Mechanisms of Biomolecular Self-Assembly Investigated Through *In Situ* Observations of Structures and Dynamics

Sakshi Yadav Schmid,[1, +] Kacper Lachowski,[2, 3, +] Huat Thart Chiang[2, +] Lilo Pozzo,[2, 3, 4, *] Jim De Yoreo,[1, 4, *] Shuai Zhang,[1, 3, 4, *]

[1]Physical Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354

[2]Chemical Engineering, University of Washington, Seattle, WA 98105

[3]Molecular Engineering and Sciences Institute, University of Washington, Seattle, WA 98105

[4]Materials Science and Engineering, University of Washington, Seattle, WA 98105

[+] Equal contribution

* Corresponding authors

Abstract: Biomolecular self-assembly of hierarchical materials is a precise and adaptable bottom-up approach to synthesizing across scales with considerable energy, health, environment, sustainability, and information technology applications. To achieve desired functions in biomaterials, it is essential to directly observe assembly dynamics and structural evolutions that reflect the underlying energy landscape and the assembly mechanism. This review will summarize the current understanding of biomolecular assembly mechanisms based on *in situ* characterization and discuss the broader significance and achievements of newly gained insights. In addition, we will also introduce how emerging deep learning/machine learning-based approaches, multiparametric characterization, and high-throughput methods can boost the development of biomolecular self-assembly. The objective of this review is to accelerate the development of *in*

situ characterization approaches for biomolecular self-assembly and to inspire the next generation of biomimetic materials.

1. Introduction

After billions of years of evolution, living systems have developed capabilities to solve a vast range of biological and chemical challenges, including mass transport, energy transfer and storage, and chemical transformation. These complex functionalities are enabled by biomolecules with a high information content and the hierarchical structures created from these building blocks via self-assembly[1–4]. Inspired by observations made in natural systems, researchers are working to mimic biomolecular structures and functions via self-assembly. Over the past few decades, significant progress has been made in the field of biomolecule assembly, including for nucleic acids[5], peptides[6], proteins [7,8], peptoids[9], and viruses[10]. Due to the unique properties of these biomolecules, and their controllable hierarchical structures with tunable functions, these self-assembled materials have extensive promising applications in bioengineering, energy production, catalysis, CO_2 capture, water purification, and nanotechnology[11–13].

One of the main challenges to achieving functional bioinspired materials is predicting and controlling the assembly of building units into desired structures with accuracy and efficiency. Previous studies have shown that assembled structures can be manipulated by carefully controlling internal interactions and external stimulation[8,14,15]. Furthermore, these assembled platforms already possess distinct properties and can act as a scaffold for functional components such as nanoparticles[16,17], carbon nanotubes[18], graphene[19], and polymers[20] for hybrid materials design.

To date, however, the outstanding biomimetic and bioinspired materials produced with biomolecular self-assembly have often been the product of intuition and serendipity rather than a mechanistic understanding of the processes by which living creatures make materials with incomparable accuracy in structure, excellent diversity and efficiency in function, and fast responses to environmental stimuli. This lack of understanding reflects the complexity of the underlying energy landscapes and the fact that the assembly dynamics and structural evolutions of biomolecules and bio-complexes have been poorly investigated.

Recently, biomolecular self-assembly processes have been directly observed and described at high spatial and temporal resolution, required to define design rules for such materials, through fast-paced developments of *in situ* characterization approaches, such as ensemble spectroscopic techniques based on scattering, absorption and emission of radiation, and microscopy-based techniques, including atomic force microscopy (AFM) and transmission electron microscopy (TEM). First, these techniques are a means to directly probe the kinetic and/or free energy barriers that determine the dynamics of these biological processes. Second, *in situ* characterization removes the ambiguity associated with deducing the assembly pathway by solely observing the final assembled structure (Fig. 1).

In this review, we first introduce *in situ* methods used to characterize biomolecular assembly. Second, we summarize the biomolecules, including DNA, peptides, proteins, and peptoids, widely used as building blocks as well as their assembly mechanisms and dynamics during conversion into higher-order hierarchical structures, elucidated via various *in situ* techniques. Third, we discuss emerging opportunities for boosting research in biomolecular assembly with simulations, deep learning, and machine learning, as well as the prospect of combining human intelligence with robotics and automation to advance the field. We aim to

present recent progress on biomolecular assembly with sophisticated *in situ* characterizations in a comprehensive manner (Table 1), both to accelerate the study of complex biological processes through utilizing the right ensemble of *in situ* techniques and to aid the efficient development of novel bioinspired materials.

2. Approaches to *In Situ* Characterization of Biomolecular Assembly

2.1 Small Angle X-Ray and Neutron Scattering

Small angle scattering (SAS) techniques have been very successful at investigating biomolecular self-assembly dynamics. One strength of SAS techniques is their ability to characterize the structure of biomolecular assemblies *in situ* across multiple length scales without chemically labeling the biomolecule systems of interest, with the possibility of being in-line with other techniques (e.g., size-exclusion chromatography and optical spectroscopy). The two most common SAS techniques used for *in situ* biomolecular analysis are small angle X-ray scattering (SAXS) and small angle neutron scattering (SANS). Data from SAS consists of the scattered intensity measured as a function of the scattering wavevector, q, which is a function of the scattering angle and incident wavelength[21]. The data interpretation of SAXS and SANS data is nearly identical, but the two methods differ in many ways. We refer readers to other sources for detailed descriptions of fundamental SAS theories applied to biomolecular samples[22,23].

Despite the similar theoretical foundations and data interpretation of SAXS and SANS, the strengths and weaknesses of each technique need to be considered when planning experiments. An advantage of SAXS is its relative accessibility through the numerous synchrotron beamlines around the world and laboratory SAXS instruments. High-flux synchrotron-based instruments

allow SAXS to achieve time resolution in the order of milliseconds. The ability to focus X-ray beams has enabled the use of small sample volumes, on the order of tens of μL, an important consideration in many materials-limited studies of biomolecular self-assembly. In contrast, the X-ray radiation of SAXS may damage biomolecular structures. Additionally, high flux X-rays cannot overcome poor contrast between biomolecules and solvents due to their similar scattering length densities[24].

Compared to the availability and high flux of X-ray sources, neutron sources have much lower brilliance[25] and are only accessible at a few facilities in the world. The lower flux results in a temporal resolution on the order of tens of seconds[26], lower than high-flux X-ray sources. An additional consequence of the lower flux of neutron sources is the need to collimate beams to a larger area, which requires larger sample volumes. Despite these disadvantages, SANS has several major advantages over SAXS in the study of biomolecular self-assembly: it is a non-destructive technique, the contrast of biomolecules in an aqueous solution is typically higher, and it has a substantial potential for contrast variation[24,27]. Differences in the scattering lengths of isotopes (hydrogen vs. deuterium in particular) can be leveraged to control the contrast in SANS experiments. Using purely deuterated solvents (e.g., D_2O) maximizes the contrast of hydrogenated structures. Alternatively, mixtures of deuterated and hydrogenated solvents or replacing hydrogen with deuterium in biomolecules can be used for contrast variation. In contrast variation, scattering from a structure of interest may be highlighted by matching the scattering length density of other structures to that of the solvent to reduce their contrast. This technique also allows researchers to study the dynamics of self-assembly or the mixing of biomolecules without fluorescent probes[27–29].

The work of Eves et al.[28] is an excellent example of how contrast-matched SANS experiments provide unique insights into the self-assembly mechanism of amyloid fibril elongation. Their SANS approach can distinguish between structural changes due to the average growth rate of fibrils and the formation of new fibrils. This is an improvement compared to traditional fluorescence and circular dichroism spectroscopy methods which solely capture the proportion of protein in fibrilized form and the rate of change of this proportion. The authors measured the elongation of existing fibrils using SANS by contrast matching the buffer to the hydrogenated fibril seeds and incubating the seeds with deuterated monomers of α-synuclein. The resulting deuterated extensions of the α-synuclein fibrils were fitted with a cylinder model to quantify the elongation rate (Fig. 2a-b). Time-resolved SANS with contrast variation has also been used to study lipid transfer dynamics in vesicles. Nguyen et al. studied the effect of methanol, commonly used to prepare lipid structures, on the dynamics of lipid rearrangement in the vesicle[27]. Two separate batches of vesicles were prepared using hydrogenated or deuterated lipids in a mixture of H_2O and D_2O that was contrast matched to a 1:1 mixture of hydrogenated and deuterated lipids. Spontaneous mixing of the vesicles resulted in lipid transfer, lipid flip-flop, and a decrease in scattering intensity as the average vesicle approached a composition of 1:1 hydrogenated and deuterated lipid. The dynamics of these processes were captured by modelling the decrease in scattering intensity over time. Poor temporal resolution is a primary limitation of many time-resolved SANS studies and can be improved by developing SANS instruments with higher flux and signal-to-noise ratios.

In addition to using geometric models like the previously mentioned cylinder model, small angle scattering data may also be analyzed using shape-independent models, peak analysis, and

comparison to computational results[24]. The Guinier approximation is a shape-independent method for estimating the radius of gyration of the largest structure in a sample. It was used to characterize the liquid-liquid phase separation of prion-like domains as a function of NaCl concentration, demonstrating that the nucleation time was related to the quench depth[30]. In an experiment using chaotic-flow mixing, the protein was quenched to the two-phase region of the protein–NaCl phase diagram. The SAXS curves from the first 20 ms of data collection were analyzed using the Guinier approximation and the trend in the radius of gyration was fitted with a model of exponential collapse. The similarity of the resulting time constant to another mechanism, attributed to a barrier-limited process, led the authors to suggest that the reorganization of the prion-like domain was also barrier limited. In separate experiments using laminar-flow mixing, the formation of mesoscopic aggregates at different NaCl quench depths was characterized by calculating an assembly metric related to the total volume of clusters. The broader implication of the work is that time-resolved SAS and shape-independent analysis are powerful tools for studying biomolecular self-assembly pathways and intermediates.

SAS instruments configured to measure scattered intensity at large angles are suited to characterize the formation of biomolecular crystalline structures, such as non-lamellar lipid phases, *in situ*[31]. The scattering curve of any macromolecule with a known atomic structure can be computationally calculated with software such as CRYSOL[32,33] or reciprocal grids[34] and then compared with the experimental scattering curve to identify the best matching structure. The reciprocal grid algorithm was used to simulate the different structures that the tubulin protein could assemble into (e.g. conical-spiral, conical-spiral tubules, and hexagonal bundles of inverted

tubules) at varying spermine concentrations[35]. The simulated structures were then compared to time-resolved SAXS data (Fig. 2c-e) to identify the hierarchically assembled structures.

While SAS data can contain a variety of information for studying the self-assembly of biomolecules, data analysis must be done cautiously because scattering curves are non-unique, meaning that many different solutions can satisfy the fit of a scattering curve. Additionally, model choice and fit parameters for unknown systems need support from complementary approaches such as microscopy or computational methods [32,33,36]. Nonetheless, SAS techniques are highly capable of monitoring the emergence of order in nanoscale structures with high temporal resolution (on the order of milliseconds) and can help determine the assembly pathway.

2.2 Static and Dynamic Light Scattering

Like SAS, static light scattering (SLS) and dynamic light scattering (DLS) utilize scattering that arises from inhomogeneities in a medium[24]. Light scattering occurs due to differences in the refractive index between a scatterer and its medium. A review of SLS, DLS, and other light scattering methods in biological and biopharmaceutical sciences was put together by Minton[37]. In this review, we will discuss the light scattering techniques in the context of studying the dynamics of biomolecular self-assembly.

Like SAS, SLS and multi-angle light scattering (MALS) measure the scattering intensity as a function of angle. Specifically, SLS requires scanning a detector across multiple angles, whereas MALS consists of multiple detectors at fixed angles[37]. MALS is particularly useful for studying the dynamics of biomolecular self-assembly because it enables time-resolved static light

scattering experiments with a time resolution on the order of seconds. Light scattering experiments can extract the molar mass, radius of gyration, second virial coefficient, and fractal dimension of biomolecular suspensions. Unlike SLS, in which the intensity of scattering is averaged over time, DLS measures fluctuations in scattering intensity over very short intervals, down to microseconds[38]. The dynamics of colloids, such as proteins, are determined by Brownian motion when thermal fluctuations are more significant than any other forces acting on the colloid (e.g., gravitational). In this scenario, the dynamics of scattering can be described by an autocorrelation function which can be fit to extract a distribution of diffusion coefficients. In turn, the hydrodynamic radii of spherical structures can be obtained by applying the Stokes-Einstein equation[39]. While the diffusion of structures such as micelles, globular proteins, and vesicles are well approximated using a spherical model, the approximation is inappropriate for more complex or anisotropic structures.

Recently, Narang et al.[40] characterized the liquid-liquid phase separation (LLPS) and self-assembly of FMOC-L-lysine (fluorenylmethyloxycarbonyl L-lysine) in aqueous DMSO using simultaneous time-resolved SLS and DLS. The authors showed that droplets of about 100 nm in diameter initially formed upon the addition of water to FMOC-L-lysine dissolved in DMSO, which eventually evolved into anisotropic structures. The importance of LLPS and biomolecular condensates in biomolecular self-assembly and biological functions continues to be uncovered[30,41]. Therefore, despite their limitations at resolving nanoscale changes, SLS and DLS will continue to be useful in studying structures at larger length scales, such as in LLPS and similar phenomena that may lead to order at the nanoscale. Other notable applications of SLS and DLS include the study of fibrinogen assembly as a function of ionic strength[42] and controlled vesicle

aggregation[43] (Fig. 3a-b). In practice, SLS and DLS are more widely available than SANS or SAXS and are non-destructive, *in situ* techniques that can provide information to complement or guide additional methods.

2.3 Optical Spectroscopy

Optical spectroscopy studies how matter interacts with electromagnetic radiation as a function of wavelength. The stimulation of an analyte with incident radiation elevates it from the ground state to an excited state. Spectroscopic measurements are made of either the electromagnetic emission, absorption, or scattering from the species as it returns to the ground state. The four commonly used optical spectroscopy methods are absorption, emission, luminescence, and scattering. These analytical techniques have been essential for the study of biomolecules and their structural chemistry for several decades[44]. Recently, optical spectroscopy has been used in conjunction with other *in situ* techniques. Absorption techniques such as ultraviolet-visible (UV-vis) spectroscopy are typically limited in providing information about biomolecular self-assembly. Techniques that study infrared absorption, circular dichroism, or fluorescence spectra are able to capture time-resolved structural changes of biomolecules[45,46]. In addition to quantifying the concentration of a given analyte, spectroscopic methods report on changes in the molecular conformation[46], environment[30], and bonding[47] of the analyte. Optical spectroscopy methods may also be enhanced by coupling them with other tools such as pump-probe[23], stopped-flow[47], or confocal microscopy[30].

2.3.1 Fourier Transform Infrared Spectroscopy (FTIR) and Raman Spectroscopy

Infrared (IR) light matching the resonant frequency of a bond in a molecule can lead to specific information about the chemical structure. Resonance occurs in bonds that have a vibrational mode associated with a change in the bond's dipole moment. A key design feature of most FTIR instruments is the Michelson interferometer. The Fourier transform of the resulting interferogram yields the spectra of the sample and is typically reported in units of inverse cm. In the rapid scan mode, the mirror is moved back and forth quickly to obtain spectra with time resolution on the order of ~10 ms[48].

In contrast to ultraviolet (UV) and visible light spectroscopy, the sensitivity of IR spectra to both intra- and intermolecular bonds makes it a powerful tool in studying molecular structures in biomolecular self-assembly. For example, hydrogen bonds, essential in protein structure and hydrogels, are detectable by IR spectroscopy. Additionally, FTIR can detect coordination between inorganic and biomolecular species[45]. The vibrational mode of a molecule is sensitive to both the bond strength and difference in mass between the two atoms. Selective deuteration experiments can exploit this mass dependence to probe specific bonds in a system, help with peak assignment, or study dynamics through the exchange of hydrogen and deuterium[47,49–51]. Adams et al.[47] performed time-resolved stopped-flow FTIR studies on the gelation of MAX1, a peptide designed to mimic amphiphilic β-sheet-containing peptides that is being explored in therapeutic hydrogel applications. They designed five variants, replacing one of five valines with a perdeuterated valine to separate the absorption of the C–D peak from other species in the system. Changes in the peak position and width in the dissolved and gelled state allowed the identification of differences in the

environment of each side chain after self-assembly. Fitting time-resolved changes in peak position during gelation with exponential curves, as shown in Fig. 4 a and b, yielded similar time constants for each variant and was consistent with cooperative self-assembly.

In addition, *in situ* FTIR has been used to study thermodynamic, kinetic, and structural changes associated with protein adsorption processes[52–54]. For instance, *in situ* attenuated total reflection FTIR was used to study the dynamics of changes in the conformation of bovine serum albumin adsorbing on a hematite surface in response to initial protein concentration, ionic strength, and the pH of the medium (Fig. 4c-e)[55]. In another study, *in situ* FTIR showed that Hfq CTR (E. Coli strain) formed an amyloid-like structure both in vitro and in vivo[56]. *In situ* FTIR studies of peptides and proteins are enabled by the sensitivity to changes in the intensity and position of amide I and amide II bands caused by changes in hydrogen bonding associated with secondary structure. Another method of studying protein secondary structure was recently proposed[57], where the researchers could resolve the contributions of the amide I band by subtracting the buffer and analyzing the data with a second derivative transformation and Fourier self-deconvolution. Besides practical limitations, such as inherent difficulties with the characterization of samples with non-negligible scattering or the overlap of bands that cannot be addressed by deuteration, the density of information in the IR spectra of biomacromolecules is an ongoing challenge in data analysis and an opportunity for the future.

Raman spectroscopy, in which inelastic light scattering is used to probe molecular vibrational spectra, gives insights into the changes in intermolecular bonding that accompany structural changes in biomolecules[58]. This technique is rarely used in time-resolved studies of self-assembly due to the low signal intensity associated with biomolecules that typically undergo

self-assembly. However, it has great promise because Raman spectroscopy is substantially sensitive to chromophore within a protein if the exciting laser line is in resonance with its electronic transition, even when biomolecules are otherwise optically transparent. Future time-resolved studies of self-assembly may be realized by enhancing the latter through resonance[58,59] surface-enhanced Raman techniques[60] or stimulated Raman spectroscopy (SRS). The latter technique is well suited for characterizing changes in photoactive biomolecules such as green fluorescent protein or correlating chemical signatures to structural features in SRS microscopy. While it has been employed to study protein aggregation *ex situ*[61], it could be applied to future *in situ* studies of self-assembly and provide multimodal information on the structure and chemical nature of protein aggregates in pharmaceutical formulations.

2.3.2 Circular Dichroism Spectroscopy

Circular dichroism (CD), which measures the difference in absorption between left-handed and right-handed circularly polarized light, is sensitive to molecular configuration and conformation. The electric dipole moment between an initial and final state during light absorption is called the transition dipole moment. In cases where the electric and magnetic dipole transition moments are non-zero and not perpendicular to one another, the response will depend on the polarization of the light. This response is known as rotational strength and is described by the Rosenfeld equation[62]. The rotational strength is directly proportional to the dissymmetry factor, which is the difference in the molar absorption coefficient between left and right circularly polarized light divided by the molar absorption coefficient. The nature of the transition dipole moment, and thereby rotational strength, depends on molecular structure. In turn, CD is widely

used to probe optically active (chiral) changes in biomolecular conformation during self-assembly with high sensitivity.

CD is routinely used to study changes in the secondary or tertiary structure of proteins and peptides in response to the concentration of denaturant, temperature, pH, and other factors. The feasibility of using time-resolved CD to study biomolecular self-assembly depends on the assembly rate and time needed to record a quality spectrum. In the case of proteins, CD spectra can be fit to extract the proportion of structural motifs, such as β-sheets and α-helices, present in the sample. Time-resolved CD has been used to study the self-assembly of protein structures such as amyloid fibrils[46], chiral supramolecular structures and hydrogels[47,63–65], and protein-surfactant complexes[66].

A typical CD spectrum requires several minutes to acquire. Hence, characterizing a fast assembly process requires incorporating additional techniques, such as stopped-flow methods[66]. Pump-probe techniques may be used to study photoinitiated mechanisms and can produce nanosecond temporal resolution with a narrow spectral bandwidth [67]. Synchrotron radiation CD (SRCD) is a promising approach for recording time-resolved CD more quickly with more stable far-UV radiation[68,69]. Recently, Auvray et al. have made significant progress in utilizing the natural polarization present at synchrotrons for CD measurements[23]. The authors measured photoisomerization events in an azobenzene-modified peptide during alternating pulses of light exposure to trigger *cis-to-trans* transitions and decreased the temporal resolution of broadband (120 nm to 600 nm) CD measurements by several orders of magnitude (10 mHz to 130 HZ). This allowed CD spectra to be collected with 520 μs temporal resolution (Fig. 5a-d).

Significant limitations of CD include its incompatibility with turbidity, difficulties resolving species absorbing at similar wavelengths, and its restriction to chiral molecules or supramolecular assemblies with optically active chromophores (e.g., protein backbone carbonyls and amide bonds). Moreover, CD spectra alone are insufficient for characterizing structures if the connection between the structure and CD spectra is poorly understood. Fortunately, complementary datasets of experimentally determined CD spectra and protein structures are readily accessible and enable the fitting of experimental data to help predict structures. Therefore, CD is an indispensable technique for studying self-assembly pathways where protein or peptide secondary structures change during self-assembly. It is highly complementary to SAXS, AFM, and other spectroscopic methods. Given that the function of some therapeutic peptides is connected with changes in the secondary structure, a number that is expected to increase in the future[70], CD will remain an important technique for studying biomolecular self-assembly.

2.3.3 Fluorescence Spectroscopy

Fluorescence occurs when a molecule reaches an excited state by absorbing light at a particular wavelength and then emits light at a higher wavelength due to the dissipation of energy into environment. Fluorophores are chemical species that can fluoresce and have different Stokes shifts (the difference in the wavelength of peak excitation and emission). Fluorophores may be categorized as intrinsic if they are a naturally occurring component of a biomolecule (e.g., aromatic amino acids, fluorescent protein) or extrinsic if they are fluorescent dyes or modified biomolecules not found in nature. Fluorophores may either be covalently bound or adsorbed to the biomolecule

undergoing self-assembly. Conventional fluorescence measures the emission intensity at either a specific wavelength or across a spectrum of wavelengths when excited by a single wavelength.

There are several variations of fluorescence spectroscopy methods that enable unique insight into biomolecular self-assembly, including Förster resonance energy transfer (FRET), fluorescence correlation spectroscopy (FCS), and fluorescence lifetime correlation spectroscopy (FLCS). In FRET, a donor fluorophore transfers energy to an acceptor fluorophore that has an excitation overlapping with the donor's emission spectra. This process depends on the two components being in sufficient proximity and alignment, making it particularly useful for studying self-assembled structures. FRET was used to probe the tendency of different oligopeptides to co-assemble into supramolecular structures with fluorenylmethyloxycarbonyl-diphenylalanine (Fmoc-FF) early on in the assembly process[71].

FCS records the intensity of emissions from fluorophores as a function of time with a fluorescent confocal microscope. Like DLS, the signal is analyzed using correlation functions to extract similar information (e.g., diffusion coefficients, hydrodynamic radii), but is not broadly applicable because it requires the use of a fluorophore. On the other hand, FCS and its variants have unique abilities, such as the ability to distinguish molecular complexes[72]. FLCS is an advancement of FCS, utilizing the differences in the lifetimes of fluorophores (with overlapping excitation and emission spectra) to separately track their dynamics. Martin et al.[30] used FCS to show that the viscosity increased inside droplets during liquid-liquid phase separation in proportion to the concentration of NaCl. Mañas-Torres et al.[73] proposed that Fmoc-FF followed different assembly pathways based on fitting FLCS data to extract the proportion of signal contributions from slow and fast diffusing species (Fig. 6a-b). One limitation of correlation

spectroscopy and related techniques is the need for relatively long measurement times, on the order of minutes in some cases. Bayesian non-parametric methods were used to analyze single-molecule fluorescence confocal microscopy data. They yielded similar molecular properties to those attainable by FCS, but with shorter measurement times[74]. As these studies demonstrate, fluorescence-based absorption spectroscopy, alone or in combination with other methods, enables the study of biomolecular self-assembly methods with high temporal resolution. Combined with techniques like confocal microscopy and through experimental techniques such as FRET, it can also report on structural changes in highly specific regions of interest.

Changes to fluorescent emissions may occur due to quenching, such as during the stacking of aromatic residues, or due to the adsorption of a fluorescent dye to a biomolecular structure (e.g., amyloid fibrils)[46] (Fig. 6c-e). The limitations of fluorescence spectroscopy include the difficulty of dye conjugation to certain biomolecules and the fact that fluorescent imaging techniques have limited spatial resolution relative to electron microscopy or AFM. Moreover, care should be taken when evaluating whether an introduced fluorophore interferes with a biomolecular self-assembly process or unaccounted-for factors in the system interfere with the fluorescence signal[75,76]. Finally, fluorescence is limited to detecting changes in local chemical environments and is therefore unable to directly provide certain mechanistic details, such as distinguishing between fibril elongation and an increase in the total number of fibrils.

2.4 Atomic Force Microscopy

AFM was invented in 1986[77] and early efforts to image the biological samples were made with contact mode AFM (CM-AFM). In CM-AFM, a cantilever with a very sharp probe (with a radius on the order of nanometers) is in continuous contact with the substrate in a raster scanning motion, while the vertical and horizontal motions of the cantilever are monitored via a laser beam reflected off the cantilever. Most commonly, CM-AFM is operated in the constant force mode using the deflection of the reflected laser beam as feedback where the aggressiveness or gentleness of the probe-sample interaction is controlled. However, due to the "dragging" motion of the probe on the sample, fragile samples are damaged by the cantilever[78].

To rectify drawbacks of CM-AFM, amplitude modulation AFM (AM-AFM) was developed and quickly became a widely used mode in biomolecular studies. In conventional AM-AFM, the cantilever is oscillated at its resonance frequency, and, after engagement, the probe-sample interactions cause the oscillation amplitude to decrease. Feedback is based on monitoring the reduced amplitude due to the probe-sample interactions. AM-AFM has much faster feedback and better efficiency in tuning probe-sample interactions than CM-AFM. It allows the characterization of fragile and sensitive biological samples with a lower imaging force and higher spatial resolution. In the last three decades, AFM has been used to obtain high-resolution images of biomolecular assemblies in physiological conditions[79]. In addition, AFM-based multiparametric methods have enabled the correlation of mechanical properties of biomolecular assemblies with the underlying structure and provided quantitative data on inter- and intra-biomolecular interactions[80]. The dynamics of biomolecules, such as DNA diffusion across a surface[81], protein assembly and phase transition[82,83], and cell movement[84] have all been

captured via time-lapse AFM. However, the temporal resolution of conventional AFM is typically in the sub-minute range, much slower than the dynamics of single biomacromolecules.

One of the pioneering biology-based research groups, Ando et al., developed high-speed AFM (HS-AFM) to visualize the dynamic structural changes and interactions occurring in biomolecular complexes[85–87]. They also directly visualized the motion of proteins, such as myosin V molecules walking along actin tracks[88] (Fig. 7a-b). This breakthrough created a pathway to study the structure and dynamics of biomolecules in action. This technique has been used to capture nuances of biomolecule dynamics in response to stimuli, including temperature[85], pH[86], and photo stimulation[87]. In addition, high-speed AFM height spectroscopy (HS-AFM-HS), inspired by fluorescence spectroscopy, can detect the motion of molecules underneath an AFM tip held at a constant x-y position. It has been used to monitor the surface concentration, diffusion rate, and oligomer size of highly mobile annexin-V molecules during their self-assembly into 2D crystals in the presence of calcium ions[89].

Recently, HS-AFM captured the dynamic process of protein nanorods assembling in 2D on a mica surface[90]. This study highlights the importance of high spatiotemporal resolution to visualize the anisotropic motion and dynamics of proteins during the assembly process to allow extraction of the underlying physical mechanisms of assembly. The adsorption and desorption of the protein nanorods and their assembly into threefold symmetric patterns as a function of protein number density were observed *in situ*. Other nuances of the assembly process such as the diffusional motion of rods—translational and rotational—and the emergence of nematic order and fiber formation by end-to-end protein interaction were also observed (Fig. 7c). Observing and quantifying these dynamic factors, which controlled the 2D assembly process, were essential for

constructing a theoretical model to explain the substrate mediated 2D-assembly of proteins. In another study, HS-AFM was used to study the assembly of hexagonal DNA origami blocks into a honeycomb lattice on a mica-supported lipid bilayer membrane[91]. The assembly process captured by HS-AFM included boundary reorganization, defect diffusion, and defect filling in the assembled honeycomb structure (Fig. 7d). Also, the honeycomb lattices made by the customized DNA monomers with different features were observed by time-resolved imaging.

These experimental studies demonstrate the versatility of AFM to visualize complex biomolecular processes in real time with atomic-level spatial and microsecond temporal resolution. However, AFM is only applicable to processes occurring at interfaces and is not a suitable technology for characterizing assembly in 3D.

2.5 Transmission Electron Microscopy

Transmission electron microscopy (TEM) is another widely used imaging-based method. Unlike optical imaging techniques, it uses electrons to probe a specimen and achieves atomic resolution in a vacuum environment. In a conventional TEM experiment, specimens must be dehydrated or fixed in resin before being characterized, making *in situ* imaging of biomolecular processes and assembly impossible. Over the last two decades, advances in microfabrication and ultrathin electron-transparent membranes have made *in situ* liquid-phase transmission electron microscopy (LP-TEM) possible. LP-TEM has been utilized to visualize the dynamic processes of nanoparticle assembly, crystallization, mineralization, phase transition, and diffusion with spatial resolution down to the atomic scale and millisecond temporal resolution[92–94].

Recent reports include pioneering studies of biomolecular crystallization, assembly, and phase transitions with LP-TEM[95–97]. LP-TEM can capture the dynamics of assembly and the formation of transient states, thus providing insights into the design of self-assembling biomimetic systems. Jin et al.[93] used *in situ* liquid phase TEM to monitor the peptoid-induced formation of novel five-fold twinned Au nanostars exhibiting five uniformly shaped star points (Fig. 8a-b). The study showed that by tuning the peptoid-peptoid and peptoid-Au interactions, the Au nanocrystal morphology could be controlled by biasing the crystallographic face on which particle attachment occurs and introducing anisotropy into the growth kinetics. Pre-formed spheroidal single-crystalline Au nanoparticles and peptoids were sealed in a TEM liquid cell and imaged continuously over time. As time progressed, particle aggregation and attachment events dominated the growth process rather than Oswald ripening, wherein small particles dissolve as large particles grow. Initially, the free-moving Au nanoparticles exhibited repeated collision and attachment events, followed by coalescence to form larger, five-fold twinned spheroidal particles. Peptoid binding to the (111) faces poisoned their growth, leading to elongation along the five [100] directions. Subsequent attachment events were biased to occur on these [100] points. Together, these effects led to the formation of the five-fold twinned nanostars, using the selected peptoid sequence as well as an appropriately adjusted pH and peptoid concentration.

However, LP-TEM still has several technical challenges. A better understanding of the adverse effects of electron beams on biomolecular samples, restrictions imposed by the volume limitations of the liquid cell, and the potential impact of confining macromolecules in the nanoscale space between the liquid cell membrane is required[98,99]. In addition, electron-induced radiolysis and illumination can significantly change the properties of the liquid environment, including pH values, and generate harmful by-products with large diffusivities[100]. Graphene

liquid cells have emerged as a less damaging alternative to conventional silicon nitride liquid cells. Graphene-encapsulated hydrated biological samples can be imaged for several minutes without any visible structural damage to the samples[101,102].

Another way around the problem of radiolytic damage to biomolecules is to use cryogenic electron microscopy (Cryo-EM)[103], which enables imaging biomolecules in their native state and obtaining high-resolution structural information. The basic principle of cryo-EM is to fix the samples in amorphous ice and obtain 2D images of the sample over an extensive range of orientations. These 2D images are then used to reconstruct a 3D image of biomolecules or biomolecular assemblies. Cryo-EM has proven suitable for imaging proteins, their large complexes, and assemblies. Moreover, multiple structural conformations can be captured in one experiment.

Single-particle cryo-EM imaging can be combined with machine learning-based data processing to classify and average 2D images into 3D maps in order to visualize complex structural properties[104,105]. First, 2D classification of the images is performed by computationally merging images of individual particles in identical (or similar) conformations. For this purpose, tens to hundreds of thousands of particle images with randomly orientated 2D projections of the particle that contain all the structural detail of the 3D structure are used. Angular orientation parameters are obtained by comparing these 2D projections with spatially defined re-projections of an initial 3D model filtered at low resolution[106]. A 3D map of the particle structure is then obtained by "back-projection" where the 2D projections are combined into a single 3D map. 3D reconstructions are obtained by iteratively performing the projection-matching process to improve the resolution until no further improvement is possible. Finally, the known biomolecular sequence is fitted into the 3D map and the constructed 3D atomic model is further refined[106,107]. We

emphasize that 3D EM imaging, although potent for the reconstruction of nucleation and attachment events in inorganic systems and imaging of biological sample structure, has not yet been used for *in situ* imaging of assembly processes in biological samples.

It must be noted that cryo-EM is a time-lapse technique and does not offer the temporal resolution achievable with LP-TEM or AFM, nor does it enable one to follow the structural evolution of a single object. Thus, assembly pathways must be inferred from the evolution of the overall distribution of structures and dynamic information is largely unavailable. Nevertheless, there are some noteworthy studies that used cryo-EM to observe biomolecular self-assembly. For example, Van Driessche et al. used this method to monitor the nucleation pathway of glucose isomerase crystals, leading to two crystalline states and one gelled state [108]. They did not observe a metastable dense liquid as the precursor to the crystalline state, rather they proposed that the nucleation was driven by oriented attachments between subcritical clusters already exhibiting a degree of crystallinity. In another study by the same group[109], cryo-EM was used to observe oriented attachment (OA) in a protein system where facetted nanocrystals merge their lattices via OA alignment well before they make contact and then the assemblies grow into mesocrystals. This study shows the potential of cryo-EM to capture previously unobserved complexities of OA in protein crystals and predict an assembly pathway.

The studies discussed in this section show that, despite limitations, TEM has a high potential to study the complex process of assembly and nucleation in three-dimension with nanoscopic resolution. In addition, a ~400fps rate of LP-TEM imaging has been reported for *in situ* inorganic crystals nucleation[110,111] that will likely be applied to biomaterials in the near future.

3. Self-Assembly Dynamics Resolved by *In Situ* Characterization

3.1 Self-Assembly of Biomolecules

Inspired by nature, scientists have long been developing and fabricating artificial nanostructures that can mimic the hierarchical functional materials found in biological systems. To achieve this, researchers have investigated a wide range of material systems, including DNA, peptides, proteins, peptoids, and lipids. This section will examine scientific achievements based on *in situ* research in the field of biomolecular self-assembly and discuss how the fundamental knowledge gained through these achievements defines steppingstones for ongoing research.

3.1.1 Self-Assembly of DNA

The self-assembly of DNA is a promising approach for building novel bio-hybrid materials[4,5,112,113]. In 1982, Seeman first envisioned combining branched DNA molecules with complementary sticky ends to construct 2D nano-arrays[113]. This proposal is regarded as the seminal step in the development of DNA nanotechnology. A large variety of DNA tiles with different geometries and topologies have been synthesized in recent decades[4,114,115].

DNA nanotechnology has now undergone a radical change from answering questions regarding why individual components self-organize into ordered structures to designing programmable and precisely controlled stable nanoscale architectures. For example, Liu et al.[116] used a combination of AFM and *in situ* X-ray scattering to study the integration of gold nanoparticles (AuNP) with DNA origami frames to lead to DNA encoded interactions and

assembly into diverse planar architectures. They created building blocks of DNA combined with gold nanoparticles, which can be viewed as patchy particles with selective and fully prescribed anisotropic interactions. These patchy particles assembled into diverse morphologies (cross-shaped, zigzag, and 1D arrays), which were directly visualized with AFM. Interparticle spacing, chain length, and the long-range order of the 1D assemblies and 2D lattices were determined via *in situ* SAXS. The same group also used DNA origami technology to design tetrahedral cages to create a superlattice of nano-objects and studied the assembled structure via cryo-TEM and *in situ* SAXS[117]. They showed that particles coated with corresponding complementary DNA can either individually interact with the tetrahedral cages to form tetravalent caged particles and FCC superlattices or hybridize with the tetrahedral cages to create a diamond lattice. The structure of the DNA origami tetrahedron was resolved by cryo-TEM and *in situ* SAXS. The 2D scattering pattern and associated structure factor revealed a series of sharp scattering peaks and the ratio of the position of the peaks indicated the formation of a well-defined lattice. Another similar work on DNA crystals was performed by Zhang et al., wherein 3D rhombohedral crystalline lattices were created and characterized with SAXS and electron microscopy[118]. Approximately 90% of the volume of these crystals consisted of empty space, large enough to host 20 nm spherical gold nanoparticles. This size is large enough for some biomolecules, such as ribosomes. An additional study used SAXS to investigate the changes in DNA origami at different temperatures and salt concentrations to obtain information on its annealing and melting behavior. It was discovered that DNA can assemble in sheet, brick, or cylinder structures depending on temperature and ionic strength, which can be helpful in the retention and release of drugs[119].

SAXS has also demonstrated great potential for capturing time-resolved conformational changes in DNA origami structures. For example, Bruetzel et al.[120] studied a system composed of DNA monomer blocks that assemble into dimers and can undergo reversible conformational switches, depending on the concentration of $MgCl_2$. They used SAXS from a high-flux synchrotron beamline to differentiate between structures that undergo changes at significantly different rates. This included the formation of dimers from monomers, which took place on the timescale of minutes, as well as switching between the open and closed DNA dimer conformations (Fig. 9a-b), which occurred on the timescale of milliseconds. Information was obtained on the percentage of each structural conformation (e.g., monomer vs. dimer or opened vs. closed) at different times and $MgCl_2$ concentrations. This was then used to determine a rate constant for the formation of dimers. These results show that SAXS has the necessary spatial (~100 nanometers) and temporal (milliseconds) resolution to observe the dynamics of self-assembly in a DNA origami system and quantify its assembly kinetics.

A study by Song et al.[121] captured the temperature-dependent dynamic characteristics of DNA origami via thermally controlled *in situ* AFM (Fig. 9c). Topographical changes in the DNA structure were observed with an increase in temperature from 50ºC to 75ºC, resulting in the eventual dissociation of the DNA. Interestingly, the damage was reversed during a cooling cycle and the reversible DNA conformational changes were monitored *in situ*. Recently, *in situ* AFM imaging capabilities have advanced substantially with millisecond temporal resolution proving extremely successful for time-resolved visualization of the dynamic structural changes and interactions in individual DNA building blocks. Suzuki et al.[122] used the cavities of the 2D DNA origami framework to incorporate square-shaped DNA origami structures (SQ-origamis). The

real-time adsorption/desorption behavior of these SQ-origamis were imaged at a rate of 0.2 frames/sec and the dynamic docking and undocking of the SQ-origamis were studied under a range of experimental conditions, including varying the Mg^{2+} concentration, SQ-origami shape, and interaction between SQ-origamis and the pre-assembled DNA framework. High-speed AFM enabled the capture of a sudden transition of the SQ-origamis between adjacent cavities, which occurred on a timescale of the order of a few tens of milliseconds. The results provided insights into the assembly of higher-order DNA origami architectures and their response to external stimuli. In addition, some studies show bottom-up synthesis of higher-order DNA structures from DNA origami seeds. Mohammed et al. [123] used time-lapse fluorescence microscopy to capture the dynamic process of DNA nanotube nucleation from seeds, utilizing point-to-point assembly. The growth occurred by connecting 2 seeds attached to a glass surface at random locations. Nanotubes grew from both seeds by monomer addition and the connection formed once the growing ends of the two nanotubes were close enough to each other (Fig. 9d).

The above studies are just some of several that demonstrate the high programmability of DNA as a building block and reveal how oligonucleotide-nanoparticle conjugates self-assemble into complex nanostructured architectures and hybrid materials. More importantly, the complexity of the DNA self-assembly process is now largely understood because of imaging techniques with high spatial and temporal resolution. These unique properties of DNA and insights into the molecular structure and assembly of DNA are expected to enable a wide range of applications, such as molecular electronics, biosensing, and DNA-based computing.

3.1.2 Self-Assembly of Peptides

Peptides are an essential building block for biomimetic material synthesis. Peptides can self-assemble into various architectures, ranging from nanotubes to fiber bundles on the macro-scale, with different secondary structures, including β-sheet and α-helix[124,125]. Compared to proteins, the short length of peptides reduces the complexity in their design and synthesis, simplifying the study of the relationship between sequences, secondary structures, and their functionalities. Important applications of self-assembled peptides include targeted drug delivery[126], supramolecular hydrogels for biomedical applications[127], and control over the structure of inorganic nanomaterials[128].

An in-depth understanding of peptide assembly dynamics and mechanisms is of utmost importance in the pharmaceutical and biomedical field, especially for amyloidogenic proteins under pathological conditions (e.g., Alzheimer's and Parkinson's diseases). The self-assembly of different peptides into amyloid-like fibrils is a consequence of backbone hydrogen bonding for the single β-sheet layer, side-chain interactions (e.g., hydrophobic interactions and π-π stacking), and result in "cross-β" structures[129].

Amyloid-like fibrils of different peptides have a typical cross-β structure and the self-assembly process is a consequence of backbone hydrogen bonding for the single β-sheet layer formation and side-chain interaction (e.g., hydrophobic interactions and π-π stacking)[129]. Understanding amyloid nucleation and the pathways of structural changes from oligomers to fibrils will lead to a better understanding of diseases involving amyloids. An interesting study by Huang et al.[130] monitored the amyloid aggregation process via *in situ* AFM and observed the nucleation of soluble amyloid monomers to oligomers to protofibrils and fibrils. Time-resolved observation

of the nucleation process led to the clear demarcation of protofibril and fibril stages during the aggregation of amylin. The fibrillation dynamics between the two stages under different experimental conditions were monitored *in situ* (Fig. 10a). The growth dynamics of protofibrils and fibrils follow different pathways—protofibril growth is concentration dependent. In contrast, the fibril growth depends on molecular rearrangements and β-sheet optimization. Beyond insights into growth dynamics, *in situ* techniques are essential to understand the mechanism that inhibits the formation of amyloid fibrils. Ghadami et al. sought to identify how the endogenous protein transthyretin (TTR) affects the fibrillization of amyloid-β peptides (Aβ)[131]. The fluorescence of thioflavin T (ThT) increases in proportion to the concentration of amyloid fibrils because of ThT binding. Time-resolved fluorescence spectra of Aβ fibrillization were fit using kinetic models at different concentrations of TTR and different initial states of Aβ (i.e., preformed Aβ fibrils or monomers). One key takeaway from the analysis was that a decrease in the rate of primary nucleation, rather than elongation or secondary nucleation, could be attributed to the presence of TTR. However, not all amyloid fibrils appear to follow a multistep mechanism, especially those exhibiting structural reversibility, as reported by Gobeaux et al.[46] through their use of *in situ* fluorescence and CD to characterize the fibril assembly of atosiban nonapeptide (Fig. 10b). They observed that the peptide monomers quickly aggregated into short fibrils with imperfect β-sheets before the fibrils mature into longer fibrils richer in β-sheets. In addition, the secondary structure evolution of the peptide was observed *in situ* via synchrotron radiation CD, which gives a unique signal for each state of assembly due to its sensitivity to conformational changes. The linear dependence of the ellipticity at wavelengths corresponding to random coils to the ellipticity at wavelengths corresponding to β-sheets is consistent with the assembly of monomers to dimers and then fibrils without any intermediate oligomeric species with different conformations. On the other

hand, a break in the linearity would have indicated the existence of an intermediate species with a specific structure.

Peptides can also assemble into 2D arrays, as shown by Chen et al. [132]. This study not only captured the assembly dynamics via HS-AFM but also provided an in-depth insight into the nucleation mechanism, which occurred without a size barrier (the critical nuclei size was zero) and one row at a time. Peptides assembled into elongated islands on molybdenum disulfide (MoS_2) row by row without a transient precursor phase or large cluster attachment. The nucleation rate as a function of peptide concentration revealed that nucleation began as soon as peptide concentration reached equilibrium concentration and thus, the critical island size was zero. Although the growth dynamics of this system appear in violation of classical nucleation theory, they verify its long-standing prediction of 1D structure nucleation without a free energy barrier.

Peptide self-assembly may also be controlled by the conjugation of an alkyl tail to one end of peptide molecule to form a peptide amphiphile (PA). Qiu et al. studied the co-assembly of short peptides with a PA and evaluated the use of the latter as a therapeutic delivery method[133]. Wide-angle X-ray scattering (WAXS) revealed that the integrated area of the diffraction peaks associated with β-sheets increased and then decreased as the proportion of one of the peptides was increased relative to the PA. These results were coupled with molecular dynamics, FTIR, and fluorescence spectroscopy to show that guest peptides displace water molecules between the β-sheets formed by PA. Thermal annealing of the co-assembled structures, which leads to more thermodynamically stable states, was followed using *in situ* WAXS. The changes in peak intensity during annealing were consistent with a decrease in the amount of guest peptide co-assembled with PA and were corroborated by FTIR and fluorescence studies. The metastability of the co-assembled system was

exploited to control the release of a therapeutic peptide that interferes with the formation of amyloid fibrils, as demonstrated by time-resolved ThT fluorescence.

These studies are examples showing the functionality of peptides as building blocks for biomimetic material synthesis. Moreover, these studies demonstrate the significance of *in situ* techniques to observe the time-resolved assembly, mineralization, and folding of peptides at high resolution. This provides an understanding of assembly kinetics and mechanisms to modify or replicate peptide-based nanostructures.

3.1.3 Self-Assembly of Proteins

The self-assembly of proteins has long been the focus of scientific research due to its impact on medical science and technology. However, the sophisticated and complex hierarchical structures of proteins make it challenging to study and understand their self-assembly behavior. In nature, polypeptide chains coil and fold into proteins possessing diverse secondary structures. Further interaction and folding of these chains lead to tertiary and quaternary structures with tremendous morphological diversity. Moreover, the vast manifold of amino acid sequences and folding architectures allows the construction of an equally vast array of hierarchical assemblies through the precise arrangement of these protein-building units. Recently, numerous studies have been performed to better understand protein design and assembly[134,135]. The limited in-depth understanding of the pathway to the assembled structures is still the bottleneck to successfully replicating the hierarchical nature of biological structures. The advancement of time-resolved

techniques to visualize the different phases of protein assembly can help fill this knowledge gap for the development of biomimetic materials.

Collagen, one of the most prevalent extracellular matrix and structural proteins, is an extensively researched protein system. Collagens can assemble into hierarchical structures that span molecular to macroscopic scales[136]. A dynamic study of collagen type I fibrillogenesis was done by Stamov et al.[137] using buffers with different pH values and potassium concentrations. The process was monitored via AFM with high temporal resolution and the D-banding of the collagen fibrils became apparent within 15-17 min of fibrillogenesis on mica substrates (Fig. 11a). Initially, the fibrillar intermediates had a random orientation on the substrate, but they gradually fused into larger collagen fibrils, eventually forming a dispersed matrix of aligned ribbon-like collagen fibrils. *In situ* observation of the growing tips of the fibrils showed that the assembly process was mediated by the longitudinal stepwise addition of single building blocks protruding from the main fibrillar body.

Mineralization of protein self-assembled scaffolds, in the formation of bone, is a complex process where collagen fibrils serve as building blocks. This process involves nucleation of minerals in confined (intrafibrillar mineralization, IM) as well in unconfined (extrafibrillar mineralization, EM) spaces in the collagenous matrix. To understand the nuances of this complex process, Kim et al.[138] used *in situ* SAXS/WAXS to examine calcium phosphate (CaP) nucleation rates during EM and IM in simulated body fluids (SBF) with and without polyaspartic acid (pAsp), an extrafibrillar nucleation inhibitor (Fig. 11b-c). The SAXS patterns during EM and IM revealed aggregates of thin apatite crystals and individual plates separately arranged within the collagen, respectively. WAXS indicated slower development of nuclei crystallinity during EM than IM,

even though the total particle volume, as quantified from the absolute SAXS intensity, was slightly higher for EM. This suggests that the pathway for EM involves an amorphous intermediate that must undergo a phase transformation, thus delaying the appearance of hydroxyapatite crystals. The nucleation rate was obtained from plots of the slope of the scattering invariant vs. time and used in the classical nucleation model to obtain the interfacial energies between CaP nuclei and SBF in EM and IM. Unlike in previous studies, the interfacial energy and nucleation barrier was found to vary significantly depending on the nucleation site. This finding is significant in identifying important mechanisms governing biomineralization and replicating them for bone regeneration.

Another important protein assembly process is that of cytoskeletal proteins, which are essential for cellular processes. The dynamic assembly for one such protein that assembles into versatile morphologies, tubulin, was studied by Dharan et al. using time-resolved X-ray scattering and time-lapse cryo-TEM [35]. Tetravalent spermine in varying concentrations was used to effectively promote tubulin assembly into helical structures, double helical structures, and bundles. Multiple hierarchical structures were visualized as a function of spermine concentration and incubation time (Fig. 11d). Based on the observed information, it was determined that as the spermine concentration increased, hierarchical structures formed from less ordered and transient structures. Since there is a lack of theory to explain the formation of conical-spiral-based assemblies, a study such as this with time-lapsed tracking of the tubulin assembly pathway can be used to model the dynamics of the process. The versatility of tubulin can be used to create a wide range of building blocks for bioinspired material applications.

The studies discussed in this section show the protein functionality and versatility in biomineralization and novel biological nanomaterials synthesis. Protein assemblies may have

various potential applications in materials science, biomedical engineering, and bio-nanotechnology[134]. However, to realize these applications, the assembly dynamics must be well understood, making the advancement of *in situ* techniques extremely beneficial.

3.1.4 Self-Assembly of Peptoids

Peptoids (or poly-N-substituted glycines) are biocompatible materials with similar backbones to peptides; however, they exhibit much higher chemical and thermal stabilities than peptides and proteins[139]. Unlike natural peptides, peptoid side chains are appended to the backbone nitrogen atoms rather than the α-carbons, eliminating backbone hydrogen bond donors. Hence, peptoid-peptoid and peptoid-surface interactions can be simply manipulated through the variation of side-chain chemistry [140,141]. Moreover, the availability of a wide variety of functional groups for use as side chains allows peptoid designs that can self-assemble into a diverse array of hierarchical structures[142].

A study by Kang et al.[143] highlights the importance of time-resolved monitoring of peptoids to understand the mechanism of supramolecular self-assembly of diblock copolypeptoids in solution. They designed and synthesized two types of diblock copolypeptoids, composed of a core-forming block with either linear or branched n-alkyl side chains, namely, PNMG-b-PNOG (poly(N-methylglycine)-b-poly(N-octylglycine)) and PNMG-b-PNEHG (poly(N-methylglycine)-b-poly(N-2-ethyl-1-hexylglycine)). The copolypeptoids were dissolved by heating in methanol and then returned to room temperature to induce self-assembly. The formation of structures after being returned to room temperature was tracked at different lag times using time-resolved SAXS/WAXS and imaged after assembly was complete using cryo-TEM and *ex situ* AFM. They observed that PNMG-b-PNOG with linear n-octyl side chains slowly self-assembled into

hierarchical flowerlike structures composed of radially distributed nanoribbons seen in the AFM data (Fig. 12a). The intermediate and low-q power law exponents of the time-resolved SAXS data indicated the presence of structures with 2D symmetry consistent with ribbons and a larger self-assembled structure consistent with the imaged assemblies, respectively (Fig. 12b). By contrast, PNMG-b-PNEHG bearing branched racemic 2-ethyl-1-hexyl side chains self-assemble into well-defined hexagonal nanosheets. The authors tracked crystallinity formation in PNMG-b-PNOG using time-resolved WAXS (Fig. 12c) and contrasted this with the relative lack of long-range order in PNMG-b-PNEHG. Based on these time-dependent observations the authors provided a growth mechanism: PNMG-b-PNOG molecules associate to form amorphous spherical micelles, followed by the onset of crystallization resulting in the formation of crystal nuclei. The growth of the flower petals (i.e., nanoribbons subunits) occurs by the addition of the PNMG-b-PNOG molecules from the amorphous micelles to the nuclei following 2D crystallization kinetics. The self-assembly process of PNMG-b-PNEHG hexagonal nanosheets is relatively faster, attributed to the less defined molecular packing of PNEHG segments within the micellar core relative to the PNOG segment.

Apart from the synthesis of hierarchical structures, peptoids have demonstrated great potential for controlling the growth and nucleation of inorganic particles as demonstrated in a study by Jin et al. [93] (discussed in detail in section 2.4). Feng et al. [144] showed peptoid designs that lead to the controlled synthesis of highly branched plasmonic gold nanoparticles. They used *in situ* LP-TEM to investigate the influence of peptoid sequence hydrophobicity and side-chain chemistry on the evolution of nanoparticles from spherical to coral shaped. These studies and others with different peptoid designs[9] show that control over peptoid–particle interactions provides diverse possibilities for developing new functional nanomaterials.

Along with diverse functionality, peptoids also exhibit robustness and an ability to self-repair[145]. Jiao et al. used *in situ* AFM to first deliberately damage assembled peptoid sheets with an AFM probe and then capture the repair process upon introduction of peptoid-containing solution (Fig. 12d). They used amphiphilic peptoids to form bilayer membranes in the bulk solution that were then deposited on mica or highly ordered pyrolytic graphite (HOPG). An AFM probe scribed the peptoid sheets by applying a vertical load to remove the peptoids before freshly made peptoid solution at a suitable pH was introduced into the AFM fluid cell. Repair was performed both with the same peptoid used to grow the membrane as well as those with different hydrophilic heads. The effect of pH, surface charge, and crystallographic direction of the damage on the membrane repair rate was determined by *in situ* AFM. The study led to the conclusion that the kinetics of peptoid attachment and detachment at the growing edge of the membrane are impacted by the peptoid–peptoid interaction, the peptoid–membrane interaction, and the pH-controlled protonation state of the carboxyl groups in the hydrophilic head of the peptoid, which influences the inter-peptoid electrostatic repulsion.

In situ studies of similar amphiphilic peptoids to which other sidechains or headgroups, as well as distinct non-amphiphilic sequences were added, have been performed using AFM, DLS and SAXS[9,142,145–147]. These studies have documented the formation of helices, tubes, cylindrical micelles, and nanoribbons. They have determined both the formation pathways and kinetics for the different structures. Peptoids have also been investigated to control calcium carbonate mineralization[148]. These studies show that peptoids can mimic lipid amphiphilicity for self-assembly into highly stable 2D nanosheets and provide a robust matrix for the development of biomimetic membranes tailored to specific applications or be selected to form other

supramolecular motifs and exhibit protein-like functionality, such as self-repair or control over inorganic morphogenesis.

3.2 Self-Assembly Pathways

The self-assembly of biomolecules in living organisms can be complex because they can operate out of equilibrium in precisely regulated environments. These systems are thermodynamically and kinetically controlled, giving rise to special characteristics such as dynamic, adaptive, and self-regulating features. Identifying the thermodynamic and kinetic pathways in biomolecular assembly is critical for developing approaches to fabricate pre-designed hierarchical features[149]. However, biomolecular self-assembly frequently goes through nonclassical multi-step pathways, involving transient metastable states with oligomers or other species. Exceptions to this multi-step nucleation process where assembly follows the classical monomer-by-monomer pathway have been reported[9,82,132].

Two-step (or multi-step) nucleation is the general biomolecular crystallization and self-assembly pathway. In such cases, an intermediate state, either clusters, amorphous aggregates, or a dense liquid phase, forms between the initial and final ordered states and may serve as the nucleation precursor[150,151] (Fig. 13a). While this pathway was proposed to describe protein crystallization based on optical microscopy studies of lysozymes[151], later studies provided a molecular-scale view of such pathways for a broader range of macromolecules. In this context, a time-dependent cryo-TEM study by Houben et al.[152] about ferritin nucleation is noteworthy. It provided evidence for nonclassical crystallization by capturing the key stages in the structural-order evolution (Fig. 13b). The wide range of intermediate states with different degrees of order

implies that crystallinity evolves via continuous structural optimization and eventually converges into the final crystalline phase.

In an AFM study by Chung et al.[153], the S-layer protein, SbpA from Lysinibacillus sphaericus, was found to assemble on supported lipid bilayers via a multi-step process. The monomers first form a mobile adsorbed phase on the surface and then condense into amorphous or liquid-like clusters. These clusters further undergo a phase transition into crystalline arrays composed of compact tetramers through simultaneous folding and rearrangement of the tetramers in the clusters. Subsequent growth occurs by the formation of new tetramers at the edges of the arrays (Fig. 13c), where the existing matrix of tetrameric units acts as a template to guide the folding and assembly of the monomers into new tetramers. In a separate study, they showed that by changing the solid-liquid interface, i.e., using mica instead of a lipid bilayer, SbpA assembly exhibited a kinetic trap, leading to two distinct phases with conformational differences. First is a long-lived transient phase followed by a final stable phase[83] that differed only in the degree to which the individual proteins folded into the final lowest-energy conformation. Furthermore, assembly pathways and outcomes are not only modulated by changes in the solid-liquid interface but also by changes in the biomolecule design sequence, as demonstrated by Ma et al.[9]. They used *in situ* AFM to study the change of the direct crystallization pathway into a two-step nucleation pathway by the addition of a short hydrophobic region to a peptoid sequence.

Besides the multi-step assembly, biomolecular self-assembly can also occur via oriented attachment[109,132,154,155]. In several studies of protein and peptide nucleations, the nucleation rate is many orders of magnitude lower than the one predicted by the classical nucleation theory. Hence, for such systems, hierarchical pathways dominate and are similar to those widely observed

in inorganic systems—growth occurs first through the nucleation of crystalline nanoparticles, called primary particles, that then become the building blocks for the next level of assembly via a process of oriented attachment (OA). In OA, the primary particles rotate into crystallographic alignment before attaching, thus leading to the growth of larger single crystals. A study by Van Driessche et al.[109] explored crystallization of glucose isomerase via time-resolved cryo-EM and found that the material properties of the final crystalline structure are the result of the interactions between nuclei mediated by OA (Fig. 13d). The authors proposed a model based on time-lapsed observations of the growth process and showed how freely diffusing nanocrystals approach each other and undergo rotational and translational adjustments to align both lattices. The nanocrystals jump to contact and any violations of symmetry lead to grain boundary formation. Similarly, several other studies of organic and inorganic assembly via OA utilize cryo-EM[156] because the high resolution offered by this technique can resolve the grain boundaries in a mesocrystal and provide evidence of OA without the complication of drying effects or electron beam damage associated with *ex situ* and *in situ* TEM, respectively.

These studies reveal that unassembled biomolecules can transform from initial, high-entropy conformations to final ordered states by passing through metastable states with local energetic minimums. Conformational changes, which enable the emergence of ordered states during the nucleation process, need to be carefully considered and incorporated into future models of biomolecular material design.

4. Emerging Opportunities

4.1. Simulated Models to Predict Biomolecular Self-Assembly

In situ techniques can capture the emergence of order and the evolution of hierarchy. To drive hierarchical growth toward desired outcomes, dynamic control of reaction pathways via predictive physics-based models and real-time control of synthesis parameters is required. Experimental quantities measured by microscopy, spectroscopy, and scattering can be directly or indirectly linked to simulation models, via quantum mechanical or semi-empirical techniques. Combining simulations with data science tools elucidates the kinetics of building block growth and their assembly into ordered hierarchies, thus controlling the ultimate outcomes of synthesis. The most widely used simulation technique to study this process is molecular dynamics (MD), in which Newton's equations of motion are solved for a system of interacting particles representing atoms, coarse-grained groups of atoms, and their bonded and non-bonded interactions.

An all-atom model (AA) can be used to observe the formation and equilibration of hierarchical structures where all atoms are considered interaction centers. For example, in the empirical energy function in CHARMM22[157], the force field for peptides, nucleic acids, and lipids was modified for peptoids into MFTOID force field[158]. In a recent study, MFTOID was modified to study the assembly of peptoids into nanosheets during solvent evaporation[159]. The authors used different experimental techniques to validate the MD simulations of self-assembling peptoids at different solvent compositions. An extended intermediate structure was identified through SAXS experiments and confirmed to be a plausible intermediate in the assembly pathway to peptoid nanosheets through MD simulations. AA simulation has also been used to characterize

the structure and growth of drug amphiphiles via π–π stacking[160]. The ion-mediated assembly and folding of peptoids has been studied using a combination of ab initio molecular dynamics (AIMD) and atomistic resolution classical force field (FF) to span the relevant time and length scales[161].

Some biological processes, such as protein folding, aggregation, and biological assembly, occur at time scales beyond the scope of all-atom simulations. Therefore, processes involving large-size biological systems and longer timescale dynamics are approached by coarse-grained simulations that unite groups of atoms into effective interaction centers and provide a more comprehensive understanding of the length- (micrometers) and timescales (microseconds) for self-assembly, at the expense of atomistic accuracy. A study by Zhao et al. developed a bottom-up coarse-grained peptoid force field compatible with the popular suite of MARTINI[162] force fields for biomolecules. They constructed the model for the assembly of two representative peptoid sequences and demonstrated that the coarse-grained model was up to 25-fold more computationally efficient than the all-atom model. More importantly, the model is better suited for nanomaterial-relevant time and length scales that cannot be well simulated by an atomistic model (Fig. 14a). Similarly, coarse-grained models for DNA tile-based architectures have proven accurate in predicting the equilibrium structures[163]. They can simulate systems that contain thousands of nucleotides at nanoscopic level and generate ensembles of configurations at extended time periods representative of the equilibrium structure.

Simulation and artificial intelligence are also proving immensely successful in designing proteins with prespecified functions [164,165]. De novo design of proteins enables the synthesis of protein architectures from the ground up in a timely manner and has been highly impactful in bioengineering and biomedicine[166]. Recently, Dauparas et al.[167] developed a deep learning–

based protein sequence design method applicable to the design of monomers, cyclic oligomers, protein nanoparticles, and protein–protein interfaces (Fig. 14b). Moreover, it overcomes the protein design problems of Rosetta or AlphaFold in a fraction of the time. For example, in Rosetta, hydrophobic amino acids are often restricted on the protein surface because they can stabilize undesired multimeric states. Additionally, there may be considerable ambiguity about the extent to which restrictions should be applied to hydrophobic amino acids at the intermediate region between the core and surface of a protein. Despite the black-box nature of deep learning approaches relative to the physical transparency of Rosetta, the former avoids ambiguities in the application of restrictions by leveraging the Protein Data Bank to predict the most likely amino acid from a set of protein backbone coordinates.

In the last two decades, there have been many ingenious approaches to simulations to develop predictive models to gain nanostructure design control. However, there have been no literature reports of complete synthetic assembly pathways designed by a computer and then successfully executed in the laboratory. Nevertheless, significant progress has been made in recent years and terabytes of data have been generated through simulations. Difficulties comprehending simulations due to human limitations will become the shortcomings of the past and automation will potentially streamline the entire field of biomolecule dynamics and functionality.

4.2 Multi-Parametric *In Situ* Techniques

This review discusses several examples of using *in situ* characterization methods to characterize biomolecular assembly. However, each technique has its own strengths and

weaknesses. For example, X-ray scattering and AFM characterize structure, but do not provide chemical information. Hence, it would be advantageous to employ *in situ* or *in vitro* measurements colocalized with quantitative mapping of physical, chemical, and biological information to maximize the amount of multi-parametric information obtained to develop a complete understanding of self-assembly processes[168]. AFM-based measurement of force-distance curves is a classic example that enables researchers to simultaneously collect topological and mechanical maps of biological samples. By using probes functionalized with chemical groups and ligands, this method can further help map the chemical groups across biological matrixes[169,170]. Such information is essential for delineating the sophisticated functions of hierarchical biomolecular structures.

In recent years, a combination of AFM and IR spectroscopy to simultaneously perform morphological and chemical analysis has been widely used in studies of biomacrmolecular assembly. AFM-IR is a promising technique in biomolecular self-assembly[171] to colocalize AFM and IR spectroscopy for multi-parametric measurements. In general, samples can either be irradiated from below through an IR transparent prism (Fig. 15a) or from above by using a gold-coated tip (Fig. 15c). AFM-IR uses an oscillating AFM probe to detect the thermal expansion of biomolecules upon adsorption of light. At each wavenumber, the modulated amplitude is proportional to the IR adsorption coefficient[172](Fig. 15b). The second AFM-IR detection mode measures the plasmonic enhancement of the IR field between a gold-coated AFM tip and biomolecules on a gold substrate (Fig. 15d)[173–175]. AFM-IR has been widely used to detect the secondary structures and intermolecular interactions of peptide and protein assemblies with sub-

10-nm resolution[172]. It is also a powerful technique for distinguishing different material species across hybrid bio-materials and -complexes[172,176,177].

However, AFM-IR usually measures dehydrated samples to avoid signal interference from water, making it hard to measure biomolecular self-assembly *in situ/in vitro*. To solve this challenge, it is possible to use the bottom illumination configuration while the sample and AFM-IR tip are immersed in solvent[174,175,178]. Recent work tried to encapsulate biomolecules and water with graphene to *in vitro* investigate S-layer protein, SbpA, assembly with AFM-IR[177]. They used AFM-IR to directly monitor the amide I and II absorption bands as a function of time from the outer side of the graphene liquid cell, revealing the changes in the chemical bonds of the assembling protein in response to its environment, including Ca^{2+} and H_2O vs. D_2O. In a parallel experiment, they used *in situ* AFM to record the assembly on the inner side of the graphene window immersed in the imaging solution. After correlating the IR chemical information and AFM dynamics, they concluded that the protein assembles nonlinearly with respect to time and that the assembly is affected by changes in ionic strength or solvent.

Given advancements in nanofabrication, we believe that the spatial resolution of AFM-IR will be further improved and will enable the simultaneous characterization of structural and chemical information in a broader range of biomolecular self-assembly studies. In addition, developing colocalized AFM with other chemical mapping approaches, such as AFM-secondary ion mass spectrometry[179], could advance scientific understanding of the chemical evolution and distribution of biomolecules within assembly matrixes corresponding to structural changes.

4.3 Automated Experiment Workflows with High Throughput Screening

A primary goal of biomolecular self-assembly is to fabricate hierarchically structured materials. This is challenging because the assembly pathways of many biomolecular systems are multi-dimensional with design variables with multivariate relationships. To manipulate the pathways of the biomolecular systems into desirable assemblies, changes in experimental conditions (e.g., temperature, reagent concentration, pH, electric fields) must be made at specific times and orders. While performing these changes is often simple, achieving a targeted final structure is difficult due to the huge number of interventions. Consequently, many approaches employ machine learning-boosted simulation algorithms because of their ability to model complex multivariate relationships[180], as discussed in section 4.1.

Given the high number of iterations needed to train the neural networks, the previously mentioned method may become too costly or laborious to be feasible. However, experiments can be streamlined and human labor and time can be reduced through automated experimental workflows. These workflows include (1) a robotic system that is able to physically perform experiments, such as robotic arms[181] or liquid handling robots[182]; (2) a method to characterize the experiments being performed, such as an *in situ* rapid screening method that would deliver critical information about the reaction or assembly pathway and the final reaction outcome; and (3) a machine learning algorithm to generate and test hypotheses based on the collected data. The kind of algorithm required depends on the objective, such as causal analysis to infer complex relationships in datasets[183] or optimization algorithms such as Bayesian optimization[184] or reinforcement learning[185] to optimize targeted properties (Fig. 16a).

Data generated in high throughput experiments can often be hard to interpret, many dimensional, or in a large enough quantity to be impractical for a human or an algorithm to analyze. Therefore, it is common to process data first, as using machine learning techniques on unprocessed data can lead to poor performance[186]. For example, UV-vis spectra and SAS curves often contain hundreds to thousands of data points. They traditionally rely on expert knowledge to extract features from the data (e.g., peak wavelength positions, Guinier slopes) to make the data more interpretable and usable for models or algorithms.

Recently, functional principal component analysis was used to simplify data by transforming high-dimensional datasets into a few functional principal components[187]. Many optimization algorithms (e.g., Bayesian optimization and reinforcement learning) require distance metrics to rank how close samples are to a target and, in functional datasets, new metrics based on Riemannian geometry have been shown to be effective[188]. For example, the amplitude-phase metric used by Vaddi et al.[188] was demonstrated on UV-vis spectroscopy curves, where it ranked sample curves on how close they were to a targeted curved based on shape, often the most important feature when determining structural features from functional data. Another method of reducing data dimensionality is using variational autoencoders (VAE) that perform this reduction by sampling from the latent space (i.e., a distribution). In Valleti et al.[189], a combination of a VAE with Bayesian optimization was used on functional data, the electric field as a function of time, to maximize the total curl of the polarization field of a system.

Apart from functional datasets, data can be obtained from images and extracting meaningful information from them can also be automated. For instance, a convolutional neural network, U-net, was used to track and quantify the large-volume data of the distributions and

dynamics of protein nanorods at solid-liquid interfaces captured by HS-AFM (Fig. 16b) [190]. Upon comparison with kinetic Monte Carlo simulations, it was found that the rotational dynamics of proteins could be either a nonclassical random walk or classical Brownian motion. This resolved dual-model rotational dynamics of biomolecules at solid-liquid interfaces defines a general procedure for manipulating the self-assembly of hybrid biomolecular–inorganic materials. Ensemble learning and iterative training is another method of extracting information from microscopy [191]. Like the previous example, this method could be trained to identify biomolecules in images to automate data analysis. Combining human intelligence with the speed and dexterity of robotics and automation will enhance our scientific advancement immensely[192].

5. Summary and Outlook

In this review, we discussed and summarized the utilization of X-ray/Neutron/light scattering, light spectroscopy, AFM, and TEM for the *in situ* characterization of biomolecular assembly across spatial and temporal scales (Table 1). We further summarized a wide array of examples demonstrating the potential of cutting-edge *in situ* approaches, emerging data-science methodologies, and automated high throughput experiments to improve the accuracy and efficiency of biomaterials design by capturing and predicting bioinspired material synthesis pathways. The studies demonstrate that the fascinating functionalities of biomolecular materials are rooted in their hierarchical nature, enabled by the high information content encoded in their building blocks. With the accumulated understanding of self-assembly pathways and mechanisms using *in situ* approaches, self-assembling oligonucleotides, peptides, peptoids, and other foldamers

have been employed to create quasi-1D, 2D, and 3D structural motifs, including fibers, ribbons, sheets, tubes, and vesicles, for diverse applications.

To date, however, the outstanding biomimetic and bioinspired materials produced with these macromolecules have often been the product of intuition and serendipity and significant knowledge gaps remain. Specifically, we lack sufficient knowledge regarding the mechanisms of how sequence and chemistry translate into the interactions and assembly dynamics from which order emerges. In addition, we are still unable to connect atomistic descriptions of intermolecular interactions with coarse-grained models of building blocks to bridge the time and length scales required for predicting assembly. We do not know how the interplay of molecular interactions, solvent, electrolytes, disparate blocks, and dynamic processing conditions controls the energy landscapes across which hierarchy develops. Finally, we do not know how to predict the metastable states that exist in the pathway to an ordered state or how to encode a balance of forces that will, by design, create multi-well potentials for out-of-equilibrium reconfiguration in response to external stimuli.

To bridge these knowledge gaps and predict and manipulate the pathways and outcomes of biomolecular assembly, we need to utilize all the advances we have in the areas of macromolecular design, theory, simulations, artificial intelligence, and *in situ* characterization. Specifically, we need to both adopt the classical ideas of shape complementarity and packing interactions, H-bonds, electrostatic bonds, covalent cross-linking, entropic effects, dipole–dipole interaction, etc. and, in addition, pay attention to the profound impact of solvent–electrolyte–building block interactions, modulated by molecular details, on the energy landscapes across which synthesis proceeds. Ultimately, we envision a day when we can utilize biomolecules as building blocks for the construction of sophisticated molecular machines, energy storage devices, drug delivery agents,

and other dynamic materials comparable to those employed by nature.

Acknowledgements

The authors acknowledge the support of the US Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences as part of the Center for the Science of Synthesis Across Scales, an Energy Frontier Research Center, under Award DE-SC0019288. PNNL is a multiprogram national laboratory operated for the DOE by Battelle under Contract DE-AC05-76RL01830.

References

[1] C. Gong, S. Sun, Y. Zhang, L. Sun, Z. Su, A. Wu, G. Wei, *Nanoscale* **2019**, *11*, 4147–4182.

[2] S. Zhang, J. Chen, J. Liu, H. Pyles, D. Baker, C. L. Chen, J. J. De Yoreo, *Adv. Mater.* **2021**, *33*, 11–13.

[3] Q. Luo, C. Hou, Y. Bai, R. Wang, J. Liu, *Chem. Rev.* **2016**, *116*, 13571–13632.

[4] P. W. K. Rothemund, *Nature* **2006**, *440*, 297–302.

[5] H. Yan, S. H. Park, G. Finkelstein, J. H. Reif, T. H. LaBean, *Science (80-.).* **2003**, *301*, 1882–1884.

[6] J. M. Fletcher, R. L. Harniman, F. R. H. Barnes, A. L. Boyle, A. Collins, J. Mantell, T. H. Sharp, M. Antognozzi, P. J. Booth, N. Linden, M. J. Miles, R. B. Sessions, P. Verkade, D.

N. Woolfson, *Science (80-.).* **2013**, *340*, 595–599.

[7] J. Zhu, N. Avakyan, A. Kakkis, A. M. Hoffnagle, K. Han, Y. Li, Z. Zhang, T. S. Choi, Y. Na, C. J. Yu, F. A. Tezcan, *Chem. Rev.* **2021**, *121*, 13701–13796.

[8] P. S. Huang, S. E. Boyken, D. Baker, *Nature* **2016**, *537*, 320–327.

[9] X. Ma, S. Zhang, F. Jiao, C. J. Newcomb, Y. Zhang, A. Prakash, Z. Liao, M. D. Baer, C. J. Mundy, J. Pfaendtner, A. Noy, C. L. Chen, J. J. De Yoreo, *Nat. Mater.* **2017**, *16*, 767–774.

[10] P. J. Yoo, K. T. Nam, J. Qi, S. K. Lee, J. Park, A. M. Belcher, P. T. Hammond, *Nat. Mater.* **2006**, *5*, 234–240.

[11] X. Zhang, C. Gong, O. U. Akakuru, Z. Su, A. Wu, G. Wei, *Chem. Soc. Rev.* **2019**, *48*, 5564–5595.

[12] X. Hu, M. Liao, H. Gong, L. Zhang, H. Cox, T. A. Waigh, J. R. Lu, *Curr. Opin. Colloid Interface Sci.* **2020**, *45*, 1–13.

[13] Y. Zhao, W. Yang, C. Chen, J. Wang, L. Zhang, H. Xu, *Curr. Opin. Colloid Interface Sci.* **2018**, *35*, 112–123.

[14] T. S. Choi, F. A. Tezcan, *Nature* **2022**, *603*, 522–527.

[15] H. Jin, Y.-H. Ding, M. Wang, Y. Song, Z. Liao, C. J. Newcomb, X. Wu, X.-Q. Tang, Z. Li, Y. Lin, F. Yan, T. Jian, P. Mu, C.-L. Chen, *Nat. Commun.* **2018**, *9*, 270.

[16] A. D. Merg, J. C. Boatz, A. Mandal, G. Zhao, S. Mokashi-Punekar, C. Liu, X. Wang, P. Zhang, P. C. A. Van Der Wel, N. L. Rosi, *J. Am. Chem. Soc.* **2016**, *138*, 13655–13663.

[17] J. Zhang, Y. Liu, Y. Ke, H. Yan, *Nano Lett.* **2006**, *6*, 248–251.

[18] R. H. Tunuguntla, R. Y. Henley, Y.-C. Yao, T. A. Pham, M. Wanunu, A. Noy, *Science (80-.).* **2017**, *357*, 792–796.

[19] C. Li, J. Adamcik, R. Mezzenga, *Nat. Nanotechnol.* **2012**, *7*, 421–427.

[20] Q. Zhang, M. Li, C. Zhu, G. Nurumbetov, Z. Li, P. Wilson, K. Kempe, D. M. Haddleton, *J. Am. Chem. Soc.* **2015**, *137*, 9344–9353.

[21] C. J. Gommes, S. Jaksch, H. Frielinghaus, *J. Appl. Crystallogr.* **2021**, *54*, 1832–1843.

[22] S. Da Vela, D. I. Svergun, *Curr. Res. Struct. Biol.* **2020**, *2*, 164–170.

[23] F. Auvray, D. Dennetiere, A. Giuliani, F. Jamme, F. Wien, B. Nay, S. Zirah, F. Polack, C. Menneglier, B. Lagarde, J. D. Hirst, M. Réfrégiers, *Struct. Dyn.* **2019**, *6*, 054307.

[24] P. L. Th. Zemb, Ed. , *Neutrons, X-Rays and Light. Scattering Methods Applied to Soft Condensed Matter*, **1988**.

[25] T. C. Hansen, H. Kohlmann, *Zeitschrift für Anorg. und Allg. Chemie* **2014**, *640*, 3044–3063.

[26] E. Mahieu, Z. Ibrahim, M. Moulin, M. Härtlein, B. Franzetti, A. Martel, F. Gabel, *EPJ Web Conf.* **2020**, *236*, 03002.

[27] M. H. L. Nguyen, M. DiPasquale, B. W. Rickeard, C. B. Stanley, E. G. Kelley, D. Marquardt, *Biophys. J.* **2019**, *116*, 755–759.

[28] B. J. Eves, J. J. Doutch, A. E. Terry, H. Yin, M. Moulin, M. Haertlein, V. T. Forsyth, P.

Flagmeier, T. P. J. Knowles, D. M. Dias, G. Lotze, A. M. Seddon, A. M. Squires, *RSC Chem. Biol.* **2021**, *2*, 1232–1238.

[29] Y. T. Lee, L. D. Pozzo, *Langmuir* **2019**, *35*, 15192–15203.

[30] E. W. Martin, T. S. Harmon, J. B. Hopkins, S. Chakravarthy, J. J. Incicco, P. Schuck, A. Soranno, T. Mittag, *Nat. Commun.* **2021**, *12*, 4513.

[31] N. Mertz, A. Yaghmur, J. Østergaard, H. Amenitsch, S. W. Larsen, *J. Colloid Interface Sci.* **2021**, *602*, 415–425.

[32] D. Svergun, C. Barberato, M. H. J. Koch, *J. Appl. Crystallogr.* **1995**, *28*, 768–773.

[33] D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, D. I. Svergun, *J. Appl. Crystallogr.* **2017**, *50*, 1212–1225.

[34] A. Ginsburg, T. Ben-Nun, R. Asor, A. Shemesh, I. Ringel, U. Raviv, *J. Chem. Inf. Model.* **2016**, *56*, 1518–1527.

[35] R. Dharan, A. Shemesh, A. Millgram, R. Zalk, G. A. Frank, Y. Levi-Kalisman, I. Ringel, U. Raviv, *ACS Nano* **2021**, *15*, 8836–8847.

[36] B. R. Pauw, *J. Phys. Condens. Matter* **2014**, *26*, 239501.

[37] A. P. Minton, *Anal. Biochem.* **2016**, *501*, 4–22.

[38] J. C. Berg, *An Introduction to Interfaces and Colloids*, World Scientific, **2009**.

[39] J. Stetefeld, S. A. McKenna, T. R. Patel, *Biophys. Rev.* **2016**, *8*, 409–427.

[40] N. Narang, T. Sato, *Polym. J.* **2021**, *53*, 1413–1424.

[41] E. Korkmazhan, P. Tompa, A. R. Dunn, *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 647–648.

[42] B. Hämisch, A. Büngeler, C. Kielar, A. Keller, O. Strube, K. Huber, *Langmuir* **2019**, *35*, 12113–12122.

[43] N. De Lange, F. A. M. Leermakers, J. M. Kleijn, *Soft Matter* **2020**, *16*, 2379–2389.

[44] H. A. Havel, R. S. Chao, R. J. Haskell, T. J. Thamann, *Anal. Chem.* **1989**, *61*, 642–650.

[45] A. Pragya, S. Mutalik, M. W. Younas, S. K. Pang, P. K. So, F. Wang, Z. Zheng, N. Noor, *RSC Adv.* **2021**, *11*, 10710–10726.

[46] F. Gobeaux, F. Wien, *Langmuir* **2018**, *34*, 7180–7191.

[47] Z. C. Adams, E. J. Olson, T. L. Lopez-Silva, Z. Lian, A. Y. Kim, M. Holcomb, J. Zimmermann, R. Adhikary, P. E. Dawson, *Chem. Sci.* **2022**, *13*, 10020–10028.

[48] B. Süss, F. Ringleb, J. Heberle, *Rev. Sci. Instrum.* **2016**, *87*, 063113.

[49] J. De Meutter, E. Goormaghtigh, *Eur. Biophys. J.* **2021**, *50*, 613–628.

[50] V. G. Ostapchenko, M. R. Sawaya, N. Makarava, R. Savtchenko, K. P. R. Nilsson, D. Eisenberg, I. V. Baskakov, *J. Mol. Biol.* **2010**, *400*, 908–921.

[51] J. S. Stevanic, L. Salmén, *Carbohydr. Polym.* **2020**, *230*, 115615.

[52] M. P. Schmidt, C. E. Martínez, *Langmuir* **2016**, *32*, 7719–7729.

[53] A. Bouhekka, T. Bürgi, *Appl. Surf. Sci.* **2012**, *261*, 369–374.

[54] H. Tiernan, B. Byrne, S. G. Kazarian, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*

2020, *241*, 118636.

[55] M. S. C. Barreto, E. J. Elzinga, L. R. F. Alleoni, *Sci. Rep.* **2020**, *10*, 1–13.

[56] D. Partouche, V. Militello, A. Gomez-Zavaglia, F. Wien, C. Sandt, V. Arluison, *Pathogens* **2019**, *8*, 36.

[57] J. Zhao, J. K. Cui, R. X. Chen, Z. Z. Tang, Z. L. Tan, L. Y. Jiang, F. Liu, *Biochem. Eng. J.* **2021**, *176*, 108225.

[58] D. Buhrke, P. Hildebrandt, *Chem. Rev.* **2020**, *120*, 3577–3630.

[59] G. Ramachandran, E. A. Milán-Garcés, J. B. Udgaonkar, M. Puranik, *Biochemistry* **2014**, *53*, 6550–6565.

[60] M. Banchelli, M. De Angelis, C. D'Andrea, R. Pini, P. Matteini, *Sci. Rep.* **2018**, *8*, 1–8.

[61] B. Wong, X. Zhao, Y. Su, D. Fu, **2023**, 35.

[62] M. Urbanová, P. Maloň, in *Anal. Methods Supramol. Chem.*, Wiley, **2012**, pp. 337–369.

[63] O. El Hamoui, K. Gaudin, S. Battu, P. Barthélémy, G. Lespes, B. Alies, *Langmuir* **2021**, *37*, 297–310.

[64] S. L. Thorpe, G. N. Snyder, A. Mammana, *Chirality* **2020**, *32*, 5–16.

[65] T. Zhou, J. J. Vallooran, R. Mezzenga, *Nanoscale* **2019**, *11*, 5891–5895.

[66] S. Takaramoto, Y. Nakasone, K. Sadakane, S. Maruta, M. Terazima, *RSC Adv.* **2020**, *11*, 1086–1097.

[67] F. Hache, P. Changenet, *Chirality* **2021**, *33*, 747–757.

[68] A. J. Miles, B. A. Wallace, *Chem. Soc. Rev.* **2006**, *35*, 39–51.

[69] B. A. Wallace, R. W. Janes, *Curr. Opin. Chem. Biol.* **2001**, *5*, 567–571.

[70] N. Taricska, D. Horváth, D. K. Menyhárd, H. Ákontz-Kiss, M. Noji, M. So, Y. Goto, T. Fujiwara, A. Perczel, *Chem. – A Eur. J.* **2020**, *26*, 1968–1978.

[71] C. Gila-Vilchez, M. C. Mañas-Torres, J. A. González-Vera, F. Franco-Montalban, J. A. Tamayo, F. Conejero-Lara, J. M. Cuerva, M. T. Lopez-Lopez, A. Orte, L. Álvarez De Cienfuegos, *Polym. Chem.* **2021**, *12*, 6832–6845.

[72] M. A. Digman, E. Gratton, *Annu. Rev. Phys. Chem.* **2011**, *62*, 645–668.

[73] M. C. Mañas-Torres, C. Gila-Vilchez, J. A. González-Vera, F. Conejero-Lara, V. Blanco, J. M. Cuerva, M. T. Lopez-Lopez, A. Orte, L. Álvarez De Cienfuegos, *Mater. Chem. Front.* **2021**, *5*, 5452–5462.

[74] S. Jazani, I. Sgouralis, O. M. Shafraz, M. Levitus, S. Sivasankar, S. Pressé, *Nat. Commun.* **2019**, *10*, 3662.

[75] J. Wägele, S. De Sio, B. Voigt, J. Balbach, M. Ott, *Biophys. J.* **2019**, *116*, 227–238.

[76] E. Coelho-Cerqueira, A. S. Pinheiro, C. Follmer, *Bioorganic Med. Chem. Lett.* **2014**, *24*, 3194–3198.

[77] G. Binnig, C. F. Quate, C. Gerber, *Phys. Rev. Lett.* **1986**, *56*, 930–933.

[78] H. G. Hansma, *J. Vac. Sci. Technol. B Microelectron. Nanom. Struct.* **1991**, *9*, 1282.

[79] Y. F. Dufrêne, T. Ando, R. Garcia, D. Alsteens, D. Martinez-Martin, A. Engel, C. Gerber,

D. J. Müller, *Nat. Nanotechnol.* **2017**, *12*, 295–307.

[80] D. J. Müller, A. C. Dumitru, C. Lo Giudice, H. E. Gaub, P. Hinterdorfer, G. Hummer, J. J. De Yoreo, Y. F. Dufrêne, D. Alsteens, *Chem. Rev.* **2021**, *121*, 11701–11725.

[81] M. Guthold, X. Zhu, C. Rivetti, G. Yang, N. H. Thomson, S. Kasas, H. G. Hansma, B. Smith, P. K. Hansma, C. Bustamante, *Biophys. J.* **1999**, *77*, 2284–2294.

[82] M. Sleutel, J. Lutsko, A. E. S. Van Driessche, M. A. Durán-Olivencia, D. Maes, *Nat. Commun.* **2014**, *5*, 5598.

[83] S. H. Shin, S. Chung, B. Sanii, L. R. Comolli, C. R. Bertozzi, J. J. De Yoreo, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 12968–12973.

[84] C. A. Schoenenberger, J. H. Hoh, *Biophys. J.* **1994**, *67*, 929–936.

[85] H. Takahashi, A. Miyagi, L. Redondo-Morata, S. Scheuring, *Small* **2016**, *12*, 6106–6113.

[86] Y. Ruan, K. Kao, S. Lefebvre, A. Marchesi, P. J. Corringer, R. K. Hite, S. Scheuring, *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 10333–10338.

[87] M. Shibata, H. Yamashita, T. Uchihashi, H. Kandori, T. Ando, *Nat. Nanotechnol.* **2010**, *5*, 208–212.

[88] N. Kodera, D. Yamamoto, R. Ishikawa, T. Ando, *Nature* **2010**, *468*, 72–76.

[89] G. R. Heath, S. Scheuring, *Nat. Commun.* **2018**, *9*, 4983.

[90] K. Kikuchi, T. Fukuyama, T. Uchihashi, T. Furuta, Y. T. Maeda, T. Ueno, *Small* **2022**, *5*, 1–9.

[91] Y. Suzuki, I. Kawamata, K. Watanabe, E. Mano, *iScience* **2022**, *25*, 104292.

[92] K. He, M. Sawczyk, C. Liu, Y. Yuan, B. Song, R. Deivanayagam, A. Nie, X. Hu, V. P. Dravid, J. Lu, C. Sukotjo, Y. P. Lu, P. Krá, T. Shokuhfar, R. Shahbazian-Yassar, *Sci. Adv.* **2020**, *6*, 1–12.

[93] B. Jin, F. Yan, X. Qi, B. Cai, J. Tao, X. Fu, S. Tan, P. Zhang, J. Pfaendtner, N. Y. Naser, F. Baneyx, X. Zhang, J. J. DeYoreo, C. Chen, *Angew. Chemie Int. Ed.* **2022**, *61*, DOI 10.1002/anie.202201980.

[94] D. Li, M. H. Nielsen, J. R. I. Lee, C. Frandsen, J. F. Banfield, J. J. De Yoreo, *Science (80-.).* **2012**, *336*, 1014–1018.

[95] T. Yamazaki, A. E. S. Van Driessche, Y. Kimura, *Soft Matter* **2020**, *16*, 1955–1960.

[96] H. Wang, B. Li, Y. J. Kim, O. H. Kwon, S. Granick, *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 1283–1292.

[97] M. A. Touve, A. S. Carlini, N. C. Gianneschi, *Nat. Commun.* **2019**, *10*, 1–12.

[98] R. Kröger, A. Verch, *Minerals* **2018**, *8*, 21.

[99] L. R. Parent, E. Bakalis, M. Proetto, Y. Li, C. Park, F. Zerbetto, N. C. Gianneschi, *Acc. Chem. Res.* **2018**, *51*, 3–11.

[100] N. M. Schneider, M. M. Norton, B. J. Mendel, J. M. Grogan, F. M. Ross, H. H. Bau, *J. Phys. Chem. C* **2014**, *118*, 22373–22382.

[101] H. Cho, M. R. Jones, S. C. Nguyen, M. R. Hauwiller, A. Zettl, A. P. Alivisatos, *Nano Lett.* **2017**, *17*, 414–420.

[102] J. Park, H. Park, P. Ercius, A. F. Pegoraro, C. Xu, J. W. Kim, S. H. Han, D. A. Weitz, *Nano Lett.* **2015**, *15*, 4737–4744.

[103] C. Lei, Y. H. Wang, P. X. Zhuang, Y. T. Li, Q. Q. Wan, Y. X. Ma, F. R. Tay, L. N. Niu, *J. Dent. Res.* **2022**, *101*, 505–514.

[104] S.-C. Chung, H.-H. Lin, P.-Y. Niu, S.-H. Huang, I.-P. Tu, W.-H. Chang, *Commun. Biol.* **2020**, *3*, 508.

[105] K. M. Roccapriore, S. V. Kalinin, M. Ziatdinov, *Adv. Sci.* **2022**, *9*, 1–10.

[106] P. A. Penczek, in *Methods Enzymol.* (Ed.: G. Jensen), **2010**, pp. 1–33.

[107] M. V. De Ruiter, R. Klem, D. Luque, J. J. L. M. Cornelissen, J. R. Castón, *Nanoscale* **2019**, *11*, 4130–4146.

[108] A. E. S. Van Driessche, N. Van Gerven, P. H. H. Bomans, R. R. M. Joosten, H. Friedrich, D. Gil-Carton, N. A. J. M. Sommerdijk, M. Sleutel, *Nature* **2018**, *556*, 89–94.

[109] A. E. S. Van Driessche, N. Van Gerven, R. R. M. Joosten, W. L. Ling, M. Bacia, N. Sommerdijk, M. Sleutel, *Nat. Commun.* **2021**, *12*, 6–13.

[110] H. G. Liao, D. Zherebetskyy, H. Xin, C. Czarnik, P. Ercius, H. Elmlund, M. Pan, L. W. Wang, H. Zheng, *Science (80-.).* **2014**, *345*, 916–919.

[111] S. Jeon, T. Heo, S.-Y. Hwang, J. Ciston, K. C. Bustillo, B. W. Reed, J. Ham, S. Kang, S. Kim, J. Lim, K. Lim, J. S. Kim, M.-H. Kang, R. S. Bloom, S. Hong, K. Kim, A. Zettl, W. Y. Kim, P. Ercius, J. Park, W. C. Lee, *Science (80-.).* **2021**, *371*, 498–503.

[112] W. M. Shih, J. D. Quispe, G. F. Joyce, *Nature* **2004**, *427*, 618–621.

[113] N. C. Seeman, *Nature* **2003**, *421*, 427–431.

[114] K. Lund, Y. Liu, S. Lindsay, H. Yan, *J. Am. Chem. Soc.* **2005**, *127*, 17606–17607.

[115] F. A. Aldaye, P. K. Lo, P. Karam, C. K. McLaughlin, G. Cosa, H. F. Sleiman, *Nat. Nanotechnol.* **2009**, *4*, 349–352.

[116] W. Liu, J. Halverson, Y. Tian, A. V. Tkachenko, O. Gang, *Nat. Chem.* **2016**, *8*, 867–873.

[117] W. Liu, M. Tagawa, H. L. Xin, T. Wang, H. Emamy, H. Li, K. G. Yager, F. W. Starr, A. V. Tkachenko, O. Gang, *Science (80-.).* **2016**, *351*, 582–586.

[118] T. Zhang, C. Hartl, K. Frank, A. Heuer-Jungemann, S. Fischer, P. C. Nickels, B. Nickel, T. Liedl, *Adv. Mater.* **2018**, *30*, 1800273.

[119] S. Fischer, C. Hartl, K. Frank, J. O. Rädler, T. Liedl, B. Nickel, *Nano Lett.* **2016**, *16*, 4282–4287.

[120] L. K. Bruetzel, P. U. Walker, T. Gerling, H. Dietz, J. Lipfert, *Nano Lett.* **2018**, *18*, 2672–2676.

[121] J. Song, J. M. Arbona, Z. Zhang, L. Liu, E. Xie, J. Elezgaray, J. P. Aime, K. V. Gothelf, F. Besenbacher, M. Dong, *J. Am. Chem. Soc.* **2012**, *134*, 9844–9847.

[122] Y. Suzuki, H. Sugiyama, M. Endo, *Angew. Chemie - Int. Ed.* **2018**, *57*, 7061–7065.

[123] A. M. Mohammed, P. Šulc, J. Zenk, R. Schulman, *Nat. Nanotechnol.* **2017**, *12*, 312–316.

[124] M. C. Branco, D. M. Sigano, J. P. Schneider, *Curr. Opin. Chem. Biol.* **2011**, *15*, 427–434.

[125] E. Tayeb-Fligelman, O. Tabachnikov, A. Moshe, O. Goldshmidt-Tran, M. R. Sawaya, N.

Coquelle, J. P. Colletier, M. Landau, *Science (80-.).* **2017**, *355*, 831–833.

[126] E. B. Peters, M. R. Karver, K. Sun, D. C. Gillis, S. Biswas, T. D. Clemons, W. He, N. D. Tsihlis, S. I. Stupp, M. R. Kibbe, **2021**, *2100103*, 1–15.

[127] P. Chakraborty, L. Adler-Abramovich, E. Gazit, in *Multifunct. Hydrogels Biomed. Appl.*, Wiley, **2022**, pp. 127–147.

[128] B. A. Krajina, A. C. Proctor, A. P. Schoen, A. J. Spakowitz, S. C. Heilshorn, *Prog. Mater. Sci.* **2018**, *91*, 1–23.

[129] R. Nelson, M. R. Sawaya, M. Balbirnie, A. Madsen, C. Riekel, R. Grothe, D. Eisenberg, *Nature* **2005**, *435*, 773–778.

[130] Q. Huang, H. Wang, H. Gao, P. Cheng, L. Zhu, C. Wang, Y. Yang, *J. Phys. Chem. Lett.* **2019**, *10*, 214–222.

[131] S. A. Ghadami, S. Chia, F. S. Ruggeri, G. Meisl, F. Bemporad, J. Habchi, R. Cascella, C. M. Dobson, M. Vendruscolo, T. P. J. Knowles, F. Chiti, *Biomacromolecules* **2020**, *21*, 1112–1125.

[132] J. Chen, E. Zhu, J. Liu, S. Zhang, Z. Lin, X. Duan, H. Heinz, Y. Huang, J. J. De Yoreo, *Science (80-.).* **2018**, *362*, 1135–1139.

[133] R. Qiu, I. R. Sasselli, Z. Álvarez, H. Sai, W. Ji, L. C. Palmer, S. I. Stupp, *J. Am. Chem. Soc.* **2022**, *144*, 5562–5574.

[134] P. Katyal, M. Meleties, J. K. Montclare, *ACS Biomater. Sci. Eng.* **2019**, *5*, 4132–4147.

[135] D. S. Goodsell, A. J. Olson, *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 105–153.

[136] M. D. Shoulders, R. T. Raines, *Annu. Rev. Biochem.* **2009**, *78*, 929–958.

[137] D. R. Stamov, E. Stock, C. M. Franz, T. Jähnke, H. Haschke, *Ultramicroscopy* **2015**, *149*, 86–94.

[138] D. Kim, B. Lee, S. Thomopoulos, Y.-S. Jun, *Nat. Commun.* **2018**, *9*, 962.

[139] J. Sun, R. N. Zuckermann, *ACS Nano* **2013**, *7*, 4715–4732.

[140] H. K. Murnen, A. M. Rosales, J. N. Jaworski, R. A. Segalman, R. N. Zuckermann, *J. Am. Chem. Soc.* **2010**, *132*, 16112–16119.

[141] H. Jin, F. Jiao, M. D. Daily, Y. Chen, F. Yan, Y. H. Ding, X. Zhang, E. J. Robertson, M. D. Baer, C. L. Chen, *Nat. Commun.* **2016**, *7*, 1–8.

[142] C. L. Chen, R. N. Zuckermann, J. J. Deyoreo, *ACS Nano* **2016**, *10*, 5314–5320.

[143] L. Kang, A. Chao, M. Zhang, T. Yu, J. Wang, Q. Wang, H. Yu, N. Jiang, D. Zhang, *J. Am. Chem. Soc.* **2021**, *143*, 5890–5902.

[144] F. Yan, L. Liu, T. R. Walsh, Y. Gong, Z. Zhu, J. J. De Yoreo, M. H. Engelhard, P. Z. El-khoury, Y. Zhang, X. Zhang, C. Chen, *Nat. Commun.* **2018**, 1–8.

[145] F. Jiao, Y. Chen, H. Jin, P. He, C. L. Chen, J. J. De Yoreo, *Adv. Funct. Mater.* **2016**, *26*, 8960–8967.

[146] P. Armand, K. Kirshenbaum, A. Falicov, R. L. Dunbrack, K. A. Dill, R. N. Zuckermann, F. E. Cohen, *Fold. Des.* **1997**, *2*, 369–375.

[147] J. A. Hammons, M. D. Baer, T. Jian, J. R. I. Lee, T. M. Weiss, J. J. De Yoreo, A. Noy, C.

L. Chen, A. Van Buuren, *J. Phys. Chem. Lett.* **2021**, *12*, 6126–6133.

[148] C.-L. Chen, J. Qi, R. N. Zuckermann, J. J. DeYoreo, *J. Am. Chem. Soc.* **2011**, *133*, 5214–5217.

[149] J. J. De Yoreo, *Rev. Mineral. Geochemistry* **2003**, *54*, 57–93.

[150] A. Sauter, F. Roosen-Runge, F. Zhang, G. Lotze, R. M. J. Jacobs, F. Schreiber, *J. Am. Chem. Soc.* **2015**, *137*, 1485–1491.

[151] P. G. Vekilov, *Nanoscale* **2010**, *2*, 2346–2357.

[152] L. Houben, H. Weissman, S. G. Wolf, B. Rybtchinski, *Nature* **2020**, *579*, 540–543.

[153] S. Chung, S. H. Shin, C. R. Bertozzi, J. J. De Yoreo, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16536–16541.

[154] Ç. Ç. Cenker, P. H. H. Bomans, H. Friedrich, B. Dedeoğlu, V. Aviyente, U. Olsson, N. A. J. M. Sommerdijk, S. Bucak, *Soft Matter* **2012**, *8*, 7463–7470.

[155] L. Yang, H. Yang, Z. Yang, Y. Cao, X. Ma, Z. Lu, Z. Zheng, *J. Phys. Chem. B* **2008**, *112*, 9795–9801.

[156] Z. Chen, K. Higashi, K. Ueda, K. Moribe, *Nano Lett.* **2022**, *22*, 6841–6846.

[157] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

[158] D. T. Mirijanian, R. V. Mannige, R. N. Zuckermann, S. Whitelam, *J. Comput. Chem.* **2014**, *35*, 360–370.

[159] M. Zhao, K. J. Lachowski, S. Zhang, S. Alamdari, J. Sampath, P. Mu, C. J. Mundy, J. Pfaendtner, J. J. De Yoreo, C. L. Chen, L. D. Pozzo, A. L. Ferguson, *Biomacromolecules* **2022**, *23*, 992–1008.

[160] M. Kang, P. Zhang, H. Cui, S. M. Loverde, *Macromolecules* **2016**, *49*, 994–1001.

[161] M. D. Daily, M. D. Baer, C. J. Mundy, *J. Phys. Chem. B* **2016**, *120*, 2198–2208.

[162] M. Zhao, J. Sampath, S. Alamdari, G. Shen, C. Chen, C. J. Mundy, J. Pfaendtner, A. L. Ferguson, *J. Phys. Chem. B* **2020**, *124*, 7745–7764.

[163] J. S. Schreck, F. Romano, M. H. Zimmer, A. A. Louis, J. P. K. Doye, *ACS Nano* **2016**, *10*, 4236–4247.

[164] C.-H. Yu, M. J. Buehler, *APL Bioeng.* **2020**, *4*, 016108.

[165] Z. Qin, L. Wu, H. Sun, S. Huo, T. Ma, E. Lim, P. Y. Chen, B. Marelli, M. J. Buehler, *Extrem. Mech. Lett.* **2020**, *36*, 100652.

[166] J. Wang, S. Lisanza, D. Juergens, D. Tischer, J. L. Watson, K. M. Castro, R. Ragotte, A. Saragovi, L. F. Milles, M. Baek, I. Anishchenko, W. Yang, D. R. Hicks, M. Expòsit, T. Schlichthaerle, J.-H. Chun, J. Dauparas, N. Bennett, B. I. M. Wicky, A. Muenks, F. DiMaio, B. Correia, S. Ovchinnikov, D. Baker, *Science (80-.).* **2022**, *377*, 387–394.

[167] J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D.

Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, D. Baker, *Science (80-.).* **2022**, *378*, 49–56.

[168] A. Belianinov, A. V. Ievlev, M. Lorenz, N. Borodinov, B. Doughty, S. V. Kalinin, F. M. Fernández, O. S. Ovchinnikova, *ACS Nano* **2018**, *12*, 11798–11818.

[169] Y. F. Dufrêne, D. Martínez-Martín, I. Medalsy, D. Alsteens, D. J. Müller, *Nat. Methods* **2013**, *10*, 847–854.

[170] D. Alsteens, D. J. Müller, Y. F. Dufrêne, *Acc. Chem. Res.* **2017**, *50*, 924–931.

[171] J. J. Schwartz, D. S. Jakob, A. Centrone, *Chem. Soc. Rev.* **2022**, *51*, 5248–5267.

[172] D. E. Otzen, M. S. Dueholm, Z. Najarzadeh, T. P. J. Knowles, F. S. Ruggeri, *Small Methods* **2021**, *5*, 2001002.

[173] F. S. Ruggeri, J. Habchi, S. Chia, R. I. Horne, M. Vendruscolo, T. P. J. Knowles, *Nat. Commun.* **2021**, *12*, 688.

[174] G. Ramer, F. S. Ruggeri, A. Levin, T. P. J. Knowles, A. Centrone, *ACS Nano* **2018**, *12*, 6612–6619.

[175] C. Vitali, R. J. B. Peters, H. G. Janssen, M. W. F. Nielen, F. S. Ruggeri, *TrAC - Trends Anal. Chem.* **2022**, *157*, 116819.

[176] C. Paluszkiewicz, N. Piergies, P. Chaniecki, M. Rękas, J. Miszczyk, W. M. Kwiatek, *J. Pharm. Biomed. Anal.* **2017**, *139*, 125–132.

[177] X. Zhao, D. Li, Y. H. Lu, B. Rad, C. Yan, H. A. Bechtel, P. D. Ashby, M. B. Salmeron, *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, 1–7.

[178] J. Heberle, E. Pfitzner, *J. Phys. Chem. Lett.* **2020**, *11*, 8183–8188.

[179] J. P. Cosas Fernandes, C. E. Federico, E. Lentzen, N. Valle, B. Basterra-Beroiz, M. Weydert, R. Quintana, *Polym. Test.* **2023**, *121*, 107996.

[180] S. Whitelam, I. Tamblyn, **2020**, *052604*, 1–13.

[181] C. W. Coley, D. A. Thomas, J. A. M. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. Rogers, H. Gao, R. W. Hicklin, P. P. Plehiers, J. Byington, J. S. Piotti, W. H. Green, A. J. Hart, T. F. Jamison, K. F. Jensen, *Science (80-.).* **2019**, *365*, DOI 10.1126/science.aax1566.

[182] M. J. Tamasi, R. A. Patel, C. H. Borca, S. Kosuri, H. Mugnier, R. Upadhya, N. S. Murthy, M. A. Webb, A. J. Gormley, *Adv. Mater.* **2022**, *34*, 1–12.

[183] Y. Liu, M. Ziatdinov, S. V. Kalinin, *ACS Nano* **2022**, *16*, 1250–1259.

[184] M. Christensen, L. P. E. Yunker, F. Adedeji, F. Häse, L. M. Roch, T. Gensch, G. dos Passos Gomes, T. Zepel, M. S. Sigman, A. Aspuru-Guzik, J. E. Hein, *Commun. Chem.* **2021**, *4*, 1–12.

[185] Z. Zhou, X. Li, R. N. Zare, *ACS Cent. Sci.* **2017**, *3*, 1337–1344.

[186] C. Zhu, C. U. Idemudia, W. Feng, *Informatics Med. Unlocked* **2019**, *17*, 100179.

[187] K. J. Lachowski, K. Vaddi, N. Y. Naser, F. Baneyx, L. D. Pozzo, *Digit. Discov.* **2022**, *1*, 427–439.

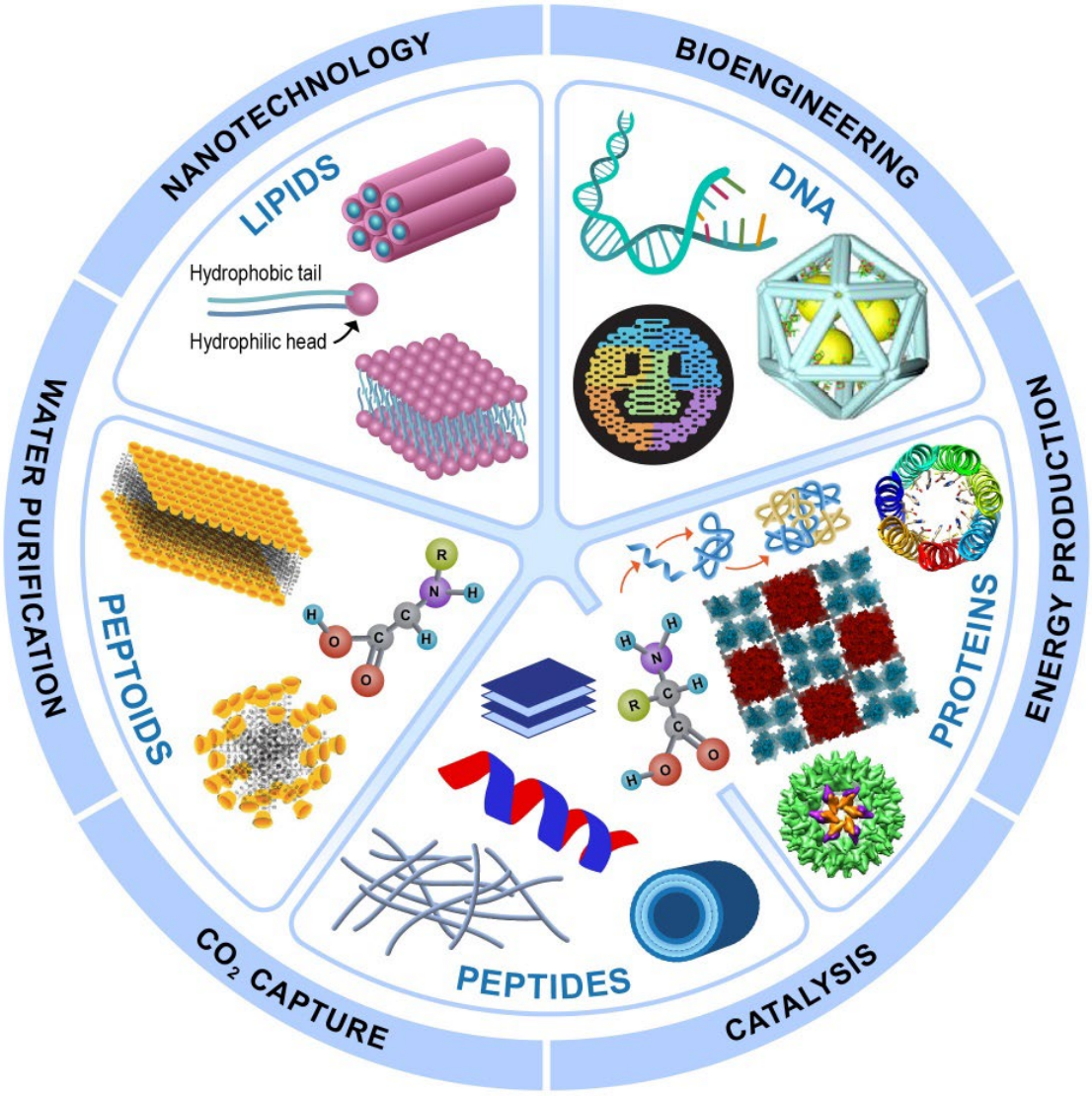
[188] K. Vaddi, H. T. Chiang, L. D. Pozzo, *Digit. Discov.* **2022**, *1*, 502–510.

[189] M. Valleti, R. K. Vasudevan, M. A. Ziatdinov, S. V. Kalinin, *Digit. Discov.* **2022**, *1*, 910–925.

[190] S. Zhang, R. Sadre, B. A. Legg, H. Pyles, T. Perciano, E. W. Bethel, D. Baker, O. Rübel, J. J. De Yoreo, *Proc. Natl. Acad. Sci.* **2022**, *119*, DOI 10.1073/pnas.2020242119.

[191] A. Ghosh, B. G. Sumpter, O. Dyck, S. V. Kalinin, M. Ziatdinov, *npj Comput. Mater.* **2021**, *7*, 1–8.

[192] H. S. Stein, J. M. Gregoire, *Chem. Sci.* **2019**, *10*, 9640–9649.



Frontispiece. Biomolecular assembly from monomers to hierarchical structures with complex functionalities and applications in various fields.

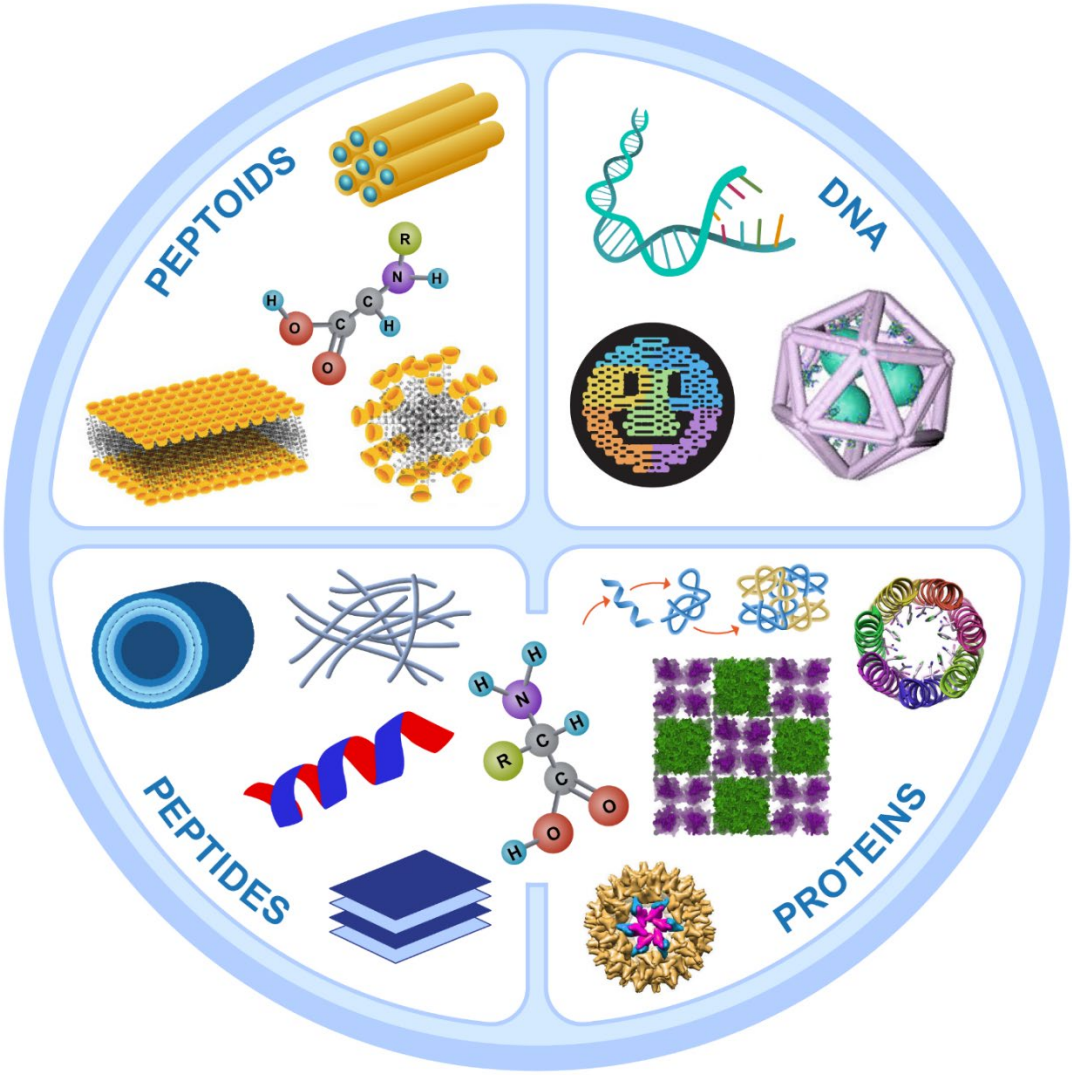


Table of Content: Inspired by nature, biomolecular self-assembly is a precise and adaptable bottom-up approach to synthesizing hierarchical biomaterials across scales. Scientists have long been developing artificial nanostructures, via biomolecular self-assembly, using DNA, peptides, proteins and peptoids as the building blocks to mimic the hierarchical functional materials found in biological systems for considerable energy, health, environment, sustainability, and

information technology applications. In addition to characterize the assembly outcomes, it is essential to directly observe the assembly dynamics and structural evolutions, which determine the biomolecular-assembly structures and properties, using *in situ* characterization approaches. Those measurements reflect the underlying energy landscape and the assembly mechanism for improving the efficiency and accuracy of biomaterials design.

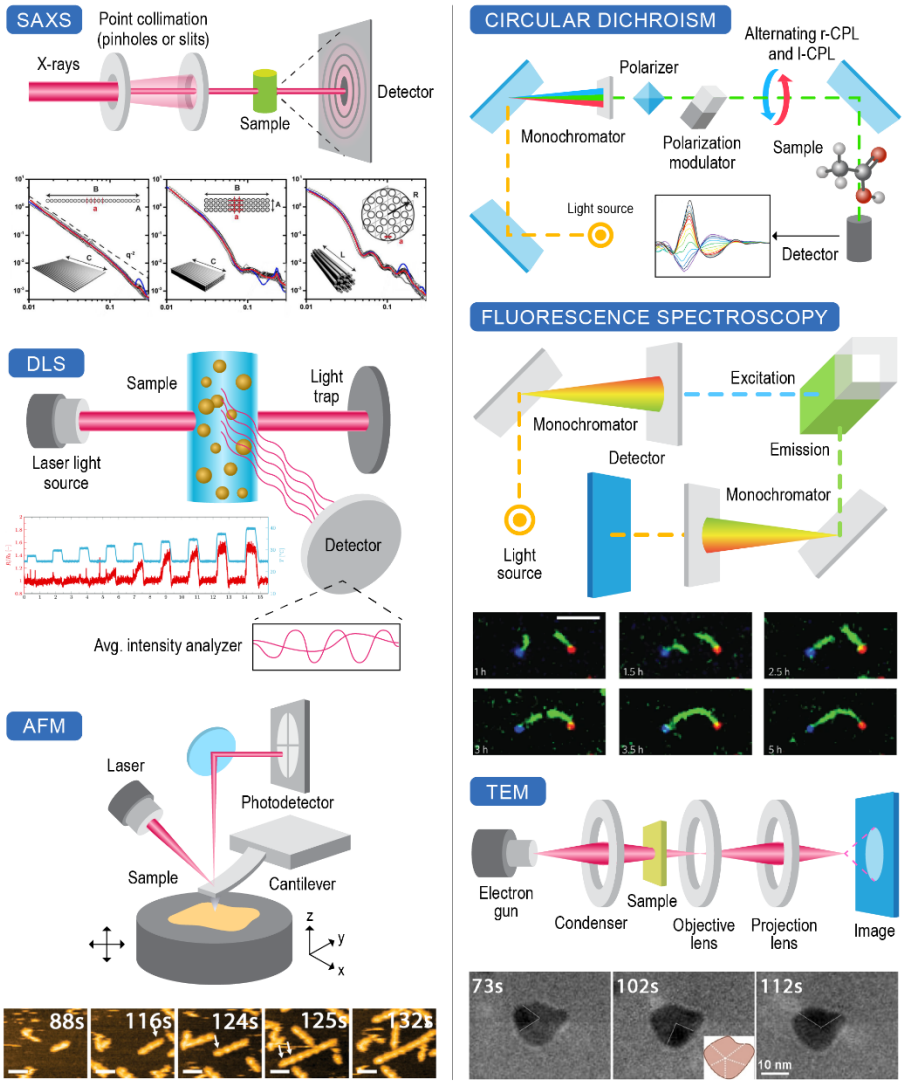


Figure 1. Schematic of various *in situ*, scattering, absorption, emission, and microscopy-based techniques used in the study of assembly dynamics.

Table 1. Comparison of techniques for *in situ* characterization of biomolecule self-assembly

Technique	Basic Principle	Resolution	Limitations
SAXS	· Measures elastically scattered X-rays from individual electrons · High x-ray flux from synchrotron radiation allows for high temporal resolution and low sample volumes	· Temporal resolution (TR): Milliseconds · Spatial resolution (SR): 1nm to 1 µm	· Destructive and weak sensitivity to biomolecules
SANS	· Measures elastically scattered neutrons from atomic nuclei · Can obtain a temporal resolution of seconds and obtain structural information over multiple length scales · Contrast variation and deuteration can be used to obtain the scattering of selected biomolecules · Neutrons are nondestructive to biomolecules	· TR: Seconds to minutes · SR: 1nm to 1 µm	· Lower temporal resolution compared to SAXS · Larger required sample size compared to SAXS
SLS & DLS	· Visible light is elastically scattered by suspensions due to differences in refractive indices · Characterization of molar mass, radius of gyration, second virial coefficient, fractal dimension, and hydrodynamic diameter	· SR: 10s to 100s of nm · TR: depends on sample, typically at least 1 min	· Limited information of fine changes in structures smaller than the wavelength of light
FTIR	· Infrared absorption spectra report on molecular vibrational modes · Changes in hydrogen bonding of backbone can be related to protein secondary structure	· TR: Rapid-scan mode measurements on order of 10s of ms	· Significant overlap of absorption bands · Resolving biomolecular signal from background is not trivial, and may require high concentrations

CD	· Differential absorption of polarized light is sensitive to molecular conformation · Databases of protein spectra can be used to analyze protein secondary structure	· TR: low bandwidth 10s of ms, high bandwidth 10s of ms possible with SR-CD	· Limited to chiral molecules or supramolecular assemblies · Connecting spectra to structural changes difficult for systems other than proteins/peptides
Fluorescence	· Light emission reports on changes in local environment and proximity to other probes as result of self-assembly · Changes in fluorescence quantifies the concentration of self-assembled structures	· Transient grating allows TR on order of 100s of ns · Typical measurements on order of minutes	· Introduction of probes can alter assembly pathways
AFM	· A cantilever with a sharp probe scans the surface and gives topographical and mechanical properties of the surface · Amplitude modulated “tapping” mode is suitable for sensitive biological samples	· SR: Few Å to microns · TR: 50-100 fps for HS-AFM	· Cannot image bulk processes
TEM	· Uses a well-defined electron beam to probe a specimen · Transmitted electrons are projected on a screen to obtain a visible image · Graphene liquid cells prevent radiation damage	· SR: Nano to micrometer · TR: 5-10 fps	· Organic samples are sensitive to electron beam damage · Not suitable for imaging bio interfaces

Cryo-EM	· Uses electron beam to image samples fixed in glassy ice · Prevents the damage caused by electron radiation to the organic matter	· SR: As low as few Å	· Not an *in situ* technique

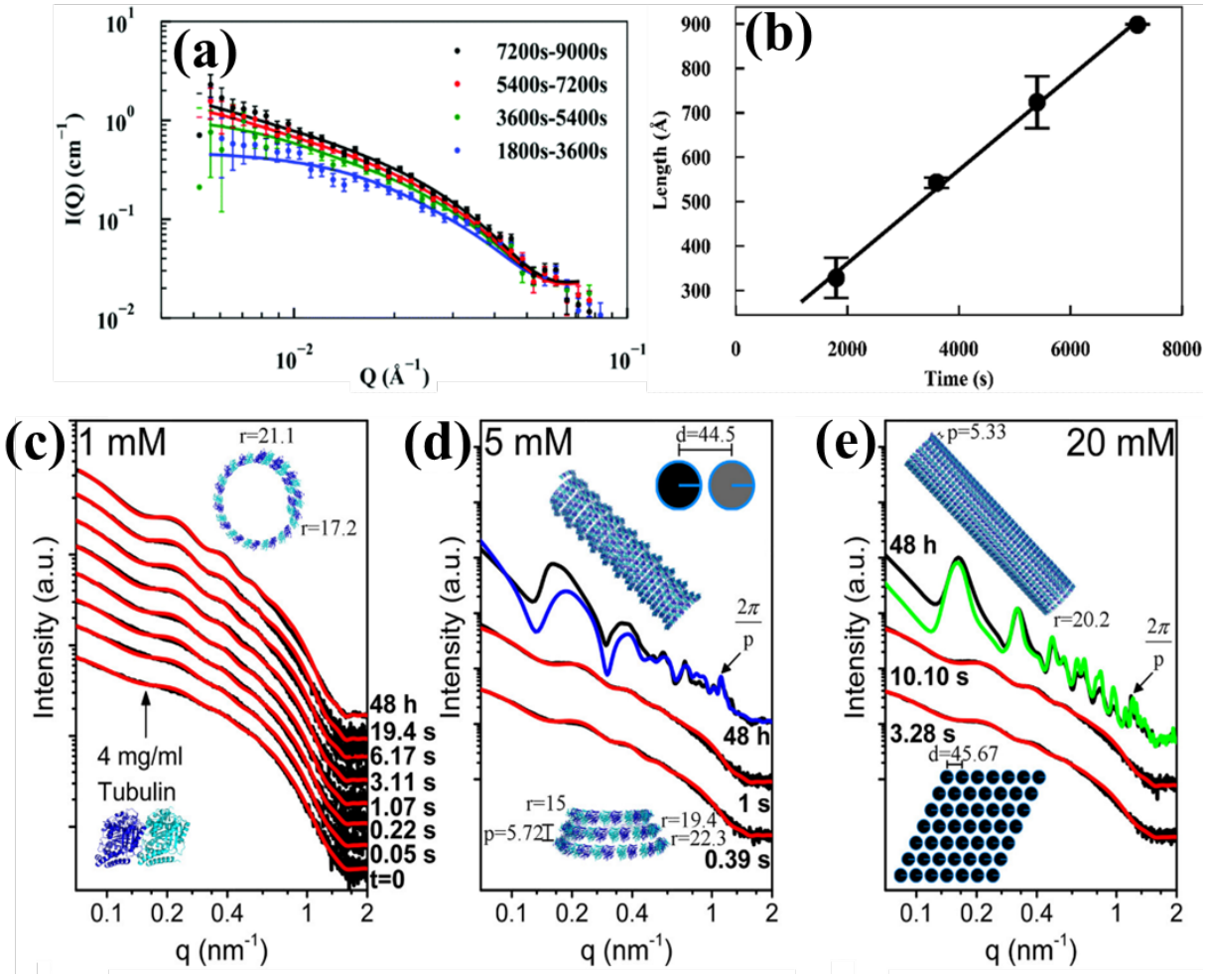


Figure 2. (a-b) Time resolved SANS of hydrogenated α-synuclein fibril seeds after mixing with α-synuclein monomers (scatter plot). The data were fitted (solid lines) with a cylindrical model in order to extract the elongation rate over time, as shown in **(b)** (adapted with permission from Eves et al.[28] with permission from the Royal Society of Chemistry). **(c-e)** Time-resolved SAXS results for the assembly of the tubulin protein (black curves) along with modeled curves (colored curves) of short conical-spirals **(c)**, conical-spiral tubules **(d)**, and inverted tubules **(e)** (adapted with permission from Dharan et al.[35]).

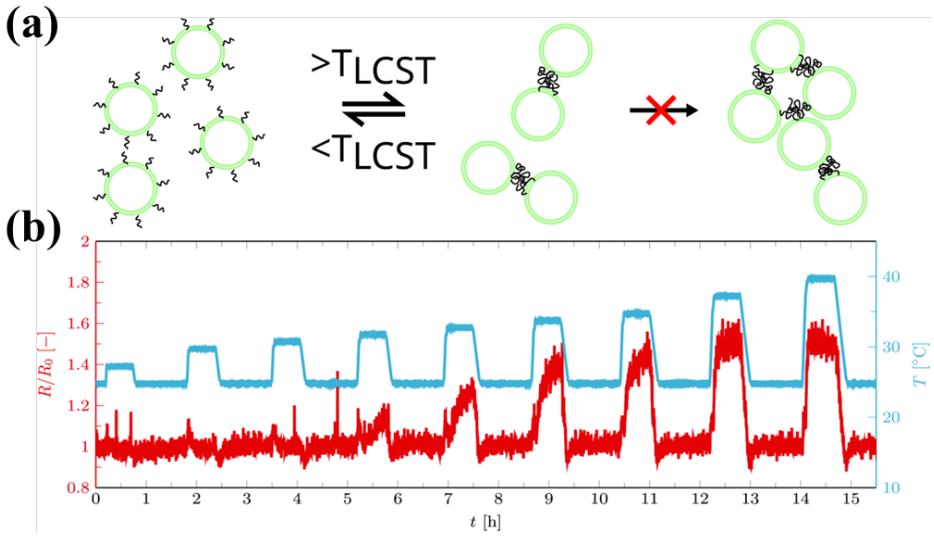


Figure 3. (a) Vesicles composed of a mixture of a lipid and an alkyl tail conjugated temperature sensitive polymer underwent controlled aggregation into doublets above the lower critical solution temperature of the polymer. **(b)** Continuous measurement of the normalized vesicle radius (red) during a temperature program (blue) using DLS which indicates the formation of doublets based on normalized hydrodynamic radius (adapted with permission from De Lange et al.[43]).

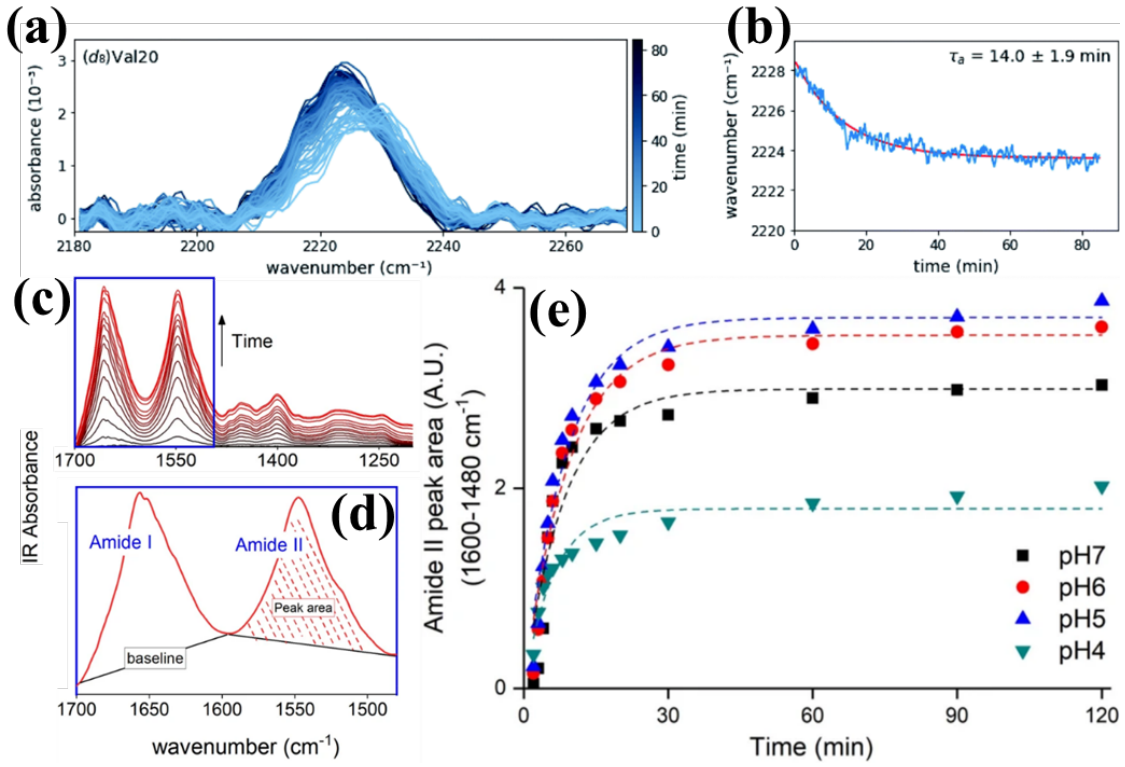


Figure 4. **(a)** Time resolved FTIR of MAX1 gelation with valine at sequence position 20 substituted with deuterated valine. **(b)** Exponential fit to the changes in peak position shown in **(a)** as a function of time to extract a time constant associated with self-assembly (adapted with permission from Adams et al. [47]). **(c-e)** Analysis procedure for FTIR data of bovine serum albumin (BSA) adsorbing to a hematite surface. **(c)** shows the changes in FTIR spectra over time. The method for integrating the area of the amide II peak is shown in panel **(d)**. This analysis was performed for multiple pH conditions to track changes in BSA secondary structure over time as shown in panel **(e)** (adapted with permission from Barreto et al.[55]).

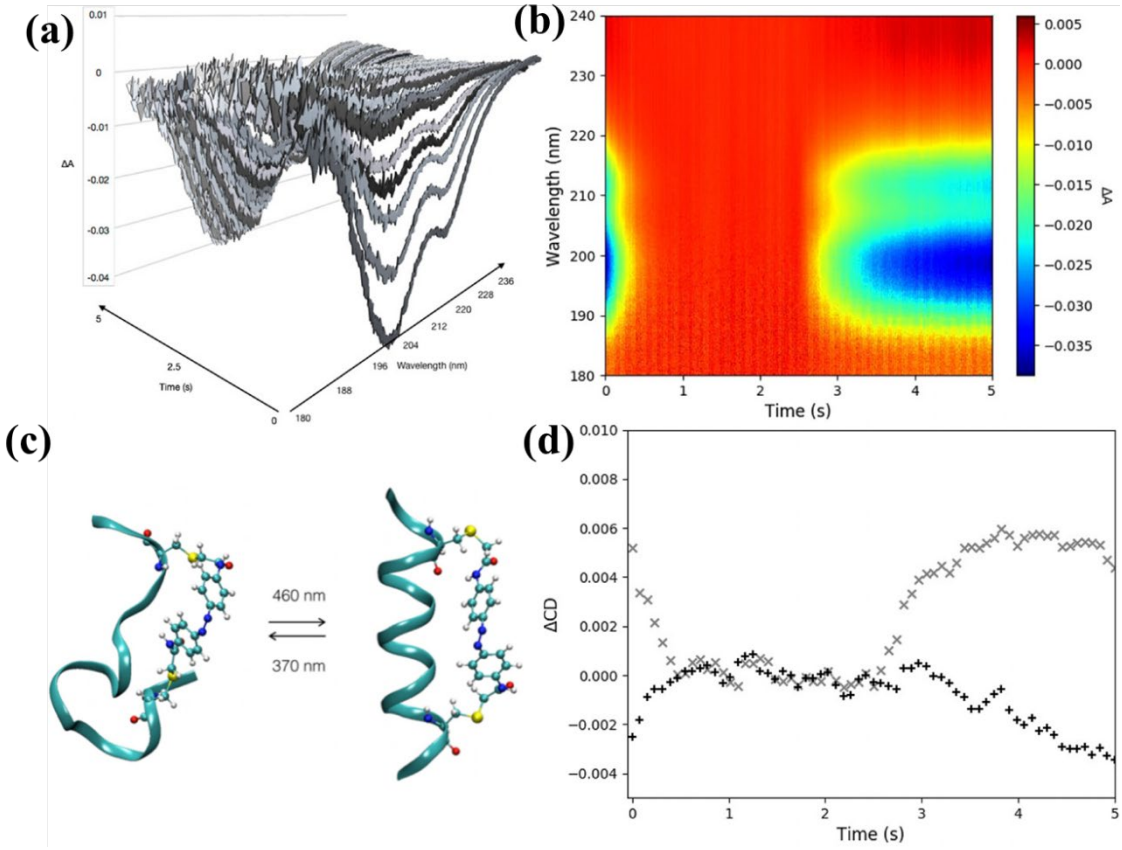


Figure 5. Time-resolved synchrotron radiation circular dichroism measurements of photoisomerization events in an azobenzene modified peptide. The spectra presented in **(a)** show the changes in CD with 70 ms resolution. A 2D representation of the data in **(b)** shows the change in peaks at 198 nm and 212 nm indicating changes to the secondary structure of the peptide as a result of cis to trans isomerization, which occurs after the pulse at 2.5s. **(c)** Schematic of conformational changes to the FK-11-X peptide because of azobenzene photoisomerization. **(d)** Changes in CD at 190 nm (x) and 204 nm (+) that demonstrate the reversibility of conformational changes (adapted with permission from Auvray et al.[23]).

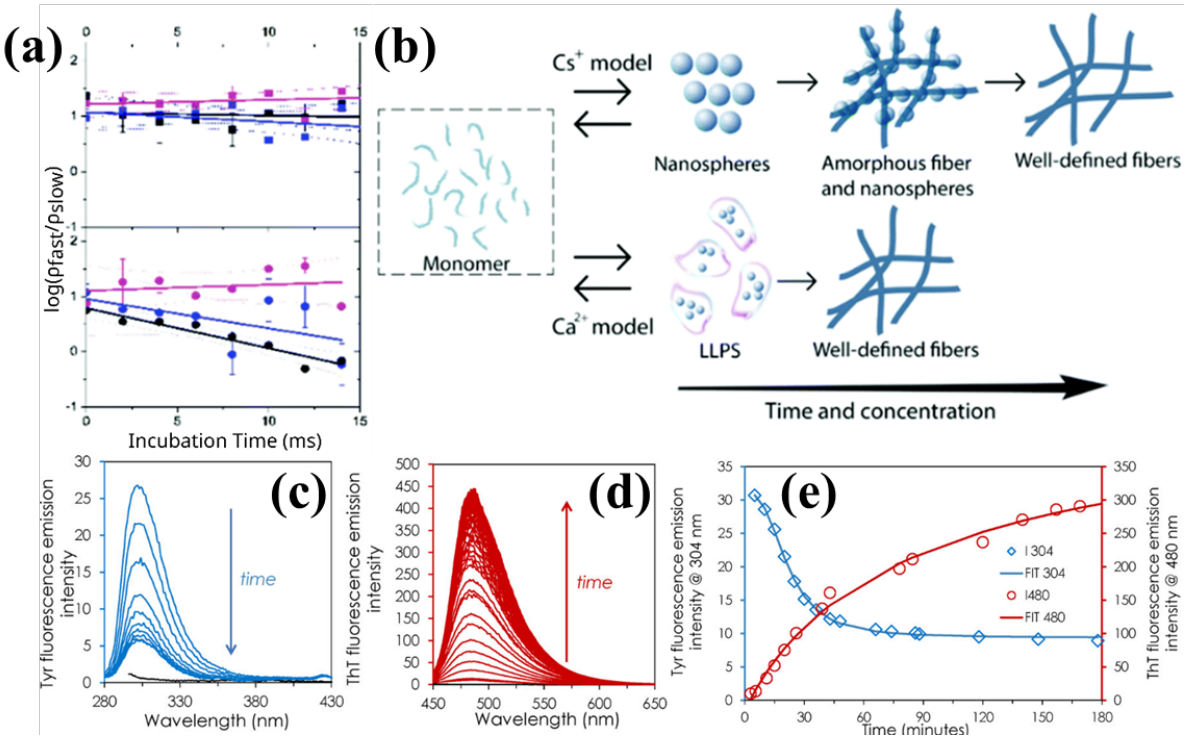


Figure 6. (**a**) The diffusion times of fast diffusing species (e.g., monomers) and slow diffusing species (e.g., nuclei, aggregates) were obtained from FLCS (6.25 mM magenta, 12.5 mM blue, or 25 mM black) of Ca^{2+} (squares) or Cs^{+} (circles). A logarithm of the ratio of fast to slow diffusing species is fit to obtain the kinetics of the equilibrium shift. **(b)** The model for self-assembly pathways in the presence of either Ca^{2+} or Cs^{+} (adapted with permission from Mañas-Torres et al.[73]). **(c-d)** show the decrease and increase in fluorescence in intensity at wavelengths corresponding to tyrosine and ThT, respectively. The former reports on the stacking and quenching of tyrosine fluorescence and the latter indicates an increase in the mass of amyloid fibrils. **(e)** A logistic equation was used to fit the changes in fluorescence in figures **(c)** and **(d)** to quantitatively evaluate self-assembly dynamics (adapted with permission from Gobeaux et al. [46]).

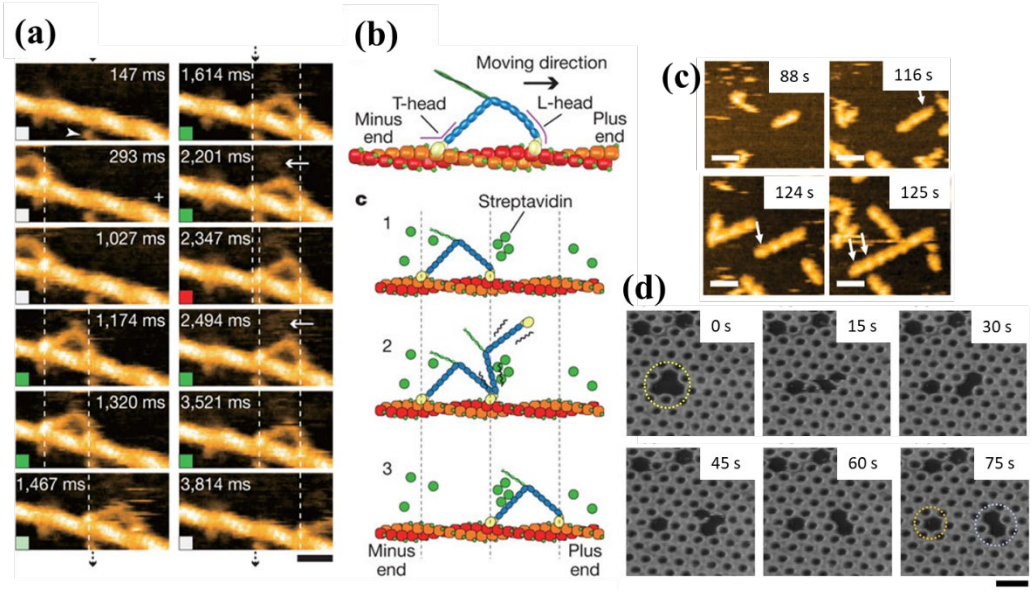


Figure 7. (a) HS-AFM images showing progressive movement of two headed Myosin V motor protein. Scale bar is 30nm. **(b)** Schematic of Myosin V movement captured by HS-AFM (adapted with permission from Kodera et al.[88]). **(c)** End-to-end interaction of protein nanorods showing elongation along the longitudinal direction. Scale bars are 25nm (adapted with permission from Kikuchi et al.[90]). **(d)** Defect splitting and defect diffusion of a DNA origami honeycomb lattice monitored via HS-AFM. Scale bar is 100nm (adapted with permission from Suzuki et al.[91]).

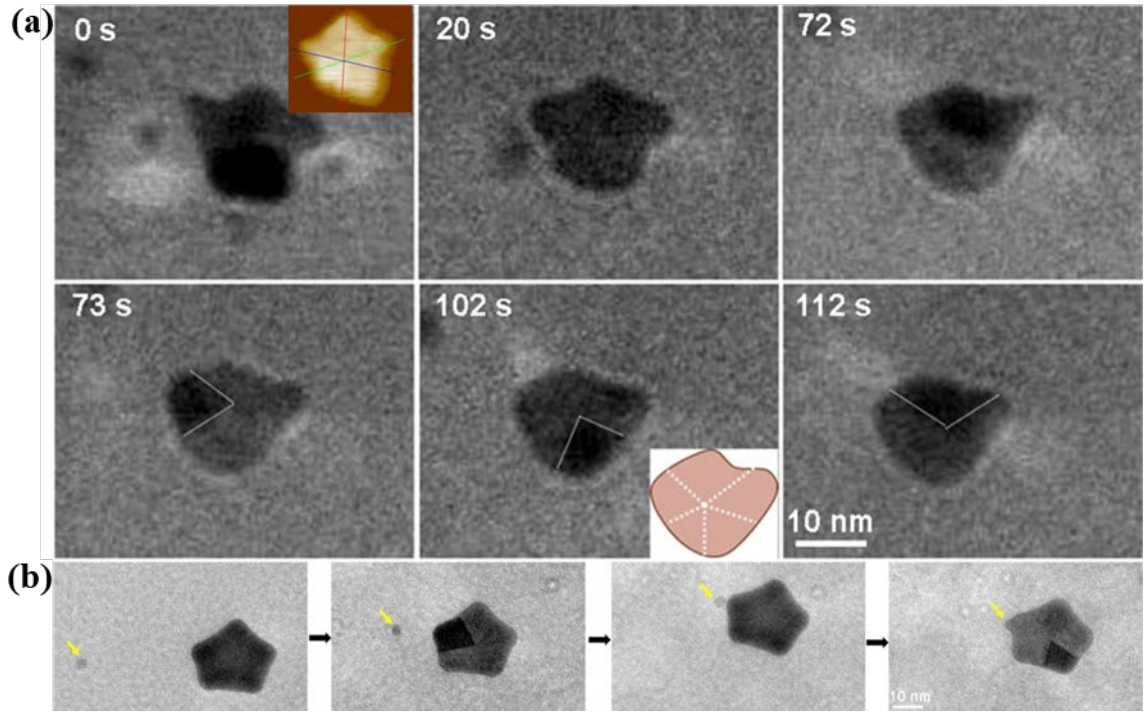


Figure 8. (a) Time-dependent LP-TEM reveal the stages of peptoid-mediated Au nanostar formation through particle attachment. Inset: AFM image of the assembled five-fold twinned Au nanostar. **(b)** The growth of a regular five-fold twinned nanocrystal into a concave Au star by the attachment of small nanocrystals (marked by a yellow arrow) on the corners (adapted with permission from Jin et al.[93]).

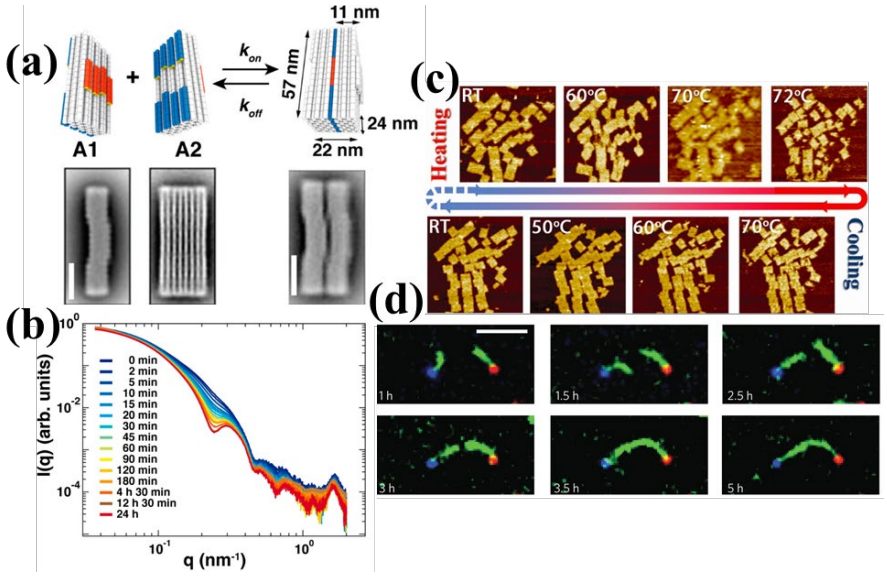


Figure 9. (a) Schematic of DNA origami brick monomer transformation into a dimeric block in the presence of 20mM $MgCl_2$. Scale bar is 20nm. **(b)** Time evolution of scattering profiles after 1:1 mixing of monomeric brick samples (adapted with permission from Bruetzel et al.[120]). **(c)** *In situ* AFM images under continuous heating from room temperature (RT) to 75 °C at a rate of 2 °C/min, followed by natural cooling to RT. The scan area was 3 μm × 3 μm (adapted with permission from Song et al.[121]). **(d)** Time-lapse multi-color fluorescence microscopy shows point-to-point assembly of a nanotube via polymer nucleation at molecular landmarks and growth by monomer attachment. Scale bar is 5μM (adapted with permission from Mohammed et al.[123]).

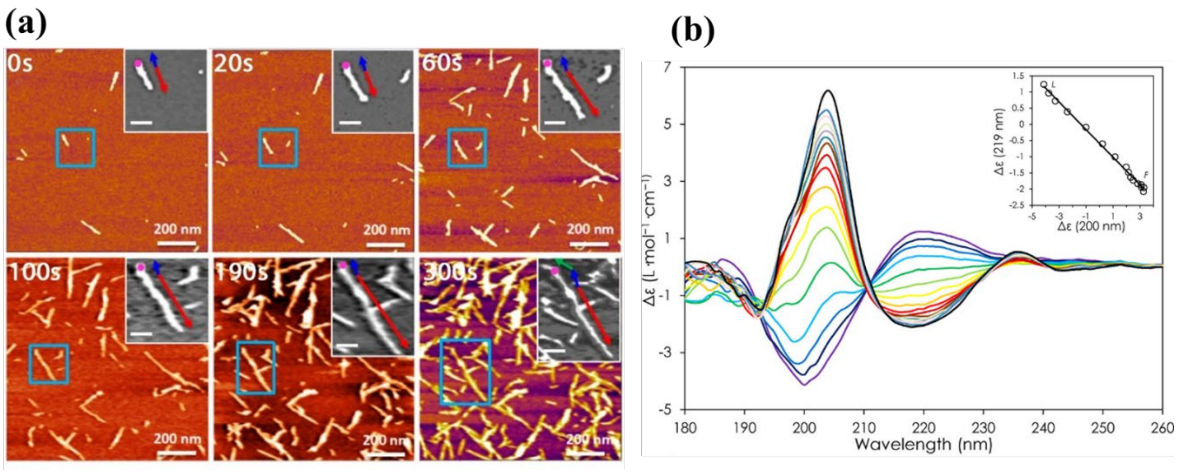


Figure 10. (a) *In situ* AFM images of amylin fibrillation (adapted with permission from Huang et al.[130]). **(b)** SRCD monitoring of the self-assembly of 5% atosiban solubilized in NH4Ac at pH 9. The first spectrum is plotted in purple and the last one in black. Inset: Phase diagram analysis of fibril assembly. L indicates the initial liquid state (monomers and dimers) and F the final gel state (mostly fibrils) (adapted with permission from Gobeaux et al. [46]).

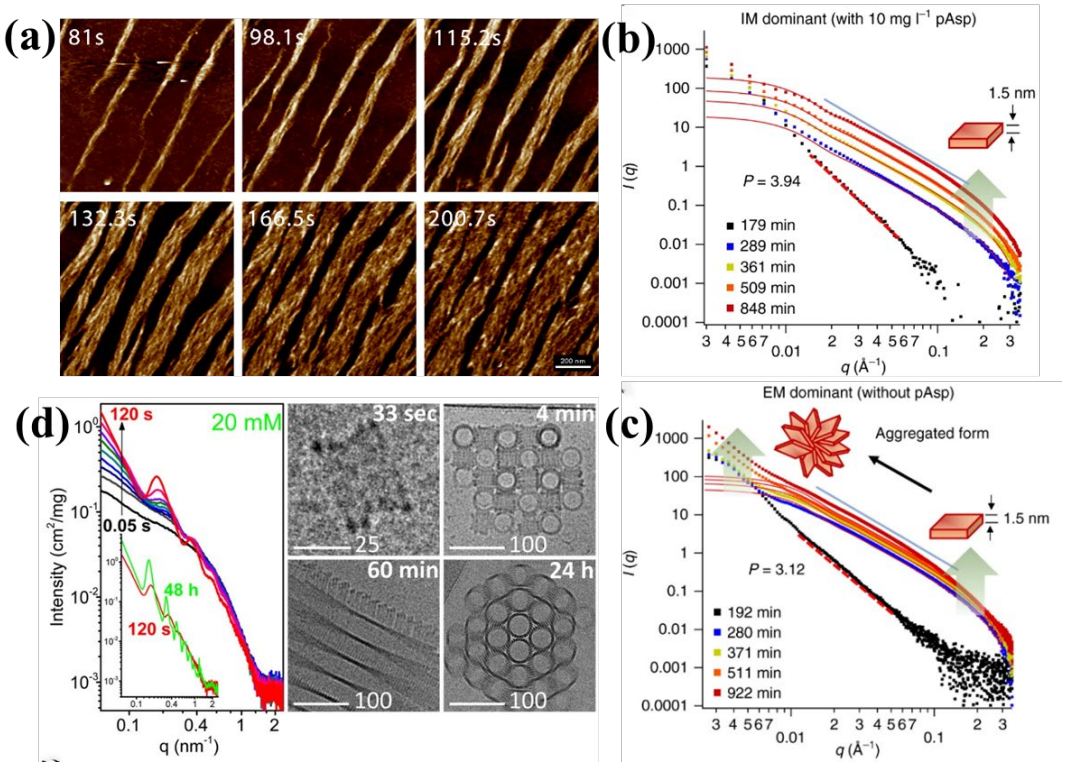


Figure 11. (a) Collagen type I fibrillogenesis imaged via *in situ* AFM (adapted with permission from Stamov et al.[137]). **(b-c)** SAXS patterns collected during mineralization with and without pAsp (adapted with permission from Kim et al.[138]). **(d)** Spermine induced tubulin assembly pathway captured via time-resolved synchrotron SAXS and time-lapse cryo-TEM (adapted with permission from Dharan et al.[35]).

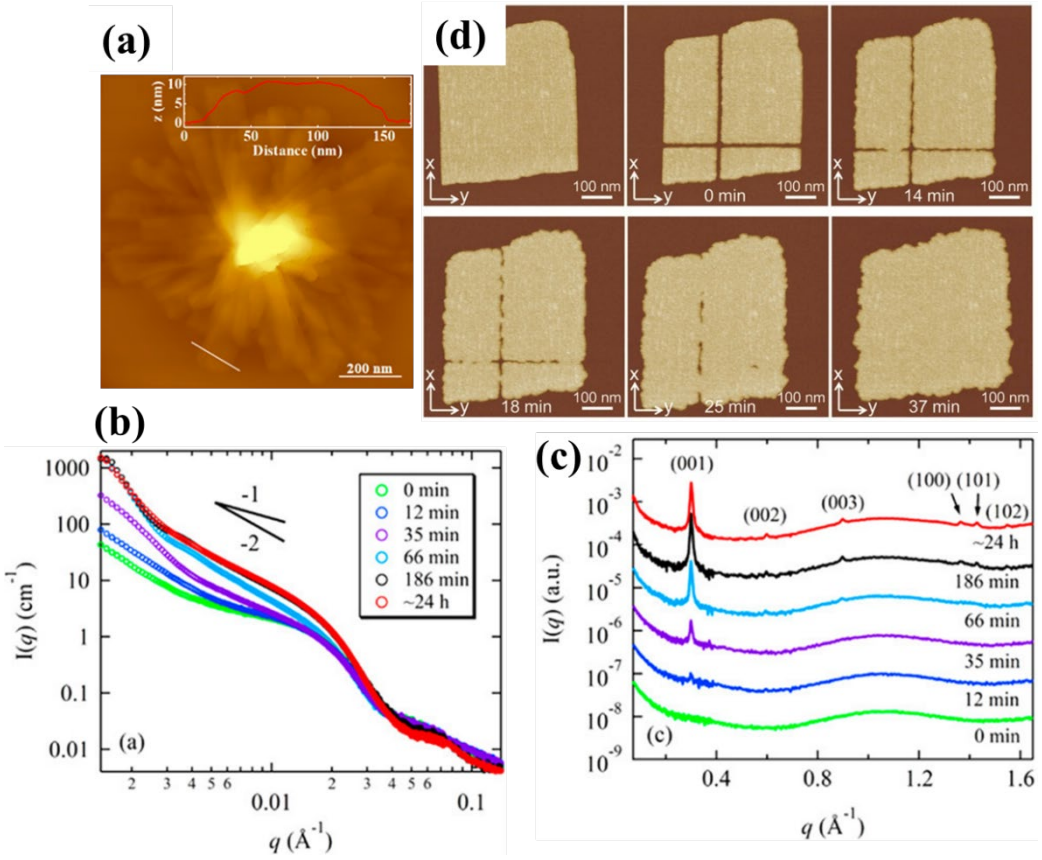


Figure 12. **(a)** AFM image of PNMG-b-PNOG showing the height profile (inset). **(b)** SAXS profiles of the PNMG-b-PNOG methanol solution at different times after being cooled down to room temperature. **(c)** Time-resolved WAXS data of the sample shown in **(b)** demonstrating long range order (adopted with permission from Kang et al.[143]). **(d)** *In situ* AFM images that show peptoid membrane repair (adapted with permission from Jiao et al.[145]).

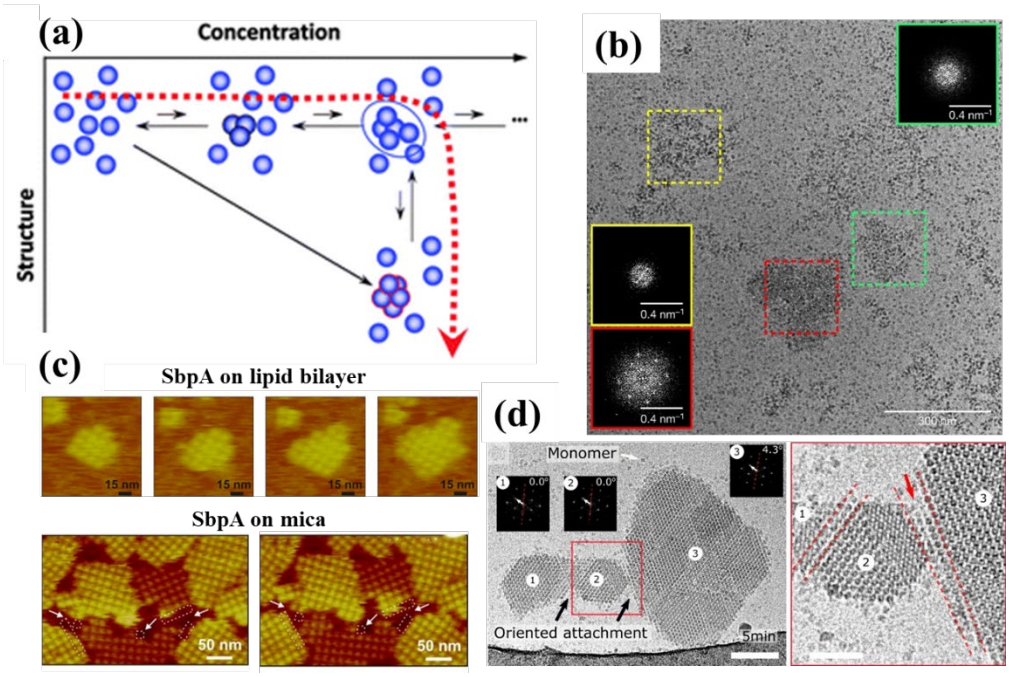


Figure 13. (a) Schematic illustration of the two-step mechanism of crystal nucleation (adapted with permission from Vekilov et al.[151]). **(b)** Ferritin aggregates with different ordering, observed by cryo-TEM (adapted with permission from Houben et al.[152]). **(c)** Sequential *in situ* AFM height images showing SbpA assembly on a lipid bilayer (adapted with permission from Chung et al.[153]) and on a mica surface (adapted with permission from Shin et al.[83]) . **(d)** Oriented attachment of individually nucleated glucose isomerase nanocrystals into a larger, merged lattice composed of domains 1 and 2. Scale bar is 100nm (left) and 50nm (right) (adapted with permission from Van Driessche et al.[109]).

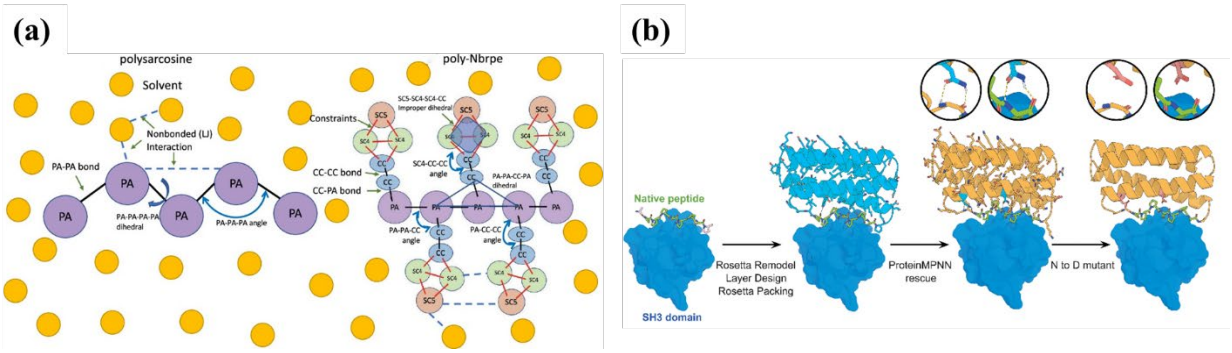


Figure 14. (a) Schematic illustration of the bonded and nonbonded interactions in coarse-grained models of polysarcosine and poly-Nbrpe in the solvent (adapted with permission from Zhao et al.[162]). **(b)** Protein designed by a message-passing neural network (adapted with permission from Dauparas et al.[167]).

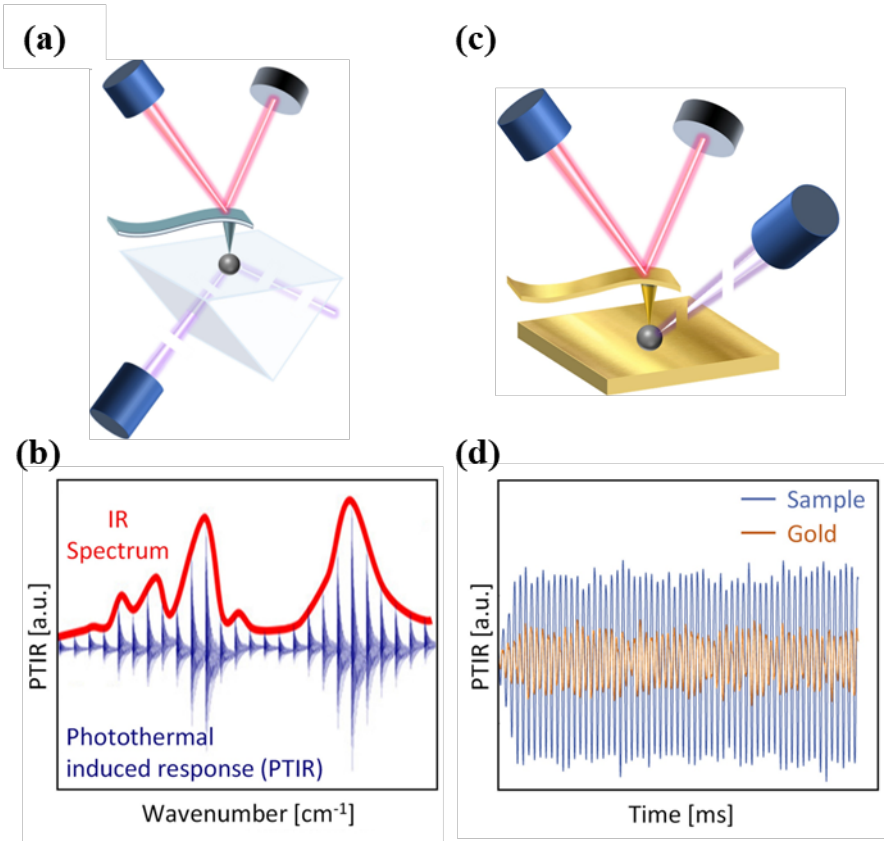


Figure 15. Schematic diagram of AFM-IR. **(a)** AFM-IR in the bottom illumination optical configuration. **(b)** Photothermal induced response (PTIR) as a function of the wavenumber results in an IR absorbance spectrum in the ring-down excitation scheme. **(c)** AFM-IR in top illumination requires the use of gold-coated probes to minimize self-absorption of IR light by the cantilever. **(d)** Typical PTIR in an off-resonance excitation scheme plotted against time. (adapted with permission from Vitali et al.[175]).

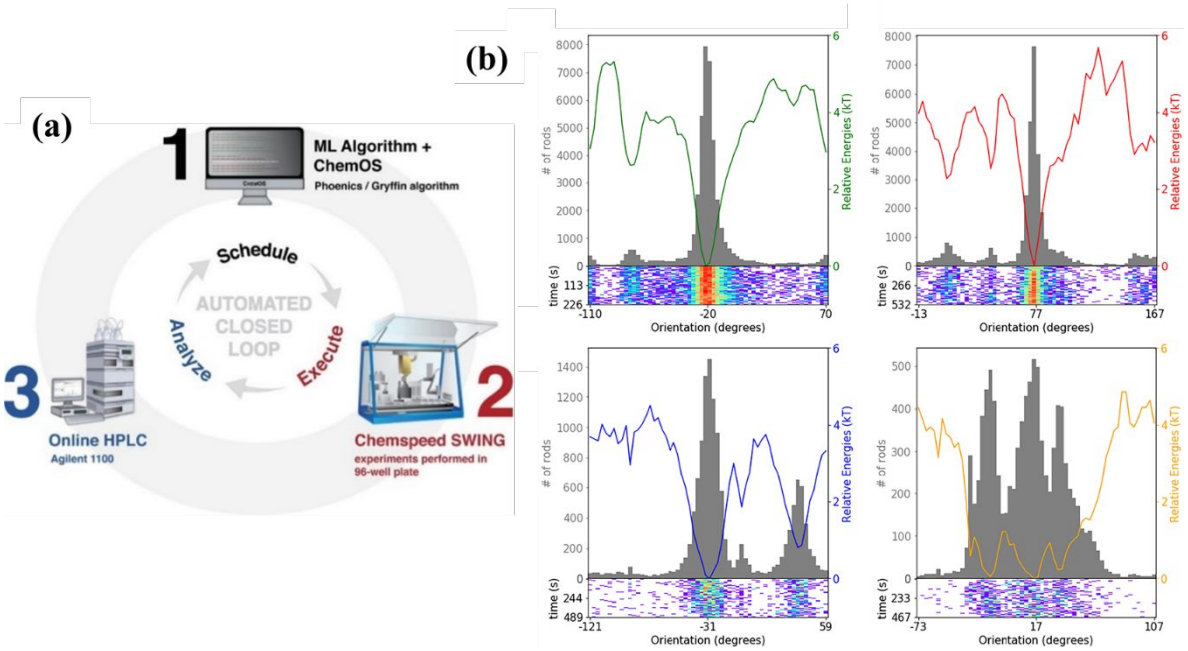


Figure 16. (a) Example of an automated self-driving platform using ChemOS to coordinate and plan experiments, Chemspeed SWING for the automated synthesis of experiments, and online HPLC for characterization (adapted with permission from Christensen et al.[184]). **(b)** The accumulated orientation distributions, derived relative energy landscapes, and orientation heat maps over the imaging time of protein nanorods elucidated from an HS-AFM video using a convolutional neural network (adapted with permission from Zhang et al[190]).