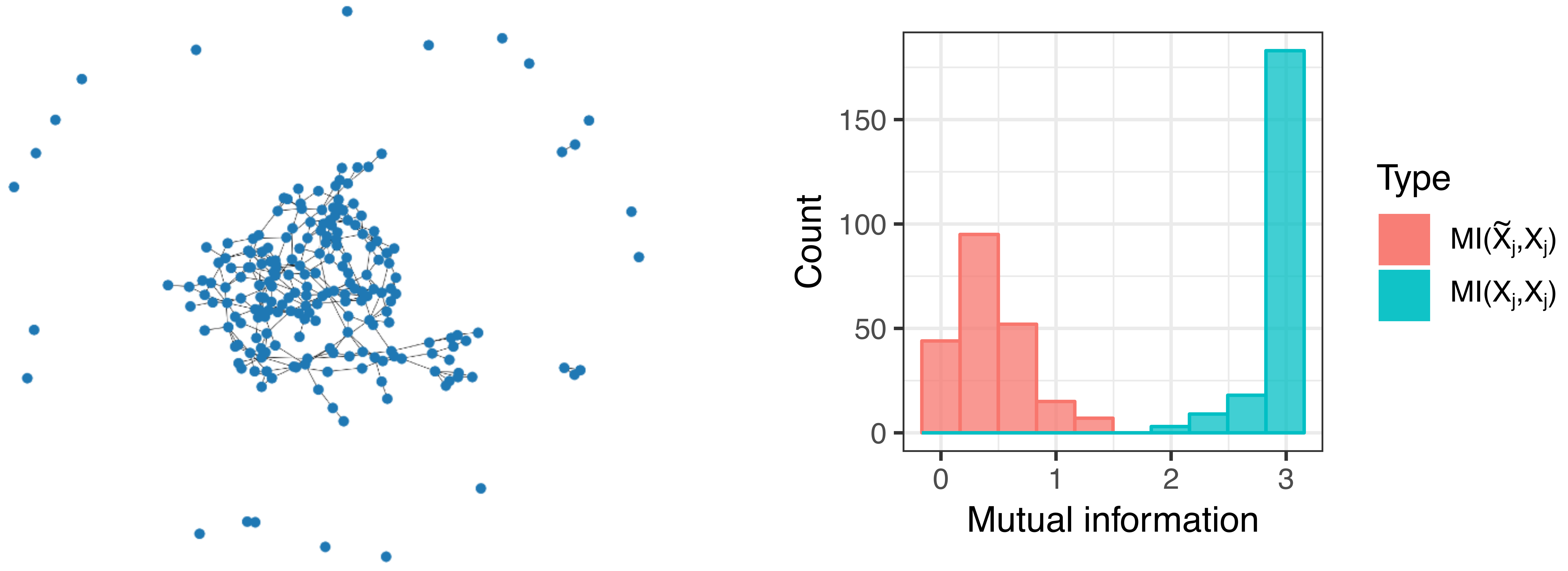

Count
150
100
50
0
0
1
2
3
Mutual information
Type
MI($\tilde{X}_j$,X_j)
MI(X_j,X_j)

$$\mathbb{P}_{h,J}(X) = \frac{1}{Z(h,J)} \exp\left(\sum_{j=1}^{p} h_i(X_i) + \sum_{1 \leq i < j \leq p} J_{ij}(X_i, X_j) \right)$$

Simulation:
Potts model for protein contact prediction

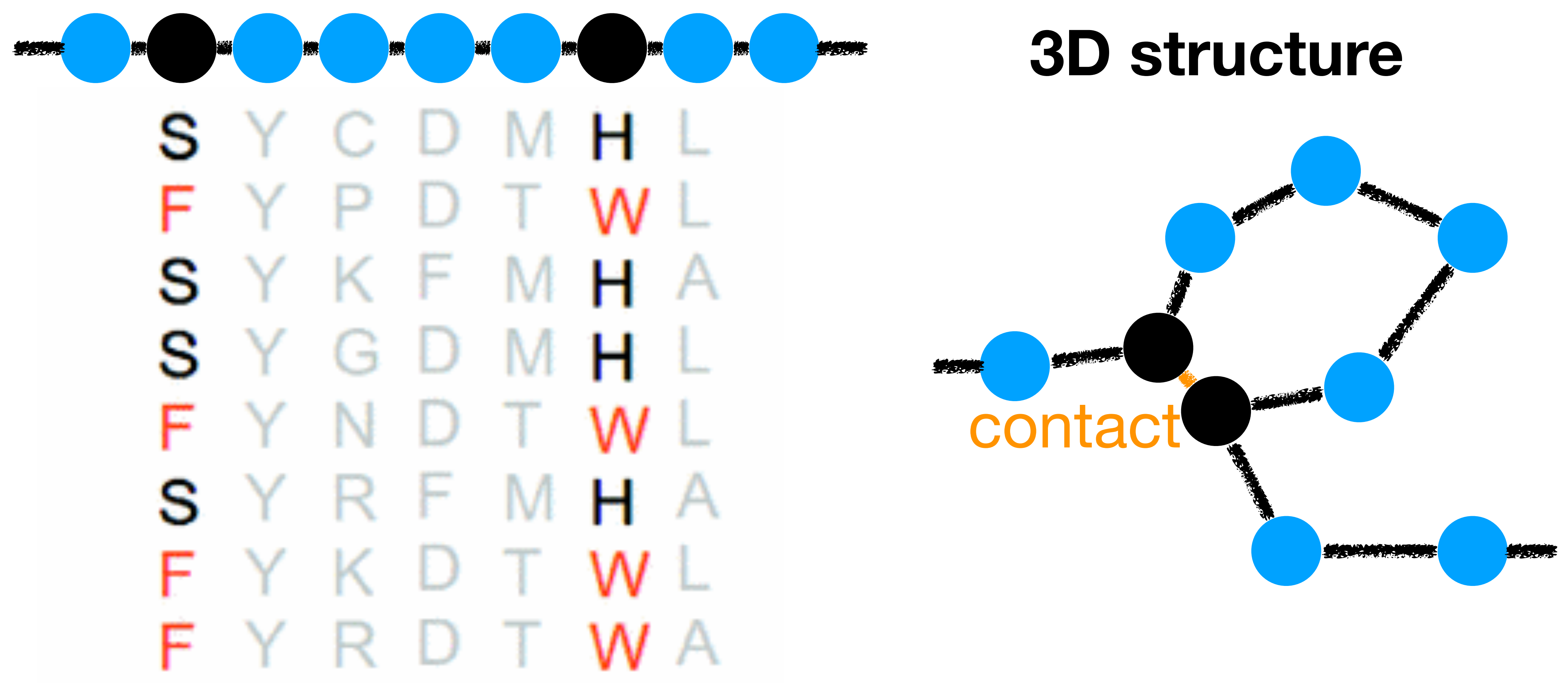
Ekeberg, Löfkvist, Lan, Weigt, and Aurell (2013)

$$\|J_{ij}\| \neq 0$$

protein family PF00006, $p = 213$

$$X \in \{0,1,\ldots,20\}^p$$

(20 amino acids, one gap)

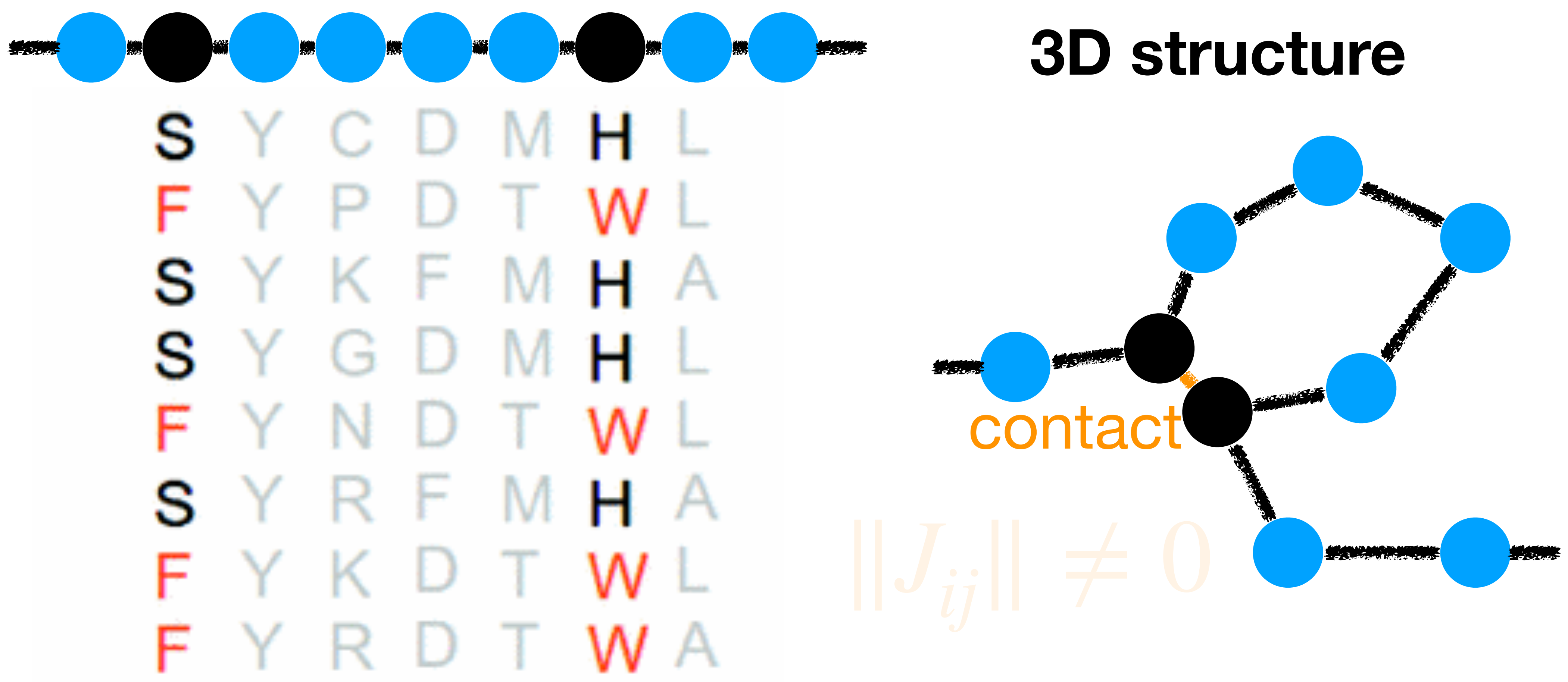


3D structure
S Y C D M H L
F Y P D T W L
S Y K F M H A
S Y G D M H L
F Y N D T W L
S Y R F M H A
F Y K D T W L
F Y R D T W A
contact

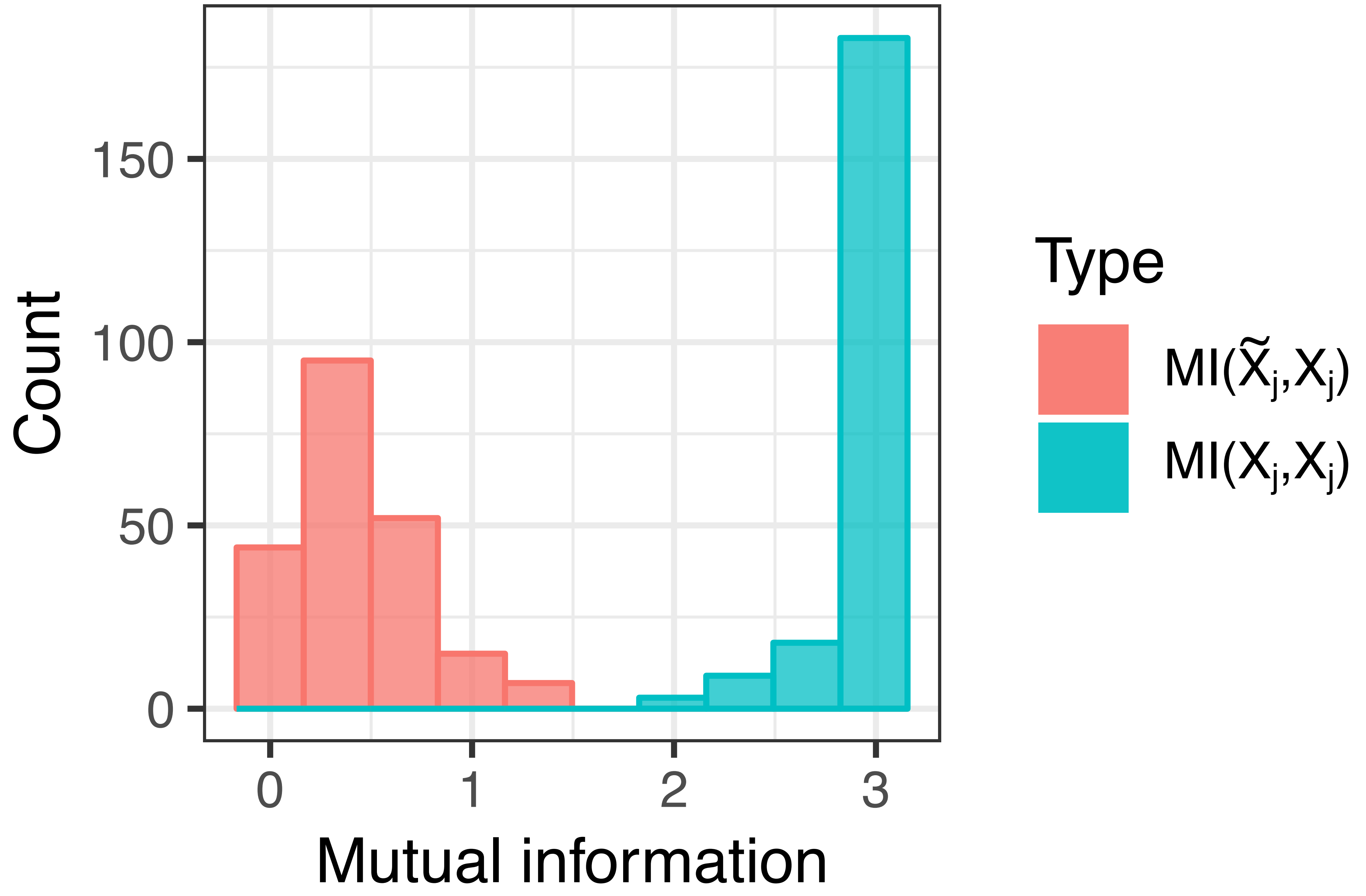
Simulation:
Potts model for protein contact prediction

$$\mathbb{P}_{h,J}(X) = \frac{1}{Z(h,J)} \exp\left(\sum_{j=1}^{p} h_i(X_i) + \sum_{1 \le i < j \le p} J_{ij}(X_i, X_j) \right)$$

$X \in \{0,1,\ldots,20\}^p$ (20 amino acids, one gap)



protein family PF00006, $p = 213$



Ekeberg, Löfkvist, Lan, Weigt, and Aurell (2013)

1. Introduction
    - Variable selection and model-X knockoffs
    - Knockoff sampling is difficult
2. Characterizing knockoff distributions
    - The characterization theorem
    - Connection to Markov chain Monte Carlo (MCMC)
3. Metropolized knockoff sampling (Metro)
    - How it works
    - Time complexity and graphical structure
4. Good proposals inspired by the MCMC literature
    - Covariance-guided proposal
    - Multiple-try Metropolis (MTM)
5. Simulation results
6. **Discussion**