Functional Stimulated Raman Imaging for Quantitative Cell Biology with Small Bioorthogonal Tags

Thesis by
Xiaotian Bi

In Partial Fulfillment of the Requirements for
the degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended May 22nd, 2025

Xiaotian Bi
ORCID: 0000-0003-4449-7488

ACKNOWLEDGEMENTS

Pursuing a Ph.D. abroad once seemed like an unattainable dream during my childhood. Yet now, six years in Pasadena have passed in the blink of an eye. The excitement and curiosity I felt upon first arriving remain vivid in my memory, and it is difficult to believe that the time has come to say farewell. These six years have been profoundly meaningful, and I have no regrets about spending this chapter of my life at Caltech. I would like to take this opportunity to express my heartfelt gratitude to all those who have supported me throughout this journey.

First and foremost, I would like to express my deepest gratitude to my Ph.D. advisor, Prof. Lu Wei. Throughout my Ph.D. journey, she has provided me with tremendous support, insightful guidance, and constant encouragement. Her mentorship has been instrumental to my academic and personal growth, and I would not have achieved what I have without her. She entrusted me with the freedom to explore my own interests while consistently providing thoughtful suggestions in both science and career development. Additionally, her example as a rigorous, innovative, hard-working, and dedicated scientist has served as a powerful inspiration, motivating me to pursue my own path with passion and determination.

Next, I would like to sincerely thank other members in my thesis committee for their time, thoughtful guidance, and invaluable support during each of our meetings, including my committee chair, Prof. Dennis A. Dougherty, and committee members, Prof. Shu-ou Shan and Prof. Viviana Gradinaru. I also want to extend my heartfelt thanks to my candidacy committee members, Prof. Jesse L. (Jack) Beauchamp and Prof. Jacqueline K. Barton, both

of whom are now esteemed emeritus professors. Their advice and insights during the early stages of my graduate studies laid a strong foundation for my growth.

Then I would like to thank the Biotechnology Leadership Pre-Doctoral Training Program (BLP) at the Donna and Benjamin M. Rosen Bioengineering Center at Caltech. Since the third year of my graduate studies, BLP has provided invaluable interdisciplinary training and generous support. I was also honored to receive the Barbara J. Burger Graduate Fellowship, and I sincerely thank Dr. Barbara J. Burger for her generous support and steadfast commitment to graduate education and scientific research.

I am also deeply grateful to all members of the Wei Lab for their support and collaboration throughout my doctoral studies. Dr. Kun Miao was my mentor when I joined the lab and played a key role in helping me quickly acquire essential skills. Amy Vo was a talented undergraduate mentee with whom I had the privilege of mentoring early in my Ph.D. Another excellent mentee, Berea Suen, joined the lab during my final year and is now continuing the research direction I mainly focused on. Mentoring them was an incredibly meaningful experience that allowed me to grow both scientifically and personally. Dr. Jiajun Du was my longest-standing labmate and provided me with enduring support in both research and life. Dr. Chenxi Qian and Adrian Colazo provided me with valuable technical assistance in resolving issues related to the SRS microscope. Additionally, I closely collaborated with Li-En Lin, Dr. Dongkwan Lee, and Dr. Haomin Wang on multiple projects and am grateful for the spirit of teamwork we shared. I would also like to thank Ziguang Yang for his contributions to chemical synthesis, Noor Naji, Joe Ambarian, and Rahuljeet (RJ) Chadha

for co-working in the cell culture room, and Dr. Ryan Leighton and Philip Kocheril for scientific discussions in the office.

I would also like to thank the various facilities and individuals at Caltech who supported my research. I am especially grateful to Prof. Henry A. Lester from the Biology Division and Charlene Kim from his lab, who taught me neuron dissection techniques. Moreover, I would like to thank Dr. Hao-Hsuan Hsieh from Prof. Shu-ou Shan's lab, who mentored me during my rotation early in my Ph.D. journey. In addition, I appreciate the support from the Beckman Institute Biological Imaging Facility (BIF), particularly Dr. Andres Collazo and Dr. Giada Spigolon; the Office of Laboratory Animal Resources (OLAR), particularly Gloria E. Miranda and Pedro Zaragoza; the Institutional Animal Care and Use Committee (IACUC), particularly Hsiaoni Amy Chang; the Optogenetics and Vector Engineering Research (CLOVER) Center, particularly Dr. Tim Shay and Dr. Zhe Qu; and the Proteome Exploration Laboratory (PEL), particularly Dr. Tsui-Fen Chou and Dr. Baiyi Quan, for their technical assistance. Without them, I would not have been able to complete my research smoothly.

I must also express my sincere gratitude to those who supported me outside of Caltech. My undergraduate advisor, Prof. Peng Zou at Peking University, was the first supervisor to guide me toward a career in science, and he has continued to offer invaluable advice throughout my Ph.D. During my undergraduate years, Prof. Adam Cohen at Harvard University and my mentor, Dr. Linlin Fan, supervised my summer research, an experience that was instrumental in shaping my decision to pursue a Ph.D. in the United States. I am also deeply grateful to Dr. Jacob Morgan, my mentor during my internship at Genentech, for his generous support

on both career development and life decisions. I truly appreciate the mentorship and thoughtful guidance I received from each of them.

Then, I want to express my heartfelt gratitude to the friends who made my time at Caltech so much more enjoyable and meaningful. I shared countless memorable moments with my roommate for five years, Heyun Li, and roommate for two years, Jolena (Jie) Zhou. Chuqi Lu generously drove us on weekend outings to explore restaurants and interesting places. I am also fortunate to work with the inspiring community of Women in Chemistry at Caltech. Additional thanks go to my peer Dr. Linqing Peng and Dr. Zhenqi Zhao, as well as to Dr. Xinhong Chen, Dr. Xiaozhe Ding, Dr. Zikun Zhu, Ruilin Qian, Ziyang Qin, and Qinyu Han, who kindly lent me experimental equipment during critical stages of my research. Outside of Pasadena, I want to especially thank Dr. Cong Liu, as co-hosting our podcast was one of the most fulfilling parts of my Ph.D. journey. I also deeply appreciate the support and encouragement I received from Jianing Zhu, Zhaojun Gao, Bailu Geng, Dr. Bai Hei, Dr. Shuzhang Liu, Dr. Ruirui Ma, Dr. Luxin Peng, Dr. Pengchong Wang, Dr. Wei Tang, Dr. Junqi Yang, Dr. Ruixuan Wang, Dr. Liyuan Zhu, Dr. Ziqi Ren, and Dr. Chang Lin. I am truly thankful for their presence in my life.

I also want to express my deepest gratitude to my partner, Ziyi Wang. We met during our undergraduate years and have sustained a long-distance relationship for six years with countless trips between Pasadena and the Bay Area. Throughout my Ph.D. journey, Ziyi has been by my side through every high and low, offering love, patience, and steadfast support. He has celebrated my accomplishments, comforted me in moments of doubt, and encouraged me through every challenge. My life is brighter and more meaningful because he is in it.

Last but most importantly, I wish to thank my parents, Jun Bi and Xuhua Gao. I feel incredibly fortunate and deeply proud to be their daughter. Their endless, unconditional love has been the foundation of everything I am. They have always given me the strength and courage to explore both myself and the world with confidence. One of my greatest regrets is not being able to see them in person for a long time during my Ph.D. years, especially throughout the pandemic. I hope to make up for that time in the years to come and cherish more moments together as a family.

Looking back, these six years have been filled with invaluable experiences, challenges, and moments of growth. I am truly thankful for every mentor, colleague, friend, and family member who accompanied me on this journey. I will carry all the things I have learned and gained at Caltech with me as I move forward into the next chapter of my life.

Dedicated to my parents
Jun Bi and Xuhua Gao

ABSTRACT

The development of imaging techniques, particularly optical imaging, has significantly advanced the field of cell biology. Compared to conventional fluorescence imaging, vibrational imaging leverages the intrinsic chemical bond information of molecules, providing multidimensional insights into molecular structures and local environments. So far, stimulated Raman scattering (SRS) microscopy has emerged as a powerful tool for quantitative measurements in biological research. It overcomes several fundamental limitations associated with fluorescence-based techniques and offers high spatial and temporal resolution along with excellent compatibility for live-cell imaging.

In this thesis, I mainly focus on utilizing small bioorthogonal vibrational tags for quantitative investigations in cell biology. These tags, such as alkyne and carbon-deuterium (C-D) bonds, are absent in endogenous biomolecules and smaller than 1 nm in size, enabling minimally perturbative labeling with high molecular specificity. Another key advantage is that the SRS signal scales linearly with bond concentrations, allowing for robust and quantitative analysis. Moreover, because their vibrational modes distinctly reside in the cellular silent region (1800-2700 cm^{-1}), they provide a high signal-to-background ratio, making them particularly well-suited for quantitative applications in complicated cellular environments.

In Chapter 2, we explored the potential of alkyne-tagged probes to serve as environment-sensitive vibrational sensors, extending their utility beyond imaging markers. We developed a generalizable sensing platform based on hydrogen-deuterium exchange (HDX) at terminal alkynes. This subtle isotopic substitution induces a detectable shift in the alkyne vibrational

frequency, allowing for real-time monitoring of exchange kinetics. These kinetics, in turn, provide insight into the chemical structures and local environments. We conducted a comprehensive study of the HDX process through both theoretical analysis and experimental validation. This platform was further applied to detect structural changes in DNA and to indicate pH within live cells, demonstrating the broader applicability of alkyne-tagged Raman probes for local environmental sensing in complex biological systems.

In Chapter 3, we utilized deuterated glutamine to label and study polyglutamine (polyQ) aggregates, a pathological hallmark of Huntington's disease, in neurons. Traditional imaging approaches typically rely on tagging with bulky fluorescent proteins such as EGFP, which can perturb aggregation behavior with their non-negligible sizes. Through deuterium labeling, we achieved EGFP-free imaging of polyQ aggregates, allowing for a more native characterization with live-cell compatibility. This strategy facilitated quantitative analysis of the aggregate composition and growth dynamics of polyQ aggregates in live neurons. Our results revealed significant variations in polyQ aggregates depending on cell types, subcellular localizations, aggregate sizes, and protein constructs. Notably, we identified a previously unknown type of nuclear aggregates, shedding light on the heterogeneity of polyQ pathology.

In Chapter 4, we applied deuterium-labeled small molecules to study neuronal metabolism and its dynamic interactions with neuronal activity. As neuronal firing requires high and tightly regulated metabolic input, it is critical to understand the coupling between neuronal activity and metabolism for elucidating brain function. Using deuterated glucose and fatty acids, we were able to track their downstream metabolites for metabolism studies with high

spatial and temporal resolution via SRS microscopy. In parallel, we employed optogenetic stimulation through Channelrhodopsin to achieve precise control of neuronal activity and also used neurotransmitters for longer-term modulation. By correlating different states of neuronal activation with metabolic flux changes, we gained valuable insights into how neuronal activity dynamically regulated glucose and lipid metabolism, advancing our understanding of neuroenergetic mechanisms in live neurons.

Through these studies, I demonstrate that the integration between small bioorthogonal vibrational tags and the advanced vibrational imaging technique, SRS microscopy, can provide powerful, minimally invasive, and highly quantitative tools for tackling fundamental questions in cell biology with high spatial and temporal resolution.

PUBLISHED CONTENT AND CONTRIBUTIONS

Bi, X.[†]; Miao, K.[†]; Wei, L.* Alkyne-Tagged Raman Probes for Local Environmental Sensing by Hydrogen–Deuterium Exchange. *J. Am. Chem. Soc.* **2022,** 144 (19), 8504–8514. Doi: 10.1021/jacs.2c01991.

[†]*equal contribution. X.B. participated in the conceptualization of the project, designed and performed the experiments, prepared the data, and wrote the manuscript.*

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

1.1 Stimulated Raman scattering microscopy for vibrational imaging

Bioimaging techniques have fundamentally expanded our ability to study cell biology by converting invisible molecular events into directly observable phenomena. The ability to obtain direct visual readouts of complex biological processes with high spatial and temporal resolution and molecular specificity provides an irreplaceable platform for advancing quantitative cell biology.[1-2] Among various bioimaging modalities, optical microscopy, especially the fluorescence-based techniques, plays an indispensable role due to its exceptional sensitivity and molecular precision, and compatibility with live-cell imaging.[3] It enables dynamic tracking with high spatial resolution, especially when coupled with advanced super-resolution approaches beyond the diffraction limit. Despite their tremendous utility, fluorescence imaging also suffers from inherent limitations.[4-5] The relatively large size of fluorescent labels can perturb the native behavior of biomolecules, necessitating careful validation and control experiments for analysis. Moreover, fluorescent signals are easily influenced by photophysical processes such as blinking, photobleaching, and environmental quenching, which complicate quantitative measurements, particularly for long-term studies or in crowded subcellular environments.[3]

An alternative imaging strategy has emerged from vibrational spectroscopy, which derives signal contrast based on the intrinsic chemical bond properties. Instead of using a bulky

fluorophore as the reporter for fluorescent imaging, vibrational spectroscopy only needs a tiny chemical bond for detection, allowing for minimal perturbation for labeling. Additionally, vibrational spectroscopy, mainly including infrared (IR) absorption and Raman scattering, enables the direct label-free imaging of endogenous biomolecules. Because there are no electronic excited states involved in the process, it is inherently free from common fluorescence-related issues such as photobleaching and environmental quenching, suitable for long-term imaging with excellent photostability. While IR absorption typically has significantly higher cross sections compared to Raman scattering, Raman-based measurements are generally preferred for biological applications, especially for live cell systems.[6] The main reason is that water exhibits very strong IR absorption, but weak Raman scattering based on their different selection rules. As a result, IR-based measurements often suffer from substantial background interference in aqueous environments. In addition, Raman scattering provides intrinsically better spatial resolution determined by the diffraction limit, as the IR beam has much longer wavelengths than those employed in Raman scattering.

However, conventional spontaneous Raman scattering faces major limitations due to the inherently weak signal, approximately ~10^{-10} weaker than fluorescence, with cross sections typically ranging between 10^{-28} and 10^{-30} cm^2.[7] To overcome the fundamental sensitivity limitations, coherent Raman scattering (CRS) was invented to significantly boost the weak signal.[8] CRS, prominently including the stimulated Raman scattering (SRS) and the coherent anti-Stokes Raman scattering (CARS), is a third-order nonlinear optical process that involves two spatially and temporally overlapped pulsed laser beams, named as pump and Stokes beams.[9] Unlike CARS, SRS offers detection free from non-resonant background with strict

linear molecular concentration dependence, enabling quantitative image analysis. In SRS microscopy, when the frequency difference between the pump and the Stokes lasers matches the vibrational frequency of chemical bonds of interest, vibrational activation rates can be drastically amplified through a quantum stimulation effect analogous to stimulated emission.[10] With this amplification of a factor up to 10^8, SRS microscopy has become the most suitable far-field Raman imaging modality for quantitative cell biology, offering image quality comparable to fluorescence microscopy with label-free, chemically specific contrast.[8, 11]

Enabled by advances in physical instrumentation, the use of the high-repetition-rate picosecond laser source and high-frequency (MHz-range) modulation transfer detection scheme in SRS microscopy has proven fast imaging speeds up to video rate, sub-millisecond pixel dwell times, diffraction-limited spatial resolution, and shot-noise-limited sensitivity. In addition, since pump and Stokes beams are typically in the near-infrared region, the relatively low phototoxicity makes it highly suitable for live cells requiring gentle imaging conditions. Another key advantage of SRS microscopy is its strictly linear relationship between signal intensity and the concentration of chemical bonds, which makes it an ideal imaging modality for quantitative measurements.[12] Furthermore, SRS is inherently compatible with other microscopic imaging techniques, enabling multimodal optical imaging within the same field of view in biological samples. With the powerful capability for quantifying biomolecular information, SRS microscopy has been generally established as a noninvasive and high-resolution imaging platform for complex biological systems.[13]

1.2 Label-free imaging for biological systems

SRS microscopy was originally developed as a label-free imaging modality through directly visualizing the intrinsic vibrations of endogenous biomolecules.[10] By eliminating the need for extrinsic labeling, SRS enables a more authentic observation of biological processes in their native states. The characteristic vibrational modes encoded by specific chemical structures can serve as unique molecular fingerprints, allowing for bond-selective identification in complicated biological environments.[14-15] Based on that, label-free SRS microscopy has successfully imaged a wide range of intrinsic biomolecules, including total proteins and lipids, DNA, RNA, and metabolites, by targeting endogenous chemical bonds in the cellular fingerprint region (500-1700 cm^{-1}) and high wavenumber carbon-hydrogen (C-H) stretching region (2800-3100 cm^{-1}).[8, 16-18] The detection sensitivity of label-free SRS microscopy is typically in the millimolar (mM) range for most chemical bonds commonly used in biological systems.[8] While it is sufficient for detecting relatively abundant endogenous biomolecules, it presents challenges to resolve low-abundance ones and to distinguish molecular species with overlapping vibrational spectra.[19-20] Emerging data processing strategies, particularly those based on machine learning and deep learning algorithms,[21] hold a significant promise for enhancing imaging quality and sensitivity, potentially enabling the detection and unmixing of weaker signals of complex molecular mixtures.

In this thesis, I mainly utilized the C–H stretching vibrations (~2800-3100 cm^{-1}) as the target spectral window to obtain general molecular information from cells. Specifically, SRS

imaging at the CH_2 channel (2850 cm^{-1}) allows label-free visualization of lipid-rich components, while SRS imaging at the CH_3 channel (2940 cm^{-1}) highlights protein-rich regions. In Chapter 3, I used SRS signals at the CH_3 channel for the quantification of total protein concentrations within the polyglutamine (polyQ) aggregates, providing insights into protein compositions of pathological inclusions in neurons. In Chapter 4, I leveraged SRS signals at the CH_2 channel to assess total lipid concentrations in the neuronal metabolism study, revealing lipid dynamics under different neuronal activity. Overall, label-free SRS imaging provides a powerful combination of chemical specificity, non-invasiveness, and quantitative capability, enabling the direct visualization of endogenous molecules without the limitations associated with conventional labeling strategies.

1.3 Quantitative analysis with small biorthogonal Raman probes

To overcome the limited molecular specificity of label-free SRS imaging, the development of Raman probes has become essential. A wide range of vibrational probes, including small vibrational tags (<1 nm), Raman dyes (a few nanometers), and Raman-active nanomaterials (>10 nm), have been designed for SRS imaging to advance its further applications.[8, 22-23] In this thesis, I mainly focus on small bioorthogonal Raman probes, which offer molecule-specific labeling, minimal perturbations, and compatibility with quantitative analysis due to the strict linear concentration dependence and non-quenching nature of Raman signals.[13] This bioorthogonal SRS imaging with small Raman tags is particularly suitable for live-cell investigations of small molecules, such as metabolites, whose physiological functions are easily perturbed by conventional labeling strategies for fluorescent measurements.[24]

Three commonly used small biorthogonal Raman probes are alkynes (carbon-carbon triple bond, C≡C), nitriles (carbon-nitrogen triple bond, C≡N), and carbon-deuterium (C-D) bonds. These functional groups are absent in endogenous biomolecules to enable the labeling specificity without background interference from other cellular components. They all exhibit distinct Raman peaks (~2100-2300 cm^{-1}) in the vibrational cell-silent spectral region (~1800-2700 cm^{-1}), where there are almost no intrinsic Raman signals from cells, allowing for detection with high signal-to-background ratios and high selectivity.[24-25]

In Chapter 2, I used terminal alkynes, which have the highest Raman signals among small bioorthogonal vibrational tags.[26] The alkyne-tagged small molecular precursors are well-established in click chemistry for cellular macromolecular incorporation. SRS microscopy allows for direct detection with sensitivities in the hundreds of micromolar (μM) range, eliminating the need for subsequent click reactions with azide-functionalized dyes.[8]

In Chapter 3 and Chapter 4, I targeted deuterium-labeled C-D bonds. As a stable isotope of hydrogen, deuterium retains nearly identical physicochemical properties to hydrogen. While the Raman cross section of C-D is smaller than that of triple bonds, the multiple labeling sites help with signal amplification, as one deuterated macromolecule can contain up to thousands of C-D bonds, compensating for the lower signal for the single bond. In Chapter 3, I utilized deuterated glutamine to label polyQ aggregates in neurons to avoid the structural interference introduced by bulky EGFP tagging and enable quantitative studies close to the native conditions. In Chapter 4, I applied deuterated glucose and deuterated fatty acids to label and investigate the neuronal metabolism in response to neuronal activity in live neurons.

1.4 Vibrational-based sensing applications

In addition to serving as a molecular marker for bioimaging, Raman probes, especially the triple-bond tags, hold huge potential for sensing the local environments in biological systems. Different from traditional fluorescent imaging, vibrational spectroscopy provides an additional dimension for measurements through peak frequency, which is highly sensitive to chemical structures and the local environments. Over the past decade, substantial progress has been made in the development of Raman-based sensing strategies, particularly those relying on chemical reactions. For example, chemical reactions between the probes with alkynes and various substrates have been designed to sense the presence of small molecules,[27] ions,[28-30] or enzymes[31-32] based on the Raman frequency shifts due to chemical structure change. Other than Raman peak shift, peak enhancement, peak generation, and peak switching can also work as a readout to reflect on the sensing targets based on different types of reactions.[33] Another strategy relies on the vibrational Stark effect[34-35], as noncovalent interactions between probes and their adjacent environments result in the Raman peak shift. This effect reflects changes in local electric fields and intermolecular interactions, allowing for the use of existing alkyne-tagged probes to sense protein motions[36] and conformational changes[37] in live cells. However, current strategies have their own limitations. A general strategy that makes full use of existing Raman probes towards quantitative environmental sensing with high spatiotemporal resolution would be highly desirable when combined with SRS microscopy.

In Chapter 2, I present a novel platform for local environmental sensing based on hydrogen–deuterium exchange (HDX) on terminal alkynes. This method enables chemical structure

and cellular environmental sensing by leveraging the dramatic Raman peak shift of over 130 cm^{-1}, due to the significant mass difference when hydrogen is replaced by deuterium in terminal alkynes through HDX kinetics analysis.

1.5 Metabolism studies with deuterium labeling

SRS imaging offers significant advantages for metabolic profiling, as it enables the noninvasive probing of metabolic activities with high spatial and temporal resolution. While the triple bond tagging is generally considered non-perturbative in most bioimaging applications, it might not meet the high requirements of metabolism studies in some cases. For example, the alkyne-tagged glucose analogue, 3-OPG, is frequently used to quantify the glucose uptake in live cells.[38-39] However, the presence of the alkyne group blocks phosphorylation, thereby inhibiting entry into the glucose metabolic pathway for glycolysis. This highlights the importance of selecting even smaller and less perturbative labeling strategies for accurate metabolic studies while maintaining the native metabolic fluxes.

Stable isotope-labeled metabolites, particularly deuterium labeling, are superior tools for tracing metabolites into downstream biomolecules in live cells with lower toxicity and higher fidelity, as deuterium exhibits nearly identical chemical properties to hydrogen.[40-41] As a result, deuterated glucose (i.e., d_7-glucose) can participate naturally in cellular enzymatic reactions and be incorporated into diverse downstream products, such as DNA/RNA, proteins, lipids, and glycogen.[39, 42] Deuterium labeling has become a versatile strategy for probing comprehensive metabolic processes for quantifying metabolite uptake, biosynthesis, and turnover, when coupled with SRS microscopy for high-resolution mapping of their

distribution across cells.[43] Moreover, the deuterated downstream biomolecular products exhibit distinct C-D vibrational features due to various chemical environments surrounding the incorporated deuterium atoms, which can be spectrally separable in hyperspectral SRS imaging with a linear unmixing algorithm.[44] The species-resolved analysis of newly synthesized macromolecules based on their unique spectral fingerprints allows for quantitative dynamic monitoring and directly accessing the allocation of metabolites into different metabolic pathways.

In Chapter 4, I leveraged these advantages of deuterated labeling to investigate neuronal metabolism coupled with neuronal activity. I mainly used deuterated glucose and deuterated fatty acids to investigate the uptake dynamics and downstream metabolic allocation associated with glucose and lipid metabolism. These deuterated small molecular precursors offer minimal perturbations and exhibit distinct vibrational signatures in the cell-silent region, particularly suitable for quantitative, noninvasive analysis with SRS microscopy.

1.6 Conclusion

As a powerful nonlinear optical imaging technique with enhanced sensitivity, SRS microscopy has significantly advanced the development of quantitative cell biology in recent years. Owing to the inherent chemical specificity of vibrational spectroscopy, SRS microscopy provides rich molecular information by targeting the corresponding vibrations of chemical bonds. Importantly, the strict linear relationship between SRS signals and molecular concentrations enables quantitative imaging. Its microscopic features provide high spatial and temporal resolution, and live-cell compatibility, making SRS microscopy

extremely suitable for quantitative and dynamic investigations of cellular processes under near-physiological conditions. Furthermore, the ability to perform label-free imaging, along with the development of small bioorthogonal vibrational probes, has greatly expanded the toolkits for studying biomolecules in complex environments with minimal perturbations.

This thesis primarily highlights the versatile applications of small bioorthogonal vibrational tags in biological systems, organized into three thematic sections including works on sensing, quantitative imaging, and metabolic studies. In Chapter 2, I used terminal alkynes to detect local environment changes. In Chapter 3, I used deuterated glutamine for precise measurements of polyQ aggregates in neurons. In Chapter 4, I used deuterated small molecule precursors to track neuronal metabolism coupled with neuronal activity. Together, these studies demonstrate the unique potential of small bioorthogonal vibrational probes for investigations in quantitative cell biology with live-cell compatibility, minimal perturbations, high spatiotemporal resolution, and molecular specificity. Looking ahead, we anticipate that continued advances in SRS imaging and analytical methodologies will play a pivotal role in driving future discoveries in cell biology.

1.7 Reference

1.Cole, R., Live-cell imaging. *Cell Adh Migr* **2014,** *8* (5), 452-9.
2.Galas, L.; Gallavardin, T.; Bénard, M.; Lehner, A.; Schapman, D.; Lebon, A.; Komuro, H.; Lerouge, P.; Leleu, S.; Franck, X., "Probe, Sample, and Instrument (PSI)": The Hat-Trick for Fluorescence Live Cell Imaging. *Chemosensors* **2018,** *6* (3).
3.Ntziachristos, V., Fluorescence molecular imaging. *Annu Rev Biomed Eng* **2006,** *8*, 1-33.
4.Ettinger, A.; Wittmann, T., Fluorescence live cell imaging. *Methods Cell Biol* **2014,** *123*, 77-94.
5.Specht, E. A.; Braselmann, E.; Palmer, A. E., A Critical and Comparative Review of Fluorescent Tools for Live-Cell Imaging. *Annu Rev Physiol* **2017,** *79*, 93-117.

6.Wang, H.; Lee, D.; Wei, L., Toward the Next Frontiers of Vibrational Bioimaging. *Chem Biomed Imaging* **2023,** *1* (1), 3-17.
7.Vlasov, A. V.; Maliar, N. L.; Bazhenov, S. V.; Nikelshparg, E. I.; Brazhe, N. A.; Vlasova, A. D.; Osipov, S. D.; Sudarev, V. V.; Ryzhykau, Y. L.; Bogorodskiy, A. O.; Zinovev, E. V.; Rogachev, A. V.; Manukhov, I. V.; Borshchevskiy, V. I.; Kuklin, A. I.; Pokorný, J.; Sosnovtseva, O.; Maksimov, G. V.; Gordeliy, V. I., Raman Scattering: From Structural Biology to Medical Applications. *Crystals* **2020,** *10* (1), 38.
8.Hu, F.; Shi, L.; Min, W., Biological imaging of chemical bonds by stimulated Raman scattering microscopy. *Nat. Methods* **2019,** *16* (9), 830-842.
9.Min, W.; Freudiger, C. W.; Lu, S.; Xie, X. S., Coherent nonlinear optical imaging: beyond fluorescence microscopy. *Annu Rev Phys Chem* **2011,** *62*, 507-30.
10.Freudiger, C. W.; Min, W.; Saar, B. G.; Lu, S.; Holtom, G. R.; He, C.; Tsai, J. C.; Kang, J. X.; Xie, X. S., Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy. *Science* **2008,** *322* (5909), 1857-61.
11.Fu, D.; Lu, F. K.; Zhang, X.; Freudiger, C.; Pernik, D. R.; Holtom, G.; Xie, X. S., Quantitative chemical imaging with multiplex stimulated Raman scattering microscopy. *J Am Chem Soc* **2012,** *134* (8), 3623-6.
12.Shen, Y.; Xu, F.; Wei, L.; Hu, F.; Min, W., Live-cell quantitative imaging of proteome degradation by stimulated Raman scattering. *Angew Chem Int Ed Engl* **2014,** *53* (22), 5596-9.
13.Du, J.; Wang, H.; Wei, L., Bringing vibrational imaging to chemical biology with molecular probes. *ACS Chem Biol* **2022,** *17* (7), 1621-1637.
14.Lu, F. K.; Basu, S.; Igras, V.; Hoang, M. P.; Ji, M.; Fu, D.; Holtom, G. R.; Neel, V. A.; Freudiger, C. W.; Fisher, D. E.; Xie, X. S., Label-free DNA imaging in vivo with stimulated Raman scattering microscopy. *Proc Natl Acad Sci U S A* **2015,** *112* (37), 11624-9.
15.Cheng, J. X.; Xie, X. S., Vibrational spectroscopic imaging of living systems: An emerging platform for biology and medicine. *Science* **2015,** *350* (6264), aaa8870.
16.Wang, P.; Liu, B.; Zhang, D.; Belew, M. Y.; Tissenbaum, H. A.; Cheng, J. X., Imaging lipid metabolism in live Caenorhabditis elegans using fingerprint vibrations. *Angew Chem Int Ed Engl* **2014,** *53* (44), 11787-92.
17.Hsu, C. C.; Xu, J.; Brinkhof, B.; Wang, H.; Cui, Z.; Huang, W. E.; Ye, H., A single-cell Raman-based platform to identify developmental stages of human pluripotent stem cell-derived neurons. *Proc Natl Acad Sci U S A* **2020,** *117* (31), 18412-18423.
18.Zhang, X.; Roeffaers, M. B.; Basu, S.; Daniele, J. R.; Fu, D.; Freudiger, C. W.; Holtom, G. R.; Xie, X. S., Label-free live-cell imaging of nucleic acids using stimulated Raman scattering microscopy. *Chemphyschem* **2012,** *13* (4), 1054-9.
19.Ozeki, Y.; Umemura, W.; Otsuka, Y.; Satoh, S.; Hashimoto, H.; Sumimura, K.; Nishizawa, N.; Fukui, K.; Itoh, K., High-speed molecular spectral imaging of tissue with stimulated Raman scattering. *Nature Photonics* **2012,** *6* (12), 845-851.
20.Greig, J. C.; Tipping, W. J.; Graham, D.; Faulds, K.; Gould, G. W., New insights into lipid and fatty acid metabolism from Raman spectroscopy. *Analyst* **2024,** *149* (19), 4789-4810.
21.Ho, C. S.; Jean, N.; Hogan, C. A.; Blackmon, L.; Jeffrey, S. S.; Holodniy, M.; Banaei, N.; Saleh, A. A. E.; Ermon, S.; Dionne, J., Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nat Commun* **2019,** *10* (1), 4927.

22.Lu, F. K.; Ji, M.; Fu, D.; Ni, X.; Freudiger, C. W.; Holtom, G.; Xie, X. S., Multicolor stimulated Raman scattering (SRS) microscopy. *Mol Phys* **2012,** *110* (15-16), 1927-1932.
23.Qian, N.; Gao, X.; Lang, X.; Deng, H.; Bratu, T. M.; Chen, Q.; Stapleton, P.; Yan, B.; Min, W., Rapid single-particle chemical imaging of nanoplastics by SRS microscopy. *Proc Natl Acad Sci U S A* **2024,** *121* (3), e2300582121.
24.Wei, L.; Hu, F.; Chen, Z.; Shen, Y.; Zhang, L.; Min, W., Live-Cell Bioorthogonal Chemical Imaging: Stimulated Raman Scattering Microscopy of Vibrational Probes. *Acc Chem Res* **2016,** *49* (8), 1494-502.
25.Yamakoshi, H.; Dodo, K.; Palonpon, A.; Ando, J.; Fujita, K.; Kawata, S.; Sodeoka, M., Alkyne-tag Raman imaging for visualization of mobile small molecules in live cells. *J. Am. Chem. Soc.* **2012,** *134* (51), 20681-20689.
26.Wei, L.; Hu, F.; Shen, Y.; Chen, Z.; Yu, Y.; Lin, C.-C.; Wang, M. C.; Min, W., Live-cell imaging of alkyne-tagged small biomolecules by stimulated Raman scattering. *Nat. Methods* **2014,** *11* (4), 410-412.
27.Zeng, C.; Hu, F.; Long, R.; Min, W., A ratiometric Raman probe for live-cell imaging of hydrogen sulfide in mitochondria by stimulated Raman scattering. *Analyst* **2018,** *143* (20), 4844-4848.
28.Tipping, W. J.; Wilson, L. T.; Blaseio, S. K.; Tomkinson, N. C. O.; Faulds, K.; Graham, D., Ratiometric sensing of fluoride ions using Raman spectroscopy. *Chem. Commun.* **2020,** *56* (92), 14463-14466.
29.Wilson, L. T.; Tipping, W. J.; Jamieson, L. E.; Wetherill, C.; Henley, Z.; Faulds, K.; Graham, D.; Mackay, S. P.; Tomkinson, N. C. O., A new class of ratiometric small molecule intracellular pH sensors for Raman microscopy. *Analyst* **2020,** *145* (15), 5289-5298.
30.Talley, C. E.; Jusinski, L.; Hollars, C. W.; Lane, S. M.; Huser, T., Intracellular pH sensors based on surface-enhanced Raman scattering. *Anal. Chem.* **2004,** *76* (23), 7064-7068.
31.Qin, X.; Lyu, M.; Si, Y.; Yang, J.; Wu, Z.; Li, J., Alkyne-based surface-enhanced Raman scattering nanoprobe for ratiometric imaging analysis of caspase 3 in live cells and tissues. *Anal. Chim. Acta* **2018,** *1043*, 115-122.
32.Si, Y.; Bai, Y.; Qin, X.; Li, J.; Zhong, W.; Xiao, Z.; Li, J.; Yin, Y., Alkyne-DNA-Functionalized Alloyed Au/Ag Nanospheres for Ratiometric Surface-Enhanced Raman Scattering Imaging Assay of Endonuclease Activity in Live Cells. *Anal. Chem.* **2018,** *90* (6), 3898-3905.
33.Dong, T.; Yu, P.; Zhao, J.; Wang, J., Probing the local structure and dynamics of nucleotides using vibrationally enhanced alkynyl stretching. *Phys Chem Chem Phys* **2022,** *24* (48), 29988-29998.
34.Fried, S. D.; Boxer, S. G., Measuring electric fields and noncovalent interactions using the vibrational stark effect. *Acc. Chem. Res.* **2015,** *48* (4), 998-1006.
35.Marr, J. M.; Schultz, Z. D., Imaging Electric Fields in SERS and TERS Using the Vibrational Stark Effect. *J. Phys. Chem. Lett.* **2013,** *4* (19), 3268-3272.
36.Epstein, S. C.; Huff, A. R.; Winesett, E. S.; Londergan, C. H.; Charkoudian, L. K., Tracking carrier protein motions with Raman spectroscopy. *Nat. Commun.* **2019,** *10* (2227), 1-6.
37.Flynn, J. D.; Gimmen, M. Y.; Dean, D. N.; Lacy, S. M.; Lee, J. C., Terminal Alkynes as Raman Probes of alpha-Synuclein in Solution and in Cells. *Chembiochem* **2020,** *21* (11), 1582-1586.

38.Jamieson, L. E.; Greaves, J.; McLellan, J. A.; Munro, K. R.; Tomkinson, N. C. O.; Chamberlain, L. H.; Faulds, K.; Graham, D., Tracking intracellular uptake and localisation of alkyne tagged fatty acids using Raman spectroscopy. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2018,** *197*, 30-36.
39.Zhang, L.; Shi, L.; Shen, Y.; Miao, Y.; Wei, M.; Qian, N.; Liu, Y.; Min, W., Spectral tracing of deuterium for imaging glucose metabolism. *Nat Biomed Eng* **2019,** *3* (5), 402-413.
40.Xu, J.; Yu, T.; Zois, C. E.; Cheng, J. X.; Tang, Y.; Harris, A. L.; Huang, W. E., Unveiling Cancer Metabolism through Spontaneous and Coherent Raman Spectroscopy and Stable Isotope Probing. *Cancers (Basel)* **2021,** *13* (7).
41.Du, J.; Su, Y.; Qian, C.; Yuan, D.; Miao, K.; Lee, D.; Ng, A. H. C.; Wijker, R. S.; Ribas, A.; Levine, R. D.; Heath, J. R.; Wei, L., Raman-guided subcellular pharmaco-metabolomics for metastatic melanoma cells. *Nat Commun* **2020,** *11* (1), 4830.
42.Lee, D.; Du, J.; Yu, R.; Su, Y.; Heath, J. R.; Wei, L., Visualizing Subcellular Enrichment of Glycogen in Live Cancer Cells by Stimulated Raman Scattering. *Anal Chem* **2020,** *92* (19), 13182-13191.
43.Shi, L.; Zheng, C.; Shen, Y.; Chen, Z.; Silveira, E. S.; Zhang, L.; Wei, M.; Liu, C.; de Sena-Tomas, C.; Targoff, K.; Min, W., Optical imaging of metabolic dynamics in animals. *Nat Commun* **2018,** *9* (1), 2995.
44.Zhang, D.; Wang, P.; Slipchenko, M. N.; Ben-Amotz, D.; Weiner, A. M.; Cheng, J. X., Quantitative vibrational imaging by hyperspectral stimulated Raman scattering microscopy and multivariate curve resolution analysis. *Anal Chem* **2013,** *85* (1), 98-106.

Chapter 2

ALKYNE-TAGGED RAMAN PROBES FOR LOCAL ENVIRONMENTAL SENSING BY HYDROGEN-DEUTERIUM EXCHANGE

The content of this chapter has been published in:

Bi, X.†; Miao, K.†; Wei, L.* Alkyne-Tagged Raman Probes for Local Environmental Sensing by Hydrogen–Deuterium Exchange. *J. Am. Chem. Soc.* **2022,** 144 (19), 8504–8514. https://doi.org/10.1021/jacs.2c01991.

†These authors contributed equally.

2.1 Summary

Alkyne-tagged Raman probes have shown high promise for noninvasive and sensitive visualization of small biomolecules to understand their functional roles in live cells. However, the potential for alkynes to sense cellular environments that goes beyond imaging remains to be further explored. Here, we report a general strategy for Raman-imaging based local-environment sensing by hydrogen–deuterium exchange (HDX) of terminal alkynes (termed alkyne-HDX). We first demonstrate, in multiple Raman probes, that deuterations of the alkynyl hydrogens lead to remarkable shifts of alkyne Raman peaks for about 130 cm^{-1}, providing resolvable signals suited for imaging-based analysis with high specificity. Both

our analytical derivation and experimental characterizations subsequently establish that HDX kinetics are linearly proportional to both alkyne pKas and environmental pDs. After validating the quantitative nature of this strategy, we apply alkyne-HDX to sensing local chemical and cellular environments. We establish that alkyne-HDX exhibits a high sensitivity to various DNA structures and demonstrate the capacity to detect DNA structural changes in situ from UV-induced damage. We further show that this strategy is also applicable to resolve subtle pD variations in live cells. Altogether, our work lays a foundation for utilizing alkyne-HDX strategy to quantitatively sense the local environments for a broad spectrum of applications in complex biological systems.

2.2 Introduction

The emergence of diverse imaging probes, including organic molecules and genetically encodable proteins (e.g. GFP) has significantly advanced the development of modern optical microscopy. For vibrational spectro-microscopy, alkynes have become one of the most popular functional Raman imaging probes, especially for non-invasive visualization of small biomolecules in live cells, which presented as a fundamental limit in fluorescence microscopy due to the large size of fluorophores.[1-8] The tiny, exogenous, and biorthogonal alkyne moiety exhibits a distinct, sharp, and strong Raman peak in the cell-silent spectral window (1800-2600 cm^{-1}, Figure 2.1a, 2.2a), boosting Raman detection sensitivity and specificity and providing superb biocompatibility with minimal perturbation. In particular, when coupled with stimulated Raman scattering (SRS) microscopy, alkyne-tagged small precursors or analogs have been applied to interrogating a wide range of biomolecules in

living organisms with subcellular resolution, including nucleic acids, proteins, glycans, lipids, small metabolites, natural products, and drug candidates.[9-13]

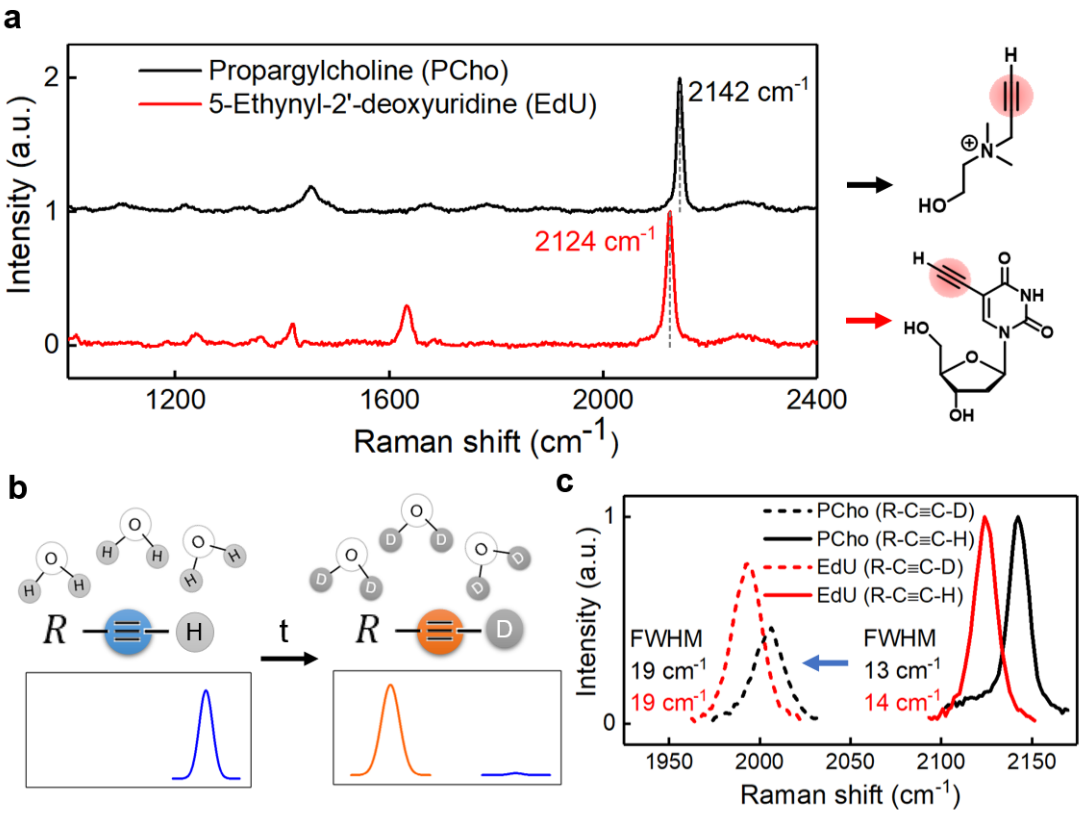

Figure 2.1. Characterization of hydrogen-deuterium exchange in terminal alkyne groups (alkyne-HDX) from live-cell compatible Raman probes by spontaneous Raman spectroscopy. (a) Chemical structures and corresponding spontaneous Raman spectra of EdU (10 mM) and PCho (40 mM) solutions. Terminal alkynes are highlighted in pink. (b) Scheme of our designed alkyne-HDX-based Raman sensing. (c) Spontaneous Raman peaks of alkynes in EdU (10 mM, red) and PCho (40 mM, black) solutions before (solid lines, normalized) and after (dashed lines) alkyne-HDX.

In addition to imaging, alkyne-tagged Raman probes have found wide utilities in sensing local chemical and cellular environments. For instance, reactions between the probes and various substrates, which lead to shifts in Raman frequency of alkynes due to chemical structure change, have been designed to sense the presence of small molecules,[14] ions,[15-17] or enzymes[18-19]. One major issue here is that a new alkyne-bearing probe needs to be devised and synthesized for each specific chemical transformation to achieve high selectivity of targeted reactions. An alternative sensing strategy is to measure the Raman peak shift produced by the vibrational Stark effect[20-21], which is caused by noncovalent interactions between an alkyne probe and its adjacent environment. This phenomenon allows exploiting diverse existing alkyne probes with proven live-cell compatibility to sense protein motions[22] and conformational changes[23]. However, the reported vibrational frequency shifts are rather limited (~ 5 cm^{-1}) and hence require prolonged data acquisition and fine spectrum fitting to determine the alkyne peak positions. The small peak shift also poses challenges for combing vibrational Stark sensing with the fast imaging-based SRS microscopy, which often has a spectral resolution of around 10-15 cm^{-1} without complex instrument customization.[12] A general strategy that further explores the potential of existing alkyne probes towards quantitative environmental sensing with high spatial and temporal resolution would hence be highly desirable.

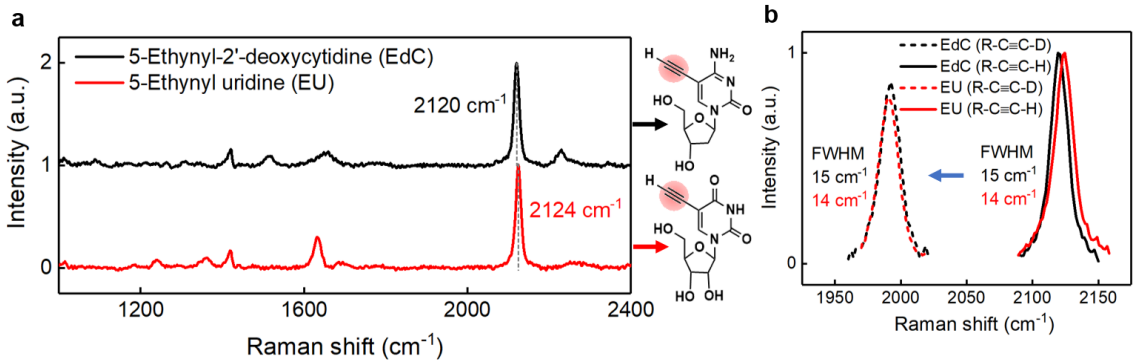

Figure 2.2. Characterizations of hydrogen-deuterium exchange on terminal alkyne groups (alkyne-HDX) by spontaneous Raman spectroscopy. (a) Chemical structures and the corresponding spontaneous-Raman spectra of EdC (10 mM) and EU (10 mM) solutions. Pink-shade highlights the terminal alkynes. (b) Spontaneous-Raman peaks of alkyne in EU (10 mM, red) and EdC (10 mM, black) solutions before (solid lines) and after (dashed lines) HDX.

Here, we report a general imaging-based platform that utilizes terminal alkyne-tagged Raman probes for local environmental sensing by alkyne hydrogen–deuterium exchange (HDX) (Figure 2.1b). Conventionally, amide HDX has been adopted by NMR[24-25] and mass spectrometry[26-29] to examine protein conformations and micro-environments, through detecting the spectroscopic or mass signatures that correspond to the deuteration at the amide backbone of proteins after exposure to deuterated solvents. The deuteration sites correlate with the accessibility of the amino acid residues at various positions of the protein secondary or tertiary structures. Similar spectroscopic characterizations were demonstrated by IR[30-34] and resonance Raman spectroscopy[35-37], yet with limited resolvability due to peak crowdedness in the fingerprint spectral region. Inspired by these studies, we here explore the

Raman spectral shift and the kinetics of deuterium exchange on alkynyl hydrogen of terminal alkynes in the background-free cell-silent region. We reasoned that HDX should occur for terminal alkynes in suitable deuterium sources (e.g. D_2O, Figure 2.1b), because the known pK_as of alkynes (~ 20-25) are much smaller than those of sp^3 CH bonds (~ 40-60) and are close to those of amides (~ 20)[38-39]. Therefore, by tracking alkyne-HDX dynamics with corresponding Raman signatures, we could investigate micro-environments of diverse alkyne-tagged probes in heterogeneous biological systems. We first theoretically establish and experimentally validate the generality of the proposed alkyne-HDX strategy. We then demonstrate unique imaging-based sensing applications of local chemical and cellular environments, including investigation of UV-induced DNA damage in cells and visualization of subtle variations in intracellular pHs. Our approach provides a novel dimension to existing sensing strategies and offers new opportunities to investigate complex biological microenvironments in their native states.

2.3 Raman peak shift upon alkyne-HDX

Using established alkyne-tagged cell-imaging probes, the thymidine analog 5-ethynyl-2'-deoxyuridine (EdU) and choline analog propargylcholine (PCho) (Figure 2.1a), we first examined the achievable alkyne Raman peak shift before and after HDX in D_2O. After EdU and PCho being diluted and equilibrated in the D_2O solutions, spontaneous Raman spectra showed a distinct red-shift over 100 cm^{-1} for the alkyne stretching frequency. 130 cm^{-1} shift was observed for EdU from 2124 cm^{-1} to 1994 cm^{-1} and 136 cm^{-1} shift was observed for PCho from 2142 cm^{-1} to 2006 cm^{-1} (Figure 2.1c). This extended peak change is a result of

the combined effect of adding more mass to the alkyne "oscillator" (similar to introducing heavier-mass ^{13}C labeling[40]) and the coupling between the C≡C and adjoining C-D bonds. Such significant shifts are consistent with previous reports of deuterated acetylide compounds[41-42]. We also performed DFT calculations on EdU after alkyne deuteration, in which the peak shift was predicted to be 149 cm^{-1}, close to our experimental data. Similar peak shift values were detected for other alkyne-tagged molecules, including the cytidine analog ethynyl-2'-deoxycytidine (EdC, 128 cm^{-1} shift) and the uridine analog 5-ethynyl-uridine (EU, 134 cm^{-1} shift) (Figure 2.2a, 2.2b). In addition to the peak shift, we observed a slight decrease in peak intensities comparing fully converted alkyne-D peaks with the original alkyne-H ones (Figure 2.1c, 2.2b, dashed lines). The ratios of the peak intensity after and before alkyne-HDX are 0.77, 0.46, 0.86, 0.79 for EdU, PCho, EdC and EU, respectively. As relative peak intensities decrease, the corresponding full-width-at-half-maximums (FWHMs) increase slightly, maintaining approximately identical integrated peak areas between alkyne peaks from before and after alkyne-HDX (integrated peak-area ratios: EdU, 0.94; PCho, 0.81; EdC, 0.96; EU, 0.93). These spectroscopic characterizations validate our first hypothesis that the alkyne-HDX strategy offers a clear spectral separation for robust peak distinction and analysis across diverse live-cell compatible alkyne probes.

2.4 Theoretical analysis and experimental validation of alkyne-HDX kinetics

We next sought to establish the theoretical basis for alkyne-HDX processes. The analytical framework of amide-HDX in proteins has been founded as acid-, base- or water-catalyzed.[43-45] We modeled the alkyne-HDX process as a base-catalyzed reaction in D_2O buffer solutions,

as the physiological pD (pD=-lg[D^+]) is higher than 7. The rate-determining step (RDS) in HDX between alkyne-tagged probes (R-C≡C-H) and catalytic base OD^- is shown in Eq. (2.1) below (more details in Materials and methods). It includes three elementary steps: a) diffusional collision, b) equilibrium redistribution of the hydrogen in the intermediate state, and c) dissociation.

$$R-C\equiv C-H+OD^- \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} R-C\equiv C-H\ldots O-D^- \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} R-C\equiv C^-\ldots H-O-D \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} R-C\equiv C^-+HOD \tag{2.1}$$

We can assume that 1) the temperature remains constant, 2) the reaction is diffusion-limited, and 3) the intermediates are in steady states.[43] The overall rate (k) for transferring a proton from R-C≡C-H to OD^- then becomes (details in Materials and Methods):

$$k = k_1 \cdot (10^{pK_a(\,donor,\ R-C\equiv C-H)-pK_a(acceptor,\ HOD)}+1)^{-1}\cdot[OD^-]\cdot[R-C\equiv C-H] \tag{2.2}$$

k_1 is the diffusion-limited collision constant, upper-bounded by 10^{10} $M^{-1}s^{-1}$.[43] As fluctuation of OD^- concentration is negligible during HDX, the RDS is considered as a pseudo-first-order reaction. Since the acceptor pK_a is much smaller than the donor pK_a, the corresponding exchange half-life ($t_{1/2}$) is reduced to Eq. (2.3):

$$\lg(t_{1/2}) = pK_a - pD + constant \tag{2.3}$$

Here pK_a designates the donor pK_a, i.e. $pK_a(R-C\equiv C-H)$. Taking the pK_a of alkynyl hydrogen as 20-25 and the pK_a of HOD as ~15,[46] we estimated the $t_{1/2}$ of alkyne-HDX in

physiological pD (7.6) to be on the order of minutes, reasonable kinetics to be captured by both spontaneous Raman spectroscopy and live-cell SRS microscopy.

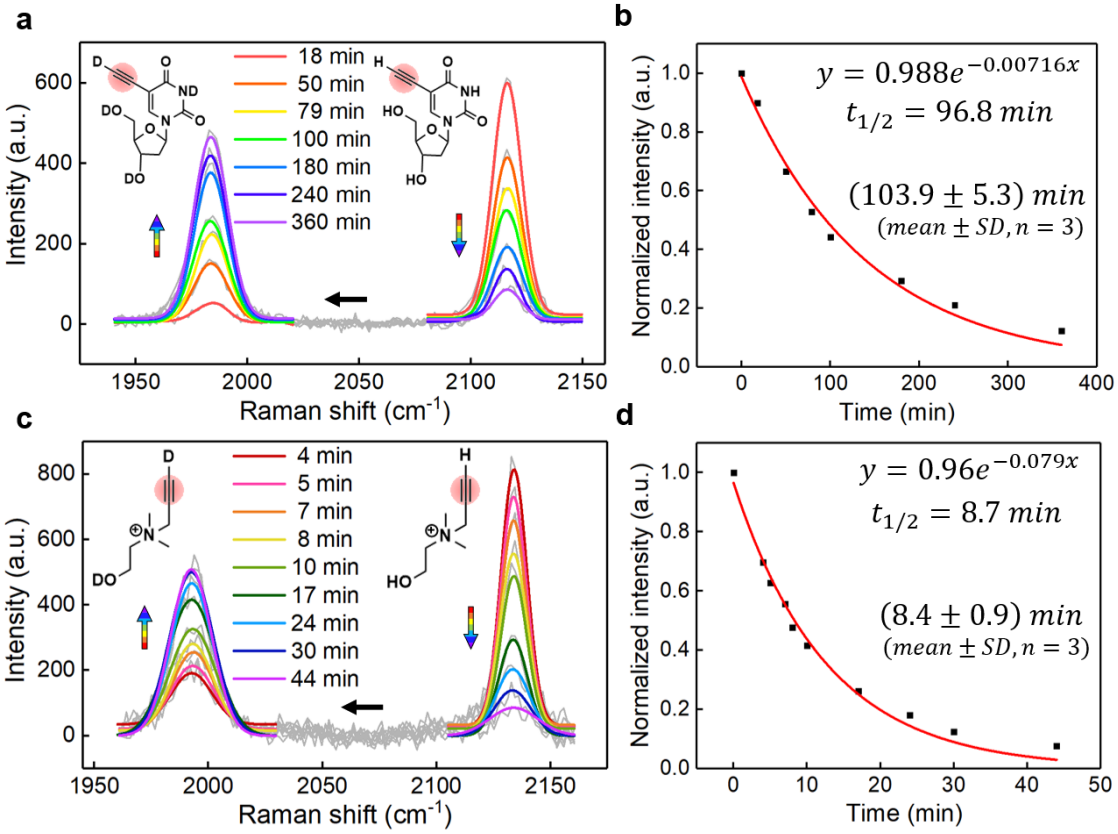

Figure 2.3. Distinct alkyne-HDX rates for EdU (top) and PCho (bottom) solutions. (a) The kinetics trace of the spontaneous Raman spectra for the decrease of alkyne-H peaks and the increase of alkyne-D peaks from EdU (10 mM) during alkyne-HDX in pD=7.6 D_2O buffer solution. (b) Exponential curve fitting of normalized alkyne peak intensities for EdU (10 mM) in (a). The average of $t_{1/2}$ over three independent measurements is also shown. (c) The kinetics trace of the spontaneous Raman spectra for the decrease of alkyne-H peaks and the increase of alkyne-D peaks from PCho (40 mM) during the alkyne-HDX process in

pD=7.6 D_2O buffer solution. (d) Exponential curve fitting of normalized alkyne peak intensities for PCho (~40 mM) in (c). The average of $t_{1/2}$ over three independent measurements is also shown.

With the above theoretical derivation, we next acquired time series of spontaneous Raman spectra for the four alkyne imaging probes (i.e. EdU, PCho, EU, and EdC), to experimentally examine the sensitivity of HDX kinetics on alkynes with minor changes of their surrounding chemical structures. By diluting each stock solution into a pD=7.6 D_2O buffer (DPBS-D_2O buffer solution), a gradual decrease of alkyne-H peak intensity (I(R-C≡C-H)) and an increase of alkyne-D peak intensity (I(R-C≡C-D)) were observed as time elapsed (Figure 2.3a, 2.3c and Figure 2.4a, 2.4c). For each peak, a Gaussian peak fitting was implemented to generate a precise peak height as the intensity. Complying with pseudo-first-order reactions, the decrease of I(R-C≡C-H) indeed follows an exponential decay. The obtained average $t_{1/2, EdU}$ is 103.9±5.3 min (Figure 2.3b, n=3), consistent with our above theoretical estimations. Interestingly, the average exchange $t_{1/2, PCho}$ is 8.4±0.9 min (Figure 2.3d, n=3), only about one-tenth of the $t_{1/2}$ for EdU, indicating that the pK_a of the terminal alkyne in PCho is smaller than that in EdU. For the other two alkyne-tagged nucleoside probes, EU and EdC[47], which have similar structures and close alkyne Raman peaks to those of EdU, while $t_{1/2, EU}$ is close to $t_{1/2, EdU}$, which is expected due to the shared uracil structure (Figure 2.4a, b), $t_{1/2, EdC}$ is about 5 times shorter than $t_{1/2, EdU}$, although EdC is only the amination product on the beta carbon to the alkyne of EdU (Figure 2.4c, d). These kinetics results demonstrate that our alkyne-HDX strategy is highly sensitive to resolve the subtle changes of local chemical structures, which is reflected in the varying alkyne pK_a values.

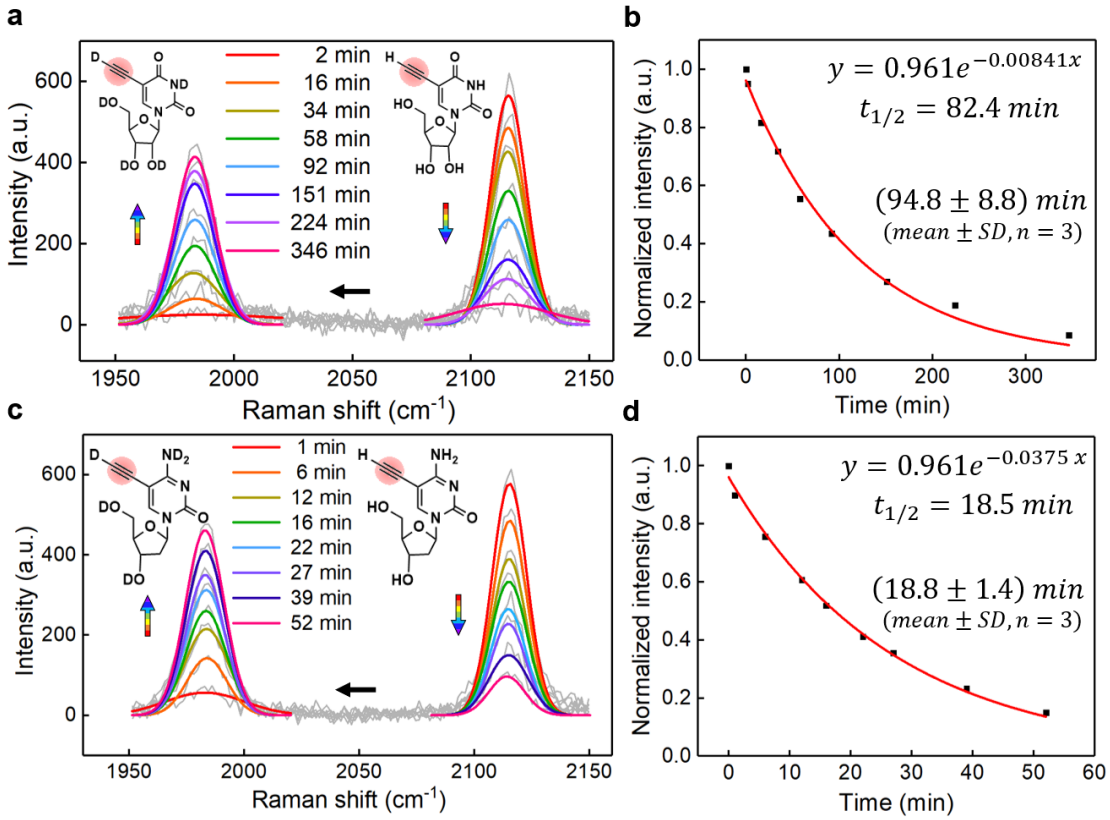

Figure 2.4. The alkyne-HDX rates for EdC and EU solutions. (a) The Kinetics trace of the spontaneous Raman spectra for the decrease of alkyne-H peaks and the increase of alkyne-D peaks from EU (10 mM) during alkyne-HDX in the pD=7.6 D_2O buffer solution. (b) Exponential curve fitting of normalized alkyne peak intensities for EU (10 mM) from (a). The average of $t_{1/2}$ over three independent measurements is also shown. (c) The kinetics trace of the spontaneous Raman spectra for the decrease of alkyne-H peaks and the increase of alkyne-D peaks from EdC (20 mM) during alkyne-HDX in the pD=7.6 D_2O buffer solution. (d) Exponential curve fitting of normalized alkyne peak intensities for EdC (20 mM) in (c). The average of $t_{1/2}$ over three independent measurements is also shown.

2.5 Establishing a general rule between alkyne-HDX and the probe structures

Guided by Eq. (2.3), it would be highly desirable to experimentally establish a quantitative relationship between the pK_as of alkynes and their exchange kinetics to inform the future structural design for alkyne-HDX probes. Even though powerful pK_a prediction methodologies have been developed, precise calculation of the pK_a for a given alkynyl hydrogen remains challenging, limited by the solvent leveling effect.[39] Hence, we adopted the Hammett substituent constant σ (Figure 2.5a), defined in Eq. (2.4) below,[48] as a surrogate for pK_a, to explore the relationship between the pK_a values and the HDX kinetics.

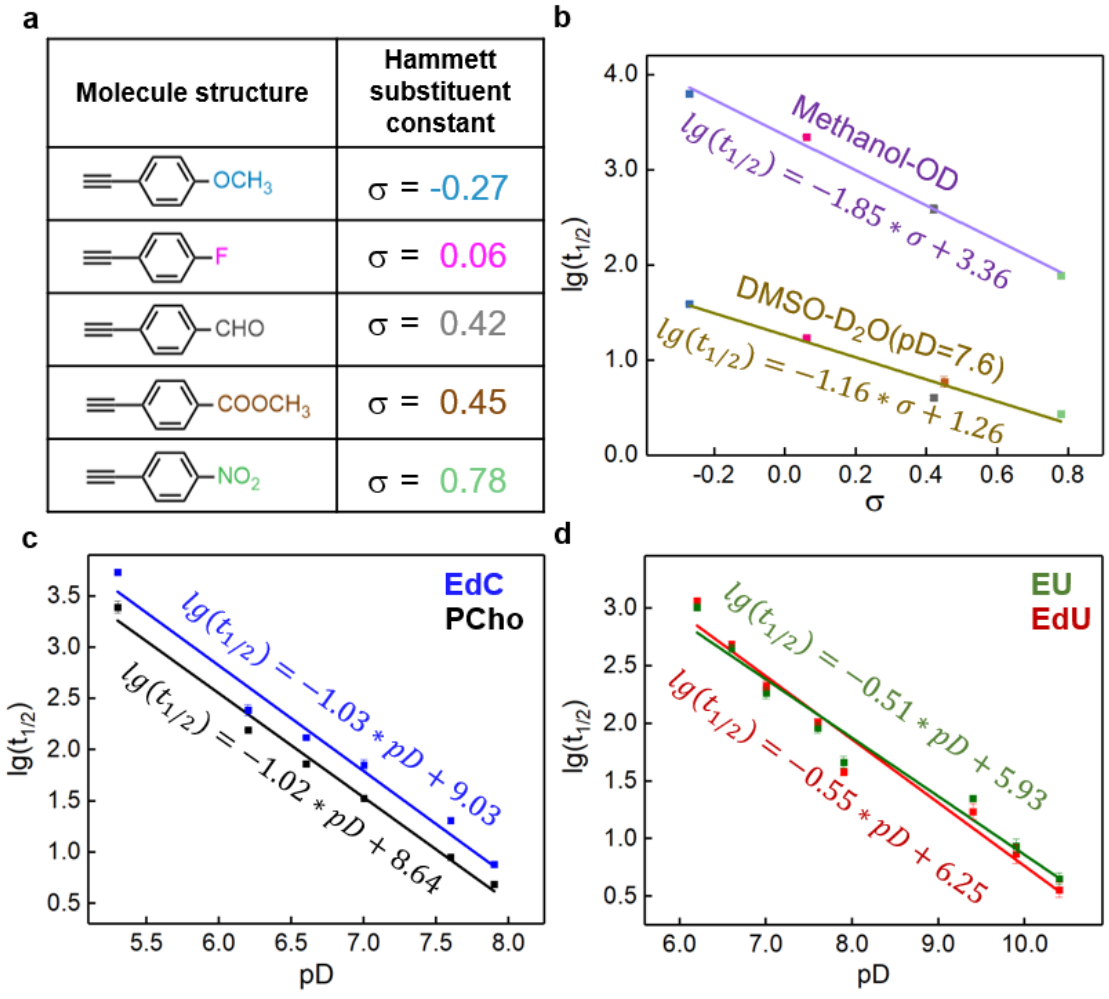

Figure 2.5. The effect of pK_a and pD on the alkyne-HDX kinetics. (a) Structures of model

molecules and the corresponding Hammett substituent constant σ. (b) Linear relationship between $\lg(t_{1/2})$ and σ for model molecules in the methanol-OD solution and 1:1 D_2O-DMSO buffer solution with pD=7.6. Corresponding data points for respective molecules shown in (a) are color-coded. (c-d) Linear relationship between $\lg(t_{1/2})$ and pD for EdC (10 mM) and PCho (40 mM) (c) and EdU (10 mM) and EU (10 mM) (d) in D_2O buffer solutions with different pD values. Error bar: SD.

$$\sigma_X = \lg K_X - \lg K_H \tag{2.4}$$

K_H is the ionization constant for benzoic acid in water at 25 °C and K_X is the corresponding constant for a para-substituted benzoic acid. A positive (negative) σ value represents an electron-withdrawing (donating) group (Figure 2.5a), and more positive σ values were reported to correlate with lower pK_as. The same correlation should also hold for alkynyl hydrogen in alkyne probes.[48]

We chose five phenyl-acetylene derivatives with various σ as our model molecules (4-fluorophenylacetylene, methy-4-ethynylbenzoate, 4-ethynylbenzaldehyde, 1-ethynyl-4-nitrobenzene, 1-ethynyl-4-methoxybenzene). Due to limited solubility of substituted phenylacetylene molecules in aqueous solutions, we measured their alkyne-HDX kinetics in two different deuterium conditions of 1:1 DMSO-D_2O solutions (pD=7.6) and methanol-OD. We observed a nice linear relationship between $\lg(t_{1/2})$ and σ for the model phenylacetylenes in both deuterium solutions (Figure 2.5b). With the lowest σ value, 1-ethynyl-4-methoxybenzene with an electron-donating group (-OMe) indeed shows the slowest kinetics. Similarly, with the highest σ value, 1-ethynyl-4-nitrobenzene with the

strongest electron-withdrawing group (-NO_2) has the fastest kinetics. The observed linear relationship between lg($t_{1/2}$) and σ together with Eq. (2.3) confirms that σ is a quantitative surrogate for alkyne pK_as. These data hence offer a general interpretation for the dependence of alkyne-HDX kinetics on chemical structure changes and would help guide the design of alkyne probes with specific alkyne-HDX kinetics.

2.6 Experimentally confirming a linear relationship between alkyne-HDX and pD

In addition to pK_a, another factor that directly influences HDX rate is the pD in exchange solutions, as shown in Eq. (2.3). To experimentally establish the relationship between pD and HDX kinetics, we measured the $t_{1/2}$ on all above alkyne probes dissolved in D_2O buffers with varying pD values. Consistently, lg($t_{1/2}$) displays a nice linear relationship with pDs for all four probes (Figure 2.5c, d). It is worth pointing out that the theoretical lg($t_{1/2}$) - pD slope is -1 in the simplified Eq. (2.3). While our data for EdC and PCho meet this criterion well, the slopes for EdU and EU (e.g. -0.5) deviate from the prediction. A similar mismatch in slopes has been shown in previous work on amide-HDX in proteins.[49-50] Further efforts to understand other factors that lead to the deviation are still ongoing.

2.7 Fast and homogenous D_2O diffusion throughout the whole cells

After establishing the quantitative nature of the alkyne-HDX approach, we now sought to apply it for cellular sensing. One potential issue moving from *in vitro* homogenous solution measurements to heterogeneous cells is the cellular uptake and intracellular diffusion of D_2O, which could influence the alkyne-HDX kinetics. Previous studies by magnetic field gradient

NMR have reported the outer membrane water permeability as ~0.001 cm/s and the intracellular diffusion coefficient as ~10^{-6} cm^2/s.[51-52] Taking the typical cell size (~10 μm) into consideration, we calculated that cells should experience a full D_2O environment within a second of D_2O incubation. Such rapid D_2O diffusion timescale should not influence alkyne-HDX rates in cells.

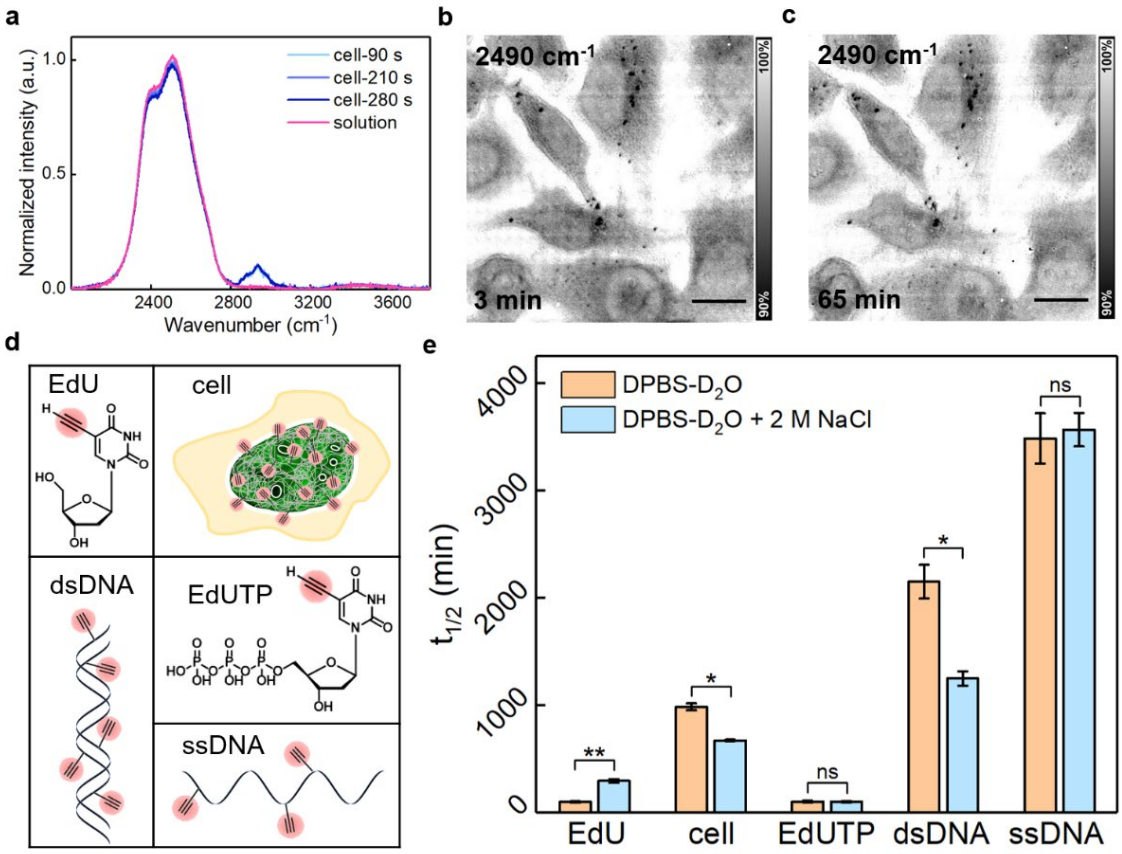

Figure 2.6. Fast D_2O diffusion across cells and the alkyne-HDX-based DNA structure sensing. (a) Time-trace spontaneous Raman spectra in the nuclear region of a living cell at the indicated time points after incubation in the DPBS-D_2O solution (90 s, 210 s, 280 s). The solution spectrum (magenta) is taken in the surrounding region without cells in the same

sample. (b-c) SRS image of O-D vibrational peak (2490 cm^{-1}) for live cells in DPBS-D_2O buffer solution after 3 min (b) and 65 min (c) incubation. Scale bar: 20 μm. (d) Cartoon structures of EdU, EdU-incorporated cells (cell), 5-Ethynyl-dUTP (EdUTP), EdU-labeled double-stranded DNA (dsDNA) and EdU-labeled single-stranded DNA (ssDNA). (e) The corresponding $t_{1/2}$ values of alkyne-HDX for EdU, EdU-incorporated cells, EdUTP, EdU-dsDNA, EdU-ssDNA in DPBS-D_2O buffer solution (light blue) and in the high salt buffer solution (DPBS-D_2O + 2 M NaCl, orange). Error bar: SD. *$p < 0.05$, **$p < 0.01$, ns $p > 0.1$ from unpaired t-tests.

To experimentally justify that D_2O diffusion would not influence alkyne-HDX, we performed time-lapse Raman spectroscopy and imaging by targeting the O-D stretching peak (~2490 cm^{-1}) as an indicator to trace the D_2O uptake and diffusion in both live and fixed cells after immersion in the DPBS-D_2O buffer. Our spontaneous Raman spectra confirmed that D_2O reaches cell nuclei (evidenced by the hump at ~ 2490 cm^{-1}) and achieves equilibrium (> 98% compared to the D_2O solution signal outside the cells) within tens of seconds of incubation in live cells (Figure 2.6a). Additionally, the signal from O-H stretching peak (~3430cm^{-1}) is nearly negligible, indicating that there is almost no residual H_2O in cells (Figure 2.6a), consistent with the fast water diffusion. To exclude the potential issue of heterogeneous intracellular distribution of D_2O, we also performed time-lapse SRS imaging at 2490 cm^{-1}. Images taken after 3 min and 65 min of incubation are almost identical (Figure 2.6b, 2.6c), indicating that the intracellular D_2O level reaches equilibrium within 3 min, similar to what is shown in spontaneous Raman (Figure 2.6a). The D_2O distribution across cells is mostly homogenous with an observed heterogeneity less than 5%, consistent with the

normal distribution of H_2O recorded by O-H stretching peak.[53] We note that the dark spots shown in Figure 2.6b-c are hydrophobic lipid droplets. Similar spontaneous Raman and SRS measurements for diffusions were also captured in fixed cells (Figure 2.7a-c). All above results confirm that D_2O diffusion is a fast and homogeneous process that would not interfere with the measurements of the alkyne-HDX kinetics in cells.

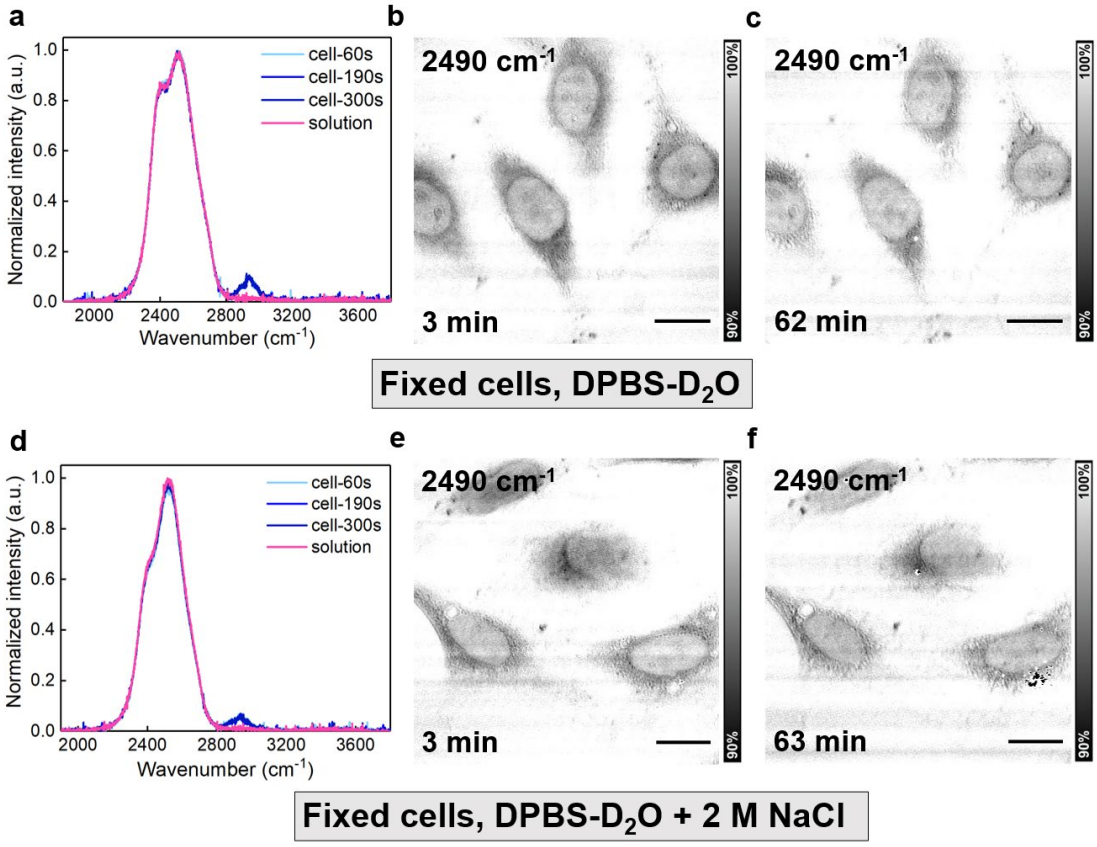

Figure 2.7. Fast D_2O diffusion across cells in solutions with different osmolarities. (a) Time-trace spontaneous Raman spectra in the nuclear region of a fixed cell in the DPBS-D_2O solution. The solution spectrum is taken in the surrounding region without cells from the same sample. (b-c) SRS image of O-D vibrational peak (2490 cm^{-1}) for the same set of

fixed cells in DPBS-D_2O buffer solution after 3 min (b) and 62 min (c) incubation. (d) Spontaneous Raman spectra in the nuclear region of a fixed cell in the high osmolarity solution (DPBS-D_2O + 2 M NaCl). The solution spectrum is taken in the surrounding region without cells from the same sample. (e-f) SRS image of O-D vibrational peak (2490 cm^{-1}) for the same set of fixed cells in DPBS-D_2O + 2 M NaCl buffer solution after 3 min (e) and 63 min (f) incubation. Scale bar: 20 μm.

2.8 Sensing the changes of DNA structures by alkyne-HDX

The ability to analyze DNA structural changes is essential yet challenging. The majority of existing methodologies rely on solution characterizations *in vitro*, which may cause perturbations to the original structures during extraction, purification, and reconstitution processes[54-55]. Understanding DNA structures in their native environments should gain new insights for scenarios such as DNA damage, genome instability, and DNA-protein interactions. Toward this goal, EdU has been proven as a well-established live-cell compatible SRS reporter for DNA where alkynes on EdU presumably extends into the major grooves of double helix[56] . We hypothesize that the changes in DNA structures would modulate the local microenvironments of alkynes and this process could be reflected and monitored by the alkyne-HDX of EdU. To verify the hypothesis, we acquired the HDX kinetics in cellular DNA by collecting a time series of SRS images at the alkyne-H (2224 cm^{-1}) and the alkyne-D (1992 cm^{-1}) frequencies on EdU-labeled cells immersed in the DPBS-D_2O buffer (Figure 2.8b). Surprisingly, the alkyne-HDX kinetics for EdU in cells are

significantly (~ 10 times) slower (Figure 2.8a, $t_{1/2}$ = 16.5 hours) compared to that for EdU in the DPBS-D_2O buffer solution.

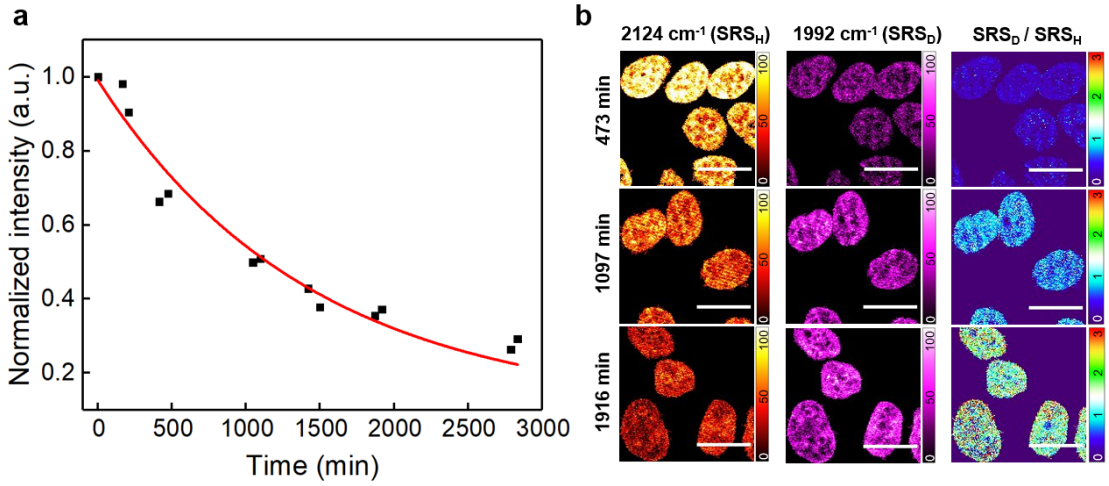

Figure 2.8. Representative data sets for SRS imaging-based alkyne-HDX kinetics on EdU-labeled cells. (a) Exponential curve fitting of normalized alkyne peak intensities for EdU-incorporated cells in a pD=7.6 D_2O buffer solution. (b) Representative ratiometric imaging of SRS_D / SRS_H (right) generated by dividing the SRS images at the alkyne-H channel (2224 cm^{-1}) (left) by those at the alkyne-D channel (1992 cm^{-1}) (middle) for cells immersed in DPBS-D_2O from a series of alkyne-HDX time points (473 min, 1097 min, 1916 min).

To better understand this slow-down effect in cells, we evaluated the alkyne-HDX for potential EdU-labeled downstream products (Figure 2.6d) *in vitro* in the same DPBS-D_2O buffer (Figure 2.6e, orange bars). We first tested the HDX kinetics of EdUTP (EdU triphosphate), because EdU needs to be phosphorylated before being incorporated into

genomic DNA and the phosphate groups bring extra negative charges to the vicinity of the alkynes. The $t_{1/2}$ for EdUTP is 105.5 min that is virtually the same as EdU (Figure 2.6e, Figure 2.9c). Hence, we ruled out the possible contribution of added phosphate groups to the slower alkyne-HDX kinetics. We then hypothesized that the formation of polynucleotides where EdU was compressed into conformationally restricted polymer chains leads to the slowdown. Consequently, we generated EdU-labeled double-stranded DNA (dsDNA, 746 bp) and single-stranded DNA (ssDNA, 746 nt) (Figure 2.6d) by replacing dTTP with EdUTP in PCR mixture and using the EdUTP-tolerating pwo DNA polymerase.[57] The amplification of ssDNA was based on asymmetric PCR, in which concentrations of forward and reverse primers are different.[58] The successful formation of dsDNA and ssDNA was confirmed through gel electrophoresis and ssDNA-specific Exonuclease I (ExoI) digestion[58] (Figure 2.9f, g). The purified EdU-labeled ssDNA and dsDNA were then used for alkyne-HDX kinetics measurements (Figure 2.6e, $t_{1/2}$ (EdU-dsDNA) = 36 hours and $t_{1/2}$ (EdU-ssDNA) = 58 hours). We also attempted to measure the alkyne-HDX kinetics of purified EdU-labeled genomic DNA extracted from cells in solutions, but unfortunately failed due to the limited solubility and lower labeling efficiency of EdU inside cells compared to PCR reactions. These data indicate that the formation of polynucleotide chains plays a major role in the slowdown of HDX kinetics in EdU-incorporated cell.

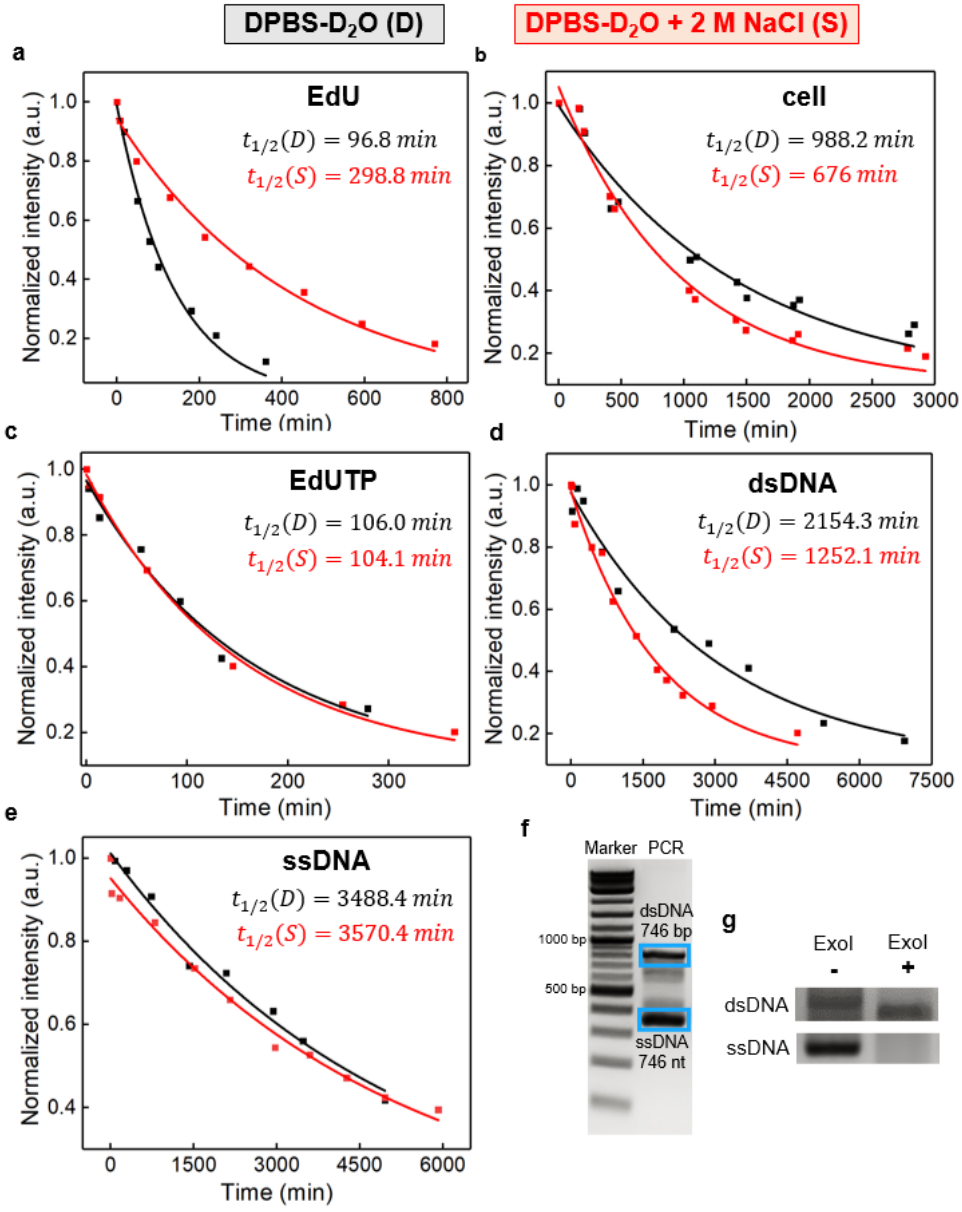

Figure 2.9. The alkyne-HDX kinetics for EdU and EdU-labeled structures at the regular (DPBS-D_2O, D, black) and high (DPBS-D_2O+2M NaCl, S, red) salt concentration. (a-e) Exponential curve fitting of normalized alkyne peak intensities for EdU (10 mM) (a); EdU-incorporated cells (b); EdUTP (10 mM) (c); EdU-labeled dsDNA (d); and EdU-labeled ssDNA (e) in a pD=7.6 D_2O buffer solution (black) and pD=7.6 D_2O buffer

solution with 2M NaCl (red), respectively. (f) DNA gel electrophoresis for confirming the EdU-incorporated products of dsDNA and ssDNA. Left: 1 kb DNA ladder; right: EdU-labeled PCR products, containing 746 bp dsDNA and 746 nt ssDNA. (g) DNA gel electrophoresis confirming the presence of EdU-dsDNA and EdU-ssDNA with Exonuclease I (ExoI) digestion. Left: control PCR products; right: PCR products after Exonuclease I (ExoI) digestion.

Interestingly, among all EdU labeled polynucleotides, EdU-ssDNA showed the slowest alkyne-HDX, whereas EdU-labeled cellular DNA has the fastest kinetics (Figure 2.6e, orange bars). This result implies that our alkyne-HDX strategy is indeed sensitive to the change of DNA structures. To gain more insights on the causes of varying alkyne-HDX kinetics across different EdU-labeled DNA structures, we next investigated how the high-salt condition (DPBS-D_2O buffer with 2 M NaCl),[59] which could perturb the DNA electrostatic interactions, modulates the alkyne-HDX of EdU (Figure 2.6e, light blue bars). We confirmed that the high osmolarity wouldn't alter the D_2O diffusion rate across cells by both spontaneous Raman and SRS measurements (Figure 2.7d-f). Interestingly, while the high-salt condition accelerated the HDX for both EdU-cells and EdU-dsDNA (Figure 2.6e, 2.9), it did not influence the HDX kinetics for both EdUTP and EdU-ssDNA. As a control, alkyne-HDX for the EdU was instead slowed down in the high-salt buffer compared to that in the regular DPBS-D_2O buffer. This means that the accelerating effect of the high-salt condition should be mostly attributed to the DNA structural changes.

Based on above data, one possible explanation for the significant HDX slowed-down of ssDNA could be the formation of intramolecular hydrogen-bonds (H-bonds) around alkynes,[60] which effectively block the alkyne-HDX by shielding. Similar H-bonds formation is known to slow down the HDX.[36] This postulation for forming strong H-bonding is also consistent with the minimum influence on the alkyne-HDX for ssDNA in high-salt condition (Figure 2.6e). As a comparison, in dsDNA (and also cellular DNA) with well-organized helical structures, the formation of Watson–Crick base pairs pushes alkynes towards major grooves, likely eliminating H-bonding around EdU alkynes. They both hence present faster alkyne-HDX compared to ssDNA (Figure 2.6e). The high-salt condition is also known to compromise the double helical structure and lead to partial denaturation of dsDNA.[59] This change of structural integrity was then reflected in the alkyne-HDX speed-up for both EdU-labeled dsDNA and cellular DNA. Collectively our results for different DNA species under varied conditions establish the viability of employing alkyne-HDX to sense and study the local DNA environments.

2.9 Sensing the changes of cellular DNA structures through UV-induced damage *in situ* by alkyne-HDX

DNA damage inhibits DNA replication, interrupts the normal function of cells and potentially leads to cell death. Studies of structural changes upon DNA damage have been mostly limited to *in vitro* examinations.[54, 61-62] As alkyne-HDX of EdU is a viable approach to monitor DNA double helices, next we aimed to apply our strategy to investigate UV-induced DNA damage in native cellular environments. DNA damage induced by UV

irradiation involves the formation of cyclobutene pyrimidine dimers (CPDs), in which the thymine dimer (T-T dimer) is the most common one (Figure 2.10a).[63-66] As a thymidine analogue, EdU would be a perfect reporter for interrogating the T-T dimer formation. To establish the proper dimerization condition, we first implemented a reported procedure to induce thymine dimerization in solutions by UV irradiation.[62] The reported change in spontaneous Raman spectra was observed (Figure 2.11a), confirming the dimer formation.[62] Similar UV condition was also suited for inducing T-T dimer formation in live cells, which was confirmed by immunostaining (anti-T dimer, Figure 2.11b, c). With the established condition, we next examined the formation of T-T dimer in EdU-labeled cellular DNA by alkyne-HDX (Figure 2.10b). Interestingly, we observed an about 30% faster alkyne-HDX kinetics for UV-cells (Figure 2.10c, $t_{1/2}$(UV-cell) = 11.6 h vs $t_{1/2}$(cell, with no UV) = 16.5 h). We speculated that the increase of HDX kinetics could originate from two non-mutually exclusive factors: first, the explicit chemical structure change upon EdU-EdU dimer formation, and second, the change in local DNA structures and thus alkyne microenvironments upon formations of T-T (or EdU-EdU or T-EdU) dimers (Figure 2.10a).

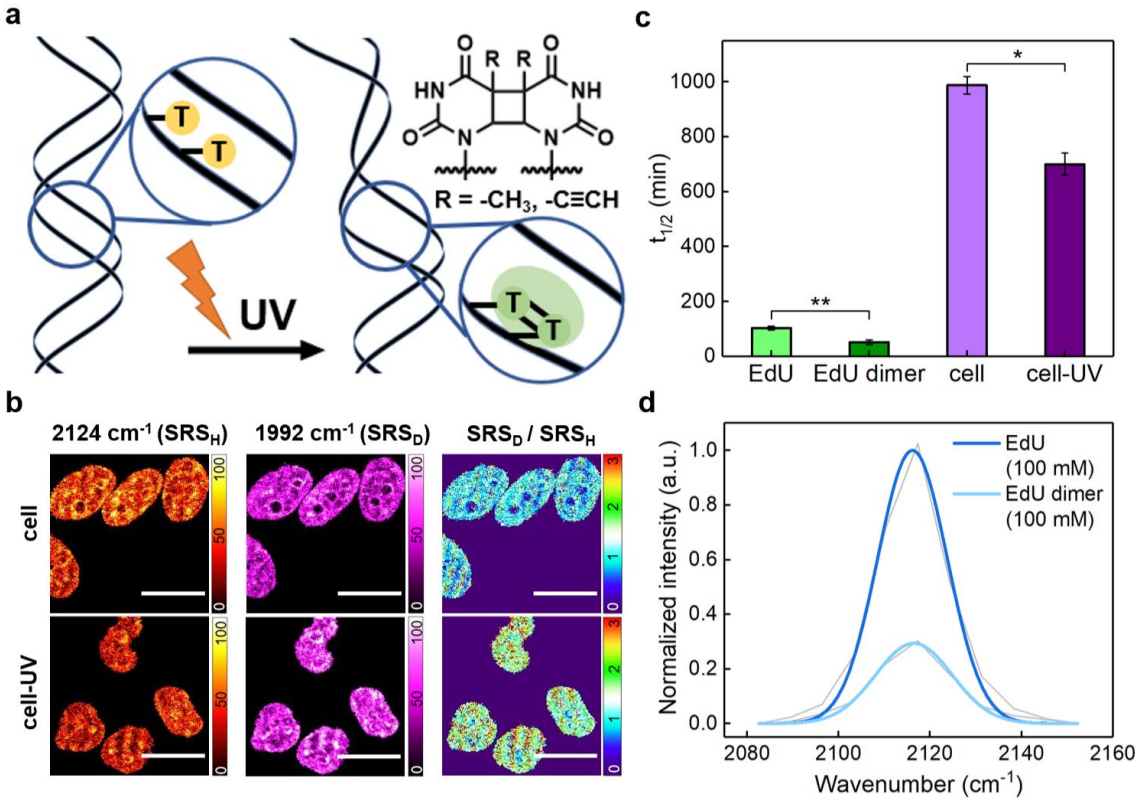

Figure 2.10 Sensing the UV-induced damage on DNA structures by alkyne-HDX from EdU-labeled cells. (a) The dimerization scheme in regular or EdU-labeled DNA induced by UV (254 nm) irradiation. (b) Ratiometric imaging of SRS_D / SRS_H (right) generated by dividing the SRS imaging at the alkyne-H channel (2224 cm^{-1}) (left) by that at the alkyne-D channel (1992 cm^{-1}) (middle) for cells without (top, average ratio of 1.11±0.47) and with (bottom, average ratio of 1.61±0.53) UV irradiation immersed in DPBS-D_2O after 23.5h incubation. Experiments were repeated with three independent measurements. Scale bar: 20 μm. (c) SRS-imaging measured $t_{1/2}$ of alkyne-HDX from EdU and EdU dimer solutions, EdU-incorporated cells, and EdU-incorporated cells after UV irradiation. Error bar: SD. (d) Raman spectra for characterization of alkyne Raman intensity in EdU (100 mM in DMSO)

and EdU dimer (100 mM in DMSO) based on SRS measurements. Error bar: SD. *$p < 0.05$, **$p < 0.01$, ns $p > 0.1$ from unpaired t-tests.

To test our hypothesis, we purified the EdU dimer using reverse-phase column chromatography and further characterized the spectral properties and the alkyne-HDX kinetics of EdU dimer. While no obvious Raman peak shift was observed, the SRS alkyne intensity of EdU dimer became only one-quarter of that from EdU. This means the intensity of alkyne in each EdU becomes 1/8 of the original intensity, probably due to the loss of conjugation to alkyne (Figure 2.10d). The alkyne-HDX kinetics of EdU dimer solution increased by almost two folds, indicating a decreasing alkyne-pK_a upon dimer formation (Figure 2.10c, EdU vs EdU dimer). With the observed 30% faster HDX kinetics in cells, we would expect a decent percentage of dimer formation that would result in a significant drop in SRS signals. This contradicts our observed SRS images for UV-irradiated cells whose intensities remain approximately unchanged (Figure 2.10b). Furthermore, the limited EdU-labeling efficiency in live cells should also restrict the probability of the EdU dimer formation, which requires two EdU residing in proximity. Collectively, our data largely ruled out that direct EdU dimer formation is the primary reason for the increased alkyne-HDX kinetics. We proposed that the accelerating HDX kinetics instead senses the disruption of double-helical structures caused by UV-induced T-T dimerization (Figure 2.10a). This is also consistent with previous reports that the formation of T-T dimers would unwind the DNA double helix about 10° and lead to a bend of about 30°.[61] In this scenario, the EdU essentially serves as a reporter for T-T dimer formation: the EdU could experience a compromised double-helical structure due to the formation of CPDs and shows faster alkyne-HDX kinetics

(Figure 2.10c). The high salt treatment and UV-irradiation result in similar EdU-HDX in cellular DNA, suggesting both dimer formation and the high-salt condition might have similar levels of disruptions in DNA double-helical structures.

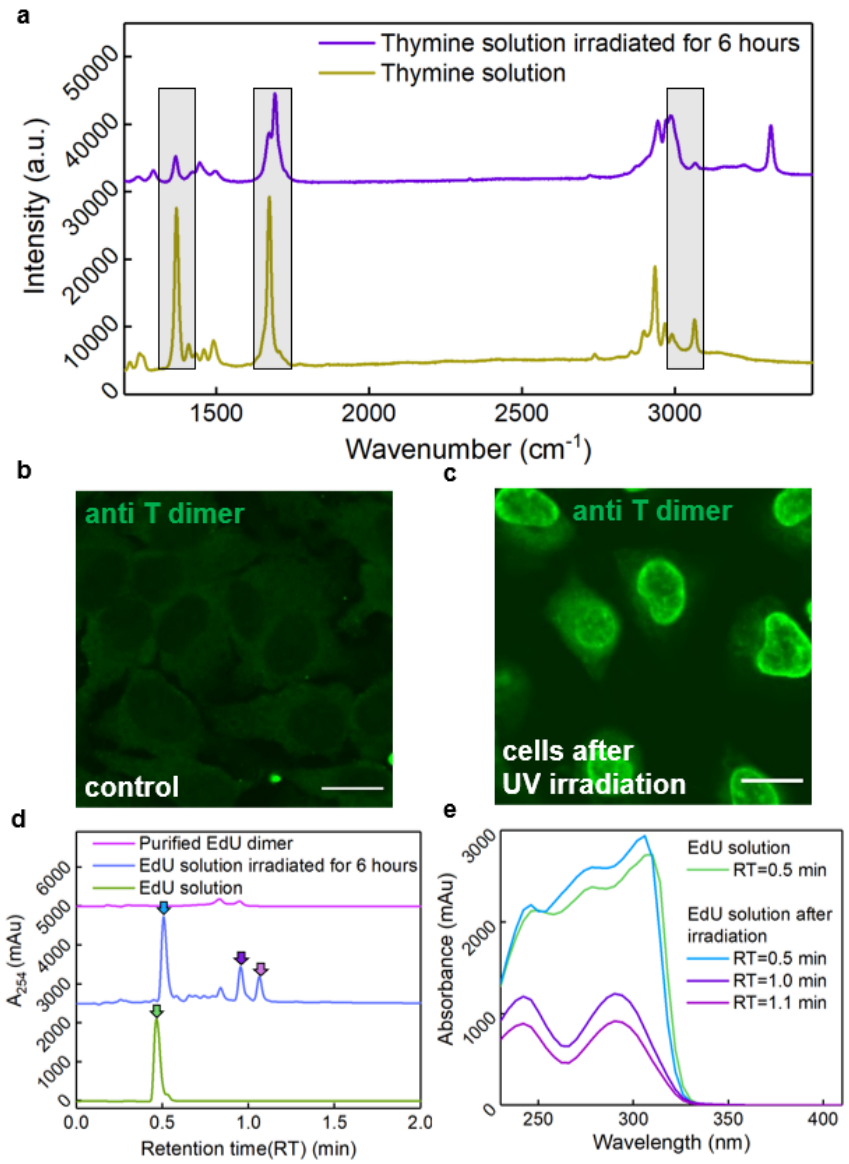

Figure 2.11 Confirmation for the UV-induced dimer formation. (a) Spontaneous Raman spectra of thymine solution (yellow) and thymine solution irradiated for 6 hours (purple).

The gray-boxed regions show the featured Raman spectra changes for thymine dimer formation, consistent with what was reported in Ref. 62. (b-c) Immunofluorescence staining images with anti-T dimer-Alexa Fluor 647 for control cells without UV irradiation (control, b) and UV-irradiated cells (c). (d) HPLC trace of EdU solution (light green), EdU solution irradiated for 6 hours (light purple) and purified EdU dimer (magenta). Arrowed peaks indicate the corresponding color-coded UV absorption traces in (e). e) UV absorption traces of peaks indicated by color-coded arrows in the HPLC trace (d) of the EdU solution before and after UV irradiation.

2.10 Sensing minor pD variations in live cells by alkyne-HDX

In addition to sensing chemical environments, our alkyne-HDX is also highly sensitive to the environmental pDs, as shown above with the logarithmic dependence of $t_{1/2}$ on pD. We therefore implemented alkyne-HDX to potentially resolve subtle pD changes in the cellular environment. Such detectability would find wide utility, particularly in differentiating the local and heterogeneous acidity in acidic diseases including cancer, inflammation and others.[67] To achieve such a goal, we first calibrated the intracellular pDs with a fluorescent ratiometric pH sensor BCECF (2',7'-is-(2-Carboxyethyl)-5-(and-6)-Carboxyfluorescein) in both D_2O buffer solutions (Figure 2.13a) and live cells immersed in these buffers with varying pD values (Figure 2.12a, left, control). The relative constant ratios (F_{488}/F_{445}) for cells indicate a cellular buffering capacity to the external pD changes (Figure 2.12a, control, left). We further adopted the treatment of nigericin, an ionophore that is utilized to modulate intracellular pH by changing the efflux of K^+/H^+. It could hence help overcome the cellular

buffer capacity and modulates the cytosolic pH by extracellular buffer pH (Figure 6a, right, nigericin; Figure 2.13b-e).[68-69]

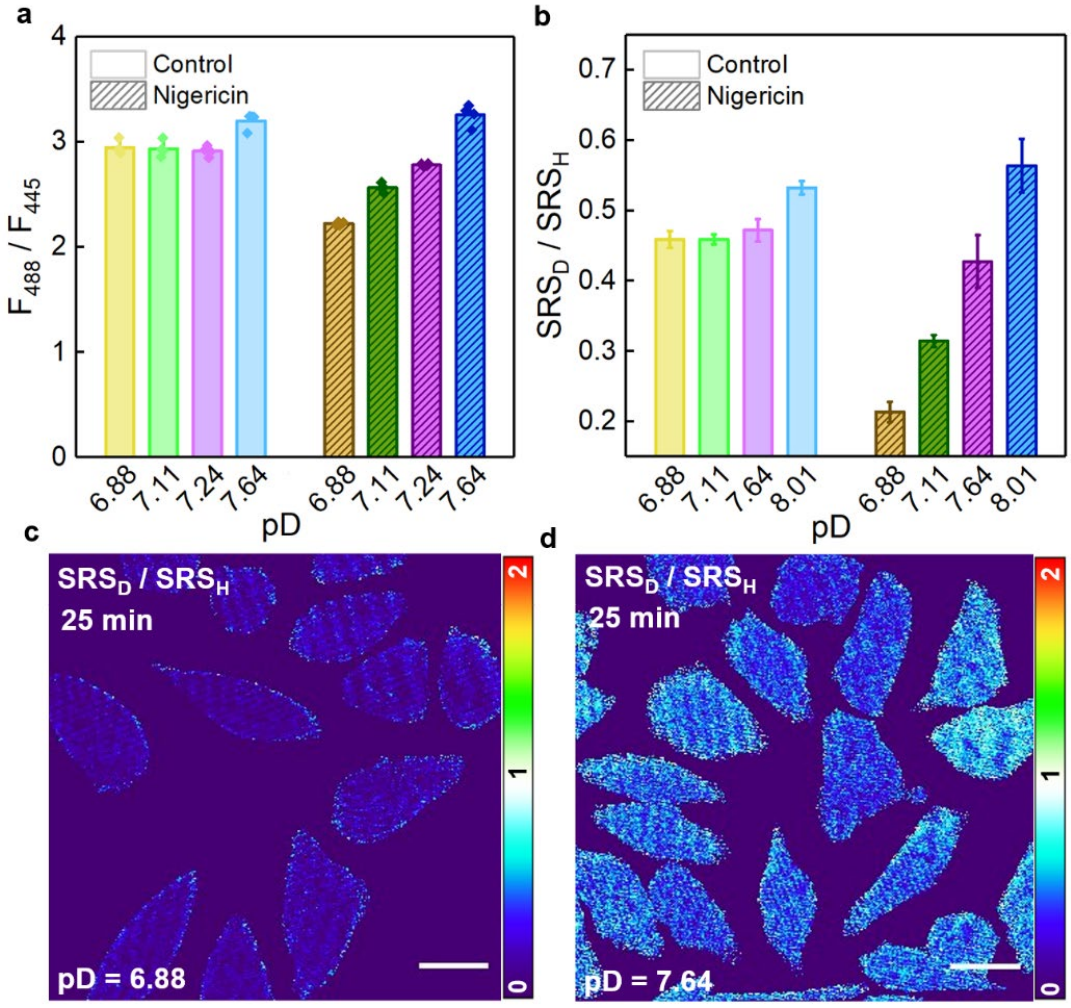

Figure 2.12. Sensing intracellular pH changes by alkyne-HDX. (a) Benchmarking of intracellular pD values by the standard fluorescence ratiometric (F_{488} / F_{445}) pH indicator BCECF from live cells immersed in DPBS-D_2O buffers with minor changes around the physiological pD range with (right) and without (left) extra nigericin for intracellular pD calibration from the extracellular environment. (b) The corresponding SRS-imaging ratios

(SRS_D / SRS_H) of PCho in live cells immersed in the DPBS-D_2O buffers with similar pD range with (right) and without (left) extra nigericin. Error bar: SD. (c-d) Ratiometric imaging of SRS_D / SRS_H for PCho at 25 min exchange in live cells in pD=6.88 (c) and pD=7.64 (d) DPBS-D_2O buffer solutions. The average ratio for (c) is 0.19±0.13 and for (d) is 0.47±0.21. Scale bar: 20 μm.

We next captured alkyne-HDX live-cell ratiometric images (I(R-C≡C-D)/I(R-C≡C-H), i.e. SRS_D/SRS_H) for PCho-labeled cells in D_2O buffers without (control) and with nigericin (Figure. 2.12b-d) at the same exchange time point. The same buffering capacity is shown with SRS ratios (Figure 2.12b, left, control). The increase in SRS_D/SRS_H ratio indicates faster alkyne-HDX. The trend of SRS_D/SRS_H is also consistent with the nigericin-balanced intracellular pD values (Figure 2.12b, left, nigericin). The fold change on our SRS ratios on pD changes is much higher than that from fluorescence ratios in BCECF (Figure 2.12a, 2.12b), demonstrating our ability to readout small cellular pD differences in live cells. Ideally, we would acquire exponential decay kinetics for PCho alkyne-HDX in cells and derive a $t_{1/2}$. Instead, we discovered interesting exchange kinetics of a biphasic trend with a fast (close to what is shown for PCho solution) and a much slower-exchanging component. One possible explanation for the observed 2-stage HDX is that PCho experiences two different environments as the headgroup of phospholipids: phospholipids exposed to an aqueous environment exhibit a faster exchange rate and those buried in a hydrophobic environment exhibit a slower rate with almost no exchange in the detect time window ($t_{1/2}$ ~2 hours).

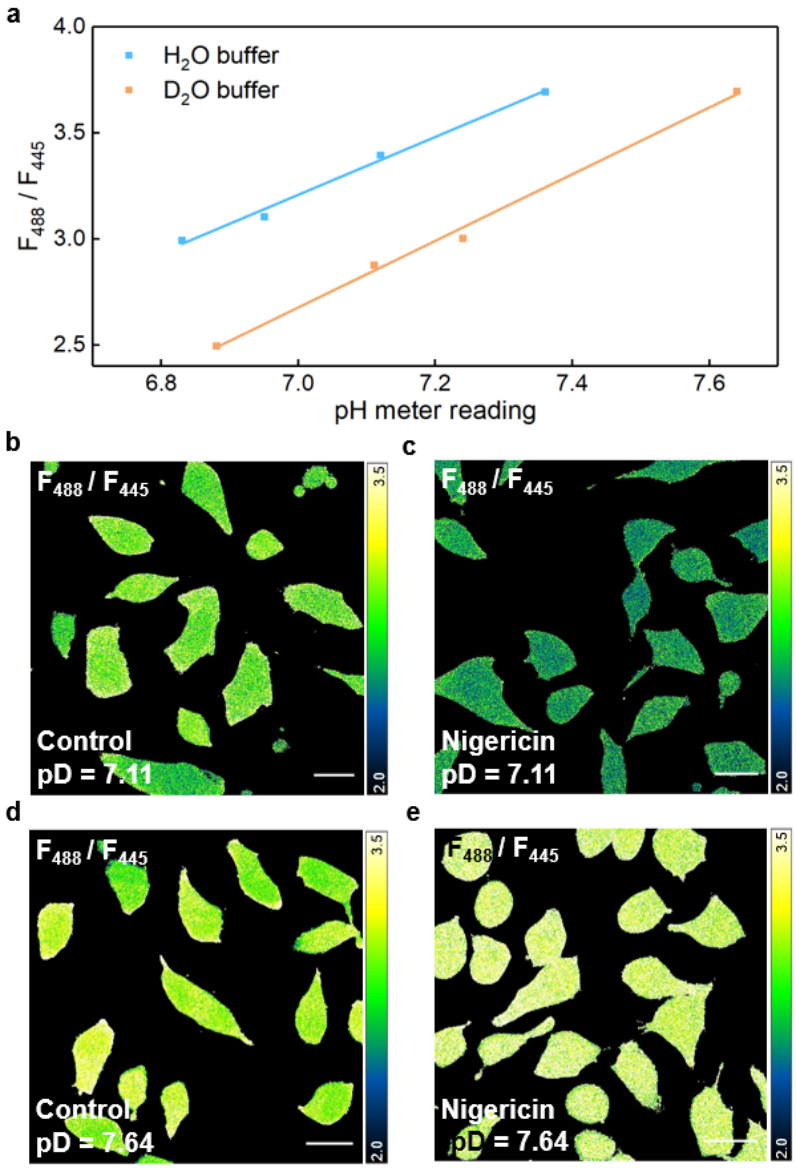

Figure 2.13. Calibration of solution and intracellular pD by a ratiometric fluorescent pH sensor BCECF. (a) Linear relationship between ratios (F_{488} / F_{445}) taken at the 488 nm excitation and 445 nm excitation of BCECF in both H_2O (light blue) and D_2O (orange) buffers with varying reading results from the pH meter over the physiological-relevant range. (b-c) Representative ratiometric images (F_{488} / F_{445}) for BCECF in live cells in pD=7.11

DPBS-D_2O buffer without (b, control, average ratio 2.93±0.38) and with nigericin (c, average ratio 2.55±0.31). (d-e) Representative ratiometric images (F_{488} / F_{445}) for BCECF in live cells in pD=7.64 DPBS-D_2O buffer, without (d, control, average ratio 3.01±0.31) and with nigericin (e, average ratio 3.34±0.44). Scale bar: 20 μm.

2.11 "Two-color" mapping of DNA and lipids in live cells via distinct kinetics of the subcellular alkyne-HDX maps

As EdU and PCho have vastly distinct alkyne-HDX kinetics in cells, we expected to differentiate EdU and PCho in ratiometric (SRS_D/SRS_H) SRS images. We first labeled DNA and lipids in HeLa cells with EdU and PCho and then incubated cells with custom-made DMEM-D_2O. We chose to take SRS images at the overlapping frequency channels 2137 cm^{-1} (SRS_H) and 1998 cm^{-1} (SRS_D) (Figure 2.1c), so that both targets could be mapped at the same channels, and we could avoid acquiring four images, which takes longer time and may introduce complications from cell movement. As EdU-DNA now goes through HDX about 100 times slower than PCho in cells, pixels from EdU-enriched DNA in the nuclear regions should show much smaller ratios than those from PCho-labeled lipid membranes. In ratiometric maps at 7 min, 11 min, and 33 min of D_2O incubation, nuclei indeed always have a smaller ratio compared to cytoplasm (Figure 2.14). Our live-cell ratiometric imaging hence provides an additional dimension to perform "two-color" imaging based on the differentiation of alkyne-exchange kinetics, but not from frequency difference. This might be potentially useful to extract multi-component information in cells for alkyne-tagged Raman probes with close vibrational frequency, but distinct alkyne-HDX kinetics.

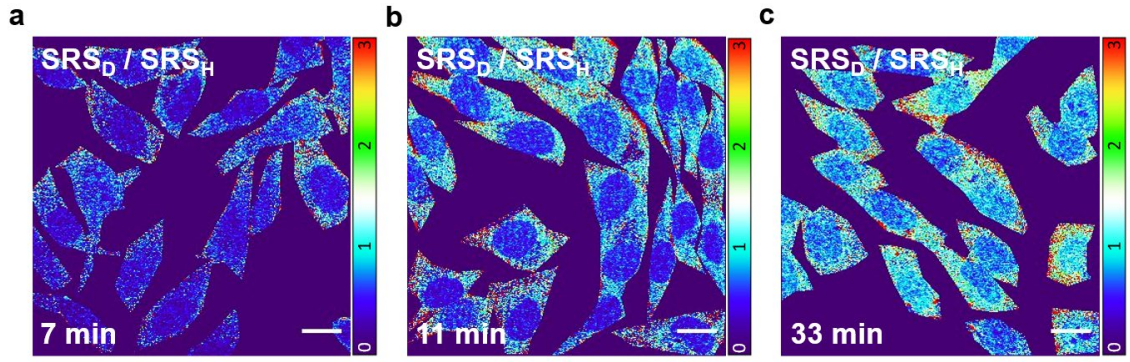

Figure 2.14 Ratiometric imaging for "two-color" identification of EdU-labeled DNA and PCho-labeled membranes by distinct alkyne-HDX kinetics between the two labeled species (a-c) Ratiometric imaging (SRS_D (1998 cm^{-1})/SRS_H (2137 cm^{-1})) for both EdU and PCho in live cells at three exchange time points. Scale bar: 20 μm.

2.12 Discussion

We reported a novel Raman-based imaging platform for sensing local chemical and environmental changes in cells by analyzing HDX kinetics of alkyne-tagged probes. We theoretically investigated the alkyne-HDX mechanism and quantitatively established the relationship for the alkyne-HDX kinetics (i.e. $lg(t_{1/2})$) with probe structures, and local chemical and cellular environments (i.e. pK_a and pD). The drastic peak shift accompanied with alkyne-HDX finds its perfect marriage with SRS microscopy, enabling sub-cellular exchange analysis with high spatial and temporal resolution. The exchange also provides an additional dimension for multiplex imaging for alkyne-tagged probes with large spectral cross-talks, but distinct alkyne-HDX kinetics. These properties render alkyne-HDX a uniquely suited imaging-based strategy for *in situ* sub-cellular sensing applications,

including sensing small pD changes and DNA structure changes, which all remained challenging to be interrogated previously. Despite limited methods for detecting thymine dimer lesions, our alkyne-HDX provides a noninvasive approach to understand thymine dimerization *in situ*.

It is worth discussing the use of pD values in our work. pD was reported to be directly measured by pH meter with an offset $pD_{real}=pH_{meter\ reading} + 0.4$.[70] We also extensively evaluated the relationship between our own pH meter readings and the actual pD values by both making the standard DPBS buffer in H_2O and D_2O, and by comparing the pD and pH values using fluorescent pH indicator, BCECF (Figure 2.13a). Our own results reliably established that $pD_{real}=pH_{meter\ reading}-0.25$ (more details in Materials and methods). The small difference of the offsets from our investigation and prior report might arise from different pH meter electrode probes or room temperatures. To avoid discrepancy and confusion, we reported all our pD values above directly from the pH meter reading results.

2.13 Future directions

As an outlook, there are a number of fronts that could be further explored for alkyne-HDX. Imaging-wise, when coupled with spectrally resolvable alkyne-probes, parallel multiplex alkyne-HDX sensing could be implemented to simultaneously investigate the local cellular environment in the vicinity of multiple alkyne-tagged biological species and structures. Chemically, alkyne-HDX should find wider utility with further chemical derivatization of the alkyne probes. We expect alkyne-HDX could serve as a general strategy for sensing enzymatic reactions in cells with properly designed alkyne probes. Biologically, alkyne-

HDX should provide unique insights on compartmentalized pDs in various cellular states with further studies.

We also envision alkyne-HDX to be further extended to investigate other local heterogeneity, such as phase separation. Our preliminary data (not shown) detected a weak but reproducible EdU signal in the non–membrane-bound sub-compartment in nucleoli, likely originated from ribosomal DNA, with almost no observable exchange kinetics. We suspect this non-exchange behavior is due to that rDNA is located in either fibrillar centers (FCs) or dense fibrillar components (DFCs) of the nucleoli,[71-72] the crowding of which blocks alkyne-HDX of EdU in rDNA in nucleolar regions.

Instead of DNA phase separation, we would like to expand the alkyne-HDX toolbox to understand the intracellular RNA phase separation, as RNA can be labeled with terminal alkynes with the incorporation of 5-ethynyl uridine (EU) through *in vitro* transcription. Nucleotide repeat expansion disorders are a group of genetic diseases caused by the expansion of specific DNA sequences consisting of repeating nucleotide triplets within certain genes.[73-75] How those repeat expansions lead to neurological disease is still unknown. Recently, people have demonstrated that RNA with certain repeat expansion sequences can undergo gel-like phase separation, forming condensates that may contribute to neurodegeneration and neurotoxicity.[76] Based on this foundation, we plan to apply alkyne-HDX analysis with *in vitro* formed RNA condensates to quantitatively unravel the local heterogeneity of the composition, chemical structures, and molecular interactions.[77-78] It holds the huge potential to advance our understanding of the RNA phase separation

mechanisms under varying conditions and potentially shine light on their role in disease pathology.

We are furthering our investigations along this line. Considering all the unique properties of our strategy, we believe our alkyne-HDX could extend the utility of alkynes from imaging to sensing and would present a valuable addition to the optical-imaging-based sensing toolbox in understanding complex biology.

2.14 Materials and Methods

Chemicals

All chemicals were purchased from Sigmal-Aldrich unless otherwise specified. 5-Ethynyl-2′-deoxyuridine (EdU, CAS# 61135-33-9) was purchased from TCI America. 5'-Ethynyl-2'-deoxycytidine (EdC, CAS# 69075-47-4) was purchased from Cayman Chemicals. 5-Ethynyl Uridine (EU, CAS# 69075-42-9) was purchased from Sigma-Aldrich. 5-Ethynyl-2'-deoxyuridine 5'-triphosphate (5-EdUTP) was purchased from Jena Bioscience. BCECF, AM (2',7'-Bis-(2-Carboxyethyl)-5-(and-6)-Carboxyfluorescein, Acetoxymethyl Ester) and BCECF, free acid were purchased from ThermoFisher scientific. Pwo DNA polymerase was purchased from Roche. Exonuclease I (ExoI) was purchased from Sigma-Aldrich. Nigericin sodium salt (CAS# 28643-80-3) was purchased from Sigma-Aldrich. Propargylcholine bromide (PCho, CAS# 111755-76-1) was synthesized according to Ref. 3. Briefly, propargyl bromide (80 wt. % solution in toluene, Sigma-Aldrich) was added dropwise to a stirring solution of 2-dimethylaminoethanol (Sigma-Aldrich) in anhydrous THF in ice bath under

argon gas protection. The reaction mixture was slowly warmed up to room temperature and stirred overnight. The resulting white solids were filtered and washed extensively with cold anhydrous THF to obtain pure PCho.

Buffer preparation

For solution experiments, we used different buffering systems according to the final pD region to prepare D_2O buffers. (Buffer range: citric acid-Na_2HPO_4, 2.6-7.6; NaH_2PO_4-Na_2HPO_4, 6.2-8.2; CAPSO-CAPSO sodium salt, 8.9-10.3). We prepared D_2O buffers by directly diluting corresponding buffer powders to the same amount of D_2O as that for the H_2O system. To keep ion strength consistent for all buffers to avoid the influence on the alkyne-HDX kinetics, we added extra NaCl to those buffers with less salt concentrations. The detailed buffer recipes are shown below: pD= 5.3 (46.4 mM citric acid, 107.2 mM Na_2HPO_4); pD= 6.2 (33.9 mM citric acid, 132.2 mM Na_2HPO_4); pD= 6.6 (27.25 mM citric acid, 145.5 mM Na_2HPO_4); pD= 7.0 (17.65 mM citric acid, 164.7 mM Na_2HPO_4); pD= 7.6 (DPBS); pD= 7.9 (3 mM NaH_2PO_4, 30 mM Na_2HPO_4, 60 mM NaCl); pD= 9.4 (28mM CAPSO, 7 mM CAPSO sodium salt, 143 mM NaCl); pD= 10.4 (19 mM CAPSO, 16 mM CAPSO sodium salt, 134 mM NaCl); pD= 10.4 (10 mM CAPSO, 25 mM CAPSO sodium salt, 125 mM NaCl). The final pD values were determined by the pH-meter. For salt concentration experiment, 2 M NaCl was added into the DPBS-D_2O buffer solution.

For BCECF experiments, high K^+ conditions were used to meet the requirement of nigericin. The detailed high K^+ buffer recipes are shown below: High K^+ D_2O buffer, pD = 7.64 (120 mM KCl, 5 mM NaCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4 in D_2O); pD=6.88, 7.11, 7.24

buffers were made by adjusting pD of pD=7.64 buffer with drops of 0.1 M HCl in D_2O. pD= 8.01 buffer was made by adjusting pD of pD=7.64 with drops of 0.1 M $NaHCO_3$ in D_2O. High K^+ H_2O buffer, pH = 7.35 (120 mM KCl, 5 mM NaCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4 in H_2O); pH = 6.83, 6.95, 7.12 buffers were made by adjusting pH of pH=7.64 buffer using 0.1 M HCl in H_2O.

pH/pD determination

pD/pH reading for all D_2O buffers and H_2O buffers were acquired on a pH meter (Mettler Toledo FiveEasy Plus FP20, with LE 410 sensor) at room temperature (~ 22 °C). The pH meter was calibrated with standard solutions (pH=4.01, 7.00, 10.01, Mettler Toledo) before measuring the custom-made buffers. As we discussed in the main manuscript, to avoid confusion, we show all the pD values in our manuscript from direct pH meter reading (i.e. the $pH_{meter\ reading}$) for all D_2O buffers. As shown below, the calibration factor is small and would not influence any of our conclusions.

The offset between $pH_{meter\ reading}$ and pD values was obtained by two control experiments. First, we compared the $pH_{meter\ reading}$ of DPBS-H_2O solution (7.35, considered as real pD if H_2O is replaced with D_2O) to that of DPBS-D_2O solution (7.6, considered as $pH_{meter\ reading}$). Second, we used the pH-sensitive ratiometric fluorophore BCECF to provide additional calibration between $pH_{meter\ reading}$ and pD values reported through BCECF ratios. In brief, we dissolved BCECF acid in water and made a 2 mM stock solution. We then diluted the stock BCECF solution in the high K^+ H_2O/D_2O buffers with adjusted $pH_{meter\ reading}$ (directly reading from the pH meter) to be in the range 6.8-8.1. The final BCECF concentration is 4 μM. We

then acquired fluorescence images using 445 nm and 488 nm excitation lasers (ZEISS LSM 980). The ratios of fluorescence intensity at 488 nm over that at 445 nm were plotted against respective readings from the pH meter for H_2O buffers and D_2O buffers (Figure 2.13a). The offset between linearly fitted ratio vs pH/pD curves is 0.25. The reading from the pH meter for D_2O buffers is shown as $pH_{meter\ reading}$ (and are reported as the pD values in our manuscript). The determined relationship is consistent with the above two different experiments: $pD=pH_{meter\ reading} - 0.25$.

DFT calculation

DFT calculations were performed using the Gaussian09 software. Structures were optimized and then characterized using frequency calculations at the B3LYP/6-311(G)++(d,p) level of theory.

Spontaneous Raman Spectroscopy

Spontaneous Raman spectra were acquired using an upright confocal Raman spectrometer (Horiba Raman microscope; Xplora plus). A 532 nm YAG laser was used to illuminate the sample with a power of 12 mW through a 100×, N.A. 0.9 objective (MPLAN N; Olympus). Data acquisition was performed with 10 s integration by the LabSpec6 software. For whole spectra recording, background was subtracted by measuring signal from the same solution without probe molecules. The spectra shown in Figure 2.1a and 2.2a are normalized to the alkyne peak.

EdU, EdC, and EU were dissolved into DMSO to make 100 mM stock solutions. PCho was dissolved into H_2O to make 2M stock solution. For measurement, EdU, EdC, and EU are 1:10 diluted into corresponding H_2O buffers or D_2O buffers, while PCho is 1:50 diluted into corresponding H_2O buffers or D_2O buffers.

Model molecules (4-fluorophenylacetylene (CAS# 766-98-3), methyl-4-ethynylbenzoate (CAS# 3034-86-4), 4-ethynylbenzaldehyde (CAS# 63697-96-1), 1-ethynyl-4-nitrobenzene (CAS# 937-31-5), and 4-Ethynylanisole (CAS# 768-60-5)) were dissolved into DMSO to make 100 mM stock solutions. For measurement in DMSO-D_2O system, model molecules were diluted into the 1:1 DMSO-D_2O (DPBS- D_2O, pD=7.6) solution to ensure good dissolvability for all model molecules with corresponding dilution factors. For measurement in methanol-OD system, model molecules were 1:10 diluted into methanol-OD (CAS# 1455-13-6).

All data are confirmed by at least three sets of independent experiments.

Stimulated Raman Scattering (SRS) Microscopy

A picoEmerald laser system (Applied Physics and Electronics) was used as the light source for SRS microscopy. Briefly, it produces 2 ps pump (tunable from 770 nm–990 nm, bandwidth 0.5 nm, spectral bandwidth ~ 7 cm^{-1}) and Stokes (1031.2 nm, spectral bandwidth 10 cm^{-1}) pulses with 80 MHz repetition rate. Stokes beam is modulated at 20 MHz by an internal electro-optic modulator. The spatially and temporally overlapped Pump and Stokes beams are introduced into an inverted laser-scanning microscope (FV3000, Olympus), and

then focused onto the sample by a 25X water objective (XLPLN25XWMP, 1.05 N.A., Olympus). Transmitted Pump and Stokes beams are collected by a high N.A. condenser lens (oil immersion, 1.4 N.A., Olympus) and pass through a bandpass filter (893/209 BrightLine, 25mm, Semrock) to filter out Stokes beam. A large area (10×10 mm) Si photodiode (S3590-09, Hamamatsu) is used to measure the pump beam intensity. A 64 V reverse-biased DC voltage is applied on the photodiode to increase the saturation threshold and reduce response time. The output current is terminated by a 50Ω terminator and pre-filtered by a 19.2-23.6-MHz band-pass filter (BBP-21.4+, MiniCircuits) to reduce laser and scanning noise. The signal is then demodulated by a lock-in amplifier (SR844, Stanford Research Systems) at the modulation frequency. The in-phase X output is fed back to the Olympus IO interface box (FV30-ANALOG) of the microscope. 30 μs time constant is set for the lock-in amplifier. Correspondingly, 80 μs pixel dwell time is used, which gives a speed of 21.3 s/frame for a 512-by-512-pixel image, with two frame-averaging. Laser powers are monitored throughout image acquisition by an internal power meter and power fluctuations are controlled within 1%. The power for the pump and Stocks beam is about 25 mW and 220 mW, respectively. 16-bit greyscale images were acquired by Olympus Fluoview 3000 software. To minimize the line-pattern issue likely due to an interfering Radio frequency (RF) picked up by our lock-in amplifier detection (demodulation at 20 MHz), we have optimized our alignment and replaced a few bandpass filters between our photodiode and lock-in amplifier.

For EdU and PCho measurement, the wavelengths of pump lasers for SRS_H are 845.9 and 844.6 nm, respectively. For ratiometric imaging of EdU and PCho after the exchange, the wavelengths of pump laser for SRS_D are 855.5 and 854.5 nm, respectively. Off-resonance

images were taken under 851.3 nm pump wavelength. For D_2O diffusion measurement, the pump wavelength is 820.5 nm. For EdU/EdU dimer spectra recordings, the wavelengths of pump lasers tuned from 843.9 to 847.9 nm with a 0.5 nm interval. For two-color ratiometric imaging of EdU and PCho during the exchange, the wavelengths of pump lasers for SRS_H and SRS_D are 845 and 855 nm, respectively.

Cell culture, sample preparation and alkyne-HDX in cells

For all SRS imaging experiments, cultured HeLa-CCL2 (ATCC) cells were seeded onto coverslips (12mm, #1.5, Fisher) with a density of 1×10^5/mL in 4 well plate with 0.3 mL DMEM culture medium (DMEM+10%FBS+1% penicillin-streptomycin) for 20 h at 37 °C and 5% CO_2. Prior to imaging, coverslips were collected and attached to a microscope slide (1mm thick, VWR) with an imaging spacer (0.12mm thick, Sigma-Aldrich).

For the EdU experiment, DMEM culture medium was then changed to DMEM medium (FBS-free, Gibco) for 20-22 h for cell cycle synchronization. After synchronization, the medium was replaced back to DMEM culture medium and EdU (10 mM stock in DPBS) was simultaneously added to a concentration of 100 μM for 20-24 h. Then 4% PFA was added for 20 min for fixation. After that, DPBS was used to wash away PFA and fixed cells could be stored in DPBS at 4 °C for several days.

For UV-irradiation on live cells, cells were put inside the biosafety cabinet (BSC) with UV (254 nm) on for one hour. Morphologies were quickly checked with no severe abnormality

under a transmission light microscope. The cells were fixed by 4% PFA immediately after the UV irradiation.

For all the fixed cells alkyne-HDX experiments, corresponding D_2O buffers were used to wash the cells three times and then the coverslip was taken out to make an imaging chamber filled with designating D_2O buffers for SRS imaging.

For live-cell BCECF experiments, cells were first loaded with 2 μM BCECF-AM in HBSS for 20 min in the CO_2 incubator and were subsequently treated with 10 μM nigericin in pH=7.35 high K^+ H_2O buffer for 5 min. Cells were switched into high K^+D_2O buffers containing 10 μM nigericin with designated pD and put onto the microscope slide. The image was collected on an inverted confocal microscope (ZEISS LSM 980) using 40x water immersion objective (NA 1.2) with either 445 or 488 nm excitation laser. Both images were acquired using the same PMT which was set to collect photons from 530-570 nm. Control experiments were done under similar conditions yet with the addition of DMSO instead of nigericin in all the buffer solutions.

For live-cells alkyne-HDX experiment for pD sensing, HeLa cells were first incubated with 10 μM nigericin in high K^+ H_2O buffer for 5 min at 37 ˚C. The high K^+ H_2O buffer was then removed, and cells were washed with high K^+ D_2O buffer with 10 μM nigericin of different pD values. We started timing as the cells were washed with high K^+ D_2O buffers. The coverslip with cells was placed onto a microscope slide with spacer filled with 10 μM nigericin high K^+ D_2O buffers. Control experiments were done under similar conditions but with the addition of DMSO instead of nigericin in all the buffer solutions.

For the PCho and EdU experiment, DMEM culture medium was changed to DMEM medium (FBS-free, Gibco) for synchronization. After synchronization, medium was replaced back to DMEM culture medium by simultaneously adding both propargylcholine (100 mM stock in DPBS) and EdU (10 mM stock in DPBS) to the culture medium with a final concentration of 1 mM and 100 μM, respectively, for 20-24 h. For the live cells alkyne-HDX experiments, home-made DMEM-D_2O buffer through dissolving DMEM powder into D_2O, was used to wash the cells three times. Then the coverslip was taken out to make an imaging chamber filled with the DMEM-D_2O buffer for SRS imaging.

Synthesis and purification of dsDNA and ssDNA

Pwo DNA polymerase, an enzyme showing good performance to accept modified nucleotides, was used to incorporate EdUTP into DNA through PCR process. A typical PCR reaction contained ~100 ng plasmid template, 0.05 mM each of the forward and reverse primers, 2.5 U polymerase and 10x polymerase buffer with magnesium, 100μM of dNTPs (dATP, dCTP, and dGTP, NEB) instead of dTTP, and 100μM modified EdUTP. The reactions were done in an overall volume of 50 μL with the addition of UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen). For ssDNA synthesis, asymmetric PCR was used. The concentration of the forward primer and the reverse primer is 0.001 mM and 0.05 mM, respectively, while other conditions are the same with that indicated above. PCR experiments were performed on a Bio-Rad C1000 thermal cycler.

Primers and the template employed for PCR experiments:

Forward primer: 5′- GGAAATCGGTACTGGCTTTCCATTCGAC

reverse primer: 3′- GTGAGTTAAAGTTGTACTCGAGTTTGTGTCCG

Template (sequence of 746 bp, contains 382 T):

GGAAATCGGTACTGGCTTTCCATTCGACCCCCATGATGGTTCCGTTCAACTAGC
AGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAG
ACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGATCCCAACGAAAAG
CGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGC
ATGGATGAGCTCTACAAAGGCGGTGGGTCGGGCGGGGGCTCCCCCGGGGGTG
GCGGTTCATGATCAGGTGGAGGGTCAGGGGGCGGATCAATGAGCAAAGGAGA
AGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAA
TGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGA
AAACTCACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCA
ACACTTGTCACTACTCTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGAT
CACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACA
GGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACGCGTGCTGAA
GTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGA
TTTTAAAGAAGATGGAAACATTCTCGGACACAAACTCGAGTACAACTTTAACT
CAC

PCR products were purified with a Monarch PCR & DNA Cleanup Kit (NEB) and confirmed by gel electrophoresis and Sanger sequencing. The products of PCR reaction were separated

by 1% agarose (UltraPure™ Low Melting Point Agarose, Invitrogen) gel electrophoresis. dsDNA and ssDNA were purified with a Monarch Gel Extraction Kit (NEB). Exonuclease I (Thermo Scientific, 20 U/μL) was used to digest ssDNA in 37℃, which was confirmed by gel electrophoresis as well. Before enzyme digestion, 5 min 95℃ heat shock was used for denaturation. The images were recorded with Bio-Rad Gel and Blot Imaging Systems. The combined PCR products (~50 tubes per sample) were loaded into Microcon-10kDa Centrifugal Filter (Millipore Sigma) and washed with UltraPure™ DNase/RNase-Free Distilled Water three times to remove the remaining salts and EDTA in the elusion buffer through buffer exchange. Then the PCR products in pure water were concentrated into ~1 uL through vacufuge (Vacufuge Plus Concentrator, Eppendorf). The concentrated PCR products were diluted into DPBS-D_2O / DPBS-D_2O + 2 M NaCl buffer solutions for alkyne-HDX kinetics measurements.

Formation and purification of EdU dimers

For solution samples, EdU (or thymine) was dissolved into water to make saturated solutions. After frozen into ice, the EdU (or thymine) solution was put into a homemade dry ice chamber and was irradiated with a UV lamp (254nm, UVLS-24 EL Series UV Lamp, 4 Watt) for 6 hours. Analytical HPLC coupled with mass spectrometry (LC-MS) for the UV-irradiated EdU product was performed on Agilent 1290 infinity LC system using ZORBAX RRHD Eclipse Plus C18, 95Å, 2.1 x 50 mm, 1.8 μm column with Agilent 6140 Series Quadrupole LCMS/LC-MS/MSD/Mass Spectrometer System. The mobile phase is water (0.1% AcOH) and acetonitrile with a running method of gradient 40% - 95% acetonitrile (1

ml/min, 10 min for total running time). The data shown in the Figure 2.11e) are the absorption (254 nm) intensity traces.

The aqueous reaction mixture (100 mg) post UV-radiation was first concentrated under reduced pressure at 40 ˚C and was then loaded on a reverse phase Biotage cartridge (12g SNAP Ultra C18). The flash column chromatography was automated by the Biotage Isolera System, with acetonitrile and water as the mobile phases. The flushing gradient was set to 5% to 40% acetonitrile over 10 column volume (CV) and the EdU dimer should be the second UV active portion eluted around the second to third CV. The collect portion was confirmed by normal phase TLC. EdU dimer would have an R_f value of 0.15 and EdU has an R_f of 0.6 when running with pure Ethyl Acetate. The purified EdU dimer was confirmed through high-resolution mass spectrometry (calculated exact mass: 505.1571, detected mass: 505.1575) and dissolved into H_2O to prepare a stock solution for further alkyne-HDX measurements.

Immunofluorescence staining

The UV-irradiated cells and control cells were first fixed and treated with 0.2% Triton-X-100 in 1x DPBS (no calcium, no magnesium, Gibco) at room temperature for 30 min. Then the Triton-X-100 was removed, and cells were washed with 1x DPBST (DPBS + 0.1% Tween 20) three times. Cells were incubated with 1% BSA, 22.52 mg/mL glycine in PBST at room temperature for 60 min for blocking. After washed with PBST, cells were incubated with 1:300 diluted anti-thymine dimer antibody (Mouse monoclonal, ~2 mg/mL, MilliporeSigma) in 0.1% BSA, PBST overnight at 4°C. The primary antibody solution was

removed, and cells were washed three times with DPBST. Then cells were incubated with 3% BSA in PBST at room temperature for 60 min for blocking. After washed with PBST, cells were incubated with 1:400 diluted Goat anti-Mouse IgG (H+L), (Invitrogen, Superclonal™ Recombinant Secondary Antibody, Alexa Fluor 647, 1 mg/mL) in 0.1% BSA, PBST for 2 hours at room temperature. The secondary antibody solution was removed, and cells were washed three times with DPBST. Fluorescent imaging was conducted immediately after sample preparation through the same Olympus FV3000 confocal microscope with CW laser excitation (640 nm, Coherent OBIS LX laser) and standard bandpass filter sets.

RNA *in vitro* transcription and phase separation

The RNA was synthesized via *in vitro* transcription with the corresponding DNA template using HiScribe® T7 High Yield RNA Synthesis Kit (NEB) and purified by Monarch® Kits for RNA Cleanup following the manufacturer's protocol. The template DNA for 47×CAG with flank sequences (47× RNA sequence: GGGAGACCCAAGCUGGCUAGUUAA GCUUGGUACCGGGCCCCCCCUCGAGGUCGACCGCCAUGGCGACCCUGGAAA AGCUGAUGAAGGCGGUCUCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC AGCAGCAGAGACGGAGCUCACUUCCUCAACCUCCUCCACAGGCACAGCC) was purchased from Integrated DNA Technologies (IDT) as single-stranded DNA. The double-strand DNA with flank sequences was generated by RCP reaction with Phusion® High-

Fidelity DNA Polymerase (NEB). The 20×CAG RNA (CAGCAGCAGCAG CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG) was directly purchased from Integrated DNA Technologies (IDT) as single-stranded RNA.

For phase separation assays, 47×CAG RNAs were diluted to concentrations of 500 ng/μl in 10 mM $MgCl_2$, 25 mM NaCl, 10 mM Tris, pH=7.0 buffer with UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen). RNA was denatured at 95°C for 3min and cooled down at 1°C per minute to 37 °C final temperature in a thermocycler. Despite the shorter CAG repeat region, 20×CAG RNA exhibited the ability to form RNA condensates at high RNA concentration (>9 uM) in buffer with a high concentration of Mg^{2+} (35 mM $MgCl_2$, 10 mM Tris, pH=7.0 buffer). The RNA sample was sealed inside an imaging spacer (0.12mm thick, Sigma-Aldrich) with a coverslip and a microscope slide (1mm thick, VWR) and immediately imaged by the SRS microscope.

Statistical analysis

All the statistical analyses were done using Origin software. All data are shown as means ± SEM. One-way analyses of variance (ANOVAs) followed by Tukey's multiple comparisons test were used for single-variable comparisons. The statistical significance level was defined as: ns not significant, * $p<0.05$, ** $p < 0.01$, *** $p < 0.001$.

Data processing

All spectra were processed using LabSpec6 software. Spectral baselines were subtracted. Peak centers and intensity were read out by Gaussian peak fitting. All images were processed using ImageJ software. Corresponding off-resonance images were subtracted.

For precise $t_{1/2}$ fitting, we obtained the time-zero intensity (i.e. I_0(R-C≡C-H)) by the normalization of I_1(R-C≡C-H) and I_1(R-C≡C-D), the intensity measured from the first exchange data point at the alkyne-H and the alkyne-D channels, respectively. I_0(R-C≡C-H) is defined as I_1(R-C≡C-H)+I_1(R-C≡C-D)/r, in which r is the intensity correction ratio, defined as dividing I(R-C≡C-D), the alkyne-probe solution after equilibrium in D_2O (i.e. finished with exchange), by I(R-C≡C-H), the corresponding alkyne-probe solution in H_2O with the same concentration without any exchange. This strategy was used for both the spontaneous Raman measurement of solution samples and the SRS imaging recording of cell samples. I(R-C≡C-H) and I(R-C≡C-D) from spontaneous Raman measurements were read out by Gaussian peak fitting. I(R-C≡C-H) and I(R-C≡C-D) from SRS measurements (also refer as SRS_H and SRS_D) were read out from all the nucleus regions in one field of view based on SRS images.

Derivation of alkyne-HDX kinetics

The rate-determining step (RDS) in HDX between alkyne-tagged probes (R-C≡C-H) and catalytic base OD^- is shown in Eq. (2.1) below. It includes three elementary steps: a) diffusional collision, b) equilibrium redistribution of the hydrogen in the intermediate state, and c) dissociation.

$$R-C\equiv C-H+OD^- \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} R-C\equiv C-H\ldots O-D^- \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} R-C\equiv C^-\ldots H-O-D \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} R-C\equiv C^-+HOD$$

(2.1)

We can assume that 1) the temperature remains constant, 2) the reaction is diffusion-limited, and 3) the intermediates are in steady states.[43] So the concentrations of the two intermediates remain the same in the reaction, shown in Eq. (2.5).

$$\frac{d[(R-C\equiv C-H\cdots OD)^-]}{dt}=\frac{d[(R-C\equiv C\cdots HOD)^-]}{dt}=0 \qquad (2.5)$$

i.e.

$$k_1[R-C\equiv C-H][OD^-]-k_{-1}[(R-C\equiv C-H\cdots OD)^-]+k_{-2}[(R-C\equiv C\cdots HOD)^-]-k_2[(R-C\equiv C-H\cdots OD)^-]=0$$

$$k_2[(R-C\equiv C-H\cdots OD)^-]-k_{-2}[(R-\equiv C\cdots HOD)^-]-k_3[(R-C\equiv C\cdots HOD)^-]=0$$

So we can get the concentration of the intermediate shown in Eq. (2.6).

$$[(R-C\equiv C\cdots HOD)^-]=\frac{k_1k_2}{k_{-1}k_{-2}+k_{-1}k_3+k_2k_3}[R-C\equiv C-H][OD^-] \qquad (2.6)$$

The overall rate (k) for transferring a proton from R-C≡C-H to OD^- then becomes:

$$k=\frac{d[R-C\equiv C^-]}{dt}=k_3[(R-C\equiv C\cdots HOD)^-]=\frac{k_1k_2k_3}{k_{-1}k_{-2}+k_{-1}k_3+k_2k_3}[R-C\equiv CH][OD^-]$$

(2.7)

Since the reaction is diffusion-limited, there are some restrictions on rate constants:

$$k_{-1} \ll k_2 \text{ or } k_3 \ll k_{-2}, \qquad k_{-1} \approx k_3$$

So the overall rate can be simplified as Eq. (2.8).

$$k = \frac{k_1}{\frac{k_{-2}}{k_2}+1}[R-C\equiv C-H][OD^-] \tag{2.8}$$

In addition, $\frac{k_{-2}}{k_2} = 10^{pK_a(R-C\equiv C-H)-pK_a(HOD)} \gg 1$, so Eq. (2.8) can be further simplified as Eq. (2.2).

$$k = k_1 \cdot (10^{pK_a(\,donor,\ R-C\equiv C-H)-pK_a(acceptor,\ HOD)} + 1)^{-1} \cdot [OD^-] \cdot [R-C\equiv CH] \tag{2.2}$$

k_1 is the diffusion-limited collision constant, upper-bounded by 10^{10} $M^{-1}s^{-1}$. As fluctuation of OD^- concentration is negligible during HDX, the RDS is considered as a pseudo-first-order reaction. Since the acceptor pK_a is much smaller than the donor pK_a, the corresponding exchange half-life ($t_{1/2}$) is reduced to Eq. (2.9):

$$\lg(t_{1/2}) = pK_a - pD + lg\,(ln2) - 10 - pK_a\,(HOD) + pK_a(D_2O) \tag{2.9}$$

Here pK_a designates the donor pK_a, i.e. $pK_a(R-C\equiv C-H)$. Taking the pK_a of alkynyl hydrogen as 20-25 and the pK_a of HOD as ~15, we estimated the $t_{1/2}$ of alkyne-HDX in physiological pD (7.6) to be on the order of minutes. For example, if we used 20 for pK_a of alkynyl hydrogen, the calculated $t_{1/2}$ could be 174 s.

2.15 References

1.Yamakoshi, H.; Dodo, K.; Palonpon, A.; Ando, J.; Fujita, K.; Kawata, S.; Sodeoka, M., Alkyne-tag Raman imaging for visualization of mobile small molecules in live cells. *J. Am. Chem. Soc.* **2012,** *134* (51), 20681-20689.
2.Hong, S.; Chen, T.; Zhu, Y.; Li, A.; Huang, Y.; Chen, X., Live-cell stimulated Raman scattering imaging of alkyne-tagged biomolecules. *Angew. Chem. Int. Ed.* **2014,** *53* (23), 5827-5831.
3.Wei, L.; Hu, F.; Shen, Y.; Chen, Z.; Yu, Y.; Lin, C.-C.; Wang, M. C.; Min, W., Live-cell imaging of alkyne-tagged small biomolecules by stimulated Raman scattering. *Nat. Methods* **2014,** *11* (4), 410-412.
4.Ando, J.; Asanuma, M.; Dodo, K.; Yamakoshi, H.; Kawata, S.; Fujita, K.; Sodeoka, M., Alkyne-Tag SERS Screening and Identification of Small-Molecule-Binding Sites in Protein. *J. Am. Chem. Soc.* **2016,** *138* (42), 13901-13910.
5.Chen, Y.; Ren, J. Q.; Zhang, X. G.; Wu, D. Y.; Shen, A. G.; Hu, J. M., Alkyne-Modulated Surface-Enhanced Raman Scattering-Palette for Optical Interference-Free and Multiplex Cellular Imaging. *Anal. Chem.* **2016,** *88* (12), 6115-6119.
6.Prince, R. C.; Frontiera, R. R.; Potma, E. O., Stimulated Raman Scattering: From Bulk to Nano. *Chem. Rev.* **2017,** *117* (7), 5070-5094.
7.Jamieson, L. E.; Greaves, J.; McLellan, J. A.; Munro, K. R.; Tomkinson, N. C. O.; Chamberlain, L. H.; Faulds, K.; Graham, D., Tracking intracellular uptake and localisation of alkyne tagged fatty acids using Raman spectroscopy. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2018,** *197*, 30-36.
8.Shen, Y.; Hu, F.; Min, W., Raman Imaging of Small Biomolecules. *Annu. Rev. Biophys.* **2019,** *48*, 347-369.
9.Hu, F.; Chen, Z.; Zhang, L.; Shen, Y.; Wei, L.; Min, W., Vibrational Imaging of Glucose Uptake Activity in Live Cells and Tissues by Stimulated Raman Scattering. *Angew. Chem. Int. Ed.* **2015,** *54* (34), 9821-9825.
10.Lee, H. J.; Zhang, W.; Zhang, D.; Yang, Y.; Liu, B.; Barker, E. L.; Buhman, K. K.; Slipchenko, L. V.; Dai, M.; Cheng, J. X., Assessing cholesterol storage in live cells and C. elegans by stimulated Raman scattering imaging of phenyl-Diyne cholesterol. *Sci. Rep.* **2015,** *5* (7930), 1-10.
11.Tipping, W. J.; Lee, M.; Serrels, A.; Brunton, V. G.; Hulme, A. N., Imaging drug uptake by bioorthogonal stimulated Raman scattering microscopy. *Chem. Sci.* **2017,** *8* (8), 5606-5615.
12.Hu, F.; Shi, L.; Min, W., Biological imaging of chemical bonds by stimulated Raman scattering microscopy. *Nat. Methods* **2019,** *16* (9), 830-842.
13.Seidel, J.; Miao, Y.; Porterfield, W.; Cai, W.; Zhu, X.; Kim, S. J.; Hu, F.; Bhattarai-Kline, S.; Min, W.; Zhang, W., Structure-activity-distribution relationship study of anti-cancer antimycin-type depsipeptides. *Chem. Commun.* **2019,** *55* (63), 9379-9382.
14.Zeng, C.; Hu, F.; Long, R.; Min, W., A ratiometric Raman probe for live-cell imaging of hydrogen sulfide in mitochondria by stimulated Raman scattering. *Analyst* **2018,** *143* (20), 4844-4848.
15.Tipping, W. J.; Wilson, L. T.; Blaseio, S. K.; Tomkinson, N. C. O.; Faulds, K.; Graham, D., Ratiometric sensing of fluoride ions using Raman spectroscopy. *Chem. Commun.* **2020,** *56* (92), 14463-14466.

16.Wilson, L. T.; Tipping, W. J.; Jamieson, L. E.; Wetherill, C.; Henley, Z.; Faulds, K.; Graham, D.; Mackay, S. P.; Tomkinson, N. C. O., A new class of ratiometric small molecule intracellular pH sensors for Raman microscopy. *Analyst* **2020,** *145* (15), 5289-5298.
17.Talley, C. E.; Jusinski, L.; Hollars, C. W.; Lane, S. M.; Huser, T., Intracellular pH sensors based on surface-enhanced Raman scattering. *Anal. Chem.* **2004,** *76* (23), 7064-7068.
18.Qin, X.; Lyu, M.; Si, Y.; Yang, J.; Wu, Z.; Li, J., Alkyne-based surface-enhanced Raman scattering nanoprobe for ratiometric imaging analysis of caspase 3 in live cells and tissues. *Anal. Chim. Acta* **2018,** *1043*, 115-122.
19.Si, Y.; Bai, Y.; Qin, X.; Li, J.; Zhong, W.; Xiao, Z.; Li, J.; Yin, Y., Alkyne-DNA-Functionalized Alloyed Au/Ag Nanospheres for Ratiometric Surface-Enhanced Raman Scattering Imaging Assay of Endonuclease Activity in Live Cells. *Anal. Chem.* **2018,** *90* (6), 3898-3905.
20.Fried, S. D.; Boxer, S. G., Measuring electric fields and noncovalent interactions using the vibrational stark effect. *Acc. Chem. Res.* **2015,** *48* (4), 998-1006.
21.Marr, J. M.; Schultz, Z. D., Imaging Electric Fields in SERS and TERS Using the Vibrational Stark Effect. *J. Phys. Chem. Lett.* **2013,** *4* (19), 3268-3272.
22.Epstein, S. C.; Huff, A. R.; Winesett, E. S.; Londergan, C. H.; Charkoudian, L. K., Tracking carrier protein motions with Raman spectroscopy. *Nat. Commun.* **2019,** *10* (2227), 1-6.
23.Flynn, J. D.; Gimmen, M. Y.; Dean, D. N.; Lacy, S. M.; Lee, J. C., Terminal Alkynes as Raman Probes of alpha-Synuclein in Solution and in Cells. *Chembiochem* **2020,** *21* (11), 1582-1586.
24.Craven, C. J.; Derix, N. M.; Hendriks, J.; Boelens, R.; Hellingwerf, K. J.; Kaptein, R., Probing the nature of the blue-shifted intermediate of photoactive yellow protein in solation by NMR: Hydrogen-deuterium exchange data and pH studies. *Biochemistry* **2000,** *39* (47), 14392-14399.
25.Yagi-Utsumi, M.; Chandak, M. S.; Yanaka, S.; Hiranyakorn, M.; Nakamura, T.; Kato, K.; Kuwajima, K., Residual Structure of Unfolded Ubiquitin as Revealed by Hydrogen/Deuterium-Exchange 2D NMR. *Biophys. J.* **2020,** *119* (10), 2029-2038.
26.Englander, J. J.; Del Mar, C.; Li, W.; Englander, S. W.; Kim, J. S.; Stranz, D. D.; Hamuro, Y.; Woods, V. L., Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2003,** *100* (12), 7057-7062.
27.Hoofnagle, A. N.; Resing, K. A.; Ahn, N. G., Protein analysis by hydrogen exchange mass spectrometry. *Annu. Rev. Biophys. Biomol. Struct.* **2003,** *32*, 1-25.
28.Konermann, L.; Pan, J.; Liu, Y. H., Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **2011,** *40* (3), 1224-1234.
29.Jansson, E. T.; Lai, Y. H.; Santiago, J. G.; Zare, R. N., Rapid Hydrogen-Deuterium Exchange in Liquid Droplets. *J. Am. Chem. Soc.* **2017,** *139* (20), 6851-6854.
30.Vigano, C.; Smeyers, M.; Raussens, V.; Scheirlinckx, F.; Ruysschaert, J. M.; Goormaghtigh, E., Hydrogen-deuterium exchange in membrane proteins monitored by IR spectroscopy: a new tool to resolve protein structure and dynamics. *Biopolymers* **2004,** *74* (1-2), 19-26.

31.DeFlores, L. P.; Tokmakoff, A., Water penetration into protein secondary structure revealed by hydrogen-deuterium exchange two-dimensional infrared spectroscopy. *J. Am. Chem. Soc.* **2006,** *128* (51), 16520-16521.
32.Brielle, E. S.; Arkin, I. T., Site-Specific Hydrogen Exchange in a Membrane Environment Analyzed by Infrared Spectroscopy. *J. Phys. Chem. Lett.* **2018,** *9* (14), 4059-4065.
33.Smith, J. R.; Cicerone, M. T.; Meuse, C. W., Measuring hydrogen-deuterium exchange in protein monolayers. *Surf. Interface Anal.* **2009,** *41* (11), 878-885.
34.Dunkelberger, E. B.; Woys, A. M.; Zanni, M. T., 2D IR cross peaks reveal hydrogen-deuterium exchange with single residue specificity. *J. Phys. Chem. B.* **2013,** *117* (49), 15297-15305.
35.Doukas, A. G.; Pande, A.; Suzuki, T.; Callender, R. H.; Honig, B.; Ottolenghi, M., ON THE MECHANISM OF HYDROGEN-DEUTERIUM EXCHANGE IN BACTERIORHODOPSIN. *Biophys. J.* **1981,** *33* (2), 275-279.
36.Deng, H.; Huang, L. W.; Callender, R.; Ebrey, T., Evidence For a Bound Water Molecule Next to the Retinal Schiff Base in Bacteriorhodopsin and Rhodopsin: A Resonance Raman Study of the Schiff Base Hydrogen/Deuterium Exchange. *Biophys. J.* **1994,** *66* (4), 1129-1136.
37.Shashilov, V.; Xu, M.; Makarava, N.; Savtchenko, R.; Baskakov, I. V.; Lednev, I. K., Dissecting structure of prion amyloid fibrils by hydrogen-deuterium exchange ultraviolet Raman spectroscopy. *J. Phys. Chem. B.* **2012,** *116* (27), 7926-7930.
38.Streitwi.A; Reuben, D. M. E., Acidity of hydrocarbons. XXXV. Equilibrium acidities of phenylacetylene and tert-butylacetylene in cyclohexylamine. *J. Am. Chem. Soc.* **1971,** *93* (7), 1794-1795.
39.Yang, Q.; Li, Y.; Yang, J. D.; Liu, Y.; Zhang, L.; Luo, S.; Cheng, J. P., Holistic Prediction of the pKa in Diverse Solvents Based on a Machine-Learning Approach. *Angew. Chem. Int. Ed.* **2020,** *59* (43), 19282-19291.
40.Chen, Z.; Paley, D. W.; Wei, L.; Weisman, A. L.; Friesner, R. A.; Nuckolls, C.; Min, W., Multicolor live-cell chemical imaging by isotopically edited alkyne vibrational palette. *J. Am. Chem. Soc.* **2014,** *136* (22), 8027-8033.
41.Bell, R. A.; Chisholm, M. H., Reactions of alkynyl- and alkenylplatinum(II) compounds. 2. Study of the reaction between anhydrous hydrogen chloride and diethynylbis (dimethylphenylphosphine) platinum(II). Evidence for platinum-stabilized vinyl cations. *Inorg. Chem.* **1976,** *16* (3), 687-697.
42.Egoshi, S.; Dodo, K.; Ohgane, K.; Sodeoka, M., Deuteration of terminal alkynes realizes simultaneous live cell Raman imaging of similar alkyne-tagged biomolecules. *Org. Biomol. Chem.* **2021,** *19* (38), 8232-8236.
43.Eigen, M., Proton Transfer, Acid-Base Catalysis, and Enzymatic Hydrolysis. *Angew. Chem. Int. Ed.* **1964,** *3* (1), 1-19.
44.Englander, S. W.; Downer, N. W.; Teitelbaum, H., HYDROGEN-EXCHANGE. *Annu. Rev. Biochem.* **1972,** *41*, 903-924.
45.Englander, S. W.; Mayne, L.; Bai, Y.; Sosnick, T. R., Hydrogen exchange: The modern legacy of Linderstrom-Lang. *Protein Sci.* **1997,** *6* (5), 1101-1109.
46.Krezel, A.; Bal, W., A formula for correlating pKa values determined in D2O and H2O. *J. Inorg. Biochem.* **2004,** *98* (1), 161-166.

47.Qu, D.; Wang, G.; Wang, Z.; Zhou, L.; Chi, W.; Cong, S.; Ren, X.; Liang, P.; Zhang, B., 5-Ethynyl-2'-deoxycytidine as a new agent for DNA labeling: detection of proliferating cells. *Anal. Biochem.* **2011,** *417* (1), 112-121.
48.Hansch, C.; Leo, A.; Taft, R. W., A survey of Hammett substituent constants and resonance and field parameters. *Chem. Rev.* **1991,** *91* (2), 165-195.
49.Dempsey, C. E., pH dependence of hydrogen exchange from backbone peptide amides in apamin. *Biochemistry* **1986,** *25* (13), 3904-3911.
50.Houliston, R. S.; Liu, C. S.; Singh, L. M. R.; Meiering, E. M., pH and urea dependence of amide hydrogen-deuterium exchange rates in the beta-trefoil protein hisactophilin. *Biochemistry* **2002,** *41* (4), 1182-1194.
51. JE, T., Intracellular diffusion of water. *Arch. Biochem. Biophys.* **1983,** *224* (2), 416-428.
52.Mills, R., Self-diffusion in normal and heavy water in the range 1 to 450. *J. Phys. Chem. C* **1973,** *77* (5), 685-688.
53.Shi, L.; Hu, F.; Min, W., Optical mapping of biological water in single live cells by stimulated Raman excited fluorescence microscopy. *Nat. Commun.* **2019,** *10* (4764), 1-8.
54.Zwang, T. J.; Tse, E. C. M.; Barton, J. K., Sensing DNA through DNA Charge Transport. *ACS Chem. Biol.* **2018,** *13* (7), 1799-1809.
55.Maeshima, K.; Imai, R.; Tamura, S.; Nozaki, T., Chromatin as dynamic 10-nm fibers. *Chromosoma* **2014,** *123* (3), 225-237.
56.Fantoni, N. Z.; El-Sagheer, A. H.; Brown, T., A Hitchhiker's Guide to Click-Chemistry with Nucleic Acids. *Chem. Rev.* **2021,** *121* (12), 7122-7154.
57.Burley, G. A.; Gierlich, J.; Mofid, M. R.; Nir, H.; Tal, S.; Eichen, Y.; Carell, T., Directed DNA Metallization. *J. Am. Chem. Soc.* **2006,** *128* (5), 1398-1399.
58.Veneziano, R.; Shepherd, T. R.; Ratanalert, S.; Bellou, L.; Tao, C.; Bathe, M., In vitro synthesis of gene-length single-stranded DNA. *Sci. Rep.* **2018,** *8* (1), 6548.
59.Maity, A.; Singh, A.; Singh, N., Differential stability of DNA based on salt concentration. *Eur. Biophys. J.* **2017,** *46* (1), 33-40.
60.Baillargeon, P.; Lussier, T.; Dory, Y. L., Hydrogen Bonds between Acidic Protons from Alkynes (C–H···O) and Amides (N–H···O) and Carbonyl Oxygen Atoms as Acceptor Partners. *J. Appl. Crystallogr.* **2014,** *2014*, 1-5.
61.Park, H.; Zhang, K. J.; Ren, Y. J.; Nadji, S.; Sinha, N.; Taylor, J. S.; Kang, C. H., Crystal structure of a DNA decamer containing a cis-syn thymine dimer. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99* (25), 15965-15970.
62.Nagpal, A.; Dhankhar, D.; Cesario, T. C.; Li, R.; Chen, J.; Rentzepis, P. M., Thymine dissociation and dimer formation: A Raman and synchronous fluorescence spectroscopic study. *Proc. Natl. Acad. Sci. U. S. A.* **2021,** *118* (6), 1-7.
63.Rastogi, R. P.; Richa; Kumar, A.; Tyagi, M. B.; Sinha, R. P., Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *J. Nucleic Acids* **2010,** *2010*, 1-32.
64.Sinha, R. P.; Hader, D. P., UV-induced DNA damage and repair: a review. *Photochem. Photobiol. Sci.* **2002,** *1* (4), 225-236.
65.Douki, T.; Court, M.; Sauvaigo, S.; Odin, F.; Cadet, J., Formation of the main UV-induced thymine dimeric lesions within isolated and cellular DNA as measured by high performance liquid chromatography-tandem mass spectrometry. *J. Biol. Chem.* **2000,** *275* (16), 11678-11685.

66.DeRosa, M. C.; Sancar, A.; Barton, J. K., Electrically monitoring DNA repair by photolyase. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102* (31), 10788-10792.
67.Huber, V.; Camisaschi, C.; Berzi, A.; Ferro, S.; Lugini, L.; Triulzi, T.; Tuccitto, A.; Tagliabue, E.; Castelli, C.; Rivoltini, L., Cancer acidity: An ultimate frontier of tumor immune escape and a novel target of immunomodulation. *Semin. Cancer Biol.* **2017,** *43*, 74-89.
68.Ozkan, P.; Mutharasan, R., A rapid method for measuring intracellular pH using BCECF-AM. *Biochim. Biophys. Acta* **2002,** *1572* (1), 143-148.
69.Bassnett S, R. L., Beebe DC, Intracellular pH measurement using single excitation-dual emission fluorescence ratios. *Am. J. Physiol.* **1990,** *258* (1), C171-C178.
70.Paabo, M.; Bates, R. G., Standards for a practical scale of pD in heavy water. *Anal. Chem* **1969,** *41* (2), 283-285.
71.Cerqueira, A. V.; Lemos, B., Ribosomal DNA and the Nucleolus as Keystones of Nuclear Architecture, Organization, and Function. *Trends Genet.* **2019,** *35* (10), 710-723.
72.Lafontaine, D. L. J.; Riback, J. A.; Bascetin, R.; Brangwynne, C. P., The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell. Biol.* **2021,** *22* (3), 165-182.
73.Mirkin, S. M., Expandable DNA repeats and human disease. *Nature* **2007,** *447* (7147), 932-40.
74.Fan, H. C.; Ho, L. I.; Chi, C. S.; Chen, S. J.; Peng, G. S.; Chan, T. M.; Lin, S. Z.; Harn, H. J., Polyglutamine (PolyQ) diseases: genetics to treatments. *Cell Transplant* **2014,** *23* (4-5), 441-58.
75.Malik, I.; Kelley, C. P.; Wang, E. T.; Todd, P. K., Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat Rev Mol Cell Biol* **2021,** *22* (9), 589-607.
76.Lieberman, A. P.; Shakkottai, V. G.; Albin, R. L., Polyglutamine repeats in neurodegenerative diseases. *Annu Rev Pathol* **2019,** *14*, 1-27.
77.Jain, A.; Vale, R. D., RNA phase transitions in repeat expansion disorders. *Nature* **2017,** *546* (7657), 243-247.
78.Ma, Y.; Li, H.; Gong, Z.; Yang, S.; Wang, P.; Tang, C., Nucleobase Clustering Contributes to the Formation and Hollowing of Repeat-Expansion RNA Condensate. *J Am Chem Soc* **2022,** *144* (11), 4716-4720.

Chapter 3

GFP-FREE LIVE-NEURON QUANTITATIVE IMAGING REVEALS COMPARTMENTALIZATION AND GROWTH DYNAMICS OF POLYQ AGGREGATES

3.1 Summary

Huntington's Disease (HD), the most prevalent polyglutamine (polyQ) neurodegenerative disorder, features brain aggregates induced by mutant huntingtin (mHtt) proteins harboring expanded polyQ tracts. Despite extensive efforts, molecular mechanisms of polyQ aggregates remain elusive. Here we establish quantitative stimulated Raman scattering imaging of polyQ aggregates (q-aggSRS) for non-invasive investigations in live neuronal co-cultures using deuterated glutamine labeling. Q-aggSRS allows for specific visualization by targeting the distinct Raman peak from carbon-deuterium bonds, eliminating the need for bulky GFP tagging. Coupled with analysis from aggregate-tailored expansion microscopy, newly designed two-color imaging, and pulse-chase visualization, we comprehensively quantified the mHtt and non-mHtt proteins within the same aggregates across varying sizes, cell types, mHtt constructs, and subcellular locations. Our findings demonstrate a two-phase aggregate growth model with a distinct core-shell spatial organization, reveal significant heterogeneity in nucleus/cytoplasm compartmentalization, and identify previously unrecognized "gel-like" aggregates specifically in neuronal nuclei. These insights should

advance our understanding of native polyQ aggregates and our proposed interaction coefficients may offer new quantitative parameters for developing effective HD therapies.

3.2 Introduction

Polyglutamine (polyQ) diseases represent a group of genetic neurodegenerative disorders that cause cognitive and physical impairments, including spinocerebellar ataxias, Huntington's Disease (HD), spinal and bulbar muscular atrophy (SMBA), and dentatorubral-pallidoluysian atrophy (DRPLA).[1-3] These diseases are characterized by the pathological expansion of CAG trinucleotide repeats, which encode polyQ tracts that exceed a critical threshold of approximately 35 repeats,[4-5] leading to the formation of aggregates mainly in the brain. Among this family, HD is the most prevalent and widely studied, resulting from a genetic mutation in the HTT gene that induces the aggregation of mutant huntingtin (mHtt) proteins. Over the decades, extensive investigations have provided various fundamental insights into the potential mechanism of mHtt aggregation and its molecular composition.[6-9] Despite these advances, there remains no effective cure for HD. Moreover, significant phenomena, such as the predominance of nuclear aggregates in younger patients and cytoplasmic aggregates in older patients, remain poorly understood at the cellular level.[1, 8, 10] Recent research suggests that soluble mHtt oligomers may be more toxic than the aggregates, with aggregation potentially acting as a protective mechanism to sequester soluble mHtt proteins rather than directly causing cellular toxicity.[5, 9, 11-14] Nevertheless, such a mechanistic picture remains controversial, and our understanding of the structural

composition and pathological roles of polyQ aggregates, particularly in the native states within neuronal environments, is limited.

Traditional biochemistry methods with *in vitro* models[15] and emerging mass spectroscopy analysis from purified aggregates offer detailed molecular views. However, these approaches involve extensive postprocessing steps such as cell lysis, centrifugation, washing, and extraction, which can introduce severe sample perturbations.[12, 16] In addition, these studies rarely employ neuronal models, relying instead on non-neuronal systems such as yeast, [14, 17-19] E. Coli,[20] or mammalian cell lines,[11-12, 16, 21-22] which may lead to complications in interpretation as they fail to recapitulate the neuronal environment relevant to polyQ diseases. Fluorescence microscopy allows unprecedented dynamic analysis of polyQ aggregates in live cells with high spatial resolution.[21, 23-24] However, the common use of large fluorescent proteins such as EGFP in fusion with mHtt can disrupt its native aggregation behavior, potentially altering the original mechanism. This issue is especially critical for the widely studied disease-causing mHtt-exon1 peptide model, [25] an aberrantly spliced variant of mHtt that is less than half the size of EGFP. With the emergence of cryo-electron tomography (Cryo-ET), it is now feasible to perform ultrastructural analysis of native polyQ aggregates in unperturbed cellular environments without EGFP tagging.[26-27] However, Cryo-ET is limited by its low-throughput, lack of quantification, and inability to capture dynamic information in live cells. An effective and high-throughput platform that allows quantitative examination of native aggregates in live and physiological-relevant models would hence be invaluable in providing new and unbiased perspectives into the previously “invisible”

phenomenon of polyQ aggregates and accelerating the development of promising therapies for polyQ diseases.[2, 28-30]

Stimulated Raman scattering (SRS) microscopy, a nonlinear Raman imaging modality, has been proven to be a non-invasive optical technique for live-cell and tissue imaging with high spatial resolution.[31-34] Recently, we introduced an effective platform integrating deuterated glutamine (Gln-d_5) labeling with SRS imaging for highly specific and quantitative analysis of polyQ aggregates in live HeLa cells, eliminating the need for EGFP tagging.[35] We demonstrated significant compositional changes caused by EGFP tagging in the live-cell context and discovered an unexpectedly high concentration of non-mHtt proteins sequestered in the aggregates.[35] However, similar to many other studies relying on non-neuronal systems, concerns arise regarding whether the observed aggregation processes fully recapitulate those occurring in live nervous systems which have a much more complex cellular environment. Not only is the intracellular environment of neurons distinct from other cell types, but their communication with glia might also play crucial roles in the formation, propagation, and potential clearance of polyQ aggregates,[36-37] making it essential to conduct studies in a native system.

In this work, we established a quantitative SRS imaging platform of mHtt aggregates (q-aggSRS) in live primary hippocampal neuronal co-cultures through coupled AAV transduction and Gln-d_5 metabolic incorporation to trace the aggregation formation, compartmentalization, dynamics, and composition for polyQ aggregates (Figure 3.1a). By targeting the carbon-deuterium (CD) vibrational frequency in the desired cell-silent region

(1800-2600 cm^{-1}), SRS allowed for specific visualization of polyQ aggregates labeled by Gln-d_5 in neuronal cells. This unique atom-labeling strategy via stable isotope replacement of carbon-hydrogen (CH) bonds with CD bonds is highly non-perturbative with minimal size difference and maintains the integrity of the original hydrogen bonding network. The linear relationship between SRS signals and bond concentrations coupled with the known mHtt protein sequence uniquely allowed us to calculate absolute concentrations for both mHtt and non-mHtt proteins (c(mHtt) and c(non-mHtt)) within the same aggregates in live neurons. We comprehensively examined the influence of varied mHtt constructs, including those with and without EGFP labeling, varying polyQ tract lengths, and the deletion of the proline-rich region (Figure 3.1b). Our findings also confirmed that EGFP tagging causes more severe perturbations to polyQ aggregates in neurons compared to HeLa cells.

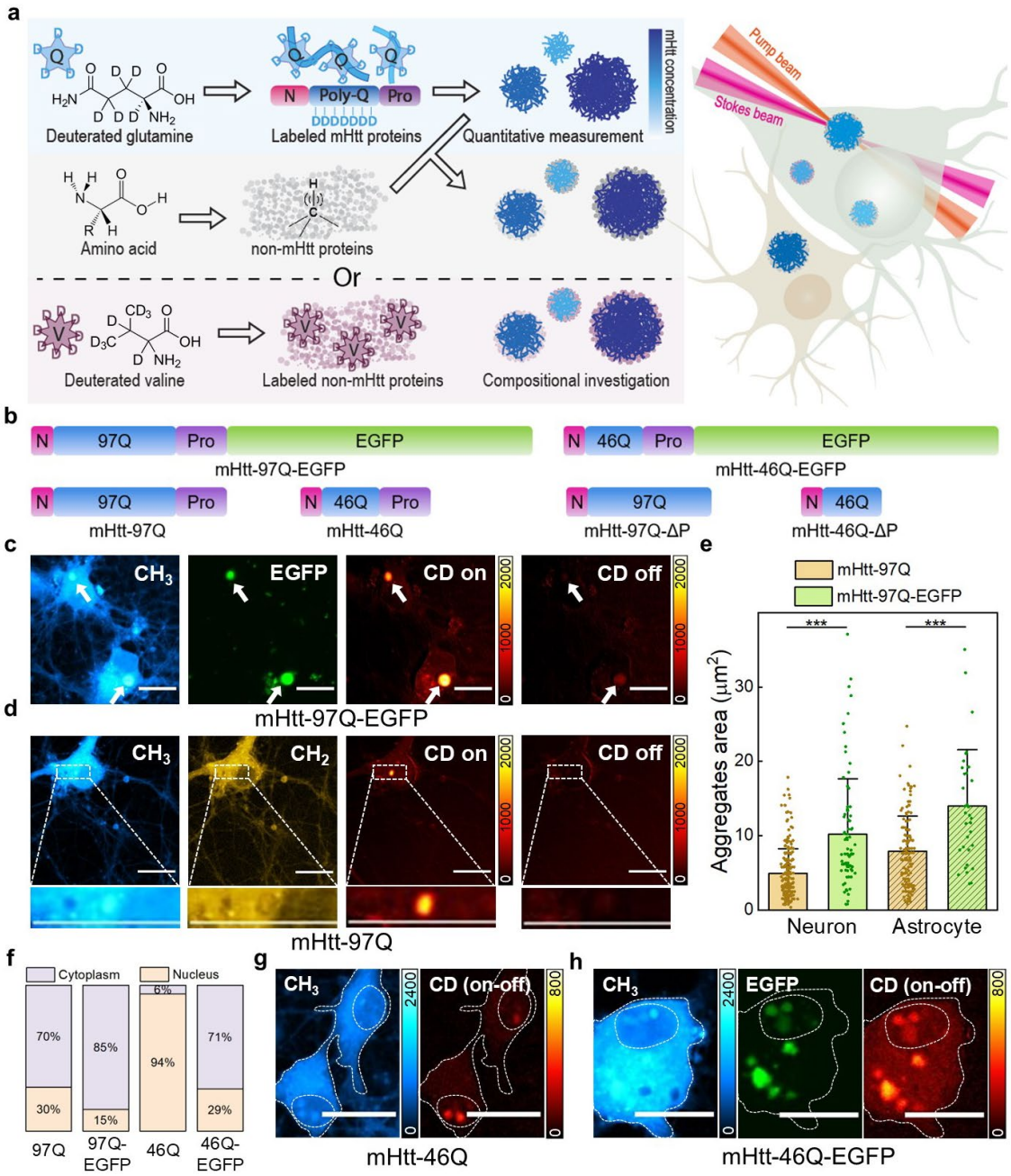

Figure 3.1. Experimental scheme for q-aggSRS imaging in hippocampal neuronal co-cultures with deuterium labeling. (a) Left: schematic illustration showing the deuterium labeling strategy for polyQ aggregates. Top: the use of deuterated glutamine (Gln-d_5) to label mHtt proteins; middle: label-free detection of non-mHtt proteins with C-H vibrations;

bottom: the use of deuterated valine (Val-d_8) to label non-mHtt proteins. Right: Stimulated Raman Scattering (SRS) microscopy with pump and Stokes beams to image the aggregates in neuron co-cultures. (b) Plasmid constructs of various model mutant Huntingtin (mHtt) exon 1 proteins, including different polyQ lengths (46Q or 97Q), with/without the fusion of EGFP at the C terminus, and the deletion of the proline-rich region. N: N-terminal 17-amino acid fragment; 46Q/97Q: polyQ tracts with different lengths; Pro: the proline-rich region; EGFP: enhanced green fluorescent protein. (c) Fluorescence and SRS images in the same set of neurons containing mHtt-97Q-EGFP aggregates with Gln-d_5 labeling. White arrows indicate polyQ aggregates. From left to right: the CH_3 channel (2940 cm^{-1}), EGFP fluorescence channel, CD on channel (2164 cm^{-1}), and CD off channel (2036 cm^{-1}). (d) SRS images in the same set of neurons containing mHtt-97Q aggregates with Gln-d_5 labeling. The inset shows a zoomed-in view of the aggregates. From left to right: the CH_3 channel (2940 cm^{-1}), CH_2 channel (2845 cm^{-1}), CD on channel (2164 cm^{-1}), and CD off channel (2036 cm^{-1}). (e) Comparison of aggregate areas for mHtt-97Q and mHtt-97Q-EGFP aggregates in neurons and astrocytes, error bar: SD, *** $p < 0.001$. (f) Distribution of aggregates between the cytoplasm and nucleus in neurons for various mHtt constructs, including mHtt-97Q, mHtt-97Q-EGFP, mHtt-46Q, and mHtt-46Q-EGFP. (g) SRS images in the same set of neurons containing mHtt-46Q aggregates with Gln-d_5 labeling. The dashed line outlines the cell shape and nucleus. Left: CH_3 channel (2940 cm^{-1}); right: CD on-off channel (on, 2164 cm^{-1}; off, 2036 cm^{-1}). (h) Fluorescence and SRS images in the same set of neurons containing mHtt-46Q-EGFP aggregates with Gln-d_5 labeling. The dashed line outlines the cell shape and nucleus. Left: the CH_3 channel (2940 cm^{-1}); middle: EGFP fluorescence channel; right: CD on-off channel (on, 2164 cm^{-1}; off, 2036 cm^{-1}). Scale bars: 20 μm.

We further developed a two-color imaging strategy with co-labeling of Gln-d_5 and deuterated valine (Val-d_8) to unequivocally investigate mHtt and non-mHtt proteins based on the absence of valine in mHtt constructs and their distinct CD spectral features (Figure 3.1a and Figure 3.2). By combining expansion microscopy and live-cell pulse-chase experiments for high spatial and temporal resolution imaging, we developed a two-phase core-shell model that elucidates the spatial organization of the aggregates and illuminates a potential dynamic mechanism underlying their growth. Notably, our data offered new insights into the heterogeneous subcellular nucleus/cytoplasm compartmentalization of polyQ aggregates, phenomenon unique to neurons but not observed in astrocytes. Moreover, we identified a distinctive "gel-like" type of aggregates within neuronal nuclei that have different concentration and dynamics features compared to cytoplasmic aggerates or the previously observed "liquid-like" states.[24] All these new findings underscore the importance of using live neuronal systems with a high-throughput strategy to investigate native aggregates. Finally, we proposed an interaction coefficient (IC), defined as the concentration ratio between non-mHtt and mHtt proteins within the same aggregate to serve as a quantitative perturbation index. The IC may find utility for future drug screening needed to develop effective therapies for HD.

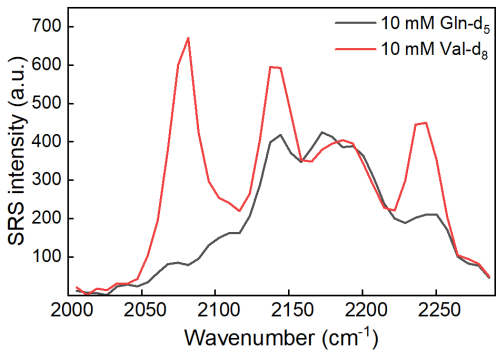

Figure 3.2. SRS spectra of 10 mM Gln-d_5 and Val-d_8 in water solutions.

3.3 The neuronal q-aggSRS imaging platform confirms severe perturbations caused by EGFP tagging

The hippocampus, a critical region involved in learning and memory, is a frequently detected site for polyQ aggregate accumulation in HD patients.[38-39] Hippocampal atrophy and dysfunction are believed to contribute to the cognitive decline observed in the disease. Here, we adopt primary neuronal co-cultures from the rat hippocampus to obtain a stable and reproducible source of neurons for investigating polyQ aggregate formation. We utilized AAV transduction for effective gene delivery with the established disease-causing model of mHtt-exon 1, containing N-terminal 17-amino acid fragments, the polyQ tract, and the proline-rich region, which has been demonstrated to significantly contribute to mHtt-induced toxicity.[27, 40] To explore the impact of polyQ tract lengths, we selected the commonly used 97Q construct and a shorter 46Q construct, which is just above the threshold for disease initiation.[41] To validate our q-aggSRS imaging strategy in live neurons, we first benchmarked it against fluorescence imaging using mHtt-97Q-EGFP. Following AAV transduction, the medium was changed to one containing Gln-d_5. SRS imaging targeting the CD vibration frequency at 2164 cm^{-1} (Figure 3.1c, CD on) indeed clearly captured mHtt aggregates (white arrowed), consistent with that shown in fluorescence (Figure 3.1c, EGFP). The much higher occurrence of glutamine in mHtt proteins compared to the general proteome provides a decent SRS signal-to-background ratio with high imaging specificity (Figure 3.1c, CD on vs. CD off, 2036 cm^{-1}). In parallel, SRS leveraged the established CH_3 label-free channel (Figure

3.1c, CH_3, 2940 cm^{-1}) to identify the cell morphology and map the non-labeled total proteome.

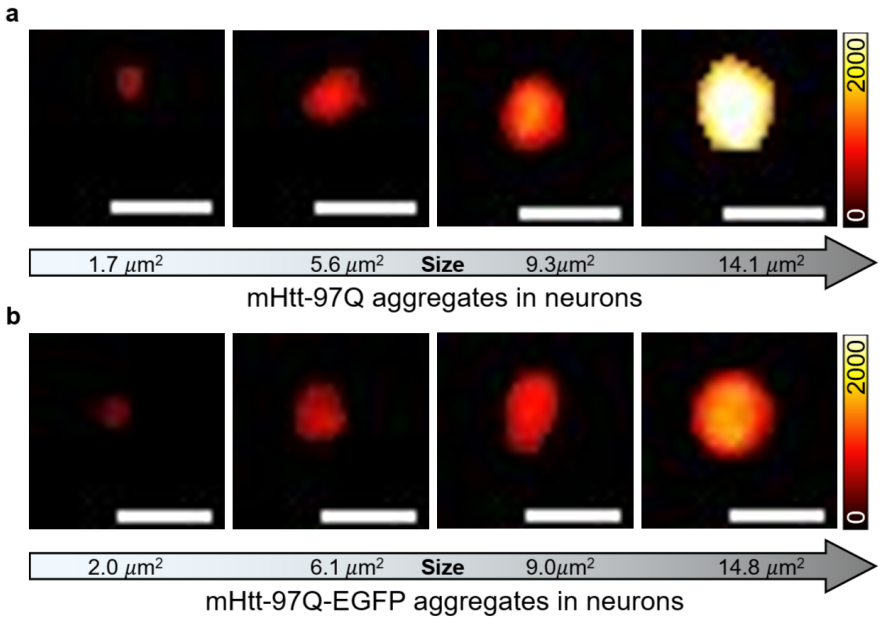

Figure 3.3. Comparison of SRS CD intensity of polyQ aggregates in neurons for mHtt-97Q (top, a) or mHtt-97Q-EGFP (bottom, b) sequences. (a) SRS images at the CD on-off channel (on, 2164 cm^{-1}; off, 2036 cm^{-1}) for 4 representative mHtt-97Q aggregates with increasing sizes in neurons. (b) SRS images at the CD on-off channel (on, 2164 cm^{-1}; off, 2036 cm^{-1}) for 4 representative mHtt-97Q-EGFP aggregates of close sizes to that shown in (a) in neurons. Scale bar: 5μm. Color bars are kept the same for (a) and (b).

We next proceeded to image native mHtt-97Q proteins without EGFP and evaluated the impact of EGFP tagging in live neurons. While the perturbation caused by fluorescent protein tagging has been revealed by studies such as Cyro-ET and immunofluorescence, they are limited to fixed cells and purified proteins, which are subject to uncertainties from labeling

specificity, antibody penetration, and protein purification length limits [25-26, 42]. Clear mHtt-97Q aggregates were detected at the CD channel (Figure 3.1d, CD on, 2164 cm^{-1} vs. CD off, 2036 cm^{-1}). The corresponding label-free CH_3 and CH_2 images provided additional information about the total proteome (Figure 3.1d, CH_3, 2940 cm^{-1}) and lipidome (Figure 3.1d, CH_2, 2845 cm^{-1}) distribution, with the latter indicating low lipid content within polyQ aggregates (Figure 3.1d, zoomed-in view from the white-boxed area). Our results revealed that polyQ aggregates formed by mHtt-97Q were about twice as bright at the CD channel compared to those of similar sizes formed by mHtt-97Q-EGFP (Figure 3.1c, 3.1d, and Figure 3.3). This indicates that mHtt-97Q forms denser aggregates with higher c(mHtt) in neurons compared to those with EGFP tagging, consistent with previous findings in HeLa cells[35], and a Cryo-ET study reporting a 50% reduction in fibril density of mHtt-97Q-EGFP aggregates.[26] In addition, we observed that polyQ aggregates formed by mHtt-97Q were overall significantly smaller than those formed by mHtt-97Q-EGFP in both neurons and astrocytes (Figure 3.1e and Figure 3.3; see Figure 3.4 and Materials and methods for live neuron and astrocyte identification). Interestingly, we also found that EGFP largely influenced the subcellular location of aggregates in live neurons (Figure 3.1f-h). Notably, while mHtt-46Q primarily formed nuclear aggregates (Figure 3.1f, 3.1g, 94% in nucleus), the EGFP-labeled counterpart dominantly formed cytoplasmic aggregates (Figure 3.1f, 3.1h, 29% in nucleus). This suggests that EGFP not only alters aggregate structure but also disrupts the intracellular translocation of mHtt proteins. As a comparison, astrocytes exclusively formed cytoplasmic aggregates regardless of the mHtt construct. Our results hence highlight the importance of conducting measurements in live neurons and avoiding EGFP perturbation to accurately investigate the polyQ aggregation phenomenon.

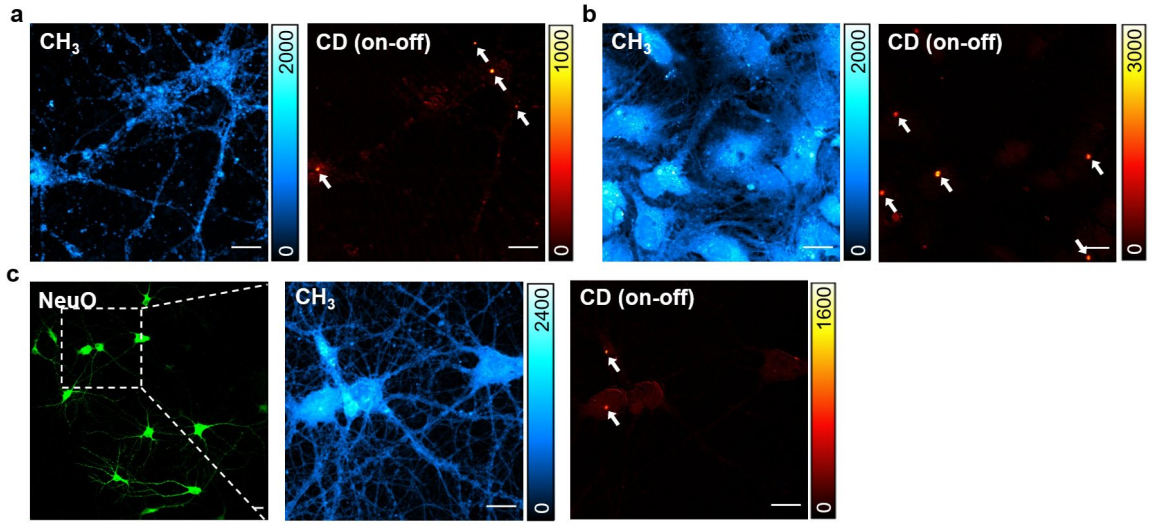

Figure 3.4. Q-aggSRS images of polyQ aggregates in neurons and astrocytes with different cell-distinguishing strategies. (a) SRS images at the CH_3 channel (left, 2940 cm^{-1}) and CD on-off channel (right, on, 2164 cm^{-1}; off, 2036 cm^{-1}) in the same set of pure neuron culture containing mHtt-97Q aggregates (arrowed in the CD image). (b) SRS images at the CH_3 channel (left, 2940 cm^{-1}) and CD on-off channel (right, on, 2164 cm^{-1}; off, 2036 cm^{-1}) in the same set of pure astrocyte culture containing mHtt-97Q aggregates (arrowed in the CD image). (c) Fluorescence image (left, NeuO) in live neuronal co-cultures containing mHtt-97Q aggregates with NeuO staining. Corresponding SRS images at the CH_3 channel (middle, 2940 cm^{-1}) and CD on-off channel (right, on, 2164 cm^{-1}; off, 2036 cm^{-1}) at the zoomed-in region in fluorescence (white-dashed box) are shown. mHtt-97Q aggregates are arrow-indicated in the CD image, indicating the top one in the astrocyte and the bottom one in the neuron. Scale bar: 20 μm.

3.4 Q-aggSRS reveals a concentration dependence of mHtt proteins, but not non-mHtt proteins, on aggregate sizes with distinct differences in neurons and astrocytes

After establishing the q-aggSRS imaging strategy in live neuron co-cultures for studying native mHtt aggregates, we aimed to perform quantitative investigations on mHtt-97Q with high spatial resolution (Figure 3.5a-d). We observed polyQ aggregates of varying sizes in the cytoplasm (Figure 3.5a, arrowed) that correspond to drastically different calculated c(mHtt) (Figure 3.5c, arrowed; see Materials and methods and Figure 3.6 for calculation details) and CH/CD ratios (Figure 3.5d, arrowed), directly measured through dividing the aggregate signals at the CH_3 channel (Figure 3.5b, arrowed) by those at the CD channel (Figure 3.5a, arrowed). CH/CD ratios serve as a surrogate indicator for the ratios of non-mHtt/mHtt proteins, since CD signals from Gln-d_5 mainly originate from mHtt proteins, while the CH_3 signals from non-Gln amino acids (a.a.), mainly represent the sequestered non-mHtt proteins within the aggregates. With CH/CD ratios and the precise sequence of mHtt-exon1, absolute c(mHtt) and c(non-mHtt) within the same aggregates could be calculated correspondingly (see Materials and methods for calculation details). The c(non-mHtt) is quantified using the average proteome protein size as a reference scale.[43-44] Compared to the fibril density measurement by cyro-ET,[26] our absolute protein concentrations should provide a complementary quantitative assessment that reflects the local density of protein monomers.

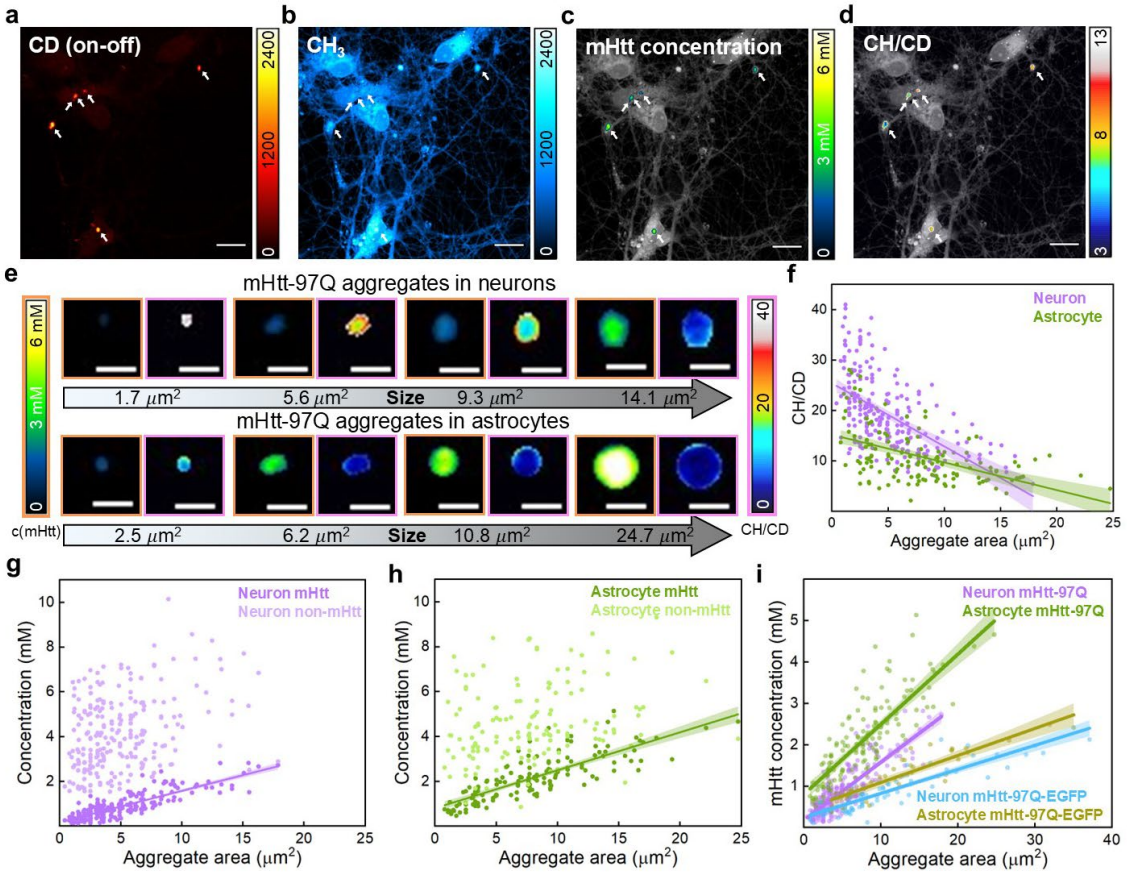

Figure 3.5. Quantitative measurement of mHtt-97Q cytoplasmic aggregates with q-aggSRS. (a-b) SRS image at the CD on-off channel (a, on, 2164 cm^{-1}; off, 2036 cm^{-1}) and the CH_3 channel (b, 2940 cm^{-1}) in the same set of neurons containing mHtt-97Q aggregates with Gln-d_5 labeling. (c-d) Image displaying the calculated c(mHtt) (c) and CH/CD ratios (d) within aggregates with the grey color representing the CH_3 channel (2940 cm^{-1}) in the same set of neurons as (a). Scale bar: 20 μm, white arrows indicate aggregates of different sizes. (e) Representative aggregates with increasing sizes, top: in neurons; bottom: in astrocytes; left: c(mHtt) image with orange border; right: CH/CD ratiometric image with pink border, scale bar: 5 μm. (f) Scatter plots with linear fittings between CH/CD ratios and aggregate

areas of polyQ aggregates in neurons (purple, $y = -1.25x + 25.42$, Pearson's $r = -0.56$) and astrocytes (green, $y = -0.55x + 15.22$, Pearson's $r = -0.50$). (g) Scatter plots with linear fittings of c(mHtt) (dark purple, $y = 0.14x + 0.18$, Pearson's $r = 0.84$) and c(non-mHtt) (light purple, Pearson's $r = 0.35$) over aggregate areas of polyQ aggregates in neurons. (h) Scatter plots with linear fittings of c(mHtt) (dark green, $y = 0.17x + 0.81$, Pearson's $r = 0.81$) and c(non-mHtt) (light green, Pearson's $r = 0.28$) over aggregate areas of polyQ aggregates in astrocytes. (i) Scatter plots with linear fittings of c(mHtt) over aggregate areas of polyQ aggregates (mHtt-97Q/ mHtt-97Q-EGFP) in neurons and astrocytes. mHtt-97Q-EGFP: neuron c(mHtt), cyan blue: $y = 0.058x + 0.25$, Pearson's $r = 0.88$; astrocyte c(mHtt), mustard yellow: $y = 0.065x + 0.44$, Pearson's $r = 0.88$. (f-i) A 95% confidence band is plotted with the fitted line.

Zoomed-in views illustrate aggregates of increasing sizes (Figure 3.5e, left panels with orange borders: c(mHtt); right panels with pink borders: CH/CD) in both neurons (Figure 3.5e, top row) and astrocytes (Figure 3.5e, bottom row). Comprehensive analysis of hundreds of aggregates across different sizes reveals that CH/CD ratios (Figure 3.5f) exhibit a weak negative correlation with aggregate sizes, which originates from a strong positive correlation between calculated c(mHtt) and aggregate sizes (Pearson's r >0.8) in both neurons (Figure 3.5g, dark purple) and astrocytes (Figure 3.5h, dark green). In contrast, c(non-mHtt) sequestered in the same aggregates is intriguingly high, reaching up to 10 mM, but shows a much weaker and scattered dependence on aggregate sizes (Pearson's r <0.4) (Figure 3.5g, light purple, and Figure 3.5h, light green). This indicates that the amount of trapped non-mHtt proteins is independent of aggregate size and the protein entrapment mechanisms may

be different for mHtt and non-mHtt proteins. To our knowledge, this is the first time that quantitative mapping of polyQ aggregates has been achieved in live neurons and astrocytes.

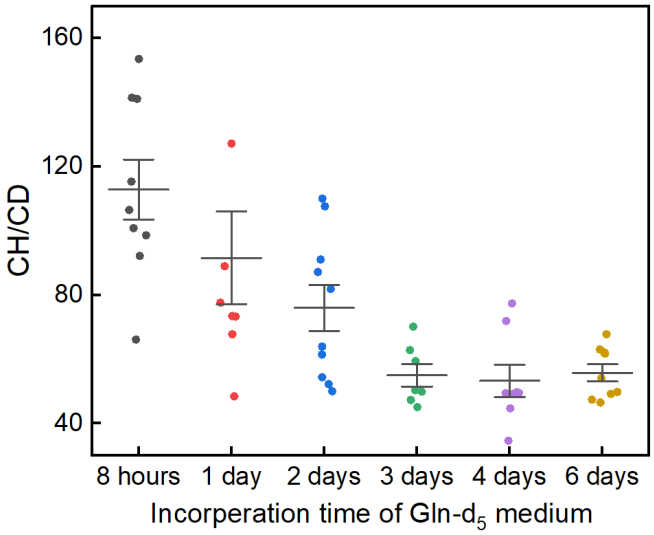

Figure 3.6. CH/CD ratios dependence on Gln-d_5 medium incubation time without transductions of AAV packed mHtt plasmids.

We further found that aggregates of the same sizes consistently present a 2-fold higher c(mHtt) in astrocytes (Figure 3.5i, dark green) compared to those in neurons (Figure 3.5i, dark purple). Interestingly, the fitting curves showed lower intercepts in neurons compared to astrocytes (Figure 3.5i, dark purple vs. dark green, and Figure 3.7a, 3.7b), which implies that neurons may have a lower concentration threshold for aggregate formation, and are likely more susceptible to perturbations of mHtt proteins.[36-37, 45] Such heterogeneity between neurons and astrocytes became much smaller for the mHtt-97Q-EGFP aggregates (Figure 3.5i, cyan blue vs. mustard yellow), again emphasizing the influence of EGFP labeling in a direct live-cell context. Additionally, aggregates were found to appear in both the cell body

and axons/dendrites regions. While the c(mHtt) vs. aggregate size relationship remains the same across different subcellular locations (Figure 3.8a), the CH/CD ratios in the cell body were slightly higher (Figure 3.8b), likely reflecting differences in the local subcellular environment. However, due to the similar quantification matrix, we will focus on the analysis of aggregates in the cell bodies.

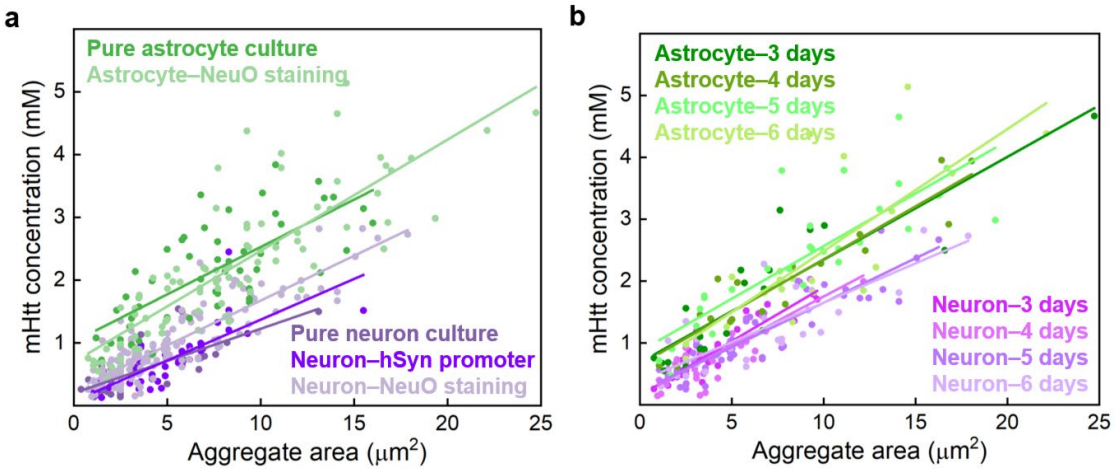

Figure 3.7. Demonstration of validity of calculated mHtt concentration (c(mHtt)) with minimal influence from cell-differentiation methods and expression time. (a) Scatter plots with linear fittings between c(mHtt) and aggregate areas of polyQ aggregates in neurons and astrocytes by different cell-differentiation methods mentioned above. Pure neuron culture: $y = 0.20x + 0.10$, Pearson's $r = 0.74$; Neuron-hSyn promoter: $y = 0.13x + 0.074$, Pearson's $r = 0.78$; Neuron-NeuO staining: $y = 0.14x + 0.25$, Pearson's $r = 0.88$; Pure astrocyte culture: $y = 0.15x + 1.02$, Pearson's $r = 0.74$; Neuron-NeuO staining: $y = 0.18x + 0.70$, Pearson's $r = 0.84$. (b) Scatter plots with linear fittings between c(mHtt) and aggregate areas of polyQ aggregates in neurons and astrocytes in

different post-AAV incubation times. Neuron-3 days: $y = 0.18x + 0.15$, Pearson's $r = 0.93$; Neuron-4 days: $y = 0.16x + 0.18$, Pearson's $r = 0.82$; Neuron-5 days: $y = 0.14x + 0.22$, Pearson's $r = 0.90$; Neuron-6 days: $y = 0.13x + 0.38$, Pearson's $r = 0.81$; Astrocyte-3 days: $y = 0.17x + 0.68$, Pearson's $r = 0.88$; Astrocyte-4 days: $y = 0.17x + 0.65$, Pearson's $r = 0.94$; Astrocyte-5 days: $y = 0.17x + 0.86$, Pearson's $r = 0.82$; Astrocyte-6 days: $y = 0.20x + 0.51$, Pearson's $r = 0.75$.

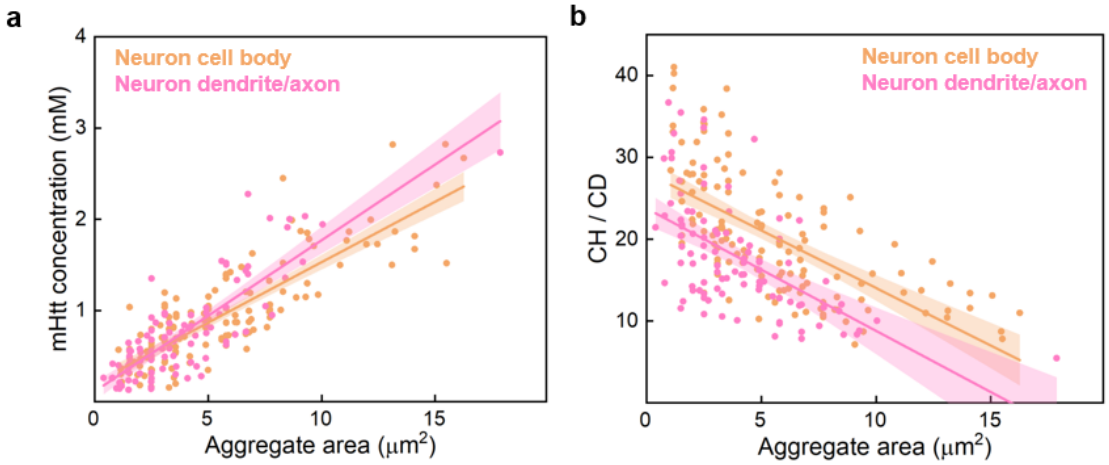

Figure 3.8. Quantitative analysis of mHtt-97Q aggregates in neuronal subcellular localizations of the cell body and dendrite/axon. (a) Scatter plots with linear fittings between c(mHtt) and aggregate areas of polyQ aggregates in the neuron cell body (orange, $y = 0.13x + 0.20$, Pearson's $r = 0.84$) and dendrite/axon (pink, $y = 0.17x + 0.11$, Pearson's $r = 0.84$). (b) Scatter plots with linear fittings between CH/CD ratios and aggregate areas of polyQ aggregates in the neuron cell body (orange, $y = -1.41x + 28.12$, Pearson's $r = -0.65$) and dendrite/axon (pink, $y = -1.50x + 23.78$, Pearson's $r = -0.58$). A 95% confidence band is plotted with the fitted line.

3.5 Expansion and two-color q-aggSRS imaging reveal a core-shell aggregate structure

As shown in Figure 3.5e, the CH/CD ratiometric images (right panels with pink borders) imply a core-shell structure, where mHtt proteins are concentrated in the core of aggregates and non-mHtt proteins are relatively more prominent at the periphery. We hence sought to confirm and understand such a core-shell spatial distribution of mHtt and non-mHtt proteins within the same aggregates with enhanced spatial and spectral resolution. First, we applied Raman-tailored sample-expansion microscopy to perform q-aggSRS VISTA (Vibrational Imaging of Swelled Tissues and Analysis),[46] a super-resolution Raman imaging technique previously developed in our laboratory. VISTA has been applied to probe both polyQ and amyloid beta protein aggregates in cells and tissues with validated isotropic expansion[47]. Neurons containing mHtt-97Q aggregates were embedded in polymer hydrogels and then expanded isotropically by 2-fold to reach an effective resolution below 200 nm. Indeed, the overlay image (Figure 3.9a, right) of the CH_3 (Figure 3.9a, left) and the CD (Figure 3.9a, middle) channel clearly confirms a core-shell structure with non-mHtt proteins being relatively more abundant in the shell.

Second, we devised a new two-color q-aggSRS strategy (Figure 3.9b, top) by co-labeling live neurons with Gln-d_5 and deuterated valine (Val-d_8), harnessing differences in Raman spectral features arising from their distinct chemical structures (Figure 3.9b, bottom). Since valine is absent in mHtt protein sequences, all Val-d_8 signals originate exclusively from newly synthesized non-mHtt proteins, eliminating any crosstalk contributed by mHtt proteins, which exist in the CH_3 signals. Such a two-color strategy more clearly resolved the

core-shell analysis. By incubating cells with Gln-d_5 and Val-d_8 substituted culture medium and performing two-color q-aggSRS imaging (Figure 3.9b, and Materials and methods for unmixing details), our results consistently confirmed clear core-shell structures (Figure 3.9c, right, merged), with non-mHtt proteins (Figure 3.9c, middle, Val-d_8) more enriched in the shell and mHtt in the core (Figure 3.9c, left, Gln-d_5). This spatial organization becomes more pronounced as aggregate sizes increase (Figure 3.9c, top two rows).

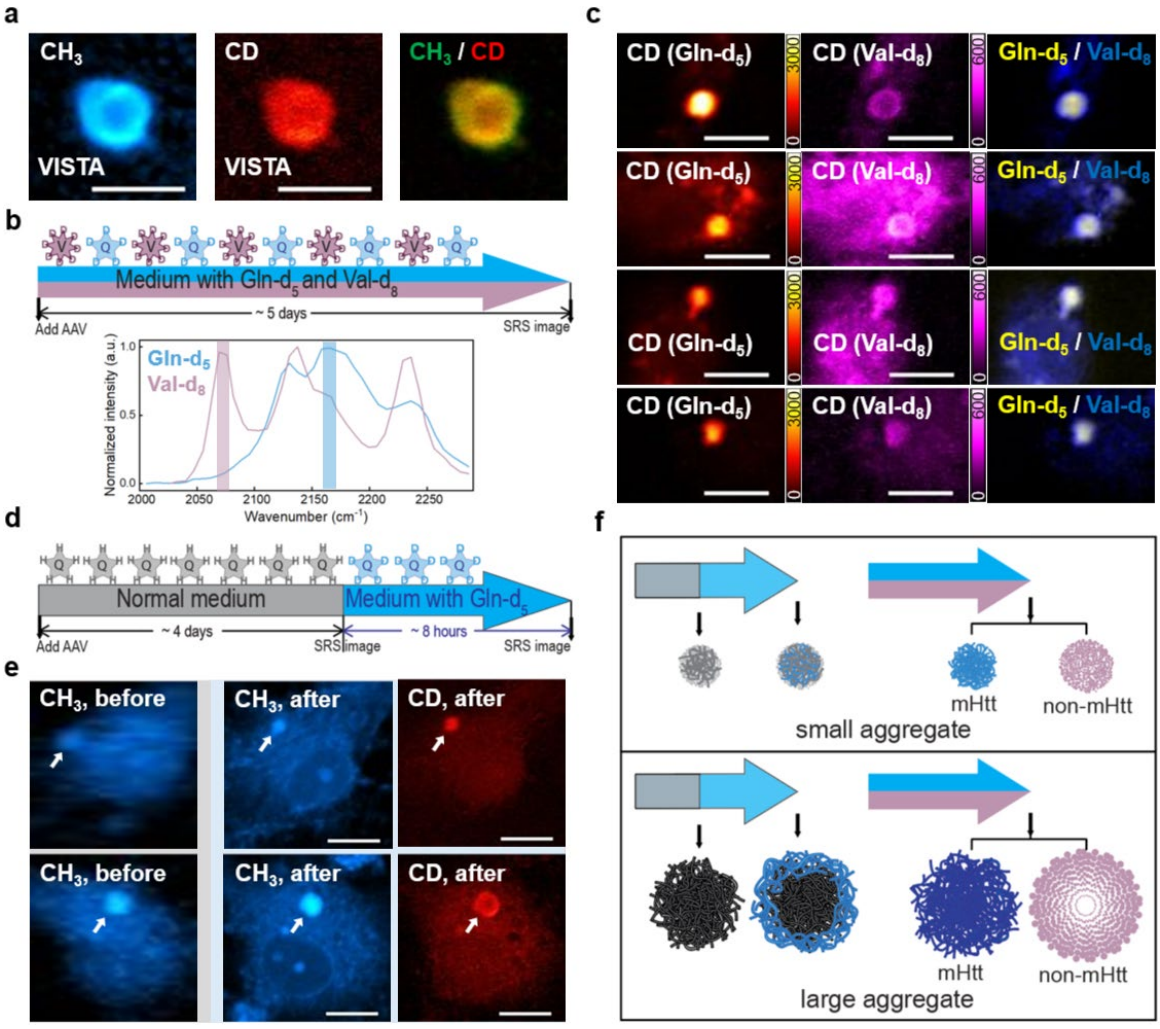

Figure 3.9. Structural investigation and dynamic tracking of mHtt-97Q cytoplasmic aggregates. (a) Q-aggSRS images from the VISTA sample (2-fold expansion). From left to right: the CH_3 channel (2940 cm^{-1}), the CD channel (2164 cm^{-1}), merged image of the CH_3 and CD channels, scale bar: 10 μm. (b) Top: Experimental design of two-color imaging by co-incubation of Gln-d_5 and Val-d_8. Bottom: SRS spectra of polyQ aggregates with either Gln-d_5 (blue) or Val-d_8 (purple) incorporation. The two colored bands indicate SRS acquisition positions (2068 cm^{-1}, 2164 cm^{-1}) for unmixing. (c) Two-color imaging of Gln-d_5 labeled mHtt proteins and Val-d_8 labeled non-mHtt proteins for four sets of aggregates, scale bar: 10 μm. (d) Experimental design for the pulse-chase labeling experiment. (e) Q-aggSRS images at the CH_3 channel (2940 cm^{-1}) and CD channel (2164 cm^{-1}) in the same set of neurons before and after the medium replacement with white arrows indicating polyQ aggregates. The top row shows a diffused CD signal across the aggregate, while the bottom row shows a shell structure of the CD signal in the aggregate, scale bar: 10 μm. (f) Proposed two-stage aggregate growth model for small aggregates (top) and large aggregates (bottom), building upon the pulse-chase experiment (left) and the spatial distribution of mHtt proteins and non-mHtt proteins within the same polyQ aggregate (right).

3.6 Pulse-chase q-aggSRS reveals two-phase growth dynamics

One unique advantage of our metabolic labeling strategy is the ability to distinguish newly synthesized proteins from the pre-existing pool by a pulse-chase medium change over time. To investigate the growth dynamic of cytoplasmic mHtt-97Q aggregates in live neurons, we designed a pulse-chase time-lapse experiment (Figure 3.9d). Neurons were initially cultured

in a regular medium for 4 days post-AAV addition to allow sufficient protein expression, followed by first round of SRS imaging at the CH_3 channel (Figure 3.9e, CH_3, before) to confirm the formation of polyQ aggregates (arrowed). Here, we chose very low laser power and rapid acquisition time to minimize potential photodamage to neurons (Figure 3.9e, CH_3, before). The regular medium was then replaced with the Gln-d_5 medium for 8 hours (Figure 3.9d), after which the same neurons were re-imaged at both the CH_3 and the CD channels (Figure 3.9e, CH_3, after and CD, after). Here, the CD signals indicate newly synthesized mHtt proteins incorporated into the same aggregates after the medium change. Interestingly, representative q-aggSRS images revealed two distinct patterns of CD signal distribution: a uniform distribution throughout smaller aggregates (Figure 3.9e, top row, arrowed) and an exclusive localization to the periphery in larger aggregates (Figure 3.9e, bottom row, arrowed). This result suggests that large aggregates represent a more compact and inert state as newly sequestered mHtt proteins are mostly restricted to the outer layer of large aggregates, whereas small aggregates allow newly sequestered mHtt proteins to diffuse throughout the core, implying a more flexible structural organization.

Collectively, we propose a two-phase composition and growth model between the small and large polyQ aggregates (Figure 3.9f). According to Cryo-ET high-resolution structural studies, mHtt forms polyQ aggregates in the form of fibrils.[27, 48-49] Based on this, we propose that small aggregates have a relatively loose fibril structure with higher flexibility and diffusion rate, hence allowing newly sequestered proteins to freely diffuse into the core (Figure 3.9, top row, left). In addition, they form a weak core-shell compositional structure featuring a minor exclusion of non-mHtt proteins in the center (Figure 3.9f, top row, right).

In large aggregates with higher c(mHtt), the core is denser and more compact, preventing further diffusion of newly formed mHtt fibrils into the core and leading to a shell-like distribution of newly sequestered mHtt surrounding the dense core (Figure 3.9f, bottom row, left). This effect is likely more pronounced for non-mHtt proteins, hence resulting in an obvious core-shell compositional structure between mHtt and non-mHtt cytosolic proteins in large aggregates (Figure 3.9f, bottom row, right). Our core-shell model provides new insights and aids in the understanding of polyQ aggregate formation, growth, and structural composition.

3.7 Nuclear polyQ aggregates are dominant for the shorter Q length and feature lower c(mHtt) than cytoplasmic aggregates

The above discussions feature cytoplasmic polyQ aggregates, which are the major form in mHtt-97Q (Figure 3.1f, 70% in cytoplasm). For mHtt-46Q with shorter Q length, nuclear aggregates became dominant (Figure 3.1f, 94% in nucleus, Figure 3.10a). We confirmed that the observed aggregates were not nucleoli, which would appear more prominent in the CH_3 channel (Figure 3.10a). We further found that mHtt-46Q aggregates exclusively localized in the cytoplasm of astrocytes (Figure 3.10a, 3.10b), revealing stark differences in aggregation compartmentalization between neurons and astrocytes. The c(mHtt) of nuclear aggregates was also found to be significantly lower and the CH/CD ratios were significantly higher compared to cytoplasmic ones (Figure 3.10c, 3.10d). This suggests a more preferential sequestration of non-mHtt proteins in nuclear aggregates (Figure 3.11), likely contributing to higher toxicity as one suggested toxicity mechanism involves the depletion of functional

proteins by aggregate formation[4, 16, 18, 50-53]. In addition, we found that while there was typically one cytoplasmic aggregate per cell (Figure 3.10e, green and purple histograms with distribution curves), there were usually multiple nuclear aggregates per cell (Figure 3.10e, light blue histogram with distribution curve). These findings indicate distinct features between nuclear and cytoplasmic aggregates,[42] emphasizing the role of the subcellular environment in influencing aggregate formation.

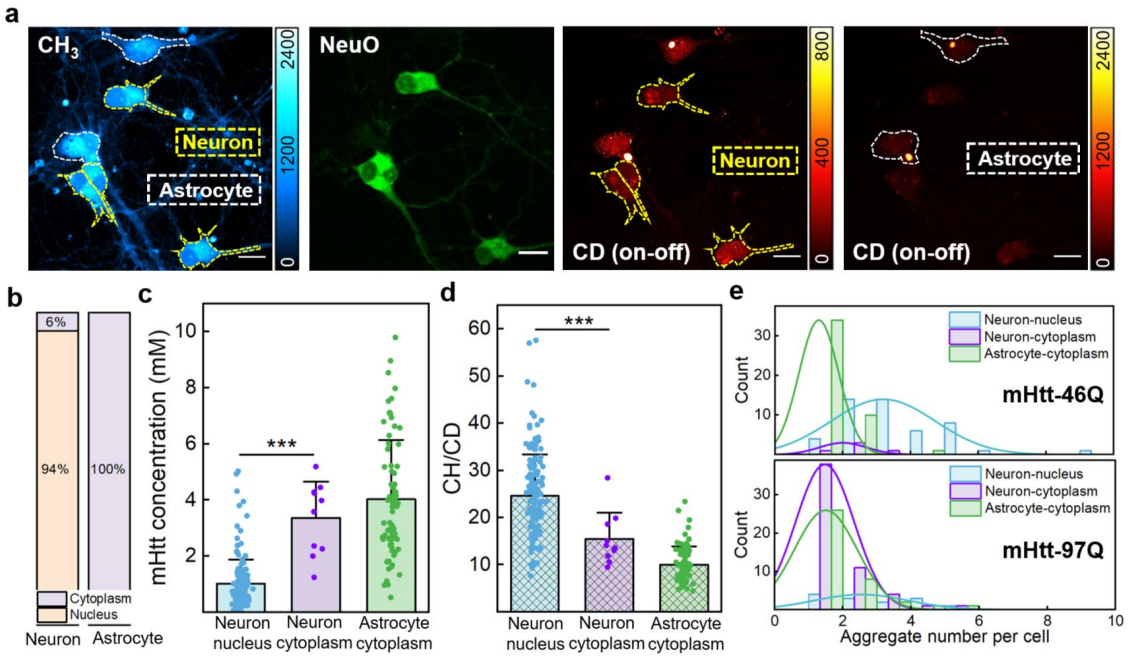

Figure 3.10. Quantitative measurement of mHtt-46Q nuclear aggregates with q-aggSRS. (a) Fluorescence and SRS images in the same set of neuronal co-cultures containing mHtt-46Q aggregates. From left to right: the CH_3 channel (2164 cm^{-1}) with the outline for neurons (dashed, yellow boxes) and astrocytes (dashed, white boxes), the fluorescence image showing NeuO[54] staining for neurons, the CD on-off channel (on, 2164 cm^{-1}; off, 2036 cm^{-1}) image showing nuclear aggregates in neurons, the CD on-off channel (on, 2164 cm^{-1}; off,

2036 cm^{-1}) image showing cytoplasmic aggregates in astrocytes under different contrast, scale bar: 20 μm. (b) Distribution of nuclear and cytoplasmic aggregates for mHtt-46Q in neurons and astrocytes. (c-d) c(mHtt) (c) and CH/CD ratios (d) of mHtt-46Q aggregates in different subcellular compartments, including neuron nucleus, neuron cytoplasm, and astrocyte cytoplasm, *** $p < 0.001$. (e) Histogram plots showing the distribution of aggregate numbers per cell for mHtt-46Q and mHtt-97Q in different subcellular compartments, including neuron nucleus, neuron cytoplasm, and astrocyte cytoplasm.

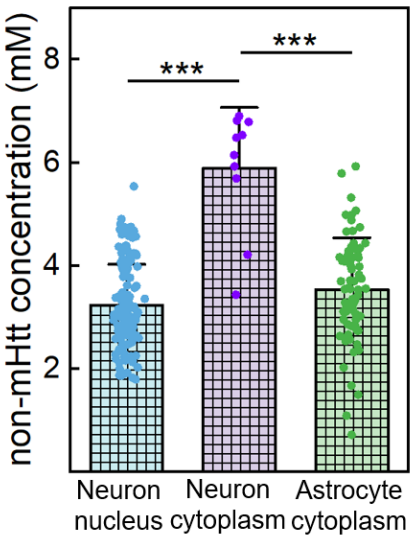

Figure 3.11. Non-mHtt concentrations of mHtt-46Q aggregates in different subcellular compartments, including neuron nucleus (blue), neuron cytoplasm (purple), and astrocyte cytoplasm (green), *** $p < 0.001$. No plot is shown for astrocyte nucleus due to almost 0% aggregate appearance.

3.8 The deletion of the proline-rich region significantly changes c(non-mHtt) and aggregate structures

We next sought to examine the effect of deleting the proline-rich region in mHtt constructs, using mHtt-46Q-ΔP and mHtt-97Q-ΔP. Previous yeast survival studies showed that deleting the proline-rich region enhances the toxicity of mHtt.[17, 55] However, the relationship between mHtt sequences and their toxicity remains unclear. Our experiments revealed that the deletion of the proline-rich region slightly increased the percentage of nuclear aggregates (Figure 3.12a vs. Figure 3.1f). Representative q-aggSRS images at both the CD and CH_3 channels of mHtt-46Q-ΔP aggregates in a neuron show a much higher contrast for aggregates at the CH_3 channel (Figure 3.12b), compared to mHtt-46Q aggregates (Figure 3.10a, CH_3), indicating a much-increased sequestration of non-mHtt proteins in the mHtt-46Q-ΔP aggregates. Subsequent statistical analysis indeed confirmed the observed increase in c(non-mHtt) for both mHtt-46Q-ΔP (Figure 3.12c) and mHtt-97Q-ΔP (Figure 3.12d) across different subcellular compartments in neurons and astrocytes, while only minor changes were seen in c(mHtt) (Figure 3.13a, 3.13b). This finding is again consistent with the proposed toxicity mechanism of aggregates, which involves the depletion of essential structural (i.e. actins) and functional (i.e. chaperones) cytosolic proteins.[4, 16, 18, 50-53] The observed increase in absolute c(non-mHtt) may provide quantitative support for the enhanced toxicity associated with the deletion of the proline-rich region.

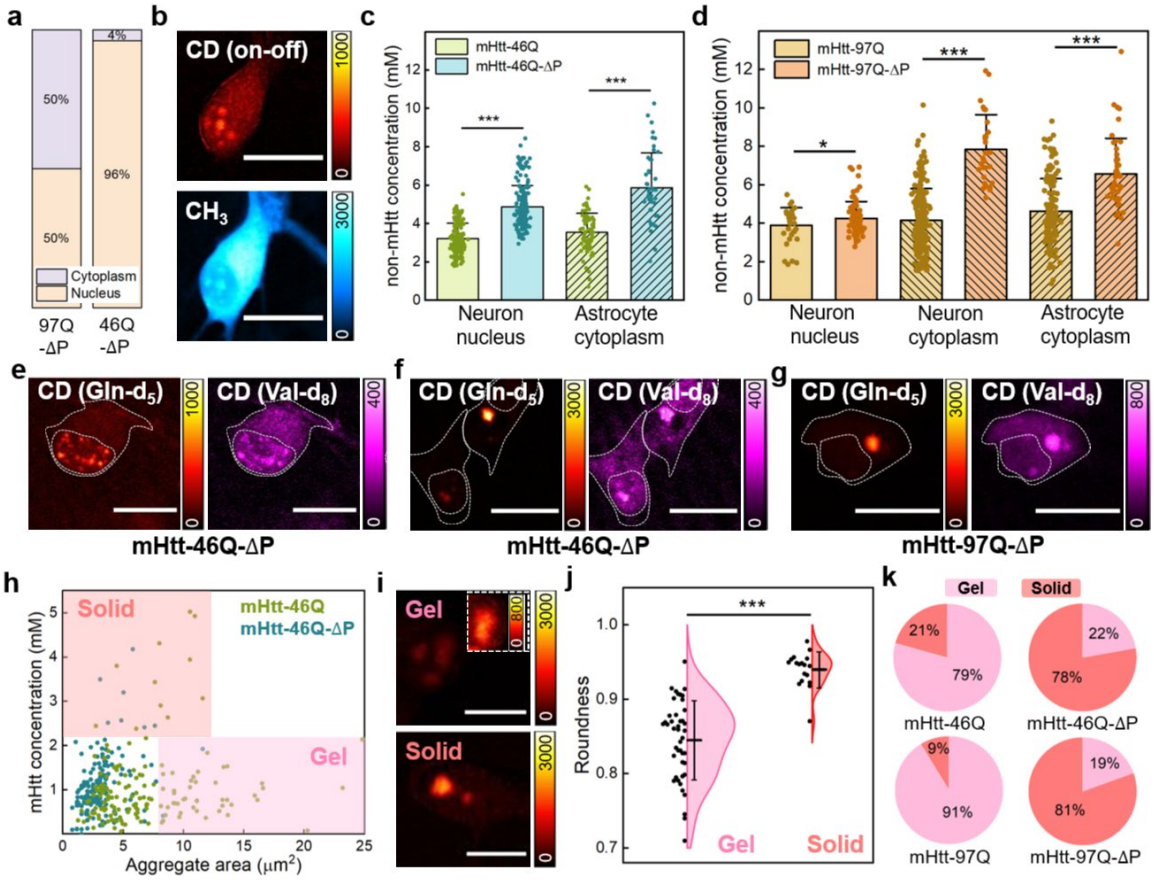

Figure 3.12. Quantitative measurement of mHtt-46Q-ΔP and mHtt-97Q-ΔP aggregates with q-aggSRS. (a) Distribution of nuclear and cytoplasmic aggregates in neurons for mHtt-46Q-ΔP and mHtt-97Q-ΔP. (b) q-aggSRS images at the CD (top, 2164 cm^{-1}) and CH_3 (bottom, 2940 cm^{-1}) channels in the same set of neuron containing mHtt-46Q-ΔP aggregates with Gln-d_5 labeling, scale bar: 20 μm. (c-d) Comparison of c(non-mHtt) in different subcellular compartments, including neuron nuclei, neuron cytoplasm, and astrocyte cytoplasm, for mHtt-46Q/mHtt-46Q-ΔP (c) and mHtt-97Q/mHtt-97Q-ΔP (d), * $p < 0.05$, *** $p < 0.001$. No neuronal cytoplasm data is shown in (c) due to an insufficient number of aggregates detected, as the majority of aggregates are in neuronal nuclei. (e-g) Two-color imaging of Gln-d_5 labeled mHtt proteins and Val-d_8 labeled non-mHtt proteins for mHtt-

46Q-ΔP nuclear aggregates (e), mHtt-46Q-ΔP cytoplasmic aggregates (f), mHtt-97Q-ΔP cytoplasmic aggregates (g). The dashed lines outline the cell shape and nucleus, scale bar: 20 μm. (h) Scatter plot of c(mHtt) vs nuclear aggregate areas for mHtt-46Q and mHtt-46Q-ΔP, indicating a bifurcation of concentration-area dependence. The pink region (lower right quadrant) indicates the nuclear gel-like aggregates, which feature low concentration-area dependence; the coral region (upper left quadrant) indicates the nuclear solid-like aggregates, with high concentration-area dependence. (i) Q-aggSRS images of representative nuclear gel-like aggregates (top) and solid-like aggregates (bottom) at the CD on-off channel (on, 2164 cm^{-1}; off, 2036 cm^{-1}), scale bar: 10 μm. The white box inset shows the same gel-like aggregate but with enhanced contrast. (j) Violin plot comparing the roundness of nuclear gel-like and solid-like mHtt aggregates, *** $p < 0.001$. (k) Pie charts showing the proportion of gel-like and solid-like nuclear aggregates across mHtt constructs of mHtt-46Q, mHtt-46Q-ΔP, mHtt-97Q, and mHtt-97Q-ΔP.

Interestingly, our two-color imaging through the co-incubation of Gln-d_5 and Val-d_8 demonstrated that ΔP sequences always resulted in a homogenous aggregate structure between mHtt and non-mHtt proteins in both nucleus and cytoplasm, regardless of aggregation sizes and Q length (Figure 3.12e-g). This drastically differs from the previously characterized core-shell structure observed from mHtt-97Q, even for large cytoplasmic aggregates formed by mHtt-97Q-ΔP (Figure 3.12g). Previous protein structural studies identified the proline-rich region as a helix breaker that disrupts the helical structure formed by polyQ tracts. For ΔP sequences, it is plausible that the helical structures remain relatively intact and hence promote mHtt aggregation.[56-59] The increased aggregation capability in ΔP

sequences might also strengthen the interaction between mHtt and non-mHtt proteins and cause more effective sequestration of non-mHtt proteins to the core regardless of aggregate sizes, highlighting the importance of further understanding of the interaction between sequestered mHtt and non-mHtt proteins.

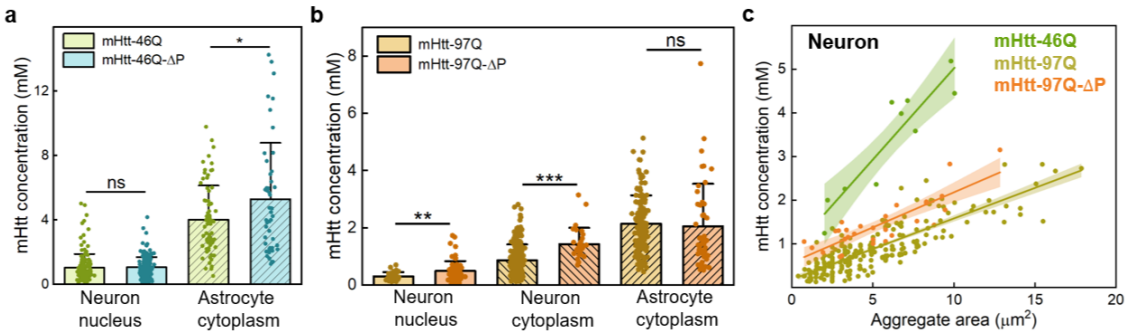

Figure 3.13. Quantitative measurement of polyQ aggregates with different mHtt constructs in neurons and astrocytes. (a-b) c(mHtt) of polyQ aggregates in different subcellular compartments, including neuron nucleus, neuron cytoplasm, and astrocyte cytoplasm, for mHtt-46Q/mHtt-46Q-ΔP (a) and mHtt-97Q/mHtt-97Q-ΔP (b), ns: not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. No neuronal cytoplasm data is shown in (a) due to an insufficient number of aggregates detected, as the majority of aggregates are in neuronal nuclei. (c) Scatter plots with linear fittings between c(mHtt) and aggregate areas of cytoplasmic aggregates in neurons. mHtt-46Q: $y = 0.42x + 0.84$, Pearson's $r = 0.92$; mHtt-97Q: $y = 0.14x + 0.18$, Pearson's $r = 0.84$; mHtt-97Q-ΔP: $y = 0.16x + 0.57$, Pearson's $r = 0.84$.

3.9 An unrecognized "gel-like" state is revealed for nuclear aggregates

Similar to our analysis for cytoplasmic aggregates, we next plotted the c(mHtt) dependence on the nuclear aggregate sizes across all four sequences: mHtt-46Q and mHtt-46Q-ΔP (Figure 3.12h), and mHtt-97Q and mHtt-97Q-ΔP (Figure 3.14a). Surprisingly, we observed a drastic bifurcation, revealing two distinct populations, highlighted in separate quadrants (Figure 3.12h and 3.14a). The upper-left population (Figure 3.12h, shaded coral) has a strong positive correlation between c(mHtt) and aggregate sizes, similar to the trend shown in cytoplasmic aggregates. We hence defined this population as the "solid-like" aggregates, featuring dense mHtt core with lower mobility. In contrast, the lower-right population (Figure 3.12h, shaded pink) shows no correlation between c(mHtt) and aggregate sizes. Since their density remains uncorrelated with variations in size, but their protein mobility is lower compared to the known "liquid" form, [24] we defined this low-concentration type of aggregates as "gel-like", close to the semi-solid properties of a gel. Representative images show that the "gel-like" aggregates appear elongated (Figure 3.12i, top, contrast-adjusted image shown in the while-dashed box), whereas the "solid-like" aggregates are more circular with significantly higher CD intensities (Figure 3.12i, bottom). We next calculated the roundness, defined as a ratio of 4π*area divided by the square of the perimeter, where a value closer to 1 indicates a shape closer to a perfect circle. Statistical analysis confirmed a significant difference in roundness between the two types, with "solid-like" aggregates being more circular (Figure 3.12j). Based on the roundness, we again note that this "gel-like" state is distinct from previously reported "liquid-like" state which exhibits a more circular shape due to its fluidity.[24] These "gel-like" aggregates we discovered for the first time, defined based on their physical properties of concentration, fluidity, and morphology, are expected

have higher diffusion flexibility for sequestration compared to “solid-like” ones, but possess less fluidity compared to the “liquid-like” state. Interestingly, we found that the percentage of “gel-like” nuclear aggregates largely decreases for both mHtt-46Q-ΔP and mHtt-97Q-ΔP compared to their counterparts (Figure 3.12k). Along with the above two-color imaging results (Figure 3.12e-g), this again indicates that the deletion of the proline-rich region should promote the binding interactions between sequestered mHtt and non-mHtt proteins, further changing the properties of polyQ aggregates.

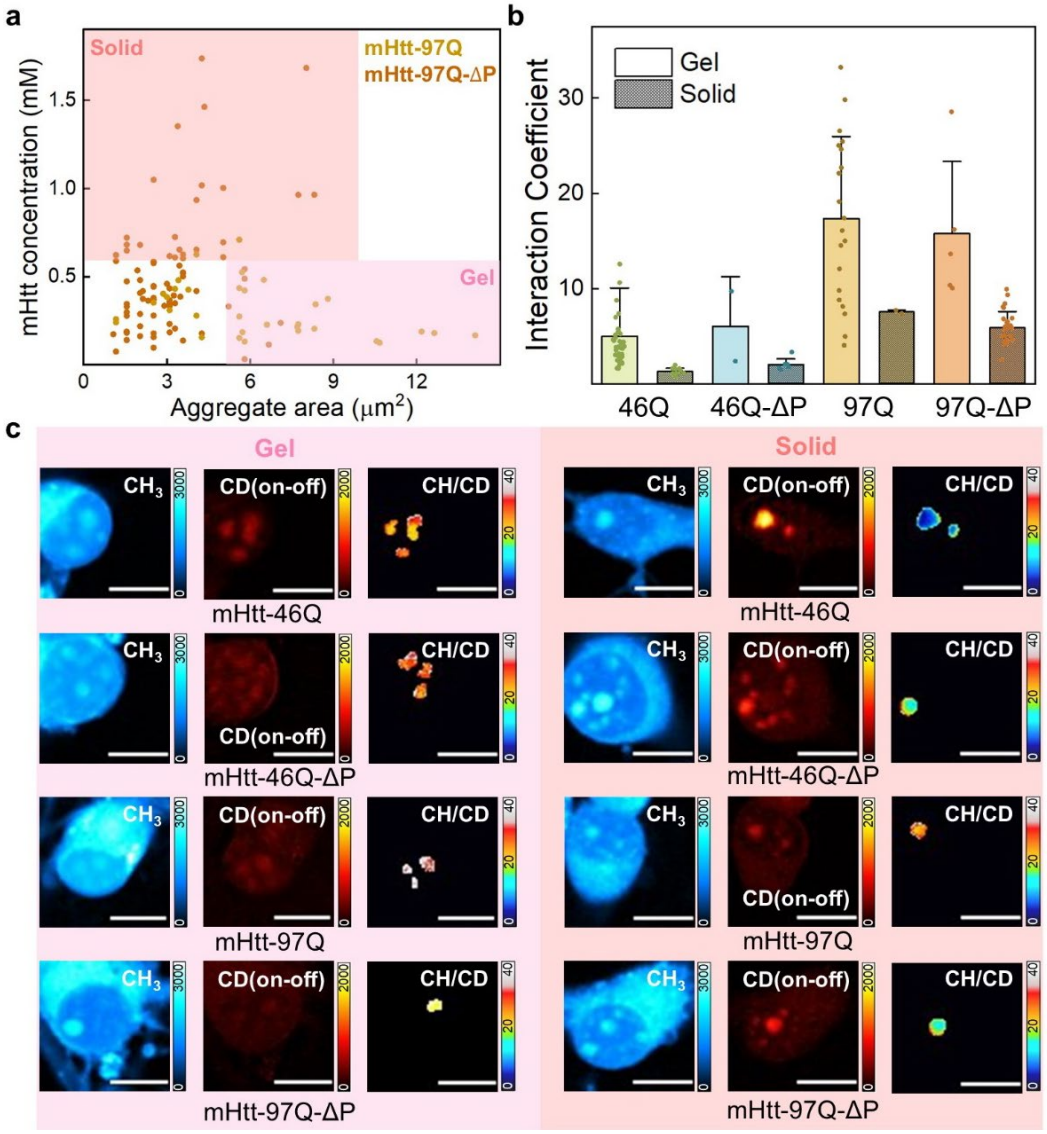

Figure 3.14. Investigation of gel-like and solid-like nuclear aggregates with different mHtt constructs. (a) Scatter plots of c(mHtt) vs aggregate areas for mHtt-97Q and mHtt-97Q-ΔP, indicating a bifurcation of concentration-area dependence. The pink region (lower right quadrant) indicates the nuclear gel-like aggregates, which feature low concentration-area dependence; the coral region (upper left quadrant) indicates the nuclear solid-like aggregates, which feature high concentration-area dependence. (b) Bar diagram of interaction coefficients (ICs, c(non-mHtt)/c(mHtt))) of gel-like (solid-fill) and solid-like (diagonal-line patterned) nuclear aggregates for different mHtt constructs of mHtt-46Q, mHtt-46Q-ΔP, mHtt-97Q, and mHtt-97Q-ΔP. (c) Q-aggSRS images of gel-like (left panel, pink) and solid-like (right panel, coral) nuclear aggregates across constructs of mHtt-46Q, mHtt-46Q-ΔP, mHtt-97Q, and mHtt-97Q-ΔP. For each set of three images: left: the CH_3 channel (2940 cm^{-1}), middle: the CD on-off channel (on, 2164 cm^{-1}; off, 2036 cm^{-1}), right: CH/CD ratios of the masked representative gel-like (left panel) or solid-like (right panel) aggregates. scale bar: 10 μm.

3.10 Discussion

We established the neuronal q-aggSRS microscopy platform for live hippocampal neuron co-cultures and conducted a comprehensive and quantitative analysis of mHtt polyQ aggregates under various conditions in an EGFP-free manner. Adding more evidence to previous studies with fixed-cells and purified proteins, we substantiated the increased perturbation of EGFP tagging for interrogating the aggregate properties in live-cell context with high specificity and more pathologically relevant models. We calculated the

concentrations of mHtt and non-mHtt proteins and revealed varied concentration dependence on sizes. Utilizing expansion q-aggSRS imaging, along with two-color and pulse-chase time-lapse imaging, our data established a two-phase growth model with detailed compositional structures for small and large polyQ aggregates *in situ* from live neurons. Interestingly, we revealed notable heterogeneity in subcellular location for nuclear and cytoplasmic aggregates across different mHtt constructs and between neurons and astrocytes. Specifically, we only detected cytoplasmic aggregates in astrocytes, while both nuclear and cytoplasmic aggregates were found to exist in neurons (Figure 3.15a). This highlights the importance of using neurons for polyQ studies, as other cell models may not fully replicate the native states and subcellular environment. In neuronal nuclei, we identified unique "gel-like" aggregates, which have c(mHtt) independent of aggregate sizes and feature a homogeneous spatial distribution of mHtt and non-mHtt proteins (Figure 3.15b, pink, and Figure 3.14c). This is drastically different from cytoplasmic aggregates, which have a strong linear correlation between c(mHtt) and aggregate sizes and show a relatively homogeneous composition in small aggregates but a clear core-shell structure in large aggregates (Figure 3.15b, coral).

The strong linear correlation between c(mHtt) and aggregate sizes enables precise quantification of aggregate growth. We observed that the slopes in the fitting curves remained consistent in both neurons and astrocytes, suggesting a similar growth trend across different cell types. However, the intercepts vary, which points to different critical concentrations required for aggregation initiation and likely reflects the heightened susceptibility to mHtt perturbations in neurons compared to astrocytes. Notably, we found that c(mHtt) in mHtt-46Q aggregates is approximately twice that of similarly sized mHtt-

97Q aggregates, resulting in a two-fold steeper slope in the mHtt-46Q fitting curve (Figure 3.13c). This suggests that mHtt-46Q has a weaker aggregation ability with a higher c(mHtt) requirement for aggregation, consistent with the fact that longer polyQ repeats result in higher aggregation propensity and earlier age of disease onset for HD.[41, 58] Interestingly, the Q count in mHtt-97Q is about twice that of mHtt-46Q, which raises the possibility that aggregation might be driven by overall Q concentrations which are the same between mHtt-46Q and mHtt-97Q aggregates rather than c(mHtt).

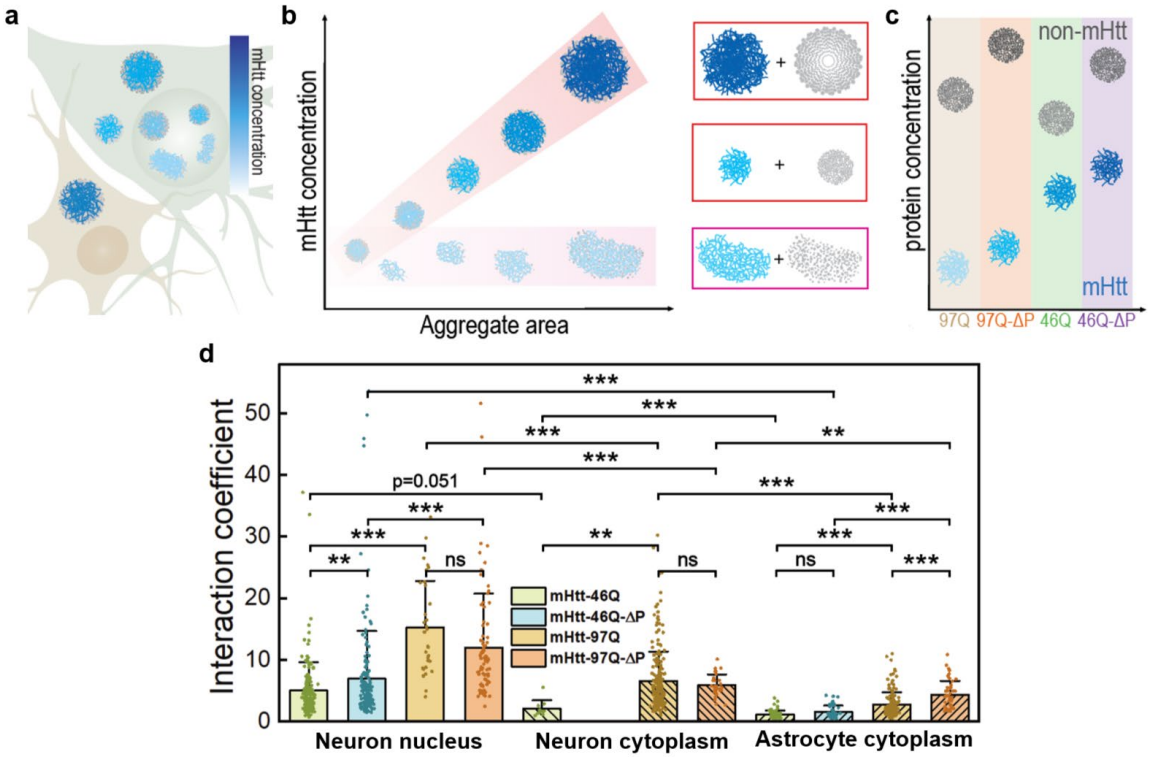

Figure 3.15. Proposed model of native neuronal polyQ aggregates. (a) Subcellular distribution of polyQ aggregates in neurons and astrocytes. (b) Our proposed quantitative model for polyQ aggregate formation and growth, pink: gel-like aggregates; coral: solid-like aggregates. Right panel shows the spatial distribution of mHtt proteins (Blue) and non-mHtt

proteins (Gray) within corresponding different types of polyQ aggregates. (c) Comparison of c(mHtt, blue) and c(non-mHtt, gray) for polyQ aggregates across different mHtt constructs, including mHtt-97Q, mHtt-46Q, mHtt-97Q-ΔP, and mHtt-46Q-ΔP. (d) Bar diagram of proposed interaction coefficients (ICs, defined as c(non-mHtt)/c(mHtt))) in different cell types, subcellular compartments across different mHtt constructs, including mHtt-97Q, mHtt-46Q, mHtt-97Q-ΔP, and mHtt-46Q-ΔP. ns: not significant, * $p<0.05$, ** $p < 0.01$, *** $p < 0.001$.

One major debate for poly-Q research is whether the formation of polyQ aggregates is harmful or protective. Recent evidence suggests that soluble mHtt oligomers interact more extensively with cellular proteins compared to mHtt in polyQ aggregates, making soluble oligomers more toxic. [5, 9, 11-14] However, previous experiments have struggled to accurately measure the total amount of non-mHtt proteins increasing within the aggregates due to the need for extensive polyQ extraction. Our platform offers a live-cell solution to this challenge. We hence introduced an interaction coefficient (IC), defined as the concentration ratios of non-mHtt over mHtt proteins (Figure 3.15c). This number quantitatively represents the "dragging" potential of each mHtt protein in pulling multiple copies of non-mHtt structural or functional proteins (Figure 3.15d), which may serve as a quantitative indicator of mHtt toxicity for its impact on cellular function.

ICs are higher in neurons compared to those in astrocytes, consistent with the notion that mHtt exerts higher toxicity in neurons than astrocytes (Figure 3.15d). Interestingly, ICs are much higher for nuclear aggregates than for cytoplasmic aggregates, further supporting the

hypothesis that nuclear aggregates are more toxic, as demonstrated previously by cellular function studies.[22, 60-61] As mentioned in the introduction, younger patients tend to exhibit nuclear aggregates, while older patients more commonly show cytoplasmic aggregates. Our findings highlight the distinct difference between nuclear and cytoplasmic aggregates, suggesting a potential pathway to better understand how the subcellular localization of polyQ aggregates may impact HD progression. In addition, as size increases, ICs decrease for most aggregates other than the gel-like ones (Figure 3.15d), suggesting that the growth of aggregates may help mitigate cellular functional loss, supporting the protective role of aggregate formation. Furthermore, ICs of mHtt-46Q-ΔP and mHtt-97Q-ΔP aggregates are higher than those of mHtt-46Q and mHtt-97Q (Figure 3.15d), consistent with increased toxicity after deleting the proline-rich region. All this evidence points to the potential effectiveness of ICs as a new quantitative indicator of aggregate toxicity, which may find utility in high-throughput screening of drug discovery for HD. Interestingly, we observed that gel-like nuclear aggregates exhibited significantly higher IC values compared to solid-like nuclear aggregates (Figure 3.14b), suggesting potential differences in the toxicity associated with these distinct aggregation types. Of course, more functional verifications are needed to confirm the role of IC, which is an important direction for our future work. Multidimensional information from tandem mass spectrometry with q-aggSRS should further confirm the protein identity in the non-mHtt protein pool.

3.11 Future directions

To advance our understanding of how various types of aggregates across different subcellular locations contribute to the pathological development of diseases, we plan to decipher their detailed molecular components through spatial proteomic profiling of polyQ aggregates. Previous studies with mass spectrometry (MS)-based proteomics analysis have highlighted the unprecedented complexity of protein compositions within polyQ aggregates and linked the widespread loss of functional protein to aggregate-mediated toxicity and HD pathogenesis.[12, 42, 50] However, current proteomic studies rely on *in vitro* purification methods, resulting in the loss of spatial information and limiting the ability to investigate heterogeneity across neurons.

The challenge would be how to specifically analyze the proteome within selected polyQ aggregates to potentially uncover why different types of aggregates exhibit distinct effects in patients. Enzyme-catalyzed proximity labeling,[62] such as APEX and BioID, provides a genetically encoded, high-selectivity labeling strategy for proteomics studies. However, these approaches face critical limitations: the large size of the enzymes can interfere with polyQ aggregate formation, and their diffuse labeling can result in non-specific background signals. More importantly, these strategies are unable to differentiate between polyQ aggregate types as defined by q-aggSRS imaging.

To correlate the spatial information with the proteomic data, we propose integrating photo-cross-linking (PXL) with precise light control to enable highly selective labeling for spatial proteomics.[63-66] Diazirine, a widely used functional group in photochemistry, will generate highly reactive carbenes for proximity labeling upon UV light irradiation. Notably, diazirine

is small and minimally disruptive to the native structure of polyQ aggregates, and diazirine's photoreaction can also be triggered using a 405 nm laser, offering versatility in experimental setups and enabling the combination with q-aggSRS.

We will use a probe that combines diazirine with an MS-detectable mass tag[67], allowing for the sensitive detection in MS analysis. Another challenge involves optimizing MS sample preparation procedures since the imaging-based PXL throughput is inherently limited, largely increasing the burden on MS analysis. The isolation of individual target cells from cultures and the removal of non-labeled cells using techniques such as fluorescence-activated cell sorting (FACS) and Photostick[68] would significantly reduce background signals and largely enhance the detection sensitivity in PXL-MS. [69]

By providing insights into how differences in protein components drive polyQ aggregate heterogeneity, this work seeks to establish a direct link between aggregate composition and its impact on cellular processes. These findings could help identify molecular markers of disease progression and pinpoint potential therapeutic targets for intervention, advancing our understanding of the molecular mechanisms underlying polyQ aggregate-mediated toxicity and dysfunction. Furthermore, this approach holds promise to evolve into a versatile platform for studying spatial proteomics guided by optical microscopy. With its exceptionally high spatial resolution and precise subcellular localization control, this platform could be broadly applicable to a variety of biologically relevant investigations, driving innovations in proteomics and beyond.

3.12 Materials and Methods

Primary neuron culture

Primary rat hippocampal neurons were isolated from neonatal Sprague–Dawley rat (CD (Sprague–Dawley) IGS rat, Charles River) pups with a protocol (IA22-1835) approved by Caltech's Institutional Animal Care and Use Committee (IACUC). The brains were dissected from the skull and placed into a 10-cm Petri dish with ice-chilled Hanks' balanced salt solution (Gibco). The hippocampus was isolated from the brains under a dissection scope, cut into small pieces (~0.5 mm), and incubated with 5 ml of Trypsin-EDTA (0.25%, Gibco) at 37 °C with 5% CO_2 for 15 min. The Trypsin-EDTA liquid was aspirated and replaced with 2 ml of DMEM containing 10% FBS to stop the digestion. The tissue fragments were moved into 2 ml of neuronal culture medium (Neurobasal A medium, B-27 supplement, 2 mM GlutaMAX supplement, Thermos Fisher, and 1× penicillin-streptomycin) and dispersed by repeated pipetting several times. The supernatant was collected and further diluted by neuronal culture medium to a final cell density of 9 × 104 cells ml^{-1}. A 0.7-ml volume of cell suspension was added to each well of a 24-well plate on coated 12-mm circular cover glass. For pre-coating, sterile 12-mm circular cover glass was incubated with 100 μg/ml poly-d-lysine (Sigma) solution at 37 °C with 5% CO_2 for 24 h in a 24-well plate. The 12-mm circular cover glass was washed twice with ddH_2O and incubated with 10 μg/ml laminin mouse protein (Gibco) solution at 37 °C with 5% CO_2 overnight. Thereafter, the 12-mm circular cover glass was washed twice with ddH_2O and allowed to dry at room temperature inside a biosafety cabinet. Half of the neuron culture medium was replaced with fresh medium every four days. All the medium changes were done after pre-heating to 37°C.

AAV packaging

We used the pAAV backbone for building plasmid constructs. The inserted gene was a mutant HTT gene encoding the exon 1 of mHtt protein. The model mHtt protein contains an N-terminal 17 a.a. fragments (MATLEKLMKAFESLKSF), the polyQ tract (46Q or 97Q), and the proline-rich region PPPPPPPPPPPQLPQPPPQAQPLLPQP QPPPPPPPPPPGP). For mHtt protein without EGFP, the C-terminal is AVAEEPL HRP*, while for mHtt-EGFP protein, there is a linker (AVAEEPLHRPGSSPVAT) between mHtt and EGFP (MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFIC TTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDD GNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQ KNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNE KRDHMVLLEFVTAAGITLGMDELYK*). The expression of mHtt protein was initiated by either a general promoter, CMV (cytomegalovirus) promoter or a neuron-specific promoter hSYN promoter (Human synapsin 1 gene promoter). The plasmids were sent to the CLOVER center at Caltech to produce AAV-DJ vectors. The typical titer is around 1E+13 genome copies (GCs)/mL with a normal yield of AAV vectors.

AAV transduction and medium change

AAV vectors were added to neurons at DIV 6-8 for about 5E9 GCs per well for the incubation of about 24 hours. A full-medium change from the normal medium to the medium with Gln-d_5 or Gln-d_5+Val-d_8 was performed after the AAV incubation, and cells were incubated with the new medium for 3-5 days. Gln-d_5 neuron culture medium was made by

replacing GlutaMAX with 2 mM Gln-d_5 (Cambridge Isotope Laboratories) in Neurobasal A medium which is glutamine-deficient. Gln-d_5 and Val-d_8 complete neuron culture medium was made from an amino-acid-deficient Neurobasal powder (BioTrend) with adding other normal amino acids and Gln-d_5 & Val-d_8 (Cambridge Isotope Laboratories) corresponding to the concentrations listed in Neurobasal A medium formula. For the pulse-chase experiment, the full-medium change from the normal medium to the medium with Gln-d_5 was done after the first round of SRS imaging. Cells were incubated with the new medium for an additional 8 h before the second round of SRS imaging. All the medium changes were done after pre-heating to 37°C.

Stimulated Raman Scattering (SRS) Microscopy

A picoEmerald laser system (Applied Physics and Electronics) was used as the light source for SRS microscopy. Briefly, it produces 2 ps pump (tunable from 770 nm – 990 nm, bandwidth 0.7 nm, spectral bandwidth ~ 10 cm^{-1}) and Stokes (1032 nm, spectral bandwidth 10 cm^{-1}) pulses with 80 MHz repetition rate. Stokes beam is modulated at 20 MHz by an internal electro-optic modulator. The spatially and temporally overlapped Pump and Stokes beams are introduced into an inverted laser-scanning microscope (FV3000, Olympus), and then focused onto the sample by a 25X water objective (XLPLN25XWMP, 1.05 N.A., Olympus). Transmitted Pump and Stokes beams are collected by a high N.A. condenser lens (oil immersion, 1.4 N.A., Olympus) and pass through a bandpass filter (893/209 BrightLine, 25mm, Semrock) to filter out Stokes beam. A large area (10×10 mm) Si photodiode (S3590-09, Hamamatsu) is used to measure the pump beam intensity. A 64 V reverse-biased DC

voltage is applied to the photodiode to increase the saturation threshold and reduce response time. The output current is terminated by a 50Ω terminator and pre-filtered by a 19.2-23.6-MHz band-pass filter (BBP-21.4+, MiniCircuits) to reduce laser and scanning noise. The signal is then demodulated by a lock-in amplifier (SR844, Stanford Research Systems) at the modulation frequency. The in-phase X output is fed back to the Olympus IO interface box (FV30-ANALOG) of the microscope. 30 μs time constant is set for the lock-in amplifier. Correspondingly, 80 μs pixel dwell time is used, which gives a speed of 21.3 s/frame for a 512-by-512-pixel image, with two frame-averaging. Laser powers are monitored throughout image acquisition by an internal power meter and power fluctuations are controlled within 1%. 16-bit greyscale images were acquired by Olympus Fluoview 3000 software.

Right before imaging, the cover glass with neurons was taken out to make an imaging chamber (Grace Bio-Labs SecureSeal™ imaging spacer, diam. × thickness 9 mm × 0.12 mm) filled with the original medium for SRS imaging. For the channel of CH_3, CH_2, CD on (for Gln-d_5), CD off (for Gln-d_5), CD on (for Val-d_8), and CD off (for unmixing), the wavelengths of pump lasers for SRS are 791.8 nm, 797.8 nm, 843.6 nm, 852.8 nm, 850.5 nm, and 855 nm. For hyperspectral measurements, the wavelengths of the pump laser were tuned from 835 nm to 855 nm with a 0.5 nm interval. All images were processed using ImageJ software. Corresponding off-resonance images were subtracted.

Materials and sample preparation for VISTA

Sodium acrylate (SA), N,N′-methylenebisacrylamide (BIS, 2%), ammonium persulfate (APS), tetramethylethylenediamine (TEMED), and sodium dodecyl sulfate (SDS) were

purchased from Sigma-Aldrich. Acrylamide (AA) was purchased from Ambeed, Inc. The 1.0 M Tris solution was obtained from Biosolve, and nuclease-free water was from Ambion. The expansion protocol follows our previously published methodology. Stock solutions were prepared and stored under the following conditions: an incubation solution containing 30% AA in 4% paraformaldehyde (PFA) in 1× phosphate-buffered saline (PBS) was stored at 4°C. A gelation solution consisting of 7% SA, 20% AA, and 0.1% BIS in 1× PBS was stored at −20°C. APS and TEMED were dissolved in nuclease-free water to create a 10% (w/w) solution and stored at −20°C.

For hydrogel embedding, cell samples were incubated in a 30% AA and 4% PFA solution without prior fixation. This incubation was carried out for 16 hours at 37°C, followed by a wash with gelation solution at 4°C. The gelation solution, APS, and TEMED were thawed and kept at 4°C until use. Coverslips containing cells were placed at the bottom of a 4°C gelation chamber with the cell side facing up. A mixture of 90% gelation solution, 5% APS solution, and 5% TEMED solution was prepared and immediately pipetted onto the sample to ensure complete immersion before covering it with a glass slide. The chamber was maintained at 4°C for 1 minute before being transferred to a humidified incubator for 1 hour at 37°C.

Post-gelation, the coverslips with gels were incubated in a denaturing buffer (200 mM SDS, 200 mM NaCl, and 50 mM Tris in nuclease-free water, pH 8) in petri dishes for 15 minutes at room temperature. Subsequently, they were transferred to 2 ml Eppendorf centrifuge tubes and heated in 1.5 ml of denaturing buffer at 95°C for 30 minutes. Initial expansions were

performed immediately after denaturation at room temperature in water, with the water changed every hour for three times. The expanded gel was then shrunk in PBS overnight in the dark at 4°C. Under these conditions, the cells underwent a 2-fold linear isotropic expansion.

Statistical analysis

All the statistical analyses were done using OriginPro 2024b software. All data are shown as means ± SEM. One-way analyses of variance (ANOVAs) followed by Tukey's multiple comparisons test were used for single-variable comparisons. The statistical significance level was defined as: ns not significant, * $p<0.05$, ** $p < 0.01$, *** $p < 0.001$.

Identification of live neurons and astrocytes for aggregate analysis

To image aggregates specifically in live neurons and astrocytes, we adopted Neuron-NeuO staining in live neuronal co-cultures for experimental analysis. NeuroFluor™ NeuO (STEMCELL) is a green membrane-permeable fluorescent probe that selectively labels primary and pluripotent stem cell-derived neurons in the presence of other brain cells.[54] The neurons were incubated with 0.25 μM NeuO for 1 hour and then were changed back to normal medium without NeuO for another 2 hours. In this case, neurons present clear green fluorescence with corresponding images in the CH_3 SRS channel; astrocytes present as cells in the CH_3 SRS channel but have no green fluorescence contrast (Figure 3.4).

We further benchmarked the quantitative results with other cell distinguishing strategies and found good consistency (Figure 3.7a). First, we used the neuron-specific promoter, hSYN.

Second, we added 2 uM 1-β-D-arabinofuranosylcytosine (Ara-C) to the culture medium at DIV 3 to obtain pure neuron cultures without astrocytes (major species other than neurons) or oligodendrocytes (minor species) (Figure 3.4a). To specifically image aggregates in astrocytes, we also performed pure astrocyte cultures (Figure 3.4b), which were obtained by using Basal Eagle Media containing 10% fetal bovine serum (FBS, Invitrogen) and GlutaMAX supplement.

Calculation of mHtt/non-mHtt concentrations

We use 10 mM Gln-d_5 solution as a standard to calibrate the SRS signal of CD bonds to the concentration of CD bonds (Figure 3.2). We confirmed that SRS signals of CD bonds have a linear relationship with Gln-d_5 concentration, and the CH Raman cross-section is the same as CD (confirmed by DMSO-d_6 and DMSO comparison). All signals are processed with power normalization.

For non-mHtt proteins, the average length of proteins in eukaryotes is 438 amino acids.[43] On average there are 6 C-H bonds of each amino acid, with the consideration of relative abundance. To calculate the concentration of mHtt proteins (c(mHtt)) and non-mHtt proteins (c(non-mHtt)) within the same aggregates from measured CD signals and the CH/CD ratios, we can write the following two Equations (3.1) and (3.2):

$$c(CD) = 5 * c(mHtt) * n(Q) + 5 * c(non - mHtt) * 438 * X\% \quad (3.1)$$

$$c(CH) = 6 * c(mHtt) * n(non - Q\ a.a.) + 6 * c(non - mHtt) * 438 * (1 - X\%) \quad (3.2)$$

Here $X\%$ indicates the incorporation ratio for Gln-d_5 in the regular proteome of non-Htt proteins. We determine this number experimentally by incubating neurons with Gln-d_5 medium without AAV transfection. After 0-6 days' incubation, we took SRS images at both CH_3 and CD channels to determine $X\%$. We found that the CH/CD ratios plateau from day 3 to day 6, shown below in Figure 3.6, with an average CH/CD ratio of 55. We confirmed through d_7-glucose labeling that *de novo* glutamine synthesis is negligible. So we have the following equation for $X\%$ as:

$$\frac{6 * (1 - X\%)}{5 * X\%} = 55$$

Yielding $X\%$ of 2.1% after 3-6 days of Gln-d_5 incubation, consistent with the number that Gln-d_5 makes up 4.2% of the overall proteome.[44] We use 2.1% for all calculations in Eq (3.1) and Eq (3.2) for c(mHtt) and c(non-mHtt).

For different mHtt-exon1 constructs, $n(Q)$, the number of glutamines, and $n(non-Q\ a.a.)$, the number of non-glutamine amino acids, are exactly known and shown in the table below in Table 3.1. Taking these numbers and each measured CH/CD ratio back into Eq. (1) and (2) can solve for c(mHtt) and c(non-mHtt), corresponding to the mHtt concentration and non-mHtt cytosolic protein concentrations within the same aggregates.

Table 3.1: Number of glutamine (Q) and non-glutamine amino acid (non-Q a.a.) in each mHtt construct sequence.

	97Q	97Q-ΔP	97Q-EGFP	46Q	46Q-ΔP	97Q-EGFP
$n(Q)$	103	97	111	52	46	60
$n(non-Q\ a.a.)$	61	27	299	61	27	299

Unmixing for two-color imaging of Gln-d_5 and Val-d_8

Two images were taken at the channel of 2164 cm^{-1} (pump laser 843.6 nm) and 2068 cm^{-1} (pump laser 850.5 nm). Background at the channel of 2006 cm^{-1} (pump laser 855 nm) was subtracted before further analysis. According to the spectra of aggregates incorporated with Gln-d_5 and Val-d_8 (Figure 3.9, bottom), we can obtain two linear equations below.

$$SRS_{2164} = SRS_{Gln-d_5} + 0.6 * SRS_{Val-d_8}$$

$$SRS_{2068} = 0.1 * SRS_{Gln-d_5} + SRS_{Val-d_8}$$

So we can obtain the equations for unmixing to get pure signals from Gln-d_5 and Val-d_8 labeling channels as:

$$SRS_{Gln-d_5} = \frac{10 * SRS_{2164} - 6 * SRS_{2068}}{9.4}$$

$$SRS_{Val-d_8} = \frac{10 * SRS_{2068} - SRS_{2164}}{9.4}$$

3.13 Reference

1.Hands, S.; Sinadinos, C.; Wyttenbach, A., Polyglutamine gene function and dysfunction in the ageing brain. *Biochim Biophys Acta* **2008,** *1779* (8), 507-21.
2.Fan, H. C.; Ho, L. I.; Chi, C. S.; Chen, S. J.; Peng, G. S.; Chan, T. M.; Lin, S. Z.; Harn, H. J., Polyglutamine (PolyQ) diseases: genetics to treatments. *Cell Transplant* **2014,** *23* (4-5), 441-58.
3.Lieberman, A. P.; Shakkottai, V. G.; Albin, R. L., Polyglutamine repeats in neurodegenerative diseases. *Annu Rev Pathol* **2019,** *14*, 1-27.
4.Bates, G., Huntingtin aggregation and toxicity in Huntington's disease. *Lancet* **2003,** *361* (9369), 1642-4.

5.Arrasate, M.; Mitra, S.; Schweitzer, E. S.; Segal, M. R.; Finkbeiner, S., Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **2004,** *431* (7010), 805-810.
6.Ding, Q.; Lewis, J. J.; Strum, K. M.; Dimayuga, E.; Bruce-Keller, A. J.; Dunn, J. C.; Keller, J. N., Polyglutamine expansion, protein aggregation, proteasome activity, and neural survival. *J Biol Chem* **2002,** *277* (16), 13935-42.
7.Khare, S. D.; Ding, F.; Gwanmesia, K. N.; Dokholyan, N. V., Molecular origin of polyglutamine aggregation in neurodegenerative diseases. *PLoS Comput Biol* **2005,** *1* (3), 230-5.
8.Hoffner, G.; Soues, S.; Djian, P., Aggregation of expanded Huntingtin in the brains of patients with Huntington disease. *Prion* **2007,** *1* (1), 26-31.
9.Arrasate, M.; Finkbeiner, S., Protein aggregates in Huntington's disease. *Exp Neurol* **2012,** *238* (1), 1-11.
10.DiFiglia, M.; Sapp, E.; Chase, K. O.; Davies, S. W.; Bates, G. P.; Vonsattel, J. P.; Aronin, N., Aggregation of Huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* **1997,** *277* (5334), 1990-1993.
11.Leitman, J.; Ulrich Hartl, F.; Lederkremer, G. Z., Soluble forms of polyQ-expanded huntingtin rather than large aggregates cause endoplasmic reticulum stress. *Nat Commun* **2013,** *4*, 2753.
12.Kim, Y. E.; Hosp, F.; Frottin, F.; Ge, H.; Mann, M.; Hayer-Hartl, M.; Hartl, F. U., Soluble oligomers of polyQ-expanded Huntingtin target a multiplicity of key cellular factors. *Mol Cell* **2016,** *63* (6), 951-64.
13.Todd, T. W.; Lim, J., Aggregation formation in the polyglutamine diseases: protection at a cost? *Mol Cells* **2013,** *36* (3), 185-94.
14.Berglund, L. L.; Hao, X.; Liu, B.; Grantham, J.; Nystrom, T., Differential effects of soluble and aggregating polyQ proteins on cytotoxicity and type-1 myosin-dependent endocytosis in yeast. *Sci Rep* **2017,** *7* (1), 11328.
15.Siu, H. W.; Stritt, P.; Zhao, H.; Hauser, K., PolyQ aggregation studied by model peptides with intrinsic tryptophan fluorophores. *Biophys Chem* **2022,** *284*, 106782.
16.Suhr, S. T.; Senut, M.-C.; Whitelegge, J. P.; Faull, K. F.; Cuizon, D. B.; Gage, F. H., Identities of sequestered proteins in aggregates from cells with induced polyglutamine expression. *Journal of Cell Biology* **2001,** *153* (2), 283-294.
17.Gruber, A.; Hornburg, D.; Antonin, M.; Krahmer, N.; Collado, J.; Schaffer, M.; Zubaite, G.; Luchtenborg, C.; Sachsenheimer, T.; Brugger, B.; Mann, M.; Baumeister, W.; Hartl, F. U.; Hipp, M. S.; Fernandez-Busnadiego, R., Molecular and structural architecture of polyQ aggregates in yeast. *Proc Natl Acad Sci U S A* **2018,** *115* (15), E3446-E3453.
18.Park, S. H.; Kukushkin, Y.; Gupta, R.; Chen, T.; Konagai, A.; Hipp, M. S.; Hayer-Hartl, M.; Hartl, F. U., PolyQ proteins interfere with nuclear degradation of cytosolic proteins by sequestering the Sis1p chaperone. *Cell* **2013,** *154* (1), 134-45.
19.Ripaud, L.; Chumakova, V.; Antonin, M.; Hastie, A. R.; Pinkert, S.; Korner, R.; Ruff, K. M.; Pappu, R. V.; Hornburg, D.; Mann, M.; Hartl, F. U.; Hipp, M. S., Overexpression of Q-rich prion-like proteins suppresses polyQ cytotoxicity and alters the polyQ interactome. *Proc Natl Acad Sci U S A* **2014,** *111* (51), 18219-24.

20.Onodera, O.; Rose, A. D.; Tsuji, S.; Vance, J. M.; Strittmatter, W. J.; Burke, J. R., Toxicity of expanded polyglutamine-domain proteins in Escherichia coli. *FEBS Letters* **1996,** *399* (1-2), 135-139.
21.Kim, S.; Nollen, E. A.; Kitagawa, K.; Bindokas, V. P.; Morimoto, R. I., Polyglutamine protein aggregates are dynamic. *Nat Cell Biol* **2002,** *4* (10), 826-31.
22.Iwata, A.; Christianson, J. C.; Bucci, M.; Ellerby, L. M.; Nukina, N.; Forno, L. S.; Kopito, R. R., Increased susceptibility of cytoplasmic over nuclear polyglutamine aggregates to autophagic degradation. *Proceedings of the National Academy of Sciences* **2005,** *102* (37), 13135-13140.
23.Li, L.; Liu, H.; Dong, P.; Li, D.; Legant, W. R.; Grimm, J. B.; Lavis, L. D.; Betzig, E.; Tjian, R.; Liu, Z., Real-time imaging of Huntingtin aggregates diverting target search and gene transcription. *Elife* **2016,** *5*.
24.Peskett, T. R.; Rau, F.; O'Driscoll, J.; Patani, R.; Lowe, A. R.; Saibil, H. R., A liquid to solid phase transition underlying pathological Huntingtin exon1 aggregation. *Mol Cell* **2018,** *70* (4), 588-601 e6.
25.Pandey, N. K.; Varkey, J.; Ajayan, A.; George, G.; Chen, J.; Langen, R., Fluorescent protein tagging promotes phase separation and alters the aggregation pathway of huntingtin exon-1. *J Biol Chem* **2024,** *300* (1), 105585.
26.Bauerlein, F. J. B.; Saha, I.; Mishra, A.; Kalemanov, M.; Martinez-Sanchez, A.; Klein, R.; Dudanova, I.; Hipp, M. S.; Hartl, F. U.; Baumeister, W.; Fernandez-Busnadiego, R., In situ architecture and cellular interactions of polyQ inclusions. *Cell* **2017,** *171* (1), 179-187 e10.
27.Galaz-Montoya, J. G.; Shahmoradian, S. H.; Shen, K.; Frydman, J.; Chiu, W., Cryo-electron tomography provides topological insights into mutant huntingtin exon 1 and polyQ aggregates. *Commun Biol* **2021,** *4* (1), 849.
28.Eisele, Y. S.; Monteiro, C.; Fearns, C.; Encalada, S. E.; Wiseman, R. L.; Powers, E. T.; Kelly, J. W., Targeting protein aggregation for the treatment of degenerative diseases. *Nat Rev Drug Discov* **2015,** *14* (11), 759-80.
29.Minakawa, E. N.; Nagai, Y., Protein Aggregation Inhibitors as Disease-Modifying Therapies for Polyglutamine Diseases. *Front Neurosci* **2021,** *15*, 621996.
30.Llamas, E.; Koyuncu, S.; Lee, H. J.; Wehrmann, M.; Gutierrez-Garcia, R.; Dunken, N.; Charura, N.; Torres-Montilla, S.; Schlimgen, E.; Mandel, A. M.; Theile, E. B.; Grossbach, J.; Wagle, P.; Lackmann, J. W.; Schermer, B.; Benzing, T.; Beyer, A.; Pulido, P.; Rodriguez-Concepcion, M.; Zuccaro, A.; Vilchez, D., In planta expression of human polyQ-expanded huntingtin fragment reveals mechanisms to prevent disease-related protein aggregation. *Nat Aging* **2023,** *3* (11), 1345-1357.
31.Freudiger, C. W.; Min, W.; Saar, B. G.; Lu, S.; Holtom, G. R.; He, C.; Tsai, J. C.; Kang, J. X.; Xie, X. S., Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy. *Science* **2008,** *322* (5909), 1857-61.
32.Hu, F.; Shi, L.; Min, W., Biological imaging of chemical bonds by stimulated Raman scattering microscopy. *Nat. Methods* **2019,** *16* (9), 830-842.
33.Du, J.; Wang, H.; Wei, L., Bringing vibrational imaging to chemical biology with molecular probes. *ACS Chem Biol* **2022,** *17* (7), 1621-1637.

34.Devitt, G.; Howard, K.; Mudher, A.; Mahajan, S., Raman spectroscopy: An emerging tool in neurodegenerative disease research and diagnosis. *ACS Chem Neurosci* **2018,** *9* (3), 404-420.
35.Miao, K.; Wei, L., Live-cell imaging and quantification of polyQ aggregates by stimulated Raman scattering of selective deuterium labeling. *ACS Cent Sci* **2020,** *6* (4), 478-486.
36.Khakh, B. S.; Beaumont, V.; Cachope, R.; Munoz-Sanjuan, I.; Goldman, S. A.; Grantyn, R., Unravelling and exploiting astrocyte dysfunction in Huntington's disease. *Trends Neurosci* **2017,** *40* (7), 422-437.
37.Palpagama, T. H.; Waldvogel, H. J.; Faull, R. L. M.; Kwakowsky, A., The role of microglia and astrocytes in Huntington's disease. *Front Mol Neurosci* **2019,** *12*, 258.
38.Michalik, A.; Van Broeckhoven, C., Pathogenesis of polyglutamine disorders: aggregation revisited. *Hum Mol Genet* **2003,** *12 Spec No 2*, R173-86.
39.Proskura, A. L.; Vechkapova, S. O.; Zapara, T. A.; Ratushniak, A. S., Protein–protein interactions of huntingtin in the hippocampus. *Molecular Biology* **2017,** *51* (4), 647-653.
40.Ghosh, R.; Wood-Kaczmar, A.; Dobson, L.; Smith, E. J.; Sirinathsinghji, E. C.; Kriston-Vizi, J.; Hargreaves, I. P.; Heaton, R.; Herrmann, F.; Abramov, A. Y.; Lam, A. J.; Heales, S. J.; Ketteler, R.; Bates, G. P.; Andre, R.; Tabrizi, S. J., Expression of mutant exon 1 huntingtin fragments in human neural stem cells and neurons causes inclusion formation and mitochondrial dysfunction. *FASEB J* **2020,** *34* (6), 8139-8154.
41.Bravo-Arredondo, J. M.; Kegulian, N. C.; Schmidt, T.; Pandey, N. K.; Situ, A. J.; Ulmer, T. S.; Langen, R., The folding equilibrium of huntingtin exon 1 monomer depends on its polyglutamine tract. *J Biol Chem* **2018,** *293* (51), 19613-19623.
42.Riguet, N.; Mahul-Mellier, A. L.; Maharjan, N.; Burtscher, J.; Croisier, M.; Knott, G.; Hastings, J.; Patin, A.; Reiterer, V.; Farhan, H.; Nasarov, S.; Lashuel, H. A., Nuclear and cytoplasmic huntingtin inclusions exhibit distinct biochemical composition, interactome and ultrastructural properties. *Nat Commun* **2021,** *12* (1), 6579.
43.Milo, R., What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays* **2013,** *35* (12), 1050-5.
44.Kozlowski, L. P., Proteome-pI: proteome isoelectric point database. *Nucleic Acids Res* **2017,** *45* (D1), D1112-D1116.
45.Yang, J.; Yang, X., Phase transition of Huntingtin: factors and pathological relevance. *Frontiers in Genetics* **2020,** *11*.
46.Qian, C.; Miao, K.; Lin, L. E.; Chen, X.; Du, J.; Wei, L., Super-resolution label-free volumetric vibrational imaging. *Nat Commun* **2021,** *12* (1), 3648.
47.Lin, L. E.; Miao, K.; Qian, C.; Wei, L., High spatial-resolution imaging of label-free in vivo protein aggregates by VISTA. *Analyst* **2021,** *146* (13), 4135-4145.
48.Sahl, S. J.; Weiss, L. E.; Duim, W. C.; Frydman, J.; Moerner, W. E., Cellular inclusion bodies of mutant huntingtin exon 1 obscure small fibrillar aggregate species. *Sci Rep* **2012,** *2*, 895.
49.Zhao, D. Y.; Bauerlein, F. J. B.; Saha, I.; Hartl, F. U.; Baumeister, W.; Wilfling, F., Autophagy preferentially degrades non-fibrillar polyQ aggregates. *Mol Cell* **2024,** *84* (10), 1980-1994 e8.
50.Hosp, F.; Gutierrez-Angel, S.; Schaefer, M. H.; Cox, J.; Meissner, F.; Hipp, M. S.; Hartl, F. U.; Klein, R.; Dudanova, I.; Mann, M., Spatiotemporal proteomic profiling of

Huntington's disease inclusions reveals widespread loss of protein function. *Cell Rep* **2017,** *21* (8), 2291-2303.
51.Ramdzan, Y. M.; Trubetskov, M. M.; Ormsby, A. R.; Newcombe, E. A.; Sui, X.; Tobin, M. J.; Bongiovanni, M. N.; Gras, S. L.; Dewson, G.; Miller, J. M. L.; Finkbeiner, S.; Moily, N. S.; Niclis, J.; Parish, C. L.; Purcell, A. W.; Baker, M. J.; Wilce, J. A.; Waris, S.; Stojanovski, D.; Bocking, T.; Ang, C. S.; Ascher, D. B.; Reid, G. E.; Hatters, D. M., Huntingtin inclusions trigger cellular quiescence, deactivate apoptosis, and lead to delayed necrosis. *Cell Rep* **2017,** *19* (5), 919-927.
52.Taran, A. S.; Shuvalova, L. D.; Lagarkova, M. A.; Alieva, I. B., Huntington's disease-An outlook on the interplay of the HTT protein, microtubules and actin cytoskeletal components. *Cells* **2020,** *9* (6).
53.Wang, J.-Y.; Liu, Y.-J.; Zhang, X.-L.; Liu, Y.-H.; Jiang, L.-L.; Hu, H.-Y., PolyQ-expanded ataxin-2 aggregation impairs cellular processing-body homeostasis via sequestering the RNA helicase DDX6. *Journal of Biological Chemistry* **2024,** *300* (7).
54.Er, J. C.; Leong, C.; Teoh, C. L.; Yuan, Q.; Merchant, P.; Dunn, M.; Sulzer, D.; Sames, D.; Bhinge, A.; Kim, D.; Kim, S. M.; Yoon, M. H.; Stanton, L. W.; Je, S. H.; Yun, S. W.; Chang, Y. T., NeuO: a fluorescent chemical probe for live neuron labeling. *Angew Chem Int Ed Engl* **2015,** *54* (8), 2442-6.
55.Dehay, B.; Bertolotti, A., Critical role of the proline-rich region in Huntingtin for aggregation and cytotoxicity in yeast. *J Biol Chem* **2006,** *281* (47), 35608-15.
56.Elena-Real, C. A.; Sagar, A.; Urbanek, A.; Popovic, M.; Morato, A.; Estana, A.; Fournet, A.; Doucet, C.; Lund, X. L.; Shi, Z. D.; Costa, L.; Thureau, A.; Allemand, F.; Swenson, R. E.; Milhiet, P. E.; Crehuet, R.; Barducci, A.; Cortes, J.; Sinnaeve, D.; Sibille, N.; Bernado, P., The structure of pathogenic huntingtin exon 1 defines the bases of its aggregation propensity. *Nat Struct Mol Biol* **2023,** *30* (3), 309-320.
57.Guo, Q.; Bin, H.; Cheng, J.; Seefelder, M.; Engler, T.; Pfeifer, G.; Oeckl, P.; Otto, M.; Moser, F.; Maurer, M.; Pautsch, A.; Baumeister, W.; Fernandez-Busnadiego, R.; Kochanek, S., The cryo-electron microscopy structure of huntingtin. *Nature* **2018,** *555* (7694), 117-120.
58.Mier, P.; Andrade-Navarro, M. A., Between interactions and aggregates: The polyQ balance. *Genome Biol Evol* **2021,** *13* (11).
59.Caterino, M.; Squillaro, T.; Montesarchio, D.; Giordano, A.; Giancola, C.; Melone, M. A. B., Huntingtin protein: A new option for fixing the Huntington's disease countdown clock. *Neuropharmacology* **2018,** *135*, 126-138.
60.Havel, L. S.; Li, S.; Li, X. J., Nuclear accumulation of polyglutamine disease proteins and neuropathology. *Mol Brain* **2009,** *2*, 21.
61.Woerner, A. C.; Frottin, F.; Hornburg, D.; Feng, L. R.; Meissner, F.; Patra, M.; Tatzelt, J.; Mann, M.; Winklhofer, K. F.; Hartl, F. U.; Hipp, M. S., Cytoplasmic protein aggregates interfere with nucleocytoplasmic transport of protein and RNA. *Science* **2016,** *351* (6269), 173-176.
62.Qin, W.; Cho, K. F.; Cavanagh, P. E.; Ting, A. Y., Deciphering molecular interactions by proximity labeling. *Nat Methods* **2021,** *18* (2), 133-143.
63.Tanaka, Y.; Bond, M. R.; Kohler, J. J., Photocrosslinkers illuminate interactions in living cells. *Mol Biosyst* **2008,** *4* (6), 473-80.
64.Brieke, C.; Rohrbach, F.; Gottschalk, A.; Mayer, G.; Heckel, A., Light-controlled tools. *Angew Chem Int Ed Engl* **2012,** *51* (34), 8446-76.

65.Pham, N. D.; Parker, R. B.; Kohler, J. J., Photocrosslinking approaches to interactome mapping. *Curr Opin Chem Biol* **2013,** *17* (1), 90-101.
66.Lundberg, E.; Borner, G. H. H., Spatial proteomics: a powerful discovery tool for cell biology. *Nat Rev Mol Cell Biol* **2019,** *20* (5), 285-302.
67.Gutierrez, C.; Salituro, L. J.; Yu, C.; Wang, X.; DePeter, S. F.; Rychnovsky, S. D.; Huang, L., Enabling Photoactivated Cross-Linking Mass Spectrometric Analysis of Protein Complexes by Novel MS-Cleavable Cross-Linkers. *Mol Cell Proteomics* **2021,** *20*, 100084.
68.Chien, M. P.; Werley, C. A.; Farhi, S. L.; Cohen, A. E., Photostick: a method for selective isolation of target cells from culture. *Chem Sci* **2015,** *6* (3), 1701-1705.
69.Zhu, Y.; Akkaya, K. C.; Ruta, J.; Yokoyama, N.; Wang, C.; Ruwolt, M.; Lima, D. B.; Lehmann, M.; Liu, F., Cross-link assisted spatial proteomics to map sub-organelle proteomes and membrane protein topologies. *Nat Commun* **2024,** *15* (1), 3290.

Chapter 4

MAPPING NEURONAL METABOLISM COUPLED WITH NEURONAL ACTIVITY

This chapter is temporarily embargoed.