Biomathematics: Proteins, Graphs and Clocks, Oh My!

Tanya Leise and Amy Wagaman

Department of Mathematics

Outline

- Introduction: What is Biomathematics?
- Biomathematics in the Pioneer Valley
- NSF Funding
- Investigating Protein Folding Landscapes with Sheila Jaswal
- Applications of Statistics in the Protein Folding Problem with Amy Wagaman
- Investigating Circadian Clocks with Tanya Leise

What is Biomathematics?

- Intersection of biological research and mathematics & statistics.

- Mathematical modeling can reveal properties and structures of biological systems.

- Models must be simplified versions of reality, ignoring many details and complexities.

- However, without mathematical models and statistical and computational tools to discover and test principles underlying biological systems, progress would be greatly limited.

All models are wrong, but some are useful.

--George E. P. Box

Why should students pursue Biomathematics?

- Biologists generate enormous amounts of data.
- Biologists can observe some patterns in their data, but need tools to quantitatively detect, describe, and validate them, for example, patterns buried in huge datasets.
- The American Association for the Advancement of Science strongly recommends quantitative, analytical, and modeling skills for biology majors.
- For mathematics students, biomathematics offers exciting opportunities to see math/stats used to solve real problems in neat applications.

April 2011 – Mathematics Awareness Month

- Mathematics, Statistics, and the Data Deluge

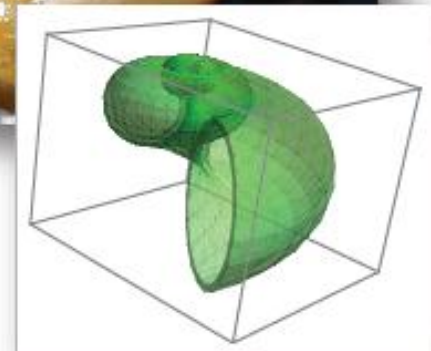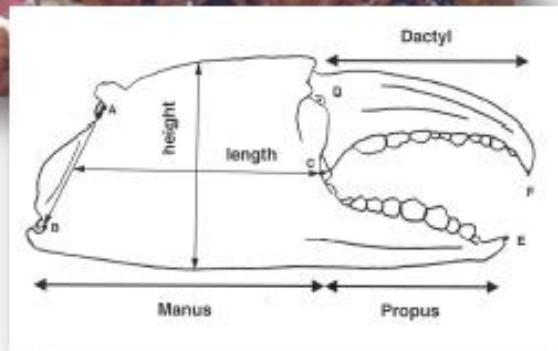# Biomathematics in the Pioneer Valley

- Biomathematics is strong and active in the Valley.

- Smith College now has a Biomathematics Concentration.

- Recently funded NSF grant to the Four Colleges has highlighted biomathematics projects and encouraged new collaborations.

- Six Amherst students participated in the first Frontiers in Biomathematics course (held in the evenings at Smith) this past semester.

Biomathematics in the Pioneer Valley

Dynamics of a crab and snail arms race

David Smith, Biological Sciences & Christophe Golé, Mathematics & Statistics

Smith College

Goals: Experimentally test and mathematically model patterns of growth and predator- and diet-induced responses between an invasive crab and a native snail

Biomathematics in the Pioneer Valley

How to turn a mechanical lizard model into a truly autonomous robot

Green anole (*Anolis carolinensis*) dewlapping

GOAL: We would like to add computer vision to our robotic lizard (bottom images) so that it can recognize the live lizard's dewlap display (top images) and respond autonomously.

Sarah Partan from Hampshire College

NSF Initiative

From the National Science Foundation (NSF):

"There has been an **explosion of knowledge in the life sciences** over the past twenty years that cuts across all levels from molecules to ecosystems. Current research is often characterized by **integrative and interdisciplinary** approaches. At the center of this explosion of knowledge is a revolution in instrumentation, **computational abilities, information systems, and mathematical tools**."

[...] However, currently there are comparatively **few people able to work in this intersection**. The UBM *[NSF]* program aims to **transcend traditional boundaries** in educating **biological and mathematical scientists**.

NSF Grant for Student Biomath Research

- 4 College NSF grant awarded in 2011
- Designed as Junior-Senior biomath research training program
  - Covers junior and senior years and summer in between
  - Students are paired (as are the faculty) with one from the life sciences and one from the quantitative sciences

Amherst Students in Biomath

- Ian Mellis, Class of 2011, wrote an interdisciplinary thesis in Bioinformatics.

- Michelle Li, Class of 2012, wrote an interdisciplinary thesis in Biomathematics.

- Shennon Lu, Class of 2014, worked with Amy and Sheila on a comparison of methods for computing protein unfolded surface area.

- Winthrop Harvey, Class of 2013, worked with Tanya on wavelet analysis of mouse activity patterns.

Proteins

- Perform vital functions
- Made up of a linear sequence of amino acids
- Function is determined by protein shape and motion
- The *protein folding problem* refers to the problem of determining a folded structure from a list of amino acids
- Many proteins fold spontaneously, some need help from other proteins
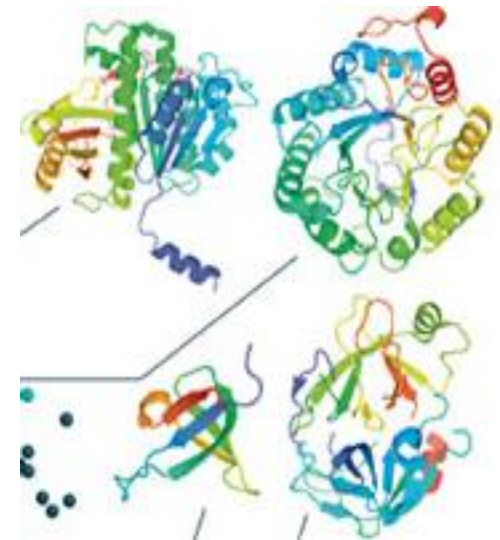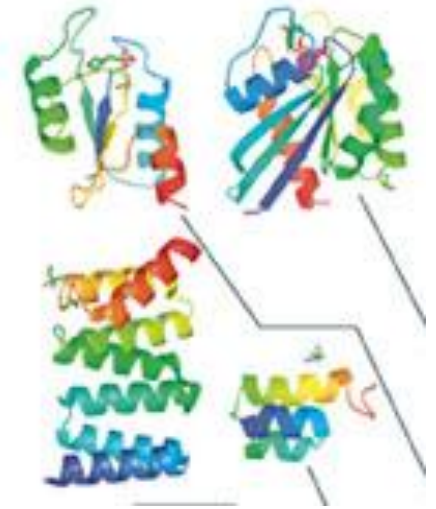
Protein Folding and Disease

Diseases due to inability to fold

Protein involved	Disease
CFTR	Cystic fibrosis
p53 ("guardian of the genome")	Cancer
α-fetoprotein	Down's Syndrome
Superoxide dismutase	Lou Gehrig's Disease

Diseases due to toxic misfolding

Protein involved	Disease
Crystallin	Cataracts
$A\beta$	Alzheimer's
Prion	Mad Cow Disease
Amylin	Type II Diabetes
$\beta 2$-microglobulin	Dialysis-related Amyloidosis

folds very slowly

CFTR

most degraded

Most common cystic fibrosis mutant

small active fraction

partial unfolding

Sticky intermediate

over time

protein in **Dialysis Related Amyloidosis**

amyloid fibers

9
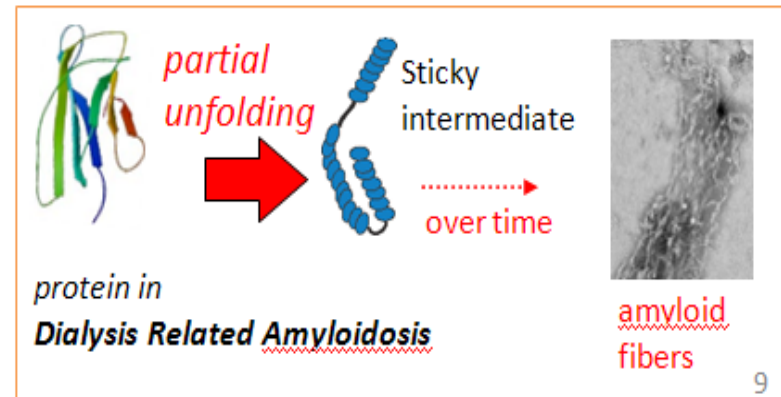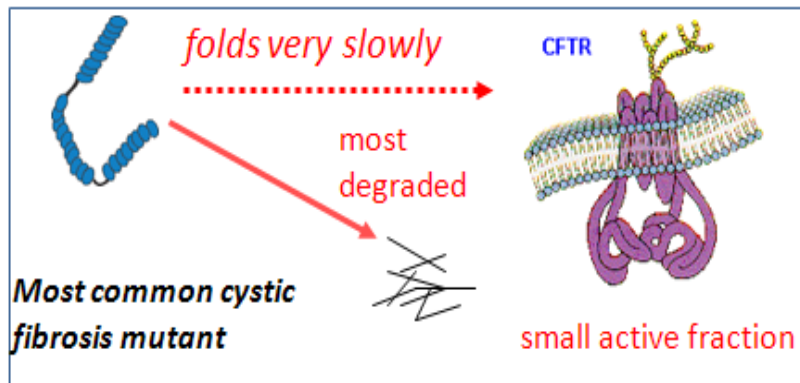
Investigating Protein Landscapes (Sheila Jaswal)

- Unfolded proteins are in a high energy state, and their folded states are (except for some rare exceptions) much lower in energy.

- But how do proteins know what to fold into and how do they do it?

Random search

100 aa protein : 10^{10} years!

© www.SciStyle.com
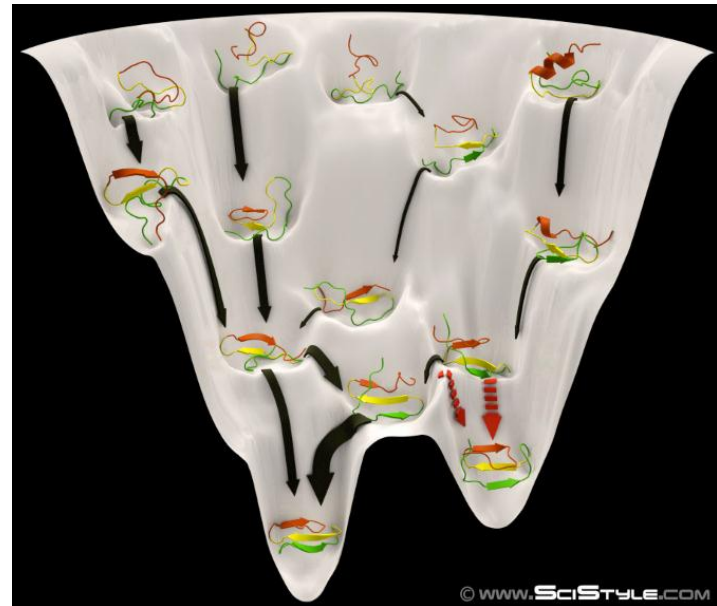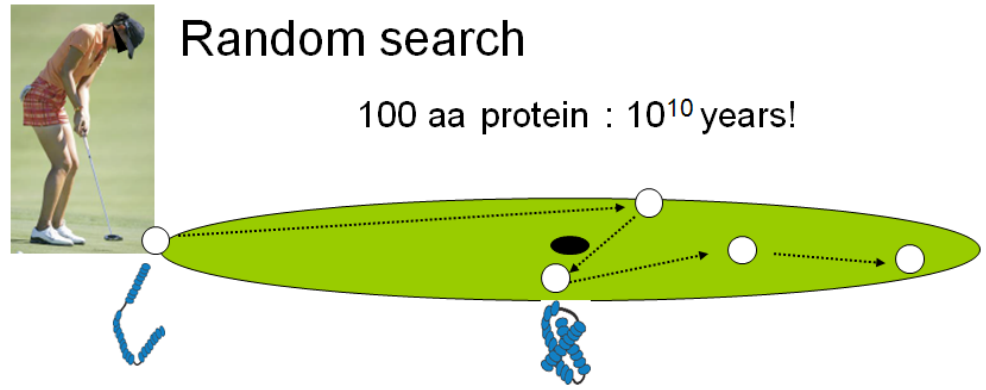
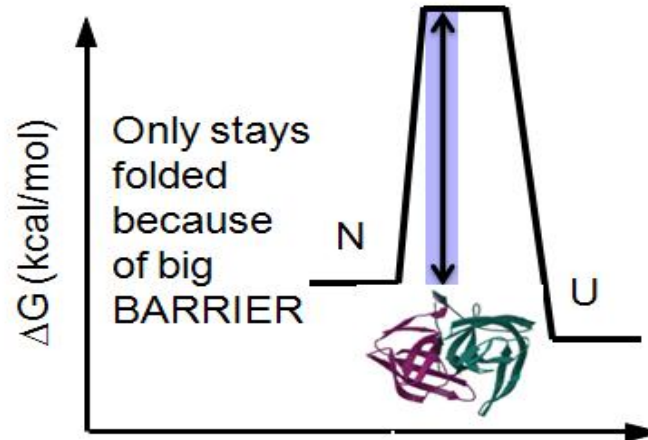# Typical Protein Landscape
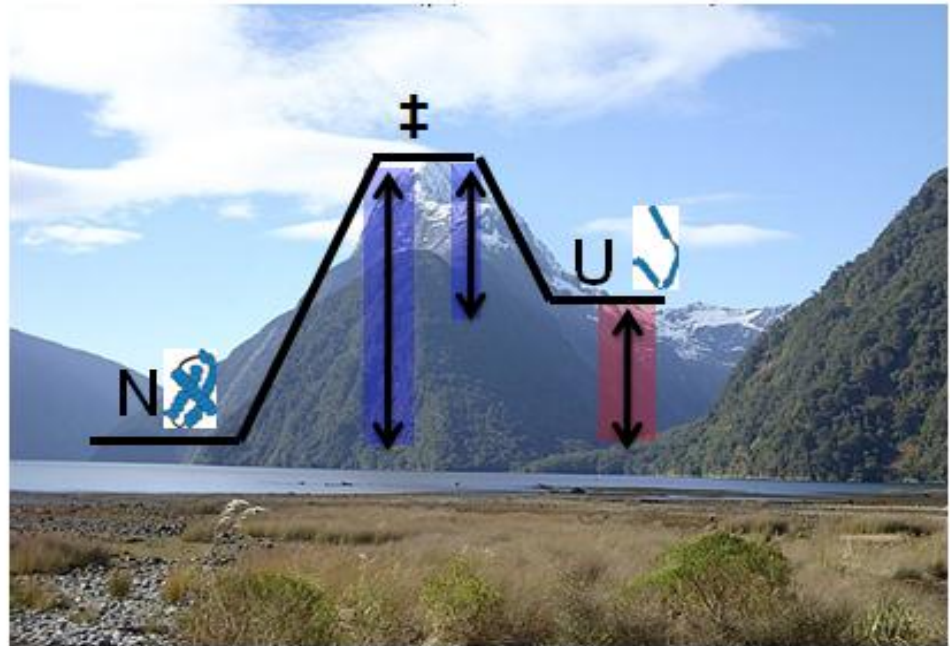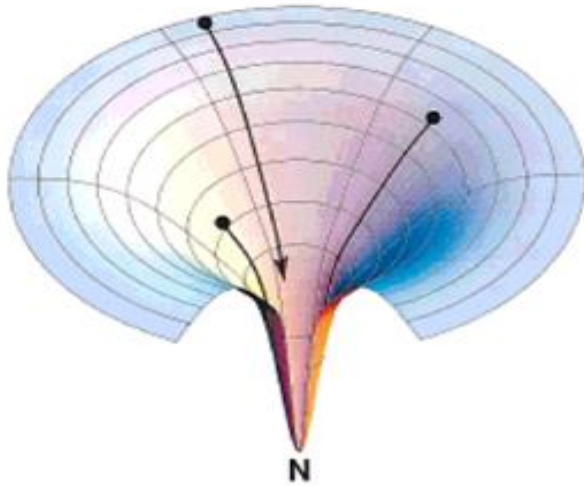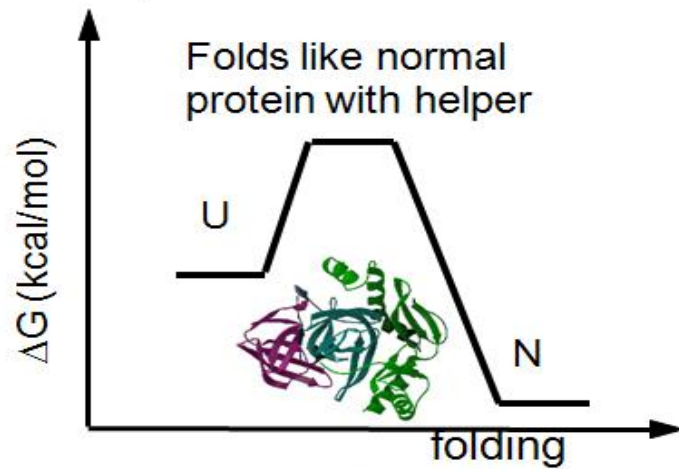

Folding is generally downhill (to lower energy).

Can determine landscape by monitoring unfolding / folding in experiments

Heat/denaturant must be added for proteins to unfold, but then refolding is spontaneous once the proteins are cooled/denaturant removed.
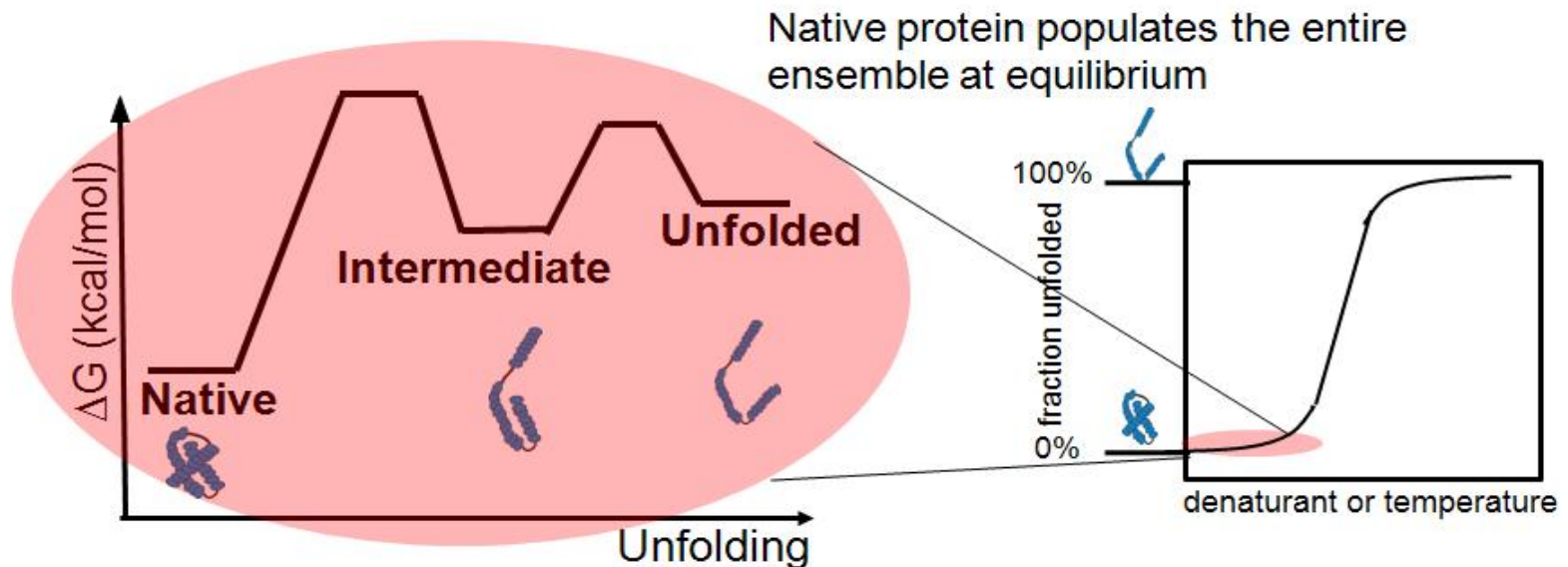
Not so Typical Protein Landscape
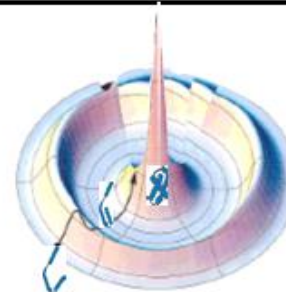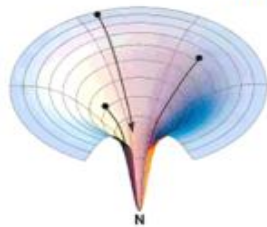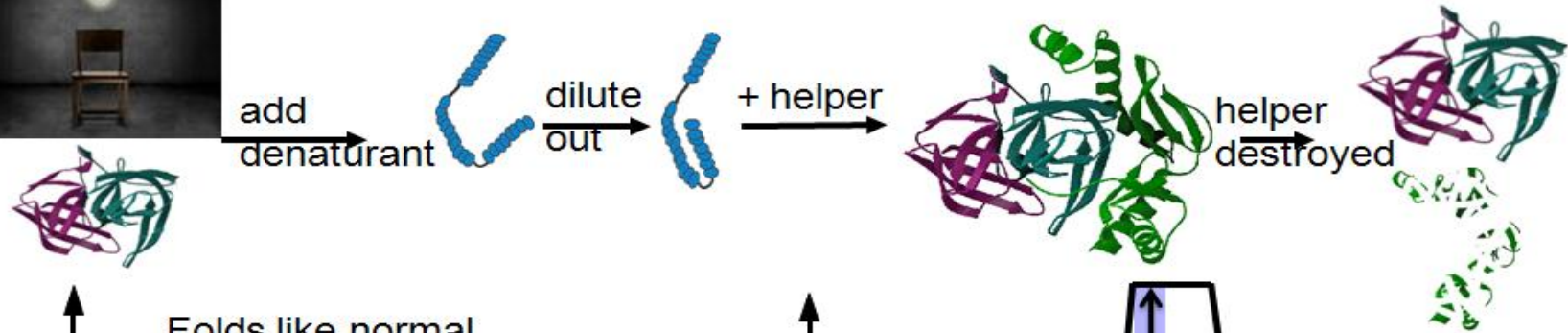
New Landscape Mapping Tools

Landscape-mapping tools:
• Patient stalking vs. harsh interrogation

Hydrogen Exchange Mass Spectrometry:
(How to interrogate proteins without torturing them)

Native protein populates the entire ensemble at equilibrium

Applications of Statistics in the Protein Folding Problem (Amy Wagaman)

- Collaboration with Laufer Center at Stonybrook University to help with prediction of folded protein structures via clustering and dimension reduction methods

- Collaboration with Sheila Jaswal to investigate characteristics of proteins and identify patterns in folding/unfolding behaviors

- Viewing proteins as graphs is relevant to both collaborations

Proteins as Graphs

- A graph is a collection of:
  - vertices/nodes (V)
  - edges (E)
- We can build graphs related to proteins in many ways.
  - Protein-protein interaction networks
  - Graph of protein differences based on some distance measure
  - Graph for an individual protein based on 3-D structure
- There are several choices to make when building each graph.
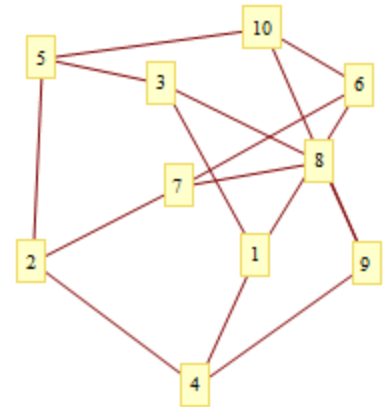
Clustering in Protein Structure Prediction

- For protein structure prediction, a simulation is run which generates thousands of possible structures for the protein.

- From those thousands, how do you determine what the likely folded protein structure is?

- The researchers believe the folded state occurs fairly often in the simulations, so you want to find a large group of similar structures.

- This is done via clustering.

Graph Partitioning for Clustering

- Many different clustering algorithms exist and have been tried on these simulated protein trajectories, but the typical ones used are not graph-related.

- As part of my collaboration with the Laufer group we are trying spectral clustering and some additional cluster pruning methods to try to get better results

- Spectral clustering can be viewed as a graph partitioning method.
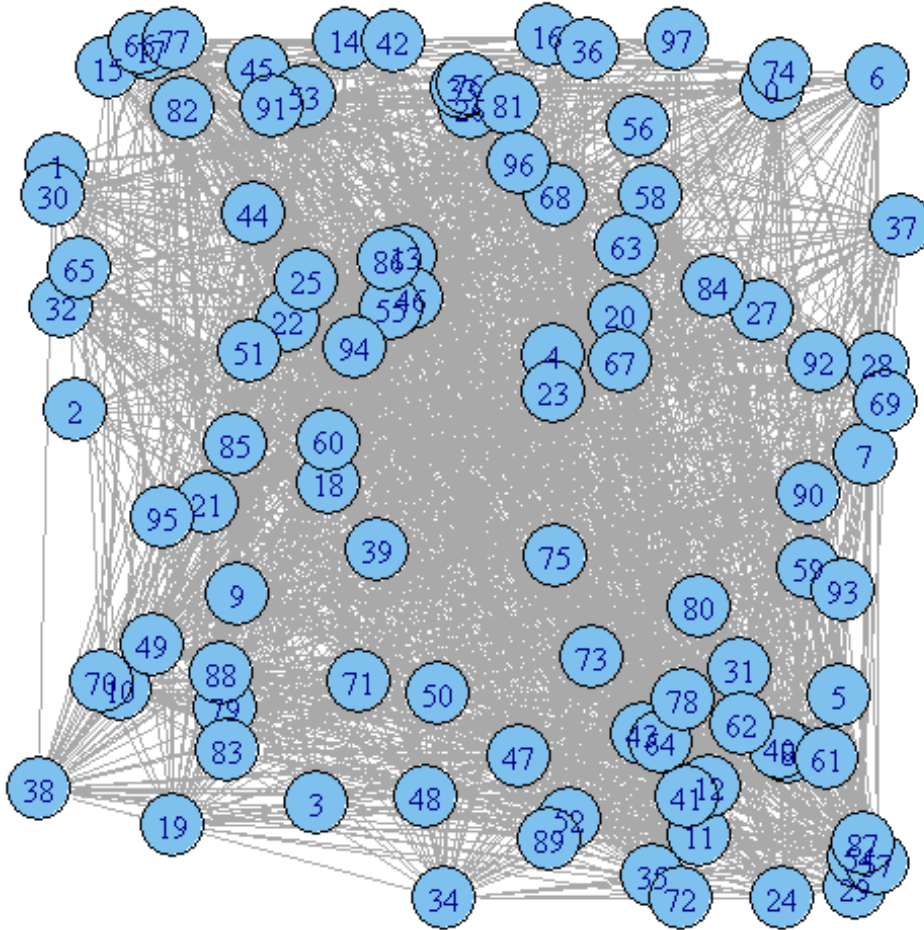 - Example: Shi-Malik Normalized Cuts algorithm

Example Clustering Results
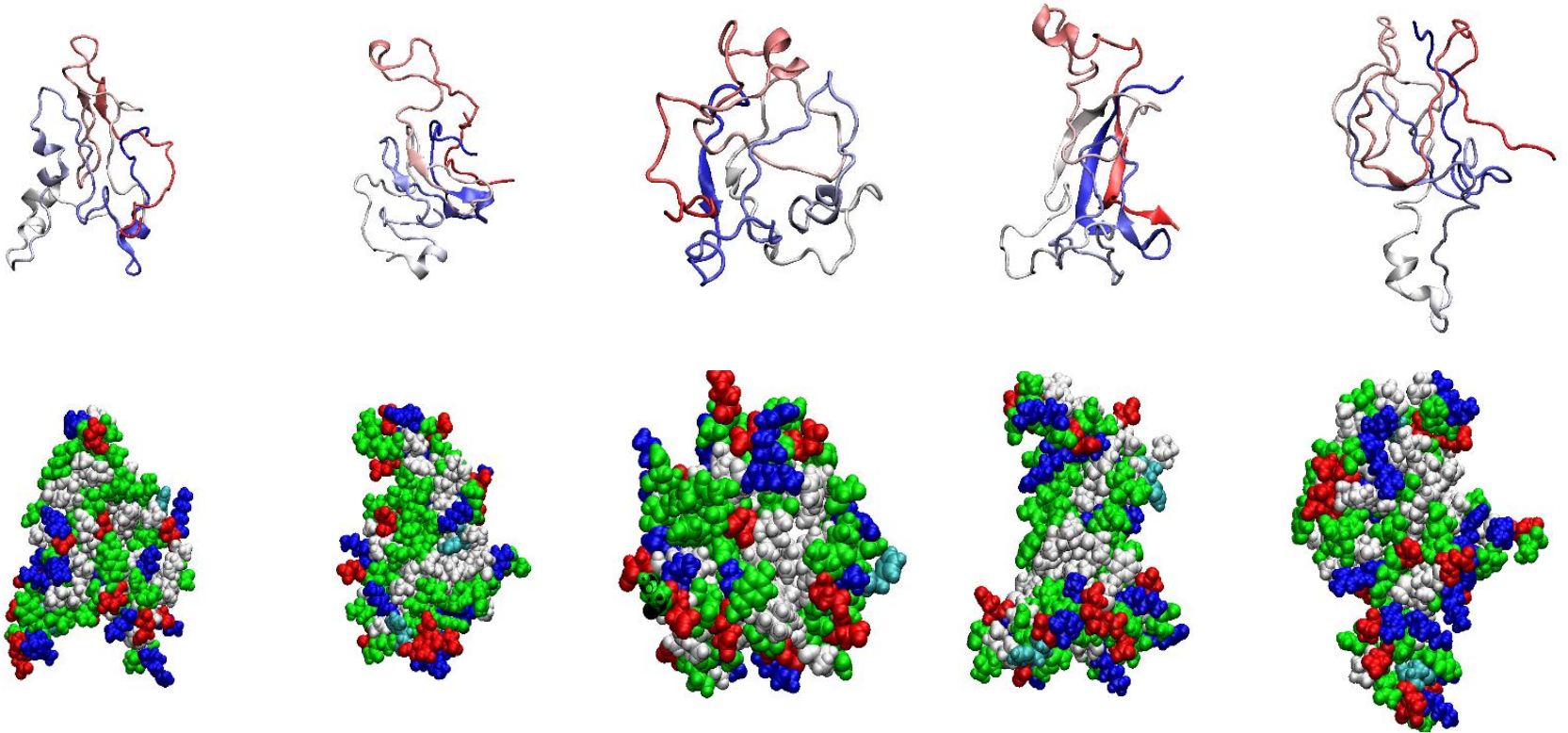
Representative protein structures from the largest 5 clusters from a molecular dynamics simulation of several thousand frames

Graphs for Individual Proteins

- This is a graph formed using protein 3-D structure. An edge is placed if a pair of non-hydrogen atoms from the amino acids in question are within 8 Angstroms when the protein is folded.
- This is 1APS. It is 98 amino acids in length.

Example Properties of Graphs

- Degree – number of edges each vertex has

- Can compute average degree, or look at degree distribution

- Measures of centrality, rigidity, stability

- Investigating how these properties are related to folding/unfolding properties

Histogram of degree(Adj)
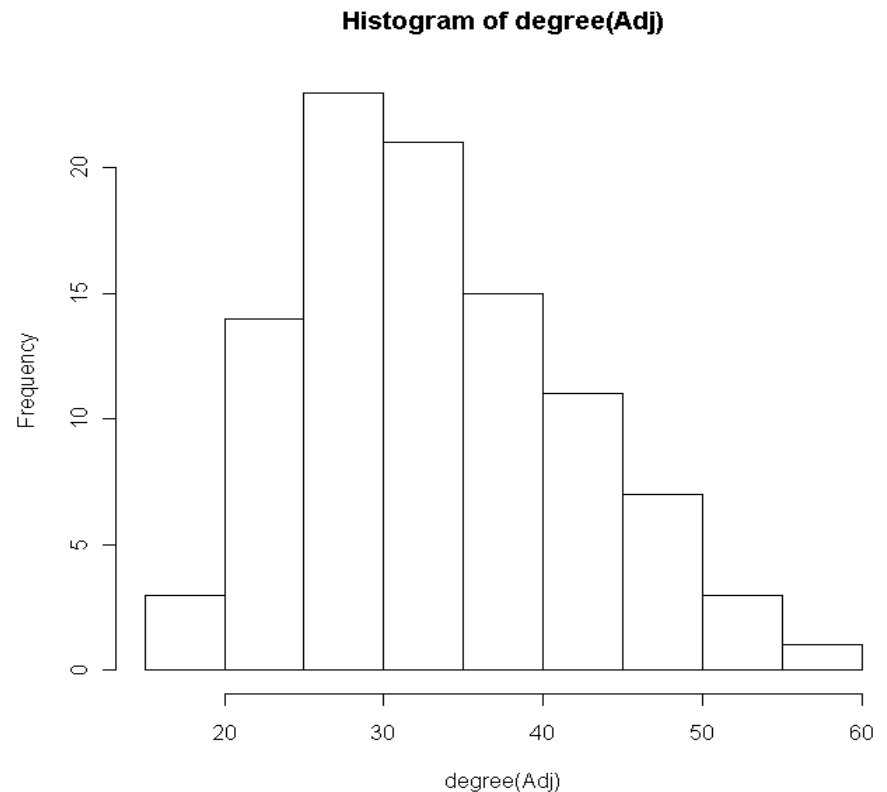
Future Research Directions

- Continue building database of protein folding/unfolding information (with help of Amherst IT)

- Continue exploring relationships between protein structural/graph/composition properties and protein folding/unfolding properties

- Explore additional projects with Laufer Center where dimension reduction methods may help with data analysis
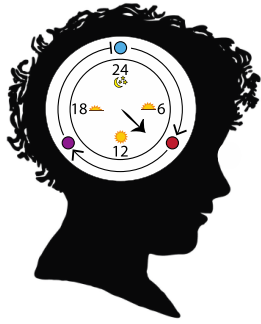
Biology + Mathematics (Tanya Leise)

- Interdisciplinary effort to understand a very complex physiological mechanism: the **internal circadian clock**

  – What is the circadian clock?

  – What is its role in our lives?

  – The importance of noise for the clock

What is the circadian clock?

- *Circa dies* = about a day

Midnight

Melatonin secretion starts

Lowest body temperature

24

Highest body temperature

18

Sharp rise in blood pressure

6

Greatest cardiovascular and muscle strength

Melatonin secretion stops

Best coordination

12

High alertness

Noon
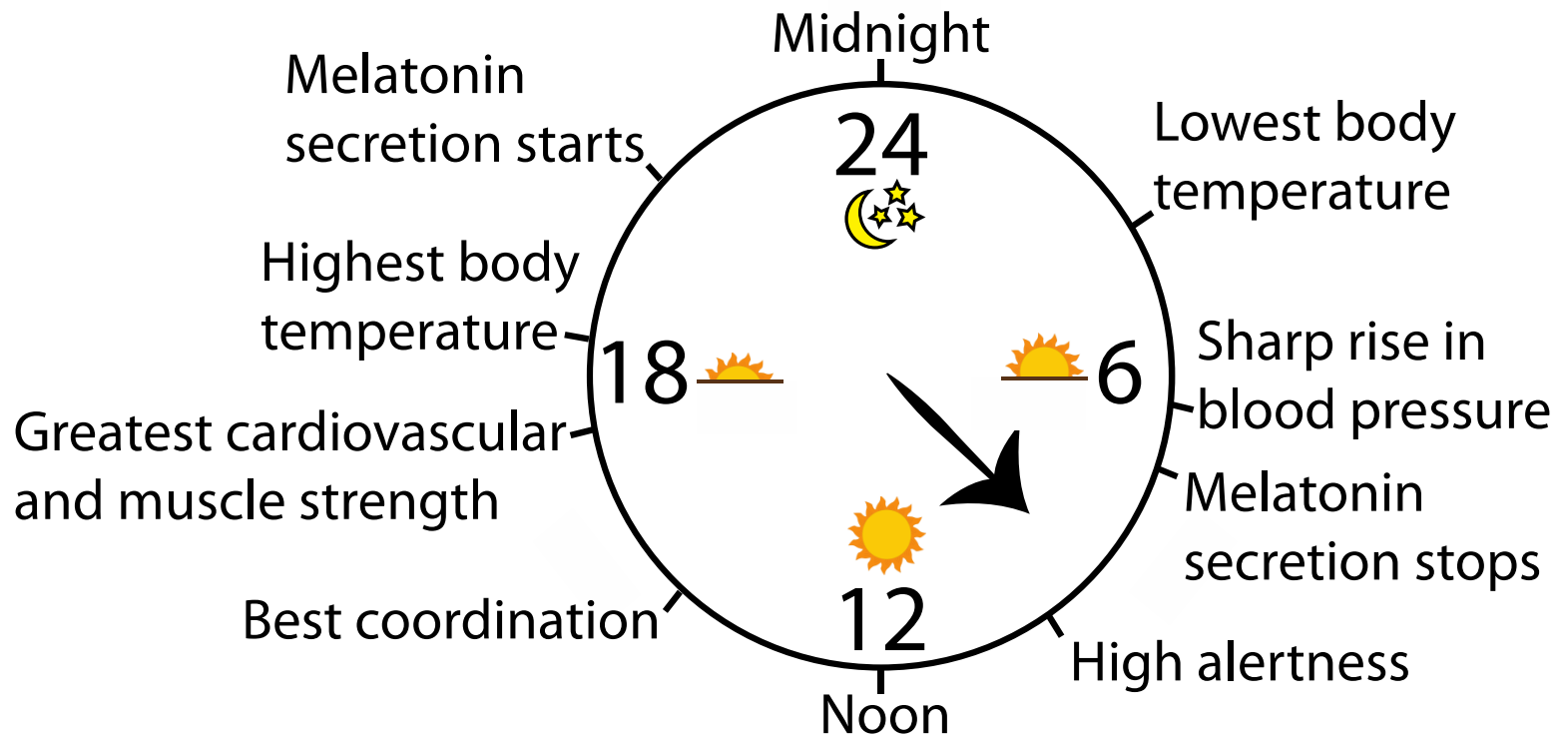
What is the circadian clock?

- Master pacemaker:
  - SCN in the hypothalamus

Light

Output rhythms:
Physiology
Behavior

Suprachiasmatic
Nucleus (SCN)

National Institutes of Health - NIGMS
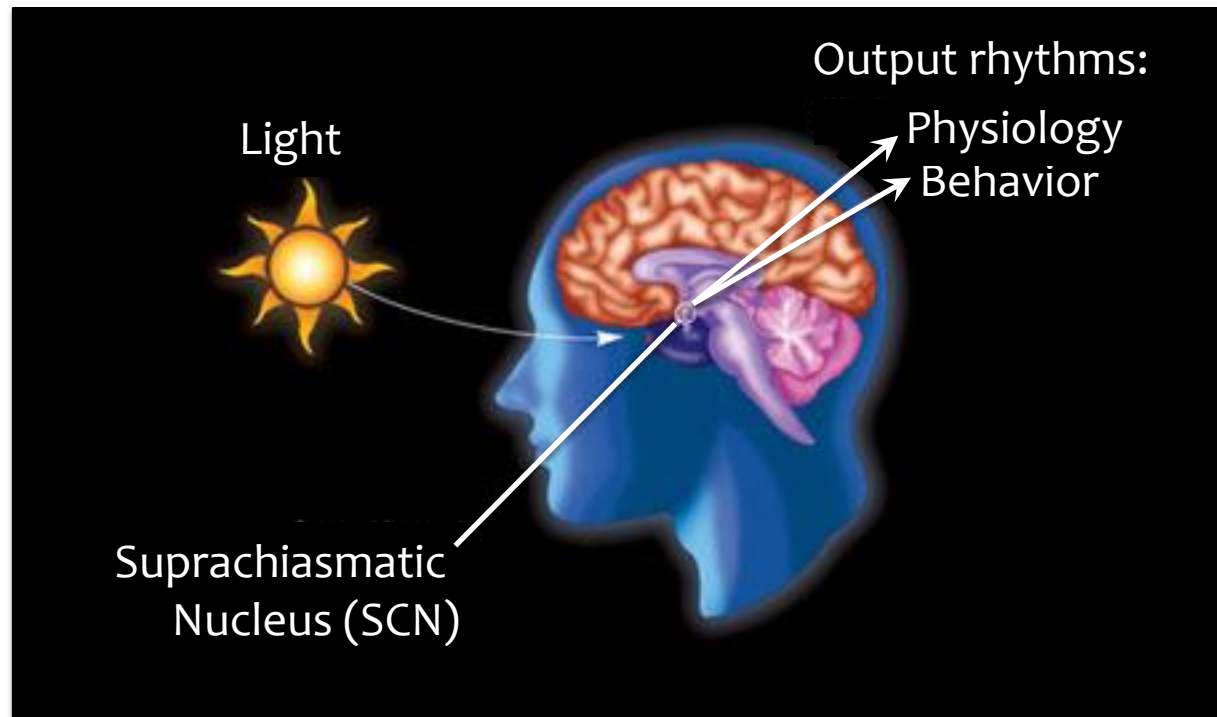
Circadian clock's role in our lives

xkcd

Effects of not entraining to the external time:

- Jet lag due to travel crossing time zones

- "Social jet lag" due to job shift mismatch with person's chronotype

Circadian clock's role in our lives

Foster & Roenneberg, 2008

MSF=mid-time of sleep on free days
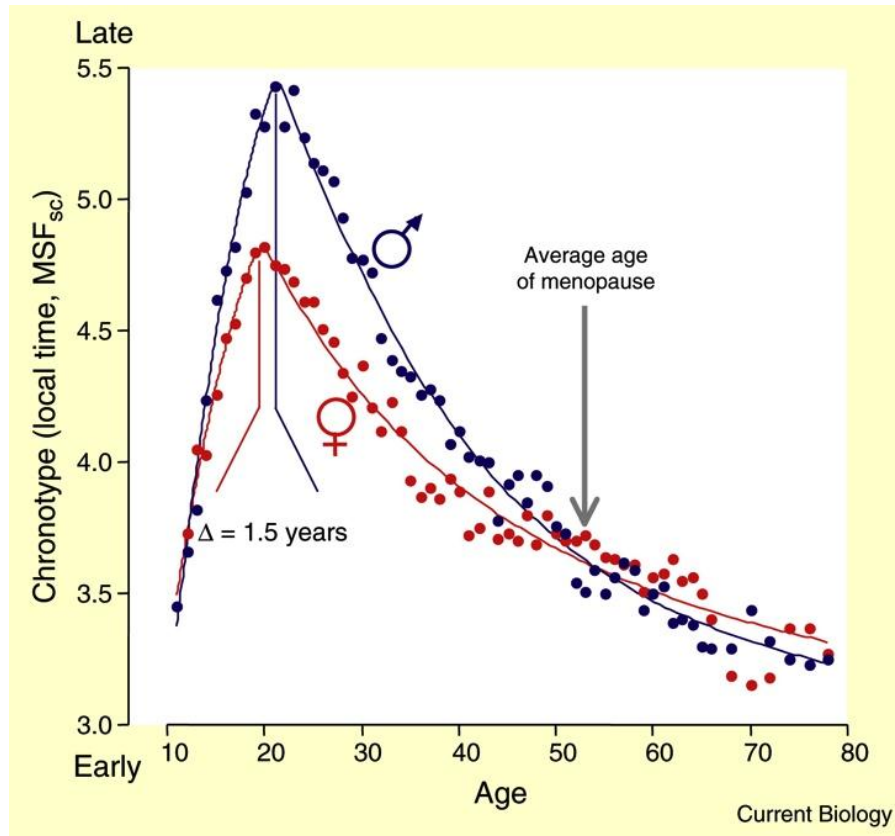
Social jet lag:
people sleep at different
times on the weekend
than during the week

Circadian clock's role in our lives

- Consequences of long-term lack of synchrony between our internal clock and external time:
  - Increased cancer risk
  - Increased cardiovascular disease
  - Increased metabolic syndrome
  - Difficulty sleeping (insomnia)
  - Deficits in cognitive function, long-term memory, and alertness
  - Reduced immune function
  - Contributes to depression and bipolar disorder

A mathematician's view of the clock

- Oscillations of clock gene expression generated by a feedback loop
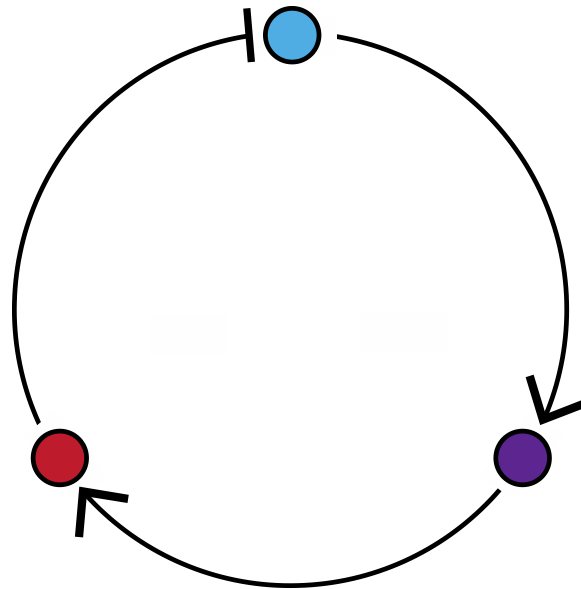
Negative feedback loop

- Foxes and rabbits analogy:
  - Low population of foxes leads to increasing numbers of rabbits

  - Large population of rabbits allows foxes to prosper and increase their numbers

  - Overpopulation of foxes impacts rabbit population, insufficient to sustain foxes

  - Fox population plummets, allowing rabbit population to recover, starting cycle over again

Negative feedback loop

Clock gene **protein**

Clock gene **mRNA**

6am 6pm 6am 6pm
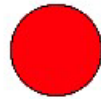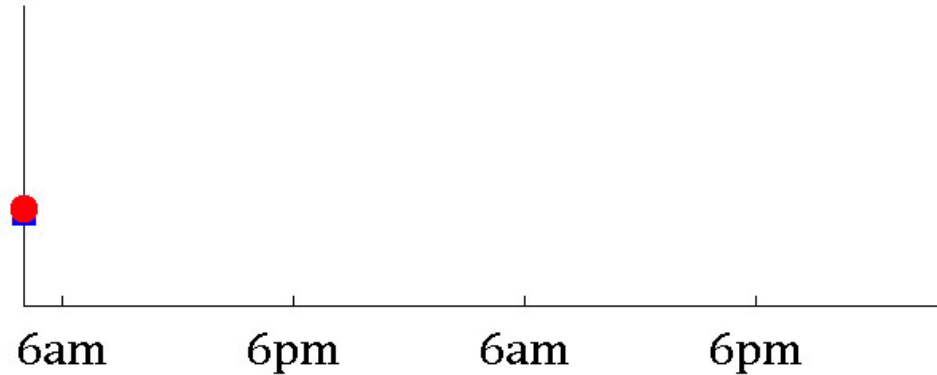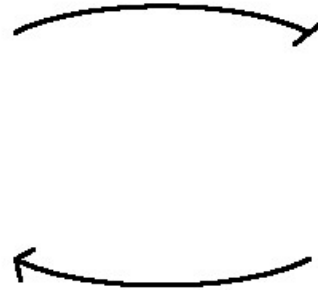
The role of "noise" in the clock
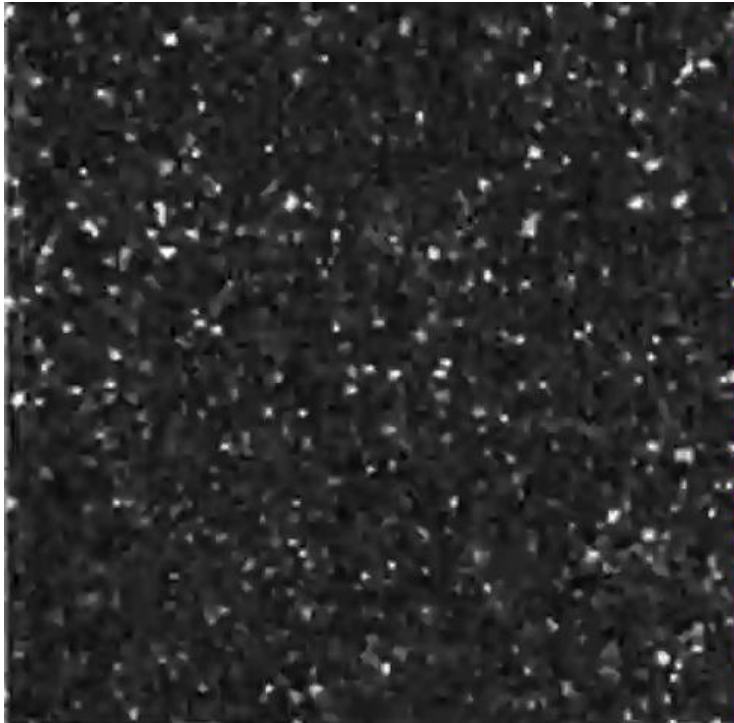
Oscillations in PERIOD2 (core component of internal circadian clock) over 42 days
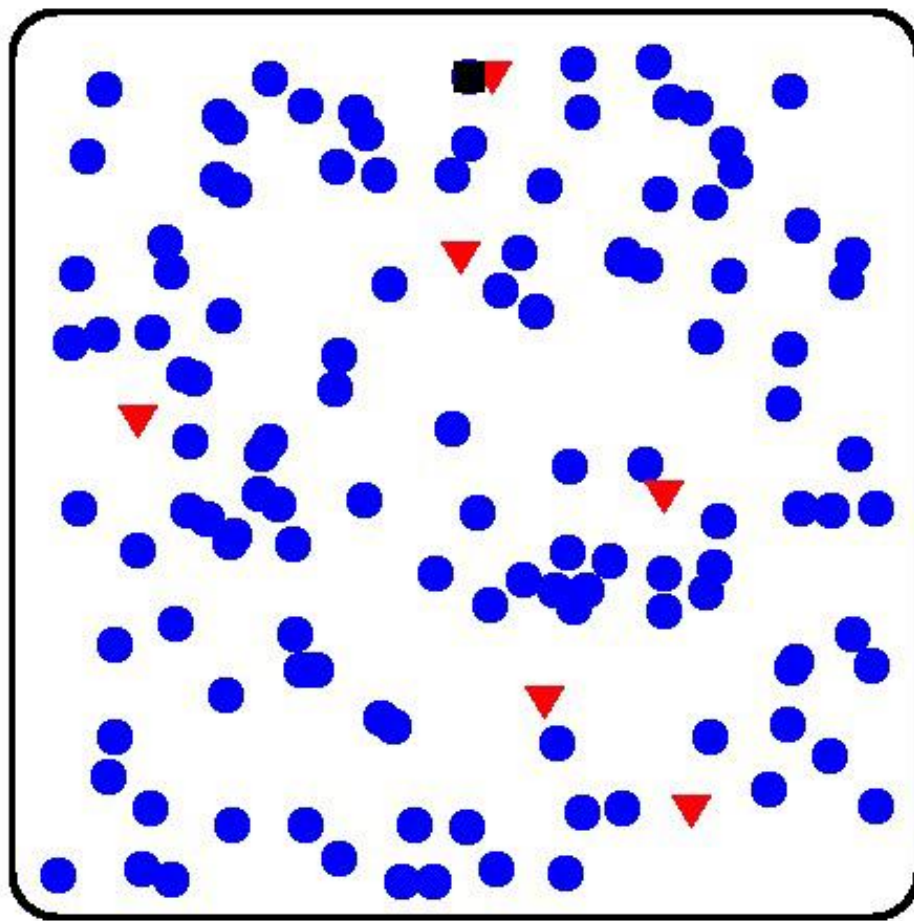
Fibroblast cells from mouse tails recorded in David Welsh's lab at U of California – San Diego (Leise at al. PLoS ONE 2012)
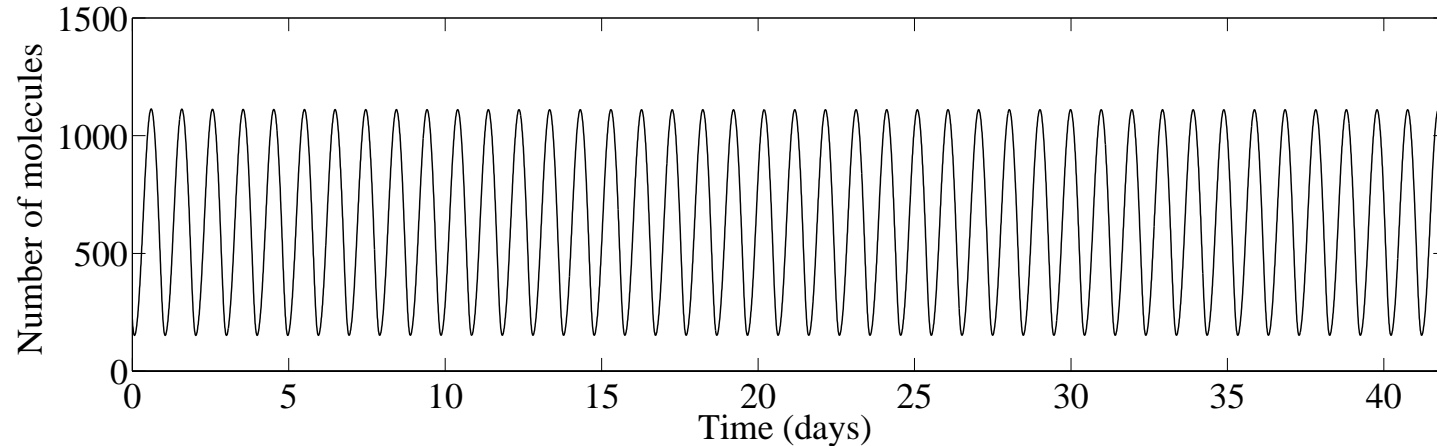
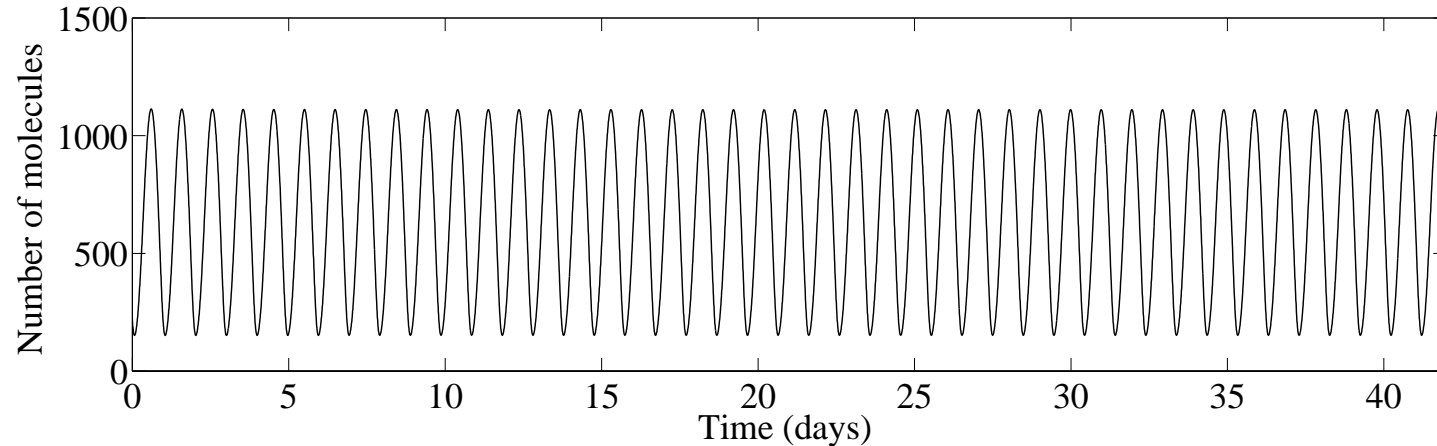# The role of "noise" in the clock
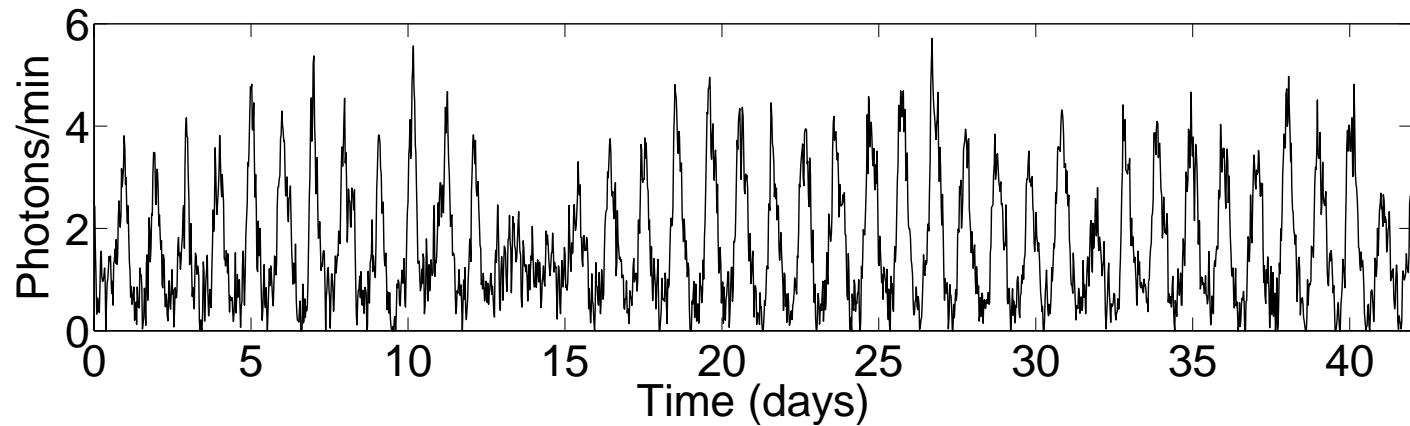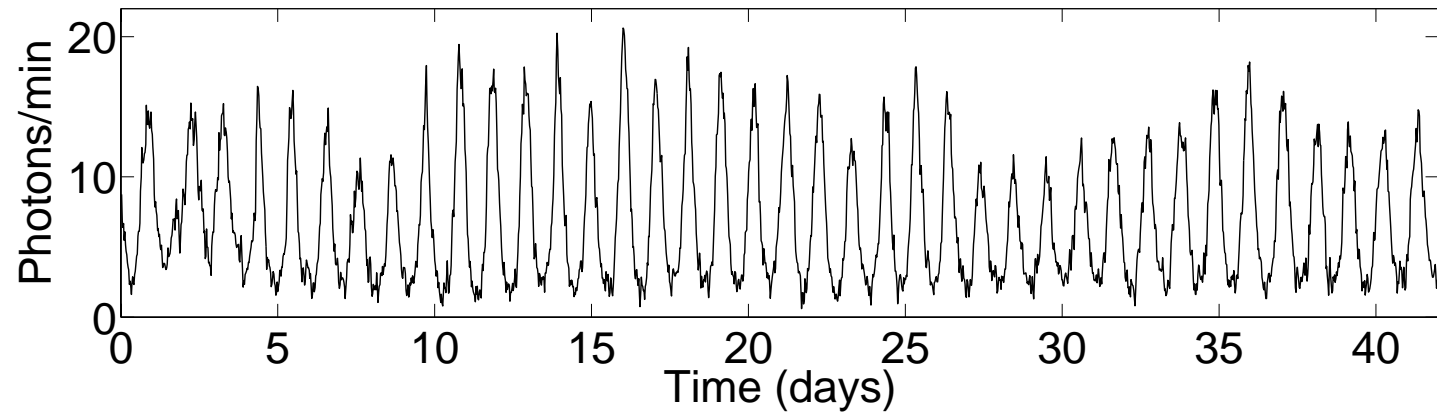
The role of "noise" in the clock

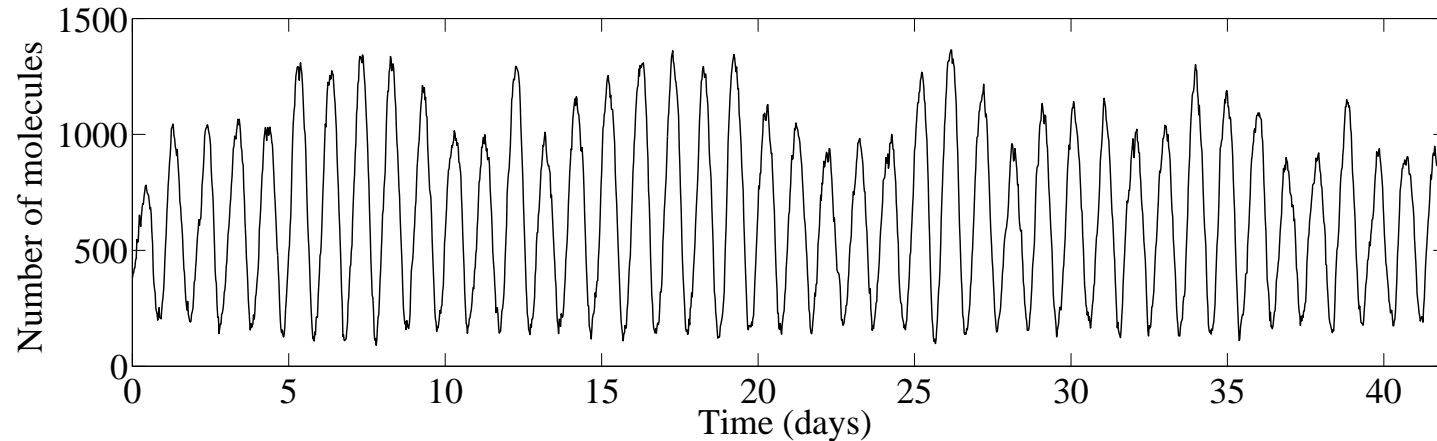# The role of "noise" in the clock

Idealized simulation:

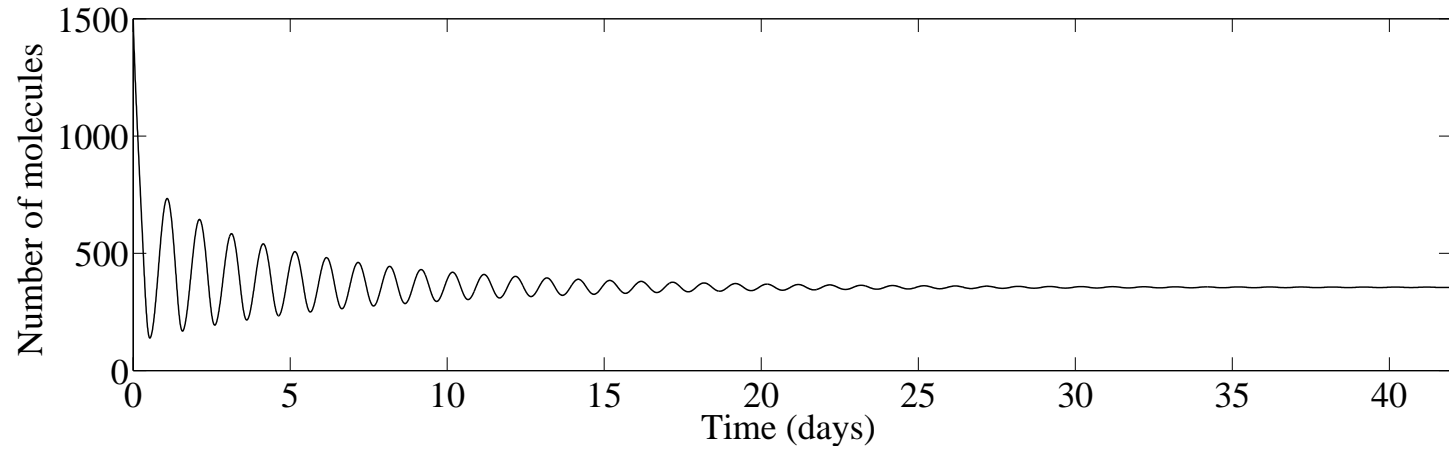# The role of "noise" in the clock

Idealized simulation:
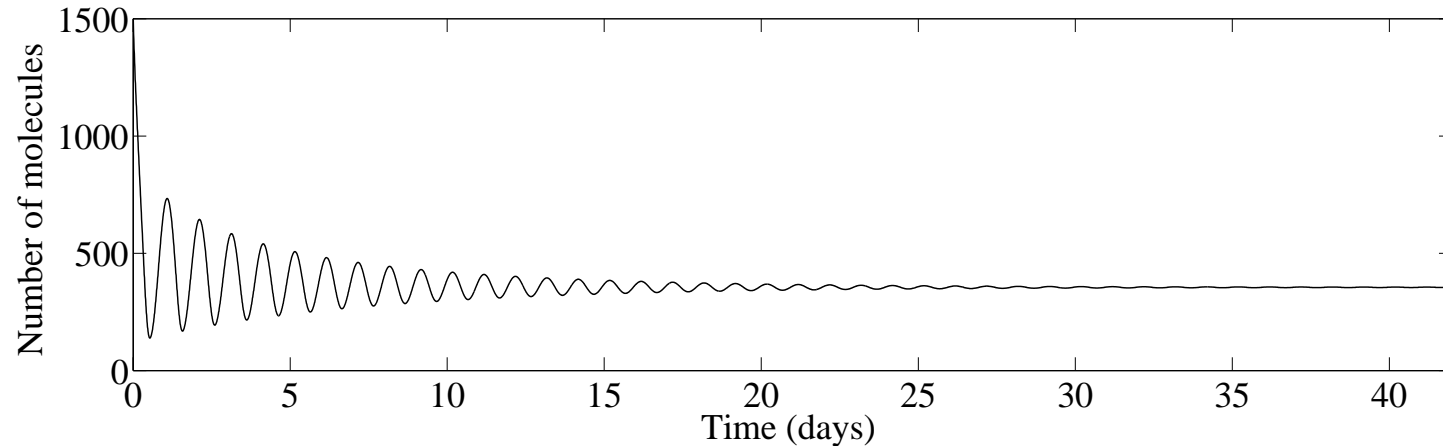

Stochastic simulation:

The role of "noise" in the clock
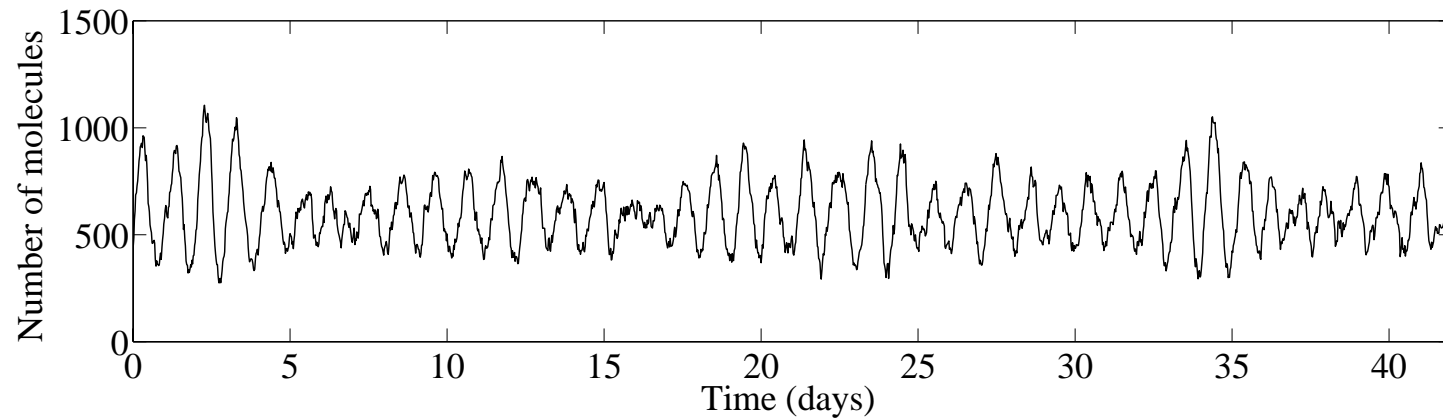
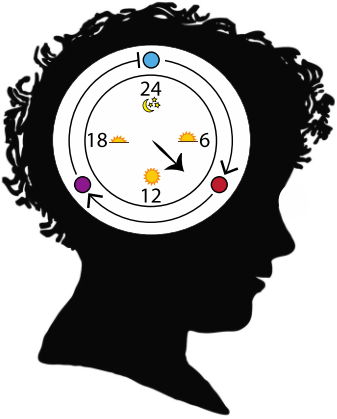Idealized simulation:

The role of "noise" in the clock

Idealized simulation:

Stochastic simulation:

Our Internal Clock

- We are hard-wired with an internal clock that coordinates daily internal rhythms.

- Disruption of this clock increases risk of significant health problems.

- Biologists working together with mathematicians have improved our understanding of the circadian clock and are developing ways to mitigate negative health consequences and to use the clock to improve medical treatments.

Concluding Remarks

- Biomathematics is an emerging field with lots of exciting opportunities for faculty and student research, as shown by these snapshots of biomathematics projects here at Amherst.

- Recent NSF funding and development of interdisciplinary collaborations has provided new opportunities for student research in intriguing and important problems.

- Thank you! We're happy to answer any questions.