論文 / 著書情報
Article / Book Information

題目(和文)	PR繰り返し単位を持つポリペプチドとRNAによるLLPS液滴の構成及び固液界面における拡散の分子理解
Title(English)	Molecular understanding on LLPS droplets consisting of poly(PR) dipeptide repeats and RNA and its diffusion at solid/liquid interface
著者(和文)	ChenChen
Author(English)	Chen Chen
出典(和文)	学位:博士（工学）, 学位授与機関:東京工業大学, 報告番号:甲第12037号, 授与年月日:2021年6月30日, 学位の種別:課程博士, 審査員:早水 裕平,VACHA　MARTIN,松本 英俊,石川 謙,道信 剛志,金蔵 孝介
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12037号, Conferred date:2021/6/30, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

学位論文　　博士(工学)

Molecular understanding on LLPS droplets consisting of poly(PR) dipeptide repeats and RNA and its diffusion at solid/liquid interface

PR 繰り返し単位を持つポリペプチドと RNA による
LLPS 液滴の構成及び固液界面における拡散の分子理解

2021 年度

東京工業大学大学院　理工学研究科

有機・高分子物質専攻

Chen Chen

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

It would have been difficult for me to write this thesis without the support of many people.

Firstly, I am very grateful for my supervisor, Prof. Yuhei Hayamizu, for his instructive advice and useful suggestions in the proposal stages and helped make the thesis a reality. Moreover, during my doctoral studying periods, he encouraged me countlessly not only for my research but favored me to adapt myself well to the living in Japan.

High tribute shall be paid to Prof. Kohsuke Kanekura from Tokyo Medical University, whose profound knowledge of LLPS and earnest attitude towards research left me deep impression. Meanwhile, he also did his best to guide my research and shared me with his precious experience in biomedical experiments and even postdoctoral position hunting. Besides, I also want to express my thanks to Prof. Kohei Takahashi in Department of Chemistry at Tokyo Institute of Technology, who kindly trained me doing ITC experiments and provided me with his good opinions in finding postdoctoral positions also.

I owe a special debt of gratitude to all members in Hayamizu Lab. Ms. Wei Luo, and Mr. Han Jia, keep forward towards your goals and regard each challenge as a chance to become stronger; Mr. Robert Jose Ccorahua Santo, whose concentrative attitude towards research leaves me very deep impression, hope you achieve your expectations; Mr. Kazunori Motai and Mr. Chishu Homma, hope you be more confident in the research; Mr. Yoshiki Nakamura, thank you for the kind support for my experiments; Ms. Vanessa Seudo Dimo, who is very diligent in the research and also encourages me a lot during the whole period; Ms. Rio Takizama, hope you find a good job; Ms. Yui Yamazaki, many thanks to the continuous assistance and encouragement to me. I still need to say thank you to all other new members, including Mr. Yuxiang Mao, Mr. Shogo Hayashi, Mr. Shun Shimizu and Mr. Kai Saito.

I am of course indebted many former members in the lab: Ms. Tomoko Oonishi, Dr.

Linhao Sun, Dr. Peiying Li, Dr. Takakazu Seki, Mr. Hironaga Noguchi, Ms. Sayaka Tezuka, Ms. Tsukiiwa Minaro, Ms. Wenbi Lai Hirakawa, Mr. Ryan Yoo and Mr. Abubakar Mohamed and other all graduates. Wish all of you have a bright future in the life.

On the personal side, I would finally like to express my deep gratitude to my beloved parents who have been helping me out of difficulties and supporting me without a word of complaint. They always taught me to follow my dreams and never settle for less.

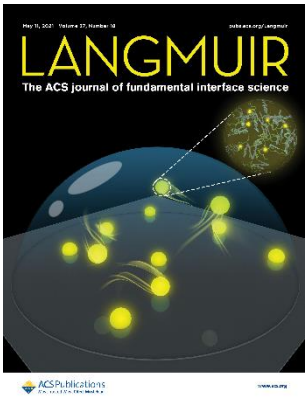
The doctoral studying comes to end fluently. I cannot finish it without all kind and great assistance from all supporters above. I will regard what I have learned during this unforgettable period as the driving force to go ahead in the future.

PUBLICATIONS AND CONFERENCES

Publications

1. **Chen Chen,** Yamanaka Yoshiaki, Ueda Koji, Li, Peiying, Harada Yuichiro, Tezuka, Sayaka, Narumi Satoshi, Sugimoto Masahiro, Kuroda Masahiko, Hayamizu Yuhei, Kanekura Kohsuke. Phase Separation and Toxicity of C9orf72 poly(PR) depends on alternate distribution of Arginine. (***Journal of Cell Biology***, under review, 2021)

2. **Chen Chen,** Peiying Li, Wei Luo, Yoshiki Nakamura, Vanessa Seudo Dimo, Kohsuke Kanekura, Yuhei Hayamizu. Diffusion of LLPS Droplets Consisting of Poly(PR) Dipeptide Repeats and RNA on Chemically Modified Glass Surface. ***Langmuir*** (2021).



3. Wei Luo, **Chen Chen,** Ivan V. Korendovych, Yuhei Hayamizu. Self-assembling peptide functionalized graphite to promote electrocatalytic activity as mimetic efficient peroxidase electrode. (under preparation, 2021)

4. Kazunori Motai, Takuma Narimatsu, **Chen Chen,** Yuhei Hayamizu. Oriented crystal growth of phenylalanine and a dipeptide by solution shearing. ***Journal of Materials Chemistry C*** 8, 8585–8591 (2020).

Conferences

1. **Chen Chen,** Yuhei Hayamizu, Kohsuke Kanekura. Molecular interactions of poly(PR) dipeptides in liquid-liquid phase separation. The 81st JSAP Autumn Meeting. 2020, Online, JAPAN. (Oral)

2. **Chen Chen,** Yuhei Hayamizu, Kohsuke Kanekura. Determinants of molecular interactions of poly(PR) dipeptides in liquid-liquid phase separation. 第 4 回先端光機能計測研究会プログラム, 2020, Online, JAPAN. (Oral)

3. **Chen Chen,** Peiying Li, Koishihara Nao, Kazunori Motai, Tomoko Ohnishi, Kohsuke Kanekura, Yuhei Hayamizu. Molecular mechanisms underlying cytotoxicity of C9orf72-encoding Arginine-rich dipeptides. 第 3 回先端光機能計測研究会プログラム, 2019, Tokyo, JAPAN. (Poster)

4. **Chen Chen,** Peiying Li, Kazunori Motai, Hironaga Noguchi, Yuhei Hayamizu. Self-assembly of amphiphilic tetrapeptides on 2D nanomaterials surface. 第 2 回先端光機能計測研究会プログラム, 2019, Ikaho, JAPAN. (Poster)

5. **Chen Chen,** Peiying Li, Kazunori Motai, Hironaga Noguchi, Yuhei Hayamizu. Self-assembly of amphiphilic tetrapeptides on 2D nanomaterials surface. 2nd A3 Foresight Meeting about Functionalization and Flexible Device Application of Atomic Scale Organic and Inorganic Material. 2019, Anhui, CHINA. (Poster)

6. **Chen Chen,** Peiying Li, Koishihara Nao, Kazunori Motai, Tomoko Ohnishi, Kohsuke Kanekura, Yuhei Hayamizu. Surface enhanced Raman spectroscopy for understanding the interaction between RNA and Arginine-rich peptides. CHEMINAS39. 2019, Kanazawa, JAPAN.

7. **Chen Chen,** Koishihara Nao, Kazunori Motai, Peiying Li, Tomoko Ohnishi,

Kohsuke Kanekura, Yuhei Hayamizu. Surface enhanced Raman spectroscopy for understanding the interaction between RNA and Arginine-rich peptides. 1st A3 Foresight Meeting about Functionalization and Flexible Device Application of Atomic Scale Organic and Inorganic Material. 2019, Atami, JAPAN. (Best Young Researcher Poster Award)

8. **Chen Chen,** Koishihara Nao, Kazunori Motai, Peiying Li, Tomoko Ohnishi, Kohsuke Kanekura, Yuhei Hayamizu. Surface enhanced Raman spectroscopy for understanding the interaction between RNA and Arginine-rich peptides. 2nd International Workshop by the 174th Committee JSPS on Symbiosis of Biology and Nanodevices. 2019, Kyoto, JAPAN. (Student Poster Award)

9. **Chen Chen,** Peiying Li, Kazunori Motai, Hironaga Noguchi, Yuhei Hayamizu. Self-assembly of amphiphilic tetrapeptides on 2D nanomaterials surface. The 80th. JSAP Autumn Meeting. 2019, Sapporo, JAPAN. (Poster)

ABSTRACT

This doctoral thesis entitled “Molecular understanding on LLPS droplets consisting of poly(PR) dipeptide repeats and RNA and its diffusion at solid/liquid interface” describes the patterning effect of positively-charged Arg and rigid Pro residues on the interaction between the peptide and protein/nucleic acid to form liquid-liquid phase separation (LLPS) and the dynamic diffusion process of LLPS droplets at two respective solid substrates: untreated solid cover glass and chemically modified solid cover glass.

LLPS of proteins and RNA molecules emerges as the ubiquitous phenomenon related to the condensed organization of living cells and cellular functions. Recently, Arg-rich dipeptides repeat proteins (DRPs), especially poly(PR), encoded by a hexanucleotide expansion in *C9ORF72* gene are responsible for the neurodegeneration in amyotrophic lateral sclerosis (ALS). Most importantly, these poly(PR) dipeptide repeats are known to undergo LLPS and affect multiple biological processes. Moreover, these LLPS droplets exhibit liquid-like features, including round spherical shape, fusion and fission, wetting surface, dynamic fluidic properties and so forth. However, there still exists some typical challenges for these droplets research. One is the intermolecular interaction for forming LLPS based on the structural distributions of positively-charged Arg and rigid Pro residues in polypeptide. The other one is the interaction between LLPS interface with outer medium, which could affect LLPS uptake process for surrounding biomolecules.

Here, we conducted a series of computations and experiments and found that the primary peptide sequence of dipeptides caused unique charge distributions and molecular rigidity which correlated to their binding energy to RNA. Inserted Pro residues functioned as spacers and provided multivalency to form LLPS. Meanwhile,

the alternate Arg structure of DRPs could efficiently trap proteins with acidic stretches, including NPM1, by electrostatic interactions. In addition, we utilized the wetting properties of LLPS droplets at interface to observe LLPS droplets on two different solid substrates through single particle tracking method. The results revealed that there were mainly fix and diffusion modes for LLPS droplets diffusing at solid/liquid interface. Especially the chemically modified cover glass surface with amine groups could trap more LLPS droplets surface and less of them could diffuse at interface while most LLPS droplets exhibit dynamic diffusion behavior on untreated solid cover glass surface.

In summary, our results will provide a fundamental explanation for understanding LLPS formation and dynamics, which would be helpful for future research on ALS disease. Meanwhile, we also provide an improved method to control LLPS diffusion at solid/liquid interface, so that the fundamental understanding for LLPS droplets nature could be advanced. Most importantly, the LLPS uptake process is desired to be controlled according to the its interface interaction with outer medium, which may pave a way for ALS disease curation in the future.

ABBREVIATIONS

MBO	membrane-bound organelle
ER	endoplasmic reticulum
MLO	membrane-less organelle
LLPS	liquid-liquid phase separation
FRAP	fluorescence recovery after photobleaching
ALS	amyotrophic lateral sclerosis
FTD	frontotemporal dementia
PD	Parkinson's disease
HD	Huntington's disease
AD	Alzheimer disease
DPR	dipeptide repeat protein
NPM1	nucleolar nucleophosmin
IVT	*in vitro* translation
MD	molecular dynamics
LJ	Lennard-Jones
CHARMM	Chemistry at Harvard Macromolecular Mechanics
AMBER	Assisted Model Building with Energy Refinement
OPLS-AA	Optimized Potentials for Liquid Simulations All-Atom
VMD	Visual Molecular Dynamics
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
MM-PBSA	molecular mechanics Poisson Boltzmann surface area
SASA	solvent accessible surface area
Rg	radius of gyration
IDP	intrinsically disordered protein
TAMRA	tetramethyl rhodamine
ROI	region of interest
N.A.	numerical aperture
CSC	critical salt concentration
OD	oligomerization domain
NBD	nucleic-acid-binding domain
IDR	intrinsically disordered region
IPTG	Isopropyl-β-D-thiogalactopyranoside
GFP	green fluorescent protein
LC-MS	liquid chromatography-mass spectrometry
IP	immunoprecipitation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
EDTA	Ethylenediaminetetraacetic acid

TCEP	tris(2-carboxyethyl)phosphine
SDS-PAGE	Sodium dodecyl sulfate-Polyacrylamide gel electrophoresis
CBB-staining	Coomassie Brilliant Blue-staining
SPT	Single particle tracking
FITC	fluorescein isothiocyanate
APTMS	3-aminomethoxysilane
PDMS	polydimethylsiloxane
MSD	mean square displacement

CHAPTER 1. General Introduction

1.1 Cell compartmentalization

In the biological world, cells are the smallest units of life that have to coordinate a myriad of complex biological reactions spatiotemporally by establishing distinct compartments or organelles, for the purpose of regulating cellular metabolism and responding to changes in the living environment.[1,2] The canonical compartment surrounded by a closed lipid boundary has always been synonymous with intracellular organization and is defined as membrane-bound organelles (MBOs).[1] Typically, the mitochondria, which contains the chemical environment necessary for the production of ATP;[3] the lysosomes, which contain the components necessary for the destruction of other proteins;[4] the nucleus, which contains a majority of cell's genetic material;[5] the endoplasmic reticulum (ER), which serves as the largest MBO in eukaryotic cells and performs a variety of important cellular functions, including protein synthesis and processing, lipid synthesis and calcium ion storage and release;[6] Golgi apparatus, which processes membrane and secretory proteins at ER and distributes them to appropriate destinations.[7]

The intriguing class of organelles discovered so far is out of paradigm, which does not have discrete enclosing membranes yet remains coherent structures compartmentalizing and concentrating specific collective molecules.[8] Therefore, these organelles are defined as membrane-less organelles (MLOs), such as nucleolus, Cajal bodies, P-bodies, stress granules and germ granules, etc.[8,9-13] Most importantly, they have spatially patterned and multilayered characteristics, thus exhibiting diverse non-uniform structural and functional material properties not only in the cellular processes shown in Fig. 1, such as stress response, regulation of gene expression, and control of signal transduction, but also in biomedical and biotechnological applications.[14-18]

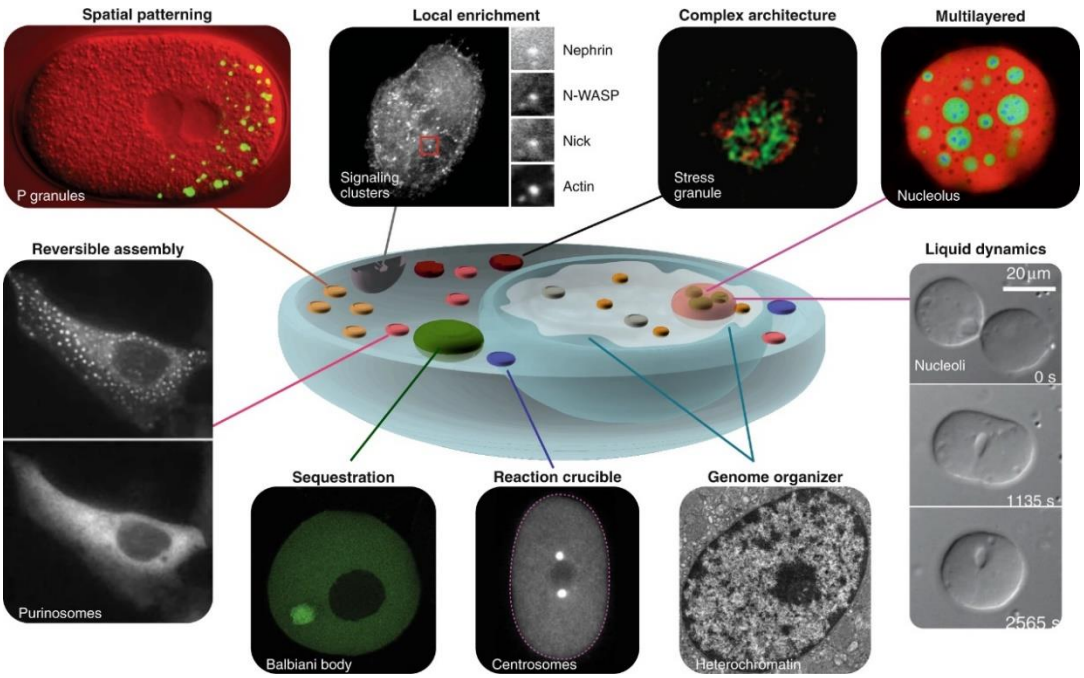


Figure 1. Cases of MLOs within living cells and their cellular functions.[14]

1.2 Liquid-liquid phase separation in biology

The process of assembling dozens of MLOs is defined as liquid-liquid phase separation (LLPS), which increasingly denotes biomolecular condensates as shown in Fig. 2, an example of a highly robust and ubiquitous phenomenon responsible for the insoluble mixtures of oil and water in our daily lives. [8,19] The prerequisite for flexible polymers making up LLPS whether in the cytoplasm or nucleoplasm is that their concentrations must exceed saturation concentration. Finally, the polymers inside and outside LLPS droplets come into a dynamic equilibrium state.

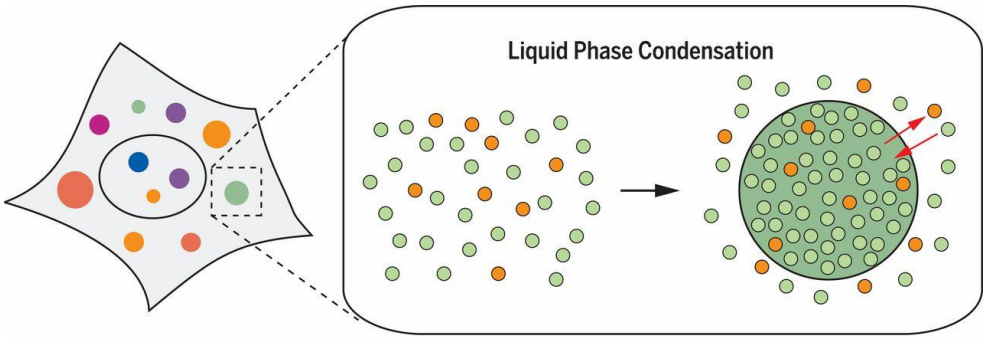


Figure 2. Liquid phase separation condensation within living cells.[8]

As for the characteristics of MLOs from LLPS, we take the germline P granules of *Caenorhabditis elegans* in Fig. 3A and the nucleoli of *Xenopus laevis* germinal vesicles in Fig. 3B as examples. Both of them exhibit remarkable liquid-like features that resemble conventional liquids rather than hard kernels because their round shape and, coalescence to form a larger spherical droplet once contact each other.[12,20] MLOs also possess wetting property resembling water droplets beading up when in contact with the hydrophobic surface so that these droplets drip off in response to shear stress shown in Fig. 3A.[12] In addition, MLOs have rapid reversible internal rearrangements as shown in Fig. 3C, which are assessed by fluorescently labeling the constitutive biomolecules during fluorescence recovery after photobleaching (FRAP) measurments.[12] Therefore, based on the observations above, the dynamic fluid properties of MLOs from LLPS have been increasingly demonstrated.

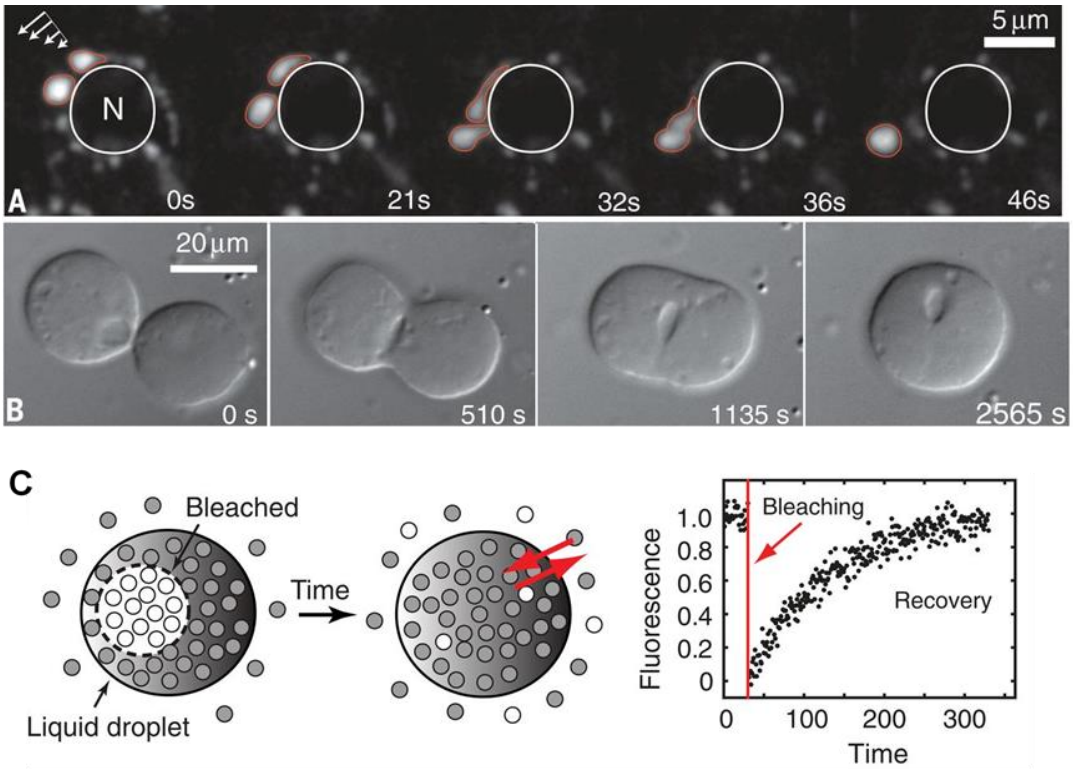


Figure 3. Dynamic fluid properties of MLOs from LLPS. (A) P granules flow, coalesce, and drip off the surface of the nucleus responding to shear forces.[12] (B) Spherical nucleoli from *Xenopus laevis* germinal vesicles coalesce to form a larger sphere.[20] (C) Molecules within the liquid phase show dynamic reorganization process detected by FRAP experiments.[8]

1.3 Principles underlying LLPS formation

With the knowledge about LLPS functioning as organelles, these phase-separated states selectively enrich certain components while excluding others, thereby accelerating or inhibiting cellular biochemical reactions. According to previous studies both *in vitro* or *in vivo*, the most important thing for LLPS is to establish a dense network of interacting macromolecules, with key determinants coming from the valence of interacting constituents and the number of interaction motifs or domains.[19,21]

A meaningful concept when describing how constituents enriched inside condensates lies in scaffolds and clients.[22] Scaffold proteins usually appear in cells with high concentrations and possess many valences, which can make condensate formation; client proteins usually appear with lower concentrations and have fewer valences, which partition into condensates formed by scaffold proteins.[21] Furthermore, the physical properties of LLPS can be particularly achieved through reversible weak or strong interactions during a brief time.[21] Under the energetically favorable conditions, the thermodynamic driving force underlying LLPS of macromolecules solutions (mainly composed of proteins/nucleic acids) includes the replacement of macromolecule/water interactions with macromolecule/macromolecule and water/water interactions.[23] The whole dynamic process of liquid macromolecular rearrangement is filled with diffusion and mixing to equalize the system concentration. As for the principles of LLPS, it has been reported that the internal energy change coming from homophilic interactions negatively contributes during mixing to the free energy change in overall system.[2,24] For mixing events without liquid volume change, a variation of Helmholtz free energy ΔF_{mix} in general can be described as below:

$$\Delta F_{mix} = \Delta U_{mix} - T\Delta S_{mix} \quad (1)$$

where ΔU_{mix}, T and ΔS_{mix} denote a change in internal energy, temperature and entropy, respectively.

Flory-Huggins theory has been considered to provide a lattice-based, mean field approximation model for polymer solutions.[25,26] For a two-component system composed of solute A and solvent B, ΔU_{mix} and ΔS_{mix} have the below expressions:

$$\Delta U_{mix} = \chi(T)\,\phi_a\phi_b N k_B T \quad (2)$$

$$= \chi(T)\,\phi_a(1-\phi_a)nRT \quad (3)$$

where ϕ_a and ϕ_b refer to the volume fraction of molecular species A and B. $\chi(T)$ denotes Flory's χ parameter; N denotes the total molecular number; n denotes the amount of substance; k_B denote the Boltzmann constant; and R denote the gas constant. For the simplified model, χ reflects the interaction strength between different molecular species and is defined:

$$\chi(T) = \frac{\mathrm{z}}{2k_BT}(2\varepsilon_{ab} - \varepsilon_{aa} - \varepsilon_{bb}) \quad (4)$$

where z denotes a constant relying on the model, ε_{ij} denotes the internal energy change through the interaction between molecules i and j. The effective χ parameter value is related to some factors, including temperature, salt concentration and pH, etc., which might influence effective interaction energies. When the phase separation is sensitive to these environmental changes, ΔS_{mix} can be described as:

$$\Delta S_{mix} = -k_B N\left(\frac{\phi_a}{m} ln\phi_a + \phi_b ln\phi_b\right) \quad (5)$$

$$= -nR\left(\frac{\phi_a}{m} ln\phi_a + (1-\phi_a)\ln(1-\phi_a)\right) \quad (6)$$

where m refers to the number of lattice units occupied by a molecule of solute A. Combining the equations above, ΔF_{mix} can be defined:

$$\Delta F_{mix} = nRT\{\chi(T)\,\phi_a(1-\phi_a) + \frac{\phi_a}{m} ln\phi_a + (1-\phi_a)ln(1-\phi_a)\} \quad (7)$$

For χ parameter with a small value, the intermolecular interactions favor the mixed state, there exists a concave curve for the dependence of free-energy on the volume fraction. As shown in Fig. 4 (a), suppose A solution follows a concave free energy cave and the volume fraction of A is ϕ_0 and the free energy is F_0 in the initial state.[27] After forming mixture phase separation states into two phases, we define the volume fraction of A in the bulk phase as ϕ_1 with free energy $F(\phi_1)$ and the same fraction in condensate is ϕ_2 with free energy $F(\phi_2)$. The free energy after phase separation F_{sep} is on the line segment between $F(\phi_1)$ and $F(\phi_2)$. Finally, once the free energy curve is concave, due to F_{sep} is higher than that in the initial state, so the whole system will be destabilized and return to the initial single-phase solution.

For χ parameter with a large value, there is a strong homophilic molecular interactions, which means the system obtains energy in close proximity for the same type of molecules and loses energy in close proximity for different types of molecules. Therefore, the internal energy variation is not favorable for the formation of mixed state, leading to the convex region of free energy curve shown in Fig. 4 (b).[27] For the phase separation occasion, the total free energy in the system F_{sep} is smaller than the initial one F_0. The free energy of system will be the smallest for the volume fractions of when the dilute and dense phase reach the contact point at the same tangent line, namely binodal point. As a result, the minimal requirement for phase separation is that the initial volume fraction is within the binodal points and the phase separation can be diverse on the basis of the volume fraction.

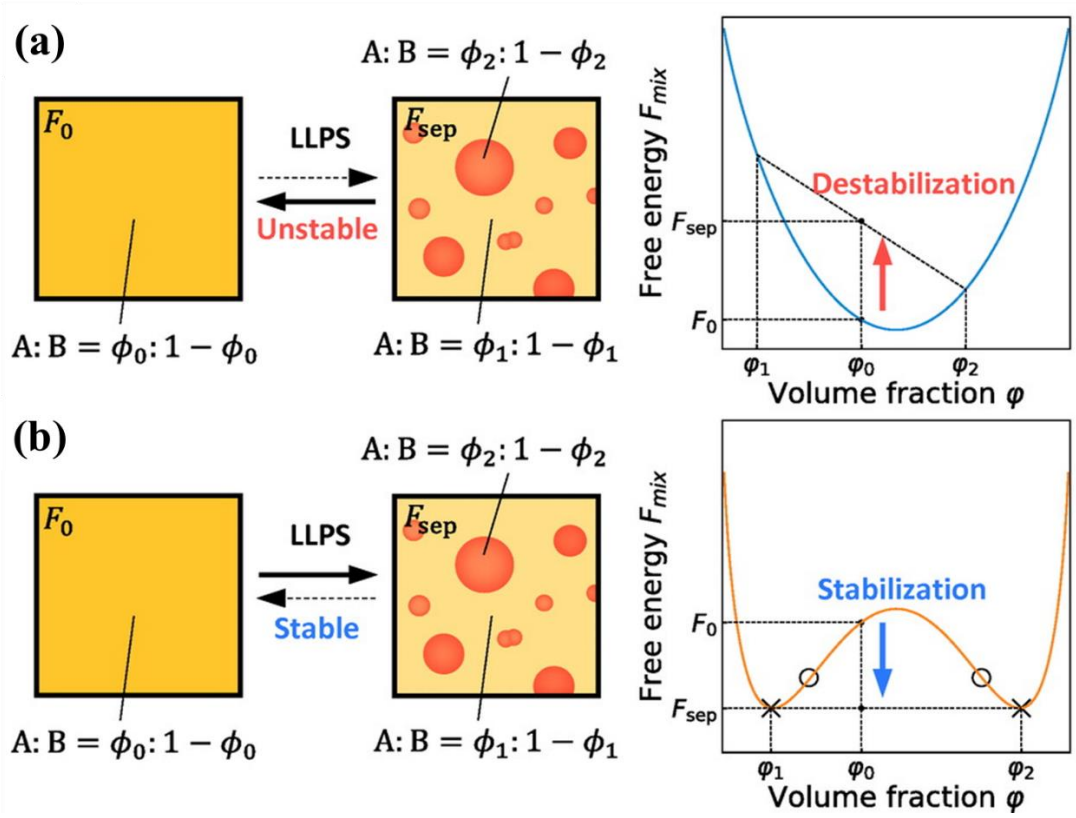


Figure 4. (a) Behavior of the mixture with a concave free energy curve. LLPS destabilizes the system, leading to a single phase state;[27] (b) Behavior of the mixture with a convex region on the free energy curve. Markers with a crosses denote the binodals and circles denote the spinodals.[27]

For χ is sufficiently large with a higher internal energy shown in Fig. 5, the volume fraction value affects phase separation states when it is even inside the binodal points. There are two occasions: one is the initial volume fraction is ϕ_{ms}, there are two phases with different volume fractions as $\phi_{ms}{}'$ and $\phi_{ms}{}''$. Then, the free energy $F_{ms}{}'$ at the phase separation state is larger than that of F_{ms}, the metastable two phases will soon return to the initial mixing state; the other is the initial volume fraction φ_{us}, there are two phases with different volume fractions as $\varphi_{us}{}'$ and $\varphi_{us}{}''$. Then, the free energy $F_{us}{}'$ at the phase separation state is smaller than that of F_{us}, the two phases will continue to further separate. Besides, the solution is always at the mixing state for the convex curve region. Finally, there are two spinodal points for the whole system in total.

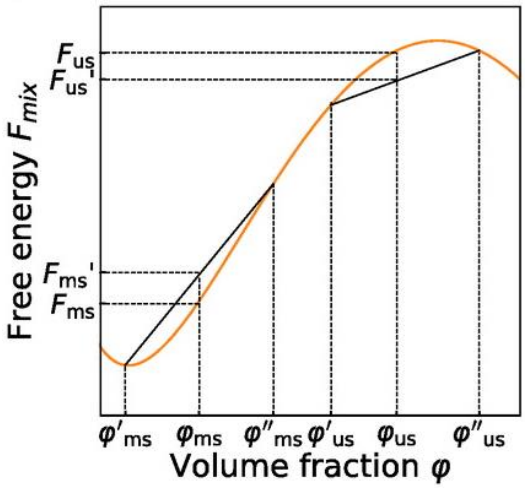


Figure 5. Part of free energy curve to show the relation between volume fraction and free energy of the system for large χ case.[27]

Based on the above facts, the phase diagram of the two-component system is plotted in Fig. 6 to depict the phase behavior in solution when connecting binodal points as binodal curve and spinodal points as spinodal curve at each temperature, salt concentration, etc. as the function as the volume fraction. When the initial conditions are outside the binodal curve and above the critical point, only one-phase mixing state exists without any phase separation; when it lies between the binodal and spinodal curve,

phase separation can occur through nucleation and growth processes; when it lies within the spinodal curve, continuous and spontaneous phase separation occurs as spinodal decomposition.

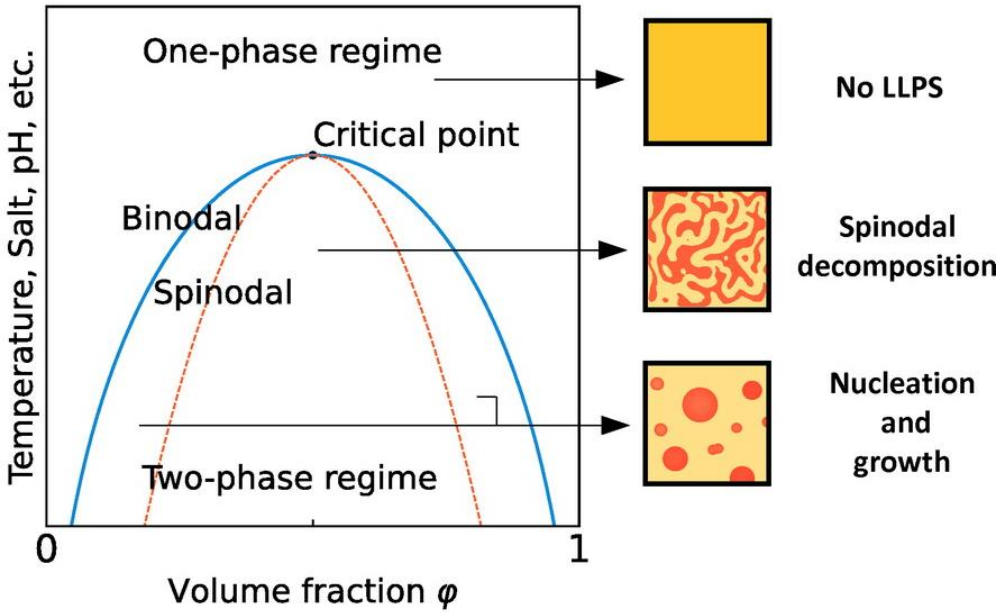


Figure 6. Phase diagrams and phase behavior of mixtures in various occasions.[27]

1.4 LLPS connects with the pathology of neurodegenerative disease

Neurodegeneration is the progressive degeneration and/or death of neurons that leads to the occurrence of diverse neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS),[28] frontotemporal dementia (FTD),[29] Parkinson's disease (PD),[30] Huntington's disease (HD)[31] and Alzheimer disease (AD),[32] etc. This leads to permanent problems related to dementia, language abnormalities, personality changes and progressive muscle weakness and currently there are no known treatments on these neurodegenerative diseases.[33-35] In contrast to the normal cellular function of LLPS, abnormal phase separation produces local aggregation of cytosolic or nuclear proteins, which has been demonstrated as a hallmark of neurodegenerative disease pathogenesis.[36,37] A spate of studies have reported some typical neurodegenerative disease-associated proteins, including α-synuclein, FUS, TDP-43, and Tau, within

biomolecular condensates both *in vivo* and *in vitro* as shown in Fig. 7.[37]

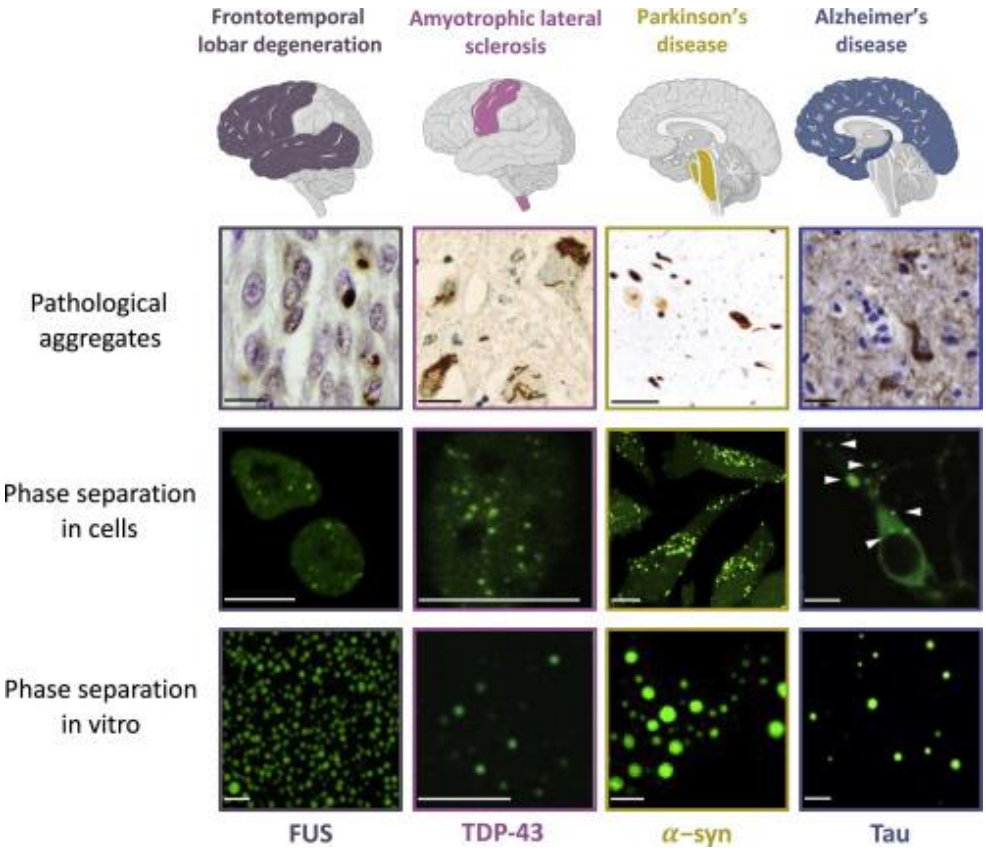


Figure 7. Proteins related to neurodegenerative diseases indicate LLPS behavior. First row: Schematic diagram of brain regions having pathological aggregationss of FTD-associated FUS in purple, ALS-associated TDP-43 in pink, PD-associated α-synuclein in yellow, and AD associated Tau in blue. Second row: typical immunohistochemistry illustrations of pathological aggregations appeared in postmortem brains: FUS and TDP-43,[34] α-synuclein,[38] and Tau.[39] Scale bars represent 20 μm. Third and fourth rows: phase separation of proteins *in vivo* and *in vitro*. Illustrations taken from: FUS,[40,41] TDP-43,[42] α-synuclein,[43] and Tau.[44,45] Scale bars represent 10 μm.

During the insoluble protein aggregations formation, there exists a phase transition from liquid to solid phase, which is an irreversible process over time so that biomolecular condensates eventually lost their previous reversible dynamic properties as shown in Fig. 8.[36,46,47] This phase transition phenomenon helps to explain the origin

of protein aggregates that may be related to the aberrant LLPS behavior.

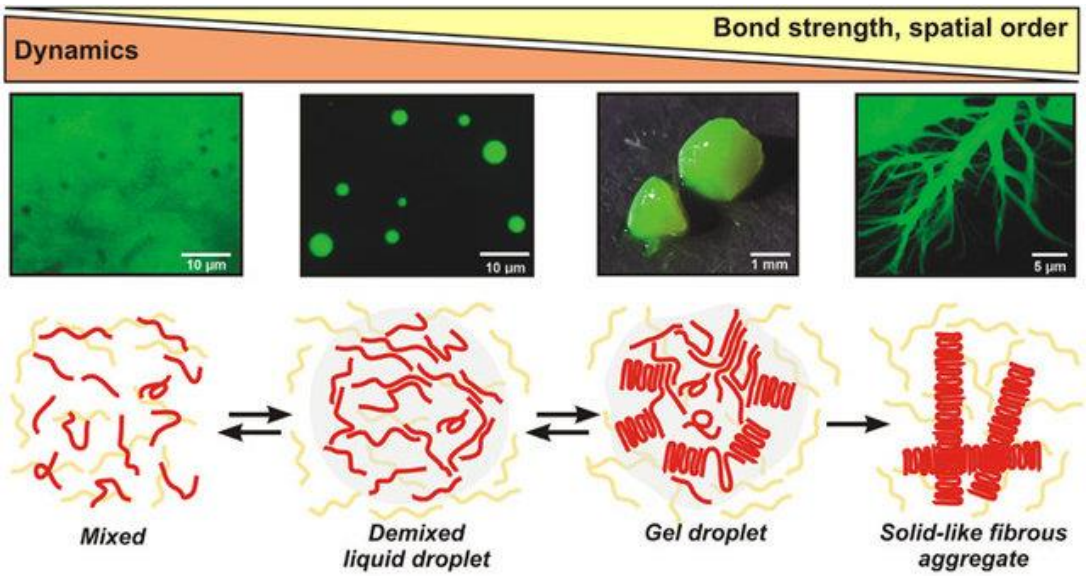


Figure 8. Continuous model of intrinsically disordered proteins undergoing phase separation.[48] Top: FUS proteins form liquid droplets, gels and fibrous aggregates.[49] Bottom: An intrinsically disordered protein in red has a phase transition to form liquid droplet in the presence of LLPS. Then, the droplet evolutes into gel droplet structure and final irreversible aggregates over time.[48]

1.4.1 Arg-rich dipeptide repeat proteins makes cell death

To date, the amplification of a hexanucleotide repeat (GGGGCC) in the gene *C9orf72* has been identified as the typical pathogenic mutation in the sporadic and familial cases of FTD and ALS diseases.[50,51] Non-conventional repeat-associated non-ATG-initiated translation (RAN translation) causes aberrant expression of dipeptide repeat proteins (DPRs).[52] Among the five hexanucleotide amplifications from possible DPRs [poly(GA), poly(PA), poly(GP), poly(GR) and poly(PR)],[52,53] two Arg-rich DPRs [poly(PR) and poly(GR)] tend to be toxic for various potential targets, both in cellar and animal models.[54-56] For example, (i) protein translation, (ii) dynamics of MLOs, (iii) RNA metabolism, (iv) nucleocytoplasmic transport, and (v) DNA damage responses to both poly(PR) and poly(GR).[55, 57-63] (vi) ubiquitin-proteasome systems, and (vii) heterochromatin anomalies, for poly(PR) alone;[64,65] and (viii) mitochondrial

function for poly(GR) alone.[66] Thus, the processes impaired by Arg-rich DPRs are highly diverse and involved in ALS disease-related regulation through independent mechanisms.

1.4.2 Arg-rich dipeptide repeat proteins phase separation mechanism

As a highly polar amino acid with positive charge and a dipole moment, Arginine plays a key role in biomolecular interactions (protein-protein and protein-nucleic acid).[67] The high repeatability and abundant positive charge allow toxic DPRs to interact with nucleic acids and proteins at multiple sites through long-range electrostatics and directional cation-π interactions, forming LLPS.[60, 68]

The key biophysical features of protein with the ability to undergo LLPS include multivalent, relatively low-affinity interactions through intrinsically disordered regions (IDRs) and/or frequent occurrence of charged amino acids.[69] Indeed, Arg-rich *C9orf72* DPRs phase-separated with proteins or nucleic acids to form droplets *in vitro*, and preferentially accumulate in MLOs, such as nucleoli and stress granules cells.[59,60] The DPRs regulate dynamics of phase-separated proteins and thereby impede functions of MLOs as well as proteins composing MLOs. In particular, poly(PR) and poly(GR) inhibit stress granules dynamics and nucleolar nucleophosmin (NPM1) diffusion.[70]

Although many peptides undergo LLPS, not all of them show cytotoxicity. One of the primary sequences is important for cytotoxicity.[71] For example, although the highly cationic Arg_{12} (R_{12}) peptide phase-separates with RNA under *in vitro* conditions, it is located in the cytosol and has no toxic effect.[72,73] In contrast, Arg-rich peptide with alternating Pro residues such as $(PR)_{20}$, is located mainly in the nucleolus and exerts toxicity by inhibiting protein translation shown in Fig. 9.[74] Therefore, the scientific challenge is the rules governing the structural amino acid sequential differences among Arg-rich peptides interacting with nucleic acids/proteins for LLPS formation are still unknown.

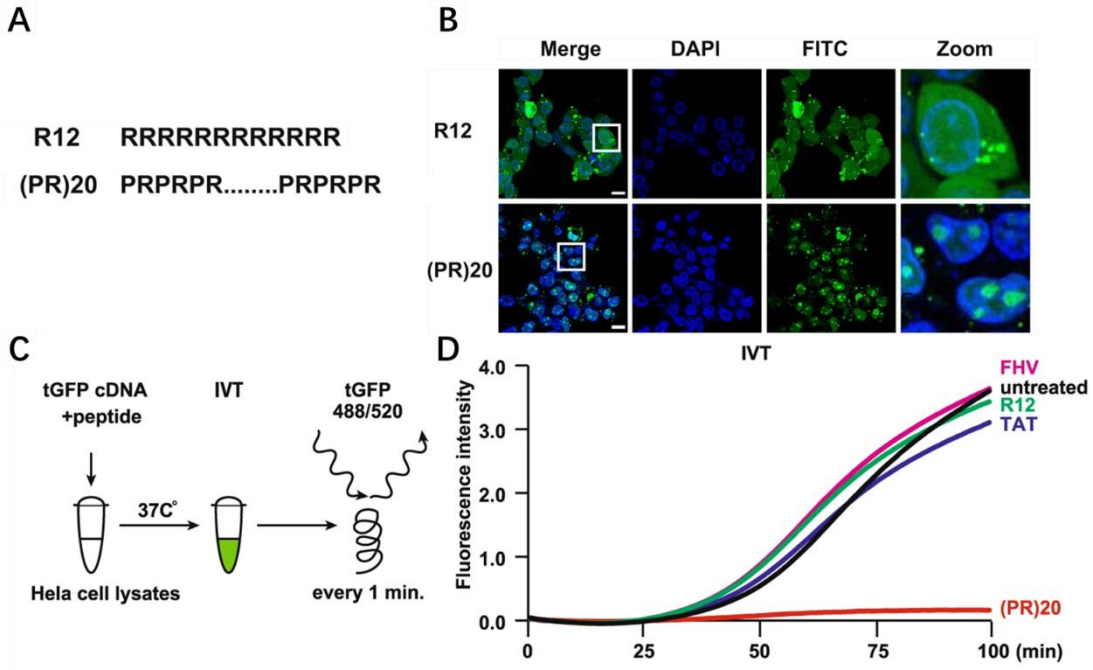


Figure 9. Poly-PR localizes to the nucleoli and causes cell death by inhibiting protein translation.[74] A. Amino acid sequences of R_{12} and $(PR)_{20}$. B. Confocal imaging of HEK293 cells treated with 10 μM of FITC-labeled R_{12} and $(PR)_{20}$, respectively. Nuclei were visualized by Nucblue live dyes. C. Real time monitoring of protein translation via an *in vitro* translation (IVT) assay. tGFP refers to turboGFP. D. Real time monitoring of the tGFP signal in the IVT assay. The signal of tGFP was detected every one minute.

1.5 LLPS diffusion on the solid/liquid interface

Commonly, LLPS droplets have round morphologies due to minimal surface tension, coalesce to form a single spherical droplet after touching each other and wetting interface.[12,20] Until now, there are two main concepts to describe the droplet growth process. One is Ostwald ripening showing small molecules diffuse in search of larger droplets in Fig. 10 (a).[75] This process is spontaneous to decrease the surface free energy because of the poor energy stability within molecules located near the droplet surface.[75] The other is the passive coalescence for droplets Brownian movement in Fig. 10 (b).[76] The process happens if two similar droplets in size come into contact and then fuse to form a larger droplet. The LLPS droplet interface is the decisive factor in these

processes. In addition, the constituent molecules within LLPS droplets dynamically interact with their surroundings via the interface on a time scale of a few seconds.[77,78,79]

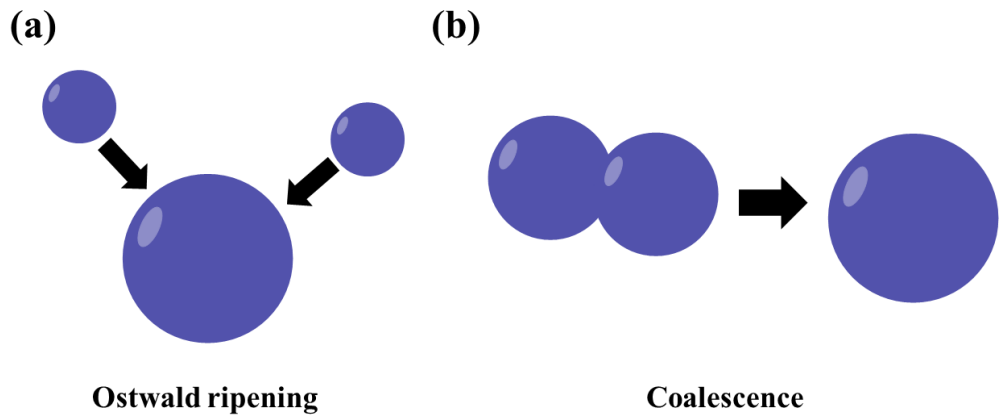


Figure 10. Droplet growth process. (a) Ostwald ripening; (b) Coalescence.

Importantly, the role of zeta potential of individual LLPS droplets has been utilized as a fundamental quantity to reflect the relation between droplet surface electrostatic forces and its stability, indicating the tendency to diffuse and grow.[80] These conclusions suggest that droplet interfacial interactions influence the dynamic properties of LLPS and its diffusion pattern in bulk solutions.

Compared to the current progress about the physical properties of LLPS droplets in bulk solutions, the knowledge of their interaction with solid surfaces has lagged relatively behind. Moreover, the optical microscopy remains the usual method for characterizing LLPS droplets. Utilizing the transparent substrates in optical measurements, the interactions of LLPS droplets with glass surface result in wetting or de-wetting of the interface and variations for droplet material properties. The LLPS wetting property is also crucial for fluorescence microscopy and immobilized droplets, providing a platform for quantitative studies of the droplet shape and its fluorescence intensity, such as FRAP. Currently, it has been proposed how to immobilize droplets on glass surfaces.[81] The wetting assay has proposed that spherical droplets in liquids wet the surface and form irregular shapes over time if the protein phases are separated while hydrogels and protein aggregations cannot wet or vary shape over time. In addition, a variety of coatings have been suggested, such as polyethylene glycol (PEG) or lipids

on the glass surface to maintain the material properties of the droplets and imaging capability for long enough.[82] Although some recommended protocols have been put forward about droplet wetting, the interfacial interaction between droplets and the glass surface remains poorly understood.

1.6 Focus of this thesis

As a ubiquitous phenomenon occurred in cells, LLPS plays a key role in a variety of cellular functions, i.e., making up MLOs in the cytoplasm or nucleoplasm and contributing to different biochemical reactions. Due to this phase separation possesses droplet-like structures, specific proteins and nucleic acid molecules can reversibly be organized to form biomolecular condensates, dynamically inside and outside the droplets. However, LLPS sometimes makes biomolecular aggregation, which is characteristic of the origin of cytotoxicity in several neurodegenerative diseases and support their pathological activity. Arg-rich poly(PR) dipeptides repeats has been reported to interact with some negatively charged proteins or nucleic acids and then form LLPS droplets in cells, impeding the functions of MLOs including nucleoli. However, there are still some main mysteries to be explored:

(1) Intermolecular interactions during LLPS formatio.

Although Arg-rich poly(PR) dipeptide repeats has the structural characteristics with high contents of positively charged amino acids and indeed forms LLPS interacting with other negatively charged biomolecules, the molecular interaction mechanism for this phase separation formation is unclear, especially the effect of primary sequences for Arg-rich poly(PR) dipeptide repeats on LLPS formation process. This prompted us to do structural decoding for the principles defining the LLPS by Arg-rich poly(PR) dipeptide repeats. Meanwhile, we also took the distribution of Pro residue into function, especially the effect of Pro rigidity on intermolecular interactions.

(2) The interface of the LLPS droplet with outside medium.

It is also unclear until now that the LLPS interface, which is involved their interactions with other components, including other droplets, surface and the uptake of proteins and other biomolecules into the droplets. Due to LLPS droplets have liquid and wetting properties on glass surface, we then study LLPS dynamics at solid/liquid surface and utilize engineering method to chemically modify glass surface to adjust the interaction strength between LLPS droplets and solid surface.

Collectively, on the one hand, the study on molecular interaction mechanism provides the information about how the structural distribution of both Arg and Pro residues of Arg-rich poly(PR) affects LLPS formation. On the other hand, there exists LLPS droplets growth and diffusion events after forming the initial droplet. According to the study on LLPS diffusion mode at solid/liquid interface through engineering method, a novel insight could be provided to advance fundamental studies related to LLPS formation understanding, especially the LLPS uptake process for surrounding biomolecules

1.7 References

1. Alberti, S. Phase separation in biology. *Curr. Biol.* **27**, R1097–R1102 (2017).
2. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
3. Friedman, J. R. & Nunnari, J. Mitochondrial form and function. *Nature* **505**, 335–343 (2014).
4. Luzio, J. P., Pryor, P. R. & Bright, N. A. Lysosomes: Fusion and function. *Nat. Rev. Mol. Cell Biol.* **8**, 622–632 (2007).
5. Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL, Darnell J (2004). *Mol. Cell Biol.* (5th ed.). New York: WH Freeman.
6. Phillips, M. J. & Voeltz, G. K. Structure and function of ER membrane contact sites with other organelles. *Nat. Rev. Mol. Cell Biol.* **17**, 69–82 (2016).
7. Fernández, C. J. & Warren, G. In vitro synthesis of sulfated glycosaminoglycans coupled to inter- compartmental Golgi transport. *J. Biol. Chem.* **273**, 19030–19039 (1998).
8. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, (2017).
9. Pederson, T. The nucleolus. Cold Spring Harb. Perspect. Biol. 3, a000638 (2011).
10. Gall, J. G. The centennial of the Cajal body. *Nat. Rev. Mol. Cell Biol.* **4**, 975–980 (2003).
11. Molliex, A. *et al.* Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **163**, 123–133 (2015).
12. Brangwynne, Clifford P., et al. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729-1732 (2009).
13. Nott, T. J. *et al.* Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* **57**, 936–947 (2015).
14. Bracha, D., Walls, M. T. & Brangwynne, C. P. Probing and engineering liquid-phase organelles. *Nat. Biotechnol.* **37**, 1435–1445 (2019).

15. Boulay, G. *et al.* Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. *Cell* **171**, 163-178.e19 (2017).
16. Li, P. *et al.* Phase transitions in the assembly of multivalent signaling proteins. *Nature* **483**, 336–340 (2012).
17. Wheeler, J. R., Matheny, T., Jain, S., Abrisch, R. & Parker, R. Distinct stages in stress granule assembly and disassembly. *eLife* **5**, 1–25 (2016).
18. Su, Xiaolei, et al. Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **352**, 595-599 (2016).
19. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
20. Brangwynne, C. P., Mitchison, T. J. & Hyman, A. A. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proc. Natl. Acad. Sci. U. S. A*. **108**, 4334–4339 (2011).
21. Alberti, S. & Dormann, D. Liquid-Liquid Phase Separation in Disease. *Annu. Rev. Genet.* **53**, 171–194 (2019).
22. Banani, S. F. *et al.* Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651–663 (2016).
23. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).
24. Gomes, E. & Shorter, J. The molecular language of membraneless organelles. *J. Biol. Chem*. **294**, 7115–7127 (2019).
25. Huggins, Maurice L. Some properties of solutions of long-chain compounds. *J. Phys. Chem.* **46**, 151-158. (1942)
26. Flory, P. Thermodynamics of high polymer solutions. *J. Chem. Phys.* **22**, 415–426 (1942).
27. Kamimura, Y. R. & Kanai, M. Chemical Insights into Liquid-Liquid Phase Separation in Molecular Biology. *Bull. Chem. Soc. Jpn*. 1045–1058 (2020).
28. Julien, J. P. ALS: Astrocytes move in as deadly neighbors. *Nat. Neurosci.* **10**, 535–

537 (2007).

29. Birx, H. *et al. Dsm V Tr*. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* **52**, (2011).
30. Chen, S. Y. & Tsai, S. T. The epidemiology of Parkinson's disease. *Tzu Chi Medical Journal* **22**, 73–81 (2010).
31. Lobsiger, C. S. & Cleveland, D. W. Glial cells as intrinsic components of non-cell-autonomous neurodegenerative disease. *Nat. Neurosci.* **10**, 1355–1360 (2007).
32. Serrano-Pozo, A., Frosch, M. P., Masliah, E. & Hyman, B. T. Neuropathological alterations in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* **1**, 1–23 (2011).
33. Josephs, K. A. *et al.* Neuropathological background of phenotypical variability in frontotemporal dementia. *Acta Neuropathol.* **122**, 137–153 (2011).
34. Mackenzie, I. R. A., Rademakers, R. & Neumann, M. TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. *Lancet Neurol.* **9**, 995–1007 (2010).
35. Rademakers, R., Neumann, M. & MacKenzie, I. R. Advances in understanding the molecular basis of frontotemporal dementia. *Nat. Rev. Neurol.* **8**, 423–434 (2012).
36. Taylor, J. P., Hardy, J. & Fischbeck, K. H. Toxic proteins in neurodegenerative disease. *Science* **296**, 1991–1995 (2002).
37. Zbinden, A., Pérez-Berlanga, M., De Rossi, P. & Polymenidou, M. Phase Separation and Neurodegenerative Diseases: A Disturbance in the Force. *Dev. Cell* **55**, 45–68 (2020).
38. Mamais, A. *et al.* Divergent α-synuclein solubility and aggregation properties in G2019S LRRK2 Parkinson's disease brains with Lewy Body pathology compared to idiopathic cases. *Neurobiol. Dis.* **58**, 183–190 (2013).
39. Gibbons, G. S. *et al.* Detection of Alzheimer disease (AD)-specific tau pathology in AD and nonAD tauopathies by immunohistochemistry with novel conformation-selective tau antibodies. *J. Neuropathol. Exp. Neurol.* **77**, 216–228 (2018).
40. Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918–921 (2018).

41. Qamar, S. *et al.* FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. *Cell* **173**, 720-734.e15 (2018).
42. Gasset-Rosa, F. *et al.* Cytoplasmic TDP-43 De-mixing Independent of Stress Granules Drives Inhibition of Nuclear Import, Loss of Nuclear TDP-43, and Cell Death. *Neuron* **102**, 339-357.e7 (2019).
43. Ray, S. *et al.* α-Synuclein aggregation nucleates through liquid–liquid phase separation. *Nat. Chem.* **12**, 705–716 (2020).
44. Boyko, S., Qi1, X., Chen, T. H., Surewicz, K. & Surewicz, W. K. Liquid-liquid phase separation of tau protein: The crucial role of electrostatic interactions. *J. Biol. Chem.* **294**, 11054–11059 (2019).
45. Wegmann, S. *et al.* Tau protein liquid–liquid phase separation can initiate tau aggregation. *EMBO J.* **37**, 1–21 (2018).
46. Li, Y. R., King, O. D., Shorter, J. & Gitler, A. D. Stress granules as crucibles of ALS pathogenesis. *J. Cell Biol.* **201**, 361–372 (2013).
47. Polymenidou, M. & Cleveland, D. W. The seeds of neurodegeneration: Prion-like spreading in ALS. *Cell* **147**, 498–508 (2011).
48. Alberti, S. *et al.* A User's Guide for Phase Separation Assays with Purified Proteins. *J. Mol. Biol*. **430**, 4806–4820 (2018).
49. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015).
50. Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268 (2011).
51. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron* **72**, 245–256 (2011).
52. Mori, K. *et al.* The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339**, 1335–1339 (2013).
53. Zu, T. *et al.* Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. U. S. A*. **108**, 260–265 (2011).

54. Mizielinska, S. *et al*. C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins." *Science* **345**, 1192–1195 (2014).
55. Zhang, Y. J. *et al.* Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Nat. Med.* **24**, 1136–1142 (2018).
56. Hao, Z. *et al.* Motor dysfunction and neurodegeneration in a C9orf72 mouse line expressing poly-PR. *Nat. Commun.* **10**, (2019).
57. Kanekura, K. *et al.* Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. *Hum. Mol. Genet*. **25**, 1803–1813 (2016).
58. Moens, T. G. *et al.* C9orf72 arginine-rich dipeptide proteins interact with ribosomal proteins in vivo to induce a toxic translational arrest that is rescued by eIF1A. *Acta Neuropathol.* **137**, 487–500 (2019).
59. Boeynaems, S. *et al.* Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. *Mol. Cell* **65**, 1044-1055.e5 (2017).
60. Lee, K. H. *et al.* C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. *Cell* **167**, 774-788.e17 (2016).
61. Kwon, I. *et al.* Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**, 1139–1145 (2014).
62. Zhang, K. *et al.* The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**, 56–61 (2015).
63. Andrade, N. S. *et al.* Dipeptide repeat proteins inhibit homology-directed DNA double strand break repair in C9ORF72 ALS/FTD. *Mol. Neurodegener.* **15**, 1–18 (2020).
64. Gupta, R. *et al.* The proline/arginine dipeptide from hexanucleotide repeat expanded C9ORF72 inhibits the proteasome. *eNeuro* **4**, (2017).
65. Zhang, Y. J. *et al.* Heterochromatin anomalies and double-stranded RNA accumulation underlie C9orf72 poly(PR) toxicity. *Science* **363**, (2019).
66. Choi, S. Y. *et al.* C9ORF72-ALS/FTD-associated poly(GR) binds Atp5a1 and compromises mitochondrial function in vivo. *Nat. Neurosci.* **22**, 851–862 (2019).

67. Boeynaems, S. *et al.* Spontaneous driving forces give rise to protein−RNA condensates with coexisting phases and complex material properties. *Proc. Natl. Acad. Sci. U. S. A*. **116**, 7889–7898 (2019).
68. Lin, Y. *et al.* Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. *Cell* **167**, 789-802.e12 (2016).
69. Babinchak, W. M. *et al.* The role of liquid-liquid phase separation in aggregation of the TDP-43 low-complexity domain. *J. Biol. Chem.* **294**, 6306–6317 (2019).
70. White, M. R. *et al.* C9orf72 Poly(PR) Dipeptide Repeats Disturb Biomolecular Phase Separation and Disrupt Nucleolar Function. *Mol. Cell* **74**, 713-728.e6 (2019).
71. Ukmar-Godec, T. *et al.* Lysine/RNA-interactions drive and regulate biomolecular condensation. *Nat. Commun.* **10**, (2019).
72. Meloni, B. P. *et al.* The neuroprotective efficacy of cell-penetrating peptides TAT, penetratin, Arg-9, and Pep-1 in glutamic acid, kainic acid, and in vitro ischemia injury models using primary cortical neuronal cultures. *Cell. Mol. Neurobiol.* **34**, 173–181 (2014).
73. Meloni, B. P. *et al.* Poly-arginine and arginine-rich peptides are neuroprotective in stroke models. *J. Cereb. Blood Flow Metab.* **35**, 993–1004 (2015).
74. Kanekura, K. *et al.* Characterization of membrane penetration and cytotoxicity of C9orf72-encoding arginine-rich dipeptides. *Sci. Rep*. **8**, 1–11 (2018).
75. Li, Y., Lubchenko, V., Vorontsova, M. A., Filobelo, L. & Vekilov, P. G. Ostwald-like ripening of the anomalous mesoscopic clusters in protein solutions. *J. Phys. Chem. B* **116**, 10657–10664 (2012).
76. Boeynaems, S. *et al.* Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* **28**, 420–435 (2018).
77. Phair, R. D. & Misteli, T. High mobility of proteins in the mammalian cell nucleus. *Nature* **404**, 604-609 (2000).
78. Snaar, S., Wiesmeijer, K., Jochemsen, A. G., Tanke, H. J. & Dirks, R. W. Mutational analysis of fibrillarin and its mobility in living human cells. *J. Cell Biol.* **151**, 653–662 (2000).

79. Kedersha, N. *et al.* Dynamic shuttling of TIA-1 accompanies the recruitment of mRNA to mammalian stress granules. *J. Cell Biol.* **151**, 1257–1268 (2000).
80. Welsh, T. J. *et al.* Single particle zeta-potential measurements reveal the role of electrostatics in protein condensate stability. *bioRxiv* 1–35 (2020).
81. Wang, Z., Zhang, G. & Zhang, H. Protocol for analyzing protein liquid–liquid phase separation. *Biophys. Reports* **5**, 1–9 (2019).
82. Alberti, S. *et al.* A User's Guide for Phase Separation Assays with Purified Proteins. *J. Mol. Biol*. **430**, 4806–4820 (2018).

CHAPTER 2. Molecular understanding on the interaction between RNA and poly(PR) dipeptide repeats with various periodicities

2.1 Introduction

The most remarkable biochemical properties of Arg-rich DPRs, such as poly(PR) and poly(GR), are repetitive structures and extraordinarily highly charged nature due to the abundant Arg residues. Arg is a highly polar amino acid with a positive charge interacting with molecules with negative charges (acidic proteins and nucleic acids). Biomolecular interactions, including protein-protein and protein-nucleic acid, affect biomaterial properties and functions of Arg-rich DPRs. Instead of the cytotoxicity occurred due to poly(PR) or poly(GR), it has been reported that poly-Arg peptide, often used as a drug deliver carrier accompanied with its cell-permeable properties, is highly charged and interacts with anionic molecules in a similar manner,[1] but does not exert a cytotoxic effect.[2,3] Given the above points, we speculated that the distribution of charged Args in the peptide could affect its interaction with target biomolecules, which may determine its toxicity.

Due to the rapid development of scientific technology, computer simulations have become an effective tool for solving scientific problems and complementing experiments. On the one hand, computer simulations can simulate experiments and illuminate invisible microscopic details, thus explaining the experimental results. On the other hand, simulations also help to predict experimental results. There are two common simulation methods for molecular systems, Monte Carlo[4] and molecular dynamics simulation.[5] Although Monte Carlo simulations are simpler than Molecular Dynamics (MD) simulations because the force calculations can be ignored in the system, they cannot produce better meaningful statistics than MD within a specific calculation time.[6] Thus, MD becomes more general method that analyzes the physical motion of atoms and molecules, allowing them to interact with each other for a fixed amount of

time. MD allows the dynamic evolution of a system to be observed at atomic resolution. Here, we utilized MD simulations to study the molecular interactions between RNA and poly(PR) variants with the same length, while the only difference lies in the distribution of Arg residues.

2.2 Molecular dynamics simulation

2.2.1 General principle

MD simulations method bases on Newton's second law under the assumption that each particle is similar to a Newtonian particle, and the quantum behavior is wholly neglected in the system. To be precise, the electron motion is not taken into account and it is assumed that electrons always keep in the ground state and then adjust the dynamic properties immediately if the atomic position changes. Thus, MD merely uses classical mechanics to describe the motion of an interacting atomic system:

$$F = ma \tag{1}$$

where F denotes the force; m denotes the mass; a denotes the atomic acceleration. If the position and velocity of every atom are given, the system state will be predicted, new position and velocity can be finally calculated. After repeating the same process over a period of time, the trajectory of atomic motion can be tracked consequently.

2.2.2 Empirical force field

In MD simulations, the key step is to describe atomic interactions in the system, i.e., force field, which is built to introduce all necessary molecular interactions and then model the important freedom degrees.[7] The initial step in building a real atomistic model is for the forces evaluation. Then, the force F at time t will be obtained according to the gradient of the potential energy V related to the position coordinates for particle i via the following equation:

$$F_i = -\nabla_i V \tag{2}$$

In the general way, the total potential energy V_{total} is subdivided into non-bonded (inter-molecular) and bonded (intra-molecular) contributions:

$$V_{total} = V_{non-bonded} + V_{bonded} \tag{3}$$

Here, non-bonded interactions involve van der Waals and electrostatic interactions; bonded interactions are made up of bond stretching potentials, angular bending potentials and torsional potentials shown in Fig. 1.

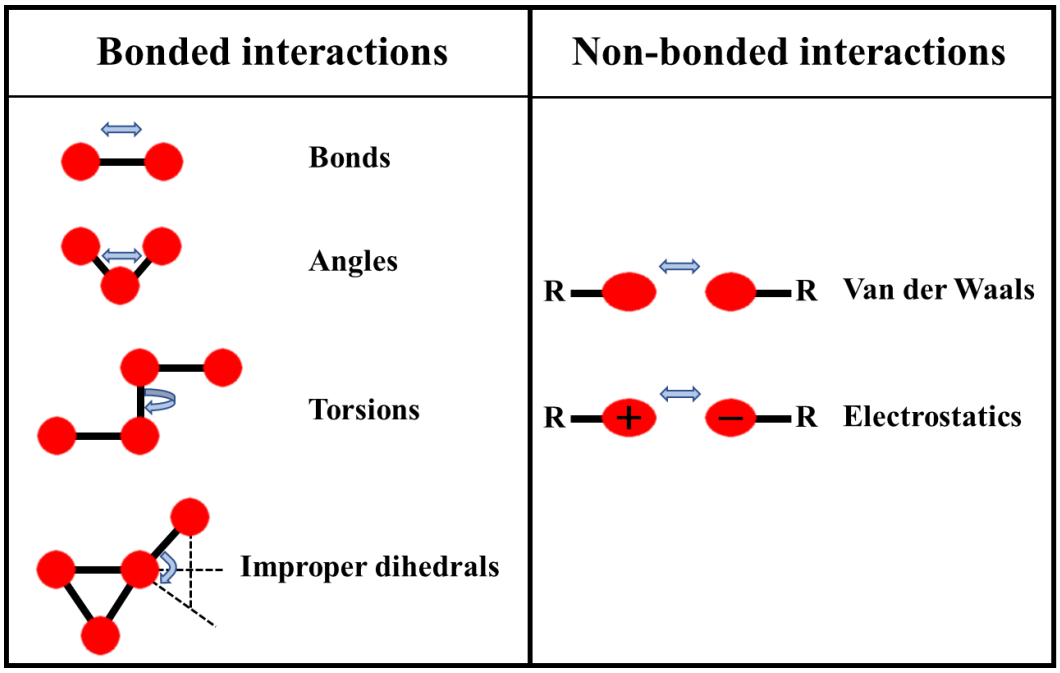


Figure 1. Bonded and non-bonded interactions that constitute force field.

For bonded interactions, there are three main interactions of bonded atoms described above. Bond stretching potentials V_{str} is the energy required to compress or stretch a covalent bond. Thinking of a bond like a spring with the equilibrium length r_0 and the energy required to compress or stretch it can be predicted via Hooke's law:

$$V_{str} = \frac{1}{2} k_{s,ij} \left(r_{ij} - r_{0,ij} \right)^2 \tag{4}$$

V_{angle} is the energy required for a bond bending from the equilibrium angle θ_0. Also,

bending can be modeled like a spring and the energy is approximated via Hooke's law for the angle:

$$V_{angle} = \frac{1}{2} k_{b,ijk} \left(\theta_{ijk} - \theta_{0,ijk}\right)^2 \quad (5)$$

$V_{torsion}$ denote the energy required to rotate with the bond. The torsional interaction is modeled through the electric potential:

$$V_{torsion} = \sum K_{\phi} \left(1 - \cos\left(n\phi\right)\right) \quad (6)$$

$V_{improper}$ denote the energy for the atomic planar group deformation from the equilibrium angle ω_0, which is zero commonly. This system can also be modeled in terms of a spring with an energy given via Hooke's law for the plane angle:

$$V_{improper} = \frac{1}{2} k_{o,ijkl} \left(\omega_{ijkl} - \omega_{0,ijkl}\right)^2 \quad (7)$$

For non-bonded interactions, two main types are significant. In van der Waals interactions, the dispersive force has an attractive contribution at long-range distances and a repulsive contribution at short-range distances. The most well-known function to build the van der Waals model is the Lennard-Jones function $V_{LJ}(r_{ij})$ for the distance r_{ij} between atoms i and j, which is usually defined as:

$$V_{LJ}(r_{ij}) = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (8)$$

where ε_{ij} and σ_{ij} are adjustable parameters that represent the well depth of the potential profile; the distance for particle-particle potential energy V equals to zero, respectively. As shown in Fig. 2, two interacting particles repel each other at too close distances, attract each other at intermediate distances, and finally do not interact at the infinite distance. The Lennard-Jones potential (LJ potential) has a minimum value when the distance of r is equal to r_m (nearly $2^{1/6}\sigma$), where the value of potential energy V is equal to $-\varepsilon$.

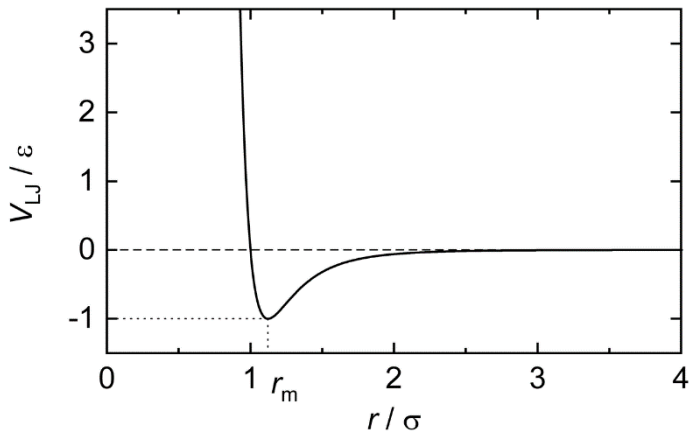


Figure 2. Plot of the Lennard-Jones potential function: Intermolecular potential energy V_{LJ} versus the distance between a pair of particles.

A variety of reliable force fields have been continuously developed and utilized as a computation method for estimating the forces between atoms within molecules and between molecules. In particular, force fields widely used in biomolecular simulations include Chemistry at Harvard Macromolecular Mechanics (CHARMM),[8] Assisted Model Building with Energy Refinement (AMBER),[9,10] Optimized Potentials for Liquid Simulations All-Atom (OPLS-AA)[11] and others.

2.2.3 MD simulation process

Based on calculations on a molecular system over time, MD simulations provide detailed information on the conformational changes and thermodynamics of biomolecules and their complexes. The whole process of MD simulations can be briefly described as follows:

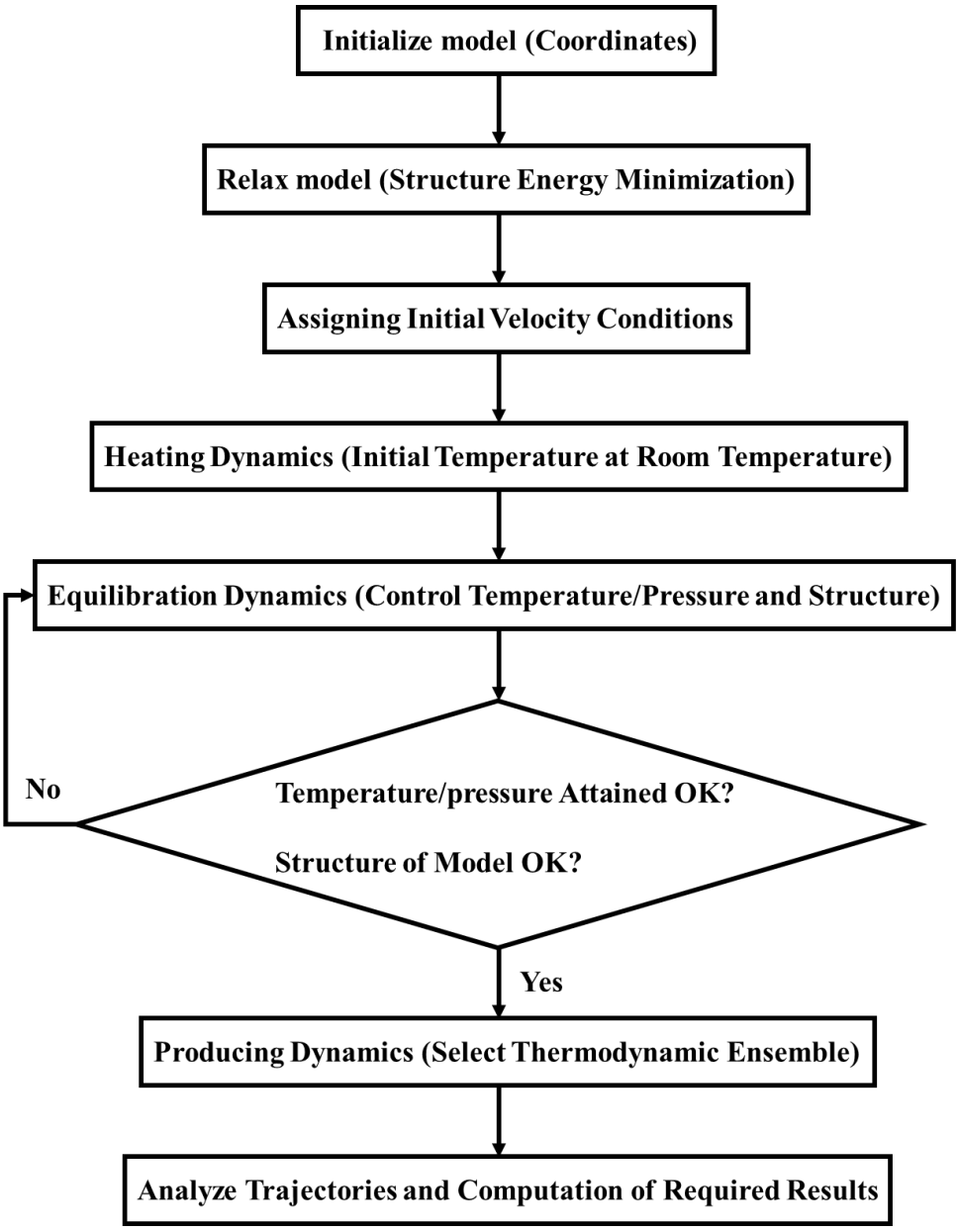


Figure 3. Flow chart of MD process

2.3 MD modelling and simulation conditions

We firstly designed a series of poly(PR) variants ($(PR)_4$, $(P_2R_2)_2$ and P_4R_4) with the same total length at 8 and Arg density but different periodicities (periodicity size t = 2, 4, 8) and homopolymeric adenine poly-rA as a model for RNA, as shown in Fig. 4.

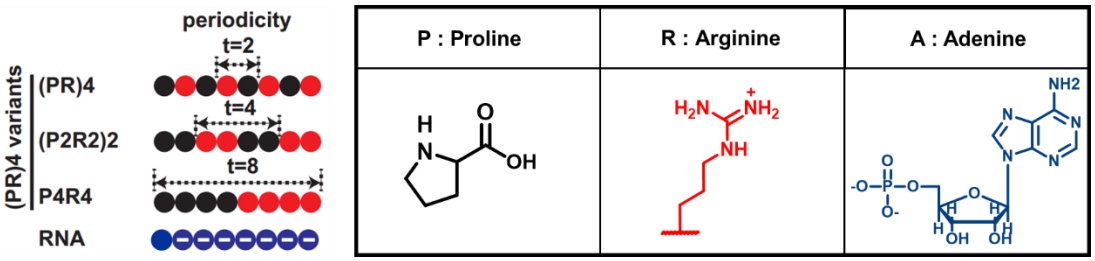


Figure 4. Designed poly(PR) variants with different periodicities (t = 2, 4, 8) and RNA sequence and composition group. Red circles indicate positions of Arg.

We then modeled these biomolecules using Avogadro software, which has been developed as an advanced tool to edit and visualize biomoelcules.[12] All $(PR)_4$ variants modelling structures and RNA (rA_8) are shown below:

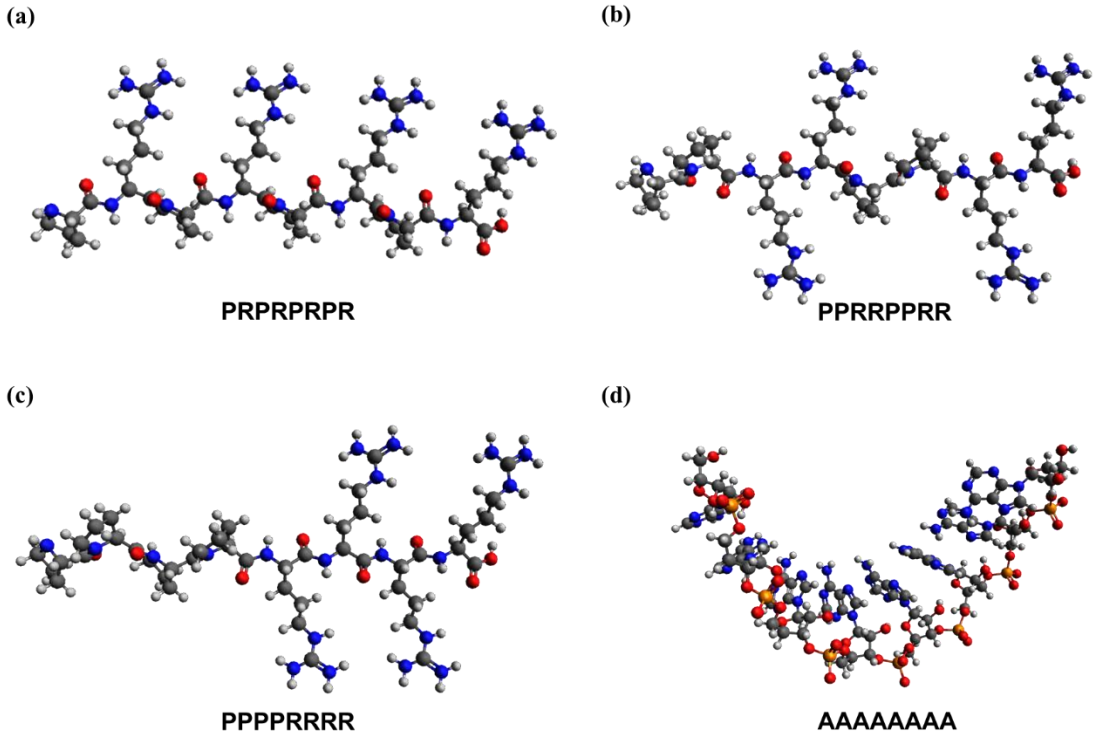


Figure 5. $(PR)_4$ variants (a) PRPRPRPR; (b) PPRRPPRR; (c) PPPPRRRR and RNA (rA_8) with the sequence AAAAAAAA.

Next, we performed MD simulations through GROMACS software in version 5.1.4, one of the most widely used open-source and freeware codes in chemistry and primarily

used for dynamic simulation of biomolecules.[13,14] Amber99sb-ildn force field was used for all energy parameters of inter and intramolecular interactions.[15,16] Simulations were performed in a cubic box (50 × 50 × 50 $Å^3$) with a periodic boundary condition in three-dimensional orientations, where all complexes between $(PR)_4$ variants and RNA were centralized in the box. The entire system was solubilized with an explicit TIP3P water model and neutralized via the addition of sodium ions.[17,18] All simulated systems were studied at pH 7, with both N-terminal Pro residues and C-terminal Arg residues protonated, and final initial entire simulation system is shown below:

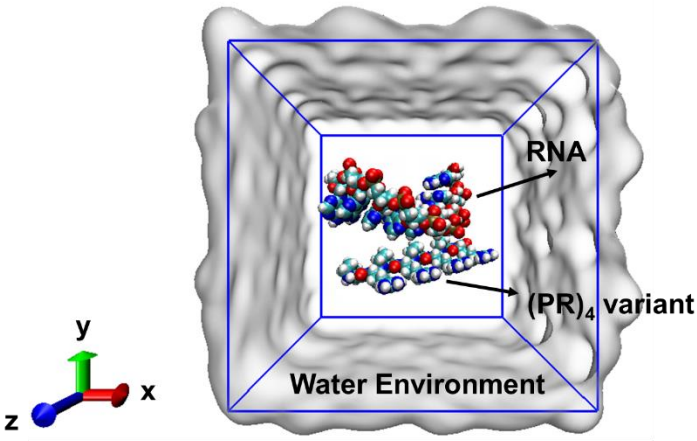


Figure 6. Scheme of MD simulation including one $(PR)_4$ variants and one rA_8 in a water box.

Subsequently, the system energetic minimization was performed by the steepest descent method to reach a maximum force below 1000 kJ/mol. The system was pre-equilibrated to 1 ns in the canonical (NVT) ensemble using the V-rescale coupling method at a constant 300 K temperature and to 1 ns in the Isothermal-isobaric (NPT) ensemble using the Parrinello-Rahman coupling method at a constant 1 bar pressure.[19] The coupling constants were defined at 0.1 and 2.0 ps for temperature and pressure, respectively. Electrostatic interactions were obtained by the particle mesh Ewald method[20]; Van der Walls interactions were calculated at a distance smoothing cutoff of 10 Å. The LINCS algorithm[21] was utilized to constrain the length of bonds, thus allowing the use of an integration time step of 2 fs. Finally, a 30 ns MD simulation was performed for all complexes. The simulation trajectories were visualized with the

Visual Molecular Dynamics (VMD) program, which belongs to one molecular graphics program displaying and analyzing combinations of molecules, in particular biopolymers like nucleic acids and proteins.[22] This MD calculation focuses on the one-by-one interactions between a peptide and RNA in an aqueous environment.

2.4 MD simulation analysis

2.4.1 Root-mean-square deviation and root-mean-square fluctuation

The root-mean-square deviation (RMSD) of atomic coordinates after superposition of optimal rigid bodies of the two structures has proven to be the most popular choice for quantifying structural and dynamic differences in macromolecules.[23] The equation of RMSD calculation is defined as:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\delta_i)^2} \tag{9}$$

where δ_i denotes the distance between atom *i* and reference structure of *N* equivalent atoms. Thus, the smaller the RMSD between two structures is, the more similar these two structures are. If we calculate the RMSD between two groups of atomic coordinates throughout the trajectory, this value can be used as a measurement for the degree of conformational variation of the molecule.

In addition, the other calculated parameter is root-mean-square fluctuation (RMSF), which evaluates the regional dynamical changes for the molecule, and thus one can specifically distinguish regions of the molecule experiencing huge structural fluctuations through the simulation, with higher RMSF values implying higher flexibility.[24] The RMSF function is defined as:

$$RMSF = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(r_i - r_{i,ref})^2} \tag{10}$$

where r_i and $r_{i,ref}$ denote the instantaneous coordinate vector and the reference coordinate vector of the *i*th atom, respectively. *T* is the statistical whole simulation time.

Given the RMSD function, we performed calculations on the biomolecular backbone to evaluate structural deviations along the trajectories for all complex systems. We analyzed the RMSD of each peptide by comparing it with the reference structure, as depicted in Fig. 7 (a), and also for the RMSF, we calculated the fluctuation of all atoms in the peptide structure over the whole simulation time range, as shown in Fig. 7 (b).

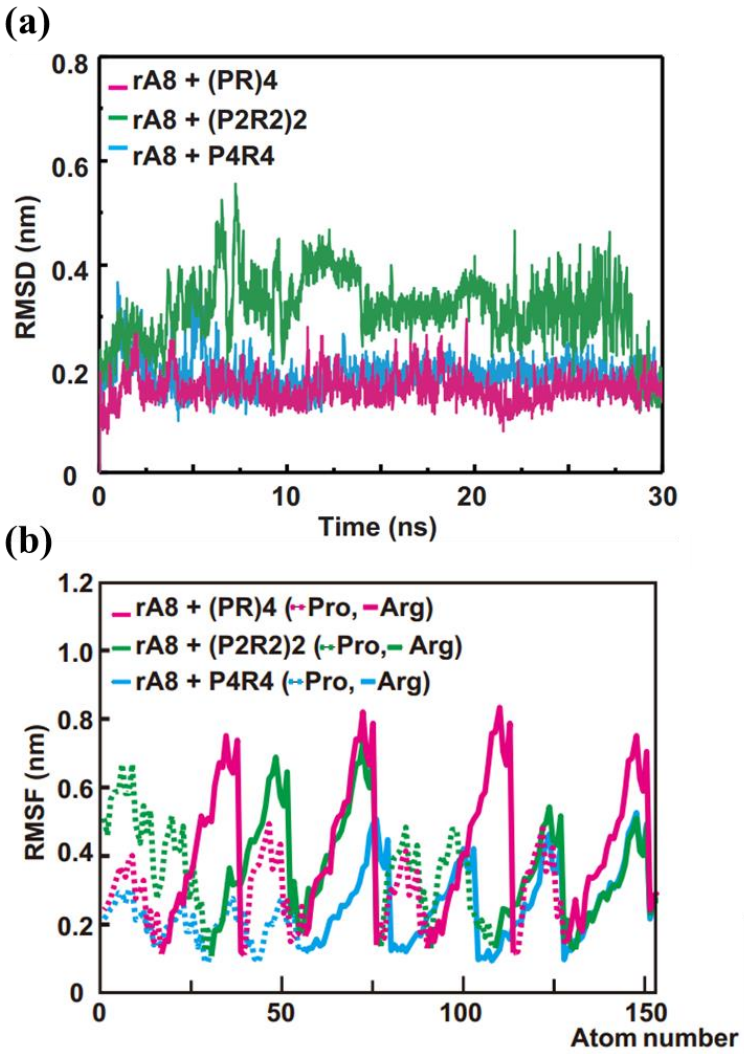


Figure 7. (a) Root-mean-square deviation (RMSD) and (b) root-mean-square fluctuation (RMSF) of $(PR)_4$ variants and RNA complex systems during MD simulations. The solid lines refer to Arg amino acid and the dashed lines refer to Proline amino acid.

Apparently, the RMSD values of $(PR)_4$ and P_4R_4 were relatively small and constant throughout 30 ns, while the value of $(P_2R_2)_2$ was large, reflecting that proline plays a key role in the peptide rigidity during its interaction process with RNA. The RMSD of

$(PR)_4$ was the smallest among the peptides because prolines were uniformly inserted into the sequence. On the other hand, the RMSD of $(P_2R_2)_2$ was the largest, indicating that RR was more flexible than PR. Furthermore, RMSF showed that Arg residues mainly caused fluctuations when interacting with RNA.

2.4.2 Binding energy calculation

In order to investigate the primary driving forces of rA_8 and $(PR)_4$ variants interactions, molecular mechanics Poisson Boltzmann surface area (MM-PBSA) analysis was conducted through g_mmpbsa tool in GROMACS.[25,26] This method has been widely utilized in biomolecular research, including protein folding, protein-protein interactions and protein-ligand binding, etc.[27] Ideally, the binding process of two solubilized molecules (in the case of peptide and RNA) can be illustrated as follows:

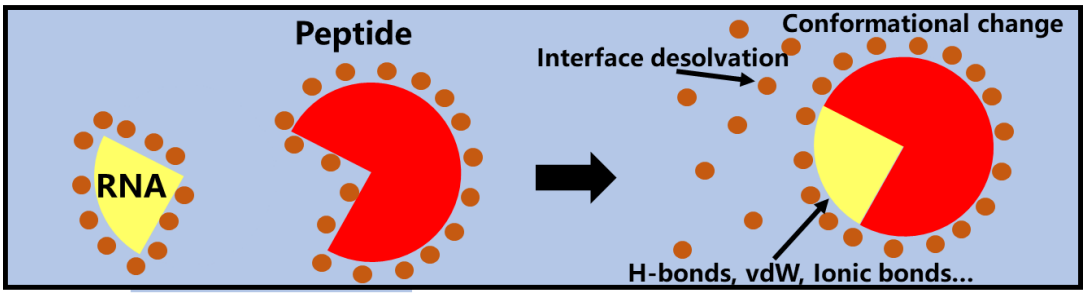


Figure 8. Binding process of peptide and RNA in the solution

Hence, the binding energy in general can be expressed as

$$\Delta G_{binding} = G_{complex} - (G_{RNA} + G_{Peptide}) \quad (11)$$

where $\Delta G_{binding}$ is the binding free energy, $G_{complex}$, G_{RNA} and $G_{Peptide}$ refer to free energies of the complex, RNA and peptide in the solvent, respectively.

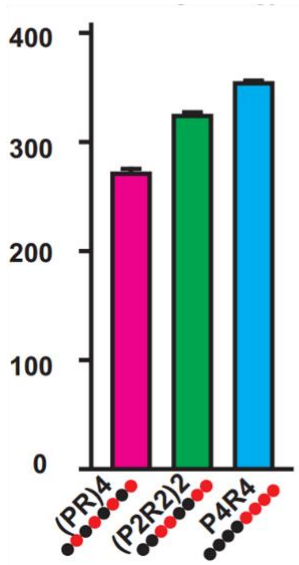
$$\Delta G_{binding} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (12)$$

where ΔE_{MM} denotes the intramolecular energy difference in vacuum, ΔG_{sol} denotes the solvation free energy difference, T denotes the absolute temperature, and ΔS

denotes the entropic change. ΔE_{MM} is calculated through the molecular mechanics (MM) method; ΔG_{sol} includes the polar solvation free energy and the nonpolar solvation energy difference, and the former one is calculate through the Poisson-Boltzmann (PB) equation and the latter one is fitted through solvent accessible surface area (SASA); $T\Delta S$ is calculated using normal mode method and is usually ignored altogether in the ranking of relative binding affinities, since its involvement cannot enhance the agreement with experiment.[28]

According to the g_mmpbsa method, the average binding energy was obtained based on the final 10 ns in whole MD trajectories. The dielectric constant of solvent was set to 80; the internal dielectric constant was set to 4; and the surface tension constant was set to 0.022 kJ/mol. The binding energy results for all simulated systems, including the energy contributions in term of all components, i.e., MM energy, polar energy and nonpolar energy, which are shown in Fig. 9 (a) and (b).

(a)



(b)

Energy contribution

	(PR)4	(P2R2)2	P4R4
vdW	-81.6 +/-3.1	-100.5 +/-1.9	-190.4 +/-1.7
Electrostatic	-624.3 +/-15.5	-811.9 +/-7.9	-900.2 +/-5.5
Polar solv.	443.5 +/-15.7	601.5 +/-8.0	758.8 +/-7.0
SASA	-9.06 +/-0.35	-13.12 +/-0.20	-22.42 +/-0.19
Binding	-270.9 +/-4.5	-323.8 +/-3.4	-353.8 +/-2.5

Figure 9. (a) Binding energy between each $(PR)_4$ variant and rA_8; (b) Energy contribution of each component (vdW, electrostatics, polar solvation energy, non-polar solvation energy) for bind energy.

According to the results in Fig. 9, P_4R_4 had the highest binding energy, while $(PR)_4$ had the lowest one. Then, we examined each binding energy component according to the

equation:

$$\Delta G_{binding} = \Delta G_{vdw} + \Delta G_{electrostatic} + \Delta G_{polar} + \Delta G_{non-polar} \quad (13)$$

where ΔG_{vdw} is the van der Waals energy; $\Delta G_{electrostatic}$ is the electrostatic energy; ΔG_{polar} is the polar solvation energy and $\Delta G_{nonpolar}$ is the non-polar energy. The first two represent molecular interactions, while the last two are related to solvation energy. Both van der Waals forces and electrostatic forces showed negative values, and absolute values increased in the order of $(PR)_4$, $(P_2R_2)_2$ and P_4R_4 in Fig. 9 (b). Interestingly, all peptides had positive values of ΔG_{polar}, while $(PR)_4$ displayed significantly smaller values than the others, indicating that water solubilization was significantly modulated by the presence and spatial distribution of Pro.

Finally, we decomposed the energy of each residue in the peptide sequence and RNA (poly-rA) shown in Fig. 10 (a) - (c). The N-terminal Pro residue and all Arg residues except the C-terminal one mainly contributed to the binding energy. In addition, the ΔG_{polar} of each Arg depended on the peptide sequence. $(PR)_4$ displayed a relatively suppressed ΔG_{polar}, due to the presence of water molecules between $(PR)_4$ and RNA. However, P_4R_4 showed a relatively large ΔG_{polar}, indicating that water molecules around the peptide were excluded from the interaction between the peptides and RNA.

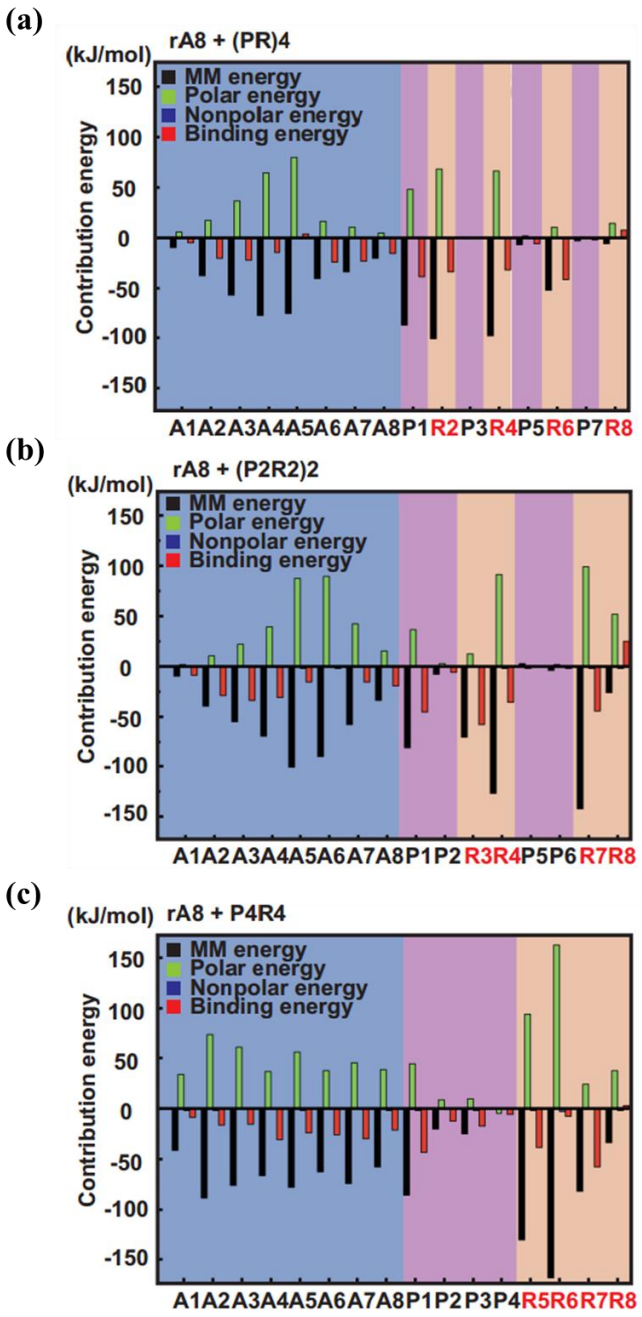


Figure 10. Dissection of energetic contribution into each component of binding energy. MM energy: molecular mechanic energy.

2.4.3 Free energy landscape

The free energy calculation is crucial for understanding the dynamics and the structural determinants in biomolecular processes, like ligand binding to receptors and enzymes, protein folding and unfolding, and the transportation of small molecules through biomolecular channels.[29] Here, the free energy landscapes of the dynamic behaviors of

rA_8 and $(PR)_4$ variant complexes were approximated via the sampling method. To obtain a energy landscape in two-dimension, we chose two reaction coordinates for each complex system: one is RMSD and the other is radius of gyration (Rg), which indicates the size and compactness for each complex system and is calculated using the equation:

$$Rg = \sqrt{\frac{\sum_{i=1}^{N} m_i {r_i}^2}{\sum_{i=1}^{N} m_i}} \tag{14}$$

where m_i denotes the atomic mass of the *i*th atom, r_i denotes the distance of atom *i* from the center of mass of molecule.

Then, the energy landscape was obtained along these two reaction coordinates using the following equation[30]:

$$G\alpha = -kTln\left[\frac{P(q_\alpha)}{P_{max}(q)}\right] \tag{15}$$

where k denotes the Boltzmann constant, T denotes the temperature in simulation system, $P(q_\alpha)$ denotes the probability density estimation through MD data histogram, $P_{max}(q)$ denotes the most probable state probability. q_i and q_j denote two different response coordinates. Finally, the two-dimensional free energy landscapes can be calculated through the joint probability distributions $P(q_i, q_j)$ in simulation system.

To make better comparisons, we also simulated the complex system containing rA_8 and R_8 (sequence: RRRRRRRR) and all under the same conditions as the other $(PR)_4$ variants of the simulated system. Finally, we plotted the free energy for system with respect to Rg and RMSD shown in Fig. 11 (a) - (d).

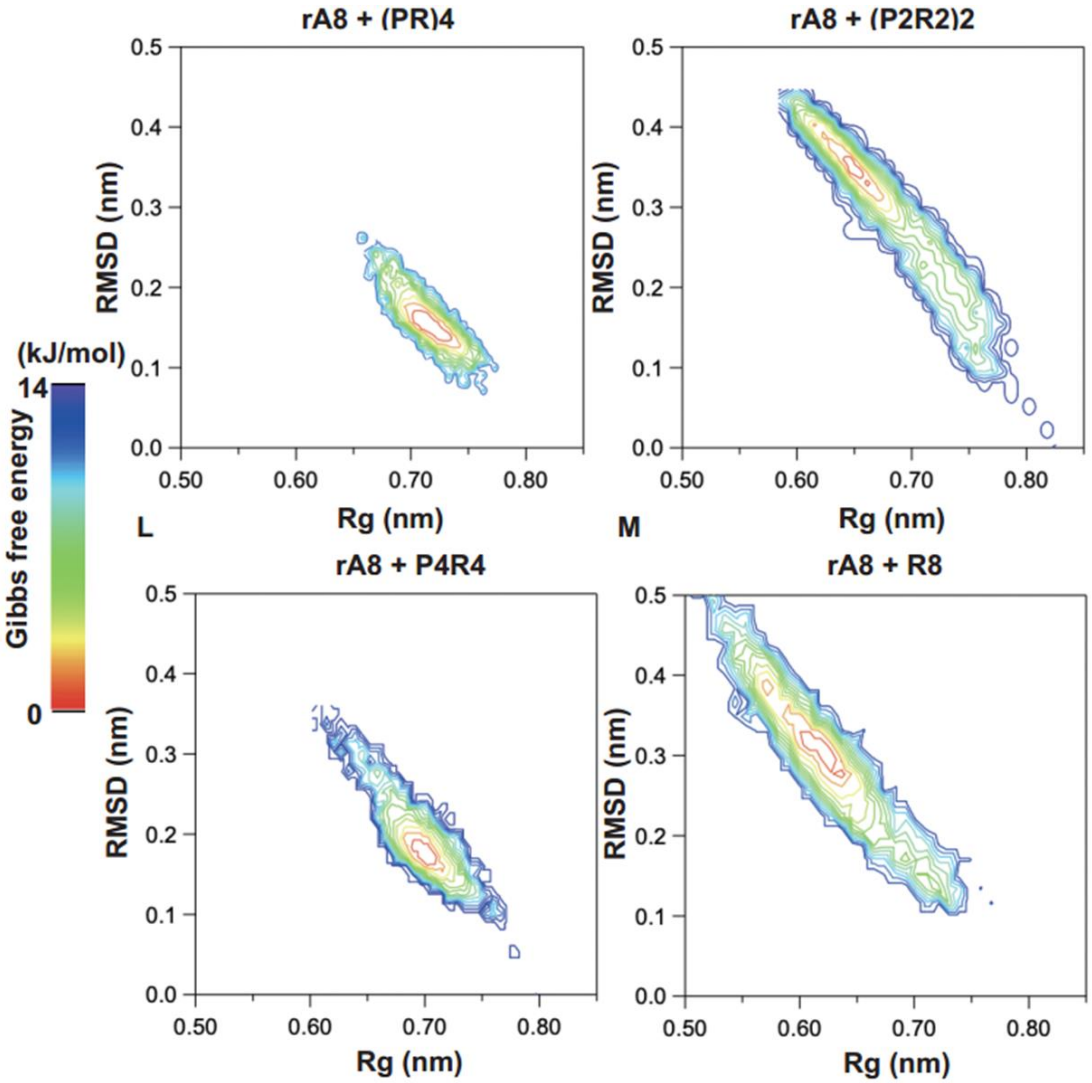


Figure 11. Free energy landscape of rA_8 and $(PR)_4$ variants and R_8.

Clearly, the energy minimum of $(PR)_4$ had a larger Rg value and a smaller RMSD, while the distribution of R_8 has a smaller Rg and a larger RMSD, reflecting that Pro residues conferred rigidity to the structure of the polypeptide. Therefore, we propose that this rigidity intervenes in the interaction between cationic amino acids and anionic nucleic acids. Our proposal would also be supported by related work suggesting that PR folds into a helical structure.[31]

2.5 Conclusion

In this section, we performed MD simulations to model the binding process between RNA and $(PR)_4$ variants. We found the alternating charge distribution in poly(PR) leads to a lower binding energy, while poly(PR) variants with higher periodicities (e.g. PnRn) have a higher binding energy when interacting with RNA. Additionally, we tested the function of Pro residues, which conferred structural rigidity to the polypeptide and thus interfered with the interaction between cationic amino acids and anionic nucleic acids.

2.6 References

1. Kosuge, M., Takeuchi, T., Nakase, I., Jones, A. T. & Futaki, S. Cellular internalization and distribution of arginine-rich peptides as a function of extracellular peptide concentration, serum, and plasma membrane associated proteoglycans. *Bioconjug. Chem.* **19**, 656–664 (2008).
2. Meloni, B. P. *et al.* The neuroprotective efficacy of cell-penetrating peptides TAT, penetratin, Arg-9, and Pep-1 in glutamic acid, kainic acid, and in vitro ischemia injury models using primary cortical neuronal cultures. *Cell. Mol. Neurobiol.* **34**, 173–181 (2014).
3. Meloni, B. P. *et al.* Poly-arginine and arginine-rich peptides are neuroprotective in stroke models. *J. Cereb. Blood Flow Metab.* **35**, 993–1004 (2015).
4. Sadus, Richard J. *Molecular simulation of fluids*. Elsevier, (2002).
5. Allen, Michael P., and Dominic J. Tildesley. *Computer simulation of liquids*. Oxford university press, (2017).
6. Lindahl, E., Abraham, M. J., Hess, B. & van der Spoel, D. GROMACS Documentation – Release, **2**–607 (2019).
7. Ponder, J. W. & Case, D. A. Protein Simulations. *Adv. Protein Chem.* **66**, 27–85 (2003).
8. Brooks, B. R. et al. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
9. Weiner, S. J. *et al.* A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **106**, 765–784 (1984).
10. Bayly, C. I. *et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
11. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).
12. Hanwell, Marcus D., *et al*. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform* **4**, 1-17 (2012).

13. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**–**2**, 19–25 (2015).
14. Pronk, S. *et al.* Molecular Simulation Workflows as Parallel Algorithms: The Execution Engine of Copernicus, a Distributed High-Performance Computing Platform. *J. Chem. Theory Comput.* **11**, 2600–2608 (2015).
15. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma.* **78**, 1950–1958 (2010).
16. Hornak, Viktor, *et al*. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinforma.* **65**, 712-725, (2006).
17. Mahoney, M. W. & Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **112**, 8910–8922 (2000).
18. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
19. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
20. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
21. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
22. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
23. Sargsyan, K., Grauffel, C. & Lim, C. How Molecular Size Impacts RMSD Applications in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **13**,

1518–1524 (2017).

24. Zhao, D., Peng, C. & Zhou, J. Lipase adsorption on different nanomaterials: A multi-scale simulation study. *Phys. Chem. Chem. Phys.* **17**, 840–850 (2015).
25. Kumari, R., Kumar, R. & Lynn, A. G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* **54**, 1951–1962 (2014).
26. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037–10041 (2001).
27. Wang, E. *et al.* End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* **119**, 9478–9508 (2019).
28. Yang, T. *et al.* Virtual screening using molecular simulations. *Proteins Struct. Funct. Bioinforma.* **79**, 1940–1951 (2011).
29. Sekijima, M. *et al.* Free Energy Landscape Analysis System Based on Parallel Molecular Dynamics Simulation. *IPSJ Digit. Cour.* **3**, 757–766 (2007).
30. Papaleo, E., Mereghetti, P., Fantucci, P., Grandori, R. & De Gioia, L. Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: The myoglobin case. *J. Mol. Graph. Model.* **27**, 889–899 (2009).
31. Edun, D. N., Flanagan, M. R. & Serrano, A. L. Does liquid-liquid phase separation drive peptide folding? *Chem. Sci*. **12**, 2474–2479 (2021).

CHAPTER 3. Molecular interaction affects phase separation formation of poly(PR) dipeptide repeats with various periodicities

3.1 Introduction

Complex coacervation is known as the phase separation through interactions among complexing polyions.[1] It has been shown PR polycation having 30 PR unit repeats formed phase separation when mixed with negatively-charged RNA, and the morphology of the resulting condensates were actually spherical droplets with minimized surface energy, which was consistent with LLPS driven by complex coacervation.[2] Taking the homopolymeric RNA (poly-rA) as an example, the liquid condensates formed by poly-rA and $(PR)_{30}$ not only had spherical shapes and glassy surface wettability, but also fused with each other to form larger condensates.[3]

Meanwhile, biomolecules rely on the proper placement of monomers to encode structural, functional, and physiological information. The use of charge in sequence-controlled biomaterials is common.[1,4,5,6] For instance, charge pattern has been reported to affect conformational behavior for intrinsically disordered proteins (IDPs), accompanied by theoretical work linking IDP sequence to charge-driven phase separation.[7,8] In addition, other previous studies have testified the distribution of charged amino acids affects electrostatic force and phase separation, particularly for blocky sequences of charged residues exhibit a strong charge correlation due to the sequence alignment of the two nearby chains, which facilitated the binding process.[4]

Here, we focused mainly on understanding how the periodicity of poly(PR) variants with the same length regulates LLPS formation when mixed with poly-rA as a model for RNA.

3.2 Two-body interaction between RNA and poly(PR) variants

3.2.1 Materials and experimental condition

We synthesized a series of $(PR)_{12}$ variants with different periodicities (periodicity size t = 2, 4, 8 and 24) with purity higher than 85% shown in Fig.1 by Genscript (Piscataway NJ), and poly-rA RNA was purchased from ThermoFisher Scientific (Waltham, MA). We also prepared chamber coverglasses (uncoated or hydrophilic polymer MAS-coated) obtained from Matsunami glass (Osaka, Japan). In addition, we mainly used fluorescence technique to observe LLPS formation, so TAMRA-labeled rA_{15} was synthesized by Fasmac (Japan).

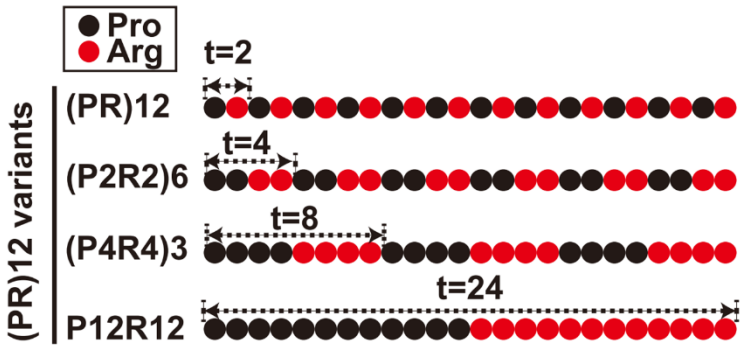


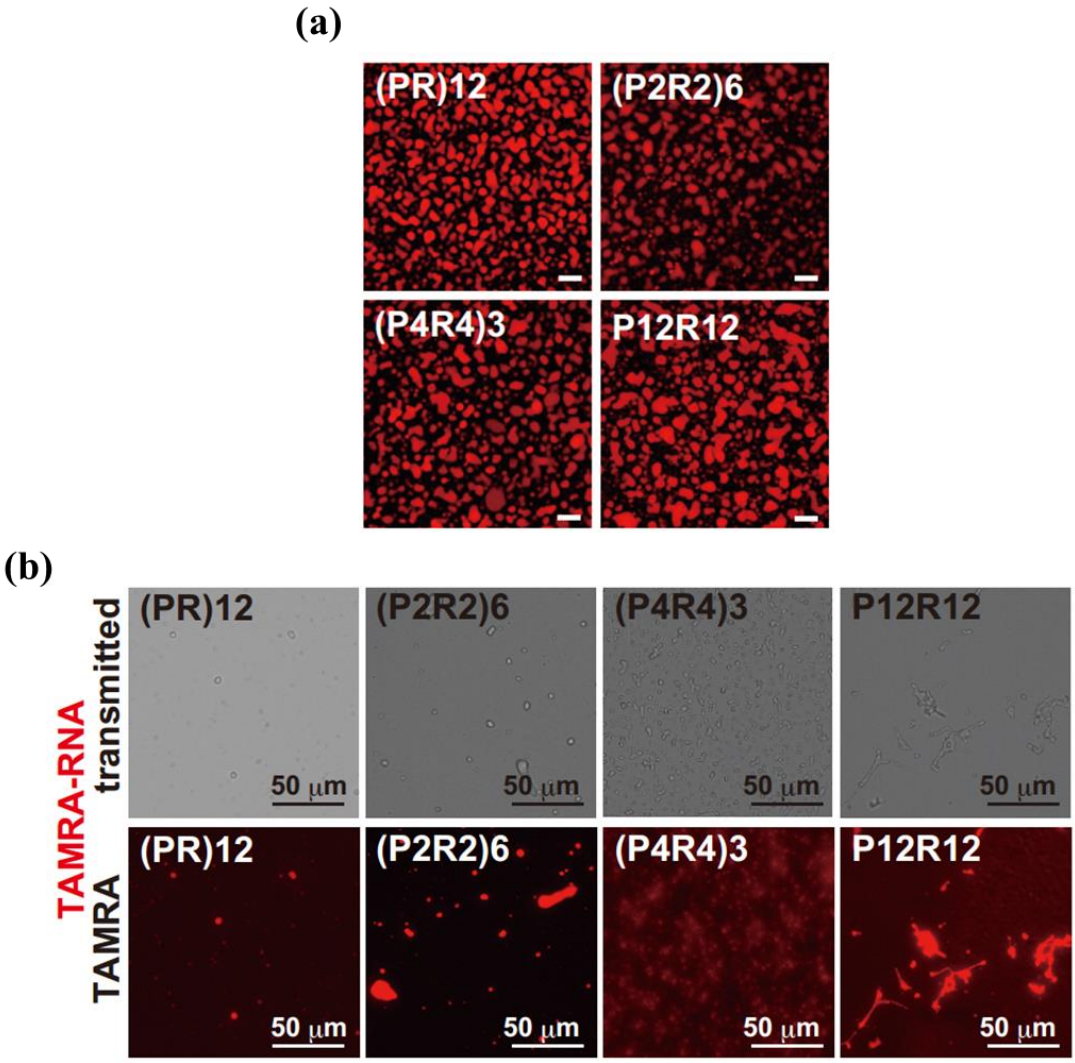
Figure 1. Structures of $(PR)_{12}$ variants with different periodicities (repeat size: t = 2, 4, 8 24).

As for the LLPS formation of $(PR)_{12}$ variants and poly-rA, the experimental concentrations were set as each peptide (final concentration 100 μM) and poly-rA (final concentration 0.5 mg/mL) containing 100 nM of tetramethylrhodamine (TAMRA)-labeled TAMRA-rA_{15}. After mixing with each other, we used an LSM-710 confocal microscope (Carl Zeiss, Oberkochen, Germany) to observe LLPS droplets on the chamber coverglass.

3.2.2 LLPS of poly(PR) variants with poly-rA on the substrate

We first conducted experiments to understand the molecular interactions between $(PR)_{12}$ variants and poly-rA. If charged molecules were present in a blocky sequence, it was expected that nearby molecules would enhance the interaction by electrostatic

forces.[4] After making freshly mixed solutions, we dropped them on the surface of hydrophobic uncoated coverglass and hydrophilic MAS-coated coverglass. When we observed the phase separated droplets of $(PR)_{12}$ variants and RNA on the hydrophobic glass surface, all these droplets similarly wetted the surface shown in Fig. 2 (a). However, when using a glass coated with a hydrophilic polymer, most droplets floated and fewer droplets adhered to the surface shown in Fig. 2 (b). Droplets with the larger periodic $(PR)_{12}$ variants lost their spherical structure under this condition, while $P_{12}R_{12}$ formed sticky non-spherical condensates that failed to minimize surface tension, which



is a typical characteristic of liquids.

Figure 2. $(PR)_{12}$ variants with poly-rA containing TAMRA-labeled rA_{15} on the surface of (a) uncoated coverglass with LLPS, scale bar is 10 μm and (b) hydrophilic MAS-coated coverglass surface.

3.2.3 LLPS fluidity of poly(PR) variants with poly-rA

3.2.3.1 FRAP principle

Fluorescence recovery after photobleaching (FRAP) have been widely used as a fluorescence-based biophysical method, especially the study on diffusion kinetics in living cells.[9] It was first described by Axelrod *et al* in 1976[10] and is capable of quantitatively and qualitatively examining the diffusion of membrane-bound or soluble particles containing fluorescently labeled probes. In this thesis work, the method was used to quantify the diffusion coefficient of LLPS droplets. The principle is schematically represented as below:

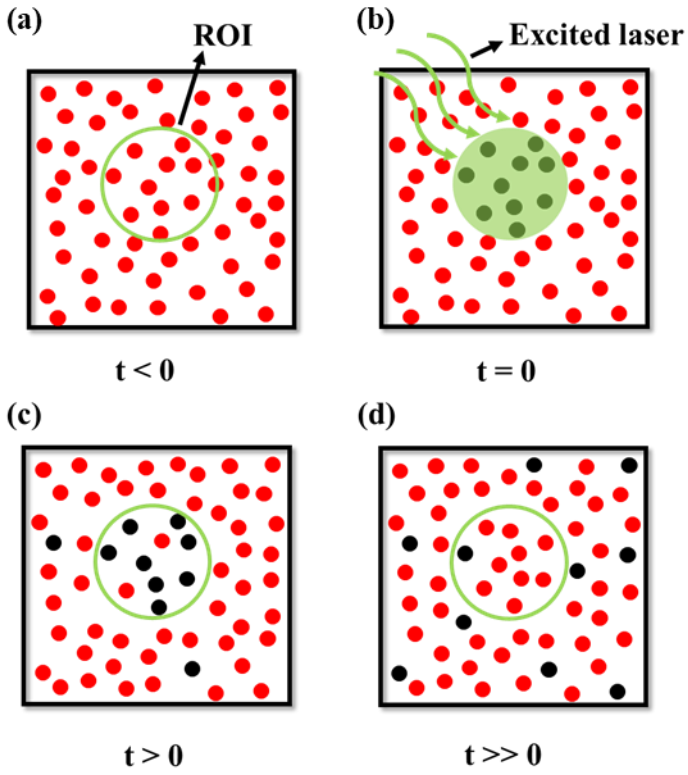


Figure 3. Schematic illustration of a fluorescence recovery after photobleaching (FRAP) experiment. The molecular system consisting of fluorescently labeled ones (represented by the red color) is prepared.

At time point t = 0, all fluorophores within a defined region of interest (ROI) are photobleached by exposing them to a laser pulse with high intensity at their excitation wavelength shortly. Thus, the overall fluorescence intensity within RIO will be much

smaller than compared to when t < 0 due to the rapid fading of the fluorescence lifetime. Due to Brownian motion, the bleached molecules exchange with the unbleached molecules in the surrounding area over time, the photobleached fluorophores start to diffuse out of ROI, while other molecules with fluorescently active tags diffuse into it when t > 0. As a result, the fluorescence intensity in ROI returns to a certain saturation value when t >> 0. Based on the FRAP plot, we can get information about the diffusion constant and binding dynamics of the fluorescently tagged molecules.[11] A typical FRAP plot is depicted in Fig. 4:

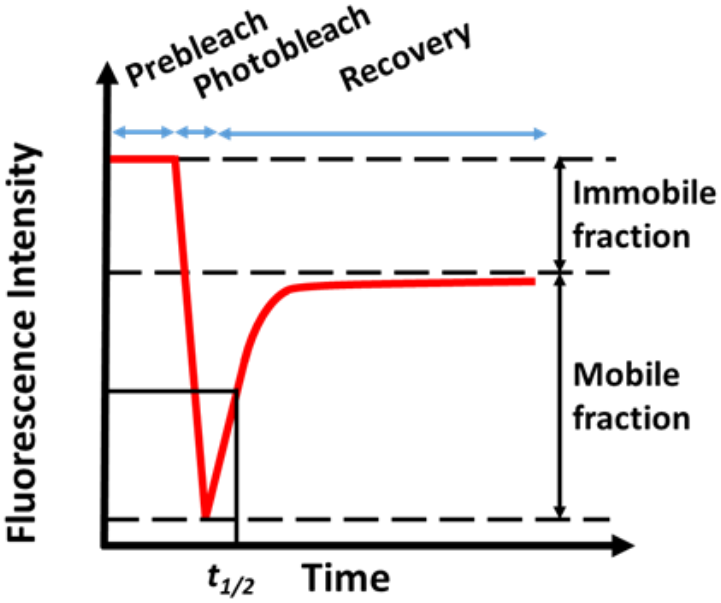


Figure 4. A typical FRAP plot. Various information, including the mobile and immobile fractions, the half time of equilibrium ($t_{1/2}$) can be calculated from the plot as shown.

A series of FRAP measurements were performed using the LSM-710 confocal microscope, and the experimental data were analyzed with ZEN software (Carl Zeiss) with a lens of numerical aperture (N.A.) at 1.2. The excitation wavelength was 561 nm.

3.2.3.2 Periodicity affects diffusion of LLPS droplets

After mixing each $(PR)_{12}$ variant with poly-rA containing TAMRA-labeled rA_{15}, FRAP

measurements were therefore conducted and repeated six times for each mixture group. The final results were shown in Fig. 5:

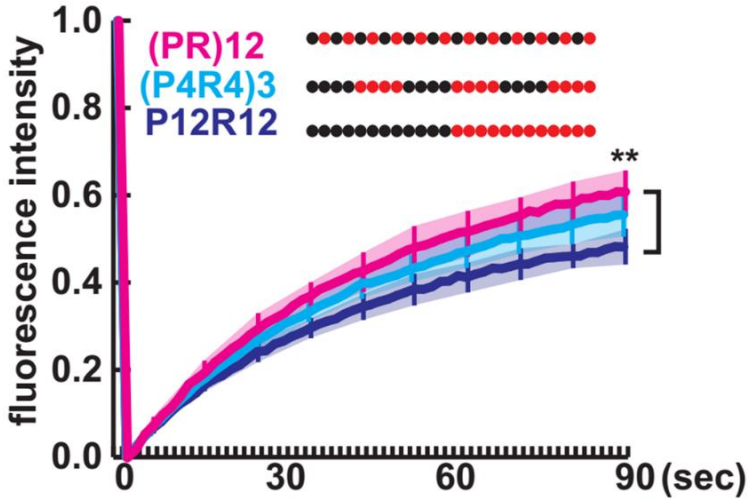


Figure 5. FRAP analysis of TAMRA-RNA phase separation using the $(PR)_{12}$ variant. The lines are the means of five droplets. Asterisks mean significant differences derived from one-way ANOVA with Tukey's test (**: $p < 0.01$).

Apparently, RNA mixed with $(PR)_{12}$ had higher fluidity than RNA mixed with $(P_4R_4)_3$ or $P_{12}R_{12}$ throughout the whole recovery time 90 s, reflecting the distribution of charged amino acids in $(PR)_{12}$ modulates the biochemical behavior of client molecules in LLPS. Meanwhile, it was also shown the periodicity value was inversely proportional to the diffusion ability of LLPS droplets.

3.2.3.3 Proline loosens Arg-RNA interaction

Considering that the fluidity of $(PR)_{12}$ droplets is higher than that of $(PR)_{12}$ variants with large periodicities, we then hypothesized that Pro residues might be expected to act as looseners for the Arg-RNA interactions. To demonstrate this, whole droplet bleaching showed that RNA in the $(PR)_{12}$ droplet fraction was more capable of entering and exiting than RNA in R_{12} droplets shown in Fig. 6. According to the curve tendency, the fluorescence intensity of R_{12} droplets quickly becomes flat with much lower values, which is mainly due to the strong electrostatic interaction between positive Arg residues

and the negative RNA. On the contrary, the fluorescence intensity of $(PR)_{12}$ droplets gradually increases until relatively much higher value since there were Pro residues inserted into the peptide structure, which can adjust the interaction strength between $(PR)_{12}$ and RNA.

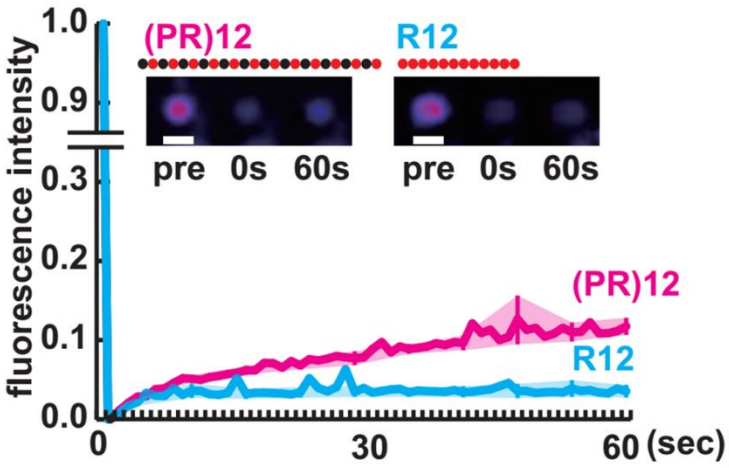


Figure 6. FRAP analysis of whole droplets of R_{12} or $(PR)_{12}$ mixed with RNA. The lines show the average of six droplets/conditions. The scale bar is 1 μm.

To further validate the function of Pro as looseners, we tested the effects of $(PR)_{12}$ variants with different proportions of Proline on LLPS formation. In this research, we synthesized different ratios of $(PR)_{12}$ variants with the same amount of Arg, $(P_1R_2)_6$ (Pro : Arg ratio = 1 : 2) and $(P_1R_3)_4$ (Pro : Arg ratio = 1:3), or $(PR)_{12}$ variants with the same amount of Pro, and evaluated their biochemical properties. When mixed with RNA, $(PR)_{12}$, $(P_1R_2)_6$ and $(P_1R_3)_4$ formed droplets shown in Fig. 7 (a). Additionally, FRAP analysis showed the fluidity in the droplet increased when the proportion of Pro residues increased, as shown in Fig. 7 (b). Based on the analysis of these results, we can summarize alternating Pro residues increase the internal mobility of LLPS droplets.

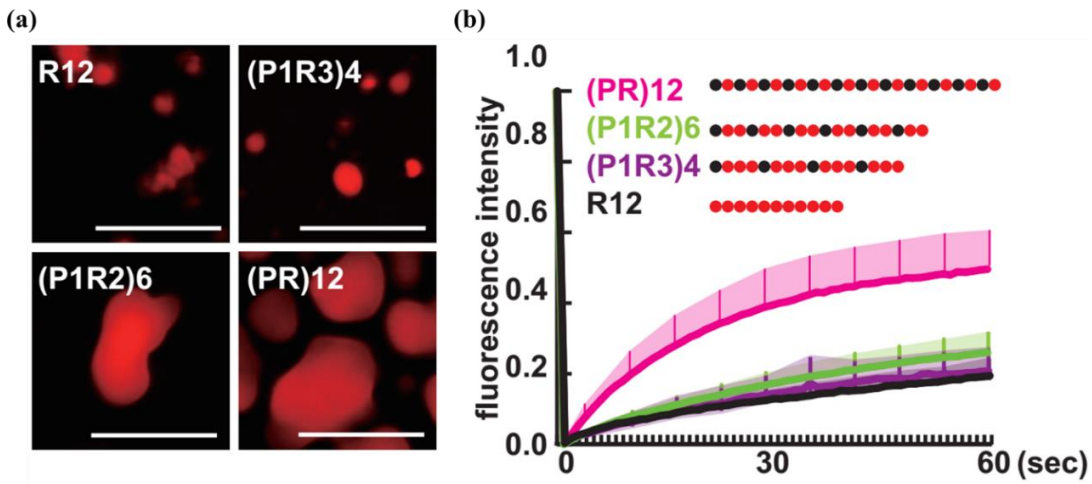


Figure 7. (a) Droplets of phase-separated droplets of $(PR)_{12}$ variants mixed with poly-rA containing TAMRA-labeled rA_{15} (scale bar is 10 μm); (b) FRAP analysis of TAMRA-RNA phase-separated with $(PR)_{12}$ ratio variants. Lines show the average of 8 - 10 droplets/condition.

3.2.3.4 CSC experiment on the binding strength between poly(PR) variants with poly-rA

The critical salt concentration (CSC) has been shown to be closely related to the size of the period block of the charged polyelectrolytes, which allows phase separation from oppositely charged polymers in aqueous solution.[4] Increasing the periodicity value leaded to a significant increase in CSC, and the binding strength was determined using CSC.[4,12] Increasing the salt concentration may interfere with ionic interactions and thus hinder LLPS. To validate the results of MD simulations, we performed CSC experiments. As shown in Fig. 8, we compared four types of $(PR)_{12}$ variants, $(PR)_{12}$, $(P_2R_2)_6$, $(P_4R_4)_3$ and $P_{12}R_{12}$ mixed with RNA. Consequently, when the concentration of NaCl was increased in a mixture containing phase-separated droplets composed of $(PR)_{12}$ variants and RNA, $P_{12}R_{12}$ droplets survived up to 800 mM, while the $(PR)_{12}$ droplets solubilized up to 700 mM, indicating that the stronger interactions were caused by blocky charge sequences. At the same time, higher values of blocky size were associated with higher turbidity, especially at NaCl concentration up to 700 mM and 800 mM.

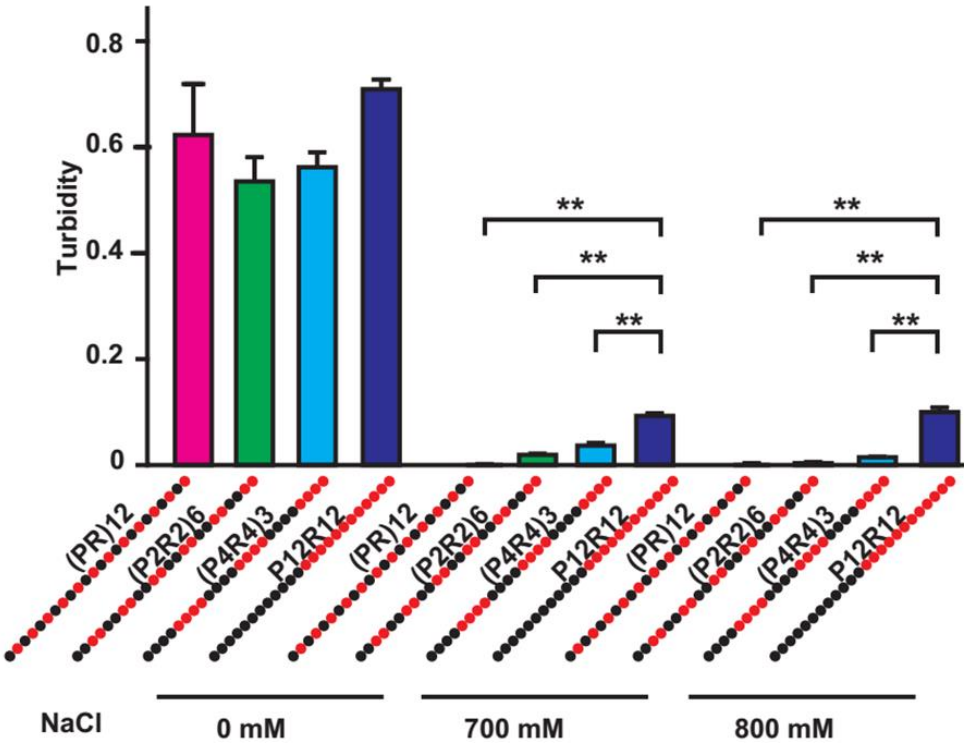


Figure 8. Critical salt concentration (CSC) of $(PR)_{12}$ variants reduced phase separation with RNA. Turbidity was determined by OD600; N = 3 biological replicates. All data are shown as mean ±S.D. Asterisks indicate significant differences derived from one-way ANOVA with Tukey's test (**: $p < 0.01$).

In conclusion, combining all experimental results and previous MD simulations, we plotted the two-body interaction process between $(PR)_n$ variants and RNA shown in Fig. 9. On the one hand, the electrostatic interaction between positively charged $(PR)_n$ variant and negatively charged RNA molecule will lead to the binding event with the binding strength directly proportional to the blocky size. On the other hand, Pro residues contribute to the structural rigidity of the peptide

and the increased fluidity of complex coacervates formed by $(PR)_n$ variants and RNA, so the alternation of Pro residues leads to higher multivalency over other distribution events.

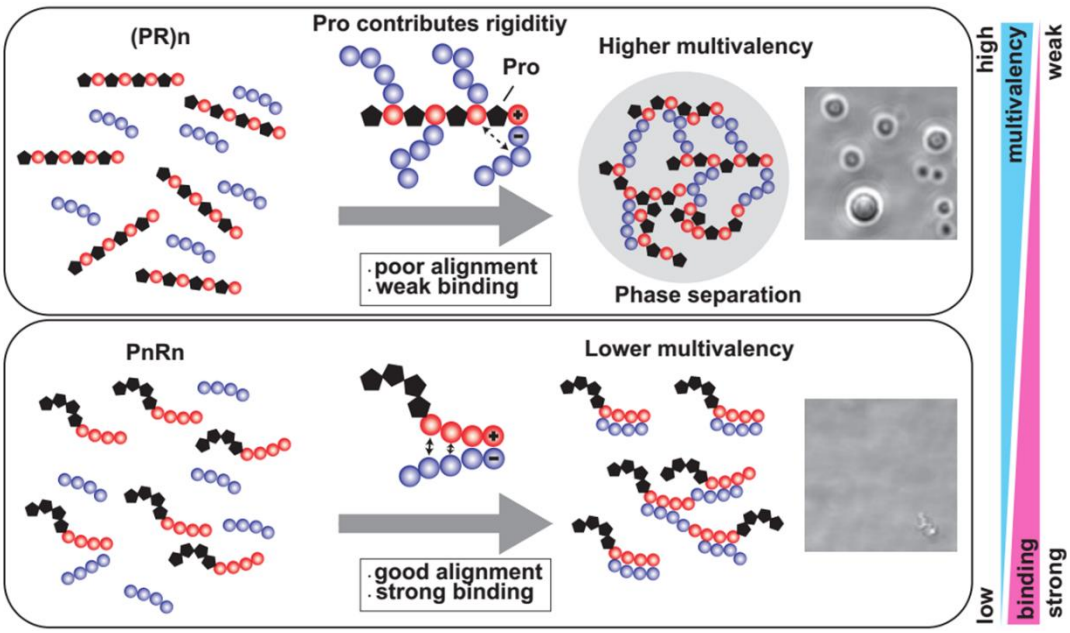


Figure 9. Multivalent interactions and phase separation schemes of $(PR)_n$ and P_nR_n.

3.3 Three-body interaction among RNA, poly(PR) variants and protein

Poly(PR) peptides have been reported to concentrate nearly exclusively within the outermost layer of nucleolus, where they co-localize with an abundant, multifunctional phosphoprotein, namely nucleophosmin (NPM1).[13,14,15] Finally, the nucleolus had the liquid-like properties.[5,16] As for NPM1with its overall structure shown in Fig. 10, the protein consists mainly of an N-terminal oligomerization domain (OD), a C-terminal nucleic-acid-binding domain (NBD) and a central intrinsically disordered region (IDR).[17] Furthermore, NPM1 was shown to undergo heterotypic LLPS with 20 PR repeats $(PR)_{20}$ as well as 23 PR repeats $(PR)_{23}$ of the peptide exhibiting a sharp, concentration-dependent transitions from the primary single-phase (monodisperse state) to the two-phase state (demixed state) to the final single phase (dissolved state).[13,17] Here, we first investigated the Arg distribution in LLPS during diffusion by mixing each poly(PR) variant and NPM1 protein.

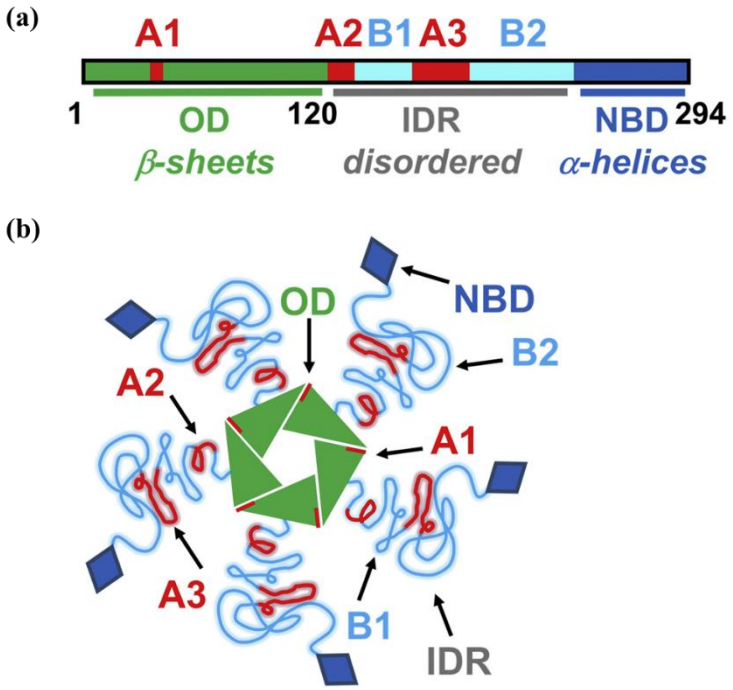


Figure 10. Schematic organization of (a) NPM1 structural domain: acidic bundles (A1-A3, red), basic bundles (B1 and B2, cyan), the oligomeric structural domains (OD; green), intrinsically disordered regions (IDR; gray), and the nucleic acid binding domains (NBD; blue);[17] (b) NPM1 oligomeric structure with the same color as (a), came from the crystal structure of OD, NMR structure of NBD, the disordered nature of IDR and protein assembly as a pentamer.[16,18-20]

3.3.1 Materials and experimental condition

Poly(PR) variants $(PR)_{50}$, $(P_4R_4)_{12}$ and $(P_8R_8)_6$ with the following sequence were synthesized by Genscript (Piscataway NJ) with a purity > 85% in Fig. 11. Human rRNA was extracted from HEK293 cells by the RNeasy kit (Qiagen). Human NPM1 cDNA was amplified from NPM1-GFP gifted from Xin Wang (Addgene plasmid #17578; http://n2t.net/addgene:17578; RRID: Addgene_17578) by PCR and subcloned into pET vector. The recombinant NPM1 expression was induced by 0.1 mM Isopropyl-β-D-thiogalactopyranoside (IPTG) at 16 ℃ for overnight, and the expressed NPM1 was purified using Ni-sepharose columns (Clontech, Mountain View, CA). The fluorescent labeling of recombinant NPM1 was achieved using the HiLyte-Fluor 555 labeling kit (Dojindo, Kumamoto, Japan), following manufacturer's protocol. FRAP analysis was

also conducted by LSM-710 confocal microscope with a lens of N.A. = 1.2, data analysis was conducted via ZEN software (Carl Zeiss). To observe LLPS of $(PR)_{50}$ variants and human NPM1, each peptide (final concentration 50 μM) and NPM1 (final concentration 20 μM) containing 0.2 μM of NPM1 labeled with HiLyte555 were mixed. To observe LLPS of $(PR)_{12}$ variants, human NPM1 and human rRNA, each peptide (final concentration 50 μM), NPM1 (final concentration 20 μM) containing 0.2 μM of NPM1 labeled with HiLyte555 and 50 ng/μL of human rRNA were mixed. RNA was visualized using SYTO-RNASelect (ThermoFisher).

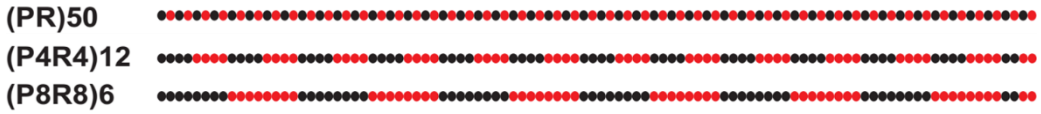


Figure 11. Structures of $(PR)_{50}$ variants with different periodicities (repeat size: t = 2, 8 and 16). To equalize the numbers of Pro and Arg to $(PR)_{50}$, P_2R_2 was added at the C-terminus of $(P_4R_4)_{12}$ and $(P_8R_8)_6$.

3.3.2 LLPS fluidity of poly(PR) variants with protein

Likewise, we did several FRAP experiments to observe the fluidity of green fluorescent protein (GFP)-tagged $(PR)_{50}$ variants expressed in HEK293 cells, and the results are shown in Fig. 12. As a result, GFP-$(PR)_{50}$ had higher fluidity than that of GFP-$(P_4R_4)_{12}$ and GFP-$(P_8R_8)_6$, indicating that compared to the other groups, the fluidity of $(PR)_{50}$ with the nucleosomes in the interactions between peripheral molecules were relatively loose.

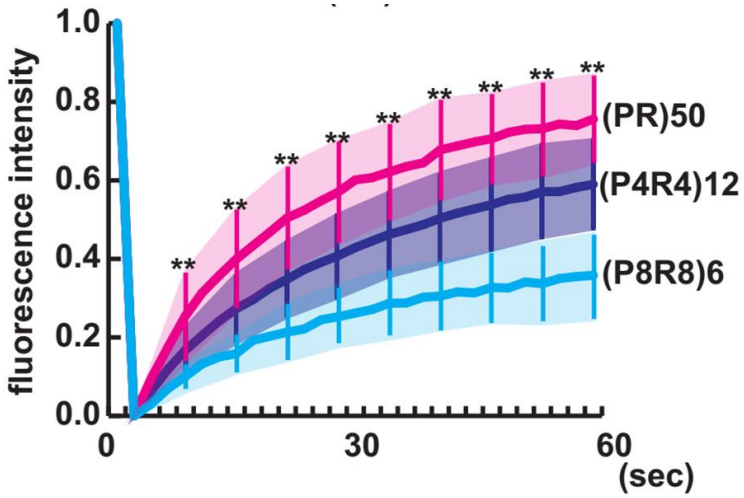


Figure 12. FRAP analysis of GFP-$(PR)_{50}$, GFP-$(P_4R_4)_{12}$ or GFP-$(P_8R_8)_6$ expressed in HEK293 cells; N = 10. Asterisks indicate significant differences analyzed by Student-*t* test (**: $p<0.01$).

Next, we tested the effect of $(PR)_{50}$ variants on GFP-NPM1 diffusion. As shown in Fig. 13, $(PR)_{50}$ inhibited the GFP-NPM1 mobility; however, the effect of $(P_4R_4)_{12}$ was not significant and $(P_8R_8)_6$ had little effect.

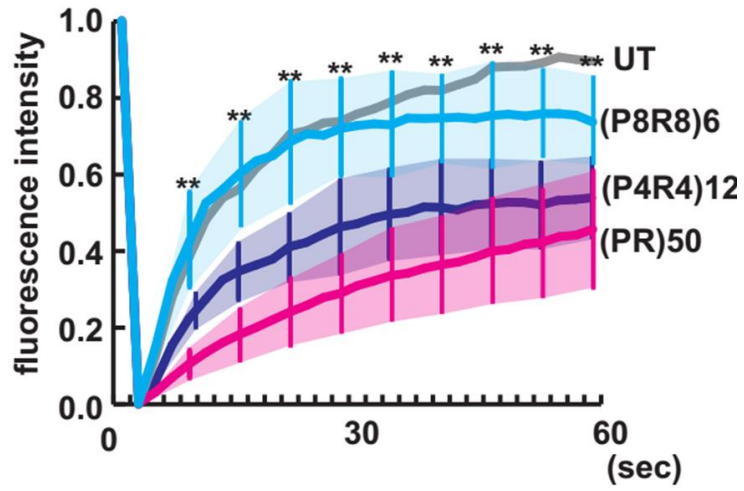


Figure 13. FRAP analysis of GFP-NPM1 co-expression with mCherry-$(PR)_{50}$, mCherry-$(P_4R_4)_{12}$ or mCherry-$(P_8R_8)_6$. N = 10 cells/variant. Error bars indicate ±S.D. Asterisks indicate significant differences analyzed by Student-*t* test (**: $p<0.01$).

To quantitatively evaluate FRAP results, we used a simplified equation to extract

diffusion coefficient from FRAP data through the half time of recovery.[21,22]

$$F(t) = \frac{F_0 + F_\infty (t/\tau_{1/2})}{1 + (t/\tau_{1/2})} \quad (1)$$

Where $F(t)$ denotes the fluorescence intensity of bleached ROI, F_0 denotes the initial fluorescence intensity after bleaching, F_∞ denotes the steady-state fluorescence intensity after bleaching, and $\tau_{1/2}$ denotes half of the recovery time.

Finally, the calculation results were shown in Fig. 14. Half recovery time indicated that diffusion of NPM1 co-expressed with $(PR)_{50}$ was much slower than that affected by $(P_8R_8)_6$, which implied that more NPM1 molecules are affected by $(PR)_{50}$ than by $(P_8R_8)_6$.

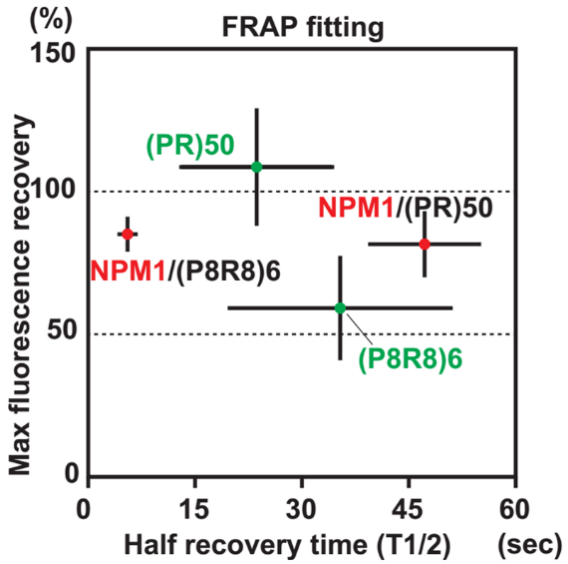


Figure 14. Fitting of FRAP curves for co-expression of GFP-$(PR)_{50}$ variants or NPM1 with $(PR)_{50}$ variants.

3.3.3 Poly(PR) variants affect the phase separation between protein and RNA

Based on the results of LLPS formation between $(PR)_{12}$ variants and RNA, we further explored the effects of $(PR)_{12}$ variants on LLPS with proteins containing acidic stretches. NPM1 was separated from poly(PR) by its acidic bundle-3 structural domain composing of 20 consecutive acidic amino acids.[17] Meanwhile, the periodicity of $(PR)_{50}$ regulates the diffusion of NPM1 in cells. Thus, we hypothesized that the periodicity of

$(PR)_{12}$ may regulate the phase separation of NPM1 and that modulation of LLPS underlies the impaired diffusion of NPM1 in cells. When recombinant human NPM1 was mixed with $(PR)_{12}$ variants, NPM1 phase-separated from $(PR)_{12}$ variants and the droplets exhibited wetting properties on the glass surface, except for the droplets of $P_{12}R_{12}$, which did not phase-separate with NPM1, as in Fig. 15 (a). Furthermore, previous studies have verified the phase separation of NPM1 from ribosomal RNA (rRNA).[17] Recombinant human NPM1 was phase-separated from human rRNA, and the signal intensity of NPM1 and rRNA in the droplets increased when $(PR)_{12}$ was added, indicating that more NPM1 molecules were captured in the phase separated droplets shown in Fig. 15 (b) and (c). When mixed with $(P_4R_4)_3$, NPM1 formed large droplets, but rRNA in Fig. 15 (b) formed fibrillar structures without liquid properties. This suggested adding $(PR)_{12}$ leaded to more involvement of NPM1 and rRNA in LLPS droplets, causing the droplets to become condensed, whereas the addition of $(PR)_{12}$ variants with large periodicities did not.

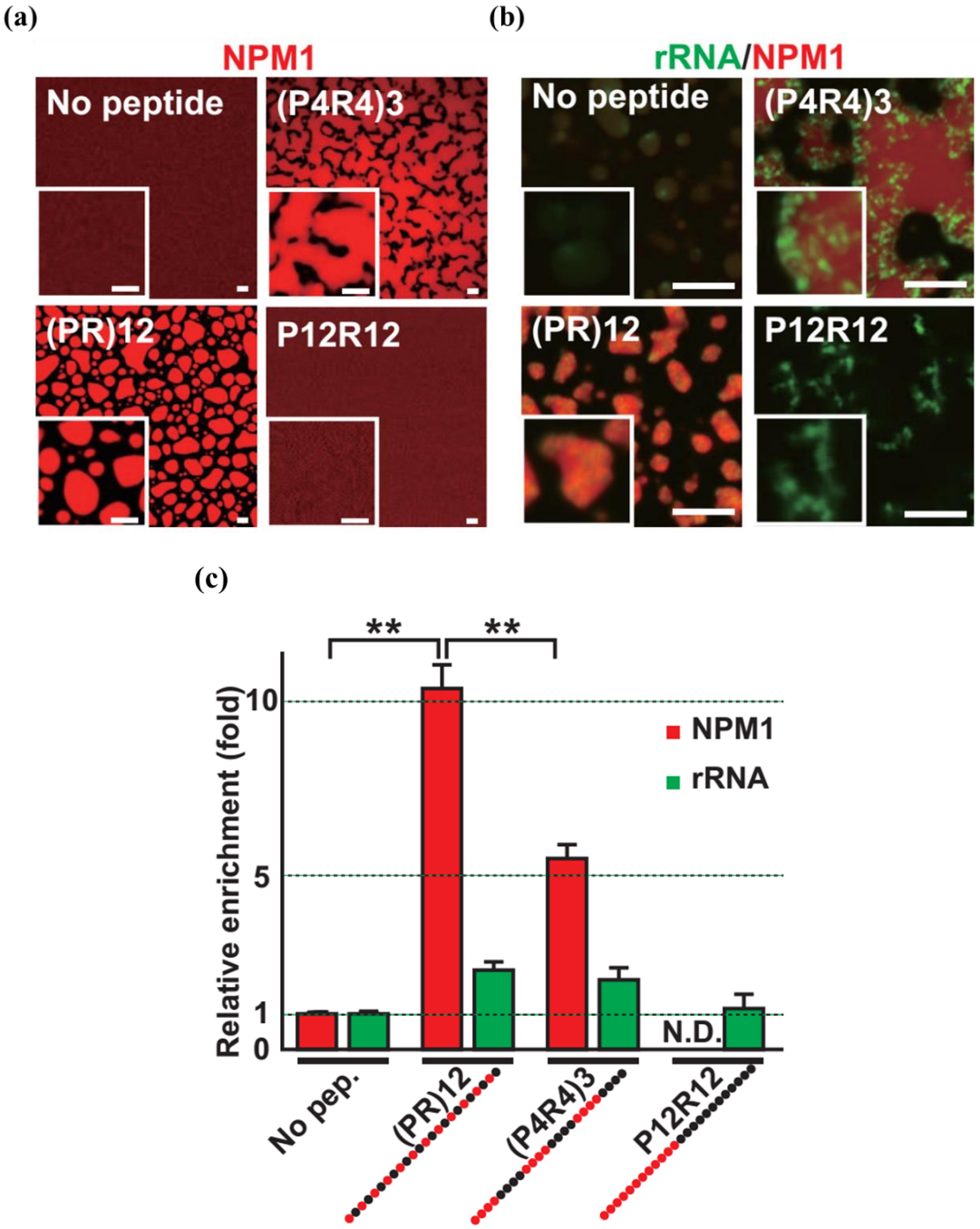


Figure 15. (a) Recombinant human NPM1 labeled with HyLite555 mixed with $(PR)_{12}$ variants; (b) recombinant human NPM1 mixed with $(PR)_{12}$ variants and human ribosomal RNA (rRNA); (c) signal enrichment of NPM1 and rRNA in the phase-separated droplets; each condition N = 10. Scale bars are shown for 10 μm. All data are expressed as mean ±S.D. Asterisks indicate significant differences derived from one-way ANOVA with Tukey's test (**: $p<0.01$).

3.4 Conclusion

In this section, based on the results of *in vitro* and *in vivo* experiments, on the one hand, we found the mobility of LLPS was inversely proportional to the periodicity value of poly(PR). The smaller the periodicity of poly(PR) variant, the weaker its binding energy to RNA. Furthermore, Pro residue confers structural rigidity to the peptide, which tuned the intermolecular interactions between poly(PR) variants and RNA and provided higher multivalency to form LLPS. On the other hand, we speculated the dynamic exchange rate of NPM1 protein in phase equilibrium was disrupted by the coexistence of complex coacervation with $(PR)_n$. $(PR)_n$ affected the mobility of NPM1 in the cell by condensing NPM1 and rRNA in phase-separated droplets through multivalent interaction, whereas P_nR_n was unable to condense NPM1 and accumulated to rRNA. Thus, P_nR_n had no significant effect on the mobility of NPM1. A schematic diagram is depicted in Fig. 16.

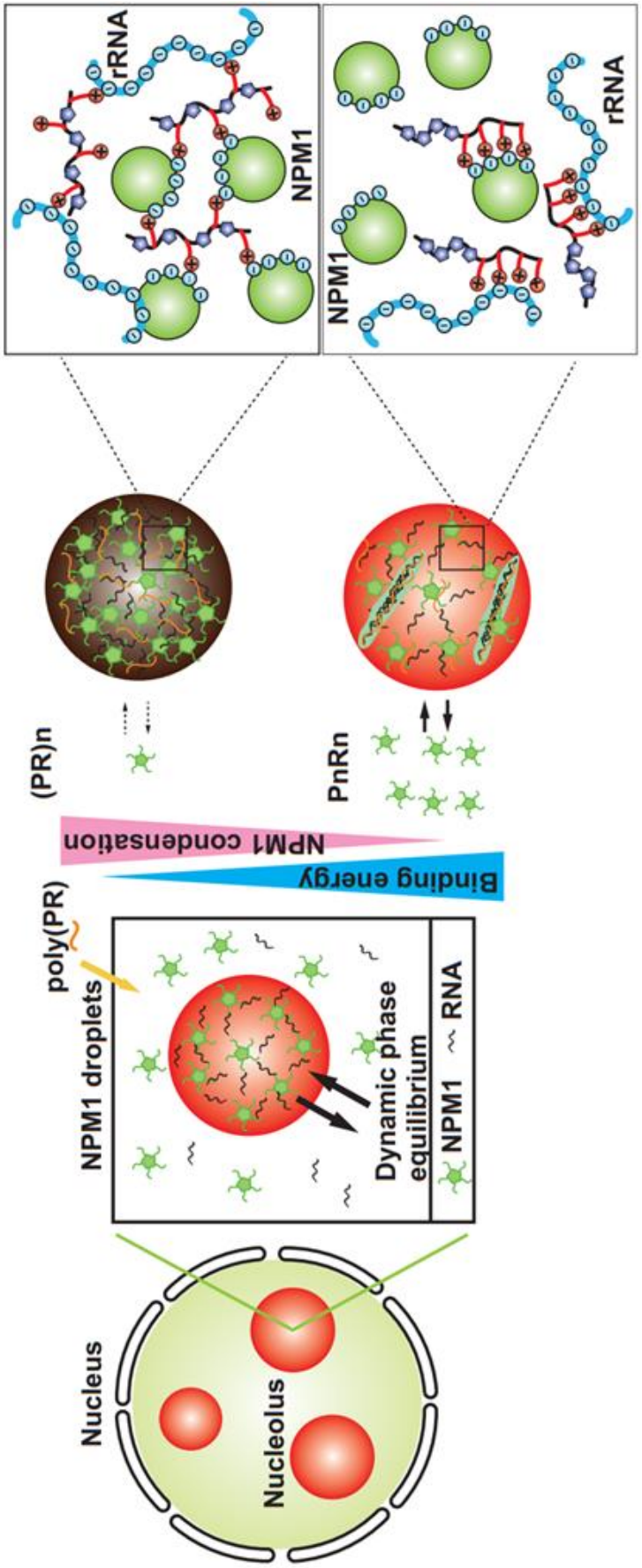


Figure 16. Schematic diagram of three-body interaction in cells

3.5 References

1. Pak, C. W. *et al.* Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol. Cell* **63**, 72–85 (2016).
2. Aumiller, W. M., Pir Cakmak, F., Davis, B. W. & Keating, C. D. RNA-Based Coacervates as a Model for Membraneless Organelles: Formation, Properties, and Interfacial Liposome Assembly. *Langmuir* **32**, 10042–10053 (2016).
3. Boeynaems, S. *et al.* Spontaneous driving forces give rise to protein−RNA condensates with coexisting phases and complex material properties. *Proc. Natl. Acad. Sci. U. S. A*. **116**, 7889–7898 (2019).
4. Chang, L. W. *et al.* Sequence and entropy-based control of complex coacervates. *Nat. Commun.* **8**, 1–7 (2017).
5. Feric, M. *et al.* Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165**, 1686–1697 (2016).
6. Miller, T., Goude, M. C., Mcdevitt, T. C. & Temenoff, J. S. Molecular engineering of glycosaminoglycan chemistry for biomolecule delivery. *Acta Biomater.* **10**, 1705–1719 (2014).
7. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A*. **110**, 13392–13397 (2013).
8. Lin, Y. H., *et al*. Phase Separation and Single-Chain Compactness of Charged Disordered Proteins Are Strongly Correlated. *Biophys. J.* **112**, 2043–2046 (2017).
9. Day, C. A., Kraft, L. J., Kang, M. & Kenworthy, A. K. Analysis of protein and lipid dynamics using confocal fluorescence recovery after photobleaching (FRAP). *Curr. Protoc. Cytom.* 1–29 (2012).
10. Axelrod, D., Koppel, D. E., Schlessinger, J., Elson, E. & Webb, W. W. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophys. J.* **16**, 1055–1069 (1976).
11. Kang, M., Day, C. A., Di Benedetto, E. & Kenworthy, A. K. A quantitative approach to analyze binding diffusion kinetics by confocal FRAP. *Biophys. J.* **99**,

2737–2747 (2010).

12. Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).
13. Lee, K. H. *et al.* C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. *Cell* **167**, 774-788.e17 (2016).
14. Box, J. K. *et al.* Nucleophosmin: From structure and function to disease development. *BMC Mol. Biol.* **17**, (2016).
15. Colombo, E., Alcalay, M. & Pelicci, P. G. Nucleophosmin and its complex network: A possible therapeutic target in hematological diseases. *Oncogene* **30**, 2595–2609 (2011).
16. Mitrea, D. M. *et al.* Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. *eLife* **5**, 1–33 (2016).
17. White, M. R. *et al.* C9orf72 Poly(PR) Dipeptide Repeats Disturb Biomolecular Phase Separation and Disrupt Nucleolar Function. *Mol. Cell* **74**, 713-728.e6 (2019).
18. Mitrea, D. M. *et al.* Structural polymorphism in the N-terminal oligomerization domain of NPM1. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4466–4471 (2014).
19. Gallo, A. *et al.* Structure of nucleophosmin DNA-binding domain and analysis of its complex with a G-quadruplex sequence from the c-MYC promoter. *J. Biol. Chem.* **287**, 26539–26548 (2012).
20. Mitrea, D. M. *et al.* Self-interaction of NPM1 modulates multiple mechanisms of liquid-liquid phase separation. *Nat. Commun.* **9**, 1–13 (2018).
21. S. Ladha, A. R. Mackie, D. C. Clark, Cheek cell membrane fluidity measured by fluorescence recovery after photobleaching and steady-state fluorescence anisotropy. *J. Membr. Biol.* (1994).
22. J. Yguerabide, J. A. Schmidt, E. E. Yguerabide, Lateral mobility in membranes as detected by fluorescence recovery after photobleaching. *Biophys. J.* **40**, 69–75 (1982).

CHAPTER 4. Molecular interaction motif for poly(PR) dipeptide repeats

4.1 Introduction

Traditional proteomics suggested poly(PR) preferentially interacts with RNA-binding proteins and ribosomal proteins, influencing protein translation.[1,2] This is in contrast to poly(GR), which interacts with mitochondrial proteins and impairing its functions.[3,4] As shown above, alternating insertion of Pro residues into consecutive R_{12} conferred protein-protein/RNA in phase-separated droplets the ability to loosen the interaction. This suggests the alternating distribution of Arg and the alternating amino acids of Arg sequence are determinants of the interactome of Arg-rich DRPs, which may simultaneously lead to different cytotoxic mechanisms. The $P_{12}R_{12}$ variant is compositionally identical to $(PR)_{12}$, such as the number of Args, the number of Pros and the length of peptide, which may enhance or inhibit protein-protein interactions, thus attenuate the deleterious nature of $(PR)_{12}$.

Liquid chromatography-mass spectrometry (LC-MS) has been one of the popular scientific method in proteomics for measuring and differentiating the components of complex mixtures.[5] Thus, we performed quantitative LC-MS to determine the interactomes of these peptides. Additionally, we also explored the common features of $(PR)_{12}$-enriched interactome.

4.2 Materials and experimental condition

One nmol of HA peptide, HA-$(PR)_{12}$, or HA-$P_{12}R_{12}$ were mixed with 200 μg of HEK293 cell lysates. Immunoprecipitation (IP) was performed with IP buffer (150 mM NaCl, 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.4, 1 mM Ethylenediaminetetraacetic acid (EDTA), 0.5% Triton-X100 and protease inhibitor cocktails) and 10 μl of anti-HA antibody beads (Fujifilm). After spinning for 4 h at 4°C,

the beads were washed 4 times with IP buffer and the precipitated proteins were eluted with 2x Laemmli sample buffer (BioRad, Carlsburg, CA). The extracted samples were analyzed by LC-MS. Briefly, samples were reduced with 10 mM tris(2-carboxyethyl)phosphine (TCEP) for 10 minutes at 100°C, and alkylated with 50 mM iodoacetamide at ambient temperature for 45 min, followed by Sodium dodecyl sulfate-Polyacrylamide gel electrophoresis (SDS-PAGE). Meanwhile, electrophoresis was terminated at a migration distance of 2 mm from the top edge of the separation gel. After Coomassie Brilliant Blue-staining (CBB-staining), protein bands were excised, distained, and finely cut prior to in-gel digestion with Trypsin/LysC Mix (Promega) at 37°C for 12 hours. The final peptides were extracted from gel fragments and then analyzed by an Orbitrap Fusion Lumos Mass Spectrometer (ThermoFisher Scientific) combined with UltiMate 3000 RSLC nano-flow HPLC system (ThermoFisher Scientific) in HCD MS/MS mode. Peptides were identified and quantified using Proteome Discoverer 2.4 software (ThermoFisher Scientific), where the MS/MS spectra were searched against the *Homosapiens* protein database in SwissProt (https://www.uniprot.org/), with a false discovery rate of 1% as an identification filter. The results from the HA control peptide were omitted due to the very low abundance of proteins in the extract. The total counts of peptide fragments were 5.861×10^{10} for all HA-peptides.

4.3 Alternating Arg distribution regulates the $(PR)_{12}$ interactome

Liquid chromatography-mass spectrometry (LC-MS) refers to a tandem analytical chemistry technique combining the physical separation capability of liquid chromatography with the mass analysis capability of mass spectrometry (MS). Liquid chromatography allows the separation of mixtures of multiple components, and mass spectrometry provides structural characterization on individual components with high molecular specificity and detection sensitivity. Due to these advantages, LC-MS has been widely used in various fields, such as biotechnology, environment monitoring, pharmaceutical industries, proteomics and so on.[6,7]

Immunoprecipitation with HEK293 cell lysates followed by LC-MS identified approximately 2000 interacting proteins shown in Fig. 1 (a) and supplementary material Fig. S1. The $(PR)_{12}$ interactome is composed mostly of nuclear and cytosolic proteins, where poly(PR) mainly localizes as shown in Fig. 1(b).

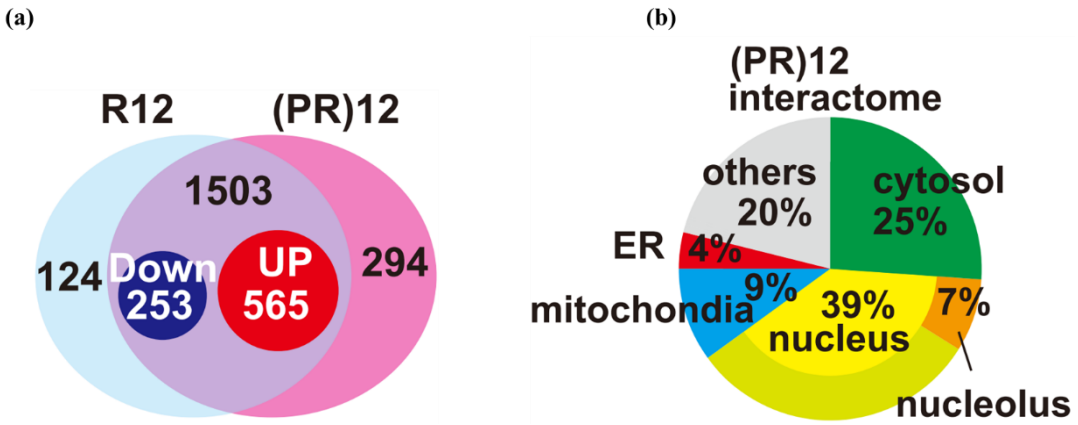


Figure 1. (a) Venn diagram of proteins identified in the R_{12} and $(PR)_{12}$ interactomes; (b) subcellular localization of the $(PR)_{12}$ interactome.

As a type of scatter plot, the volcano plot has been applied to efficiently identify the most significant changes among thousands of replicate data points in statistics. Currently, it has been widely utilized in omics experiments, including genomics, proteomics and so on.[8-10] It plots horizontally coordinated relationship curves in terms of the magnitude of change (fold change) and vertically coordinated relationship curves in terms of statistical significance (p value), respectively. For fold change, it is typically the simplest way to identify differentially expressed genes by assessing the log ratio between the two occasions and considering the entire gene to differ above a random cutoff value to be differentially expressed.[11-13] For p value, it is a measure of the evidence for the null hypothesis in statistical tests, i.e., the probability of having a test statistic equal to or exceeding the observed value under the assumption that the null hypothesis is true.[14] Here, first we will create a volcano plot highlighting all significant proteins between $(PR)_{12}$ and R_{12} interactomes. Off-axial vertical lines indicate ± 2.0 - fold changes, while the off-axial horizonal lines indicate $p = 0.005$. In this plot, proteins

are colored if they pass the threshold for log fold changes and log p values. The color is red when they are upregulated, blue when they are downregulated, and gray when they are insignificant. Therefore, the volcano plot shown in Fig. 2 reflected the enrichment of $(PR)_{12}$ in the interactome was completely different from that of R_{12}.

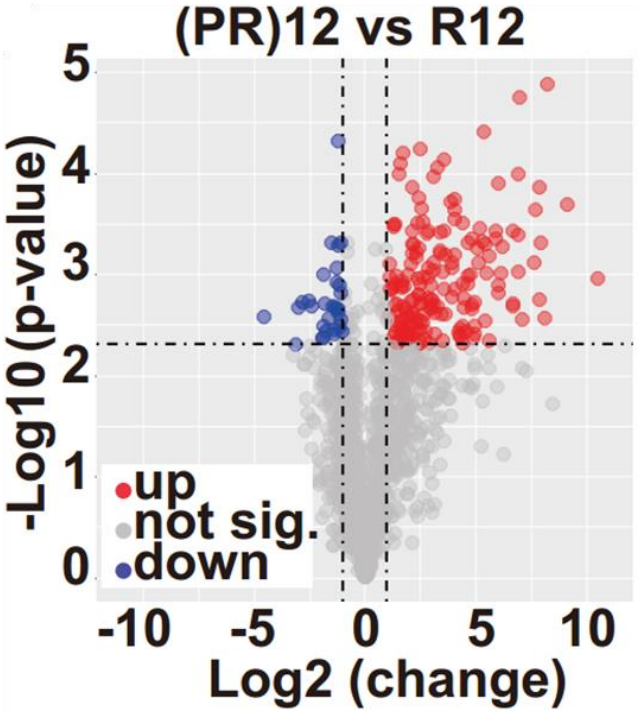


Figure 2. Volcano plot illustrates the significant different abundance of proteins between R_{12} and $(PR)_{12}$ interactors.

Comparing the interactomes of R_{12} and that of $(PR)_{12}$, 1503 proteins out of 2049 identified proteins shared a common signal with the interactomes of R_{12} and $(PR)_{12}$. Compared to R_{12} shown in Fig. 3(a), 171 proteins (11.4 %) showed at least a 10-fold increase in signal, while only 7 proteins (0.5 %) showed a decrease in signal to less than 10 % in $(PR)_{12}$. The signal intensity of the $(PR)_{12}$-enriched interactome (> 10 fold compared to R_{12}) was reduced in most cases in the $P_{12}R_{12}$ interactome shown in Fig. 3 (b), suggesting that the enrichment of the interactome was dependent on the periodicity of $(PR)_{12}$. These findings indicated that (i) peptide-protein interactions were influenced by the size of $(PR)_{12}$ periodicity, and (ii) the number of molecules interacting with 12 Arg residues increased after alternate Pro residues insertion.

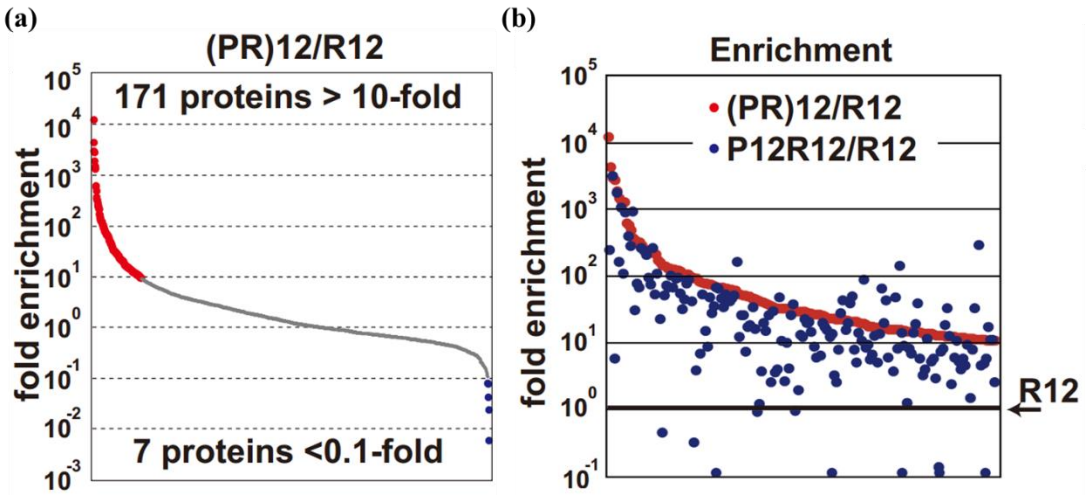


Figure 3. (a) $(PR)_{12}$ interactome enriched more than 10-fold (171 proteins: red circles) or enriched less than one-tenth of 7 proteins: (blue circles) compared to R_{12} interactome. (b) Diminished enrichment or 171 proteins shown in (a) in the P12R12 interactome [$(PR)_{12}/R_{12}$ > 10].

4.4 Alternate Pro residues enhance multivalent interactions with acidic motifs

The common features of $(PR)_{12}$-enriched interactome were analyzed to clarify the affinity of $(PR)_{12}$ for these proteins. The sequences of the first 20 enriched proteins tended to have long acidic stretches composing of consecutive D/E or repetitive clusters of D/E, as shown in Fig. 4.

Figure 4. Amino acid composition of ANP32A and LEO1 was visualized by PepCalc.com (https://pepcalc.com). ANP32A exemplifies proteins with long acidic stretches and LEO1 exemplifies proteins with repetitive D/E clusters.

Next, we screened for consecutive acidic stretches in protein sequences with at least a 5-fold increase in $(PR)_{12}$ or more than a 50% decrease in $(PR)_{12}$ compared to R_{12}, and found $(PR)_{12}$ preferred proteins with longer acidic stretches as shown in Fig. 5.

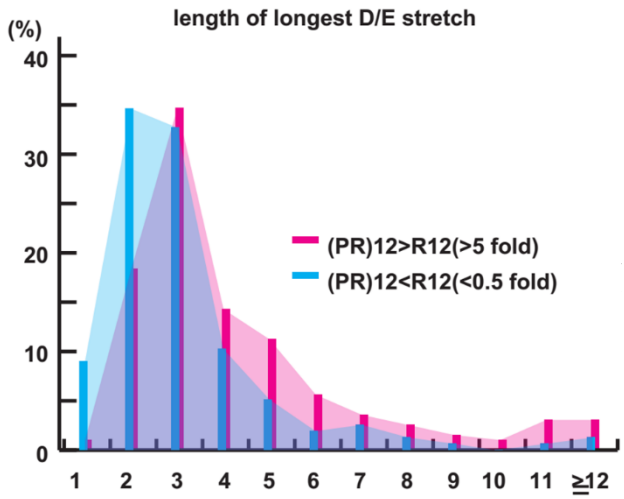


Figure 5. Length histogram of the longest D/E segment in the enriched interactors.

To extract valid features having important discriminatory information from protein sequence data, we conducted data mining to acquire accurate predictions. In this work, we evaluated the protein sequence data based on the data science platform Anaconda by importing the data into the platform's Spyder (Python 3.7.3).

Step1: Detection of consecutive acidic stretches

Take a portion of the entire amino acid sequence of a protein (i.e., EIF 3M-Human) as an example. We fished consecutively from the start codon to the stop codon, with each fish consisting of six consecutive amino acids. Likewise, the same fishing process was followed for all other selected proteins, so that all amino acids of each protein were covered.

Step2: Scoring

After fishing for all the proteins, we scored all the sequences. Here, we focused on the occurrence of D/E amino acids in each sequence made up of six consecutive amino acids. Particularly, if D/E are consecutive, they are expected to have a greater impact on the electrostatic force as in the previous analysis. Therefore, we adapted a weighted scoring system as below. For each sequence: a score of 0 is assigned if there is no D/E amino acids; a score of 1 is assigned if there is one D/E amino acid; a score of 2 is assigned if there are two D/E amino acids; a score of 4 is assigned if there are three D/E amino acids; a score of 6 is assigned if there are four D/E amino acids; a score of 8 is assigned if there are five D/E amino acids; a score of 10 is assigned if there are six D/E amino acids. As a result, all consecutive 6 amino acids of proteins were scored consecutively and summed to obtain the final score for each protein.

Fig. 6 depicts the D/E weighted scoring system (D/E score) using six A.A. peptides, and we calculated the score from the start codon to the end codon of each protein.

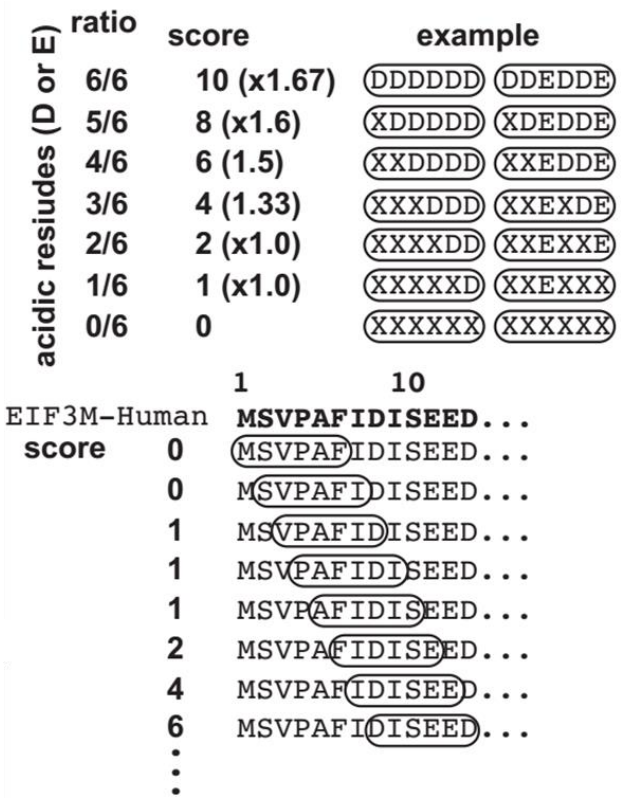


Figure 6. D/E scoring weighing continuum scheme. Six consecutive amino acids were scored consecutively and summed.

According to this data mining procedure, the total D/E scores of proteins with greater affinity for $(PR)_{12}$ was significantly higher than that of proteins with greater affinity for R_{12} ($p < 0.0001$), as shown in Fig. 7 (a) and (b).

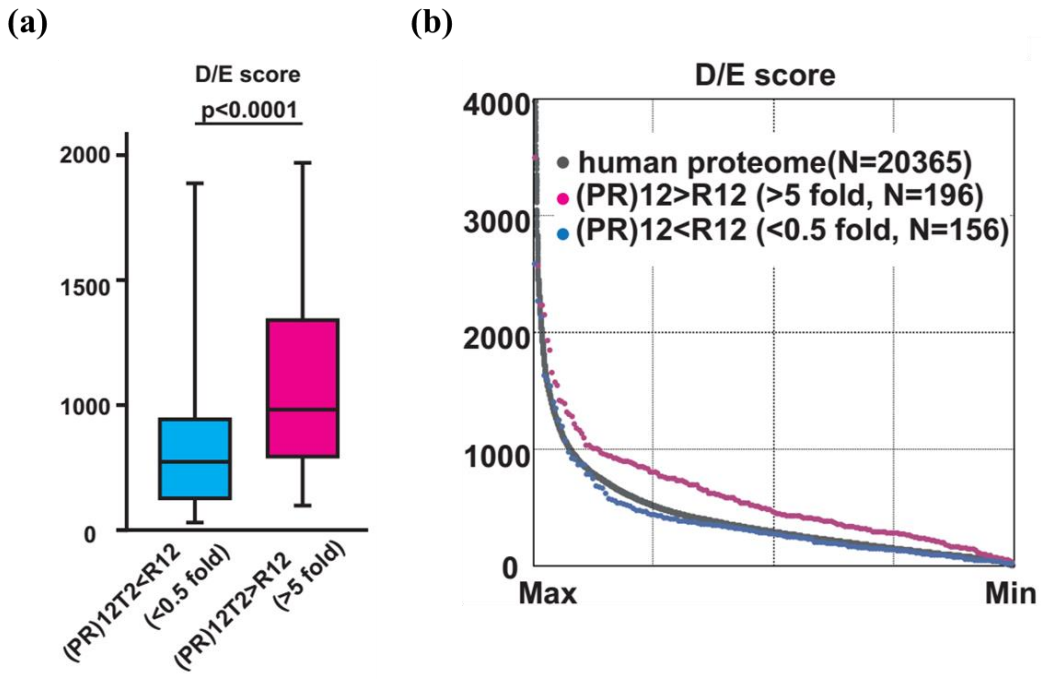


Figure 7. (a) Box plots of D/E scores for the R_{12} and $(PR)_{12}$ enriched interaction groups. Whiskers are drawn to the 5th and 95th percentiles. (b) D/E score of $(PR)_{12} < R_{12}$ (< 0.5 fold) (blue circles), $(PR)_{12} > R_{12}$ (> 5 fold) (red circles) and human proteome for the enriched interaction group. Thirteen proteins with D/E scores higher than 4000 were omitted from the human proteome for graphical clarity.

At last, we calculated the overall amino acid occurrence in the interactome. There was no significant difference between the overall interactome of R_{12} and $(PR)_{12}$, but the occurrence of charged residues (D, E, K and R) was higher than that of the human proteome shown in Fig. 8 (a). When we calculated the occurrence of amino acids in the four subgroups (> 500 fold, > 200 fold, > 5 fold or < 0.5 fold, respectively enriched in $(PR)_{12}$ compared to R_{12}), we found acidic amino acids were overrepresented in the $(PR)_{12}$ interactome compared to R_{12} shown in Fig. 8 (b). The absence of enrichement of aromatic acids suggested that electrostatic instead of cation-π interactions play a key role in determining the $(PR)_{12}$ interaction group.

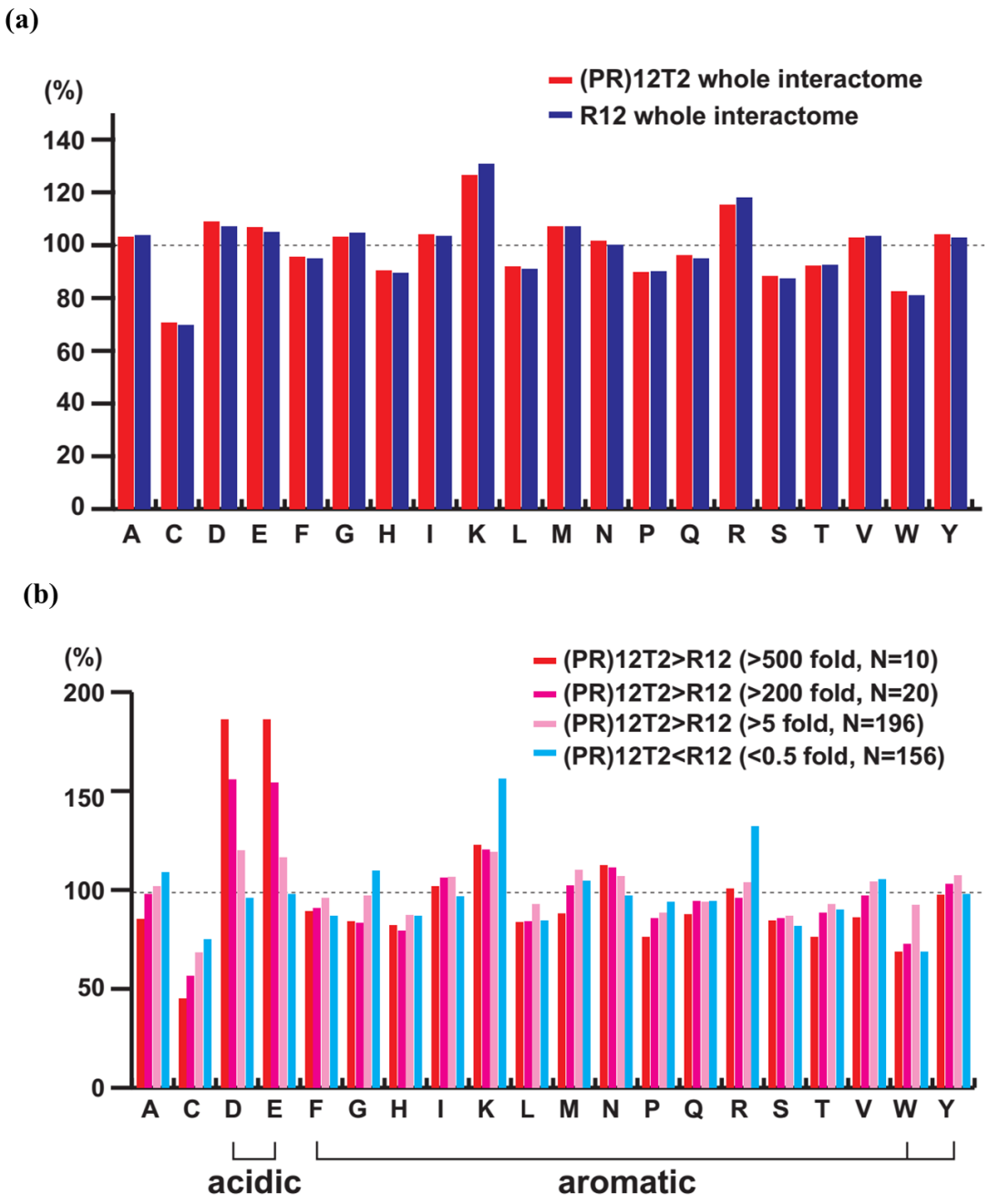


Figure 8. (a) Amino acid occurrence rates in the entire interactome of $(R)_{12}$ and $(PR)_{12}$, normalized to the human proteome; (b) Amino acid occurrence rates in the enriched interactome of $(PR)_{12}$ (> 500 fold, > 200 fold, > 5 fold or < 0.5 fold) compared to $(R)_{12}$, normalized to the human proteome.

4.5 Conclusion

Based on the cellular proteomic analysis, we demonstrated that the alternating Arg structure of DPRs can efficiently enhance dipeptide-protein interactions. Most importantly, the alternating Pro residues enhanced multivalent interactions with acidic groups via electrostatic interactions. These results suggest that the enrichment of interactors can be at least partially explained by LLPS.

4.6 References

1. Kanekura, K. *et al.* Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. *Hum. Mol. Genet*. **25**, 1803–1813 (2016).
2. Moens, T. G. *et al.* C9orf72 arginine-rich dipeptide proteins interact with ribosomal proteins in vivo to induce a toxic translational arrest that is rescued by eIF1A. *Acta Neuropathol.* **137**, 487–500 (2019).
3. Choi, S. Y. *et al.* C9ORF72-ALS/FTD-associated poly(GR) binds Atp5a1 and compromises mitochondrial function in vivo. *Nat. Neurosci.* **22**, 851–862 (2019).
4. Lopez-Gonzalez, R. *et al.* Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. *Neuron* **92**, 383–391 (2016).
5. Braisted, J. C. *et al.* The APEX quantitative proteomics tool: Generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics* **9**, 1–11 (2008).
6. Dass, C. Hyphenated Separation Techniques. *Fundamentals of Contemporary Mass Spectrometry* 151–194 (2007).
7. Jacob, C., et al. *Recent advances in redox active plant and microbial products: From basic chemistry to widespread applications in medicine and agriculture*. (2014).
8. Cui, X. & Churchill, G. A. Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol.* **4**, (2003).
9. Li, W. Volcano plots in analyzing differential expressions with mRNA microarrays. *J. Bioinform. Comput. Biol.* **10**, 1–24 (2012).
10. Li, W., Freudenberg, J., Suh, Y. J. & Yang, Y. Using volcano plots and regularized-chi statistics in genetic association studies. *Comput. Biol. Chem.* **48**, 77–83 (2014).
11. Schena, M. *et al.* Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. U. S. A*. **93**, 10614–10619 (1996).
12. DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Chemtracts* **12**, 148–152 (1999).
13. Draghici, S. Statistical intelligence: Effective analysis of high-density microarray

data. *Drug Discov. Today* **7**, 55–63 (2002).

14. Wasserstein, R. L. & Lazar, N. A. The ASA's Statement on p-Values: Context, Process, and Purpose. *Am. Stat.* **70**, 129–133 (2016).

CHAPTER 5. Diffusion of LLPS droplets consisting of poly(PR) dipeptide repeats and RNA at solid/liquid interface

5.1 Introduction

In recent years, LLPS formed by proteins and RNA molecules has become an essential physicochemical process that can describe the organization of and cellular function of MLOs in living cells[1,2], and even several fatal neurodegenerative diseases, such as ALS due to the spontaneous coalescence and growth of LLPS droplets.[3] Normally, LLPS characteristics have been performed by optical microscopy, and transparent substrates are necessary. Take advantage of liquid and wetting properties of LLPS droplets on glass surfaces, several technical protocols have thus been put forward for immobilizing droplets on surfaces.[4,5] Nevertheless, interactions between LLPS droplets and glass surfaces still remain unclear.

To address this issue, we explored the surface diffusion of LLPS droplets using single particle tracking method (SPT) with fluorescence microscopy. We assumed the zeta potential of LLPS droplets probably interact uniquely with the glass surface in a Coulombic manner.[6] Thus, we investigated the dynamic diffusion process of LLPS droplets on the untreated cover glass at first and made a comparison between two substates to control the interaction of LLPS droplets at solid/liquid interface next. One is an untreated cover glass and the other one is a chemically modified cover glass with positively charged amine groups on the surface. Previously, arginine has been reported to play a key role in protein-protein and protein-nucleic acid interactions and generally has typical characteristics with positive charge and dipole moment.[7] Proline-arginine (poly(PR)) is an arginine-rich dipeptide repeat with high repeatability and an abundant positive charges, which allows it to bind to nucleic acids/proteins at multiple sites by cation-π interactions and electrostatic forces to form LLPS droplets.[8-10] We used poly$(PR)_{12}$, which belongs to the proline-arginine dipeptide repeats, and a homopolymeric adenine poly-A as RNA model that is demonstrated to have LLPS

formation in our research. A fraction of poly(PR)$_{12}$ was labeled with fluorescein isothiocyanate (FITC). We then utilized single particle tracking to examine the droplets diffusing at the interface of two substrates, conducted the corresponding statistical analysis of the size distribution of LLPS droplets and calculated their diffusion coefficients.

5.2 Materials and experimental condition

Reagents. Poly-A RNA was purchased from Sigma-Aldrich company (USA) under product number 10108626001. Poly(PR)$_{12}$ and FITC-poly(PR)$_{12}$ were both chemically synthesized by Genscript (Piscataway NJ) with > 85% purity. After synthesis, the trifluoroacetate was replaced by acetate.

Surface chemical modification. The cover glass utilized as substrate (thickness 0.13 ~ 0.17 mm for No. 1, MATSUNAMI GLASS IND., LTD.) was firstly cleaned by plasma treatment approximately 10 min (Plasma cleaner PDC-32G, Harrick Plasma). Next, the substrates were immersed in an aqueous solution of 10% (3-Aminopropyl) trimethoxysilane (APTMS with purity nearly 97%, purchased from Sigma-Aldrich company (USA)) and left for 30 min at room temperature. After that, the substrates were rinsed with deionized water and cured at 120 ℃ for 10 min. We successfully made the glass surface positively charged with amine groups.[11] All modification processes are depicted in Fig. 1.

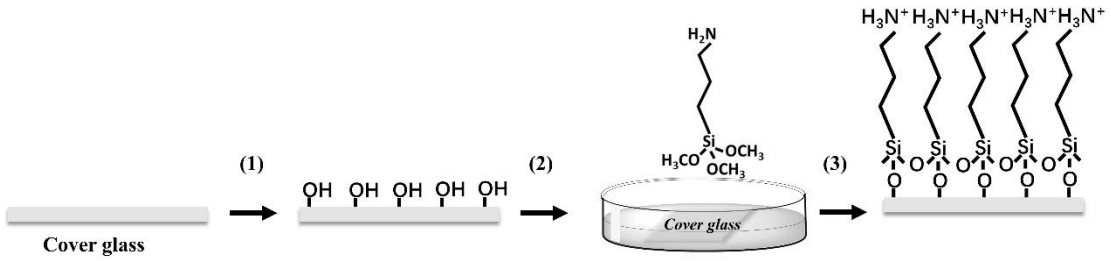


Figure 1. Schematic diagram of the process about modifying the cover glass surface. (1) Plasma treatment for approximately 10 min, (2) 10% 3-aminomethoxysilane (APTMS) treatment for approximately 30 min, (3) deionized water rinse, and dry heat at 120 ℃ for 10 min.

LLPS droplet formation. We mixed solutions of poly$(PR)_{12}$ (final concentration ranging from 100 μM, 50 μM to 10 μM) and poly-A RNA (final concentration ranging from 0.5 mg/ml, 0.25 mg/ml, 0.1 mg/ml 0.05 mg/ml) in phosphate buffer solution (final concentration of 10 mM) at a volume ratio of 1 : 1. After preparing the freshly mixed solution, we dropped 2 μL solution on the surface of cover glass and clamped with another same one to make a thin film of the solution. Therefore, the droplets were examined under an oil-immersion lens (N.A. = 1.4) in bright filed through an inverted fluorescence microscope (Olympus IX73) with an electron multiplying charge-coupled device (iXon-Ultra888 EMCCD, Oxford instruments). The exposure time was set as 2s, and the light source was white LED light located above the sample.

Fluorescence imaging. The experimental setup for fluorescence detection was illustrated in Fig. 2 (a) and (b). We placed a polydimethylsiloxane (PDMS) film (thickness equals to 3 mm thickness) with a hole (diameter equals to 5 mm) on the substrate surface. The same volume of poly$(PR)_{12}$ (final concentration 100 μM) was mixed with 100 nM of FITC-poly$(PR)_{12}$ and poly-A RNA (final concentration 0.5 mg/ml, M.W. = 100 ~ 500 kDa) under phosphate buffer solution conditions (final concentration 10 mM) at room temperature. The hole was filled with the 20 μL of the mixed solution. We then covered the top of PDMS wells with another piece of glass to avoid solution evaporation during the measurement. To study the interaction between LLPS droplets and the solid interface, we used an inverted fluorescence microscope (IX73P2F, Olympus, JP) equipped with a fluorescence filter cube (emission: 420 ~ 460 nm; excitation: > 515 nm). A mercury lamp functioned as the input power (870 μW) to excite the fluorescent dye. The sample was then placed on the fluorescence microscope equipped with an oil-immersion lens (N.A. = 1.4) and a CMOS camera (Neo sCMOS/Solis, Andor, JP). Series of images were recorded constantly to study the interaction between LLPS droplets and the solid glass surface with an exposure time of 10 ms.

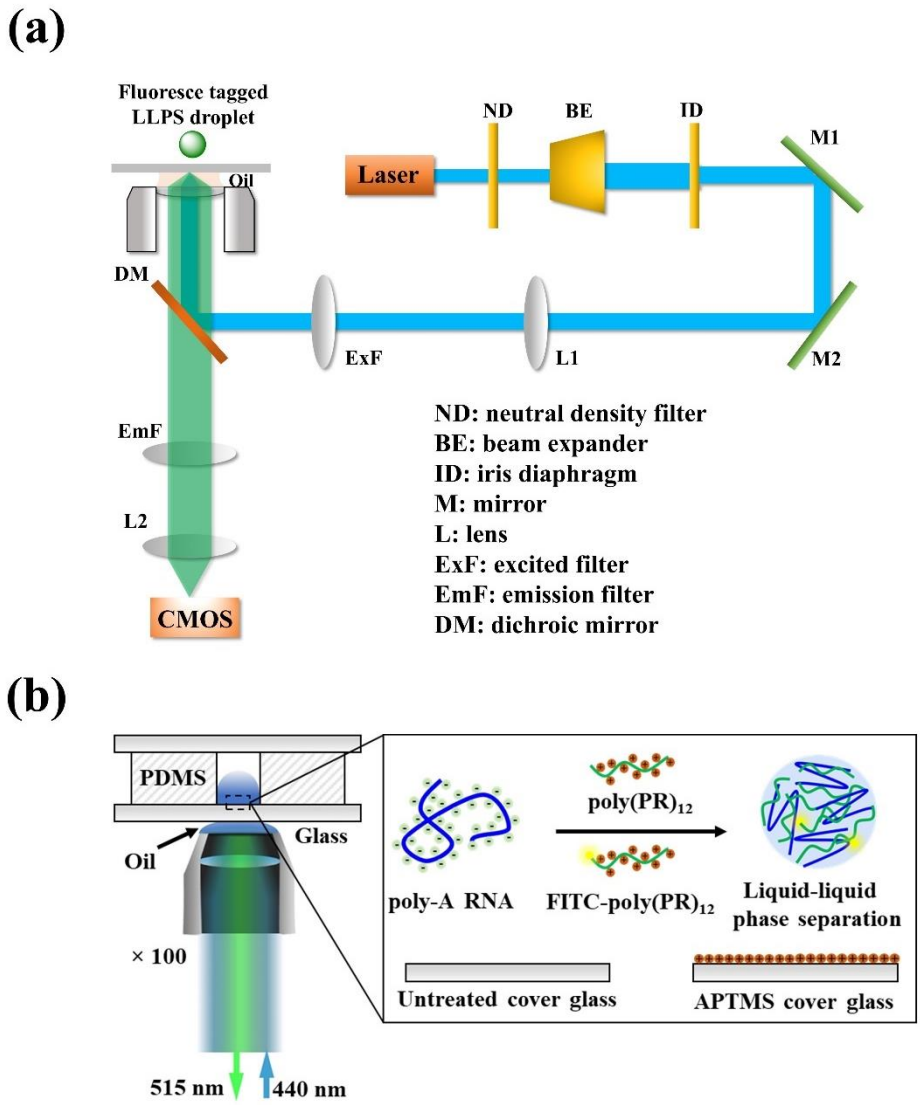


Figure 2. (a) Optical path of fluorescence microscope; (b) Schematic diagram of the interaction between LLPS droplets and solid surface detected by fluorescence microscope. Magnified view indicates LLPS droplet consisting of poly-A RNA, poly$(PR)_{12}$ and FITC- poly$(PR)_{12}$; Solid surface includes untreated cover glass and APTMS-modified cover glass.

Single particle tracking. As a effective method for studying the dynamics of heterogenous systems, SPT has been utilized to locate single particles with subdiffraction resolution and measure their individual dynamics instead of ensemble averaging.[12] In this research, multi-particle detection and tracking was perfoemed using a plug-in for the open-source biological image processing software Fiji[13] to automatically differentiate the density centers of fluorescent spots from the obtained

fluorescence images and extract the position of LLPS droplets on the solid surface. The droplet trajectories were obtained by connecting the closest positions in consecutive frames. The coordinates of the LLPS droplets could be acquired by SPT method, and the total fluorescence intensity of individual LLPS droplets was measured accordingly and mean square displacement (MSD) was finally performed.

5.3 LLPS droplet formation on substrates

As is depicted in Fig. 1, LLPS droplets were formed by the mixture of poly$(PR)_{12}$, FITC-labelled poly$(PR)_{12}$ and homopolymeric poly-A RNA in phosphate buffer solution through cation-π interactions and electrostatic forces. To testify whether the mixing concentration can tune the phase separation of polypeptide/RNA, we performed LLPS formation experiments with diverse concentration ranges of poly$(PR)_{12}$ and poly-A RNA. Once the same volumes of poly$(PR)_{12}$ (final concentration range of 100 μM, 50 μM to 10 μM) and poly-A RNA (final concentration range of 0.5 mg/ml, 0.25 mg/ml, 0.1 mg/ml to 0.05 mg/ml) were mixed directly in phosphate buffer solution (final concentration of 10 mM) at room temperature, we then used optical microscopy to reveal the occurrence of spherical droplets in the mixed solution. We observed LLPS droplets formation for all concentration samples of the mixture shown in Fig. 3. We found many spherical particles in the images, especially under conditions of high peptide and RNA concentrations.

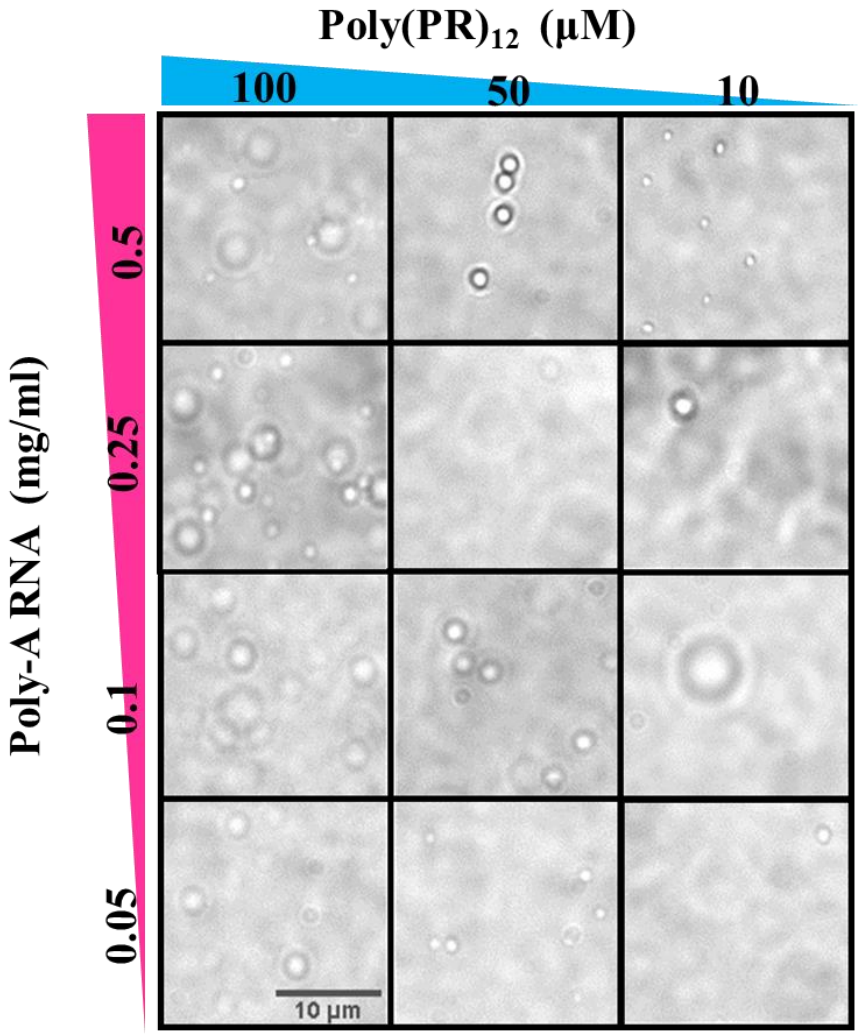


Figure 3. Optical transmission images of LLPS formed by mixing poly$(PR)_{12}$ and poly-A RNA in phosphate buffer solution.

In order to quantify the extent of light scattering by polypeptide/RNA LLPS, the absorbance of droplets in the phosphate buffer solution at 500 nm was measured using a Nanophotometer (Implen NanoPhotometer, N60) shown in Fig. 4a, i.e. turbidity measurement. In general, differences in the refractive indices of different structures within a complex medium partially reflect light from the internal interface, resulting in a light scattering component (Fig. S2). In this work, by examining the light scattering effect in the absorbance spectra of all mixture concentration ranges in Fig. 4a, we found that in the case of 10 μM poly$(PR)_{12}$ mixed with poly-A RNA, the absorbance was weak

due to the low scattering effect; in the case of 50 μM poly(PR)$_{12}$ mixed with poly-A RNA, the absorbance increased and no significant difference between each other; in the case of 100 μM poly(PR)$_{12}$ mixed with poly-A RNA, the absorbance was significantly maximum with a huge peak, indicating that in the mixed group (100 μM poly(PR)$_{12}$ + 0.25 mg/ml poly-A RNA). We also plotted the heat map shown in Fig. 4b. It shows that their tends are in good agreement with the optical microscope images in Fig. 3.

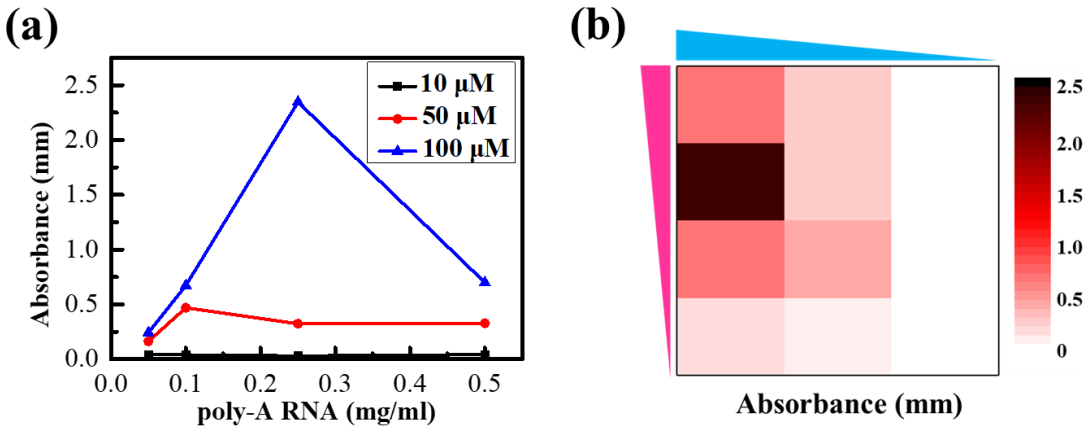


Figure 4. (a) Optical absorbance of each mixture solution at 500 nm to show the light scattering effect due to LLPS droplet formation; (b) Heat map of light absorbance of LLPS droplet in each mixture concentration.

5.4 Surface diffusion of LLPS droplet

To explore the intrinsic diffusion behavior of individual LLPS droplets on the surface, we chose a mixed solution (final concentration: 100 μM poly(PR)$_{12}$ + 0.5 mg/ml poly-A RNA mixed with 100 nM FITC-poly(PR)$_{12}$ in 10 mM phosphate buffer), considering that the group forms relative larger number density and smaller size LLPS droplets, as in Fig. 4 (a) and (b). We measured the fluorescence images of LLPS droplets on the untreated surface of the glass substrate with an inverse microscope. The exposure time of the images was 10 ms, which was short enough to resolve the location of LLPS at each moment. We recorded the trajectory of each droplet, analyzed it with SPT method[12] and compared them with each other on the solid surface. When illustrating

the trajectories of several typical cases on the untreated glass surface and their displacements over time, we found whole trajectories can be divided into two main events during their diffusion: (1) the fix mode, where the displacement is less than 50 nm over the whole measurement time, and (2) the diffusion mode, where the displacement is much larger than 0.5 μm over the same time frame shown in Fig. 5 (a). The droplets in the fix mode seemed to be trapped in certain positions, but the droplets in the diffusion mode were moving, as shown in Fig. 5 (b). Their displacement fluctuations on the two substrates are obviously different with time, as shown in Fig. 5 (c). Additionally, more typical samples distinguishing between stationary and diffusive events are also shown in Fig. S3. It is worth noting that the histogram of the displacements shows multiple peaks in the fix mode shown in Fig. 5 (c). This may imply a complex diffusion process. To explore and compare the differences between the two modes, we illustrated the velocity distribution of each droplet in both modes. Apparently, we found the velocity distribution of the diffusion mode follows the two-dimensional Maxwell-Boltzmann distribution shown in Fig. 5 (d). On the contrary, the velocity distribution of the fix mode slightly mismatches with the Maxwell-Boltzmann model, especially in the high velocity region shown in Fig. 5 (e). This may be caused by two processes: (1) fluctuation of the center of mass, resulting from the oscillation of the droplet, and (2) jumping of the droplet with a larger displacement.

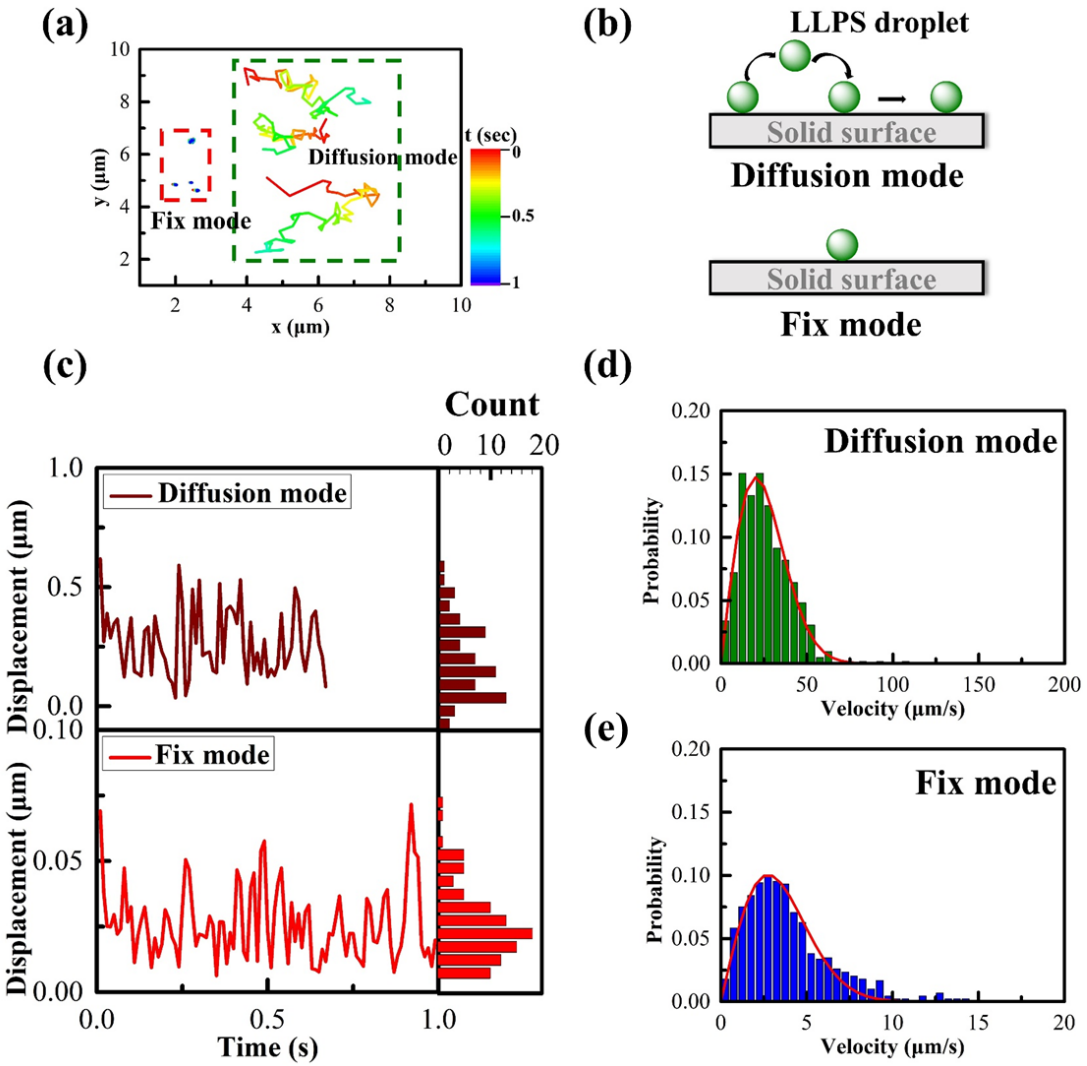


Figure 5. SPT of LLPS droplets diffusing on tan untreated glass surface. (a) Representative trajectories of observed LLPS with color-coded observation times; (b) Schematic diagrams of the fix and diffusion modes of LLPS on the glass surface; (c) Typical cases of LLPS droplet surface displacement fluctuations in the fix mode and diffusion modes and their respective displacement histograms; (d) and (e) Particle velocity probability distribution for the above cases above and their fitting curves.

To explore the effect of chemical modification on the glass surface, we analyzed the immobilization/diffusion pattern of LLPS droplets on untreated and APTMS-modified glass substrates by further quantitative analysis of the LLPS droplet surface kinetics, i.e., the mean square displacement (MSD). We derived the MSD using the following

equation:

$$MSD(\Delta t) = <\Delta r(\Delta t)^2> = <(r(t+\Delta t) - r(t))^2> = 4\,D \cdot \Delta t \qquad (1)$$

where Δr denotes the displacement of the particle, t denotes the diffusion time, and Δt denotes the lag time between the two images. The relationship between the MSD and the diffusion coefficient D applies to the case of two-dimensional Brownian motion.[15] Fig. 6 (a) and (b) indicate the MSD results for the trajectories of 100 LLPS droplets diffusing on untreated cover glass and APTMS-modified cover glass.

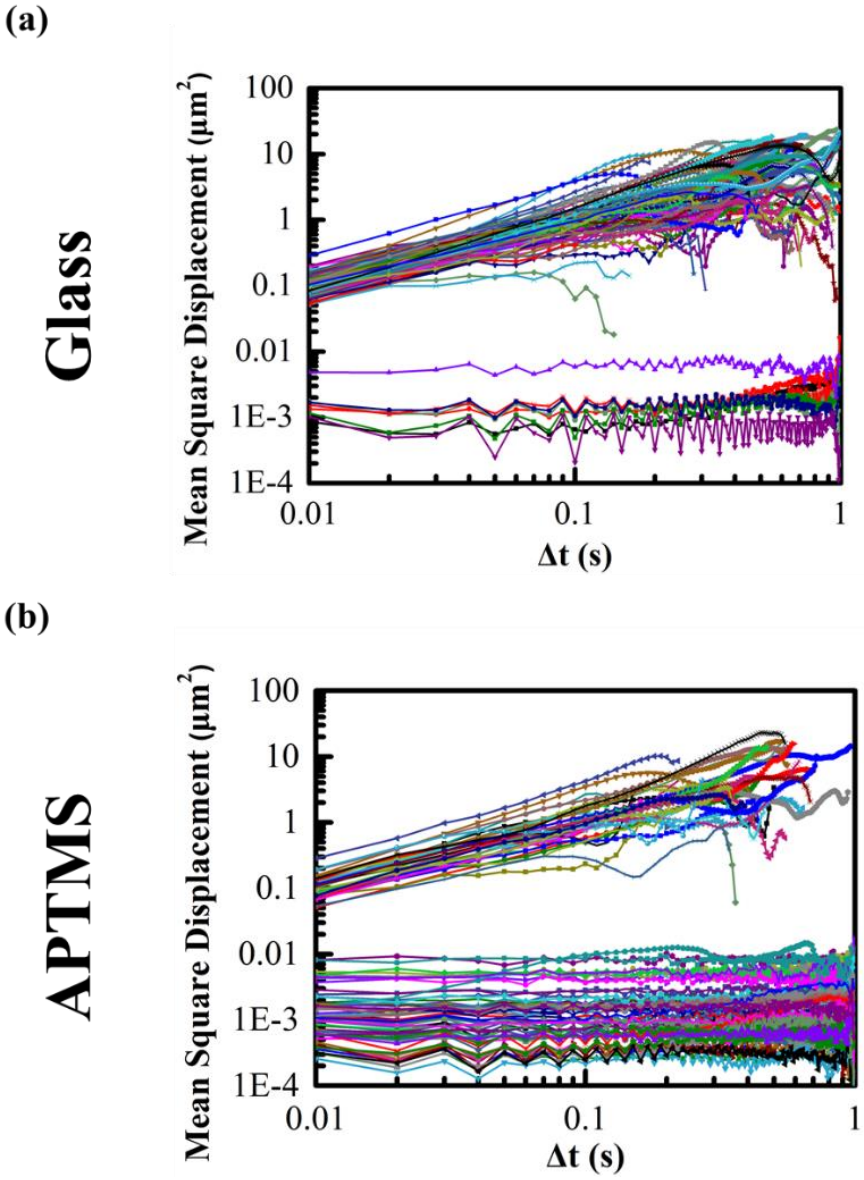


Figure 6. MSD of LLPS droplets diffusing on (a) untreated cover glass and (b) APTMS-modified cover glass surface.

We found that in the former case, the diffusion of most LLPS follows a 2D Brownian motion because the MSD increases linearly with Δt and a small number of droplets are fixed on the surface. Nonetheless, in the latter case, most of the droplets are fixed on the surface, while only a few droplets follow 2D Brownian motion. Moreover, the MSD values become more random as Δt increases because LLPS droplets have diverse sizes and volumes from each other. To explain these phenomena, we assumed the surface zeta potential of individual droplets can reflect the surface electrostatic properties of the condensate and affect the diffusion process. One current work indicated that LLPS droplets formed from 25 replicates of the dipeptide proline-arginine $(PR)_{25}$ and poly-U RNA had a negative average zeta potential and established more favorable electrostatic interactions with counterions.[6] Since we employed the same PR dipeptide and poly-RNA, we hypothesized the zeta potential of LLPS droplets in this research was also negative, which contributed to the difference in size with more attractive electrostatic interactions with positively-charged amine groups on the APTMS-modified glass surface. Here there will be a possibility that the droplets on the LLPS surface are coexisting, which may be influenced by the chemical modification of the glass surface. Although we cannot neglect the effect of LLPS surface coexistence, we believe the electrostatic effect of droplet-surface interaction can be dominant during the droplet diffusion process, because the particle tracking measurements were performed within 10 min after LLPS droplets formation in solution. The measurement time is shorter enough to reduce the main effect of coacervation.

Assuming a two-dimensional random diffusion, we derived the diffusion coefficients for both substrates by fitting eq. (1) shown in Fig.7 (a) and (b).

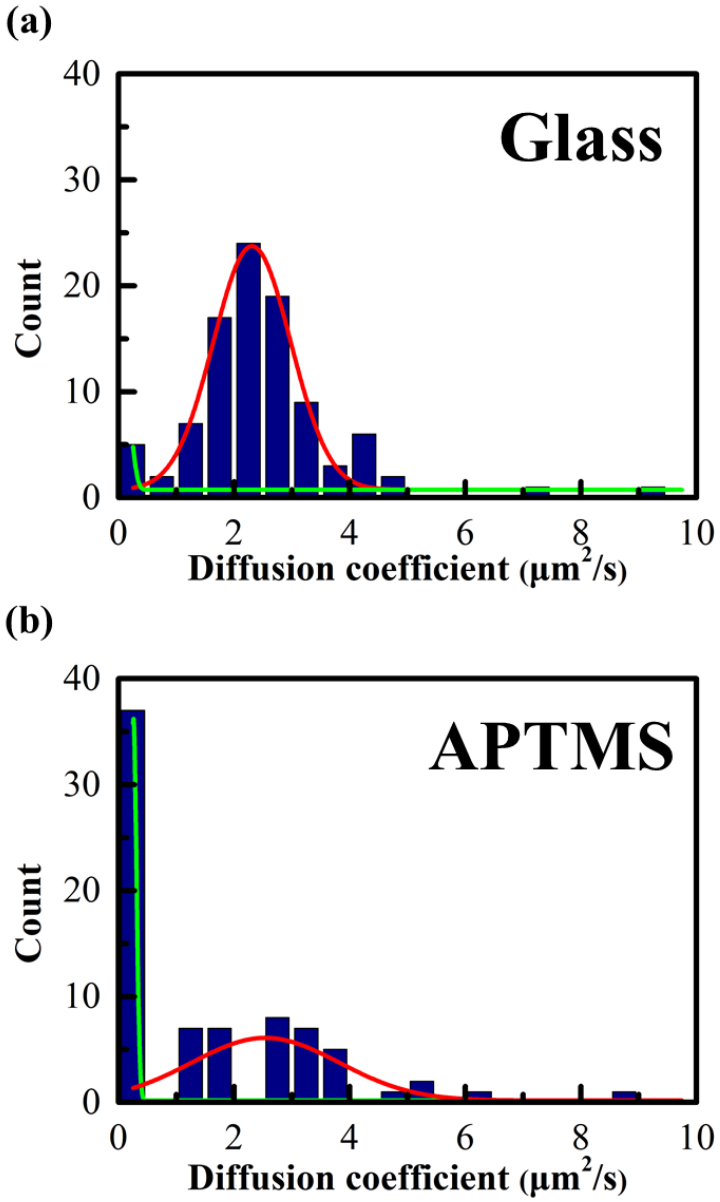


Figure 7. LLPS droplets diffusion coefficient distribution on (a) untreated cover glass and (b) APTMS-modified cover glass surface.

The obtained diffusion coefficients of LLPS droplets on the untreated cover glass surface were 0.189 $\mu m^2/s$ for fix mode and 2.311 $\mu m^2/s$ for the diffusion mode; while on the APTMS-modified cover glass surface, the fix mode was 0.258 $\mu m^2/s$ and the diffusion mode was 2.544 $\mu m^2/s$ for diffusion mode. Then, we also calculated the area percentage of each mode: on the untreated cover glass surface, the fix mode was 2.746 % and the diffusion mode was 97.254 % while on the APTMS-modified cover glass

surface, the fix mode was 20.368 % and the diffusion mode was 79.632 %. Therefore, we summarized the diffusion pattern of LLPS on the surface can be adjusted between fix and diffusion patterns, depending on the chemical modification of the surface. Exceptionally, we also analyzed the diffusion pattern in more detail using the van Hove correlation function Gs [16]:

$$G_s(\Delta x, \Delta t) = \frac{1}{N} < \sum_{i=1}^{N} \delta(x + x_i(t) - x_i(t + \Delta t)) > \quad (2)$$

where <·> denotes the ensemble averaging. This distribution shows the probability that one molecule has moved a distance Δx along the x or y coordinate during time Δt. We chose droplet trajectories with average diffusion coefficient from the results in Fig. 7 (a) and (b), and then plotted the van Hove correlation functions for the untreated and APTMS-modified glasses shown in Fig. 8 (a) and (b). The obtained peaks were in good agreement with the Gaussian distribution, reflecting the droplet diffusion on the untreated and APTMS-modified glasses followed a Brownian motion with a time scale less than 0.1 s. Based on the results of multiple displacement patterns shown in the histogram in Fig. 5 (c), we suggest that the transition between diffusion and fix modes may have longer time scales, in the order of 10 s or even longer.

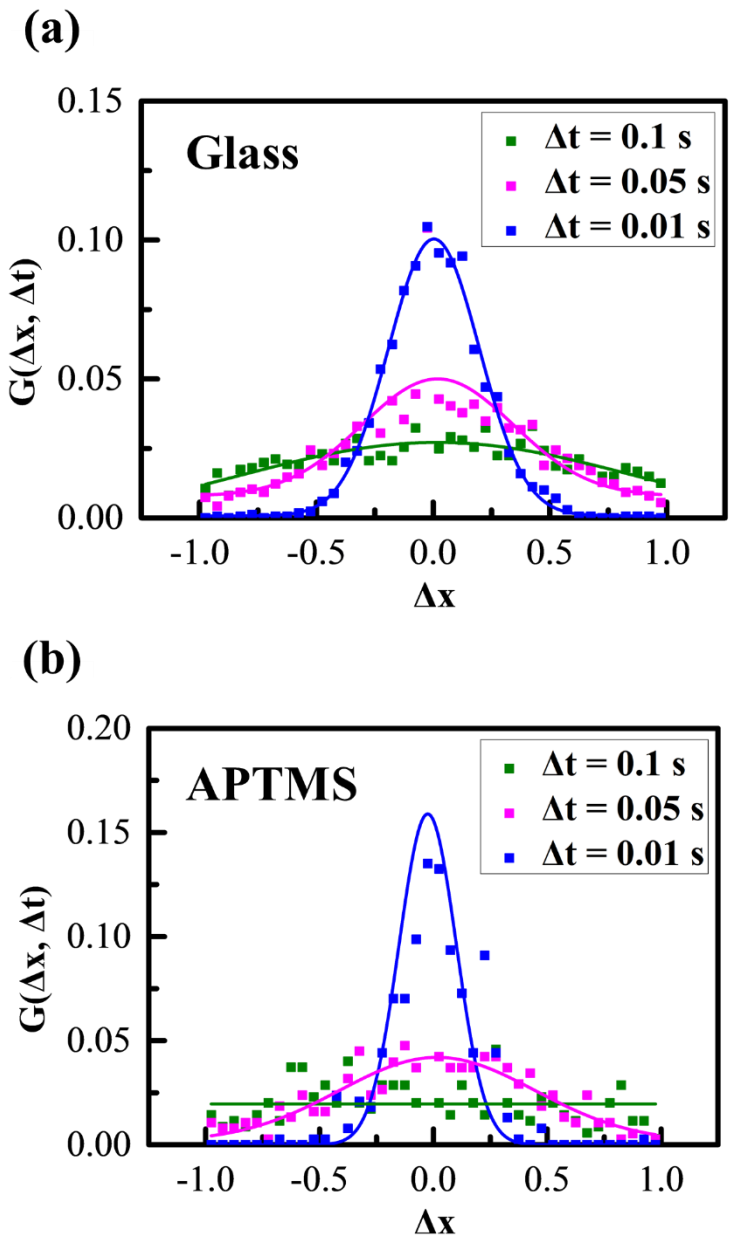


Figure 8. van Hove correlation function derived from LLPS droplets trajectories on (a) untreated and (b) APTMS-modified cover glass surface. The solid curve represents a fitting result with Gaussian distribution at Δt = 10, 50 and 100 milliseconds.

After SPT on the diffusion of 100 LLPS droplets on both substrates, we proceeded to statistically analyze the relation between LLPS droplet size and fluorescence intensity, as shown in Fig. 9 (a) and (b). To differentiate the differences in LLPS diffusion patterns on these two substrates, we coordinated their average area and average total intensity,

as shown in Fig. 9 (c). Obviously, for the untreated glass surface, LLPS particles have a broader scattering distribution with a dominant diffusion pattern, while LLPS particles on the APTMS-modified surface have a narrower distribution with a more fixed pattern in Fig. S4. Because the zeta potential of LLPS droplets is negative, we recognize the interaction between LLPS and the substrate may have a dominant effect on its diffusion coefficient and diffusion pattern.

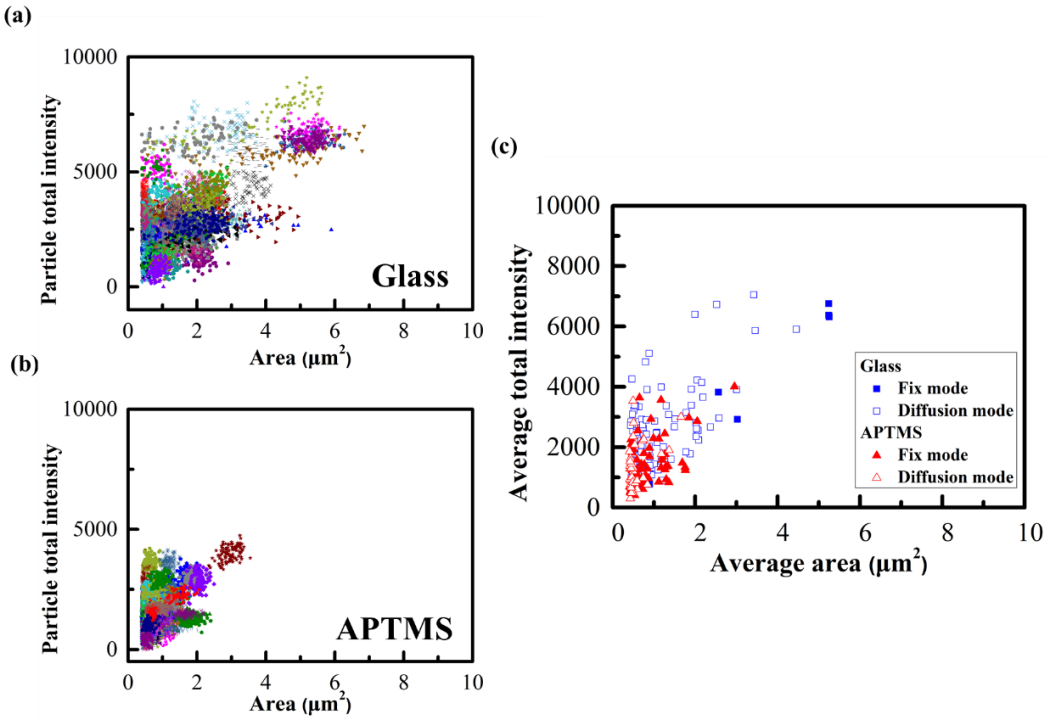


Figure 9. Scattered plots showing the relationship between area and fluorescent intensity of droplets. (a) Untreated and (b) APTMS-modified cover glass. Each color refers to each droplet group; (c) Relationship between average area and average total intensity for fix and diffusion mode on both two substrates.

5.5 Conclusion

In this part, we successfully analyzed the diffusion of LLPS droplets consisting of poly(PR) dipeptide repeats and poly-A RNA on the glass surface. Consequently, we found two diffusion modes, namely the fix mode and diffusion mode. The distribution of these diffusion modes was adjusted by chemical modification of the glass surface with amines. The negative zeta potential of LLPS droplet surface may tend to increase its interaction with positively charged glass surface, leading to more droplets in the fix mode. Understanding the dynamics of LLPS droplets at the solid/liquid interface may provide a new approach to advance the fundamental studies of LLPS properties. In addition to the previously reported knowledge of LLPS detection, this insight into its diffusion control can be used in engineering approaches to tune LLPS coalescence and fusion.

5.6 References

1. Prouteau, M. & Loewith, R. Regulation of cellular metabolism through phase separation of enzymes. *Biomolecules* **8**, 1–14 (2018).
2. Sheu-Gruttadauria, J. & MacRae, I. J. Phase Transitions in the Assembly and Function of Human miRISC. *Cell* **173**, 946-957.e16 (2018).
3. Dao, T. P. *et al*. ALS-Linked Mutations Affect UBQLN2 Oligomerization and Phase Separation in a Position- and Amino Acid-Dependent Manner. *Structure* **27,** 937-951.e5 (2019).
4. Wang, Z., Zhang, G. & Zhang, H. Protocol for analyzing protein liquid–liquid phase separation. *Biophys. Reports* **5**, 1–9 (2019).
5. Alberti, S. *et al*. A User's Guide for Phase Separation Assays with Purified Proteins. *J. Mol. Biol*. **430**, 4806–4820 (2018).
6. Welsh, T. J. *et al.* Single particle zeta-potential measurements reveal the role of electrostatics in protein condensate stability. *bioRxiv* 1–35 (2020).
7. Boeynaems, S. *et al*. Spontaneous driving forces give rise to protein − RNA condensates with coexisting phases and complex material properties. *Proc. Natl. Acad. Sci. U. S. A*. **116**, 7889–7898 (2019).
8. Kanekura, K. et al. Characterization of membrane penetration and cytotoxicity of C9orf72-encoding arginine-rich dipeptides. *Sci. Rep*. **8**, 1–11 (2018).
9. Kanekura, K. et al. Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. *Hum. Mol. Genet*. **25**, 1803–1813 (2016).
10. Lin, Y. *et al.* Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. *Cell* **167**, 789-802.e12 (2016).
11. Wang, Y. Q., Xu, S. Bin, Deng, J. G. & Gao, L. Z. Enhancing the efficiency of planar heterojunction perovskite solar cells via interfacial engineering with 3-aminopropyl trimethoxy silane hydrolysate. *R. Soc. Open Sci*. **4**, (2017).
12. Saxton, M. J. Single-particle tracking: Connecting the dots. *Nat. Methods* **5**, 671–672 (2008).
13. Schindelin, J. *et al.* Fiji: An open-source platform for biological-image analysis. *Nat.*

Methods **9**, 676–682 (2012).

14. Sbalzarini, I. F. & Koumoutsakos, P. Feature point tracking and trajectory analysis for video imaging in cell biology. *J. Struct. Biol.* **151**, 182–195 (2005).
15. Berg, H. C. Random walks in biology. Princeton University Press, 1993.
16. Hansen, J. P.; McDonald, I. R. Theory of Simple Liquids (New York: Academic.), 2006.

CHAPTER 6. Summary and outlook

In the present thesis, to understand the process of Arg-rich dipeptide repeat proteins (DPRs) to form liquid-liquid phase separation (LLPS) and its wetting property, which have close correlation to its biological function and even neurodegenerative diseases, like ALS, the molecular interactions of DPRs in the LLPS and also the LLPS droplets diffusion at solid/liquid interface in a dynamic manner have been investigated.

In chapter 2, we designed a series of poly(PR) variants: $(PR)_4$, $(P_2R_2)_2$ and P_4R_4 with the same total length and density of positive residue Arg but different periodicities (the size of periodicity t = 2, 4 and 8) and homopolymeric adenine poly-rA as a model for RNA. Then, we utilized molecular dynamics (MD) simulations on one-by-one interactions, which can provide detailed information on the conformational changes and thermodynamics of biological molecules and their complexes at atomic level. As a result, we found the binding energy strength was inversely proportional to the periodicity value of poly(PR) variants. Alternating charge distribution in poly(PR) results in lower binding energy while poly(PR) variants with higher periodicities (e.g. PnRn) have higher binding energy when interacting with RNA. Besides, we also calculated free energy landscape for the interacting complexes in order to explore Pro residues contribution towards structural property of poly(PR) variants. We found that Pro residues not only modulated solvation of water on the biomolecular surface but could confer rigidity for peptide structures and thus intervened the interaction between cationic amino acids and anionic nucleic acids.

In chapter 3, we *in vitro* and *in vivo* experimentally demonstrated the relationship between Arg distribution of poly(PR) variants and their interaction with RNA and protein. We firstly synthesized poly(PR) variants: $(PR)_{12}$, $(P_2R_2)_6$, $(P_4R_4)_3$ and $P_{12}R_{12}$ and poly-rA as RNA. Droplets of $(PR)_{12}$ variants with larger periodicities lost their spherical structure under this condition, and $P_{12}R_{12}$ formed sticky non-spherical

condensates losing surface tension minimization as the typical characteristic of liquids. Next, we conducted fluorescence recovery after photobleaching (FRAP) experiments to investigate the dynamic diffusion process of poly(PR) variants LLPS droplets from the perspective of two-body interaction (poly(PR) variants and RNA) and three-body interaction (poly(PR) variants, RNA and protein). For two-body interaction, we found RNA mixed with $(PR)_{12}$ had a higher fluidity than RNA mixed with $(P_4R_4)_3$ or $P_{12}R_{12}$ and also Pro residue could loosen complex coacervation structure, which provide enough multivalency to form LLPS droplet according to critical salt concentration (CSC) experiments. For three-body interaction, we observed that the rate of dynamic exchange of NPM1 protein in phase equilibrium was disrupted by complex coacervation with $(PR)_n$. $(PR)_n$ condensates NPM1 protein and rRNA in phase-separated droplets via multivalent interaction, thereby affecting the mobility of NPM1 protein in cells, whereas P_nR_n failed to condense NPM1 protein and accumulated to rRNA. Thus, P_nR_n did not significantly affect NPM1 protein mobility.

In chapter 4, we did cellular proteomic analysis to explore the interaction motif of protein for poly(PR) peptide. We found that dipeptide-protein interaction was affected by the size of $(PR)_{12}$ periodicity and alternate Arg structure of DPRs could efficiently enhance dipeptide-protein interaction. Next, we did structural decoding for these proteins and demonstrated the electrostatic interactions between positive Arg amino acids and acidic motifs were the main driving force for interactome enrichments, which could be explained, at least partially, by LLPS.

In chapter 5, we focused on the dynamic diffusion process of LLPS droplets consisting of ploy(PR) dipeptide repeats, fluorescence tagged poly(PR) dipeptide repeats and poly-rA RNA at solid/liquid interface. We explored two solid surfaces, one was untreated cover glass and the other was positive APTMS chemically modified cover glass, and then utilized fluorescence microscope to observe the diffusion of LLPS droplets on these two solid surfaces, respectively. As the result, after using single

particle analysis, we found two diffusion modes, fix mode and diffusion mode. The chemical modification of the glass surface by amine termination tuned the distribution of these diffusion modes. Meanwhile, the negative zeta potential of LLPS droplet surface probably tended to increase its interaction with positively charged glass surface, resulting in exhibiting more droplets in the fix mode. These results could provide an improved method to advance fundamental studies for understanding nature of LLPS droplets.

Investigation on the molecular interactions of poly(PR) dipeptide repeats in the LLPS and the droplets diffusion at solid/liquid interface, we found the alternate distribution of Arg and Pro in poly(PR) controls for phase separation of poly(PR) with proteins /nucleic acids and determines NPM1 protein mobility, which may be helpful to explain the impedance of functions of membrane-less organelles including nucleoli and ALS disease pathology. Besides, we also focused on the LLPS diffusion modes at two different solid surfaces in aqueous environment, one is untreated cover glass and the other is chemically modified cover glass with amine groups, which can provide an improved method to control LLPS diffusion at solid/liquid interface by engineering methods to tune LLPS coalescence and fusion. Most importantly, the study of LLPS interface interaction with outer medium is closely related to LLPS uptake for surrounding molecules. The control of this uptake process is desired to be helpful for the curation of ALS disease in the near future.

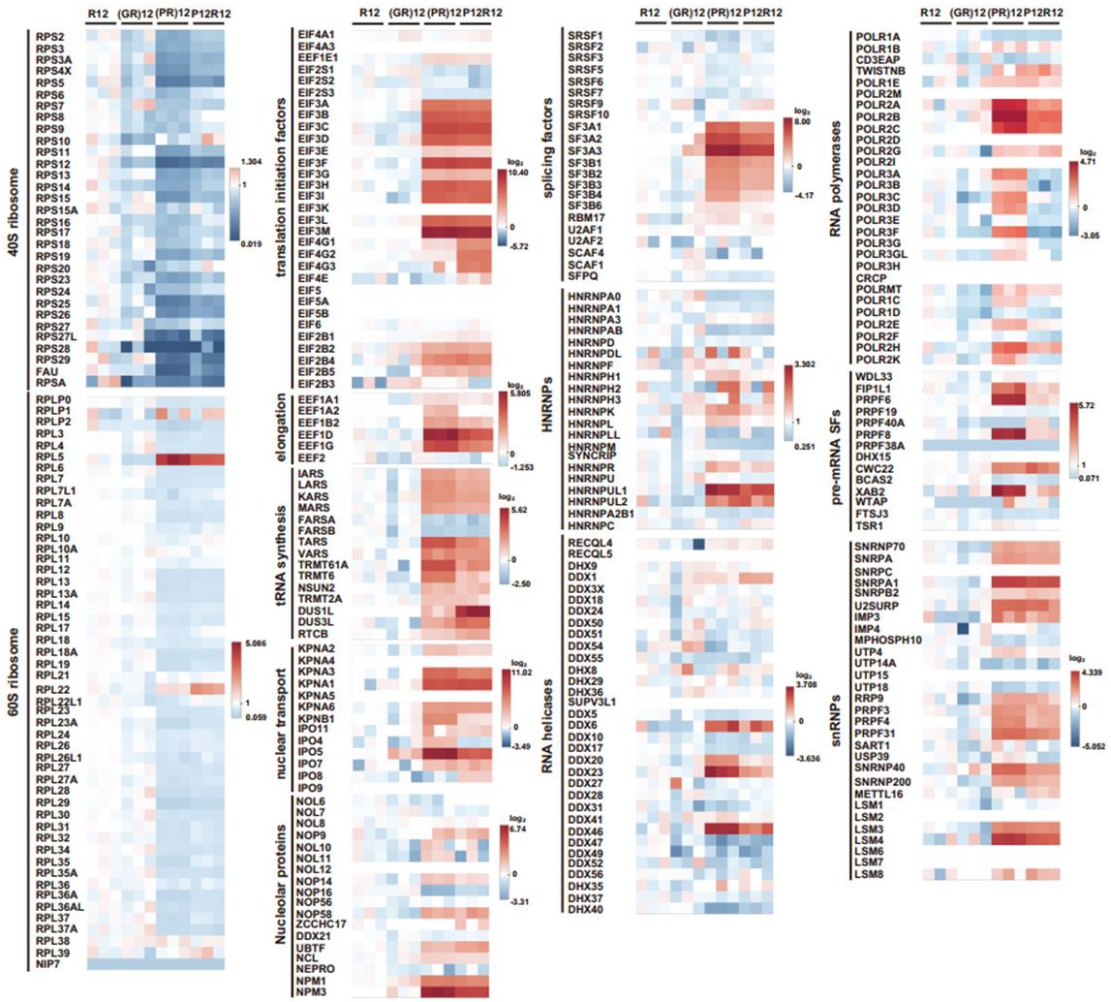


Figure S1 The heatmap showing all three classes of RNA polymerases could be targeted by $(PR)_{12}$. To examine whether RNA polymerase activity is indeed impaired by poly(PR), 5-Ethynyl uridine was incorporated into newly synthesized RNA during transcription and visualized with Alexa Fluor 594 azide by click chemistry reaction

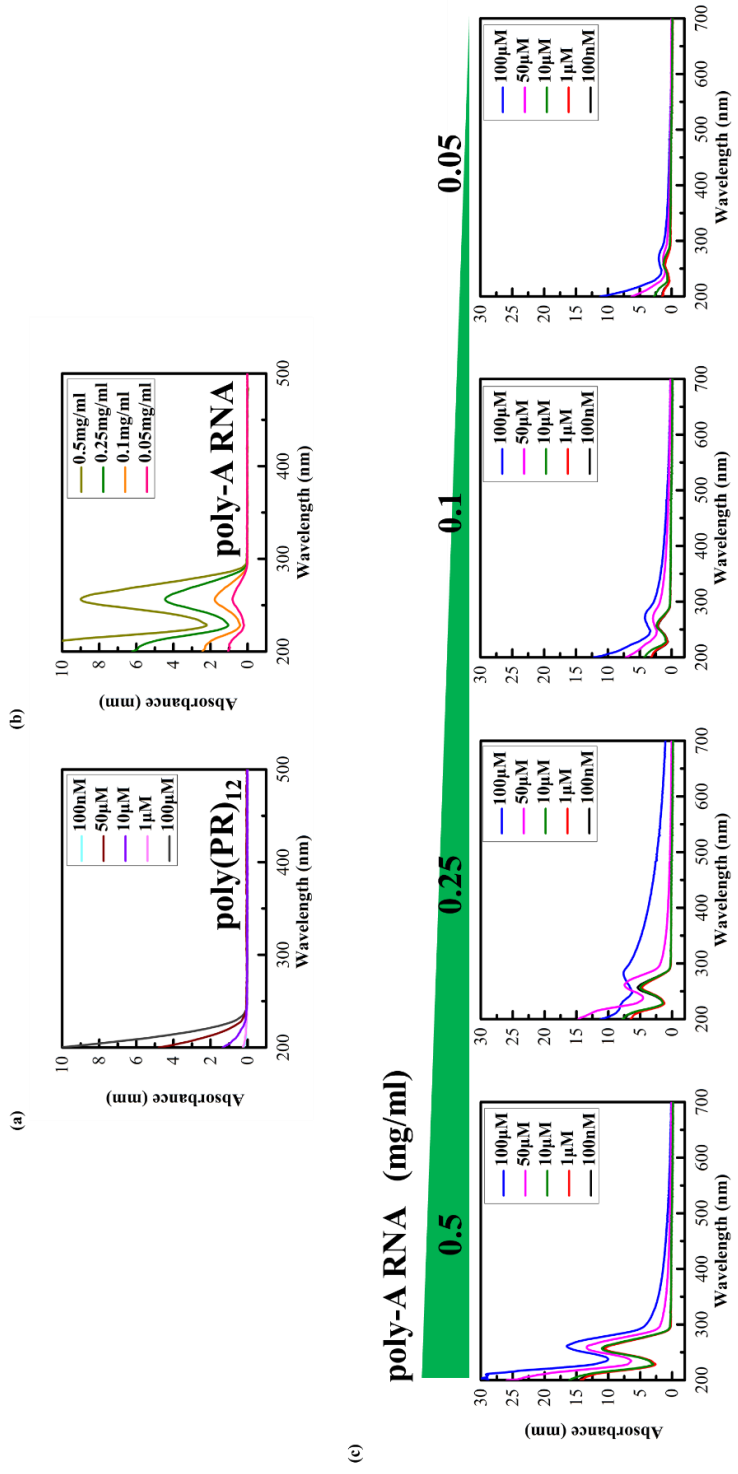


Figure S2. UV absorbance spectra for pure peptide/RNA and mixture solution. (a) poly(PR)$_{12}$ with concentrations ranging from 100 μM, 10 μM, 1 μM to 100 nM; (b) poly-A RNA with concentrations ranging from 0.5 mg/ml, 0.25 mg/ml, 0.1 mg/ml to 0.05 mg/ml; (c) Mixture solutions for poly(PR)$_{12}$ and poly-A RNA in each respective concentration above.

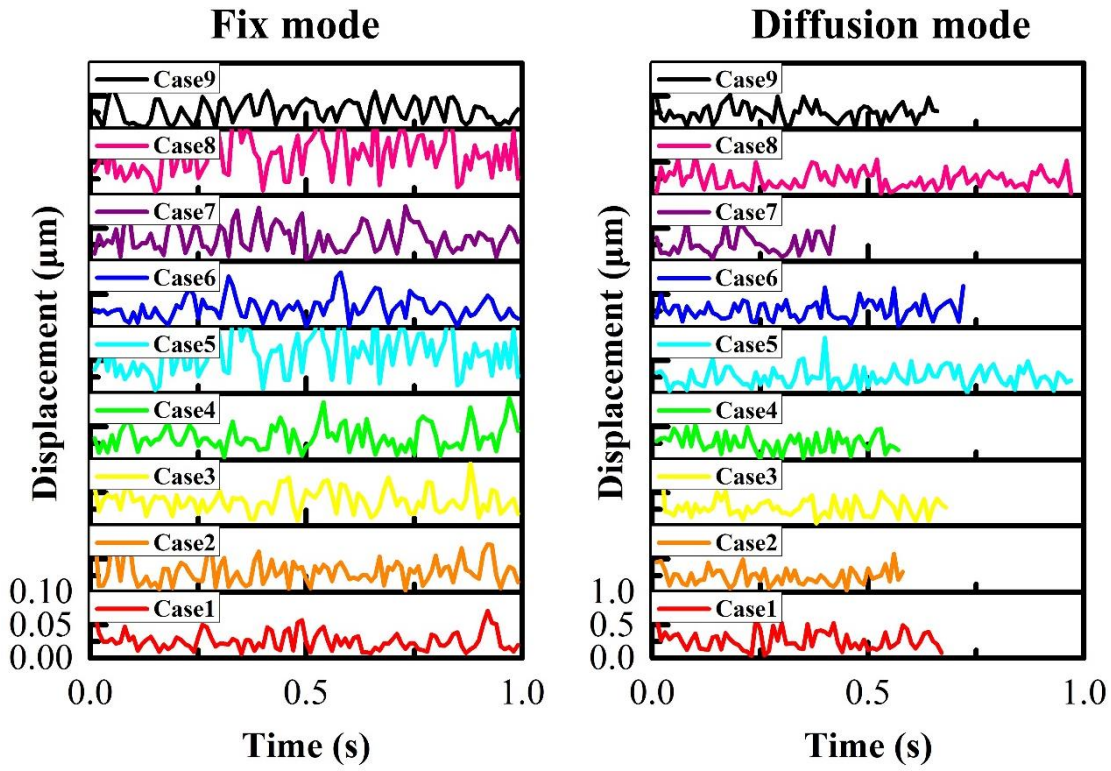


Figure S3. Cases about the fluctuations of LLPS droplet surface displacement over time under fix mode and diffusion mode.

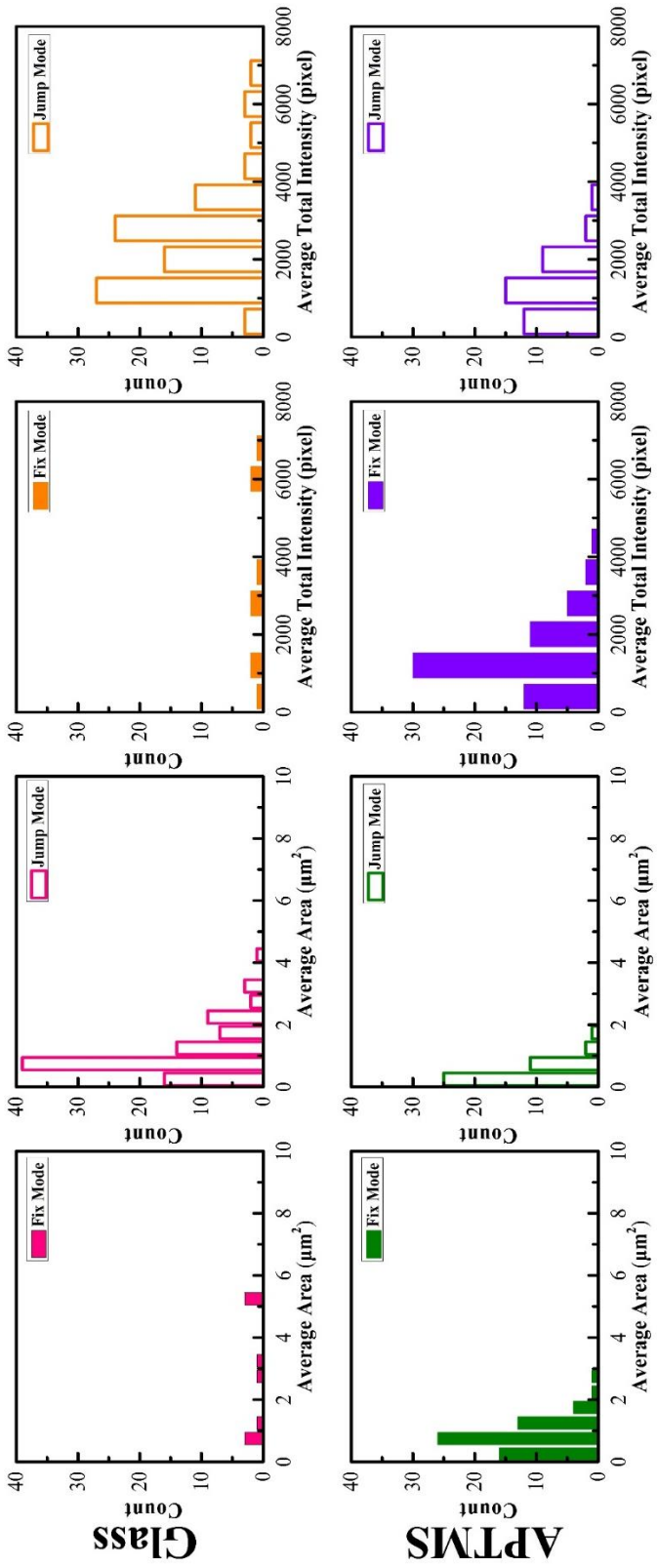


Figure S4. The distribution of average area and average total intensity for LLPS droplets diffusing at both two substrates (Glass and APTMS).