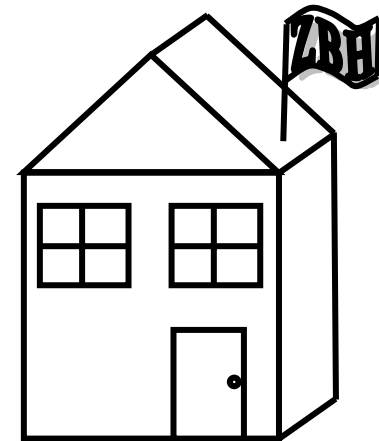
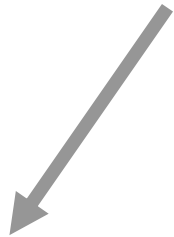


# Centre for Bioinformatics and Inhabitants

- 3 groups
- long and boring overview
- my interests – numerical methods, simulations
- corporate identity



# Corporate identity



# Overview

## Teaching

- Masters in Bioinformatics
  - after Bachelors in chemistry, biology, informatics +...

## Research

- Matthias Rarey – drug design, chemo-informatics
- Stefan Kurtz – genome comparisons, DNA microarrays
- Andrew Torda - numerical modelling, simulation proteins, RNA

## More info

`www.bioinformatics.uni-hamburg.de`

Is this really being filmed by ZDF ?

**Hi**

**Mum**

# Why do we like bioinformatics ?

Masses of biological data

- DNA sequences, protein structures

Either

- find patterns in gigantic data sets

or

- propose experiments

To be useful

- Do something that could not be done (patterns)
- Do something cheaper
  - drug proposal
  - predict protein structure

# Rarey: molecular docking

## Problems

- A protein is involved in disease
  - should be blocked
  - propose candidates for experiments (leads)
- Technical issues
  - how similar are molecules

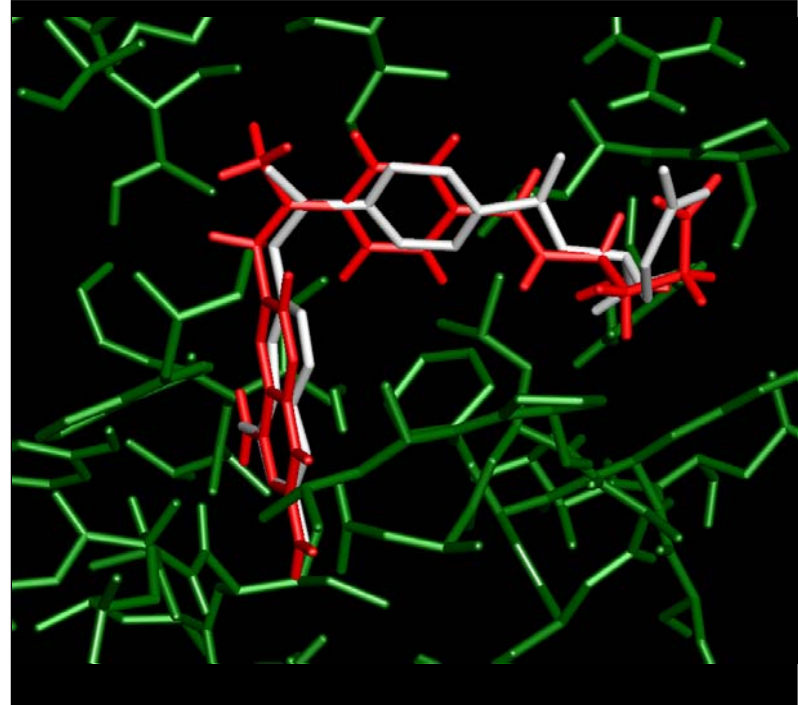
# Rarey group: Molecular Docking

## Incremental Construction:

- Select base fragments
- Place base fragments
- Sequentially grow ligand into active site

## Advantages:

- detailed interaction model
- full ligand flexibility (incl. rings)



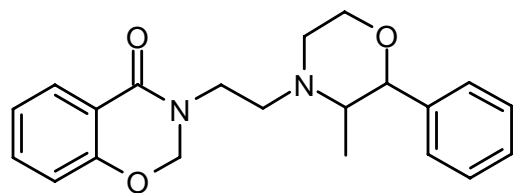
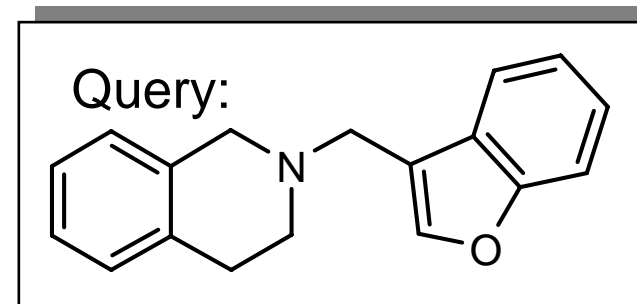
## FlexX software

- computing time ~1 min/cpd
- very widely used code for molecular docking

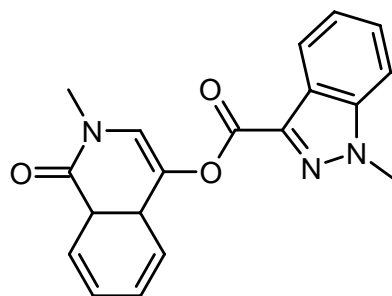
[www.biosolveit.de/flexx](http://www.biosolveit.de/flexx)

# Example : Target Similarity Values

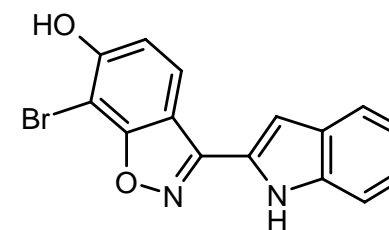
- Dopamine D4 antagonists
- Assessing target similarity values
- Search in WDI+D4 space @ **0.85**



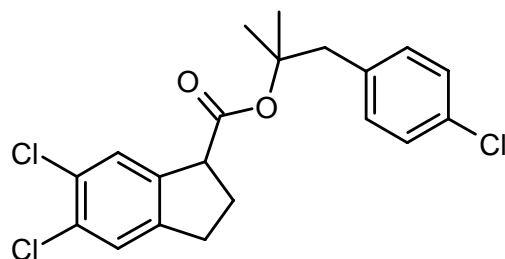
1



2

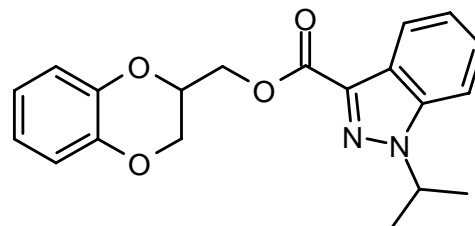


3



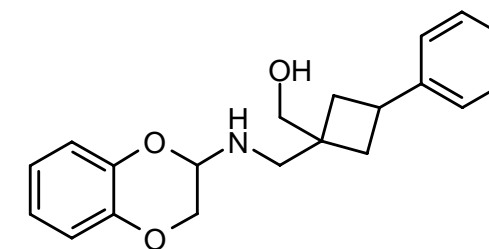
4

4



5

5



6

6

# Kurtz group - genomes

## Genomes

- huge strings
- index structures for bio-sequences
- similarity based gene prediction
- genome comparisons
- analysis of repeat structures
- search for complex patterns

# Kurtz group – metabolism / pathways

## Applications

- Techniques
  - statistical evaluation
  - database design
  - pathway modelling
  
- Topics
  - prediction of splice variants
  - mRNA expression data
  - MPSS data
  - identification of metabolic targets
  - search for transposons

# Kurtz group - software

## Suffix trees

- REPuter: repeat structures in genomes
- MUMmer: pairwise comparison of genomes
- Vmatch: versatile software for matching large sequence sets
- MGA: multiple alignment of genomes
- PossumSearch: index based matching – position specific scoring matrices
- GenomeThreader: pipeline for genome structure prediction (homology based)
- LTRharvest: prediction of LTR transposons

**[www.zbh.uni-hamburg.de/research/GI](http://www.zbh.uni-hamburg.de/research/GI)**

**Kurtz group – [kurtz@zbh.uni-hamburg.de](mailto:kurtz@zbh.uni-hamburg.de)**

# Kurtz group - applications

## Collaborations with University Clinic

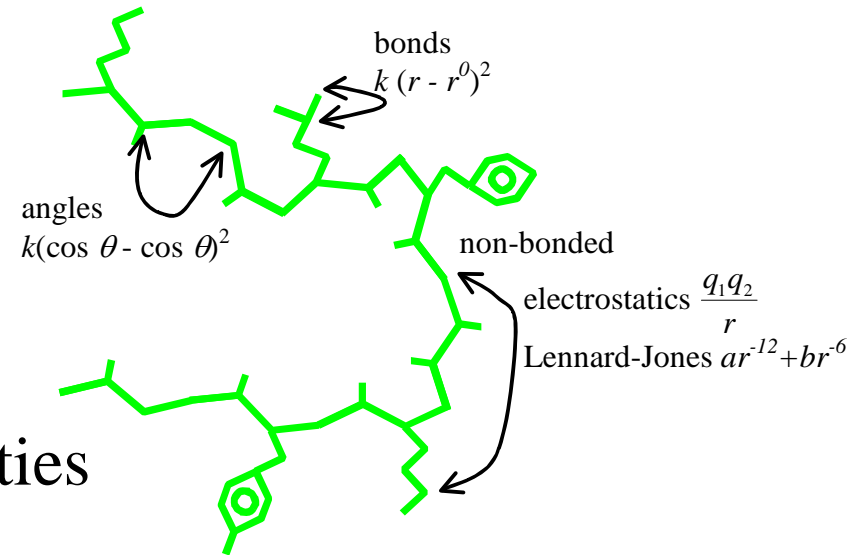
- identification of copy number variations (amplicons / deletions)
- combination of genomic and expression identification
- analysis of differential expression
- identification of splice variants
- global location of tyr phosphorylation in human proteome

# Torda group – numerical modelling

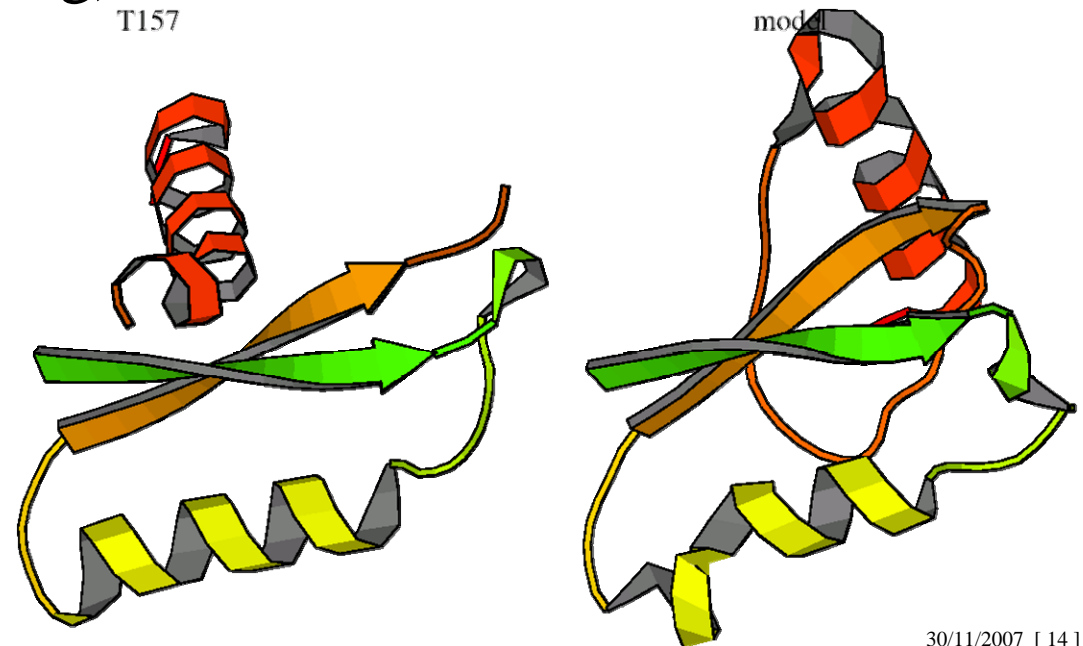
- protein
  - structures / force fields / comparisons
- RNA
  - structure prediction
  - design
- side projects
  - multi-time scale evolution, ecosystems

# Protein score functions

- what is a force field ?
- can we use probabilistic methods ?
  - Bayesian classification
  - from known to unknown properties



- structure prediction (threading)



- something more useful..

Torda group (me)

# Structure alignments

- 47 000 protein structures known
  - comparison methods
  - finding similarities
    - function prediction
  - normally NP-complete

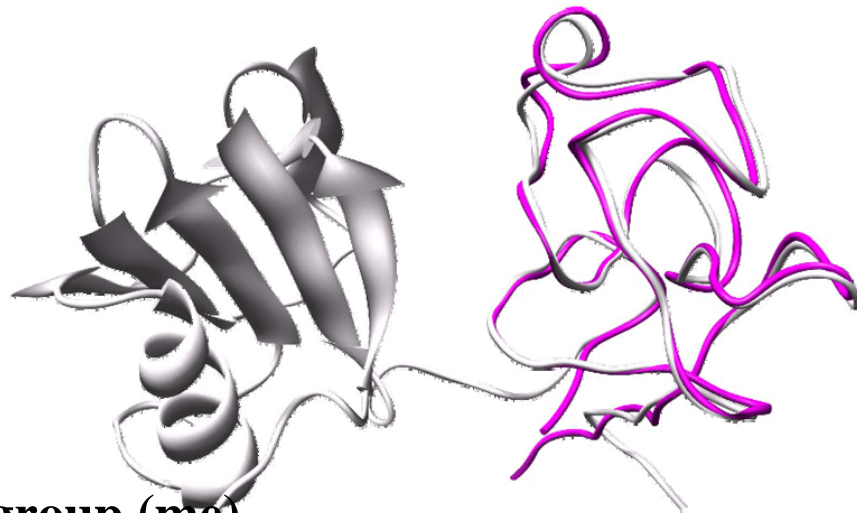
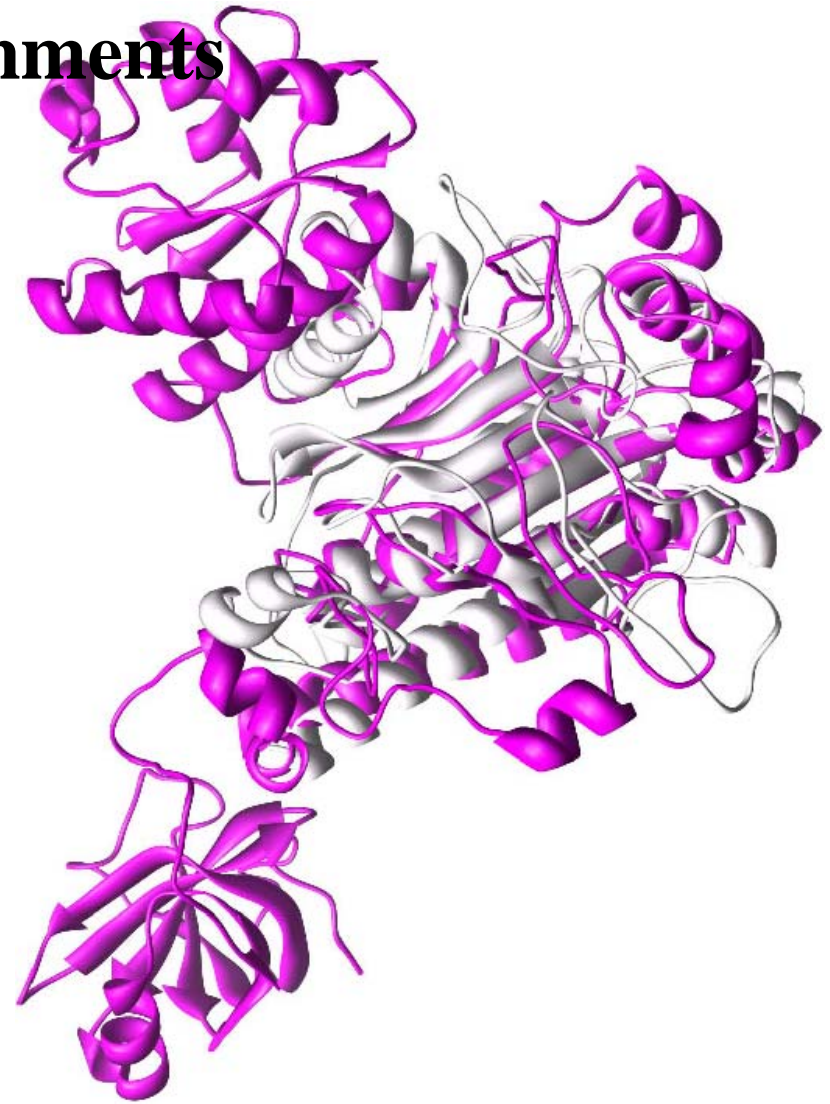
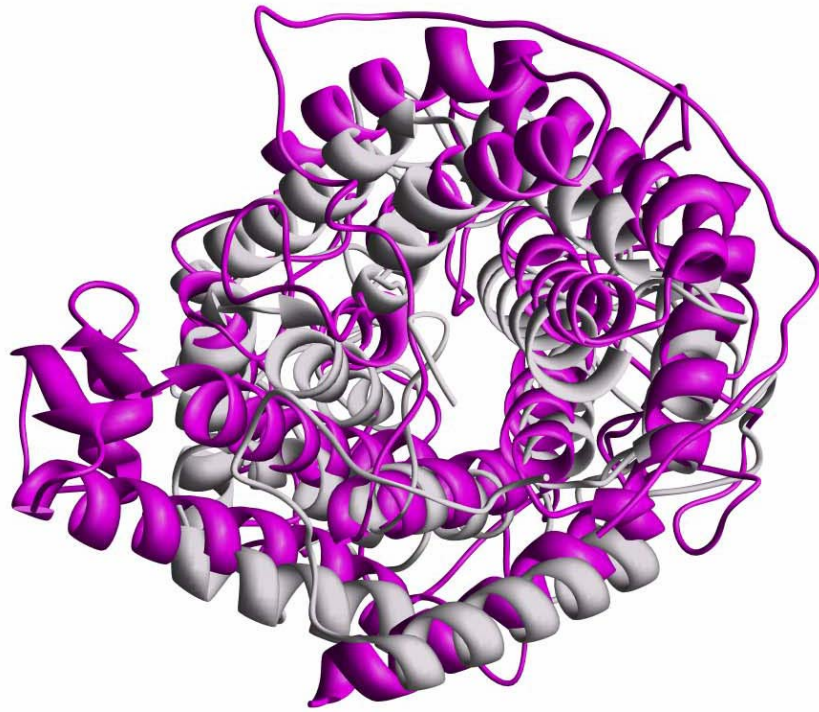


Gundolf Schenk

`temporary cardigan.zbh.uni-hamburg.de/~tmargraf`

**Torda group (me)**

# Structure alignments



**Torda group (me)**

- common domains
- repetitive structures
- irregular structures

# Structure alignments

- now used for bulk phylogenies, classification
- play today  
    temporary `cardigan.zbh.uni-hamburg.de/~tmargraf`
- very fast methods
- code being packaged for release now

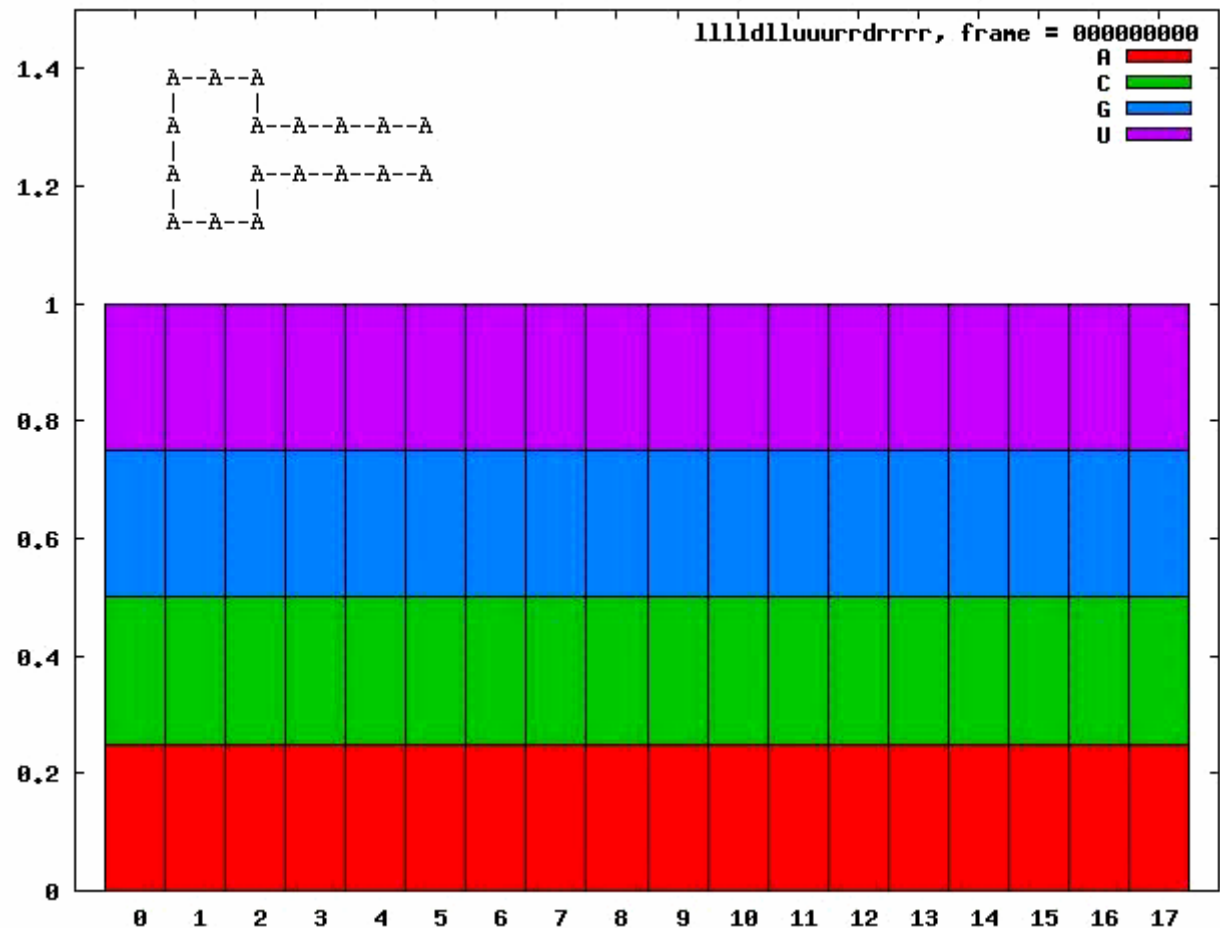
# Molecular design

- A nature in molecule is useful
  - enzyme, selective binding (sensors)
- Can we change the building blocks of the molecule ?
  - maintain shape
  - more stable
- RNA example
  - molecule length 50
  - possible sequences  $4^{50}$  about  $10^{30}$
  - too many
- naughty movie ...

# RNA design

- Given a structure can I design a new sequence ?
- Dynamics in sequence space

- newtonian dynamics
- read out sequence at end



# Summary

- Rarey - Drug design, small molecule similarity
- Kurtz - Genomes – fast string comparison
- Torda - Numerical methods, molecular simulation / design