On Incorporating Non-canonical Amino Acids into Protein Design

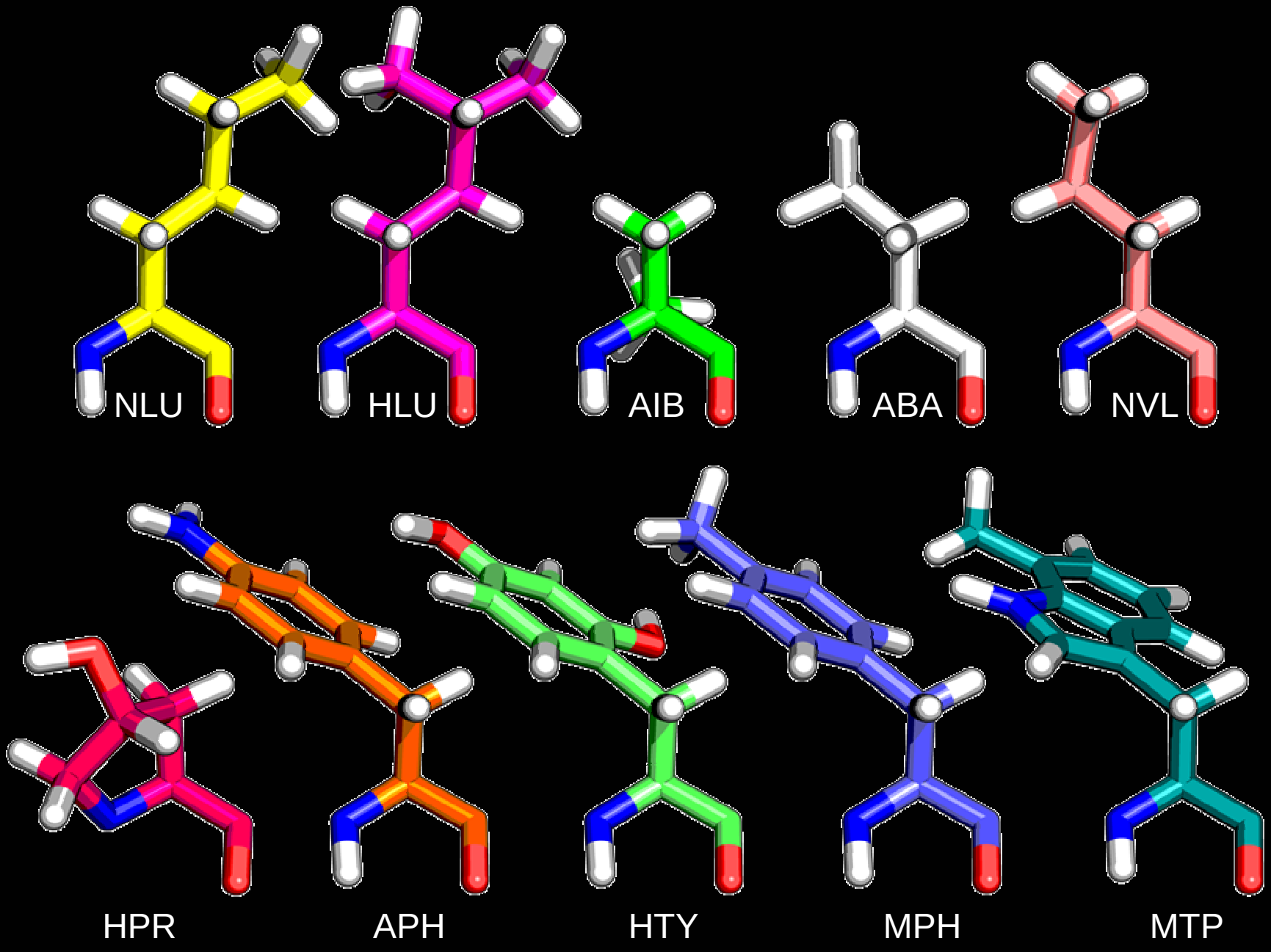
RosettaCon July 23rd 2008

P. Douglas Renfrew
Kuhlman Lab @ UNCCH

Introduction

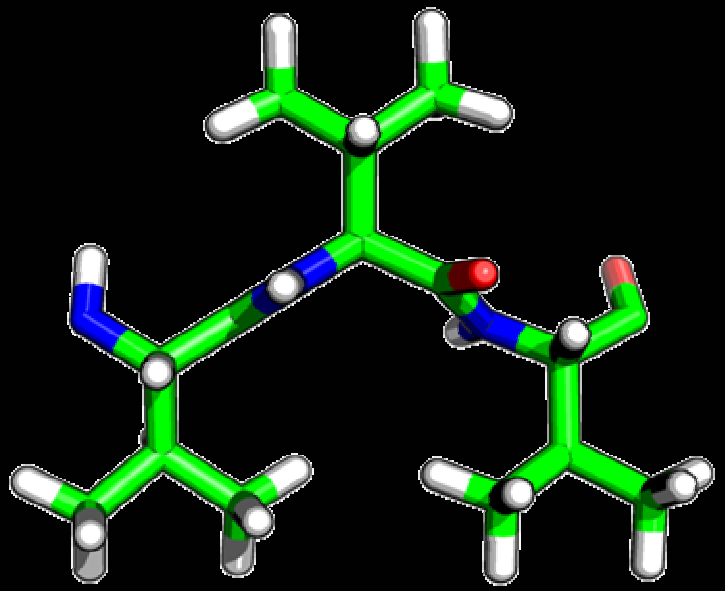
- 20 canonical amino acids (CAAs)
  - Limited to protein design
- Novel large molecule design
- Incorporate non-canonical amino acids (NCAAs) into computational protein design
- NCAAs increase number of sequences and conformations
- Tighter protein-protein interactions

NCAA: Side Chains

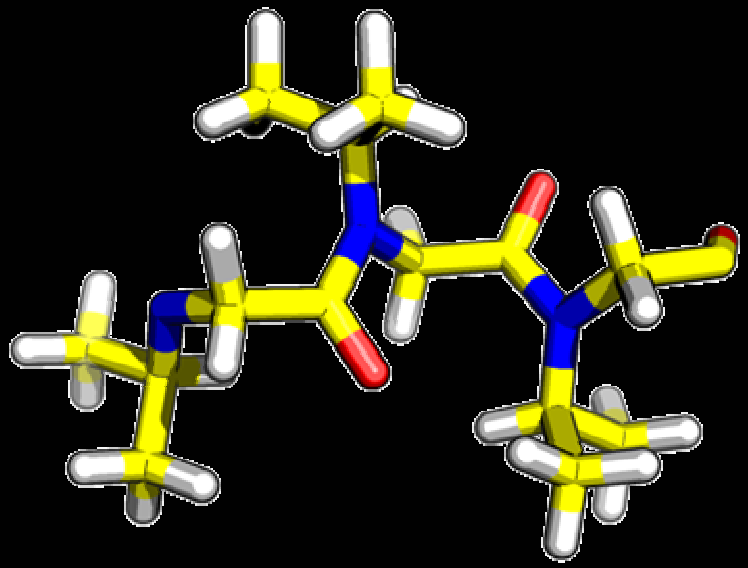


NCAA: D-enantiomer & Peptoids*

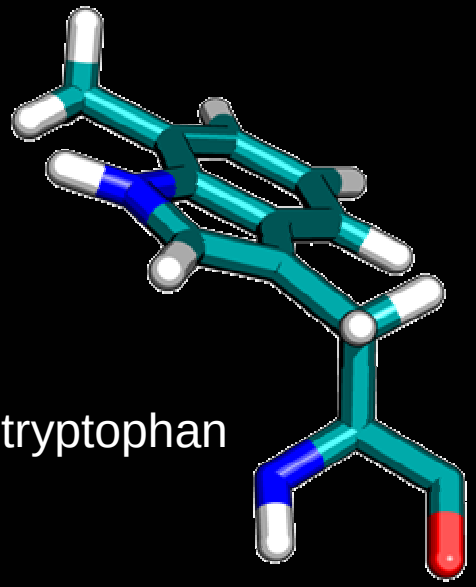
*Collaboration with Glenn Butterfoss



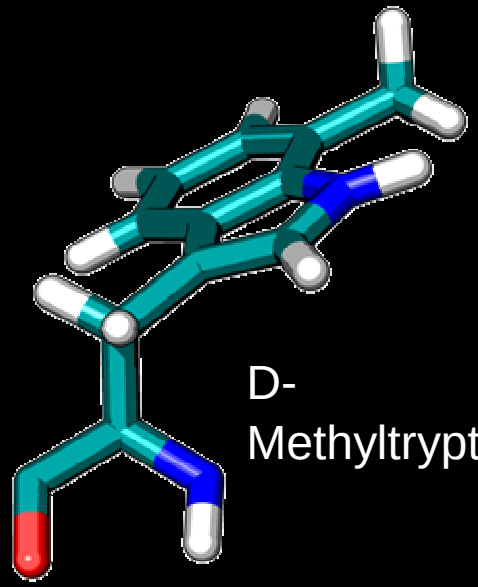
Peptide VVV



Peptoid VVV



L-
Methyltryptophan



D-
Methyltryptophan

So You Want to Add an NCAA?

- Scoring
  - Replacement for knowledge based terms
- Rotamers
  - Develop ways to generate NCAA rotamer libraries
- Reference Energies
  - Calculate from explicit model of the unfolded state

Scoring: EF Modifications

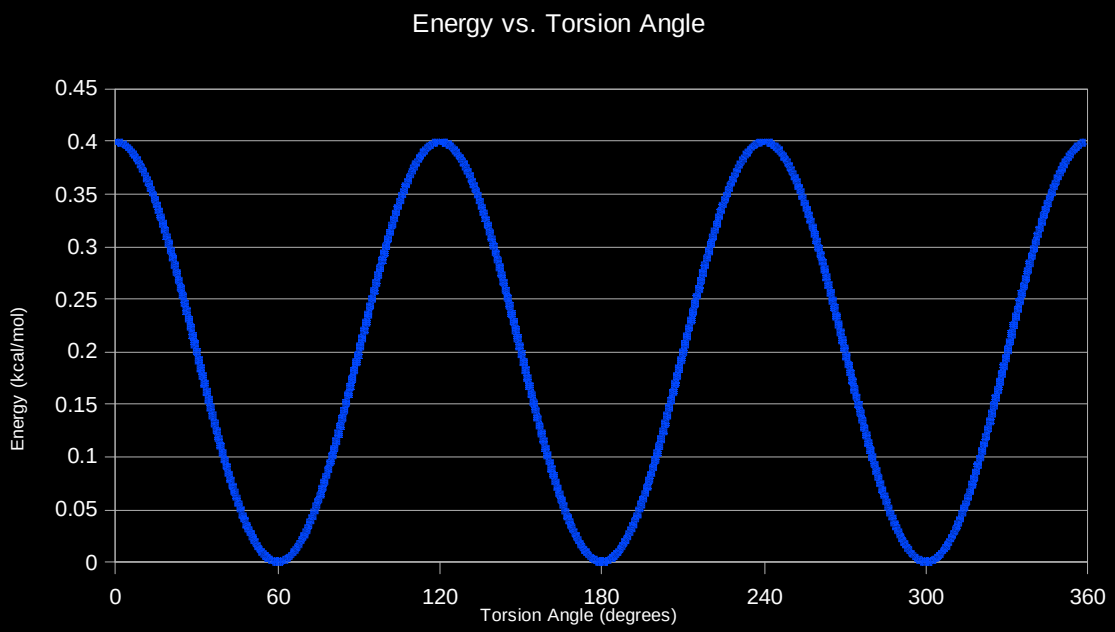
EF Term	Added	Omitted	Modified	Untouched	Basis	Reasoning
Inter residue LJ Energy				X	Physical	Evaluated on an atomistic level
Hydrogen bonding				X	Physical	Evaluated on an atomistic level
Solvation				X	Physical	Evaluated on an atomistic level
Rama Torsion Preferences		X			KB	Not compatible with NCAAs
Dunbrack Torsion Preferences		X			KB	Not compatible with NCAAs. Rotamer selection bias
Pair Electrostatics		X			KB	Not compatible with NCAAs. Possibly add a simple Coulombic term
Reference			X			Form will stay the same, just need data for NCAAs
MM Torsion Energy	X				Physical	Along with MM LJ Energy term replaces RAMA and DUN terms
MM Intra Residue LJ Energy	X				Physical	Along with MM Torsion Energy term replaces RAMA and DUN terms

Scoring: MM Torsion

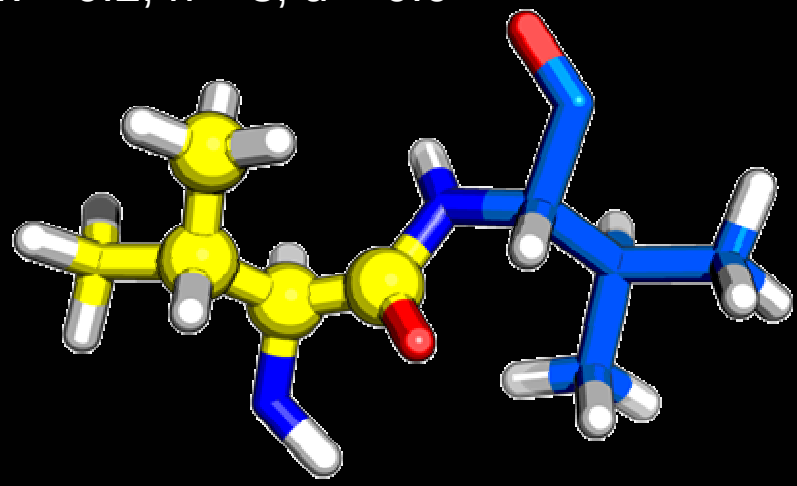
- CHARMM 27 parameters
- Computed for all sets of 4 connected atoms
- Energy associated with rotating around a bond

$$E_{tor} = k_{ijkl}\left[1 + \cos\left(n_{ijkl}\, x_{ijkl} - d_{ijkl}\right)\right]$$

i,j,k,l = atom indices type
k = constant (based on w,x,y,z)
n = multiplicity
d = minimum angle
x = torsion angle



k = 0.2; n = 3; d = 0.0

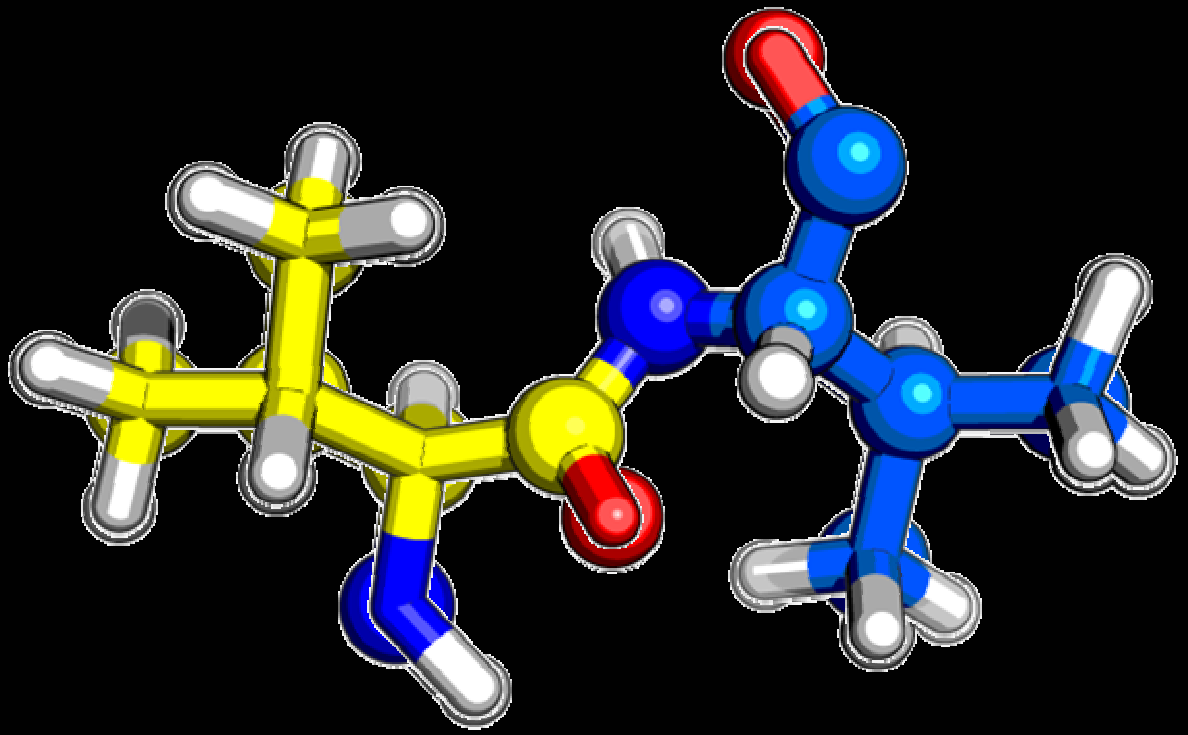


Scoring: MM Torsion

CG1	CB	CA	C
CG2	CB	CA	C
HB	CB	CA	C
CG1	CB	CA	HA
CG1	CB	CA	N
CG2	CB	CA	HA
CG2	CB	CA	N
HB	CB	CA	HA
HB	CB	CA	N
1HG1	CG1	CB	CA
1HG1	CG1	CB	CG2
1HG1	CG1	CB	HB
2HG1	CG1	CB	CA
2HG1	CG1	CB	CG2
2HG1	CG1	CB	HB
3HG1	CG1	CB	CA
3HG1	CG1	CB	CG2
3HG1	CG1	CB	HB
1HG2	CG2	CB	CA
1HG2	CG2	CB	CG1
1HG2	CG2	CB	HB
2HG2	CG2	CB	CA
2HG2	CG2	CB	CG1
2HG2	CG2	CB	HB
3HG2	CG2	CB	CA
3HG2	CG2	CB	CG1
3HG2	CG2	CB	HB
HN	N	CA	HA
HN	N	CA	C
HN	N	CA	CB
N	CA	C	O
HA	CA	C	O
CB	CA	C	O

FULL WEIGHT

FULL WEIGHT



N CA C N'	CA N C' CA'	C N' CA' HA'
HA CA C N'	CA N C' HN'	C N' CA' C'
CB CA C N'	O N C' CA'	C N' CA' CB'
	O N C' HN'	

HALF WEIGHT

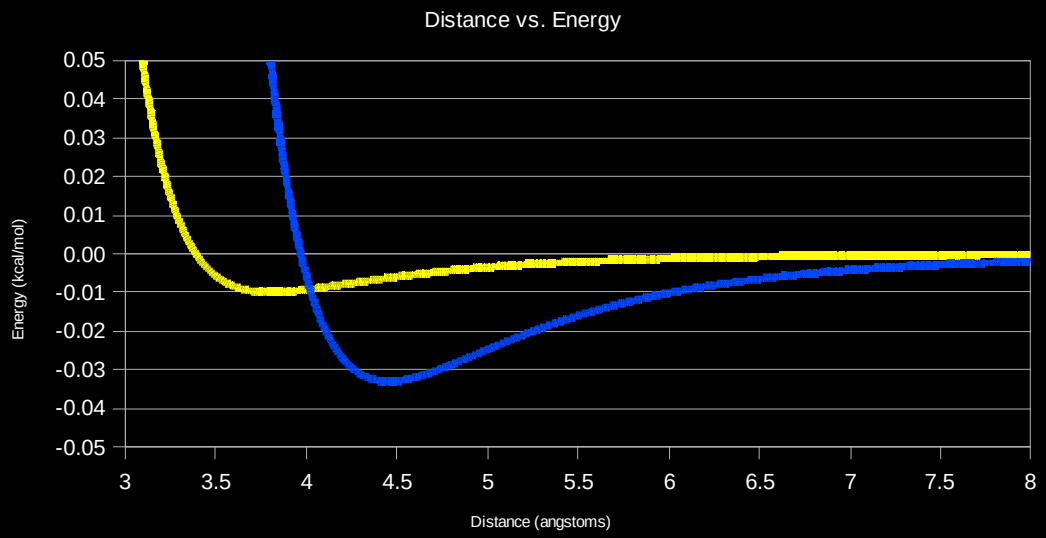
CG1'	CB'	CA'	C'
CG2'	CB'	CA'	C'
HB'	CB'	CA'	C'
CG1'	CB'	CA'	HA'
CG1'	CB'	CA'	N'
CG2'	CB'	CA'	HA'
CG2'	CB'	CA'	N'
HB'	CB'	CA'	HA'
HB'	CB'	CA'	N'
1HG1'	CG1'	CB'	CA'
1HG1'	CG1'	CB'	CG2'
1HG1'	CG1'	CB'	HB'
2HG1'	CG1'	CB'	CA'
2HG1'	CG1'	CB'	CG2'
2HG1'	CG1'	CB'	HB'
3HG1'	CG1'	CB'	CA'
3HG1'	CG1'	CB'	CG2'
3HG1'	CG1'	CB'	HB'
1HG2'	CG2'	CB'	CA'
1HG2'	CG2'	CB'	CG1'
1HG2'	CG2'	CB'	HB'
2HG2'	CG2'	CB'	CA'
2HG2'	CG2'	CB'	CG1'
2HG2'	CG2'	CB'	HB'
3HG2'	CG2'	CB'	CA'
3HG2'	CG2'	CB'	CG1'
3HG2'	CG2'	CB'	HB'
HN'	N'	CA'	HA'
HN'	N'	CA'	C'
HN'	N'	CA'	CB'
N'	CA'	C'	O'
HA'	CA'	C'	O'
CB'	CA'	C'	O'

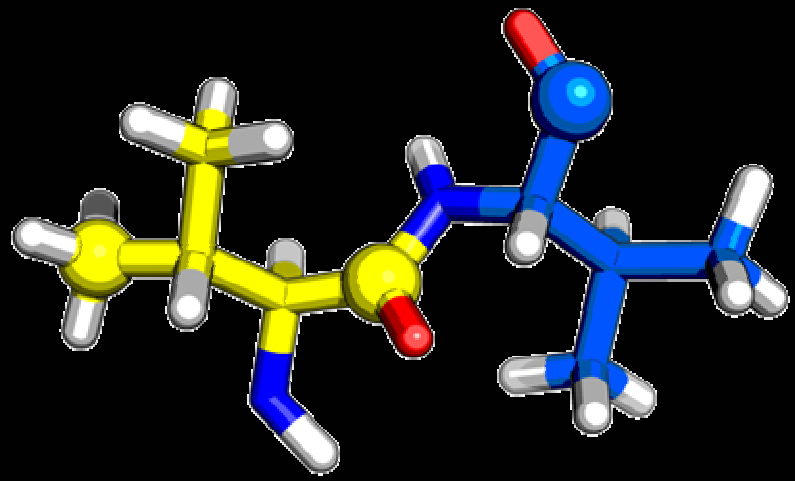
Scoring: MM Lennard-Jones

- CHARMM 27 parameters (dampened if 3 bonds)
- Attractive/Repulsive energy between atoms
- Computed for all atom pairs to 7Å (previously just inter-residue)

$$E_{LJ} = \left[\left(\frac{r_{ij}}{d_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}}{d_{ij}} \right)^{6} \right] e_{ij}$$

i,j = residue indices
d = atomic interaction distance
r = summed van der Walls radii
e = mean of atomic well depths





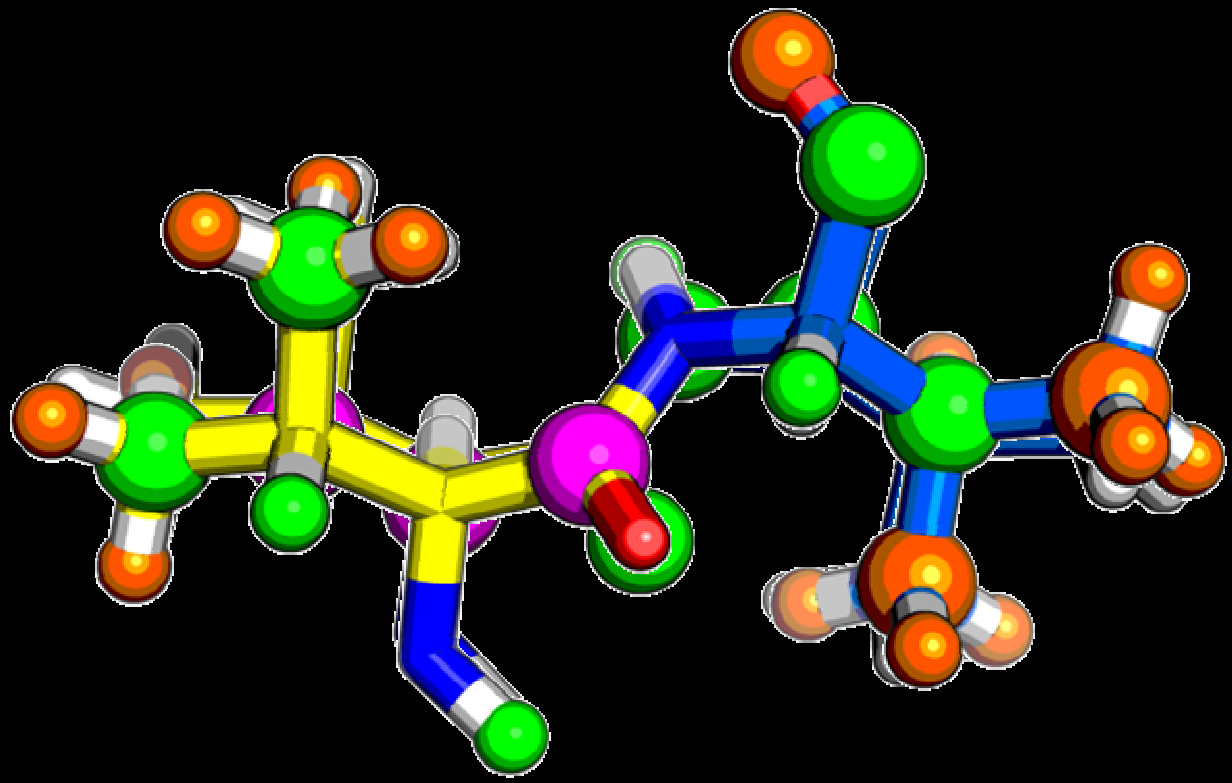
Scoring: MM LJ

- Simpler to split up in to rotamers
- Half energy for each atom pair stored in rotamer

BASE ATOM

3 BOND ATOMS

4+ BOND ATOMS



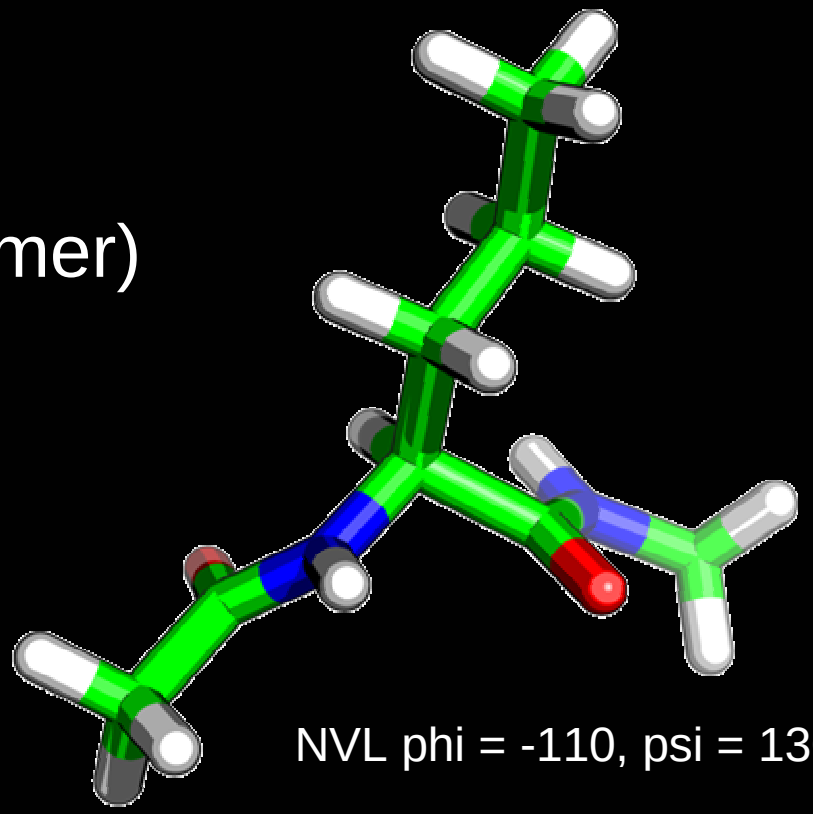
MM Term Results

- Rotamer Prediction Benchmark
  - Stock Rosetta EF: Chi1: ~87%, Chi12: ~71%
  - MM + Mod. Rosetta EF: Chi1: ~84%, Chi12: ~67%
- Sequence Recovery
  - Stock Rosetta EF: Total: 33%, Hydro 49%
  - MM + Mod. Rosetta EF: ~35% (in progress)
- Conclusions
  - Scores are not bad but could use improvement
  - Suspect improved scores with better weighting of MM terms relative to other scoring terms
  - David's suggestions

Rotamers: MakeRotLib Protocol

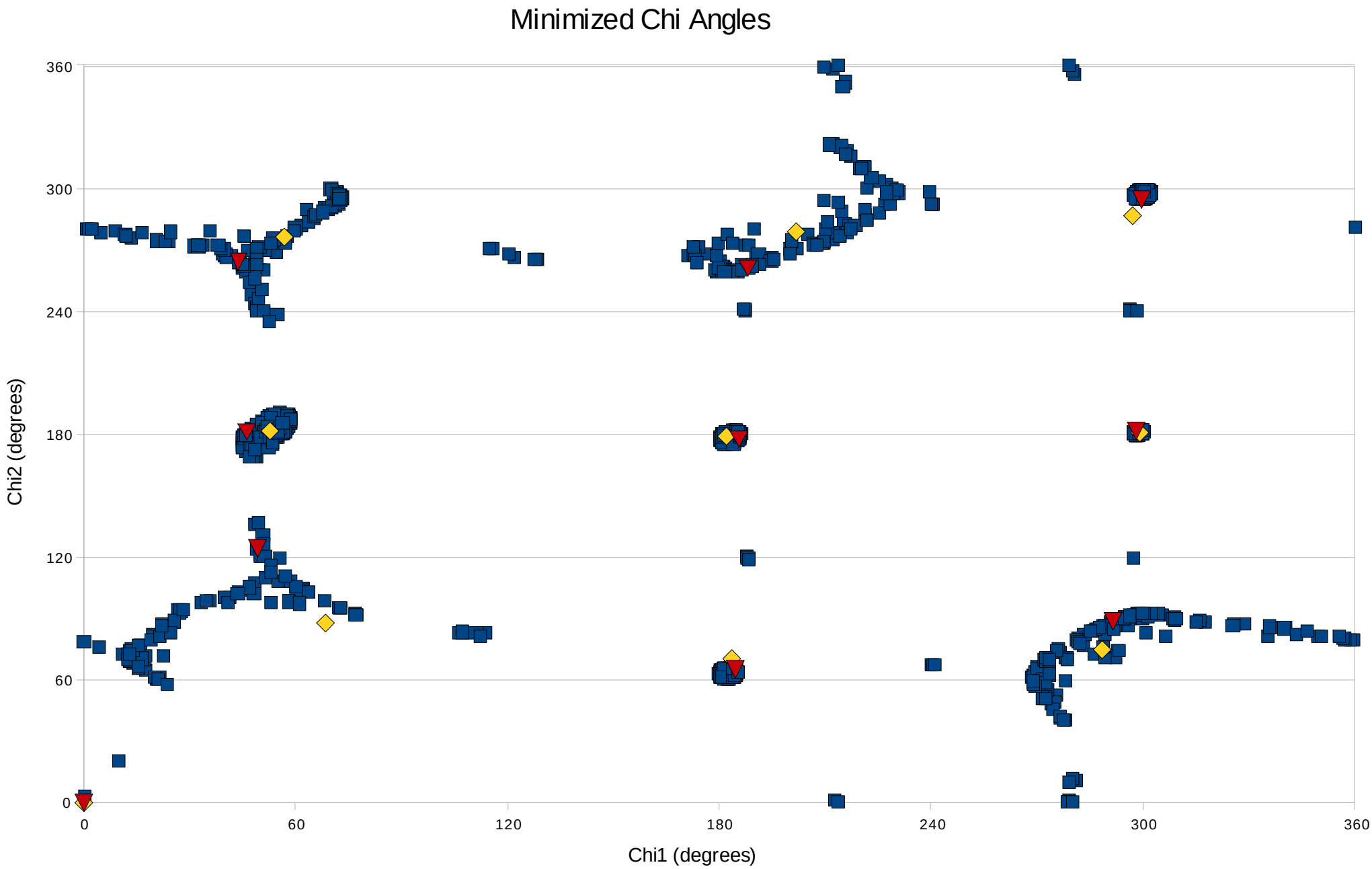
- AA Dipeptide
  - Phi, psi fixed
  - Chi(s) set
- SC minimized
  - Steepest descent
  - Local minimum (closest rotamer)
- Minimized SC clustered
  - K-Means Algorithm
- Lowest energy SC is the rotamer for that cluster

- Rot Libs are used as starting points
- Not currently used for scoring



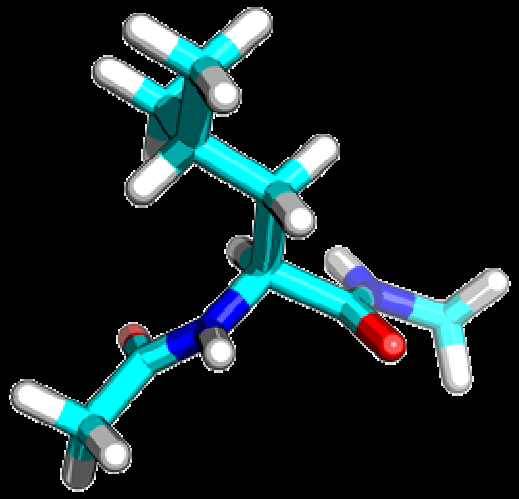
NVL phi = -110, psi = 130

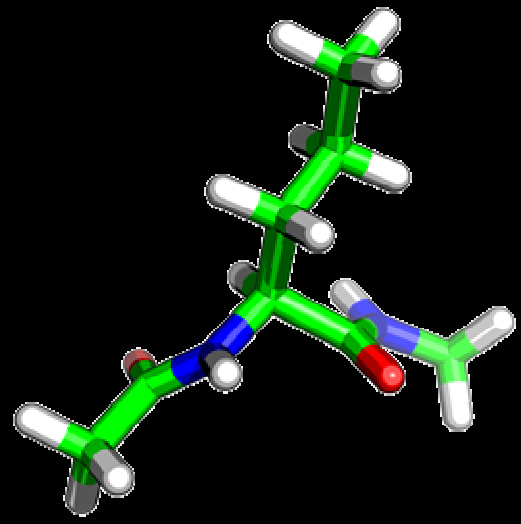
NCAA Rot Lib: Norvaline



NCAA Rot Lib: Comparison

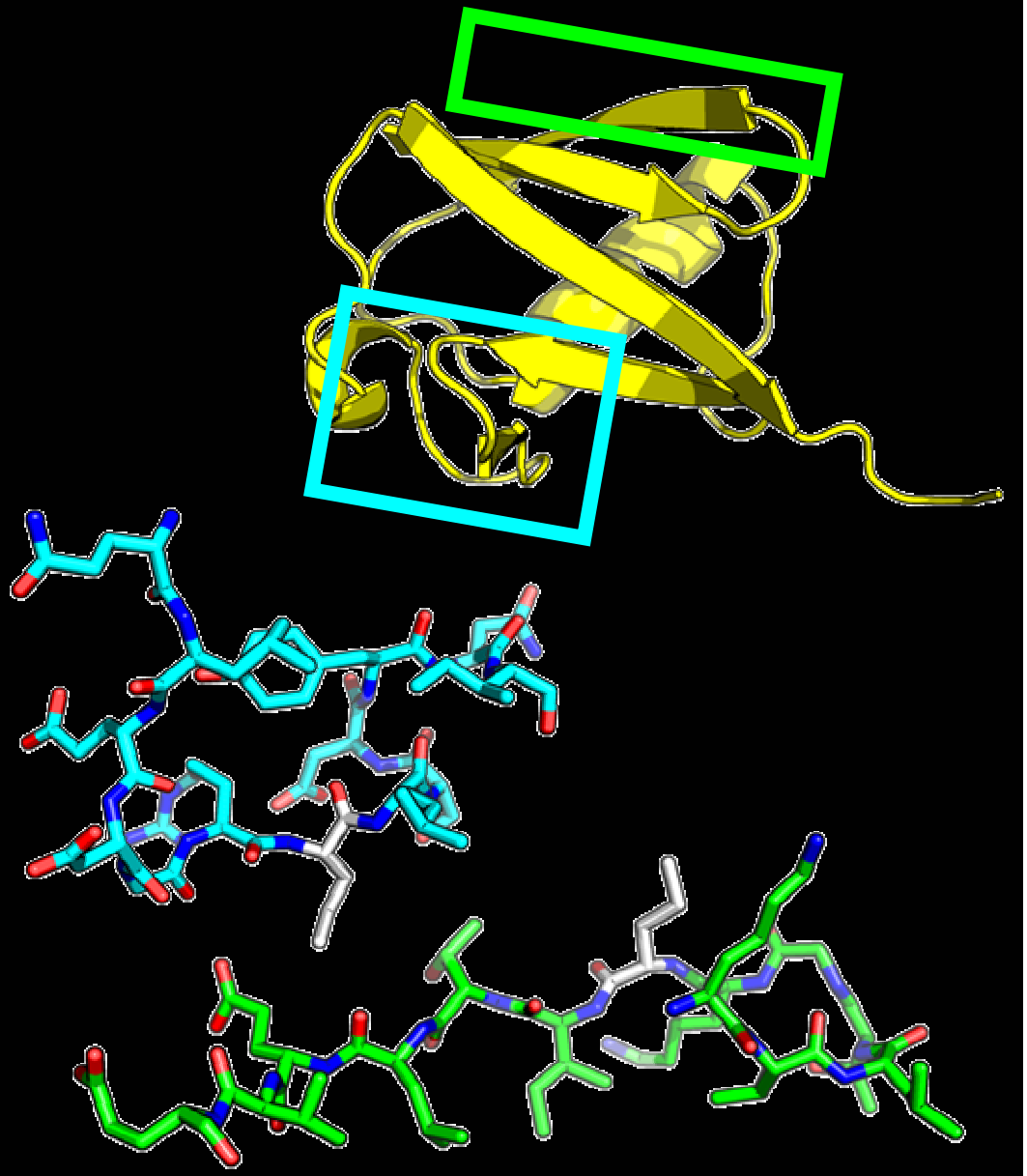
Rotamer	LEU Dunbrack			LEU MakeRotLib			NVL MakeRotLib		
	Probability	Chi1	Chi2	Probability	Chi1	Chi2	Probability	Chi1	Chi2
MT	52%	-60	177	62%	-61	169	40%	-62	-178
TP	38%	178	66	36%	186	65	10%	-175	65
MP	5%	-82	60	0%	-62	99	1%	-68	89
TT	3%	-165	177	2%	-168	139	20%	-176	177
MM	1%	-81	-54	0%	-104	-66	27%	-60	-66
TM	1%	-169	-72	0%	-117	-63	0%	-172	-99
PP	0%	59	81	0%	58	119	0%	49	124
PT	0%	68	165	0%	52	137	2%	46	-179
PM	0%	58	-74	0%	30	-77	0%	44	-96





NCAA Reference Energies

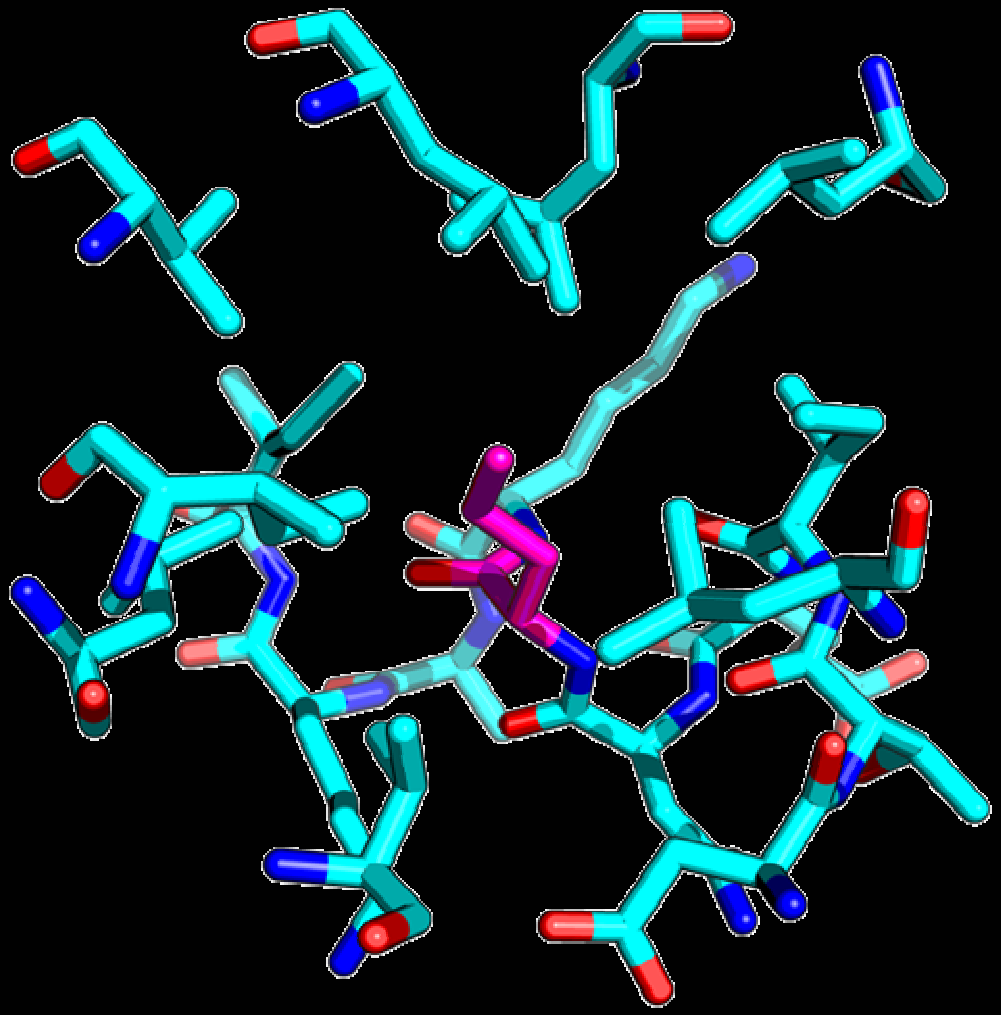
- Energy of the unfolded state
  - DOF in OptE protocol
  - random sequence structure AA energy
- Protocol
  - Randomly select 13 residue segments
  - Mutate central residue
  - Repack
  - Calc energy of central
  - Average is ref energy



Brief Design Results

- Score NCAAs
- Create Rot Libs
  - packing
- Reference Weights
  - design

- Ubiquitin, VAL26NVL

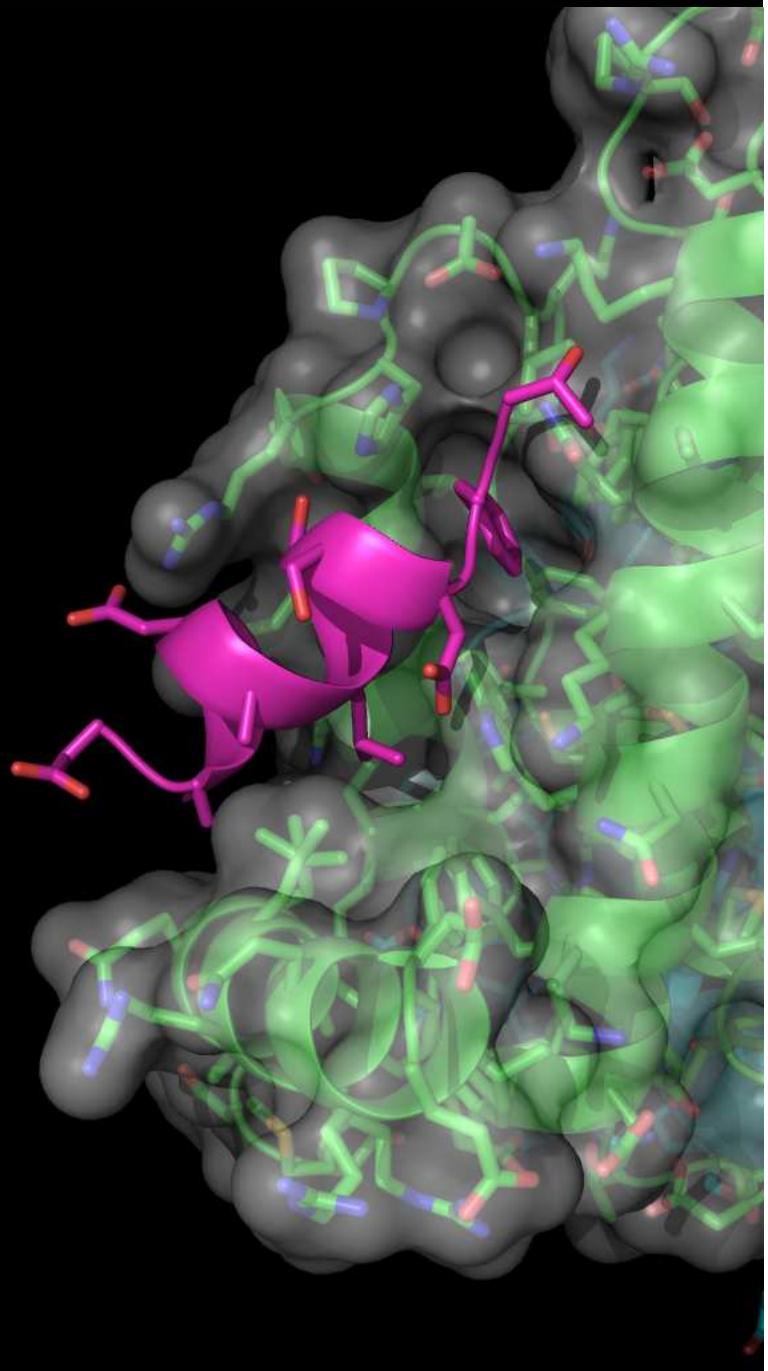


Future Directions

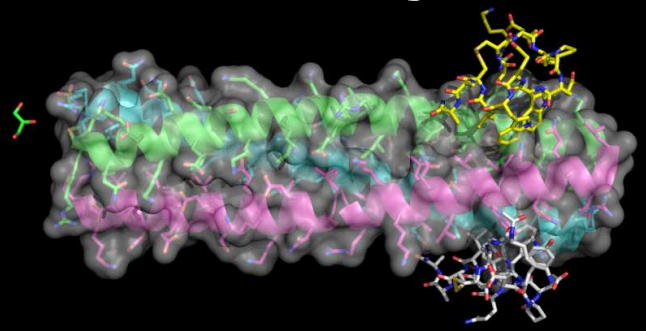
- Optimize Weights
- Try other MM potentials
  - AMBER, OPLS, MMF, etc.
  - Mini `AtomTypeSets`
- Design like crazy!!!

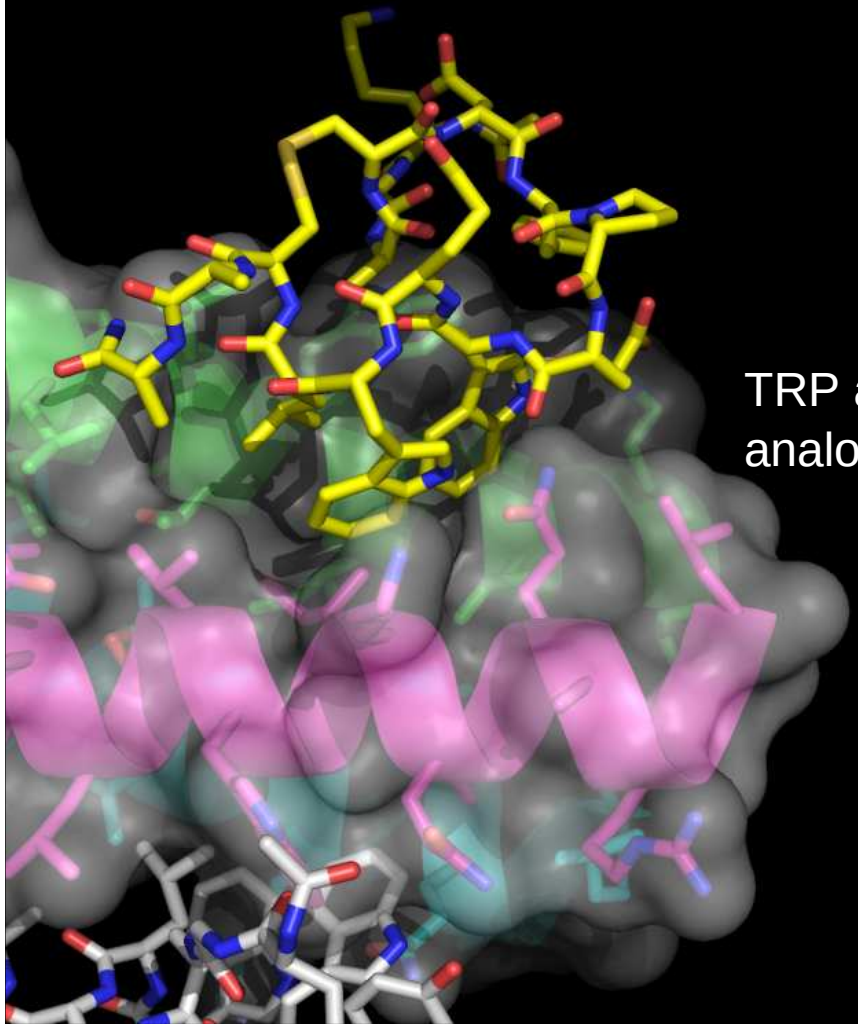
Model Systems: Calpain/Calpastatin

- Ga/GoLoco
  - GoLoco, 19AA
  - Deanne + Glenn
- Calpain/Calpastatin
  - Calpastatin, 10AA
  - Synthesized (smaller, more designs)
  - Eun Jung
    - Success with FP binding assays
- Redesigns NCAA side chains



Model Systems: gp41/PIE





TRP and PHE analogs

- Michael Kay @ U. Utah
- Directed evolution with D-peptides
- 70 pM Kd, 250 pM IC_{50}
- Explore D-AAs, NCAA side chains and length

Acknowledgments

- Brian Kuhlman
- Kuhlman Lab (UNC)
  - Ben Stranges
  - Glenn Butterfoss
  - Andrew Leaver-Fay
- Kay Lab (Utah)
  - Michael Kay
- MiniRosetta Team

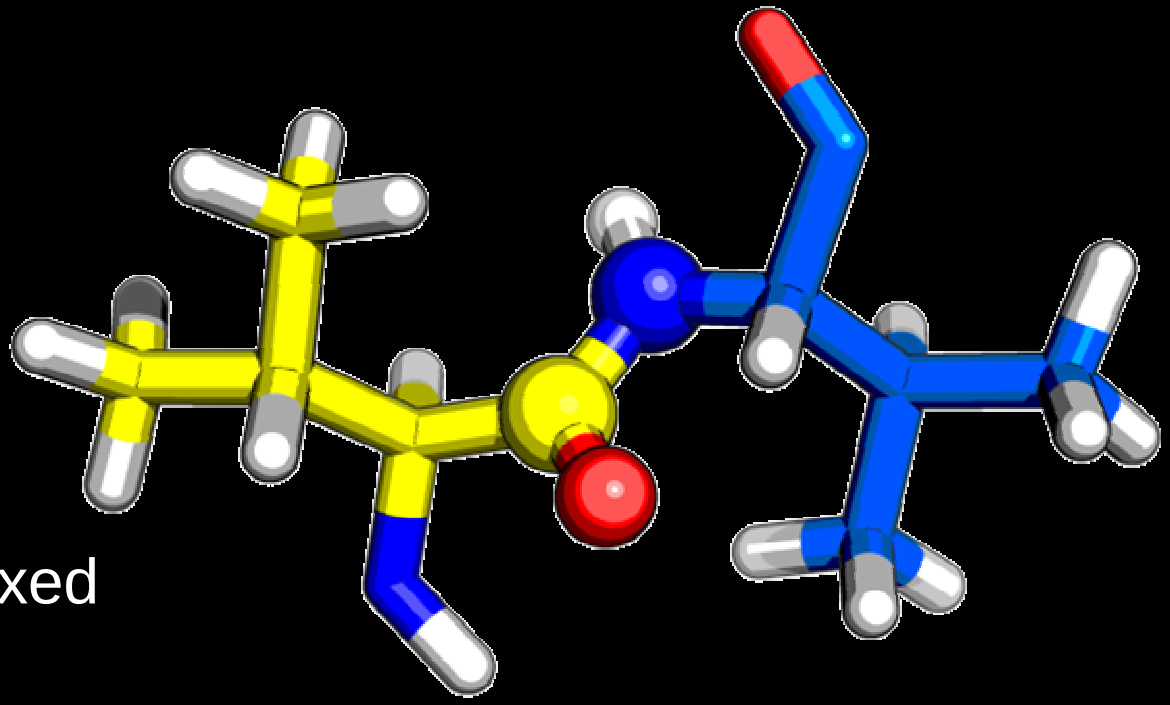
"Top7 is easy – I want to see a Greek Key!"

-BK

Questions?

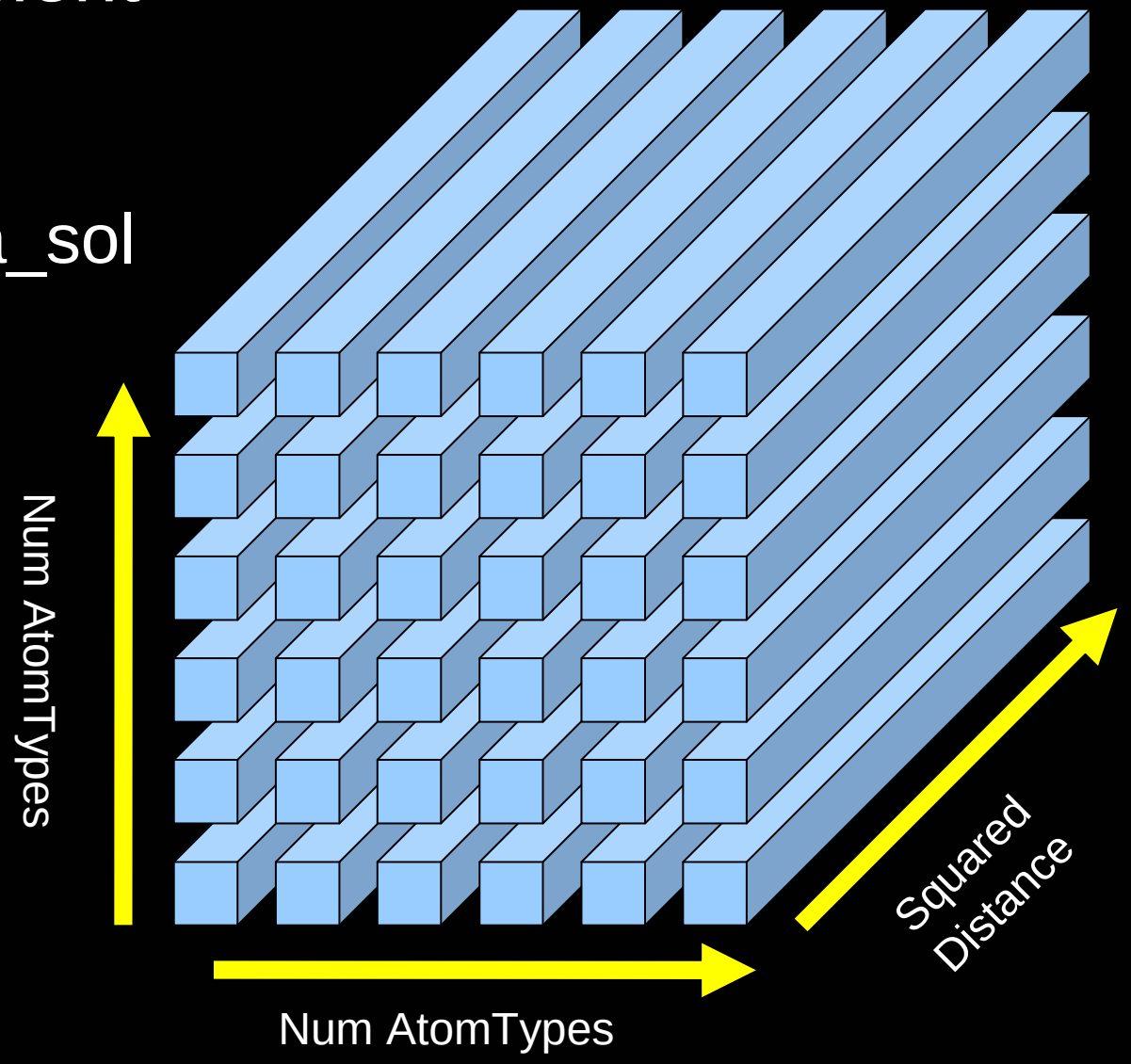
Scoring: MM Term Implementation

- Rosetta vs. MM
  - Basic Unit
    - Rosetta: rotamer
    - MM: DOF
  - Sequence
    - Rosetta: changes
    - MM: (generally) fixed
- Complications
  - Where energies stored
  - Torsion term is 2-body



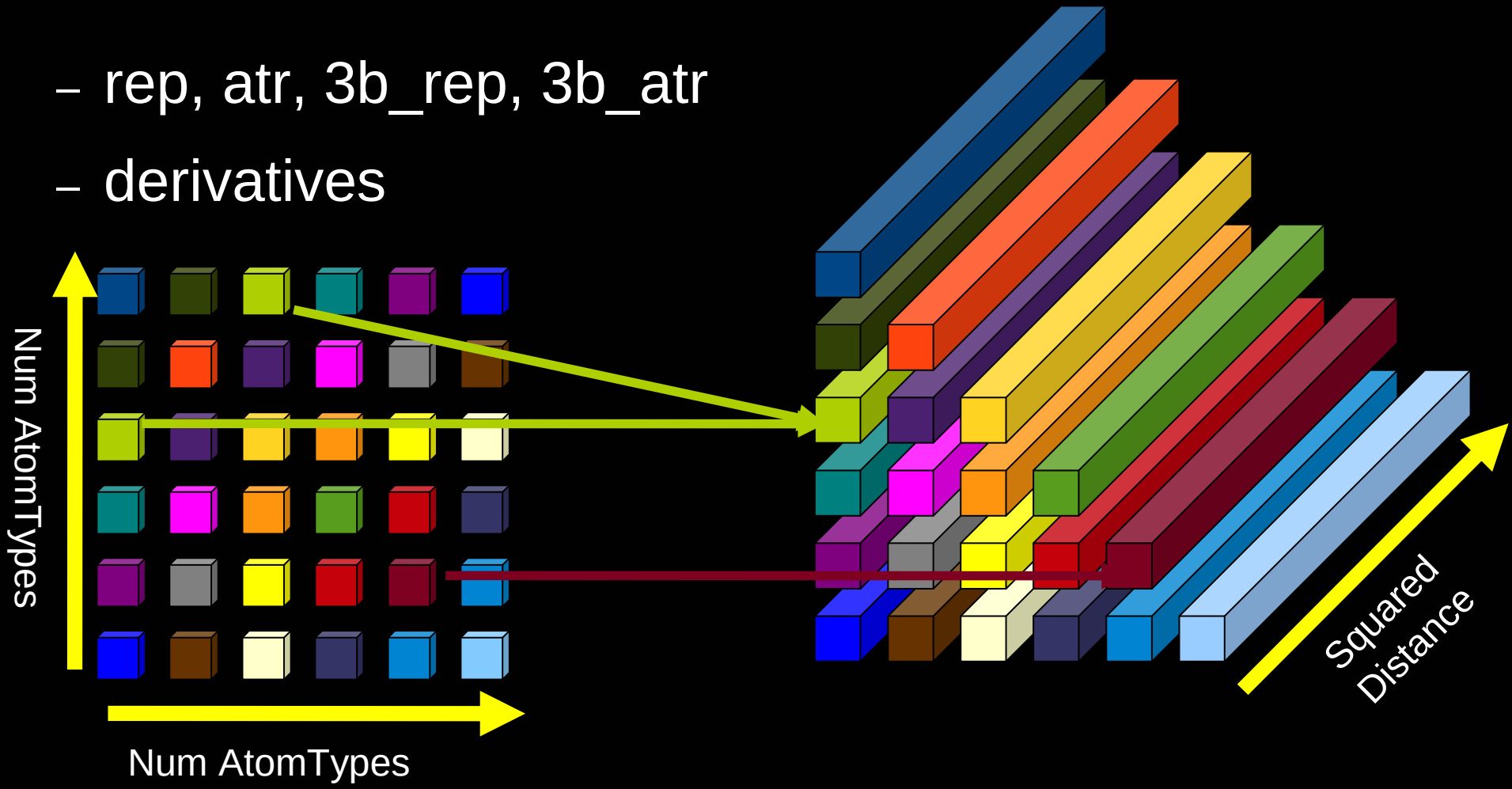
MM LJ Optimizations

- Currently ETable is `Farray3D<Reals>`
- Not memory efficient
- Six tables
  - fa_rep, fa_atr, fa_sol
  - derivatives



MM LJ Optimizations

- MM LJ Precomputed Energy Table
- Memory Efficient, 2D Table of pointer to vectors
- 8 Tables
  - rep, atr, 3b_rep, 3b_atr
  - derivatives



NCAA Rot Lib: Leucine