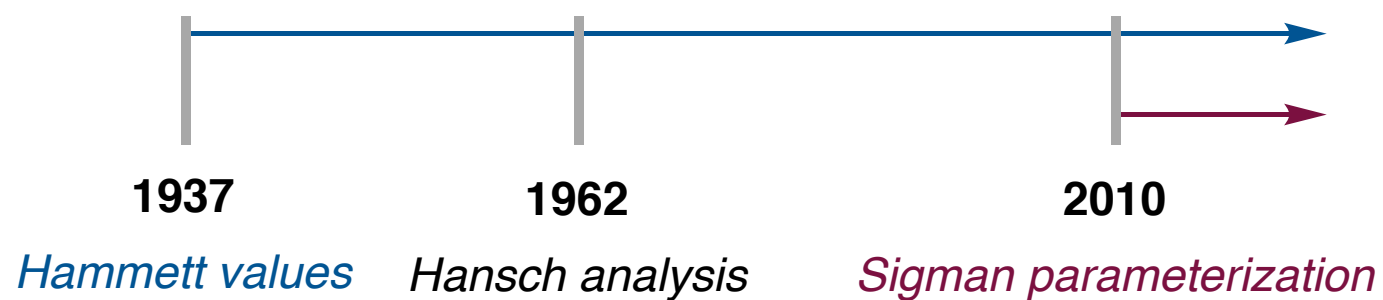
Development of Additional Electronic Parameters

■ Corwin Hansch finds a model to predict pharmacological activity with MLR, 1962
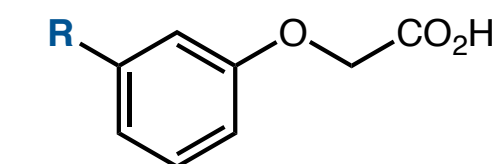
R—C₆H₄—O—CH₂—CO₂H structure → *qualitative observations:* electronic effects matter → σ values alone cannot predict activity

promotes plant growth

compound uptake by plant is necessary and related to **P** → $1/C = 4.08\pi - 2.14\pi^2 + 2.78\sigma + 3.36$ $\quad \pi = \log(P) - \log(P_H)$

successfully models plant growth activity

C: concentration of compound that induces 10% growth of plant (high potency if 1/C is high)

P is the octanol/water partition coefficient

■ Fitting the polynomial carried out on a Clary *DE*-60 computer (about the size of a desk)

■ QSAR continues to exist as a mainstay of drug discovery

Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. M. *Nature* **1962**, *194*, 178–180.

Intuition behind MLR: Simplest Case

■ Simplest case: finding a model that predicts m/z for identically charged molecules
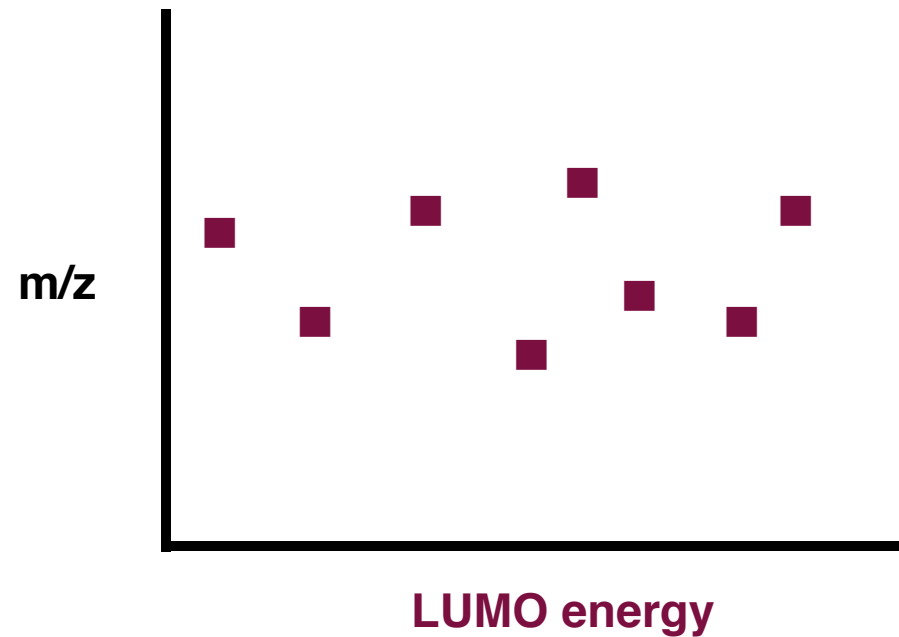
*what determines **m/z***

in mass spectrometer?
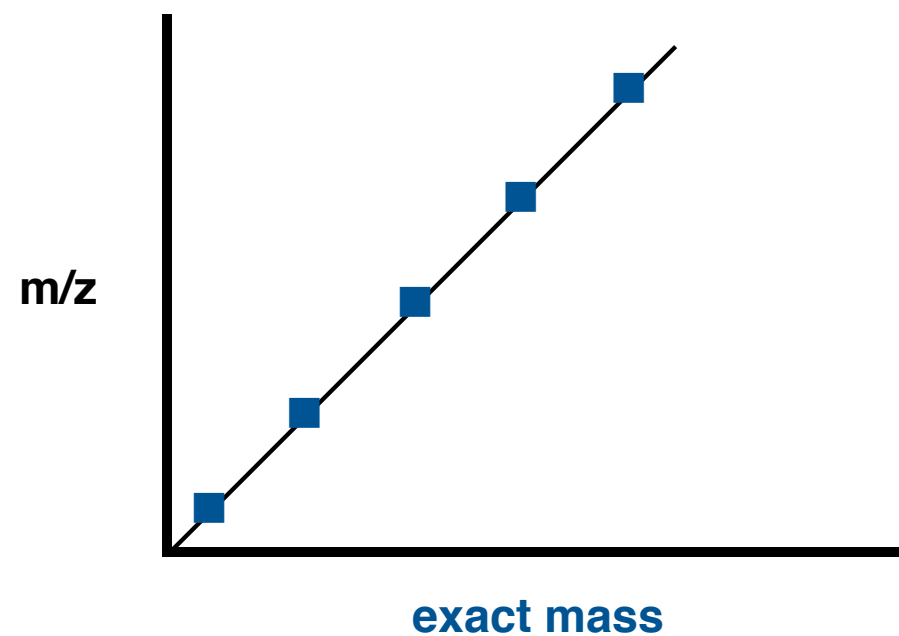
┌ known or measurable ┐

- molecular weight
- molecular formula
- net charge
- **exact mass**
- electronegativities
- calculated HOMO
- **calculated LUMO**
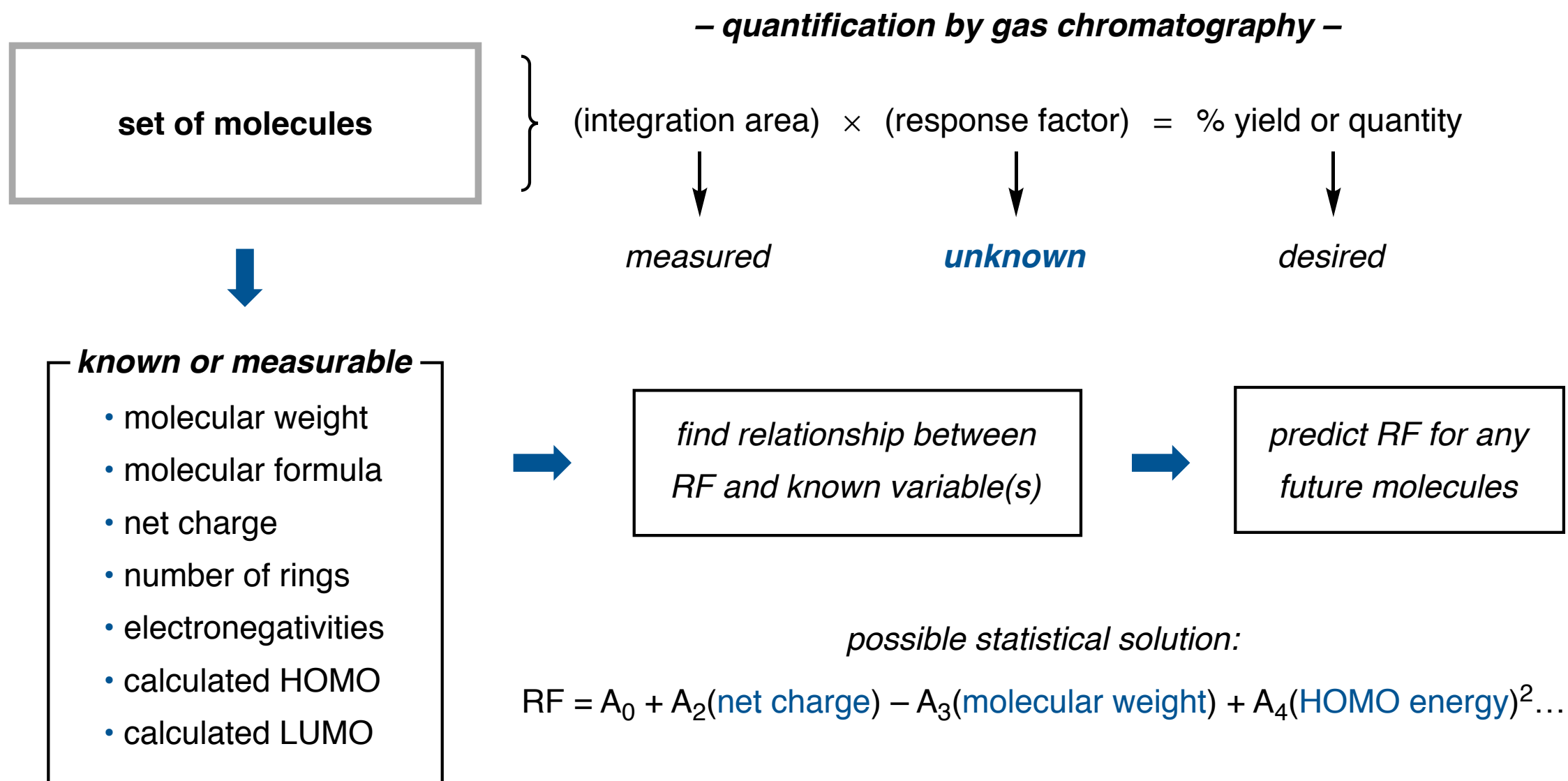
⇩
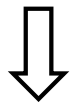
pick a parameter
*and plot against **m/z***

m/z

LUMO energy

⇨ *poor correlation*

m/z

exact mass

⇨ *perfect correlation*

model: *m/z = exact mass + C*

Intuition Behind MLR in QSAR and Sigman Parameterization

set of molecules

– quantification by gas chromatography –

(integration area) $\times$ (response factor) = % yield or quantity

measured   ***unknown***   *desired*

known or measurable

- molecular weight
- molecular formula
- net charge
- number of rings
- electronegativities
- calculated HOMO
- calculated LUMO

find relationship between RF and known variable(s)

predict RF for any future molecules

possible statistical solution:

$$RF = A_0 + A_2(\text{net charge}) - A_3(\text{molecular weight}) + A_4(\text{HOMO energy})^2\ldots$$

Fundamental principle: an organic structure contains all information needed to predict properties

Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *Chem. Soc. Rev.* **1995**, *24*, 279–287.

System Theory Definition of Catalytic Reactions

	inputs	*transforms*	*outputs*

System Theory

x = 1 ——

$$y = x + 2$$

——→ y = 3

Chemical Example

chemical
structure ——

reaction mechanism

——→ reaction
outcome

↓   ↓   ↓

known properties   **model as LFER**   **measure to train
then predict**
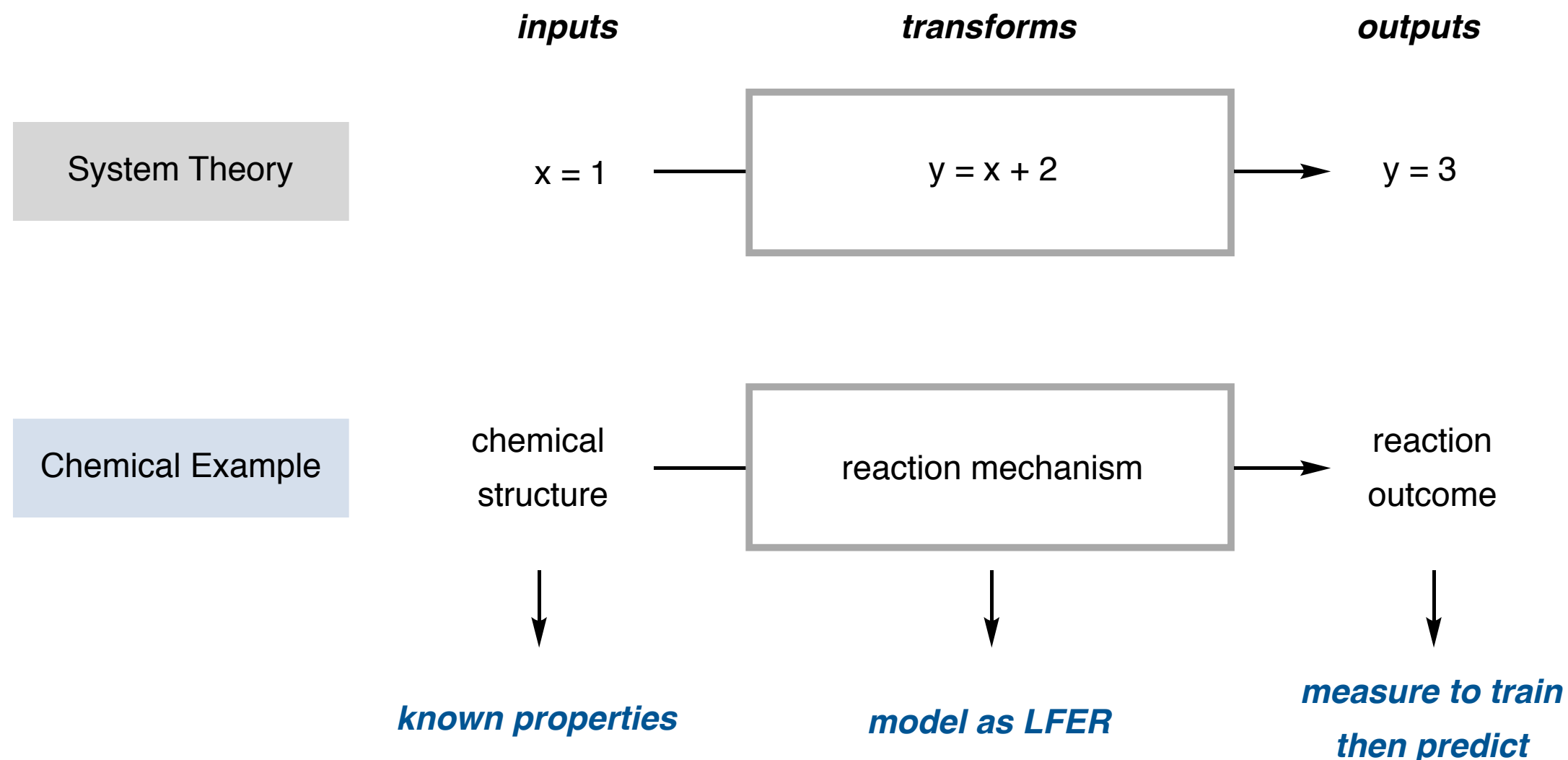
System theory provides a well-established framework for model development

Deming SN, Morgan SL (1993) Experimental design: a chemometric approach; data handling in science and technology, (Elsevier, Amsterdam-London-New York-Tokyo), 11

Mathematics of Multiple Linear Regression

parameter *j* for observation *i*

(σ for catalyst 5 for example)

multiple linear regression

in summation notation

$$y_i = b_0 + \sum_{j=1}^{p} b_j x_{ij} + e_i$$

error term

(distance of y$_i$ from fit)

yield or % ee

for catalyst *i*

regression slope

for parameter *j*

formulate X and

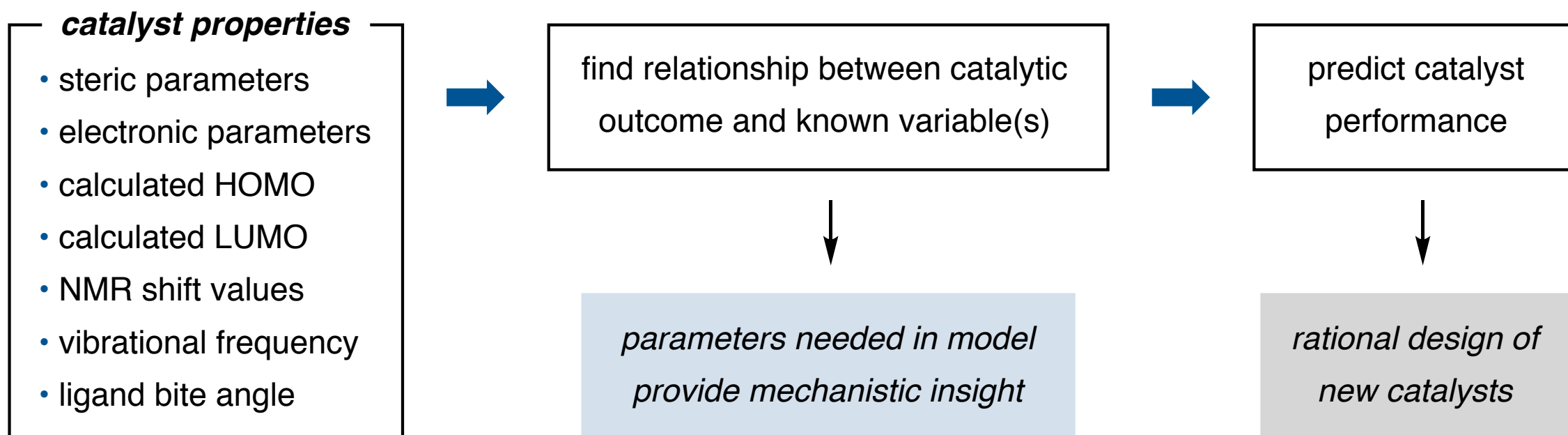**Y as matrices**   ➡   **B = (X^{T}X)$^{-1}$(X^{T}Y)**   ➡   *solve this in Matlab*

B: matrix of parameter coefficients   **X:** model matrix   **Y:** response matrix

Rationale behind the Application of MLR in Catalysis

catalyst properties

- steric parameters
- electronic parameters
- calculated HOMO
- calculated LUMO
- NMR shift values
- vibrational frequency
- ligand bite angle

→ find relationship between catalytic outcome and known variable(s)

→ predict catalyst performance

parameters needed in model provide mechanistic insight

rational design of new catalysts

■ utilizing an extensive set of catalyst properties to build a performace model maximizes predictive power

■ if catalyst properties chosen for parameterization are physically meaningful, mechanistic insight can be gained

specifically in enantioselective catalysis:

"The field of asymmetric catalysis has come to recognize that the accumulation of weak, noncovalent interactions is critical in a myriad of enantioselective reactions."

Seminal Work in Utilizing MLR for Asymmetric Catalysis

■ Kaid Harper and Matt Sigman publish parameterization of NHK allylation, 2011

10 mol% CrCl$_3$, 11 mol% L*

10 mol% Et$_3$N

TMSCl, Mn^0, THF

chiral ligand
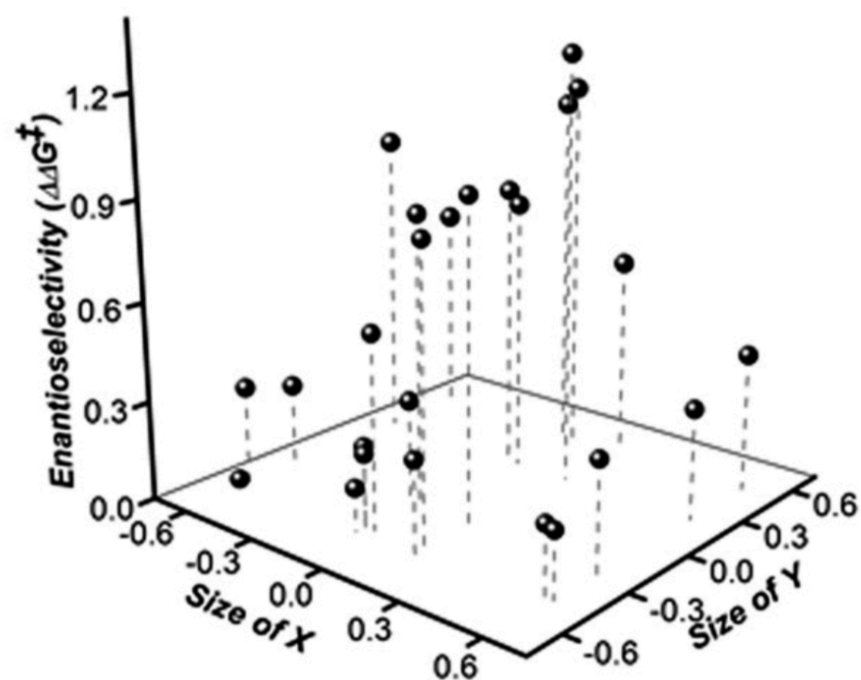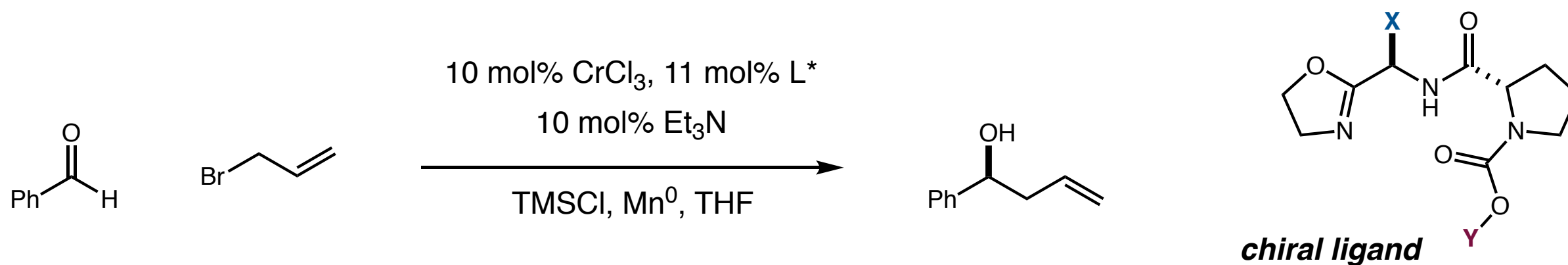
library of 25 ligands evaluated

*Charton values of **X** and **Y** plotted against e.r. (ΔΔG‡)*

raw data (no fitting) with translated Charton values

Harper, K. C.; Sigman, M. S. *PNAS* **2011**, *108*, 2179-2183.

Seminal Work in Utilizing MLR for Asymmetric Catalysis

■ Fitting Charton values of X and Y with multiple linear regression yields surface

take the **X = − 0.1**
(Methyl) slice

$R^2 = 0.99$
$\Delta\Delta G^{\ddagger} = 0.927Y + 0.623$
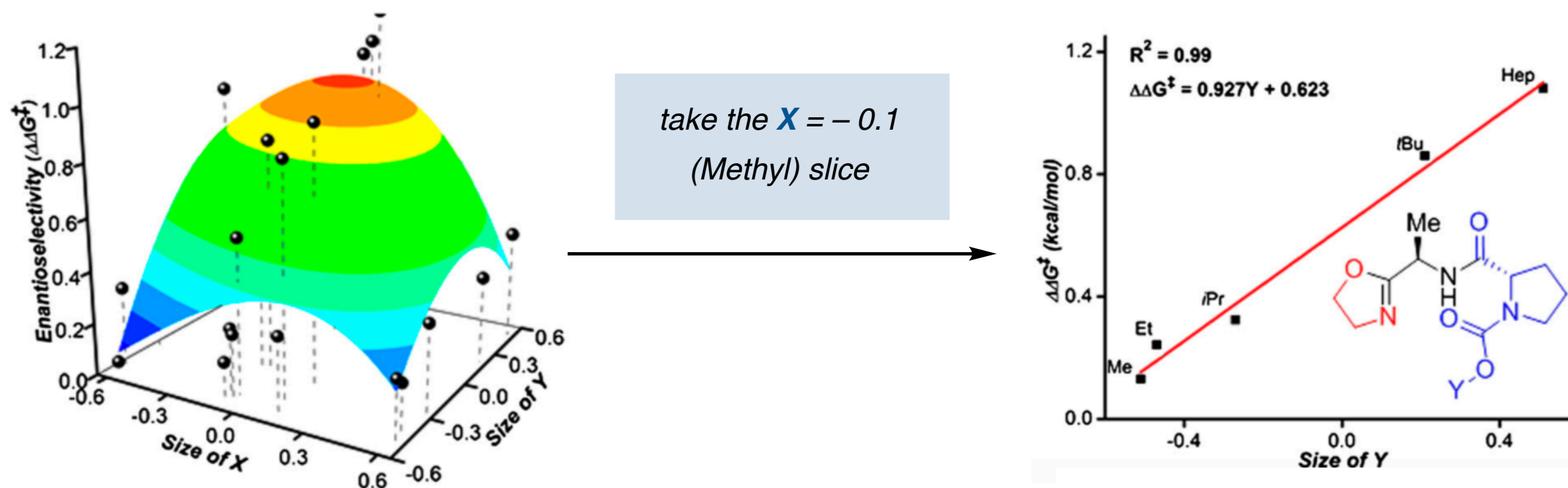
raw data (black balls) + fitted surface (overlaid)

MLR-fitted model: $\Delta\Delta G^{\ddagger} = 0.931 + 0.576Y - 0.905X^2 - 1.005Y^2 - 0.502XY - 0.407X^3 - 0.475YX^2$

3D surface model
vs.
simple linear fit

X	Predicted er		Error	Measured er
	Linear	3D		
H	NA	40.8:59.2	±3	42.5:57.5
Me	3:97	46.7:53.3	±4	36:64
iPr	NA	38.6:61.3	±3	40:60

Harper, K. C.; Sigman, M. S. *PNAS* **2011**, *108*, 2179-2183.
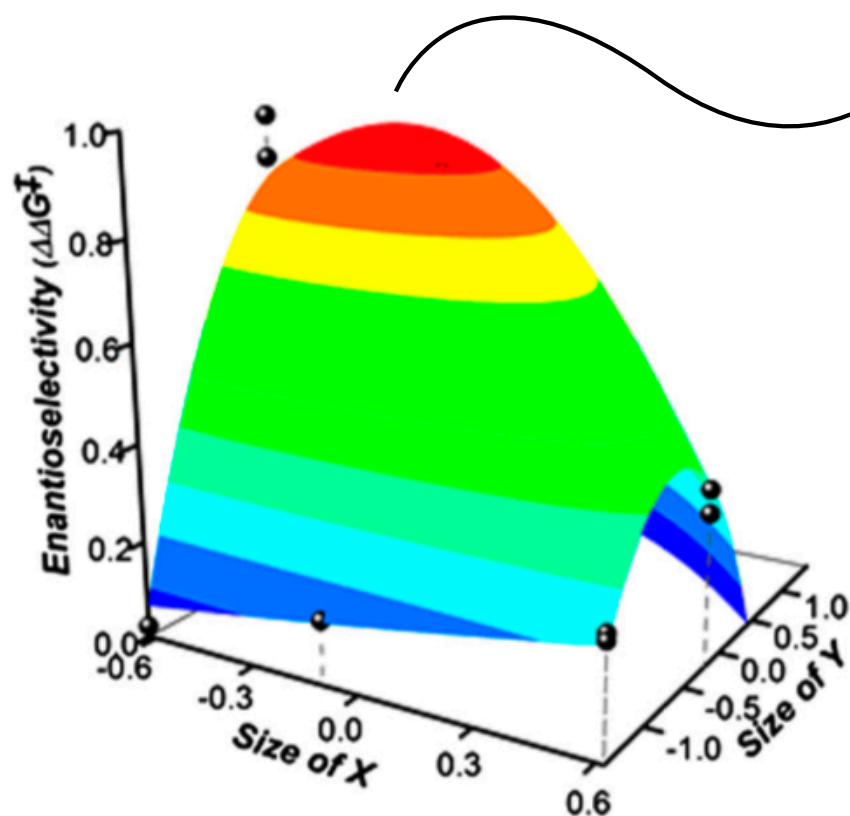
Seminal Work in Utilizing MLR for Asymmetric Catalysis

■ Paring down the ligand library to only 9 ligands still predicts optimal catalyst
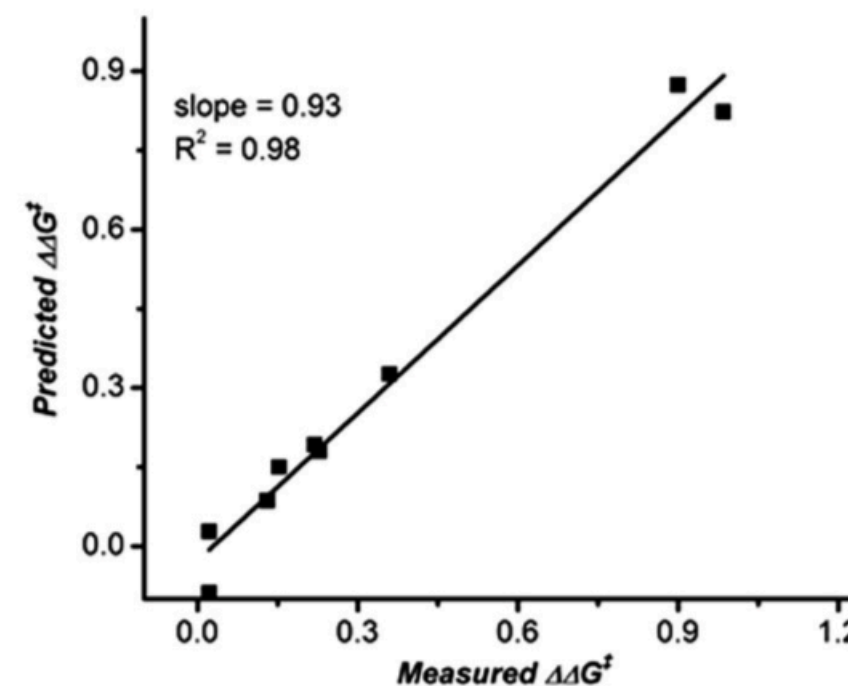
what if you don't have
a library of 25 ligands?

predicted
optimal ligands:
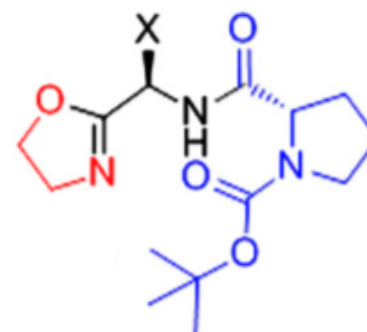
X	Predicted er	Error	Measured er
Et	16:84	±2	9.5:90.5
iPr	17:83	±3	8:92

slope = 0.93
$R^2 = 0.98$

3 x 3 matrix of ligands
(optimal ligands omitted)

measured vs. predicted plot judges model

Harper, K. C.; Sigman, M. S. *PNAS* **2011**, *108*, 2179-2183.

Extension of Multiple Linear Regression to Electronic Parameters

- Examining only steric parameters leads to a modest maximum e.r.

$$10 \text{ mol\% } CrCl_3, \ 11 \text{ mol\% } L^*$$

$$20 \text{ mol\% } Et_3N$$

TMSCl, Mn^0, THF

chiral ligand
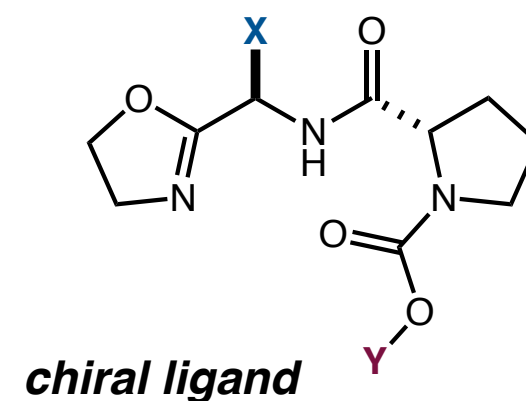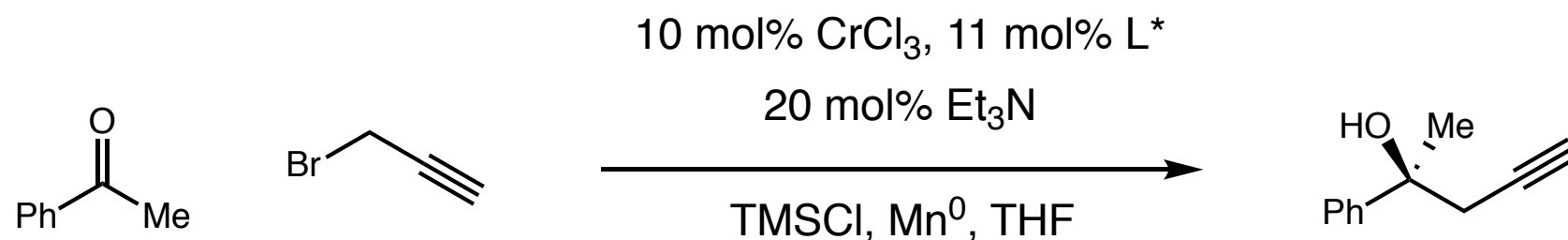
local maximum reached:

X = Me; Y = tBu
er = 75:25
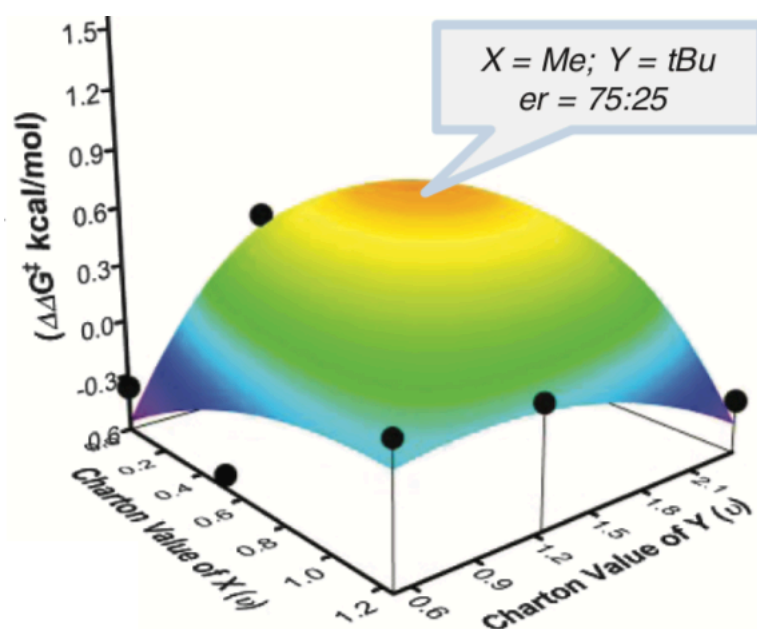
can a new ligand class surpass 50% ee?
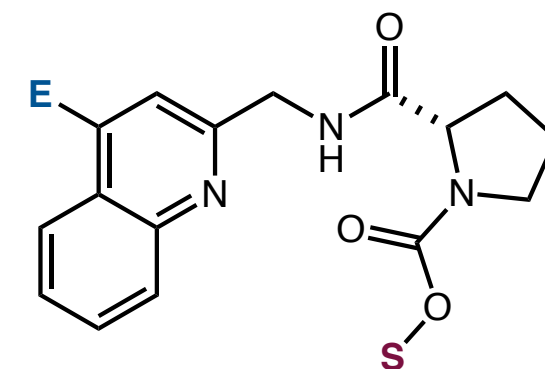
E: electronic parameter tuning   **S:** steric parameter tuning ⟹ **comprehensive model?**

Harper, K. C.; Sigman, M. S. *PNAS* **2011**, *108*, 2179-2183.

Extension of Multiple Linear Regression to Electronic Parameters

■ Changing ligand class provides route to new local maximum

chiral ligand

Optimal Ligand

6b

er = 92:8
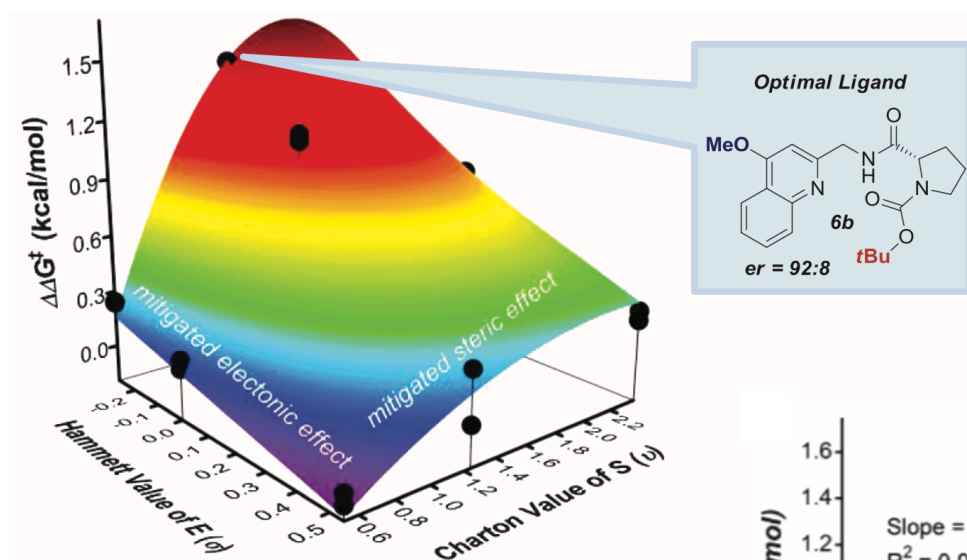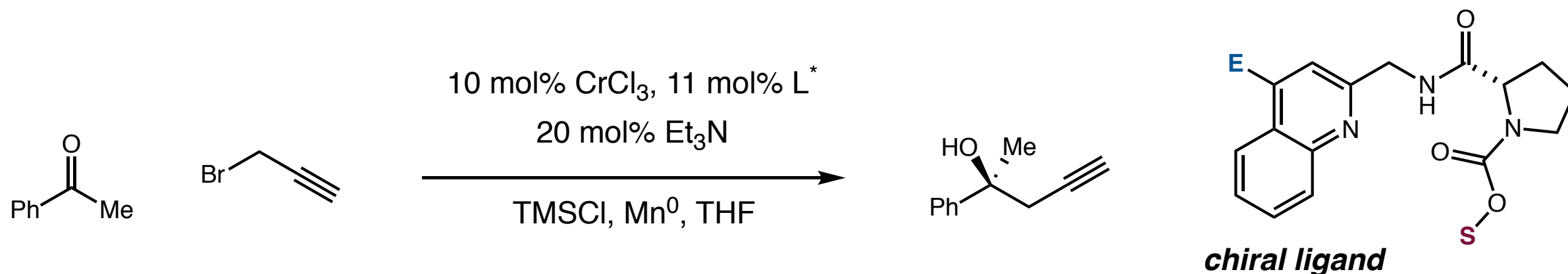
predicted vs. measured
plot shows good fit

Slope = 1.02
$R^2 = 0.96$

Predicted $\Delta\Delta G^{\ddagger}$ (kcal/mol)

Measured $\Delta\Delta G^{\ddagger}$ (kcal/mol)

E: isosteric slice shows increased % ee with more negative σ

S: isoelectronic slice shows increased % ee with steric size
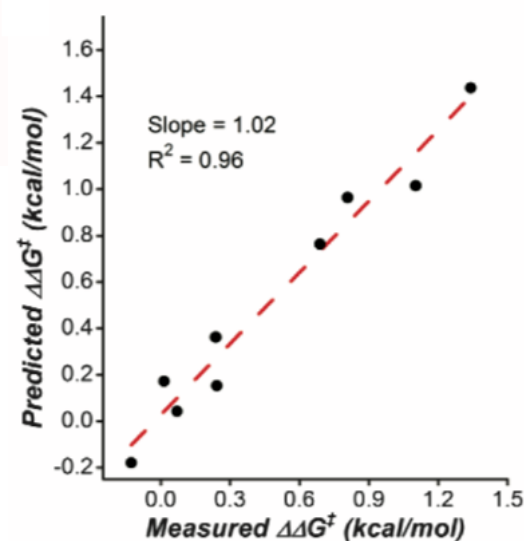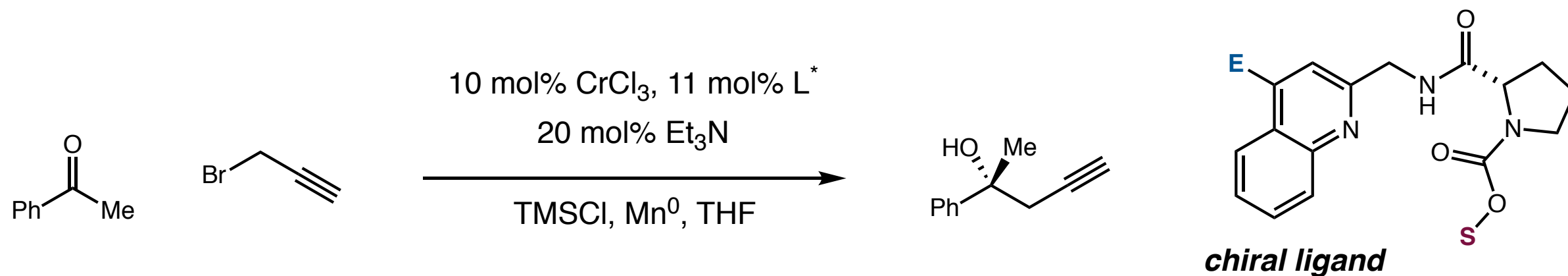
Harper, K. C.; Sigman, M. S. *Science* **2011**, *333*, 1875–1878.

Extension of Multiple Linear Regression to Electronic Parameters

■ Changing ligand class provides route to new local maximum

chiral ligand

*scope with the LFER–optimized ligand*

2a; 75% yield*
96:4 er[†]

2b; 74% yield
96:4 er

2c; 60% yield
92:8 er

2d; 74% yield
94:6 er

2e; 62% yield
96:4 er

2f; 72% yield
95:5 er

2g; 76% yield
86:14 er

2h; 80% yield
96:4 er

2i; 68% yield
95:5 er

2j; 68% yield
95:5 er

2k; 61% yield
93:7 er

2l; 70% yield
85:15 er

2m; 70% yield
85:15 er

2n; 62% yield
96:4 er

2o; 86% yield
98:2 er

2p; 72% yield
95:5 er

2q; 72% yield
92:8 er

Harper, K. C.; Sigman, M. S. *Science* **2011**, *333*, 1875–1878.
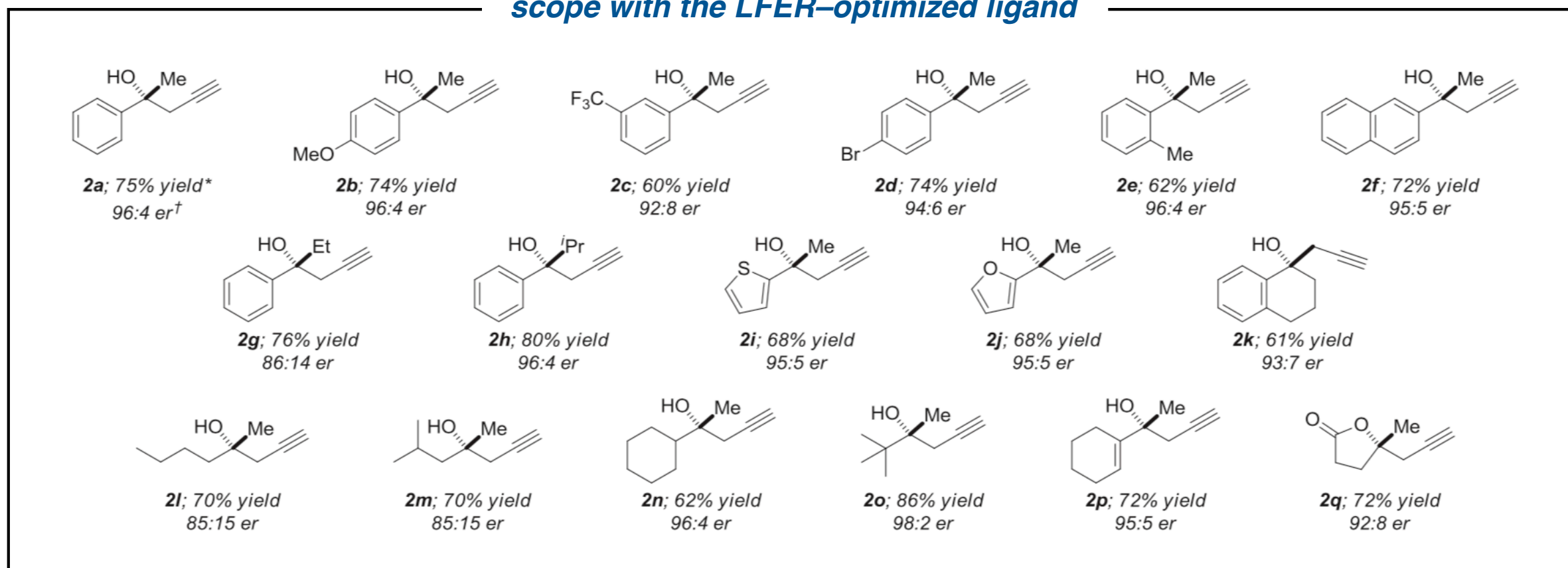
Predicting Site-Selectivity in C–H Oxidations

■ MLR LFER as a solution to predict C–H oxidation site selectivity
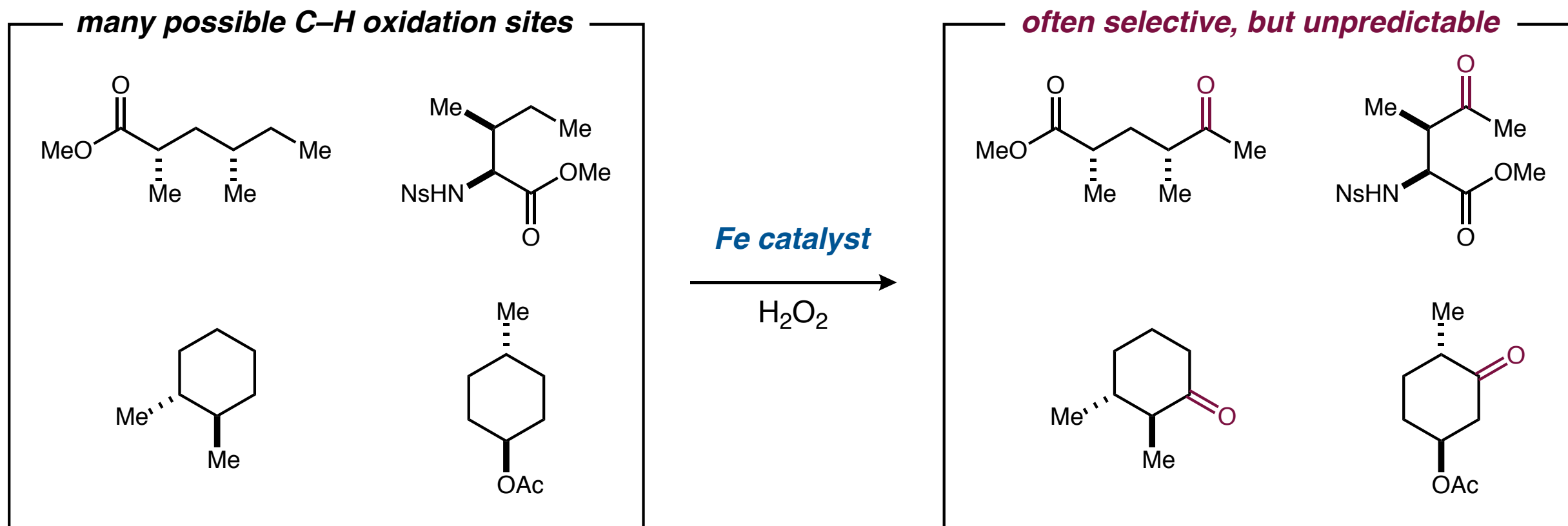
Gormisky, P. E.; White, M. C. *J. Am. Chem. Soc.* **2013**, *135*, 14052–14055.

Predicting Site-Selectivity in C–H Oxidations

■ Filtering process and model building for regioselectivity predictions

| **Full set of 25 substrates** | → | α–heteroatom C–H bonds, axial C–H bonds, and 1° C–H bonds removed | → | **set of 140 C–H sites** |
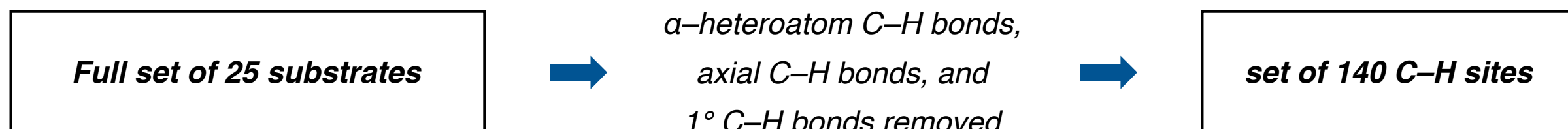
C–H bonds are assigned electronic parameters (**E**)

E: natural population analysis (NPA) calculated at B3LYP/6-311++G(d,p)

(a measure of partial charge)

C–H bonds are assigned steric parameters (**S**)

S: combined steric and stereoelectronic value

constructed from: adjusted A–value, gauche interactions, and a stereoelectronic parameter

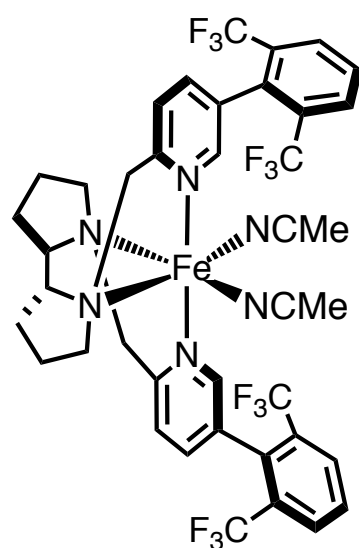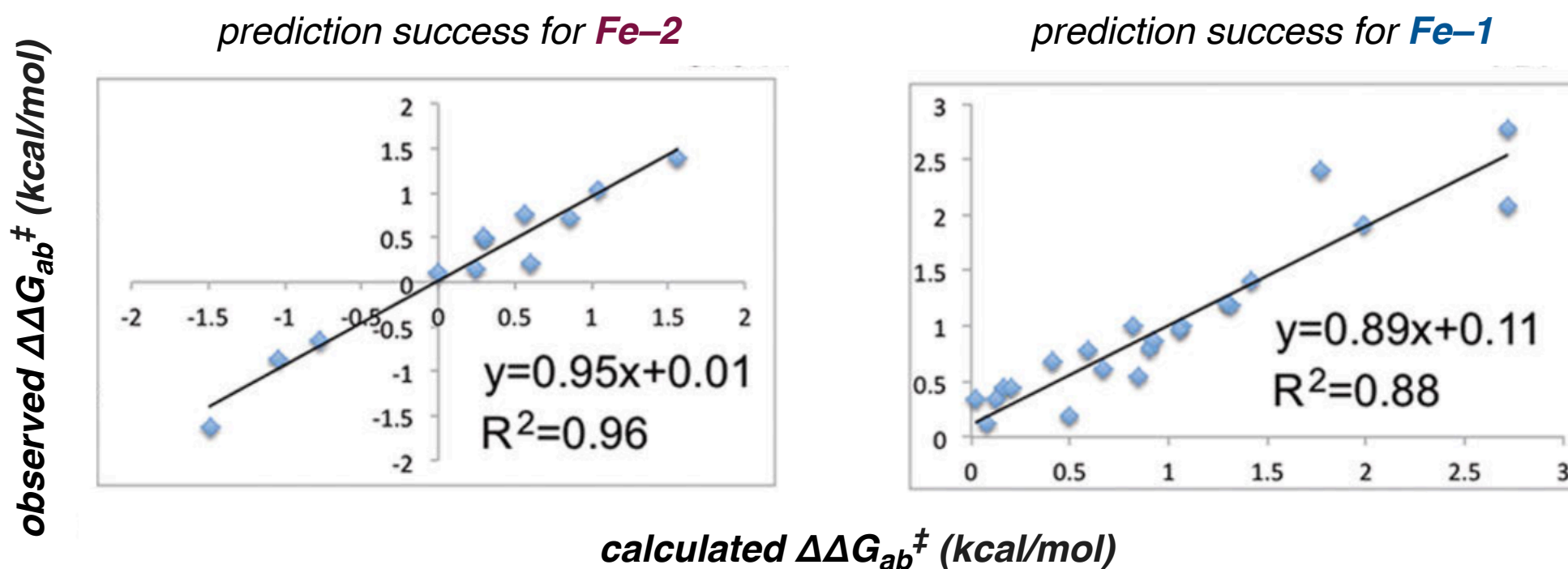| **E** and **S** values are classified as highly, moderately, or unreactive | → | 2 most reactive C–H sites taken |

plot the observed selectivity ($\Delta\Delta G^{\ddagger}$) values against calculated ΔE and ΔS values

↓

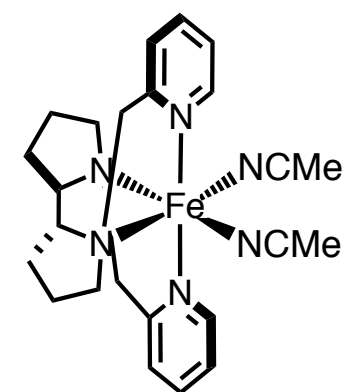$$\Delta\Delta G^{\ddagger} = 0.5 - 4.5 \cdot \Delta E - 1.3 \cdot \Delta S + 9.6 \cdot (\Delta E)^2 + 6.5 \Delta E \cdot \Delta S + 1.8(\Delta S)^2 - 2.7(\Delta E)^3 - 1.5(\Delta E)^2 \cdot \Delta S - 1.5 \Delta E \cdot (\Delta S)^2 - 0.4(\Delta S)^3$$

Gormisky, P. E.; White, M. C. *J. Am. Chem. Soc.* **2013**, *135*, 14052–14055.

Predicting Site-Selectivity in C–H Oxidations

■ Across an array of C–H oxidation substrates, both models correlate selectivity well

*prediction success for **Fe–2***

*prediction success for **Fe–1***

observed $\Delta\Delta G_{ab}^{\ddagger}$ (kcal/mol)

$y=0.95x+0.01$
$R^2=0.96$

$y=0.89x+0.11$
$R^2=0.88$

calculated $\Delta\Delta G_{ab}^{\ddagger}$ (kcal/mol)

Fe–2

Fe–1

Gormisky, P. E.; White, M. C. *J. Am. Chem. Soc.* **2013**, *135*, 14052–14055.
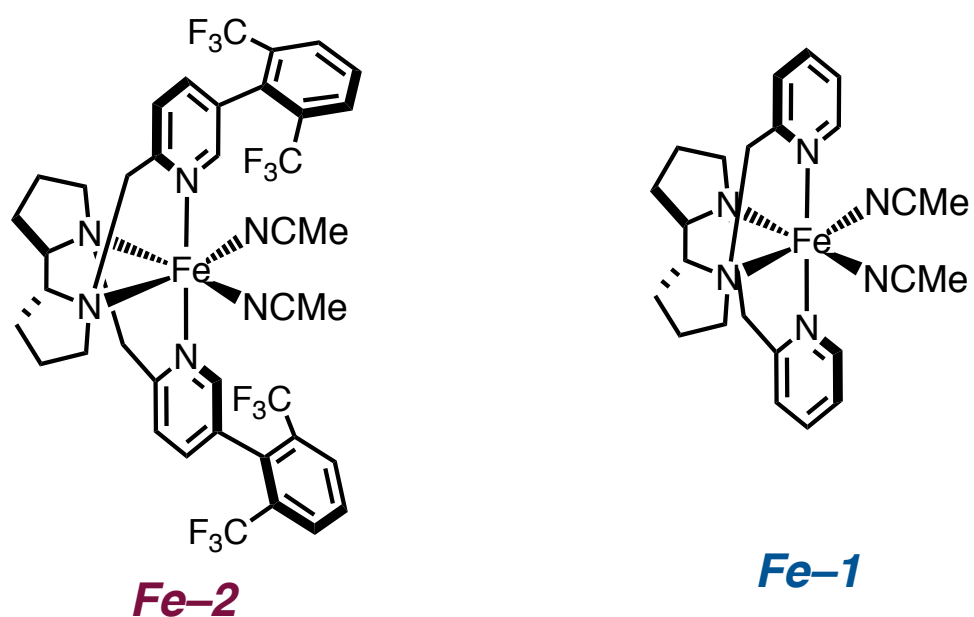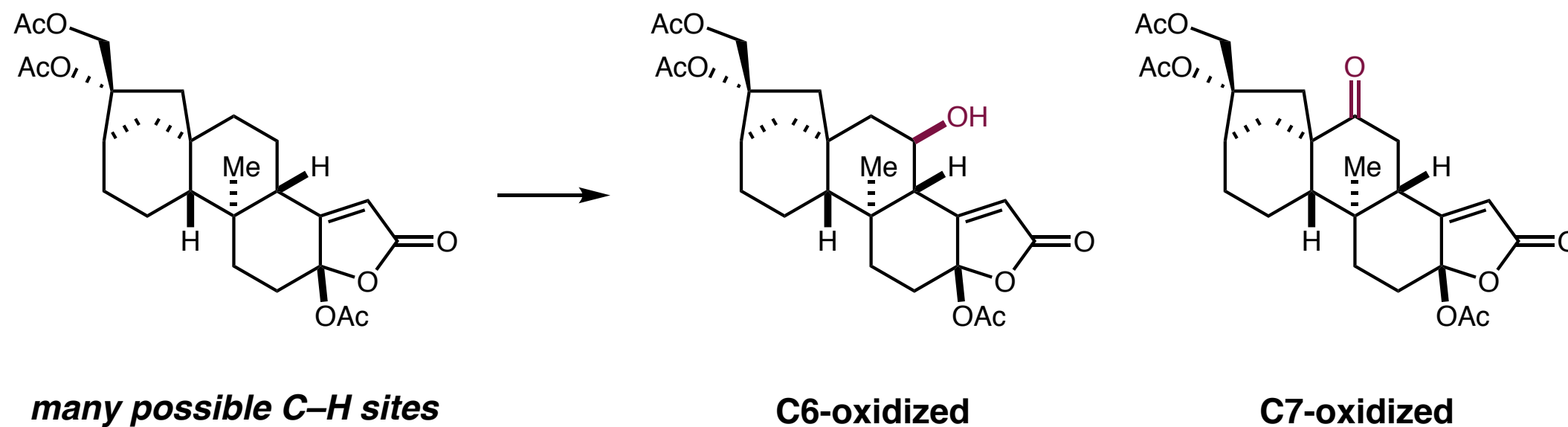
Predicting Site-Selectivity in C–H Oxidations

- White and coworkers developed an empirical model to predict sites of C–H oxidation

many possible C–H sites → **C6-oxidized** **C7-oxidized**

Fe–2

Fe–1

Fe catalyst	C6:C7 (predicted)	C6:C7 (observed)
Fe–1	1:1	1:1
Fe–2	11:1	>10:1

Gormisky, P. E.; White, M. C. *J. Am. Chem. Soc.* **2013**, *135*, 14052–14055.

Predicting Site-Selectivity in C–H Oxidations

■ Visual inspection of LFER surfaces quickly reveals catalyst performance

*selectivity surface for **Fe–2***

Fe–2
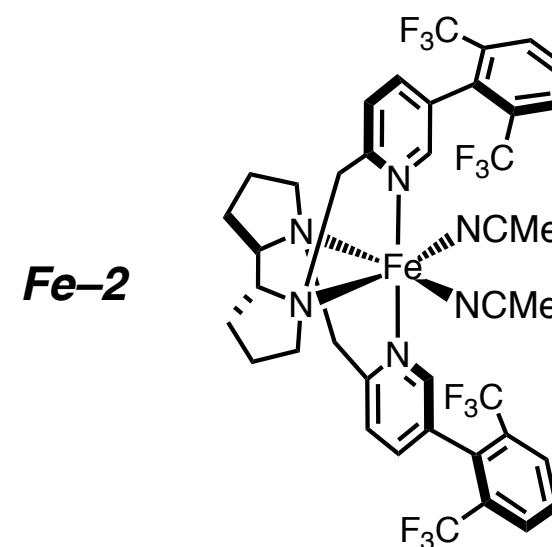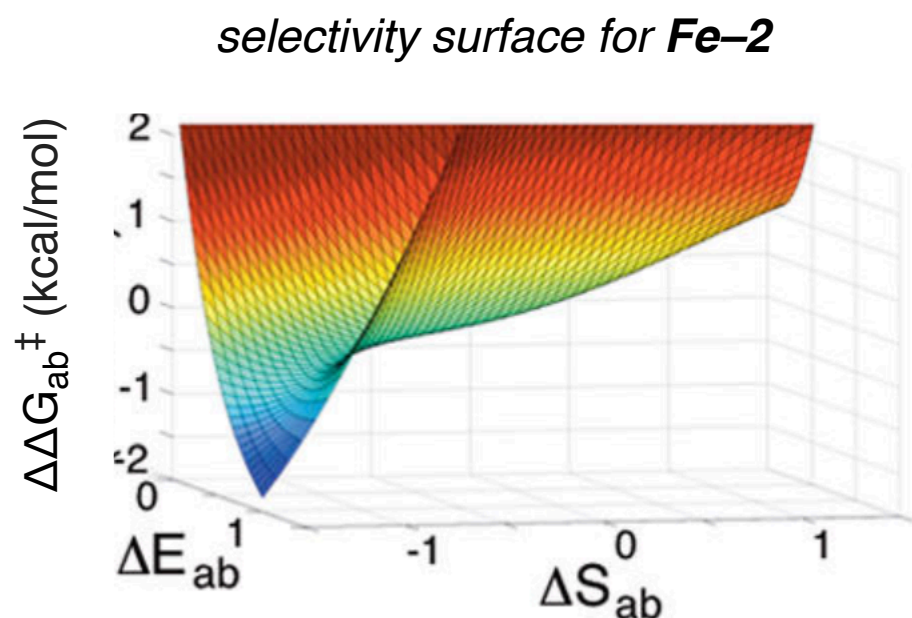
$\Delta\Delta G^{\ddagger}_{ab} = \Delta G^{\ddagger}_{b} - \Delta G^{\ddagger}_{a}$

$\Delta\Delta G^{\ddagger}_{ab} > 0$ means selective for C–H$_a$

ΔE_{ab} = NPA charge(C–H$_b$) – NPA charge(C–H$_a$)

$\Delta E_{ab} > 0$ means C–H$_a$ is more hydridic

ΔS_{ab} = sterics(C–H$_b$) – sterics(C–H$_a$)

$\Delta S_{ab} > 0$ means C–H$_a$ is more exposed

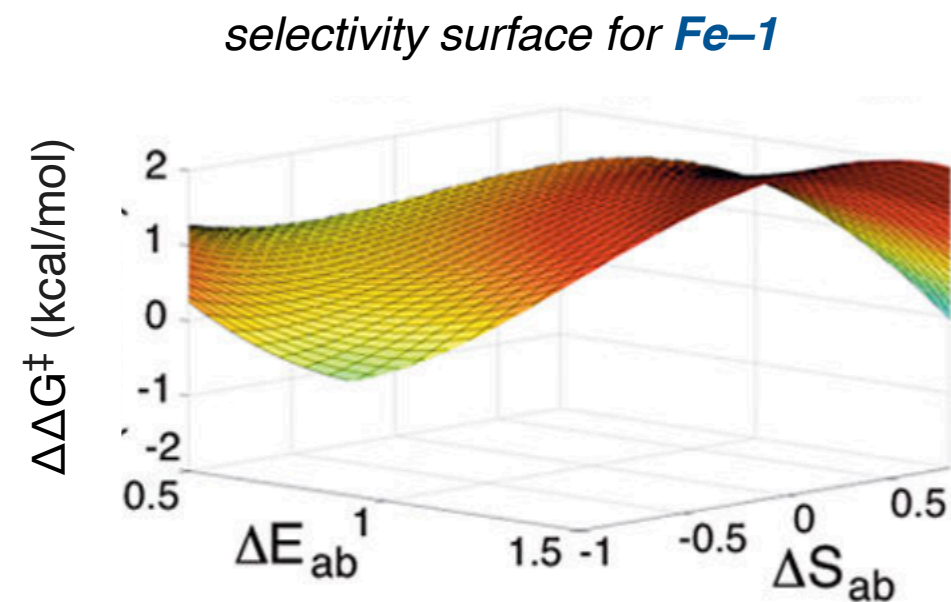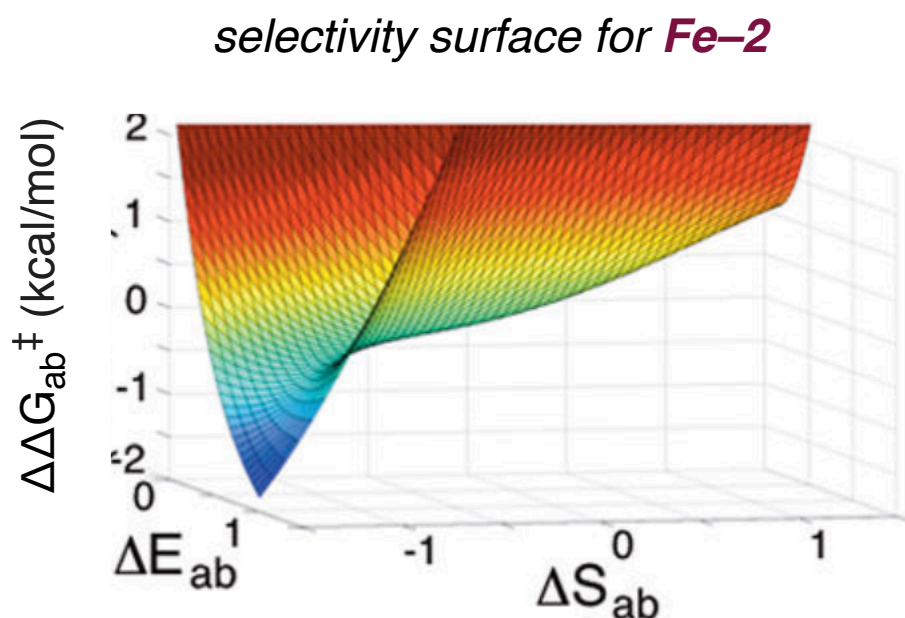***trends by visual inspection***

isosteric slice shows bimodal electronic effect

isoelectronic slice shows ~ symmetrical steric effect

Gormisky, P. E.; White, M. C. *J. Am. Chem. Soc.* **2013**, *135*, 14052–14055.

Predicting Site-Selectivity in C–H Oxidations

■ Comparing LFER surfaces to determine catalyst quality

*selectivity surface for **Fe–2***

*selectivity surface for **Fe–1***

---

desirable surface properties

high |Z| – high selectivity for a given product

(a high Z-value plane across the entire parameter space)

absent of perfect selectivity…
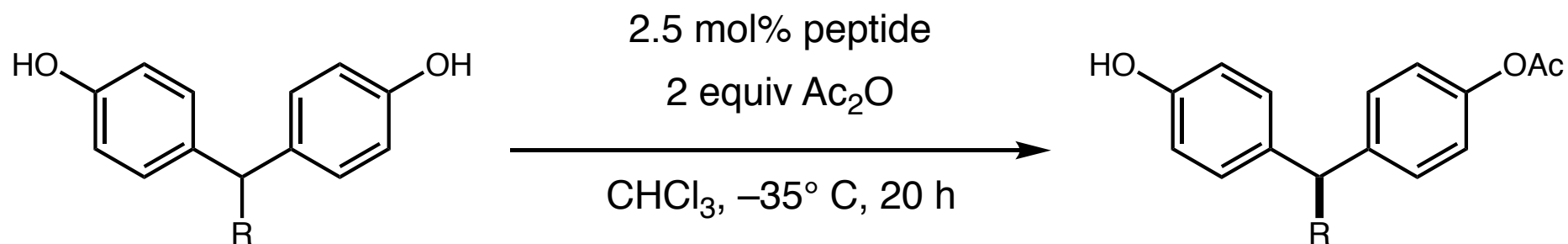
smooth surface – indicates high predictive power

■ Fairly flat, low Z-valued surface
■ High selectivities only at high values of ΔE and ΔS

Gormisky, P. E.; White, M. C. *J. Am. Chem. Soc.* **2013**, *135*, 14052–14055.

Molecular Vibration as a Steric/Electronic Parameter

peptide-catalyzed desymmetrization

2.5 mol% peptide
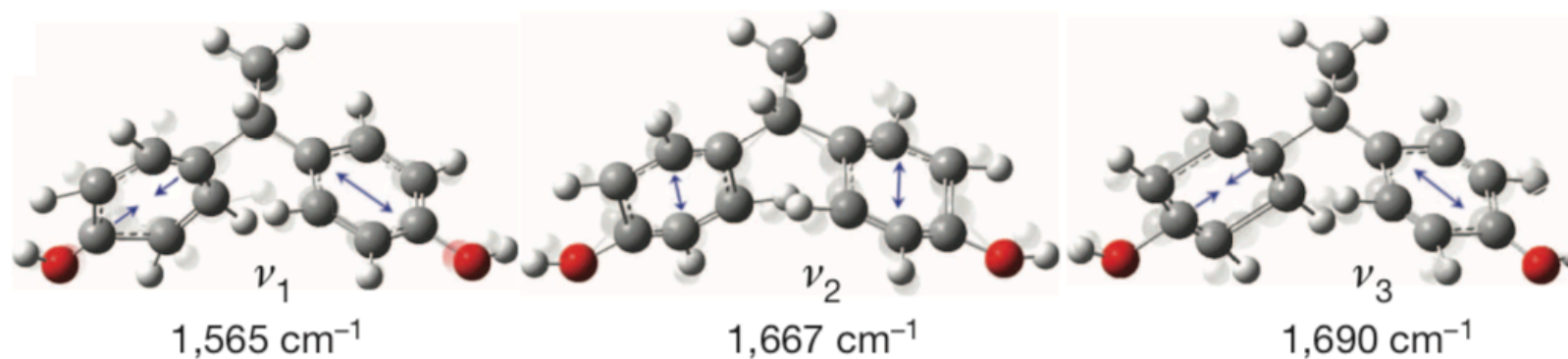
2 equiv Ac_2O

$CHCl_3$, −35° C, 20 h

vibrational energy is influenced by both sterics and electronics

can parameterizing molecular vibration be used to predict enantioselectivity?

ring vibrations:
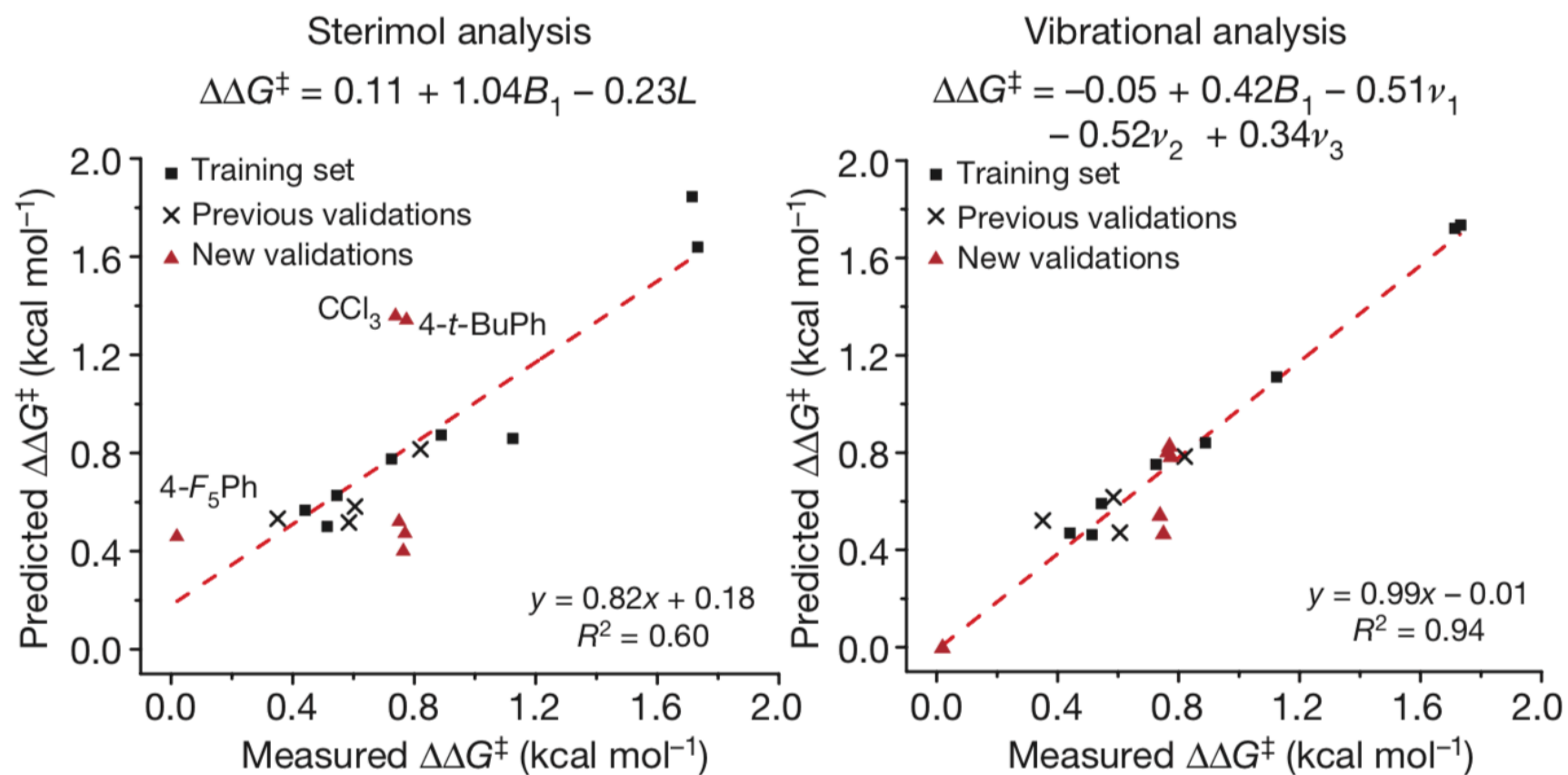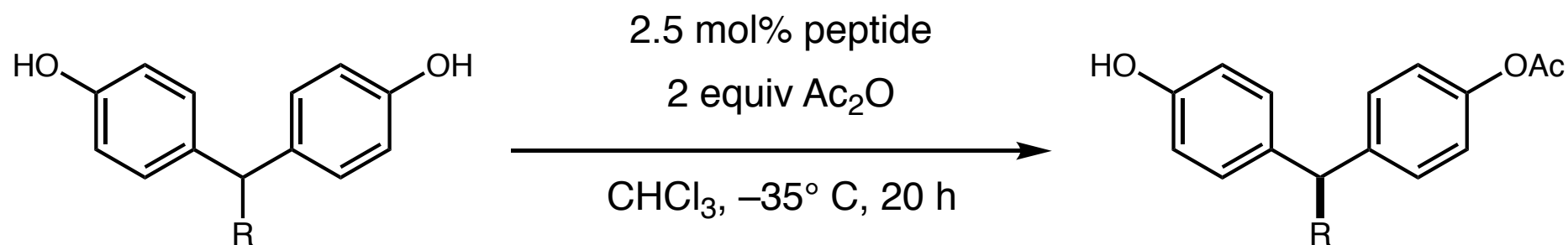

ν_1
1,565 cm^{-1}

ν_2
1,667 cm^{-1}

ν_3
1,690 cm^{-1}

3 vibrational parameters + 1 steric parameter ⟶ ***accurate prediction of enantioselectivity***

Milo, A.; Bess, E. N.; Sigman, M. S. *Nature* **2014**, *507*, 210–214.

Molecular Vibration as a Steric/Electronic Parameter

peptide-catalyzed desymmetrization

Sterimol analysis

$$\Delta\Delta G^{\ddagger} = 0.11 + 1.04B_1 - 0.23L$$

- ■ Training set
- ✕ Previous validations
- ▲ New validations

CCl_3 ▲▲ 4-t-BuPh

4-F_5Ph

$y = 0.82x + 0.18$
$R^2 = 0.60$

Predicted $\Delta\Delta G^{\ddagger}$ (kcal mol^{-1})

Measured $\Delta\Delta G^{\ddagger}$ (kcal mol^{-1})

Vibrational analysis

$$\Delta\Delta G^{\ddagger} = -0.05 + 0.42B_1 - 0.51v_1 - 0.52v_2 + 0.34v_3$$

- ■ Training set
- ✕ Previous validations
- ▲ New validations

$y = 0.99x - 0.01$
$R^2 = 0.94$

Predicted $\Delta\Delta G^{\ddagger}$ (kcal mol^{-1})

Measured $\Delta\Delta G^{\ddagger}$ (kcal mol^{-1})

vibrational parameterization outperforms purely steric parameterization

Milo, A.; Bess, E. N.; Sigman, M. S. *Nature* **2014**, *507*, 210–214.

Correlation Analysis and Pitfalls of Deducing Mechanism

■ Correlation analysis can quickly identify relationships between many parameters
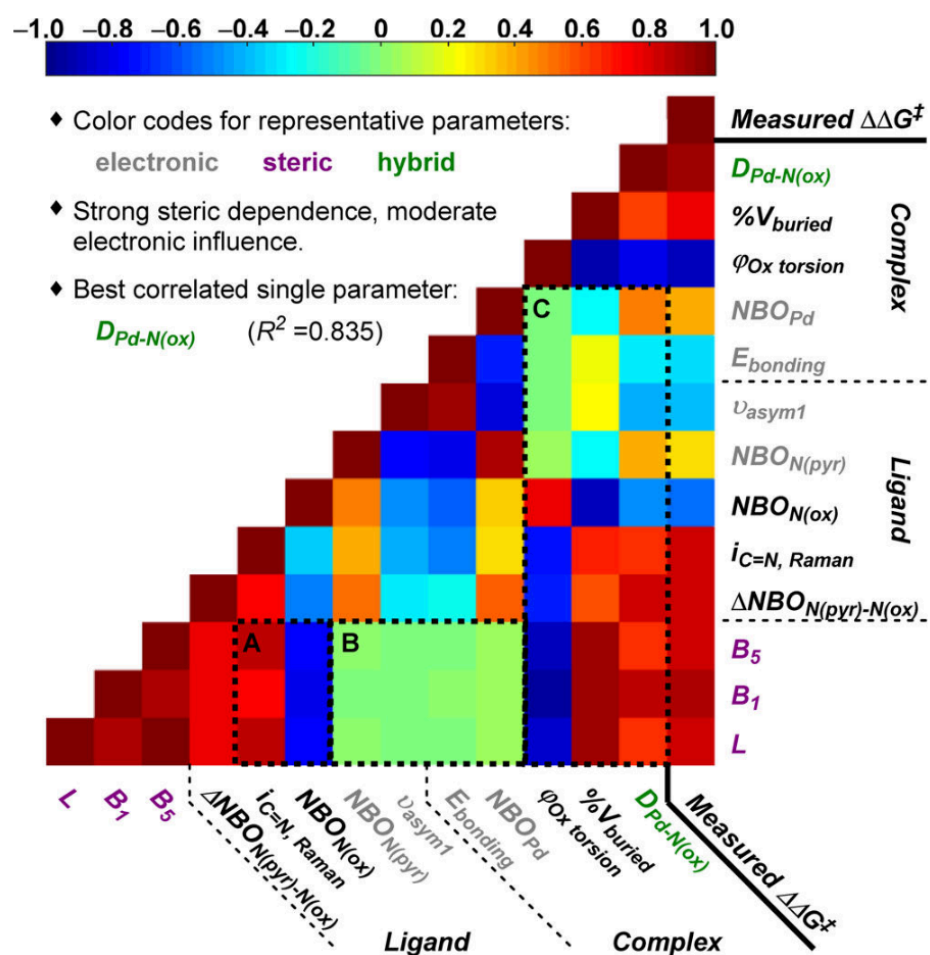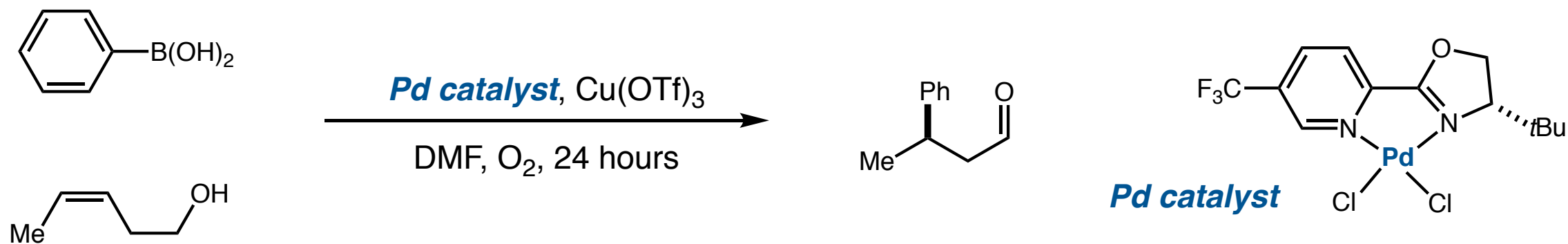
■ Pd–N bond length easily identified as strongly correlating

■ Purely steric/purely electronic parameters not correlated (in green)

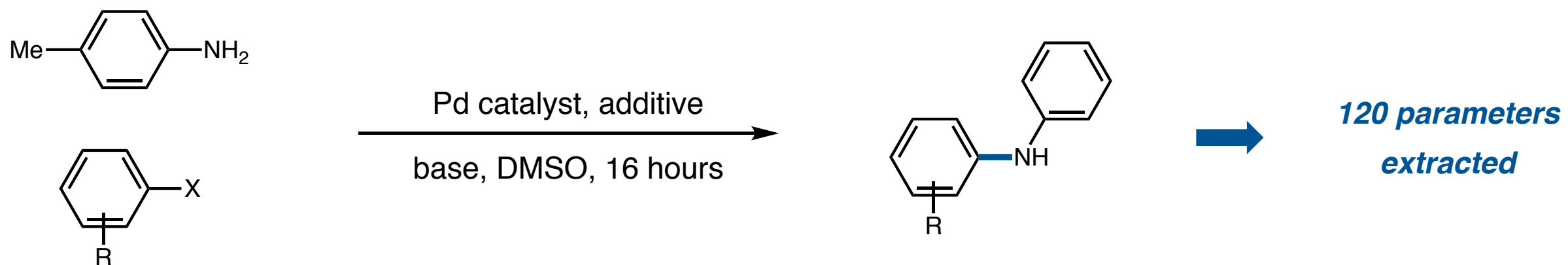■ Many parameters correlate with each other (in red)

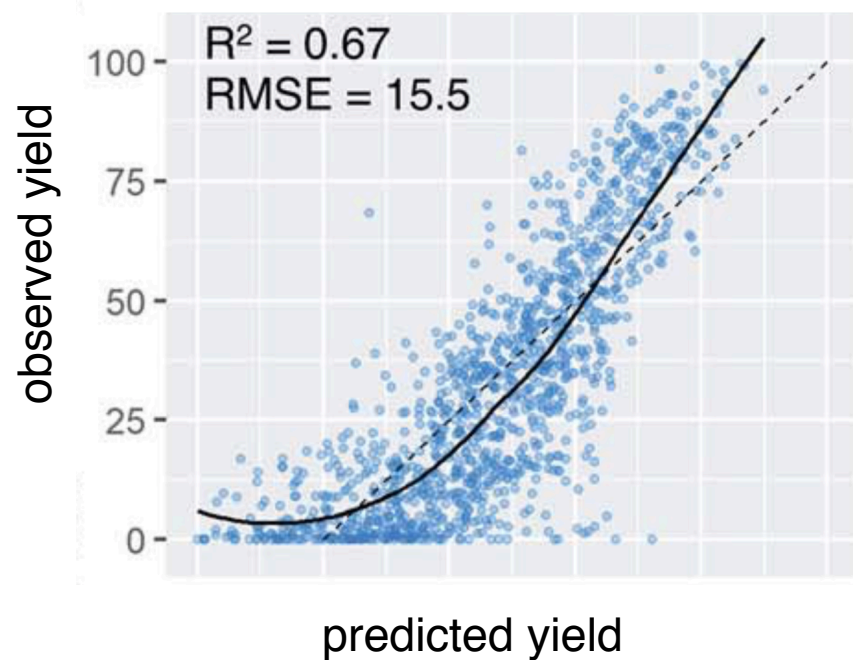**this can nullify mechanistic conclusions**

Guo, J.-Y.; Minko, Y.; Santiago, C. B.; Sigman, M. S. *ACS Catal.* **2017**, *7*, 4144–4151.

Machine Learning for Predicting Catalyst Performance

Buchwald-Hartwig Coupling

Pd catalyst, additive

base, DMSO, 16 hours

120 parameters extracted

Linear Model

$R^2 = 0.67$
RMSE = 15.5

observed yield

predicted yield

Random Forest Model

$R^2 = 0.92$
RMSE = 7.8

observed yield

predicted yield
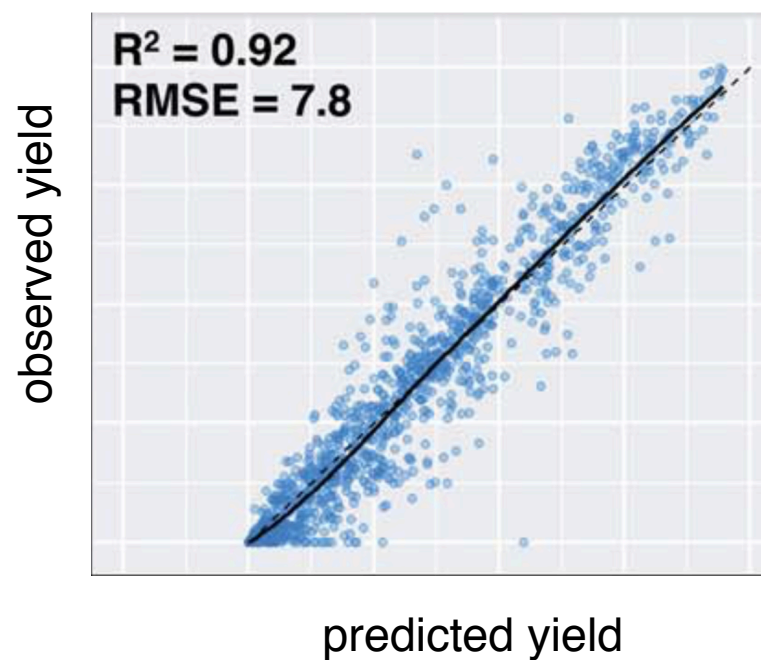
ML required to construct predictive model

Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. *Science* **2018**, *360*, 186–190.

Summary and Takeaways for Data-Intensive Reaction Development

% yield or % ee vs **reactions performed**

how do we use these data points to most effectively inform reaction optimization?

MLR LFERs?
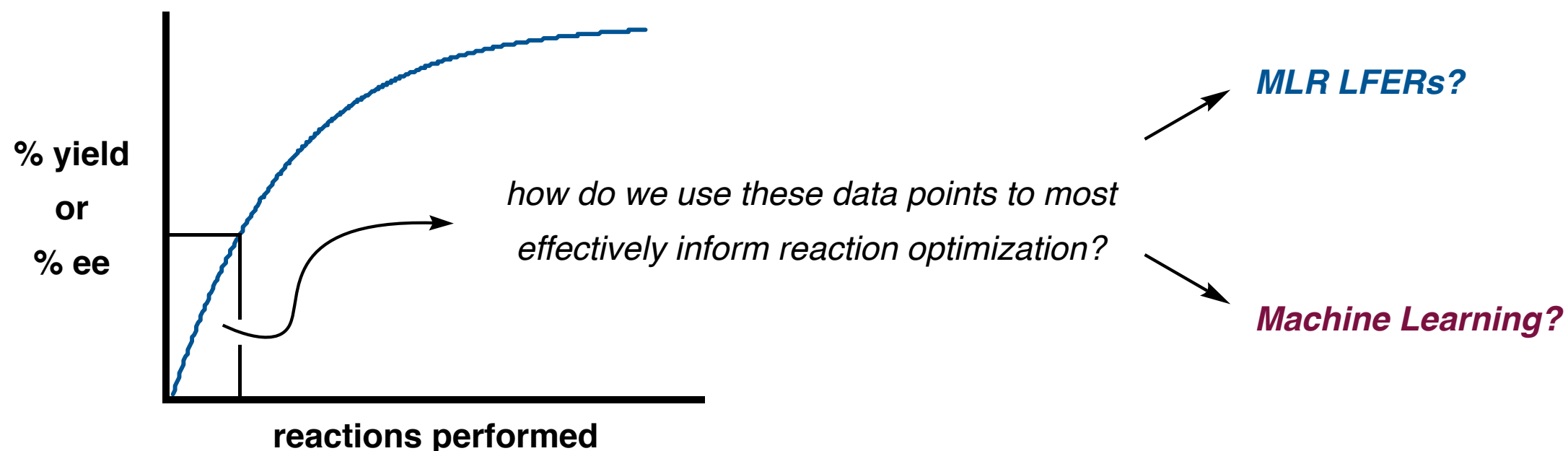
Machine Learning?

■ LFERs can be a powerful method of predicting catalyst performace, at least in a fairly confined chemical space

■ LFERs as mechanistic probes? with a carefully chosen parameter set and follow-up experiments, this is possible

■ Machine learning can offer a more powerful approach to predicting reaction performance

■ Large parameter sets and complex models come at the cost of mechanistic insight