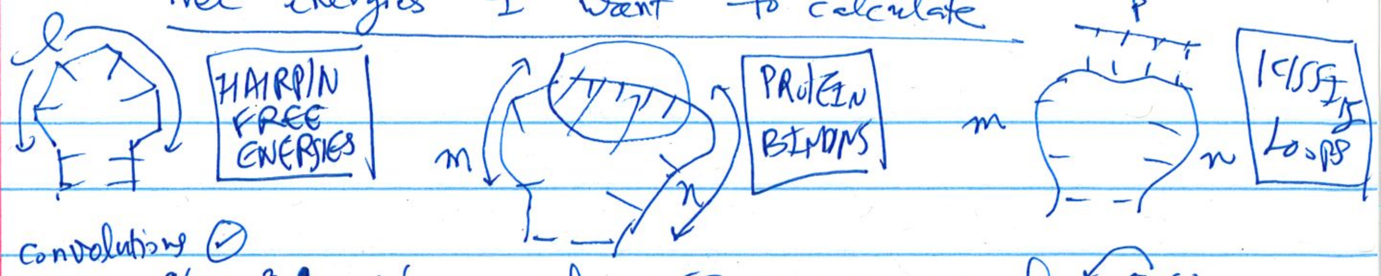
MOTIVATION: Ensembles and Free energies I want to calculate

4 July, 2019



- convolutions ✓

In SE(2), $N^6 + l\, N \times N^3 + N^6$ (could transform to basis; sum over index; back-transform)

where N = maximum order of harmonic

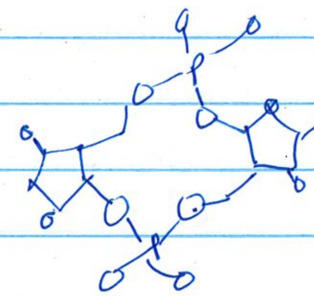
- convolutions ✓ via harmonic transform SE(3)
- products in real space $\stackrel{?}{=}$ product in harmonic space (parseval's theorem) [check in SE(2), SE(3)]

For SE(3),
- Should precalculate harmonic transform for fixed bond length, fixed bond angle.

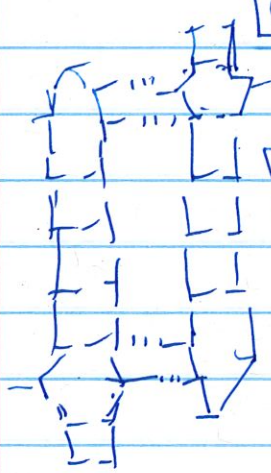
- sterics through backtrack sampling

In SE(2), N^6 (back transform)

SAMPLE RING CONFORMATIONS (Ribose)

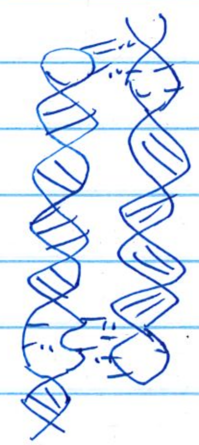


SAMPLE RING CONFORMATIONS cyclic dinucleotide → COMPARE to PDB



TERTIARY FOLDING ENERGIES

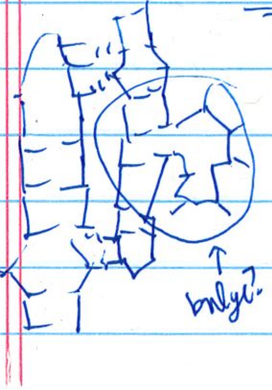
(Gaussian convolution SE(2)) ✓



(Gaussian convolution SE(3)) ✓

Could also solve through NORMAL MODES. ✓

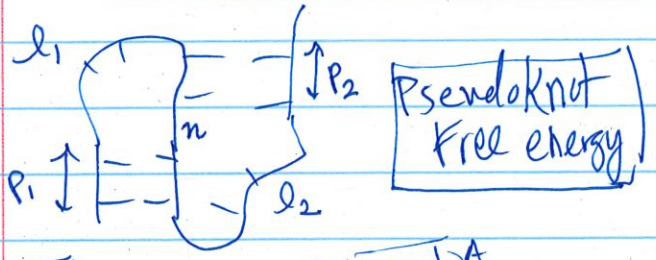
But NEED... transforms at BASE PAIRS, might be easier to leave coordinates on nucleotide bases?

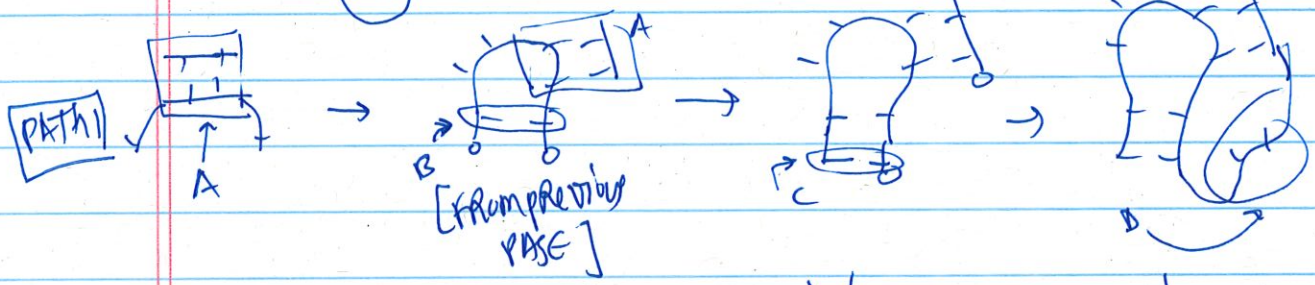


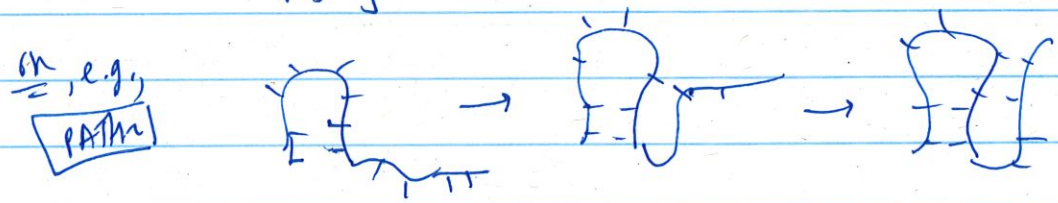
RAPID SEARCH OF PSEUDOKNOTS & A-MINOR & T-LOOP CONTACTS →

4-July, 2019

Ensembles and free energies I want to calculate

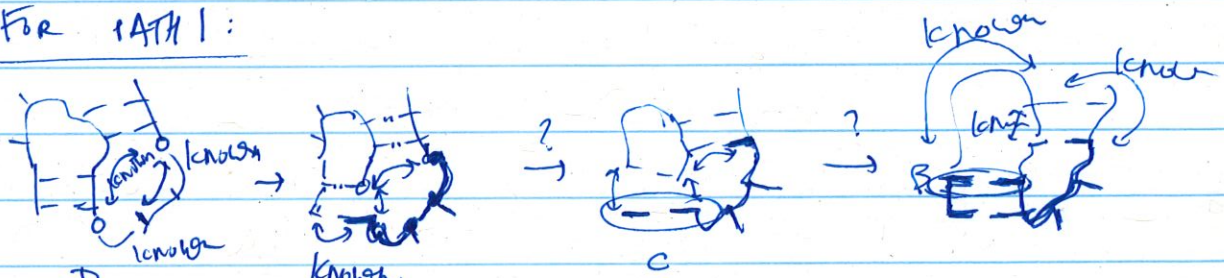


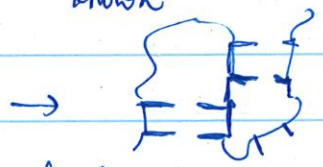




— Wow, but how would I backtrack sample? Follow the arrows backward...

FOR PATH 1:





— Hmm, tough because at several stages of choosing nucleotides there are multiple paths dictating answer.

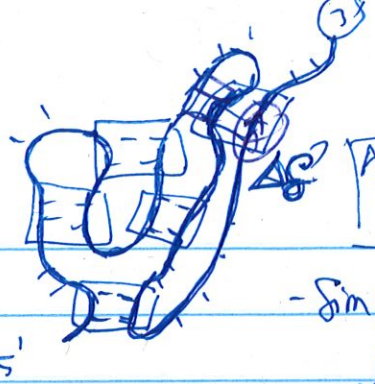
— PLAN WORK THROUGH 1D Toyfold calc.

- Alternative:

For pseudoknot, first sample each "Right" chunk, then:

$P_1^{tot} = P_1 P_2 P_3$

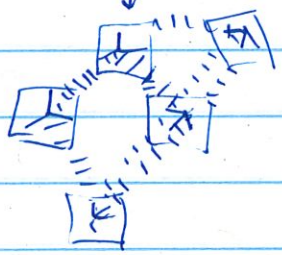
then resample loops 1, 2, 3



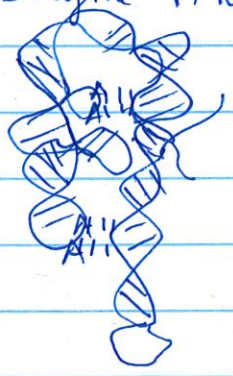
ARBITRARY PSEUDOKNOTTED & 4-MINORED NETWORK

- Simplifying Assumptions? HELICES ARE RIGID.
- What if there are a finite number of global topologies? ← Combinatorial SEARCH
  Then just sum over them with HARMONIC MODE APPROXIMATIONS

- If that's the case "Build-a-Fold" may be challenging...
  2D → 3D
  - Sample-and-minimize?
  - Enumerate?

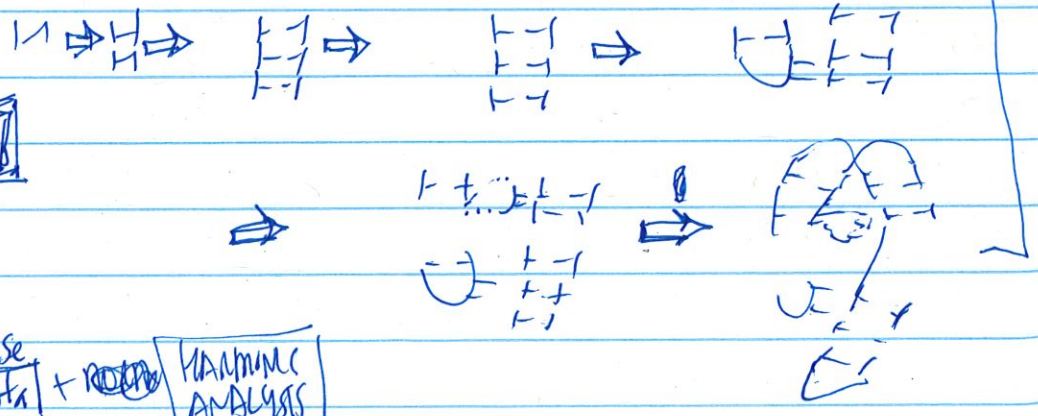


Imagine P4-P6 fold...



- REALLY COULD PREDEFINE ALL 6D ARRANGEMENTS of Helices.
- ~~AT~~ INSTANTIATE BIT-BY-BIT with 6D TRANSFORMS SET BY STERICS PREFS, MINIMIZE. REPEAT UNTIL COMPLETE

STERICS



STERICS?

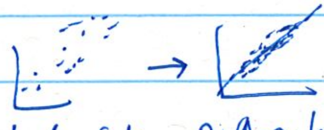
To test, could use PyRosetta + HARMONIC ANALYSIS

the "answer" will be different for different problems

So: target a paper for each problem

① TECTO RNA → OPTIMIZE k_BT_{eff} & 21 DOF TL/TR ?

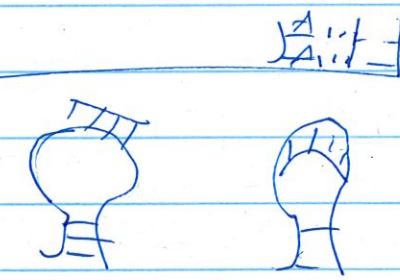
Helices. 6 + 15 TRANSFORM Σ^2



- Embed into RNAmake design calcs. with free → test via high throughput synthesis & ~~SHAPE~~ DMS-

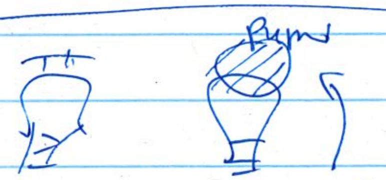
② calc. { Δs for each junction topology
↳ { CLUSTER model

How to test?



Array measurements → ~~Bend~~ loop, Kd.

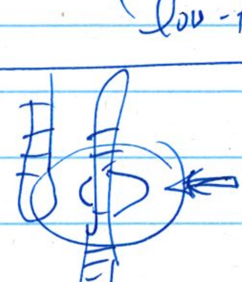
③ Pum



DATA [illegible]!

PROSPECTIVE!
(Do a few low-throughput measurements)

④



Predict REAL ensembles... Array (MaP) measurement.

ACTION ITEMS

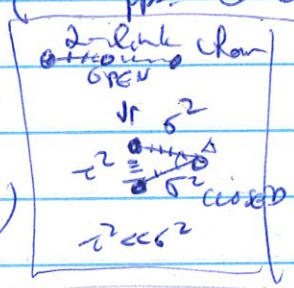
- gaussian convolutions in 1D, 2D, compare to HARMONIC TRANSFORMS
  3D → Tect RNA. Compute derivatives.
- Pum → HARMONIC Transforms, steers + samples. T[illegible]

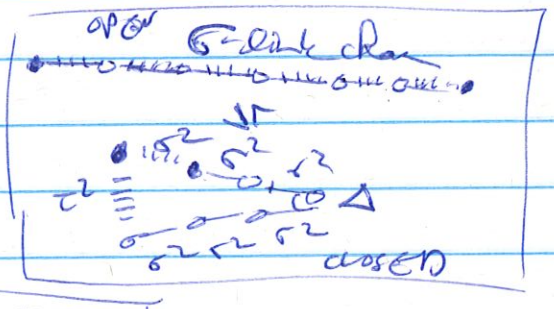
As a starting point for above, create a document summarizing worked examples of free energy estimation:

TARGET Audience: current & future group lab members

"internal memo"

SECTION 1. Motivation → { model problems in pp. ①-④





SECTION 2. 1D HARMONIC FORMATION (RING CLOSURE)

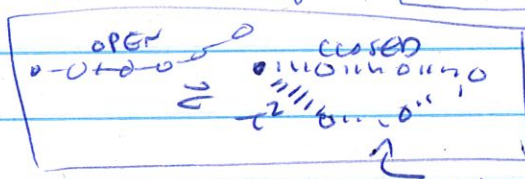
Subsections

Single CLOSED conformation

2A. ANALYTICAL Single C_{eff} [Gaussian convolution]

2B. Normal mode analysis

2C. Fourier analysis



odd # of links induces "STRAIN"

2D. Sim over STRAINED CONFORMATION (odd-link) → Repeat Single (C_{eff}), Normal mode analysis, Fourier analysis

2E. Sim over ALL OPEN & CLOSED CONFORMATIONS.

2F. Markov chain Monte Carlo [REJECTS style] → density of states.

SECTION 3. ENSEMBLE SAMPLING → $\langle X_0^2 \rangle$, $\langle r_{0,0} \rangle$, etc.

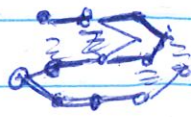
3A. Analytical

3B "BACKTRACKING"

3C. Normal mode analysis

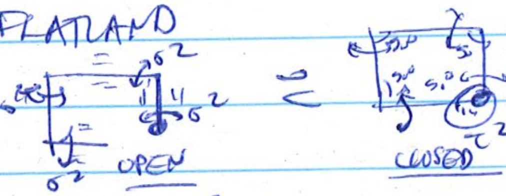
3D. Markov chain Monte Carlo.

3F. more complex 1D CASES:



Could BE A SEPARATE CHAPTER

SECTION 4. "FLATLAND"



DG.

4A. Gaussian convolution ⇒ coeff [COMPARE to STOCHASTIC samples of free chain]
↳ Also sample closed chain through MCMC

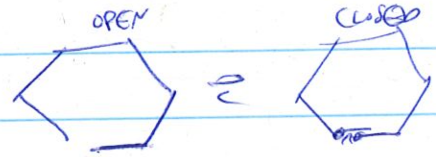
4B. Normal mode analysis
↳ Sample through normal mode expansion.

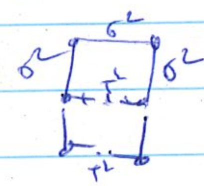
4C. SE(2) Fourier analysis [Bessel function $J(pr)e^{im\phi}e^{il\theta}$]
↳ Sample closed chain through BACKTRACKING.

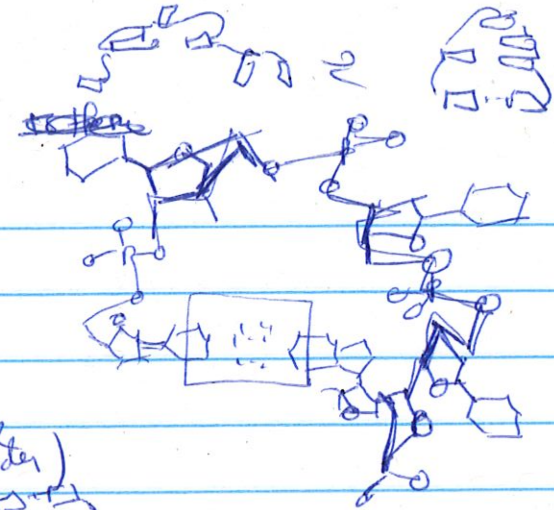
4D Allow chain to bend ±90°, 180°.

4E. STERICS (self-avoiding chain) → { Δδ d / BACKTRACK SAMPLING (numerical) }

4F. More complex cases

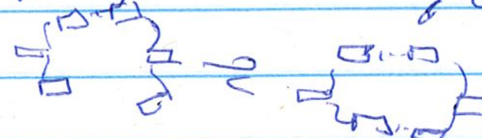




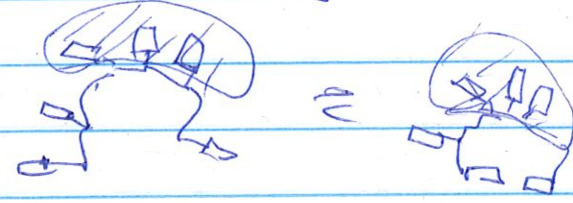


SECTION 5. THE REAL WORLD (2?)

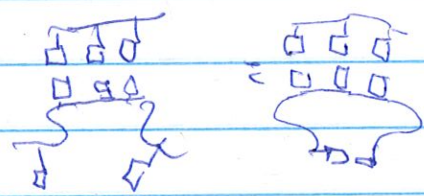
A. HAIRPIN FORMATION
(Harmonic transforms vs. Monte Carlo vs. Normal modes)
- PEAK AT n=4 (tetraloops)?
- BACKTRACK → SNRA loops? optical loops?



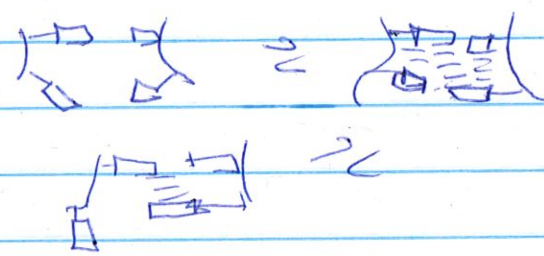
B. Internal loops & BULGES
↳ Again, do we reproduce stacking pattern in PDB?



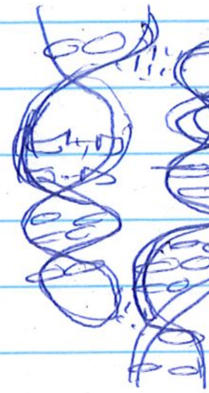
C. PROTEIN & OLIGO BINDING
- PUM1
- CREATE SERIES (?)

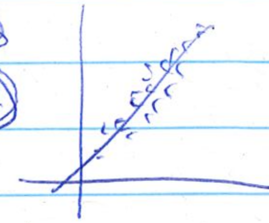


D. STACKED PAIRS
(Harmonic vs. MC Monte Carlo (RECCE?) vs. Normal modes)
→ COMPARE to PDB?



SECTION 6 HIGHER-ORDER STRUCTURE





6A: Tecto RNA
- (BASE PAIR STEP coordinates vs. BASE coordinates)
- NORMAL MODES vs. Gaussian convolution,
- NORMAL modes [All atom?]

{ Need TLTR "intrinsic" under varying conditions
[predict long lengths → PAPER

6B. Design [RNA make]
- Gaussians convolution → MOTIFS ↖ SERVER
- modularity costs.

6C. Multiple tertiary contacts
- Gaussians vs. Normal modes

SECTION 7. SEARCH OF COMPLEX GLOBAL FOLDS (2D → 3D Build-a-fold)

7A. Pseudoknots
- HARMONIC (?)
- GAUSSIAN (?)

7B. A-minor SEARCHING

7. TERTIARY CONTACT

SECTION 8.

High-Res stepwise sampling.