



BGGN 213

Structural Bioinformatics II

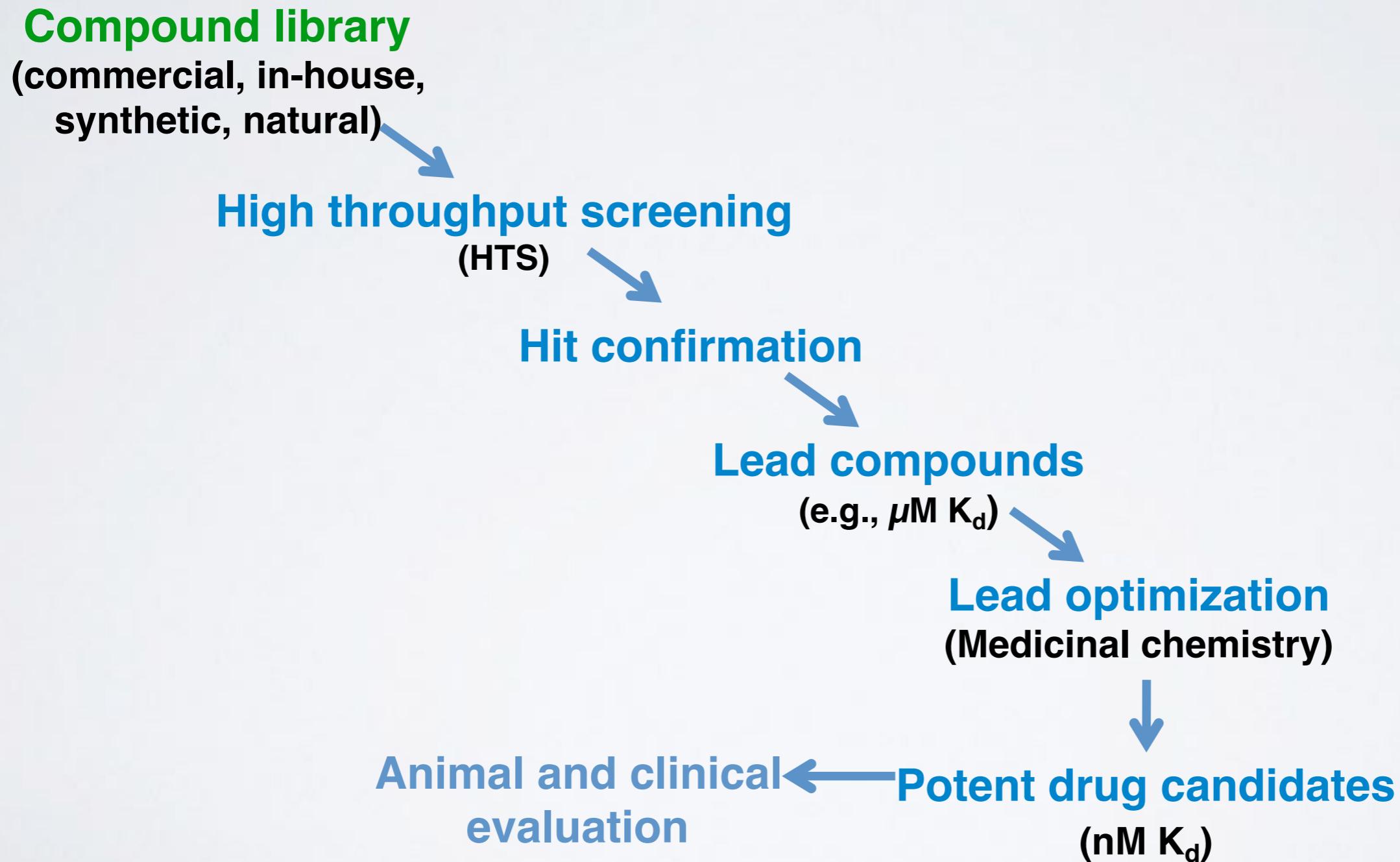
Barry Grant
UC San Diego

<http://thegrantlab.org/bggn213>

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



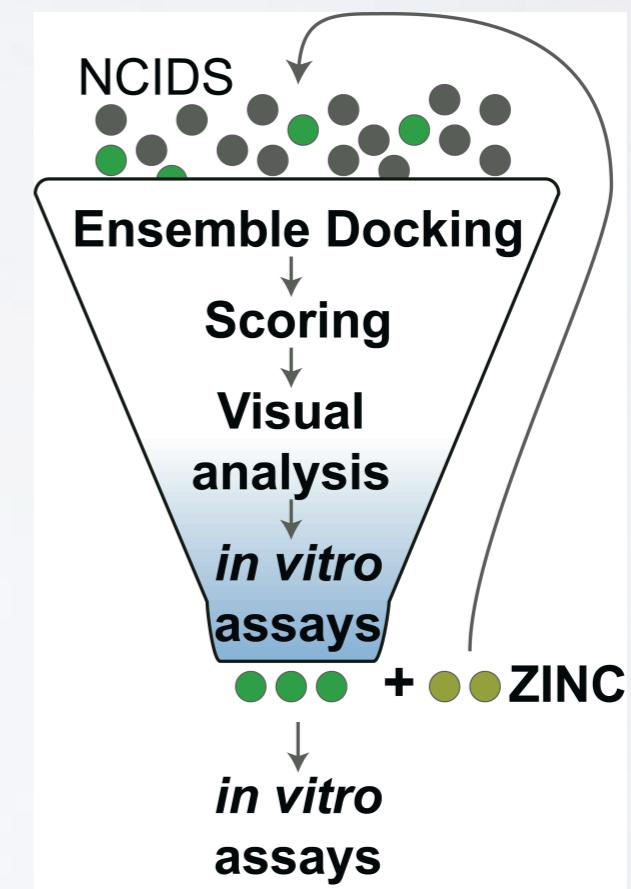
COMPUTER-AIDED LIGAND DESIGN

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**

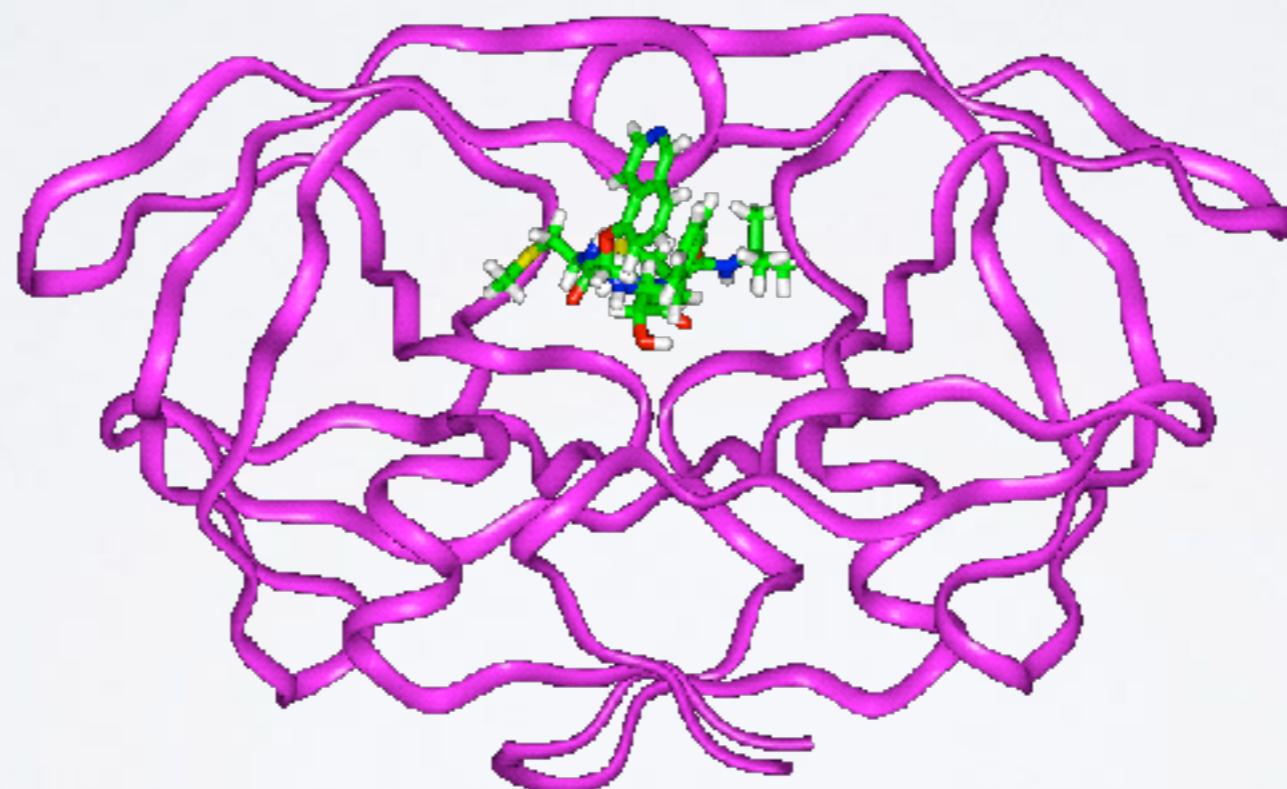
Two main approaches:

- (1). Receptor/Target-Based**
- (2). Ligand/Drug-Based**

SCENARIO I:

RECEPTOR-BASED DRUG DISCOVERY

Structure of Targeted Protein Known: **Structure-Based Drug Discovery**



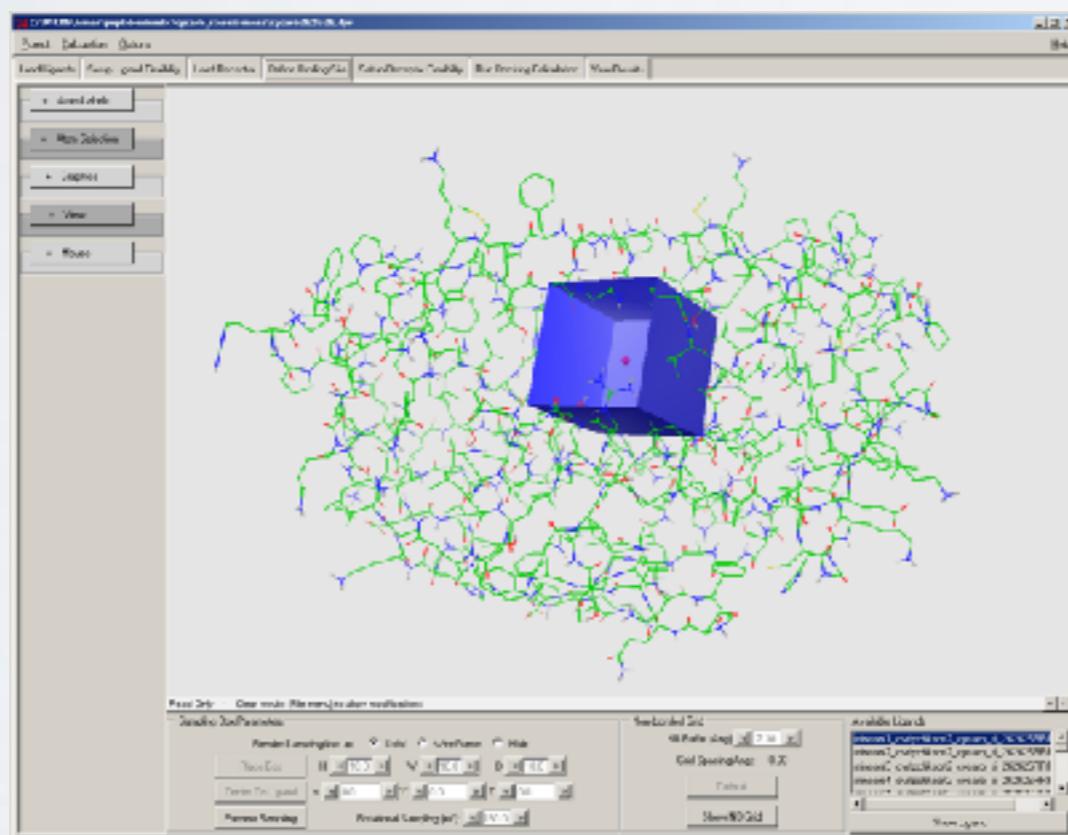
HIV Protease/KNI-272 complex

PROTEIN-LIGAND DOCKING

Structure-Based Ligand Design

Docking software

Search for structure of lowest energy



Potential function

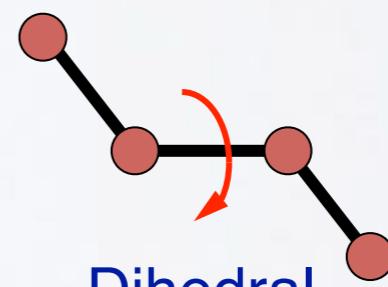
Energy as function of structure



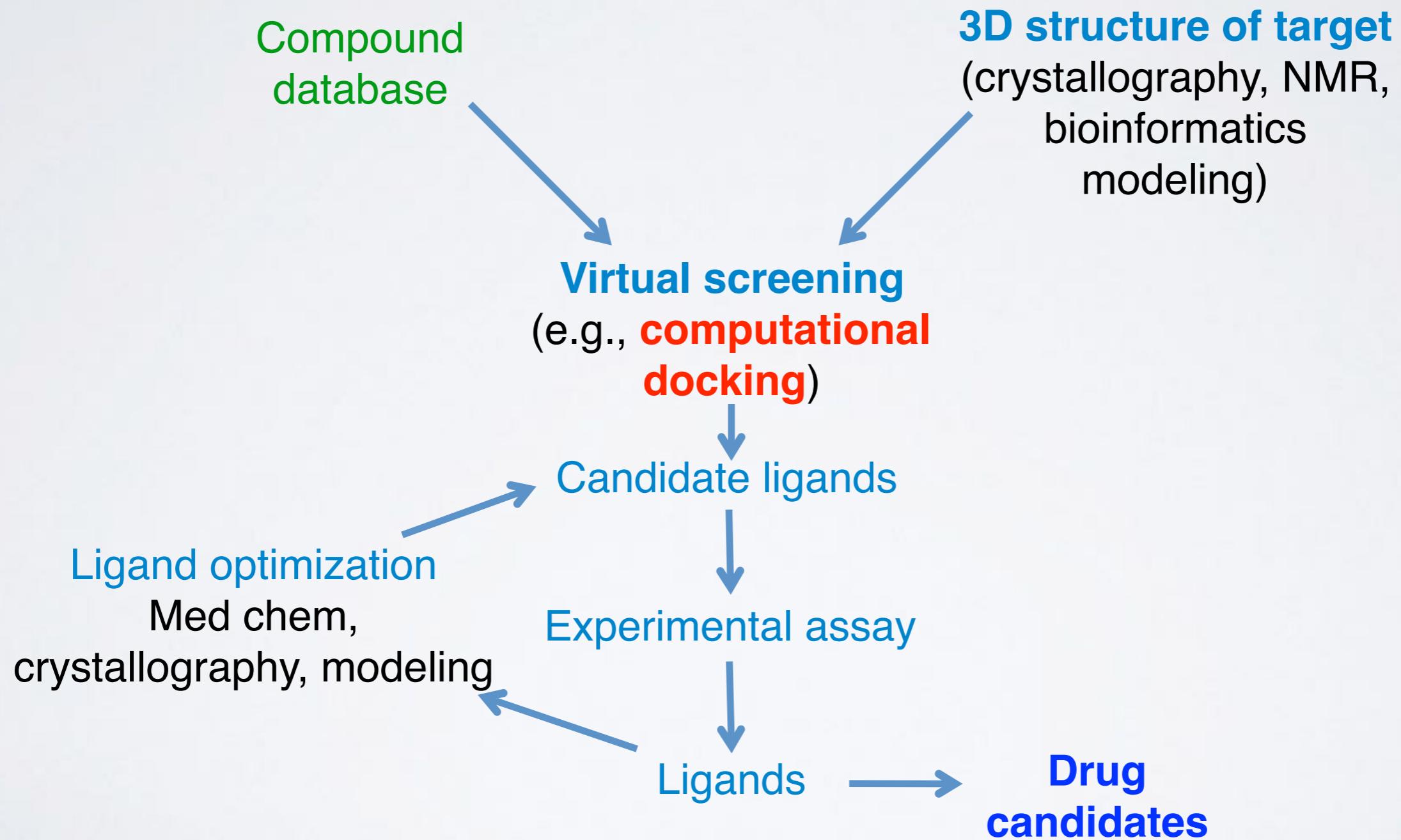
VDW



Screened Coulombic



STRUCTURE-BASED VIRTUAL SCREENING



COMPOUND LIBRARIES

The screenshot shows the Maybridge website. At the top, there's a search bar and a navigation menu with links like "HOME", "SCREENING SERVICES", "INDUSTRIAL SCREENING", "INDUSTRY", "INDUSTRY", and "CONTACT US". Below the header, there's a banner for "Maybridge HitFinder™". The main content area features a heading "Maybridge HitFinder™" and a sub-section "The pre-selected diverse screening library includes identifying potential drug leads easy, universal, and cost effective." It includes a section on "Screening quality data from your screens" and a "Ready to Screen" section with a grid of small molecule images.

The screenshot shows the NIH Molecular Libraries Small Molecule Repository website. The header includes the NIH logo and the text "NIH MOLECULAR LIBRARIES SMALL MOLECULE REPOSITORY". The main content area has a heading "A NIH Roadmap Initiative" and a "Welcome" section. It features a photograph of a scientist in a lab setting. The sidebar contains links for "Home", "MSMR Project", "MSMR Details", and "Submit Compounds". The footer includes the BioFocus logo and the text "BioFocus, a Galapagos company".

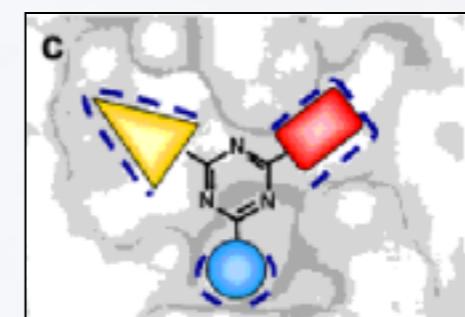
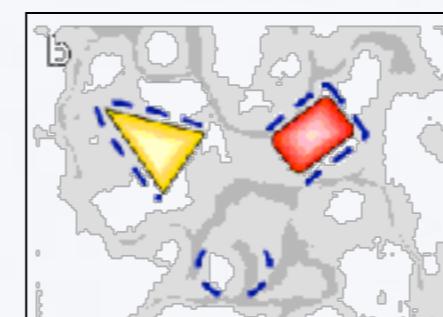
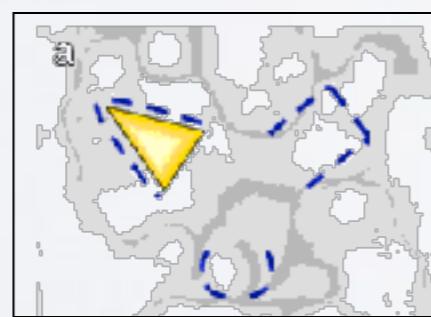
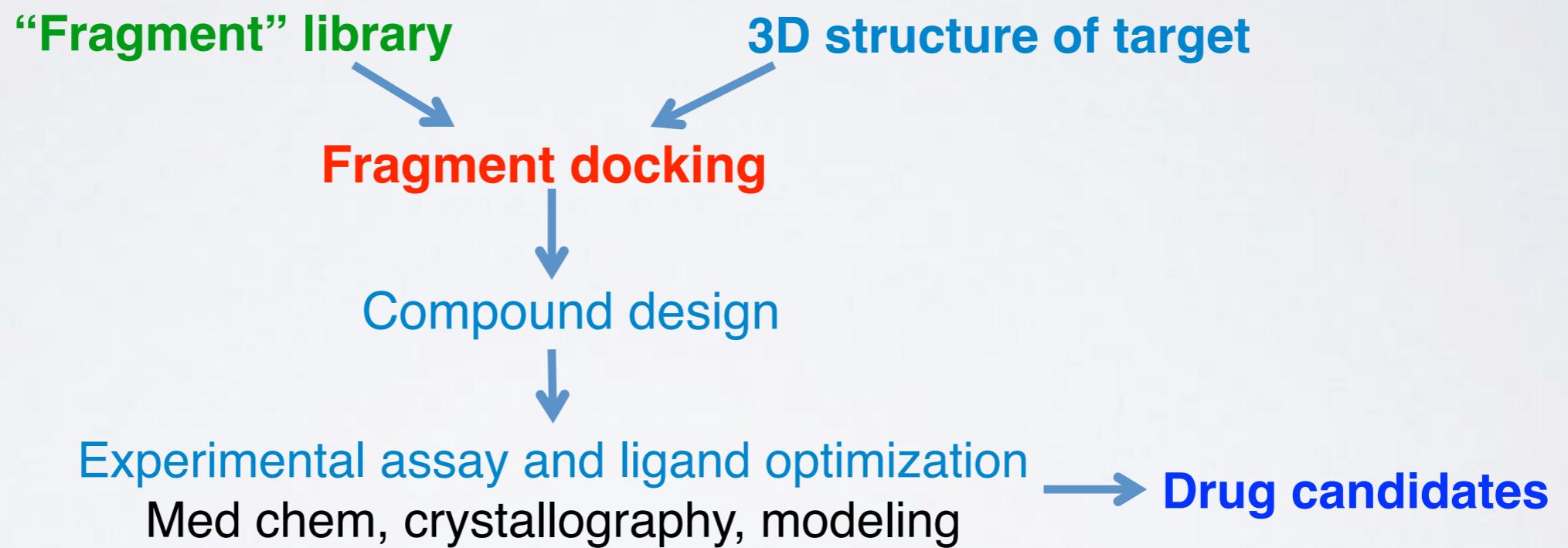
The screenshot shows the PMLSC website. The header includes the University of Pittsburgh logo and the text "University of Pittsburgh" and "Pittsburgh Molecular Libraries Screening Center". The main content area features a large image with the text "BIG DISCOVERIES FROM SMALL MOLECULES". The sidebar contains links for "HOME", "HISTORY", "PERSONNEL", "SCREENING TECHNOLOGY", "COMPONENTS", "RESEARCH & PUBLICATIONS", "LITERATURE", "ASSAY/PROTOLAB ASSAY PROTOCOLS", "PMLSC PROTO REPORTS", "LDR INDEX", "DATA ANALYSIS/INFORMATICS", "EDUCATIONAL ACTIVITIES", "MEMBERSHIPS", "LINKS", "CONTACTS", and "Corporate Search". The footer includes links for "Health Sciences & Pitt", "UPMC", "PSU", "School of Medicine", "Health Sciences Calendar", "Our News & Events", "Up or page | home | contact us", and "© 2006 by the Center for the Molecular Sciences, University of Pittsburgh. All rights reserved".

Commercial
(in-house pharma)

Government (NIH)

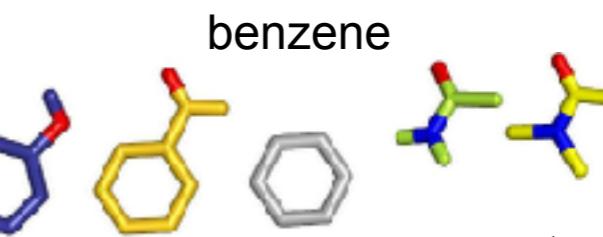
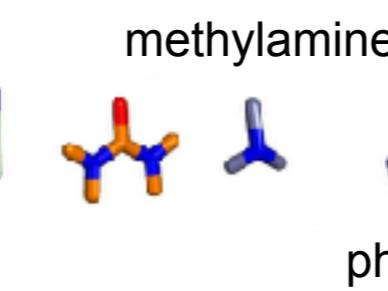
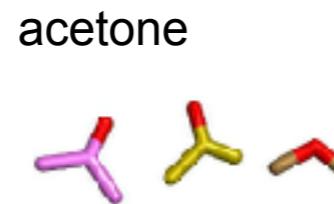
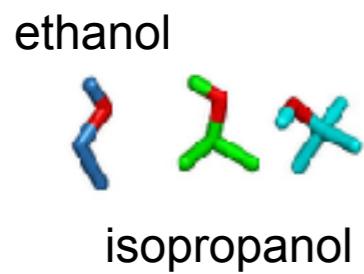
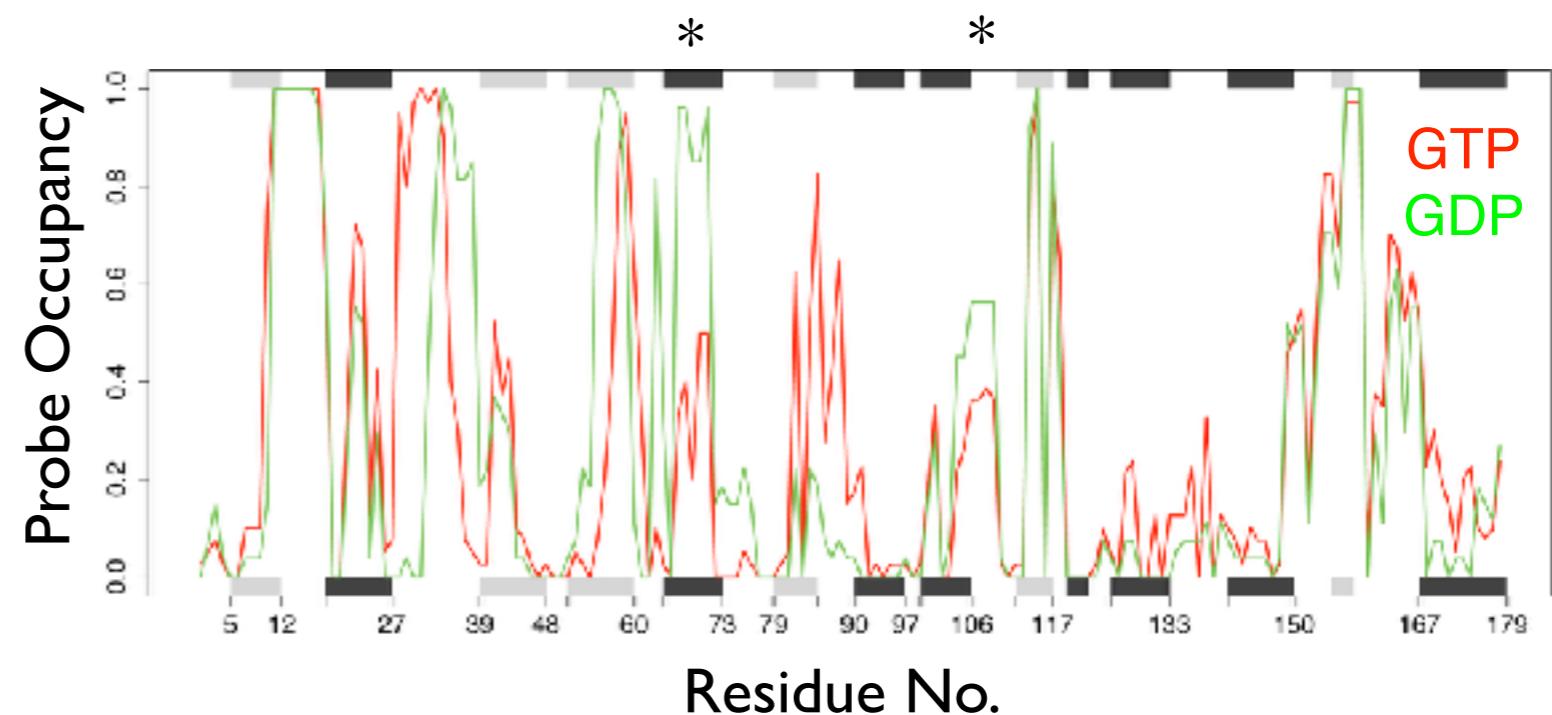
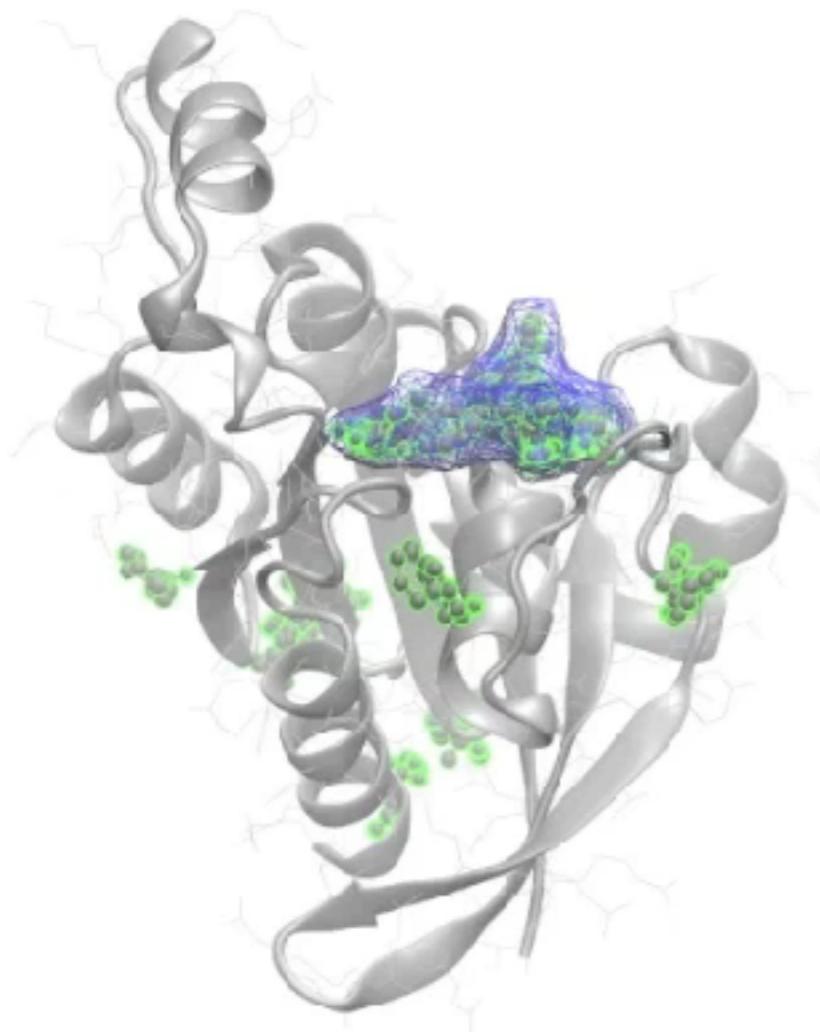
Academia

FRAGMENTAL STRUCTURE-BASED SCREENING



Multiple non active-site pockets identified

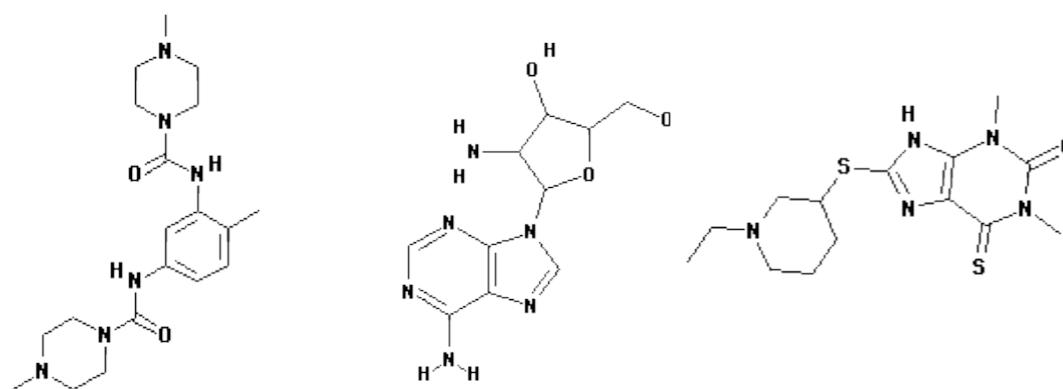
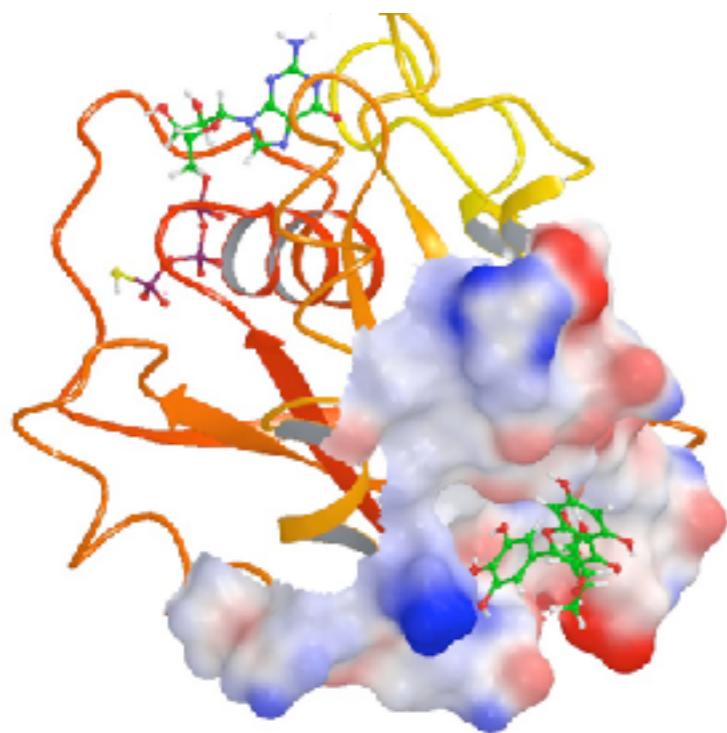
Small organic probe fragment affinities map multiple potential binding sites across the structural ensemble.



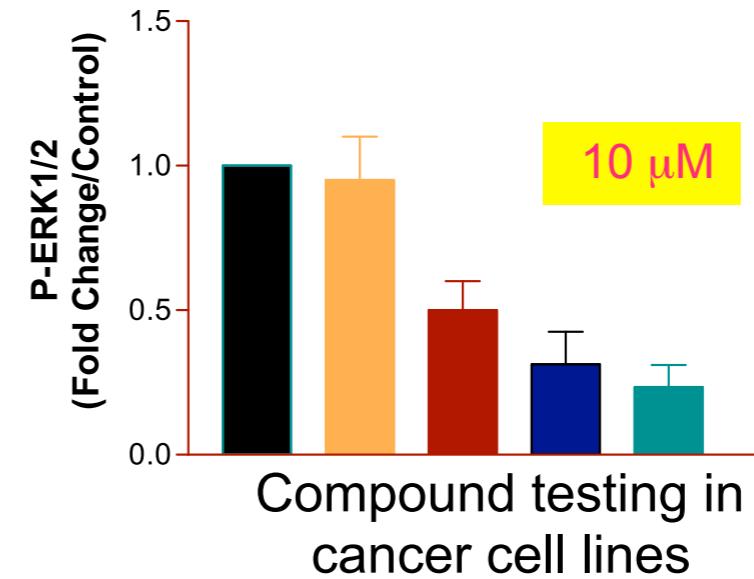
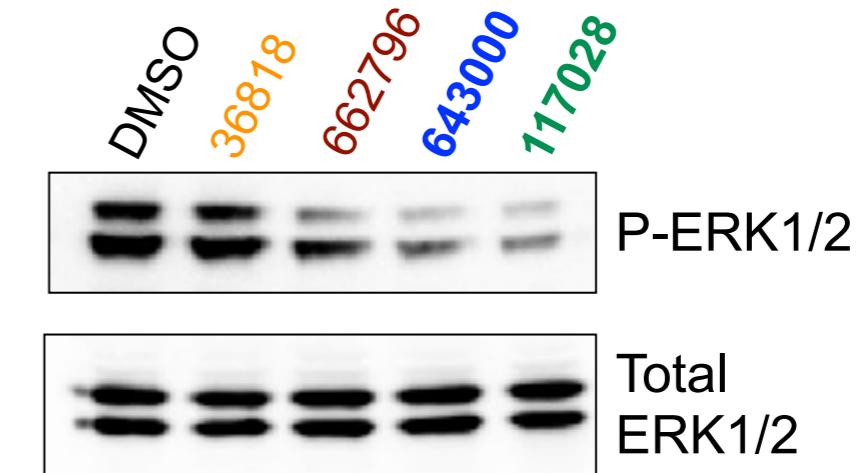
Ensemble docking & candidate inhibitor testing

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

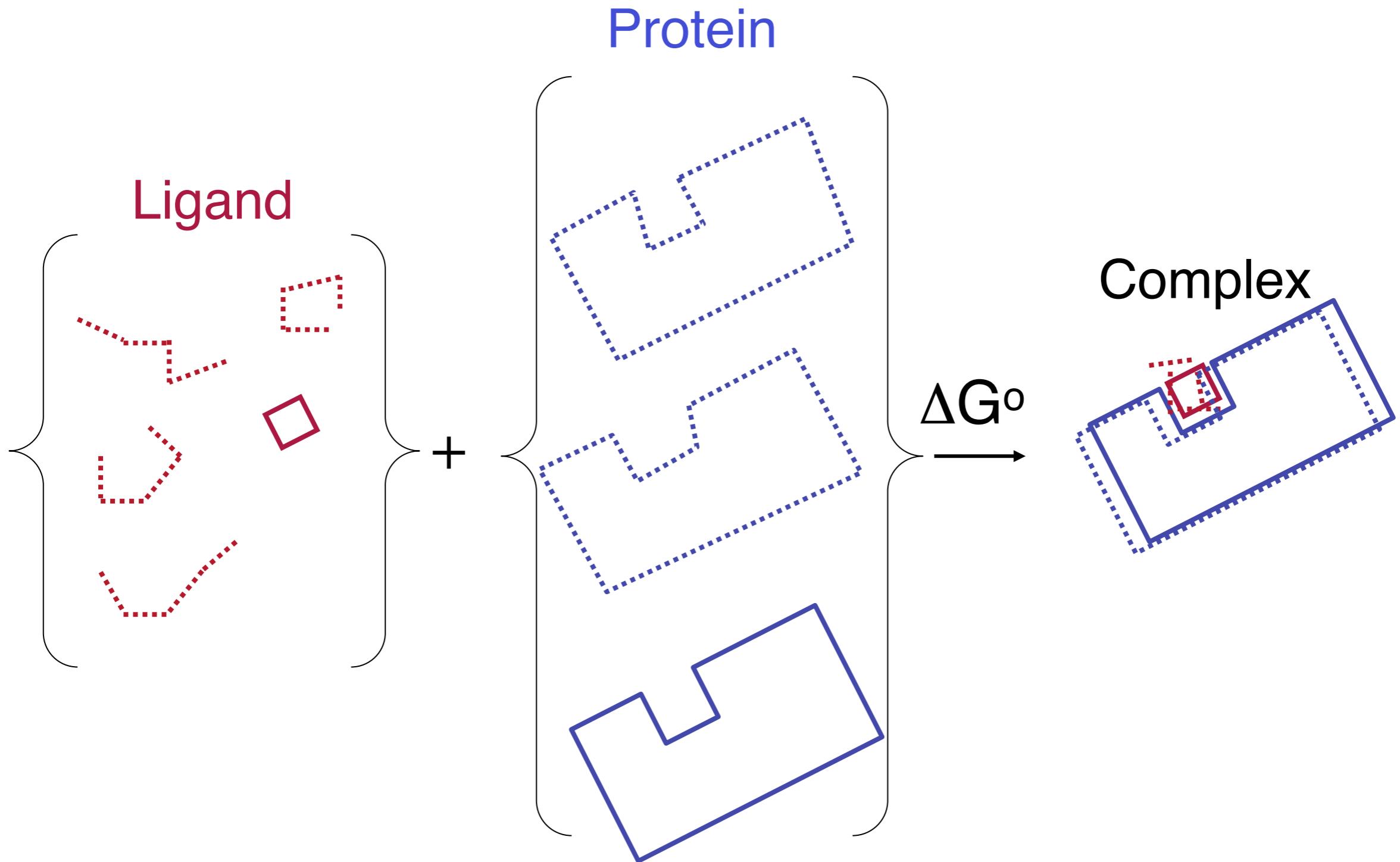
Ensemble computational docking



Compound effect on U251 cell line



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**

Do it Yourself!

Hand-on time!

https://bioboot.github.io/bggn213_f17/lectures/#12

You can use the classroom computers or your own laptops. If you are using your laptops then you will need to install **VMD** and **MGLTools**

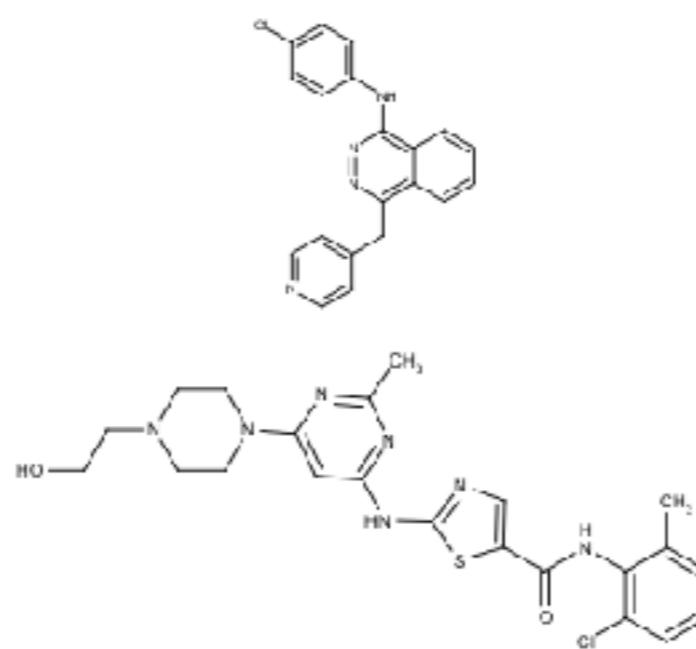
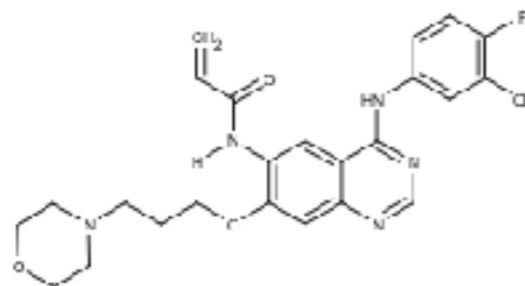
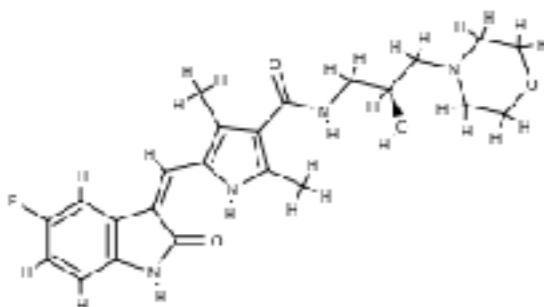
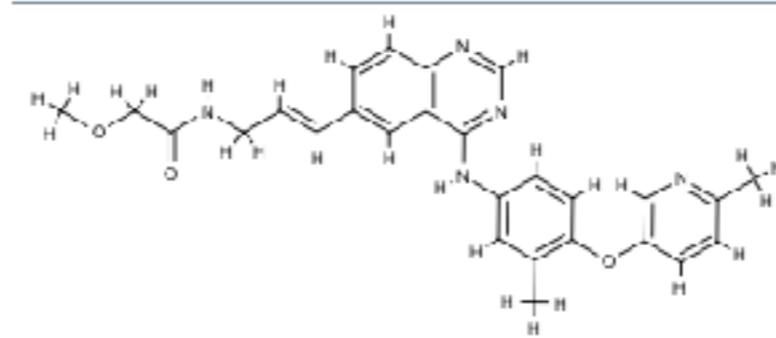
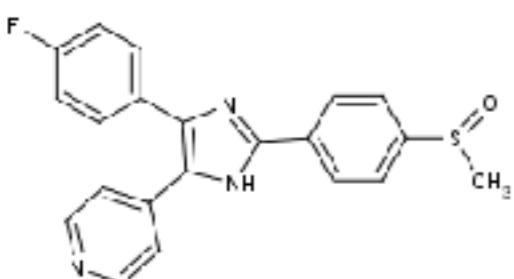
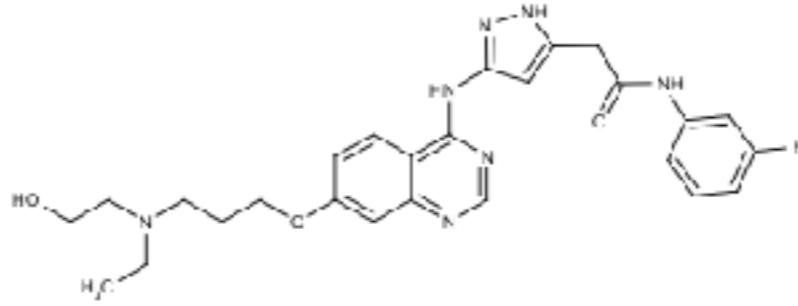
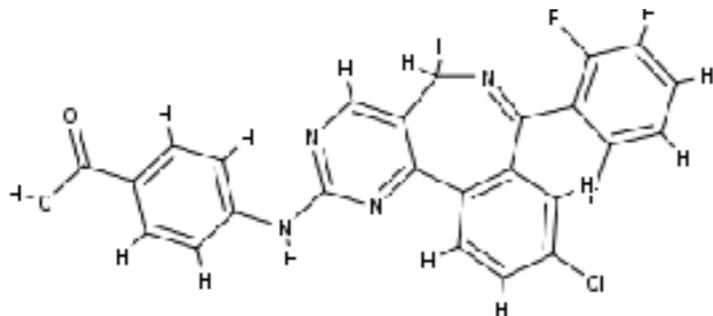
Two main approaches:

- (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

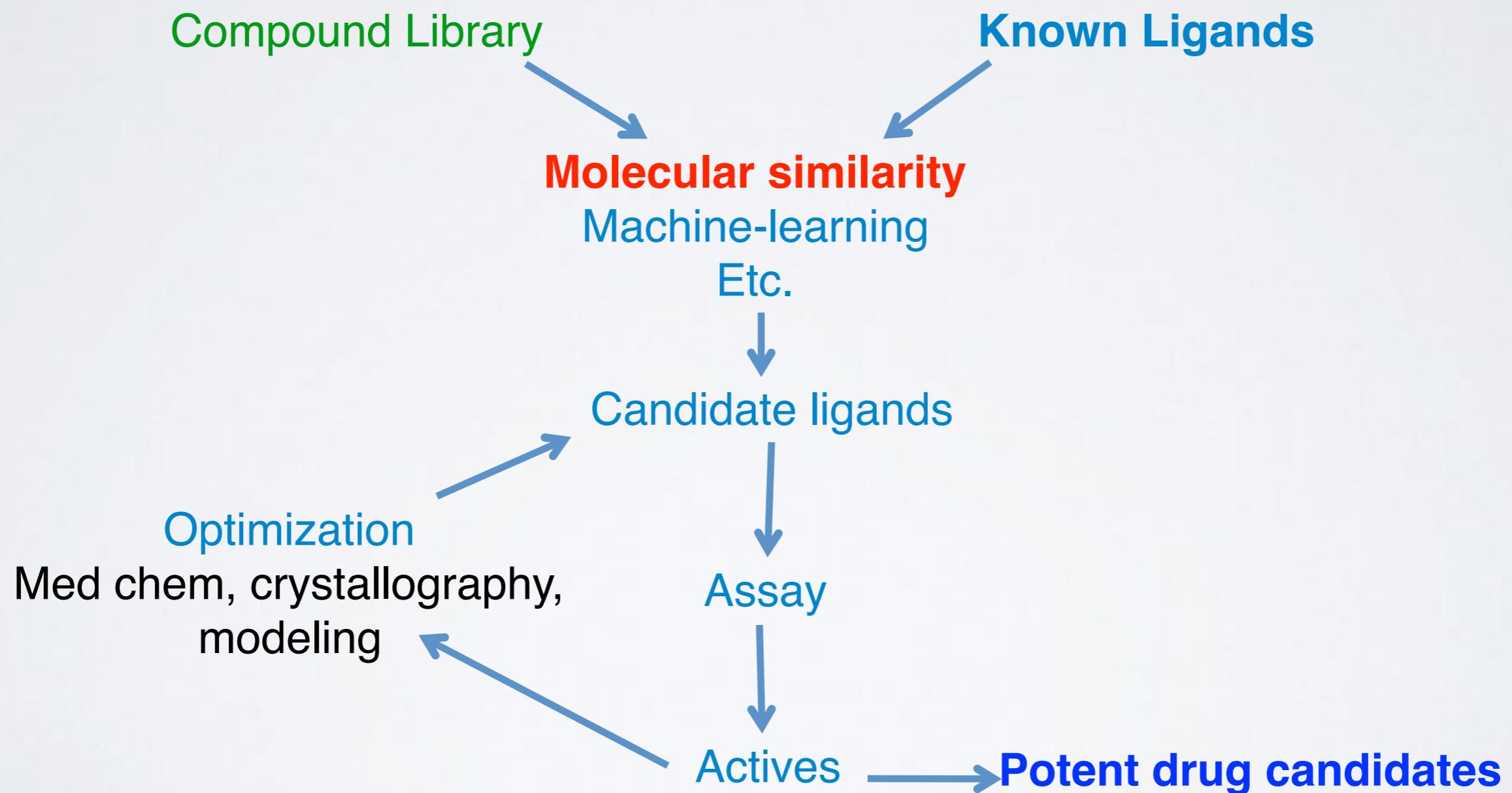
Why Look for Another Ligand if You Already Have Some?

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

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

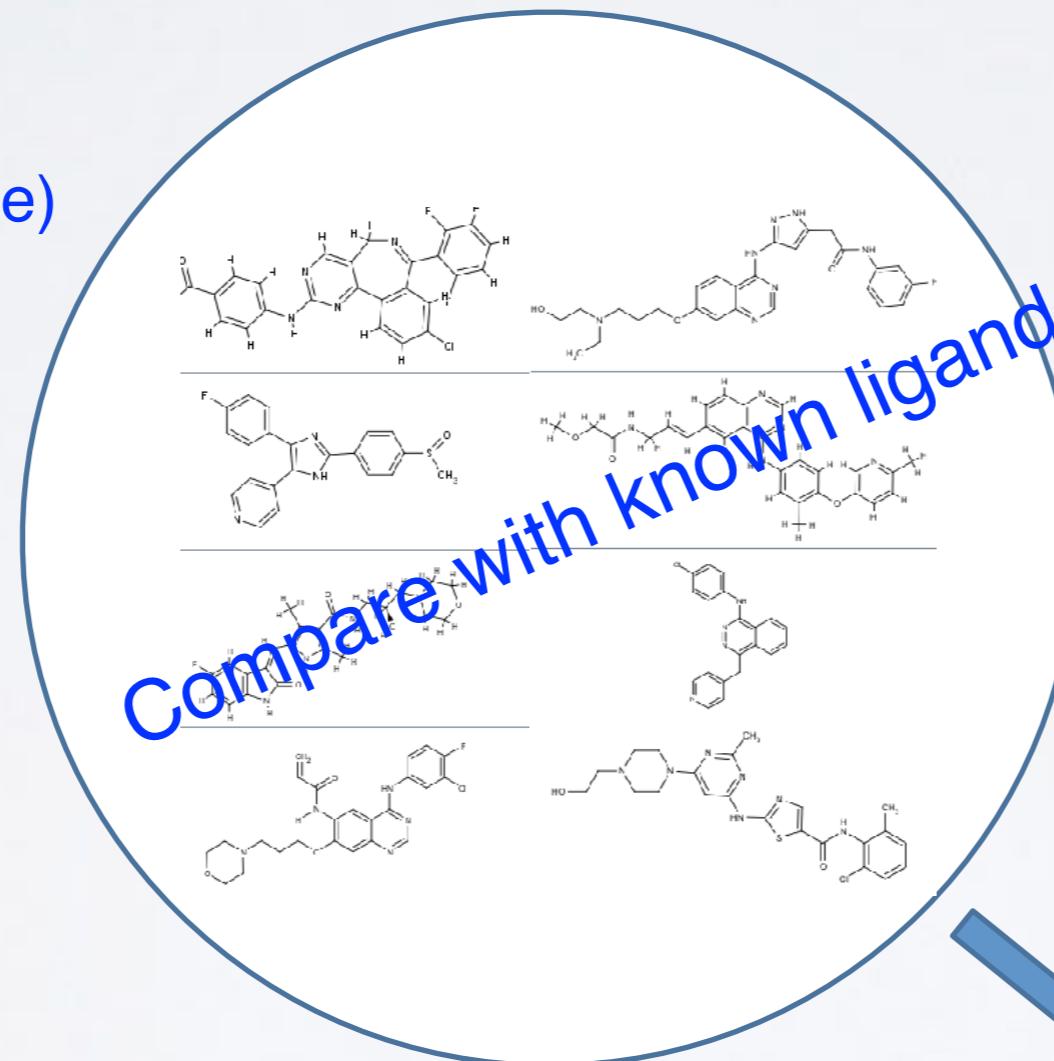
An high-affinity ligand is toxic, is not well-absorbed, difficult to synthesize etc.

LIGAND-BASED VIRTUAL SCREENING



CHEMICAL SIMILARITY LIGAND-BASED DRUG-DISCOVERY

Compounds
(available/synthesizable)



Different

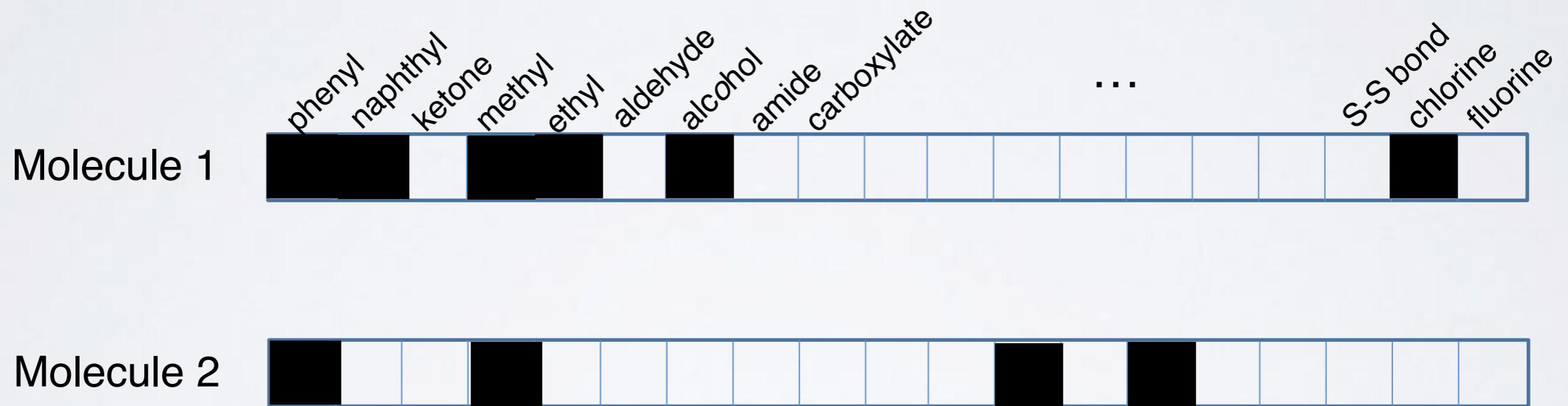
Don't bother

Similar

Test experimentally

CHEMICAL FINGERPRINTS

BINARY STRUCTURE KEYS



CHEMICAL SIMILARITY FROM FINGERPRINTS



Tanimoto Similarity
(or Jaccard Index), T

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

Intersection



$N_I=2$

Union

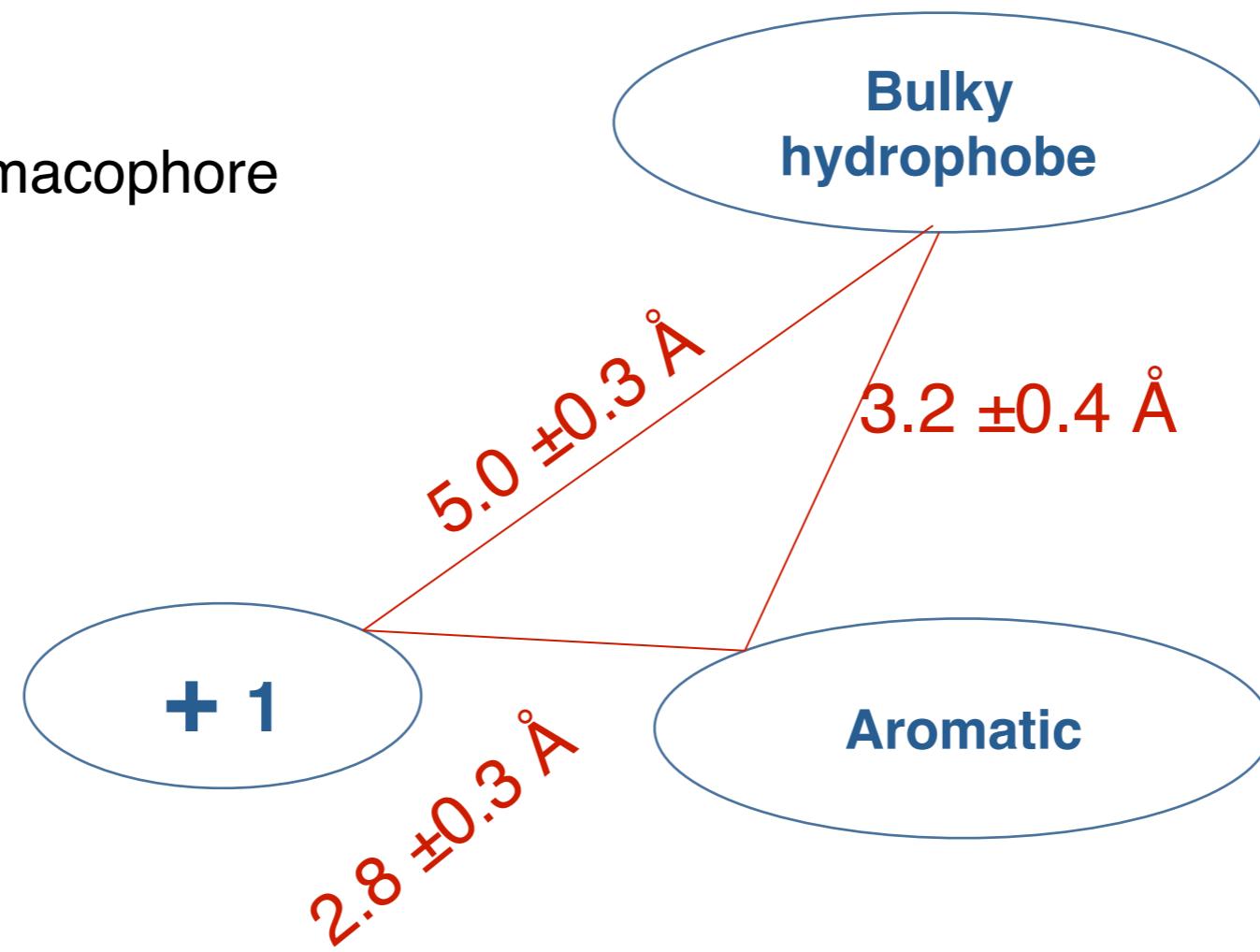


$N_U=8$

Pharmacophore Models

Φάρμακο (drug) + Φορά (carry)

A 3-point pharmacophore



Molecular Descriptors

More abstract than chemical fingerprints

Physical descriptors

molecular weight

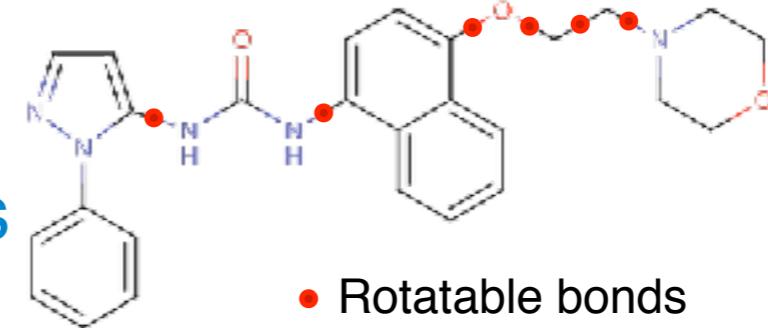
charge

dipole moment

number of H-bond donors/acceptors

number of rotatable bonds

hydrophobicity ($\log P$ and $c\log P$)



• Rotatable bonds

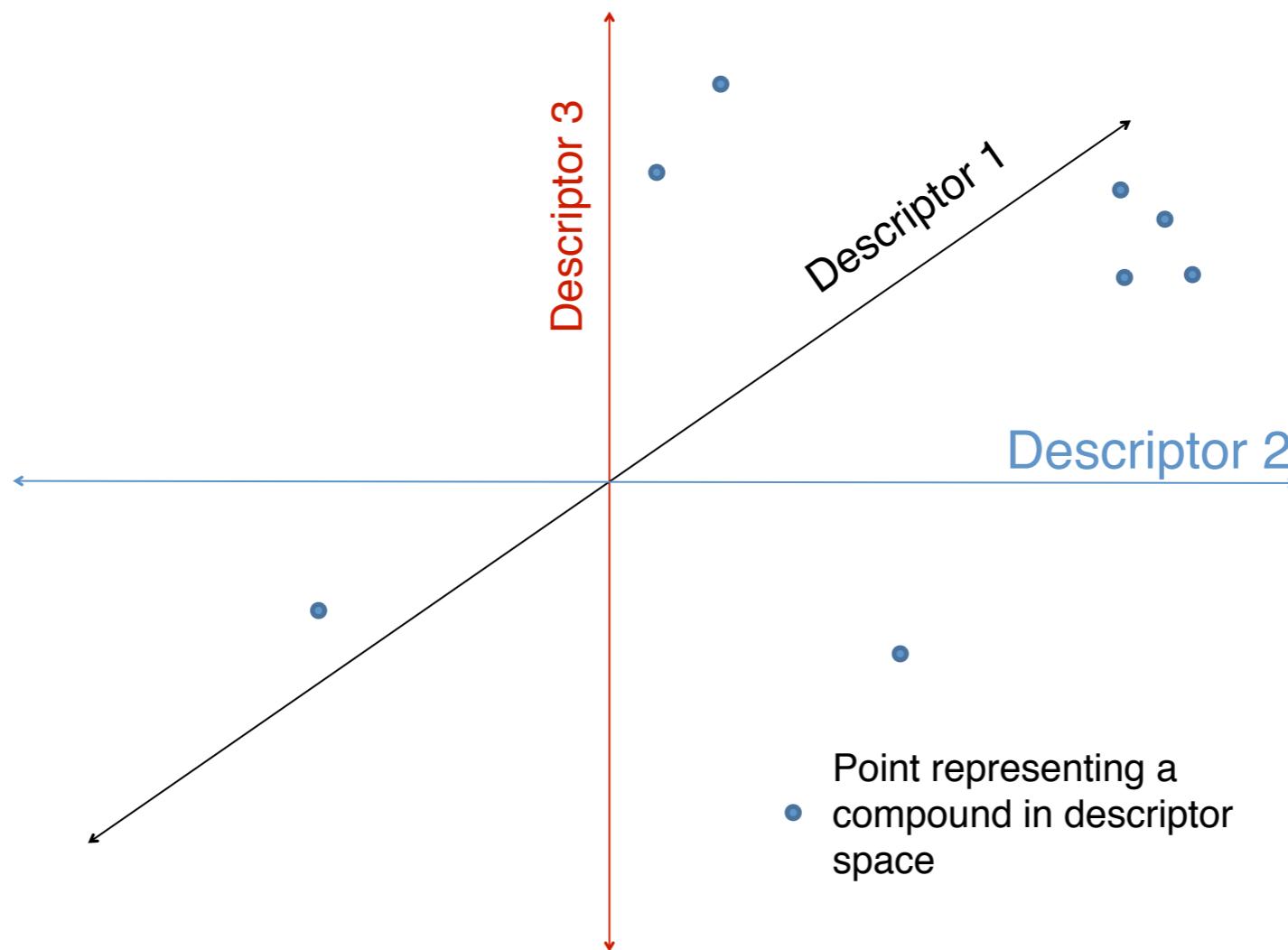
Topological
branching index
measures of linearity vs interconnectedness

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, deep learning etc.)

Approved drugs and clinical candidates

- Catalogue approved drugs and clinical candidates from FDA Orange Book, and USAN applications
- Small molecules and biotherapeutics

ChEMBL wellcome trust

EBI > Databases > Small Molecules > ChEMBL Database > Home

Search ChEMBL... Compounds Targets Assays Documents Activity Source Filter

Ligand Search Target Search Browse Targets **Browse Drugs** Browse Drug Targets Drug Approvals About

Downloads... 10 records per page Search: Show / hide columns

Parent Molecule	Synonyms	Phase	Research Codes	Applicants	USAN Stem	USAN Year	First Approval	ATC Code	Icon
	Elosulfase Alfa (INN, USAN)	4		Biomarin Pharmaceutical Inc.	-ase	2012	2014		
CHEMBL2108676									
	Tasimelteon (FDA, INN, USAN)	4	BMS-214778 VEC-162	Vanda Pharmaceuticals Inc	-melteon	2007	2014		
CHEMBL2103822									
	Apremilast (FDA, INN, USAN)	4	CC-10004	Celgene Corp	-ast	2005	2014	L04AA32	
CHEMBL514800									
	Flortetaben F-18 (FDA) Flortetaben F18 (USAN)	4	BAY-949172 UNII-TLA7312TOI	Piramal Imaging Sa		2013	2014		
CHEMBL1908906									
	Droxidopa (FDA, INN, USAN)	4	DOPS L-DOPS	Chelsea Therapeutics Inc	-dopa	2008	2014		
CHEMBL2108676									

Drug properties

	Drug Type	Rule of Five	First in Class	Chirality	Prodrug	Oral	Parenteral	Topical	Black-Box Warning	Availability Type	
synthetic small molecule		5									prescription only
natural product-derived										over-the-counter	
inorganic					chirally pure					discontinued	
polymer		Ingredient-related (USANs, candidates and approved drugs)				Product-related (approved drugs only)					
monoclonal antibody											
enzyme											
peptide/protein											
oligonucleotide											
oligosaccharide											

LIPINSKI'S RULE OF FIVE

Lipinski's rule of five states that, in general, an orally active drug has no more than one violation of the following criteria:

- Not more than 5 hydrogen bond donors (nitrogen or oxygen atoms with one or more hydrogen atoms)
- Not more than 10 hydrogen bond acceptors (nitrogen or oxygen atoms)
- A molecular mass less than 500 daltons
- An octanol-water partition coefficient $\log P$ not greater than 5

Rules for drug discovery success

- Set of approved drugs or medicinal chemistry compounds and their targets can be used to derive rules for drug discovery success (or failure):
 - What features make a successful drug target?
 - What features make a protein druggable by small molecules?
 - What features of a compound contribute to good oral bioavailability?
 - What chemical groups may be associated with toxicity?

Druggability prediction

https://www.ebi.ac.uk/chembl/druggability/domain/32655

**View cavities
(and ligands)
on structure**

↓

Domain Details:

PDB	1eeo	SCOP
Gene	P18031	View Protein Summary
Description	Tyrosine-protein phosphatase non-receptor type 1	
Fold	Phosphotyrosine protein phosphatases II	
Superfamily	Phosphotyrosine protein phosphatases II	
Family	Higher-molecular-weight phosphotyrosine protein phosphatases	
Other PDB(s)	1eeo:A - px32655 (Tractable: 1, Druggable: 0, Ensemble: -0.96)	

Average Druggability Scores:

Tractable	Druggable	Ensemble
0.97	0.02	-0.93

Tractable/Druggable ranges from low:0 to high:1. Ensemble ranges from low:-1 to high:+1.

Site Druggability Details:

Reset:	Site 1	Site 2	Site 3	Site 4
Druggable	0.00	0.00	0.00	0.00
Confidence	0.73	0.96	0.96	0.96
Tractable	1.00	0.00	0.00	0.00
Confidence	0.92	0.86	0.83	0.86
<u>Ensemble</u>	-0.96	-0.99	-0.98	-0.99
Volume [Å ³]	1535.2	1318.36	1446.61	1454.2
Buried Surface [%]	71.3	65.25	72.27	64.08
Show Site	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Show Residues	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ligand PTR (CHEMBL286939):

Green :Druggable, Yellow :Tractable, Pink :Undruggable

Use | Privacy | Cookies | EBI Funding | Contact EBI | © European Bioinformatics Institute 2012. EBI is an Outstation of the European Molecular Biology Laboratory.

Examples

nature

Vol 460 | 16 July 2009 | doi:10.1038/nature08160

ARTICLES

The genome of the blood fluke *Schistosoma mansoni*

Matthew Berriman¹, Brian J. Haas^{3†}, Philip T. LoVerde⁴, R. Alan Wilson⁵, Gary P. Dillon⁵, Gustavo C. Cerqueira^{6,7,8}, Susan T. Mashiyama^{9,10}, Bissan Al-Lazikani¹¹, Luiza F. Andrade¹², Peter D. Ashton⁴, Martin A. Aslett¹, Daniella C. Bartholomeu^{3†}, Gaelle Blandin³, Conor R. Caffrey⁹, Avril Coghlan¹³, Richard Coulson², Tim A. Day¹⁴, Art Delcher⁷, Ricardo DeMarco^{5,15,16}, Appolinaire Djikeng³, Tina Eyre¹, John A. Gamble¹, Elodie Ghedin^{3†}, Yong Gu¹, Christiane Hertz-Fowler¹, Hirohisa Hirai¹⁷, Yuriko Hirai¹⁷, Robin Houston¹, Alasdair Ivens^{1†}, David A. Johnston^{18†}, Daniela Lacerda^{3†}, Camila D. Macedo^{6,8}, Paul McVeigh¹⁴, Zemin Ning¹, Guilherme Oliveira¹², John P. Overington², Julian Parkhill¹, Mihaela Pertea⁷, Raymond J. Pierce¹⁹, Anna V. Protasio¹, Michael A. Quail¹, Marie-Adèle Rajandream¹, Jane Rogers^{1†}, Mohammed Sajid^{9†}, Steven L. Salzberg^{7,8}, Mario Stanke²⁰, Adrian R. Tivey¹, Owen White^{3†}, David L. Williams^{21†}, Jennifer Wortman^{3†}, Wenjie Wu^{4†}, Mostafa Zamanian¹⁴, Adhemar Zerlotini¹¹, Claire M. Fraser-Liggett^{3†}, Barclay G. Barrell¹ & Najib M. El-Sayed^{3,6,7,8}

Schistosoma mansoni is responsible for the neglected tropical disease schistosomiasis that affects 210 million people in 76 countries. Here we present analysis of the 363 megabase nuclear genome of the blood fluke. It encodes at least 11,809 genes, with an unusual intron size distribution, and new families of micro-exon genes that undergo frequent alternative splicing. As the first sequenced flatworm, and a representative of the Lophotrochozoa, it offers insights into early events in the evolution of the animals, including the development of a body pattern with bilateral symmetry, and the development of tissues into organs. Our analysis has been informed by the need to find new drug targets. The deficits in lipid metabolism that make schistosomes dependent on the host are revealed, and the identification of membrane receptors, ion channels and more than 300 proteases provide new insights into the biology of the life cycle and new targets. Bioinformatics approaches have identified metabolic chokepoints, and a chemogenomic screen has pinpointed schistosome proteins for which existing drugs may be active. The information generated provides an invaluable resource for the research community to develop much needed new control tools for the treatment and eradication of this important and neglected disease.

Schistosomiasis is a neglected tropical disease that ranks with malaria and tuberculosis as a major source of morbidity affecting approximately 210 million people in 76 countries, despite strenuous control efforts¹. It is caused by blood flukes of the genus *Schistosoma* (phylum Platyhelminthes), which exhibit dioecy and have complex life cycles comprising several morphologically distinct phenotypes in definitive human and intermediate snail hosts. *Schistosoma mansoni*, one of the three major human species, occurs across much of sub-Saharan Africa, parts of the Middle East, Brazil, Venezuela and some West Indian islands. The mature flukes dwell in the human portal vasculature, depositing eggs in the intestinal wall that either

pass to the gut lumen and are voided in the faeces, or travel to the liver where they trigger immune-mediated granuloma formation and peri-portal fibrosis². Approximately 280,000 deaths per annum are attributable to schistosomiasis in sub-Saharan Africa alone³. However, the disease is better known for its chronicity and debilitating morbidity⁴. A single drug, praziquantel, is almost exclusively used to treat the infection but this does not prevent reinfection, and with the large-scale control programmes in place, there is concern about the development of drug resistance. Indeed, resistance can be selected for in the laboratory and there are reports of increased drug tolerance in the field⁵.

¹Wellcome Trust Sanger Institute, ²European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. ³The Institute for Genomic Research/The J. Craig Venter Institute, 9712 Medical Center Drive, Rockville, Maryland 20850, USA. ⁴Departments of Biochemistry and Pathology, Mail Code 7760, University of Texas, Health Science Center, San Antonio, Texas 78229-3900, USA. ⁵Department of Biology, University of York, PO Box 373, York YO1 5YW, UK. ⁶Department of Cell Biology and Molecular Genetics, ⁷Center for Bioinformatics and Computational Biology, and ⁸Maryland Pathogen Research Institute, University of Maryland, College Park, Maryland 20742, USA. ⁹Sandler Center for Basic Research in Parasitic Diseases, ¹⁰Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, California Institute for Quantitative Biomedical Research (QB3), Byers Hall, 1700 4th Street, University of California, San Francisco, California 94158-2330, USA. ¹¹Cancer Research UK Centre for Cancer Therapeutics, The Institute of Cancer Research, Haddow Laboratories, 15 Cotsfold Road, Belmont, Sutton, Surrey SM2 5NG, UK. ¹²Centro de Pesquisas René Rachou (CPqRR)—FIOCRUZ, Av Augusto de Lima 1715, Belo Horizonte, MG 30190002, Brazil. ¹³Department of Microbiology, University College Cork, Western Road, Cork, Ireland. ¹⁴Department of Biomedical Sciences, Iowa State University, Ames, Iowa 50011, USA. ¹⁵Instituto de Química, ¹⁶Instituto de Física de São Carlos, Universidade de São Paulo, Brazil. ¹⁷Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan. ¹⁸Biomedical Parasitology Division, The Natural History Museum, London SW7 5BD, UK. ¹⁹Inserm, U547, Université Lille 2, Institut Pasteur de Lille, IFR 142, Lille, France. ²⁰Institut für Mikrobiologie und Genetik, Abteilung Bioinformatik, Universität Göttingen, Goldschmidtstraße 1, Göttingen 37077, Germany. ²¹Department of Biological Sciences, Illinois State University, Normal, Illinois 61790-4120, USA. [†]Present addresses: The Broad Institute, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA (B.J.H.); Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil (D.C.B. and D.L.); Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA (E.G.); Flos Genomics Ltd, ETC, King's Buildings, Edinburgh EH9 3JL, UK (A.J.). ²²Biomedical Imaging Unit, School of Medicine, University of Southampton, Southampton SO16 6YD, UK (D.A.). ²³Leibniz Institute for Age Research—Fränckel Center, Leibniz Research Institute, Berlin, Germany (M.D.). ²⁴Leiden University Medical Center, Leiden, The Netherlands (J.B.). ²⁵Universität Regensburg, Regensburg, Germany (S.M.). ²⁶University of Cambridge, Cambridge, UK (A.L.H.). ²⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁸Leiden University, Leiden, The Netherlands (A.L.H.). ²⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ³⁰Leiden University, Leiden, The Netherlands (A.L.H.). ³¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ³²Leiden University, Leiden, The Netherlands (A.L.H.). ³³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ³⁴Leiden University, Leiden, The Netherlands (A.L.H.). ³⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ³⁶Leiden University, Leiden, The Netherlands (A.L.H.). ³⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ³⁸Leiden University, Leiden, The Netherlands (A.L.H.). ³⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁴⁰Leiden University, Leiden, The Netherlands (A.L.H.). ⁴¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁴²Leiden University, Leiden, The Netherlands (A.L.H.). ⁴³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁴⁴Leiden University, Leiden, The Netherlands (A.L.H.). ⁴⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁴⁶Leiden University, Leiden, The Netherlands (A.L.H.). ⁴⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁴⁸Leiden University, Leiden, The Netherlands (A.L.H.). ⁴⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁵⁰Leiden University, Leiden, The Netherlands (A.L.H.). ⁵¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁵²Leiden University, Leiden, The Netherlands (A.L.H.). ⁵³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁵⁴Leiden University, Leiden, The Netherlands (A.L.H.). ⁵⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁵⁶Leiden University, Leiden, The Netherlands (A.L.H.). ⁵⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁵⁸Leiden University, Leiden, The Netherlands (A.L.H.). ⁵⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁶⁰Leiden University, Leiden, The Netherlands (A.L.H.). ⁶¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁶²Leiden University, Leiden, The Netherlands (A.L.H.). ⁶³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁶⁴Leiden University, Leiden, The Netherlands (A.L.H.). ⁶⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁶⁶Leiden University, Leiden, The Netherlands (A.L.H.). ⁶⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁶⁸Leiden University, Leiden, The Netherlands (A.L.H.). ⁶⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁷⁰Leiden University, Leiden, The Netherlands (A.L.H.). ⁷¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁷²Leiden University, Leiden, The Netherlands (A.L.H.). ⁷³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁷⁴Leiden University, Leiden, The Netherlands (A.L.H.). ⁷⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁷⁶Leiden University, Leiden, The Netherlands (A.L.H.). ⁷⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁷⁸Leiden University, Leiden, The Netherlands (A.L.H.). ⁷⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁸⁰Leiden University, Leiden, The Netherlands (A.L.H.). ⁸¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁸²Leiden University, Leiden, The Netherlands (A.L.H.). ⁸³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁸⁴Leiden University, Leiden, The Netherlands (A.L.H.). ⁸⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁸⁶Leiden University, Leiden, The Netherlands (A.L.H.). ⁸⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁸⁸Leiden University, Leiden, The Netherlands (A.L.H.). ⁸⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁹⁰Leiden University, Leiden, The Netherlands (A.L.H.). ⁹¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁹²Leiden University, Leiden, The Netherlands (A.L.H.). ⁹³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁹⁴Leiden University, Leiden, The Netherlands (A.L.H.). ⁹⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁹⁶Leiden University, Leiden, The Netherlands (A.L.H.). ⁹⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ⁹⁸Leiden University, Leiden, The Netherlands (A.L.H.). ⁹⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁰⁰Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁰¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁰²Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁰³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁰⁴Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁰⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁰⁶Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁰⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁰⁸Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁰⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹¹⁰Leiden University, Leiden, The Netherlands (A.L.H.). ¹¹¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹¹²Leiden University, Leiden, The Netherlands (A.L.H.). ¹¹³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹¹⁴Leiden University, Leiden, The Netherlands (A.L.H.). ¹¹⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹¹⁶Leiden University, Leiden, The Netherlands (A.L.H.). ¹¹⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹¹⁸Leiden University, Leiden, The Netherlands (A.L.H.). ¹¹⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹²⁰Leiden University, Leiden, The Netherlands (A.L.H.). ¹²¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹²²Leiden University, Leiden, The Netherlands (A.L.H.). ¹²³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹²⁴Leiden University, Leiden, The Netherlands (A.L.H.). ¹²⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹²⁶Leiden University, Leiden, The Netherlands (A.L.H.). ¹²⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹²⁸Leiden University, Leiden, The Netherlands (A.L.H.). ¹²⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹³⁰Leiden University, Leiden, The Netherlands (A.L.H.). ¹³¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹³²Leiden University, Leiden, The Netherlands (A.L.H.). ¹³³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹³⁴Leiden University, Leiden, The Netherlands (A.L.H.). ¹³⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹³⁶Leiden University, Leiden, The Netherlands (A.L.H.). ¹³⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹³⁸Leiden University, Leiden, The Netherlands (A.L.H.). ¹³⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁴⁰Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁴¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁴²Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁴³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁴⁴Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁴⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁴⁶Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁴⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁴⁸Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁴⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁵⁰Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁵¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁵²Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁵³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁵⁴Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁵⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁵⁶Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁵⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁵⁸Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁵⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁶⁰Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁶¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁶²Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁶³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁶⁴Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁶⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁶⁶Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁶⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁶⁸Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁶⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁷⁰Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁷¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁷²Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁷³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁷⁴Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁷⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁷⁶Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁷⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁷⁸Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁷⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁸⁰Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁸¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁸²Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁸³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁸⁴Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁸⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁸⁶Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁸⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁸⁸Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁸⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁹⁰Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁹¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁹²Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁹³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁹⁴Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁹⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁹⁶Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁹⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ¹⁹⁸Leiden University, Leiden, The Netherlands (A.L.H.). ¹⁹⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁰⁰Leiden University, Leiden, The Netherlands (A.L.H.). ²⁰¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁰²Leiden University, Leiden, The Netherlands (A.L.H.). ²⁰³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁰⁴Leiden University, Leiden, The Netherlands (A.L.H.). ²⁰⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁰⁶Leiden University, Leiden, The Netherlands (A.L.H.). ²⁰⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁰⁸Leiden University, Leiden, The Netherlands (A.L.H.). ²⁰⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²¹⁰Leiden University, Leiden, The Netherlands (A.L.H.). ²¹¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²¹²Leiden University, Leiden, The Netherlands (A.L.H.). ²¹³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²¹⁴Leiden University, Leiden, The Netherlands (A.L.H.). ²¹⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²¹⁶Leiden University, Leiden, The Netherlands (A.L.H.). ²¹⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²¹⁸Leiden University, Leiden, The Netherlands (A.L.H.). ²¹⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²²⁰Leiden University, Leiden, The Netherlands (A.L.H.). ²²¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²²²Leiden University, Leiden, The Netherlands (A.L.H.). ²²³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²²⁴Leiden University, Leiden, The Netherlands (A.L.H.). ²²⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²²⁶Leiden University, Leiden, The Netherlands (A.L.H.). ²²⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²²⁸Leiden University, Leiden, The Netherlands (A.L.H.). ²²⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²³⁰Leiden University, Leiden, The Netherlands (A.L.H.). ²³¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²³²Leiden University, Leiden, The Netherlands (A.L.H.). ²³³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²³⁴Leiden University, Leiden, The Netherlands (A.L.H.). ²³⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²³⁶Leiden University, Leiden, The Netherlands (A.L.H.). ²³⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²³⁸Leiden University, Leiden, The Netherlands (A.L.H.). ²³⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁴⁰Leiden University, Leiden, The Netherlands (A.L.H.). ²⁴¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁴²Leiden University, Leiden, The Netherlands (A.L.H.). ²⁴³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁴⁴Leiden University, Leiden, The Netherlands (A.L.H.). ²⁴⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁴⁶Leiden University, Leiden, The Netherlands (A.L.H.). ²⁴⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁴⁸Leiden University, Leiden, The Netherlands (A.L.H.). ²⁴⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁵⁰Leiden University, Leiden, The Netherlands (A.L.H.). ²⁵¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁵²Leiden University, Leiden, The Netherlands (A.L.H.). ²⁵³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁵⁴Leiden University, Leiden, The Netherlands (A.L.H.). ²⁵⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁵⁶Leiden University, Leiden, The Netherlands (A.L.H.). ²⁵⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁵⁸Leiden University, Leiden, The Netherlands (A.L.H.). ²⁵⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁶⁰Leiden University, Leiden, The Netherlands (A.L.H.). ²⁶¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁶²Leiden University, Leiden, The Netherlands (A.L.H.). ²⁶³Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁶⁴Leiden University, Leiden, The Netherlands (A.L.H.). ²⁶⁵Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁶⁶Leiden University, Leiden, The Netherlands (A.L.H.). ²⁶⁷Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁶⁸Leiden University, Leiden, The Netherlands (A.L.H.). ²⁶⁹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁷⁰Leiden University, Leiden, The Netherlands (A.L.H.). ²⁷¹Leiden University Medical Center, Leiden, The Netherlands (A.L.H.). ²⁷²Leiden University, Leiden, The Netherlands (A.L.H.). ²⁷³Leiden University Medical Center, Leiden

Target prediction models

- Active compounds from ChEMBL can be used to train target prediction models
- Variety of methods used
 - Multi-Category Naïve Bayesian Classifier (e.g., ChEMBL)
 - Chemical similarity between ligand sets (e.g., SEA)
 - 3D similarity between ligands (e.g., SwissTargetPrediction)
 - Protein and ligand descriptors (e.g., Proteochemometric models)
- Open source tools available for many methods
 - E.g., Scikit-learn with RDKit

Examples at: https://github.com/chembl/mychembl/blob/master/ipython_notebooks

RESEARCH ARTICLE

Mycobacterial Dihydrofolate Reductase Inhibitors Identified Using Chemogenomic Methods and *In Vitro* Validation

Grace Mugumbate¹, Katherine A. Abrahams², Jonathan A. G. Cox², George Papadatos¹, Gerard van Westen¹, Joël Lelièvre³, Szymon T. Calus², Nicholas J. Loman², Lluís Ballell³, David Barros³, John P. Overington^{1*}, Gurdyal S. Besra^{2*}

1 European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom, **2** Institute of Microbiology and Infection (IMI), School of Biosciences, University of Birmingham, Edgbaston, Birmingham, United Kingdom, **3** Diseases of the Developing World, GlaxoSmithKline, Severo Ochoa 2, 28760 Tres Cantos, Madrid, Spain

* jpo@ebi.ac.uk (JPO); g.besra@bham.ac.uk (GSB)



OPEN ACCESS

Citation: Mugumbate G, Abrahams KA, Cox JAG, Papadatos G, van Westen G, Lelièvre J, et al. (2015) Mycobacterial Dihydrofolate Reductase Inhibitors Identified Using Chemogenomic Methods and *In Vitro* Validation. PLoS ONE 10(3): e0121492. doi:10.1371/journal.pone.0121492

Academic Editor: Anil Kumar Tyagi, University of Delhi, INDIA

Received: December 4, 2014

Accepted: February 1, 2015

Published: March 23, 2015

Copyright: © 2015 Mugumbate et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: GM and GvW are grateful to EMBL and Marie Curie Actions for funding this work. GSB acknowledges support in the form of a Personal Research Chair from Mr. James Bardwick, a Royal Society Wolfson Research Merit Award, The Wellcome Trust (681569/Z/06/Z), and the Medical Research Council (MR/K012118/1). The research also received funding from the European Union's 7th framework programme (FP7-2007–2013) under grant agreement ORCHID no. 261378. The funders had no

Abstract

The lack of success in target-based screening approaches to the discovery of antibacterial agents has led to reemergence of phenotypic screening as a successful approach of identifying bioactive, antibacterial compounds. A challenge though with this route is then to identify the molecular target(s) and mechanism of action of the hits. This target identification, or deorphanization step, is often essential in further optimization and validation studies. Direct experimental identification of the molecular target of a screening hit is often complex, precisely because the properties and specificity of the hit are not yet optimized against that target, and so many false positives are often obtained. An alternative is to use computational, predictive, approaches to hypothesize a mechanism of action, which can then be validated in a more directed and efficient manner. Specifically here we present experimental validation of an *in silico* prediction from a large-scale screen performed against *Mycobacterium tuberculosis* (*Mtb*), the causative agent of tuberculosis. The two potent anti-tubercular compounds studied in this case, belonging to the tetrahydro-1,3,5-triazin-2-amine (THT) family, were predicted and confirmed to be an inhibitor of dihydrofolate reductase (DHFR), a known essential *Mtb* gene, and already clinically validated as a drug target. Given the large number of similar screening data sets shared amongst the community, this *in vitro* validation of these target predictions gives weight to computational approaches to establish the mechanism of action (MoA) of novel screening hit.

Introduction

The human pathogen, *Mycobacterium tuberculosis* (*Mtb*) is the causative agent of tuberculosis (TB), an infectious disease that is widespread, infecting around one third of the world's population [1]. The discovery of streptomycin in 1943, and the subsequent discovery and

ARTICLE

doi:10.1038/nature11159

Large-scale prediction and testing of drug activity on side-effect targets

Eugen Lounkin^{1*}, Michael J. Keiser^{2,3*}, Steven Whitebread¹, Dmitri Mikhailov¹, Jacques Hamon⁴, Jeremy L. Jenkins¹, Paul Lavan⁴, Eckhard Weber⁴, Allison K. Doak³, Serge Côté⁴, Brian K. Shoichet³ & Laszlo Urban¹

Discovering the unintended 'off-targets' that predict adverse drug reactions is daunting by empirical methods alone. Drugs can act on several protein targets, some of which can be unrelated by conventional molecular metrics, and hundreds of proteins have been implicated in side effects. Here we use a computational strategy to predict the activity of 656 marketed drugs on 73 unintended 'side-effect' targets. Approximately half of the predictions were confirmed, either from proprietary databases unknown to the method or by new experimental assays. Affinities for these new off-targets ranged from 1 nM to 30 μM. To explore relevance, we developed an association metric to prioritize those new off-targets that explained side effects better than any known target of a given drug, creating a drug-target-adverse drug reaction network. Among these new associations was the prediction that the abdominal pain side effect of the synthetic oestrogen chlorotrianisene was mediated through its newly discovered inhibition of the enzyme cyclooxygenase-1. The clinical relevance of this inhibition was borne out in whole human blood platelet aggregation assays. This approach may have wide application to de-risking toxicological liabilities in drug discovery.

Adverse drug reactions (ADRs) can limit the use of otherwise effective drugs. Next to lack of efficacy, they are the leading cause for attrition in clinical trials of new drugs^{1–3} and are more prominent still in the failure of molecules to advance from pre-clinical research into human trials⁴. Some ADRs are caused by modulation of the primary target of a drug⁵, others result from non-specific interactions of reactive metabolites⁶. In many cases, however, ADRs are caused by unintended activity at off-targets. Notorious examples of off-target toxicity include that of the appetite suppressant fenfluramine-phenetermine (fen-phen), which was withdrawn from the market after numerous patient deaths. These owed to the activation of the 5-hydroxytryptamine-2B (5-HT_{2B}) receptor by one of its metabolites, norfenfluramine, leading to proliferative valvular heart disease⁷. Similarly, well-known drugs, such as the antihistamine terfenadine, have been withdrawn because they caused arrhythmias and death, which have been attributed to their off-target inhibition of the human *ether-à-go-go*-related gene potassium channel (hERG, also known as KCNH2)^{8,9}. Prediction of unknown off-target drug interactions might prevent such disastrous drug toxicities, which are often detected only after fatalities in the clinic, and might allow safer molecules to be prioritized for pre-clinical development. Methods to systematically predict off-targets, and associate these with side effects, have thus attracted intense interest^{10–16}, frequently in the form of either chemical genomics^{17,18} or informatics^{19–26} approaches.

Whereas the informatics methods have never been tested systematically on a large scale, in principle they can be deployed against thousands of targets. Here we present a large-scale, prospective evaluation of safety target prediction using one such method, the similarity ensemble approach (SEA)^{25–27}. SEA calculates whether a molecule will bind to a target based on the chemical features it shares with those of known ligands, using a statistical model to control for random similarity. Because SEA relies only on chemical similarity, it can be applied systematically and, for those targets that have known ligands,

comprehensively. For 656 drugs approved for human use (Supplementary Table 1), targets were predicted from among 73 proteins (Supplementary Table 2 and Methods) with established association of ADRs^{22,28}, for which assays were available at Novartis. Encouragingly, many of the predictions were confirmed, often at pharmacologically relevant concentrations. This motivated us to develop a guilt-by-association metric that linked the new targets to the ADRs of those drugs for which they are the primary or well-known off-targets, creating a drug-target-ADR network. The applicability and the limitations of this approach will be considered.

Testing the predictions

The 656 drugs were computationally screened for their likelihood to bind to 73 targets (Supplementary Table 2) using SEA^{25–27}. The targets belong to the Novartis *in vitro* safety panels based on their association with ADRs^{22,28}. Here we insisted that they also be described in the ChEMBL database²⁹, enabling correspondence with SEA predictions (Supplementary Table 2). ChEMBL annotates more than 285,000 ligands modulating more than 1,500 different human targets with affinities better than 30 μM. SEA calculated the similarity of each drug versus each set of ligands for the 73 targets, comparing the overall set similarity to a model of such expected at random. For instance, the sodium channel blocker aprindine loosely resembled the set of histamine H₁ ligands; although no single H₁ ligand was strongly similar to the drug (Table 1), the overall similarity of the set was much greater than expected at random, leading to a highly significant SEA expectation value (*E* value) of 5×10^{-26} between aprindine and H₁ receptor ligands. Only 1,644 of the more than 47,000 possible drug-target pairs had significant *E* values. Of these, 403 were already known in ChEMBL, and so were trivially confirmed; we do not consider these further. Of the remaining 1,241 predictions, 348 (28%) were unknown to ChEMBL, but could be found in proprietary ligand-target databases that were unavailable to SEA (see Methods). The remaining

¹Novartis Institutes for Biomedical Research, Cambridge, Massachusetts 02139, USA. ²SeaChange Pharmaceuticals Inc, 409 Illinois Street, San Francisco, California 94158, USA. ³Department of Pharmaceutical Chemistry, University of California, San Francisco, 1700 4th Street, Byers Hall Suite 508D, California 94158-2550, USA. ⁴Novartis Institutes for Biomedical Research, 4056 Basel, Switzerland.

*These authors contributed equally to this work.

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
- Introduced both structure and ligand based bioinformatics approaches for drug discovery and design

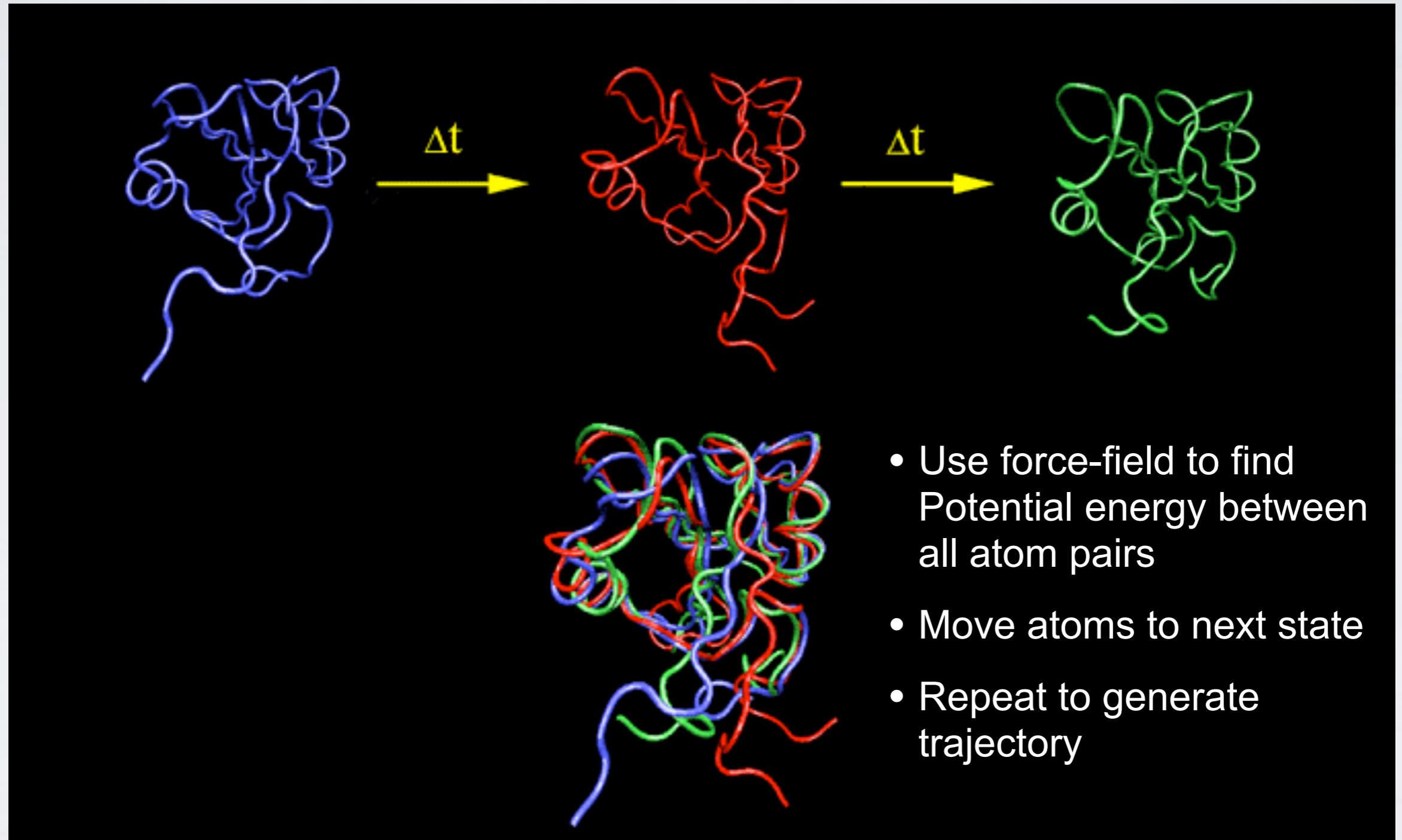
NEXT UP:

- ▶ **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 intrinsically flexible 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 mapping of structure to function
 - Molecular dynamics (MD) and normal mode analysis (NMA) are two major methods for predicting and characterizing molecular motions and their properties

MOLECULAR DYNAMICS SIMULATION



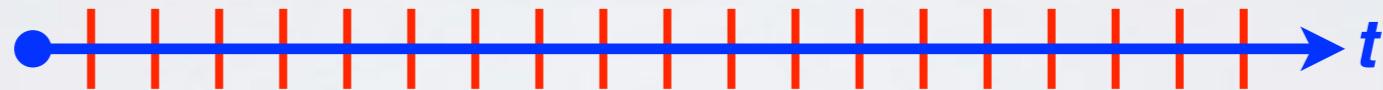
McCammon, Gelin & Karplus, *Nature* (1977)

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

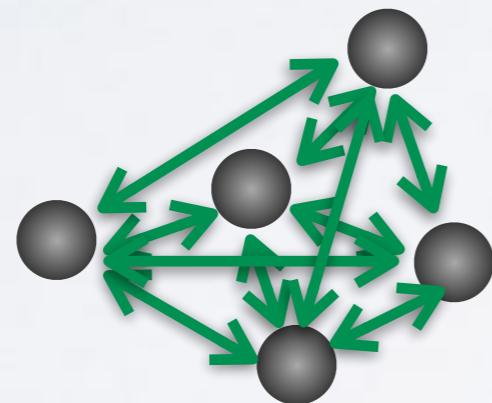
- ▶ Divide **time** into discrete ($\sim 1\text{fs}$) **time steps (Δt)**
(for integrating equations of motion, see below)



- ▶ Divide **time** into discrete ($\sim 1\text{fs}$) **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)



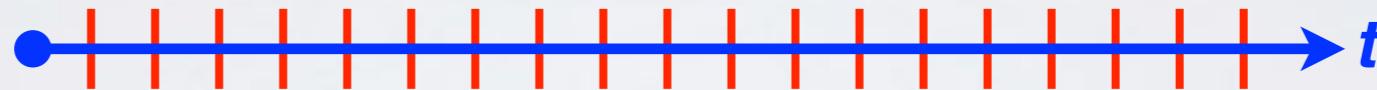
Nucleic motion described classically

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

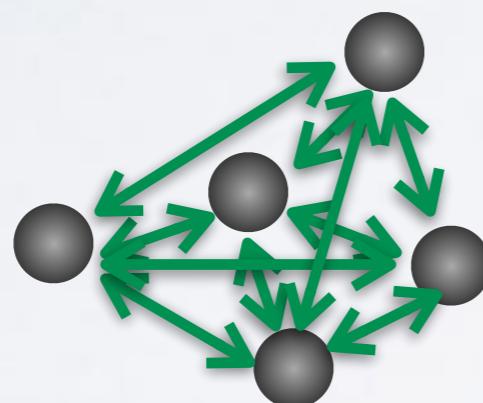
Empirical force field

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

- ▶ Divide **time** into discrete ($\sim 1\text{fs}$) **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)



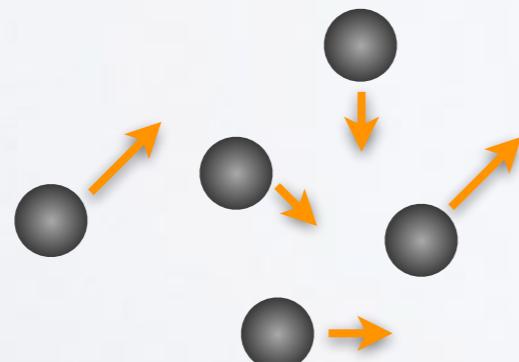
Nucleic motion described classically

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

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)



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

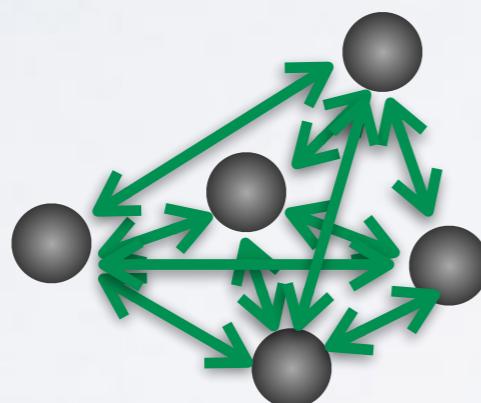
$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \boxed{v(t + \frac{\Delta t}{2})} \Delta t$$

BASIC ANATOMY OF A MD SIMULATION

- Divide **time** into discrete ($\sim 1\text{fs}$) **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)



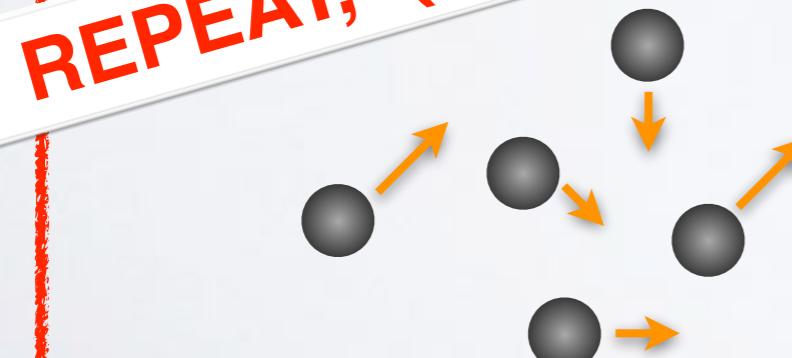
Nucleic motion described classically

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

Empirical force function

$$E(\vec{R}) = \sum_{i=1}^N \sum_{j \neq i, \text{non-bonded}} E_i(\vec{R})$$

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

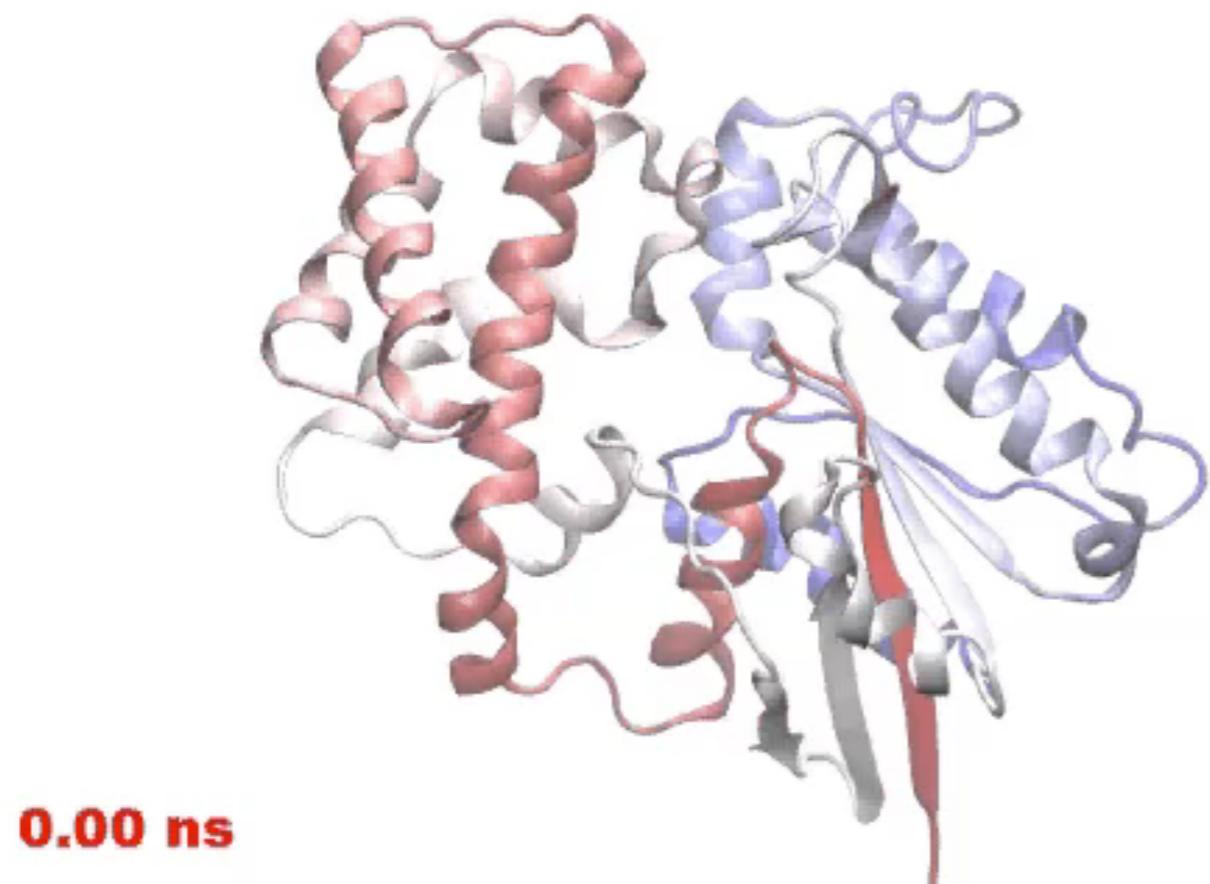


$$\begin{aligned} \mathbf{v}(t + \frac{\Delta t}{2}) &= \mathbf{v}(t - \frac{\Delta t}{2}) + \frac{\mathbf{F}(t)}{m} \Delta t \\ \mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \mathbf{v}(t + \frac{\Delta t}{2}) \Delta t \end{aligned}$$

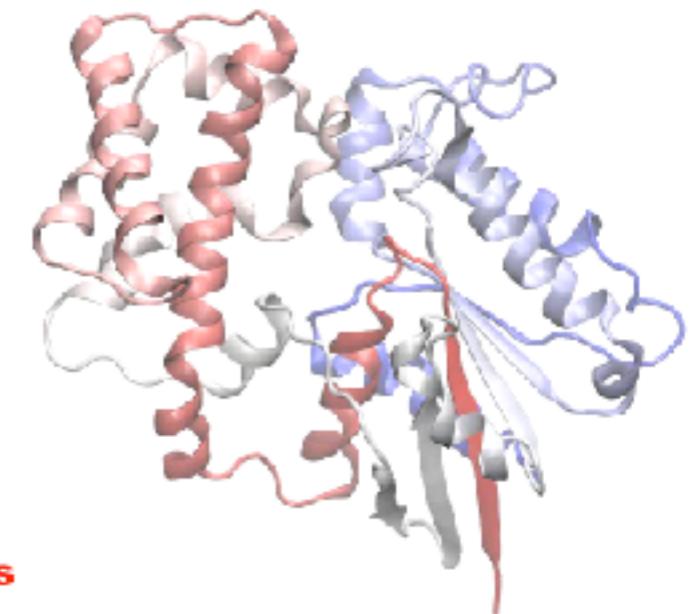
REPEAT, (iterate many, many times... 1ms = 10^{12} time steps)

MD Prediction of Functional Motions

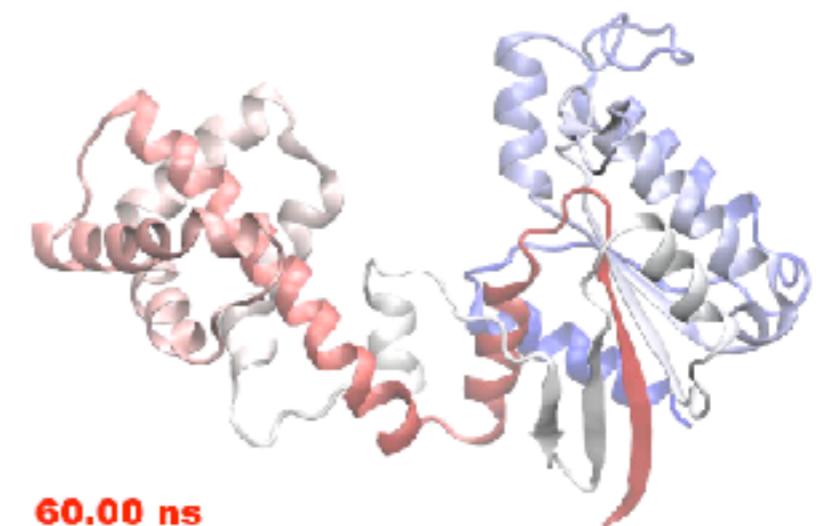
Accelerated MD simulation of
nucleotide-free transducin alpha subunit



“close”

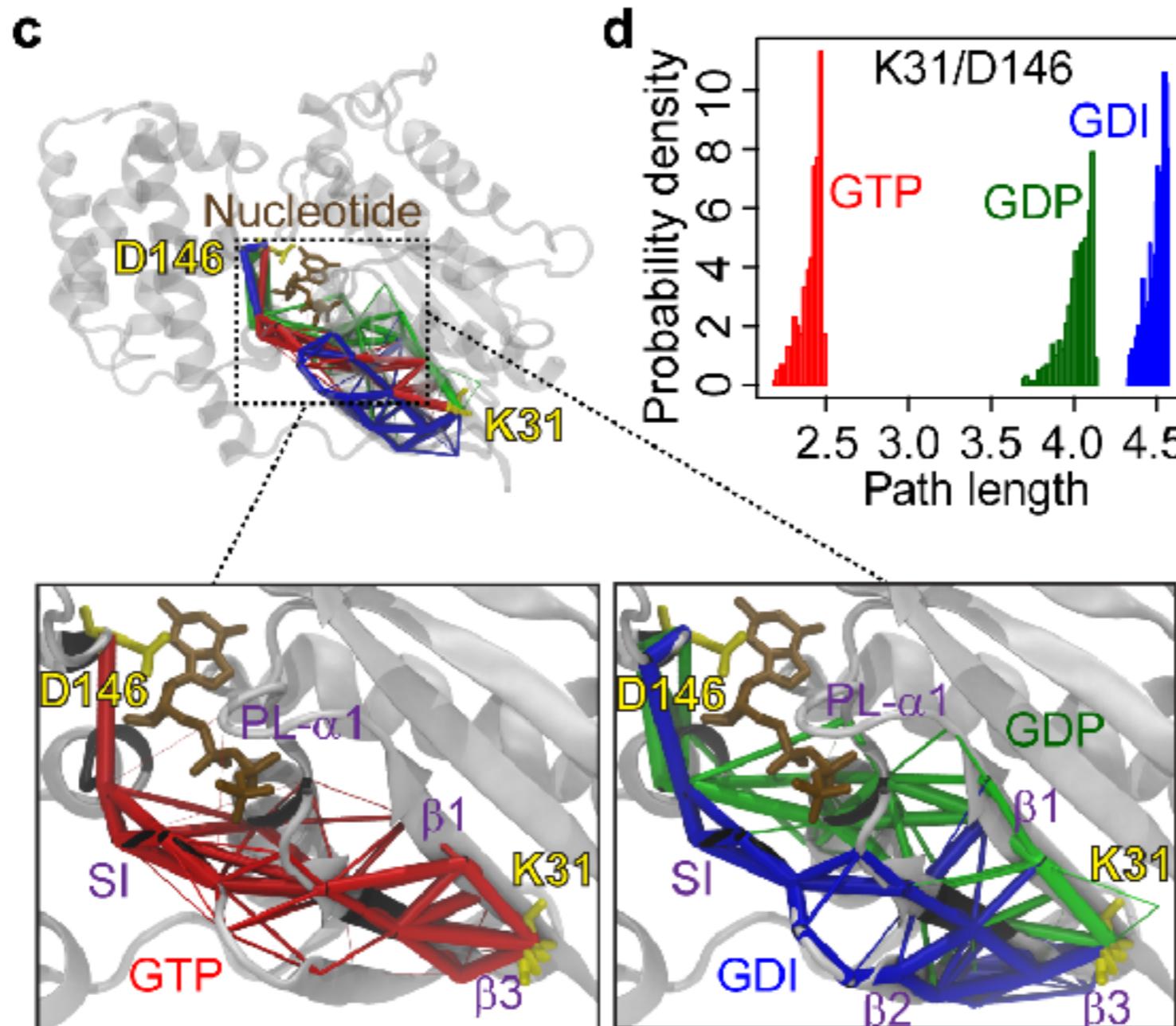


“open”

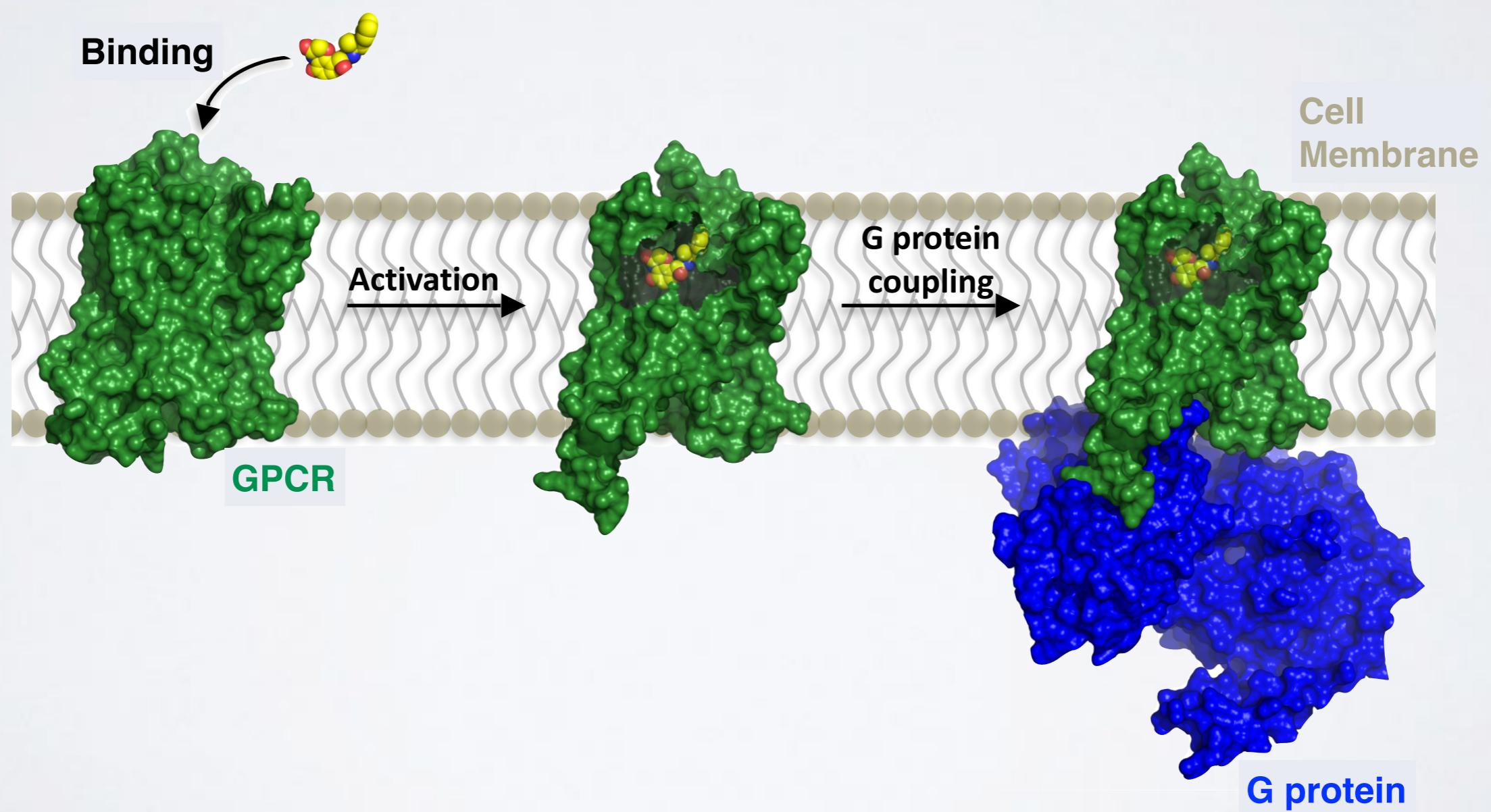


Yao and Grant, Biophys J. (2013)

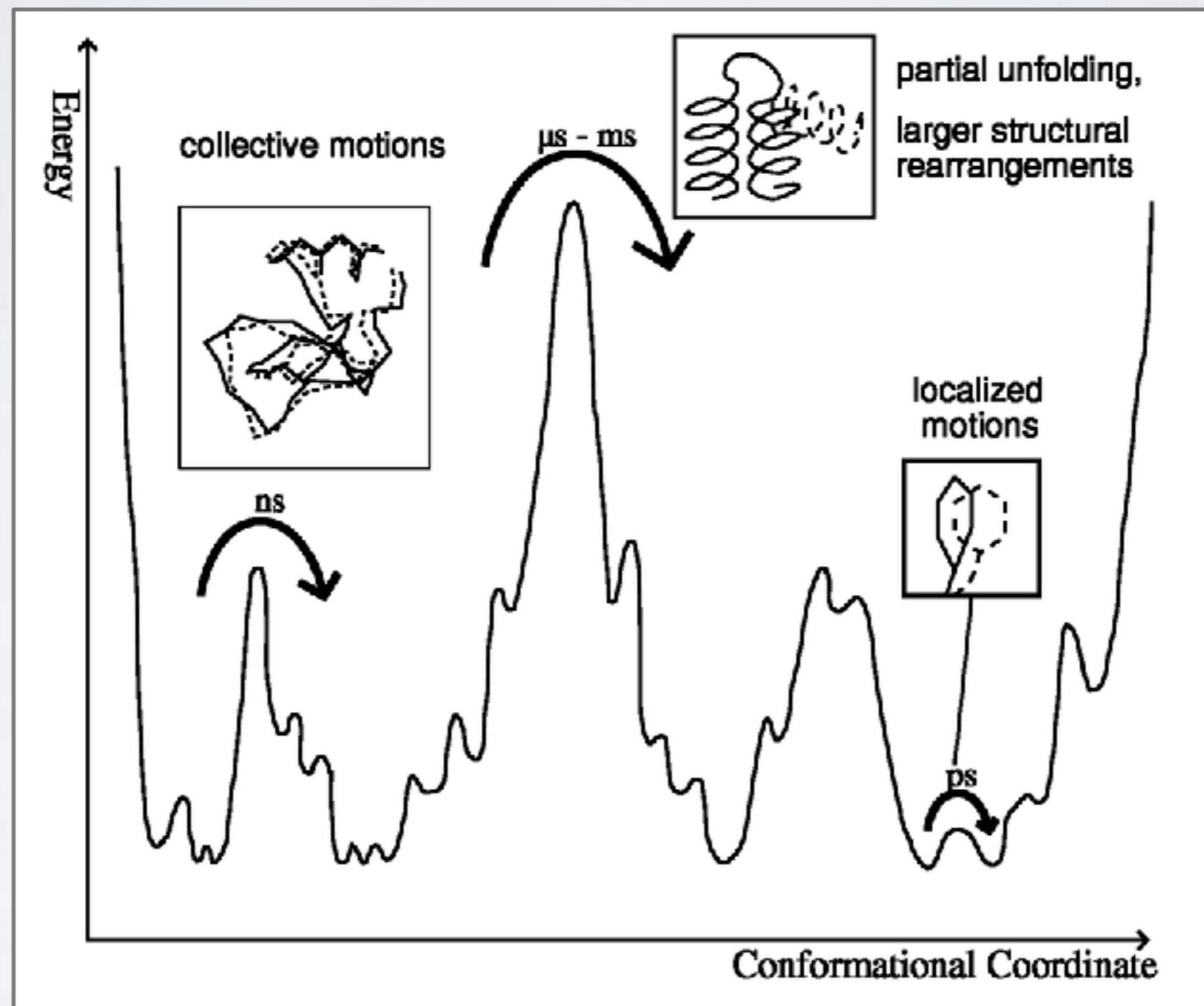
Simulations Identify Key Residues Mediating Dynamic Activation



EXAMPLE APPLICATION OF MOLECULAR SIMULATIONS TO GPCRS



PROTEINS JUMP BETWEEN MANY, HIERARCHICALLY ORDERED “CONFORMATIONAL SUBSTATES”



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

MOLECULAR DYNAMICS IS VERY

Example: F₁-ATPase in water (183,674 atoms) for 1 nanosecond:

- => 10⁶ integration steps
- => 8.4 * 10¹¹ floating point operations/step
[n(n-1)/2 interactions]

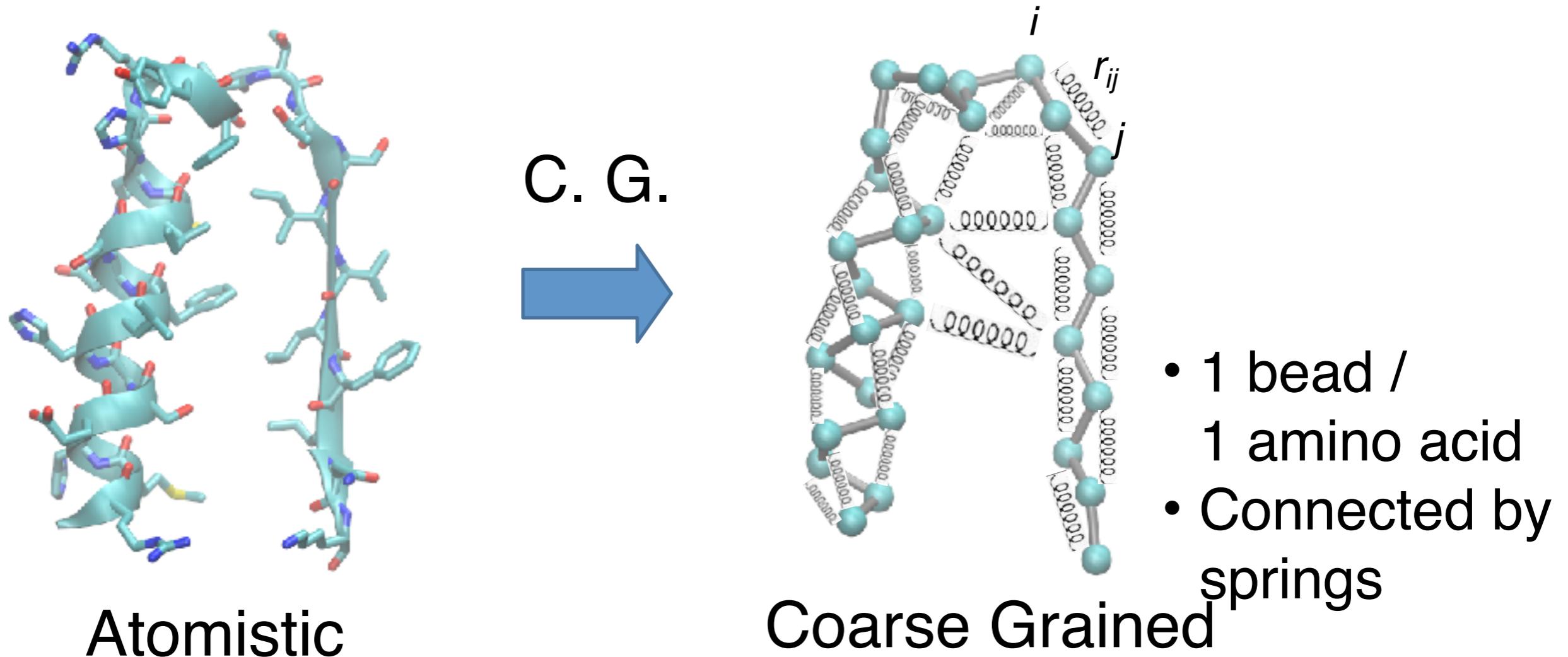
Total: 8.4 * 10¹⁷ flop
(on a 100 Gflop/s cpu: **ca 25 years!**)

... but performance has been improved by use of:

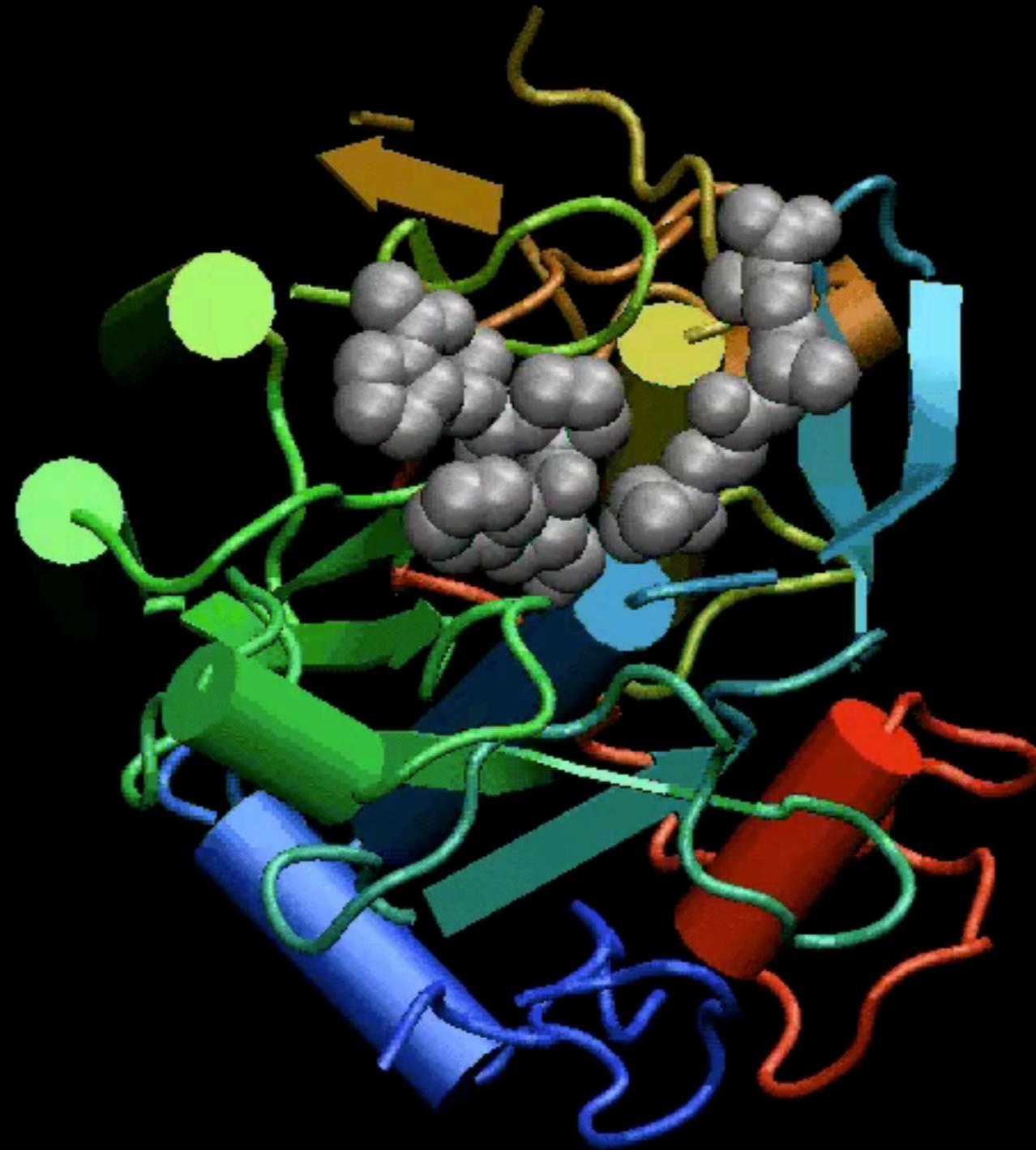
multiple time stepping	ca. 2.5 years
fast multipole methods	ca. 1 year
parallel computers	ca. 5 days
modern GPUs	ca. 1 day
(Anton supercomputer	ca. minutes)

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.



NMA models the protein as a network of elastic strings



Proteinase K

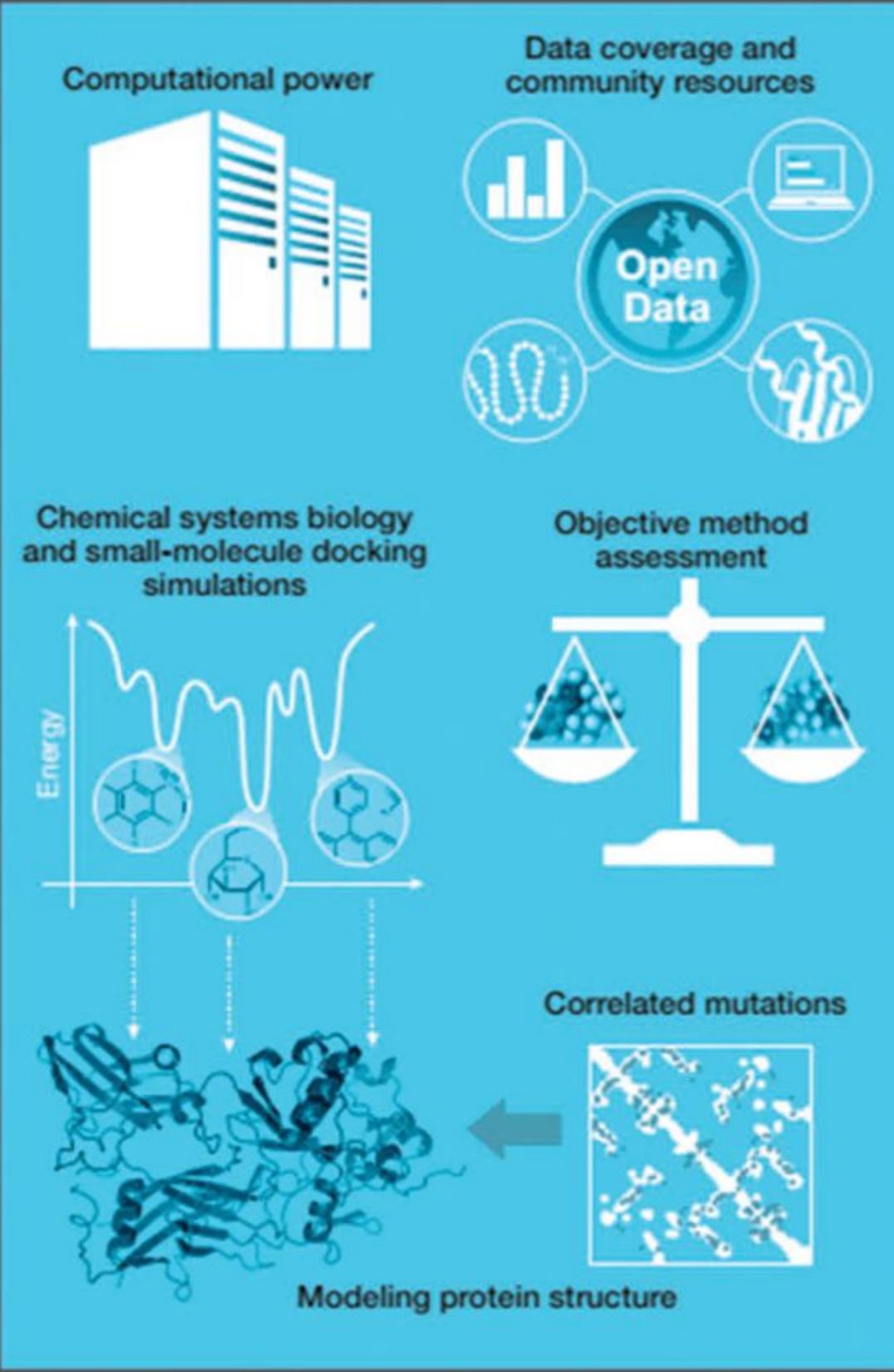
Do it Yourself!

Hand-on time!

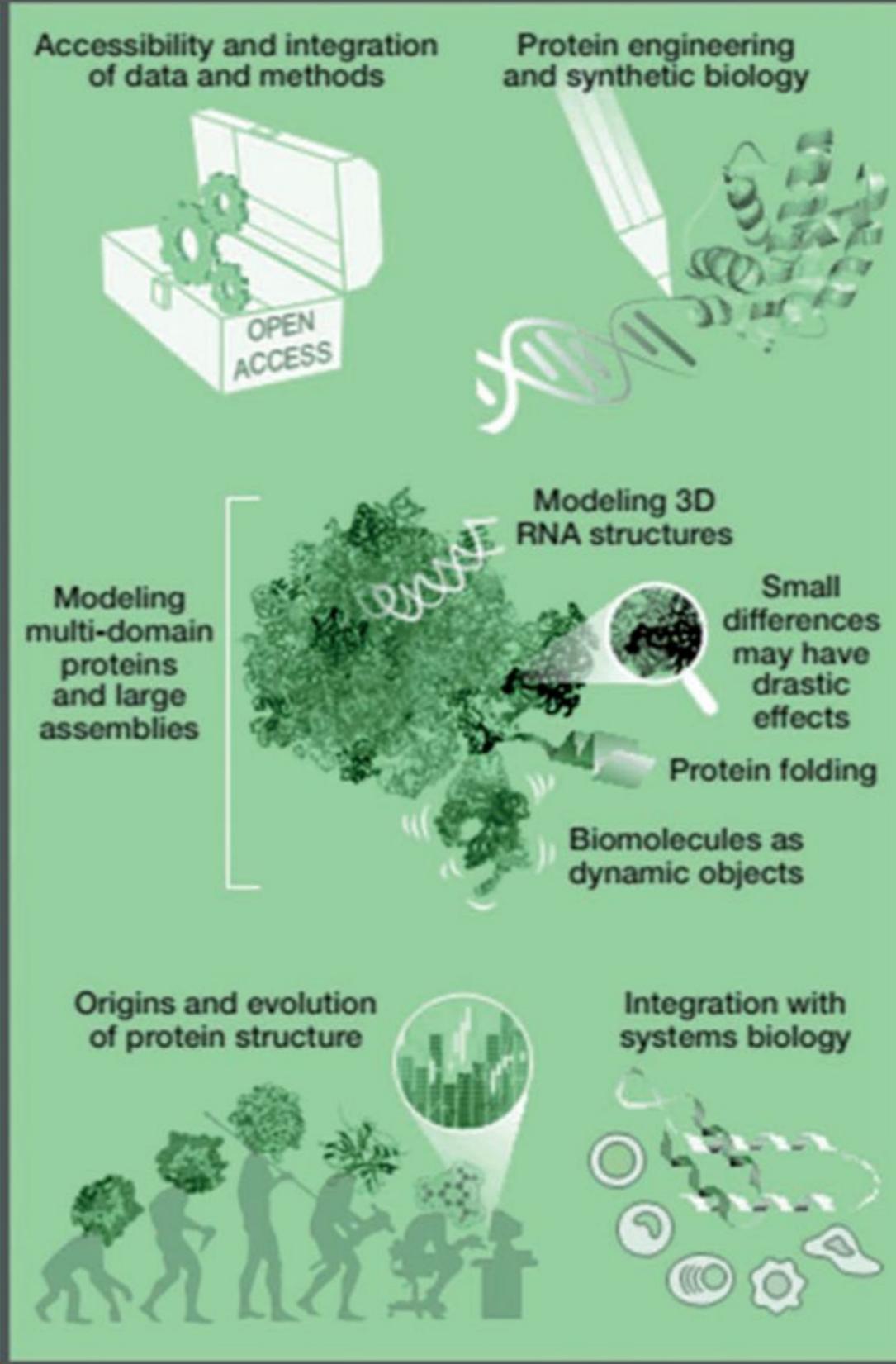
https://bioboot.github.io/bggn213_f17/lectures/#12

Focus on **section 4** exploring **PCA** and **NMA apps**

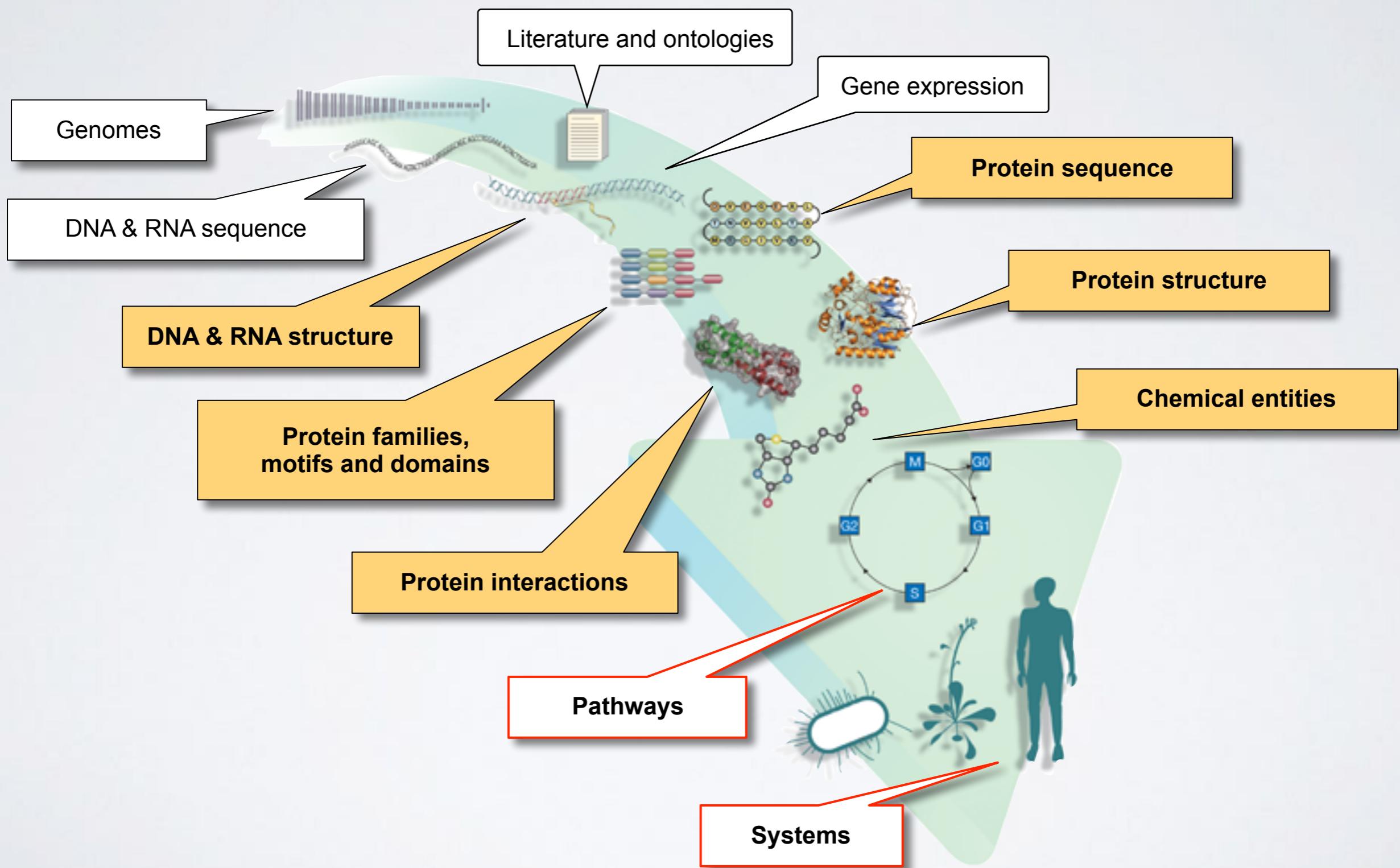
ACHIEVEMENTS



CHALLENGES



INFORMING SYSTEMS BIOLOGY?



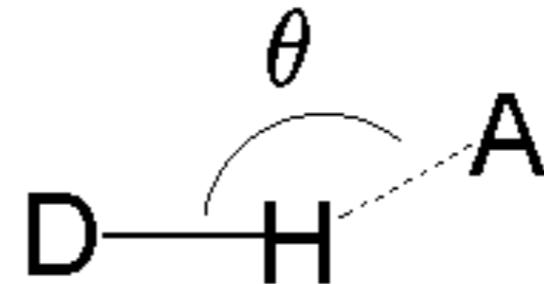
Key forces affecting structure:

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

Hydrogen-
bond donor Hydrogen-
bond acceptor



← d →

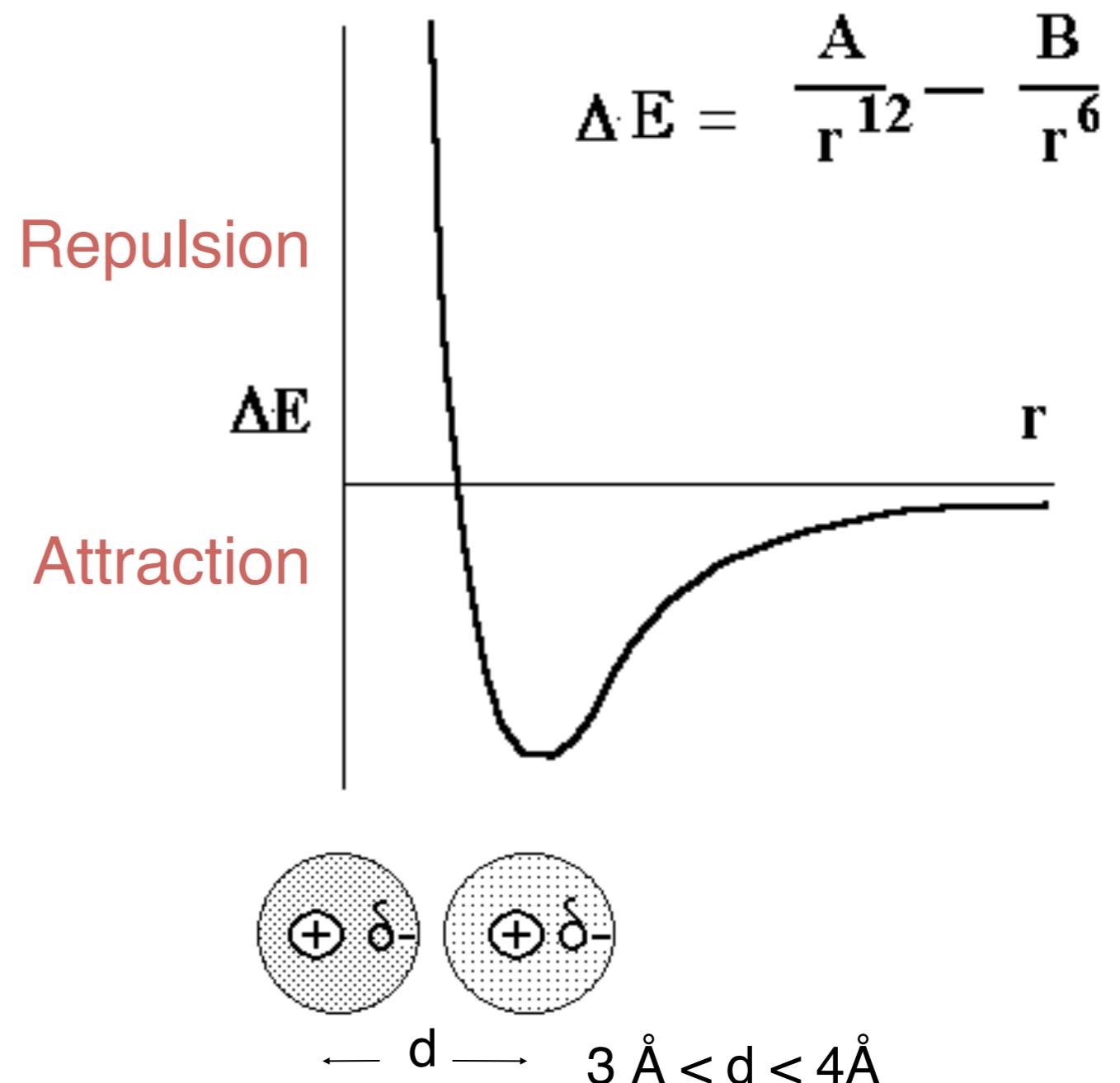


$2.6 \text{ \AA} < d < 3.1 \text{ \AA}$

$150^\circ < \theta < 180^\circ$

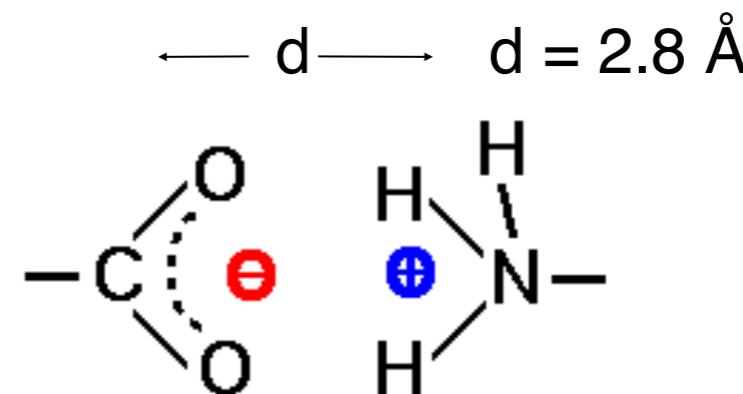
Key forces affecting structure:

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



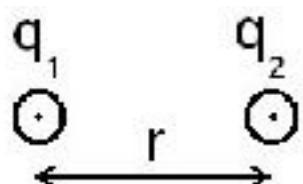
Key forces affecting structure:

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



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

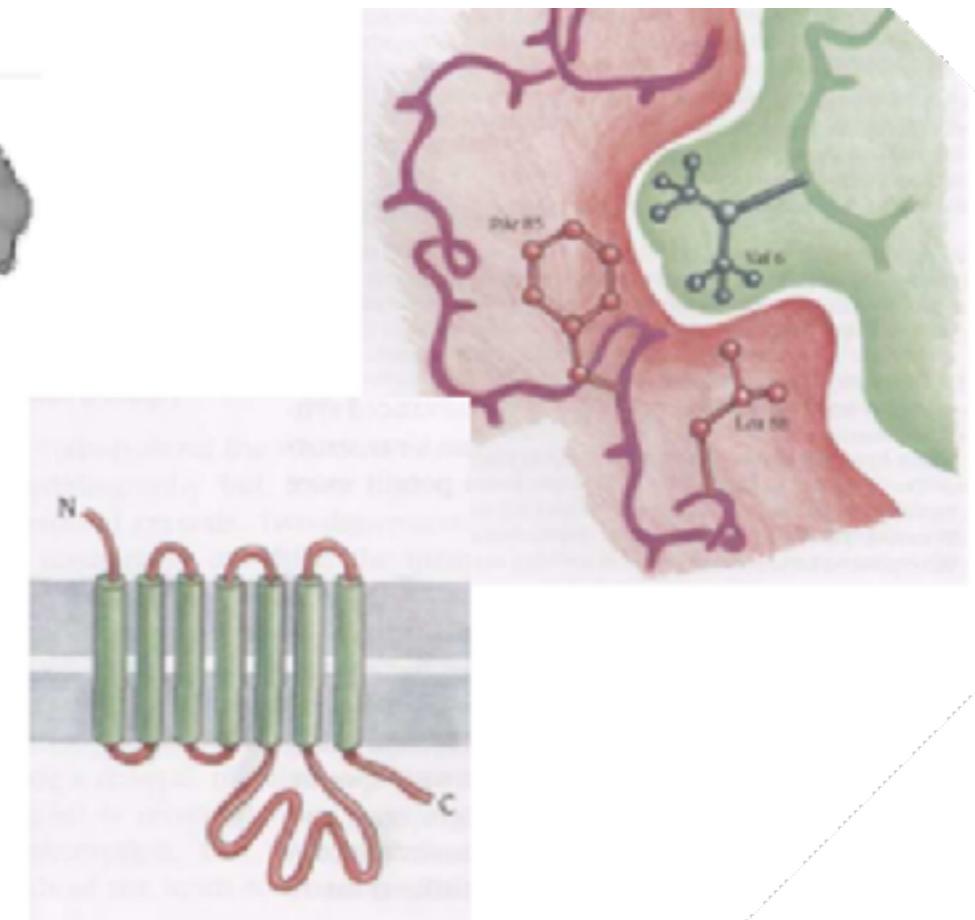
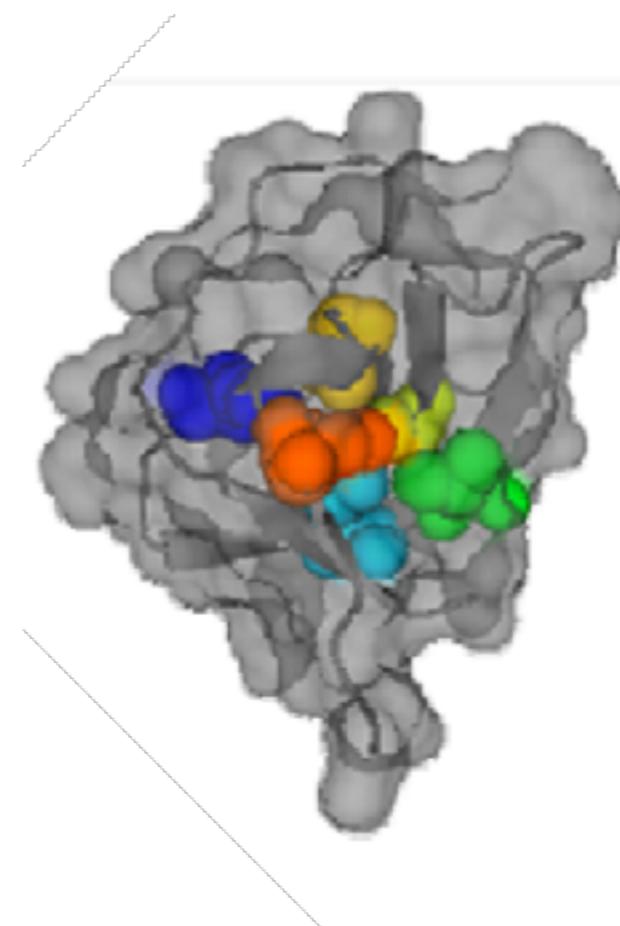
D = Dielectric constant (vacuum = 1; H₂O = 80)

q₁ & q₂ = electronic charges (Coulombs)

r = distance (Å)

Key forces affecting structure:

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



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.

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
- Introduced both structure and ligand based bioinformatics approaches for drug discovery and design