Week 7

Quantifying Features of Moieties

7.1 Parameters and Linear Regression

- Lecture 11 recap.

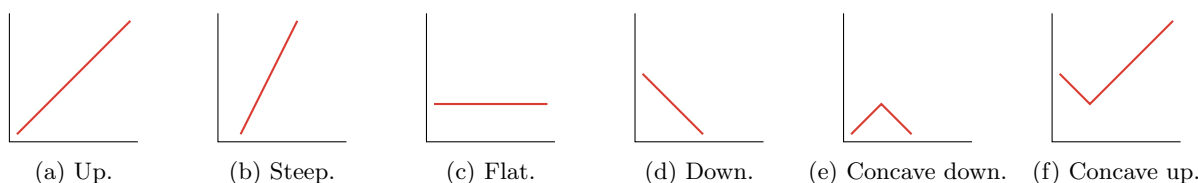| (a) Up. | (b) Steep. | (c) Flat. | (d) Down. | (e) Concave down. | (f) Concave up. |

Figure 7.1: Different types of Hammett plots.

  - LFERs and Hammett plots let us correlate substituent parameters to changes in the equilibrium (ΔG) or kinetic ($\Delta G^{\ddagger}$) energies of reaction.
    - These tools come from the observation that substituents exert common influences on reactions.
  - Substituent effects (inductive, field, resonance, polarizability, and steric). Also solvent effects.
  - The different types of Hammett plots.
    - Figure 7.1a: Some negative charge build up in the transition state.
    - Figure 7.1b: More negative charge build up in the transition state.
    - Figure 7.1c: No positive or negative charge build up in the transition state.
    - Figure 7.1d: Some positive charge build up in the transition state.
    - Figure 7.1e: Change in the rate-determining step.
    - Figure 7.1f: Change in the mechanism.
  - Remember that in a Hammett plot, our x-axis is a parameter σ that quantifies electron-donating or electron-withdrawing intensity, and our y-axis is either $\log(k_X/k_H)$ or $\log(K_X/K_H)$.
    - Remember also that stronger EWGs lie to the right, and stronger EDGs lie to the left.

- Announcements.

  - Next week: Masha's last lecture before Alex takes over. It will cover ML.
  - PSet 2: Will be graded by tomorrow or the next day, so we'll be able to study it for the exam.
  - Exam: Live on Canvas on Tuesday. Once downloaded, we'll have 90 mins to take and upload it.
    - Don't cheat; it's not open-book or open-note. Don't take it around anyone else.
    - Take the practice exam under exam-like conditions with a timer and everything.
  - Office hours: Jonathan will hold these virtually on Friday because he's a bit sick currently.

- Today: Continuing our discussion of parameters (such as σ) and linear regression.

- Lecture outline.

    - Defining two new substituent parameters (σ^+ and σ^-).
    - Other electronic parameters (Mayr, Swain-Scott, NBO, Mulliken, NMR, IR, orbital energies).
    - Steric parameters (A-values, sterimol parameters).
    - Stereoelectronic parameters (Taft parameters, Charton parameters).
    - Steric parameters in catalysis (bite angle, cone angle, PBV).
    - Higher-dimensional Hammett plots and the foundations of ML.

- We begin today with a critique of σ_p and σ_m, and the solution developed to address this critique.

- Essentially, chemists noticed that while σ_p and σ_m were good for characterizing electron-withdrawing and -donating character, they did not capture everything.
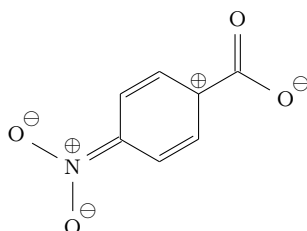

Figure 7.2: Carboxylates do not delocalize efficiently into arenes.

    - Importantly, they did not do a great job of capturing resonance effects since the benzoate anion could not delocalize efficiently into the aromatic ring.

        ■ In general, there is no resonance delocalization of carboxylates into arenes.
        ■ Recall from last lecture that the substituent can delocalize its charge up to the *ipso*-position; however, the anion can't go in.
    - This means that σ_p and σ_m underestimate π-EWG and π-EDG effects.

- As such, later chemists developed scales based on new reference reactions.

    - These new reference reactions generated anions and cations that could resonance-delocalize into the aromatic ring, and hence all the way to the substituent.
    - In particular, two new substituent parameters were developed: $\boldsymbol{\sigma^-}$ and $\boldsymbol{\sigma^+}$.

- $\boldsymbol{\sigma^-}$: A measure of a substituent's ability to stabilize (inductively and through resonance) the negative charge that builds up when a substituted phenol is deprotonated. *Reference reaction*
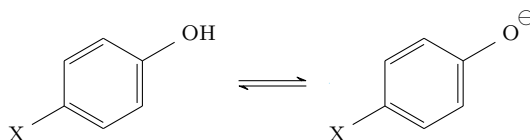
Figure 7.3: Reference reaction for σ^-.

    - This is the deprotonation of a phenol, which is nice because phenolates *can* delocalize their anion into the ring and over to the substituent.

        ■ Thus, this reaction better captures benzylic anion stabilization and π-EWG effects.
    - For reference, we set $\sigma^- := 0$ when X = H.

- σ^+: A measure of a substituent's ability to stabilize (inductively and through resonance) the positive charge that builds up when a substituted cumyl chloride is deprotonated. *Reference reaction*
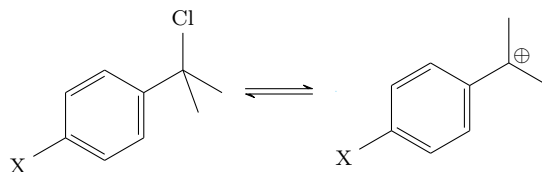
Figure 7.4: Reference reaction for σ^+.

  - This is the ionization of cumyl chloride, which is nice because cumyl groups can delocalize their cation into the ring and over to the substituent.
    - Thus, this reaction better captures benzylic cation stabilization and π-EDG effects.
  - For reference, we set $\sigma^+ := 0$ when X = H.

- Now that we have two new substituent parameters, let's compare some of their values with our old substituent parameters.

X	σ_p	σ^-	σ^+
CH_3O	-0.27	-0.26	-0.78
CH_3	-0.14	-0.17	-0.31
H	0	0	0
Cl	0.24	0.19	0.11
NO_2	0.81	1.23	0.79

Table 7.1: Comparing σ_p with σ^+/σ^- for common substituents.

  - Misc. observations.
    - σ^- reflects that nitro groups are *six* times better than chlorine, not four times as σ_p suggests.
    - σ^+ is more similar to σ_p for EWGs, and σ^- is more similar to σ_p for EDGs.
      - In the first row, σ^+ deviates (because it captures the π-EDG nature).
      - In the last row, σ^- deviates (because it captures the π-EWG nature).
  - These numbers just quantify our intuition about which groups are stronger, which way (EWG or EDG) the groups are stronger, and why!

- So now that we have several parameters, which one should we use?

  - This is a mechanistic probe; we don't know the mechanism yet!
  - So we plot all of them and see which gives us the best fit to a straight line.
  - This tells us something about the electronics of the transition state.
    - Is it positive? Negative?
    - Is it influenced by π-interactions? σ-interactions?
    - Etc.

- This concludes our discussion of Hammett substituent parameters.

- Let's now discuss some other electronic parameters, including some that describe inductive, resonance, field, and/or polarizability properties.

  - What if we want to quantify nucleophilicity or electrophilicity?

    - Use Mayr electrophilicity from Lecture 6!
      - ➤ Recall that here, we measure k for the reaction

      $$E^+ + Nuc^- \longrightarrow Nuc-E$$
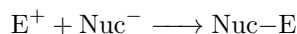
      and then set
      $$\log(k) = s(N + E)$$

    - There are also **Swain-Scott parameters**.
    - These are both empirical frameworks that have their own domains of usefulness.
      - ➤ These people spent their careers compiling tables of data so that we don't have to!
      - ➤ It's super useful.

  - What if we want to quantify atomic charges?

    - Look into **natural bond orbital** (NBO), **Mulliken**, etc.
    - Be careful here, though, because the numbers are often in the gas phase, so they may not be useful.
    - Alternatively, the *relative* values may be more useful than the absolute values.

  - Another way we can determine electronic effects is with spectroscopic data.

    - NMR shift or IR frequency can be used as a proxy for electronic character.

  - We can calculate the energies of certain bonds, orbitals (e.g., σ, σ^*), lp's, hybrid orbitals, etc.

  - We can also calculate hybridization — the percent s-character can easily be calculated exactly.

- **Swain-Scott parameter**: The rate of reaction for various nucleophiles. *Given by*

  $$CH_3-I + Nuc^- \xrightarrow{-I} CH_3-Nuc$$
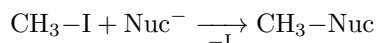
  - Swain and Scott defined two parameters (s and n_x) via the equation

  $$\log\left(\frac{k_{\text{nuc}}}{k_{\text{H}_2\text{O}}}\right) = sn_x$$

    - s is the sensitivity.
    - n_x is the substrate constant.

  - $k_{\text{H}_2\text{O}}$ indicates that we are setting the hydrolysis of methyl iodide as the reference reaction.

- Besides electronics, the other big thing in chemistry is sterics! The two main steric parameters are...

  1. **A-values**;

  2. **Sterimol parameters**.

- **A-value** (of an R group): The difference in energy between the axial and equatorial conformers of a mono-R-substitued cyclohexane.

  - This is one way of measuring the size of an R group.

  - Limitation: When the C−R bond is long.

    - Example: Cl has $A = 0.43$, and I has $A = 0.43$.
    - I is much bigger than Cl, but they have the same A-value because the C−I bond is so long; it doesn't really matter energetically if your I is axial or equatorial since it's far away from the hydrogens.

- **Sterimol parameter**: A parameter that considers the size of the R group is multiple dimensions.

(a) L/B_5 of 1-naphthyl.   (b) B_1 of 1-naphthyl.   (c) $L/B_5/B_1$ of s-butyl.
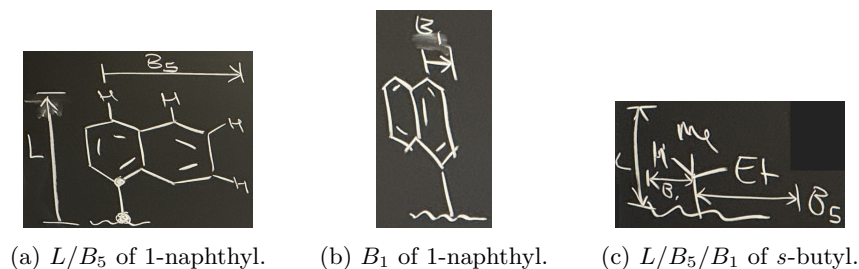
Figure 7.5: Sterimol parameters.

  – These are our best steric parameters to date.
    ■ They are very good at decoupling multiple dimensions of information.
    ■ They can be calculated by a number of programs and websites.
  – Sterimol parameters narrow down what we mean by "size."
    ■ For example, it's hard to say whether a tree or a car is bigger — what do we mean by "big?" Trees have lots of empty space, cars are more dense, trees have more mass (in general), etc.
    ■ The same is true of certain R groups.
  – Sterimol parameters define the size of a substituent dimension by dimension.
    ■ L is the length (in Å) from the "parent atom" to the end of the substituent, following the vector connecting the parent atom and the "start of the substituent" (Figure 7.5a).
      ➢ The "parent atom" and "start of the substituent" are the circled atoms in Figure 7.5a.
    ■ B_5 is the maximum size of the substituent along any vector perpendicular to $\vec{L}$ (Figure 7.5a).
    ■ B_1 is the minimum size of the substituent along any vector perpendicular to $\vec{L}$ (Figure 7.5b).
      ➢ Figure 7.5b is supposed to be a perspective drawing.
  – Masha also draws the sterimol parameters on a sec-butyl group (Figure 7.5c).

- Some parameters account for both steric *and* electronic effects.

  1. **Taft parameters**.
  2. **Charton parameters**.

- **Taft parameter**: A measure of a substituent's ability to electronically activate or deactivate as well as sterically block or expose a reactive site. *Given by*

$$\log\left(\frac{K_{\mathrm{X}}}{K_{\mathrm{H}}}\right) = \rho^*\sigma^* + \delta E_s \qquad\qquad \log\left(\frac{k_{\mathrm{X}}}{k_{\mathrm{H}}}\right) = \rho^*\sigma^* + \delta E_s$$

  – We can decouple the sterics and electronics by measuring k or K of both pathways.
    ■ This is some good, honest physical organic chemistry that some people did.
  – $\rho^*\sigma^*$ is the **polar term**.
  – δE_s is the **steric term**

- **Polar term**: The term governing electronics.

- **Steric term**: The term governing a substituent's ability to block π^* because of its anomalous size.

  – This definition is not necessarily universal; it's relevant to this reaction, specifically.
  – In some reactions, the ability to block some other antibonding orbital will be more relevant.

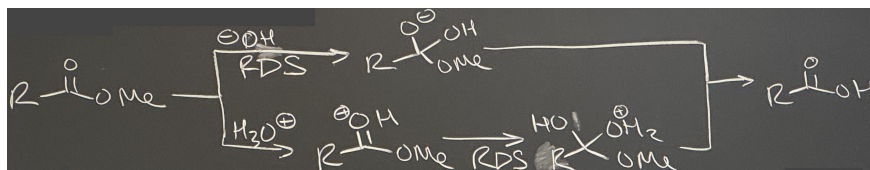• Example: The hydrolysis of a methyl ester under basic and acidic conditions.

Figure 7.6: Taft parameters characterize ester hydrolysis.

  – Under basic conditions, we have basically one step: Addition, then kicking out.
    ■ The RDS is the hydroxide adding in.
    ■ There is a negative charge buildup in the transition state.
    ■ Therefore, the sterics of R can block the addition and the electronics of R might change the carbonyl's electrophilicity. In other words, both the sterics and electronics of R matter.
  – Under acidic conditions, we could get protonation of the carbonyl followed by water addition to form the tetrahedral intermediate, and then elimination and deprotonation to the carboxylic acid product.
    ■ The RDS is the water adding in.
    ■ There is no charge buildup in the transition state (the charge is already included).
    ■ Therefore, it's *only* the sterics of R that matter.

• **Charton parameter**: A refinement of E_s with van der Waals radii. *Also known as* **Charton modification of Taft parameters**.

• Taft and Charton are both a bit historical at this point, but we still need to know them in order to read the literature.

  – Masha's never actually used them, but she does use sterimol.

• We now move onto some steric parameters in catalysis; these were motivated by the need to quantify ligand size.
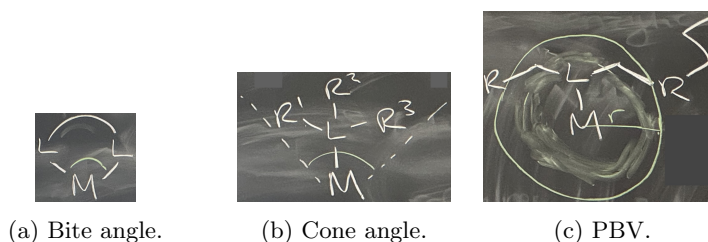
(a) Bite angle.   (b) Cone angle.   (c) PBV.

Figure 7.7: Steric parameters in catalysis.

• **Bite angle**: The $L-M-L$ angle for a bidentate ligand. *Schematic* Figure 7.7a.

  – To reiterate: Bite angle is a metric of size for bidentate ligands *only*.
  – Naturally, bite angle depends significantly on the size of the metal.
    ■ The example values below are all for the same metal.
    ■ Historically, bite angles were reported for nickel.
  – Examples.
    ■ DPPM, DPPE, and DPPP ligands have bite angles of 73°, 86°, and 91°.
    ■ TRANSphos has a 180° angle so that it sits on either side of our catalys; really useful!
  – Bite angle correlates really well to a lot of reactivity, so it's good to know.

- **Cone angle**: The angle from the metal to the outside R groups, where L is a monodentate ligand with three substituents. *Schematic* Figure 7.7b.

    - To reiterate: Cone angle is a metric of size for monodentate ligands only.
    - Cone angle also (naturally) depends on the metal.
    - Examples: Phosphane, trimethylphosphane, and triethylphosphane have 87°, 118°, and 132°.

- **Percent-buried volume**: The percent of the sphere around the metal occupied by the ligand, where the sphere has $r = 3.5\,\text{Å}$ by default. *Also known as* **PBV**. *Schematic* Figure 7.7c.

    - The radius can be changed, though, because the ligand should fit mostly in the sphere.
    - Examples: NHC ligands.

        - If R = Me, ^{i}Pr, 2,6-^{i}PrPh, then PBV $= 26, 28, 47$.

    - Example of when this is useful.

        - If you want to block your metal, you need bigger R groups.
        - But if you just choose floppy alkyl chains, that might not really block the sphere because they'll just flop away.[1]

    - We should see these in papers.
    - Use them in our work if we need!

- This is it for regular parameters at this point.

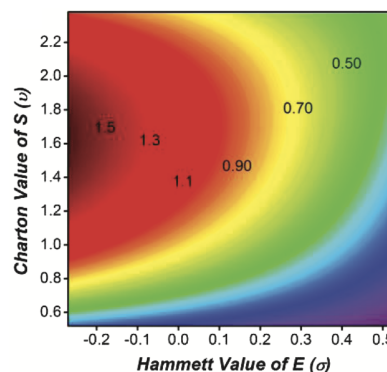- However, there's one more wrinkle: The case of multidimensional LFERs.

(a) Reaction.
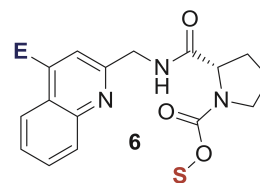
(b) Ligand template.
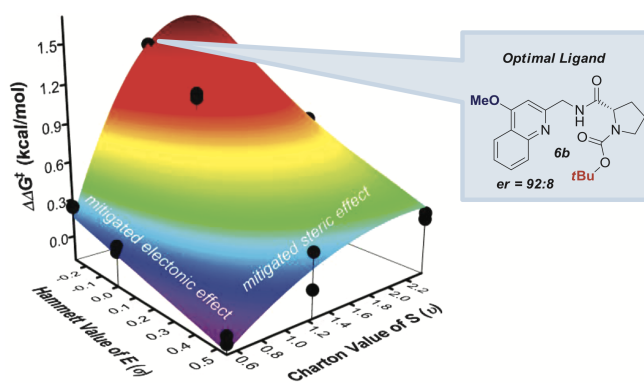
(c) 3D plot.

(d) Contour map.

Figure 7.8: Multidimensional LFERs.

---

[1]Sometimes floppiness can be the point, though, as in my research with Santa!

- To make sense of these, we use a higher-dimensional Hammett plot!
    - Even this one step up gets us to multidimensional regression, which is the dawn of ML in chemistry.
    - Back in the day, they called this multidimensional LFERs; today, we call it machine learning (ML).
- Essentially, some substituents can have synergistic or interdependent effects.
    - To conceptualize this kind of relationship, we model multiple parameters at once.
- History.
    - This work was pioneered by Matt Sigman at the University of Utah.
    - Now a lot of other people have jumped in: Abby Doyle, Connor Coley, Masha, etc.
    - "As this field gets bigger and bigger and hypier and hypier, no one should forget Matt. Don't come for Matt."
- Example: Catalytic $CrCl_3$ gets chelated to an asymmetric ligand (Figure 7.8b), and then enantioselectively combines two things (Figure 7.8a).
    - Three variables to consider: Two independent variables, and one dependent variable.
        - The authors varied the size of the substituent S, and measured its size with a Charton parameter, ν.
        - Simultaneously, they varied the electronics of the substituent E, and measured its EWG/EDG character with a Hammett parameter, σ.
        - They measured the ee, from which they could calculate er, k_{rel}, and finally $\Delta\Delta G^{\ddagger}$.
    - This all results in a 3D plot (Figure 7.8c).
        - The general shape is a sheet that's going up and down.
    - The contour map might be a bit easier to visualize (Figure 7.8d).
        - High ee toward the left and low ee toward the right.
    - We can also describe all this with an equation.

$$\Delta\Delta G^{\ddagger} = -1.20 + 1.22E + 2.84S - 0.85S^2 - 3.79ES + 1.25ES^2$$

    - The **cross terms** (ES and ES^2) in this equation are particularly important; they are a mathematical demonstration of the interdependency between sterics and electronics.
    - Thus, the $\Delta\Delta G^{\ddagger}$ depends on ligand sterics, electronics, and how those interact with each other.
    - The original constant doesn't have chemical meaning, then electronic parameter, then two steric parameters, then two cross terms.
    - What's happening here in chemical terms is that electron-poor ligands are not very sensitive to sterics, but electron-rich ligands are.
        - Note the difference in the curves in the back of the 3D plot and the front of the 3D plot. The back one (high EDG) is much more dependent on sterics! We want mid-sterics for highest ee.
        - The front one isn't great.
    - So the best ligand is when E is very electron-donating (OMe) and S is big but not too big (^{t}Bu; not something huge like adamantyl).
    - Reference: Harper and Sigman (2011).

- Overall guide/overview to building your own multidimensional free energy relationship with more parameters: Santiago et al. (2018).
    - A very accessible read!