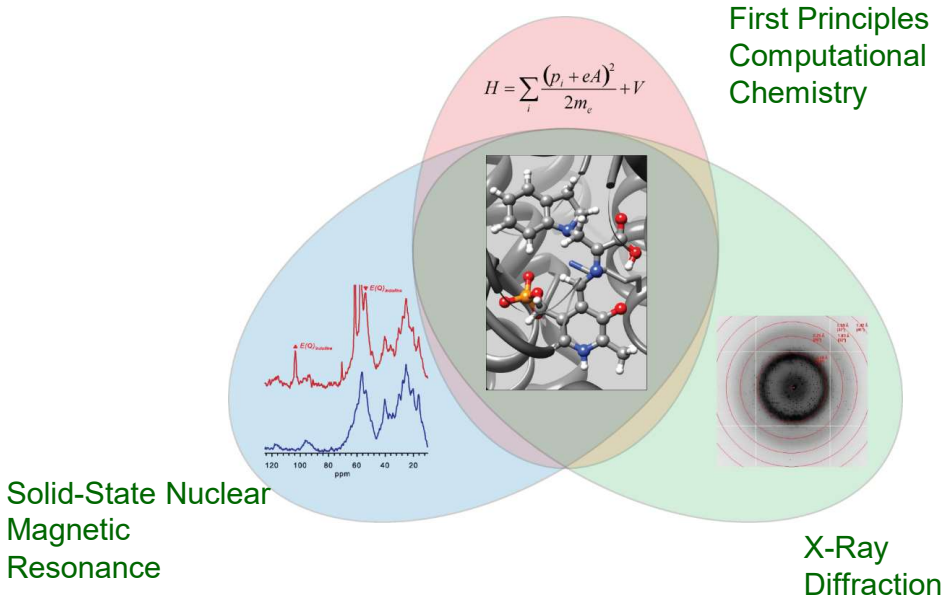
NMR Crystallography:

at the interface of solid-state NMR, X-ray diffraction, and first-principles computational chemistry

Len Mueller

Department of Chemistry

UC Riverside



NMR and Crystallography

- Crystallography
  - the study of crystal structure
  - the arrangement of atoms in crystals
- NMR complements diffraction methods
  - long range order vs. short range order/local chemical structure and dynamics
- Combination
  - chemically-detailed crystal structures
  - insight into relationship between structure, dynamics, reactivity, and function

Determination of molecular symmetry in crystalline naphthalene using solid-state NMR

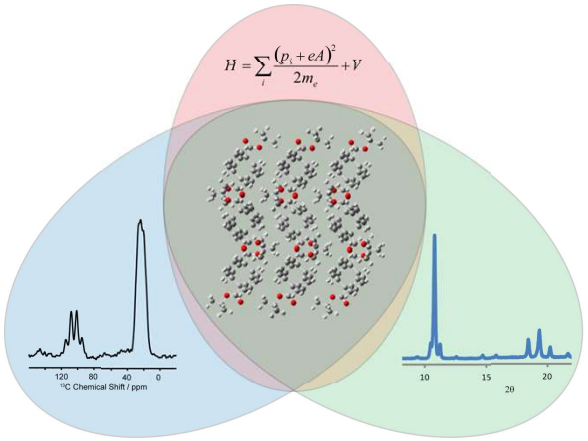
Julio C. Facelli*† & David M. Grant*

* Department of Chemistry and † Utah Supercomputing Institute, University of Utah, Salt Lake City, Utah 84112, USA

NATURE · VOL 365 · 23 SEPTEMBER 1993

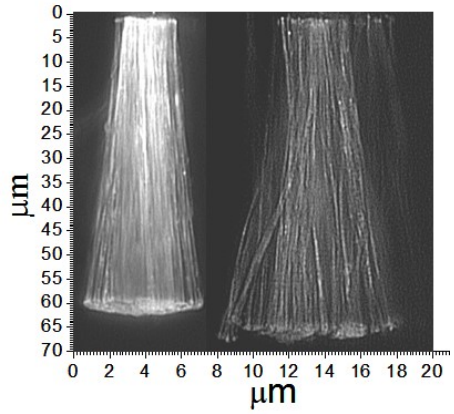
The use of solid-state NMR methods for refining structural data should be especially attractive in the study of biomolecules with molecular weights in the range $10–20 \times 10^3$, where diffraction data have even larger structural errors. Many types of crystalline imperfections, which degrade diffraction data, have no effect on chemical-shift data that are not sensitive to such imperfections as translational disorder or the absence and/or occlusion of a given molecular impurity. A ^{13}C-labelled atom in a large molecule may be observed with the magnification factor of the isotopic enrichment, allowing one to focus on the active sites of larger molecular systems while avoiding spectral interference from less relevant parts of the molecule. □

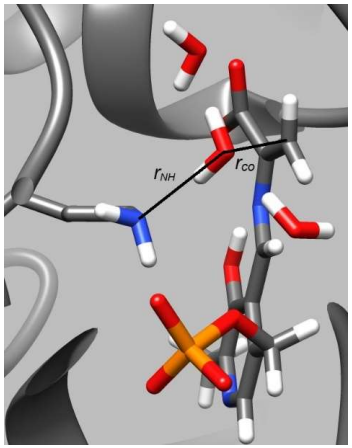
First pairing of NMR, X-ray, and ab initio comp. chemistry



Outline

- The Tools of NMR Crystallography by way of two applications …
  - Materials Science: Photomechanical Materials
  - Structural Biology: Enzyme Active Sites



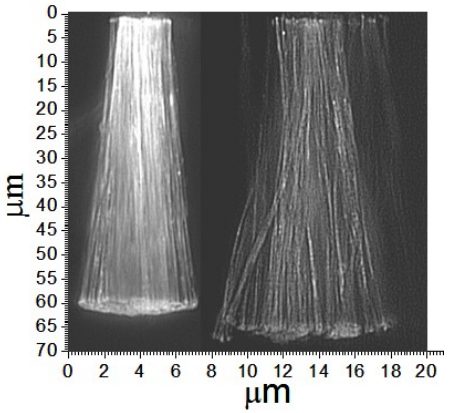


Along the Way … A Few Helpful Tools and (STRONG) Opinions

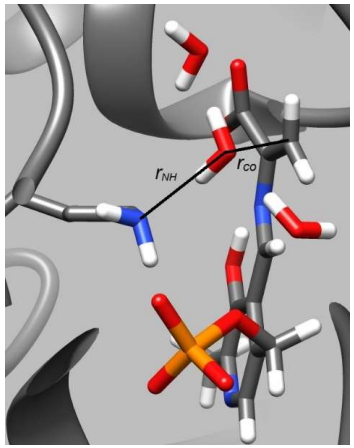
- First principles calculations, choice of functional, and *a priori* linear rescaling
- Statistics and the assignment of model probabilities
- Visualizing tensors with TensorView
- Common errors in how spherical tensor rotations are applied

Outline

- Materials Science: Photomechanical Materials



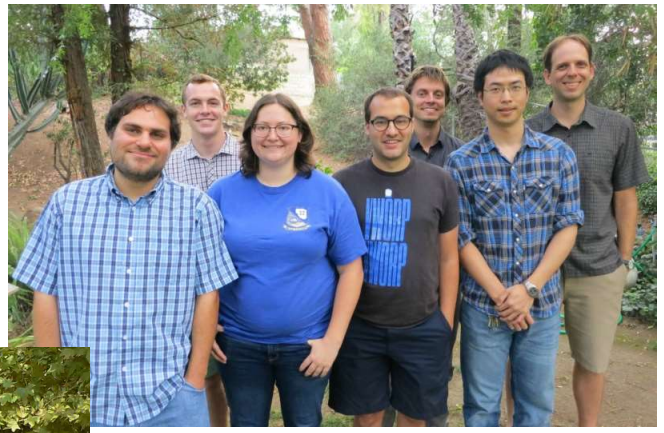
- Structural Biology: Enzyme Active Sites

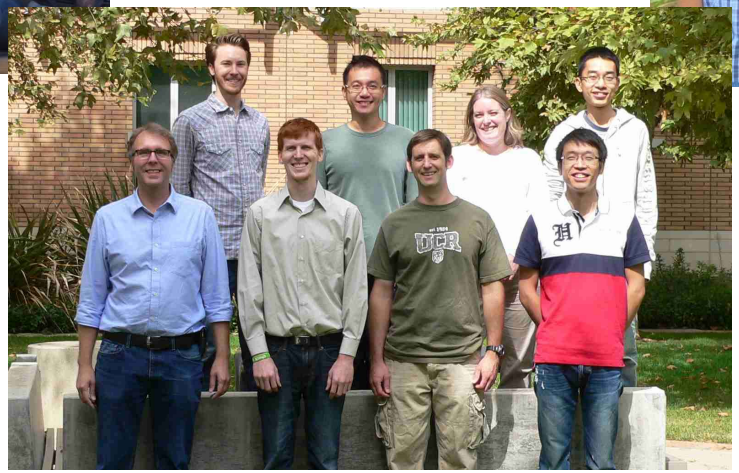


Photomechanical Materials Group

Beran Lab: *Theory and Computational Chemistry*

Bardeen Lab:
Solid-State Photochemistry





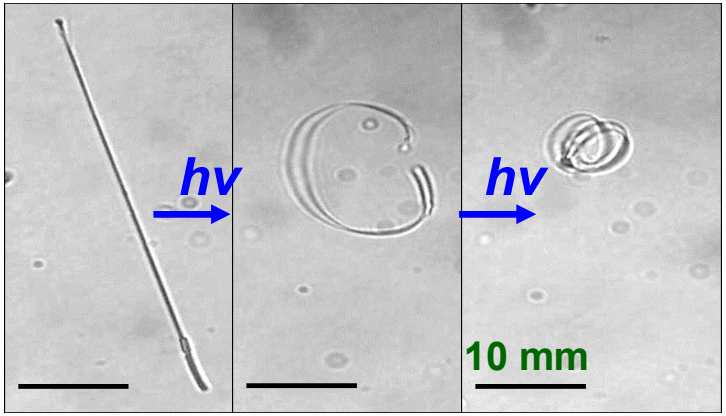
Mueller Lab: *NMR and NMR Crystallography*

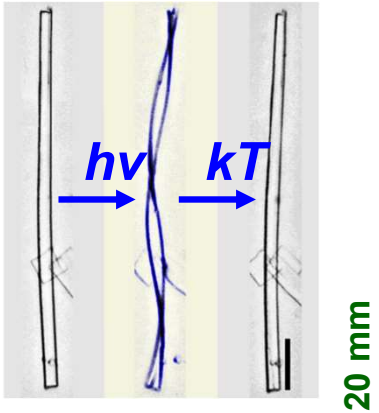
CrystEngComm **18**, 7319 (2016)
Chem Sci **12,** 453 (2021)
Chem Sci **14,** 937 (2023)

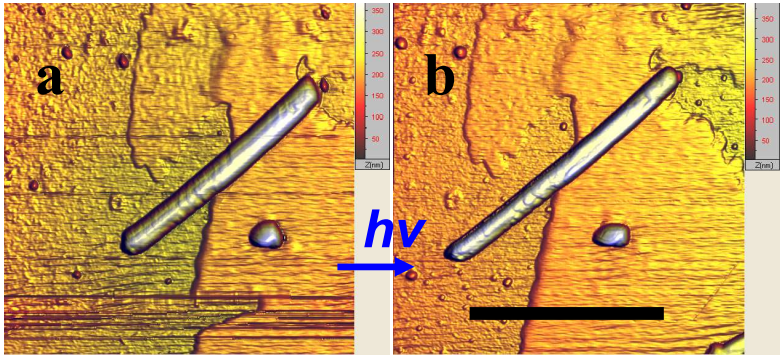
Funding: NSF

Photomechanical Materials

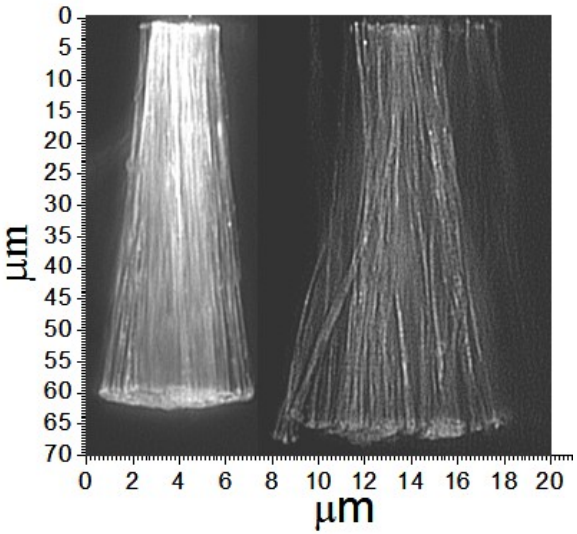
- Use photochemical reactions to turn photons into mechanical work
- Goal: atomic-level basis for the macroscopic response







9-Tertbutyl-Anthracene Ester (9TBAE) Nanorods



- 200 nm x 60 µm nanorods
  - single crystal (TEM)
  - self-organize in anodic alumina oxide (AAO) templates
- Expand ~8%
- [4+4] photodimerization (365 nm)

AAO template

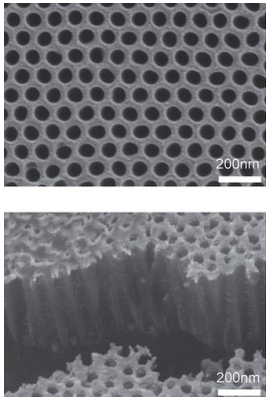
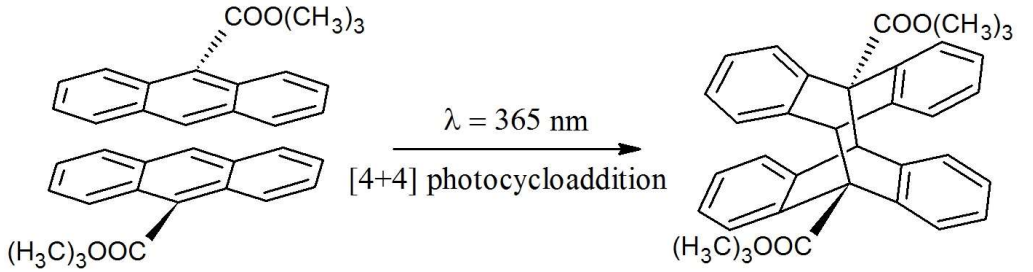
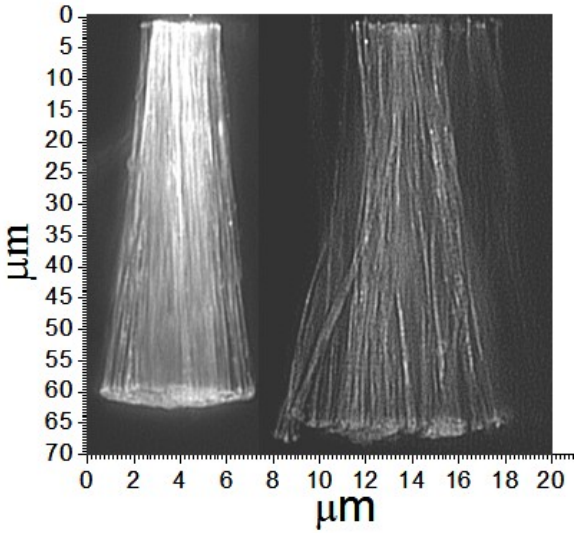


Image: Bae et al, RSC Adv 5 (2015)

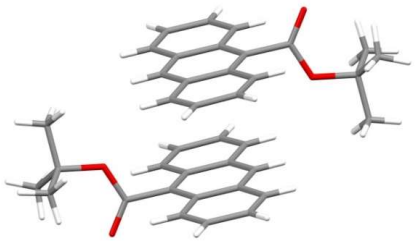


9TBAE **9TBAE head to tail photodimer**

9-Tertbutyl-Anthracene Ester (9TBAE) Nanorods

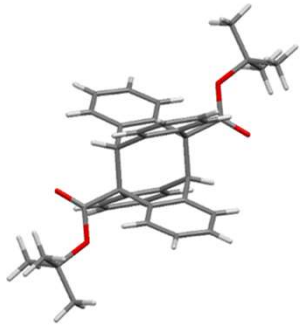


- 200 nm x 60 µm nanorods
  - single crystal (TEM)
  - self-organize in anodic alumina oxide (AAO) templates
- Expand ~8%
- [4+4] photodimerization (365 nm)



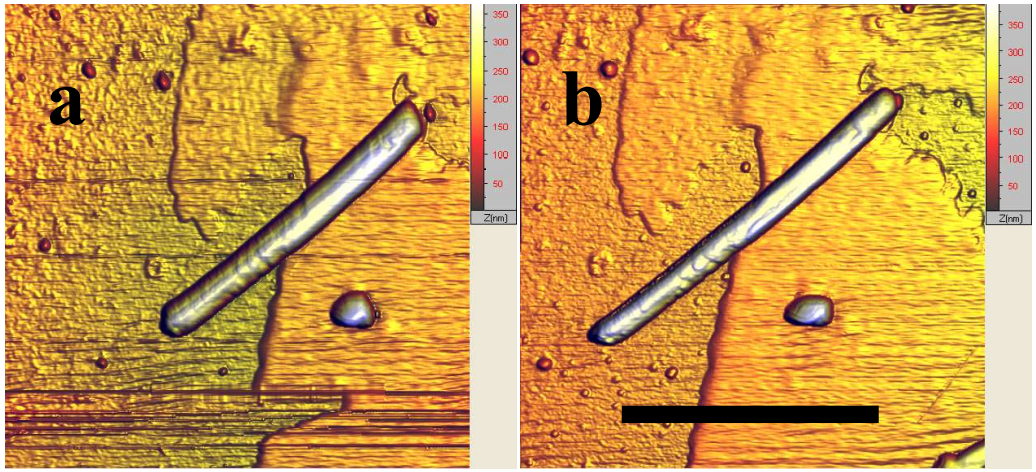
in solution

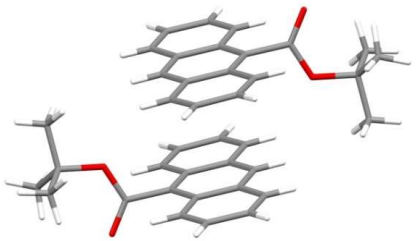
then crystalized



Solution-grown dimer (SGD)

9-Tertbutyl-Anthracene Ester (9TBAE) Nanorods



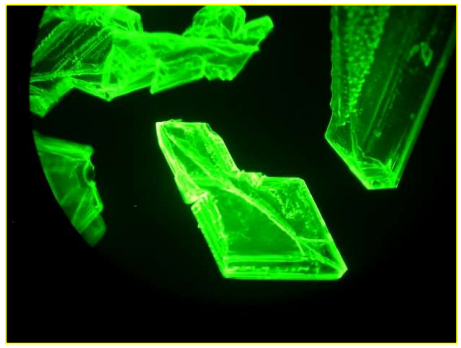


in solid-state

?

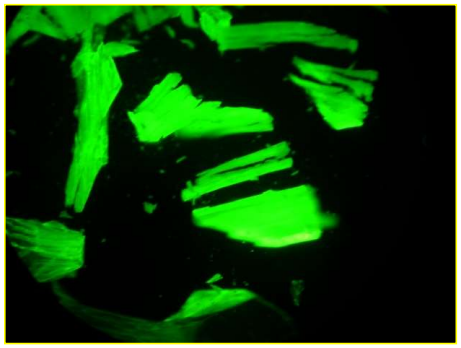
unknown molecular conformation and crystal packing

Photoresponse of Bulk Crystals

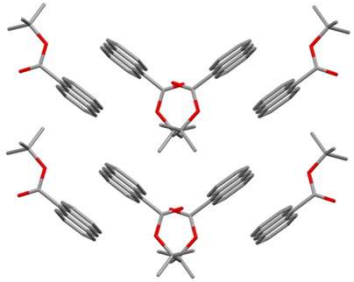


single crystals ~ 0.5 mm

λ = *365 nm*



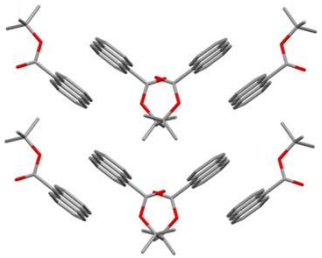
under irradiation



?

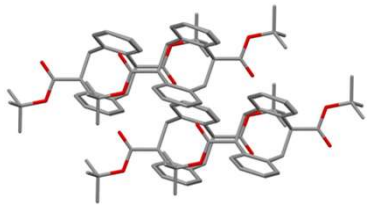
Macroscopic 9TBAE crystals shatter – single crystal X-ray is out

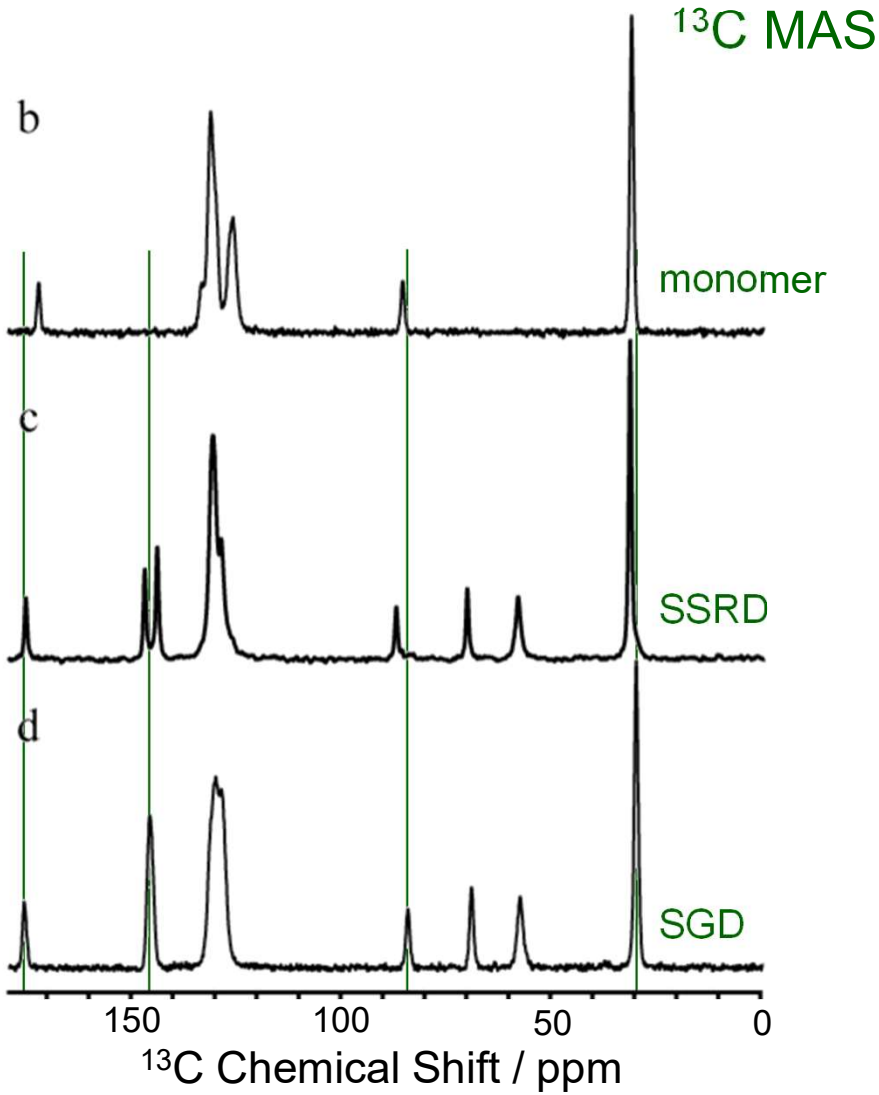
Monomer: herringbone

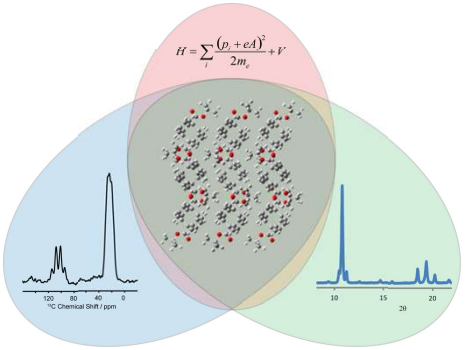


Solid-state reacted dimer (SSRD): ?

Solution-grown dimer (SGD): parallel layer structure

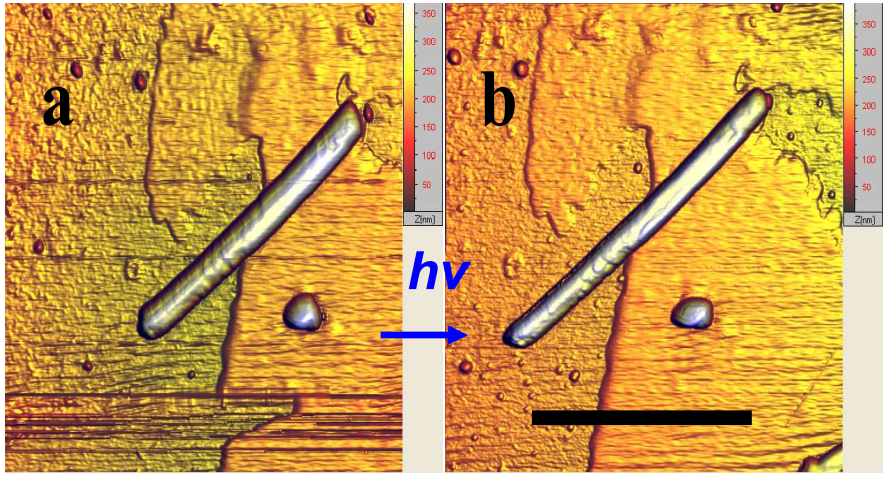


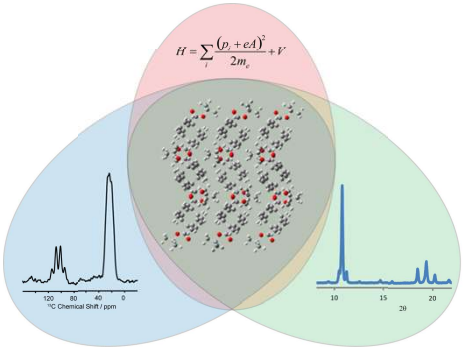




Challenge to NMR Crystallography

Identify and characterize the crystal structure of the metastable solid-state reacted dimer and provide a rationale for the photomechanical response





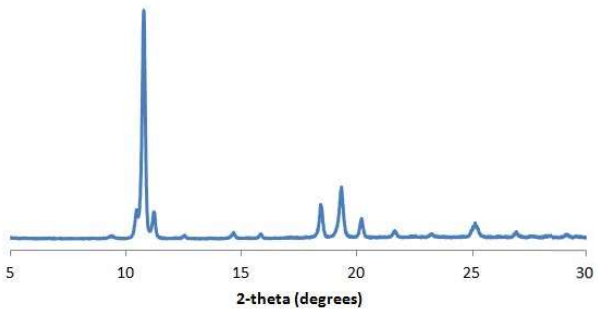
NMR Crystallography

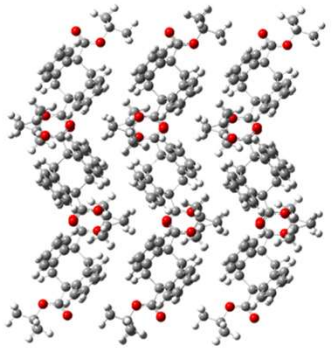
Requirements:

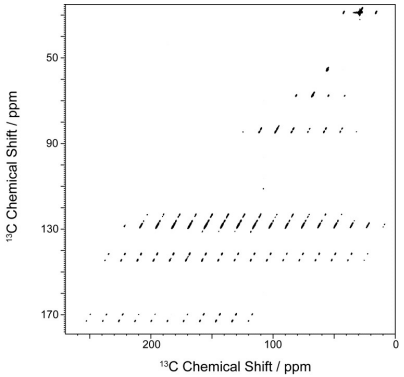
1. A good problem!

NMR Crystallography of 9TBAE Nanorods

- Structure of the SSRD
  - Powder X-ray
  - Solid-state NMR spectroscopy
  - Computational chemistry
- Mechanism of expansion
  - Orient the reactant and product unit cells with respect to the nanorod axis



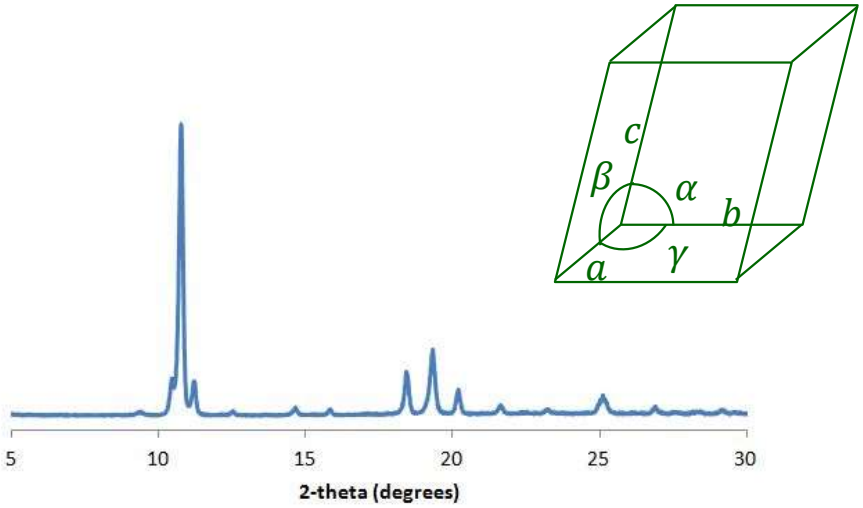




Powder X-Ray Diffraction of Solid-State Reacted Dimer

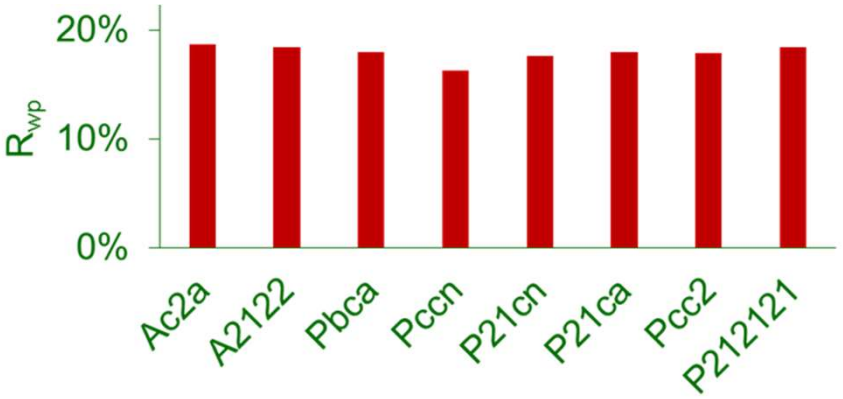
- Powder X-ray
  - Indexing
  - Pawley refinement
  - Introduce and optimize molecular geometry and packing
  - Rietveld refinement
  - Quantum Espresso solid-state DFT optimization
  - Final ranking by R_{wp} (weighted residuals)

- Suite of programs within Materials Studio



Powder X-Ray Diffraction of Solid-State Reacted Dimer

- Powder X-ray
  - Indexing
  - Pawley refinement
  - Introduce and optimize molecular geometry and packing
  - Rietveld refinement
  - Quantum Espresso solid-state DFT optimization
  - Final ranking by R_{wp} (weighted residuals)



- Identify eight candidate crystal structures for SSRD consistent with the PXRD – all orthorhombic crystal system

8 Candidate Crystals/2 Groupings

Small Torsions: Similar to monomer crystal

Small Asym:
$P2_12_12_1$
$P2_1cn$

1-sided:
Aba2
$A2_122_1$

Inversion:
Pbca
Pccn

Large Asym:
$P2_1ca$
Pcc2

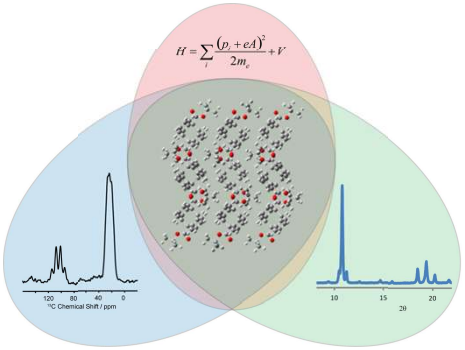
Similar to SGD

Equivalent unit cell dimensions

Differ in symmetry of the molecular packing and ester torsion

Classify into 2 groups of structures

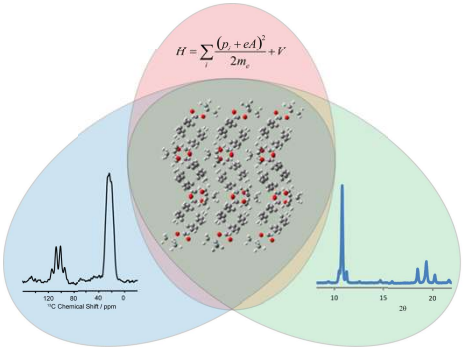
Progress. But can we distinguish these sets?



NMR Crystallography

Requirements:

1. A good problem!
2. Candidate structures



Requirements:

1. A good problem!
2. Candidate structures

i. Powder diffraction / Rietveld

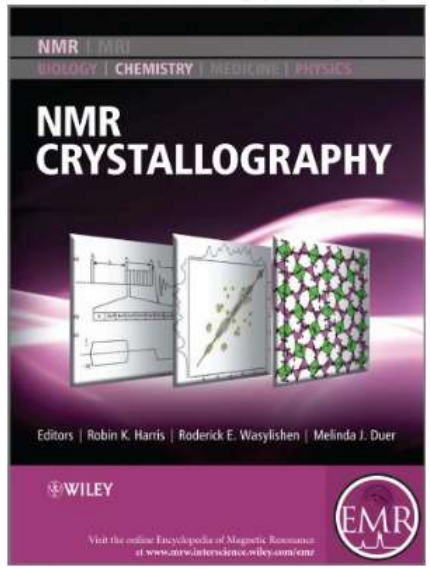
ii. Crystal structure/polymorph prediction

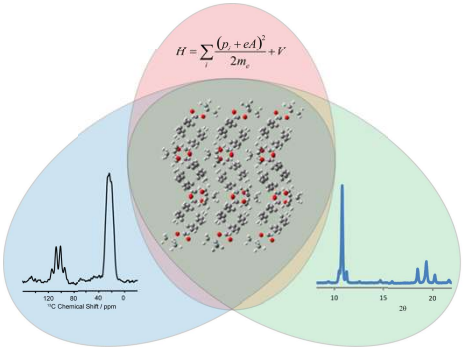
iii. Ab Initio Random Structure Searching (AIRSS)

iv. Powder NMR structural restraints, e.g., NMR spin diffusion, through-bond connectivity …

v. …

See the many approaches in Harris, Wasylishen, and Duer





Requirements:

1. A good problem!
2. Candidate structures

i. Powder diffraction / Rietveld

ii. Crystal structure/polymorph prediction

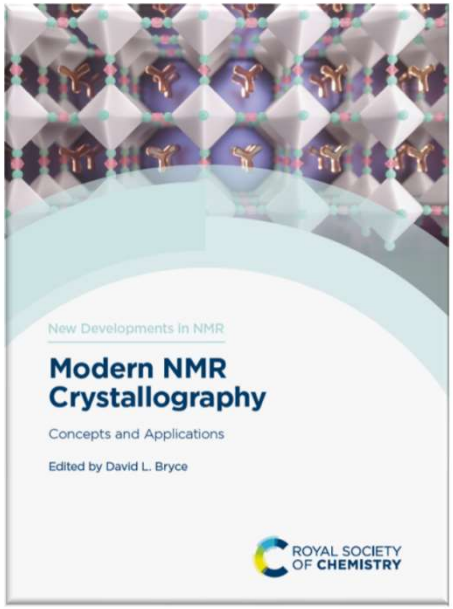
iii. Ab Initio Random Structure Searching (AIRSS)

iv. Powder NMR structural restraints, e.g., NMR spin diffusion, through-bond connectivity …

v. …

See the many approaches in Harris, Wasylishen, and Duer

… and Bryce (2024)



8 Candidate Crystals/2 Groupings

Small Torsions: Similar to monomer crystal

Small Asym:
$P2_12_12_1$
$P2_1cn$

1-sided:
Aba2
$A2_122_1$

Inversion:
Pbca
Pccn

Large Asym:
$P2_1ca$
Pcc2

Similar to SGD

Equivalent unit cell dimensions

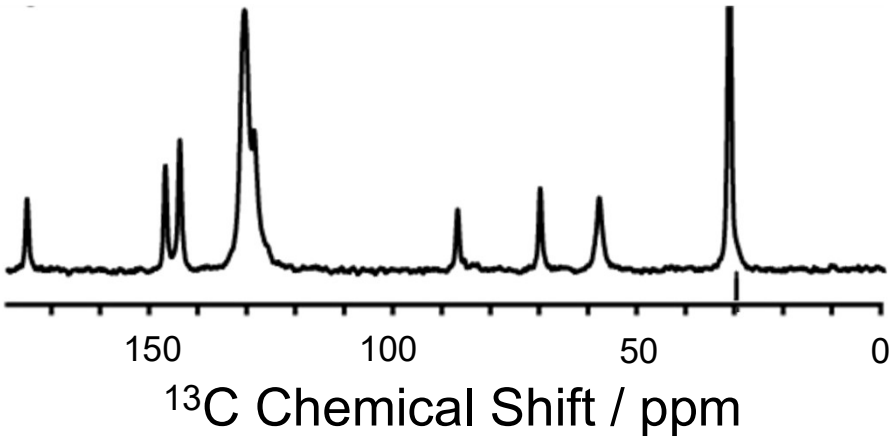
Differ in symmetry of the molecular packing and ester torsion

Classify into 2 groups of structures

Progress. But can we distinguish these sets?

NMR Crystallography of 9TBAE Nanorods

- Structure of the SSRD
  - Powder X-ray
  - Solid-state NMR spectroscopy
  - Computational chemistry
- Use the chemical shifts as restraints in a first-principles screening to sort out the candidate structures

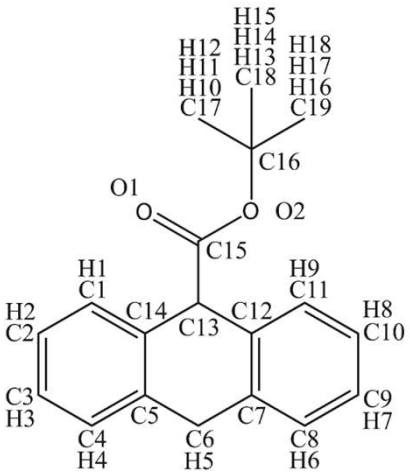


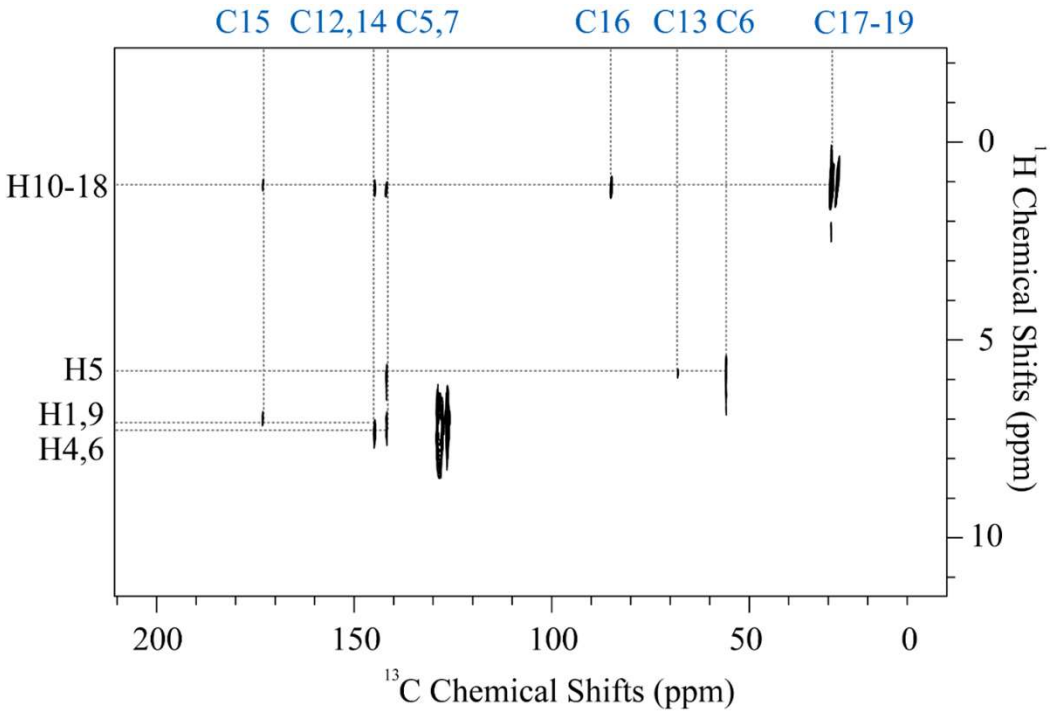
SSNMR of 9TBAE Nanorods

Isotropic shifts:

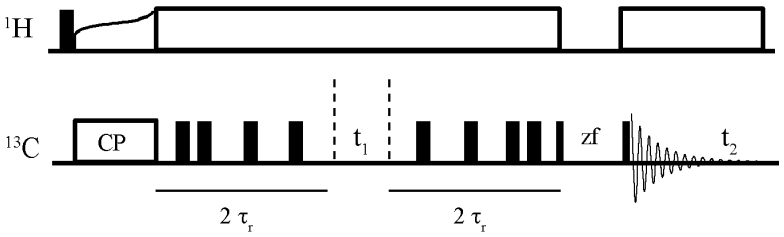
^{1}H-^{13}C HETCOR

(50 kHz MAS, 14.1 T)



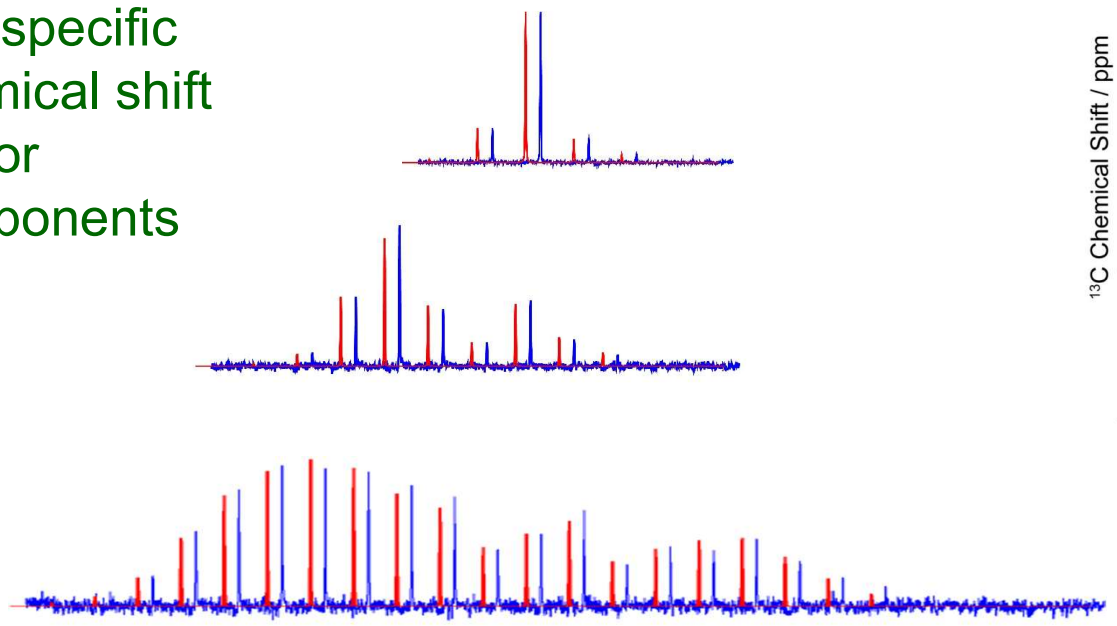


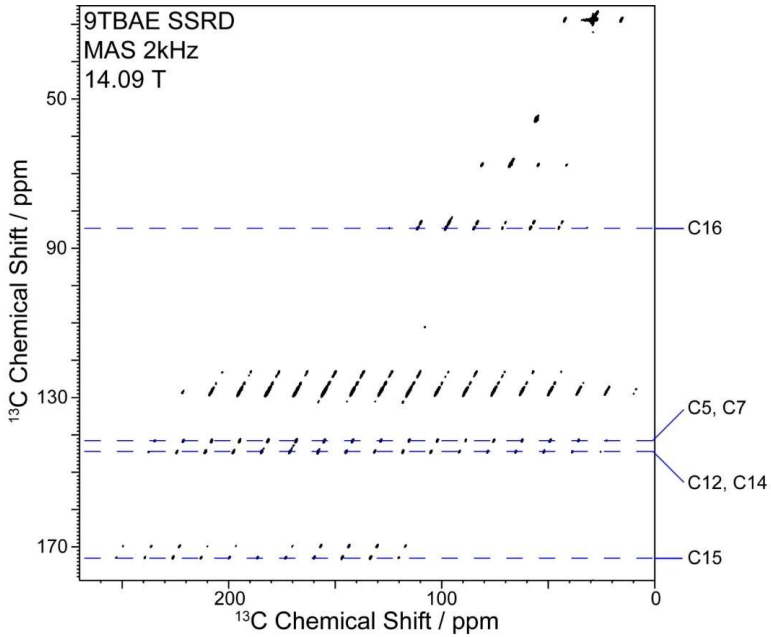
TOSS-deTOSS



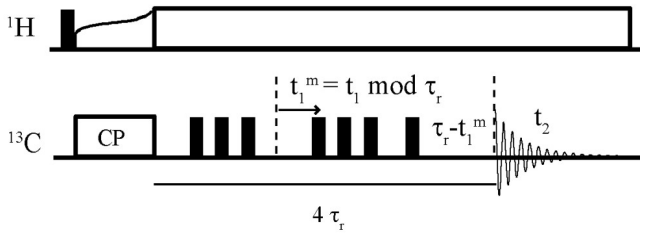
Kolbert and Griffin, *CPL* 166, 87 (1990)

Site-specific chemical shift tensor components

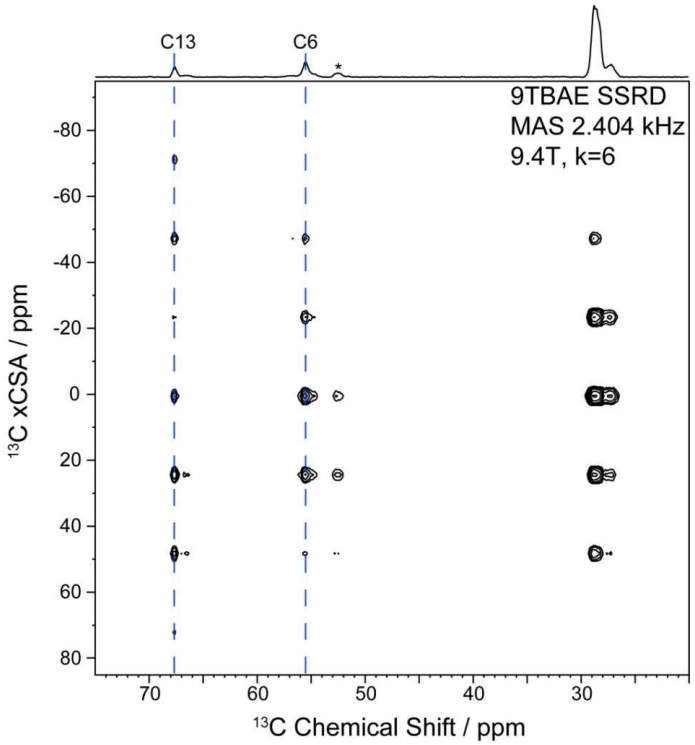


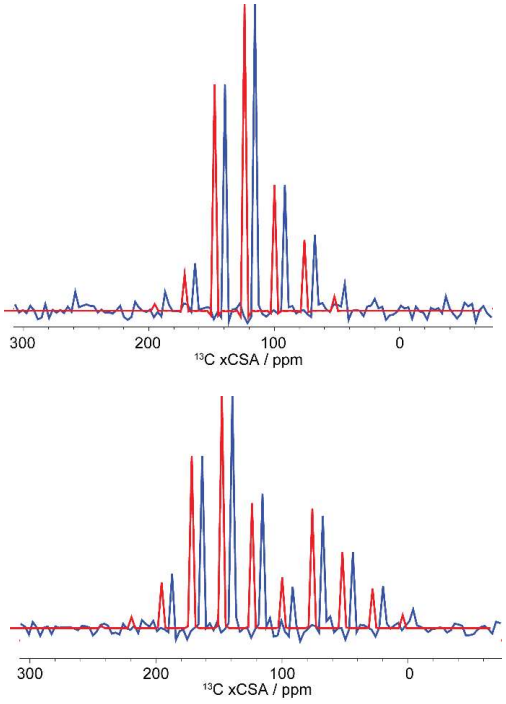


xCSA



Hung and Gan, *JMR* 213, 196 (2011)

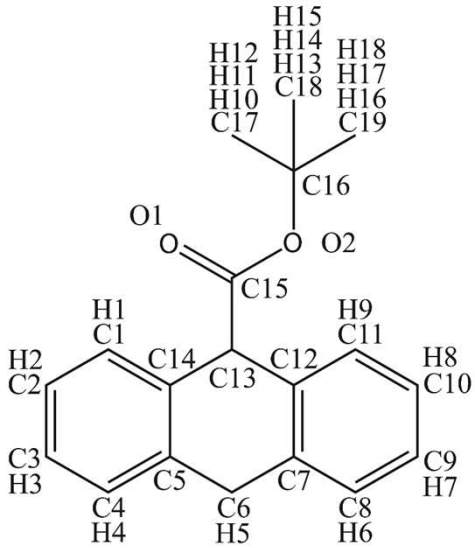


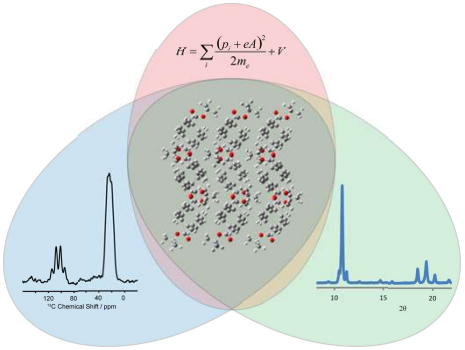


SSRD: Experimental NMR Shifts

Tensors	δ_{11} / ppm	δ_{22} / ppm	δ_{33} / ppm
C5, C7	13.2	176.2	235.9
C6	46.1	58.6	63.0
C12/C154	21,8	174.6	238.0
C13	49.7	74.2	80.3
C15	114.9	141.2	263.2
C16	29.1	110.0	115.6

Isotropic	δ_{iso} / ppm
H1/H9	6.88
H4/H6	6.72
H5	5.49
H-Me	0.80
C-Me	29.2





NMR Crystallography

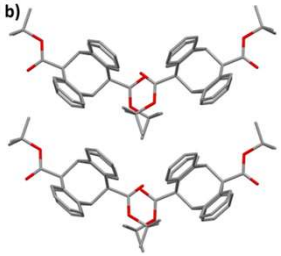
Requirements:

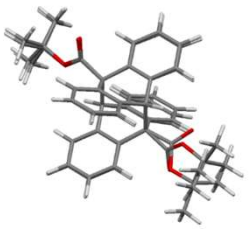
1. A good problem!
2. Candidate structures
3. NMR restraints

Selecting Crystal Structures using First-Principles Chemical Shifts

- Use the chemical shifts as restraints in a first-principles screening
- 8 candidate crystal structures: calculate shifts for each and rank by agreement with experiment
- Requires high-precision and high-accuracy first-principles chemical shift calculations for solid-state structures
- Quantitative ranking of structures

$$H = \sum_i \frac{\left(p_i + eA\right)^2}{2m_e} + V$$



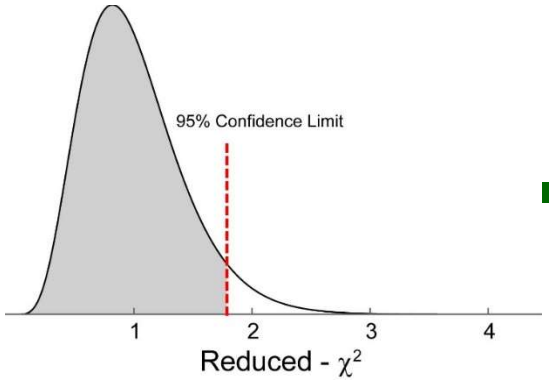


$$\chi^2 = \frac{1}{N}\sum_i \frac{\left(\delta_i^{\text{model}} - \delta_i^{\text{exp}}\right)^2}{\sigma_i^2}$$

Two Essential Components of NMRX

- First principles computational chemistry (DFT)

$$H = \sum_i \frac{\left(p_i + eA\right)^2}{2m_e} + V$$



$$\chi^2 = \frac{1}{N}\sum_i \frac{\left(\delta_i^{\text{model}} - \delta_i^{\text{exp}}\right)^2}{\sigma_i^2}$$

- Quantitative statistics

First Principles Computational Chemistry

The molecular Hamiltonian

$$H = -\sum_{A}^{\text{nuc}} \frac{1}{2M_A}\nabla_A^2 - \sum_{i}^{\text{elec}} \frac{1}{2m_i}\nabla_i^2 - \sum_{A}^{\text{nuc}}\sum_{i}^{\text{elec}} \frac{Z_A}{r_{iA}} + \sum_{i}^{\text{elec}}\sum_{j>i}^{\text{elec}} \frac{1}{r_{ij}} + \sum_{A}^{\text{nuc}}\sum_{B>A}^{\text{nuc}} \frac{Z_A Z_B}{R_{AB}}$$

- Kinetic energy for each nucleus
- Kinetic energy for each electron
- Attraction of each electron to each nucleus
- Repulsion between each pair of electrons
- Repulsion between each pair of nuclei

In a magnetic field:

$$p_i \rightarrow p_i + eA$$

Solving the Schrödinger equation gives all of the molecular properties: energy, dipole moment, **chemical shifts …**

* Adapted from lecture notes by Prof. Greg Beran, UCR

First Principles Computational Chemistry

The Schrödinger equation with the molecular Hamiltonian is too complicated to solve exactly

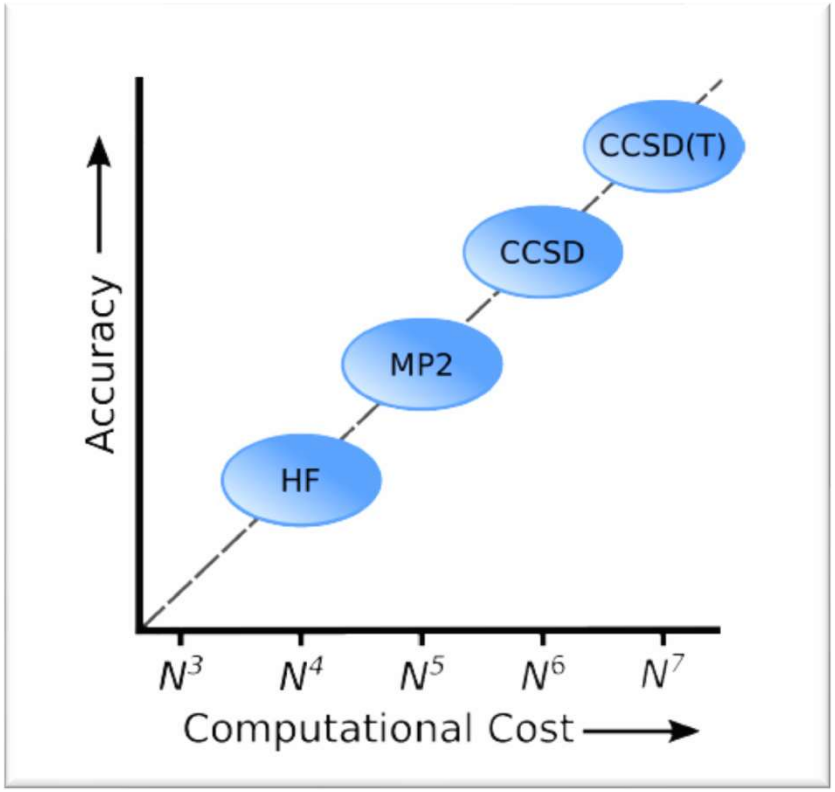
Approximate approach:

1. Born-Oppenheimer approximation – treat nuclei as fixed
2. Solve the electronic part of the Schrödinger equation using various levels of perturbation theory
   a. Wavefunction methods
   b. Density functional theory

$$H\psi(r,R) = E\psi(r,R)$$

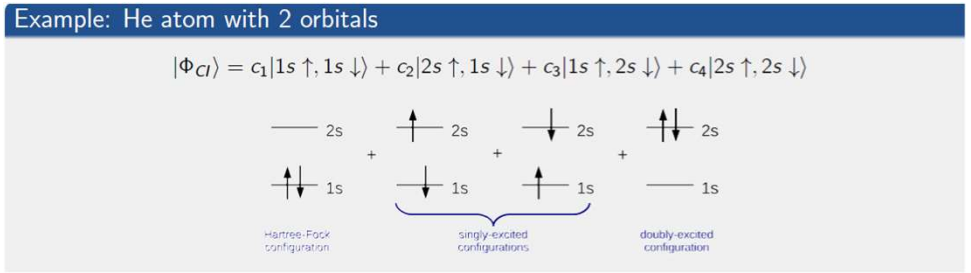
* Adapted from lecture notes by Prof. Greg Beran, UCR

Wavefunction Methods



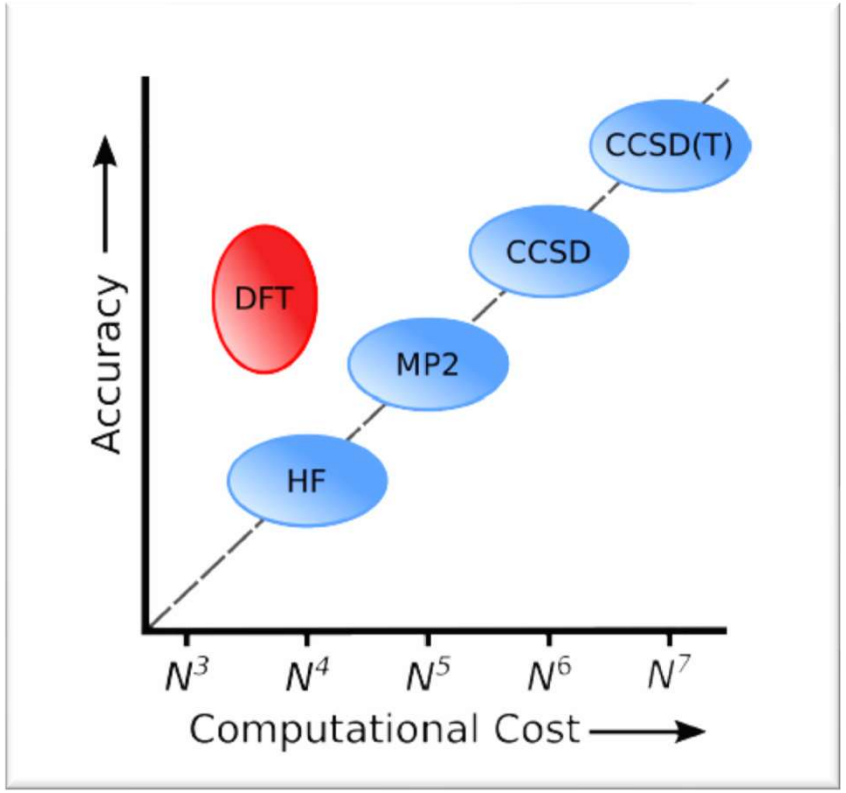
* Adapted from lecture notes by Prof. Greg Beran, UCR

- All of these methods seek to solve the electronic part of the Schrödinger equation
- Hartree-Fock (HF) is the simplest method, but not very accurate
  - Mean field approximation, ignores electron correlation $\psi(r_1, r_2) = \varphi(r_1)\varphi(r_2)$
- Higher methods introduce correlation through expanded bases



- CCSD(T) is the “gold standard" of quantum chemistry. Practical upper limit of accuracy.
- Computational cost grows steeply for better methods.

Density Functional Theory



* Adapted from lecture notes by Prof. Greg Beran, UCR

- DFT has HF-like cost, but significantly better accuracy
- Hohenberg-Kohn Theorem
  - There exists a 1:1 mapping between electron density ρ(r) and energy: E[ρ(r)]
  - Problem: we don't know what the mapping is
- Kohn-Sham DFT provides a workable solution for approximate density functionals
  - LDA, PBE, PBE0, B3LYP
- Note: Standard density functionals do not describe van der Waals dispersion, so should always augment with a dispersion correction
  - Grimme's D3, D4; Tkatchenko-Scheffler (TS) or Many-body Dispersion (MBD) …

Basis Sets

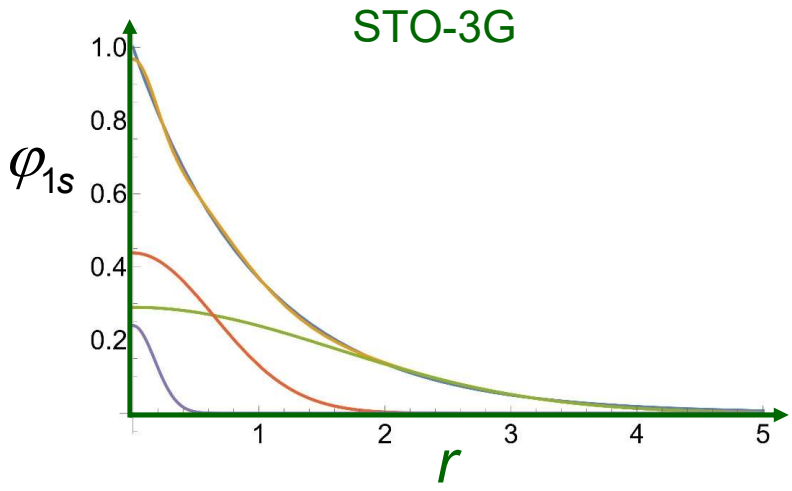
Gaussian bases frequently used in molecular problems express each MO, $|\varphi\rangle$, as a linear combination of AOs, $|\chi_n\rangle$

$$|\varphi\rangle = c_1|\chi_1\rangle + c_2|\chi_2\rangle + c_3|\chi_3\rangle + \ldots + c_n|\chi_n\rangle$$

Each atomic orbital $\chi(r) = f(r)e^{-a|r|}$

represented by a sum of

Gaussian functions $g(r)e^{-ar^2}$



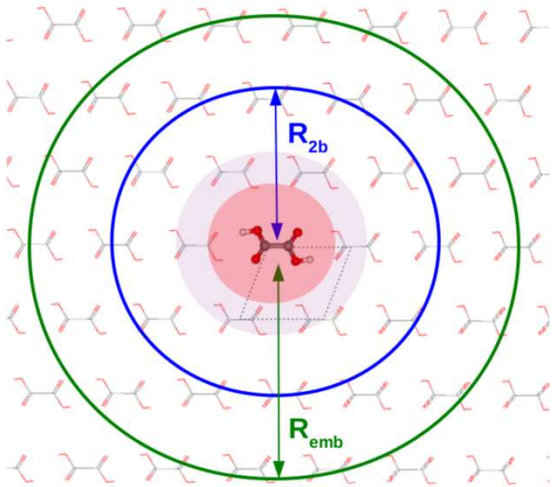
* Adapted from lecture notes by Prof. Greg Beran, UCR

Gaussian Basis Set Primer

Family	Basis Names	Comments
Minimal	STO-3G, STO-6G	DON'T USE!
Pople	6-31G(d), 6-31G(d,p) 6-311G(d), 6-311G(d,p) 6-311+G(d), 6-311++G(d,p)	Double-ζ, Smallest decent sets. Triple-ζ, moderately larger. The "+" adds diffuse basis functions.
Ahlrichs	def2-SVP, def2-TZVP, def2-QZVP	Double, triple, and quadruple-ζ. Good for DFT.
Jensen	pc-n (where $n = 1, 2, 3, 4$)	Another hierarchy that's good for DFT.
Dunning	cc-pVXZ, aug-cc-pVXZ (where X = D, T, Q)	Good hierarchies for MP2 and coupled cluster. The "aug-" adds diffuse basis functions.

* Adapted from lecture notes by Prof. Greg Beran, UCR

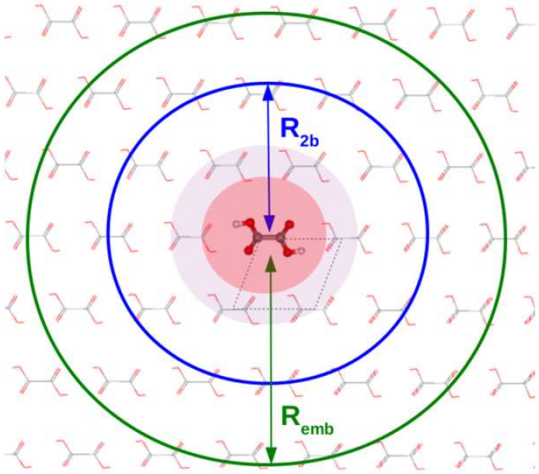
Solid-State Computational Chemistry



- Plane-wave methods
  - Plane-wave basis periodic on the crystal lattice
  - CASTEP, Quantum-Espresso, CPMD
  - Expensive to use hybrid functionals

    … but Dracinsky, Unzueta, and Beran have a nice solution to this (*PCCP* 2019)
- Cluster/Fragment-based approaches
  - Build large clusters to mimic the solid-state
  - Atom-centered Gaussian orbitals
  - Hybrid functionals more economical
  - Convergence: need large clusters

Hybrid Many-Body Interaction Fragment Approach



- Developed by Greg Beran (UCR)
- Efficient, fragment-based approach
- Intrinsically parallelizable
- Builds large clusters (30+ Å) to mimic the solid-state
- Atom-centered Gaussian orbitals allows DFT with hybrid functionals
- Highly accurate for NMR chemical shifts

References
Hartman & Beran, *JCTC* **10**, 4682 (2014)
Hartman, Monaco, Schatschneider, Beran. *JCP* **143**, 102809 (2015)
Hartman, Kudla, Day, Mueller, Beran, *PCCP* **18**, 21686 (2016)
Harman, Neubauer, Caulkins, Mueller, Beran, *JBNMR* **62**, 327 (2016)
Hartman, Balaji, Beran, *JCTC* **13**, 6043 (2017)
Dracinsky, Unzueta, Beran, *PCCP* **21**, 14992 (2019)

Benchmarks

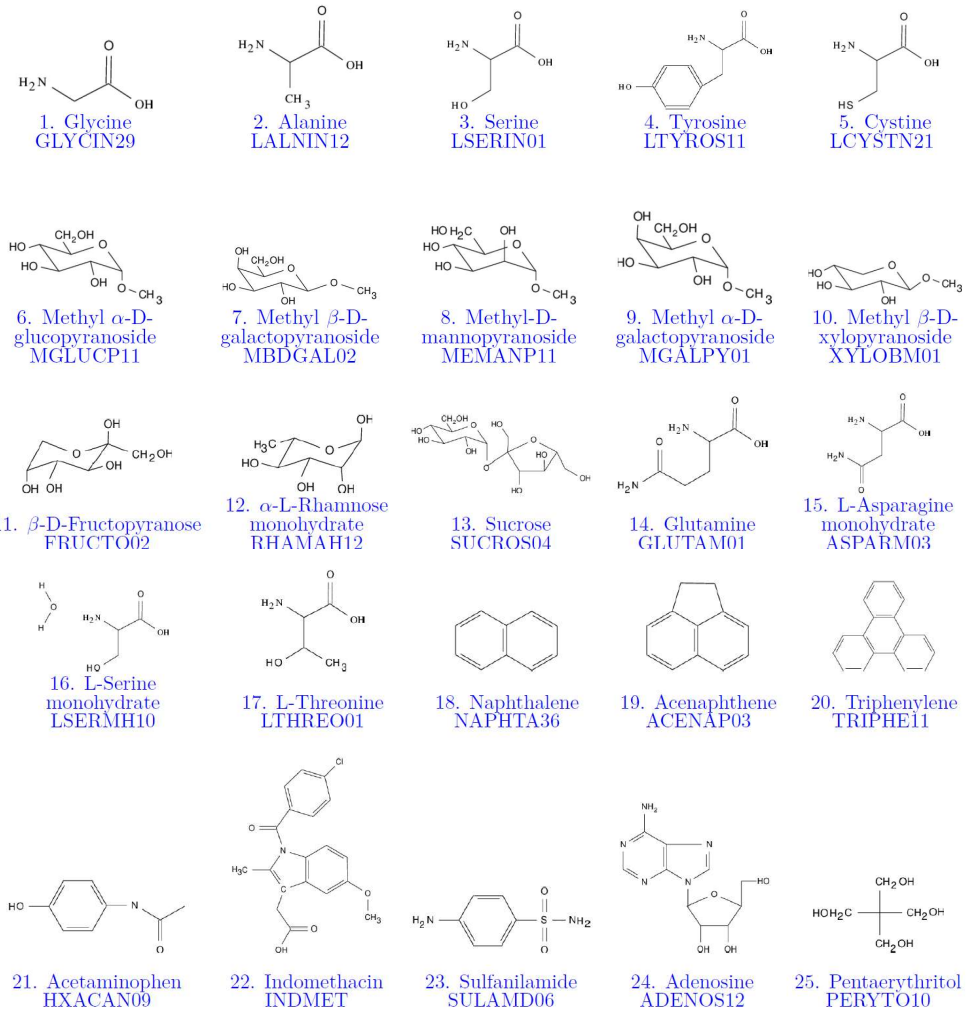
$$\delta_i = m\sigma_i + \sigma_{ref}$$

- Linear rescaling parameters are constants determined from benchmarks, not adjustable parameters
- Allow absolute, not just relative, comparison of theory and experiment

^{13}C isotropic shifts

Functional	RMS Error (ppm)	
GIPAW-PBE	2.1	CASTEP
PBE	2.0	Non-Hybrids
BLYP	2.1	
BP86	2.1	
O-PBE	1.8	
PBE0	1.4	Hybrids
B97-2	1.4	
B3PW91	1.4	
B3LYP	1.4	

Test Set



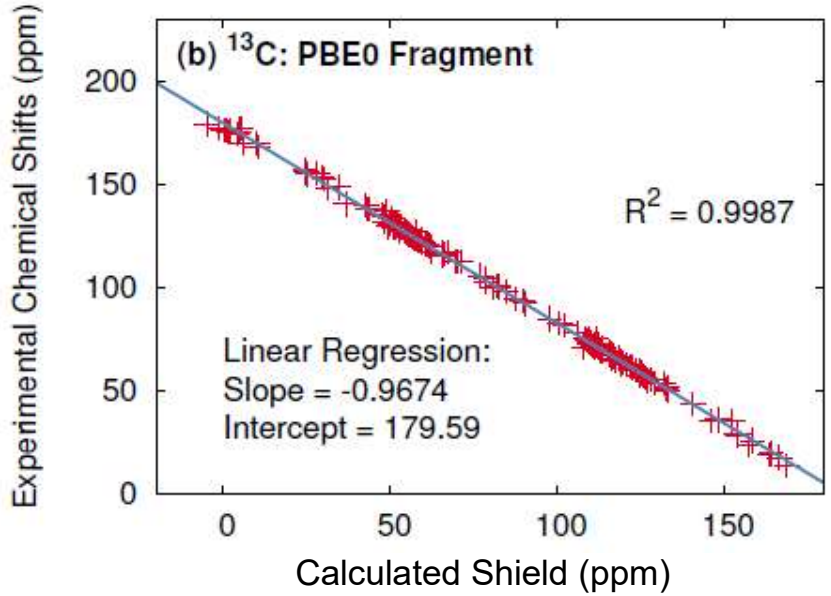
Benchmarks

$$\delta_i = m\sigma_i + \sigma_{ref}$$

- Linear rescaling parameters are constants determined from benchmarks, not adjustable parameters
- Allow absolute, not just relative, comparison of theory and experiment

^{13}C isotropic shifts

Functional	RMS Error (ppm)	
GIPAW-PBE	2.1	CASTEP
PBE	2.0	Non-Hybrids
BLYP	2.1	
BP86	2.1	
O-PBE	1.8	
PBE0	1.4	Hybrids
B97-2	1.4	
B3PW91	1.4	
B3LYP	1.4	



References:
Hartman & Beran, *JCTC* **10**, 4682 (2014)
Hartman, Monaco, Schatschneider, & Beran. *JCP* **143**, 102809 (2015)
Hartman, Kudla, Day, Mueller, Beran, *PCCP* **18**, 21686 (2016)

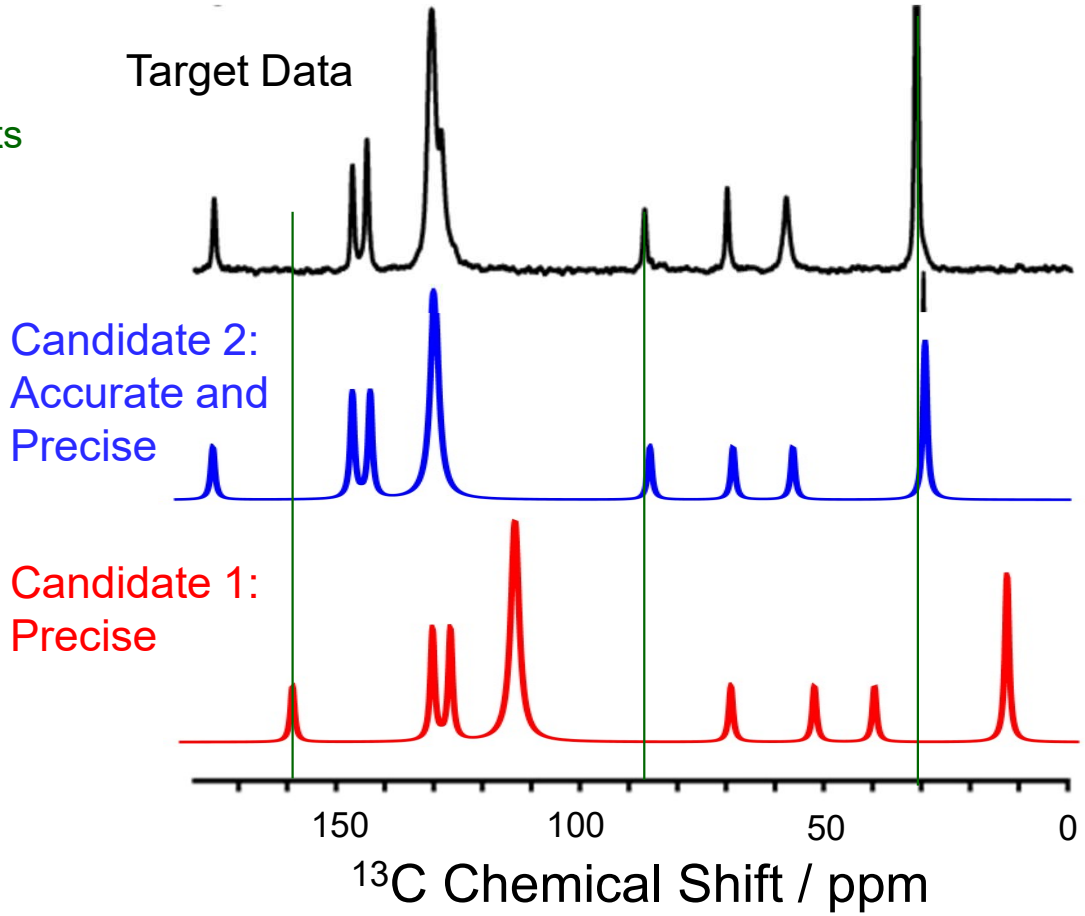
Linear Rescaling

$$\delta_i = m\sigma_i + \sigma_{ref}$$

- Linear rescaling parameters are constants determined from benchmarks, not adjustable parameters
- Allow absolute, not just relative, comparison of theory and experiment

Internal referencing:
- Candidates 1 and 2 are equivalent

Absolute referencing:
- Candidates 2 is superior

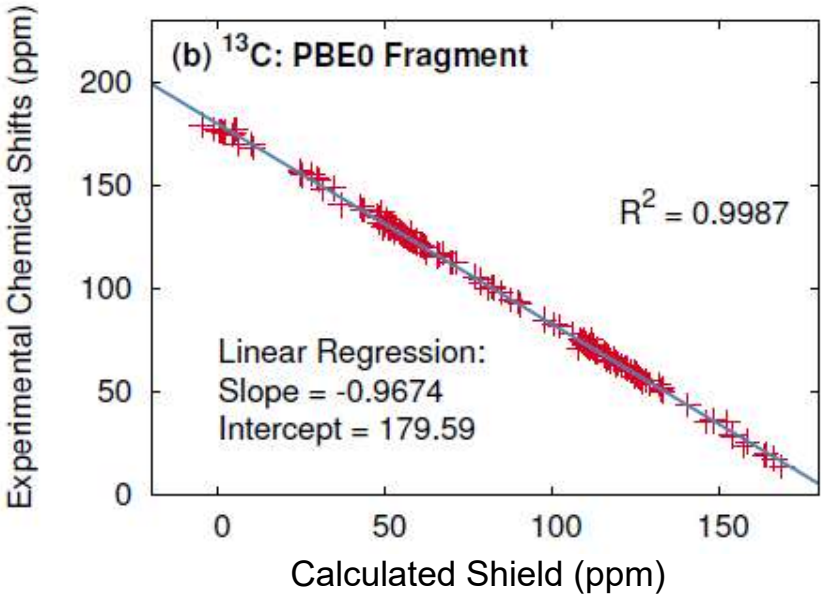


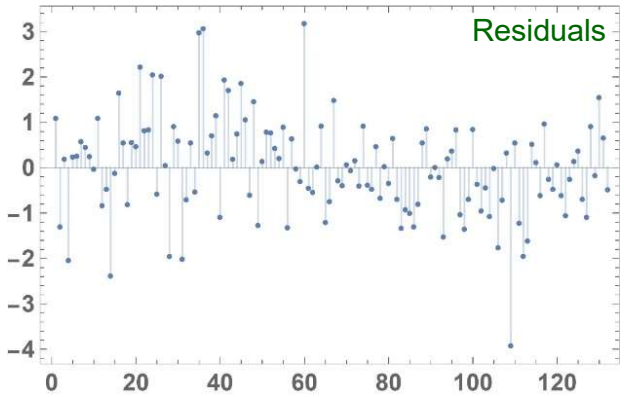
Choice of Functional

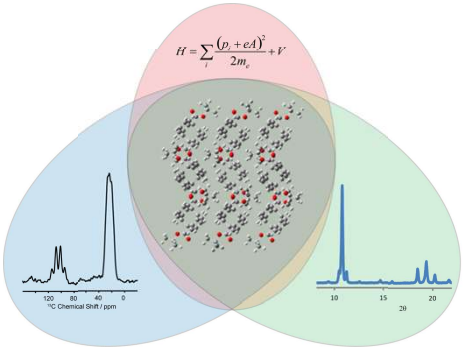
- Hybrid functionals do ~√2 better than the non-hybrids
- No significant differences within each class

^{13}C isotropic shifts

Functional	RMS Error (ppm)	
GIPAW-PBE	2.1	CASTEP
PBE	2.0	Non-Hybrids
BLYP	2.1	
BP86	2.1	
O-PBE	1.8	
PBE0	1.4	Hybrids
B97-2	1.4	
B3PW91	1.4	
B3LYP	1.4	







NMR Crystallography

Requirements:

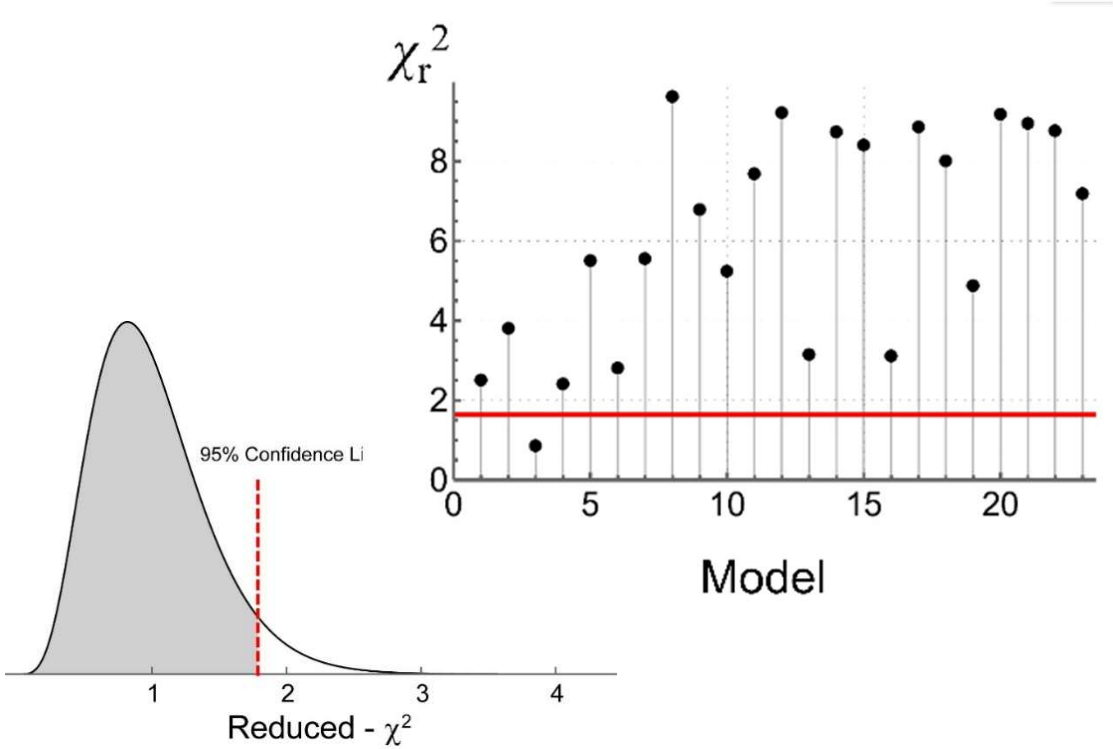
1. A good problem!
2. Candidate structures
3. NMR restraints
4. Accurate chemical shift prediction

Model Ranking and Selection

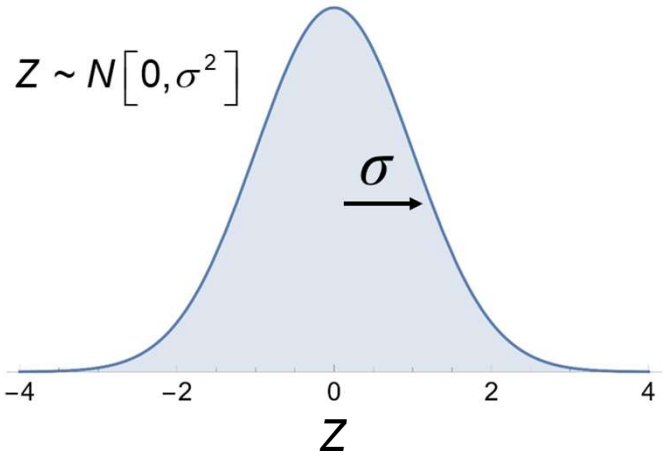
$$\chi_r^2 = \frac{1}{N}\sum_i \frac{\left(\delta_i^{\text{model}} - \delta_i^{\text{exp}}\right)^2}{\sigma_i^2}$$

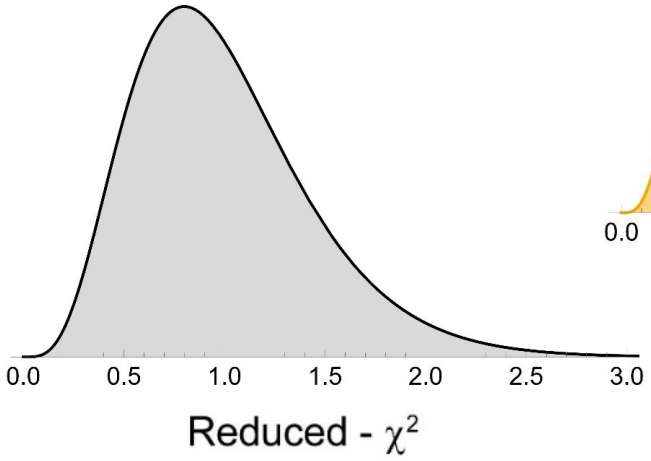
Quantitative Statistics:

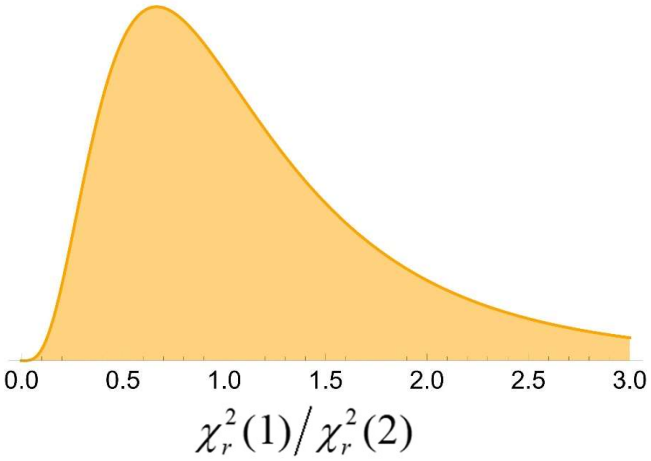
- Are the structures “good”?
- How much better is the best structure?



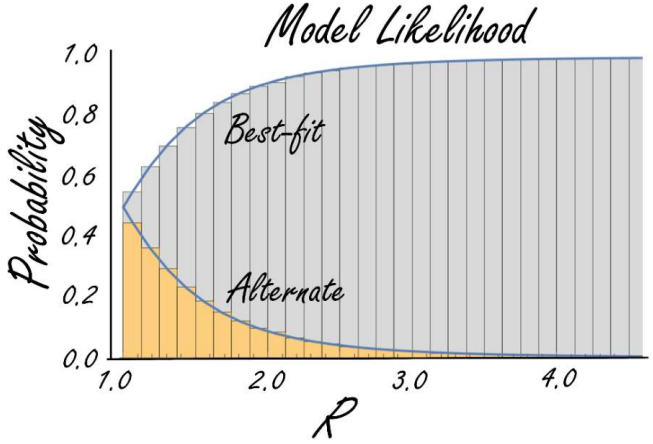
A Grad School Primer on Statistics







$$p(M|y^*) = \frac{p(y^*|M)p(M)}{p(y^*)} = \frac{p(y^*|M)p(M)}{\sum_{M'} p(y^*|M')p(M')}$$



L.J. Mueller in *Modern NMR Crystallography* (D.L. Bryce, ed; 2024)*; Faraday Disc*, **in press** (2024)

Statistical Monte Carlo Simulations

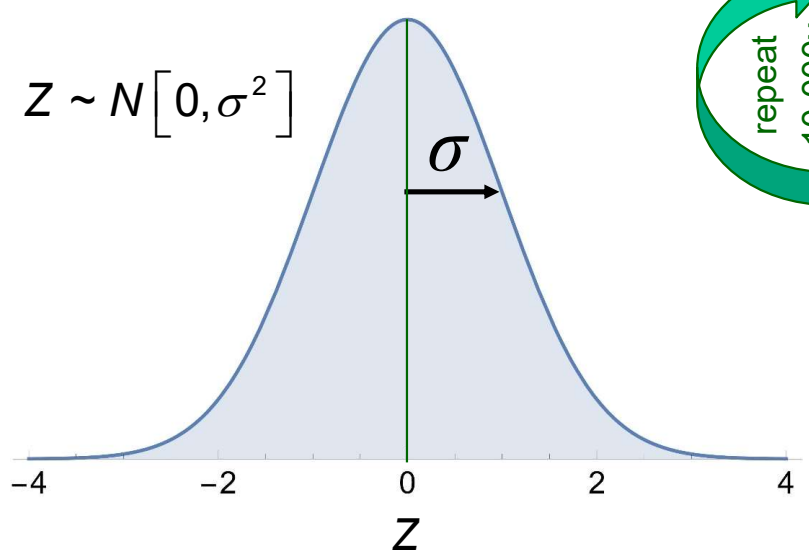
"Offered the choice between mastery of a five-foot shelf of analytical statistics books and middling ability at performing statistical Monte Carlo simulations, we would surely choose to have the latter skill."

Press, Teukolsky, Vetterling, and Flannery, *Numerical Recipes in C*

The Normal Distribution

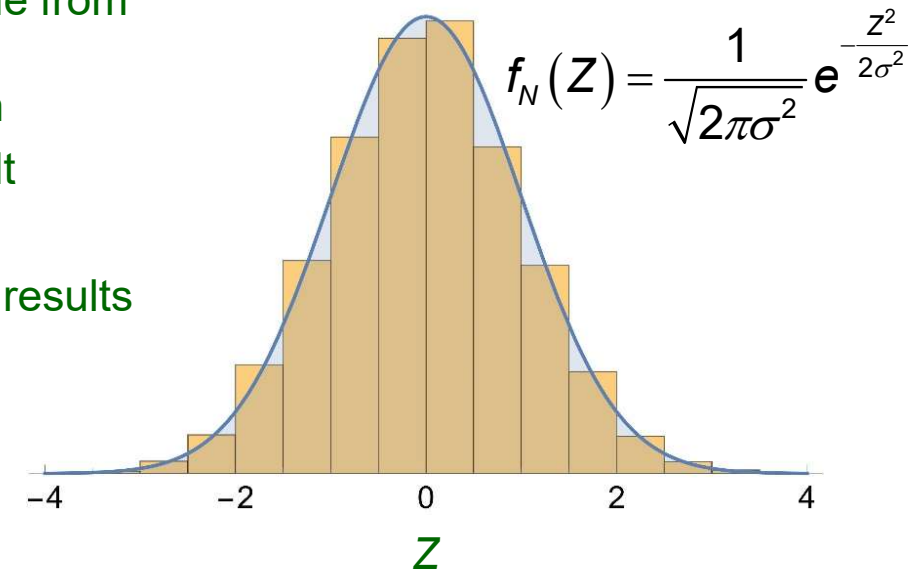
Single measurements are normally distributed

Normal Distribution



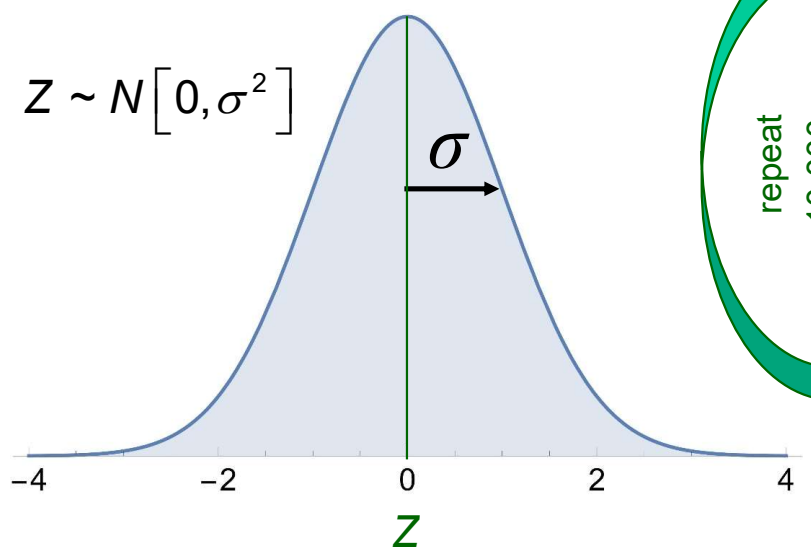
Monte Carlo Approach

1. Pick a value from the normal distribution
2. Store result
3. Histogram results



The Reduced-χ^2 Distribution

Normal Distribution



repeat 10,000x

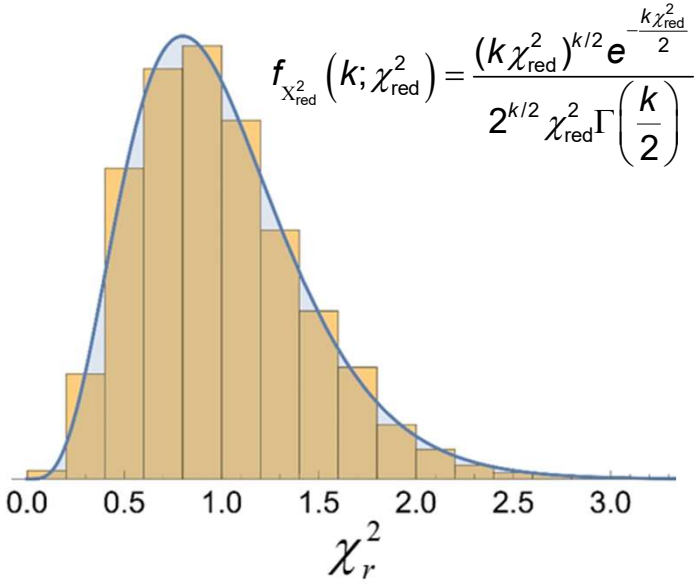
Monte Carlo Approach

1. Pick N values at a time from the normal distribution
2. Combine as the average weighted squared sum (the reduced-χ^2)

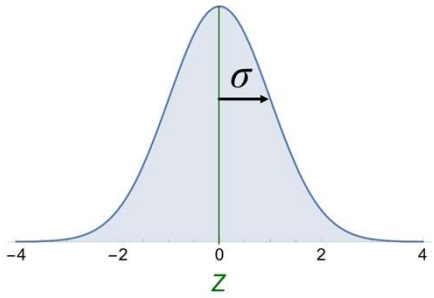
$$\chi^2_{\text{red}} = \frac{1}{N}\sum_{i=1}^{N}\frac{Z_i^2}{\sigma_i^2}$$

3. Histogram results

Average weighted squared sum is red-χ^2 distributed

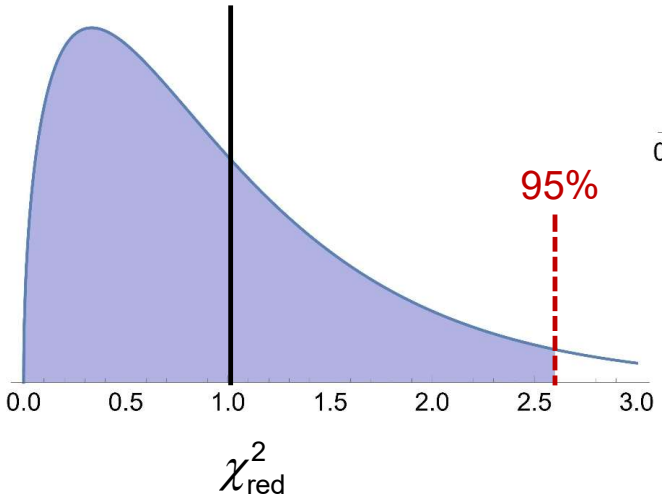


The Reduced-χ^2 Distribution

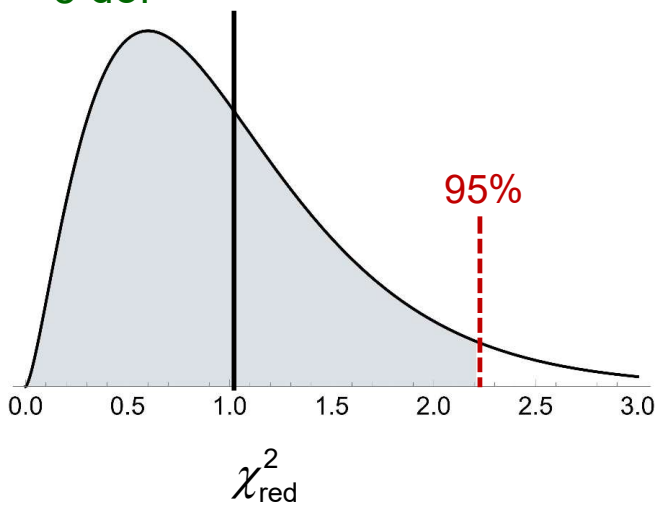


$$\chi^2_{\text{red}} = \frac{1}{N}\sum_{i=1}^{N}\frac{Z_i^2}{\sigma_i^2}$$

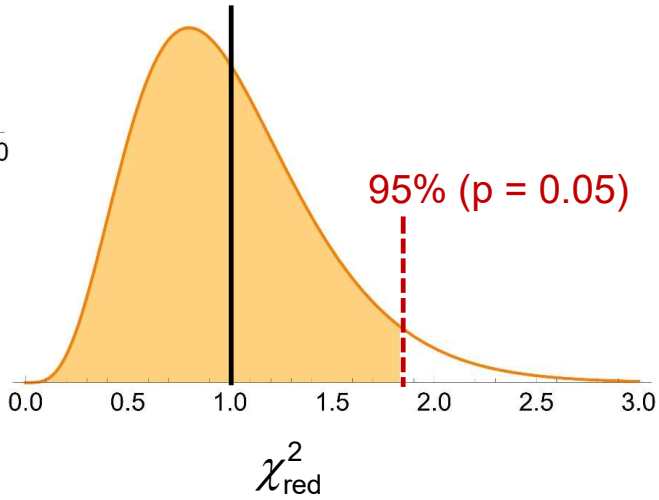
N = 3 dof



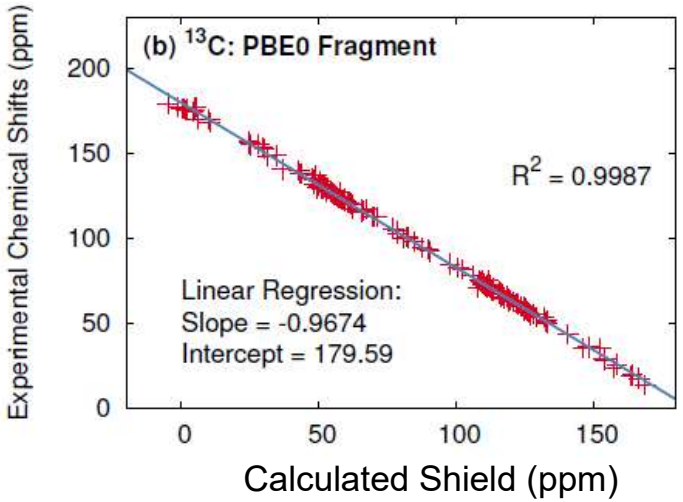
N = 5 dof

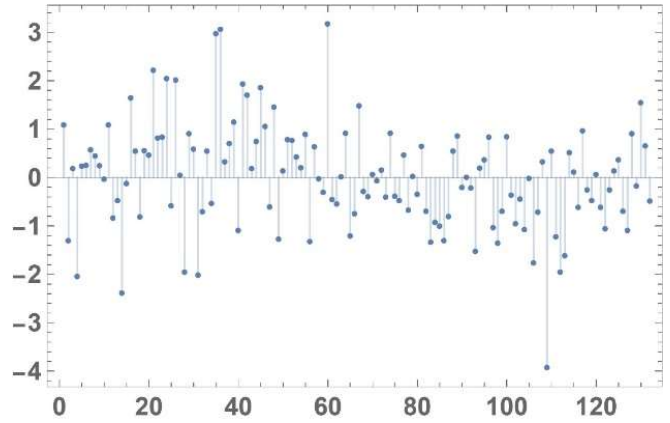


N = 10 dof

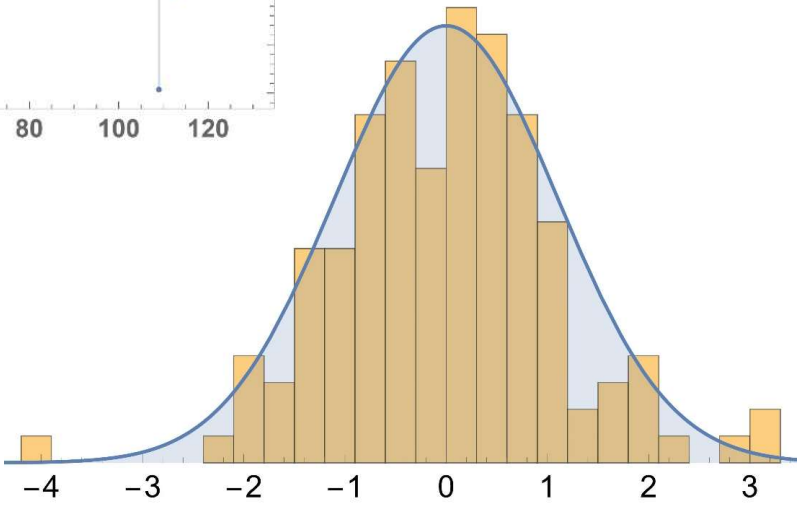


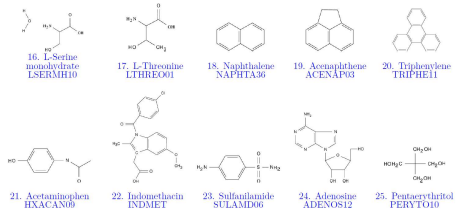
Residuals in Benchmark Studies are Normally Distributed

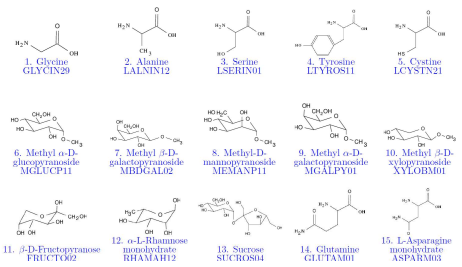




Residuals

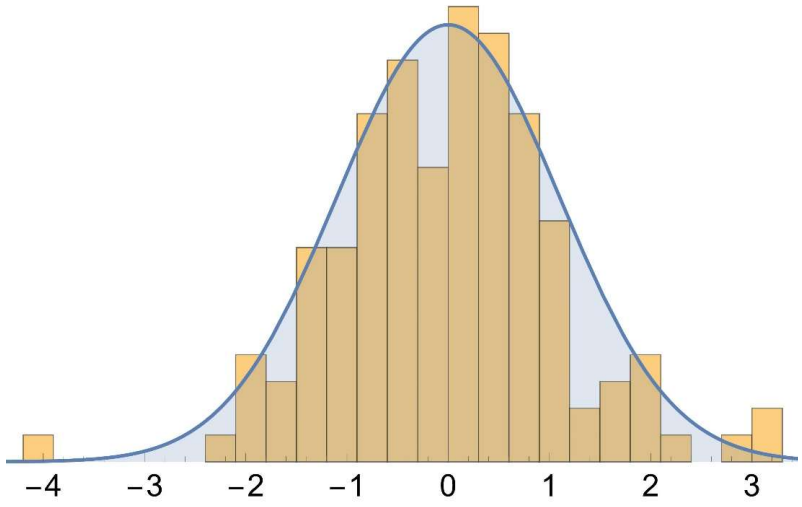


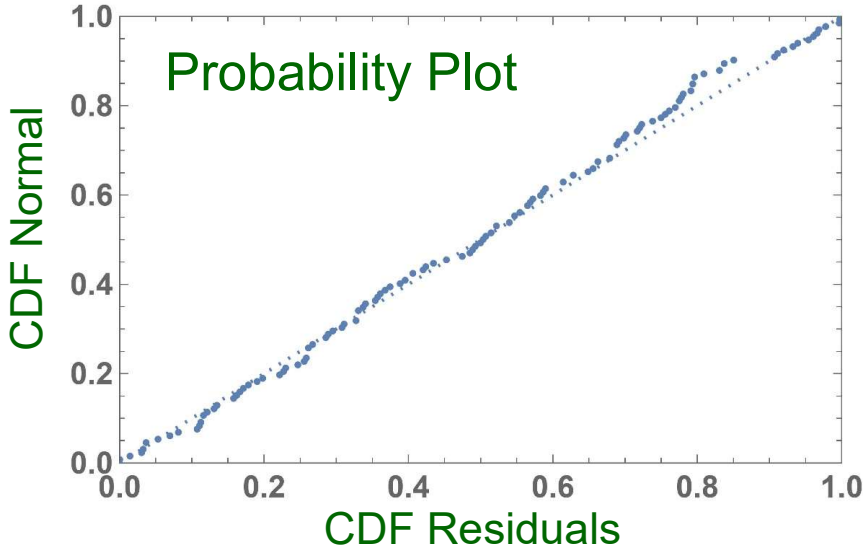


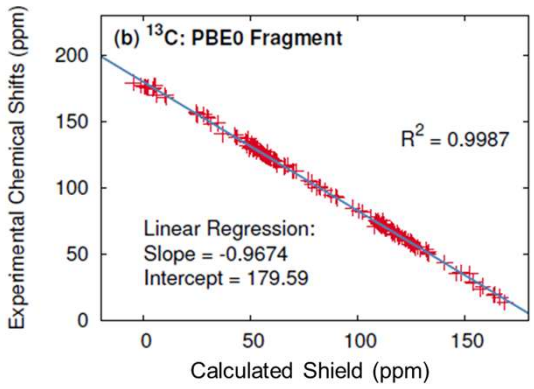


The Residuals

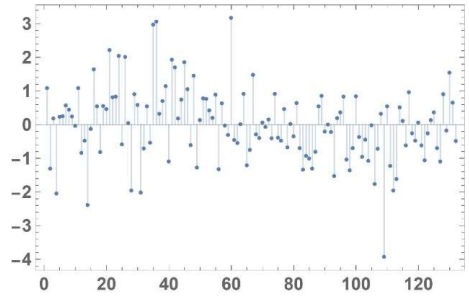
- The residuals in the test sets are normally distributed







Residuals



Model Selection in NMR Crystallography

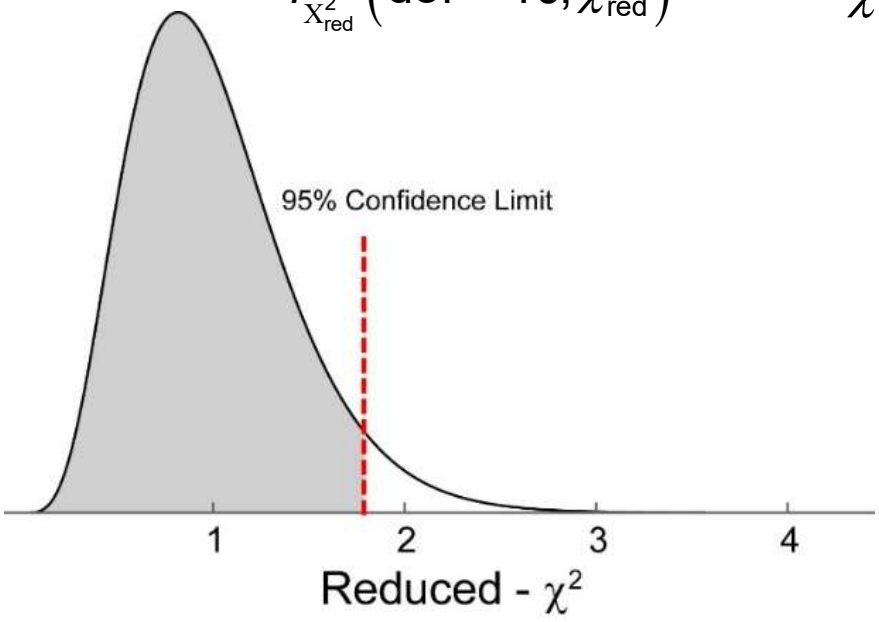
$$\chi_r^2 = \frac{1}{N}\sum_i \frac{\left(\delta_i^{\text{model}} - \delta_i^{\text{exp}}\right)^2}{\sigma_i^2}$$

estimated error from benchmark studies

- Rank models based on their agreement with experimental data using the red-χ^2
- If the residuals are normally distributed (and they are), then the above figure of merit is reduce chi-squared distributed
- Can not only compare models to each other, but can determine if the model is consistent with the data in an absolute sense

Chi-Squared Goodness-of-Fit Test

$$f_{X^2_{red}}\left(\text{dof} = 10; \chi^2_{red}\right)$$



$$\chi_r^2 = \frac{1}{N}\sum_i \frac{\left(\delta_i^{\text{model}} - \delta_i^{\text{exp}}\right)^2}{\sigma_i^2}$$

Example: If model has predictions for 10 experimental shifts, then 95 out of 100 times, a correct model will have red-χ^2<1.83.

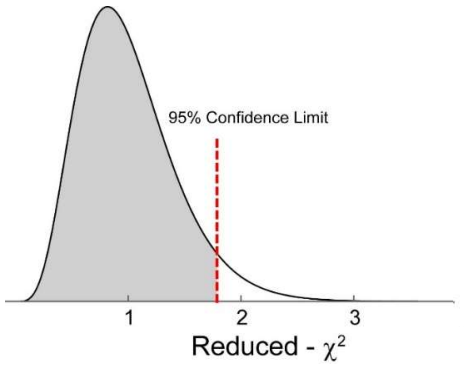
If a model has red-χ^2>1.83, it can be rejected at the 95% confidence level

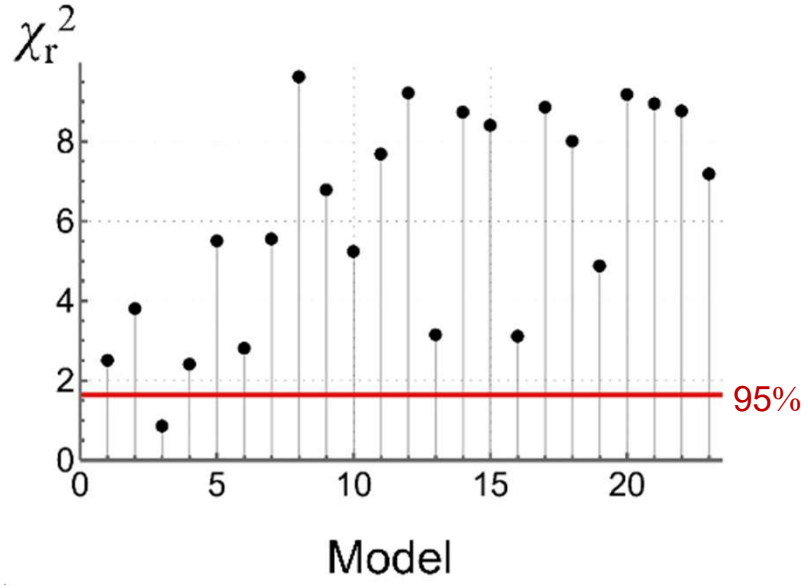
95% confidence intervals depend on the degrees-of-freedom and can be obtained from statistical software or tables

Benchmark for Structure Selection

From a comprehensive set of candidate structures, the identification of (i) a single structure or (ii) a closely related ensemble of structures that satisfy the 95% confidence limits of the red-χ^2 statistic

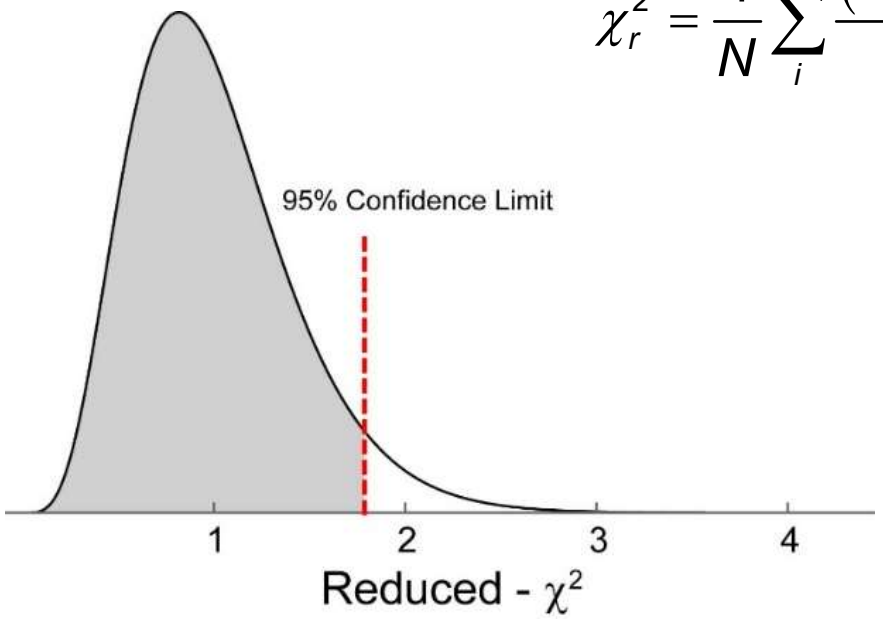
$$\chi_r^2 = \frac{1}{N}\sum_i \frac{\left(\delta_i^{\text{model}} - \delta_i^{\text{exp}}\right)^2}{\sigma_i^2}$$

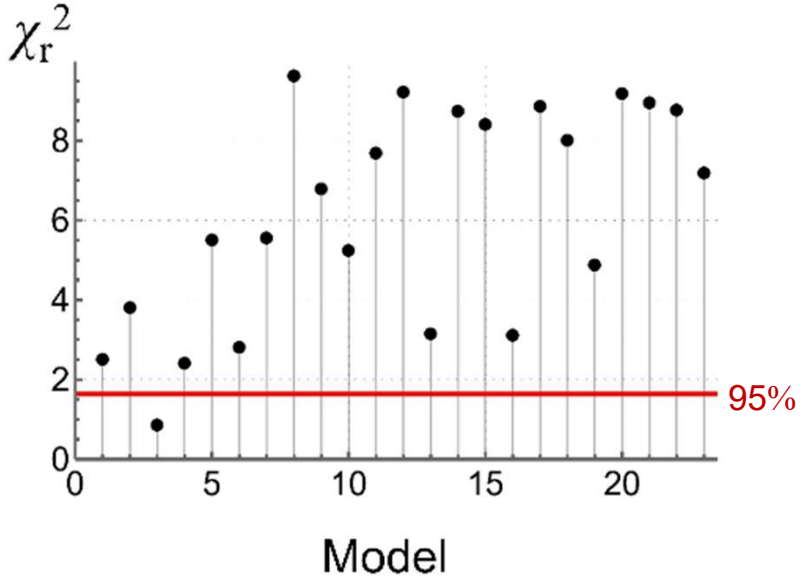




Structure Selection

$$\chi_r^2 = \frac{1}{N}\sum_i \frac{\left(\delta_i^{\text{model}} - \delta_i^{\text{exp}}\right)^2}{\sigma_i^2}$$





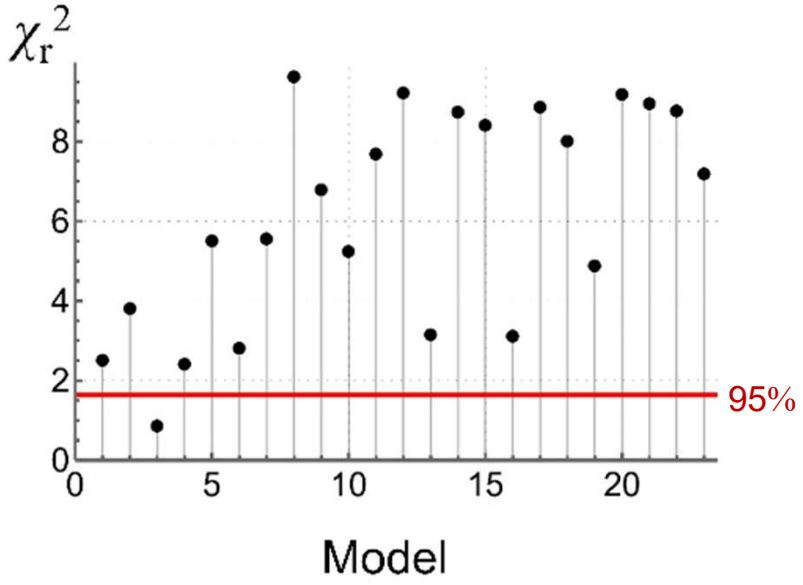
Data from Salager *et al*, JACS **132**, 2564-2566 (2010)

95% CI, red-χ^2<1.64 (16 dof)

Model Probabilities

But even when only one model satisfies the 95% confidence limits, this does not mean that there is a 95% chance that it is the correct, experimental structure!

$$\chi_r^2 = \frac{1}{N}\sum_i \frac{\left(\delta_i^{\text{model}} - \delta_i^{\text{exp}}\right)^2}{\sigma_i^2}$$



To assign model probabilities, we need Bayesian analysis

- Engel *et al.*, *PCCP* **21**, 23385 (2019)
- Mueller, *Faraday Disc* **in press** (2024)

Bayes Theorem

probability of the data, given the model:
the likelihood function
(what we typically know)

model prior probability
(a matter of life or death)

$$P\left(M \mid \mathbf{d}^*\right) = \frac{P\left(\mathbf{d}^* \mid M\right)P(M)}{\sum_{M'} P\left(\mathbf{d}^* \mid M'\right)P(M')}$$

probability of a model, given the data
(what we actually want to know)

Engel *et al.*, *PCCP* **21**, 23385 (2019)

Prior Probabilities

Professorial hyperfixation dementia (PHD)

- PHD Effects 1 in 10,000
- PHD Test sensitivity: 99%

$P(\mathrm{D}) = 0.0001 \quad P(\bar{\mathrm{D}}) = 0.9999$ *Prior prob*

$P(+|\mathrm{D}) = 0.99 \quad P(+|\bar{\mathrm{D}}) = 0.01$ *Likelihood*

Q: You test positive, what is the probability that you have PHD?

$$P(\mathrm{D}|+) = \frac{P(+|\mathrm{D})P(\mathrm{D})}{P(+|\mathrm{D})P(\mathrm{D}) + P(+|\bar{\mathrm{D}})P(\bar{\mathrm{D}})}$$

$$= \frac{(0.99)(0.0001)}{(0.99)(0.0001) + (0.01)(0.9999)}$$

$$\approx 0.0098$$

Good. So just because you have a positive PHD test, your life isn't ruined

Bayesian Approaches

Traditional Bayesian analysis

uniform

$$P\left(M \mid \mathbf{d}^*\right) = \frac{P\left(\mathbf{d}^* \mid M\right)P\left(M\right)}{\sum_{M'} P\left(\mathbf{d}^* \mid M'\right)P\left(M'\right)}$$

- Engel *et al.*, *PCCP* **21**, 23385 (2019)

Hierarchical Bayesian analysis

empirically derived

$$P\left(M \mid \mathbf{d}^*\right) = \int P\left(M, s \mid \mathbf{d}^*\right) ds = \frac{\int P\left(\mathbf{d}^* \mid M, s\right)P\left(M \mid s\right)P\left(s\right) ds}{\sum_{M'} \int P\left(\mathbf{d}^* \mid M', s\right)P\left(M' \mid s\right)P\left(s\right) ds}$$

The UC Model

- Mueller, *Faraday Disc* **in press** (2024)

… or we could use statistical Monte Carlo analysis

A Game of Model Selection

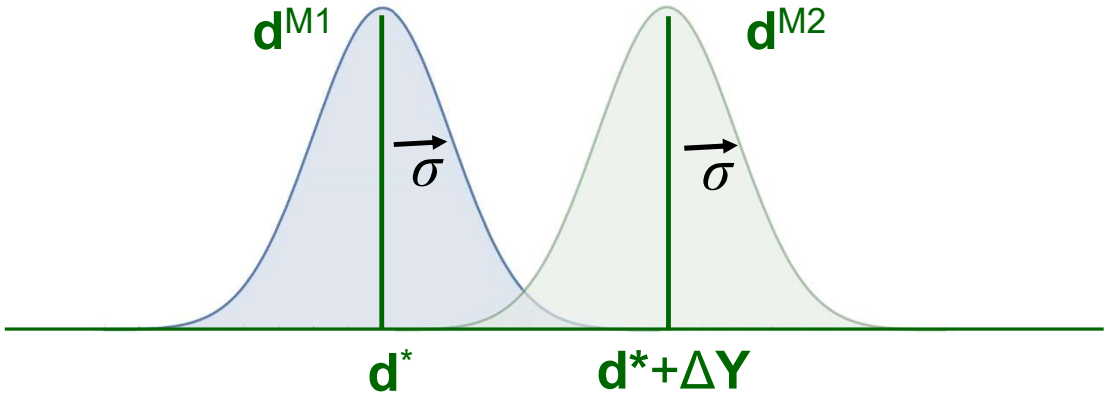
Model 1: the correct experimental structure

- In the limit of perfect theory, its first-principles predicted properties $\mathbf{d}^{M1} = \left\{d_1^{M1}, d_2^{M1}, \ldots, d_n^{M1}\right\}$ are in exact agreement with experiment $\mathbf{d}^{*} = \left\{d_1^{*}, d_2^{*}, \ldots, d_n^{*}\right\}$

Model 2: An incorrect model

- In the limit of perfect theory, its first-principles predicted properties $\mathbf{d}^{M2} = \left\{d_1^{M2}, d_2^{M2}, \ldots, d_n^{M2}\right\}$ deviate systematically from experiment by the set of differentials: $\Delta\mathbf{Y} = \left\{\Delta Y_1, \Delta Y_2, \ldots, \Delta Y_n\right\}$

Now reintroduce variable uncertainty into the predictions and ask: if the model with the smaller red-χ^2 is always selected, what is the probability that each model will be chosen?



Note: assuming all the error is in the predictions, not the experimental data. This can be relaxed.

Statistical Monte Carlo Simulations

Assumptions:

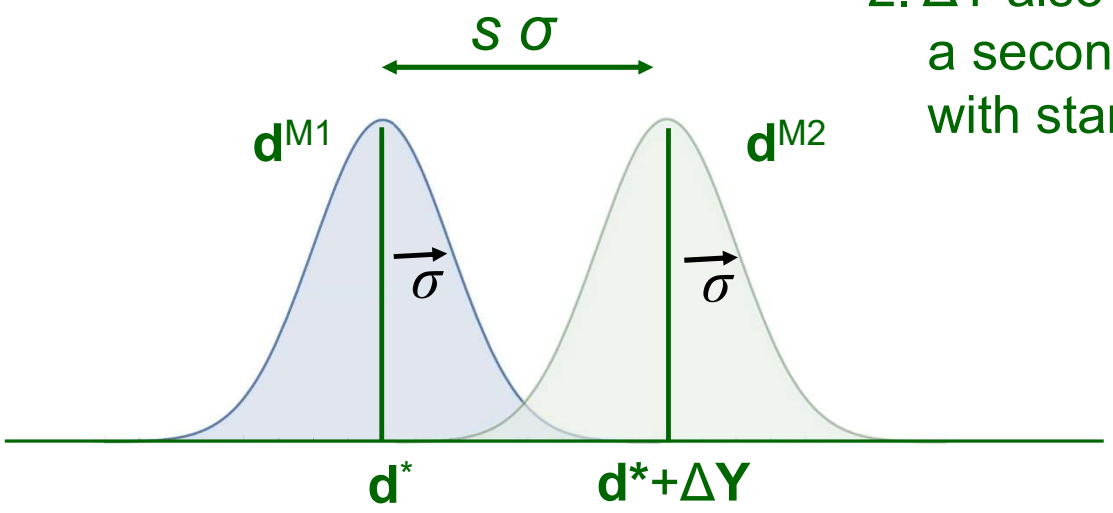
$$\mathbf{d}^{M1} - \mathbf{d}^{*} = \{X_1, X_2, \ldots, X_n\}, \quad X \sim N[0, \sigma^2]$$

$$\mathbf{d}^{M2} - \mathbf{d}^{*} = \{Y_1, Y_2, \ldots, Y_n\}, \quad Y \sim N[\Delta Y, \sigma^2]$$

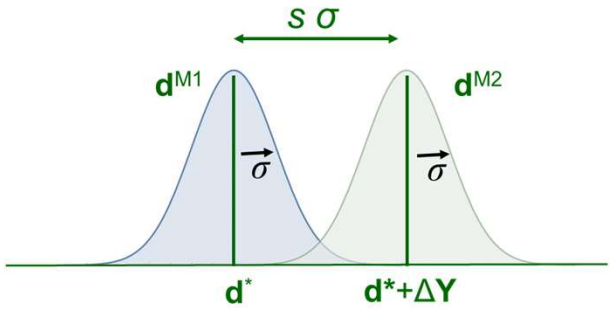
1. predications are normally distributed about 0 and ΔY

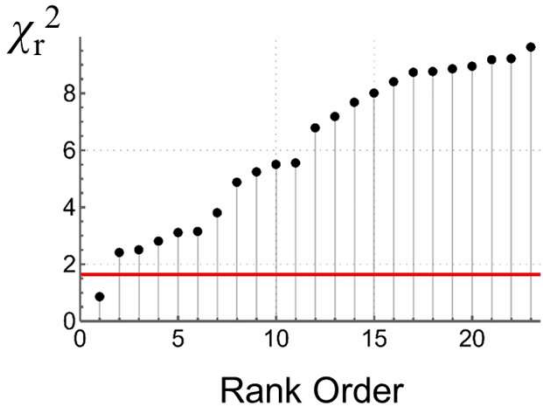
$$Y \sim N[\Delta Y, \sigma^2] = N\left[N\left[0, (s\sigma)^2\right], \sigma^2\right] = N\left[0, (s^2 + 1)\sigma^2\right]$$

2. ΔY also unknown, so pick from a second normal distribution with standard deviation $s\,\sigma$



A Game of Model Selection





Data from Salager *et al*, JACS **132**, 2564-2566 (2010)

Assumptions:

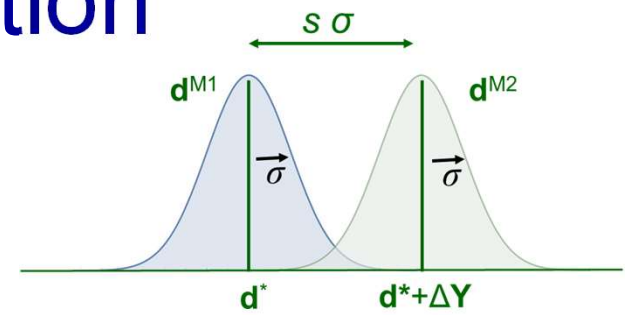
1. predications are normally distributed about 0 and ΔY

2. ΔY also unknown, so pick from a second normal distribution with standard deviation *s* σ

3. *s* is also unknown, so pick from a third distribution in which the probability of models increases linearly with *s*

   This corresponds to candidate models being uniformly distributed with respect to red-χ^2 values, as seen experimentally: the "Uniform Chi-Squared (UC) Model"

A Game of Model Selection



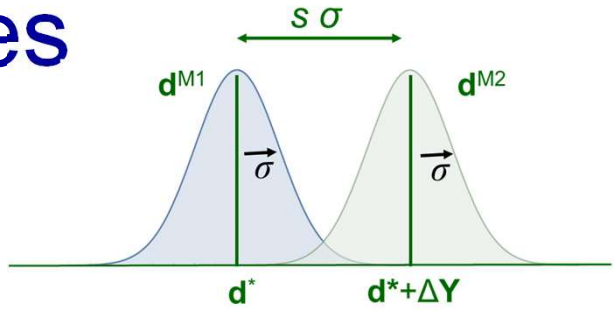
Monte Carlo Simulation for 2 models with *n* chemical shifts measured:

repeat 10,000x

1. Pick s from a distribution with linearly increasing probability
2. Pick *n* samples of *X* for model 1 and calculate: $\chi^2_{\text{red,M1}} = \frac{1}{n}\sum_{i=1}^{n}\frac{X_i^2}{\sigma^2}$
3. Pick *n* samples of *Y* for model 2 and calculate: $\chi^2_{\text{red,M2}} = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i^2}{\sigma^2}$
4. Assign best-fit structure based on lower red-χ^2
   - Best-fit = Model 1, correct assignment made
   - Best-fit = Model 2, incorrect assignment made
5. Store the ratio of the red-χ^2 in either the correct or incorrect list

$$R = \chi^2_{\text{red,Alt}} \Big/ \chi^2_{\text{red,BF}} \geq 1$$

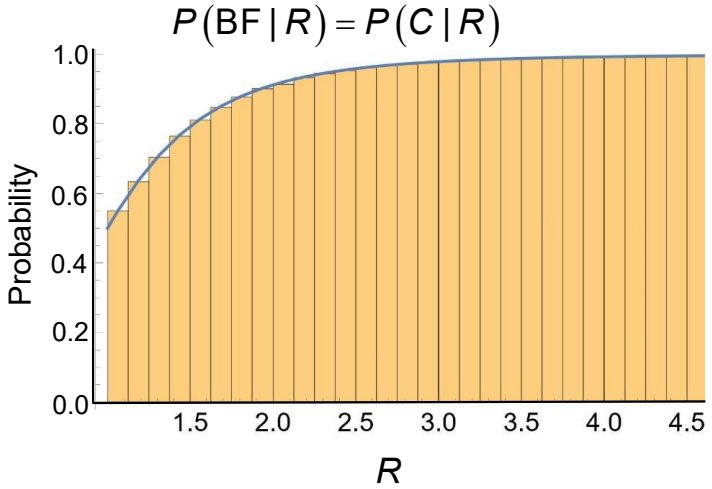
UC Model Probabilities



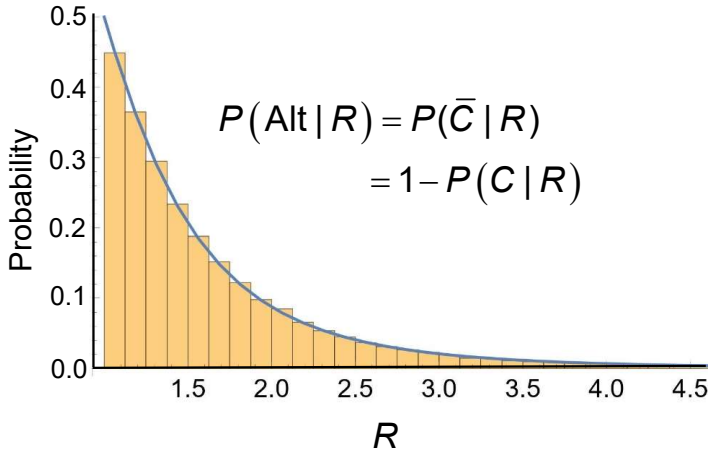
Best-Fit Model: $\chi^2_{\text{red,BF}}$

Alternate Model: $\chi^2_{\text{red,Alt}}$

$$R = \chi^2_{\text{red,Alt}} \Big/ \chi^2_{\text{red,BF}} \geq 1$$



Probability that the best-fit model is the experimental structure



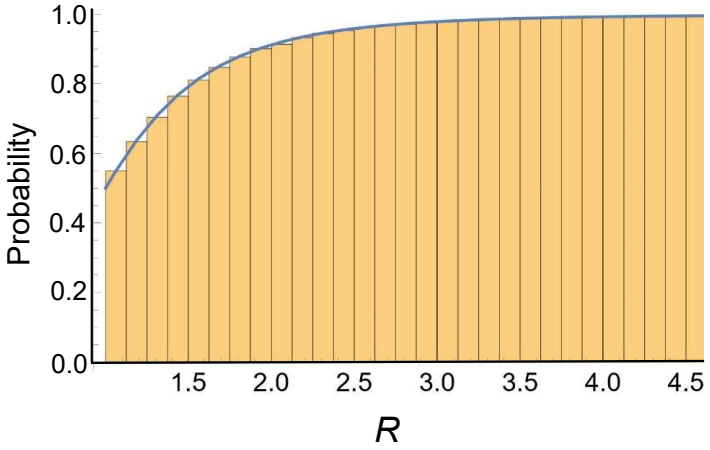
Probability that the alternate model is actually the experimental structure

UC Model Probabilities

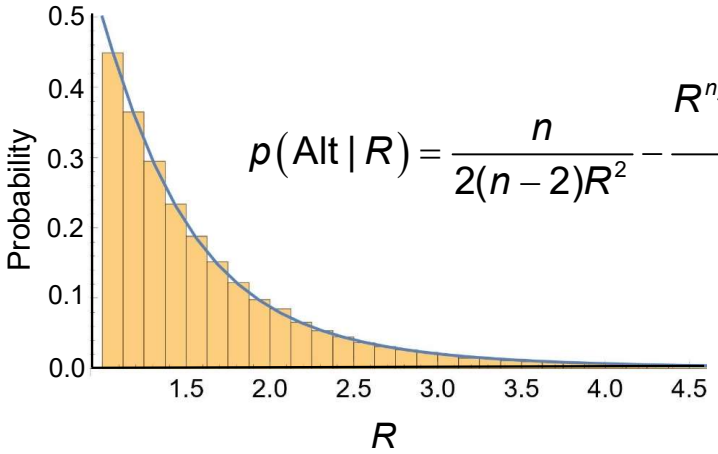
Hierarchical Bayesian inference can give these curves analytically

$$p(\mathrm{BF}|R) = \frac{R^{n/2-1}\,{}_2F_1\left(n, \frac{n}{2}-1, \frac{n}{2}, -R\right)}{(n-2)B\left(\frac{n}{2}, \frac{n}{2}\right)}$$

$$R = \chi^2_{\mathrm{red,Alt}} \Big/ \chi^2_{\mathrm{red,BF}} \geq 1$$



$$p(\mathrm{Alt}\,|\,R) = \frac{n}{2(n-2)R^2} - \frac{R^{n/2-1}\,{}_2F_1\left(n, \frac{n}{2}+1, \frac{n}{2}+2, -R\right)}{(n+2)B\left(\frac{n}{2}, \frac{n}{2}\right)}$$



Mueller, *Faraday Disc* **in press** (2024)

UC Model Probabilities: Example

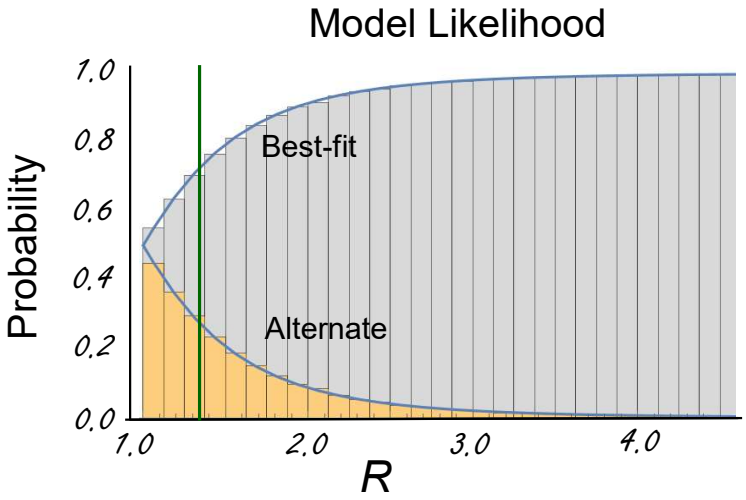
Two Models: A and B

10 NMR chemical shifts

$$\chi^2_{\text{red,A}} = 1.12$$

$$\chi^2_{\text{red,B}} = 1.51$$

$$R = 1.35$$



Probability that A is the experimental structure: 73%

Probability that B is the experimental structure: 27%

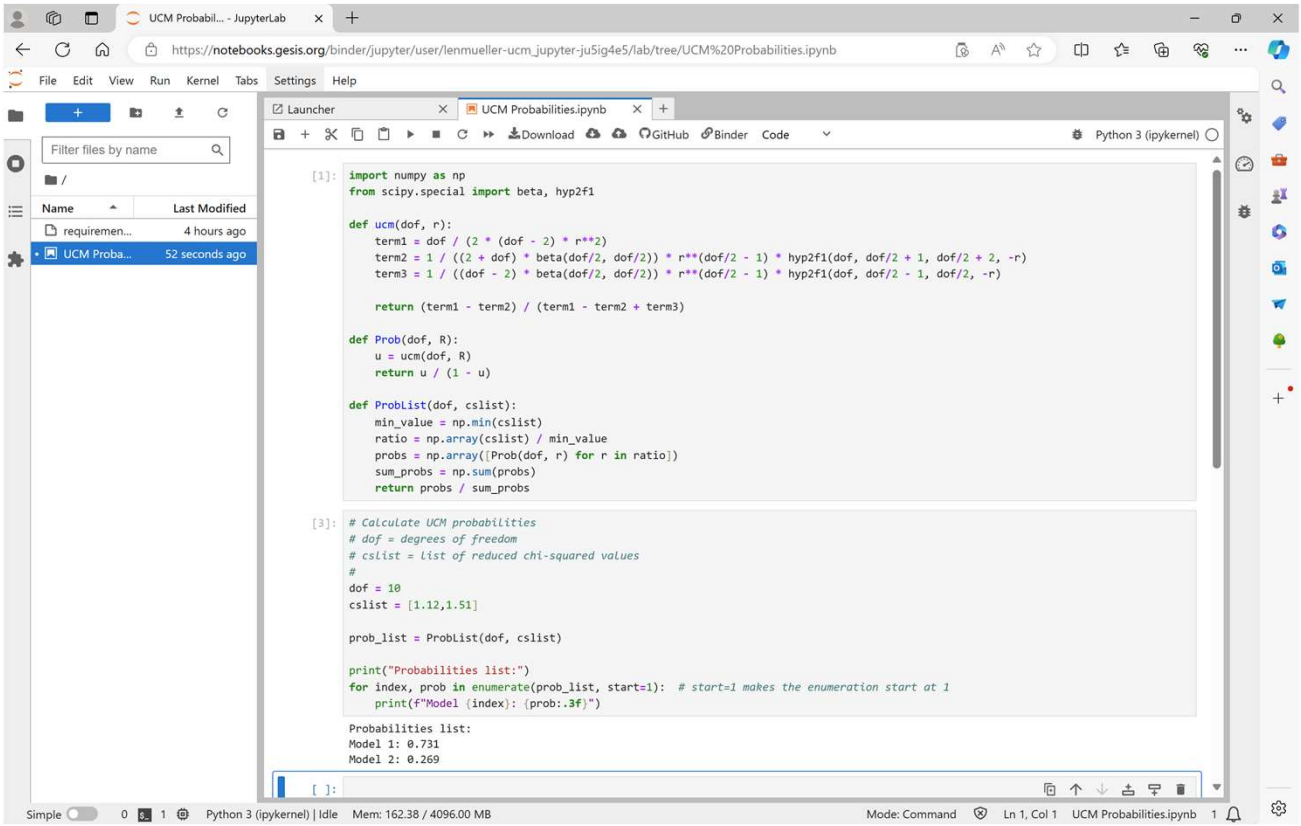
UC Model Probabilities: Example

Binder: mybinder.org/v2/gh/lenmueller/ucm_jupyter/main

github.com/Lenmueller

- Jupyter notebook
- Python script

All you need is a list of red-χ^2 values



UC Model Probabilities: Example

Binder: mybinder.org/v2/gh/lenmueller/ucm_jupyter/main

github.com/Lenmueller

- Jupyter notebook
- Python script

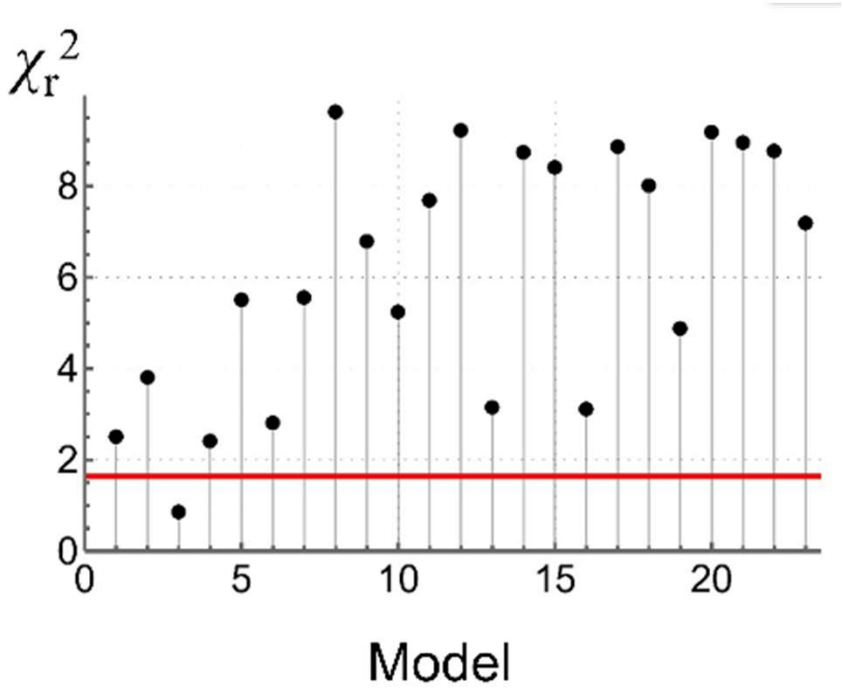
All you need is a list of red-χ^2 values

[3]:
```
# Calculate UCM probabilities
# dof = degrees of freedom
# cslist = list of reduced chi-squared values
#
dof = 10
cslist = [1.12,1.51]

prob_list = ProbList(dof, cslist)

print("Probabilities list:")
for index, prob in enumerate(prob_list, start=
    print(f"Model {index}: {prob:.3f}")
```

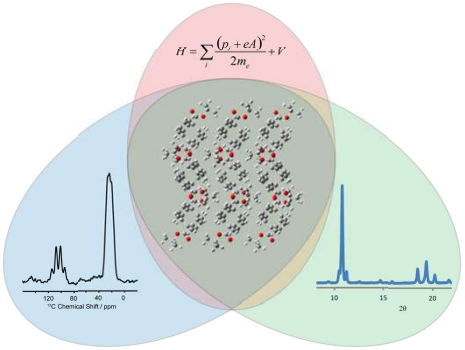
```
Probabilities list:
Model 1: 0.731
Model 2: 0.269
```

Model	All Shifts (n =18, f = 2)	
	red-χ^2	$P_{UC}(M\|R)$
1	2.50	0.0070
2	3.80	0.0006
3	0.85	0.9759
4	2.41	0.0088
5	5.50	0
6	2.81	0.0037
7	5.55	0
8	9.62	0
9	6.78	0
10	5.24	0.0001
11	7.68	0
12	9.22	0
13	3.15	0.0018
14	8.73	0
15	8.41	0
16	3.11	0.0020
17	8.86	0
18	8.00	0
19	4.88	0.0001
20	9.18	0
21	8.95	0
22	8.76	0
23	7.18	0
95% Confidence	*1.64*	

UC Model Probabilities



Data from Salager *et al*, JACS **132**, 2564-2566 (2010)

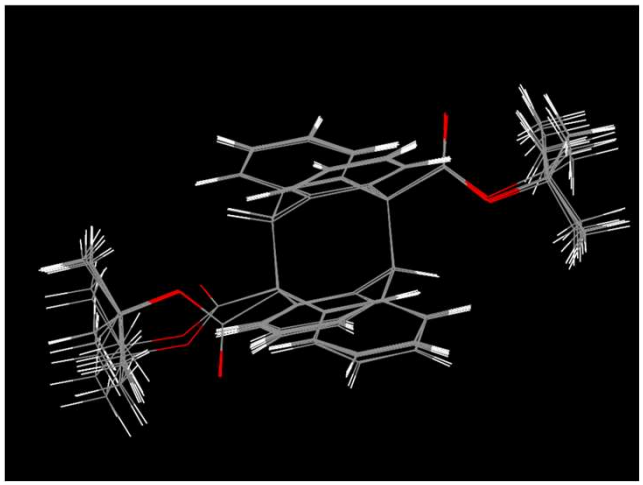


NMR Crystallography

Requirements:

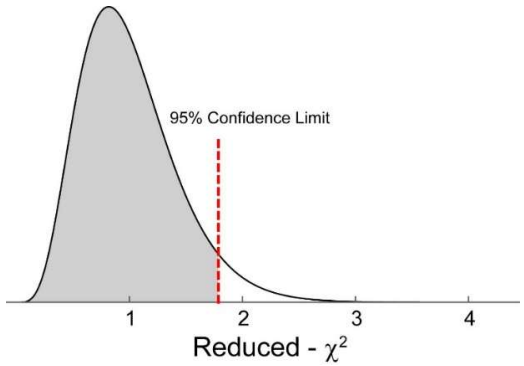
1. A good problem!
2. Candidate structures
3. NMR restraints
4. Accurate chemical shift prediction
5. Quantitative ranking of models

Solid-State Reacted Dimer



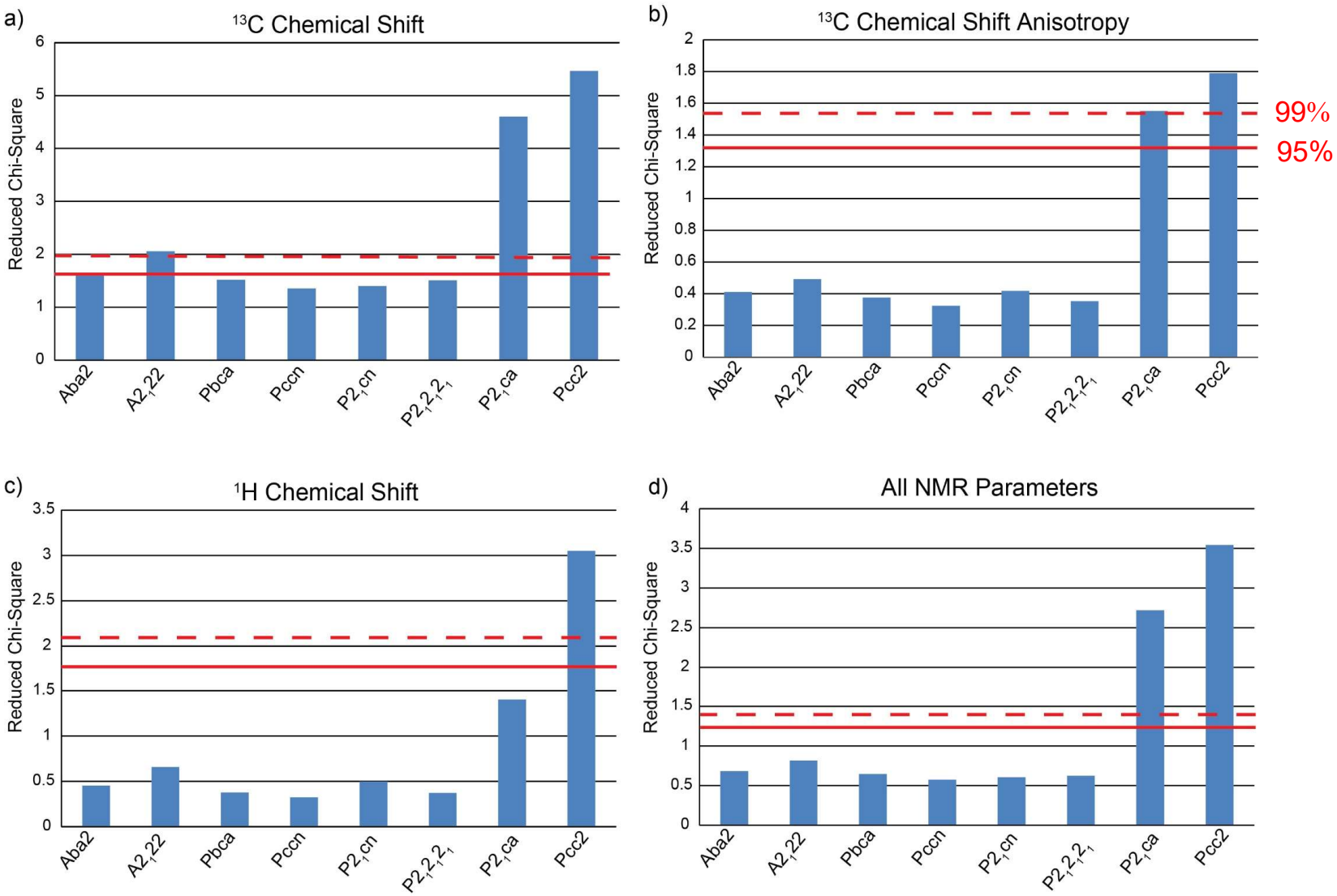
- 8 candidate crystal structures
- Calculate shifts for each and rank using reduced chi-square

$$\chi_r^2 = \frac{1}{N}\sum_i \frac{\left(\delta_i^{\text{model}} - \delta_i^{\text{exp}}\right)^2}{\sigma_i^2}$$



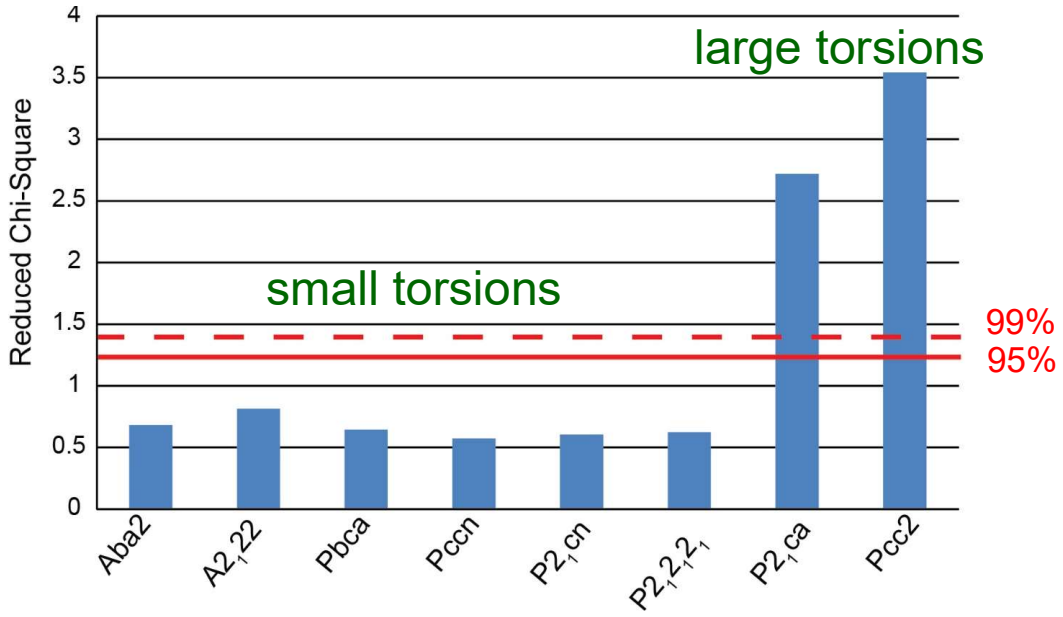
Weightings σ / ppm:

- ^{1}H isotropic: 0.32
- ^{13}C isotropic: 1.4
- ^{13}C tensor: 3.0


a)
^{13}C Chemical Shift
Reduced Chi-Square
0
1
2
3
4
5
6
Aba2
A2$_1$22
Pbca
Pccn
P2$_1$cn
P2$_1$2$_1$2$_1$
P2$_1$ca
Pcc2
b)
^{13}C Chemical Shift Anisotropy
Reduced Chi-Square
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
99%
95%
Aba2
A2$_1$22
Pbca
Pccn
P2$_1$cn
P2$_1$2$_1$2$_1$
P2$_1$ca
Pcc2
c)
^{1}H Chemical Shift
Reduced Chi-Square
0
0.5
1
1.5
2
2.5
3
3.5
Aba2
A2$_1$22
Pbca
Pccn
P2$_1$cn
P2$_1$2$_1$2$_1$
P2$_1$ca
Pcc2
d)
All NMR Parameters
Reduced Chi-Square
0
0.5
1
1.5
2
2.5
3
3.5
4
Aba2
A2$_1$22
Pbca
Pccn
P2$_1$cn
P2$_1$2$_1$2$_1$
P2$_1$ca
Pcc2

All Spectroscopic Parameters

$$\chi_r^2 = \frac{1}{N}\sum_i \frac{\left(\delta_i^{\text{model}} - \delta_i^{\text{exp}}\right)^2}{\sigma_i^2}$$

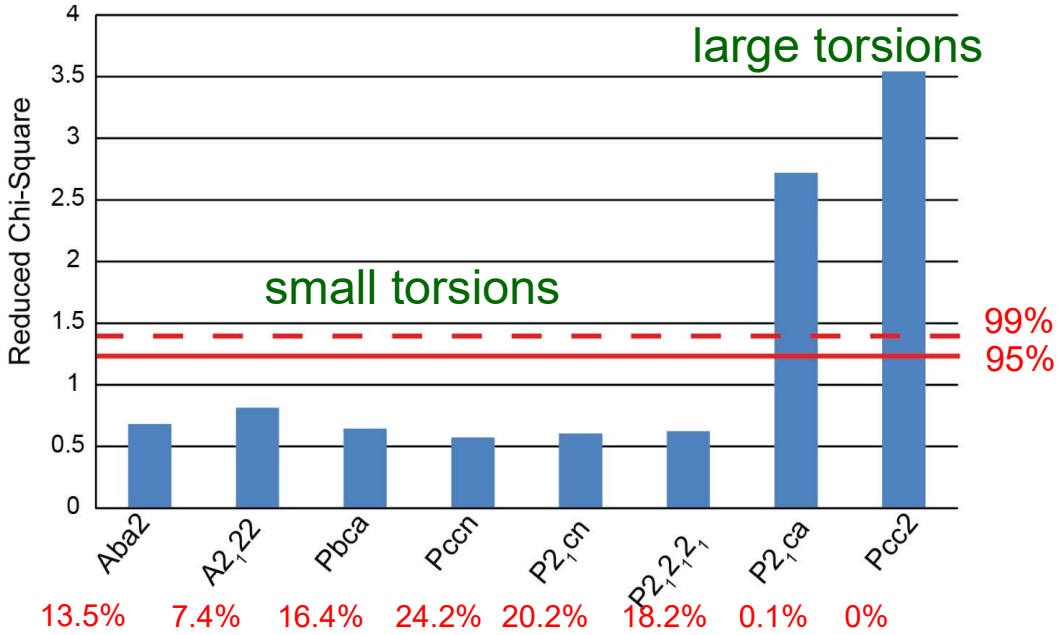


- Chi-squared goodness-of-fit can rule out large torsions for the ester groups

All Spectroscopic Parameters

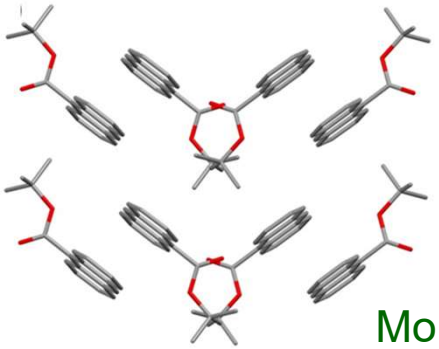
UC Model Probabilities

Model	All Data (k =78)	
	red-χ^2	$P_{UC}(M\|R)$
Aba2	0.682	13.5%
A2$_1$22	0.814	7.4%
Pbca	0.644	16.4%
Pccn	0.573	24.2%
P2$_1$cn	0.605	20.2%
P2$_1$2$_1$2$_1$	0.624	18.2%
P2$_1$ca	2.720	0.1%
Pcc2	3.541	0.0%
95% Confidence	*1.27*	

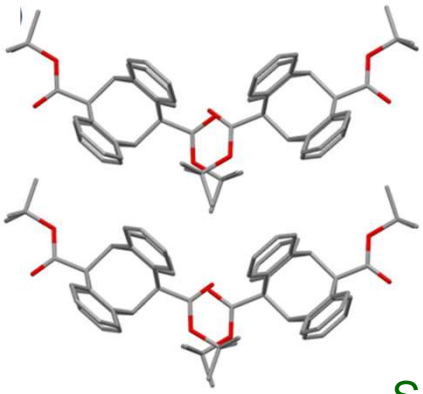


- Can rule out large torsions for the ester groups
- Best description has spacegroup *Pccn*
  (but others show essentially equivalent results)

Crystal Structure of the Solid-State Reacted Dimer



Monomer



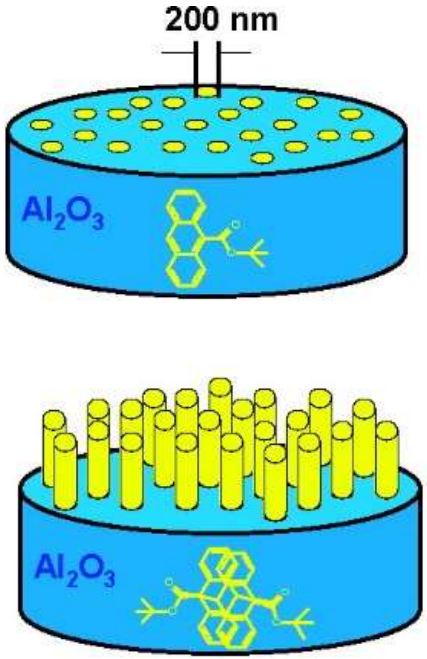
SSRD

- Success: crystal structure of the solid-state reacted dimer
- Maintains the herringbone packing of the anthracene rings
- The t-butyl ester groups are still rotated inward
- Consistent with the *Topochemical Principle*
- But no obvious mechanism for expansion
  - Volume per anthracene decreases slightly for the dimer unit cell

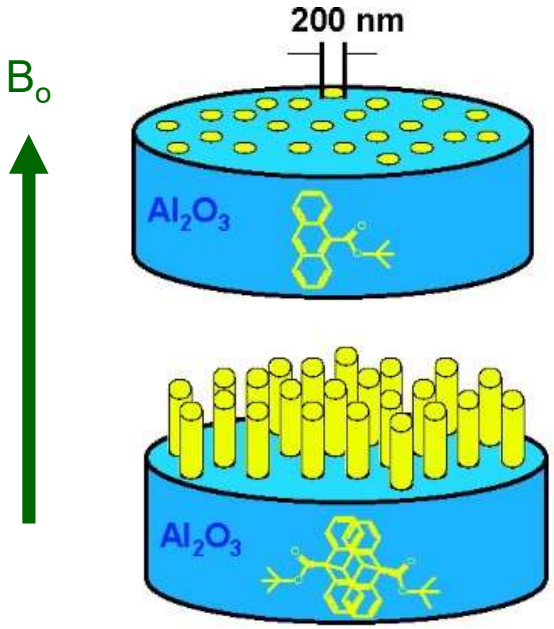
NMR Crystallography of 9TBAE Nanorods

- To determine of mechanism of expansion need to orient the monomer and dimer unit cells relative to the nanorod axis

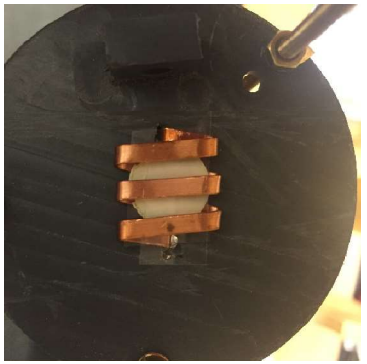
- Direct NMR measurements on an ensemble of uniformly oriented single crystals nanorods in the AAO template

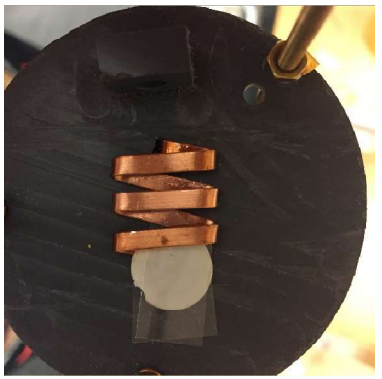


NMR Crystallography of 9TBAE Nanorods

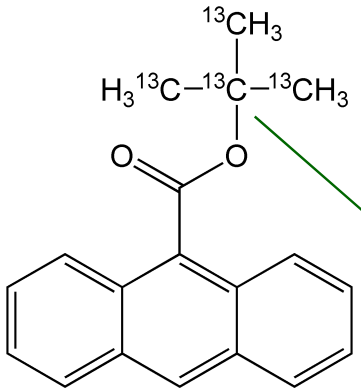


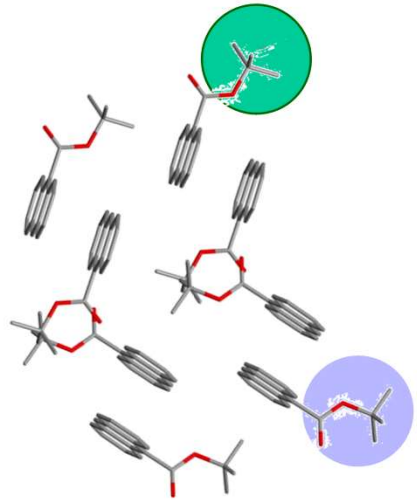
Place in flat coil NMR probe with nanorod long axis along the static magnetic field: one degree of orientation

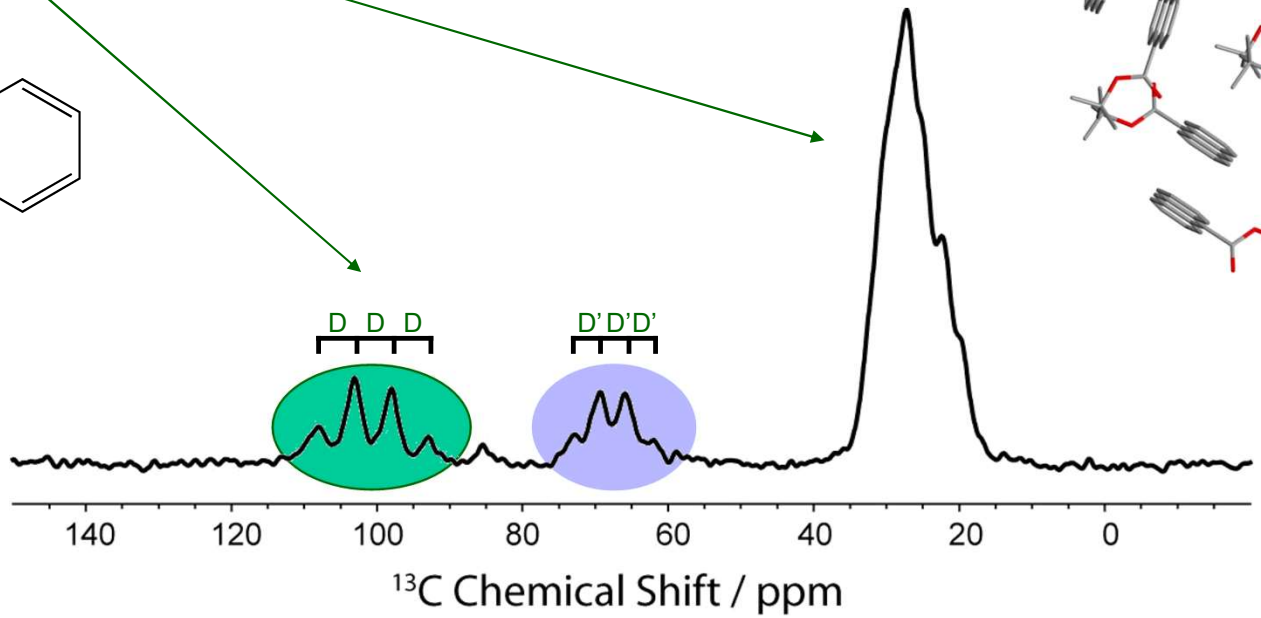




Single Crystal Solid-State NMR of 9TBAE

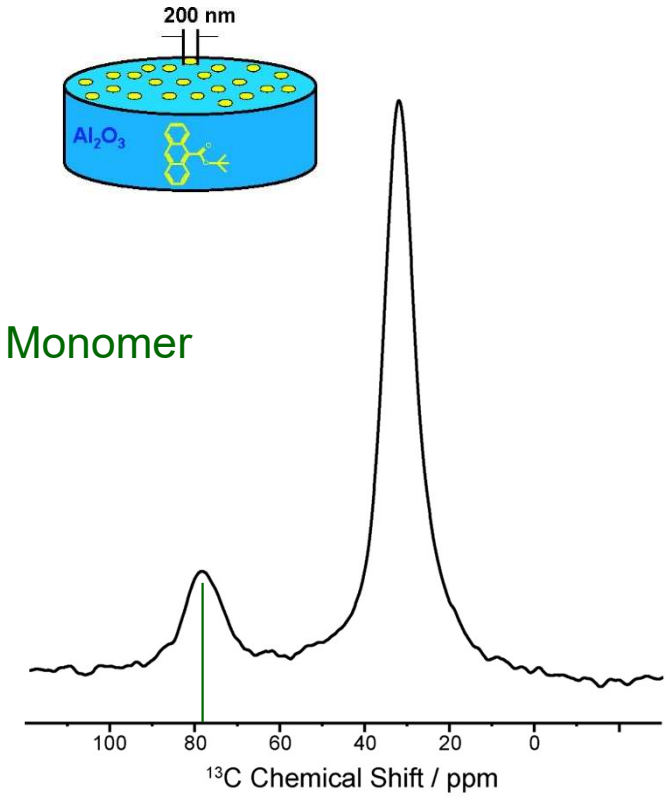


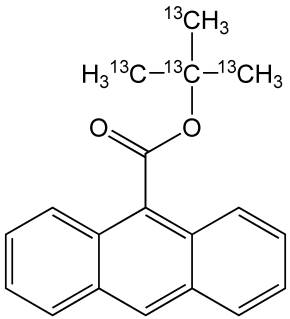




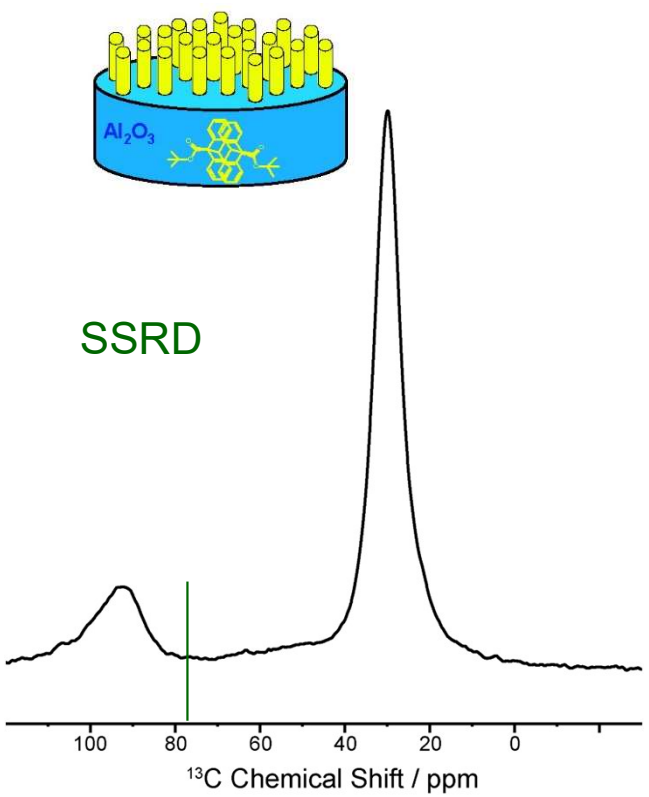
- consistent with rapid rotation of the t-butyl groups about the O-C bond
- averaged dipolar coupling along the rotation axis with a 1:3:3:1 intensity pattern

Solid-State NMR of Oriented 9TBAE Nanorods

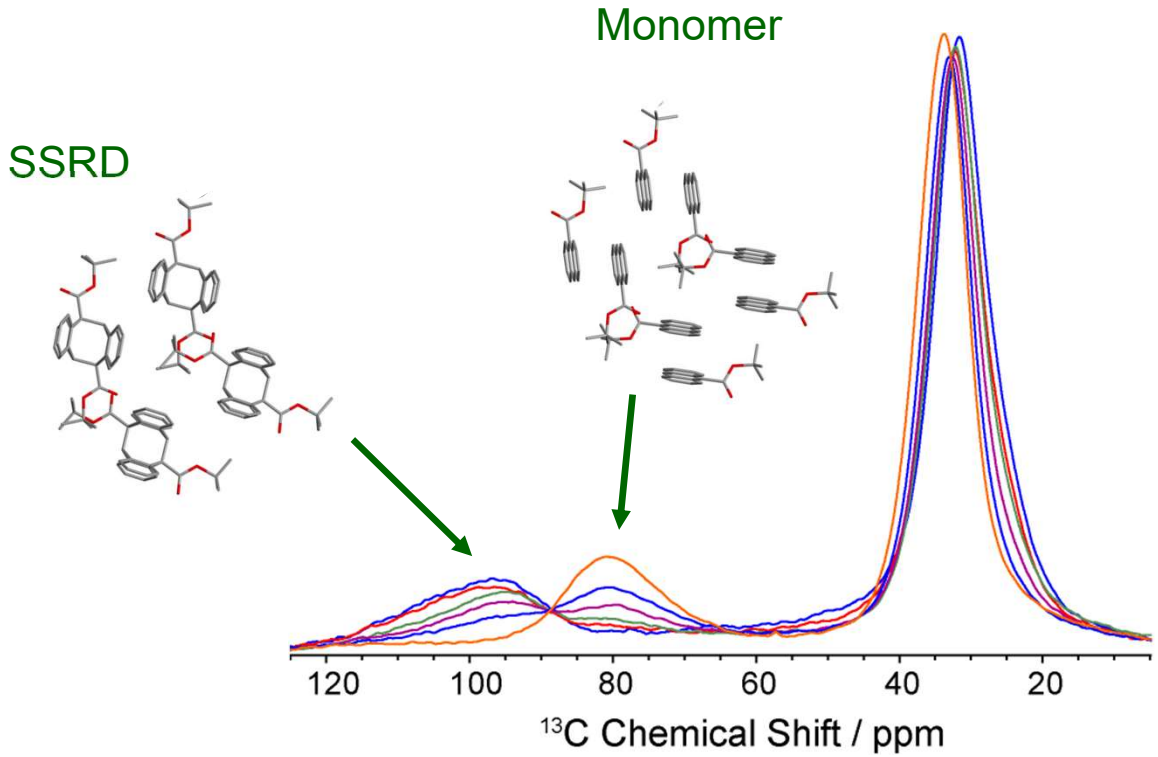




h v

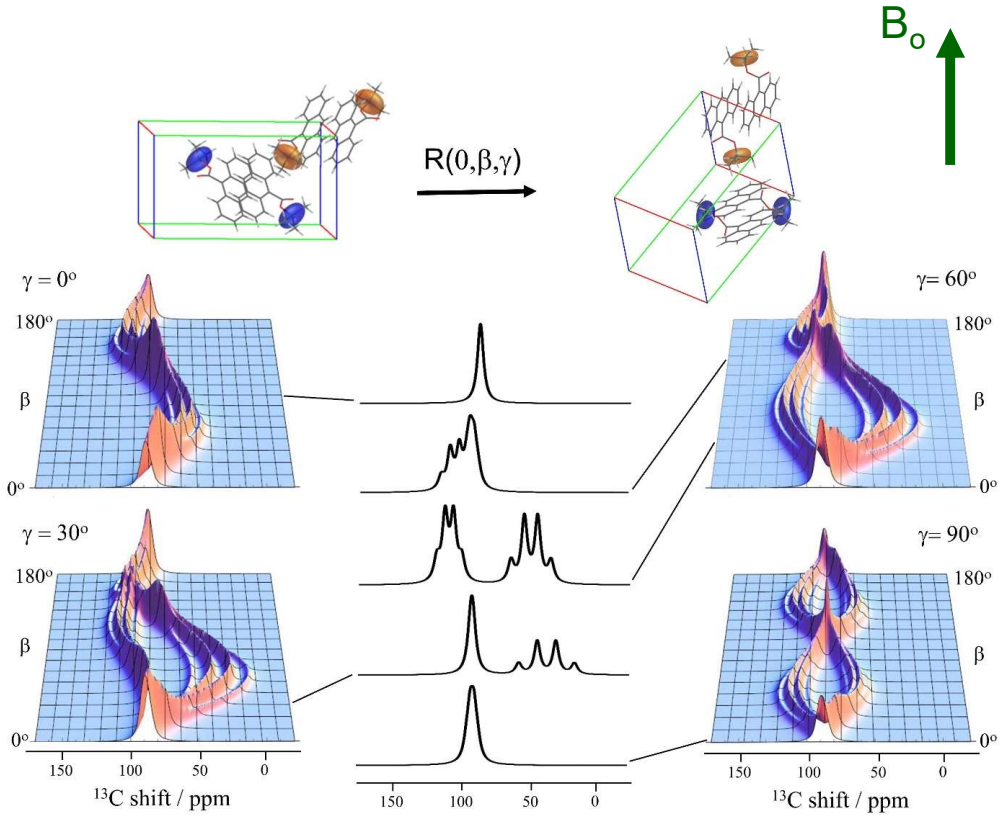


Two-State Single-Crystal to Single-Crystal Reaction



NMR Crystallography: Orienting the Monomer and SSRD Unit Cells

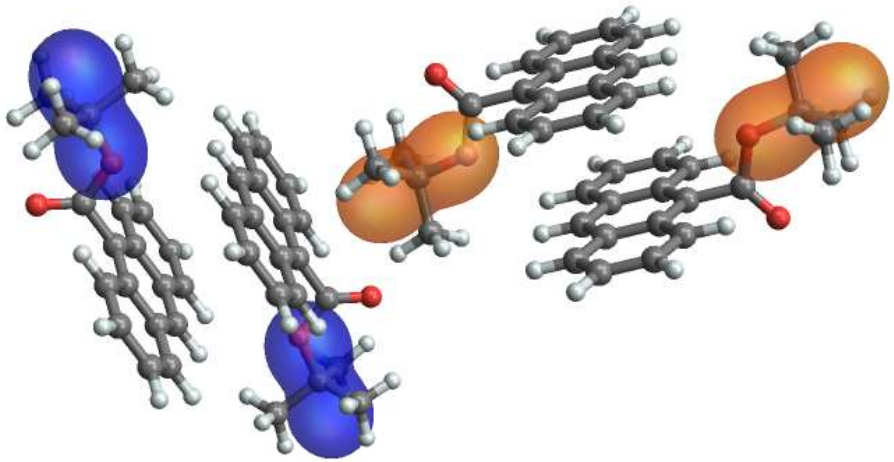
- Using the first principles shielding tensor and its alignment in the crystal frame, we can predict the spectra as a function of orientation of the unit cells in the magnetic field

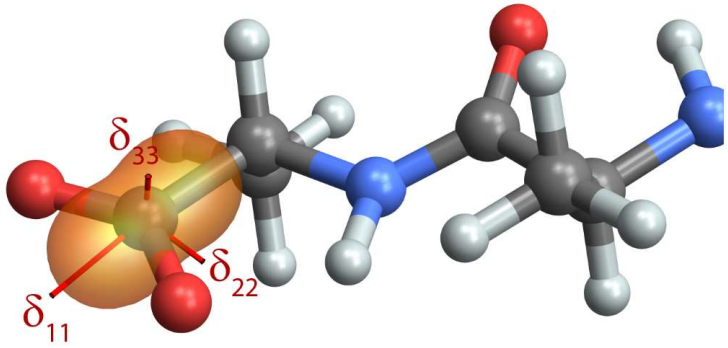


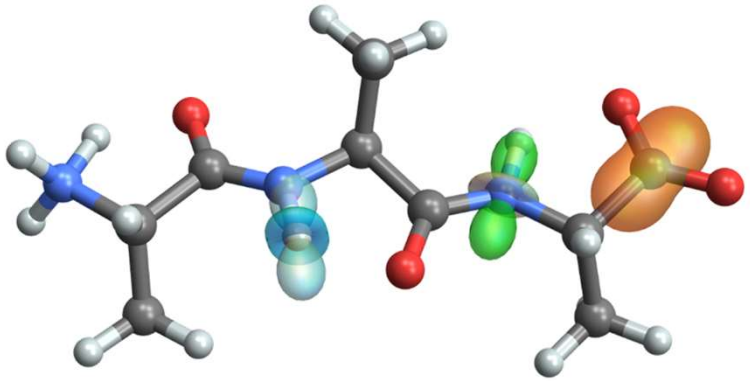
TensorView

- A software tool for displaying NMR tensors on molecular models
- Mathematica and MATLAB versions

Magn Reson Chem 2019, **57**: 211-223
SSNMR 2023, **123**: 101849







MATLAB version with Leo Svenningson
No MATLAB license required

Spherical Tensors and Rotations

- Two ways to treat this are the direct rotation in Cartesian form and the decomposition of the Cartesian tensor into irreducible spherical components that rotate in subgroups of rank 0, 1, and 2

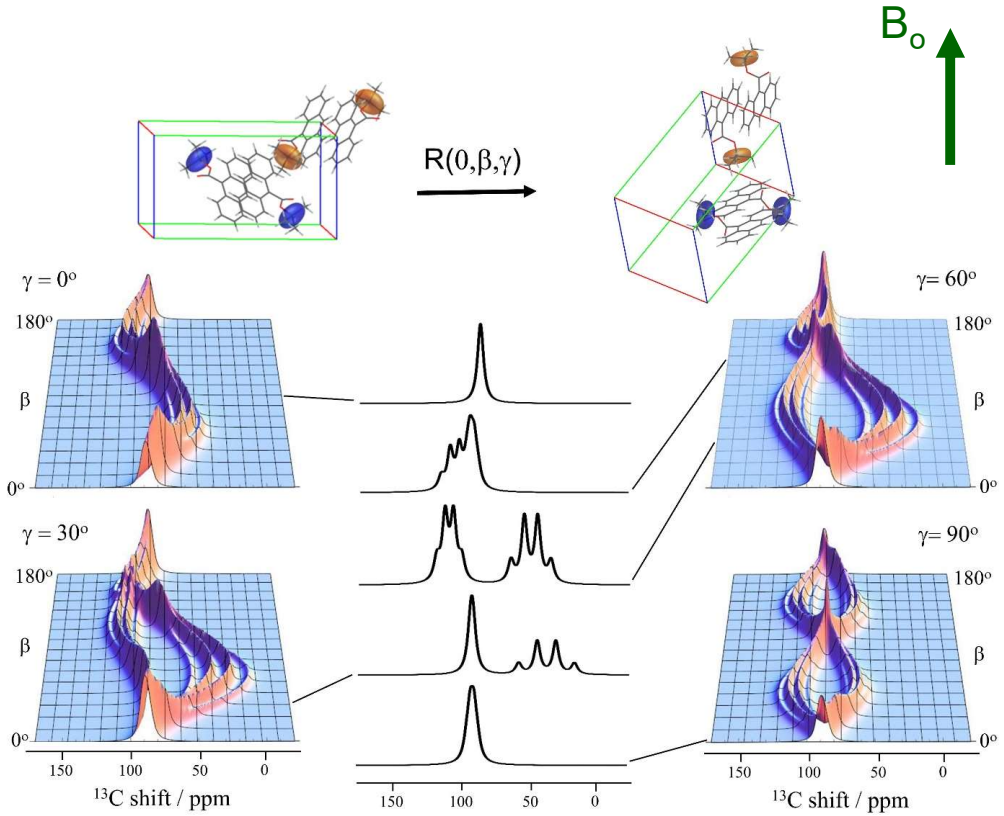
$$\mathbf{A} \xrightarrow{R} \mathbf{A}'$$

$$\mathbf{A}' = \mathbf{R}\,\mathbf{A}\,\mathbf{R}^{-1} \qquad \mathrm{a}'_{k\,q} = \sum_{p=-k}^{k} D^{(k)}_{pq}\left(\Omega_{R^{-1}}\right) a_{k\,p}$$

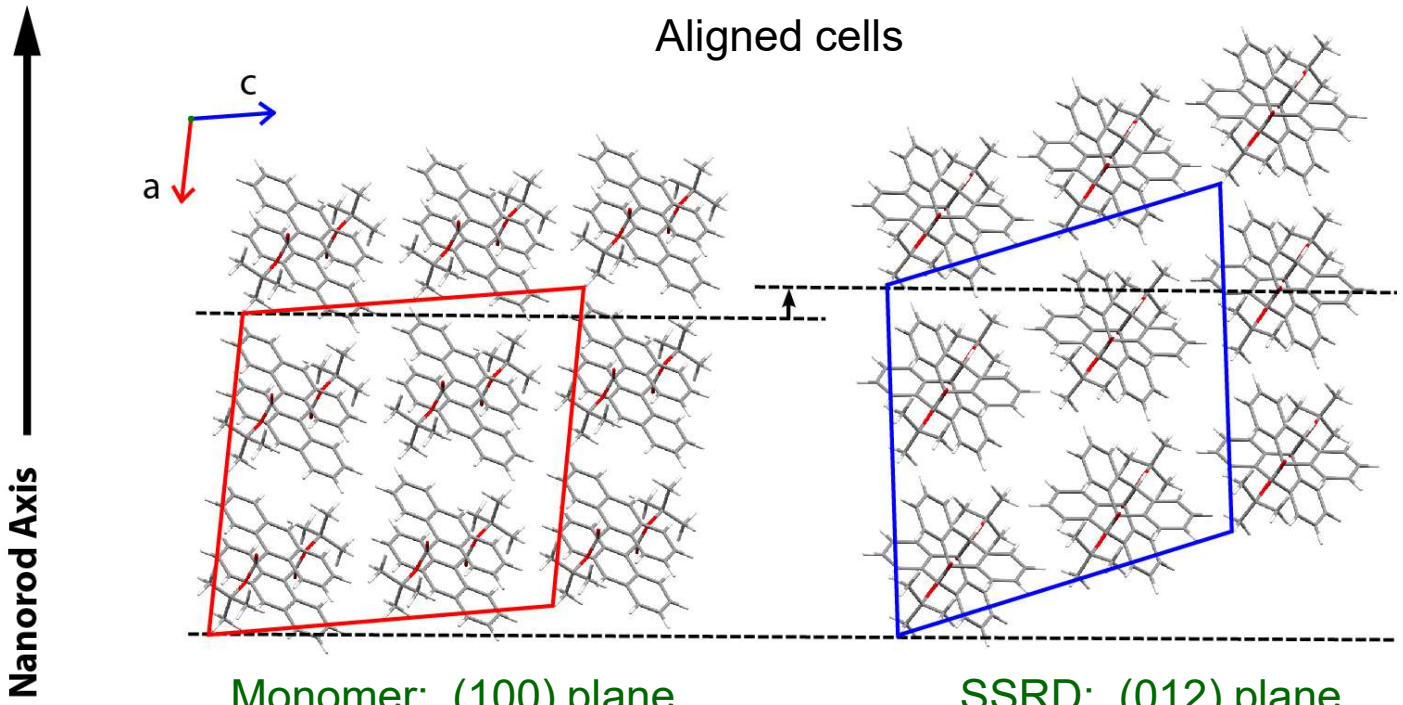
Mueller, *Concepts in Magnetic Resonance A*, **38A**, 221-235 (2011)
ENC tutorials 2015, 2019 – online at www.enc-conference.org

NMR Crystallography: Orienting the Monomer and SSRD Unit Cells

- Using the first principles shielding tensor and its alignment in the crystal frame, we can predict the spectra as a function of orientation of the unit cells in the magnetic field



Nanorod Expansion



Monomer unit cell axes

$$\mathbf{r} = u\mathbf{a}_m + v\mathbf{b}_m + w\mathbf{c}_m$$

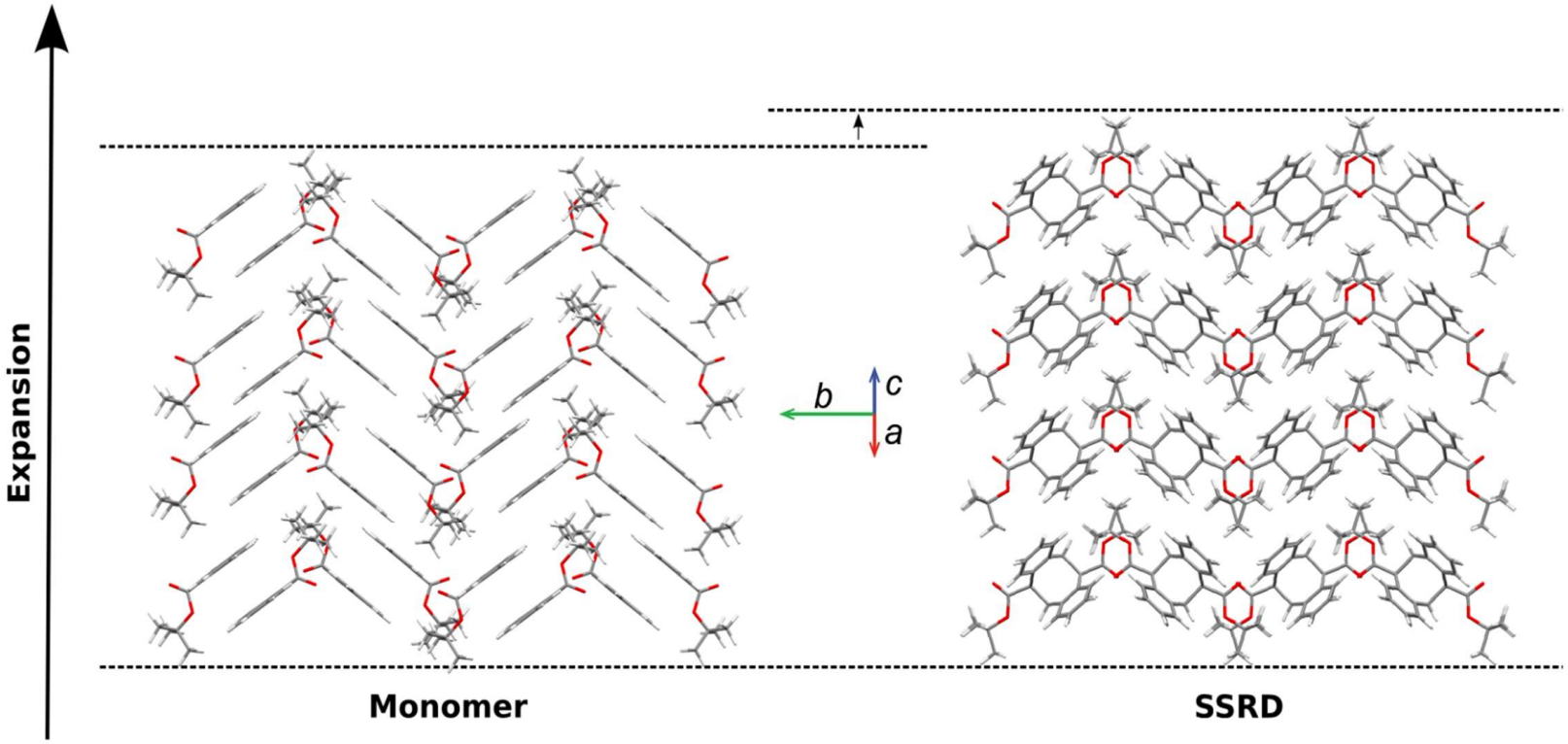
Transformed (effective) dimer unit cell axes

$$\mathbf{r}' = u\mathbf{a}'_d + v\mathbf{b}'_d + w\mathbf{c}'_d$$

Once aligned in rod frame, can measure microscopic expansion directly from equivalent lattice points

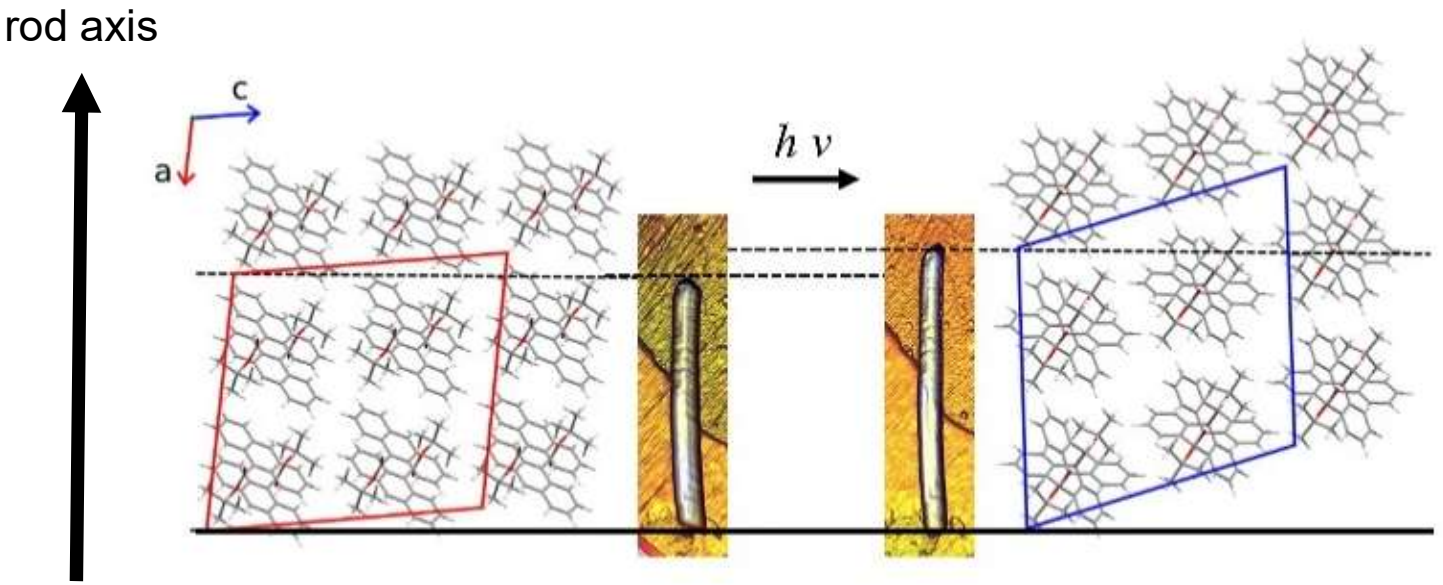
Expansion
NMR Alignment: 7.4%
Experimental: $8 \pm 2\%$

Underlying Mechanism

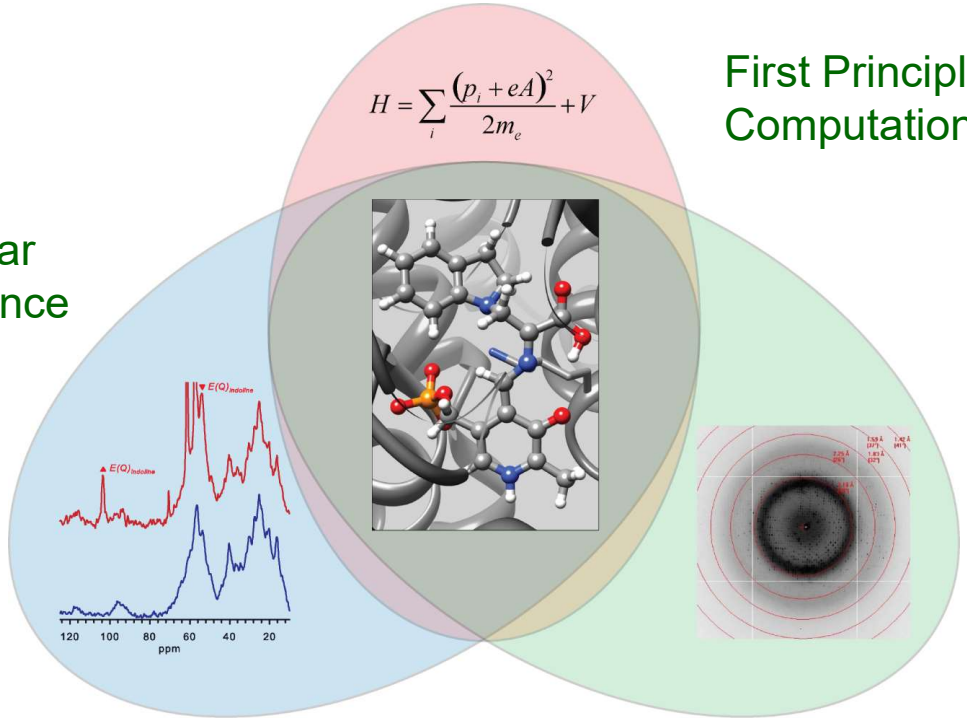


Conclusion

- NMR crystallography can establish the atomic-level basis for the macroscopic expansion
- Determines both the unit cells and their orientations relative to the shape change



Integrative Structural Biology of Enzyme Active Sites with NMR Crystallography



First Principles
Computational Chemistry

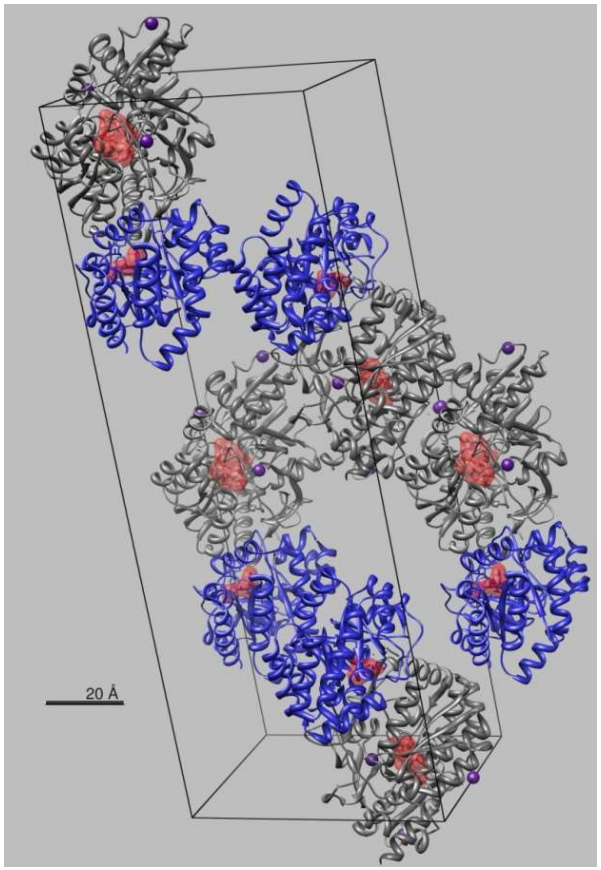
Solid-State Nuclear
Magnetic Resonance

X-Ray Diffraction

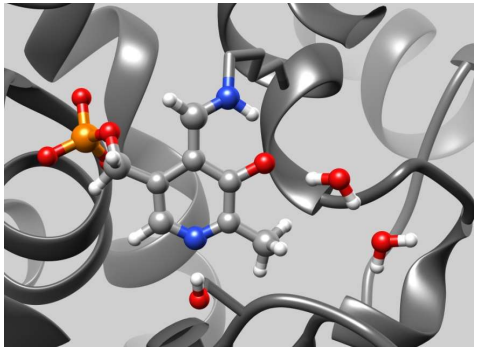
JACS **2016**, 138, 15214-15226
ACIE **2016**, 55, 1350-1354
PNAS **2022,** 119(2) e2109235119
PNAS **2022,** 119(4) e2114690119

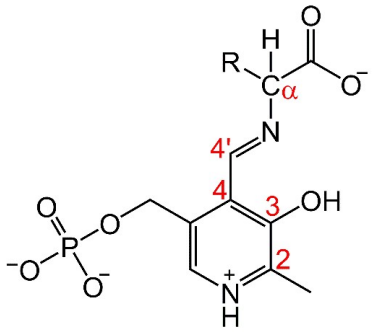
Funding: NIH MIRA

Tryptophan Synthase



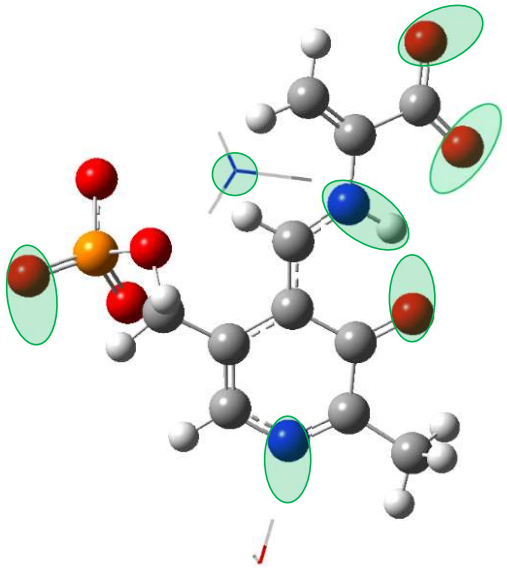
- 143 kDa, $\alpha_2\beta_2$ bi-enzyme complex
- Catalyzes the last two steps in the synthesis of L-Trp
- β-subunit cofactor: pyridoxal-5′-phophate (PLP)

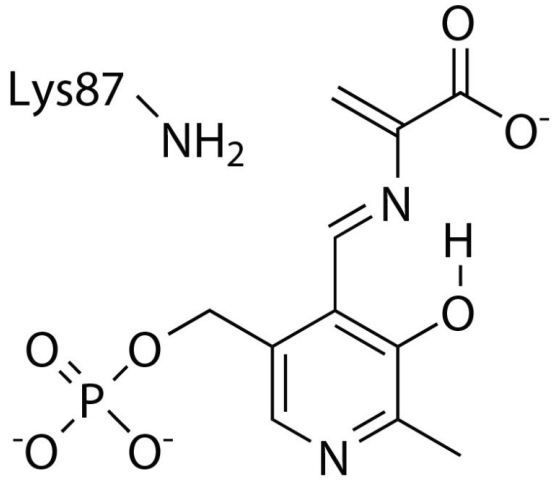




PLP - External Aldimine Form

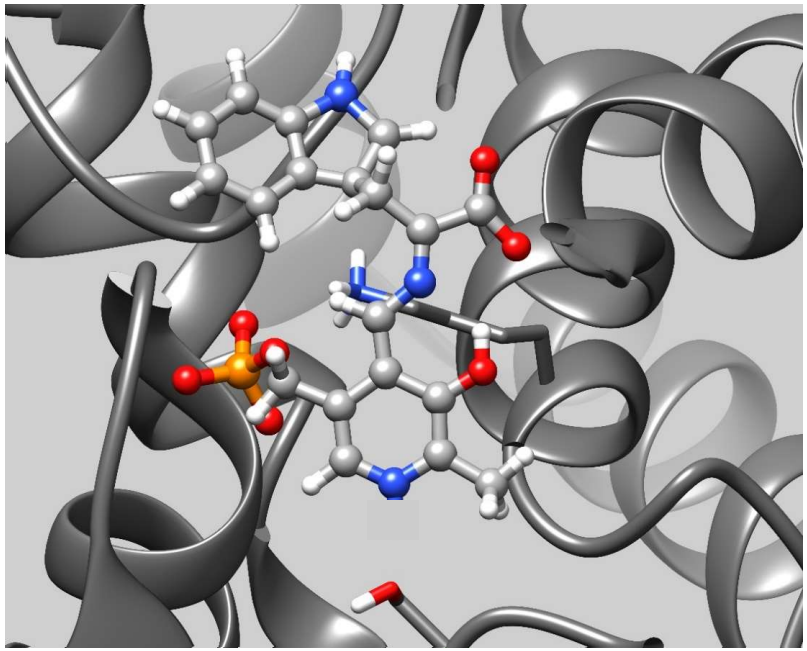
Protonation States in the Active Site: Mechanism and Inhibition





Refining Active-Site Chemical Structure in Tryptophan Synthase

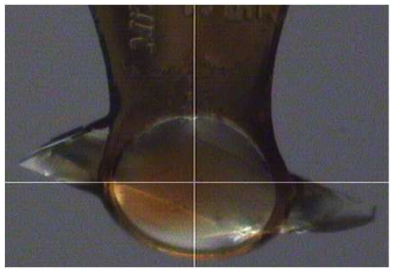
Challenge to NMR Crystallography: Identify and characterize intermediates, including their protonation states

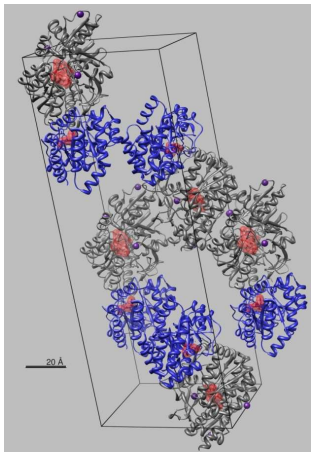


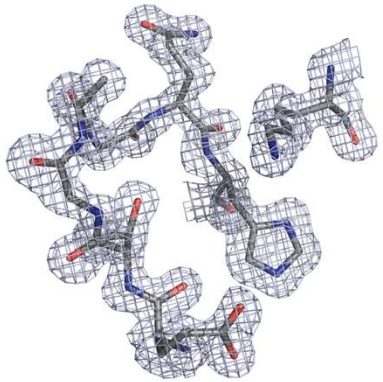
"Chemically-Rich"

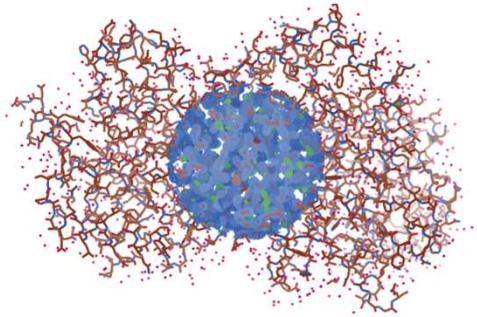
Step 1: X-Ray Crystal Structure

Intermediate/ Analog	*α-Site Ligand*	*β-Site Ligand(s)*	*Cation*	*Res (Å)*	*PDB ID*
$E(A_{in})$	F9	-	Cs^+	1.30	4HT3
$E(A_{in})$	F9	L-Trp	Cs^+	1.18	5CGQ
$E(A_{in})$	F9	L-His	Cs^+	1.60	7LV5
$E(A_{in})$	F6	-	Na^+	1.82	5BW6
$E(A_{in})$	F6	F6	Na^+	1.65	4Y6G
$E(A_{in})$	F6, F6	F6	Na^+	1.75	4WX2
$E(A_{in})$	F6, F6	F6	Na^+	1.54	4ZQC
E(A-A)	F9	L-Ser	Cs^+	1.45	4HN4
E(A-A)(BZI)	F9	LSer, BZI	Cs^+	1.75	4HPX
$E(C_3)_{2AP}$	F9	L-Ser, 2AP	Cs^+	1.45	4HPJ
$E(A_{ex})$	-	L-Ser	Na^+	1.45	6DZ4
βQ114A $E(A_{in})$	F9	-	Cs^+	1.65	6C73
βQ114A E(A-A)	F9	L-Ser	Cs^+	1.64	6D0V
βQ114A $E(A_{ex})$	F9	L-Ser	Cs^+	1.64	6DZO
βQ114A $E(C_3)$	F9	L-Ser, 2AP	Cs^+	1.70	6O1H
E(A-A)	F9	L-Ser	NH_4^+	1.40	7MT4
E(A-A)	F9	L-Ser	Cs^+	1.50	7MT5
E(A-A)(BZI)	F9	LSer, BZI	Cs^+	1.70	7MT6





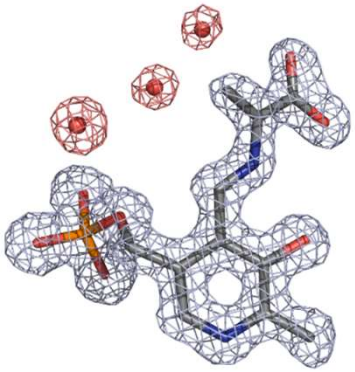


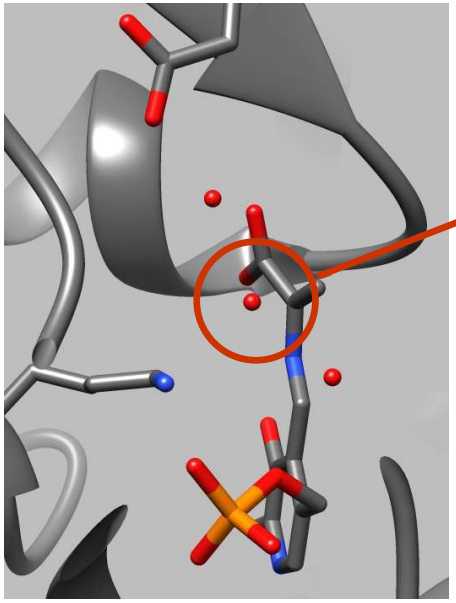


X-Ray Collaborators: Eduardo Hilario (UCR Biochemistry) and Tim Mueser/Tori Drago (U Toledo)

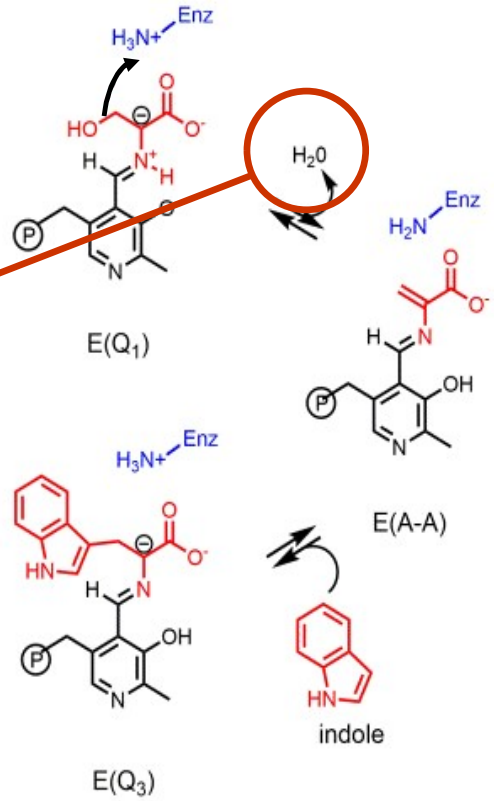
α-Aminoacrylate Crystal Structure

- Formed by the acid-catalyzed loss of hydroxide
- Structure shows crystal waters in the active site adjacent to the substrate C^{β}
- Tempting to think it could be the hydroxide!



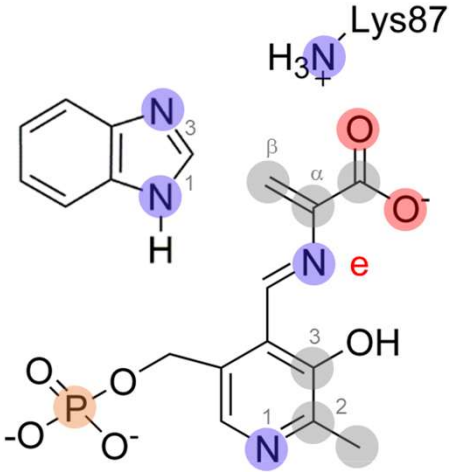


PDBID: 4HN4

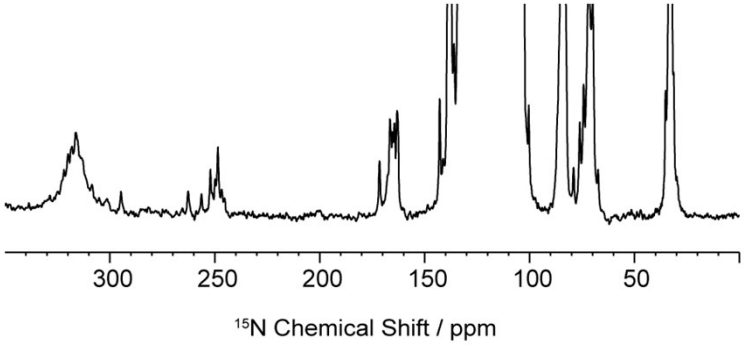


Step 2: NMR Spectroscopy

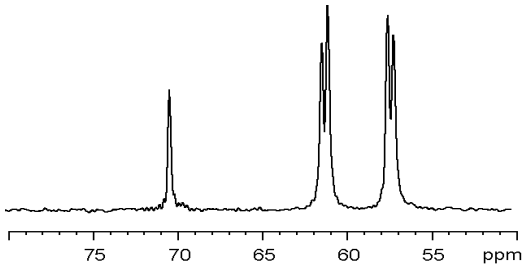
- Prepare microcrystals of enzyme for solid-state NMR under analogous conditions as X-ray
- Make use of labeled protein, cofactor, and/or substrates and establish steady-state concentration of intermediates in the catalytically-active crystals

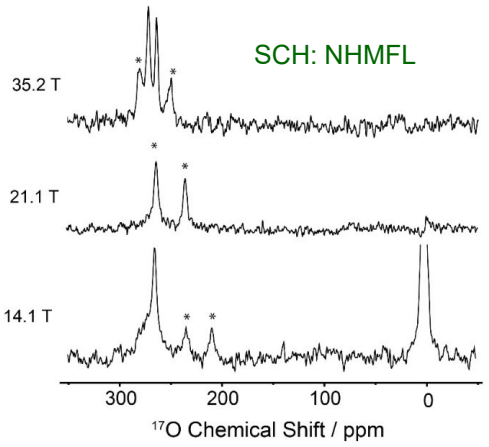


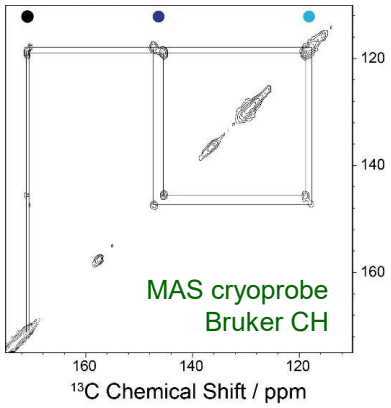
bound substrate: magic-angle-spinning solid-state NMR

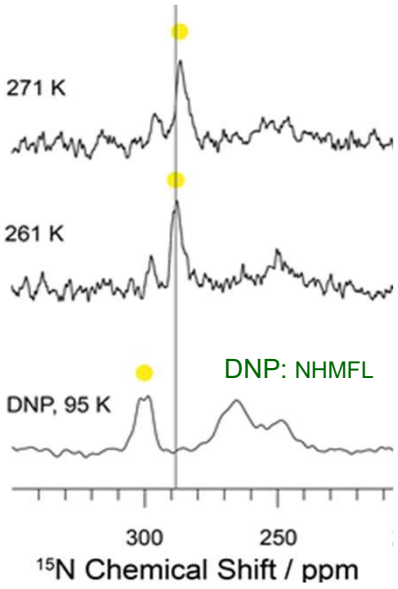


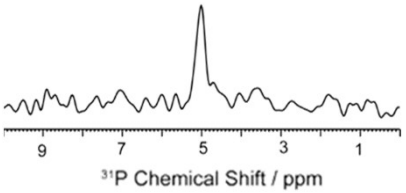
free substrate in mother liquor: solution-state NMR

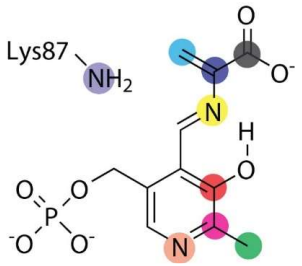


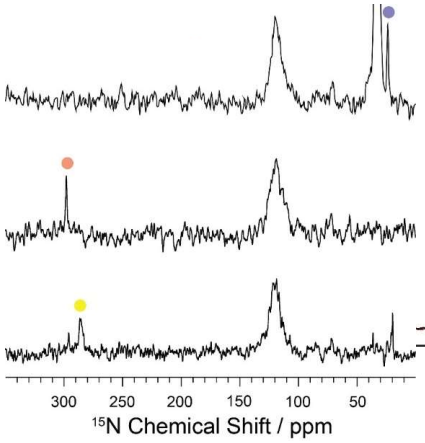


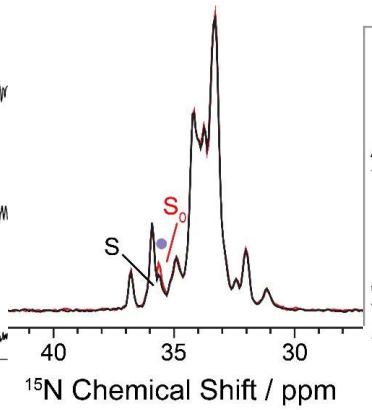


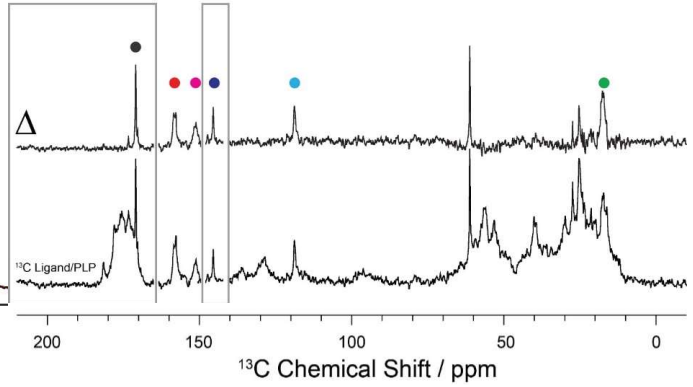








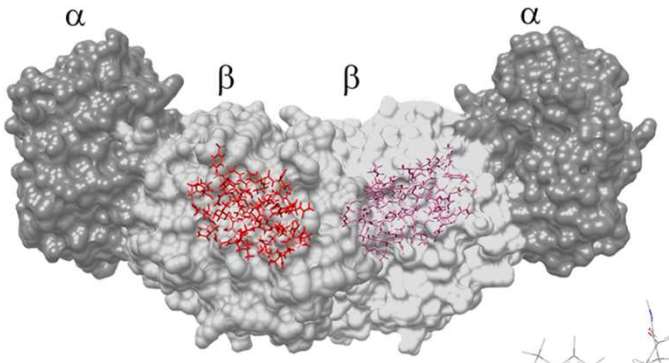




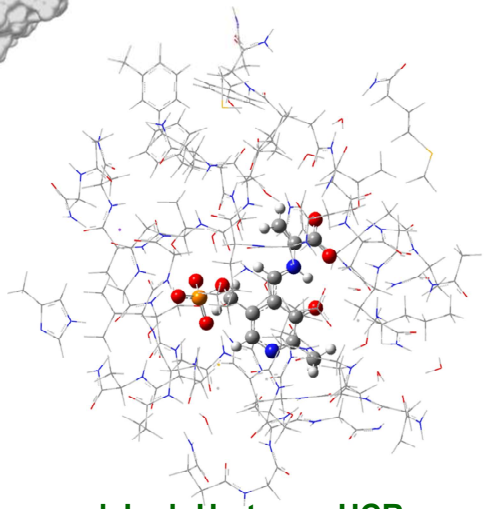
Chemical Shifts / ppm

	E(A-A)
C^{α}	145.6
C^{β}	118.8
C'	170.9
N (S.B.)	286.7
O	257 289
C2	151.2
C2'	17.5
C3	158.1
N1	297.6
P	5.2
N (Lys)	24.2

Step 3: First-Principles Computational Chemistry



PDB ID: 4HN4

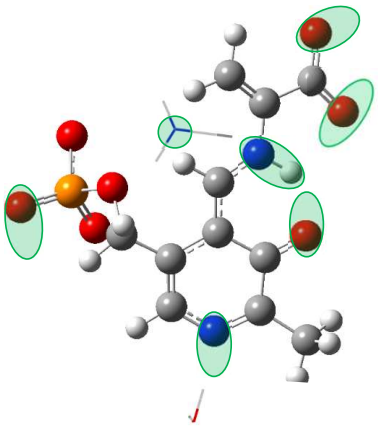


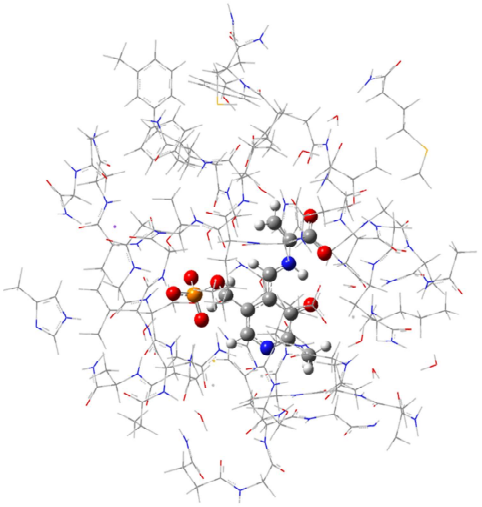
Benchmarked with Profs Greg Beran and Josh Hartman, UCR

Hartman & Beran, *JCTC* **10**, 4682{4872 (2014)
Hartman, Monaco, Schatschneider, Beran. *JCP* **143**, 102809 (2015)
Hartman, Kudla, Day, Mueller, Beran, *PCCP* **18**, 21686 (2016)
Harman, Neubauer, Caulkins, Mueller, Beran, *JBNMR* **62**, 327 (2016)

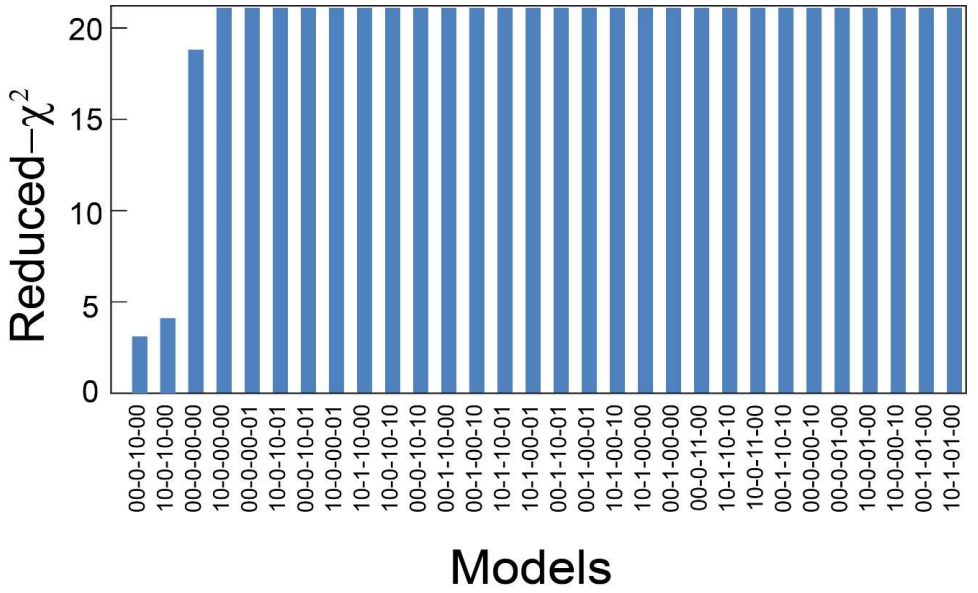
- Place the chemistry of the active site in full structural context
- Cluster model of active site: ~700 atoms
- Select residues with at least 2 atoms within 7 Å of substrate/cofactor
- Initial hydrogen-only MD scan
- Fully quantum-mechanical geometry optimization and NMR chemical shift calculation using DFT and locally-dense basis sets
- If we have the correct structure we expect
  - ^{13}C to within 1.5 ppm RMSD
  - ^{15}N to within 4.3 ppm RMSD
  - ^{17}O to within 7.5 ppm RMSD
- Linear rescaling from shielding to shift determined a priori and benchmarked across test sets
  - Quantitatively test absolute agreement of predicted shifts with experimental data

Model Rankings

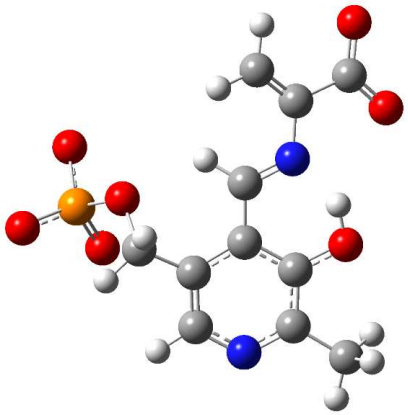




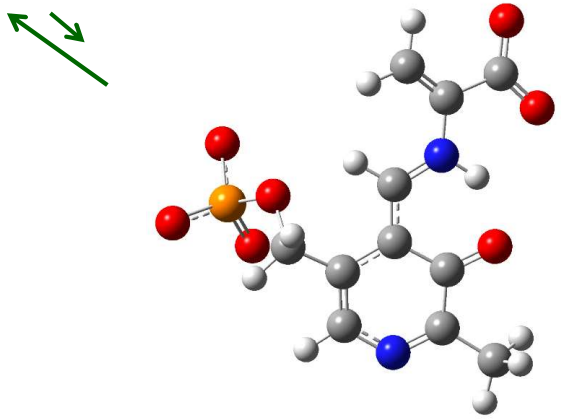
$$\chi^2 = \frac{1}{N}\sum_i \frac{\left(\delta_i^{\text{model}} - \delta_i^{\text{exp}}\right)^2}{\sigma_i^2}$$



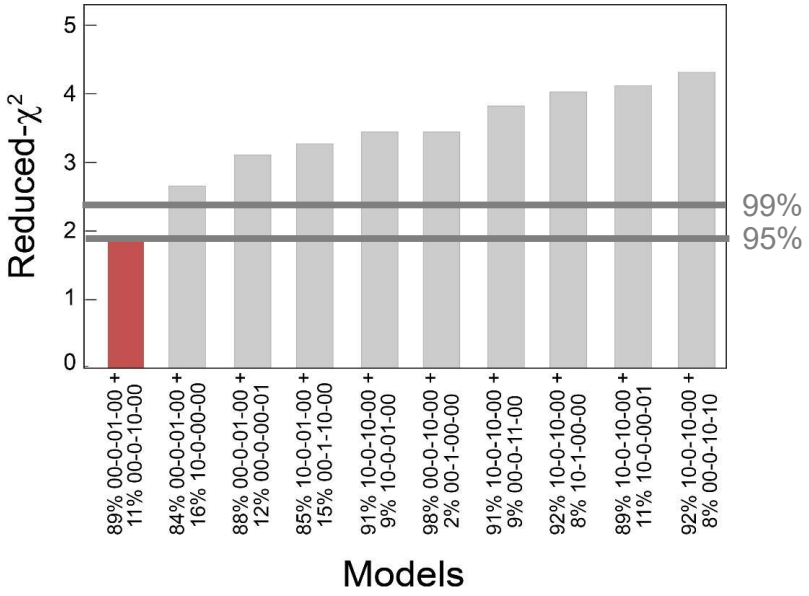
Fast-Exchange Equilibrium



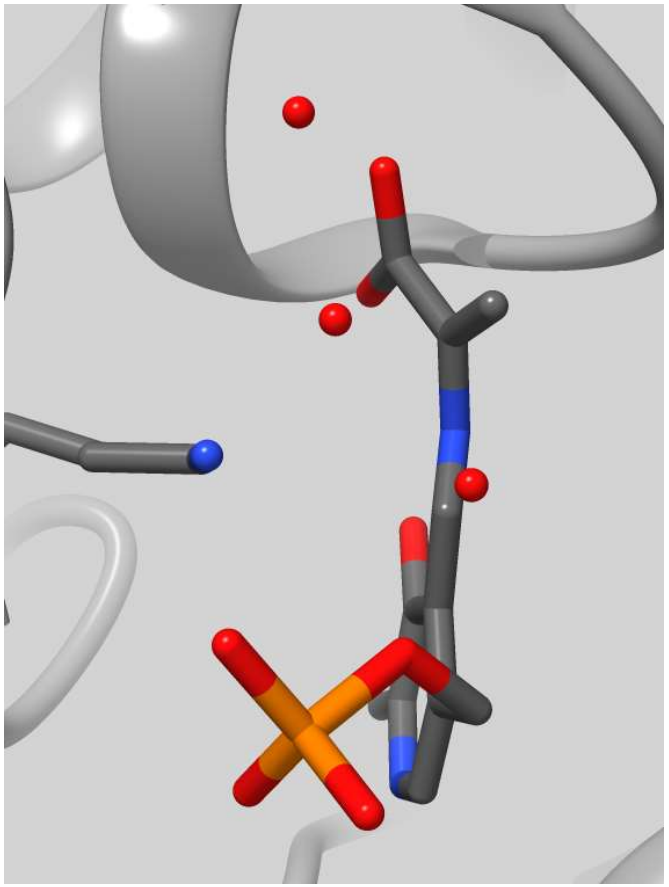
89% phenolic / enolimine



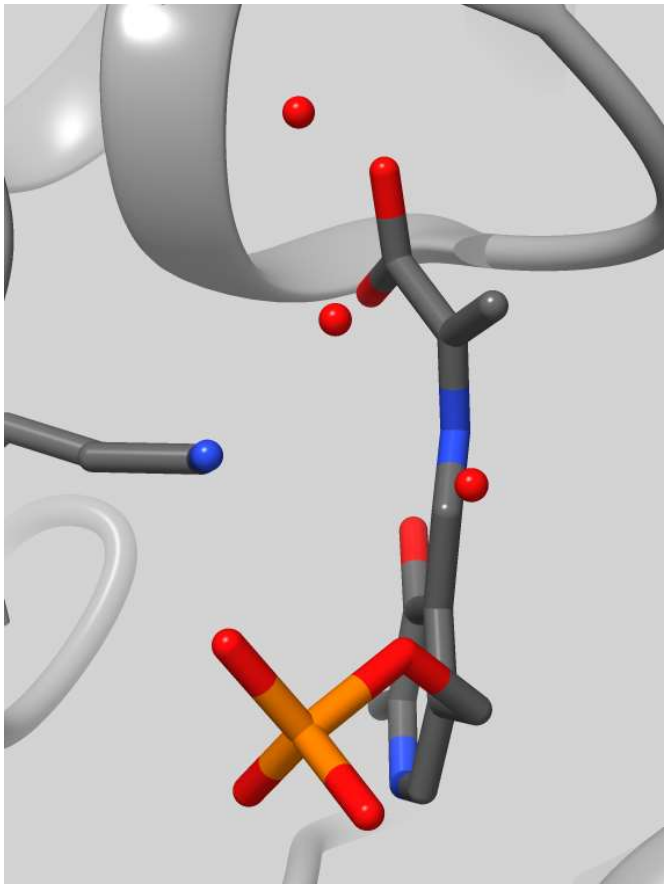
11% protonated Schiff base / ketoenamine



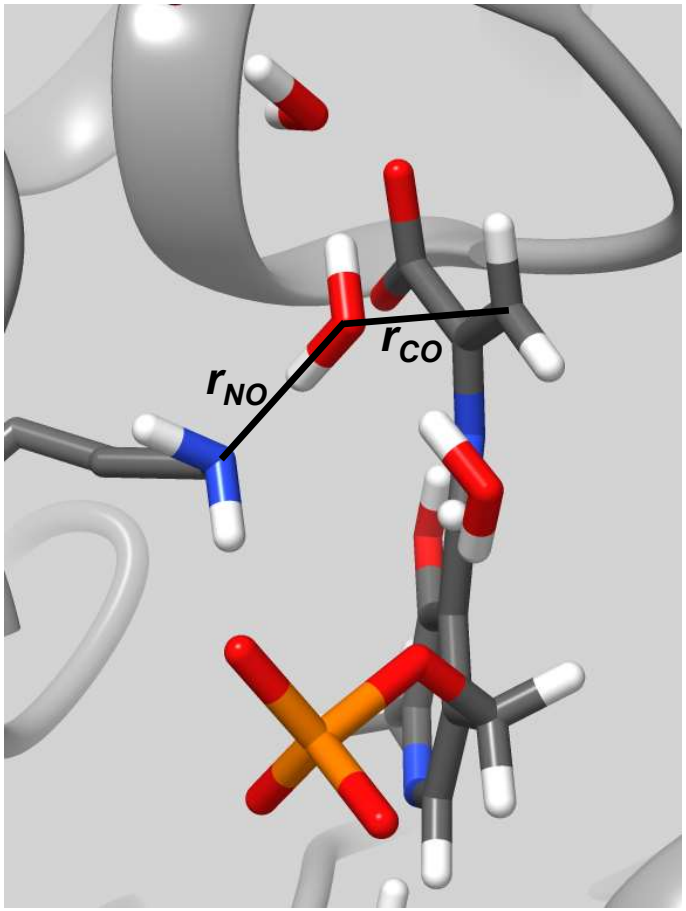
Reactivity and Transition States

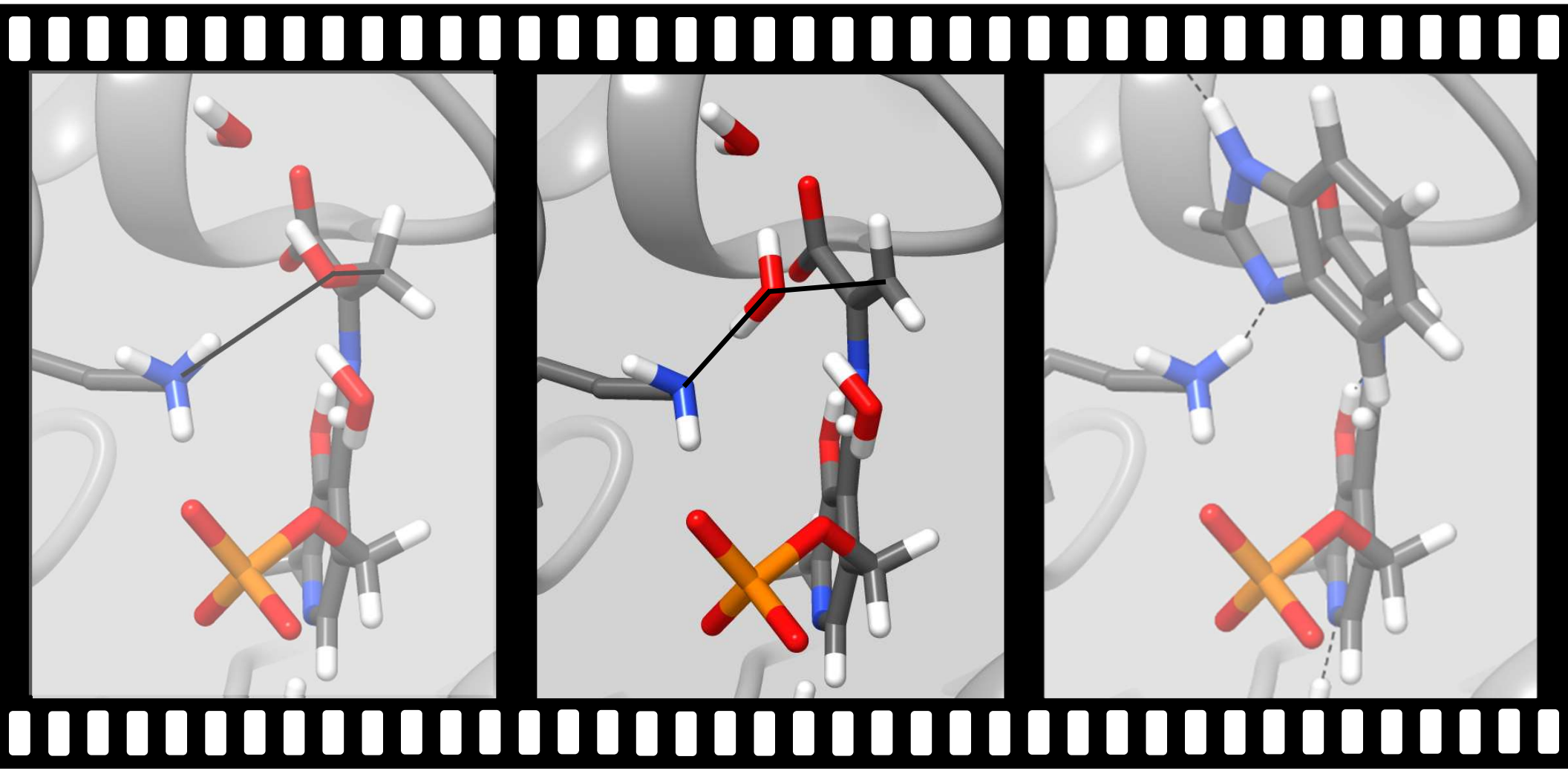


Reactivity and Transition States



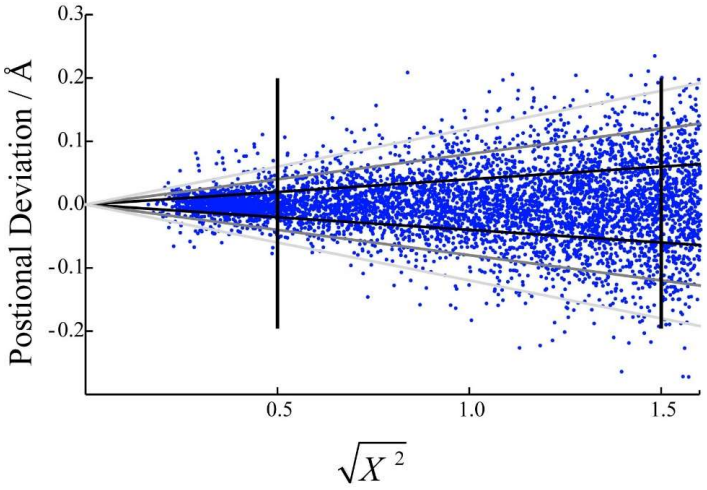
Water placement and orientation points back to the acid-base catalytic residue and along the reaction coordinate for the formation of the α-aminoacrylate intermediate





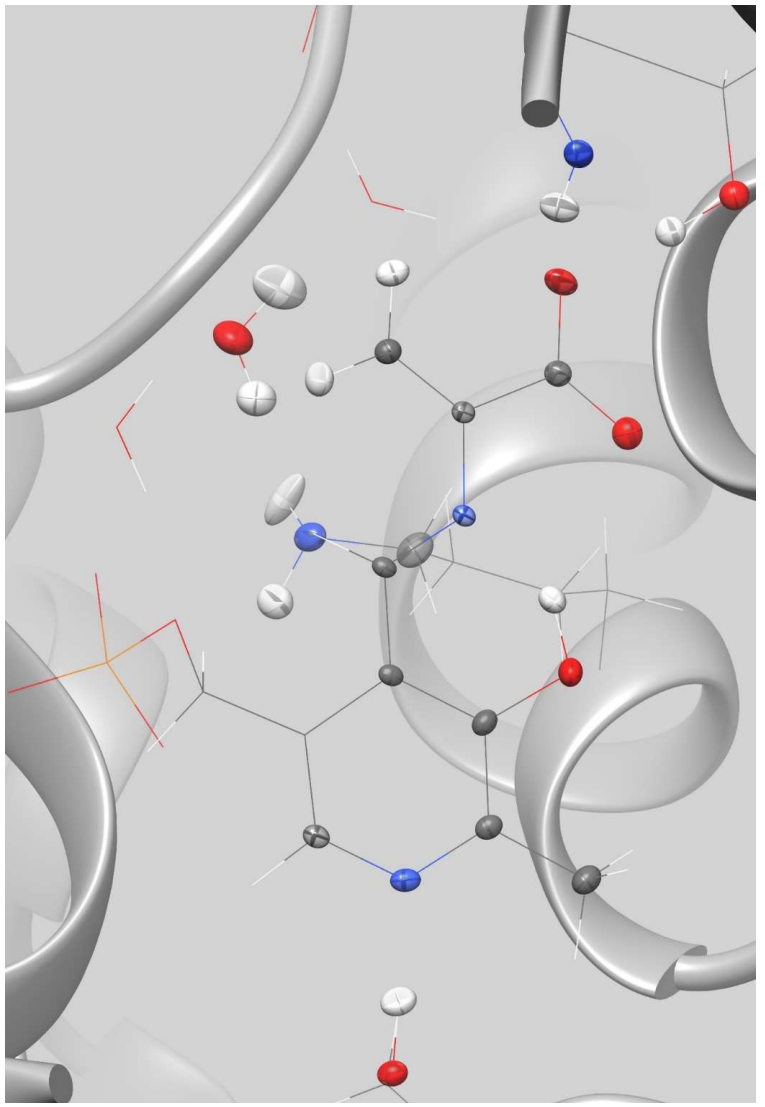
Positional Uncertainties

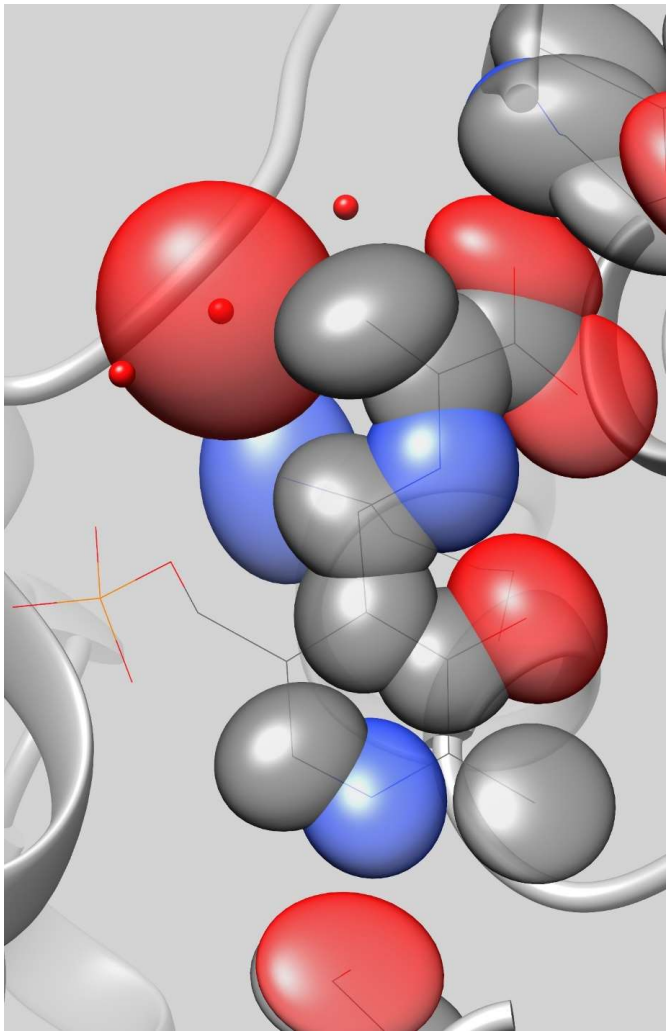
- Quantified the positional uncertainties in our structures by adapting the method for calculating ADP from Hofstetter and Emsley for molecular organic crystals (*JACS* 2017) to our cluster model approach for enzyme active sites
- Use low temperature molecular dynamics (1-150 K) to generate chemically reasonable perturbed structures and calculate their shifts
- Plot the corresponding reduced-X^2 vs. the positional deviations
- Funnel plot that allows us to find structures consistent with the chemical shift restraints at 95% certainty
- These define anisotropic displacement parameters



Average positional RMSD

- 0.11 Å for heavy atoms
- 0.17 Å for H atoms





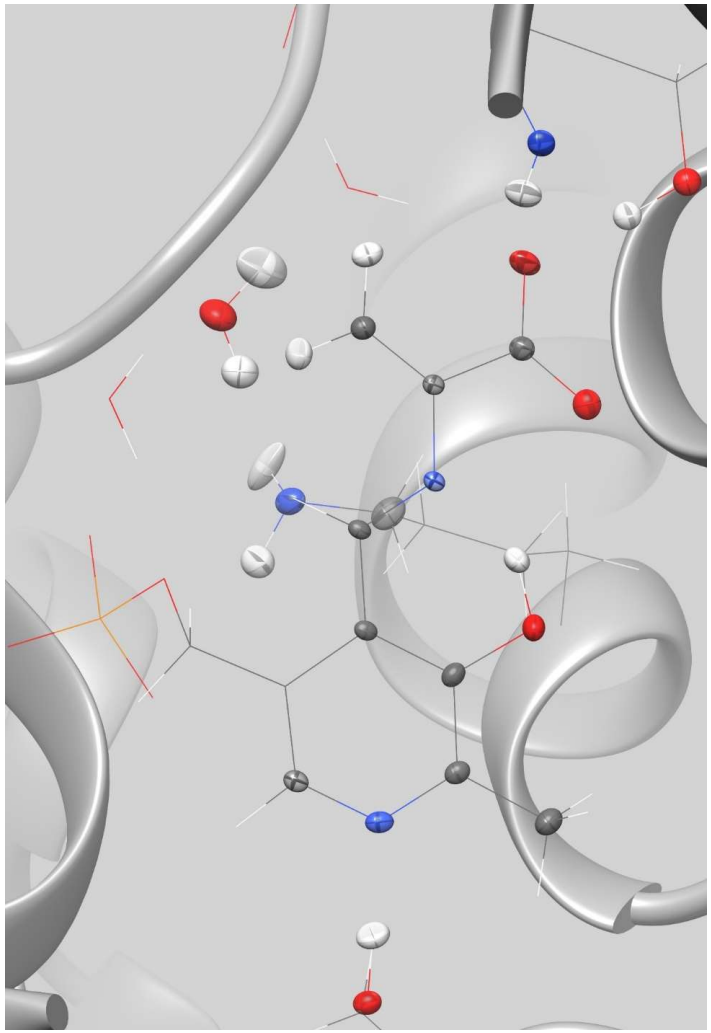
X-Ray Crystallography

Positional Uncertainties

Average positional RMSD

- 0.11 Å for heavy atoms
- 0.17 Å for H atoms
- 6.5 x smaller than X-ray

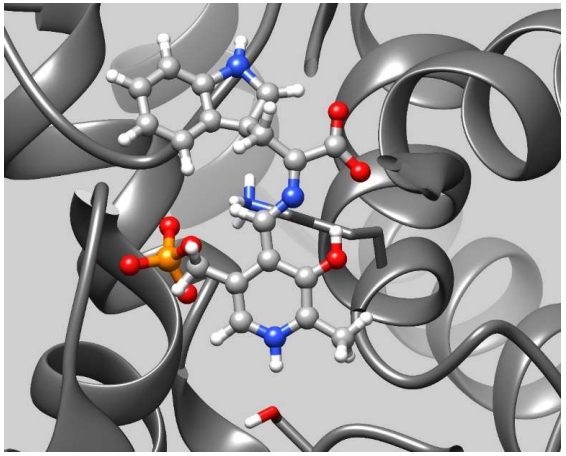
  (yes, this is not a fair comparison!)
- Similar in size to NMRX ADP for molecular organics crystals
- Suggests that NMRX ADP may be independent of molecular size

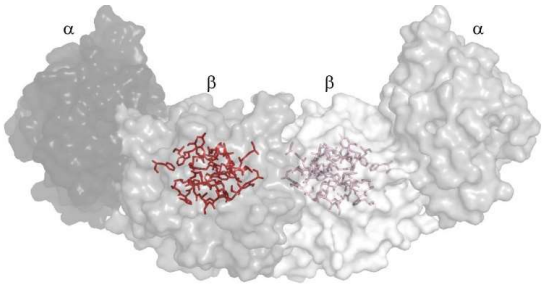


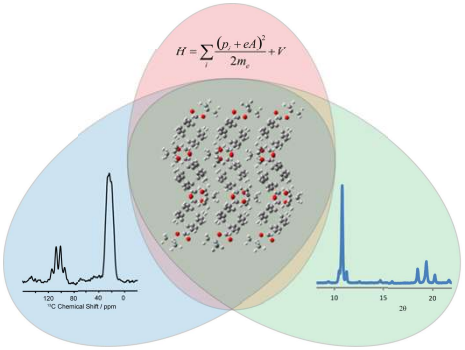
NMR Crystallography

NMR-Assisted Protein Crystallography

- Structure and dynamics
- In TS, identifies the active site protonation states and tautomeric exchange
- Informs us about transition states into and out of the aminoacrylate species







Summary

1. Pick a good problem!
2. Candidate structures: comprehensive list
3. NMR restraints: as many as possible
4. Accurate chemical shift prediction: appropriate level of theory and basis set
5. Quantitative ranking of models: Monte Carlo!