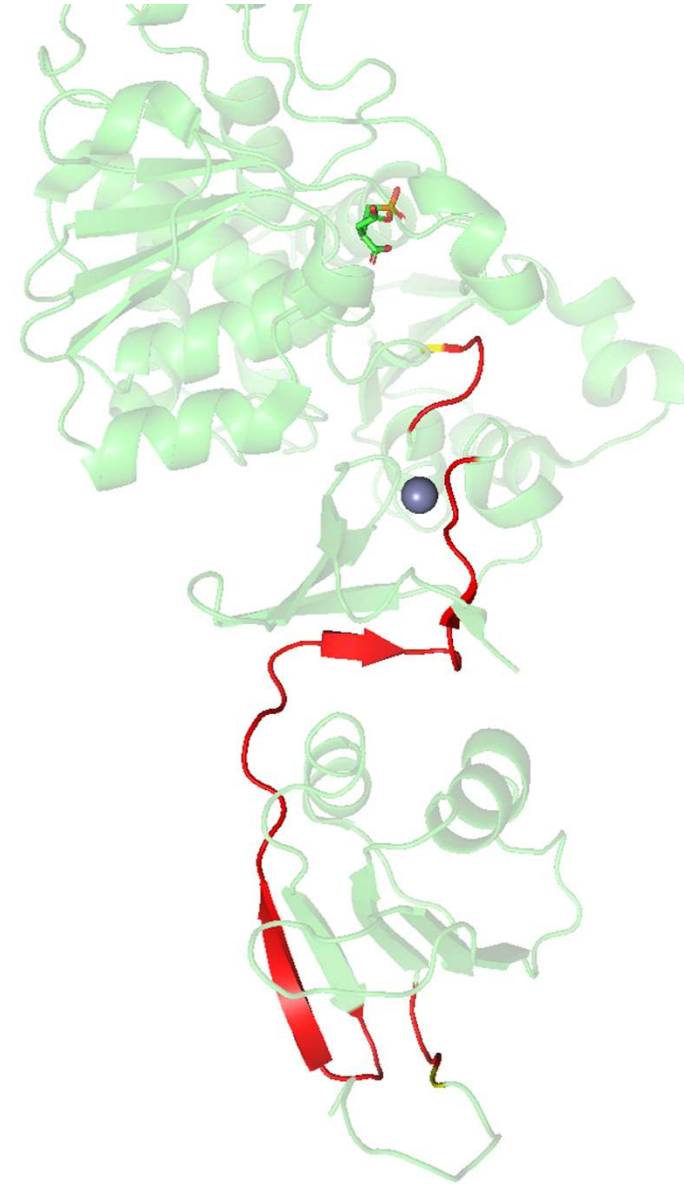# IMPERIAL

Uncovering Allosteric Pathways:
A thermodynamic framework and computational method

Ralph Holden
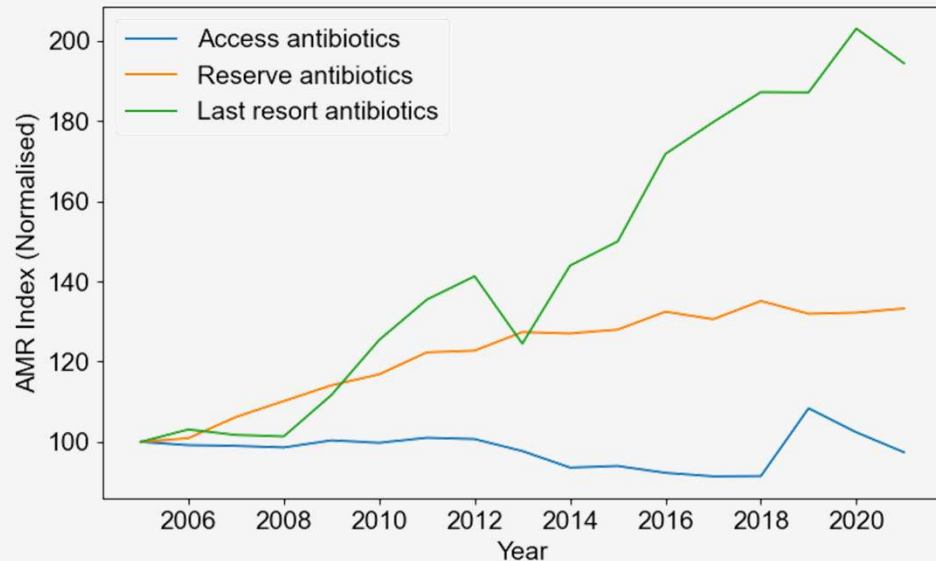PI: Prof. Sophia Yaliraki
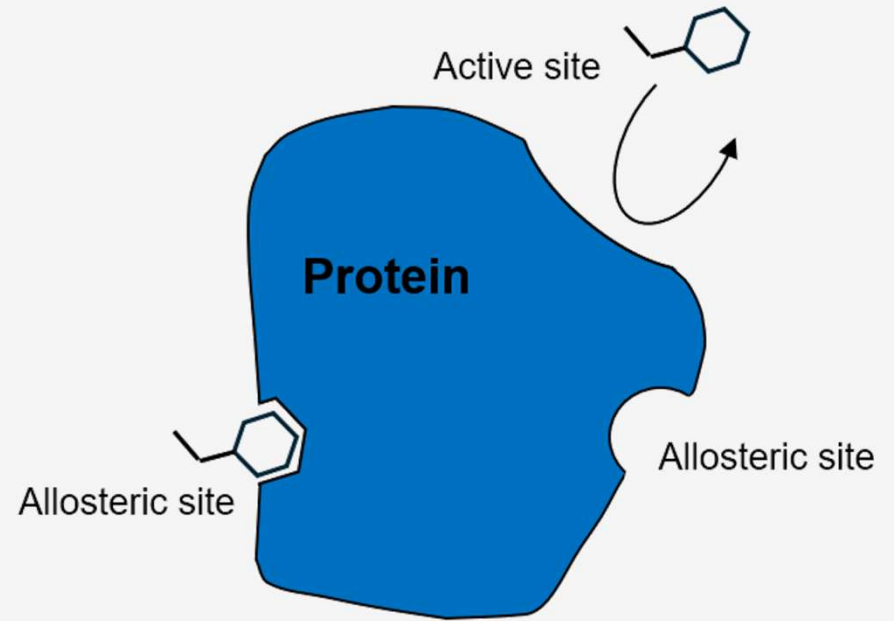Co-PI: Prof. Mauricio Barahona
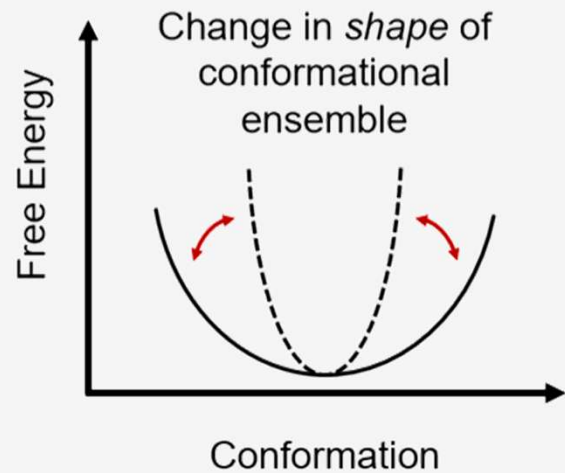02/06/2025

Motivation
Off Site Drug Targets

- Antimicrobial resistance poses major global health challenge

- Insufficient innovative antimicrobials to meet demand

- Allosteric sites influence active site

- New possibilities for off-site drug targets

- Drug the "undruggable" proteins?

Data taken from: M. Cecchini and E. Özçelik, *Embracing a One Health Framework to Fight Antimicrobial Resistance*, tech. rep., OECD Health Policy Studies, 2023

Scientific Background
Dynamic Allostery

Change in *shape* of conformational ensemble

Free Energy

Conformation

- Two extreme models of allosteric communication

Domino Model

Active site

Allosteric site

Localised path
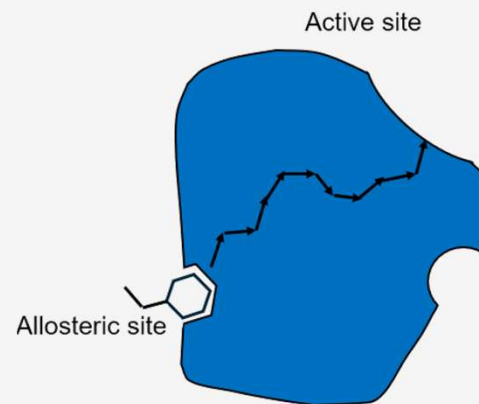
Violin Model

Active site

Allosteric site

Global rearrangement

- No change in the average structure

- Significant entropic contribution to allostery

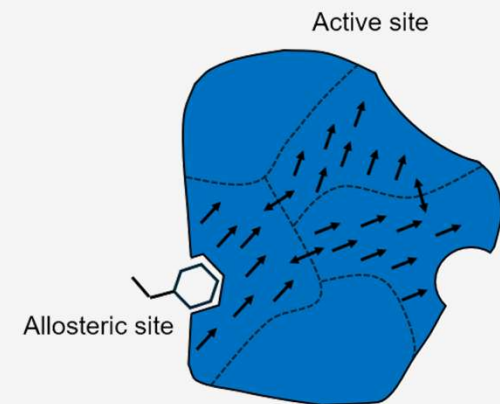- Subtle changes in the frequency and amplitude of thermal fluctuations

H. N. Motlagh, J. O. Wrabl, J. Li and V. J. Hilser, *Nature*, 2014, **508**, 331–339

Scientific Background
Graph Methods

Figure taken from: F. J. Song, Ph.D. Thesis, Imperial College London, 2021
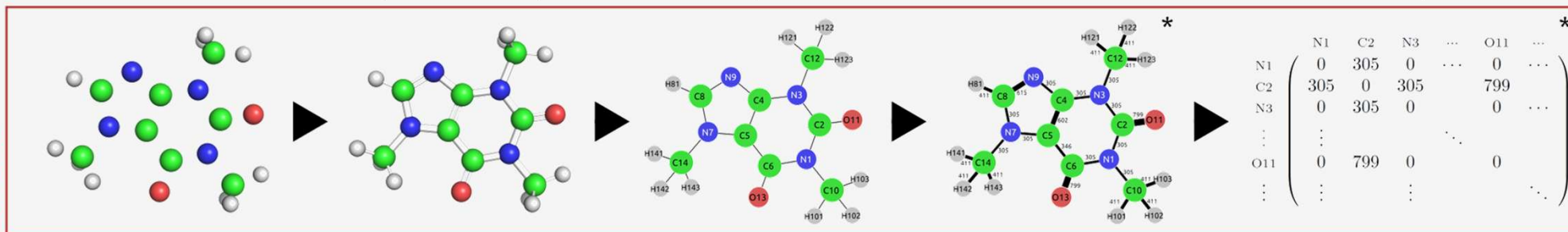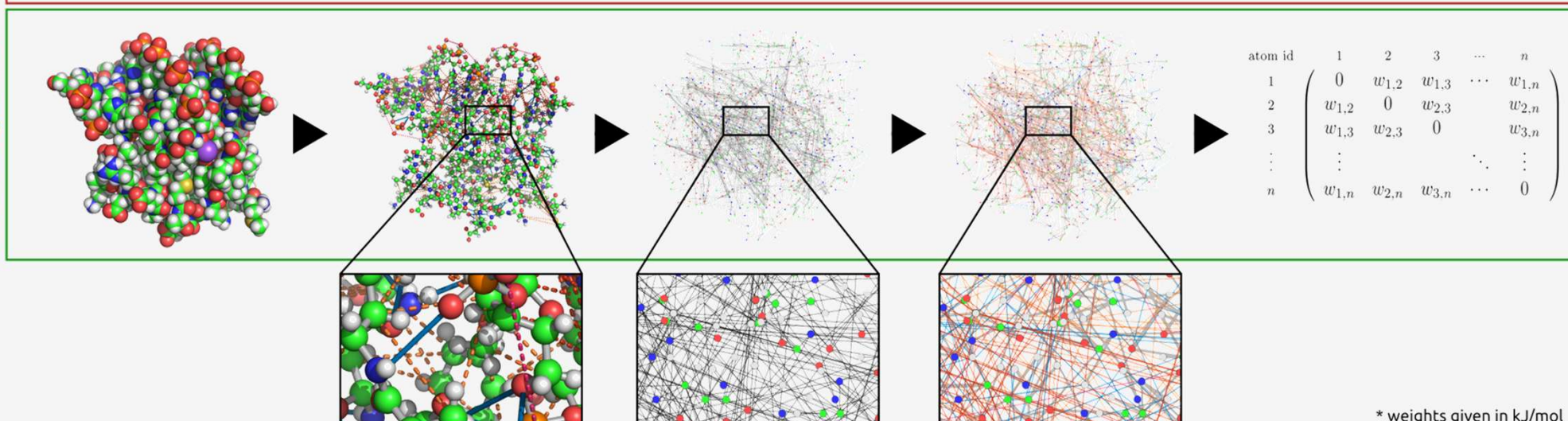
Novel methods

More principled approaches to allosteric pathways

Theoretical Framework
## Thermodynamic Perspective

- Allosteric pathway with optimum "free energy" balances two effects – energy & entropy

1) Minimum entropic cost:
   path with minimum no. steps
   ⟶ Drives path through shortcuts otherwise energetically insignificant weak interactions

2) Best energetic link:
   strongest energy weighted path
   ⟶ Imposes energetic cost: weak steps must be carefully chosen & highly specialised

1) Optimum Entropy Path   **2) Optimum Energy Path**

- Node
- Source/ target node
- Backbone interaction
- Weak interaction
- Pathway
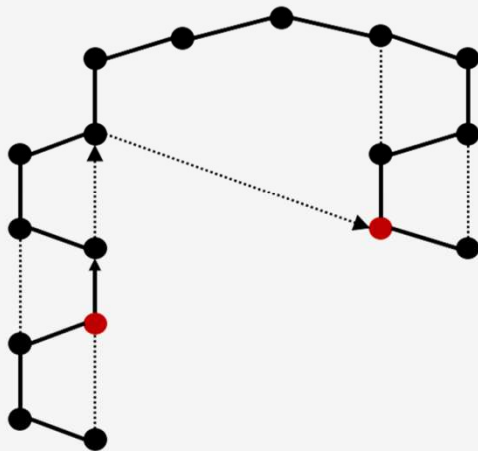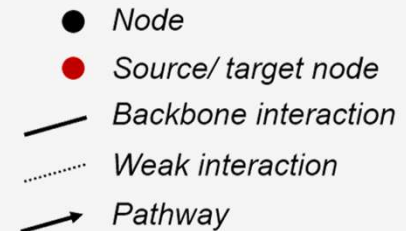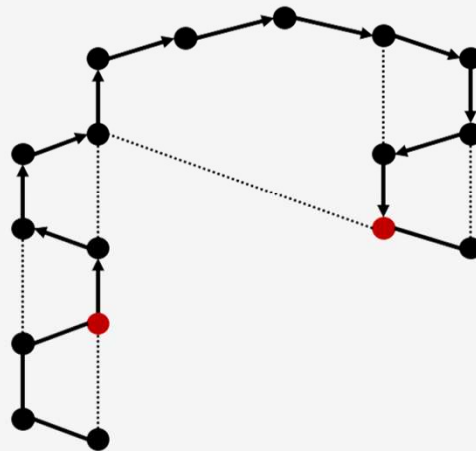
Method Design for Path Search
Maximum Flow

- Maximum flow algorithms optimise for greatest flow rate through graph edges

- Two constraints:
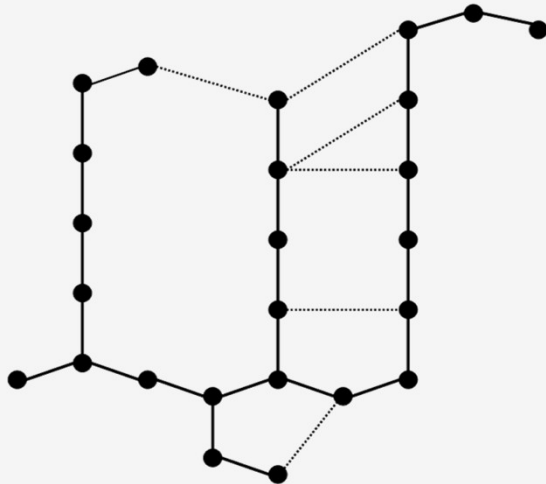 1) Flow conserved into & out of nodes
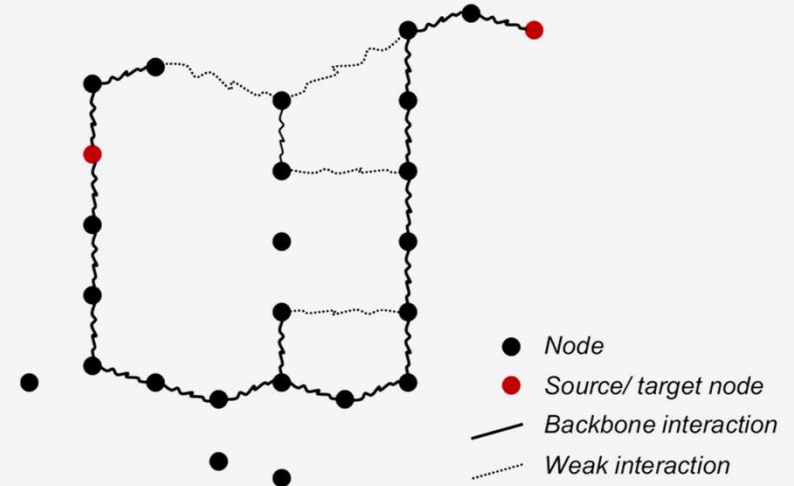 2) Flow rate through edge is limited by "capacity" $\longrightarrow$ Applied to BagPype graph with heuristic: BagPype edge energy weights are proportional to flow capacity

BagPype Graph, U

Maximum Flow Graph, Φ
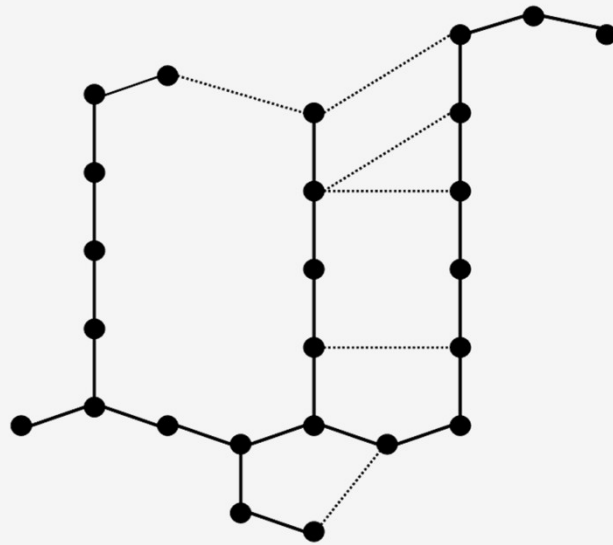
- Node
- Source/ target node
— Backbone interaction
······· Weak interaction

Method Design for Path Search
Maximum Flow

BagPype Graph, U

Maximum Flow Graph, Φ

- ● Node
- ● Source/ target node
- —— Backbone interaction
- ⋯⋯ Weak interaction

Observation:

1) Graph complexity reduced $\longrightarrow$ Importance sampling for robust flows

2) Edge max flow rate less than BagPype energy (capacity) $\longrightarrow$
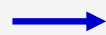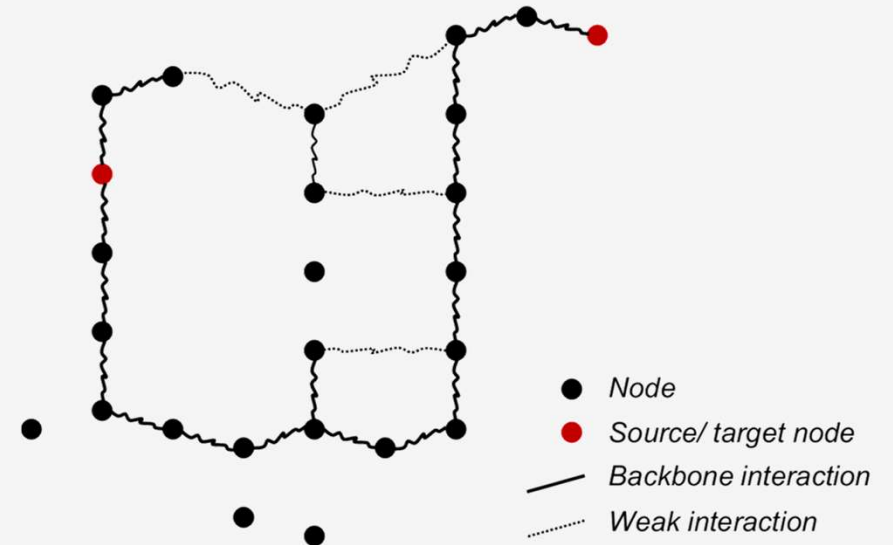
Interpretation:

Importance sampling for robust flows

Effective ability of edge to transfer energy, dependant on surrounding environment

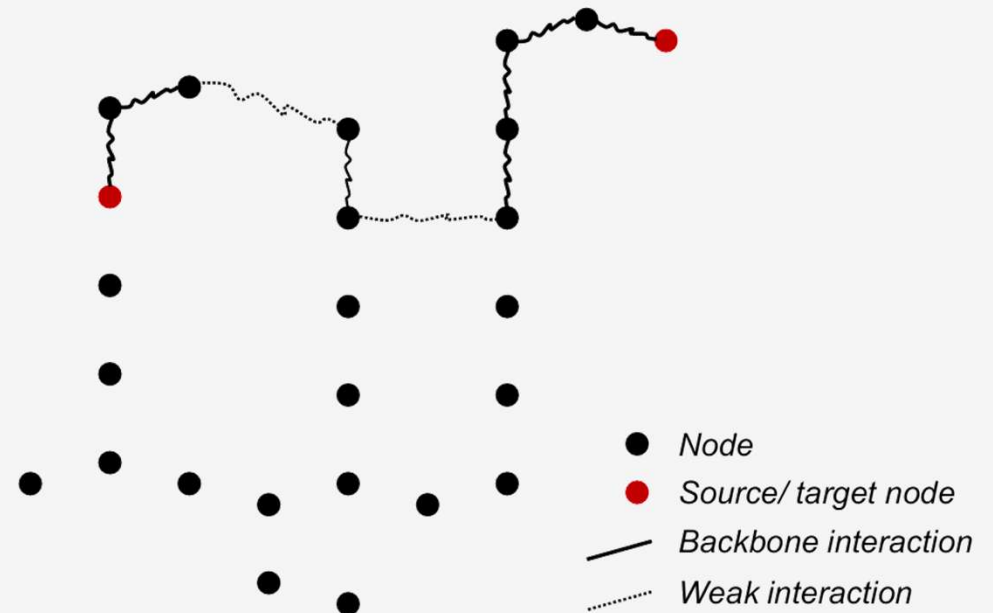# Method for Path Search
Flow Path

- Path must optimise for both:
  1) Strongest interactions (energy)
  2) Minimum steps (entropy)

- Flow, ϕ must be converted to alternative form, ϕ' for balanced optimisation by Dijkstra's algorithm

$$\phi' = {}^{1}\!/_{\phi}$$

  1) Favours backbone step-by-step
  2) "Cost" buildup over many steps – prefers shortcuts

Optimum Free Energy Path

- ● Node
- ● Source/ target node
- —— Backbone interaction
- ········ Weak interaction

Method for Path Score
Path Rank

- Ruelle-Bowen random walk stationary distribution optimises the sum of:

  1) Shannon entropy

  2) Expected energy

- Construct transition matrix, B
  - Energetic transition probabilities between nodes

- Spectral decomposition for dominant eigenvectors

$$Bv = \lambda v$$

$$u^{\mathsf{T}}B = u^{\mathsf{T}}\lambda$$

$u_i v_i$	Random walk stationary distribution (probability at node i)
λ	Spectral radius, related to topological entropy

BagPype Graph, U
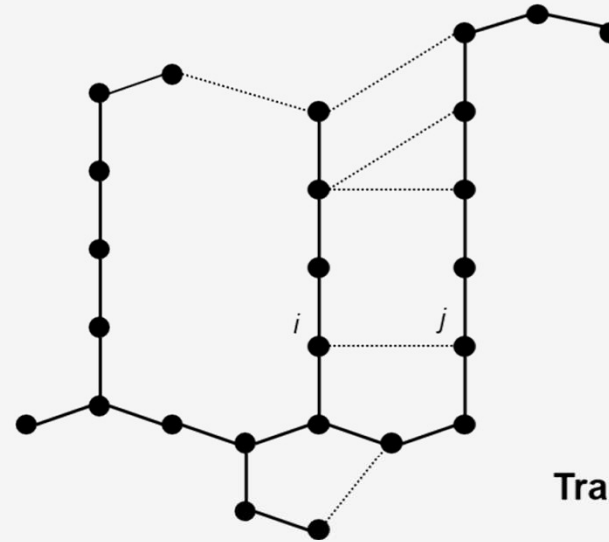
Transition Matrix, B
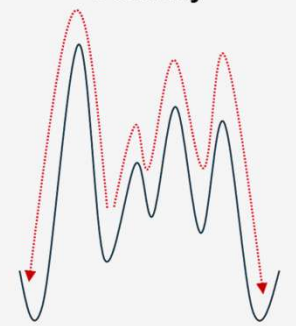
$$B_{ij} = \exp\left(-1/\beta U_{ij}\right)$$

$$B = \begin{bmatrix} \cdots & & \cdots & \exp\left(-1/\beta U_{ij}\right) \\ & \vdots & \ddots & \vdots \\ \exp\left(-1/\beta U_{ji}\right) & \cdots & & \cdots \end{bmatrix}$$

Method for Path Score
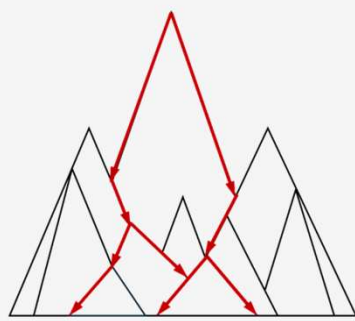## Path Rank

Interpretation

- Transition probability takes form akin to Boltzmann factor

- Crucial difference: exponent is the inverse

a) Energy Landscape Pathway

Overcoming potential barriers (Arrhenius eq.)

b) Graph Perturbation Propagation

Distribution of energy into "valleys"

— Potential landscape
⋯▸ Reaction path
— Graph edge
➡ Flow Path

Random Walker Path Probability

- Balance of energy and entropy

$$p_{k \to l} = \lambda^{-t} \left(\prod B_{k \to l} \right) u_k v_l$$

- Used to score Flow Paths

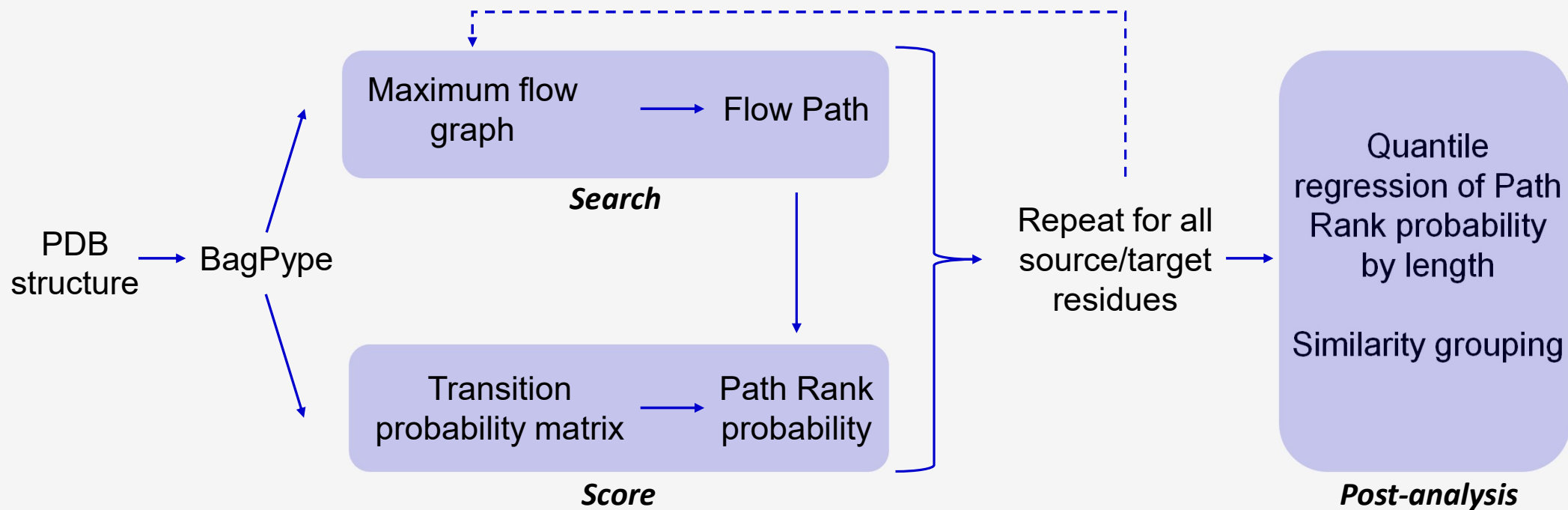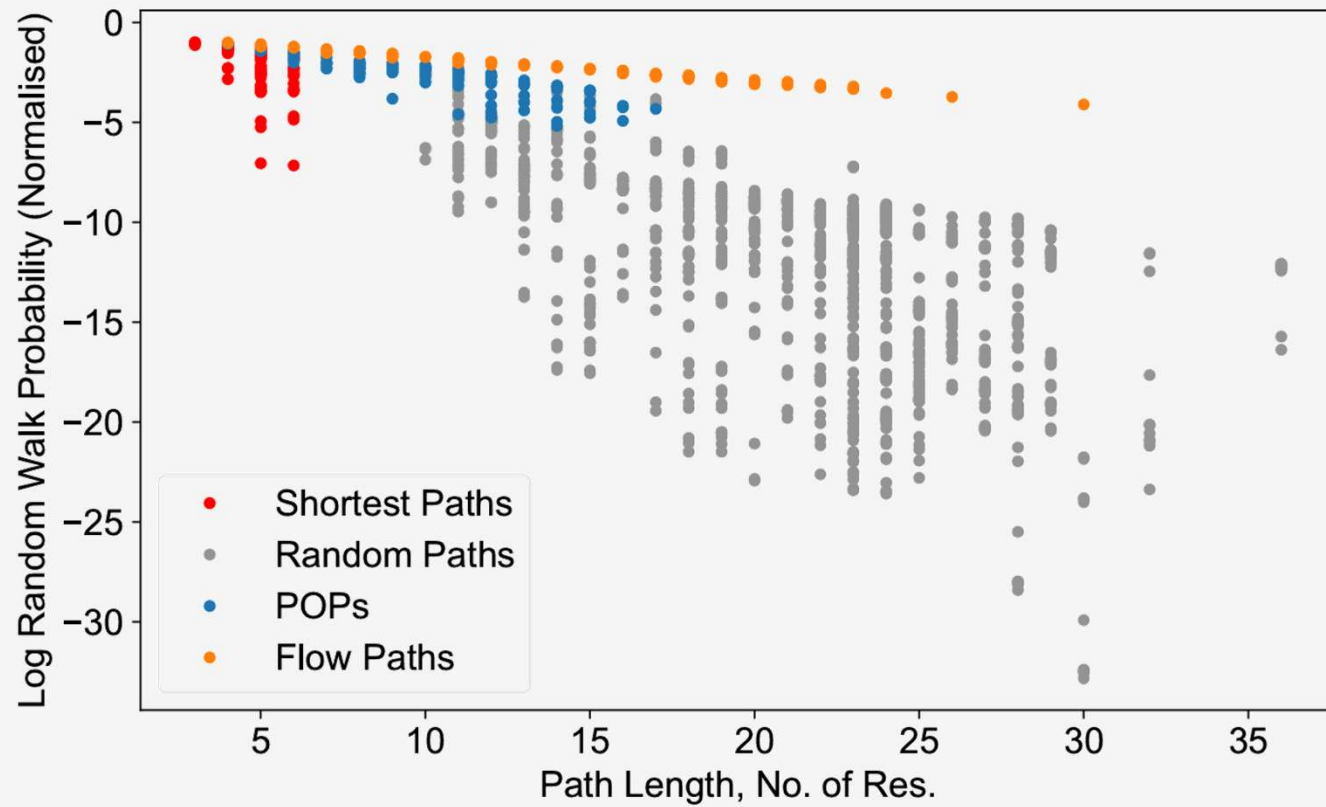|	
λ^{-t}	Entropic factor scaled by path length, global
$\prod B_{k \to l}$	Product of energetic transition probabilities, local
$u_k v_l$	Boltzmann stationary distribution of Ruelle-Bowen random walk

Workflow
Integrating Search and Score

Search

Maximum flow graph → Flow Path

Score

Transition probability matrix → Path Rank probability

PDB structure → BagPype

Repeat for all source/target residues

Post-analysis

Quantile regression of Path Rank probability by length

Similarity grouping

Validation
Benchmarking Flow Path

Validation
Effect of Temperature Scaling Factor

Low Temperature Limit

High Temperature Limit
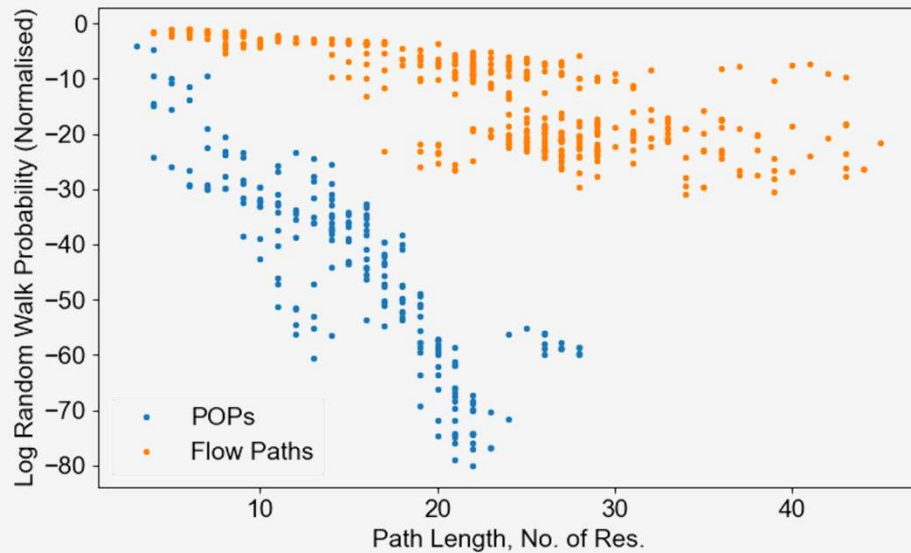
- Energy dominating, little entropy penalty from steps ⟷ Entropy dominating, recovering Ruelle-Bowen random walk on unweighted graph

- λ much reduced ⟷ λ increases

- Transition probability smaller for weak interactions ⟷ Transition probability even between interactions

Validation
Experimental Alanine Mutation Scanning
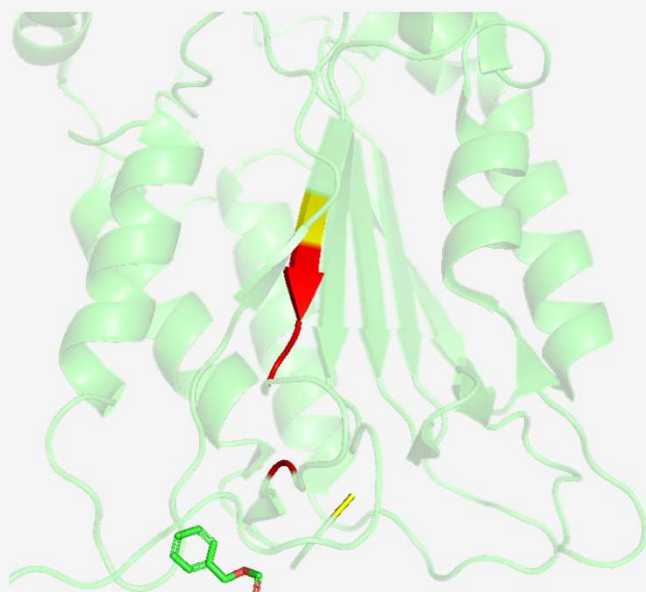
Flow path on caspase-1
(PDB: 2HBQ)

- Experimental alanine mutation "turns off" weak interactions in signalling pathways

→

- Weak interactions scored by sum of all Flow Paths weighted by Path Rank

Residue	Computational Score
SER332 [†]	9.0
ARG286 [†]	6.8
GLU390 [†]	6.3
ARG179	5.0
ALA284	4.6
SER339 [†]	4.5

[†] Highly functionally important

Uncovering Allosteric Pathways
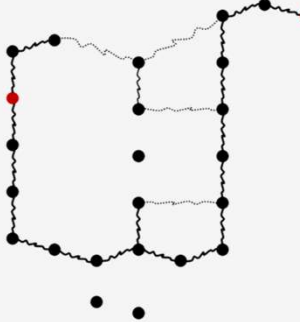## Summary

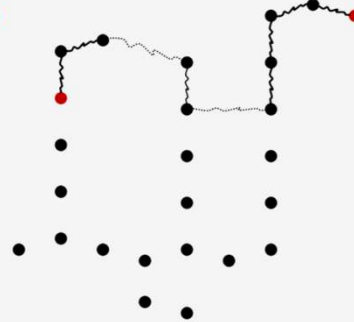| Theory	Computation	Application
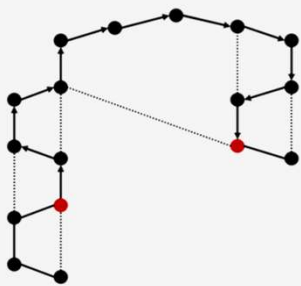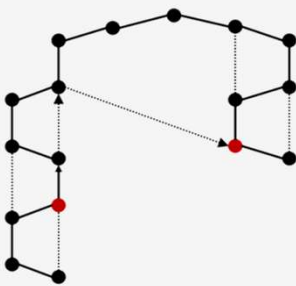
- Thermodynamic perspective of allostery

- Balance of energetic and entropic effects

- Weak interactions important via entropic effect

- Computational methods for thermodynamic path

  1) Search: "Flow Path"
  2) Score: "Path Rank"
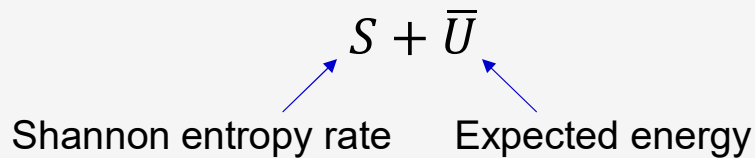
- No *a priori* choice of dominant factor

- Identify allosteric pathways

- Identify functional allosteric residues

Method Design for Path Score
Graph-thermodynamic formalism

- Ruelle-Bowen random walk has optimum free energy rate

$$S + \overline{U}$$

Shannon entropy rate   Expected energy

- Expected energy is ensemble average
  - In Path Rank formulation, contains temperature factor

- Node stationary probability distribution that maximises free energy rate: $\boldsymbol{vu}^{\mathsf{T}}$

  ! Takes form of Boltzmann distribution

- Path partition function

$$Q = \sum_{\substack{paths\ of \\ length\ t}} \exp\left(\sum U_{i \to j}\right)$$

- Total path free energy

$$F = \ln \sum_{\substack{paths\ of \\ length\ t}} \exp\left(\sum U_{i \to j}\right)$$

- Analogous to free energy of the canonical ensemble in thermodynamics

- In Path Rank, collapses to microcanonical ensemble equivalent at very large temperatures
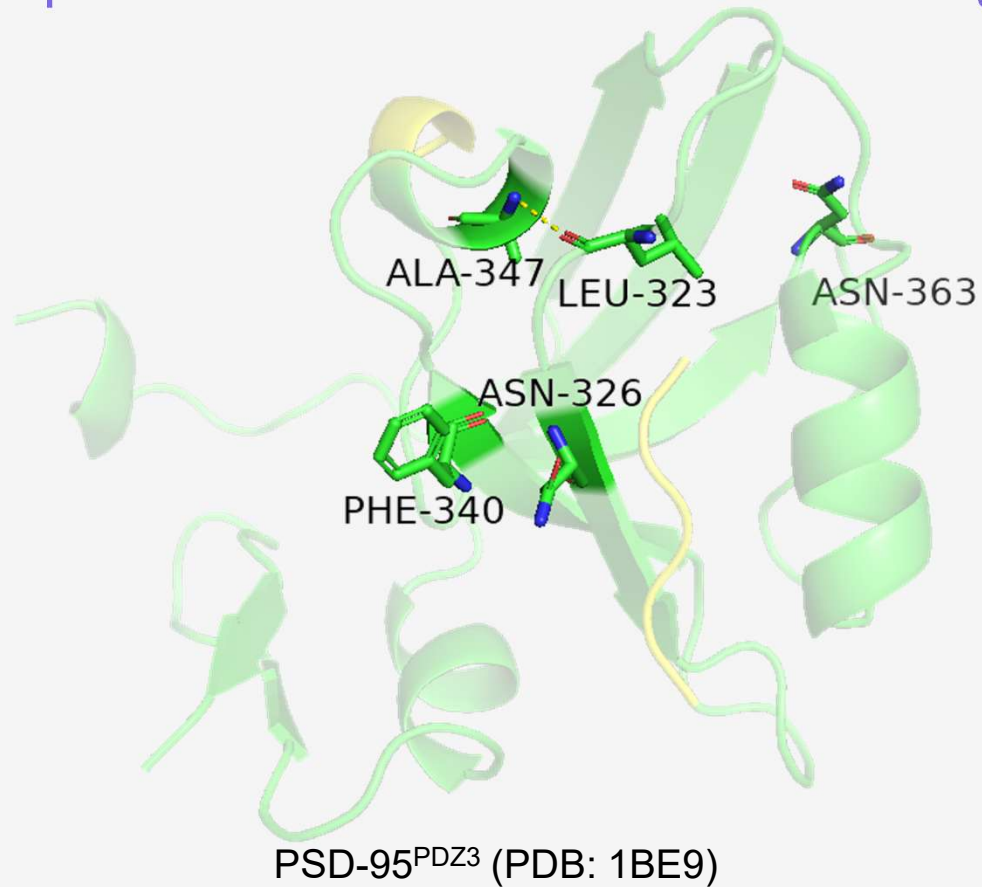
Validation
Experimental Alanine Mutation Scanning
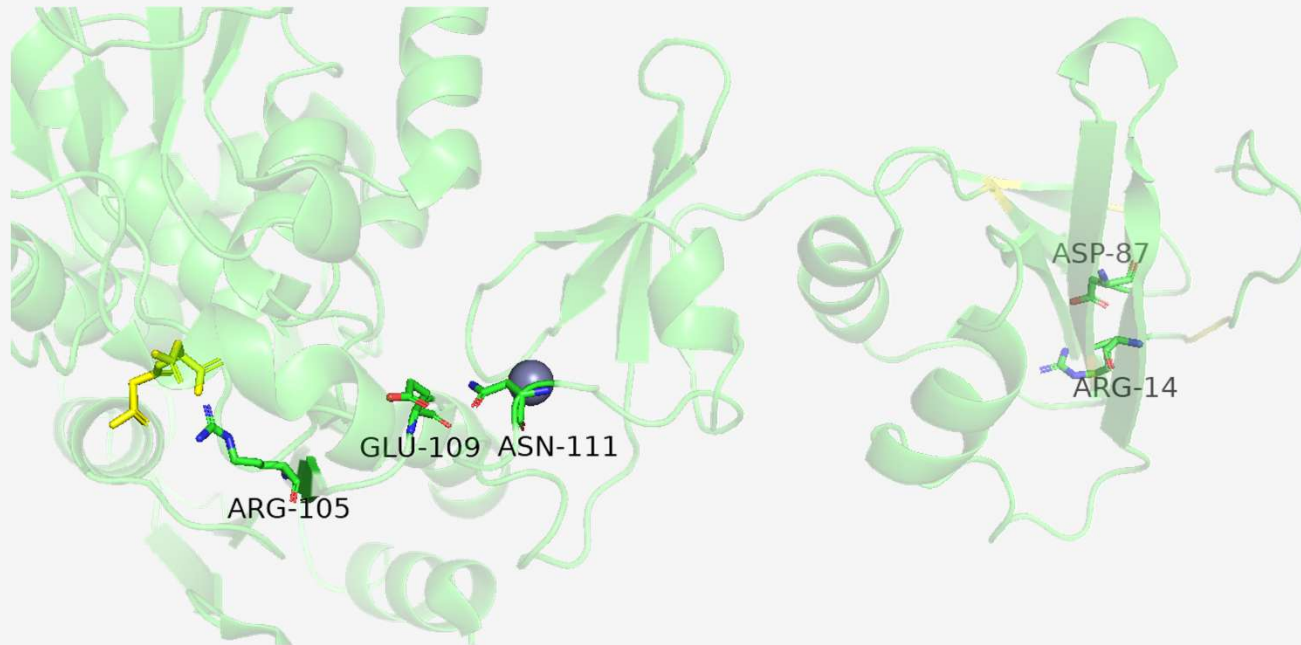
ATCase (PDB: 1D09)

Residue	Computational Score
GLU109 [†]	47.7
ASN111 [†]	46.9
ARG105 [†]	40.5
ASP87	21.2
ARG14	21.2

[†] Highly functionally important

E. Eisenstein, D. W. Markby and H. K. Schachman, *PNAS*, 1989, **86**, DOI: 10.1073/pnas.86.9.3094

Validation
Experimental Alanine Mutation Scanning

PSD-95^{PDZ3} (PDB: 1BE9)

Residue	Computational Score
ASN326 [†]	5.3
PHE340 [†]	5.3
LEU323 [†]	2.6
ALA347 [†]	2.6
ASN363	2.2

[†] Highly functionally important

Validation
Experimental Alanine Mutation Scanning
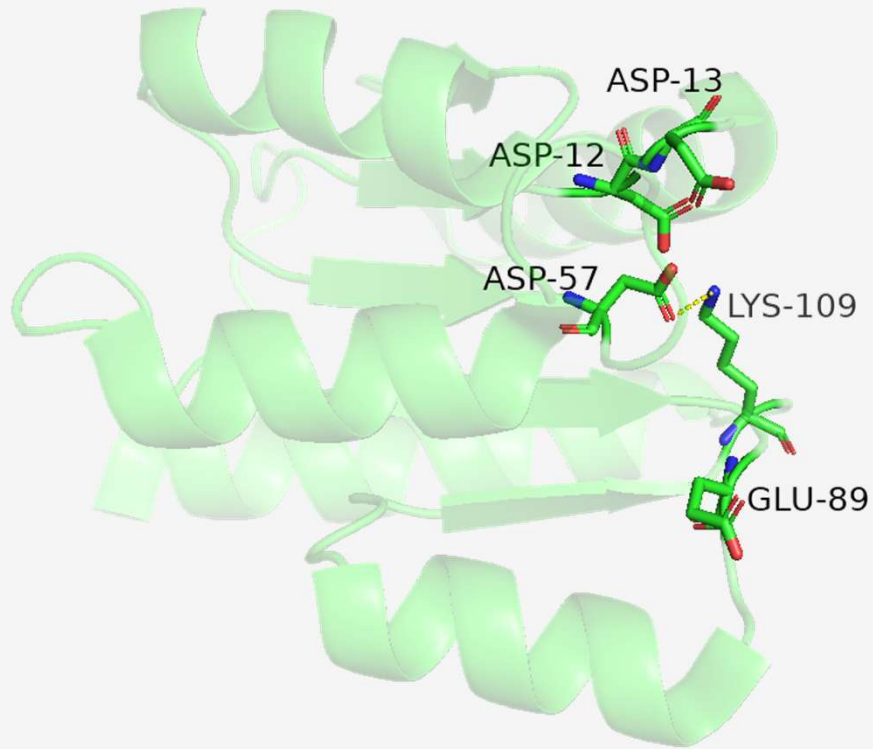
CheY (PDB: 1F4V)

Residue	Computational Score
LYS109 [†]	7.6
ASP57 [†]	5.1
ASP12 [†]	4.0
ASP13 [†]	4.0
GLU89	2.9

[†] Highly functionally important

G. S. Lukat, B. H. Lee, J. M. Mottonen, A. M. Stock and J. B. Stock, *JBC*, 1991, **266**, DOI: 10.1016/s0021-9258(18)92982-0

A. M. Stock, E. Martinez-Hackert, A. H. West, B. F. Rasmussen, D. Ringe, G. A. Petsko and J. B. Stock, *Biochemistry*, 1993, **32**, DOI: 10.1021/bi00212a001

Validation
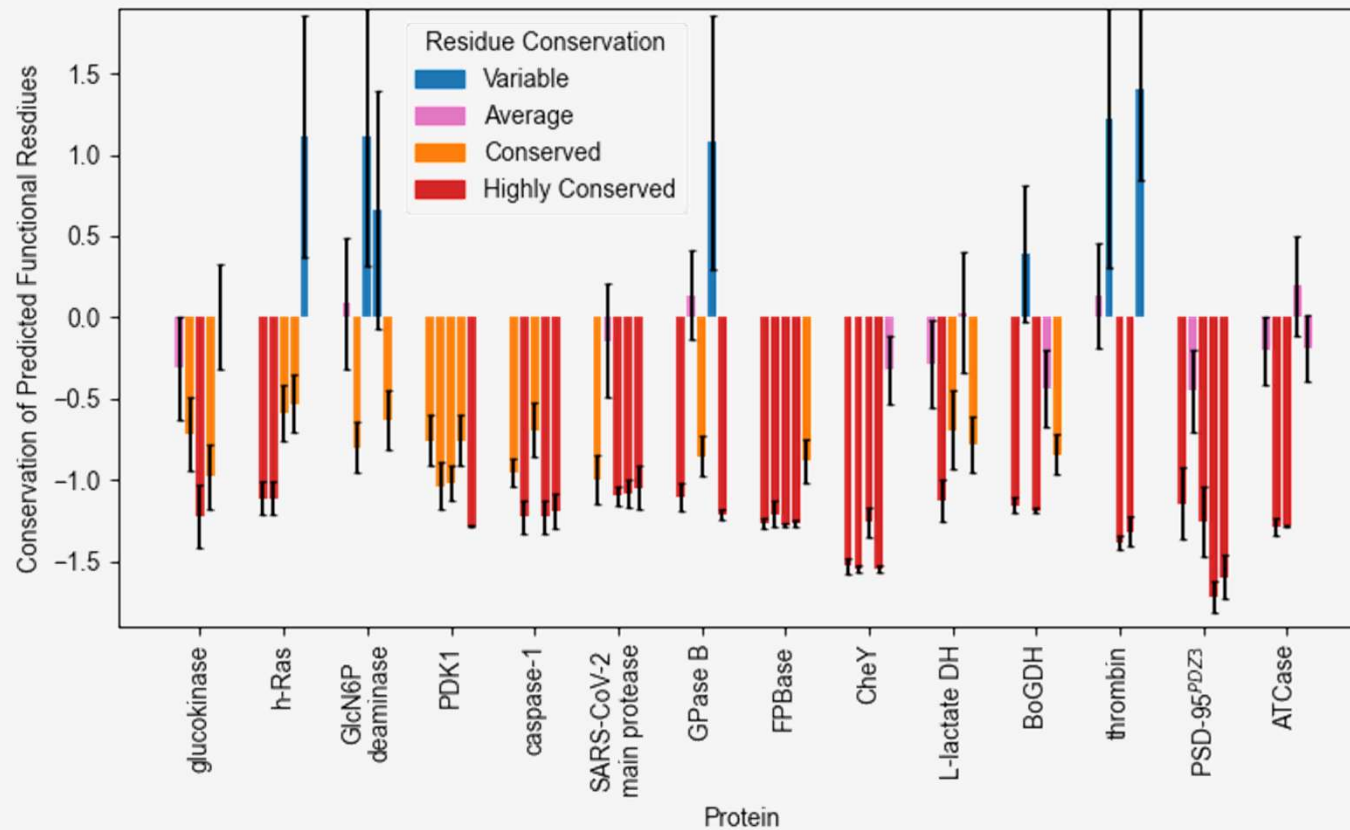## Evolutionary Conservation

- Functional residues conserved due to evolutionary pressure

- Conservation score calculated across amino acid sequences of variants and families

- Weak signalling residues in predicted allosteric pathway highly conserved