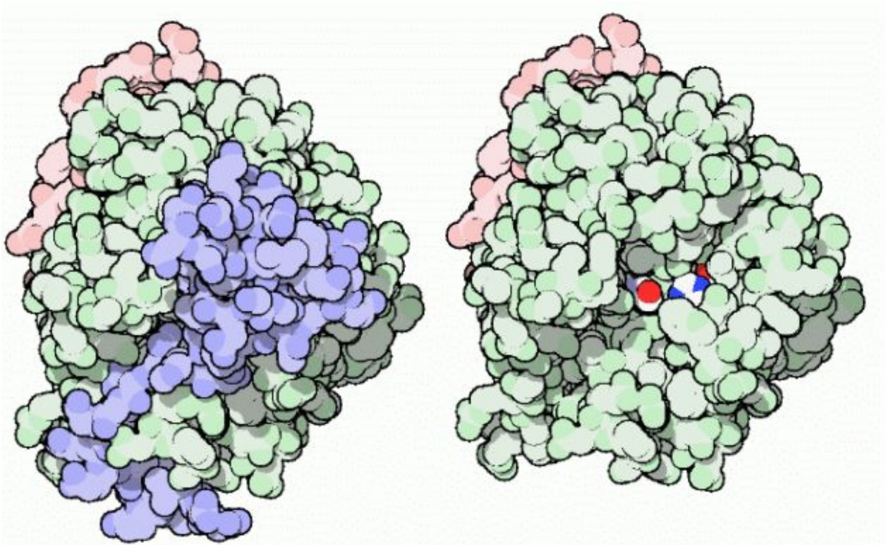
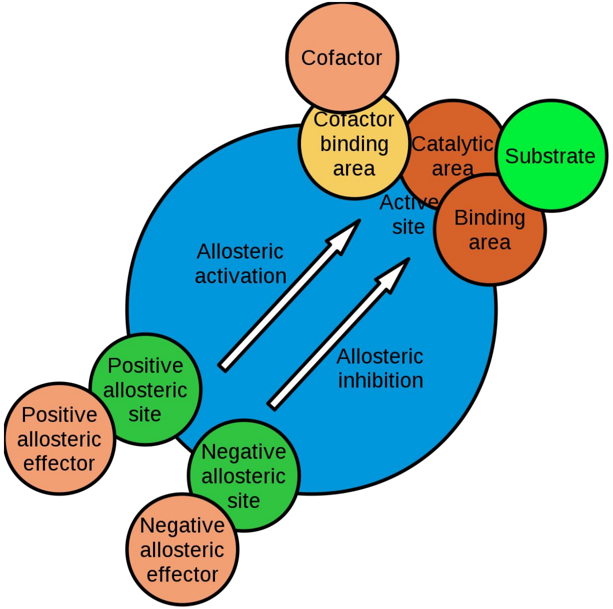
ARTEMIS: Studying the allosteric communication in biomolecules using information theory

R.A. Mallaev, R.G. Efremov, A.A. Polyansky

Moscow, 2025
Laboratory of biomolecular modeling, Shemyakin–Ovchinnikov Institute of bioorganic chemistry RAS

Allosteric regulation



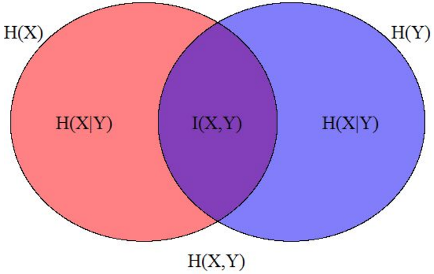


- Plays an important role in functioning of biomolecules
- Assumes the transfer of information between sites
- Has applications in the basic research and biomedicine

Application of information theory

Information theory is the mathematical treatment of the concepts, parameters and rules governing the transmission of messages through communication systems. (**International Encyclopedia of the Social & Behavioral Sciences**, 2001)

$$I(X_1; \ldots; X_n) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1 < \ldots < i_k} S(X_{i_1}, \ldots, X_{i_k})$$

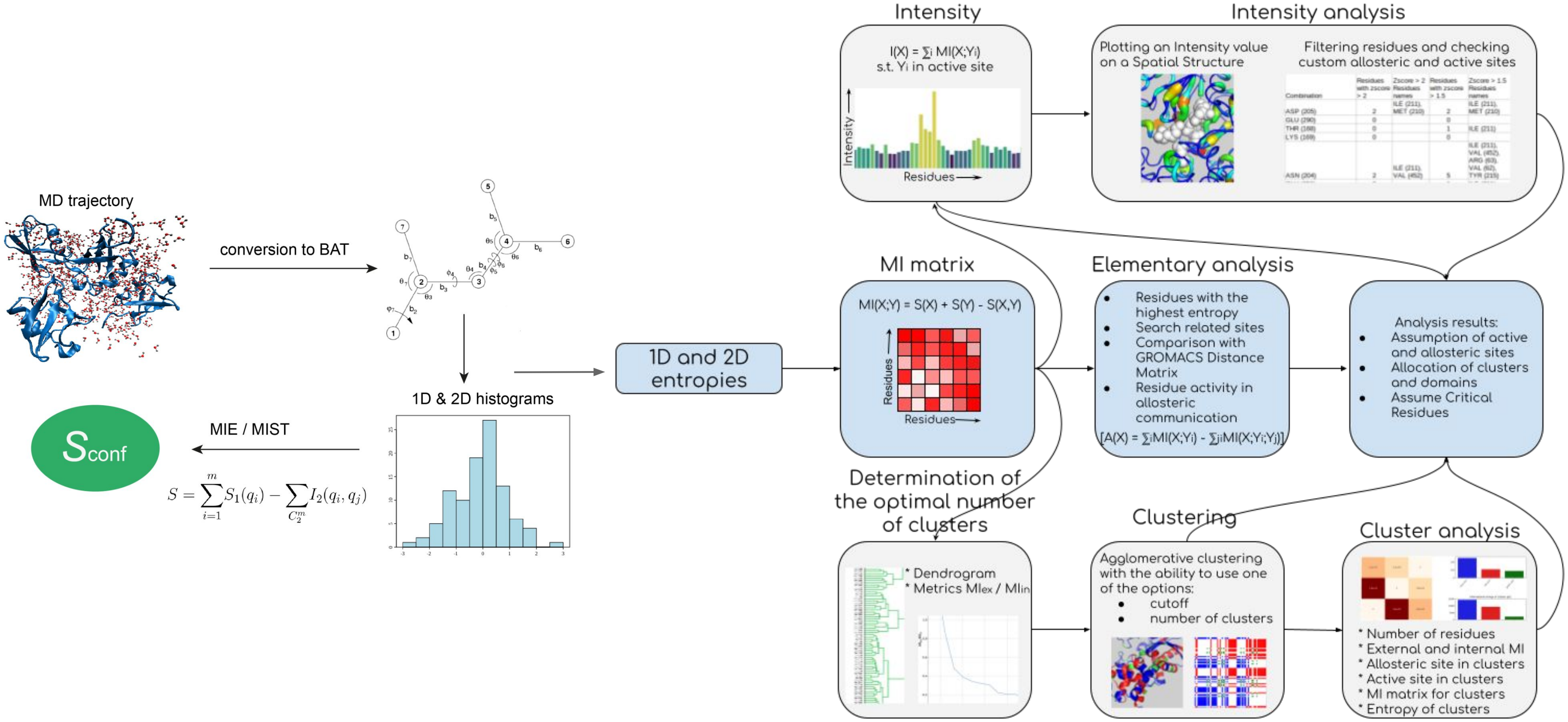


Mutual information (MI) expansion

$$S(X_1; \ldots; X_n) \approx \sum_{k=1}^{n} S(X_k) - \sum_{i<j} I(X_i; X_j)$$

H. Matsuda, Phys. Rev. E **62**, 3096 (2000)

ARTEMIS: Allostery analysis framework



MD trajectory

conversion to BAT

1D & 2D histograms

1D and 2D entropies

MIE / MIST

S_{conf}

$$S = \sum_{i=1}^{m} S_1(q_i) - \sum_{C_2^m} I_2(q_i, q_j)$$

Intensity

I(X) = $\sum_i$ MI(X;Y_i)
s.t. Y_i in active site

Intensity analysis

Plotting an Intensity value on a Spatial Structure

Filtering residues and checking custom allosteric and active sites

Combination	Residues with zscore > 2	Zscore > 2 Residues names	Residues with zscore > 1.5	Zscore > 1.5 Residues names
ASP (205)	2	ILE (211), MET (210)	2	ILE (211), MET (210)
GLU (290)	0		0	
THR (168)	0		1	ILE (211)
LYS (169)	0		0	
ASN (204)	2	ILE (211), VAL (452)	5	ILE (211), VAL (452), ARG (63), VAL (62), TYR (215)

MI matrix

MI(X;Y) = S(X) + S(Y) - S(X,Y)

Elementary analysis

- Residues with the highest entropy
- Search related sites
- Comparison with GROMACS Distance Matrix
- Residue activity in allosteric communication

[A(X) = $\sum_i$MI(X;Y_i) - $\sum_{ij}$MI(X;Y_i;Y_j)]

Analysis results:

- Assumption of active and allosteric sites
- Allocation of clusters and domains
- Assume Critical Residues

Determination of the optimal number of clusters

* Dendrogram
* Metrics MI_{ex} / MI_{in}

Clustering

Agglomerative clustering with the ability to use one of the options:

- cutoff
- number of clusters

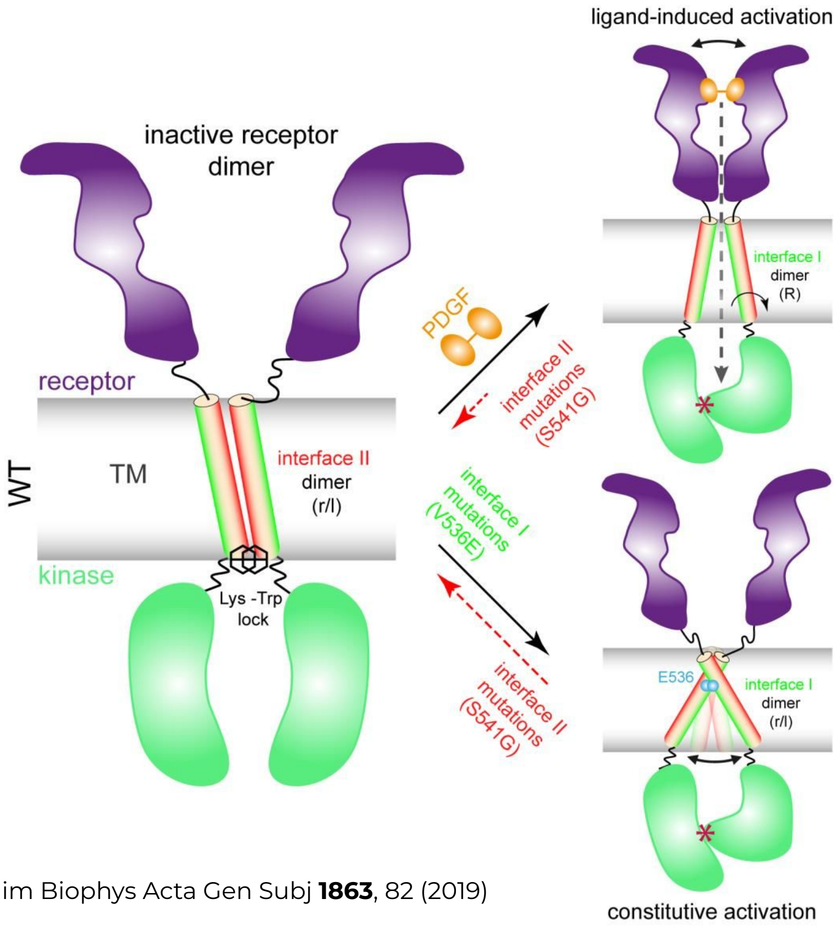
Cluster analysis

* Number of residues
* External and internal MI
* Allosteric site in clusters
* Active site in clusters
* MI matrix for clusters
* Entropy of clusters

Ruslan A. Mallaev et al, ChemRxiv, (2025)
ARTEMIS: https://github.com/nalsur-veallam/ARTEMIS

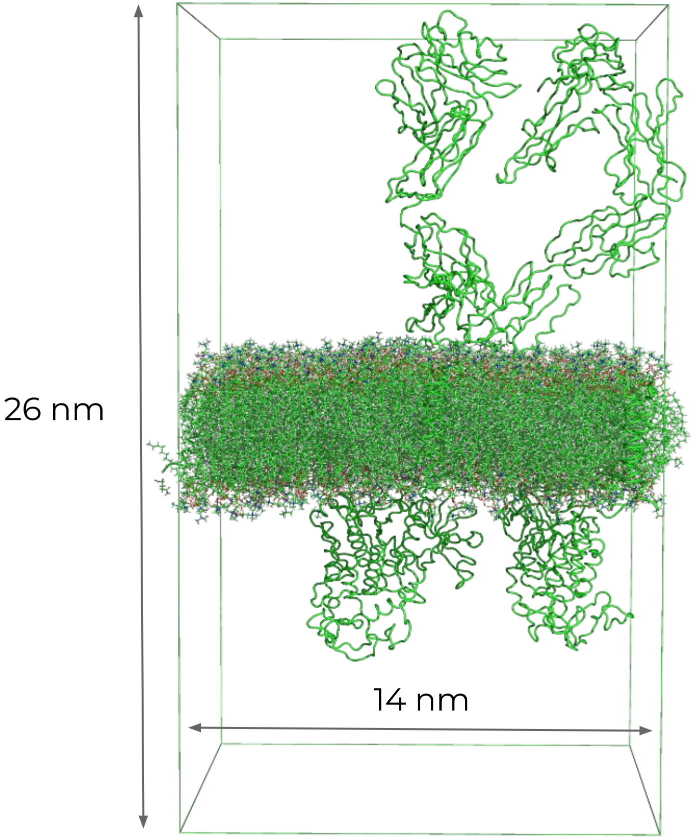
Platelet-derived growth factor receptor A (PDGFRA)

- Plays a key role in the regulation of cell growth, division, and differentiation
- Is activated upon ligand (PDGF) binding, leading to signal transmission
- Is involved in embryonic development and pathological processes, such as cancer (glioblastoma, leukemia)
- Point mutations (for instance, V536E) may cause spontaneous activation, leading to oncogenesis



Anton A Polyansky et al, Biochim Biophys Acta Gen Subj **1863**, 82 (2019)

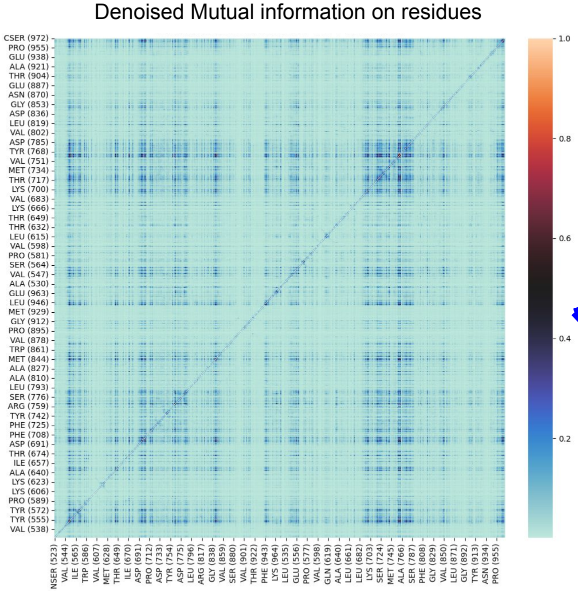
Platelet-derived growth factor receptor A (PDGFRA)

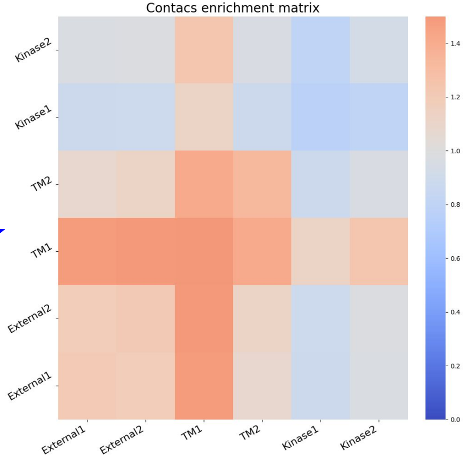


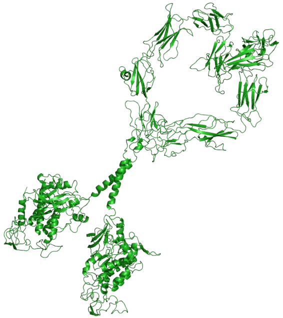
- No experimental receptor structure available. System built manually
- 6 independent MD runs (1 μs each):
  - 2x full-length wild-type receptor
  - 2x full-length V536E mutant
  - 2x wild-type receptor with truncated extracellular domain
- ~ $3*10^4$ atoms (receptor)
- > $5*10^5$ atoms (total)
- Per-trajectory analysis of intramolecular communication
- Comparison: mutant vs. wild-type vs. truncated receptor

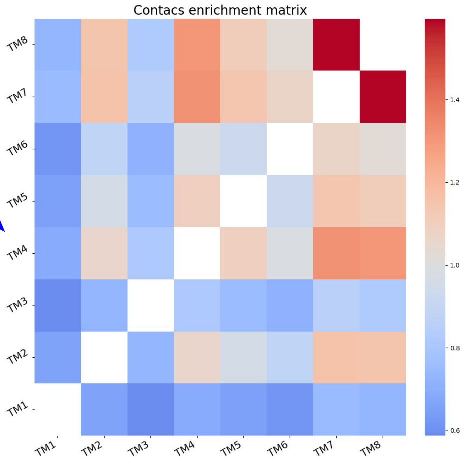
Anton A Polyansky and Roman G. Efremov, CSBJ **21,** 2837 (2023)

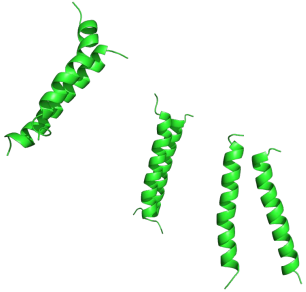
Mutual information matrix (MI)


Denoised Mutual information on residues


Contacs enrichment matrix




Contacs enrichment matrix

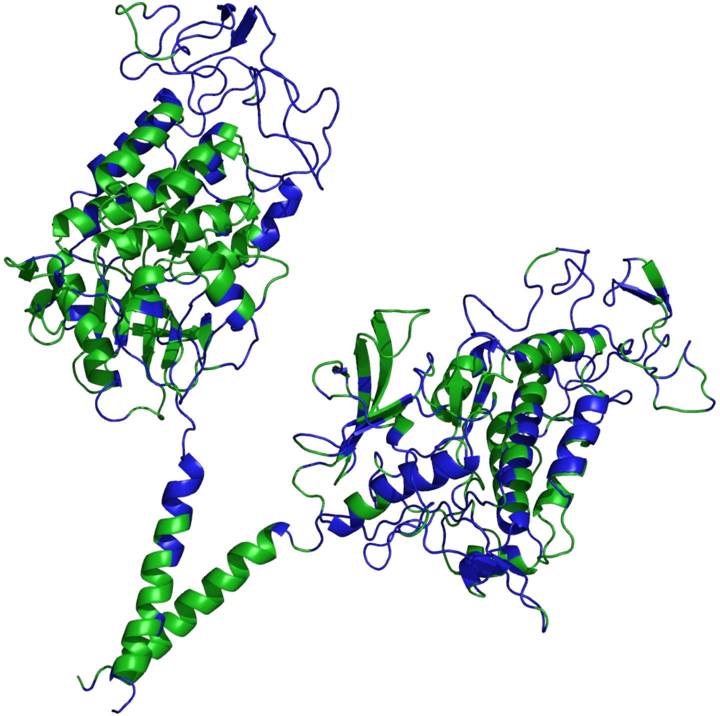


In our implementation, the mutual information for the residues Res$_1$ and Res$_2$, consisting of degrees of freedom {df$_1$} and {df$_2$}, respectively, is calculated as:

$$I(Res_1; Res_2) = \sum_{df_1, df_2} I(df_1; df_2)$$

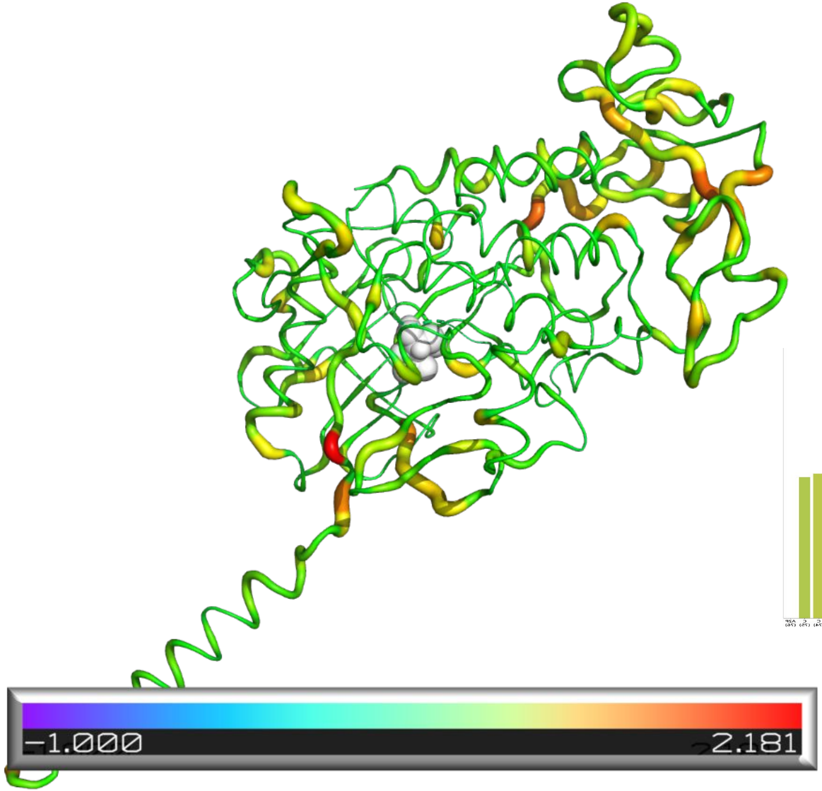
Clustering

	N Contacts	Z-score	Kinase-Kinase MI
PDGFRA WT1	153	1.71	34023
PDGFRA WT2	132	1.31	39544
TMKinase WT1	439	0.89	46116
TMKinase WT2	273	1.23	31137
PDGFRA V536E1	441	1.69	45844
PDGFRA V536E2	241	1.28	47413

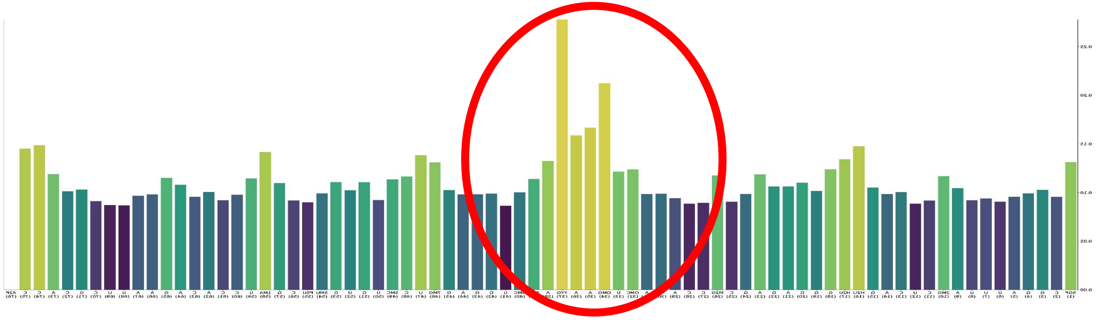


Clustering of the MI matrix allows for the identification of a group of residues that describe the interaction between kinase domains. The flow of information between these domains and the number of residue pairs involved in this interaction are both determined through analysis of this group.

Intensity of communication



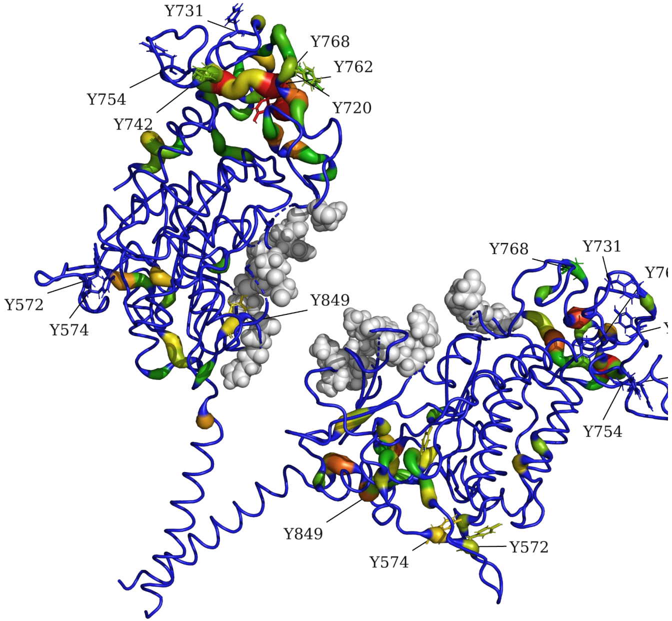
$$intensity(X) = \sum_{Res \in active\ \ site} MI(Res; X)$$



The intensity of intramolecular communication predicts the presence of highly connected positions (amino acid residues) in a protein.

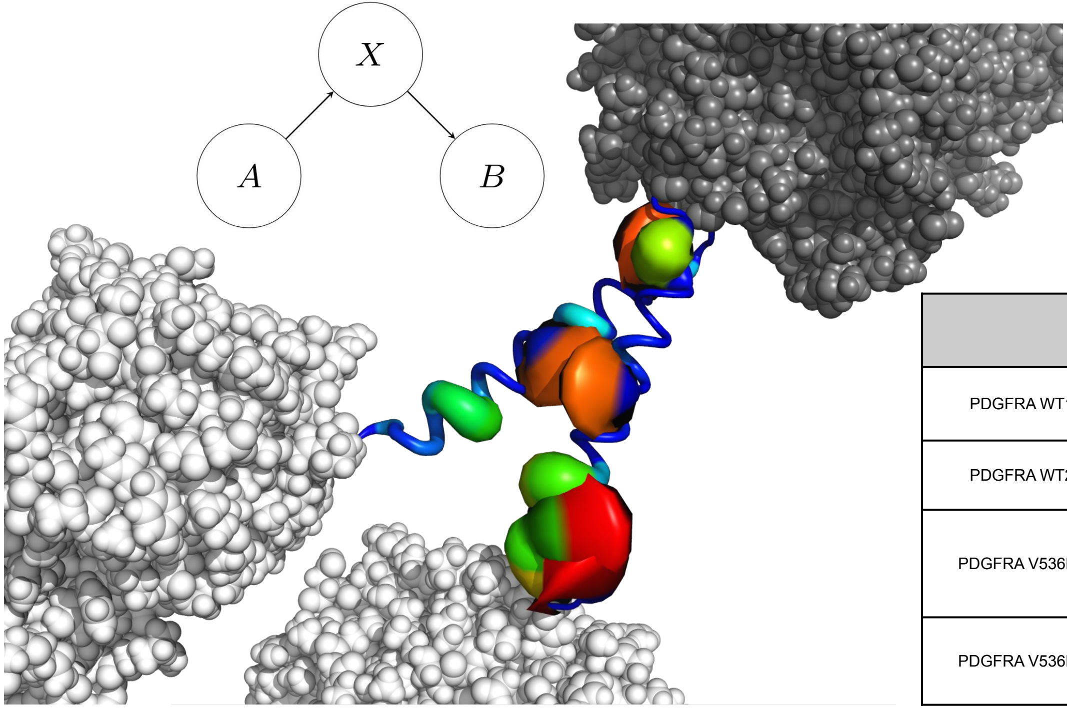
Allosteric Communication Between the Kinase Domain Interface and Phosphorylatable Tyrosines

	Enrichment	Z-score	Interface-Tyr MI
PDGFRA WT1	1.90	1.14	165
PDGFRA WT2	1.66	0.72	177
TMKinase WT1	2.38	1.59	274
TMKinase WT2	2.19	1.31	92
PDGFRA V536E1	2.16	1.67	77
PDGFRA V536E2	1.52	0.69	151



Activating mutations in V536E lead to constitutive kinase activity. We propose that this mutation rewires the allosteric network, enhancing correlation between the dimer interface and the dynamics of key tyrosine residues, thereby stabilizing the active state.

Critical residues in communication

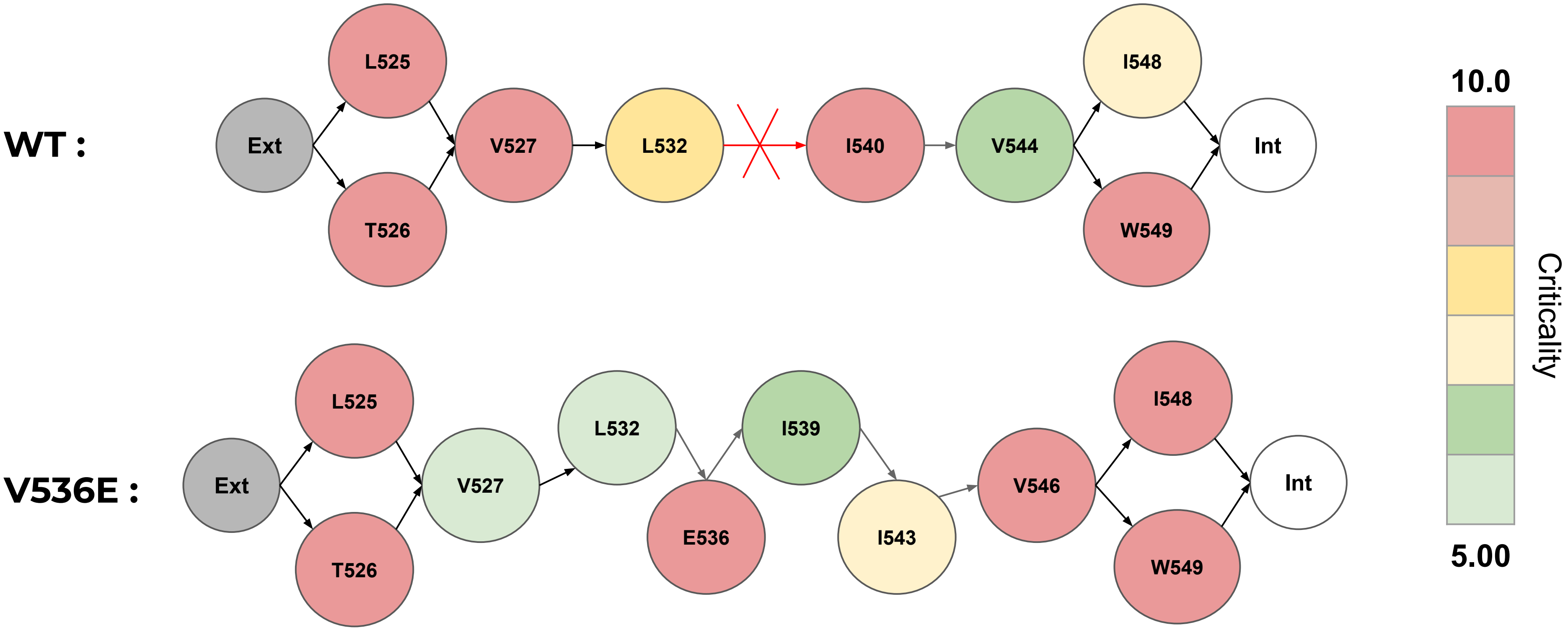


$$\frac{I(X;A)\cdot I(X;B)}{H(X)} \longrightarrow \max$$

	Residues with zscore > 10	Zscore > 10 Residues names	Residues with zscore > 7	Zscore > 7 Residues names
PDGFRA WT1	3	LEU (525)[2], THR (526)[2], LEU (532)[2]	2	THR (526)[1], VAL (544)[2]
PDGFRA WT2	4	VAL (527)[2], TRP (549)[1], THR (526)[1], ILE (540)[1]	4	ILE (548)[2], LEU (532)[2], VAL (546)[2], ILE (548)[1]
PDGFRA V536E1	7	TRP (549)[1], ILE (548)[1], THR (526)[2], VAL (547)[1], VAL (546)[1], LEU (525)[2], GLUH (536)[2]	1	GLUH (536)[1]
PDGFRA V536E2	6	TRP (549)[1], VAL (546)[1], THR (526)[2], GLUH (536)[2], GLUH (536)[1], ILE (548)[1]	7	VAL (544)[1], LEU (525)[1], THR (526)[1], LEU (545)[1], ILE (543)[1], VAL (547)[1], ILE (543)[2]

Maximizing the functional, which is equivalent to minimizing signal loss in a noisy channel, highlights residues that play a critical role in signal transmission between domains.

Critical residues of the TM domain



Therefore, the V536E mutation "closes the connection", simulating signal transmission between the catalytic and receptor regions.

Conclusions

- Mutual information analysis identified correlated residues, enabling a mechanistic communication model
- ARTEMIS reveals communication between non-adjacent transmembrane domains (TMDs) via the membrane
- We propose that the V536E mutation rewires the allosteric network, enhancing correlation between the dimer interface and the dynamics of key tyrosine residues
- The V536E mutation initiates communication between receptor and intracellular domains