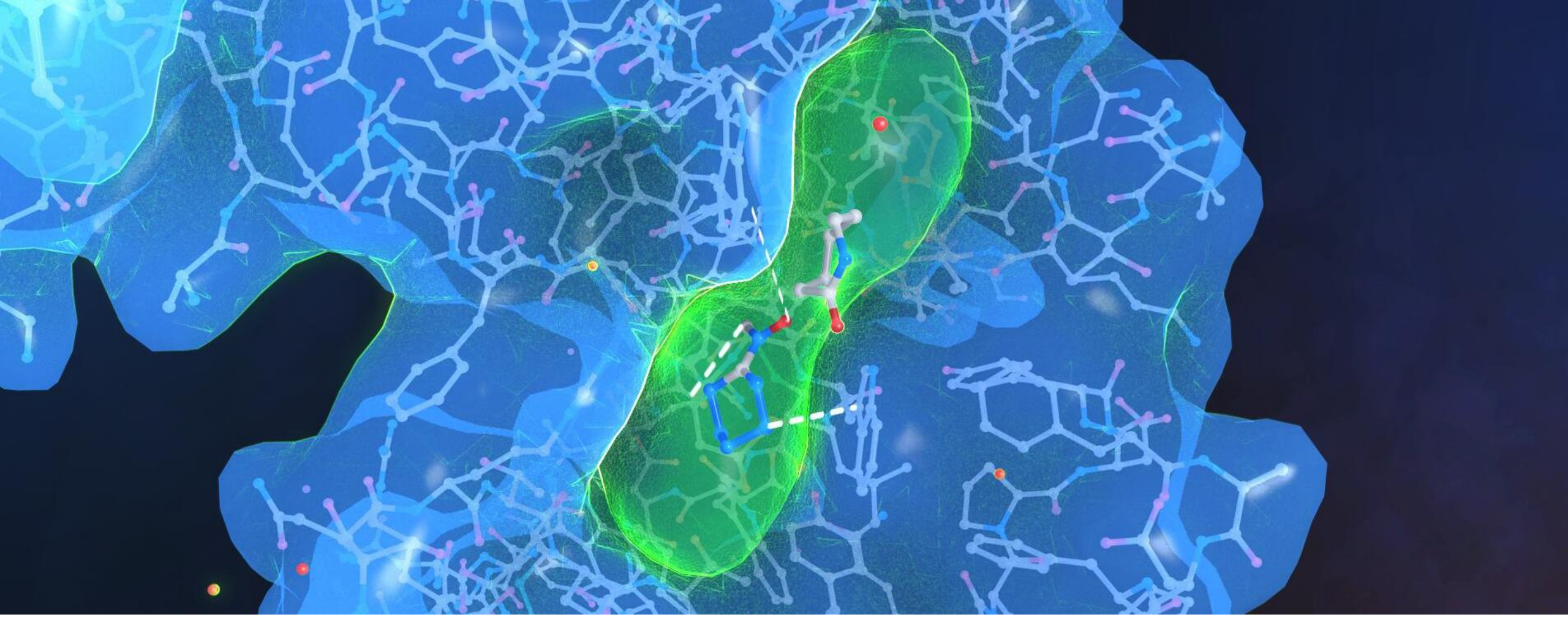


MALTAOMICS SUMMER SCHOOL

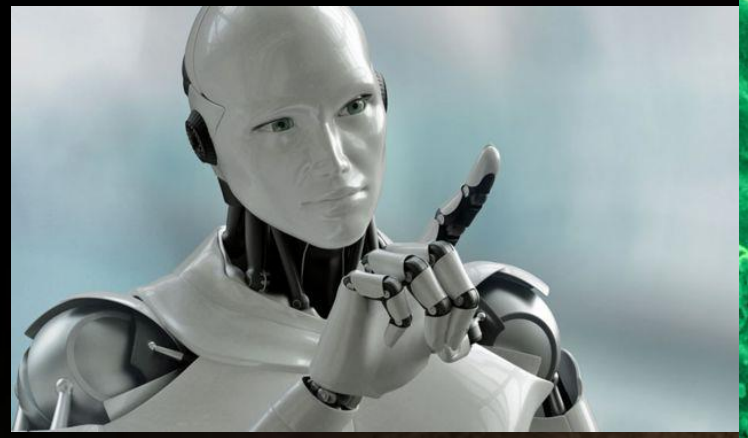
INTRODUCTION TO COMPUTER-AIDED DRUG DESIGN

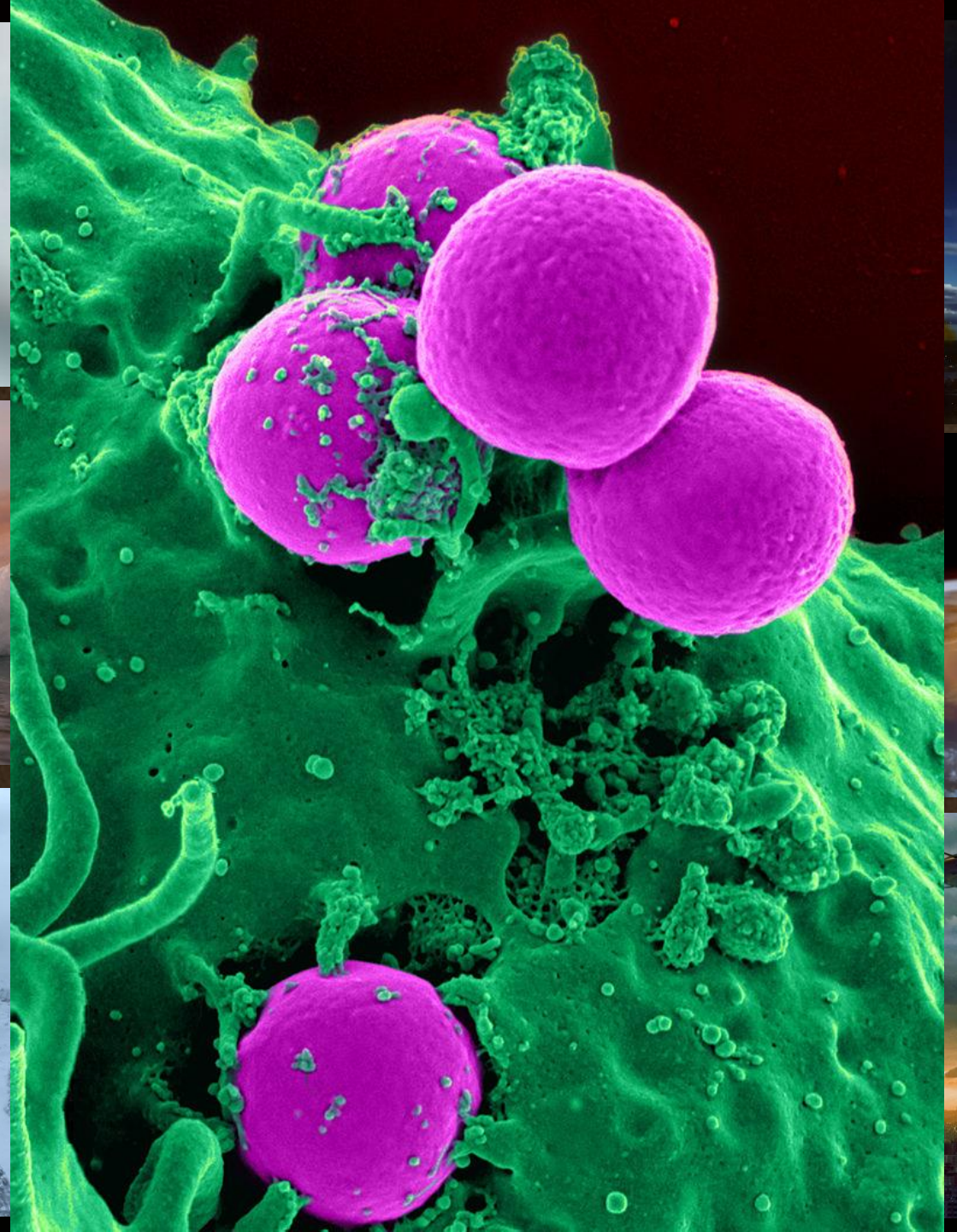
DAY 4 – 09:00-10:30

Dr Jean-Paul Ebejer
jean.p.ebejer@um.edu.mt

AGENDA

Day 4 - Thu Sep 14	
09:00 – 10:30	Introduction to Computer-Aided Drug Design
10:30 – 10:40	Break
10:40 – 11:30	Introduction to Molecular Representation
11:30 – 11:40	Break
11:40 – 13:00	Introduction to Random Forests for Virtual Screening





THINK THIS IS UNLIKELY? THINK AGAIN

- One example: Methicillin-resistant Staphylococcus aureus (MRSA)
- A growing number of infections – such as pneumonia, tuberculosis, gonorrhoea, and salmonellosis – are becoming harder to treat as the antibiotics used to treat them become less effective

World Health Organization


TIMES TRENDS
Next epidemic? Outbreak of drug-resistant bugs
World Facing Antibiotics Crunch To Fight Diseases, But Fast-Tracking Medicine Approvals A Concern
CHINA SET FOR 2ND MANNED TRIP TO SPACE
Bugs resistant to antibiotics 'will kill more than cancer'


The Miami Herald
MiamiHerald.com
WHEN DRUGS STOP WORKING
Groups must repay 'dirty money'

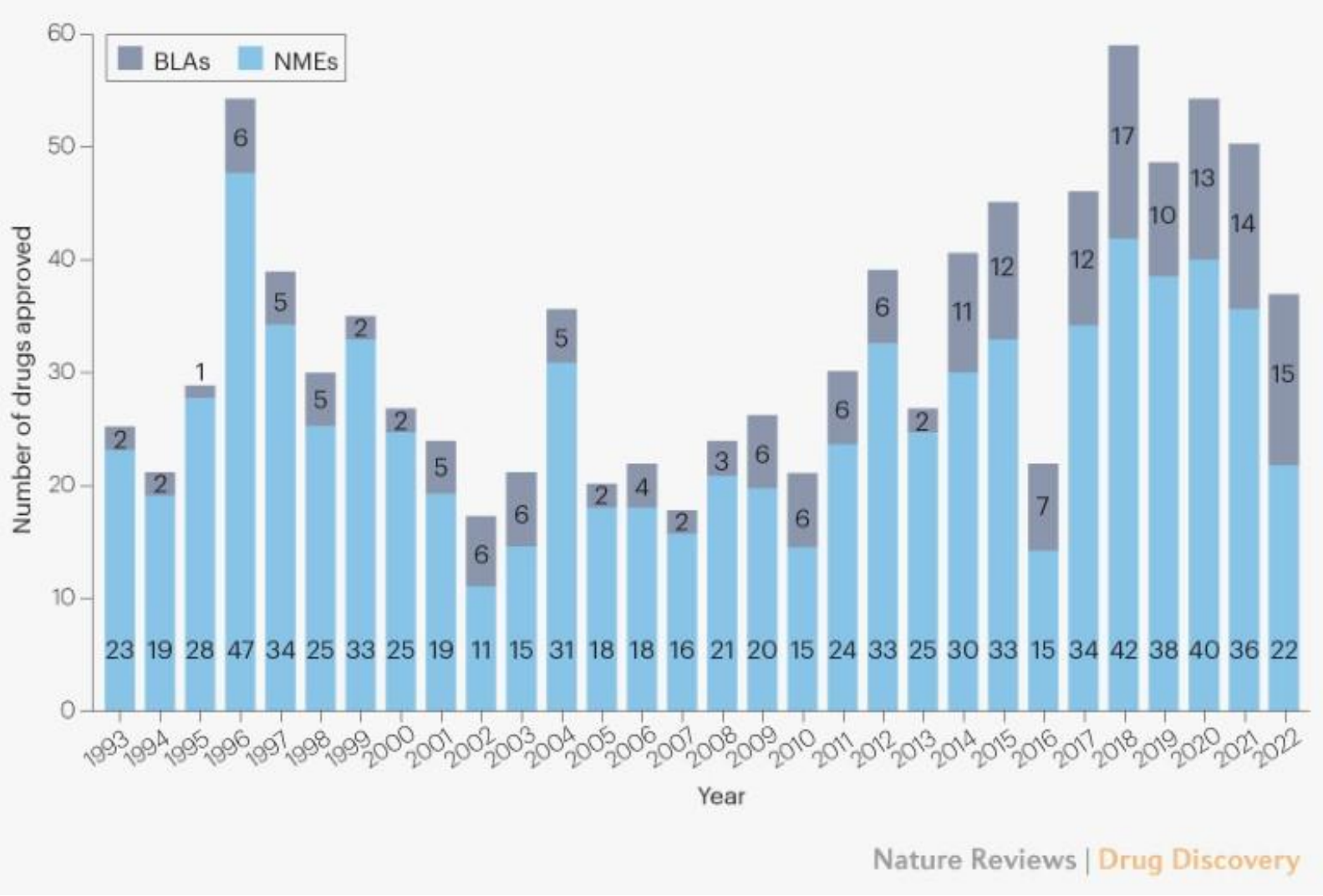

THE WORLD AT WAR
FREE DVD
GET YOUR SECOND SPECIAL BONUS DISC FREE TODAY
FREE LILY COLLECTION WORTH £25 FOR EVERY READER
Hospital infection soars by 22 per cent in just three months
1,000 SUPERBUG VICTIMS A WEEK


DAILY Mirror
EXCLUSIVE: The story that EVERY parent must read...
How I snared paedo trying to groom my girl aged 12
SUPERBUGS DEADLIER THAN CANCER
Fears 10m a year could die from infections as antibiotics may stop working


METRO
Strain of superbug 'may be now HIV'


The INDEPENDENT
Antibiotic-resistant infections spread through Europe

TAKE II - WHY COMPUTATIONAL DRUG DISCOVERY?



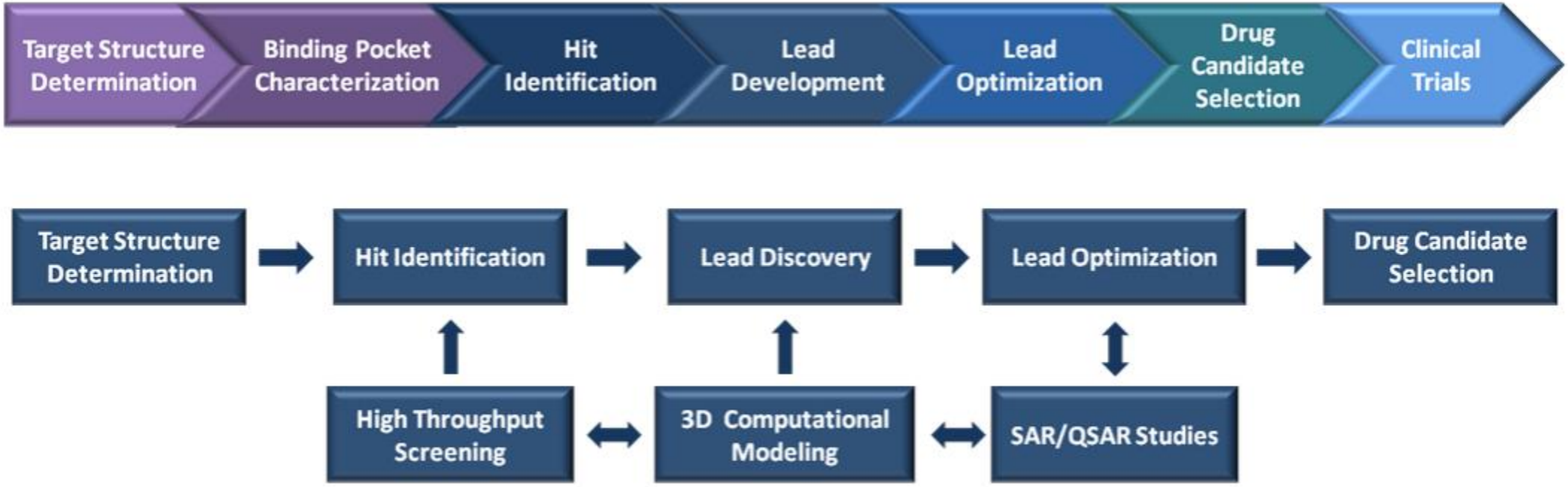
MOTIVATION

- Drug discovery is a complex challenge
- Pharmaceutical Productivity for discovering new drugs is terrible!
- Many deadly diseases still do not have an effective cure: Malaria, HIV, Hepatitis, Ebola, Meningitis, Lupus, some forms of "Cancer"
- And even many non-deadly ones; e.g. the common cold
- WHO has approx. 12,000 entries in its Classification of Diseases catalogue

WHY IS DRUG DISCOVERY HARD?

- Process which takes years (12-15 years)
- Many failure points, sometimes late in the process
- Stringent regulation (good)
- ‘Low hanging fruit’ has already been picked
- Multi-objective optimization problem
  - Simultaneous optimization of all properties make drug discovery a complex problem

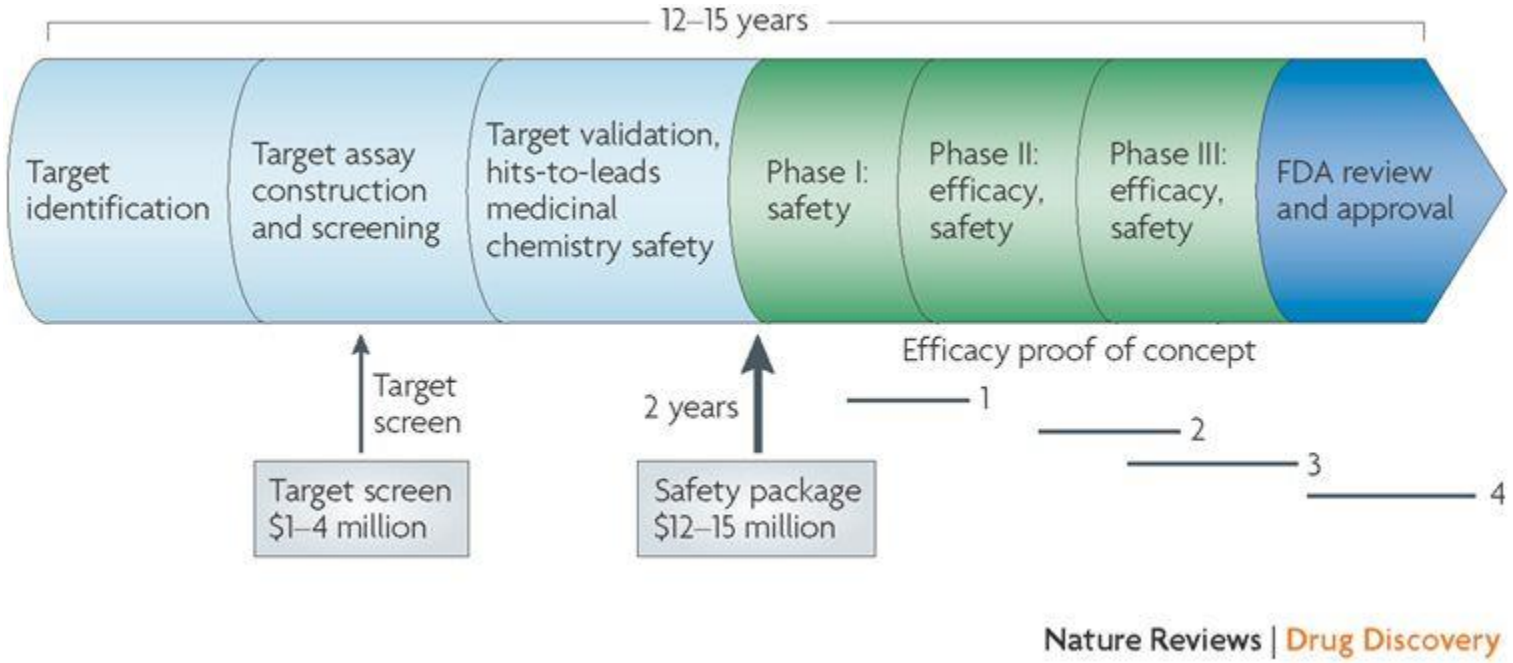
DRUG DISCOVERY PIPELINE (I)



Challenges and Opportunities for New Protein Crystallization Strategies in Structure-Based Drug Design, Jessica Grey and David Thompson, Expert Opin Drug Discov. 2010 Nov 1; 5(11): 1039–1045.

DRUG DISCOVERY PIPELINE – TIMELINE & COST

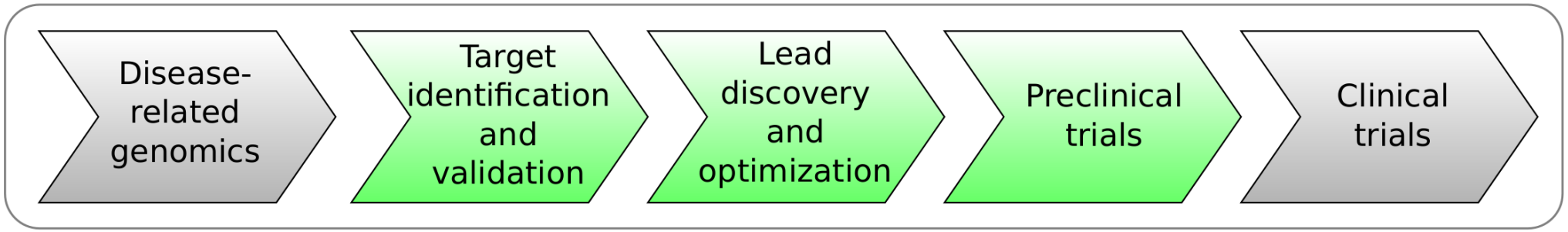
- Spoiler Alert: Takes 12-15 years!



Pharmacogenetics in drug discovery and development: a translational perspective, Allen D. Roses, Nature Reviews Drug Discovery volume 7, pages 807–817 (2008)

CADD MAIN APPLICATION AREAS

Drug Discovery and Development Process



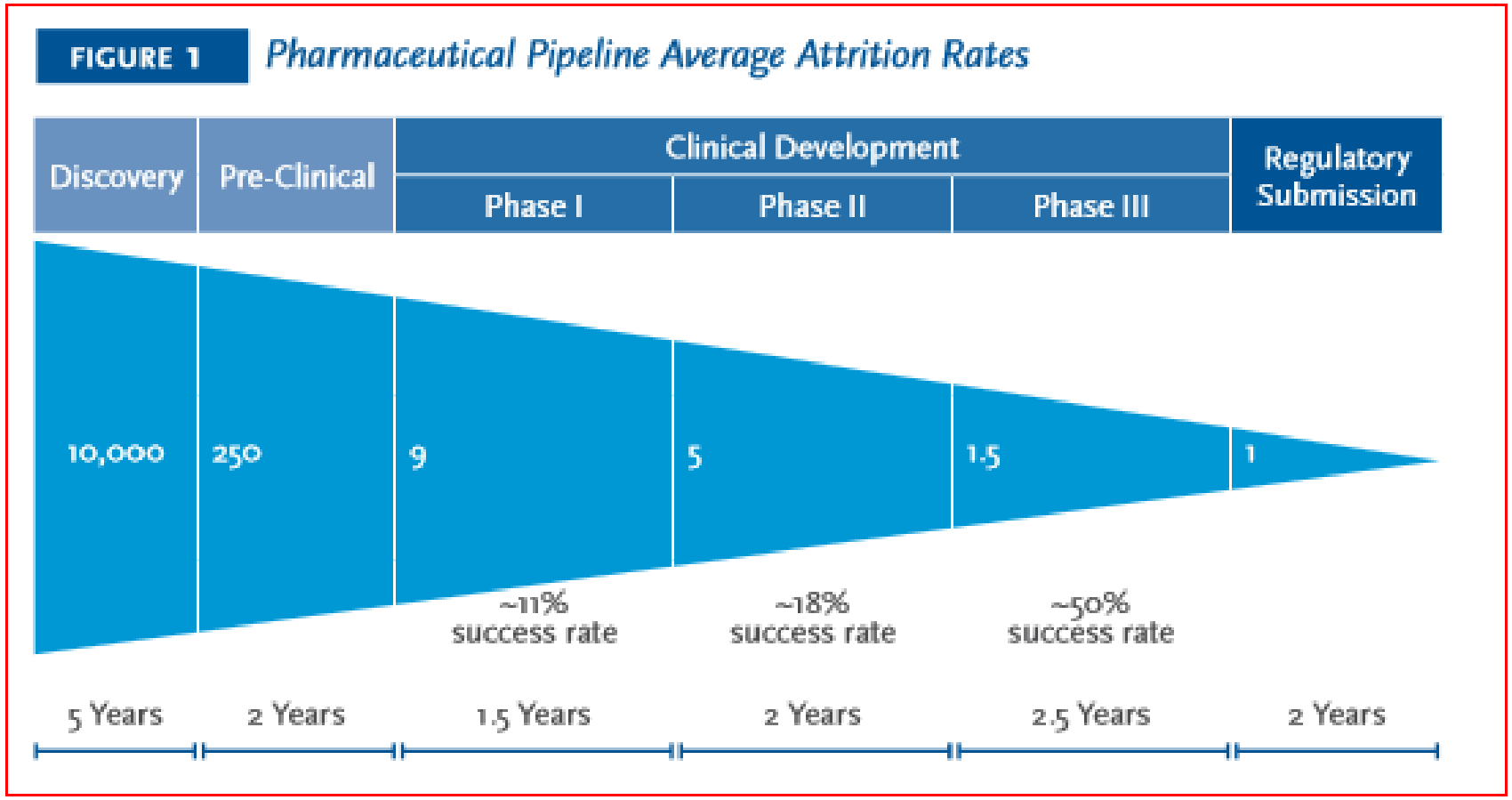
Computational Approaches

- Bioinformatics
- Reverse docking
- Pharmacophore mapping
- Protein structure prediction
- Target druggability
- Chemical probe design

- Molecular Dynamics
- Molecular docking
- Pharmacophore modelling
- Similarity searches (Fingerprints)
- De novo design
- Virtual library design
- QSAR and 3D QSAR
- Sequence-based method

- In silico ADMET prediction
- Physiologically-based pharmacokinetic (PBPK) simulations

DRUGS ATTRITION RATES



COMPUTATIONAL DRUG DISCOVERY

- Counter-intuitively, Computational Drug Discovery helps us to FAIL (fast/early)

Cheaper to fail

In silico

In vitro

In vivo

GROWING ROLE OF COMPUTATION IN DRUG DISCOVERY

- Role of computational models is to increase prediction based on existing knowledge

BBC Sign in | News | Sport | Weather | iPlayer | TV

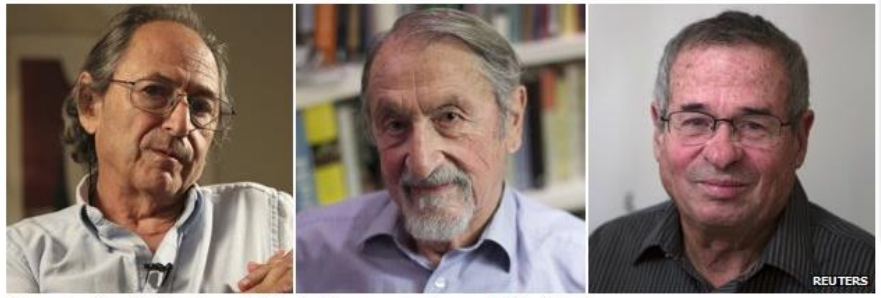
NEWS SCIENCE & ENVIRONMENT

Home | World | UK | England | N. Ireland | Scotland | Wales | Business | Politics | Health | Education | Sci/Env

9 October 2013 Last updated at 11:36

Computer chemists win Nobel prize

By James Morgan and Jonathan Amos
Science reporters, BBC News



The work of Levitt, Karplus and Warshel has spawned a worldwide industry

The Nobel Prize in chemistry has gone to three scientists who "took the chemical experiment into cyberspace".

Related Stories

PRESSMEDDELANDE
Press release
9 October 2013

The Nobel Prize in Chemistry 2013

The Royal Swedish Academy of Sciences has decided to award the Nobel Prize in Chemistry for 2013 to

Martin Karplus — Université de Strasbourg, France and Harvard University, Cambridge, MA, USA
Michael Levitt — Stanford University School of Medicine, Stanford, CA, USA
Arieh Warshel — University of Southern California, Los Angeles, CA, USA

"for the development of multiscale models for complex chemical systems"

The computer — your Virgil in the world of atoms

Chemists used to create models of molecules using plastic balls and sticks. Today, the modelling is carried out in computers. In the 1970s, Martin Karplus, Michael Levitt and Arieh Warshel laid the foundation for the powerful programs that are used to understand and predict chemical processes. Computer models mirroring real life have become crucial for most advances made in chemistry today.

Chemical reactions occur at lightning speed. In a fraction of a millisecond, electrons jump from one atomic to the other. Classical chemistry has a hard time keeping up; it is virtually impossible to experimentally map every little

This year's Nobel Laureates in chemistry took the best from both worlds and devised methods that use both classical and quantum physics. For instance, in simulations of how a drug couples to its target protein in the body, the computer performs quantum theoretical calculations on those atoms in the target protein that interact with the drug. The rest of the large protein is simulated using less demanding classical physics.

Today the computer is just as important a tool for chemists as the test tube. Simulations are so realistic that they predict the outcome of traditional experiments.

> Today the computer is just as important a tool for chemists as the test tube. Simulations are so realistic that they predict the outcome of traditional experiments.

CADD IN LEAD DISCOVERY AND OPTIMIZATION

- **Virtual screening** - libraries of small-molecules are searched to identify inhibitors against some biological target
- ***De novo* design** - small-molecule inhibitors with novel molecular structures are assembled 'from scratch' either in an atom-by-atom fashion or by using larger building blocks (e.g. a benzene ring)
  - In contrast to virtual screening, where a library of 'whole' molecules is searched for actives

DE NOVO DESIGN

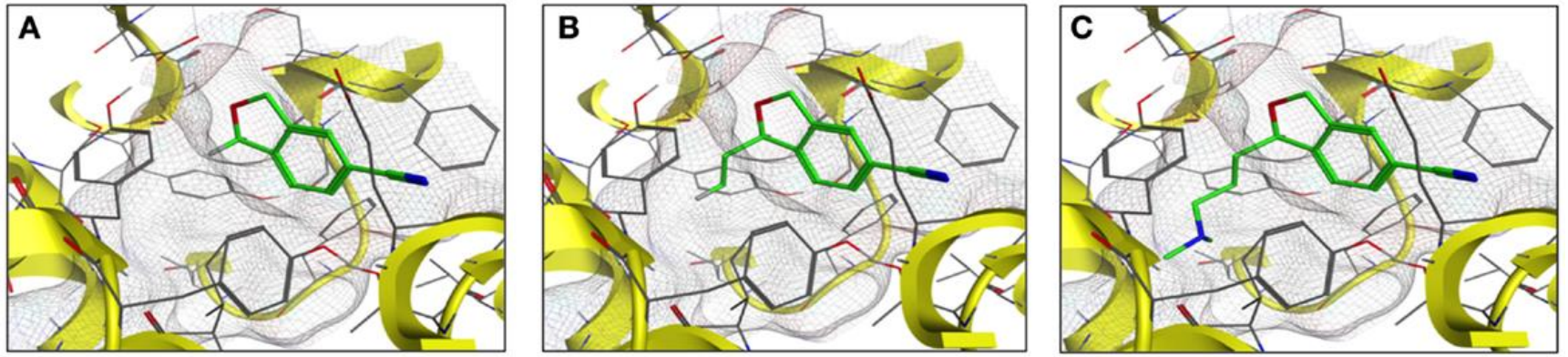
- Steps
  1. Strategy of how to assemble the compound must be selected (either atom-based or fragment-based)
  2. A scoring function is required which evaluates the molecule in its current state
  3. An algorithm which systematically visits the search space for the next molecular modification is needed
- A common criticism of *de novo* methods is that they do not always produce compounds which are amenable to chemical synthesis

DE NOVO DESIGN - STRATEGIES

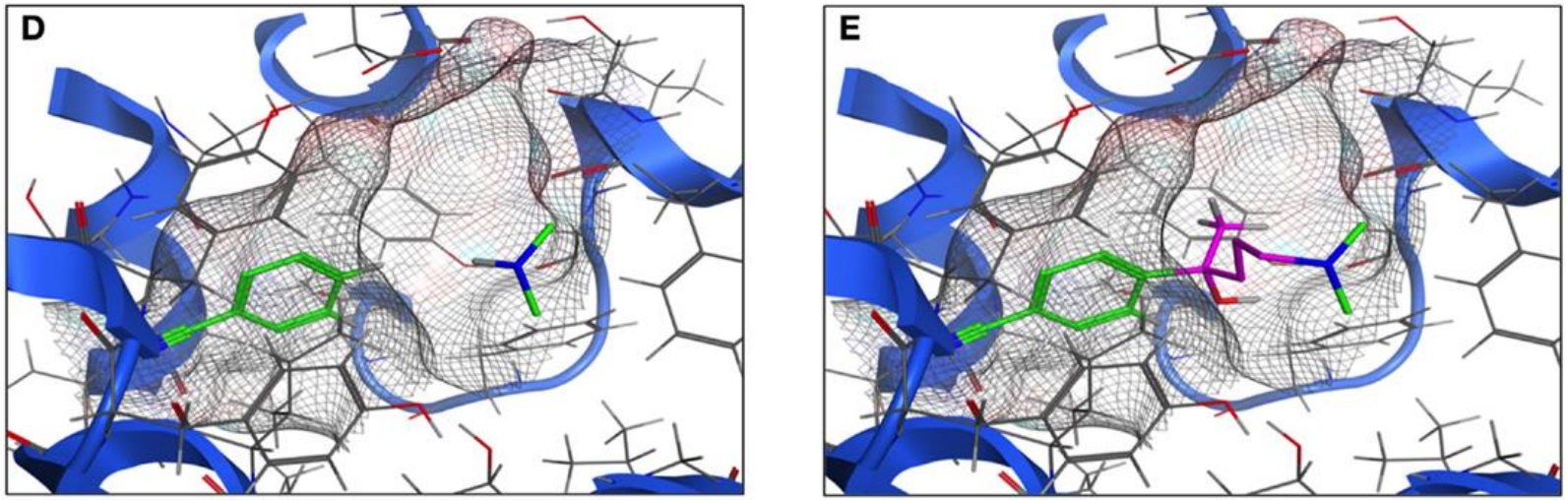
- **Linking** - multiple fragments docked in distinct parts of the protein pocket are then linked together using a linker or scaffold molecular fragment
- **Growing** - to anchor a fragment in the pocket and explore the rest of the pocket by adding more fragments which optimize binding interactions

DE NOVO DESIGN - STRATEGIES

Fragment growth



Fragment linking



A role for fragment-based drug design in developing novel lead compounds for central nervous system targets, *Wasko et al.*, 2015.

HOW DO SMALL MOLECULE DRUGS WORK? (I)

- A small-molecule is said to be active when:

"it binds to a receptor, typically a protein, and either elicits a response (agonist) or blocks that response (antagonist)"

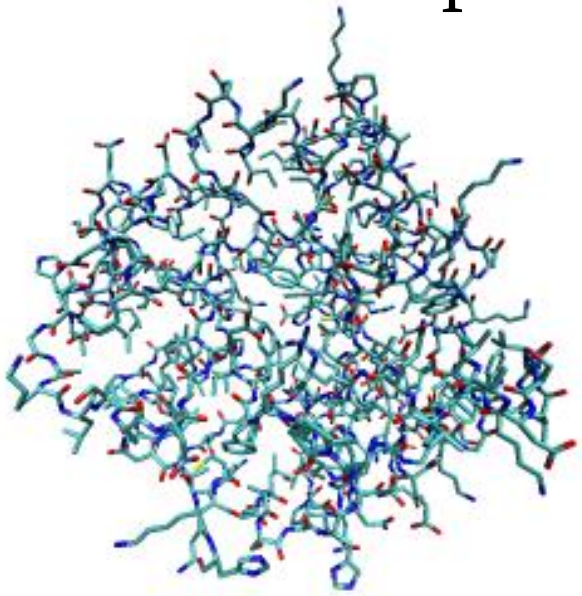
- So, either a small-molecule stops the protein from carrying out a function
- Or, it makes the protein do something

HOW DO DRUGS WORK? (II)

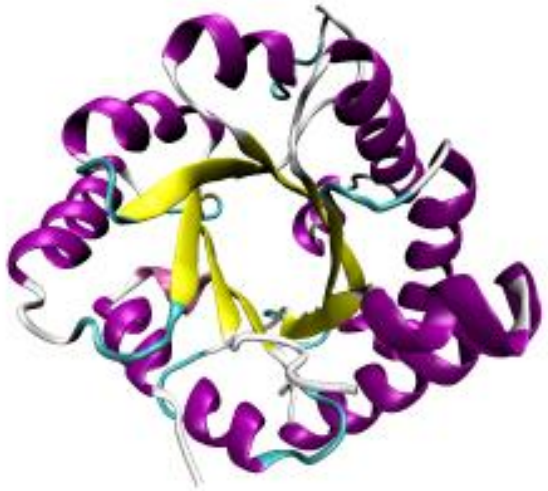
- Proteins are large (macro) molecules, which carry out a specific function – “basic building blocks of life”
- Used for:
  - protection (anti-bodies)
  - facilitate chemical reactions (enzymes)
  - signalling
  - storage
  - transport
  - structure (forms part of larger cellular structure)
- Around 20,000 protein coding genes in DNA

HOW DO DRUGS WORK? (II)

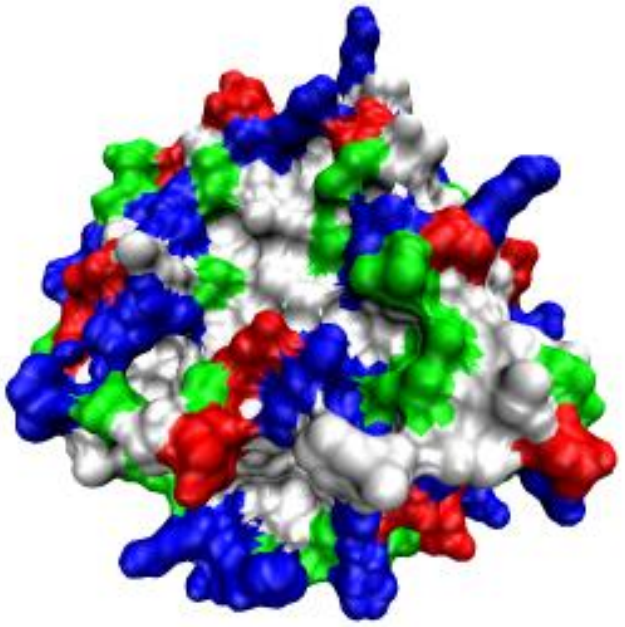
- So, what does a protein look like?
  - 3 common representations



Sticks (all atoms without hydrogens)

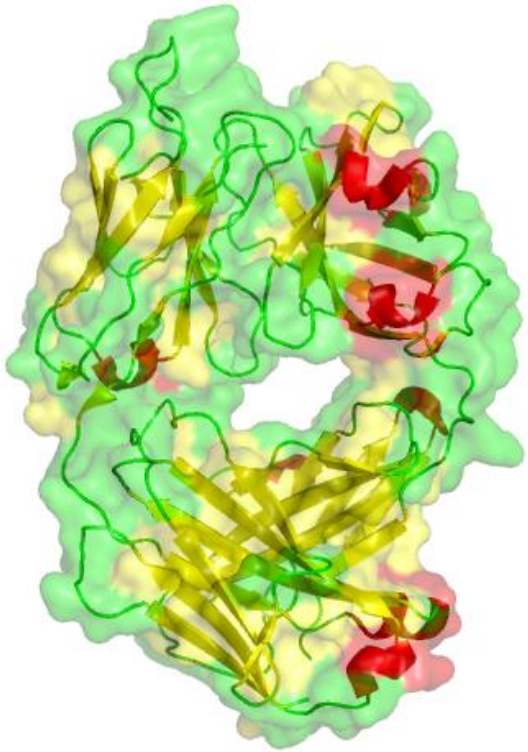


Cartoon (highlights the secondary structure of the protein)



Surface (represents the solvent-accessible surface of the protein)

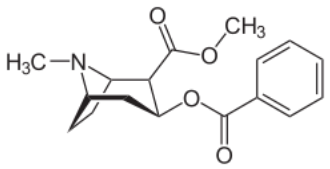
HOW DO (SMALL-MOLECULE) DRUGS WORK? (III)



2AJV

Protein

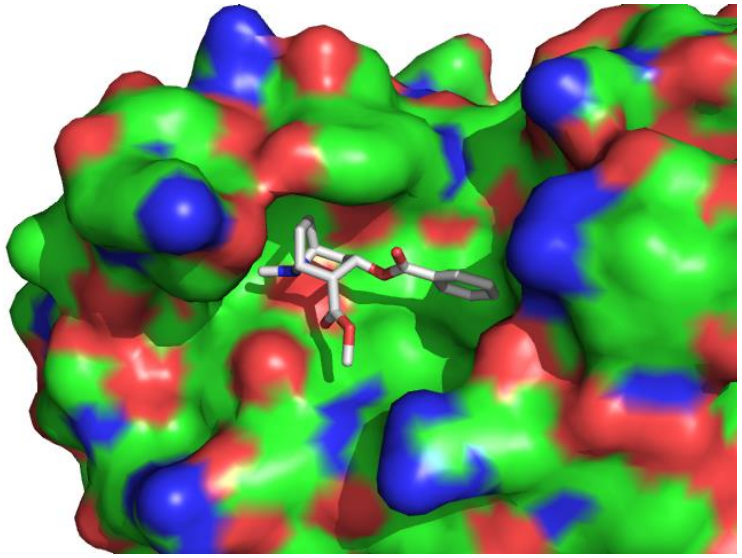
+



benzoylmethylecgonine

Small-molecule

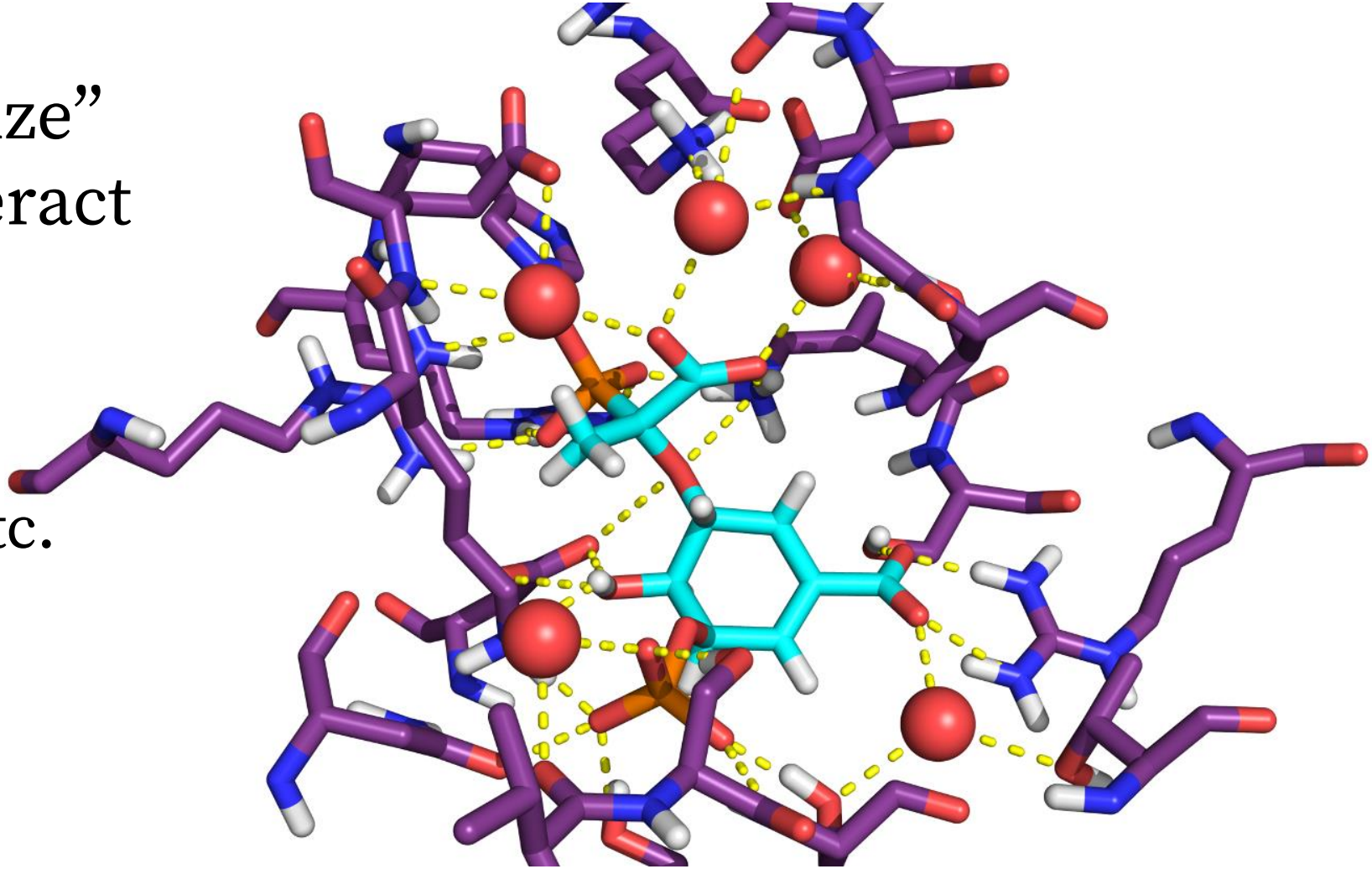
=



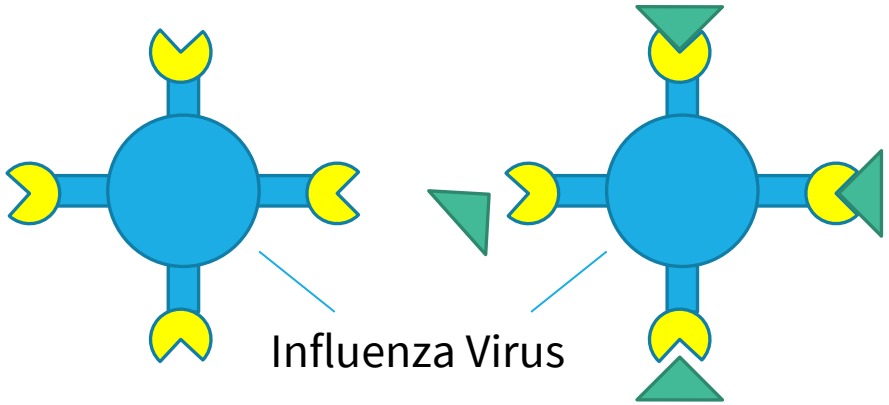
Protein + Small- Molecule Complex

MOLECULES INTERACT

- Molecules “recognize” each other and interact in many ways
  - Steric fit, polar/non-polar interactions, charges, H bonds, etc.
- This is how drugs work!

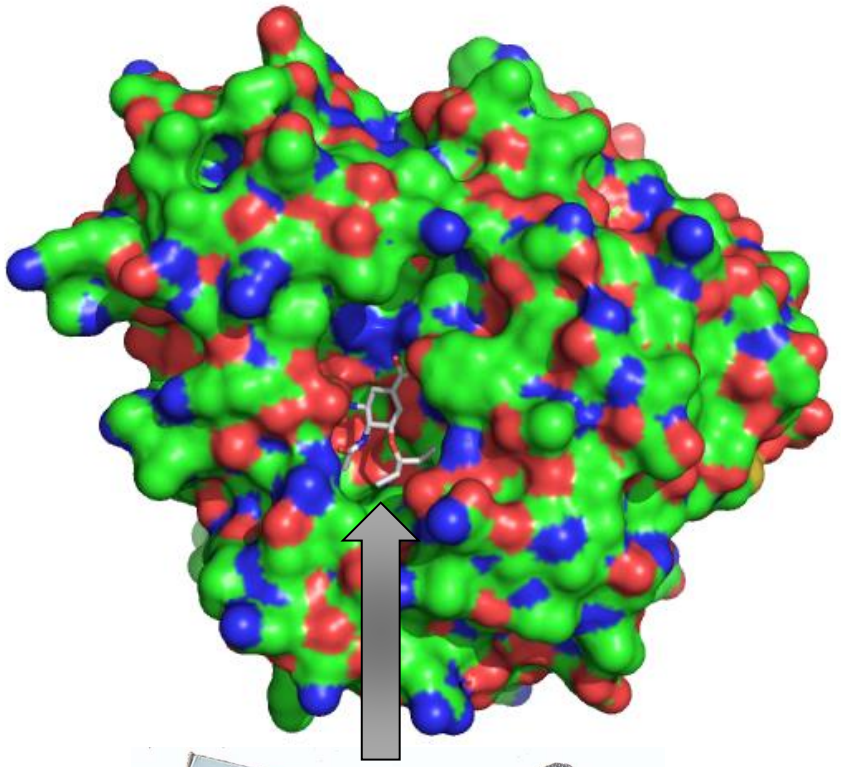


EXAMPLE: INFLUENZA

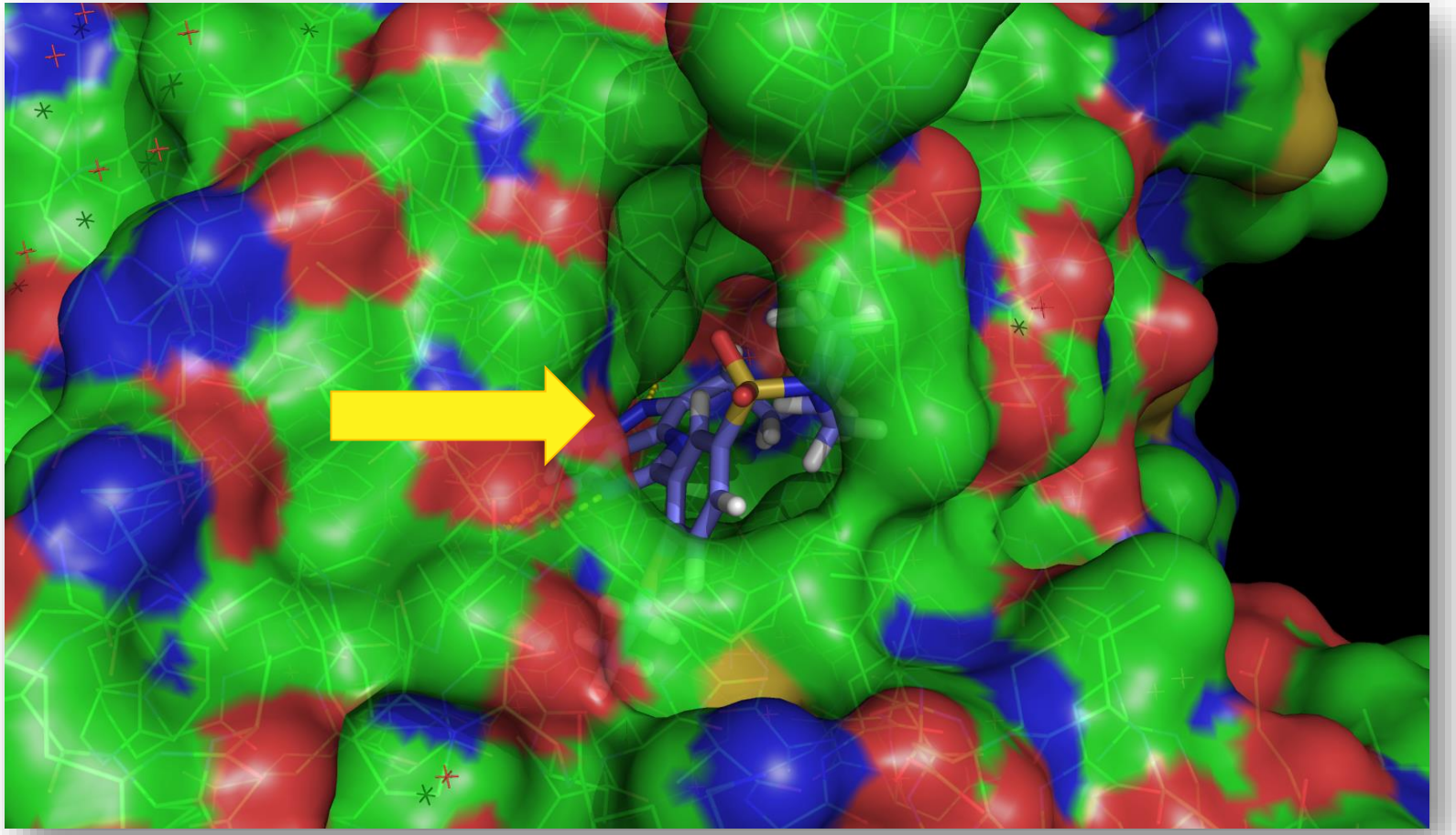


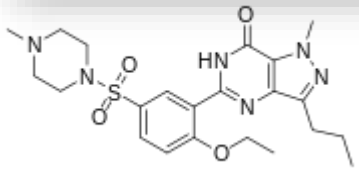
Neuraminidase Protein (part of the Influenza Virus which allows the virus to get in and out of the cell to replicate)

Oseltamivir (Trade name: Tamiflu, Roche)



ANOTHER DRUG EXAMPLE — 2H42; SILDENAFIL





Crowdsourcing Yields a New Standard for Kinks in Protein Helices

Henry R. Wilman[†], Jean-Paul Ebejer[†], Jiye Shi[‡§], Charlotte M. Deane[†], and Bernhard Knapp[*†]
[†] Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, U.K.
[‡] UCB Celltech, a branch of UCB Pharma S. A., 208 Bath Road, Slough SL1 3WE, U.K.
[§] Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800, China

J. Chem. Inf. Model., **2014**, *54* (9), pp 2585–2593
DOI: 10.1021/ci500403a
Publication Date (Web): August 20, 2014

Resolving Protein Structures computationally is still an open problem!

Journal List › Nucleic Acids Res › v.41(Web Server issue); 2013 Jul › PMC3692111

Nucleic Acids Research

Nucleic Acids Res. 2013 Jul; 41(Web Server issue): W379–W383. PMCID: PMC3692111
Published online 2013 May 2. doi: 10.1093/nar/gkt331

Memoir: template-based structure prediction for membrane proteins

Jean-Paul Ebejer,[1] Jamie R. Hill,[1] Sebastian Kelm,[1] Jiye Shi,[2] and Charlotte M. Deane[1,*]

Author information ► Article notes ► Copyright and License information ►

ABSTRACT Go to:

Membrane proteins are estimated to be the targets of 50% of drugs that are currently in development, yet we have few membrane protein crystal structures. As a result, for a membrane protein of interest, the much-needed structural information usually comes from a homology model. Current homology modelling software is optimized for globular proteins, and ignores the constraints that the membrane is known to place on protein structure. Our Memoir server produces homology models using alignment and coordinate generation software that has been designed specifically for transmembrane proteins. Memoir is easy to use, with the only inputs being a structural template and the sequence that is to be modelled. We provide a video tutorial and a guide to assessing model quality. Supporting data aid manual refinement of the models. These data include a set of alternative conformations for each modelled loop, and a multiple sequence alignment that incorporates the query and template. Memoir works with both α-helical and β-barrel types of membrane proteins and is freely available at http://opig.stats.ox.ac.uk/webapps/memoir.

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS
Explore this journal >

Article

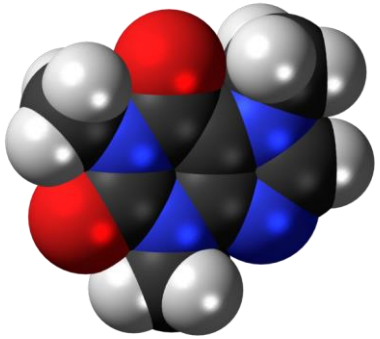
Fragment-based modeling of membrane protein loops: Successes, failures, and prospects for the future

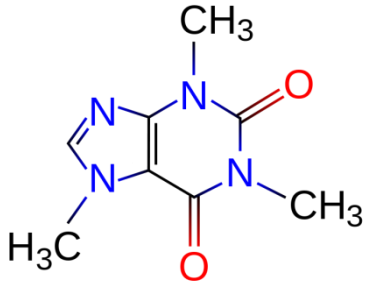
Sebastian Kelm ✉, Anna Vangone, Yoonjoo Choi, Jean-Paul Ebejer, Jiye Shi, Charlotte M. Deane

First published: 16 October 2013 Full publication history
DOI: 10.1002/prot.24299 View/save citation

WHAT DOES A DRUG LOOK LIKE?

- Various ways how to represent the same compound (molecule)
- An Example: **Caffeine** (or 1,3,7-trimethyl-3,7-dihydro-1H-purine-2,6-dione if you prefer)
- Chemical formula (1D): $C_8H_{10}N_4O_2$
- SMILES String (1D): Cn1cnc2c1c(=O)n(c(=O)n2C)C
- 2D Representation in 3D (Graph)





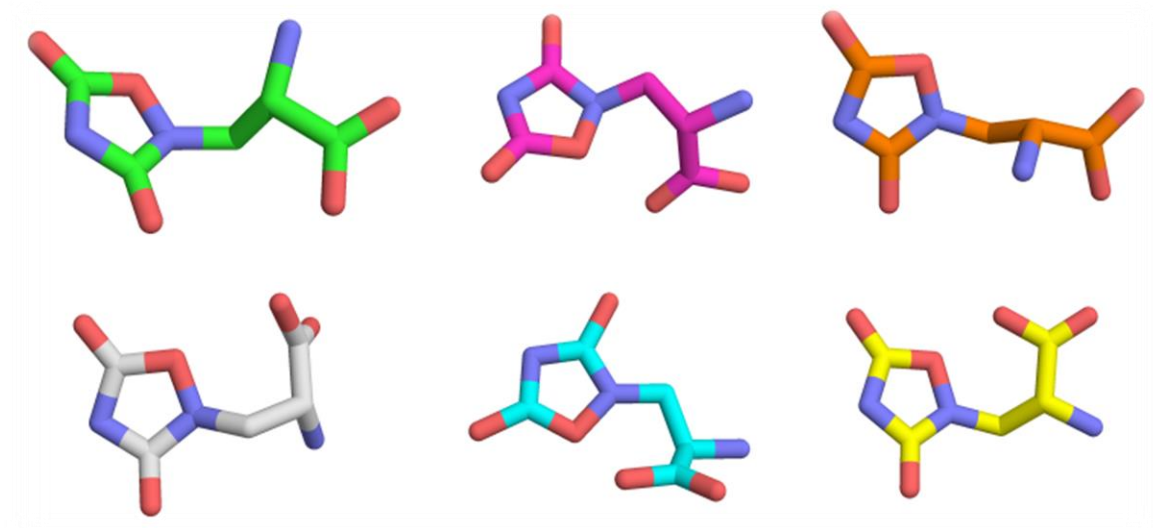
HOW MANY DRUG-LIKE MOLECULES ARE THERE?

- Drug-like chemical space estimated to be anywhere between 10^9 and 10^{180} molecules. Most often quoted figure is 10^{60} (more than the number of atoms on earth)*
- Enumeration of all possible molecules up to 17 atoms of C,N,S,O GDB-17 – 166.4 billion molecules (average drug has about 35-50 heavy atoms)
- These are molecules and do not include their conformers!

* See Table 1 from "Estimation of the size of drug-like chemical space based on GDB-17 data", Polishchuk et al. 2013.

BUT WE LIVE IN A 3D WORLD

- A molecule can take on many different shapes (“conformers”)
- Conformational space may be very large, and is a function of number of rotatable bonds
- At a resolution of 1 degree – conformer space is 360^n, where n is number of rotatable bonds (360^5 = 6,046,617,600,000)

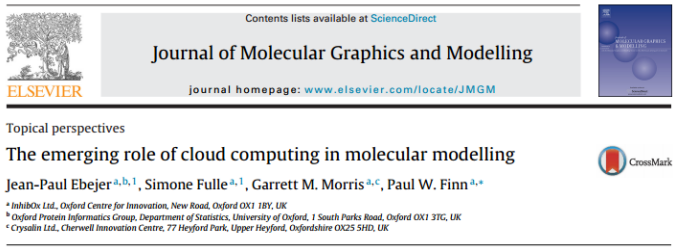


PDB code - QUS

A PRACTICAL EXAMPLE

Now we need algorithms to search this big data!

- Theoretically, each time we search for a new drug we should explore all of this space
- Luckily, we can prune this space
  - Just consider "commercially available" molecules and easily synthesised molecules
  - Use common drug-like filters (e.g. Lipinski's "Rule of 5" molecular weight < 500Da, log P < 5 etc.)
- We built a database in the cloud (Amazon AWS):
  - 28 million molecules;
  - at most 200 low-energy conformers per molecule (max: 5.6 x 10^9 3D conformers)
  - 6 TB of data


Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Topical perspectives

The emerging role of cloud computing in molecular modelling

Jean-Paul Ebejer[a,b,1], Simone Fulle[a,1], Garrett M. Morris[a,c], Paul W. Finn[a,*]

[a] InhibOx Ltd., Oxford Centre for Innovation, New Road, Oxford OX1 1BY, UK
[b] Oxford Protein Informatics Group, Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK
[c] Crysalin Ltd., Cherwell Innovation Centre, 77 Heyford Park, Upper Heyford, Oxfordshire OX25 5HD, UK

WHAT IS VIRTUAL SCREENING?

- Using computers to find small-molecule inhibitors
  - (Note: A few years need to pass for this to become a drug)
- Find a small-molecule which has a therapeutic effect against some condition/disease
- Two main kinds:
  - Structure-based virtual screening (SBVS)
  - Ligand-based virtual screening (LBVS)

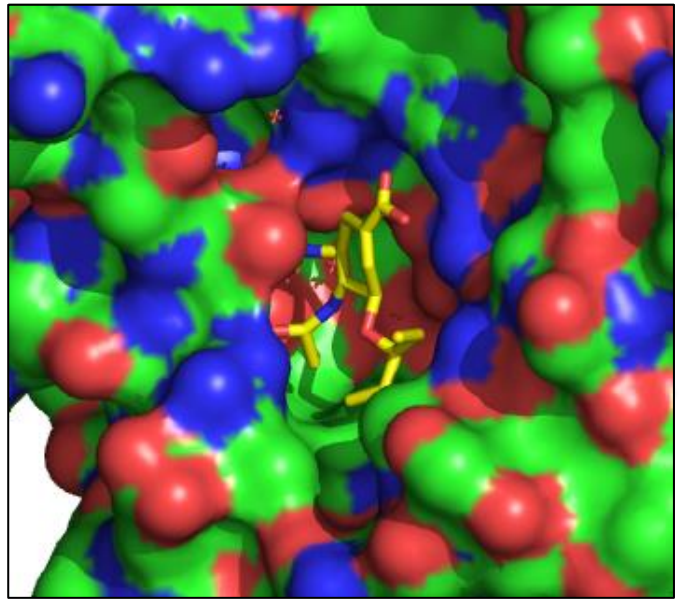
STRUCTURE BASED VIRTUAL SCREENING (SBVS)

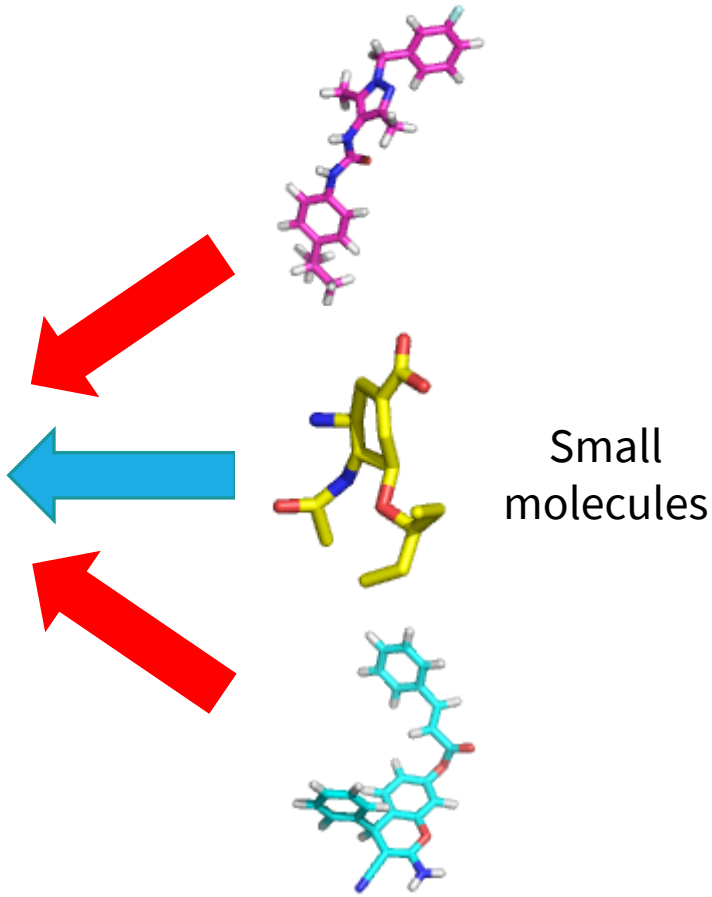
- Three things required:
  - A protein structure (hence structure-based)
  - A library of small molecules we want to test
  - A docking protocol

SBVS - DOCKING

- Lock and key analogy

Protein

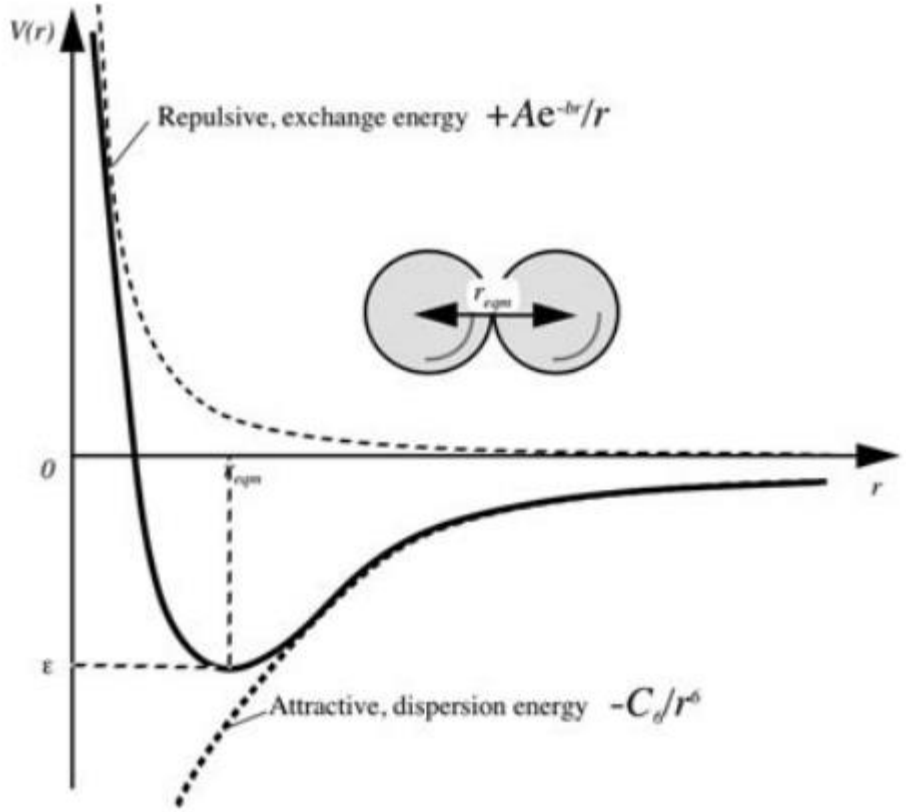




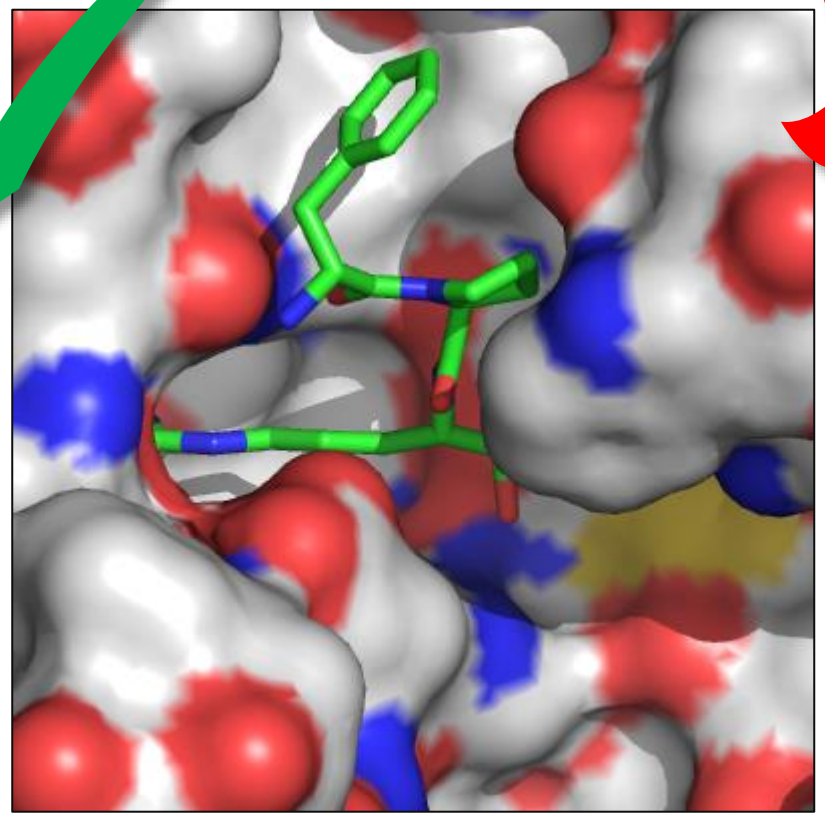
Small molecules

WHAT DETERMINES BINDING OF A SMALL-MOLECULE TO A PROTEIN?

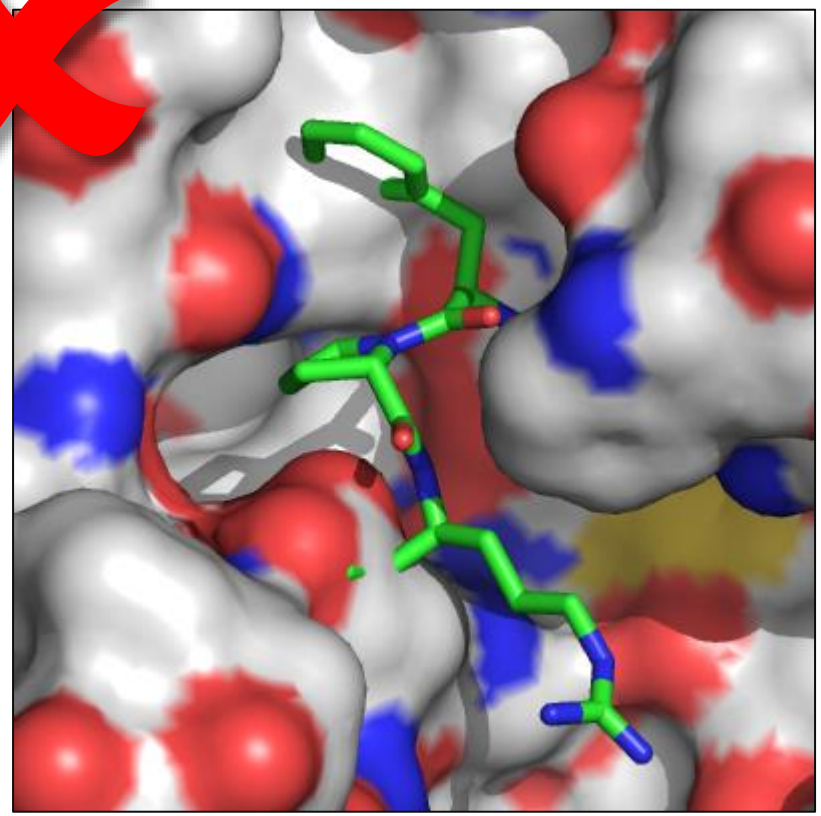
- Steric fit (how well a small molecule fits)
- Van der Waals
- Hydrogen bonding
- Electrostatics
- Hydrophobic interactions



EXAMPLE OF A BAD "BINDING POSE"



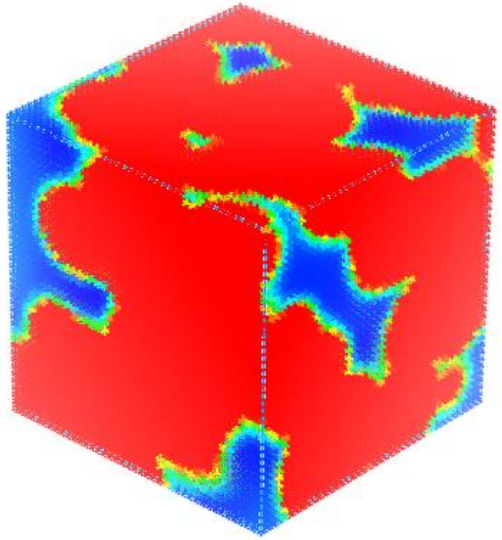
Thrombin (1PPB), with binding inhibitor 0G6

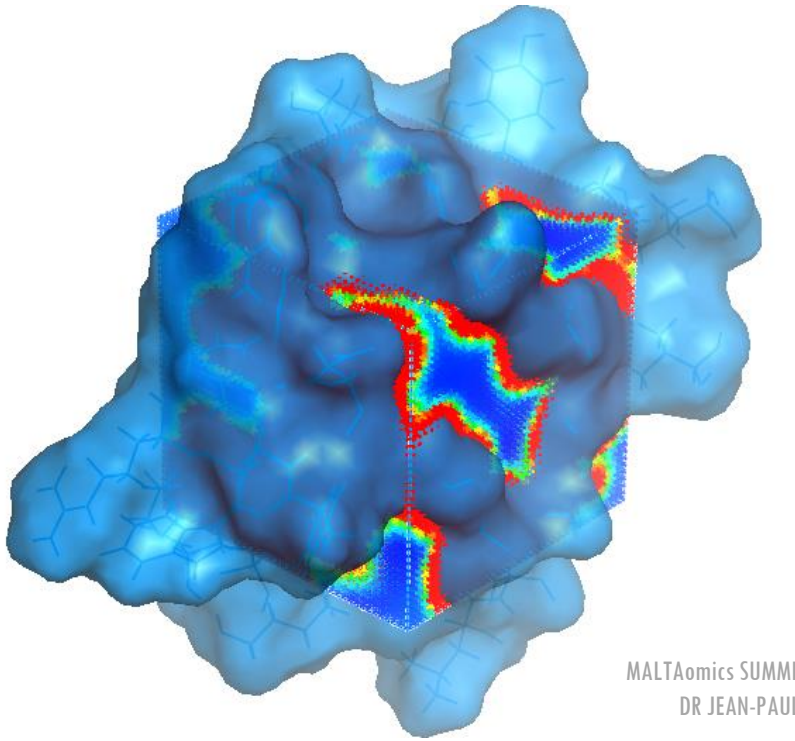


Same receptor, with different conformer of same inhibitor (0G6)

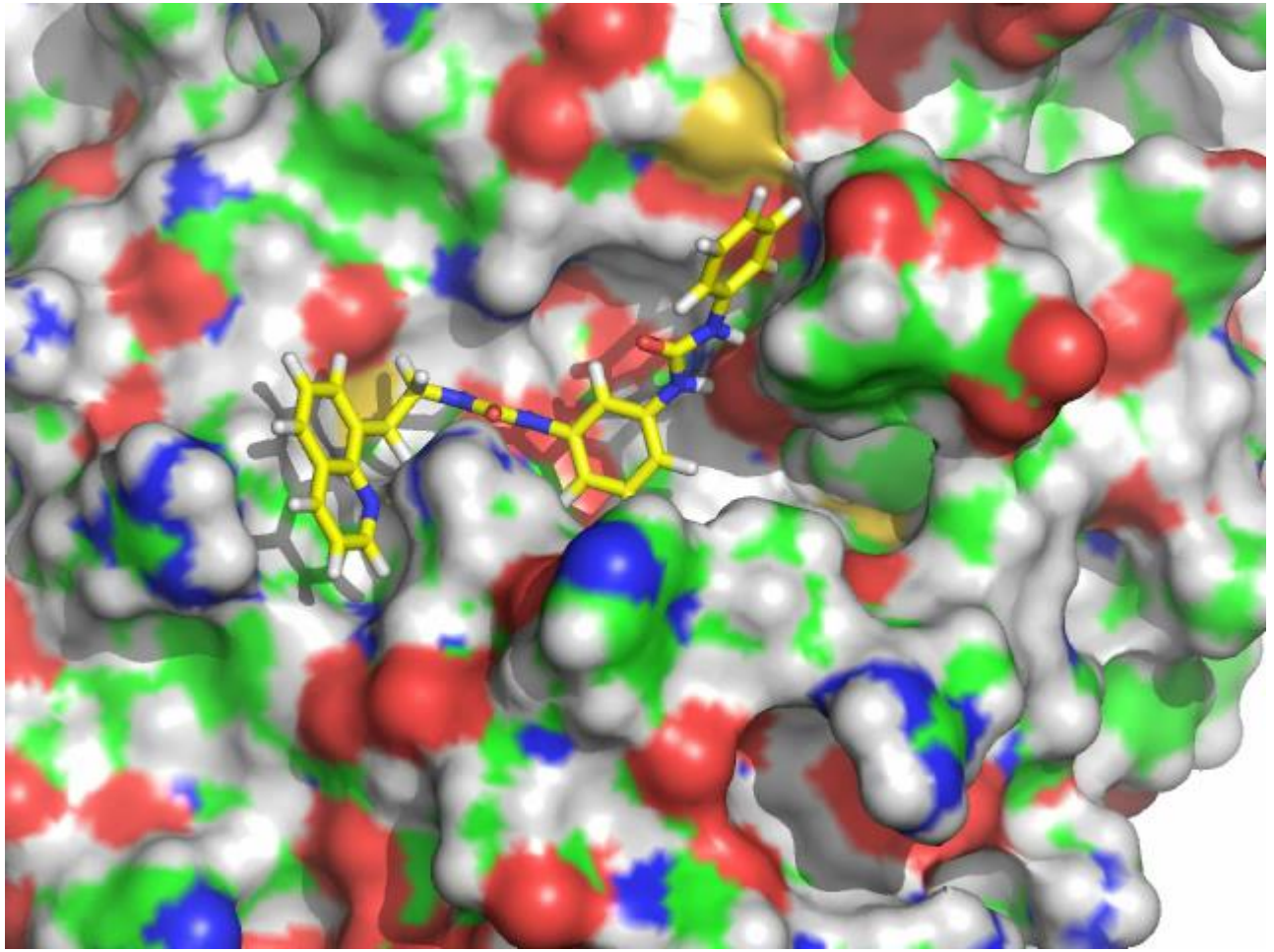
HOW DO WE NAVIGATE THE SEARCH SPACE?

- Search problem
  - Build a grid over “pocket” (area of interest in the protein)
  - Search (stochastically) for best placement of small molecule (using Monte Carlo simulations or Genetic Algorithms approaches)





SOME EXAMPLE RESULTS, ON A MALARIAL TARGET



SBVS – SOME PROBLEMS

- Need a protein structure
  - Sometimes hard to get, but some techniques exist to model this on existing structures
- Receptor flexibility (protein is considered to be static most times)
- Role of water not well modelled
- Binding affinity calculation still an open problem

MD SIMULATIONS – AN EXAMPLE

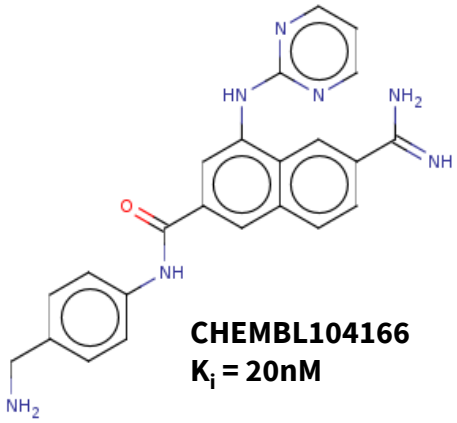
Acknowledgement: Dr. Bernhard Knapp, MHC Simulation

LIGAND-BASED VIRTUAL SCREENING

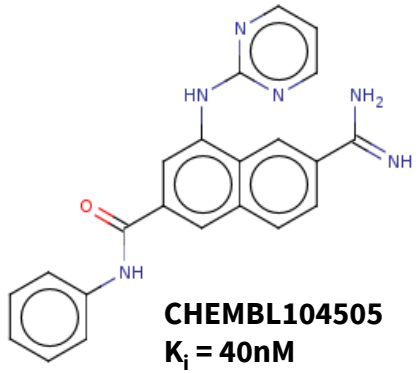
- **Ligand means "small-molecule binder"**
- Ligand based virtual screening – using information from known binders to find novel inhibitors (without considering the target protein structure)
- Find **similar** molecules to the active/binding one
- But what does "similar" mean?

SIMILAR PROPERTY PRINCIPLE[1]

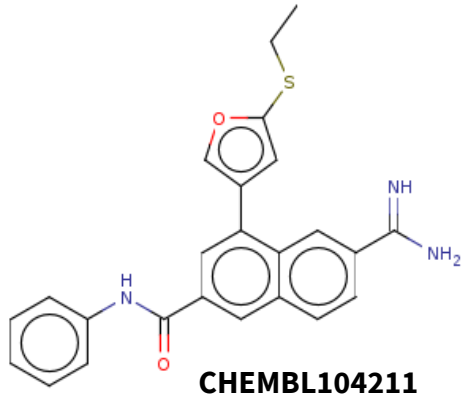
- Pillar of cheminformatics
- Similar chemical structures usually possess similar physicochemical properties and exhibit similar biological activity
- E.g. trypsin inhibitors



CHEMBL104166
K_i = 20nM



CHEMBL104505
K_i = 40nM

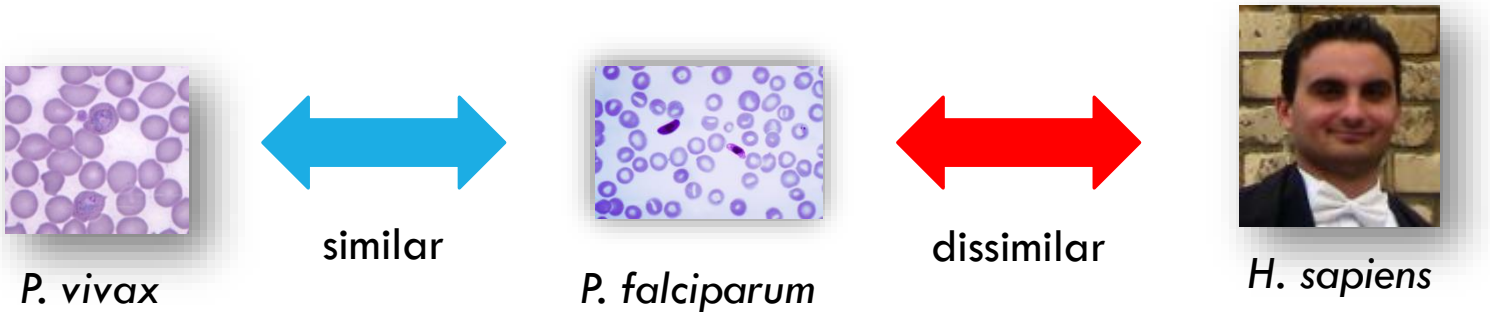


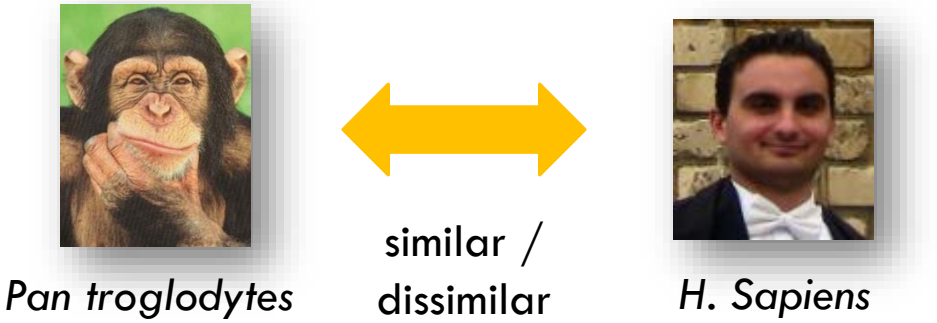
CHEMBL104211
K_i = 11nM

[1] A. M. Johnson, G. M. Maggiora (1990). *Concepts and Applications of Molecular Similarity*. New York: John Willey & Sons

BUT SIMILARITY IS A TRICKY CONCEPT TO DEFINE

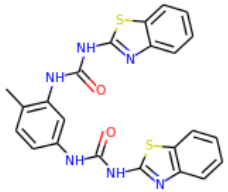
- Problem: "similarity" is a difficult concept to define





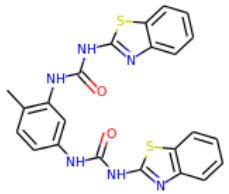
EXAMPLE LBVS SEARCH

Query:

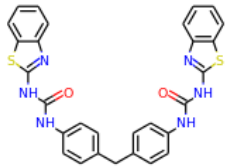


319990 1.00

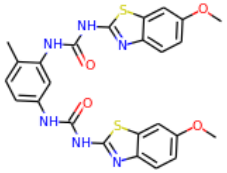
Known active molecule



319990 1.00

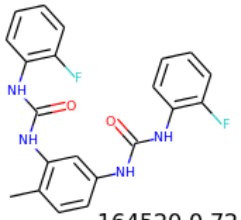


319025 0.81

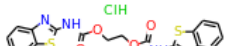


319046 0.79

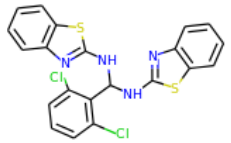
110320 0.77



164520 0.72

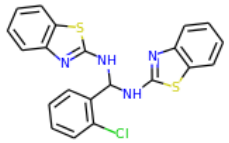


26836 0.71

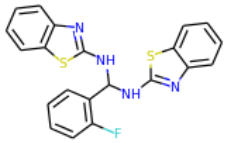


357652 0.69

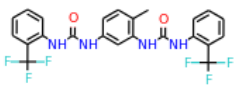
364731 0.69



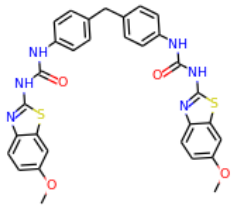
357649 0.68



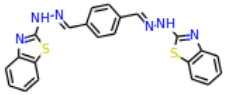
357655 0.68



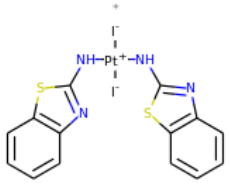
164521 0.67



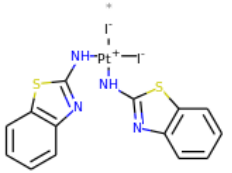
319047 0.66



151907 0.66



264148 0.65

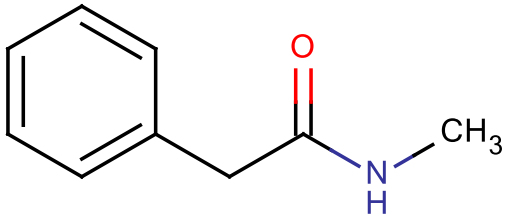
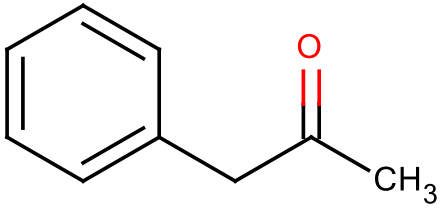


264149 0.65

DEFINING MOLECULAR SIMILARITY – 2 STEPS

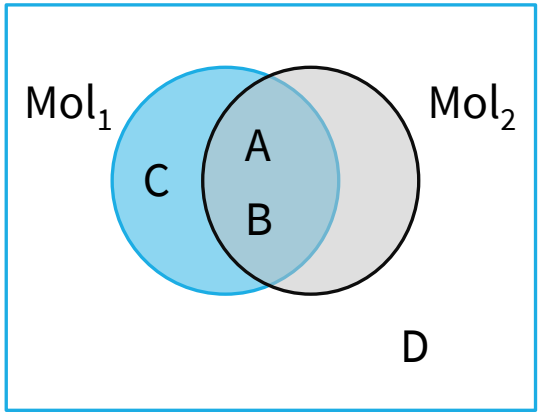
- Compute a description for every molecule
  - Physicochemical descriptors e.g. molecular weight, rotatable bonds etc.
  - 2D descriptors e.g. chemical fingerprints
  - 3D descriptors, related to molecular shape
  - Higher dimensionality (e.g. 4D = 3D + electrostatics)
- Compare descriptors
  - Simple mathematical formula gives a similarity score (e.g. Tanimoto)
  - More complex machine learning techniques

HOW DO FINGERPRINTS WORK?

		
Has aromatic 6-member ring?	YES	YES
Is charged species?	NO	NO
Has methyl group CH_3?	YES	YES
Has amide C(=O)N?	YES	NO
Resulting 4-bit fingerprint	1011	1010

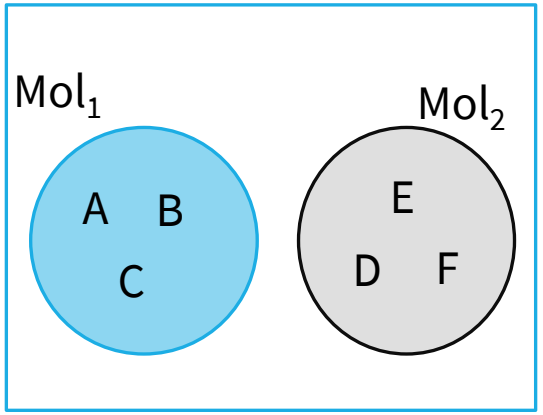
HOW DO WE COMPARE FINGERPRINTS?

Feature Space



A – 6-member ring
B – Methyl
C – Amide
D - Charges

Feature Space for two distinct molecules



- Set Theory
- Venn Diagrams are good qualitatively but we want a quantitative score
- Tanimoto (or Jaccard Index)
- Tanimoto score bound by [0,1]

$$\frac{|M_1 \cap M_2|}{|M_1 \cup M_2|} = \frac{2}{3} = 0.667$$

A MANUAL DATA MINING EXERCISE (FOOD FOR THOUGHT)

	$Feature_1$	$Feature_2$	$Feature_3$	$Feature_4$	$Feature_5$	$Feature_6$	$Feature_7$	$Feature_8$
Mol_1	1	1	0	1	1	0	1	0
Mol_2	1	1	0	0	1	0	1	0
Mol_3	0	1	1	1	0	0	1	1
Mol_4	0	0	0	1	0	0	0	0
Mol_5	0	1	1	0	1	0	1	1
Mol_6	0	0	0	0	1	1	1	1
Mol_7	1	0	0	1	0	1	1	0

Note: Can get better results using proper ML and data mining techniques (e.g. Random Forests, Artificial Neural Networks etc.)

CONCLUSIONS

- Finding a small-molecule inhibitor is one of the many steps in drug discovery:
  - Safety, potency, delivery, specificity etc.
- SBVS and LBVS screening have become widely used tools and have contributed to many drug discovery programmes
- Nonetheless, the complexity of biological systems means that improved methods are still very much required
- The rapid growth of experimental data (human genome, protein structure databases, etc.) and continuing increases in computational power provide a platform for the development of novel algorithms
  - A lot of this big data is processed on the cloud
- A few techniques, many potential application areas