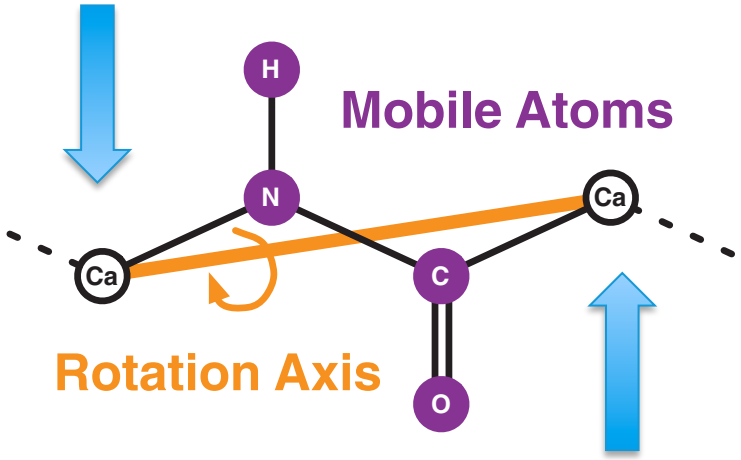
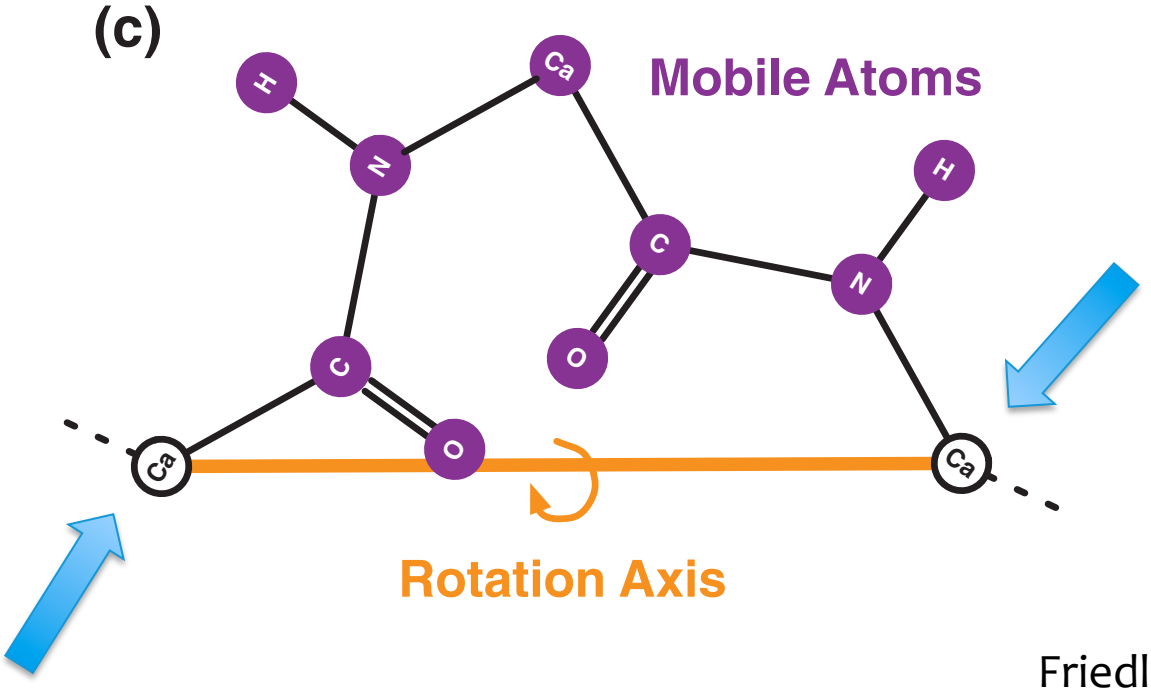
Features Analysis: Backrub Ensemble Bond Angles

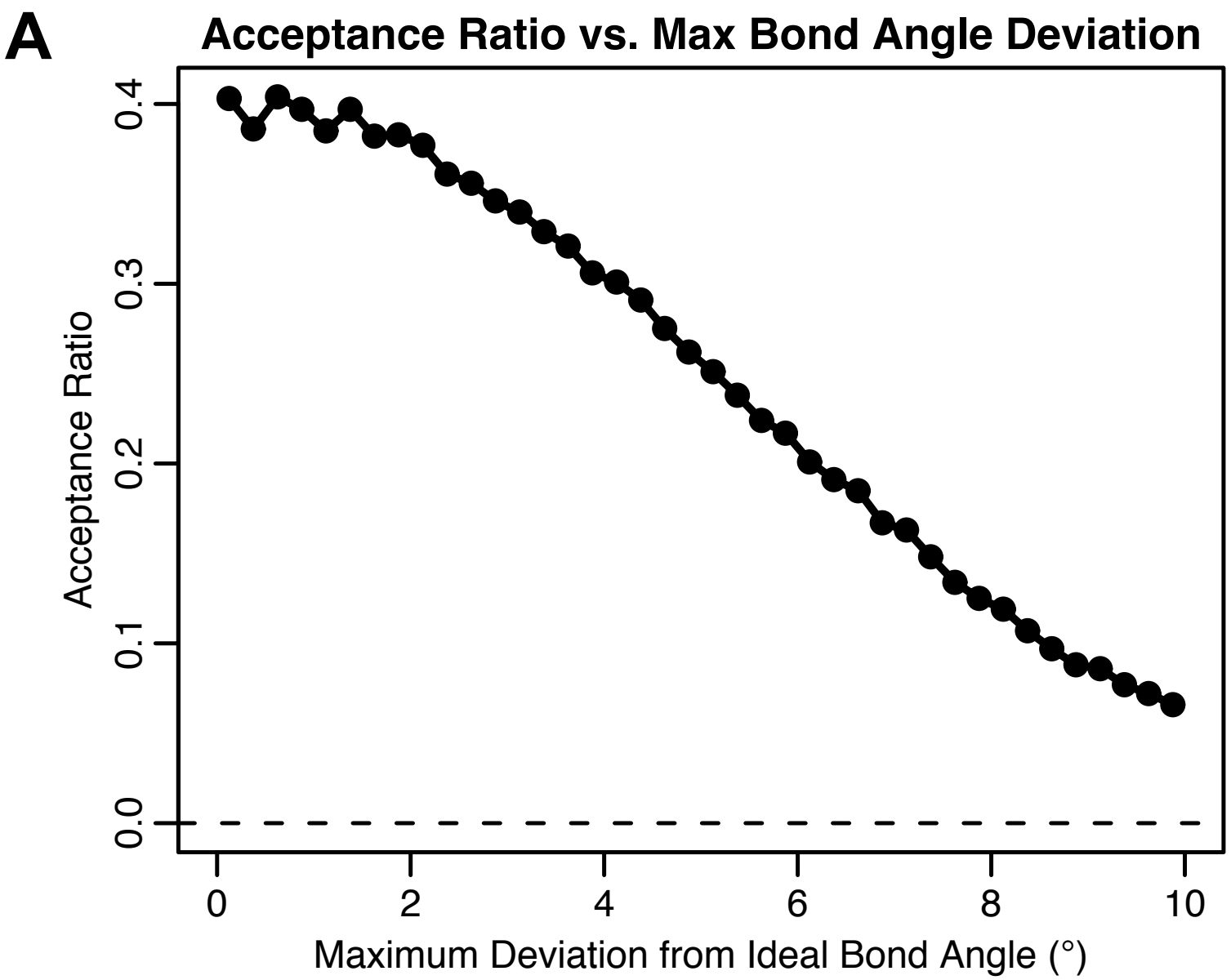
Matthew O'Meara

RosettaCon 2012



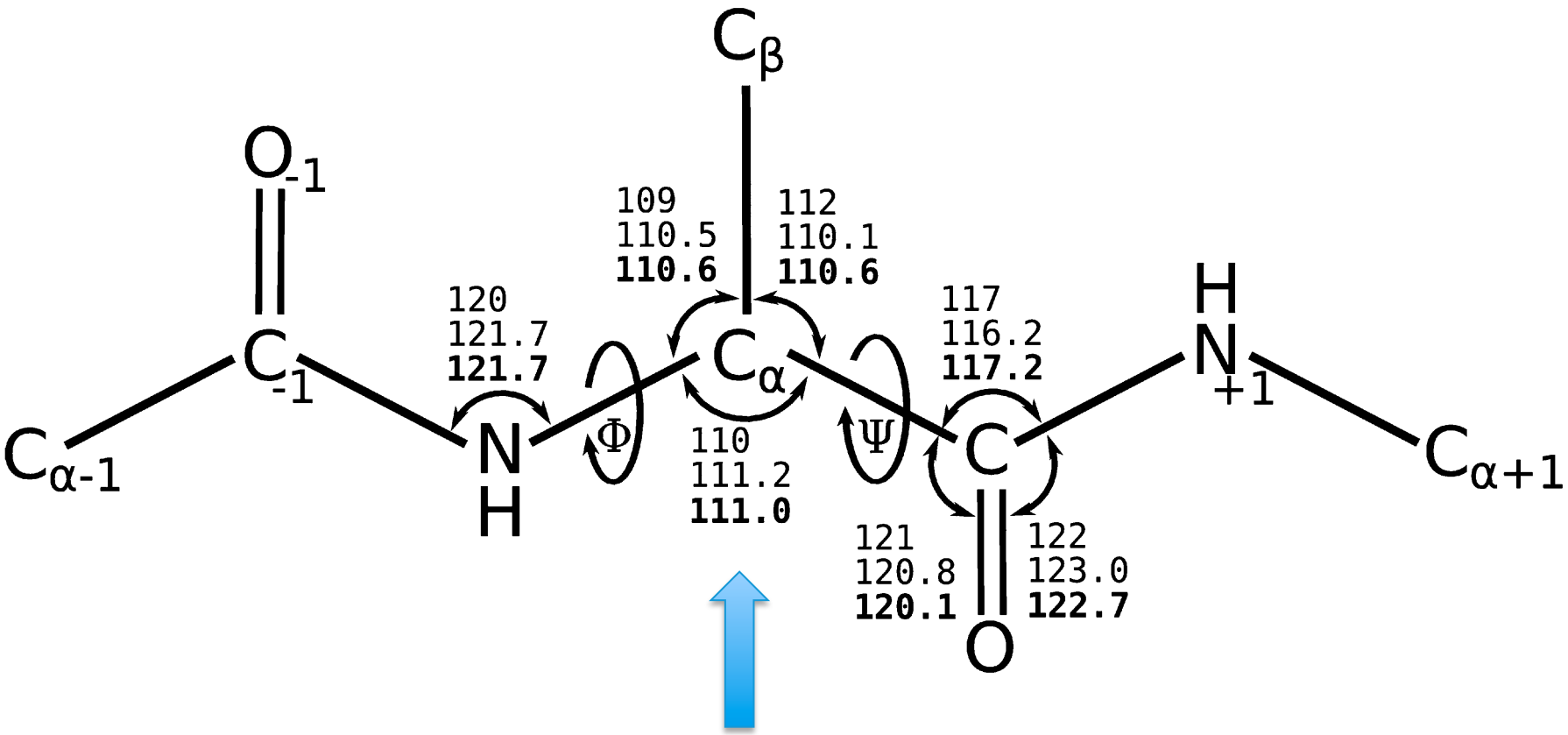


Friedland JMB 2008

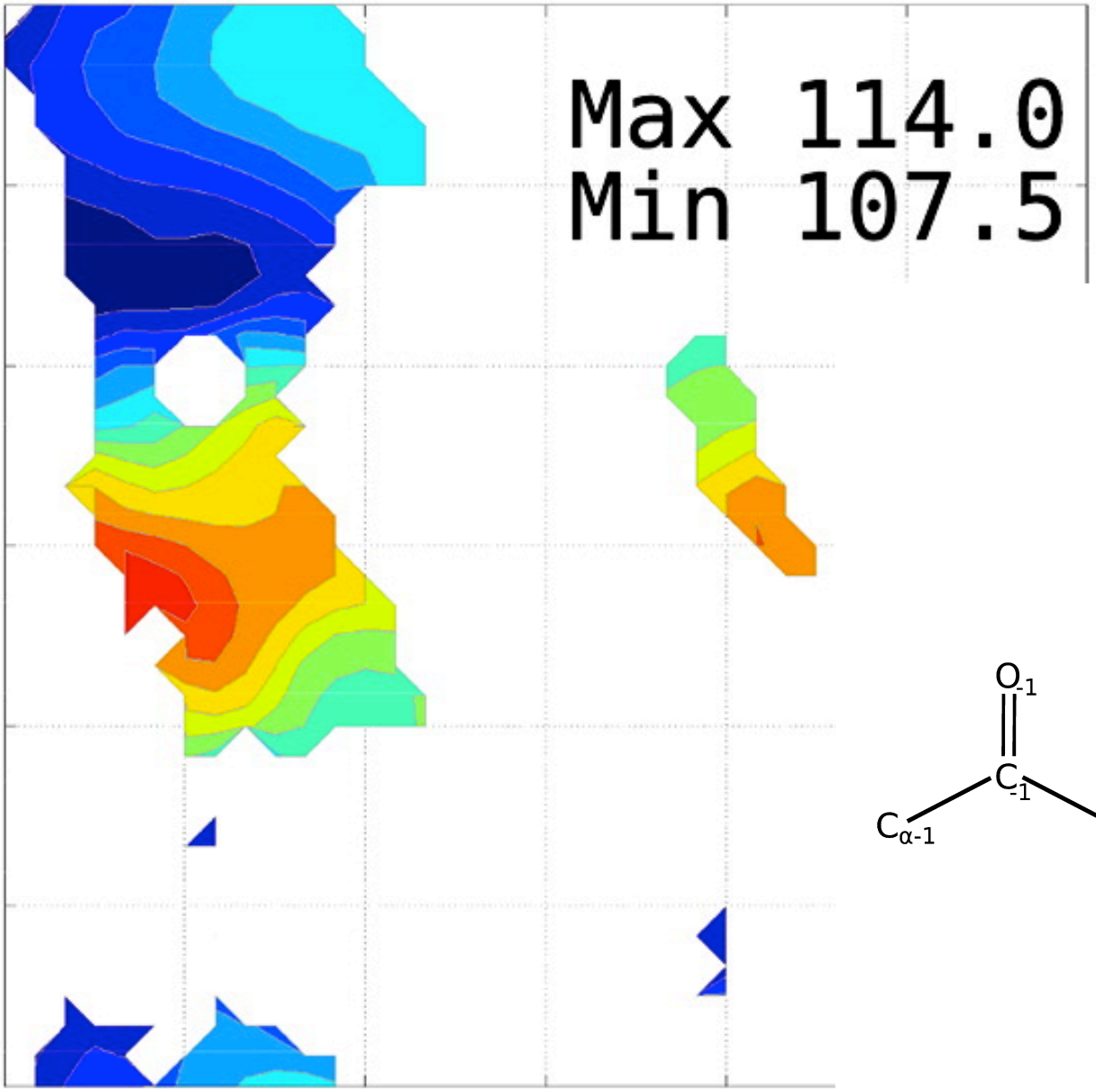


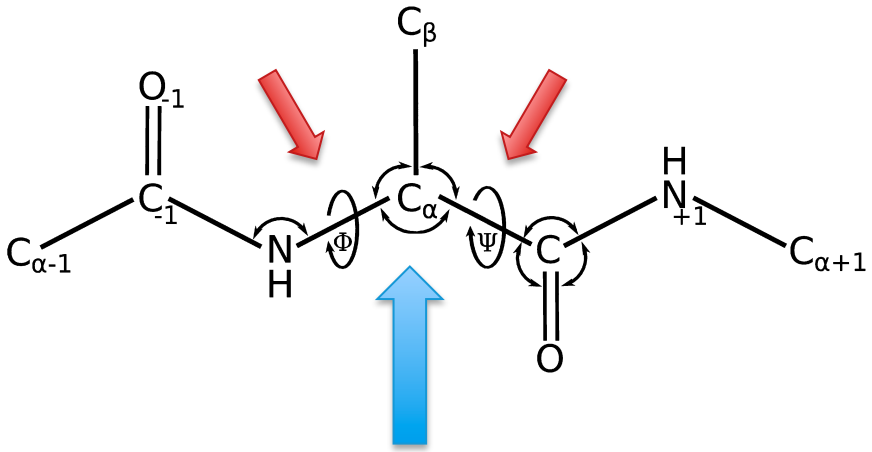
Smith JMB 2008

“Ideal” Bond Angles

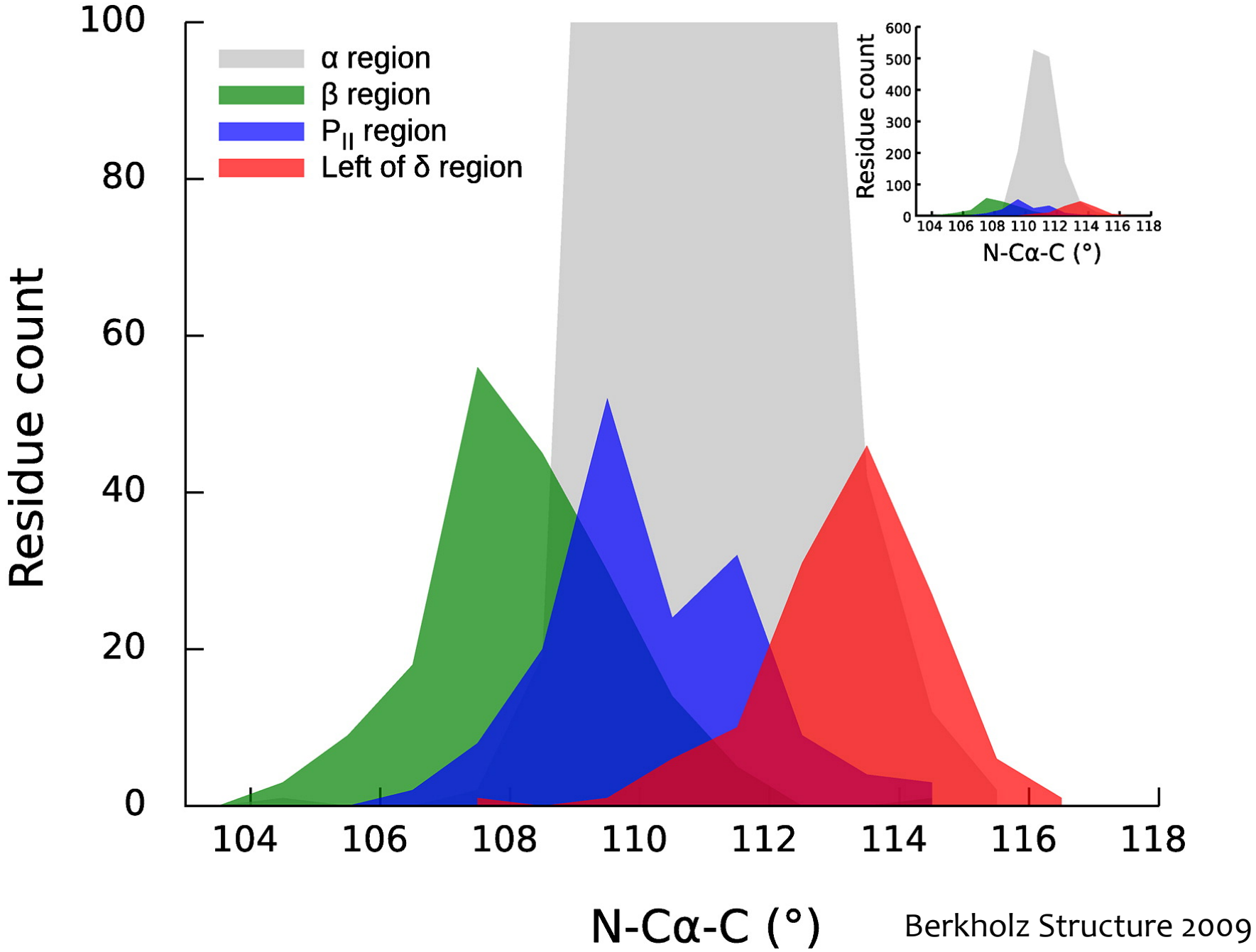


Berkholz Structure 2009



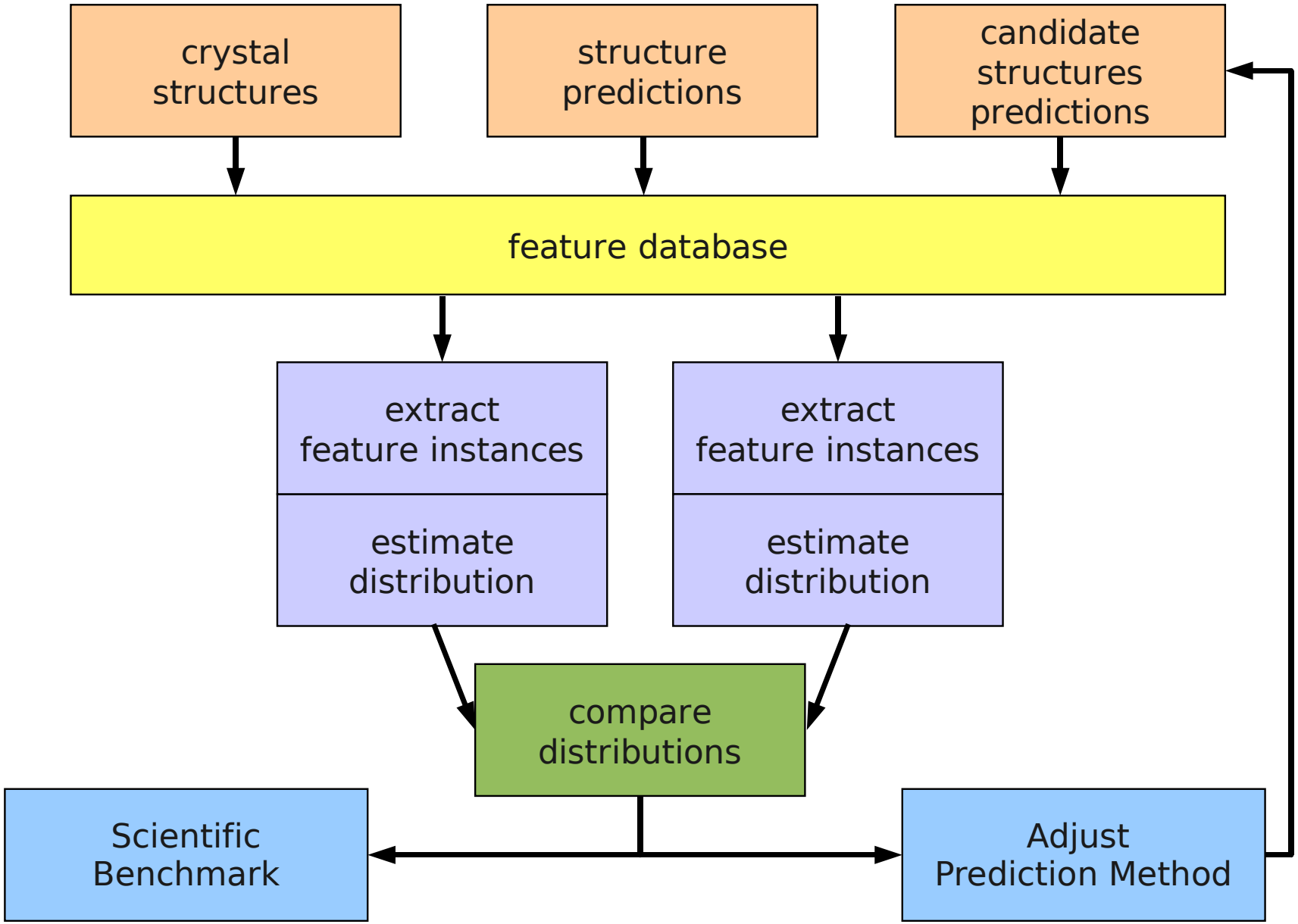


Berkholz Structure 2009



Berkholz Structure 2009

Feature-Based Workflows



Native Data

- Richardson Lab: Top8000 Chains
  - 70-seq-homology, <2A, have EDS -> 6,563 chains

Strip to relevant chain -> 1abcFH_**A**.pdb:

```
for x in \
  $(find top8000_chains_eds_70_rosetta_named_hydrogens -name "*FH_*pdb");
do
  base=${x##*/}; chain=${base:7:1};
  cat $x | grep "^ATOM.*$" | grep "^.\{21\}${chain}.*$" > \
    top8000_chains_eds_70_rosetta_named_hydrogens_single_chains/$base;
done
```

Backrub Protocol

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
    <s weights=score12_full> <Reweight scoretype=mm_bend weight=1/> </s>
  </SCOREFXNS>
  <TASKOPERATIONS>
    <RestrictToRepacking name=rtrp/> <PreserveCBeta name=preserve_cb/>
  </TASKOPERATIONS>
  <MOVERS>
     <MetropolisHastings name=mc scorefxn=s trials=10000>
  <Backrub sampling_weight=.75/>
  <Sidechain sampling_weight=0.25 task_operations=rtrp,preserve_cb/>
     </MetropolisHastings>
  </MOVERS>
  <PROTOCOLS>
    <Add mover_name=mc/>
  </PROTOCOLS>
</ROSETTASCRIPTS>
```

COMMAND LINE FLAGS:

```
-ex1 -ex2 -extrachi_cutoff 0
```

Features Reporters

```
<ReportToDB name=features_reporter database_name="features_backrub_120725.db3"
database_mode=sqlite3 database_separate_db_per_mpi_process=1
sample_source="Backrub Ensemble”>
  <feature name=ScoreTypeFeatures/>
  <feature name=StructureScoresFeatures scorefxn=s/>
  <feature name=ProteinRMSDFeatures reference_name=init_struct/>
  <feature name=RadiusOfGyrationFeatures/>
  <feature name=ResidueTypesFeatures/>
  <feature name=ResidueFeatures/>
  <feature name=PdbDataFeatures/>
  <feature name=ResidueScoresFeatures scorefxn=s/>
  <feature name=PairFeatures/>
  <feature name=ResidueBurialFeatures/>
  <feature name=ResidueSecondaryStructureFeatures/>
  <feature name=ProteinBackboneTorsionAngleFeatures/>
  <feature name=ProteinResidueConformationFeatures/>
  <feature name=ProteinBondGeometryFeatures/>
  <feature name=HBondFeatures scorefxn=s/>
  <feature name=SaltBridgeFeatures/>
</ReportToDB>
```

Analysis Script Template

```
f <- query_sample_sources(sample_sources, sql_query)

dens <- estimate_density_1d(f, id_columns, measure_column)

p <- ggplot(dens) +
  geom_line(…) +
  geom_vline(…) +
  scale_x_…(…) +
  scale_y_…(…) +
  opts(…) + theme_…()

save_plots(self, plot_id, …)
```

bond_angles.R

```
sql_query <- “
SELECT
  b_ang.ideal,
  b_ang.observed
FROM
  bond_intrares_angles AS b_ang
WHERE
  b_ang.outAtm1Num = 1 AND b_ang.cenAtmNum = 2 AND b_ang.outAtm2Num = 3;"
```

bond_angles.R

```
sql_query <- “
SELECT
  b_ang.ideal,
  b_ang.observed
FROM
  residues AS res,
  residue_pdb_confidence AS res_conf,
  bond_intrares_angles AS b_ang
WHERE
  res_conf.struct_id = res.struct_id AND
  res_conf.residue_number = res.resNum AND
  res_conf.max_temperature < 30 AND
  b_ang.struct_id = res.struct_id AND b_ang.resNum = res.resNum AND
  b_ang.outAtm1Num = 1 AND b_ang.cenAtmNum = 2 AND b_ang.outAtm2Num = 3;"
```

bond_angles.R

```
sql_query <- "
SELECT
  res.name3 AS res_type,
  dssp_code.label AS dssp_label,
  b_ang.ideal,
  b_ang.observed
FROM
  residues AS res,
  residue_pdb_confidence AS res_conf,
  residue_secondary_structure AS ss,
  dssp_codes AS dssp_code,
  bond_intrares_angles AS b_ang
WHERE
  res_conf.struct_id = res.struct_id AND res_conf.residue_number = res.resNum AND
  res_conf.max_temperature < 30 AND
  ss.struct_id = res.struct_id AND ss.resNum == res.resNum AND
  dssp_code.code = ss.code AND
  b_ang.struct_id = res.struct_id AND b_ang.resNum = res.resNum AND
  b_ang.outAtm1Num = 1 AND b_ang.cenAtmNum = 2 AND b_ang.outAtm2Num = 3;"
```

bond_angles.R

backbone_geometry_bond_angle_NCaC

Layers:

- geom_line: data=dens
  - x=x, y=y, color=sample_source
- geom_vline: data=f
  - x=ideal
- geom_indicator: data=dens
  - indicator=counts
  - color=sample_source

Analysis Configuration

```
{ "sample_source_comparisons" : [{
    "sample_sources" : [{
      "database_path" : "path/features_top8000_r50086.db3",
      "id" : "top8000",
    }, {
      "database_path" : "path/features_top8000_backrub_r50086.db3",
      "id" : "top8000_backrub",
    }],
  "analysis_scripts" : [
    "scripts/analysis/plots/backbone_geometry/bond_angles.R"
  ],
  "output_dir" : "build”,
  "output_formats" : [
    "output_slide_pdf"]}]}
```

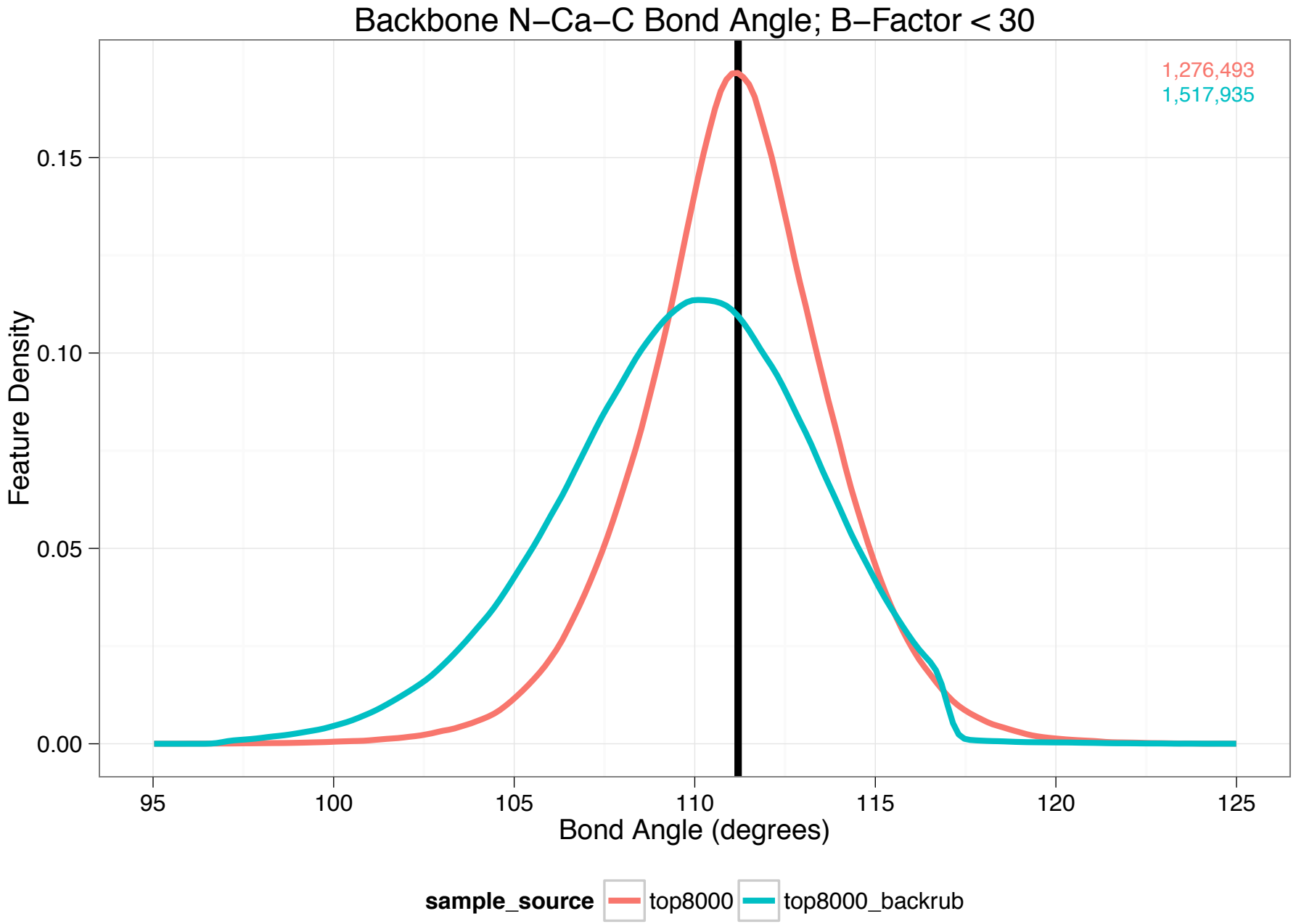
Features Analysis Output

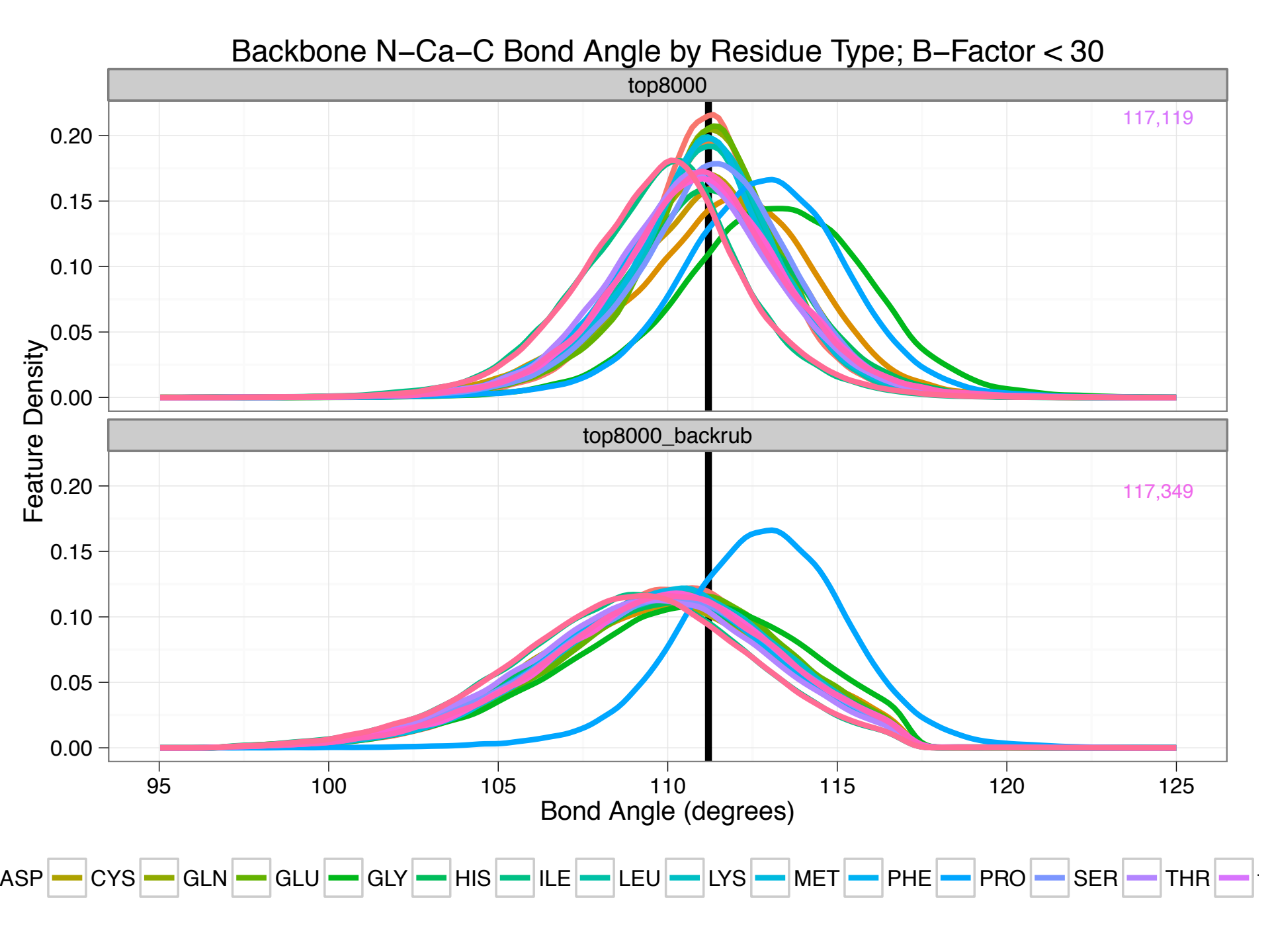
```
$~/rosetta/rosetta/rosetta_tests/features/compare_sample_sources.R --
config analysis_configurations/bond_angles.json
Sample Source Comparison:
  Output Directory <- 'path/build/top8000_top8000_backrub'
    Output Formats <- output_slide_pdf

  Sample Sources:
  top8000 <- path/features_top8000_r50086.db3
  top8000_backrub <- path/features_top8000_backrub_r50086.db3

  Analysis_scripts:
  scripts/analysis/plots/backbone_geometry/bond_angles.R

Features Analysis: bond_angles
loading: top8000 ... 23.33 s
loading: top8000_backrub ... 22.55 s
Saving Plot:  path/build/top8000_top8000_backrub/bond_angles/
output_slide_pdf/backbone_geometry_bond_angle_NCaC_120727.pdf ... 0.33s
```

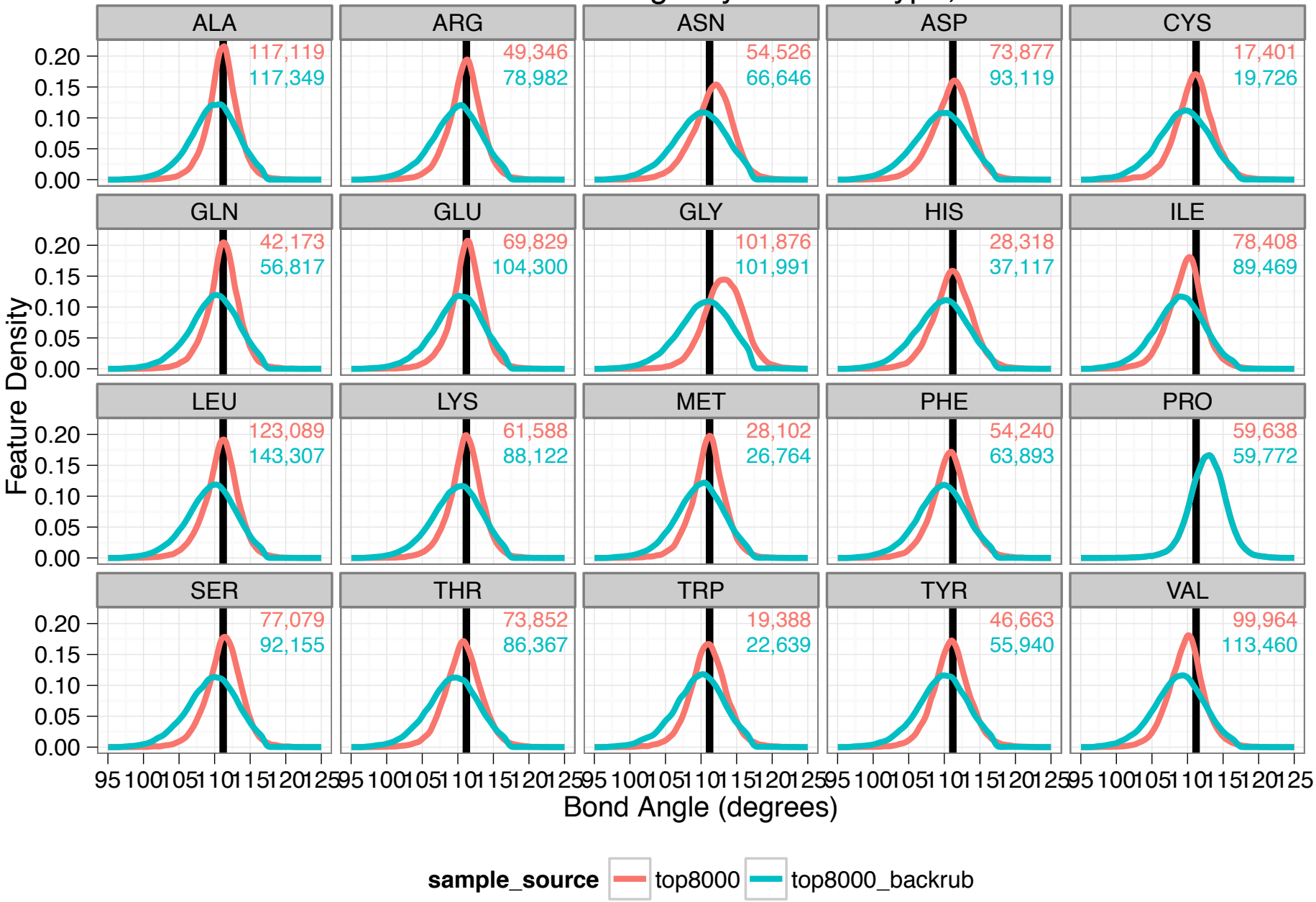
N-Ca-C Backrub vs Native

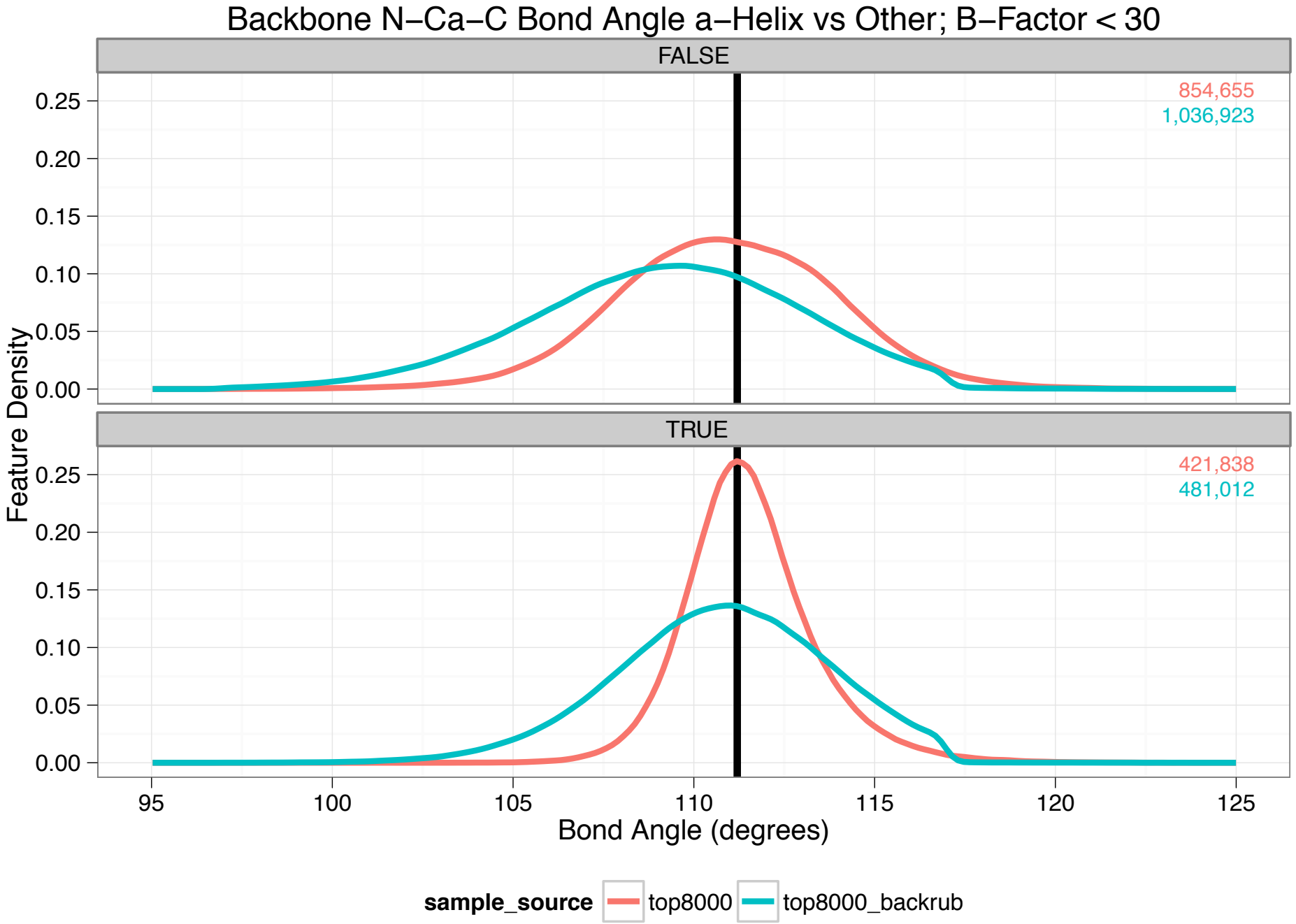
- Observations
  - Non-helix, mean shifted tighter
  - Larger standard deviation
  - Recapitulate secondary structure variation
  - Angle restriction at 117
- Questions
  - Does folded structures bias towards tighter angles?
  - Does backrub uniformly sample secondary structure?

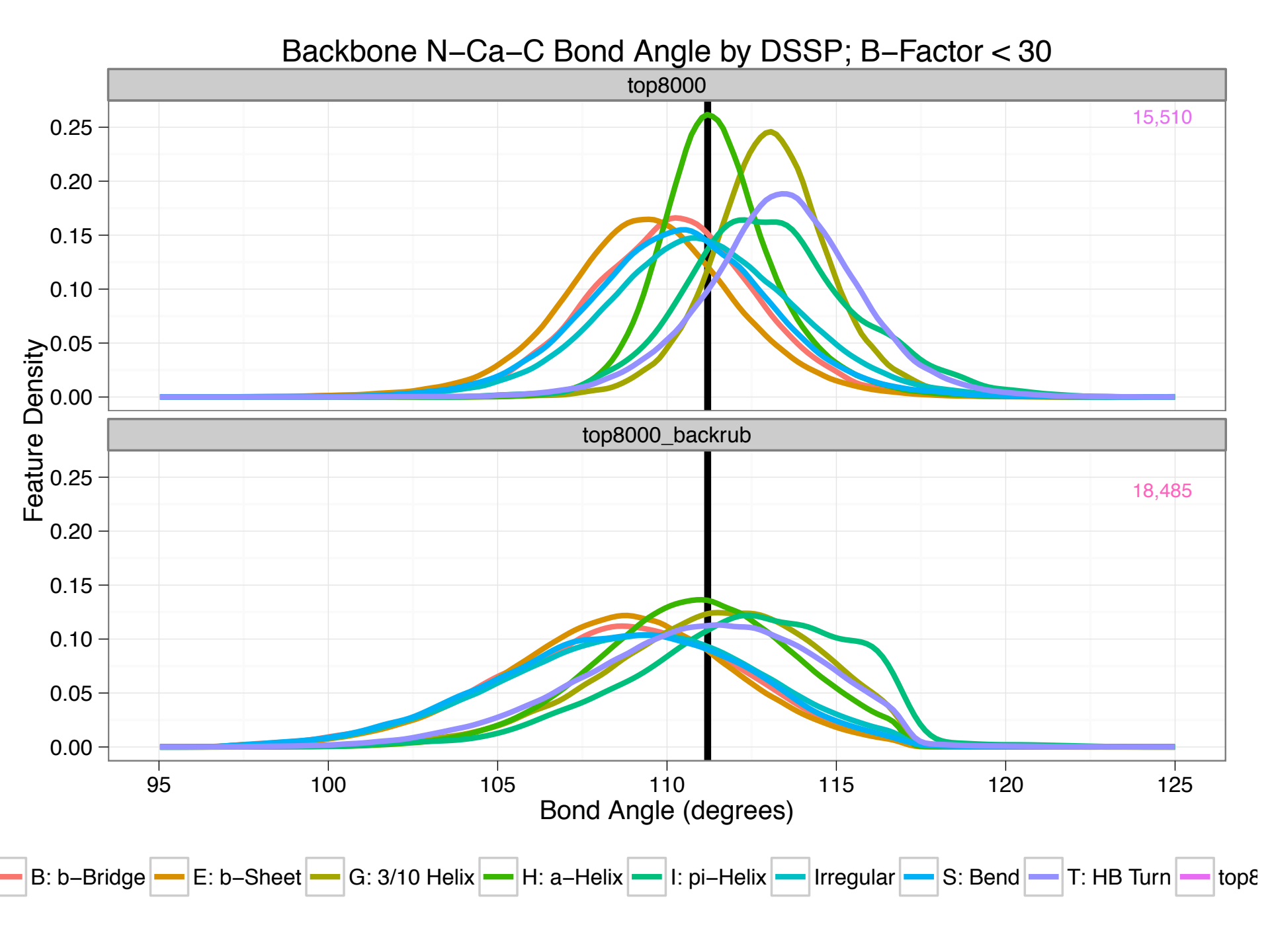
  - Backrub:
    - More variation in

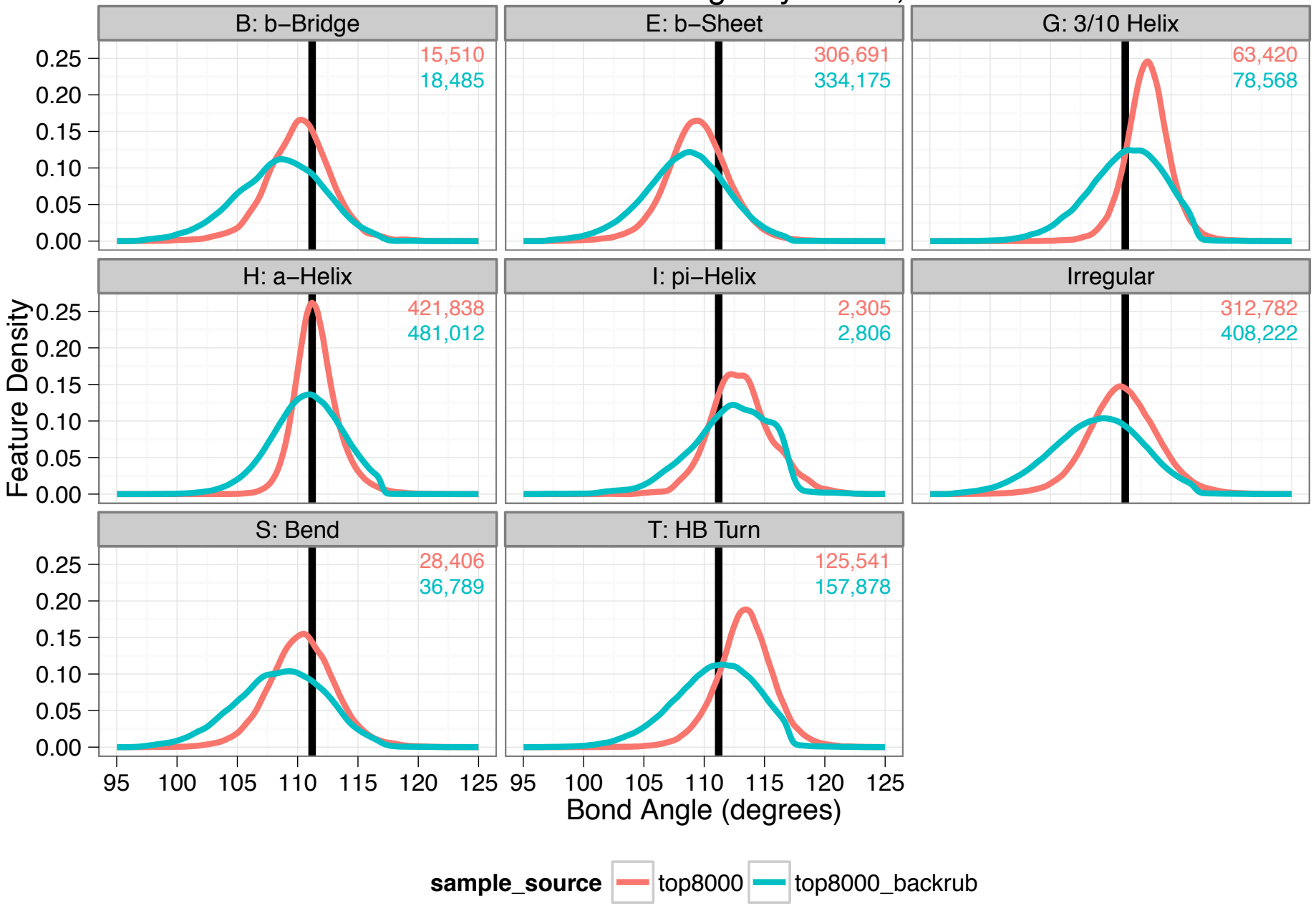

Backbone N–Ca–C Bond Angle; B–Factor < 30
1,276,493
1,517,935
0.15
0.10
0.05
0.00
Feature Density
95
100
105
110
115
120
125
Bond Angle (degrees)
sample_source
top8000
top8000_backrub

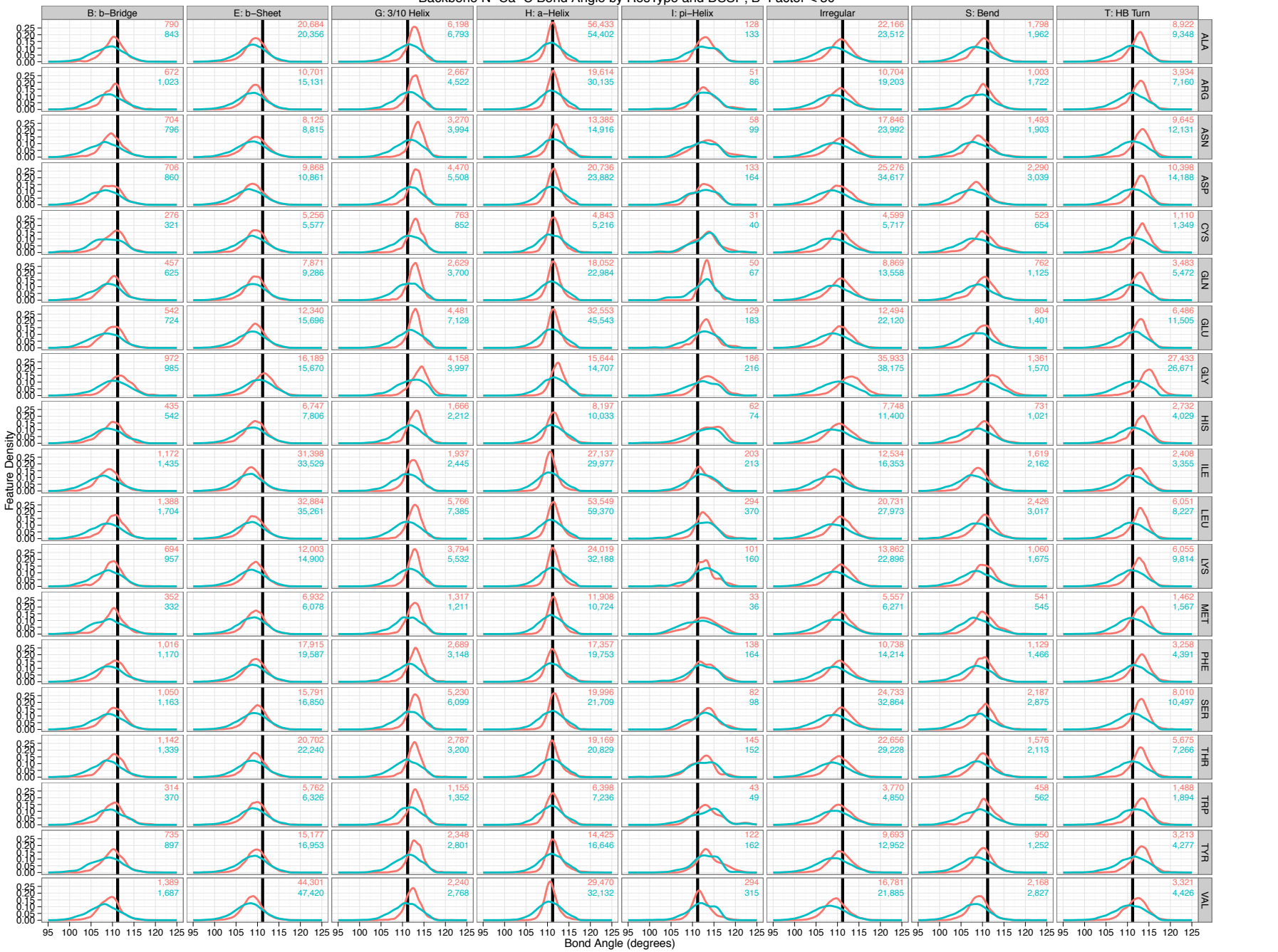

Backbone N–Ca–C Bond Angle by Residue Type; B–Factor < 30
top8000
117,119
top8000_backrub
117,349
Feature Density
0.00
0.05
0.10
0.15
0.20
95
100
105
110
115
120
125
Bond Angle (degrees)
ASP
CYS
GLN
GLU
GLY
HIS
ILE
LEU
LYS
MET
PHE
PRO
SER
THR

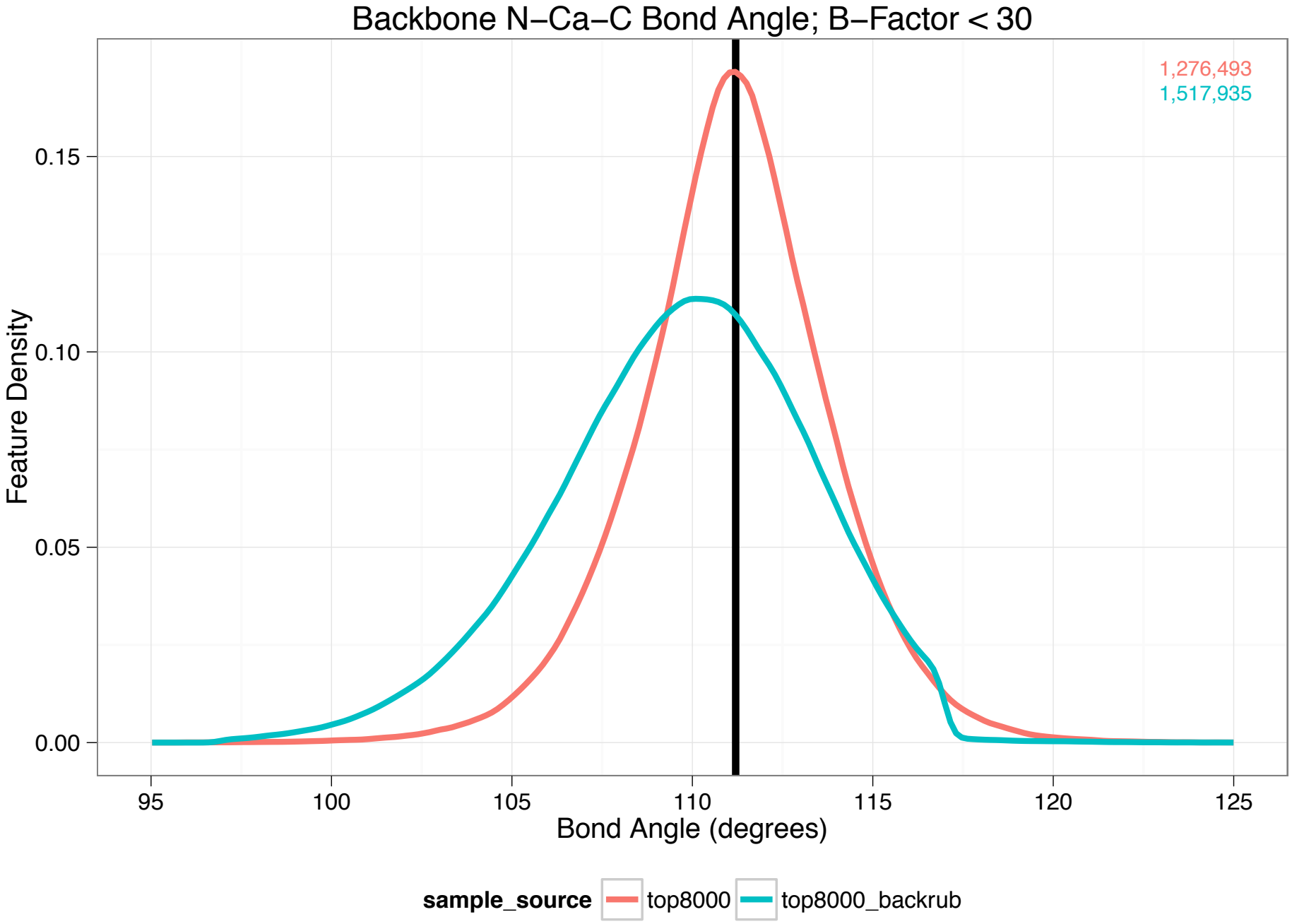

Backbone N–Ca–C Bond Angle by Residue Type; B–Factor < 30
ALA
117,119
117,349
ARG
49,346
78,982
ASN
54,526
66,646
ASP
73,877
93,119
CYS
17,401
19,726
GLN
42,173
56,817
GLU
69,829
104,300
GLY
101,876
101,991
HIS
28,318
37,117
ILE
78,408
89,469
LEU
123,089
143,307
LYS
61,588
88,122
MET
28,102
26,764
PHE
54,240
63,893
PRO
59,638
59,772
SER
77,079
92,155
THR
73,852
86,367
TRP
19,388
22,639
TYR
46,663
55,940
VAL
99,964
113,460
Feature Density
0.00
0.05
0.10
0.15
0.20
95 100 105 110 115 120 125
Bond Angle (degrees)
sample_source
top8000
top8000_backrub


Backbone N–Ca–C Bond Angle a–Helix vs Other; B–Factor < 30
FALSE
TRUE
854,655
1,036,923
421,838
481,012
Feature Density
0.25
0.20
0.15
0.10
0.05
0.00
95
100
105
110
115
120
125
Bond Angle (degrees)
sample_source
top8000
top8000_backrub

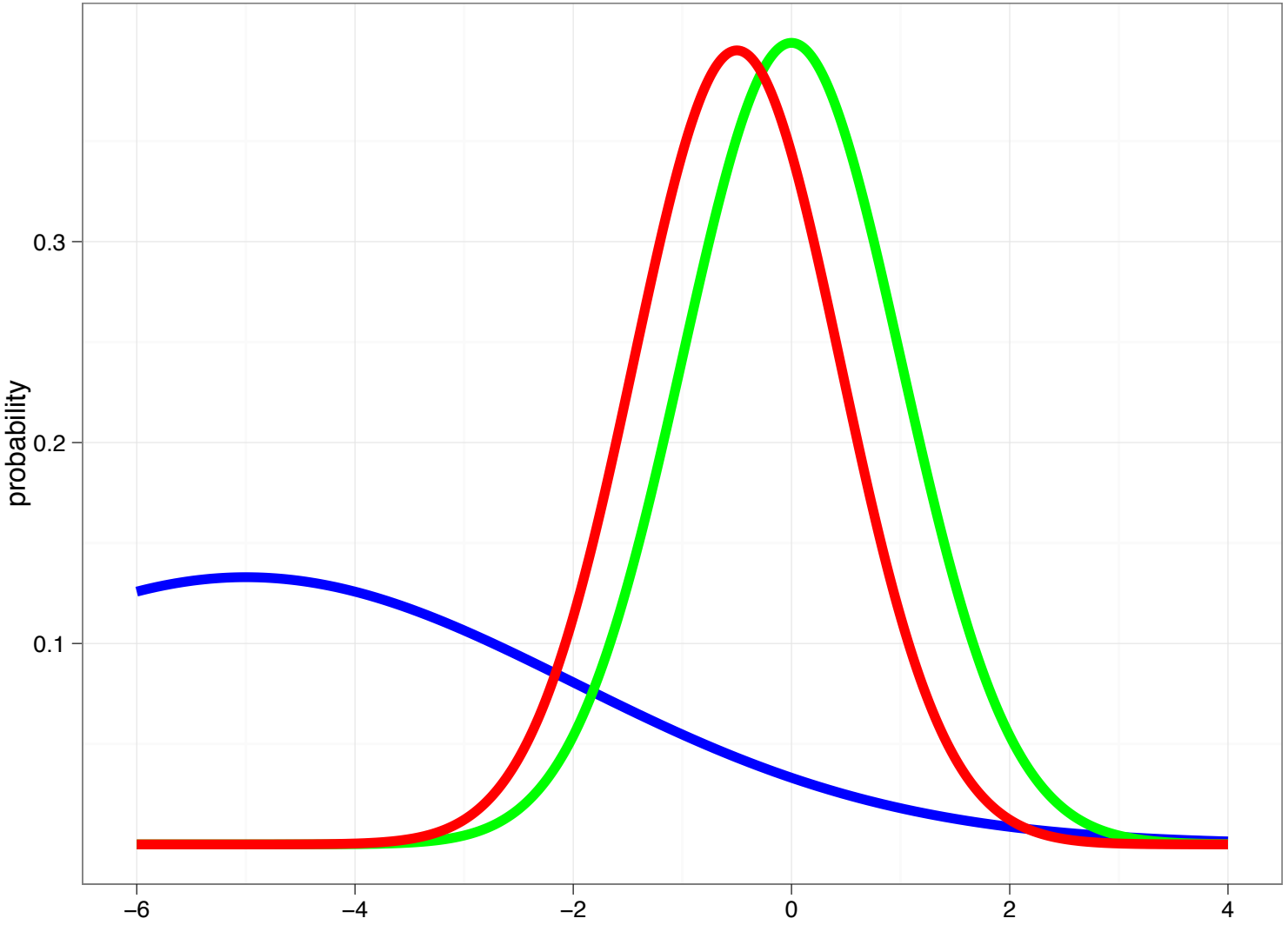

Backbone N–Ca–C Bond Angle by DSSP; B–Factor < 30
top8000
15,510
top8000_backrub
18,485
Feature Density
0.25
0.20
0.15
0.10
0.05
0.00
95
100
105
110
115
120
125
Bond Angle (degrees)
B: b–Bridge
E: b–Sheet
G: 3/10 Helix
H: a–Helix
I: pi–Helix
Irregular
S: Bend
T: HB Turn
top8

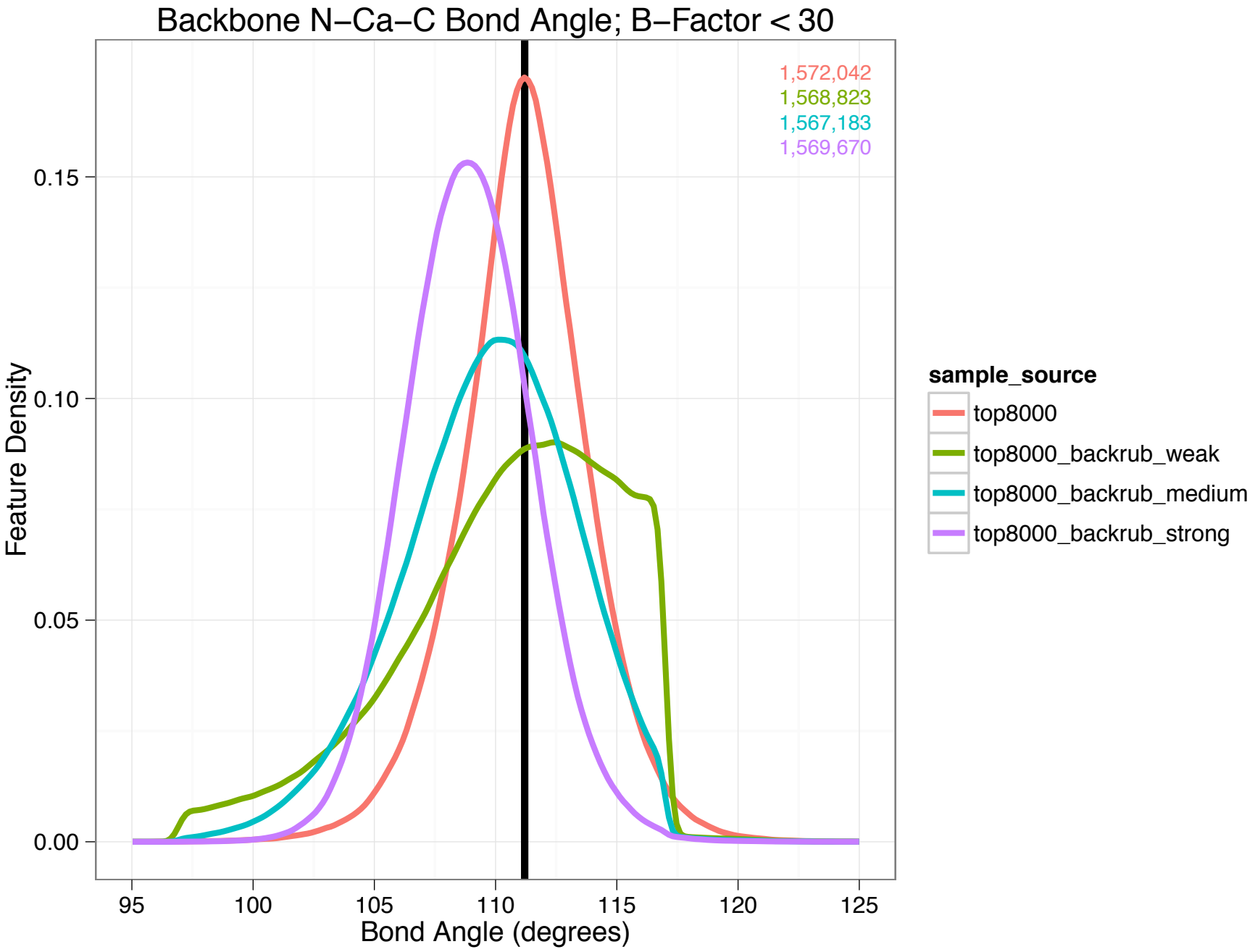

Backbone N–Ca–C Bond Angle by DSSP; B–Factor < 30
B: b–Bridge
15,510
18,485
E: b–Sheet
306,691
334,175
G: 3/10 Helix
63,420
78,568
H: a–Helix
421,838
481,012
I: pi–Helix
2,305
2,806
Irregular
312,782
408,222
S: Bend
28,406
36,789
T: HB Turn
125,541
157,878
Feature Density
0.25
0.20
0.15
0.10
0.05
0.00
95
100
105
110
115
120
125
Bond Angle (degrees)
sample_source
top8000
top8000_backrub

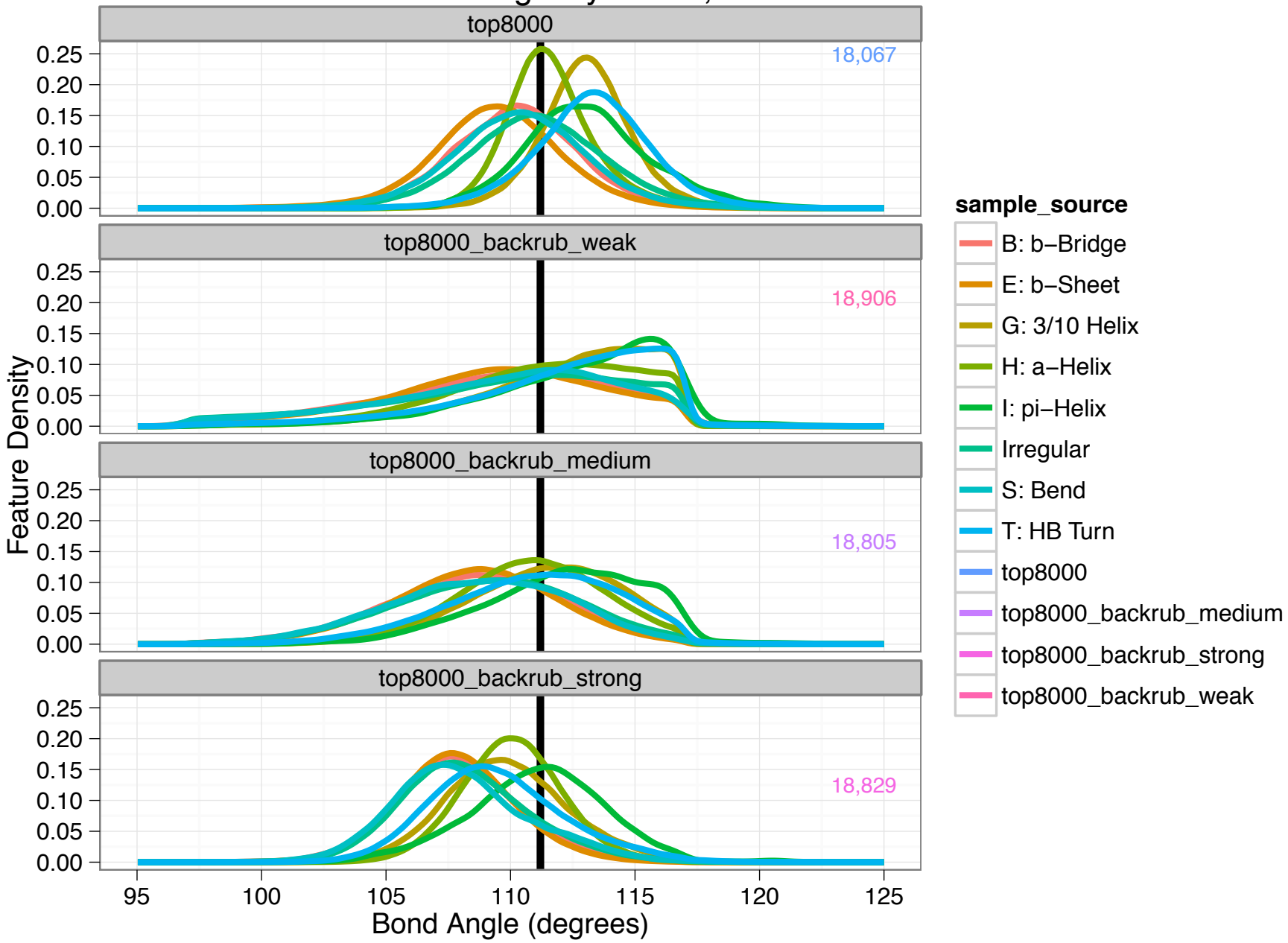

Backbone N–Ca–C Bond Angle by ResType and DSSP; B–Factor < 30
B: b–Bridge
E: b–Sheet
G: 3/10 Helix
H: a–Helix
I: pi–Helix
Irregular
S: Bend
T: HB Turn
ALA
ARG
ASN
ASP
CYS
GLN
GLU
GLY
HIS
ILE
LEU
LYS
MET
PHE
SER
THR
TRP
TYR
VAL
Feature Density
Bond Angle (degrees)
sample_source
top8000
top8000_backrub

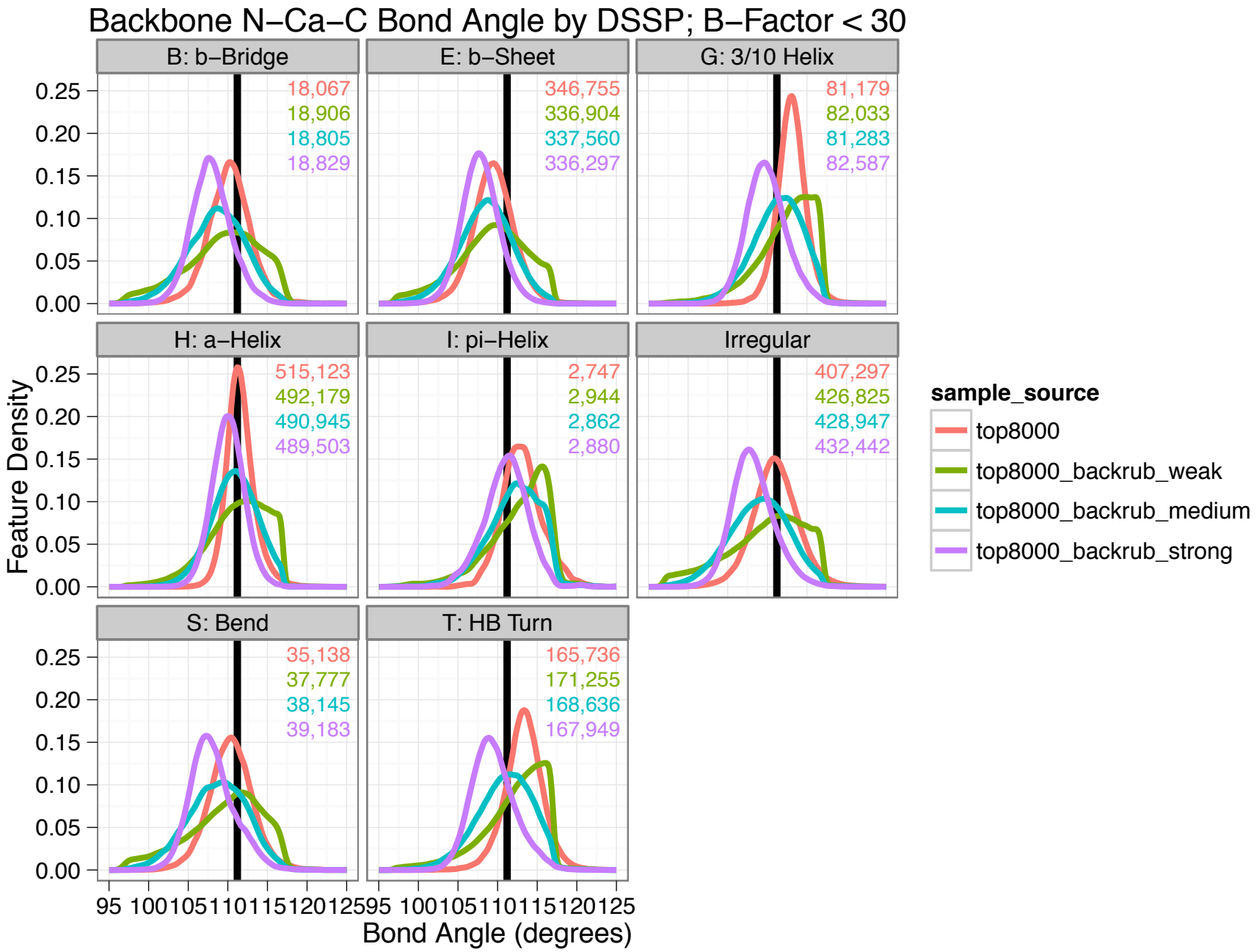

Backbone N–Ca–C Bond Angle; B–Factor < 30
1,276,493
1,517,935
Feature Density
0.15
0.10
0.05
0.00
95
100
105
110
115
120
125
Bond Angle (degrees)
sample_source
top8000
top8000_backrub

Prior Distribution on Bond Angles?




Backbone N–Ca–C Bond Angle; B–Factor < 30
1,572,042
1,568,823
1,567,183
1,569,670
Feature Density
0.15
0.10
0.05
0.00
95
100
105
110
115
120
125
Bond Angle (degrees)
sample_source
top8000
top8000_backrub_weak
top8000_backrub_medium
top8000_backrub_strong


Backbone N–Ca–C Bond Angle by DSSP; B–Factor < 30
top8000
top8000_backrub_weak
top8000_backrub_medium
top8000_backrub_strong
18,067
18,906
18,805
18,829
Feature Density
0.25
0.20
0.15
0.10
0.05
0.00
95
100
105
110
115
120
125
Bond Angle (degrees)
sample_source
B: b–Bridge
E: b–Sheet
G: 3/10 Helix
H: a–Helix
I: pi–Helix
Irregular
S: Bend
T: HB Turn
top8000
top8000_backrub_medium
top8000_backrub_strong
top8000_backrub_weak


Backbone N–Ca–C Bond Angle by DSSP; B–Factor < 30
B: b–Bridge
18,067
18,906
18,805
18,829
E: b–Sheet
346,755
336,904
337,560
336,297
G: 3/10 Helix
81,179
82,033
81,283
82,587
H: a–Helix
515,123
492,179
490,945
489,503
I: pi–Helix
2,747
2,944
2,862
2,880
Irregular
407,297
426,825
428,947
432,442
S: Bend
35,138
37,777
38,145
39,183
T: HB Turn
165,736
171,255
168,636
167,949
Feature Density
0.25
0.20
0.15
0.10
0.05
0.00
95 100105110115120125
Bond Angle (degrees)
sample_source
top8000
top8000_backrub_weak
top8000_backrub_medium
top8000_backrub_strong

Thanks

- Brian Kuhlman / Jack Snoeyink (advisors)
- Sam Deluca, Tim Jacobs (database support)
- Andrew Leaver-Fay, Steven Combs (Features Beta testers)
- Colin Smith, Frank DiMaio, Patrick Conway (Bond Angle Advice)
- Rosetta Community Community

- Friedland, Linares, Smith Kortemme, **A Simple Model of Backbone Flexibility Improves Modeling of Side-chain Conformational Variability**, JMB 2008
- Smith, Kortemme , **Backrub-Like Backbone Simulation Recapitulates Natural Protein Conformational Variability and Improves Mutant Side-Chain Prediction**, JMB 2008
- Berkholz, Shapovalov, Dunbrack Jr., Karplus, **Conformation Dependence of Backbone Geometry in Proteins**, Structure 2009