# CS/CME/Biophys/BMI 371
# Computational biology in four dimensions

Ron Dror

*Jan. 9, 2017*

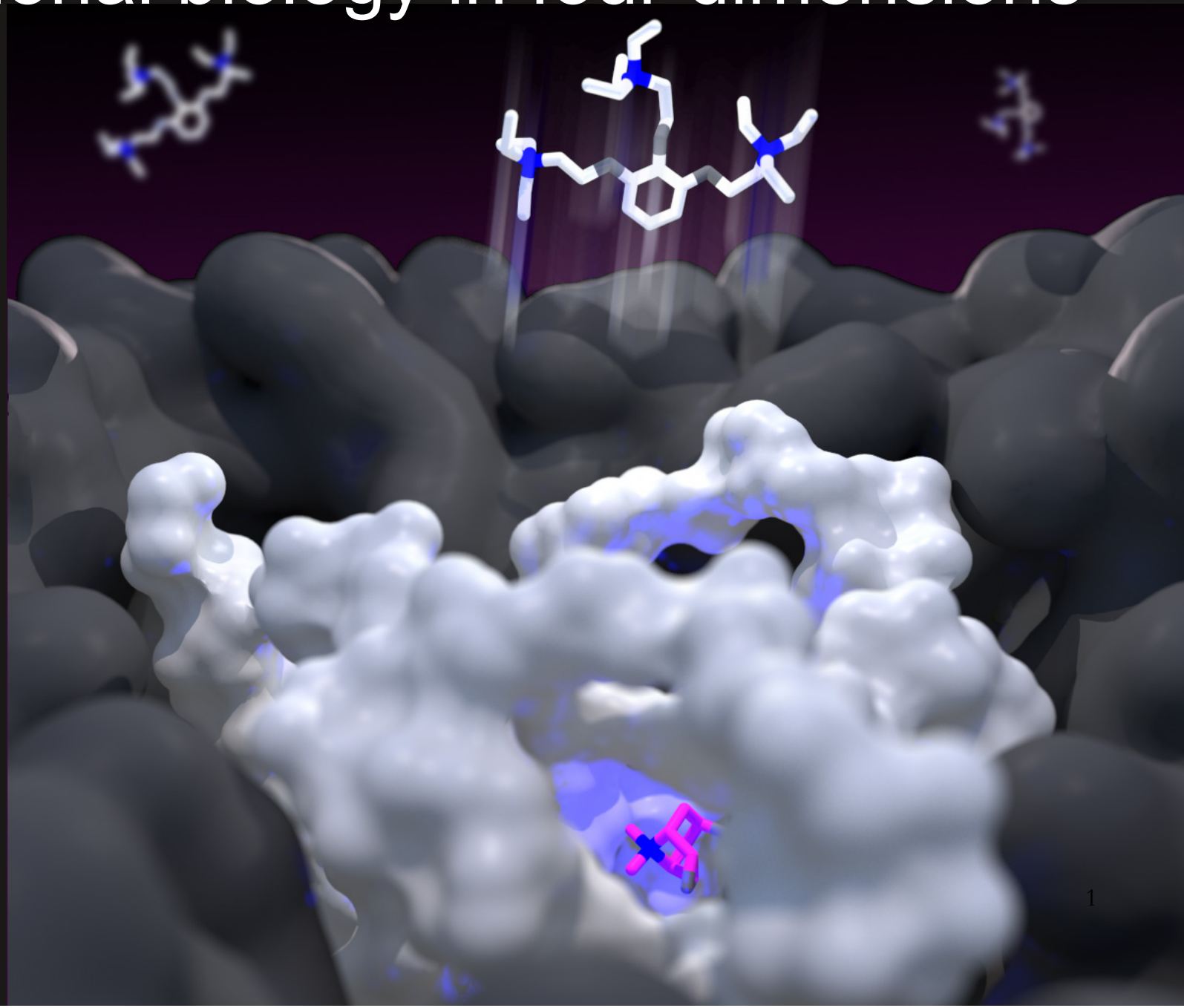


*Image credit:*
*Sam Hertig*

# Outline for lecture 1 (course overview)

- Course format and objectives
- What is structure?
  - Structure (and dynamics) at multiple spatial scales
- Why is structure important?
- Overview of course topics
  - Atomic-level modeling of biomolecules
  - Structures of macromolecular complexes
  - Cellular-level organization
- Course logistics
- Guidelines for presentations and critiques
- Immediate next steps

# Course format and objectives

# Focus is on presentation, discussion, and critique of cutting-edge research

- The majority of class time will be spent on presentations and discussion of important recent research papers
- Most presentations will be by students
  - Each student will present on a paper of interest to them
  - Professor & TAs will meet with students to review presentations
  - Professor will introduce topics in class
  - Professor and guest speaker will give initial presentations
- Professor will facilitate discussion. Students are strongly encouraged to share comments, questions

# Course requirements

- One presentation
  - 20-30 minutes each
- Three written critiques of papers being presented
- Read the remaining papers, attend class, and participate in discussions

# What do I want students to learn from this course?

1. Gain exposure to cutting-edge computational research in structural biology, broadly defined (a rapidly growing, interdisciplinary area).
2. Learn to critique and evaluate research, and practice critical reading of research papers.
3. Refine the skill of presenting deep technical material to a non-expert audience.

The latter two are broadly important skills. They're especially important to practice if they make you a bit nervous.

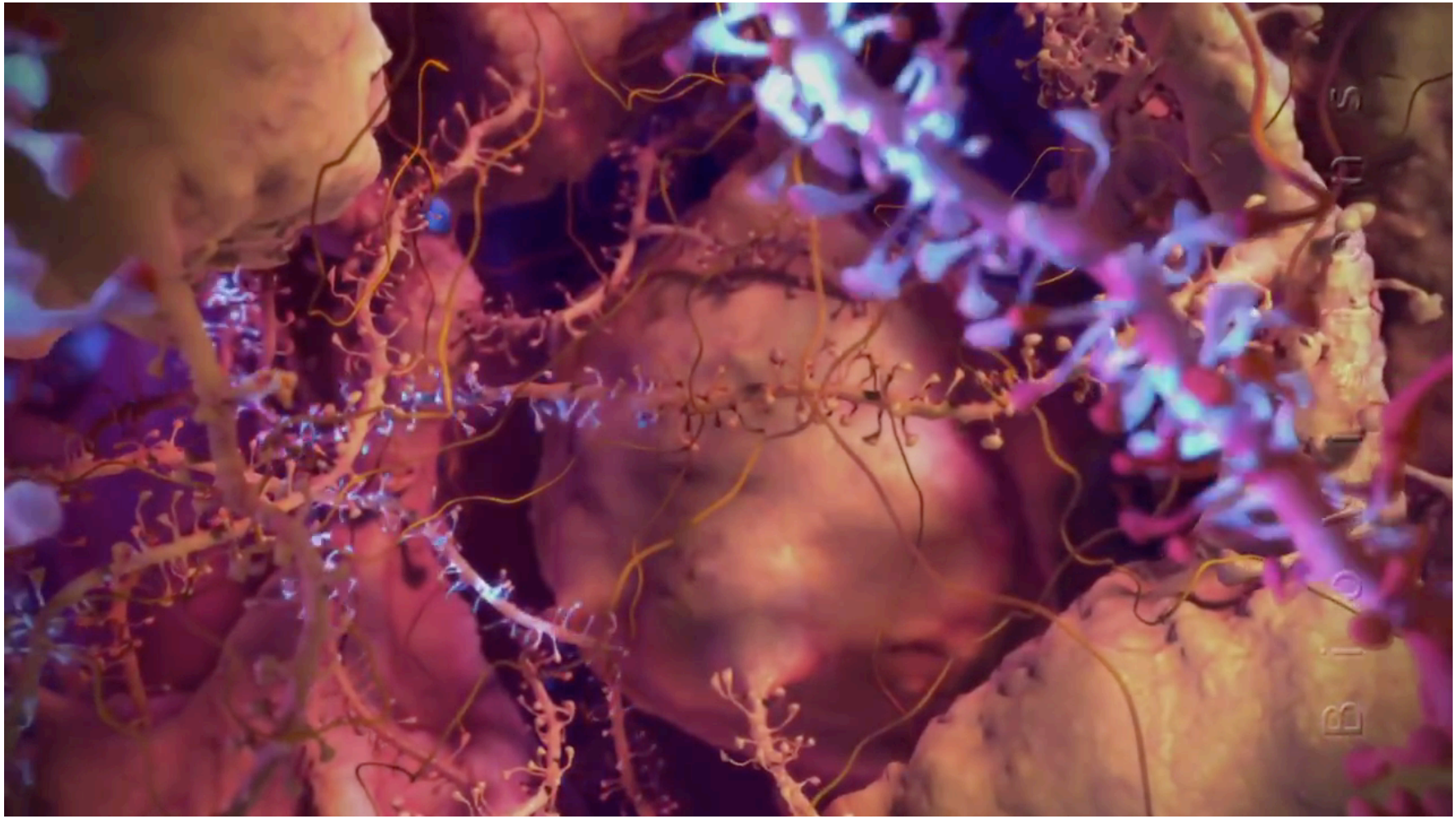
# Relationship to CS/CME/BioE/Biophys/BMI 279

- 371 and 279 cover the same general field
- 279 is a traditional lecture/homework course covering basics of the field
- 371 focuses on current research topics
- Neither is a prerequisite for the other

# What is structure?

# In daily life, we use machines with functional *structure* and *moving parts*

# Cells and biomolecules (e.g., proteins) are also machines whose function depends on structure and moving parts



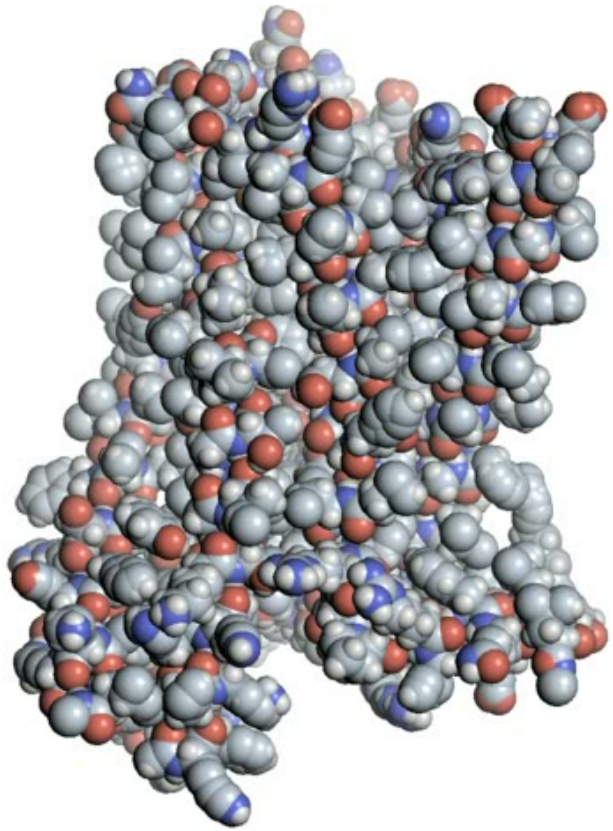


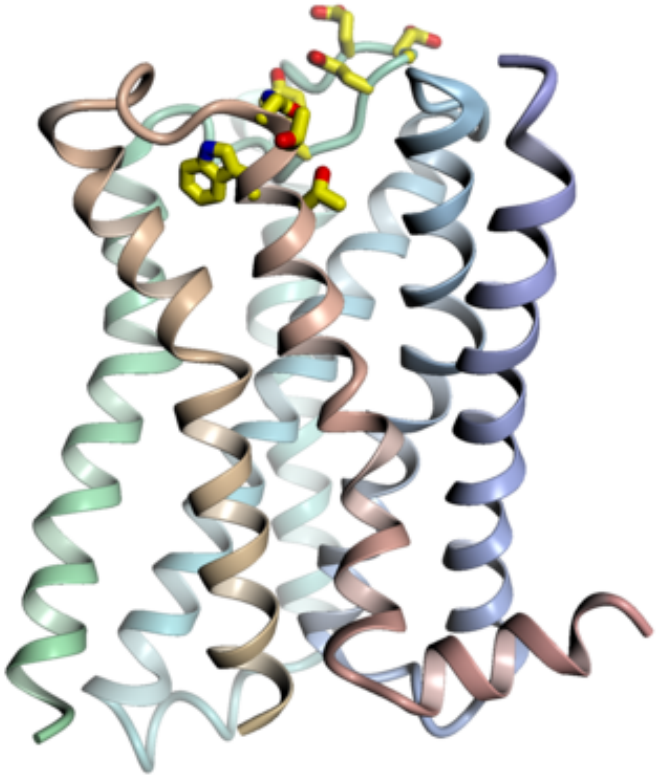
From *Inner Life of the Cell* | *Protein Packing*, XVIVO and Biovisions @ Harvard

# What is structure?

Structure (and dynamics)
at multiple spatial scales

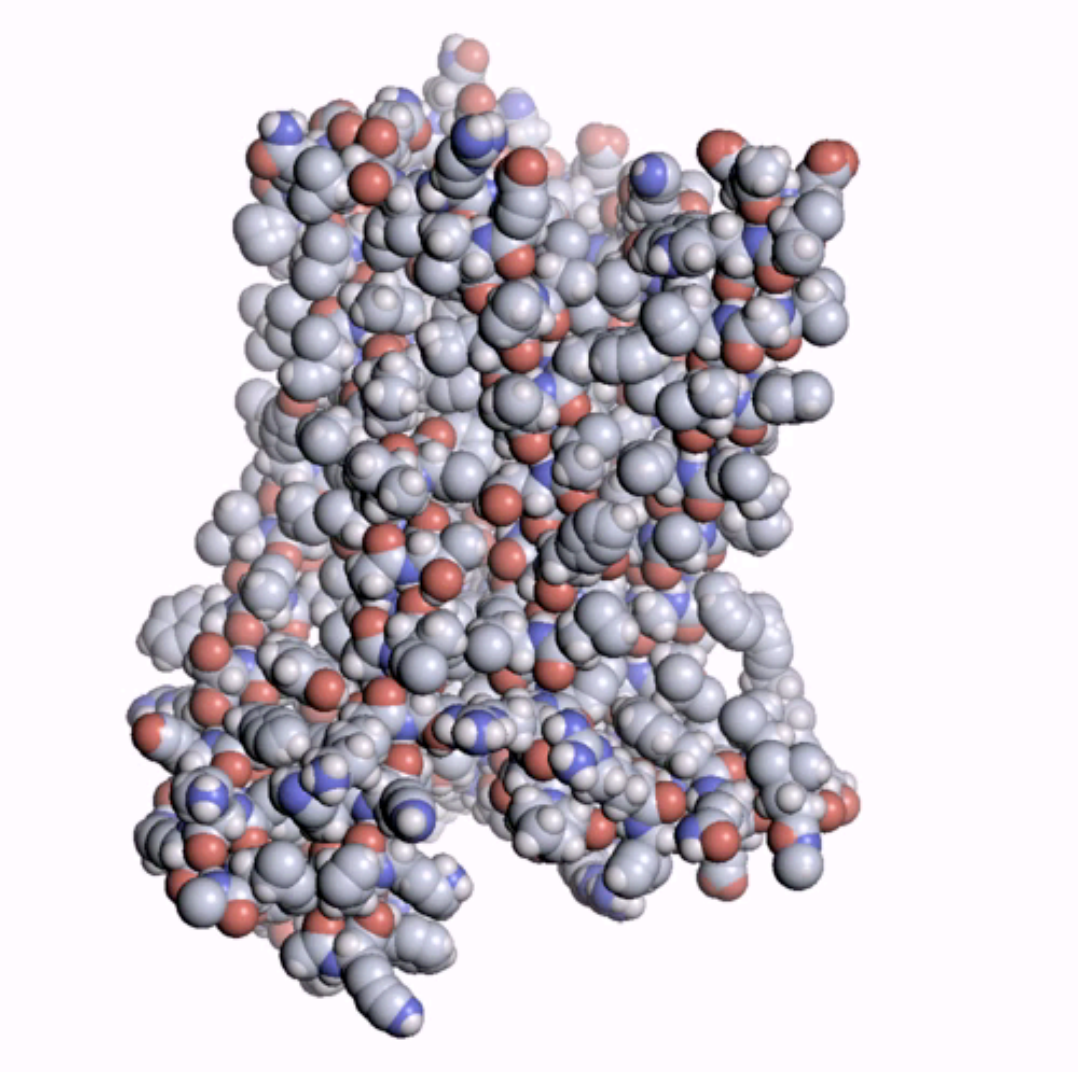
# Protein structure





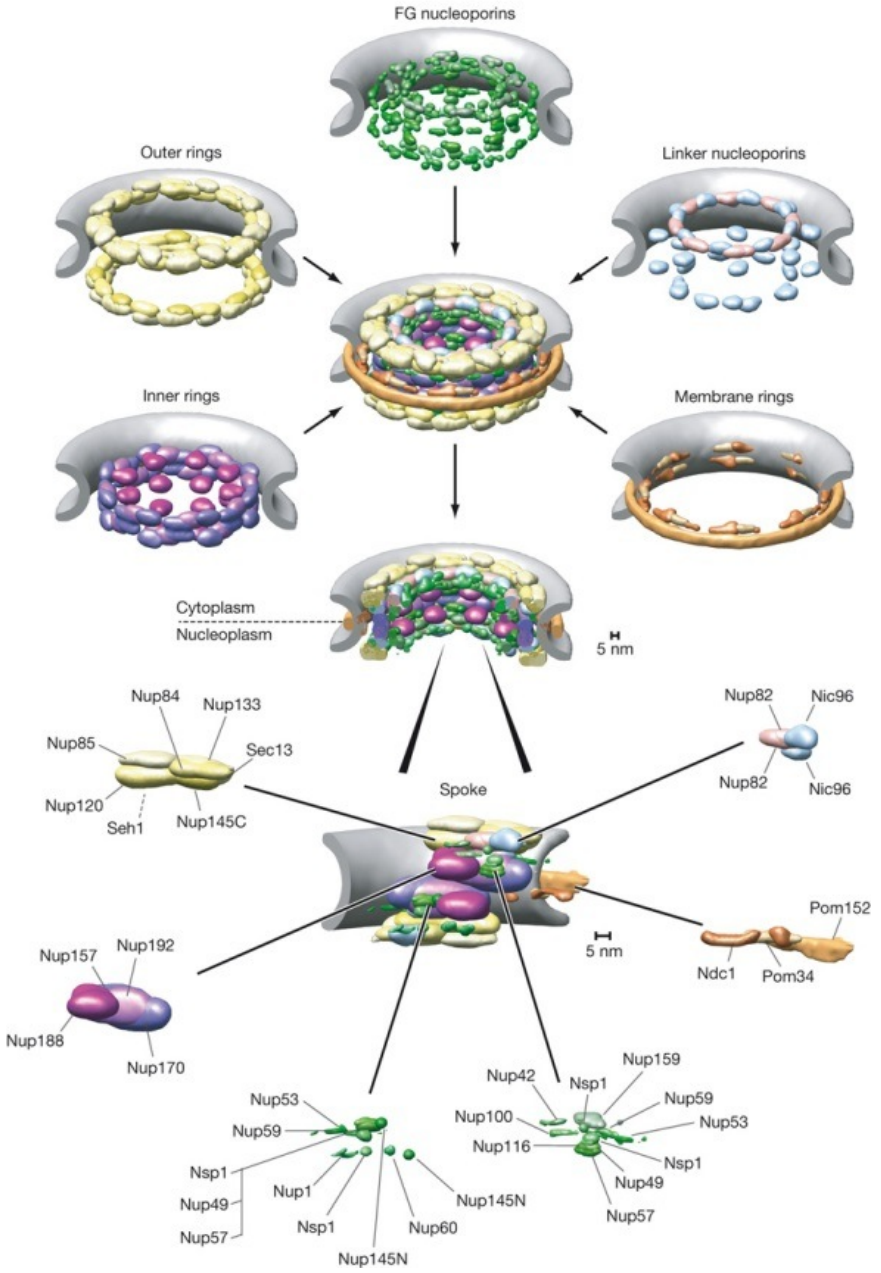
An adrenaline receptor
(the $\beta_2$ adrenergic receptor)

# Protein dynamics



$\beta_2$ adrenergic receptor

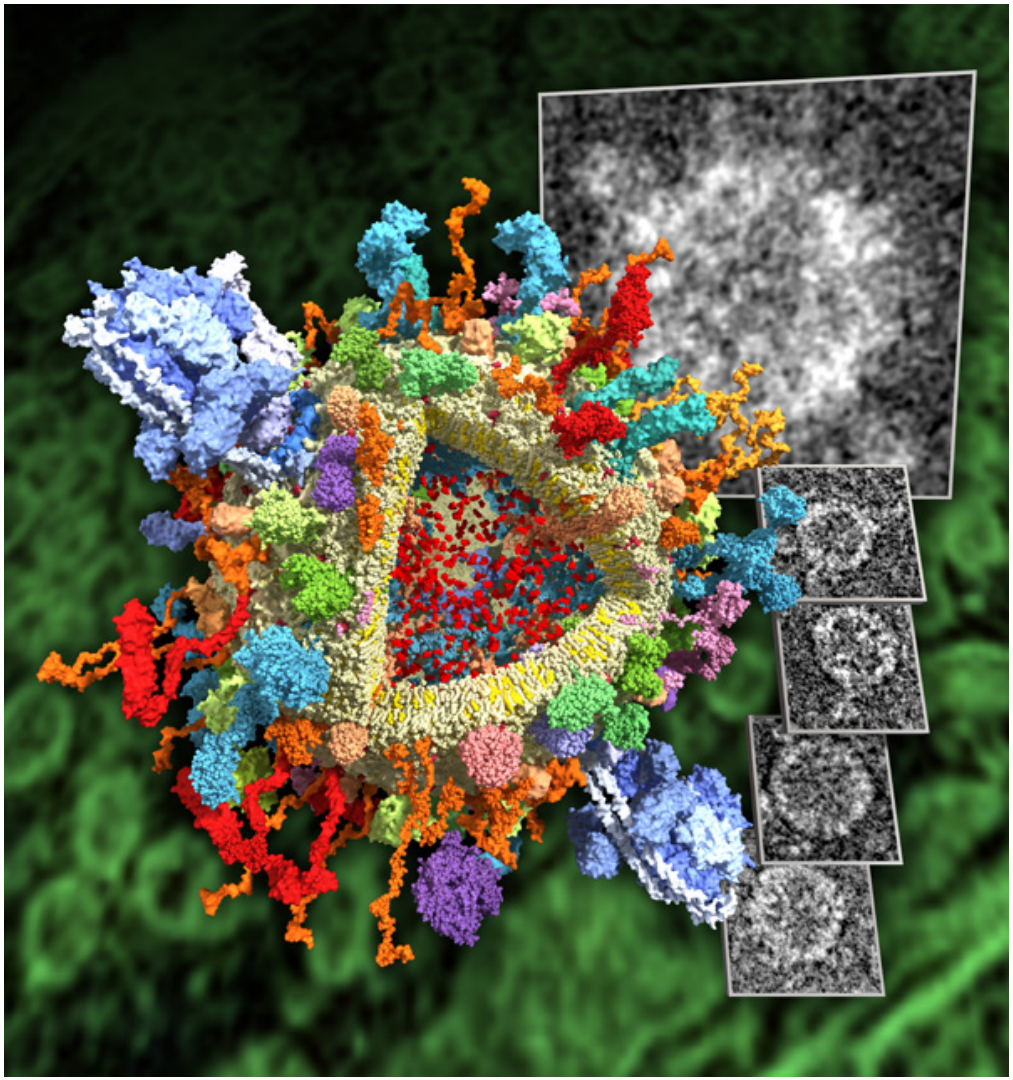
Proteins (and other molecules) often come together to form *macromolecular complexes*




Nuclear Pore Complex

Alber et al., *Nature* 2007

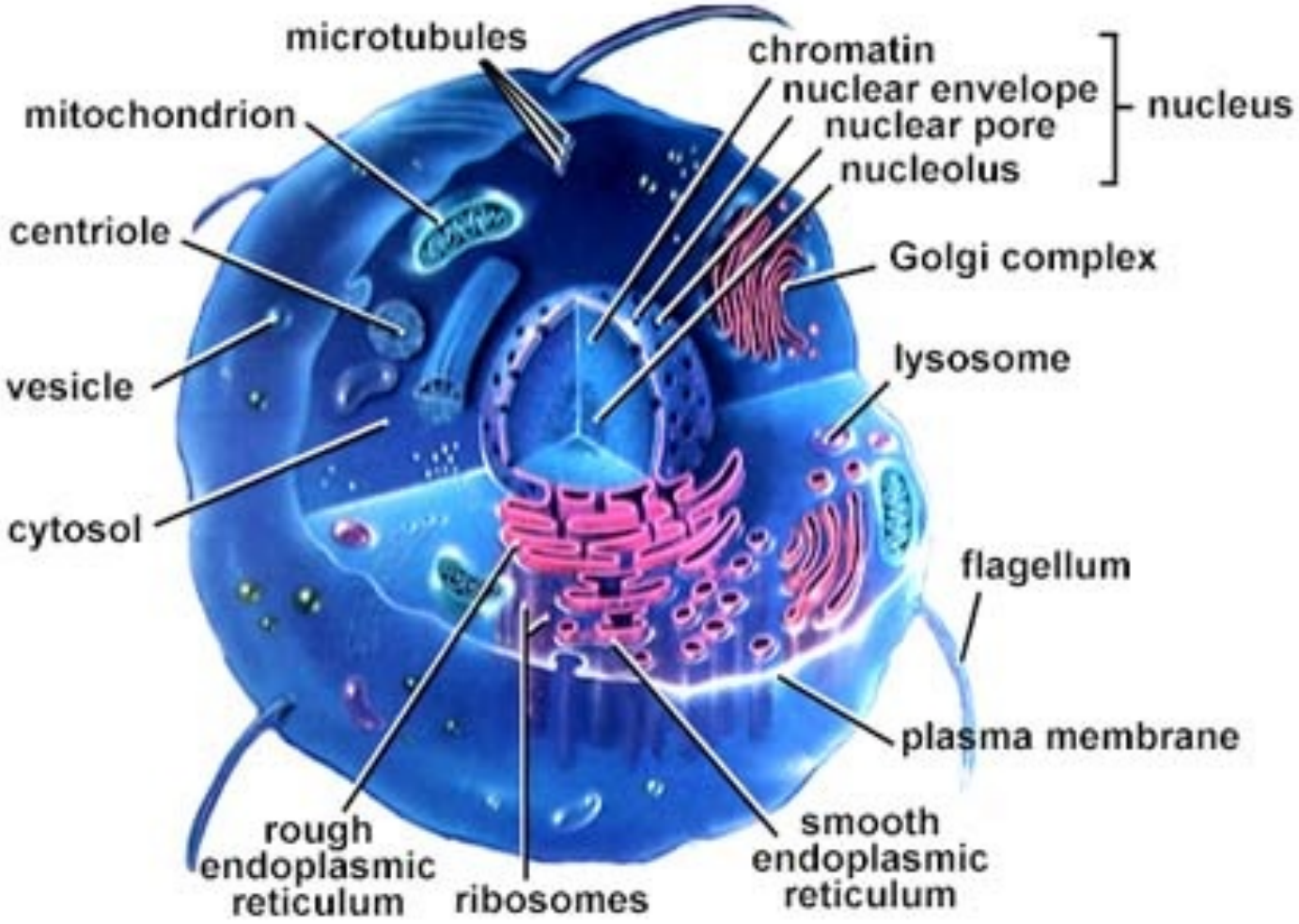
# These come together to form organelles



Synaptic vesicle

http://www.mpibpc.mpg.de/9547480/vesicle600.jpg

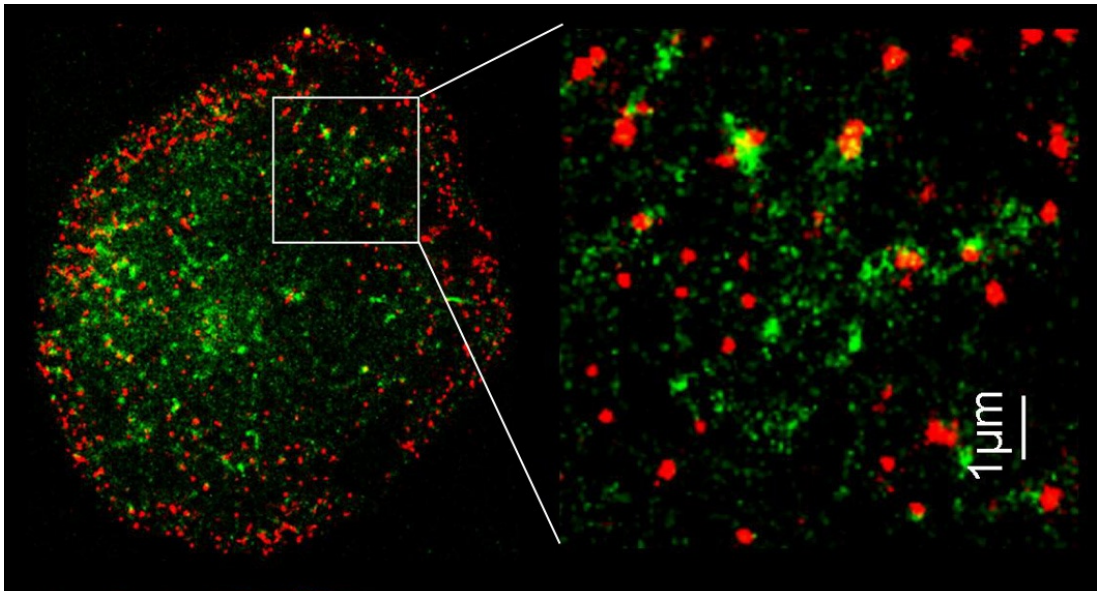
# and cells



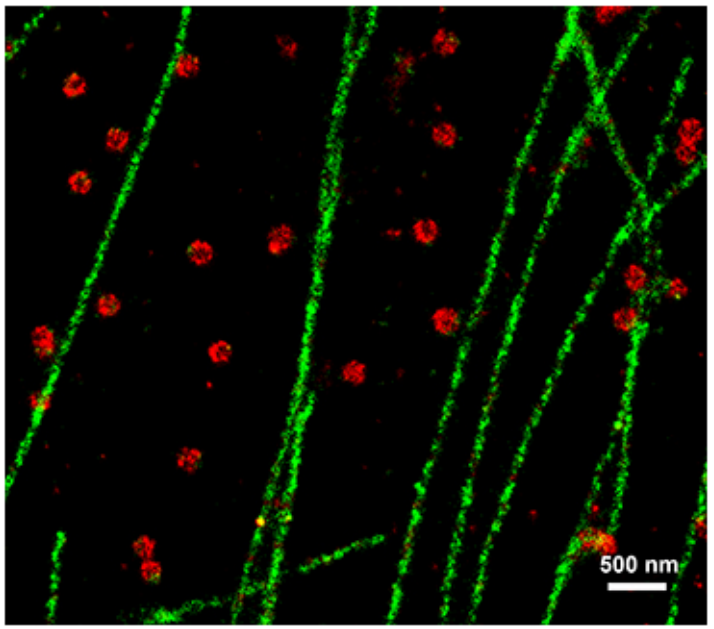



http://www.medfriendly.com/cell.html

# Intracellular structure




Chih-Jung Hsu, Janis Burkhardt and Tobias Baumgart




http://www.nikoninstruments.com/Products/Microscope-Systems/Inverted-Microscopes/N-STORM-Super-Resolution/(gallery); Zhuang group

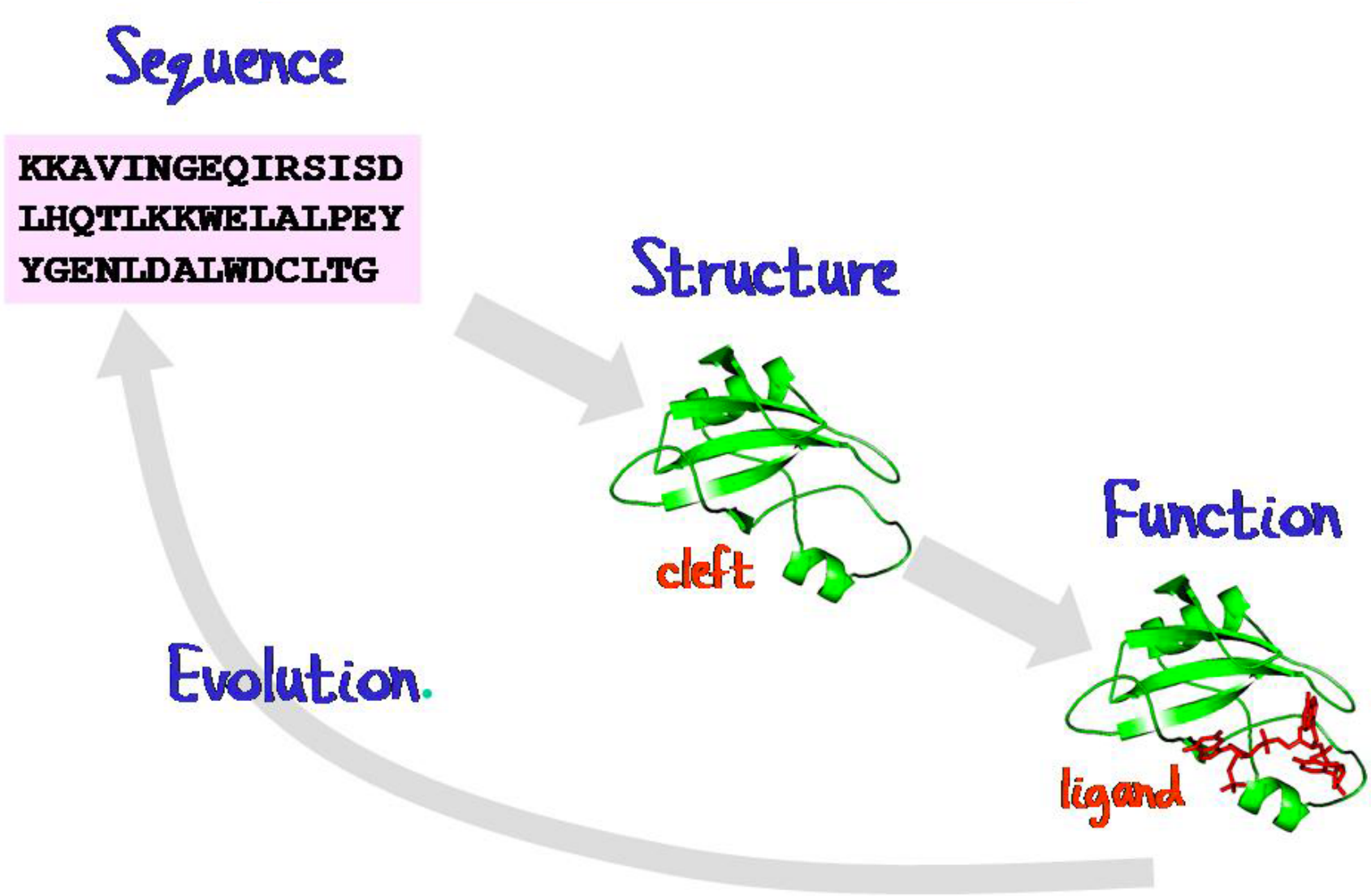
David Goodsell

# Intracellular dynamics (artist's rendition)

Janet Iwasa and Tomas Kirchhausen

# Why is structure important?

# The cycle of life




From Michael Levitt

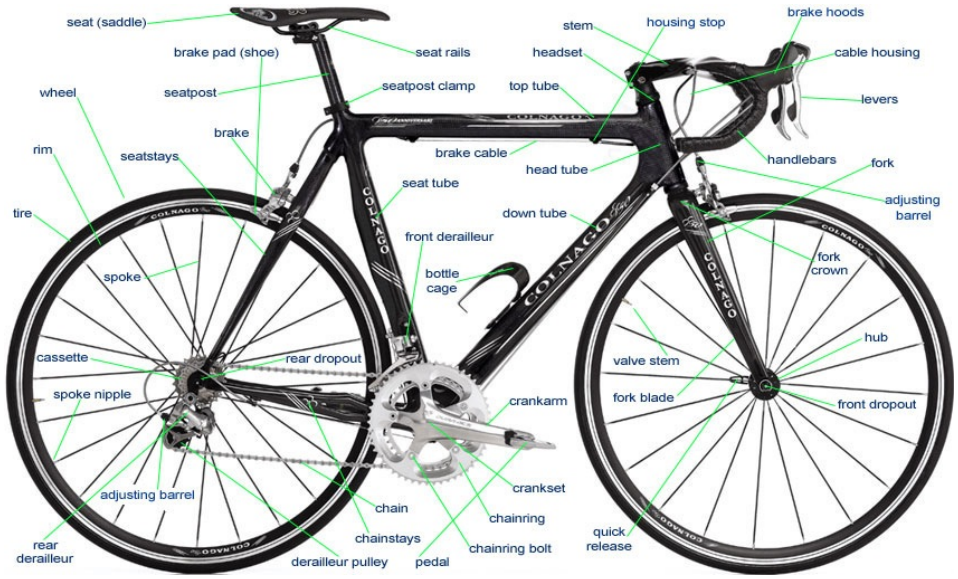
# Genomics is a great start ….

**Track Bike — DL 175**

| REF. NO. | IBM NO. | DESCRIPTION |
|---|---|---|
| 1 | 156011 | Track Frame 21", 22", 23", 24", Team Red |
| 2 | 157040 | Fork for 21" Frame |
| 2 | 157039 | Fork for 22" Frame |
| 2 | 157038 | Fork for 23" Frame |
| 2 | 157037 | Fork for 24" Frame |
| 3 | 191202 | Handlebar TTT Competition Track Alloy 15/16" |
| 4 | | Handlebar Stem, TTT, Specify extension |
| 5 | 191278 | Expander Bolt |
| 6 | 191272 | Clamp Bolt |
| 7 | 145841 | Headset Complete 1 x 24 BSC |
| 8 | 145842 | Ball Bearings |
| 9 | 190420 | 175 Raleigh Pistard Seta Tubular Prestavalve 27" |
| 10 | 190233 | Rim, 27" AVA Competition (36H) Alloy Prestavalve |
| 11 | 145973 | Hub, Large Flange Campagnolo Pista Track Alloy (pairs) |
| 12 | 190014 | Spokes, 11 5/8" |
| 13 | 145837 | Sleeve |
| 14 | 145636 | Ball Bearings |
| 15 | 145170 | Bottom Bracket Axle |
| 16 | 145838 | Cone for Sleeve |
| 17 | 146473 | L.H. Adjustable Cup |
| 18 | 145833 | Lockring |
| 19 | 145239 | Straps for Toe Clips |
| 20 | 145834 | Fixing Bolt |
| 21 | 145835 | Fixing Washer |
| 22 | 145822 | Dustcap |
| 23 | 145823 | R.H. and L.H. Crankset with Chainwheel |
| 24 | 146472 | Fixed Cup |
| 25 | 145235 | Toe Clips, Christophe, Chrome (Medium) |
| 26 | 145684 | Pedals, Extra Light, Pairs |
| 27 | 123021 | Chain |
| 28 | 145980 | Seat Post |
| 29 | | Seat Post Bolt and Nut |
| 30 | 167002 | Saddle, Brooks |
| 31 | 145933 | Track Sprocket, Specify 12, 13, 14, 15, or 16 T. |

- But a parts list is not enough to understand how a bicycle works

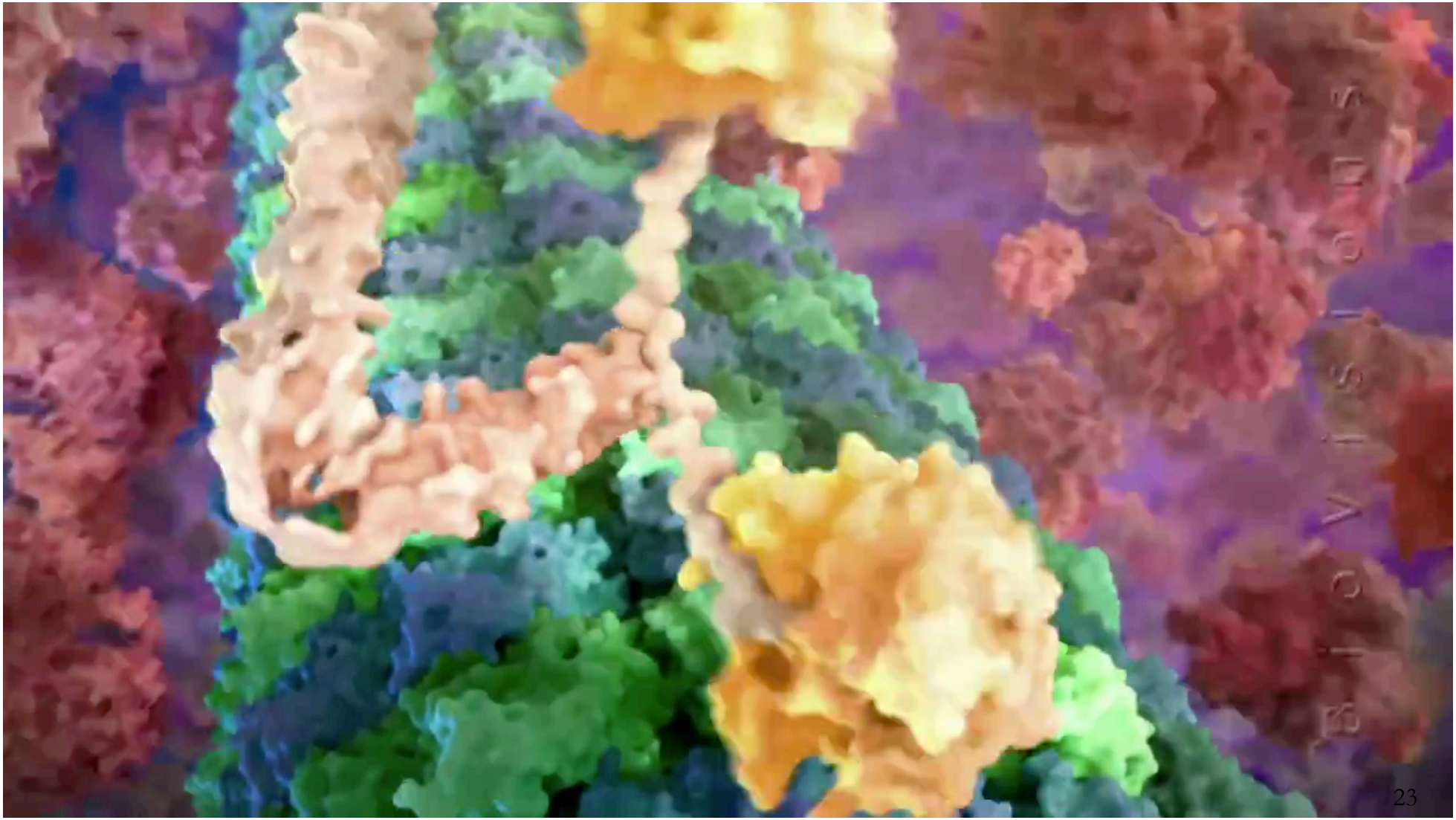
# … but not the end




- We want the full spatiotemporal picture, and an ability to control it
- Broad applications, including drug design, medical diagnostics, chemical manufacturing, and energy

# Structure determines function

- Example: Motor protein (walks along microtubules, dragging load)

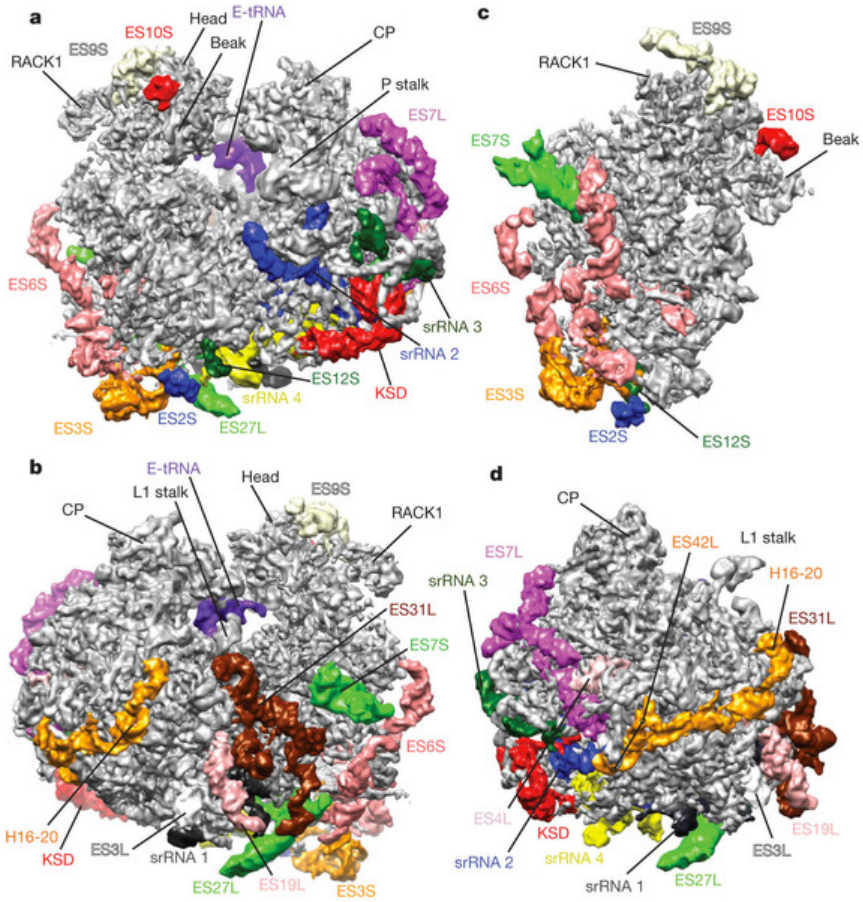



From *Inner Life of the Cell* | *Protein Packing*

# Structure determines function

- Example: Ribosome
  - Complex of many proteins and RNAs that together makes new proteins (by reading the genetic code and combining amino acids)

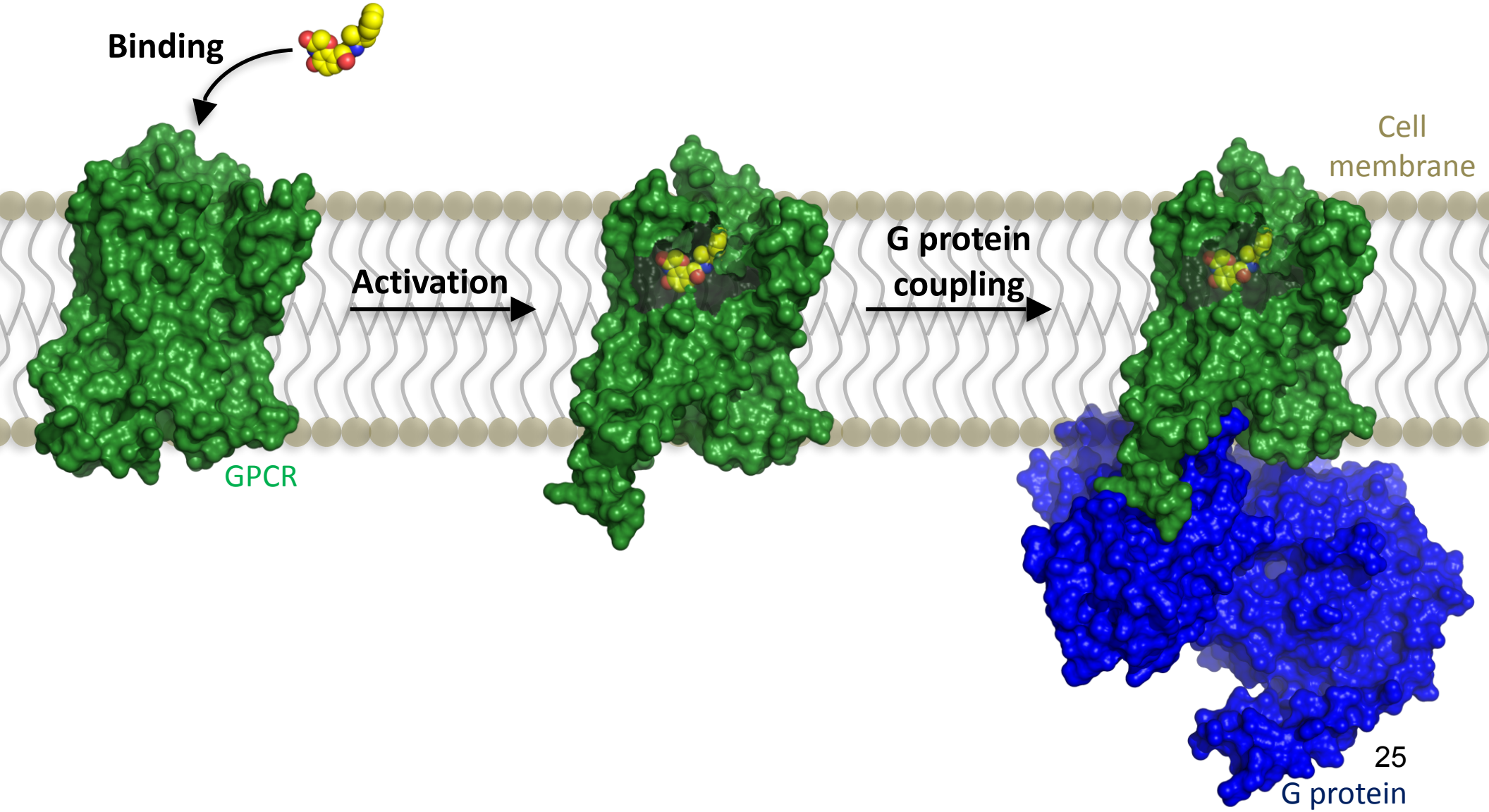
From *Inner Life of the Cell*, XVIVO and Biovisions @ Harvard




Hashem et al., Nature 494:385-9, 2013

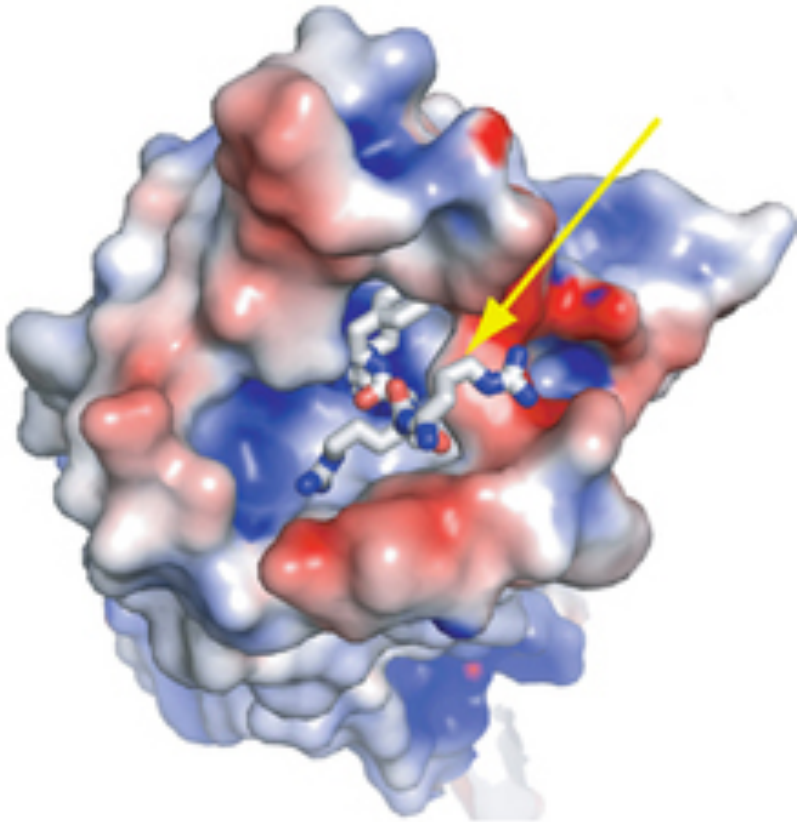
# Structure determines function

- Example: G protein-coupled receptors (GPCRs)
  - Largest class of human drug targets
  - Function: allow the cell to sense and respond to molecules outside it



# Structure-based drug design

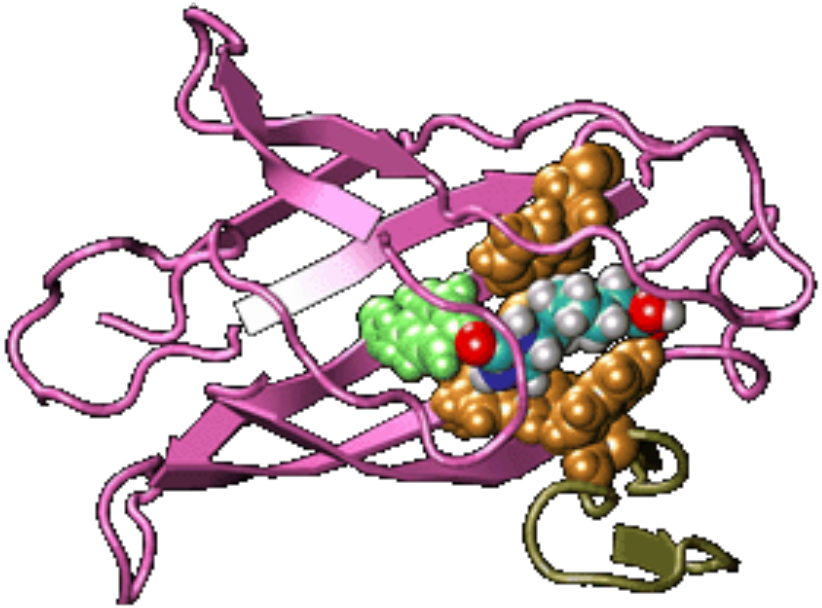
- Almost all drugs act by binding to proteins and altering their function
- Using knowledge of structures, we can design drugs that bind more tightly or more selectively, bind in different positions, alter behavior of protein in different ways, etc.



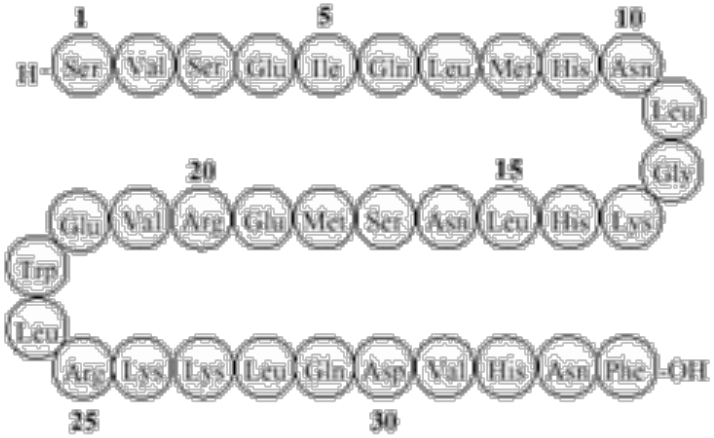
http://www.nih.gov/researchmatters/october2012/images/structure_l.jpg

# Designing new biomolecular machines

- Protein design (for health or industrial applications)
- Cell design?



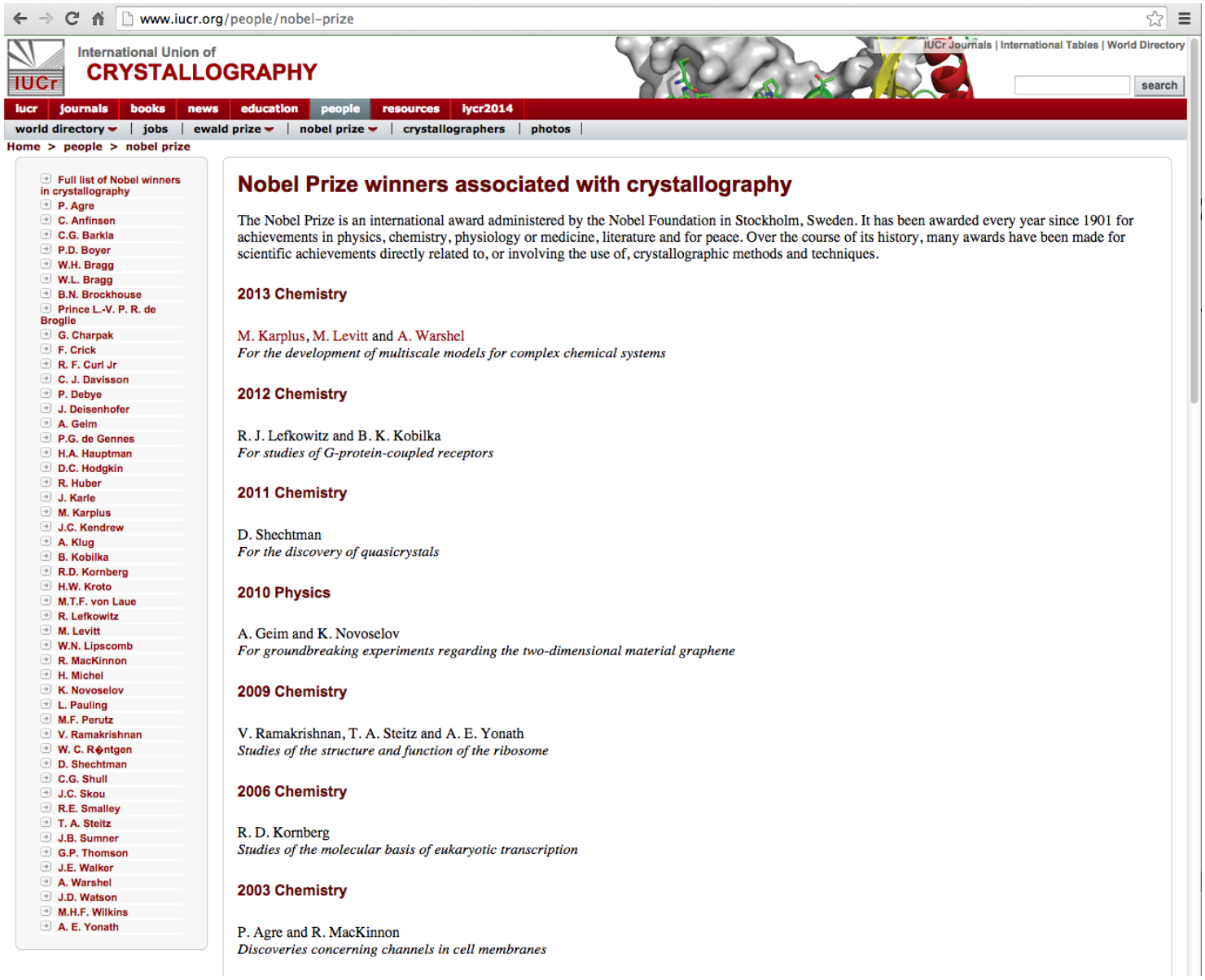
How?






http://zhanglab.ccmb.med.umich.edu/image/Protein_design.gif

# A strikingly large share of Nobel Prizes have recognized work on molecular structure



www.iucr.org/people/nobel-prize

International Union of CRYSTALLOGRAPHY

IUCr Journals | International Tables | World Directory

search

iucr | journals | books | news | education | people | resources | iycr2014

world directory | jobs | ewald prize | nobel prize | crystallographers | photos

Home > people > nobel prize

- Full list of Nobel winners in crystallography
- P. Agre
- C. Anfinsen
- C.G. Barkla
- P.D. Boyer
- W.H. Bragg
- W.L. Bragg
- B.N. Brockhouse
- Prince L.-V. P. R. de Broglie
- G. Charpak
- F. Crick
- R. F. Curl Jr
- C. J. Davisson
- P. Debye
- J. Deisenhofer
- A. Geim
- P.G. de Gennes
- H.A. Hauptman
- D.C. Hodgkin
- R. Huber
- J. Karle
- M. Karplus
- J.C. Kendrew
- A. Klug
- B. Kobilka
- R.D. Kornberg
- H.W. Kroto
- M.T.F. von Laue
- R. Lefkowitz
- M. Levitt
- W.N. Lipscomb
- R. MacKinnon
- H. Michel
- K. Novoselov
- L. Pauling
- M.F. Perutz
- V. Ramakrishnan
- W. C. R◆ntgen
- D. Shechtman
- C.G. Shull
- J.C. Skou
- R.E. Smalley
- T. A. Steitz
- J.B. Sumner
- G.P. Thomson
- J.E. Walker
- A. Warshel
- J.D. Watson
- M.H.F. Wilkins
- A. E. Yonath

## Nobel Prize winners associated with crystallography

The Nobel Prize is an international award administered by the Nobel Foundation in Stockholm, Sweden. It has been awarded every year since 1901 for achievements in physics, chemistry, physiology or medicine, literature and for peace. Over the course of its history, many awards have been made for scientific achievements directly related to, or involving the use of, crystallographic methods and techniques.

**2013 Chemistry**

M. Karplus, M. Levitt and A. Warshel
*For the development of multiscale models for complex chemical systems*

**2012 Chemistry**

R. J. Lefkowitz and B. K. Kobilka
*For studies of G-protein-coupled receptors*

**2011 Chemistry**

D. Shechtman
*For the discovery of quasicrystals*

**2010 Physics**

A. Geim and K. Novoselov
*For groundbreaking experiments regarding the two-dimensional material graphene*

**2009 Chemistry**

V. Ramakrishnan, T. A. Steitz and A. E. Yonath
*Studies of the structure and function of the ribosome*

**2006 Chemistry**

R. D. Kornberg
*Studies of the molecular basis of eukaryotic transcription*

**2003 Chemistry**

P. Agre and R. MacKinnon
*Discoveries concerning channels in cell membranes*

# 2013 Nobel Prize recognized early developments underlying modern biomolecular computation

## The Nobel Prize in Chemistry 2013

Photo: A. Mahmoud

**Martin Karplus**

**Prize share:** 1/3

Photo: A. Mahmoud

**Michael Levitt**

**Prize share:** 1/3

Photo: A. Mahmoud

**Arieh Warshel**

**Prize share:** 1/3

The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel *"for the development of multiscale models for complex chemical systems"*.

# 2014 Nobel Prize recognized microscopy techniques used to study cellular structure (which also rely on computation)

## The Nobel Prize in Chemistry 2014

Photo: A. Mahmoud

**Eric Betzig**

**Prize share:** 1/3

Photo: A. Mahmoud

**Stefan W. Hell**

**Prize share:** 1/3

Photo: A. Mahmoud

**William E. Moerner**

**Prize share:** 1/3

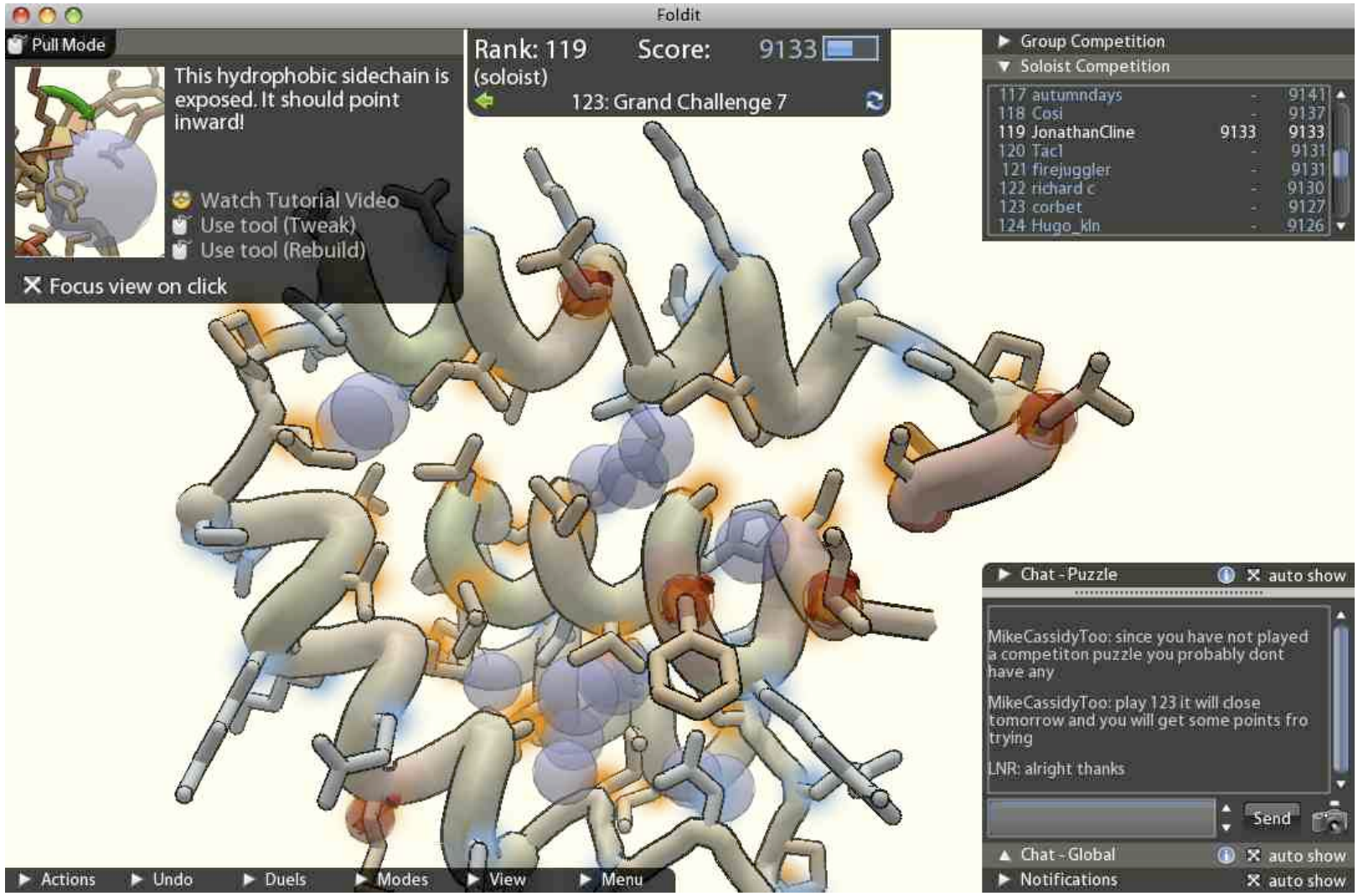
The Nobel Prize in Chemistry 2014 was awarded jointly to Eric Betzig, Stefan W. Hell and William E. Moerner *"for the development of super-resolved fluorescence microscopy"*.

# Overview of course topics

# Overview of course topics

## Atomic-level modeling of biomolecules

# Protein structure prediction and RNA/ protein design by video game



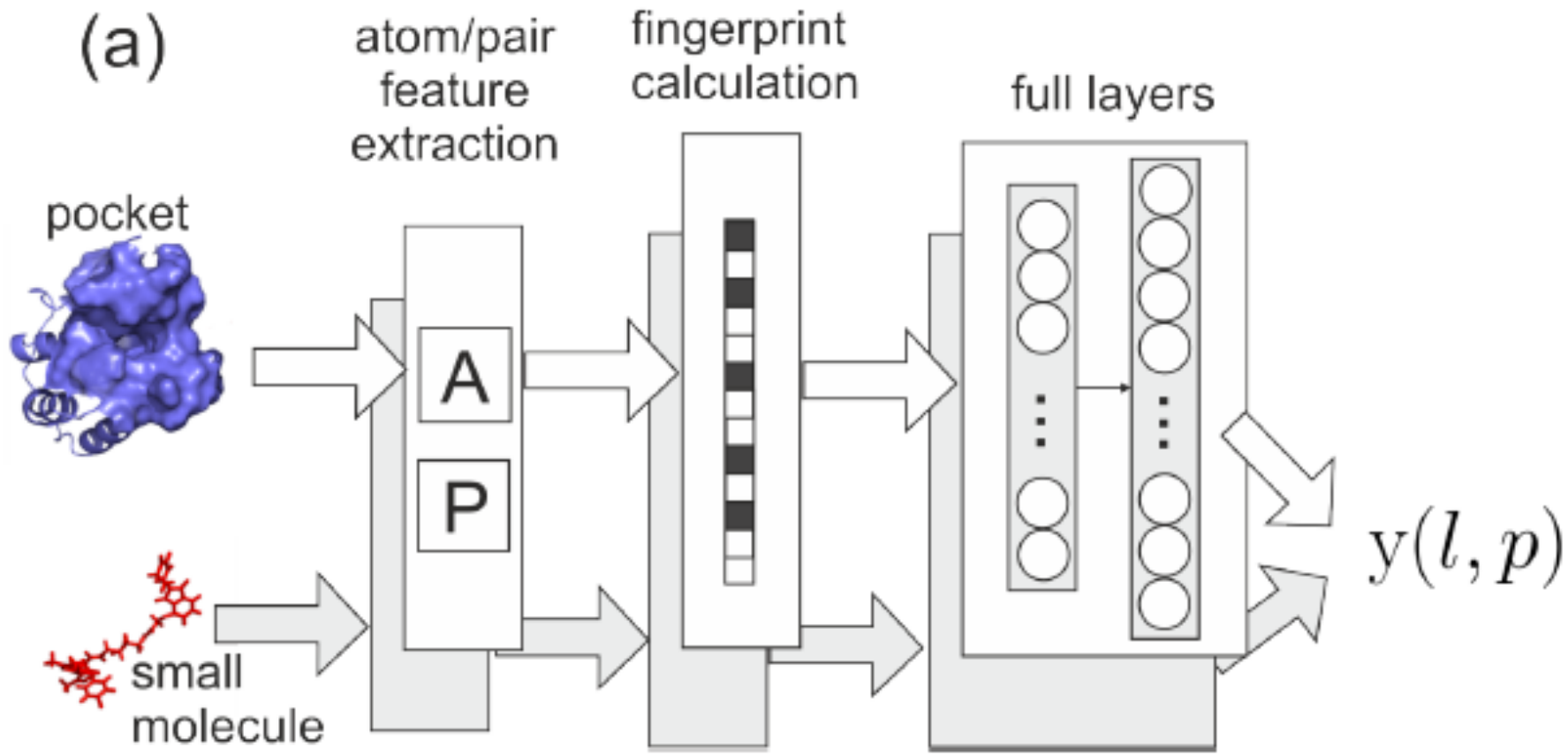



screen shot from Foldit

Foldit - 1-1: One Small Clash
Pull Mode
Score:
7940
of 7900
1-1: One Small Clash
You have completed
1 of 31 intro puzzles!
Moves:
4
Time:
0:19
Next is: 1-2: Swing It Around!
Next Puzzle
Puzzle Menu
Replay Puzzle
Reset
Puzzle
Actions
Undo
Menu

# Machine learning on structures

- Applying machine learning techniques to 3D structures of proteins or drug-like molecules — for example, for virtual drug screening

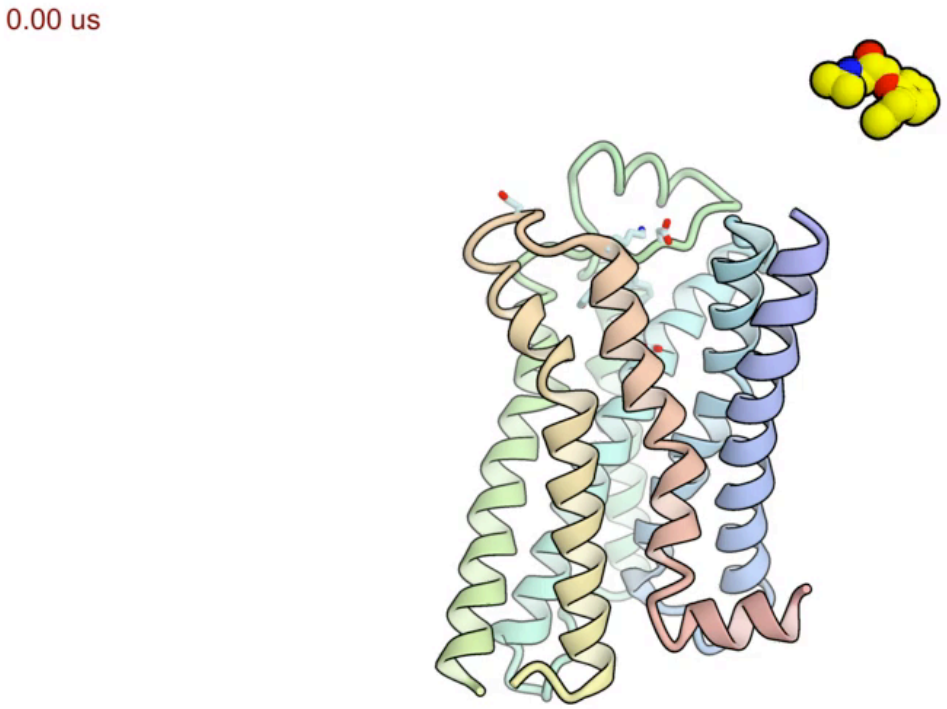





Gonczarek et al., NIPS 2016

# Drug screening: computing accurate binding strengths for drug candidates

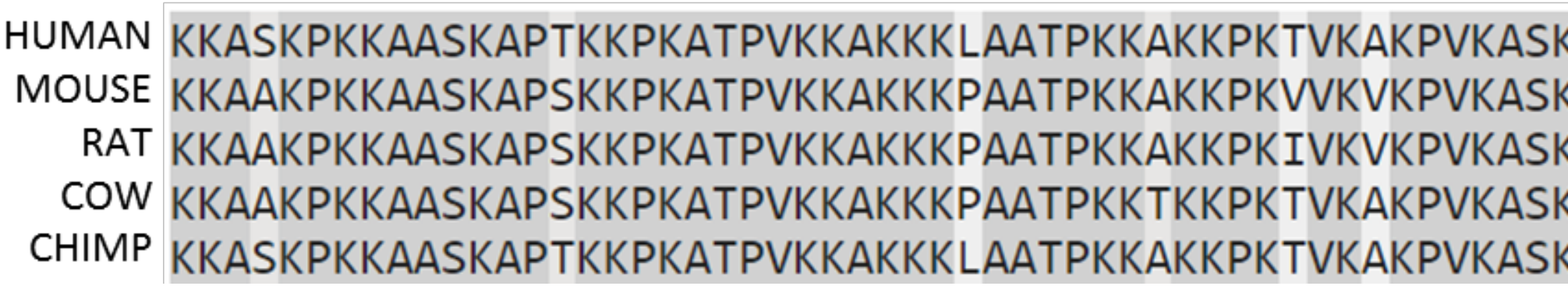
- "Alchemical" methods for computing binding free energies: simulate the drug as it gradually "disappears" from the binding pocket or "changes" into a different drug
- Sounds like magic, but it's actually much more faithful to the physics than traditional docking methods






Beta-blocker alprenolol binding to an adrenaline receptor (Dror et al., PNAS 2011)

# Coevolution methods for predicting structure from large numbers of genetic sequences




Key idea: amino acids in direct physical contact within a protein tend to mutate in a correlated fashion. Given enough sequence data, one can use this fact to predict structure.

# Modern protein design

Design of a protein that selectively transports zinc ions through a cell membrane

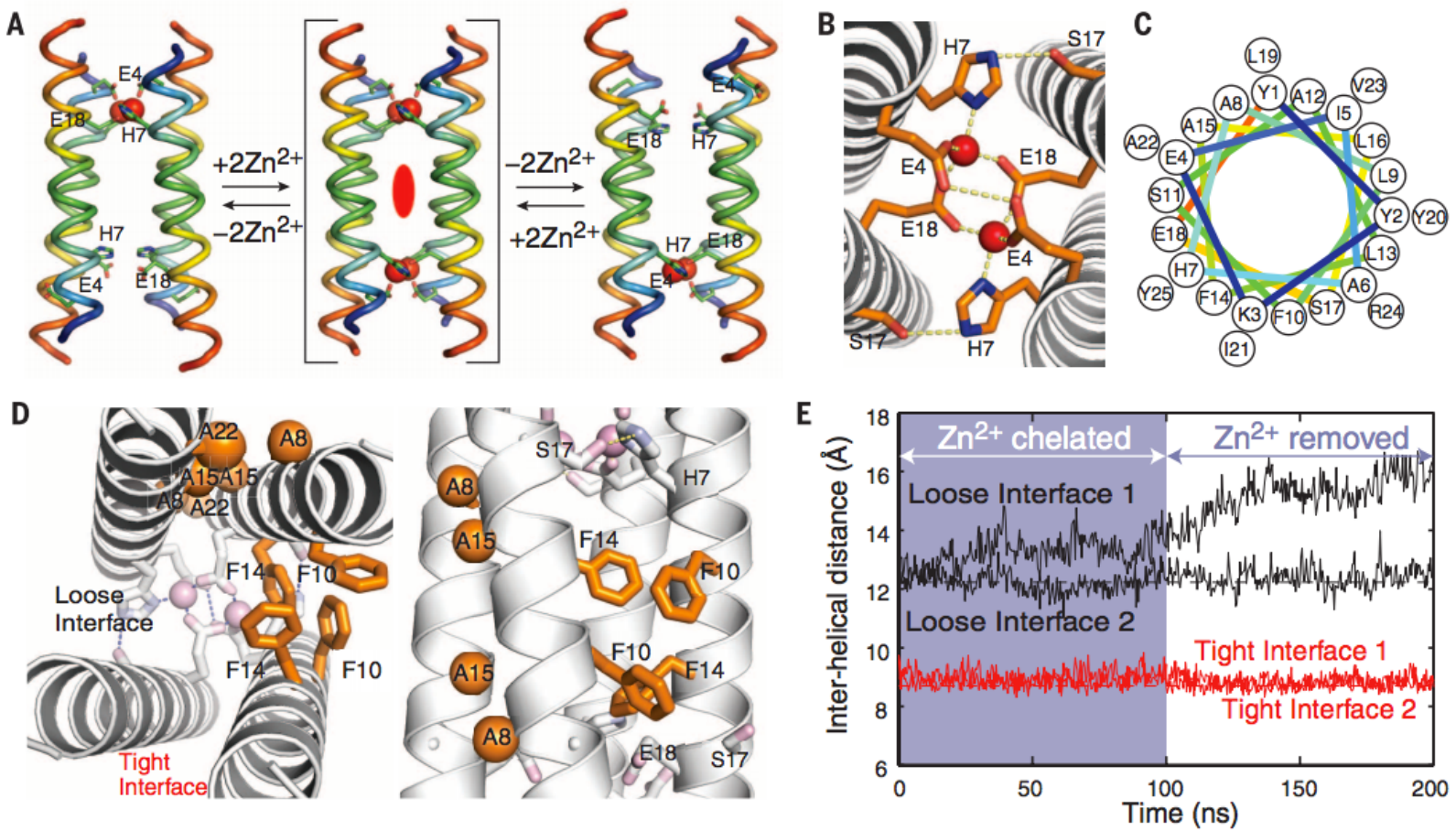
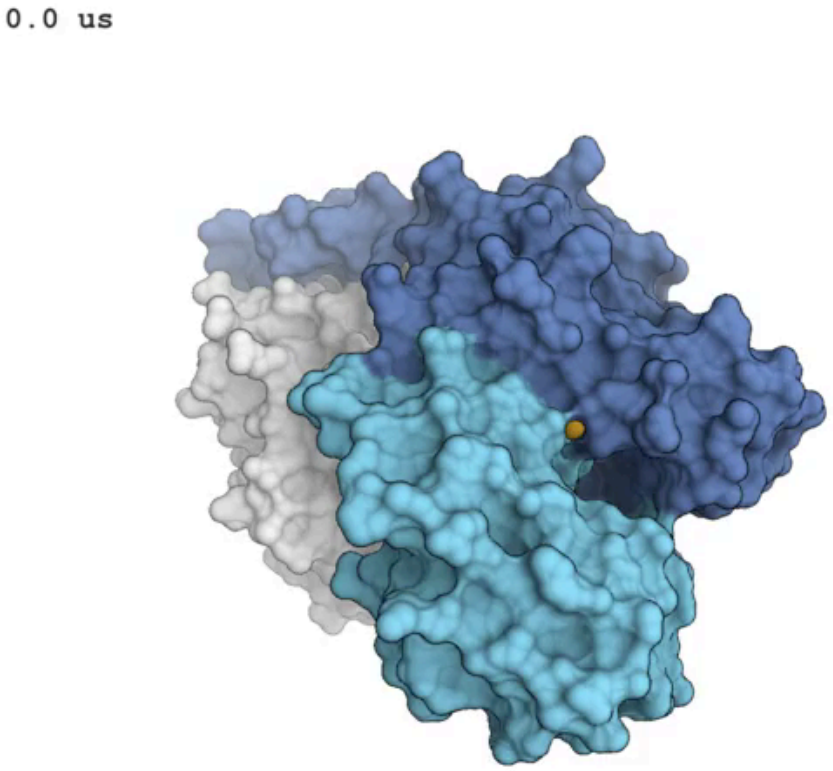



**Fig. 1. Computational design and molecular dynamics simulations of Rocker.** (**A**) Schematic of the goal of obtaining conformational exchange between two oppositely oriented symmetry-frustrated states without being trapped in a symmetric state with both sites simultaneously occupied. (**B**) Metal site consists of a set of ExxH motifs and a single Glu residue from each of the tight dimers. (**C**) Helical-wheel diagram of Rocker peptide. (**D**) The re-packing algorithm placed Ala residues at the tight interface and Phe residues at the loose interface. Empty metal site on the left is omitted for clarity. (**E**) MD simulation of the design model with two $Zn^{2+}$ ions placed at one metal site show stable interhelical distances for both tight and loose interfaces. Continuing the simulation after removing the $Zn^{2+}$ ions maintained the tight interfaces, but resulted in an increased interhelical distance at the loose interface.

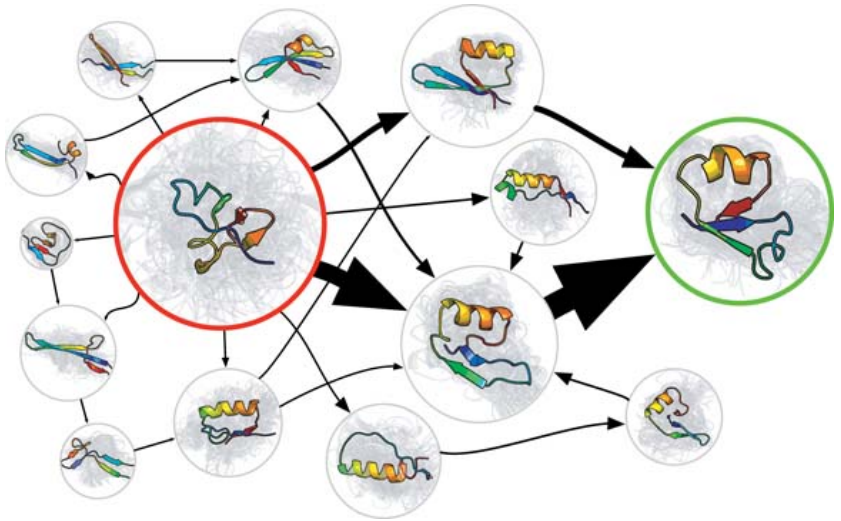


Joh et al., *Science* 346:1520-4 (2014)

# Molecular dynamics simulations and Markov State Models



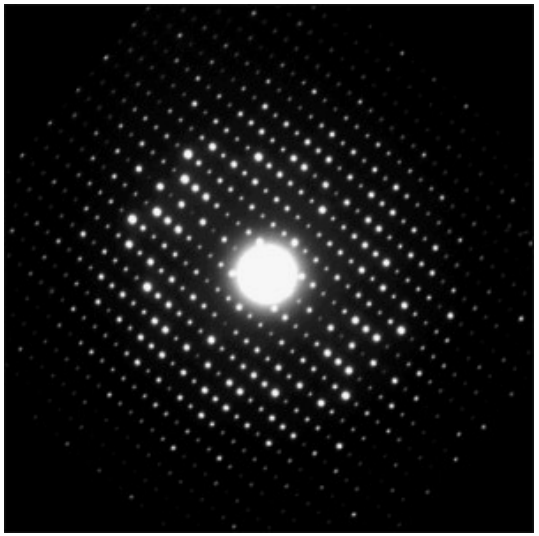

Dror et al., Science 2015

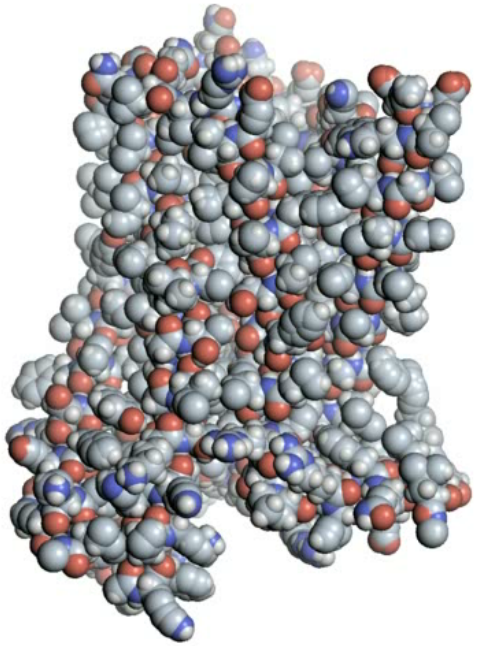


https://folding.stanford.edu/home/faq/faq-simulation/

# New methods for solving tough crystal structures

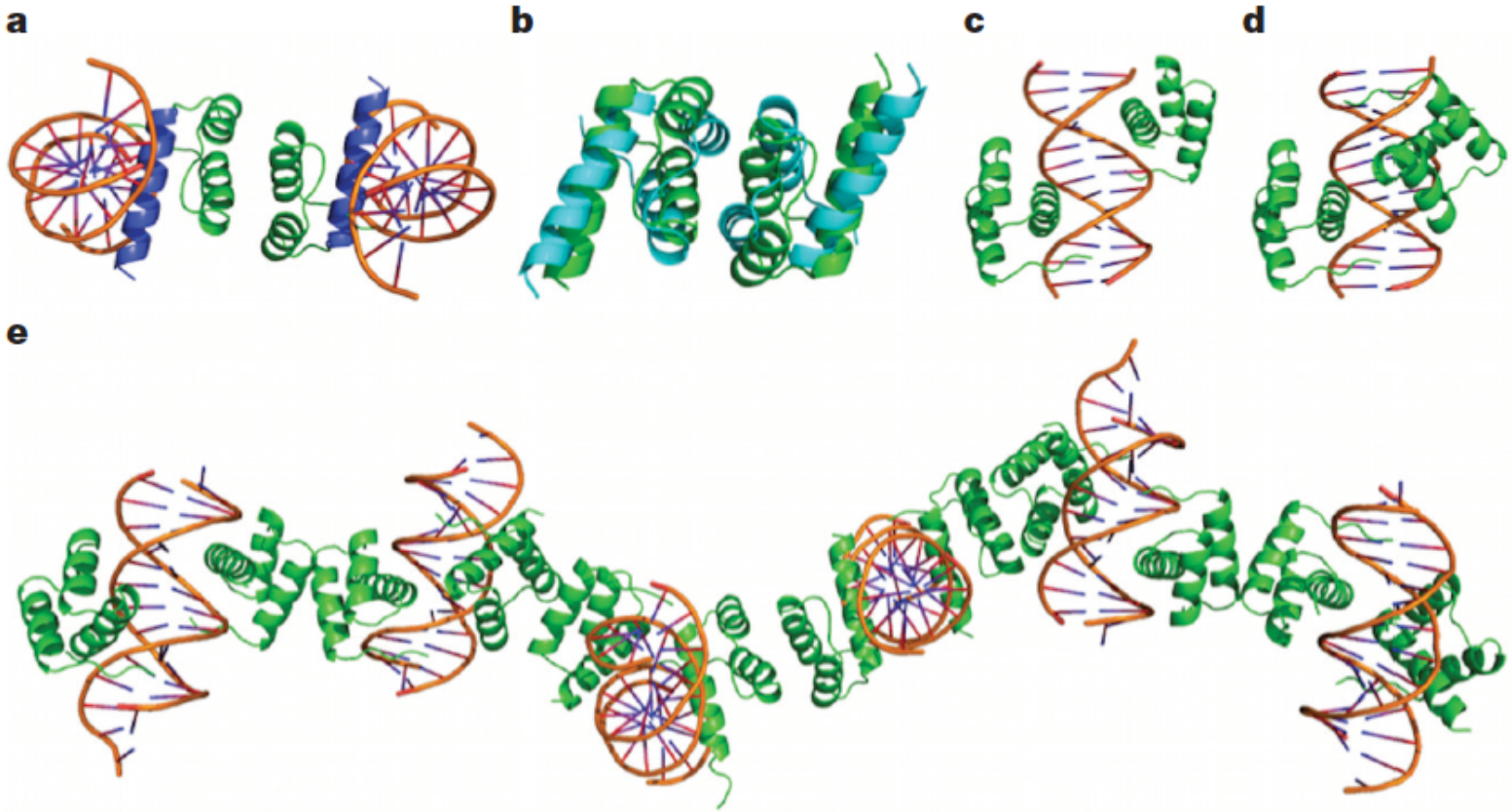
- Computational problem in crystallography: go from diffraction pattern to structure
  - Mathematically: invert a Fourier transform without phase information
- New methods for:
  - Solving low-resolution crystal structures using computational protein modeling
  - Solving structures using data from the new x-ray free electron lasers, which vaporize the crystal as they image it





http://lacasadeloscristales.trianatech.com/wp-content/uploads/2014/09/image005-300x300.jpg

# Nucleic acid (i.e., DNA and RNA) structure and design




Computational design of co-assembling protein–DNA nanowires (Mou et al., *Nature* 2015)

# Overview of course topics

## Structures of macromolecular complexes

# Single-particle electron microscopy

- The current revolution in structure determination through electron microscopy is due in part to new computational methods — e.g., Bayesian approaches

## THE REVOLUTION WILL NOT BE CRYSTALLIZED

**MOVE OVER X-RAY CRYSTALLOGRAPHY. CRYO-ELECTRON MICROSCOPY IS KICKING UP A STORM IN STRUCTURAL BIOLOGY BY REVEALING THE HIDDEN MACHINERY OF THE CELL.**

BY EWEN CALLAWAY

In a basement room, deep in the bowels of a steel-clad building in Cambridge, a major insurgency is under way.

A hulking metal box, some three metres tall, is quietly beaming terabytes' worth of data through thick orange cables that disappear off through the ceiling. It is one of the world's most advanced cryo-electron microscopes: a device that uses electron beams to photograph frozen biological molecules and lay bare their molecular shapes. The microscope is so sensitive that a shout can ruin an experiment, says Sjors Scheres, a structural biologist at the UK Medical Research Council Laboratory of Molecular Biology (LMB), as he stands dwarfed beside the £5-million (US$7.7-million) piece of equipment. "The UK needs many more of these, because there's going to be a boom," he predicts.

In labs around the world, cryo-electron microscopes such as this one are sending tremors through the field of structural biology. In the past three years, they have revealed exquisite details of protein-making ribosomes, quivering membrane proteins and other key cell molecules,

ILLUSTRATION BY VICTOR KOCH

172 | NATURE | VOL 525 | 10 SEPTEMBER 2015

© 2015 Macmillan Publishers Limited. All rights reserved

*Nature*, Sept. 10, 2015

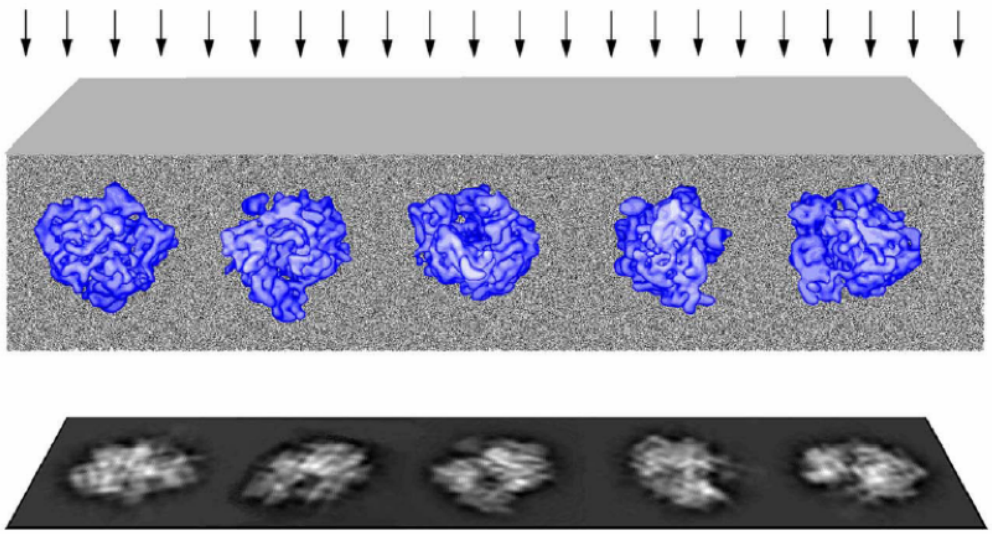
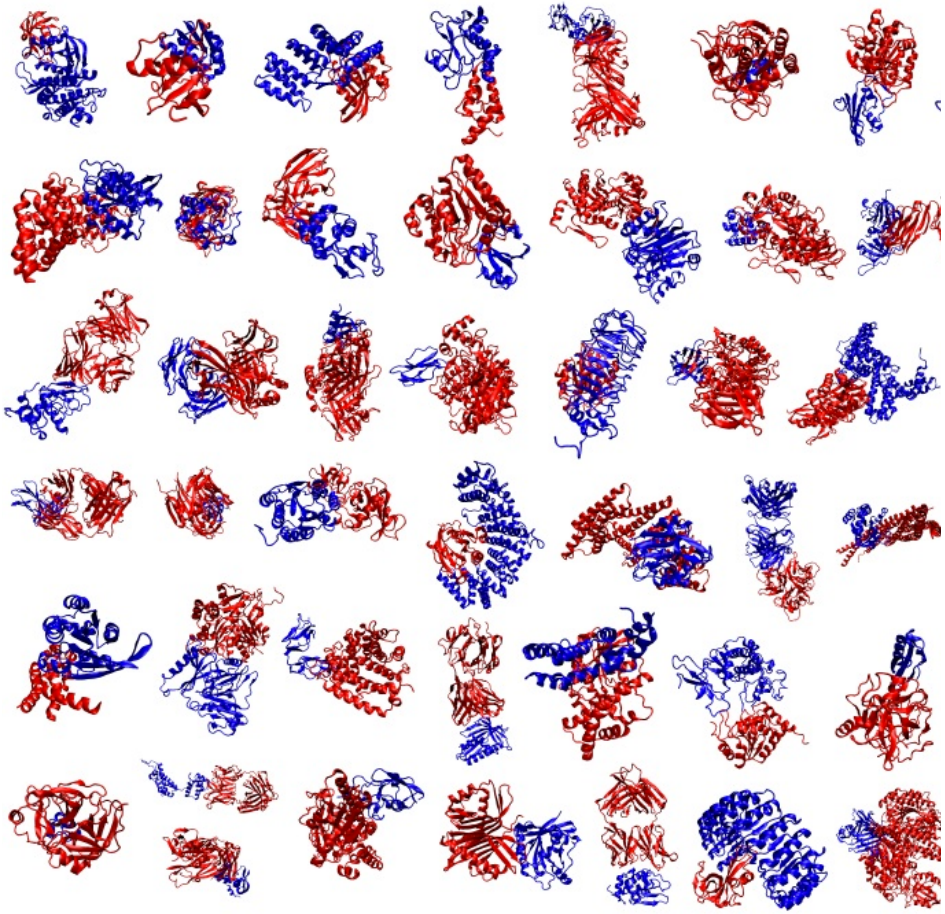


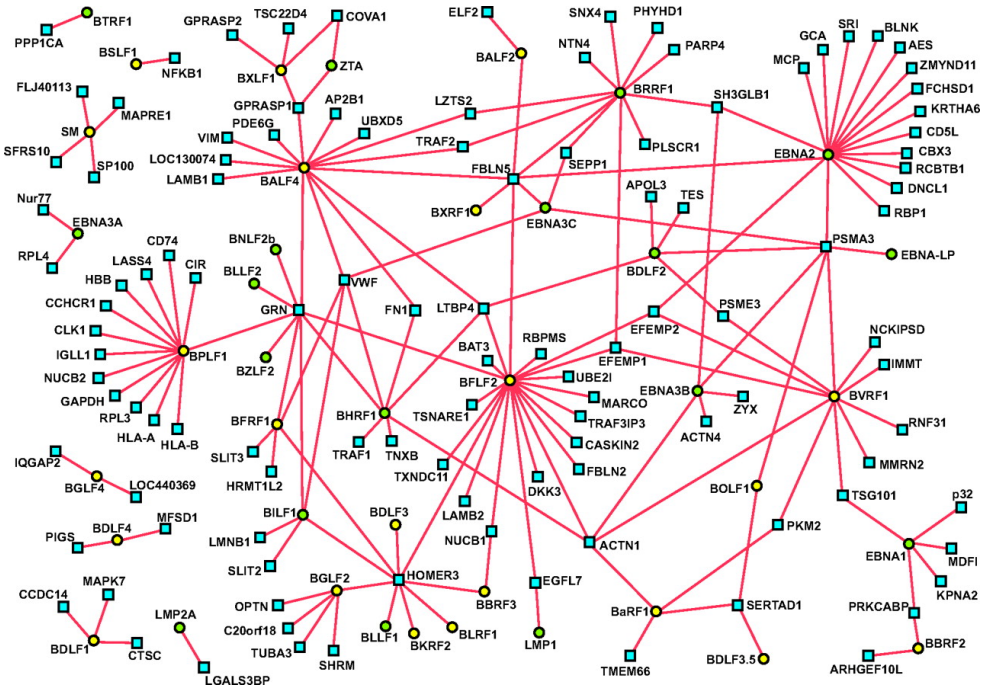

Image from Joachim Frank
http://biomachina.org/courses/structures/091.pdf

# Predicting protein–protein interactions

- Which proteins bind to one another?
- What is the structure of the bound complex?
- What types of networks do they form?

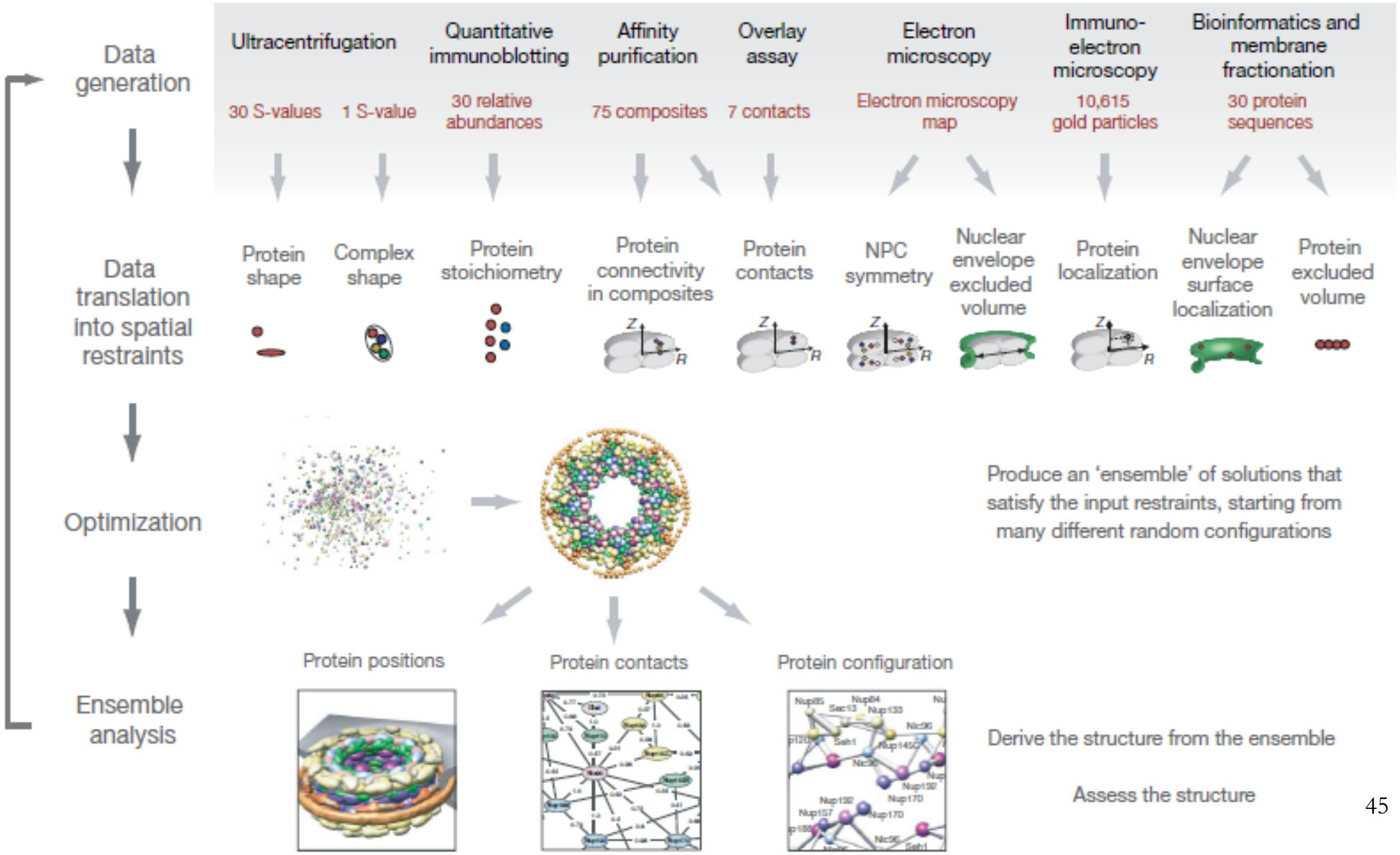


http://bmm.crick.ac.uk/~bmmadmin/Affinity/




Calderwood et al., PNAS 104:7606-11, 2007

# Integrative modeling: combining diverse experimental data to deduce structures of large complexes






Alber et al., *Nature* 2007

# Genome architecture

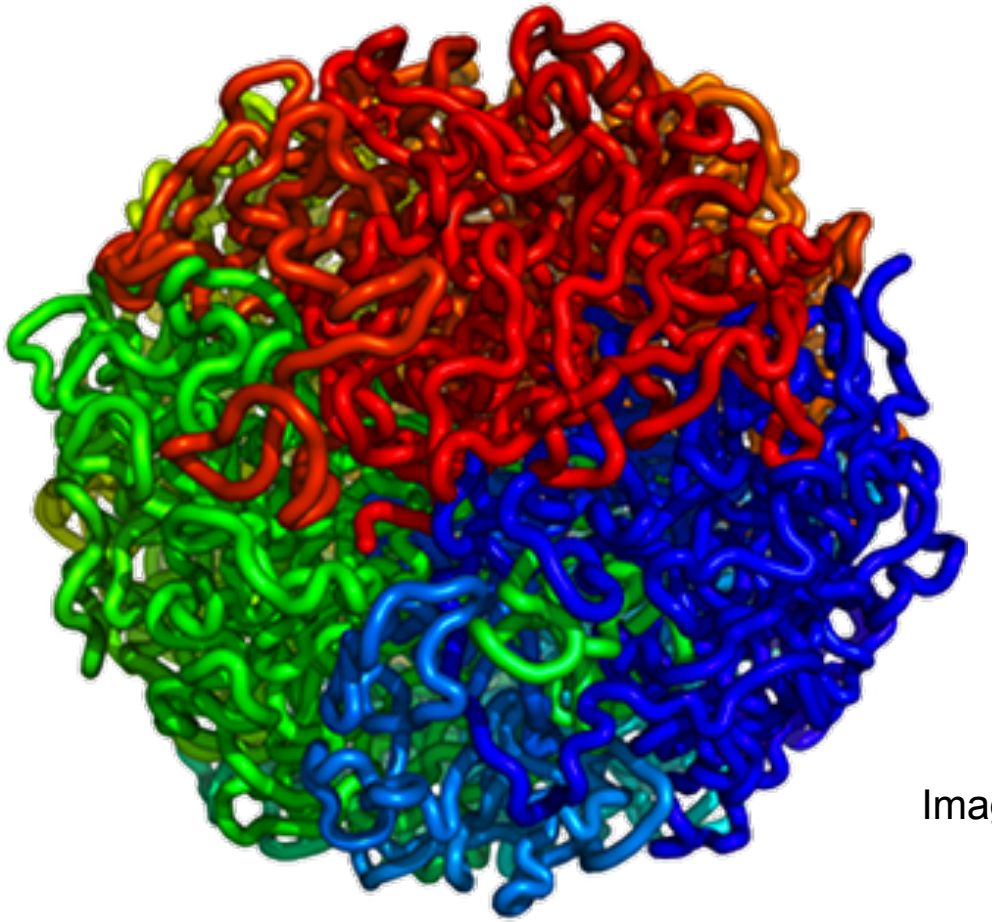


Image: http://www.aidenlab.org/

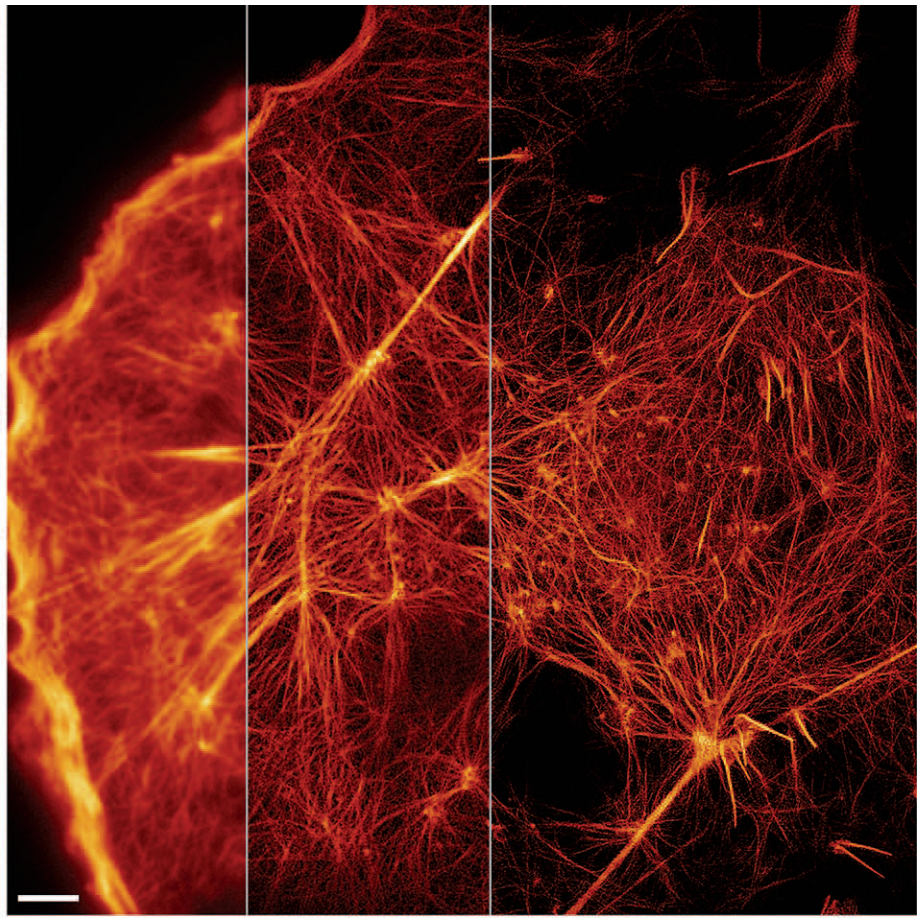
Techniques for mapping out the three-dimensional organization (and dynamics) of your DNA using a combination of computation and experiment

# Overview of course topics

## Cellular-level organization

# Super-resolution microscopy methods

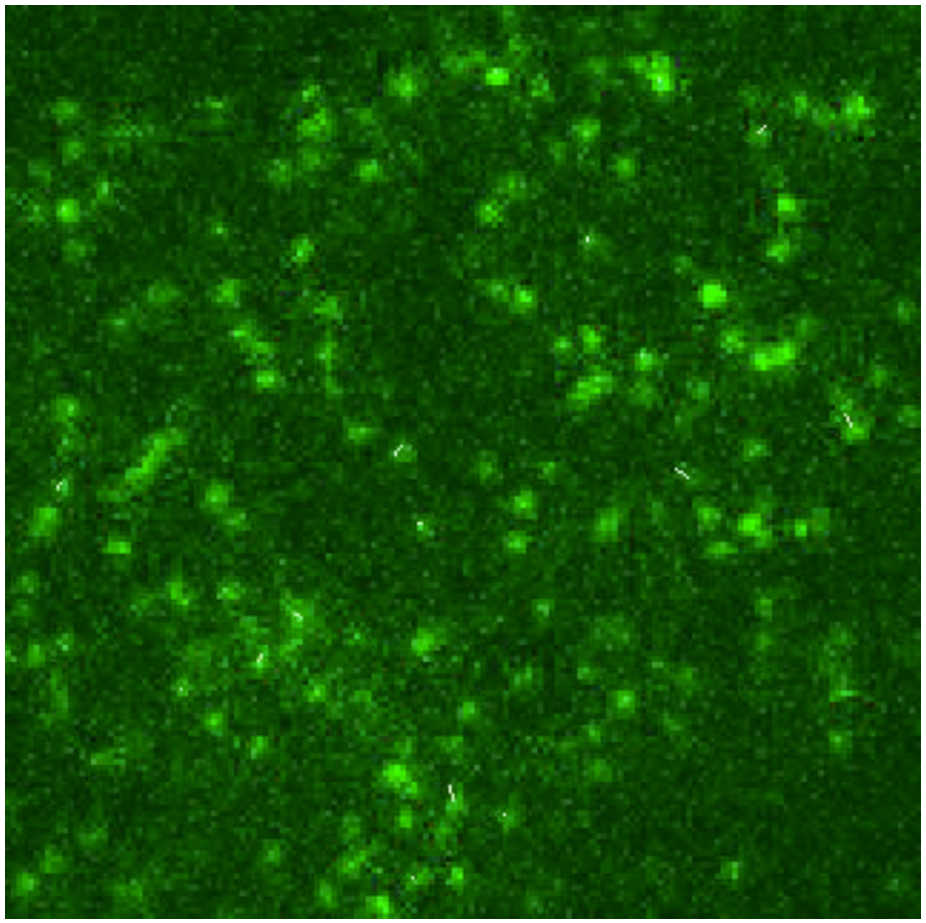
- New super-resolution microscopy methods
- Analysis methods that exploit compressed sensing





Li et al., Science 349:944 (2015)

# Tracking the motion of single molecules



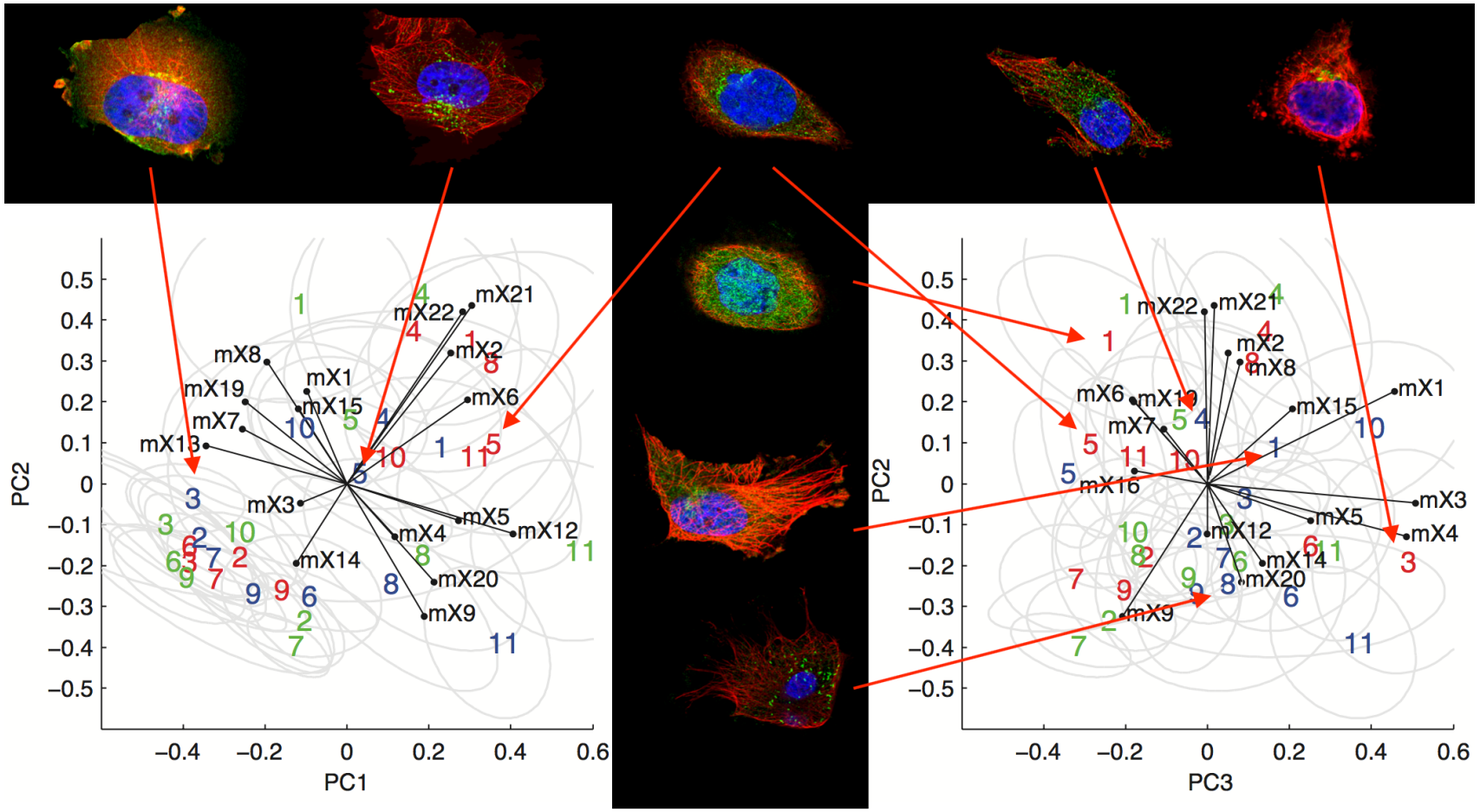
Data: Bettina van Lengerich, Natalia Jura
Tracking and movie: Robin Jia

- Fluorescence microscopy makes it possible to follow the motion of individual molecules in live cells, but actually doing the tracking is a challenging image analysis/computer vision problem

# Cellular-level simulations

http://www.readdy-project.org/

# Learning on cell shape and structure






Johnson et al., PLOS Computational Biology 11:e1004614 (2015)

# Course logistics

# Course web page

- http://cs371.stanford.edu/
- Evaluation criteria on web page
- **Please sign up on Piazza (via link on webpage) so that you get announcements**

# Course staff

- Prof. Ron Dror
  - http://drorlab.stanford.edu/rondror.html
  - Office hours: Monday and Wednesday after class (2:50-3:45), in classroom and then in Gates 204, or by appointment
- TA: Anthony Ma
  - Office hours: Tuesdays 1:30-2:30, Lathrop Tech Lounge
- TA: Osama El-Gabalawy
  - Office hours TBA

# Expected background

- Course is intended to be broadly accessible to students with computational or biological backgrounds

- You should have:
  - Introductory biology background, at the level of a first-year course, and ideally also introductory chemistry and physics background
  - Introductory computer science background (e.g., CS 106A)
  - Math through calculus, and some previous exposure to mathematical modeling or probability.

- You're *not* expected to be an expert in any of the areas we cover, but you *are* expected to be willing to read and present papers, ask questions, and share your thoughts

# Honor code

- I'm not a tough grader, but I take the honor code seriously

- In the context of this course, this means:
  - Avoid verbatim **or near-verbatim** copying of text from any source (unless it's enclosed in quotation marks and clearly attributed to the source). You need to explain ideas in your own words.
  - Make sure that all materials (including figures) in presentations and critiques are properly cited.

# Guidelines for presentations and critiques

# Giving a good presentation

- Why should you care?
  - To develop a critical skill
  - To help make this a good course!

- See “Tips for Giving Clear Talks” by Kayvon Fatahalian (Stanford CS PhD 2010)
  - http://www.cs.cmu.edu/~kayvonf/misc/cleartalktips.pdf
  - This focuses on giving talks on your own research but is applicable to the presentations for this class as well. Possible exception: I’d like you to point out both strengths and weaknesses in the work you present.

# Giving a good presentation

- Make it understandable to a diverse audience
  - Students in class are from: CS, CME, EE, BioE, ChemE, chemistry, chemical & systems biology, biophysics, and more
  - Include appropriate background material. You may need to read other, related papers as well.
- Describe the main idea *intuitively*
  - *You need to figure out what the main idea is.* (Ask course staff for help if necessary.)
  - Do not rely on equations or code to convey intuition
  - You do not need to cover everything in the paper
- Strive for clarity
  - See Kayvon's presentation
- Make it interesting and exciting!

# Preparing for presentations

- Each presenter should meet with a TA, and then with me (Ron), before their presentation
  - By default, meet with me right after the class preceding your presentation. If one of the presenters has a conflict then, schedule a different time.
- You should prepare a complete draft of your presentation before these meetings
- You need to coordinate with the other students presenting on the same day.
  - For example, you may want shared introductory material
  - I would like to meet with all of you together

# Writing a critique

- Although you should include a brief summary of the paper you're critiquing, that shouldn't be the focus
- Instead, the focus should be on:
  - Strengths and weaknesses of the approach or results
  - Other approaches that might have been applied instead
  - Potential extensions or follow-on work
- You're encouraged to read related papers to help you write this.
  - Please indicate in the critique which other papers you read.

# Immediate next steps

# Choosing topics for presentation and critique

- By next Tuesday, Jan. 17, email cs371-win1617-staff@lists.stanford.edu a list of six topics (class sessions) you'd be interested in presenting on and twelve papers (or six topics) you'd be interested in critiquing
  - My default list of papers is at http://cs371.stanford.edu/schedule.html and will be complete by Tuesday afternoon. The "Main Papers" are the ones I'm expecting students to present.
  - You're also welcome to suggest your own paper(s) of interest to present. If you do, specify any dates on which you can't present.
  - You may rank order your lists if you'd like (you should specify that you've done so)
- Also include:
  - Brief description of your background

# Next couple class periods

- Wednesday, Jan. 11: Ron to present on simulation of drug targets and simulation analysis
  - Reading materials on website
- Monday, Jan. 16: Guest lecture by Eli Groban (Autodesk) on virtual reality for biomolecules
  - *Completely optional (MLK day).*

# Volunteers for initial presentations & critiques

- Email cs371-win1617-staff@lists.stanford.edu ASAP to volunteer to present or critique during the following four class periods:
  - Jan. 18: Using multiplayer online video games for structure prediction and design
  - Jan. 23: Coevolution methods for predicting structure from large numbers of genetic sequences
  - Jan. 25: Modern protein design
  - Jan. 27: Machine learning for structure-based virtual screening