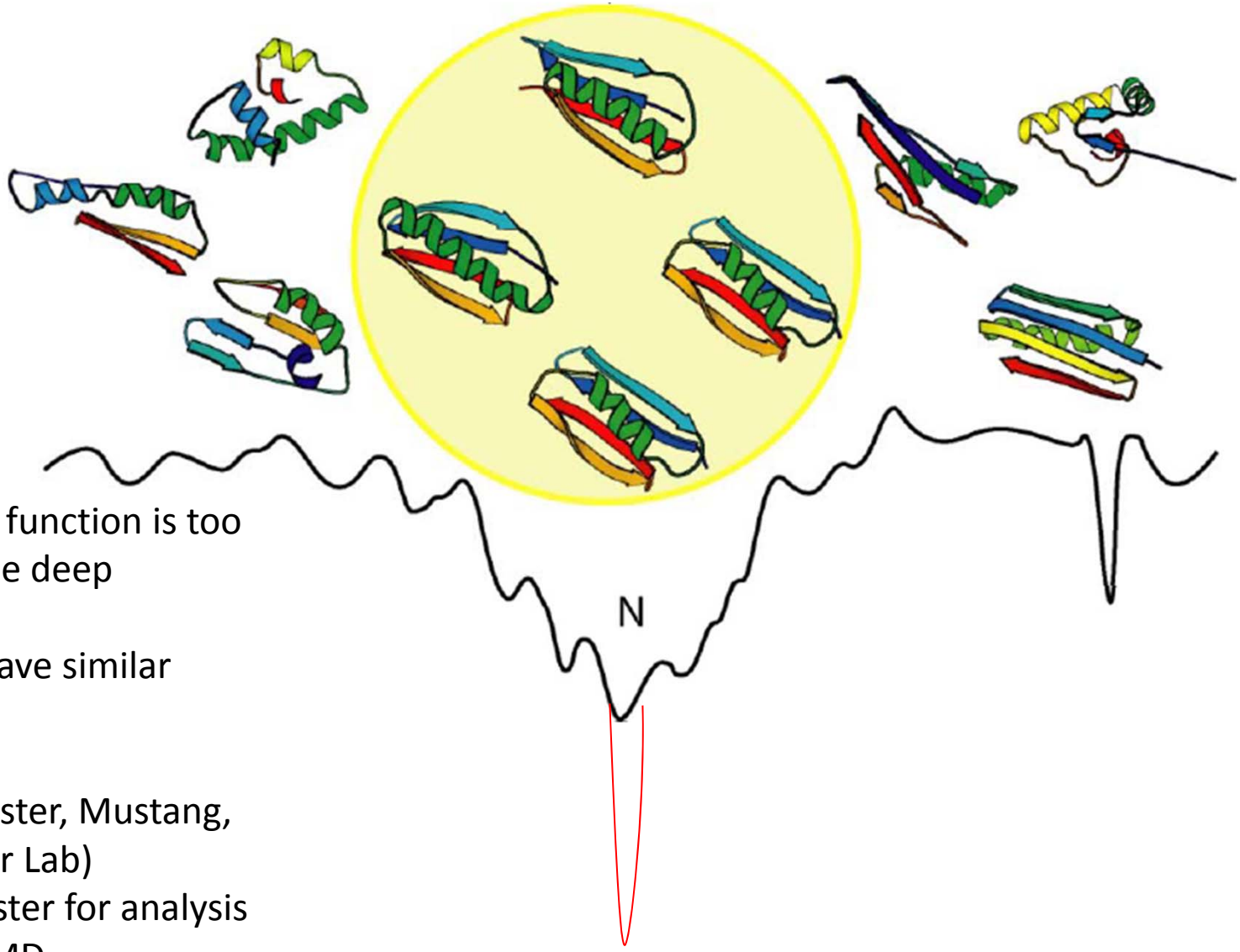
Is this Model Good?

How to pick the model for further experiments

The Problem

- You have 100,000 models, which one is the best?
- Take top 5 by energy?
  - Energy units only 0.20 difference
  - What now?

Cluster Models

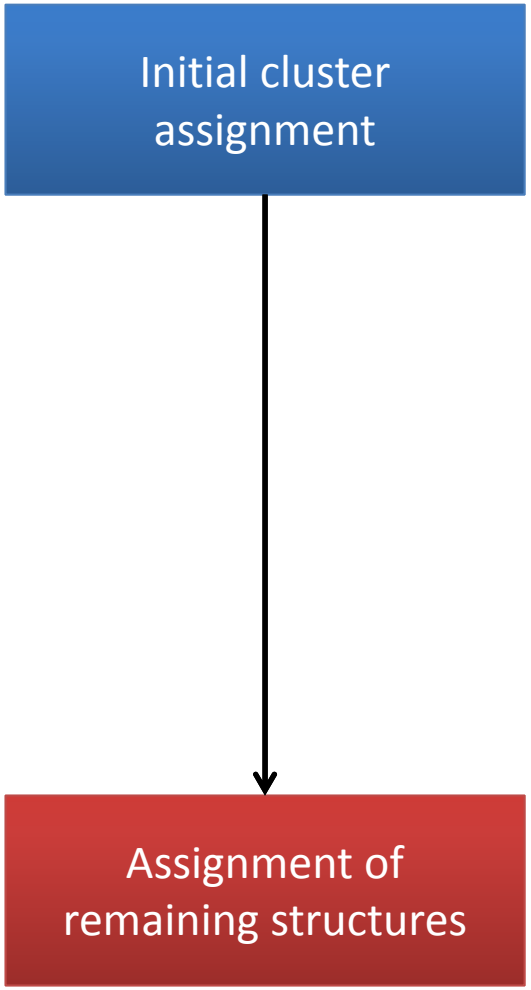


- Rosetta Scoring function is too coarse to sample deep minimum
- Many models have similar fold/dock
- Cluster models
  - RosettaCluster, Mustang, BCL (Meiler Lab)
- Take largest cluster for analysis
- Pop them into MD

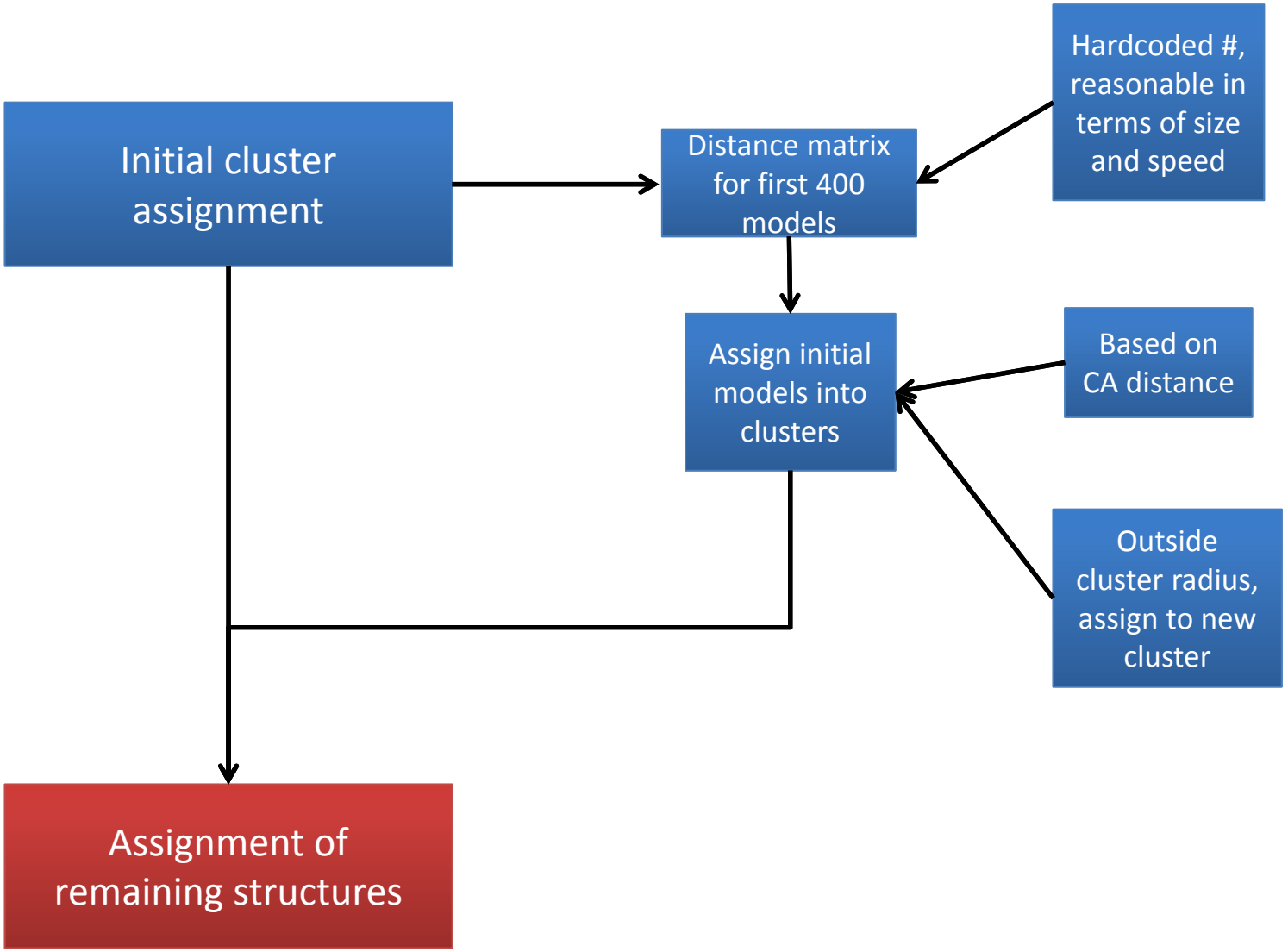
Clustering in Rosetta

- The Rosetta clustering algorithm is slightly unconventional
- Traditional clustering methods require the calculation of a pairwise distance matrix
  - The memory requirements of this method are n^2 where n is the number of models being clustered
  - For large numbers of models, these methods are therefore impractical

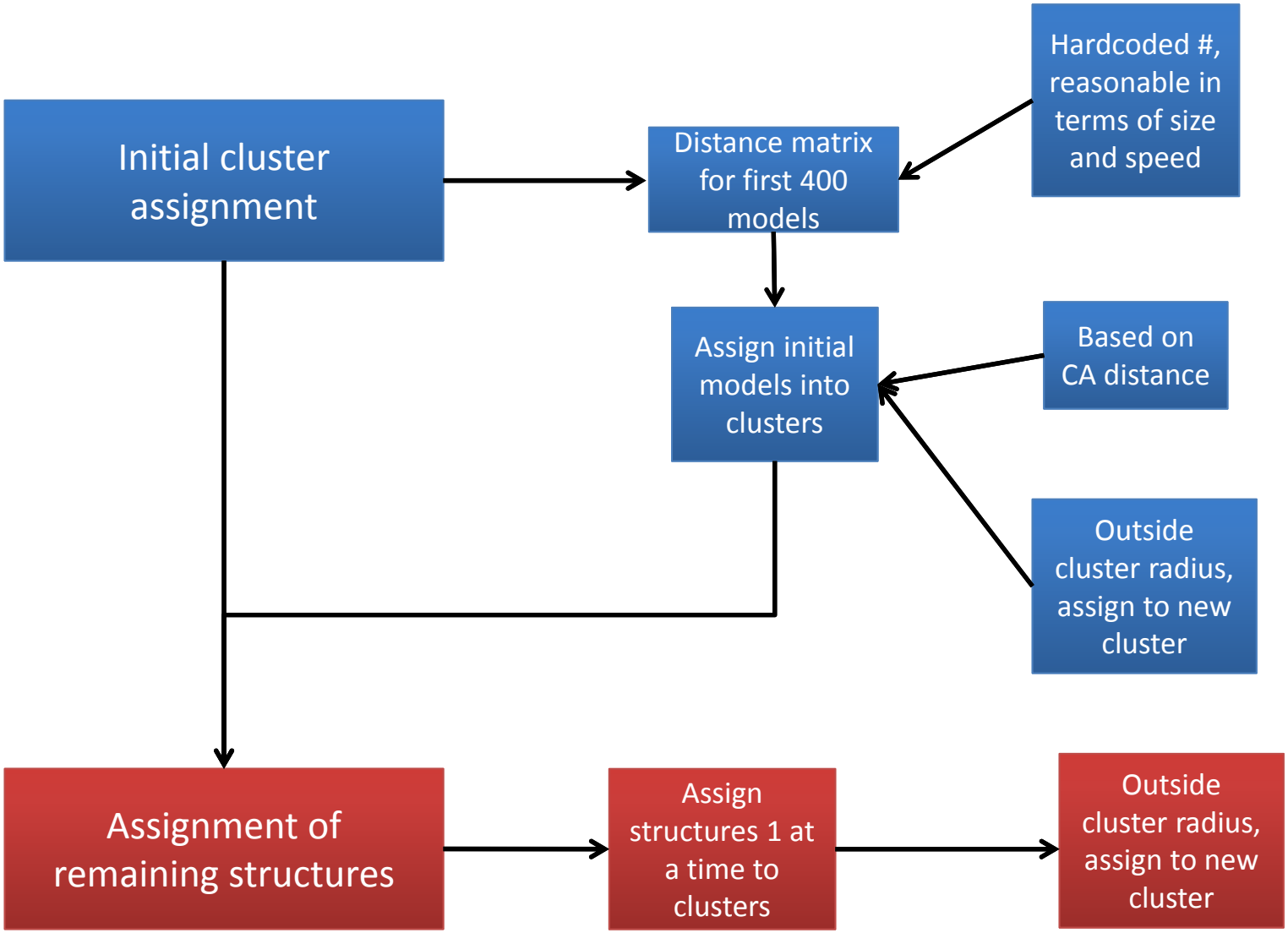
Clustering In Rosetta



Clustering In Rosetta



Clustering In Rosetta

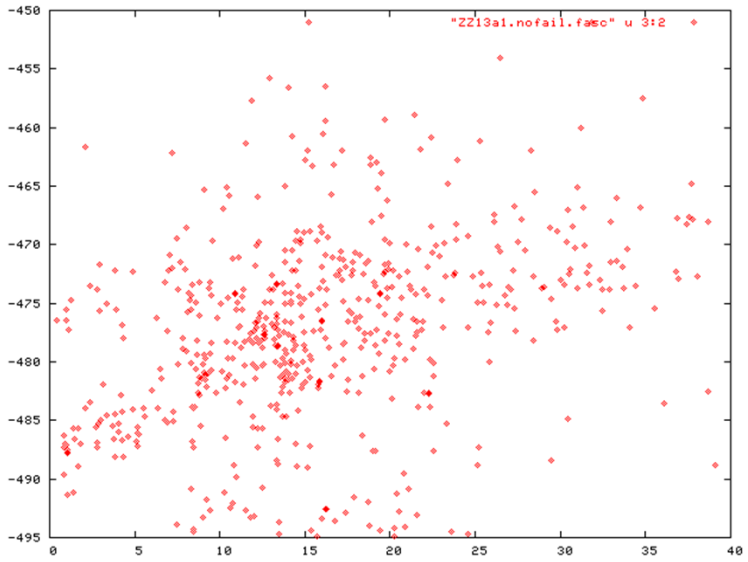


Score VS RMSD

- Mainly used for benchmarking
- Answers these questions:
  - Does score function capture correct model?
  - How much variance do you have to original structure?
  - Can show different population states

Score VS RMSD Plots

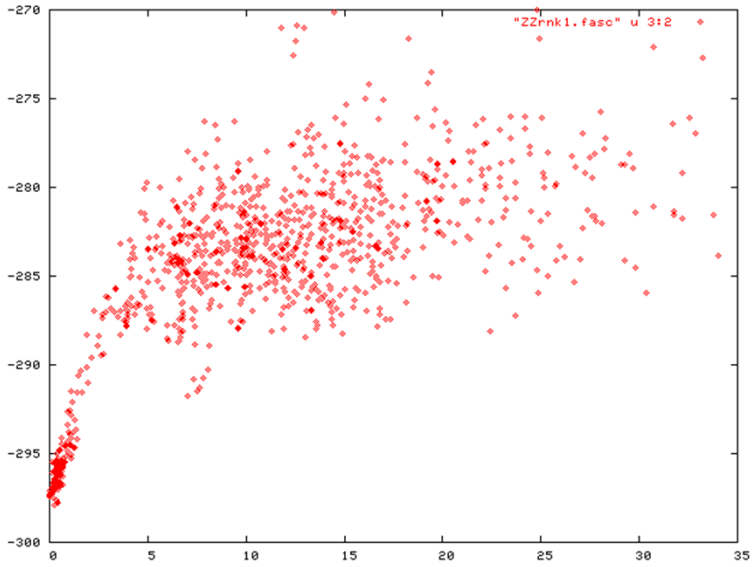
Energy



RMSD

Bad

Energy



RMSD

Good

Pose Metrics

- Calculate several metrics of proteins
  - Salt Bridges
  - Cation Pi
  - Number of Hbonds
  - Number of unsatisfied buried Hbonds
  - Packing of protein
  - SASA
  - Etc
- Lives in:

`src/protocols/toolbox/PoseMetricCalculators/`

Rosetta Holes

- Folding has known problem
  - Voids within protein
- RosettaHoles finds voids
- Outputs 4 scores

Score	Range	Interpretation
Decoys	-5 to 5	Discriminatory, low if native like
Packstat	0 to 1	How well packed a protein is. Parameterized to high resolution structures
Resolution	0 to 4	Correlates with X-ray Resolution
Holes	0 to 7	Resolution score + 3*packstats

Rosetta Holes Algorithm

- See RosettaHoles

Visualize Holes

Pose Metrics