Week 8

Molecular Relations & Quantification

8.1 Machine Learning

- Lecture 12 recap.

  - Different electronic parameters capture different features of molecules.
  - Electronic parameters.
    - Hammett parameters (σ).
      - ➤ Examples include σ_p, σ_m, σ^+, and σ^-.
    - Nucleophilicity and electrophilicity.
      - ➤ Examples include Mayr and Swain-Scott.
    - NMR or IR shifts.
    - You can also parameterize via the energy of certain electrons (e.g., σ, σ^*, lp, etc.).
  - Steric parameters.
    - A values: Historic.
    - Sterimol (L, B_1, and B_5): Common.
    - Taft (E_s) and Charton for stereoelectronic.
    - Bite angle, cone angle, and PBV for sterics in catalysis.
  - Why do we use parameters?
    - Correlating parameters to reaction outcomes (e.g., rate, selectivity, etc.) lets us. . .
      - ➤ Predict reaction outcomes;
      - ➤ Design better catalysts;
      - ➤ Learn something about the reaction mechanism (this is especially important for this class).

- Announcements.

  - Don't forget the exam!
  - This is Masha's last lecture. Fill out the teaching evaluations for Masha and Jonathan at the end of the course! Masha's evals will influence her tenure decision, and Jonathan's could help win him a teaching award.

- Today: More complex relationships between the input parameters from last time and our output.

  - This is machine learning (ML)!
  - Masha will focus on the applications of ML to organic chemistry, but please read more about the math and other applications if you're interested!

- There will be a lot of vocab in this lecture, starting with the definition of **AI**.

- **Artificial intelligence**: The development of computer systems able to perform tasks that normally require human intelligence. *Also known as* **AI**.

- Examples of such tasks.

  - Speech recognition, decision making, visual perception.
  - Not just things like calculus, but things that require a "greater" level of intelligence.

- Under the umbrella of AI falls **ML**.

- **Machine learning**: A subfield of AI that allows computer systems to learn and adapt without explicit instructions or programming. *Also known as* **ML**.

  - ML is characterized by the computer system being able to do things that we didn't explicitly program it to do.

- In the context of ML, we also have an explicit definition of **learning**.

- **Learning**: A computer program is said to form some experience (E) with respect to a task (T) and a measure of performance (P; aka the "performance metric"), if it's performance on T — as measured by P — improves with E.

  - This gets into the Turing test, and what it really means to know and to learn and to be conscious. This is more the realm of philosophy, and we won't get into that.
  - It's not like it did great from the beginning; it's that it had to get better with more experience.

- Reviews on the subject of ML in chemistry.

  - A great one to start for organic chemists: Williams et al. (2021). Four big-name corresponding authors.
    - Tobias Gensch: He's new, but we'll know him soon.
    - Sigman: The pioneer of multivariate linear regression.
    - Doyle: First to publish ML in chemistry; her 2018 *Science* paper — Ahneman et al. (2018) — exploded the field.
    - Anslyn: Wrote our textbook; the gold standard of Phys Orgo.
  - Any review publsihed by Doyle or Sigman will be great to read.
  - There are also great reviews from Connor Coley, Bill Green, and Klavs Jensen.

- Types of learning: **Supervised** and **unsupervised**.

- **Supervised** (learning): ML that has **labeled** training data.

  - This type of ML analyzes the labeled data and then makes a guess on unlabeled data. After the model guesses, we evaluate its performance.
  - Example: Show my model 100 reactions (with their yields labeled), and then have it guess the yield of a new reaction it's never seen before.
  - This is called "supervised" learning, because after the model guesses, *we* need to show it the right answer (i.e., the label).
  - Example: Spam filters.
    - These separate spam from "ham," the technical term for good emails.
    - We train such models by showing them a bunch of spam emails and a bunch of ham emails so that they "learn" what spam looks like.
    - The model looks for typos, weird email addresses, requests for money, etc.

- **Labeled** (training data): A set of data in which each data point (or datum) has an input and output label.

- **Unsupervised** (learning): ML that has **unlabeled** training data.

  – This type of ML tries to uncover relationships between data and find patterns.
  – This is "unsupervised" because there is no right answer, no guidance, no yield.
  – A common approach: **Clustering**.
  – Example: Netflix recommends movies that are similar to each other (i.e., which share common actors, common runtime, common genre labels, common people who have watched them, etc.).

- **Clustering**: Grouping together similar data.

- A really common approach is to do both of these at once in **semi-supervised** learning, our secret third option.

- **Semi-supervised** (learning): ML that splits data into a small labeled dataset and a big unlabeled dataset.

  – We group data together and assign a label to the group.
  – Example: Image classification, i.e., to answer the question, "which photos are of the same animal?"
    - An unsupervised ML finds similar images, and then a few of those get labeled "cat," so the whole group gets labeled "cat."
    - This is how self-driving cars and Captcha work. When you help Captcha find all the images with stairs, you're (nonconsensually) providing labels to help train image recognition models!

- We'll focus on supervised learning for the rest of today.

- Two types of supervised learning: **Classification** and **regression**.

- **Classification**: The output/label is a category.

  – There are a finite number of options.
  – Example: Photos are "cat," "dog," or "human."

- **Regression**: The output/label is a continuous number.

  – There are an infinite number of options.
  – Example: We could model the cost of a house as a function of house properties (e.g., the year it was built, the year we're trying to buy it, the cost of the surrounding homes, the neighborhood school system, etc.).
  – Example: Model ΔG as a function of reaction parameters.
    - This is Hammett plots! That was linear regression, so that's why we call this, "regression."[1]
  – Far more common in chemistry.
  – Formal definition: A statistical technique for determining the relationship between independent or explanatory variables (x) and dependent or response variables (y).

- **Linear regression**: Describe the relationship between x and y as a straight line.

  – Fitting to $y = mx + b$.
  – Example: $\log(k_X/k_H) = \rho\sigma$.
  – Some people don't call this ML; they call this "statistics." But that distinction is really only fought over by people who care about semantics or credit. So you may hear some strong opinions in the field (e.g., Sigman doesn't call it ML), but it's just labels at the end of the day (in Masha's opinion).

---

[1]Is there a "second time you hear it" effect in psychology that mimics ML? Unlikely to place emphasis on something the first time we hear it (e.g., Dad saying that there are crazy jobs for smart people/Maya telling me about the email tracker), but more likely when we hear it again (e.g., Carina Hong's job/Dylan Miars telling me about the email tracker). Relation to retention in learning!

- **Multivariate** (linear regression): Multiple x and y variables.
  - This is all the work of Matt Sigman, building off of his classic Hammett paper (Figure 7.8) that we reviewed last lecture.

- The ML workflow: Here are the steps if you want to go into lab and plan a project.
  0. Know or define your goal and application.
     - Why do you want this model?
     - What do you want it to do?
     - Why do you want to use it?
     - Why would anyone care about it?
     - A model that can predict yield to a decimal point will need tens of thousands of data points.
       - If you want a model to help you refine a ligand for a reaction that you've already studied pretty well, that's a good use.
     - ML is fundamentally an engineering solution, so you better have a practical use for this tool you've built.
  1. Data collection.
     - How much data?
     - Labeled or unlabeled?
     - Will this data come from the literature or from experiment?
       - This is an especially relevant consideration in chemistry.
     - Be aware of bias; we, as a field, tend to overreport high-yield reactions and underreport low-yield reactions. So if we train a model based just on the literature, it will think all chemical reactions are high yield.
     - The adage here: Garbage in = garbage out.
       - If you train a model on bad data, you're going to have a bad model.
     - So we can go into the lab and get a bunch of low-yield reactions and feel good about it, which almost never happens!
     - Warning: If you go to Reaxys and dump a bunch of data into a model, the error rate is 30-60% (typos and such).
       - Notoriously, the patent database used to be 60% wrong due to bad data-scraping.
  2. **Parameterization**.
     - Categorical descriptors (common for solvents, salts, and additives).
       - Tell your model that these 10 were run in toluene, these 10 in DCM, etc. That's a common category approach.
     - Chemically meaningful parameters.
       - This is all of last lecture: σ values, sterimol values, etc.
       - These are chemically meaningful because they capture a feature important to reactivity.
     - Graph networks: Atoms are nodes, bonds are edges, etc.
     - Molecular fingerprints: Lists of functional groups.
       - Example: My molecule contains a ketone, an ester, two methynes, etc.
     - SMILES, or its derivatives: This is how ChemDraw encodes molecules.
       - Example: Cyclohexane is "C1CCCCC1".
       - This string is just text, so then we can use LLMs.
       - Masha doesn't think these work that well, but they do exist.
     - Most ML papers mess up by this point: They either got bad data, or misrepresented it.
     - Most ML applications use millions or billions of datapoints, so chemistry is a bit unique in that it uses dozens of data points. How to make that tenable is a big question in both the chemistry and computer science communities!

3. Data preprocessing: Cleaning up the data for modeling.
   - Technique: Normalization.
     - Make all parameters lie in the range of 0-1.
       ➤ To do this, just divide by the maximum value.
     - Example: Sterimol values of 1, 4, and 8 become 0.125, 0.5, and 1; charge values of 0.01, 0.04, and 0.08 become 0.125, 0.5, and 1.
     - Normalization helps us make sure that the model doesn't think sterimol values are 100 times more important than charge. Essentially, it prevents bias toward parameters with large values.
   - Technique: Reduce the number of parameters (if needed).
     - Sigman recommends 8 data points per single parameter.
     - If we have more parameters than data points, we have an **overfit** model (see Figure 8.1c), and that is no good.
   - The simplest model architecture (Occam's razor) is the best model.

4. Data sampling: Splitting data into **training**, **validation**, and **test** sets.
   - We also sometimes have an **out-of-sample** set.
   - So we have a model, but just like in science, we need the model to make useful predictions for it to be a good model.
   - In the validation set, the model is going to show us it's performance.
     - Example: If linear regression does 70% right and multivariate linear regression does 90% right, we go with the multivariate architecture.
   - Then the performance of the model that we report is the performance on the final (test) set.
   - Overexposing your model to the test set invalidates the model; this is called **data leakage**.
   - Note: Chemists tend to mix up the "validation" and "test" sets in their writing.
     - Just make sure that we have real evidence that the model works on *unseen* data.
     - If you see people say, "we used the validation set to test the model, and then applied the model to the test set," that's not a bad thing; that's just a semantic error.

5. Training and testing: Evaluate the performance of different model architectures based on **metrics**.
   - Example metrics: Accuracy, percent data explained, R^2, RMSE (root-mean-square error).
   - Here, we run control and baseline models to predict the average and mode.

6. Interpretation and prediction: Use the trained model to predict reaction results, guide catalyst design, or present mechanistic hypotheses.
   - Best practice: Test the model experimentally.
   - Relating back to Point 0: Make sure the model is worth making.
     - Make sure you use it for something cool; you don't need a chainsaw to hammer in a nail.

- **Parameterization**: Converting chemical information to **machine-readable** formats.

- **Machine-readable** (input): A number, binary value, graph representation, etc.

- **Training** (set): The set on which the model learns the trends.

  - Roughly 70% of the data.

- **Validation** (set): The subset of the training set that help you choose between model architectures.

- **Test** (set): The set on which we evaluate the model performance.

  - Roughly 30% of the data.
    - The actual numbers here are up to you! People do everything from 90 : 10 to 60 : 40.
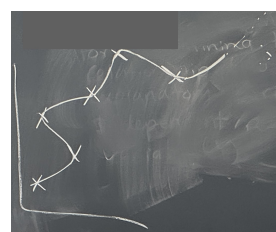 - We really only want to use the test set once or twice.

- **Out-of-sample** (set): A set that helps us further validate generalizability or **extrapolation**. *Also known as* **experimental validation** (set).

- **Data leakage**: The mixing of data between the test set and the training set. *Also known as* **poor data hygeine**.

- **Extrapolation**: The ability to make predictions beyond the training set.

- Memo: Extrapolation vs. **interpolation**?

    - There definitely is a difference.
    - For an example of true extrapolation, see Sigman's paper on extrapolating a model trained on ligands with ee below 80% to find ligands with ee beyond 80%.

- Don't report your training set performance!

- **Overfit** (model): A model that can predict the training set, but not new data.

    - Such models cannot generalize, and they definitely cannot extrapolate.
    - See Figure 8.1c.

- There are actually three types of fitting.
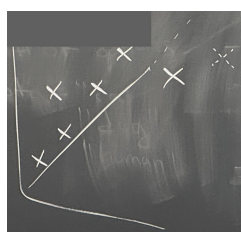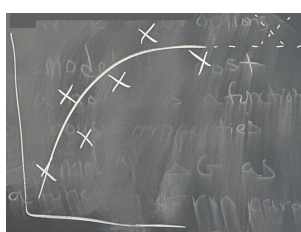
(a) Underfit.   (b) Good fit.   (c) Overfit.
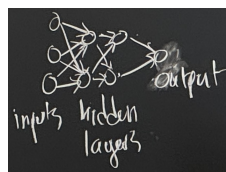
Figure 8.1: Fitting machine learning models.

    - Figure 8.1a: Underfitting.
        - Doesn't generalize to new data.
    - Figure 8.1b: Good fitting.
        - Generalizes well to new data.
    - Figure 8.1c: Overfitting.
        - Doesn't generalize at all to new data.
        - This is tempting to chemists, because it gives them a good-looking model. But that's not actual model; that's fraud!

- Model architectures.

(a) Decision tree.   (b) Neural network.

Figure 8.2: Machine learning model architectures.

- These exist on a spectrum from models with low complexity and high interpretability to models with high complexity and low interpretability.
- Lowest complexity and highest interpretability: Linear regression.
- k-nearest neighbors (knn).
  - Basic idea: Our ligand is close to something with high ee, so we'll probably get high ee, too.
- Decision trees (Figure 8.2a).
  - Answer questions such as, "high electronegativity or low electronegativity," and correlates that to ees.
- Random forest.
  - A subset of decision trees.
  - Make a lot of trees and average the result.
  - Called a "forest" because there are many trees!
- Highest complexity and lowest interpretability: Neural networks (Figure 8.2b).
  - Like the brain: Inputs, through hidden layers, that converge on an output.
  - This year's physics Nobel Prize went to neural networks; Masha's not quite sure how these are physics, but they really have changed their game.
  - They're very powerful, but extremely complex (so often overfit and not very interpretable).
    - ➢ Called "black box models."
    - ➢ Not good for mechanisms!

- Masha also has some additional notes on model architectures that she will post on Canvas.

8.2 Noncovalent Interactions

10/24:
- Logistics.

  - Alex has led this class for several years at this point, and it is his favorite one to teach!
  - Goal for the remainder of the semester: Build out an increasingly deep foundation on molecular reactivity.
  - Thus far, we've learned about the intrinsic reactivity of certain things (e.g., carbenes); from now on, though, Alex will expand beyond intrinsic properties to talk about how molecules behave not only in isolation, but with other molecules and how these bits of reactivity manifest (sometimes in statistical ways).
  - A lot of stuff going forward may be review from physical chemistry, advanced organic, etc. This is not a problem! That's a feature of the graduate curriculum.
    - Especially if you've seen it before, be more inquisitive this time! Once you're familiar, be incisive.
    - If it's new, that's why we're here! Learn it now.
  - This should be an interactive lecture: Alex will pose questions and call on people if answers are not forthcoming.
  - The class style will be chalk-talk. The majority of what we need to know goes on the board, but not all!
    - If he riffs and says something exciting, write it down! It will not be in the notes, though.
  - 2 PSets and 1 exam this half-semester as well.
  - Exam 1 will likely be returned before next Tuesday.
  - Class begins at 10:35 on the dot, and Alex will talk until (not past) noon if he has enough to say.

- Alex is quite funny!

- We've learned a bit thus far about how molecules adopt their equilibrium shape (e.g., with Walsh diagrams).

  – Electronic control elements define intrinsic properties, which can tell us a great deal about reactivity.
  – Frontier MOs are very important, "as Ken Fukui and Roald Hoffmann taught us."
  – We will now layer on top additional forces that give us greater insight into how molecules behave when in proximity to others.
  – Unless you're a microwave spectroscopist, most of the chemistry we're interested in happens in a condensed medium.

- Later on: How these interactions affect a system, and how stimuli can modify that system.

- At the end of the class: Photochemistry and electron transfer, if we move apace.

- Today: **Noncovalent interactions**.

  – We'll list a bunch of forces; a not super fun lecture. But we'll reference all these forces going forward.

- **Noncovalent interaction**: An interaction between atoms (stabilizing or destabilizing) that does not occur pairwise through a covalent bond. *Also known as* **NCI**.

  – These are enthalpic and entropic forces that modulate molecular interactions.

- **Electrostatic** (interaction): A charge-charge interaction. *Also known as* **ion-ion interaction**.

  – Even in organic chemistry, a lot of charge densities can be generated.
  – Example: Ammonium ion held near a carboxylate.
  – To learn something about the overall energy of a system subject to this interaction, use **Coulomb's law**.

- **Coulomb's law**: If a charge q_1 is held a distance r from a charge q_2, the energy of the system is given as follows. *Given by*
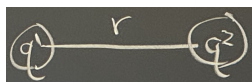
$$E = \frac{q_1 q_2}{4\pi\epsilon r}$$

Schematic

Figure 8.3: Schematic of an electrostatic interaction.

  – ϵ is the **dielectric constant**.
  – What's important here is that $\Delta E \propto 1/r$.
    ■ Not all forces have this energy relationship!

- **Dielectric constant**: The ability of a medium to isolate two static charges from each other. *Denoted by ϵ*.

  – This is *distinct* from ϵ_0, which is the ability of a *vacuum* to isolate two static charges from each other!
  – Anslyn and Dougherty (2006) refers to the permittivity of the medium as ϵ_μ, and uses ϵ to denote ϵ_μ/ϵ_0.

- Example of electrostatic interactions.

(a) Reaction.

(b) A bulky ligand (dtbpy).   (c) A charged ligand.
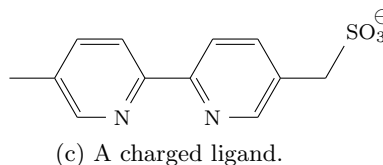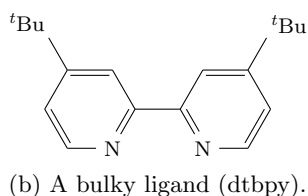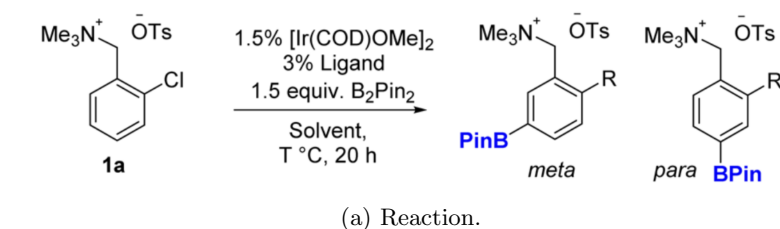
Figure 8.4: Electrostatic interactions in catalysis.

- A chloroarene bearing a pendant ammonium ion is subjected to catalytic borylation conditions (Figure 8.4a).
- When a particular ligand is employed (Figure 8.4b), the catalyst delivers the Bpin to the position remote from the charged group (*para*) favorably over the *meta*-position in a 2 : 1 ratio.
- However, use of a complementarily charged ligand (Figure 8.4c) led to the complete inverse regiochemistry — and with an improved selectivity of 10 : 1!
- The dielectric can also modify this: More polar solvents with higher dielectrics erode selectivity.
- Reference: Davis et al. (2016).
  ■ Also out of the Phipps lab at Cambridge!

- **Charge-dipole** (interaction): An NCI between a point charge and a dipole. *Given by*

$$E = \frac{\mu q_2 \cos(\theta)}{4\pi\epsilon r^2}$$
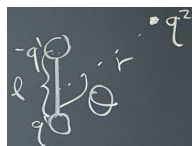
Schematic

Figure 8.5: Schematic of a charge-dipole interaction.

- μ is the **dipole moment**.
- We can number charges q with subscripts or superscripts.
- Again, higher dielectrics screen the charge better.
- Here, $\Delta E \propto 1/r^2$.
- Example: The attraction between a carbonyl and a Li^+ ion, e.g., carbonyl activation by a Lewis acid!

- **Dipole moment**: A measure of the separation of positive and negative electrical charges within a system; that is, a measure of the system's overall polarity. *Denoted by* $\boldsymbol{\mu}$. *Units* D "debye"

- These are both great, but many organic molecules are uncharged.

- **van der Waals** (interaction): An NCI between uncharged molecules.

  - Can be either attractive *or* repulsive.
  - Several subsets.
    - I.) **Dipole-dipole interactions**.
    - II.) **Dipole-induced dipole interactions**.
    - III.) **Dispersion forces**.

- **Dipole-dipole** (interaction): An NCI between two dipoles fixed in the same plane and parallel. *Given by*

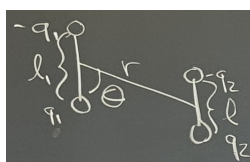$$E = -\frac{\mu_1\mu_2(3\cos^2\theta - 1)}{4\pi\epsilon r^3}$$

Schematic

Figure 8.6: Schematic of a dipole-dipole interaction.

  - Here, $\Delta E \propto 1/r^3$!
  - If the dipoles align, this can stabilize the system, building up hierarchical interactions that can even lead to self-assembly (e.g., in the case of polymers).
  - Example: Solid-state NMR with **magic angle spinning**.
  - See Anslyn and Dougherty (2006, p. 168).

- **Magic angle spinning**: When you optimize the dipole-dipole interaction energy expression to have $E = 0$ by setting $\theta = 54.7°$ (aka, the "magic angle").
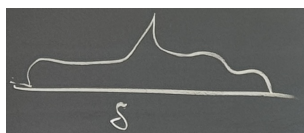
(a) Without MAS.   (b) With MAS.

Figure 8.7: Magic angle spinning.

  - Essentially, a "normal" solid-state NMR spectrum has a bunch of the anisotropy and broad linewidths.
  - However, suppose we fix the sample in the magnetic field at $54.7°$. Then we spin the crap out of it ($\approx 10\,000\,\mathrm{rpm}$). In this case, you get really fantastic tight lines from which you can read out the chemical shift tensor.

- David: Why is there only one θ in Figure 8.6?

  - Look at an intro physics or linear algebra textbook.
  - There are more degrees of freedom that we can introduce, but this is all we need for right now. If the dipoles are not parallel or in the same plane, the equation gets more complicated but the critical $\Delta E \propto 1/r^3$ relation always remains true.

- **Dipole-induced dipole** (interaction): An NCI in which one dipole induces a dipole in a **polarizable** molecule, and the two attract. *Also known as* **Debye force**. *Given by*

$$E = \frac{\mu^2 \alpha}{r^6}$$

  - Example: Taking a polar molecule (like water) and bringing it close to something polarizable like an arene, so that the two attract.
  - See Anslyn and Dougherty (2006, pp. 187–88).

- **Polarizability**: The susceptibility of an atom or molecule's electron cloud to being pushed around or distorted. *Denoted by* $\boldsymbol{\alpha}$.

  - α is called the **polarizability tensor**.
  - Indeed, polarizability can be quantified as a tensor (not a vector).
  - See Anslyn and Dougherty (2006, pp. 24–25).

- **Disperson** (interaction): An NCI between the quantum-mechanically arising, correlated motions of electrons between molecules. *Also known as* **London force**. *Given by*

$$E = \frac{\alpha_1 \alpha_2}{r^6}$$

Schematic

Figure 8.8: Schematic of a dispersion interaction.

  - Any single electron induces a small correlation, but many electrons can induce some big correlations in bigger molecules.

    - Longer molecules also maximize dispersion forces.
  - Governed by the Pauli exclusion principle: We can't fit more than two electrons into the same orbital, so electrons will still not "hop" between molecules.

- Example of dispersion forces: Silylenes.

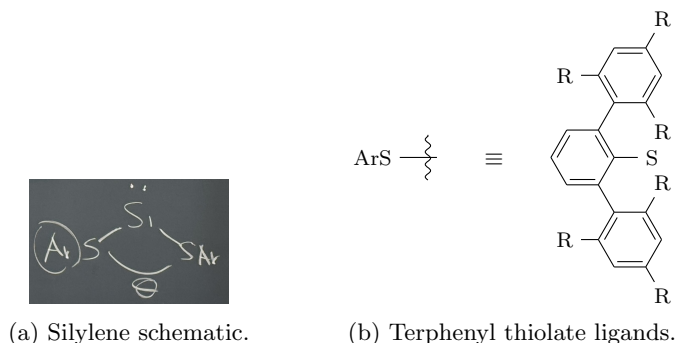(a) Silylene schematic.   (b) Terphenyl thiolate ligands.

Figure 8.9: Dispersion forces bend silylenes.

- Silylenes are the silicon analog of carbenes (Figure 8.9a).
  - We've already talked about carbene bending in this couse.
  - Here's an additional note, though: There are both electronic *and* steric reasons why a carbene would bend at a certain angle, since the electronic preference for some angle has to overcome the steric clash.
  - The same is true of silylenes.
- Consider a silylene with two terphenyl thiolate ligands (Figure 8.9b).
  - Such ligands are a lot to look at, but they're really easy to make.
  - When R = Me, $\theta \approx 90.5°$.
    - Note that this is a smaller angle than either singlet or triplet carbenes (see Figure 5.15) because the heavier **congeners** of carbon tend to be more bent.
  - When R = ^{i}Pr, $\theta \approx 84.8°$.
    - Thus, when the terphenyl thiolate ligands are bulkier, the silylenes are *more* bent!
    - Steric repulsion is being overcome somehow.
- Why this counterintuitive result?
  - Steric interactions are not necessarily repulsive! We've all been lied to.
  - Why? Recall the **Lennard-Jones potential**.
    - At large distances, two atoms have an energy as if they're not interacting at all.
    - As you bring them together, you get a stabilizing interaction until the nuclei begin repulsing, giving you an equilibrium bond distance.
    - Implication: Things must be *very* close together to get steric repulsions; there are bonding interactions that occur at far greater systems.
    - Steric properties of molecules are not always repulsive.
    - This is a universe-level phenomenon that matter tends to condense.
    - This is a profound insight that is animating a lot of method development in organic chemistry right now.

- **Congeners**: Chemical substances related to each other by origin, structure, or function.

- **Lennard-Jones** (potential): A mathematical energy potential that fairly well approximates the energy of two binding atoms as they are brought closer together. *Given by*

$$E \propto \frac{A}{r^{12}} - \frac{B}{r^6}$$
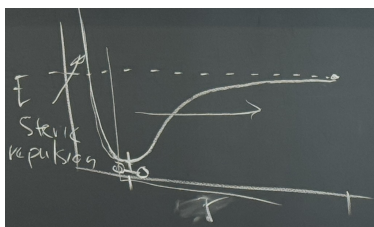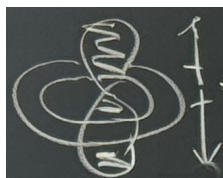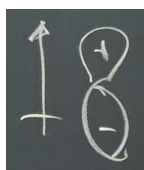
Graph

Figure 8.10: Lennard-Jones potential.

- Most Lennard-Jones potentials have the 12-6 form transcribed above.
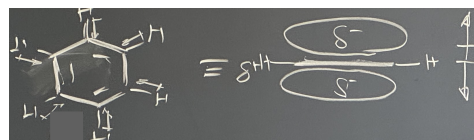- This concludes our discussion of van der Waals interactions.

- **Quadrupole**: Something that has the shape/topology of a d-orbital.

- Examples of quadrupoles.

    (a) Quadrupolar orbital.   (b) Dipolar orbital.   (c) Quadrupolar molecule.

Figure 8.11: Quadrupole examples.

- A d_{z^2} orbital has two dipoles that cancel each other out (Figure 8.11a).
  - In contrast, dipoles are like p-orbitals (Figure 8.11b).[2]
  - Note that orbitals aren't polar; these analogies are given to illustrate phasing properties.
- You don't need a d-orbital to have a quadrupole, though — they exist in organic chemistry, too!
- Example of quadrupolar molecules in organic chemistry: Benzene (Figure 8.11c).
  - All C–H bonds have dipoles that cancel each other out.
  - However, the net transfer of electron density increases the electron density in the π-cloud, even as it depletes the electron density at the hydrogens.
- Example of quadrupolar molecules in organic chemistry: CO_2.

- Just like we have charge-dipole interactions, we can have **charge-quadrupole** interactions.

- **Ion-quadrupole** (interaction): An NCI between a cation and the face of a quadrupole.

  - Different atoms in the periodic table have different electronegativities, which we know thanks to Linus Pauling's building on Dmitri Mendeleev's foundation.
  - These are most commonly encountered when we have an arene and we bring in an ion on top.
    - Leads to a net-stabilizing interaction between the ion and the charged molecule.
  - See Anslyn and Dougherty (2006, pp. 181–83).

- Example: K^+ and benzene.

Figure 8.12: Ion-quadrupole interaction of K^+ and benzene.

  - These two species' gas phase binding enthalpy is $\Delta H = -19\,\text{kcal/mol}$.
  - For comparison, the gas phase binding enthalpy of K^+ and H_2O is $\Delta H = -18\,\text{kcal/mol}$.
    - This shows that ion-dipole and ion-quadrupole interactions are about similarly intense, even though benzene doesn't really have charged regions.
    - This is surprising, in Alex's opinion.
  - Reference: Sunner et al. (1981).

---

[2]Alex drew the dipole arrow in the inverse direction here by accident.

- Example: Ions and benzene.

$\mathbf{M^+}$	$\mathbf{-\Delta H}$
Li^+	38.3
Na^+	28.0
K^+	19.2
$NH_4{}^+$	19.3
$NMe_4{}^+$	9.4

Table 8.1: Ion-quadrupole interactions with benzene.

  - Ion-quadrupole interactions are subject to a size dependence.
  - Specifically, harder ions bind better to quadrupoles, and softer ones binds worse.
  - This is consistent with the inverse distance dependence in all of our interaction equations.[3]
  - See Dougherty (1996), specifically the section entitled "The Fundamental Interaction: Gas-Phase Studies" and refs. 2-7 cited therein.

- Example: The effect of arene substituents on ion-quadrupole interactions.

  - Look it up if we're so inclined!!
  - Alex goes deep into this topic many years, "but this year, I'm like f it."
  - Reference: Mecozzi et al. (1996).
    - By Dennis Dougherty, of Anslyn and Dougherty (2006)!

- Example: Cation-π interactions in biology.

  - Proteins stabilize charged species within them, even when solvated in a high dielectric (H_2O) at physiological pH (pH $\approx$ 7). How do they do this?
    - Water is great at solvating charged species, so to incorporate charged species, proteins have to pay a desolvation penalty.
    - Proteins definitely don't work by allowing water inside them: In fact, enzymes and proteins adopt a tertiary structure that often excludes water into the bulk.
    - Thus, the solution is that proteins use ion-quadrupole interactions (with aromatic residues).
  - Takeaway: Nonpolar, hydrophobic arenes can stabilize cations even in aqueous media.
  - Example: Would benzene or a carboxylate anion stabilize the trimethylammonium cation more?
    - Gas phase: The binding of trimethylammonium to benzene ($\Delta H = -19$ kcal/mol) is far more stabilizing than its binding to acetate ($\Delta H > 100$ kcal/mol).
    - In water: There is so much dielectric stabilization that binding to benzene is weaker ($\Delta H = -5.5$ kcal/mol), and binding to acetate doesn't do that much at all ($\Delta H = -2.2$ kcal/mol).
  - Corollary: Proteins can take carboxylic acids and modulate their pK_a over multiple log-units via enclosure in a hydrophobic environment.
  - Aside: This is a case of a biological phenomenon being explained by physical chemistry.
    - If you believe that biology lends greater insight into the physical world, good for you!
    - Alex, personally, would rather just look at physical systems though.
  - Reference: Dougherty (1996).
    - By Dennis Dougherty again!

---

[3]Note: HSAB (as a polarizability- and dispersion-directed phenomenon) does indeed favor the binding of "soft" benzene to "soft" ions, but the fact that we see this relationship here implies that electrostatics (i.e., classical ion-quadrupole attractions) are *more* active in determing benzene's affinity for cations (Dougherty, 1996, pp. 163–64).

- **Apparent** (interaction): An NCI that is frequently alluded to in the literature, real, and yet has no fundamental physical basis.

  – This is Alex's term; we won't hear about "apparent" interactions anywhere else.

  – Several subsets.

    I.) **π-π interactions**.
    II.) **Hydrogen bonds**.

- **π-π** (interaction): An NCI between two (typically aromatic) π-systems.

  – The physical basis for such interactions is probably very complex.

  – See Anslyn and Dougherty (2006, p. 184).

- There are three main types of π-π interaction geometries.

(a) Parallel stack.   (b) T-shaped.   (c) Slipped stack.
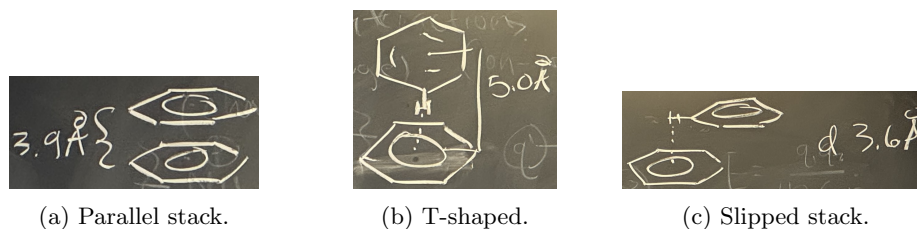
Figure 8.13: π-π interaction geometries.

1. The **parallel stack** of two centroids.
2. The **T-shaped geometry**.
3. The **slipped stack geometry**.

- **Parallel stack** (geometry): A π-π interaction geometry in which the two centroids stack their π-systems on top of each other.

  – The equilibrium distance is about 3.9 Å.

  – Important for electronic transport through adjacent π-systems, but not relevant to us today.

  – Energy: $\Delta H = -1.7$ kcal/mol.

- **T-shaped** (geometry): A π-π interaction geometry in which a region of negative electrostatic potential (the face of one ring) is in contact with a region of positive electrostatic potential (the edge of the other). *Also known as* **edge-to-face**.

  – The equilibrium distance is about 5.0 Å.

  – Energy: $\Delta H = -2.6$ kcal/mol.

    ■ This is more stable than the π-stack!

- **Slipped stack** (geometry): A π-π interaction geometry consisting of coplanar π-systems with a dislocation between the centroids. *Also known as* **displaced**.

  – The equilibrium distance is about 3.6 Å.

  – Energy: $\Delta H = -2.6$ kcal/mol.

    ■ Thus, this geometry is still more stable than the conventional parallel stack.

- Implication: There's nothing intrinsic about arenes that makes them want to π-π stack, and this phenomenon will be generalized to other types of species.

- There are two main models through which to view the phenomenon of π-π interactions.

    1. The **Wheeler-Houk model**.[4]
    2. The **Grimme dispersion model**.[5]
    3. As a possible third, Brent Iverson (UT-Austin) has many beautiful papers on π-π stacking.
        - The papers are a bit pedantic, but "I'm [Alex is] a pedant."

- **Wheeler-Houk** (model): A model of π-π stacking predicated on direct interactions.

    - Wheeler and Houk found (computationally and experimentally) that it's not the π-faces that interact, but the substituents on the arene that have a quadrupolar interaction with the other arene below.
    - Their model works for substituted arenes.
    - Reference: Wheeler and Houk (2008).

- **Grimme dispersion** (model): A model of π-π stacking which posits that more extended π-sytems facilitate and maximize disperson.

    - This model is more descriptive as we get to larger π-systems.

- This concludes our discussion of π-π interactions; we now move onto hydrogen bonding.

- **Hydrogen bond**: An NCI between a donor of the form A−H and an acceptor of the form B. *Schematic*
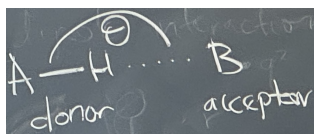

Figure 8.14: Schematic of a hydrogen bond.

- This is a really ill-defined suite of physical phenomena that — all the same — we humans can classify really well.
    - There is (at least) an electrostatic component, an ion-dipole moment, some covalency, etc.
- Review (the sacred text of hydrogen bonds): Steiner (2002).
- There is a continuum of hydrogen bonds — from strong to moderate to weak — that is characterized by the distance between the H-bond donor and acceptor as well as the angle, resulting in an energy of stabilization.
    - Strong: 1.2-1.5 Å and $\theta = 175° - 180°$, $\Delta E = 14 - 40$ kcal/mol.
        - Large degree of covalency.
    - Moderate: 1.5-2.2 Å and $\theta = 130° - 180°$, $\Delta E = 4 - 15$ kcal/mol.
        - More electrostatic in nature.
    - Weak: 2.3-3.2 Å and $\theta = 90° - 150°$, $\Delta E < 4$ kcal/mol.
        - Even 4 kcal/mol can upend selectivity.
        - Thus, even these are important!

---

[4] "HOWK"
[5] "shteh-FAHN GRIM-uh," a professor at the University of Bonne.