

1 **Small but complete: liquid-liquid phase separation in minimalistic peptides**

2
3 Yiming Tang^a, Santu Bera^b, Yifei Yao^a, Jiyuan Zeng^a, Zenghui Lao^a, Xuewei Dong^a, Ehud Gazit^{b*}, and
4 Guanghong Wei^{a*}

5
6 a Department of Physics, State Key Laboratory of Surface physics, and Key Laboratory for
7 Computational Physical Science (Ministry of Education), Fudan University, Shanghai 200433,
8 People's Republic of China

9 **E-mail:** ghwei@fudan.edu.cn

10
11 b Shmunis School of Biomedicine and Cancer Research, George S. Wise Faculty of Life Sciences,
12 Department of Materials Science and Engineering, Iby and Aladar Fleischman Faculty of
13 Engineering, Tel Aviv University, 6997801 Tel Aviv, Israel

14 **E-mail:** ehudg@post.tau.ac.il

15 16 17 **Significance**

18 Protein liquid-liquid phase separation (LLPS) is associated with human health and diseases.
19 Identifying the minimalistic building blocks and sequence determinants of LLPS is of urgent
20 importance but remains computationally challenging partially due to the lack of methodologies
21 characterizing the liquid condensates. Herein we provide approaches to evaluate LLPS ability of
22 dipeptides, and screen all 400 dipeptides by MD simulations combined with multi-bead-per-
23 residue models which capture key interactions driving LLPS that are missing in one-bead-per-
24 residue models. Three categories of LLPS dipeptides are identified and the experimentally-verified
25 QW dipeptide is by far the smallest LLPS system. Our results suggest that dipeptides, minimal but
26 complete, possess multivalent interactions sufficient for LLPS, and LLPS is a general property of
27 peptides/proteins, independent of their length.

28 29 **Abstract**

30 Liquid-liquid phase separation (LLPS) of proteins mediates the assembly of biomolecular
31 condensates involved in physiological and pathological processes. Identifying the minimalistic
32 building blocks and the sequence determinant of protein phase separation is of urgent importance
33 but remains challenging due to the enormous sequence space and difficulties of existing
34 methodologies in characterizing the phase behavior of ultrashort peptides. Here we demonstrate
35 computational tools to efficiently quantify the microscopic fluidity and density of liquid-
36 condensates/solid-aggregates and the temperature-dependent phase diagram of peptides.
37 Utilizing our approaches, we comprehensively predict the LLPS abilities of all 400 dipeptide
38 combinations of coded amino acids based on 492 micro-second molecular dynamics simulations,
39 and observe the occurrences of spontaneous LLPS. We identify 54 dipeptides that form solid-like
40 aggregates and three categories of dipeptides with high LLPS propensity. Our predictions are

1 validated by turbidity assays and differential interference contrast (DIC) microscopy on four
2 representative dipeptides (WW, QW, GF, and VI). Phase coexistence diagrams are constructed to
3 explore the temperature dependence of LLPS. Our results reveal that aromatic moieties are crucial
4 for a dipeptide to undergo LLPS, and hydrophobic and polar components are indispensable. We
5 demonstrate for the first time that dipeptides, minimal but complete, possess multivalent
6 interactions sufficient for LLPS, suggesting that LLPS is a general property of peptides/proteins,
7 independent of their sequence length. This study provides a computational and experimental
8 approach to the prediction and characterization of the phase behavior of minimalistic peptides,
9 and will be helpful for understanding the sequence-dependence and molecular mechanism of
10 protein phase separation.
11

1 **Introduction**

2 Accumulating evidence demonstrates the importance of protein phase separation in physiological
3 and pathological processes (1–3). Liquid-liquid phase separation (LLPS) of proteins underlies the
4 formation of membraneless compartments in cells (4, 5), possessing important intercellular
5 functions, such as facilitating bio-chemical reactions, activating stress response, and regulating
6 gene transcriptions (6–8). In addition, aberrant LLPS and the liquid-solid phase separation (LSPS)
7 can lead to formation of neurotoxic aggregates such as amyloid plaques in Alzheimer’s disease,
8 fibrillary aggregates in Amyotrophic Lateral Sclerosis (ALS) (9, 10), aberrant stress granules in
9 Frontotemporal Dementia (FTD) (11), and Lewy bodies in Parkinson’s disease (12). Identifying the
10 sequence determinants and the molecular mechanisms underlying phase separation is yet an
11 unmet need crucial for designing therapeutic targets to treat protein-aggregation disorders.

12

13 A rapidly increasing number of proteins have been demonstrated to have LLPS capabilities.
14 Examples include the transactive response DNA binding protein (TDP-43) (13, 14), Tau (15–17) and
15 Fused in Sarcoma (FUS) (18–20). It has been reported that LLPS is driven by multivalent interactions
16 entailed by the amino acid sequence (1, 21, 22). In addition, it is considered that a minimum
17 sequence length of protein is required and below which LLPS cannot occur. For example, the LLPS
18 of the poly-PR/poly-GA requires a minimum peptide length of 50/100 (23, 24). In contrast,
19 numerous experiments demonstrate that even very short peptides can self-assemble into ordered
20 solid-like nanoarchitectures. For example, FF, FW, and IF can respectively form nanotubes, tubular
21 and fibrous nanostructures (25–27). In this context, we suspect that ultrashort peptides, as short
22 as dipeptides, may also possess LLPS capacity. In order to identify the minimalistic building blocks
23 and the sequence determinant of protein phase separation, we decipher the LLPS/LSPS ability of
24 all 400 dipeptide combinations of coded amino acids by conducting a combined computational and
25 experimental study.

26

27 Different theoretical and computational approaches have shown success in probing the phase
28 behaviors of intrinsically disordered proteins (IDPs) or proteins with intrinsically disordered regions
29 (IDRs) (28, 29). Such methods include random phase approximation theory by Lin and Chan (30),
30 field-theoretic simulations by Shea *et al.* (31), Gibbs-ensemble simulations by Zhou *et al.* (32, 33),
31 and slab simulations by Mittal *et al.* (34–36) and others (37). Those studies greatly enhance our
32 understanding of key physical principles and sequence features crucial to phase separation of IDPs
33 or IDR-containing proteins. Following these studies, we have attempted to investigate the phase
34 separation of dipeptides by utilizing the slab simulation protocol in combination with the
35 hydrophobicity-scaled (HPS) one-bead-per-residue model. Unexpectedly, dipeptides with known
36 high aggregation capability (such as FF, FW and IF) display totally no sign of aggregation
37 (Supplementary Part 1), probably due to the fact that highly-simplified protein models significantly
38 reduces the molecular interaction sites, thus lack molecular details to explicitly describe aromatic
39 stacking, hydrogen-bonding interactions and peptide-water interactions crucial for ultrashort

1 peptide phase separation (22, 38–40). On the other hand, the slab simulation protocol, which has
2 become a *de facto* standard for computing the LLPS phase diagrams (29), underestimates the
3 density of dilute phases (i.e., the saturation concentration for phase separation) probably due to
4 the lack of explicit solvents. Thus, a more detailed description of ultrashort peptides in combination
5 with explicit solvent model is necessary for the investigation of their LLPS.

6

7 Hereby, we have carried out micro-second explicit solvent molecular dynamics (MD) simulations
8 on each of the 400 dipeptides in combination with the multiple-bead-per-residue MARTINI model
9 which explicitly capture specific physical interactions (such as aromatic stacking, and protein-water
10 interactions). A similar simulation protocol has been used by Tuttle and coworkers to screen out
11 dipeptides with high self-assembly ability (41), albeit with a short simulation time scale.
12 Nevertheless, the LLPS ability of dipeptides and their phase behaviors remain completely elusive.
13 We have developed a novel, efficient approach to predict the LLPS propensities of ultrashort
14 peptides which consists of two steps: (1) the aggregation capability of each dipeptide is quantified
15 by the clustering degree and the collapse degree of initially randomly dispersed dipeptide
16 molecules, and (2) we define two physical parameters (the fluctuation extent of clustering degree
17 and the exchange rate of inter-peptide interactions) to characterize the microscopic fluidity of the
18 aggregates and distinguish between LLPS and LSPS. Our extensive simulations demonstrate that
19 dipeptides can spontaneously phase separate into solid- or liquid-like condensates, and predict the
20 LLPS and LSPS propensities of all 400 dipeptides, among which QW possesses the highest LLPS
21 propensity. Our predictions are further verified through temperature dependent turbidity assays
22 and differential interference contrast (DIC) microscopy. Moreover, we unravel the molecular
23 mechanism underlying the QW LLPS and construct the temperature-dependent density profiles
24 (i.e., the phase diagram). To the best of our knowledge, this is the first study demonstrating the
25 LLPS behavior of dipeptides (the minimalistic peptide) which is intractable by existing theoretical
26 and computational methods for studying protein LLPS.

27

28 **Results**

29 **The aggregation capability of each of the 400 dipeptides**

30 For each of the 400 dipeptides, we have performed 1~2 individual micro-second simulations
31 starting from 720 dipeptide molecules randomly dispersed in aqueous solution. Reaching
32 equilibrium is essential to characterizing the phase behavior of peptides, so we assess the
33 convergence of each simulation by tracking the time evolution of solution accessible surface area
34 (SASA) (Fig. S4), and the length of each simulation varies from 2.5~10.0 μ s (the number of
35 simulations for each dipeptide as well as the time length of each simulation are given in Table S1-
36 3).

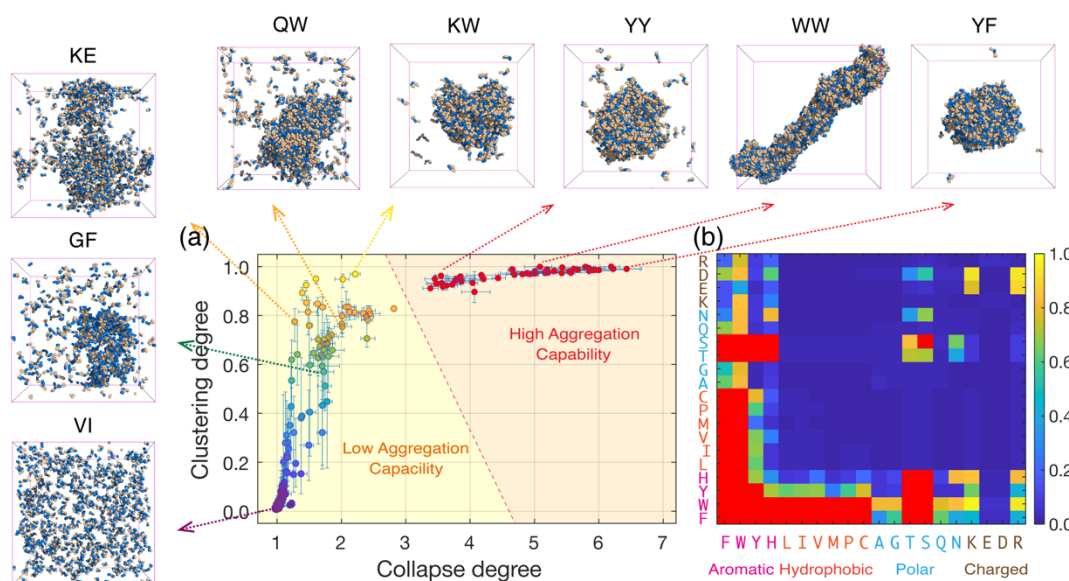
37

38 An important feature of LLPS is the separation of a solute-rich dense phase from the homogenous
39 solution, thus the first step in deciphering the LLPS propensity of a dipeptide is to identify its

1 aggregation capability. Tuttle and Ulijn have performed short time-scale MD simulations on all 400
2 dipeptides, and screened out dipeptides with high self-assembly tendencies by defining an
3 “aggregation propensity” (AP) parameter (41), which is calculated by the ratio of the solvent-
4 accessible surface area (SASA) of all dipeptide molecules in the initial state to their SASA in the final
5 configuration. Those high-AP peptides usually form solid-like aggregates (25, 27, 42), suggesting
6 their low ability to undergo LLPS. As shown in Fig. S5a, the AP parameter is highly degenerate for
7 relatively low aggregation capability peptides. We therefore introduce a new parameter: the
8 clustering degree (defined by the number of dipeptide molecules in the largest cluster divided by
9 total number of dipeptide molecules in the system), and characterize the aggregation capability of
10 each dipeptide by two parameters (clustering degree and AP). For clarity, the AP parameter (41,
11 43) is hereinafter referred as the collapse degree. Figure 1a shows the clustering degree of the 400
12 dipeptides as a function of their collapse degree. A gap exists around the collapse degree value of
13 3.0, dividing 400 dipeptides into two groups. The first group contains 54 dipeptides which present
14 large values of collapse degree (3.0~7.0) and clustering degree (>0.9), suggesting that those
15 dipeptides have high aggregation capabilities (see snapshots of YF/WW/YY). Dipeptides in the first
16 group include: (1) dipeptides composed by aromatic residues F/W/Y, (2) dipeptides containing an
17 F/W residue and a hydrophobic residue (H/L/I/V/M/P/C) or S/T residue (Fig. 1b). These results
18 reaffirm that aromatic residues are crucial for the aggregation of dipeptides, and hydrophobic
19 residues enhance the aggregation.

20
21 Dipeptides in the second group have smaller values of collapse degree (1.0~3.0), suggesting their
22 relatively lower aggregation capabilities. However, they present a broad and continuous
23 distribution of clustering degree. Dipeptides with high clustering degree values can aggregate into
24 a large but loosely-packed cluster (see snapshots of WA/QW/KE/GF), while those with low
25 clustering degree values cannot aggregate at all (see snapshots of VI). Therefore, this parameter
26 can be used to further estimate the aggregation ability of dipeptides in the second group (they are
27 colored from blue to yellow in the order of increasing clustering degree in Fig. 1b). We note that
28 there is no gap in the clustering degree value distribution, so further classification of dipeptides in
29 the second group requires alternative measurements.

30



1

2 **Fig. 1. Sequence-dependent aggregation capabilities of all 400 dipeptides.** (a) The clustering
 3 degree value of all dipeptides as a function of their collapse degree value. Final snapshots of 8
 4 representative dipeptides are given. (b) The aggregation capability map of all dipeptides.
 5 Dipeptides with high aggregation capabilities are colored red, while those with relatively low
 6 aggregation capabilities are colored from blue to yellow in the order of increasing clustering degree
 7 values. The horizontal and vertical axes correspond respectively to the N- and C-terminal residue.

8

9 **The fluidity of the dipeptide aggregates**

10 Aggregation capability is an effective measurement of whether a dipeptide forms a peptide-rich
 11 dense phase, but additional parameters are needed to understand the dynamical properties of the
 12 aggregates and to distinguish between LLPS and LSPS. Based on physical principles on state of
 13 matters, liquids can be distinguished from solids by their fluidity: molecules in solids keep their
 14 specific neighborhood for a long time, and rarely exchange with their surroundings, while
 15 molecules in liquids can easily rearrange (8, 44). Thus, the liquid-like dipeptide aggregates can be
 16 distinguished from solid-like aggregates by the rapid exchange of contents between dense and
 17 dilute phase, and the rapid exchange of inter-molecular interactions within dense phase (44) (Fig.
 18 2a). These two properties of liquid can respectively be described by the fluctuation extent of
 19 clustering degree parameter (FE , defined by the root mean square fluctuation of the dense phase
 20 size) and the exchange rate of interaction parameter (ER , defined as the fraction of inter-molecular
 21 interactions within the dense phase that are lost per unit time).

22

23

$$FE = \sqrt{\langle [CD(t) - \langle CD(t) \rangle]^2 \rangle}$$

$$ER = \left\langle \frac{\sum_{dense\ phase} \# \text{ of lost interaction } (t \rightarrow t + \Delta t)}{\sum_{dense\ phase} \# \text{ of inter-peptide interaction } (t)} \times \frac{1}{\Delta t} \right\rangle$$

24

25

26

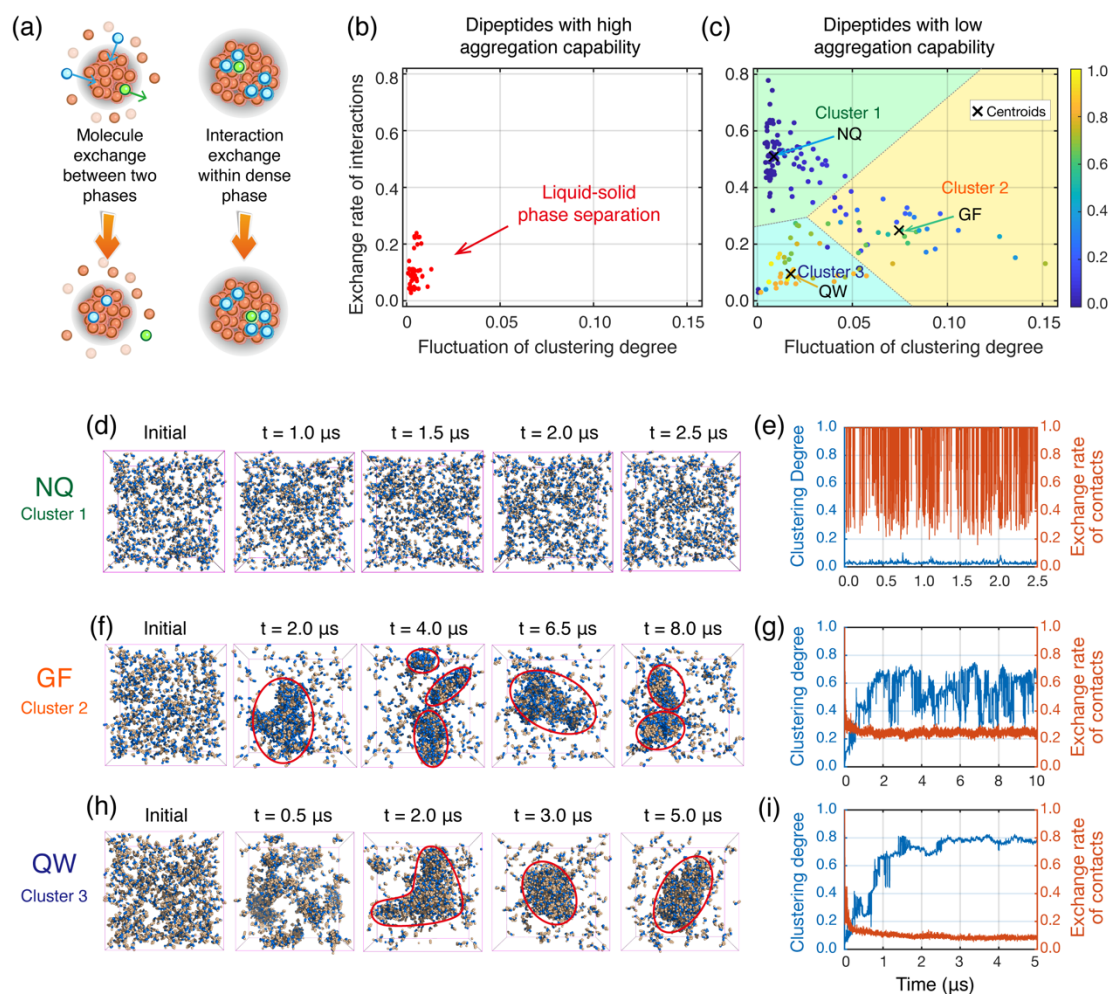
where $CD(t)$ represents the clustering degree at a given time t , the angular bracket $\langle \rangle$ indicate
 averaging over conformation and over time during the last 100 ns of each simulation trajectory,
 and the summation is performed on all molecules within the dense phase. The FE and ER

1 parameters reflect two independent aspect of the microscopic fluidity of the dipeptide aggregates,
2 and they are both positively correlated with the fluidity.

3

4 We calculate respectively the fluidity of dipeptides in group 1 and 2 (corresponding to dipeptides
5 with high and low aggregation capability, respectively). Dipeptides in group 1 display small values
6 of both *FE* and *ER*, suggesting poor fluidity and formation of solid-like aggregates (Fig. 2b and
7 Supplementary Video 1). They are classified as dipeptides with high LSPS propensities. In contrast,
8 dipeptides in group 2 present a broad distribution of both *FE* and *ER* values (Fig. 2c). They are
9 further classified into three clusters using a k-means algorithm (45). NQ/GF/QW dipeptide, which
10 is located closest to the centroid of each cluster, is taken as the representative of the corresponding
11 cluster. NQ (centroid of cluster 1) molecules do not aggregate at all (Fig. 2d and Supplementary
12 Video 2). Its clustering degree values maintain a small value of 0.02, and the interaction exchange
13 rate fluctuates around a large value of 0.6 (Fig. 2e), suggesting poor phase separation ability.
14 Interestingly, many hydrophobic dipeptides (such as VI) belong to this cluster and don't phase
15 separate (Supplementary Video 3). GF (centroid of cluster 2) molecules can self-assemble into a
16 single aggregate, but the aggregate rapidly dissociates and reassociates within a microsecond
17 timescale (Fig. 2f and Supplementary Video 4), which can be reflected by the rapid fluctuation of
18 clustering degree (Fig. 2g). The excessive liquidity of GF aggregates suggests its inability to form
19 liquid-like droplets. In contrast, QW (centroid of cluster 3) molecules self-associate into a single
20 ellipsoid-like aggregate, which maintains its overall stability albeit with a certain degree of fluidity
21 (Fig. 2h and Supplementary Video 5). The clustering degree of QW fluctuates slightly, and the *ER*
22 value of QW maintain a much smaller value of ~0.09 compared to that of NQ (~0.6) and GF (~0.25),
23 indicative of liquid-like QW aggregates and possibility of forming liquid-like droplets (Fig. 2i). We
24 thus infer that QW has a high LLPS propensity.

25



1

2 **Fig. 2. The fluidity of the dipeptide aggregates.** (a) Illustration of the two parameters (the
3 fluctuation extent of clustering degree, and the exchange rate of interactions) to characterize the
4 fluidity of dipeptide aggregates. (b) The exchange rate of interactions values of dipeptides in group
5 1 (with high aggregation propensity) as a function of their fluctuation extent of clustering degree
6 values. (c) Similar to (b) but for dipeptides in group 2 (with low aggregation propensity). The points
7 are colored from blue to yellow in the order of increasing clustering degree values. All points are
8 grouped into three clusters using a k-means algorithm. (d) Snapshots of NQ (closest to the centroid
9 of cluster 1) molecules at five different time points. (e) Degree of clustering and exchange rate of
10 contacts parameters as a function of simulation time for NQ molecules. (f-i) Similar to (d-e) but for
11 (f-g) GF (closest to the centroid of cluster 2) molecules and (h-i) QW (closest to the centroid of
12 cluster 3) molecules.

13

14 **Experimental verifications and characterization of QW LLPS**

15 To verify the differences in the phase behavior of dipeptides belong to different groups, we choose
16 four dipeptides (WW, QW, GF and VI) and perform temperature dependent turbidity assay (a
17 standard process to monitor liquid-liquid demixing (46)). We fix the peptide concentration (1.5 mM)
18 at physiological pH=7.4 and measure the turbidity at 350 nm upon variation of the solution

1 temperature from 5°C to 55°C (Fig. 3a-c). We detect very little turbidity for all four dipeptide
2 solutions below 20°C, indicating the presence of dispersed monomers. With temperature
3 increasing, the turbidity of QW gradually rises, reaches to a maximum around the temperature 37-
4 42°C and decreases again at even higher temperature (Fig. 3a). The variation of solution turbidity
5 can arise from LLPS or formation of other aggregates (LSPS). To identify the presence of specific
6 phase separated state, we employ the DIC microscopy. At 5°C, the solution is clear and no evidence
7 of liquid-demixing is identified (Fig. 3d). Upon increasing the temperature to 25°C, various sizes of
8 liquid droplets are observed with both larger diameter ~20 µm and smaller (1 µm or below) (Fig.
9 3d and Supplementary video 6). The number and size of the droplets increase with the
10 temperature rising to 37°C, in agreement with the observed high turbidity (Fig. 3d and
11 Supplementary video 7). However, further increase of temperature (65°C) results in disappearance
12 of the droplets after 1h but reappear on cooling down the solution to 37°C for 3 h, indicative of
13 temperature dependent LLPS of QW (Fig. 3d). We study the dynamics of peptide molecules in the
14 droplets during their formation and maturation over time. A time-lapse detailed inspection of the
15 droplets shows the coalescence of two droplets to form single larger droplets through droplet
16 fusion, confirming their liquid-like nature (Fig. 3e). The presence of peptide molecules in the liquid
17 droplets is confirmed from the observation that the droplets converted to solid aggregates upon
18 complete evaporation of the solvent (Fig. S6 and Supplementary video 8). On the other hand, WW
19 shows continuous increase of turbidity with rise of temperature probably due to formation of static
20 solid aggregates at higher temperature (Fig. 3b). Starting from clear solution at 5°C, DIC images of
21 WW at 25°C exhibit the formation of smaller spherical assemblies with diameter below 1 µm (Fig.
22 3f). These assemblies gradually transform into elongated structure at higher temperature, 37°C.
23 Further increase of temperature (65°C) results into the formation of solid-like bigger elongated
24 aggregates, in sharp contrast to the temperature-dependent behavior of QW. The solid-nature of
25 these aggregates is further confirmed from the fact that these structures remain undamaged and
26 the shape persists unaltered for long time after coming out from the peptide solution (Fig. S7 and
27 Supplementary video 9,10). Therefore, WW undergoes temperature dependent LSPS. Both the GF
28 and VI were unable to show any kind of phase behavior depending on the solution temperature.
29 The turbidity of the solutions slightly decreases with rise of temperature, probably due to increase
30 of solubility at higher temperature (Fig. 3c and S8). DIC images exhibit presence of clear solution
31 at a range of temperature from 5°C to 65°C (Fig. 3g and S8). As phase separation ability is
32 concentration-dependent, we increase the concentration of peptide solutions by 3 times and still
33 unable to identify any kind of structure formation in the solution. These results clearly unveil the
34 inability of GF and VI to undergo phase separation.
35

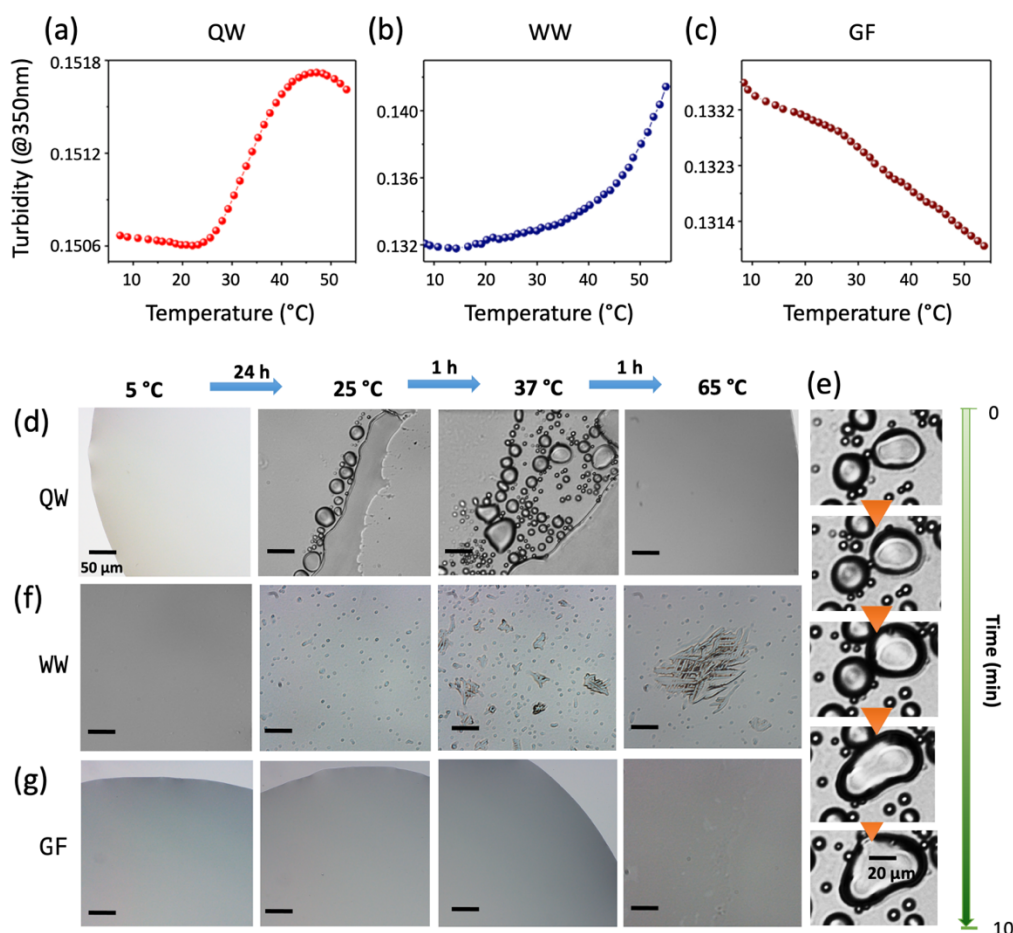
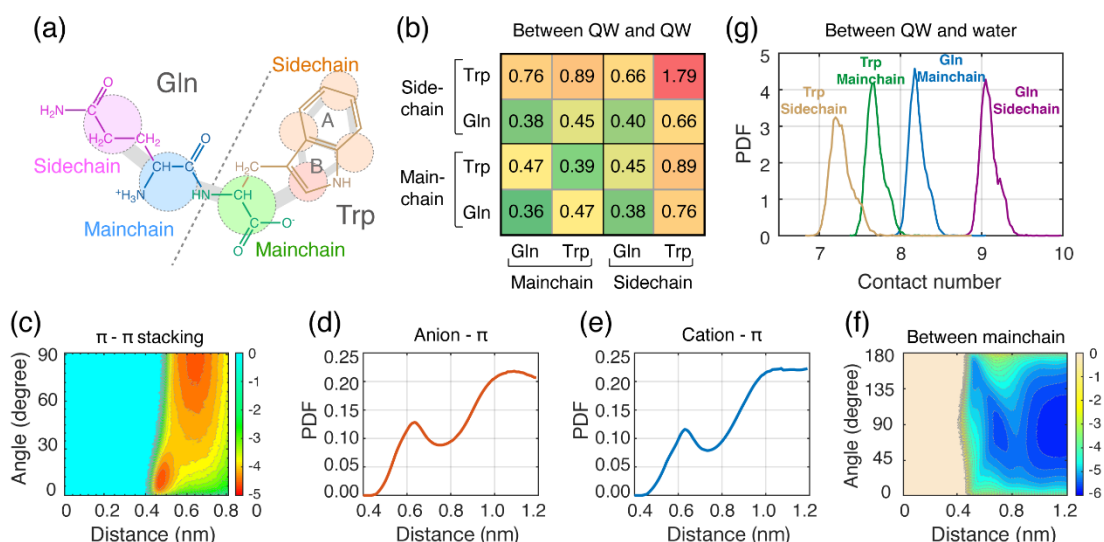


Fig. 3. Temperature dependent phase behavior of QW, WW and GF dipeptides. (a-c) Temperature-dependent changes in turbidity (at 350 nm) of (a) QW, (b) WW, and (c) GF solutions containing 1.5 mM peptide at pH 7.4 adjusted by adding dilute HCl or NaOH (20 mM final concentration). (d-g) Temperature-dependent DIC micrographs: (d) Reversible dispersed monomer and phase-separated liquid droplets formation of QW depending on the temperature of the solution of concentration 3.01 mM at pH 7.4, (e) Time-lapse images showing fusion of two small droplets and formation of a single larger droplet over time. (f) Temperature dependent LSPS behavior of WW. (g) Images of GF solution at different temperatures indicating the inability of phase separation.

The microscopic formation mechanism of QW liquid droplets

We perform interaction analysis to unravel the molecular mechanism underlying the QW LLPS. The QW molecule can be divided into four groups (Fig. 4a). The contact probability of each group pair in Fig. 4b shows that the Trp sidechain-sidechain pair has the highest contact probability (1.79). This interaction is characterized by the free energy landscape (FES) of Trp sidechain aromatic rings. The deep energy minimum basin located at (distance, angle) values of (0.45 nm, 10°) and the shallow basin centered at distance=0.6 nm with a widespread angle distribution (60~90°) suggest strong parallel and T-shaped aromatic stackings (Fig. 4c). The contact number between Trp/Gln mainchain and Trp sidechain (0.89/0.76) is ranked second/third. The probability density functions (PDF) of distances between Trp/Gln mainchain and Trp aromatic rings both exhibit a peak centered

1 at ~ 0.65 nm (Fig. 4d-e), reflecting anion-/cation- π interactions (47). The contact probability
 2 between Gln mainchain and Trp mainchain is also high (0.47), implying hydrogen-bonding
 3 interactions. An active debate exists in literatures regarding whether proteins within liquid-like
 4 condensates form cross- β structures or remain mostly disordered (9, 15, 48–50). We examine the
 5 possible existence of β structures by calculating the FES of mainchain interactions (Fig. 4f) and
 6 observe a deep (0.7 nm, 90°) basin and a much shallower (0.5 nm, 170°) basin, indicating that QW
 7 has certain preference towards forming anti-parallel β structures, albeit with a low propensity. The
 8 solute-solvent interactions are also important for LLPS (8). Single peak is observed in the PDF curves
 9 of contact numbers between water molecules and each group in QW molecule (Fig. 4g). The
 10 ranking of peak values suggests that Gln sidechain has the highest propensity to interact with water,
 11 the Gln and Trp mainchain groups also have solvent-exposure propensity, and the aromatic Trp
 12 sidechains prefer to be buried inside the aggregate.
 13
 14 Taken together, the QW LLPS is a result of the interplay among different non-covalent interactions.
 15 The inter-QW aromatic stacking interaction plays a leading role, and cooperates with the anion-
 16 /cation- π , and hydrogen-bonding interactions to drive the phase separation of QW. Meanwhile,
 17 the QW-water interaction adds to the solubility of QW, thus increase the fluidity of the aggregate.
 18 Eventually, the delicate balance between QW-QW interactions and QW-water interactions leads to
 19 the formation of liquid-like droplets.
 20



21
 22 **Fig. 4. Multivalent interactions governing the LLPS of QW.** (a) Mapping from all-atom model of
 23 QW molecules to the coarse-grained model. The atoms in each dashed circle are mapped to a
 24 single bead. (b) The contact probabilities or each pair of groups (mainchain/sidechain of Gln/Trp)
 25 in QW. (c) The FES as a function of centroid distance and angle between two aromatic rings of
 26 different QW molecules. (d-e) The PDF of centroid distance between Trp aromatic rings and
 27 mainchain bead of (d) Trp or (e) Gln in different QW molecules. (f) The FES as a function of centroid

1 distance and angle between mainchains of different QW molecules. (g) The contact number
2 between each QW group and water molecules.

3

4 **The temperature-dependent phase diagram of QW**

5

6 Many factors including temperature, salt concentration and pH condition have been shown to
7 affect the phase separation of diverse proteins (49, 51). Changes to the phase behavior of
8 dipeptides in response to temperature is of key importance to gain a fundamental understanding
9 of the thermodynamics of their phase separation. According to the phase rule, temperature and
10 the composition of the original homogeneous mixture are sufficient for determining whether
11 phase separation occurs, and for predicting the compositions of coexisting phases (52). However,
12 determining LLPS phase diagram of proteins using standard sampling techniques (such as explicit-
13 solvent MD simulation in cubic box) remains a great challenge due to the computational cost
14 needed to reach thermodynamical equilibrium (29) and the difficulty in calculating the densities of
15 dense/dilute phases. Existing computational approaches proposed as settlements are all based on
16 highly-simplified (patchy particle or one-bead-per-residue) protein models in combination with
17 slab simulation or Gibbs-ensemble simulation (33, 35) which cannot properly describe the LLPS of
18 ultrashort peptides. Compared to proteins, the greatly reduced conformational space of dipeptide
19 monomers enable the determination of phase equilibria for LLPS directly from micro-second MD
20 simulations. In addition, we develop a novel and efficient protocol to determine the densities of
21 both dense phase and dilute phase (see below for more details).

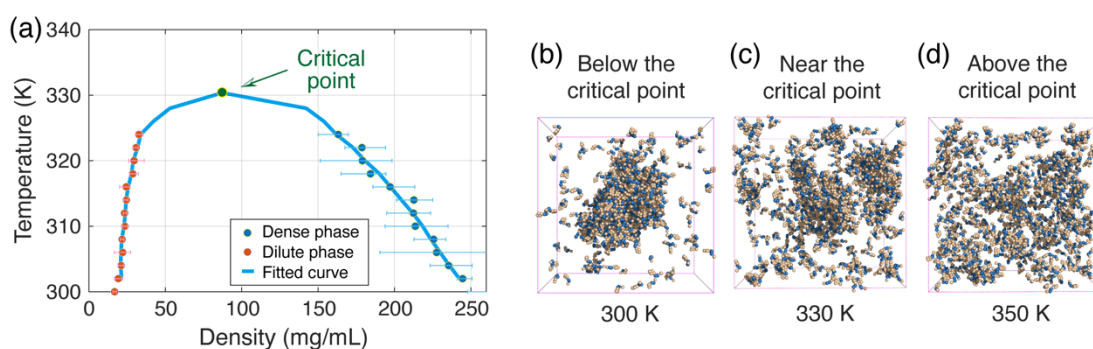
22

23 We evenly choose 26 temperature points ranging from 300 to 350 K and perform three 1.8- μ s
24 simulations at each temperature. For each simulation, we first identify the QW molecules
25 belonging to the dense/dilute phase, calculate the total mass of peptides in the dense/dilute phase
26 by summing up the molecular weight of the QW molecules in that phase, estimate the volume of
27 the dense/dilute phase by a Monte Carlo sampling algorithm (53), and obtain the density of each
28 phase by dividing its corresponding mass of peptides by volume. We then generate the coexistence
29 curve by fitting the density profiles to the critical exponent equation ($\rho_{dense} - \rho_{dilute} =$
30 $A(T_c - T)^\beta$), where β is the critical exponent which is set to 0.365 (universality class of 3D-
31 Heisenberg model). A detailed description to the methodology is given in the Method section and
32 supplementary information.

33

34 As shown in Fig. 5a, the densities of the dilute phase range from 20~50 mg/mL, and those of the
35 dense phase are within 125~250 mg/mL, in agreement with the experimental determined
36 concentrations for several proteins (the dilute/dense phase concentration for Ddx4 is within 0~50
37 and 160~320 mg/mL (48), and the FUS LC domain dense phase concentration is ~120 mg/mL (13)).
38 The concentration of the dilute/dense phase increases/decreases with the rise of temperature and
39 displays a temperature dependent phase behavior described by an upper critical solution

1 temperature (UCST) at ~330 K. QW molecules phase separate into a well-concentrated dense phase
 2 and a dilute phase at temperatures lower than the critical point (Fig. 5b). The phase separation
 3 capability weakens as temperature increases. Around the critical temperature, the QW molecules
 4 form several small loosely-packed aggregates (Fig. 5c). Above the critical temperature, QW totally
 5 loses its phase separation ability (Fig. 5d). The UCST behavior of QW is consistent with a previous
 6 study showing that proteins with abundant polar residues generally follow an UCST phase
 7 separation(54) (the sidechain of Gln, and mainchain of Gln and Trp are all polar, and the sidechain
 8 of Trp also has a polar nitrogen atom).
 9



10

11 **Fig. 5. Temperature-dependent LLPS behavior of QW dipeptide.** (a) The coexistence phase
 12 diagram as a function of temperature and density of the two phases. The critical point is obtained
 13 by fitting the density profiles to the critical exponent equation. (b-d) The final snapshot of a
 14 representative simulation at temperature (b) lower than, (c) near, and (d) higher than the critical
 15 temperature.

16

17 Prediction of the LLPS propensities of all dipeptides

18 We predict the LLPS propensity of dipeptides (excluding the 54 LSPS dipeptides) by comparing their
 19 fluidities with QW. We define the difference between a dipeptide (A) and QW as the Euclidean
 20 distance between the (FE , ER) values of the dipeptide and those of QW:

$$21 \quad d_{A \rightarrow QW} = \sqrt{(FE_A - FE_{QW})^2 + (ER_A - ER_{QW})^2}$$

22 The LLPS propensity of dipeptide A (P_A^{LLPS}) is thus defined using the equation $P_A^{LLPS} = 1 -$
 23 $d_{A \rightarrow QW} / d_{dipeptide \rightarrow QW}^{(max)}$, which is distributed between 0 (lowest LLPS propensity) and 1 (highest

24 LLPS propensity). The predicted LLPS/LSPS propensities (Fig. 6a) are supported by experimental
 25 observations by us and other groups. On one hand, some dipeptides (FF, FK, FW, IF, VF, WF, WW,
 26 and WY) have been experimentally shown to form solid-like aggregates (41). They are all classified
 27 as LSPS dipeptides in our prediction. On the other hand, QW is predicted to be the most probable
 28 LLPS candidate, GF and VI are predicted to have low phase separation ability. The ability of QW and
 29 the inability of GF and VI to form liquid droplet are verified by our experiments (Fig. 3 and Fig. S8).

30

1 From Fig. 6a, which shows the LLPS and LSPS propensities for all dipeptides, we propose that LLPS
2 propensity > 0.7 indicates a good candidate for further experimental investigation. We decipher
3 the sequence determinant of LLPS by studying the sequence characteristics of these high-LLPS-
4 propensity dipeptides. It is apparent that aromatic residues are crucial for both LLPS and LSPS of
5 dipeptides. Dipeptides with high LLPS propensities can be classified into three classes (Fig. 6b): 1)
6 dipeptides composed of Y (an aromatic residue with polarity) and a hydrophobic residue L/I/V/C,
7 2) dipeptides composed of W (an aromatic residue with high hydrophobicity) and a polar residue
8 A/G/Q/N (Fig. 6b), and 3) dipeptides with large dipole moment (KE/KD/RE/RD). Class 1 and 2 show
9 that the LLPS of dipeptides requires a cooperation among aromatic stacking, hydrophobic and polar
10 interactions. The crucial role of strong aromatic stacking interactions can be evidenced by the
11 inability of all non-aromatic hydrophobic dipeptides to undergo LLPS, and by the lower LLPS
12 propensities of F/H-containing dipeptides than Y/W-containing dipeptides (the π electron densities
13 of F/H are lower than those of Y/W). The indispensability of hydrophobic and polar interactions for
14 LLPS can be respectively shown by the inability of dipeptides composed of Y and A/G/Q/N residue
15 to undergo LLPS, and the LSPS capability of dipeptides composed of W and L/I/V/C. Dipeptides in
16 class 3 possess the largest dipole moment (both the mainchain and sidechain beads of the N-/C-
17 terminal residues are positively-/negatively-charged), revealing that strong dipole-dipole
18 interaction promotes LLPS. We notice that some dipeptides containing S and/or T residues show
19 abnormally high LLPS/LSPS propensities compared to dipeptides containing A/G/Q/N residues.
20 This is probably an artifact of force field, which needs to be investigated in future studies (the
21 amber99sb/amber03ws force field parameters of S/T residues have been shown to lead to
22 significant deviation of computational-calculated chemical shifts from experimental values (55)).

23

24 Taken together, our predictions show that LLPS of dipeptides is driven by a cooperation among
25 different types of physical interactions provided by aromatic, hydrophobic, and polar components.
26 Compared to literature-reported proteins with LLPS capability, dipeptides have extremely short
27 amino acid sequence lengths, but still encompass different non-covalent interaction sites, enabling
28 them to undergo LLPS.

29

30

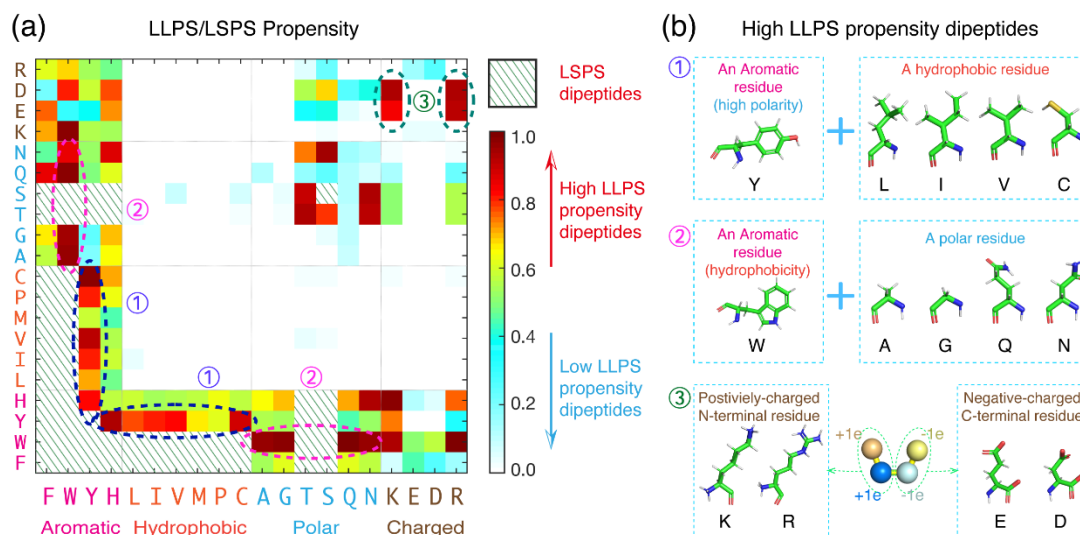


Fig. 6. Predicted phase separation propensities of all 400 dipeptides. (a) The LSPS capability or LLPS propensity of all 400 dipeptides. The horizontal and vertical axes correspond respectively to the N- and C-terminal residue. The dipeptides with LSPS capability are represented by slash lines, while other dipeptides are colored from white to red in increasing order of LLPS propensity. The three categories of high-LLPS-propensity dipeptides are highlighted in dashed circles. (b) The illustrations to the sequence characteristics of the three categories of high-LLPS-propensity dipeptides.

Discussion

On the basis of microsecond explicit-solvent MD simulations in combination with multi-bead-per-residue coarse-grained Martini model, we comprehensively explore the LLPS abilities of all 400 dipeptide combinations of coded amino acids. A similar simulation protocol has been used by Tuttle and coworkers to predict the self-assembly propensity of dipeptides/tripeptides, albeit with a short simulation time scale. They find that aromatic dipeptides have high self-assembly propensities. Nevertheless, the LLPS ability of dipeptides and their phase behaviors remain completely elusive. We have developed a novel, efficient approach to predict the LLPS propensities of ultrashort peptides which consists of two steps: (1) determine the aggregation capability of a peptide, and (2) evaluate the fluidity of the peptide aggregates to distinguish between LLPS and LSPS. Tuttle and co-workers define the AP parameter to characterize the aggregation capability of dipeptides(41). The AP parameter, defined as the fraction of decrease in SASA, which reflects the compactness of a dipeptide aggregate, is a good quantity to screen out dipeptides with high self-assembly ability within short simulation time-scales, but is highly degenerative for low-AP dipeptides. In order to quantify aggregation capabilities of both high-AP and low-AP dipeptides, we perform micro-second-long simulations and use two parameters by introducing a new parameter, i.e., the clustering degree. On the other hand, computational characterization of the

1 microscopic fluidity of peptides is an unexplored area. Based on physical principles on state of
2 matters (44), we define the exchange rate of interactions and the fluctuation extent of clustering
3 degree parameters to effectively evaluate the fluidity of the peptide aggregates.

4

5 One of the most important progresses of this work is the identification of spontaneous LLPS in
6 dipeptide systems. We have identified 54 dipeptides with LSPS capability, and predicted the LLPS
7 propensity of the other 346 dipeptides, among which QW possess the highest LLPS propensity. Our
8 predictions are validated by temperature dependent turbidity assays and DIC microscopy on four
9 representative dipeptides (WW with LSPS capacity, and QW, GF, and VI with high, medium, and low
10 LLPS propensities, respectively). The identified QW peptide is by far the simplest LLPS system. The
11 computational observation of dipeptide LLPS in this work is mainly attributed to the use of multi-
12 bead-per-residue force-field and explicit solvent model, which explicitly captures key interactions
13 (such as aromatic stacking and peptide-water interactions) driving LLPS that are missing in one-
14 bead-per-residue models.

15

16 Another important progress of this work is the success in constructing the LLPS phase diagram
17 using non-slab simulations. We determine the phase diagram of QW which displays a temperature
18 dependent phase behavior described by an UCST of ~ 330 K. Our constructed phase diagram can
19 better describe the temperature dependence of the densities of both dense and dilute phases
20 mostly due to the use of explicit solvents. The protocol provided in this work can be easily applied
21 to other short peptide systems, thus offers a way of exploring temperature dependent phase
22 behaviors of peptides. We note that a limitation of Martini force-field is the overestimation of
23 water freezing point (290 ± 5 K) (56), thus this force-field cannot properly describe phase behavior
24 at temperature around or lower than 290 K. As a result, for peptides which display a lower critical
25 solution temperature (LCST) behavior, a more refined water model (such as polarizable Martini
26 water) is needed. However, this will greatly increase computational costs.

27

28 In summary, our systematic investigation suggests that dipeptides, minimal but complete, possess
29 multivalent interactions sufficient for LLPS. Three distinct categories of dipeptides are identified to
30 have high LLPS propensities. Our results reveal that aromatic residues are crucial for a dipeptide to
31 have LLPS capacity, and hydrophobic and polar compartments are indispensable. Our findings,
32 together with previous reports showing the LLPS capacity of various IDPs or IDRs with varying
33 sequence length (57–59), suggest that LLPS is a general property of peptides/proteins,
34 independent of their length. Our protocols, which efficiently and effectively predict the LLPS
35 propensity of dipeptides, provides a comprehensive framework for predicting and characterizing
36 the phase behaviors of minimalistic peptides. This study set the basis for the prediction of phase
37 behavior at the sequence level and in-depth understanding molecular mechanisms of protein
38 phase separation.

39

1

2 **Materials and methods**

3 **Molecular dynamics simulations**

4 All dipeptide molecules are modeled using the MARTINI coarse-grained protein model (version 2.1)
5 with positive and negative charges assigned to main-chain beads at N and C terminal, respectively.
6 In the initial state of each MD simulation, 720 dipeptide molecules are randomly displaced in
7 aqueous solution containing 40,000 water beads. For dipeptides with net charges, adequate
8 number of Na⁺ or Cl⁻ ions are randomly added to the solution to neutralize the system. The details
9 of each simulation and the analysis protocols are given in the supplementary information.
10 Graphical analysis are preformed using the PyMOL package (60).

11

12 **LLPS propensity determination**

13 Two dipeptide molecules are considered to be in contact if their minimum distance is within 0.7
14 nm. The dense phase is defined as the maximized subset of all dipeptide molecules in the system
15 where each molecule forms contact with another molecule in this subset. The aggregation
16 propensity of each dipeptide is modeled using clustering degree and collapse degree parameters,
17 and the fluidity of the aggregate is modeled using two parameters: the fluctuation of cluster size
18 and the exchange rate of interactions. Definition and calculation protocols of these four parameters
19 are given in the supplementary information. All dipeptides with low aggregation propensity are
20 grouped into three clusters using a k-means clustering algorithm. The LLPS propensity of each of
21 these dipeptides are defined as the additive inverse of the Euclidean distance between the
22 (fluctuation of cluster size, exchange rate of interactions) values of this dipeptide and those of QW.
23 The LLPS propensity is scaled between 0 (lowest LLPS propensity) and 1 (highest LLPS propensity).

24

25 **Phase diagram construction**

26 Three independent 1.8- μ s simulations are performed at each of the 26 temperature points ranging
27 from 300 to 350 K. The dense phase is defined as the maximized subset of all dipeptide molecules
28 in the system where each molecule forms contact with another molecule in this subset. The mass
29 of the dense phase is calculated by summing up the molecular weight of the QW molecules in that
30 phase. The volume of the dense phase is calculated by a refined Monte Carlo sampling algorithm
31 (see details in supplementary information). The mass and volume of the dilute phase is calculated
32 by subtracting the mass and volume of the total system from those of the dense phase. The critical
33 temperature T_c is obtained using a similar method to the work by Mittal and coworkers (34). The
34 densities of the dense and dilute phase (ρ_{dilute} and ρ_{dense}) as functions of temperature are fitted to
35 the equation $\rho_{\text{dense}} - \rho_{\text{dilute}} = A(T_c - T)^\beta$, where β is the critical exponent which is set to 0.365
36 (universality class of 3D-Heisenberg model). The minimum fitting temperature is chosen as the
37 physiological temperature (310 K), and the maximum fitting temperature (326 K) is chosen to
38 maximize the coefficient of determination R^2 . The density at the critical temperature is defined as
39 the average peptide density of the simulation at that temperature.

1

2 **Preparation of peptide solutions**

3 Peptide solutions are prepared by dissolving the corresponding peptide at required concentration
4 (1.5 mM or 3.01 mM or 9 mM) in double distilled water at pH 7.4 adjusted by drop wise addition
5 of 0.1 M HCl or NaOH to a final concentration of 20 mM. The samples are first vortexed for 2 mins,
6 followed by sonication until the peptides dissolved completely.

7

8 **Turbidity analysis and DIC microscopy**

9 Turbidity analysis for the dipeptides is conducted by preparing solutions at concentrations of 1.5
10 mM in DD water at pH 7.4. Then, 200 μ L aliquots are pipetted into a 96-well plate, sealed using a
11 Breathe-Easy sealing membrane (Sigma Aldrich, Rehovot, Israel), and absorbance at 350 nm is
12 measured. All measurements are performed using a Synergy HT plate reader (Biotek, Winooski, VT,
13 USA) at temperature range 5°C-55°C. The heating or cooling is performed at a rate of 1 °C min⁻¹.
14 Temperature-dependent phase behavior of dipeptide samples is monitored by DIC microscopy. An
15 aliquot of 15 μ L peptide solutions at concentration 3.01 mM or 9 mM from said temperature are
16 loaded onto glass coverslips and DIC images are acquired on a LSM 510 META (Zeiss) microscope
17 with a x100 objective (oil immersion). Images are processed using ImageJ.

18

19 **Acknowledgements**

20 This work has been supported by the National Natural Science Foundation of China (Grant Nos:
21 12074079 and 11674065), the National Key Research and Development Program of China (Grant
22 Nos: 2016YFA0501702) and Israel Science Foundation (Grant No. 3145/19 to E.G). All simulations
23 were performed using the GPU High Performance Cluster at Fudan University.

24

25 **Author Contributions**

26 Y. T., Y. Y., and G.W. contributed to the conception of the study; Y. T. and Y. Y. performed the
27 molecular dynamics simulations; Y. T., Y. Y., J. Z., and G.W. contributed to the data analysis and
28 manuscript preparation. Z. L. and X. D. helped perform the analysis with constructive discussions.
29 S. B. and E. G. performed the turbidity assay and DIC experiments. Y. T., S. B., G. W., and E. G. co-
30 wrote the manuscript. All authors commented on the manuscript.

31

32 **Competing interests**

33 The authors declare no competing interests.

34

35 **References**

- 36 1. A. S. Holehouse, R. V. Pappu, Protein polymers: Encoding phase transitions. *Nat. Mater.* **14**,
37 1083–1084 (2015).
38 2. C. P. Brangwynne, P. Tompa, R. V. Pappu, Polymer physics of intracellular phase transitions.
39 *Nat. Phys.* **11**, 899–904 (2015).

- 1 3. B. D. Knapp, K. C. Huang, Translating the Physical Code of Life. *Cell* **174**, 253–255 (2018).
- 2 4. E. Gomes, J. Shorter, The molecular language of membraneless organelles. *J. Biol. Chem.* **294**,
- 3 7115–7127 (2019).
- 4 5. C. P. Brangwynne, *et al.*, Germline P granules are liquid droplets that localize by controlled
- 5 dissolution/condensation. *Science*. **324**, 1729–1732 (2009).
- 6 6. T. M. Franzmann, S. Alberti, Protein phase separation as a stress survival strategy. *Cold Spring*
- 7 *Harb. Perspect. Med.* **9**, 1–18 (2019).
- 8 7. S. Jiang, J. B. Fagman, C. Chen, S. Alberti, B. Liu, Protein phase separation and its role in
- 9 tumorigenesis. *Elife* **9**, 1–27 (2020).
- 10 8. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: Organizers of
- 11 cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
- 12 9. A. Patel, *et al.*, A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by
- 13 Disease Mutation. *Cell* **162**, 1066–1077 (2015).
- 14 10. X. Ding, *et al.*, Amyloid-Forming Segment Induces Aggregation of FUS-LC Domain from Phase
- 15 Separation Modulated by Site-Specific Phosphorylation. *J. Mol. Biol.* **432**, 467–483 (2020).
- 16 11. Y. Baradaran-Heravi, C. Van Broeckhoven, J. van der Zee, Stress granule mediated protein
- 17 aggregation and underlying gene defects in the FTD-ALS spectrum. *Neurobiol. Dis.* **134**,
- 18 104639 (2020).
- 19 12. S. Ray, *et al.*, α -Synuclein aggregation nucleates through liquid–liquid phase separation. *Nat.*
- 20 *Chem.* **12**, 705–716 (2020).
- 21 13. V. H. Ryan, *et al.*, Mechanistic View of hnRNPA2 Low-Complexity Domain Structure,
- 22 Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. *Mol. Cell*
- 23 **69**, 465-479.e7 (2018).
- 24 14. A. Prasad, V. Bharathi, V. Sivalingam, A. Girdhar, B. K. Patel, Molecular mechanisms of TDP-43
- 25 misfolding and pathology in amyotrophic lateral sclerosis. *Front. Mol. Neurosci.* **12**, 1–36
- 26 (2019).
- 27 15. S. Wegmann, *et al.*, Tau protein liquid–liquid phase separation can initiate tau aggregation.
- 28 *EMBO J.* **37**, 1–21 (2018).
- 29 16. Y. Lin, Y. Fichou, Z. Zeng, N. Hu, S. Han, Liquid-liquid phase separation and fibrillization of tau
- 30 are independent processes with overlapping conditions. *ACS Chem. Neurosci.*, 1–31 (2019).
- 31 17. T. Ukmar-Godec, S. Wegmann, M. Zweckstetter, Biomolecular condensation of the
- 32 microtubule-associated protein tau. *Semin. Cell Dev. Biol.* **99**, 202–214 (2020).
- 33 18. F. Luo, *et al.*, Atomic structures of FUS LC domain segments reveal bases for reversible
- 34 amyloid fibril formation. *Nat. Struct. Mol. Biol.* **25**, 341–346 (2018).
- 35 19. M. Hofweber, *et al.*, Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor
- 36 and Arginine Methylation. *Cell* **173**, 706-719.e13 (2018).
- 37 20. S. Maharana, *et al.*, RNA buffers the phase separation behavior of prion-like RNA binding
- 38 proteins. *Science*. **360**, 918–921 (2018).
- 39 21. S. Qamar, *et al.*, FUS Phase Separation Is Modulated by a Molecular Chaperone and
- 40 Methylation of Arginine Cation- π Interactions. *Cell* **173**, 720-734.e15 (2018).
- 41 22. F. G. Quiroz, A. Chilkoti, Sequence heuristics to encode phase behaviour in intrinsically
- 42 disordered protein polymers. *Nat. Mater.* **14**, 1164–1171 (2015).
- 43 23. S. Boeynaems, *et al.*, Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule
- 44 Dynamics. *Mol. Cell* **65**, 1044-1055.e5 (2017).

- 1 24. H. Jafarinia, E. van der Giessen, P. R. Onck, Phase Separation of Toxic Dipeptide Repeat
2 Proteins Related to C9orf72 ALS/FTD. *Biophys. J.* **119**, 843–851 (2020).
- 3 25. M. Reches, E. Gazit, Casting metal nanowires within discrete self-assembled peptide
4 nanotubes. *Science*. **300**, 625–627 (2003).
- 5 26. C. Guo, Y. Luo, R. Zhou, G. Wei, Probing the self-assembly mechanism of diphenylalanine-
6 based peptide nanovesicles and nanotubes. *ACS Nano* **6**, 3907–3918 (2012).
- 7 27. N. S. De Groot, T. Parella, F. X. Aviles, J. Vendrell, S. Ventura, Ile-Phe dipeptide self-assembly:
8 Clues to amyloid formation. *Biophys. J.* **92**, 1732–1741 (2007).
- 9 28. K. M. Ruff, R. V. Pappu, A. S. Holehouse, Conformational preferences and phase behavior of
10 intrinsically disordered low complexity sequences: insights from multiscale simulations. *Curr.*
11 *Opin. Struct. Biol.* **56**, 1–10 (2019).
- 12 29. J. E. Shea, R. B. Best, J. Mittal, Physics-based computational and theoretical approaches to
13 intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **67**, 219–225 (2021).
- 14 30. Y. H. Lin, J. D. Forman-Kay, H. S. Chan, Sequence-Specific Polyampholyte Phase Separation in
15 Membraneless Organelles. *Phys. Rev. Lett.* **117**, 1–6 (2016).
- 16 31. J. McCarty, K. T. Delaney, S. P. O. Danielsen, G. H. Fredrickson, J. E. Shea, Complete Phase
17 Diagram for Liquid-Liquid Phase Separation of Intrinsically Disordered Proteins. *J. Phys. Chem.*
18 *Lett.* **10**, 1644–1652 (2019).
- 19 32. S. Qin, H. X. Zhou, Fast Method for Computing Chemical Potentials and Liquid-Liquid Phase
20 Equilibria of Macromolecular Solutions. *J. Phys. Chem. B* **120**, 8164–8174 (2016).
- 21 33. A. Ghosh, K. Mazarakos, H. X. Zhou, Three archetypical classes of macromolecular regulators
22 of protein liquid-liquid phase separation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19474–19483
23 (2019).
- 24 34. G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, J. Mittal, Sequence determinants of protein phase
25 behavior from a coarse-grained model. *PLoS Comput. Biol.* **14**, 1–23 (2018).
- 26 35. G. L. Dignon, W. Zheng, J. Mittal, Simulation methods for liquid–liquid phase separation of
27 disordered proteins. *Curr. Opin. Chem. Eng.* **23**, 92–98 (2019).
- 28 36. G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, J. Mittal, Relation between single-molecule
29 properties and phase behavior of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S.*
30 *A.* **115**, 9929–9934 (2018).
- 31 37. W. T. Chu, Wang, Thermodynamic and sequential characteristics of phase separation and
32 droplet formation for an intrinsically disordered region/protein ensemble. *PLoS Comput. Biol.*
33 **17**, 1–20 (2021).
- 34 38. K. M. Ruff, R. V. Pappu, A. S. Holehouse, Conformational preferences and phase behavior of
35 intrinsically disordered low complexity sequences: insights from multiscale simulations. *Curr.*
36 *Opin. Struct. Biol.* **56**, 1–10 (2019).
- 37 39. Y. H. Lin, J. D. Forman-Kay, H. S. Chan, Theories for Sequence-Dependent Phase Behaviors of
38 Biomolecular Condensates. *Biochemistry* **57**, 2499–2508 (2018).
- 39 40. E. W. Martin, T. Mittag, Relationship of Sequence and Phase Separation in Protein Low-
40 Complexity Regions. *Biochemistry* **57**, 2478–2487 (2018).
- 41 41. P. W. J. M. Frederix, R. V. Uljijn, N. T. Hunt, T. Tuttle, Virtual screening for dipeptide
42 aggregation: Toward predictive tools for peptide self-Assembly. *J. Phys. Chem. Lett.* **2**, 2380–
43 2384 (2011).
- 44 42. X. Yan, *et al.*, Reversible transitions between peptide nanotubes and vesicle-like structures

- 1 including theoretical modeling studies. *Chem. A Eur. J.* **14**, 5974–5980 (2008).
- 2 43. P. W. J. M. Frederix, *et al.*, Exploring the sequence space for (tri-)peptide self-assembly to
3 design and discover new hydrogels. *Nat. Chem.* **7**, 30–37 (2015).
- 4 44. A. A. Hyman, C. A. Weber, F. Jülicher, Liquid-liquid phase separation in biology. *Annu. Rev. Cell*
5 *Dev. Biol.* **30**, 39–58 (2014).
- 6 45. K. Krishna, M. N. Murty, Genetic K-means algorithm. *IEEE Trans. Syst. Man, Cybern. Part B*
7 *Cybern.* **29**, 433–439 (1999).
- 8 46. S. Ambadipudi, J. Biernat, D. Riedel, E. Mandelkow, M. Zweckstetter, Liquid-liquid phase
9 separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nat.*
10 *Commun.* **8**, 1–13 (2017).
- 11 47. F. N. R. Petersen, M. Jensen, C. H. Nielsen, Interfacial tryptophan residues: A role for the
12 cation- π effect? *Biophys. J.* **89**, 3985–3996 (2005).
- 13 48. J. P. Brady, *et al.*, Structural and hydrodynamic properties of an intrinsically disordered region
14 of a germ cell-specific protein on phase separation. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E8194–
15 E8203 (2017).
- 16 49. A. C. Murthy, *et al.*, Molecular interactions underlying liquid-liquid phase separation of the
17 FUS low-complexity domain. *Nat. Struct. Mol. Biol.* **26**, 637–648 (2019).
- 18 50. I. Peran, T. Mittag, Molecular structure in biomolecular condensates. *Curr. Opin. Struct. Biol.*
19 **60**, 17–26 (2020).
- 20 51. G. L. Dignon, R. B. Best, J. Mittal, Biomolecular phase separation: From molecular driving
21 forces to macroscopic properties. *Annu. Rev. Phys. Chem.* **71**, 53–75 (2020).
- 22 52. H. Falahati, A. Haji-Akbari, Thermodynamically driven assemblies and liquid-liquid phase
23 separations in biology. *Soft Matter* **15**, 1135–1154 (2019).
- 24 53. C. R. Salama, GPU-based Monte-Carlo volume raycasting in *Proceedings - Pacific Conference*
25 *on Computer Graphics and Applications*, (2007) <https://doi.org/10.1109/PG.2007.33>.
- 26 54. G. L. Dignon, W. Zheng, Y. C. Kim, J. Mittal, Temperature-Controlled Liquid-Liquid Phase
27 Separation of Disordered Proteins. *ACS Cent. Sci.* **5**, 821–830 (2019).
- 28 55. W. S. Tang, N. L. Fawzi, J. Mittal, Refining All-Atom Protein Force Fields for Polar-Rich, Prion-
29 like, Low-Complexity Intrinsically Disordered Proteins. *J. Phys. Chem. B* **124**, 9505–9512
30 (2020).
- 31 56. S. O. Yesylevskyy, L. V. Schäfer, D. Sengupta, S. J. Marrink, Polarizable water model for the
32 coarse-grained MARTINI force field. *PLoS Comput. Biol.* **6**, 1–17 (2010).
- 33 57. S. Elbaum-Garfinkle, Matter over mind: Liquid phase separation and neurodegeneration. *J.*
34 *Biol. Chem.* **294**, 7160–7168 (2019).
- 35 58. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science.*
36 **357** (2017).
- 37 59. M. Hardenberg, A. Horvath, V. Ambrus, M. Fuxreiter, M. Vendruscolo, Widespread occurrence
38 of the droplet state of proteins in the human proteome. *Proc. Natl. Acad. Sci. U. S. A.* **117**,
39 33254–33262 (2021).
- 40 60. W. L. DeLano, The PyMOL Molecular Graphics System, Version 2.3. *Schrödinger LLC* (2020).
- 41