Molecular Dynamics Simulations in Drug Discovery

Thomas Fox

Discovery Research
Boehringer Ingelheim Pharma GmbH & Co KG
April 2017

Boehringer Ingelheim

Boehringer Ingelheim in brief

Our headquarters in Ingelheim, Germany

- Family-owned global corporation
- Founded 1885 in Ingelheim, Germany
- Focus on Human Pharmaceuticals, Animal Health and Biopharmaceuticals
- Employees worldwide ~ 46,000
- ~8000 employees in R&D+Medicine

- Net sales 2016: ~16 billion €
- Expenses for RD&M: ~3 billion €

4 major research areas:

Immunology and respiratory diseases

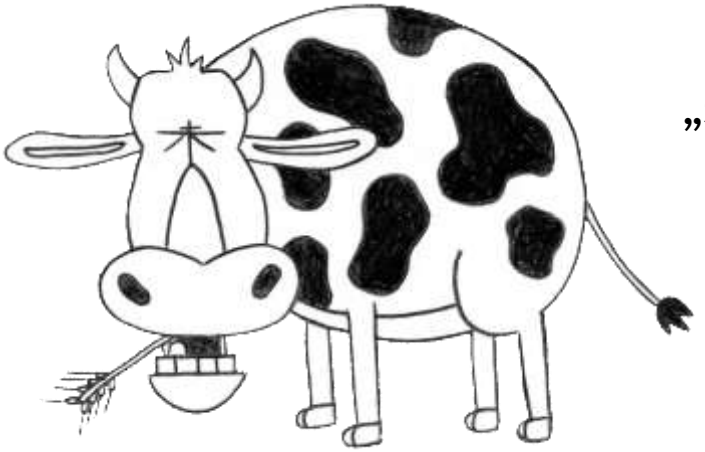
Cardiometabolic diseases

Oncology

Diseases of the central nervous system

Our global research and development sites human pharmaceuticals

From Accidental Discovery to a Drug

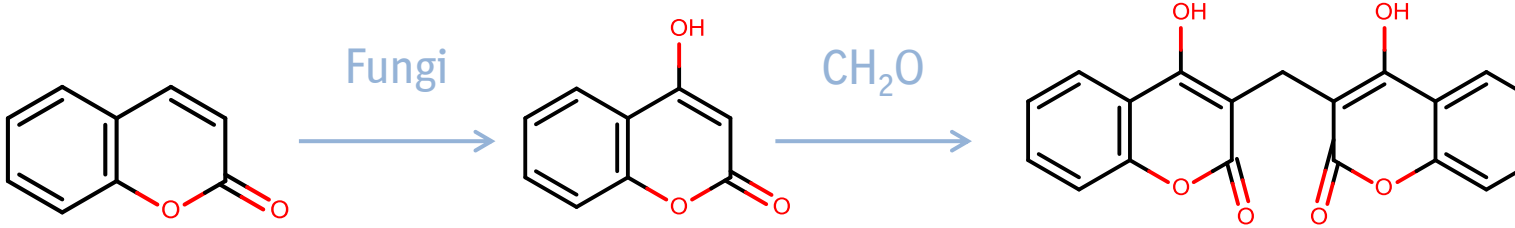


Cattle, 1920s, North Dakota

„sweet clover disease“

rotten clover

Dead Cattle, 1920s, North Dakota



Coumarin

4-Hydroxy-Coumarin

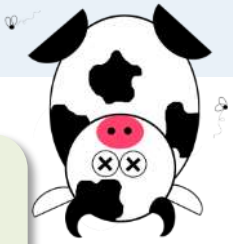
Di-Coumarol

strong anti-coagulant

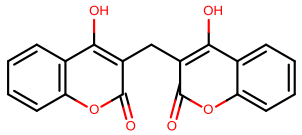
From Accidental Discovery to a Drug

Accident
(Decaying Hay)

Observation of a pharmacological effect

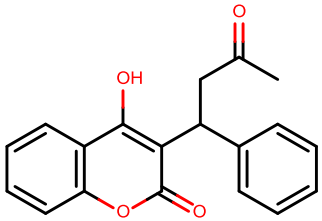


Link pharmacological effect to a substance



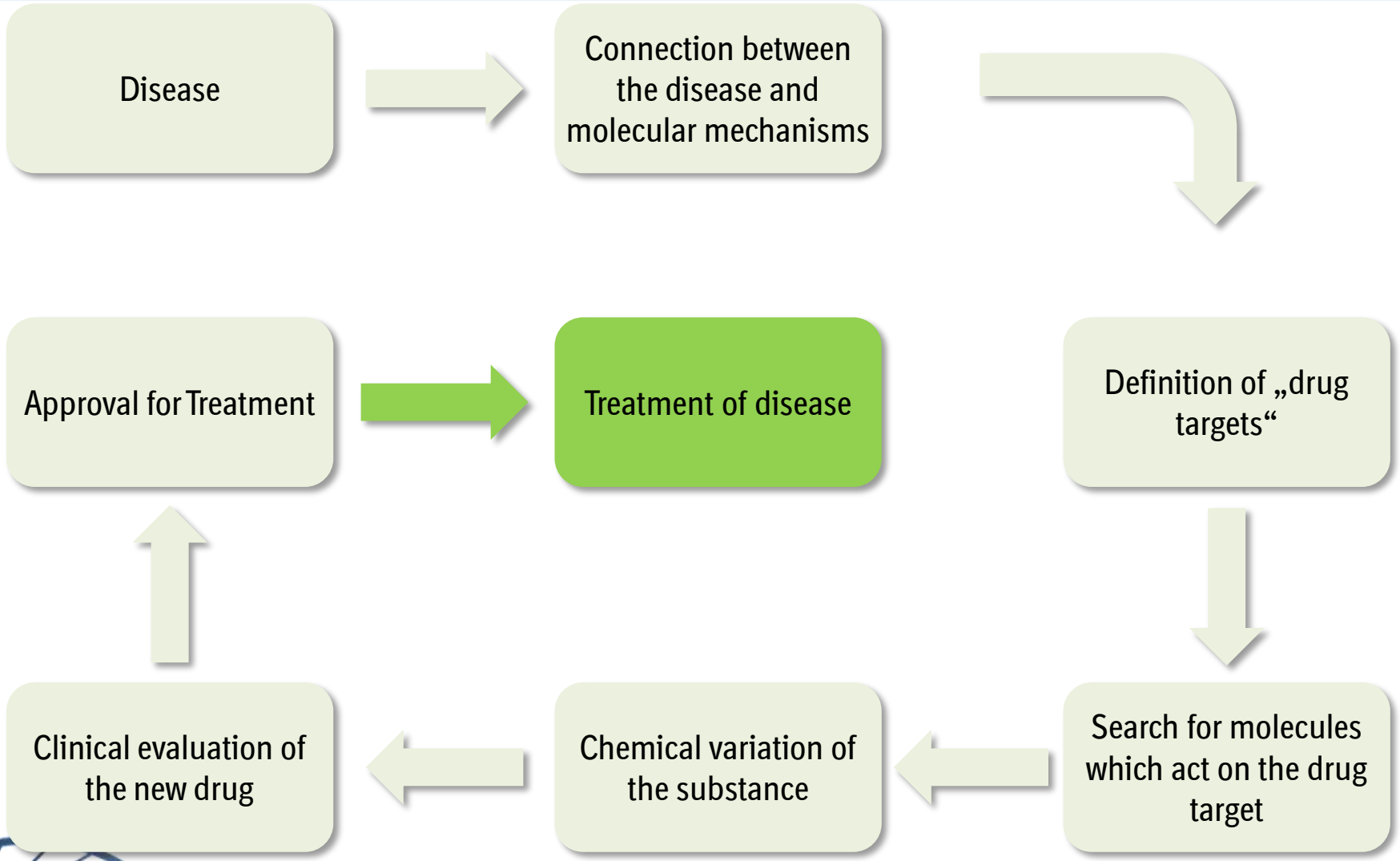
Treatment of diseases

Chemical variation of the substance



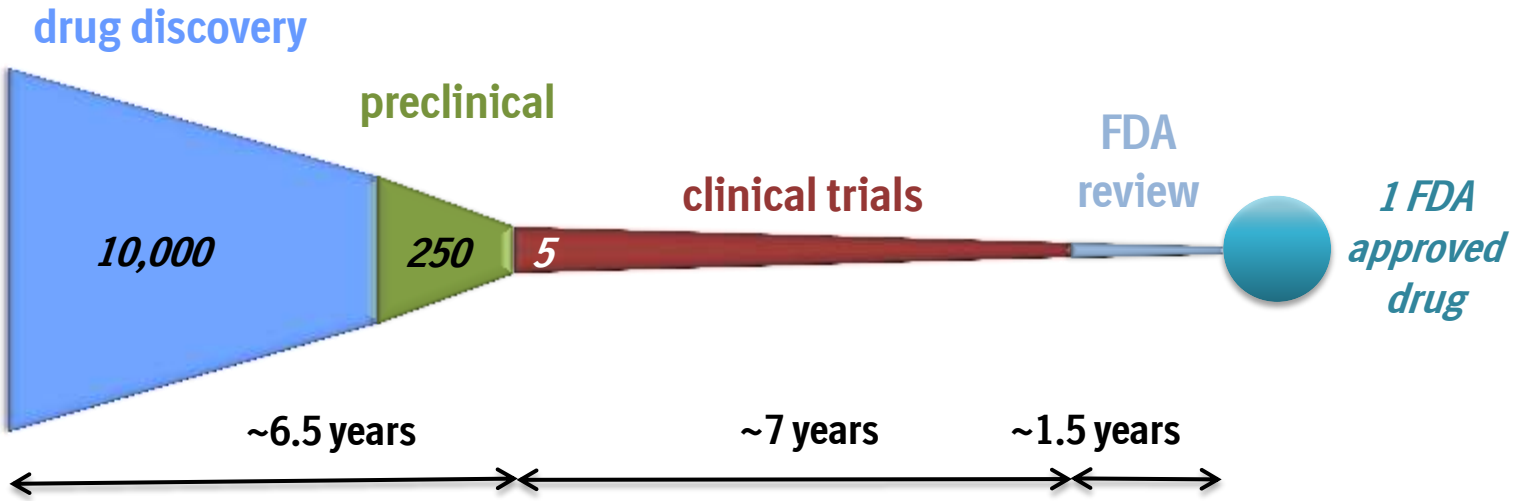
Warfarin

From the Disease to the Drug

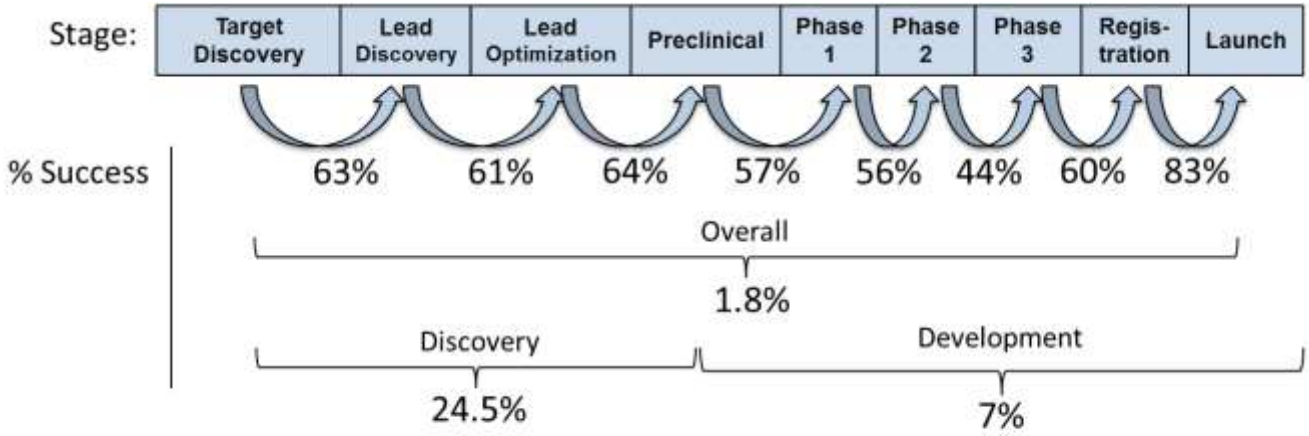


Drug Discovery Is …

a long journey



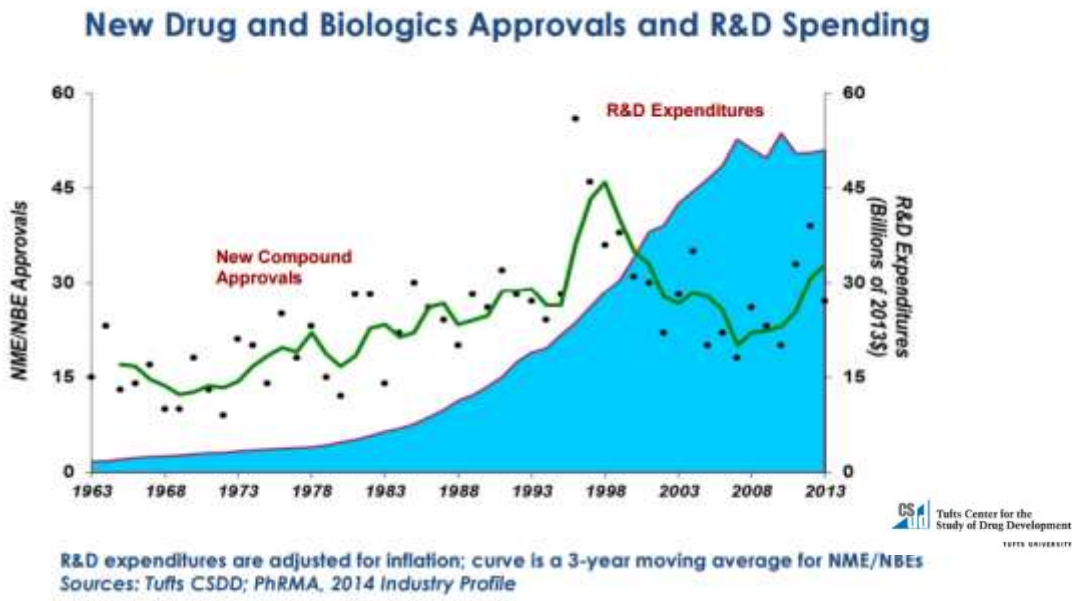
risky



- http://www.discoverymanagementsolutions.com/the-organization-of-biopharmaceutical-rd/attrition/

Drug Discovery Is …

expensive



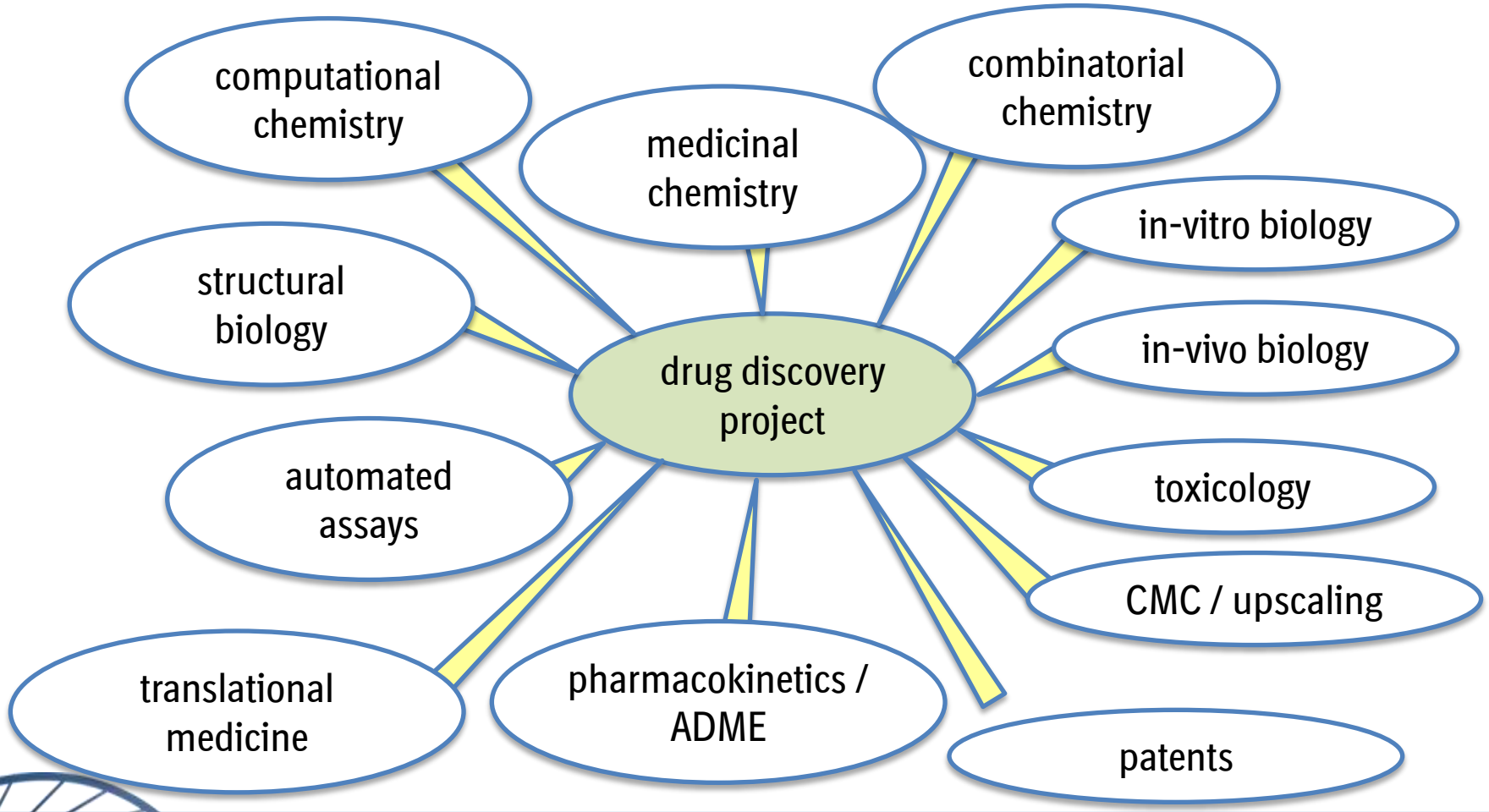
- current estimate: 800 - 1000 M$ to bring new drug to market
- 50-100 M$ preclinical cost

changing

- new, difficult targets
- new chemical matter
- biologicals

Drug Discovery Is ...

a multi-disciplinary team approach

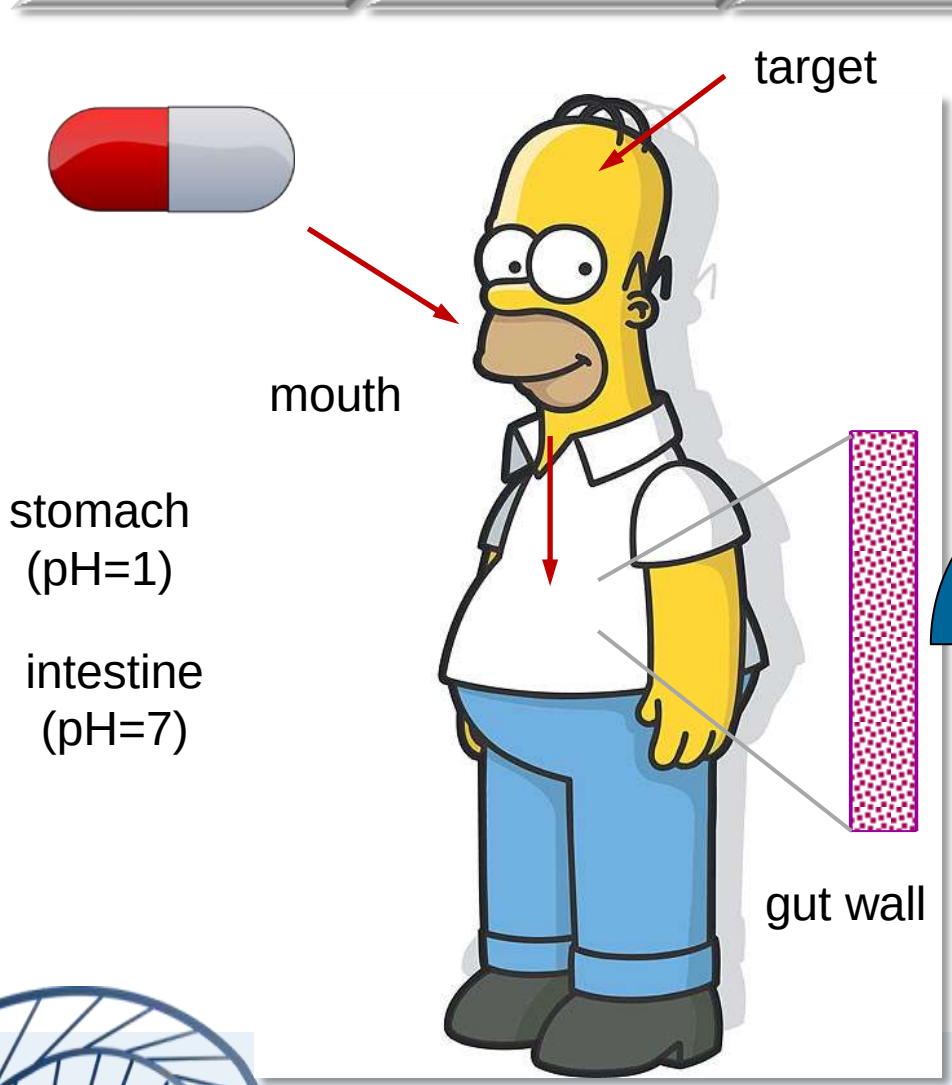


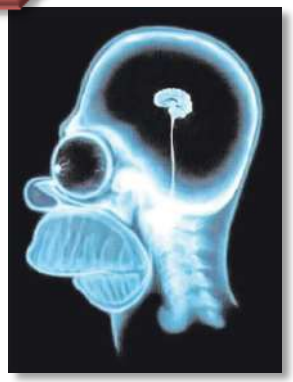
Drug Discovery Project Phases
A Typical Research Project

Target Discovery	Assay Development	Lead Identification	Lead Optimization
➢ Target ideas from literature and in-house research	➢ Cell-lines expressing the target protein	➢ High-throughput screening	➢ Affinity/Activity
➢ Target validation with knock-out animals, siRNA, tool compounds	➢ Biochemical assays	➢ Hit analysis	➢ Selectivity
	➢ Optimization and automation for high-throughput screening	➢ Selection of promising chemical matter	➢ Pharmacokinetics/ Metabolism
		➢ Testing of related compounds	➢ Chemical manufacturing control
		➢ Structure-activity relationships	

Computational Chemistry

Multiple Challenges for a Molecule





BRAIN

Portal vein

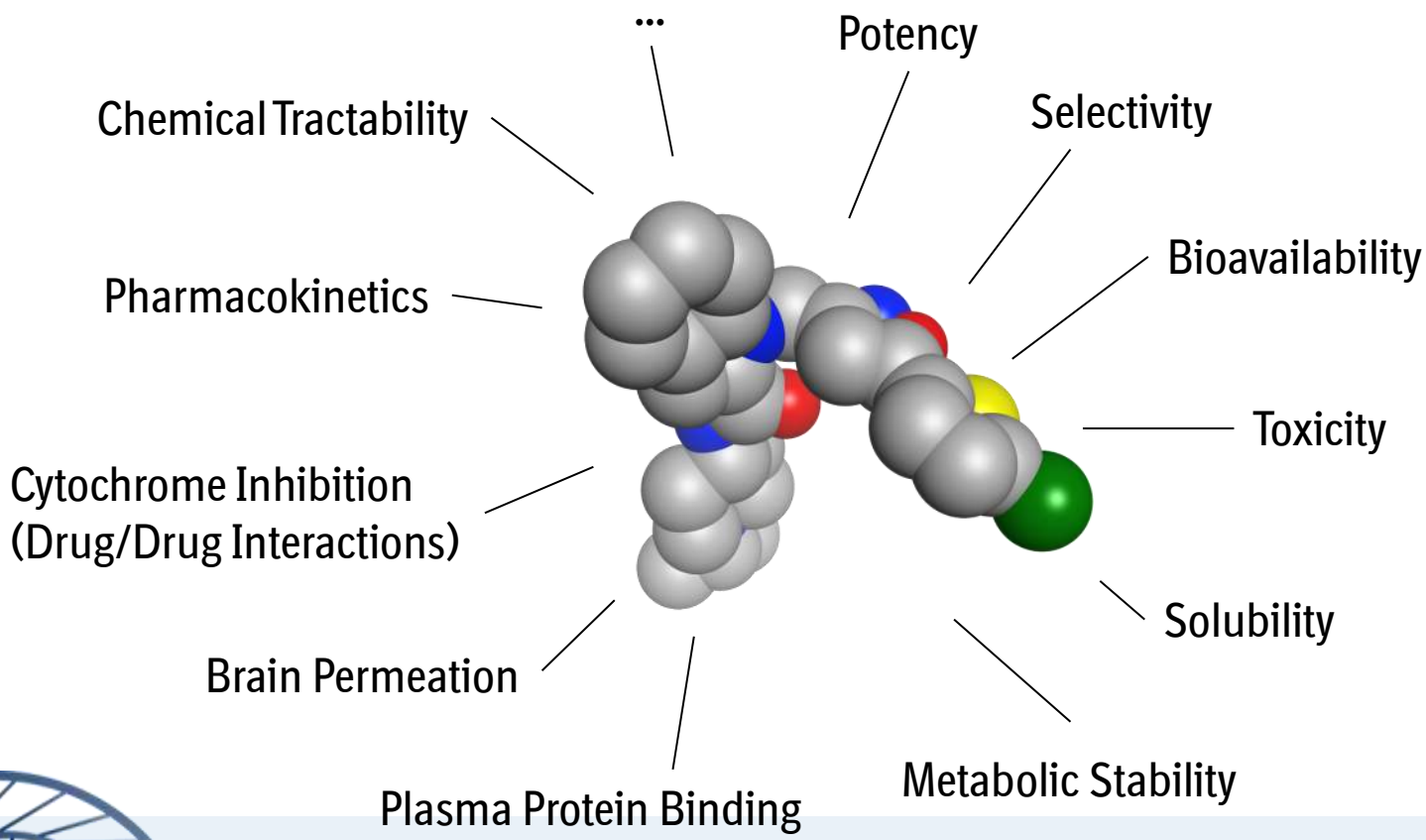
Liver

BLOOD

Metabolism

Optimization Parameters

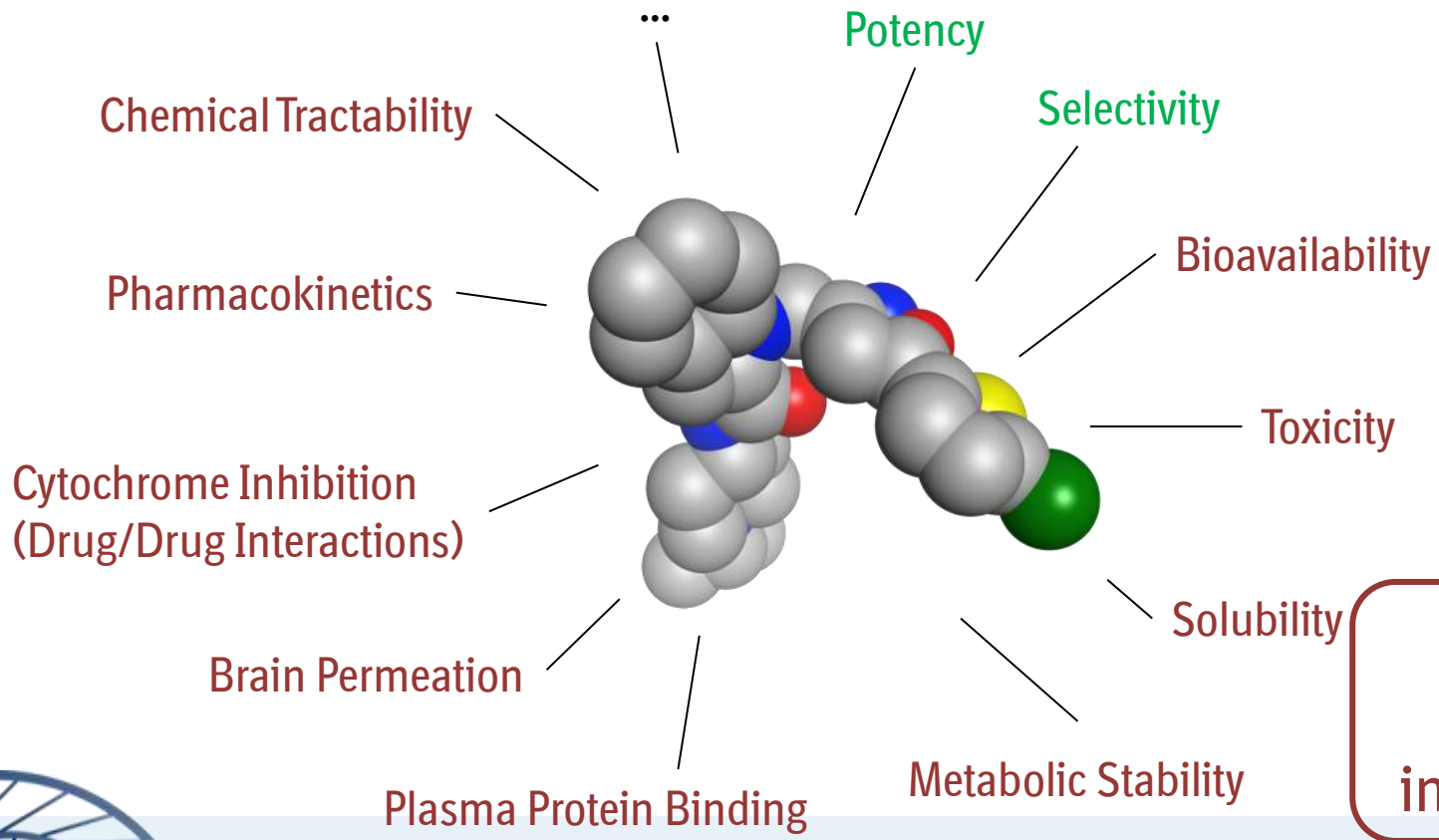
Drug Design is a Multi-Parameter Optimization



Optimization Parameters

Drug Design is a Multi-Parameter Optimization

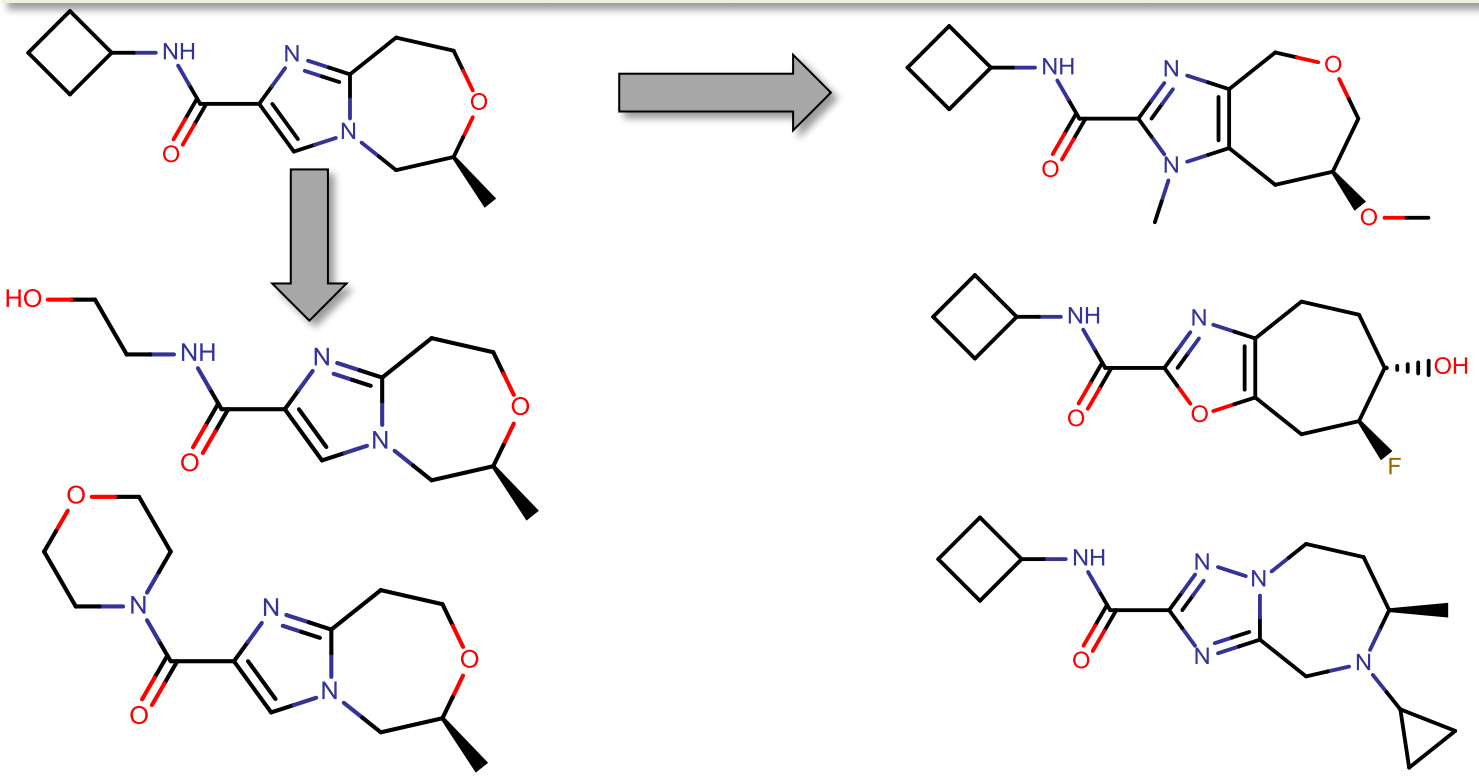
target specific



often target-independent

From a Lead to a Drug Candidate

Lead Optimization essentially means synthesis of close analogs of an active molecule.



„easy“ chemistry -> variation straightforward
cheap!

difficult chemistry, different synthetic routes
expensive!

Prediction of Molecule Properties

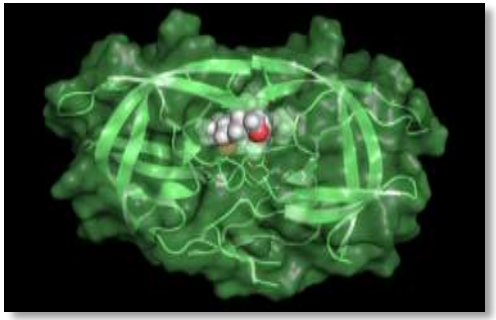
- 50-100 M$ preclinical cost
- most of it MedChem: labor-intensive, often not amenable to automation
- estimation: cost to make a compound 2000-3000$ on average

→ need to make better decisions which compounds to make

→ predict of molecular properties - prior to synthesis!

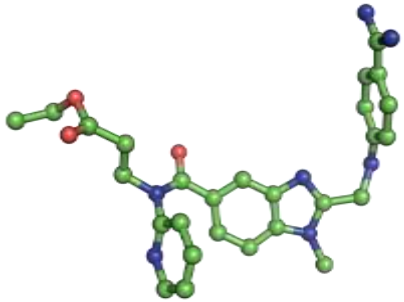
„Predictions are difficult, especially about the future“ (Niels Bohr)

Structure-based design



- Xray structure(s) required
- Physics-based approaches

Ligand-based design

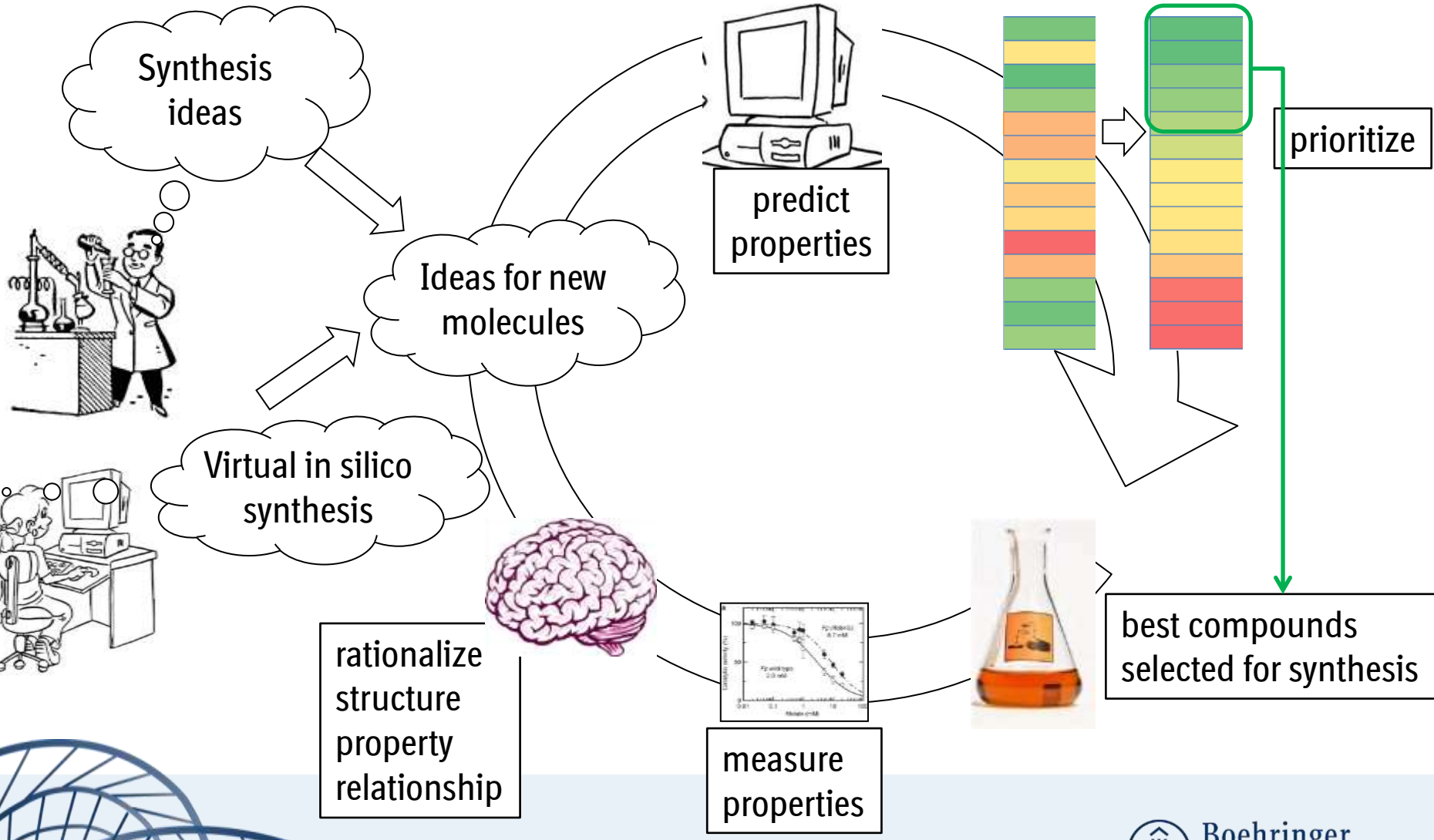


- known ligand required
- physics-based approaches
- chemoinformatics

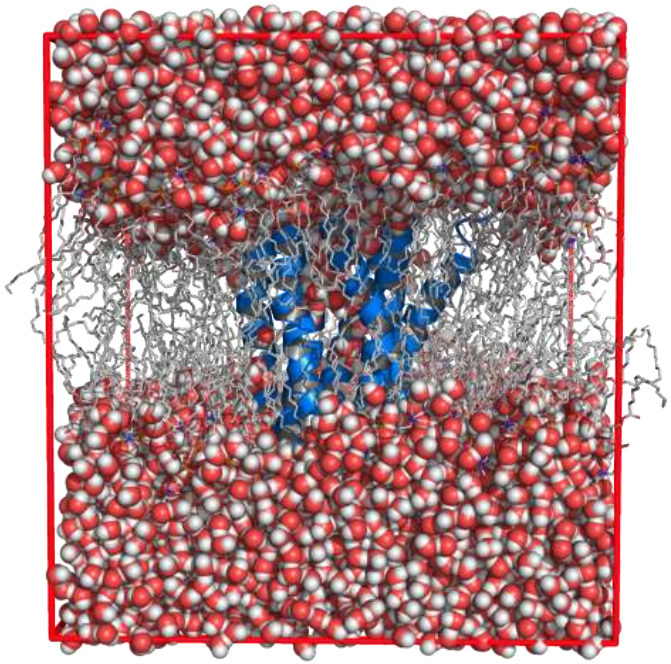
Data-driven design

- lots of data required
- chemoinformatics
- Machine Learning

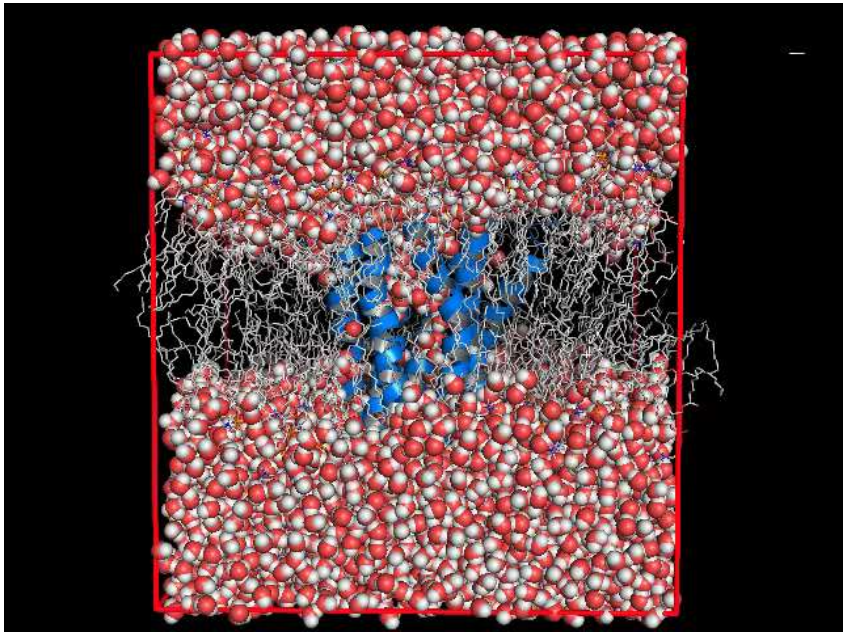
Design Cycle



What is Molecular Dynamics?



"…everything that is living can be understood in terms of jiggling and wiggling of atoms."
(R. Feynman)



- atomistic model of a biological system
- dynamics is described by Newton's equations of motion

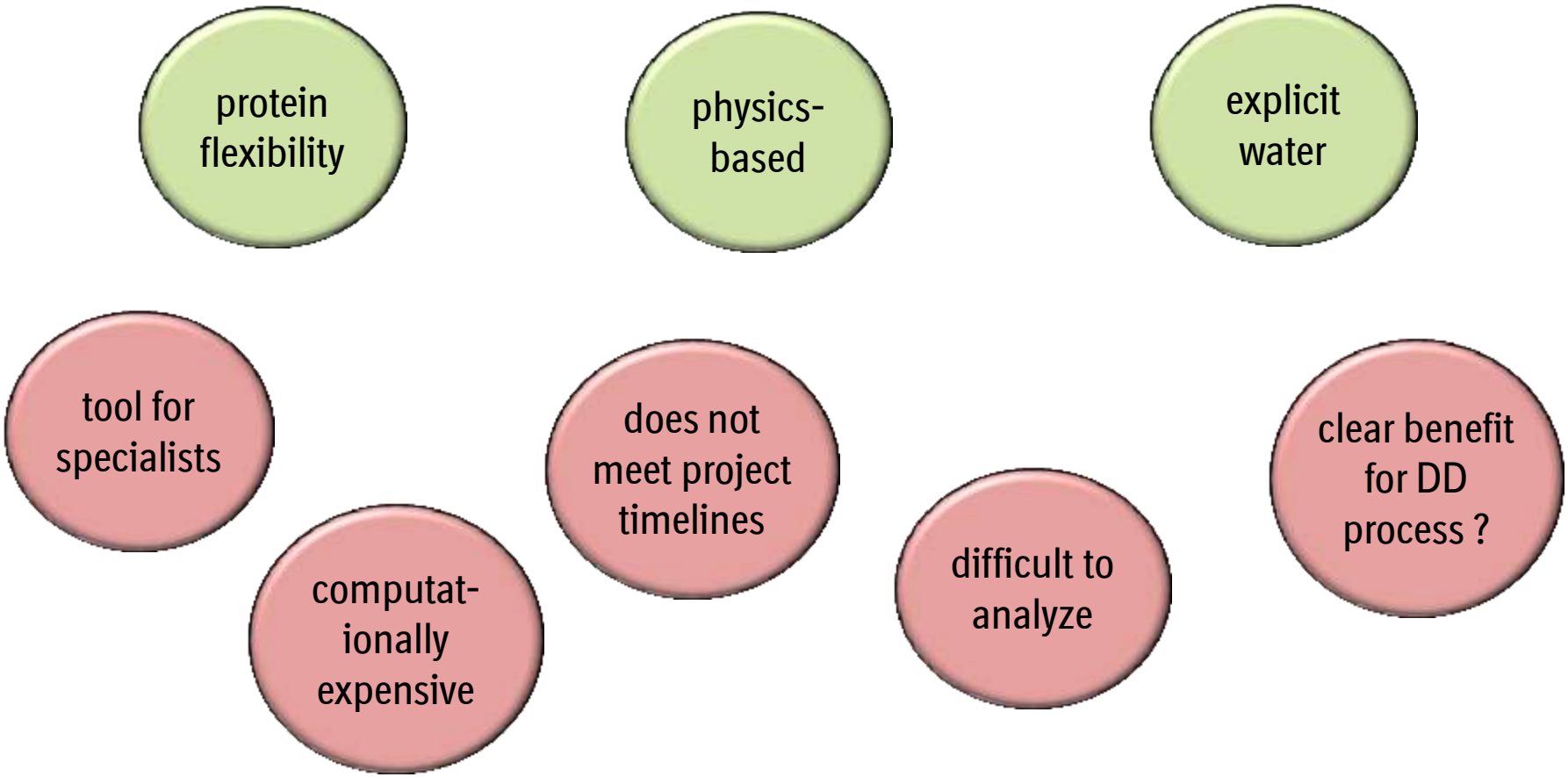
Molecular Dynamics in Drug Discovery

- Molecular Dynamics is primarily used to understand protein function
  - Ion channels
  - GPCRs
  - Aquaporins
  - .....

- Understanding protein function is important for Drug Discovery, but the **central questions** are:
  - Where and how does a ligand bind?
  - How to improve affinity?

„The computational microscope“

Molecular Dynamics in Drug Discovery

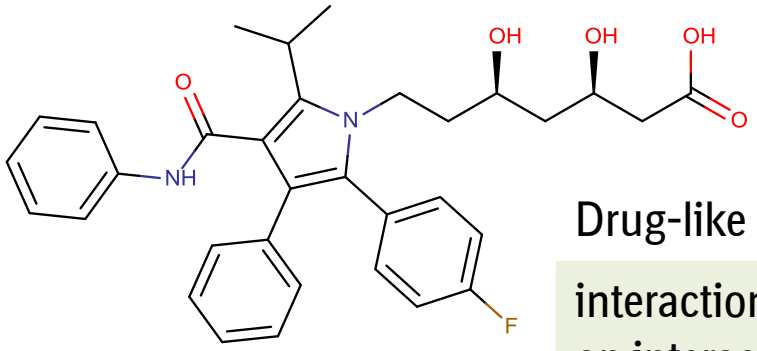


Where are current/future applications of Molecular Dynamics in Drug Discovery?

Possible Applications

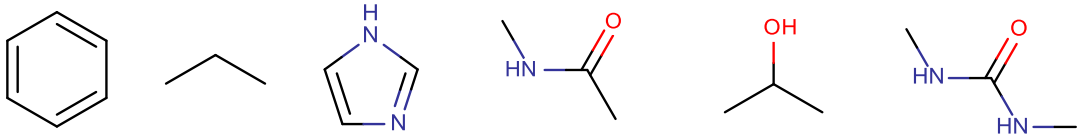
- investigate flexibility of proteins (which conformations do I have to deal with?)
- search for putative binding sites which are not obvious or not present in experimental structure
- calculate binding energies
- conformational sampling and analysis
- find and evaluate binding poses
- calculate (un)binding kinetics
- analysis of water structure and water energetics in binding site
- …

Individual Interaction Patterns



Drug-like molecule

interaction with target protein based on interactions of its fragments



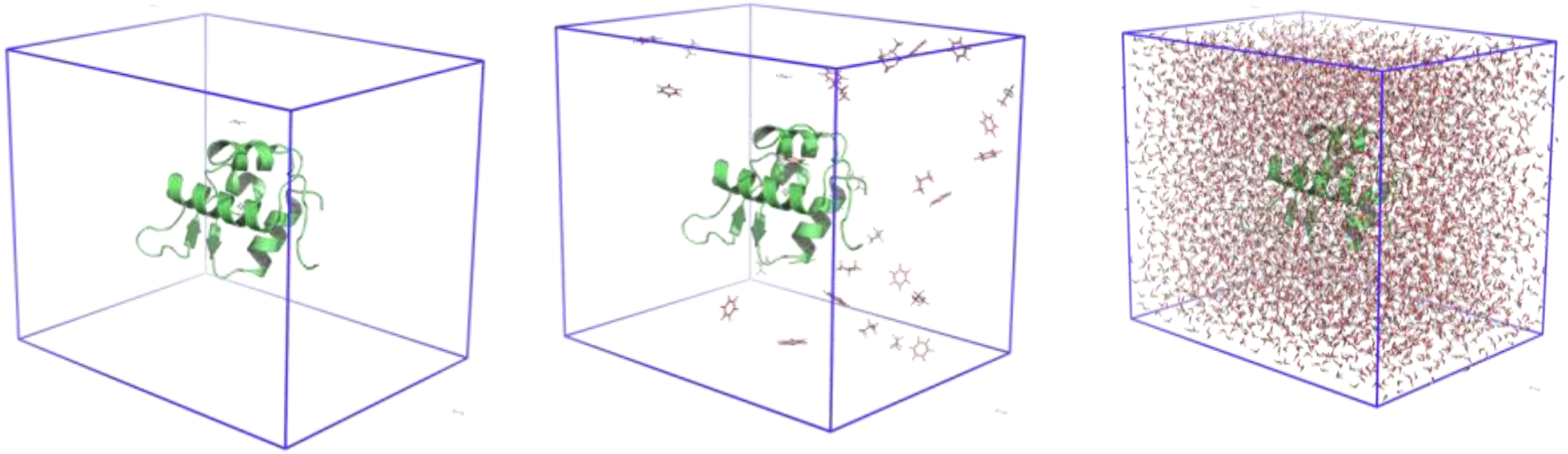
fragments (“probes“)

use these probes in MD simulations:
map protein surfaces → identify favorable interaction sites and types

Mixed-Solvent Molecular Dynamics (SILCS)

Site Identification by Ligand Competitive Saturation

MD simulation of a protein in an environment of different solvent probes (fragments)



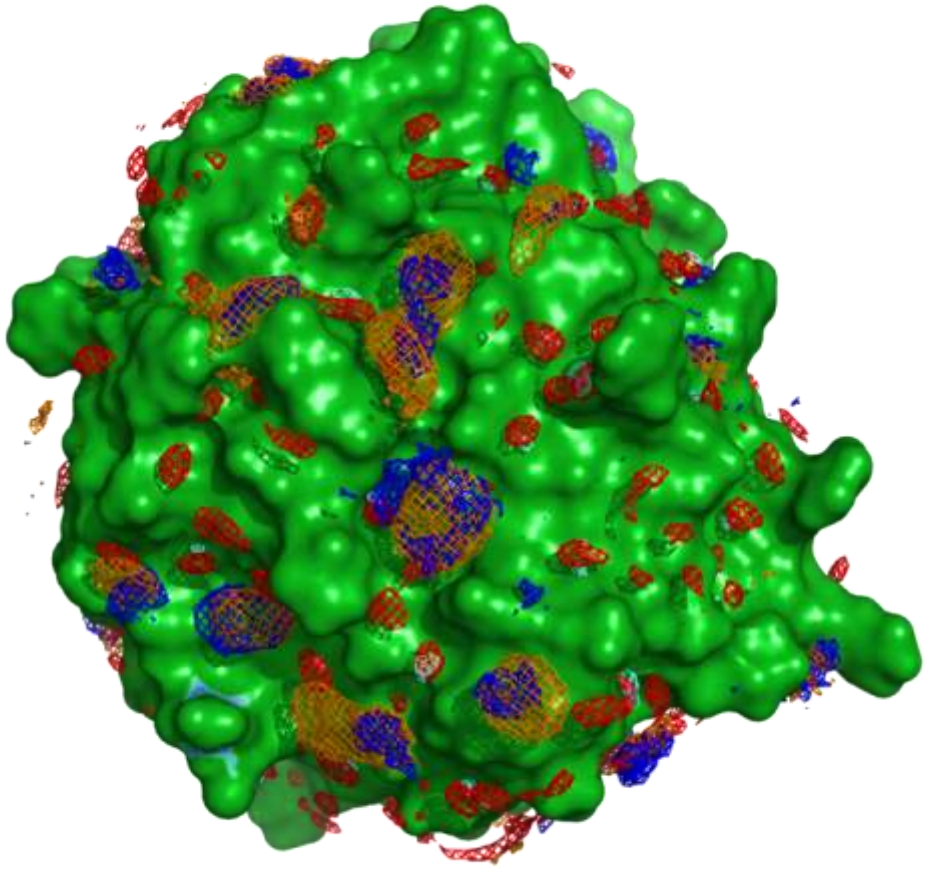
Approx. 150 mM benzene and 150mM propane (other fragments are also possible).

Basic concept:

- no assumption about a particular binding site
- fragments compete with water for binding sites at the protein surface
- flexible protein -> induced-fit, transient pockets

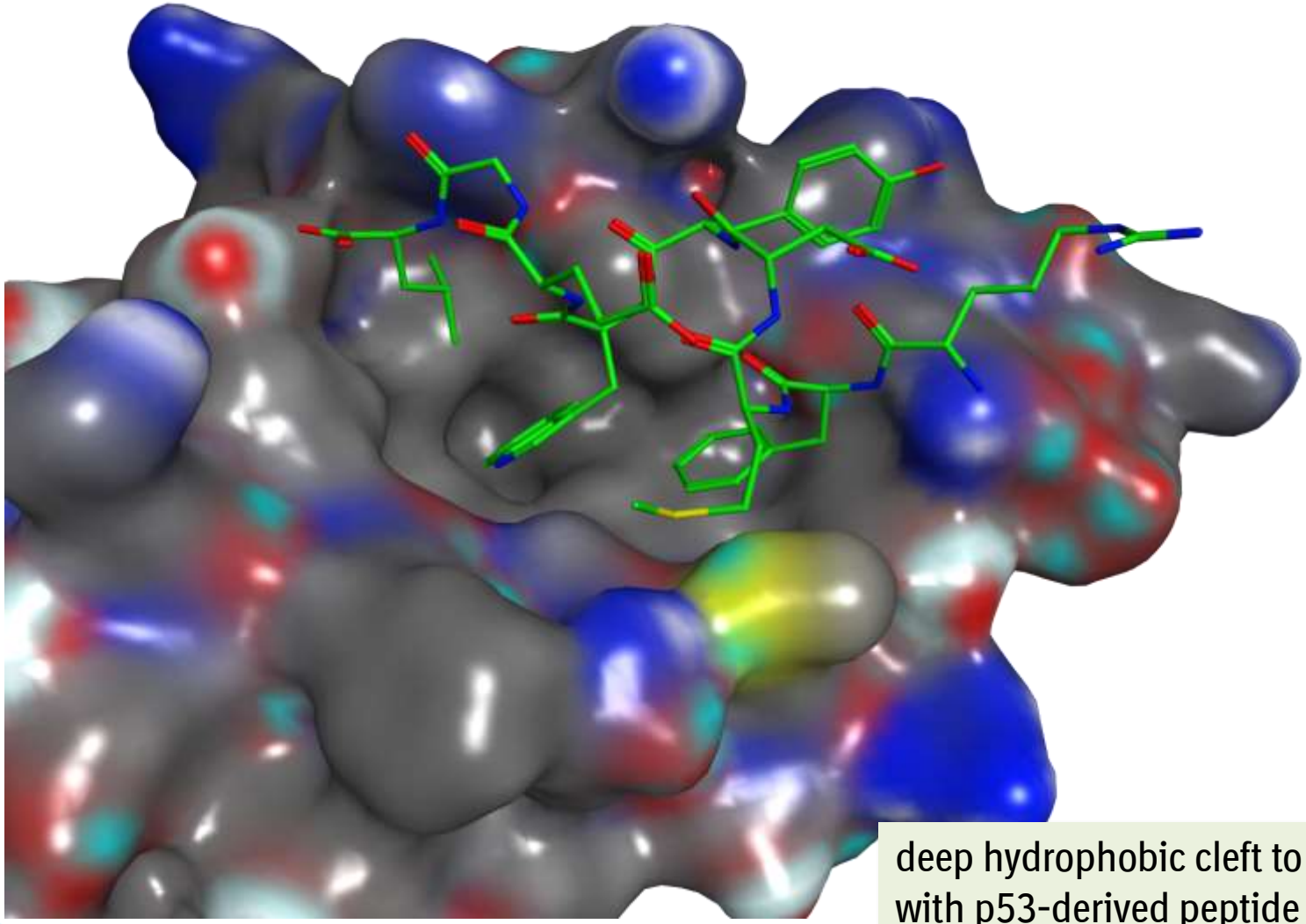
Ligand Optimization with Fragment Maps

- reasonably long mixed solvent simulations (10 x 100 ns) for converged results
- convert spatial distribution to a free-energy map for each fragment type
- same can be done for water
- use this information to suggest modifications to existing ligands



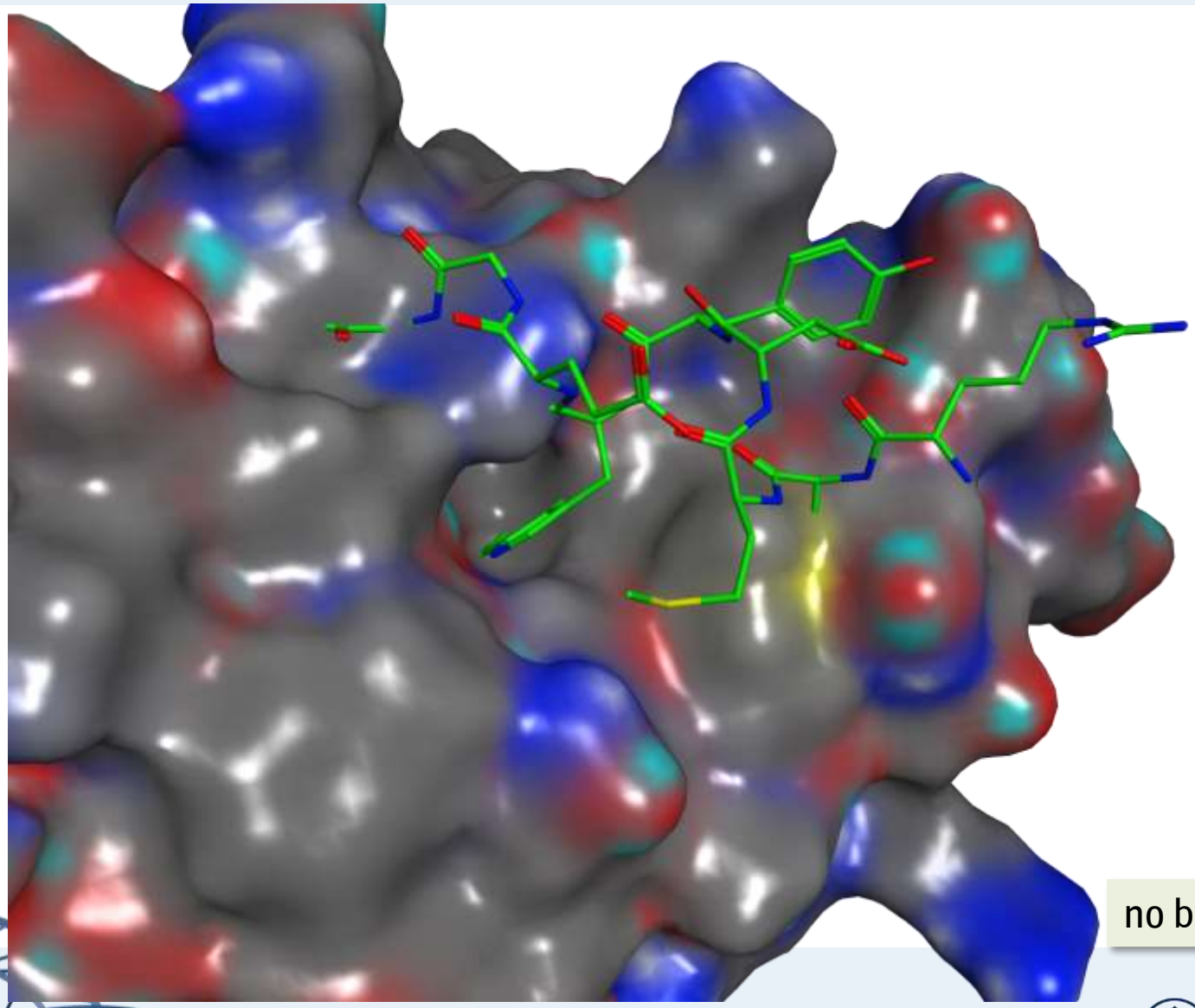
combined maps give a pharmacophore describing interaction patterns

Mdm2/p53 complex (3dac)



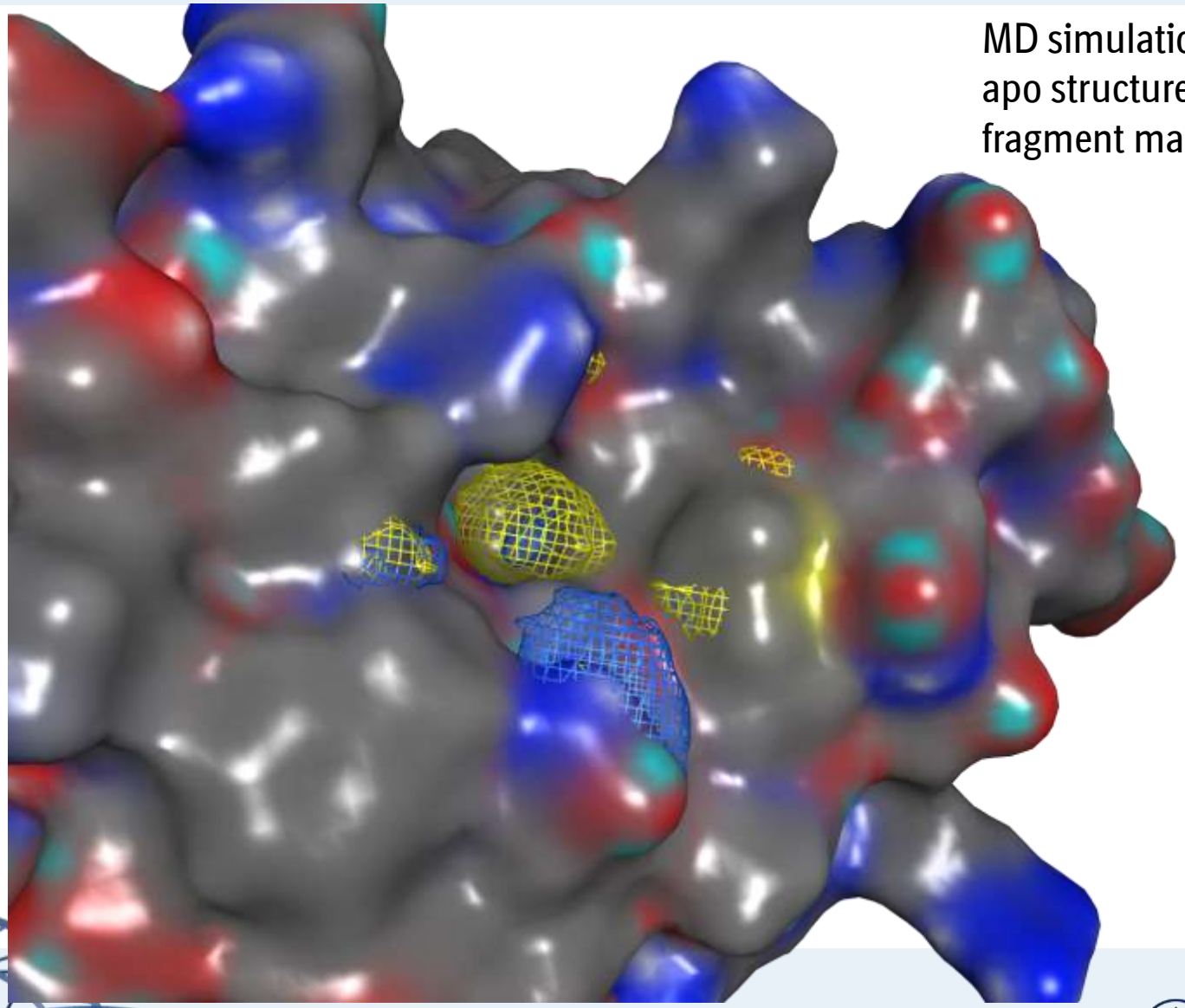
deep hydrophobic cleft to interact with p53-derived peptide

Mdm2 apo structure (1z1m)



no binding site

Fragment maps of Mdm2 (apo structure)

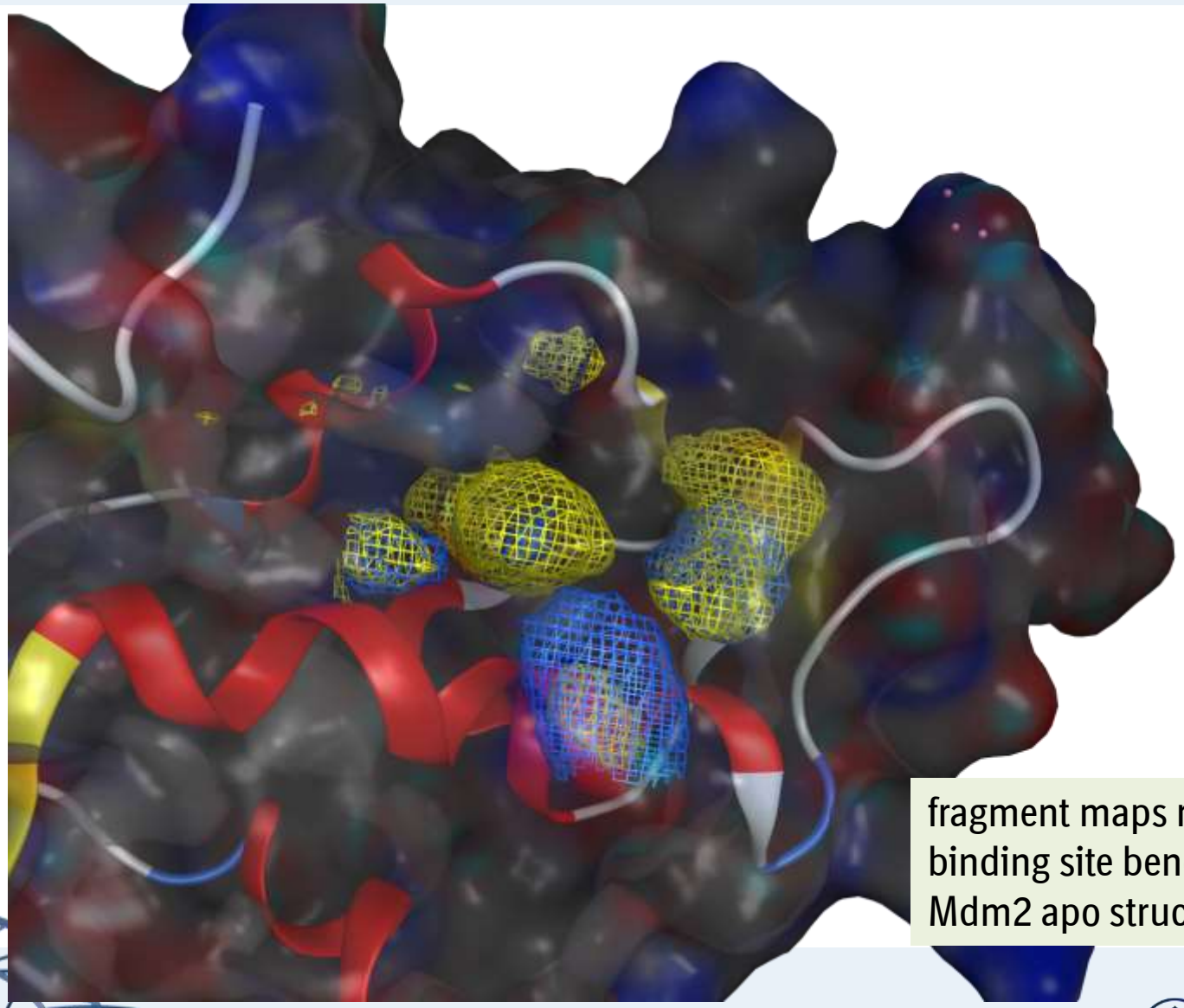


MD simulation starting from apo structure

fragment maps: benzene
propane

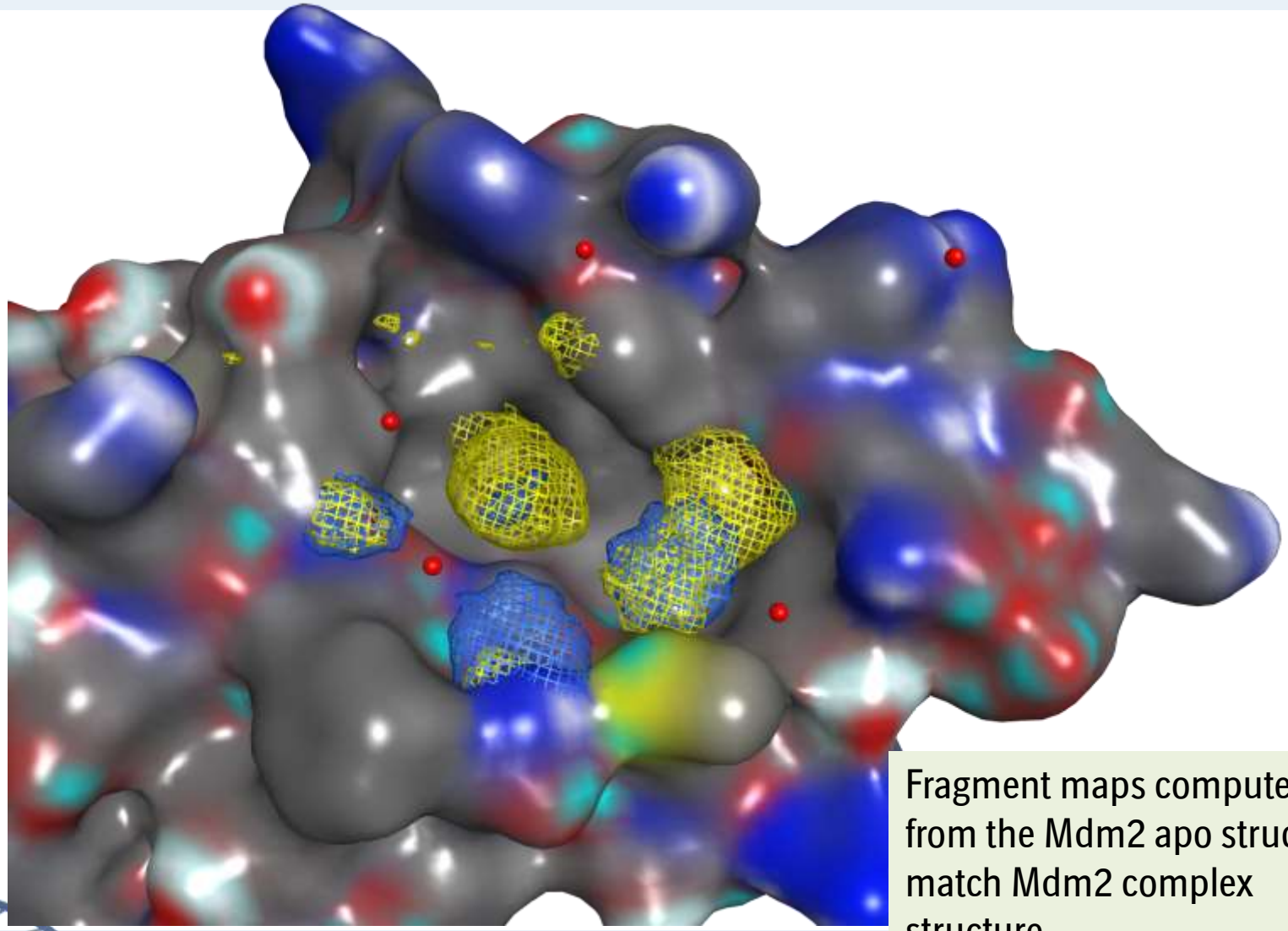
Boehringer Ingelheim

Fragment maps of Mdm2 (apo structure)



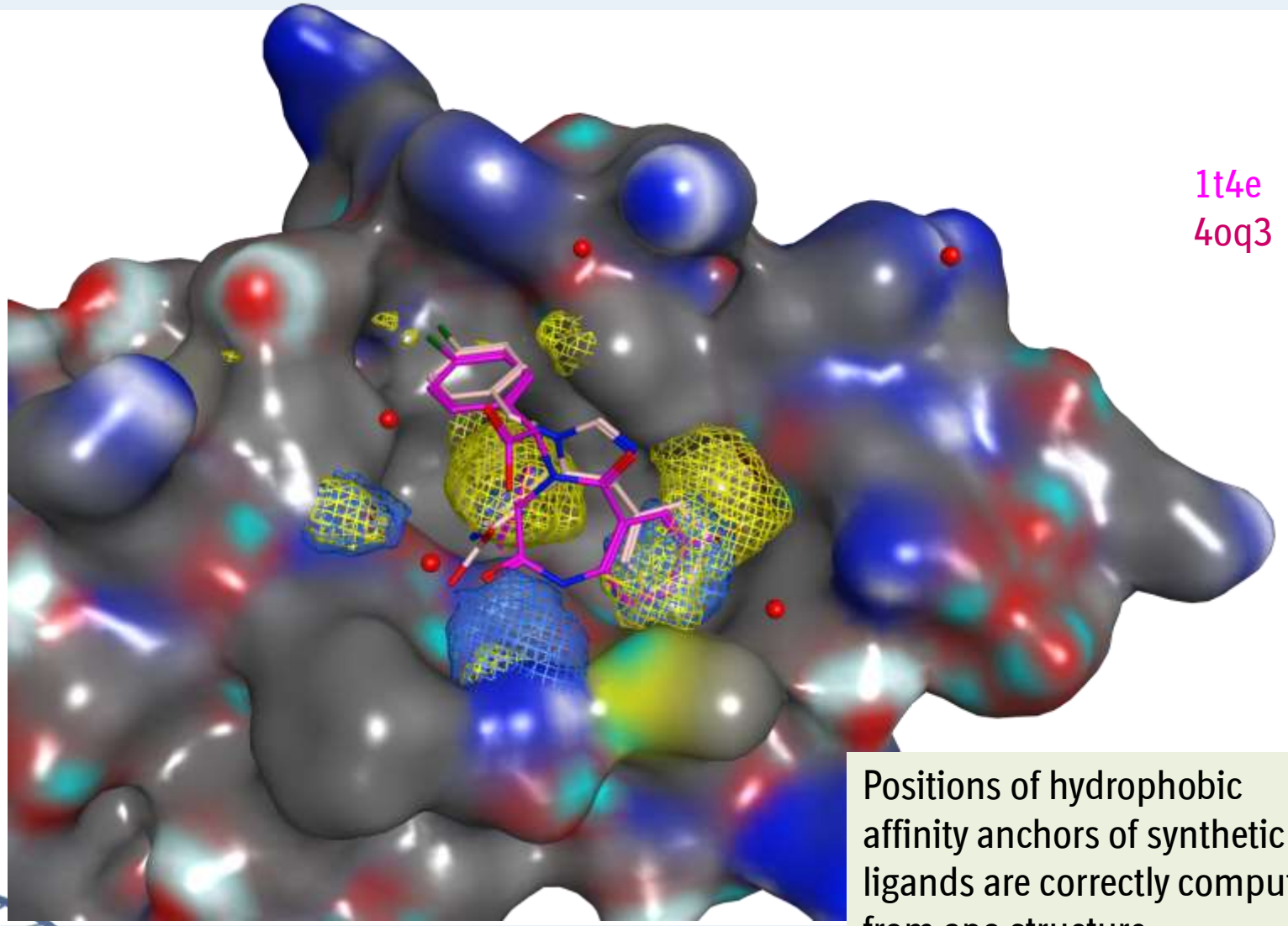
fragment maps reveal binding site beneath the Mdm2 apo structure surface

Fragment maps of compared to Mdm2 complex structure



Fragment maps computed from the Mdm2 apo structure match Mdm2 complex structure

Comparison with experimental structures



1t4e
4oq3

Positions of hydrophobic affinity anchors of synthetic ligands are correctly computed from apo structure

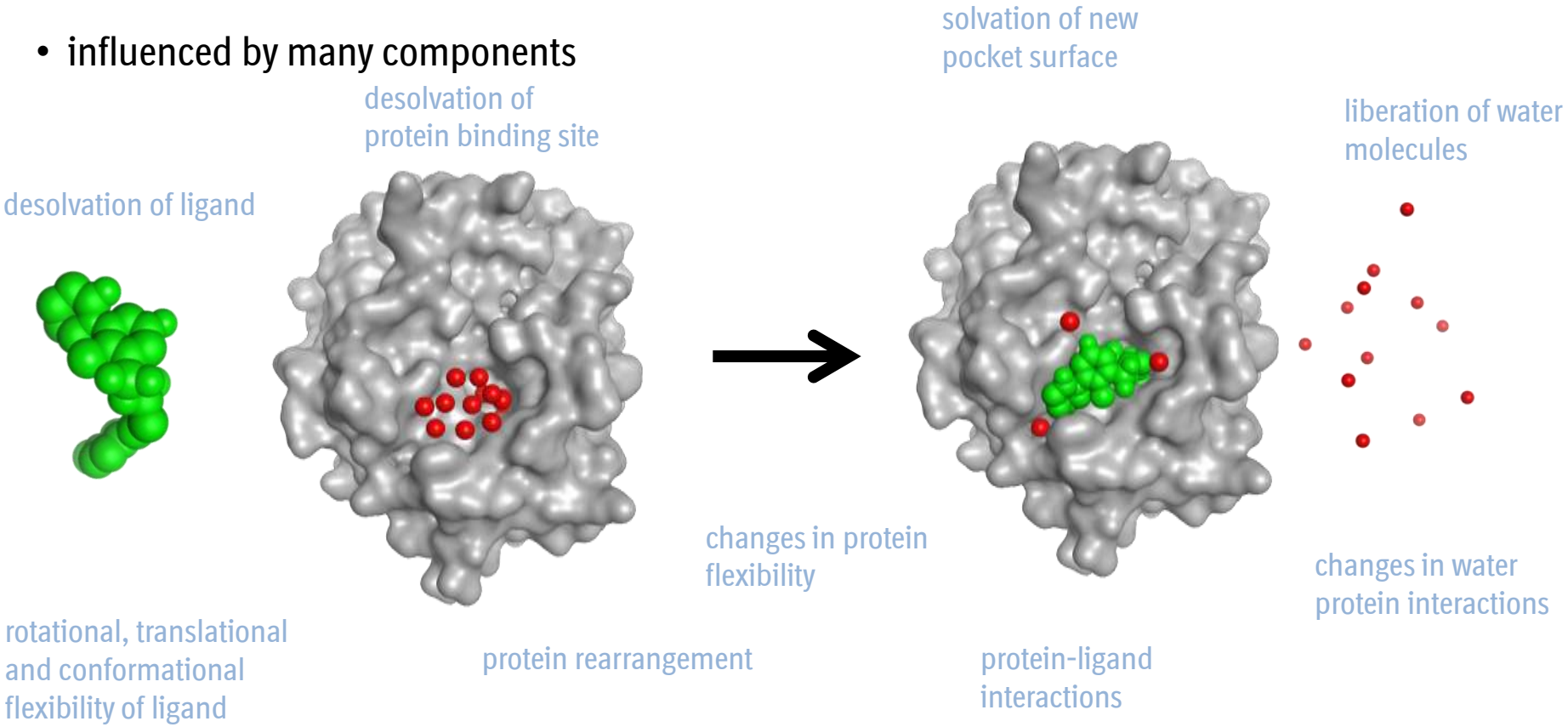
Summary

- MD simulations with small chemical probes yield fragment maps which represent favorable interaction possibilities on the protein surface
- crystal water positions are well reproduced and water networks can be analyzed
- pharmacophores derived from fragment maps agree with crystal structures of protein/ligand complexes
- induced-fit/conformational selection be observed

- methodology provides valuable hints for ligand optimization

Protein Ligand Binding Affinity Prediction

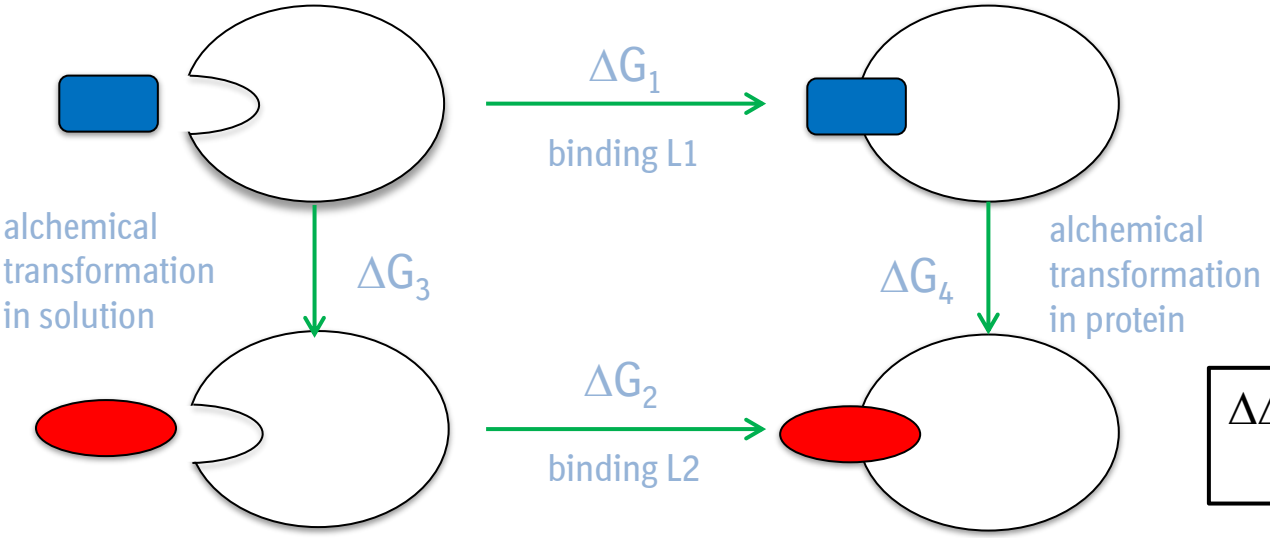
- influenced by many components



- need to be highly accurate
  Gibbs' fault: 1.4 kcal/mol ≡ factor 10 in affinity

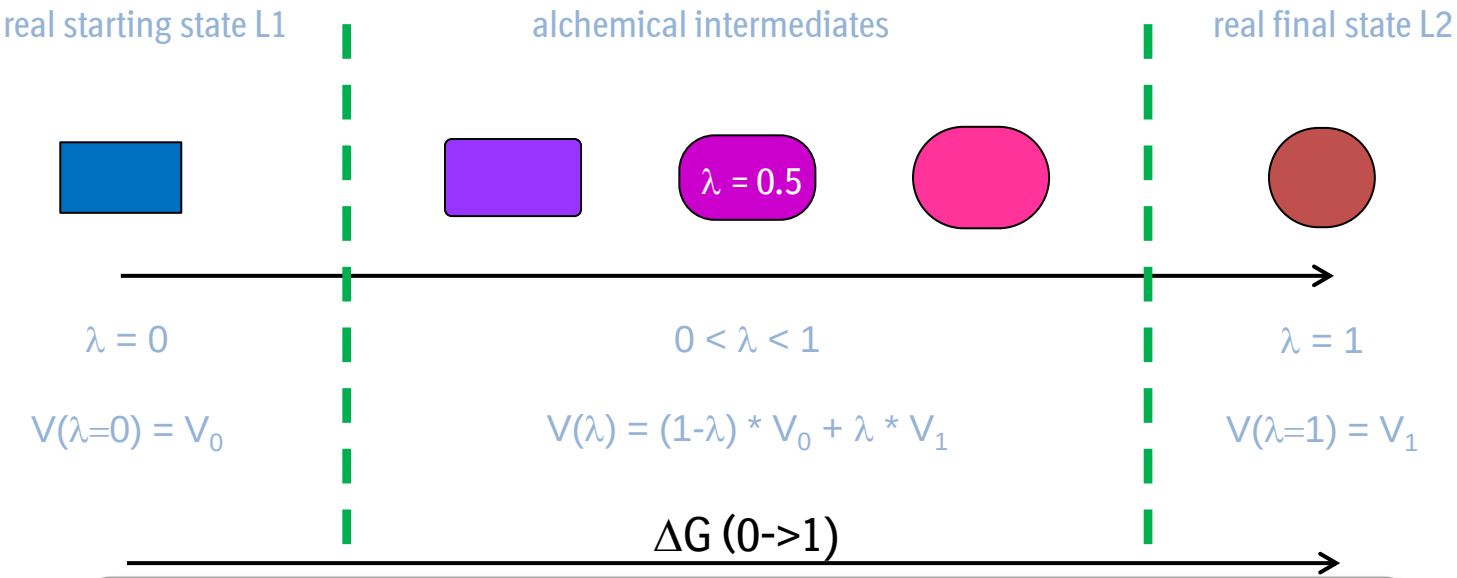
Relative Free Binding Energy Calculations

- much more efficient than absolute free energy calculations
- modeling of smaller changes should be more accurate
- relative differences probably more relevant in lead optimization
- currently the most used technique for rigorous calculations of binding energies
- compute difference between ligand 1 and 2 in a) solution and b) in the binding site



Boehringer Ingelheim

The Alchemical Transformation



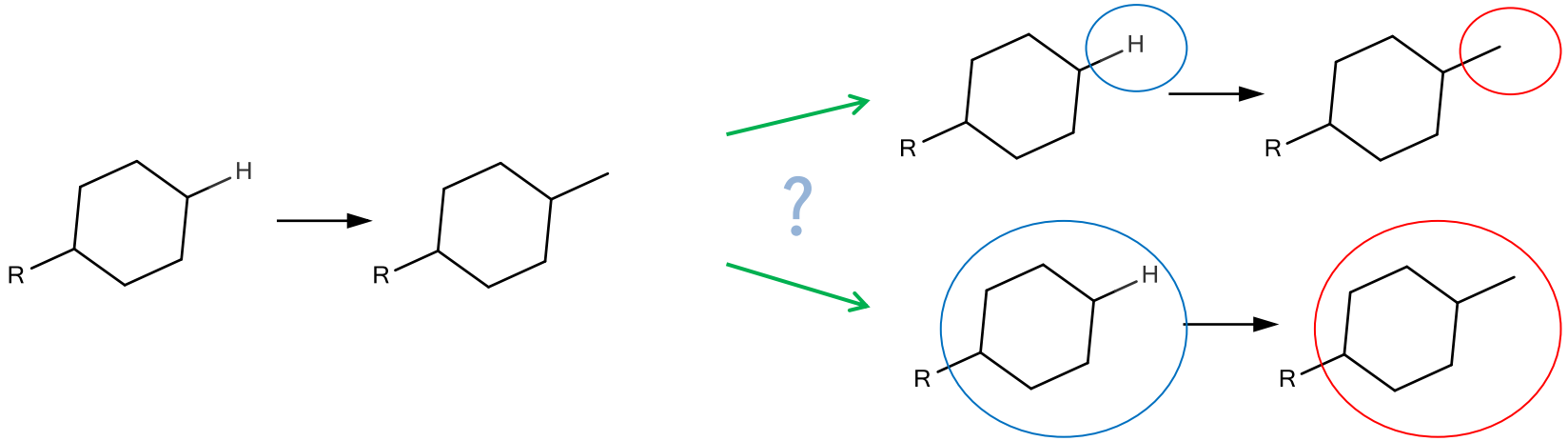
$$\Delta G = \int_{\lambda_A}^{\lambda_B} \frac{dG}{d\lambda} d\lambda = \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

Kirkwood, J. G. J Chem Phys **1935**, 3, 300

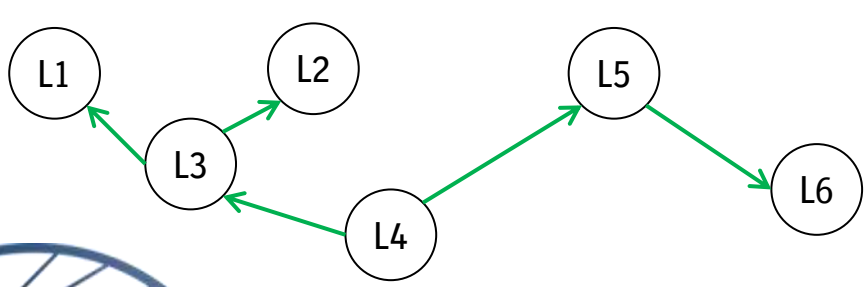
- need MD simulations at „intermediate“ steps between start and final state („λ-windows“)
- calculate ΔG either via Thermodynamic Integration or FEP (BAR or MBAR used in practice)
- absolute ΔG / ranking of ligands from a set of perturbations (solve linear equation system)

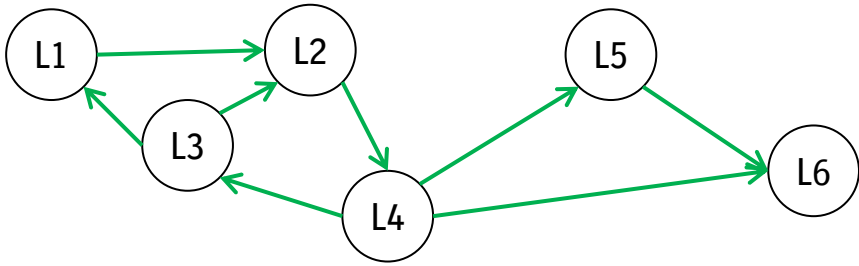
In-house Implementation / Adaption

- use TI engine as implemented in the AMBER MD software
- largely automated setup with little user intervention
- need to define size of changing part in the molecule („perturbation“):



- minimal spanning tree or closed cycles ?





Automated Setup

Input:
protein pdb file,
coordinates of bound ligands

generate amber parameter files
(prepin, frcmod) for ligands
(antechamber)

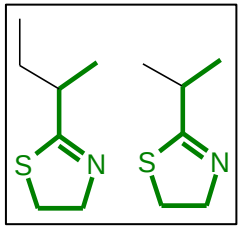
compute **maximum common substructure (MCS)**

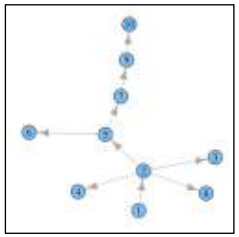
based on MCS: decide which ligands to connect via perturbation (e.g. minimal spanning tree)

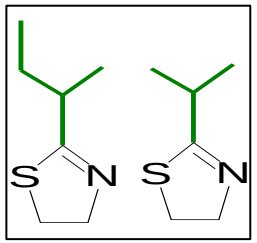
generate amber topologies and coordinates (leap, ff99SB/gaff-am1bcc)

decide on **perturbed atom selection**
(softcore regions)
Generate input and jobscript files

topologies, coordinates, input, job-scripts







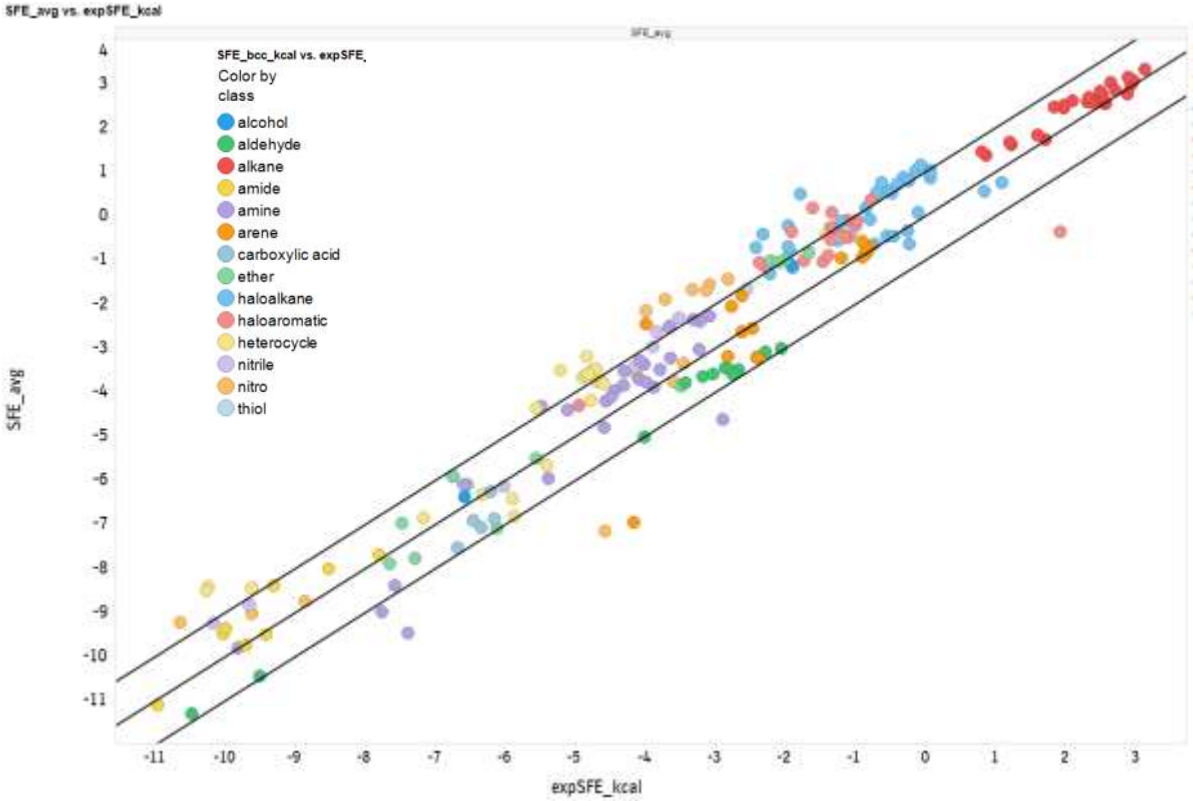
manual intervention possible/ necessary

implementation in svl, R, perl

Can it Work At All? Solvation Free Energies

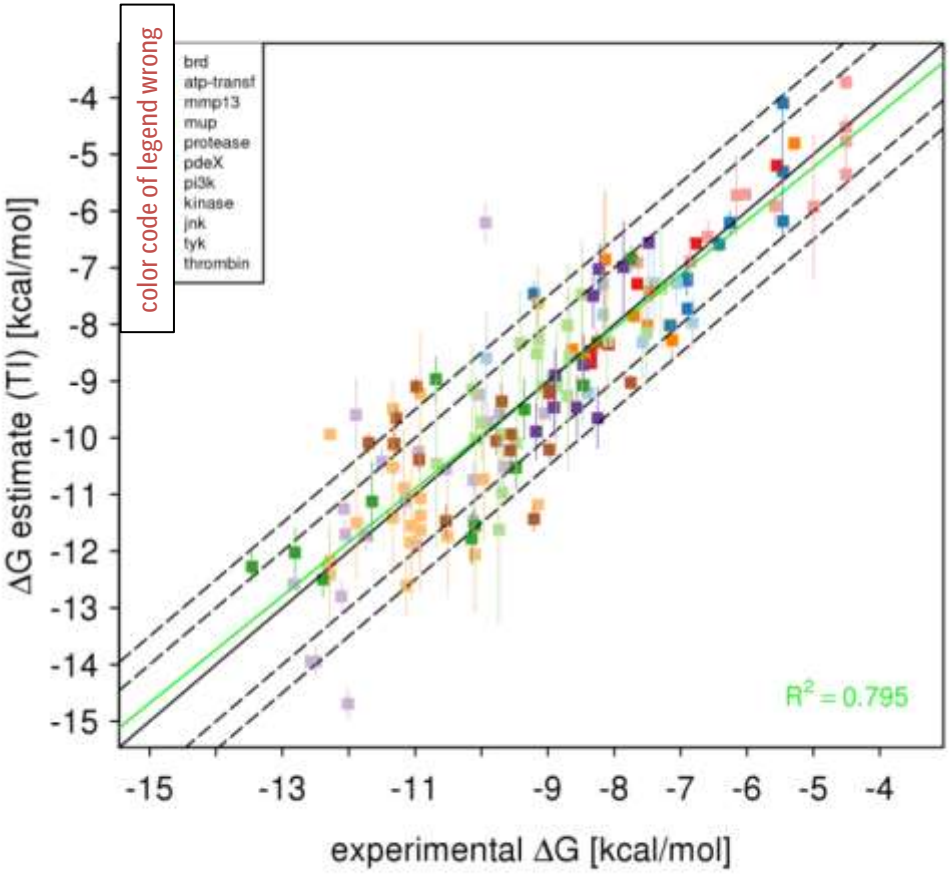
- disappear single molecule in water box
- less of a sampling issue, test of underlying FF
- dataset: 211 diverse small organic molecules
- average of 5 runs

- most compounds within the accepted error range
- discrepancies often can be explained by known deficiencies of the gaff-FF



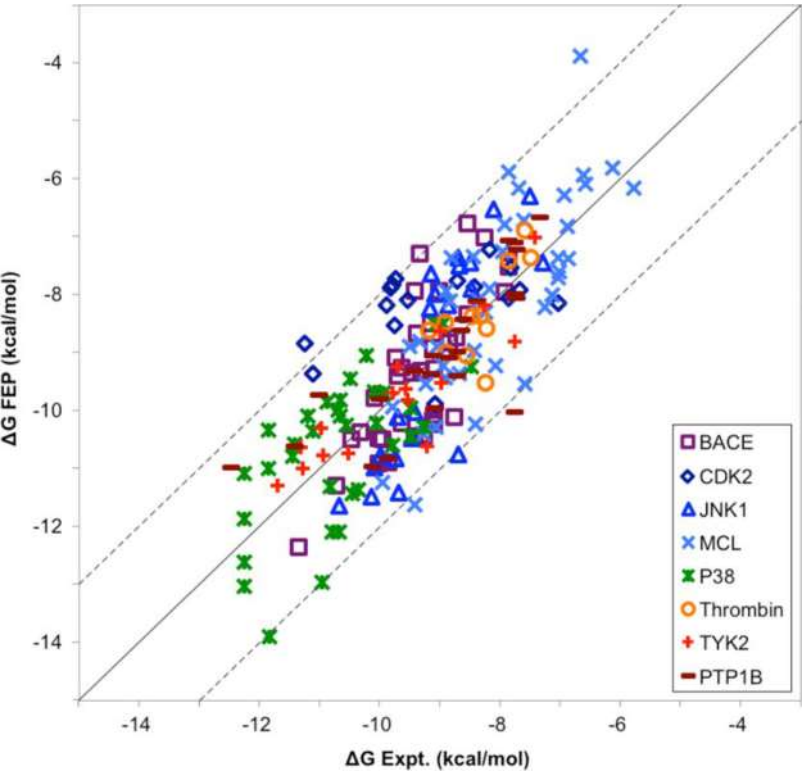
Overall Results

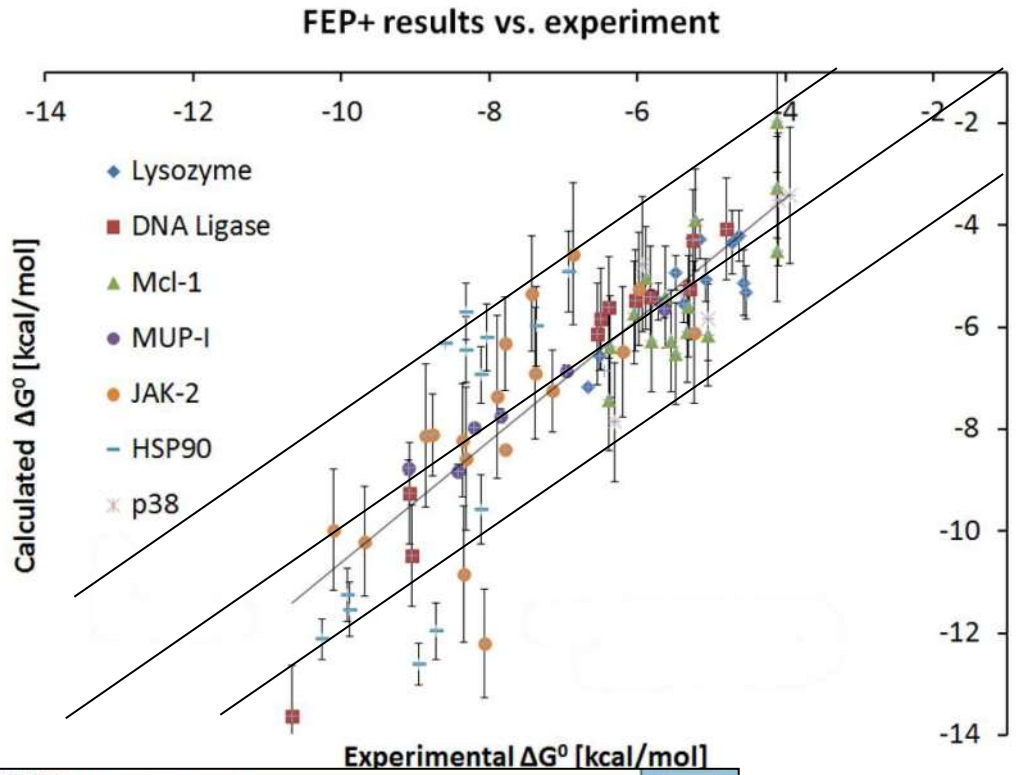
- 11 systems, 147 ligands
- overall, good agreement with experiment over large range of affinities
- 7 >2 kcal/mol error
- 130 <1.5 kcal/mol
- overall correlation: 0.8
- MUE = 0.75 kcal/mol
- RMSD = 1.0 kcal/mol



Boehringer Ingelheim

Schrödinger FEP





J|A|C|S JACS, 2015, 137, 2695–2703

Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field

Lingle Wang,† Yujie Wu,† Yuqing Deng,† Byungchan Kim,† Levi Pierce,† Goran Krilov,† Dmitry Lupyan,† Shaughnessy Robinson,† Markus K. Dahlgren,† Jeremy Greenwood,† Donna L. Romero,† Craig Masse,† Jennifer L. Knight,† Thomas Steinbrecher,† Thijs Beuming,† Wolfgang Damm,† Ed Harder,† Woody Sherman,† Mark Brewer,† Ron Wester,† Mark Murcko,† Leah Frye,† Ramy Farid,† Teng Lin,† David L. Mobley,‡ William L. Jorgensen,§ Bruce J. Berne,§ Richard A. Friesner,§ and Robert Abel*,†

JOURNAL OF CHEMICAL INFORMATION AND MODELING

JCIM, 2015, 55, 2411–2420

Accurate Binding Free Energy Predictions in Fragment Optimization

Thomas B. Steinbrecher,*,† Markus Dahlgren,‡ Daniel Cappel,† Teng Lin,‡ Lingle Wang,‡ Goran Krilov,‡ Robert Abel,‡ Richard Friesner,§ and Woody Sherman‡

†Schrödinger GmbH, Dynamostrasse 13, 68165 Mannheim, Baden-Württemberg, Germany
‡Schrödinger Inc., 120 West 45th Street, 17th Floor, New York, New York 10036, United States
§Department of Chemistry, Columbia University, 3000 Broadway New York, New York 10027, United States

-Energy

e Vries‡, Thea Mulder-Vlijmen†, Lingle

, Leiden 2300 RA, The Netherlands

‡ Schrödinger, Inc., 120 West 45th Street, New York, New York 10036, United States

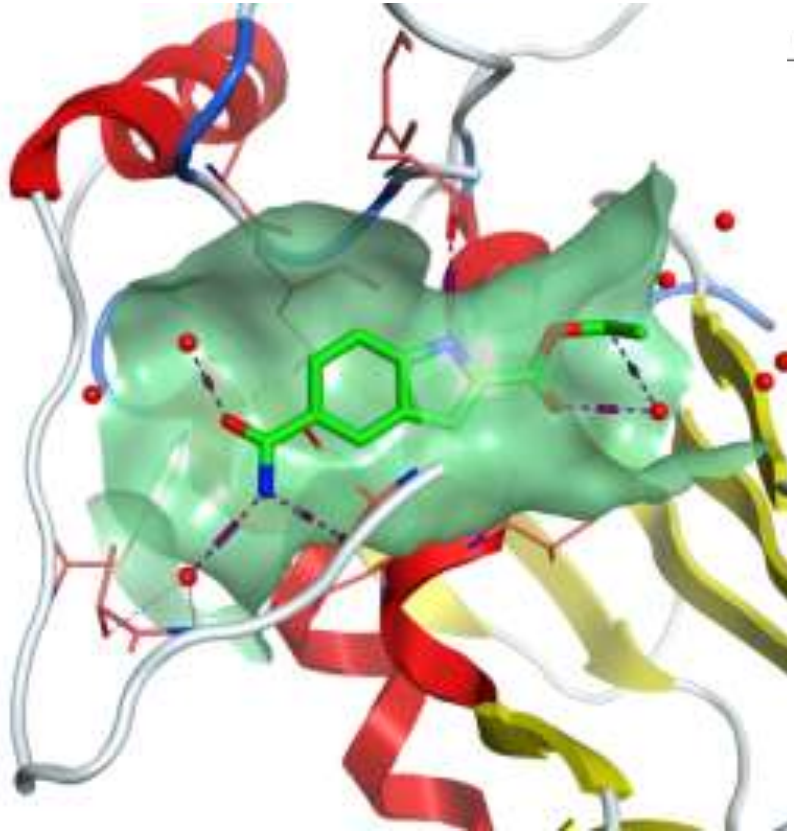
ACS Omega, 2016, 1 (2), pp 293–304

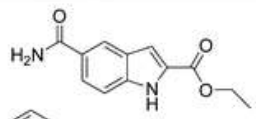
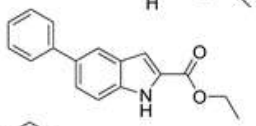
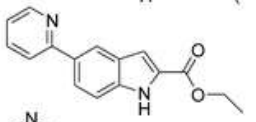
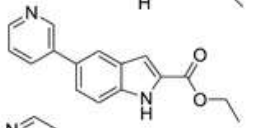
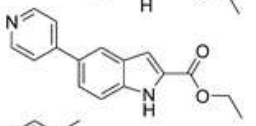
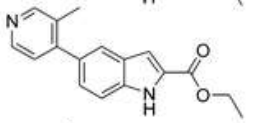
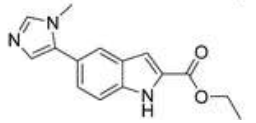
Is it really useful?

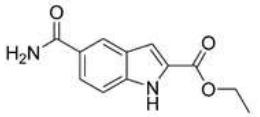
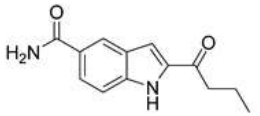
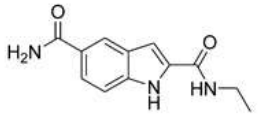
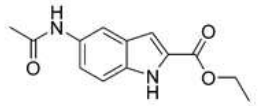
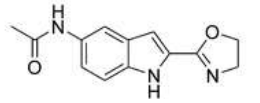
- a few examples from literature data sets and in-house projects:
  - MMP13
  - in-house protein kinase
  - Phosphodiesterase 5A

- in-house example which takes advantage of limited sampling

MMP13



Cpd.	Structure	MMP-13 IC_{50} (LE)[a]
1		39 (0.35)
6		>500
7		>500
8		15 (0.32)
9		82 (0.29)
10		9.6 (0.32)
11		2.5 (0.38)

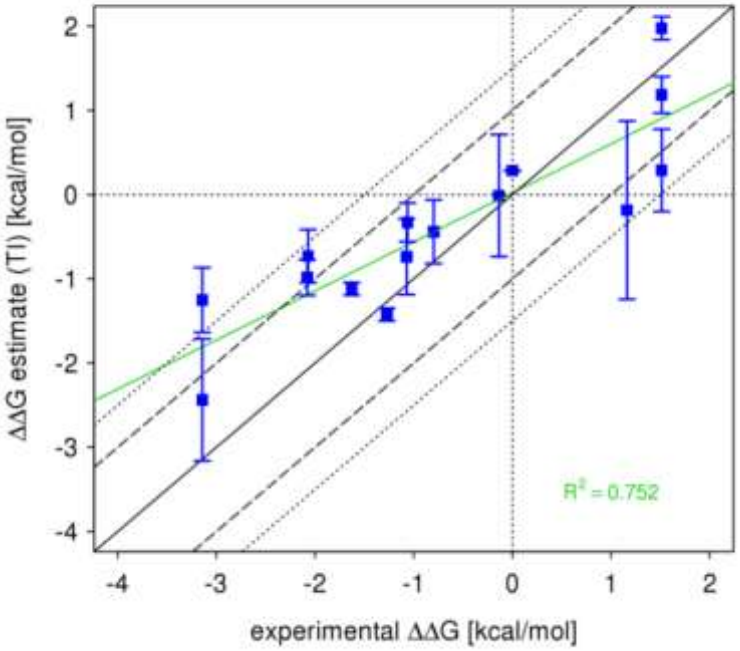
Cpd.	Structure	MMP-13 IC_{50} (LE
1		39 (0.35)
2		>500
3		>500
4		31 (0.34)
5		220 (0.27)

Taylor et al., J. Med. Chem. 2011, 54, 8174

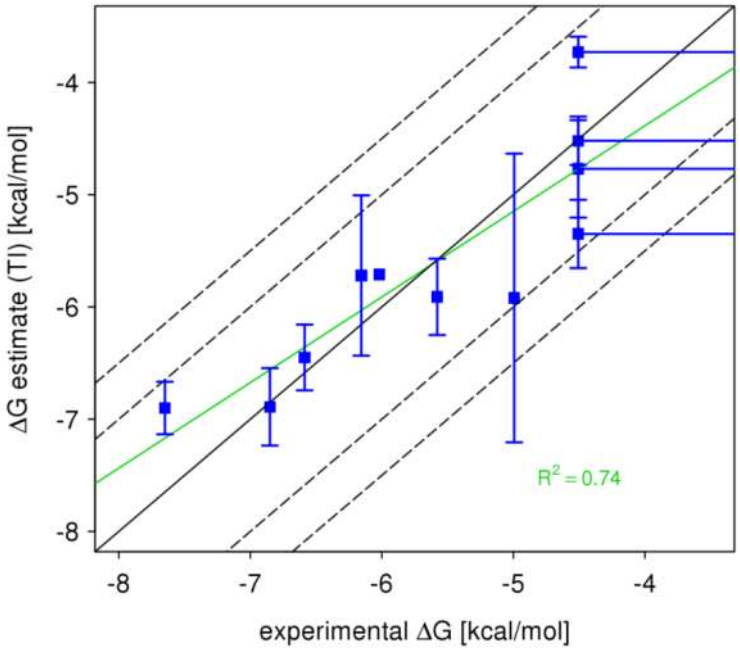
- 12 ligands from 2.5 → >500 µM
- binding modes modeled by analogy to the Xray structures of cpd1 and cpd11

MMP13 - Results

relative ΔΔG

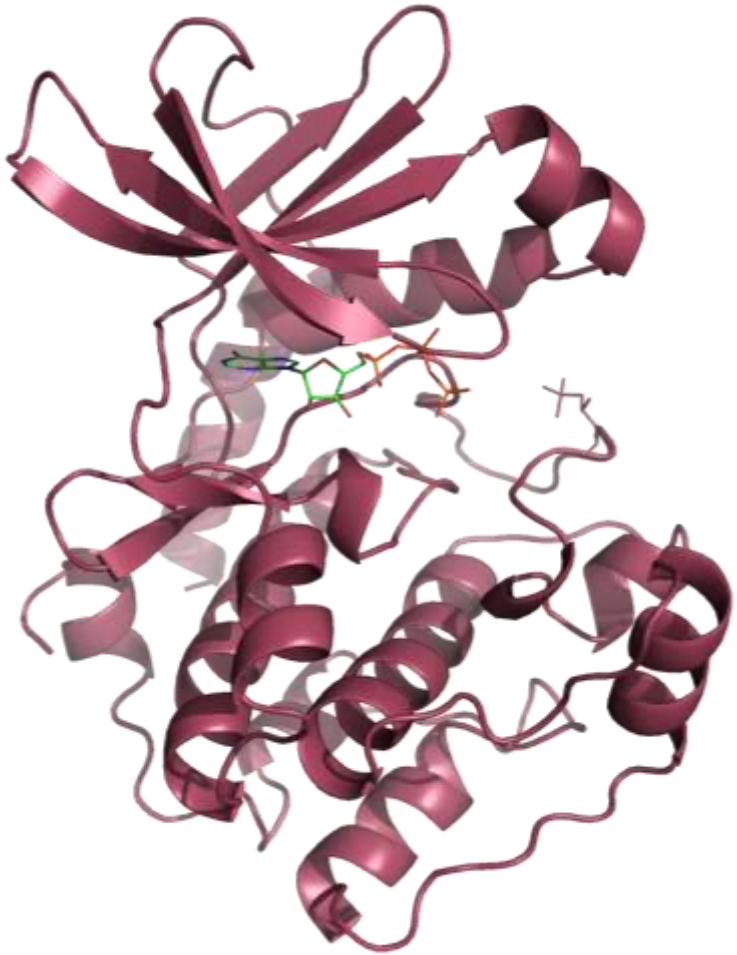


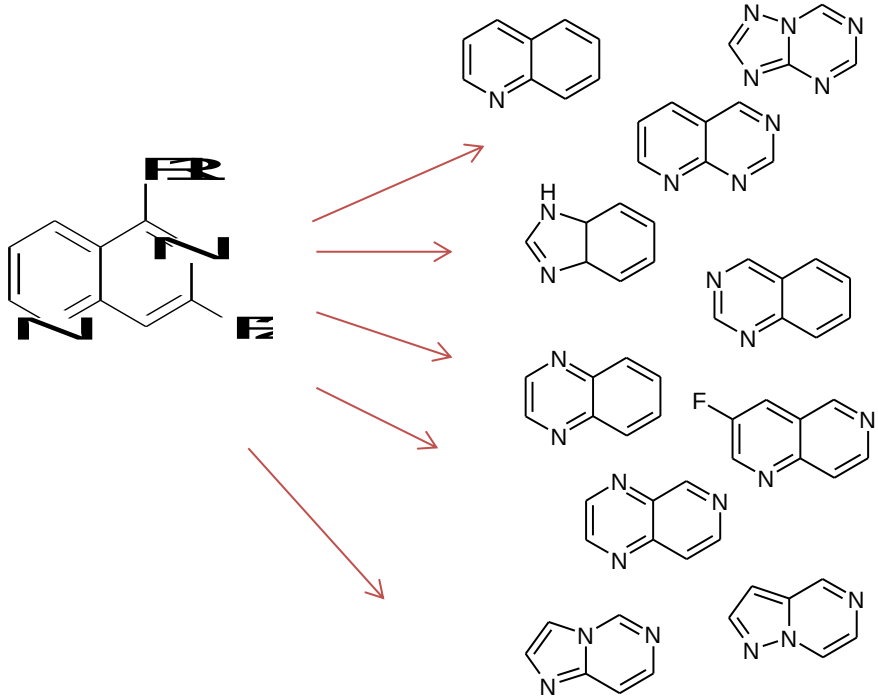
absolute ΔG



- ΔΔG: all but one within 1.5 kcal/mol, 10/15 within 1 kcal/mol
- experimental trends well reproduced
- ΔG: clear separation between high- and low-affinity ligands
- large error bars for ligands where multiple binding conformations look reasonable

The Masked Kinase („TMK“) - Core Modifications of an active ligand

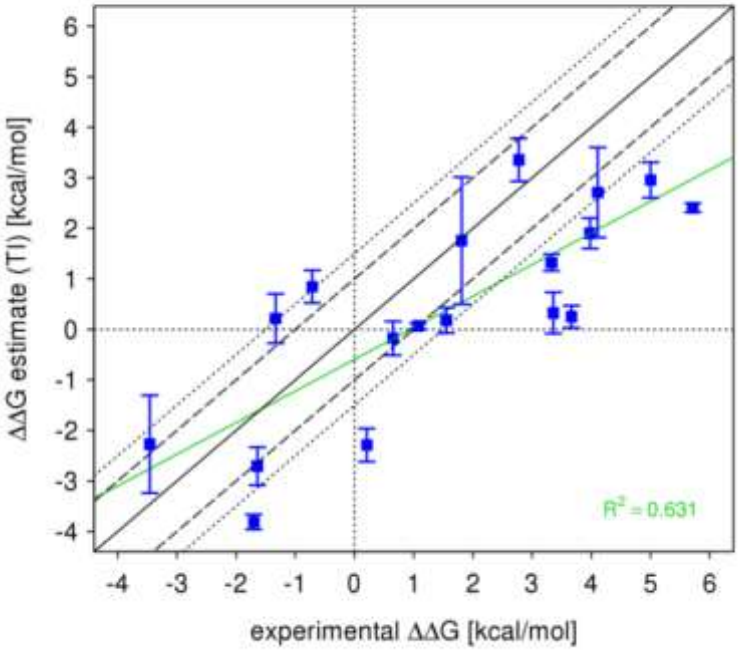




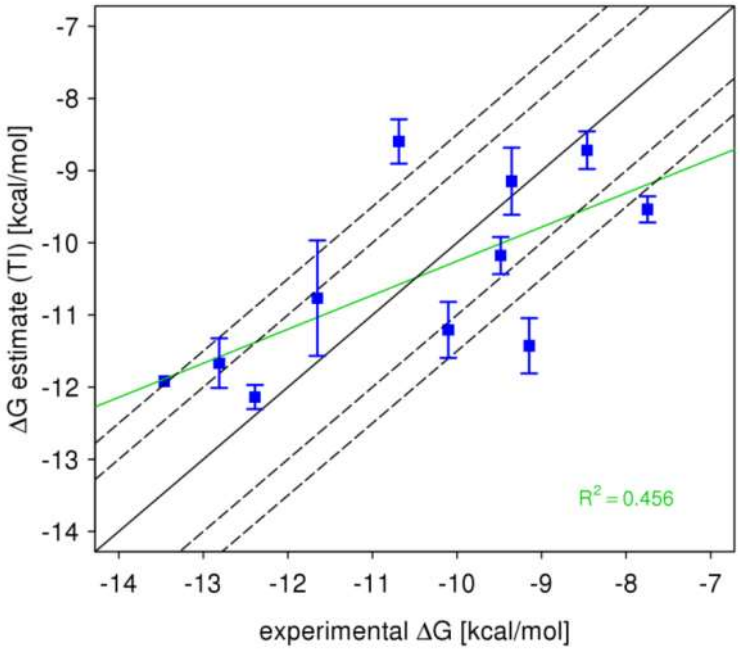
11 ligands with IC50 from 0.13 to 2100 nM

"TMK" – Core Modifications: Results

relative ΔΔG

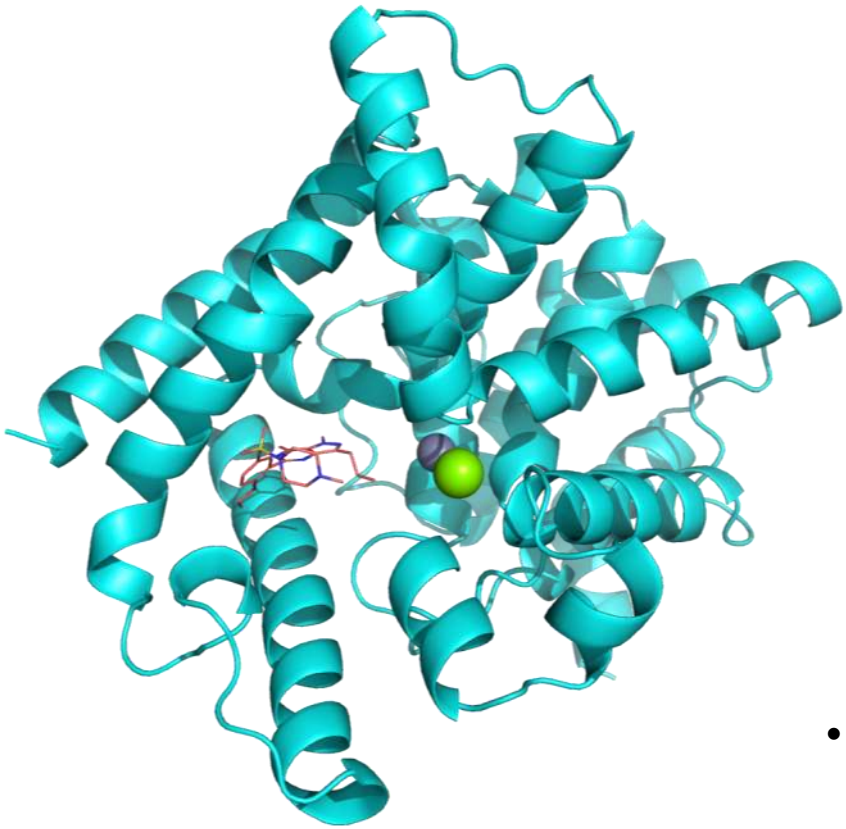


absolute ΔG

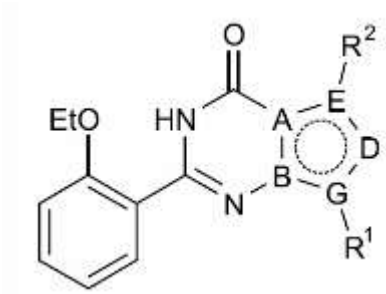


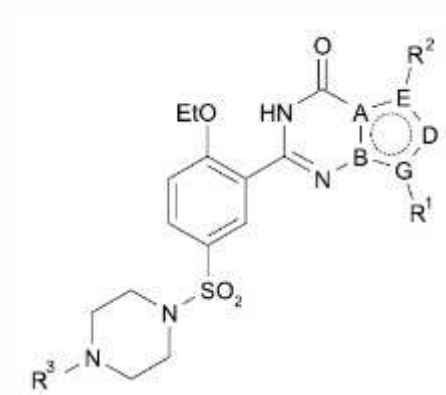
- larger errors >1.5 kcal/mol observed, nevertheless still prioritization of cores possible

PDE5A



- of different heterocyclic scaffolds as substrate analog PDE5A inhibitors



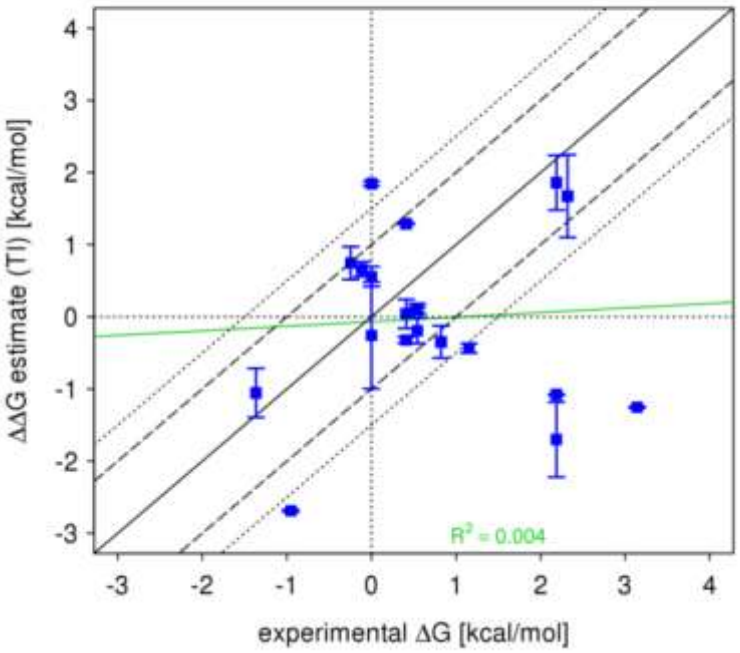


Haring, Bioorg. Med. Chem. Lett. 15 (2005) 3900)

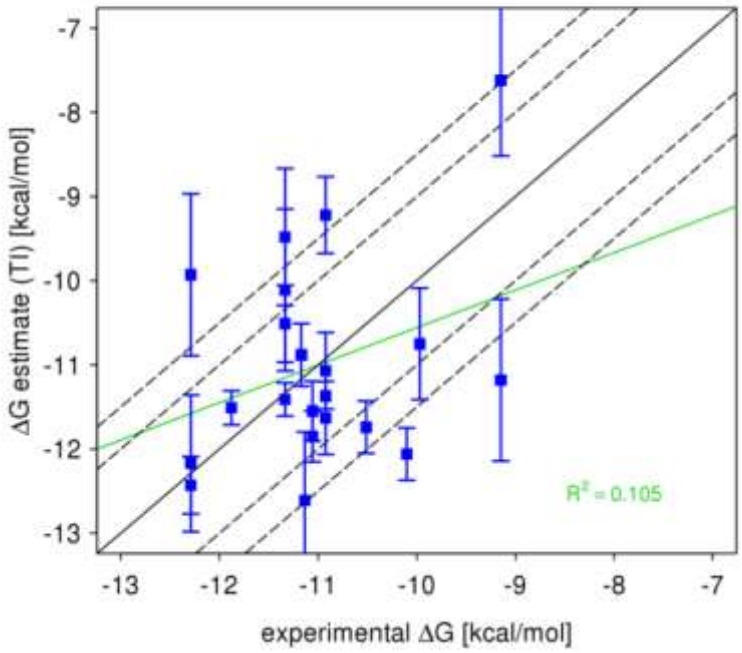
- 32 compounds with PDE5 IC50 from 1 to 200 nM

PDE5A - Results

relative ΔΔG



absolute ΔG



- most of the individual perturbations are within 1.5 kcal/mol error margin and have only small hystereses
- ΔG values with large errors, no correlation with experiment

Sources of Error

- inaccurate model ('force field') - especially for arbitrary organic molecules or metal ions in the binding site

- insufficient sampling - lack of sufficient criteria for convergence
  - statistical error estimates from a single FE simulation severely underestimate sampling error
  - forward/backward convergence is necessary but not sufficient condition for convergence
  - cannot estimate error bar from a single simulation

- role of water molecules
  - energies of displaced water molecules hard to assess
  - displacement of buried waters poses sampling issue

Summary

- setup of TI calculations largely automated
- TI calculations close to matching time lines of projects (goal ~1-2 days, achievable with GPUs)
- albeit no quantitative agreement, in most cases TI calculations very valuable in prioritizing synthetic efforts
- very difficult to track problems
- cannot estimate error bar from a single simulation
  → cannot say how confident I am about a result
  → multiple replicas

- make sure you throw your CPUs at a problem where the chemists do not outpace you with a small library
  – substituent scan ??
  – core modifications which usually mean a completely different synthetic route

MD Simulations in Drug Design
Summary

- MD simulations are starting to become a standard tool in drug discovery
- from expert tool to routine application
- combines physics based methodology, explicit water treatment, and inclusion of protein flexibility to obtain a realistic model of the system of interest
- many of the earlier obstacles that hampered its use are being overcome
  - increase in computer power and GPU/cloud computing
  - easier and more intuitive user interfaces for simulation setup and analysis

several in-house examples where MD simulations had impact on project progress

Thanks to

- University of Innsbruck (AK Klaus R. Liedl)
  - Hannes G. Wallnoefer (now Sandoz)
  - Anna S. Kamenik

- BI CompChem Group
  - Clara Christ (now Bayer Pharma)
  - Julian Fuchs (BI Wien)
  - Sandra Handschuh
  - Jan Kriegl
  - Uta Lessel
  - Daniel Seeliger
  - Christofer Tautermann
  - Alex Weber
  - Bernd Wellenzohn