Event Tracking: A systematic method for analyzing nucleation and growth in hierarchical self-assembly

Argha Chakraborty[†], Rumela Adhikary[†], Sangeeta Das[†], and Avisek Das[*]

School of Chemical Sciences, Indian Association for the Cultivation of Science, Kolkata, India

E-mail: mcsad@iacs.res.in

Abstract

Molecular self-assembly has garnered significant attention in the field of biomaterials and nanotechnology due its potential for creating novel materials with diverse applications. The entire process is guided by either classical nucleation and growth or formation of multiple nucleus and their growth and finally the fusion of the self-assembled states. Systematic way to track this nucleation, growth and fusion process is still unknown. We have developed an algorithm to systematically identify all the possible molecular events. The events provide immediate information when a cluster or individual molecule combines with another cluster or molecule, or when a cluster or molecule detaches from another, during each stage of the mechanism. By comprehensively examining the entire process, we can gain a clearer un derstanding of the molecular mechanisms involved in the assembly process. We applied this algorithm to self-assembly of some ultrashort peptides. Through a systematic analysis, we identify commonalities and differences in the self-assembly mechanism of various ultrashort peptides. This comparative analysis contributes to a deeper understanding of the mechanisms governing ultrashort peptide self-assembly, offering valuable guidance for the

rational design of biomaterials which can serve various technological and biomedical purposes.

1 Introduction

In the fascinating realm of nanotechnology, where the manipulation of matter takes place at the molecular level, one captivating phenomenon stands out molecular self-assembly. The intricate dance of these minuscule building blocks, driven by their inherent chemical properties, gives rise to complex and well-defined nanostructures with diverse applications in biomedicine, materials science, and beyond.[1–5]

In early nineties Whitesides et al. described a chemical strategy to synthesize nanostructure from molecular self-assembly.[6] Within few years Zhang et al. and Ghadiri et el. synthesized nanostructure which were self-assembled from oligopeptides.[7–9] In 2003, Reches and Gazit conducted a study wherein they demonstrated that the uncapped zwitterionic dipeptide L-phenylalanine (FF), which is a component of the Alzheimer's Aβ-peptide sequence, exhibited spontaneous self-assembly into nanotubes in an aqueous medium.[10] This breakthrough prompted thorough investigations into the self-assembly behavior of di- and tri-peptides, commonly referred to as ultrashort peptides (USPs), and their derivatives. These studies delved into their ability to spontaneously form organized structures in water and various solvent mixtures. In contrast to longer peptides, ultrashort peptides (USPs) possess a considerably simpler conformational landscape. Nonetheless, they hold a pivotal advantage due to the molecular architecture of these peptides, which grants them access to a wide array of intermolecular interactions. These interactions include electrostatic forces, hydrogen bonding, aromatic stacking, and hydrophobic interactions.[11,12] Despite their simple conformational nature, ultrashort peptides (USPs) exhibit the remarkable ability to spontaneously assemble into a diverse range of nano- to micron-scale structures. These structures include spheres, micelles, tubes, fibers, ribbons, two-dimensional sheets, and more complex

higher-order assemblies such as supramolecular gels.[10,13–16]

The entire process is guided by multiple steps where assembled state of various length scale is formed. The mechanism of these processes is still unknown and an active research interest.[17,18] But these researchs unable to provide each steps of the assembly process. But from very old days scientist wanted to get idea about nucleation and growth phenomena of various self-assembly problems. In the middle of twenth century Turkevich et al. gave an idea about nucleation and growth of colloidal gold.[19] Before their work Levine and Dube had already shown the interaction between two hydrophobic colloidal particles by approximate Debye-Hückel theory.[20] In recent days also nucleation and growth of colloidal particles and nanoparticles is still interest of many scientists.[21–25]

There are theoretical studies also in which pathways of self-assembly was analysed by statistical mechanics.[26] On the other hand, computer simulations have been employed in various domains of peptide-based self-assembly. These applications encompass areas such as cyclic peptides, amyloid-type assemblies, and peptide amphiphiles.[27–40] Molecular Dynamics (MD) simulation has emerged as the principal computational approach for investigating self-assembly phenomena due to its capacity to replicate systems closely resembling experimental conditions. Studies employing simulations on extended peptides have unveiled that the conformational distributions play a crucial role in shaping the initial phases of the assembly process.[35] The scenarios within the assemblies of ultrashort peptides (USPs) are anticipated to differ significantly owing to their limited conformational complexities. Numerous investigations on USPs have been conducted using all-atom (AA) implicit-solvent, all-atom (AA) explicit-solvent, and coarse-grained (CG) models.[14,41–46]

Very recently Zhou et al. showed comaprative work of self-assembly of peptides composed of Leucine and Isoleucine and the role of conformation and intermolecular interactions.[47] Whereas Xiong et al. also showed formation of some peptide sequence dependent nanofiber structures of Phenylalanine and Isoleucine Tripeptide.[48] Inspired from their work we had chosen Leucine(Figure 1b and Figure 1e), Isoleucine(Figure 1c and Figure 1f) and Pheny-

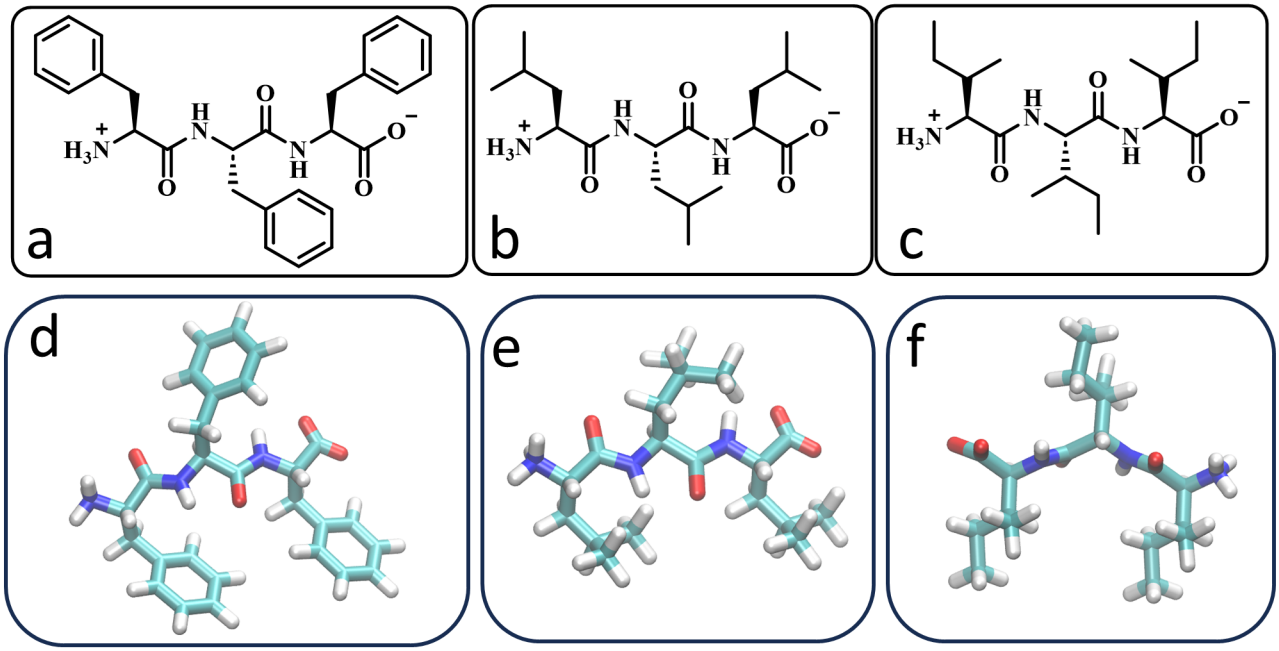


Figure 1: (a)Chemical structure of uncapped L-Phenylanaline Tripeptide (L-FFF); (b)Chemical structure of uncapped L-Leucine Tripeptide(L-LLL); (c)Chemical structure of uncapped L-Isoleucine Tripeptide (L-III); (d)Three dimensional molecular model of L-FFF; (e)Three dimesional molecular model of L-LLL; (f)Three dimensional molecular model of L-III

lalanine tripeptides(Figure 1a and Figure 1d) and compared their nucleation and growth process. For that we have developped an algorithm by which we can determine all possible molecular events. We applied this algorithm to our all atom MD simulation data of these three tripeptides. And at the end we have presented commonalities and differences in the self-assembly mechanisms of these peptides.

2 Clustering along the trajectory and storing the clusters

Understanding the underlying mechanism of nucleation and growth processes can be achieved by studying the sequential stages involved in the formation of aggregated states. To accomplish this, a thorough analysis of the aggregates is necessary at each step of the trajectory.

Now it is important to understand which is ordered aggregates or clusters. In our approach, we opted for a simpler spatial clustering algorithm to aid in the analysis. This algorithm involves defining a cut-off distance in space. If two molecules are located within this distance of each other, they are grouped together into the same cluster(Figure 2). This technique allows us to directly observe which molecules belong to which cluster in each time step.

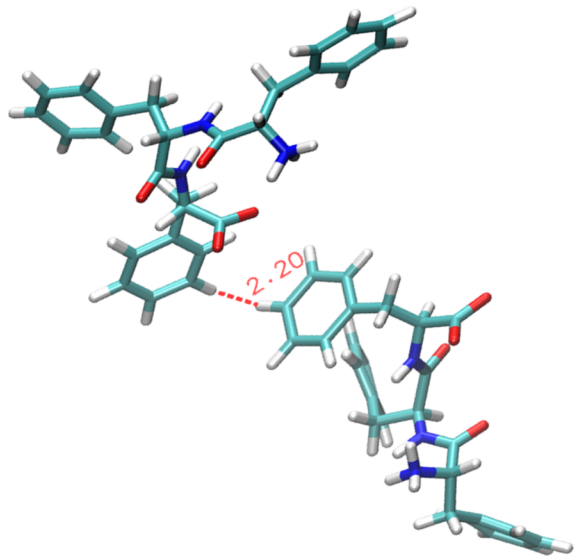


Figure 2: Two molecules are with in a cut-off distance

Let's consider a system containing N molecules, which can form various clusters. In each time frame, there can be a total of $\sum_{n=1}^{N} {}^{N}C_n (= 2^N - 1)$ possible clusters. If we consider this to be a set $\mathcal{C}_P$ -

$$\mathcal{C}_P = \{\alpha, \beta, ...\}_{(2^N-1)} \tag{1}$$

At time step t, the system contains $M_{\mathcal{C}}$ clusters, and in the subsequent time step $t+1$, these clusters may undergo different changes. Let's say at time t the set of cluster is -

$$\mathcal{C}_t = \{\gamma(t), \delta(t), ...\}_{M_{\mathcal{C}}(t)} \tag{2}$$

and at time $t+1$,

$$\mathcal{C}_{t+1} = \{\gamma(t+1), \delta(t+1), ...\}_{M_{\mathcal{C}}(t+1)} \tag{3}$$

Here $\mathcal{C}_t \subseteq \mathcal{C}_P$ and $\mathcal{C}_{t+1} \subseteq \mathcal{C}_P$ and $\mathcal{C}_t$ & $\mathcal{C}_{t+1}$ is may be or may not be equal. Also $\mathcal{C}_t \cap \mathcal{C}_{t+1}$ may be or may not be zero. If $\mathcal{C}_t \cap \mathcal{C}_{t+1} = 0$, then none of the clusters remain at the timestep $t+1$. They might lose or gain a small number of molecules, split into two or more clusters, completely disintegrate into individual molecules, or merge with different clusters that existed at the immediate previous time step, $t-1$. If $\mathcal{C}_t \cap \mathcal{C}_{t+1} \neq 0$, then some of the clusters remain same in the immediate next timestep $t+1$ and other might have same fate as discussed earlier. And if $\mathcal{C}_t = \mathcal{C}_{t+1}$, then the clusters present at timestep t remain exactly same at timestep $t+1$.

Establishing connections between the elements of $\mathcal{C}_t$ and $\mathcal{C}_{t+1}$, presents a significant challenge. One potential strategy to address this is to assign a unique cluster index to all elements of $\mathcal{C}_P$ right from the start. Subsequently, we can then identify and compare which clusters are present at each time step and track their evolution between the consecutive time steps, t and $t+1$.

While this direct approach may be effective for cluster bookkeeping, implementing the corresponding algorithm would be inherently complex. The huge number of possible clusters and the potential changes occurring from one time step to another can quickly lead to computational challenges. The algorithm would need to efficiently handle various scenarios such as cluster merging, splitting, and disintegration, while accurately tracking the molecules' movements among clusters.

To tackle this intricacy and to simplify the analysis, our approach adopts a different path. To store the clustering data effectively, we devised a method that prioritizes the clusters based on their indices. Specifically, for each frame, the cluster with the lowest index is designed to contain the maximum number of molecules. Subsequently, the subsequent clusters follow this pattern, with their indices and sizes determined accordingly. Although it may not provide unique cluster indices or explicitly track the transformation of individual clusters, it offers a more manageable and less computationally demanding solution.

In other words, our data storage scheme ensures that the cluster with the lowest index in each frame represents the largest cluster in terms of size, followed by the clusters with higher indices. This approach allows us to track the evolution of the aggregated states throughout the process and gain insights into their growth patterns and transformations.

3 Events: definitions

So if we look from point of view of a single molecule between an time interval there are three and only three fates of the single molecule in a single time interval. These categories are defined from the point of view of the the sizes of clusters the molecule is part of in times t and $t+1$. This notion applies to clusters of all size (at both times), including singleton clusters. Let's say the molecule I, is part of clusters $\mathcal{C}_\alpha(t)$ and $\mathcal{C}_\beta(t+1)$ at two time-points and the corresponding cluster sizes are $\eta_\alpha(t)$ and $\eta_\beta(t+1)$, respectively. Then the change in cluster size for the molecule I is

$$\delta_I = \eta_\beta(t+1) - \eta_\alpha(t) = s_I(t+1) + s_I(t) \tag{4}$$

Here $\alpha, \beta, \gamma, ...$ are the cluster index and the s_I is the size of cluster in which the molecule I belongs.

So the three possible fates of single molecules are : **1. Association:** Molecule has participated in an event which is at least partly associative in nature. This is the case when

$\delta_I > 0.$; **2. Dissociation:** Molecule has participated in an event which is at least partly dissociative in nature. This is the case when $\delta_I < 0$. There could be other things happening, but from the perspective of the molecule in question the resultant event is associative and dissociative for the last two events. **3. No-change:** This is for $\delta_I = 0$.

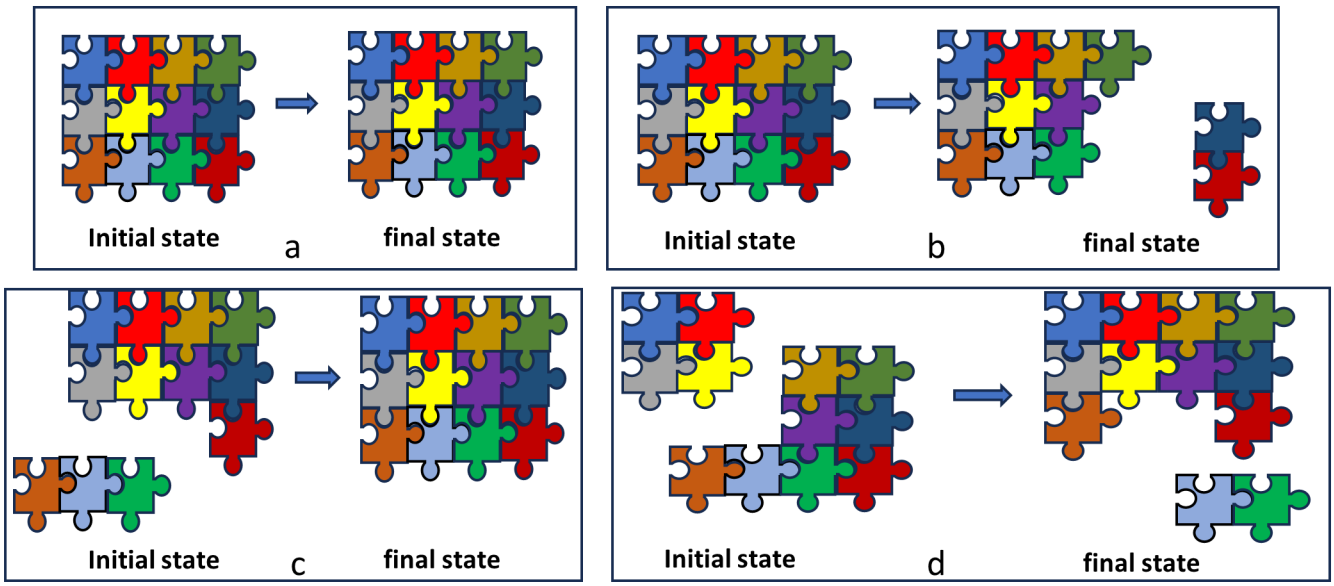


Figure 3: (a) Non interacting; (b) Pure Fission; (c) Pure Fusion and (d) Mixed Fusion-Fission events

Now if we look at from the perspective of clusters then the complete list of events will be the following.

- **Pure fusion**. Two or more clusters at time t are merging into a single cluster at time $t+1$, and nothing else. Due to the finite time interval there is a possibility that more than two clusters are involved, even though a collision other than a simple binary collision is very unlikely. For this event, minimum number of clusters involved at time t is two; number of clusters involved at time $t+1$ is one and only one and all molecules involved in this event have $\delta > 0$.

- **Pure fission**. A single cluster at time t is dissociating into at least two or more clusters at time $t+1$ and nothing else. For this event number of clusters involved at time t

is one and only one; minimum number of clusters involved at time $t+1$ two and all molecules involved in this event have $\delta < 0$.

- **No change**. Nothing happens to a single cluster. For this case number of clusters involved at time t: one and only one; number of clusters involved at time $t+1$: one and only one and all molecules in the cluster have $\delta = 0$.

- **Mixed fusion-fission**. At least two clusters at time t undergo both fission and fusion and generate at least two clusters, that are different from the initial two clusters at time $t+1$. There could be more than two clusters both at the beginning and the end of the interval. For this type of event minimum number of clusters involved at time $t$is two; minimum number of clusters involved at time $t+1$ is two; there is no maximum number of clusters, in principle, both at the beginning and end of the interval; there could be more than one molecules participating in the fission event irrespective of the participating cluster have $\delta < 0$; there could be more than one molecules participating in the fusion event irrespective of the participating cluster have $\delta > 0$ and in this type of event no molecule can remain in the unchanged state, i.e. no molecule involved in this event will have $\delta = 0$.

A cartoon diagram of the all possible events is shown in Figure 3.

4 Algorithm for tracking the time evolution of events

To explain the algorithm let us first define how two clusters are similar and how one cluster is shared with another cluster of consecutive timestep.

Two clusters are said to be same if and only if, both of them store the exact same set of molecules. The notion involves identity of molecules and not the positions of them. Therefore if two clusters are same they must have the same size. But the reverse is not true; if size of two cluster is same that doesn't necessarily means that they are same.

Now to understand how two clusters are shared let us define a parameter for the ownership of clusters. A notation for number of common elements of two clusters or sets $\mathcal{A}(t)$ and $\mathcal{B}(t')$ is $\nu(\mathcal{A}(t) \cap \mathcal{B}(t'))$ and this only makes sense under the condition $t = t'$.

Let us define two possible ownership parameter-

$$\mathscr{W}(\mathcal{A}(t)|\mathcal{B}(t')) = \frac{\nu(\mathcal{A}(t) \cap \mathcal{B}(t'))}{\eta_{\mathcal{B}}(t')} \tag{5}$$

This measures how much of cluster $\mathcal{B}(t')$ is owned by cluster $\mathcal{A}(t)$.

$$\mathscr{W}(\mathcal{B}(t')|\mathcal{A}(t)) = \frac{\nu(\mathcal{A}(t) \cap \mathcal{B}(t'))}{\eta_{\mathcal{A}}(t)} \tag{6}$$

This measures how much of cluster $\mathcal{A}(t)$ is owned by cluster $\mathcal{B}(t')$.

Here $\eta_{\mathcal{B}}(t')$ is the sizes of $\mathcal{B}(t')$ and $\eta_{\mathcal{A}}(t)$ is the size of $\mathcal{A}(t)$.

In the very first step of the algorithm, a list of molecule index named master list was constructed. Then δ_Is were calculated for all the molecules and an ownership matrix was formed. The index of cluster at time t and $t+1$ was defined to be α and α' respectively. The number of clusters were $M_{\mathcal{C}}(t)$ and $M_{\mathcal{C}}(t+1)$, respectively. The share of each cluster at time t in all clusters at $t+1$ and *vice versa* were calculated.

For each $\alpha = 1, \ldots, M_C(t)$ there was a set of ownerships

$$\left\{\mathscr{W}(\mathcal{C}_\alpha(t)|\mathcal{C}_1(t+1)), \mathscr{W}(\mathcal{C}_\alpha(t)|\mathcal{C}_2(t+1), \ldots, \mathscr{W}(\mathcal{C}_\alpha(t)|\mathcal{C}_{M_{\mathcal{C}}(t+1)}(t+1))\right\} \tag{7}$$

And also for each $\alpha' = 1, \ldots, M_{\mathcal{C}}(t+1)$ there was a set of ownerships

$$\left\{\mathscr{W}(\mathcal{C}_{\alpha'}(t+1)|\mathcal{C}_1(t)), \mathscr{W}(\mathcal{C}_{\alpha'}(t+1)|\mathcal{C}_2(t), \ldots, \mathscr{W}(\mathcal{C}_{\alpha'}(t+1)|\mathcal{C}_{M_{\mathcal{C}}(t)}(t))\right\} \tag{8}$$

This information was stored in a $M_{\mathcal{C}}(t) \times M_{\mathcal{C}}(t+1)$ matrix. And for clusters at two time-points without any common molecules the ownership value will be zero.

In the next step, at first we declared four arrays named "no-change", "pure-fission", "pure-fusion" and "mixed-fusion-fission" for the respective cluster of initial state and also we declared three arrays named "pure-fission-final", "pure-fusion-final" and "mixed-fusion-fission-final" to store the cluster of final state of the respective events. Then we took each indices from the list of molecules and calculated their δ value. Let's say for molecule index I, it is δ_I. If $\delta_I = 0$ then the event for the molecule is 'no-change' event. The algorithm to determine this event was as follows -

Algorithm 1: Algorithm for no-change event

```
Input  : master list which contains molecules
Output: A list, no-change, filled with values and modified master list
if δ_I = 0 then
    if s_I(t) = 1 then
        no-change ← α ;          // α is the cluster index of consideration
        master list → I
    end
    else if s_I(t) > 1 then
        if 𝒲(C_α(t)|C_α'(t+1)) = 1 then
            no-change ← α;
            master list → C_α(mol) ;        // C_α(mol) are the molecules of the
             cluster C_α
        end
    end
end
```

In the algorithm 1 $s_I(t) > 1$ means the cluster in which the molecule I resides contains two or molecules and lets say its index is α at time t. The algorithm 1 is checking the ownership of cluster α with some cluster α' at time $t+1$.

For $\delta_I < 0$ the events would be either pure fission or mixed fusion-fission. If the event is pure fission then the molecule in question must not be singleton at time t i.e. $s_I(t) > 1$ and also the clusters into which the molecules of $\mathcal{C}_\alpha(t)$, where $\alpha = c_I(t)$, get distributed at time $t+1$, must not have any other molecules except for the ones present in $\mathcal{C}_\alpha(t)$ which means

the sum of sizes of clusters at time $t+1$ will be equal to the size of cluster $\mathcal{C}_\alpha(t)$.

We defined a set $\mathbb{F}_{t+1} = \{\alpha'\}$ such that $\forall\ \alpha' \in \mathbb{F}_{t+1}$

Algorithm 2: Algorithm for pure-fission and mixed fission-fusion event

```
Input  : master list which contains molecules, num_cluster
Output: Lists pure-fission,pure-fission-final,mixed-fusion-fission and
          mixed-fusion-fission-final filled with values and modified master list
if δ_I < 0 then
    if s_I(t) > 1 then
        for i ← 1 to num_cluster(t + 1) do
            if 𝒲(C_α(t)|C_i(t + 1)) ≠ 0 then
                distributed_mol ← C_i(mol) ; // C_i(mol) are the molecules of the
                 cluster C_i
                distributed_index ← i ;          // distributed_index contains the
                 cluster indexes of final phase of fission event
            end
        end
        if len(distributed_mol) = η_α) then
            pure-fission-final ← distributed_index
            pure-fission ← α
            master list → C_α(mol) ;        // C_α(mol) are the molecules of the
             cluster C_α
        end
        else if len(distributed_mol) ≠ η_α) then
            mixed-fusion-fission ← α
            mixed-fusion-fission-final ← distributed_index
            master list → C_α(mol)
        end
    end
end
```

$$\mathscr{W}(\mathcal{C}_\alpha(t)|\mathcal{C}_{\alpha'}(t+1)) \neq 0. \tag{9}$$

And for a pure fission event

$$\sum_{\alpha' \in \mathbb{F}_{t+1}} \eta_{\alpha'}(t+1) = \eta_\alpha(t) \tag{10}$$

Hence if the event is pure fission, α and the set $\mathbb{F}_{t+1}$ are identified and all the molecules and clusters are removed from future consideration and the counter of pure fission is updated by one. If the event is mixed fission-fusion then the counter for mixed event is updated by one and the procedure is followed for dealing this type of event to update the master list and cluster under consideration. The algorithm is shown in algorithm 2.

For $\delta_I > 0$ the events will be either pure fusion or mixed fission-fusion event. If the event is pure fusion then the outcome of this event must be single cluster and nothing else. That means there will one and only α' at time $t+1$ such that

$$\mathscr{W}(\mathcal{C}_\alpha(t)|\mathcal{C}_{\alpha'}(t+1)) \neq 0. \tag{11}$$

And it necessarily true that

$$\mathscr{W}(\mathcal{C}_\alpha(t)|\mathcal{C}_{\beta'}(t+1)) = 0 \text{ for all } \beta' \neq \alpha' \tag{12}$$

Hence we can determine α'. Also we found out which clusters at time t has non-zero overlap with $\mathcal{C}_{\alpha'}(t+1)$. One of them was obviously α and the rest were easy to find out by looking at the ownership of $\mathcal{C}_{\alpha'}(t+1)$ in the clusters at time t. We called this set $\mathbb{P}_t$, defined as

$$\mathscr{W}(\mathcal{C}_{\alpha'}(t+1)|\mathcal{C}_\beta(t)) > 0 \text{ if and only if } \beta \in \mathbb{P}_t \tag{13}$$

Since this is a pure fusion all molecules in the fused cluster at time $t+1$, $\mathcal{C}_{\alpha'}(t+1)$ are distributed in clusters $\mathcal{C}_\beta(t)$, with $\beta \in \mathbb{P}_t$, in other words, none these clusters at time t can have molecules other than those present in $\mathcal{C}_{\alpha'}(t+1)$. The consequence of this the following.

$$\sum_{\beta \in \mathbb{P}_t} \eta_\beta(t) = \eta_{\alpha'}(t+1) \tag{14}$$

If the event is pure fusion, α' and the set $\mathbb{P}_t$ are identified and all the molecules and clusters are removed from future consideration and the counter of pure fusion is updated by one. If the event is mixed fission-fusion then the counter for mixed event is updated by one and the procedure is followed for dealing this type of event to update the master list and cluster under consideration.The algorithm is shown in algorithm 3

Algorithm 3: Algorithm for pure-fusion and mixed fission-fusion event

```
Input  : master list which contains molecules, num_cluster
Output: Lists pure-fusion,pure-fusion-final,mixed-fusion-fission and
          mixed-fusion-fission-final filled with values and modified master list
if δ_I < 0 then
    if s_I(t) > 1 then
        for i ← 1 to num_cluster(t) do
            if 𝒲(C_α'(t+1)|C_i(t)) ≠ 0 then
                distributed_mol ← C_i(mol) ; // C_i(mol) are the molecules of the
                 cluster C_i
                distributed_index ← i ;          // distributed_index contains the
                 cluster indexes of initial phase of fission event
            end
        end
        if len(distributed_mol) = n_α') then
            pure-fusion-final ← α'
            pure-fusion ← distributed_index
            master list → distributed_mol
        end
        else if len(distributed_mol) ≠ n_α') then
            mixed-fusion-fission ← distributed_index
            mixed-fusion-fission-final ← α'
            master list → C_α(mol)
        end
    end
end
```

For the mixed events, α is known and two sets $\mathbb{P}_t$ and $\mathbb{F}_{t+1}$ are defined as the sets of indices of clusters at time points t and $t+1$ that are involved in the mixed fission-fusion event. These two set contained all the clusters that are part of this event. This was done by a back and forth iterative and recursive algorithm. At the beginning of the algorithm $\mathbb{F}_{t+1}$

had several elements for which

$$\mathscr{W}(\mathcal{C}_\alpha(t)|\mathcal{C}_{\beta'}(t+1)) \neq 0 \tag{15}$$

and $\mathbb{P}_t$ had one element, namely α. The steps were, here, β is a generic index present in $\mathbb{P}_t$ and β' is a generic index in $\mathbb{F}_{t+1}$. For each β', β is identified for which

$$\mathscr{W}(\mathcal{C}_{\beta'}(t+1)|\mathcal{C}_\beta(t)) \neq 0 \tag{16}$$

if the index was not already present. Then for each β, β' is searched for which

$$\mathscr{W}(\mathcal{C}_\beta(t)|\mathcal{C}_{\beta'}(t+1)) \neq 0 \tag{17}$$

if the index was not already present. This step is iterated until one search in any direction yielded nothing or all clustersat one time-point were included. The master list of molecules and list of clusters are updated by removing these molecules from the list.

And after completing all the steps another molecule index is taken from the updated master lists and the previous steps are repeated and iterated till the master list became empty.

Handling statistical fluctuation- estimation of time scale of an assembly process: At some point in time, all the major clusters remain unchanged with only minor fluctuations. Fluctuations occur when a very small number of molecules associate with an existing cluster or when a few molecules dissociate from one. Mixed events can involve both of these occurrences, or multiple large clusters might undergo such changes.

The question that arises is what constitutes a "very small" number. Without addressing this, classification becomes difficult. To address this, we set a strict cutoff at "two." While two seems reasonable, if three molecules dissociate from a cluster of 300 molecules, that

would still be considered a fluctuation. Thus, we cannot completely rule out all possibilities. Nevertheless, two serves as a good starting point, and manual inspection can be done, accounting for variations in different situations, especially with very large assembly states.

Let's examine when events can be considered fluctuations when the smallest cluster size is two or less. If a two-molecule cluster associates with or dissociates from a five, fifteen, or fifty-molecule cluster, then the event is definitely not a fluctuation event. If a two-molecule cluster associates with or dissociates from a hundred-molecule cluster, the situation becomes more complex. In this case, it won't be considered a fluctuation event if additional changes occur within the same event, if these changes occur continuously over a time interval (resulting in linear growth of the cluster), and if a cluster of almost the same size exists at the end of the trajectory. The event will likely be considered a fluctuation event if nothing else happens to that cluster, but later along the trajectory, it undergoes a major event. In such cases, there must be a significant time gap between the initial event and the subsequent one. The event will undoubtedly be classified as a fluctuation event if a two-molecule cluster associates with or dissociates from a two-hundred-molecule cluster.

From the discussions above, it seems that to rigorously reclassify mixed events as non-change events, we need to track them over a long temporal duration. This duration could be specified as a parameter as well.

Let's say number of cluster at time t and $t+1$ are $M_{\mathcal{C}}(t)$ and $M'_{\mathcal{C}}(t+1)$ respectively. We might drop the time arguments whenever necessary. As discussed the small cluster size cutoff $\eta_{small} = 2$. The steps of this algorithm is as follows -

1. We looked at the smallest cluster sizes at both time instants, since clusters were sorted in decreasing order of size, these were $\eta_{M_{\mathcal{C}}}$ and $\eta_{M_{\mathcal{C}'}}$. We proceeded if and only if

$$\max\{\eta_{M_{\mathcal{C}}}, \eta_{M_{\mathcal{C}'}}\} \leq \eta_{small} \tag{18}$$

2. If the above test passeed we looked at the similarity data of each cluster at t with those

in $t+1$, by looking at the ownership scores $\mathscr{W}(\mathcal{C}_{\alpha'}(t+1)|\mathcal{C}_{\alpha}(t))$ This gave a matrix with some zeros. Nor for a given α' we calculated the maximum of all $\mathscr{W}(\mathcal{C}_{\alpha'}(t+1)|\mathcal{C}_{\alpha}(t))$, lets say that it happens at $\alpha = \beta$. we checked if

$$|\eta_{\alpha'}(t+1) - \eta_{\beta}(t)| \leq \eta_{small} \tag{19}$$

Then we did this for clusters at $t+1$.

3. If the answer was yes for all clusters then the mixed event could be treated as a fluctuations and for all practical purpose we could equate two clusters at different time

$$\mathcal{C}_{\alpha'}(t+1) \simeq \mathcal{C}_{\beta}(t) \tag{20}$$

4. If the there was at least one no, then it was not a fluctuation event.

In the case of fusion events, we can consider them as fluctuations if the conditions $\eta_{M_C} \leq \eta_{small}$ and $M_{\mathcal{C}} < M_{small}$ are met. Here, M_{small} represents a cutoff for the number of clusters involved, ensuring that all fusion events within this limit can be considered as fluctuations. For instance, if $M_{\mathcal{C}} = 2$, meaning only one small cluster joined a larger cluster, then the event is unquestionably a fluctuation. Similarly, if $M_{\mathcal{C}} = 3$, there were two such fluctuation events. However, if $M_{\mathcal{C}} = 6$ and $\eta_{M_{\mathcal{C}}} \leq \eta_{small} = 2$, then too many clusters were fusing in a single event, and the event would not be classified as a fluctuation. Hence, it appears that $M_{\mathcal{C}} = 2$ or $M_{\mathcal{C}} = 3$ would be appropriate, but beyond that, the event might not be considered a fluctuation.

For the fission event the idea was similar to fusion just the condition was $\eta_{M'_{\mathcal{C}}} \leq \eta_{small}$ and $M'_{\mathcal{C}} < M_{small}$.

When for all practical purpose nothing was happening we called that time to be $t_{assembly}$. In other words the following statement is true -

For all $t > t_{assembly}$, for all intervals in that time regime, all events were no-change

events.

To determine $t_{assembly}$, the algorithm was quite straightforward. It began by starting from the end time point $t = t_{final}$ and checked whether all fusion, fission, or mixed events could be classified as no-change events for $t < t_{final}$. The algorithm continued until it encountered an event that was not considered a fluctuation. At that point, the algorithm halted, and the corresponding time $t_{assembly}$ was identified as the moment when the first non-fluctuation event occurred.

5 Computational Details

Table 1: Simulation systems

peptide	number of peptides	box length (nm)	concentration (mg/ml)	number of water molecules	total number of atoms	run length (ns)
FFF	600	25	29.30	494202	1520406	400
LLL	600	20	44.52	238635	751905	1000
III	600	20	44.52	238050	750150	900

Phenylalanine tripeptide, Leucine tripeptide and Isoleucine tripeptide in their L form(Figure 1), are taken in a cubic water box and we went for computational experiment with these peptide-water systems. The details of all these systems of our interest are given in Table 1. All atom MD simulations of all the peptide water systems were carried by the NAMD 2.13[49,50] software over multiple graphics processing unit (GPU) nodes[51] and by the OpenMM package[52] on a single node with multiple GPUs. Periodic boundary conditions were implimeneted in all three directions. The interaction force field parameters were taken from CHARMM36[53] force field and TIP3P[54] water model. A cut-off of 12 Å with a switching function at 10 Å is used for short range force. Long-ranged electrostatic interactions were handled by the particle mesh Ewald method with a grid-spacing of 1 Å. For the integration the timestep was 2 fs, and the RATTLE algorithm is used for implemanting constraints. For temperature and pressure control, the Langevin thermostat and Langevin piston methods were used,

respectively. For each system at first water box is prepared and equilbrated with constant pressure and temperature(NPT) MD simulation then peptide molecules are inserted inside the box randomly. Then the peptide-water total systems were minimized with constant volume and constant temperature(NVT) MD simulation restraining all the atoms of peptides except hydrogen atom. After that to get proper equilibrated box the system was run for 5ns under NPT condition removing all restraints. After the complete equilibration was done, the NPT production run was carried out until complete self-assembled structures are formed.[40] To handle the periodic boundary condition we took the help of algorithm mentioned in our previous paper.[40]

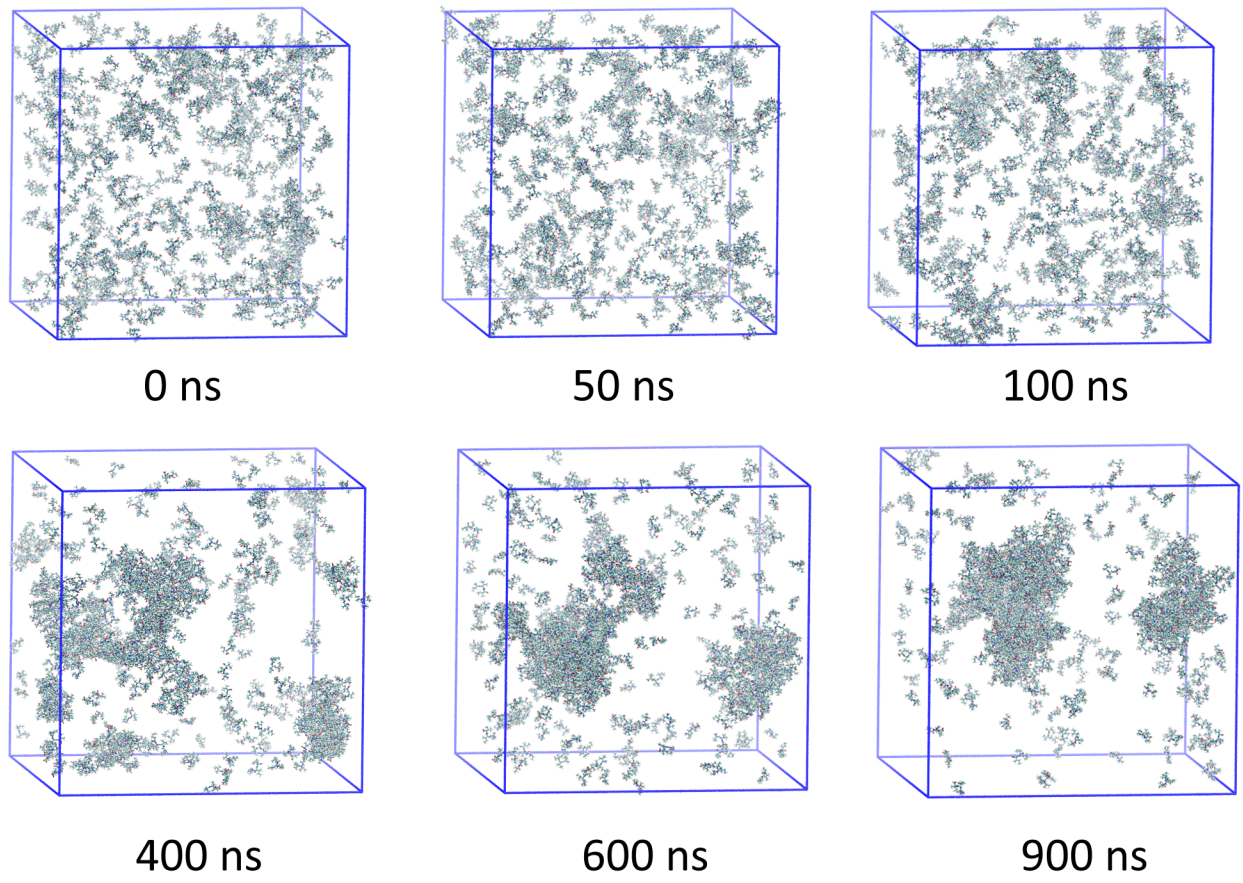


Figure 4: Time evolution Isoleucine tripeptide self-assembly

We stored the MD simulation data with a timestep of 1ns and calculated all possible peptide peptide contact with a cut-off distance of 3 Å using MDAnalysis[55] packages implimented

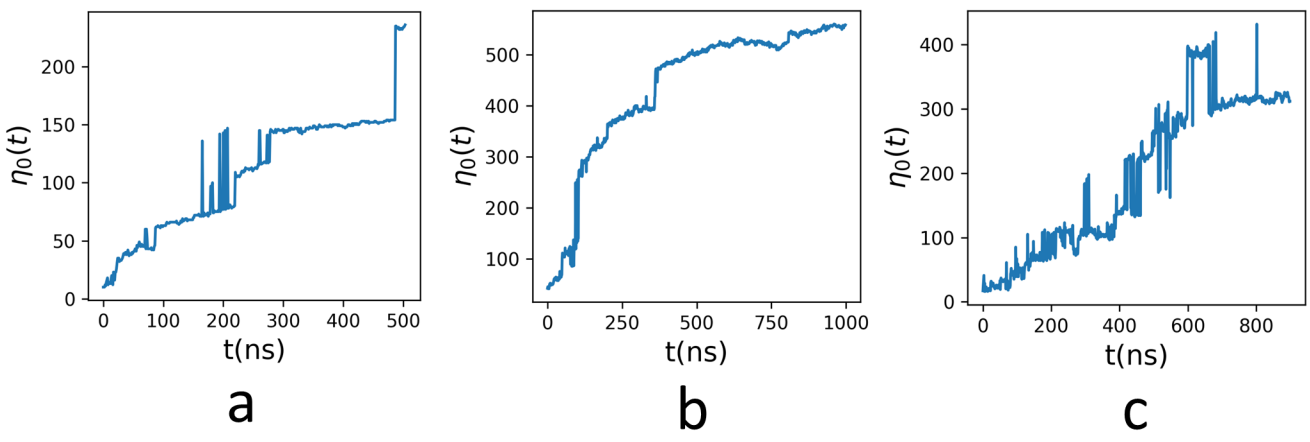
in scripts written in Python. After calculating the all possible peptide peptide contact we calculted the neighbour molecules of each molecule which are with in the cut-off distance. Taking the neighbour molecules we applied our distance based clustering algorithm which was discussed in a previous section. In the final step we applied our event-analysis algorithm to find out all possible events between two consecutive timestep. All algorithm scripts are written in Python. System preparation, analysis and image productions were performed in VMD[56] and in house code written in Tcl and Python.

6 Results and discussions

We simulated three peptide-water systems with 600 peptides of FFF,LLL and III (See Section 5 for further details of the systems). The concentration of the systems are quite a bit higher than the experimental concentration most relevent to our simulated systems.[57,58] Peptides were in their zwitter ionic form(Figure 1) and they were placed randomly inside the water box. Post-equilibration, we executed molecular dynamics (MD) simulations while meticulously maintaining a constant temperature of 300 K and pressure of 1 atm. This controlled environment facilitated the observation and analysis of the dynamic behavior of the peptide-water system until the attainment of its final self-assembled states. Importantly, these final states exhibited remarkable stability, sustaining for an extended period of over 100 ns, during which only minimal fluctuations were observed.

Following the completion of the Molecular Dynamics (MD) run, we applied our clustering algorithm, as detailed in Section 2. Subsequently, we distributed the clusters and plotted their sizes over time, as illustrated in Figure 6. Figure 5 depicts the evolution of the largest cluster for three peptides. Specifically, subfigures 5a, 5b and 5c correspond to FFF,LLL and III. At the onset of the FFF trajectory, the evolution of the largest cluster appears linear, suggesting a ripening mechanism wherein single molecules or smaller clusters fuse with this prominent cluster. Subsequently, distinct step like evolution in the cluster formation become

Figure 5: Evolution of largest cluster with time. The x axis represents time in nanosecond and in the y axis $\eta_0(t)$ stands for size of largest cluster at time t



evident, indicating the merging of larger clusters with this particular cluster. In subfigure 5b, representing the peptide sequence LLL, a linear region is evident during the initial approximately 100 nanoseconds. Following this period, there are noticeable rapid increases in cluster size. Conversely, in the case of III (Figure 5c), these rapid increase steps are not as pronounced as observed in FFF and LLL. Instead, there is a more prominent presence of linear growth in the case of III.

Upon examining the cluster size distribution for FFF (Figure 6a), it becomes apparent that in the initial timeframes, there is a higher count of smaller clusters, a reduced count of medium-sized clusters, and a complete absence of larger clusters. This trend is also observable in the cases of LLL (Figure 6b) and III (Figure 6c). This pattern is expected since, initially, all molecules exist as stray entities or in very small clusters, gradually evolving into larger clusters over time. Furthermore, in Figure 6a, it is noticeable that each cluster size range exhibits a peak value in its count. Interestingly, these peaks subtly shift with time as the cluster size range increases. This observation implies a hierarchical process in the formation of the final fiber, where clusters of a specific size range originate from the immediately preceding size range. This signature is slightly present in the case of LLL(6b) but absolutely not present in III(Figure 6c).

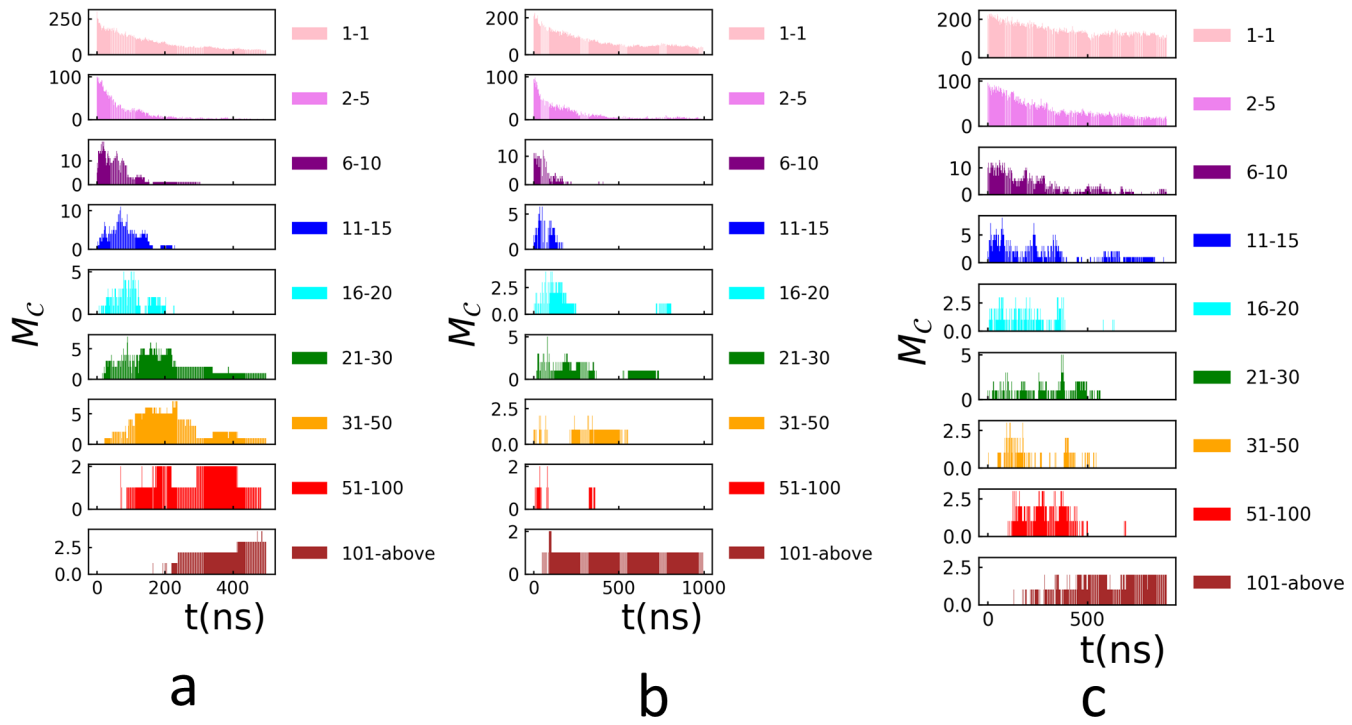


Figure 6: Cluster size distribution for FFF(a),LLL(b) and III(c). The X-axis of the plots represent time in nanosecond and the y axis of the plots represent number of clusters(M_C). The colors in the right of every plot represents size range of the cluster.

In the case of FFF, two significant bent fibers were generated, one measuring approximately 15 nanometers in length and the other approximately 8 nanometers. Meanwhile, the LLL system produced a substantial fiber measuring around 16 nanometers. In contrast, the III system resulted in the formation of a self-assembled state characterized by a length of approximately 11 nanometers on one side and approximately 10 nanometers on the other side. Our interest was to find out process by which this self-assembled states are formed. For that we applied our event analysis algorithm and found out the pure fusion, pure fission and mixed fusion-fission events present in the mechanism.

The first row of the Figure 7 illustrates pure fusion events occurring in phenylalanine tripeptide, leucine tripeptide, and isoleucine tripeptide systems. Specifically, subfigures 7a, 7b, and 7c correspond to FFF, LLL, and III, respectively. Conversely, the second row of Figure 7 depicts pure fission events for the three peptides. Notably, the labels in this

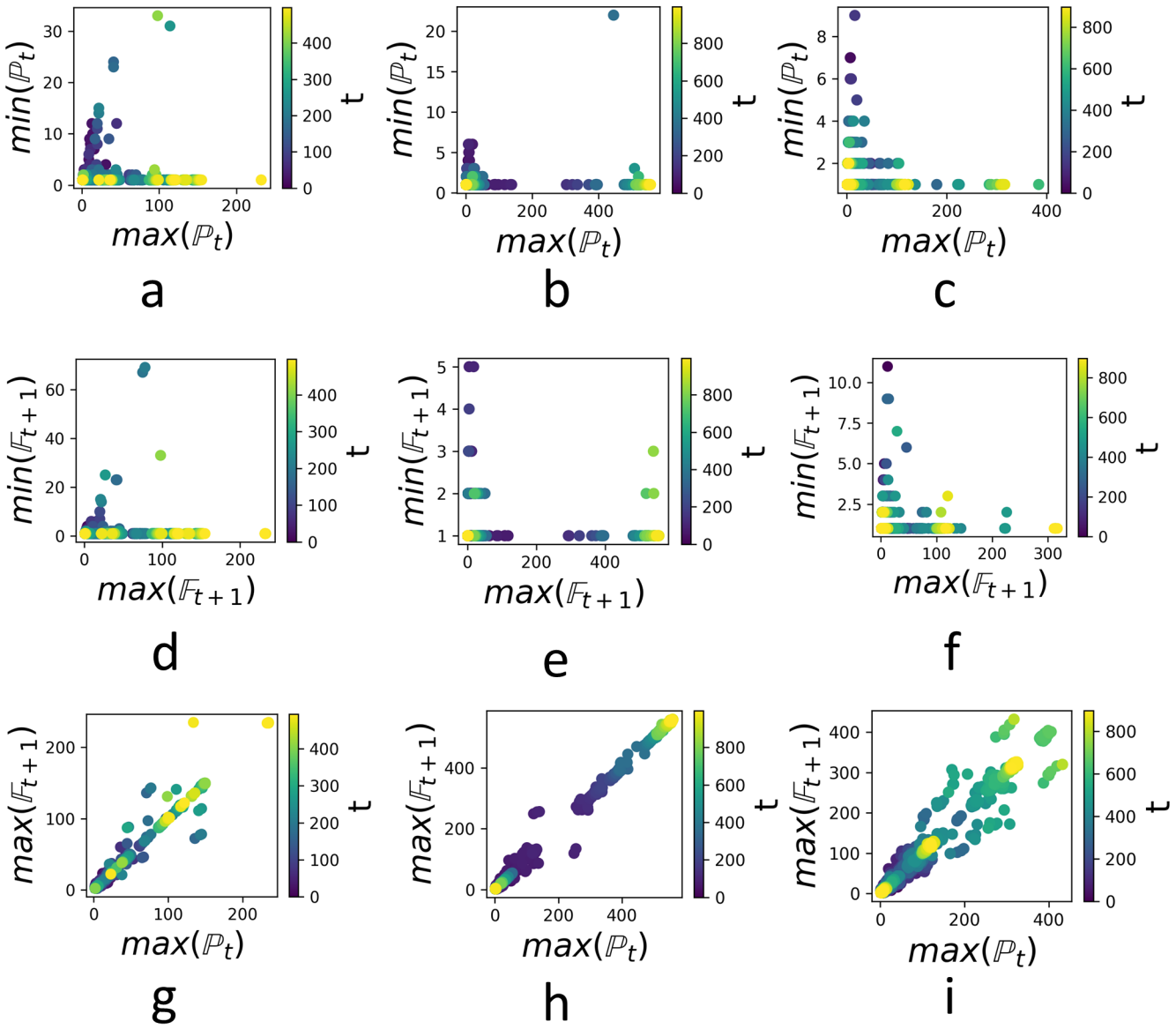


Figure 7: (a),(b) and (c) represent pure fusion events of FFF,LLL and III respectively whereas (d),(e) and (f) represent pure fission events of FFF,LLL and III respectively. The other subfigures are for mixed events. The subfigures (g) & (j) are for FFF, subfigures (h),(i) and (l) are for FFF,LLL and III respectively. $\mathbb{P}_t$ and $\mathbb{F}_{t+1}$ represent the initial and final phase cluster size sets for a particular event. The colorbar at right side of the each plot represent time in nanosecond.

figure closely mirror those in the fusion events figure. Here, the subfigures 7d, 7e, and 7f correspond to FFF, LLL, and III, respectively. The last row of the 7 is for mixed fusion-fission events. The subfigures 7g is for mixed events of FFF. Subsequently 7h is for LLL and

7i is for III. $\mathbb{P}_t$ is set of cluster size at t^{th} frame and $\mathbb{F}_{t+1}$ is set of cluster size at $(t+1)^{th}$ frame of a particular event. $min(\mathbb{P}_t)$ and $max(\mathbb{P}_t)$ are the smallest and biggest cluster of tth frame. Whereas $min(\mathbb{F}_{t+1})$ and $max(\mathbb{F}_{t+1})$ are the smallest and biggest cluster of $(t+1)^{th}$ frame. During fusion events between two consecutive timesteps, there is a notable difference in the number of clusters. At the previous timestep, there are multiple clusters, but in the subsequent timestep, only one cluster remains. In the context of these fusion events, $max(\mathbb{P}t)$ and $min(\mathbb{P}t)$ represent the largest and smallest clusters, respectively, among these multiple clusters. For the pure fusion events of FFF (Figure 7a), it is observed that predominantly smaller clusters or isolated molecules are merging with larger clusters. However, it is also evident that some clusters are fusing with others of the same size, indicating a clear hierarchy in the assembly mechanism of FFF. Contrastingly, when examining the fusion events of LLL (Figure 7b) and III (Figure 7c), the pattern is distinct. In these cases, only smaller clusters or single molecules are observed to fuse with larger clusters. Specifically, for LLL (Figure 7b), only one medium-sized cluster is observed to fuse with a larger cluster. Notably, upon careful inspection of the y-axis ($min(\mathbb{P}_t)$), there is an absence of medium or larger clusters. The addition of smaller clusters or single molecules in these fusion events suggests a growth-type mechanism.

Now for the mixed events, in the previous timestep there will be more than one cluster and for the later timestep also there will be more than one cluster. Now if we compare largest cluster of these clusters of consecutive timestep we can get an idea about the change of the largest cluster in mixed event. So in the subfigures 7g,7h and 7i the diagonal straight line of the plot represents no change in the size of largest cluster. But there are some points which are not present on the diagonal straight line and for these cases there is change in largest cluster size. For the points which are above the diagonal line there is increase in cluster size and these are definitely fusion events. In the same way the points which are below the diagonal straight line are for fission events as cluster size decreases in this case. And also if we draw a straight line from the point to the diagonal straight line, the length of that line is

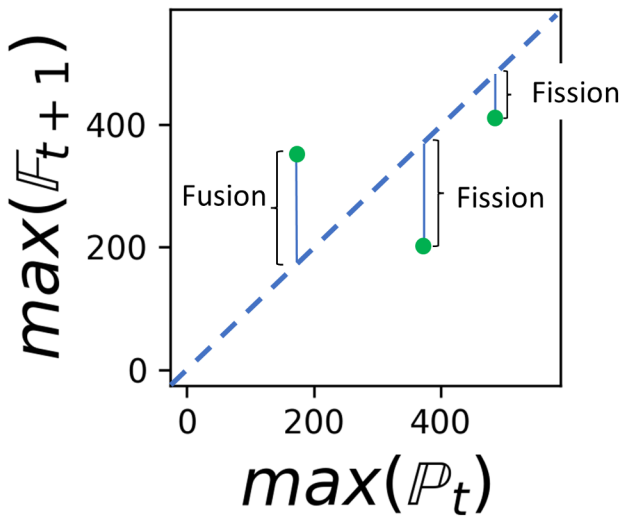


Figure 8: Further classification of mixed events. The points above the diagonal straight line actualy represent fusion and the points below the diagonal straight line represents fission

change of the size of the cluster(Figure 8). So if we look at the mixed events for FFF(Figure 7g) there is a point for which change of cluster size is large this actually represnt the large jump that was present in the very later stage of Figure 5a. In the case of mixed events of LLL (Figure 7h) there are very few points which are daviated from the diagonal straight line and also the points which are daviated corresponds very small change into cluster size. But in the case of III(Figure 7i) there are many points which are not located on the diagonal straight line but for all these cases the change of cluster size is not much.

Fission events should not be important in assembly as it is not associated with evolution of the larger cluster. But still we are interested in fission event to check stability of a cluster. If a cluster takes part in fission event then it is definitely not stable. It is interesting fact that almost all the clusters take part in fission. Some of them are fluctuations and the rests are not stable clusters. But some clusters take part in fission after sufficient time. That means the cluster is stable atleast for that time region.

The formation of largest fiber of FFF system is shown in Figure 9. Whereas the Figure 10 represents the formation of largest self-assembled state of LLL(Figure 10a) and III(Figure 10b) system. These events are observed by our event analysis algorithm and it is clear from

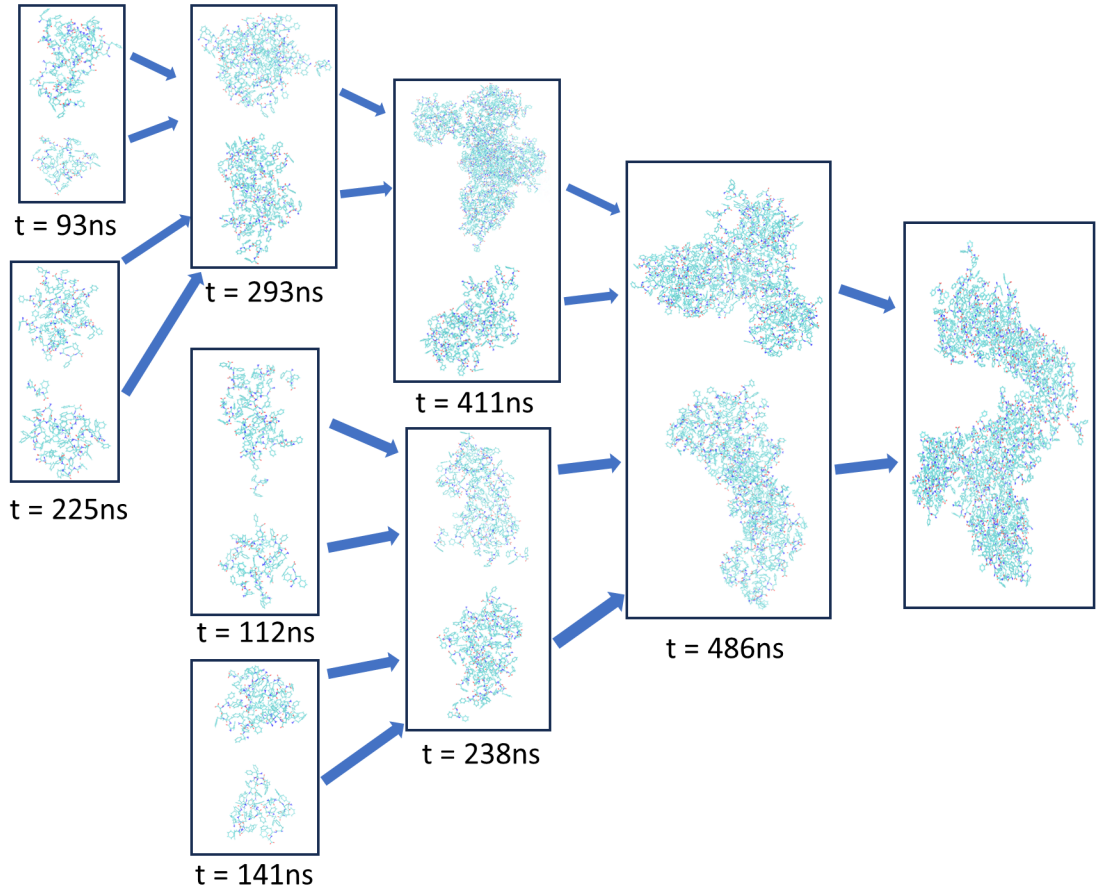


Figure 9: Formation of largest fiber of FFF system from its smaller aggregates

this figures that the mechanism of self-assembly of FFF is most hierarchical. Self-assembly of LLL system is intermediate and self-assembly of III system is least hierarchical in between these three.

The mixed events can be an artifact of duration of storing our simulation data. The timestep of our molecular dynamics integration was 2 femtosecond whereas we were storing our data in the timestep of 1 nanosecond. It might happen that there are fusion and fission events present in this 1 nanosecond time gap and we are defining them mixed event. To prove this we ran another simulation of taking FFF and maintaining the same concentration but this time we stored the date in 10 times smaller timestep. In this case we observed that the number of mixed events were significantly less. This results suggest our analogy. (Results in supplementary information)

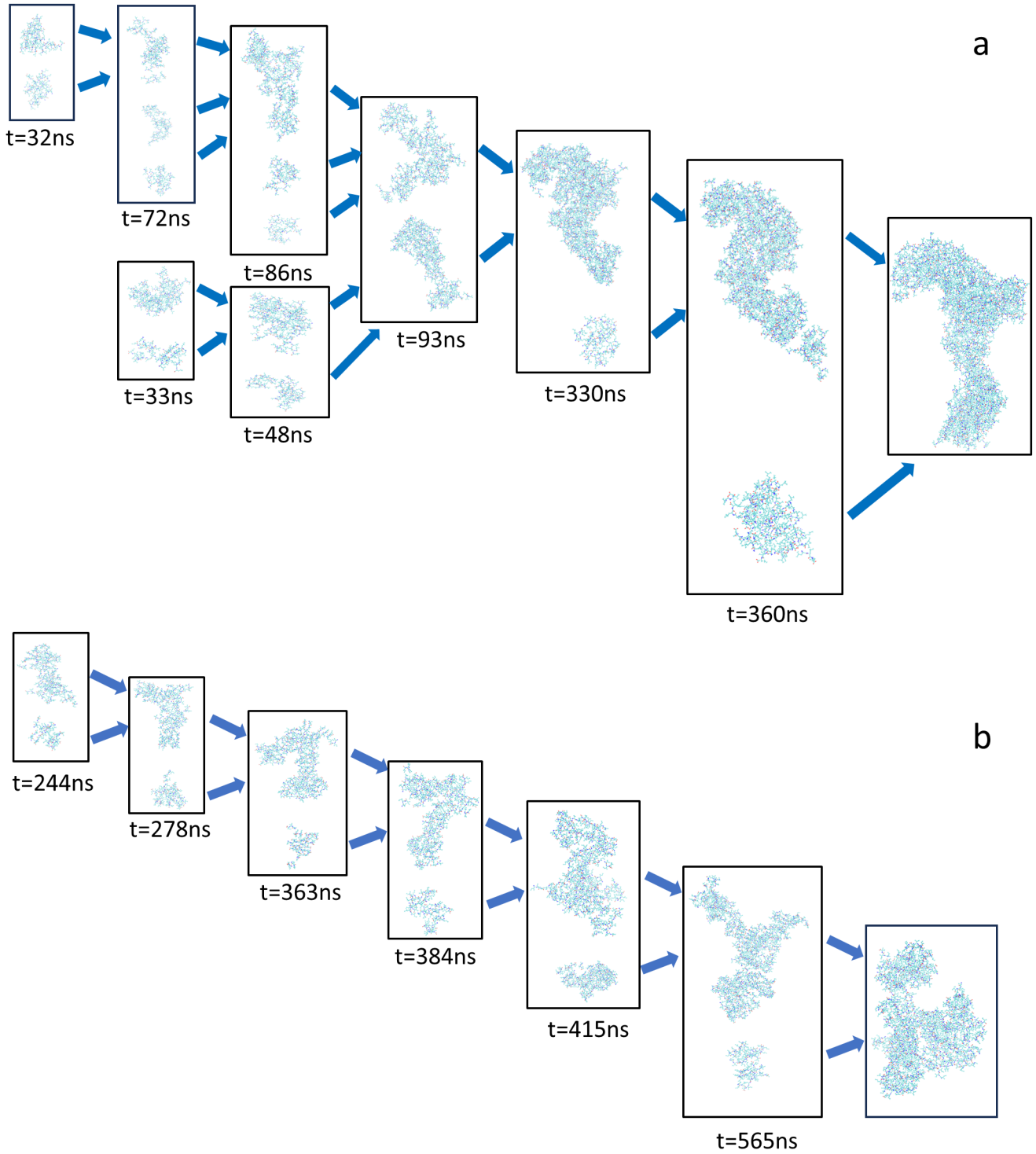


Figure 10: Formation of largest self-assembled states of (a)LLL and (b)III system from its smaller aggregates

To test our event analysis algorithm we applied the algorithm to some implicit solvents model simulation trajectory. In this case we have chosen phenylalanine dipeptide(FF) and phenylalaninetripeptide(FFF) for this case. In the case of FF a classical nucleation and growth type mechanism is observed and for the FFF the mechanism is hierarchical as it was in the case of explicit solvent simulation trajectory. Our alogorithm captured the growth mechanism and all the fusion events were either single molecule fusion or small cluster fusion. As for the implicit solvent systems timestep of storing data is very less the number of mixed events is significantly less. (Simulation methods and result in supplementary information)

7 Conclusion

We simulated three different peptide-water system by all-atom MD simulation and ended up self-assembled form in all the cases. But the mechanism of formation of these fibers was different. Our clustering and event analysis algorithm efficiently detect all the possible molecular events and with the details of molecular events we could easyly differentiate the mechanism of self-assembly of three peptides. There is no doubt that for all three peptide mechanism is hierarchical as non-classical nucleation is present in all the cases. All the nucleous ripened up by addition of single molecule or smaller clusters. Then they fused among each other. But the level of hierarchy present in the case of FFF, that is not present in case of LLL and III. In other words mechanism of FFF assembly is most hierarchical whereas mechanism of III assembly is least hierarchical. Hierarchy of assembly of LLL lies in between FFF and III.

On the other hand it is evident that our algorithm can easly determine the mechanism of formation of any self-assembled state. Not-only the case of self-assembly this algorithm can be helpful in any case where these four events are present.

Acknowledgement

The research was supported by the DST-SERB Ramanujan Fellowship (SB/S2/RJN-129/2016), DST-SERB project (SERB CRG/2019/006418) and a start-up grant from Indian Association for the Cultivation of Science (IACS) to A.D. A.C.,R.A. and S.D. acknowledges IACS for financial support. Computational resources were provided by the IACS Central Computing Cluster and the Cray supercomputing facility.

References

(1) Hamley, I. W. Small bioactive peptides for biomaterials design and therapeutics. *Chemical reviews* **2017**, *117*, 14015–14041.

(2) Hendricks, M. P.; Sato, K.; Palmer, L. C.; Stupp, S. I. Supramolecular assembly of peptide amphiphiles. *Accounts of chemical research* **2017**, *50*, 2440–2448.

(3) Lampel, A.; Ulijn, R.; Tuttle, T. Guiding principles for peptide nanotechnology through directed discovery. *Chemical Society Reviews* **2018**, *47*, 3737–3758.

(4) Makam, P.; Gazit, E. Minimalistic peptide supramolecular co-assembly: expanding the conformational space for nanotechnology. *Chemical Society Reviews* **2018**, *47*, 3406–3420.

(5) Hu, X.; Liao, M.; Gong, H.; Zhang, L.; Cox, H.; Waigh, T. A.; Lu, J. R. Recent advances in short peptide self-assembly: from rational design to novel applications. *Current Opinion in Colloid & Interface Science* **2020**, *45*, 1–13.

(6) Whitesides, G. M.; Mathias, J. P.; Seto, C. T. Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures. *Science* **1991**, *254*, 1312–1319.

(7) Zhang, S.; Holmes, T.; Lockshin, C.; Rich, A. Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane. *Proceedings of the National Academy of Sciences* **1993**, *90*, 3334–3338.

(8) Ghadiri, M. R.; Granja, J. R.; Milligan, R. A.; McRee, D. E.; Khazanovich, N. Self-assembling organic nanotubes based on a cyclic peptide architecture. *Nature* **1993**, *366*, 324–327.

(9) Ghadiri, M. R.; Granja, J. R.; Buehler, L. K. Artificial transmembrane ion channels from self-assembling peptide nanotubes. *Nature* **1994**, *369*, 301–304.

(10) Reches, M.; Gazit, E. Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science* **2003**, *300*, 625–627.

(11) Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **2005**, *437*, 640–647.

(12) Han, S.; Cao, S.; Wang, Y.; Wang, J.; Xia, D.; Xu, H.; Zhao, X.; Lu, J. R. Self-assembly of short peptide amphiphiles: the cooperative effect of hydrophobic interaction and hydrogen bonding. *Chemistry–A European Journal* **2011**, *17*, 13095–13102.

(13) Jayawarna, V.; Ali, M.; Jowitt, T. A.; Miller, A. F.; Saiani, A.; Gough, J. E.; Ulijn, R. V. Nanostructured hydrogels for three-dimensional cell culture through self-assembly of fluorenylmethoxycarbonyl–dipeptides. *Advanced materials* **2006**, *18*, 611–614.

(14) Tamamis, P.; Adler-Abramovich, L.; Reches, M.; Marshall, K.; Sikorski, P.; Serpell, L.; Gazit, E.; Archontis, G. Self-assembly of phenylalanine oligopeptides: insights from experiments and simulations. *Biophysical journal* **2009**, *96*, 5020–5029.

(15) Bera, S.; Mondal, S.; Xue, B.; Shimon, L. J.; Cao, Y.; Gazit, E. Rigid helical-like assemblies from a self-aggregating tripeptide. *Nature Materials* **2019**, *18*, 503–509.

(16) Das, R.; Gayakvad, B.; Shinde, S. D.; Rani, J.; Jain, A.; Sahu, B. Ultrashort PeptidesA Glimpse into the Structural Modifications and Their Applications as Biomaterials. *ACS Applied Bio Materials* **2020**, *3*, 5474–5499.

(17) Wang, J.; Liu, K.; Xing, R.; Yan, X. Peptide self-assembly: thermodynamics and kinetics. *Chemical Society Reviews* **2016**, *45*, 5589–5604.

(18) Mason, T. O.; Michaels, T. C.; Levin, A.; Dobson, C. M.; Gazit, E.; Knowles, T. P.; Buell, A. K. Thermodynamics of polypeptide supramolecular assembly in the short-chain limit. *Journal of the American Chemical Society* **2017**, *139*, 16134–16142.

(19) Turkevich, J.; Stevenson, P. C.; Hillier, J. A study of the nucleation and growth processes in the synthesis of colloidal gold. *Discussions of the Faraday Society* **1951**, *11*, 55–75.

(20) Levine, S.; Dube, G. Interaction between two hydrophobic colloidal particles, using the approximate Debye-Hückel theory. I. General properties. *Transactions of the Faraday Society* **1939**, *35*, 1125–1140.

(21) Puntes, V.; Bastus, N.; Pagonabarraga, I.; Iglesias, O.; Labarta, A.; Batlle, X. Nucleation phenomenon in nanoparticle self-assemblies. *International journal of nanotechnology* **2005**, *2*, 62–70.

(22) Hecht, F. M.; Bausch, A. R. Kinetically guided colloidal structure formation. *Proceedings of the National Academy of Sciences* **2016**, *113*, 8577–8582.

(23) Polte, J. Fundamental growth principles of colloidal metal nanoparticles–a new perspective. *CrystEngComm* **2015**, *17*, 6809–6830.

(24) Jun, Y.-S.; Kim, D.; Neil, C. W. Heterogeneous nucleation and growth of nanoparticles at environmental interfaces. *Accounts of chemical research* **2016**, *49*, 1681–1690.

(25) Thanh, N. T.; Maclean, N.; Mahiddine, S. Mechanisms of nucleation and growth of nanoparticles in solution. *Chemical reviews* **2014**, *114*, 7610–7630.

(26) Whitelam, S.; Jack, R. L. The statistical mechanics of dynamic pathways to self-assembly. *Annual review of physical chemistry* **2015**, *66*, 143–163.

(27) Tsonchev, S.; Troisi, A.; Schatz, G. C.; Ratner, M. A. All-atom numerical studies of self-assembly of zwitterionic peptide amphiphiles. *The Journal of Physical Chemistry B* **2004**, *108*, 15278–15284.

(28) Gnanakaran, S.; Nussinov, R.; García, A. E. Atomic-level description of amyloid β-dimer formation. *Journal of the American Chemical Society* **2006**, *128*, 2158–2159.

(29) Khurana, E.; Nielsen, S. O.; Ensing, B.; Klein, M. L. Self-assembling cyclic peptides: molecular dynamics studies of dimers in polar and nonpolar solvents. *The Journal of Physical Chemistry B* **2006**, *110*, 18965–18972.

(30) Bellesia, G.; Shea, J.-E. Self-assembly of β-sheet forming peptides into chiral fibrillar aggregates. *The Journal of chemical physics* **2007**, *126*.

(31) Krone, M. G.; Hua, L.; Soto, P.; Zhou, R.; Berne, B.; Shea, J.-E. Role of water in mediating the assembly of Alzheimer amyloid-β Aβ16- 22 protofilaments. *Journal of the American Chemical Society* **2008**, *130*, 11066–11072.

(32) Velichko, Y. S.; Stupp, S. I.; De La Cruz, M. O. Molecular simulation study of peptide amphiphile self-assembly. *The journal of physical chemistry B* **2008**, *112*, 2326–2334.

(33) Bellesia, G.; Shea, J.-E. What determines the structure and stability of KFFE monomers, dimers, and protofibrils? *Biophysical journal* **2009**, *96*, 875–886.

(34) Lee, O.-S.; Stupp, S. I.; Schatz, G. C. Atomistic molecular dynamics simulations of peptide amphiphile self-assembly into cylindrical nanofibers. *Journal of the American Chemical Society* **2011**, *133*, 3677–3683.

(35) Straub, J. E.; Thirumalai, D. Toward a molecular theory of early and late events in monomer to amyloid fibril formation. *Annual review of physical chemistry* **2011**, *62*, 437–463.

(36) Yu, T.; Schatz, G. C. Free-energy landscape for peptide amphiphile self-assembly: stepwise versus continuous assembly mechanisms. *The Journal of Physical Chemistry B* **2013**, *117*, 14059–14064.

(37) Morriss-Andrews, A.; Shea, J.-E. Simulations of protein aggregation: insights from atomistic and coarse-grained models. *The Journal of Physical Chemistry Letters* **2014**, *5*, 1899–1908.

(38) Manandhar, A.; Kang, M.; Chakraborty, K.; Tang, P. K.; Loverde, S. M. Molecular simulations of peptide amphiphiles. *Organic & biomolecular chemistry* **2017**, *15*, 7993–8005.

(39) Yuan, C.; Li, S.; Zou, Q.; Ren, Y.; Yan, X. Multiscale simulations for understanding the evolution and mechanism of hierarchical peptide self-assembly. *Physical Chemistry Chemical Physics* **2017**, *19*, 23614–23631.

(40) Adhikary, R.; Das, A. Atomistic Pictures of Self-Assembled Helical Peptide Nanofibers. *The Journal of Physical Chemistry B* **2022**, *126*, 9476–9492.

(41) Frederix, P. W.; Ulijn, R. V.; Hunt, N. T.; Tuttle, T. Virtual screening for dipeptide aggregation: toward predictive tools for peptide self-assembly. *The journal of physical chemistry letters* **2011**, *2*, 2380–2384.

(42) Guo, C.; Luo, Y.; Zhou, R.; Wei, G. Probing the self-assembly mechanism of diphenylalanine-based peptide nanovesicles and nanotubes. *ACS nano* **2012**, *6*, 3907–3918.

(43) Azuri, I.; Adler-Abramovich, L.; Gazit, E.; Hod, O.; Kronik, L. Why are diphenylalanine-based peptide nanostructures so rigid? Insights from first principles calculations. *Journal of the American Chemical Society* **2014**, *136*, 963–969.

(44) Frederix, P. W.; Scott, G. G.; Abul-Haija, Y. M.; Kalafatovic, D.; Pappas, C. G.; Javid, N.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. Exploring the sequence space for (tri-) peptide self-assembly to design and discover new hydrogels. *Nature chemistry* **2015**, *7*, 30–37.

(45) Sasselli, I.; Moreira, I.; Ulijn, R.; Tuttle, T. Molecular dynamics simulations reveal disruptive self-assembly in dynamic peptide libraries. *Organic & biomolecular chemistry* **2017**, *15*, 6541–6547.

(46) Rissanou, A. N.; Simatos, G.; Siachouli, P.; Harmandaris, V.; Mitraki, A. Self-assembly of alanine-isoleucine and isoleucine-isoleucine dipeptides through atomistic simulations and experiments. *The Journal of Physical Chemistry B* **2020**, *124*, 7102–7114.

(47) Zhou, P.; Deng, L.; Wang, Y.; Lu, J. R.; Xu, H. Interplay between intrinsic conformational propensities and intermolecular interactions in the self-assembly of short surfactant-like peptides composed of leucine/isoleucine. *Langmuir* **2016**, *32*, 4662–4672.

(48) Xiong, Q.; Liu, Z.; Han, W. Sequence-Dependent Nanofiber Structures of Phenylalanine and Isoleucine Tripeptides. *International Journal of Molecular Sciences* **2020**, *21*, 8431.

(49) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **2005**, *26*, 1781–1802.

(50) Phillips, J. C.; Hardy, D. J.; Maia, J. D.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of chemical physics* **2020**, *153*.

(51) Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. Accelerating molecular modeling applications with graphics processors. *Journal of computational chemistry* **2007**, *28*, 2618–2640.

(52) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology* **2017**, *13*, e1005659.

(53) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell Jr, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi1$ and $\chi2$ dihedral angles. *Journal of chemical theory and computation* **2012**, *8*, 3257–3273.

(54) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983**, *79*, 926–935.

(55) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry 32*, 2319–2327.

(56) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *Journal of molecular graphics* **1996**, *14*, 33–38.

(57) Rampulla, D. M.; Oncel, N.; Malcolm, S. A.; Bernasek, S. L. Higher-order complexity through R-group effects in self-assembled tripeptide monolayers. *Langmuir* **2010**, *26*, 16287–16290.

(58) Mayans, E.; Casanovas, J.; Gil, A. M.; Jiménez, A. I.; Cativiela, C.; Puiggalí, J.; Alemán, C. Diversity and hierarchy in supramolecular assemblies of triphenylalanine: from laminated helical ribbons to toroids. *Langmuir* **2017**, *33*, 4036–4048.