Draft version September 2, 2025
Typeset using LaTeX **twocolumn** style in AASTeX631

Linking Residue-Level Network Dynamics to Peptide Aggregate Stability: A Hierarchical Spectral Graph Analysis of KYFIL Self-Assembly

AstroPilot[1]

[1] *Anthropic, Gemini & OpenAI servers. Planet Earth.*

ABSTRACT

Understanding the relationship between microscopic interactions and macroscopic stability is crucial for designing self-assembling peptide materials. We propose and apply a novel hierarchical graph-based approach to analyze the self-assembly of K-Y-F-I-L pentapeptides using a molecular dynamics simulation trajectory. The method involves constructing time-evolving graphs at two levels: a peptide-level graph tracking aggregate formation and persistence, and detailed residue-level contact graphs for identified persistent aggregates. We analyze spectral properties, such as algebraic connectivity (Fiedler value λ_2), and other graph metrics including density and clustering coefficient, focusing on their time evolution within these residue-level networks. The analysis revealed that while the system forms a dominant large aggregate at the peptide level, the internal residue-level contact network within persistent aggregates exhibits consistently zero algebraic connectivity, indicating a disconnected or minimally connected global structure despite high local clustering. This finding suggests that aggregate stability in this system may arise from a collection of dynamic local interactions rather than a single, globally robust residue network, and consequently limits the direct use of global connectivity metrics like λ_2 for predicting instability. However, residue-level network density and average clustering coefficient were found to change significantly around aggregate dissolution and growth events, suggesting their sensitivity to peripheral association and dissociation dynamics. This hierarchical approach provides a multi-scale perspective on peptide self-assembly and identifies residue-level density and clustering as potential indicators of local structural changes associated with aggregate evolution.

Keywords: Detection, Time series analysis, Computational methods, Distributed computing, Algorithms

1. INTRODUCTION

Self-assembling peptides are remarkable molecular building blocks, capable of spontaneously forming diverse and functional nanostructures such as fibers, hydrogels, and nanoparticles. These structures hold significant promise for applications ranging from drug delivery and tissue engineering to catalysis and biosensing. The intricate process by which simple peptide sequences organize into these complex architectures is fundamentally governed by the collective interplay of non-covalent interactions between individual amino acid residues and peptides. A central challenge in harnessing the full potential of these materials lies in establishing a clear and quantitative understanding of the relationship between these microscopic residue-level interactions and the macroscopic stability, morphology, and dynamic behavior of the resulting peptide aggregates.

Molecular dynamics (MD) simulations provide atomic-resolution insights into the dynamic processes of peptide self-assembly, capturing the transient nature of inter-peptide and intra-aggregate interactions over time (Horowitz & Hughto 2008). However, extracting meaningful, system-level principles that govern aggregate stability from the vast and dynamic network of interactions within a simulation trajectory is a significant analytical hurdle. Traditional approaches often focus on average properties or specific types of interactions, which can fail to capture the emergent collective network behavior that underpins the structural integrity and dynamics of an aggregate (Horowitz & Hughto 2008). While graph theory offers a powerful framework for representing complex interaction networks, its application to the inherently dynamic and multi-scale nature of peptide self-assembly, particularly in linking residue-level network properties to the stability of ag-

gregates defined at a higher level, remains an active area of exploration.

The difficulty lies in developing a systematic analytical framework that can simultaneously identify and track aggregate formation at the peptide level while also analyzing the evolving internal network structure at the residue level within those aggregates, and then rigorously correlating the dynamics across these scales to understand stability (Horowitz & Hughto 2008).

In this paper, we address this challenge by proposing and implementing a novel hierarchical graph-based analysis framework designed to link residue-level network dynamics to peptide aggregate stability. Our approach leverages an MD simulation trajectory of the self-assembling K-Y-F-I-L pentapeptide system. The framework operates at two distinct levels of granularity. First, we construct a time-evolving peptide-level contact graph, where nodes represent individual peptides and edges denote inter-peptide contacts. This allows us to identify, track, and analyze the formation, growth, and persistence of peptide aggregates throughout the simulation. Second, for aggregates identified as persistent over defined time periods, we delve into their internal structure by constructing detailed residue-level contact graphs. In these graphs, nodes represent individual amino acid residues within the aggregate, and edges represent inter-residue contacts.

The core of our analysis focuses on characterizing the time evolution of various graph theoretical metrics *specifically within these dynamic residue-level networks*. We analyze spectral properties of the graph Laplacian, such as the algebraic connectivity (or Fiedler value, λ_2) and its corresponding Fiedler vector, which provide insights into the global connectivity and structural bottlenecks of the network (Yang & Yu 2023; Strey et al. 2024). We also examine traditional graph measures, including network density and average clustering coefficient, to understand the overall connectivity and local cohesiveness of the residue network (Pavlou et al. 2023; Yang & Yu 2023). By systematically analyzing how these residue-level network properties fluctuate over time and correlating these dynamics with observed changes in the parent peptide aggregate's behavior (such as growth, shrinkage, splitting, or dissolution events), we aim to uncover the specific microscopic network signatures that are indicative of macroscopic aggregate stability or propensity for structural rearrangement (Yang & Yu 2023; Strey et al. 2024).

To verify the utility and insights provided by this hierarchical approach, we apply it to a representative MD trajectory of the K-Y-F-I-L system. We demonstrate how the peptide-level graph effectively maps the overall aggregation landscape, including the formation of a dominant large aggregate. Our subsequent detailed analysis of the residue-level graphs within persistent aggregates reveals that, in this specific system, the internal residue network often exhibits consistently zero algebraic connectivity ($\lambda_2 = 0$), suggesting a lack of robust global connectivity across the entire aggregate despite the presence of significant local clustering (Pandya et al. 2025). However, we show that other residue-level metrics, such as network density and average clustering coefficient, are sensitive to changes associated with aggregate evolution, such as peripheral peptide association and dissociation events (Makinen et al. 2024,?). This analysis highlights the potential of specific residue-level network dynamics as indicators of local structural changes relevant to aggregate stability and evolution. Ultimately, this work establishes a systematic framework for bridging the gap between residue-level interactions and aggregate stability in peptide self-assembly, providing a deeper, chemically-informed understanding of these complex processes (Serjeant et al. 2024).

2. METHODS

The primary objective of this study is to establish and apply a hierarchical graph-based framework to analyze the self-assembly of K-Y-F-I-L pentapeptides, linking residue-level network dynamics to aggregate stability. This framework is applied to a molecular dynamics (MD) simulation trajectory of the peptide system. The analysis proceeds through several stages: data preparation, peptide-level aggregate analysis, residue-level graph analysis within persistent aggregates, and correlation of residue-level dynamics with aggregate stability events.

2.1. *Data Preparation and System Definition*

The analysis utilizes a pre-computed molecular dynamics simulation trajectory of 30 K-Y-F-I-L pentapeptides. The simulation data consists of a trajectory file ('stripped.nc') containing atomic coordinates over time and a topology file ('stripped.parm7') describing the system's structure and parameters. The simulation was performed for a duration of approximately 500 ns, with frames saved every 20.0 ps. The total number of atoms in the system is derived from the topology file. Prior exploratory data analysis indicated that the system reaches a state of approximate equilibrium after the initial 100.0 ns of simulation time (Krajewski et al. 2024; Voit et al. 2024; Francis et al. 2025). Consequently, our analysis focuses on the trajectory segment from 100.0 ns onwards. Computational analysis was performed using Python, leveraging libraries such as `numpy` and `scipy`

for numerical operations (Faes 2018), and `networkx` for graph construction and analysis. Trajectory reading was handled using the `netCDF4` library, while topology parsing required custom scripting to extract atom-specific information including index, name, type, residue (name and number), and parent peptide ID (Casado & García 2024). A mapping was created to link atom indices to their respective peptide and residue identities within that peptide. Heavy (non-hydrogen) atoms were identified as the basis for contact calculations. Key parameters defined for the analysis are: the time per frame (20.0 ps), the equilibrium start time (100.0 ns), the number of peptides (30), the number of residues per peptide (5: K, Y, F, I, L), a contact distance threshold for peptide-peptide interactions (4.5 Å, based on heavy atom distance), a contact distance threshold for residue-residue interactions within aggregates (4.0 Å, based on heavy atom distance), and a persistence threshold for aggregates (250 frames, equivalent to 5 ns).

2.2. *Peptide-Level Aggregate Analysis*

This stage focuses on identifying and tracking peptide aggregates throughout the simulation trajectory from the defined equilibrium start frame.

2.2.1. *Peptide-Level Graph Construction*

For each simulation frame from the equilibrium start time onwards, a time-evolving peptide-level graph, $G_{\text{peptide},t}$, was constructed (Zheng et al. 2023). Nodes in this graph represent the individual peptides in the system. An undirected edge is added between two peptide nodes i and j if the minimum distance between any heavy atom belonging to peptide i and any heavy atom belonging to peptide j is less than or equal to the peptide-level contact distance threshold (4.5 Å). Minimum distances between sets of atoms were computed efficiently using the `scipy.spatial.distance.cdist` function. This process was parallelized across frames to enhance computational efficiency. The resulting adjacency matrices or edge lists for each frame's peptide graph were stored.

2.2.2. *Aggregate Identification and Tracking*

In each frame's peptide-level graph $G_{\text{peptide},t}$, connected components were identified using the `networkx.connected_components()` function. Each connected component represents a peptide aggregate at that specific time point, defined by peptides in contact with each other (Remijan et al. 2021; Zheng et al. 2023). An aggregate tracking algorithm was implemented to follow the evolution of specific aggregates over time (Oliver & Buck 2024; Zhuang et al. 2025). An aggregate identified in frame $t+1$ was considered a continuation of an aggregate from frame t if the Jaccard index of their constituent peptide sets exceeded a defined threshold (e.g., 0.7) (Zhuang et al. 2025). This method allows for tracking aggregate growth, shrinkage, merging, splitting, and dissolution (Oliver & Buck 2024; Zhuang et al. 2025). Each tracked aggregate lineage was assigned a unique persistent ID, and its lifespan (duration in frames) was recorded (Oliver & Buck 2024). Aggregates with a lifespan greater than or equal to the defined persistence threshold (250 frames) were identified as "persistent aggregates" for subsequent detailed analysis.

2.2.3. *Peptide-Level Graph Metrics*

For each frame from the equilibrium start, various graph theoretical metrics were calculated for the peptide-level graph $G_{\text{peptide},t}$. These metrics provide insights into the overall aggregation state and connectivity landscape (Yang & Yu 2023; García et al. 2024). The metrics computed included:

- **Algebraic Connectivity (λ_2):** The second smallest eigenvalue of the graph Laplacian matrix. This metric quantifies the robustness of the graph's connectivity; a higher value indicates a more strongly connected graph that is harder to break into components. For disconnected graphs, $\lambda_2 = 0$.
- **Fiedler Vector:** The eigenvector corresponding to the algebraic connectivity λ_2. Its components can provide insight into potential bottlenecks or partitions within the graph.
- **Density:** The ratio of the number of edges in the graph to the maximum possible number of edges. It indicates how close the graph is to being a complete graph.
- **Average Clustering Coefficient:** A measure of the degree to which nodes in a graph tend to cluster together. It quantifies the density of triangles in the graph.

These metrics were calculated for the entire peptide-level graph at each frame and stored as time series data.

2.3. *Residue-Level Graph Analysis for Persistent Aggregates*

This phase focuses on analyzing the internal contact network structure *within* the persistent aggregates identified in the previous stage (Shandarin & Yess 1997; Hong & Dey 2015).

2.3.1. *Residue-Level Graph Construction*

For each persistent aggregate, a detailed residue-level graph, $G_{\text{residue,agg},f}$, was constructed for every frame f in which that aggregate existed. Nodes in this graph represent the individual amino acid residues belonging to the peptides that constitute the aggregate at frame f. Nodes were uniquely identified by a combination of their parent peptide ID and their residue index within that peptide. An undirected edge was added between two residue nodes $r1$ and $r2$ if the minimum distance between any heavy atom of residue $r1$ and any heavy atom of residue $r2$ was less than or equal to the residue-level contact distance threshold (4.0 Å). These contacts could be between residues within the same peptide (intra-peptide) or between residues in different peptides within the aggregate (inter-peptide). This computationally intensive process was parallelized across persistent aggregates and frames. The resulting adjacency matrices or edge lists were stored, indexed by persistent aggregate ID and frame number.

2.3.2. *Residue-Level Graph Metrics*

For each constructed residue-level graph $G_{\text{residue,agg},f}$, a comprehensive set of graph theoretical metrics was calculated to characterize the internal network structure (Yang & Yu 2023; García et al. 2024).

- **Algebraic Connectivity ($\lambda_{2,\mathbf{res}}$):** The second smallest eigenvalue of the graph Laplacian of $G_{\text{residue,agg},f}$. This metric indicates the global connectivity robustness of the internal residue network within the aggregate.
- **Fiedler Vector ($F_{\mathbf{vec,res}}$):** The eigenvector corresponding to $\lambda_{2,\text{res}}$, providing insights into potential partitions within the residue network.
- **Centrality Measures:** Degree centrality was calculated for each residue node, measuring the number of direct contacts it forms within the aggregate network. Optionally, betweenness centrality was also considered for its ability to identify residues acting as bridges between different parts of the network, though its computation is more intensive.
- **Density ($\rho_{\mathbf{res}}$):** The density of the residue-level graph, indicating the overall fraction of possible residue-residue contacts that are formed within the aggregate.
- **Average Clustering Coefficient ($C_{\mathbf{res}}$):** The average of the local clustering coefficients for all residue nodes, measuring the tendency of residues to form tightly knit groups within the aggregate.

These metrics were calculated for each residue graph and stored as time series associated with the specific persistent aggregate and frame.

2.4. *Correlating Residue-Level Dynamics with Aggregate Stability*

The final stage involves integrating the information from the peptide-level aggregate tracking and the residue-level network analysis to identify links between microscopic network dynamics and macroscopic aggregate behavior.

2.4.1. *Analysis of Residue-Level Metrics Time Series*

For each persistent aggregate, time series plots were generated for the calculated residue-level graph metrics ($\lambda_{2,\text{res}}$, characteristics of $F_{\text{vec,res}}$, average centrality of specific residue types (Hong & Dey 2015), ρ_{res}, and C_{res}) over its lifespan. These plots were visually inspected to identify trends, significant fluctuations, or abrupt changes in the internal network structure (Shandarin & Yess 1997; Tsizh et al. 2020).

2.4.2. *Identification of Aggregate Stability Events*

The aggregate tracking data from the peptide-level analysis (Section 2.2) was revisited to identify specific events in the history of persistent aggregates that represent changes in their stability or structure (Dirkx et al. 2015; Verma 2022). These events included splitting (an aggregate breaking into smaller ones), merging (an aggregate combining with others), and dissolution (an aggregate disassembling). For each identified event, a pre-event time window (e.g., 50-100 frames) immediately preceding the event was defined (Dirkx et al. 2015).

2.4.3. *Correlation Analysis*

The time series of residue-level graph metrics for persistent aggregates were analyzed in relation to the identified stability events. Specifically, the behavior of metrics within the pre-event windows was compared to their behavior during periods where the aggregate was stable. The analysis focused on identifying patterns such as:

- Changes in $\lambda_{2,\text{res}}$ (e.g., significant drops or increased variance) preceding splitting or dissolution.
- Localization of Fiedler vector components onto specific residues or interfaces that subsequently break during splitting.
- Shifts in the centrality distribution, indicating changes in which residues play key roles in maintaining connectivity.

- Marked changes in residue-level density (ρ_{res}) or average clustering coefficient (C_{res}) around events like peripheral peptide association or dissociation.

Statistical comparisons (e.g., comparing mean metric values in pre-event vs. stable windows) were employed to quantify the significance of observed changes. Based on these correlations, residue-level graph signatures potentially indicative of aggregate stability or propensity for structural change were documented and formulated as potential predictive indicators within the proposed framework.

3. RESULTS

This study employed a hierarchical graph-based approach to investigate the self-assembly dynamics of the KYFIL pentapeptide, aiming to link residue-level network properties to the stability and behavior of peptide aggregates. The analysis focused on the equilibrium phase of a molecular dynamics simulation, starting from 100 ns, as detailed in the Methods section.

3.1. *Peptide-level aggregate dynamics*

The initial phase of the analysis involved processing the molecular dynamics trajectory and topology files to extract structural information and define the system. As described in the Methods (Section 2.1), the system consists of 30 KYFIL pentapeptides, totaling 3422 atoms. The analysis focused on the trajectory segment from 100 ns onwards, considered the equilibrium phase, encompassing 66771 frames with a time resolution of 20 ps per frame. Heavy atoms were identified as the basis for defining contacts, with each KYFIL peptide containing 48 heavy atoms, summing to 1225 heavy atoms for the entire system.

3.1.1. *Aggregate identification and tracking*

Peptide-level graphs were constructed for each frame within the equilibrium trajectory (Section 2.2.1). In these graphs, each node represented an individual peptide, and an edge was drawn between two nodes if the minimum distance between any heavy atoms of the corresponding peptides was less than or equal to the peptide-level contact threshold of 4.5 Å. Connected components within these graphs were identified as peptide aggregates at each time point.

An aggregate tracking algorithm (Section 2.2.2), utilizing a Jaccard index threshold of 0.7 for peptide set overlap between consecutive frames, was applied to follow the evolution of aggregates over time. This process identified 1742 unique aggregate lineages across the analyzed trajectory segment. Out of these, 28 lineages were classified as "persistent" based on a minimum lifespan threshold of 250 frames (5 ns). This finding underscores the dynamic nature of the self-assembly process, where numerous transient associations occur, yet a select few structures maintain their integrity over longer timescales. The relatively low number of persistent lineages compared to the total number of lineages suggests that the system spends significant time forming and breaking smaller, less stable clusters before or while sustaining larger, more enduring assemblies. These persistent aggregates were the focus of the subsequent residue-level analysis.

3.1.2. *Time evolution of peptide-level graph metrics*

To characterize the overall aggregation state and the structural evolution of the dominant assemblies, graph theoretical metrics were calculated for the largest aggregate present in each frame from 100 ns onwards (Section 2.2.3). The summary statistics for these metrics, averaged over the entire equilibrium phase, are presented in Table 1.

Table 1. Summary of Peptide-Level Graph Metrics for the Largest Aggregate (Mean ± Std Dev over 100 ns - end of simulation)

Metric	Mean	Std Dev
Largest Aggregate Size	28.89	2.72
Algebraic Connectivity (λ_2)	0.2030	0.1632
Graph Density	0.1485	0.0370
Avg. Clustering Coefficient	0.3062	0.0682

As shown in Table 1, the largest aggregate size averaged 28.89 peptides, indicating that throughout most of the simulation's equilibrium phase, nearly all 30 peptides are incorporated into a single dominant aggregate. The standard deviation of 2.72 indicates moderate fluctuations around this average size, suggesting dynamic exchange of peptides at the aggregate periphery. The time evolution of these metrics is illustrated in Figure 1.

The algebraic connectivity (λ_2) of the largest peptide-level aggregate averaged 0.2030 with a standard deviation of 0.1632 (Table 1). As λ_2 quantifies the robustness of graph connectivity and its resistance to partitioning, a non-zero average value indicates that the largest aggregate is generally well-connected at the peptide level. However, the considerable standard deviation and the fluctuations shown in Figure 1 suggest that the connectivity is not static and can vary significantly over time, possibly reflecting structural rearrangements or transient loosening of inter-peptide contacts.

The mean peptide-level graph density was 0.1485 (Table 1). Density represents the fraction of realized edges

out of all possible edges. For an aggregate of size N, the maximum number of edges is $N(N-1)/2$. A density of 0.1485 for an aggregate of ≈ 29 peptides means that each peptide is, on average, in contact with approximately $0.1485 \times (29-1) \approx 4.16$ other peptides. This relatively low density suggests that the aggregate structure is not maximally compact; peptides are connected to a subset of their potential neighbors, likely forming an extended or non-spherical structure, as also suggested by the fluctuations in density over time (Figure 1).

The average clustering coefficient at the peptide level was 0.3062 (Table 1). This metric measures the degree to which peptides in the aggregate tend to form triangles (i.e., if peptide A contacts B and C, how likely are B and C to contact each other?). A value of 0.3062 indicates a moderate tendency for local clustering among peptides, suggesting some degree of ordered local packing within the larger aggregate structure, beyond what would be expected in a random network of similar size and density. The time evolution (Figure 1) shows dynamic fluctuations in these metrics, with periods of decreased λ_2 potentially corresponding to transient structural loosening and increases in density or clustering indicating compaction or tighter packing.

3.2. *Residue-level network dynamics within persistent aggregates*

The second phase of the analysis focused on the internal structure of the 28 identified persistent peptide aggregates by constructing and analyzing residue-level contact graphs within them (Section 2.3). For each frame in which a persistent aggregate existed, a residue-level graph was built where nodes represented individual amino acid residues of the peptides within that aggregate, and edges represented heavy-atom contacts (≤ 4.0 Å) between any two residues (intra- or inter-peptide). This analysis covered a total of 62,255 aggregate-frame instances. Graph theoretical metrics were calculated for each of these residue-level graphs, and the summary statistics, averaged over all processed instances, are presented in Table 2.

As expected, the average number of residues per graph (144.24) is consistent with the average size of the parent peptide aggregates ($\approx$ 28.89 peptides $\times$ 5 residues/peptide $\approx$ 144.45 residues).

A central and highly significant finding from the residue-level analysis is the algebraic connectivity ($\lambda_{2,\text{res}}$), which consistently averaged 0.0000 with a standard deviation of 0.0000 across all analyzed persistent aggregate frames (Table 2). An algebraic connectivity of zero indicates that the graph is either disconnected into multiple components or is connected but contains

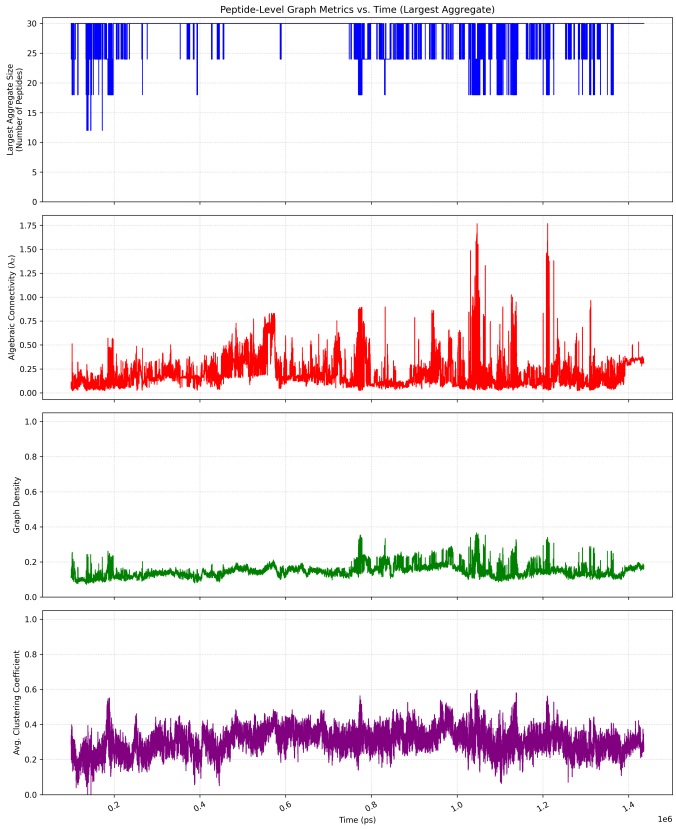

Figure 1. Time evolution of peptide-level graph metrics for the largest aggregate. The plots show the aggregate size (number of peptides), algebraic connectivity (λ_2), graph density, and average clustering coefficient over time (ps) during the simulation equilibrium phase, reflecting the dynamic nature and structural properties of the dominant peptide assembly.

Table 2. Summary of Residue-Level Graph Metrics for Persistent Aggregates (Mean ± Std Dev, averaged over all processed (aggregate, frame) pairs)

Metric	Mean	Std Dev
Number of Residues in Graph	144.24	17.32
Algebraic Connectivity ($\lambda_{2,\text{res}}$)	0.0000	0.0000
Graph Density (ρ_{res})	0.0272	0.0103
Avg. Clustering Coefficient (C_{res})	0.5733	0.0323
Avg. Degree Centrality (res)	0.0272	0.0103

articulation points (cut vertices) whose removal would disconnect it. This finding implies that, despite the persistence of the aggregate as a group of peptides at the macroscopic level, the underlying network of residue-residue contacts is not robustly connected across the entire aggregate structure. The internal residue network is fragmented or minimally connected, lacking strong links

that span the whole assembly. This is visually confirmed by the time series plots for individual aggregates, such as those shown in Figure 7 and Figure 8, where $\lambda_{2,\text{res}}$ remains at zero throughout.

The residue-level graph density (ρ_{res}) averaged 0.0272 (Table 2), which is a very low value. For a graph with an average of $\approx$ 144 nodes, this density corresponds to each residue being in contact with approximately $0.0272 \times (144 - 1) \approx 3.89$ other residues, on average. This low density further supports the notion of a sparse internal contact network within the aggregates. The average degree centrality, being directly proportional to density, reflects this same sparsity.

In stark contrast to the low density and zero algebraic connectivity, the average clustering coefficient (C_{res}) at the residue level was relatively high, averaging 0.5733 (Table 2). This indicates that while the overall residue network is sparse and globally disconnected, residues that are in contact with each other tend to form tightly knit local neighborhoods. This suggests that the aggregates are likely composed of numerous small, well-packed structural motifs (e.g., local regions of specific residue-residue interactions) that are themselves sparsely and possibly dynamically interconnected to form the larger, persistent aggregate. The time series plots (Figure 7, Figure 8) show that while $\lambda_{2,\text{res}}$ is constant at zero, ρ_{res} and C_{res} fluctuate over time, reflecting dynamic changes in these local packing arrangements.

3.3. *Correlation of residue-level dynamics with aggregate stability events*

To establish a link between the observed residue-level network dynamics and the macroscopic stability of the aggregates, specific events in the life cycle of persistent aggregates were identified and correlated with changes in residue-level graph metrics (Section 2.4). The event analysis (Section 2.4.2, based on peptide lineage tracking) for the 28 persistent aggregates revealed the following event types: 22 dissolution events (aggregate breaks down), 27 growth events (aggregate forms or significantly increases in size), 0 split events (aggregate breaks into two or more persistent sub-aggregates), and 0 merge-form events (persistent aggregate forms from two or more persistent precursors merging). The absence of split and merge-form events involving *persistent* aggregates suggests that once formed, these larger assemblies primarily evolve through the attachment or detachment of individual peptides or small, transient clusters at their periphery, rather than undergoing large-scale fragmentation or fusion with other substantial aggregates.

Statistical comparisons were performed for key residue-level metrics between "event windows" (a 50-frame window immediately preceding dissolution or following growth) and "stable windows" (periods where the aggregate size was relatively constant, Section 2.4.3). The results of these comparisons are shown in Figure 2, Figure 3, Figure 4, Figure 5, and Figure 6.

3.3.1. *Changes around dissolution events*

Comparing metrics in the 50 frames immediately preceding a dissolution event to those during stable periods revealed significant differences for some metrics. As illustrated in Figure 2, the number of residues in the aggregate was significantly lower in the pre-dissolution window (Mean 135.45) compared to stable periods (Mean 144.83, p=0.0), which is expected as the aggregate is shrinking before its complete breakdown.

However, the residue-level density (ρ_{res}) and average clustering coefficient (C_{res}) were found to be significantly *higher* in the pre-dissolution window (Mean ρ_{res}=0.0323, Mean C_{res}=0.589) compared to stable periods (Mean ρ_{res}=0.0270, Mean C_{res}=0.570, p=0.0 for both), as shown in Figure 3 and Figure 5. The average residue-level degree centrality also followed this trend (Figure 4). The residue-level algebraic connectivity ($\lambda_{2,\text{res}}$) remained effectively zero in both windows (Mean 0.0 vs 0.0, p=1.0), confirming the findings from the global analysis (Figure 6).

The observation of increased density and clustering in the shrinking aggregate just before dissolution is counter-intuitive if one expects a general loosening. It might suggest that as peripheral peptides detach, the remaining core transiently becomes more compact or reveals a more densely packed internal structure before further dissociation leads to complete dissolution. Alternatively, the "stable" state might include more loosely associated peptides that lower the overall average density and clustering, and their removal concentrates contacts within a tighter core.

3.3.2. *Changes around growth events*

For growth events, metrics in the 50 frames immediately following the event were compared to those in later stable periods. The number of residues was significantly lower in the post-growth window (Mean 135.1) compared to later stable periods (Mean 145.22, p=0.0), reflecting the aggregate growing into its larger, stable size (Figure 2). Similar to dissolution, both residue-level density (ρ_{res}) and average clustering coefficient (C_{res}) were significantly *higher* in the post-growth window (Mean ρ_{res}=0.0311, Mean C_{res}=0.5935) compared to later stable periods (Mean ρ_{res}=0.0268, Mean C_{res}=0.5716, p=0.0 for both), as seen in Figure 3 and

Figure 5. Average residue-level degree centrality also increased (Figure 4). Again, $\lambda_{2,\text{res}}$ remained zero in both windows (Mean 0.0 vs 0.0, p=1.0), as shown in Figure 6.

The higher density and clustering immediately after a growth event suggest that newly incorporated peptides rapidly establish a relatively high number of local contacts and participate in forming locally clustered structures. As the aggregate continues to grow and integrate further peptides into its larger, stable form, the overall average density and clustering might decrease slightly, potentially as new peptides occupy more peripheral or less constrained positions within the growing assembly.

Time series plots for individual persistent lineages, such as those shown in Figure 7 and Figure 8, visually support these findings, showing fluctuations in density and clustering around identified event frames while $\lambda_{2,\text{res}}$ remains consistently at zero.

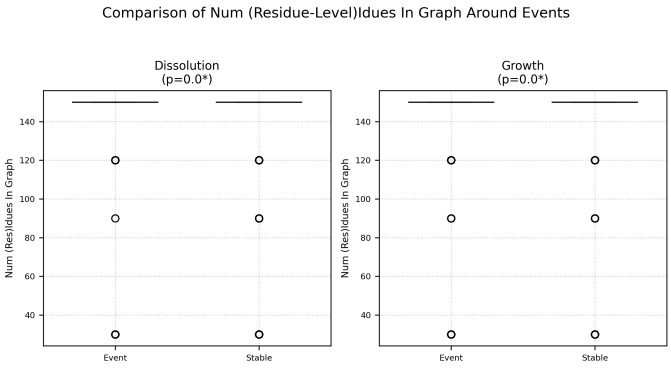

Figure 2. Box plots comparing the number of residues in residue-level graphs during event windows (pre-dissolution or post-growth) and stable periods for persistent aggregates. The number of residues is significantly lower in event windows ($p = 0.0$) for both dissolution and growth, reflecting the change in aggregate size during these events.

3.4. *Interpretation and implications*

The hierarchical graph-based analysis provided a multi-scale perspective on KYFIL self-assembly. The peptide-level analysis (Table 1, Figure 1) effectively captured the formation and persistence of a dominant large aggregate, characterizing its overall size, connectivity, and packing.

The residue-level analysis within these persistent aggregates (Table 2) yielded a critical insight: the internal network of residue-residue contacts consistently exhibited zero algebraic connectivity ($\lambda_{2,\text{res}} = 0$), as consistently shown in Figure 6, Figure 7, and Figure 8. This implies that the aggregate, while stable as a collection of peptides over nanoseconds, lacks a robust, globally connected network of residue contacts. Instead, it appears to be held together by numerous local interactions, forming many small, highly clustered regions (C_{res} is high, Table 2) that are only sparsely or tenuously linked to each other, leading to a fragmented overall residue network topology.

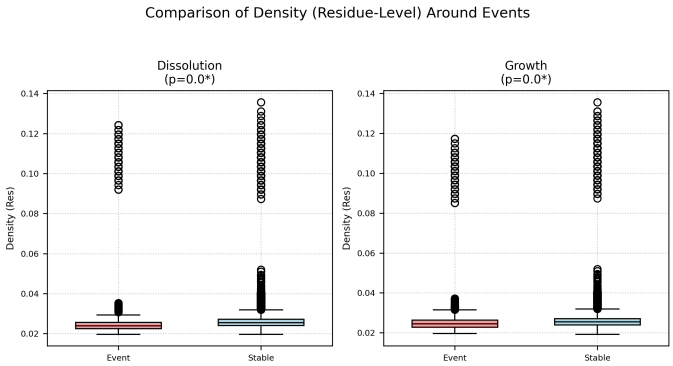

Figure 3. Comparison of residue-level graph density during aggregate dynamics. Box plots show the distribution of residue-level graph density for persistent aggregates in "Event" windows (pre-dissolution or post-growth, depending on the panel) and "Stable" windows. The left panel compares density in the window immediately preceding dissolution events to stable periods, showing significantly higher density in the event window (p=0.0). The right panel compares density in the window immediately following growth events to later stable periods, also showing significantly higher density in the event window (p=0.0). This indicates that the residue-level contact network is transiently more dense and locally ordered during periods of peptide association or dissociation compared to the overall stable aggregate state.

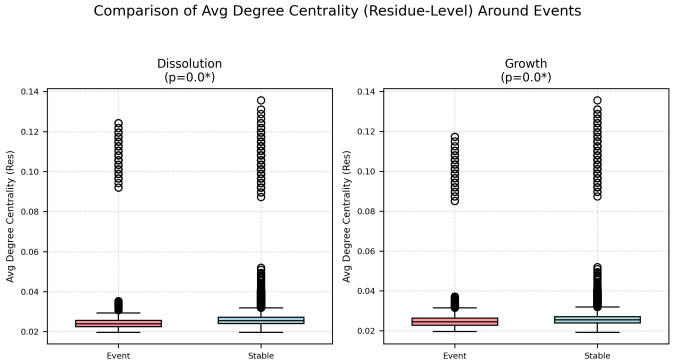

Figure 4. Box plots comparing the distribution of average residue-level degree centrality in persistent aggregates during event windows (pre-dissolution or post-growth) and stable periods. The average degree centrality is significantly higher in event windows compared to stable periods for both dissolution and growth events ($p = 0.0$), indicating changes in local residue packing around these events.

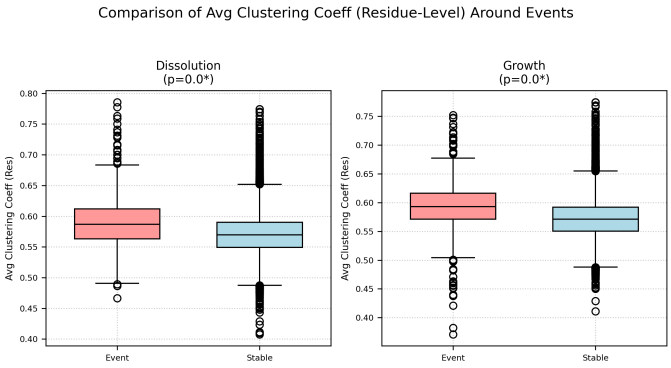

Figure 5. Box plots comparing the average residue-level clustering coefficient in windows preceding aggregate dissolution events or following aggregate growth events to stable periods. The average residue-level clustering coefficient is significantly higher ($p = 0.0$) in these event-associated windows, suggesting altered local residue packing around aggregate formation and breakdown.

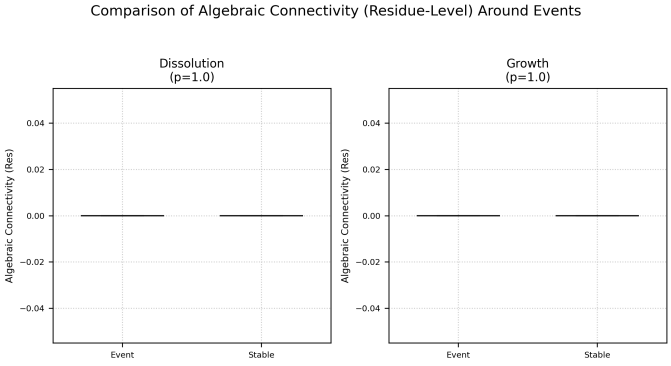

Figure 6. Comparison of residue-level algebraic connectivity ($\lambda_{2,\mathrm{res}}$) around dissolution and growth events in persistent aggregates. Box plots show that $\lambda_{2,\mathrm{res}}$ remains zero in both event and stable periods, indicating the consistently disconnected nature of the residue contact network within aggregates and its lack of change around these events.

This finding has significant implications for understanding and predicting aggregate stability in this system. The consistently zero value of $\lambda_{2,\mathrm{res}}$ means that changes in this metric cannot serve as a direct precursor or predictor for aggregate splitting or dissolution. The aggregate is already in a state of minimal global connectivity at the residue level, making it difficult to detect impending instability through changes in this specific global metric.

However, the analysis revealed that residue-level density (ρ_{res}) and average clustering coefficient (C_{res}) are sensitive indicators of changes associated with aggregate dynamics. Their significant increase in magnitude around both dissolution and growth events (Figure 3,

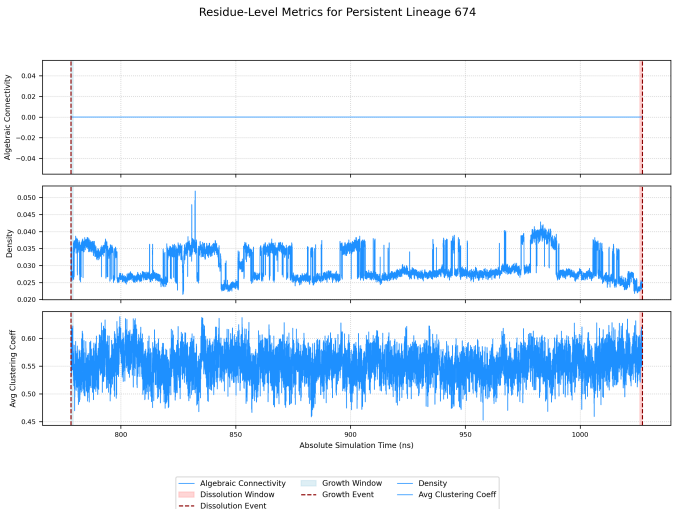

Figure 7. Time series of residue-level graph metrics for a persistent peptide aggregate lineage. The panels show the algebraic connectivity (λ_2), graph density, and average clustering coefficient, plotted against absolute simulation time. Aggregate dissolution events are indicated by vertical dashed lines, with shaded regions representing the corresponding event windows. The consistently zero algebraic connectivity highlights the fragmented nature of the internal residue contact network within the aggregate, while density and average clustering coefficient fluctuate over time and show changes around dissolution events.

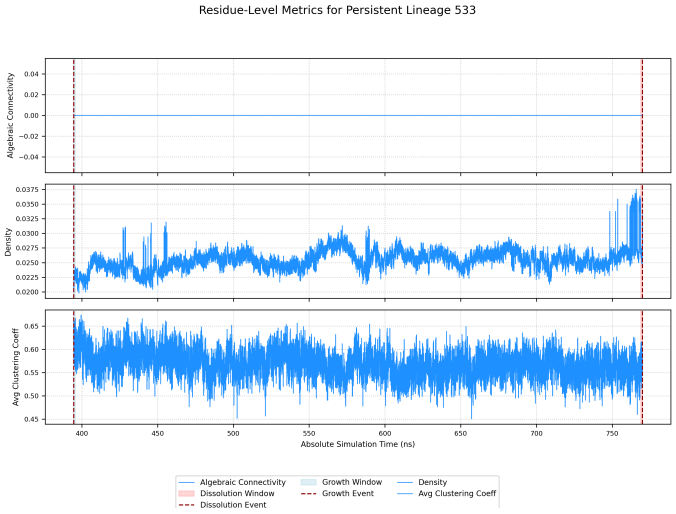

Figure 8. Time evolution of residue-level graph metrics for persistent aggregate lineage 533. The panels show the algebraic connectivity (top), graph density (middle), and average clustering coefficient (bottom) over simulation time. Vertical dashed lines indicate a dissolution event. The consistently zero algebraic connectivity demonstrates that the residue-level contact network within this persistent aggregate lacks robust global connectivity, while fluctuations in density and clustering reflect dynamic changes in local residue packing.

Figure 5) suggests that these metrics respond to the processes of peptide association and dissociation at the aggregate boundary. Specifically, the transient increase in density and clustering before dissolution might reflect a compaction of the core as peripheral peptides are lost, while the higher values immediately after growth suggest rapid local ordering upon peptide incorporation. These local metrics, unlike global connectivity, appear to capture structural changes relevant to the aggregate's dynamic evolution.

The study demonstrates that aggregate stability in this system does not rely on a single, rigid, globally interconnected residue network but rather on the collective effect of dynamic local interactions. This understanding is crucial for interpreting simulation data and for guiding future peptide design efforts, suggesting that modifying sequences to promote specific local packing motifs or to enhance global connectivity might lead to aggregates with different structural robustness or dynamic properties.

4. CONCLUSIONS

In this study, we addressed the fundamental challenge of linking microscopic residue-level interactions to macroscopic aggregate stability in self-assembling peptides using a novel hierarchical graph-based analysis framework. Leveraging a molecular dynamics simulation trajectory of the KYFIL pentapeptide, our approach involved constructing and analyzing time-evolving graphs at two distinct levels: a peptide-level graph to track aggregate formation and persistence, and detailed residue-level contact graphs within identified persistent aggregates.

The analysis of the peptide-level graphs effectively captured the dynamic self-assembly process, revealing the formation and persistence of a dominant large aggregate comprising nearly all peptides in the system during the equilibrium phase. Metrics at this level, such as non-zero algebraic connectivity, confirmed the overall connectivity of the aggregate as a collection of peptides.

Delving into the internal structure using residue-level graphs within persistent aggregates yielded key insights. Despite the macroscopic persistence of the aggregates, the residue-level contact network consistently exhibited zero algebraic connectivity ($\lambda_{2,\mathrm{res}} = 0$). This finding indicates that the aggregate's internal structure is not characterized by a single, robustly connected network spanning the entire assembly. Instead, it appears to be composed of multiple, potentially transiently connected components or contains critical bottlenecks at the residue level. This lack of global connectivity at the residue level contrasts sharply with a relatively high average clustering coefficient (C_{res}), suggesting that while globally sparse, the network is rich in local, tightly knit clusters of residues.

The correlation analysis between residue-level network dynamics and observed aggregate stability events (growth and dissolution) further illuminated the nature of stability in this system. The consistently zero $\lambda_{2,\mathrm{res}}$ meant this global connectivity metric did not serve as a precursor to instability. However, residue-level density (ρ_{res}) and average clustering coefficient (C_{res}) were found to change significantly around these events, showing increased values in the windows immediately preceding dissolution or following growth. This suggests that these local metrics are sensitive to the dynamic processes of peptide association and dissociation occurring at the aggregate periphery, reflecting transient changes in local packing or contact formation.

Collectively, our results indicate that the stability of KYFIL aggregates, at least on the observed nanosecond timescale, does not arise from a rigid, globally interconnected residue network. Instead, it appears to be sustained by a collection of dynamic local interactions and the formation of numerous small, well-packed structural motifs that are only loosely or transiently linked together. This finding limits the direct applicability of global connectivity metrics like algebraic connectivity for predicting aggregate instability in such systems, while highlighting the potential of residue-level density and clustering as indicators of local structural changes associated with aggregate growth and dissolution. This hierarchical graph-based framework provides a powerful tool for dissecting the multi-scale dynamics of peptide self-assembly, offering a chemically-informed perspective crucial for understanding and ultimately controlling the formation and stability of peptide-based materials.

REFERENCES

Casado, J., & García, B. 2024, Multimodal photometry and spectroscopy: a new approach to data analysis in Astrophysics. https://arxiv.org/abs/2404.17720

Dirkx, D., Noomen, R., Visser, P., Gurvits, L., & Vermeersen, B. 2015, Space-time Dynamics Estimation from Space Mission Tracking Data, doi: https://doi.org/10.1051/0004-6361/201527524

Faes, D. M. 2018, Use of Python programming language in astronomy and science, doi: https://doi.org/10.6062/jcis.2012.03.03.0063

Francis, A., Williams, L. L. R., & Hjorth, J. 2025, Entropy Evolution Towards DARKexp of IllustrisTNG Dark Matter Halos. https://arxiv.org/abs/2502.15880

García, C. R., Torres, D. F., Zhu-Ge, J.-M., & Zhang, B. 2024, Separating repeating fast radio bursts using the minimum spanning tree as an unsupervised methodology. https://arxiv.org/abs/2411.02216

Hong, S., & Dey, A. 2015, Network Analysis of Cosmic Structures : Network Centrality and Topological Environment, doi: https://doi.org/10.1093/mnras/stv722

Horowitz, C. J., & Hughto, J. 2008, Molecular Dynamics Simulation of Shear Moduli for Coulomb Crystals. https://arxiv.org/abs/0812.2650

Krajewski, T., Lewicki, M., & Zych, M. 2024, Bubble-wall velocity in local thermal equilibrium: hydrodynamical simulations vs analytical treatment, doi: https://doi.org/10.1007/JHEP05(2024)011

Makinen, T. L., Alsing, J., & Wandelt, B. D. 2024, Fishnets: Information-Optimal, Scalable Aggregation for Sets and Graphs. https://arxiv.org/abs/2310.03812

Oliver, W. H., & Buck, T. 2024, Galaxy Formation and Evolution via Phase-temporal Clustering with FuzzyCat ∘ AstroLink. https://arxiv.org/abs/2411.03229

Pandya, S., Patel, P., Nord, B. D., Walmsley, M., & Ćiprijanović, A. 2025, SIDDA: SInkhorn Dynamic Domain Adaptation for Image Classification with Equivariant Neural Networks. https://arxiv.org/abs/2501.14048

Pavlou, O., Michos, I., Lesta, V. P., et al. 2023, Graph Theoretical Analysis of local ultraluminous infrared galaxies and quasars, doi: https://doi.org/10.1016/j.ascom.2023.100742

Remijan, A., Xue, C., Margulès, L., et al. 2021, Expanding the submillimeter wave spectroscopy and astronomical search for thioacetamide (CH3CSNH2) in the ISM, doi: https://doi.org/10.1051/0004-6361/202142504

Serjeant, S., Pearson, J., Dickinson, H., & Jarvis, J. 2024, Citizen Science in European Research Infrastructures. https://arxiv.org/abs/2404.18635

Shandarin, S. F., & Yess, C. 1997, Detection of Network Structure in the Las Campanas Redshift Survey, doi: https://doi.org/10.1086/306135

Strey, S.-G., Castronovo, A., & Elumalai, K. 2024, Graph Theoretic Approach Identifies Critical Thresholds at which Galaxy Filamentary Structures Form. https://arxiv.org/abs/2310.20184

Tsizh, M., Novosyadlyj, B., Holovatch, Y., & Libeskind, N. I. 2020, Large-scale structures in the ΛCDM Universe: network analysis and machine learning, doi: https://doi.org/10.1093/mnras/staa1030

Verma, A. K. 2022, A Python-based tool for constructing observables from the DSN's closed-loop archival tracking data files. https://arxiv.org/abs/2208.03865

Voit, G. M., Carr, C., Fielding, D. B., et al. 2024, Equilibrium States of Galactic Atmospheres II: Interpretation and Implications. https://arxiv.org/abs/2406.07632

Yang, D., & Yu, H.-B. 2023, A graph model for the clustering of dark matter halos, doi: https://doi.org/10.1103/PhysRevResearch.5.043187

Zheng, S., Li, J., Wang, J., et al. 2023, Mapping Observations of Peptide-like molecules around Sagittarius B2. https://arxiv.org/abs/2310.16664

Zhuang, G., Song, W., Huang, J., et al. 2025, High Performance Space Debris Tracking in Complex Skylight Backgrounds with a Large-Scale Dataset. https://arxiv.org/abs/2506.02614