

Multiplex Network of a Protein Family



Nivedita Deo

Department of Physics & Astrophysics
University of Delhi,
Delhi, India.

Collaborators:

Pradeep Bhadola

Centre for Theoretical Physics & Natural Philosophy,
Mahidol University, Thailand

Rakhi Kumari

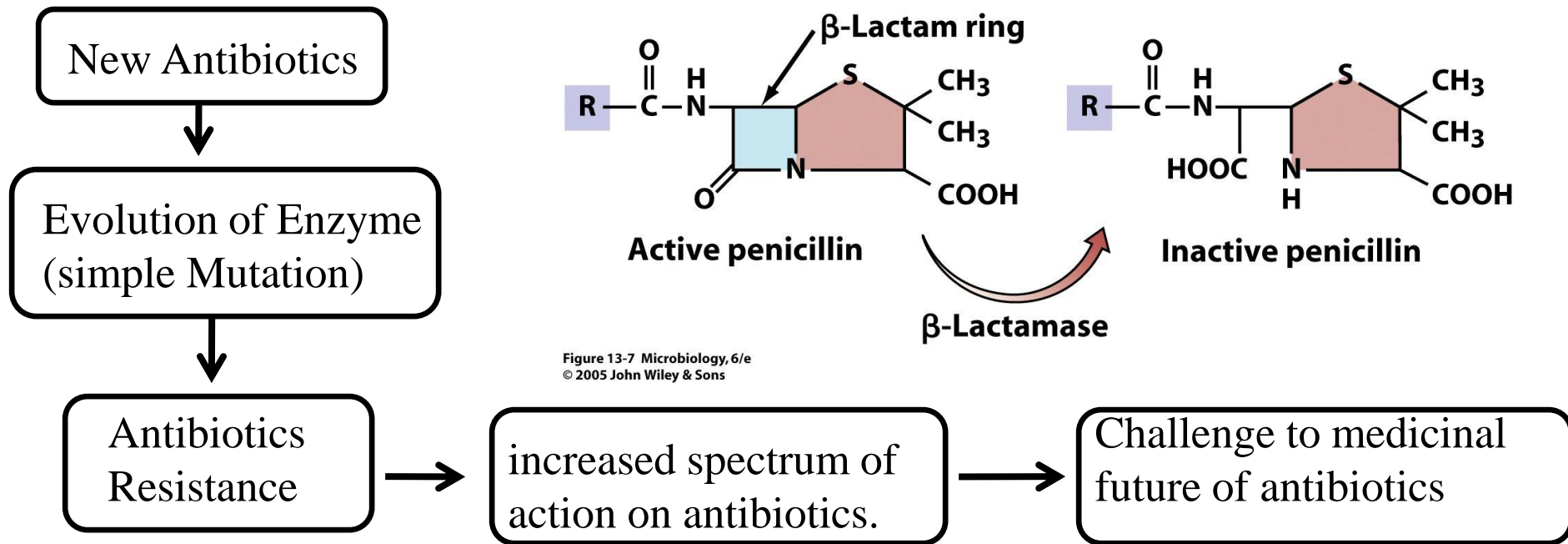
Department of Physics & Astrophysics
University of Delhi, Delhi, India.

Protein

- **Protein molecules** are strings of repeating units of **Amino acids** (20 in number).
- Amino acid order determines the protein
- Sequence of amino acids is important, it determines the 3-dimensional shape of the protein molecule.
- Structure of the protein determines its function

System : Beta-Lactamase Family

- Beta-lactamases: Enzymes secreted by bacteria in response to beta-lactam antibiotics like Penicillin & Bacterial Cephalosporins.
- Beta lactamase enzymes irreversibly hydrolyzes the amide bond of beta-lactam ring making beta-lactam antibiotics inactive.



- ❑ To identify the important motifs or sectors in the protein family for targets to control (deactivate or activate) the enzymatic actions.

Data

- ✓ Interpro entry IPR000871 comprising 5447 proteins for class A/D beta-lactamases family .
- ✓ Contains mainly Phylogentic or Historical Noise
- ✓ To reduce noise , eliminating sequences with similarity > 90 percent. This reduces number to 559 unique sequences.
- ✓ Other sequence similarities 70-95 % are also tested but the results are nearly robust in the range. Most method takes 80-90% similarity.
- ✓ The 559 sequences aligned in multiple sequence alignment (MSA)
- ✓ Positions possessing gaps greater than 20 percent are removed from MSA as these positions are important to individual protein not for the complete family

Multiple Sequence Alignment

```
ADPFLESKFSARLGVYAIDTGTERTIVAYRPPDERFAYASTFKALAAGAVLQ
-----
LRNI----RATVGVALI-VDGKDTLTVNNGFHYPTQSTYKFHLALAVLD
VEAILEQKHNASIGVFATNLASGKTVAHRAQEPFAMCSTFKAYASGRILQ
LQEL----KAQVGIAMI-LDGRDTLTVNNDVRFPMMSVFKFHQALAVAD
VSQL----VGRIGVAAQEIGSGERITVNGDEMFMMASTYKVAIAVALLE
LDAILEAQYATRIGVTAVDPQTGAVYSHRGDERFAMCSTFKAYASAAVLR
LDKLLHSANGRLGIAMINSGAGTKILYRGAQRFPPFCSTFKFMLAAAVLD
--WIECGTGGRLGINLFDSATGWRLGQREDERFPMCSTFKFVLAAGAVLQ
--RGAELPGEVSLSATDLLTGRRLEHRGERPCPVLTLASAGLALGKLLR
VRQLLERGYPGRIGLAAFAPRTGARVDYRATERFALASTFKVLTAAAVLR
LVKVAEDQLGARVGYIELDLNSGKILSFRPEERFPMMSTFKVLLCGAVLS
LEQISESLSGRVGMIEMDLASGRTLAWRADERFPMMSTFKVVLCGAVLA
VVQALEQSLGARVGLAVLDTADGHVW-YRANERFPMASTSKLVICAALLA
DADFIERSVAGRLGVCAYDLETGRSLGHRQDERFAMCSTFKWLLAAAVLA
AREFLEASSGGRLGVAALNTADGSYVGYRESERFPMCSTFKLLVVALVLK
LTQIIIEQTLDARVGLHIHDSATGWTV-HRPNERFLMASTFKAMLCGAVLS
LKMI----RATVGVAVL-YGDNRLFVN-RGDPMMSVCKFPLALAVLD
LKQI----DGTLGIGISGLNNRDTLTLNHHVHFAMQSVYKFHLGLAILN
MDGLLEQDAEITLAVCASVQGGQR--EYRATEPLAMCSTFKAILVGVVLS
-I--LETDFNGRIGVYALDTGSGKSFYSYKANERFPLCSSFKGFLAAAVLK
LTALIEGVFEGELGVYVQNLASGEGFSWRADAPWYLASLIKIPVAAQVLA
IIAAVEERLGMRVGLSVIDTGRGTVVAYRETERFAMASTFKSLACAAAL-
LKQI----KATVGISIQHLETGQTLTINNDKHYPMQSVYKFHIAMAVLN
SGRFLQREFDARLGVYAVDTGSGRTVTHRPDERFAYCSTHKALTAAGAVLE
PDELLAQRHDAEIGVYAVDLVTQDTLAFRADETFAVCSTIKTYIAAGAVLQ
AAGMLEERSGGRLGVAVLDTATGRLRGWRGDERFPMCSTFKFLLAAAVLK
```

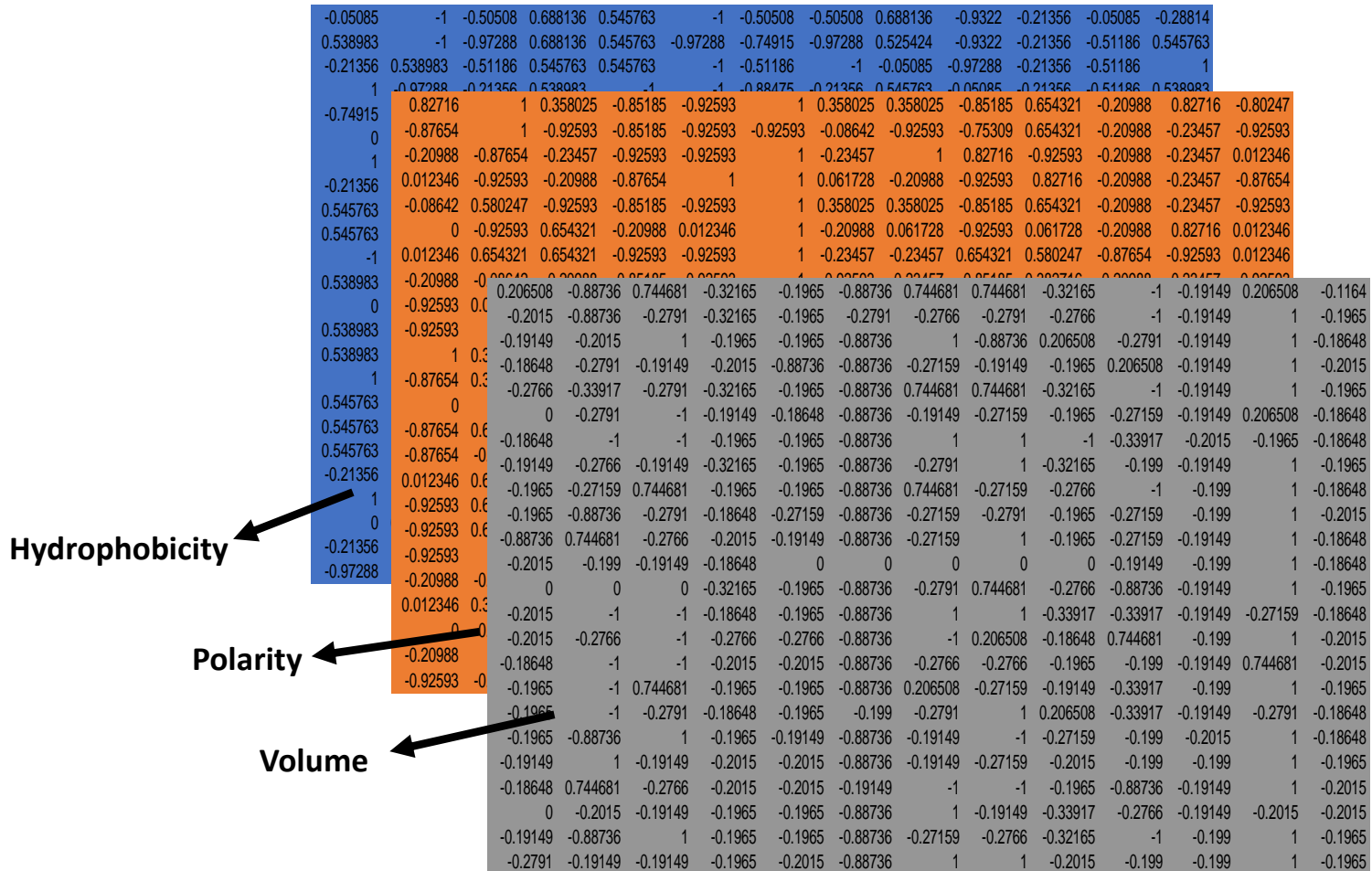
- Interpro entry IPR000871 comprising 5447 proteins for class A/D beta-lactamases family.

Physiochemical Based Datamatrices(D^α)

D_{si}^α = Property α of amino acid in the i^{th} position of sequence s .

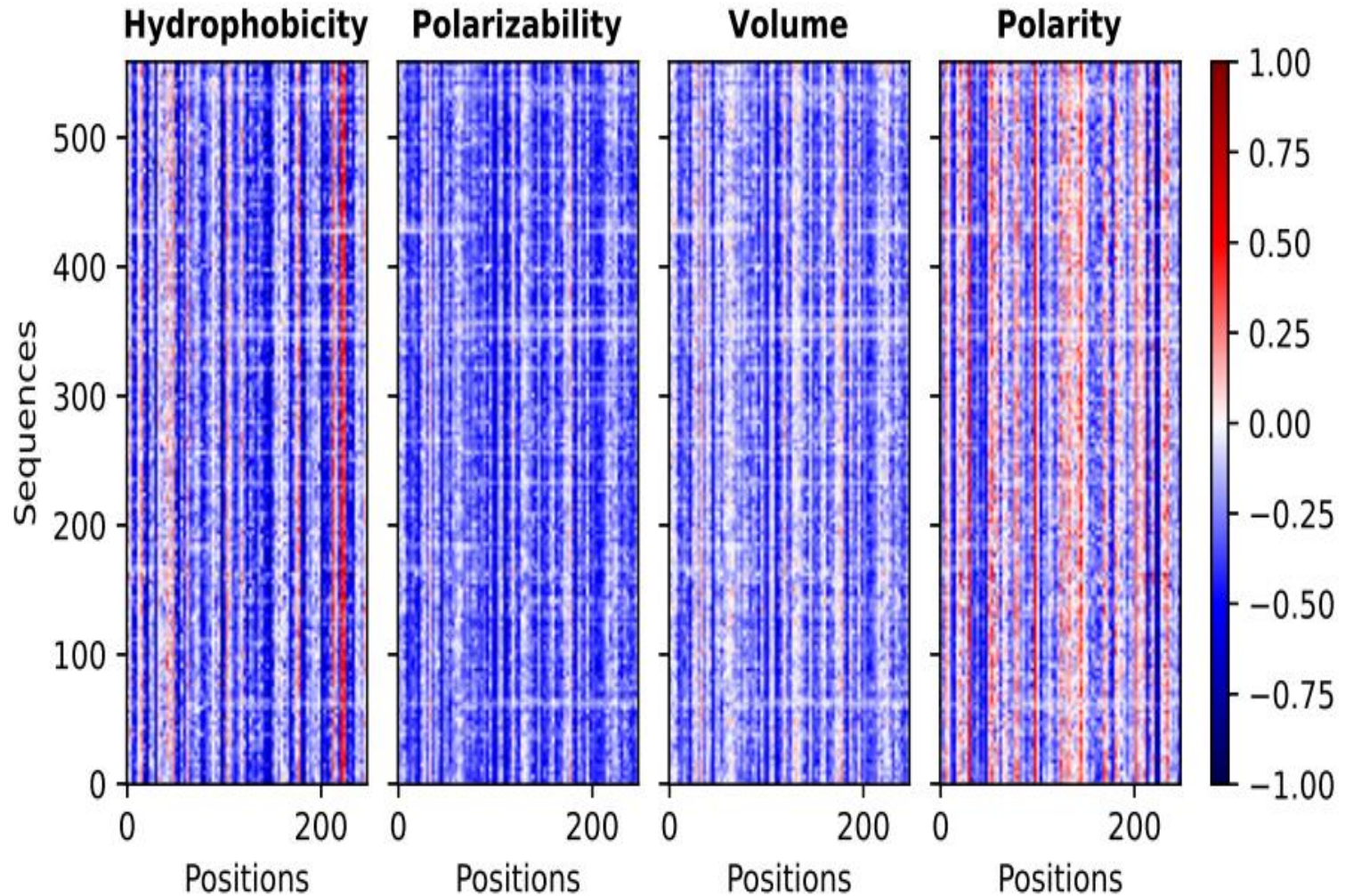
$i = 1, 2, \dots, L$ where L is the length of each Sequence = 248

$s = 1, 2, \dots, S$ where S is the number of sequences = 559



Physiochemical Based Datamatrices(D^α)

Data matrix for *Beta*-lactamase family for different properties



Data matrix changes with properties, even though it is derived from the same MSA, suggesting that each property provides unique and valuable information.

Correlations

The co-evolution between positions of the data matrix D is given by

$$C_{i,j}^{\alpha} = \frac{Cov(d_i^{\alpha}, d_j^{\alpha})}{\sigma_{d_i^{\alpha}} \sigma_{d_j^{\alpha}}}$$

Where d_i^{α} is the i^{th} column of data matrix D^{α} having standard deviation $\sigma_{d_i^{\alpha}}$

$Cov(d_i^{\alpha}, d_j^{\alpha})$ is the covariance between i^{th} and j^{th} column of data matrix D^{α} given by

$$Cov(d_i^{\alpha}, d_j^{\alpha}) = \langle (d_{s,i}^{\alpha} - \langle d_{s,i}^{\alpha} \rangle) (d_{s,j}^{\alpha} - \langle d_{s,j}^{\alpha} \rangle) \rangle$$

$\langle .. \rangle$ implies average over sequences.

Proteins as a Multiplex Network

- ✓ To characterize multiple types of interactions between positions in a protein family, We created a weighted multiplex network of evolutionary interactions between positions for the β -lactamase family.
- ✓ Multilayer networks also consist of nodes and edges, but the nodes exist in separate layers, representing different forms of interactions.
- ✓ When interlayer edges can only link nodes to nodes representing the same entity in different layers, the network is classified as a **Multiplex** network.
- ✓ Each layer realizes a different aspect of the interaction between amino acids

Threshold Multiplex Network

A multiplex network is a set of networks $G^\alpha(N,E)$, arranged in layers, with $\alpha = 1, \dots, L$ with L as the number of layers. The set of nodes N is the same in each layer whereas the set of edges E^α is layer dependent.

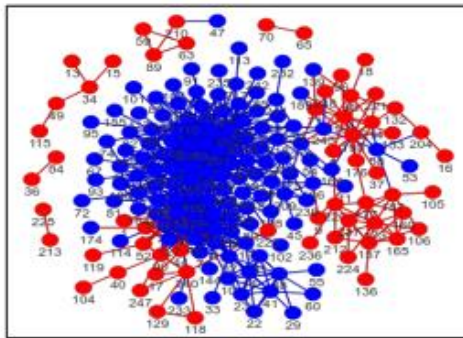
Each layer in network (α) can be represented by a graph $G^\alpha(N,E)$ with nodes (N) given by the positions in the multiple sequence alignment and edges

$$E_{i,j}^\alpha = \begin{cases} |C_{ij}^\alpha| & \text{if } i \neq j, |C_{ij}^\alpha| \geq \theta \\ 0 & \text{otherwise,} \end{cases} \quad \text{where } \theta \text{ is the threshold value.}$$

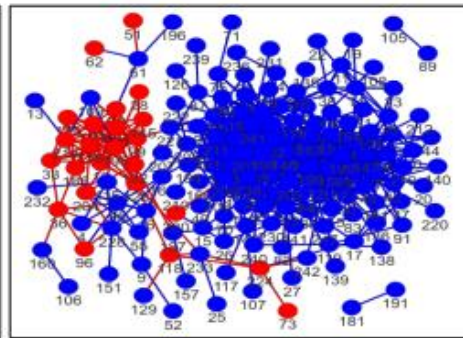
Different threshold generates different networks with same set of nodes but different edges.

Connected components for different threshold are extracted which shows both structural and functional significance

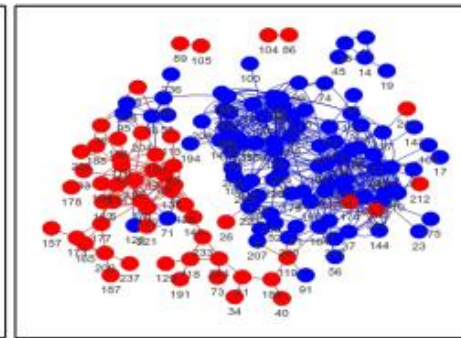
Network Layers at Different Thresholds



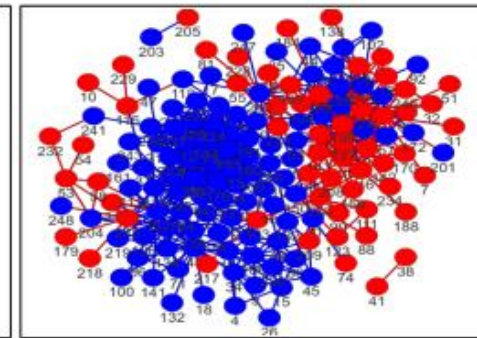
(a) Hydrophobicity(0.3)



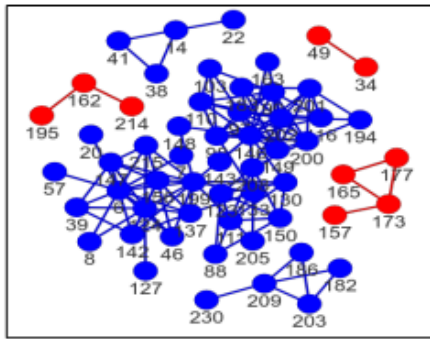
(b) Polarizability(0.3)



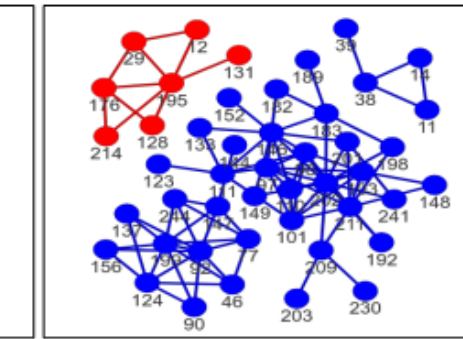
(c) Volume(0.3)



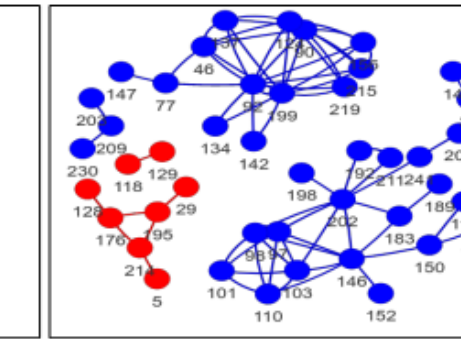
(d) Polarity(0.3)



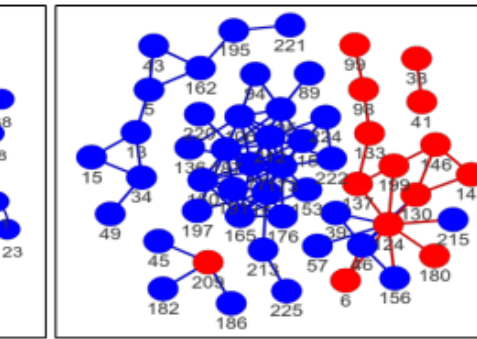
(a) Hydrophobicity(0.5)



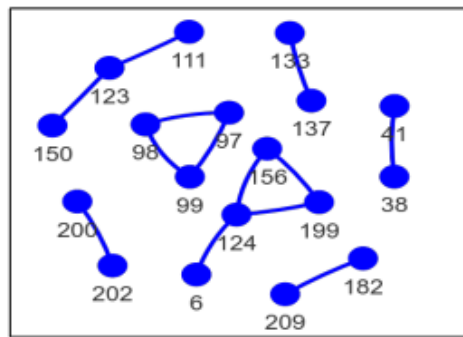
(b) Polarizability(0.5)



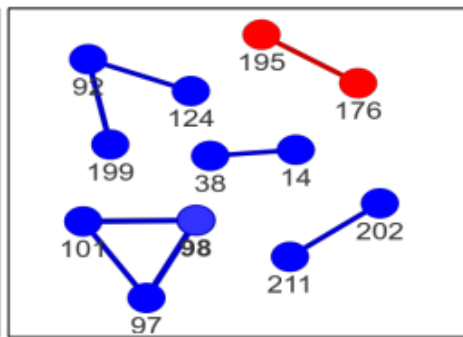
(c) Volume(0.5)



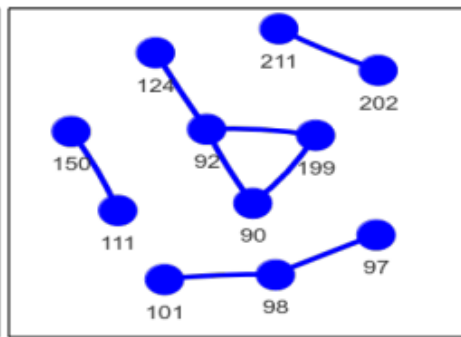
(d) Polarity(0.5)



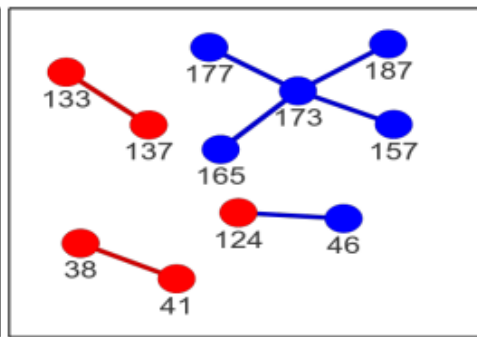
(i) Hydrophobicity(0.7)



(j) Polarizability(0.7)



(k) Volume(0.7)



(l) Polarity(0.7)

Nodes with identical property signs tend to aggregate, providing evidence of the presence of consequential functional and evolutionary constraints shaping the Beta-lactamase family.

Topological Properties of each Layer

Topological Properties of the *Beta*-lactamase family such as density, number of edges, average clustering, average degree (K^{avg}), maximum degree (K^{max}), size of largest component N^{comp} , average path length L^{avg} , Radius (R), and Average Eccentricity (ϵ) at different threshold θ

| θ | Property | Density | Edges | Avg. Clus. | K_{avg} | K_{max} | N_{comp} | L_{avg} | R | ϵ |
|----------|----------------|---------|-------|------------|-----------|-----------|------------|-----------|-----|------------|
| 0.1 | Hydrophobicity | 0.2408 | 7365 | 0.519 | 59.39 | 129 | 248 | 1.87 | 3 | 3.00 |
| | Polarizability | 0.2854 | 8740 | 0.530 | 70.484 | 145 | 248 | 1.74 | 2 | 2.81 |
| | Volume | 0.2204 | 6753 | 0.478 | 54.459 | 114 | 248 | 1.85 | 2 | 2.99 |
| | Polarity | 0.2019 | 6185 | 0.472 | 49.879 | 88 | 248 | 1.89 | 3 | 3.04 |
| 0.3 | Hydrophobicity | 0.0275 | 842 | 0.388 | 6.790 | 52 | 177 | 3.72 | 5 | 7.78 |
| | Polarizability | 0.0298 | 914 | 0.339 | 7.371 | 55 | 168 | 3.37 | 5 | 6.59 |
| | Volume | 0.0209 | 639 | 0.278 | 5.153 | 36 | 143 | 4.398 | 7 | 8.61 |
| | Polarity | 0.0237 | 726 | 0.331 | 5.855 | 33 | 158 | 3.54 | 5 | 6.25 |
| 0.5 | Hydrophobicity | 0.0042 | 129 | 0.127 | 1.040 | 17 | 39 | 2.59 | 3 | 4.13 |
| | Polarizability | 0.0031 | 95 | 0.109 | 0.766 | 14 | 24 | 2.25 | 3 | 3.67 |
| | Volume | 0.0025 | 77 | 0.080 | 0.621 | 11 | 18 | 2.39 | 3 | 3.83 |
| | Polarity | 0.0030 | 92 | 0.093 | 0.742 | 15 | 22 | 2.10 | 3 | 3.77 |
| 0.7 | Hydrophobicity | 0.0004 | 13 | 0.022 | 0.104 | 3 | 4 | 1.33 | 1 | 1.75 |
| | Polarizability | 0.0003 | 8 | 0.012 | 0.065 | 2 | 3 | 1.33 | 1 | 1.67 |
| | Volume | 0.0002 | 8 | 0.009 | 0.064 | 3 | 4 | 1.33 | 1 | 1.75 |
| | Polarity | 0.0002 | 7 | 0.000 | 0.056 | 4 | 5 | 1.60 | 1 | 1.80 |

Presence of hubs in the network, which are nodes with remarkably high degrees. For example, hubs (200, 202, 199, 124, 137, 146, 6, 97, 99, 98, 149, 133, 137, 38, 41) for the hydrophobicity ($\vartheta = 0.5$), which when compared with the literature are found to have important functional and structural role

Multilinks

A multi-link between two nodes is defined by the vector \vec{m} such that $m^\alpha = 1$ if the two nodes are connected by a link in layer α and zero otherwise.

In general, a multi-link between two nodes say i and j is given using the layer adjacency matrix as $\vec{m} = \vec{m}_{ij} = (a_{ij}^1, a_{ij}^2, \dots, a_{ij}^\alpha, \dots, a_{ij}^L)$ where a^α being the adjacency matrix for network layer α .

If two given nodes are connected in every layer then $\vec{m} = \vec{1}$ whereas multi-link $\vec{m} = \vec{0}$ signifies that the nodes are not directly connected in any layer.

Using multi-links, we define the multi-Adjacency matrix ($\mathcal{A}^{\vec{m}}$) as

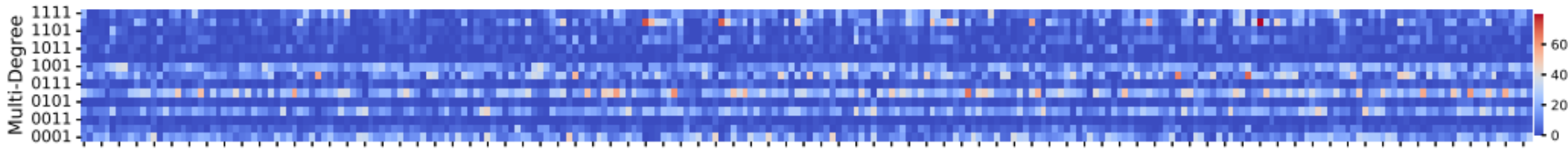
$$\mathcal{A}_{ij}^{\vec{m}} = \prod_{\alpha=1}^L [a_{ij}^\alpha m^\alpha + (1 - a_{ij}^\alpha)(1 - m^\alpha)]$$

With multi-adjacency matrix, one can define the *multi-degree* $\mathcal{K}^{\vec{m}}$ of a node i as

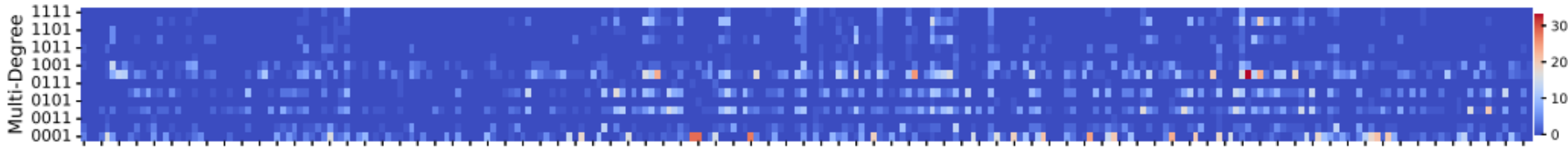
$$\mathcal{K}_i^{\vec{m}} = \sum_{j=1}^N \mathcal{A}_{ij}^{\vec{m}}$$

Polarity (0001), Volume (0010), Polarizability (0100), and Hydrophobicity (1000).

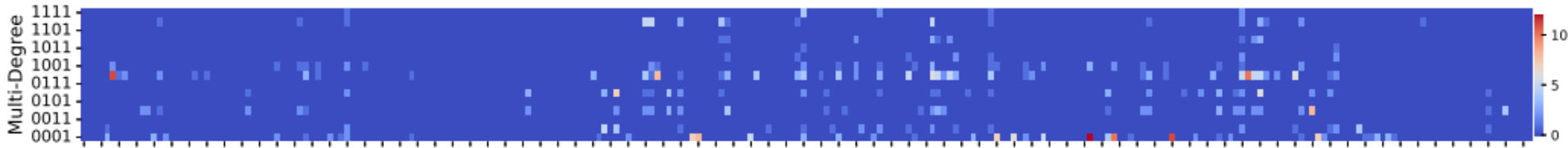
Threshold = 0.1



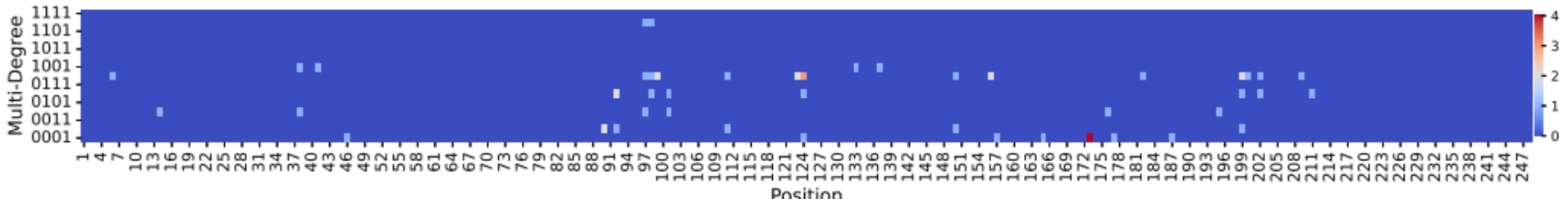
Threshold = 0.3



Threshold = 0.5



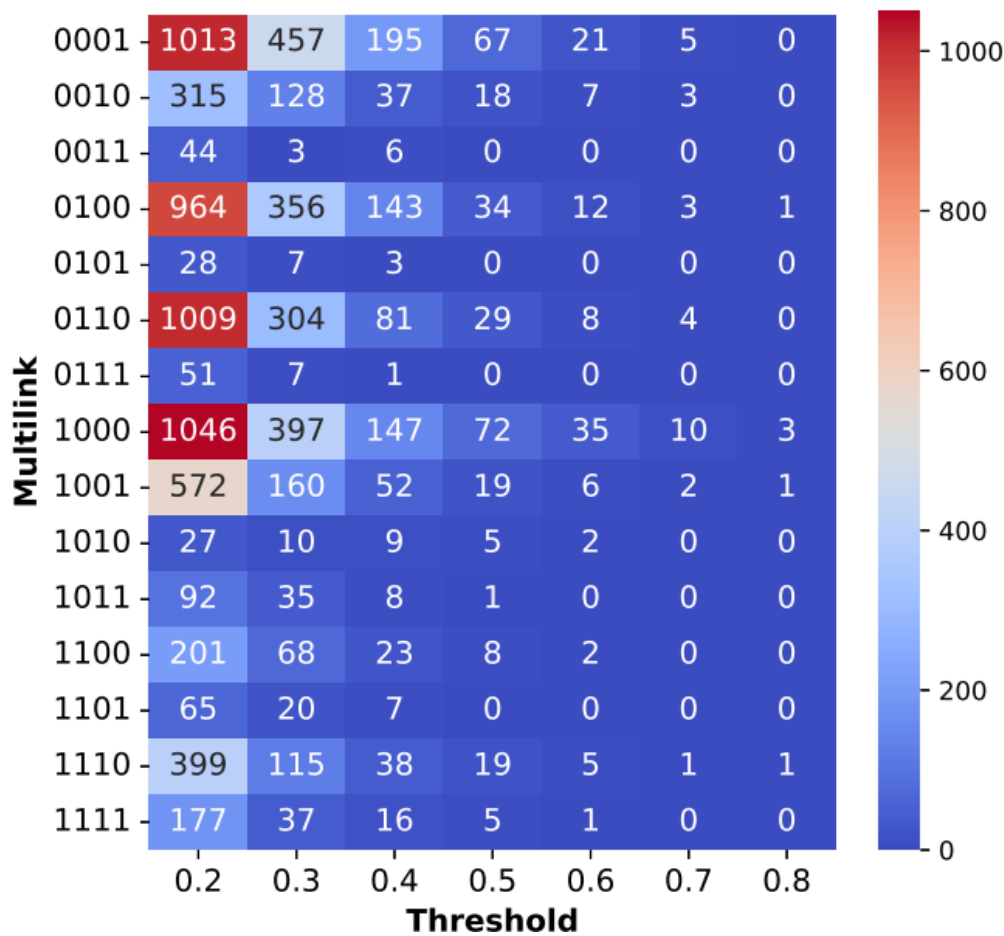
Threshold = 0.7



For a specific multi-degree say 1000, indicates links exclusive only to hydrophobicity layer where a 1001 suggests links common to both the hydrophobicity and polarity layers.

At $\theta = 0.1$ Out of the total 248 positions, 233 positions have non-zero values for the multi degree $\mathcal{K}^{\vec{1}}$ --- **Statistical Noise**.

At $\theta = 0.5$, only five nodes (46, 124, 137, 156, 199) with a non-zero multi-degree $\mathcal{K}^{\vec{1}}$.

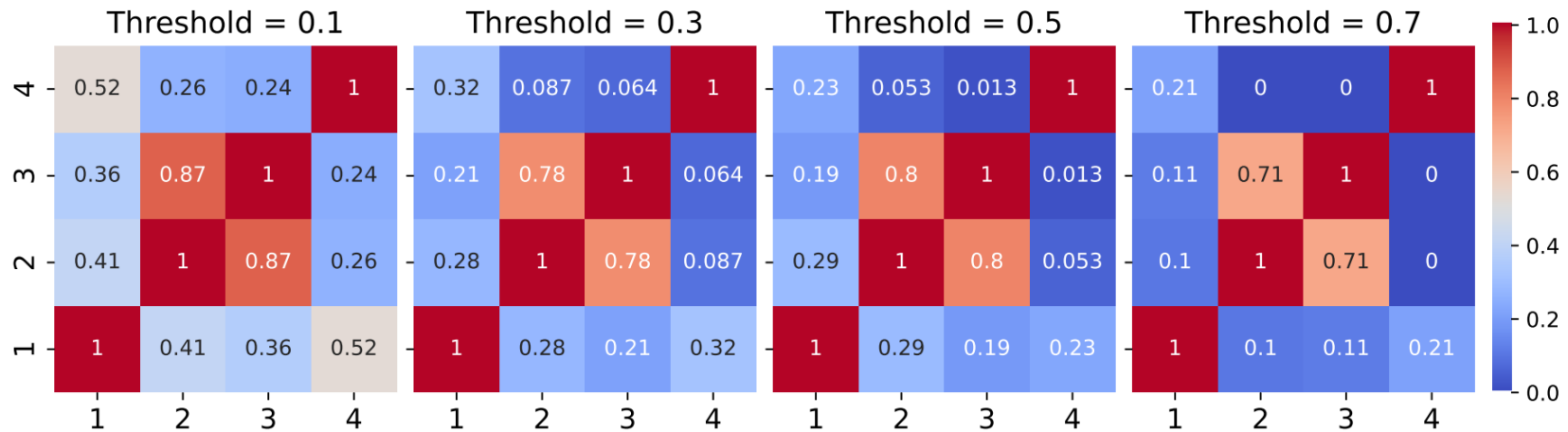


Distribution of multi-links in the network at the different thresholds. In multi-links, the properties are represented from least significant bit to most significant in order of Polarity, Volume, Polarizability, and hydrophobicity.

- At $\theta < 0.5$, Polarity (0001) consistently exhibits the highest number of unique evolutionary links.
- Volume layer (0010) demonstrates the lowest number of unique links.
- At ($\theta > 0.5$), hydrophobicity surpasses polarity a **stronger overall evolutionary interaction within the hydrophobicity compared to polarity (other layers)**.
- The combination of polarizability and volume (0110) exhibits the highest number of simultaneous connections, followed by the combinations of hydrophobicity-polarity (1001) and hydrophobicity-polarizability (1100).
- Layers (0011, 0101, 1010) exhibit almost zero multidegree.
- Hydrophobicity, polarizability, and volume (1110) reveals links that are common to all three layers, but absent in the polarity layer, even at a very high threshold of 0.8, indicating evolutionary conservation.

The multi-layer structure of the multiplex network provides valuable insights into the hierarchy of physiochemical properties influencing individual positions

Similarity between network layers



Layers displayed as 1-hydrophobicity, 2-polarizability, 3-volume, and 4-polarity.

- At a low threshold ($\theta = 0.1$), significant overlap in the interaction among all properties. Threshold range 0.0- 0.2 represents a noisy region, high overlap is attributed to the random links between positions but these links are very weak in strength.
- As the threshold increases, the overlap between layers in the network starts to decrease, although polarizability and volume continue to show significant overlap.
- In the threshold region where the information is predominant ($\theta > 0.3$), there is a sudden decrease in the overlaps and only the actual link contributes. Overlap between polarity with volume and polarizability is over 25% at threshold 0.1 but reduces to almost zero at threshold 0.7.
- Volume and polarizability consistently exhibit an high overlap of approximately 60% across all thresholds, indicating that they encode similar information.
- Consequently, it will be reasonably safe to reduce the size of the multiplex network by eliminating one of the two layers (either volume or polarizability) with minimal loss of information.

Conclusions

- By deriving each layer of the network from a correlation matrix calculated using physiochemical properties, unveil novel information about the intricate interactions between nodes, allowing us to selectively determine key positions and interactions that may act as potential targets for influencing enzymatic or catalytic activity.
- Although each layer is derived from the same MSA, but it unravels a piece of different information in terms of interaction between the nodes giving useful insight into the functionality and structure of the protein family.
- Interaction between the positions depends on the physiochemical properties where positions tend to cluster into groups with identical signs of the property.
- Reveals the hierarchy in the influence of physiochemical properties at a given position, pinpointing the most relevant property responsible for the protein's functionality.
- Link overlap analysis reveals that there are limited information exchanges between any two layers, indicating the importance of combining layers to shed light on their collective behavior.
- The combination of hydrophobicity, polarizability, and volume exhibits common links across all layers, suggesting functional or structural constraints for the beta-lactamase family.
- Furthermore, the multiplex network proposed in this study exhibits considerable potential for broader utilization across various protein families, offering invaluable insights into their structural and functional characteristics.