STRUCTURAL BIOINFORMATICS

Barry Grant
University of Michigan
www.thegrantlab.org

MODULE OVERVIEW

Objective: Provide an introduction to the practice of bioinformatics as well as a practical guide to using common bioinformatics databases and algorithms

1.1. ▸ *Introduction to Bioinformatics*

1.2. ▸ *Sequence Alignment and Database Searching*

1.3 ▸ *Structural Bioinformatics*

1.4 ▸ *Genome Informatics: High Throughput Sequencing Applications and Analytical Methods*

WEEK TWO REVIEW

☑ **Answers to last weeks homework** (19/19)**:**

Answers week 2

☑ **Muddy Point Assessment** (11/19)**:**

Responses

- *“More time to finish the assignment”*
- *“I felt there was too much material to cover in one lab”*
- *“The [NCBI] sites were so slow”*
- *“More time with HMMER would be helpful”*
- *“Very nice lab”*

Q18: NW DYNAMIC PROGRAMMING

Match: +2
Mismatch: -1
Gap: -2

		A	G	T	T	C
	0	-2	-4	-6	-8	-10
A	-2	+2	0	-2	-4	-6
T	-4	0	+1	+2	0	-2
T	-6	-2	-1	+3	+4	+2
G	-8	-4	0	+1	+2	+3
C	-10	-6	-2	-1	0	**+4**

```
A T T G C
|   |   |
A G T T C
```

```
A - T T G C
|   | |   |
A G T T - C
```

THIS WEEK'S HOMEWORK

- Check out the "**Background Reading**" material online:
  - Achievements & Challenges in Structural Bioinformatics
  - Protein Structure Prediction
  - Biomolecular Simulation
  - Computational Drug Discovery

- Complete the **lecture 1.3 homework questions**:
  http://tinyurl.com/bioinf525-quiz3

"*Bioinformatics is the application of computers to the collection, archiving, organization, and analysis of biological data.*"

… A hybrid of biology and computer science

"*Bioinformatics is the application of computers to the collection, archiving, organization, and analysis of biological data.*"

Bioinformatics is computer aided biology!

"*Bioinformatics is the application of computers to the collection, archiving, organization, and analysis of biological data.*"

Bioinformatics is computer aided biology!

Goal: Data to Knowledge

So what is **structural bioinformatics**?

So what is **structural bioinformatics**?

… computer aided structural biology!

Aims to characterize and interpret biomolecules and their assembles at the molecular & atomic level

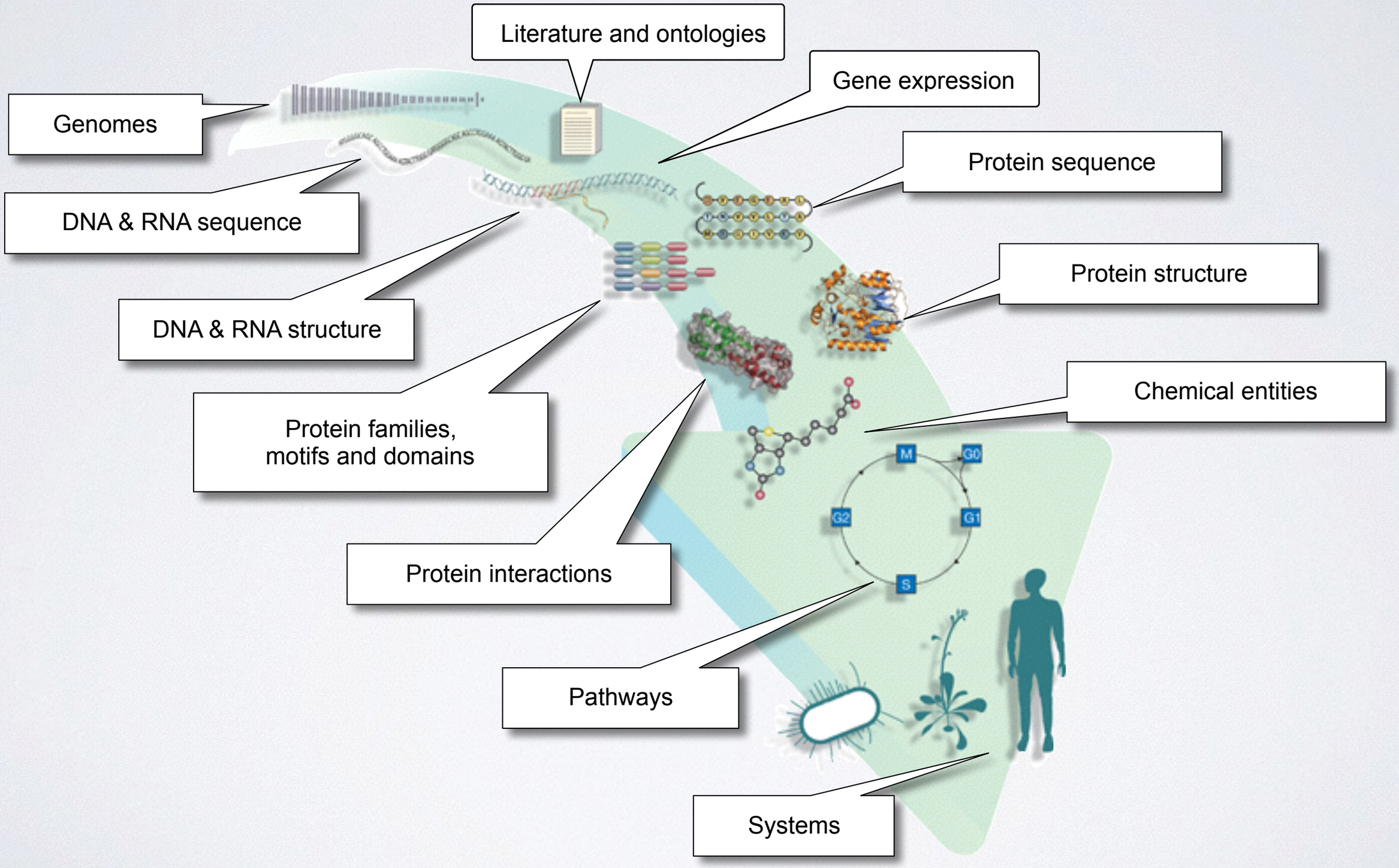
Why should we care?

Why should we care?

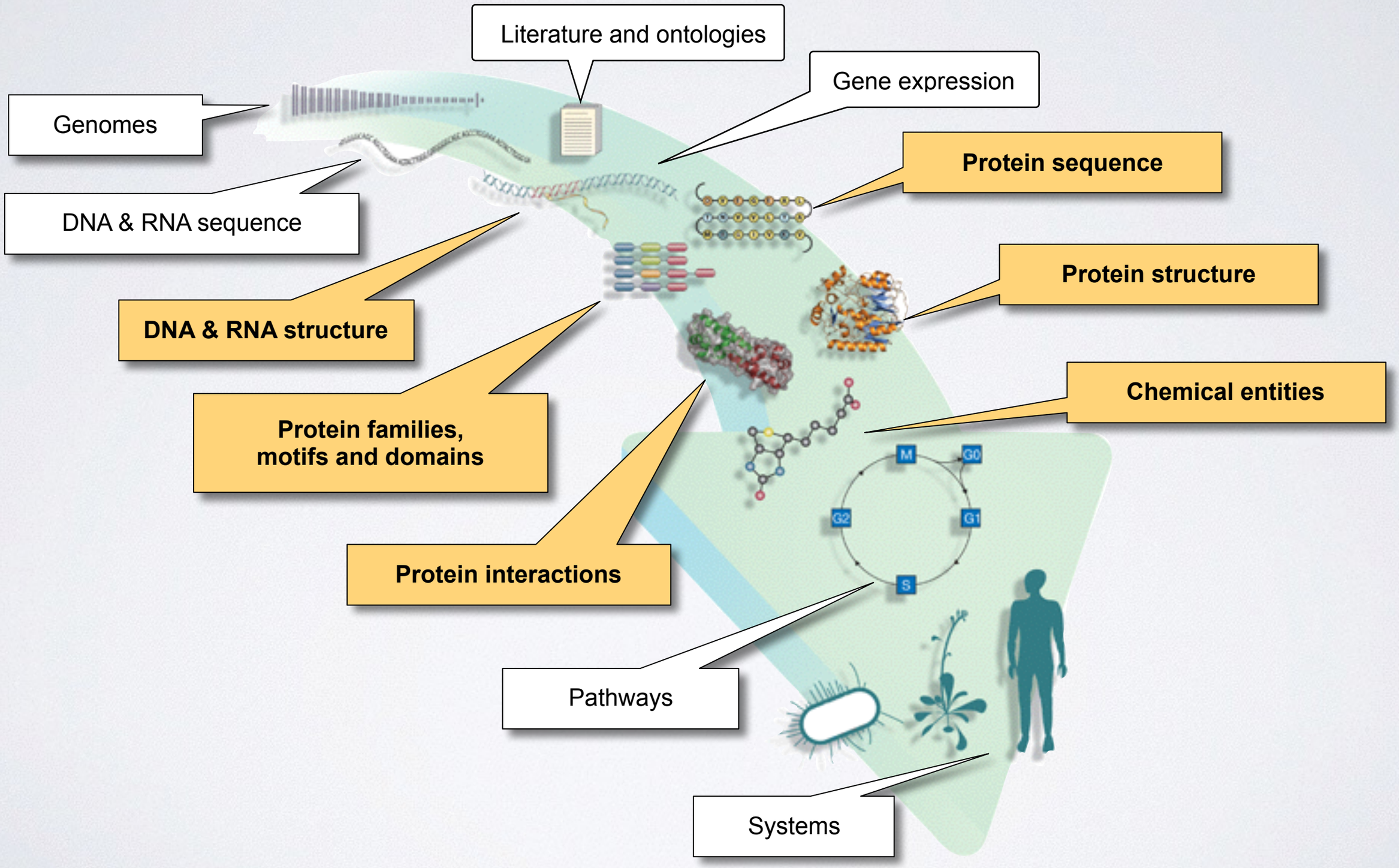
Because biomolecules are “nature’s robots”

… and because it is only by coiling into **specific 3D structures** that they are able to perform their functions

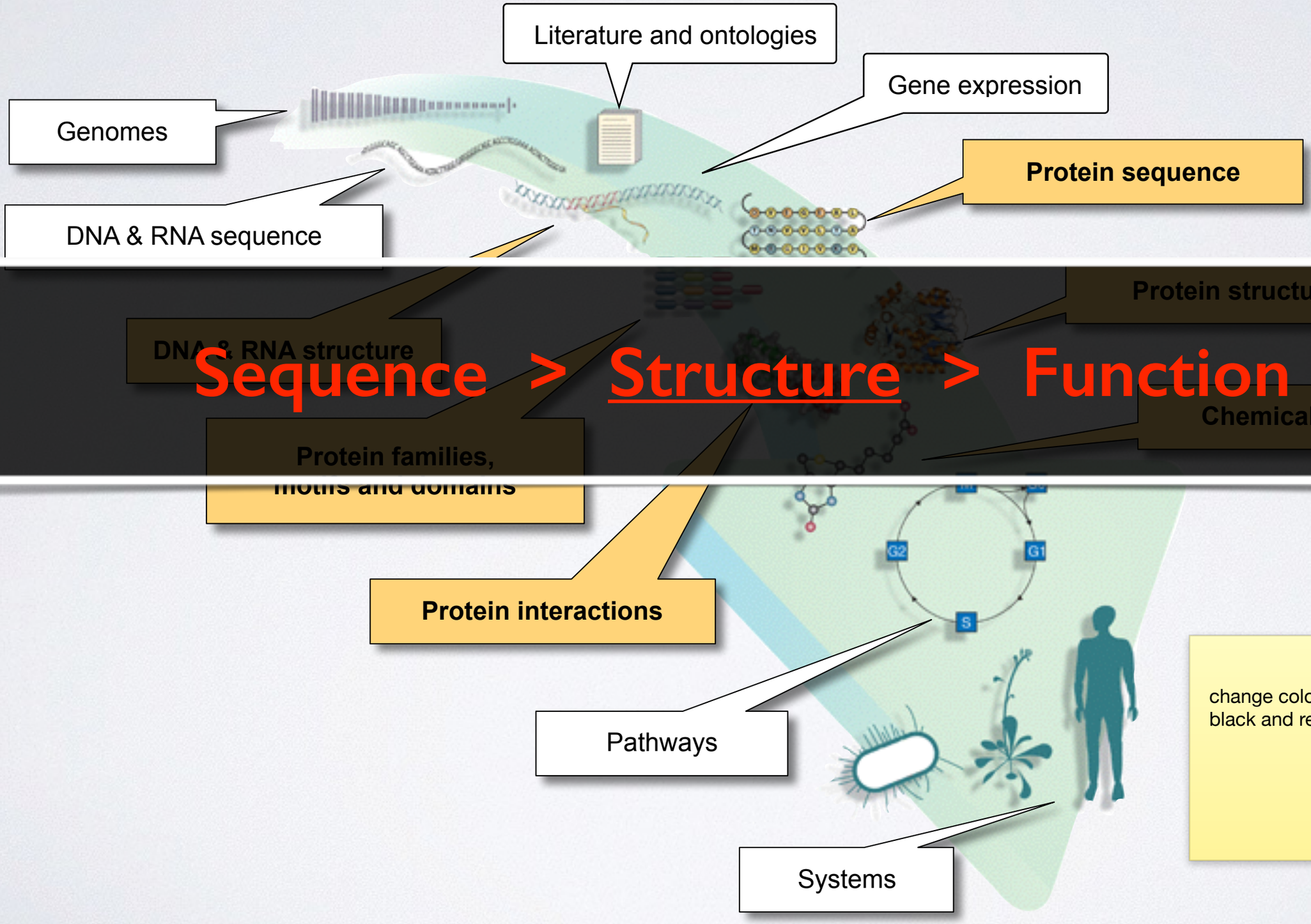
BIOINFORMATICS DATA



STRUCTURAL DATA IS CENTRAL

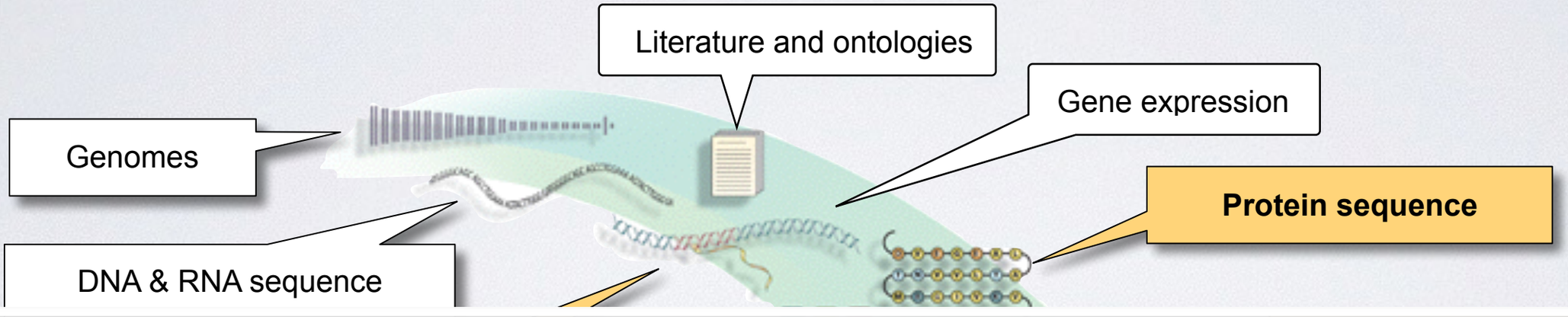


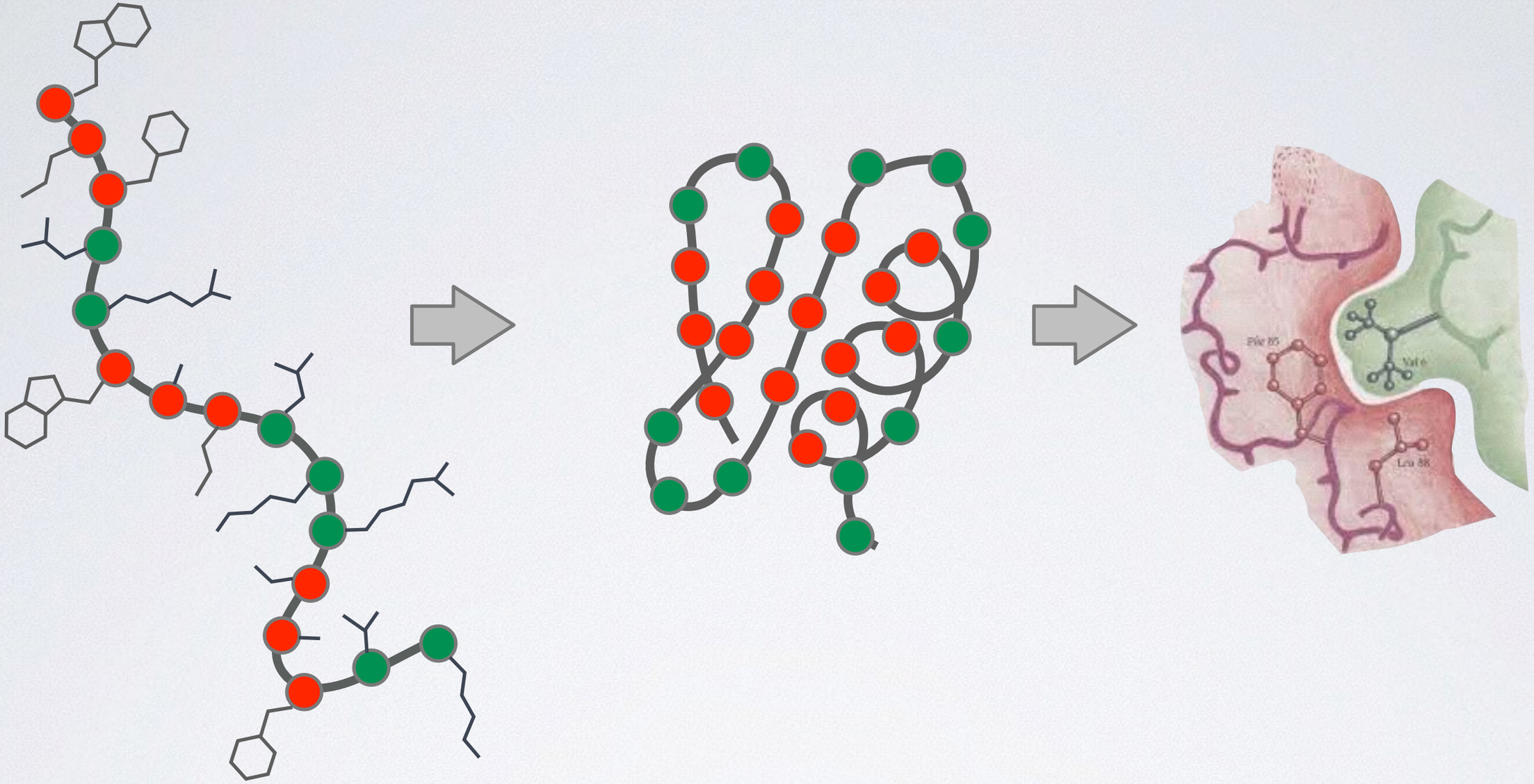
STRUCTURAL DATA IS CENTRAL



change color to gray and yellow from black and red?

STRUCTURAL DATA IS CENTRAL





Sequence	Structure	Function
• Unfolded chain of amino acid chain • Highly mobile • Inactive	• Ordered in a precise 3D arrangment • Stable but dynamic	• Active in specific “conformations” • Specific associations & precise reactions

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

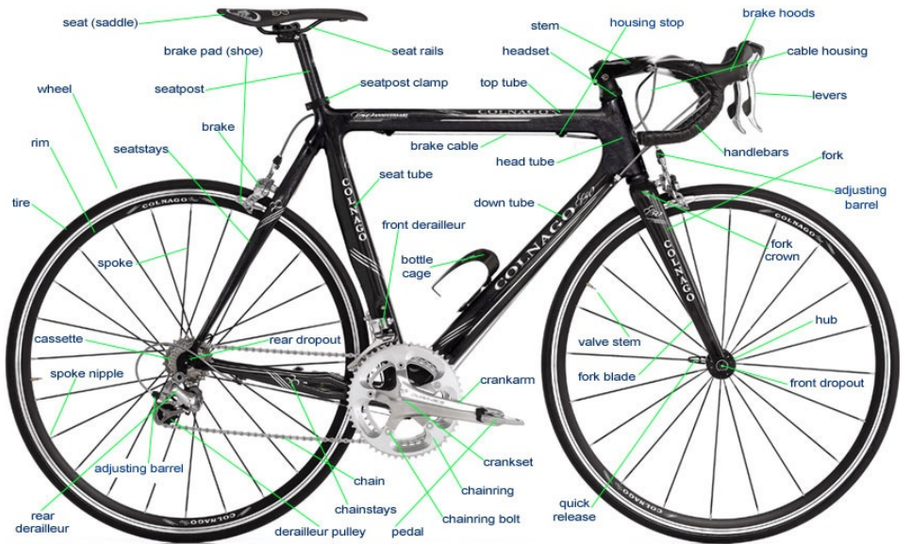
Genomics is a great start ….

Track Bike — DL 175

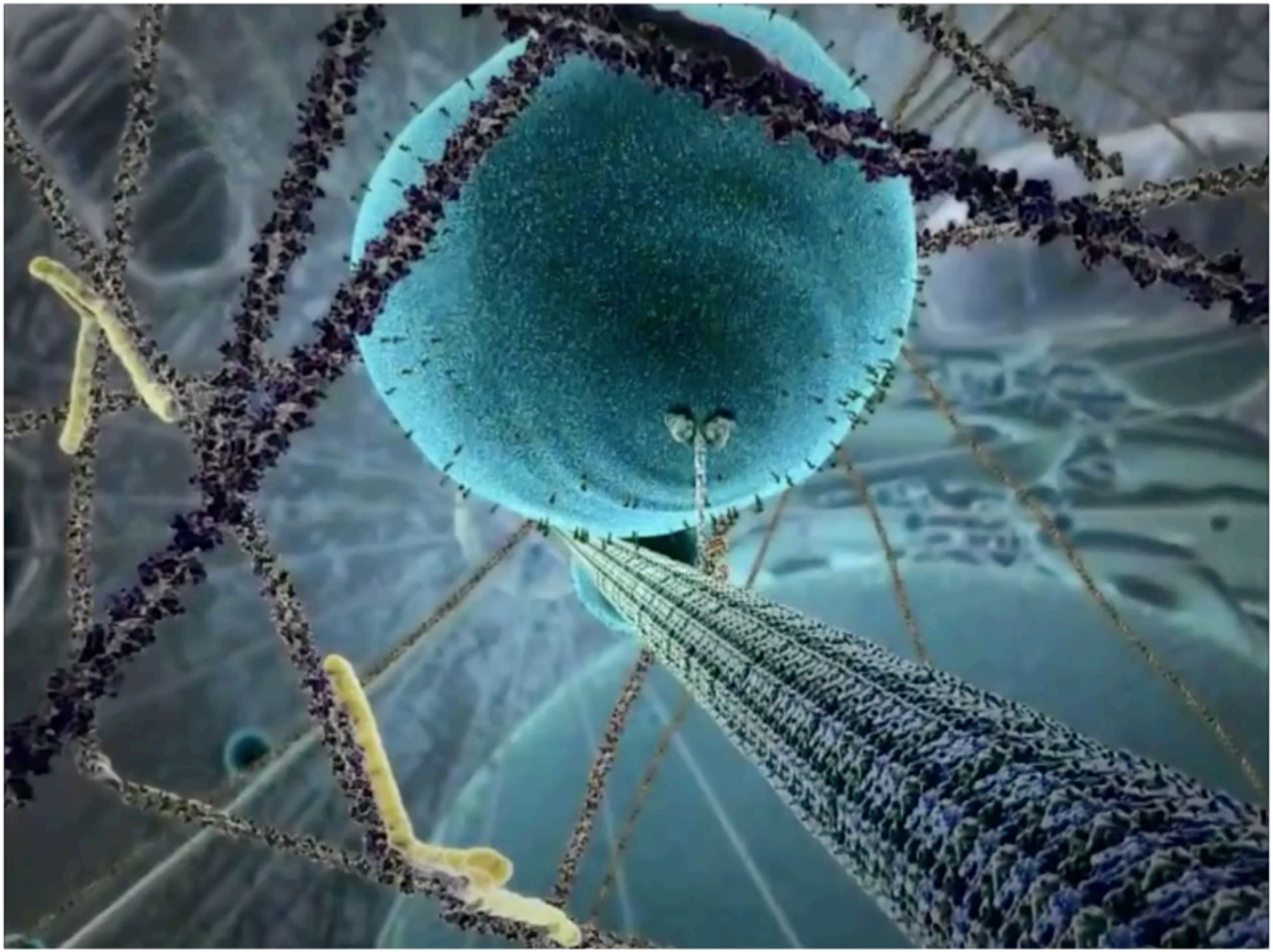
REF. NO.	IBM NO.	DESCRIPTION
1	156011	Track Frame 21", 22", 23", 24", Team Red
2	157040	Fork for 21" Frame
2	157039	Fork for 22" Frame
2	157038	Fork for 23" Frame
2	157037	Fork for 24" Frame
3	191202	Handlebar TTT Competition Track Alloy 15/16"
4		Handlebar Stem, TTT, Specify extension
5	191278	Expander 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

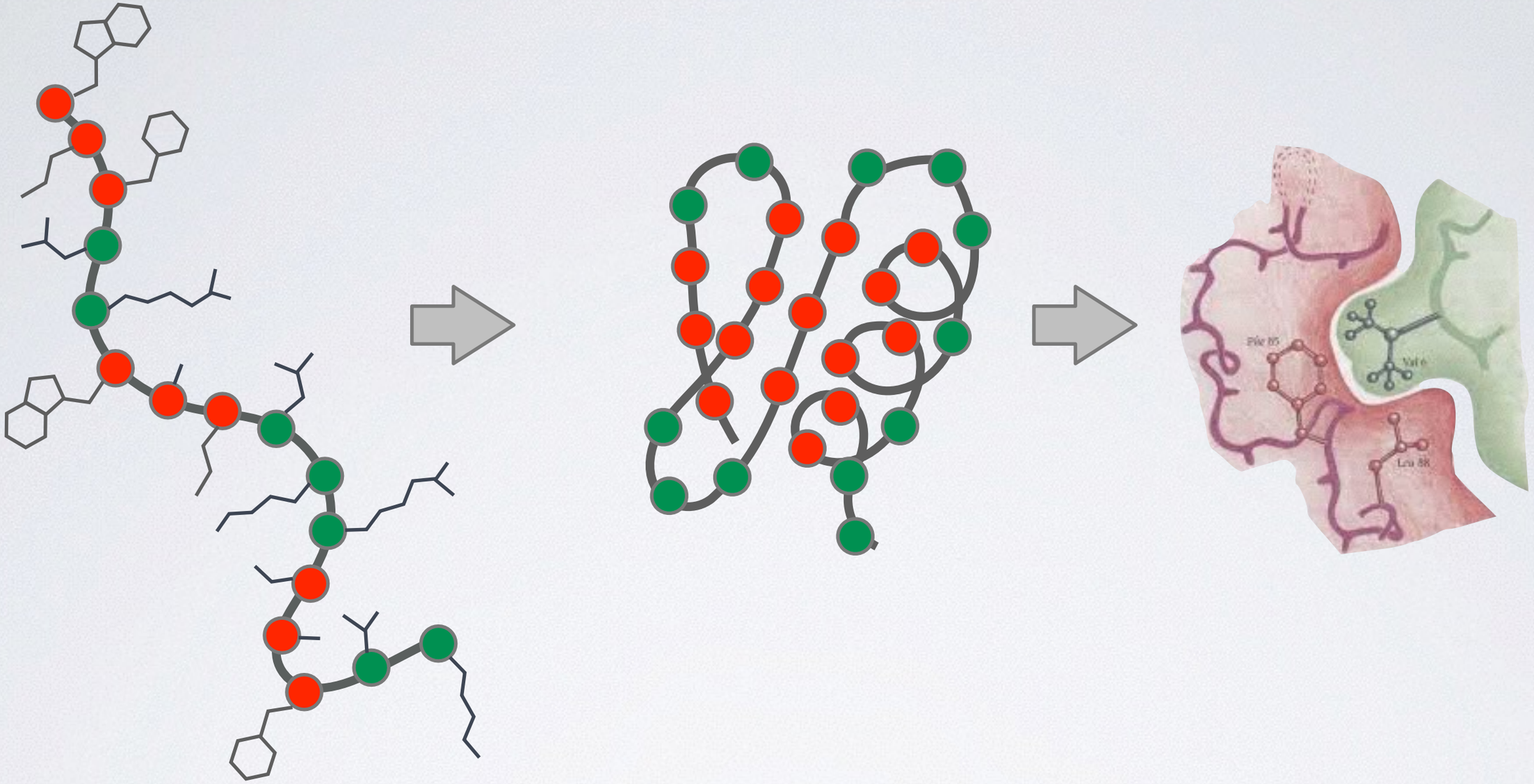
… but not the end



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

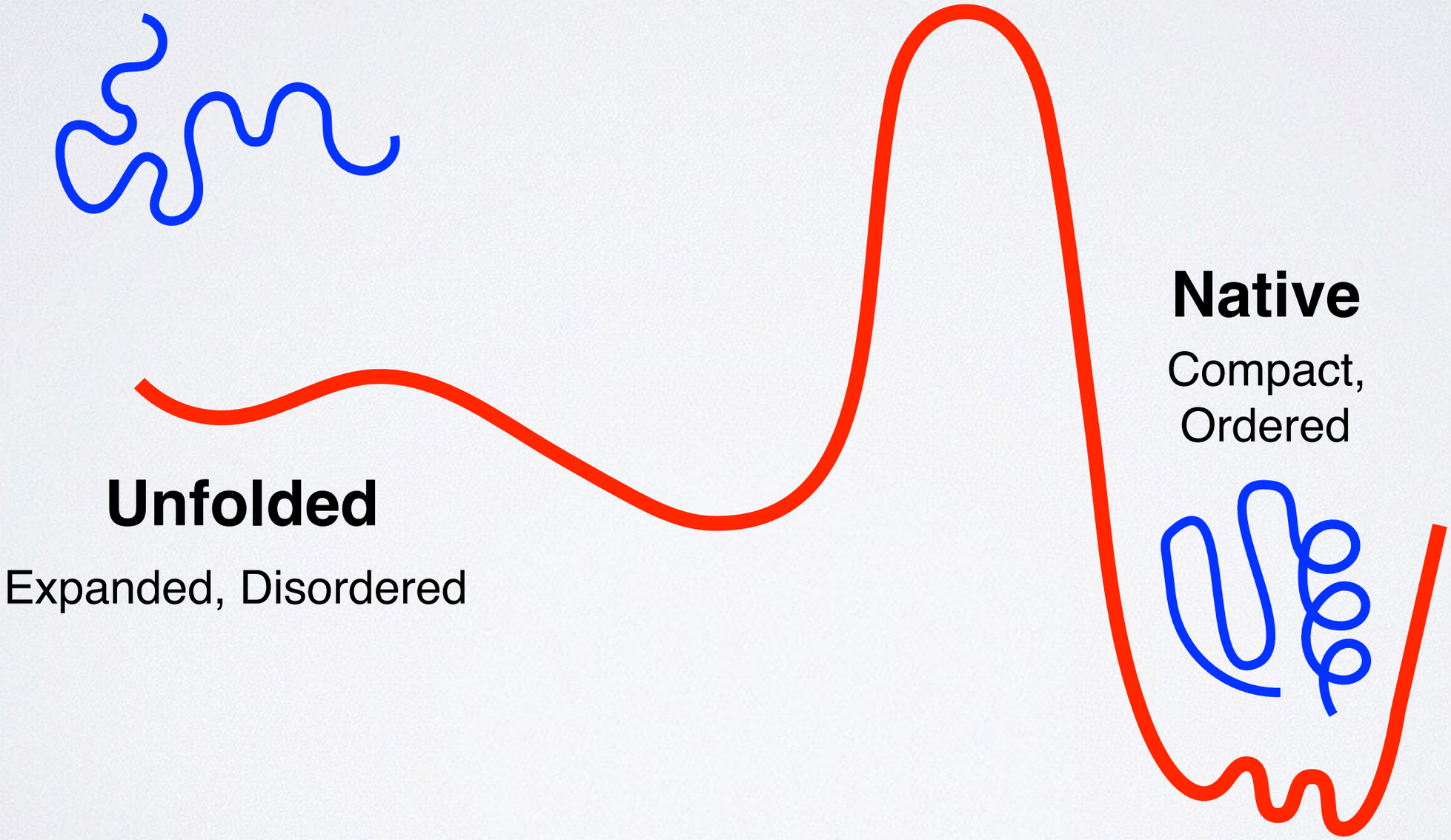


Extracted from The Inner Life of a Cell by Cellular Visions and Harvard
[YouTube link: https://www.youtube.com/watch?v=y-uuk4Pr2i8]



Sequence	Structure	Function
• Unfolded chain of amino acid chain • Highly mobile • Inactive	• Ordered in a precise 3D arrangment • Stable but dynamic	• Active in specific “conformations” • Specific associations & precise reactions

KEY CONCEPT: ENERGY LANDSCAPE



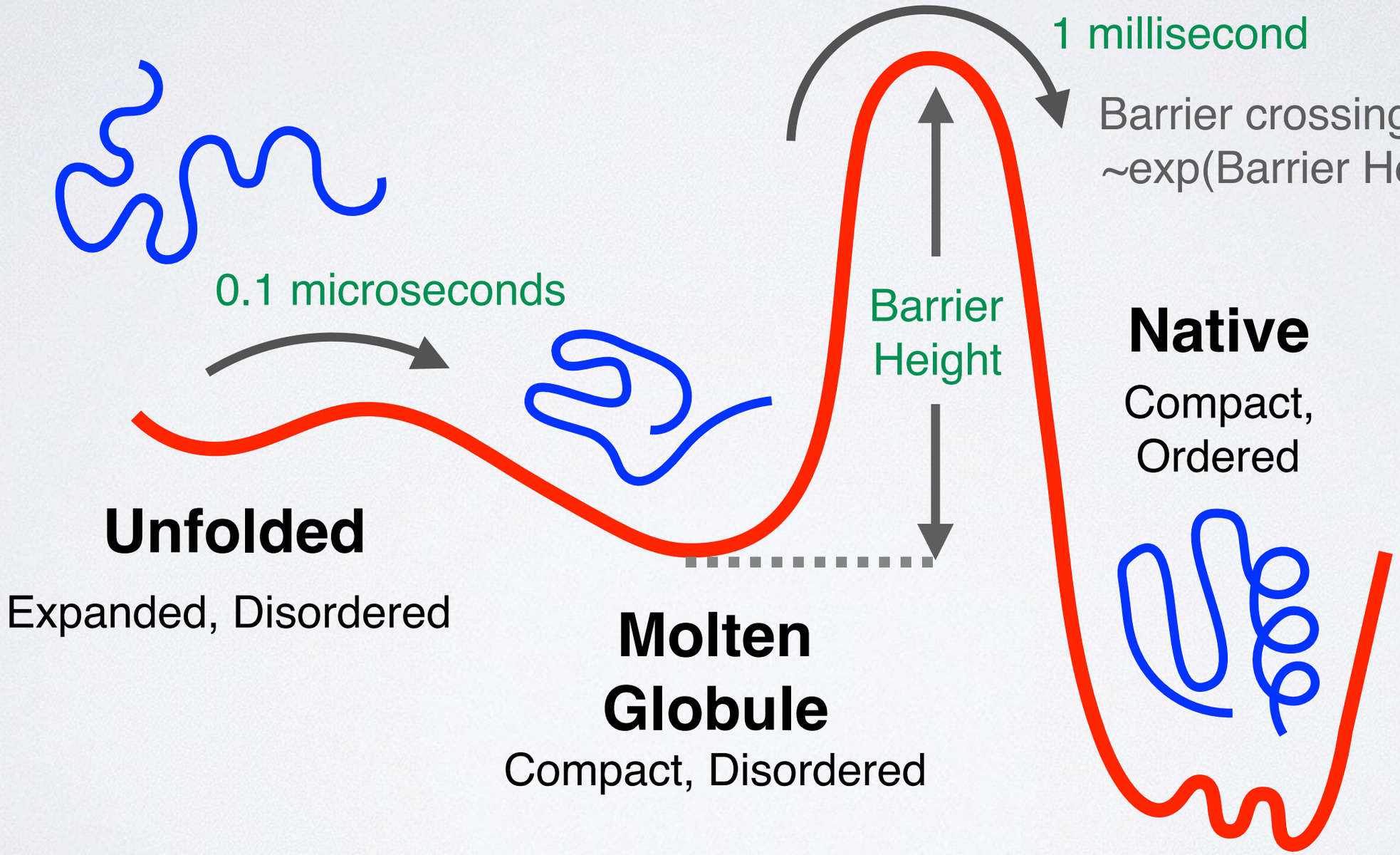
Unfolded

Expanded, Disordered

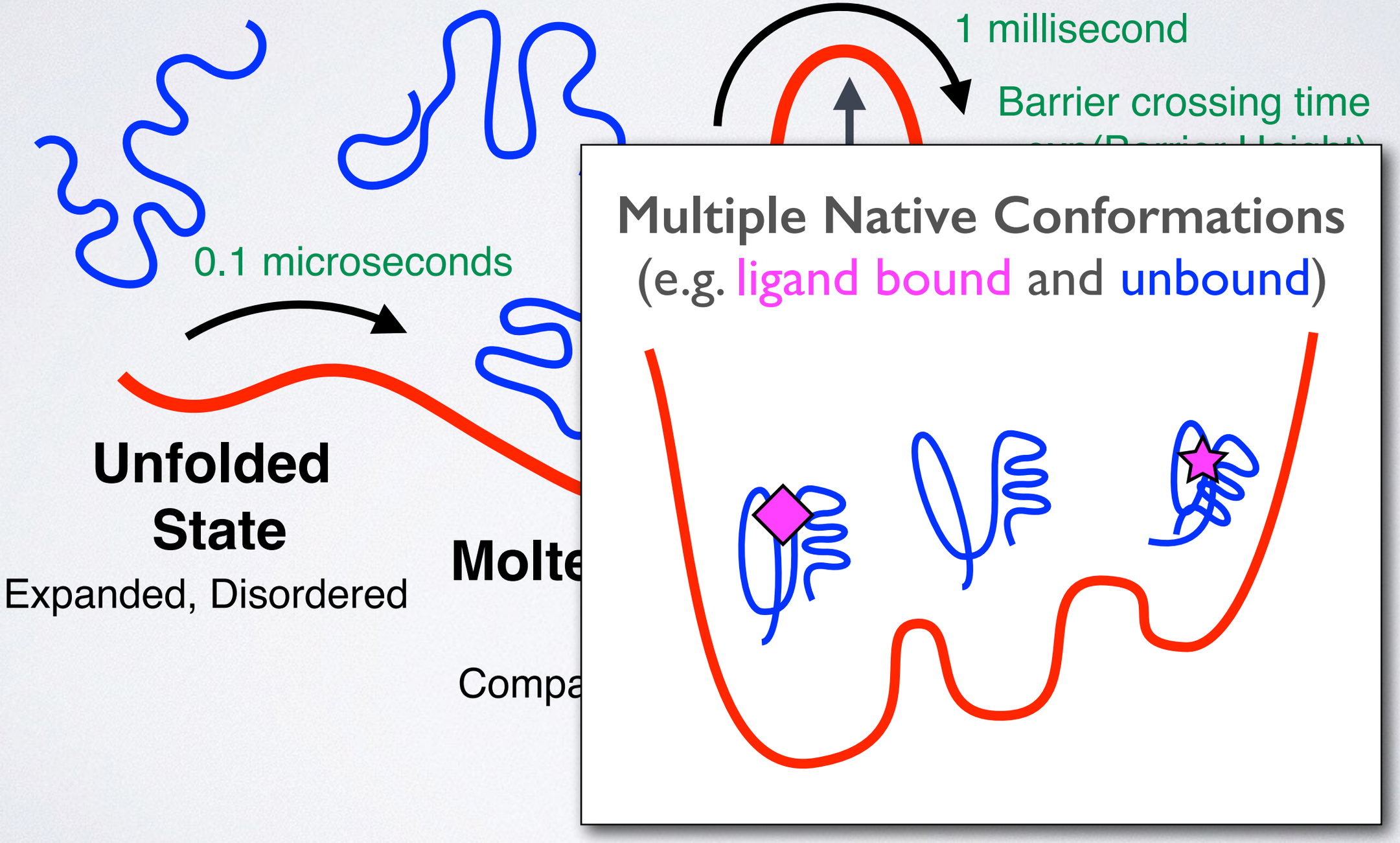
Native

Compact,
Ordered

KEY CONCEPT: ENERGY LANDSCAPE



KEY CONCEPT: **ENERGY LANDSCAPE**



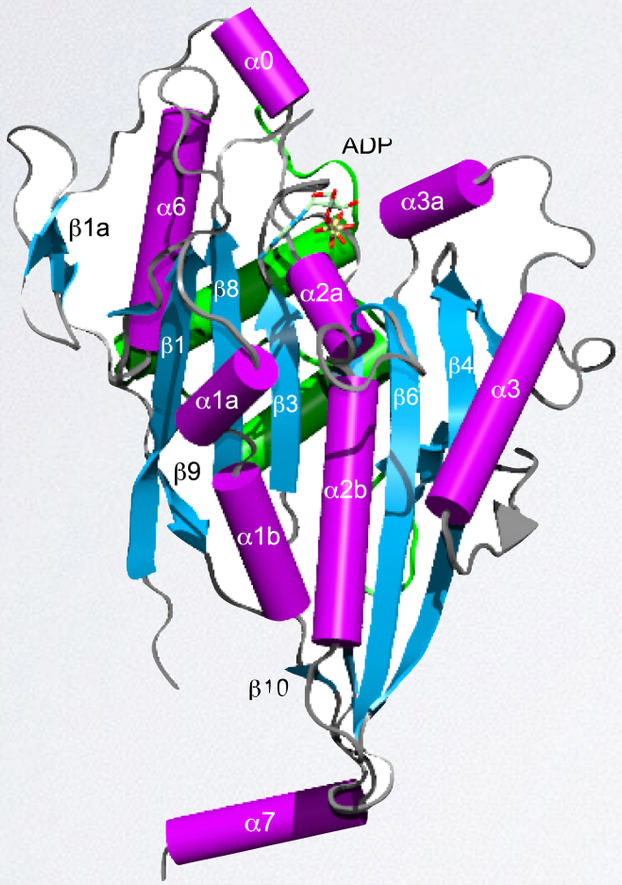
OUTLINE:

- **Overview of structural bioinformatics**
  - Major motivations, goals and challenges
- **Fundamentals of protein structure**
  - Composition, form, forces and dynamics
- **Representing and interpreting protein structure**
  - Modeling energy as a function of structure
- **Example application areas**
  - Predicting functional dynamics & drug discovery

OUTLINE:

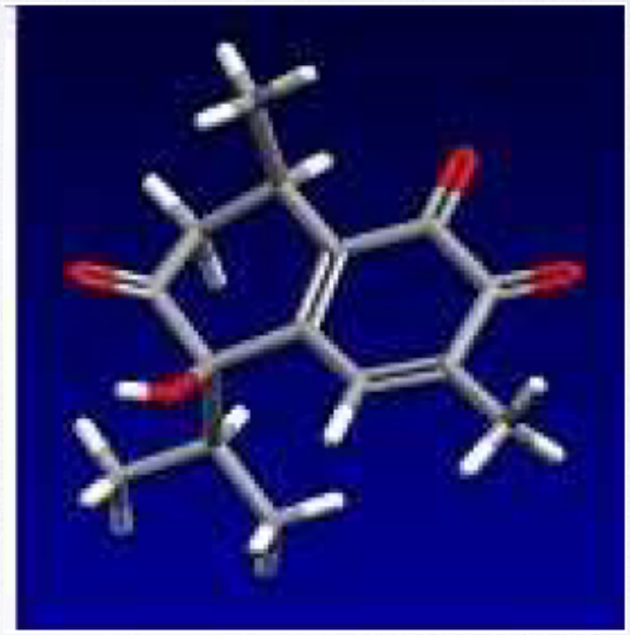
- **Overview of structural bioinformatics**
  - Major motivations, goals and challenges
- **Fundamentals of protein structure**
  - Composition, form, forces and dynamics
- **Representing and interpreting protein structure**
  - Modeling energy as a function of structure
- **Example application areas**
  - Predicting functional dynamics & drug discovery

TRADITIONAL FOCUS **PROTEIN**, **DNA** AND **SMALL MOLECULE** DATA SETS WITH **MOLECULAR STRUCTURE**



Protein
(PDB)

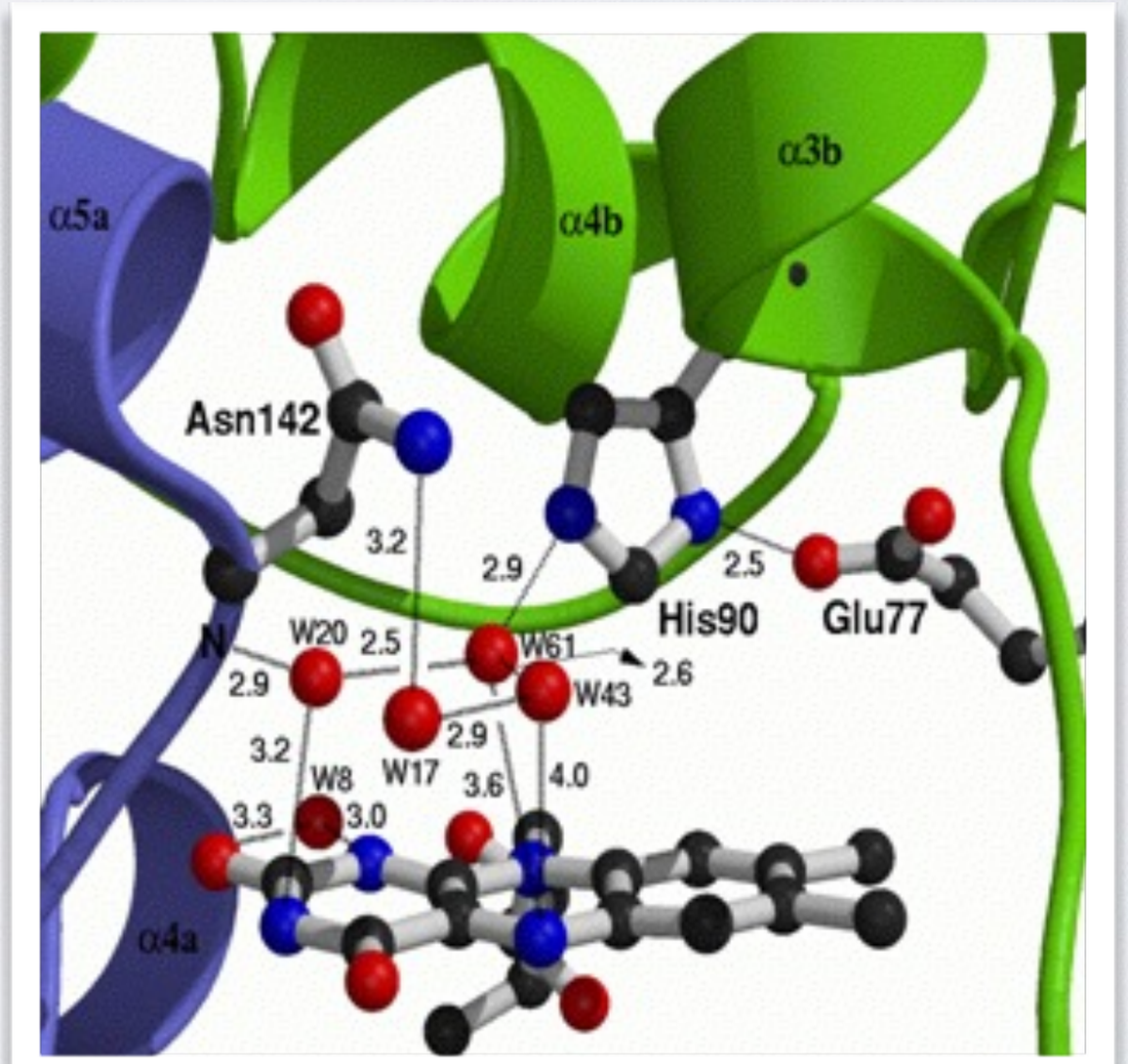
DNA
(NDB)



Small Molecules
(CCDB)

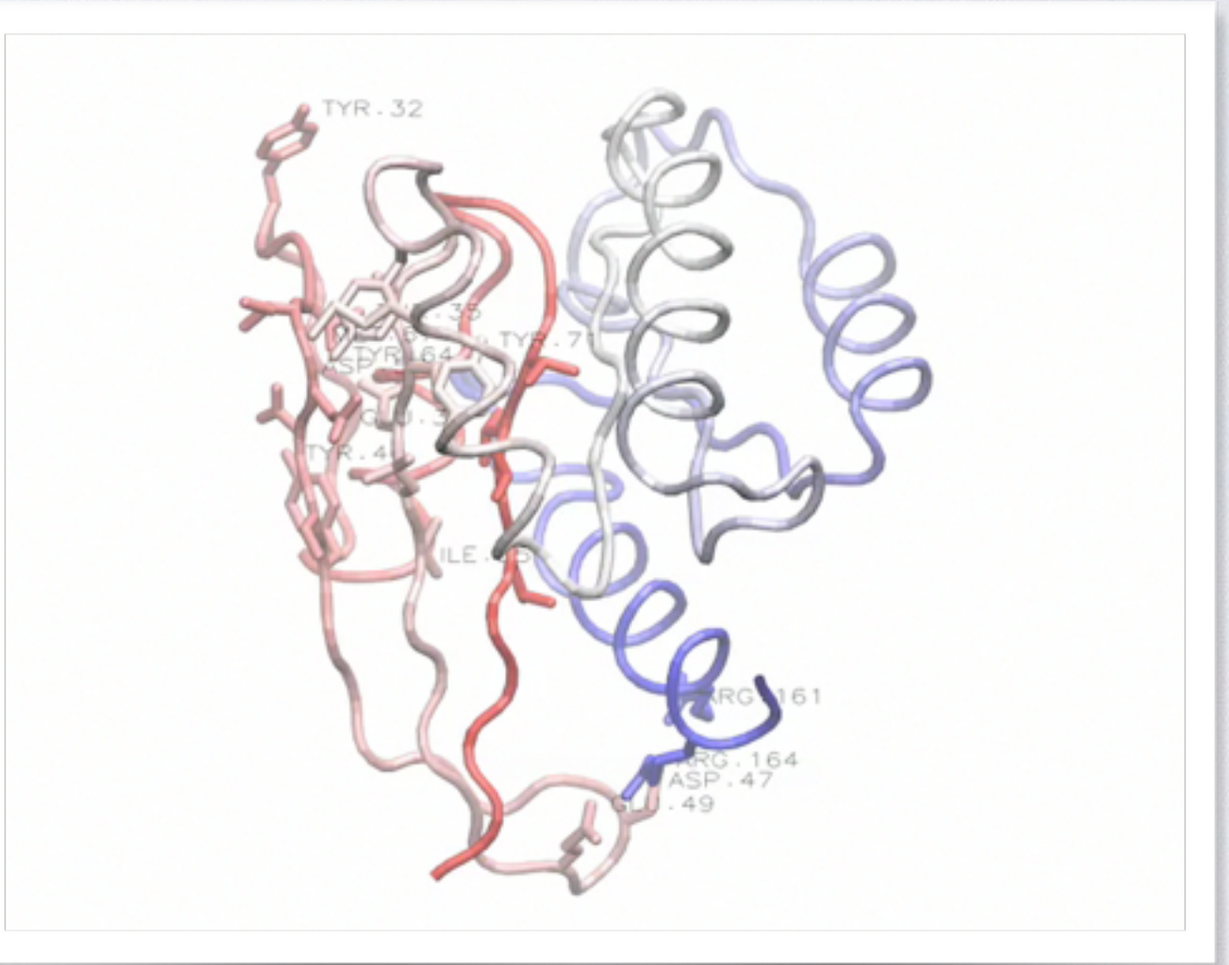
Motivation 1:
Detailed understanding of molecular interactions

Provides an invaluable structural context for conservation and mechanistic analysis leading to functional insight.



Motivation 1:
Detailed understanding of molecular interactions

Computational modeling can provide detailed insight into functional interactions, their regulation and potential consequences of perturbation.

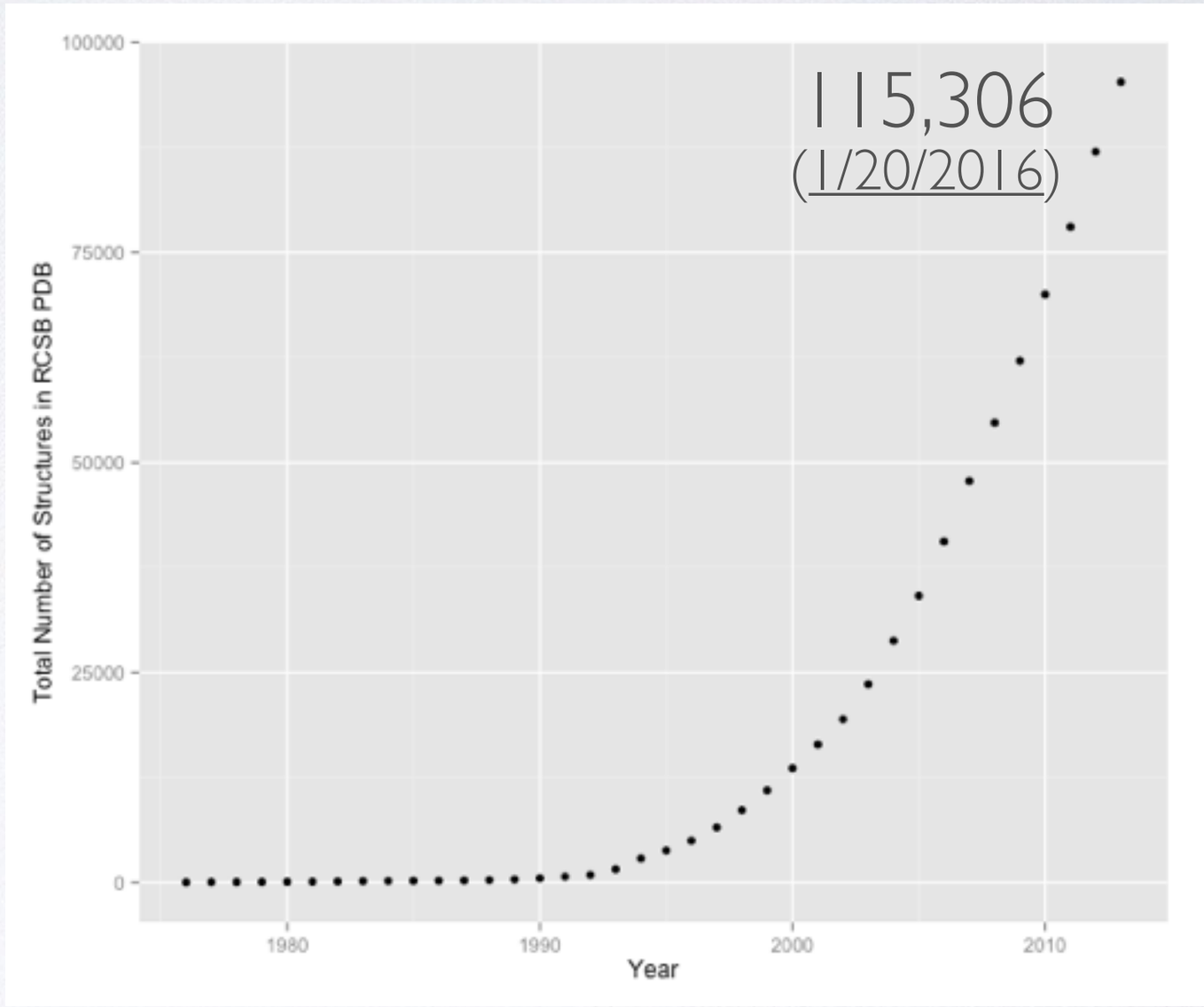


Grant *et al.* PLoS. Comp. Biol. (2010)

Motivation 2:

Lots of structural data is becoming available

Structural Genomics has contributed to driving down the cost and time required for structural determination



Data from: http://www.rcsb.org/pdb/statistics/

Motivation 2:

Lots of structural data is becoming available

Structural Genomics has contributed to driving down the cost and time required for structural determination

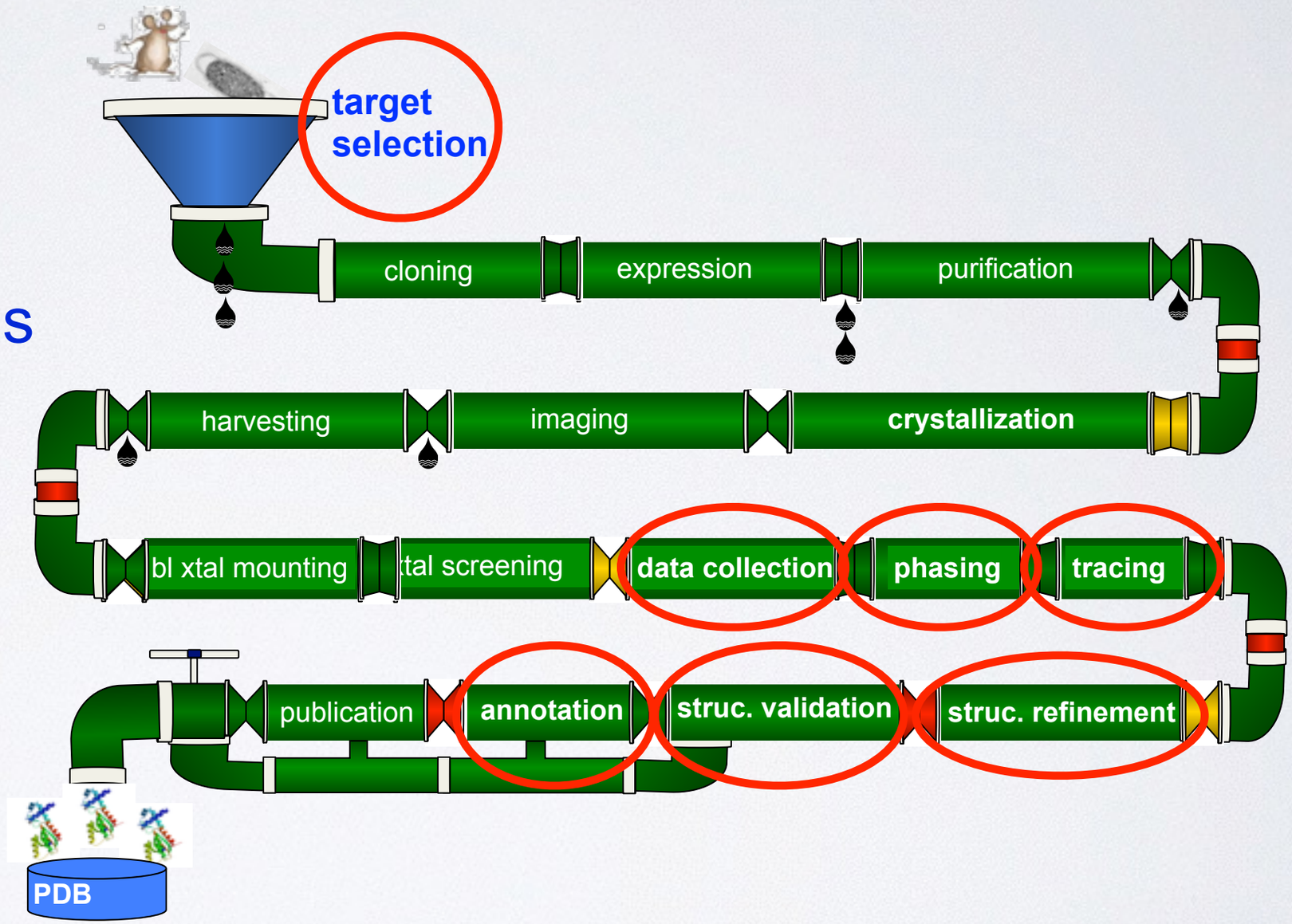
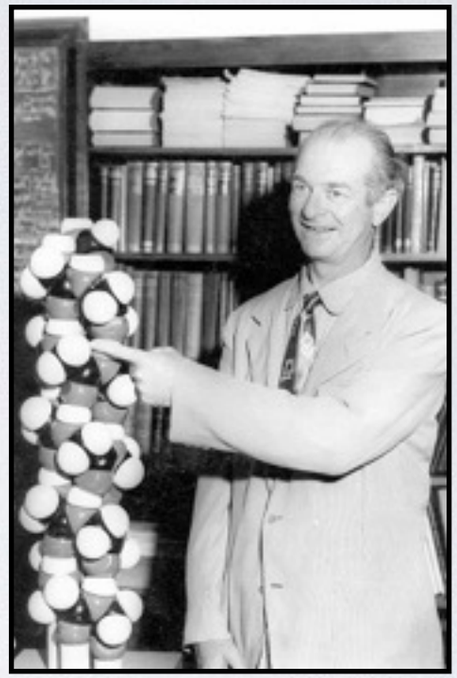


Image Credit: "Structure determination assembly line" Adam Godzik

Motivation 3:

Theoretical and computational predictions have been, and continue to be, enormously valuable and influential!



SUMMARY OF KEY **MOTIVATIONS**

Sequence > Structure > Function

- Structure determines function, so understanding structure helps our understanding of function

Structure is more conserved than sequence

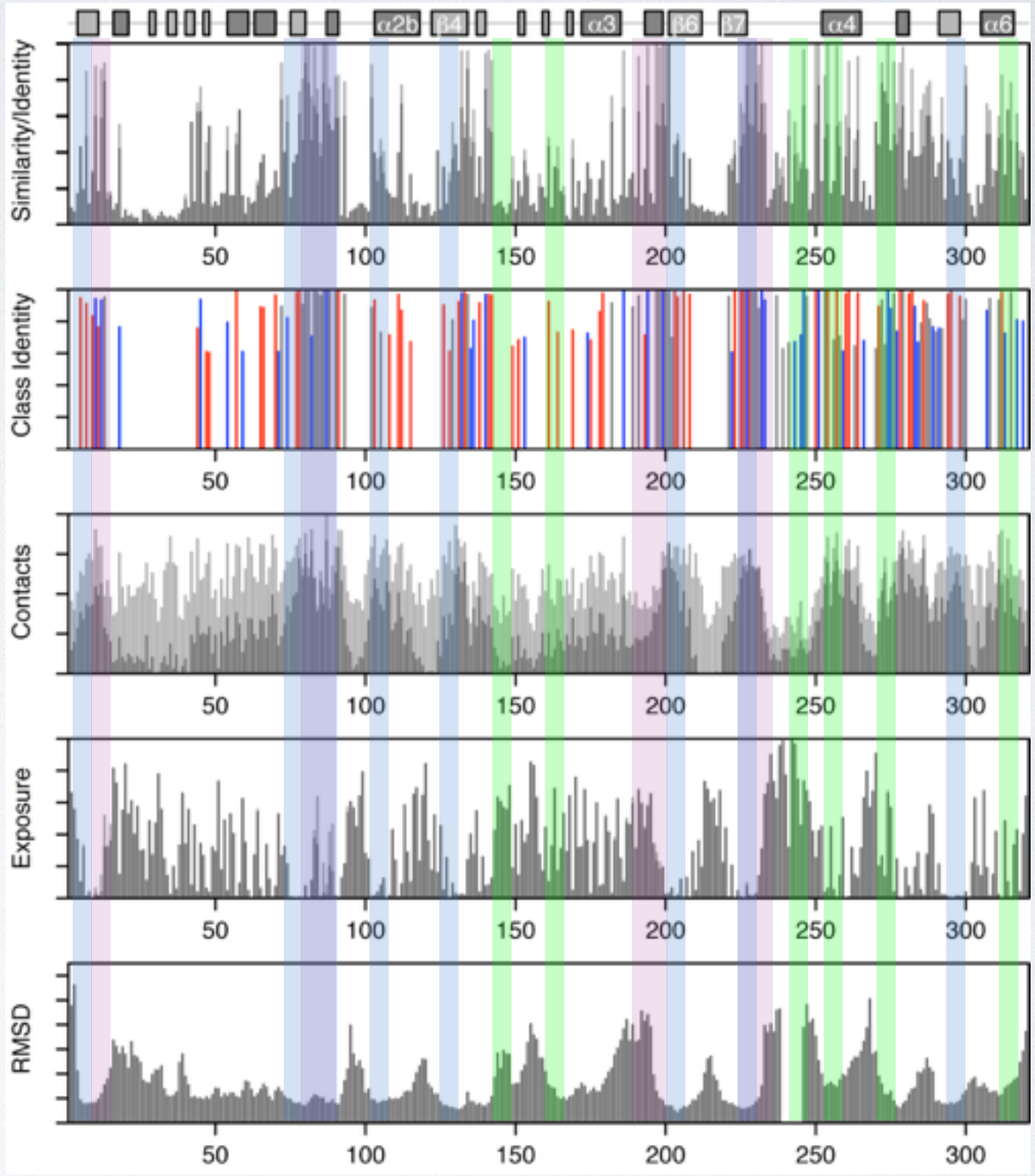
- Structure allows identification of more distant evolutionary relationships

Structure is encoded in sequence

- Understanding the determinants of structure allows design and manipulation of proteins for industrial and medical advantage

Goals:

- Analysis
- Visualization
- Comparison
- Prediction
- Design



Grant *et al.* JMB. (2007)

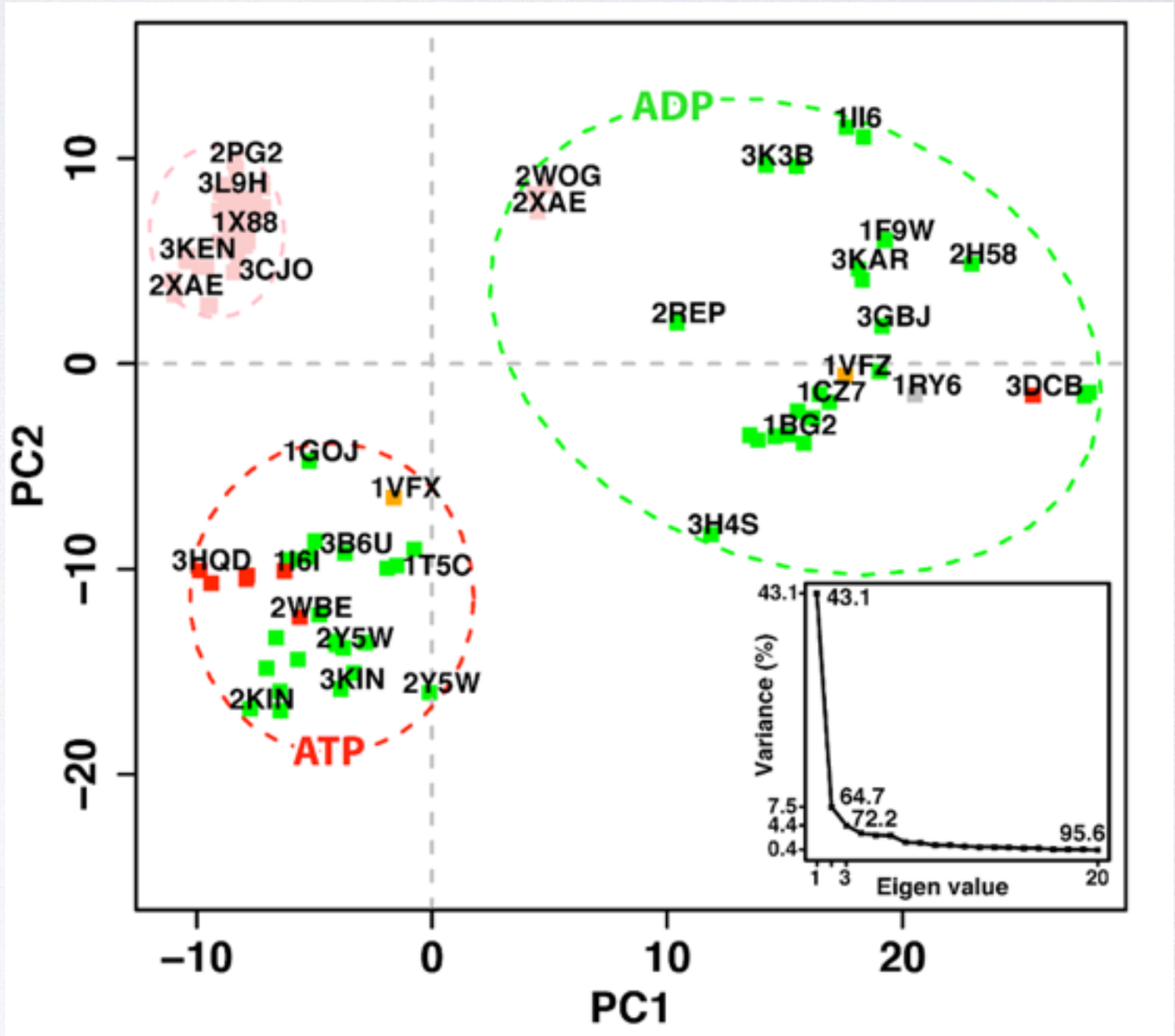
Goals:

- Analysis
- Visualization
- Comparison
- Prediction
- Design

Scarabelli and Grant. PLoS. Comp. Biol. (2013)

Goals:

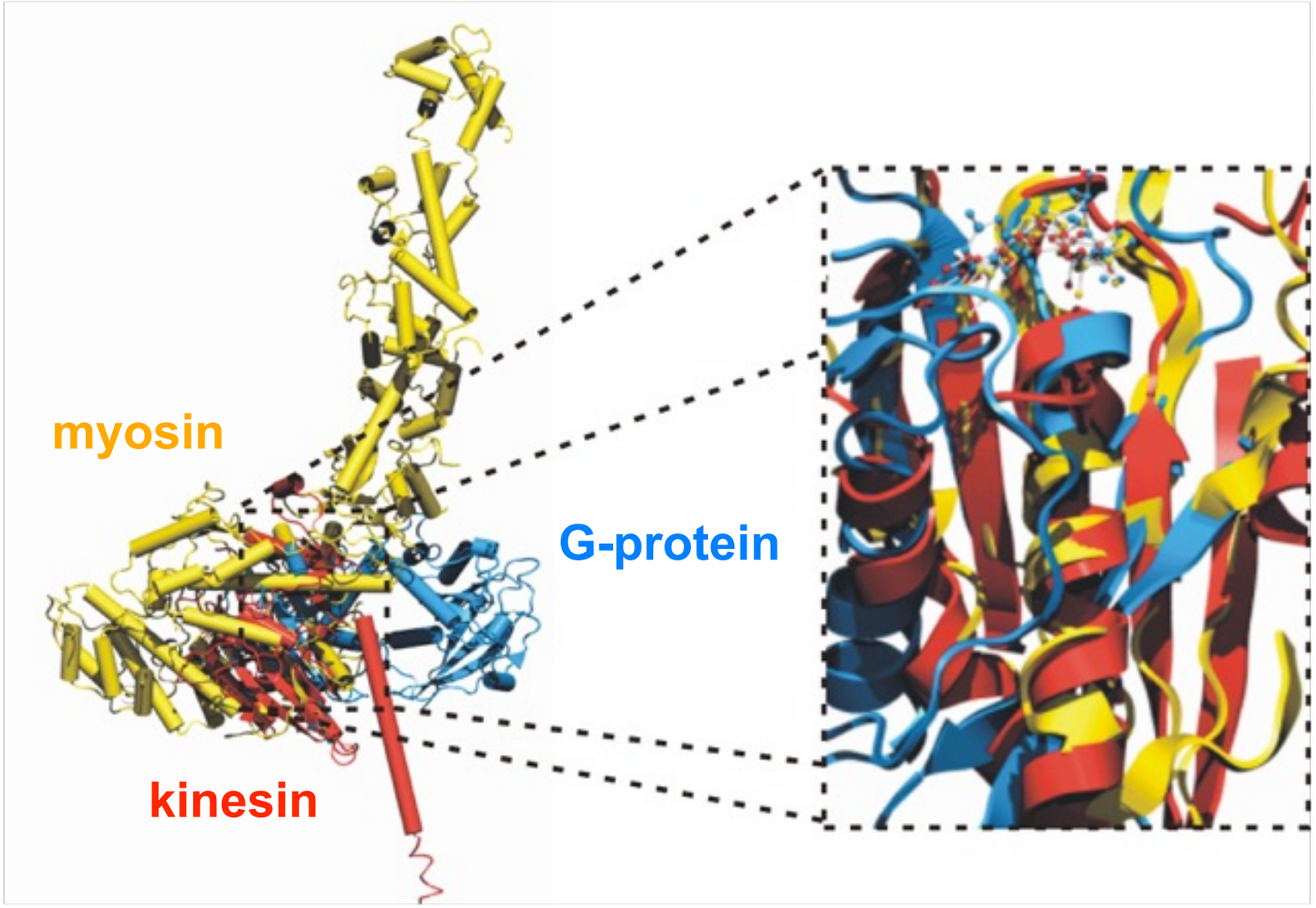
- Analysis
- Visualization
- Comparison
- Prediction
- Design



Scarabelli and Grant. PLoS. Comp. Biol. (2013)

Goals:

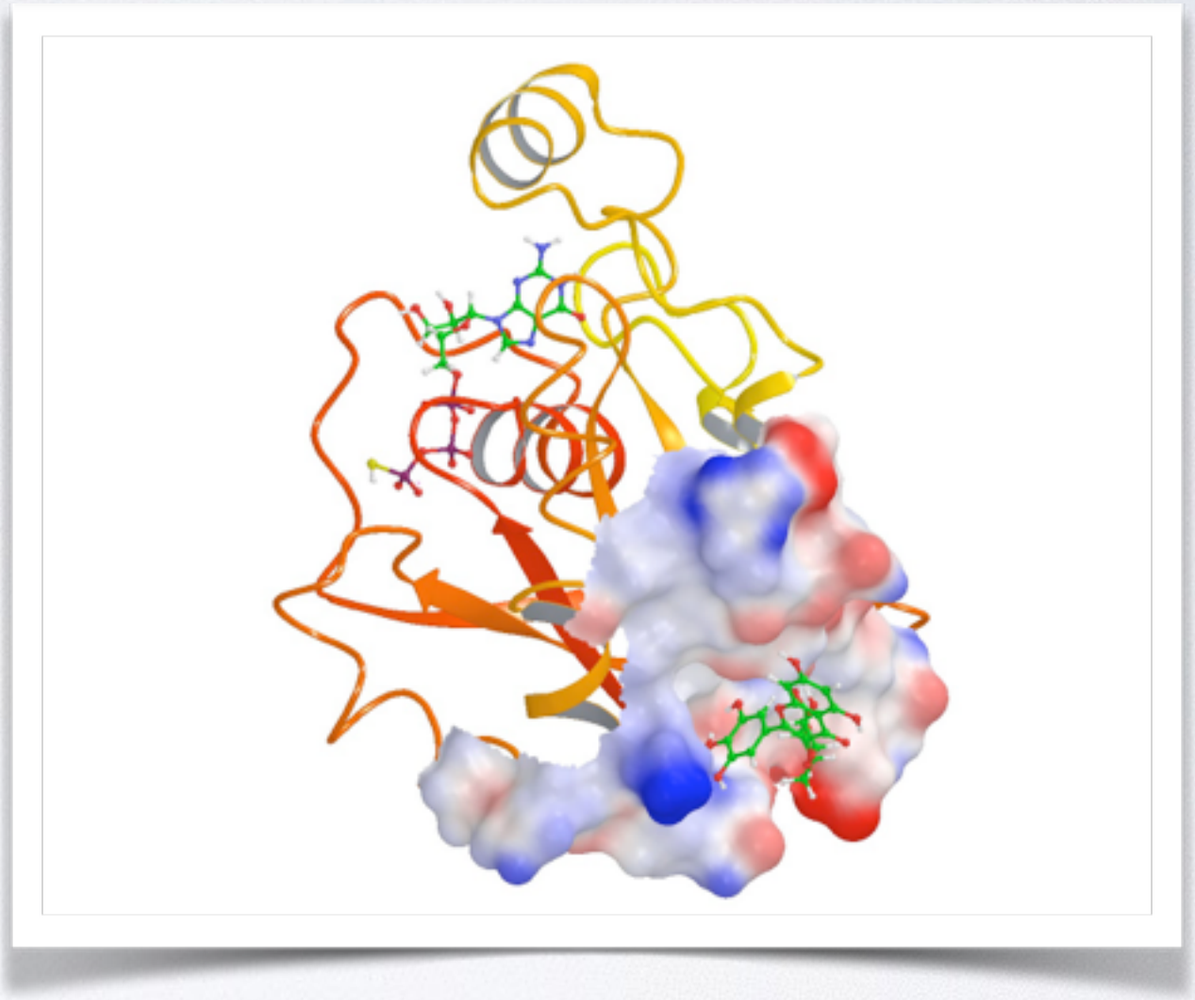
- Analysis
- Visualization
- Comparison
- Prediction
- Design



Grant *et al.* unpublished

Goals:

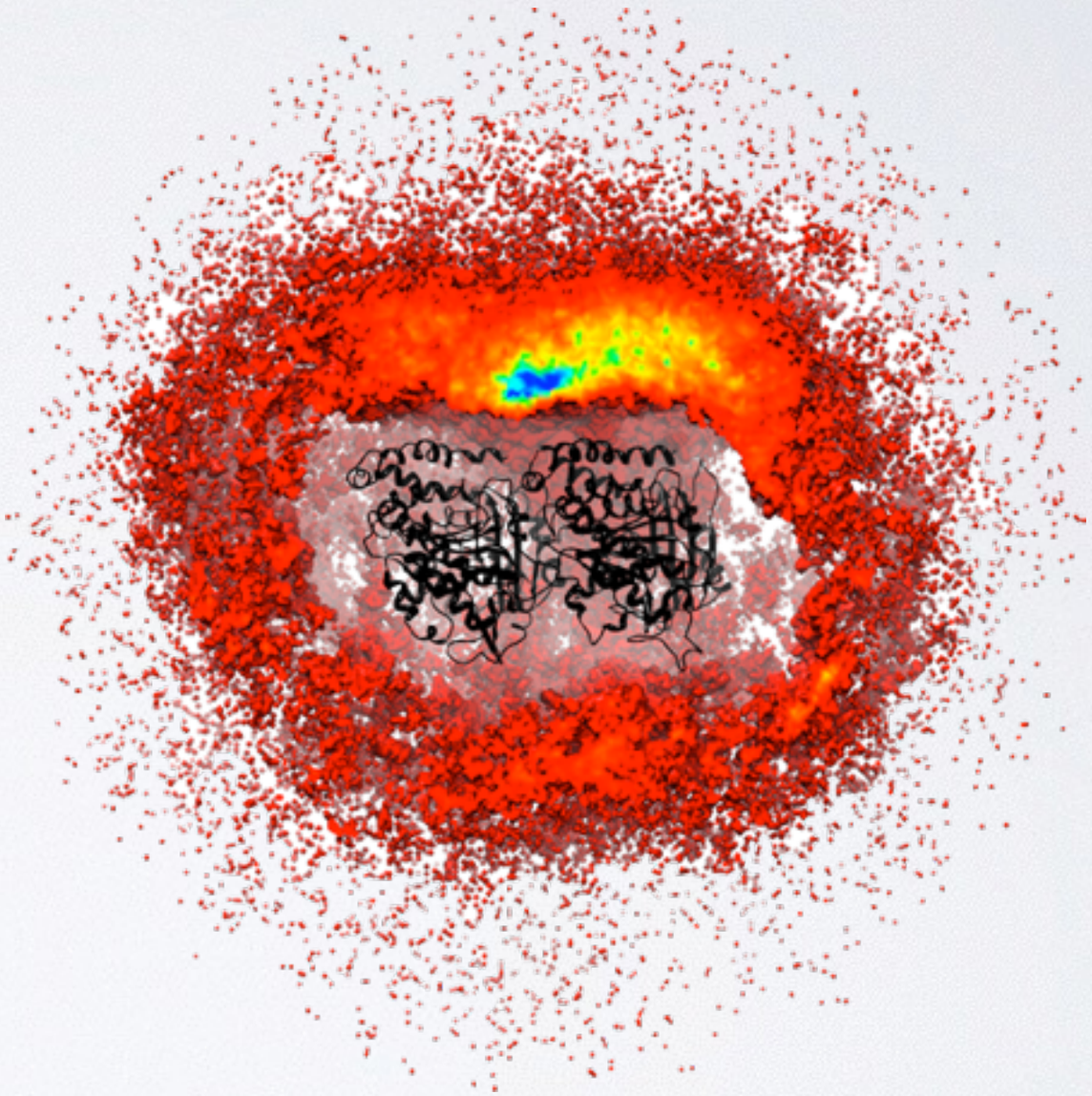
- Analysis
- Visualization
- Comparison
- Prediction
- Design



Grant *et al.* PLoS One (2011, 2012)

Goals:

- Analysis
- Visualization
- Comparison
- Prediction
- Design



Grant *et al.* PLoS Biology (2011)

MAJOR RESEARCH AREAS AND CHALLENGES

Include but are not limited to:

- Protein classification
- Structure prediction from sequence
- Binding site detection
- Binding prediction and drug design
- Modeling molecular motions
- Predicting physical properties (stability, binding affinities)
- Design of structure and function
- etc...

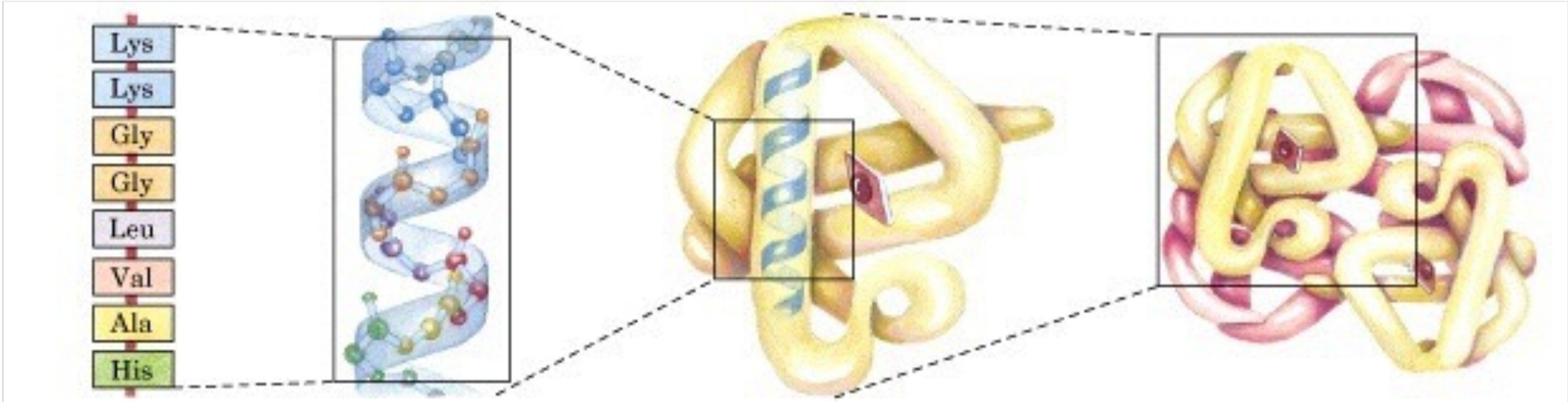
With applications to Biology, Medicine, Agriculture and Industry

NEXT UP:

- **Overview of structural bioinformatics**
  - Major motivations, goals and challenges
- **Fundamentals of protein structure**
  - Composition, form, forces and dynamics
- **Representing and interpreting protein structure**
  - Modeling energy as a function of structure
- **Example application areas**
  - Predicting functional dynamics & drug discovery

HIERARCHICAL STRUCTURE OF PROTEINS

Primary > Secondary > Tertiary > Quaternary



amino acid residues

Alpha helix

Polypeptide chain

Assembled subunits

Image from: http://www.ncbi.nlm.nih.gov/books/NBK21581/

RECAP: AMINO ACID NOMENCLATURE

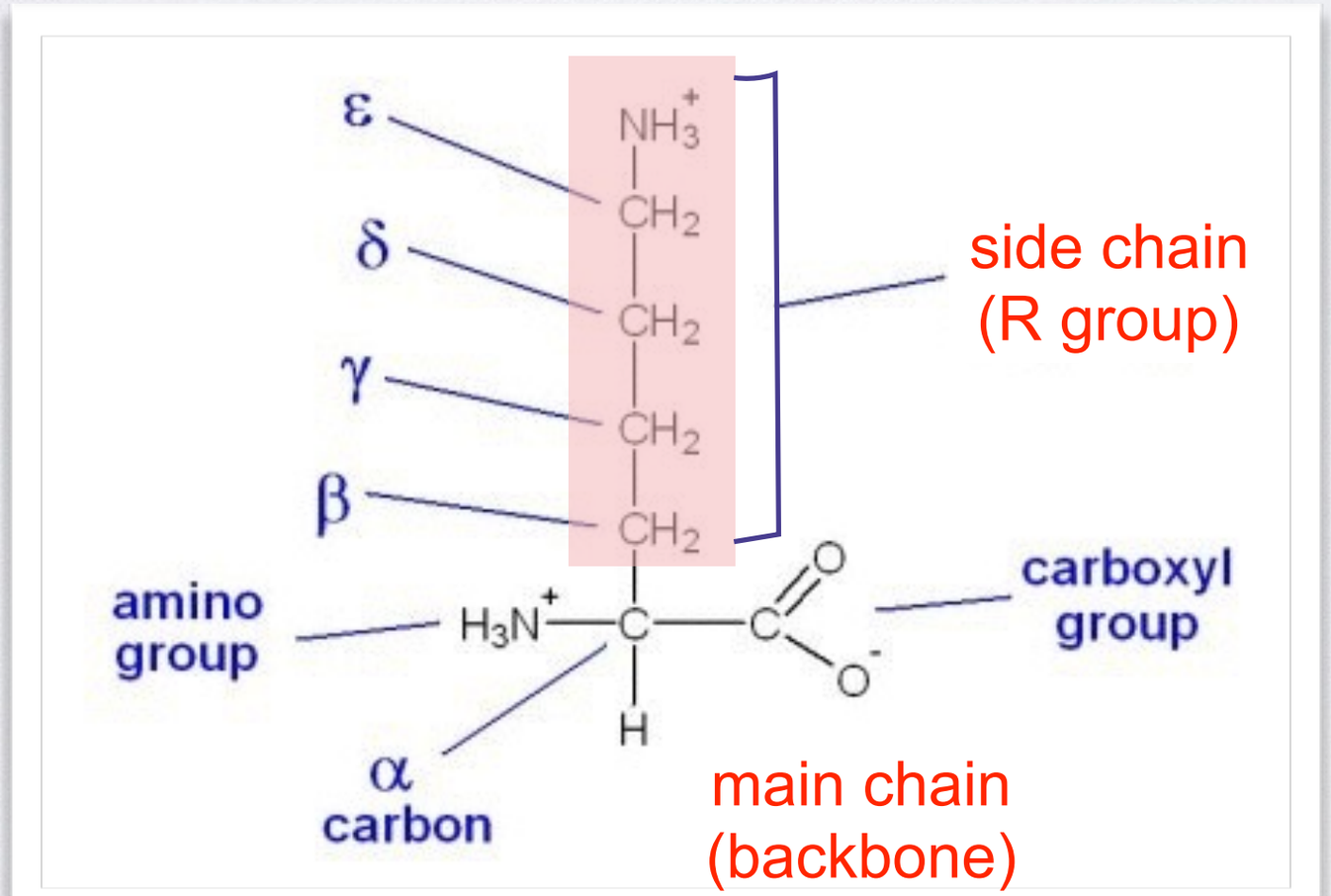
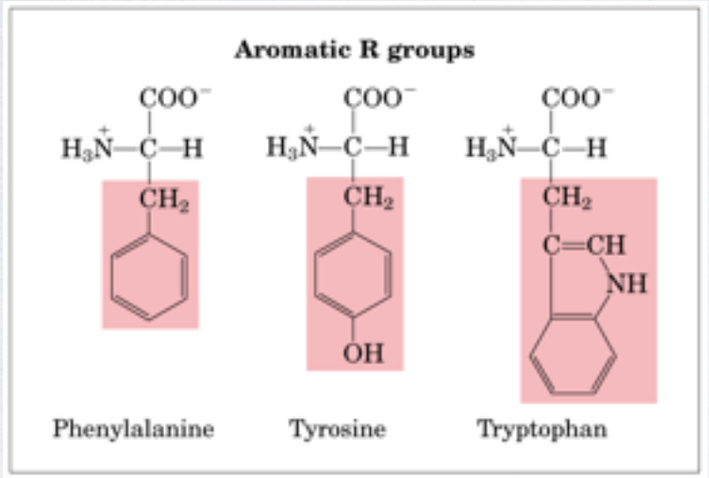
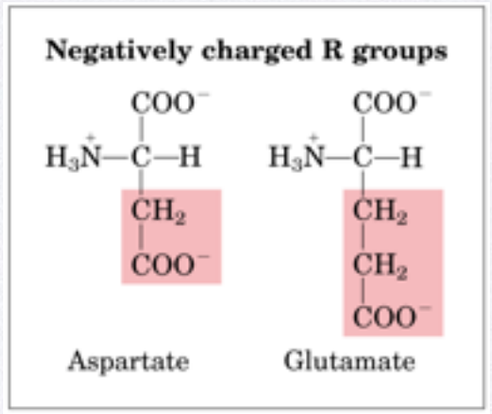
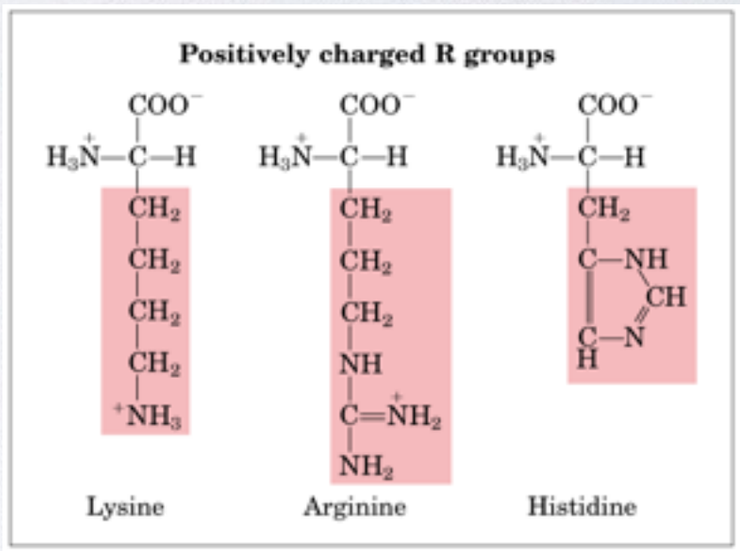


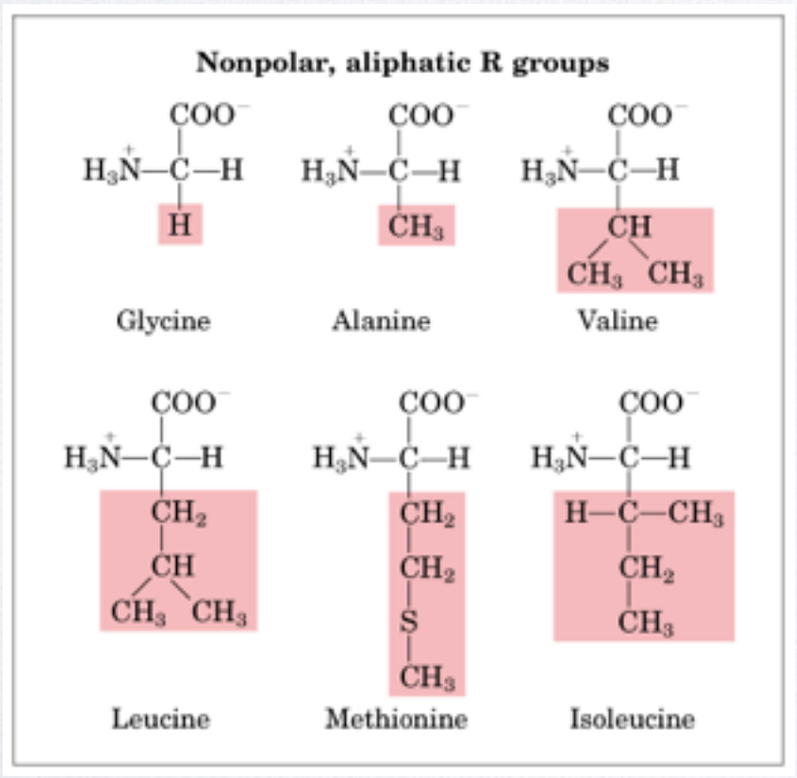
Image from: http://www.ncbi.nlm.nih.gov/books/NBK21581/

AMINO ACIDS CAN BE GROUPED BY THE **PHYSIOCHEMICAL PROPERTIES**









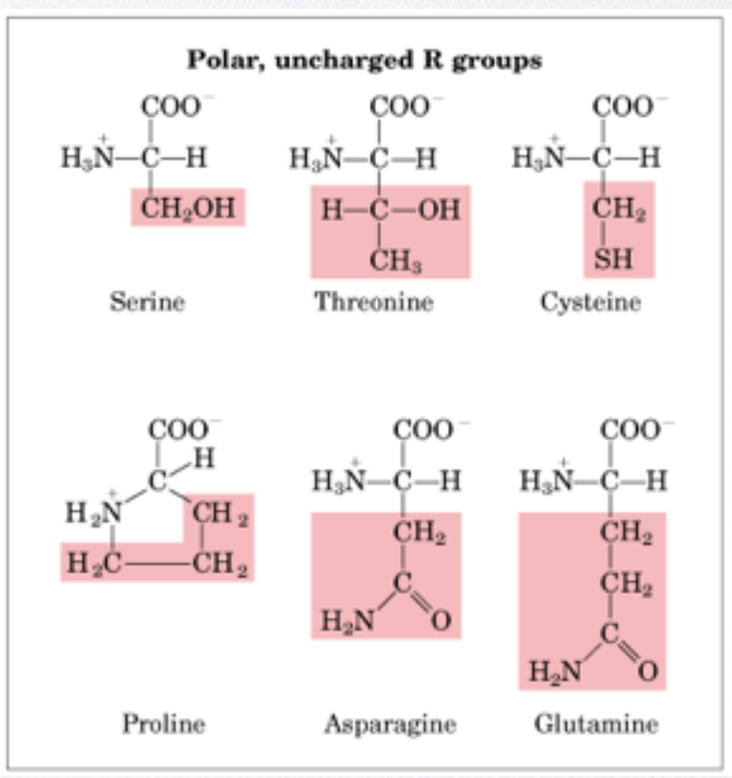
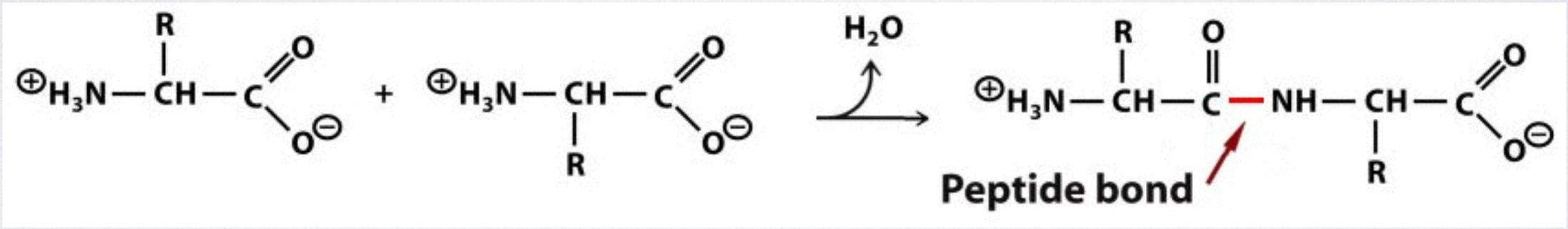


Image from: http://www.ncbi.nlm.nih.gov/books/NBK21581/

AMINO ACIDS POLYMERIZE THROUGH **PEPTIDE BOND** FORMATION



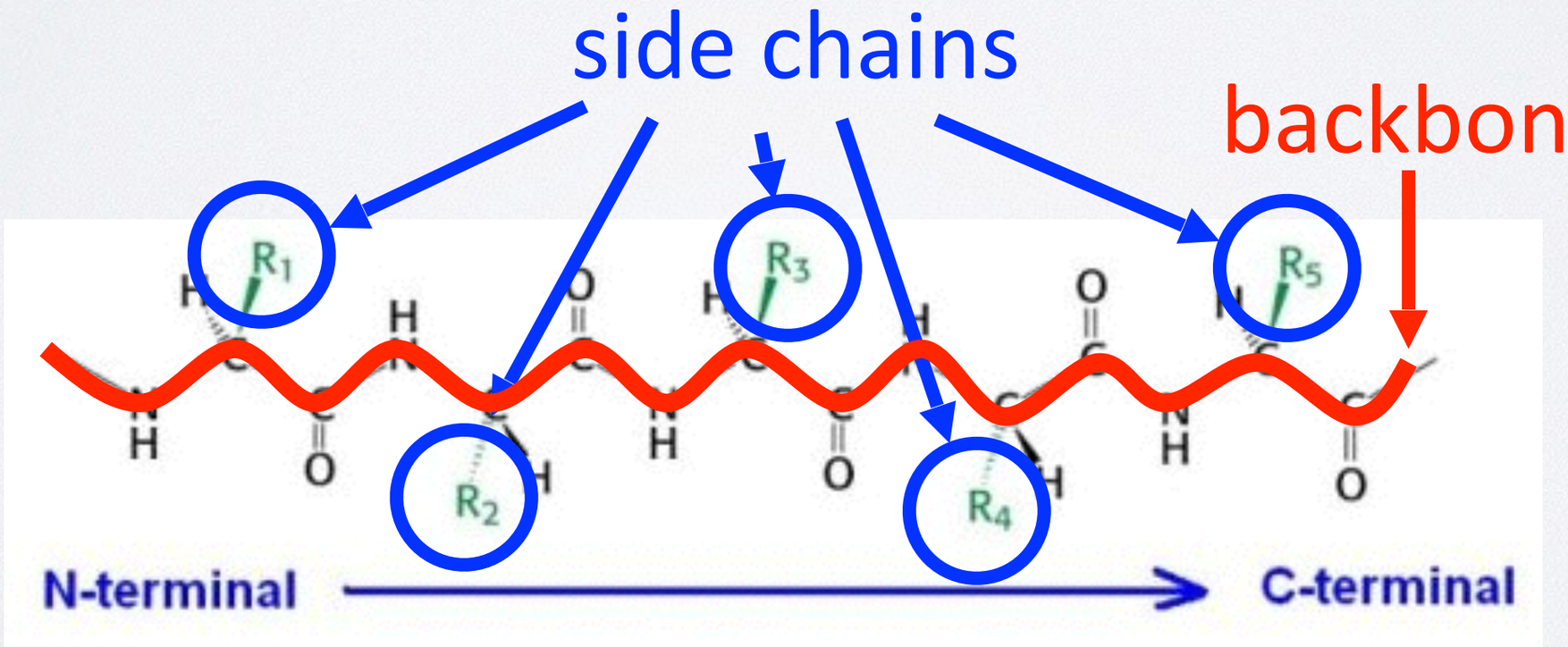
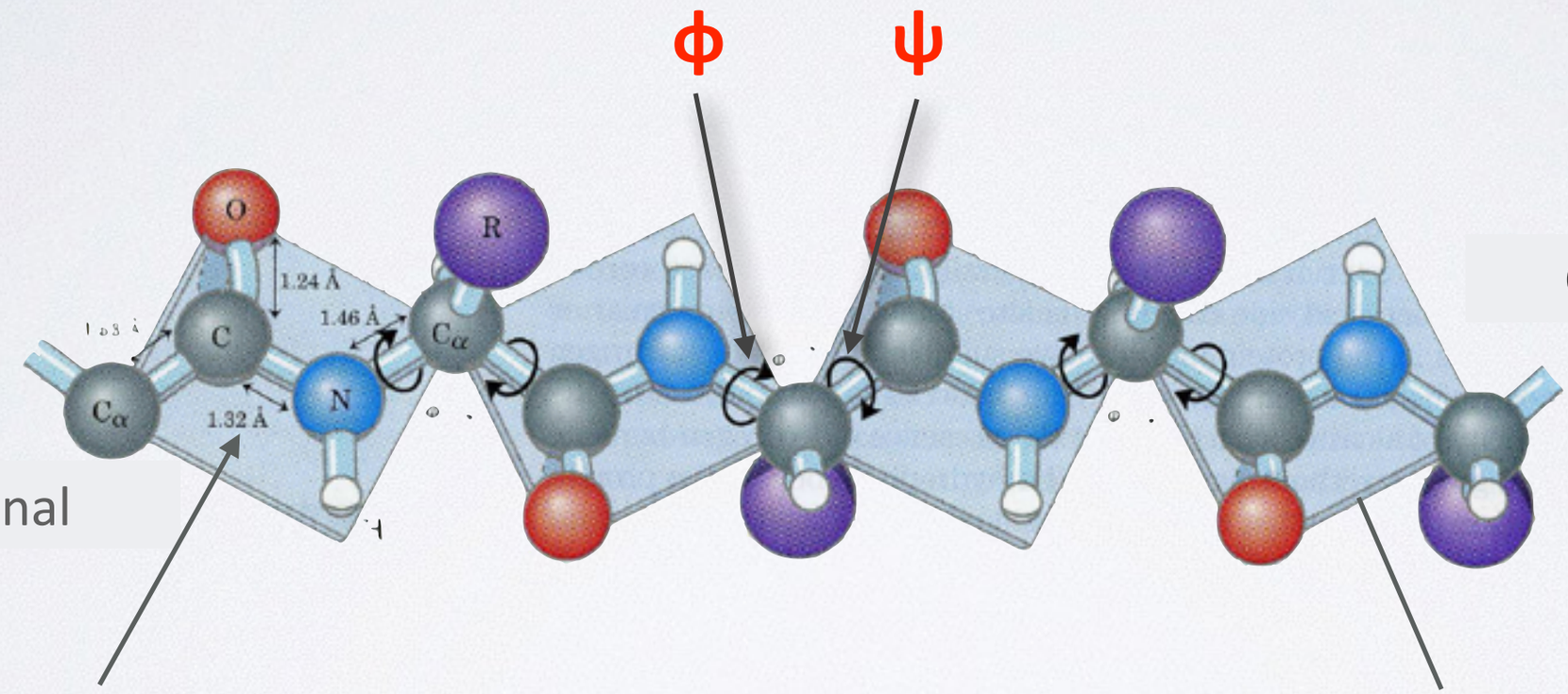


Image from: http://www.ncbi.nlm.nih.gov/books/NBK21581/

PEPTIDES CAN ADOPT DIFFERENT CONFORMATIONS BY VARYING THEIR **PHI & PSI BACKBONE TORSIONS**

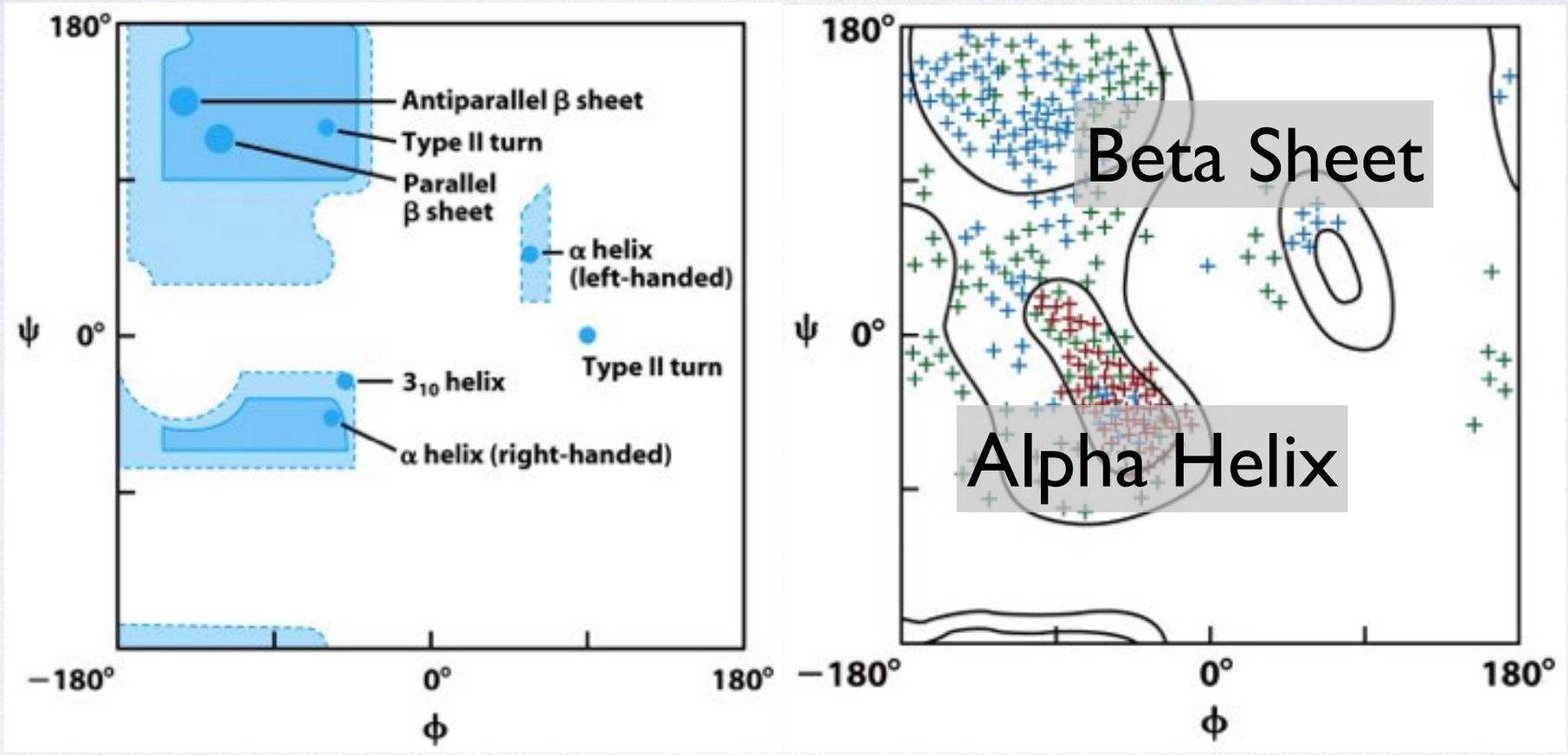


Bond angles and lengths are largely invariant

Peptide bond is planer (Cα, C, O, N, H, Cα all lie in the same plane)

Image from: http://www.ncbi.nlm.nih.gov/books/NBK21581/

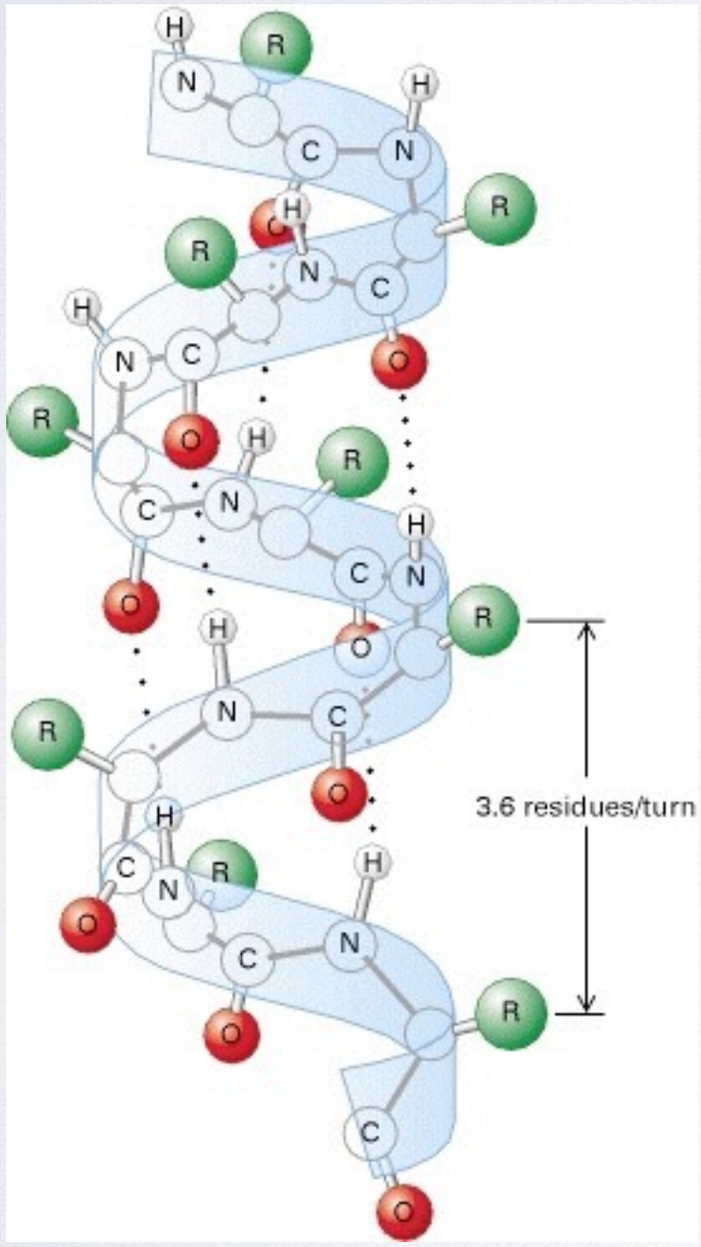
PHI vs PSI PLOTS ARE KNOWN AS **RAMACHANDRAN DIAGRAMS**



- Steric hindrance dictates torsion angle preference
- Ramachandran plot show preferred regions of ϕ and ψ dihedral angles which correspond to major forms of **secondary structure**

Image from: http://www.ncbi.nlm.nih.gov/books/NBK21581/

MAJOR SECONDARY STRUCTURE TYPES
ALPHA HELIX & BETA SHEET



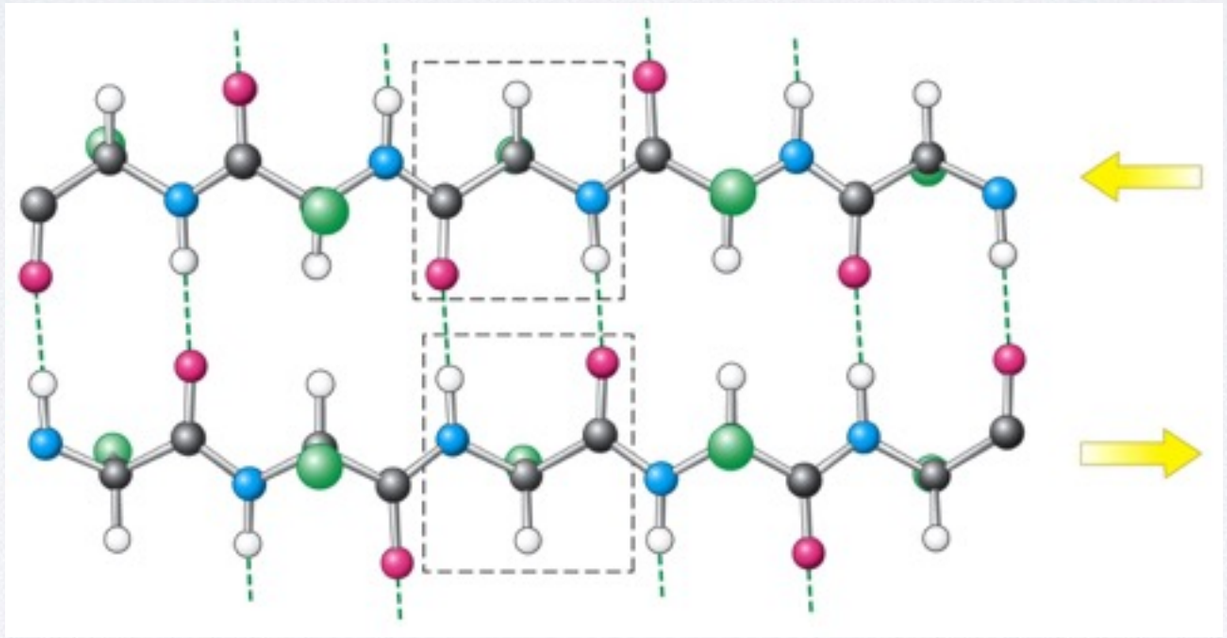
α-helix

- Most common from has <u>3.6 residues per turn</u> (number of residues in one full rotation)
- Hydrogen bonds (dashed lines) between residue <u>*i* and *i+4*</u> stabilize the structure
- The side chains (in green) protrude outward
- 3_{10}-helix and π-helix forms are less common

Hydrogen bond: **i→i+4**

Image from: http://www.ncbi.nlm.nih.gov/books/NBK21581/

MAJOR SECONDARY STRUCTURE TYPES
ALPHA HELIX & **BETA SHEET**

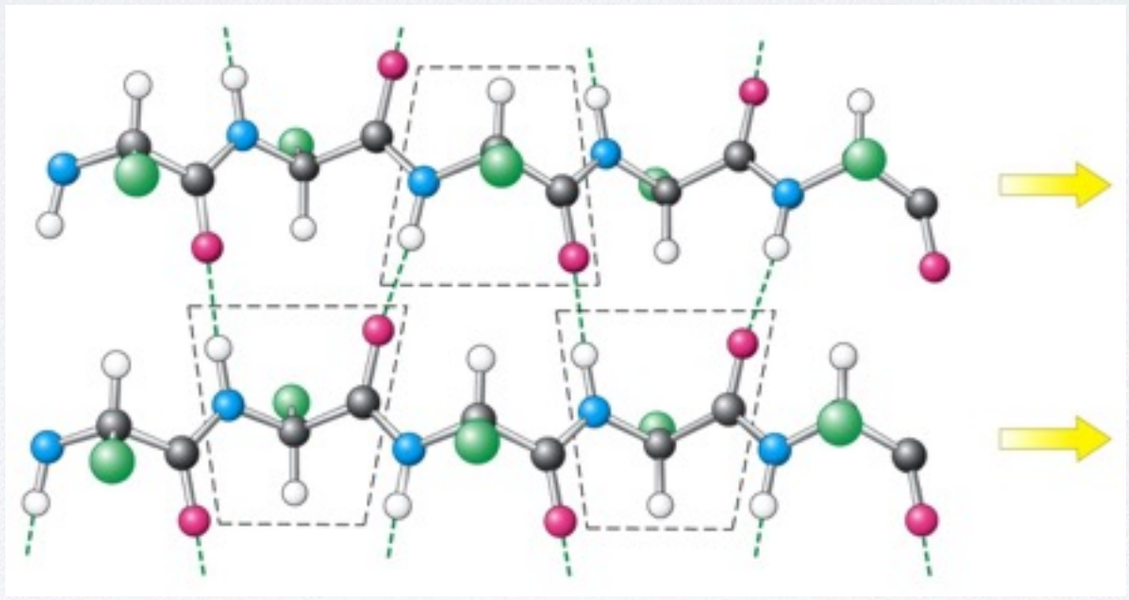


In **<u>antiparallel</u> β-sheets**

- Adjacent β-strands run in <u>opposite</u> directions
- Hydrogen bonds (dashed lines) between NH and CO stabilize the structure
- The side chains (in green) are above and below the sheet

Image from: <u>http://www.ncbi.nlm.nih.gov/books/NBK21581/</u>

MAJOR SECONDARY STRUCTURE TYPES
ALPHA HELIX & **BETA SHEET**

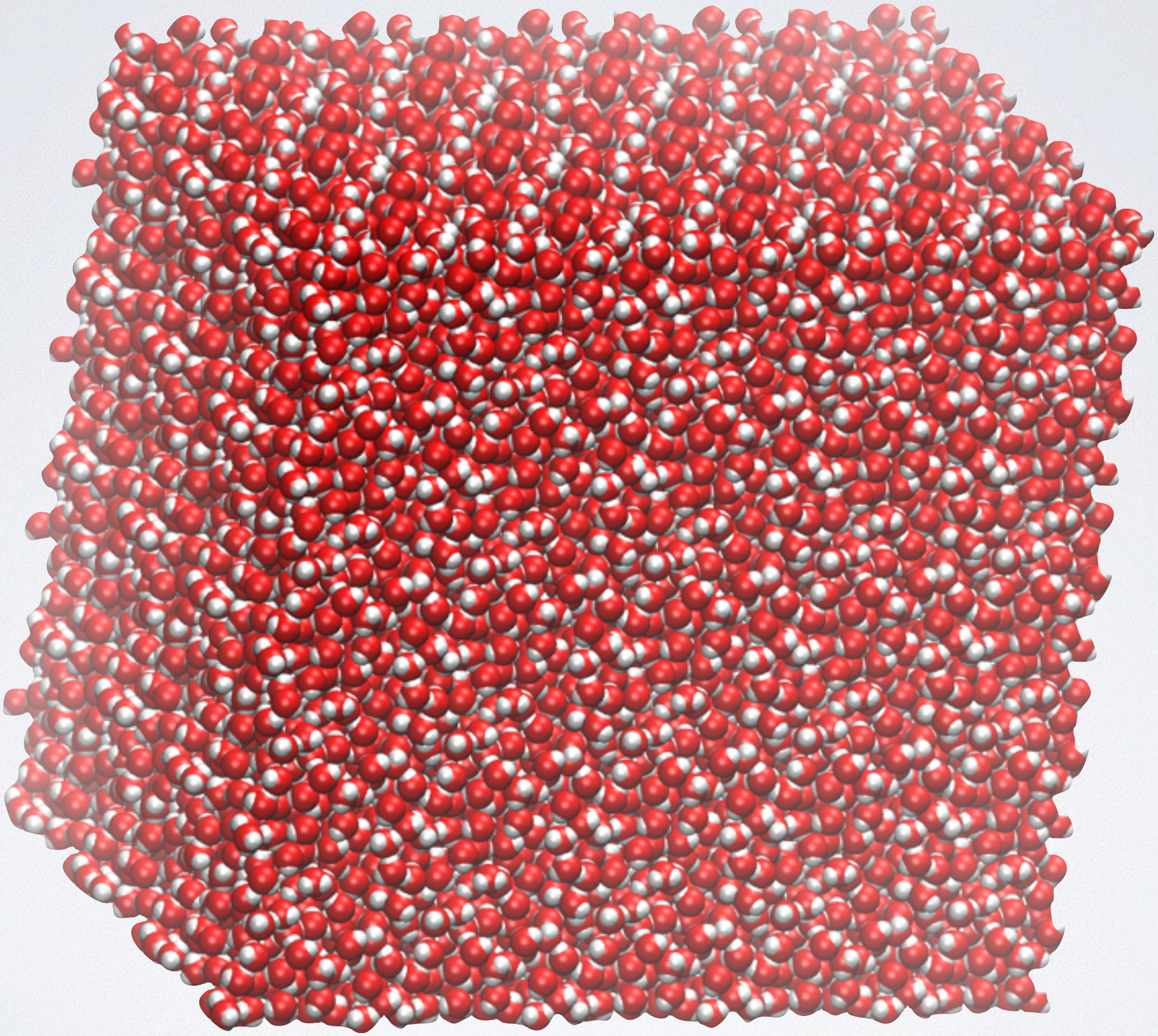


In **<u>parallel</u> β-sheets**

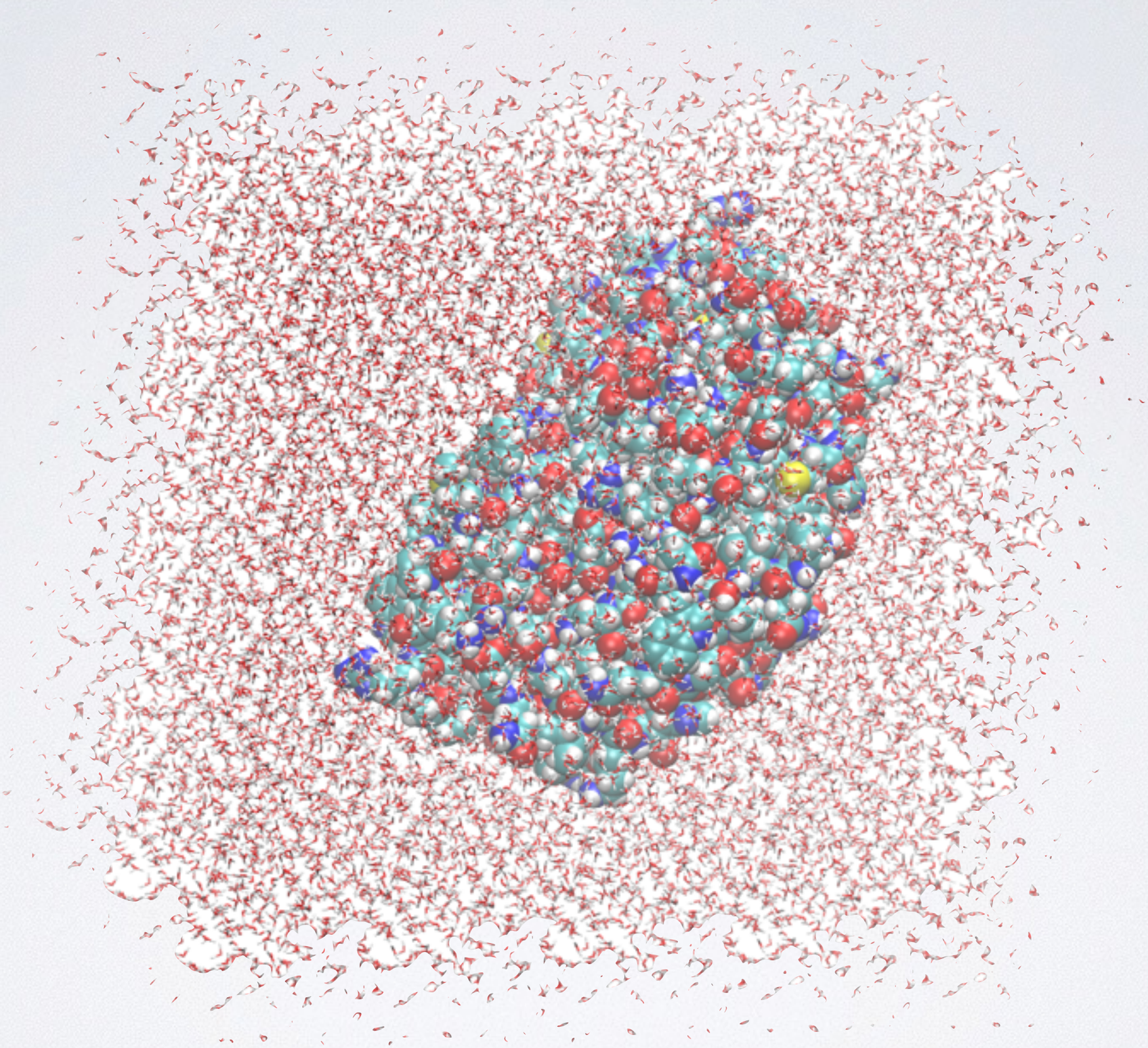
- Adjacent β-strands run in <u>same</u> direction
- Hydrogen bonds (dashed lines) between NH and CO stabilize the structure
- The side chains (in green) are above and below the sheet

Image from: <u>http://www.ncbi.nlm.nih.gov/books/NBK21581/</u>

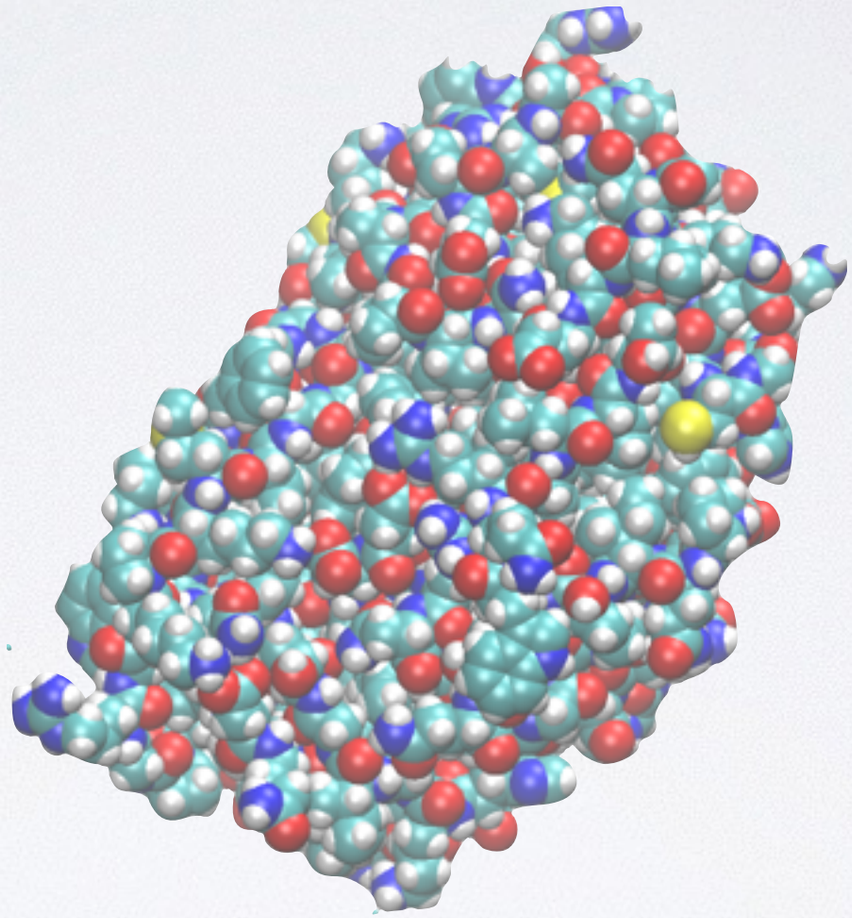
What Does a Protein Look like?



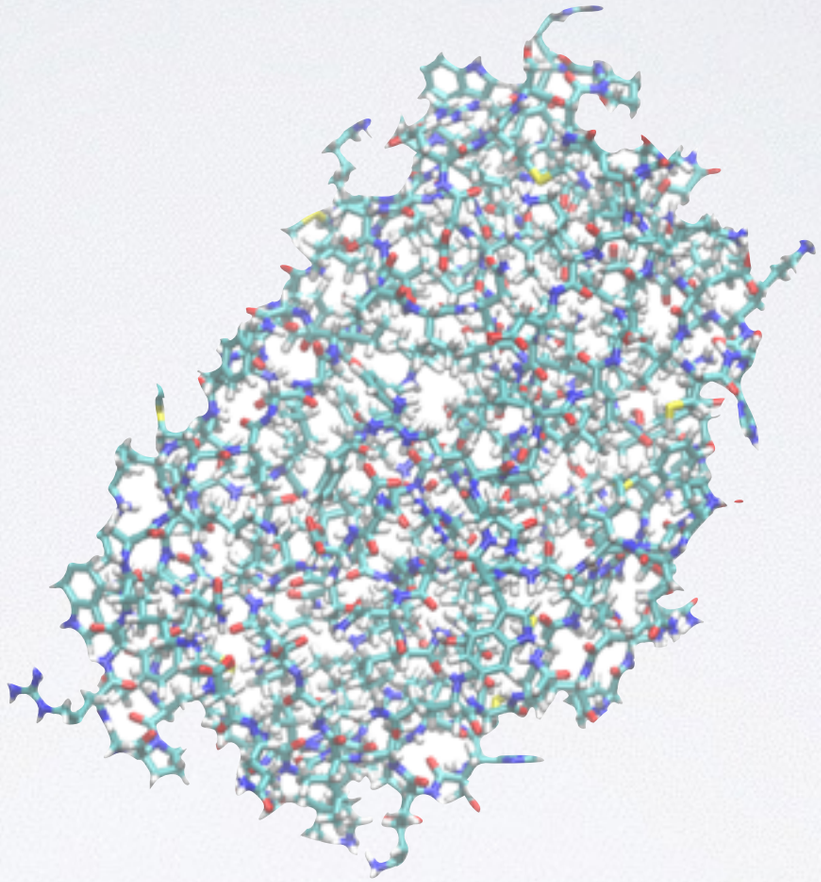
- Proteins are stable (and hidden) in water



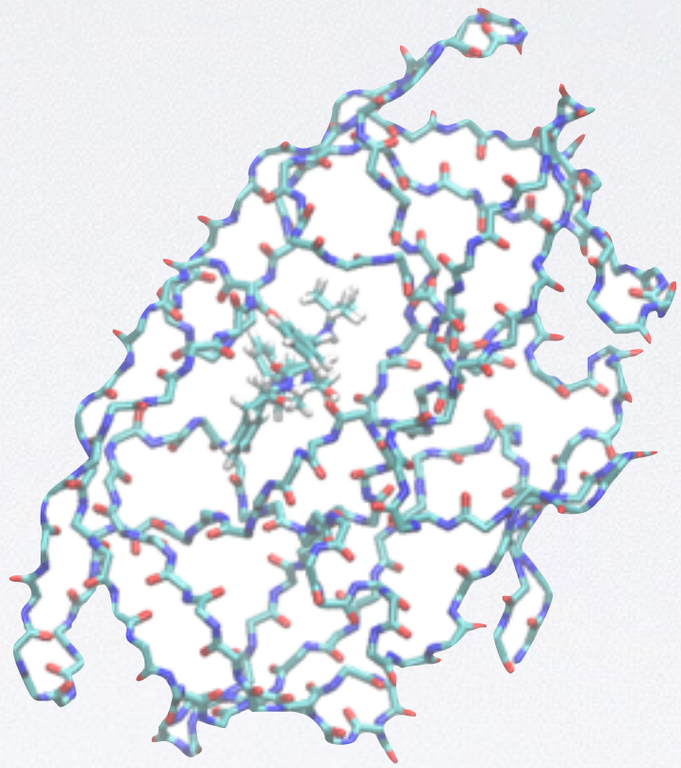
- Proteins closely interact with water



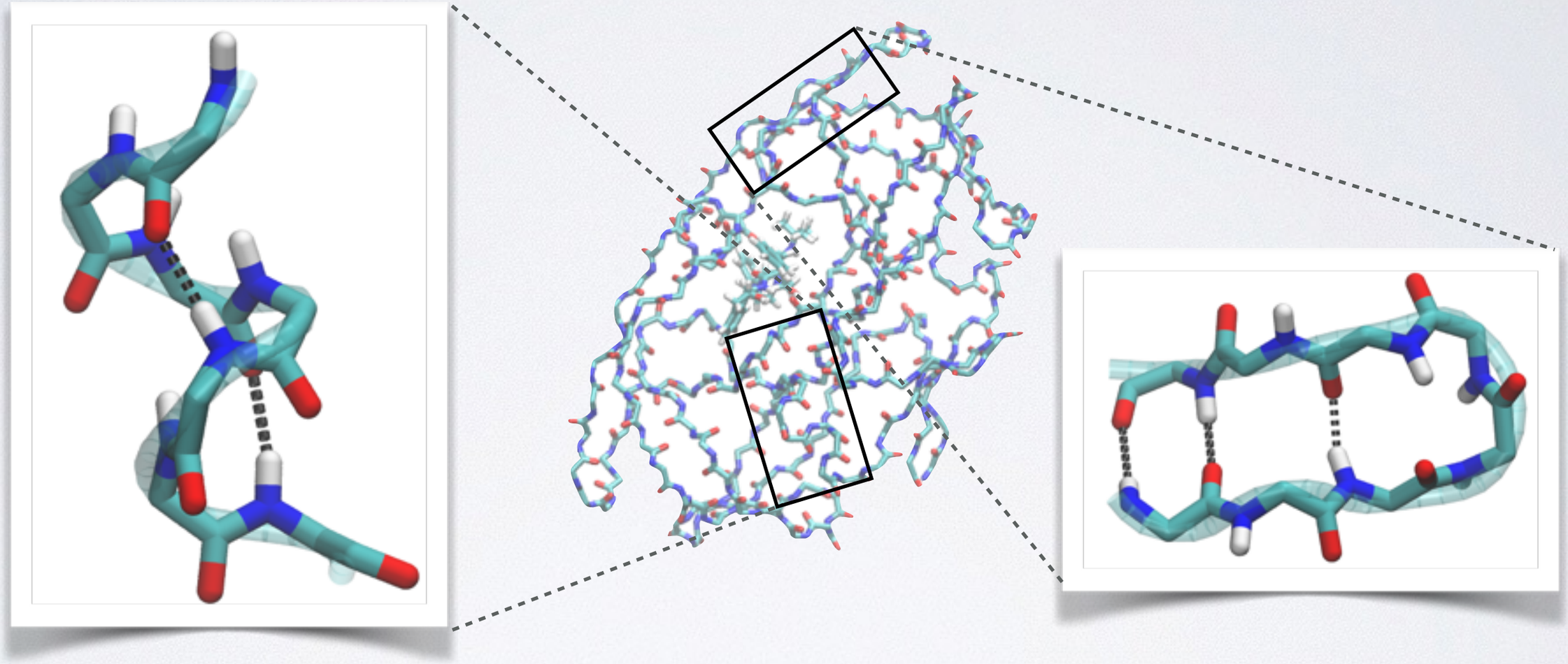
- Proteins are close packed solid but flexible objects (globular)



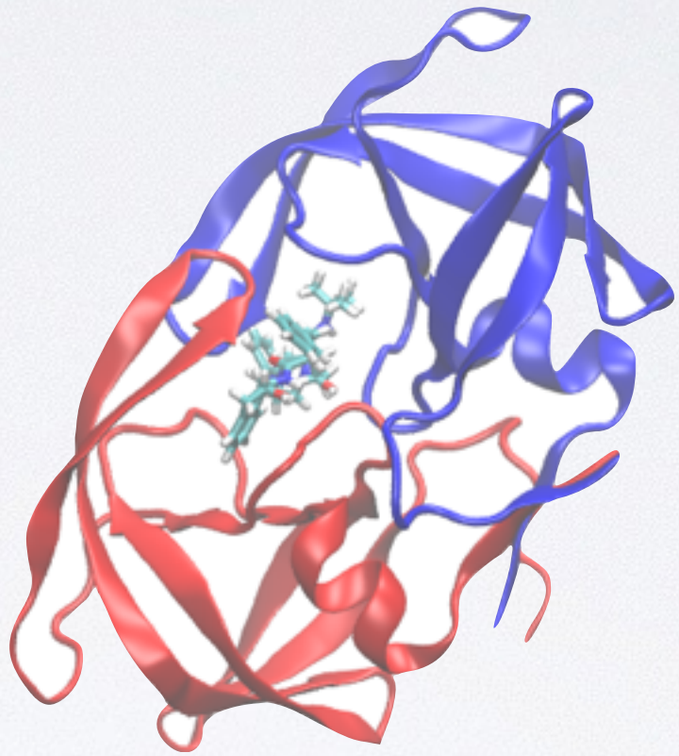
- Due to their large size and complexity it is often hard to see whats important in the structure



- Backbone or main-chain representation can help trace chain topology

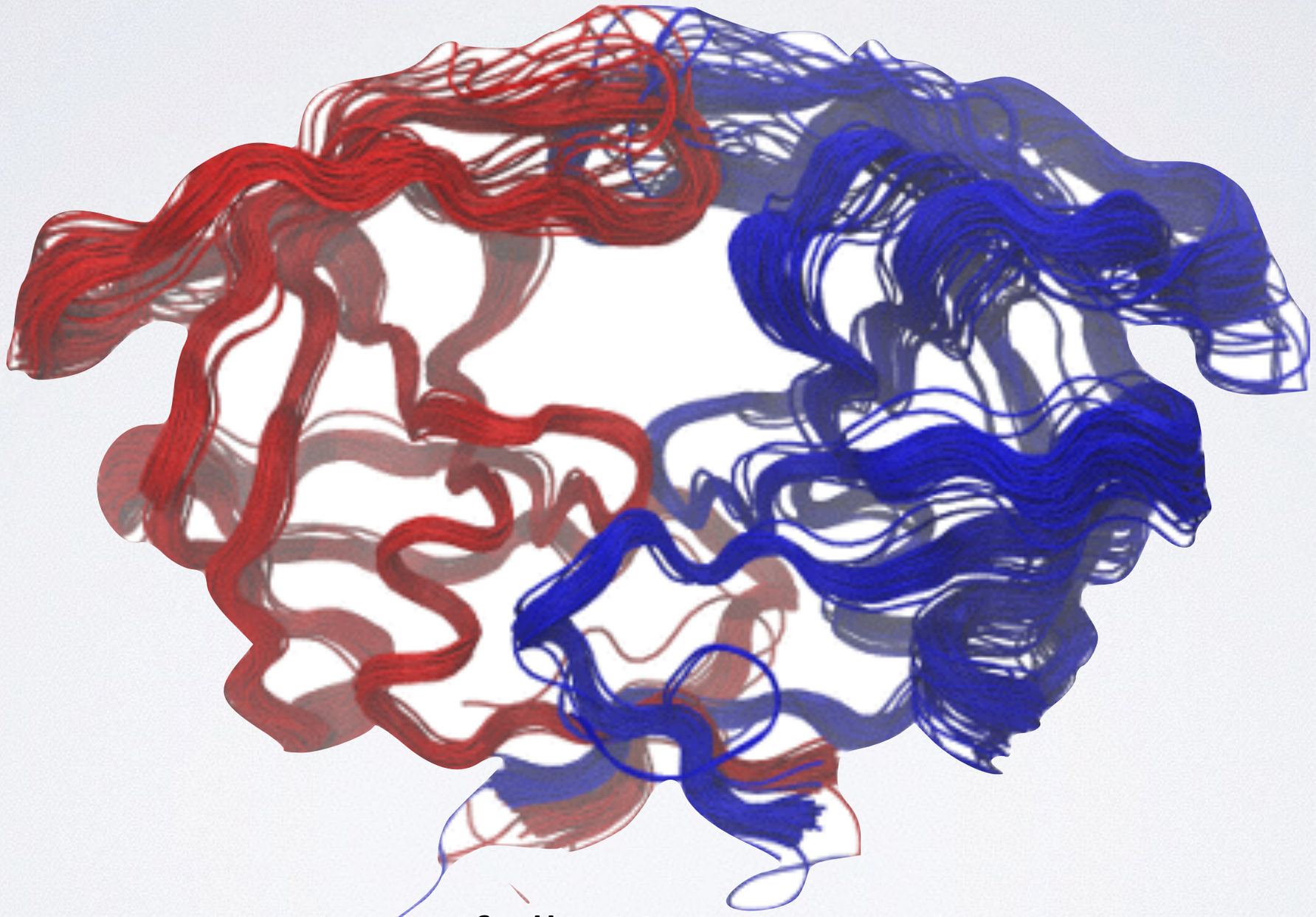


- Backbone or main-chain representation can help trace chain topology & reveal secondary structure



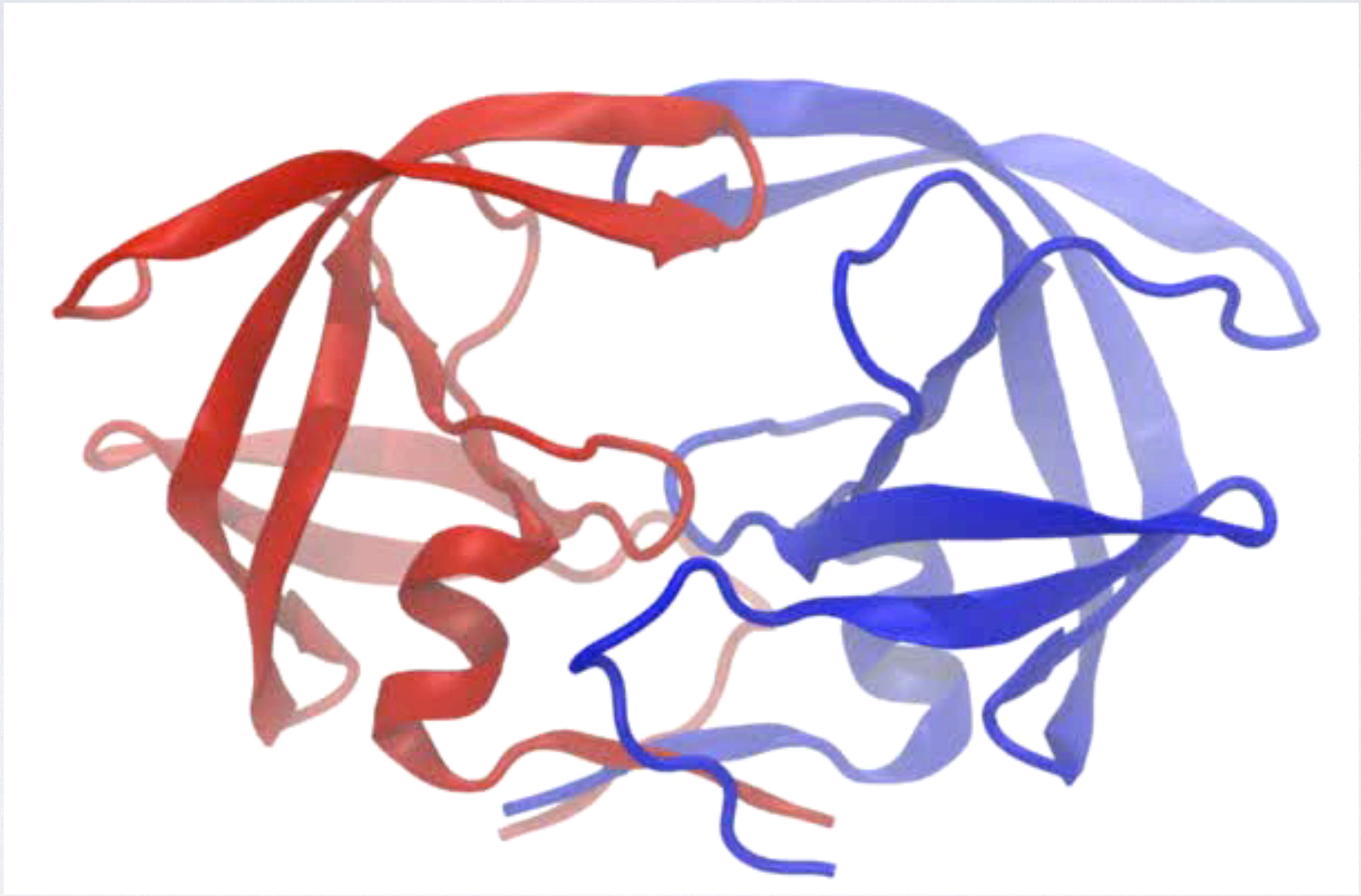
- Simplified secondary structure representations are commonly used to communicate structural details
- Now we can clearly see 2º, 3º and 4º structure
- Coiled chain of connected secondary structures

DISPLACEMENTS REFLECT INTRINSIC FLEXIBILITY



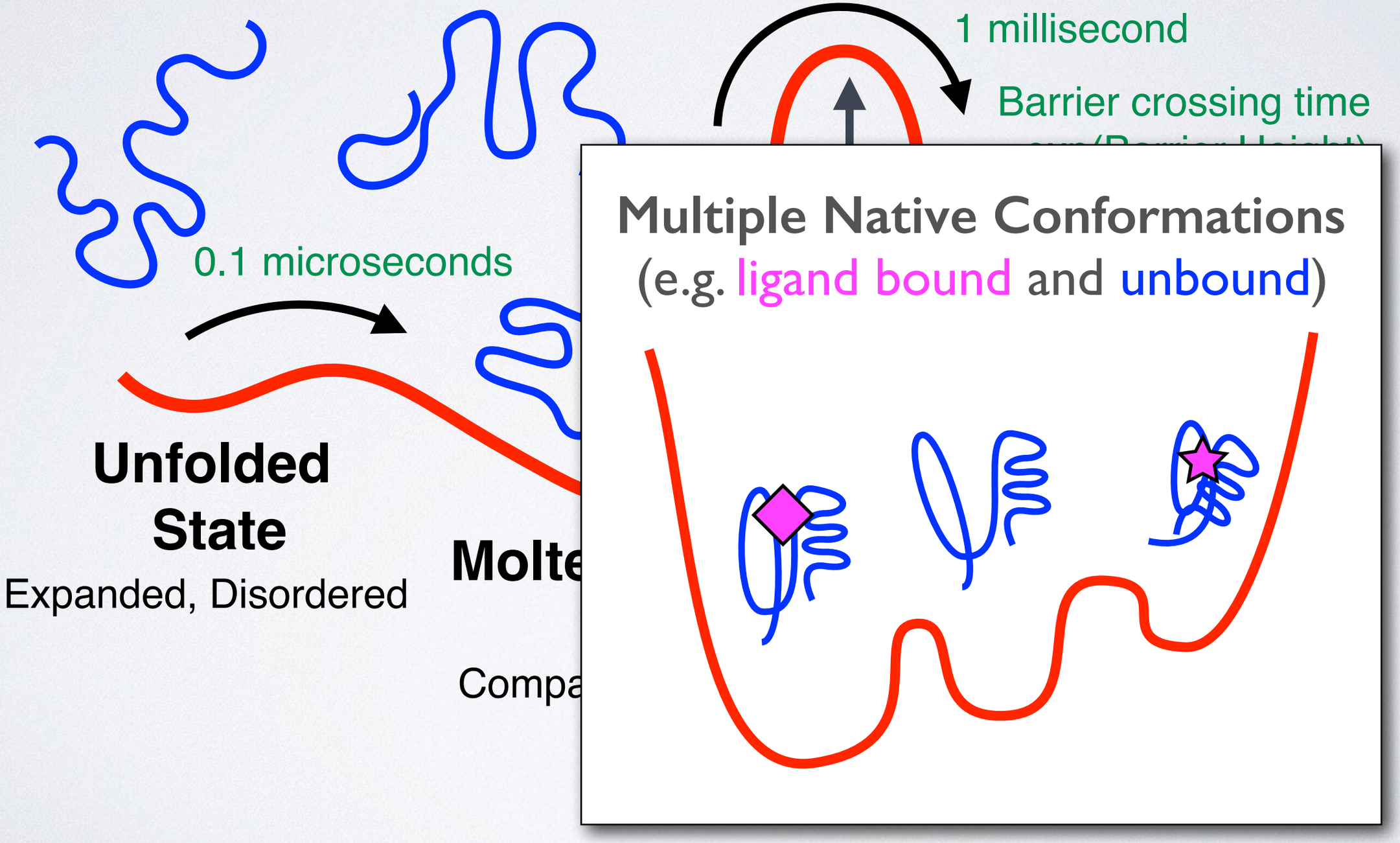
Superposition of all 482 structures in RCSB PDB (23/09/2015)

DISPLACEMENTS REFLECT INTRINSIC FLEXIBILITY



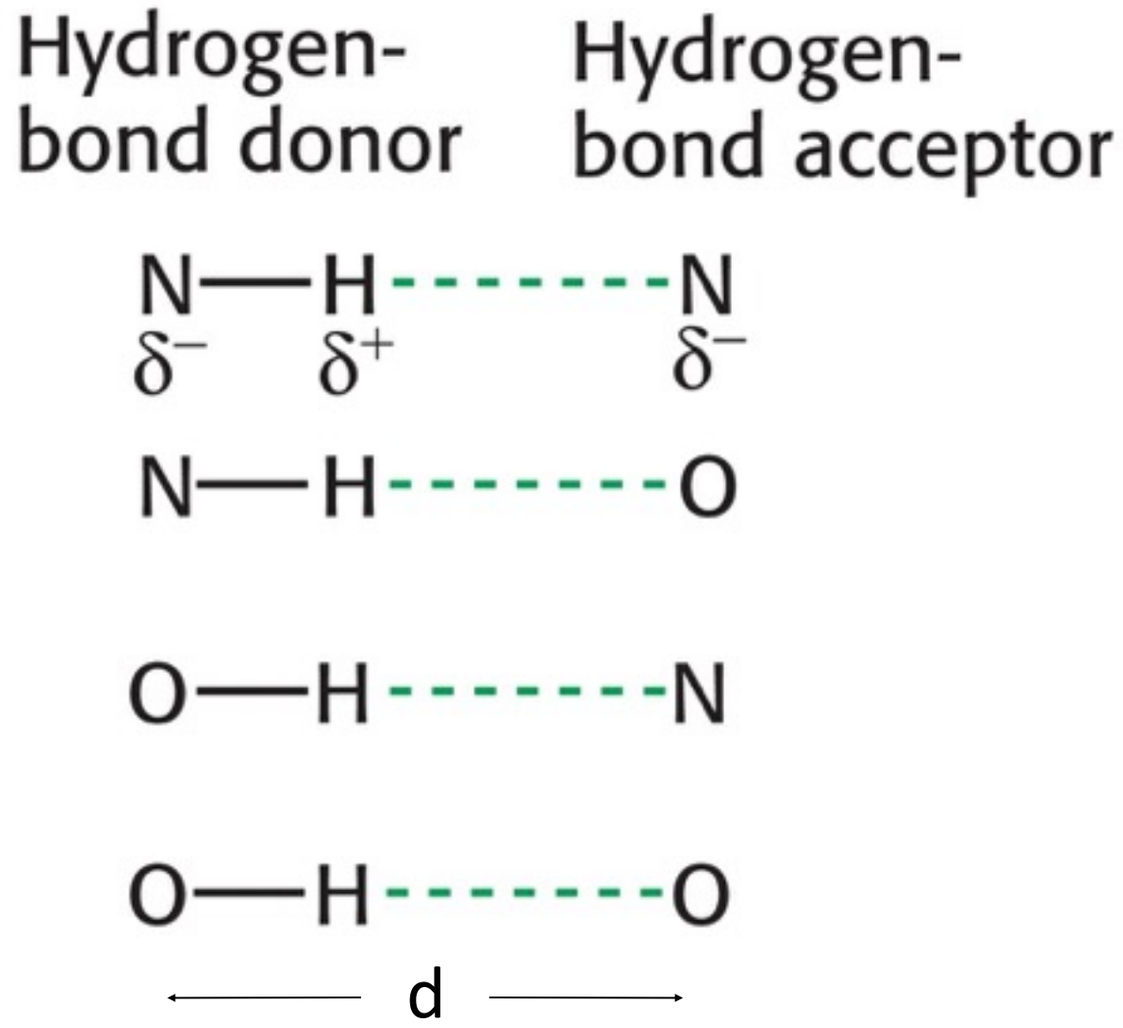
Principal component analysis (PCA) of experimental structures

KEY CONCEPT: **ENERGY LANDSCAPE**



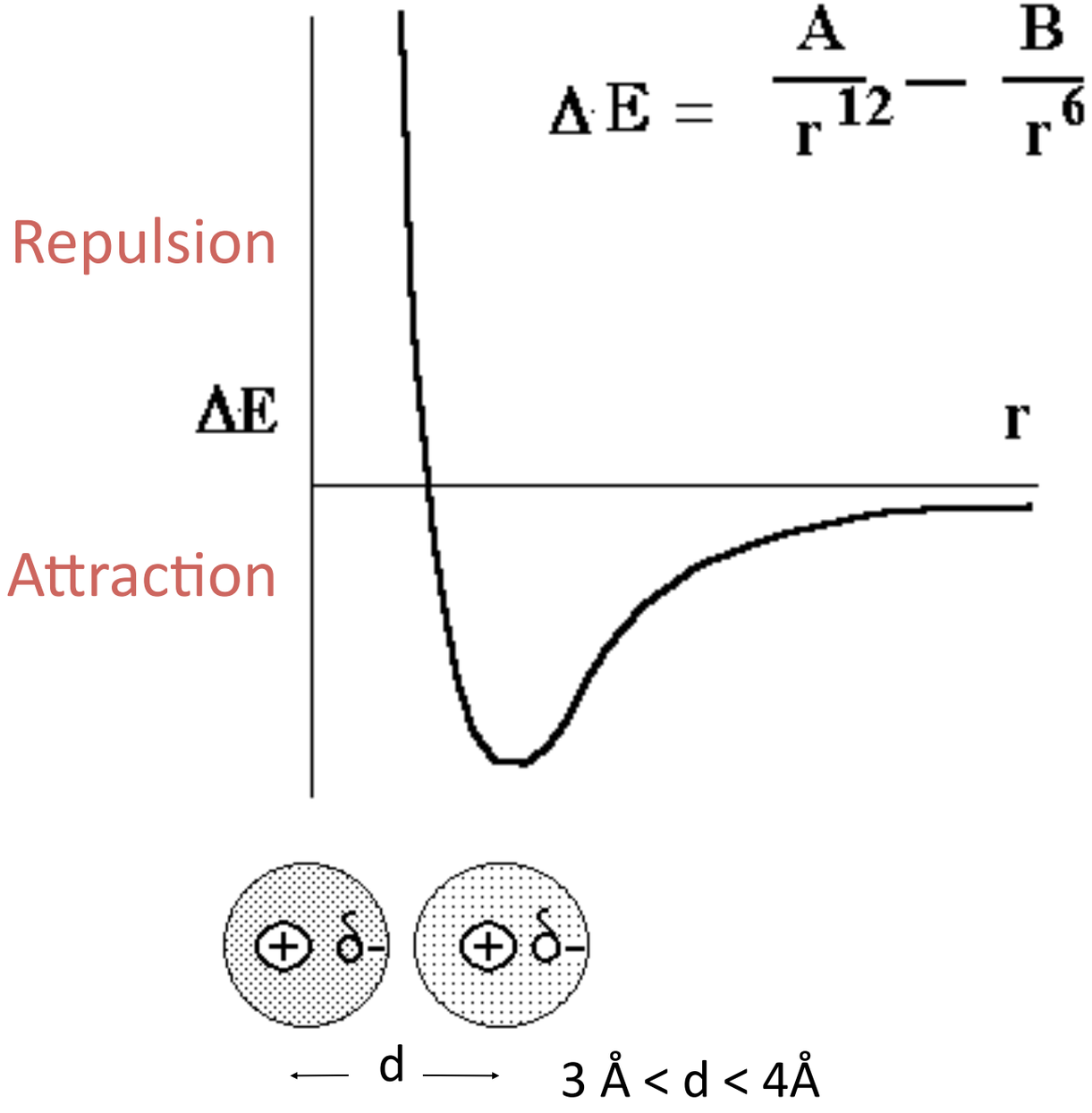
Key forces affecting structure:

- H-bonding
- Van der Waals
- Electrostatics
- Hydrophobicity
- Disulfide Bridges



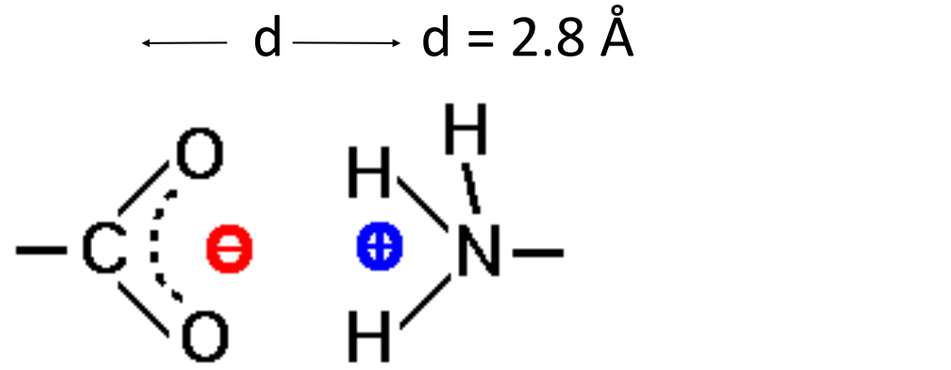
Key forces affecting structure:

- H-bonding
- Van der Waals
- Electrostatics
- Hydrophobicity
- Disulfide Bridges



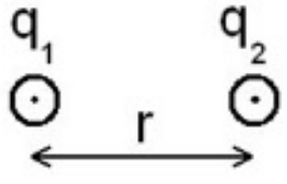
Key forces affecting structure:

- H-bonding
- Van der Waals
- Electrostatics
- Hydrophobicity
- Disulfide Bridges



carboxyl group and amino group

(some time called IONIC BONDs or SALT BRIDGEs)



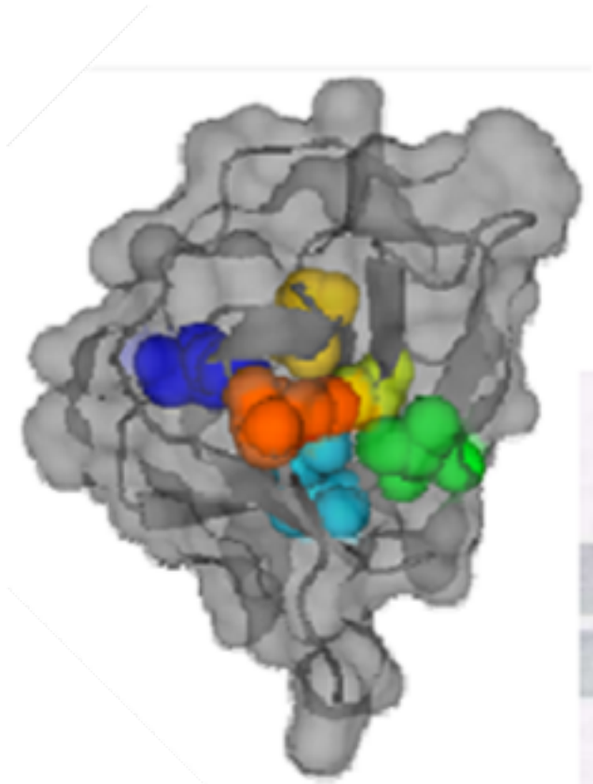
Coulomb's law

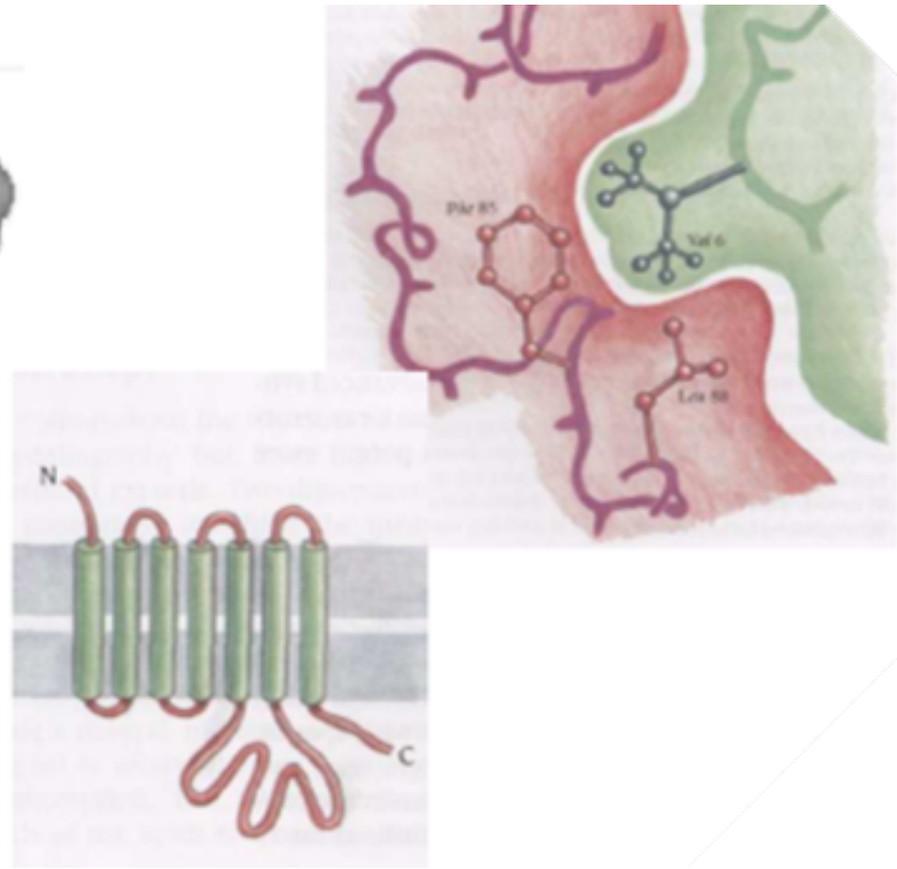
$$E = \frac{K\, q_1\, q_2}{D\, r}$$

E = Energy
k = constant
D = Dielectric constant (vacuum = 1; H_2O = 80)
q_1 & q_2 = electronic charges (Coulombs)
r = distance (Å)

Key forces affecting structure:

- H-bonding
- Van der Waals
- Electrostatics
- Hydrophobicity
- Disulfide Bridges



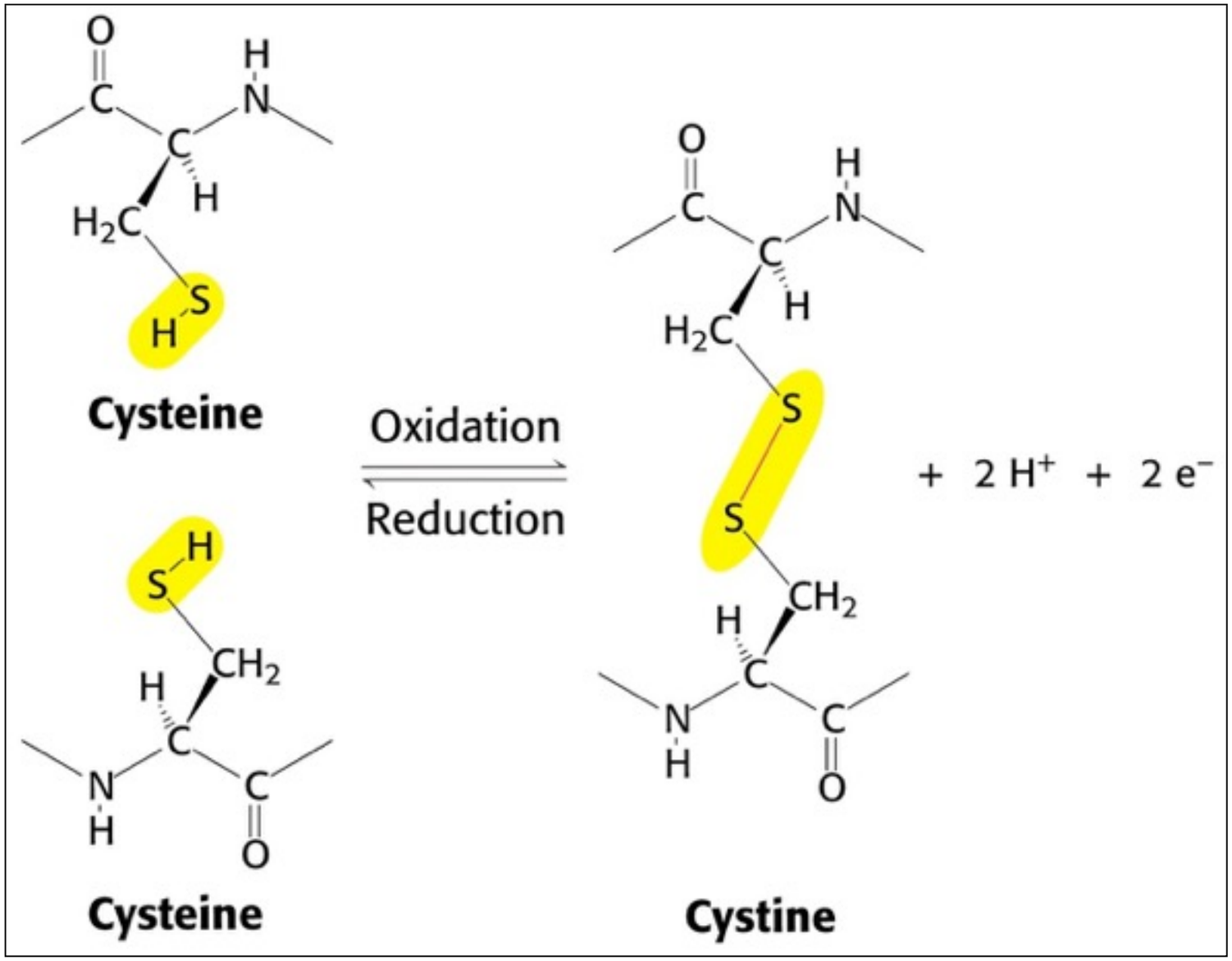


The force that causes hydrophobic molecules or nonpolar portions of molecules to aggregate together rather than to dissolve in water is called Hydrophobicity (*Greek, "water fearing"*). This is not a separate bonding force; rather, it is the result of the energy required to insert a nonpolar molecule into water.

Forces affecting structure:

- H-bonding
- Van der Waals
- Electrostatics
- Hydrophobicity
- Disulfide Bridges

Other names:
cystine bridge
disulfide bridge



Hair contains lots of disulfide bonds which are broken and reformed by heat

NEXT UP:

- **Overview of structural bioinformatics**
  - Major motivations, goals and challenges
- **Fundamentals of protein structure**
  - Composition, form, forces and dynamics
- **Representing and interpreting protein structure**
  - Modeling energy as a function of structure
- **Example application areas**
  - Predicting functional dynamics & drug discovery

RCSB Protein Data Bank

www.rcsb.org/pdb/home/home.do

Apps | Home | Docs | Gmail | Google Calendar | Access | 2CiteULike | 2Papers | 2Delicious | Journals | News&Sport | + | Research Develop | Other Bookmarks

RCSB PDB PROTEIN DATA BANK | PDB-101

A MEMBER OF THE PDB | EMDataBank

An Information Portal to Biological Macromolecular Structures

As of **Tuesday Nov 12, 2013 at 4 PM PST** there are **95475** Structures | **PDB Statistics**

Search | Advanced | Browse

Everything | Author | Macromolecule | Sequence | Ligand

e.g., PDB ID, molecule name, author

Search History , Previous Results

Customize This Page

Available on the App Store

PDB-101 Hide

- Structural View of Biology
- Understanding PDB Data
- Molecule of the Month
- Educational Resources
- Author Profiles

MyPDB Hide

- Login to your Account
- Register a New Account
- MyPDB Help Page

Home Hide

- News & Publications
- Usage/Reference Policies
- Deposition Policies
- Website FAQ
- Deposition FAQ
- Contact Us
- About Us
- Careers
- External Links
- Sitemap
- New Website Features

Deposition Hide

- All Deposit Services
- Electron Microscopy
- X-ray | NMR
- Validation Server
- BioSync Beamlines/Facilities
- Related Tools

Tools Hide

- Download Files
- Compare Structures
- Drug & Drug Target Mapping
- File Formats
- RESTful Web Services
- Widgets

Help Hide

- Launch Help System
- Display Settings
- Video Tutorials

Biological Macromolecular Resource

Full Description

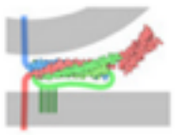
Learn: Featured Molecules Hide

Structural View of Biology

List View of Archive By: **Title** | **Date** | **Category**

PDB-101

Protein Synthesis

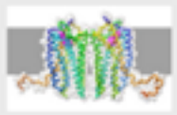


Molecule of the Month

SNARE Proteins

Small membrane-enclosed vesicles are used like cargo trucks to deliver proteins and other molecules inside and outside of cells. When these vesicles reach their proper destination, they fuse with a membrane and deliver their cargo. For instance, vesicles are used inside cells to transport digestive enzymes from the Golgi to their final location in lysosomes. They are also used to deliver molecules out of the cell: for example, neurotransmitters are released from vesicles that fuse with the cell membrane at nerve synapses. The 2013 Nobel Prize was awarded to three researchers who have revealed the central molecular machinery for this process of vesicle fusion.

Full Article



Protein Structure Initiative Featured System

G Proteins and Cancer

Miscommunication is the hallmark of cancer. Normally, our cells are in constant communication, deciding how to share resources, determining the best time to grow, and if necessary, the best time to die. Cancer cells, on the other hand, typically have corrupted these lines of communication, allowing them to grow without limits and selfishly steal resources for themselves. GPCRs (G-protein-coupled receptors) are among the many different molecules of communication that are changed when a normal cell is transformed into a cancer cell.

Full Article | **Archive** | **PSI Structural Biology Knowledgebase**

Explore Archive Hide

Organism | Taxonomy | Exp. Method | X-ray Resolution | Release Date | Polymer Type | Enzyme Classification | SCOP Classification | Protein Symmetry | Protein Stoichiometry

Show all

Experimental Method

- X-ray (84373)
- Solution NMR (10116)
- Electron Microscopy (690)
- Solid-State NMR (62)
- Hybrid (59)
- Neutron Diffraction (43)
- Fiber Diffraction (38)
- Electron Crystallography (38)
- Solution Scattering (32)
- Other (24)

Latest Structures Hide

4NCO : Crystal Structure of the BG505 SOSIP gp140 HIV-1 Env trimer in

New Features Hide

Latest release: September 2013

Visualize, browse & search symmetry / stoichiometry

Website Release Archive:

RCSB PDB News Hide

Weekly | Quarterly | Yearly

2013-11-12

Upcoming Meeting: ABRCMS

Meet the RCSB PDB at booth 715 at the Annual Biomedical Research Conference for Minority Students (Nov 13-16; Nashville, TN). **more**

- **Comparison Tool for Exploring Sequence and Structure Alignments**
- **Find Structures Using the Protein Symmetry Browser**
- **Fall Newsletter Published**

New Structures Hide

- Latest Release
- New Structure Papers

RCSB PDB - Query Results

www.rcsb.org/pdb/results/results.do?qrid=8BEC9330&tabtoshow=Current

Apps | Home | Docs | Gmail | Google Calendar | Access | 2CiteULike | 2Papers | 2Delicious | Journals | News&Sport | + | Research Develop | Other Bookmarks

RCSB PDB PROTEIN DATA BANK → PDB-101

A MEMBER OF THE PDB | EMDataBank

An Information Portal to Biological Macromolecular Structures

Nov 12, 2013 at 4 PM PST there are 95475 Structures | PDB Statistics

Search: HIV

Search
Advanced
Browse

Everything | Author

HIV

Search History (1), Previous Results

PDB-101 Hide
- Structural View of Biology
- Understanding PDB Data
- Molecule of the Month
- Educational Resources
- Author Profiles

MyPDB Hide
- Login to your Account
- Register a New Account
- MyPDB Help Page
- Query Results (2054)
- Query History (1)

Home Hide
- News & Publications
- Usage/Reference Policies
- Deposition Policies
- Website FAQ
- Deposition FAQ
- Contact Us
- About Us
- Careers
- External Links
- Sitemap
- New Website Features

Deposition Hide
- All Deposit Services
- Electron Microscopy
- X-ray | NMR
- Validation Server
- BioSync Beamlines/Facilities
- Related Tools

Tools Hide
- Download Files
- Compare Structures
- Drug & Drug Target Mapping
- File Formats
- RESTful Web Services
- Widgets

Help Hide
- Launch Help System
- Display Settings
- Video Tutorials
- Glossary of Terms
- RCSB PDB *Mobile*

2054 Structure Hits | 109 Unreleased Structures | 861 Citations | 760 Ligand Hits | 80 Web Page Hits

Query Parameters: Query Details | Save Query to MyPDB

Text Search for: hiv

Other search suggestions:

Molecule Name
- **HIV**-1 protease ... (448)
- **HIV**-1 REVERSE ... (210)
- **HIV**-2 PROTEASE ... (18)
- Anti-**HIV**-1 reverse ... (6)
- **HIV**-1 fusion ... (4)
- **HIV**-1 DIS RNA (4)

More - Find all

Structural Domains
- **HIV**-1 reverse ... (112)
- **HIV** Type ... (180)
- **HIV** RNase ... (86)
- **HIV**-1 Transactivator ... (5)
- **HIV**-1 gp41 ... (2)
- **HIV**-1 Reverse ... (6)

More

Molecule of the Month
- Integrase [**HIV**]
- **HIV**-1 Protease
- **HIV** Capsid
- Reverse Transcriptase [**HIV**]
- T-Cell Receptor [**HIV**]

Organism
- **HIV**-1 M:B_HXB2R (87)
- **HIV**-1 M:A (1)
- **HIV**-2 subtype A (9)
- **HIV**-1 M:J (1)
- **HIV**-1 M:J_SE9173 (1)
- **HIV**-1 M:B_ARV2/SF2 (52)

More

Enzyme Classification
- 3.4.23.16: **HIV**-1 retropepsin (548)
- 3.4.23.47: **HIV**-2 retropepsin (19)

UniProt Gene Names
- **HIV**EP1 (3)
- **HIV**1 ENV (4)

BIRD Molecules
- PRD_000280 - **HIV** entry ... (1)
- PRD_000281 - **HIV** ENTRY ... (2)

Find all

Chemical Name
- BE6: **HIV**-1 INHIBITOR
- BES: **HIV**-1 INHIBITOR

Find all

Ontology Terms
- HS : TAR (**HIV**-1) RNA ... (3)
- B04.820 ... .350: **HIV** [MeSH ... (1171)
- HS : TAR (**HIV**-1) RNA ... (3)
- D08.811 ... .187: **HIV** Reverse ... (218)
- HS : TAR (**HIV**-1) RNA ... (1)
- D27.505 ... Anti-**HIV** Agents ... (694)

More

Pfam Description
- PF13949 ... binding to **HIV** (10)

close

Query Refinements: ***Select an item or pie chart*** Hide

Organism
- Human immunodeficiency virus 1 (921)
- Homo sapiens (477)
- HIV-1 M:B_HXB2R (87)
- Mus musculus (85)
- Human immunodeficiency virus ty ... (60)
- Human immunodeficiency virus ty ... (58)
- HIV-1 M:B_ARV2/SF2 (52)
- Other (370)

Taxonomy
- Viruses (1464)
- Eukaryota (651)
- Unassigned (132)
- Bacteria (75)
- Other (25)
- Archaea (5)

Experimental Method
- X-ray (1735)
- Solution NMR (255)
- Electron Microscopy (56)
- Solid-State NMR (3)
- Other (2)
- Electron Crystallography (1)
- Neutron Diffraction (1)
- Hybrid (1)

X-ray Resolution
- less than 1.5 Å (183)
- 1.5 - 2.0 Å (579)
- 2.0 - 2.5 Å (496)
- 2.5 - 3.0 Å (347)
- 3.0 and more Å (131)
- more choices...

Release Date
- before 2000 (315)
- 2000 - 2005 (372)
- 2005 - 2010 (571)
- 2010 - today (796)
- this year (239)
- this month (9)
- more choices...

Polymer Type
- Protein (1885)
- RNA (79)
- Mixed (76)
- DNA (14)

Enzyme Classification
- 3: Hydrolases (841)
- 2: Transferases (304)
- 5: Isomerases (26)
- 6: Ligases (5)
- 1: Oxidoreductases (1)

SCOP Classification
- All beta proteins (529)
- Alpha and beta proteins (a/b) (128)
- Alpha and beta proteins (a+b) (119)
- Multi-domain proteins (alpha an ... (113)
- All alpha proteins (57)

Protein Symmetry
- Cyclic (913)
- Asymmetric (869)
- Dihedral (14)
- Helical (4)
- Icosahedral (1)

Protein Stoichiometry
- Homomer (917)
- Heteromer (588)
- Monomer (296)
- more choices...

Search: 1HSG (PDB ID)

www.rcsb.org/pdb/explore/explore.do?structureId=1HSG

RCSB PDB PROTEIN DATA BANK | PDB-101

A MEMBER OF THE PDB | EMDataBank
An Information Portal to Biological Macromolecular Structures
12, 2013 at 4 PM PST there are 95475 Structures | PDB Statistics

Search | Advanced | Browse

Everything | Author | Macrom

1HSG

Search History (1), Previous Results (2054)

PDB-101 Hide
Structural View of Biology
Understanding PDB Data
Molecule of the Month
Educational Resources
Author Profiles

MyPDB Hide
Login to your Account
Register a New Account
MyPDB Help Page
Query Results (2054)
Query History (1)

Home Hide
News & Publications
Usage/Reference Policies
Deposition Policies
Website FAQ
Deposition FAQ
Contact Us
About Us
Careers
External Links
Sitemap
New Website Features

Deposition Hide
All Deposit Services
Electron Microscopy
X-ray | NMR
Validation Server
BioSync Beamlines/Facilities
Related Tools

Tools Hide
Download Files
Compare Structures
Drug & Drug Target Mapping
File Formats
RESTful Web Services
Widgets

Help Hide
Launch Help System
Display Settings
Video Tutorials
Glossary of Terms
RCSB PDB *Mobile*

Summary | 3D View | Sequence | Annotations | Seq. Similarity | 3D Similarity | Literature | Biol. & Chem. | Methods | Geometry | Links

CRYSTAL STRUCTURE AT 1.9 ANGSTROMS RESOLUTION OF HUMAN IMMUNODEFICIENCY VIRUS (HIV) II PROTEASE COMPLEXED WITH L-735,524, AN ORALLY BIOAVAILABLE INHIBITOR OF THE HIV PROTEASES

1HSG

Display Files
Download Files
Share this Page

DOI:10.2210/pdb1hsg/pdb

Primary Citation

Crystal structure at 1.9-A resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases.

Chen, Z., Li, Y., Chen, E., Hall, D.L., Darke, P.L., Culberson, C., Shafer, J.A., Kuo, L.C.

Journal: (1994) J.Biol.Chem. **269:** 26344-26348

PubMed: 7929352
Search Related Articles in PubMed

PubMed Abstract:

L-735,524 is a potent, orally bioavailable inhibitor of human immunodeficiency virus (HIV) protease currently in a Phase II clinical trial. We report here the three-dimensional structure of L-735,524 complexed to HIV-2 protease at 1.9-A resolution, as well as the structure of the native HIV-2 protease at 2.5-A resolution. The structure of HIV-2 protease is found to be essentially identical to that of HIV-1 protease. In the crystal lattice of the HIV-2 protease complexed with L-735,524, the inhibitor is chelated to the active site of the homodimeric enzyme in one orientation. This feature allows an unambiguous assignment of protein-ligand interactions from the electron density map. Both Fourier and difference Fourier maps reveal clearly the closure of the flap domains of the protease upon L-735,524 binding. Specific interactions between the enzyme and the inhibitor include the hydroxy group of the hydroxyaminopentane amide moiety of L-735,524 ligating to the carboxyl groups of the essential Asp-25 and Asp-25' enzymic residues and the amide oxygens of the inhibitor hydrogen bonding to the backbone amide nitrogen of Ile-50 and Ile-50' via an intervening water molecule. A second bridging water molecule is found between the amide nitrogen N2 of L-735,524 and the carboxyl oxygen of Asp-29'. Although other hydrogen bonds also add to binding, an equally significant contribution to affinity arises from hydrophobic interactions between the protease and the inhibitor throughout the pseudo-symmetric S1/S1', S2/S2', and S3/S3' regions of the enzyme. Except for its pyridine ring, all lipophilic moieties (t-butyl, indanyl, benzyl, and piperidyl) of L-735,524 are rigidly defined in the active site.

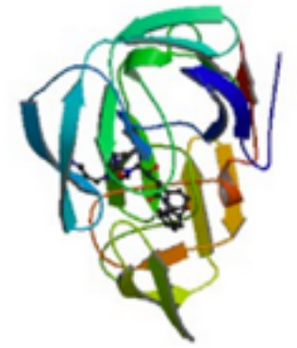
Keywords:
Aspartic Acid Endopeptidases, Binding Sites, Crystallography, X-Ray, Drug Resistance, HIV Protease, HIV Protease Inhibitors, Indinavir, Pyridines

Related Structures:
Primary Citation of: 1HSG 1HSH 1HSI

Organizational Affiliation:
Department of Biological Chemistry, Merck Research Laboratories, West Point, Pennsylvania 19486.

Click on abstract words and keywords to add them to the search box.

Biological Assembly



3D View | More Images...

Symmetry: **C2 view**
Stoichiometry: **Homo 2-mer - A2**
Biological assembly 1 assigned by authors and generated by PISA (software)

Downloadable viewers:
Simple Viewer | Protein Workshop | Kiosk Viewer

MyPDB Personal Annotations Hide

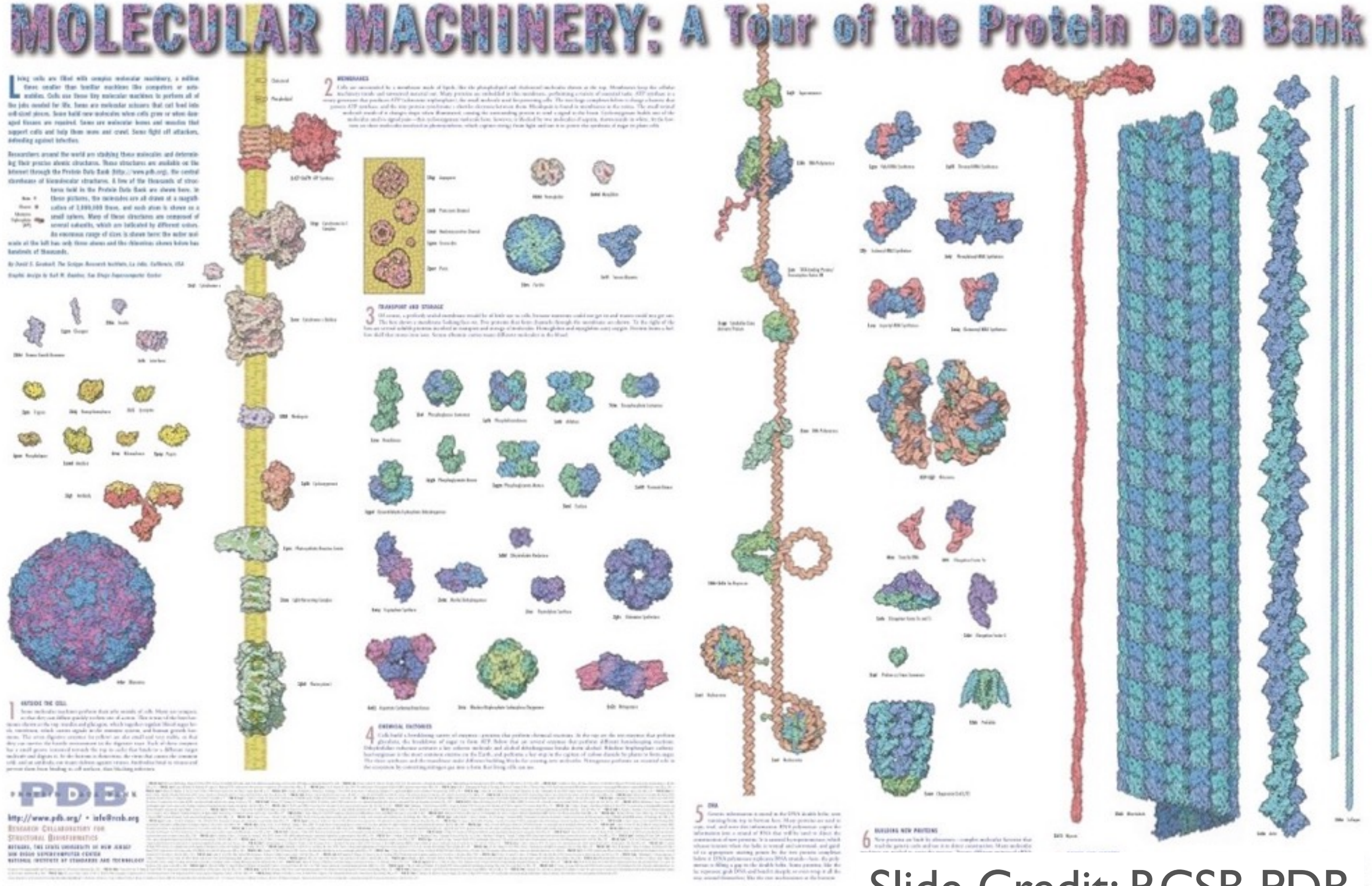
To save personal annotations, please **login** to your MyPDB account.

Deposition Summary Hide

Authors: Chen, Z.

Deposition:	1995-03-31
Release:	1996-04-03

MOLECULAR MACHINERY: A Tour of the Protein Data Bank



Slide Credit: RCSB PDB

PDB FILE FORMAT

```
                                  Chain name
    Amino Acid                   /   Sequence Number
                \               /   /
     Element     \             /   /      -----Coordinates-----
                  \   \       /   /       X        Y        Z       (etc.)
ATOM       1   N    ASP L    1           4.060    7.307    5.186   ...
ATOM       2   CA   ASP L    1           4.042    7.776    6.553   ...
ATOM       3   C    ASP L    1           2.668    8.426    6.644   ...
ATOM       4   O    ASP L    1           1.987    8.438    5.606   ...
ATOM       5   CB   ASP L    1           5.090    8.827    6.797   ...
ATOM       6   CG   ASP L    1           6.338    8.761    5.929   ...
ATOM       7   OD1 ASP L    1           6.576    9.758    5.241   ...
ATOM       8   OD2 ASP L    1           7.065    7.759    5.948   ...
                  \\
                   Element position within amino acid
```

- **PDB files** contains atomic coordinates and associated information.

KEY CONCEPT: POTENTIAL FUNCTIONS DESCRIBE A SYSTEMS **ENERGY** AS A FUNCTION OF ITS **STRUCTURE**

Two main approaches:

(1). **Physics-Based**

(2). **Knowledge-Based**

KEY CONCEPT: POTENTIAL FUNCTIONS DESCRIBE A SYSTEMS **ENERGY** AS A FUNCTION OF ITS **STRUCTURE**

Two main approaches:

(1). **Physics-Based**

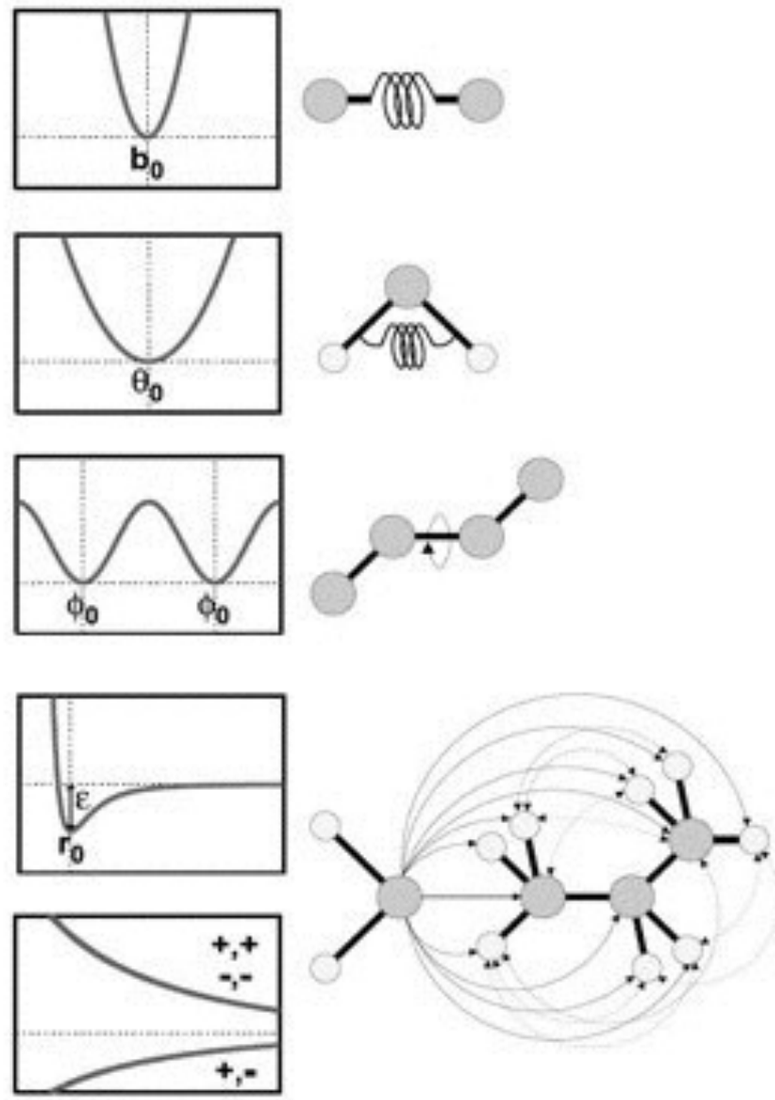
(2). **Knowledge-Based**

PHYSICS-BASED POTENTIALS
ENERGY TERMS FROM PHYSICAL THEORY

$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} +$$

$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} +$$

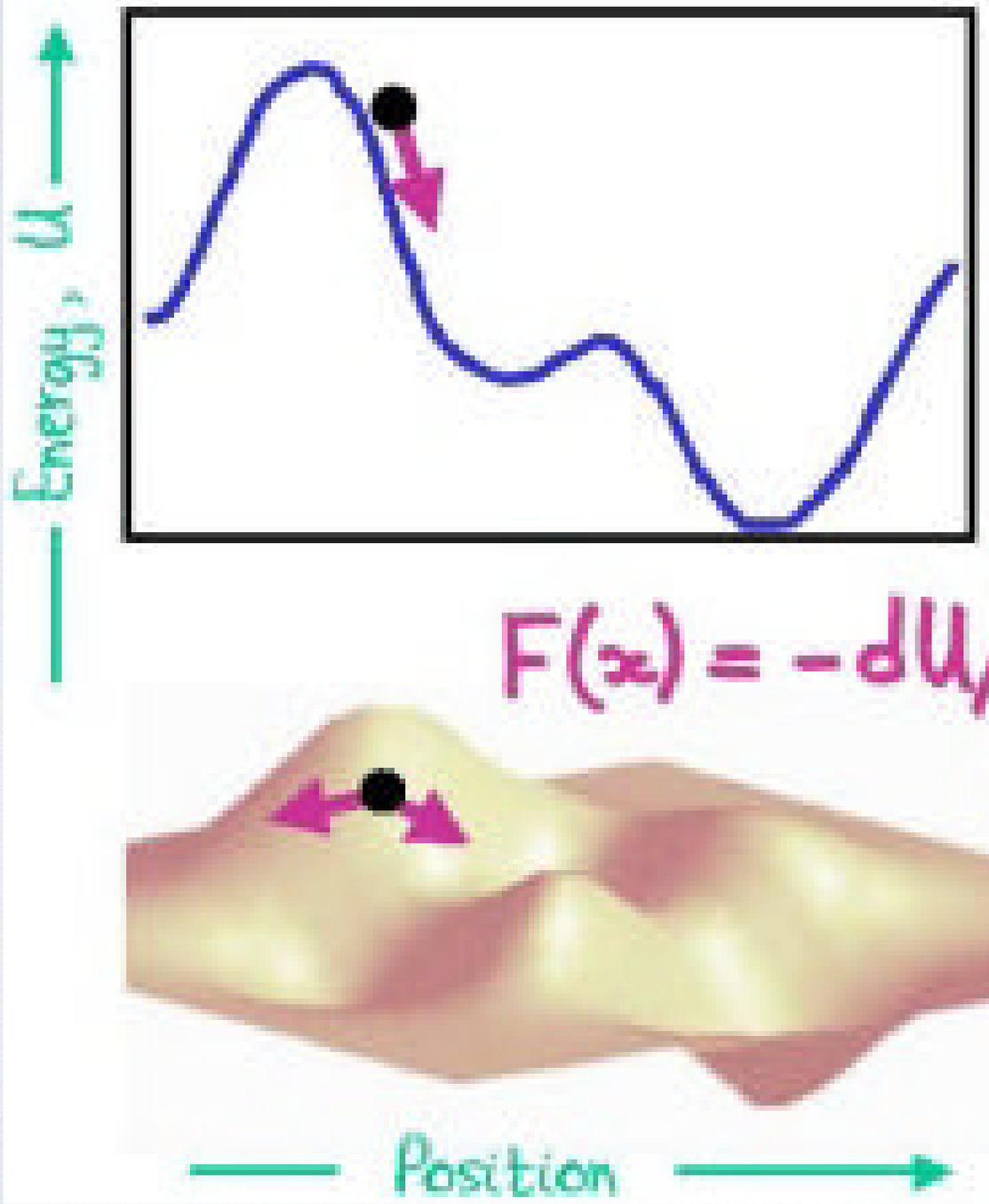
$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$



U_{bond} = oscillations about the equilibrium bond length
U_{angle} = oscillations of 3 atoms about an equilibrium bond angle
$U_{dihedral}$ = torsional rotation of 4 atoms about a central bond
$U_{nonbond}$ = non-bonded energy terms (electrostatics and Lenard-Jones)

CHARMM P.E. function, see: http://www.charmm.org/

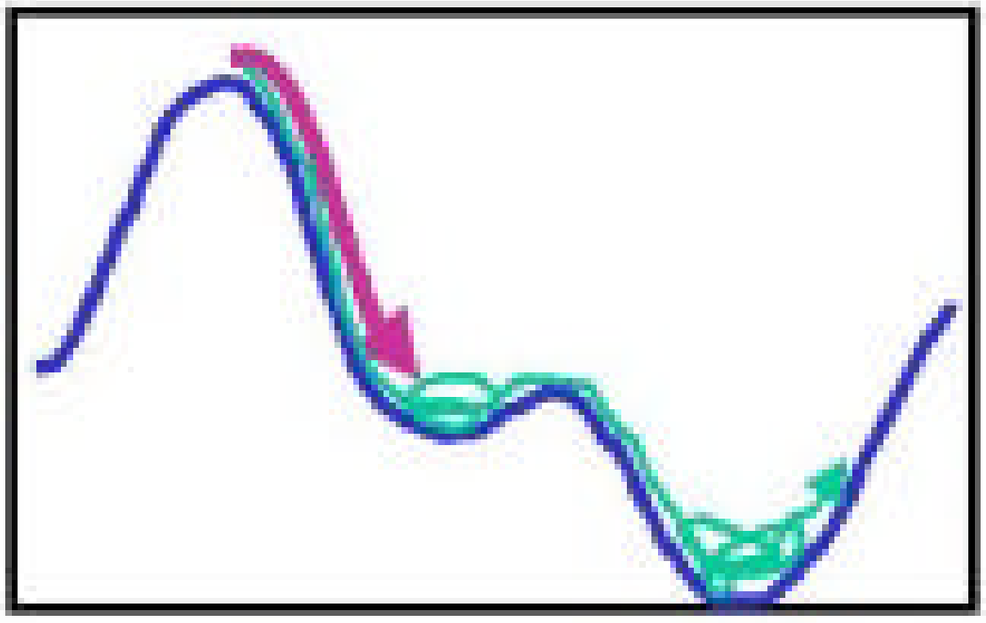
TOTAL POTENTIAL ENERGY



- The total potential energy or enthalpy fully defines the system, U.
- The forces are the gradients of the energy.
- The energy is a sum of independent terms for: Bond, Bond angles, Torsion angles and non-bonded atom pairs.

Slide Credit: Michael Levitt

MOVING OVER THE ENERGY SURFACE



- Energy Minimization drops into local minimum.
- Molecular Dynamics uses thermal energy to move smoothly over surface.
- Monte Carlo Moves are random. Accept with probability $\exp(-\Delta U/kT)$.

Slide Credit: Michael Levitt

PHYSICS-ORIENTED APPROACHES

Weaknesses

Fully physical detail becomes computationally intractable

Approximations are unavoidable

(Quantum effects approximated classically, water may be treated crudely)

Parameterization still required

Strengths

Interpretable, provides guides to design

Broadly applicable, in principle at least

Clear pathways to improving accuracy

Status

Useful, widely adopted but far from perfect

Multiple groups working on fewer, better approxs

Force fields, quantum

entropy, water effects

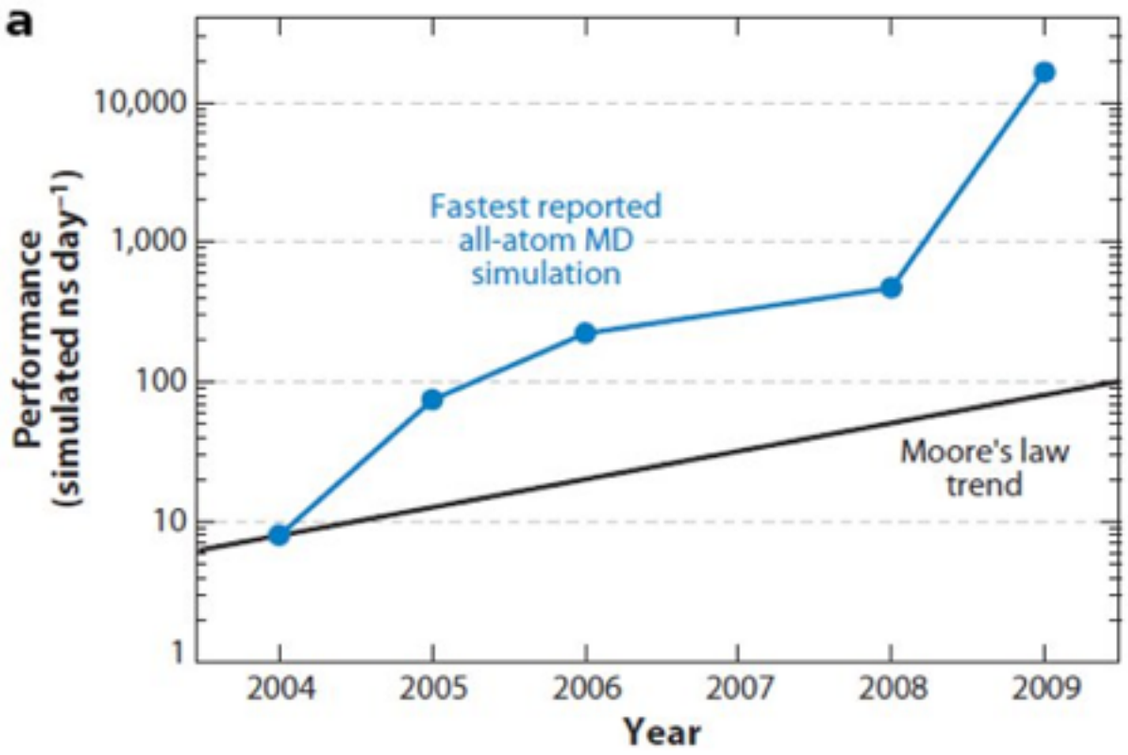
Moore's law: hardware improving

HOW COMPUTERS HAVE CHANGED

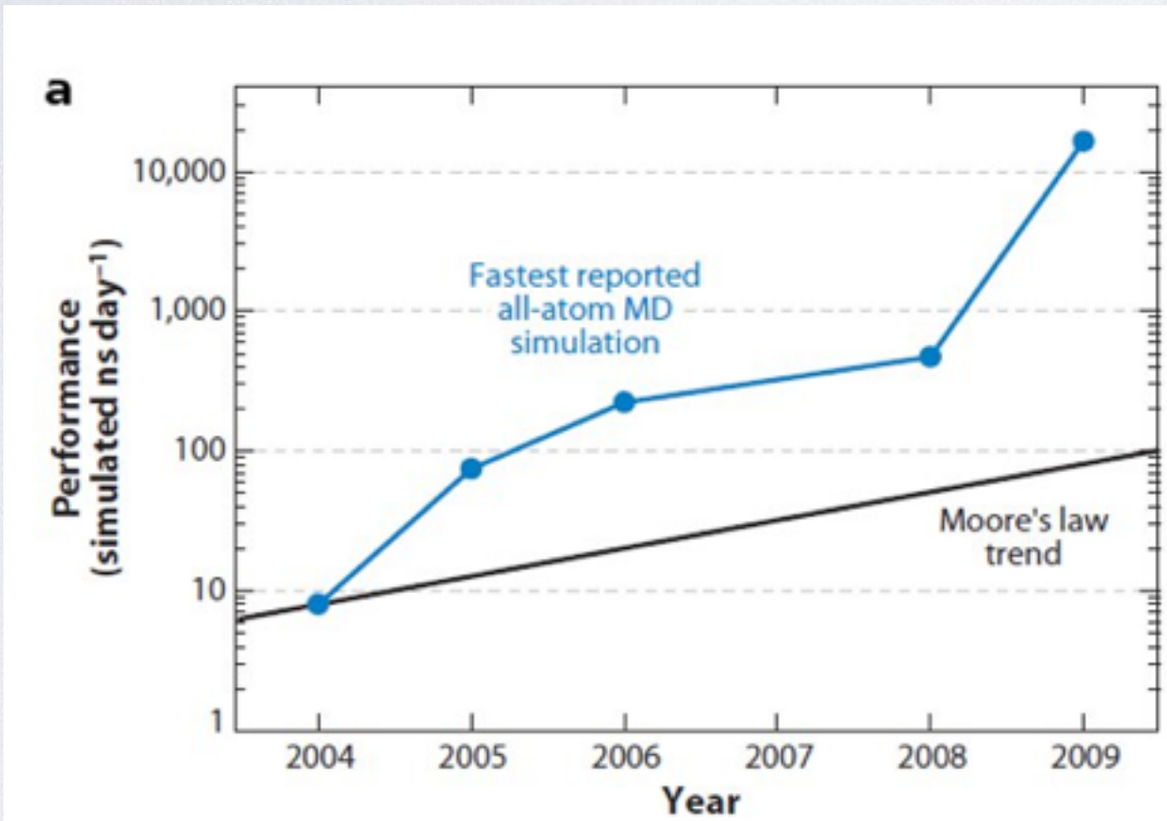
DATE	COST	SPEED	MEMORY	SIZE
1967	$4OM	0.1 MHz	1 MB	HALL
2013	$4,000	1 GHz	10 GB	LAPTOP
CHANGE	10,000	10,000	10,000	10,000

If cars were like computers then a new Volvo would cost $3, would have a top speed of 1,000,000 Km/hr, would carry 50,000 adults and would park in a shoebox

SIDE-NOTE: GPUS AND ANTON SUPERCOMPUTER



SIDE-NOTE: GPUS AND ANTON SUPERCOMPUTER



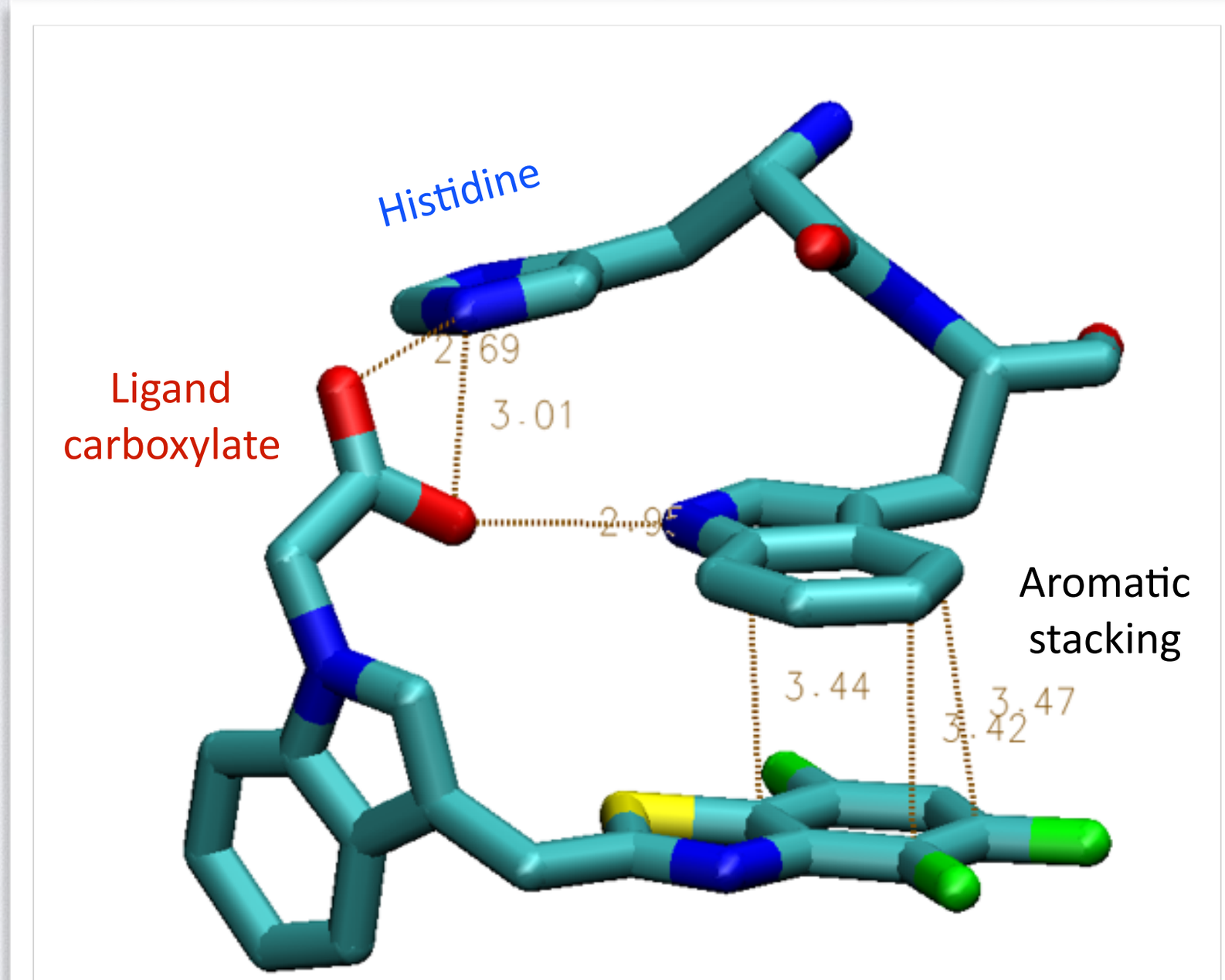
KEY CONCEPT: POTENTIAL FUNCTIONS DESCRIBE A SYSTEMS **ENERGY** AS A FUNCTION OF ITS **STRUCTURE**

Two main approaches:

(1). **Physics-Based**

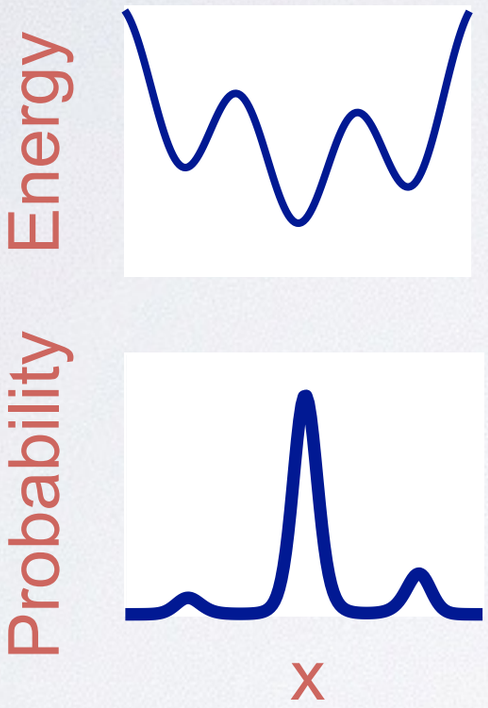
(2). **Knowledge-Based**

KNOWLEDGE-BASED DOCKING POTENTIALS



ENERGY DETERMINES **PROBABILITY** (STABILITY)

Basic idea: Use probability as a proxy for energy



Boltzmann:

$$p(r) \propto e^{-E(r)/RT}$$

Inverse Boltzmann:

$$E(r) = -RT \ln\left[p(r)\right]$$

Example: ligand carboxylate O to protein histidine N

Find all protein-ligand structures in the PDB with a ligand carboxylate O

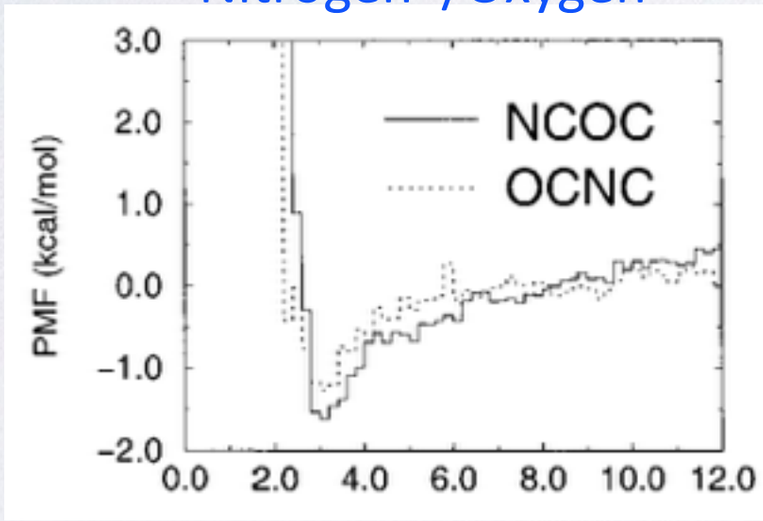
1. For each structure, histogram the distances from O to every histidine N
2. Sum the histograms over all structures to obtain $p(r_{O\text{-}N})$
3. Compute $E(r_{O\text{-}N})$ from $p(r_{O\text{-}N})$

KNOWLEDGE-BASED DOCKING POTENTIALS

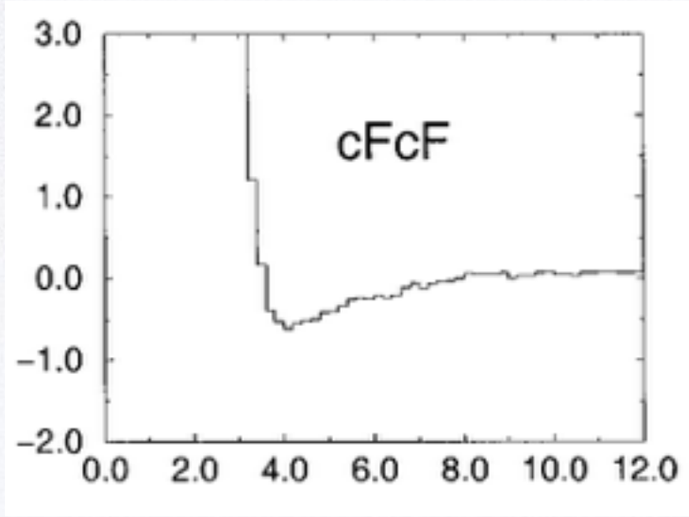
"PMF", Muegge & Martin, J. Med. Chem. (1999) 42:791

A few types of atom pairs, out of several hundred total

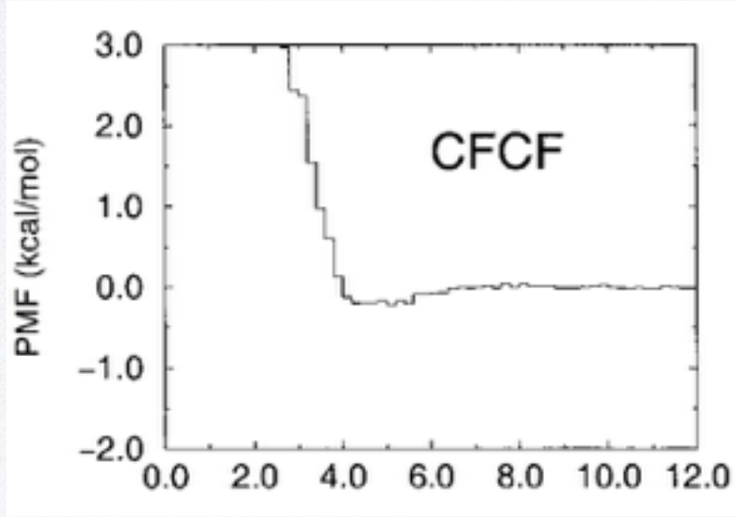
Nitrogen^{+}/Oxygen^{-}



Aromatic carbons



Aliphatic carbons



Atom-atom distance (Angstroms)

$$E_{prot-lig} = E_{vdw} + \sum_{pairs\ (ij)} E_{type(ij)}(r_{ij})$$

KNOWLEDGE-BASED POTENTIALS

Weaknesses

Accuracy limited by availability of data

Strengths

Relatively easy to implement

Computationally fast

Status

Useful, far from perfect

May be at point of diminishing returns

(not always clear how to make improvements)

NEXT UP:

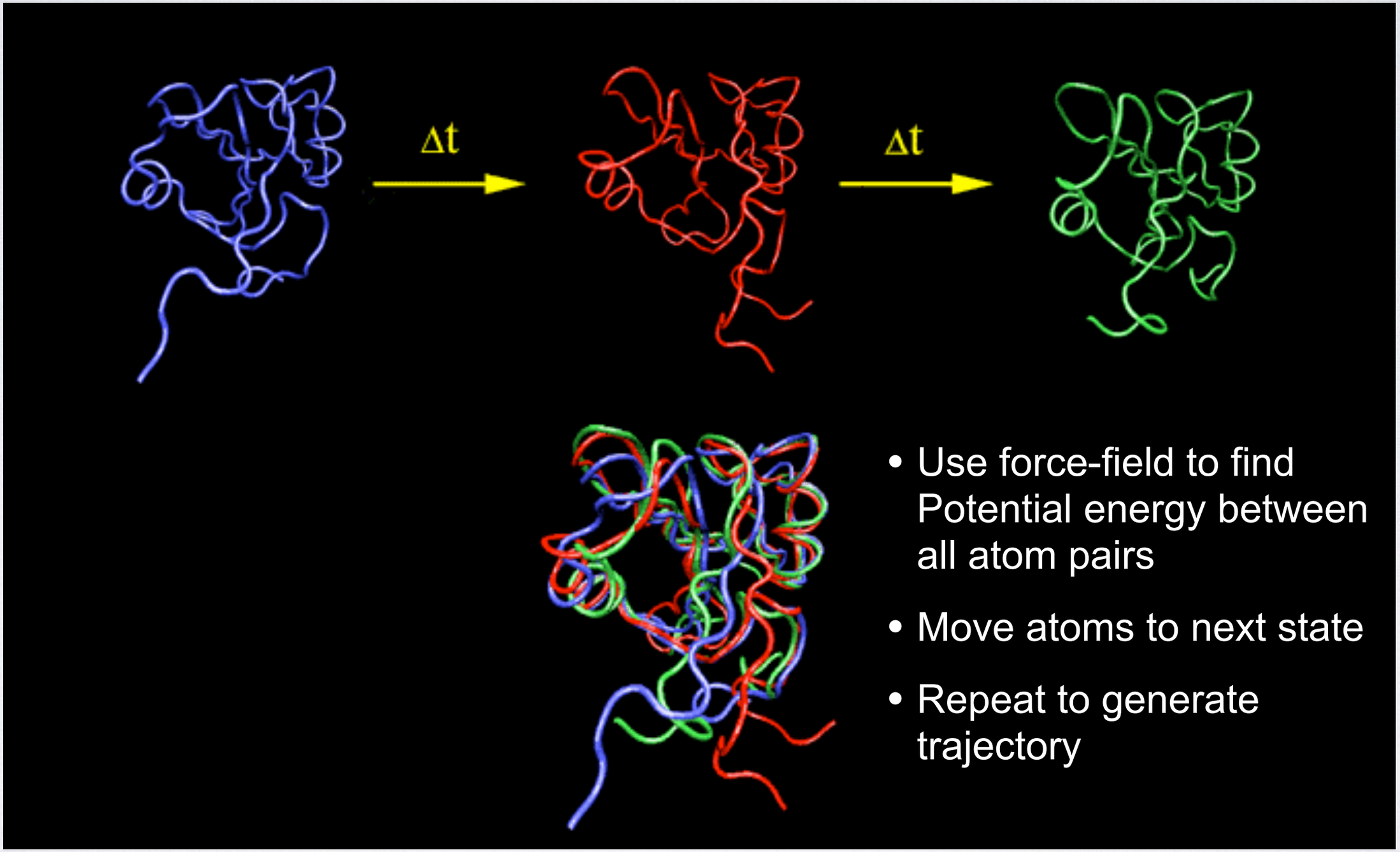
- **Overview of structural bioinformatics**
  - Major motivations, goals and challenges
- **Fundamentals of protein structure**
  - Composition, form, forces and dynamics
- **Representing and interpreting protein structure**
  - Modeling energy as a function of structure
- **Example application areas**
  - Predicting functional dynamics & drug discovery

PREDICTING FUNCTIONAL DYNAMICS

- **Proteins are <u>intrinsically flexible</u> molecules with internal motions that are often intimately coupled to their biochemical function**
  - E.g. ligand and substrate binding, conformational activation, allosteric regulation, etc.

- **Thus knowledge of dynamics can provide a deeper understanding of the <u>mapping of structure to function</u>**
  - **<u>Molecular dynamics</u>** (MD) and **<u>normal mode analysis</u>** (NMA) are two major methods for predicting and characterizing molecular motions and their properties

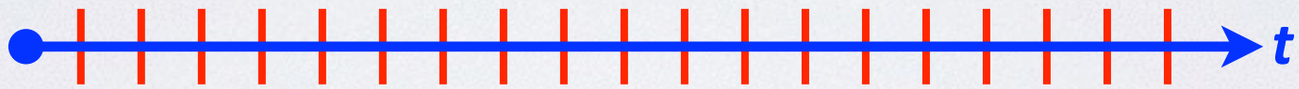
MOLECULAR DYNAMICS SIMULATION



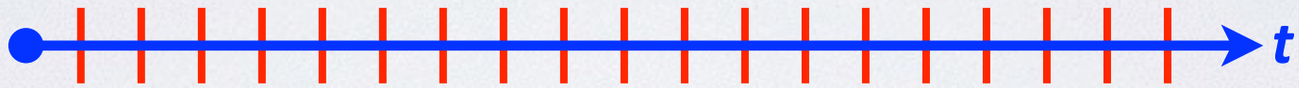
McCammon, Gelin & Karplus, *Nature* (1977)

[See: https://www.youtube.com/watch?v=ui1ZysMFcKk]

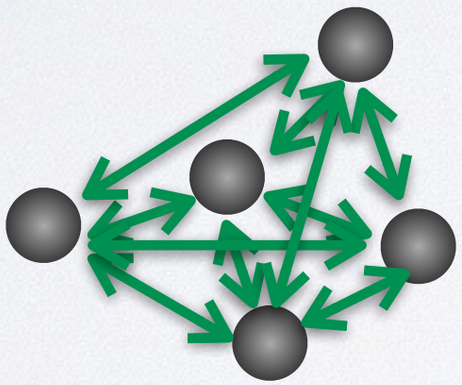
- Divide **time** into discrete (~1fs) **time steps** (**Δt**)
  (for integrating equations of motion, see below)



- Divide **time** into discrete (~1fs) **time steps** (**Δt**)
  (for integrating equations of motion, see below)



- At each time step calculate pair-wise atomic **forces** (***F(t)***)
  (by evaluating **force-field** gradient)



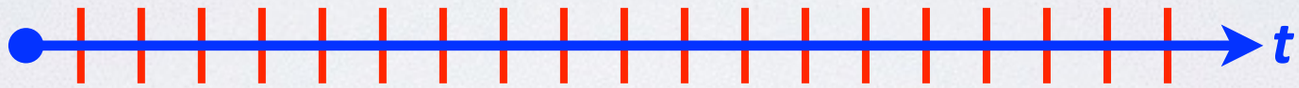
Nucleic motion described classically

$$m_i \frac{d^2}{dt^2} \vec{R}_i = -\vec{\nabla}_i E(\vec{R})$$

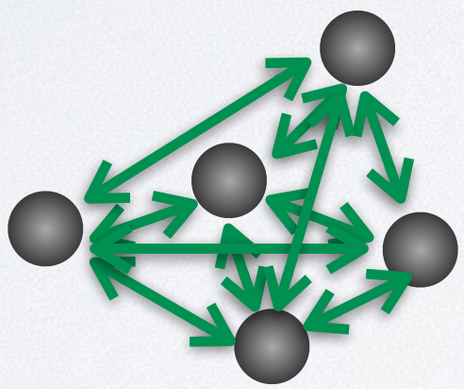
Empirical force field

$$E(\vec{R}) = \sum_{\text{bonded}} E_i(\vec{R}) + \sum_{\text{non-bonded}} E_i(\vec{R})$$

- Divide **time** into discrete (~1fs) **time steps** (**Δt**)
  (for integrating equations of motion, see below)



- At each time step calculate pair-wise atomic **forces** (***F(t)***)
  (by evaluating **force-field** gradient)



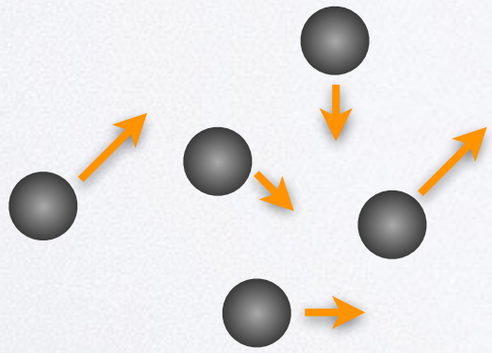
Nucleic motion described classically

$$m_i \frac{d^2}{dt^2}\vec{R}_i = -\vec{\nabla}_i E(\vec{R})$$

Empirical force field

$$E(\vec{R}) = \sum_{\text{bonded}} E_i(\vec{R}) + \sum_{\text{non-bonded}} E_i(\vec{R})$$

- Use the forces to calculate **velocities** and move atoms to new **positions**
  (by integrating numerically via the “leapfrog” scheme)

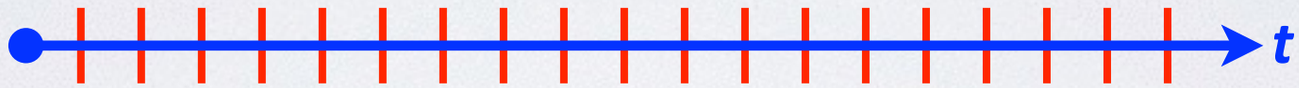


$$\boldsymbol{v}(t+\frac{\Delta t}{2}) = \boldsymbol{v}(t-\frac{\Delta t}{2}) + \frac{\boldsymbol{F}(t)}{m}\Delta t$$

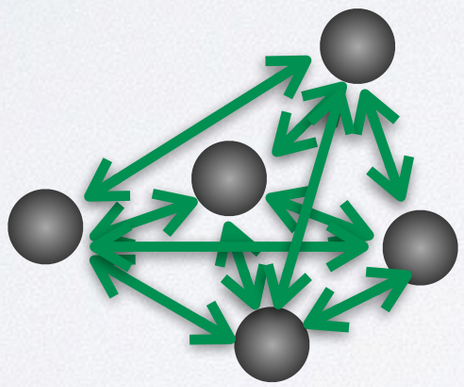
$$\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t+\frac{\Delta t}{2})\Delta t$$

BASIC ANATOMY OF A MD SIMULATION

- Divide **time** into discrete (~1fs) **time steps** (**Δt**)
  (for integrating equations of motion, see below)



- At each time step calculate pair-wise atomic **forces** (***F(t)***)
  (by evaluating **force-field** gradient)



Nucleic motion described classically

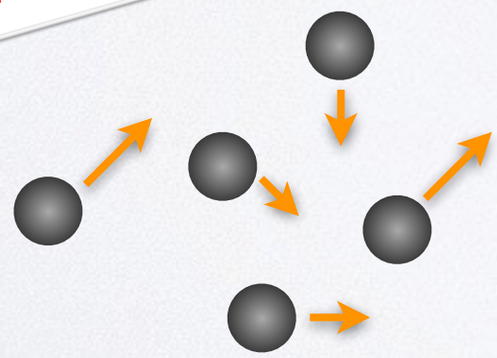
$$m_i \frac{d^2}{dt^2}\vec{R}_i = -\vec{\nabla}_i E(\vec{R})$$

Empirical f

$$E(\vec{R}) = \sum_{\text{non-bonded}} E_i(\vec{R})$$

REPEAT, (iterate many, many times… 1ms = 10^{12} time steps)

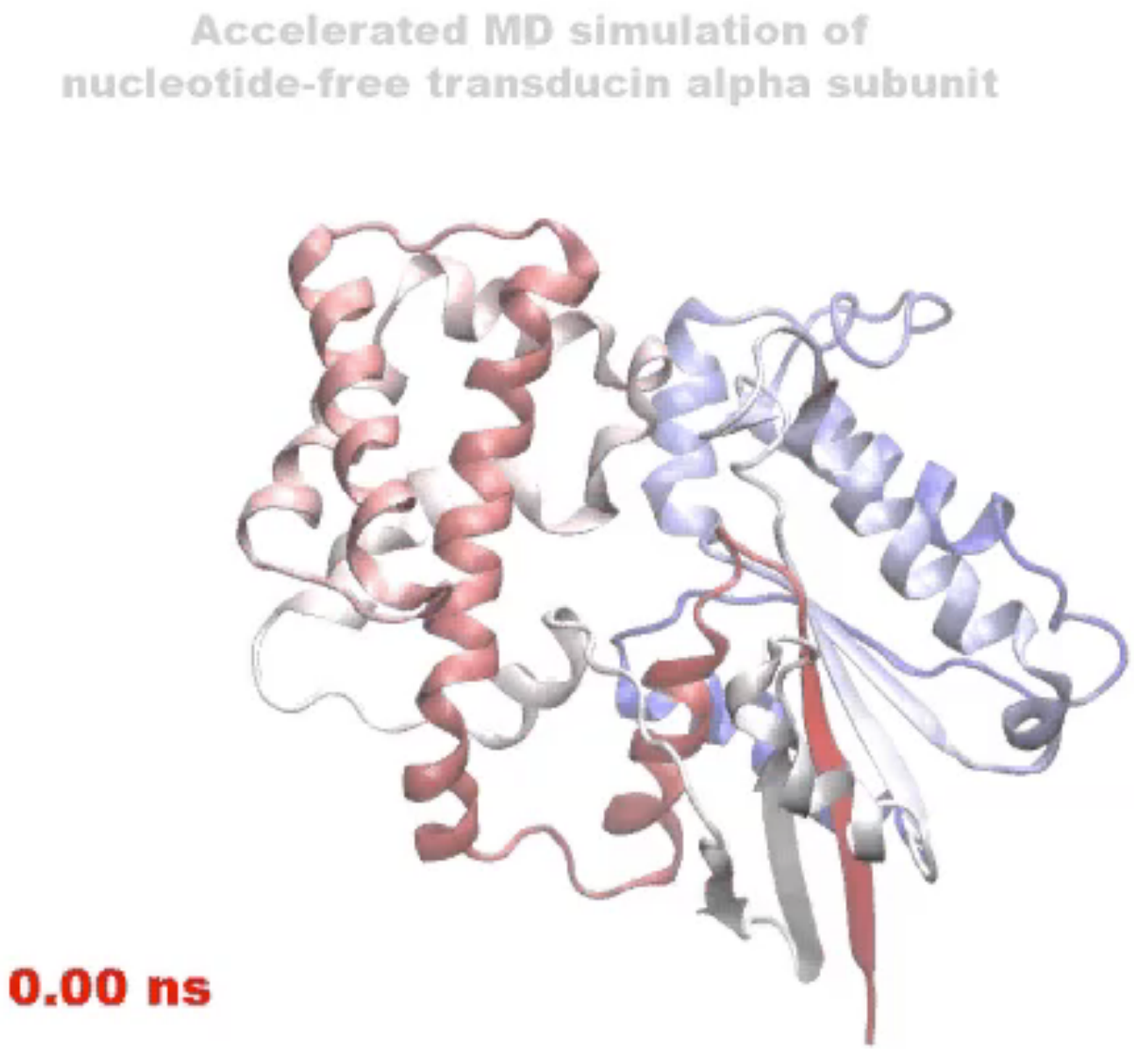
- Use the f… **velocities** and move atoms to new **positions**
  …ng numerically via the "leapfrog" scheme)



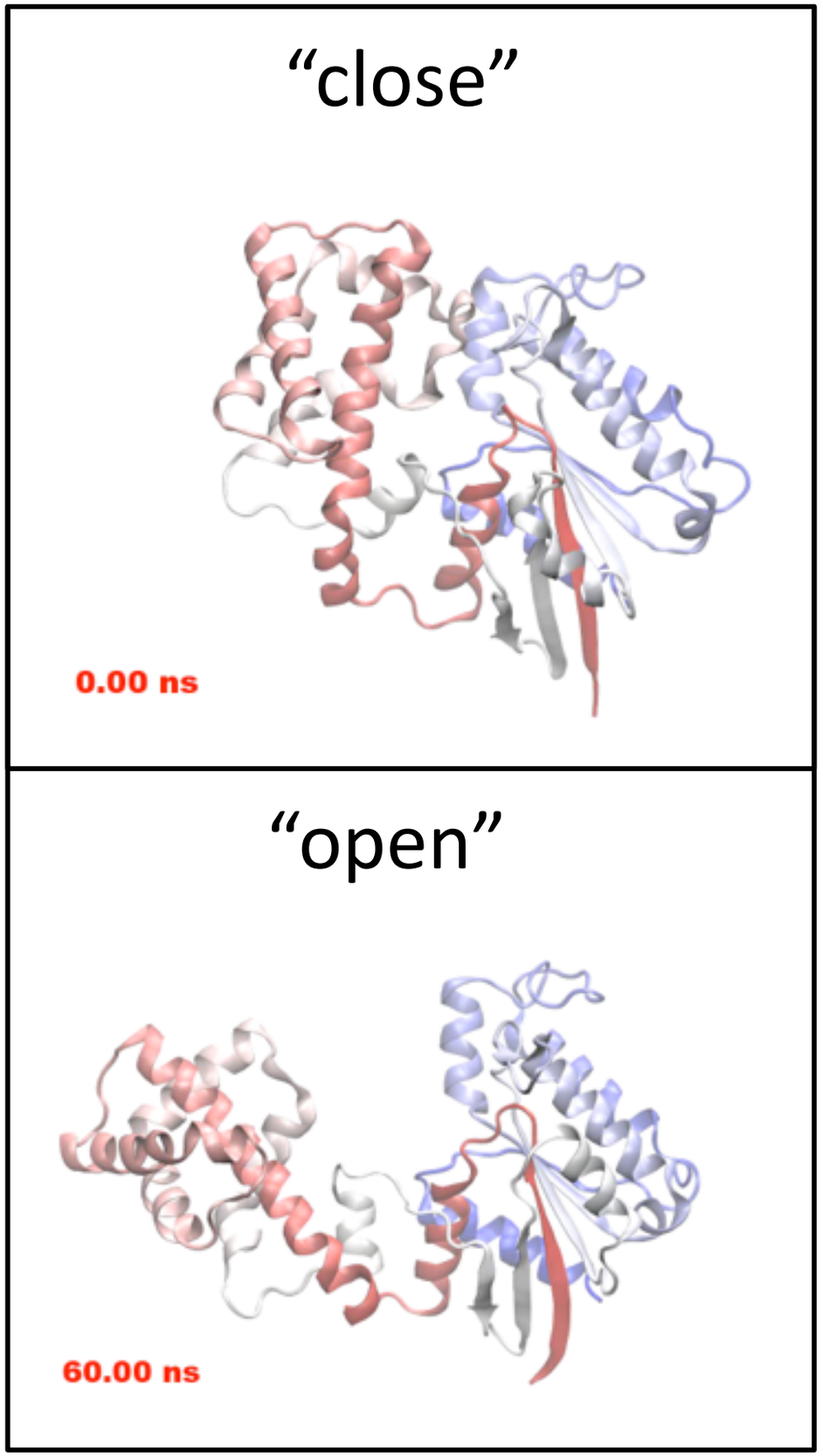
$$\boldsymbol{v}(t+\frac{\Delta t}{2}) = \boldsymbol{v}(t-\frac{\Delta t}{2}) + \frac{\boldsymbol{F}(t)}{m}\Delta t$$

$$\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t+\frac{\Delta t}{2})\Delta t$$

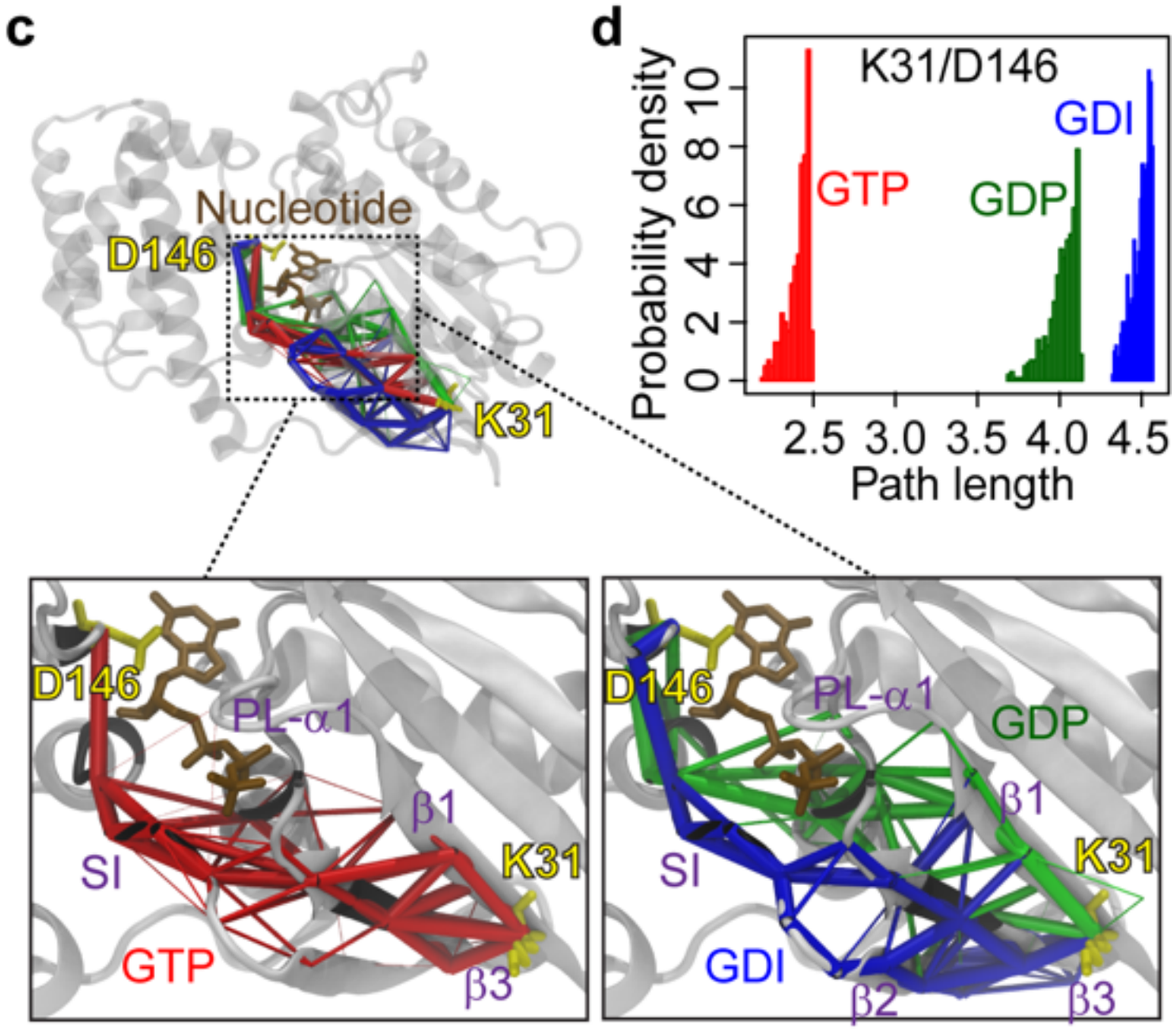
MD Prediction of Functional Motions



Yao and Grant, Biophys J. (2013)

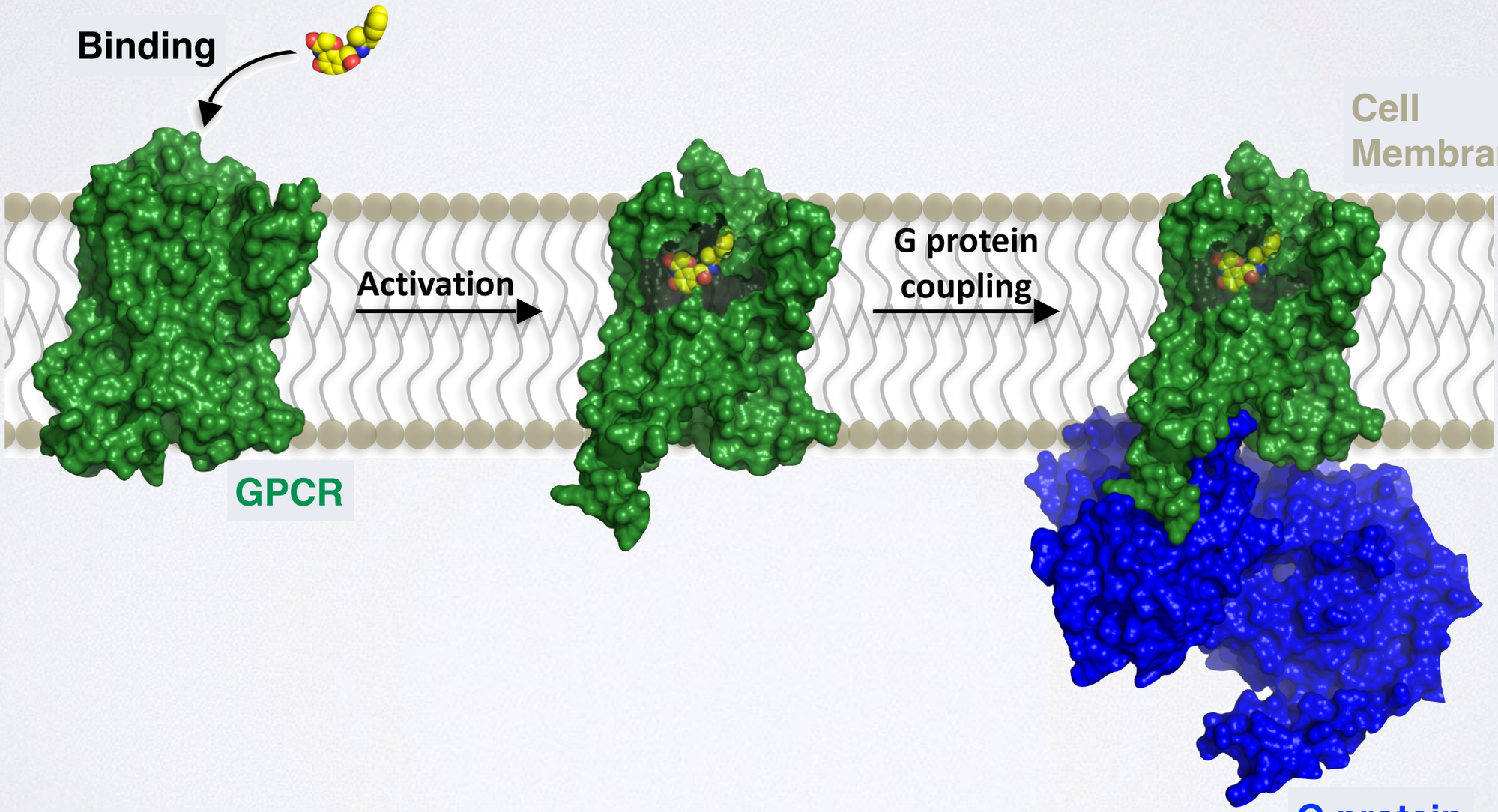


Simulations Identify Key Residues Mediating Dynamic Activation

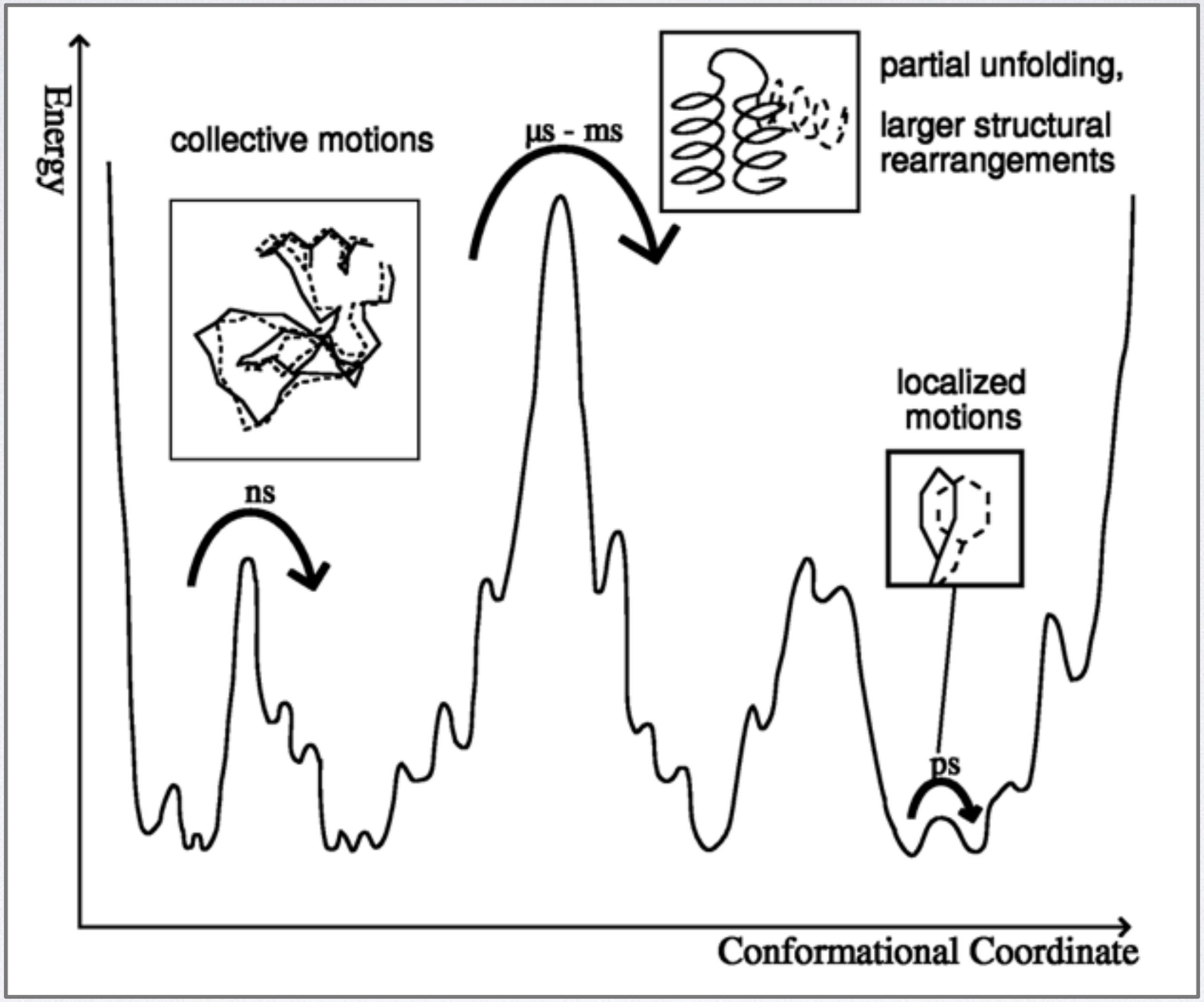


Yao … Grant, Journal of Biological Chemistry (2016)

EXAMPLE APPLICATION OF MOLECULAR SIMULATIONS TO GPCRS



PROTEINS JUMP BETWEEN MANY, HIERARCHICALLY ORDERED "CONFORMATIONAL SUBSTATES"



H. Frauenfelder et al., *Science* **229** (1985) 337

Improve this slide

MOLECULAR DYNAMICS IS VERY EX

Example: F_1-ATPase in water (183,674 atoms) for 1 nanosecond:

=> 10^6 integration steps

=> $8.4 * 10^{11}$ floating point operations/step

[n(n-1)/2 interactions]

Total: $8.4 * 10^{17}$ flop

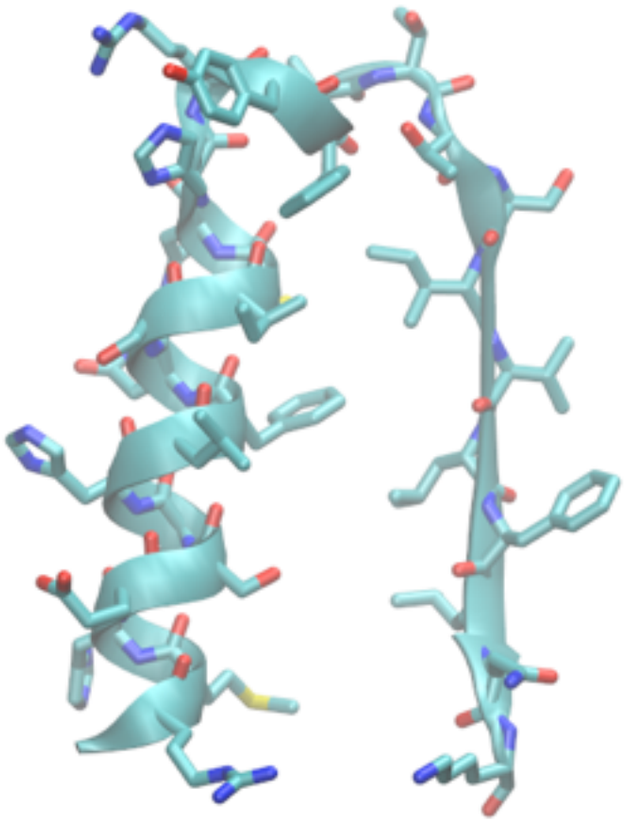
(on a 100 Gflop/s cpu: **ca 25 years!**)

… but performance has been improved by use of:

multiple time stepping	ca. 2.5 years
fast multipole methods	ca. 1 year
parallel computers	ca. 5 days
modern GPUs	**ca. 1 day**
(Anton supercomputer	**ca. minutes)**

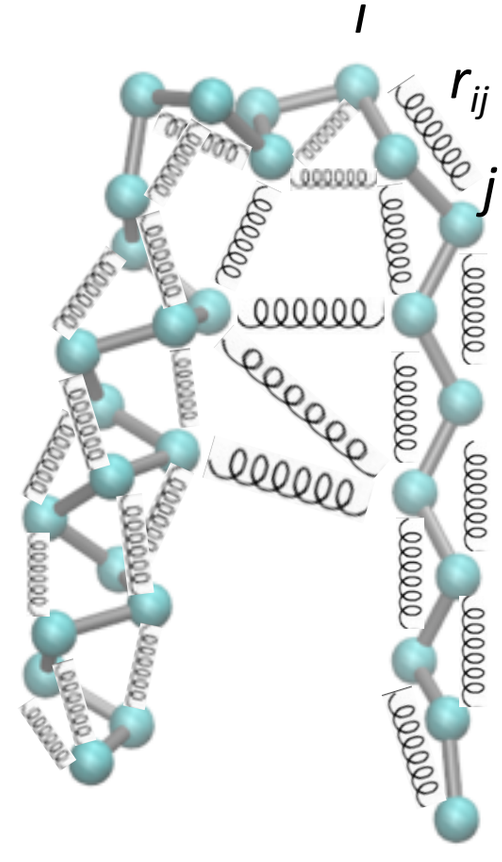
COARSE GRAINING: **NORMAL MODE ANALYSIS** (NMA)

- MD is still time-consuming for large systems
- Elastic network model NMA (ENM-NMA) is an example of a lower resolution approach that finishes in seconds even for large systems.



Atomistic

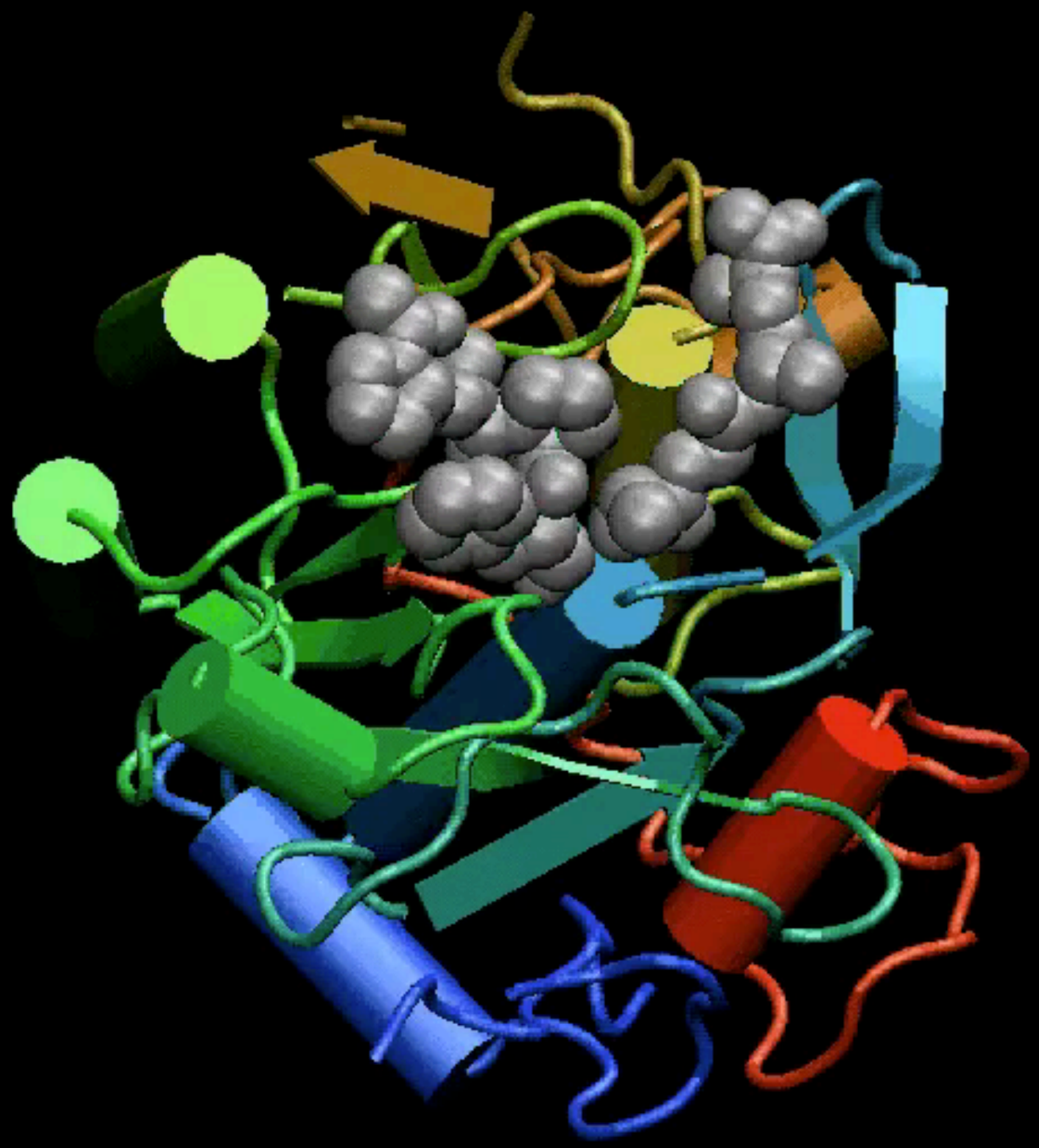
C. G.



Coarse Grained

- 1 bead / 1 amino acid
- Connected by springs

NMA models the protein as a network of elastic strings



Proteinase K

NEXT UP:

- **Overview of structural bioinformatics**
  - Major motivations, goals and challenges
- **Fundamentals of protein structure**
  - Composition, form, forces and dynamics
- **Representing and interpreting protein structure**
  - Modeling energy as a function of structure
- **Example application areas**
  - Predicting functional dynamics & **drug discovery**

THE TRADITIONAL EMPIRICAL PATH TO DRUG DISCOVERY

Compound library
(commercial, in-house, synthetic, natural)

↘

High throughput screening
(HTS)

↘

Hit confirmation

↘

Lead compounds
(e.g., μM K_d)

↘

Lead optimization
(Medicinal chemistry)

↓

Potent drug candidates
(nM K_d)

←

Animal and clinical evaluation

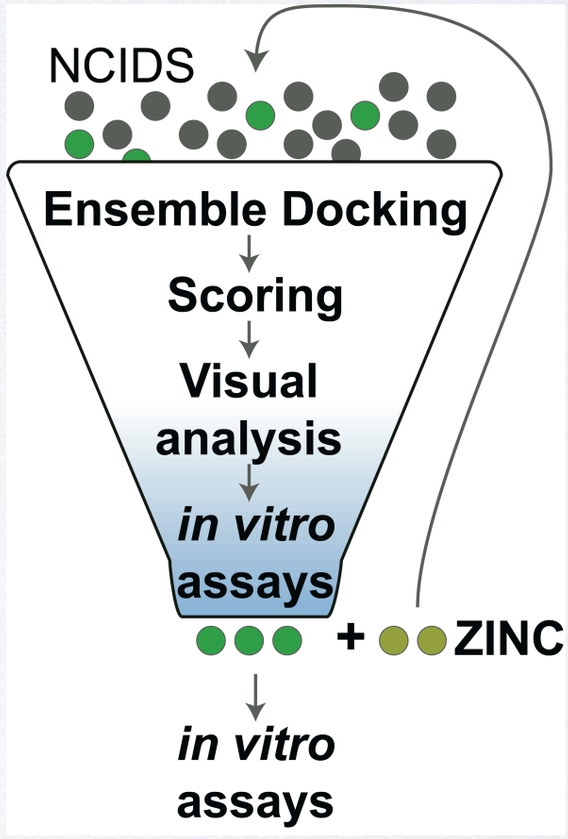
COMPUTER-AIDED LIGAND DESIGN

Aims to reduce number of compounds synthesized and assayed

Lower costs

Reduce chemical waste

Facilitate faster progress



Two main approaches:

(1). **Receptor/Target-Based**

(2). **Ligand/Drug-Based**

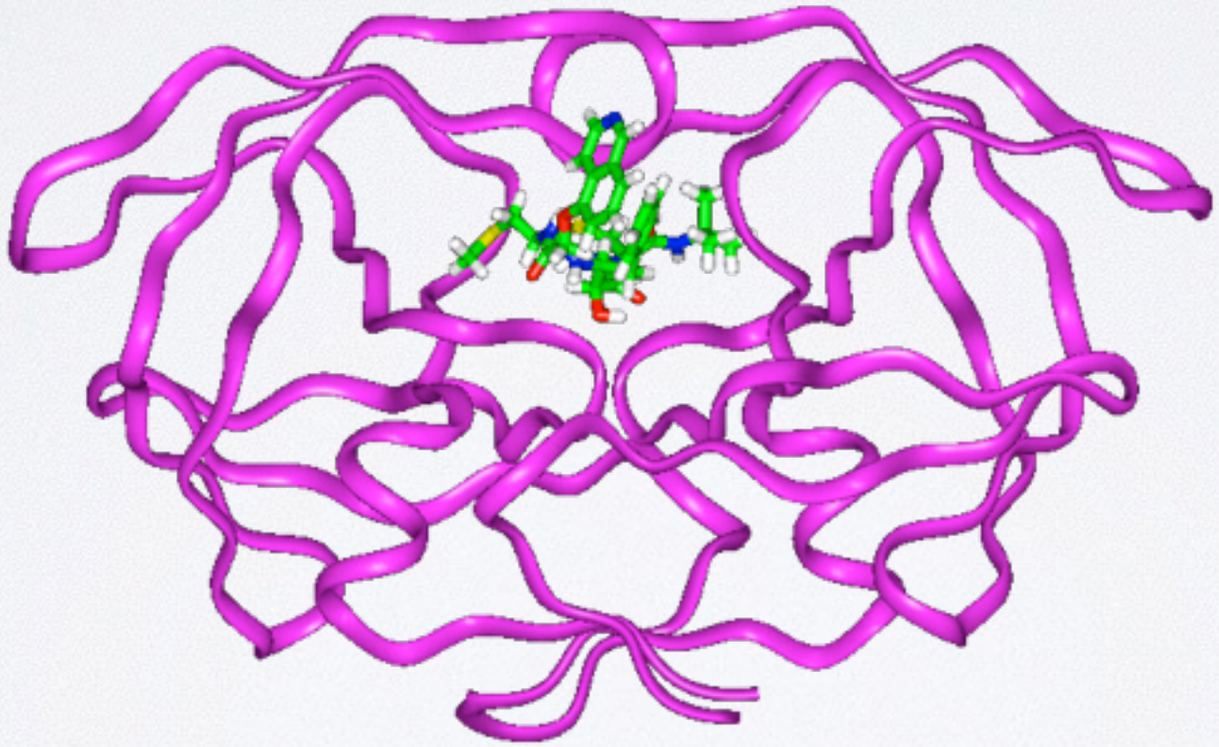
Two main approaches:

(1). **Receptor/Target-Based**

(2). **Ligand/Drug-Based**

SCENARIO I:
RECEPTOR-BASED DRUG DISCOVERY

Structure of Targeted Protein Known: Structure-Based Drug Discovery



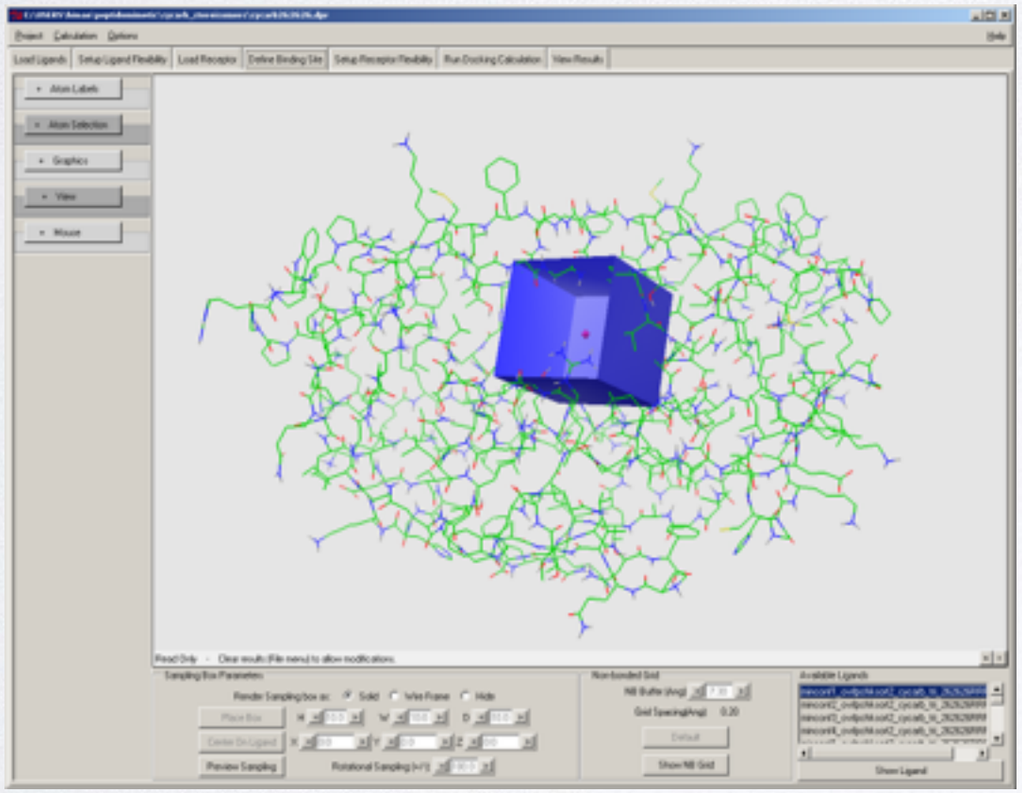
HIV Protease/KNI-272 complex

PROTEIN-LIGAND DOCKING

Structure-Based Ligand Design

Docking software

Search for structure of lowest energy

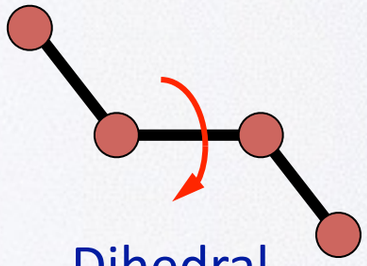


Potential function

Energy as function of structure

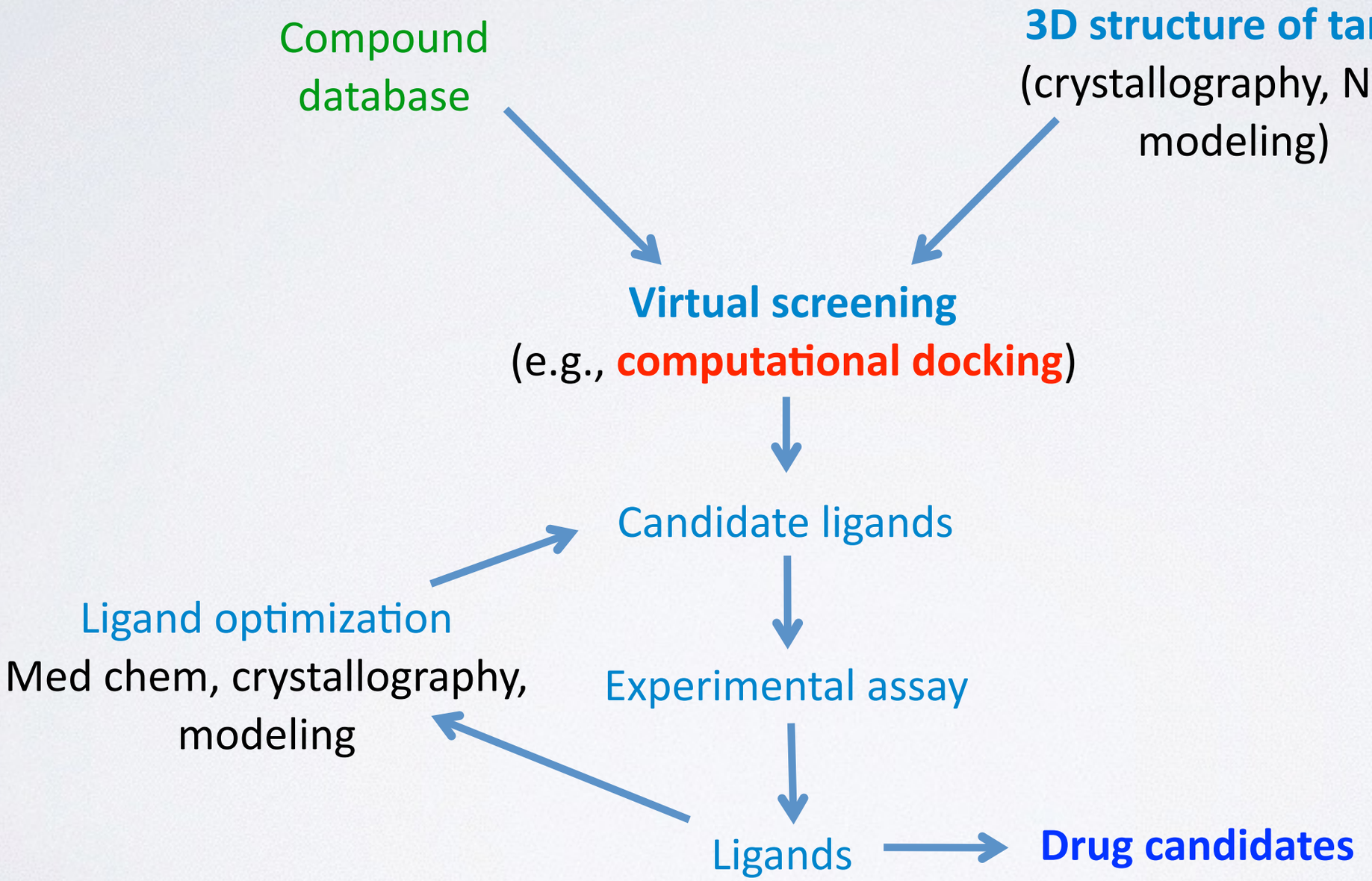
VDW

Screened Coulombic



Dihedral

STRUCTURE-BASED VIRTUAL SCREENING



COMPOUND LIBRARIES

Commercial
(in-house pharma)

Government (NIH)

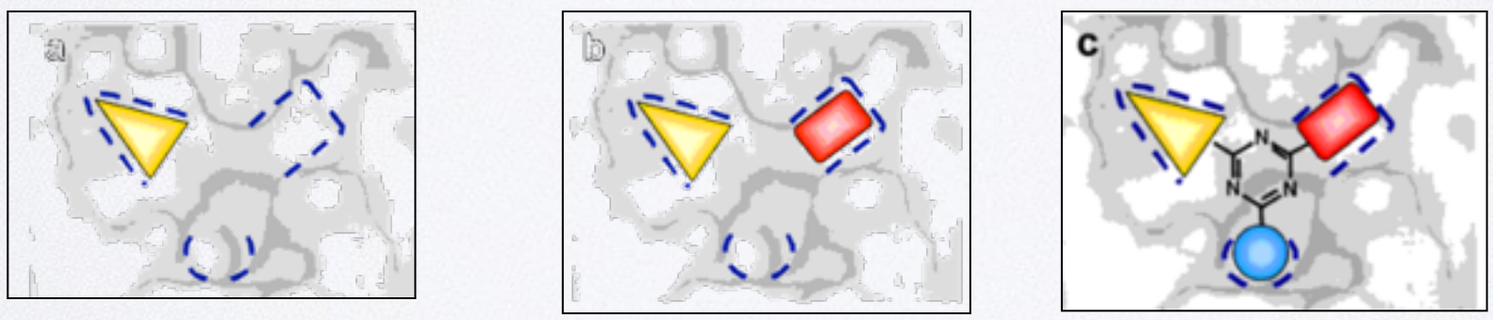
Academia

FRAGMENTAL STRUCTURE-BASED SCREENING

“Fragment” library → **Fragment docking** ← **3D structure of target**

Fragment docking → Compound design → Experimental assay and ligand optimization → **Drug candidates**

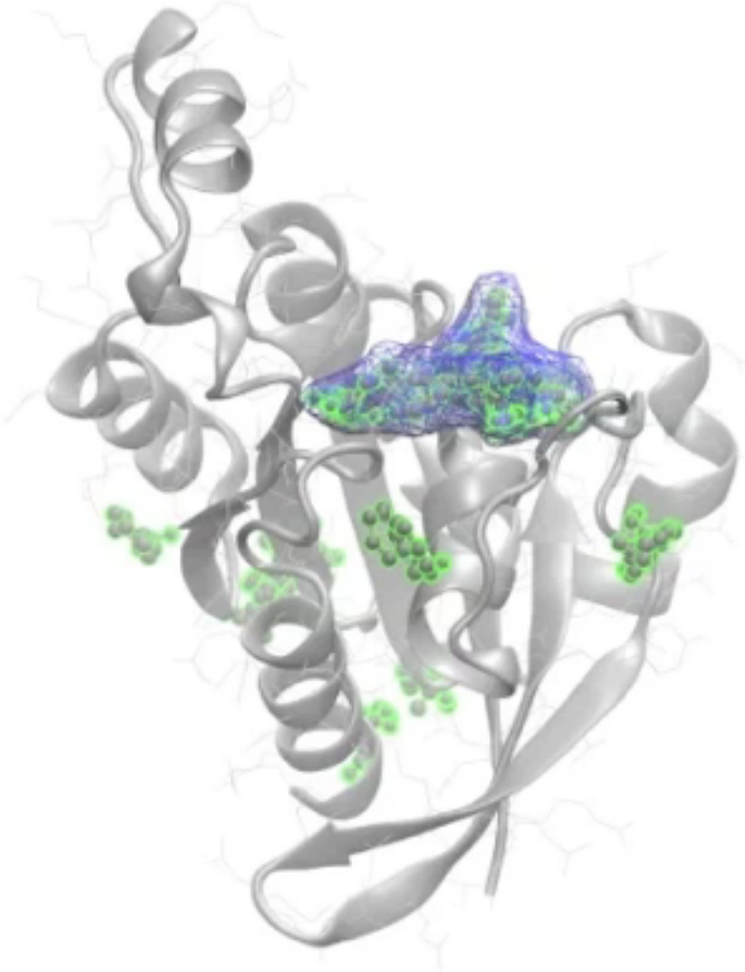
Med chem, crystallography, modeling

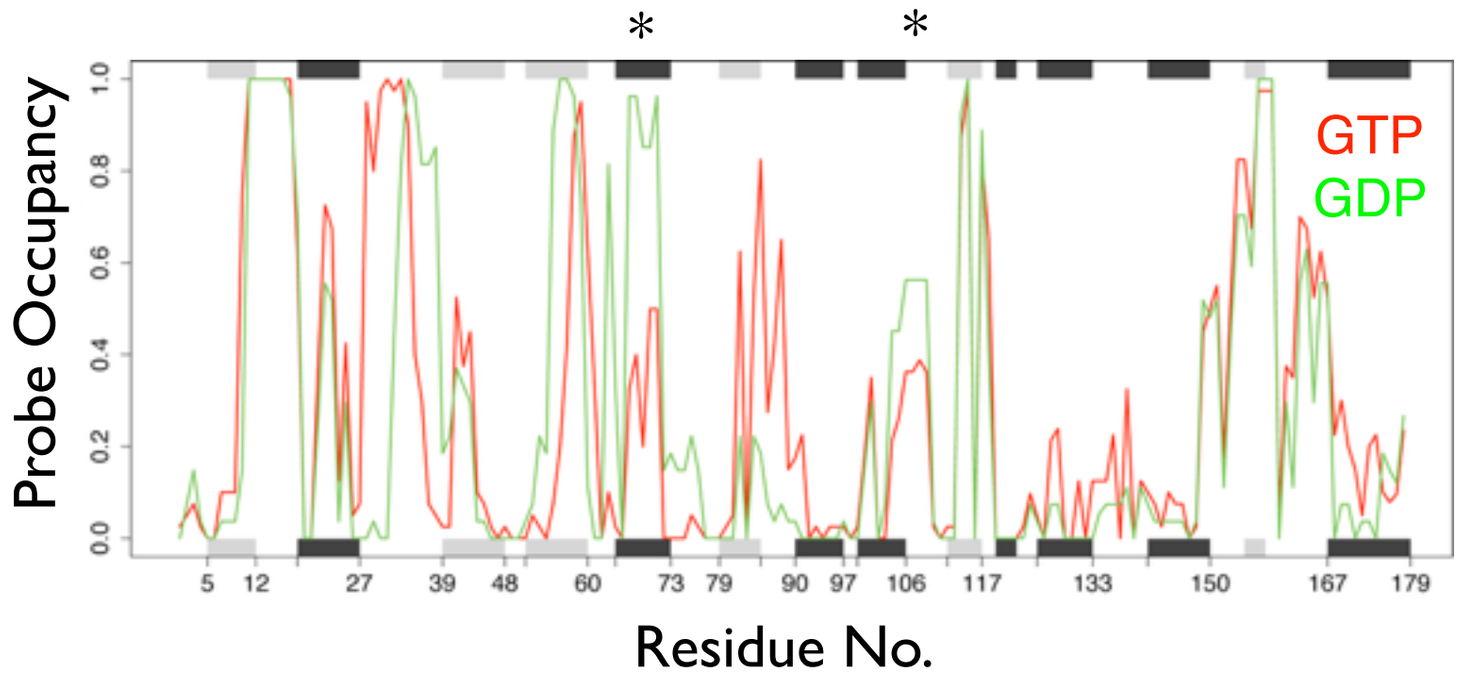


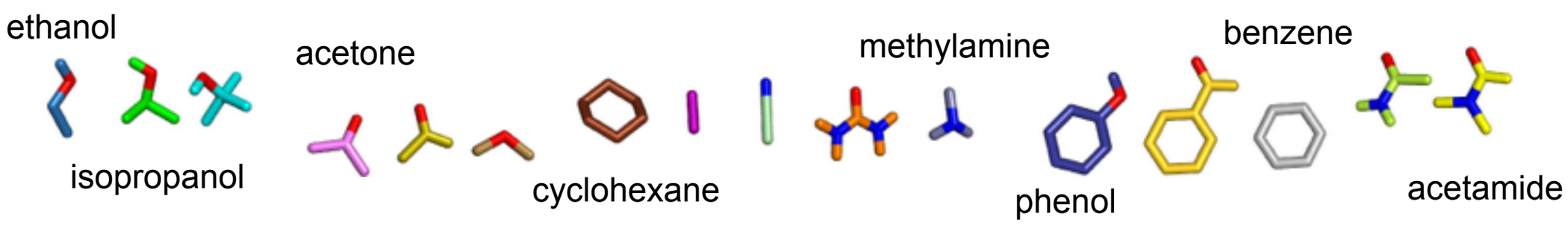
http://www.beilstein-institut.de/bozen2002/proceedings/Jhoti/jhoti.html

Multiple non active-site pockets identified

Small organic probe fragment affinities map multiple potential binding sites across the structural ensemble.



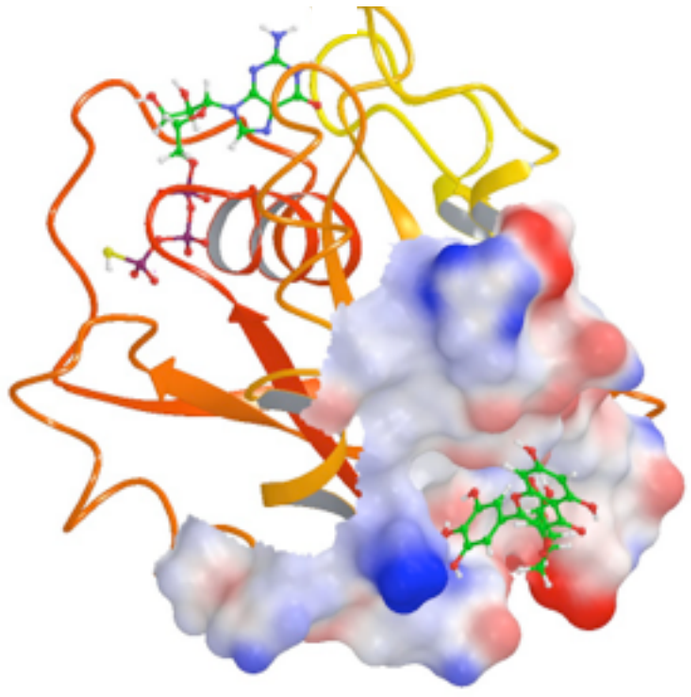


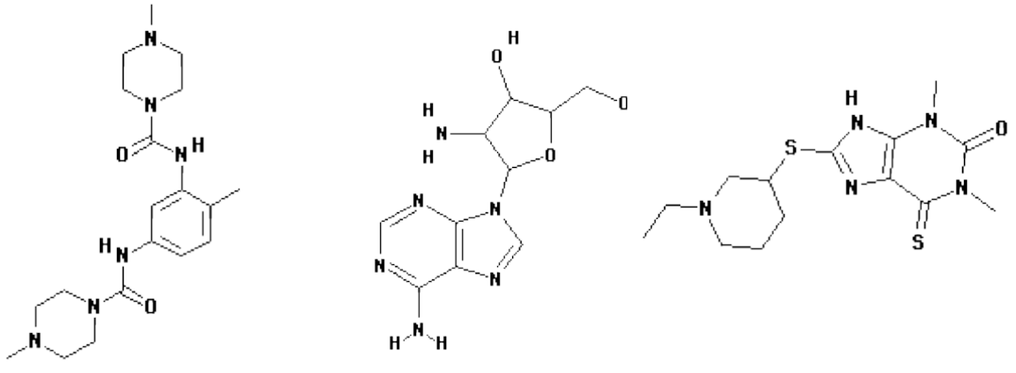


Ensemble docking & candidate inhibitor testing

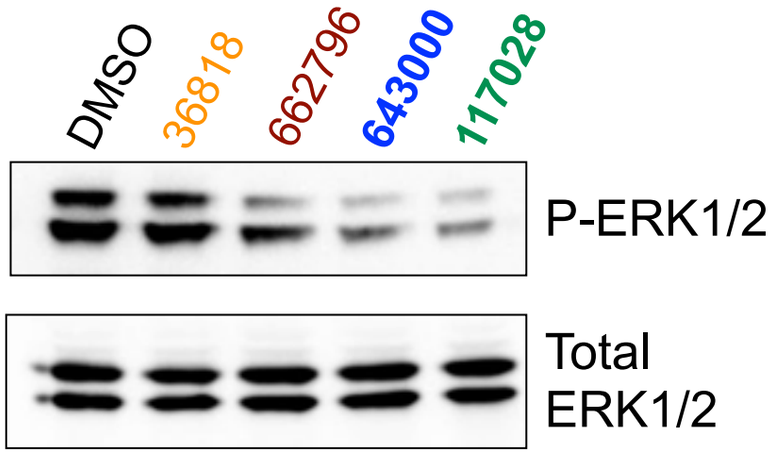
Top hits from ensemble docking against distal pockets were tested for inhibitory effects on basal ERK activity in glioblastoma cell lines.

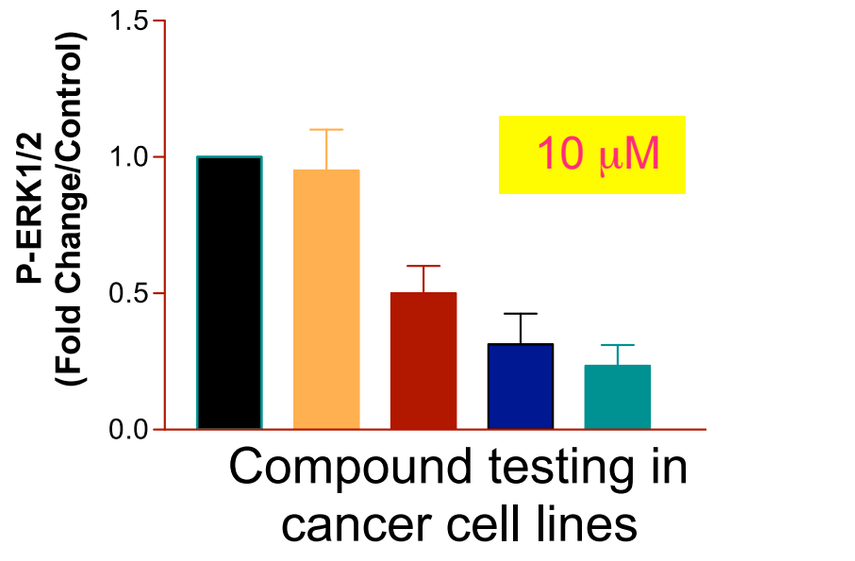
Ensemble computational docking





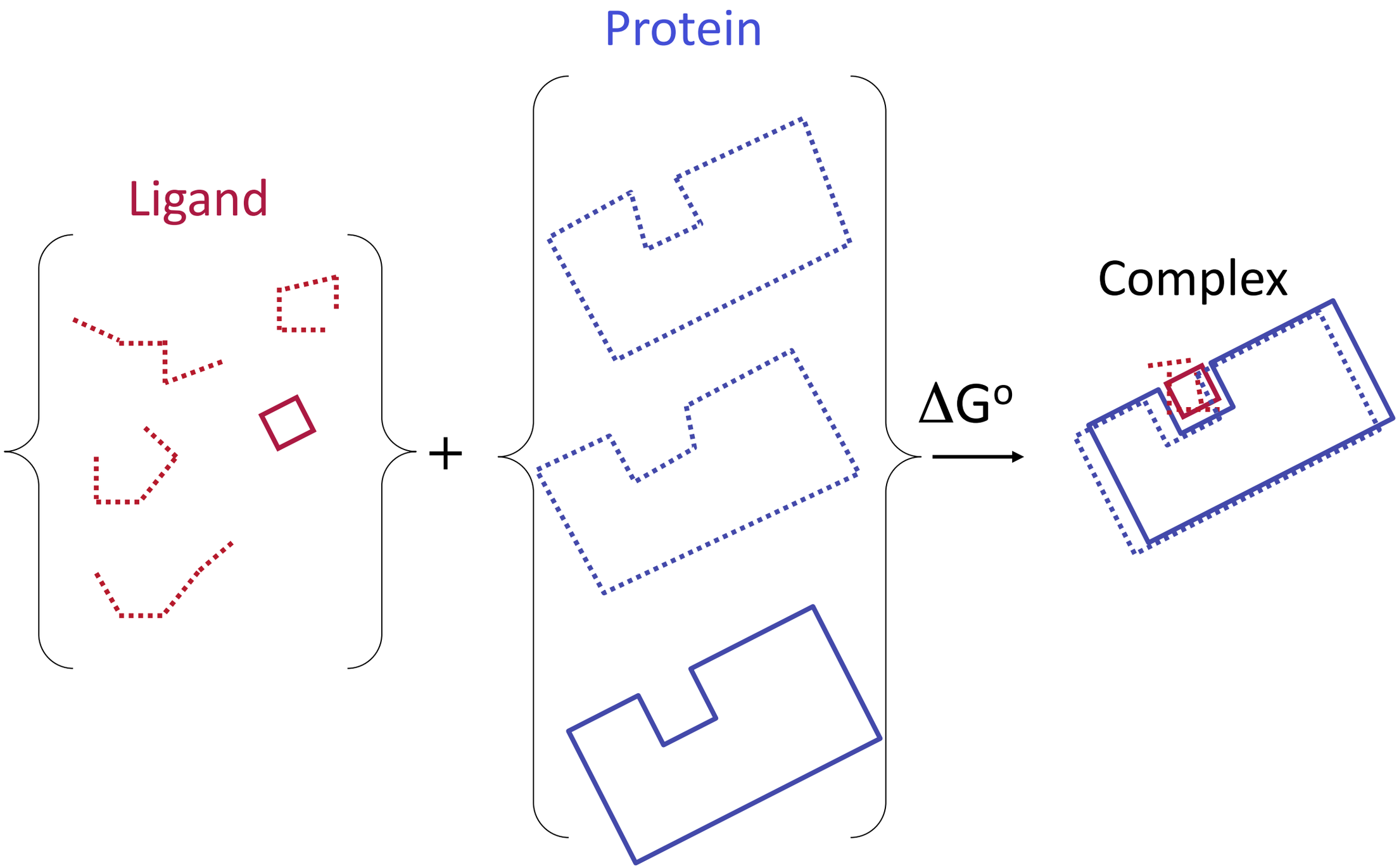
Compound effect on U251 cell line





PLoS One (2011, 2012)

Proteins and Ligand are Flexible



COMMON SIMPLIFICATIONS USED IN PHYSICS-BASED DOCKING

Quantum effects approximated classically

Protein often held rigid

Configurational entropy neglected

Influence of water treated crudely

Two main approaches:

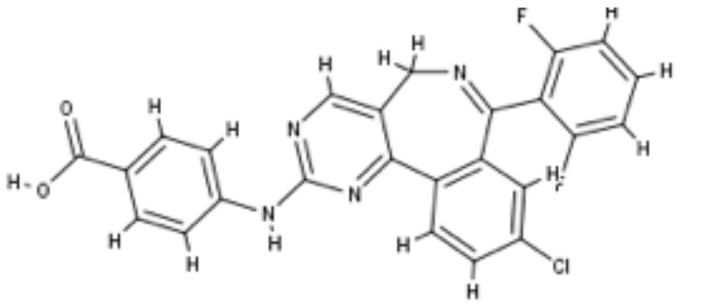
(1). **Receptor/Target-Based**

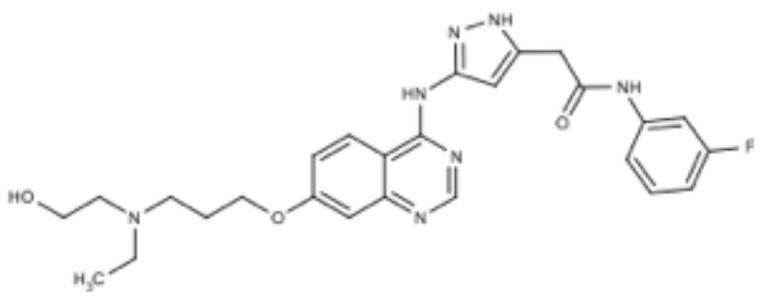
(2). **Ligand/Drug-Based**

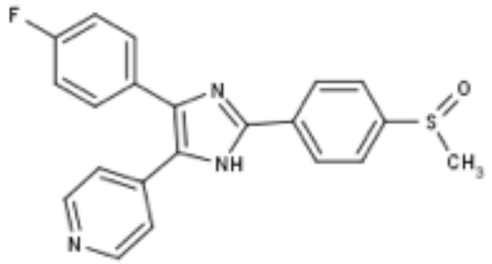
Scenario 2

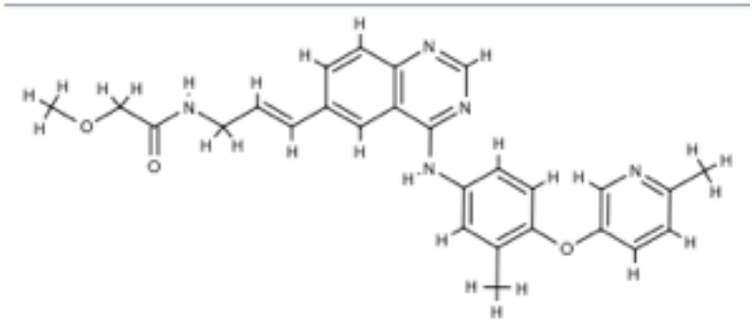
Structure of Targeted Protein Unknown: Ligand-Based Drug Discovery

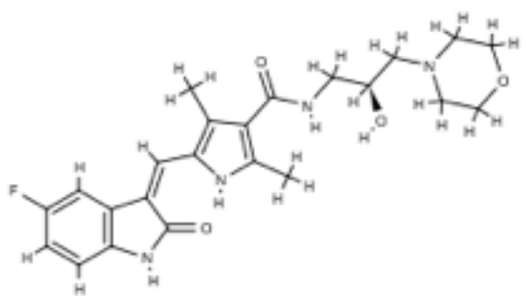
e.g. MAP Kinase Inhibitors

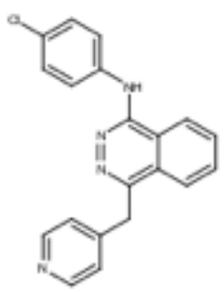


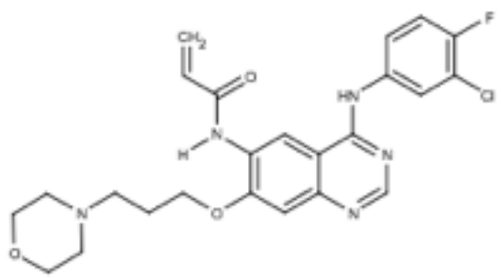


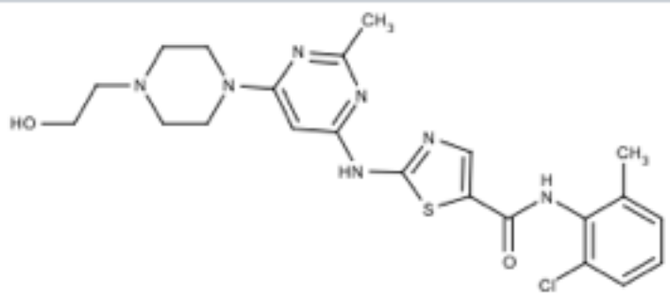












Using knowledge of existing inhibitors to discover more

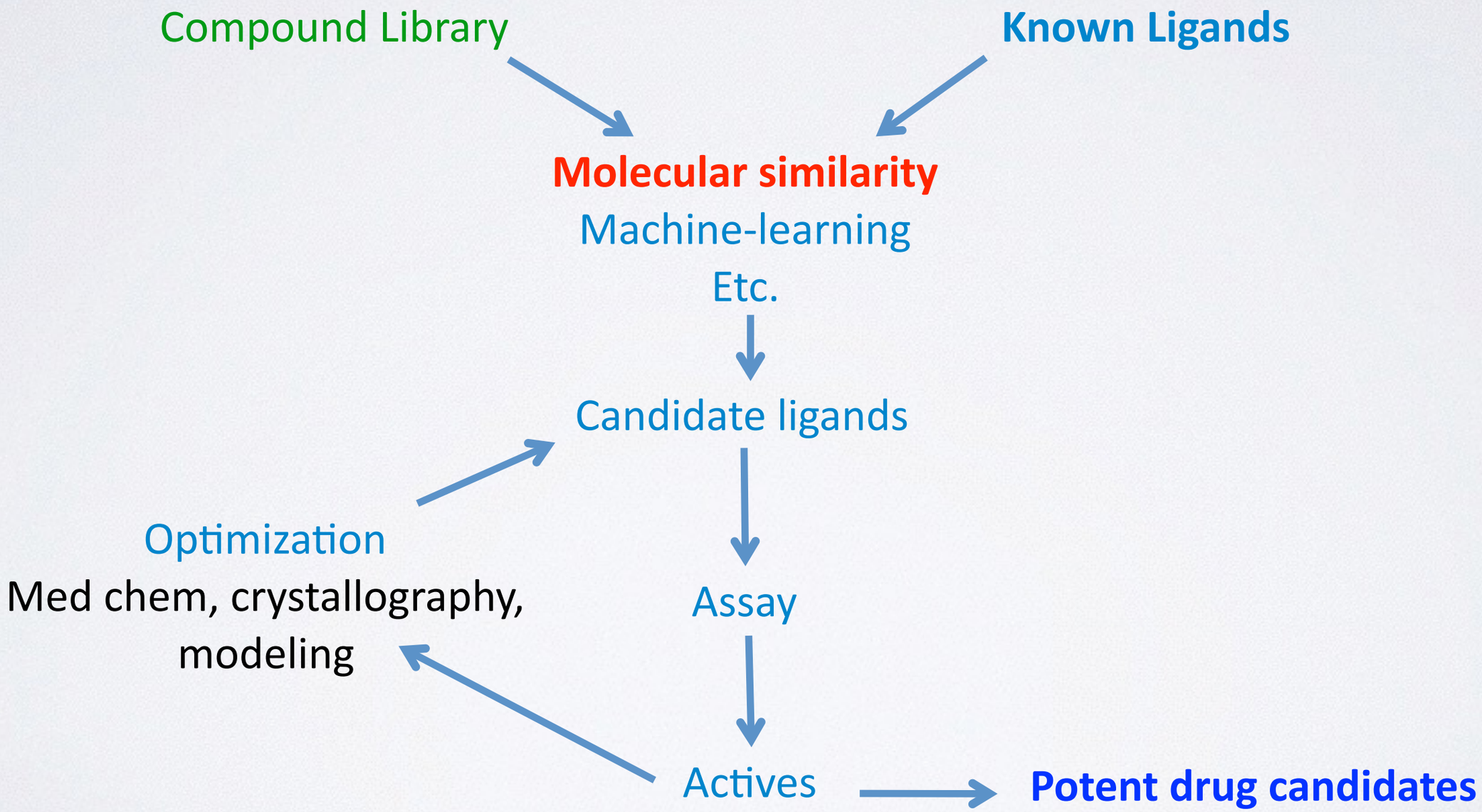
Why Look for Another Ligand if You Already Have Some?

Experimental screening generated some ligands, but they don’t bind tightly

A company wants to work around another company’s chemical patents

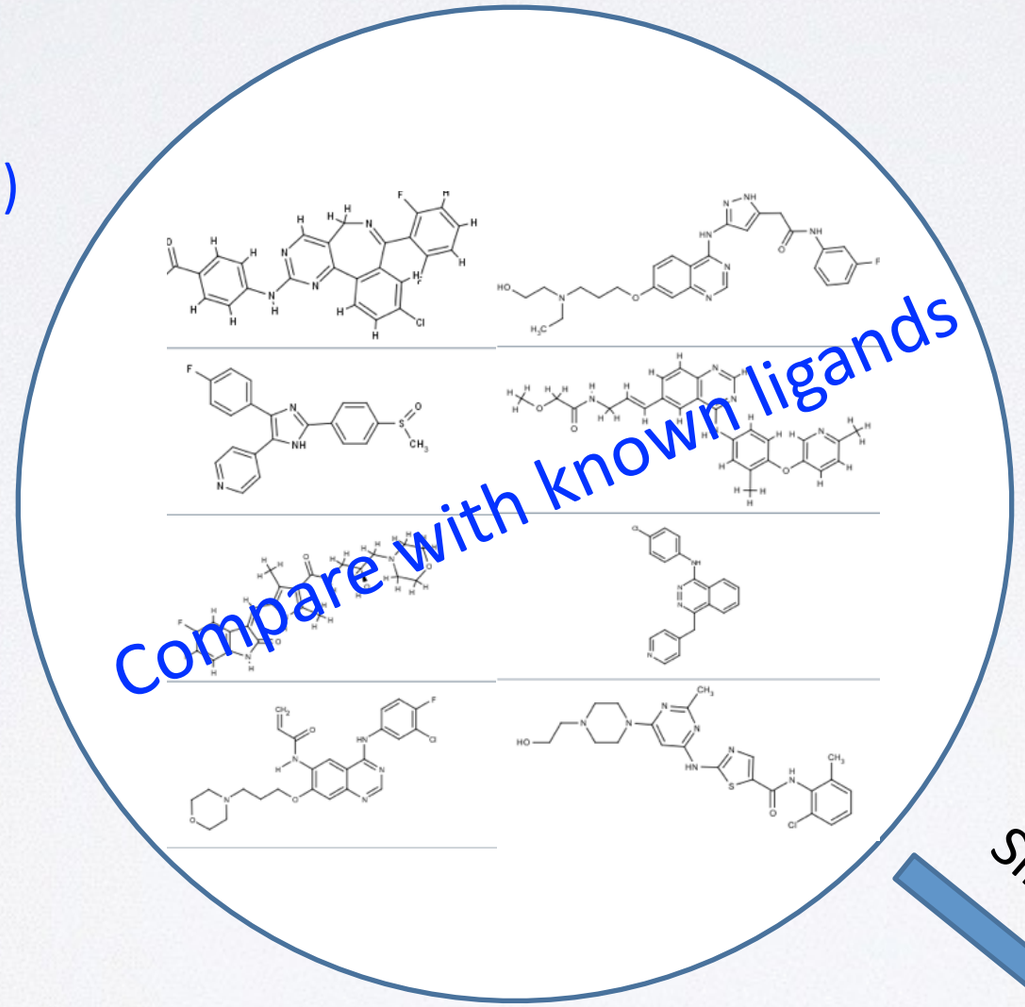
An high-affinity ligand is toxic, is not well-absorbed, etc.

LIGAND-BASED VIRTUAL SCREENING



CHEMICAL SIMILARITY
LIGAND-BASED DRUG-DISCOVERY

Compounds
(available/synthesizable)



Different

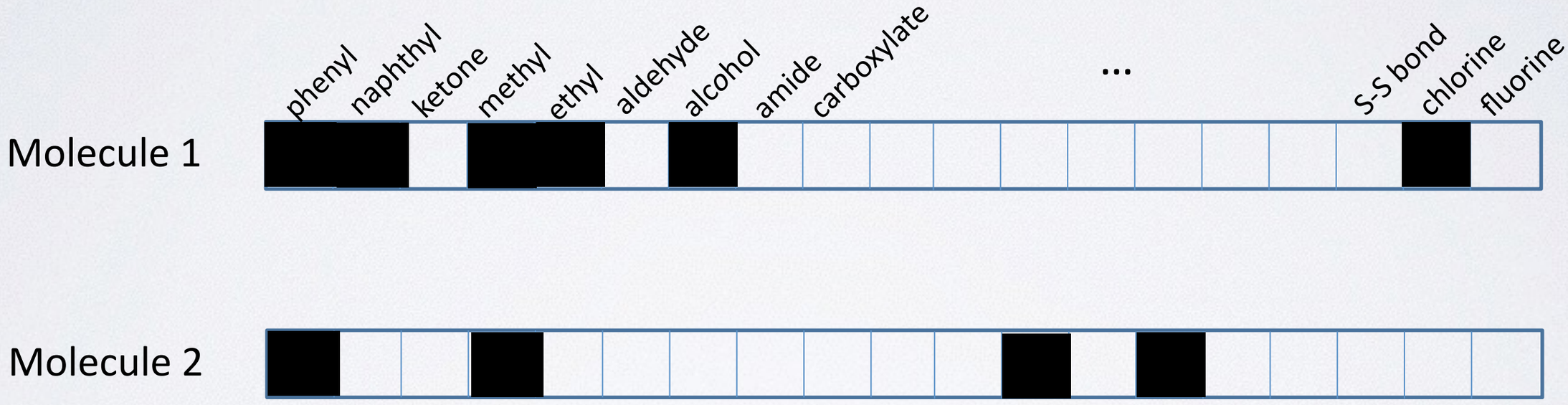
Don’t bother

Similar

Test experimentally

CHEMICAL FINGERPRINTS

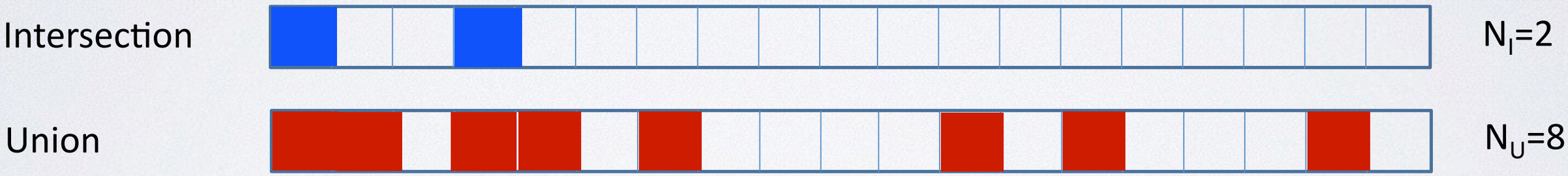
BINARY STRUCTURE KEYS



CHEMICAL SIMILARITY FROM FINGERPRINTS

Tanimoto Similarity or Jaccard Index, T

$$T \equiv \frac{N_I}{N_U} = 0.25$$



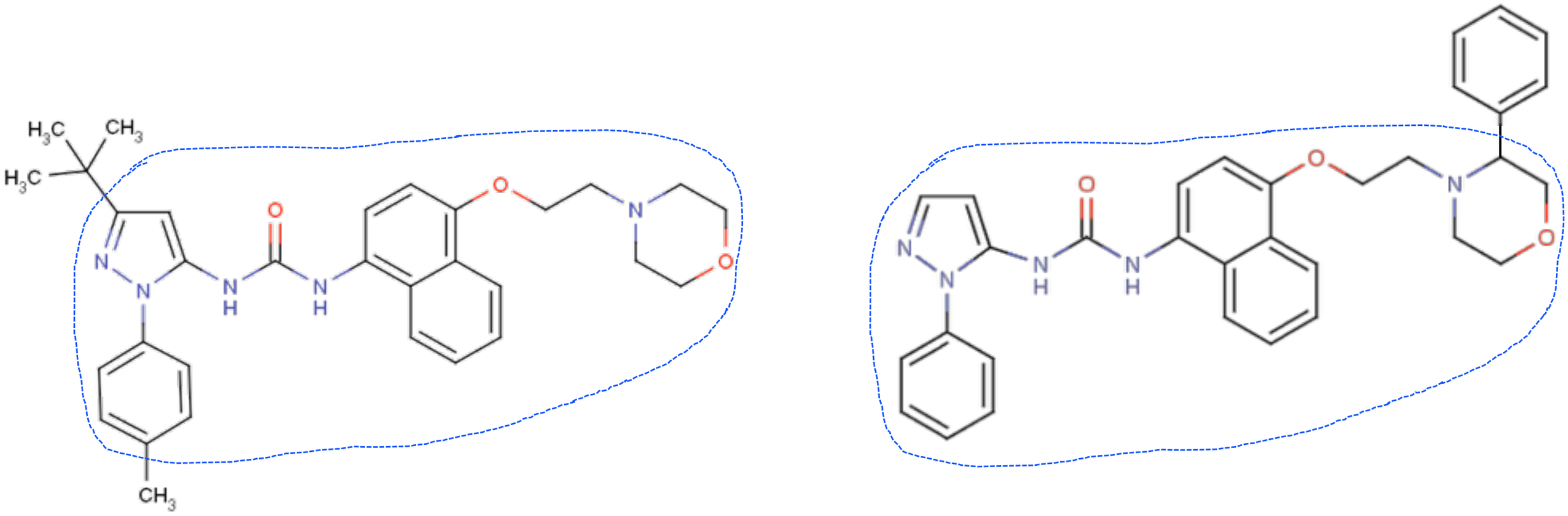
POTENTIAL DRAWBACKS OF PLAIN CHEMICAL SIMILARITY

May miss good ligands by being overly conservative

May put too much weight on irrelevant details

- Examine ligand shape and common substructures
- Build pharmacophore models
- Statistics and machine learning on chemical descriptors

Maximum Common Substructure

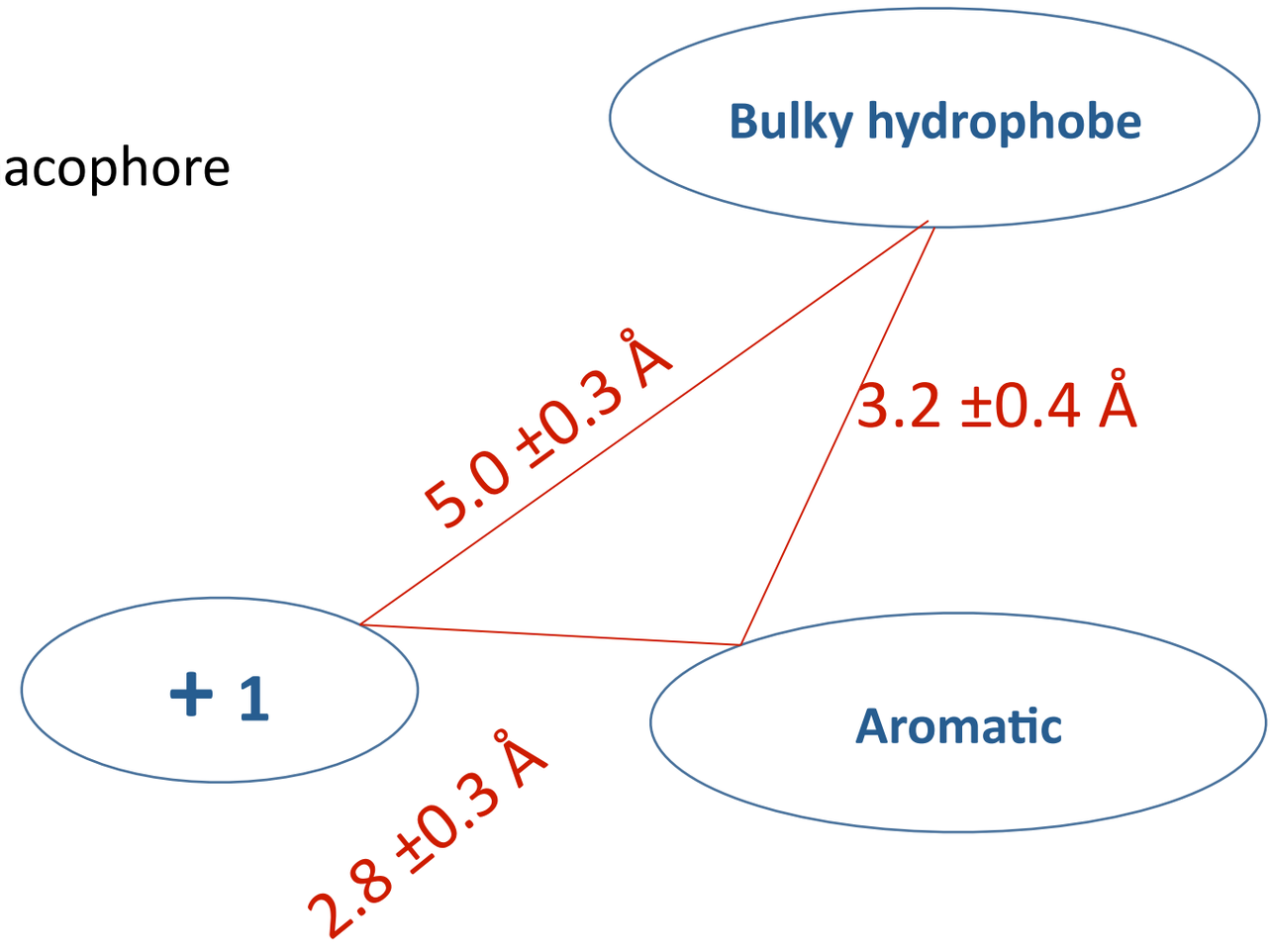


$N_{common}=34$

Pharmacophore Models

Φάρμακο (drug) + Φορά (carry)

A 3-point pharmacophore

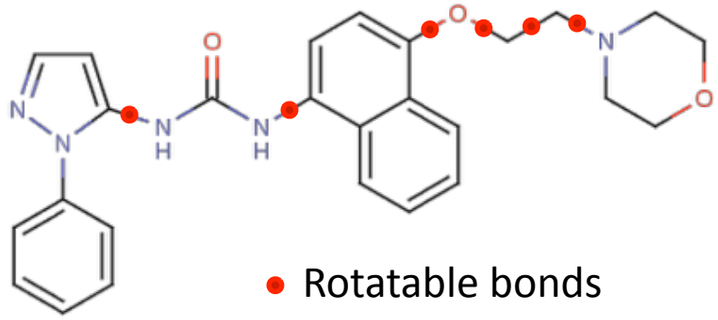


Molecular Descriptors

More abstract than chemical fingerprints

Physical descriptors

- molecular weight
- charge
- dipole moment
- number of H-bond donors/acceptors
- number of rotatable bonds
- hydrophobicity (log P and clogP)



Topological

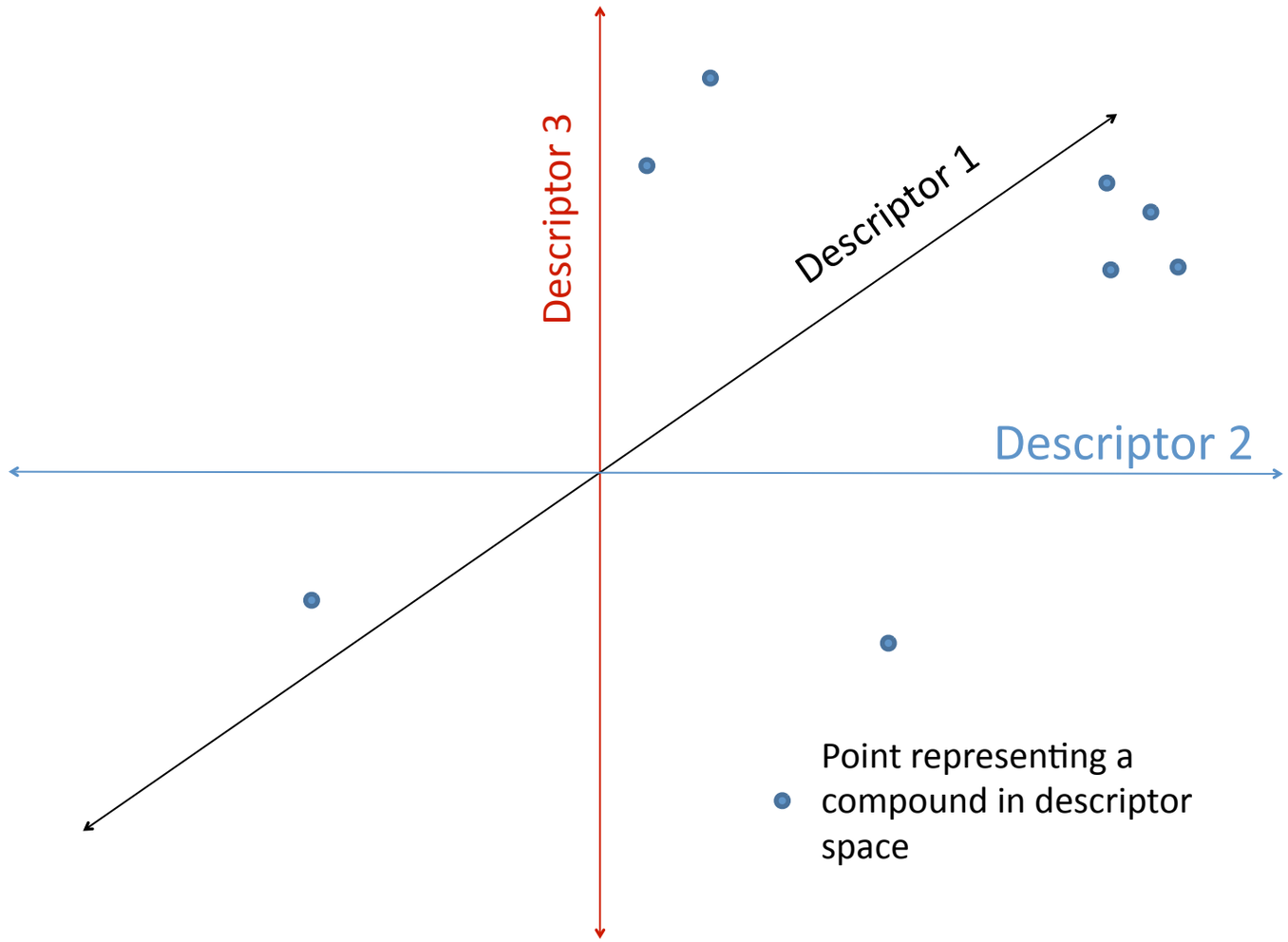
- branching index
- measures of linearity vs interconnectedness

Etc. etc.

A High-Dimensional "Chemical Space"

Each compound is at a point in an n-dimensional space

Compounds with similar properties are near each other



Apply **multivariate statistics** and **machine learning** for descriptor-selection.
(e.g. partial least squares, support vector machines, random forest, etc.)

CAUTIONARY NOTES

- **"Everything should be made as simple as it can be but not simpler"**

  A model is **never perfect**. A model that is not quantitatively accurate in every respect does not preclude one from establishing results relevant to our understanding of biomolecules as long as the biophysics of the model are properly understood and explored.

- **Calibration of the parameters is an ongoing and imperfect process**

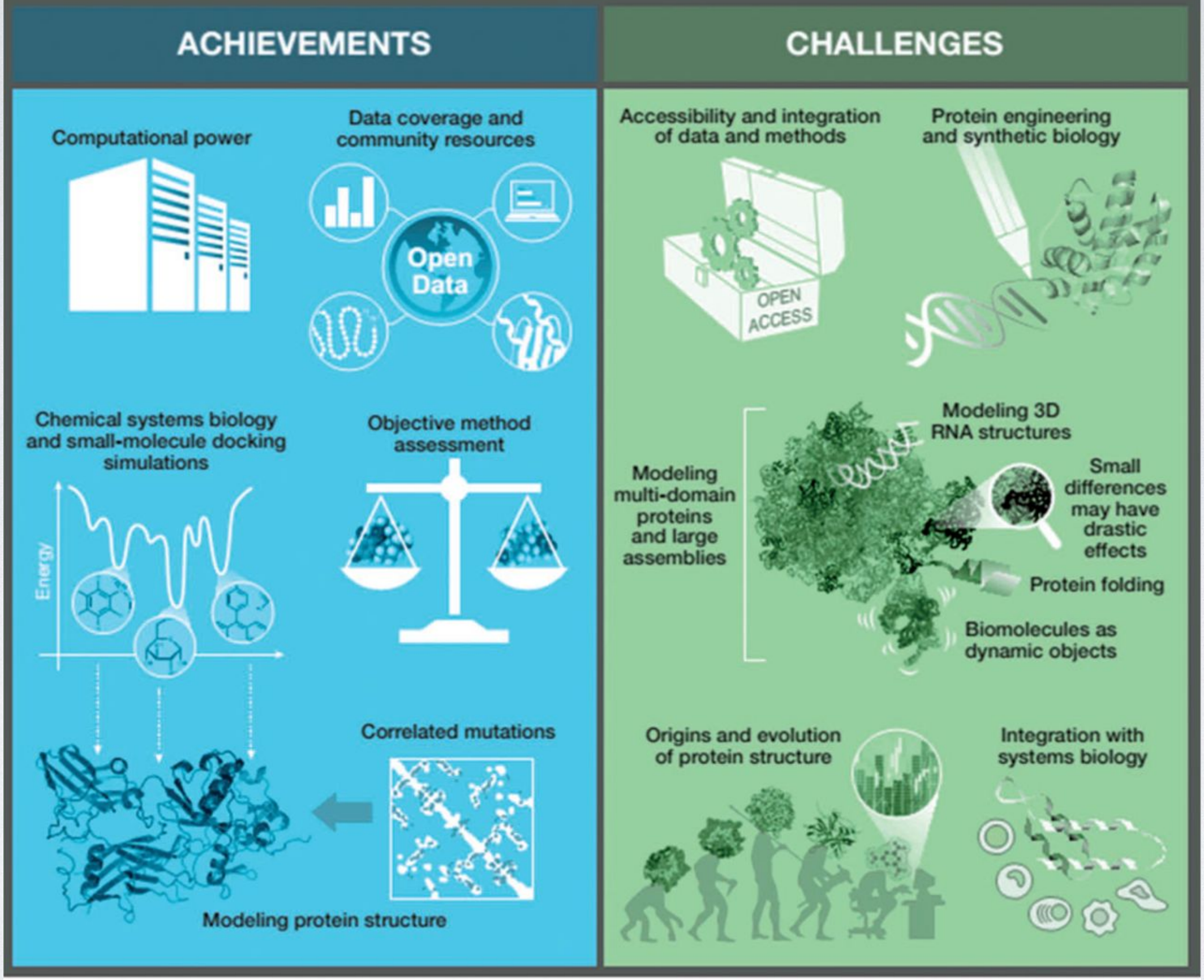
  Questions and hypotheses should always be designed such that they do not depend crucially on the precise numbers used for the various parameters.

- **A computational model is rarely universally right or wrong**

  A model may be accurate in some regards, inaccurate in others. These subtleties can only be uncovered by comparing to all available experimental data.

SUMMARY

- Structural bioinformatics is computer aided structural biology
- Described major motivations, goals and challenges of structural bioinformatics
- Reviewed the fundamentals of protein structure
- Introduced both physics and knowledge based modeling approaches for describing the structure, energetics and dynamics of proteins computationally



Ilan Samish et al. Bioinformatics 2015;31:146-150

INFORMING SYSTEMS BIOLOGY?