Advanced Building

SASSIE CCP-SAS Workshop

January 23-25, 2017

ISIS Neutron and Muon Source
Rutherford Appleton Laboratory UK

EPSRC
Engineering and Physical Sciences Research Council

NSF

Barriers: BUILD; EQUILIBRATE; PROPAGATE; ANALYZE

What software package(s) and force-fields do I use?

Starting structure?

How do I clean up the structure?

How do I set up a trajectory (time or space)?

How do I calculate scattering observables correctly?

EPSRC
Engineering and Physical Sciences Research Council

NSF

Overview

Intermediate Building

Building coordinates for missing atoms/residues

Dealing with domain / domain orientation
SASSIE-web Build
A trick for topological issues

Advanced Building

FF Development (not in list of “RESI”)

CGenFF (https://cgenff.paramchem.org)

FFTK (http://www.ks.uiuc.edu/Research/vmd/plugins/fftk/)

CHARMM-GUI (http://charmm-gui.org)

Antechamber (Amber)

EPSRC
Engineering and Physical Sciences Research Council

NSF

Overview

Missing bits

Typical scenarios:

N-terminus & C-terminus missing (random or cloning artifacts)
Single amino acid missing (or small loop)
Large internal loop missing

Two simple solutions:

(1) Use psfgen to add internal coordinates to fill gaps & then minimize

(2) Use CHARMM (MD) to add internal coordinates & minimize

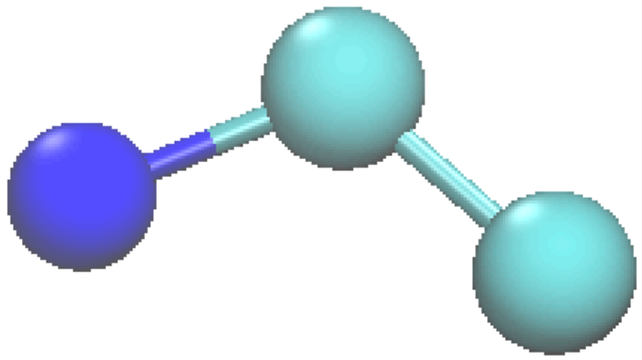
(3) ROSETTA, Modeller ($UCSF$), . . .

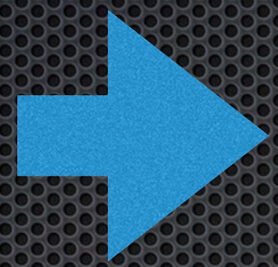
Internal loops: one could run some “simulated annealing” and/or Torsion Angle MD dynamics to sample configurations to relax large manually inserted loops.

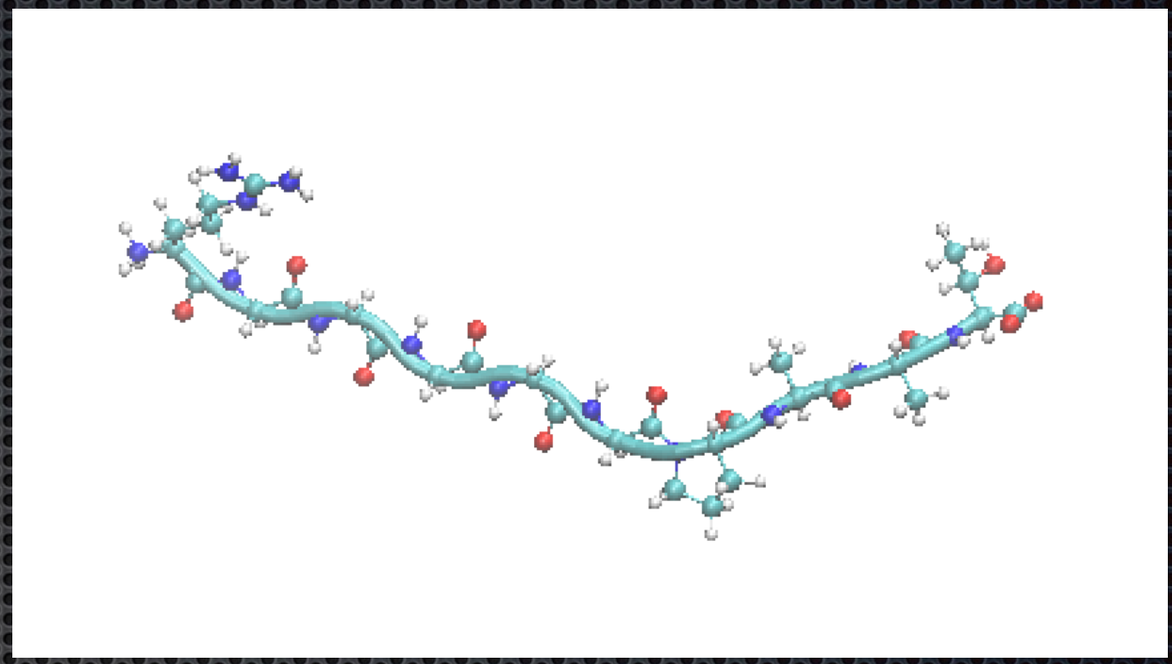
EPSRC
Engineering and Physical Sciences Research Council

NSF

missing bits







linear peptide from three atoms

```
ATOM      1  N   ARG X   1      -21.525 -67.562  86.759  1.00  0.00      PEP  N
ATOM      2  CA  ARG X   1      -21.725 -66.910  85.457  1.00  0.00      PEP  C
ATOM      3  C   ARG X   1      -23.103 -66.411  85.215  1.00  0.00      PEP  C
```

EPSRC
Engineering and Physical Sciences Research Council

NSF

PSFGEN to build missing bits

linear peptide from three atoms

```
ATOM      1  N   ARG X   1      -21.525 -67.562  86.759  1.00  0.00      PEP  N
ATOM      2  CA  ARG X   1      -21.725 -66.910  85.457  1.00  0.00      PEP  C
ATOM      3  C   ARG X   1      -23.103 -66.411  85.215  1.00  0.00      PEP  C
```

```
topology /usr/local/bin/toppar/top_all27_prot_na.inp

alias residue HIS HSE
alias atom ILE CD1 CD
alias atom SER HG HG1
alias atom CYS HG HG1

segment PEP {
 first NTER
 pdb output_building/peptide_sequence_from_fasta.pdb
 last CTER
}

coordpdb output_building/peptide_stub.pdb PEP

guesscoord

writepsf output_building/linear_peptide.psf

writepdb output_building/linear_peptide.pdb
```

sequence file only needs one atom per residue

path to pdb file (sequence)

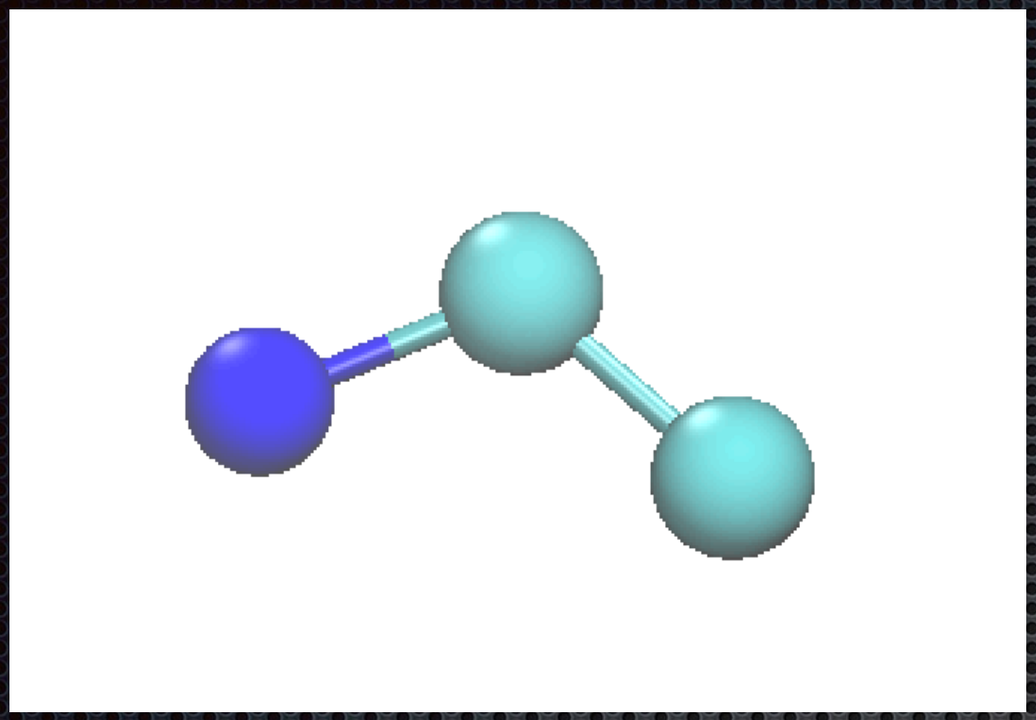
path coordinates file (pdb)

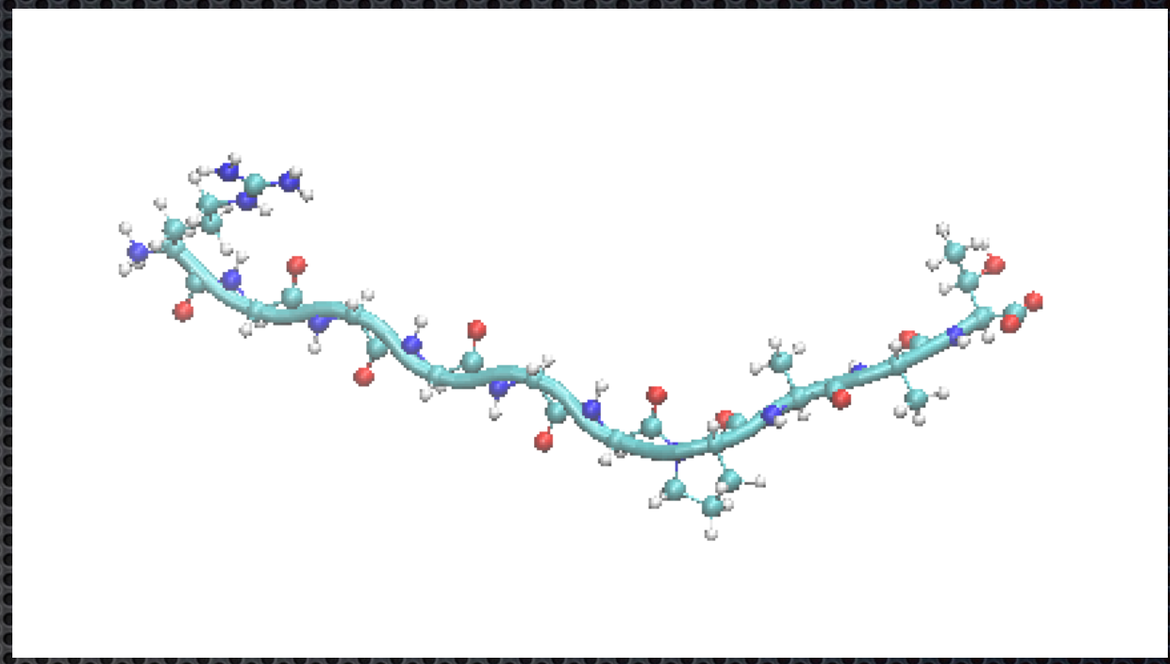
use IC to add H etc.

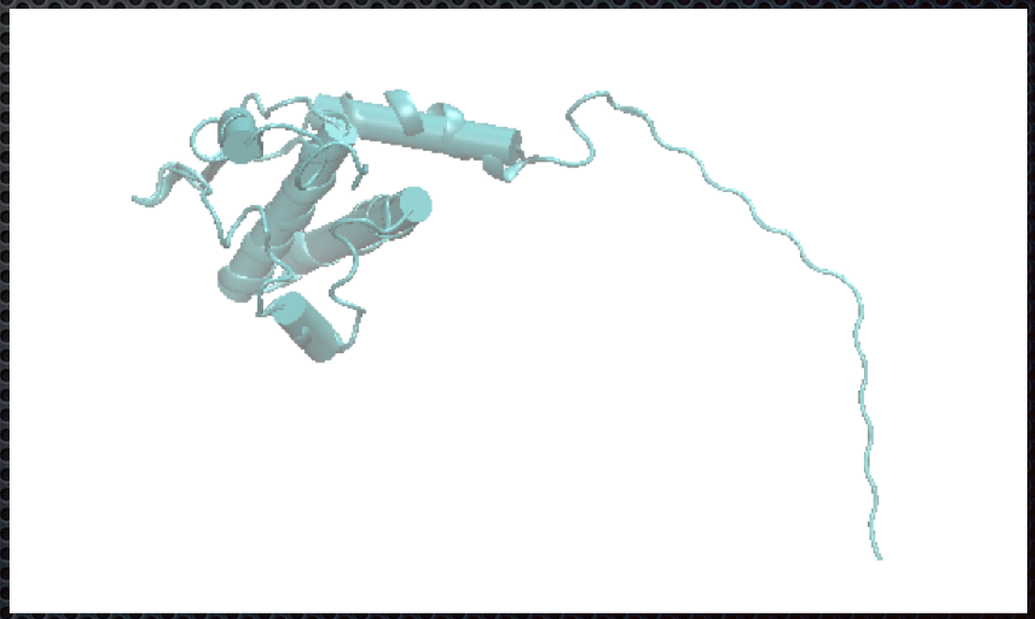
EPSRC
Engineering and Physical Sciences Research Council

NSF

missing bits







EPSRC
Engineering and Physical Sciences Research Council

NSF

Toy system

a piece of HIV1 Gag Matrix

“missing” residues 21-24 (but we know the sequence)

```
ATOM    314  C   LEU X  20      -9.122 -84.992  81.008  1.00  0.00       GAG  C
ATOM    315  O   LEU X  20      -9.647 -85.603  80.079  1.00  0.00       GAG  O
ATOM    361  N   GLY X  24      -5.700 -90.839  86.003  1.00  0.00       GAG  N
ATOM    362  HN  GLY X  24      -5.897 -91.523  85.304  1.00  0.00       GAG  H
```

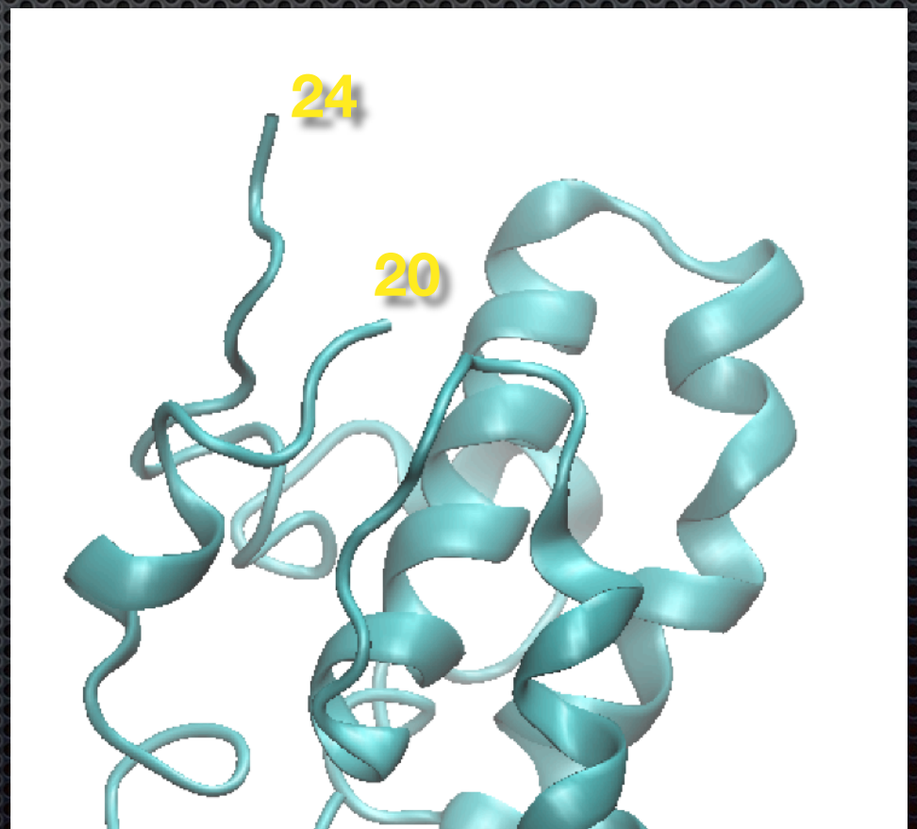
```
>> hiv1 gag matrix sequence 1 - 140
GARASVLSGGELDKWEKIRLRPGGKKQYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTIAVLYCVHQRIDV
KDTKEALDKIEEEQNKSKKKAQQAAADTG
```

EPSRC
Engineering and Physical Sciences Research Council

NSF

```
ATOM    314  C   LEU X  20      -9.122 -84.992  81.008  1.00  0.00       GAG  C
ATOM    315  O   LEU X  20      -9.647 -85.603  80.079  1.00  0.00       GAG  O
ATOM    361  N   GLY X  24      -5.700 -90.839  86.003  1.00  0.00       GAG  N
ATOM    362  HN  GLY X  24      -5.897 -91.523  85.304  1.00  0.00       GAG  H
```

```
>> hiv1 gag matrix sequence 1 - 140
GARASVLSGGELDKWEKIRLRPGGKKQYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTIAVLYCVHQRIDV
KDTKEALDKIEEEQNKSKKKAQQAAADTG
```



EPSRC
Engineering and Physical Sciences Research Council

NSF

PSFGEN to build missing bits

missing loop in hiv1 gag matrix

```
ATOM    314  C   LEU X  20      -9.122 -84.992  81.008  1.00  0.00      GAG  C
ATOM    315  O   LEU X  20      -9.647 -85.603  80.079  1.00  0.00      GAG  O
ATOM    361  N   GLY X  24      -5.700 -90.839  86.003  1.00  0.00      GAG  N
ATOM    362  HN  GLY X  24      -5.897 -91.523  85.304  1.00  0.00      GAG  H
```

```
topology /usr/local/bin/toppar/top_all27_prot_na.inp

alias residue HIS HSE
alias atom ILE CD1 CD
alias atom SER HG HG1

segment GAG {
 first None
 pdb output_building/hiv1_gag_matrix_sequence_from_fasta.pdb
 last CTER
}

coordpdb output_building/hiv1_gag_matrix_missing_loop.pdb GAG
patch GLYP GAG:1

guesscoord

writepsf output_building/linear_peptide.psf

writepdb output_building/linear_peptide.pdb
```

sequence file only needs one atom per residue

path to pdb file (sequence)

path coordinates file (pdb)

use IC to add H etc.

EPSRC
Engineering and Physical Sciences Research Council

NSF

PSFGEN (filling gaps)

STEP 1: Create new psf and pdb files where psfgen will add missing bits using edited PDB files as input.

LOOK AT THE STRUCTURE BEFORE MOVING TO STEP 2.

STEP 2: Set up a “fixed-atoms” PDB file so that structure is not pulled apart when you minimize the structure

STEP 3: Write NAMD input file with the addition of:
fixed atoms
IMD (so you can watch)
more minimization steps

STEP 4: Minimize the structure using NAMD (repeat w/o fixed atoms)

STEP 5: If all goes well, run a short NVE dynamics run (10 ps) . . .

If step 5 fails then go back to step 4 or maybe back to step 1 !!!

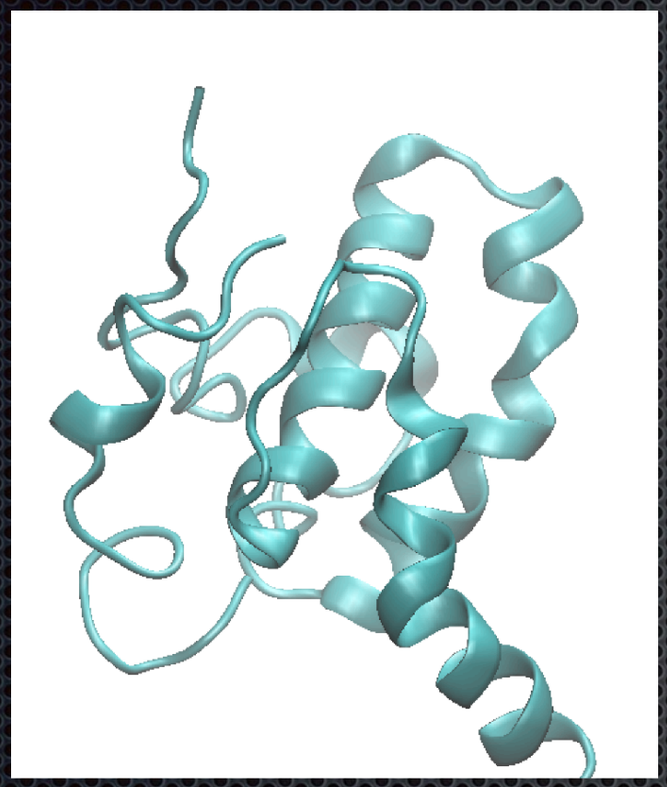
EPSRC
Engineering and Physical Sciences Research Council

NSF

Toy system

a piece of HIV1 Gag Matrix

initial structure

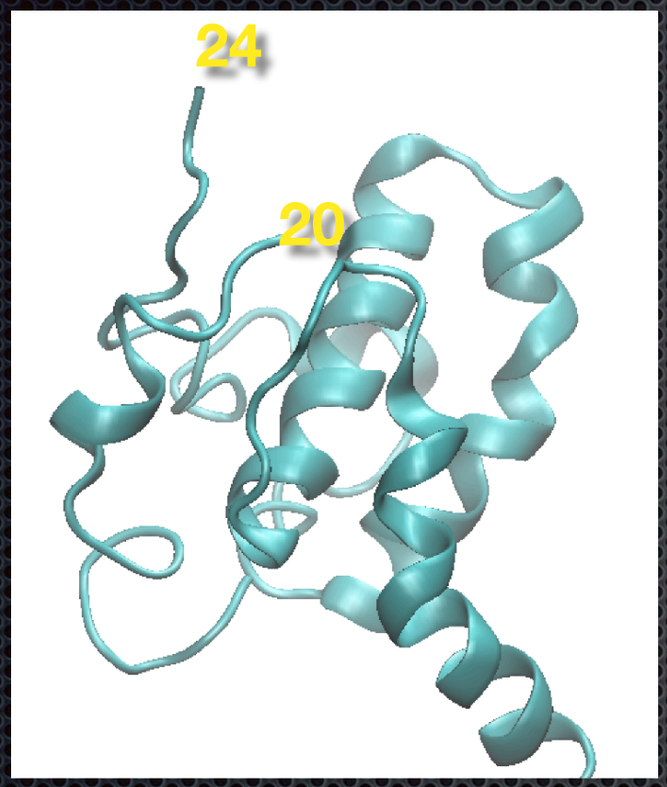


psfgen build (IC)

energy minimization & 10 ps MD

EPSRC
Engineering and Physical Sciences Research Council

initial structure



psfgen build (IC)

energy minimization & 10 ps MD

EPSRC
Engineering and Physical Sciences Research Council

NSF

PSFGEN (filling gaps)

STEP 1: Create new psf and pdb files where psfgen will add missing bits using edited PDB files as input.

Have the sequence at hand . . .

```
TSGIHVLDEL SVRALSRDIM KQNRVTVHPE KSVPRTAGYS DAVSVLAQDR      50
PSLAIVSGQG GAAGQRERVA ELVMMAREQG REVQIIAADR RSQMNMKQDE     100
RLSGELITGR RQLLEGMAFT PGSTVIVDQG EKLSLKETLT LLDGAARHNV     150
QVLITDSGQR TGTGSALMAM KDAGVNTYRW QGGEQRPATI ISEPDRNVRY     200
ARLAGDFAAS VKAGEESVAQ VSGVREQAIL TQAIRSELKT QGVLGLPEVT     250
MTALSPVWLD SRSRYLRDMY RPGMVMEQWN PETRSHDRYV IDRVTAQSHS     300
LTLRDAQGET QVVRISSLDS SWSLFRPEKM PVADGERLRV TGKIPGLRVS     350
GGDRLQVASV SEDAMTVVVP GRAEPATLPV SDSPFTALKL ENGWVETPGH     400
SVSDSATVFA SVTQMAMDNA TLNGLARSGR DVRLYSSLDE TRTAEKLARH     450
```

(a) you have one pdb file with known coordinates

(b) make a copy of (a) with only the CA atoms, add in unknown residues to make a "sequence.pdb" file. Coordinates in the sequence file do no matter.

EPSRC
Engineering and Physical Sciences Research Council

NSF

PSFGEN (filling gaps)

STEP 1: Create new psf and pdb files where psfgen will add missing bits using edited PDB files as input.

> cat coords.pdb | grep CA > ca_sequence.pdb

then, copy known coordinate line (say, ATOM 163 CA VAL ...) and manually change the index, resname, and resid fields. Do this for all missing atoms/AA.

```
ATOM      1   CA THR     1      35.989 -22.055 248.035  1.00  0.00      TRAI C
ATOM      2   CA SER     2      35.989 -22.055 248.035  1.00  0.00      TRAI C
ATOM      3   CA GLY     3      35.989 -22.055 248.035  1.00  0.00      TRAI C
ATOM      4   CA ILE     4      35.989 -22.055 248.035  1.00  0.00      TRAI C
ATOM      5   CA HSE     5      35.989 -22.055 248.035  1.00  0.00      TRAI C
ATOM      6   CA VAL     6      35.989 -22.055 248.035  1.00  0.00      TRAI C
ATOM      7   CA LEU     7      35.989 -22.055 248.035  1.00  0.00      TRAI C
ATOM      8   CA ASP     8      35.989 -22.055 248.035  1.00  0.00      TRAI C
ATOM      9   CA GLU     9      35.989 -22.055 248.035  1.00  0.00      TRAI C
ATOM     10   CA LEU    10      35.989 -22.055 248.035  1.00  0.00      TRAI C
ATOM     11   CA SER    11      35.989 -22.055 248.035  1.00  0.00      TRAI C
ATOM    163   CA VAL    12      35.989 -22.055 248.035  1.00  0.00      TRAI C
ATOM    179   CA ARG    13      39.541 -23.591 248.463  1.00  0.00      TRAI C
ATOM    203   CA ALA    14      41.070 -23.037 251.954  1.00  0.00      TRAI C
ATOM    213   CA LEU    15      42.765 -25.002 254.835  1.00  0.00      TRAI C
```

EPSRC
Engineering and Physical Sciences Research Council

NSF

PSFGEN (filling gaps)

now, create a temporary pdb/psf to build the structure

Sample input file (here named: build_system_protein.psfgen):

>psfgen build_system_protein.psfgen >& build_protein.out &

```
topology /home/mdschool/research/toppar/top_all27_prot_na.inp

alias residue HIS HSE

segment TRAI {
 first NTER
 last CTER
 pdb output_building/ca_sequence.pdb
}

coordpdb output_building/coords_traI_recd_10_481_temp.pdb TRAI

guesscoord

writepsf output_building/temp_traI.psf

writepdb output_building/temp_traI.pdb
```

path to pdb file (sequence)

path coordinates file (pdb)

use IC to add H and missing AA

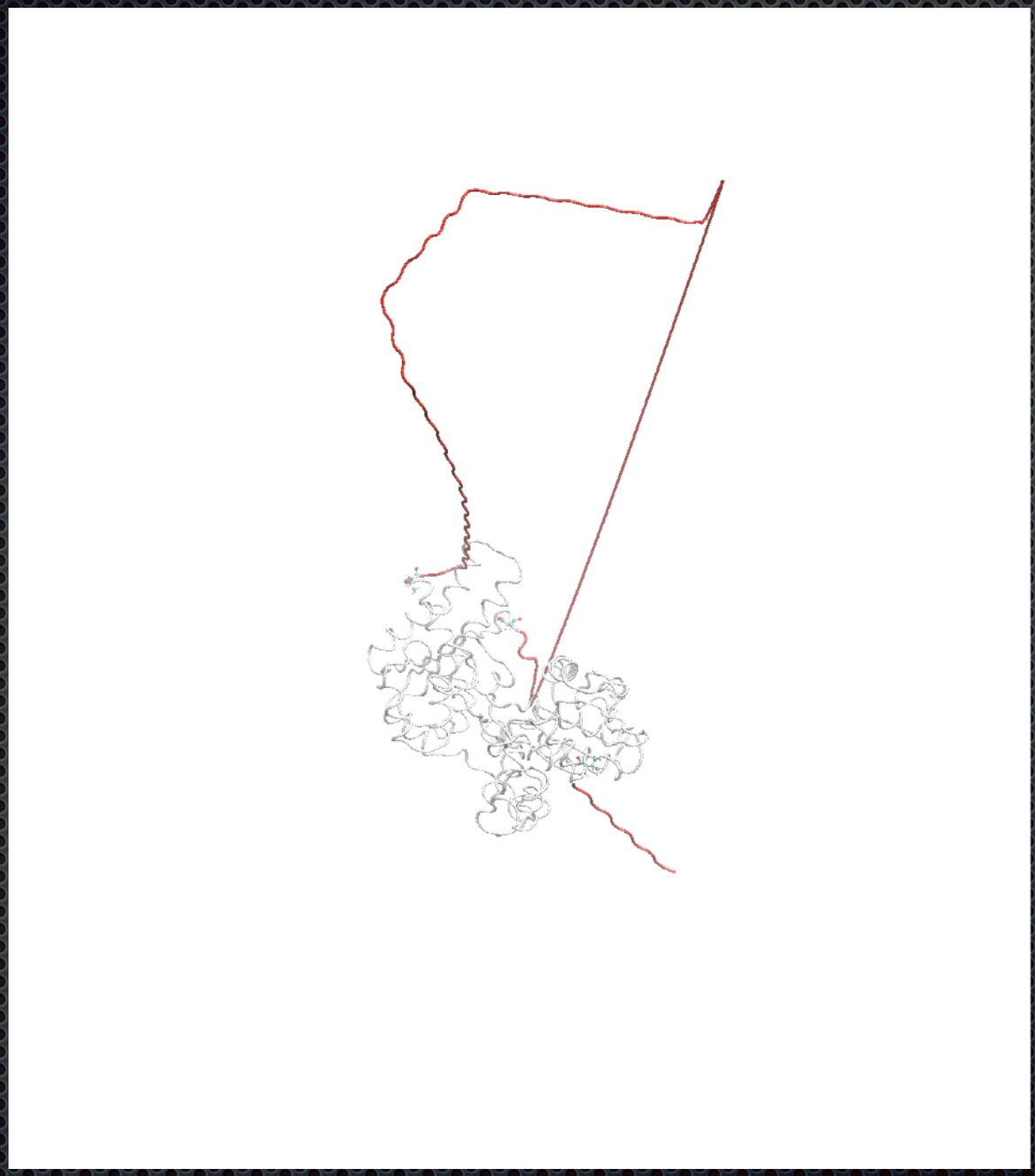
IC will “grow” the chain **blindly** . . .

EPSRC
Engineering and Physical Sciences Research Council

NSF

PSFGEN (filling gaps)

LOOK AT THE STRUCTURE BEFORE MOVING TO STEP 2.

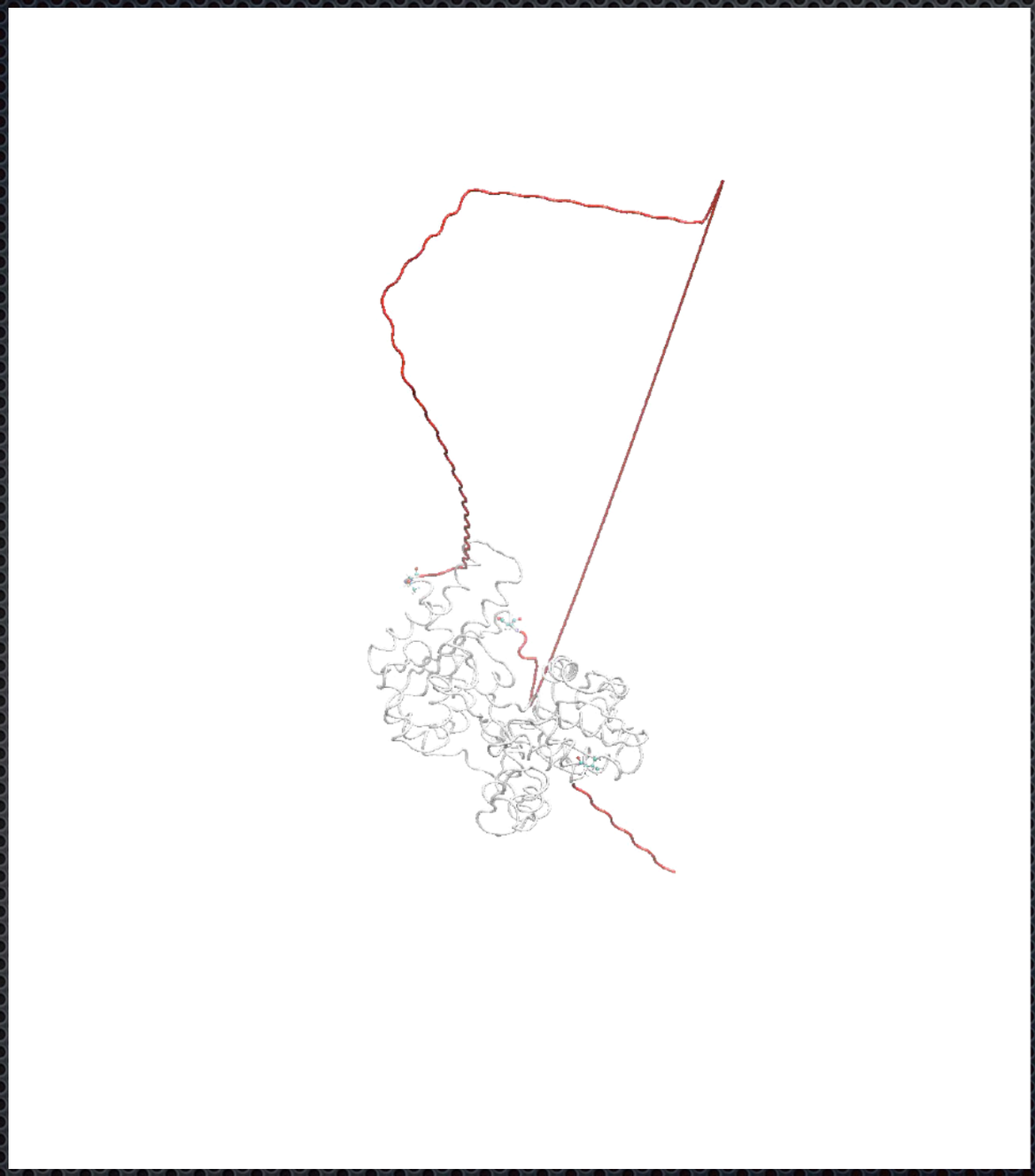


EPSRC
Engineering and Physical Sciences Research Council

NSF

PSFGEN (filling gaps)

LOOK AT THE STRUCTURE BEFORE MOVING TO STEP 2.

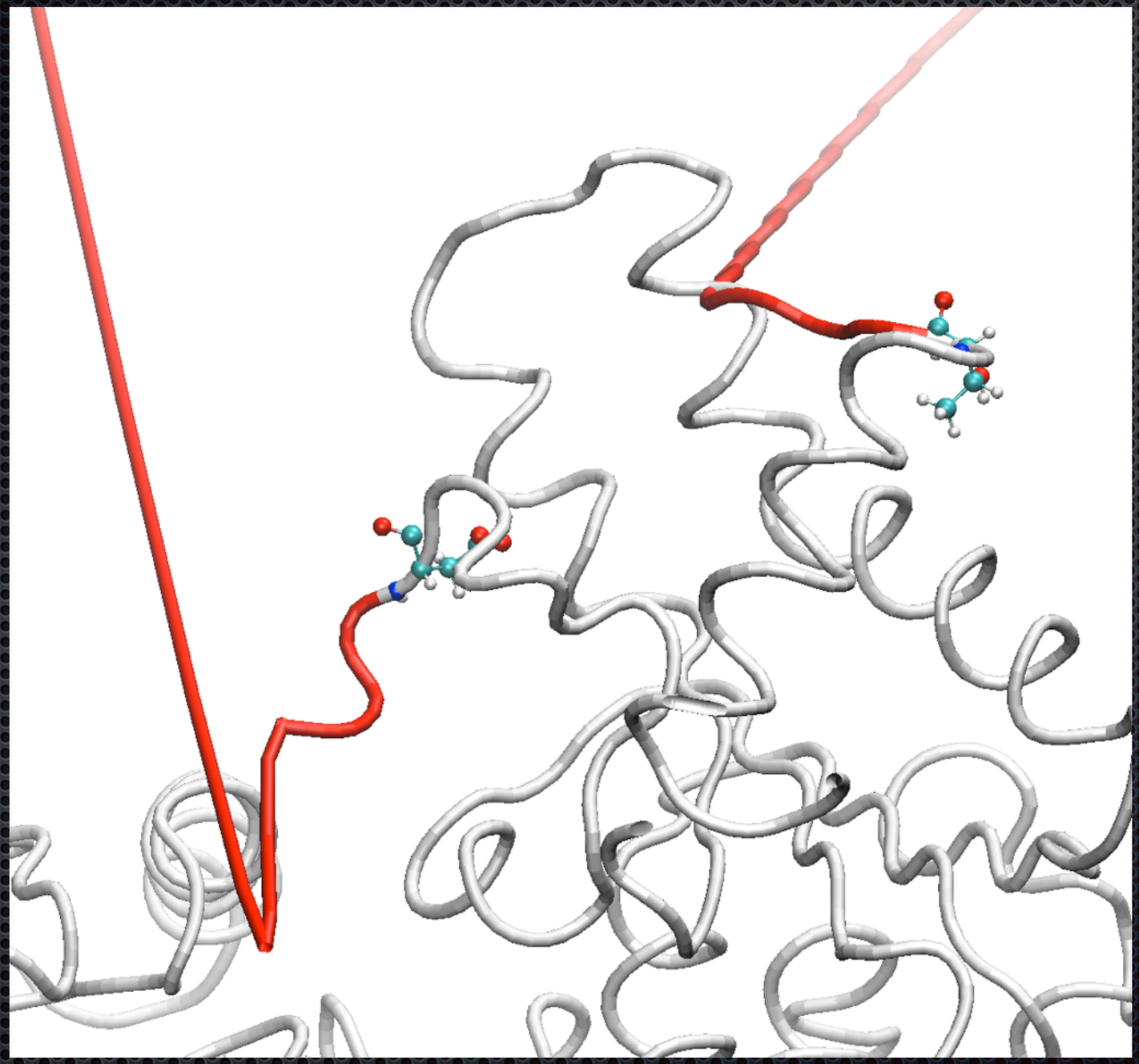


EPSRC
Engineering and Physical Sciences Research Council

NSF

PSFGEN (filling gaps)

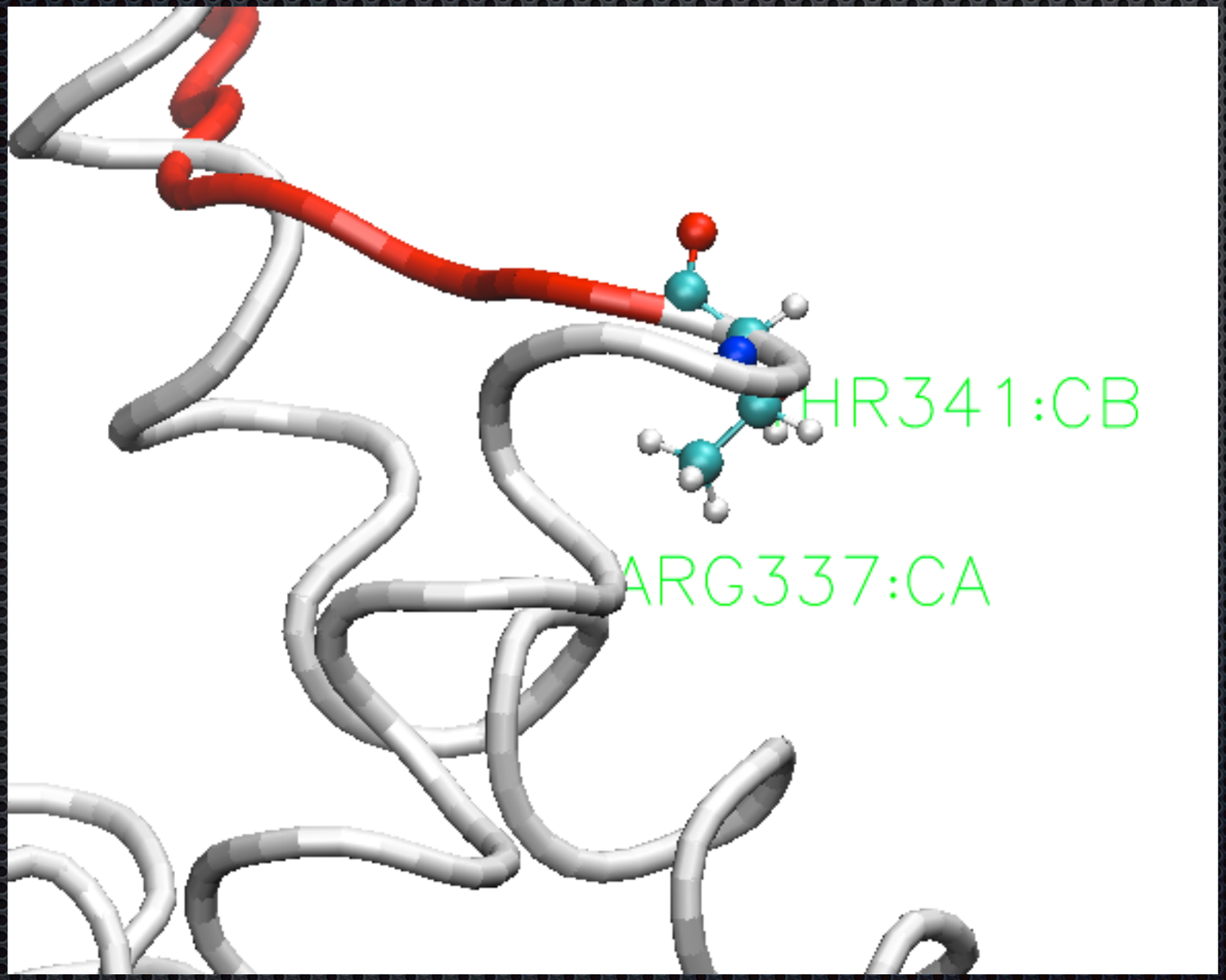
LOOK AT THE STRUCTURE BEFORE MOVING TO STEP 2.



EPSRC
Engineering and Physical Sciences Research Council

PSFGEN (filling gaps)

YIKES! Welcome to the real world!!! We have overlap. Let's remove some known coordinates near the overlap and re-build the structure. You don't need to change the sequence file. Call the new coordinate file: coords_tral_10_gap_481.pdb



EPSRC
Engineering and Physical Sciences Research Council

NSF

PSFGEN (filling gaps)

YIKES! Welcome to the real world!!! We have overlap. Let's remove some known coordinates near the overlap and re-build the structure. You don't need to change the sequence file. Call the new coordinate file: coords_tral_10_gap_481.pdb

```
. . .

ATOM   1621    N LEU D 338     -62.123 -15.211 -65.673      207.57           N
ATOM   1622   CA LEU D 338     -62.146 -15.489 -64.287      207.57           C
ATOM   1623    C LEU D 338     -62.455 -14.115 -63.597      207.57           C
ATOM   1629    N ARG D 339     -63.258 -13.329 -64.312      274.04           N
ATOM   1630   CA ARG D 339     -63.760 -11.991 -63.952      274.04           C
ATOM   1631    C ARG D 339     -65.237 -12.041 -64.220      274.04           C
ATOM     49    N ALA D 406     -47.335 -18.204 -62.224      192.55           N
ATOM     50   CA ALA D 406     -47.752 -17.411 -63.397      192.55           C
ATOM     51    C ALA D 406     -46.685 -16.573 -64.078      192.55           C
ATOM     59    N THR D 407     -46.016 -17.138 -65.075      244.03           N
ATOM     60   CA THR D 407     -45.050 -16.361 -65.852      244.03           C
ATOM     61    C THR D 407     -45.999 -16.025 -66.999      244.03           C

. . .
```

EPSRC
Engineering and Physical Sciences Research Council

NSF

PSFGEN (filling gaps)

now, create a temporary pdb/psf to build the structure

Sample input file (here named: build_system_protein.psfgen):

>psfgen build_system_protein_II.psfgen >& build_protein_II.out &

```
topology /home/mdschool/toppar/top_all27_prot_na.inp

alias residue HIS HSE

segment TRAI {
 first NTER
 last CTER
 pdb output_building/ca_sequence.pdb
}

coordpdb output_building/coords_traI_10_gap_481.pdb TRAI

guesscoord

writepsf output_building/temp_2_traI.psf

writepdb output_building/temp_2_traI.pdb
```

path to pdb file (sequence)

path coordinates file (pdb)

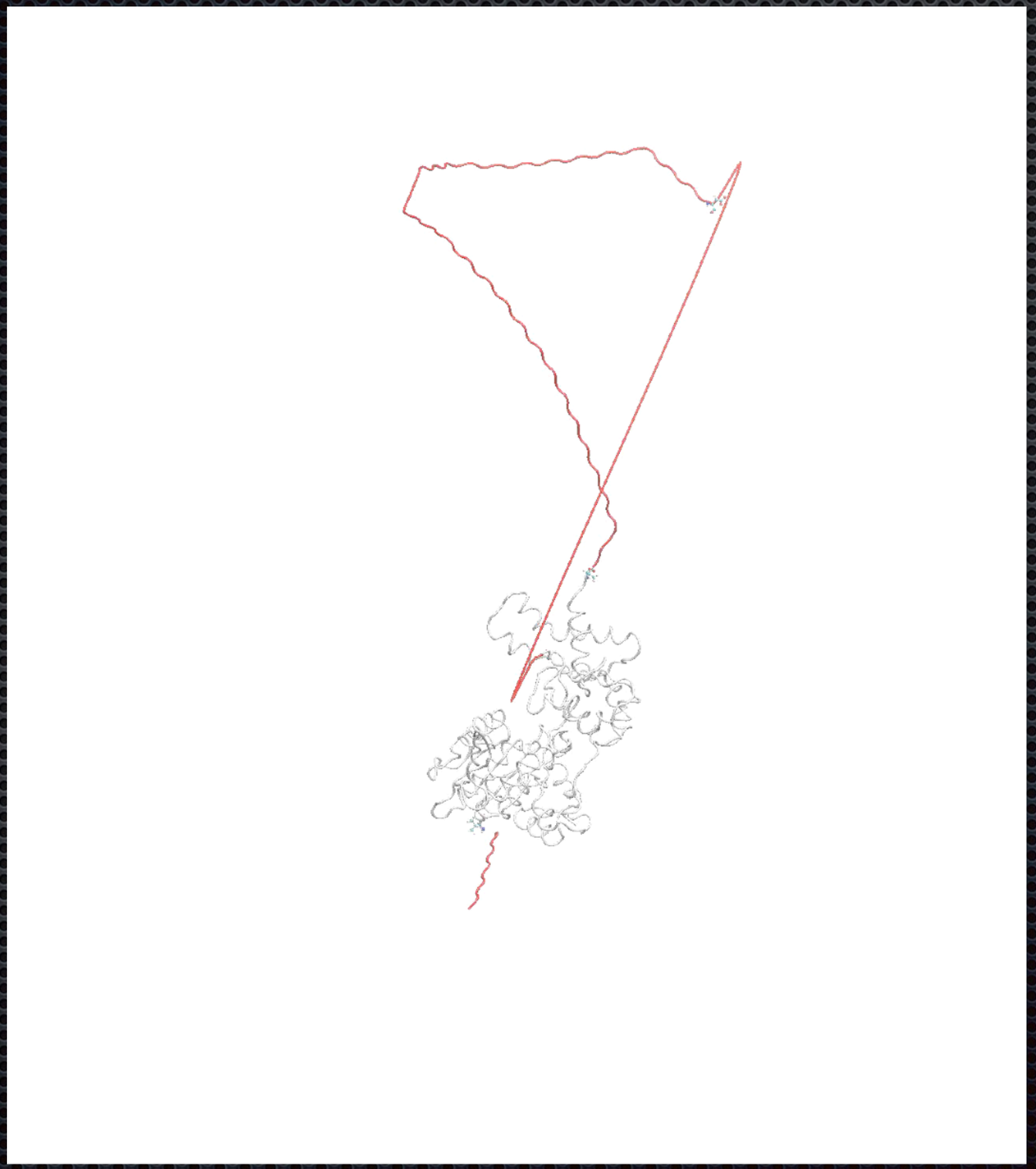
use IC to add H and missing AA

EPSRC
Engineering and Physical Sciences Research Council

NSF

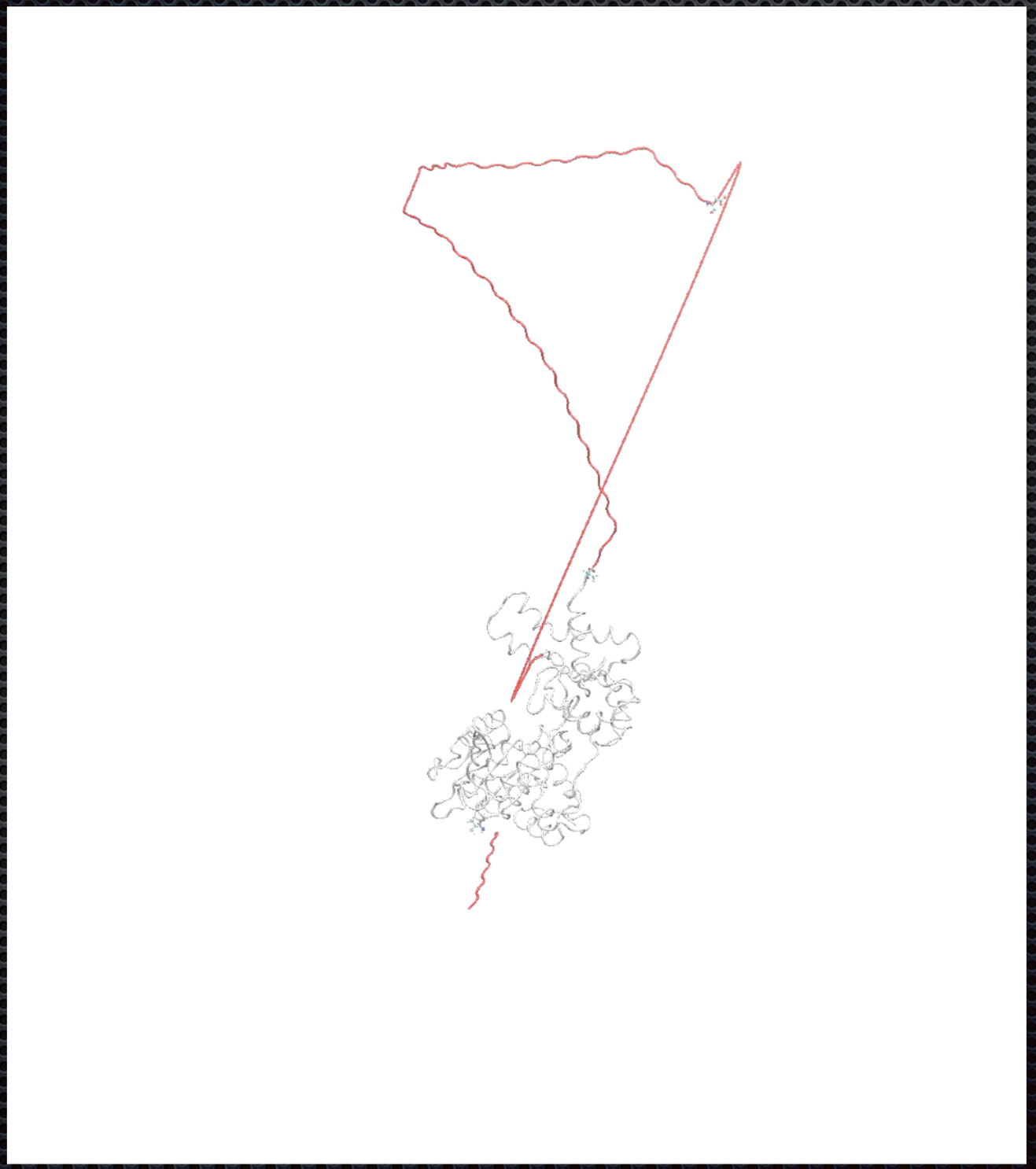
PSFGEN (filling gaps)

LOOK AT THE STRUCTURE BEFORE MOVING TO STEP 2.



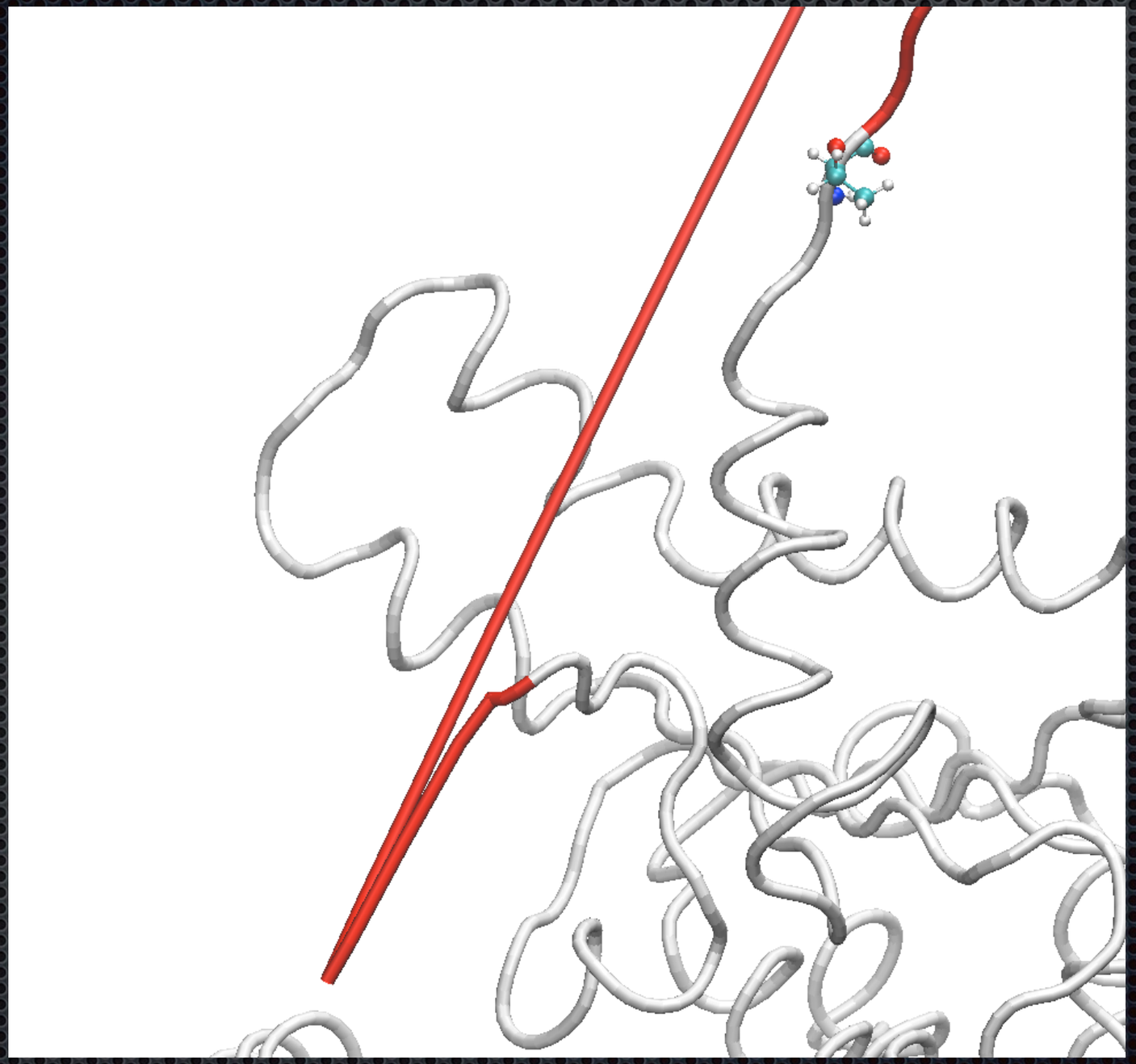
PSFGEN (filling gaps)

LOOK AT THE STRUCTURE BEFORE MOVING TO STEP 2.



PSFGEN (filling gaps)

LOOK AT THE STRUCTURE BEFORE MOVING TO STEP 2.



EPSRC
Engineering and Physical Sciences Research Council

NSF

PSFGEN (filling gaps)

STEP 1: Create new psf and pdb files where psfgen will add missing bits using edited PDB files as input.

LOOK AT THE STRUCTURE BEFORE MOVING TO STEP 2.

STEP 2: Set up a "fixed-atoms" PDB file so that structure is not pulled apart when you minimize the structure

STEP 3: Write NAMD input file with the addition of:
fixed atoms
IMD (so you can watch)
more minimization steps

STEP 4: Minimize the structure using NAMD (repeat w/o fixed atoms)

STEP 5: If all goes well, run a short NVE dynamics run (10 ps) . . .

If step 5 fails then go back to step 4 or maybe back to step 1 !!!

EPSRC
Engineering and Physical Sciences Research Council

NSF

New PDB I

With the **temp_2_tral_piece.pdb (psf)** file open in VMD . . . at the console type:

```
vmd > set everyone [atomselect top all]
atomselect32

vmd > $everyone set beta 1.00

vmd > set sel1 [atomselect top "(resid > 341 and resid < 406) or (resid < 12)"]
atomselect33

vmd > $sel1 set beta 0.00

vmd > $everyone writepdb fixed_atoms.pdb

Info) Opened coordinate file fixed_atoms.pdb for writing.
Info) Finished with coordinate file fixed_atoms.pdb.
```

New PDB I

With the **temp_2_tral_piece.pdb (psf)** file open in VMD . . . at the console type:

```
vmd > set everyone [atomselect top all]
atomselect32

vmd > $everyone set beta 1.00

vmd > set sel1 [atomselect top "(resid > 341 and resid < 406) or (resid < 12)"]
atomselect33

vmd > $sel1 set beta 0.00
```

```
vmd > $everyone writepdb fixed_atoms.pdb

Info) Opened coordinate file fixed_atoms.pdb for writing.
Info) Finished with coordinate file fixed_atoms.pdb.
```

```
. . .

ATOM   6203  C   SER P 405     -46.159 -18.127 -61.621  1.00  0.00       PCLN C
ATOM   6204  O   SER P 405     -45.252 -17.351 -61.914  0.00  0.00       PCLN O
ATOM   6205  N   ALA P 406     -47.335 -18.204 -62.224  1.00  1.00       PCLN N
ATOM   6206  HN  ALA P 406     -48.017 -18.846 -61.875  0.00  1.00       PCLN H

. . .
```

PSFGEN (filling gaps)

STEP 1: Create new psf and pdb files where psfgen will add missing bits using edited PDB files as input.

LOOK AT THE STRUCTURE BEFORE MOVING TO STEP 2.

STEP 2: Set up a "fixed-atoms" PDB file so that structure is not pulled apart when you minimize the structure

STEP 3: Write NAMD input file with the addition of:
fixed atoms
IMD (so you can watch)
more minimization steps

STEP 4: Minimize the structure using NAMD (repeat w/o fixed atoms)

STEP 5: If all goes well, run a short NVE dynamics run (10 ps) . . .

If step 5 fails then go back to step 4 or maybe back to step 1 !!!

EPSRC
Engineering and Physical Sciences Research Council

NSF

Overview

Intermediate Building

Building coordinates for missing atoms/residues

Dealing with domain / domain orientation

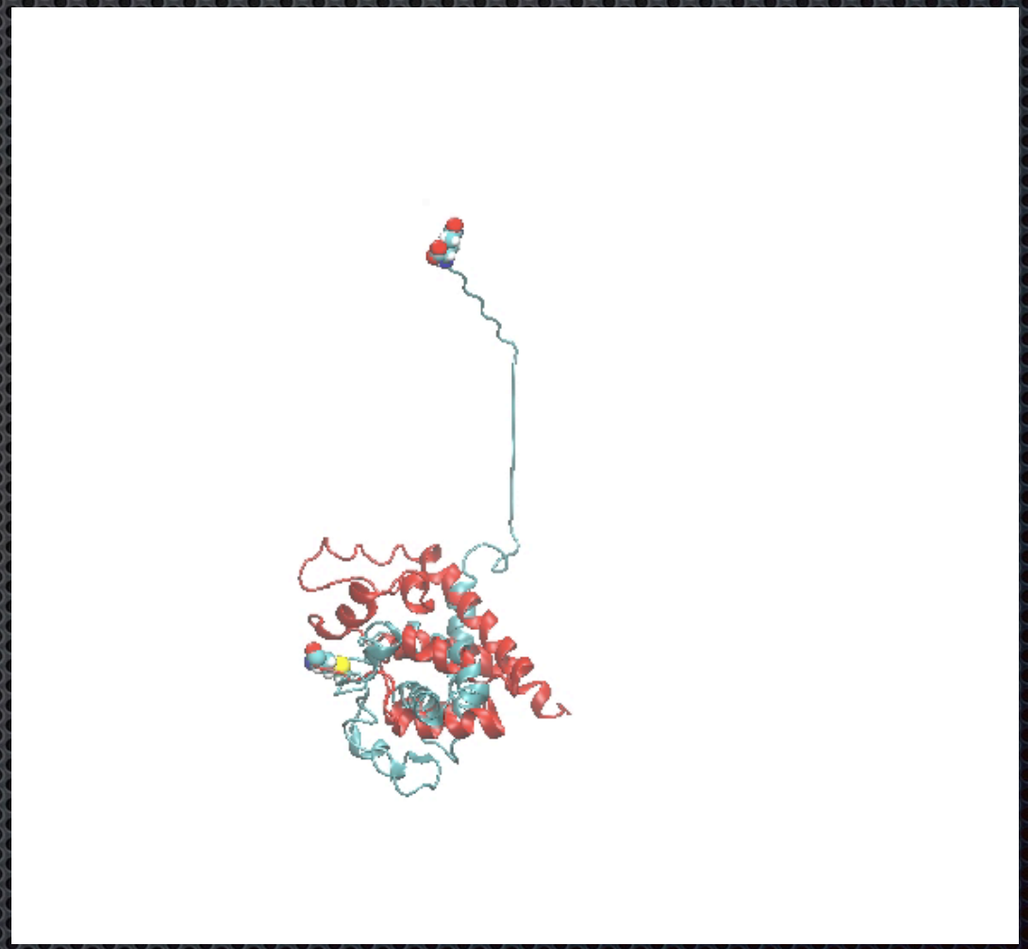
SASSIE-web Build
A trick for topological issues

EPSRC
Engineering and Physical Sciences Research Council

NSF

multi domains

vmd: move / molecule OR “8”



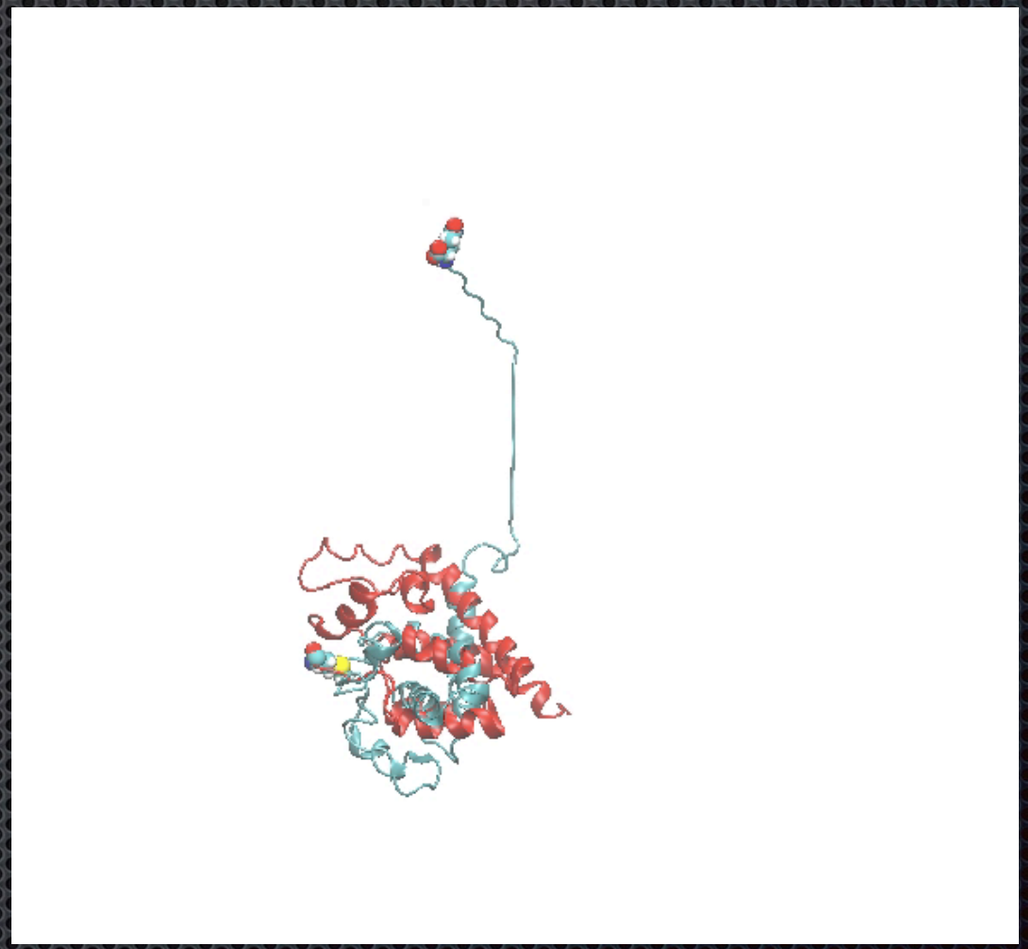
save coordinates after moving

EPSRC
Engineering and Physical Sciences Research Council

NSF

multi domains

vmd: move / molecule OR “8”



save coordinates after moving

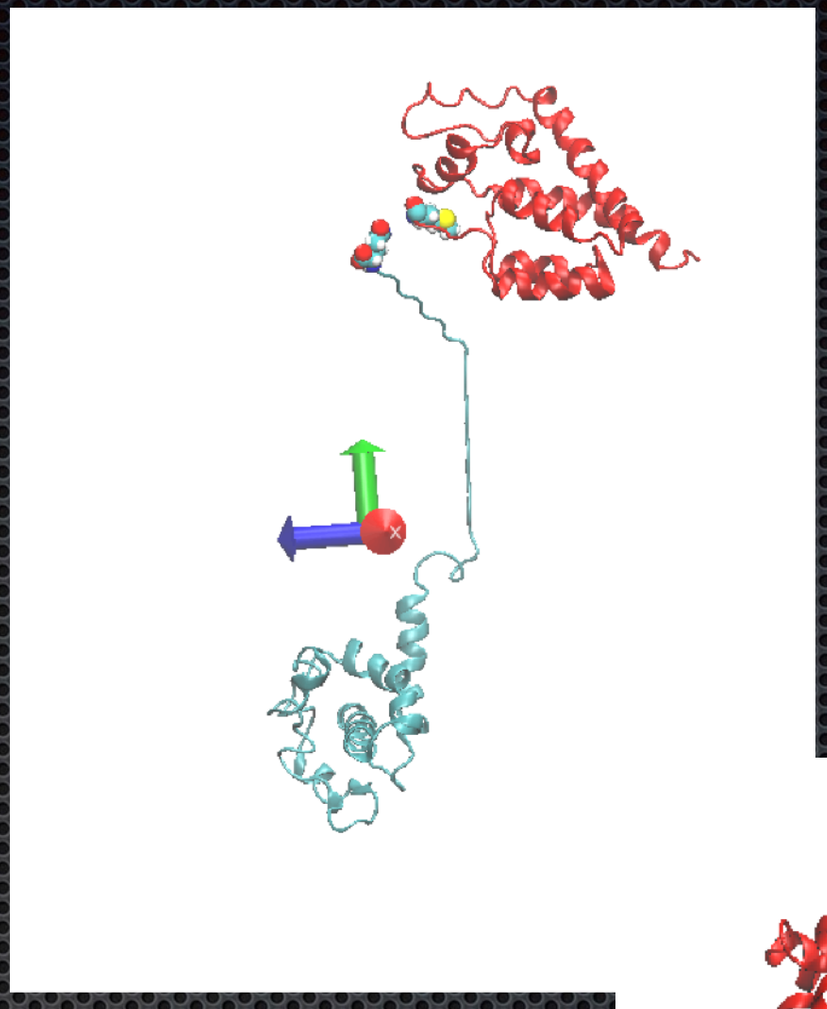
EPSRC
Engineering and Physical Sciences Research Council

NSF

vmd matrix operations

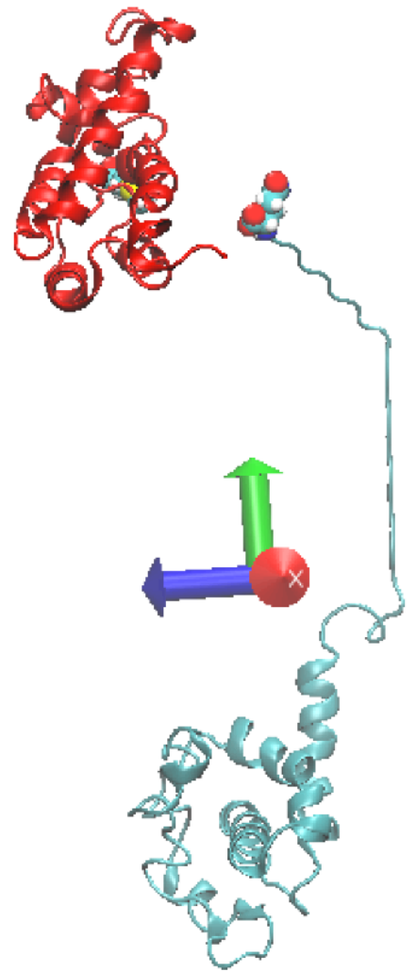
translate:

vmd >set sel [atomselect top all]
vmd> $sel moveby {0 60 0}
vmd> $sel writepdb moved_rn_d2_new.pdb



rotate -90 degrees:

vmd> set sel [atomselect top all]
vmd> set M [transaxis y -90]
vmd> $sel move $M



http://www.ks.uiuc.edu/Research/vmd/vmd-1.7/ug/node164.html

EPSRC
Engineering and Physical Sciences Research Council

IDPs

Piece structures together (1) in lab VIII

1. Load individual PDB files into VMD. Use “8” to grab one and sequentially move along X, rotate system, move along Y, rotate system, move along Z. Save coordinates using VMD. ((will probably have to execute rotate TCL commands))

“align by eye” & hope for the best. Concatentate files, renumber amino acids and atoms, make new PDB/PSF pair, minimize, 10 ps NVE MD . . .

2. Add 3 to 5 amino acids to N-terminal of the C-terminal part, then overlap these amino acids (using backbone as a basis).

Then you align the two structures (C-terminal of “A” to modified N-terminal of “B”)

Concatentate PDB files, remove the “bridging” 3 to 5 AA, renumber remaining amino acids and atoms, make new PDB/PSF pair, minimize, 10 ps NVE MD . . .

3. Insert your method here . . . etc. (pymol, mdtools, etc.)

EPSRC
Engineering and Physical Sciences Research Council

NSF

Overview

Intermediate Building

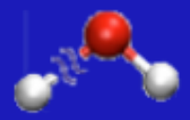
Building coordinates for missing atoms/residues
Dealing with domain / domain orientation

SASSIE-web Build

A trick for topological issues

EPSRC
Engineering and Physical Sciences Research Council

NSF

Build 

PDB Scan: to help understand what is in your PDB file (no-header or header)

Build Utilities: collection of PDB & FASTA methods

Beta β

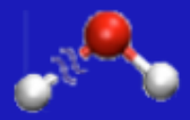
PDB Rx: interactive / automatic repair of PDB files
proteins & nucleic acids
missing bits w/ PyRosetta

INPUT: PDB
OUTPUT: new PDB w/ PSF

Alpha α

EPSRC
Engineering and Physical Sciences Research Council

NSF

Build 

Build Utilities: collection of PDB & FASTA methods

Beta 

Docs / Demo

EPSRC
Engineering and Physical Sciences Research Council

NSF

Overview

Intermediate Building Tricks

Building coordinates for missing atoms/residues
Simulated annealing
Building IDPs

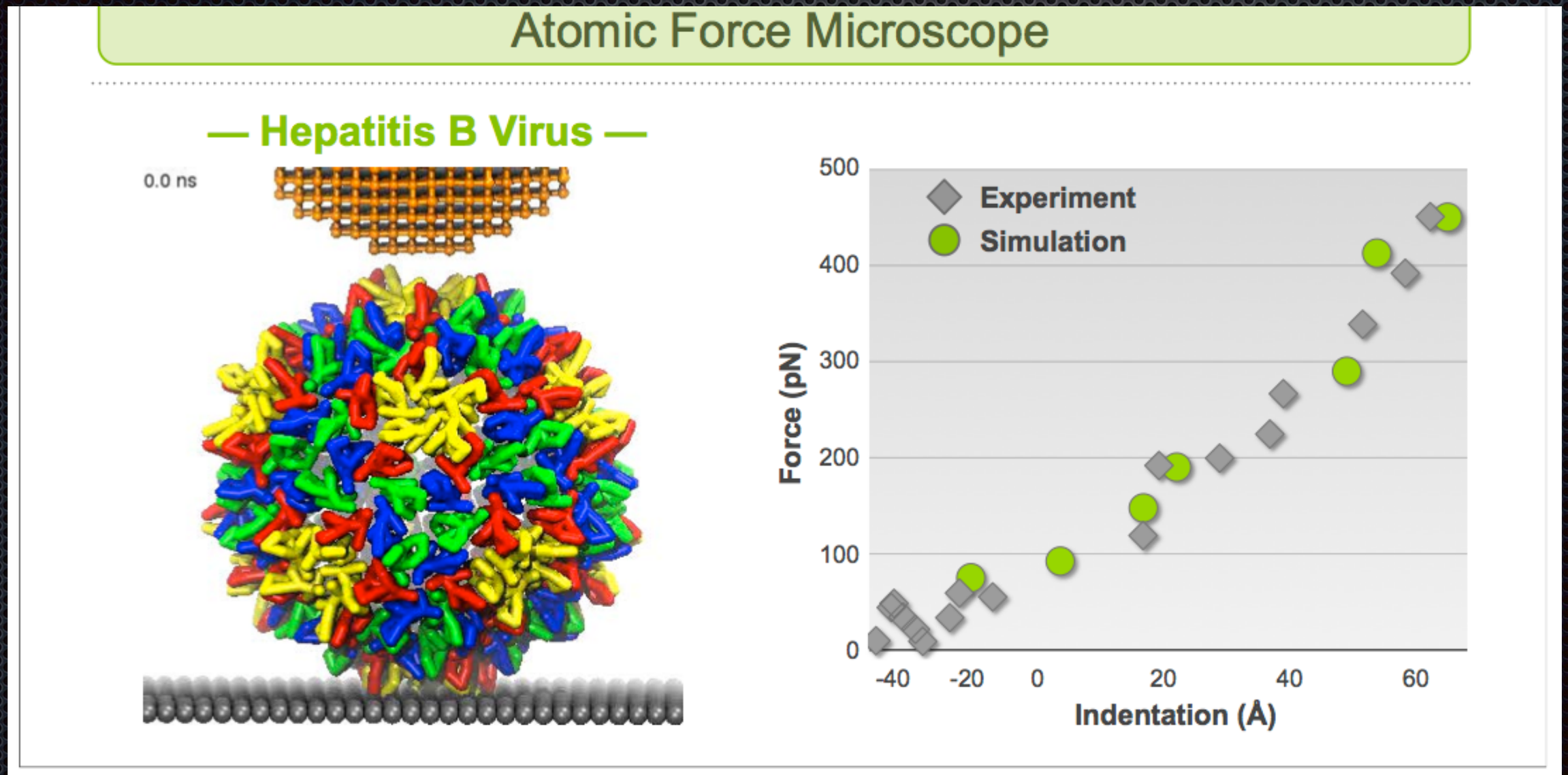
A trick for topological issues

EPSRC
Engineering and Physical Sciences Research Council

NSF

Steered MD

http://www.ks.uiuc.edu/Research/smd_imd/

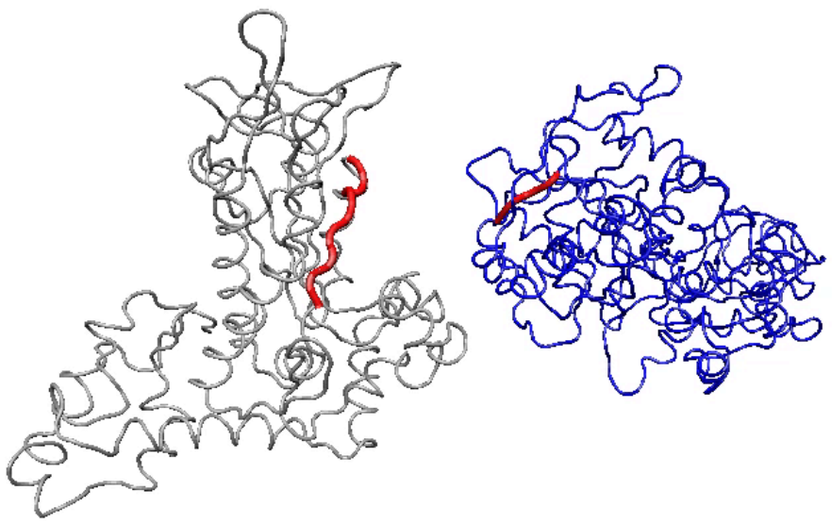


EPSRC
Engineering and Physical Sciences Research Council

NSF

Steered MD for building

Buried disordered tails ... have to pull them out to link two structures together

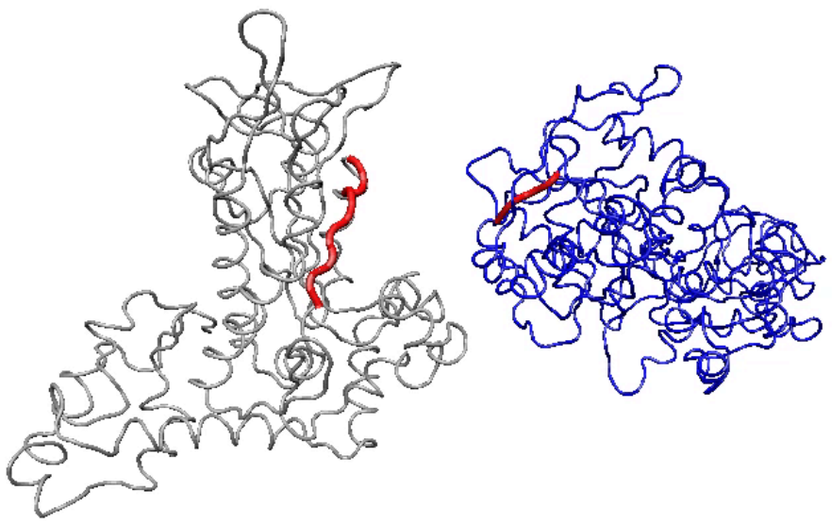


EPSRC
Engineering and Physical Sciences Research Council

NSF

Steered MD for building

Buried disordered tails ... have to pull them out to link two structures together



EPSRC
Engineering and Physical Sciences Research Council

NSF

Steered MD for building

So, in the following movie you will see:

1. files that we are working with
2. a view of the input file (min0)
3. loading the psf and then pdb into vmd
4. initiating namd2 min0
5. starting the run via IMD in VMD
6. applying the residue specifc force to pull the N-terminal out
7. a few thousand steps of MD

NOTE:

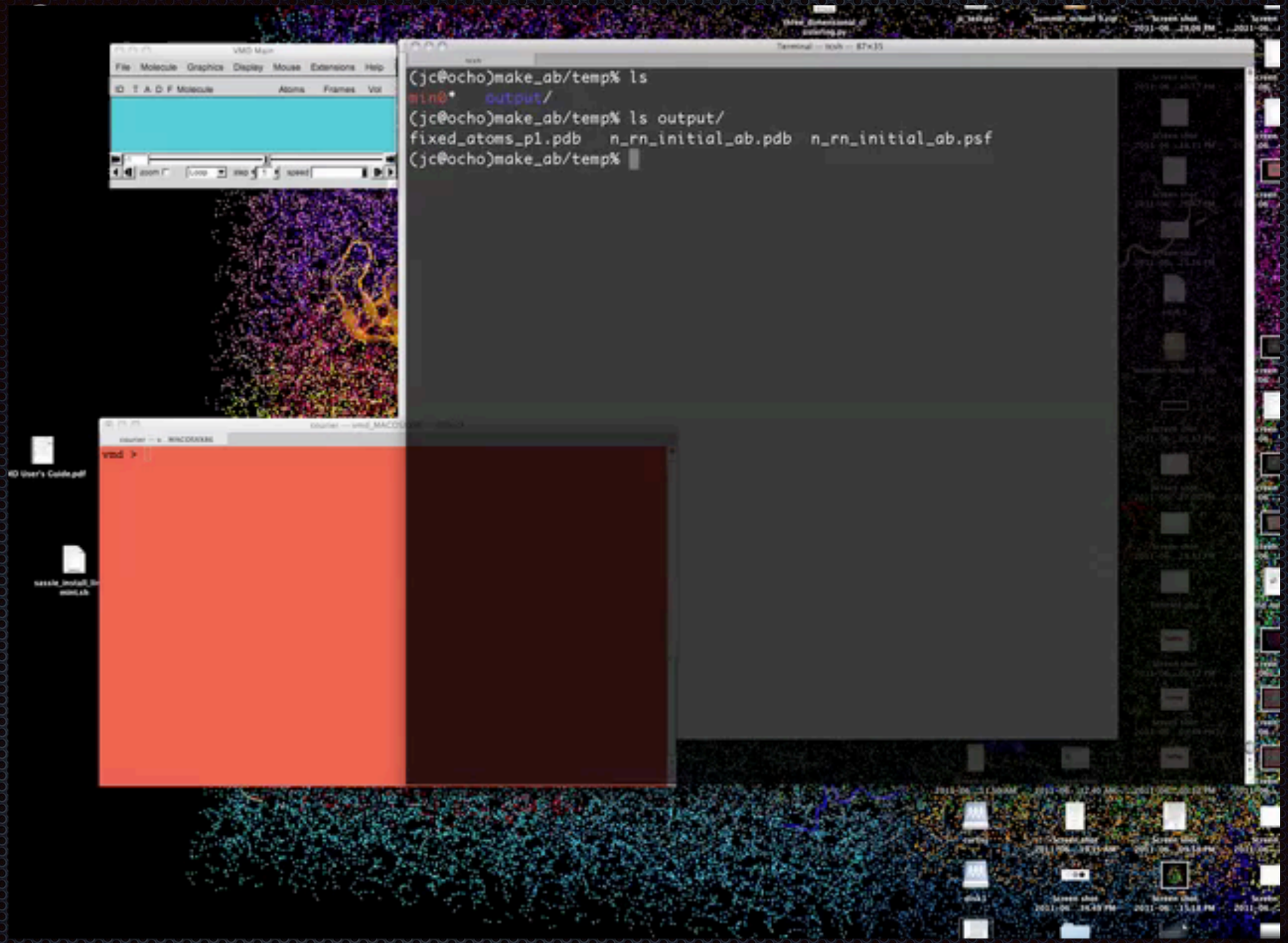
This N-terminal was known NOT to be buried and that it was unstructured. The resulting structure was minimized prior to it's use in further building steps!

This example was done in the process of building a larger IDP with many known and unknown bits.

It allowed us to overlap the N-terminal AA with the "other" piece etc.

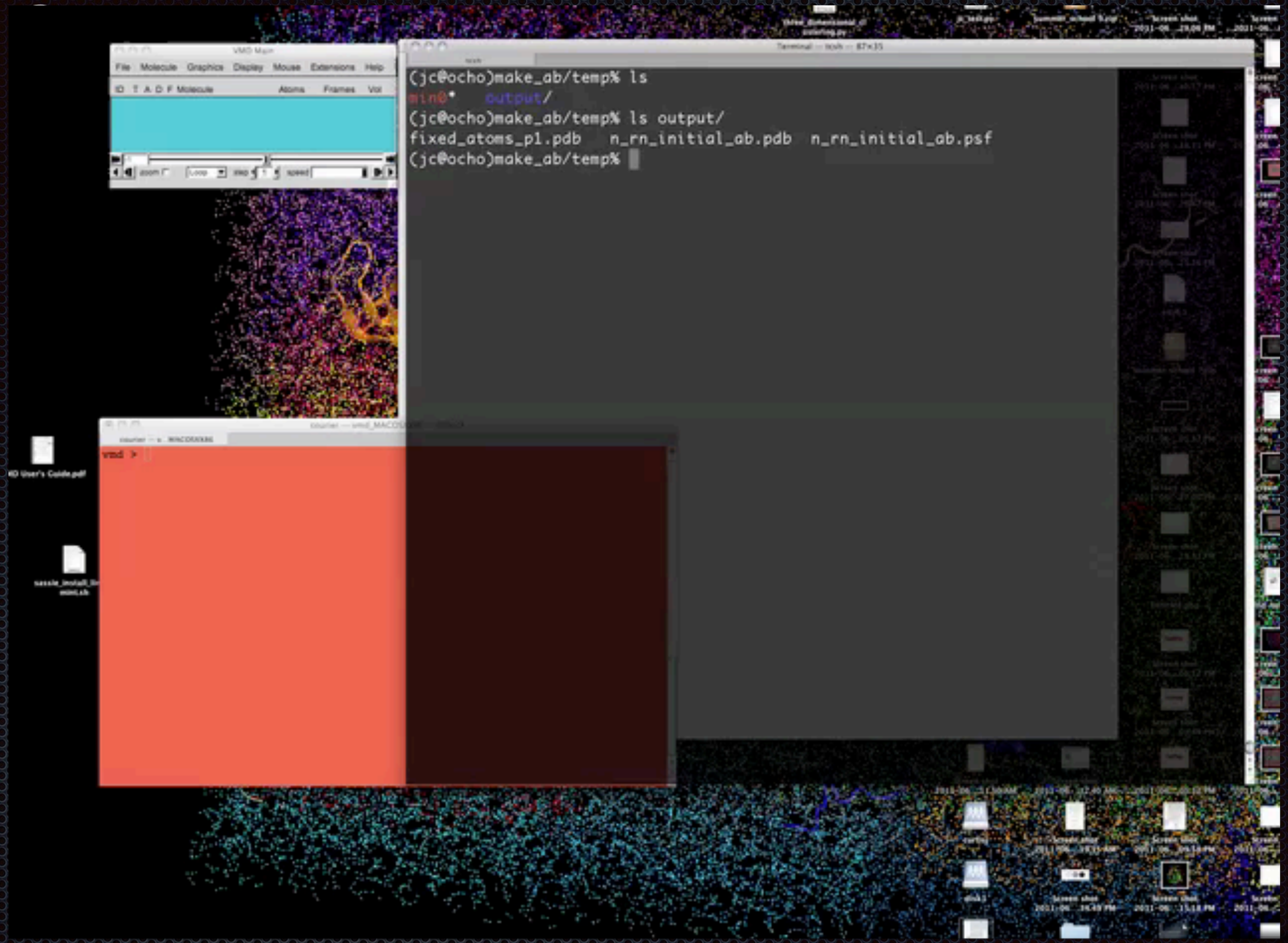
EPSRC
Engineering and Physical Sciences Research Council

NSF



EPSRC
Engineering and Physical Sciences Research Council

NSF



EPSRC
Engineering and Physical Sciences Research Council

NSF

SUMMARY

What have we accomplished?

We showed some examples of how to complete structures and covered some ways to build larger structures/models from bits of known structure.

What is next?

LAB VIII --> Advanced Building

LABS IX/X & STUDENT PROJECTS

Take home “final exam” -> antibody

EPSRC
Engineering and Physical Sciences Research Council

NSF