University of Illinois at Urbana-Champaign
Luthey-Schulten Group
NIH Center for Macromolecular Modeling and Bioinformatics

Assembly of the h18 Pseudoknot in *Escherichia coli* Using Structure-Based Models
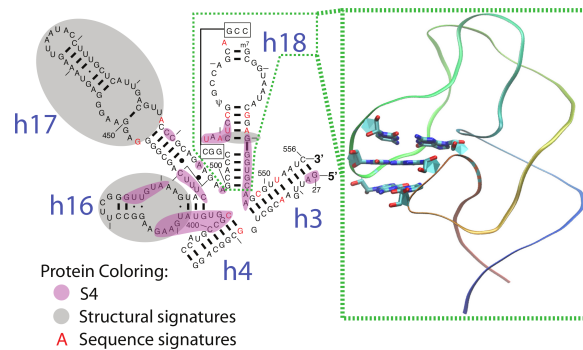

Figure 1: Structure of the h18 pseudoknot taken from the *Escherichia coli* crystal structure: 2I2P

VMD Developer: Jim Phillips

NAMD-SMOG Developers

Jonathan Lai
Jim Phillips

Tutorial Authors

Jonathan Lai
Ke Chen
Zan Luthey-Schulten

January 21, 2014

A current version of this tutorial is available at:
http://vidar.scs.uiuc.edu/jlai7/Tutorial/GoDemo.pdf

Contents

1 Introduction

This tutorial is designed to guide users of VMD and NAMD in the use of NAMD Go. It is assumed that one already has a working knowledge of VMD and NAMD, as well as, Shell and Tcl languages. For the accompanying VMD and NAMD tutorials go to: `http://www.ks.uiuc.edu/Training/Tutorials`

This tutorial has been designed specifically for NAMD versions newer than November 20th, 2013. It should take about 5 hours to complete in its entirety. NAMD can be downloaded from: `http://www.ks.uiuc.edu/~jim/tmp/NAMD_build.latest/`

Description of problem

Because the timescale of ribosome assembly is on the order of minutes (a timescale inaccessible to MD simulations), we will use structure-based Go models to study the assembly of part of the ribosome—namely the h18 pseudoknot. Although these structure-based approaches were used originally in protein folding studies, they are now being applied to protein/nucleic acid systems. In these Go models, the final structure is known and is used to apply a potential within a molecular dynamics (MD) simulation that biases the system towards its native state. This allows for the observation of intermediates and transition barriers in a short time.

We will first give an overview of ribosome assembly and the Go potentials. Then we will provide scripts to analyze several order parameters that lend themselves to experiment.

You should download the following files:

- Starting.h18.gro   Coordinate file of the h18 pseudoknot rRNA at the decoding site of the small subunit
- Reference.h18.top   Topology file used to run the Go-NAMD simulations that fold the pseudoknot from an unfolded conformation
- go.namd   configuration file
- (Scripts)   Scripts to analyze the RNA folding trajectory
- (Intermediate files)   Directory of intermediate files from a previously run trajectory

All of the tutorial files can be downloaded from the following website: `http://vidar.scs.uiuc.edu/jlai7/Tutorial/GoDemo.zip`.

The ribosome, a heteromeric dimer, is the core of cellular protein synthesis. In order to translate new proteins, the ribosome reads and decodes a mRNA message. The heart of the decoding center of the small subunit is the h18 pseudoknot. This complex tertiary structure is vital for facilitating the base pairing between the A-site tRNA and the codon on the mRNA. Because of the complex topology of the pseudoknot, the struc-

ture is highly frustrated. This frustration is thought to be alleviated by the binding of two r-proteins, S4 and S12, implicating protein binding with accelerating rRNA folding. A more complete analysis of the h18 pseudoknot structure will involve looking at changes in the torsions and secondary structure [1, 2]; however, for the sake of the tutorial, we will limit our analysis to studying the formation of the base pairs that are most frustrated.

Since we are only interested in seeing what topological barriers arise during the folding of h18, we could use several enhanced sampling techniques, such as replica-exchange. However, because the ribosome assembly takes about a minute to complete, the folding of h18 occurs on a timescale that is inaccessible to these MD techniques. Instead, we use a structure-based model. These models have been widely used in the field of protein folding and have also been applied to nucleic acid folding [1, 2]. Since the energetics of the system will be captured via a structure-based approached, we will not need to include solvent or ions. For this work, we will use the following potential from the Onunchic group (Figure 2) [3]:

To setup the input files needed for a structure-based simulation, please see the SMOG website (`http://smog-server.org`) [4]; however, for this tutorial, we have provided both the topology and the coordinate files.

If there are any questions or comments on this tutorial, please email jlai7@illinois.edu with the subject heading "Go tutorial - (name here)".

Required programs

In order to fully complete this tutorial, please make sure that you have the following software installed:

$$V = \sum_{bonds} \varepsilon_r (r - r_0)^2 + \sum_{angles} \varepsilon_\theta (\theta - \theta_0)^2 + \sum_{impropers/planar} \varepsilon_\chi (\chi - \chi_0)^2$$

$$+ \sum_{backbone} \varepsilon_{BB} F_D(\phi) + \sum_{sidechain} \varepsilon_{SC} F_D(\phi)$$

$$+ \sum_{contacts} \varepsilon_C \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - 2\left(\frac{\sigma_{ij}}{r} \right)^6 \right] + \sum_{non-contacts} \varepsilon_{NC} \left(\frac{\sigma_{ij}}{r} \right)^{12}$$

where,

$$F_D(\phi) = [1 - \cos(\phi - \phi_0)] + \frac{1}{2}[1 - \cos(3(\phi - \phi_0))]$$

Figure 2: Structure-based potential: The key features of this potential are: 1) the reference bond lengths, angles, and torsions are derived from the crystal structure, and 2) all long range interactions are replaced by a Lennard-Jones potential between native contacts in the crystal structure.

- **NAMD beta build**   (Required) Program used to run MD simulations. `http://www.ks.uiuc.edu/~jim/tmp/NAMD_build.latest/`
- **VMD**   (Required) Program used to view MD simulations. `http://www.ks.uiuc.edu/Research/vmd/`
- **Any file archiver**   (Required) Program used to open zip files.
- X3DNA-DSSR   (Optional) Useful for calculating base pairing. `http://forum.x3dna.org/site-announcements/download-instructions/`
- Q   (Optional) Useful for calculating fraction of native contact. `http://vidar.scs.uiuc.edu/jlai7/Tutorial/Libbiokit.zip`

Note that the only required software are NAMD, VMD, and any file archiver that can read zip files; however, the other software will make life easier (ranked in decreasing importance).

2   How do we run a structure-based simulation?

Before one can run a structure-based simulation, one should make sure that the above software is installed and added to the $PATH variable. Because the NAMD beta build is different from the NAMD executable that one downloaded previously, please go to `http://www.ks.uiuc.edu/~jim/tmp/NAMD_build.latest/` to download the newest version. For the rest of the tutorial, we will be using the terminal (or command prompt in Windows), so please launch the terminal now. On a Linux/Mac/Windows Cygwin system, one should type the following:

```
$which namd2
```

On Windows XP and earlier versions, one should type:

```
C:\>for %i in (namd2.exe) do @echo.   %~$PATH:i}
```

On a Windows 2003 or later, one should type:

```
C:\>where namd2.exe
```

If the command returns something similar to:
``/Volumes/HomeRAID/Homes/jlai7/Programs/NAMD2/namd2'', then one should be ready to go. Otherwise, please go back to the NAMD tutorial for more information on how to add NAMD to the path. It is also advisable that one checks if the other programs are added to the $PATH.

After software is installed...

After checking that NAMD is installed, download the `http://vidar.scs.uiuc.edu/jlai7/Tutorial/GoDemo.zip` and extract the files using one's file archiver of choice. Then proceed to navigate to the extracted directory. Once in the extracted

directory, one will notice that there are five files of interest (and a directory with the intermediate files):

- REF.pdb

  A PDB file of the h18 pseudoknot from the PDB 2I2P with *E. coli* numbering. This is also used as an input to tell NAMD how to restrain the center of mass as well as how to apply the Berendsen thermostat.
- Starting.h18.gro

  A Gromacs coordinate file for the h18 pseudoknot and contains the same information as REF.pdb
- Reference.h18.top

  A Gromacs topology file for the h18 pseudoknot. This has all of structure based potential information.
- Starting.h18.gro

  A Gromacs coordinate file for the h18 pseudoknot and is unfolded.
- go.namd

  A NAMD configuration file used to describe the simulation to be perform.

Please type (On Linux/Mac/Windows) "more go.namd" to read through the go.namd file. Below, the text of go.namd is also provided.

Configuration file

```
###############################################################
## ADJUSTABLE PARAMETERS                                     ##
###############################################################

set outputname      namd2
set temperature     60

gromacs     on
grotopfile      Reference.h18.top
grocoorfile     Starting.h18.gro

###############################################################
## SIMULATION PARAMETERS                                     ##
###############################################################

# Force-Field Parameters
cutoff              15
switching     off
pairlistdist        17

# Integrator Parameters
timestep            0.5
nonbondedFreq       1
stepspercycle       20
staticAtomAssignment yes
twoAwayX            no
twoAwayY            no
twoAwayZ            no

# Constant Temperature Control - Berendsen thermostat
tCouple             on
tCoupleTemp         $temperature
tCoupleFile         REF.pdb
```

```
tCoupleCol          B
temperature      $temperature
seed     151069

# Fixed atoms - to prevent system from flying away
fixedAtoms      true
fixedAtomsForces    on
fixedAtomsFile      REF.pdb
fixedAtomsCol     O

# Disables NAMD internal check that assumes all 1 amu atoms are hydrogens
ignoreMass      true

# Turns on GromacsPair code to read in pair potentials
GromacsPair true
```

The biggest difference between the go.namd and most normal MD runs are the following lines:

- margin 1000   make sure that all nodes have the full set of atoms
- fullElectFrequency 5500000   because electrostatics are not needed for the structure based model, we can disable the electrostatic code by making it run VERY infrequently
- ignoreMass true   Disables NAMD internal check that assumes all 1 amu atoms are hydrogens
- GromacsPair true   Turns on GromacsPair code to read in pair potentials

For the astute molecular dynamic users, we are going to run a constant temperature *in vacuo* simulation. One should notice that the Berendsen thermostat and fixed atoms are used. The reasoning for this is that the Langevin thermostat tends to slow down the dynamics of a simulation and would make observing the transitions difficult in the timespan of a single workshop. To make sure that the system does not fly off into space, the center of mass of the folded pseudoknot is restrained using fixedAtoms.

To launch the simulation, enter the following command:
(On Linux/Macs)

```
$namd2 go.namd >LOG &
```

(On Windows)

```
C:\start \B namd2 go.namd >LOG
```

As the simulation runs, we will work on the workflow to analyze the system.

3 Analysis

The goal for this tutorial is to understand the main topological barrier to the folding of the h18 pseudoknot. To do so, we will use a combination of RMSD and fraction of

native contacts to determine when the topological barrier occurs in the MD simulation
and then analyze which base pairs form around that time. To do this, first launch VMD
and load up the trajectories. One can do this by typing the command:
(On Linux/Macs)

```
vmd -dispdev win REF.pdb namd2.dcd
```

On Windows, one will have to use the GUI and the standard file input.

Once VMD launches, a blank screen might appear. This might occur because the
molecule has migrated off of the screen. The first thing one should do is to center the
molecule by pressing the "=" key on the keyboard. After the molecule is centered,
please use the analysis.tcl script to get RMSD (and optionally fraction of native con-
tacts (Q), and base-pairing). You can do this by launching the Tcl/Tk console and
typing the following:

```
source analysis.tcl
```

The analysis script is reproduced below:

Analysis script

```
set all [atomselect top all]
set refall [atomselect top all]
set nf [molinfo top get numframes]
set wp [open "RMSD" "w"]
###############################################################################
## To turn on the X3DNA and Q commands, uncomment the following lines
## by removing the proceeding "#" characters after this comment
###############################################################################
#set wpA [open "DSSR.log" "w"]
#set wpB [open "Q.log" "w"]
$refall frame 0
for {set i 1} {$i < $nf} {incr i} {
    $all frame $i
    $all writepdb temp.pdb
    $all move [measure fit $all $refall]
    puts $wp [measure rmsd $refall $all]
    #puts $wpA [concat $i [exec -ignorestderr x3dna-dssr -i=temp.pdb --OUTPUT=DSSR.$i.log]]
    #puts $wpB [concat $i [exec -ignorestderr q native_contacts -c 4 -s 2 -smax 1000 \
# -d 1.2 -ct1 RNA -r 1 temp.pdb REF.pdb | grep "Fraction" ]]
}
close $wp
#close $wpA
#close $wpB
```

Afterwards, you should get a log file containing the rmsd. If you have included Q
and the base-pairing lines, you will get individual log files that you will need to parse
before plotting (Figure: 3).

Now that we have several reaction coordinates, what do we see?

In the sample trajectory provided, one can see a transition in the h18 pseudoknot at
around frame 300. To see which bases are giving rise to this effect, one might want to
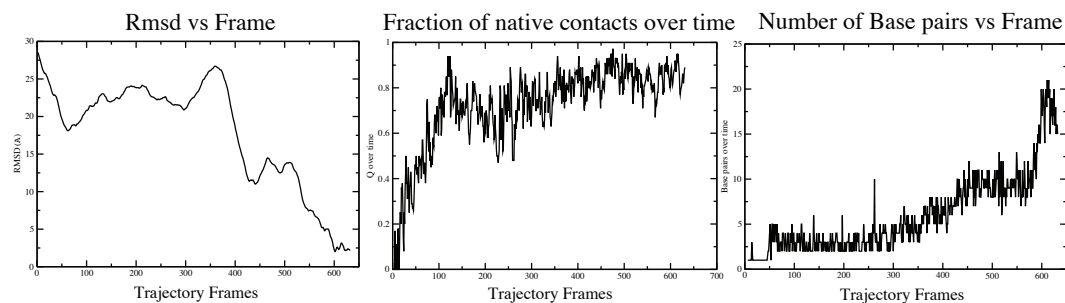consider measuring:

Figure 3: Plot of RMSD, fraction of native contacts (Q), and number of base pairs over time

- Distance between "residue" 15 and 37
- Distance between "residue" 5 to 26
- Distance between "residue" 6 to 25
- Distance between "residue" 7 to 24

Note that these are distances between "residues" and NOT "resid".

4   Conclusion

After completing the analysis, please plot the data using any graphing software and move on to the next tutorial.

References

[1] Ke Chen, John Eargle, Jonathan Lai, Hajin Kim, Sanjaya Abeysirigunawardena, Megan Mayerle, Sarah Woodson, Taekjip Ha, and Zaida Luthey-Schulten. Assembly of the Five-way Junction in the Ribosomal Small Subunit using Hybrid MD-Gō Simulations. *J. Phys. Chem. B*, 116(23):6819–31, June 2012.

[2] Jonathan Lai, Ke Chen, and Zaida Luthey-Schulten. Structural intermediates in the early assembly of the ribosomal small subunit. *J. Phys. Chem. B*, 42:1333513345, 2013.

[3] Paul C. Whitford, Alexander Schug, John Saunders, Scott P. Hennelly, Jos N. Onuchic, and Kevin Y. Sanbonmatsu. Nonlocal helix formation is key to understanding s-adenosylmethionine-1 riboswitch function. *Biophys. J.*, 96(2):L7 – L9, 2009.

[4] Jeffrey K Noel, Paul C Whitford, Karissa Y Sanbonmatsu, and José N Onuchic. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucl. Acids Res.*, 38(Web Server issue):W657–61, July 2010.