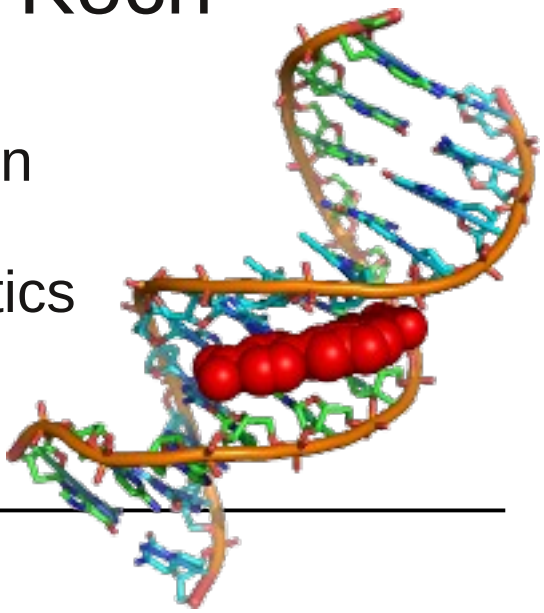
GCB 2012

Computation and visualization of protein topology graphs including ligands

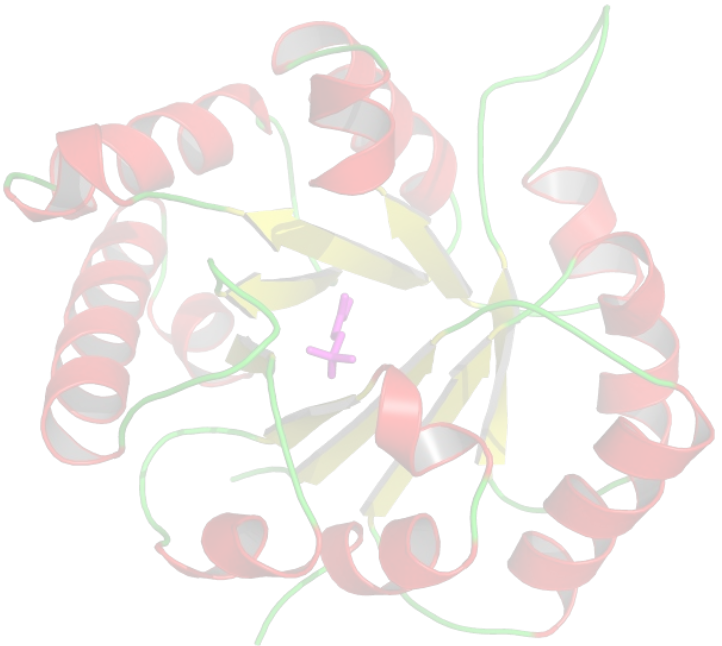
Tim Schäfer, Patrick May, Ina Koch

Goethe-University Frankfurt am Main
Institute for Computer Science
Department for Molecular Bioinformatics



Contents

- Introduction
  - Protein structure, ligands and graphs
- Methods
  - Computing protein ligand graphs from PDB and DSSP data
  - Types of protein ligand graphs
  - Visualization of protein ligand graphs
- Results
  - The VPLG software
  - Case studies

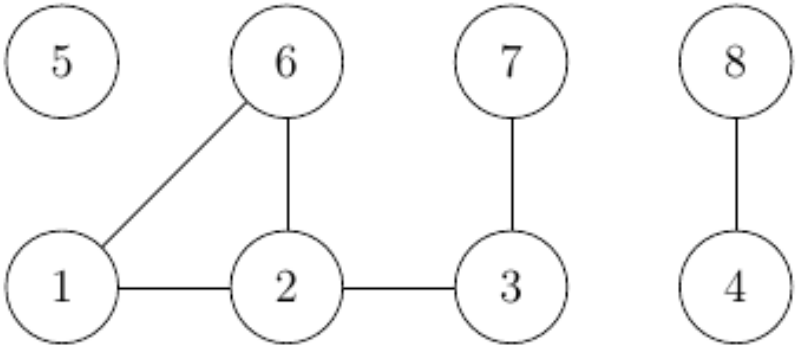


Protein structure

- Primary structure
  - String of 20 AAs
  - 3D: peptide bonds, steric constraints, Ramachandran
- Secondary structure
  - local, H-bonds, SSE types
  - super-secondary structure
  - protein domains
- Tertiary structure
  - atom coordinates
- Quaternary structure
  - multi-chain proteins

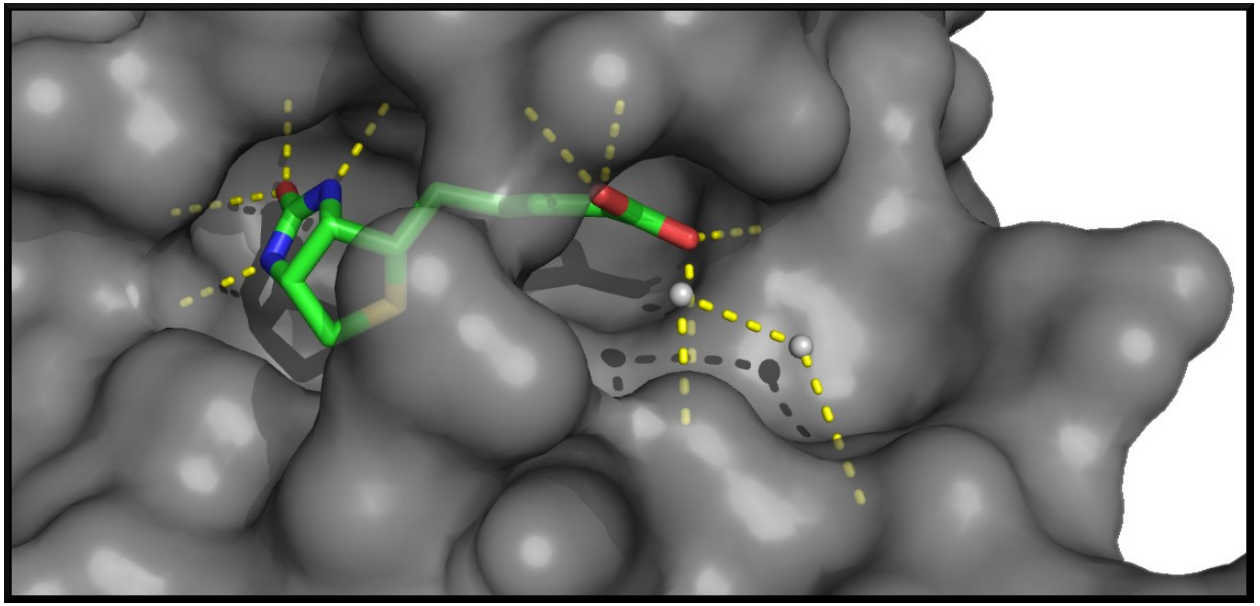
Modeling protein structure as a graph

- Graph G = (V, E)
- Pervasive data structure in computer science and bioinformatics
- Usage of graps to model proteins
  - CATH [Orengo et al.]
  - SCOP [Murzin et al.]
  - TOPS+ [Gilbert, Veeramalai]
  - PTGL [May, Koch]



Protein ligand interactions

- Proteins interact with their environment to fulfill their biological function: other proteins, ligands, ...
- Knowledge on PLI is important in many applications, e.g. drug design, molecular medicine
- > 10,000 different ligands occur in PDB files



A graph model of proteins on the super-secondary structure level

- Secondary Structure Elements (SSEs)
  - Few: 40,000 atoms => 400 residues => 40 SSEs
  - Automated assignment from 3D data possible (e.g., DSSP)
- Protein model
  - undirected, labeled **graph** for each chain of a protein
- Protein graph
  - vertices: SSEs or ligands
  - edges: contacts and spatial relations between them
  - graph types: alpha-beta-graph, alpha-graph or beta-graph

Protein ligand graphs

- Extension of protein graph model
- Protein ligand graph G = (V, E)
  - Vertex: SSE || Ligand
  - Edge: Spatial contact (parallel, antiparallel, mixed, ligand, backbone)

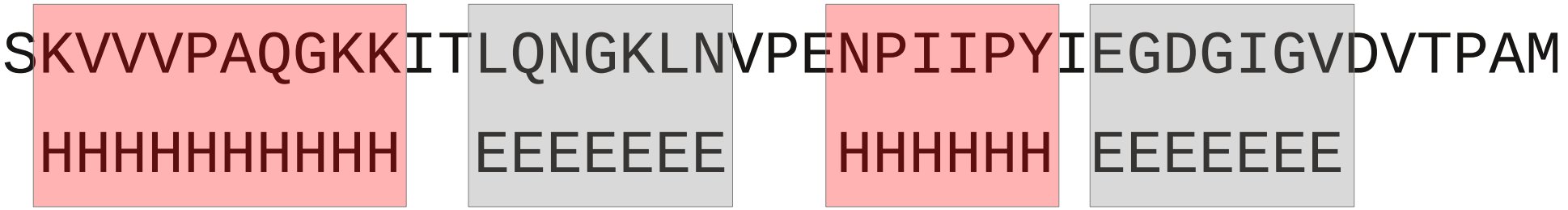
Computing protein ligand graphs

1. Obtain sequence from PDB file

SKVVVPAQGKKITLQNGKLNVPENPIIPYIEGDGIGVDVTPAM

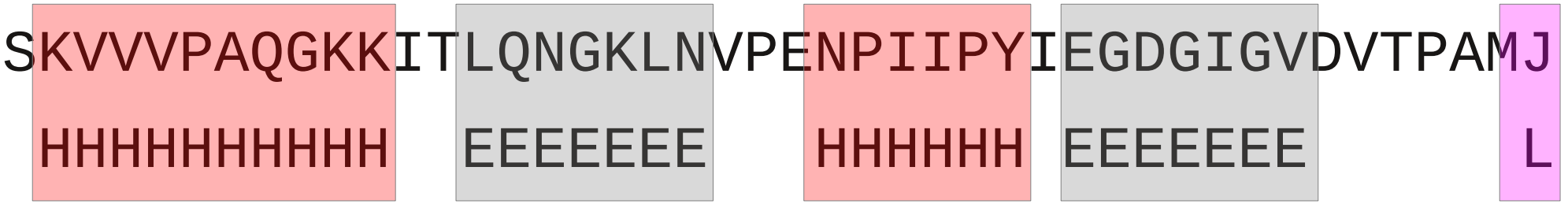
Computing protein ligand graphs

1. Obtain sequence from PDB file
2. Obtain secondary structure assignments from DSSP file



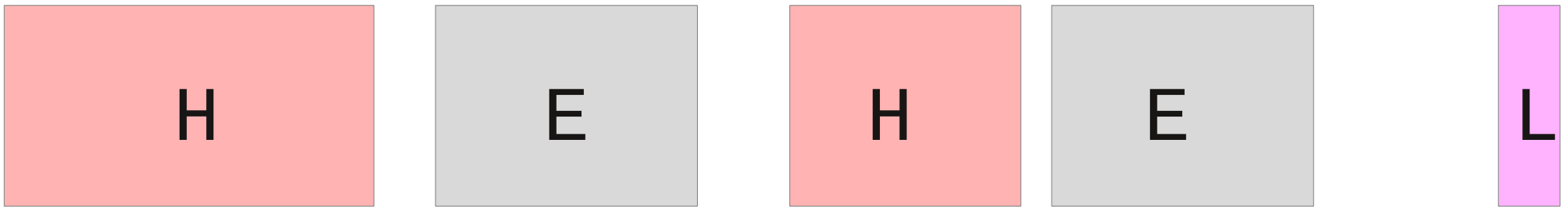
Computing protein ligand graphs

1. Obtain sequence from PDB file
2. Obtain secondary structure assignments from DSSP file
3. Add ligand information from PDB file



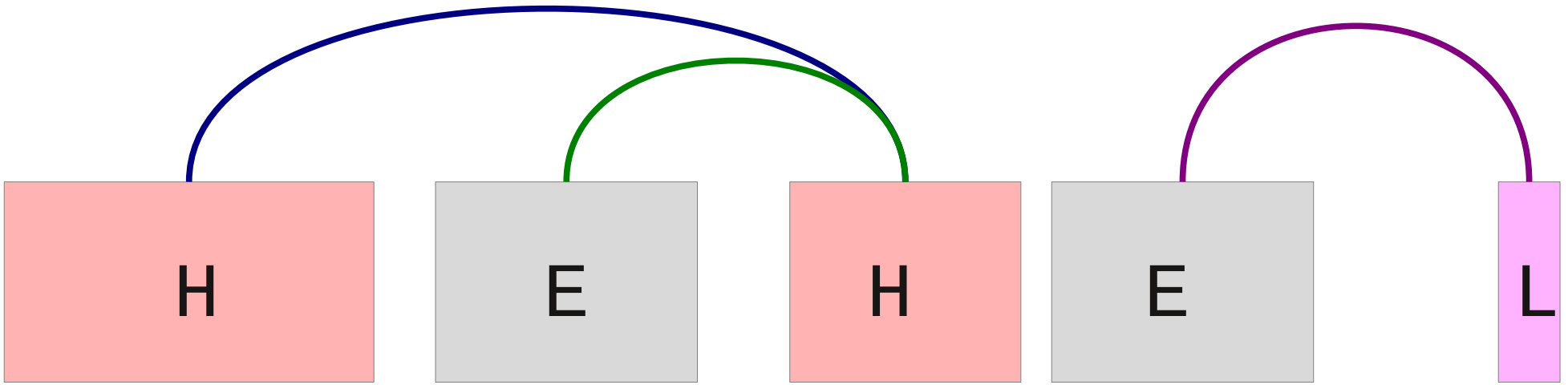
Computing protein ligand graphs

1. Obtain sequence from PDB file
2. Obtain secondary structure assignments from DSSP file
3. Add ligand information from PDB file
4. Build graph: add a vertex for each SSE

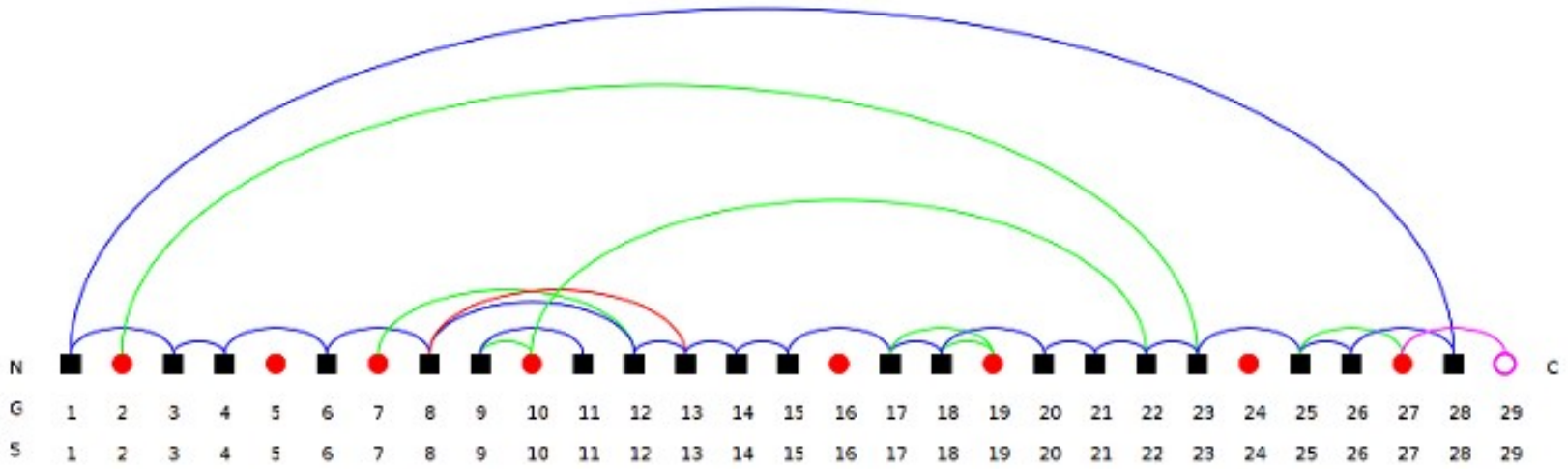


Computing protein ligand graphs

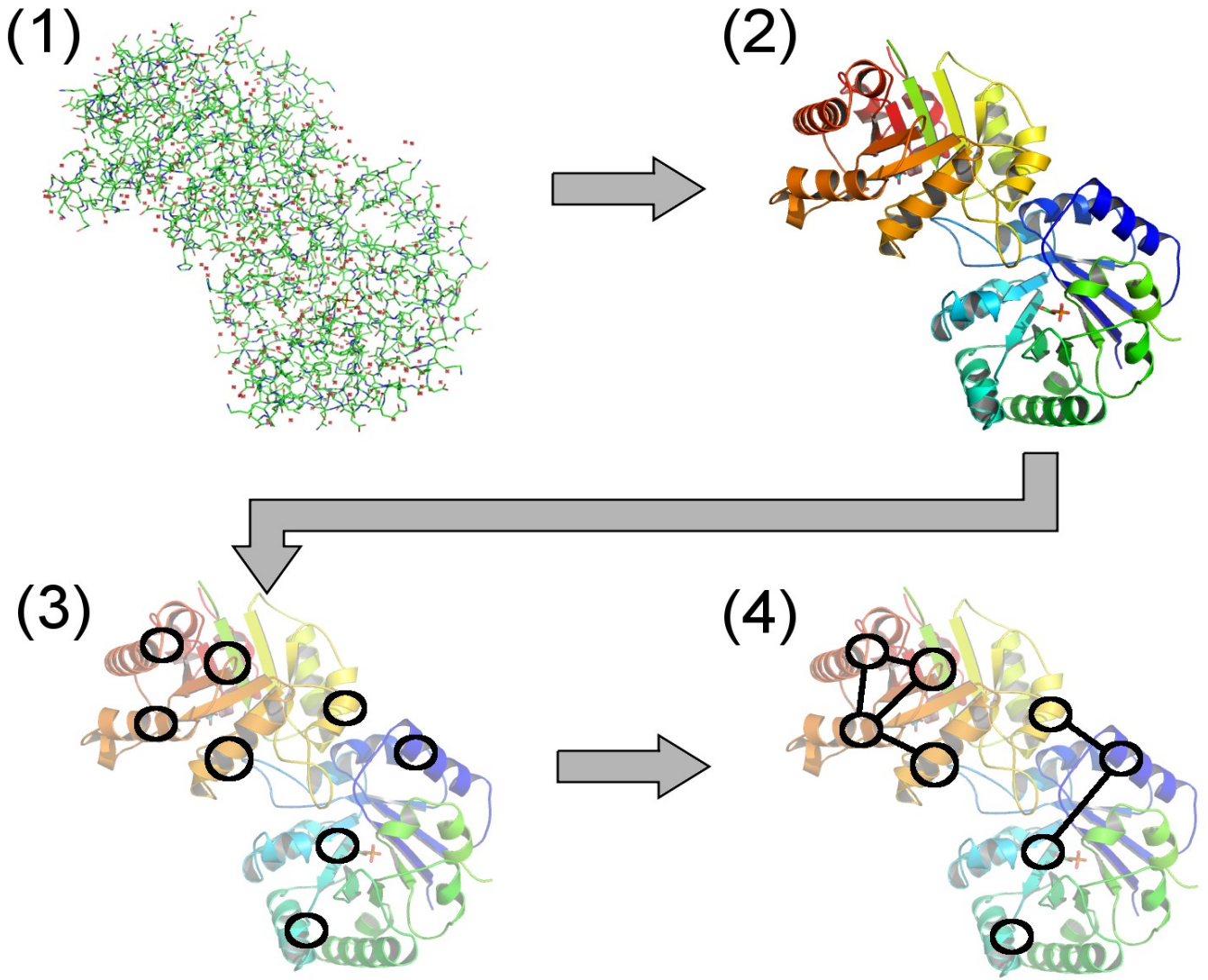
1. Obtain sequence from PDB file
2. Obtain secondary structure assignments from DSSP file
3. Add ligand information from PDB file
4. Build graph: add a vertex for each SSE
5. Compute spatial contacts and add edges between SSEs



VPLG – an open source implementation in Java

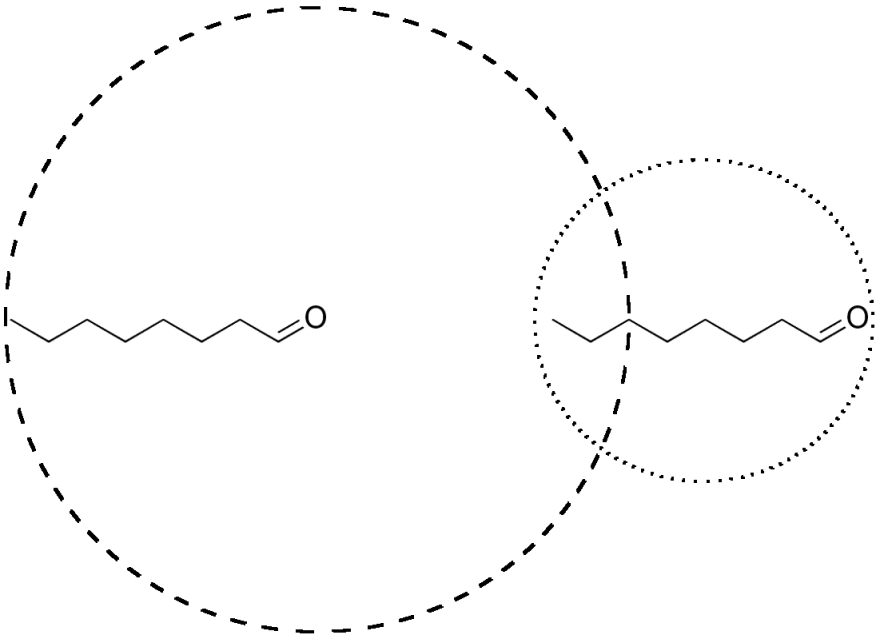


Overview:
Computing protein ligand graphs



Contact computation

- Parser for PDB and DSSP files
- Atom level contacts
  - vdW radius overlap, radius 2 A
  - complexity for n atoms is O (n*n)
- Residue level contacts
  - collision spheres, CA is center
- Protein-ligand contacts
  - collision spheres, min max cente
- SSE level contacts
  - rules depending on SSE type, dif

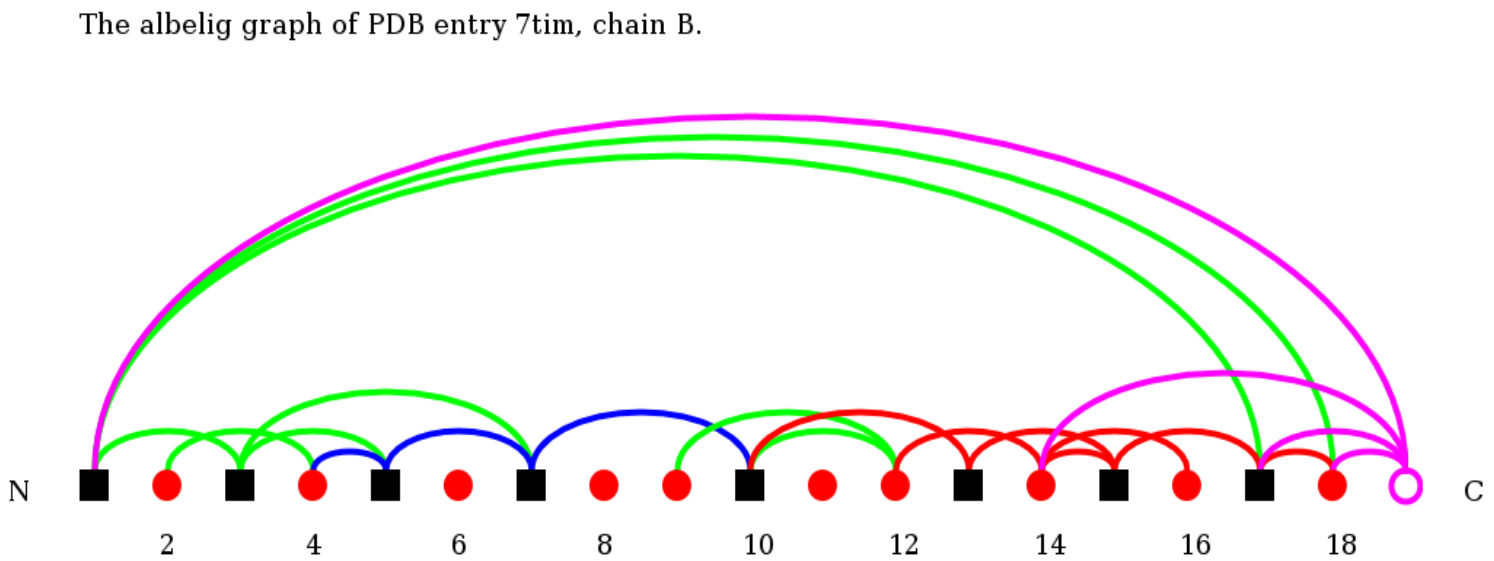


Types of protein graphs

- Alpha
- Beta
- Alpha-beta
- Ligands
- Not ness. connected
  - Connected components
  - Backbone option

The VPLG software

- Command line tool
- GUI in Java/Swing
- Screenshots
- Graph file formats
- Database support



Protein graphs for the whole PDB – Some statistics

- 62,364 proteins consisting of 151,025 chains
- 1.4 million helices, 1.2 million beta strands, 300k ligands
- Contacts
  - Mixed: 710.000
  - Parallel: 450.000
  - Antiparallel: 870.000
  - Total non-ligand: 2.000.000 => 0.76 contacts per SSE
  - Ligand: 760.000 => 2.50 contacts per ligand
- High number of isolated ligands (~30 %) at first run
  - Atom radius too small?
  - Contact definition?
  - Small ligands (few atoms?)

Outlook and possible improvements

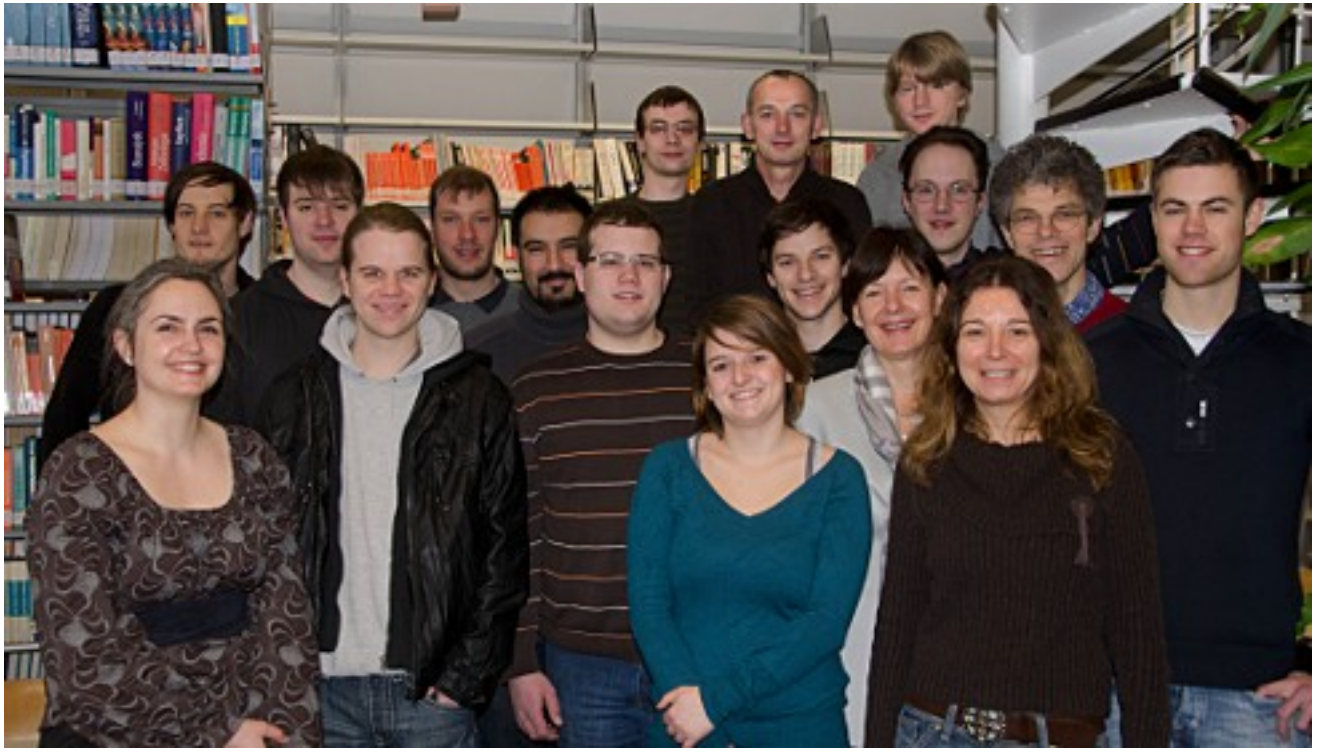
- VPLG software
  - Post-processing of DSSP output (short SSEs)
  - New spatial relation 'close to' for ligands
  - RNA / DNA as ligands
- Backend
  - Get Web server online
  - Protein structure comparison (not necessarily graph-based)
  - Use VPLG to get ligands into the PTGL

Summary

- Ligands are part of protein graph
- Visualization based on primary structure ordering of SSEs
- VPLG software is open source implementation
  - Visualization of protein graphs from PDB and DSSP files
  - http://www.bioinformatik.uni-frankfurt.de
    https://sourceforge.net/projects/vplg/
- Evaluation
  - Statistics
  - Hemoglobin as an example

Acknowledgments

- Supervisor Prof. Ina Koch
- Molecular Bininformatics group at Goethe University Frankfurt am Main (MolBI)



Thanks for your attention!

Questions?

(Appendix slides follow)