

# Protein Hypernetworks

Johannes Köster

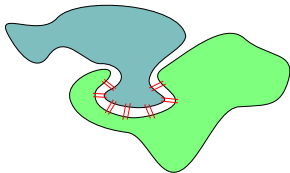
TU Dortmund, Informatik LS 11  
Max-Planck-Institute of Molekular Physiology Dortmund

4. 5. 2010

## Motivation

### Proteins

- building blocks of cells
- execution of cellular functions
- three-dimensional structure
- binding domains for other proteins
- form networks of interactions



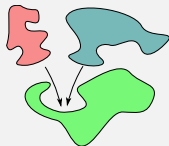
## Motivation

### Interaction dependencies

- allosteric effects



- competition on binding domain



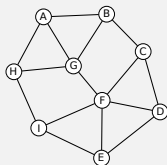
# Structure

- 1** Protein Hypernetworks
- 2** Prediction of Protein Complexes
- 3** Prediction of Functional Importance

## Idea

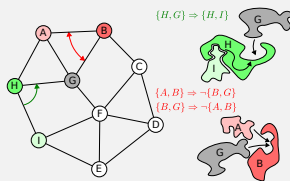
### Protein network $(P, I)$

- Set  $P$  of proteins as nodes
- Set  $I \subseteq \binom{P}{2}$  of interactions as edges
- ⚡ Interaction dependencies not considered



### Protein hypernetwork $(P, I, C)$

- Protein Network  $(P, I)$
- Set  $C$  of propositional logic constraints  $q \Rightarrow \psi$  with  $q \in P \cup I$



## Constraints

### Allosteric effects

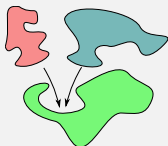
$$\{C, B\} \Rightarrow \{A, B\}$$



### Competition on binding domain

$$\{C, B\} \Rightarrow \neg\{A, B\}$$

$$\{A, B\} \Rightarrow \neg\{C, B\}$$



## Minimal network states

### Minimal network states (*Nec*, *Imp*) for $q \in P \cup I$

$$q \wedge \bigwedge_{c \in C} c$$

- Satisfying model  $\alpha : P \cup I \rightarrow \{0, 1\}$  by tableau algorithm
- Constraint  $q' \Rightarrow \psi$  active iff  $\alpha(q') = 1$
- For each constraint, the inactive case is expanded first
- Contains simultaneously necessary (*Nec*) and impossible (*Imp*) proteins and interactions

$$Nec := \{q' \in P \cup I \mid \alpha(q') = 1\}$$

$$Imp := \{q' \in P \cup I \mid \alpha(q') = 0 \text{ by active } c \in C\}$$

## Proof: Tableau needs only $\mathcal{O}(|C|)$ expansions

$$f = q \wedge \bigwedge_{c \in C} c$$

- Assumption: constraints  $c$  of the form  $q_1 \Rightarrow l$ ,  $l \in \{\neg q_2, q_2\}$  and  $f$  is satisfiable.
- Observation: Active constraint cannot become inactive again: Assume contradiction by  $l$ .  $l$  is backtracked and  $\neg q_1$  is expanded again. Now  $\neg q_1$  contradicts either  $q$  or another active constraint (apply argument recursively), so both branches are unsatisfiable  $\zeta$ .
- ▶ Each  $c$  is expanded at most 2 times:
  - Never activated: 1 expansion
  - Immediate activation: 2 expansions
  - Activation by backtracking: 2 expansions



## Minimal network states

### Clashes

Two minimal network states  $(Nec, Imp)$  and  $(Nec', Imp')$  are clashing iff

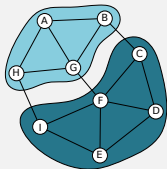
$$Nec \cap Imp' \neq \emptyset \text{ or } Nec' \cap Imp \neq \emptyset.$$

If a not clashing pair of minimal network states of two proteins or interactions exists, then the proteins or interactions are simultaneously possible.

## Prediction of Protein Complexes

### Network based

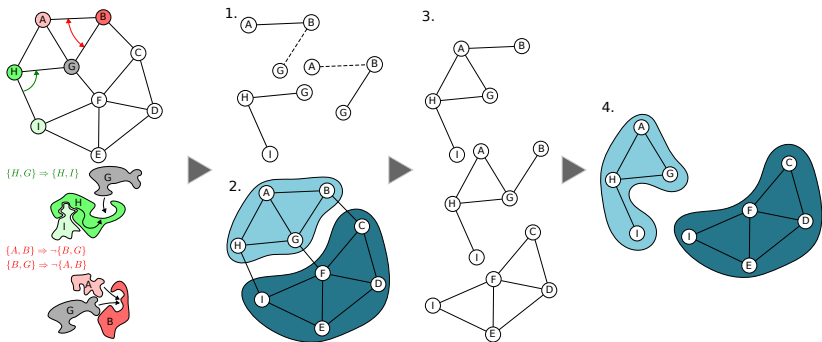
- Find dense regions in graph (e.g. clustering)
- ⚡ May violate interaction dependencies



### Hypernetwork based

- Network based complex prediction
- For each complex: calculate simultaneous subnetworks
- Perform network based complex prediction on the subnetworks
- Add all necessary interactions to complexes

## Prediction of Protein Complexes



## Results on the Yeast Protein Network

	precision	recall
plain (no constraints)	0.142	0.792
458 constraints	0.206	0.792
458 rand. constraints, mean (SD)	0.149 (0.005)	0.782 (0.02)

$$\text{recall: } \frac{B-FN}{B}, \text{ precision: } \frac{P-FP}{P}$$

- Network: CYGD (4579 proteins, 12576 interactions)
- Constraints: Competition on binding sites (Jung et al. 2010)
- Complexes: CYGD (55 connected complexes)

## Prediction of Functional Importance

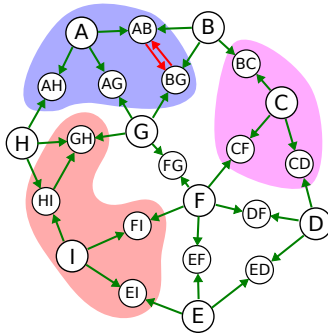
### Network based

- Plain node degree (Jeong et al. 2001)
- ⚡ Interaction dependencies?

### Hypernetwork based

- Minimal network state graph  $G_{MNS} = (P \cup I, E)$   
 $(q', q) \in E$  for  $q \in P \cup I$  and  $q' \in Nec_q \cup Imp_q$
- BFS from each node
- Perturbation Impact Score

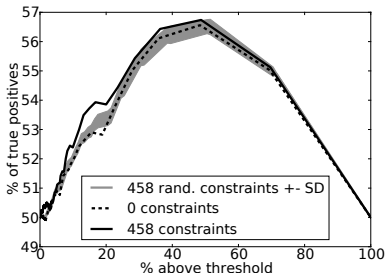
## Prediction of Functional Importance



### Perturbation Impact Score

$$PIS_{(P,I,C)}(Q_{\downarrow}) := \sum_{q \in \text{reach}_{Q_{\downarrow}}^{\text{BFS}}} \text{dist}_{Q_{\downarrow}}^{\text{BFS}}(q)$$

## Results



TP: lethal/sick and  $PIS \geq t$ , viable and  $PIS < t$

- Network: CYGD (4579 proteins, 12576 interactions)
- Constraints: Competition on binding sites (Jung et al. 2010)
- Perturbations classified as lethal/sick and viable (SGD)

## Conclusion

- Hypernetworks as an extension of graph based network models
- Propositional logic constraints
- Minimal network states by tableau algorithm
- Improvements in complex prediction quality
- Improvements in functional importance prediction quality