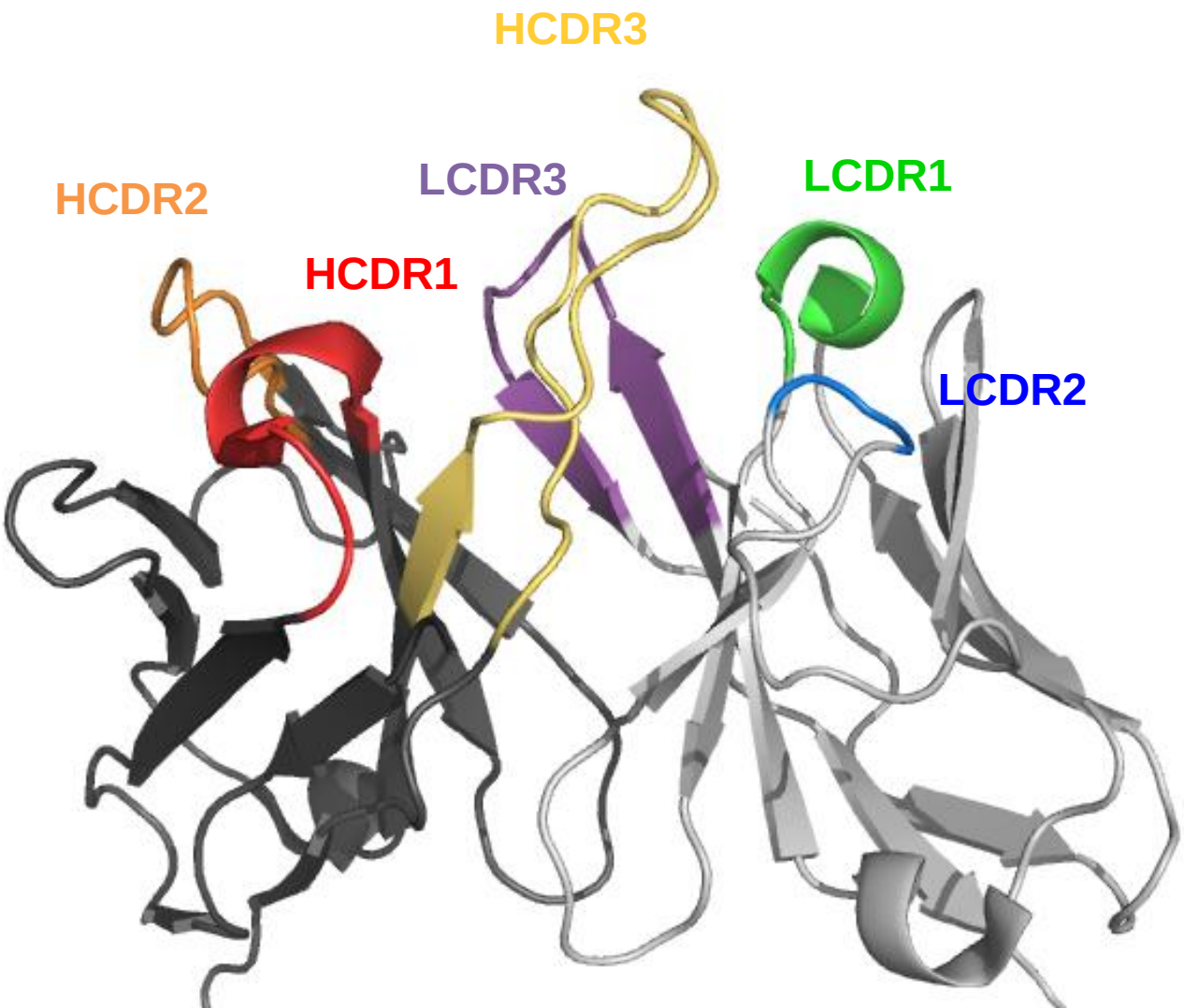
# HCDR3 Loop Modeling

Jessica Finn

Rosetta Workshop

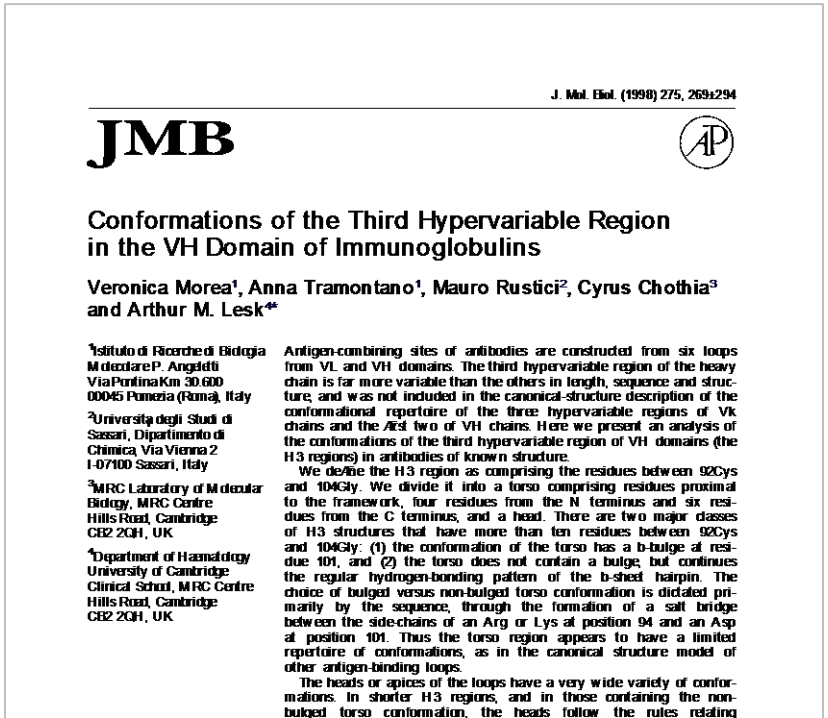
April 25th, 2017

# Antibodies recognize and bind targets using six complementarity determining region loops



# CDR loops have been extensively studied and characterized

- For 5 of 6 CDR loops, there are robust rules that define canonical structures
- The HCDR3 loop is extremely variable and only canonical domains can be identified



J. Mol. Biol. (1998) 275, 269±294

JMB

## Conformations of the Third Hypervariable Region in the VH Domain of Immunoglobulins

Veronica Morea[1], Anna Tramontano[1], Mauro Rustici[2], Cyrus Chothia[3] and Arthur M. Lesk[4]*

[1]Istituto di Ricerche di Biologia Molecolare P. Angeletti Via Pontina Km 30.600 00040 Pomezia (Roma), Italy

[2]Università degli Studi di Sassari, Dipartimento di Chimica, Via Vienna 2 I-07100 Sassari, Italy

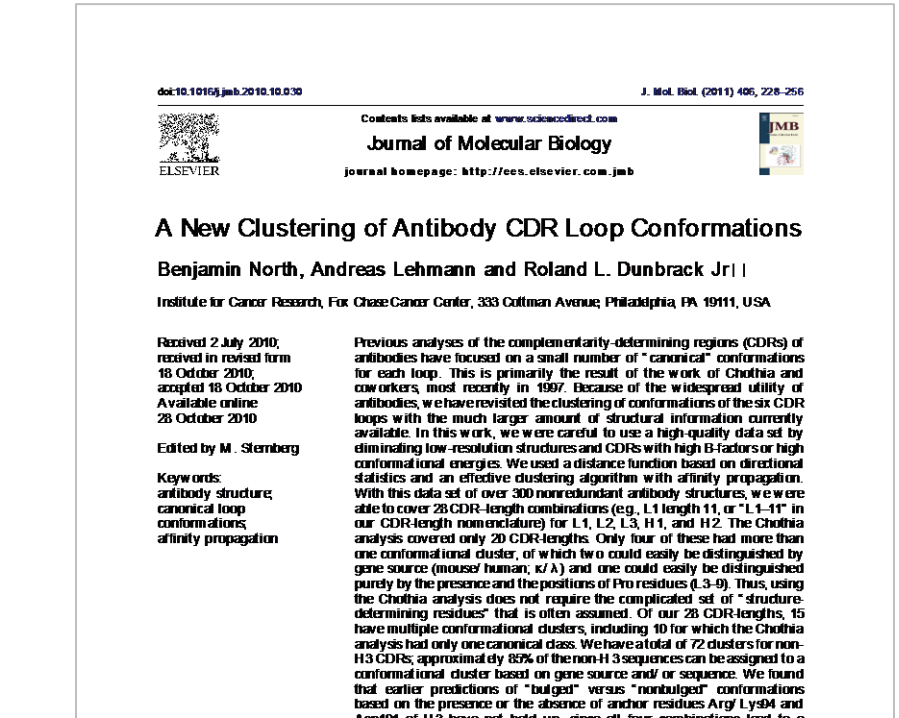
[3]MRC Laboratory of Molecular Biology, MRC Centre Hills Road, Cambridge CB2 2QH, UK

[4]Department of Haematology University of Cambridge Clinical School, MRC Centre Hills Road, Cambridge CB2 2QH, UK

Antigen-combining sites of antibodies are constructed from six loops from VL and VH domains. The third hypervariable region of the heavy chain is far more variable than the others in length, sequence and structure, and was not included in the canonical-structure description of the conformational repertoire of the three hypervariable regions of Vk chains and the first two of VH chains. Here we present an analysis of the conformations of the third hypervariable region of VH domains (the H3 regions) in antibodies of known structure.

We define the H3 region as comprising the residues between 92Cys and 104Gly. We divide it into a torso comprising residues proximal to the framework, four residues from the N terminus and six residues from the C terminus, and a head. There are two major classes of H3 structures that have more than ten residues between 92Cys and 104Gly: (1) the conformation of the torso has a b-bulge at residue 101, and (2) the torso does not contain a bulge, but continues the regular hydrogen-bonding pattern of the b-sheet hairpin. The choice of bulged versus non-bulged torso conformation is dictated primarily by the sequence, through the formation of a salt bridge between the side-chains of an Arg or Lys at position 94 and an Asp at position 101. Thus the torso region appears to have a limited repertoire of conformations, as in the canonical structure model of other antigen-binding loops.

The heads or apices of the loops have a very wide variety of conformations. In shorter H3 regions, and in those containing the non-bulged torso conformation, the heads follow the rules relating



doi:10.1016/j.jmb.2010.10.030 J. Mol. Biol. (2011) 406, 228–256

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

## A New Clustering of Antibody CDR Loop Conformations

Benjamin North, Andreas Lehmann and Roland L. Dunbrack Jr⁎

Institute for Cancer Research, Fox Chase Cancer Center, 333 Cottman Avenue, Philadelphia, PA 19111, USA

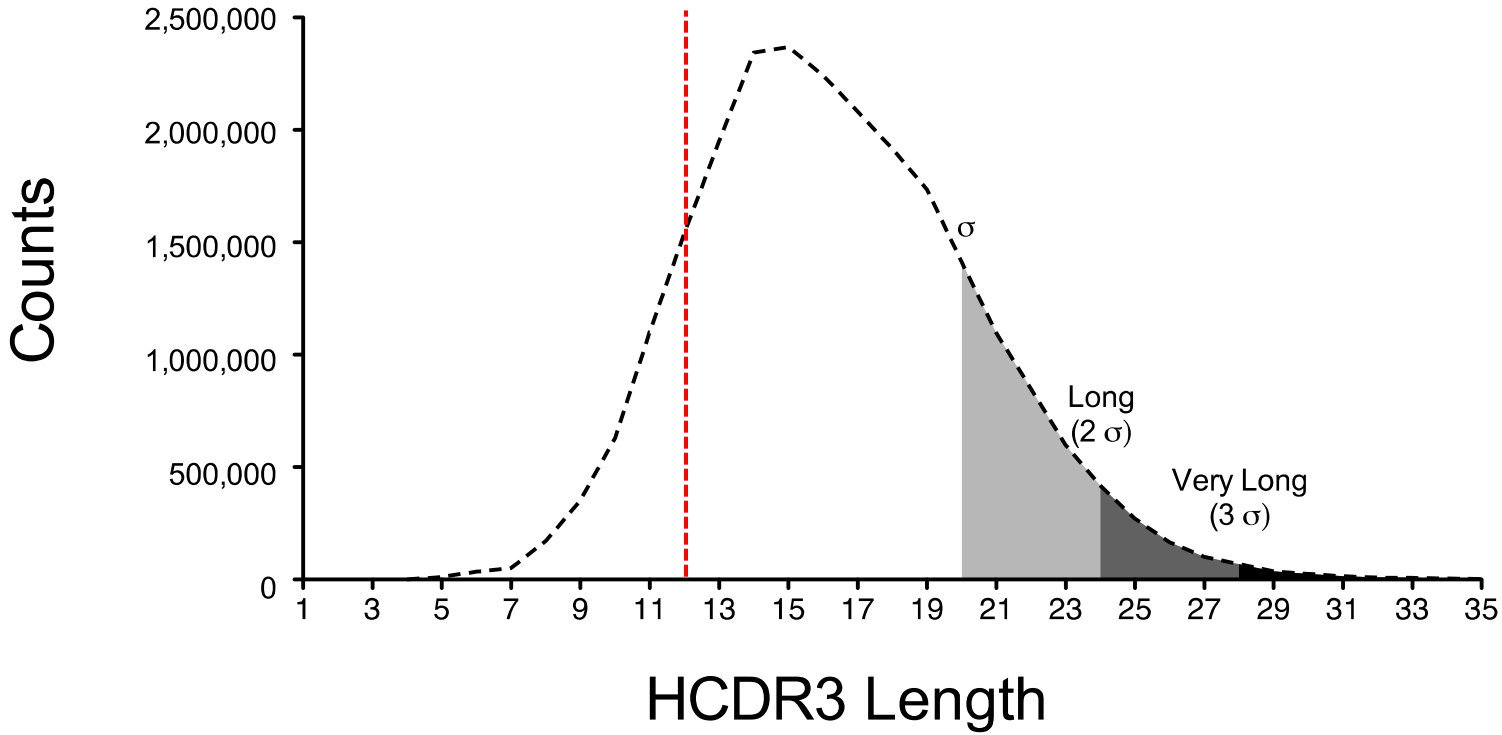
Received 2 July 2010;
received in revised form
18 October 2010;
accepted 18 October 2010
Available online
28 October 2010

Edited by M. Sternberg

Keywords:
antibody structure;
canonical loop
conformations;
affinity propagation

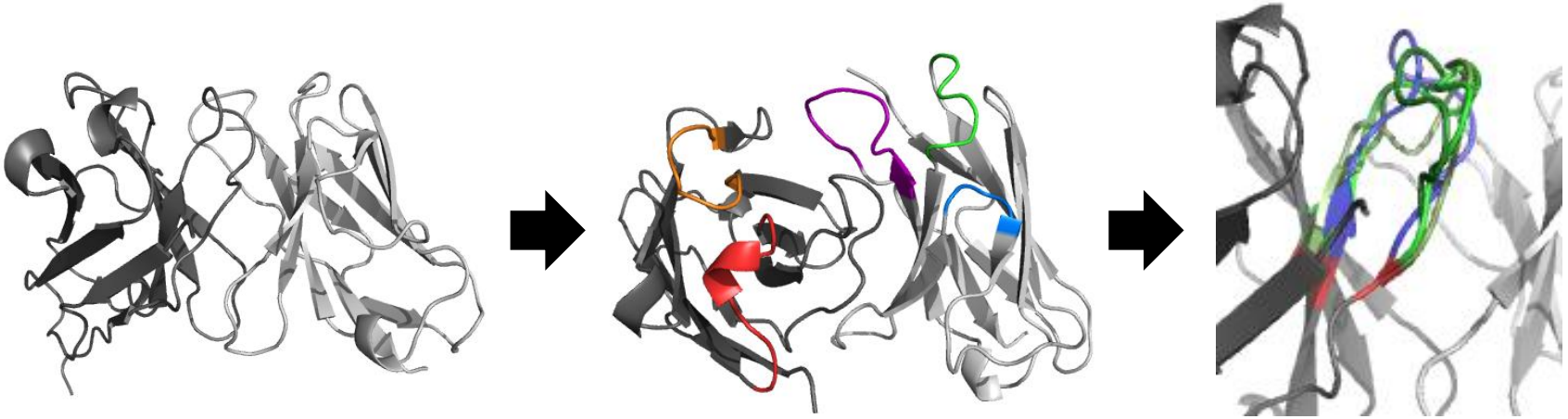
Previous analyses of the complementarity-determining regions (CDRs) of antibodies have focused on a small number of "canonical" conformations for each loop. This is primarily the result of the work of Chothia and coworkers, most recently in 1997. Because of the widespread utility of antibodies, we have revisited the clustering of conformations of the six CDR loops with the much larger amount of structural information currently available. In this work, we were careful to use a high-quality data set by eliminating low-resolution structures and CDRs with high B-factors or high conformational energies. We used a distance function based on directional statistics and an effective clustering algorithm with affinity propagation. With this data set of over 300 nonredundant antibody structures, we were able to cover 28 CDR-length combinations (e.g., L1 length 11, or "L1–11" in our CDR-length nomenclature) for L1, L2, L3, H1, and H2. The Chothia analysis covered only 20 CDR-lengths. Only four of these had more than one conformational cluster, of which two could easily be distinguished by gene source (mouse/human; κ/λ) and one could easily be distinguished purely by the presence and the positions of Pro residues (L3–9). Thus, using the Chothia analysis does not require the complicated set of "structure-determining residues" that is often assumed. Of our 28 CDR-lengths, 15 have multiple conformational clusters, including 10 for which the Chothia analysis had only one canonical class. We have a total of 72 clusters for non-H3 CDRs; approximately 85% of the non-H3 sequences can be assigned to a conformational cluster based on gene source and/or sequence. We found that earlier predictions of "bulged" versus "nonbulged" conformations based on the presence or the absence of anchor residues Arg/Lys94 and

# Many modeling techniques cannot predict HCDR3 loops




Jordan Willis

# How can we use Rosetta to model antibodies?



Identify homologous framework regions

Graft canonical CDR loops

*De novo* model HCDR3

# *De novo* loop modeling in Rosetta

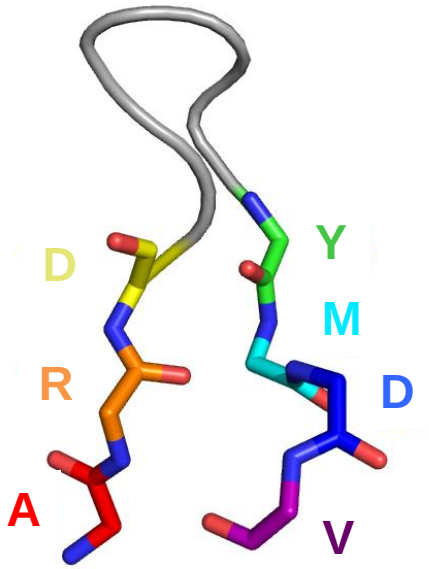
ARDYLD

AYYMDV

**Primary Sequence**

2° Structure Prediction,
PDB Fragments,
Energy Function,
Restraints

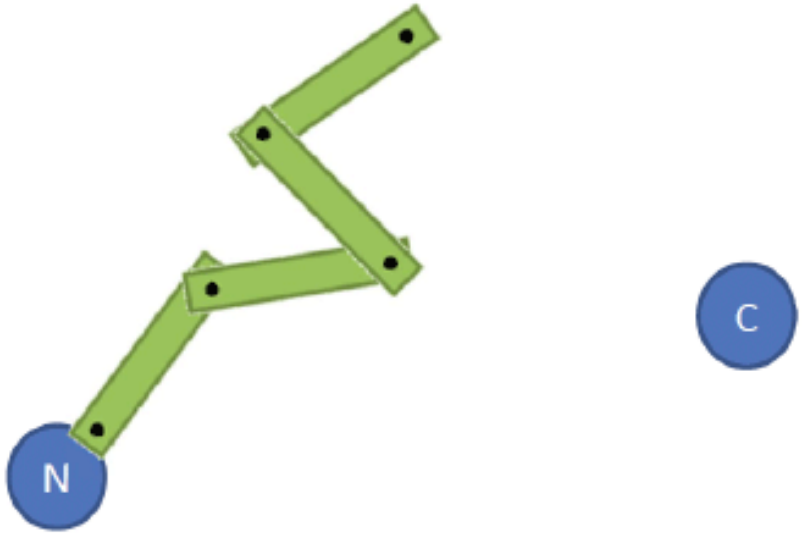
*de novo* modeling




**Tertiary Structure**

# *De novo* loop modeling in Rosetta

- **Stage 1: Remodel (Cyclic Coordinate Descent (CCD))**
  - Fast broad sampling of backbone conformations (centroid)
- **Stage 2: Refine (Kinematic Loop Closure Method (KIC))**
  - All-atom side chains, evaluated with Rosetta's high-resolution scoring function.



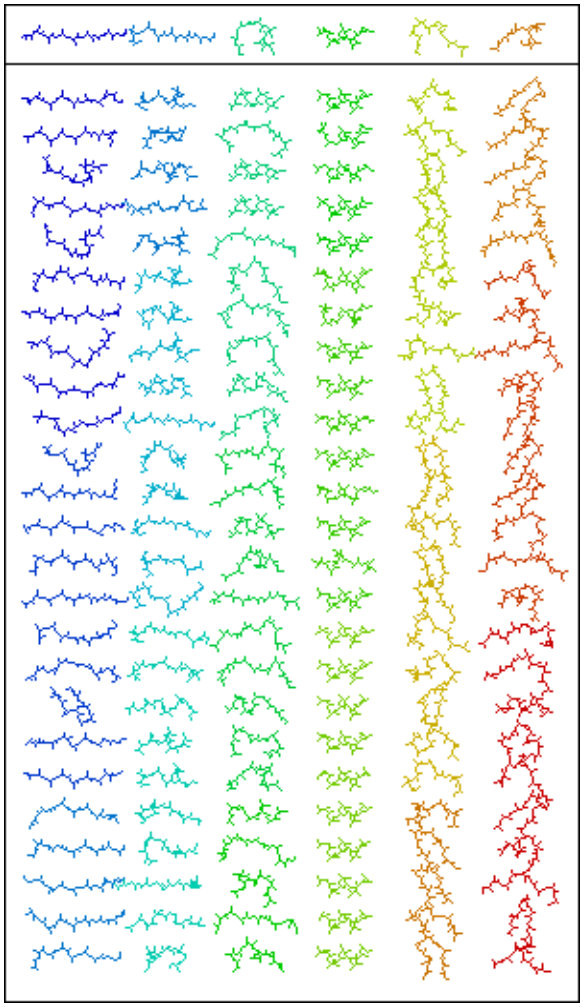

Canutescu, A. A., & Dunbrack, R. L. (2003). Protein Science, 12(5), 963-972.

# Necessary input files for the *de novo* loop modeling protocol

- FASTA file of your antibody sequence
- Secondary structure prediction files
- **Fragment library files**
- **Options file**
- ***(Optional) Constraints file***

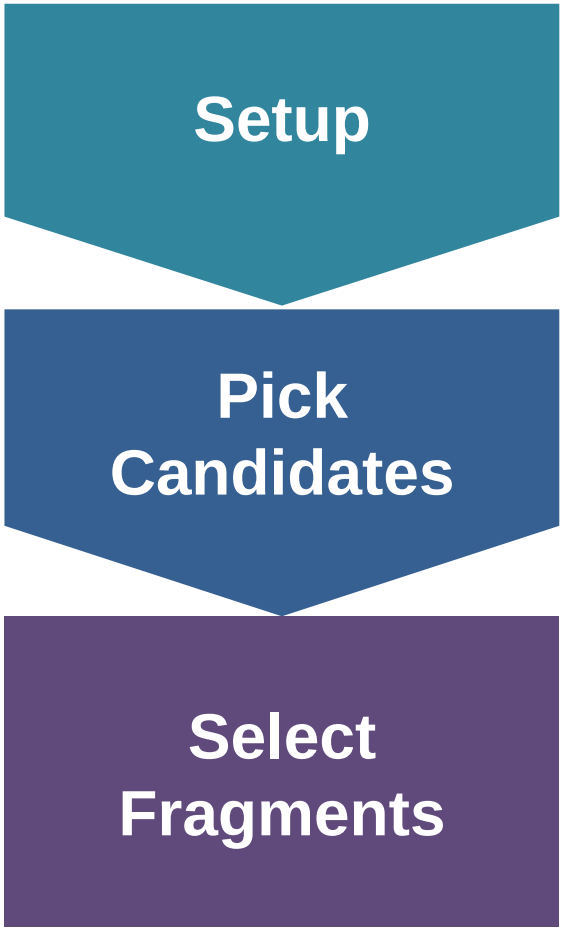
# Fragment library generation

# What are fragments?



- 3 and 9 amino acid peptides generated from the PDB
- Fragments change the geometry of the protein
- Scoring functions identify and maintain good fragments

# Fragment file generation



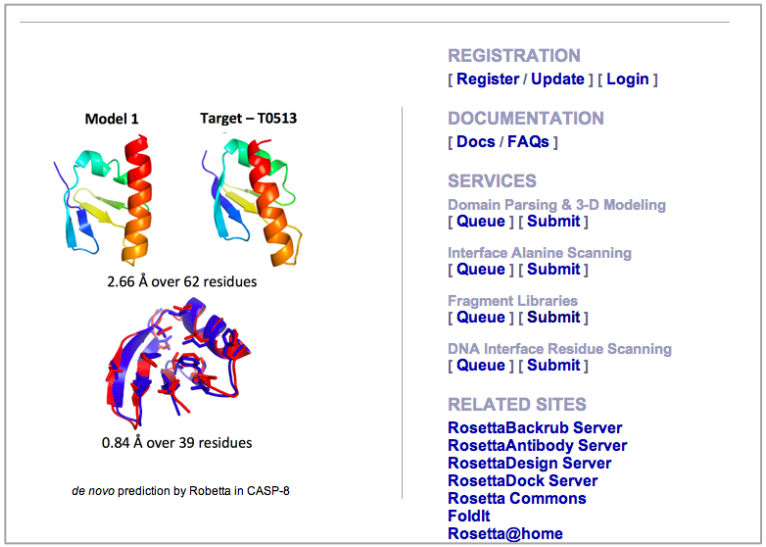

- Vall database
- Primary sequence
- Secondary structure prediction
- NMR data (if applicable)

- Gather all possible fragments
- Score candidates based on input

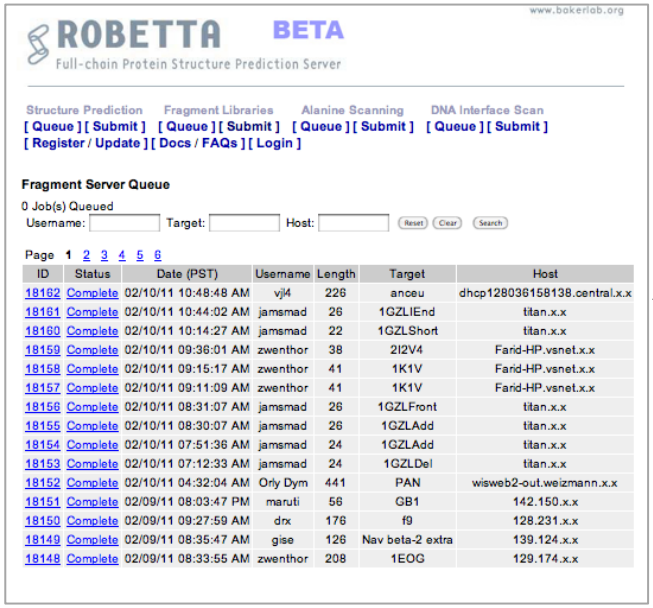
- Keep the best N fragments
- Default = 200 per sequence
- Write to fragment files

# Making fragments with Robetta

http://robetta.bakerlab.org/








| ID | Status | Date (PST) | Username | Length | Target | Host |
|---|---|---|---|---|---|---|
| 18162 | Complete | 02/10/11 10:48:48 AM | vjl4 | 226 | anceu | dhcp128036158138.central.x.x |
| 18161 | Complete | 02/10/11 10:44:02 AM | jamsmad | 26 | 1GZLIEnd | titan.x.x |
| 18160 | Complete | 02/10/11 10:14:27 AM | jamsmad | 22 | 1GZLShort | titan.x.x |
| 18159 | Complete | 02/10/11 09:36:01 AM | zwenthor | 38 | 2I2V4 | Farid-HP.vsnet.x.x |
| 18158 | Complete | 02/10/11 09:15:17 AM | zwenthor | 41 | 1K1V | Farid-HP.vsnet.x.x |
| 18157 | Complete | 02/10/11 09:11:09 AM | zwenthor | 41 | 1K1V | Farid-HP.vsnet.x.x |
| 18156 | Complete | 02/10/11 08:31:07 AM | jamsmad | 26 | 1GZLFront | titan.x.x |
| 18155 | Complete | 02/10/11 08:30:07 AM | jamsmad | 26 | 1GZLAdd | titan.x.x |
| 18154 | Complete | 02/10/11 07:51:36 AM | jamsmad | 24 | 1GZLAdd | titan.x.x |
| 18153 | Complete | 02/10/11 07:12:33 AM | jamsmad | 24 | 1GZLDel | titan.x.x |
| 18152 | Complete | 02/10/11 04:32:04 AM | Orly Dym | 441 | PAN | wisweb2-out.weizmann.x.x |
| 18151 | Complete | 02/09/11 08:03:47 PM | maruti | 56 | GB1 | 142.150.x.x |
| 18150 | Complete | 02/09/11 09:27:59 AM | drx | 176 | f9 | 128.231.x.x |
| 18149 | Complete | 02/09/11 08:35:47 AM | gise | 126 | Nav beta-2 extra | 139.124.x.x |
| 18148 | Complete | 02/09/11 08:33:55 AM | zwenthor | 208 | 1EOG | 129.174.x.x |

# Making fragments with the Fragment Picker

- Using the fragment picker provides more control over fragment generation
- The fragment picker is an application within Rosetta, and is run from the command line
- Output files from Robetta can be used as preparation files for the fragment picker

# Rosetta options file

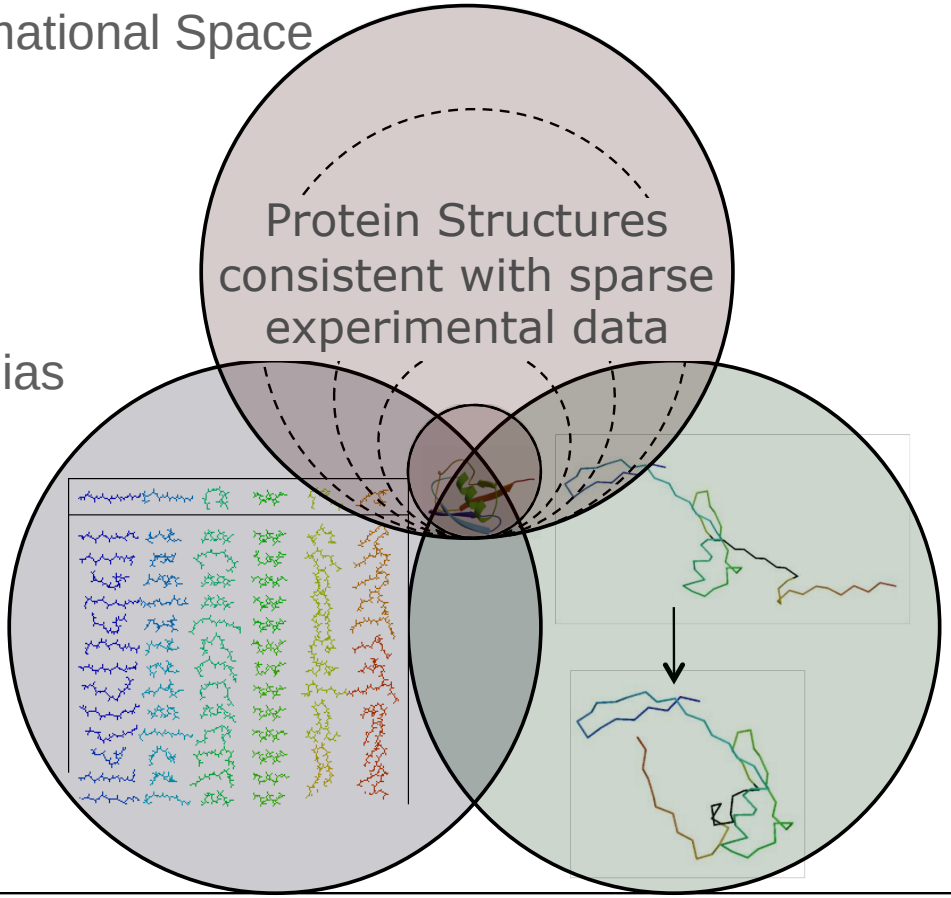
# Loop building options

-s 4m5y_renum.pdb  #Starting pdb with loops to rebuild
-loops:loop_file 4m5y_.loops #Loop file, defining which loop to rebuild
-loops:extended true #Force phi-psi angles to be set to 180 degrees
-loops:idealize_after_loop_close #Give idealized phi and psi angles after loop is closed
-loops:relax fastrelax #Does a minimization of the structure in the torsion space
-loops:fast #Decreases the monte carlo cycles of loop rebuilding, decreasing comp time
-loops:frag_sizes 9 3 1 #Fragment lengths
-loops:frag_files 4m5y_frags.200.9mers 4m5y_frags.200.3mers none
-loops:remodel quick_ccd #Use CCD for coarse loop modeling
-loops:refine refine_kic #Refine with KIC
-score:weights talaris2014.wts
-out:file:fullatom #Output file will be full-atom
-ex1 #Include extra chi1 rotamers
-ex2
-nstruct 2 #Number of models to build. 1000 recommended for production runs

# Rosetta constraint files

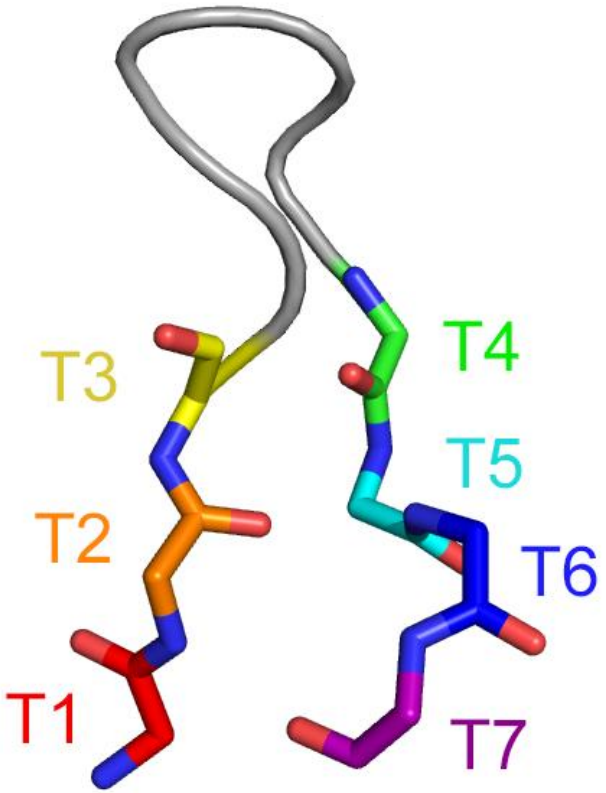
# Why combine experimental restraints with Rosetta?

- Complete Conformational Space
- Local Sequence Bias
- Energy Evaluation of non-Local Interactions

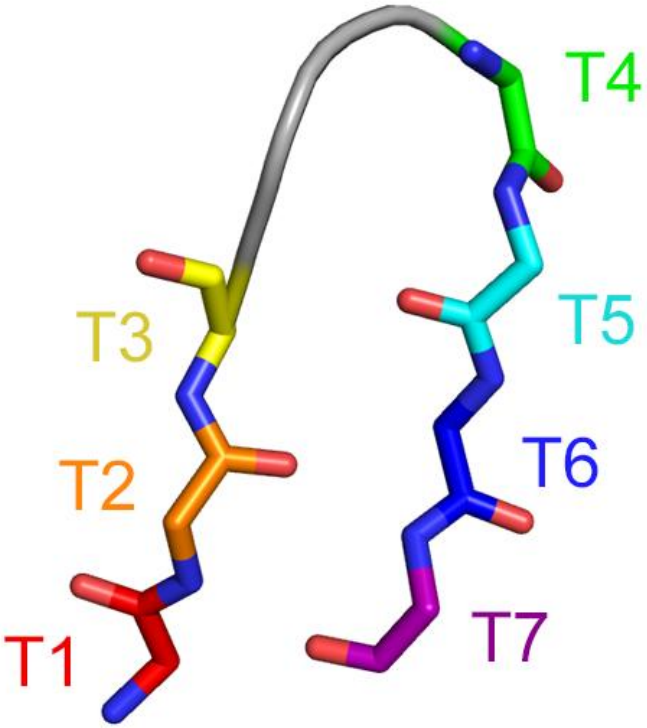



Jens Meiler

# HCDR3 torso domain conformations

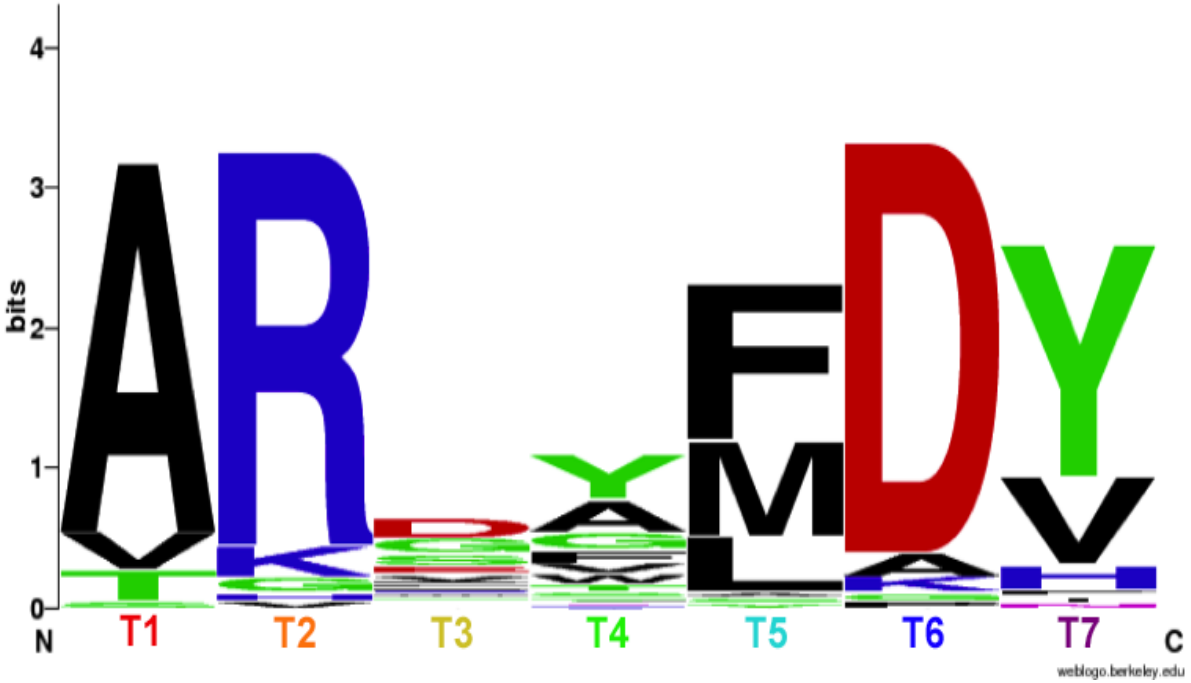


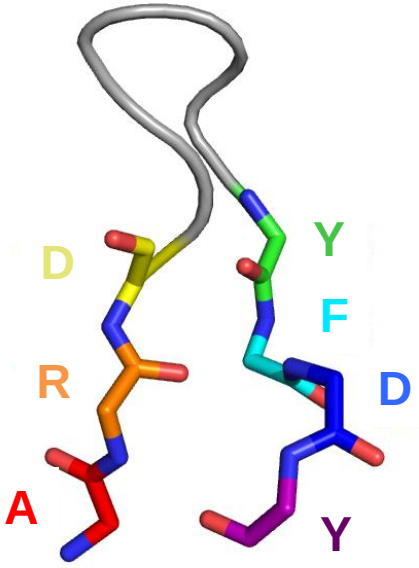

Bulged




Non-bulged or extended

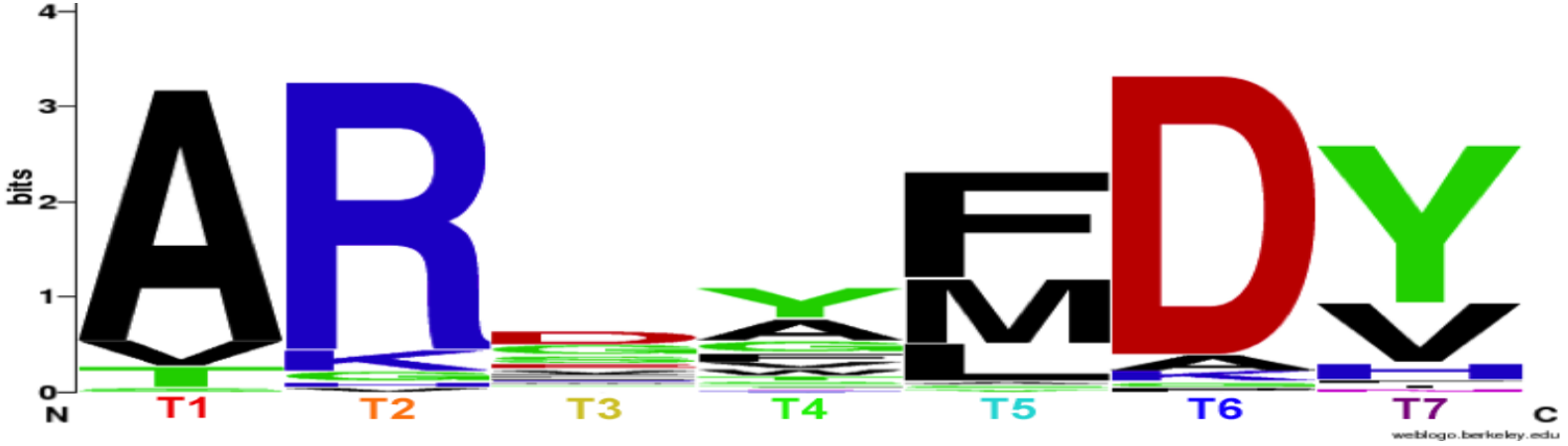
# Bulged torso structures share similar sequences



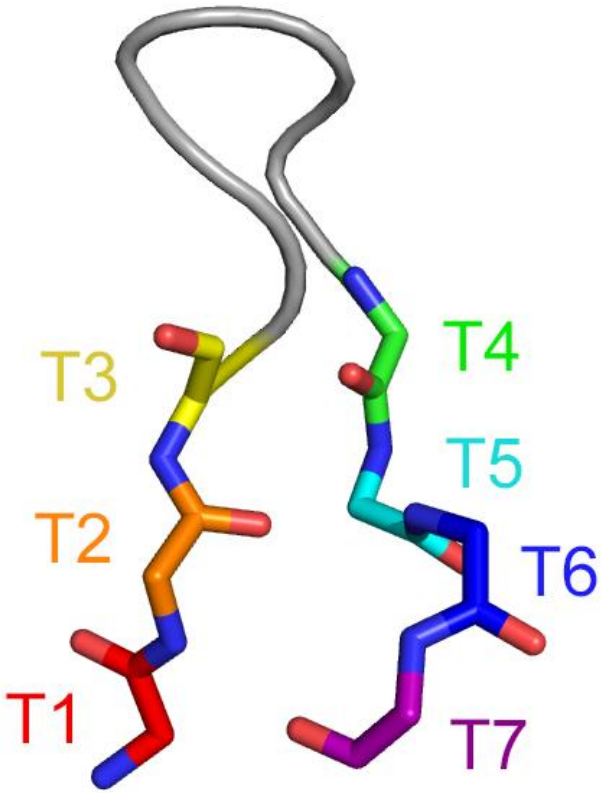
# The bulged torso sequence motif is germline encoded

**Bulged Consensus**

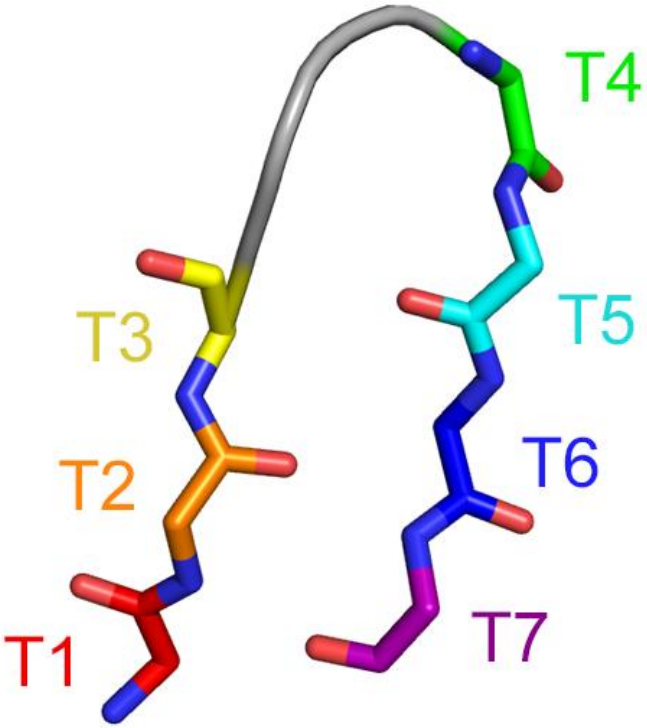



**Germline Consensus**

# HCDR3 torso domain conformations

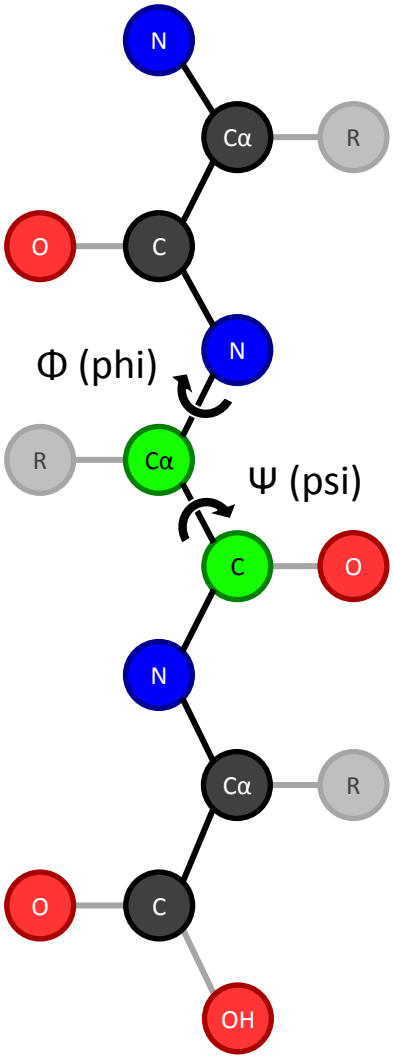


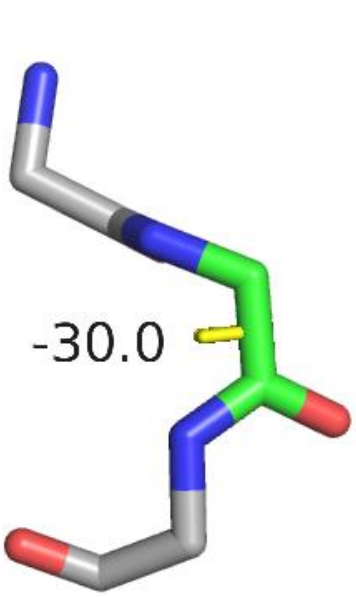

Bulged



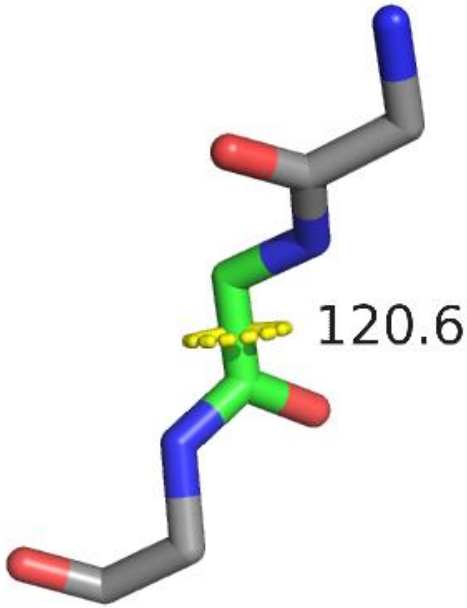

Non-bulged or extended

# Protein backbone structure



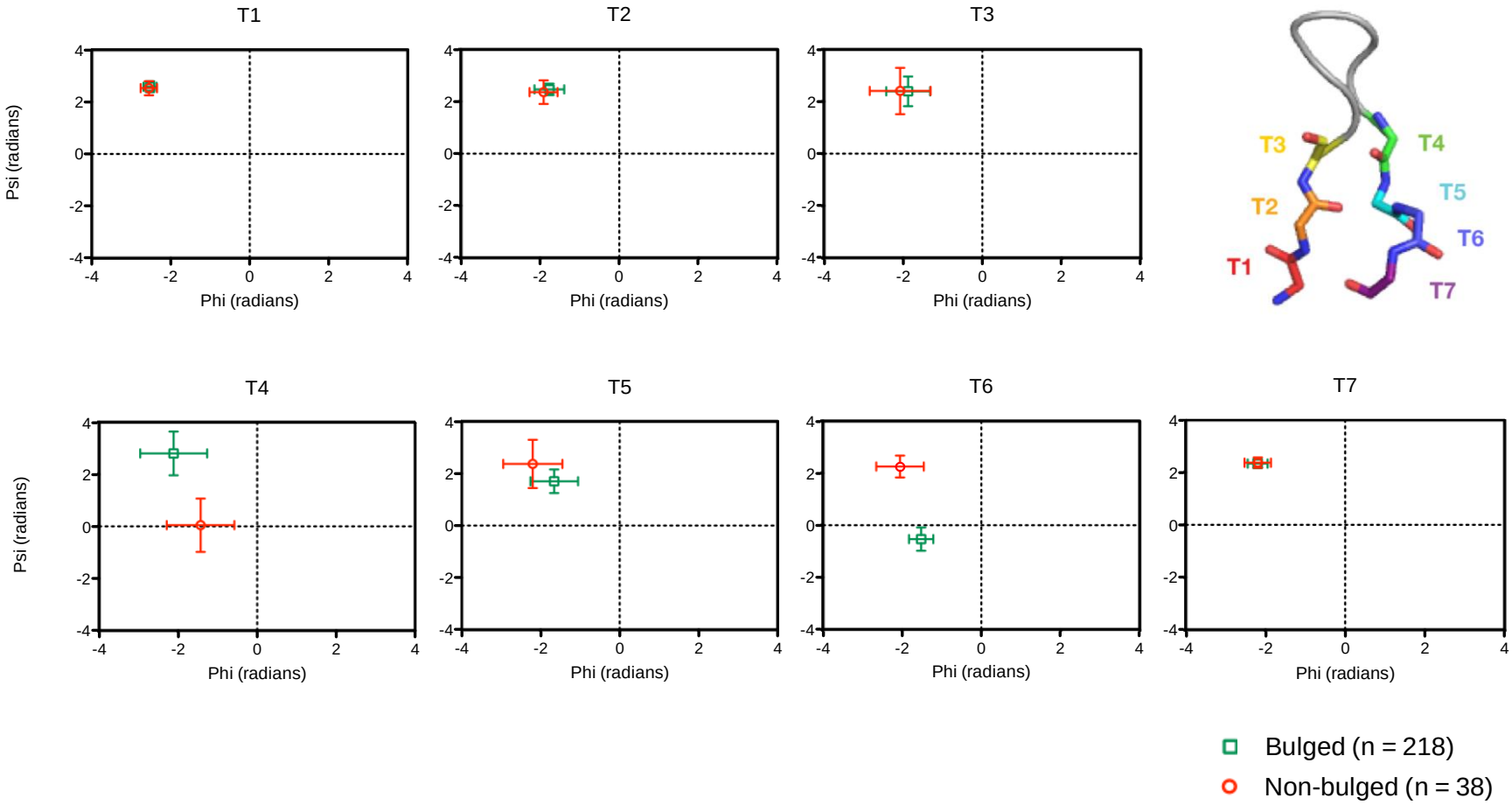




Bulged




Non-bulged or extended

# Bulged and non-bulged dihedral angle measurements differ



# Rosetta uses knowledge-based restraints to restrict modeling

- Rosetta constraint files apply a score penalty to models that break defined rules

$$f(x) = \left( \frac{NearestAngleRadians(x, x0) - x0}{sd} \right)^2$$

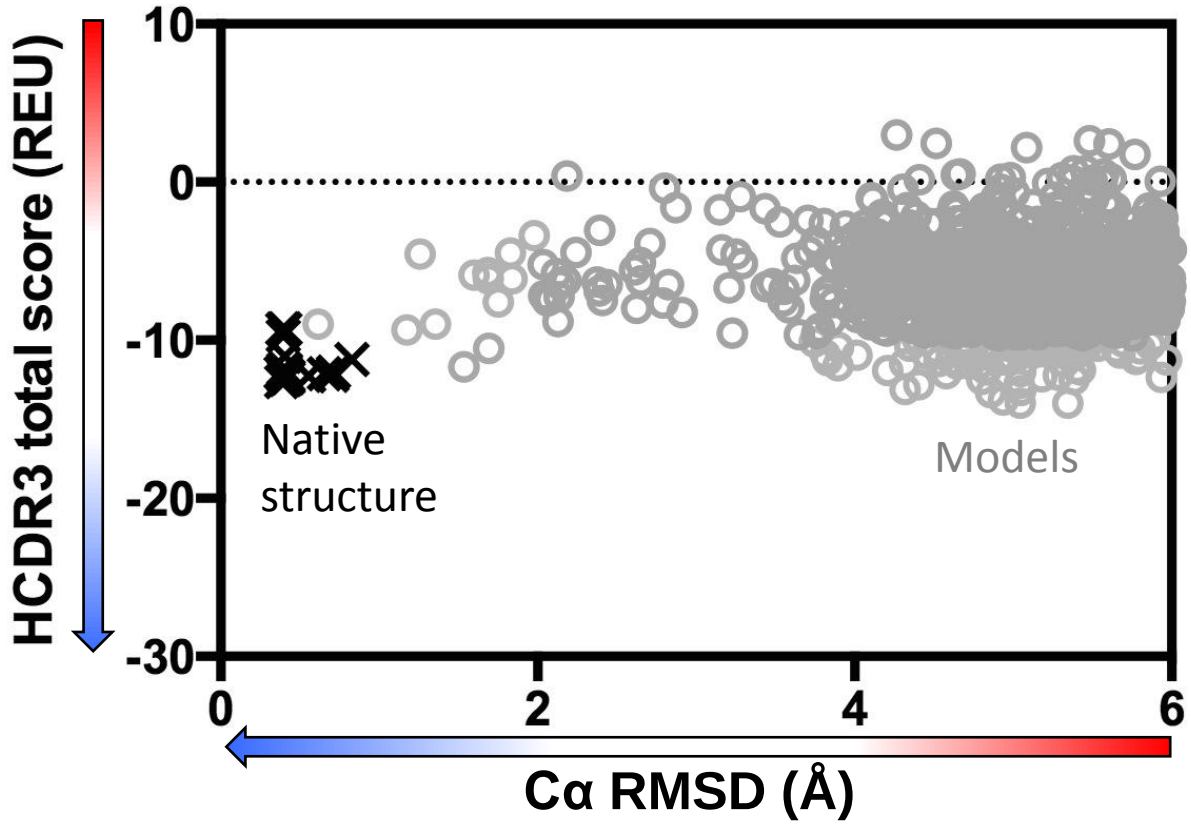
```
Dihedral C 95  N 96  CA 96  C 96  CIRCULARHARMONIC -2.5 0.2
Dihedral C 96  N 97  CA 97  C 97  CIRCULARHARMONIC -1.7 0.3
Dihedral C 97  N 98  CA 98  C 98  CIRCULARHARMONIC -1.8 0.6
Dihedral C 121 N 122 CA 122 C 122 CIRCULARHARMONIC -2.0 0.8
Dihedral C 122 N 123 CA 123 C 123 CIRCULARHARMONIC -1.7 0.5
Dihedral C 123 N 124 CA 124 C 124 CIRCULARHARMONIC -1.5 0.3
Dihedral C 124 N 125 CA 125 C 125 CIRCULARHARMONIC -2.3 0.4

Dihedral N 96  CA 96  C 96  N 97  CIRCULARHARMONIC  2.6 0.2
Dihedral N 97  CA 97  C 97  N 98  CIRCULARHARMONIC  2.5 0.2
Dihedral N 98  CA 98  C 98  N 99  CIRCULARHARMONIC  2.4 0.5
Dihedral N 122 CA 122 C 122 N 123 CIRCULARHARMONIC  2.8 0.5
Dihedral N 123 CA 123 C 123 N 124 CIRCULARHARMONIC  1.8 0.3
Dihedral N 124 CA 124 C 124 N 125 CIRCULARHARMONIC -0.6 0.5
Dihedral N 125 CA 125 C 125 N 126 CIRCULARHARMONIC  2.3 0.3
```

# Additional loop building options needed to apply restraints

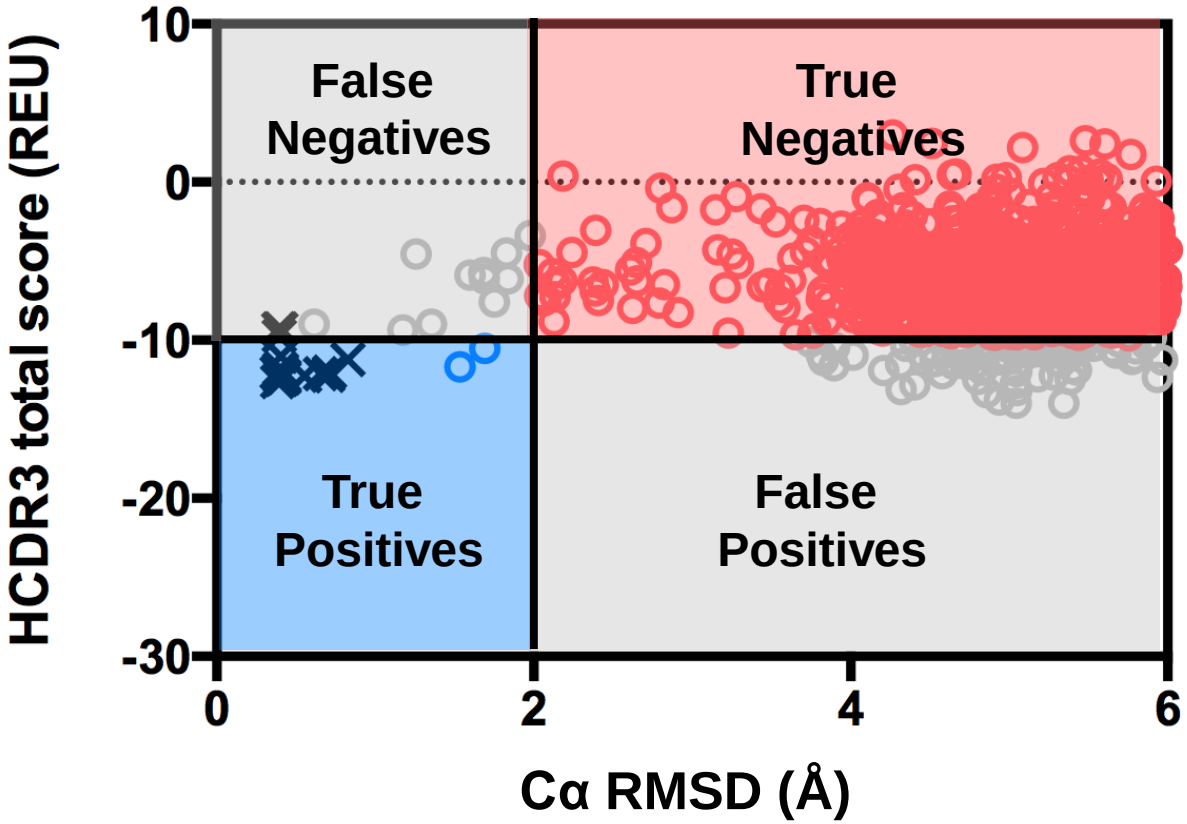
-score:weights talaris2014.wts
-score:patch dihedral_cst.wts_patch #Patch to turn on dihedral term
-constraints:cst_file 4m5y_bulged.constraints #Your restraints file
-constraints:cst_weight 1 #Weight all of the constraints by 1
-constraints:viol
-constraints:viol_level 101

# Analyzing your results

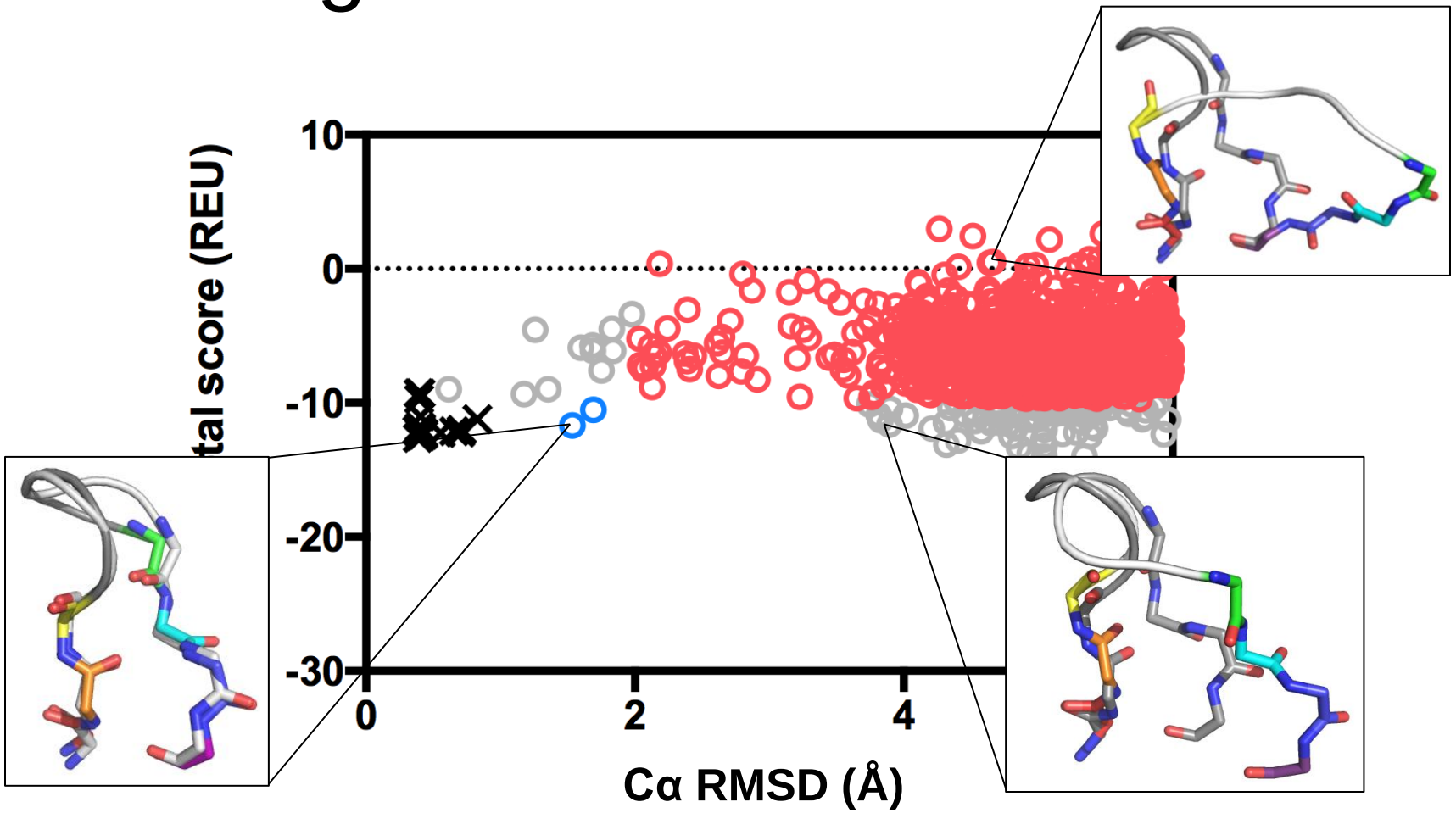
# Analyzing Rosetta HCDR3 modeling results: score vs. rmsd



# Analyzing Rosetta HCDR3 modeling results: score vs. rmsd



# Analyzing Rosetta HCDR3 modeling results: visualize it!



# References

**A new clustering of antibody CDR loop conformations.** North B, Lehmann A, Dunbrack RL. J Mol Biol. 2011.

**Improving loop modeling of the antibody complementarity-determining region 3 using knowledge-based restraints.** Finn JA, Koehler Leman J, Willis JR, Cisneros A III, Crowe JE Jr., Meiler J. PLOS ONE. 2016.

**Accurate structure prediction of CDR H3 loops enabled by a novel structure-based C-terminal constraint.** Weitzner BD, Gray JJ. J Immuno. 2017.

# Today’s tutorial

Antibody: 5J8

HCDR3 length: 17

PDB ID: 4M5Y

HCDR3 torso: Bulged