This work was written as part of one of the author's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law. Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it’s important to you. Thank you.
Research review paper

Protein folding and assembly in confined environments: Implications for protein aggregation in hydrogels and tissues

Laura W. Simpson\textsuperscript{a}, Theresa A. Good\textsuperscript{b}, Jennie B. Leach\textsuperscript{a,⁎}

\textsuperscript{a} Department of Chemical, Biochemical and Environmental Engineering, University of Maryland Baltimore County, Eng 314, 1000 Hilltop Circle, Baltimore, MD 21250, USA

\textsuperscript{b} Division of Molecular and Cellular Biosciences, National Science Foundation, 2415 Eisenhower Ave, Alexandria, VA 22314, USA

A R T I C L E   I N F O

Keywords:
Protein aggregation
Confinement
Crowding
Hydrogel
Protein stabilization

A B S T R A C T

In the biological milieu of a cell, soluble crowding molecules and rigid confined environments strongly influence whether the protein is properly folded, intrinsically disordered proteins assemble into distinct phases, or a denatured or aggregated protein species is favored. Such crowding and confinement factors act to exclude solvent volume from the protein molecules, resulting in an increased local protein concentration and decreased protein entropy. A protein’s structure is inherently tied to its function. Examples of processes where crowding and confinement may strongly influence protein function include transmembrane protein dimerization, enzymatic activity, assembly of supramolecular structures (e.g., microtubules), nuclear condensates containing transcriptional machinery, protein aggregation in the contexts of disease and protein therapeutics. Historically, most protein structures have been determined from pure, dilute protein solutions or pure crystals. However, these are not the environments in which these proteins function. Thus, there has been an increased emphasis on analyzing protein structure and dynamics in more “in vivo-like” environments. Complex in vitro models using hydrogel scaffolds to study proteins may better mimic features of the in vivo environment. Therefore, analytical techniques need to be optimized for real-time analysis of proteins within hydrogel scaffolds.

1. Introduction

Proteins are the building blocks of life and have highly diverse functions. The biophysical dogma of proteins is that structure predicts function (Burr and Koshland, 1964; Dunker et al., 2001; Guerrucci and Belle, 1995; Jez, 2017; Kendrew, 1959; McLachlan, 1972; Olesen, 1970; Orengo et al., 1999; Sandberg et al., 2002; Zheng et al., 2018): proper protein folding and supramolecular assembly yield appropriate function while misfolding and aggregation yield dysfunction. Indeed, processes of protein folding and aggregation are ubiquitous. Numerous theories and experimental tools have been established during the past century to study structure-function relationships of proteins. While these theories and tools have allowed great understanding of proteins in dilute solutions, extrapolation of protein structure-function relationships determined in dilute solution to their structure and dynamics in vivo has been problematic. As a stepping stone between the simplified system of a single component in dilute solution and the complexity of living organisms, scientists and engineers are exploring a number of complex 3D in vitro models that are designed to mimic features of the in vivo environment. Currently, these models are enabling great advancements in basic science (e.g., to unravel the roles of heterogeneous networks of factors on cellular processes and disease), as well as create new technologies (e.g., biopharmaceuticals designed for greater therapeutic efficacy).

In the early to mid 1900’s, new tools to visualize molecular structures were established, notably x-ray crystallography to determine the atomic crystal structure of molecules, electron microscopy (EM) to image specimens that are dried and stained with heavy metals, nuclear
magnetic resonance spectroscopy (NMR) to determine molecular structures without crystallization, and light scattering techniques to characterize molecular sizes and interactions in pure solutions (Bragg, 1914; Darrow, 1953; Marton, 1934; Ovannall and Peaker, 1959). One such method is circular dichroism (CD) which measures changes in polarized light that are signatures of different protein structures depending on the light wavelength: far-ultraviolet (UV) CD measures secondary structure, whereas near UV CD measures tertiary folding. Fourier transform infrared (FTIR) microspectroscopy measures the absorption or emission of IR light associated with secondary structure such as cross-β-pleated sheets. Conformational sensitive dyes monitor specific protein conformations; Thioflavin T (ThT) only fluoresces when bound in the grooves of stacked cross-β-sheets associated with amyloid proteins. With NMR, the backbone of each amino acid is labeled to identify the localization of secondary structural domains such as β-sheets and α-helices (Ami et al., 2013; Dobson, 2001; Kuznetsova et al., 2014). Identification of protein structure, size and function inform tissue physiology or disease pathology. These methods provide complementary information regarding the molecular resolution of static structures, morphology of static structures, near molecular resolution of dynamic structures, and dynamic measures of molecular assembly.

Building on ideas in physics as well as colloidal and polymer science, theorists have described the processes involved in protein folding by models of random packing of hard spheres, lattice statistical mechanics, and the driving forces of hydrophobic interactions and conformational entropy (Bernal and Finney, 1967; Chan and Dill, 1990; Dill, 1985; Dill et al., 1989; Stigter and Dill, 1989; Wolynes et al., 1995). Unfolded proteins are highly disordered and have high entropy; as hydrophobic regions are hidden during folding, the protein takes on a more highly ordered, lower entropy native state (Fig. 1). Models for protein-protein association and aggregation were developed using similar principles, enabling the prediction of populations of particles in solution (Fields et al., 1992; Fink, 1998; Hansen and Müller, 1996; Nyeo and Chu, 1989). These tools were developed for proteins in ideal solutions and thus it is uncertain whether the findings are reflective of protein structure and function in living organisms. For example, the interior of a cell is a crowded environment where proteins are confined and influenced by their dynamic and heterogeneous local environment. Therefore, to move from ideal solutions towards theories and measurements that better describe protein behaviors in vivo, it is critical to consider how a protein’s local environment influences its structure and function.

1.1. Crowding by a protein’s local environment

Solutions in living organisms contain a vast array of macromolecules. Inside the cell, proteins assemble into rigid structures (e.g., F-actin, microtubules, and intermediate filaments) (Ellis, 2001a; Zimmerman and Minton, 1993) and signaling complexes (e.g., DNA replication complex and RNA replication complex) (Cebecauer et al., 2010; Smith and Scott, 2002). In the nucleus, proteins condense around DNA, forming chromatin; protein transcriptional and repair machinery assemble around DNA (Pliss et al., 2019). Outside the cell, complex heterogeneous matrices composed of collagen, laminin, fibronectin, and other macromolecules aid in maintaining higher order cellular and tissue structures (Ellis, 2001a; Frantz et al., 2010). In these compartments, individual macromolecules only contribute a minute fraction of the total solution volume. However, collectively, the population of macromolecules occupy a substantial fraction (~30%) of the total solution volume (Zimmerman and Minton, 1993). Therefore, observations of proteins in pure dilute solutions do not fully represent how these molecules behave in intracellular and extracellular milieu in which there are substantial ranges in compositions and concentrations. Notably, macromolecules in vivo are physically crowded in concentrated “non-ideal” solutions; this crowding has profound effects on protein assembly and stability.

The phenomenon of macromolecular crowding was first applied in the 1940’s to describe how proteins precipitate in complex solutions (Cohen, 1942; Juckles, 1971; Laurent, 1963). However, protein folding theories at that time approximated the protein population as a set of non-interacting rigid spheres and apply well only to dilute solutions (Ellis, 2001a, 2001b; Zimmerman and Minton, 1993). Later, theories such as the available volume theory and the scaled particle theory were developed to better describe the non-ideal behaviors that macromolecules demonstrate in crowded solutions (Cheung et al., 2005; Hall and Minton, 2003; Zimmerman and Minton, 1993). These theories apply the idea that a macromolecule’s access to “available” solvent is compromised in crowded high concentration solution; in other words, a volume of solvent is excluded from interacting with the macromolecule.

The net effect of this excluded volume of solvent is a reduced available space for a protein to rearrange and fold, thereby provoking a structural state that is compact and low entropy (Fig. 2). These theories have been verified experimentally using a variety of crowding factors (e.g., albumin, dextran, polyethylene glycol, and ficoll), which were observed to exclude the protein from interacting with a volume of solvent, thus altering protein folding and function versus what is predicted and observed in “ideal” solutions of pure protein at dilute concentrations (Comper and Laurent, 1978; Edmond and Ogston, 1968; Laurent, 1963; Laurent and Ogston, 1963; Minton, 1980, 1981, 1983).

These theories that describe macromolecular crowding consider a macromolecule’s altered diffusivity and solute-solute interactions that contribute to the non-ideal nature of the solution. Classical methods of measuring the thermodynamic activity (non-ideal correction parameter) of macromolecules in solution include sedimentation...
equilibrium, osmometry, and Rayleigh scattering of light (Rivas et al., 1999; Zimmerman and Minton, 1993). The crowding effect is specific to the macromolecular crowder and the intermolecular interactions that arise between crowder and protein; charge, hydrophobicity, viscosity, and rigidity of the crowder (or environment) all play a role (Batra et al., 2009; Kuznetsova et al., 2014). Generally, high concentrations of large molecular weight crowding factors significantly increase protein-protein interactions. However, when the available volume is decreased, molecular interactions become more limited, and thus, the effects from hindered diffusion overcomes thermodynamic forces, resulting in reduced rates of enzymatic reactions, protein assembly, and aggregation (Ellis, 2001a; Minton, 2000).

1.2. Aggregation

Protein aggregation may occur when proteins are misfolded, by exposing sites for non-specific protein interactions, thereby resulting in higher-order structures (e.g., disordered clumps or organized fibrils). In vivo, genetic mutations can result in protein structures that are more likely to aggregate than in wild-type organisms, resulting in altered protein function and possibly disease (Chiti and Dobson, 2017; Sipe et al., 2016). Protein aggregation rarely occurs in healthy tissues but is often correlated with disease and may readily occur within in vitro solutions when protein structure is disrupted or destabilized (Hatters et al., 2002; Van den Berg et al., 1999).

Aggregation prone proteins often require a molecular chaperone to avoid unwanted aggregation. The protein lysozyme in pure solution is prone to aggregation, which is exacerbated in the presence of crowders bovine serum albumin (BSA) and ovalbumin. However, when protein disulfide isomerase (molecular chaperone) is added to the denatured lysozyme solution, lysozyme refolds into its native enzymatically active state. This refolding process is accelerated in the presence of crowders due to increased molecular associations between lysozyme and chaperone protein disulfide isomerase (Van den Berg et al., 1999). The excluded volume effect stabilizes protein structure; however, the structure may be unwanted aggregation.

As described before, the excluded volume effect imparted by crowding factors stabilizes the lower entropy conformation of the native protein. However, aggregation prone proteins and intrinsically disordered proteins (IDP) may stabilize lower entropy nuclei for

Fig. 2. Schematic free energy surface (F) of hen lysozyme consisting of two folding domains designated as α and β (red and yellow respectively). The unfolded state has high energy and is highly disordered. As the α and β domains take on intermediate conformations, the energy decreases and the fraction of native connections approaches unity forming the low entropy native state. Reproduced from (Dobson, 2001) with copyright permission from The Royal Society. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
aggregation. In a crowded solution, the effective protein concentration is increased, resulting in a higher likelihood of protein-protein interactions and formation of nuclei, which accelerates the growth of filamentous aggregates (Batra et al., 2009; Bokvist and Grobner, 2007; Borrero and Escobedo, 2006; Luo et al., 2016; Munishkina et al., 2004; Zhou et al., 2009).

1.3. Intrinsically disordered proteins

The importance of IDPs both in normal biological processes and in abnormal pathophysiology is becoming increasingly evident (Banks et al., 2018; Dunker et al., 2001; Shin and Brangwynne, 2017). Intrinsically disordered proteins (or regions) lack defined structure. This conformational freedom allows IDPs to promiscuously bind to diverse partners, playing key roles in protein complex assembly, function and regulation (Uversky, 2015). NMR 3D structural determination has led to the discovery of numerous proteins with functional, yet disordered regions (e.g., amyloid-β, α-synuclein and prion protein) (Chiti and Dobson, 2017; Dunker et al., 2001).

IDPs are particularly susceptible to the effect of crowders given their conformational flexibility. As stated above for proteins in general, the crowding effect is specific to the crowder and the intermolecular interactions that arise between crowder and protein; this concept also applies to IDPs (Banks et al., 2018; Goldenberg and Argyle, 2014; Miller et al., 2016; Nguemaha et al., 2018; Qin and Zhou, 2013). To ensure the continued advances in our understanding of IDPs in normal and pathological biological function, IDPs studied in vitro must consider well-characterized crowded and confined environments.

1.4. Confinement

Protein folding, assembly, and aggregation occur within the cell, at the cell membrane, and in the extracellular space. The terms “crowding” and “confinement” are often misused interchangeably to describe an increase in protein stability due to excluded solvent volume accessible for protein folding, resulting in more ordered (lower entropy) protein conformations. However, these terms refer to different mechanisms by which protein structure and diffusivity are altered by interactions with neighboring molecules or structures: “crowding” is volume exclusion by neighboring soluble macromolecules, but “confinement” is volume exclusion by rigid or fixed structures (Fig. 3). Said another way, theoretical models of crowding consider effects of varied concentrations of a soluble crowding species, whereas confinement models include molecular meshes, pores and channels as mimics of cytoskeletal and extracellular matrix structures (Bolis et al., 2004; Borrero and Escobedo, 2006; Eggers and Valentine, 2001; Minton, 1992, 2005; Mittal and Best, 2008; Rathore et al., 2006; Wang et al., 2009).

Because “crowding” and “confinement” are mechanistically different, it follows that these influences have different effects on protein structure. Confining structures limit molecular diffusion to a greater degree than soluble crowders; thus, theoretical models of crowding often predict faster protein diffusion rates than what is measured experimentally in the confined environment existing within a cell. (Eggers and Valentine, 2001; Mastro et al., 1984; Zimmerman and Minton, 1993). Other theoretical models suggest that confinement imparts a greater protein folding and stabilization effect compared to crowding (Cheng et al., 2018; Ping et al., 2004; Zhou, 2004, 2008); this idea is supported by experimental studies reporting that protein confinement within crosslinked alginate imparts a greater stabilization effect on denatured proteins compared to protein crowding within alginate that is not crosslinked (Long et al., 2017). Confinement is a unique phenomenon essential to physiological function that should be incorporated in in vitro models to better mimic features of the in vivo environment that influence protein structure and function.

This review should be of interest to many areas of research relating to proteins in living systems. These areas include, for example, dimerization of transmembrane proteins, enzymatic activity, assembly of supramolecular structures such as microtubules, disease-related protein aggregation in the ECM and inside the cell, protein therapeutics such as antibody drugs, and even nuclear condensates and membraneless organelles (Aljohani et al., 2018; Choi et al., 2014; Fisher et al., 2015; Rianna et al., 2017; Shin and Brangwynne, 2017; Wei et al., 2017; Wieland et al., 2007). We take inspiration from the experimental and theoretical studies described above to suggest tools and measurement techniques to better mimic properties of the in vivo environment relevant to protein structure, folding, and aggregation. As such, we ask: What tools can we adapt from the fields of tissue engineering and hydrogels to create crowded and confined environments? And what challenges are encountered when applying established biophysical techniques to analyze the structure, function, and aggregation of proteins when they are entrapped in hydrogel scaffolds?

1.5. Hydrogels

Hydrogels are ideally suited to mimic key features of tissues for applications as in vitro platforms for cell culture, tissue engineering, and drug testing. Hydrogels are derived from natural and synthetic sources and are composed of macromolecules (polymers, proteins, glycans) that interact non-covalently or are crosslinked to form an insoluble macrostructure (Edgar et al., 2016; Ruedinger et al., 2015; Stratakis, 2018; Xu et al., 2012). When applied toward an in vitro model or an in vivo implant, hydrogels are associated with proteins, and most often, proteins are located within the molecular structure of the hydrogel. Because the molecular properties of hydrogels can be controlled via chemical and physical tools (Fig. 4), these structures are an ideal platform to probe and tune protein crowding and confinement as vehicles for protein stabilization, formulation and delivery in biopharmaceutical applications, as well as for applications in models of disease (e.g., to determine molecular mechanisms of amyloid aggregation) and models of physiological functions (e.g., enzymatic activity and nuclear condensates containing transcriptional machinery) (Gregoritza and Brandl, 2015; Jin et al., 2010; Khan et al., 2013; Leach et al., 2003; Lutolf et al., 2001; Place et al., 2009; Wu et al., 2013).

When modeling disease, selection of the most appropriate hydrogel model depends on the particular problem being investigated and should consider the heterogeneous compositions and structures that exist within the particular tissue or organ of interest (Franz et al., 2010; Marieb and Hoehn, 2018; Tong et al., 2015). Further, the structure and function of proteins that exist intracellularly and extracellularly drive cell behavior; therefore, it is essential to consider how the milieu within a particular in vitro model influences protein crowding and confinement.

Both static and dynamic measurements of hydrogel microstructure are informative of the crowding and confinement features of a system. Static measurements give general insight into architectures such as mesh size and microdomains (e.g., fibers, network heterogeneity). Dynamic measurements may also provide this information, but because they can also provide insight into solute transport within the hydrogel, dynamic measurements are vital for studying protein crowding and confinement.

The architecture of the static hydrogel scaffold can be imaged using atomic force microscopy (AFM), scanning electron microscopy (SEM) and transmission electron microscopy (TEM). AFM is applicable for hydrated or dry samples and may be modified with a receptor functionalized tip to combine topography mapping and molecular recognition mapping (Abuellaf et al., 2015; Creasey et al., 2010). However, For SEM and TEM, the sample must be dehydrated; therefore, results may not be representative of the hydrated state of the hydrogel (Narayanan et al., 2006). Static light scattering (SLS), small angle X-ray scattering (SAXS), and neutron scattering (SANS) are more suitable tools to probe the static properties of a hydrogel in its hydrated state (Bohidar, 2001; Khan et al., 2013). However, hydrogel microstructures
are typically not static—molecular interactions and random structural fluctuations will dynamically influence the microenvironment of an embedded protein.

Dynamics of hydrogel microstructures can be characterized by a) analysis of probe diffusion via fluorescence recovery after photobleaching (FRAP) and dynamic light scattering (DLS), b) estimation of mesh size from swelling data and Flory-Rehner-Huggins theory, and c) examination of mechanical properties by rheology and dielectric relaxation (Aggelis et al., 1997, 2001; Barretta et al., 2000; Bohidar, 2001; Burne and Sellen, 1994; Canal and Peppas, 1989; Johnson et al., 1995; Pluen et al., 1999; Pusey and Van Megen, 1989; Simpson et al., 2020). Unfortunately, these techniques are most often incapable of distinguishing between features of the hydrogel and features of the protein because signatures related to the hydrogel typically overshadows the small signal generated by features of individual proteins. Therefore, this review addresses the following related questions: How can we adapt techniques to distinguish structures of a protein from structures of its complex local environment? What new tools could be applied to characterize these structures individually? What insights and opportunities are unveiled through the study of proteins in crowded and confined environments?

2. Current challenges and opportunities

When adapting analytical techniques to study crowded or confined proteins in hydrogel microstructural environments, the major challenge faced is the potential for the hydrogel itself to interfere with measurements of the embedded protein. It may be tempting to analyze protein extracts isolated from supernatants or digests of the hydrogel. However, considering that conditions of crowding and confinement stabilize proteins into low entropy states, analysis of protein structure and within a well-characterized hydrogel environment. The biophysical techniques available that are not dependent upon large background signal (Bolis et al., 2004; Cheng et al., 2015; Dobson, 2001; Edgar et al., 2016; Staskus and Johnson Jr., 1988). Achiral polymers (e.g. polyethylene glycol (PEG), polybutylene terephthalate (PBT), polycaprolactone) minimally interfere with entrapped protein CD signal (Li et al., 2015a; Munishkina et al., 2004). When using CD to

![Fig. 3. Schematic of crowding vs. confinement.](image)

Crowding consists of soluble macromolecules that exclude volume from other macromolecules in the solution. Confinement consists of structures that are crosslinked, forming a 3D scaffold.
study protein structure, it is crucial to select appropriate achiral biomaterials for the hydrogel.

Fourier Transform Infrared Spectroscopy (FTIR) detects the vibrational spectra of a protein to determine secondary structure. An infrared light beam passes through the sample and detects the attenuation of the beam (Ami et al., 2016). Beam attenuation occurs when light interacts with vibrational transitions in covalent bonds, thereby informing protein secondary structure (Derenne and Goormaghtigh, 2017). FTIR can differentiate between parallel and antiparallel β-sheets as well as native β-sheets vs. amyloid (Ami et al., 2006; Zolls et al., 2012). However, this technique is limited to a sample depth of 0.5–2 μm (Mirabella, 1993).

2D NMR is a powerful technique that analyzes the molecular rotational correlation time to identify protein tertiary structure (Ahmed et al., 2010; Luhrs et al., 2005). However, large molecular weight (> 25 kDa) proteins and polymers tumble slowly causing increased linewidths and spectral overlap from the large number of unique signals (Foster et al., 2007). Protein assemblies and aggregates are not suitable for 2D NMR analysis. Although NMR has been used to characterize amyloid proteins, aggregates are intentionally broken down into monomers to analyze protein structure of a single protein (Luhrs et al., 2005).

High-resolution techniques such as AFM and electron microscopy can be used to study protein tertiary structure, but only within a very thin (< 150 nm) section or a surface plane of a hydrogel. AFM uses a transducer to measure deflections of a probe that is tapped or dragged across a surface, allowing visualization of the molecular topography at resolutions as small as a fraction of a nanometer. Additionally, AFM may be conducted on surfaces of hydrated samples (Harke et al., 2012), but cannot probe structure within a 3D material unless thin sections are prepared (Graham et al., 2010).

Electron microscopy techniques take place under vacuum which requires samples to be dry. SEM captures 10–100 nm of depth into the sample with a resolution of 2 nm but requires sample dehydration and a coating of a conductive material. TEM can be used to examine protein structure within a 3D hydrogel but the sample must be resin-embedded, sectioned, and the protein is stained with a heavy metal. Electrons are passed through an ultra-thin (~100 nm) section of hydrogel to capture images with subnanometer resolution. In TEM array tomography, serial sections are imaged then reconstructed to generate a 3D image to visualize structures (e.g. protein fibrils) that are larger than the 100-nm thick section (Micheva et al., 2010). For resin embedding, samples undergo repetitive ethanol and resin washes that compromise the hydrated state of the hydrogel (Narayanan et al., 2006; Suri and Schmidt, 2010) and potentially washes out small molecules from the hydrogel (Waters, 2010). These treatments may yield samples where smaller protein species are removed and thus measurements of protein size distribution are skewed towards large protein aggregates. As such, systems meant to mimic in vivo milieu require characterization techniques where samples are examined in their hydrated state with

Fig. 4. Schematic of hydrogel crosslinking capabilities. Chemical, biological, and physical crosslinking options provides tunability for the hydrogel physiochemical and mechanical properties. Reproduced from (Place et al., 2009) with permissions from The Royal Society of Chemistry.
negligible washing steps.

It is possible to approximate a hydrogel’s structure in its hydrated state by cryo-preservation EM (cryo-EM). In this technique, the water at a hydrogel surface is removed so that the structural detail may be obtained up to 10–100 nm deep into the structure. However, removal of water necessarily removes the protein’s solvent and likely disrupts the locations of the protein in dry samples vs when they are hydrated. Prior chemical fixation of the hydrogel may bind the soluble proteins to the hydrogel scaffold, but at this time we are unaware of cryo-SEM studies of small soluble macromolecules when encapsulated in hydrogels.

In TEM, electrons pass through the sample, thus frozen water does not need to be removed. However, when the high energy electron beam hits the ice-organic material interface, a volatile reaction occurs and organic sample is destroyed (Glaser, 1978; Talmon et al., 1986). Samples are vitrified by rapid freezing either by plunging samples into liquid ethane or freezing under high pressures (2100 bar). Vitrification is possible in samples up to 200-μm thick which then can be cryo-sectioned (Lucic et al., 2013). Cryo-sections of hydrogel can also be imaged dry on support grids. However, thin sections (100 nm) will cut most structures not parallel to the thin plane (protein assemblies and aggregates) making size information unobtainable. In addition, the nanometer scale of proteins and protein assemblies are near impossible to differentiate from the interconnected/crosslinked hydrogel scaffold.

To differentiate protein from hydrogel scaffold, the protein may be labeled with nanoparticles to give contrast to aggregate clusters. The nanoparticle labeling should not to hinder protein diffusion, protein folding or interfere with protein assembly. Nanoparticles < 5 nm in diameter cannot be resolved as individual species, but when clustered, provide sufficient contrast for imaging (Milligan et al., 1990). It is important to remember, however, that nanoparticles can interfere with protein-protein interactions, such β-sheet structures that stack into large filaments that are associated with amyloid diseases. In these cases, a small ratio of labeled species may be mixed with the unlabeled species to allow enough space between bulky groups to allow proteins to assemble or aggregate (Amaro et al., 2014; Jungbauer et al., 2009).

Scattering techniques including static light scattering (SLS), dynamic light scattering (DLS), Raman scattering, small angle neutron scattering (SANS), and small angle X-ray scattering (SAXS) are powerful techniques to analyze unlabeled protein structures in pure solutions. These scattering techniques measure fluctuations of light intensity generated when a species undergoes Brownian motion within a focused light beam. However, when the species is a protein embedded within a hydrogel, the identification of unique signals from the protein vs the hydrogel can be challenging (Ami et al., 2006; Amin et al., 2014; Liiyestrom et al., 2012; Loh et al., 2012; Stradner et al., 2006). Hydrogel scaffolds are interconnected networks that are relatively immobile but scatter intensely, while small molecules diffuse quickly and scatter minimally. Scattering techniques have been used to study the behavior of a single non-interacting protein species within a hydrogel scaffold by separating the fast and slow fluctuation times (Kloster et al., 1998; Kloster et al., 2000; Rochas and Geissler, 2014). When a heterogeneous mixture of proteins are in a hydrogel scaffold, the intense light scattering signal from the hydrogel and larger species overshadows the weaker signal from the smaller species, making it challenging to differentiate various aggregate sizes.

Fluorophore labeling has been used for decades in various fluorescence microscopy and spectroscopy techniques. The advantage of fluorescence-based tools is that they detect only the fluorescent-label species without interference from other molecules or the hydrogel itself. Live cell imaging can capture the diffusion and interactions of proteins inside a cell, in the cell membrane, and in the extracellular space in a hydrogel or tissue. Total internal reflection fluorescence microscopy (TIRFM) characterizes fluorescently-labeled proteins that are located on a glass surface with a maximum imaging depth of ~100 nm from the glass surface (Ban et al., 2004; Yagi et al., 2007) and a diffraction-limited resolution of ~240 nm. Fluorescence resonance energy transfer (FRET) uses two distinct fluorophores that transfer energy when located within 10-100 Å of each other and are close enough to associate (Morell et al., 2008). Fluorescence polarization and anisotropy (FP/FA) detect unidirectional light emission and can be used to capture molecular rotational movement and rotation (Robert et al., 2011). While some proteins can utilize intrinsic tryptophan or tyrosine fluorescence to measure protein aggregation and conformation (Amaro et al., 2011; Amaro et al., 2013; Luo et al., 2016; Rolinski et al., 2015), other species may require labeling with a fluorophore. As with nanoparticles, these labeling techniques can interfere with protein-protein interactions.

Fluorescence correlation spectroscopy (FCS) detects the individual photons of fluorescently-labeled molecules that pass through a very small confocal volume. Photon fluctuation data can be used to generate a correlation function, which can then be used to calculate molecular diffusion and size (Michelman-Ribeiro et al., 2007; Sengupta et al., 2003; Vagias et al., 2013; Vagias et al., 2017; Zusiak et al., 2010; Simpson et al., 2020). In cases where fluorophore photobleaching is a concern, raster image correlation spectroscopy (RICS) is similar to FCS but applies multiphoton excitation and rasterized scanning to limit fluorophore exposure to light (Fig. 5) (Dong et al., 2010; Shah et al., 2017).

Fluorescent or enzymatically-active reporter proteins that are sensitive to aggregation can be expressed in cells to measure protein aggregation that occurs within cells (Ami et al., 2013; Schultz et al., 2006). Also, dyes that are sensitive to protein conformation such as Congo Red (CR), Thioflavin T (ThT), and Thioflavin S (ThS) fluoresce when bound to β-sheet structures. While light microscopy is diffraction-limited to a resolution of ~300 nm, these dyes provide a rather easy-to-
use approach to follow protein aggregation kinetics and image large aggregate deposits (Ami et al., 2013; Groenning, 2010; Rodina et al., 2017; Simpson et al., 2020).

The development of new fluorescence-based single particle tracking (SPT) and super-resolution microscopy (SRM) techniques allow study of dynamic protein interactions resolved at < 100 nm. SRM uses illumination patterns or fluorescence photo-switching to overcome diffraction limitations. An excellent example of enhancing fluorescence resolution is demonstrated by Li et al., where traditional TIRFM was demonstrated to resolve features as small as 220 nm, TIRF-structured illumination microscopy (SIM) can resolve 97 nm features, and patterned activation of a reversibly photo-switchable fluorescent protein using non-linear SIM (PA NL-SIM) resolves 62 nm features (Fig. 6) (Li et al., 2015b). Combining cryo-EM with SRM can achieve resolutions of < 100 nm to yield images with high structural detail and identification of specific species by fluorescence (Hauser et al., 2017).

SRM techniques and SPT have shed light on transcription factors (TF) and their dynamic functions. TFs assemble to form the transcription preinitiation complex through a multi-step process, but little is known of the dynamics of assembly. TFs also contain intrinsically disordered regions (IDRs) that regulate transcription but underlying mechanisms are unclear. By live-cell slimfield single molecule imaging, stochastic optical reconstruction microscopy (STORM), FCS, and lattice lightsheet microscopy, IDRs have been visualized to cluster into oligomeric assemblies, forming liquid-liquid phase transitions, stabilizing DNA binding, recruiting RNA polymerase, and activating transcription (Chong et al., 2018; Leake, 2018).

Many SRM techniques require long capture times and high photon input. Movement can yield false results, therefore, specimen can be fixed to capture higher quality images (Dersch and Graumann, 2018). SPT in live cells tracks the dynamic movement using EM-CCD cameras that have high frame rates and high aperture objectives to capture freely diffusing proteins. Combining SRM techniques such as SIM or stimulated emission depletion (STED) with SPT obtains the ultimate spatiotemporal tracking resolution. MINFLUX tracking uses STED concepts with 80% reduced photon input, which minimizes photobleaching, and has achieved a resolution of 6 nm (Balzarotti et al., 2017; Dersch and Graumann, 2018). Current drawbacks of SPT and SRM include limited viewing at depth and of larger samples, many require static, fixed samples, and many require specialized fluorophore labeling.

3. Concluding remarks and perspectives

Tissue development, disease and repair involve highly complex and dynamic processes. Though the in vivo scenario is ultimately the most relevant to understand, in vitro models allow control of the composition and spatial arrangement of cells and biologically-active molecules with specificity not possible in living organisms. In these complex 3D models, cellular behaviors have been the primary subject of interest, whereas investigations of protein stability, structure, and assembly have been relatively limited. Yet, environments both inside and outside cells are composed of protein structures that drive cellular behavior and shape tissues properties (e.g., stiffness, local charge, pH, and gradient features). The ECM also contains various motifs that are involved with cell binding, gene activation, and protein-protein interactions (e.g., sequestration of growth factors, localization of mitogens and cytokines).

Simple identification of the presence or general location of a protein can be accomplished through fixation and immunolabeling. However, definition of protein structures in dynamically interacting systems requires real-time tools that can detect features within the three-dimensional hydrated structures of cells, tissues and hydrogels. As reviewed herein, many analytical tools to study protein structure in solutions can be extrapolated for use within hydrated structures. However, delineation between signals from the protein of interest vs its physical microenvironment can be very challenging. Advanced tools and approaches that apply to the contexts of crowded and confined microenvironments will undoubtedly shed new insights into protein structure, folding and aggregation processes that will advance our understanding of disease states and the stability of biopharmaceutical drugs.
Funding

This work was supported by funding from National Science Foundation (NSF) [grant number EAGER CBET-1447057] and by the National Institute of Health [grant number R01GM117159]. NSF provided support for TAG to contribute to this project through their Independent Research and Development program. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References


