Dynamic Multiscale Graph Analysis Reveals Structural Signatures of Peptide Aggregate Stability and Splitting

AstroPilot[1]

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

ABSTRACT

Understanding the structure, dynamics, and stability of peptide aggregates formed during self-assembly is crucial for designing functional biomaterials. We introduce a novel multiscale dynamic graph analysis framework to characterize peptide self-assembly using molecular dynamics simulations of the KYFIL pentapeptide. Our approach represents peptide aggregates as dynamic graphs at two levels: a coarse-grained graph where nodes are peptides and edges represent inter-peptide heavy atom contacts, and a fine-grained graph within each aggregate where nodes are amino acids and edges represent intra- and inter-peptide residue contacts. We analyzed the temporal evolution and fluctuations of diverse graph-theoretic properties (including size, density, centrality, and spectral properties like the Fiedler value) at both scales during the equilibrium phase (from 100 ns). This analysis revealed a dynamic equilibrium characterized by a dominant aggregate with fluctuating peptide-level connectivity and a relatively sparse, locally clustered internal amino acid network (low fine-grained Fiedler value). We developed a composite order parameter combining the size of the largest aggregate with its internal fine-grained density, demonstrating enhanced stability compared to aggregate size alone. Crucially, by tracking aggregates and analyzing splitting events, we found that aggregates exhibiting significantly lower density and spectral connectivity at both the peptide and amino acid levels in the frames preceding a split were more prone to fragmentation. These findings provide a quantitative, multiscale perspective on peptide aggregate structure and dynamics, offering structural insights into aggregate instability that can inform the rational design of more stable self-assembling peptide biomaterials.

1. INTRODUCTION

Peptide self-assembly, the spontaneous organization of peptides into ordered structures, is a fundamental process in biological systems and a powerful strategy for the rational design of functional biomaterials. These self-assembled structures span a wide range of morphologies, from transient oligomers to extensive fibrils and hydrogels, exhibiting diverse structural motifs and physical properties. A comprehensive understanding of the molecular mechanisms governing the formation, intricate structure, dynamic evolution, and ultimately, the stability of these aggregates is paramount for controlling their functional outcomes in applications such as drug delivery, tissue engineering, and for addressing protein aggregation-related diseases.

However, characterizing the complex and dynamic nature of peptide aggregates at a molecular level poses significant challenges. Aggregates are inherently heterogeneous and undergo continuous structural rearrangements, growth through monomer addition or aggregate merging, and fragmentation or dissolution. Experimental techniques often provide ensemble-averaged information or static snapshots, lacking the detailed, dynamic atomic-level insight. Molecular dynamics (MD) simulations offer atomic-resolution trajectories of self-assembly, but extracting meaningful, quantitative measures that capture the complex internal connectivity and its temporal changes, and critically, relating these measures to aggregate stability and dynamic events like splitting or merging, remains a major hurdle. Traditional MD analysis often focuses on macroscopic properties such as aggregate size or shape, or on simple contact counts, but lacks a comprehensive framework to describe the system's structure across different organizational scales simultaneously and dynamically.

To overcome this challenge, we introduce a novel framework based on dynamic multiscale graph analy-

sis applied to molecular dynamics simulations of peptide self-assembly. Graph theory provides a powerful mathematical language to represent complex networks of interactions and quantify their structural properties beyond simple proximity (Pavlou et al. 2023; Sun et al. 2024). By applying graph-theoretic concepts to peptide aggregates, we can move beyond simple distance cutoffs or aggregate size to characterize the dynamic network of contacts at different levels of resolution (Sun et al. 2024).

Our approach represents peptide aggregates as dynamic graphs at two distinct scales: a coarse-grained (CG) level where each node represents an entire peptide and edges denote inter-peptide heavy atom contacts, and a fine-grained (FG) level within each aggregate where nodes are individual amino acids and edges represent intra- or inter-peptide residue contacts. This multiscale representation allows us to simultaneously probe the peptide-level organization and the detailed internal amino acid packing and interaction network within aggregates (Sun et al. 2024).

In this work, we leverage extensive molecular dynamics simulations of the self-assembling pentapeptide KY-FIL to demonstrate the utility and power of this multiscale dynamic graph analysis framework. We analyze the temporal evolution and fluctuations of a comprehensive set of graph-theoretic properties at both the CG and FG scales during the equilibrium phase of self-assembly (from 100 ns onwards). These properties include standard measures like size, density, and centrality, as well as spectral properties derived from the Graph Laplacian, such as the Fiedler value (algebraic connectivity), which provides insights into graph fragmentation. By tracking aggregates over time and identifying specific dynamic events, particularly splitting events, we investigate how distinct multiscale structural features correlate with aggregate stability and fragmentation proneness. We extract graph properties in the frames preceding a split and compare them to stable aggregates to identify structural signatures of instability. Furthermore, we explore the development of a more robust order parameter for aggregation by combining information from both scales, such as aggregate size from the CG graph and internal density or spectral connectivity from the FG graph, and validate its enhanced stability compared to size alone.

Through this rigorous, quantitative analysis, we aim to reveal structural signatures at both the peptide and amino acid levels that dictate aggregate stability and dynamic behavior. This work establishes a novel, quantitative, and multiscale perspective on peptide aggregate structure and dynamics, providing fundamental insights that can inform the rational design of peptide sequences with tailored assembly and stability properties for a wide range of biomaterial applications.

2. METHODS

2.1. *Molecular Dynamics Simulation Data*

The dataset for this study consists of a 500 ns molecular dynamics (MD) simulation trajectory of 30 KYFIL pentapeptides in explicit water solvent, previously generated using standard simulation protocols. The simulation data included atomic coordinates, velocities, and topology information. The trajectory and topology files, named `stripped.nc` and `stripped.parm7` respectively, were loaded using the `MDAnalysis` Python library for initial processing and data extraction. All subsequent analyses were performed using custom scripts leveraging `NumPy`, `SciPy`, and graph analysis libraries.

2.2. *Data Preparation and Equilibrium Identification*

To ensure our analysis focuses on the system's behavior at or near equilibrium (Wang et al. 2015; Voit et al. 2024), we first performed exploratory data analysis (EDA) on the full 500 ns trajectory.

2.2.1. *Loading and Trajectory Segmentation*

The simulation trajectory was loaded, and per-frame atomic coordinates, atom types, residue identities, and peptide segment information were extracted. Based on preliminary analysis (detailed below), the system reached a stable state after approximately 100 ns. Consequently, all subsequent detailed graph analyses were performed exclusively on the trajectory segment starting from 100 ns up to the end of the simulation (500 ns). The frame corresponding to the 100 ns mark was identified by examining the time attribute of the trajectory frames.

2.2.2. *Exploratory Data Analysis for Equilibration and Contact Definition*

To confirm system equilibration and determine a suitable distance cutoff for defining inter-atomic contacts, we performed the following analyses:

Root Mean Square Deviation (RMSD):—The RMSD of all heavy atoms (Carbon, Nitrogen, Oxygen) across all 30 peptides was calculated relative to the heavy atom coordinates of the first frame (0 ns) as a reference structure. For each subsequent frame, the heavy atom coordinates were aligned to the reference using the Kabsch algorithm (implemented using `SciPy`'s capabilities) to minimize translational and rotational differences before calculating the RMSD. The time evolution of the heavy atom RMSD was plotted to assess global system stability. A plateau in RMSD over time indicates that the

system has reached a relatively stable conformational ensemble, supporting the choice of the equilibrium analysis window.

Radius of Gyration (Rg):—The radius of gyration was calculated for the entire system (all heavy atoms of all 30 peptides) and averaged over individual peptides (heavy atoms of each peptide) for every frame of the full trajectory (Wadhwa et al. 2024,?). The system Rg reflects the overall compactness of the peptide ensemble and aggregate formation (Claret 2023), while the average individual peptide Rg indicates changes in peptide conformation. Plotting Rg values over time provided further insights into the system's dynamic state and supported the identification of the equilibrium phase marked by a plateau in system Rg (Li et al. 2021).

Determination of Heavy Atom Contact Cutoff:—To define the edges in our graph representations, we needed a consistent distance cutoff for heavy atom contacts. We calculated the pairwise distances between all non-bonded heavy atoms belonging to *different* peptides for a representative set of frames sampled from the equilibrium phase (after 100 ns). A Radial Distribution Function (RDF), g(r), was computed for these inter-peptide heavy atom pairs (Pan et al. 2011; Dong et al. 2025). The first minimum after the first peak in the RDF, which typically corresponds to the boundary of the first solvation shell, was used to define the maximum distance at which two heavy atoms are considered to be in "contact". Based on this analysis, a cutoff distance of 4.0 Å was determined and used consistently throughout the study for defining heavy atom contacts.

2.3. *Dynamic Multiscale Graph Construction*

For each frame within the equilibrium trajectory segment (from 100 ns to 500 ns), we constructed dynamic graph representations of the peptide system at two distinct scales: a coarse-grained (CG) peptide level and a fine-grained (FG) amino acid level.

2.3.1. *Coarse-Grained Peptide-Level Graph Construction*

At the coarse-grained level, each of the 30 KYFIL peptides was represented as a single node in the graph. An undirected edge was placed between two peptide nodes (Peptide i and Peptide j) if the minimum distance between any heavy atom in Peptide i and any heavy atom in Peptide j was less than or equal to the predetermined contact cutoff distance of 4.0 Å. This was implemented by extracting the heavy atom coordinates for each peptide in a given frame and computing the pairwise distance matrix between all heavy atoms of Peptide i and Peptide j using `scipy.spatial.distance.cdist`.

If any distance within this matrix was found to be less than or equal to 4.0 Å, an edge was added between nodes i and j. The resulting graph for each frame was stored as an adjacency matrix.

2.3.2. *Fine-Grained Amino-Acid-Level Graph Construction*

Within each identified peptide aggregate from the CG graph, a fine-grained graph was constructed to represent the internal connectivity at the amino acid level.

Aggregate Identification:—For each frame, connected components in the CG peptide graph were identified. Each connected component represents a distinct peptide aggregate.

FG Graph Construction per Aggregate:—For each identified aggregate, a new graph was constructed. The nodes of this graph were the individual amino acids (Lysine, Tyrosine, Phenylalanine, Isoleucine, Leucine) belonging to the peptides within that specific aggregate. An undirected edge was placed between two amino acid nodes (Amino Acid A and Amino Acid B) if the minimum distance between any heavy atom of Amino Acid A and any heavy atom of Amino Acid B was less than or equal to the 4.0 Å contact cutoff. This definition included both intra-peptide contacts (between amino acids within the same peptide) and inter-peptide contacts (between amino acids belonging to different peptides but within the same aggregate). The heavy atom coordinates for all amino acids within the aggregate were extracted, and pairwise distances between heavy atoms of potentially interacting amino acid pairs were calculated to determine edge presence.

2.4. *Calculation of Graph-Theoretic Properties*

For each frame in the equilibrium trajectory, a comprehensive set of graph-theoretic properties was calculated for both the CG peptide graph and the FG amino acid graphs (one for each aggregate) (Pavlou et al. 2023; Strey et al. 2024; Sun et al. 2024).

2.4.1. *Properties for CG Peptide Graphs*

For the CG graph representing the entire system of 30 peptides, we calculated: the number of nodes (fixed at 30), the total number of edges, the graph density (ratio of existing edges to maximum possible edges), the number of connected components (representing the number of distinct aggregates) (Airale et al. 2025,?; Coupette et al. 2025), and the size (number of peptide nodes) of each connected component. For each peptide node, its degree (number of direct peptide neighbors) was calculated. For the largest connected component (the dominant aggregate), we calculated its average clustering coefficient, which measures the local density of connections

around nodes. We also computed the Graph Laplacian matrix ($L = D - A$, where D is the diagonal degree matrix and A is the adjacency matrix) for the largest component. The eigenvalues of the Laplacian were calculated using `scipy.linalg.eigh` for symmetric matrices. The second smallest eigenvalue (λ_2), also known as the algebraic connectivity or Fiedler value, was extracted as a measure of graph connectivity and robustness to fragmentation. The corresponding eigenvector (Fiedler vector) was also computed. Additional centrality measures, such as betweenness, closeness, and eigenvector centrality, were calculated for each peptide node within the largest aggregate using `NetworkX`.

2.4.2. *Properties for FG Amino-Acid Graphs*

For each fine-grained graph representing an individual aggregate, we calculated: the number of nodes (total amino acids in the aggregate), the number of edges (total intra-aggregate amino acid contacts), and the graph density (reflecting the packing efficiency within the aggregate) (Zhang et al. 2025). For each amino acid node, its degree was calculated. We computed the average clustering coefficient for the FG graph. Similar to the CG graph, the Graph Laplacian and its eigenvalues, including the Fiedler value (λ_2), were calculated for each FG graph using `NumPy` and `scipy.linalg.eigh` (Makinen et al. 2024). Centrality measures were also computed for amino acid nodes (Zhang et al. 2025).

2.4.3. *Implementation Details for Graph Analysis*

Graph objects were constructed and manipulated using either the `NetworkX` or `igraph_python` libraries, which provide efficient algorithms for calculating standard graph metrics like density, clustering coefficients, and centrality measures. Spectral properties (Laplacian, eigenvalues, eigenvectors) were computed using `NumPy` for matrix operations and `scipy.linalg.eigh` for efficient eigenvalue decomposition of the symmetric Laplacian matrix (Eldén 2023).

2.5. *Temporal Analysis of Graph Properties*

The calculated graph properties for the CG and FG graphs were recorded for each frame from 100 ns to 500 ns, generating time series for each property (e.g., CG graph density vs. time, FG λ_2 of the largest aggregate vs. time) (Yang & Yu 2023). For each time series, we calculated the mean value, standard deviation (as a measure of fluctuation), and coefficient of variation (standard deviation divided by the mean) over the equilibrium time window (Roncoli et al. 2024). Histograms were generated to visualize the distributions of key properties, such as aggregate sizes and FG densities, over the

equilibrium phase (Yang & Yu 2023). The time evolution of selected properties was plotted to identify trends, dynamic fluctuations, and transitions (Yang & Yu 2023).

2.6. *Development of a Stable Aggregation Order Parameter*

Recognizing that simple aggregate size can be highly dynamic, we aimed to develop a more stable order parameter for characterizing the aggregated state by combining information from both the CG and FG scales. Candidate internal stability measures derived from the FG graph properties of aggregates were considered, including the average Fiedler value (λ_2), average density, and average clustering coefficient of the FG graph(s) within an aggregate. We proposed composite order parameters (OPs) that combine the size of the largest CG aggregate with a measure of its internal FG structural integrity. Examples included OPs defined as the product of the size of the largest CG aggregate and the FG λ_2 or density of that same aggregate. A global OP was also considered, defined as the sum of the products of aggregate size and its FG property (e.g., λ_2), normalized by the total number of peptides. The time evolution of these proposed OPs was analyzed, and their stability was quantitatively compared to that of simpler measures like the size of the largest aggregate alone, primarily using the coefficient of variation over the equilibrium time window.

2.7. *Correlation of Multiscale Graph Features with Aggregate Dynamics*

To understand the relationship between multiscale structural features and aggregate stability, we tracked the dynamics of individual peptide aggregates over time.

2.7.1. *Identification of Aggregate Events*

Aggregates (connected components in the CG graph) were tracked across consecutive frames by identifying maximum overlap in their constituent peptide lists. This allowed us to follow the "life history" of individual aggregates. Specific dynamic events were identified and timestamped: formation (appearance of a new aggregate), dissolution (disappearance of an aggregate), splitting (an aggregate at frame t fragments into two or more smaller aggregates at frame $t+1$), and merging (two or more aggregates at frame t combine into a single larger aggregate at frame $t+1$). The longevity of aggregates was defined as the duration from their formation to their dissolution or splitting.

2.7.2. *Feature Extraction and Correlation Analysis*

For each identified aggregate instance that persisted for a minimum duration, we calculated the average val-

ues of its key CG and FG graph properties (e.g., CG λ_2, FG density, FG λ_2) over its lifetime. To investigate the structural signatures preceding aggregate fragmentation, we specifically focused on splitting events. For each aggregate that underwent a split, we extracted its CG and FG graph properties within a defined time window immediately preceding the split frame. These properties were then compared to the properties of aggregates that remained stable throughout the same time window or for a significantly longer duration. Statistical comparisons (e.g., t-tests or Mann-Whitney U tests) were performed to identify properties that were significantly different in aggregates prone to splitting compared to stable ones. Correlation analysis (e.g., Pearson or Spearman correlation coefficients) was used to quantify the relationship between aggregate longevity and its average multiscale graph properties.

2.8. *Computational Workflow and Data Management*

The entire analysis pipeline was implemented using Python 3.x (Parsotan et al. 2023, 2025). Core libraries included `NumPy` for numerical computations and array manipulation, and `SciPy` for spatial distance calculations, sparse matrices, and advanced linear algebra routines, particularly eigenvalue decomposition. Graph construction and property calculation relied on `NetworkX` or `igraph_python`. Data visualization was performed using `Matplotlib` and `Seaborn` (Deal & Espinoza 2024). Given the frame-by-frame nature of the analysis, the computational workload was parallelized across the available 128 CPU cores using Python's `multiprocessing` or `joblib` library. The loop over trajectory frames was distributed among worker processes, with each process handling a batch of frames independently. Calculated time series data and summary statistics were saved in structured formats such as CSV or HDF5 for efficient storage and subsequent analysis (Parsotan et al. 2023, 2025). Comprehensive logging was implemented using print statements to track the progress of the analysis, report completion of key stages, and output essential summary statistics and shapes of generated data arrays for verification.

3. RESULTS

Our analysis employed a novel multiscale dynamic graph approach to investigate the structural and dynamic properties of KYFIL pentapeptide aggregates during the equilibrium phase of a 500 ns molecular dynamics simulation. The framework allowed us to represent peptide aggregates as dynamic networks at both the coarse-grained (peptide) and fine-grained (amino acid) levels, providing a quantitative means to characterize their evolving structure and relate it to dynamic events.

3.1. *System equilibration and simulation setup*

Prior to detailed graph analysis, we assessed the simulation trajectory for equilibration. The Root Mean Square Deviation (RMSD) of heavy atoms, calculated relative to the initial frame, showed an initial rise followed by a plateauing behavior, as depicted in Figure 1. The standard deviation of the RMSD significantly decreased after approximately 100 ns (from 9.496 Å in the first 100 ns to 5.080 Å in the subsequent 400 ns), indicating that the system reached a relatively stable conformational ensemble. Similarly, the system's Radius of Gyration (Rg), reflecting the overall compactness, also stabilized after 100 ns, as shown in Figure 2. While the standard deviation of the system Rg increased slightly in the latter phase, this is consistent with the expected dynamic fluctuations of large aggregates forming and rearranging in equilibrium. Based on these indicators, the trajectory segment from 100 ns–500 ns was identified as the equilibrium phase and used for all subsequent analyses, encompassing 66,772 frames.
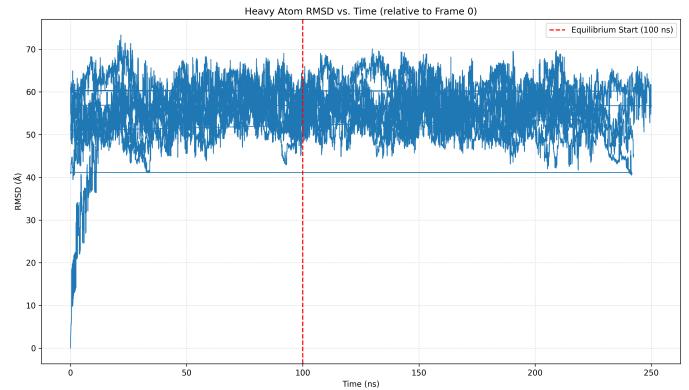

Figure 1. Heavy atom Root Mean Square Deviation (RMSD) of the entire system relative to the initial frame (Frame 0) as a function of simulation time. The red dashed line at 100 ns marks the defined start of the equilibrium phase. The plot shows that the RMSD stabilizes after approximately 100 ns, indicating system equilibration, while continued fluctuations in the equilibrium phase reflect the dynamic nature of aggregate formation and rearrangement.

Defining inter-atomic contacts is crucial for graph construction. We attempted to determine an optimal heavy atom contact cutoff by analyzing the Radial Distribution Function (RDF) of inter-peptide heavy atom pairs. However, this analysis did not yield a distinct first solvation shell minimum. Consequently, based on common practice in molecular simulations for defining nonbonded contacts, a standard cutoff distance of 4.0 Å was adopted for all graph constructions. Peptide entities were successfully defined as contiguous blocks of 5
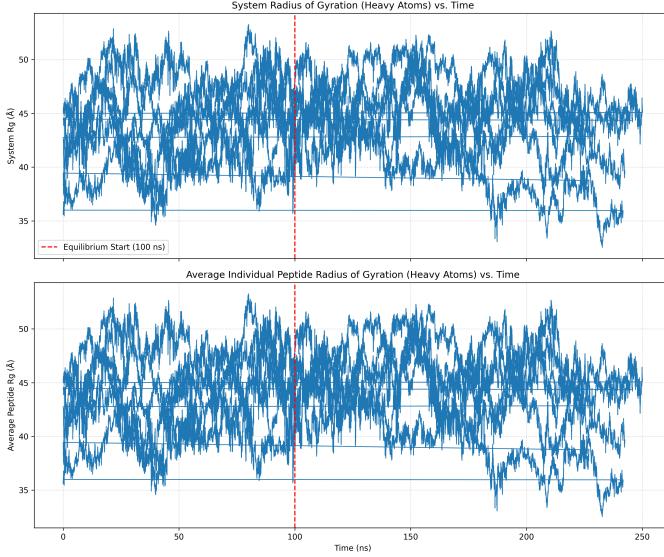
Figure 2. Radius of gyration over time for the molecular dynamics simulation of KYFIL pentapeptide self-assembly. Top panel shows the system radius of gyration (Rg) of heavy atoms. Bottom panel shows the average radius of gyration for individual peptides (heavy atoms). The dashed red line at 100 ns indicates the start of the equilibrium phase. The system Rg stabilizes overall but exhibits increased fluctuations after 100 ns, reflecting dynamic aggregate formation and rearrangement, while the average individual peptide Rg remains relatively constant, supporting the designation of 100 ns as the onset of equilibrium.

residues, ensuring the correct identification of 30 peptides for the coarse-grained analysis.

3.2. *Multiscale graph representation of peptide aggregates*

For each frame within the equilibrium phase, we constructed dynamic graphs at two levels of resolution, as described in the Methods. The coarse-grained (CG) graph represented the system of 30 peptides, with edges indicating inter-peptide heavy atom contacts ($\leq$ 4.0 Å). The fine-grained (FG) graphs were constructed for each identified peptide aggregate (connected component in the CG graph), with nodes representing individual amino acids and edges representing intra- or inter-peptide amino acid heavy atom contacts ($\leq$ 4.0 Å) within that aggregate.

This graph construction process was successfully applied to all equilibrium frames. For example, in the first equilibrium frame (100 ns), the CG graph had 30 nodes (peptides) and 35 edges, forming two connected components: a large aggregate comprising 24 peptides and a smaller one of 6 peptides. The fine-grained analysis of the 24-peptide aggregate yielded an FG graph with 120 nodes (amino acids) and 169 edges, while the 6-peptide

aggregate resulted in an FG graph with 30 nodes and 40 edges. These initial snapshots illustrate the ability of the framework to capture the system's organization at both the peptide-level assembly and the internal amino acid packing within aggregates.

3.3. *Temporal evolution and statistical properties of graph metrics*

We calculated a suite of graph-theoretic properties for both the CG and FG graphs across the entire equilibrium trajectory (100–500 ns). Analysis of the time series and distributions of these properties provides a quantitative description of the dynamic equilibrium.

3.3.1. *Coarse-grained peptide network properties*

The CG graph properties characterize the peptide-level organization and aggregation state. The average number of edges in the CG graph remained relatively stable over time, with a mean of 35.392 and a low coefficient of variation (CoV) of 10.40%, indicating consistent overall inter-peptide contact levels. This stability is visible in the time series plot shown in Figure 3. The graph
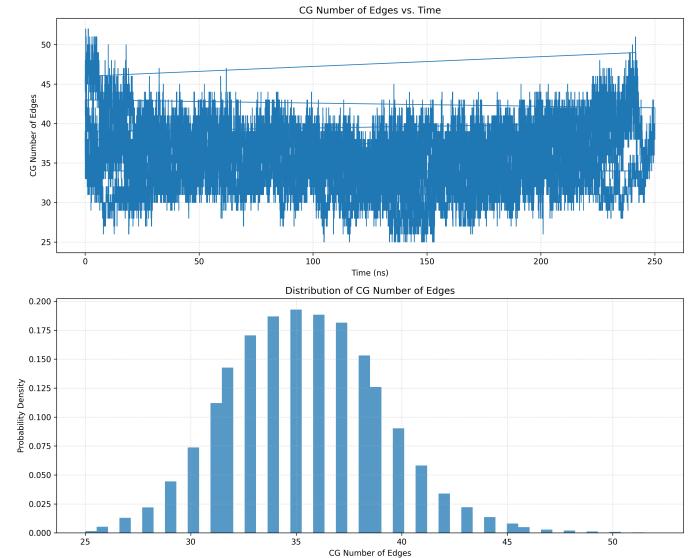

Figure 3. Time evolution (top) and distribution (bottom) of the number of edges in the coarse-grained (peptide) network. CG edges represent inter-peptide heavy atom contacts. The time series shows fluctuations around a stable mean during the equilibrium phase (100 ns onwards), indicating a relatively stable overall level of inter-peptide contact.

density, directly proportional to the number of edges for a fixed number of nodes, showed a similar stability (mean 0.081, CoV 10.40%), as illustrated in Figure 4.

The number of connected components in the CG graph averaged 1.857 (CoV 44.46%), reflecting the presence of a primary dominant aggregate alongside a small
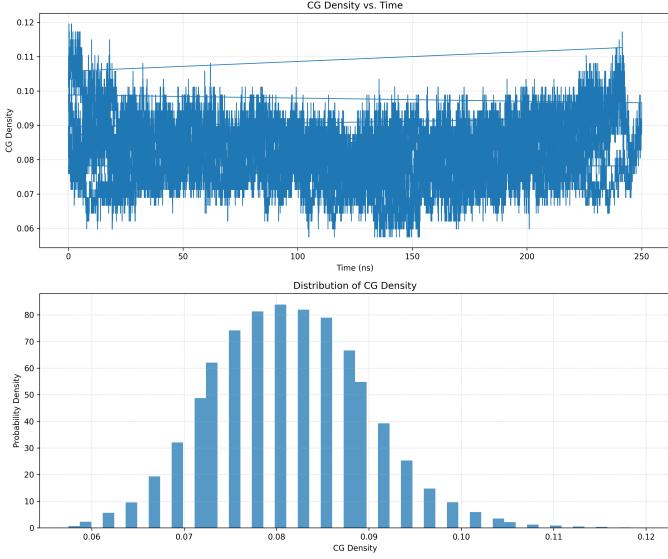
Figure 4. Time evolution and distribution of the coarse-grained (CG) density, a measure of peptide-level connectivity. The plots show that peptide aggregates maintain a moderately connected network during the equilibrium phase, with CG density fluctuating around a stable mean.

number of transient smaller clusters or isolated peptides. The size of the largest connected component (LCC) averaged 24.196 peptides (out of 30) with a standard deviation of 5.529, resulting in a CoV of 22.85%. This confirms that a large aggregate is typically present, but its size undergoes significant fluctuations, likely due to dynamic processes like monomer/oligomer exchange or aggregate merging/splitting. The temporal evolution and distribution of the LCC size are shown in Figure 5 and Figure 6.

The Fiedler value (λ_2) of the LCC's Laplacian matrix, a measure of its structural robustness and resistance to partitioning, averaged 0.107 but exhibited a high CoV of 88.01%. This high fluctuation, visible in Figure 7, indicates that the structural integrity of the dominant aggregate at the peptide level varies considerably over time, suggesting periods of relative rigidity interspersed with more flexible states. The average clustering coefficient for the CG LCC was 0.113 (CoV 46.09%), indicative of a modest level of transitivity in peptide contacts – if peptide A is connected to B and C, B and C have a moderate probability of also being connected.

3.3.2. *Fine-grained amino acid network properties (within the LCC)*

The FG graph properties within the LCC provide insights into the internal packing and interaction network at the amino acid level. The number of nodes in the FG graph of the LCC directly corresponded to its size
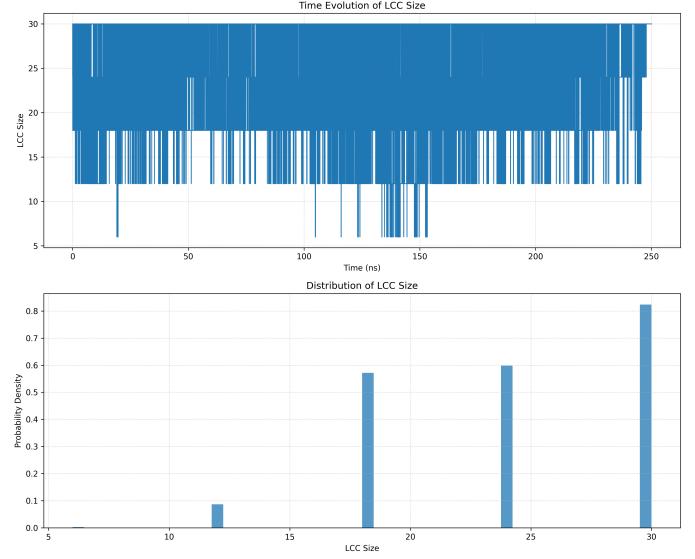

Figure 5. Time evolution (top) and probability density distribution (bottom) of the size of the Largest Connected Component (LCC) in the coarse-grained peptide network during the equilibrium phase. The figure shows that a dominant aggregate, typically containing most peptides, is present but exhibits significant fluctuations in size over time.
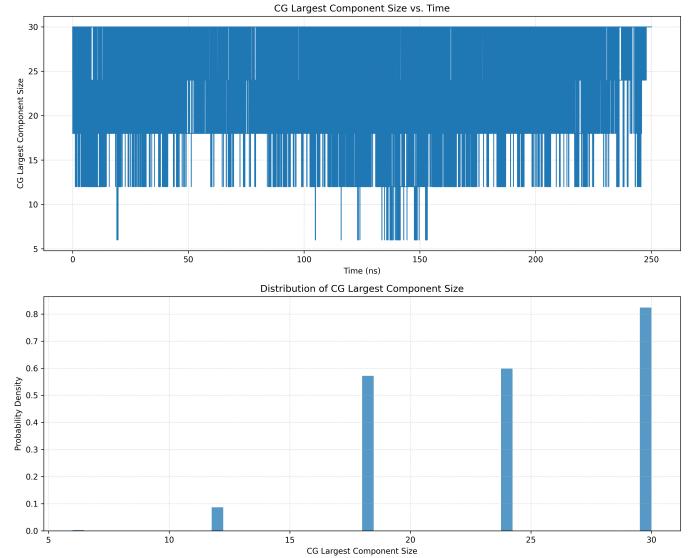

Figure 6. Time evolution (top) and distribution (bottom) of the largest coarse-grained (peptide-level) aggregate size. The system primarily forms a large aggregate, frequently including all 30 peptides, but the size fluctuates dynamically, reflecting the transient presence of smaller aggregates.

(mean 120.979 amino acids, CoV 22.85%). The number of edges averaged 174.896 (CoV 22.45%), as shown in Figure 8. The FG density averaged 0.026 (CoV 30.64%), visible in Figure 9. These relatively low density values
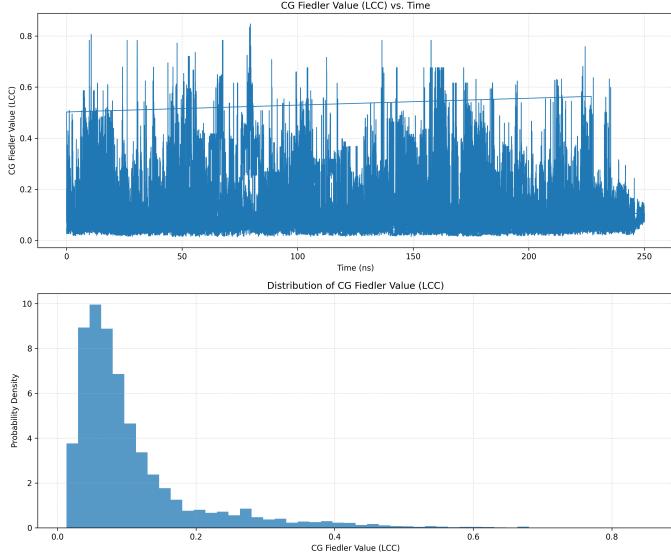
Figure 7. Time series (top) and distribution (bottom) of the Coarse-Grained (CG) Fiedler value for the Largest Connected Component (LCC) of peptide aggregates. The plot shows that the LCC's structural robustness is highly dynamic over time, with lower values indicating increased susceptibility to aggregate splitting.
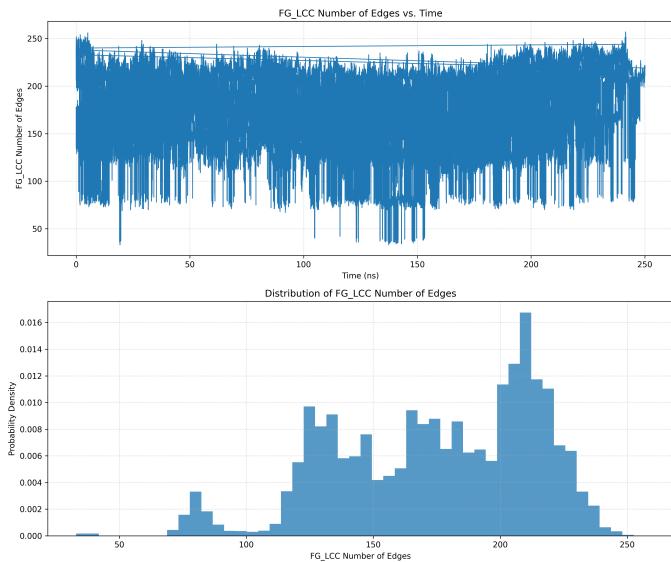

Figure 8. Time series (top) and distribution (bottom) of the number of edges in the fine-grained (amino-acid level) graph of the largest coarse-grained aggregate (FG_LCC_Number_of_Edges) over the 250 ns simulation. The number of edges represents the total number of heavy atom contacts within the largest peptide aggregate. The plots illustrate the dynamic fluctuations and the range of internal contacts, indicating a relatively sparse amino acid network within the aggregate.

suggest that the internal amino acid contact network within the largest aggregate is rather sparse.
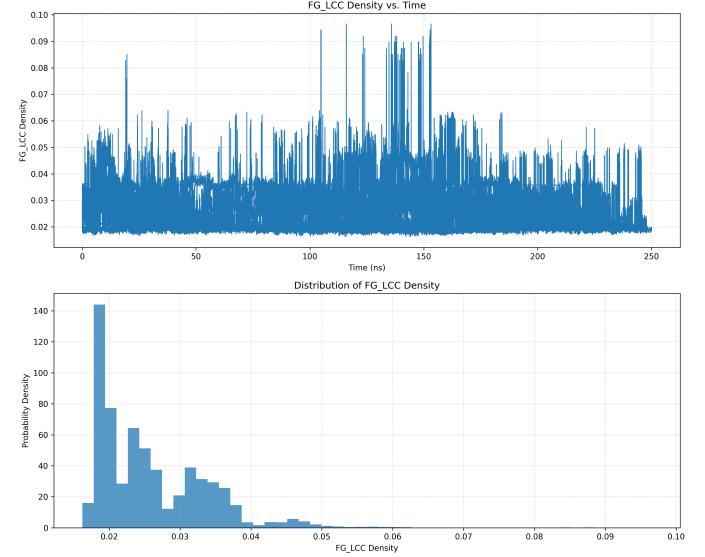

Figure 9. Time series (top) and distribution (bottom) of the fine-grained (amino acid level) network density within the largest peptide aggregate. Reflecting the density of internal amino acid contacts, this property is relatively sparse during the equilibrium phase, contributes to a stable aggregation order parameter, and is lower in aggregates preceding splitting events.

A key finding is the average FG Fiedler value of the LCC, which was extremely close to zero (mean 0.000, CoV 1716.73%, driven by the mean being near zero). As shown in Figure 10, values are consistently low or zero. A Fiedler value close to zero indicates that the graph is either disconnected or very weakly connected, meaning it can be easily partitioned into multiple components. This strongly suggests that even within the largest peptide aggregate, the amino acid contact network is not a single tightly integrated unit but is likely composed of multiple loosely connected or even transiently disconnected subgraphs. This implies that while peptides associate at the CG level, the specific amino acid contacts within the aggregate might be localized or fragmented.

Despite the low overall FG density and near-zero Fiedler value, the average FG clustering coefficient within the LCC was 0.559 (CoV 7.49

Averaging FG properties over all aggregates present in each frame yielded similar results: the average FG density across all aggregates was 0.043 (CoV 49.67%), as shown in Figure 11, and the average FG Fiedler value was 0.000 (CoV 662.51%), reinforcing the observation of
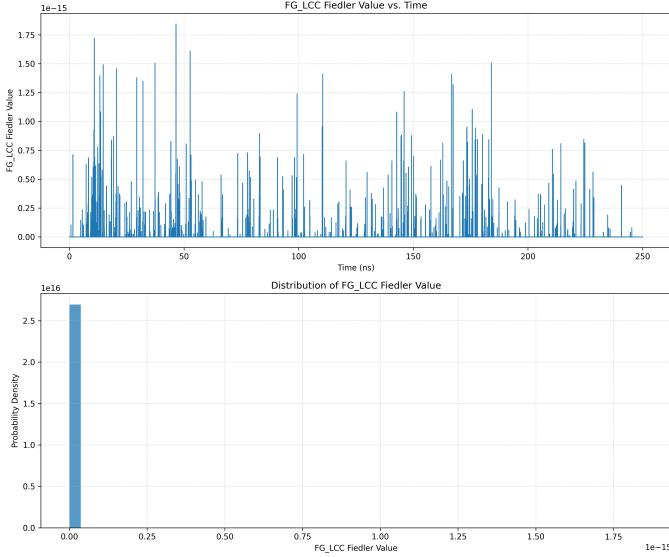
Figure 10. Time series (top) and distribution (bottom) of the fine-grained Fiedler value for the largest peptide aggregate. Values are consistently low or zero, indicating a fragmented internal amino acid contact network.

sparse and potentially fragmented internal amino acid networks within aggregates of all sizes.
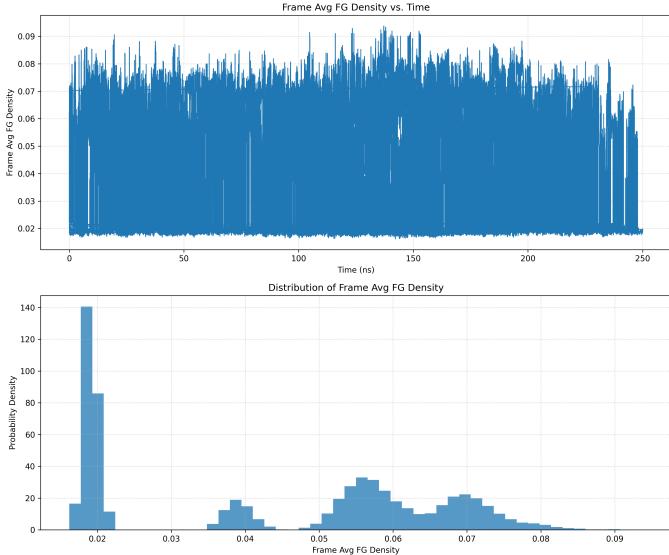

Figure 11. Time evolution (top) and distribution (bottom) of the average fine-grained (amino acid) contact density across peptide aggregates. The consistently low values observed indicate sparse internal packing of amino acids within the aggregates.

3.4. *Development and validation of stable aggregation order parameters*

Recognizing the significant fluctuations in the size of the largest aggregate (LCC_Size, shown in Figure 5 and Figure 6), we explored combining information from the CG and FG scales to develop more stable order parameters (OPs) for characterizing the aggregated state. We tested composite OPs that multiply the LCC size by different FG properties of the LCC.

The traditional OP, LCC_Size, had a CoV of 22.85% over the equilibrium trajectory. The OP combining LCC size with the FG_Fiedler value resulted in a mean near zero due to the consistently low FG Fiedler values (Figure 10), rendering it uninformative as formulated. This is illustrated in Figure 12. However, the OP defined
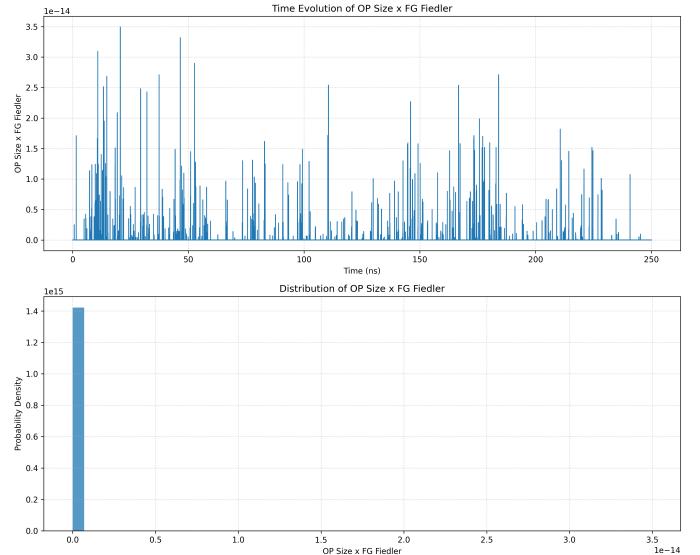

Figure 12. Time evolution (top) and distribution (bottom) of the order parameter combining the size of the largest peptide aggregate with its internal amino acid network Fiedler value (OP_Size $\times$ FG_Fiedler) over the equilibrium simulation phase. The consistently near-zero values indicate that this parameter is not informative for characterizing the aggregation state in this system.

as the product of LCC Size and the FG Density of the LCC (`OP_Size_x_FG_Density`) showed a mean of 0.585 with a remarkably low CoV of 7.07%. This is significantly lower than the CoV of LCC_Size alone. The time evolution and distribution of this stable OP are presented in Figure 13. The OP combining LCC Size with FG_Clustering (`OP_Size_x_FG_Clustering`) had a CoV of 23.75%, comparable to LCC_Size, as shown in Figure 14.

The `OP_Size_x_FG_Density` thus emerged as the most stable order parameter among those tested. Its reduced fluctuation suggests that incorporating the internal amino acid packing density provides a more ro-
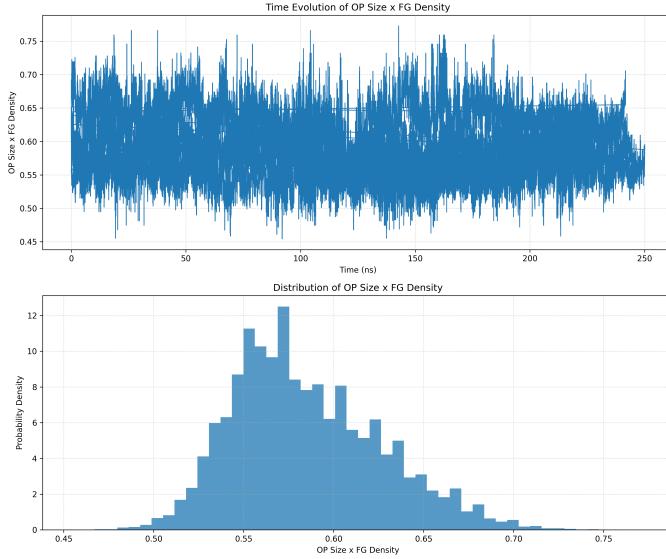
Figure 13. Time evolution and distribution of the composite order parameter (OP) defined as the size of the largest coarse-grained peptide aggregate multiplied by the fine-grained amino acid density within that aggregate. The top panel shows the value of this OP over the equilibrium phase of the simulation (100–250 ns). The bottom panel shows the corresponding probability distribution. This OP provides a stable measure of the overall peptide aggregation state, exhibiting less fluctuation than the largest aggregate size alone.

bust measure of the overall aggregated state compared to simply counting the number of peptides in the largest cluster. This indicates that the internal density of the aggregate fluctuates less drastically than its size, or that changes in size are often compensated by inverse changes in density, leading to a more stable product. This composite parameter offers a potentially more reliable metric for monitoring and comparing aggregation states across different simulation conditions or peptide sequences.

3.5. *Relationship between multiscale graph features and aggregate dynamics*

To understand how structural features relate to aggregate stability, we tracked the dynamics of individual aggregates using a Jaccard index threshold to identify constituent peptides across frames. Over the equilibrium phase, 2832 aggregate instances lasting at least 5 frames were identified, highlighting the continuous formation, dissolution, and rearrangement occurring within the system. Analysis of events identified numerous formation (2830) and dissolution (2831) events, alongside 48 potential splitting events.
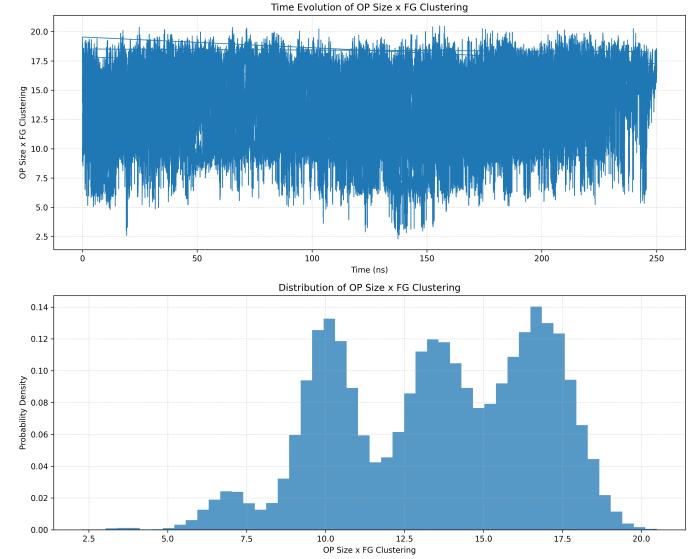

Figure 14. Time evolution (top) and probability density distribution (bottom) for the order parameter combining the size of the largest peptide aggregate with its internal amino acid clustering coefficient (OP_Size_x_FG_Clustering). The time series shows fluctuations throughout the equilibrium simulation, and the distribution captures the range and frequency of observed values, indicating its dynamic nature compared to density-based order parameters.

3.5.1. *Correlation of aggregate longevity with graph properties*

We investigated the relationship between the longevity of aggregates and their time-averaged CG and FG graph properties. Weak negative correlations were observed between aggregate longevity and average CG_Density (Pearson $r = -0.228$), average CG_Fiedler_Value ($r = -0.124$), average FG_Density ($r = -0.233$), and average FG_Fiedler_Value ($r = -0.034$). A weak positive correlation was observed between aggregate longevity and average FG_Clustering ($r = 0.098$). These relationships are visualized in Figure 15, Figure 16, Figure 17, Figure 20, Figure 18, and Figure 19, respectively.

These results are somewhat counterintuitive, suggesting that aggregates that are, on average, denser or more structurally connected (higher Fiedler value) at either scale do not necessarily persist longer. While the correlations are weak, they might hint that very compact or rigid structures could be less able to adapt to dynamic stresses or exchange monomers efficiently, potentially making them slightly more susceptible to dissociation or rearrangement over extended periods compared to more dynamic, moderately connected aggregates. However, the weak nature of these correlations indicates that aggregate longevity is likely a complex phenomenon influ-
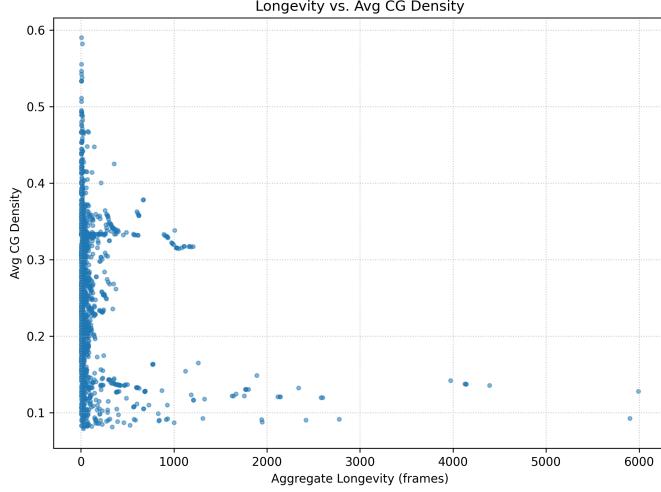
Figure 15. Scatter plot showing the relationship between the longevity of individual peptide aggregates (measured in simulation frames) and their time-averaged coarse-grained (CG) density over the equilibrium trajectory. Each point represents a distinct aggregate instance tracked across frames. A weak negative correlation is observed, suggesting that aggregates with higher average peptide-level density do not exhibit greater longevity.
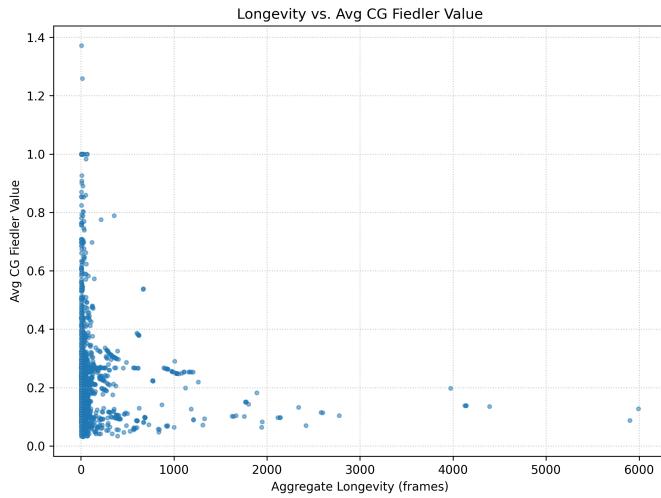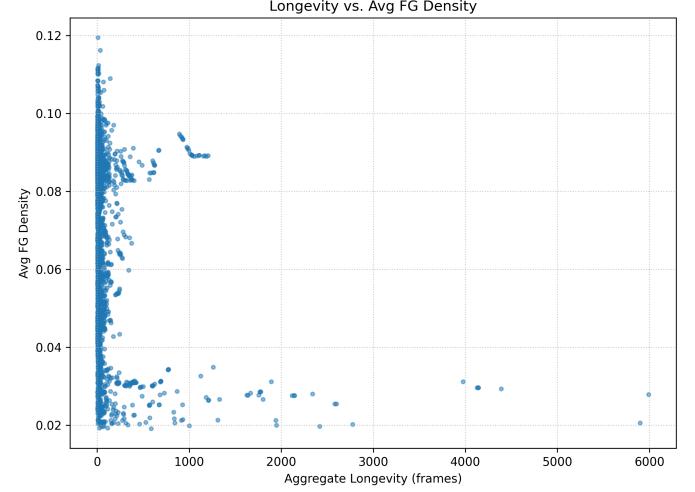

Figure 17. Scatter plot showing the relationship between aggregate longevity (in frames) and the average fine-grained (FG) density of the aggregate over its lifetime. A weak negative correlation is observed, indicating that aggregates with higher average internal packing density do not necessarily exhibit greater longevity.
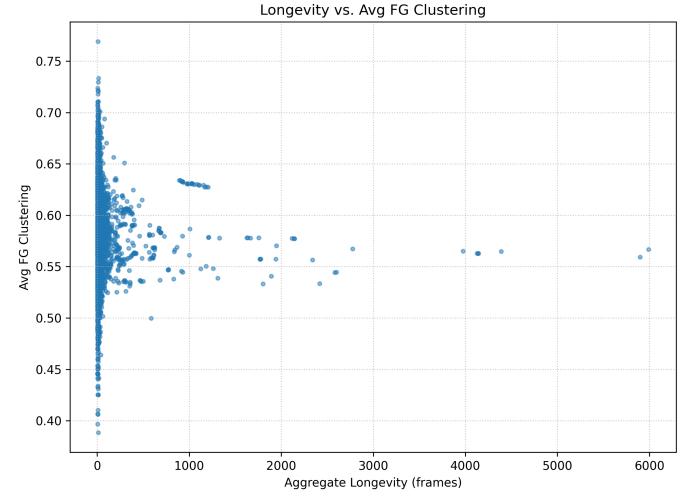

Figure 16. Scatter plot showing the relationship between aggregate longevity (in frames) and the average coarse-grained (CG) Fiedler value, a measure of peptide-level structural robustness, for detected aggregates. The plot illustrates a weak negative correlation, indicating that aggregates with higher average peptide network robustness do not necessarily exhibit greater longevity.

Figure 18. Scatter plot showing the relationship between aggregate longevity (frames) and the average fine-grained (amino acid) clustering coefficient for tracked aggregates. The plot visually confirms a weak correlation, indicating that the degree of local amino acid packing within aggregates does not strongly influence their lifespan.

enced by multiple factors beyond simple time-averaged structural properties.

3.5.2. *Structural signatures of aggregate splitting events*

To pinpoint structural features predictive of instability, we specifically compared the multiscale graph properties of aggregates in the 5 frames immediately preceding a splitting event to those of stable aggregates (those that did not split within the observed window).

Aggregates that were about to split exhibited significantly different structural characteristics compared to
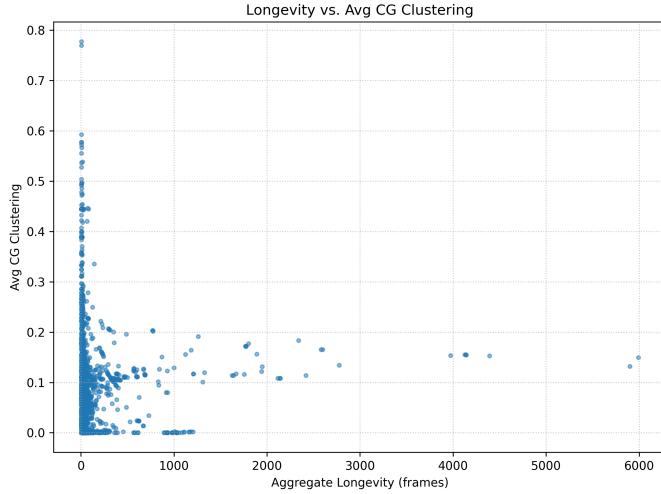
Figure 19. Scatter plot showing the relationship between aggregate longevity (in simulation frames) and the time-averaged coarse-grained (CG) clustering coefficient for individual peptide aggregates. A weak positive correlation (Pearson $r = 0.098$) is observed, indicating that higher peptide-level clustering is only weakly associated with increased aggregate lifetime.
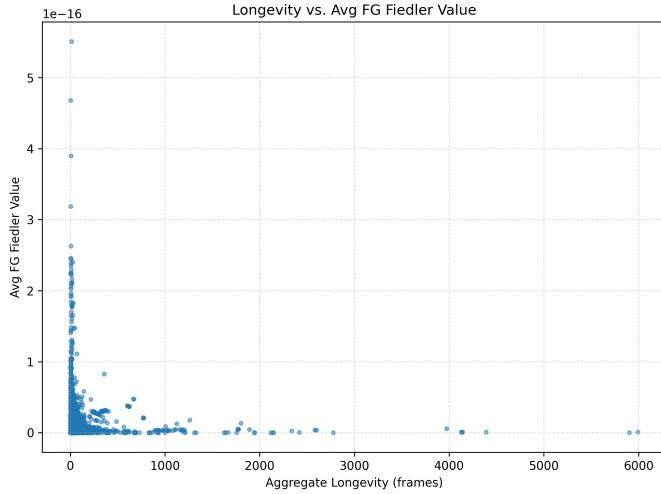

Figure 20. Scatter plot of aggregate longevity (number of frames) versus the time-averaged fine-grained (amino acid) Fiedler value for individual aggregates. The plot shows a weak correlation between these properties, indicating that the internal structural robustness of aggregates at the amino acid level, as measured by the Fiedler value, is not a strong predictor of aggregate longevity.

stable aggregates. The average CG_Density of pre-split aggregates (0.1259) was substantially lower than that of stable aggregates (0.2833). Similarly, the average CG_Fiedler_Value was much lower for pre-split aggregates (0.0758) compared to stable ones (0.2524). At the fine-grained level, the average FG_Density of pre-split

aggregates (0.0307) was also lower than that of stable aggregates (0.0715). The average FG_Fiedler_Value remained near zero for both groups (0.0000), consistent with the overall observation of fragmented internal networks and thus not discriminatory for splitting.

These findings demonstrate that aggregates poised for fragmentation are characterized by reduced connectivity and packing at both the peptide and amino acid levels. A lower CG_Density implies fewer interpeptide contacts holding the aggregate together, while a lower CG_Fiedler_Value indicates a less robust, more easily partitionable peptide network. Concurrently, a lower FG_Density suggests looser internal packing at the amino acid level. Together, these structural deficits across scales appear to predispose an aggregate to splitting. This provides crucial structural signatures of instability that can be potentially used to predict fragmentation events.

3.6. *Summary of key findings*

This multiscale graph analysis framework has provided a quantitative lens through which to view peptide aggregate structure and dynamics. We found that the system exhibits a dynamic equilibrium with a prominent aggregate. The CG peptide network is moderately connected with fluctuating stability (e.g., Figure 4, Figure 7). The FG amino acid network within aggregates is notably sparse (Figure 9, Figure 11) and often structurally fragmented (low Fiedler value, Figure 10), although localized dense contacts (high clustering) exist. The composite order parameter combining LCC_size (Figure 5) and FG_density (Figure 9) proved to be a more stable measure of the aggregated state than size alone (Figure 13). While average structural properties showed only weak correlations with longevity (Figure 15–20), the analysis of splitting events revealed significant structural precursors: aggregates with lower CG_Density and lower algebraic connectivity at the peptide level, and lower FG_Density at the amino acid level, were more prone to fragmentation. These results highlight the importance of considering structure across multiple scales to understand peptide aggregate behavior and stability.

4. CONCLUSIONS

In this study, we addressed the challenge of quantitatively characterizing the complex, dynamic structure of peptide aggregates across multiple scales using molecular dynamics simulations. We introduced a novel framework based on dynamic multiscale graph analysis, representing peptide aggregates as evolving networks at both a coarse-grained (peptide-level) and fine-grained (amino

acid-level) resolution. This approach allowed us to move beyond simple measures and probe the intricate connectivity and structural properties of aggregates over time.

Using extensive simulations of the KYFIL pentapeptide, we applied this framework to analyze the system's behavior during its equilibrium phase. Our analysis revealed a dynamic equilibrium characterized by the persistent presence of a dominant aggregate undergoing continuous structural fluctuations. At the coarse-grained level, the peptide network exhibited moderate connectivity and significant temporal variations in its structural integrity, as reflected by fluctuations in the Fiedler value. At the fine-grained level within aggregates, the amino acid contact network was found to be relatively sparse and structurally fragmented (indicated by near-zero Fiedler values), yet possessing significant local structuring (high clustering coefficient), suggesting localized dense interactions within a loosely connected overall network.

A valuable outcome of this multiscale perspective was the development of a composite order parameter, combining the size of the largest aggregate with its internal fine-grained density. This composite parameter demonstrated significantly greater stability over time compared to aggregate size alone, offering a potentially more robust metric for characterizing the aggregated state in dynamic systems.

Crucially, by tracking individual aggregates and analyzing dynamic events, we established a direct link between multiscale structural properties and aggregate stability. Specifically, we found that aggregates exhibiting significantly lower density and lower algebraic connectivity (Fiedler value) at the coarse-grained peptide level, coupled with lower density at the fine-grained amino acid level, were considerably more prone to splitting events. These findings provide clear, quantitative structural signatures of aggregate instability across different levels of organization.

Overall, this work demonstrates the power of dynamic multiscale graph analysis as a quantitative tool for dissecting the complex structure and dynamics of peptide self-assembly. We have learned that aggregate stability is not solely determined by size or overall compactness but is intimately linked to the specific network of contacts at both the peptide and amino acid levels. Structural deficits, particularly reduced connectivity and packing density across these scales, predispose aggregates to fragmentation. These insights provide a fundamental understanding of the molecular underpinnings of aggregate instability and offer valuable guidance for the rational design of peptide sequences with tailored self-assembly pathways and enhanced aggregate stability for applications in biomaterials and beyond.

REFERENCES

Airale, L., Longa, A., Rigon, M., Passerini, A., & Passerone, R. 2025, Simple Path Structural Encoding for Graph Transformers. https://arxiv.org/abs/2502.09365

Claret, A. 2023, Theoretical tidal evolution constants for stellar models from the pre-main sequence to the white dwarf stage Apsidal motion constants, moment of inertia, and gravitational potential energy, doi: https://doi.org/10.1051/0004-6361/202346250

Coupette, C., Wayland, J., Simons, E., & Rieck, B. 2025, No Metric to Rule Them All: Toward Principled Evaluations of Graph-Learning Datasets. https://arxiv.org/abs/2502.02379

Deal, D., & Espinoza, N. 2024, Spelunker: A quick-look Python pipeline for JWST NIRISS FGS Guide Star Data, doi: https://doi.org/10.21105/joss.06202

Dong, X., Xie, H., Chen, Y., et al. 2025, Discovering dense hydrogen solid at 1200K with deep variational free energy approach. https://arxiv.org/abs/2501.09590

Eldén, L. 2023, Multiway Spectral Graph Partitioning: Cut Functions, Cheeger Inequalities, and a Simple Algorithm. https://arxiv.org/abs/2302.03615

Li, K., Xia, Q.-Q., Kim, C.-H., et al. 2021, Two contact binaries with mass ratios close to the minimum mass ratio, doi: https://doi.org/10.3847/1538-4357/ac242f

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

Pan, L., Padoan, P., Scalo, J., Kritsuk, A. G., & Norman, M. L. 2011, Turbulent Clustering of Protoplanetary Dust and Planetesimal Formation, doi: https://doi.org/10.1088/0004-637X/740/1/6

Parsotan, T., Laha, S., Palmer, D. M., et al. 2023, BatAnalysis – A Comprehensive Python Pipeline for Swift BAT Survey Analysis. https://arxiv.org/abs/2303.06255

14

Parsotan, T., Palmer, D. M., Ronchini, S., et al. 2025, BatAnalysis – A Comprehensive Python Pipeline for Swift BAT Time-Tagged Event Data Analysis. https://arxiv.org/abs/2502.00278

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

Roncoli, A., Ćiprijanović, A., Voetberg, M., Villaescusa-Navarro, F., & Nord, B. 2024, Domain Adaptive Graph Neural Networks for Constraining Cosmological Parameters Across Multiple Data Sets. https://arxiv.org/abs/2311.01588

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

Sun, Z., Ting, Y.-S., Liang, Y., et al. 2024, Knowledge Graph in Astronomical Research with Large Language Models: Quantifying Driving Forces in Interdisciplinary Scientific Discovery. https://arxiv.org/abs/2406.01391

Voit, G. M., Pandya, V., Fielding, D. B., et al. 2024, Equilibrium States of Galactic Atmospheres I: The Flip Side of Mass Loading. https://arxiv.org/abs/2406.07631

Wadhwa, S. S., Landin, N. R., Arbutina, B., et al. 2024, Low Mass Contact Binaries: Orbital Stability at Extreme Low Mass Ratios. https://arxiv.org/abs/2411.02020

Wang, X., Jiang, Y., & Gong, S. 2015, Analysis of the Potential Field and Equilibrium Points of Irregular-shaped Minor Celestial Bodies, doi: https://doi.org/10.1007/s10509-014-2022-8

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

Zhang, S., Ye, Q., & Hu, H. 2025, Structure-Preference Enabled Graph Embedding Generation under Differential Privacy. https://arxiv.org/abs/2501.03451