

# Comparative Modeling of Protein Structures

James Thompson  
Baker Lab

# Topics

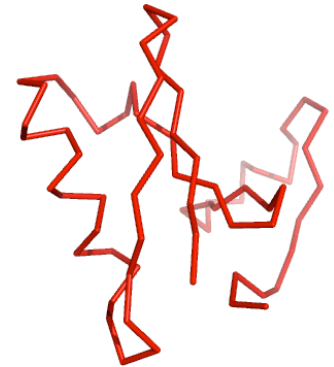
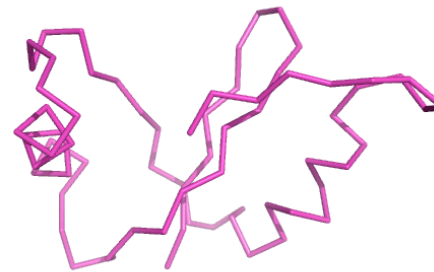
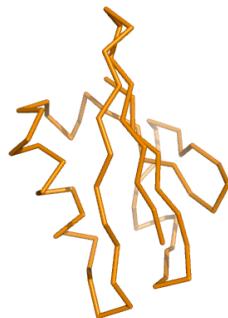
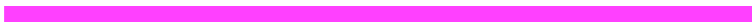
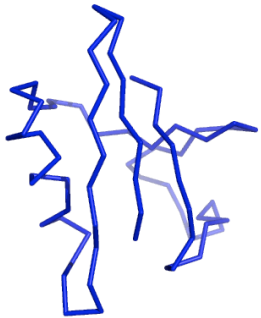
- Comparative Modeling
- Deriving Constraints
- Scoring Proteins
- FoldConstraints

# Comparative Modeling

- Protein structure is determined by its amino acid sequence, and similar sequences have similar structures.
- Comparative modeling uses information from proteins of known structure to build models of proteins with similar structure.

# Comparative Modeling

Query Sequence



Known Structures

# Sequence Alignment

Query (unknown structure)

LISFATALLKGEDITCFALD



LTAAFATKLAEHGISANVIA

Template 1zvp\_A

# Sequence Alignment

Query (unknown structure)

LISFATALLKGEDITCFALD



LTAAFATKLAEHGISANVIA

Template 1zvp\_A



# Sequence Alignment

Query (unknown structure)

LISFATALLKGEDITCFALD



LTAAFATKLAEHGISANVIA

Template 1zvp\_A



# Sequence Alignment

Query (unknown structure)

LISFATALLKGEDITCFALD



LTAAFATKLAEHGISANVIA

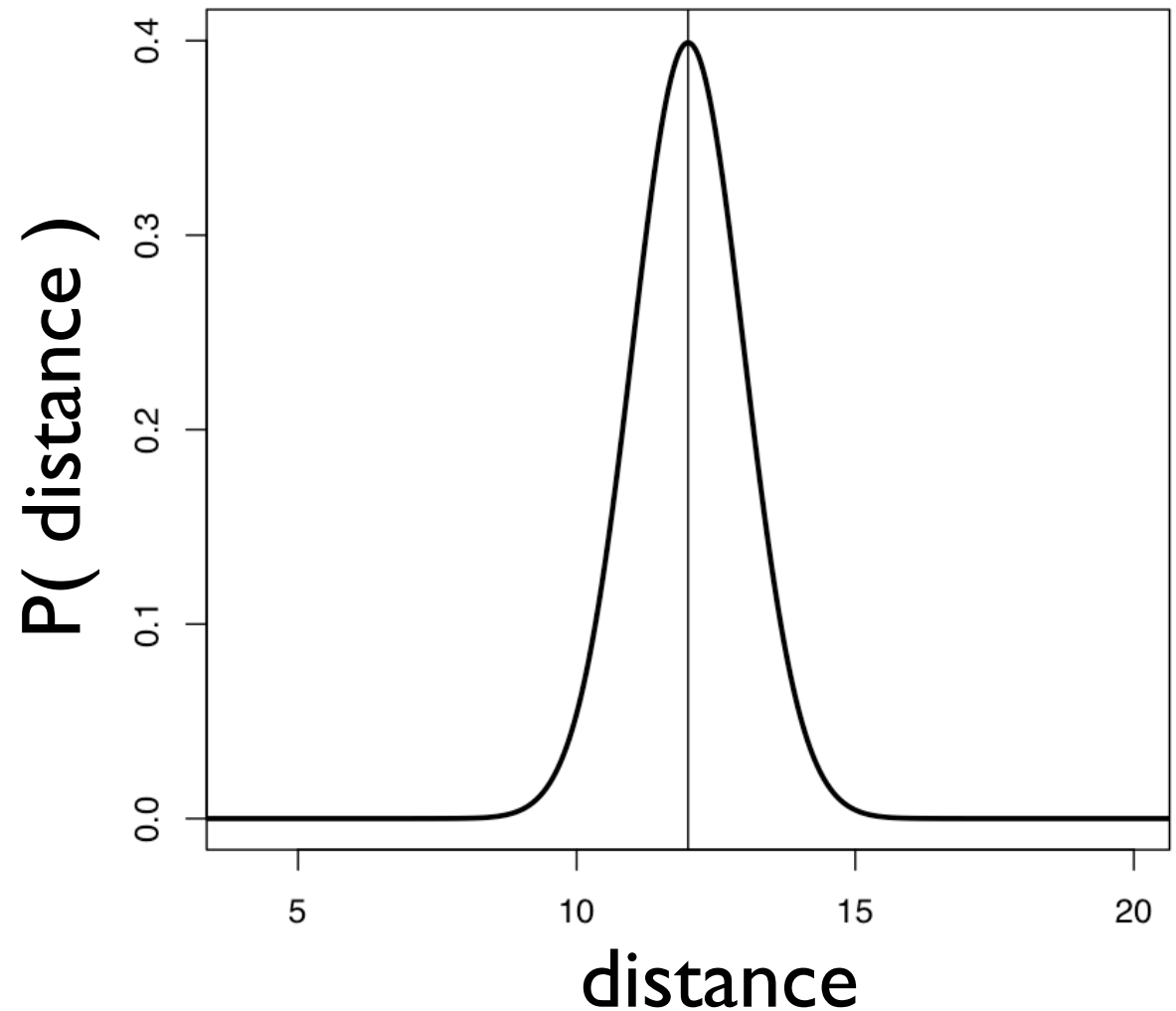
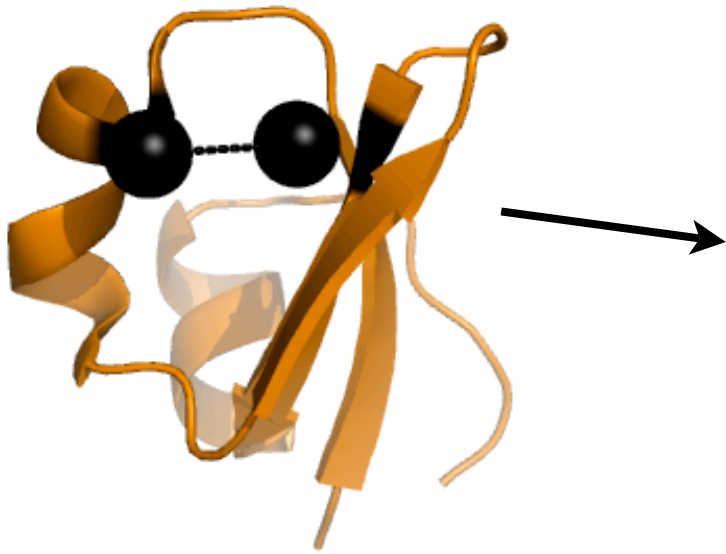
Template 1zvp\_A



What does the distance in the template structure tell us about the distance in the unknown structure?



# Inferring Probabilities



$P(\text{distance} \mid \text{template, alignment})$

# Alignment Variables

Query Sequence



Template Sequence

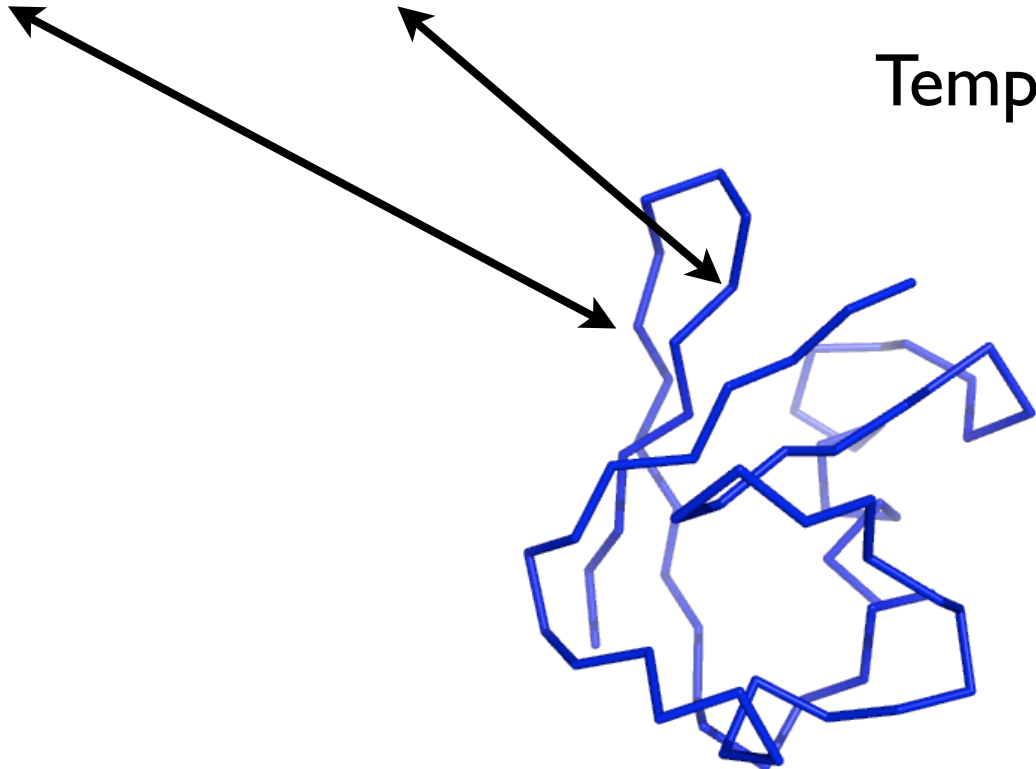
- Alignment e-value
- BLOSUM score (5-residue window)
- Average distance to alignment gap

# Structural Variables

Query Sequence

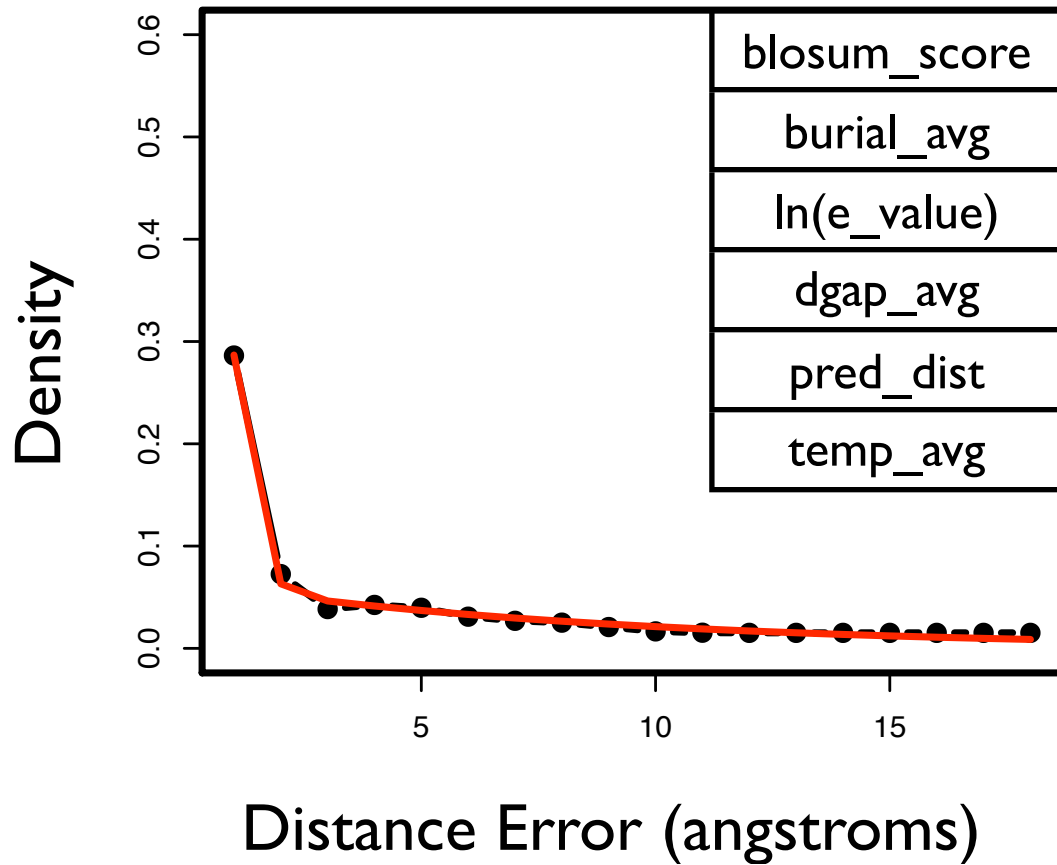


Template Structure



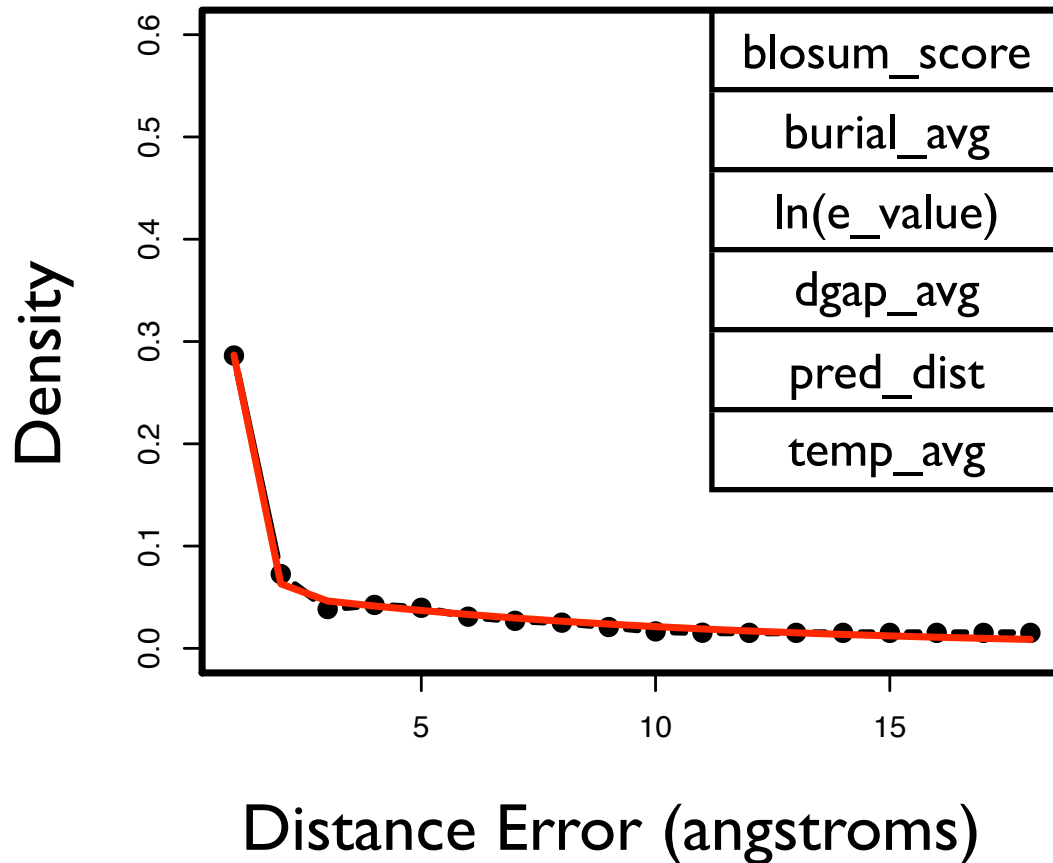
- Average Burial
- Distance

# Error Distributions

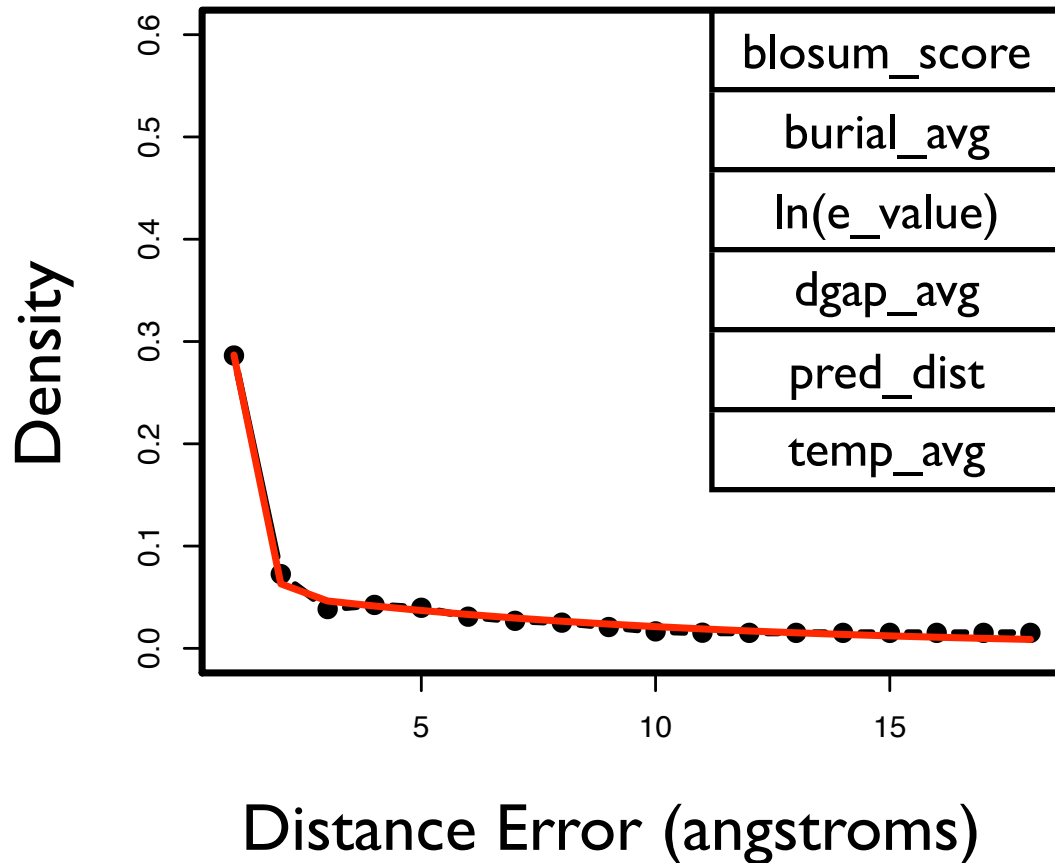


# Error Distributions

- Different distributions depending on predictor variables.

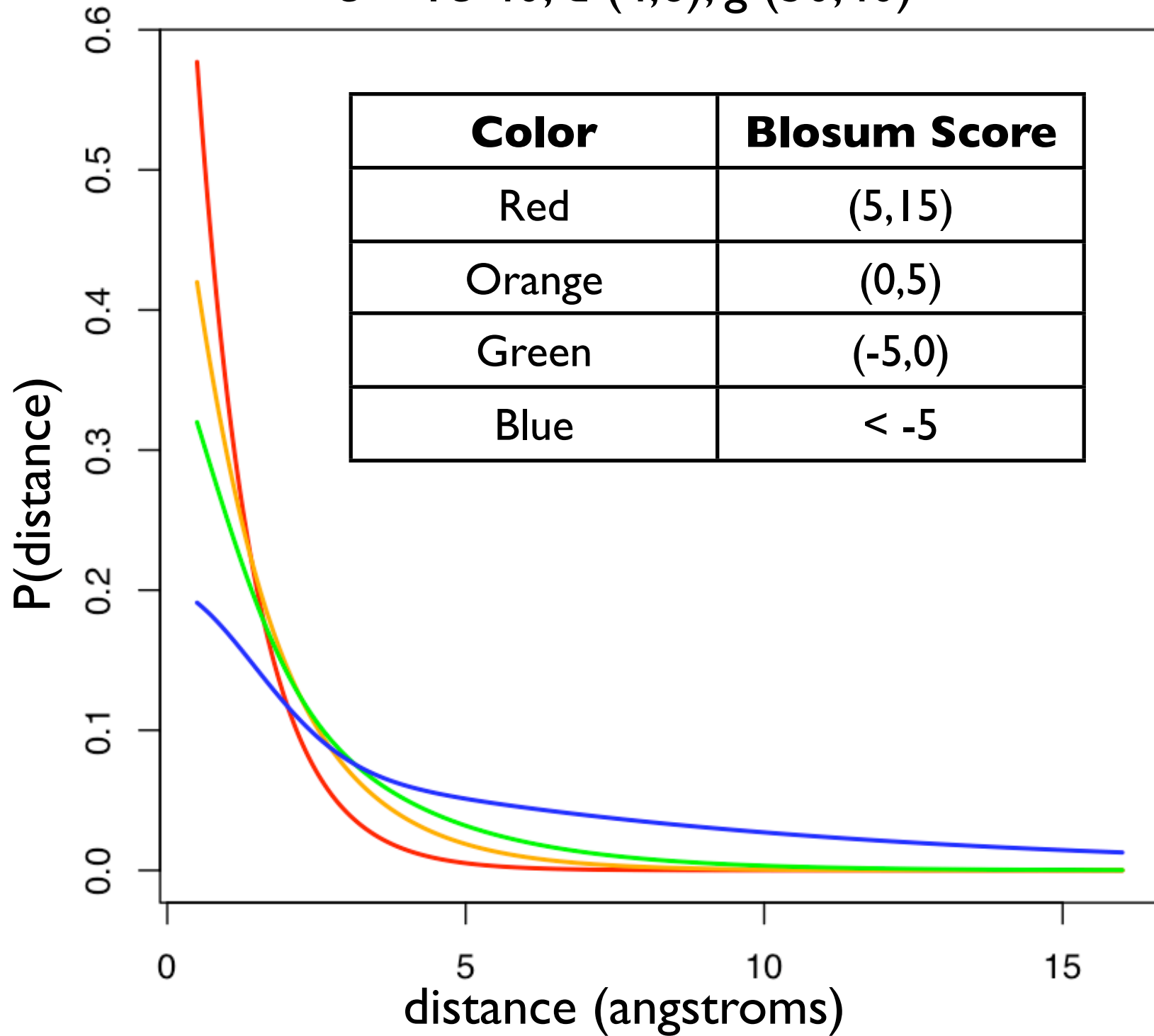


# Error Distributions



- Different distributions depending on predictor variables.
- We can derive pair potentials from a target-template alignment using these distributions.

$e < 1e-40$ , d (4,6), g (30,40)



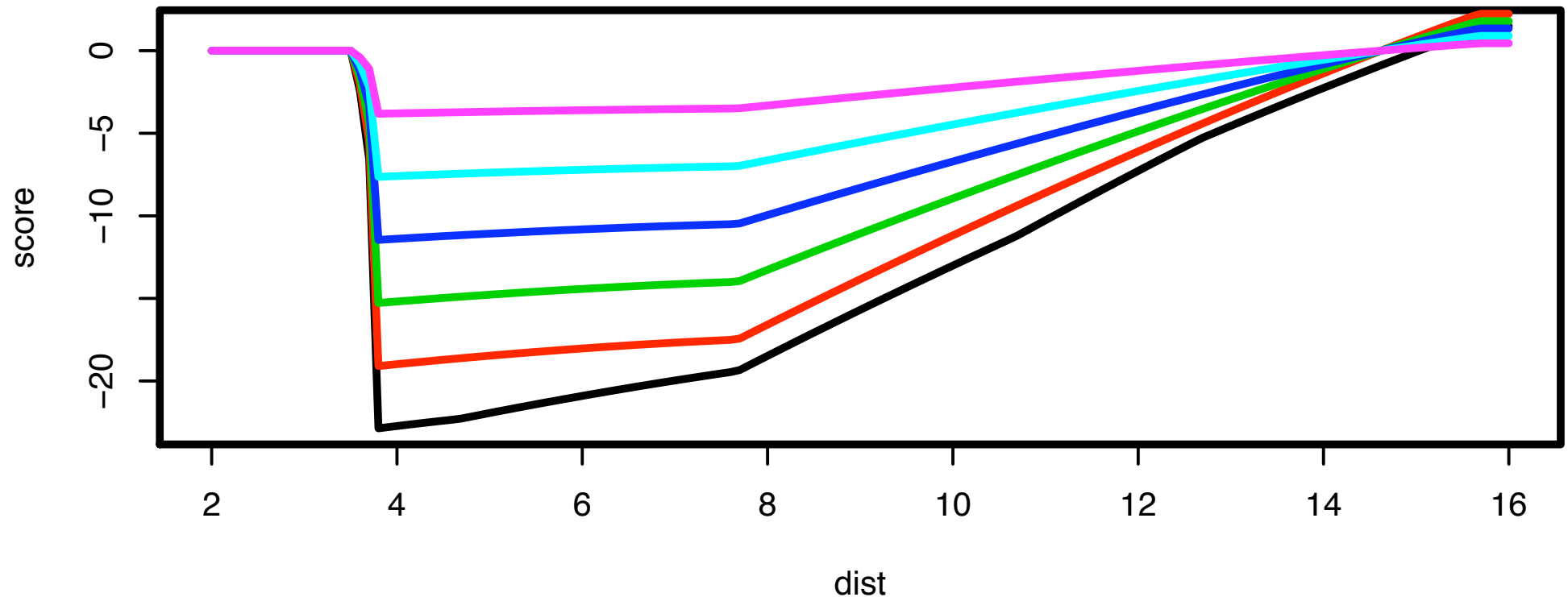
# Combing Predictions from Multiple Templates

Distance predictions from different templates on the same pairs of atoms are simply added together.



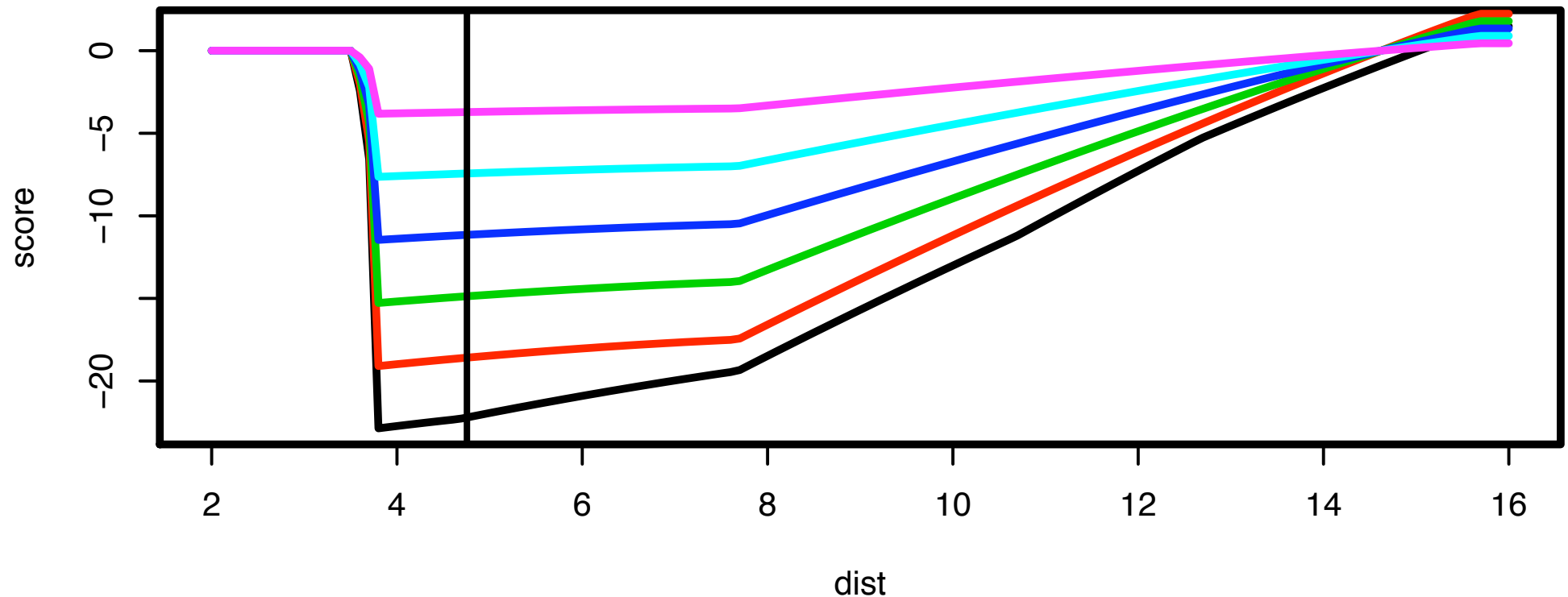
# Combing Predictions from Multiple Templates

Distance predictions from different templates on the same pairs of atoms are simply added together.



# Combining Predictions from Multiple Templates

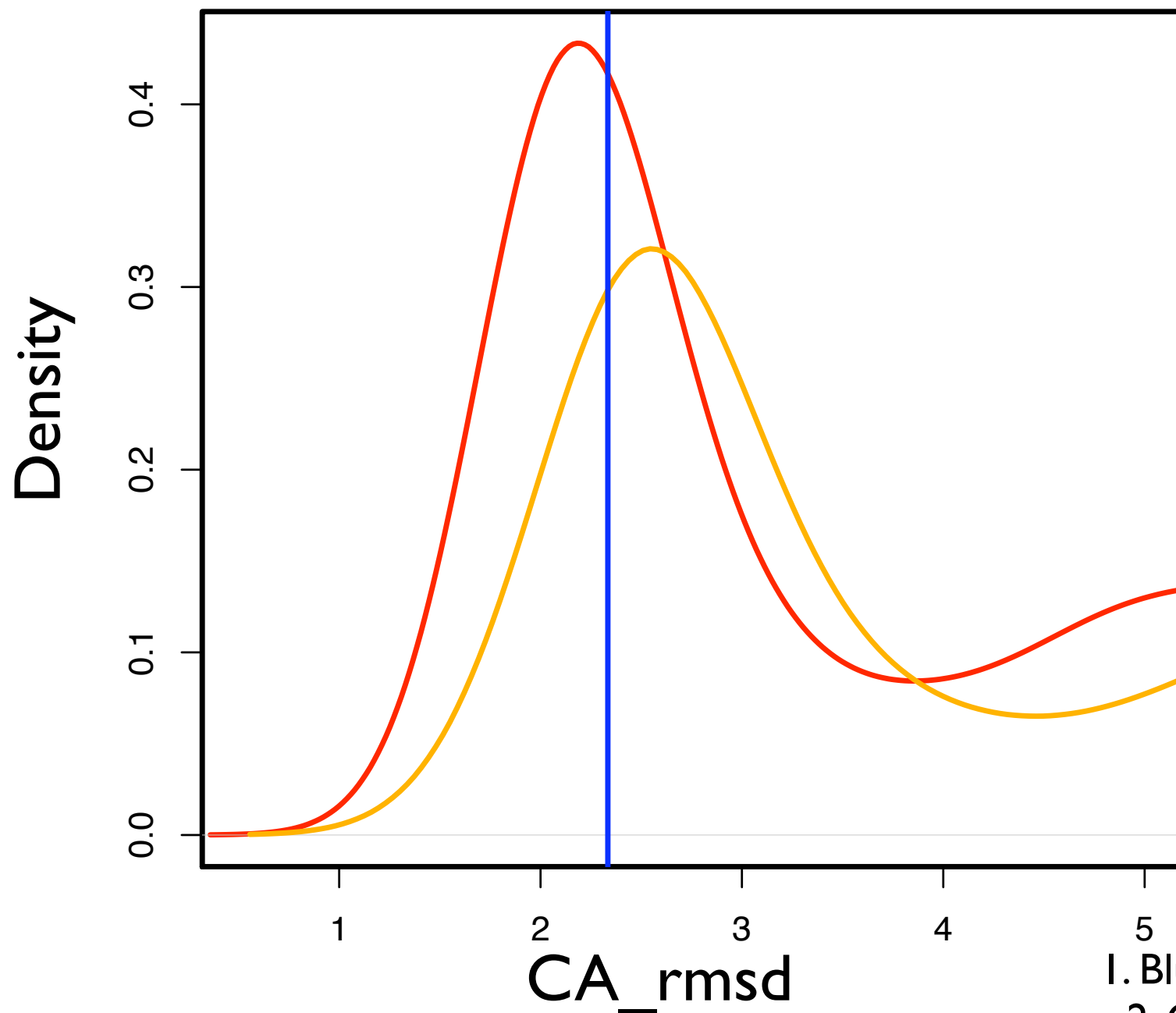
Distance predictions from different templates on the same pairs of atoms are simply added together.



# High Sequence Identity Comparative Modeling

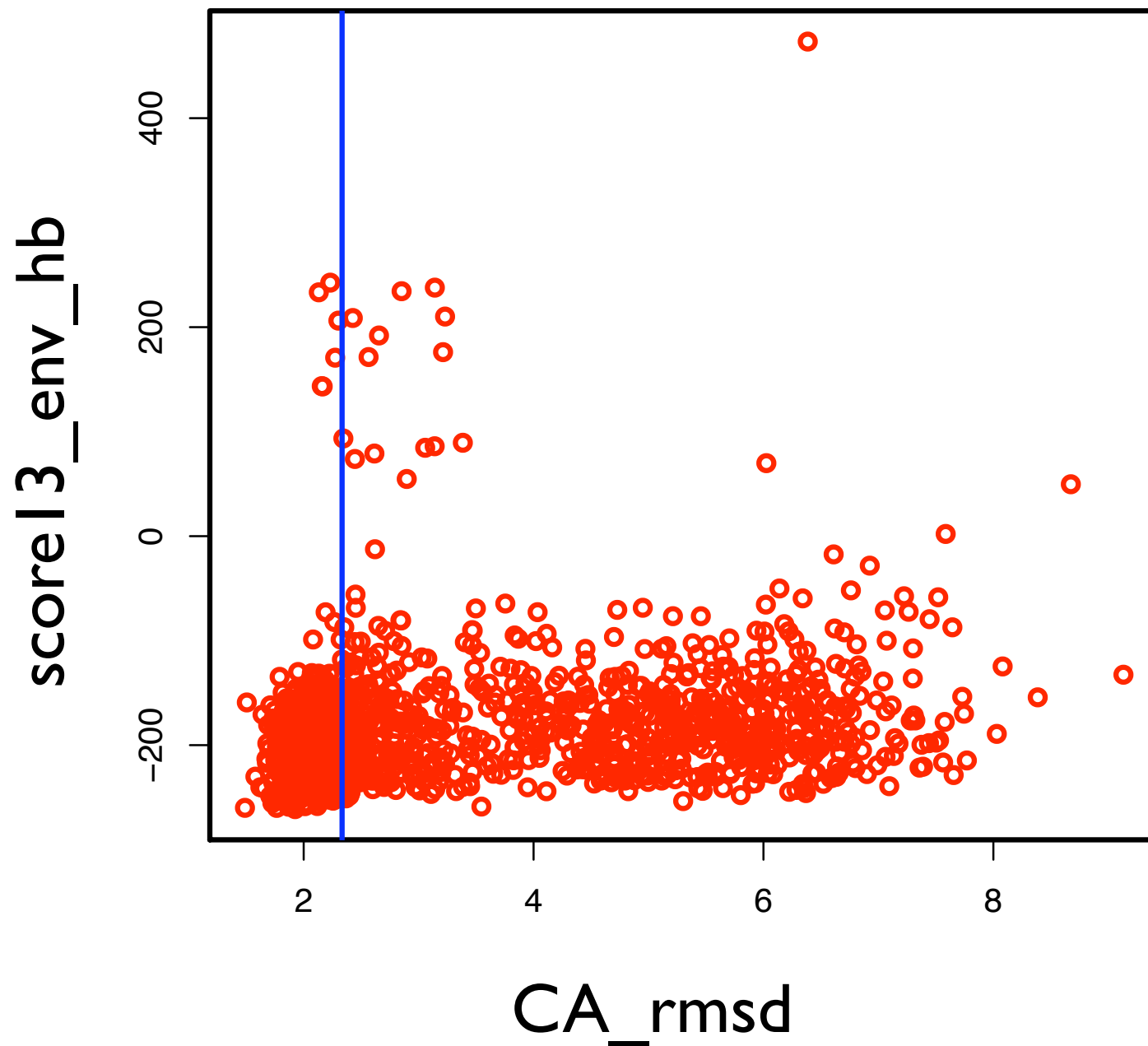
- For targets that have  $>30\%$  sequence identity, we'd like to be very conservative in moving our model away from the template.
- Threading
- Loop Modeling
- Minimization with Constraints

# CASP7 Target T380



1. Blue (Best Starting Model)
2. Orange (Loop Modeling)
3. Red (Minimization + Csts)

# CASP7 Target T380



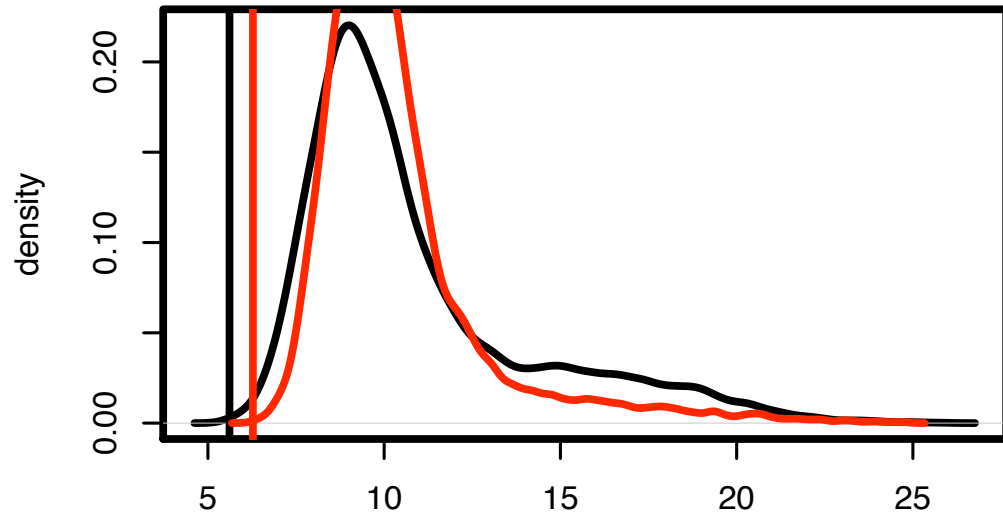
# FoldConstraints Benchmark

- FoldConstraints - fragment assembly in the presence of constraints (Carol Rohl, Oliver Lange).
- Benchmark ran on several CASP7 cases of various difficulty.
- Concoord constraints derived from the same alignments were compared with my constraints.

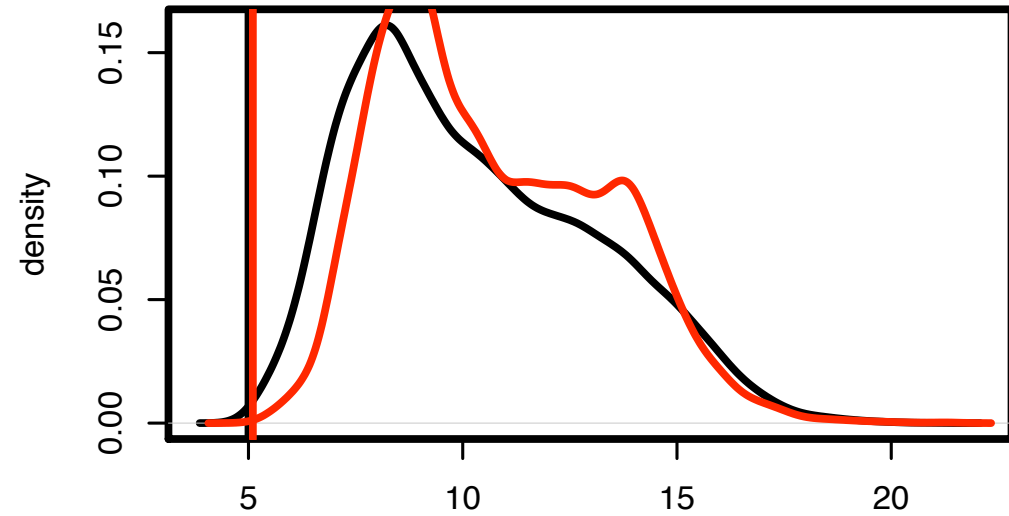
# Somewhat Depressing ...

t331

t357



rms  
black = james, red = concoord



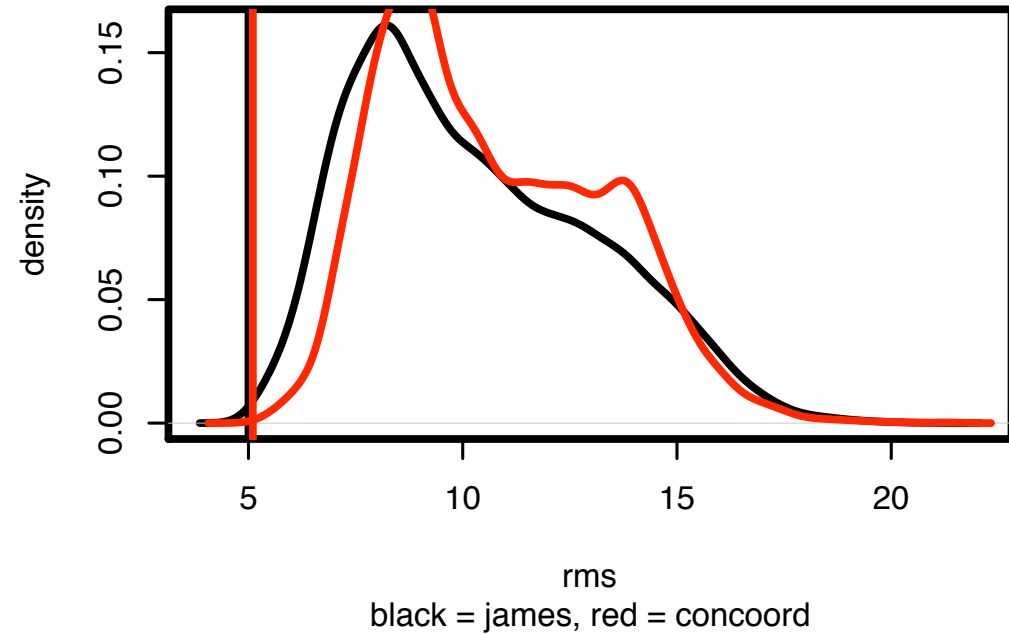
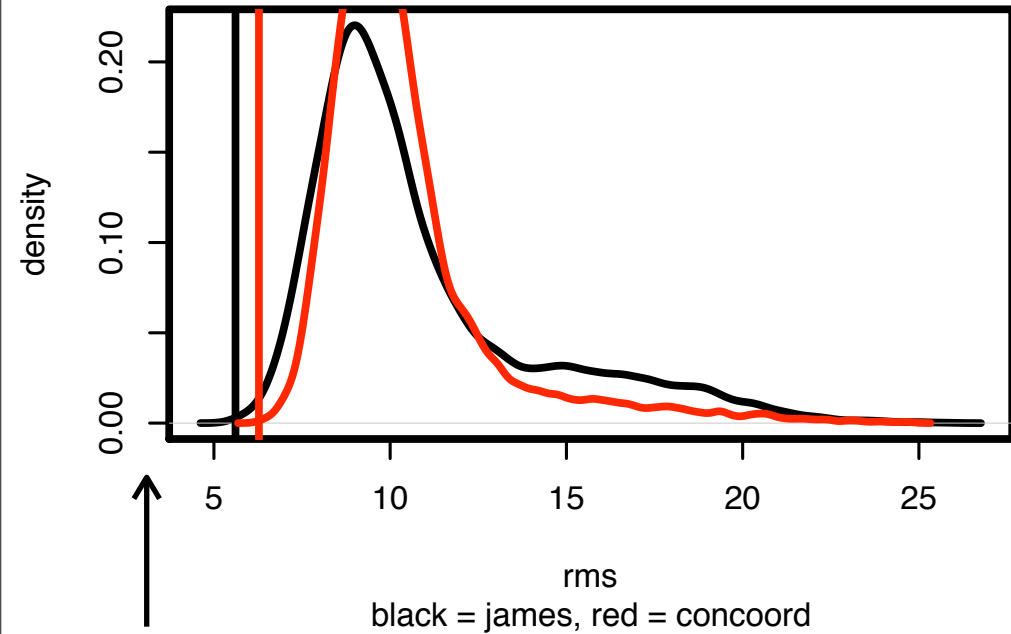
rms  
black = james, red = concoord

Black line = My method  
Red line = Concoord method

# Somewhat Depressing ...

t331

t357



## CASP7 Submission

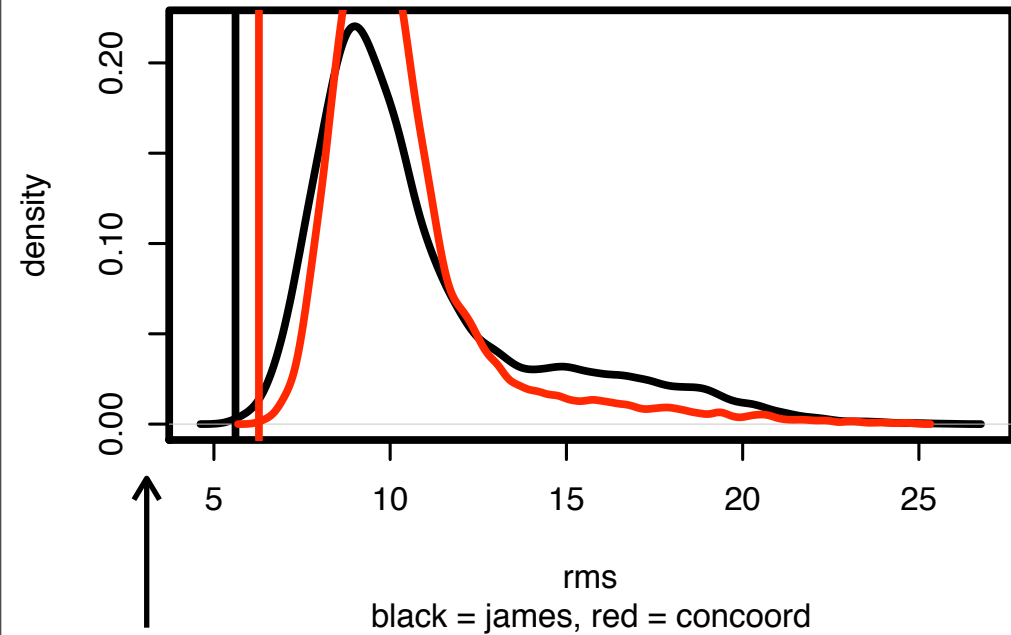
Black line = My method  
Red line = Concoord method



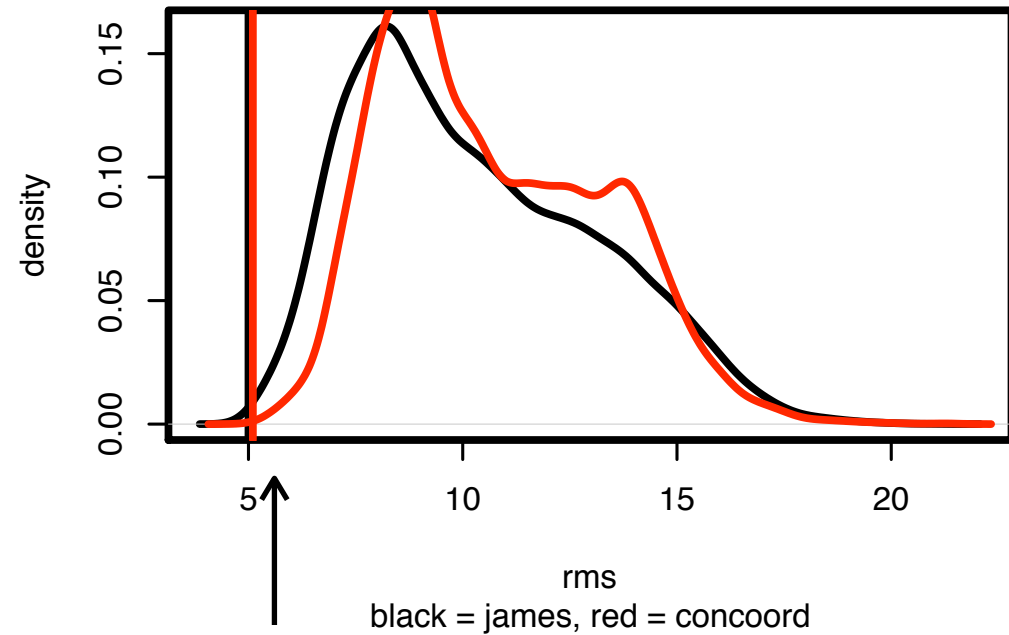
# Somewhat Depressing ...

t331

t357



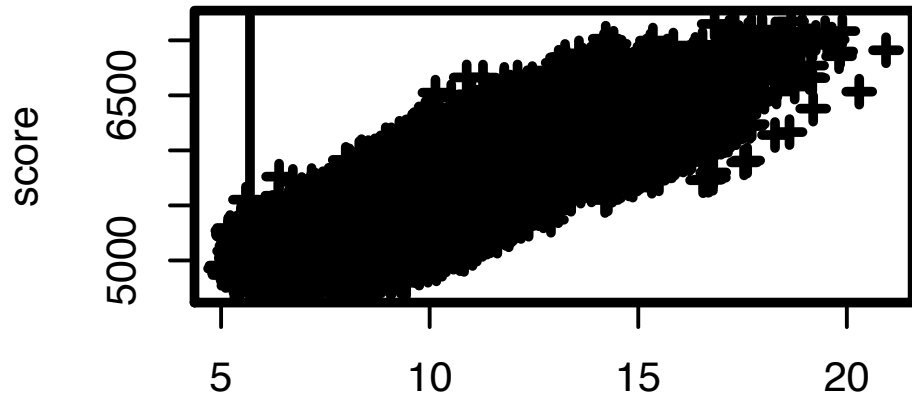
CASP7 Submission



CASP7 Submission

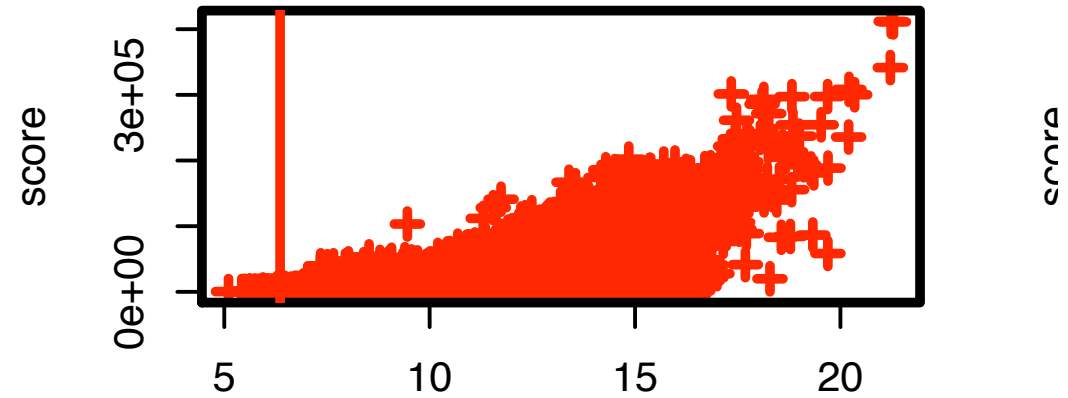
Black line = My method  
Red line = Concoord method

t357



RMS

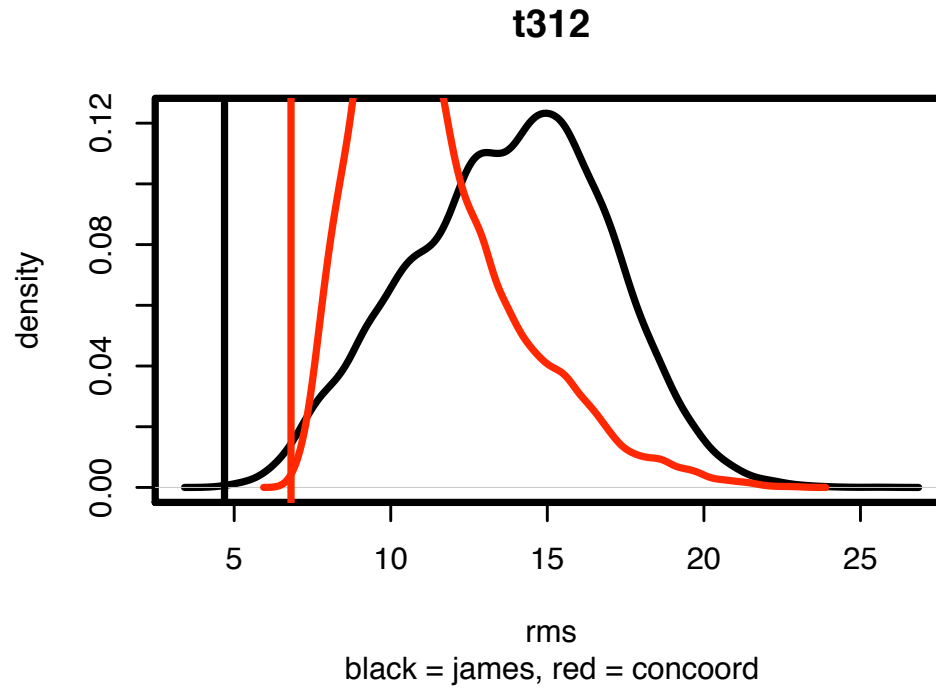
t357



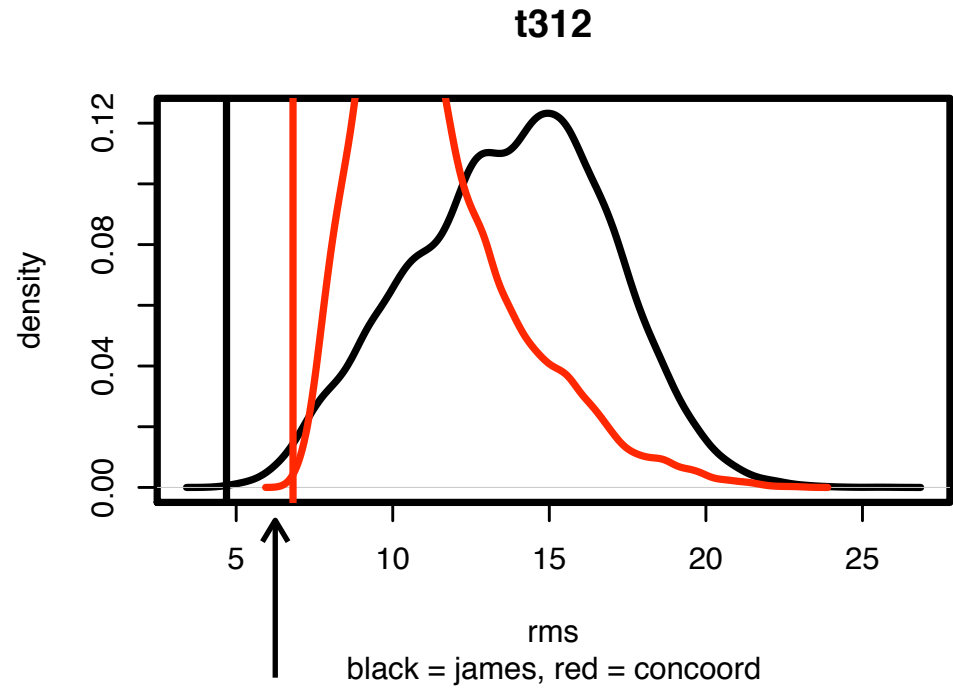
RMS

Line marks 10th percentile of structure by RMS

# Interesting!



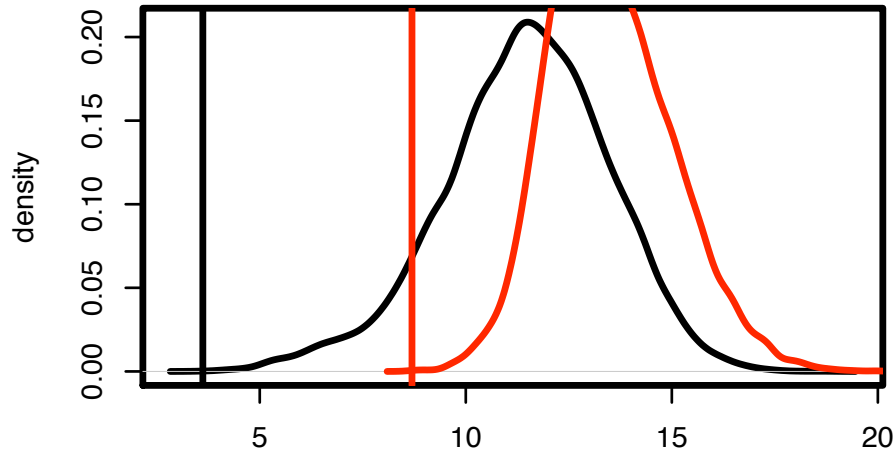
# Interesting!



CASP7 Submission

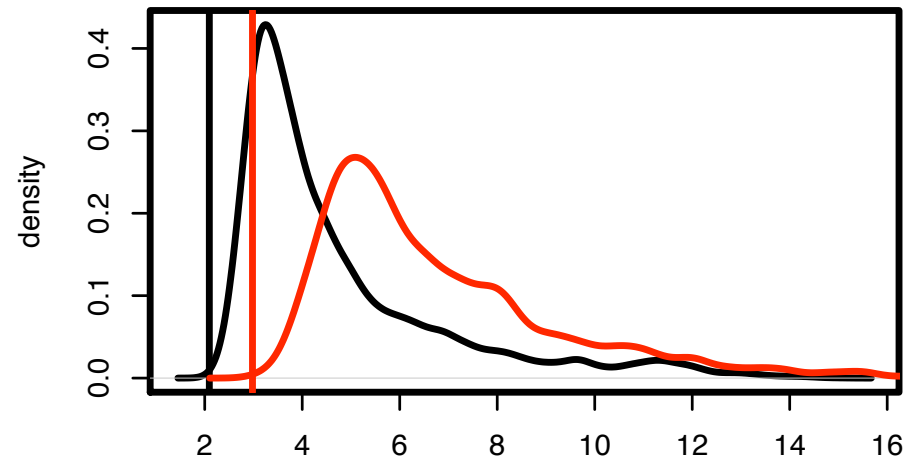
# Exciting!

t360



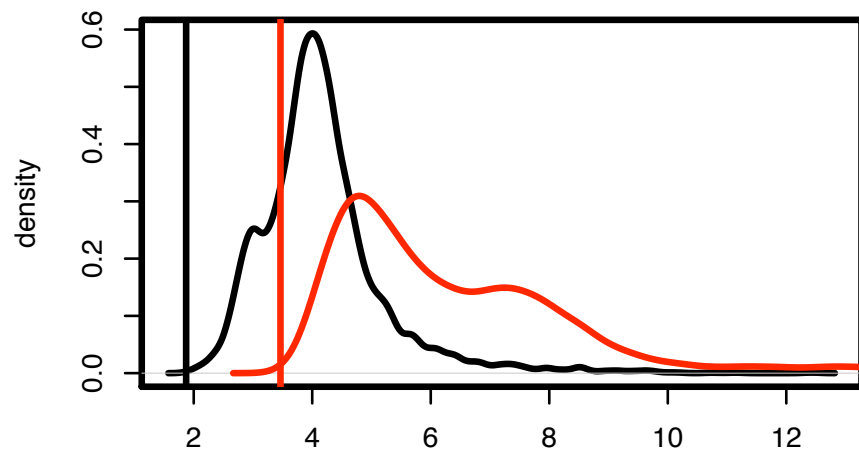
rms  
black = james, red = concoord

t374



rms  
black = james, red = concoord

t363



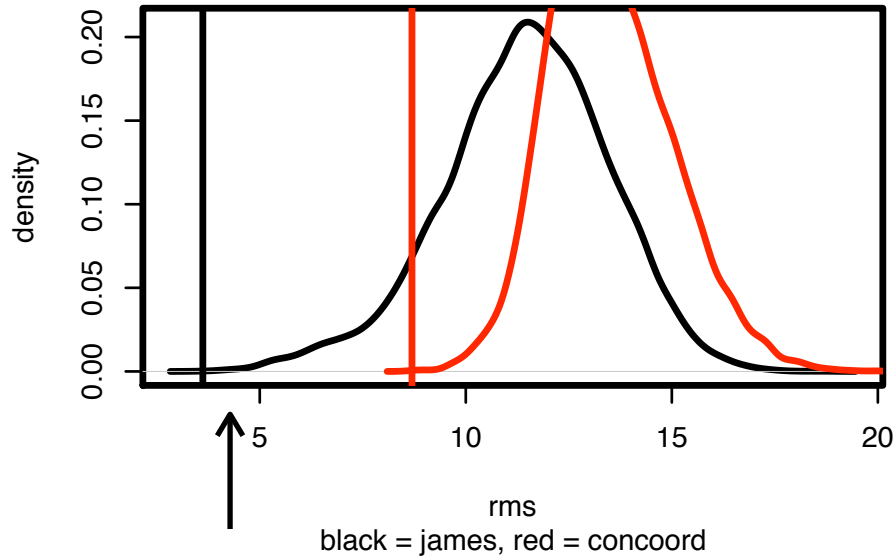
rms  
black = james, red = concoord

## Summary:

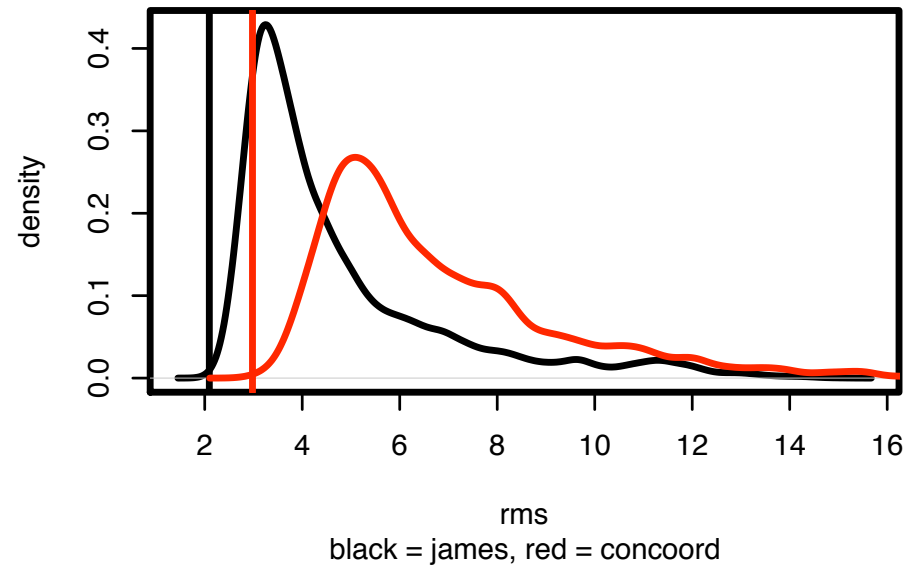
- 3 major improvements
- 2 minor improvements
- 1 improvement in the opposite direction

# Exciting!

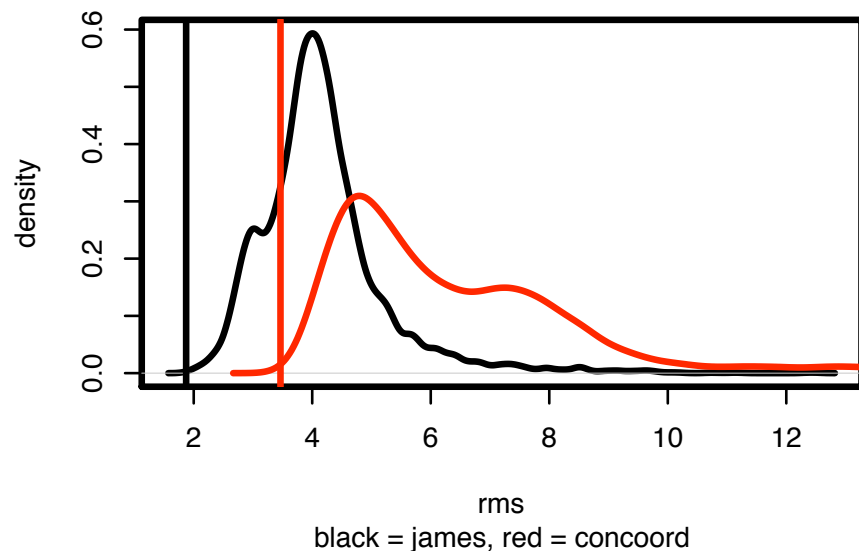
t360



t374



t363

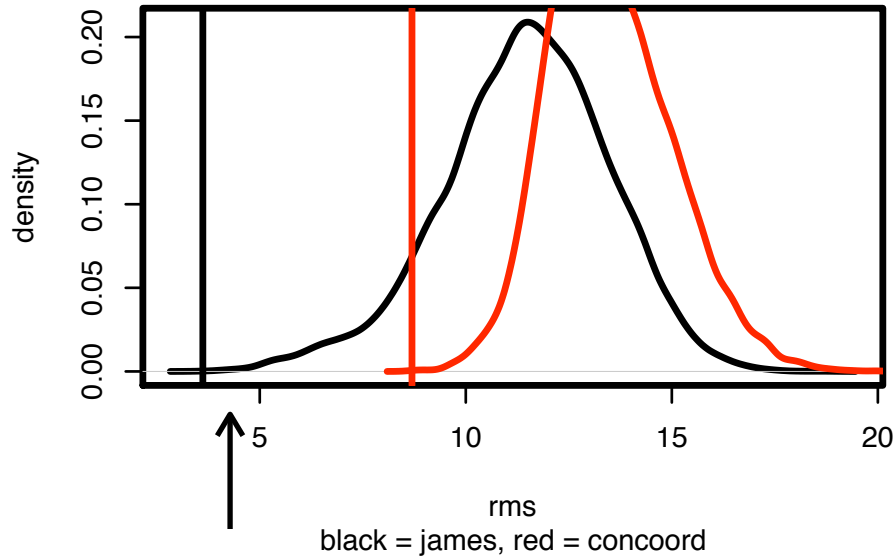


## Summary:

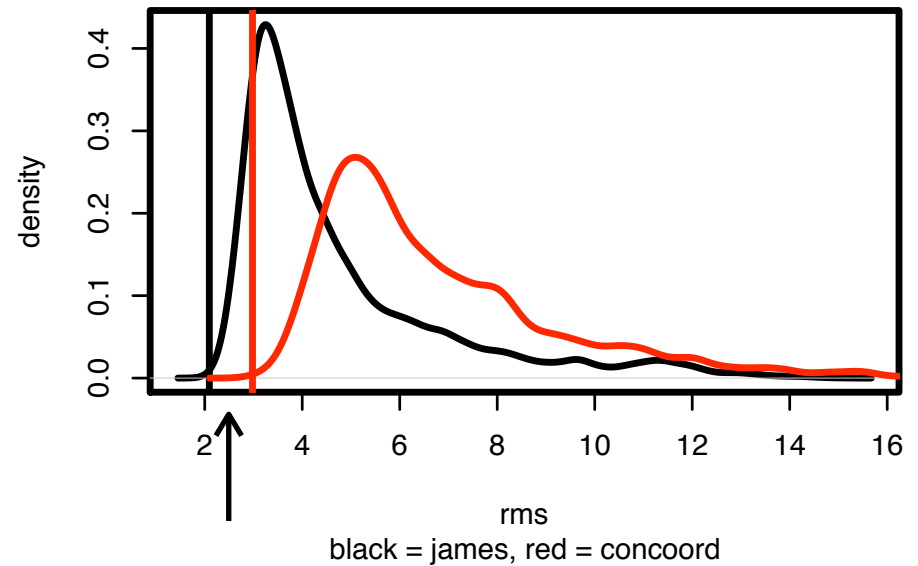
- 3 major improvements
- 2 minor improvements
- 1 improvement in the opposite direction

# Exciting!

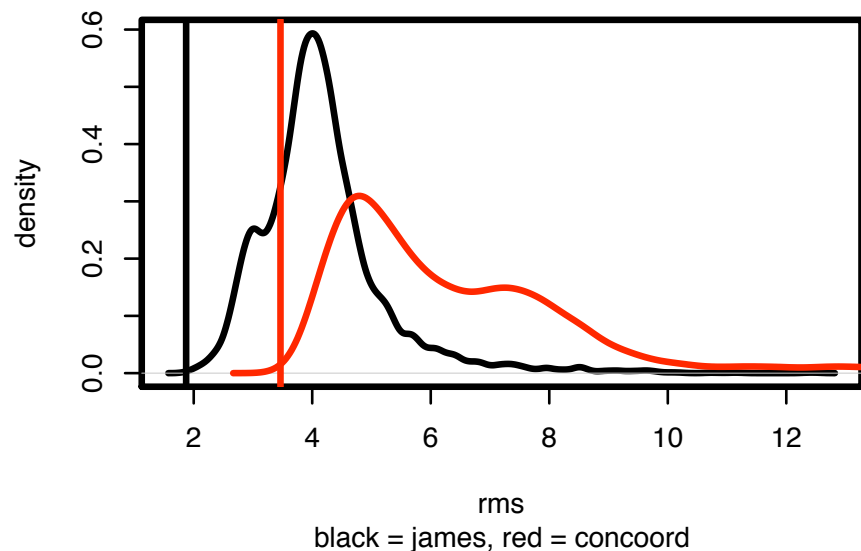
t360



t374



t363

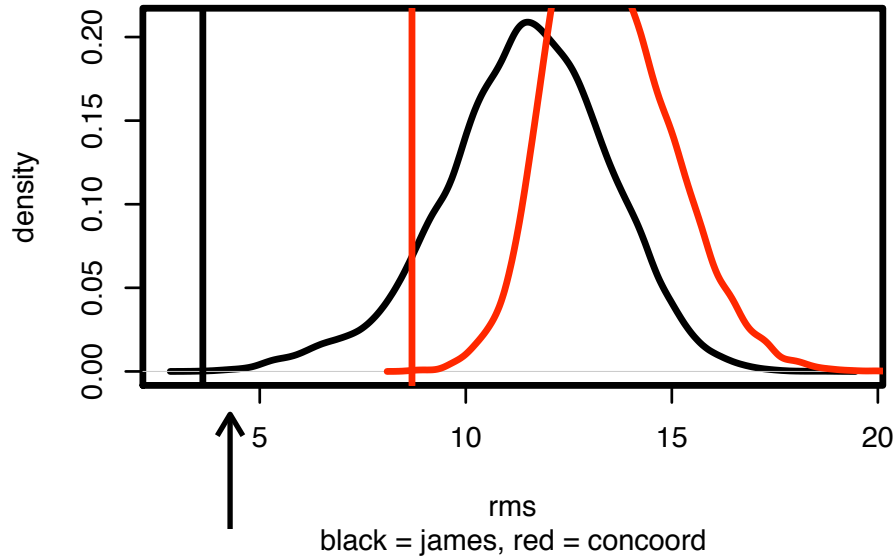


## Summary:

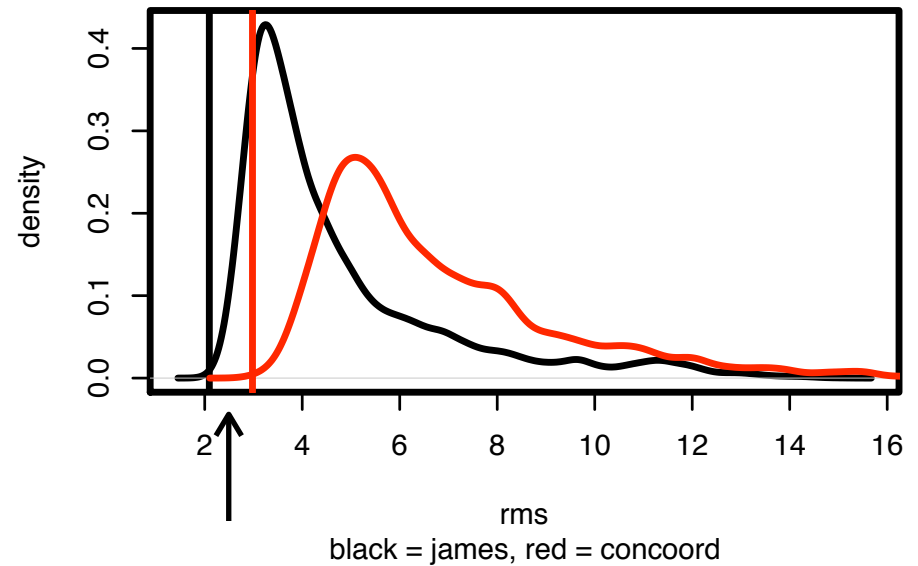
- 3 major improvements
- 2 minor improvements
- 1 improvement in the opposite direction

# Exciting!

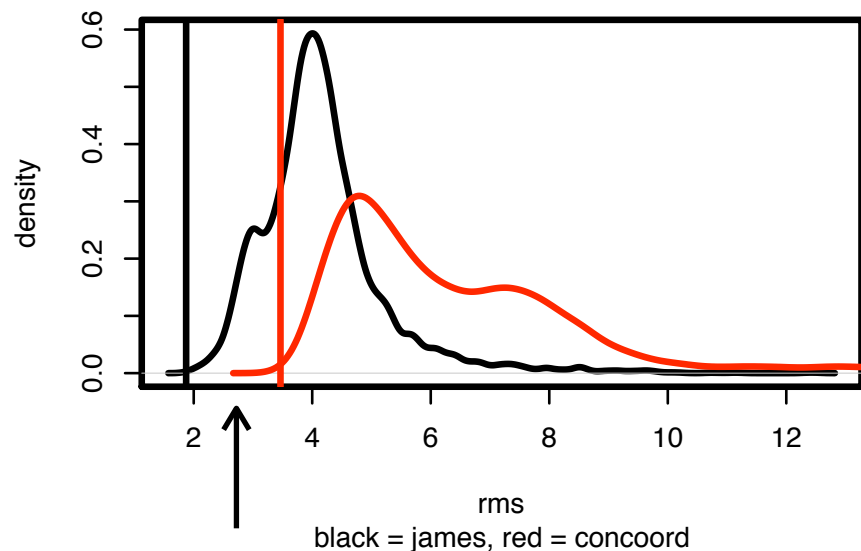
t360



t374



t363



## Summary:

- 3 major improvements
- 2 minor improvements
- 1 improvement in the opposite direction



# Future Goals

- More Benchmarks
- Better Understand Utility of Constraints in Structure Prediction
- Using Comparative Modeling
  - Docking Ligands and Comparative Models
  - Models of Disease-Related Proteins

# Acknowledgements

- Will Sheffler
- Elizabeth Kellogg
- Ian Davis
- Michael Tyka
- Oliver Lange
- Andrew Leaver-Fay
- Rhiju Das
- David Baker

