Relationship between liquid-liquid phase separation and epigenetic modification based on the Boltzmann distribution

Fei Xiong[1*]

[1] Department of Gastroenterology, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Chengdu, China. * feiji19841984@gmail.com

Abstract

Liquid-liquid phase separation is pervasive in the interior of cells. Elucidating the mechanism of its formation is important both for understanding its physiological function and for predicting prospective drug targets. According to the Boltzmann distribution, we describe a physical relationship between liquid-liquid phase separation and epigenetic modification, suggesting that biochemical metabolism may play an important role in phase separation. It also suggests that liquid-liquid phase separation is universal in the origins of early life because alkaline hydrothermal vents are full of biochemical reactions.

Keywords: Liquid-liquid phase separation, epigenetic modification, Boltzmann distribution, entropy, hydrothermal vent.

The frequent occurrence of liquid-liquid phase separation (LLPS)in cell biology has been documented since Cliff Brangwynne identified P granules in *C. elegans* cells as liquid droplets[1]. There are numerous studies suggesting that LLPS has a vast impact on cells, influencing processes such as chromatin remodeling [2], transcriptional control [3],autophagy [4], and innate immunity [5], but the mechanism of LLPS formation remains unknown. Phase separation is in essence a physics concept [6], so the next logical step would be to explain the origin of LLPS with the help of thermodynamic concepts and mathematical tools.

Recently, based on the work of Cliff Brangwynne's team[2], the free energy cost of droplet nucleation has been expressed by the following formula:

$$\Delta F(R) = 4\pi\gamma R^2 - \frac{4}{3}\pi R^3\left(\Delta\mu \cdot c_{drop} - \frac{5}{6}G\right) \quad (1)$$

Please see the publication by Cliff Brangwynne team's [2] for the meaning of each mathematical symbol. LLPS is caused by the resistance of surface tension, the cohesive energy of a molecule, and the stiffness of euchromatin and heterochromatin based on Eq. (1). It is by definition two coexisting phases in equilibrium. Let us define protein and RNA as particles in the cell, to simplify these particles with

varying densities into a single density. Eq. (2) is given for LLPS by the Boltzmann distribution [7]:

$$X_2 = X_1 \exp\left[-\left(\Delta\mu \cdot c_{drop} + mg\Delta z + e\Delta\psi\right)/KT\right] \quad (2)$$

Where X_1 is the concentration of water molecules before droplet formation, X_2 is the concentration of water molecules in the droplet, m is the mass of a water molecule, g is the gravitational acceleration, z is altitude, e is charge, and $\Delta\psi$ is electrical potential.

The chemical potential of these particles inside droplets $\Delta u \cdot c_{drop}$ can be given as

$$\Delta\mu \cdot c_{drop} = -mg\Delta z - e\Delta\psi + KT\log(X_1 - X_2) \quad (3)$$

Thus, Eq. (1) can be written as

$$\Delta F(R) = 4\pi\gamma R^2 - \frac{4}{3}\pi R^3\left(-mg\Delta z - e\Delta\psi + KT\log(X_1 - X_2) - \frac{5}{6}G\right) \quad (4)$$

Let us suppose that the droplet is in equilibrium with liquid outside the droplet, and the surface tension and size are therefore constant. Let us consider gravity first. It is well known that gravity has a great influence on the spatial distribution of droplets [8]. At the macro level, the NASA twin study also hints at some impacts of microgravity on life forms [9]. However, for water molecules, we have

$$mg = \rho g V \quad (5)$$

Where ρ is the fluid density, and V is the particle volume.

Returning to Eq. (1), although the spatial distribution of large particles is influenced by gravity, we can ignore Δz because the height of the cell is negligible relative to that of the atmosphere[7], and the density of the droplet is far greater than that of water molecules under static conditions. In short, the influence of gravity can be ignored for water molecules.

According to thermodynamics, the motion of water molecules is a random walk-like jiggling motion [10]. At the moment of LLPS formation, it is certain that the

motion of water molecules has directionality on the droplet's surface (Fig.1). The most logical way to extrapolate is to find a force that acts upon water molecules in the droplet based on the second law of thermodynamics.

Intermolecular forces can be classified into three kinds according to their different origins. Let us consider quantum forces first, such as the forces that give rise to covalent or chemical bonding. If water molecules form covalent bonds in the droplet, a solid phase forms. We can remove this possibility from consideration because the solid phase is a pathological state [11,12].A second consideration is that some forces have a purely entropic origin [7]. Let us assume that the temperature is constant. According to Eq. (4), "$X_1 - X_2$"suggests an osmotic force acting on phase separation. Again, the liquid is at the point of LLPS formation. The second law of thermodynamics tells us that when an osmotic force acts on a system, the particles reach a uniform distribution at the end. It is impossible that the directional movement of water molecules originating from the relevant forces has a purely entropic origin, so we can rule out this possibility. Electrostatic force [7] is the only choice based on the term "$-e\Delta\psi$". In addition, "G" suggests that elastic force also plays an important role in our system. Studies in epigenetic modification have shown that some types of histone tail posttranslational modification can affect chromatin's electric potential [13]. The flexibility of chromatin structure can also be affected by epigenetic modification [14,15]. This combination allows us to associate epigenetic modification with LLPS. Hence, it is possible that the relationship exists between them is based on the definition of entropy.

According to the first law of thermodynamics [7], for an isolated system, entropy is constant. Let us suppose that time approaches zero, which allows the assumption that a cell is an isolated system because it does not exchange with their environment in the instantaneous state. The entropy of the cell is constant. On the basis of the second law of thermodynamics[7], the entropy of the droplet decreases because of the directed movement of water molecules. Therefore, we have

$$\sum_{k=1}^{n} S_{external-k} = \sum_{k}^{n}(S_{total-k} - S_{droplet-k}) \quad (6)$$

Where n is the number of instantaneous states, $S_{external-k}$ is the entropy outside the droplet in a certain instantaneous state, $S_{total-k}$ is the entropy of the total system in the instantaneous state, and $S_{droplet-k}$ is the entropy of the droplet in the instantaneous state.

According to the study by Cliff Brangwynne et al[8], the nuclear F-actin scaffold can protect against the force of gravity acting on the larger size of the droplet. Since larger droplets tend to merge together, it is logical that the scaffold can also stop the horizontal motion of these droplets. The effect of gravity on the cell [8] suggests that these droplets are suspended at different heights because of their different sizes. As previously described, we consider the random motion of water molecules only on the horizontal plane, ignoringΔz. The logical conclusion is the presence of isolated and closed focal areas caused by the nuclear F-actin scaffold, so we have

$$S_{external-n} = S_{total-n} - S_{droplet-n} \quad (7)$$

Where $S_{external-n}$ is the entropy outside the droplets in the closed focal area , $S_{total-n}$ is the entropy of the total system in the closed focal area, and $S_{droplet-n}$ is the entropy of the droplets in the closed focal area.

According to the work of Feinberg et al, entropy increases with both hypermethylation in the CpG island and hypomethylation in the shore[16,17]. The entropy in epigenetics is known as Shannon entropy. This presents a paradox: if p is the probability of DNA methylation at a certain site, it is logical that hypomethylation in the shore corresponds to low entropy. It is possible that information theory cannot explain the problem; thus, the paradox may need to be explained by the essence of entropy: a measure of uncertainty.

It is well known that the occurrence of signaling molecules depends on molecular collision. From the information theory, Shannon entropy is related to the probability of molecular collision at a particular site. Hence, increasing Shannon entropy decreases the probability of molecular collision. This decrease in probability increases the uncertainty of the motion direction selection of a single molecule. In fact, the motion direction selection is equivalent to a possible microscopic configuration for a

single molecule. Returning to thermodynamics, based on a statistical ensemble, we have an equivalence relation between Shannon entropy and thermodynamic entropy:

$$\Omega \Leftrightarrow r_{direction} \Rightarrow \Delta S_{thermo} \Leftrightarrow \Delta S_{shannon} \quad (8)$$

Where Ω is the number of possible microscopic configurations , $r_{direction}$ is the number of possible motion direction selections of a single molecule, ΔS_{thermo} is the thermodynamic entropy, and $\Delta S_{shannon}$ is the information entropy.

As described above, “hypomethylated shores” and “hypermethylated CpG islands” may show higher Shannon entropy than other regions of the genome[16,17]. DNA methylation results in some conformational changes in DNA. When DNA methylation reaches a certain level, it turns B-DNA into Z-DNA[18]. It is clear that Z-DNA decreases the probability of molecular collision at a particular site in the gene sequence because the disappearing major groove and the narrow and deep minor groove[19] inhibit the binding of transcription factors to gene elements. From the information theory perspective, this can explain higher entropy in “hypermethylated CpG islands” but not in “hypomethylated shores”. The distance between CpG islands and the CpG shore is approximately 2 kb[20]. It is known that a nucleosome core region and its linear region contain approximately 200 nucleotide pairs[21]. A corollary of this characteristic is that CpG islands are close to the CpG shore in space because the DNA in a nucleosome forms a squat disc-like structure approximately 5.5 nm in height and 11 nm in diameter[22]. Let us suppose that the squat disc-like structure is spherical with a diameter of 11 nm and that the nucleus of a simple columnar cell is 6 μm[28].Thus, we have

$$n_{ratio} = \frac{\frac{4}{3}\pi \frac{2kb}{0.2kb}\left(\frac{d_{DNA}}{2}\right)^3}{\frac{4}{3}\pi\left(\frac{d_{nucleus}}{2}\right)^3} = 0.00000006 \quad (9)$$

Where d_{DNA} is the diameter of the squat disc-like structure, and $d_{nucleus}$ is the diameter of the nucleus of a simple columnar cell.

From the ratio reaching one in one hundred-million , we can conclude that CpG islands are near the CpG shore in space. In general, it is known that DNA methylation can increase the rigidity of DNA[23]. Returning to f Eq. (1), increasing rigidity

increases the free energy cost of droplet nucleation, so phase separation is impossible on hypermethylated CpG islands[1]. This suggests that hypermethylated CpG islands inhibit DNA-protein intercalation by the mechanical signal of rigidity[24]. On hypomethylated shores, $\frac{5}{6}G$ is too low to increase the rigidity. Instead, a small quantity of DNA methylation and other epigenetic modifications can increase $e\Delta\psi$. Clearly, this suggests that phase separation tends to aggregate in hypomethylated shores. According to Eq. (7), the uncertainty of the motion direction selection of single molecules outside the droplets increases because of the first law of thermodynamics. Hence, hypomethylated shores silence gene expression because the uncertainty of the motion direction selection reduces the collision probability of molecules at any particular locus of the closed focal area. On the basis of Eq. [7],the characteristics of CpG islands near CpG shores suggest that DNA methylation silences gene expression via two possible mechanisms: 1. DNA methylation changes the physical properties of two DNA strands; 2. LLPS leads to a change in the collision probability of molecules. In other words, the intrinsic connection between epigenetic modification and LLPS can explain why hypomethylation in shore regions corresponds to high entropy.

In recent years, induced pluripotent stem cell research [25] has revealed a relationship between metabolic pathway changes and epigenetic remodeling. Changes in intermediary metabolites can regulate epigenetic modification[26], which has a vast impact on LLPS based on the Boltzmann distribution. According to the work of Michael J Russell et al[27],the hydrothermal vents of the early earth might have been full of metabolic biochemical reactions before the formation of RNA and protein. Returning to Eq. (4), the "$\frac{5}{6}G$" and "$e\Delta\psi$" values of amino acids, pyridines, and purines may be changed by certain metabolic reactions. It is possible that the accumulation of amino acids, pyrimidine and purine accompanies LLPS. The droplets can provide a stable environment for forming RNA and protein because droplet formation is a process of decreasing entropy based on Eq. (8). According to the hydrothermal vent theory[27], such vents were full of biochemical reactions that contributed to the origins of early life. Returning to Eq. (4), the relationship between

LLPS and epigenetic modification suggests that LLPS was necessary for the origins of early life because the droplet physically separated the intra-droplet components from the extra-droplet environment before the appearance of the cell membrane.

In short, on the basis of the Boltzmann distribution [7], it may be concluded from the work of Cliff Brangwynne's team that LLPS originates from electrostatic and elastic forces. Changes in entropy suggest a relationship between LLPS and epigenetic modifications. In addition, LLPS may play an important role in the origins of early life.

Conflict interests: The author declare no conflict interests;

References

1. C.P. Brangwynne, C.R. Eckmann, D.S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F . Jülicher, A.A Hyman. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, *324*, 1729-1732(2009).
2. Y .Shin,Y. C. Chang, D. S. Lee, J. Berry, D. W. Sanders, P. Ronceray, N.S. Wingreen, M. Haataja, C.P. Brangwynne. Liquid nuclear condensates mechanically sense and restructure the genome. *Cell*, *175*, 1481-1491(2018).
3. B. R.Sabari, A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K.Shrinivas, B.J. Abraham, N.M. Hannett, A.V. Zamudio, J.C. Manteiga, C.H. Li, Y.E. Guo, D.S. Day, J. Schuijers, E. Vasile, S. Malik, D.Hnisz, T.I. Lee, I.I. Cisse, R.G. Roeder, P.A. Sharp, A.K. Chakraborty, R.A. Young. Coactivator condensation at super-enhancers links phase separation and gene control. *Science*, *361*, eaar3958 (2018).
4. G.hang , Z. Wang, Z. Du , H. Zhang. mTOR regulates phase separation of PGL granules to modulate their autophagic degradation. *Cell*, *174*, 1492-1506 (2018).
5. M.Du , Z. J. Chen, DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science*, *361*, 704-709 (2018).
6. H.Falahati , A. Haji-Akbari,. Thermodynamically driven assemblies and liquid–liquid phase separations in biology. *Soft matter*, *15*, 1135-1154 (2019).
7. J .N. Israelachvili Ed. Intermolecular and surface forces: revised third edition. (Cambridge Academic press2011, 2011).
8. M.Feric , C. P.Brangwynne, A nuclear F-actin scaffold stabilizes ribonucleoprotein droplets against gravity in large cells. *Nature cell biology*, *15*, 1253(2013).
9. F. E.Garrett-Bakelman, M.Darshi, S. J.Green, R. C.Gur, L. Lin, B. R. Macias,et al. The NASA Twins Study: A multidimensional analysis of a year-long human spaceflight. *Science*, *364*, eaau8650 (2019).

10. R.P. Feynman, R.B. Leighton, M. Sands Ed.The Feynman lecture on Physic.(Addison Wesley Longman, London 1964).
11. G.Boulay, G.J. Sandoval, N.Riggi, S. Iyer, R.Buisson, B.Naigles, M.E. Awad, S.Rengarajan, A.Volorio, M.J. McBride, L.C. Broye, L. Zou, I. Stamenkovic, C. Kadoch, M.N. Rivera.. Cancer-specific retargeting of BAF complexes by a prion-like domain. *Cell*, *171*(1), 163-178(2017).
12. S.Alberti, A. A. Hyman. Are aberrant phase transitions a driver of cellular aging?. *BioEssays*, *38*(10), 959-968(2016).
13. R.Cortini, M.Barbi, B. R.Caré, C. Lavelle, A. Lesne, J. Mozziconacci, J. M. Victor. The physics of epigenetics. *Reviews of Modern Physics*, *88*, 025002(2016).
14. T. T.Ngo , J.Yoo, Q.Dai, Q.Zhang, C. He , A. Aksimentiev , T. Ha, Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nature communications*, *7*, 10813(2016).
15. A.Santos, L.Ferreira , M.Oliveira. Concerted flexibility of chromatin structure, methylome, and histone modifications along with plant stress responses. *Biology*, *6*, 3(2017).
16. A. P.Feinberg, The key role of epigenetics in human disease prevention and mitigation. *New England Journal of Medicine*, *378*, 1323-1334(2018).
17. G.Jenkinson, E. Pujadas, J. Goutsias, A. P. Feinberg, Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nature genetics*, *49*(5), 719(2017).
18. N. A.Temiz, D. E.Donohue, A.Bacolla, B. T.Luke, J. R. Collins. The role of methylation in the intrinsic dynamics of B-and Z-DNA. *PLoS One*, *7*, e35558(2012).
19. S.Ravichandran, V. K.Subramani, K. K.Kim. Z-DNA in the genome: from structure to disease. *Biophysical reviews*, *11*, 383–387 (2019).
20. R.A. Irizarry, C. Ladd-Acosta, B.Wen, Z.Wu, C. Montano, P. Onyango, H.Cui, K.Gabo, M .Rongione, M. Webster, H. Ji, J. Potash, S. Sabunciyan, A.P. Feinberg. The human colon cancer methylome shows similar hypo-and hypermethylation at conserved tissue-specific CpG island shores. *Nature genetics*, *41*, 178(2009).
21. Susan Standring, Ed.Gray's Anatomy 40th edition the Anatomical Basis of Clinical Practice. (Elsevier London,2009).
22. T. J.Richmond, J. T.Finch, B. Rushton, D. Rhodes, A. Klug. Structure of the nucleosome core particle at 7 Å resolution. *Nature*, *311*, 532(1984).
23. A.Pérez, C.L. Castellazzi, F. Battistini, K. Collinet, O. Flores, O. Deniz, M.L. Ruiz, D. Torrents, R. Eritja, M. Soler-López, M. Orozco. Impact of methylation on the physical properties of DNA. *Biophysical journal*, *102*, 2140-2148(2012).

24. G.Portella, F.Battistini, M. Orozco, (2013). Understanding the connection between epigenetic DNA methylation and nucleosome positioning from computer simulations. *PLoS computational biology*, *9*, e1003354.

25. J.Wu, A. Ocampo, J. C. I. Belmonte. Cellular metabolism and induced pluripotency. *Cell*, *166*, 1371-1385 (2016).

26. J. P. Etchegaray , R. Mostoslavsky, Interplay between metabolism and epigenetics: a nuclear adaptation to environmental changes. *Molecular cell*, *62*, 695-711 (2016).

27. W.Martin,M. J. Russell,. On the origin of biochemistry at an alkaline hydrothermal vent. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*, 1887-1926(2006).

28. Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. (New York: Garland Science; 2002).

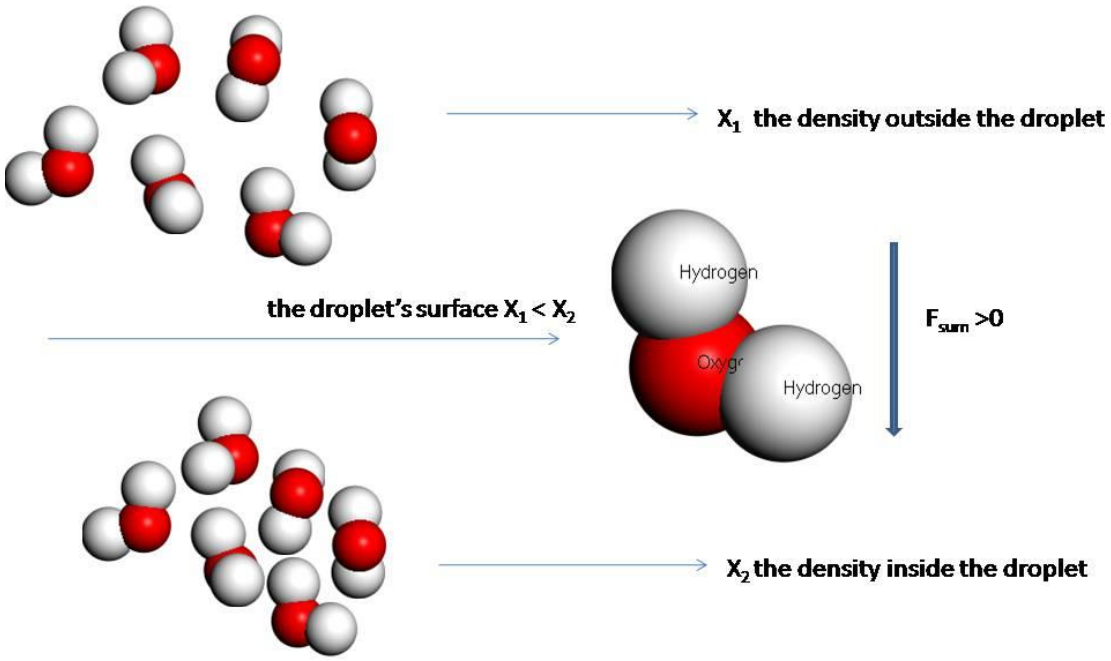

Fig. 1. The red ball is an oxygen atom. The white ball is a hydrogen atom. A red ball and two white balls combine into a water molecule. The largest water molecule model is at the droplet's surface. Every molecule is attracted or repulsed by the intermolecular forces of the surrounding molecules(attraction is dominant) because the molecules are uniformly distributed, which means that the sum force F_{sum}on each molecule must be zero. When liquid-liquid phase separation occurs ($X_2 > X_1$), the sum of the forces F_{sum} must be not zero. The motion direction of the molecule at the surface is oriented toward the inside of the droplet.