UNIVERSITAT DE
BARCELONA

Calculations of water free energy in bulk and large biological systems

Luis Enrique Coronas Serna

Aquesta tesi doctoral està subjecta a la llicència ***Reconeixement- NoComercial – SenseObraDerivada 4.0. Espanya de Creative Commons.***

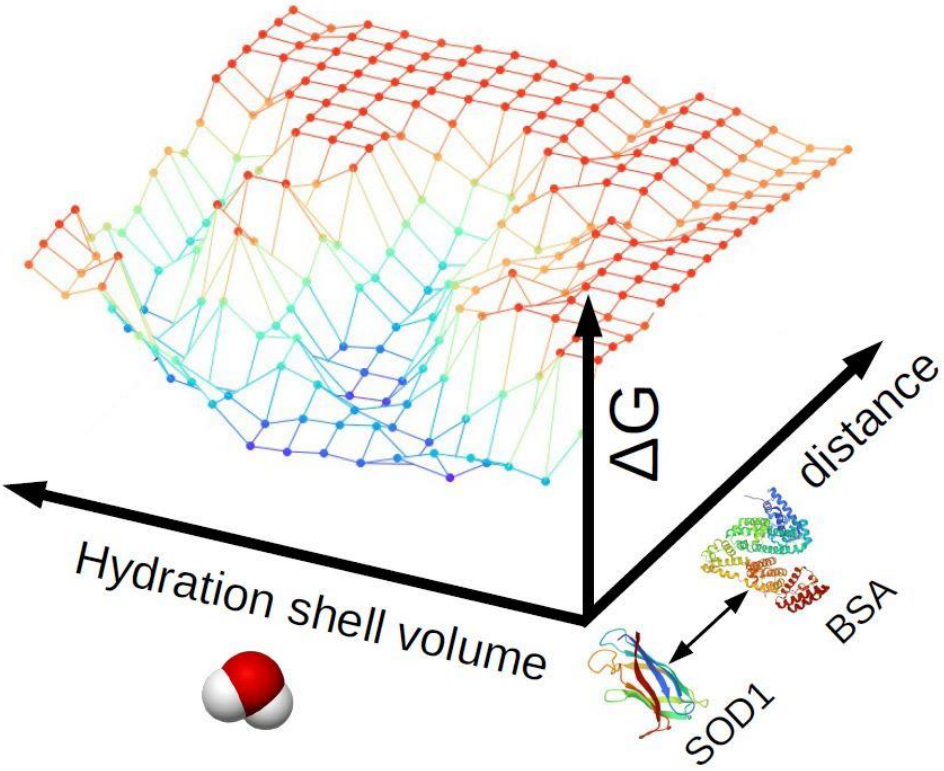
Esta tesis doctoral está sujeta a la licencia ***Reconocimiento - NoComercial – SinObraDerivada 4.0. España de Creative Commons.***

This doctoral thesis is licensed under the ***Creative Commons Attribution-NonCommercial-NoDerivs 4.0. Spain License.***

Tesis Doctoral

Calculations of water free energy in bulk and large biological systems

Luis Enrique Coronas Serna


ΔG
distance
Hydration shell volume
BSA
SOD1

Director: Dr. Giancarlo Franzese

UNIVERSITAT DE BARCELONA

Calculations of water free energy in bulk and large biological systems

Memoria presentada para optar al grado de doctor por la
Universitat de Barcelona

Programa de doctorado en Física

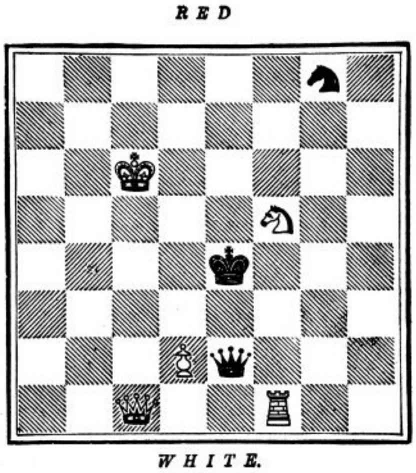
Autor:
Luis Enrique Coronas Serna

Director y Tutor:
Dr. Giancarlo Franzese

El científico tiende a saber todo de nada y el filósofo a saber nada de todo.
John Zimann

El exceso de especialización empieza cuando las facultades de ciencias dejan de compartir la cafetería y se agrava cuando cada cátedra instala su propia cafetera.

Jorge Wagensberg: "Si la naturaleza es la respuesta, ¿cuál era la pregunta?"



White Pawn (Alice) to play and win in eleven moves

Lewis Carroll "Through the looking glass and what Alice found there"

Contents

Agradecimientos **9**

Acknowledgements **11**

Summary in English **13**

Resumen en castellano **15**

Resum en català **19**

List of Abbreviations **23**

1 Introduction **27**

1.1 Atomistic and coarse-grained water models 29
1.2 The Franzese-Stanley coarse grained water model 31
1.3 Dynamic behavior of the FS water monolayer 35
1.3.1 Two dynamic crossovers in FS water 35
1.3.2 Effect of pressure and temperature on the dynamics 37
1.3.3 The diffusion anomaly 38
1.4 Phase Diagram of the FS water monolayer 41
1.4.1 The critical point analysis 44
1.4.2 The Widom line 45
1.4.3 The effect of cooperativity in the phase diagram 46
1.5 The Bianco-Franzese model for water and protein interfaces 47
1.5.1 Applications of the BF model 48
1.6 Motivation and Objectives of the Thesis 50

2 Methodology: Parallel distributed algorithms **53**

2.1 Introduction 53
2.2 Bulk FS model and simulation method 55
2.2.1 Extension from monolayer to bulk FS model 55
2.2.2 Simulation method: Definition of the MC step 57
2.2.3 Parallel code setup 58

2.3 Parallel Metropolis 58
2.4 Parallel Swendsen-Wang 62
2.5 Dynamic behavior of the algorithms 64
2.6 Conclusions 67

3 Parametrization and Phase Diagram 69
3.1 Introduction 69
3.2 Parametrization of the model 71
  3.2.1 Estimate of J_σ/J 71
  3.2.2 Choice of parameters and fitting to experimental data 73
3.3 Simulation Method 74
3.4 Results 75
3.5 Discussion 80
3.6 Conclusions 80

4 Supercooled Water 83
4.1 Introduction 83
4.2 Methods 85
  4.2.1 MC simulations for the phase diagram 85
  4.2.2 MC simulations for the estimation of the LLCP 86
  4.2.3 Estimation of the LLCP 86
4.3 Results 88
  4.3.1 Bulk FS thermodynamics at the supercooled region 88
  4.3.2 Finite size analysis for the LLPT and the LLCP 93
  4.3.3 FS water polymorphism at the critical region 95
4.4 Conclusions 99

5 Hydration effect in biomolecular simulations: the von Willebrand factor case. 101
5.1 Introduction 101
5.2 Model and methods 104
  5.2.1 The OPEP protein model 104
  5.2.2 Extension to bulk of the BF model 105
  5.2.3 Mapping of OPEP protein configurations into the BF volume partition 107
  5.2.4 Simulation method 108
5.3 Results 109
5.4 Conclusions 115

6 Effect of crowding in FUS condensates 117
6.1 Introduction 118

6.2 Methods 119
6.3 Results 120
6.4 Discussion 126
6.5 Conclusions 127

7 Conclusions 129

A Generation and usage of random numbers 133

B Scaling of the algorithms 135

C Rescale of the phase diagram under different parameters 141

D The fluctuation-dissipation theorem 143

E The histogram reweighting method 145

F Estimation of the order parameter *x(m)* for the LLPT 147

G Free energy landscape in terms of BF water enthalpy 151

H Supporting Information for Chapter 5 159

I Supporting Information for Chapter 6 163

Bibliography 171

Curriculum Vitae 209

Agradecimientos

Esta tesis arrancó en octubre de 2017, alargándose durante algo más de cinco años. Los inicios fueron turbulentos, con muchos momentos de desapego por la ciencia, pero compensados por la experiencia de dar clases en la universidad y de estudiar Filosofía. Con el paso del tiempo la balanza se fue invirtiendo. Aunque todo el edificio tuviera que caer al menos un par de veces, de los errores es de donde más se aprende, y el proyecto de investigación se erguiría cada vez sobre cimientos más sólidos. Hace apenas un año no hubiera siquiera apostado por una tesis como esta. Visto en perspectiva ha sido un camino muy arduo, pero gracias al cual he podido aprender mucho. No solo ciencia, sino, muy importante, acerca de mí mismo. En estas líneas quiero reconocer a los que me han acompañado y ayudado a discurrir ese camino, y a quienes pertenece también esta tesis. Honestamente, sin vosotros no lo hubiera conseguido.

A mi director, Giancarlo Franzese, quien apostó por mí el primero y me animó a embarcarme en el doctorado. Por instruirme en los usos y costumbres de la academia (becas, congresos, artículos, etc.) y por abrirme un camino en ella.

A mis compañeros de grupo de investigación, en especial a Arne, Oriol y Valentino, a quienes debo los primeros códigos del modelo en 3D, con los que empecé el TFM y la tesis, y de quienes he aprendido los entresijos de la investigación que no aparecen en los papers. Asimismo, quiero agradecer a Fabio Sterpone y a los miembros de su grupo, particularmente a Emeline y Stepan, por la calurosa acogida en el Institut de Biologie Physico-Chimique de París y por iniciarme en el vasto campo de la modelización de proteínas.

También quiero reconocer la ayuda recibida por parte de investigadores de otras universidades, con quienes he intercambiado discusiones en congresos, videoconferencias y/o correspondencia. Hago una mención especial a Thomas Kühne y Frederik Zysk (U. de Paderborn) por presentarnos los cálculos ALMOEDA y, aunque finalmente no se incluya en la presente memoria, a Pablo Debenedetti (U. de Princeton) y Thomas Gartner (Georgia Tech) por ayudarnos en el ajuste de la ecuación de estado TSEOS a nuestros resultados.

A mi familia: mis padres, mis hermanos Javi y Julia, y mis abuelos Celi, Lala, Luis y Paco. Por haber estado conmigo desde el principio y por haberme inculcado el valor del esfuerzo y la perseverancia. Y, por descontado, a mi tía María, quien me ha acogido en su casa durante los últimos siete años.

Finalmente, gracias a Andrés, Juan, Lucas y Maria, por tantas horas compartidas en el despacho y tantos planes fuera de él. A Cristina y Pablo, que me perdonarán aquella visita al MACBA. A Miguel, Mónica y Xavi, por tantos paseos, cañas y bravas. Por las giras por Europa, a Alberto, Arturo, Daniel, Eduardo, Néstor, Óscar y Pablo. Desde Burgos, a Álvar, Enrique y Óscar por la larga amistad que nos une.

A Isabel, por todo lo que hemos vivido y aprendido.

A todos y según corresponda: muchas gracias, moltes gràcies, grazie mille, thank you, merci beaucoup, danke schön.

Barcelona, enero de 2023

Acknowledgements

This thesis started in October 2017, stretching out for just over five years. The beginnings were turbulent, with many moments of detachment from science, but compensated by the experience of teaching at the university and studying Philosophy. As time went by, the balance was reversed. Although the whole structure fell at least a couple of times, it is from mistakes that one learns the most, and the research project would stand on increasingly solid foundations. Just a year ago, I would not have even bet on a thesis like this one. Seen in perspective, it has been a very arduous path, but thanks to which I have been able to learn a lot. Not only science, but, most relevantly, about myself. In these lines, I would like to acknowledge those who have accompanied and helped me along this path, and to whom this thesis also belongs. Honestly, without you, I would not have made it.

To my director, Giancarlo Franzese, who bet on me first and encouraged me to embark on the doctorate. For instructing me in the ways and customs of academia (grants, conferences, articles, etc.) and for opening a path for me in it.

To my colleagues in the research group, especially Arne, Oriol and Valentino, to whom I owe the first 3D model codes, with which I started the TFM and the thesis, and from whom I have learned the intricacies of research that do not appear in the papers. I would also like to thank Fabio Sterpone and the members of his group, particularly Emeline and Stepan, for the warm welcome at the Institut de Biologie Physico-Chimique in Paris and for initiating me into the vast field of protein modeling.

I would also like to acknowledge the help I have received from researchers from other universities with whom I have exchanged discussions at conferences, videoconferences and/or correspondence. A special mention goes to Thomas Kühne and Frederik Zysk (U. Paderborn) for introducing us to the ALMOEDA calculations and, although not included in this report, to Pablo Debenedetti (Princeton U.) and Thomas Gartner (Georgia Tech) for helping us in fitting the TSEOS equation of state to our results.

To my family: my parents, my siblings Javi and Julia, and my grandparents Celi, Lala, Luis and Paco. For having been with me from the beginning, and for having instilled in me the value of effort and perseverance. And, of course, my aunt Maria, who has taken me into her home for the past seven years.

Finally, thanks to Andrés, Juan, Lucas and Maria, for so many hours shared in the office and so many plans outside of it. To Cristina and Pablo, who will forgive me for

that visit to the MACBA. To Miguel, Mónica and Xavi, for so many walks, beers and bravas. To Alberto, Arturo, Daniel, Eduardo, Néstor, Óscar and Pablo, for the tours around Europe. From Burgos, to Álvar, Enrique and Óscar for the long friendship that unites us.

To Isabel, for all that we have lived and learned.

To all and as appropriate: muchas gracias, moltes gràcies, grazie mille, thank you, merci beaucoup, danke schön.

Barcelona, January 2023

Summary in English

This thesis develops, validates, and explores applications of the Franzese-Stanley (FS) coarse-grained (CG) model for bulk water, ultimately intended to represent the solvent in large-scale simulations of biological systems. The model coarse-grains the positions of the water molecules through a discrete density field, but keeps a detailed account of the hydrogen bond (HB) contributions to local density heterogeneity. It was initially proposed for hydrophobically confined water monolayers and is suitable for analytic calculations and Monte Carlo simulations. The FS monolayer displays dynamic and thermodynamic anomalies qualitatively consistent with experiments and atomistic calculations. Moreover, its extension to hydration water in protein solutions unveils the water contribution to protein folding, protein design, and protein aggregation, including effects arising from changes in temperature, pressure, salinity, and pH.

Here, we extend to bulk the FS model. Although going from monolayer to bulk could drastically change the thermodynamic behavior of the model, we show that the bulk FS reproduces well the experimental data of water. Furthermore, we prove that the model is suitable for simulations of large biological systems in explicit solvent. On the one hand, we account for water interactions at the molecular level, at variance with commonly employed coarse-grained models. On the other hand, we equilibrate systems at length scales and times so large that it would be unaffordable for current atomistic simulations.

First, we develop efficient parallel distributed algorithms for Metropolis and Swendsen-Wang Monte Carlo simulations. We benchmark accessible size systems of $260^3 \sim 10^7$ water molecules using parallel Metropolis and $128^3 \sim 10^6$ with Swendsen-Wang. Second, we calibrate the model parameters according to ab initio calculations and experimental results. We find that the model fits the experimental data in a wide range of temperatures T and pressures P around ambient conditions. Third, we transfer the model to deep supercooled conditions and compare the model's predictions to those obtained from atomistic models. We find that the FS model displays a liquid-liquid critical point (LLCP) at $T_C = (186 \pm 2)$ K and $P_C = (174 \pm 14)$ MPa, consistent with estimates from the most accepted atomistic models.

Finally, we investigate applications of the bulk FS model to hydrate large biological systems. More specifically, we develop a simulation approach to estimate the contribution of the solvent to the free energy of hydrated systems previously calculated with implicit solvent models. We consider two cases of application. First, we studied the unfolding upon increasing hydrodynamic stress of the globular domains A1, A2, and A3 of the von Willebrand factor (vWf). Here, we identify four characteristic states of the system and estimate the free energy barriers that separate them as a function of the shear rate. Second, we consider the SOD1 protein sequestration into FUS biocondensates and the anti-sequestration action of the BSA crowder, mimicking the high cytoplasm concentration. BSA are globular proteins, while FUS are intrinsically disordered proteins. Our results suggest that the decrease of the SOD1 partitioning coefficient into FUS biocondensates when BSA is present is a consequence of the increase in water's entropy. In both the vWf and SOD1 examples, we show that the explicit calculation of water contribution to the free energy is essential to interpret the phenomena' physics correctly. Within this approach, we overcome the intrinsic limitation of implicit solvent models that, by construction, average over the solvent degrees of freedom.

Resumen en castellano

El agua es de suma importancia para los sistemas biofísicos y bioquímicos debido a que ejerce un papel determinante en la funcionalidad de proteínas, membranas o nanopartículas. Por ejemplo, las interacciones agua-proteína modifican el paisaje de energía libre en el que interactúan los aminoácidos, afectando al plegamiento y a la agregación de las proteínas o a los procesos de selección natural de sus secuencias.

Desde las últimas décadas, la investigación por simulación ha contribuido a nuestra comprensión acerca de los mecanismos subyacentes al comportamiento de sistemas biológicos a escalas micro y mesoscópoicas, incluida la influencia del entorno (confinamiento, presión, temperatura, salinidad, etc.). En este contexto, la modelización del agua es esencial: a mayor precisión en la representación de sus propiedades, mayor fiabilidad de los resultados obtenidos. Desafortunadamente, una mayor precisión suele conllevar un mayor coste computacional. En las simulaciones que representan el agua expícitamente, el modelo de agua puede convertirse en el factor limitante de las escalas de longitud y tiempo accesibles. El motivo es que por cada unidad biomolecular (átomo, aminoácido, etc.) se deben incluir cientos o miles de moléculas de agua, incrementando enormemente el número de interacciones que se deben calcular.

Dado el interés por abordar problemas que requieren mayores escalas de longitud y tiempo, muchos científicos han propuesto modelos de grano grueso [coarse-grained models] que, promediando sobre los grados de libertad menos relevantes para el problema en cuestión, permiten reducir enormemente el coste de las simulaciones. En 2002, Franzese y Stanley (FS) propusieron un modelo de grano grueso para monocapas de agua hidrofóbicamente confinadas (2D). El modelo promedia sobre las posiciones de las moléculas mientras que mantiene una descripción detallada de la red de puentes de hidrógeno y cómo su presencia introduce heterogeneidades en la densidad local. El estudio de la monocapa de agua FS muestra la presencia de anomalías dinámicas y termodinámicas cualitativamente consistentes con resultados experimentales y los cálculos de modelos atomísticos. En particular, destacamos que el modelo presenta una transición de fase líquido-líquido entre dos fases líquido-de-alta-densidad y líquido-de-baja-densidad que termina en un punto crítico. Este resultado contribuye al debate sobre el origen de las anomalías del agua, apoyando la hipótesis del punto crítico líquido-líquido. Además, la extensión del modelo llevada a cabo por Bianco y Franzese (BF) para reproducir la hidratación de proteínas ayuda a desvelar las contribuciones

del agua a los procesos de plegamiento, diseño y agregación de proteínas, teniendo en cuenta además los efectos derivados de cambios de temperatura, presión, salinidad y pH.

Objetivos de la tesis

A lo largo de la presente tesis, extendemos a 3D el modelo FS, lo validamos demostrando que se ajusta en un amplio rango de presiones y temperaturas a los resultados experimentales, y lo aplicamos para calcular la contribución del agua a la energía libre de sistemas biológicos de gran extensión.

El objetivo principal es demostrar que el modelo en 3D es aplicable a simulaciones de gran escala, calculando explícitamente las contribuciones del solvente. Para ello, el modelo debe reproducir las propiedades del agua a condiciones ambiente, y en otras condiciones relevantes para los procesos biológicos, y debe ser posible, mediante simulaciones, equilibrar sistemas de tamaño muy grande. La consecución de este objetivo abriría un nuevo método para llevar a cabo simulaciones de sistemas biológicos muy grandes, con una descripción sin precedentes de las interacciones agua-residuo, y de los puentes de hidrógeno en la interfaz.

Por otro lado, la extensión del modelo de 2D a 3D puede afectar drásticamente al comportamiento del sistema. Por tanto, es pertinente estudiar las propiedades dinámicas y termodinámicas del modelo en 3D y, a pesar de ser menos relevante para la cuestión de las simulaciones de sistemas biológicos, si el modelo en 3D muestra o no el punto crítico líquido-líquido. De ser así, el modelo FS quedaría reforzado por su consistencia con los modelos atomísticos más aceptados.

Estructura de la tesis y Resultados

- Capítulo 1: Contextualizamos el problema de la modelización del agua y las simulaciones de sistemas biológicos y revisamos los resultados obtenidos en investigaciones de la monocapa FS.

- Capítulo 2: Extendemos el modelo a 3D introduciendo una partición de la red FS que garantiza que todas las moléculas de agua puedan formar hasta cuatro puentes de hidrógeno. Además, desarrollamos algoritmos Monte Carlo paralelizados que tienen en cuenta las propiedades específicas de la red FS. Las simulaciones se ejecutan en GPUs y muestrean eficientemente sistemas de agua de tamaño sin precedentes de hasta 17.576.000 moléculas mediante Metrópolis y hasta 2.097.152 con Swendsen-Wang. Para los sistemas más grandes las simulaciones se realizan 140 y 70 veces más rápido respecto a su implementación secuencial en CPUs, con Metrópolis y Swendsen-Wang respectivamente.

- Capítulo 3: Calibramos los parámetros del modelo de manera que sus predicciones se ajusten a la densidad experimental y las funciones respuesta del agua, al menos en condiciones ambiente. Para ello, nos servimos de cálculos ab initio y resultados experimentales. Ajustando a funciones de rescalado lineales, el modelo reproduce cuantitativamente la ecuación de estado del agua (experimental) presión ambiente $P = 0.1$ MPa y temperaturas $270 \leq T/$ K ≤ 330. Aumentando la presión P, el intervalo de T se reduce hasta $290 \leq T/$ K ≤ 310 para $P = 50$ MPa. Dicho rango de T y P cubre sobradamente las condiciones de trabajo de la mayoría de proteínas.

- Capítulo 4: Estudiamos el comportamiento del modelo en la región sobreenfriada. En particular, mostramos presencia del punto crítico líquido-líquido a $P_C = (174 \pm 14)$ MPa y $T_C = (186 \pm 2)$ K en el límite termodinámico, condiciones cuantitativamente comparables a las estimaciones de los modelos atomísticos más aceptados. Además, el diagrama de fases en la región crítica es consistente con la presencia de polimorfismo (líquido) que recuerda al poliamorfismo (hielo amorfo) observado experimentalmente.

- Capítulo 5: Proponemos un método de simulación que toma trayectorias de sistemas biológicos previamente calculadas con modelos de modelos de agua implícita y mapea las configuraciones de proteínas en su representación correspondiente en la red FS. El objetivo es calcular la contribución del agua a la energía libre de solvatación de grandes sistemas biológicos. En este capítulo aplicamos el método al desplegamiento de los dominios globulares A1, A2 y A3 del factor de von Willebrand (vWf), bajo condiciones de estrés hidrodinámico. Gracias a la inclusión de la componente del agua, identificamos los cuatro estados característicos del sistema y calculamos las barreras de energía libre que los separan.

- Capítulo 6: Aplicamos el método de simulación desarrollado en el Capítulo 5 al caso del secuestro de proteínas SOD1 en condensados de FUS y el efecto antisecuestro producido por el medio de BSA que imita el citoplasma celular. La comparación entre ambos casos es relevante porque FUS es una proteína intrínsecamente desordenada y BSA, globular. Mediante la descripción del sistema incluyendo la contribución del agua, explicamos el efecto de antisecuestro producido cuando los condensados de FUS están en un medio de BSA por el aumento de la entropía del agua en la solución de BSA.

  Tomados en conjunto, los Capítulos 5 y 6 muestran que es necesario incluir los grados de libertad del solvente para describir de forma adecuada la física del sistema y, en consecuencia, para calcular las barreras de energía libre que separan los estados característicos de los procesos examinados. Así, superamos la limitación intrínseca de los modelos de agua implícita que, por construcción, promedian sobre los grados de libertad del solvente.

- Capítulo 7 Reunimos los principales resultados y conclusiones generales de la tesis. Concluimos que la tesis ha demostrado satisfactoriamente que el modelo FS en 3D, junto con la parametrización propuesta, es capaz de simular grandes sistemas de agua con resultados consistentes con los datos experimentales. Además, hemos aplicado con éxito el modelo a problemas de interés biológico, mostrando que la inclusión del agua de forma explícita es necesaria para caracterizar adecuadamente dichos sistemas.

Resum en català

L'aigua és de summa importància per als sistemes biofísics i bioquímics a causa del seu paper determinant en la funcionalitat de proteïnes, membranes o nanopartícules. Per exemple, les interaccions aigua-proteïna modifiquen el paisatge d'energia lliure en el qual interactuen els aminoàcids, influint sobre el plegament i l'agregació de proteïnes o els processos de selecció natural de les seves seqüències.

Des de les últimes dècades, la recerca per simulació d'aquesta mena de sistemes ha contribuït a la nostra comprensió al voltant dels mecanismes subjacents al seu comportament a escales micro i mesoscòpiques, inclosa la influència de l'entorn (confinament, pressió, temperatura, salinitat, etc.). En aquest context, la modelització de l'aigua és essencial: a major precisió en la representació de les seves propietats, major fiabilitat dels resultats obtinguts. Malauradament, una major precisió sol comportar un major cost computacional. En les simulacions que representen l'aigua explícitament, el model d'aigua pot esdevenir el factor limitant de les escales de longitud i temps accessibles. El motiu és que per cada unitat biomolecular (àtom, aminoàcid, etc.) s'han d'incloure centenars o milers de molècules d'aigua, incrementant enormement el nombre d'interaccions que s'han de calcular.

A causa de l'interès de tractar problemes que requereixen majors escales de longitud i temps, molts científics han proposat models de gra gruixut [coarse-grained models] que, fent una mitjana de sobre els graus de llibertat menys rellevants per al problema en qüestió, redueixen enormement el cost de les simulacions. En 2002, Franzese i Stanley (FS) van proposar un model de gra gruixut per a monocapes d'aigua hidrofòbicament confinades (2D). El model fa una mitjana de sobre les posicions de les molècules mentre que manté una descripció detallada de la xarxa de ponts d'hidrogen i com la seva presència introdueix heterogeneïtats a la densitat local. L'estudi de la monocapa d'aigua FS mostra la presència d'anomalies dinàmiques i termodinàmiques qualitativament coherents amb resultats experimentals i càlculs atomístics. En particular destaquem que el model presenta una transició de fase líquid-líquid entre dues fases líquid-d'alta-densitat i líquid-de-baixa-densitat que acaba en un punt crític. Aquest resultat contribueix al debat sobre l'origen de les anomalies de l'aigua, donant suport a la hipòtesi del punt crític líquid-líquid. A més, l'extensió del model duta a terme per Bianco i Franzese (BF) per a reproduir la hidratació de proteïnes ajuda a revelar les contribucions de l'aigua als processos de plegament, disseny i agregació de proteïnes, tenint en compte a més els efectes derivats de canvis de temperatura, pressió, salinitat i pH.

Objectius de la tesi

Al llarg de la present tesi, ampliem a 3D el model FS, el validem demostrant que s'ajusta en un ampli rang de pressions i temperatures als resultats experimentals, i l'apliquem per a calcular la contribució de l'aigua a l'energia lliure de sistemes biològics de gran grandària.

L'objectiu principal és demostrar que el model en 3D és aplicable a simulacions de gran escala, calculant explícitament les contribucions del solvent. Per a això, el model ha de reproduir les propietats de l'aigua a condicions ambient, i en altres condicions rellevants per als processos biològics, i ha de ser possible, mitjançant simulacions, equilibrar sistemes de grandària molt gran. La consecució d'aquest objectiu obriria un nou mètode per a dur a terme simulacions de sistemes biològics molt grans, amb una descripció sense precedents de les interaccions aigua-residu, i dels ponts d'hidrogen a la interfície.

D'altra banda, l'extensió del model de 2D a 3D pot afectar dràsticament el comportament del sistema. Per tant, és rellevant estudiar les propietats dinàmiques i termodinàmiques del model en 3D i, malgrat ser menys rellevant per a la qüestió de les simulacions de sistemes biològics, si el model en 3D mostra o no el punt crític líquid-líquid. En aquest cas, el model FS quedaria reforçat per la seva consistència amb els models atomístics més acceptats.

Estructura de la tesi i Resultats

- Capítol 1: Contextualitzem el problema de la modelització de l'aigua i les simulacions de sistemes biològics i revisem els resultats obtinguts en recerques de la monocapa FS.

- Capítol 2: Ampliem el model a 3D introduint una partició de la xarxa FS que garanteix que totes les molècules d'aigua puguin formar fins a quatre ponts d'hidrogen. A més, desenvolupem algorismes Monte Carlo paral·lelitzats que tenen en compte les propietats específiques de la xarxa FS. Les simulacions s'executen en GPUs i mostregen eficientment sistemes d'aigua de grandària sense precedents de fins a 17.576.000 molècules mitjançant Metròpolis i fins a 2.097.152 amb Swendsen-Wang. Per als sistemes més grans les simulacions es realitzen 140 i 70 vegades més ràpid respecte a la seva implementació seqüencial en CPUs, amb Metròpolis i Swendsen-Wang respectivament.

- Capítol 3: Calibrem els paràmetres del model de manera que les seves prediccions s'ajustin a la densitat experimental i les funcionis resposta de l'aigua, almenys en condicions ambient. Per a això, fem servir càlculs ab initio i resultats experimentals. Ajustant a funcions de rescalat lineals, el model reprodueix quantitativament l'equació d'estat de l'aigua (experimental) a pressió ambient $P = 0.1$ MPa i temperatures $270 \leq T/$ K ≤ 330. Augmentant la pressió P, l'interval de T es redueix fins a $290 \leq T/$ K ≤ 310 per a $P = 50$ MPa. Aquest rang de T i P engloba àmpliament les condicions de treball de la majoria de proteïnes.

- Capítol 4: Estudiem el comportament del model a la regió sobrerefredada. En particular, vam mostrar presència del punt crític líquid-líquid a $P_C = (174 \pm 14)$ MPa y $T_C = (186 \pm 2)$ K en el límit termodinàmic, condicions quantitativament comparables a les estimacions dels models atomístics més acceptats. A més, el diagrama de fases a la regió crítica és consistent amb la presència de polimorfisme (líquid) que recorda al poliamorfisme (gel amorf) observat experimentalment.

- Capítol 5: Proposem un mètode de simulació que pren trajectòries de sistemes biològics prèviament calculades amb models de models d'aigua implícita i amb el qual obtenim les configuracions de proteïnes en la seva representació corresponent en la xarxa FS. L'objectiu és calcular la contribució de l'aigua a l'energia lliure de solvatació de grans sistemes biològics. En aquest capítol apliquem el mètode al desplegament dels dominis globulars A1, A2 i A3 del factor de von Willebrand (vWf), sota condicions d'estrès hidrodinàmic. Gràcies a la inclusió de la component de l'aigua, identifiquem els quatre estats característics del sistema i calculem les barreres d'energia lliure que els separen.

- Capítol 6: Apliquem el mètode de simulació desenvolupat en el Capítol 5 al cas del segrest de proteïnes SOD1 en condensats de FUS i l'efecte antisegrest produït pel medi de BSA que imita el citoplasma cel·lular. La comparació entre tots dos casos és rellevant perquè FUS és una proteïna intrínsecament desordenada i BSA, globular. Mitjançant la descripció del sistema incloent-hi la contribució de l'aigua, expliquem l'efecte d'antisegrest produït quan els condensats de FUS estan en un medi de BSA per l'augment de l'entropia de l'aigua en la solució de BSA.

  Conjuntament, els Capítols 5 i 6 mostren que és necessari incloure els graus de llibertat del solvent per a descriure de forma adequada la física del sistema i, en conseqüència, per a calcular les barreres d'energia lliure que separen els estats característics dels processos estudiats. Així, superem la limitació intrínseca dels models d'aigua implícita que, per construcció, fan una mitjana de sobre els graus de llibertat del solvent.

- Capítol 7: Reunim els principals resultats i conclusions generals de la tesi. Concloem que la tesi ha demostrat satisfactòriament que el model FS en 3D, juntament amb la parametrització proposada, és capaç de simular grans sistemes d'aigua amb resultats consistents amb les dades experimentals. A més, hem aplicat amb èxit el model a problemes d'interès biològic, mostrant que la inclusió de l'aigua de manera explícita és necessària per a caracteritzar adequadament els sistemes.

List of Abbreviations

ALMOEDA Absolutely Localized Molecular Orbitals - Energy Decomposition Analysis

ALS Amyotrophic Lateral Sclerosis

BF Bianco-Franzese

BSA Bovine Serum Albumin

c.c. checkerboard cube

CG Coarse-grain

CPF Critical point free

CPU Central Processing Unit

CUDA Compute Unified Device Architecture

CWD Crowder

FS Franzese-Stanley

FUS Fused in Sarcoma

GPU Graphic Processing Unit

HB Hydrogen bond

HDA High density amorphous

HDL High density liquid

IS Implicit solvent

IU Internal Units

LBMD Lattice-Boltzmann Molecular Dynamics

LDA Low density amorphous

LDL Low density liquid

LG Spinodal Liquid-Gas Spinodal

LJ Lennard-Jones

LLCP Liquid-liquid critical point

LLPS Liquid-liquid phase separation

LLPT Liquid-liquid phase transition

MC Monte Carlo

MD Molecular Dynamics

MLO Membraneless Organelle

MO Molecular Orbital

n.n. nearest neighbor

OPEP Optimized Potential for Efficient Protein structure prediction

PBCs Periodic Bondary Conditions

PC Partition coefficient

PHI Hydrophilic

PHO Hydrophobic

PIMD Path-Integral Molecular Dynamics

SF Speedup Factor (Chapter 2, Appendix B) ; Singularity free (elsewhere)

SG Stress granule

SI International System of Units

SL Stability limit

SOD1 Superoxide Dismutase 1

SW Swendsen-Wang

TMD Temperature of Maximum Density

TminD	Temperature of minimum Density
vdW	van der Waals
VHDA	Very high density amorphous
VHDL	Very high density liquid
vWf	von Willebrand factor
WL	Widom line

Chapter 1

Introduction

Chapter adapted from Ref. [1].

Water plays a fundamental role in the behavior of biomolecular systems, since its unique structural and dynamical properties strongly affect their macromolecular conformations and functioning [2,3]. For example, water hydrating proteins play a key role in denaturation [4–7] or in the regulation of protein aggregation [8, 9]. On the other hand, water between cell membranes is responsible for self-assembly of phospholipids into bilayers [10, 11] and stabilizes the membrane structure [12–14]. To rationalize the effects of water on such systems, we must pay attention to the so-called *biological* or *hydration layer*, a thin shell of two to three layers of water molecules that envelopes biological systems [15]. It is commonly understood that the hydration layer is constitutive of the behavior of such systems, since water-amino acid interactions and water-water hydrogen bonds (HBs) modulate the free energy landscape [2]. From the thermodynamic point of view, the total free energy that governs the system results from the balance between enthalpic solute-solute and solute-solvent interactions (van der Waals, electrostatic, HB) and the entropic cost of breaking and forming the (distorted) HB network at the hydration shell [16]. Therefore, it is clear that a more in-depth understanding of water properties is essential in biomolecular modeling [17, 18].

Despite the extensive research effort, many features of the behavior of water are still poorly understood [19]. Water has more than 60 thermodynamic, dynamic, and structural anomalies [2] whose origin is largely debated [20–24]. For example, in contrast with normal fluids, water has the property of *polyamorphism*, that is, it has at least three amorphous solids [20, 25–27], whose formation depends on the preparation route [23], and numerous (crystal) ice polymorphs: 17 have been confirmed experimentally and others are predicted computationally [28]. Water can be supercooled in its liquid state almost 50 degrees below its melting temperature [29]. Ice has a lower density than liquid water at ambient pressure, and its density decreases below 4°C. Experiments [30–32] have shown that water's isobaric specific heat C_P and isothermal compressibility K_T have a non-monotonic behavior with minima at approximately 35°C and 46°C, respectively, at ambient pressure, while isobaric thermal expansivity α_P turns negative at approximately 4°C. All the anomalies of water become more relevant in the supercooled region, where the fluctuations increase upon cooling [22, 24, 31, 33], instead of decreasing as in normal liquids.

Several thermodynamic scenarios have been proposed to explain the origin of the anomalies and polyamorphism. The *singularity free* (SF) scenario hypothesizes that the uncommon volume-entropy and volume-energy anticorrelations of water, due to the water hydrogen bonds (HBs) properties, are responsible for the increase of fluctuations in the supercooled region with no singular behavior [34,35]. Three other scenarios, the *stability limit* (SL) conjecture [36], the *liquid-liquid critical point* (LLCP) scenario [37], and the *critical point free* (CPF) hypothesis [38, 39], instead postulate a singular behavior that enhances the fluctuations at low temperature: a reentrant spinodal for the first, a critical point for the second, and a first-order phase transition for the third scenario. Stokely et al. demonstrated that all the scenarios belong to the same theoretical framework and that it is possible to go from one to another by tuning a single parameter related to the water's cooperativity [40]. Nevertheless, which of these scenario holds for water is still a matter of debate, because so far no definitive experimental evidence has been found, although many recent experiments are contributing to enrich our insight [29, 41–44]. One of the issues is that water freezes before experimental measurements are made in the region, conventionally called 'no-man's land' [23,45,46], where the different scenarios predict different behaviors.

Several strategies have been explored to overcome the inevitable crystallization of supercooled water, including strong confinement [47–49] and anti-freezing solutions [43, 50, 51] with results possibly related to the bulk case [52–55]. Experiments [56, 57] and simulations on confined water [58–60] show controversial phenomena, as water under confinement can have properties significantly different from those found in bulk water. For instance, experiments on water confined between flat crystals of graphite and hexagonal boron nitride find a decay of two orders of magnitude in the dielectric constant of water as a function of the sample thickness (down to 1 nanometer) compared to the bulk case [61]. The mobility of water is also strongly affected by confinement, as different regimes can be observed depending on the structure, composition and geometry of the confining surfaces, going from subdiffusive to superdiffusive [62–64]. Moreover, confinement also modifies the phase diagram, as crystallization can be avoided, at least partially, at temperatures down to 160 K for water confined in silica MCM-41 nanopores [55, 65, 66].

The peculiar properties of confined water are also relevant for the behavior of biological systems, e.g. in biomolecular crowded environments or in membrane channels [67, 68]. A noteworthy example is that of water in the cellular cytoplasm environment, although its study is challenging due to the large complexity of such system. NMR experiments in living *Escherichia coli* and *Haloarcula marismortui* (Hmm) show that 85% of the water molecules have bulk-like dynamics while the remaining 15%, in contact with biomolecules, are retarded by a factor 15 [69]. However, the results for Hmm are in apparent conflict with neutron scattering experiments showing a translational diffusion 250 times slower compared to the bulk-case, whereas they are consistent for E. coli [70]. A possible rationale of this difference is that NMR and neutron scattering resolve different motions, rotational and translational respectively, that could decouple upon such environmental conditions [70].

In this chapter, we describe in detail the state-of-the-art of the Franzese-Stanley (FS) coarse-grained (CG) model for water [71, 72] and its contributions to the ongoing debate on the origin of water anomalies and the role of water in the behavior of biological systems. In Sec. 1.1, we motivate the introduction of the FS model in the context of other atomistic and CG water models. Next, we describe the FS model for water monolayers (2D) in Sec. 1.2. The Secs. 1.3 and 1.4 review how the FS monolayer accounts for the dynamic and thermodynamic anomalies of water, relating the cooperativity of the HB network to the occurrence of a possible liquid-liquid phase transition (LLPT) in the supercooled water region. In Sec. 1.5, we describe the Bianco-Franzese (BF) extension of the FS monolayer to represent water at biological interfaces [5, 6]. The BF model rationalizes how the HBs at the hydration shell contribute to protein folding [5, 7], design [6] and aggregation [8, 9, 73].

The goal of this Thesis is to extend the FS (BF) model from monolayers (Secs. 1.2-1.5) to bulk, i.e. from 2D to 3D systems. In Sec. 1.6, we motivate the interest of this project, describe our objectives, and relate the chapters structure of the Thesis.

1.1 Atomistic and coarse-grained water models

Modeling water is a notorious difficult problem [74, 75], in particular because it is not settled how to include the quantum many-body, or cooperative, effects in water interactions. In the literature there are more than a hundred water models [76], from those parametrized based on experimental data to those fitting ab-initio calculations, and each model is coarse-grained at a different level, from atomistic non-polarizable and rigid, to polarizable and/or flexible, from spatially resolved to spatially coarse-grained. For a number of these models, it is possible to explore the no-man's land, although it could require considerably large computational times [77] and elaborated analysis [78–82].

Specifically, for those models belonging to the family of ST2, TIP4P and TIP5P potentials [22], the LLCP hypothesis is, among the different scenarios, the one that better adjusts to the low-temperature phase diagram. In particular, if the correct order parameter, defined as a combination of energy and density [83], is considered, it has been shown by all-atoms simulations that the ST2 [77,84], TIP4P/2005, TIP4P/Ice [85] and WAIL [86] models have a LLCP between two metastable liquid phases, low-density liquid (LDL) and high-density liquid (HDL), that belongs to the 3D Ising universality class. Furthermore, ST2 water show different local structures of LDL and HDL that are better discriminated by structural parameters quantifying the amount of diamond structure in the first shell and the amount of hcp structure in the second shell [87].

Similar conclusions have been reached for different models, for example recently in Refs. [88–90], with results that are in principle model-dependent. Therefore, understanding which feature of these models regulates the occurrence of the anomalies of water and its peculiar properties is a task that requires a detailed analysis for each of them [88].

Regarding simulations of biological systems, the inclusion of water contributions is essential to adequately describe the behavior of the system and functioning, including macromolecular conformations, protein dynamics, or unfolding and aggregation paths. Atomistic simulations of hydrated systems extend over time scales going from 10 fs to 1 μs and length scales from 1 Å to 100 nm, adopting water models such as SCP/E, ST2, TIP3P, TIP4P/2005 and TIP5P [12–14, 77, 87, 91–93]. Recently, Zheng et al. probed μs-long atomistic simulations of 98912 TIP4P/2005 water molecules and 40 chains of 163-residue Fused in Sarcoma (FUS) proteins, previously equilibrated with a CG model [94]. Nevertheless, these models have extremely large equilibration times at very low temperature or at extreme pressure [95–97] and have free-energy minima that differs from those of more realistic polarizable models [98, 99].

Alternatively, CG models represent the system in a simplified manner, keeping the degrees of freedom that are essential for its behavior. Typically, they represent multiple atoms or molecules into single beads that interact through effective potentials. The most straightforward strategy consists in mapping one molecule into one interacting bead. For example, core-softened potentials [100–102] exhibit water-like structural and thermodynamic anomalies, mW employs a Stillinger-Weber potential with explicit three-body contributions [103], and BUMPER maps three-body interactions into a pure basis set of two-body terms [104, 105]. Other models consider interacting beads representing more than one water molecule. For instance, the MARTINI model maps four molecules into a single bead [106] and SIRAH [107, 108], eleven molecules into four beads [109]. Their reduced computational cost has prompted its use in biological simulations [110]. However, they neglect polarization and cooperative contributions to the HBs, which makes them unsuitable to accurately describe the role of the solvent in the behavior of biological systems [111]. Recently, machine learning techniques have been employed to develop CG water models [89, 112]. The major advantage is that neural networks can be trained to optimize the representation of large sets of reference data, in a clear departure from traditional model fitting [113].

In the following we will describe our approach to define a model of water that

- is suitable for large-scale, long-time simulations as needed in biologically relevant problems;
- is suitable for theoretical calculations, important to understand general properties;
- includes many-body interactions as required for a proper water model;
- can be equilibrated at extreme conditions.

1.2 The Franzese-Stanley coarse grained water model

In 1979, in a seminal paper that was marking the conclusions of a CECAM workshop about computer simulations of liquid water, Barnes, Finney Nicholas and Quinn stated that the use of pair-additive interactions for water is a "serious oversimplification" that "is necessary to abandon" because it does not account for many-body forces, to which "solution and interfacial properties of aqueous systems are particularly sensitive" [74]. For example, they reviewed quantum calculations for small water clusters, showing that HB energy in trimers and tetramers is 20-30% stronger than in dimers. More recently, James et al. [98] and Hernández-Rojas et al. [99] have studied the configurations that minimize the energy of water clusters made of up to 21 molecules adopting non-polarizable and polarizable models. The comparison shows structural differences for clusters with more than 5 molecules, in particular with 6 or more than 10. This observation emphasizes that many-body effects in water are especially important when there are at least 5 molecules.

To account for these many-body effects, in 2002 Franzese and Stanley (FS), proposed a Hamiltonian model[1] that is analytically tractable [71, 72, 115] and is suitable for MC calculations at constant pressure P and temperature T [116–122]. Thanks to a percolation mapping [123], the model can be equilibrated with efficient MC methods (cluster MC) [40, 124, 125], at extreme $T < 125$ K and $P > 10$ GPa or negative pressures [126–133], for very large systems with more than 160.000 molecules [134]. On the other hand, by adopting *local* MC dynamics [117] and a rescaling of time units based on experimental data, it is possible to simulate the model for times up to 100 s [135]. Furthermore, using diffusive MC dynamics, transport properties can be evaluated [121, 122] offering new interpretations for the related anomalies [125, 126, 132].

---

[1] A preliminary version of the model and its mean field solution at zeroth order was proposed in 2000 [114].

The FS model is defined as follows. For N water molecules distributed in a volume V, we partition V into N equal cells, each with an average volume $v \equiv V/N \geq v_0$, where v_0 is the van der Waals volume of a water molecule and v_0/v is the cell density in van der Waals units. Each molecule $i \in [i, \ldots, N]$ has an index $n_i = 0$ if $v_0/v \leq 0.5$, and $n_i = 1$ otherwise. If, for sake of simplicity, we consider the case of a water monolayer with height $h = 0.5$ nm, then $v \equiv hr^2$ with $r(v_0) \equiv r_0 \simeq 2.9$ Å being the van der Waals diameter, and $n_i = 0$ when $r \geq 3.7$ Å$\equiv r_{\max}$. Because $r_{\max}$ is the maximum O–O elongation of a straight HB, as calculated by *ab initio* molecular dynamics simulations from the proton-transfer coordinate [136] assuming a covalent distance O–H $\simeq 1$ Å, when $n_i = 0$ the molecule i cannot form HBs because its neighbor molecules are too far. Hence, the first condition to form HBs for molecule i is that $n_i = 1$.

It is worth noticing at this point that the discretized *density field* n_i reminds the two-states variable of a lattice-gas model for the liquid-gas phase transition, where $\langle n_i \rangle$ is the order parameter. However, here by definition it is $n_i = n_j$, $\forall i \neq j$, hence $\langle n_i \rangle = n_i$ in any configuration and the order parameter, as we will discuss, is a more complex function of the molecular configuration. The index n_i is introduced in this model to check the first condition for the HB formation in a simple way. As we will discuss, there is also a second condition.

FS, following previous works [34], make the reasonable assumption that the formation of HBs is the primary source of local density fluctuations. A water molecule fully bonded to its hydration shell, formed by other four water molecules in a tetrahedral configuration, occupies approximately the same volume as a hydrated water molecule with no HBs and larger coordination number [137]. Therefore, to each HB we can associate a proper volume v_{HB} given by 1/4 of the proper volume fluctuation in the hydration shell. A reasonable choice for this parameter is $v_{\mathrm{HB}}/v_0 = 0.5$, equal to the average volume increase between (high density) ices VI and VIII and (tetrahedral, low density) ice Ih. Hence, the total volume V_{tot} occupied by the system increases linearly with the number of HBs, N_{HB}, i.e.,

$$V_{\mathrm{tot}} \equiv V + N_{\mathrm{HB}} v_{\mathrm{HB}}. \tag{1.1}$$

While V is the volume without HBs, used for partitioning the system, and is homogeneously distributed among the N water molecules, V_{tot} includes the local heterogeneities in the density field due to the HBs.

The FS Hamiltonian is by definition

$$\mathcal{H} \equiv \mathcal{H}_{\mathrm{vdW}} + \mathcal{H}_{\mathrm{HB}} + \mathcal{H}_{\mathrm{Coop}} \tag{1.2}$$

where the first term accounts for the van der Walls (dispersive) attraction and hard core (electron) repulsion between water molecules and is

$$\mathcal{H}_{\mathrm{vdW}} = \sum_{i,j} U(r_{ij}), \tag{1.3}$$

summed over all the water molecules i and j at O–O distance r_{ij}, with

$$U(r) \equiv \begin{cases} \infty & \text{if } r \leq r_0 \\ 4\epsilon \left[\left(\frac{r_0}{r}\right)^{12} - \left(\frac{r_0}{r}\right)^{6} \right] & \text{if } r_0 \leq r \leq 25 r_0 \\ 0 & \text{if } r > 25 r_0, \end{cases} \tag{1.4}$$

a truncated Lennard-Jones (LJ) interaction with $\epsilon \equiv 5.8$ kJ/mol, close to the estimate based on isoelectronic molecules at optimal separation $\simeq 5.5$ kJ/mol [138].

The LJ potential is a function of the continuous variable r and it is truncated at large distance and at short distance for numerical efficiency. Our previous analysis [126] shows that both truncations do not affect the results. This interaction is necessary to reproduce the liquid-gas phase transition, and dominates above the liquid-gas spinodal temperature.

Because the formation of HBs, does not affect the distance r between a molecule and those in its hydration shell [137], the van der Waals interaction is not affected by the HBs. This observation is crucial to state that the FS model is not mean-field [2].

The second term accounts for the directional, covalent [139], short-range two-bodies component of the HB, with $J = 0.5 \times 4\epsilon \simeq 11$ kJ/mol, close to the estimate from the optimal HB energy and a HB cluster analysis [40]

$$\mathcal{H}_{\text{HB}} = -J \sum_{<i,j>_{n.n.}} n_i n_j \delta_{\sigma_{ij},\sigma_{ji}} = -J N_{\text{HB}}, \tag{1.5}$$

where N_{HB} is the number of HBs, with the sum performed over nearest neighbors (n.n.) pairs of water molecules i and j at a distance $r \leq r_{\max}$ [140, 141] (i.e., with $n_i n_j = 1$) and in the same *bonding state* ($\delta_{\sigma_{ij},\sigma_{ji}} = 1$), where $\sigma_{ij} = 1, \ldots, q$ is the *bonding variable* of molecule i facing the n.n. molecule j, with $\delta_{ab} = 1$ if $a = b$, 0 otherwise. While the factor $n_i n_j = 1$ enforces the first condition for the HB formation that we already discussed, the factor $\delta_{\sigma_{ij},\sigma_{ji}} = 1$ represents the second condition once we set $q = 6$. Indeed, consistent with Debye-Waller factors estimates [141, 142], a HB is broken if the angle between two oxygen atoms and the intermediate hydrogen atom is $\widehat{\text{OOH}} > 30°$, or $\widehat{\text{OOH}} < -30°$. Hence, only 1/6 of the entire range of possible values $[0, 360°]$ of the $\widehat{\text{OOH}}$ is associated to a bonded state, as required by $\delta_{\sigma_{ij},\sigma_{ji}} = 1$ with $q = 6$. Furthermore, each HB formation leads to an entropy decrease equal to $-k_B \ln 6$ (k_B is the Boltzmann constant) that is consistent with the entropy loss due to the formation of a HB. In its tetrahedral configuration, a water molecule has 4 n.n. in its hydration shell. Therefore, in the FS model each molecule has 4 bonding variables, both in 2D and 3D, regardless of the number of n.n.

The third term of the Hamiltonian accounts for the HB cooperativity due to many-body correlations [74] that lead to the local order in the hydration shell [137] and are a consequence of the quantum nature of the HB [74, 143, 144]. This term is modeled in classical atomistic potentials with a long range dipolar interaction, for example in Ref. [145]. In the FS Hamiltonian, it is modeled as a five-bodies term in which each molecule i is interacting with all its hydration shell

---

[2] We acknowledge the late Professor David Chandler for noticing this.

$$\mathcal{H}_{\text{Coop}} = -J_\sigma \sum_i n_i \sum_{(k,l)_i} \delta_{\sigma_{ik},\sigma_{il}} \tag{1.6}$$

where $(l,k)_i$ indicates each of the six different pairs of the four indexes σ_{ij} of the molecule i and J_σ is the extra energy-gain provided by each HB when it is locally ordered with the other HBs in the hydration shell. A direct experimental evaluation of such a term is not available, however it can be estimated by attributing to it [146] the 3 kJ/mol increase in strength of the HBs in ice Ih with respect to liquid water [147]. Considering that each HB participates to three terms in Eq.(1.6), we can estimate the value of J_σ to be $\simeq$ 1.0 kJ/mol. Because $J_\sigma \ll J$, this term is relevant only when $N_{\text{HB}} \gg 1$. This asymmetry between the two components of the HB interaction is necessary for the model to represent water.

In the NPT ensemble, the partition function of the system is

$$Z(T,P) \equiv \sum_{\{\sigma\}\{v\}} e^{-H/k_BT}, \tag{1.7}$$

where the sum is over all the possible configurations of bonding variables $\{\sigma\}$ and cell volumes $\{v\}$, and

$$H \equiv \mathcal{H} + PV_{\text{tot}} \equiv \mathcal{H}_{\text{vdW}} - J_{\text{eff}} \sum_{\langle i,j \rangle} n_i n_j \delta_{\sigma_{ij},\sigma_{ji}} \mathcal{H}_{\text{Coop}} + Pv \tag{1.8}$$

is the enthalpy and where $J_{\text{eff}} \equiv J - Pv_{HB}$ is the effective interaction between σ-variables of n.n. molecules that depends on P.

Eq. (1.7) can be rewritten as

$$Z(T,P) = \sum_{\{v\}} e^{-[U(v)+Pv]/k_BT} \times Z_{\{\sigma\}}, \tag{1.9}$$

where

$$\begin{aligned} Z_{\{\sigma\}} &\equiv \textstyle\sum_{\{\sigma\}} \; e^{(J_{\text{eff}}/k_BT)\sum_{\langle i,j\rangle} n_i n_j \delta_{\sigma_{ij},\sigma_{ji}}} \times e^{(J_\sigma/k_BT)\sum_i n_i \sum_{(k,l)_i} \delta_{\sigma_{ik},\sigma_{il}}} \\ &= \textstyle\sum_{\{\sigma\}} \; \prod_{\langle i,j\rangle} \left[1 + \left(e^{(J_{\text{eff}}/k_BT)} - 1\right) n_i n_j \delta_{\sigma_{ij},\sigma_{ji}}\right] \\ &\quad \times \textstyle\prod_{i=1}^N \prod_{(k,l)_i} \left[1 + \left(e^{(J_\sigma/k_BT)} - 1\right) n_i \delta_{\sigma_{ik},\sigma_{il}}\right]. \end{aligned} \tag{1.10}$$

Here $\prod_{\langle i,j\rangle}$ runs over all the n.n. molecules j of the molecule i, $\prod_{i=1}^N$ runs over all molecules and $\prod_{(k,l)_i}$ extends over all the six pairs of bonding variables of a specific molecule i.

Due to its simplicity, it is possible to perform a thorough analysis of how the macroscopic properties of the model depend on its limited number of parameters, each describing a molecular mechanism, both with theoretical calculations under mean field assumptions and with MC simulations [40, 71, 72, 115, 117, 120, 124, 125, 131, 133, 135, 148–151].

1.3 Dynamic behavior of the FS water monolayer

1.3.1 Two dynamic crossovers in FS water

Data on protein hydration water and confined water show the presence of a dynamical crossover from a non-Arrhenius to an Arrhenius regime [120] at ~ 220 K. For example, this crossover is found in the translational correlation time of water molecules hydrating lysozyme proteins [152] or in the structural relaxation time of water confined in silica pores [65, 153]. Also, simulations of the TIP5P water model show a dynamic crossover from a non-Arrhenius to an Arrhenius regime in the diffusivity of water hydrating lysozyme and DNA [91]. As this crossover takes place at much higher temperatures than T_G, the so-called *glass transition temperature*, Kumar *et al.* discard any relation with the glass state. According to these numerical and experimental evidences, a possible hypothesis, among others [154, 155], for the origin of this crossover is the local rearrangement of the HB network at low temperatures [117, 156].

In particular, MC simulations of the FS model display a dynamic crossover from a non-Arrhenius to an Arrhenius regime that is a consequence of the local rearrangement of the water HBs [117, 119]. Kumar *et al.*, by mean field calculations and MC simulations of the FS water monolayer, estimate the orientational correlation time, i.e. the relaxation time τ of the autocorrelation function $C_S(t) \equiv \left\langle S_i(t)S_i(0)\right\rangle / \left\langle S_i^2\right\rangle$ where $S_i \equiv \sum_j \sigma_{ij}/4$, for an FS monolayer. The physical meaning of S_i is the total bond ordering of the i-th water molecule, and τ is defined as the time at which $C_S(t)$ decays by a factor $1/e$. Kumar *et al.* find a non-Arrhenius regime at high-T upon cooling at constant pressure, where τ can be fitted with the Vogel-Fulcher-Tamman (VTF) function

$$\tau^{\mathrm{VTF}} \equiv \tau_0^{\mathrm{VTF}} \exp\left[\frac{T_1}{T-T_0}\right], \tag{1.11}$$

with τ_0^{VTF}, T_0 and T_1 fitting parameters. At low-T, τ displays Arrhenius behavior, $\tau = \tau_0 \exp[E_A/k_BT]$, where τ_0 is the limiting time at high-T and E_A is the T-independent activation energy. The crossover occurs at the same temperature where the specific heat displays a maximum. The authors discuss that they interpret this dynamic crossover as an effect of the breaking and reorientation of HBs, leading to a more tetrahedral structure for the HB network at low-T. In Ref. [117], Kumar *et al.* also investigate how the cooperative term affects the dynamics. At constant $J = 0.5$, they compare the results for $J_\sigma = 0.05$ and $J_\sigma = 0$, corresponding to the LLCP scenario and the SF scenario, respectively, as we will discuss in the following. In both scenarios, they find that (i) the crossover time τ_C is approximately P-independent, (ii) the Arrhenius activation energy $E_A(P)$ decreases upon increasing P and (iii) the temperature $T_A(P)$, at which τ reaches a fixed macroscopic time $\tau_A \geq \tau_C$, decreases upon increasing P. Furthermore, they show that (iv) while $E_A/(k_B T_A)$ increases upon pressurization in the LLCP scenario, it remains constant in the SF scenario.

These new predictions have been tested in experiments in a protein hydration layer [120,157]. In particular, quasi-elastic neutron scattering (QENS) [120,157] verifies that the predictions (i)-(iii) are correct for a water monolayer hydrating lysozyme. Nevertheless, these experiments cannot settle which among the LLCP and SF scenario is satisfied by proteins' hydration water based on prediction (iv) because the difference between the two forecasted behaviors lies within the error bars of the measurements [120].

Further research on water monolayers at lower T, surprisingly, has shown the presence of not one, but two dynamic crossovers [135]. Mazza *et al.* measure the dielectric relaxation time of water protons τ_{WP}, as it is sensitive to breaking and formation of HBs [158], in a water monolayer hydrating lysozyme proteins, and compare it with MC simulation of an FS water monolayer. In both cases, the authors find two crossovers at ambient pressure. The first is at $T \sim 252$ K, as previously reported by Kumar and coworkers. The second is at $T \sim 181$ K.

They find that the high-T crossover is from VTF behavior to another VTF behavior. This can be interpreted as a change in the diffusion of water protons, between a high-T diffusive regime to a low-T subdiffusive regime.

At the second dynamic crossover, τ_{WP} changes from a VTF regime to an Arrhenius regime, corresponding to the rearrangement of the HB network structure.

The experimental measurements of τ_{WP} compare well with FS water calculations of the S_i relaxation time from MC simulations τ_{MC}. Mazza and coworkers consider the autocorrelation function

$$C_M(t) \equiv \frac{1}{N} \sum_i \frac{\langle S_i(t_0+t) S_i(t_0) \rangle - \langle S_i \rangle^2}{\langle S_i^2 \rangle - \langle S_i \rangle^2}, \tag{1.12}$$

which decays to 0 as $t \to \infty$ and, by definition, $C_M(0) = 1$. Following Ref. [117], τ_{MC} is defined as $C_M(\tau_{\mathrm{MC}}) = 1/e$. They find two dynamic crossovers and relate them to the presence of two specific heat maxima: the high-T weak maximum and the low-T strong maximum. At high-T, the C_P weak maximum occurs when the fluctuations of N_{HB} are maximum. At low-T, the C_P strong maximum is due to the maximum in the fluctuations of the cooperative term (Eq. 1.6) of the Hamiltonian. Then, the high-T crossover is associated with the formation of the HB network and the low-T crossover is due to the local rearrangement of the HBs in an ordered structure, as close as possible to tetrahedral in 3D.

1.3.2 Effect of pressure and temperature on the dynamics

Protein hydration water undergoes a liquid-glass transition (LGT) with the glass state characterized by a huge increment of viscosity and the freezing of long-range translational diffusion [159]. The macroscopic structural arrest of the glass state emerges from the slowing down of nearest neighbors HBs dynamics. Doster and coworkers performed neutron scattering experiments on myoglobin at hydration level $h = 0.35$ $\mathrm{g_{H_2O}/g}$, as this technique allows for monitoring displacements at the micorscopic scale. They measure the incoherent intermediate scattering functions $I(q,t)$ at $T = 320$ K (above the LGT but relevant for translational diffusion) and for scattering vector $0.4 \leq q/\text{Å}^{-1} \leq 2$ [160]. They find a two-step time-decay in $I(q,t)$ that, at high-q and long times t, can be fitted to a stretched exponential

$$C(t) = C_0 \exp[-(t/\tau_0)^\beta], \tag{1.13}$$

where τ_0 is the correlation time, $0 < \beta \leq 1$ is the stretched exponent and C_0 is a normalization factor. For large q their results show that $0.3 \leq \beta \leq 0.4$.

When they measure $I(q,t)$ at constant $q = 1.8\text{Å}^{-1}$ and $180 \leq T/\text{K} \leq 320$, as the temperature decreases, the two-step decay turns into a plateau, which leads to a relaxation time that exceeds the observation time [159].

Calculations from MC calculations of the FS model are consistent with these two experimental results. In particular, to study the microscopic origins of the complex dynamics of water on low-hydrated proteins, de los Santos and Franzese consider a monolayer of FS water adsorbed on a generic inert substrate at 75% hydration [121, 126, 132]. At such hydration level, adsorbed water molecules are restricted to diffuse on a surface geometry with an up to four coordination number. They calculate $C_M(t)$ at different T and P [121] finding that at high pressure $P \geq 1\epsilon/v_0$ the correlation function decays exponentially for any T, allowing the system to equilibrate easily. At these pressures, the HB network is inhibited inducing rapid dewetting and large dry cavities with decreasing temperature.

At lower pressure $P = 0.7\epsilon/v_0$ and low-T the behavior of $C_M(t)$ can be fitted with a stretched exponential function, with no strong increase in the correlation time as T decreases (Figure 1.1). The authors associate this behavior with (i) the rapid ordering of the HBs that generates heterogeneities and (ii) with the lack of a single timescale due to the vicinity of the liquid-liquid critical point, as we will discuss in the next section.

At even lower pressures, the gradual formation of the HB network, starting at higher T, is responsible for the dynamic slowing down as T decreases and for the dynamical arrest at $(P,T) = (0.1\epsilon/v_0, 0.05\epsilon/k_B)$, with an increase in τ_0 of more than four orders of magnitude, as in a glass. Under these conditions, the dewetting process is strikingly different, with the formation of many small cavities.

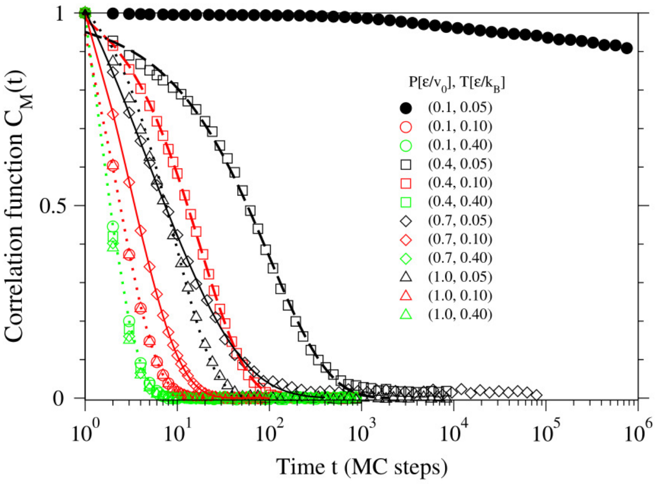


Figure 1.1: The correlation function $C_M(t)$ for pressures $P \leq 1.0\epsilon/v_0$ and temperatures $T \leq 0.4\epsilon/k_B$. $C_M(t)$ decays exponentially (dotted lines) at $P = 1.0\epsilon/v_0$ for all T (triangles) and at $P < 1.0\epsilon/v_0$ for any $T \geq 0.4\epsilon/k_B$ (not shown). For $T \leq 0.1\epsilon/k_B$ and $P = 0.7\epsilon/v_0$ (diamonds, continuous line) or $P = 0.4\epsilon/v_0$ (squares, dashed line), $C_M(t)$ can be described by a stretched exponential, with a exponent β that decreases as T is lowered. At low pressure $P = 0.1\epsilon/v_0$ (circles), $C_M(T)$ is exponential for $T \geq 0.1\epsilon/k_B$ and non-exponential at $T = 0.05\epsilon/k_B$ (solid circle). Figure reprinted from Ref. [121] "Dynamically slow processes in supercooled water confined between hydrophobic plates". Copyright (2009) IOP Publishing. Reproduced with permission. All rights reserved.

Comparison between FS results and experiments [159, 160] show that the complex dynamic behavior of protein hydration water at low h can be well reproduced by solely taking into account the dynamics of the HB network. The LGT emerges as a consequence of the slowing of the HBs by decreasing T, that at extreme low-T results in a dynamic arrest of the system.

1.3.3 The diffusion anomaly

Experiments and simulations of the diffusion of confined water show controversial results. For water confined in NaX and NaA zeolites and for T between 310 K and 260 K, experiments observe a reduction of two orders of magnitude of the translational diffusion coefficient D in respect to the bulk case [161]. Other researchers show that D decreases upon increasing the confinement in either hydrophilic [162] or hydrophobic conditions [162, 163]. However, for confined water in carbon nanotubes with diameter smaller than 2 nm, experiments find an extremely fast mass transport [164]. Furthermore, water confined in smooth graphene capillaries shows a fast flow (of ~ 1 m/s) that is enhanced if the height of the channel can accommodate only a few water layers [165]. The authors relate the fast flow to the great capillary pressure, and the enhancement to the increased structural order of nanoconfined water.

Several models, e.g. [166, 167], can reproduce numerically the diffusion anomaly, but they display a variety of different results still controversial. Classical molecular dynamics simulations of SPC/E water in hydrophilic MCM-41 or Vycor show that the mobility of water molecules decreases as the hydration level is lowered [63]. This can be interpreted as a consequence of the greater proportion of molecules bonded to the surface at low hydration level. A decrease of diffusion by two orders of magnitude with respect to the bulk is found also in simulations of TIP5P water between hydrophobic, smooth, planar plates, for which diffusion is normal in the direction orthogonal to the walls, while it is anomalous for the parallel direction [168, 169]. However, first principle molecular dynamics simulations of SPC/E water confined in graphene sheets and carbon nanotubes show a faster diffusion with respect to the bulk case [170], possibly due to weaker hydrogen bonding at the interface.

In order to shed light into this controversy, de los Santos and Franzese perform MC simulations of a water monolayer confined in a smooth slit pore at partial hydration. Their results provide a description of the origins of the diffusion anomaly [126, 132] in terms of *Cooperative Rearranging Regions* (CRR) of water. They calculate the diffusion coefficient $D_{||}$ parallel to the walls by means of Einstein's formula

$$D_{||} = \lim_{t\to\infty} \frac{\langle |\vec{r_i}(t+t_0) - \vec{r_i}(t)|^2 \rangle}{4t} \tag{1.14}$$

where $\vec{r_i}$ is the projection of the position of molecule i onto the plates. The average $\langle\cdot\rangle$ is over all molecules and different times t_0. The analysis of $D_{||}$ shows the presence of maxima and minima along isotherms at high temperature (Figure 1.2).

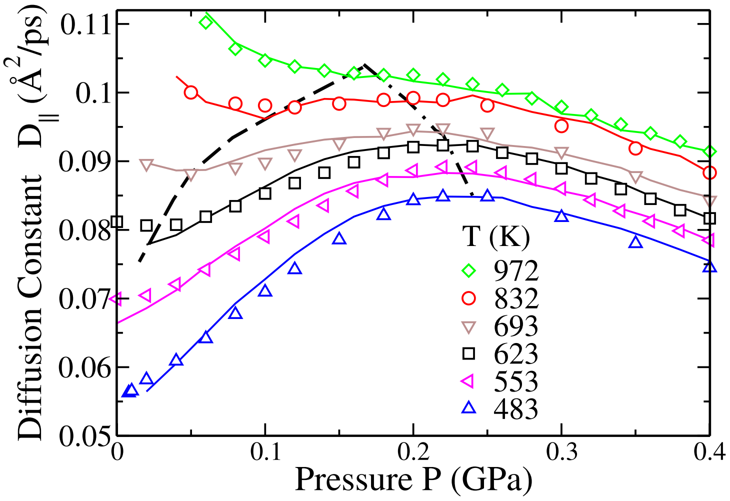


Figure 1.2: Diffusion coefficient $D_{||}$ from MC simulations (symbols) as a function of pressure along isotherms. For $T < 972$ K, $D_{||}$ has maxima (dotted-dashed line) and minima (dashed line). Solid lines are from $W_{\nu,\mu}$ calculations (defined in the text). Reprinted figure with permission from Ref. [132]. Copyright (2012) by the American Physical Society.

The authors describe the anomaly in terms of the joint probability

$$W_{\nu,\mu}(P,T) \equiv \mathcal{P}_{\rm F}^{\nu}\ \mu\mathcal{P}_b \frac{1}{Z}\exp[-H(T,P)/(k_B T)] \tag{1.15}$$

of finding ν molecules with a free cell available for diffusion within a region with a number $\mu\mathcal{P}_b$ of HBs, where Z is the partition function, $\mathcal{P}_{\rm F}$ and $\mathcal{P}_b$ are the probability for each cell to have a free n.n. and the probability for each HB to be formed, respectively, with $\mathcal{P}_{\rm F} \equiv \langle n_{\rm F}\rangle/4$, $\mathcal{P}_b \equiv \langle n_{\rm HB}\rangle/4$, $\langle n_{\rm F}\rangle$ the average number of free n.n. cells per molecule and $\langle n_{\rm HB}\rangle$ the average number of HBs per molecule.

The authors find that $D_{||}$ is proportional to $W_{\nu,\mu}$, implying that the diffusion is dominated by the cooperativity of water (Figure 1.3 main panel). The resulting value for $\nu = 12.5 \pm 0.5$ suggests that diffusion requires a CRR that reaches ~ 3.5 molecules (Figure 1.3 lower inset). Thus, the FS model clarifies that the diffusion anomaly, at constant T by increasing P, originates from the competition between, on the one hand, the increase of free volume $\langle n_{\rm F}\rangle$ and the decrease of the energy cost for a molecule to move due to the reduction of $\langle n_{\rm HB}\rangle$ (Figure 1.3 upper inset), and, on the other hand, the decrease of free volume due to the increase of density. The diffusion is favored by pressurizing at constant T until the HB formation is unfavorable for enthalpic reasons, giving origin to the $D_{||}$ maxima.

The diffusion coefficient $D_{||}$ correlates to the phase diagram of FS water (Figure 1.4). The loci of $D_{||}^{\max}$ and $D_{||}^{\min}$ along isotherms lie between the temperature of maximum density (TMD) line and the liquid-gas spinodal and the resulting constant

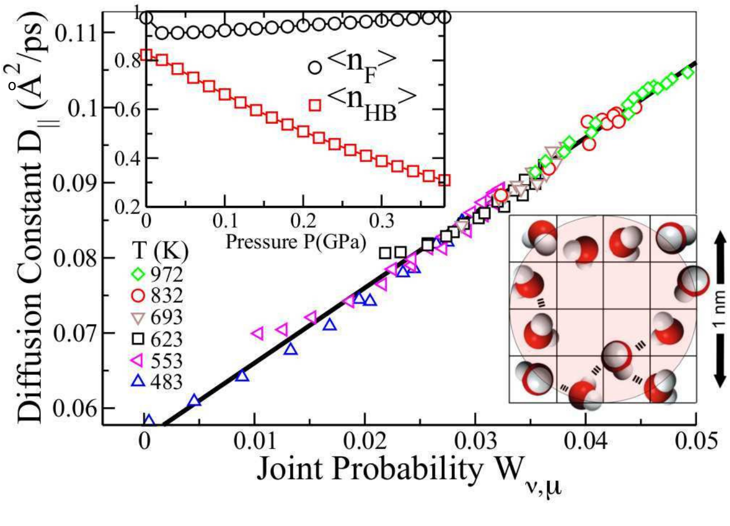


Figure 1.3: Upper inset: Average number of free n.n. cells around a molecule $\langle n_{\rm F}\rangle$ and average number of HBs formed by a molecule $\langle n_{\rm HB}\rangle$ at $T = 693$ K. There is a discontinuity in $\langle n_{\rm F}\rangle$ at low P corresponding to the gas-liquid phase transition. At high P they are both monotonic. Main panel: Linear dependency of $D_{||}$ vs. $W_{\nu,\mu}$ along the isotherms represented in Figure 1.2. Lower inset: Example of a CRR (shaded) of about 1 nm size, with $\nu = 12$ molecules and $\mu < n_{\rm HB} > /4 = 5$ HBs. Reprinted figure with permission from Ref. [132]. Copyright (2012) by the American Physical Society.

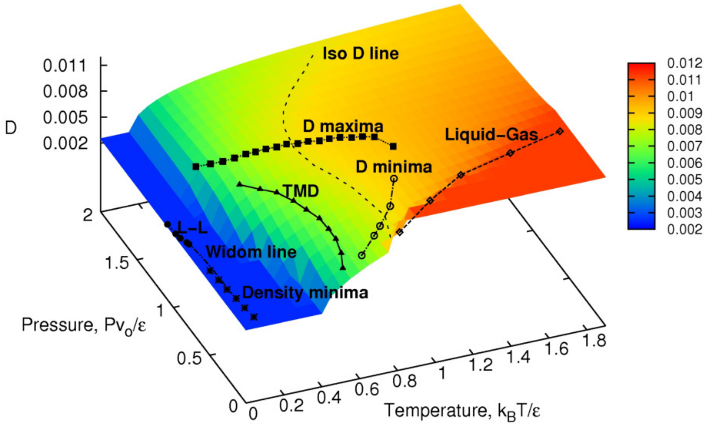


Figure 1.4: Phase diagram of a water monolayer nanoconfined between hydrophobic plates. Z-axis and color scale represent the diffusion constant $D_{||}$ for values $0.12 \geq D_{||}/(\text{Å}/\text{ps}^2) \geq 0.03$. At high T and low P there is the liquid-gas phase transition line (open diamonds) ending in the liquid-gas critical point (diamond symbol at highest T), as discussed in section 5. The loci of isothermal $D_{||}$ extrema, $D_{||}^{\max}$ (solid squares), and minima, $D_{||}^{\min}$ (open circles), envelope the TMD line (solid triangles). Loci at constant $D_{||}$ (e.g., the dashed line marked as "Iso D") resemble in their reentrant behavior the water melting line. Reprinted with permission from Ref. [126]. Copyright (2011) American Chemical Society

$D_{||}$ lines resemble the melting line of bulk water. In the deep supercooled region there is a subdiffusive regime due to the increment of the relaxation time (that turns eventually in a dynamincal arrest of the system at low P and T, as we have already discussed). This is a consequence of the HB network dynamics and can be associated with the observed amorphous glassy water [23].

1.4 Phase Diagram of the FS water monolayer

The phase diagram of the FS model has been extensively studied in the case of a water monolayer by analytic [71, 72, 114] and numerical methods [116]. By changing the model's parameters [122] Stokely et al. reveal the relations among the different scenarios for the anomalies of water [40].

In particular, Bianco and Franzese (Figure 1.5) find in the T-P plane (i) the TMD line along isobars with negative slope at high P, as in the experiments, turning into a line with positive slope near $P = 0$ that (ii) asymptotically approaches the liquid-gas (LG) spinodal at $P < 0$. (iii) At the turning point the TMD is crossed by the line of minima of the isothermal compressibility along isobars, $K_T^{min}(T)$. (iv) At low T, the TMD line turns into a negatively-sloped line of temperature of minimum density along isobars, as suggested by experiments [47], avoiding crossing with the LG spinodal line. (v) The low-T turning point of the locus of extrema of density occurs where the line of (weak) minima of α_P along isobars, $\alpha_P^{wmin}(T)$, crosses it. This point is also where the line of specific heat (weak) maxima along isotherms, $C_P^{wMax}(P)$, is aiming for. (vi) The locus of (weak) maxima of compressibility along isobars, $K_T^{wMax}(T)$ turns, at its minimum P, into the locus of minima of K_T along isobars, $K_T^{min}(T)$ and, all together, they coincide within the error bar (not shown) with the locus of (weak) minima of α_P along isotherms, $\alpha_P^{wmin}(P)$. All the observations (i)-(vi) are consistent with thermodynamic relations [171], confirming the correctness of the numerical calculations performed in Ref [134] and showing that the model recovers what has been shown to be valid also in atomistic models [171].

However, thanks to its mapping onto a percolation formulation [123], the FS water can be easily simulated adopting a fast (cluster) MC algorithms [124] that allows us to explore in detail the deeply supercooled states [115] and the high pressure region [120, 125, 148]. In particular, in the deeply supercooled region, Mazza et al. [131, 135] and Bianco et al. [133,134] reveal the presence of strong extrema, for the response functions C_P, K_T and α_P, occurring at temperatures lower than those for the (weak) extrema known also from atomistic water models. In fact, the deeply supercooled region is inaccessible to atomistic water models because their dynamics slows down by orders of magnitude, transforming the supercooled liquids into glasses [77, 78, 87].

Mazza et al. [135]. show that the weak and strong maxima of the response functions correspond to the two dynamic crossovers discussed in section 1.3.1. In this way they establish a connection between thermodynamics and dynamics that extends up to moderate supercooling at low P.

As initially observed by Franzese and coworkers [115,117,119,120,131,134,148], all the loci of extrema of response functions converge toward a point where the extrema reach their maximum values, as expected at a critical point in a finite-size system. This point corresponds to the LLCP, as shown by analyzing the fluctuations at extreme supercooling [135] and extreme pressure [131]. In particular, as described with more detail in the next section, Bianco et al. study the liquid-liquid critical point region showing that it belongs to the 2D Ising universality class [134].

This result is consistent with those for ST2 water [77, 78, 172], TIP4P/2005 water [97, 173], but not with those for the three-body interactions mW model [81]. The difference with the mW model is understood as a consequence of the fact that the non-ideality of mixing is entropy driven, instead of energy-driven as in the models with the LLCP, and it is not strong enough to induce the liquid-liquid phase separation [174, 175]. Furthermore, although pair-wise additive atomistic models have the LLCP, upon pressurization they do not show fluctuations as large as those extrapolated from experiments [29, 176, 177], while models with many-body interactions, including the FS model, do, revealing an important difference between those water models with explicit many-body interactions and those without it.

Above the critical pressure, at very low T, the FS water monolayer presents the

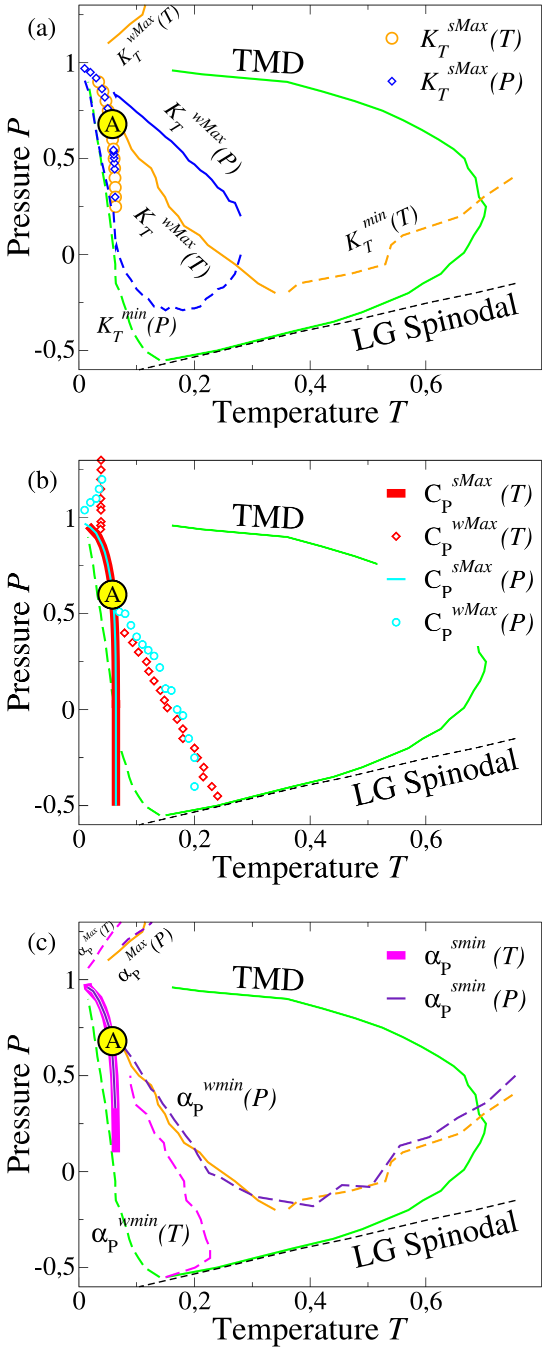


Figure 1.5: Phase diagram of a water monolayer. In all panels, the black dashed line represents the LG spinodal, while all the colored dashed lines turning into solid lines of the same color represent loci of minima and maxima, respectively, of different quantities: e.g., in green the temperature of minimum density and the TMD. Each panel focuses on one response function: (a) on K_T along isobars (isotherms) in orange (blue), with dashed (continuous) lines for weak minima (maxima) and symbols for strong maxima; (b) on C_P along isobars (isotherms) in red (turquoise), with lines (symbols) for strong (weak) maxima; (c) on α_P along isobars (isotherms) in magenta (violet), with dashed (continuous) lines for weak (strong) minima. All the loci of extrema of response functions converge toward the large circle with label A corresponding to the LLCP. Reprinted with permission from Ref. [134].

first-order LLPT between two liquid phases with different energy and local structure. A similar result is found also in atomistic models that adopt a two-state description of water [22,88,174,175,178,179]. In the FS model, the phase transition occurs between a HDL with a disordered HB network and high energy, and a LDL with an ordered HB network and low energy [116].

1.4.1 The critical point analysis

Bianco and Franzese study the universality class of the LLCP [134]. To define the correct order parameter M for the LLCP, they first analyze the free-energy landscape. At the finite-size LLCP the landscape has two equivalent minima, each with a basin of attraction, with different energy and density, corresponding to two coexisting HDL and LDL phases (Figure 1.6.a). They find that the two basins are separated by a free energy barrier $\sim k_BT$. Hence, as expected near a critical point, under these conditions the system has enough thermal energy to cross the barrier between the two basins.

The order parameter M is by definition the reaction coordinate between the two minima, i.e. the linear combination of energy and density $M \equiv \rho + sE$, as in the mixed-field approach [180–182]. M is related to the diagonal connecting the centers of the two wells, as its probability distribution must be symmetric at the LLCP. By

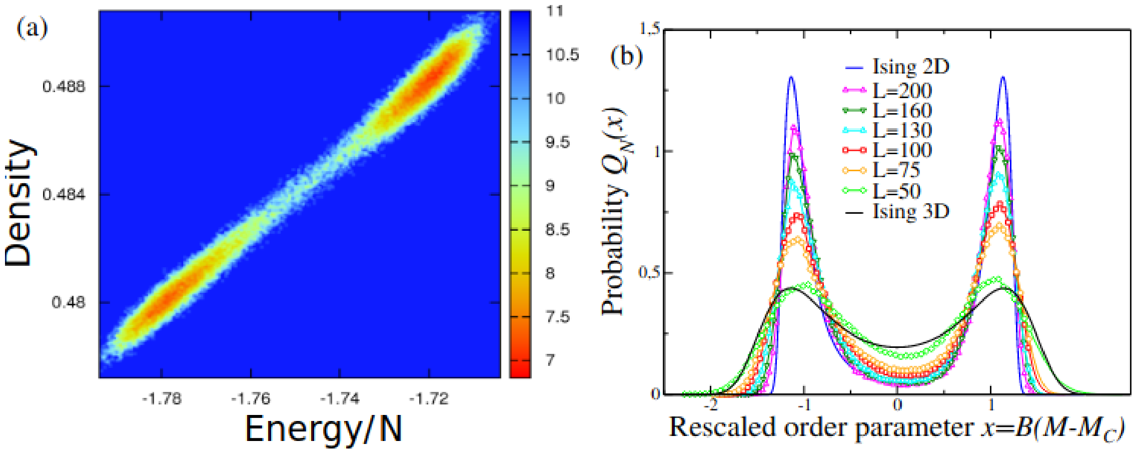


Figure 1.6: The free-energy landscape in the vicinity of the LLCP. (a) The Gibbs free energy G in units of k_BT is represented as heat-map in the energy-density plane, showing two basins separated by a barrier $\Delta G \sim k_BT$, as expected near a critical point. The phase transition is described by the order parameter M given by the reaction coordinate (a straight line in the plane) between the two minima. Reprinted with permission from Springer Nature Customer Service Centre GmbH: Springer Food Biophysics Ref. [180], Copyright (2013). (b) The size-dependent probability distribution, Q_N, of the rescaled order parameter M at the LLCP approximates the 2D Ising universality class (blue solid line) as the size L of the system increases. Here, B is a scaling factor and the black reference distribution holds for the 2D Ising model. Reprinted with permission from Ref. [134].

performing a finite-size scaling analysis at fixed sample thickness $h = 0.5$ nm, and by varying the number of water molecules from 2500 to 40000 at fixed density, Bianco and Franzese find that the FS monolayer in the thermodynamic limit displays a LLCP belonging to the 2D Ising universality class.

The probability distribution of M shows a crossover from 3D to 2D Ising universality class as the number of water molecules in the monolayer increases (Figure 1.6.b). The system crosses from 3D to 2D behavior for $L/h > 50$, where L is the system lateral size, while for normal liquids this crossover takes place for $L/h > 5$. This can be interpreted as a consequence of the high cooperativity, and the low coordination number, of the water molecules. For small L, the strong cooperativity at the LLCP increases the HBs fluctuations, resulting in a (3D-like) probability distribution for M broader than the one in 2D. The water coordination number–four both in 3D and 2D–emphasizes the effect because it reduces the fluctuations differences between the two cases when the system is small [134, 180].

1.4.2 The Widom line

The Widom line (WL) is defined as the locus of maxima of the statistical correlation length ξ, emanated from the critical point as analytic continuation of the first order transition line, and spanning into the supercritical region [115, 183, 184]. To the best of our knowledge, the expression "Widom line" was first used in a work by Xu and coworkers in 2005 as the locus where the lines of the maxima for different response functions asymptotically converge approaching the critical point from the supercritical region [185].

Close to the critical point the thermodynamic response functions can be expressed as power law functions of ξ. Since ξ exhibits a maximum along the WL, the loci of maxima of thermodynamic response functions converge toward the WL on approaching the critical point and near (T_c, P_c) are often used as *proxies* for the Widom line [173, 186].

Following ref. [134], within the FS model it is possible to define the correlation length ξ by means of the spatial correlation function $G(r)$ defined

$$G(r) = \frac{1}{4N} \sum_{|\vec{r_i}-\vec{r_l}|=r} \left[\langle \sigma_{ij}(\vec{r_i}) \sigma_{lk}(\vec{r_l}) \rangle - \langle \sigma_{ij} \rangle^2 \right]. \tag{1.16}$$

Far from the critical region, at high T and for any P, $G(r) \sim e^{-r/\xi}$ at high temperatures in a broad range of pressures. By decreasing T, the correlation function can be written as $G(r) \sim e^{-r/\xi}/r^{d-2+\eta}$, being d the system dimension and η a (critical) positive exponent. The correlation length, computed along isobars, exhibits a maximum, ξ^{Max} at all the P explored. The T–P locus of ξ^{Max} overlaps with the LLPT at high P, and with the locus of strong maxima of the specific heat at lower P. This line, at low P, identifies the WL, with a large slope in the T–P plane (it is almost

T independent), consistent with extrapolations from experiments [176, 187]. Previous experiments [188], simulations [120, 173, 183, 189] and analytic calculations [115], revealed a WL with a larger T-dependence. A possible argument for this incongruence is that several approaches could identify the WL along the lines of weak extrema of the thermodynamic response functions predicted by the FS model. In a recent work, Bianco and Franzese have used a percolation approach [190, 191] to identify the regions (clusters) of statistically correlated water molecules [123]. According to this mapping, the FS model undergoes to a percolation transition along a T–P locus that is numerically consistent with the line of extrema of ξ^{Max}. A detailed cluster analysis reveals that at high P, along the LLPT line, the percolation transition is due to the building up of HB network. At lower P the origin of the percolation transition is related to the local tetrahedral reordering of the water molecules. This percolation approach allows to compute the connectivity length ξ_C, that measures the average size of the connected clusters. This quantity shows maxima along the T–P line of ξ^{Max}. Nevertheless, as pointed out by the authors, the equivalence $\xi \sim \xi_C$ can be guaranteed only in the critical region [123, 190, 191].

1.4.3 The effect of cooperativity in the phase diagram

To conclude this section, we discuss how the phase diagram of liquid water depends on the ratio between the directional and cooperative component of the HB, respectively encoded by the parameters J and J_σ of the FS model. Stokely et al. have shown that, by tuning J_σ/J the FS model accounts for most of the scenarios proposed for the origin of the water anomalies.

By setting $J_\sigma = 0$, meaning that there is no cooperativity effect among the HBs, the FS model reproduces the SF scenario independently of the J value. By increasing $J_\sigma/J > 0$ the model presents a LLPT. If J_σ increases keeping J constant, the LLPT terminates in a LLCP. By increasing the ratio $J_\sigma/J > 0$ the LLCP moves from positive pressures and low temperatures to negative pressures and larger temperatures. By further increasing $J_\sigma/J > 0$ the LLPT cross the limit of stability of the liquid phase respect to the gas phase, consistent with the CPF scenario.

According to estimations made from experimental data on the structural and dynamical properties of liquid water [138, 146, 147, 192, 193], the hypothesis of the LLCP scenario, with the critical point occurring at positive pressure, seems to be the most probable for water.

1.5 The Bianco-Franzese model for water and protein interfaces

We motivated the introduction of the FS water model for its ability to equibrate large water samples, as required for simulations of biological systems, including a detailed description of HBs and cooperativity (Sec. 1.1). Here, we review the efforts by Bianco, Franzese and coworkers to extend the FS model to reproduce water and protein systems [5–7].

The BF model accounts for water-water interactions as in the FS model, adopting the same volume partition into cells, but assumes that some cells are occupied by protein residues. BF proteins are self-avoiding lattice polymers that keep sequence specificity. Although BF lattice proteins seem a serious oversimplification compared with real proteins, multiple lattice models have been proposed for macromolecules [194], proteins [195–197], hydrated ions [198], membranes [199], or bacteria nucleoids [200]. This sort of model takes advantage of its versatility and the simplified account of the physical interactions, easing the description of the mechanisms underlying the behavior of the systems.

The Hamiltonian of the BF model can be separated into residue-residue, residue-water, and water-water terms:

$$\mathcal{H} \equiv \mathcal{H}_{\mathrm{R,R}} + \mathcal{H}_{\mathrm{R,w}} + \mathcal{H}_{\mathrm{w,w}}. \tag{1.17}$$

The protein interactions are given by

$$\mathcal{H}_{\mathrm{R,w}} + \mathcal{H}_{\mathrm{R,R}} \equiv -\sum_{i}^{N^{\mathrm{R}}} \left[\sum_{j}^{N^{\mathrm{w}}} C_{ij} S_{ij}^{\mathrm{w}} + \sum_{k \neq i}^{N^{\mathrm{R}}} C_{ik} S_{ik} \right], \tag{1.18}$$

where N^{R} is the number of protein residues, N^{w} is the number of water molecules, C_{ij} is a contact matrix, i. e. $C_{ij} = 1$ if the cells i and j are nearest neighbors and $C_{ij} = 0$ otherwise, and S_{ij}^{w} and S_{ij} are the water-residue and residue-residue interaction energy. In Ref. [6], Bianco et al. expressed S_{ij} in terms of a rescaled Miyazawa-Jernigan solvent-independent interaction matrix [201]. Regarding S_{ij}^{w}, they adopted the Doolitle hydropathy scale [202], assigning $S_{ij}^{\mathrm{w}} = J$ if the residue is hydrophobic and $S_{ij}^{\mathrm{w}} = 0$ otherwise.

Experiments and numerical calculations support the hypothesis that HBs at the hydration shell are more stable and correlated for longer times than those in the bulk [203–207]. The BF model defines the hydration shell as those water molecules that are first neighbors of protein residues. Hence, HBs that are formed at the hydration shell have different properties compared with bulk:

$$\mathcal{H}_{\mathrm{w,w}} \equiv \mathcal{H}_{\mathrm{w,w}}^{\mathrm{b}} + \mathcal{H}_{\mathrm{w,w}}^{\mathrm{h}}, \tag{1.19}$$

where $\mathcal{H}^{\mathrm{b}}_{\mathrm{w,w}} \equiv \sum_{ij} U(r_{ij}) - J^{\mathrm{b}} N^{\mathrm{b}}_{\mathrm{HB}} - J^{\mathrm{b}}_{\sigma} N^{\mathrm{b}}_{\sigma}$ is the Hamiltonian of bulk water as in the FS model (Eq. 1.2), and

$$\mathcal{H}^{\mathrm{h}}_{\mathrm{w,w}} \equiv - \left[J^{\mathrm{PHO}} N^{\mathrm{PHO}}_{\mathrm{HB}} + J^{\mathrm{PHO}}_{\sigma} N^{\mathrm{PHO}}_{\sigma} \right] - \left[J^{\mathrm{PHI}} N^{\mathrm{PHI}}_{\mathrm{HB}} + J^{\mathrm{PHI}}_{\sigma} N^{\mathrm{PHI}}_{\sigma} \right], \tag{1.20}$$

where $J^{\mathrm{PHO(PHI)}}$ and $J_{\sigma}^{\mathrm{PHO(PHI)}}$ substitute J^{b} and J^{b}_{σ} for the HBs formed at the hydrophobic (hydrophilic) interface. Analogously, the volume attributed to HB formation at the hydration shell is $v_{\mathrm{HB}}^{\mathrm{PHO(PHI)}}$ and the total volume

$$V \equiv Nr^3 + N^{\mathrm{b}}_{\mathrm{HB}} v^{\mathrm{b}}_{\mathrm{HB}} + N^{\mathrm{PHO}}_{\mathrm{HB}} v^{\mathrm{PHO}}_{\mathrm{HB}} + N^{\mathrm{PHI}}_{\mathrm{HB}} v^{\mathrm{PHI}}_{\mathrm{HB}}. \tag{1.21}$$

The water-water HBs at the hydrophobic hydration shell are stronger and more stable than those in the bulk [204–208]. The BF model accounts for this behavior setting $J^{\mathrm{PHO}} > J^{\mathrm{b}}$ and $J^{\mathrm{PHO}}_{\sigma} > J^{\mathrm{b}}_{\sigma}$, ensuring enthalpy compensation upon cold-denaturation [209]. In addition, the compressibility of water at the PHOB interface increases upon pressurization [205, 210–212]. The BF model assumes that the volume fluctuations associated to HB formation increase linearly in P

$$v^{\mathrm{PHO}}_{\mathrm{HB}} \equiv (1 - k \cdot P)\, v^{\mathrm{PHO}}_{\mathrm{HB},0}, \tag{1.22}$$

where $k > 0$ and $v^{\mathrm{PHO}}_{\mathrm{HB},0}$ corresponds to the volume change associated with HB formation at the PHO interface for $Pv_0/(4\epsilon) = 0$.

Finally, the BF model assumes that HBs at thy PHI interface behave as those in the bulk. Hence, $J^{\mathrm{PHI}} = J^{\mathrm{b}}$, $J^{\mathrm{PHI}}_{\sigma} = J^{\mathrm{b}}_{\sigma}$, $v^{\mathrm{PHI}}_{\mathrm{HB}} = v^{\mathrm{b}}_{\mathrm{HB}}$. For those HBs formed between a water molecule at a PHO interface and another molecule at PHI interface, the associated parameters are $\left(J^{\mathrm{PHO}} + J^{\mathrm{PHI}}\right)/2$, $\left(J^{\mathrm{PHO}}_{\sigma} + J^{\mathrm{PHI}}_{\sigma}\right)/2$, and $\left(v^{\mathrm{PHO}}_{\mathrm{HB}} + v^{\mathrm{PHI}}_{\mathrm{HB}}\right)/2$.

1.5.1 Applications of the BF model

In the first place, the BF model was applied to study protein denaturation [5, 7]. According to Hawley's theory, protein surfaces show elliptical iso-free energy contours in the P-T phase diagram [213]. Hence, proteins unfold upon heating, cooling, pressurization and depressurization [214–219]. Heat unfolding is well understood in terms of the increase of thermal fluctuations and entropy, depressurization unfolding has been experimentally observed [220], and high-P unfolding can be attributed to the loss of cavities in the native state [221]. The rationale for cold unfolding is debated [222], although it can be attributed to changes in the total free energy balance of the protein and solvent [223].

The BF model shows that the change of enthalpy and density fluctuations of the HB at the hydration shell is enough to predict the elliptical shape of protein stability regions (Fig. 1.7), consistently with Hawley's theory [213]. In particular, Bianco and Franzese attributed high-T unfolding to the increase of water and protein entropy and high-P unfolding to the decrease of enthalpy $H \equiv U + PV$ when PHO residues are

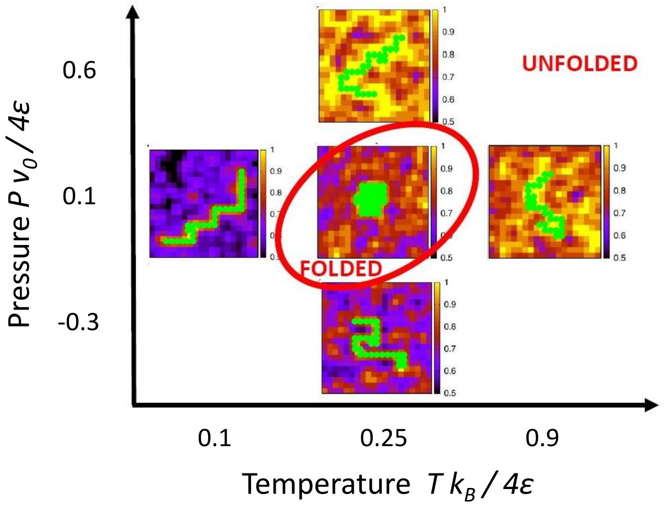


Figure 1.7: Example of the stability region (red ellipse) of a BF homopolymer of hydrophobic residues (green). Color code: density field of water from black (low-ρ) to yellow (high-ρ). Figure adapted from Ref. [5]. Reprinted figure with permission from Ref. [5]. Copyright (2015) by the American Physical Society.

exposed thanks to the higher compressibility of water at PHO interfaces. At low-T , unfolding is a consequence of the decrease of the hydration water internal energy when more PHO residues are exposed, and at low-P to the decrease of enthalpy of hydration water. Both mechanisms are due to the increase of $N_{\mathrm{HB}}^{\mathrm{PHO}}$. Further analysis showed that these results are robust against changes of model parameters (properties of the solvent) or protein sequence [7].

Next, Bianco and coworkers addressed the protein design problem, i. e. finding the protein sequence that folds into a target backbone structure [6]. Despite the hardness of the problem [224], it is of great interest in Biophysics for its implications in drug design engineering and a to better understand biological evolutionary processes. The BF model allows to study the role of the solvent in the selection of protein sequences thanks to its ability to rapidly sample a large amount of water configurations, and comparing for different protein sequences. Following a previous approach by Coluzza [225, 226], Bianco et al. developed two protocols for protein design based on 1) minimizing the enthalpy calculated for a fixed protein sequence by averaging over equilibrium water configurations, or 2) maximizing the enthalpy gap between folded and unfolded protein configurations, averaged over equilibrium water configurations [6].

Their results show that the stability regions of designed proteins are positively correlated with the design temperature T_d. More precisely, proteins designed at high-T_d are superstable and their sequence is less segregated into hydrophobic core and hydrophilic surface. Surprisingly, they find that an increase of segregation induces less stable proteins upon changes of T and P, suggesting that there is an optimal fraction of hydrophobic residues exposed to the solvent [227, 228]. Bianco et al. interpreted that the exposure of hydrophobic residues arises as a consequence of water properties, rather than a compromise between stability and biological functionality [6].

Finally, Bianco and coworkers studied protein aggregation of homogeneous [9] and heterogeneous [8] protein mixtures. Protein and peptide aggregation is related to the development of neurological diseases such as Alzheimer's or Parkinson's [229–231]. Simulations of the BF model show that proteins first partially unfold and then aggregate upon increasing concentration [9]. In this process, water mediates protein-protein interactions, prior to protein contact. Also, they find and induces a free energy barrier against aggregation [9], a relevant observation for drug engineering. By considering heterogeneous mixtures, Bianco et al. find that proteins tend to aggregate with other proteins of the same species, a result confirmed by light-scattering expriments [8]. More recently, March, Bianco and Franzese studied the effects of salt concentration and hydrophobic confinement on the folded $\leftrightarrow$ unfolded $\leftrightarrow$ aggregation process [73]. They reproduced the increase of salt concentration by reducing J and J^{PHO}, representing the screening effect of ions on HB interactions and hydrophobic confinement as a flat surface with excluded-volume interactions with both water and proteins. Their results show that the threshold concentrations for unfolding and aggregation increase upon decreasing salt concentration and that the presence of a hydrophobic wall decreases them [73].

1.6 Motivation and Objectives of the Thesis

In this chapter, we have presented a Hamiltonian CG approach, proposed by Franzese and Stanley (FS) [71, 72], to model water under confinement and its extension, by Bianco and Franzese (BF), to model hydrated protein systems [5, 6]. The FS model is essentially based on a discretization of the position of water molecules, and on the description of the water-water hydrogen bonding in terms of two terms, accounting for the covalent and cooperative components of the HB interaction, indicated with J and J_σ respectively. The discrete nature of the water degrees of freedom allows the implementation of efficient Monte Carlo algorithms that makes accessible to simulation a wide range of temperatures and pressures, often prohibitive to correctly sample for atomistic water models. Furthermore, the BF model describes proteins as a self avoiding chains of beads, at a resolution of one bead per residue. By assuming that the HBs at the hydration shell have different properties than those in the bulk, the BF model rationalizes the role of water in protein denaturation, design, and aggregation.

This Thesis is devoted to extend the FS model from monolayer (2D) to bulk water systems (3D). The main objective is opening a new path for simulations of large biological systems with unprecedented detail, including explicit water-residue, and interfacial cooperative HBs. To achieve this goal, we must first develop a simulation method that allows to equilibrate systems including a large number of FS water molecules, reproducing the water's experimental density and response functions in quantitative agreement with experiments, at least at conditions of biological interest.

From the thermodynamic point of view, the change of dimension modifies the balance between entropy and enthalpy that determines the free energy of the system. More specifically, there is a large increase in entropy that is not followed by a significant variation of enthalpy. Thus, the second objective of the Thesis is to investigate which of the aforementioned properties of the FS monolayer are intrinsic to the model and which result from the hydrophobic confinement of the monolayer, if any. We are particularly interested in investigating the thermodynamic anomalies at the supercooled regime and further contribute to the debate on the origin of the anomalies of water and the possible thermodynamic scenarios. By the time we started this project, the LLCP had only been located in simulations of the ST2 model [77]. At present, it has also been rigorously located for atomistic TIP4P/2005, TIP4P/Ice [85], and WAIL [86] water models, highlighting the interest of this research field [24]. Although the LLCP occurs at T conditions far below those of biological interest, its identification at the bulk FS model further validates it by showing its transferability to deep supercooled conditions, with results comparable to atomistic models.

After benchmarking and validation, we consider the application of the 3D BF model to hydrate large biological systems. In particular, we develop a simulation approach that maps configurations of proteins simulated with implicit solvent models, e.g. OPEP [232, 233], and equilibrates BF water keeping the proteins fixed. Our objective is to calculate how water thermodynamic properties depend on the protein configuration and the water contribution to the free energy of the system. We find that the inclusion of solvent coordinates is necessary to adequately describe the characteristic configurations of the system. Our approach overcomes an intrinsic limitation of implicit solvent models, since by construction they integrate over the degrees of freedom of the solvent. This work was carried out in collaboration with the group of Dr. Fabio Sterpone, at the Institute de Biologie Physico-Chimique (CNRS, Paris), during a three-month research stay.

The Thesis is structured as follows. In Chapter 2, we extend the FS model to bulk and design parallel distributed algorithms for local and cluster MC simulations. Our implementation of the algorithms enormously reduce the computational cost of the simulations, making accessible large water samples of at least 10^7 water molecules. In Chapter 3, we calibrate the parameters of the bulk FS model to reproduce experimental data around ambient conditions, a prerequisite for its use in simulations of biological interest. The following chapters describe applications of the bulk FS model. In Chapter 4, we study the transferability of the model to the supercooled regime. Here, we localize the LLCP belonging to the Ising 3D universality class at T and P conditions comparable to the estimates in atomistic models. Next, in Chapter 5, we extend the

BF model to bulk. We consider two cases of application: the conformational changes of the von Willebrand factor (vWf) upon extreme shear stress (Chapter 5) and the sequestration of Superoxide Dismutase 1 (SOD1) proteins in Fused in Sarcoma (FUS) and Bovine Serum Albumin (BSA) highly concentrated solutions (Chapter 6). We summarize our conclusions in Chapter 7.

Chapter 2

Methodology: Parallel distributed algorithms

As we have discussed in Chapter 1, water modeling is a challenging task due to the difficulty of reproducing its thermodynamic, dynamic and structural anomalies. They arise from weakly covalent hydrogen bonding (HB) interactions that also exhibit cooperative effects. In this chapter, we extend to bulk the FS model and develop efficient parallel MC algorithms specifically designed for it. The algorithms are executed in GPUs using CUDA. In particular, we develop two parallel algorithms: local Metropolis and cluster Swendsen-Wang (SW). Our implementation of Metropolis considers a layered partition of the lattice that, to our knowledge, is novel in simulations using CUDA. On the other hand, for cluster SW simulations, we adapt the multi-cluster labeling algorithm by Hawick and Kalentev to the FS model. The Metropolis algorithm allow simulating large size systems of $260^3 \sim 2 \cdot 10^7$ water molecules, whereas we benchmark $128^3 \sim 2 \cdot 10^6$ for SW simulations. The model offers a good compromise between accuracy, as it accounts for the HBs and their cooperativity, and access to large-size systems, as required for problems of biological relevance.

2.1 Introduction

Water models play an essential role in simulations of biological systems. First, the properties of the solvent strongly affect the behavior of the whole system [2]. Second, the computational cost to equilibrate water around the proteins and estimate water-protein interactions determines the accessible length and time scales of the representation [234]. All-atom simulations that include explicit solvent such as TIP4P/2005 or AMOEBA water models provide the most accurate descriptions of the systems [235], but they are not affordable for large-scale simulations. On the contrary, coarse-grained (CG) models come at a reduced computational cost but represent the system in a simplified manner. For each particular problem, researchers must reach a compromise between the desired accuracy of the representation and the computational cost of simulating the characteristic length and timescale of the problem [234].

The quest for a water model that both accounts for a detailed description of the HB network, including cooperativity, and being suitable for large-scale simulations is still open. Within this effort, the Franzese-Stanley (FS) model (Sec. 1.2, Refs. [1, 72, 120, 132, 134, 135]) is a promising approach. It is suitable for analytic calculations [115] and Monte Carlo (MC) simulations [131] that reach scales from 1 nm to 1 μm in space and from 10 ns to 0.1 s in time [1]. The model describes in detail the HB network, including cooperativity [40], while coarse-grains the positions of the molecules through a discrete density field. Moreover, as we discussed in Sec. 1.5, it has been applied to biological problems such as protein folding [5, 7], protein design [6], and aggregation [8, 9, 73]. In these studies, the model has helped to unveil the role of HB interactions in the complex behavior of the proteins at different thermodynamic conditions.

In this chapter, we extend to bulk the FS model and benchmark the accessible scales using MC simulations. To this aim, we develop efficient parallel distributed algorithms using CUDA, a C-style programming language, to develop kernels that are executed in the GPU [236]. During the last decade, CUDA has been extensively used in Computational Physics, including lattice spin models simulated with local and cluster MC [237–239], molecular engines [240], stochastic differential equations [241], or Brownian motors [242]. Among these models, the architectures of the GPUs are especially useful for performance gains in spin models on a regular lattice [237]. This feature holds for the FS model, as it accounts for the HB formation and breaking through a regular lattice of Potts variables, as described in Sec. 1.2.

The algorithms developed here are designed considering the specific structural properties of the 3D FS lattice, i.e. how the bonding variables σ_{ij} are spatially distributed and connected. Hence, our algorithms cannot be straightforwardly generalized to other models. However, our work may inspire the search for parallel algorithms for other models. In the context of water, this holds for the model proposed by Cerdeiriña et al. [243] or the model for the hydration of ions proposed in Ref. [198].

The algorithms described in this chapter are significant because they open the possibility of performing large-scale simulations of a CG water model that includes a detailed description of the HB network. They clear the way towards more realistic simulations of large protein systems in explicit solvent, including the effects that arise from individual HBs and their cooperativity.

2.2 Bulk FS model and simulation method

2.2.1 Extension from monolayer to bulk FS model

In the bulk FS model, each molecule has six nearest neighbors (n.n.) with which it can form a HB. However, a water molecule can form up to four HBs. For this reason, the model must introduce a constraint to avoid a molecule forming more than four HBs. A possible solution consists in setting to zero the enthalpy change of the system if a molecule forms or breaks a fifth or sixth HB. In other words, if $\delta_{\sigma_{ij},\sigma_{ji}} = 1$, but the molecule i or j already forms four HBs, this bond does not contribute to the energy nor the volume of the system.

Unfortunately, this solution is not free from problems. First, for a given configuration of σ variables, the calculation of N_{HB} depends on the order in which the molecules are visited. Let us consider a configuration for which all the σ variables are in the same state. Hence, $N_{\text{HB}} = 2N$. However, when estimating N_{HB}, the algorithm may find that, for a given molecule, three or more of its neighbors have already formed four HBs. Thus, this molecule cannot form its corresponding four HBs, and the calculated N_{HB} is $< 2N$. Second, although this solution is easy to implement in the Metropolis algorithm, it cannot be adapted to cluster MC straightforwardly. Conflicts occur when deciding which pairs of adjacent σ variables should be added to the cluster. Moreover, after updating a cluster, many molecules at the boundaries may have formed more than four HBs at once. If we check the border to eliminate the extra bonds, we face the first problem again.

Therefore, we look for a criterion that defines N_{HB} uniquely and that can be uniformly implemented in both local and cluster MC algorithms. We introduce this criterion through the allowing η variables. Thus, we rewrite N_{HB} of the FS monolayer (defined in Eq. 1.5) as

$$N_{\text{HB}} \equiv \sum_{<i,j>} n_i n_j \eta_{ij} \delta_{\sigma_{ij},\sigma_{ji}}, \tag{2.1}$$

where η_{ij} *switches on and off* connections between n.n. so that all molecules in the system can form up to four HBs. We describe how we accomplish this in the following.

Let us regard the FS lattice as a graph $\mathcal{G}(\mathcal{V}, \eta)$, where each cell i is a vertex, $i \in \mathcal{V}$, and η is the set of edges ($\eta_{ij} \in \eta$). Here, η_{ij} links the pair of n.n. molecules (cells i and j) in a 3D simple cubic lattice. Next, we split η into two sets $\eta_0 \equiv \{\eta_{ij} = 0\}$ and $\eta_1 \equiv \{\eta_{ij} = 1\}$, with $\eta = \eta_0 \cup \eta_1$ and $\eta_0 \cap \eta_1 = \emptyset$. We use these sets to define the graphs $\mathcal{G}_0(\mathcal{V}, \eta_0)$ and $\mathcal{G}_1(\mathcal{V}, \eta_1)$. Our goal is to generate sets η_1 so that all the vertices (molecules) i are connected to exactly four n.n.. We achieve this by letting $\mathcal{G}_0$ be a Hamiltonian cycle [244]. A Hamiltonian cycle is a path that visits all the vertices once. In other words, if we start from any vertex i, we can follow the edges in η_0 so that we visit all the vertices once and, eventually, return to the initial vertex. We note that if $\mathcal{G}_0$ is a Hamiltonian cycle, then all the cells have exactly two allowing variables $\eta_{ij} = 0$

and four $\eta_{ij} = 1$, which is the desired result.

There are algorithms that find all edge sets of a graph which are vertex-disjoint unions of cycles [245]. In particular, the algorithm outputs all Hamiltonian cycles (if any). However, we are not interested in listing all the possible pairs (η_0, η_1) of the system, but in generating different sets during the simulation. Then, the molecules will form up to four HBs with alternating neighbors. For this reason, we develop a method specifically designed for our model.

We partition the system into a set of 3D checkerboard cubes (c.c.), in such a way that (i) the eight vertices of a cube correspond to eight molecules $i \in \mathcal{V}$, and (ii) each molecule belongs to two, and only two, cubes (Fig. 2.1 left). A Hamiltonian path $\mathcal{G}_0$ is built by setting four edges of each cube to $\eta_{ij} = 0$, in such a way that each vertex i has one edge with $\eta_{ij} = 0$ and two with $\eta_{ij} = 1$. If this is satisfied in all cubes, and each vertex (water molecule) belongs to two cubes, then we find that $\mathcal{G}_0$ is a Hamiltonian cycle. We find nine possible ways of setting the η_{ij} edges of the cubes that fulfill the desired conditions, as shown in Fig. 2.1 (right). The restrictions for the use of this method are (i) the vertices must be in a cubic simple structure with periodic boundary conditions PBCs, and (ii) the number of vertices at each lateral edge L_X, L_Y, and L_Z of the lattice must be multiple of 4. Otherwise, the checkerboard partition cannot be formed.

Regarding the cooperative interaction, we keep the same expression as for the FS monolayer (Eq. 1.6). The only difference arises from the fact that, in 3D, there are six σ_{ij} variables per molecule instead of four. Thus, a water molecule can form up to 15 cooperative bonds in the bulk, instead of 6 as in the monolayer.

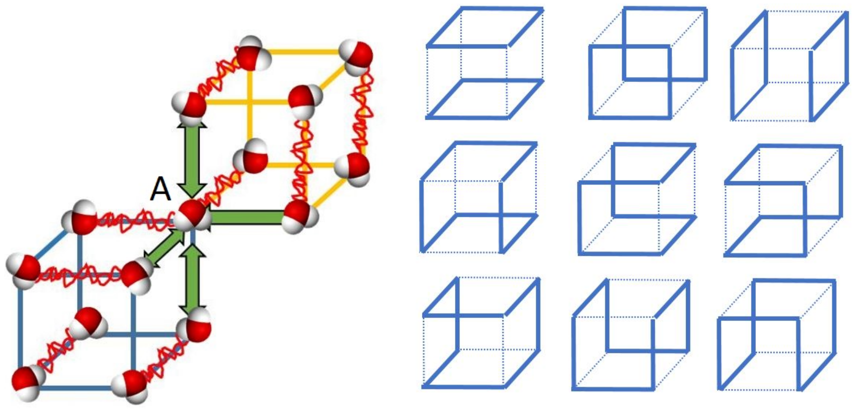


Figure 2.1: Left panel: Example of the Checkerboard partition. Water molecules mark the center of the cells in the system. The molecule labeled “A” belongs to the blue and yellow cubes in the checkerboard partition. Red broken lines represent $\eta_{ij} = 0$ edges, while all other edges correspond to $\eta_{ij} = 1$. Green double arrows highlight the allowed bonds for the molecule “A”. Right panel: The nine possible sets of η_{ij} for a checkerboard cube, where thick lines represent $\eta_{ij} = 1$, and dotted lines $\eta_{ij} = 0$. The yellow cube in the example of the left panel corresponds to the central set, while the blue one, to the bottom-central set.

2.2.2 Simulation method: Definition of the MC step

We perform MC simulations in the NPT ensemble. The configuration of the system is fully determined by the radius of the cells r, the σ variables, and the η variables. At each simulation step, we perform the following tasks to generate new configurations:

1. We update r keeping σ and η fixed, according to the Metropolis algorithm [246].
2. We update η keeping r and σ fixed, according to the Metropolis algorithm.
3. We update σ keeping r and η fixed. We perform this update according to either parallel Metropolis, parallel Swendsen-Wang (SW) [247] algorithms.

In this section, we briefly describe the application of the Metropolis algorithm to update r and the η variables. In Sections 2.3 and 2.4, we explain in detail the parallelization of Metropolis and SW algorithms to update σ variables.

The Metropolis algorithm in the NPT ensemble accepts a new configuration according to the probability

$$p^{\mathrm{MET}} = \min\left\{1, \exp\left(-\beta \Delta H\right)\right\}, \tag{2.2}$$

where $\beta \equiv 1/k_B T$ and ΔH is the enthalpy difference between the new and the old configurations.

To update r, we perform a single proposal of change $r' = r + \delta r$, where r' and r are the new and old radius, respectively, and δr is chosen at random within the range $-0.01 \leq \delta r / r_0 \leq 0.01$. The move is accepted with the probability given in Eq. 2.2, taking the enthalpy change as

$$\Delta H(r \to r') = \left[\sum_{i,j} U({r'}_{ij}) - U(r_{ij})\right] + PN({r'}^3 - r^3), \tag{2.3}$$

where $U(r)$ is the Lennard-Jones (LJ) potential

$$U(r) \equiv \begin{cases} 4\epsilon\left[\left(\frac{r_0}{r}\right)^{12} - \left(\frac{r_0}{r}\right)^{6}\right] - U_c, & \text{if } r < r_c \\ 0 & \text{if } r \geq r_c, \end{cases} \tag{2.4}$$

where $r_c = 6r_0$ is a cutoff distance introduced to save computational cost. We shift the potential by $U_c \equiv 4\epsilon\left[\left(r_0/r_c\right)^{12} - \left(r_0/r_c\right)^{6}\right]$ to avoid a discontinuity at r_c. The LJ potential considered here differs to the considered for FS monolayers (Eq. 1.4) but the results are qualitatively consistent.

To generate new configurations of η variables, we apply the algorithm described in Section 2.2.1. In particular, we propose a new state for each c.c. among the nine possible states in Fig. 2.1. The enthalpy change due to a single c.c. update is

$$\Delta H\left(\eta \rightarrow \eta'\right) = -J^{\mathrm{eff}} \Delta N_{\mathrm{HB}}, \tag{2.5}$$

where J^{eff} is the enthalpy change due to HB formation, as defined in Eq. 1.8, and

$$\Delta N_{\mathrm{HB}} = \sum_{\langle i,j\rangle_{\mathrm{c.c.}}} \left({\eta'}_{ij} - \eta_{ij}\right) \delta_{\sigma_{ij},\sigma_{ji}}. \tag{2.6}$$

The sum is performed over the $\langle i,j\rangle_{\mathrm{c.c.}}$ pairs of n.n. cells that belong to the c.c.. We note that the update of a c.c. is independent of the others. For this reason, we parallelize this step by assigning the update of a single c.c. to a single thread.

2.2.3 Parallel code setup

In CUDA applications, tasks are distributed between the GPU (device) and the CPU (host), depending on whether they are parallelized or not. Arrays that are read/written by both of them must be allocated in both of them. To distinguish them, the usual criterion is to add the `dev_` prefix if it is allocated in the device. In order to improve the performance of the application, communication between host and device must be done only if necessary.

We show the workflow scheme of our application in Fig. 2.2. We start generating the initial configuration of the system $\{\sigma, \eta\}$ in the host, initializing a vector of seeds that will be used by the CUDA random number generator. Next, we allocate the vectors `dev_`σ, `dev_`η, and `dev_seed` in the device and copy the values stored in the host. Then, we generate arrays of random numbers vectors in the device using the `cuRand` library. These arrays contain the random numbers which will be read by the CUDA kernels, following the principle of deferred decisions. In Appendix A we report more details on the random number generation and their usage. The main loop performs `N_RUN` MC steps (update r, η, and σ) as described in Section 2.2.2.

2.3 Parallel Metropolis

The Metropolis algorithm can be easily parallelized for a regular spin-lattice. The procedure consists in dividing the lattice into domains of σ variables that are updated simultaneously. To satisfy detailed balance, if σ_{ij} and σ_{kl} belong to the same domain, then

$$\begin{aligned}\Delta H(\sigma_{ij} \rightarrow \sigma'_{ij}) &= -J^{\mathrm{eff}} \Delta N_{\mathrm{HB}} - J_\sigma \Delta N_\sigma \\ &= -J^{\mathrm{eff}} \eta_{ij} (\delta_{\sigma'_{ij},\sigma_{ji}} - \delta_{\sigma_{ij},\sigma_{ji}}) - J_\sigma \sum_{k\neq j} (\delta_{\sigma'_{ij},\sigma_{ik}} - \delta_{\sigma_{ij},\sigma_{ik}})\end{aligned} \tag{2.7}$$

must not depend on σ_{kl} and vice versa. For the Ising model, there are two widely used partition schemes: layered [248] and checkerboard [249]. Another method for

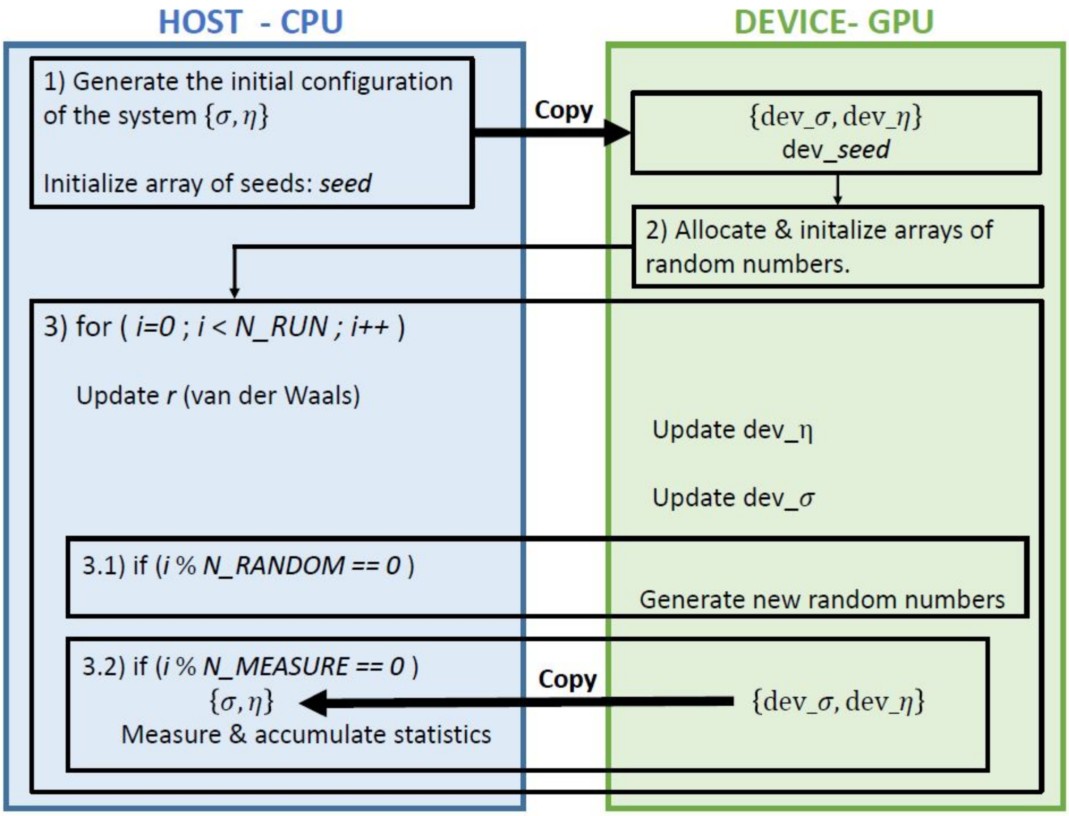

HOST - CPU
DEVICE- GPU
1) Generate the initial configuration of the system {σ, η}
Initialize array of seeds: seed
Copy
{dev_σ, dev_η}
dev_seed
2) Allocate & initalize arrays of random numbers.
3) for (i=0 ; i < N_RUN ; i++)
Update r (van der Waals)
Update dev_η
Update dev_σ
3.1) if (i % N_RANDOM == 0)
Generate new random numbers
3.2) if (i % N_MEASURE == 0)
{σ, η}
Measure & accumulate statistics
Copy
{dev_σ, dev_η}

Figure 2.2: Schematic workflow of the CPU-GPU program.

improving the efficiency of the code is storing multiple spins (σ variables in our model) in a single memory word [237, 250]. For example, eight spins can be stored in an `uint8_t` word, where each bit is 0 or 1 depending on whether the spins are up or down. This technique is useful in GPU computing due to the reduced cost of memory transactions. Unfortunately, it holds for two-states spins, while the FS model considers six-states spins.

CUDA implementations of the checkerboard parallelization scheme for the 2D and 3D Ising model can be found in [237, 238, 251]. These algorithms cannot be adapted straightforwardly to the FS model due to the specific decoration of the σ variables at the lattice. Alternatively, we consider a layered partition at σ variables level in combination with a smart sorting of the arrays that allow memory coalescing, as described in the following. To our knowledge, the parallelization of the Metropolis algorithm for the Ising model following the layered scheme has not been implemented in CUDA yet.

We partition the σ variables into six domains, each containing all the σ variables with the same sense and direction (forward X, backward X, etc.). We can update at the same time all the spins of a block since all the σ variables involved in the ΔH estimate belong to other blocks, as shown in Fig. 2.3.

We implement a CUDA kernel `gpu_metropolis(sp)` that launches one thread per water molecule, where `sp` indicates which of the six independent domains will be updated. We define a parallel Metropolis update as six sequential calls to `gpu_mertopolis(sp)`, where `sp` is chosen randomly to mimic the random selection of σ variables in the sequential Metropolis, and to avoid the propagation of correlation waves.

In CUDA applications, the main bottleneck in the execution comes from data-accessing latency [252]. The performance can be improved by smartly sorting the memory, taking advantage of memory coalescing [236,253,254]. The GPU creates, manages, schedules, and executes simultaneously blocks of 32 threads called *warps* [237]. When a kernel reads (writes) global memory directions, it performs a single coalesced read (write) transaction every half-warp of 16 threads. In view of that, we are interested in sorting the vectors in such a way that consecutive threads read (write) consecutive memory addresses. We achieve this by sorting the vectors `dev_`σ and `dev_`η according to the index `idx` $=$ `arm`$\cdot N+$`cell`, where `arm` $\in \{0, 1, ..., 5\}$, and `cell` $\in \{0, 1, ..., N-1\}$. The index `arm` stands for the six possible neighbors of the cell (from 0 to 5: left, right,

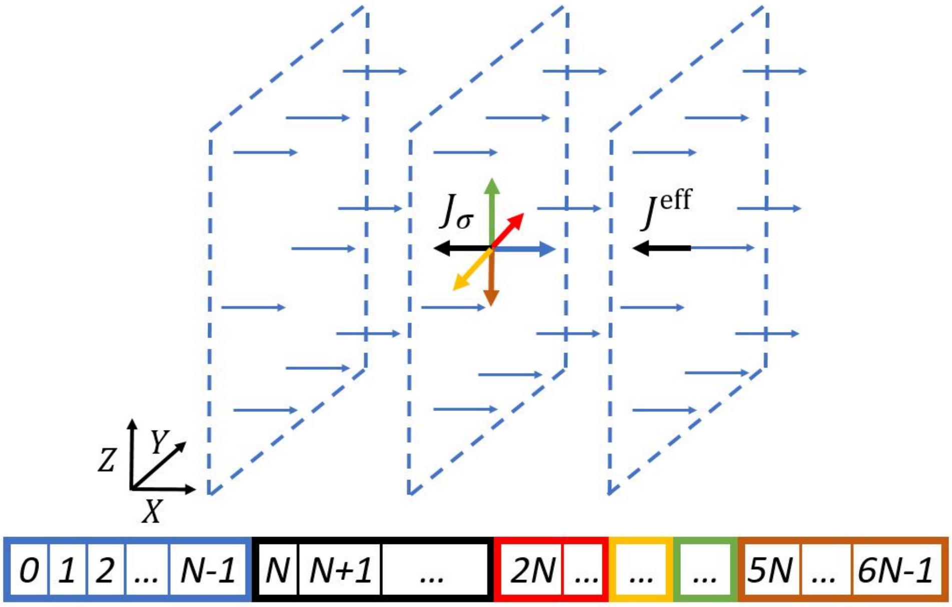


Figure 2.3: Metropolis parallelization in layered domains. Schematic depiction of the system, where the arrows represent the σ variables' lattice. Dashed lines represent cells in planes with constant X. Arrows with the same color belong to the same domain. For clarity, the six σ variables are represented only for the central cell. We update in parallel all the σ variables of the same domain, while the others remain fixed. Here we consider the case in which all the σ_{ij} in blue (forward X direction) are updated. The enthalpy variation $\Delta H(\sigma_{ij} \to \sigma'_{ji})$ depends on η_{ij}, σ_{ji} (black, J^{eff}), and the five $\sigma_{i,k}$, $k \neq i$ (black, red, yellow, green, and brown, J_σ). The bottom bar shows how the vectors `dev_`σ and `dev_`η are nested, according to the index formula described in the text. This ordering is relevant, as the algorithm performs coalesced reads, improving its performance.

front, back, top, bottom), and `cell` $= x + yL_X + zL_XL_Y$, where x, y, and z are the Cartesian coordinates of the cell (see Fig. 2.3).

We illustrate how the kernel `gpu_metropolis(sp)` performs coalesced memory transactions with the following example. We consider a half-warp that updates the left arm (forward X direction, `sp`$= 0$) of the water cells `idx`$=0$ to 15. When the kernel estimates ΔN_{HB}, it reads the right arms of the neighboring cells 1 to 16, i.e., the (consecutive) positions `sp` $= 65$ to 80. The same happens when estimating ΔN_σ. These reads go from `arm` $\cdot N + 0$ to `arm` $\cdot N + 15$, where `arm` $\neq 0$. We note that an exception to this rule comes when the neighboring cell is placed on the opposite side of the simulation box, due to the PBCs.

We measure the performance of our parallel Metropolis algorithm and compare it to the results obtained in a sequential implementation in the CPU. The results were obtained in a workstation with an AMD Ryzen 7 5800H processor, with 4.4 GHz, and one NVIDIA GeForce RTX 3060 with 6 GB of global memory and 3840 CUDA cores. The CUDA Toolkit version used is 11.5.1 (released on November 2021).

For the sake of a fair comparison between the sequential and parallel algorithms, the measurement of the time cost of the GPU algorithm includes the GPU time spent on the generation of the random numbers (Appendix A) and the cost of GPU → CPU memory transactions. The time cost of the CPU algorithm includes only the CPU time spent on updating the system (including the generation of random numbers when needed), as there is no need for memory transactions. We define the speedup factor (SF) as the ratio $t^{\mathrm{CPU}}/t^{\mathrm{GPU}}$. Thus, the GPU-parallel algorithm is more efficient than the CPU-sequential if SF > 1.

Table 2.1 reports the estimated SF for the Metropolis algorithm. We fix the model's parameters to $J/4\epsilon = 0.5$, $J_\sigma/4\epsilon = 0.03$, and $v_{\mathrm{HB}}/v_0 = 0.5$. The thermodynamic conditions are $Tk_B/4\epsilon = 0.5$, $Pv_0/4\epsilon = 0.0$. Then, we estimate SF for a wide range of sizes of the system $64 < 6N < 2,097,152$. Depending on the size, we perform between 2 and 10 independent simulations of 5000 MC steps. We find that the results shown in Table 2.1 are robust against a change of thermodynamic conditions in the range $0.05 \leq Tk_B/4\epsilon \leq 0.5$ and $0.0 \leq Pv_0/4\epsilon \leq 0.6$. This was expected, since the change of thermodynamic conditions only modify the values of β and J^{eff}, leaving unaffected the computational cost of the algorithm.

The results show that, for small-size systems ($N = 64$ molecules), the parallel algorithm is less efficient than the sequential one. This is not surprising, as a sufficiently large number of threads must be executed to take full advantage of the GPU resources [240]. In Ref. [251], Wojtkiewicz and Kalinowski also find SF < 1 for small-size systems. In Appendix B, we study the scaling of the sequential and parallel Metropolis algorithms. We benchmark accessible parallel Metropolis simulations of $N = 260^3 \sim 1.8 \cdot 10^7$ molecules (Fig. B.2). Moreover, we show that SF exhibits a crossover between two power laws at $N = 32,768$ (Fig. B.3). Within the range of N

Metropolis Speedup Factor (SF)		
Lattice size L	Number of molecules $N \equiv L^3$	SF= $t^{\mathrm{CPU}}/t^{\mathrm{GPU}}$
4	64	0.1159(6)
16	4,096	7.09(3)
20	8,000	13.42(9)
32	32,768	37.8(3)
52	140,608	63.8(3)
64	262,144	63.5(3)
128	2,097,152	136.72(3)

Table 2.1: Speedup factor (SF) of the GPU Metropolis algorithm respect to the sequential implementation in the CPU. Results were obtained using model parameters $J/4\epsilon = 0.5$, $J_\sigma/4\epsilon = 0.03$, and $v_{\mathrm{HB}}/v_0 = 0.5$. Shown data correspond to $Tk_B/4\epsilon = 0.5$ and $Pv_0/4\epsilon = 0.0$, but the SF does not change at other thermodynamic conditions, as described in the text.

considered here, SF does not reach a plateau.

2.4 Parallel Swendsen-Wang

Local MC algorithms such as Metropolis suffer from a critical slowdown of the dynamics when the correlation length of the spins is comparable with the size of the system. On the contrary, cluster MC algorithms smartly update correlated regions of bonded spins (clusters) at once. Thus, they generate statistically independent configurations at a much lower computational cost. This is crucial at the supercooled region, where the FS monolayer exhibits a liquid-liquid phase transition ending in a liquid-liquid critical point [123, 134]. Here, we consider the SW multi-cluster algorithm [247]. At each step, we generate clusters of bonded σ variables that cover the whole system (unbounded σ_{ij} variables will form clusters of size 1). The new configuration is obtained by updating all the σ variables in the same cluster accordingly. The sequential implementation of the SW algorithm for the bulk FS model reads

1. Visit all the cells i. For each i, loop over all the pairs of variables $(\sigma_{ij}, \sigma_{ik})$. If they are in the same state, place a bond between them with probability $p_\sigma = 1 - \exp(-\beta J_\sigma)$.

2. Visit all the pairs of n.n. cells $\langle i, j\rangle$. If $J^{\mathrm{eff}} > 0$, $\eta_{ij} = 1$, and $\delta_{\sigma_{ij},\sigma_{j,i}} = 1$, place a bond with probability $p_{\mathrm{eff}} = 1 - \exp\left(-\beta\left|J^{\mathrm{eff}}\right|\right)$. Instead, if $J^{\mathrm{eff}} < 0$, place a bond with probability p_{eff} if $\delta_{\sigma_{ij},\sigma_{j,i}} = 0$.

3. Use the Hoshen-Kopelman algorithm [255] to identify the clusters.

4. Visit all the clusters. For each, choose a random integer `RND_int` $\in \{0, ..., q-1\}$. Change the state of all the σ variables in the cluster to $\sigma_{ij} \leftarrow \big(\sigma_{ij} + \texttt{rnd\_int}\big) \% q$, where $\leftarrow$ is the assignation operator.

The SW algorithm performs three independent tasks. First, it places bonds between connected σ variables if they belong to the same cluster. Second, it identifies all the clusters. Third, it updates each cluster. The first and third tasks are highly local and can be parallelized straightforwardly. However, this does not hold for the cluster labeling operation. A possible way to parallelize it consists in dividing the lattice into domains that are assigned to different threads. Then, each thread computes the labels of the lattice site in its domain by applying a sequential algorithm. Obviously, this strategy fails whenever a cluster crosses the domain barriers. To solve this problem, another process must check the borders sequentially. This overall strategy, with some modifications, is followed in Refs. [238, 248, 256].

In our approach, we follow the efforts of Hawick et al., who implemented different parallel labeling algorithms for arbitrary and lattice graphs in CUDA [257]. Among these, the label equivalence algorithm was sophisticated by Kalentev et al. [258] and later applied by Komura and Okabe to SW simulations of the 2D Potts model [239]. Here, we adapt the Hawick-Kalentev label-equivalence algorithm to the 3D FS model.

For a given SW step, we first generate the clusters. We apply a direct parallelization of this task, in which each thread works on one water cell. The thread is responsible for placing the J_σ interactions within its cell and the J interactions in the left, front and top directions.

Once the bonds are placed, we apply the label equivalence algorithm. We allocate in the device the `label`[1] vector of size $6N$ that indicates to which cluster each σ_{ij} belongs. Thanks to the Kalentev sophistication, this array is also used to resolve the label equivalences [258]. The advantage is the reduction of the memory cost of the algorithm, which is relevant due to the limited storage resources of the GPUs. We initialize `label` as, for each k, $\texttt{label}[k] = k$, where k is the index `idx` of a σ variable defined in Section 2.3. The algorithm resolves the label equivalences through iterative callings to the *scanning* and *analysis* functions [239,258]. When the algorithm converges, all the σ variables in the same cluster will take the same `label` value.

The scanning function compares the label of a lattice site k to the label of all the bonded sites (nearest neighbors within the cluster). For each k, $\texttt{label}[k]$ is updated to the minimum value among all the labels of the bonded sites, including itself. In Ref. [239], Komura and Okabe implemented this function with a single kernel for the 2D Potts model. However, due to the particular decoration of the the σ variables in the 3D FS model, we find it more convenient to split this function into two kernels. First, in `gpu_scanning_covalent` each thread scans through left, front, and up J

---

[1] Since there is no copy of the array `label` in the host, we omit the prefix `dev_`

interactions of the cell Second, `gpu_scanning_cooperative` scans the J_σ interactions. An alternative implementation in a single kernel leads to race conditions when two threads attempt to update the same element of `label`. To avoid them, we could use the CUDA `atomic_min` function, but we found a slower performance due to the greater thread divergence [258].

Next, the analysis function updates `label`$[k]$ according to

```
if ( label[label[k]] != label[k] )
   label[k] = label[label[k]].
```

This step further propagates the minimum value of `label` to other σ variables in the same cluster. The parallel implementation of the analysis function suffers from race conditions, but these collisions between threads will be resolved eventually in successive applications of the scanning and analysis functions [258]. In any case, to minimize the occurrence of thread conflicts, we define the `gpu_analysis`(`arm`) kernel to only update `label` of those σ variables in the direction of `arm`. We make a loop of six calls to the kernel, to include all the lattice sites.

To check whether the algorithm has converged or not, we copy the `label` vector to `prev_label` before calling the scanning function. Then, after the calling to `analysis`, we compare the present value of `label` to `prev_label`. To parallelize this task, each thread compares the six σ variables in the cell. The algorithm converges whenever `label`[:] = `prev_label`[:]. As an example, Table 2.2 shows how `label` converges after successive applications of the scanning and analysis functions.

We estimate the speedup factor of the parallel SW algorithm relative to the sequential one (see Table 2.3). We use the same workstation and model parameters as in Section 2.3. In this case, we show results for four different thermodynamic conditions: $Tk_B/4\epsilon = 0.05$ and 0.06, both temperatures at $Pv_0/4\epsilon = 0.0$ and 0.6. Contrary to the Metropolis algorithm, the time cost of the SW algorithm depends on the size of the clusters, which in turn depends on the thermodynamic conditions. For simulations at $T = 0.05(4\epsilon)/k_B$, there is a percolating cluster, while at $T = 0.06(4\epsilon)/k_B$, there is not [123]. The effect in P is that at low-P the average cluster size is larger than at high-P. For each thermodynamic condition and algorithm, we perform between 2 and 10 independent simulations of 10^3 MC steps.

In Appendix B, we study the scaling of the SW algorithms. As in the Metropolis case, we find that SF displays a crossover between two power laws at the same size $N = 32,768$ and that SF does not saturate for the large N considered here.

2.5 Dynamic behavior of the algorithms

In this section, we calculate the autocorrelation time τ of the σ variables, in both parallel Metropolis and SW algorithms. This time is an estimate of the number of

cell index ↔ (x, y, z)	0 ↔(0,0,0)		1↔(1,0,0)		17↔(1,0,1)		18↔(2,0,1)
σ_{ij} index	64	0	65	257	337	17	82
initial label	64	0	65	257	337	17	82
scan covalent	64	0	0	257	257	17	17
scan cooperative	0	0	0	0	17	17	17
analysis	0	0	0	0	17	17	17
converged?	No						
scan covalent	0	0	0	0	0	17	17
scan cooperative	0	0	0	0	0	0	17
analysis	0	0	0	0	0	0	0
converged?	No						
scan + analysis	0	0	0	0	0	0	0
converged?	Yes						

Table 2.2: Example of parallel label equivalence algorithm. We consider a small cluster of seven σ variables with indices in "σ_{ij} index" row. The lattice has $N = 4^3 = 64$ cells. Each pair of σ variables in the same cell (cell index ↔ Cartesian coordinates row) are bonded through a J_σ interaction. The pairs of spins $(0, 65)$, $(257, 337)$, and $(17, 82)$ are bonded through a J interaction. The initial value of label coincides with the spin index. The following lines show the resulting label after the application of the kernels scan covalent, scan cooperative, and analysis. At the third iteration, `label` does not change, so this cluster has converged. The SW step ends when all the clusters converge.

Swendsen-Wang Speedup Factor SF= $t^{\mathrm{CPU}}/t^{\mathrm{GPU}}$					
L	$N \equiv L^3$	Temperature and pressure ($Tk_B/4\epsilon$, $Pv_0/4\epsilon$)			
		(0.05,0.0)	(0.05,0.6)	(0.06,0.0)	(0.06,0.6)
4	64	0.0617(12)	0.060(2)	0.052(12)	0.0549(9)
16	4,096	3.17(4)	2.6(8)	2.04(3)	2.14(3)
20	8,000	7.38(10)	6(2)	4.57(3)	4.77(4)
32	32,768	30.3(5)	22(8)	16.57(13)	17.2(2)
52	140,608	41.5(3)	25(6)	20.7(2)	21.6(2)
64	262,144	47.8(7)	49.1(7)	21.89(11)	22.79(11)
128	2,097,152	65.0(2)	72(2)	24.99(6)	26.21(9)

Table 2.3: Speedup factor (SF) of the GPU Swendsen-Wang algorithm respect to the sequential implementation in the CPU. The thermodynamic conditions are specified in the second row.

MC steps that are necessary to obtain statistically uncorrelated configurations. As we shall see, τ depends on the sampling algorithm and the thermodynamic conditions.

We estimate the time correlation function (Ref. [121], Sec. 1.3)

$$C_M(t) \equiv \frac{\langle M_i(t_0+t)M(t_0)\rangle - \langle M\rangle^2}{\langle M^2\rangle - \langle M\rangle^2}, \tag{2.8}$$

where $M(t) \equiv \frac{1}{6N}\max\{N_{q'}\}$, and $N_{q'}$ is the number of σ variables in the state $q' \in \{0, ..., q-1\}$. We define τ as $C(\tau) = 1/e$.

We plot $C_M(t)$ in Fig. 2.4. We consider the same parameters of the model as in Tables 2.1 and 2.3. The thermodynamic conditions are $T = 0.05(4\epsilon/k_B)$ for a wide range of pressures $-0.6 \leq P(v_0/4\epsilon) \leq 0.8$, corresponding to deep supercooled conditions. For the Metropolis algorithm, we find dynamic arrest of the system at low $Pv_0/(4\epsilon) = -0.6$, with $\tau \to \infty$, corresponding to a glassy state. At intermediate $P(v_0/4\epsilon) = 0.1$ and $P(v_0/4\epsilon) = 0.6$, close to the Widom line, we find $\tau \sim 300$ MC Steps. For larger $P(v_0/4\epsilon) = 0.7$ and $P(v_0/4\epsilon) = 0.8$, the system is in the HDL region, where the σ variables decorrelate fast ($\tau < 100$). Comparison with SW shows that cluster MC avoids the critical slowdown of the dynamics, even at the glassy state and close to the Widom line ($\tau \sim 20$).

Finally, we measure the time cost of generating statistically independent configurations in our workstation (Table 2.4). We find that, for $Tk_B/4\epsilon = 0.05$ (supercooled conditions), SW is faster than Metropolis at all P. We note that the time gain is notable at low and intermediate-P. However, at high-P, the time difference almost

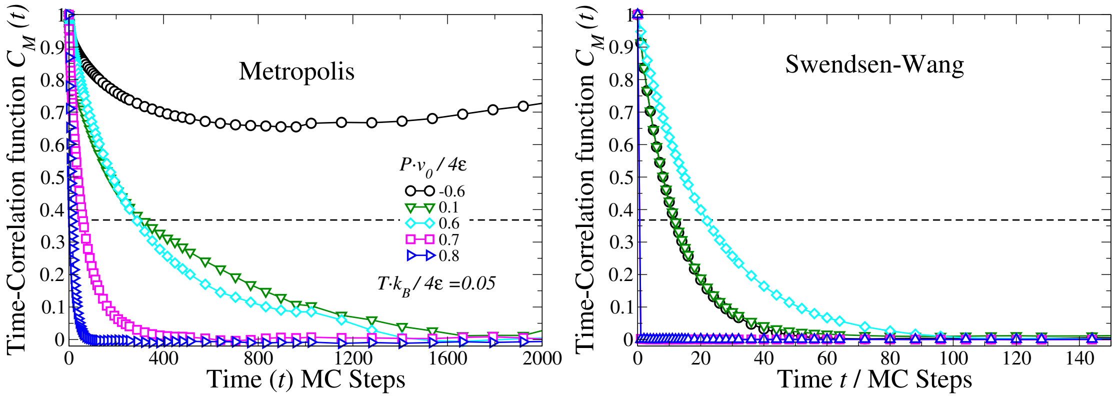


Figure 2.4: Time correlation function $C_M(t)$ at $T \cdot k_B/4\epsilon = 0.05$ for selected pressures (see legend). Dashed line is set to $C_M = 1/e$, as a guide to eye to estimate τ, $C_M(t = \tau) \equiv 1/e$. Left: Parallel Metropolis algorithm. We distinguish three regimes for the HB dynamics. At $Pv_0/4\epsilon \geq 0.7$, fast C_M decay corresponds to liquid water. At intermediate pressures $0.6 \geq Pv_0/4\epsilon \geq 0.1$, the larger correlation time indicates we are close to the LDL-HDL transition. At extreme low-$Pv_0/4\epsilon = -0.6$, we find dynamical arrest of the system corresponding to a glassy state ($\tau \to \infty$). Right: Swendsen-Wang algorithm. Critical slowdown of the dynamics is avoided. The timescale of the x-axis has been zoomed to distinguish the curves better. The size of the system is $N = 32,768$.

$N = 32,768$		Parallel Metropolis		Parallel Swendsen-Wang	
$Tk_B/4\epsilon$	$Pv_0/4\epsilon$	τ/ MC Steps	Time cost / ms	τ/ MC Steps	Time cost / ms
0.05	-0.6	>2000	>772.0	12	29.64
	0.1	320	119.04	12	30.12
	0.6	288	107.14	22	56.1
	0.7	60	22.32	1	3.04
	0.8	12	4.46	1	2.31
0.1	0.7	2	0.732	1	1.50

Table 2.4: Correlation time τ for parallel Metropolis and Swendsen-Wang algorithms in MC Steps ($Tk_B/4\epsilon = 0.05$ in Fig. 2.4) and the computational cost of simulating τ MC steps in ms. The time cost of a single Metropolis update is $\sim$ 0.37-0.39 ms, while for Swendsen-Wang is $\sim$ 2.3-3.0 ms for $Tk_B/4\epsilon = 0.05$ and 1.5 ms for $Tk_B/4\epsilon = 0.1$.

vanishes. This corresponds to a system forming small clusters. At higher T we find that the time cost of computing these small clusters is not compensated by the faster dynamics. For example, at $Tk_B/4\epsilon = 0.1$ and $Pv_0/4\epsilon = 0.7$ the Metropolis algorithm offers a better performance compared with SW.

2.6 Conclusions

In this chapter, we extended to bulk the FS water model, implement a parallel Metropolis algorithm based on a layered partition, and adapt the label equivalence algorithm from Hawick [257] and Kalentev [258] for simulations with the SW algorithm. The parallel sampling of the FS model allows for reaching unprecedentedly large-size water systems, while keeping a detailed description of the HB network. This work, along with the parametrization around ambient conditions (Chapter 3), opens the path for using of the FS model in problems of biological significance, as demonstrated for monolayers (Sec. 1.5 and Refs. [5–9,73]). Employing the Metropolis algorithm, we have benchmarked size systems of 17,576,000 water molecules, corresponding to a simulation box of 75x75x75 nm^3. This algorithm is suitable for simulations around ambient conditions of hydrated large biological systems. On the other hand, we reach systems of 2,097,152 molecules using SW, corresponding to a simulation box of 37x37x37 nm^3.

Finally, we compare the performance of cluster and local MC at T corresponding to the deep supercooled region. We estimate the computational time cost of generating independent configurations, finding that SW is faster at the supercooled conditions considered here. However, for increasing T Metropolis overcomes SW, since computing small clusters is more expensive than local updates of the σ variables.

Chapter 3

Parametrization and Phase Diagram

In Chapter 2, we have developed parallel distributed Monte Carlo (MC) algorithms that allow to simulate systems of the order of 10^7 Franzese-Satanley (FS) water molecules. In this chapter, we adjust the parameters of the bulk FS model: J, J_σ, and v_{HB} (defined in Sec. 1.2) to quantitatively reproduce water experimental results (density and response functions) around ambient conditions, a prerequisite for its use as solvent in biological simulations. To this aim, we consider ab initio calculations to estimate the ratio between J_σ and J and experimental results for bulk water to estimate the definitive set parameters. The region of quantitative agreement extends from atmospheric to 50 MPa, covering almost all pressure conditions allowing life on Earth, in a temperature range that can be as large as $270 \leq T/\mathrm{K} \leq 330$ at ambient pressure.

3.1 Introduction

As we have discussed in Chapter 1, water is of the utmost importance for biophysical and biochemical systems due to its role in the function of proteins [259, 260], membranes [68, 235, 261] or nanoparticles [262, 263]. Computational research on these systems has proven its power to unveil the relevant mechanisms underlying their behavior [264]. In this context, water modeling is essential: greater accuracy of the solvent properties implies more reliable systems' representations [235]. However, accuracy typically comes at a larger computational cost. In explicit solvent simulations, the water model can become the limiting factor for accessible length and time scales [234].

In Chapter 2, we have extended to bulk the Franzese-Stanley (FS) water model that represents the hydrogen bond (HB) network at the resolution of the water molecule [72, 134]. Thanks to efficient parallel distributed algorithms, we benchmarked accessible size systems of $\sim 1.76 \cdot 10^7$ water molecules. To our knowledge, this limit is far beyond the already obtained in any other explicit solvent model with hydrogen-bonded molecular resolution and cooperative effects. Hence, we have accomplished the first requisite for the use of the FS model to hydrate large biological systems: reaching large-size systems.

In this chapter, we address the second requisite for the use of the model in biological simulations: its ability to reproduce the experimental equation of state of water, at least around ambient conditions. To this aim, we parametrize the FS model according to quantum ab initio calculations and experimental data (water density and response functions) under life-relevant conditions. We validate the model and describe the temperature T and pressure P thresholds of quantitative agreement. Beyond these limits, the bulk FS phase diagram is qualitatively consistent with water, showing dynamic and thermodynamic anomalies as the FS monolayer (Sections 1.3 and 1.4).

Concerning the Thermodynamics of the model, we note that the Gibbs free energy of the bulk FS system is not the same as in the FS monolayer. More precisely, in 2D, each water molecule has four σ_{ij} variables and the number of accessible states is $\Omega = q^4 \sim 1.3 \cdot 10^3$ ($q = 6$) whereas, in 3D, each molecule has six nearest neighbors, thus $\Omega = q^6 \sim 4.7 \cdot 10^5$. In terms of entropy, $S \propto \ln(\Omega)$, it corresponds to an increase of $\sim 50\%$. Regarding the enthalpy, it is controlled by N_{HB} and N_σ. Since a water molecule cannot form more than four tetrahedral HBs, the coordination number of the HB network is 4 in both monolayer and bulk cases (Sec. 2.2.1). As a consequence, the covalent contribution to the HB enthalpy remains upon change from monolayer to bulk. On the other hand, the change of dimension increases the total number of cooperative bonds from 6 to 15. However, the impact of this change on the enthalpy can be controlled by tuning J_σ. In summary, the enthalpy of the system is approximately the same regardless the dimension of the sample but the entropy is larger for the bulk case. Hence, the change of Gibbs free energy $G \equiv H - TS$ implies that the results obtained in 2D cannot be straightforwardly generalized to the 3D system. In any case, the results of this chapter will show that this effect has no qualitative impact at the thermodynamic conditions of biological interest.

The significance of the work developed in this chapter, along with Chapter 2, lies in opening a path to developing simulations for biological systems in explicit-solvent, using lattice representations. In particular, in Chapter 5 we consider models that represent proteins at residue or sub-residue resolution, e.g. the OPEP force field [232], and map protein configurations into their corresponding lattice representation. This approach allows including effects that arise from HB interactions and cooperativity, thus improving the reliability of the predictions without a significant increase of the computational effort. As a future perspective, the bulk FS model developed here can be adapted for its use in combination with lattice representations of proteins [265], membranes [199] or bacteria nucleoids [200].

3.2 Parametrization of the model

3.2.1 Estimate of J_σ/J

To estimate the relative contributions from covalent (Eq. 1.5) and cooperative (Eq. 1.6) terms to the total HB energy, we pay attention to energy decomposition analysis (EDA) methods. EDA is employed in Computational Quantum Chemistry to divide total interaction energies obtained from ab initio methods into physically meaningful terms. Here, we follow absolutely localized molecular orbitals (ALMO) EDA [266] applied to density functional theory (DFT) calculations on water clusters [267]. ALMOs are molecular orbitals expressed in the basis set of atomic orbitals of a single molecule.

ALMOEDA separates the total interaction energy of a water cluster into three terms [266]:

$$E_{\mathrm{TOT}} = E_{\mathrm{FRZ}} + E_{\mathrm{POL}} + E_{\mathrm{DEL}}. \tag{3.1}$$

The first is the *frozen* energy term, defined as the energy change due to bringing infinitely separated molecules into the complex geometry of the system. It collects all isotropic contributions to the total energy: molecular orbitals (MOs) relaxation to satisfy the Pauli exclusion principle, permanent electrostatic interactions and van der Waals dispersion, if included in the DFT potential. This term collects all isotropic contributions to the total energy.

E_{POL} is the *polarization* energy, defined as the energy decrease due to ALMO's relaxation in the field of all other molecules in the system. The MOs are constrained to remain absolutely localized; hence no charge transfer is accounted for. Three-body contributions are dominated by E_{POL} [267].

E_{DEL} accounts for the remaining part of the total energy. It corresponds to the *delocalization* energy, defined as the energy difference between the state of polarized ALMOs and the fully optimized state of delocalized MOs. This term can be further decomposed into

$$E_{\mathrm{DEL}} = E_{\mathrm{CT}} + E_{\mathrm{HO}}, \tag{3.2}$$

charge transfer and high-order energy terms, respectively. The latter are small and can be safely neglected. Then, delocalization energy can also be called *charge transfer*.

In the following, we assume that E_{DEL} corresponds to the FS covalent term, while E_{POL} corresponds to the FS cooperative term. Although E_{DEL} includes cooperative effects to some extent [268], we justify our assumption by the strong directional dipole-dipole component of the HB interaction that is originated by the charge transfer in MOs, which is included in the FS covalent term. Moreover, ALMOEDA polarization energy accounts for the orbital relaxation in the field of all other molecules in the system. This energy decrease is dominated by many-body effects, precisely those in the FS cooperative term.

Recently, ALMOEDA has been applied to clusters of ~ 125 water molecules at ambient conditions in bulk liquid [268], air-water interface [269, 270], and water and ions solutions [271]. However, the authors restricted their analysis to the charge transfer energy, while frozen and polarization terms were not explicitly computed. For this reason, we focus on the research performed by Erika A. Cobar et al. in Ref. [267]. They estimate minimum-energy configurations of water clusters of $N = 2 - 5, 13, 17$ molecules, compute E_{FRZ}, E_{DEL}, and E_{POL}, and further decompose them into many-body expansions. Based on their results, we compute the ratio between total polarization and delocalization energies. Letting $E_{\text{POL}}(N) = -J_\sigma N_\sigma(N)$ and $E_{\text{DEL}}(N) = -J N_{\text{HB}}(N)$, we find

$$\frac{J_\sigma}{J} = \frac{E_{\text{POL}}(N)}{E_{\text{DEL}}(N)} \frac{N_{\text{HB}}(N)}{N_\sigma(N)}, \tag{3.3}$$

where $N_{\text{HB}}(N)$ and $N_\sigma(N)$ are the number of directional and cooperative bonds in the minimum-energy configuration of the FS cluster of size N. More precisely, $N_{\text{HB}}(N)$ is the maximum number of HBs that N water molecules in a simple cubic lattice can form and $N_\sigma(N) = 15N$. Table 3.1 reports the estimates of J_σ/J_{HB} for each N. We find that $E_{\text{POL}}/E_{\text{DEL}}$ apparently reaches a plateau for increasing N. Hence, we assume that $(E_{\text{POL}}/E_{\text{DEL}})(N = 17) = 1.25$ is valid in the thermodynamic limit $N \to \infty$. Then, for arbitrary large N:

$$\frac{J_\sigma}{J} = \frac{E_{\text{POL}}}{E_{\text{DEL}}} \frac{N_{\text{HB}}^{\text{Max}}}{N_\sigma^{\text{Max}}} = 1.25 \frac{2N}{15N} = 0.167, \tag{3.4}$$

which is the desired result.

	ALMOEDA			FS MODEL		
N	$E_{\text{POL}}/(\text{kcal} \cdot \text{mol}^{-1})$	$E_{\text{DEL}}/(\text{kcal} \cdot \text{mol}^{-1})$	$E_{\text{POL}}/E_{\text{DEL}}$	N_σ	N_{HB}	J_σ/J
2	-1.6	-1.3	1.231	30	1	0.04
3	-7.5	-4.9	1.530	45	2	0.10
4	-16.6	-10.9	1.523	60	4	0.10
5	-23.5	-14.8	1.589	75	5	0.11
13	-65.8	-54.1	1.216	195	21	0.13
17	-101.5	-81.2	1.250	255	29	0.14
$\gg 17$			1.250	$15N$	$2N$	0.17

Table 3.1: Estimate of J_σ/J based on ALMOEDA results. For each cluster size N, we calculate E_{POL} and E_{DEL} summing up the many-body contributions of Cobar et al.'s Ref. [267] obtained with the ωB97X-D functional. The fourth column reports the ratio $E_{\text{POL}}/E_{\text{DEL}}$. Next, the table contains the number of $N_\sigma(N)$ and $N_{\text{HB}}(N)$ defined in the text. The last column shows the resulting J_σ/J using Eq. 3.3. Finally, assuming $E_{\text{DEL}}/E_{\text{POL}} = 1.25$ is valid for arbitrary large N, we estimate $J_\sigma/J = 0.17$ (last row).

3.2.2 Choice of parameters and fitting to experimental data

In this section, we describe our procedure to find a set of parameters and rescaling functions that translate the equation of state from internal units into SI units. We note that the rescaling functions cancel all the uncertainty encoded in the choice of parameters, so the obtained equation of state is quantitatively consistent with water around ambient conditions. Comparison with experiments (Sec. 3.4) determines the P and T thresholds for the agreement between model predictions and experimental results.

We restrict the fitting to linear rescaling functions:

$$f = a \cdot \hat{f} + b, \tag{3.5}$$

where a and b are constants, f is any thermodynamic observable, and $\hat{f}$ is the dimensionless f. Despite it may be seen as a serious oversimplification, we obtain good results under the linear assumption. Alternatively, machine-learning approaches could be employed to fit more complex functions, but this is beyond the scope of this work.

We perform simulations with different sets of parameters, where J and J_σ are chosen under the constraint in Eq. 3.4 and v_{HB} is changed freely. In particular, we consider $\{J/4\epsilon = 0.5, J_\sigma/4\epsilon = 0.08\}$ with $v_{\mathrm{HB}}/v_0 \in \{0.5, 0.6, 0.7\}$, and $\{J/4\epsilon = 0.2, J_\sigma/4\epsilon = 0.03\}$ with $v_{\mathrm{HB}}/v_0 \in \{0.2, 0.5\}$. For each set of parameters, we find rescaling functions for T and P fitting best the T-P phase diagram of the model with experiments (Appendix C). First, we calculate the density[1] $\rho \equiv N/V$, the isothermal compressibility

$$K_T \equiv -\frac{1}{\langle V\rangle}\left(\frac{\partial\langle V\rangle}{\partial P}\right)_T = \frac{\langle V^2\rangle - \langle V\rangle^2}{k_B T\langle V\rangle}, \tag{3.6}$$

and the specific heat

$$C_P \equiv \left(\frac{\partial\langle H\rangle}{\partial T}\right)_P = \frac{\langle H^2\rangle - \langle H\rangle^2}{k_B T^2}. \tag{3.7}$$

Second, we describe the phase diagram by the maxima of ρ (Temperature of maximum Density, TMD) and the minima of K_T and C_P along isobars. To identify the atmospheric pressure, we calculate the ratio

$$\hat{r}(\hat{P}) \equiv \frac{\hat{T}_{K_T^{\min}}(\hat{P}) - \hat{T}_{\mathrm{TMD}}(\hat{P})}{\hat{T}_{C_P^{\min}}(\hat{P}) - \hat{T}_{\mathrm{TMD}}(\hat{P})}, \tag{3.8}$$

and compare it to the experimental value $r(1\ \text{atm}) = (319.65 - 277.15)\ \text{K}/(309.15 - 277.15)\ \text{K} = 1.33$ [272]. Based on the $\hat{P}$-value corresponding to atmospheric pressure, we propose linear functions that rescale $\hat{P}$ and $\hat{T}$ to fit the phase diagram of water to the experimental results. After comparison among different sets of parameters (Appendix C), we find that the best candidate is

---

[1] For simplicity, we write V instead of $V_{\mathrm{tot}} \equiv Nr^3 + N_{\mathrm{HB}}v_{\mathrm{HB}}$ (Eq. 1.1) hereafter.

$$\begin{aligned} J/4\epsilon &= 0.5 \\ J_\sigma/4\epsilon &= 0.08 \\ v_{\mathrm{HB}}/v_0 &= 0.6. \end{aligned} \tag{3.9}$$

Next, we perform extensive simulations with the set of parameters in Eq. 3.9, following the method described in Sec. 3.3. The equation of state relates the thermodynamic conditions $(\hat{T}, \hat{P})$ with the enthalpy and density $(\hat{h}, \hat{\rho})$ of the system. To obtain the equation of state in SI units, we independently fit T, P, ρ, and C_P to experimental data according to the formulas

$$\begin{aligned} T &= (\hat{T} \cdot 140.57 + 185.47)\mathrm{K} \\ P &= (\hat{P} \cdot 469.46 - 217.89)\ \mathrm{MPa} \\ \rho &= (\hat{\rho} \cdot 1527.3 - 23.102)\ \mathrm{kg/m^3} \\ C_P &= (\hat{C_P} \cdot 4.455 + 3.568)\ \mathrm{J} \cdot \mathrm{g}^{-1} \cdot \mathrm{K}^{-1}. \end{aligned} \tag{3.10}$$

We fit C_P instead of h because there is no experimental evaluation of *absolute* h, since only enthalpy differences Δh bear physical meaning. To estimate h in SI units, we integrate C_P along T, according to Eq. 3.7.

3.3 Simulation Method

We perform Monte Carlo (MC) simulations of $N = 32,768$ water molecules, at constant P and T in a cubic (variable) volume V with periodic boundary conditions. We update the HB network with the parallel Monte Carlo (MC) algorithm designed to the present model, as explained in Chapter 2. For high temperatures $Tk_B/4\epsilon \geq 0.16$, we apply parallel local Metropolis algorithm, otherwise parallel cluster Swendsen-Wang (SW). We recall that SW avoid the critical slowdown of the dynamics at low-T, but the performance of Metropolis is better at the high-T regime (see Sec. 2.5). We calculate the equation of state along isobars in the range of $Pv_0/(4\epsilon) \in [-0.8, 1.0]$ separated by intervals of $\Delta P \leq 0.1(4\epsilon)/v_0$. The temperature ranges from $Tk_B/(4\epsilon) \in [0.01, 2.0]$ with $\Delta T \leq 0.1(4\epsilon)/k_B$ if $Tk_B/(4\epsilon) \in (0.1, 2.0]$, and $10^{-4}\Delta T k_B/(4\epsilon) \leq 10^{-2}(4\epsilon)/k_B$, if $Tk_B/(4\epsilon) \in [0.01, 0.1]$. We calculate each isobar by sequential annealing starting at $T = 2(4\epsilon)/k_B$ and letting the system equilibrate. For every state point, we average over 10^4 - 10^5 MC steps after equilibration, with a number of independent configurations between 10^3 - 10^4 depending on the state point. We check that the system is in equilibrium probing the fluctuation-dissipation theorem as discussed in Appendix D.

3.4 Results

First, we analyze the behavior of density. We consider that the model results agree with the experiments if their relative difference is smaller than 0.4%. According to this criterion, the density calculations from the model agree with the experimental results in the range $233 \leq T/\text{ K} \leq 425$, $0.101 \leq P/\text{ MPa} \leq 120$ (Fig. 3.1). At negative pressures $P = -20$ MPa and -70 MPa, the model calculations agree with the complete range of temperatures available in Ref. [273]: $257 \leq T/\text{ K} \leq 333$. For $P = -100$ MPa, we find no quantitative agreement. At high-P 160 MPa and 200 MPa, the low-T thresholds are 315 K and 323 K, respectively, and the high-T threshold is 425 K.

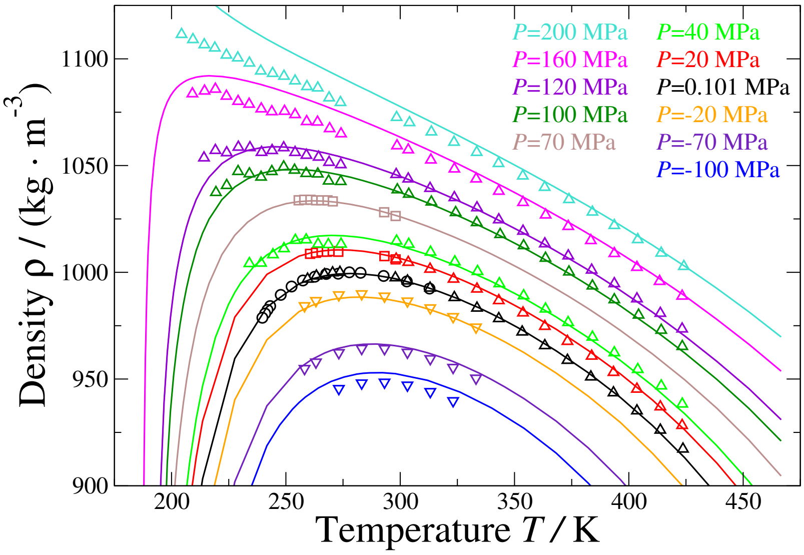


Figure 3.1: Isobaric density: comparison with experiments. Lines correspond to FS calculations and symbols to experimental results reported in Refs. [274] △, [275] ○, [273] ▽, and [276, 277] □. The colors indicate the pressure as listed in the legend.

The FS model can equilibrate large water systems at low-T conditions far from the experimental limit [1, 123, 134]. To benchmark the accessible thermodynamic conditions of the model, we extend our analysis of the density to the deep supercooled region $T \geq 180$ K, for a wide range of pressures $-590 \leq P/\text{ MPa} \leq 260$ (Fig. 3.2). For $P \geq 180$ MPa, ρ increases monotonously, as an Argon-like liquid. In the model, this corresponds to an enthalpy penalty on HB formation $\Delta H_{\text{HB}} = -J + \hat{P} v_{\text{HB}} < 0$. Under these conditions, N_{HB} decreases upon cooling, resulting in a monotonous increase of ρ. For $P < 180$ MPa, we find the TMD, as in water. By further decreasing T, we find that ρ displays a sharp decrease that becomes weaker as P decreases. We understand this behavior as a consequence of the dependence of ρ on N_{HB} [123], since N_{HB} displays a sharp increase at high-P that becomes smooth at low-P. This apparent discontinuity of ρ is consistent with the presence of a first order liquid-liquid phase transition ending in a liquid-liquid critical point, as demonstrated for the FS monolayer [123, 134]. By further lowering T, we find a locus of density minima along isobars (Temperature of minimum Density, TminD), as in atomistic models [171]. Experimental results in confined water are consistent with the occurrence of the TminD line [47,49], but they are controversial [278]. At low-P and high-T we find the liquid-gas (LG) spinodal, defined as the stability limit of the liquid with respect to the gas phase.

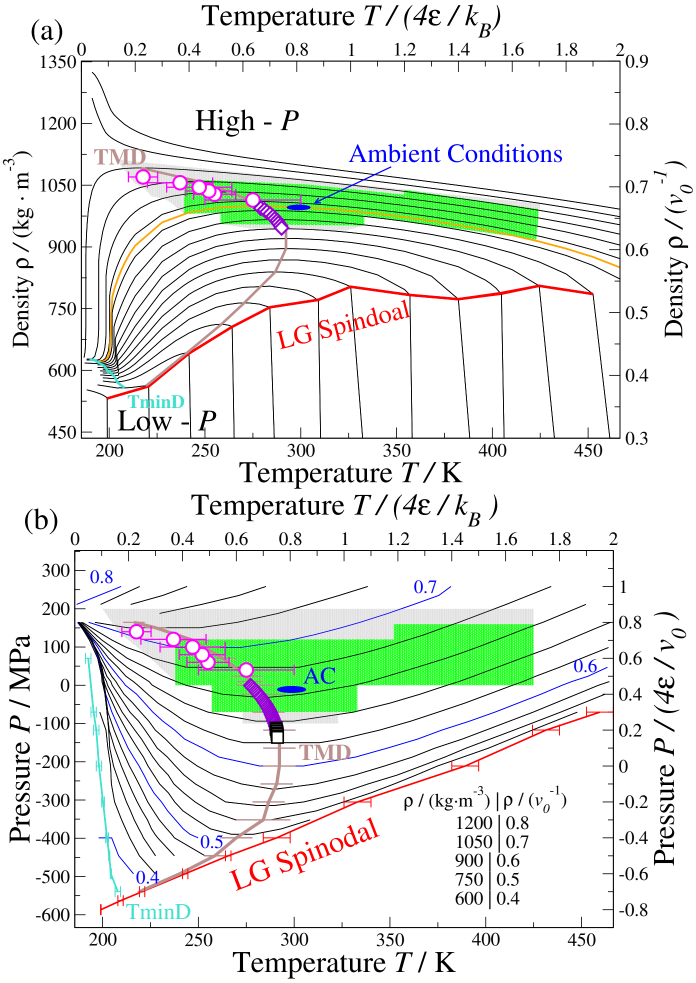


Figure 3.2: Equation of state (T, P, ρ) of the system, including supercooled conditions. (a): Black lines represent the isobaric density variation of the FS model. From top to bottom: $P/4\epsilon v_0^{-1} = 1$ (260 MPa) to $P/4\epsilon v_0^{-1} = -0.8$ (-590 MPa) every $\Delta P/4\epsilon v_0^{-1} = 0.1$ (~ 50 MPa). In addition, the orange line corresponds to the density variation at atmospheric pressure, with $P/4\epsilon v_0^{-1} = 0.45$. The red line indicates the LG spinodal, the brown line the TMD, and the turquoise, the TminD. Symbols correspond to experimental results for the TMD in Refs. [274] ◯ and [273] ◇. The gray shadow highlights the region of experimental results considered in Fig. 3.1. On top of it, the green area marks the region where the model results agree with the experiments, as described in the text. The blue ellipse is a guide for the eye marking ambient conditions (300 K, 0.1 MPa). (b): Black and blue lines correspond to isochores at the T-P phase diagram calculated with the FS model. From top to bottom, $\rho v_0 = 0.8$ ($1200\ \mathrm{kg \cdot m^{-3}}$) to $\rho v_0 = 0.4$ ($600\ \mathrm{kg \cdot m^{-3}}$) every $\Delta \rho v_0 = 0.02$ ($30\ \mathrm{kg \cdot m^{-3}}$). Symbols correspond to experimental results for the TMD line, from Refs. [274] ◯, [279] ◇, and [280] □. Any other element in this panel bears the same meaning as in (a).

We further characterize the quality of the model calculations through the response functions K_T, α_P, and C_P. We estimate the isothermal compressibility by numerical derivation of ρ in Eq. 3.10: $K_T = 1/\langle\rho\rangle \left(\partial\langle\rho\rangle/\partial P\right)_T$. We consider that the model calculations and experimental results agree if they coincide within the error bar (Fig. 3.3 a). At $P = 0.101$ MPa and 10 MPa, the calculated K_T agrees with experiments within the temperature range $255 \leq T/$ K ≤ 354. At larger pressures $P = 50$ MPa and 100 MPa the T-range shifts towards higher values, being $298 \leq T/$ K ≤ 368 and $350 \leq T/$ K ≤ 376, respectively. We do not find agreement at higher-P. The model also reproduces the position of the K_T minima in the phase diagram for $-130 \leq P/$ MPa ≤ 70 (Fig. 3.3 b). However, the locus of K_T maxima is shifted toward lower-T than in experiments [280].

Next, we calculate the thermal expansivity

$$\alpha_P \equiv \frac{1}{\langle V\rangle}\left(\frac{\partial\langle V\rangle}{\partial T}\right)_P = \langle\rho\rangle\left(\frac{\partial 1/\langle\rho\rangle}{\partial T}\right)_P = \left(\frac{\partial \ln\left(1/\langle\rho\rangle\right)}{\partial T}\right)_P \tag{3.11}$$

as a numerical derivative of the density. We manipulate Eq. 3.11 to avoid numerical errors arising from finite difference calculations. We find better agreement for α_P compared with K_T (Fig. 3.4). At atmospheric pressure, the temperature thresholds are 253 K and 383 K. For higher-P, $10 \leq P/$ MPa ≤ 100, the model agrees with the experiments at the range of temperature $273 \leq T/$ K ≤ 373, covering the entire range of available data for $P \leq 70$ MPa. At $P = 160$ MPa, we find that the model agrees

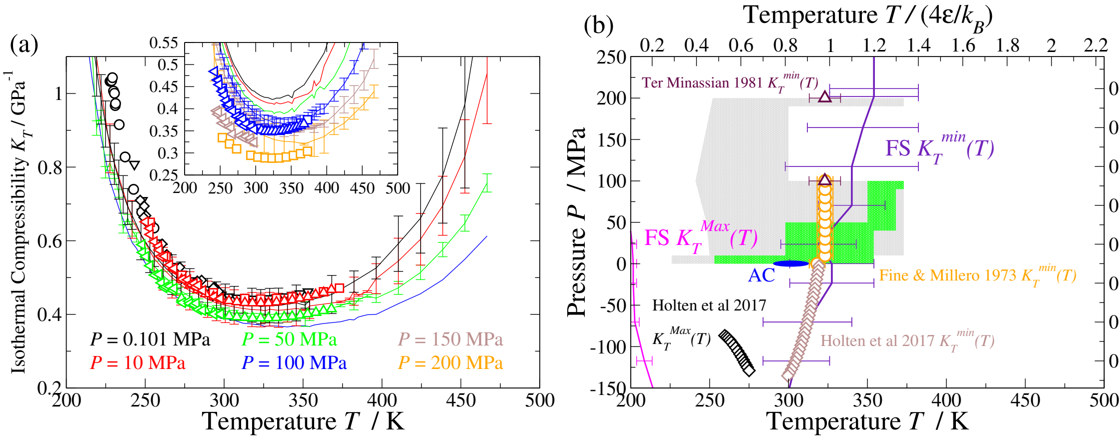


Figure 3.3: Isobaric K_T calculated as the numerical derivative of the density along isotherms. (a): Comparison of FS calculations (lines) with experiments (symbols), as in Fig. 3.1. The experimental results were obtained from Refs. [281, 282] △, [29] ○, [277, 283] ◁, [284] ▽, [31, 277] ◇, and [282, 285] □. (b): Regions of experimental data considered in panel (a) (grey) and agreement between model and experiments (green) for K_T in the P-T phase diagram. We plot the locus of K_T minima (indigo line) and maxima (magenta line) along isobars calculated with the model. Experimental results for K_T minima were obtained from Ref. [286] ○, [285] △, and [280] brown ◇ and K_T maxima from Ref. [280] black ◇. The blue ellipse indicates the ambient conditions, as in Fig. 3.2.

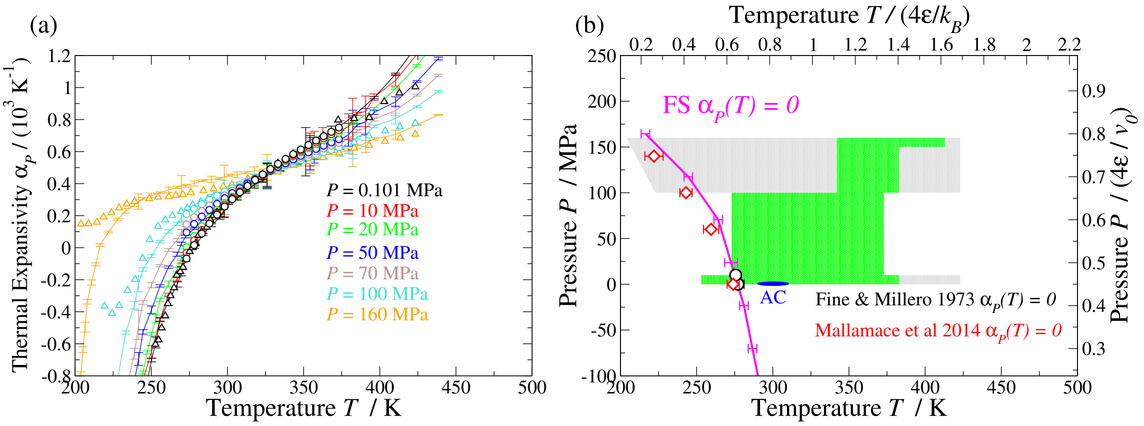


Figure 3.4: Isobaric α_P calculated as the numerical derivative of the density along isobars. Lines and symbols in both panels are as in Fig. 3.3. (a): We consider experimental results in Refs. [286] △, [287] ○. (b): Magenta line is the FS calculation of the locus $\alpha_P(T) = 0$, while symbols correspond to experimental measurements in Refs. [286] ○, and [287] ◊.

within the range $324 \leq T/$ K ≤ 410. Unlike K_T, α_P does not display maxima or minima around 300 K. Therefore, we plot the locus of T where α_P changes its sign, i.e. $\alpha_P(T) = 0$. By definition, this locus is the TMD line. Our results show that the FS estimate of $\alpha_P(T) = 0$ follows the same trend as the experimental results, but shifted to a slightly higher T. Quantitative agreement is only found for atmospheric pressure and 10 MPa (Fig. 3.4 b).

Finally, we calculate the specific heat as the numerical derivative of the enthalpy H with respect to T and as H fluctuations (Eq. 3.7). After rescaling $\hat{C}_P$ according to Eq. 3.10, we compare results with experiments in Fig. 3.5 (a). We find that model calculations agree with experiments in a range of $270 \leq T/$ K ≤ 333 at atmospheric pressure. The T range decreases as P increases, up to $297 \leq T/$ K ≤ 307 K for $P = 100$ MPa. At all P, we find the locus of C_P minima along isobars. However, experiments only show a minimum at atmospheric pressure [288].

We summarize the results in Fig. 3.6 and plot the phase diagram containing the experimental results and model calculations for the TMD and the loci of minima of C_P and K_T. The gray shadow highlights the T conditions where experimental data along isobars are available, i.e. the union of the gray regions in Figs. 3.2-3.5. We advert that it does not cover experimental results on the projection of extrema of the thermodynamic observables into the T-P phase diagram. In light green, we indicate the conditions where the model agrees with at least one observable, i.e. the union of the green regions in Figs. 3.2-3.5. We note that it coincides with ρ agreement. Last, in dark green, we highlight the region where the model is validated, i.e., where the model agrees with all the experiments under consideration. We find that it corresponds to the intersection of K_T and C_P regions of agreement. As a consequence, the model is validated in a range of T $270 \leq T/$ K ≤ 333 at ambient pressure. This interval decreases as P increases up to $297 \leq T/$ K ≤ 307 for $P = 50$ MPa.

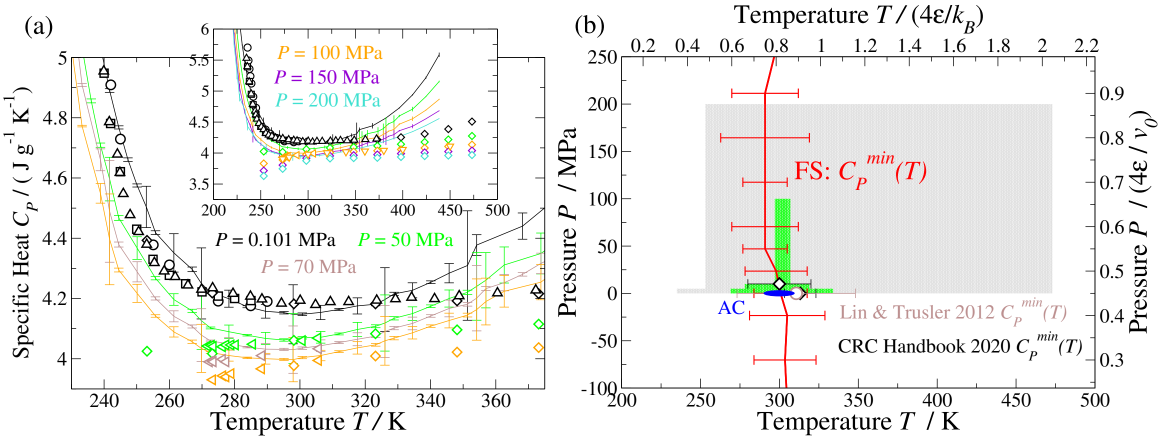


Figure 3.5: Isobaric C_P calculated as numerical derivative of the enthalpy. Lines and symbols in both panels are as in Fig. 3.3. (a): Experimental results are obtained from Refs. [287] △, [30] ○, [289, 290] ◁, [288] ◇, and [291] □. (b): Red line is the locus of C_P minima along isobars, calculated with the FS model, while symbol corresponds to the experimental value in Ref. [288] ○.

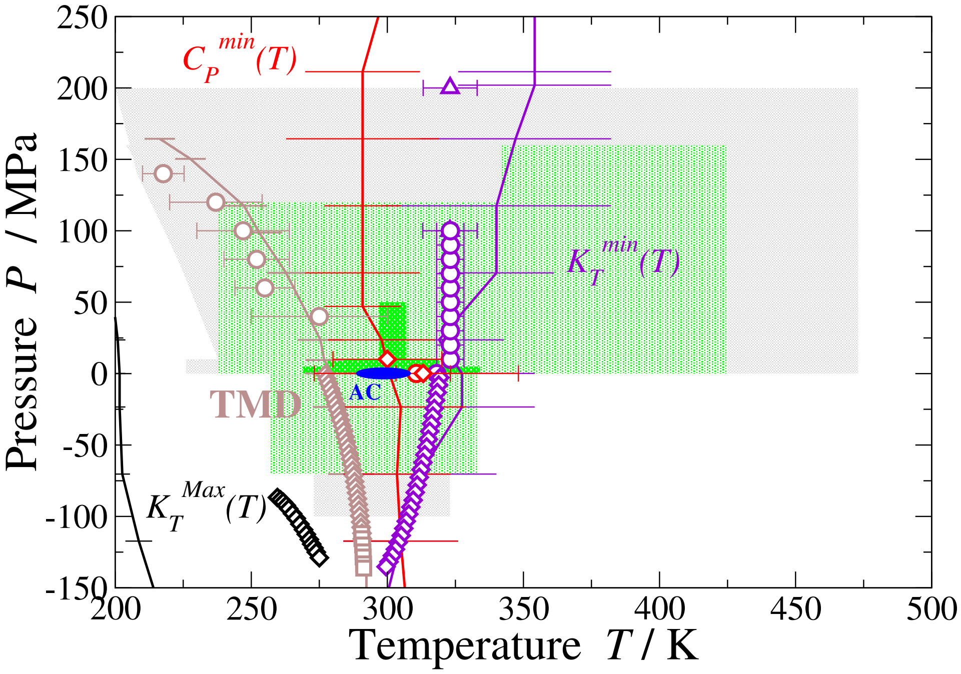


Figure 3.6: FS model validation. In grey, the P-T conditions where experimental data are available, for at least some observable f. This shadow covers only data relating (T, P, f), excluding the projections of extrema of f into the phase diagram. Light-green highlights the region where the FS model calculations agree with at least some observable (density). On top of it, the green region describes the thermodynamic conditions where there is a full agreement between the model and the experiments. The blue ellipse indicates ambient conditions, as in Fig. 3.2. Lines (FS model) and symbols (experiment) correspond to the loci of TMD (brown), $C_P^{\min}(T)$ (red), $K_T^{\min}(T)$ (violet), and $K_T^{\mathrm{Max}}(T)$ (black). For each locus we use the same symbol code as in Figs. 3.2, 3.3, and 3.5.

3.5 Discussion

The results show that the FS model with parameters $J/4\epsilon = 0.5$, $J_\sigma/4\epsilon = 0.08$, and $v_{\rm HB}/v_0 = 0.6$, with the rescaling functions in Eq. 3.10, reproduce the experimental equation of state of water in a range that extends for 60 degrees around ambient conditions (300 K, 0.1 MPa). For increasing P, the range of T decreases to nearly 40 and 10 degrees, for $P = 10$ MPa and $10 < P/$ MPa ≤ 50, respectively. This region covers almost all thermodynamic conditions allowing life on Earth, prompting the FS model for its use as a solvent in biological simulations. The average ocean depth is 3.7 km [292] , which corresponds to a pressure of ~ 40 MPa.

According to our results, the FS model represents the density of water in good agreement for a wide region of the phase diagram (Fig. 3.2). In contrast, the specific heat (enthalpy) poorly agrees with experiments. In particular, we note that for P above 50 MPa, the model results are not in qualitative agreement with the experiments (Fig. 3.5): whereas experimental C_P monotonously decreases upon cooling, the model finds a minimum. We argue that this different behavior of the model compared with the experiments arises from the hypothesis that the cooperativity in the model does not depend on P (Sec. 1.2). The model decouples the formation of directional and cooperative bonds, the former governed by $J^{\rm eff}$, the latter by J_σ. To represent water, cooperative bonds must be formed once the HB network is already built, i.e. $J_\sigma \ll J^{\rm eff}$. Our results indicate that this assumption fails at 50 MPa ($\hat{P} = 0.55$), where $J^{\rm eff} = 0.17 \sim J_\sigma$.

We note that the atomistic model q-TIP4P/F shows a similar trend compared to FS, with increasing C_P upon cooling even at extremely high-P 400 MPa [90]. Moreover, theoretical equations of state of water display minima of C_P for P up to 50 MPa [293] or 70 MPa [290], but at lower T. From these considerations, we conclude that introducing a P dependence on the cooperative interaction could be sufficient to shift the C_P minima toward lower T to be consistent with experiments and in accordance with atomistic models and equations of state.

3.6 Conclusions

In this chapter, we parametrize the FS model for bulk water according to quantum ab initio calculations [267] and experimental data [29–31, 272–277, 279–291]. The results show good agreement between model results and experimental data around ambient conditions: from $270 \leq T/$ K ≤ 330 at $P = 0.1$ MPa to $295 \leq T/$ K ≤ 305 at $P = 50$ MPa. Beyond these thresholds, the model is still qualitatively consistent with water. Thanks to the efficient parallel Monte Carlo algorithms developed in Chapter 2, the bulk FS model can equilibrate unprecedentedly large-size systems of the order of $1.76 \cdot 10^7$ water molecules, reproducing the experimental density and response functions. Thus, the model is suitable for simulating large systems in explicit solvent, as necessary for biological problems. Contrary to other explicit solvent CG models, the FS model includes a full description of the HB network at molecular resolution. This is essential to study the possible effects of polarization and cooperativity of the HBs in the behavior of proteins. We can conclude that the extended FS model is suitable for simulations of large biological systems.

Chapter 4

Supercooled Water

Water is of great interest due to its unusual behavior compared to Argon-like fluids, showing thermodynamic, structural, and dynamic anomalies. To explain its unique properties, several thermodynamic scenarios for supercooled water have been proposed. Recent experiments and simulations support the hypothesis that water exhibits a liquid-liquid phase transition (LLPT) between low density liquid (LDL) and high density liquid (HDL) phases, ending in a liquid-liquid critical point (LLCP). However, definitive experimental evidence is elusive due to rapid crystallization at the low-temperature conditions at which the LLCP is predicted. In this chapter, we study the Franzese-Stanley (FS) model for bulk water at deep supercooled conditions. Our results show the presence of a LLPT ending in a LLCP at $T_C = (186 \pm 2)$ K and $P_C = (174 \pm 14)$ MPa, consistent with estimates from atomistic models. Moreover, the FS phase diagram recalls the structural transformation occurring among experimentally observed LD-amorphous (LDA), HD-amorphous (HDA) and VHD-amorphous (VHDA). We rationalize the transition between VHDA-HDA as a continuous structural change, and between HDA-LDA as a first order phase transition.

4.1 Introduction

Despite the importance of water in many aspects of life, there are still open questions concerning its complex nature [20, 22–24]. As we have already mentioned in Chapter 1, water exhibits more than 60 anomalies [2], like the existence of a density maximum in the liquid phase at ambient pressure and temperature $T \sim 4°$C or the anomalous increase in the specific heat C_P, isothermal compressibility K_T, and absolute value of the thermal expansivity α_P upon cooling liquid water toward the melting line and below it, in the supercooled liquid state [20, 22, 24, 294]. Another intriguing feature of water is polyamorphism [23, 25]. Experiments on amorphous ice reveal the presence of three structurally and dynamically distinct amorphous states, the low density amorphous (LDA), the high density amorphous (HDA), and the very high density amorphous

(VHDA) [295]. However, the nature of the transitions between them is unclear: the LDA → HDA resembles a first order phase transition, whereas HDA → VHDA looks like a weakly discontinuous or continuous transition [23, 296].

The origin of anomalous properties of water has been largely debated since the 80s [23, 24, 36, 40, 78, 85, 297] and a series of thermodynamic scenarios have been proposed [34, 36, 37, 39]. In particular, Poole et al., based on molecular dynamic simulations, proposed a liquid-liquid second critical point (LLCP) in the supercooled region [37]. According to this scenario, the LLCP is located at the end of a first order phase transition separating low density liquid (LDL) and high density liquid (HDL) metastable water phases with different density, structure and energy. The experimentally observed polyamorphism is usually interpreted as a signature of liquid polymorphism, as predicted in the LLCP scenario [45,298]. However, these phenomena are not necessarily connected, as LDA-HDA transition is mechanically induced in non-ergodic states, while LDL-HDL is thermodynamically induced in ergodic states [299].

Multiple experiments and computational approaches support the LLCP hypothesis. On the experimental side, structure factor measurements can identify the fraction of HDL(HDA) and LDL(LDA) in a sample through neutron diffraction, neutron scattering or X-ray scattering techniques [137]. For instance, X-ray scattering experiments on amorphous ice show results consistent with a first order phase transition [300, 301]. More recently, Kim et al. conducted experiments on micro-sized liquid water droplets [29, 41, 42] and bulk [44]. Their results show a sudden change in the structure factor at one order of magnitude shorter times than subsequent crystallization, consistent with LLPT for $T = (205 \pm 10)$ K and pressures between 1 atm and 350 MPa [44], and the presence of a LLCP at positive pressure [29, 44]. On the computational side, the presence of the LLCP has been proven for the ST2 [302], hydrophobically confined Franzese-Stanley (FS) monolayer [134], rigid TIP4P/2005 and TIP4P/Ice [85], and flexible and polarizable WAIL [86] water models, or Stillinger-Weber model for silicon [303]. Moreover, the phenomenological two-states equation of state (TSEOS) assumes the LLCP hypothesis [175] and fits remarkably well to a number of water models [86, 89, 90]. Although the TSEOS fitting method does not constitute a definite proof, it is particularly useful if simulation data are not available at the hypothesized critical point, showing that the model is at least consistent with the presence of a LLCP.

In this chapter, we study the phase diagram of supercooled bulk FS water, that has been parametrized according to ab initio calculations and experiments to reproduce the experimental equation of state of water around ambient conditions (Chapter 3). Thanks to a percolation mapping [123], cluster Monte Carlo (MC) simulations of the bulk FS model are suitable to equilibrate large samples of $N \sim 2 \cdot 10^6$ water molecules at deep supercooled conditions (Sections 2.4,2.5), far above the accessible sizes in atomistic models. Such large-size systems are necessary to simulate problems of biological relevance, including explicit solvent (Chapters 5 and 6).

Following a state-of-the-art procedure [85, 86, 134, 304], we show the presence of a LLCP in the bulk FS model belonging to the Ising 3D universality class. The bulk FS LLCP is located at thermodynamic conditions in accordance to the calculated in atomistic models. Regarding the phase diagram at the critical region, the FS model offers a coherent interpretation of water's polymorphism in terms of a continuous structural change between very high density liquid (VDHL) and HDL and a first order LLPT between HDL and LDL phases. Below the critical pressure, we distinguish among three states of water: VHDL-like, HDL-like, and LDL-like separated by continuous transitions. The distinct forms of water differ in their density, energy, and dynamical properties of the HBs.

4.2 Methods

4.2.1 MC simulations for the phase diagram

We follow the same protocol as in Sec. 3.3. For the sake of consistency, we describe it here using SI units.

We perform MC simulations at constant N, P and T in a cubic (variable) volume V with periodic boundary conditions, adopting the parametrization described in Eqs. 3.9 and 3.10. We conduct extensive simulations for $N = 32,768$ water molecules. We calculate the equation of state along isobars in the range of $-540 \leq P/$ MPa ≤ 160, separated by intervals of $\Delta P \leq 50$ MPa. The range of temperatures is $186 \leq T/$ K ≤ 470, with simulated thermodynamic points at variable resolution $0.014 \leq \Delta T/$ K ≤ 14, depending on the region of interest. The selected minimum resolution corresponds to $\Delta T \cdot k_B/4\epsilon = 10^{-4}$, two orders of magnitude below the characteristic T-scale given by J_σ. We apply the sequential annealing protocol along isobars, starting at high T and letting the system equilibrate. Then, the last configuration, equilibrated at temperature T, is employed as the starting configuration of the following temperature $T - \Delta T$. We average over 10^4 - 10^5 MC steps after equilibration, with a number of independent configurations between 10^3 - 10^4 depending on the state point. For high temperatures $T \geq 215$ K, we apply parallel local Metropolis algorithm, otherwise parallel cluster Swendsen-Wang (SW) to avoid the critical slowdown of the dynamics (Sec. 2.5).

For all pressures P and the entire range of simulated T, we check that the system is in equilibrium probing the fluctuation-dissipation theorem, as discussed in Appendix D.

4.2.2 MC simulations for the estimation of the LLCP

To analyze the finite size effects associated with the divergence of the fluctuations near criticality, we perform simulations of systems of $N = 512$, 1,728, and 4,096 molecules. Considering such small-size systems is necessary to obtain trajectories that show HDL and LDL coexistence, thus adequately describing the free energy barrier between them. Larger size systems display a single HDL→LDL transition upon cooling within accessible simulation times.

For $512 \leq N \leq 4,096$, we apply the sequential annealing protocol starting at $T = 270$ K, as we are interested in the low-T regime. We perform sequential Wolff cluster simulations [305], since we find that this algorithm enhances the crossing of the free energy barrier compared with SW. The minimum resolution in T is $\Delta T = 0.014$ K, the same as for $N = 32,768$, to consistently compare results obtained at all sizes. For each N and P, we tune the simulation time at T conditions close to coexistence so that the system displays multiple transitions between the two states. In particular, we find that the time window required to observe multiple transitions increases in at least one order of magnitude for increasing N, ranging from 10^5 to 10^8 MC steps (Table 4.1). For $N = 4,096$ we only observe coexistence for $P \geq 110$ MPa.

Finally, we attempted to observe multiple transitions at larger sizes. In particular, we performed additional simulations for $N = 8,000$ at $P \geq 95$ MPa. According to our results for smaller-size systems, the expected number of MC steps required to observe multiple transitions should be $10^9 - 10^{10}$. However, even using faster parallel SW algorithm we reach 10^8 MC steps within accessible computational effort (Table 4.1).

4.2.3 Estimation of the LLCP

To rigorously prove the presence of the LLCP, we need to find the correct order parameter x describing the phase transition. In fluid-fluid phase transitions of liquids with short-range interactions, the critical point belongs to the 3D Ising universality class [33, 86]. Hence, a definite proof that the FS model displays a LLCP is that the fluctuations of x behave as expected the expected for the magnetization of the Ising 3D model at the critical point.

According to the NPT-ensemble finite size scaling theory for fluids [83], the order parameter describing a fluid-fluid phase transition is a linear combination $M \equiv \bar{\rho} + s\bar{e}$, where $\bar{\rho}$ and $\bar{e}$ are dimensionless density and energy and s is the so-called mixing parameter. This linear combination is necessary to symmetrize the probability distribution of the states corresponding to HDL and LDL phases, as required at the LLCP.

We apply the histogram reweighting method [306] to find the combination of T_C, P_C, and s for which the fluctuations of M behave as expected for the Ising 3D critical point. First, we perform simulations at T_i and P_i conditions close to the phase

N	Pressure P/ MPa	Multiple transitions?	MC steps (Time Cost)*
512	$P \geq -120$	YES	1e05 (1 min)
	$P < -120$		1e06 (10 min)
1,728	$P \geq -20$		1e07 (2-3 h)
	$P < -20$		1e08 (20-30 h)
4,096	$P \geq 110$		1e08 (2-3 days)
	$P < 110$	NO	
8,000	$P \geq 95$		1e08 (1-4 days)**

Table 4.1: Simulation times considered to analyze the HDL-LDL transition, depending on N and P. We increase the simulation time in one order of magnitude by decreasing P at constant N and by one or two orders by increasing N at constant P. Hence, the analysis is limited to $N = 4,096$ at high $P \geq 120$ and $N \leq 1,728$ for any P. The time cost in parentheses corresponds to a workstation with CPU XenonW-2155 of 3.5 GHz (sequential Wolff update of the σ_{ij} variables) and GPU NVIDIA RTX 2080Ti (parallel Metropolis update of the η_{ij} variables, Sec. 2.2.2).
* "MC Steps (Time Cost)" stand for a single T. For each isobar, we simulate at least five temperatures around the transition.
** Parallel Swendsen-Wang simulations ($N = 8,000$).

transition and estimate the histogram of visited configurations with a given e and ρ, $H_i(T_i, P_i; e, \rho)$. H_i is an approximate calculation of the correct probability density distribution corresponding to a simulation run of infinite time. Next, we combine a set of histograms H_i and calculate the corresponding $H(T, P; e, \rho)$ at T and P conditions close to the simulated T_i and P_i. Further details on the histogram calculation are described in Appendix E. Then, we integrate $H(T, P; e, \rho)$ along the direction of M and rescale the order parameter $m \equiv B(M - M_C)$ so that its probability density distribution $Q(m)$ have zero mean and unit variance. $Q(m)$ is ultimately compared to the probability density distribution of the order parameter at the critical point belonging to the Ising 3D class of universality, $Q_3(m)$.

This method allows for consistently calculating both the location of the critical point (T_C, P_C) and the s defining M, constituting rigorous proof of the occurrence of the LLCP. It has been previously applied in a number of cases, including the FS monolayer [134], confined Lennard-Jones fluid [304], rigid [77,85] and flexible [86] water models, and the Stillinger-Weber model for silicon [303].

Here we find that the calculated $Q(m)$ systematically deviates from $Q_3(m)$ (Appendix F, Fig. F.1). To overcome this issue, we add a correction term to m, testing a new order parameter $x = x(m)$ such that $Q(x)$ fits well to $Q_3(m)$. The corrected order parameter is

$$x \equiv B_0 \left(m + \lambda^2 \sin(m/\lambda) - M_{C,0} \right), \tag{4.1}$$

where λ is a new free parameter in the fitting procedure, and B_0 and $M_{C,0}$ are calculated so that $Q(x)$ has zero mean and unit variance. Further details on the functional form of x and its justification are discussed in Appendix F.

4.3 Results

4.3.1 Bulk FS thermodynamics at the supercooled region

At temperatures below the gas-liquid phase transition (LG Spindoal), we find along isobars a temperature of maximum density (TMD) as in water, see Fig. 4.1 a). By further decreasing T, we observe a sharp decrease in the isobaric density $\rho \equiv N/V_{\rm Tot}$ (we write V instead of $V_{\rm Tot}$ hereafter) and a temperature of minimum density (TminD). The sharp decrease seems to turn into a continuous decrease for decreasing P. This behavior is consistent with the LLPT between HDL and LDL as postulated for supercooled liquid water [37]. However, a similar behavior, but without any discontinuity, is predicted also by the "singularity free" scenario [34, 35, 40]. We therefore analyze the enthalpy behavior in detail.

We find that H follows the density, but with sharper changes at low-P (Fig. 4.1 b). At high-P, the sharp changes in H seemingly turn into continuous changes as P increases, inverting the trend with respect to the density.

We understand the behavior of ρ as a consequence of its dependence on $N_{\rm HB}$ from Eq. 1.1. A direct calculation shows that, by decreasing T, the model displays a rapid increase of $N_{\rm HB}$ at high-P, while the increase is progressive at low-P, as we show in Fig. 4.1 c). In particular, we find that $N_{\rm HB}$ saturates at low-T to two HBs per molecule ($N_{\rm HB}^{\rm Max} \equiv 2N$), corresponding to the case where every water molecule is involved in four HBs. Nevertheless, the changes of $N_{\rm HB}$ at low-P are not discontinuous as in the enthalpy, clearly showing that the contribution to H coming from the other terms in Eq. 1.8 are relevant. The explicit calculation of these terms shows that the dominant contribution comes from the behavior of N_σ (Fig. 4.1 d). We find that N_σ has a sharp increase at low-P that becomes continuous for $P > 70$ MPa. Furthermore, at variance with what observed for $N_{\rm HB}$, the temperature of the largest increase of N_σ is almost independent on P and coincides with the largest variation of H at low-P. Therefore, the large decrease of H is associated with the cooperative contribution that, in turn, is a consequence of a large structural rearrangement of the HBs toward a more tetrahedral configuration. However, this reorganization implies only a minor change in ρ at low-P, as it occurs when the number $N_{\rm HB}$ of HBs is almost saturated. On the other hand, at high-P the restructuring of the internal degrees of freedom of the molecules occurs at higher-T than the formation of a large amount of HBs. Therefore, at high-P the effect of $N_{\rm HB}$ on the density is large and collective, as expected at a critical phase transition, while the effect of N_σ is local (for individual water molecules).

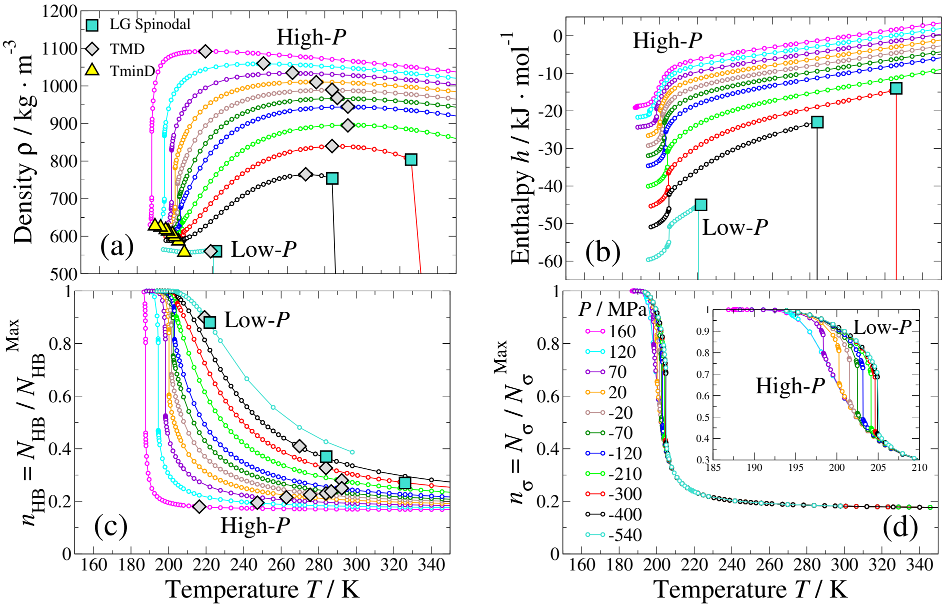


Figure 4.1: Isobaric variation of a) density ρ, b) enthalpy per molecule h, c) normalized number of HBs n_{HB}, and d) normalized number of cooperative bonds n_σ. In panels c) and d), we normalize the number of bonds with respect to the maximum number of bonds the system can form. Each color corresponds to a different P, as shown in the legend. Squares indicate the LG spinodal, diamonds the temperature of maximum density along isobars (TMD) line, and triangles, the temperature of minimum density along isobars (TminD). The results in this plot correspond to the system of size $N = 32,768$.

A further way to clarify if the observed thermodynamic behavior is consistent with the occurrence of a LLPT ending in a critical point is to calculate the response functions C_P, K_T and α_P and to study if they diverge at the hypothesized LLCP. We find sharp maxima in $C_P \equiv \left(\langle H^2\rangle - \langle H\rangle^2\right)/k_B T^2$ at any P and low-T. For $P \leq 20$ MPa, the maxima occur all at (approximately) the same T and increase as the pressure increases, with an apparent divergence of C_P at $P = (-20 \pm 50)$ MPa (Fig. 4.2 a). However, for $P > 20$ MPa, the sharp maxima decrease in intensity and move toward lower-T. This behavior is consistent with a LLCP where $C_P(T)$ apparently diverge, at the end of a first-order phase transition occurring at higher-P along a line with a negative slope in the P-T thermodynamic plane (Fig. 4.2 b), as expected in the LLCP scenario [37].

At T between the TMD line and the LG spinodal, C_P displays a locus of minima (Fig. 4.2 a, inset). The minima occur at a T that is approximately independent on P for $P \geq 70$ MPa. For decreasing P, the locus $C_P^{min}(T)$ follows the TMD line, but shifted at higher-T. By further decreasing P, the locus of minima asymptotically approaches the TMD line and the LG spinodal. At extreme low-P, the locus of minima of C_P merges to the locus of maxima of C_P (Fig. 4.2 b).

At $P \geq 70$ MPa, for thermodynamic consistency [307], the locus of minima of C_P must merge with a locus of maxima of C_P occurring at lower T. However, this locus of maxima cannot be the one we find at very low-T with strong dependence on P

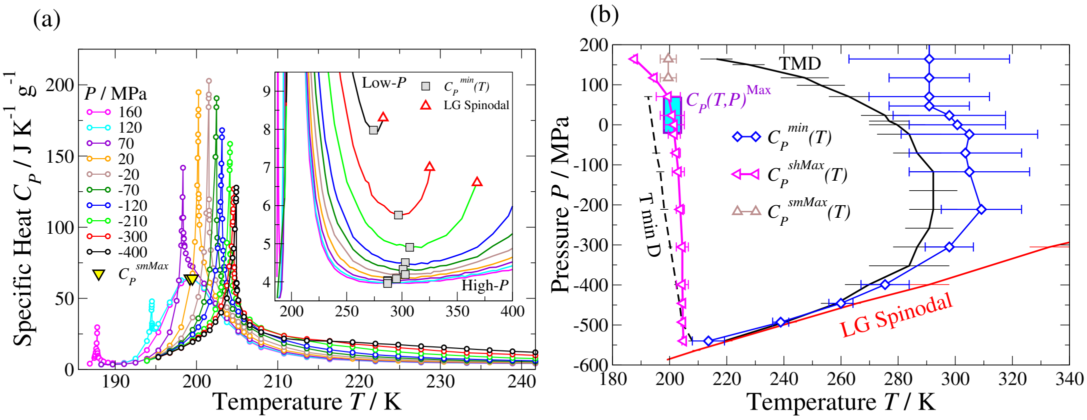


Figure 4.2: (a): Isobaric specific heat C_P calculated as function of T for different P. For the sake of clarity we show only the results from the fluctuation-dissipation theorem. We find loci of minima ($C_P^{min}(T)$, inset) and sharp maxima $C_P^{shMax}(T)$ along isobars. For $P \geq 120$ MPa, we find smooth maxima $C_P^{smMax}(T)$ at $T \sim 200$K. (b): Projection of the loci of extrema of C_P into the P-T phase diagram. Above the temperature of maximum density (TMD) line (black solid line) and below the liquid-to-gas (LG) spinodal (red line) at low-P, the locus of $C_P^{min}(T)$ (blue diamonds). At low-T and above the temperature of minimum density (TminD) line (black dashed line), the loci of $C_P^{shMax}(T)$ (magenta left triangles) and $C_P^{smMax}(T)$ (brown upper triangles). The smooth maxima merge to the sharp maxima where C_P apparently diverges ($C_P(T,P)^{\mathrm{Max}}$, cyan highlighted region).

(and discussed above) because this latter locus moves toward $\hat{T} = 0$ ($T = 185.47$ K according to Eq. 3.10) for increasing P, consistent with its interpretation as the mark of the first-order LLPT. Hence, there must be another locus of maxima of C_P in this high range of P at intermediate T below the temperature of the minima of C_P. This intermediate locus of maxima encloses the anomalous region of C_P together with the locus of minima of C_P at higher T. Indeed, we find that C_P displays a smooth maximum at $T \sim 200$ K at constant $P \geq 70$ MPa. By decreasing P, the locus of smooth maxima merge to the sharp maxima. Although we do not reach the high-P region where the smooth maxima and the minima of C_P merge as in water-like models [307], the likely existence of this merging region goes beyond the scope of this work.

It is worth noticing that two maxima of C_P were found by Mazza et al. [131] in the monolayer case. However, the two maxima occur in the monolayer at any P, also approaching the LG spinodal, at variance with what we observe here. The two low P maxima found by Mazza et al. were consistent with experimental results for a water monolayer hydrating a protein surface [135]. Our understanding is that the two maxima found by Mazza et al. are relevant for monolayer water, but are not present in bulk water or cases that resemble the bulk. This interpretation is consistent with atomistic models for bulk water [90,97,178] and experiments of confined water [308,309] showing that the C_P displays a single maximum upon cooling.

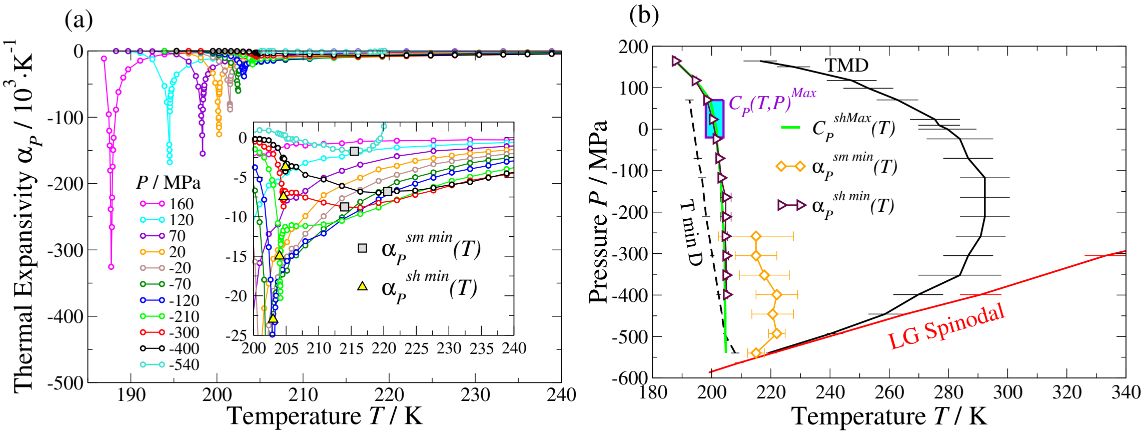


Figure 4.3: (a): The isobaric thermal expansivity α_P displays sharp minima along isobars that decrease for decreasing P. For $P \leq -300$ MPa, the sharp minima become weaker up to disappear (inset, yellow triangles). At these low-P conditions, α_P displays smooth minima (inset, gray squares) at higher-T. (b): Projection of the loci of $\alpha_P^{sh\ min}(T)$ (maroon right triangles) and $\alpha_P^{sm\ min}(T)$ (orange diamonds) into the P-T phase diagram. The locus of $\alpha_P^{sh\ min}(T)$ follow the sharp maxima of C_P (green line). The locus of $\alpha_P^{sm\ min}(T)$ shifts into higher-T and retracts to lower-T as it approaches the LG Spinodal (red line). The TMD, TminD and LG Spinodal lines and the region $C_P^{Max}(T, P)$ are defined as in Fig. 4.2.

Next, we calculate the thermal expansivity $\alpha_P \equiv (1/V)(\partial V/\partial T)_P$ along isobars (Fig. 4.3 a). We find sharp minima at low-T and high-P that decrease for decreasing P, occurring at the same T as the sharp maxima of C_P (Fig. 4.3 b). At $P = -210$ MPa, α_P develops a flat shoulder at T above the sharp minimum. For decreasing P, the shoulder turns into a smooth minimum occurring at higher-T. At the range of P between -250 MPa and -400 MPa, α_P exhibits two minima: a sharp (local) minimum that follows the sharp maxima of C_P, and a smooth (global) minimum at higher-T. We interpret the sharp minimum of α_P at low-T as a result of the increase of crossed enthalpy-volume fluctuations due to the HB rearrangement towards a more tetrahedral ordering.

Finally, we calculate the isothermal compressibility $K_T \equiv \langle V^2 \rangle / k_B T \langle V \rangle$, along isobars (Fig. 4.4 a). We find maxima that resemble the minima of α_P. In particular, K_T displays sharp maxima for that decrease for decreasing P and occur at the same T as the sharp extrema of C_P and α_P. As for α_P, at P below the region where C_P apparently diverge, K_T displays smooth maxima at higher-T. However, at variance with α_P, K_T does not display sharp and smooth maxima along isobars but a kink and a smooth maximum. Furthermore, we observe that the smooth maxima of K_T that, approaching $P \simeq -350$ MPa, turn into minima (Fig. 4.4 b). The minima of K_T occur at T increasing with $P > -350$ MPa and cross the TMD line at its turning point, as can be demonstrated by thermodynamic argument [171], showing that our results are thermodynamically consistent [134].

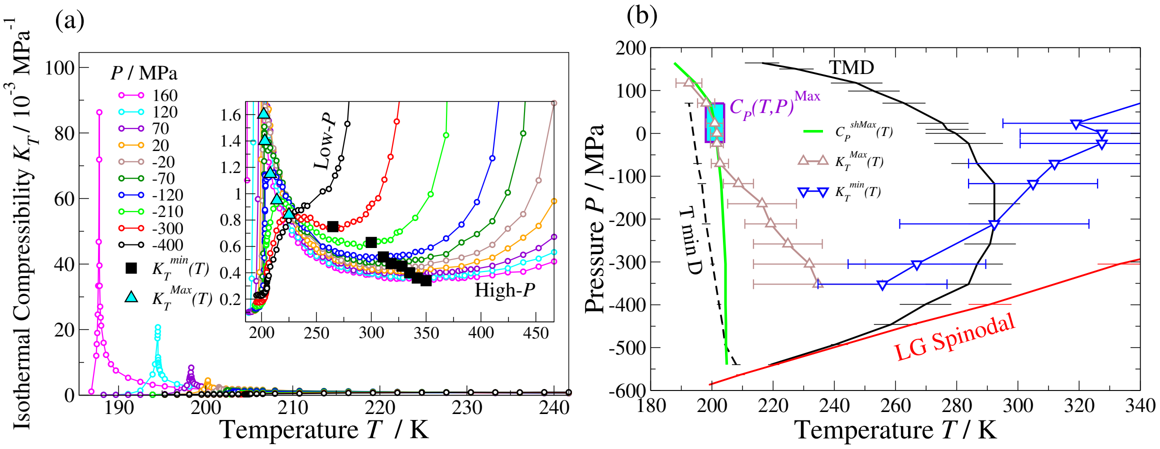


Figure 4.4: (a): Isothermal compressiblity K_T along isobars. The maxima rapidly decrease upon decreasing P and they turn into minima (inset, triangles indicate the maxima and squares, the minima). (b): Projection of the loci of maxima $K_T^{Max}(T)$ (brown upper triangles) and minima $K_T^{min}(T)$ (blue down triangles) along isobars. The TMD, TminD and LG Spinodal lines, and the region $C_P^{Max}(T,P)$ are defined as in Fig. 4.2. For thermodynamic consistency, the locus of $K_T^{min}(T)$ must cross the TMD line in its turning point (point of maximum slope), as we observe in our calculations within the error bars.

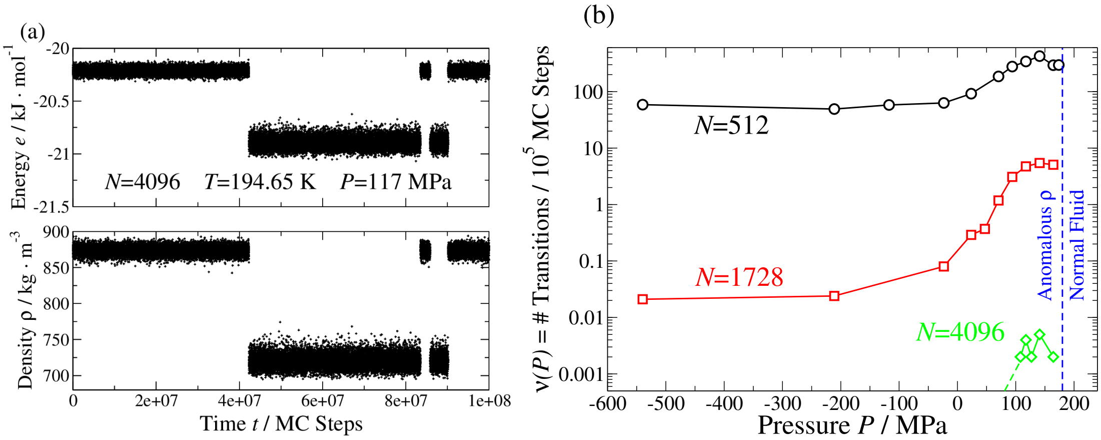


Figure 4.5: (a): Trajectory of the simulation for $L = 16$, $T = 194.65$ K, and $P = 117$ MPa. The system undergoes four transitions between HDL and LDL during the simulation time t. The energy and density are positively correlated, showing that the probability density distribution $Q(e, \rho)$ is bimodal (two distinct HDL and LDL phases). (b): Frequency of transitions $\nu(P)$ along iso-N lines $N = 512$ (black circles), $N = 1,728$ (red squares), and $N = 4,096$ (green diamonds). Green dashed line hypothesizes the behavior at lower-P for $N = 4,096$, as we cannot calculate ν within accessible simulation times. The dashed blue line indicates the maximum P where FS water displays density anomaly. Above this limit, there cannot be a LLPT.

4.3.2 Finite size analysis for the LLPT and the LLCP

Along the LLPT, two distinct phases coexist at fixed T and P, one characterized by high-e and high-ρ (HDL), the other by low-e and low-ρ (LDL). We directly observe phase coexistence in MC simulations as flips of e and ρ during the simulation trajectory (Fig. 4.5 a). The clear correlation between energy and density shows that the probability density distribution $Q(T, P; \rho, e)$ is bimodal, as required at the phase transition. We find coexistence at the entire range of P for small size systems $N = 512$ and $N = 1,728$, and at $P \geq 110$ MPa for $N = 4,096$, within accessible simulation times (Table 4.1). The estimate of the frequency of transitions ν as a function of N and P shows that, for the investigated N, ν is maximum at $P \simeq 120$ MPa and rapidly decreases for decreasing P (Fig. 4.5 b). Furthermore, ν decreases in at least two orders of magnitude for increasing N at constant P. We interpret this behavior at low-P as an apparent phase coexistence induced by finite size effects that become a continuous transition at the thermodynamic limit. On the other hand, the fact that $\nu(N, P)$ exhibits a maximum along iso-N lines is consistent with a LLCP at the P where ν is maximum.

To rigorously localize the LLCP, we need to define the correct order parameter describing the LLPT. To this aim, the method described in Sec. 4.2.3 is well-established [85, 86, 134, 304] and allows for consistently calculating the mixing parameter s that defines the order parameter $M \equiv \rho + s \cdot e$, and the location of the LLCP $\{T_C, P_C\}$. Due to the reasons discussed in Appendix F, we consider an alternative order parameter x with an additional parameter λ, but the method still holds. Regardless the chosen order parameter, the claim that a model exhibits a LLCP is justified by finding an order parameter $x = x(e, \rho)$ whose fluctuations are critical at a thermodynamic point $\{T_C, P_C\}$.

Our results show the presence of the LLCP belonging to the Ising 3D universality class (Fig. 4.6 a). As expected, the free energy landscape $\Delta G(e, \rho)/k_BT = -\log(Q(e, \rho))$ at the LLCP has two basins corresponding to HDL and LDL phases separated by a barrier of $\Delta G \sim 2k_BT$ that thermal fluctuations easily overcome (Fig. 4.6 b).

We report in Table 4.2 the critical parameters calculated for each size L. For $N = 512$, we find fluctuations of x that are critical for a wide range of P and a limited range of T. This region follows the locus of sharp maxima of C_P, which is almost constant in T. The range of P compatible with a LLCP becomes narrower for increasing N, indicating that the fluctuations become less relevant for large-size systems and that finite size effects are relevant.

Given the strong dependence of T_C and P_C on N, we analyze how they extrapolate to the thermodynamic limit. This analysis is necessary to assess whether the observed LLCP results from finite size effects or is intrinsic of the model. According to the finite size scaling theory [83], the scaling laws of $T_C(N)$ and $P_C(N)$ are governed by

N	T_C/ K	P_C/ MPa	s	λ
512	202.7 ± 2.3	-141 ± 164	-36 ± 30	$1.61/\pi$
1,728	200.4 ± 3.4	59 ± 35	-76 ± 38	$1.61/\pi$
4,096	194.6 ± 1.0	117 ± 10	-142 ± 15	$1.53/\pi$

Table 4.2: Critical parameters considered in Fig. 4.6. The rescaled order parameter is $x \equiv B_0 \left(\lambda^2 \sin(m/\lambda) + m - M_{C,0}\right)$, with $m \equiv B(\rho + s \cdot e - M_C)$ as described in Sec. 4.2.3 and further discussed in Appendix F. We fix λ for a given N, then it has no associated error.

the critical exponents of the class of the class of universality. For the 3D Ising model $\theta = 0.53$ is the correction to scaling, and $\nu = 0.63$ is the critical exponent that governs the correlation length ξ, and $d = 3$ [304]. The fitting of

$$\begin{aligned} P_C(N) &= \mathcal{A}_P N^{-(\theta+1)/d\nu} + P_C \\ T_C(N) &= \mathcal{A}_T N^{-(\theta+1)/d\nu} + T_C \end{aligned} \tag{4.2}$$

to the three calculated LLCP(N) gives $P_C = (174 \pm 14)$ MPa and $T_C = (196 \pm 5)$ K (Fig. 4.7). We find that the calculated $P_C(N)$ follow well the expected power law. However this does not hold for $T_C(N)$. Indeed, the estimate of T_C is biased by the smallest-size system $N = 512$. Since we need at least three points to fit the two independent parameters in Eq. 4.2, we cannot estimate T_C considering only the results for $N = 1,728$ and 4,096. To overcome this problem, we include a third point (T_C^*, red diamond in Fig. 4.7 b) calculated under the assumption that the extrapolated P_C is correct and the thermodynamic argument that the LLCP must occur at the locus of sharp maxima of the response functions. We find that $T_C(N = 1,728)$, $T_C(N = 4,096)$, and T_C^* follow the expected power law, giving $T_C = (186 \pm 2)$ K.

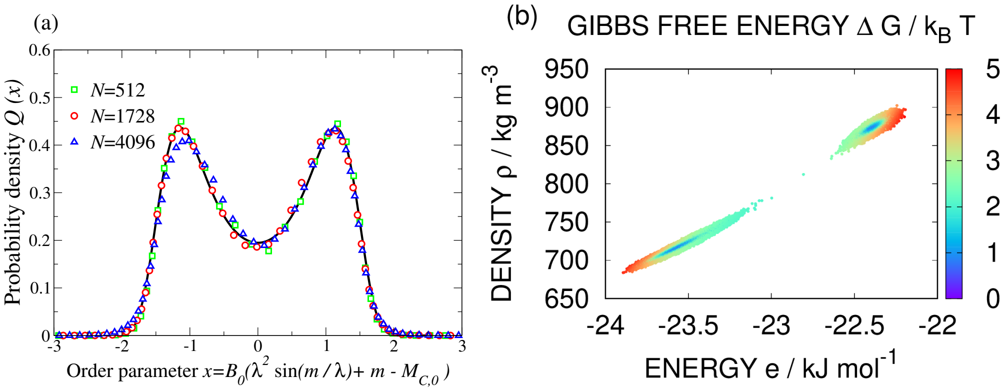


Figure 4.6: (a): Best fit of the probability density distribution of the order parameter $Q(x)$ to the distribution of the Ising 3D at the critical point. The critical parameters are reported in Table 4.2. (b): Free energy landscape at the located LLCP for $N = 4,096$, calculated with the histogram reweighting method.

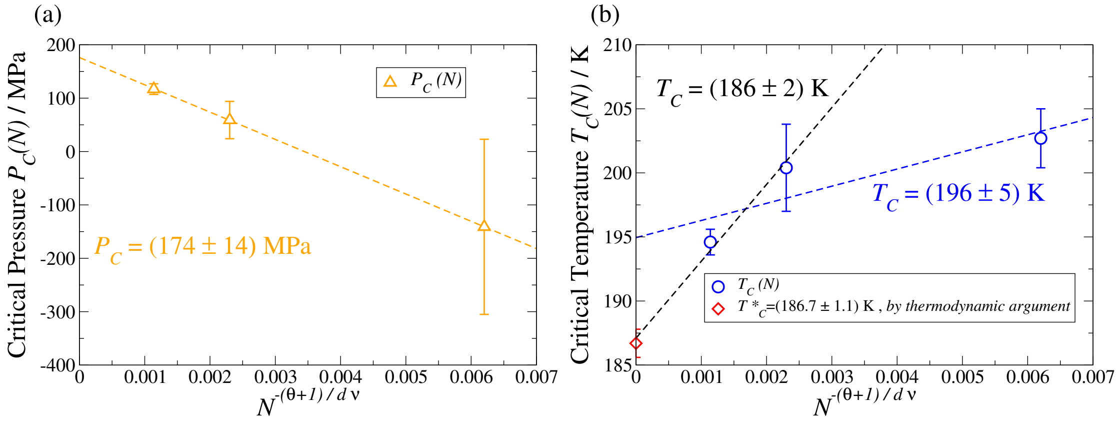


Figure 4.7: The size-dependent $P_C(N)$ (a) and $T_C(N)$ (b) extrapolate to $P_C = (176 \pm 14)$ MPa and $T_C = (196 \pm 5)$ K. We consider the scaling laws in Eq. 4.2 where $\theta = 0.53$, $\nu = 0.63$ are the critical exponents of the Ising 3D universality class and $d = 3$. Panel (b): By thermodynamic argument, we add the critical temperature $T^*_C = (186.7 \pm 1.1)$ K (red diamond) corresponding to the sharp maxima of C_P at the calculated P_C. The three temperatures $T_C(N = 1,728)$, $T_C(N = 4,096)$ and T^*_C extrapolate to $T_C = (186 \pm 2)$ K (black dashed line). Hence, we conclude that the bulk FS model shows a LLCP at $P_C = (176 \pm 14)$ MPa and $T_C = (186 \pm 2)$ K.

Remarkably, the prediction of the LLCP for the FS model compares well with the one calculated in rigid TIP4P/Ice [85] and the approximate estimation from polarizable iAMOEBA [310] (Fig. 4.8). According to analytic calculations and MC simulations on the FS monolayer, the location of the LLCP is sensitive to the choice of J_σ and J, and thus to HB cooperativity [40]. Therefore, the fact that the bulk FS LLCP compares well to models that include cooperative effects (iAMOEBA) or are fitted to reproduce the low-T phase diagram (TIP4P/Ice [311]) supports the calculation of HB cooperativity in FS water. We recall that our choice of J_σ and J is based on ab initio energy decomposition analysis of small water clusters (Sec. 3.2.1).

Finally, we observe that according to our choice of parameters, the limiting P for the density anomaly is $P = 180$ MPa. Above this pressure the LLPT is unlikely. Since we estimate the critical pressure as $P_C = (174 \pm 14)$ MPa, our results support the LLCP scenario [37] (Fig. 4.8).

4.3.3 FS water polymorphism at the critical region

The phase diagram of the bulk FS model recalls that of non-crystalline water [20]. We propose a coherent scenario for supercooled water including three liquid states at pressures above and below P_C with different dynamic and thermodynamic properties (Fig. 4.9). At very low-T and $P > P_C$, there is LDL separated from HDL by a phase transition ending in a LLCP. The LLPT coincides with the locus where the response functions display sharp maxima. We note that the pressure for which the

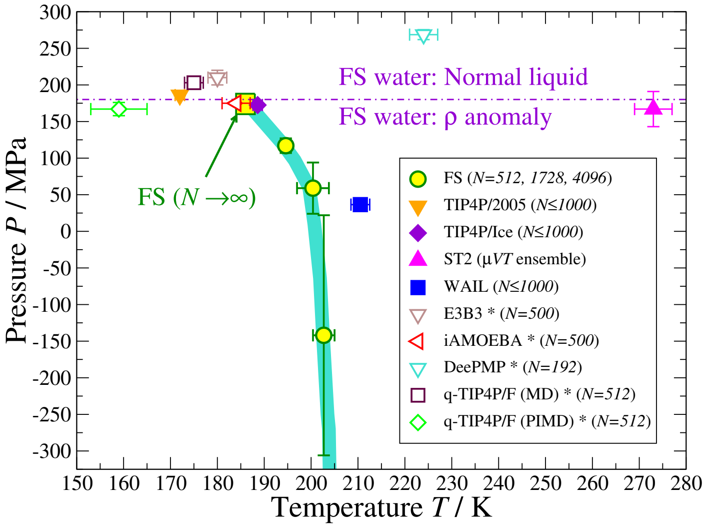


Figure 4.8: Estimations of the LLCP: TIP4P/2005 and TIP4P/Ice [85], ST2 [302], WAIL [86], E3B3 [312], iAMOEBA [310], DeePMP [89], q-TIP4P/F (MD and PIMD) [90] and FS (this work). The thick turquoise line is a guide for the eye indicating the locus of sharp maxima of C_P. By thermodynamic argument, the LLCP (FS model) must occur along this line. Dashed-dotted violet line indicates the limit of density anomaly in the FS model. Above this P, the FS water behaves as a normal liquid, without LLPT.
* void symbols correspond to approximate calculations.

sharp maxima of C_P seem to diverge does not coincide with the estimated P_C. We argue that the LLCP is located, by definition, where the correlation length ξ diverges, rather than C_P. Hence, the divergence of C_P should be considered as an independent but approximate estimate of the location of the LLCP. At high-P, we find smooth maxima for C_P that converge to the sharp maxima where they apparently diverge. We interpret the smooth maxima as a continuous structural change between HDL and VHDL states of water without first or second order singularity in the free energy (Fig. 4.9).

Below P_C we distinguish three states of water: LDL-like, HDL-like, and VHDL-like that reflect the polymorphism above P_C. From the LLCP and following the locus of sharp maxima of C_P, the Widom line separates LDL-like and HDL-like forms of water without phase transition. In particular, we find that, upon crossing the Widom line at extreme low-P, the system undergoes a change in energy (through the cooperative term) with almost no change in density. Thus, $Q(e,\rho)$ is bimodal in energy but unimodal in density. By increasing P, the change in density upon crossing the Widom line increases monotonously, giving rise to a bimodal distribution in both density and energy. We show this by estimating the Challa-Landau-Binder parameter [313] of the density $U_\rho \equiv 1 - \langle\rho^4\rangle/3\langle\rho^2\rangle^2$ upon crossing the locus of sharp maxima of C_P along isobars (Fig. 4.10). U_ρ *measures* the bimodality of the probability density distribution $Q(\rho)$: $U_\rho \to 2/3$ if unimodal, otherwise $U_\rho < 2/3$. We find that for $P \geq 20$ MPa

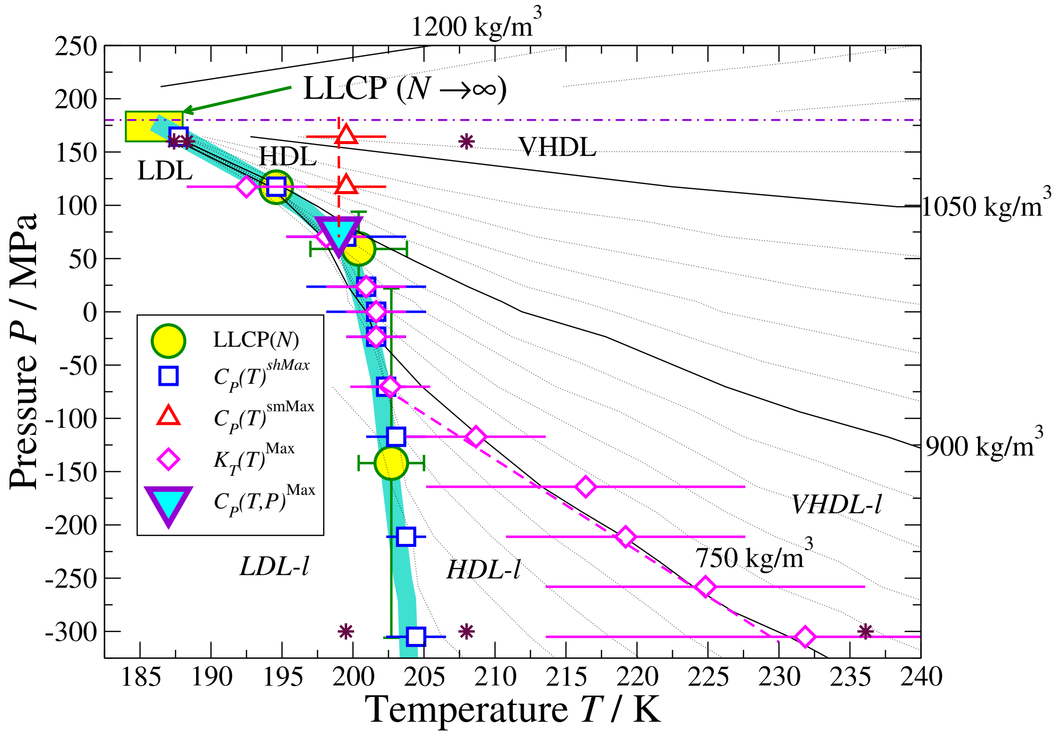


Figure 4.9: Zoom of the critical region in T-P phase diagram. The LLCP(N), thick turquoise line (C_P-maxima line), and dashed-dotted violet line (limit of density anomaly) are as in Fig. 4.8. All the data in this plot correspond to $N = 32,768$, except the locus of LLCP. Black solid lines correspond to isochores separated by 30 kg/m^3. We highlight $\rho = 750$, 900, 1050, and 1200 kg/m^3, as indicated in the panel. On top of the C_P-maxima line, the region where the sharp maxima of C_P seem to diverge (large down triangle). Empty symbols correspond to extrema of the response functions along isobars, as shown in the legend. For the sake of clarity, we do not include the minima of α_P. The LLCP separates the C_P-maxima line into the LLPT between LDL and HDL ($P > P_C$) and the Widom line between the LDL-like and the HDL- like liquid ($P < P_C$). Dashed lines are a guide for the eye that mark the continuous structural change between VHDL and HDL (red) and between VHDL-like and HDL-like (magenta). Maroon stars indicate the thermodynamic conditions for which we calculate the HB autocorrelation function (Fig. 4.11).

(N = 512 and 1,728) and $P \geq 120$ MPa ($N = 4,096$), the distribution is bimodal in density, but unimodal at lower P. This is indicative of a LLPT at $P > P_C$ and a continuous transition along the Widom line. Regarding the loci of (smooth) extrema of K_T and α_P, they initially follow the Widom line but separate as P decreases. We interpret these loci as a reflection of the locus of smooth maxima of C_P above P_C. Hence, we argue that they mark a continuous transition between HDL-like and VHDL-like forms of water (Fig. 4.9).

Finally, we focus on the dynamic behavior of the HB network. To this aim, we perform local Metropolis simulations, as cluster MC avoids the critical slowdown of the dynamics. We calculate the time correlation function $C_M(t)$ defined in Eq. 2.8. and

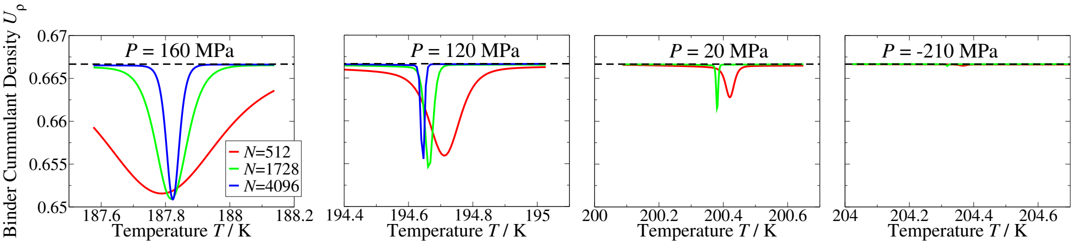


Figure 4.10: Binder cumulant of the density ρ. Each panel corresponds to a different P, decreasing from left to right. Lines correspond to system sizes $N = 512$ (red), $N = 1,728$ (green), and $N = 4,096$ (blue). Results for $N = 4,096$ are not available for $P < 110$ MPa because the system does not undergo multiple transitions within the accessible simulation time (Sec. 4.2.2). Black dashed line indicates $U_\rho = 2/3$, i. e. unimodality of the probability distribution.

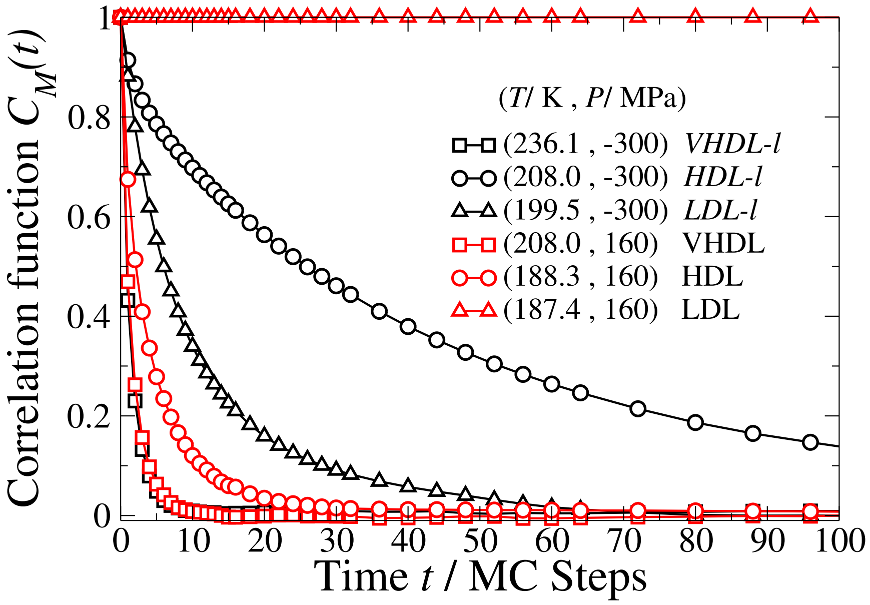


Figure 4.11: Time correlation function $C_M(t)$ for water at the six regions of the phase diagram, as indicated in the legend.

estimate the autocorrelation time of the HB network τ as $C_M(\tau) = 1/e$ (see Sec. 2.5).

At the LDL phase, the system is dynamically arrested, as expected for amorphous ice, while HDL and VHDL have fast dynamics (Fig. 4.11). Our results show that $\tau_{\text{LDL}} \gg \tau_{\text{HDL}} > \tau_{\text{VHDL}}$, which is consistent with LDA, HDA, and VHDA ices corresponding to our LDL, HDL and VHDL [300], although we recall that liquid polymorphism does not necessarily imply ice polyamorphism nor vice versa [299]. Below the LLCP, we do not find dynamic arrest of the system but it could be expected for lower-T in the LDL-like region. The slow dynamics of the HDL-like compared with VHDL-like can be explained due to the proximity of the smooth minimum of α_P to the simulated HDL-like sample. Franzese and de los Santos found a similar non-monotonous behavior of τ at low-P for the FS monolayer [121].

4.4 Conclusions

In this chapter, we show the presence of LDL and HDL phases separated by a LLPT ending in a LLCP in the coarse-grained (CG) FS model for bulk water. This result follows recent simulations of atomistic water models, such as TIP4P/2005, TIP4P/Ice [85] and flexible and polarizable WAIL [86]. Altogether, these models advocate the LLCP scenario for supercooled water [37]. Noteworthy, we find that the estimated $T_C = (186 \pm 2)$ K and $P_C = (174 \pm 14)$ MPa are close to estimations from TIP4P/Ice [85] and iAMOEBA [310]. Hence, we argue that this work reinforces the use of the coarse- grained FS model as a significant water model, not only around ambient (Chapter 3) but also at supercooled conditions.

Regarding the phase diagram in the vicinity of the critical region, we propose a coherent scenario for water polymorphism. More precisely, we distinguish three states of water in terms of their thermodynamic and dynamic properties: the LDL and HDL phases separated by a LLPT, and VHDL separated from HDL by a continuous structural change. The polymorphism above the LLCP is reflected below P_C in terms of LDL-like, HDL-like, and VHDL-like forms of water, separated by structural changes without discontinuities in the free energy.

Since in the FS model all water anomalies arise from the presence of (cooperative) HBs, this work highlights the relevance of HBs and their cooperativity in water modeling. This observation is especially relevant for CG approaches, as models that pack several water molecules into single interacting beads cannot adequately account for HBs nor cooperativity at molecular resolution, most likely impacting their predictions [111].

Chapter 5

Hydration effect in biomolecular simulations: the von Willebrand factor case.

Calculations of conformational free energy barriers for large-scale hydrated biological systems are challenging due to the large number of water interactions and the difficulty of generating statistically uncorrelated configurations. Coarse-grained (CG), implicit solvent (IS) models reduce the computational cost by averaging over the water degrees of freedom and incorporating the water contributions as effective interaction potentials. A noteworthy example is the OPEP model for amino acids. However, by definition, CG-IS models cannot explicitly calculate the solvent contribution to the free energy landscape. In this chapter, we show how to overcome this issue by hydrating CG-IS-generated configurations with the Bianco-Franzese (BF) CG water model (Sec. 1.5), extended to bulk. As proof of concept, we consider the case of the von Willebrand factor (vWf)—a protein involved in cardiovascular diseases made by globular domains connected by intrinsically disordered regions—that expands under hydrodynamic stress. We fully hydrate OPEP configurations of the vWf with $\sim 4 \cdot 10^5$ BF water molecules. We consider three vWf globular (A1, A2, A3) domains and calculate how their hydration free energy landscape depends on the externally applied shear rate. We estimate the solvation contribution to the barriers separating collapsed, detached, and extended conformations of three domains for each shear rate.

5.1 Introduction

Biomolecular systems, such as proteins or membranes, have evolved to adapt their functionality to the aqueous media in which they live [6]. Therefore, considering the properties of water is essential to adequately describe the behavior of these systems [2,3]. According to Thermodynamics, the free energy governs the system. As we discussed in Chapter 1, for biomolecular systems it results from the interplay between enthalpy (solute-solute and solute-solvent: van der Waals, hydrogen bonding (HB), and electrostatic interactions) and entropy (formation and breaking of distorted HBs at the hydration shell) [16].

Computer simulations are a powerful tool to uncover the physicochemical mechanisms that underline the behavior of biological systems [264]. Depending on the level of description of the solvent, models can be classified into explicit or implicit solvent [234]. Explicit solvent models account directly for solute-solvent interactions (see also Sec. 1.1). Atomistic water models, such as the TIPnP family [314, 315] and polarizable AMOEBA [235, 316], are among the most accurate. Unfortunately, their applications to solvate large biological systems suffer from two main drawbacks that limit the accessible length and time scales. First, the number of water molecules in the simulation can be ten times larger than the number of protein atoms [234]. Therefore, most of the computational cost is spent on calculations of water interactions. To overcome this problem, explicit coarse-grained (CG) water models as MARTINI [110] and SIRAH [107, 108] represent the water molecules in a simplified manner. More precisely, MARTINI maps four water molecules into a single interacting bead (4:1), while SIRAH represents tetrahedral structures of eleven molecules as four interacting beads (11:4) [109]. However, their level of description is not suitable to account for the HBs, nor their cooperativity. Due to the relevance of HB interactions, these models are unsuitable to accurately describe the role of the solvent in the behavior of biological systems [111].

The second drawback of atomistic simulations concerns the free energy calculations for complex biological systems, as they typically display local minima separated by large barriers [317]. Direct sampling of the conformational space is challenging because of the huge computational cost to generate uncorrelated configurations [318]. To improve the sampling of the conformational space, including metastable states, enhanced sampling methods have been proposed [317, 318], as umbrella sampling [319, 320] or replica exchange [321].

In contrast to explicit solvent, implicit solvent models represent solute-solvent interactions as effective solute-solute energy terms [232,322–324]. These models enormously reduce the computational cost of the simulations because the number of interactions is reduced. However, by neglecting solvent configurations, the solvent contribution to the free energy landscape remains unspecified.

In this context, the CG Franzese-Stanley (FS) model for water describes the HB network at water's molecular resolution, including cooperative effects [1,134]. In Chapter 2, the FS model has been extended from monolayers to bulk. Thanks to efficient parallel distributed Monte Carlo (MC) algorithms, the model is suitable for large-scale simulations up to at least 10^7 water molecules. Moreover, it reproduces the experimental equation of state of water around thermodynamic conditions of biological interest

(Chapter 3). The FS model for water allows defining the Bianco-Franzese (BF) model for water and proteins (Sec. 1.5, Refs. [5, 6]). The BF model has been applied to investigate protein folding [5, 7], design [6], and aggregation [8, 9, 73]. By assuming that the enthalpy and density fluctuations associated with the HBs at the hydration shell are different from those at the bulk [5, 6], the BF model accounts for the decisive role of water-protein interactions in macromolecular protein conformations.

In this chapter, we investigate the role of water in the free energy landscape of large biological systems previously simulated with implicit solvent models. We solvate each configuration along the trajectory with bulk BF water. After equilibrating, we estimate water-water and water-protein interactions, keeping the protein fixed. Hence, we evaluate how water thermodynamic observables depend on the configuration of the proteins. Assuming that the implicit solvent force field includes all water and protein contributions, this method is suitable to subtract the water interaction energy, provided the energy scales are balanced. Furthermore, as we describe the system in terms of protein and water configurations, we can estimate both the contributions from the protein and the solvent to the free energy of the system. We note that the combination of implicit and explicit solvent models is not novel. For example, Zheng et al. equilibrate condensates of intrinsically disordered proteins using a CG model before mapping the final configuration into all-atom description with explicit water [94]. Similarly, Timr et al. backmap from CG to all-atom selected states of local packing of BSA around SOD1 [325] and BSA and lysozyme around CI2 [326]. However, our approach has a major advantage, as we consider the whole trajectory of the large biomolecular system, rather than a single configuration.

As a proof of concept, we study the rotational dynamics of the A1, A2, and A3 neighboring domains of the von Willebrand factor (vWf) under extreme shear flow [327], previously simulated with the implicit solvent model OPEP [232] and using the Lattice Boltzmann Molecular Dynamics technique [233]. The vWf is a blood plasma giant glycoprotein involved in hemostasis (prevention of blood loss from an injured vessel) and is mainly active in conditions of high blood flow and shear stress as in organs with extensive small vessels, such as skin, gastrointestinal tract, and uterus. A defective function of the vWf protein could lead to blood disorders (e.g., the vW disease [328], or the Heyde's syndrome [329]), especially in those capillaries with higher shear stress, and an increased risk of thrombosis. In vivo it is composed of 40 to 200 monomers, each of 2050 amino acids [330, 331]. Its large size enables mechanical response upon hydrodynamic stress, as shown in experiments [331, 332] and simulations [327, 333–335]. In particular, vWf exhibits abrupt conformational changes from collapsed to extended conformations [336] under high shear rate conditions such as those produced in atherosclerosis or vessel injuries, prompting collagen and platelet adhesion. Platelets (or thrombocytes) are the blood complement in charge of initiating a blood clot. On the contrary, in healthy blood vessels (low shear rate conditions) vWf remains collapsed, hiding adhesion sites to platelets [337].

Several experimental studies consider the neighboring A1, A2, and A3 globular domains as an elementary model for vWf functioning [338,339], as they enable platelet adhesion under shear stress conditions. First, the A1 domain binds GPIbα platelets receptors, a first step to produce a hemostatic plug at the injured site [340]. Second, the A2 domain is responsible for the size control of vWf [341]. Third, collagen binding sites for initial thrombosis are located in the A3 domain. The domains are connected by intrinsically disordered regions, often called linkers. Their flexibility is essential for the mechanical response of the vWf upon hydrodynamic stress [327,335,336,342,343]. In summary, vWf plays a key role in hemostasis and subsequent thrombosis, while its dysfunction is related to cardiovascular diseases [336].

Based on the OPEP simulations on the A1A2A3 fragment [327], we calculate the free energy barriers that separate collapsed and extended conformations, including solvent configurations, and rationalize how they change for increasing shear rate. The comparison to the free energy landscape calculated with implicit solvent data reveals that the solvent contribution is necessary to characterize properly the free energy basins associated to different conformations. Moreover, we estimate the force required to overcome the solvent free energy barrier that separates the A1 domain from the attached pair A2A3. Further experimental approaches could test this result.

5.2 Model and methods

We distribute this section into three subsections. In Sec. 5.2.1 we briefly outline the implicit solvent (IS) CG model for proteins OPEP and in Sec. 5.2.2 we extend to bulk the BF model. Next, in Sec. 5.2.3 we describe in detail our algorithm to map arbitrary OPEP protein configurations into their corresponding bulk BF representation. This algorithm can be easily generalized to other IS models for proteins, membranes or nanoparticles. Finally, in Sec. 5.2.4 we relate the simulation protocol.

5.2.1 The OPEP protein model

OPEP is a multi-resolution CG model for proteins that has been developed to reproduce peptide and protein folding without ad-hoc biases and aggregation [232,344–346]. It is based on a single-bead description of the amino acid side-chain, while the backbone is represented at atomistic resolution [232]. There are two exceptions to this rule: glycine and proline are described at full atomistic resolution. The Hamiltonian of the OPEP model can be split into two sets of potential energy terms that account for bonded (neighboring beads) and non-bonded (both intra- and inter-protein) interactions, and is described in details in Ref. [232]. A fully flexible force-field is necessary to describe large conformational changes, as in the analysis of the exposure of the A2 cleavage site [327]. However, the A1, A2 and A3 domains considered in this work do not undergo such conformational changes [327]. For these cases, an elastic network model based on the OPEP force field, at a reduced amino acid resolution, allows both to describe the dynamics of the system and save computational effort. The elastic network version of the OPEP model represents each amino acid at two-beads resolution: one for the side chain and one for the backbone placed at the carbon alpha (C_α). Exceptions to this rule are alanine, proline and glycine, that are represented using only the C_α bead. The bonded interactions are represented by harmonic springs, while non-bonded interactions are modeled through a variety of short range potentials depending on the amino-acid type.

The combination of Lattice-Boltzmann molecular dynamics (LBMD) with the OPEP force field [233] allows incorporating hydrodynamic effects in the simulations. This method has been exploited in a number of cases, including amyloid aggregation [347, 348], crowded protein solutions [325, 326, 349], nanoscale vesicles [350], and protein unfolding under shear flow [351, 352]. The trajectories considered here correspond to LBMD simulations of the elastic-network OPEP representation of the A1, A2 and A3 domains of the vWf under extreme shear flow [327].

5.2.2 Extension to bulk of the BF model

We extend the monolayer BF model to bulk by introducing the allowance variables $\eta_i j$, as in Eq. 2.1, to guarantee that all molecules form up to four HBs (Sec. 2.2.1). To ensure that the MC algorithms developed in Chapter 2 are still valid, the η_{ij} are defined for every cell in the system, regardless of its occupation by a water molecule or a protein residue.

Regarding the proteins, the BF model for monolayers considers self-avoiding polymers embedded in the lattice partition of the FS water monolayer [5,6]. Here we depart from this approach, adopting the OPEP coarse-grained description at a resolution of two beads per protein amino acid (Sec. 5.2.1). Depending on its size, each OPEP bead occupies a different number of cells, as we will discuss in Sec.5.2.3.

The Hamiltonian term describing water-residue interactions reads

$$\mathcal{H}_{\mathrm{R,w}} \equiv \sum_{i}^{N^{\mathrm{R}}} \left[\sum_{j}^{N^{\mathrm{w}}} C_{ij} S_i^{\mathrm{w}} \right], \tag{5.1}$$

where N^{R} is the number of cells occupied by amino acids, N^{w} is the number of water molecules, S_i^{w} is the water-residue interaction energy, and $C_{ij} = 1$ if the cells i and j are nearest neighbors and 0 otherwise. If the residue is hydrophilic, then $S_i^{\mathrm{w}} < 0$; otherwise $S_i^{\mathrm{w}} = 0$. The model adopts the hydropathy scale from the OPEP force field, considering the non-polar beads as hydrophobic, and polar and charged beads as hydrophilic [349]. The model assumes that the total water-residue interaction energy is equal for all hydrophilic residues, if they are fully exposed to the solvent. However, beads of larger size occupy more cells, and so the number of water-residue contacts increases. Then, S_i^{w} depends on the radius of the beads, as described in Table G.1. At variance with the BF monolayer case, we do not account for residue-residue interactions because these interactions, already accounted for by the OPEP force field, do not contribute to the hydration free energy, the goal of our calculations. Hence, we set $S_{ij} = 0$ (Eq. 1.18).

The presence of the hydrophobic or hydrophilic protein interface affects the water-water hydrogen bonding in the hydration shell [6, 204–208, 353–356]. The model assumes that the hydration shell extends to $n_l = 3$ layers, consistent with recent all atoms molecular dynamics studies [261]. The first is composed by water molecules that are first and second neighbors of protein cells (Fig. 5.1). The subsequent i-th layer is defined analogously as the first and second neighbors of water molecules in the $(i-1)$-th layer. We classify the water molecules in the hydration shell as hydrophobic, hydrophilic, or mixed depending on whether they hydrate hydrophobic, hydrophilic or both classes of residues, respectively (Fig. 5.1).

The parameters for the bulk water have been calibrated to reproduce the experimental water enthalpy and density around ambient conditions (Chapter 3). However,

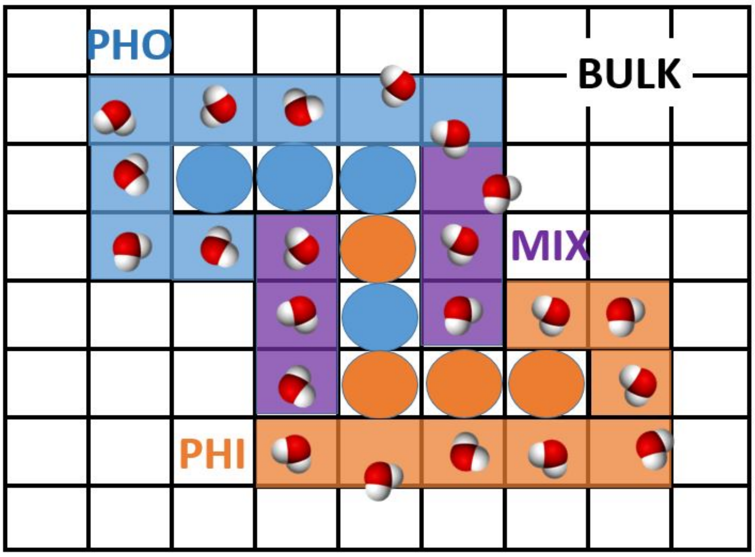


Figure 5.1: Schematic representation of the BF model. Circles indicate protein residues; otherwise, the cell is occupied by a water molecule. For the sake of simplicity: i) protein amino acids occupy a single cell, as in the BF monolayer case (Sec. 1.5), instead of a few cells (bulk BF model), ii) we show a hydration shell of $n_l = 1$ layer and iii) we do not depict the water molecules at the bulk (white cells). Colored cells indicate the hydration shell. Water molecules in blue cells hydrate hydrophobic (PHO) residues (blue circles), molecules in red cells hydrate hydrophilic (PHI) residues (red circles), and molecules in violet cells hydrate both PHO and PHI residues (mixed, MIX shell).

the parametrization of water-water (in the hydration shell) and water-residue interactions is not trivial, and it is beyond the scope of this work, that is limited to a proof of concept. The results described in Sec. 5.3 were obtained in terms of the number of water molecules at the hydration shell N^{h}, thus they are independent on the choice of parameters. In order to explore the role of the parameters, in Appendix G we consider two possible sets of parameters for the HB interactions at the hydration shell. We find that the results obtained in terms of the enthalpy of water are robust against change of parameters.

5.2.3 Mapping of OPEP protein configurations into the BF volume partition

We develop an algorithm to map arbitrary configurations of OPEP proteins into the lattice volume partition of the BF model. In particular, we consider amino acids coarse-grained at two beads resolution (side-chain and C_α), as described in Sec. 5.2.1. The goal is to keep the protein conformation, i.e. the relative distances and angles among the OPEP beads, in the BF representation as possible. We are limited by the spatial resolution of the BF lattice $r_0 = 2.9$ Å, the vdW diameter of a water molecule.

Prior to describe the algorithm, we need to assign a characteristic radius to each OPEP bead, since the radius determines the volume that a bead occupies in the BF lattice. We find it convenient to consider the list of van der Waals volumes V_i for each amino acid reported in Ref. [357] (Table G.1). We assign to each side-chain bead the radius $r_i = \left(3V_i/4\pi\right)^{1/3}$, where i stands for each amino acid. The mapping algorithm is described in the following.

First we map all the side-chain beads, as we assume that they take precedence over the C_α beads. For each side-chain bead we draw a virtual sphere around its center, of radius r_i. Those BF cells whose center is encapsulated by the sphere will be occupied by that side-chain, provided that no other side-chain beads overlap. If several beads overlap at the same cell, we calculate the probability that the OPEP bead i will occupy the cell j as

$$p_{ij} \propto \frac{\exp\left(-d_{ij}^2/2r_i^2\right)}{2\pi r_i}, \tag{5.2}$$

where r_i is the radius of the bead and d_{ij} is the distance between the center of the bead and the center of the cell. For each cell j, the probabilities are normalized so that $\hat{p}_{ij} = p_{ij}/M_j$, $\sum_i \hat{p}_{ij} = 1$, where the sum is performed over the beads that overlap at the cell j. This choice of p_{ij} assigns larger probability to those beads that are small and close to the cell, reducing the possibility that small beads disappear as a consequence of the mapping algorithm. In the limiting case that there is no overlap, the normalization is over a single term, and $\hat{p}_{ij} = 1$.

Next, we map the C_α beads. Regardless of the amino acid, C_α beads are mapped into alanine residues, except the C_α of glycine and proline. We consider that backbones are hydrophobic, as alanine residues. To map the C_α beads we apply the same procedure as for the side-chains, with the only difference that if a side-chain and a C_α overlap, then the cell will be always occupied by the side-chain.

We note that this mapping algorithm can be easily generalized to any hydrated system as nanoparticles, membranes, surfaces, etc. The only restriction is that the radius of the building units (OPEP beads in our case) must be larger (or at least of the same order) than the wdW radius of the water molecule $r_0/2$. Otherwise, the resolution of the lattice is too large to capture the building units. As an example, we show in Fig. 5.2 the application of the mapping method to a collapsed configuration of the OPEP vWf, excluding the two flexible chains that link the three globular domains [327].

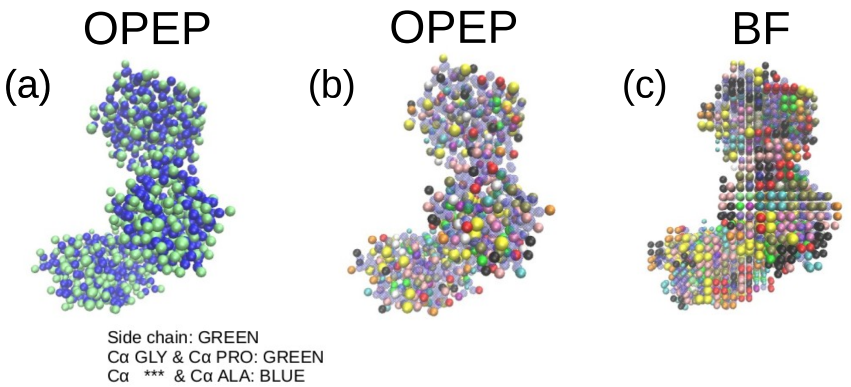


Figure 5.2: Mapping of an OPEP protein configuration into the BF lattice. This example corresponds to the three globular domains A1, A2 and A3 of the vWf, excluding the linkers [327]. (a) OPEP protein where green beads correspond to side-chain, PRO-C_α (i.e., C_α of proline amino acid), or GLY-C_α (i.e., C_α of glycine amino acid) beads; while blue beads represent any other C_α. Blue beads in this panel will be mapped into ALA amino acids, while green beads will be mapped into their respective amino acid. (b) OPEP protein where each color correspond to a different amino acid. Blue beads correspond to the same beads as in the left panel but are plotted in transparent, for the sake of clarity. (c) Result of the mapping into the BF lattice, using the same color code as in the central panel. In this panel each color bead correspond to a BF cell. We clearly note the correspondence between OPEP and BF conformations. Single OPEP beads are mapped into one or a few BF cells, depending on their size.

5.2.4 Simulation method

We consider LBMD-OPEP elastic-network trajectories for the A1, A2, and A3 domains of the vWf [327]. The authors performed simulations with shear rates that expand over four orders of magnitude $1.4 \cdot 10^5 \leq \dot{\gamma}\ \mathrm{s} \leq 1.1 \cdot 10^9$. The simulation times range from 400 to 600 ns, with protein configurations collected each 0.04 ns. In this work we map the OPEP configurations each 0.4 ns, to save computational cost. The total number of configurations mapped into their BF representation is $13,668$.

For each configuration, we perform a parallel Metropolis (Sec. 2.3) MC simulation at fixed $T = 300$ K and $P = 1$ atm, following our parametrization of the model around ambient conditions (Chapter 3). The initial configuration of the water molecules $\{\sigma_{ij}, \eta_{ij}\}$ is set at random. We average over 10^4 independent configurations of water after equilibration.

In Ref. [327], Languin-Cattoën et al. considered a simulation box of 800 Å in X, and 200 Å in Y and Z directions. Here we consider a BF lattice of 152x80x80 cells containing $\sim 4 \cdot 10^5$ water molecules. The size of the box is 440 Å in X, and 230 Å in Y and Z directions. We find this size is sufficient for our purposes because the maximum elongation of the protein system in the X direction is 340 Å. Extending the simulation box merely increases the volume of bulk reservoir, and the computational cost, without affecting the results. We include periodic boundary conditions.

5.3 Results

Results in Ref. [327] show that there is a shear threshold $\dot{\gamma}_C \sim 10^8$ s^{-1} below which the rotational dynamics of the A1 and A3 domains respect to the central A2 critically slow down. For shear rates below the critical $\dot{\gamma} \leq \dot{\gamma}_C$ the three domains remain collapsed (C), while at high $\dot{\gamma}$ the system fluctuates between C and expanded (E) conformations. To characterize these conformations, Languin-Cattoën et al. estimated the distances between A1(A3) and A2 centers of masses. If both distances are short, then the system is collapsed, while if both are large, then it is extended. The system transits between C and E conformations through intermediate detached (D) states in which the A1 domain separates from A2 and A3. The three conformations are shown in Fig. 5.3. The goal of this work is to estimate the free energy barriers due to the solvent, between the three conformations, and analyze how these barriers change as a function of the shear rate.

We first analyze the time evolution of the number of water molecules at the hydration shell N^{h} (Fig. 5.4). We identify three characteristic levels of hydration (dashed lines) corresponding to C, D and E conformations. If the domains are separated, then each domain accommodate a full hydration shell with three layers. Thus, N^{h} is maximum. On the contrary, if the domains approach the hydration shells overlap, reducing N^{h}.

At low shear rates $\dot{\gamma} \lesssim 10^6$ s^{-1} the system equilibrates at the C conformation, while D conformations are metastable. By increasing the shear rate $\dot{\gamma} \sim 10^6$-10^7 s^{-1}, D conformations become more frequent. We find a critical $\dot{\gamma} \sim 10^8$ s^{-1} above which the system cyclically fluctuates between C, E and intermediate D conformations. Our

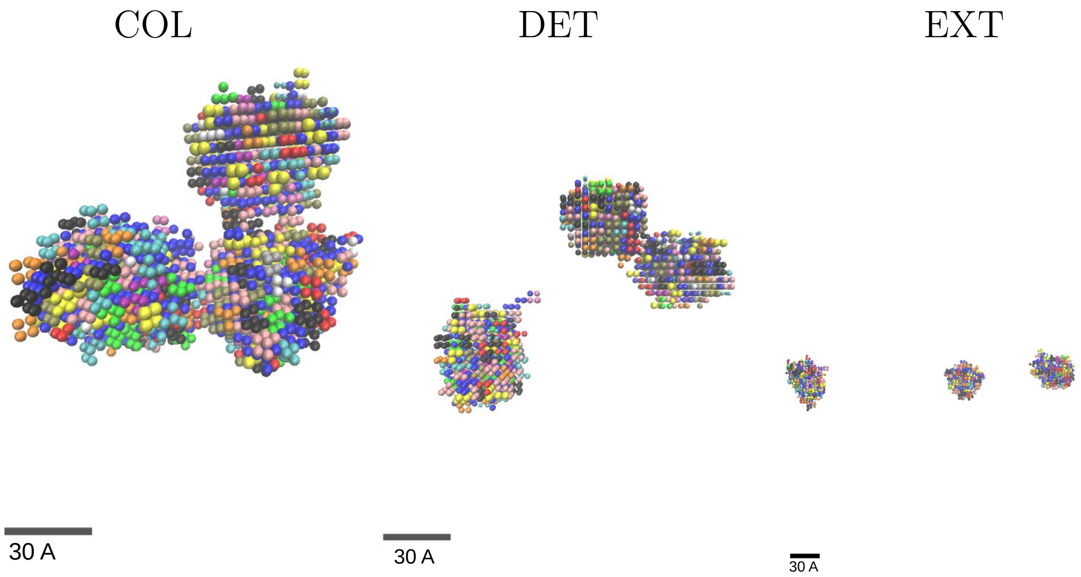


Figure 5.3: Example of collapsed (COL or C), detached (DET or D) and extended (EXT or E) conformations of the A1, A2 and A3 domains of the vWf. Each color correspond to a different amino acid, using the same color scale as in Fig. 5.2 right. We do not consider the linkers between the domains for the BF representation. Scale bar: 30 Å.

results are clearly consistent with those of the original work [327], showing that BF water is sensitive to changes in protein conformations.

For comparison, we estimate the inter-domain OPEP interaction energy ΔE, summing over all pairs of beads at different domains. By definition, the reference $\Delta E = 0$ corresponds to E conformations, since the OPEP potential is expressed in terms of short-range interactions. ΔE accounts for all protein and water interactions in the system, excluding intra-domain interactions and the work exerted by the flow. We argue that intra-domain contributions remain approximately constant because the conformations of the domains are preserved. The time evolution of ΔE is shown in Fig. 5.5.

For low shear rates $\dot{\gamma} < 10^8$ s^{-1}, ΔE progressively decreases towards a minimum. This is consistent with the observation in Ref. [327] that below the shear threshold, the behavior of the system is dominated by thermal fluctuations. For $\dot{\gamma} > 10^8$, ΔE fluctuates between 0 (E conformations) and negative values corresponding to either C or D conformations. For all the values of $\dot{\gamma}$ considered here, ΔE displays sharp spikes with $\Delta E > 0$. The spikes correspond to configurations in which one or more pairs of beads approach to such short distances that they experiment strong electrostatic repulsion. We argue that this is a consequence of the dragging of the amino acids by the hydrodynamic flow. For clarity, high values of ΔE have been truncated in Fig. 5.5. We plot the complete range of ΔE in Fig. H.1.

Qualitative comparison of ΔE and N^{h} leads to the conclusion that ΔE is not a good observable for distinguishing between C, D and E conformations, due to both

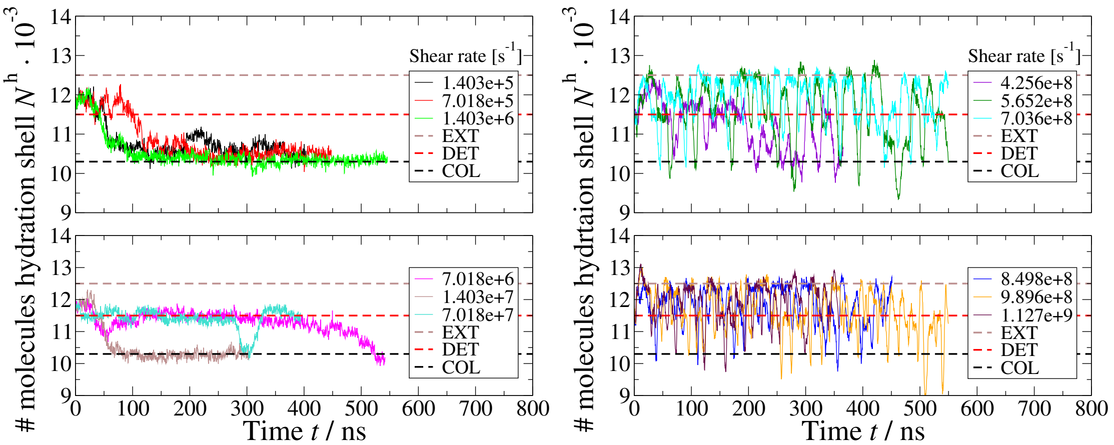


Figure 5.4: Number of water molecules at the hydration shell N^{h} (with $n_l = 3$ water layers) along the trajectory obtained from OPEP simulations. The shear rates are shown in the legend. Dashed lines are a guide for the eye to identify the characteristic hydration levels for (from bottom to top) collapsed, detached and extended conformations. For shear rates $\dot{\gamma} < \dot{\gamma}_C \sim 10^8$ s^{-1} (left panels) the system is dominated by thermal fluctuations, while for $\dot{\gamma} > \dot{\gamma}_C$ (right panels) the system fluctuates among C, D and E conformations.

the absence of characteristic levels (as in N^{h}) and the occurrence of spikes.

As we have already mentioned, LBMD-OPEP simulations of proteins under Couette shear flow incorporate the hydrodynamic effects as effective drag forces $\vec{F}_\gamma$ on the amino acids [352]. Since $\vec{F}_\gamma$ are not conservative, they cannot be derived from the Hamiltonian and the system is out-of-equilibrium. However, we can approximate the shear flow as an external work $W(\dot{\gamma}) \equiv \vec{F}_\gamma(\dot{\gamma}) \cdot \vec{\Delta x}$ exerted on the amino acids [352]. From the point of view of water, the move of the proteins occur at larger timescales than the relaxation of the HB network, so water molecules regard the proteins as if

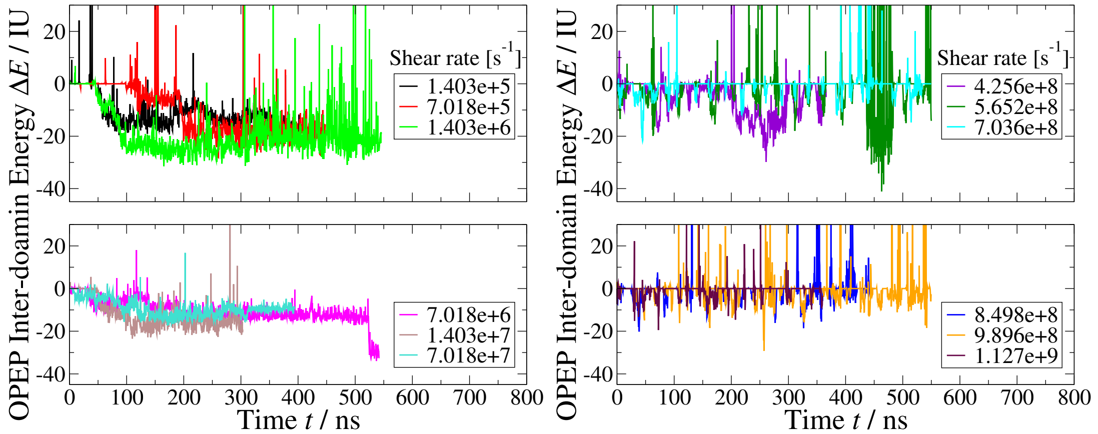


Figure 5.5: Inter-domain energy calculated with the OPEP force field along the OPEP trajectories, expressed in internal units (IU). The shear rates follow the same color code as in Fig. 5.4. We truncate the $\Delta E > 0$ axis to highlight the $\Delta E < 0$ region of interest. The complete ΔE axis is shown in Fig. H.1

they were at equilibrium. Hence, from the equilibrium Thermodynamics point of view, the behavior of the system is governed by the total free energy

$$\Delta G \equiv \Delta U + P\Delta V - T\Delta S + W(\dot{\gamma}), \tag{5.3}$$

where $\Delta U \equiv \Delta U^{\mathrm{w}} + \Delta U^{\mathrm{p}}$ is the internal energy, $\Delta V \equiv \Delta V^{\mathrm{w}}$ is the volume change due to breaking and formation of HBs, $\Delta S \equiv \Delta S^{\mathrm{w}} + \Delta S^{\mathrm{p}}$ is the entropy. Here, the superindex 'w' indicates water-water and water-protein contributions, while 'p' indicates intra- and inter-protein interactions. On the one hand, the OPEP force field accounts for ΔU and ΔS, while $W(\dot{\gamma})$ is included through effective collisions in LBMD simulations. As the OPEP model is calibrated to implicitly include water interactions, we cannot separate the internal energy nor the entropy computed with OPEP into water and protein contributions. On the other hand, the BF model accounts for ΔU^{w}, ΔV^{w}, and ΔS^{w} terms. However, quantitative evaluation of ΔU^{w} requires previous calibration of water-protein interactions, as discussed in Appendix G. Anyhow, direct comparison between energy estimates from OPEP and BF force fields is not allowed, as the energy scales have not been balanced. For these reasons we calculate the global free energy with the Statistical Physics approach:

$$\Delta G(\vec{q}) = k_B T \log\left(Q\left(\vec{q}\right)\right), \tag{5.4}$$

where k_B is the Boltzmann constant, $\vec{q}$ are generalized coordinates that define a macrostate of the system, and $Q(\vec{q})$ is the probability density of finding the system in the macrostate $\vec{q}$. As we consider a system including both the protein and the solvent, we adopt a two-components $\vec{q} \equiv \{R_G, N^{\mathrm{h}}\}$, where R_G is the radius of gyration of the domains. We argue that R_G itself is not a good descriptor of the protein conformation, as it cannot distinguish between C and D conformations if the detached A1 domain is relatively close to A2 and A3. Likewise, if A1 is far from A2 and A3, R_G does not distinguish well between D and E conformations. However, the combination of R_G and N^{h} is a good descriptor of the system. C conformations are characterized by a short R_G and a low N^{h}. When A1 detaches, N^{h} increases to the characteristic D hydration level while R_G remains approximately constant. Next, if A1 separates further from the other two but A2 and A3 remain attached, then R_G increases while N^{h} remains constant. Hence, $\vec{q}$ allows to further distinguish D conformations into detached-collapsed (DC) and detached-extended (DE), see Fig. H.2. Lastly, E conformations are characterized by both long R_G and large N^{h}.

In Fig. 5.6 (central panels) we show the results of $\Delta G(\vec{q})$ for four characteristic shear rates. In Figs. H.3 and H.4, we reproduce the results for all the shear rates considered in this work. To clearly show which conformations the system explores, we plot N^{h} as a function of minimum and maximum distances, where $d^{\mathrm{min(max)}}$ is the minimum (maximum) distance between the centers of masses of the three domains (Fig. 5.6 left panels). This is a slightly different from the criterion adopted in Ref. [327] of estimating the distances between A1(A3) and A2, but the spirit is the same. If both $d^{\max}$ and $d^{\min}$ are short, then the domains are collapsed, if both are large they are extended; otherwise they are in a detached conformation.

For the lowest value of the shear rate $\dot{\gamma} = 1.4 \cdot 10^5$ s^{-1}, the free energy landscape shows a barrier of $\sim 5k_BT$ separating two basins corresponding to C (stable) and DC (metastable) conformations. For increasing $\dot{\gamma} = 7 \cdot 10^7$ s^{-1}, we find three basins corresponding to C, DC and DE states. The barrier separating C and DC decreases for increasing $\dot{\gamma}$, while the barrier separating DC and DE is of the order of 2.5 k_BT. By further increasing $\dot{\gamma} = 4.3 \cdot 10^8$ s^{-1}, we find that the barrier separating C and DC vanishes, and the two basins merge. The free energy landscape for $R_G \gtrsim 50$ Å is almost flat, at ~ 2 k_BT above the basin. For this shear rate, the system does not explore E conformations, but DE, as we appreciate in the left panel. Lastly, for the highest $\dot{\gamma} = 1.1 \cdot 10^9$ s^{-1} we barely appreciate any barrier, and the system explores a rugged free energy landscape covering all the possible conformations. This corresponds to a system dominated by the effect of the hydrodynamic flow that completely flattens the free energy landscape.

We can estimate the force F required to overcome the free energy barrier ΔG(C↔DC). F(C→DC) is the force necessary to separate the A1 domain from the attached A2 and A3 in the field of all protein and water interactions, i.e., the force required to activate the hemostasis response. On the contrary, F(DC→C) is the force that thermal fluctuations must exert to deactivate the system. We calculate the thermodynamic force $F \equiv \Delta G/\Delta l$, where ΔG is the free energy barrier and Δl is the increase in the distance between A1 and the nearest domain A2 upon detachment (attachment), as a function of the shear rate. We find that the maximum force that the collapsed system stands prior to A1 detachment is

$$F^{\mathrm{Max}}(\mathrm{C} \rightarrow \mathrm{DC}) \equiv \frac{(4 \pm 0.5) \cdot 300 \text{ K} \cdot k_B}{(17 \pm 6) \text{ Å}} = (10 \pm 5) \text{ pN}, \tag{5.5}$$

calculated for $\dot{\gamma} = 7 \cdot 10^5$ s^{-1}. We report the calculated forces for all shear rates $\dot{\gamma} < 10^{-8}$ s^{-1} in Table H.1. These predictions could be tested in future experiments.

Finally, we compare $\Delta G(\vec{q})$ with an alternative description of the system that only considers information from the implicit solvent simulations: $\vec{q}^{\,\mathrm{IS}} = \{R_G, \Delta E\}$. As we have already discussed, ΔE is less sensitive than N^{h} to changes of protein conformations, particularly between C and D. Moreover, the sharp spikes in Figs. 5.5, H.1 impede the interpretation of the free energy landscape in terms of C, D and E basins. Snapshots with $\Delta E > 0$ correspond to either C or D conformations, but they are disconnected in the $\vec{q}^{\,\mathrm{IS}}$ landscape to other macrostates corresponding to the same conformations. We plot $\Delta G(\vec{q}^{\,\mathrm{IS}})$ in Fig. 5.6 (right panels). For the lowest shear rate $\dot{\gamma} = 1.4 \cdot 10^5$ we distinguish two basins separated by a barrier. These are consistent with the free energy landscape ΔG projected into $\vec{q}$ coordinates. For increasing $\dot{\gamma} = 7 \cdot 10^7$ s^{-1}, we find a single basin for C and DC conformations separated by a

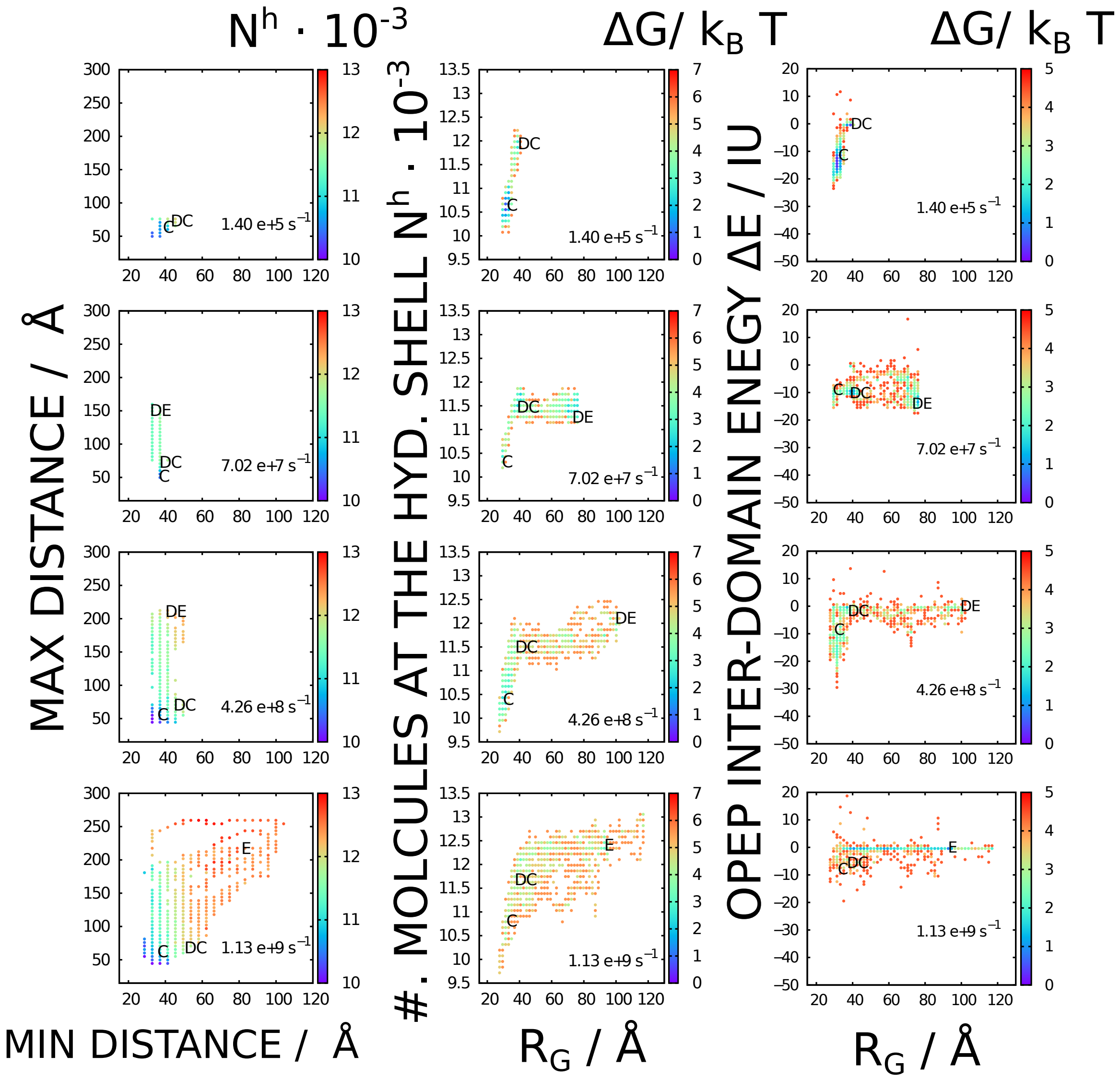


Figure 5.6: From left to right: number of water molecules in the hydration shell N^{h} as a function of minimum and maximum distances, $\Delta G(\vec{q}) \equiv \Delta G(R_G, N^{\mathrm{h}})$, and $\Delta G(\vec{q}^{\,\mathrm{IS}}) \equiv \Delta G(R_G, \Delta E)$, where R_G is the radius of gyration of the protein system and ΔE is the inter-domain energy calculated with the OPEP force field. From top to bottom: four characteristic shear rates $\dot{\gamma} \cdot \mathrm{s} \in \{1.4 \cdot 10^5, 7.0 \cdot 10^7, 4.3 \cdot 10^8, 1.1 \cdot 10^9\}$. The results corresponding to other shear rates are plotted in Figs. H.3 and H.4

barrier form DE. Here we note two clear differences between the ΔG projections into $\vec{q}$ and $\vec{q}^{\,\mathrm{IS}}$. While for the description in terms of $\vec{q}$ we still find a free energy barrier between C and DC, it vanishes in the $\vec{q}^{\,\mathrm{IS}}$ diagram. Moreover, we find that DC conformations correspond to either $\Delta E \sim 0$ ($\dot{\gamma} = 1.4 \cdot 10^5$ s^{-1} panel) or $\Delta E \sim -10$ ($\dot{\gamma} = 7 \cdot 10^7$ s^{-1} panel). For higher $\dot{\gamma} = 4.3 \cdot 10^8$ s^{-1} we find a basin corresponding to C and DC conformations, and a flat free energy plateau for large R_G, as in the $\vec{q}$ description. Next, for the largest shear rate $\dot{\gamma} = 1.1 \cdot 10^9$ s^{-1}, we find that this description gives no relevant information. We find a narrow region where G is minimum around $\Delta E/\mathrm{IU} \sim 0$ for all R_G, collecting configurations at D and E states. We do not interpret this locus as the collapse of the basin in the $\vec{q}$ projections as many of these conformations correspond to $-10 < \Delta E/\mathrm{IU} < 0$. Thus, ΔE is unable to distinguish properly among C, DC and DE conformations.

Comparison between $\Delta G(\vec{q})$ and $\Delta G(\vec{q}^{\,\mathrm{IS}})$ landscapes shows that the former is much more informative than the latter. We find that the inclusion of a coordinate for the solvent is necessary to separate C and DC conformations into distinct free energy basins. Otherwise they collapse into a single basin, i.e. if the macrostate is described in terms of R_G and/or ΔE. R_G accounts for the separation of DC and DE, as they are indistinguishable from either energetic or hydration shell points of view. Last, separation between DE and E should be accounted for by the solvent coordinate. However, the change of N^{h} depends on the conformation of the globular domains A2 and A3 before they separate. The more the hydration shells overlap, the larger the increase in N^{h} will be. We find a broadening of the accessible regions in the $\vec{q}$ landscape with a plateau in $\Delta G(\vec{q})$ for large values of $\dot{\gamma}$, instead of clearly distinct DE and E basins.

In Appendix G (Figs. G.3 and G.4) we reproduce the results in this section in terms of the enthalpy of water ΔH^{Sol} (Eq. G.1), calculated with two different sets of parameters of the BF model, instead of N^{h}. We find that the results are robust against changes of the solvent descriptor and parametrization.

5.4 Conclusions

In this chapter, we introduce a theoretical method to estimate the solvent contribution to the free energy landscape of large biological systems. We consider the BF model for water and proteins that accounts explicitly for water-residue interactions and describes the HB network at molecular resolution, including cooperative effects [1,134]. The model is suitable for large-scale simulations, up to the order of 10^7 water molecules (Chapter 2 and Appendix B) around ambient conditions (Chapter 3). Our method maps arbitrary solute (proteins, membranes, nanoparticles, etc.) configurations calculated with implicit solvent models into the CG representation of the BF water and protein model. After equilibrating the HB network, keeping the solute fixed, we calculate the enthalpy of water. Assuming that implicit solvent trajectories already include all water effects, our method makes explicit the contributions coming from water to the total free energy.

As a proof of concept, we consider implicit solvent OPEP model simulations of the rotational dynamics of the A1, A2, and A3 domains of the vWf under extreme shear flow [327]. Our results show that BF water is sensitive to changes in protein conformations through changes in the shape (number of hydration water molecules) and enthalpy of the hydration shell. We find that these solvent descriptors, together with the protein radius of gyration, allows us to characterize properly the free energy basins of the different conformations. On the one hand, our approach allows us to estimate the free energy barriers between the conformations as a function of the shear rate, showing that the barriers flatten out above the critical shear rate $\dot{\gamma}_C \sim 10^8$ s^{-1}. On the other hand, it conveys that, below the critical shear rate, thermal fluctuations are enough to overcome these free energy barriers. In both cases, although the quantitative estimate of the free energy barriers depends on the model parametrization and should be optimized comparing with experiments, the results in terms of the number of hydration water molecules is robust. Also, we demonstrated that the experiment-based quantitative optimization of the parameters would allow us to better estimate the maximum tensile strength the protein can sustain as a function of the shear rate.

Overall, our approach demonstrates that the inclusion of the explicit solvent allows a better description of the free energy landscape, compared with descriptions involving only the solute. More precisely, the solvent coordinate is necessary for characterizing properly different free energy basins. Therefore, we overcome an intrinsic limitation of implicit solvent CG models since, by construction, they average over the solvation coordinates.

Chapter 6

Effect of crowding in FUS condensates

Water is a constitutive part of biomolecular materials, as it plays a decisive role in balancing energy and entropy. In particular, hydrogen bonds (HBs) within the hydration shell modulate the amino acid interactions and control local physical properties such as dielectric permittivity, viscosity, etc. Despite the water relevance, coarse-grained (CG) models for proteins with implicit solvent, e.g., the OPEP, are popular because they allow us to study large length and time scales that are unreachable for atomistic simulations with water. Here, we consider an alternative approach based on the Bianco-Franzese (BF) model for water (Sections 5.2.2 and 5.2.3). In this chapter, we combine the OPEP and the BF models to study the effect of crowding on Superoxide Dismutase 1 (SOD1) sequestration into Fused in Sarcoma (FUS) biomolecular condensates. SOD1 proteins are involved in the disease progression of amyotrophic lateral sclerosis (ALS) and, under heat stress, are sequestered into Stress Granules (SGs) in vivo and into FUS biocondensates in vitro. A crowded in vitro cytomimetic medium –a Bovine Serum Albumin (BSA) solution– decreases the SOD1 partition coefficient (PC) even after 60 min of heat stress [349]. OPEP Lattice Boltzmann Molecular Dynamics simulations of SOD1 with FUS Low-Complexity Domain and BSA show almost no preferential interactions of SOD1 with BSA, despite the conformational difference of (intrinsically disordered) FUS and (globular) BSA. Samanta et al. concluded that the decrease in PC with BSA is due to the lack of preference for SOD1 between FUS and BSA [349]. By studying the hydration contribution to the system's free energy balance, we show that the hydration free energy landscape of SOD1 in solution with BSA is entirely different from the case of SOD1 in a FUS condensate. We find for SOD1 three preferred associative states in BSA and only one in FUS and show that the hydration controls the transition rates and the residency times of each associative state. Our study concludes that the in vitro SOD1 PC decrease in FUS condensate in the presence of BSA crowders is due to the hydration entropy increase in BSA. This mechanism could be relevant not only in vitro but also in vivo.

6.1 Introduction

Biomolecular condensates organize the cellular fluid and play an important role in cell functioning as ribosome biogenesis, DNA damage response, signal processing and control [358,359]. They are often called membraneless organelles (MLOs) and correspond to the droplet phase upon protein liquid-liquid phase separation (LLPS) [358,360–363]. Experimental studies have shown that a common feature of many MLOs is that they exhibit liquid-like properties, such as fusion, dripping [364], high viscosity [365,366], or wetting [367]. Other studies addressed physicochemical factors that affect droplet stability, such as temperature [368,369], pH [370,371], or ionic strength [372].

Stress granules (SGs) are MLOs formed at cell stress conditions, such as starvation [361] or heat stress [373]. Proteins are known to global or partially unfold upon heating, and misfolded proteins can aggregate into toxic assemblies [374]. SGs constitute a cytoprotective mechanism, sequestering misfolded proteins or aggregates [375–378] and thus preventing the development of diseases. For example, Superoxide Dismutase 1 (SOD1) proteins, which are related to the progress of amyotrophic lateral sclerosis (ALS) [379,380], are sequestered upon heating in cell into SGs and in vitro into Fused in Sarcoma (FUS) condensates [349].

Low complexity domains of intrinsically disordered proteins, such as FUS [381–383], heterogeneous nuclear ribonucleoprotein A1 (nhRNPA1) [384,385], or DEAD-box helicase protein LAF-1 [365], commonly drive LLPS [365,386,387]. Although simplistic in vitro experiments show that homotypic FUS interactions are enough to induce phase separation [381], the cellular environment is a complex, crowded, multi-component mixture of biomolecules. Crowded environments typically favor LLPS by excluded volume effects, co-condensation of crowding agents and phase-separating biomolecules, or by inducing segregative phase separation [362]. In in vitro experiments, synthetic polymer molecules such as Ficoll 70 or polyethylene glycol (PEG) are employed to mimic cellular crowding [388]. However, these molecules are relatively inert respect to the phase separating protein, neglecting the chemical specificity of the protein-crowder (CWD) interactions. Bovine Serum Albumin (BSA) constitutes a step toward a more realistic representation of cytoplasm crowding in in vitro experiments [349,389,390], since it is highly water-soluble and introduces, at some extent, protein-protein interactions as in the cellular milieu [391].

In this work, we study by simulations the SOD1 sequestration in FUS condensates and in BSA crowded mediums. The motivation for comparing SOD1 sequestration into these CWDs is two-folded. On the one hand, this simulation study helps to rationalize in vitro experiments comparing SOD1 partition into FUS condensates in synthetic polymer Ficoll 70 and in globular protein BSA crowded mediums [349]. These results show that the partition coefficient (PC) of SOD1 in FUS is reduced for the case of FUS condensate in BSA compared to Ficoll 70. On the other hand, the comparison is relevant due to the different sequence-specific interactions, crowding and excluded volume effects, and solvent composition in both environments. For BSA (globular), most of the volume is occupied by a reservoir of bulk water, while FUS chains (intrinsically disordered) are homogeneously distributed over the entire volume of the systems, with SOD1s experiencing larger excluded volume effects.

Despite these differences, Lattice-Boltzmann molecular dynamics (LBMD) simulations of the OPEP model [233, 352] -an implicit-solvent coarse-grained model for proteins that keeps sequence specificity and assumes short range interactions [232]-, yielded no relevant SOD1-CWD interaction energy differences but slightly stronger preference for partitioning into BSA compared to FUS [349]. Samanta et al. concluded that the minor PC of SOD1 into FUS condensates when BSA crowders are present, compared to Ficoll 70, is rationalized by the finding that SOD1 interaction energy with BSA is, at least, of the same order of magnitude as the interaction energy with FUS.

Water and, in particular, hydrogen bonds (HBs) at the hydration shell modulate the free-energy landscape of biomolecular systems [2, 16]. However, implicit-solvent models integrate over the solvation degrees of freedom of the system. Hence, their level of description hides the solvation coordinates of the system. In this work, we hypothesize that including explicit solvent interactions is essential to elucidate the different behavior of SOD1 sequestered into FUS and BSA environments. Following the method described in Sec. 5.2.3, we map configurations sampled with the implicit-solvent OPEP model [232, 233, 349] into their corresponding BF water-protein representation and estimate the solvent contributions to the free energy of the SOD1s. We find that the hydration free energy landscape of a) SOD1 into FUS and b) SOD1 with BSA are entirely different, something that was revealed in the implicit solvent calculations. In particular, we rationalize the experimental preference of SOD1 adsorption into BSA by the increase in the hydration entropy in the BSA environment.

6.2 Methods

Lattice Boltzmann molecular dynamics (LBMD) simulations of the implicit solvent OPEP model [232, 233, 352] have been performed to analyze SOD1 sequestration into BSA [325, 389], as well as Chymotrypsin Inhibitor 2 (CI2) sequestration into BSA and Lysozyme [326] crowded environments. In this chapter we consider the sequestration of folded SOD1 in BSA and FUS highly concentrated solutions [349]. Following the method described in detail in Sec. 5.2.3, we map the configurations of the system calculated with the OPEP model into their corresponding BF protein and water representation and sample water configurations keeping the protein fixed. Then, we analyze how the thermodynamic properties of water, namely the enthalpy, density, and response functions, change with the protein configuration.

The size of the BF simulation box and the temperature T and pressure P conditions are set to mimic those of the original work [349]. We consider a cubic box of 88 cells of lateral size, corresponding to 255 Å. Periodic boundary conditions (PBC) apply in all directions. The LBMD-OPEP simulations extend for ~ 1 μs and configurations were collected every 0.2 and 5 ns for the SOD1 in BSA and FUS cases, respectively. Here, we map the configurations into their respective BF representation every 2 ns (BSA) and 5 ns (FUS). The system is then solvated with $N^{\mathrm{w}} \sim 6.3 \cdot 10^5$ water molecules, filling the free volume left after mapping the proteins. Among the water molecules, those located at the first three water layers from the protein surface belong to the hydration shell. For every configuration, we equilibrate the solvent at ambient conditions ($T = 300$ K, $P = 1$ atm as in Chapter 3) and sample over 10^4 independent water configurations. The sampling algorithm is Metropolis Monte Carlo in the NPT-ensemble, parallelized in GPUs (Sec. 2.3).

The simulated systems are of comparable concentration, containing 10 SOD1 in 70 FUS chains (150 g/L) and 10 SOD1 in 15 BSA proteins (100 g/L) [349]. In Fig. 6.1, we represent a typical configuration of the systems, mapped into their corresponding BF representation, as described in Sec. 5.2.3.

6.3 Results

We classify the N^{w} water molecules in the system into bulk and hydration shell, $N^{\mathrm{w}} \equiv N^{\mathrm{b}} + N^{\mathrm{h}}$, and further distribute the water molecules at the hydration shell

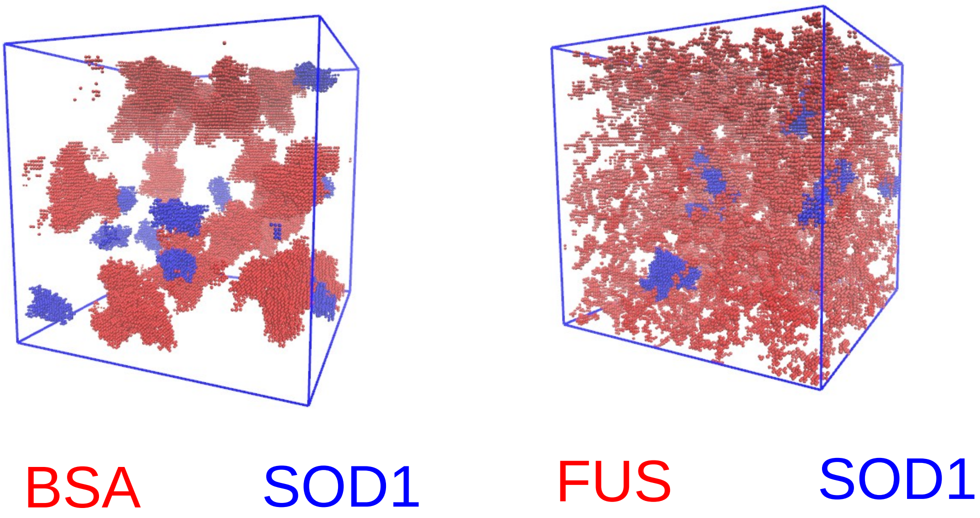


Figure 6.1: SOD1 sequestration in BSA (left) and FUS (right) condensates, mapped into the BF lattice. Blue beads correspond to cells occupied by SOD1, while red correspond to cells occupied by the other protein. Water molecules fill all the available volume left by the proteins. For the sake of clarity, they are not represented. Blue lines represent the limits of the simulation box (255 Å), where PBC apply.

into three groups depending on whether they hydrate SOD1, CWD, or both. Hence, $N^{\rm h} \equiv N^{\rm h}_{\rm SOD1} + N^{\rm h}_{\rm CWD} + N^{\rm h}_{\rm mix}$. More precisely, $N^{\rm h}_{\rm mix}$ is the number of water molecules in the hydration shell that are at the same distance from the SOD1 and the CWD. In Fig. 6.2, we depict a section of the system to illustrate the classification of water molecules into bulk and SOD1, CWD, and mixed hydration shells.

As expected from the different CWD conformational properties, Fig. 6.2 shows that the composition of the solvent in BSA and FUS solutions is entirely different. In the BSA solution, most of the volume of water is in the bulk, while the hydration shell extends to a limited region surrounding the proteins. Oppositely, in the FUS solution, most of the water molecules hydrate FUS chains. It is a consequence of the different conformational properties of the crowders: BSA is globular, while FUS is intrinsically disordered and flexible. Therefore, the amino acids of FUS are both homogeneously distributed in the system and more exposed to the solvent.

Despite the different composition of the solvent, our estimates of the total enthalpy of water (water-water and water-protein interactions) are similar for both cases, being just ~ 0.6 kJ/mol larger (more negative) for the FUS condensate (see Table I.1 in Appendix I). We rationalize this difference to the larger enthalpy of HBs at the hydration shell and to the increase in the number of water-protein contacts in the FUS solution compared to BSA.

Following the Chapter 5, we estimate the free energy landscape of an individual SOD1

$$\Delta G(\vec{q_i}) = k_B T \log \left(\rho_P \left(\vec{q_i} \right) \right) , \tag{6.1}$$

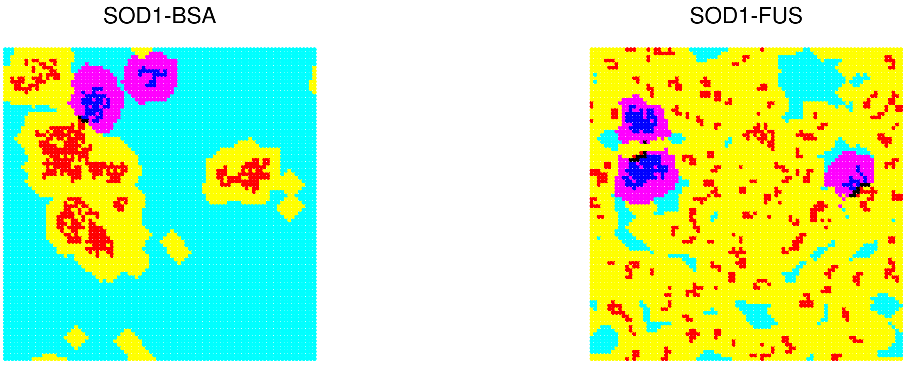


Figure 6.2: Typical configuration of the system, including proteins and water. Left: SOD1 and BSA, right: SOD1 and FUS. For clarity, we represent a section of the system. Each color point represents a BF cell, i.e., either a water molecule or a fragment of an amino acid. Color code: proteins are represented in blue (SOD1) or red (BSA and FUS in the left and right panel, respectively). Bulk water molecules are in cyan. We depict the hydration layer in magenta (molecules hydrating a SOD1), yellow (molecules hydrating the other protein in solution) and black (molecules hydrating both proteins in solution).

where k_B is the Boltzmann constant and $\rho_P(\vec{q_i})$ is the probability density that the i-th SOD1 is in a macrostate described by the generalized coordinates $\vec{q_i}$. We consider a descriptor of the macrostate of the SOD1 with two components, where the first describes the configuration of the proteins and the second, the solvent. Hence, the first component collects information from OPEP implicit solvent simulations, while the second accounts for new information arising from BF explicit water. We let $\vec{q_i} \equiv \{\mathcal{C}_i, N^{\mathrm{h}}_{\mathrm{SOD1},i}\}$, where $N^{\mathrm{h}}_{\mathrm{SOD1},i}$ is the number of water molecules hydrating the i-th SOD1, and $\mathcal{C}_i$ is the adsorption factor, as defined in the following.

For each SOD1 i and configuration of the system at the time t, we define the adsorption profile as $c_i(d,t) \equiv N_i(d,t)/N_i(t)$, where $N_i(d,t)$ is the number of BF cells of the i-th SOD1 exposed to the solvent and at a distance d from the crowder, and $N_i(t)$ is the total number of BF cells of the i-th SOD1 exposed to the solvent. The physical meaning of $c_i(d,t)$ is the fraction of the i-th SOD1 surface that, at the time t, is at the distance d from the CWD. In Fig. I.1, we plot the thermodynamic average of SOD1-BSA and SOD1-FUS adsorption profiles $c(d) \equiv \langle c_i(d,t)\rangle$, where $\langle\cdot\rangle$ indicate the average over time and the ten SOD1s. We find that $c(d)$ reflects the different conformational properties of the CWDs. In particular, $c(d)$ of SOD1 in FUS exhibits a single peak at short distances d, while that of SOD1 in BSA broadens to large d values. As we discuss in the following, this result indicates that the SOD1s can separate from the BSA, contrary to the FUS case. We define the adsorption factor as the integral

$$\mathcal{C}_i(t) \equiv \int d \cdot c_i(d,t) \cdot \mathrm{d}d. \tag{6.2}$$

Hence, $\mathcal{C}_i(t)$ is the average distance between the surface of the i-th SOD1 and the surface of the CWD at the time t.

Our results, including the solvent, show that the free energy landscape of SOD1 in BSA and SOD1 FUS are entirely different (Fig. 6.3 top panels). The SOD1s in the BSA solution explore three characteristic states connected in a single free energy well. We label the states 'A', 'B', and 'C'. We define the states A as those with adsorption factors shorter than two hydration shells $\mathcal{C}_i < 2d_{\text{Hyd. shell}} \sim 25$ Å(six water layers). At these short $\mathcal{C}_i$, the hydration shell of the SOD1 overlaps with the hydration shell of the BSA. At these short $\mathcal{C}_i$, the hydration shell of the SOD1 overlaps with the hydration shell of the BSA. Hence, the SOD1s in state A interact with the BSA, at least through water-mediated interactions. For larger $\mathcal{C}_i$, we find two basins (B and C) separated by a free energy barrier of $\lesssim 3k_BT$. B and C states differ in the number of water molecules at the hydration shell. For $N^{\mathrm{h}}_{\mathrm{SOD1},i} > 3,350$ the SOD1 is surrounded by a completely-formed hydration shell (state C). On the contrary, if $N^{\mathrm{h}}_{\mathrm{SOD1}} < 3,350$, the hydration shell of SOD1 overlaps with another hydration shell. As $C_i > 2d_{\text{Hyd. shell}}$, the only possibility is that two SOD1s interact between them (state B). Regarding the free energy landscape of SOD1 in FUS, we find a single basin corresponding to A states. This was expected, as FUS is homogeneously distributed in the system, and there is no bulk reservoir for the SOD1 to escape (states B and C). In Fig. 6.4, we depict the three typical configurations of the SOD1s in BSA.

For the sake of comparison, we calculate the free energy landscape using generalized coordinates that only account for information obtained from OPEP implicit solvent simulations. We consider $\vec{q}_i^{\,\mathrm{IS}} \equiv \{\Delta E_i, \mathcal{C}_i\}$, where ΔE_i is the interaction energy of the i-th SOD1 with all the other proteins in the system, calculated with the OPEP force field (see Fig. 6.3 bottom panels). The free energy landscape of SOD1 in BSA resembles the calculated including a coordinate of the solvent. However, configurations with $\Delta E \lesssim 0$ are much more populated that those with large and negative ΔE. In other words, states of SOD1s with different hydration volume $N^{\mathrm{h}}_{\mathrm{SOD1}}$ collapse to similar values of ΔE. We rationalize this observation to the fact that the short range interactions calculated with the OPEP force field extend up to 10 Å, while water-

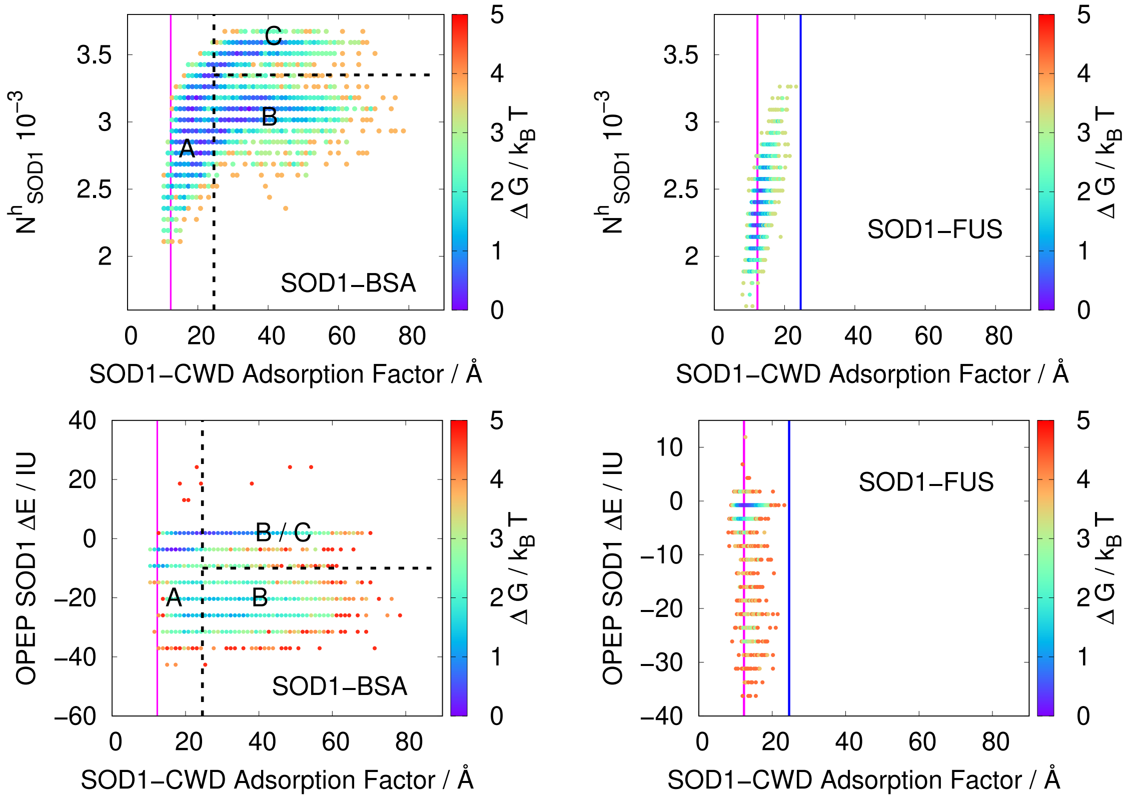


Figure 6.3: Hydration free energy landscape of SOD1 into BSA (left panels) and FUS (right panels) solutions. Top panels: $\Delta G(\vec{q})$, calculated from explicit solvent simulations. Magenta and blue solid lines indicate $\mathcal{C} = d_{\mathrm{Hyd.\ shell}}$ and $\mathcal{C} = 2d_{\mathrm{Hyd.\ shell}}$, respectively. Black dashed lines separate the landscape into three regions corresponding to A, B and C states, as described in the text. The dashed line separating A from B and C states is at $\mathcal{C} = 2d_{\mathrm{Hyd.\ shell}}$. Bottom panels: $\Delta G\left(\vec{q}^{\,\mathrm{IS}}\right) \equiv \Delta G(\Delta E, \mathcal{C})$, calculated from implicit solvent simulations. Lines bear the same meaning as in top panels. “B/C” indicates that this basin of ΔG includes both B and C states, as discussed in the text.

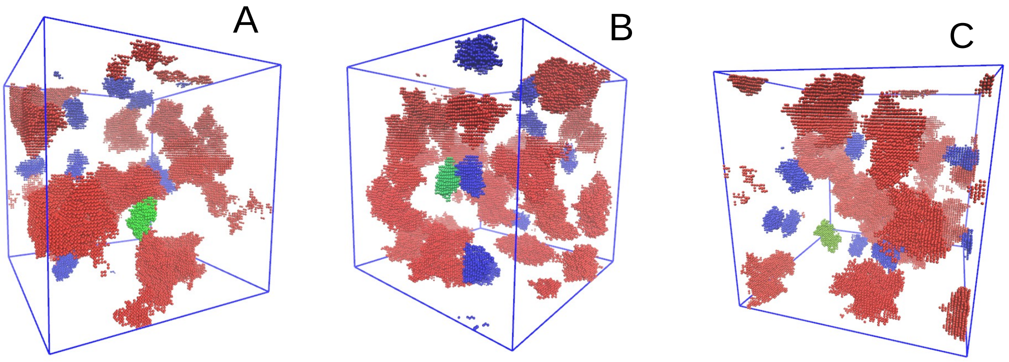


Figure 6.4: The three associative states for a SOD1 in solution with BSA. The states are defined for individual SOD1 (green), whereas the remaining SOD1s in the system are represented in blue. BSA proteins are plotted in red. A: the green SOD1 is close to BSA protein(s). B: the green SOD1 is in contact with another SOD1. C: the green SOD1 moves freely in bulk water.

mediated interactions calculated with the BF model reach distances up to 25 Å(two hydration shells). Although this is not relevant for A and C basins (A states are characterized by short $\mathcal{C}_i$, and the interaction energy of SOD1s in C is $\Delta E = 0$ by definition), it strongly affects the distribution of the B states in the $\vec{q}_i^{\,\mathrm{IS}}$ diagram. We find that, at the projection of the free energy landscape on $\vec{q}_i^{\,\mathrm{IS}}$, the basin above the barrier collects both B and C states. For the case of SOD1 in FUS solution, the free energy landscape exhibits a single basin corresponding to A states, as in ΔG including the solvent. Consistently with the BSA case, we also find that configurations of different $N^{\mathrm{h}}_{\mathrm{SOD1}}$ collapse to $\Delta E \sim 0$.

Next, we focus on the kinetics of the SOD1s in the BSA solution. We are interested in analyzing how the SOD1s explore the free energy landscape, to assess whether the SOD1s rapidly shift among the three states or tend to remain for long time is the same state. First, we estimate the fraction of the total time spent by the SOD1s in A, B and C states (Fig. 6.5 c). We find that SOD1s are preferably at A or B rather than C. We argue that the SOD1s lower their interaction energy when they approach to other protein, regardless it is a BSA (A state) or another SOD1 (B state). Second, we estimate the frequency of transitions between the different states $\nu_{X\to Y}$ as the number of transitions per unit time, where X and Y stand for A, B and C (see Fig. 6.5 d). We find that the transition rates are symmetric ($\nu_{X\to Y} \sim \nu_{Y\to X}$), with the highest frequency between A and B, followed by A and C, consistent with the fact that there are no free energy barriers between these states. Transitions between B and C also happen but at a lower frequency, showing that thermal energy is large enough to let the SOD1s overcome the free energy barrier between these states. Third, we estimate the histogram of uninterrupted times of residence t_R, i e. the elapsed time in which a SOD1 remains in the same state (see Fig. 6.5 b). We find that accessible t_R span over three orders of magnitude (from a few ns up to 200 ns), showing that there is no characteristic time t_R. Based on this observation and reckoning that negative values of t_R bear no physical meaning, we consider the log-normal distribution to fit our data. The probability density function of the log-normal distribution with parameters μ and σ is

$$f(t_R) = \frac{1}{t_R \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln(t_R) - \mu)^2}{2\sigma^2} \right). \tag{6.3}$$

We fit data in the range $t_R \leq 20$ ns, as the histograms estimated from the simulations are poorly sampled for larger times (the total simulation time is 1 μs). For each state, Table 6.1 reports the parameters resulting from the fitting of the log-normal dis-

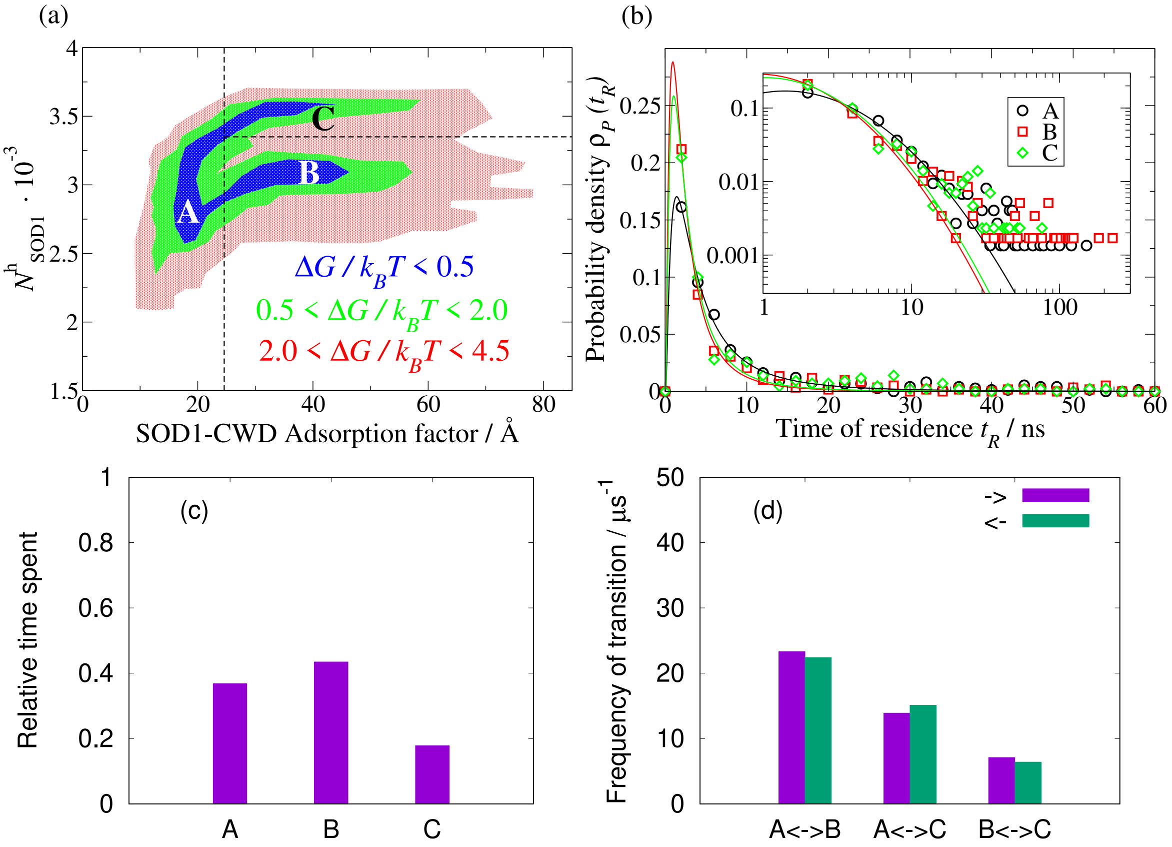


Figure 6.5: (a): Schematic hydration free energy landscape for SOD1 into the BSA solution. The blue region corresponds to $\Delta G/k_BT \lesssim 0.5$, green to $0.5 \lesssim \Delta G/k_BT \lesssim 2$, and red to $2 \lesssim \Delta G/k_BT \lesssim 4.5$. A, B, and C letters mark the associative states described in Fig. 6.4. (b): Symbols correspond to the probability density that a SOD1 resides a time t_R at the state A, B, or C. Lines represent the fit of log-normal distribution to $\rho_P(t_R)$ (see Table 6.1). (c): Time spent in each of the three states relative to the total simulation time. (d): Frequency of transitions between the three states, calculated as the number of transitions per unit time. All the results in this figure correspond to the average over ten SOD1s. Figs. I.2-I.6 in Appendix I describe individual trajectories.

	μ	σ	$\langle t_R/\mathrm{ns}\rangle$	$\mathrm{Var}(t_R/\mathrm{ns})$	α
A	1.35 ± 0.03	1.00 ± 0.03	6.4 ± 0.4	16 ± 2	-1.6 ± 0.2
B	0.83 ± 0.09	0.95 ± 0.03	3.6 ± 0.2	4 ± 2	-1.8 ± 0.3
C	0.94 ± 0.11	0.95 ± 0.05	4.0 ± 0.6	5 ± 2	-1.6 ± 0.2

Table 6.1: Parameters of log-normal distribution (μ, σ) obtained from the fit of $f(t_R)$ in Eq. 6.3 to the estimated probability density function $\rho_P(t_R)$. The fourth and fifth column report the expected value $\langle t_R\rangle = \exp\left(\mu + \sigma^2/2\right)$ and the variance $\mathrm{Var}(t_R) = (\exp\left(\sigma^2\right) - 1) \cdot (\exp\left(2\mu + \sigma^2\right))$ of t_R. The last column shows the exponent of the power law decay $\rho_P(t_R) \propto t_R^\alpha$. We fit μ, σ, and α to data in the range $2 \leq t_R/$ ns ≤ 20.

tribution, the corresponding mean and variance of t_R, and the exponent of the power law decay $\rho_P(t_R) \propto t_R^\alpha$.

The results for the kinetics of SOD1 in the BSA solution reported in this section were obtained averaging over the ten SOD1s. Hence, it describes the expected behavior of a SOD1 protein. However, a more detailed analysis shows that individual SOD1s describe different trajectories. For example, 1-st and 2-nd SOD1s spend approximately 60 % of the time in the state B, while 8-th SOD1 less than 10 %. Figs. I.2-I.6 show the same results as in Fig. 6.5, but for the individual trajectories of the SOD1s Moreover, we plot the time evolution of $\mathcal{C}_i(t)$, $N_{\mathrm{h,SOD1}_i}(t)$, and the visited states (A, B, C) in Fig. I.7.

6.4 Discussion

In Ref. [349], Samanta et al. concluded that the interaction energies of the SOD1s with FUS and BSA crowders, calculated with the implicit solvent force field OPEP, were comparable. Thus, they argued that SOD1s show no essential preference to adsorb into one or another solution. Although their result helped to rationalize in vitro experiments on the PC of SOD1 into FUS condensates in BSA and Ficoll 70 crowded solutions [349], it was unexpected. From the physicochemical point of view, we would anticipate that the different sequence-specific and crowding interactions should have an impact on the preference of the SOD1s to partition into BSA or FUS solutions.

Indeed, our results including explicit solvent show that the SOD1s behave entirely different in the two condensates. The SOD1s in the BSA solution explore three typical associative states (SOD1-BSA, SOD1-SOD1, and SOD1 *free-in-the-bulk*), while SOD1s in the FUS solution only interact with the FUS. We argue that the origin of the different behavior of the SOD1s rests on the solvent composition: in the BSA solution there is a large reservoir of bulk water, while most of the water molecules in the FUS solution hydrate FUS chains. Hence, the SOD1s in the BSA solution can escape from the crowder by cooperating with another SOD1 or spontaneously separating and moving into the bulk. Since the implicit solvent SOD1-CWD energy [349] and the water enthalpy (Table I.1) are comparable for both systems, we argue that the solvation free energy difference between both systems arises from a) the larger entropy of water and b) the role of homotypic SOD1 interactions in the BSA solution.

Regarding the kinetics of the SOD1s in the BSA solution, our analysis reveals that SOD1s undergo, on average, 90 transitions per μs between the three associative states. We also find that accessible times t_R span over at least three orders of magnitude, between 2 and 200 ns. However, shorter and longer t_R are possible, as we are limited by the selected time resolution (2 ns) and the total time of the implicit solvent simulation (1 μs). We suggest that the kinetic analysis could be biased due to our *artificial* definition of the borders separating A, B, and C regions of the free energy landscape. If this holds, high transition frequencies and short residence times indicate that (some) individual SOD1s move close to the borders.

6.5 Conclusions

We study the effects of crowding on the behavior of SOD1 sequestered in BSA and FUS highly concentrated solutions. The comparison is motivated by the different conformational properties of the crowders (BSA is globular and FUS, intrinsically disordered) and to rationalize in vitro experiments showing a decrease in the PC of SOD1s in FUS condensates when BSA is present compared to Ficoll 70 crowded solutions [349]. Our method, described in detail in Sec. 5.2, considers trajectories previously calculated with the OPEP implicit solvent model [232,349], and maps the configurations into their corresponding BF water and protein representation. We then calculate water-water and water-protein contributions to the thermodynamic observables (density, enthalpy, and response functions) along the trajectory of the system, as well as the contribution from the solvent to the Gibbs free energy.

At variance with implicit-solvent energy-calculations [349], we find that the explicit-solvent free energy landscape reveals an entirely different behavior of the SOD1s in FUS and BSA solutions. While SOD1 proteins only interact with FUS chains, in the BSA solution they cyclically fluctuate among three associative states: SOD1-BSA (A), SOD1-SOD1 (B), and SOD1 *free in the bulk* (C). Our results show that A and B are more frequent than C states, and that B and C states are separated by a free energy barrier. We argue that crowding effects, and their impact on the composition of the solvent, rationalize the behavior of the SOD1s. While most of the volume of the BSA solution is occupied by a bulk water reservoir, FUS chains are distributed homogeneously over the whole system. Therefore, most of the water molecules hydrate FUS amino acids, and the excluded volume for the SOD1s in the FUS solution increases compared to the BSA solution. We conclude that the experimentally observed preference of SOD1 proteins to partition into BSA solutions [349] arises from the larger entropy of water compared to FUS solutions.

Chapter 7

Conclusions

In this Thesis, we i) extend to bulk the FS model, ii) validate the model through comparison with experimental data around ambient conditions, iii) analyze its transferability to deep supercooled conditions, and iv) present a theoretical approach to calculate the water contribution to the free energy of large biological systems starting from configurations generated by an implicit solvent model.

According to the partition of the volume into cubic lattice, FS water molecules in each cell have six nearest neighbors with which they can form HBs. However, water molecules cannot form more than four HBs. In Chapter 2, we introduce a checkerboard partition that ensures that all the water molecules can form up to four HBs. Moreover, we design parallel-distributed MC algorithms taking advantage of the specific topology of the cubic lattice. The simulations are executed in GPUs and efficiently sample unprecedentedly large-size water systems of the order of $17,576,000$ molecules if sampled with Metropolis and $2,097,152$ with Swendsen-Wang. These systems correspond to a simulation box of 75x75x75 nm^3 and 37x37x37 nm^3, respectively.

As a prerequisite for the use of the bulk FS model in biological simulations, in Chapter 3 we calibrate the parameters so that the model fits the experimental density and response functions of water around ambient conditions. To this aim, we estimate the ratio J_σ/J between the cooperative and covalent contributions to the HB interaction energy based on ALMOEDA ab initio calculations of minimum-energy water clusters [267]. Then, fixed J_σ/J to the estimated value, we look for a set of linear rescaling functions such that the model calculations best-fit the experimental data of water around ambient conditions. The bulk FS model under the selected parametrization reproduces quantitatively the experimental equation of state of water at ambient pressure for the range $270 \leq T/\text{ K} \leq 330$. By increasing P the range of T reduces up to $290 \leq T/\text{ K} \leq 310$ for $P = 50$ MPa. This range of T and P cover the working conditions of most proteins. Hence, the bulk FS model is a remarkable candidate for simulations of biological interest with explicit solvent, since i) it equilibrates large-size water systems ii) it is quantitatively consistent with water around ambient conditions, and iii) accounts for HBs at molecular resolution, including cooperativity, at variance with spatially based coarse-grained models such as MARTINI and SIRAH [106, 107].

In Chapter 4, we study the transferability of the model to deep supercooled conditions. As shown for the monolayer case, cluster MC simulations of the FS model are suitable to equilibrate the system at extreme low-T conditions for a wide range of P (Sec. 1.4 and Refs. [123, 134]). Here, we show the presence of a LLCP at the end of a first-order LLPT between LDL and HDL phases for the bulk FS model, in accordance with atomistic models [77, 85, 86]. The estimated critical pressure P_C and critical temperature T_C exhibit a strong dependence on the size of the system, and extrapolate to $P_C = (174 \pm 14)$ MPa and $T_C = (186 \pm 2)$ K at the thermodynamic limit. Furthermore, the phase diagram of the bulk FS model around the critical conditions offers a coherent picture of water polymorphism, consistent with the experimentally observed polyamorphism [20, 23]. Above P_C, we identify three forms of water with different thermodynamic and dynamic properties: LDL and HDL phases separated by a LLPT and VHDL separated from HDL by a continuous structural change marked by the internal ordering of the degrees of freedom of the water molecules. Below P_C, the Widom line emerges from the LLCP separating LDL-like and HDL-like states, without discontinuity in the free energy, following the locus of strong maxima of C_P. The loci of maxima of K_T and smooth minima of α_P emerge from the Widom line, separating HDL-like and VHDL-like forms of water. This continuous structural change is originated by the formation of the HB network, that in turn implies an increase in density fluctuations. The two maxima of the response functions that are described for the FS monolayer are not present in bulk because they are a consequence of the lower dimensionality. At high-P, above the LLPT, for thermodynamic consistency, the bulk C_P has smooth maxima matching the high-T minima and enclosing the anomalous region for this response function.

In Chapter 5, we extend to bulk the BF model for water and protein systems. We propose a theoretical approach that considers trajectories of large biological systems previously calculated with implicit solvent models, e.g., the OPEP [232], and maps the protein configurations into their corresponding BF representation (Sec. 5.2). The aim of this method is to estimate the water contribution to the free energy of large biological systems. We successfully apply this approach to two cases of study: the conformational changes of the A1, A2, and A3 domains of the von Willebrand factor (vWf) under extreme shear rate (Chapter 5), and the analysis of the effects of BSA-crowding on SOD1 sequestration into FUS highly concentrated solutions (Chapter 6).

Thanks to the combination of solvent and protein descriptors, we characterize four typical configurations of the globular A1, A2, and A3 domains of the vWf: collapsed (C), detached-collapsed (DC), detached-extended (DE), and extended (E). Our results show that, for low shear rates $\dot{\gamma}$, C conformations are stable and CD metastable, separated by a free energy barrier of $\sim 5\ k_BT$. This barrier decreases upon increasing $\dot{\gamma}$. For intermediate-$\dot{\gamma}$, the system explores DE conformations. At high-$\dot{\gamma}$, the free energy landscape becomes almost flat and the system cyclically visits all the possible conformations. We interpret that, under these conditions, the work exerted by the shear rate is the dominating contribution to the free energy of the system.

In Chapter 6, we find that the inclusion of the solvent is necessary to unveil the different behavior of SOD1 proteins sequestered into crowded (globular) BSA and (intrinsically disordered) FUS environments. Previous calculations with an implicit-solvent force-field show that SOD1-BSA and SOD1-FUS interaction energies are comparable, contrary to the intuition that the specific protein interactions should have a larger impact on the energy [349]. Here, we find three characteristic associative states for SOD1 in the BSA solution, namely SOD1-BSA (A), SOD1-SOD1 (B), and SOD1 *free in the bulk* (C), whereas SOD1s always interact with FUS chains. The free energy landscape of SOD1 in BSA solution shows a free energy basin connecting A with B and A with C, without any barrier, and a barrier separating B and C of approximately 3 $k_B T$. Since the enthalpy of BF water is comparable in both BSA and FS solutions, we interpret that the different behavior of SOD1 in the BSA solution is a consequence of the increase in water's entropy. We argue that our results rationalize the experimental observation that the partition coefficient of SOD1 in FUS condensates decreases for FUS solvated in a BSA medium instead of Ficoll 70 [349].

Our results on biological systems reveal that we can adequately characterize their free energy landscapes by including the solvent explicitly to interpret and understand their behavior. Therefore, the proposed theoretical approach overcomes an intrinsic limitation of implicit solvent models, that coarse-grain the degrees of freedom of the solvent, and paves the way for future studies of large-scale biological systems in explicit solvent.

Appendix A

Generation and usage of random numbers

Random numbers are required for tasks such as proposing σ and η new states or deciding whether to accept or reject a new configuration. In sequential MC, one usually generates a random number only if needed. However, this strategy is not convenient in parallel algorithms, as it is desirable to avoid thread divergence. This occurs if some of the threads that are executed in parallel must generate a random number while others do not.

In our parallelization scheme, we adopt the *principle of deferred decisions* strategy. For each computation that requires a random number, we allocate in the device an array of random numbers of size $N \cdot$`N_RANDOM`. This array contains the maximum number of random numbers required for `N_RANDOM` executions of the kernels, where each kernel launches a thread per molecule. We index the arrays of random numbers according to `rnd_idx` $= r \cdot N +$ `cell`, where $0 \leq r <$ `N_RANDOM`. Then, for a given calculation that involves a random number, each thread reads its corresponding random number, taking advantage of memory coalescing. To avoid thread divergence, the computation is done even if unnecessary. An example of this is whenever $\Delta H \leq 0$ in a Metropolis update.

We generate random numbers in the GPU using the `cuRAND` library. In particular, we generate N uncorrelated sequences of random numbers and overwrite the arrays of random numbers every `N_RANDOM` steps, when they have been spent (see Fig. 2.2).

Appendix B

Scaling of the algorithms

We study the scaling laws (time cost versus N) of the four implementations considered in this work: Metropolis and SW algorithms, both in sequential and in parallel. We find that all scale linearly in a limited range of N, as explained in the following.

A sequential Metropolis step is defined as $6N$ trials of updating a σ_{ij} variable chosen at random. Although the time spent in a single update depends on the sign of ΔH in Eq. 2.7, the probability of finding it positive or negative does not depend on N. Then, we expect that the time cost of the algorithm scales linearly with N, as shown in Fig. B.1. However, we find that for large $N \gtrsim 8.8 \cdot 10^5$, the time cost of a MC step deviates from linearity. For even larger $N \sim 2.1 \cdot 10^6$, it becomes twice the time expected. We rationalize this behavior on the extra time cost of loading the large σ and η arrays in the RAM (16 GB). For the largest $N = 128^3 \sim 2.1 \cdot 10^6$ considered here, the size of the σ and η arrays is `sizeof`(σ) =`sizeof`(η) $= 6 \cdot N \cdot 1$ byte $= 12$ GB, if the arrays are allocated as `uint8_t`. Thus, the total size of the arrays (24 GB) exceeds by far the storage capacity of the RAM. Under such circumstances, the RAM loads each time the portion of the arrays needed for calculation. The extra computational cost is spent in loading portions of arrays.

Regarding the parallel Metropolis algorithm, we measure the time spent in the generation of arrays of random numbers (Appendix A), the execution of the kernels, and the time spent in copying the σ and η arrays to the host. In parallel algorithms, data access by the device is the most time-consuming task. We find that the performance scales linearly in an intermediate range of N: $3.3 \cdot 10^4 \sim 32^3 \leq N \leq 128^3 \sim 2.1 \cdot 10^6$ (Fig. B.2). This can be explained as the time spent in data accessing scales linearly with N if the GPU resources are not saturated (large N), nor under-exploited (small N). For small $N \leq 8000$, the computational resources of the GPU are not optimized. We obtain that in this range, the time cost of a MC step remains approximately constant. For large $N \geq 216^3 \sim 10^7$, `N_RANDOM` must be reduced to fill in the GPU global memory. The extra time cost owes to both the increasing number of executions of the kernels for generating random numbers and the time cost of memory transactions.

Next, we find that two power laws $\mathrm{SF}^{\mathrm{MET}} \propto N^\gamma$ fit the speedup factor of the

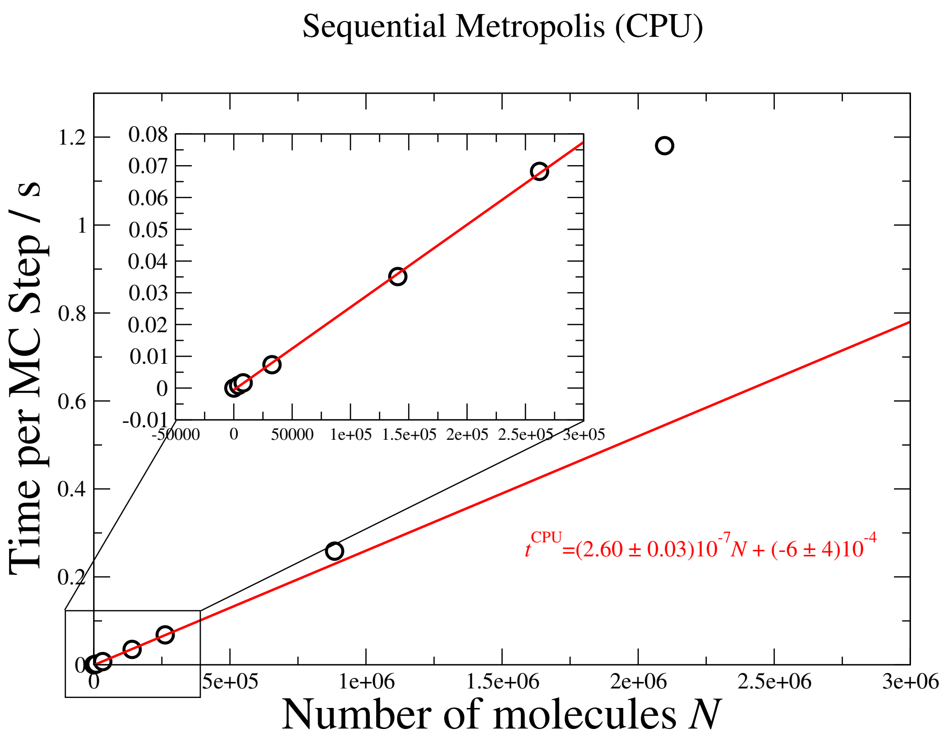

Sequential Metropolis (CPU)

Figure B.1: Performance of the sequential implementation of the Metropolis algorithm (CPU). The red line is the linear fit of the data in circles ($N \leq 262,144$), using least squares method. For larger N (diamonds) the performance deviates from linear scaling, as explained in the text. Error bars are smaller than the size of the symbols.

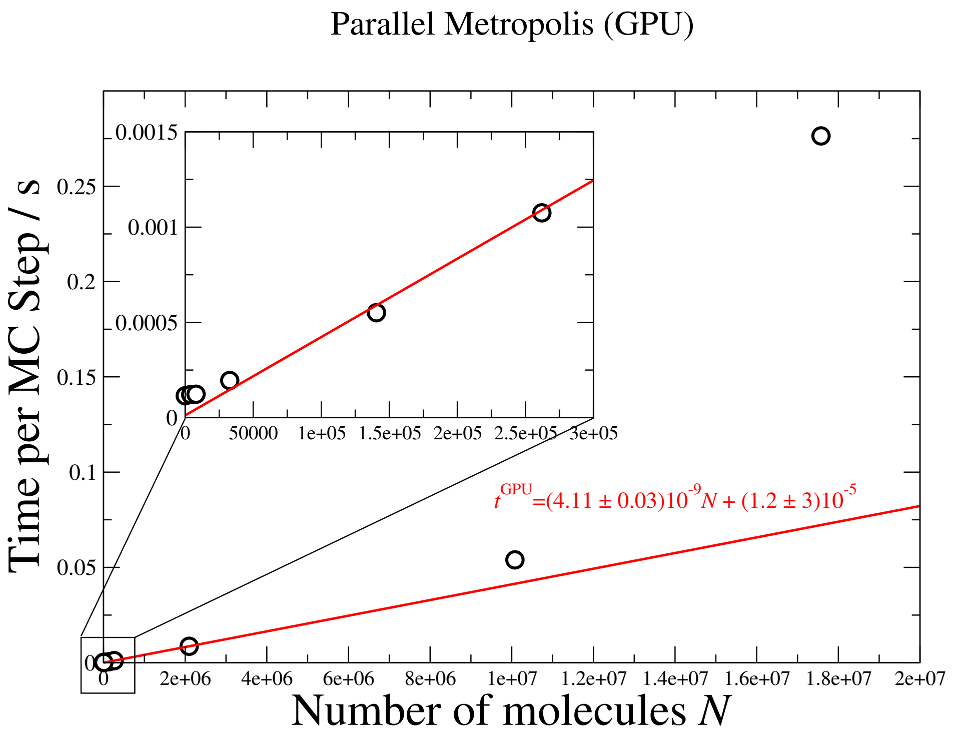

Parallel Metropolis (GPU)

Figure B.2: Performance of the parallel implementation of the Metropolis algorithm (GPU). We use the same symbol and color code as in Fig. B.1. Error bars are smaller than the size of the symbols.

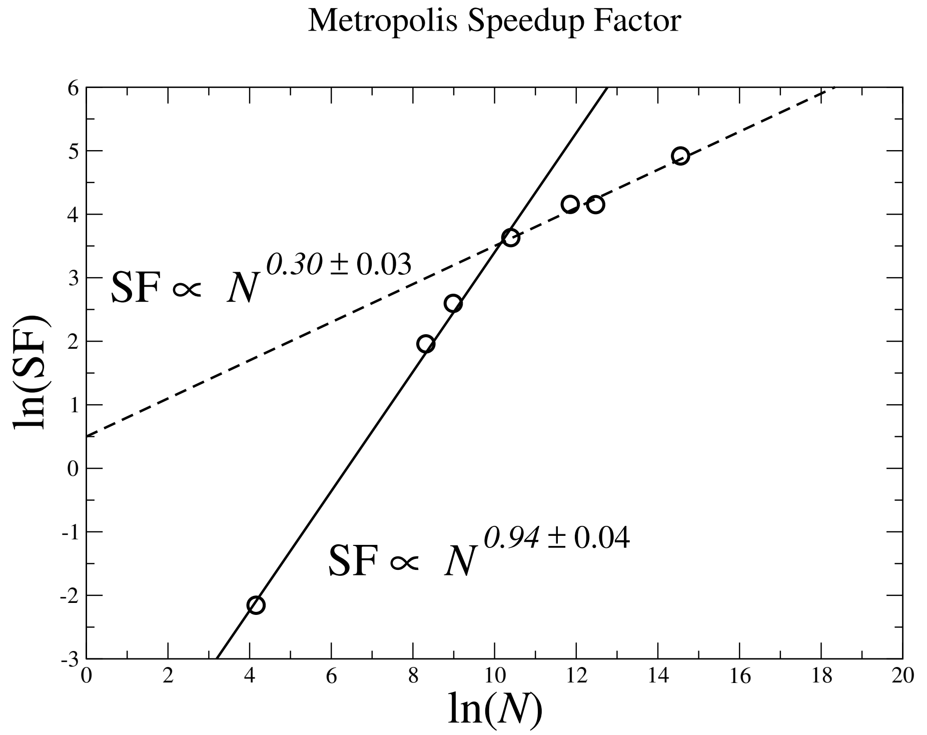


Figure B.3: Scaling of the Metropolis speedup factor respect to N. We find a crossover from $N^{0.94}$ to $N^{0.3}$ at $N = 32,768$. Error bars are smaller than the size of the symbols.

Metropolis algorithm (Table 2.1), see Fig. B.3. For small $N \leq 32^3 \sim 3.3 \cdot 10^4$, $\gamma = 0.94 \pm 0.04$ is close to 1, i.e., SF increases approximately linearly with N. The reason is that in this range of N, t^{GPU} is almost constant, while t^{CPU} increases linearly. For $3.3{\cdot}10^4 \leq N \leq 128^3 \sim 2.1{\cdot}10^6$, SF follows the power law with exponent $\gamma = 0.30{\pm}0.03$. In this range t^{GPU}, increases linearly with N, but t^{CPU} increases with $\gamma > 1$ due to the extra time cost of loading large arrays in the RAM.

We compare the performance of the SW algorithms in two temperatures at $P = 0.0(4\epsilon/v_0)$, corresponding to percolation $T = 0.05(4\epsilon/k_B)$ and non-percolation $T = 0.06(4\epsilon/k_B)$. At both T, the performance of the sequential SW algorithm scales linearly with N for $N \leq 64^3 \sim 2.6 \cdot 10^5$ (Fig. B.4). For larger $N = 128^3 \sim 2.1 \cdot 10^6$, the computational time increases above linearity due to the extra time cost of managing large arrays, as we have described in the sequential Metropolis case. We find that the sequential SW algorithm is faster under non-percolating than under percolating conditions. This is explained as the Hoshen-Kopelman algorithm is applied iteratively until it converges. The total time of the update depends on the size of the largest cluster, as it requires more iterations to converge.

Concerning the parallel SW algorithm, the time spent in a MC update increases linearly for $N \leq 64^3 \sim 2.6 \cdot 10^5$, although data at small values of N are noisy. As in the parallel Metropolis algorithm, we attribute this to the fact that the usage of GPU resources is not optimized. At large N, we find extra time cost compared with linearity, as in the other three algorithms. Again, we attribute this to the limited resources

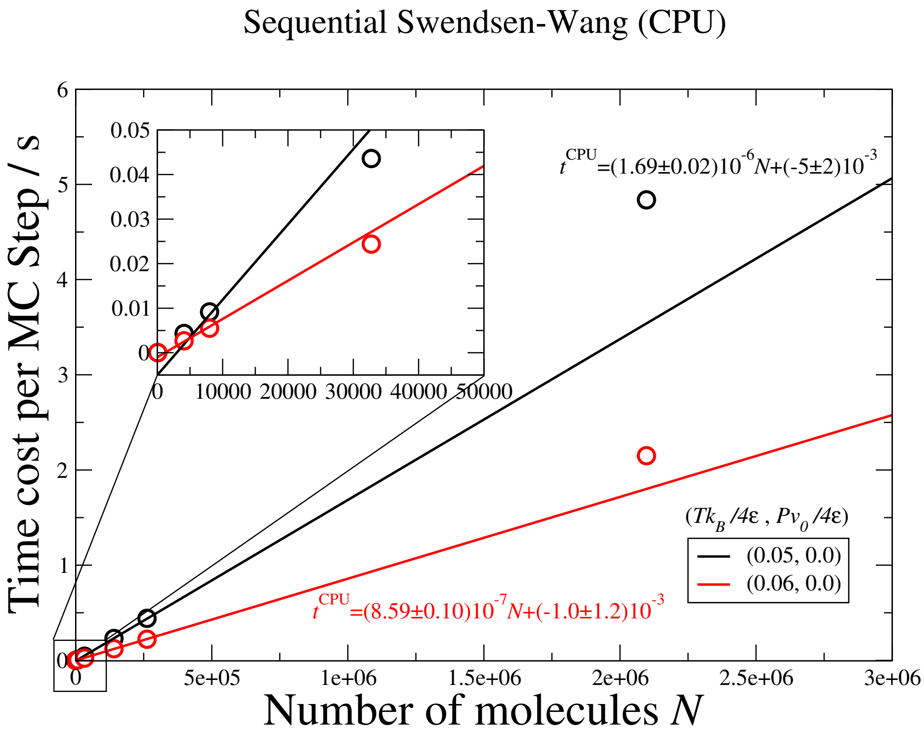


Figure B.4: Time cost of a sequential SW update. We use the same color and symbol code as in Fig. B.1

to store large arrays. We note that, contrary to the sequential SW algorithm, the parallel SW update is faster at percolating conditions than without percolation. This is surprising since we would expect that the total time cost of the update is governed by the largest-size cluster, as in the sequential algorithm. A possible explanation is that the analysis function converges rapidly regardless of the size of the cluster, making not so relevant the size of the largest cluster. Then, the different time cost between percolation and non-percolation must arise from less efficient memory readings of the label array in small clusters by the scanning and labeling functions.

Finally, we find that the speedup factor of the SW algorithm shows a crossover at $N = 32,768$, as in Metropolis. For low-N, the scaling is almost linear without percolation, $\gamma = 0.92 \pm 0.02$, and consistent with linearity under percolation, $\gamma = 0.99 \pm 0.02$. For larger N, the exponents are $\gamma = 0.182 \pm 0.015$ and $\gamma = 0.10 \pm 0.02$ with and without percolation, respectively.

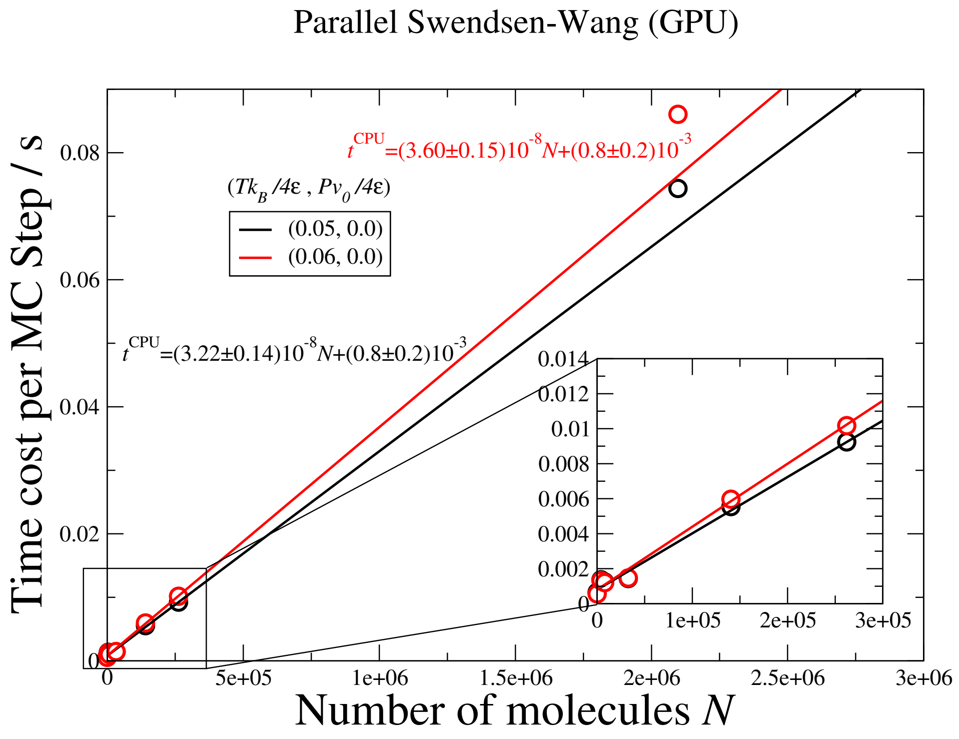


Figure B.5: Time cost of a parallel SW update. We use the same color and symbol code as in Fig. B.1

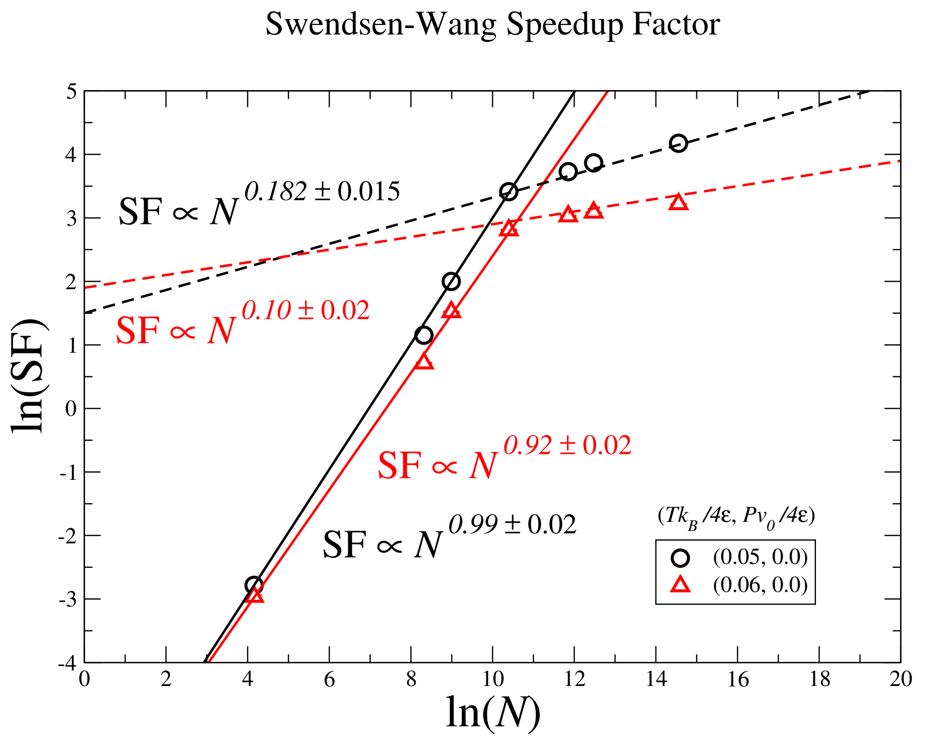


Figure B.6: Scaling of the SW speedup respect to N. Error bars are smaller than the size of the symbols.

Appendix C

Rescale of the phase diagram under different parameters

In this appendix, we show the rescaling of the phase diagram under different sets of parameters (Sec. 3.2.2). In Fig. C.1, we plot the phase diagram of five sets of parameters and compare them with experimental data. The slope and intercept of the rescaling functions are reported in Table C.1.

We find that the set of parameters with $J/4\epsilon = 0.2$, $J_\sigma/4\epsilon = 0.03$, and $v_{\mathrm{HB}}/v_0 = 0.2$ fails to reproduce the density anomaly at high-P. In particular, the model predicts that the density does not display a maximum along isobars above $P = 46$ MPa (dashed line in Fig. C.1). However, the experimental TMD line reaches up to 140 MPa. Regarding the other sets of parameters, we find better agreement with experiments, with the best compromise between the loci of the TMD and K_T^{min} obtained with $J/4\epsilon = 0.5$, $J_\sigma/4\epsilon = 0.08$, and sets$v_{\mathrm{HB}} = 0.6$.

FS Parameters		Temperature		Pressure	
$(J/4\epsilon, J_\sigma/4\epsilon)$	v_{HB}/v_0	a/K	b/K	a/MPa	b/MPa
$(0.5, 0.08)$	0.5	151.79	175.45	390.13	-200.22
	0.6	140.57	185.47	469.46	-217.89
	0.7	212.50	117.78	405.47	-100.64
$(0.2, 0.03)$	0.2	141.67	241.73	113.30	-67.01
	0.5	204.00	188.76	450.80	-22.68

Table C.1: Coefficients of the linear rescaling functions for T and P, where a is the slope and b is the intercept (Eq. 3.5).

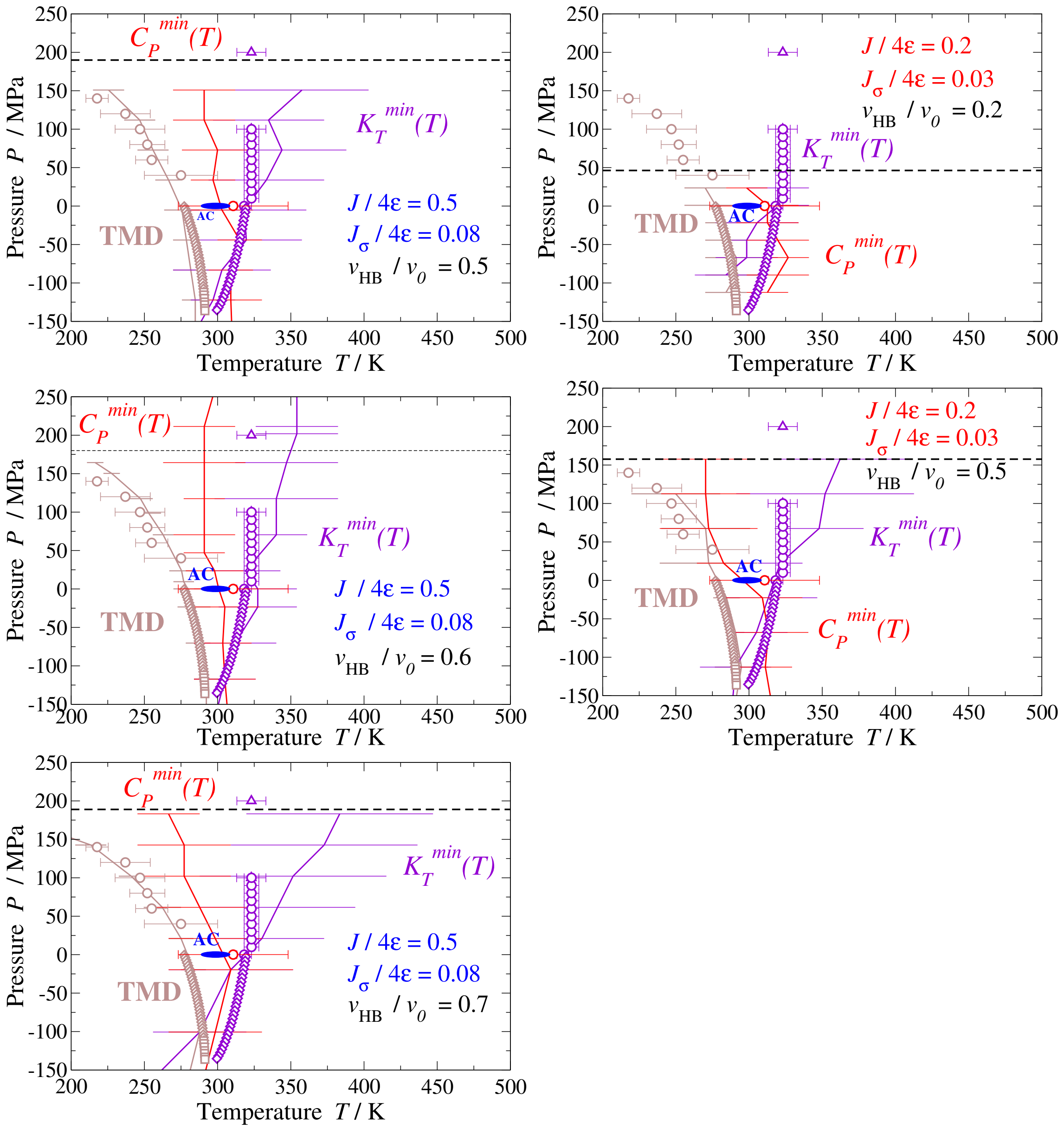


Figure C.1: Phase diagram of the FS model under different sets of parameters (see each panel), with rescaling functions from Table C.1. Lines and symbols are defined as in Fig. 3.6, omitting the locus of maxima of K_T. The dashed line indicates the P-threshold for the density anomaly. In the FS model, this threshold corresponds to $J^{\rm eff}(\hat{P}) = 0 \rightarrow \hat{P} = J/v_{\rm HB}$.

Appendix D

The fluctuation-dissipation theorem

At equilibrium, the response functions correspond to thermodynamic fluctuations. Therefore, we calculate C_P, α_P, and K_T along isobars both by numerical derivative and the fluctuation-dissipation theorem:

$$C_P \equiv \left(\frac{\partial H}{\partial T}\right)_P \equiv \frac{\langle H\rangle^2 - \langle H^2\rangle}{k_B T^2} \tag{D.1}$$

$$\alpha_P \equiv \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_P \equiv \frac{\langle VH\rangle - \langle V\rangle\langle H\rangle}{k_B T^2 \langle V\rangle} \tag{D.2}$$

$$K_T \equiv -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \equiv \frac{\langle V\rangle^2 - \langle V^2\rangle}{k_B T V}. \tag{D.3}$$

In this way, we guarantee that the system is equilibrated by verifying the validity of the fluctuation-dissipation theorem, i.e. if both estimates hold within the error bar. We find a good agreement between both estimates, for all the response functions and for the entire range of simulated T and P (Fig. D.1).

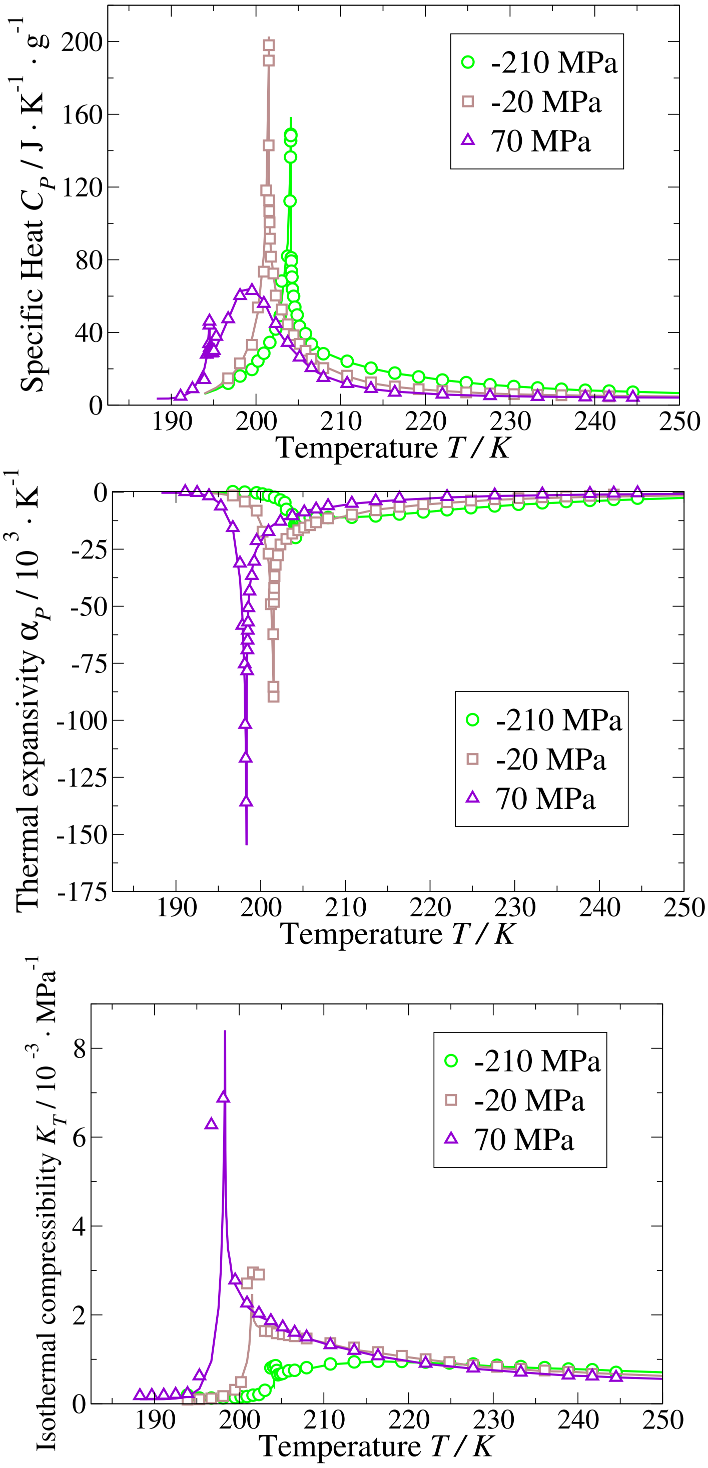


Figure D.1: Response functions C_P (top), α_P (center), and K_T (bottom), calculated as numerical derivative (symbols) and fluctuations (lines). Each color corresponds to a different P, as shown in the legend. The results correspond to the system of $N = 32,768$ molecules.

Appendix E

The histogram reweighting method

Histogram reweighting is an enhanced sampling method that predicts the partition function of the system Z, provided a set of independent estimates Z_i. This method can be employed to calculate the unbiased distribution of a system from a set of simulations performed with biased potentials, as in umbrella sampling [303, 317]. Here, we apply the histogram reweighting method in the NPT-ensemble to predict the probability distribution $Q(T, P, e, \rho)$ given $Q(T_i, P_i, e, \rho)$ calculated in simulations at temperature T_i and P_i sufficiently close to T and P [86, 134]. Then, the integral of $Q(T, P, e, \rho)$ along the direction $M = \rho + s \cdot e$ gives $Q(T, P, M)$ that is compared to the probability distribution of the LLCP of the Ising 3D class of universality, $Q_3(M)$.

The partition function of the system is

$$Z(\beta, P) \equiv \sum_E \sum_V \Omega(E, V) \exp\big(-\beta(E + PV)\big) \equiv \exp\big(-\beta G(\beta, P)\big)\,, \tag{E.1}$$

where $\Omega(E, V)$ is the density of states, $\beta = 1/k_B T$, and $G(\beta, P)$ is the Gibbs free energy. From simulations, we estimate the histogram $H(E, V)$ as the frequency of configurations having energy E and volume V along the trajectory. For the run i at thermodynamic conditions (β_i, P_i), the histogram is

$$H_i(\beta_i, P_i, E, V) = \frac{N_i(\beta_i, P_i, E, V)}{N_{c,i}} \sim \frac{\Omega(E, V) \exp(-\beta_i(E + P_i V))}{\exp(-\beta_i G(\beta_i, P_i))}, \tag{E.2}$$

where $N_{c,i}$ is the total number of sampled configurations during the run i. H_i becomes the correct probability distribution (right-hand term) in the limit of infinite time simulations $N_{c,i} \to \infty$. Hence, given a set of independent simulations at different thermodynamic conditions, we have a set of approximate estimates of the correct density of states $\Omega(E, V)$. The new histogram at β and P is given by a combination of the simulated histograms

$$H(\beta, P, E, V) = \frac{\sum_i H_i(\beta_i, P_i, E, V) \exp(-\beta(E + PV))}{\sum_i N_{c,i} \exp(-\beta_i[E + PV - G(\beta_i, P_i)])}, \tag{E.3}$$

where the Gibbs free energy is unknown a priori. To overcome this problem, the Eqs. E.3 and E.4 are solved iteratively

$$\exp(-\beta_i G(\beta_i, P_i)) = \sum_E \sum_V H(\beta_i, P_i, E, V). \tag{E.4}$$

Appendix F

Estimation of the order parameter *x(m)* for the LLPT

We addressed two major issues during the search of the order parameter $M \equiv \rho + s \cdot e$ and the location of the LLCP. First, we find that only negative values of the mixing parameter s were acceptable to symmetrize the joint probability distribution $Q(e, \rho)$ at variance with previous results in the literature [85,86,134,303]. Second, the probability distribution of the order parameter $m \equiv B(M - M_C)$ systematically fail to fit the Ising 3D critical point $Q_3(m)$ for the entire range of simulated N and P. More precisely, we find that the tails of the distributions are overrated and the peaks underrated. We show the estimated $Q(m)$ for $N = 4,096$ in Fig. F.1 (a, red squares).

We relate the origin of both issues to the shape of the probability distribution $Q(e, \rho)$ in the bulk FS model. To estimate s, we rotate $Q(e, \rho)$ with the 2D Euclidean rotation matrix, so that $\mathcal{M} = e\sin(\alpha) + \rho\cos(\alpha)$. Clearly, $M \equiv \mathcal{M}/\cos(\alpha)$ and $s \equiv \tan(\alpha)$. The only possible α that symmetrize $Q(m)$ lie within $\pi/2 < \alpha < \pi$ and $3\pi/2 < \alpha < 2\pi$, leading to negative s. We find that by tuning the parameters, the slope of $Q(e, \rho)$ in the energy-density plane changes, making $s > 0$.

Regarding the second issue, we argue that what is relevant to assess the presence of the LLCP is the existence of an order parameter whose fluctuations are critical, rather than its functional form $M = M(\rho, e)$. Therefore, we tested alternative order parameters $X(m) = X(\rho, e)$ such that $Q(X)$ best fit to $Q_3(m)$.

We start the quest for $X = X(m)$ noting that for $m = 0$ and $m = \pm m_0 = \pm 1.53$, $Q(m)$ coincides with $Q_3(m)$. Thus, we look for $x = x(m)$ invariant with respect to $m = 0$ and $\pm m_0$. Next, we note that $x(m)$ should slightly shrink the tails ($|m| > m_0$) and stretch the peaks ($|m| < m_0$) of the distribution $Q(m)$.

Under these considerations, we find that the function

$$X(m) = \alpha_0 \lambda^2 \sin\left(\frac{m}{\lambda}\right) + \alpha_1 m \tag{F.1}$$

fits well to our purposes. By letting $\alpha_1 = 1$ and $\lambda = m_0/\pi$, the set of $m = k \cdot m_0$, with $k \in \mathbb{Z}$, are fixed points. Moreover, the sinus alternatively changes its sign shifting m

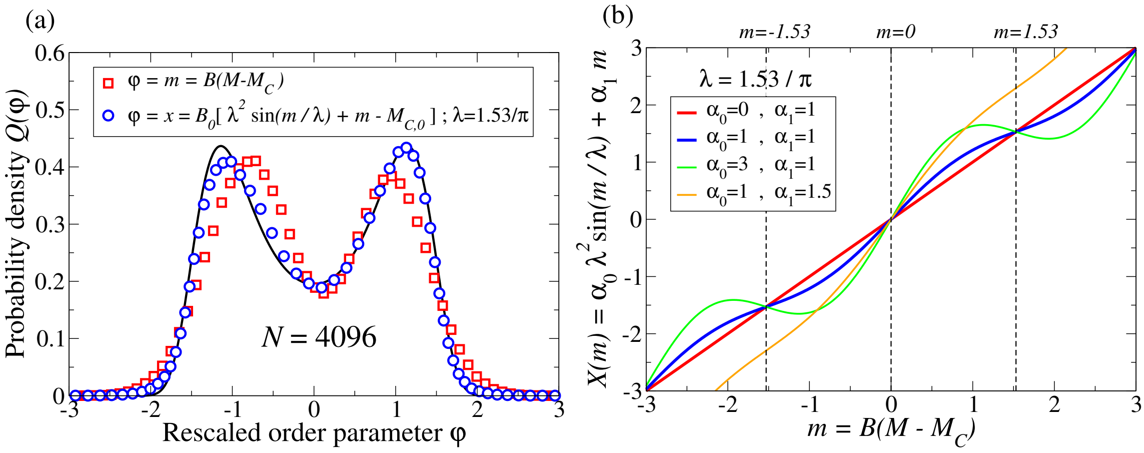


Figure F.1: (a): Best fit to $Q_3(\varphi)$ (black line) for $N = 4,096$. Red squares correspond to $Q(m)$, following the usual definition of the order parameter $M \equiv \rho + s \cdot e$. Blue circles correspond to $Q(x)$, the order parameter considered in this work. (b): Correction of the order parameter $X = X(m)$, assuming $\lambda = 1.53/\pi$ and different α_0 and α_1. The red line corresponds to the usual order parameter, and the blue line to the corrected x, following the same color code as in the panel (a). Dashed lines highlight the fixed points upon $X = X(m)$, provided $\alpha_1 = 1$.

as convenient. The parameter α_0 modulates the correction term, as shown in Fig. F.1 (b).

We rescale the order order parameter X

$$x \equiv B_0(X - M_{C,0}) \equiv B_0 \left[\alpha_0 \lambda^2 \sin\left(m/\lambda\right) + \alpha_1 m - M_{C,0}\right], \tag{F.2}$$

where B_0 and $M_{C,0}$ are calculated so that $Q(x)$ has mean zero and unit variance. By setting $\lambda = 1.53/\pi$, $\alpha_0 = 1$, and $\alpha_1 = 1$, $Q(x)$ best fits to $Q_3(x)$ (Fig. F.1 a, blue circles). Hence, the selected order parameter is

$$x = B_0 \left[\lambda^2 \sin\left(m/\lambda\right) + m - M_{C,0}\right] = B_0 \lambda^2 \sin\left(m/\lambda\right) + B_0\left(m - M_{C,0}\right), \tag{F.3}$$

that is a monotonic function of m (Fig. F.1 b) that deviates from m less than 30% for $\pm m_0/2$ (where the sinus is ± 1).

We test whether the normalization factor $B \cdot B_0$ follows or not the power law estimated in the NPT-ensemble finite size scale theory for fluids, $B \propto N^{\beta/d\nu}$ [83], finding a negative result (Fig. F.2). We argue that there are at least two possible reasons that explain this result: i) that we are considering small size systems compared to the thermodynamic limit and ii) that the scaling law is only valid for the order parameter m, but not for x. We argue that ii) is the most relevant, since $B_0 = 0.99 \pm 0.02$.

We understand that the negative sign of s and the inclusion of the correcting term are particular features of the bulk FS model arising from the calculated fluctuations of ρ and e. However, in this appendix, we have shown that these problems can be

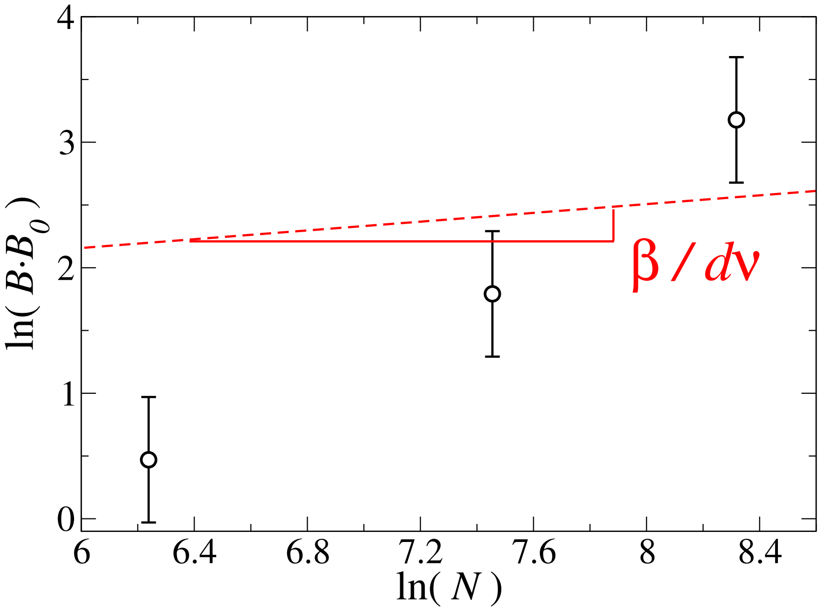


Figure F.2: The normalization factor for X, $B \cdot B_0$ apparently does not follow the scaling law $B \propto N^{\beta/d\nu}$, estimated for m. We consider $d = 3$, and the critical exponents $\beta = 0.33$ and $\nu = 0.63$ of the Ising 3D universality class.

addressed without affecting the conclusion of this work: the presence of the LLCP in the bulk FS model.

Appendix G

Free energy landscape in terms of BF water enthalpy

At variance with the bulk FS parameters, that are set to reproduce the experimental equation of state of water around ambient conditions (Chapter 3), the calibration of the BF parameters at the hydration shell is still missing. In a first approximation, we consider two possible sets of parameters for the HBs at the hydration shell, labeled 'A' and 'B'. On the other hand, as the configurations of the domains are preserved, we assume that the fluctuations of the water-protein contact interaction term ($\mathcal{H}_{\mathrm{R,w}}$ in Eq. 5.1) are not relevant. For this reason, we consider a single set of S_i^{w} parameters (Table G.1). We describe the sets A and B, and their motivation, in the following.

Set A adapts the parameters considered in a previous work for BF protein in a water monolayer [6] to the bulk case. The hydrophobic interface strengthens the water-water hydrogen bonding in the first hydration layer [204–208], therefore we assume $J^{\mathrm{PHO,1}} > J$ and $J_\sigma^{\mathrm{PHO,1}} > J_\sigma$. Moreover, the local density of water at the hydrophobic interface increases upon pressurization [205, 210–212]. The model reproduces this behavior assuming that the local volume change associated with the formation of a HB at the first hydration shell depends linearly on P as $v_{\mathrm{HB}}^{\mathrm{PHO,1}} \equiv (1 - k \cdot P) v_{\mathrm{HB,0}}^{\mathrm{PHO,1}}$, where $v_{\mathrm{HB,0}}^{\mathrm{PHO,1}} = v_{\mathrm{HB}}$ is the volume change at $P(4\epsilon)/v_0 = 0$, and $k > 0$. We note that the variation of $v_{\mathrm{HB}}^{\mathrm{PHO,1}}$ on P is not relevant for our purposes, as we do not change the thermodynamic conditions in this work. However, we apply this formula to estimate the corresponding $v_{\mathrm{HB}}^{\mathrm{PHO,1}}$ at $P = 1$ atm. Following Ref. [6], set A assumes $J^{\mathrm{PHO,1}} = 4J$, $J_\sigma^{\mathrm{PHO,1}} = 4J_\sigma$, and $v_{\mathrm{HB}}^{\mathrm{PHO,1}}(1\ \mathrm{atm}) = 4v_{\mathrm{HB}}$. Regarding the hydrophilic hydration shell, the set A assumes that the HBs are not affected by the interface, so $J^{\mathrm{PHI,1}} = J$, $J_\sigma^{\mathrm{PHI,1}} = J$, and $v_{\mathrm{HB}}^{\mathrm{PHI,1}} = v_{\mathrm{HB}}$ (see Table G.2).

According to the set A, changes in the enthalpy of water are mainly due to the changes in the shape of hydrophobic and mixed interfaces. Naively, we may expect that the volume of the hydrophobic interface is much smaller than that of the hydrophilic interface. If this were the case, the set A may not be sufficiently sensitive to changes in the shape of the hydration shell. For the case of the A1, A2 and A3 domains, we find

Amino acid	vdW radius [Å] Ref. [357]	$S^{\text{w}}/4\epsilon$
ARG	3.3	-0.1
LYS	3.2	-0.1
ASP	2.8	-0.5
GLU	3.0	-0.1
ASN	2.8	-0.5
GLN	3.0	-0.1
CYS	2.7	-0.5
MET	3.1	0
HIS	3.0	-0.1
SER	2.6	-0.5
THR	2.8	-0.5
VAL	2.9	0
LEU	3.1	0
ILE	3.1	0
PHE	3.2	0
TYR	3.2	-0.1
TRP	3.4	0
GLY	2.3	-0.5
ALA	2.5	0
PRO	2.8	0

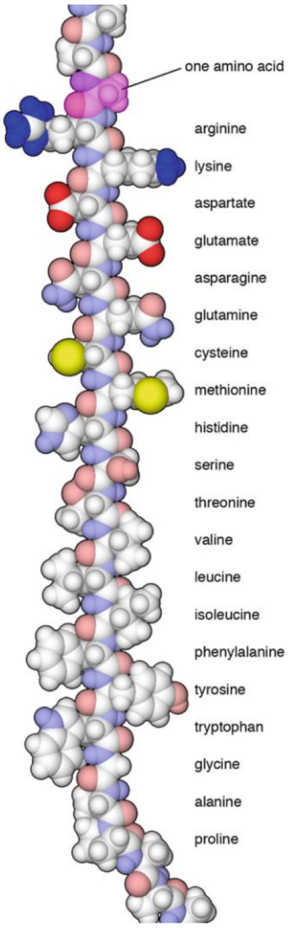

one amino acid
arginine
lysine
aspartate
glutamate
asparagine
glutamine
cysteine
methionine
histidine
serine
threonine
valine
leucine
isoleucine
phenylalanine
tyrosine
tryptophan
glycine
alanine
proline

Table G.1: Table of amino acid-water interactions. In the second column, we report the van der Waals (vdW) radius of each amino acid. In the third column, we report the interaction energy per contact with BF water. If the residue is hydrophobic, then $S^{\text{w}} = 0$. Otherwise, $S^{\text{w}}/4\epsilon = -0.5$ if the radius of the amino acid is $r_i < r_0 = 2.9$ Å and $S^{\text{w}}/4\epsilon = -0.1$ if $r_0 < r_i < 2r_0 = 5.8$ Å, with $\epsilon \equiv 5.5$ kJ/mol [1,138]. Right: Schematic representation of the amino acids, following the same ordering as in the first column. Figure adapted from Ref. [392]: Copyright (2009) Springer-Verlag New York.

BF WATER PARAMETERS - SET A				
	BULK	PHO	PHI	MIX
$J/4\epsilon$	0.5	2	0.5	1.25
$J_\sigma/4\epsilon$	0.08	0.32	0.08	0.2
$v_{\mathrm{HB},0}/v_0$	0.6	4	0.6	2.3
$k\cdot(4\epsilon/v_0)$	0	0.889	0	0.773
$J^{\mathrm{eff}}(P=1\ \mathrm{atm})/4\epsilon$	0.23	0.92	0.23	0.575
$v_{\mathrm{HB}}(P=1\ \mathrm{atm})/v_0$	0.6	2.4	0.6	1.5

Table G.2: Set of parameters A. The values reported for BULK have been parametrized to reproduce the experimental density and enthalpy of water, and fluctuations, around ambient conditions (Chapter 3). The parameters for PHOB, PHIL and MIX follow Ref. [6], and are equal for the three hydration layers. The last two rows show the resulting enthalpy and volume change due to HB formation at ambient pressure: $J^{\mathrm{eff}} \equiv J - Pv_{\mathrm{HB}}$ and $v_{\mathrm{HB}} \equiv (1 - k\cdot P)v_{\mathrm{HB},0}$ respectively, with $P = 0.45(4\epsilon/v_0) = 0.1$ MPa.

that 62% of the exposed surface is hydrophilic, while 38% is hydrophobic. This large extent of hydrophobic exposed surface is consistent with recent findings, based on the BF model, showing that globular proteins at ambient conditions only have $\approx 70\%$ of their hydrated surface that is hydrophilic [6]. However, it may be an artifact of the mapping algorithm, as it maps any backbone into hydrophobic alanine residues (Sec. 5.2.3). In particular, 15% of the total protein exposed surface correspond to hydrophobic side chains, while 23% correspond to either alanine residues or backbones.

To explore how the results are affected by considering HBs at the hydrophilic interface different from those at the bulk, we propose the set B of parameters. The set B assumes that the enthalpy change due to HB formation at the PHO interface is 1.8 times stronger than that of the bulk, while that of HBs at the PHI interface is 1.5 times stronger (see Table G.3).

Finally, for both sets A and B we let $J^{\mathrm{MIX},1} \equiv (J^{\mathrm{PHO},1} + J^{\mathrm{PHI},1})/2$, $J_\sigma^{\mathrm{MIX},1} \equiv (J_\sigma^{\mathrm{PHO},1} + J_\sigma^{\mathrm{PHI},1})/2$, and $v_{\mathrm{HB}}^{\mathrm{MIX},1} \equiv (v_{\mathrm{HB}}^{\mathrm{PHO},1} + v_{\mathrm{HB}}^{\mathrm{PHI},1})/2$. Moreover, the parameters for the second and third layers of the hydration shell are equal to those at the first layer.

We reproduce the results obtained with N^{h} (Sec. 5.3). Here, we consider the BF solvation enthalpy as observable:

$$\Delta H^{\mathrm{Sol}} \equiv \left(H^{\mathrm{b}} + H^{\mathrm{h}} + \mathcal{H}_{\mathrm{R,w}}\right) - H_{\mathrm{E}}, \tag{G.1}$$

that corresponds to the enthalpy difference between the total BF enthalpy of water and the reference level of E conformations. For the set A $H_E = -8.57$ kJ/mol, while for the set B, $H_E = -8.51$ kJ/mol, corresponding to the average total BF enthalpy at E conformations. ΔH^{Sol} includes all the enthalpy terms arising from water, thus excluding intra- and inter-protein interactions, and the work exerted by the hydrodynamic flow. The time evolution of ΔH^{Sol} is plotted in Figs. G.1 (set A) and G.2 (set

BF WATER PARAMETERS - SET B				
	BULK	PHOB	PHIL	MIX
$J/4\epsilon$	0.5	0.56	0.5	0.53
$J_\sigma/4\epsilon$	0.08			
$v_{\mathrm{HB},0}/v_0$	0.6			
$k \cdot (4\epsilon/v_0)$	0	1		
$J^{\mathrm{eff}}(P = 1\ \mathrm{atm})/4\epsilon$	0.23	0.4115	0.3515	0.3815
$v_{\mathrm{HB}}(P = 1\ \mathrm{atm})/v_0$	0.6	0.33		

Table G.3: Same as Table G.2, but for the set of parameters B.

B). For both cases we identify three levels of enthalpy (dashed lines) corresponding to C, D and E conformations, as for N^{h}. The changes of ΔH^{Sol} during the OPEP trajectory are dominated by H^{h}, while fluctuations of the other terms are not relevant. We find that $H_C > H_D > H_E$, consistent with the variation of N^{h}. If N^{h} is large (E conformations), then there is a higher number of HBs at the hydration shell, and the enthalpy decreases.

Qualitative comparison of the time evolution of ΔE (Fig. 5.5) and ΔH^{Sol} (Figs. G.1, G.2) shows a competing effect between water and protein contributions to the total enthalpy. While enthalpic HB interactions tend to separate the three domains (the minimum ΔH^{Sol} is found at E conformations), inter-domain energy minimizes when the beads of different domains approach to their equilibrium distances. As the domains tend to collapse due to thermal fluctuations (for low shear rate), we may conclude that the contribution from water-water HBs is weaker than residue-residue interactions. However, this observation neglects the role of entropic contributions to the Gibbs free energy that governs the system.

Free energy calculations $\Delta G(R_G, \Delta H^{\mathrm{Sol}})$ are in agreement with the results obtained considering N^{h} (see Figs. G.3 and G.4). We find that both sets of parameters succeed in separating the protein conformations into distinct free energy basins. Hence, the calculated free energy barriers considering either N^{h} or ΔH^{Sol} are comparable.

SET OF PARAMETERS A

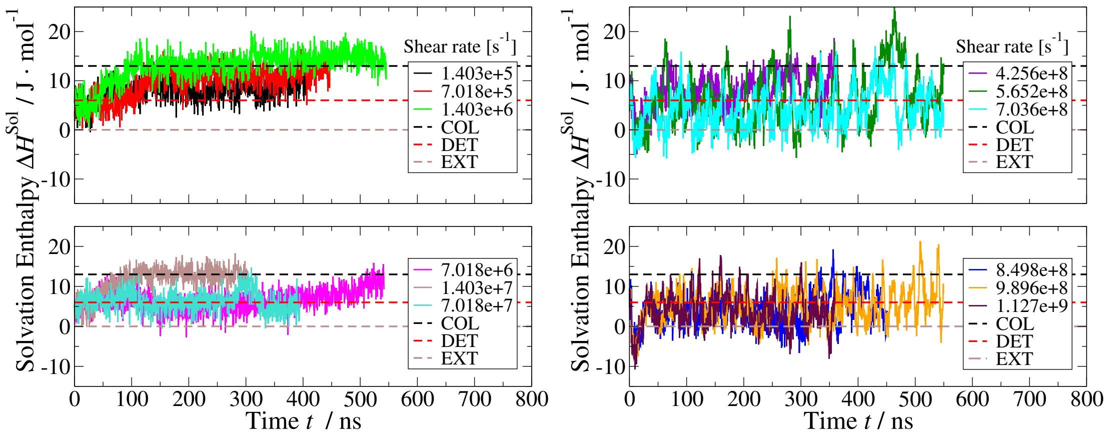


Figure G.1: Time evolution of the solvation enthalpy ΔH^{Sol} along the trajectories obtained from OPEP simulations, calculated with the set of parameters A. The shear rates are shown in the legend. Dashed lines are a guide for the eye to identify the characteristic enthalpy levels for (from top to bottom) collapsed, detached and extended conformations.

SET OF PARAMETERS B

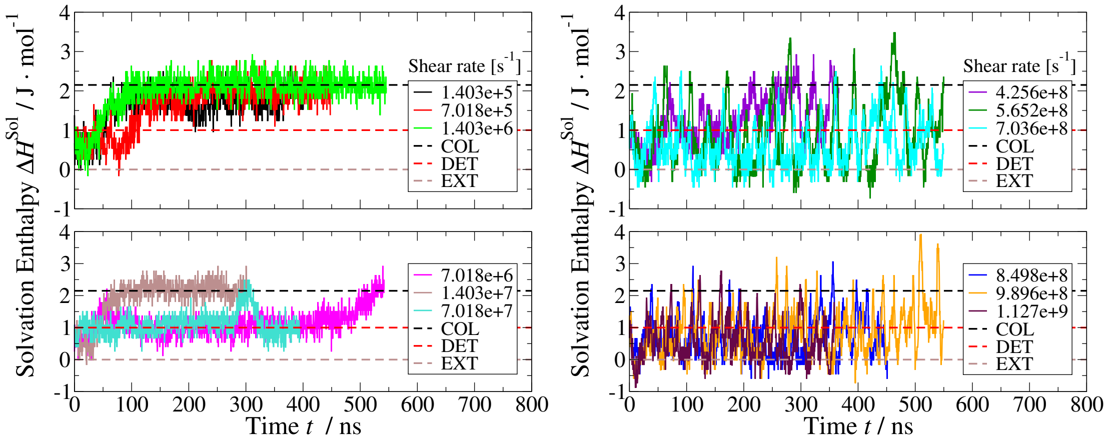


Figure G.2: Same as in Fig. G.1 but ΔH^{Sol} is calculated with the set of parameters B. Note that the y-axis scale here is one order of magnitude smaller than in Fig. G.1.

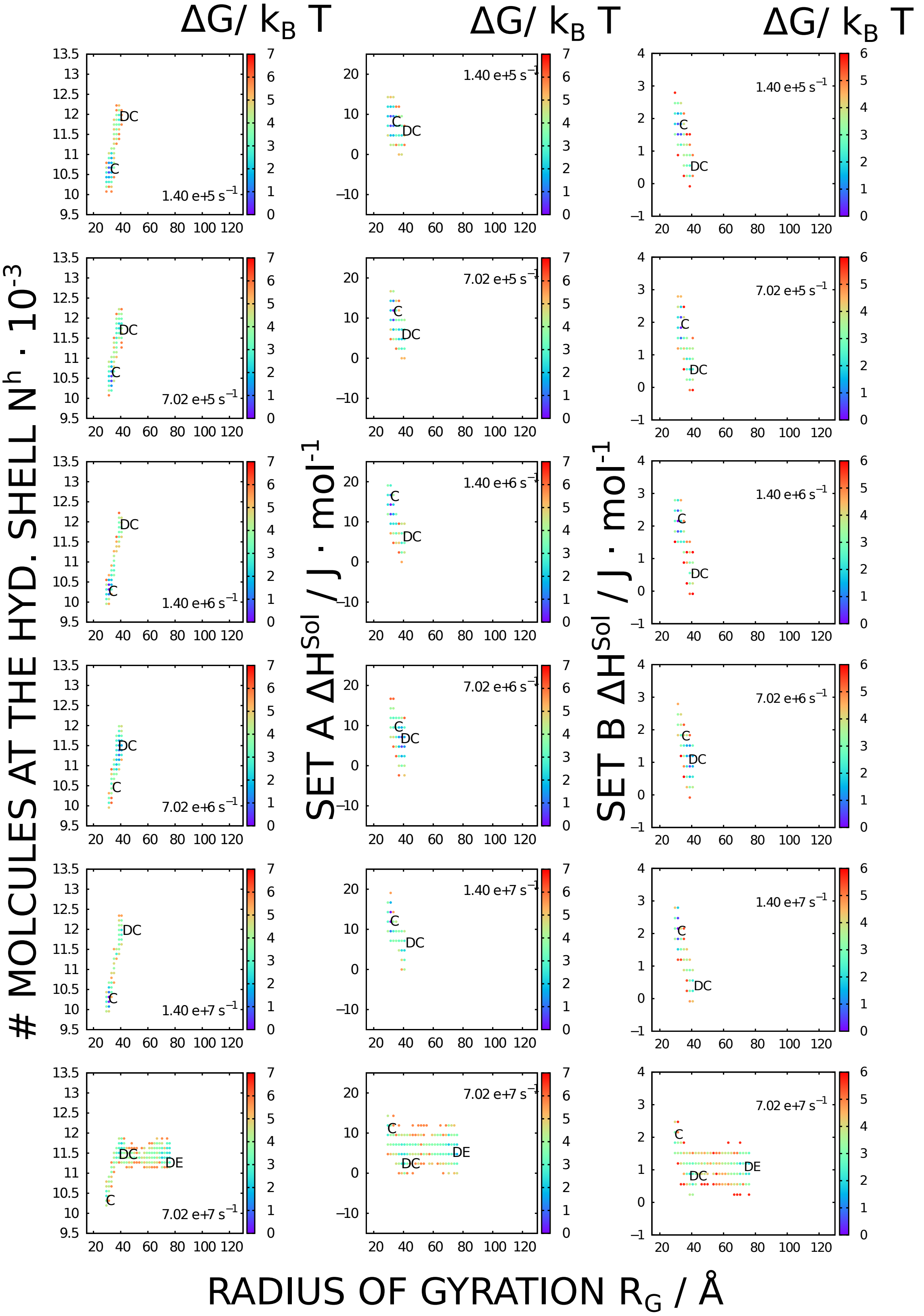


Figure G.3: From left to right: $\Delta G(R_G, N^{\mathrm{h}})$, $\Delta G(R_G, \Delta H^{\mathrm{Sol}})$ calculated with the set of parameters A, and $\Delta G(R_G, \Delta H^{\mathrm{Sol}})$ calculated with the set of parameters B. From top to bottom: shear rates $\dot{\gamma} < 10^8$ s^{-1}.

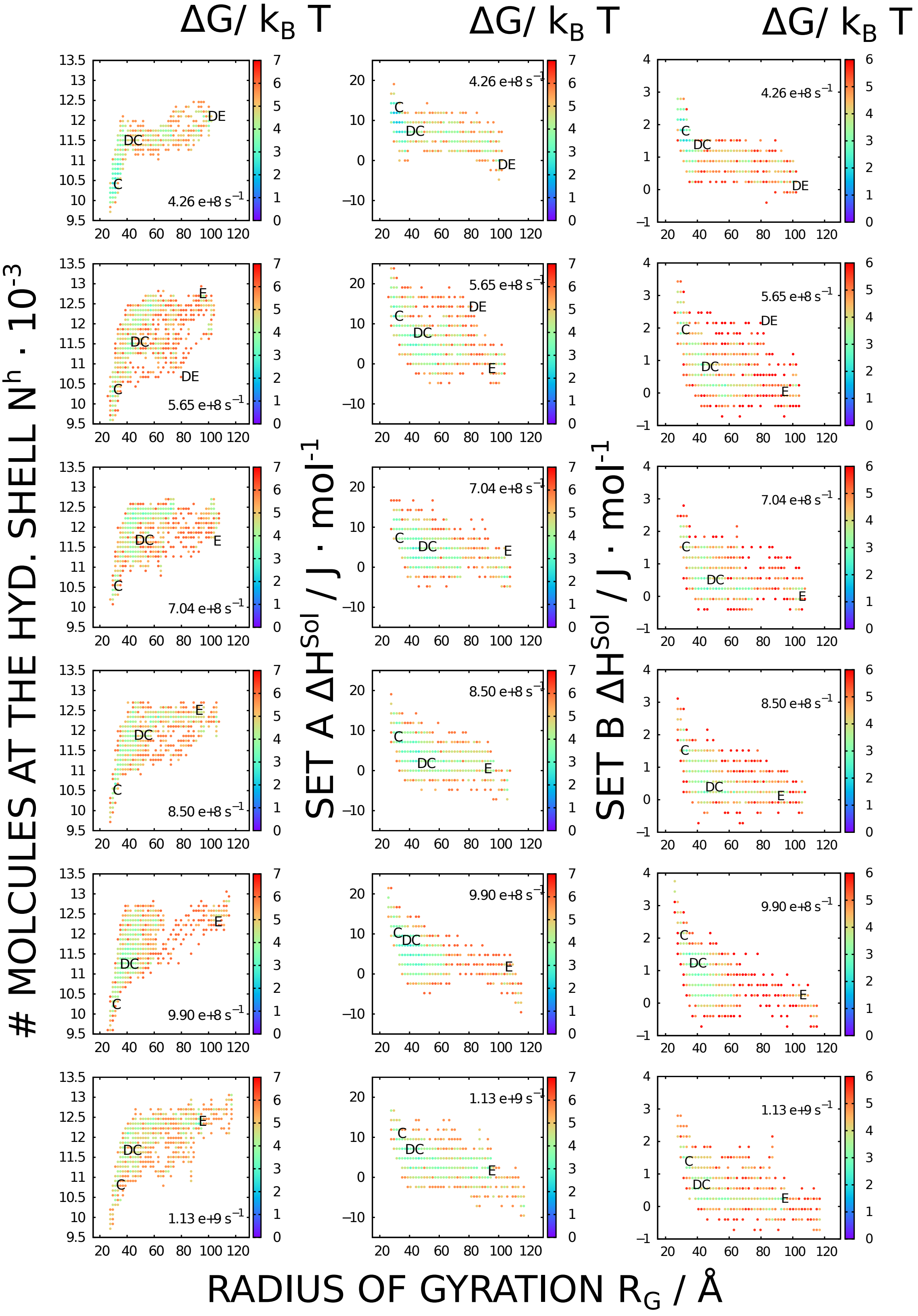


Figure G.4: Same as Fig. G.3 but for shear rates $\dot{\gamma} > 10^8$ s^{-1}.

Appendix H

Supporting Information for Chapter 5

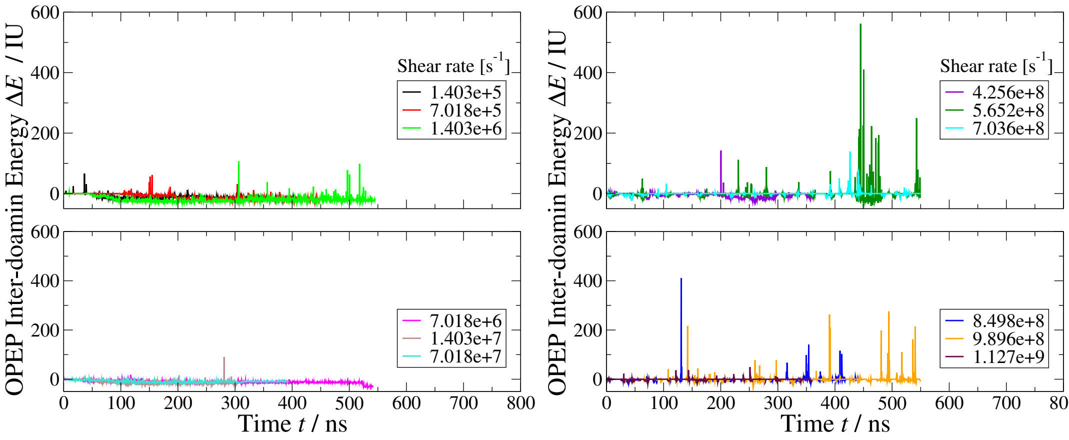


Figure H.1: Same as in Fig. 5.5 but showing the entire range of ΔE. The spikes are more frequent and reach higher values of ΔE as the shear rate increases, supporting our hypothesis that they are a result of the interaction between the beads and the hydrodynamic flow.

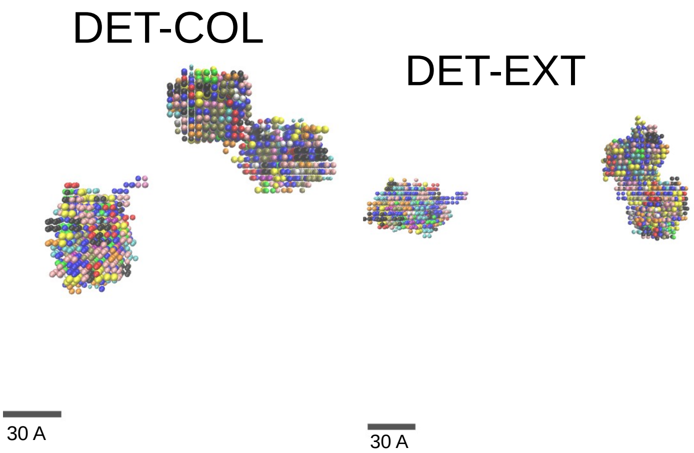


Figure H.2: Example of detached-collapsed (DET-COL or DC) and detached-extended (DET-EXT or DE) conformations of A1, A2 and A3 domains of the vWf. They are indistinguishable from both hydration and energetic points of view, but they differ in the radius of gyration. For the configurations in this figure, R_G = 46 Å and 73 Å, respectively.

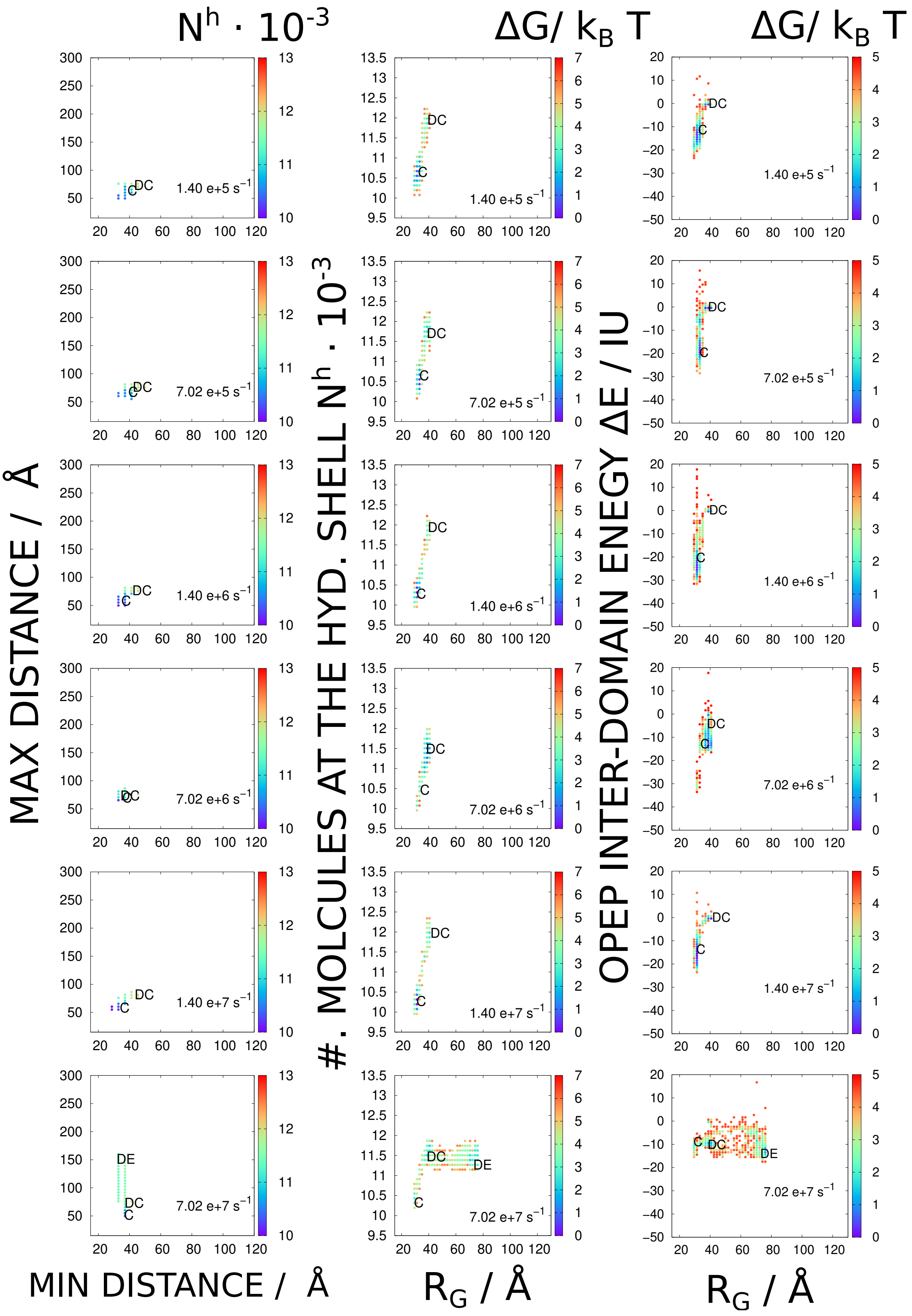


Figure H.3: From left to right: as in Fig. 5.6. From top to bottom: shear rates $\dot{\gamma} < 10^8$ s$^-1$.

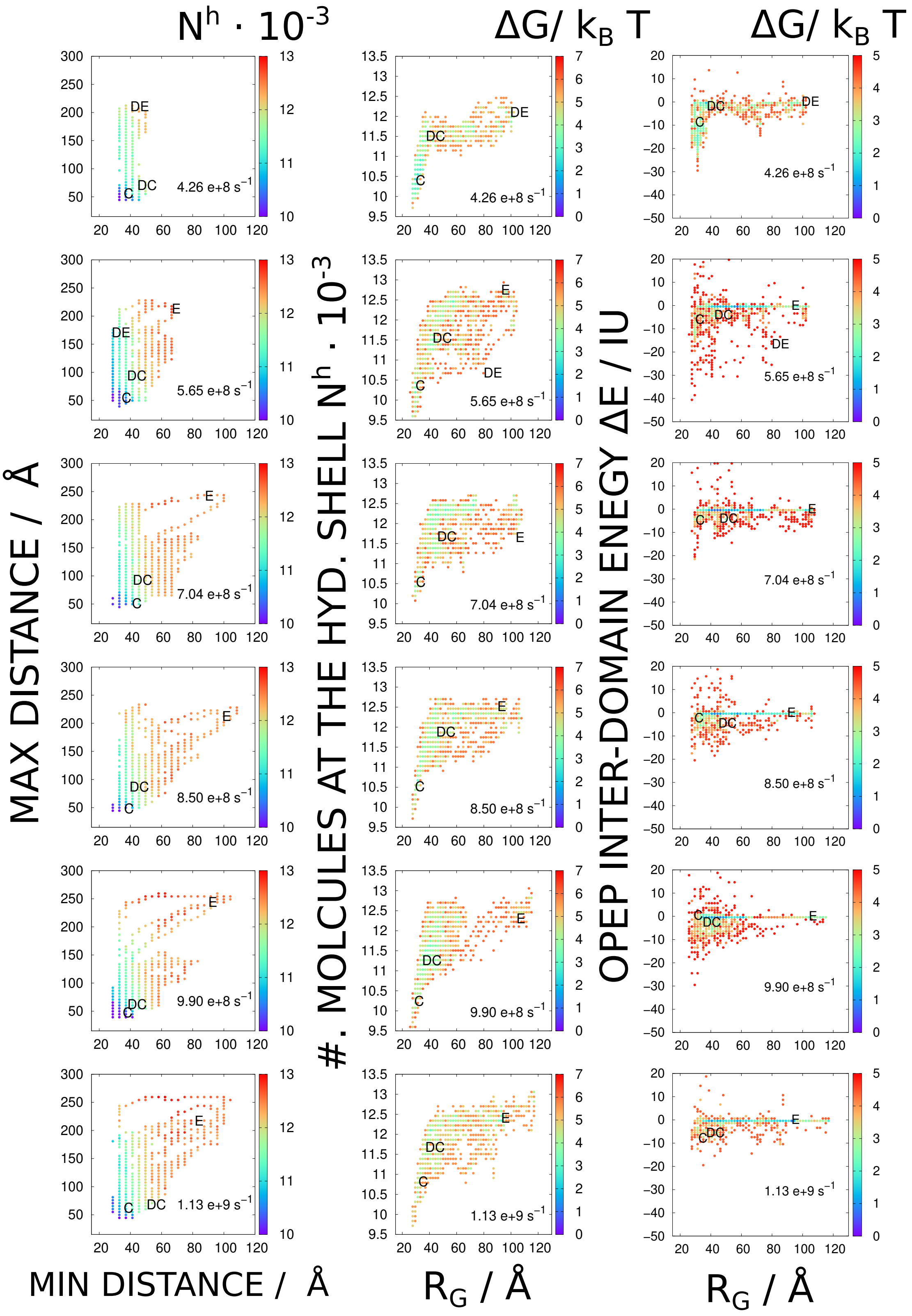


Figure H.4: From left to right: as in Fig. 5.6. From top to bottom: shear rates $\dot{\gamma} > 10^8$ s^{-1}.

	C→DC		DC→C	
Shear rate $\dot{\gamma}$·s	$\Delta G/k_BT$	F/ pN	$\Delta G/k_BT$	F/ pN
$1.4 \cdot 10^5$	4.5 ± 0.5	8.3 ± 3.6	2.0 ± 0.5	3.7 ± 2.1
$7.0 \cdot 10^5$	4.0 ± 0.5	9.8 ± 4.6	1.3 ± 0.8	3.2 ± 2.9
$1.4 \cdot 10^6$	4.5 ± 0.5	8.2 ± 3.4	0.5 ± 0.5	0.9 ± 1.1
$7.0 \cdot 10^6$	2.3 ± 0.5	2.9 ± 1.1	2.0 ± 0.5	2.6 ± 1.1
$1.4 \cdot 10^7$	3.3 ± 0.8	7.5 ± 3.1	2.0 ± 0.5	4.6 ± 2.0
$7.0 \cdot 10^7$	1.5 ± 0.3	1.7 ± 0.6	2.0 ± 0.5	2.3 ± 1.0

Table H.1: Thermodynamic force F (Eq. 5.5) calculated for each shear rate $\dot{\gamma} < \dot{\gamma}_C$ and in both directions: C→DC (hemostatic activation) and DC→C (deactivation).

Appendix I

Supporting Information for Chapter 6

ENTHALPY [kJ/mol] - SET OF PARAMETERS A			
	SOD1-BSA	SOD1-FUS	Δ [FUS-BSA]
Total: $H \equiv H^{\text{bulk}} + H^{\text{hyd}}$	-14.0248(9)	-14.6546(14)	-0.630(2)
$H^{\text{hyd}} \equiv H^{\text{hyd}}_{\text{SOD1}} + H^{\text{hyd}}_{\text{CWD}} + H^{\text{hyd}}_{\text{mix}}$	-1.922(2)	-3.933(2)	-2.011(4)
$H^{\text{hyd}}_{\text{SOD1}} \equiv H^{\text{w,w}}_{\text{SOD1}} + E^{R,w}_{\text{SOD1}}$	-0.2816(8)	-0.1958(9)	+0.084(2)
$H^{\text{hyd}}_{\text{CWD}} \equiv H^{\text{w,w}}_{\text{CWD}} + E^{R,w}_{\text{CWD}}$	-1.6298(14)	-3.689(3)	-2.059(4)
$H^{\text{hyd}}_{\text{mix}} \equiv H^{\text{w,w}}_{\text{mix}}$	-0.0107(2)	-0.0478(2)	-0.0371(4)

Table I.1: BF water enthalpy calculations expressed in kJ/mol. First line: total enthalpy of water, including isotropic van der Waals, water-water HBs and water-residue interactions. Second line: enthalpy of water within the hydration shell, including water-water HBs and water-residue interactions. Third-to-fifth lines: enthalpy of water within the hydration shell separated into SOD1, crowder and mixed contributions. The fourth column corresponds to the difference between the third and the second column. Results calculated considering the set of parameters A in Appendix G.

ENTHALPY [kJ/mol] - SET OF PARAMETERS B			
	SOD1-BSA	SOD1-FUS	Δ [FUS-BSA]
Total: $H \equiv H^{\text{bulk}} + H^{\text{hyd}}$	-13.1789(2)	-14.0982(3)	-0.9193(5)
$H^{\text{hyd}} \equiv H^{\text{hyd}}_{\text{SOD1}} + H^{\text{hyd}}_{\text{CWD}} + H^{\text{hyd}}_{\text{mix}}$	-1.3133(10)	-4.105(2)	-2.792(3)
$H^{\text{hyd}}_{\text{SOD1}} \equiv H^{\text{w,w}}_{\text{SOD1}} + E^{R,w}_{\text{SOD1}}$	-0.1675(4)	-0.1150(5)	+0.0525(9)
$H^{\text{hyd}}_{\text{CWD}} \equiv H^{\text{w,w}}_{\text{CWD}} + E^{R,w}_{\text{CWD}}$	-1.140(9)	-3.962(2)	-2.822(11)
$H^{\text{hyd}}_{\text{mix}} \equiv H^{\text{w,w}}_{\text{mix}}$	-0.00562(8)	-0.02766(12)	-0.0220(2)

Table I.2: Same as Table I.1 but considering the set of parameters B in Appendix G.

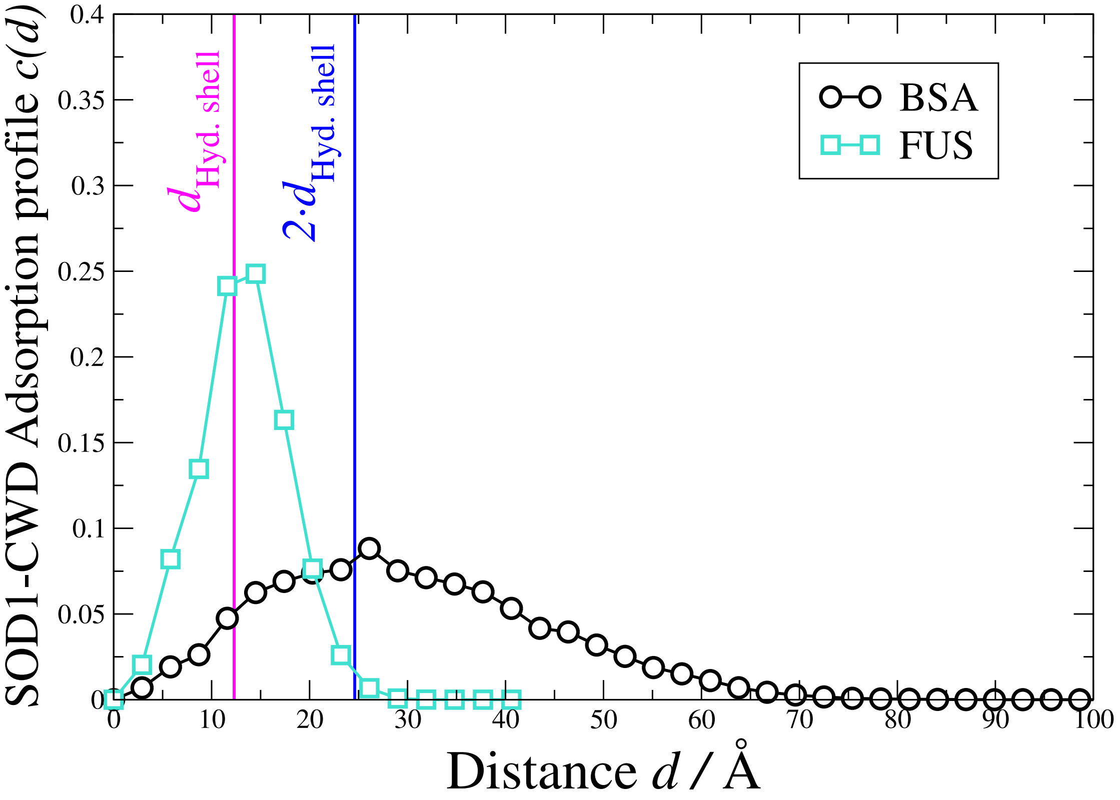


Figure I.1: Adsorption profiles $c(d)$ of SOD1 in BSA (circles) and SOD1 in FUS (squares) solutions, as defined in the text. Magenta and blue lines indicate the distances of one and two hydration shells, respectively.

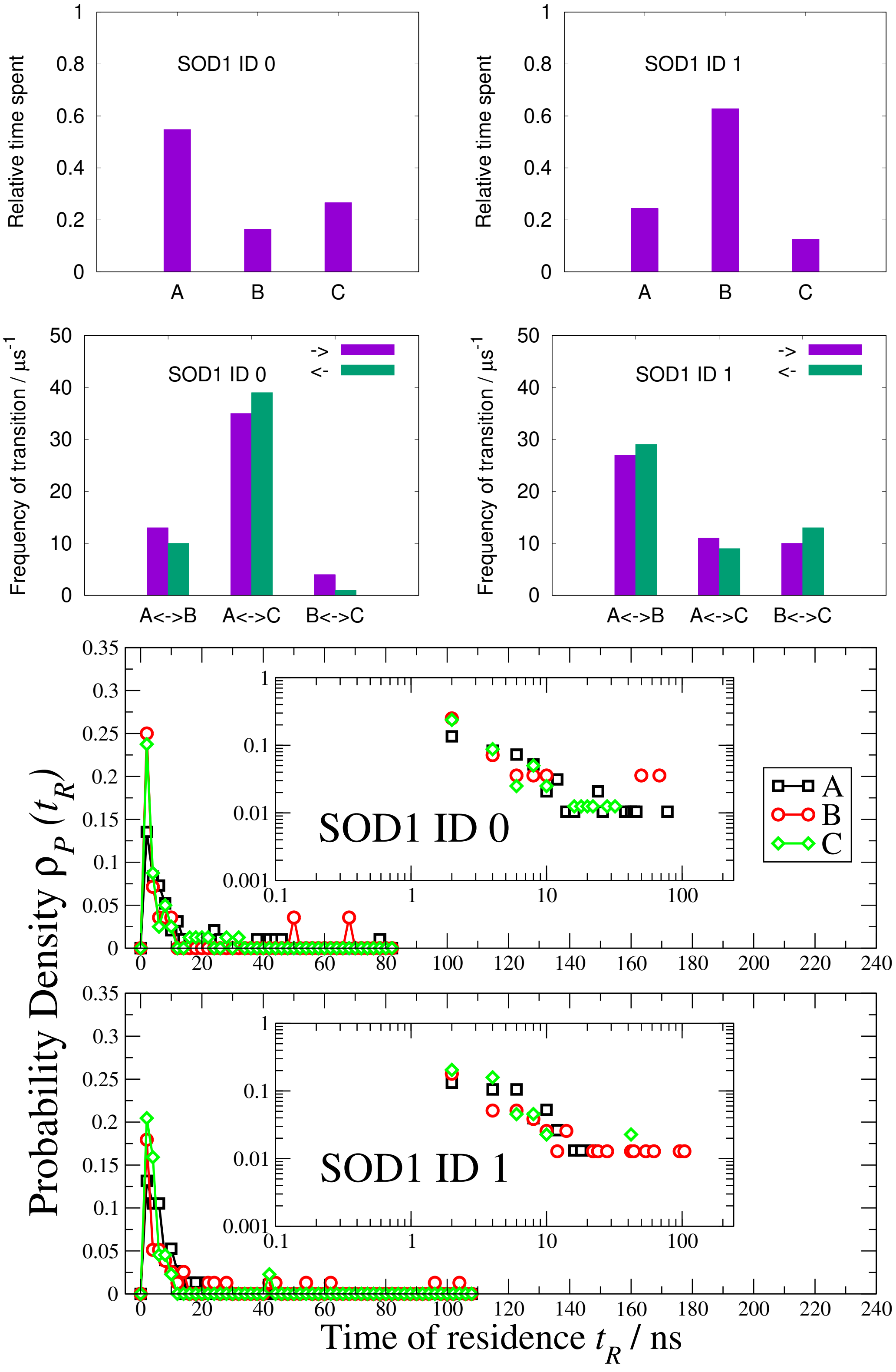


Figure I.2: Kinetic analysis of individual SOD1s. This figure refers to SOD1-ID 0 and 1, as labeled in each panel. Top: Time spent in each state, relative to the total simulation time. Center: Frequency of transition $\nu_{X \to Y}$ defined as the number of transitions per unit time. Bottom: Probability density of the times of residence for each configuration A, B or C. Lines are guides to the eyes connecting the symbols.

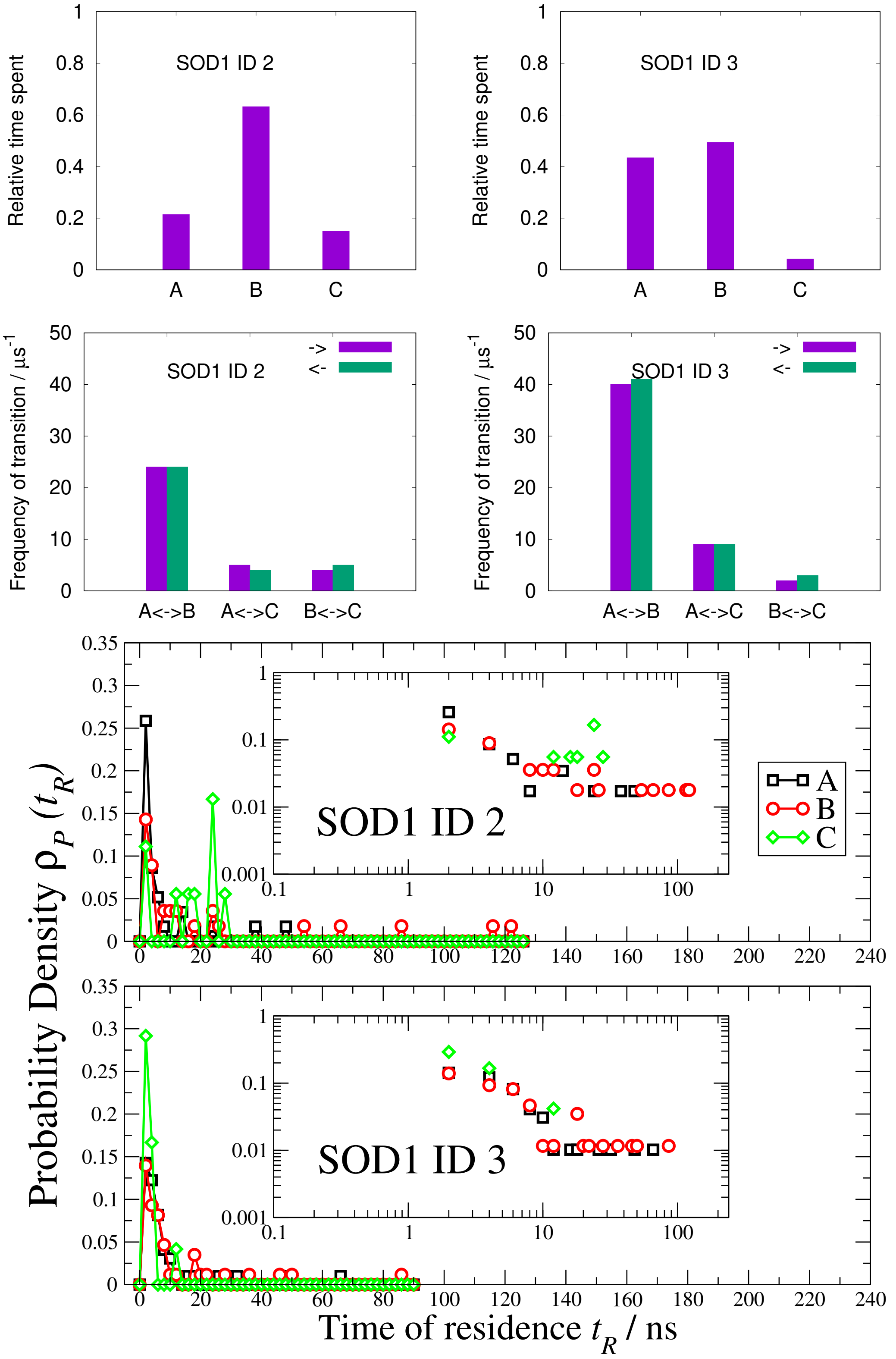


Figure I.3: Same as in Fig. I.2, but for SOD1s ID 2 and 3.

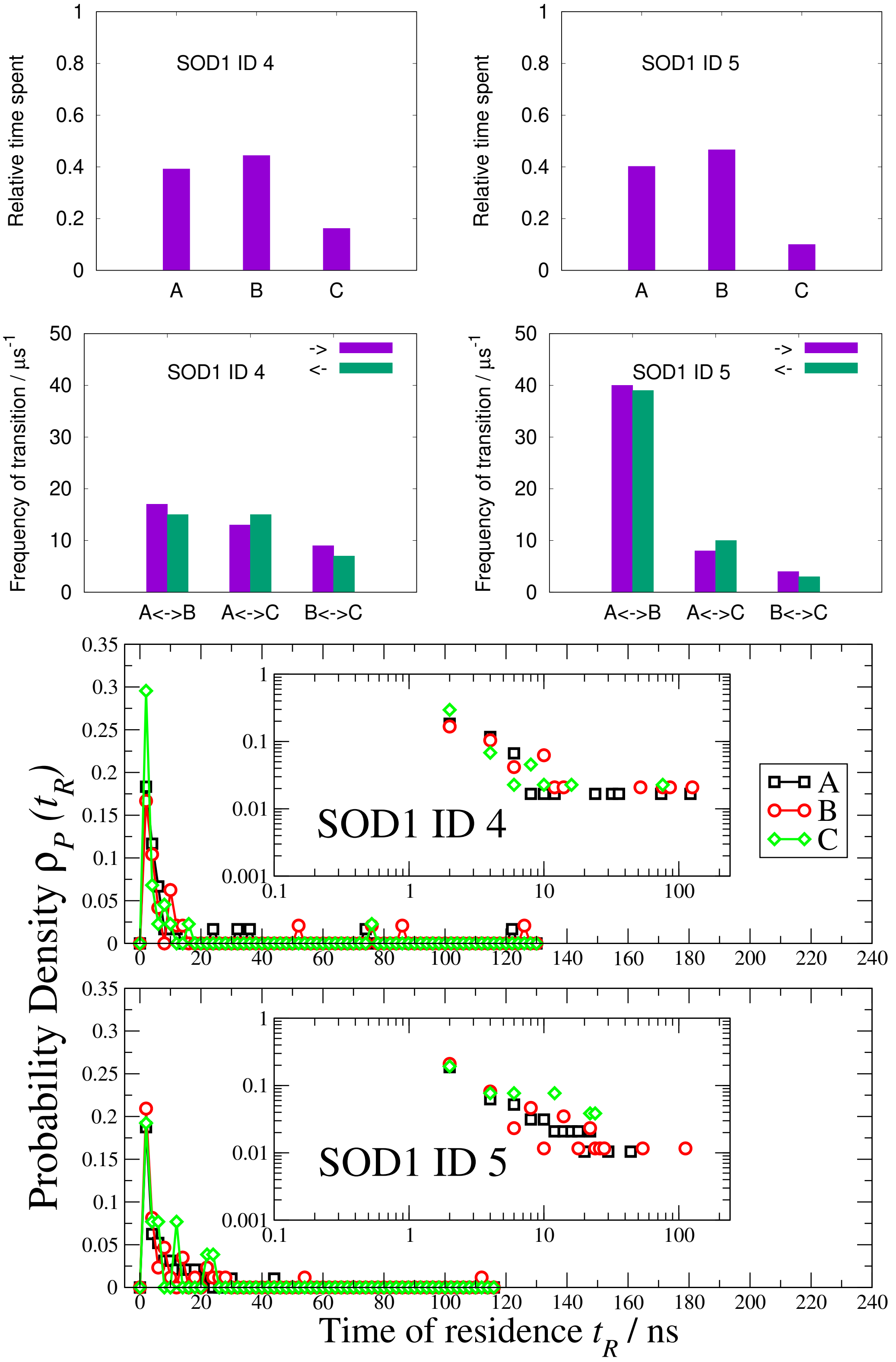


Figure I.4: Same as Fig. I.2 but for SOD1s ID 4 and 5.

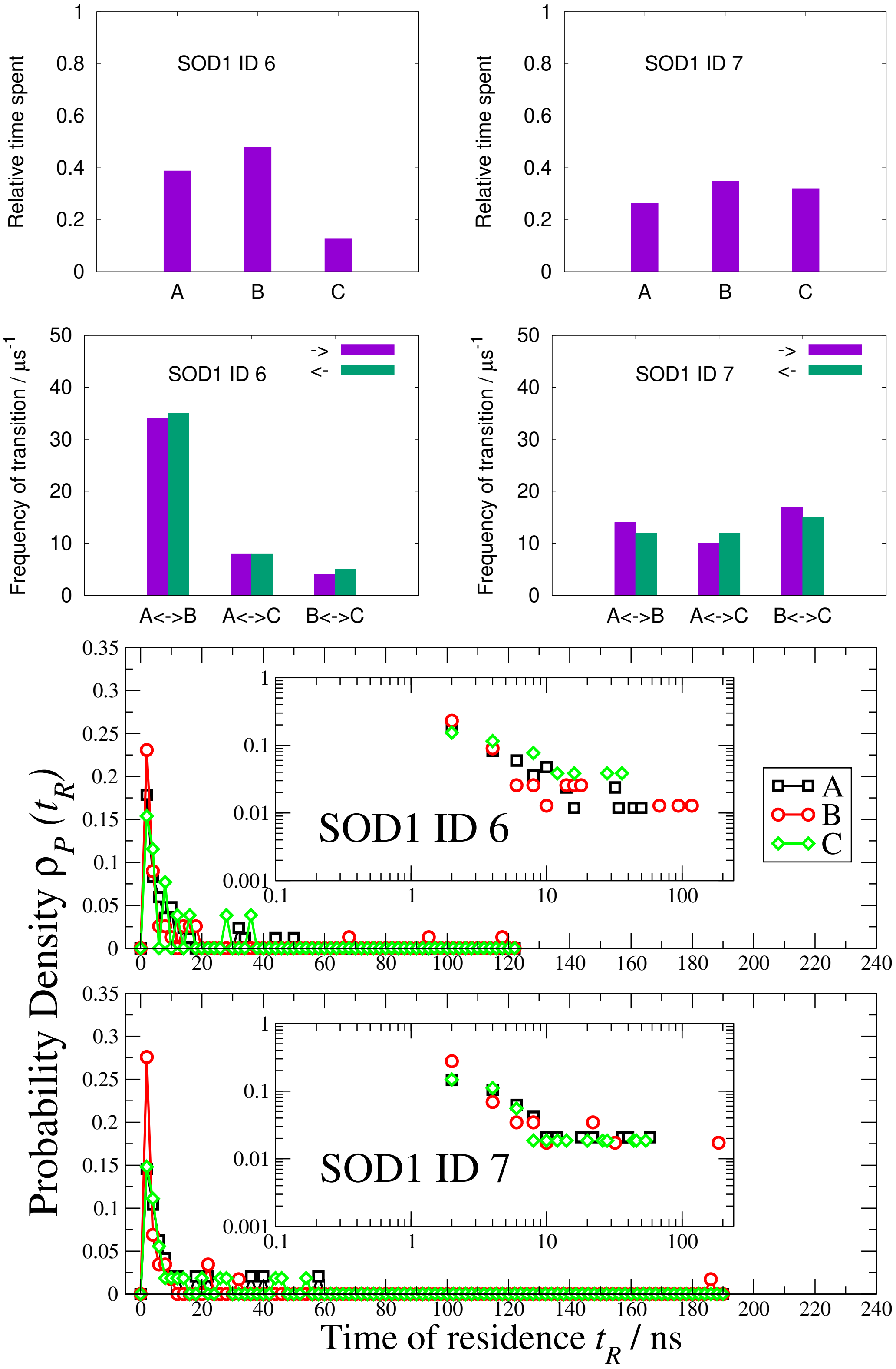


Figure I.5: Same as in Fig. I.2, but for SOD1s ID 6 and 7.

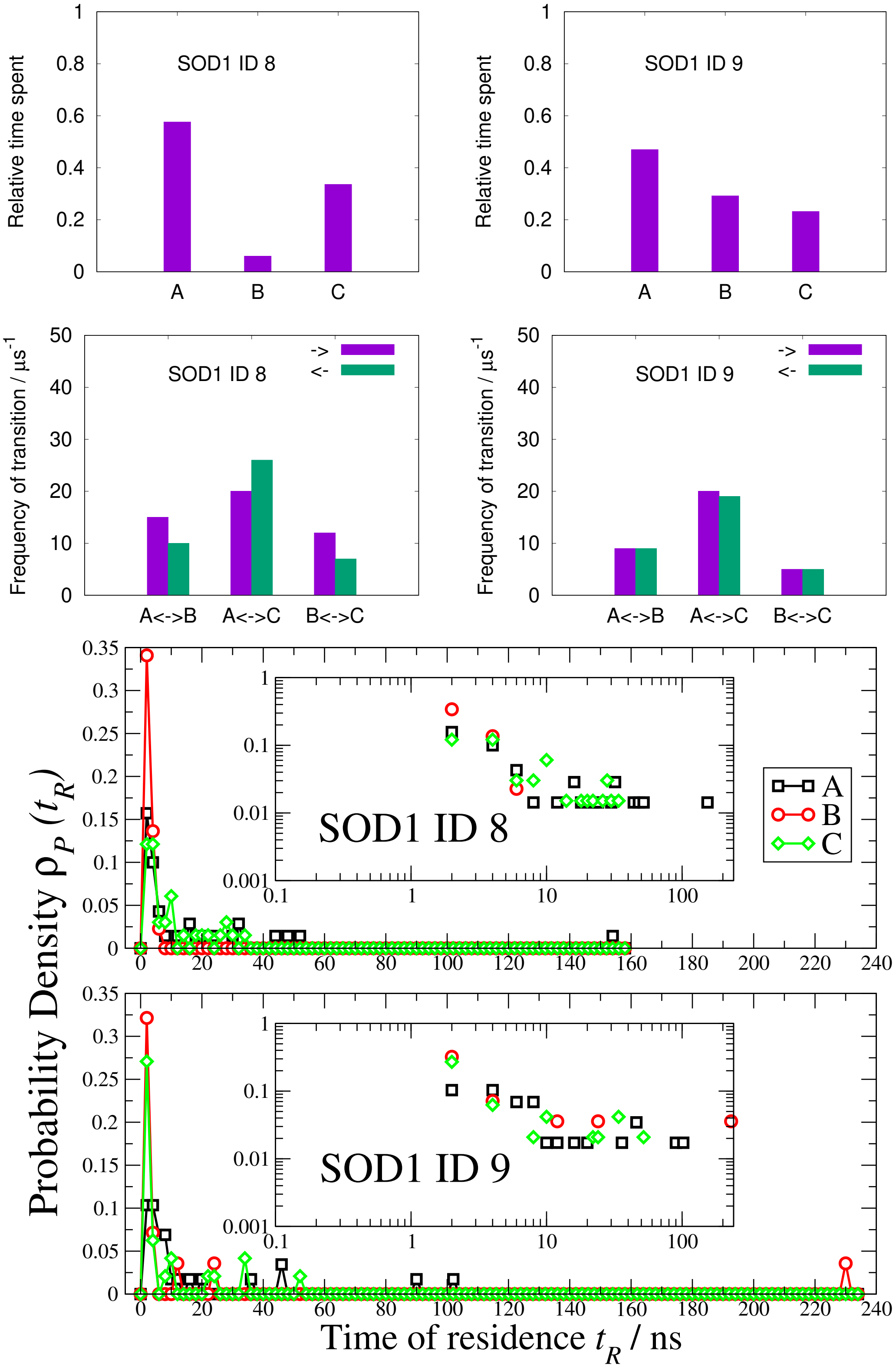


Figure I.6: Same as in Fig. I.2, but for SOD1s ID 8 and 9.

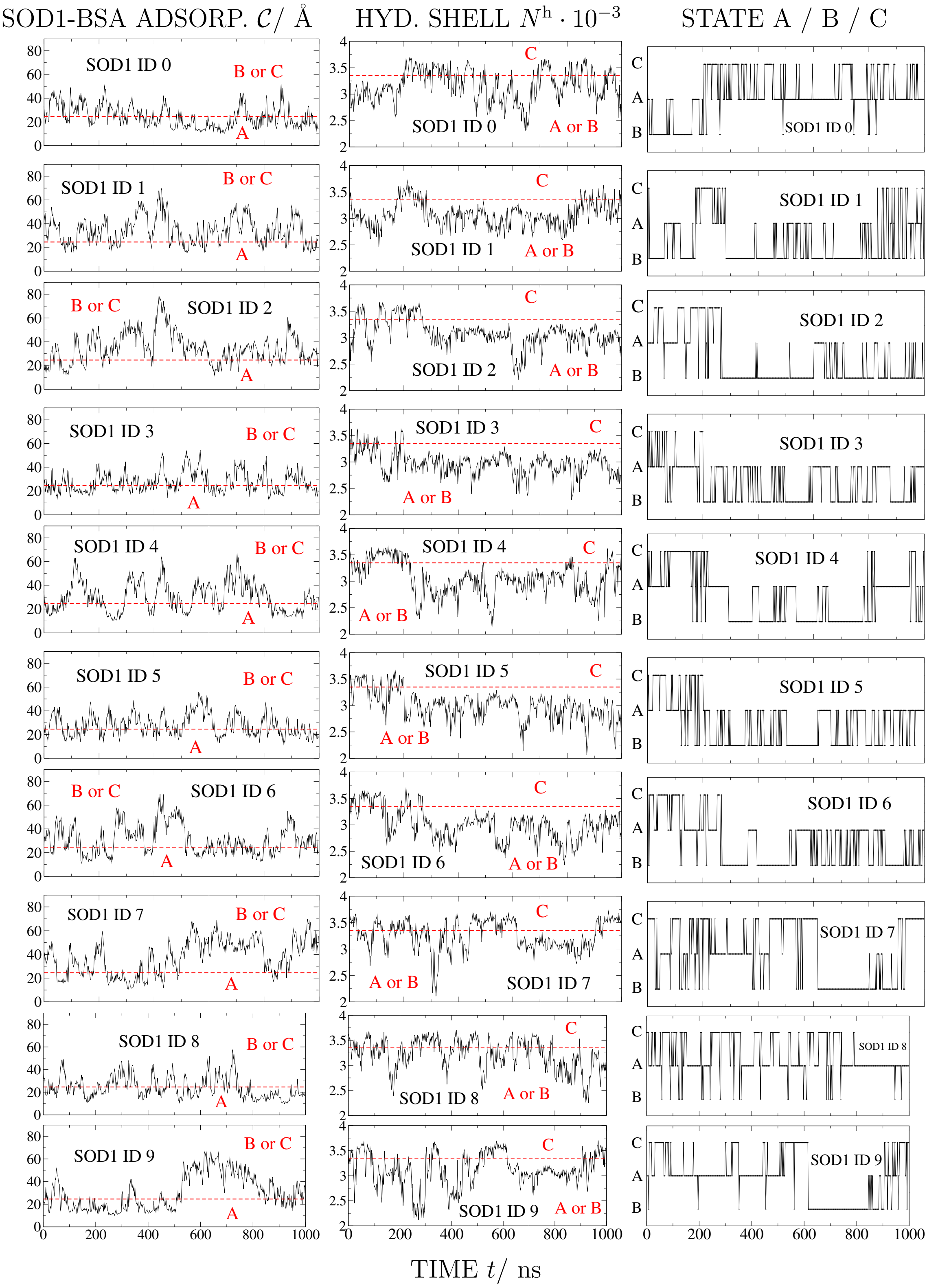


Figure I.7: Time evolution of individual SOD1 in BSA solution (x-axis). From top to bottom, each of the SOD1s (ID from 0 to 9). The y-axis shows (left) the adsorption factor $\mathcal{C}_i$/ Å, (center) the number of water molecules that hydrate the SOD1 $N^{\mathrm{h}}_{\mathrm{SOD1},i} \cdot 10^{-3}$, and (right) whether the SOD1 is in the A, B, or C state, as indicated at the top of the figure.

Bibliography

[1] L. E. Coronas, O. Vilanova, V. Bianco, F. de los Santos, and G. Franzese, "The Franzese-Stanley Coarse Grained Model for Hydration Water," in *Properties of Water from Numerical and Experimental Perspectives* (F. Marelli, ed.), ch. 3, pp. 51–79, Boca Raton: CRC Press, 1 ed., 2022.

[2] M. Chaplin, "Do we underestimate the importance of water in cell biology?," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 11, pp. 861–866, 2006.

[3] D. Laage, T. Elsaesser, and J. T. Hynes, "Water dynamics in the hydration shells of biomolecules," *Chemical Reviews*, vol. 117, no. 16, pp. 10694–10725, 2017. PMID: 28248491.

[4] S. Piana and D. E. Shaw, "Atomic-level description of protein folding inside the groel cavity," *The Journal of Physical Chemistry B*, vol. 122, no. 49, pp. 11440–11449, 2018. PMID: 30277396.

[5] V. Bianco and G. Franzese, "Contribution of water to pressure and cold denaturation of proteins," *Physical Review Letters*, vol. 115, p. 108101, Sep 2015.

[6] V. Bianco, G. Franzese, C. Dellago, and I. Coluzza, "Role of water in the selection of stable proteins at ambient and extreme thermodynamic conditions," *Physical Review X*, vol. 7, p. 021047, Jun 2017.

[7] V. Bianco, N. Pagès-Gelabert, I. Coluzza, and G. Franzese, "How the stability of a folded protein depends on interfacial water properties and residue-residue interactions," *Journal of Molecular Liquids*, vol. 245, pp. 129–139, 2017. Recent Progresses on the Experimental & Theoretical-Computational Techniques for the Study of Liquids and Supercritical Fluids. From Simple to Complex Systems.

[8] V. Bianco, M. Alonso-Navarro, D. Di Silvio, S. Moya, A. L. Cortajarena, and I. Coluzza, "Proteins are Solitary! Pathways of Protein Folding and Aggregation in Protein Mixtures," *The Journal of Physical Chemistry Letters*, vol. 10, no. 17, pp. 4800–4804, 2019.

[9] V. Bianco, G. Franzese, and I. Coluzza, “In silico evidence that protein unfolding is a precursor of protein aggregation,” *ChemPhysChem*, vol. 21, no. 5, pp. 377–384, 2020.

[10] J. F. Nagle and S. Tristram-Nagle, “Structure of lipid bilayers,” *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes*, vol. 1469, no. 3, pp. 159 – 195, 2000.

[11] W. Zhao, D. E. Moilanen, E. E. Fenn, and M. D. Fayer, “Water at the surfaces of aligned phospholipid multibilayer model membranes probed with ultrafast vibrational spectroscopy,” *Journal of the American Chemical Society*, vol. 130, no. 42, pp. 13927–13937, 2008. PMID: 18823116.

[12] F. Martelli, H.-Y. Ko, C. Calero Borallo, and G. Franzese, “Structural properties of water confined by phospholipid membranes,” *Frontiers of Physics*, vol. 13, p. 136801, Jul 2017.

[13] C. Calero and G. Franzese, “Membranes with different hydration levels: The interface between bound and unbound hydration water,” *Journal of Molecular Liquids*, vol. 273, pp. 488–496, 2019.

[14] S. Samatas, C. Calero, F. Martelli, and G. Franzese, “Water Between Membranes: Structure and Dynamics,” in *Biomembrane Simulations: Computational Studies of Biological Membranes* (M. L. Berkowitz, ed.), ch. 4, CRC Press, 2021.

[15] D. Zhong, S. K. Pal, and A. H. Zewail, “Biological water: A critique,” *Chemical Physics Letters*, vol. 503, no. 1, pp. 1–11, 2011.

[16] P. Ball, “Water is an active matrix of life for cell and molecular biology,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 51, pp. 13327–13335, 2017.

[17] N. Kobko, L. Paraskevas, E. del Rio, and J. J. Dannenberg, “Cooperativity in amide hydrogen bonding chains: Implications for protein-folding models,” *Journal of the American Chemical Society*, vol. 123, no. 18, pp. 4348–4349, 2001. PMID: 11457207.

[18] J. F. Ouyang and R. P. A. Bettens, “When are many-body effects significant?,” *Journal of Chemical Theory and Computation*, vol. 12, no. 12, pp. 5860–5867, 2016. PMID: 27779845.

[19] P. Ball, “Water — an enduring mystery,” *Nature*, vol. 452, pp. 291–292, Mar 2008.

[20] K. Amann-Winkel, R. Böhmer, F. Fujara, C. Gainaru, B. Geil, and T. Loerting, "Colloquium: Water's controversial glass transitions," *Reviews of Modern Physics*, vol. 88, p. 011002, Feb 2016.

[21] G. Franzese and H. E. Stanley, *Understanding the Unusual Properties of Water*, ch. 7. CRC Press, 2010.

[22] P. Gallo, K. Amann-Winkel, C. A. Angell, M. A. Anisimov, F. Caupin, C. Chakravarty, E. Lascaris, T. Loerting, A. Z. Panagiotopoulos, J. Russo, J. A. Sellberg, H. E. Stanley, H. Tanaka, C. Vega, L. Xu, and L. G. M. Pettersson, "Water: A tale of two liquids," *Chemical Reviews*, vol. 116, no. 13, pp. 7463–7500, 2016. PMID: 27380438.

[23] P. H. Handle, T. Loerting, and F. Sciortino, "Supercooled and glassy water: Metastable liquid(s), amorphous solid(s), and a no-man's land," *Proceedings of the National Academy of Sciences*, vol. 114, no. 51, pp. 13336–13344, 2017.

[24] P. Gallo, J. Bachler, L. E. Bove, R. Böhmer, G. Camisasca, L. E. Coronas, H. R. Corti, I. de Almeida Ribeiro, M. de Koning, G. Franzese, V. Fuentes-Landete, C. Gainaru, T. Loerting, J. M. M. de Oca, P. H. Poole, M. Rovere, F. Sciortino, C. M. Tonauer, and G. A. Appignanesi, "Advances in the study of supercooled water," *European Physical Journal E*, vol. 44, no. 11, p. 143, 2021.

[25] O. Mishima, L. D. Calvert, and E. Whalley, "'Melting ice' I at 77 K and 10 kbar: a new method of making amorphous solids," *Nature*, vol. 310, pp. 393–395, 1984.

[26] O. Mishima, L. D. Calvert, and E. Whalley, "An apparently first-order transition between two amorphous phases of ice induced by pressure," *Nature*, vol. 314, pp. 76–78, 1985.

[27] O. Mishima, "Reversible first-order transition between two H2O amorphs at $\sim$0.2 GPa and $\sim$135 K," *The Journal of Chemical Physics*, vol. 100, no. 8, pp. 5910–5912, 1994.

[28] C. G. Salzmann, "Advances in the experimental exploration of water's phase diagram," *The Journal of Chemical Physics*, vol. 150, no. 6, p. 060901, 2019.

[29] K. H. Kim, A. Späh, H. Pathak, F. Perakis, D. Mariedahl, K. Amann-Winkel, J. A. Sellberg, J. H. Lee, S. Kim, J. Park, K. H. Nam, T. Katayama, and A. Nilsson, "Maxima in the thermodynamic response and correlation functions of deeply supercooled water," *Science*, vol. 358, no. 6370, pp. 1589–1593, 2017.

[30] C. A. Angell, W. J. Sichina, and M. Oguni, "Heat capacity of water at extremes of supercooling and superheating," *The Journal of Physical Chemistry*, vol. 86, no. 6, pp. 998–1002, 1982.

[31] R. J. Speedy and C. A. Angell, "Isothermal compressibility of supercooled water and evidence for a thermodynamic singularity at -45°C," *The Journal of Chemical Physics*, vol. 65, no. 3, pp. 851–858, 1976.

[32] C. A. Angell, J. Shuppert, and J. C. Tucker, "Anomalous properties of supercooled water. Heat capacity, expansivity, and proton magnetic resonance chemical shift from 0 to -38%," *The Journal of Physical Chemistry*, vol. 77, no. 26, pp. 3092–3099, 1973.

[33] H. Stanley, J. Teixeira, A. Geiger, and R. Blumberg, "Interpretation of the unusual behavior of h2o and d2o at low temperature: Are concepts of percolation relevant to the "puzzle of liquid water"?," *Physica A: Statistical Mechanics and its Applications*, vol. 106, no. 1, pp. 260–277, 1981.

[34] S. Sastry, P. G. Debenedetti, F. Sciortino, and H. E. Stanley, "Singularity-free interpretation of the thermodynamics of supercooled water," *Physical Review E*, vol. 53, pp. 6144–6154, Jun 1996.

[35] H. E. Stanley and J. Teixeira, "Interpretation of the unusual behavior of H_2O and D_2O at low temperatures: Tests of a percolation model," *The Journal of Chemical Physics*, vol. 73, no. 7, pp. 3404–3422, 1980.

[36] R. J. Speedy, "Limiting forms of the thermodynamic divergences at the conjectured stability limits in superheated and supercooled water," *The Journal of Physical Chemistry*, vol. 86, no. 15, pp. 3002–3005, 1982.

[37] P. H. Poole, F. Sciortino, U. Essmann, and H. E. Stanley, "Phase behaviour of metastable water," *Nature*, vol. 360, pp. 324–328, 1992.

[38] P. H. Poole, F. Sciortino, T. Grande, H. E. Stanley, and C. A. Angell, "Effect of hydrogen bonds on the thermodynamic behavior of liquid water," *Physical Review Letters*, vol. 73, pp. 1632–1635, Sep 1994.

[39] C. A. Angell, "Insights into phases of liquid water from study of its unusual glass-forming properties," *Science*, vol. 319, no. 5863, pp. 582–587, 2008.

[40] K. Stokely, M. G. Mazza, H. E. Stanley, and G. Franzese, "Effect of hydrogen bond cooperativity on the behavior of water," *Proceedings of the National Academy of Sciences*, vol. 107, no. 4, pp. 1301–1306, 2010.

[41] F. Caupin, V. Holten, C. Qiu, E. Guillerm, M. Wilke, M. Frenz, J. Teixeira, and A. K. Soper, "Comment on "Maxima in the thermodynamic response and correlation functions of deeply supercooled water"," *Science*, vol. 360, no. 6390, 2018.

[42] K. H. Kim, A. Späh, H. Pathak, F. Perakis, D. Mariedahl, K. Amann-Winkel, J. A. Sellberg, J. H. Lee, S. Kim, J. Park, K. H. Nam, T. Katayama, and A. Nilsson, “Response to Comment on “Maxima in the thermodynamic response and correlation functions of deeply supercooled water”,” *Science*, vol. 360, no. 6390, 2018.

[43] S. Woutersen, B. Ensing, M. Hilbers, Z. Zhao, and C. A. Angell, “A liquid-liquid transition in supercooled aqueous solution related to the HDA-LDA transition,” *Science*, vol. 359, no. 6380, pp. 1127–1131, 2018.

[44] K. H. Kim, K. Amann-Winkel, N. Giovambattista, A. Späh, F. Perakis, H. Pathak, M. L. Parada, C. Yang, D. Mariedahl, T. Eklund, T. J. Lane, S. You, S. Jeong, M. Weston, J. H. Lee, I. Eom, M. Kim, J. Park, S. H. Chun, P. H. Poole, and A. Nilsson, “Experimental observation of the liquid-liquid transition in bulk supercooled water under pressure,” *Science*, vol. 370, no. 6519, pp. 978–982, 2020.

[45] O. Mishima and H. E. Stanley, “The relationship between liquid, supercooled and glassy water,” *Nature*, vol. 396, pp. 329–335, Nov 1998.

[46] H. E. Stanley, S. V. Buldyrev, G. Franzese, N. Giovambattista, and F. W. Starr, “Static and dynamic heterogeneities in water,” *Philosophical Transactions Of The Royal Society Of London Series A-Mathematical Physical And Engineering Sciences*, vol. 363, pp. 509–523, Feb 2005.

[47] F. Mallamace, C. Branca, M. Broccio, C. Corsaro, C.-Y. Mou, and S.-H. Chen, “The anomalous behavior of the density of water in the range 30 K $< T <$ 373 K,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 47, pp. 18387–18391, 2007.

[48] F. Mallamace, M. Broccio, C. Corsaro, A. Faraone, D. Majolino, V. Venuti, L. Liu, C.-Y. Mou, and S.-H. Chen, “Evidence of the existence of the low-density liquid phase in supercooled, confined water,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 2, pp. 424–428, 2007.

[49] D. Liu, Y. Zhang, C.-C. Chen, C.-Y. Mou, P. H. Poole, and S.-H. Chen, “Observation of the density minimum in deeply supercooled confined water,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 23, pp. 9570–9574, 2007.

[50] K.-i. Murata and H. Tanaka, “Liquid–liquid transition without macroscopic phase separation in a water–glycerol mixture,” *Nature Materials*, vol. 11, pp. 436–443, May 2012.

[51] K.-i. Murata and H. Tanaka, “General nature of liquid–liquid transition in aqueous organic solutions,” *Nature Communications*, vol. 4, p. 2844, Nov 2013.

[52] R. Mancinelli, S. Imberti, A. K. Soper, K. H. Liu, C. Y. Mou, F. Bruni, and M. A. Ricci, “Multiscale Approach to the Structural Study of Water Confined in MCM41,” *The Journal of Physical Chemistry B*, vol. 113, no. 50, pp. 16169–16177, 2009.

[53] A. K. Soper, “Density profile of water confined in cylindrical pores in MCM-41 silica,” *Journal of Physics: Condensed Matter*, vol. 24, p. 064107, jan 2012.

[54] Z. Wang, K. Ito, and S.-H. Chen, “Detection of the liquid-liquid transition in the deeply cooled water confined in MCM-41 with elastic neutron scattering technique,” *Nuovo Cimento C Geophysics Space Physics C*, vol. 39, p. 299, May 2016.

[55] F. Leoni and G. Franzese, “Effects of confinement between attractive and repulsive walls on the thermodynamics of an anomalous fluid,” *Physical Review E*, vol. 94, p. 062604, Dec 2016.

[56] Y. Zhang, A. Faraone, W. A. Kamitakahara, K.-H. Liu, C.-Y. Mou, J. B. Leão, S. Chang, and S.-H. Chen, “Density hysteresis of heavy water confined in a nanoporous silica matrix,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 30, pp. 12206–12211, 2011.

[57] D. R. Paul, “Creating New Types of Carbon-Based Membranes,” *Science*, vol. 335, no. 6067, pp. 413–414, 2012.

[58] R. Zangi and A. E. Mark, “Monolayer Ice,” *Physical Review Letters*, vol. 91, no. 2, p. 25502, 2003.

[59] S. Han, M. Y. Choi, P. Kumar, and H. E. Stanley, “Phase transitions in confined water nanofilms,” *Nature Physics*, vol. 6, pp. 685–689, Sep 2010.

[60] G. Camisasca, M. De Marzio, M. Rovere, and P. Gallo, “Slow dynamics and structure of supercooled water in confinement,” *Entropy*, vol. 19, no. 4, p. 185, 2017. Copyright - Copyright MDPI AG 2017; Última actualización - 2018-01-17.

[61] L. Fumagalli, A. Esfandiar, R. Fabregas, S. Hu, P. Ares, A. Janardanan, Q. Yang, B. Radha, T. Taniguchi, K. Watanabe, G. Gomila, K. S. Novoselov, and A. K. Geim, “Anomalously low dielectric constant of confined water,” *Science*, vol. 360, no. 6395, pp. 1339–1342, 2018.

[62] P. Gallo, M. Rovere, and E. Spohr, “Supercooled confined water and the mode coupling crossover temperature,” *Physical Review Letters*, vol. 85, pp. 4317–4320, Nov 2000.

[63] P. Gallo, M. Rovere, and S.-H. Chen, "Dynamic crossover in supercooled confined water: Understanding bulk properties through confinement," *The Journal of Physical Chemistry Letters*, vol. 1, no. 4, pp. 729–733, 2010.

[64] F. Leoni and G. Franzese, "Structural behavior and dynamics of an anomalous fluid between attractive and repulsive walls: Templating, molding, and superdiffusion," *The Journal of Chemical Physics*, vol. 141, no. 17, p. 174501, 2014.

[65] A. Faraone, L. Liu, C.-Y. Mou, C.-W. Yen, and S.-H. Chen, "Fragile-to-strong liquid transition in deeply supercooled confined water," *The Journal of Chemical Physics*, vol. 121, no. 22, pp. 10843–10846, 2004.

[66] E. Stefanutti, L. E. Bove, G. Lelong, M. A. Ricci, A. K. Soper, and F. Bruni, "Ice crystallization observed in highly supercooled confined water," *Physical Chemistry Chemical Physics*, vol. 21, pp. 4931–4938, 2019.

[67] M.-C. Bellissent-Funel, A. Hassanali, M. Havenith, R. Henchman, P. Pohl, F. Sterpone, D. van der Spoel, Y. Xu, and A. E. Garcia, "Water determines the structure and dynamics of proteins," *Chemical Reviews*, vol. 116, no. 13, pp. 7673–7697, 2016. PMID: 27186992.

[68] C. I. Lynch, S. Rao, and M. S. P. Sansom, "Water in nanopores and biological channels: A molecular simulation perspective," *Chemical Reviews*, vol. 120, no. 18, pp. 10298–10335, 2020. PMID: 32841020.

[69] E. Persson and B. Halle, "Cell water dynamics on multiple time scales," *Proceedings of the National Academy of Sciences*, vol. 105, no. 17, pp. 6266–6271, 2008.

[70] M. Jasnin, A. Stadler, M. Tehei, and G. Zaccai, "Specific cellular water dynamics observed in vivo by neutron scattering and NMR," *Physical Chemistry Chemical Physics*, vol. 12, pp. 10154–10160, 2010.

[71] G. Franzese and H. Eugene Stanley, "A theory for discriminating the mechanism responsible for the water density anomaly," *Physica A: Statistical Mechanics and its Applications*, vol. 314, no. 1, pp. 508–513, 2002. Horizons in Complex Systems.

[72] G. Franzese and H. E. Stanley, "Liquid-liquid critical point in a Hamiltonian model for water: analytic solution," *Journal of Physics: Condensed Matter*, vol. 14, pp. 2201–2209, mar 2002.

[73] D. March, V. Bianco, and G. Franzese, "Protein unfolding and aggregation near a hydrophobic interface," *Polymers*, vol. 13, no. 1, 2021.

[74] P. Barnes, J. L. Finney, J. D. Nicholas, and J. E. Quinn, "Cooperative effects in simulated water," *Nature*, vol. 282, pp. 459–464, 11 1979.

[75] J. L. Finney, “The water molecule and its interactions: the interaction between theory, modelling, and experiment,” *Journal of Molecular Liquids*, vol. 90, no. 1, pp. 303–312, 2001.

[76] J. F. Ouyang and R. P. A. Bettens, “Modelling water: A lifetime enigma,” *CHIMIA*, vol. 69, p. 104, Mar. 2015.

[77] T. A. Kesselring, G. Franzese, S. V. Buldyrev, H. J. Herrmann, and H. E. Stanley, “Nanoscale Dynamics of Phase Flipping in Water near its Hypothesized Liquid-Liquid Critical Point,” *Scientific Reports*, vol. 2, p. 474, Jun 2012.

[78] J. C. Palmer, F. Martelli, Y. Liu, R. Car, A. Z. Panagiotopoulos, and P. G. Debenedetti, “Metastable liquid-liquid transition in a molecular model of water.,” *Nature*, vol. 510, pp. 385–8, jun 2014.

[79] D. Chandler, “Metastability and no criticality,” *Nature*, vol. 531, pp. E1–E2, Mar 2016.

[80] J. C. Palmer, F. Martelli, Y. Liu, R. Car, A. Z. Panagiotopoulos, and P. G. Debenedetti, “Palmer et al. reply,” *Nature*, vol. 531, pp. E2–E3, Mar 2016.

[81] D. T. Limmer and D. Chandler, “The putative liquid-liquid transition is a liquid-solid transition in atomistic models of water. II,” *The Journal of Chemical Physics*, vol. 138, no. 21, p. 214504, 2013.

[82] J. C. Palmer, A. Haji-Akbari, R. S. Singh, F. Martelli, R. Car, A. Z. Panagiotopoulos, and P. G. Debenedetti, “Comment on “The putative liquid-liquid transition is a liquid-solid transition in atomistic models of water” [I and II: J. Chem. Phys. 135, 134503 (2011); J. Chem. Phys. 138, 214504 (2013)],” *The Journal of Chemical Physics*, vol. 148, no. 13, p. 137101, 2018.

[83] N. Wilding and K. Binder, “Finite-size scaling for near-critical continuum fluids at constant pressure,” *Physica A: Statistical Mechanics and its Applications*, vol. 231, no. 4, pp. 439–447, 1996.

[84] E. Lascaris, T. A. Kesselring, G. Franzese, S. V. Buldyrev, H. J. Herrmann, and H. E. Stanley, “Response functions near the liquid-liquid critical point of ST2 water,” *AIP Conference Proceedings*, vol. 1518, no. 1, pp. 520–526, 2013.

[85] P. G. Debenedetti, F. Sciortino, and G. H. Zerze, “Second critical point in two realistic models of water,” *Science*, vol. 369, no. 6501, pp. 289–292, 2020.

[86] J. Weis, F. Sciortino, A. Z. Panagiotopoulos, and P. G. Debenedetti, “Liquid–liquid criticality in the WAIL water model,” *The Journal of Chemical Physics*, vol. 157, no. 2, p. 024502, 2022.

[87] T. A. Kesselring, E. Lascaris, G. Franzese, S. V. Buldyrev, H. J. Herrmann, and H. E. Stanley, “Finite-size scaling investigation of the liquid-liquid critical point in ST2 water and its stability with respect to crystallization,” *The Journal of Chemical Physics*, vol. 138, no. 24, p. 244506, 2013.

[88] R. Shi, J. Russo, and H. Tanaka, “Common microscopic structural origin for water’s thermodynamic and dynamic anomalies,” *The Journal of Chemical Physics*, vol. 149, no. 22, p. 224502, 2018.

[89] T. E. Gartner, L. Zhang, P. M. Piaggi, R. Car, A. Z. Panagiotopoulos, and P. G. Debenedetti, “Signatures of a liquid-liquid transition in an ab initio deep neural network model for water,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 42, pp. 26040–26046, 2020.

[90] A. Eltareb, G. E. Lopez, and N. Giovambattista, “Evidence of a liquid–liquid phase transition in H_2O and D_2O from path-integral molecular dynamics simulations,” *Scientific Reports*, vol. 12, no. 1, p. 6004, 2022.

[91] P. Kumar, G. Franzese, S. V. Buldyrev, and H. E. Stanley, “Molecular dynamics study of orientational cooperativity in water,” *Physical Review E*, vol. 73, p. 041505, Apr 2006.

[92] C. Calero, H. E. Stanley, and G. Franzese, “Structural interpretation of the large slowdown of water dynamics at stacked phospholipid membranes for decreasing hydration level: All-atom molecular dynamics,” *Materials*, vol. 9, no. 5, 2016.

[93] J. Martí, C. Calero, and G. Franzese, “Structure and dynamics of water at carbon-based interfaces,” *Entropy*, vol. 19, no. 3, 2017.

[94] W. Zheng, G. L. Dignon, N. Jovic, X. Xu, R. M. Regy, N. L. Fawzi, Y. C. Kim, R. B. Best, and J. Mittal, “Molecular details of protein condensates probed by microsecond long atomistic simulations,” *Journal of Physical Chemistry B*, vol. 124, no. 51, pp. 11671–11679, 2020.

[95] C. Vega, J. L. F. Abascal, M. M. Conde, and J. L. Aragones, “What ice can teach us about water interactions: a critical comparison of the performance of different water models,” *Faraday Discussions*, vol. 141, pp. 251–276, 2009.

[96] T. Yagasaki, M. Matsumoto, and H. Tanaka, “Spontaneous liquid-liquid phase separation of water,” *Physical Review E*, vol. 89, p. 020301, Feb 2014.

[97] M. A. González, C. Valeriani, F. Caupin, and J. L. F. Abascal, “A comprehensive scenario of the thermodynamic anomalies of water using the TIP4P/2005 model,” *The Journal of Chemical Physics*, vol. 145, no. 5, p. 054505, 2016.

[98] T. James, D. J. Wales, and J. Hernández-Rojas, "Global minima for water clusters (H2O)n, $n \leq 21$, described by a five-site empirical potential," *Chemical Physics Letters*, vol. 415, no. 4, pp. 302–307, 2005.

[99] J. Hernández-Rojas, F. Calvo, F. Rabilloud, J. Bretón, and J. M. Gomez Llorente, "Modeling Water Clusters on Cationic Carbonaceous Seeds," *The Journal of Physical Chemistry A*, vol. 114, no. 27, pp. 7267–7274, 2010.

[100] J. Y. Abraham, S. V. Buldyrev, and N. Giovambattista, "Liquid and glass polymorphism in a monatomic system with isotropic, smooth pair interactions," *The Journal of Physical Chemistry B*, vol. 115, no. 48, pp. 14229–14239, 2011. PMID: 21992558.

[101] P. Vilaseca and G. Franzese, "Isotropic soft-core potentials with two characteristic length scales and anomalous behaviour," *Journal of Non-Crystalline Solids*, vol. 357, no. 2, pp. 419–426, 2011. 6th International Discussion Meeting on Relaxation in Complex Systems.

[102] M. L. de Haro, A. Rodríguez-Rivas, S. B. Yuste, and A. Santos, "Structural properties of the Jagla fluid," *Physical Review E*, vol. 98, p. 012138, Jul 2018.

[103] V. Molinero and E. B. Moore, "Water modeled as an intermediate element between carbon and silicon," *The Journal of Physical Chemistry B*, vol. 113, no. 13, pp. 4008–4016, 2009. PMID: 18956896.

[104] J. Jin, Y. Han, A. J. Pak, and G. A. Voth, "A new one-site coarse-grained model for water: Bottom-up many-body projected water (BUMPer). I. General theory and model," *The Journal of Chemical Physics*, vol. 154, no. 4, p. 044104, 2021.

[105] J. Jin, A. J. Pak, Y. Han, and G. A. Voth, "A new one-site coarse-grained model for water: Bottom-up many-body projected water (BUMPer). II. Temperature transferability and structural properties at low temperature," *The Journal of Chemical Physics*, vol. 154, no. 4, p. 044105, 2021.

[106] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations," *The Journal of Physical Chemistry B*, vol. 111, no. 27, pp. 7812–7824, 2007. PMID: 17569554.

[107] M. R. Machado, E. E. Barrera, F. Klein, M. Sóñora, S. Silva, and S. Pantano, "The SIRAH 2.0 Force Field: Altius, Fortius, Citius," *Journal of Chemical Theory and Computation*, vol. 15, no. 4, pp. 2719–2733, 2019. PMID: 30810317.

[108] F. Klein, E. E. Barrera, and S. Pantano, “Assessing SIRAH’s Capability to Simulate Intrinsically Disordered Proteins and Peptides,” *Journal of Chemical Theory and Computation*, vol. 17, no. 2, pp. 599–604, 2021. PMID: 33411518.

[109] L. Darré, M. R. Machado, P. D. Dans, F. E. Herrera, and S. Pantano, “Another coarse grain model for aqueous solvation: Wat four?,” *Journal of Chemical Theory and Computation*, vol. 6, no. 12, pp. 3793–3807, 2010.

[110] M. Tsanai, P. W. J. M. Frederix, C. F. E. Schroer, P. C. T. Souza, and S. J. Marrink, “Coacervate formation studied by explicit solvent coarse-grain molecular dynamics with the Martini model,” *Chemical Science*, vol. 12, pp. 8521–8530, 2021.

[111] F. Aydin, R. Sun, and J. M. Swanson, “Mycolactone toxin membrane permeation: Atomistic versus coarse-grained martini simulations,” *Biophysical Journal*, vol. 117, no. 1, pp. 87–98, 2019.

[112] B. Cheng, E. A. Engel, J. Behler, C. Dellago, and M. Ceriotti, “Ab initio thermodynamics of liquid and solid water,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 4, pp. 1110–1115, 2019.

[113] H. Chan, M. J. Cherukara, B. Narayanan, T. D. Loeffler, C. Benmore, S. K. Gray, and S. K. Sankaranarayanan, “Machine learning coarse grained models for water,” *Nature Communications*, vol. 10, no. 1, p. 379, 2019.

[114] G. Franzese, M. Yamada, and H. E. Stanley, “Hydrogen-bonded liquids: Effects of correlations of orientational degrees of freedom,” *AIP Conference Proceedings*, vol. 519, no. 1, pp. 281–287, 2000.

[115] G. Franzese and H. E. Stanley, “The Widom line of supercooled water,” *Journal of Physics: Condensed Matter*, vol. 19, p. 205126, apr 2007.

[116] G. Franzese, M. I. Marqués, and H. Eugene Stanley, “Intramolecular coupling as a mechanism for a liquid-liquid phase transition,” *Physical Review E*, vol. 67, p. 011103, Jan 2003.

[117] P. Kumar, G. Franzese, and H. E. Stanley, “Predictions of dynamic behavior under pressure for two scenarios to explain water anomalies,” *Physical Review Letters*, vol. 100, p. 105701, Mar 2008.

[118] P. Kumar, G. Franzese, S. V. Buldyrev, and H. E. Stanley, *Dynamics of Water at Low Temperatures and Implications for Biomolecule*, vol. 752 of *Lecture Notes in Physics*, pp. 3–22. Springer Berlin Heidelberg, 2008.

[119] P. Kumar, G. Franzese, and H. E. Stanley, “Dynamics and thermodynamics of water,” *Journal of Physics: Condensed Matter*, vol. 20, p. 244114, may 2008.

[120] G. Franzese, K. Stokely, X. qiang Chu, P. Kumar, M. G. Mazza, S.-H. Chen, and H. E. Stanley, "Pressure effects in supercooled water: comparison between a 2D model of water and experiments for surface water on a protein," *Journal of Physics: Condensed Matter*, vol. 20, p. 494210, nov 2008.

[121] G. Franzese and F. de los Santos, "Dynamically slow processes in supercooled water confined between hydrophobic plates," *Journal of Physics: Condensed Matter*, vol. 21, p. 504107, nov 2009.

[122] F. de los Santos and G. Franzese, "Influence of intramolecular couplings in a model for hydrogen-bonded liquids," in *MODELING AND SIMULATION OF NEW MATERIALS: Proceedings of Modeling and Simulation of New Materials: Tenth Granada Lectures* (J. Marro, P. L. Garrido, and P. I. Hurtado, eds.), vol. 1091, (Granada (Spain)), pp. 185–197, AIP, 2009.

[123] V. Bianco and G. Franzese, "Hydrogen bond correlated percolation in a supercooled water monolayer as a hallmark of the critical region," *Journal of Molecular Liquids*, vol. 285, pp. 727–739, 2019.

[124] M. G. Mazza, K. Stokely, E. G. Strekalova, H. E. Stanley, and G. Franzese, "Cluster monte carlo and numerical mean field analysis for the water liquid-liquid phase transition," *Computer Physics Communications*, vol. 180, no. 4, pp. 497–502, 2009.

[125] G. Franzese, A. Hernando-Martínez, P. Kumar, M. G. Mazza, K. Stokely, E. G. Strekalova, F. de los Santos, and H. E. Stanley, "Phase transitions and dynamics of bulk and interfacial water," *Journal of Physics: Condensed Matter*, vol. 22, p. 284103, jun 2010.

[126] F. de los Santos and G. Franzese, "Understanding diffusion and density anomaly in a coarse-grained model for water confined between hydrophobic walls," *The Journal of Physical Chemistry B*, vol. 115, no. 48, pp. 14311–14320, 2011.

[127] E. G. Strekalova, M. G. Mazza, H. E. Stanley, and G. Franzese, "Large decrease of fluctuations for supercooled water in hydrophobic nanoconfinement," *Physical Review Letters*, vol. 106, p. 145701, Apr 2011.

[128] E. G. Strekalova, M. G. Mazza, H. E. Stanley, and G. Franzese, "Hydrophobic nanoconfinement suppresses fluctuations in supercooled water," *Journal of Physics: Condensed Matter*, vol. 24, p. 064111, jan 2012.

[129] E. G. Strekalova, D. Corradini, M. G. Mazza, S. V. Buldyrev, P. Gallo, G. Franzese, and H. E. Stanley, "Effect of hydrophobic environments on the hypothesized liquid-liquid critical point of water," *Journal of Biological Physics*, vol. 38, pp. 97–111, Jan 2012.

[130] V. Bianco, G. Franzese, R. Ruberto, and S. Ancherbak, "Water and anomalous liquids," in *Complex Materials in Physics and Biology* (F. Mallamace and H. E. Stanley, eds.), vol. 176 of *Proceedings of the International School of Physics "Enrico Fermi"*, pp. 113–128, IOS Press, 2012.

[131] M. G. Mazza, K. Stokely, H. E. Stanley, and G. Franzese, "Effect of pressure on the anomalous response functions of a confined water monolayer at low temperature," *The Journal of Chemical Physics*, vol. 137, no. 20, p. 204502, 2012.

[132] F. de los Santos and G. Franzese, "Relations between the diffusion anomaly and cooperative rearranging regions in a hydrophobically nanoconfined water monolayer," *Physical Review E*, vol. 85, p. 010602, Jan 2012.

[133] V. Bianco, M. G. Mazza, K. Stokely, , F. Bruni, H. E. Stanley, and G. Franzese, "Prediction and observation of two dynamic crossovers in hydration water allows the reconciliation of simulations and experiments," in *Proceedings of "Symposium on the Fragility of Glass-formers: A Conference in Honor of C. Austen Angell"* (R. Ganapathy, A. L. Greer, K. F. Kelton, and S. Sastry, eds.), 2013.

[134] V. Bianco and G. Franzese, "Critical behavior of a water monolayer under hydrophobic confinement," *Scientific Reports*, vol. 4, no. 1, p. 4440, 2014.

[135] M. G. Mazza, K. Stokely, S. E. Pagnotta, F. Bruni, H. E. Stanley, and G. Franzese, "More than one dynamic crossover in protein hydration water," *Proceedings of the National Academy of Sciences*, vol. 108, no. 50, pp. 19873–19878, 2011.

[136] M. Ceriotti, J. Cuny, M. Parrinello, and D. E. Manolopoulos, "Nuclear quantum effects and hydrogen bond fluctuations in water," *Proceedings of the National Academy of Sciences*, vol. 110, no. 39, pp. 15591–15596, 2013.

[137] A. K. Soper and M. A. Ricci, "Structures of high-density and low-density water," *Physical Review Letters*, vol. 84, pp. 2881–2884, Mar 2000.

[138] M. Henry, "Nonempirical quantification of molecular interactions in supramolecular assemblies," *ChemPhysChem*, vol. 3, no. 7, pp. 561–569, 2002.

[139] Y. A. Galkina, N. A. Kryuchkova, M. A. Vershinin, and B. A. Kolesov, "Features of strong O–H···O and N–H···O hydrogen bond manifestation in vibrational spectra," *Journal of Structural Chemistry*, vol. 58, pp. 911–918, Sep 2017.

[140] A. Soper and M. Phillips, "A new determination of the structure of water at 25 °C ," *Chemical Physics*, vol. 107, no. 1, pp. 47–60, 1986.

[141] A. Luzar and D. Chandler, "Effect of environment on hydrogen bond dynamics in liquid water," *Physical Review Letters*, vol. 76, pp. 928–931, Feb 1996.

[142] J. Teixeira and M.-C. Bellissent-Funel, "Dynamics of water studied by neutron scattering," *Journal of Physics: Condensed Matter*, vol. 2, pp. SA105–SA108, dec 1990.

[143] R. Ludwig, "Water: From clusters to the bulk," *Angewandte Chemie International Edition*, vol. 40, no. 10, pp. 1808–1827, 2001.

[144] L. Hernández de la Peña and P. G. Kusalik, "Temperature Dependence of Quantum Effects in Liquid Water," *Journal of the American Chemical Society*, vol. 127, no. 14, pp. 5246–5251, 2005.

[145] L. X. Dang and T.-M. Chang, "Molecular dynamics study of water clusters, liquid, and liquid–vapor interface of water with many-body potentials," *The Journal of Chemical Physics*, vol. 106, no. 19, pp. 8149–8159, 1997.

[146] M. I. Heggie, C. D. Latham, S. C. Maynard, and R. Jones, "Cooperative polarisation in ice ih and the unusual strength of the hydrogen bond," *Chemical Physics Letters*, vol. 249, no. 5, pp. 485–490, 1996.

[147] D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water*. New York: Oxford University Press, 1969.

[148] K. Stokely, M. G. Mazza, H. E. Stanley, and G. Franzese, *Metastable Systems under Pressure*, pp. 197–216. NATO Science for Peace and Security Series A: Chemistry and Biology, Springer, 2010.

[149] H. E. Stanley, P. Kumar, S. Han, M. G. Mazza, K. Stokely, S. V. Buldyrev, G. Franzese, F. Mallamace, and L. Xu, "Heterogeneities in confined water and protein hydration water," *Journal of Physics: Condensed Matter*, vol. 21, p. 504105, nov 2009.

[150] H. E. Stanley, S. V. Buldyrev, G. Franzese, P. Kumar, F. Mallamace, M. G. Mazza, K. Stokely, and L. Xu, "Liquid polymorphism: water in nanoconfined and biological environments," *Journal of Physics: Condensed Matter*, vol. 22, no. 28, p. 284101, 2010.

[151] H. E. Stanley, S. V. Buldyrev, P. Kumar, F. Mallamace, M. G. Mazza, K. Stokely, L. Xu, and G. Franzese, "Water in nanoconfined and biological environments: (Plenary Talk, Ngai-Ruocco 2009 IDMRCS Conf.)," *Journal of Non-Crystalline Solids*, vol. 357, no. 2, pp. 629–640, 2011.

[152] S.-H. Chen, L. Liu, E. Fratini, P. Baglioni, A. Faraone, and E. Mamontov, "Observation of fragile-to-strong dynamic crossover in protein hydration water," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 24, pp. 9012–9016, 2006.

[153] L. Liu, S.-H. Chen, A. Faraone, C.-W. Yen, and C.-Y. Mou, "Pressure dependence of fragile-to-strong transition and a possible second critical point in supercooled confined water," *Physical Review Letters*, vol. 95, p. 117802, Sep 2005.

[154] J. Swenson, "Comment on "Pressure Dependence of Fragile-to-Strong Transition and a Possible Second Critical Point in Supercooled Confined Water"," *Physical Review Letters*, vol. 97, p. 189801, Nov 2006.

[155] S. Cerveny, J. Colmenero, and A. Alegría, "Comment on "Pressure Dependence of Fragile-to-Strong Transition and a Possible Second Critical Point in Supercooled Confined Water"," *Physial Review Letters*, vol. 97, p. 189802, Nov 2006.

[156] S.-H. Chen, L. Liu, and A. Faraone, "Chen, Liu, and Faraone Reply:," *Physical Review Letters*, vol. 97, p. 189803, Nov 2006.

[157] X.-q. Chu, A. Faraone, C. Kim, E. Fratini, P. Baglioni, J. B. Leao, and S.-H. Chen, "Proteins remain soft at lower temperatures under pressure," *The Journal of Physical Chemistry B*, vol. 113, no. 15, pp. 5001–5006, 2009. PMID: 19323465.

[158] M. Peyrard, "Glass transition in protein hydration water," *Physical Review E*, vol. 64, p. 011109, Jun 2001.

[159] W. Doster, "The protein-solvent glass transition," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1804, no. 1, pp. 3–14, 2010. Includes Special Section: Protein-Water Interactions.

[160] M. Settles and W. Doster, "Anomalous diffusion of adsorbed water: a neutron scattering study of hydrated myoglobin," *Faraday Discuss.*, vol. 103, pp. 269–279, 1996.

[161] W. A. Kamitakahara and N. Wada, "Neutron spectroscopy of water dynamics in NaX and NaA zeolites," *Physical Review E*, vol. 77, p. 041503, Apr 2008.

[162] S. Takahara, M. Nakano, S. Kittaka, Y. Kuroda, T. Mori, H. Hamano, and T. Yamaguchi, "Neutron Scattering Study on Dynamics of Water Molecules in MCM-41," *The Journal of Physical Chemistry B*, vol. 103, no. 28, pp. 5814–5819, 1999.

[163] N. Naguib, H. Ye, Y. Gogotsi, A. G. Yazicioglu, C. M. Megaridis, and M. Yoshimura, "Observation of water confined in nanometer channels of closed carbon nanotubes," *Nano Letters*, vol. 4, no. 11, pp. 2237–2243, 2004.

[164] J. K. Holt, H. G. Park, Y. Wang, M. Stadermann, A. B. Artyukhin, C. P. Grigoropoulos, A. Noy, and O. Bakajin, "Fast Mass Transport Through Sub-2-Nanometer Carbon Nanotubes," *Science*, vol. 312, no. 5776, pp. 1034–1037, 2006.

[165] B. Radha, A. Esfandiar, F. C. Wang, A. P. Rooney, K. Gopinadhan, A. Keerthi, A. Mishchenko, A. Janardanan, P. Blake, L. Fumagalli, M. Lozada-Hidalgo, S. Garaj, S. J. Haigh, I. V. Grigorieva, H. A. Wu, and A. K. Geim, "Molecular transport through capillaries made with atomic-scale precision," *Nature*, vol. 538, pp. 222–225, Oct 2016.

[166] P. A. Netz, F. W. Starr, H. E. Stanley, and M. C. Barbosa, "Static and dynamic properties of stretched water," *The Journal of Chemical Physics*, vol. 115, no. 1, pp. 344–348, 2001.

[167] M. M. Szortyka and M. C. Barbosa, "Diffusion anomaly in an associating lattice gas model," *Physica A: Statistical Mechanics and its Applications*, vol. 380, pp. 27–35, 2007.

[168] P. Kumar, S. V. Buldyrev, F. W. Starr, N. Giovambattista, and H. E. Stanley, "Thermodynamics, structure, and dynamics of water confined between hydrophobic plates," *Physical Review E*, vol. 72, p. 051503, Nov 2005.

[169] S. Han, P. Kumar, and H. E. Stanley, "Absence of a diffusion anomaly of water in the direction perpendicular to hydrophobic nanoconfining walls," *Physical Review E*, vol. 77, p. 030201, Mar 2008.

[170] G. Cicero, J. C. Grossman, E. Schwegler, F. Gygi, and G. Galli, "Water confined in nanotubes and between graphene sheets: a first principle study," *Journal of the American Chemical Society*, vol. 130, no. 6, pp. 1871–1878, 2008. PMID: 18211065.

[171] P. H. Poole, I. Saika-Voivod, and F. Sciortino, "Density minimum and liquid–liquid phase transition," *Journal of Physics: Condensed Matter*, vol. 17, pp. L431–L437, oct 2005.

[172] F. Smallenburg and F. Sciortino, "Tuning the liquid-liquid transition by modulating the hydrogen-bond angular flexibility in a model for water," *Physical Review Letters*, vol. 115, p. 015701, Jul 2015.

[173] J. L. F. Abascal and C. Vega, "Widom line and the liquid–liquid critical point for the TIP4P/2005 water model," *The Journal of Chemical Physics*, vol. 133, no. 23, p. 234502, 2010.

[174] V. Holten, D. T. Limmer, V. Molinero, and M. A. Anisimov, "Nature of the anomalies in the supercooled liquid state of the mw model of water," *The Journal of Chemical Physics*, vol. 138, no. 17, p. 174501, 2013.

[175] R. S. Singh, J. W. Biddle, P. G. Debenedetti, and M. A. Anisimov, "Two-state thermodynamics and the possibility of a liquid-liquid phase transition in

supercooled TIP4P/2005 water," *The Journal of Chemical Physics*, vol. 144, no. 14, p. 144504, 2016.

[176] D. A. Fuentevilla and M. A. Anisimov, "Scaled equation of state for supercooled water near the liquid-liquid critical point," *Physical Review Letters*, vol. 97, p. 195702, Nov 2006.

[177] V. Holten and M. A. Anisimov, "Entropy-driven liquid–liquid separation in supercooled water," *Scientific Reports*, vol. 2, p. 713, Oct 2012.

[178] V. Holten, J. C. Palmer, P. H. Poole, P. G. Debenedetti, and M. A. Anisimov, "Two-state thermodynamics of the ST2 model for supercooled water," *The Journal of Chemical Physics*, vol. 140, no. 10, p. 104502, 2014.

[179] J. Russo and H. Tanaka, "Understanding water's anomalies with locally favoured structures," *Nature Communications*, vol. 5, p. 3556, Apr 2014.

[180] G. Franzese and V. Bianco, "Water at biological and inorganic interfaces," *Food Biophysics*, vol. 8, pp. 153–169, Sep 2013.

[181] A. D. Bruce and N. B. Wilding, "Scaling fields and universality of the liquid-gas critical point," *Physical Review Letters*, vol. 68, pp. 193–196, Jan 1992.

[182] N. B. Wilding, "Critical-point and coexistence-curve properties of the Lennard-Jones fluid: A finite-size scaling study," *Physical Review E*, vol. 52, pp. 602–611, Jul 1995.

[183] P. Kumar, S. V. Buldyrev, S. R. Becker, P. H. Poole, F. W. Starr, and H. E. Stanley, "Relation between the Widom line and the breakdown of the Stokes-Einstein relation in supercooled water," *Proceedings of the National Academy of Sciences*, vol. 104, no. 23, pp. 9575–9579, 2007.

[184] V. Holten, C. E. Bertrand, M. A. Anisimov, and J. V. Sengers, "Thermodynamics of supercooled water," *The Journal of Chemical Physics*, vol. 136, no. 9, p. 094507, 2012.

[185] W.-X. Xu, J. Wang, and W. Wang, "Folding behavior of chaperonin-mediated substrate protein," *Proteins: Structure, Function, and Bioinformatics*, vol. 61, no. 4, pp. 777–794, 2005.

[186] G. G. Simeoni, T. Bryk, F. A. Gorelli, M. Krisch, G. Ruocco, M. Santoro, and T. Scopigno, "The Widom line as the crossover between liquid-like and gas-like behaviour in supercritical fluids," *Nature Physics*, vol. 6, pp. 503–507, Jul 2010.

[187] A. Taschin, P. Bartolini, R. Eramo, R. Righini, and R. Torre, “Evidence of two distinct local structures of water from ambient to supercooled conditions,” *Nature Communications*, vol. 4, p. 2401, Sep 2013.

[188] F. Mallamace, C. Corsaro, M. Broccio, C. Branca, N. González-Segredo, J. Spooren, S.-H. Chen, and H. E. Stanley, “NMR evidence of a sharp change in a measure of local order in deeply supercooled confined water,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 35, pp. 12725–12729, 2008.

[189] L. Xu, P. Kumar, S. V. Buldyrev, S.-H. Chen, P. H. Poole, F. Sciortino, and H. E. Stanley, “Relation between the Widom line and the dynamic crossover in systems with a liquid-liquid phase transition,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 46, pp. 16558–16562, 2005.

[190] P. Kasteleyn and C. M. Fortuin, “Phase Transitions in Lattice Systems with Random Local Properties,” *Physical Society of Japan Journal Supplement, Vol.˜26.˜ Proceedings of the International Conference on Statistical Mechanics held 9-14 September, 1968 in Koyto., p.11*, vol. 26, p. 11, 1969.

[191] A. Coniglio and W. Klein, “Clusters and Ising critical droplets: a renormalisation group approach,” *Journal of Physics A: Mathematical and General*, vol. 13, pp. 2775–2780, aug 1980.

[192] S. J. Suresh and V. M. Naik, “Hydrogen bond thermodynamic properties of water from dielectric constant data,” *The Journal of Chemical Physics*, vol. 113, no. 21, pp. 9727–9732, 2000.

[193] N. Chumaevskii and M. Rodnikova, “Some peculiarities of liquid water structure,” *Journal of Molecular Liquids*, vol. 106, no. 2, pp. 167–177, 2003. Contribution to the Seminar on the Structure of Liquids and Liquid Solutions of the Russian Academy of Sciences in the Honour of Professor O. Ya. Samoilov, Moscow 2001.

[194] K. Kremer and K. Binder, “Monte Carlo simulation of lattice models for macromolecules,” *Computer Physics Reports*, vol. 7, no. 6, pp. 259–310, 1988.

[195] T. Wang, J. Miller, N. S. Wingreen, C. Tang, and K. A. Dill, “Symmetry and designability for lattice protein models,” *The Journal of Chemical Physics*, vol. 113, no. 18, pp. 8329–8336, 2000.

[196] L. M. Contreras Martínez, F. J. Martínez-Veracoechea, P. Pohkarel, A. D. Stroock, F. A. Escobedo, and M. P. DeLisa, “Protein translocation through a tunnel induces changes in folding kinetics: A lattice model study,” *Biotechnology and Bioengineering*, vol. 94, no. 1, pp. 105–117, 2006.

[197] S. Matysiak, P. G. Debenedetti, and P. J. Rossky, “Role of Hydrophobic Hydration in Protein Stability: A 3D Water-Explicit Protein Model Exhibiting Cold and Heat Denaturation,” *The Journal of Physical Chemistry B*, vol. 116, no. 28, pp. 8095–8104, 2012. PMID: 22725973.

[198] S. Dutta, Y. Lee, and Y. S. Jho, “Hydration of ions in two-dimensional water,” *Physical Review E*, vol. 92, p. 042152, Oct 2015.

[199] I. P. Sugár and P. L.-G. Chong, “A statistical mechanical model of cholesterol/phospholipid mixtures: Linking condensed complexes, superlattices, and the phase diagram,” *Journal of the American Chemical Society*, vol. 134, no. 2, pp. 1164–1171, 2012. PMID: 22196210.

[200] D. S. Goodsell, L. Autin, and A. J. Olson, “Lattice models of bacterial nucleoids,” *The Journal of Physical Chemistry B*, vol. 122, no. 21, pp. 5441–5447, 2018.

[201] S. Miyazawa and R. L. Jernigan, “Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation,” *Macromolecules*, vol. 18, no. 3, pp. 534–552, 1985.

[202] J. Kyte and R. F. Doolittle, “A simple method for displaying the hydropathic character of a protein,” *Journal of Molecular Biology*, vol. 157, no. 1, pp. 105–132, 1982.

[203] C. L. Dias, T. Ala-Nissila, M. Karttunen, I. Vattulainen, and M. Grant, “Microscopic mechanism for cold denaturation,” *Physical Review Letters*, vol. 100, p. 118101, Mar 2008.

[204] C. Petersen, K.-J. Tielrooij, and H. J. Bakker, “Strong temperature dependence of water reorientation in hydrophobic hydration shells,” *The Journal of Chemical Physics*, vol. 130, no. 21, p. 214511, 2009.

[205] S. Sarupria and S. Garde, “Quantifying water density fluctuations and compressibility of hydration shells of hydrophobic solutes and proteins,” *Physical Review Letters*, vol. 103, p. 037803, Jul 2009.

[206] Y. I. Tarasevich, “State and structure of water in vicinity of hydrophobic surfaces,” *Colloid Journal*, vol. 73, pp. 257–266, Apr 2011.

[207] J. G. Davis, K. P. Gierszal, P. Wang, and D. Ben-Amotz, “Water structural transformation at molecular hydrophobic interfaces,” *Nature*, vol. 491, pp. 582–585, Nov 2012.

[208] N. Giovambattista, P. J. Rossky, and P. G. Debenedetti, "Effect of pressure on the phase behavior and structure of water confined between nanoscale hydrophobic and hydrophilic plates," *Physical Review E*, vol. 73, p. 041604, Apr 2006.

[209] N. Muller, "Search for a realistic view of hydrophobic effects," *Accounts of Chemical Research*, vol. 23, no. 1, pp. 23–28, 1990.

[210] P. Das and S. Matysiak, "Direct characterization of hydrophobic hydration during cold and pressure denaturation," *The Journal of Physical Chemistry B*, vol. 116, no. 18, pp. 5342–5348, 2012. PMID: 22512347.

[211] T. Ghosh, A. E. García, and S. Garde, "Molecular dynamics simulations of pressure effects on hydrophobic interactions," *Journal of the American Chemical Society*, vol. 123, no. 44, pp. 10997–11003, 2001. PMID: 11686704.

[212] C. L. Dias and H. S. Chan, "Pressure-dependent properties of elementary hydrophobic interactions: Ramifications for activation properties of protein folding," *The Journal of Physical Chemistry B*, vol. 118, no. 27, pp. 7488–7509, 2014. PMID: 24933471.

[213] S. A. Hawley, "Reversible pressure-temperature denaturation of chymotrypsinogen," *Biochemistry*, vol. 10, no. 13, pp. 2436–2442, 1971. PMID: 5557794.

[214] L. Smeller, "Pressure–temperature phase diagrams of biomolecules," *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, vol. 1595, no. 1, pp. 11–29, 2002.

[215] R. Ravindra and R. Winter, "On the temperature–pressure free-energy landscape of proteins," *ChemPhysChem*, vol. 4, no. 4, pp. 359–365, 2003.

[216] A. Pastore, S. R. Martin, A. Politou, K. C. Kondapalli, T. Stemmler, and P. A. Temussi, "Unbiased cold denaturation: low- and high-temperature unfolding of yeast frataxin under physiological conditions," *Journal of the American Chemical Society*, vol. 129, no. 17, pp. 5374–5375, 2007. PMID: 17411056.

[217] T. Sumi and H. Sekino, "Possible mechanism underlying high-pressure unfolding of proteins: formation of a short-period high-density hydration shell," *Physical Chemistry Chemical Physics*, vol. 13, pp. 15829–15832, 2011.

[218] N. V. Nucci, B. Fuglestad, E. A. Athanasoula, and A. J. Wand, "Role of cavities and hydration in the pressure unfolding of T_4 lysozyme," *Proceedings of the National Academy of Sciences*, vol. 111, no. 38, pp. 13846–13851, 2014.

[219] D. Sanfelice, A. Politou, S. R. Martin, P. De Los Rios, P. Temussi, and A. Pastore, “The effect of crowding and confinement: a comparison of Yfh1 stability in different environments,” *Physical Biology*, vol. 10, p. 045002, aug 2013.

[220] E. Larios and M. Gruebele, “Protein stability at negative pressure,” *Methods*, vol. 52, no. 1, pp. 51–56, 2010. Protein Folding.

[221] J. Roche, J. A. Caro, D. R. Norberto, P. Barthe, C. Roumestand, J. L. Schlessman, A. E. Garcia, B. G.-M. E., and C. A. Royer, “Cavities determine the pressure unfolding of proteins,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 18, pp. 6945–6950, 2012.

[222] P. L. Privalov, “Cold denaturation of protein,” *Critical Reviews in Biochemistry and Molecular Biology*, vol. 25, no. 4, pp. 281–306, 1990. PMID: 2225910.

[223] B. J. Sirovetz, N. P. Schafer, and P. G. Wolynes, “Water mediated interactions and the protein folding phase diagram in the temperature–pressure plane,” *The Journal of Physical Chemistry B*, vol. 119, pp. 11416–11427, Aug 2015.

[224] I. Samish, C. M. MacDermaid, J. M. Perez-Aguilar, and J. G. Saven, “Theoretical and computational protein design,” *Annual Review of Physical Chemistry*, vol. 62, no. 1, pp. 129–149, 2011. PMID: 21128762.

[225] I. Coluzza, H. G. Muller, and D. Frenkel, “Designing refoldable model molecules,” *Physical Review E*, vol. 68, p. 046703, Oct 2003.

[226] I. Coluzza, “A coarse-grained approach to protein design: Learning from design to understand folding,” *PLOS ONE*, vol. 6, pp. 1–8, 07 2011.

[227] L. Lins, A. Thomas, and R. Brasseur, “Analysis of accessible surface of residues in proteins,” *Protein Science*, vol. 12, no. 7, pp. 1406–1417, 2003.

[228] S. Moelbert, E. Emberly, and C. Tang, “Correlation between sequence hydrophobicity and surface-exposure pattern of database proteins,” *Protein Science*, vol. 13, no. 3, pp. 752–762, 2004.

[229] M. Mahmoudi, H. R. Kalhor, S. Laurent, and I. Lynch, “Protein fibrillation and nanoparticle interactions: opportunities and challenges,” *Nanoscale*, vol. 5, pp. 2570–2588, 2013.

[230] T. P. J. Knowles, M. Vendruscolo, and C. M. Dobson, “The amyloid state and its association with protein misfolding diseases,” *Nature Reviews Molecular Cell Biology*, vol. 15, pp. 384–396, Jun 2014.

[231] C. Mathieu, R. V. Pappu, and J. P. Taylor, “Beyond aggregation: Pathological phase transitions in neurodegenerative disease,” *Science*, vol. 370, no. 6512, pp. 56–60, 2020.

[232] F. Sterpone, S. Melchionna, P. Tuffery, S. Pasquali, N. Mousseau, T. Cragnolini, Y. Chebaro, J.-F. St-Pierre, M. Kalimeri, A. Barducci, Y. Laurin, A. Tek, M. Baaden, P. H. Nguyen, and P. Derreumaux, “The OPEP protein model: from single molecules, amyloid formation, crowding and hydrodynamics to DNA/RNA systems,” *Chemical Society Reviews*, vol. 43, pp. 4871–4893, 2014.

[233] F. Sterpone, P. Derreumaux, and S. Melchionna, “Protein Simulations in Fluids: Coupling the OPEP Coarse-Grained Force Field with Hydrodynamics,” *Journal of Chemical Theory and Computation*, vol. 11, no. 4, pp. 1843–1853, 2015. PMID: 26574390.

[234] A. V. Onufriev and S. Izadi, “Water models for biomolecular simulations,” *WIREs Computational Molecular Science*, vol. 8, no. 2, p. e1347, 2018.

[235] G. Klesse, S. Rao, S. J. Tucker, and M. S. Sansom, “Induced Polarization in Molecular Dynamics Simulations of the 5-HT3 Receptor Channel,” *Journal of the American Chemical Society*, vol. 142, no. 20, pp. 9415–9427, 2020. PMID: 32336093.

[236] NVIDIA, *Cuda C Programming Guide. Version 11.6.* February 2022.

[237] K. A. Hawick, A. Leist, and D. P. Playne, “Regular Lattice and Small-World Spin Model Simulations Using CUDA and GPUs,” *International Journal of Parallel Programming*, vol. 39, no. 2, pp. 183–201, 2011.

[238] M. Weigel and T. Yavorskii, “GPU accelerated Monte Carlo simulations of lattice spin models,” *Physics Procedia*, vol. 15, pp. 92–96, 2011. Proceedings of the 24th Workshop on Computer Simulation Studies in Condensed Matter Physics (CSP2011).

[239] Y. Komura and Y. Okabe, “GPU-based Swendsen–Wang multi-cluster algorithm for the simulation of two-dimensional classical spin systems,” *Computer Physics Communications*, vol. 183, no. 6, pp. 1155–1161, 2012.

[240] C. Hall, W. Ji, and E. Blaisten-Barojas, “The Metropolis Monte Carlo method with CUDA enabled Graphic Processing Units,” *Journal of Computational Physics*, vol. 258, pp. 871–879, 2014.

[241] M. Januszewski and M. Kostur, “Accelerating numerical solution of stochastic differential equations with CUDA,” *Computer Physics Communications*, vol. 181, no. 1, pp. 183–188, 2010.

[242] J. Spiechowicz, M. Kostur, and L. Machura, “GPU accelerated Monte Carlo simulation of Brownian motors dynamics with CUDA,” *Computer Physics Communications*, vol. 191, pp. 140–149, 2015.

[243] C. A. Cerdeiriña, J. Troncoso, D. González-Salgado, P. G. Debenedetti, and H. E. Stanley, “Water’s two-critical-point scenario in the Ising paradigm,” *The Journal of Chemical Physics*, vol. 150, no. 24, p. 244509, 2019.

[244] R. Diestel, *Graph Theory.* Springer-Verlag, 5th ed., 2017.

[245] M. Wild, “Generating all cycles, chordless cycles, and Hamiltonian cycles with the principle of exclusion,” *Journal of Discrete Algorithms*, vol. 6, no. 1, pp. 93–102, 2008. Selected papers from AWOCA 2005.

[246] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of State Calculations by Fast Computing Machines,” *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[247] R. H. Swendsen and J.-S. Wang, “Nonuniversal critical dynamics in Monte Carlo simulations,” *Physical Review Letters*, vol. 58, pp. 86–88, Jan 1987.

[248] G. T. Barkema and T. MacFarland, “Parallel simulation of the Ising model,” *Physical Review E*, vol. 50, pp. 1623–1628, Aug 1994.

[249] D. W. Heermann and A. N. Burkitt, “Parallelization of the Ising model and its performance evaluation,” *Parallel Computing*, vol. 13, no. 3, pp. 345–357, 1990.

[250] D. Hassani and S. Rafibakhsh, “Parallelization and implementation of multi-spin Monte Carlo simulation of 2D square Ising model using MPI and C++,” *Journal of Theoretical and Applied Physics*, vol. 12, no. 3, pp. 199–208, 2018.

[251] J. Wojtkiewicz and K. Kalinowski, “Monte Carlo Simulations of the Ising Model on GPU,” *Computational Methods in Science and Technology*, vol. 21, no. 2, pp. 69–98, 2015.

[252] J. J. Tapia and R. M. D’Souza, “Parallelizing the Cellular Potts Model on graphics processing units,” *Computer Physics Communications*, vol. 182, no. 4, pp. 857–865, 2011.

[253] A. Leist, D. P. Playne, and K. A. Hawick, “Exploiting graphical processing units for data-parallel scientific applications,” *Concurrency and Computation: Practice and Experience*, vol. 21, no. 18, pp. 2400–2437, 2009.

[254] J. Sanders and E. Kandrot, *CUDA by Example: An Introduction to General Purpose Graphical Processing Unit.* Addison-Wesley, 1st ed., 2010.

[255] J. Hoshen and R. Kopelman, “Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm,” *Physical Review B*, vol. 14, pp. 3438–3445, Oct 1976.

[256] M. Flanigan and P. Tamayo, “Parallel cluster labeling for large-scale Monte Carlo simulations,” *Physica A: Statistical Mechanics and its Applications*, vol. 215, no. 4, pp. 461–480, 1995.

[257] K. Hawick, A. Leist, and D. Playne, “Parallel graph component labelling with GPUs and CUDA,” *Parallel Computing*, vol. 36, no. 12, pp. 655–678, 2010.

[258] O. Kalentev, A. Rai, S. Kemnitz, and R. Schneider, “Connected component labeling on a 2D grid using CUDA,” *Journal of Parallel and Distributed Computing*, vol. 71, no. 4, pp. 615–620, 2011.

[259] A. D. Stephens and G. S. Kaminski Schierle, “The role of water in amyloid aggregation kinetics,” *Current Opinion in Structural Biology*, vol. 58, pp. 115–123, oct 2019.

[260] P. L. Clark, K. W. Plaxco, and T. R. Sosnick, “Water as a good solvent for unfolded proteins: Folding and collapse are fundamentally different,” *Journal of Molecular Biology*, vol. 432, no. 9, pp. 2882–2889, 2020. Integrative Biophysics: Protein Interaction and Disorder.

[261] F. Martelli, C. Calero, and G. Franzese, “Redefining the concept of hydration water near soft interfaces,” *Biointerphases*, vol. 16, no. 2, p. 020801, 2021.

[262] A. Rajabpour, R. Seif, S. Arabha, M. M. Heyhat, S. Merabia, and A. Hassanali, “Thermal transport at a nanoparticle-water interface: A molecular dynamics and continuum modeling study,” *The Journal of Chemical Physics*, vol. 150, no. 11, p. 114701, 2019.

[263] A. Gotzias, “Umbrella sampling simulations of carbon nanoparticles crossing immiscible solvents,” *Molecules*, vol. 27, no. 3, 2022.

[264] E. Brini, C. Simmerling, and K. Dill, “Protein storytelling through physics,” *Science*, vol. 370, no. 6520, p. eaaz3041, 2020.

[265] O. Collet and C. Chipot, “Non-Arrhenius Behavior in the Unfolding of a Short, Hydrophobic α-Helix. Complementarity of Molecular Dynamics and Lattice Model Simulations,” *Journal of the American Chemical Society*, vol. 125, no. 21, pp. 6573–6580, 2003. PMID: 12785798.

[266] R. Z. Khaliullin and T. D. Kühne, “Microscopic properties of liquid water from combined ab initio molecular dynamics and energy decomposition studies,” *Physical Chemistry Chemical Physics*, vol. 15, no. 38, pp. 15746–15766, 2013.

[267] E. A. Cobar, P. R. Horn, R. G. Bergman, and M. Head-Gordon, “Examination of the hydrogen-bonding networks in small water clusters (n = 2–5, 13, 17) using absolutely localized molecular orbital energy decomposition analysis,” *Physical Chemistry Chemical Physics*, vol. 14, pp. 15328–15339, 2012.

[268] H. Elgabarty, N. K. Kaliannan, and T. D. Kühne, “Enhancement of the local asymmetry in the hydrogen bond network of liquid water by an ultrafast electric field pulse,” *Scientific Reports*, vol. 9, no. 1, p. 1002, 2019.

[269] D. Ojha, N. K. Kaliannan, and T. D. Kühne, “Time-dependent vibrational sum-frequency generation spectroscopy of the air-water interface,” *Communications Chemistry*, vol. 2, no. 1, p. 116, 2019.

[270] D. Ojha and T. D. Kühne, “Hydrogen bond dynamics of interfacial water molecules revealed from two-dimensional vibrational sum-frequency generation spectroscopy,” *Scientific Reports*, vol. 11, no. 1, p. 2456, 2021.

[271] V. Balos, N. K. Kaliannan, H. Elgabarty, M. Wolf, T. D. Kühne, and M. Sajadi, “Time-resolved terahertz–Raman spectroscopy reveals that cations and anions distinctly modify intermolecular interactions of water,” *Nature Chemistry*, vol. 14, pp. 1031–1037, Sep 2022.

[272] M. Chaplin, “Water structure and science.” Available at `https://water.lsbu.ac.uk/water/` Last updated 4 Jan 2022.

[273] G. Pallares, M. A. Gonzalez, J. L. F. Abascal, C. Valeriani, and F. Caupin, “Equation of state for water and its line of density maxima down to -120 MPa,” *Physical Chemistry Chemical Physics*, vol. 18, pp. 5896–5900, 2016.

[274] O. Mishima, “Volume of supercooled water under pressure and the liquid-liquid critical point,” *The Journal of Chemical Physics*, vol. 133, no. 14, p. 144503, 2010.

[275] D. E. Hare and C. M. Sorensen, “The density of supercooled water. II. Bulk samples cooled to the homogeneous nucleation limit,” *The Journal of Chemical Physics*, vol. 87, no. 8, pp. 4840–4845, 1987.

[276] T. Sotani, J. Arabas, H. Kubota, M. Kijima, and S. Asada, “Volumetric behaviour of water under high pressure at subzero temperature,” *High Temperatures - High Pressures*, vol. 32, no. 4, pp. 433–440, 2000.

[277] F. Caupin and M. A. Anisimov, “Thermodynamics of supercooled and stretched water: Unifying two-structure description and liquid-vapor spinodal,” *The Journal of Chemical Physics*, vol. 151, no. 3, p. 034503, 2019.

[278] A. K. Soper, “Density minimum in supercooled confined water,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 47, pp. E1192–E1192, 2011.

[279] G. Pallares, M. El Mekki Azouzi, M. A. González, J. L. Aragones, J. L. Abascal, C. Valeriani, and F. Caupin, “Anomalies in bulk supercooled water at negative pressure,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 22, pp. 7936–7941, 2014.

[280] V. Holten, C. Qiu, E. Guillerm, M. Wilke, J. Rička, M. Frenz, and F. Caupin, “Compressibility anomalies in stretched water and their interplay with density anomalies,” *The Journal of Physical Chemistry Letters*, vol. 8, no. 22, pp. 5519–5522, 2017. PMID: 29043801.

[281] C. Chen, R. A. Fine, and F. J. Millero, “The equation of state of pure water determined from sound speeds,” *The Journal of Chemical Physics*, vol. 66, no. 5, pp. 2142–2144, 1977.

[282] S. Asada, T. Sotani, J. Arabas, H. Kubota, S. Matsuo, and Y. Tanaka, “Density of water at subzero temperature under high pressure: measurements and correlation,” *Journal of Physics: Condensed Matter*, vol. 14, pp. 11447–11452, oct 2002.

[283] H. Kanno and C. A. Angell, “Water: Anomalous compressibilities to 1.9 kbar and correlation with supercooling limits,” *The Journal of Chemical Physics*, vol. 70, no. 9, pp. 4008–4016, 1979.

[284] G. S. Kell, “Density, thermal expansivity, and compressibility of liquid water from 0.deg. to 150.deg.. correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale,” *Journal of Chemical & Engineering Data*, vol. 20, no. 1, pp. 97–105, 1975.

[285] L. Ter Minassian, P. Pruzan, and A. Soulard, “Thermodynamic properties of water under pressure up to 5 kbar and between 28 and 120 °C. Estimations in the supercooled region down to -40 ° C,” *The Journal of Chemical Physics*, vol. 75, no. 6, pp. 3064–3072, 1981.

[286] R. A. Fine and F. J. Millero, “Compressibility of water as a function of temperature and pressure,” *The Journal of Chemical Physics*, vol. 59, no. 10, pp. 5529–5536, 1973.

[287] F. Mallamace, C. Corsaro, D. Mallamace, S. Vasi, C. Vasi, and H. E. Stanley, “Thermodynamic properties of bulk and confined water,” *The Journal of Chemical Physics*, vol. 141, no. 18, p. 18C504, 2014.

[288] C.-W. Lin and J. P. M. Trusler, “The speed of sound and derived thermodynamic properties of pure water at temperatures between (253 and 473) K and at pressures up to 400 MPa,” *The Journal of Chemical Physics*, vol. 136, no. 9, p. 094511, 2012.

[289] A. Sirota, A. Grishkov, and T. A.G. *Thermal Engeneering*, vol. 17, no. 9, p. 90, 1970.

[290] M. Duška, “Water above the spinodal,” *The Journal of Chemical Physics*, vol. 152, no. 17, p. 174501, 2020.

[291] D. G. Archer and R. W. Carter, “Thermodynamic Properties of the NaCl + H2O System. 4. Heat Capacities of H2O and NaCl(aq) in Cold-Stable and Supercooled States,” *The Journal of Physical Chemistry B*, vol. 104, no. 35, pp. 8563–8584, 2000.

[292] “How deep is the ocean?.” Available at `https://www.nhm.ac.uk/discover/quick-questions/how-deep-is-the-ocean.html` Last accessed: 5 Dec 2022.

[293] I. Stepanov, “Anomalous dependence of the heat capacity of supercooled water on pressure and temperature,” *Results in Physics*, vol. 4, pp. 28–30, 2014.

[294] P. G. Debenedetti, *Metastable Liquids. Concepts and Principles.* Princeton, NJ: Princeton University Press, 1996.

[295] T. Loerting, C. Salzmann, I. Kohl, E. Mayer, and A. Hallbrucker, “A second distinct structural “state” of high-density amorphous ice at 77 K and 1 bar,” *Physical Chemistry Chemical Physics*, vol. 3, pp. 5355–5357, 2001.

[296] G. Shen, J. S. Smith, and C. Kenney-Benson, “Nature of polyamorphic transformations in H_2O under isothermal compression and decompression,” *Physical Review Materials*, vol. 3, p. 073404, Jul 2019.

[297] E. M. Mollica, J. Russo, H. E. Stanley, and F. Sciortino, “Decompression dynamics of high density amorphous ice above and below the liquid-liquid critical point,” *Journal of Non-Crystalline Solids: X*, vol. 13, p. 100081, 2022.

[298] C. U. Kim, M. W. Tate, and S. M. Gruner, “Glass-to-cryogenic-liquid transitions in aqueous solutions suggested by crack healing,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 38, pp. 11765–11770, 2015.

[299] H. Tanaka, “Liquid–liquid transition and polyamorphism,” *The Journal of Chemical Physics*, vol. 153, no. 13, p. 130901, 2020.

[300] K. Winkel, E. Mayer, and T. Loerting, “Equilibrated high-density amorphous ice and its first-order transition to the low-density form,” *The Journal of Physical Chemistry B*, vol. 115, no. 48, pp. 14141–14148, 2011. PMID: 21793514.

[301] F. Perakis, K. Amann-Winkel, F. Lehmkühler, M. Sprung, D. Mariedahl, J. A. Sellberg, H. Pathak, A. Späh, F. Cavalca, D. Schlesinger, A. Ricci, A. Jain, B. Massani, F. Aubree, C. J. Benmore, T. Loerting, G. Grübel, L. G. M. Pettersson, and A. Nilsson, “Diffusive dynamics during the high-to-low density transition in amorphous ice,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 31, pp. 8193–8198, 2017.

[302] Y. Liu, A. Z. Panagiotopoulos, and P. G. Debenedetti, “Low-temperature fluid-phase behavior of ST2 water,” *The Journal of Chemical Physics*, vol. 131, no. 10, p. 104508, 2009.

[303] Y. Goswami and S. Sastry, “Liquid–liquid phase transition in deeply supercooled Stillinger–Weber silicon,” *PNAS Nexus*, vol. 1, 09 2022. pgac204.

[304] Y. Liu, A. Z. Panagiotopoulos, and P. G. Debenedetti, “Finite-size scaling study of the vapor-liquid critical properties of confined fluids: Crossover from three dimensions to two dimensions,” *The Journal of Chemical Physics*, vol. 132, no. 14, p. 144107, 2010.

[305] U. Wolff, “Collective Monte Carlo Updating for Spin Systems,” *Physical Review Letters*, vol. 62, pp. 361–364, Jan 1989.

[306] A. Z. Panagiotopoulos, “Monte Carlo methods for phase equilibria of fluids,” *Journal of Physics: Condensed Matter*, vol. 12, p. R25, jan 2000.

[307] S. V. Buldyrev, G. Malescio, C. A. Angell, N. Giovambattista, S. Prestipino, F. Saija, H. E. Stanley, and L. Xu, “Unusual phase behavior of one-component systems with two-scale isotropic interactions,” *Journal of Physics: Condensed Matter*, vol. 21, p. 504106, nov 2009.

[308] A. Cupane, M. Fomina, I. Piazza, J. Peters, and G. Schirò, “Experimental evidence for a liquid-liquid crossover in deeply cooled confined water,” *Physical Review Letters*, vol. 113, p. 215701, Nov 2014.

[309] F. Mallamace, C. Corsaro, D. Mallamace, E. Fazio, S.-H. Chen, and A. Cupane, “Specific heat and transport functions of water,” *International Journal of Molecular Sciences*, vol. 21, no. 2, 2020.

[310] H. Pathak, J. C. Palmer, D. Schlesinger, K. T. Wikfeldt, J. A. Sellberg, L. G. M. Pettersson, and A. Nilsson, “The structural validity of various thermodynamical

models of supercooled water," *The Journal of Chemical Physics*, vol. 145, no. 13, p. 134507, 2016.

[311] J. L. F. Abascal, E. Sanz, R. García Fernández, and C. Vega, "A potential model for the study of ices and amorphous water: TIP4P/Ice," *The Journal of Chemical Physics*, vol. 122, no. 23, p. 234511, 2005.

[312] Y. Ni and J. L. Skinner, "Evidence for a liquid-liquid critical point in supercooled water within the E3B3 model and a possible interpretation of the kink in the homogeneous nucleation line," *The Journal of Chemical Physics*, vol. 144, no. 21, p. 214501, 2016.

[313] G. Franzese and A. Coniglio, "Phase transitions in the potts spin-glass model," *Physical Review E*, vol. 58, pp. 2753–2759, Sep 1998.

[314] S. Piana, A. G. Donchev, P. Robustelli, and D. E. Shaw, "Water dispersion interactions strongly influence simulated structural properties of disordered protein states," *The Journal of Physical Chemistry B*, vol. 119, no. 16, pp. 5113–5123, 2015. PMID: 25764013.

[315] A. Emperador, R. Crehuet, and E. Guàrdia, "Effect of the water model in simulations of protein–protein recognition and association," *Polymers*, vol. 13, no. 2, 2021.

[316] R. T. Bradshaw, J. Dziedzic, C.-K. Skylaris, and J. W. Essex, "The role of electrostatics in enzymes: Do biomolecular force fields reflect protein electric fields?," *Journal of Chemical Information and Modeling*, vol. 60, no. 6, pp. 3131–3144, 2020. PMID: 32298113.

[317] A.-h. Wang, Z.-c. Zhang, and G.-h. Li, "Advances in enhanced sampling molecular dynamics simulations for biomolecules," *Chinese Journal of Chemical Physics*, vol. 32, no. 3, pp. 277–286, 2019.

[318] M. Chen, "Collective variable-based enhanced sampling and machine learning," *The European Physical Journal B*, vol. 94, p. 211, Oct 2021.

[319] J. Vymětal and J. Vondrášek, "Iterative landmark-based umbrella sampling (ilbus) protocol for sampling of conformational space of biomolecules," *Journal of Chemical Information and Modeling*, vol. 62, no. 19, pp. 4783–4798, 2022. PMID: 36122323.

[320] D. Jiang, T. M. Gamal El-Din, C. Ing, P. Lu, R. Pomès, N. Zheng, and W. A. Catterall, "Structural basis for gating pore current in periodic paralysis," *Nature*, vol. 557, pp. 590–594, May 2018.

[321] A. Srivastava, F. Tama, D. Kohda, and O. Miyashita, “Computational investigation of the conformational dynamics in Tom20-mitochondrial presequence tethered complexes,” *Proteins: Structure, Function, and Bioinformatics*, vol. 87, no. 1, pp. 81–90, 2019.

[322] R. D. Hills, Jr, L. Lu, and G. A. Voth, “Multiscale coarse-graining of the protein energy landscape,” *PLOS Computational Biology*, vol. 6, pp. 1–12, 06 2010.

[323] V. V. H. Giri Rao, R. Desikan, K. G. Ayappa, and S. Gosavi, “Capturing the membrane-triggered conformational transition of an α-helical pore-forming toxin,” *The Journal of Physical Chemistry B*, vol. 120, no. 47, pp. 12064–12078, 2016. PMID: 27797514.

[324] C. De Michele, P. De Los Rios, G. Foffi, and F. Piazza, “Simulation and Theory of Antibody Binding to Crowded Antigen-Covered Surfaces,” *PLOS Computational Biology*, vol. 12, pp. 1–17, 03 2016.

[325] S. Timr, D. Gnutt, S. Ebbinghaus, and F. Sterpone, “The Unfolding Journey of Superoxide Dismutase 1 Barrels under Crowding: Atomistic Simulations Shed Light on Intermediate States and Their Interactions with Crowders,” *The Journal of Physical Chemistry Letters*, vol. 11, no. 10, pp. 4206–4212, 2020. PMID: 32364389.

[326] S. Timr and F. Sterpone, “Stabilizing or Destabilizing: Simulations of Chymotrypsin Inhibitor 2 under Crowding Reveal Existence of a Crossover Temperature,” *The Journal of Physical Chemistry Letters*, vol. 12, no. 6, pp. 1741–1746, 2021. PMID: 33570420.

[327] O. Languin-Cattoën, E. Laborie, D. O. Yurkova, S. Melchionna, P. Derreumaux, A. V. Belyaev, and F. Sterpone, “Exposure of Von Willebrand Factor Cleavage Site in A1A2A3-Fragment under Extreme Hydrodynamic Shear,” *Polymers*, vol. 13, no. 22, 2021.

[328] S. L. Haberichter, M. Balistreri, P. Christopherson, P. Morateck, S. Gavazova, D. B. Bellissimo, M. J. Manco-Johnson, J. C. Gill, and R. R. Montgomery, “Assay of the von Willebrand factor (VWF) propeptide to identify patients with type 1 von Willebrand disease with decreased VWF survival,” *Blood*, vol. 108, no. 10, pp. 3344–3351, 2006.

[329] S. Mondal, K. N. Hollander, S. O. Ibekwe, B. Williams, and K. Tanaka, “Heyde syndrome–pathophysiology and perioperative implications,” *Journal of Cardiothoracic and Vascular Anesthesia*, vol. 35, no. 11, pp. 3331–3339, 2021.

[330] T. A. Springer, “von Willebrand factor, Jedi knight of the bloodstream,” *Blood*, vol. 124, no. 9, pp. 1412–1425, 2014.

[331] H. Fu, Y. Jiang, D. Yang, F. Scheiflinger, W. P. Wong, and T. A. Springer, "Flow-induced elongation of von Willebrand factor precedes tension-dependent activation," *Nature Communications*, vol. 8, p. 324, Aug 2017.

[332] A. Löf, P. U. Walker, S. M. Sedlak, S. Gruber, T. Obser, M. A. Brehm, M. Benoit, and J. Lipfert, "Multiplexed protein force spectroscopy reveals equilibrium protein folding dynamics and the low-force response of von Willebrand factor," *Proceedings of the National Academy of Sciences*, vol. 116, no. 38, pp. 18798–18807, 2019.

[333] M. J. Morabito, M. Usta, X. Cheng, X. F. Zhang, A. Oztekin, and E. B. Webb, "Prediction of Sub-Monomer A2 Domain Dynamics of the von Willebrand Factor by Machine Learning Algorithm and Coarse-Grained Molecular Dynamics Simulation," *Scientific Reports*, vol. 9, p. 9037, Jun 2019.

[334] K. Rack, V. Huck, M. Hoore, D. A. Fedosov, S. W. Schneider, and G. Gompper, "Margination and stretching of von Willebrand factor in the blood stream enable adhesion," *Scientific Reports*, vol. 7, p. 14278, Oct 2017.

[335] A. V. Belyaev, "Intradimer forces and their implication for conformations of von Willebrand factor multimers," *Biophysical Journal*, vol. 120, no. 5, pp. 899–911, 2021.

[336] S. Okhota, I. Melnikov, Y. Avtaeva, S. Kozlov, and Z. Gabbasov, "Shear Stress-Induced Activation of von Willebrand Factor and Cardiovascular Pathology," *International Journal of Molecular Sciences*, vol. 21, no. 20, 2020.

[337] J. T. B. Crawley, R. de Groot, Y. Xiang, B. M. Luken, and D. A. Lane, "Unraveling the scissile bond: how ADAMTS13 recognizes and cleaves von Willebrand factor," *Blood*, vol. 118, pp. 3212–3221, 09 2011.

[338] T. Wu, J. Lin, M. A. Cruz, J.-f. Dong, and C. Zhu, "Force-induced cleavage of single VWFA1A2A3 tridomains by ADAMTS-13," *Blood*, vol. 115, pp. 370–378, 01 2010.

[339] Z. Li, J. Lin, T. Sulchek, M. A. Cruz, J. Wu, J.-f. Dong, and C. Zhu, "Domain-specific mechanical modulation of VWF–ADAMTS13 interaction," *Molecular Biology of the Cell*, vol. 30, no. 16, pp. 1920–1929, 2019. PMID: 31067148.

[340] T. Yago, J. Lou, T. Wu, J. Yang, J. J. Miner, L. Coburn, J. A. López, M. A. Cruz, J.-F. Dong, L. V. McIntire, R. P. McEver, and C. Zhu, "Platelet glycoprotein Ibα forms catch bonds with human WT vWF but not with type 2B von Willebrand disease vWF," *The Journal of Clinical Investigation*, vol. 118, pp. 3195–3207, 9 2008.

[341] Q. Zhang, Y.-F. Zhou, C.-Z. Zhang, X. Zhang, C. Lu, and T. A. Springer, "Structural specializations of A2, a force-sensing domain in the ultralarge vascular protein von Willebrand factor," *Proceedings of the National Academy of Sciences*, vol. 106, no. 23, pp. 9226–9231, 2009.

[342] Y.-F. Zhou, E. T. Eng, N. Nishida, C. Lu, T. Walz, and T. A. Springer, "A pH-regulated dimeric bouquet in the structure of von Willebrand factor," *The EMBO Journal*, vol. 30, no. 19, pp. 4098–4111, 2011.

[343] C. Aponte-Santamaría, V. Huck, S. Posch, A. Bronowska, S. Grässle, M. Brehm, T. Obser, R. Schneppenheim, P. Hinterdorfer, S. Schneider, C. Baldauf, and F. Gräter, "Force-Sensitive Autoinhibition of the von Willebrand Factor Is Mediated by Interdomain Interactions," *Biophysical Journal*, vol. 108, pp. 2312–2321, May 2015.

[344] Y. Chebaro, S. Pasquali, and P. Derreumaux, "The Coarse-Grained OPEP Force Field for Non-Amyloid and Amyloid Proteins," *The Journal of Physical Chemistry B*, vol. 116, no. 30, pp. 8741–8752, 2012. PMID: 22742737.

[345] F. Sterpone, P. H. Nguyen, M. Kalimeri, and P. Derreumaux, "Importance of the Ion-Pair Interactions in the OPEP Coarse-Grained Force Field: Parametrization and Validation," *Journal of Chemical Theory and Computation*, vol. 9, no. 10, pp. 4574–4584, 2013.

[346] M. Kalimeri, P. Derreumaux, and F. Sterpone, "Are coarse-grained models apt to detect protein thermal stability? The case of OPEP force field," *Journal of Non-Crystalline Solids*, vol. 407, pp. 494–501, 2015. 7th IDMRCS: Relaxation in Complex Systems.

[347] M. Chiricotto, S. Melchionna, P. Derreumaux, and F. Sterpone, "Hydrodynamic effects on β-amyloid (16-22) peptide aggregation," *The Journal of Chemical Physics*, vol. 145, no. 3, p. 035102, 2016.

[348] M. Chiricotto, S. Melchionna, P. Derreumaux, and F. Sterpone, "Multiscale aggregation of the amyloid aβ16–22 peptide: From disordered coagulation and lateral branching to amorphous prefibrils," *The Journal of Physical Chemistry Letters*, vol. 10, no. 7, pp. 1594–1599, 2019.

[349] N. Samanta, S. S. Ribeiro, M. Becker, E. Laborie, R. Pollak, S. Timr, F. Sterpone, and S. Ebbinghaus, "Sequestration of proteins in stress granules relies on the in-cell but not the in vitro folding stability," *Journal of the American Chemical Society*, vol. 143, no. 47, pp. 19909–19918, 2021. PMID: 34788540.

[350] A. F. Brandner, S. Timr, S. Melchionna, P. Derreumaux, M. Baaden, and F. Sterpone, "Modelling lipid systems in fluid with lattice boltzmann molecular dynamics simulations and hydrodynamics," *Scientific Reports*, vol. 9, p. 16450, 2019.

[351] O. Languin-Cattoën, S. Melchionna, P. Derreumaux, G. Stirnemann, and F. Sterpone, "Three weaknesses for three perturbations: Comparing protein unfolding under shear, force, and thermal stresses," *The Journal of Physical Chemistry B*, vol. 122, no. 50, pp. 11922–11930, 2018.

[352] F. Sterpone, P. Derreumaux, and S. Melchionna, "Molecular Mechanism of Protein Unfolding under Shear: A Lattice Boltzmann Molecular Dynamics Study," *The Journal of Physical Chemistry B*, vol. 122, no. 5, pp. 1573–1579, 2018. PMID: 29328657.

[353] K. A. Dill, T. M. Truskett, V. Vlachy, and B. Hribar-Lee, "Modeling water, the hydrophobic effect, and ion solvation," *Annual Review of Biophysics and Biomolecular Structure*, vol. 34, no. 1, pp. 173–199, 2005. PMID: 15869376.

[354] A. Badasyan, Y. S. Mamasakhlisov, R. Podgornik, and V. A. Parsegian, "Solvent effects in the helix-coil transition model can explain the unusual biophysics of intrinsically disordered proteins," *The Journal of Chemical Physics*, vol. 143, no. 1, p. 014102, 2015.

[355] C. J. Fennell, C. W. Kehoe, and K. A. Dill, "Modeling aqueous solvation with semi-explicit assembly," *Proceedings of the National Academy of Sciences*, vol. 108, no. 8, pp. 3234–3239, 2011.

[356] C. Camilloni, D. Bonetti, A. Morrone, R. Giri, C. M. Dobson, M. Brunori, S. Gianni, and M. Vendruscolo, "Towards a structural biology of the hydrophobic effect in protein folding," *Scientific Reports*, vol. 6, p. 28285, Jul 2016.

[357] T. E. Creighton, "Physical Interactions That Determine the Properties of Proteins," in *Proteins: Structures and molecular properties*, ch. 4, pp. 139–169, New York: W. H. Freeman, 1993.

[358] S. F. Banani, H. O. Lee, A. A. Hyman, and M. K. Rosen, "Biomolecular condensates: organizers of cellular biochemistry," *Nature Reviews Molecular Cell Biology*, vol. 18, pp. 285–298, May 2017.

[359] A. Klosin, F. Oltsch, T. Harmon, A. Honigmann, F. Jülicher, A. A. Hyman, and C. Zechner, "Phase separation provides a mechanism to reduce noise in cells," *Science*, vol. 367, no. 6476, pp. 464–468, 2020.

[360] A. A. Hyman, C. A. Weber, and F. Jülicher, "Liquid-liquid phase separation in biology," *Annual Review of Cell and Developmental Biology*, vol. 30, no. 1, pp. 39–58, 2014. PMID: 25288112.

[361] W. van Leeuwen and C. Rabouille, "Cellular stress leads to the formation of membraneless stress assemblies in eukaryotic cells," *Traffic*, vol. 20, no. 9, pp. 623–638, 2019.

[362] A. A. M. André and E. Spruijt, "Liquid–liquid phase separation in crowded environments," *International Journal of Molecular Sciences*, vol. 21, no. 16, 2020.

[363] M. Hondele, S. Heinrich, P. De Los Rios, and K. Weis, "Membraneless organelles: phasing out of equilibrium," *Emerging Topics in Life Sciences*, vol. 4, pp. 343–354, 08 2020.

[364] C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Jülicher, and A. A. Hyman, "Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation," *Science*, vol. 324, no. 5935, pp. 1729–1732, 2009.

[365] S. Elbaum-Garfinkle, Y. Kim, K. Szczepaniak, C. C.-H. Chen, C. R. Eckmann, S. Myong, and C. P. Brangwynne, "The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics," *Proceedings of the National Academy of Sciences*, vol. 112, no. 23, pp. 7189–7194, 2015.

[366] L. Jawerth, E. Fischer-Friedrich, S. Saha, J. Wang, T. Franzmann, X. Zhang, J. Sachweh, M. Ruer, M. Ijavi, S. Saha, J. Mahamid, A. A. Hyman, and F. Jülicher, "Protein condensates as aging Maxwell fluids," *Science*, vol. 370, no. 6522, pp. 1317–1323, 2020.

[367] H. Kusumaatmaja, A. I. May, and R. L. Knorr, "Intracellular wetting mediates contacts between liquid compartments and membrane-bound organelles," *Journal of Cell Biology*, vol. 220, 08 2021. e202103175.

[368] Y. Wang, M. Sarkar, A. E. Smith, A. S. Krois, and G. J. Pielak, "Macromolecular crowding and protein stability," *Journal of the American Chemical Society*, vol. 134, no. 40, pp. 16614–16618, 2012. PMID: 22954326.

[369] A. W. Fritsch, A. F. Diaz-Delgadillo, O. Adame-Arana, C. Hoege, M. Mittasch, M. Kreysing, M. Leaver, A. A. Hyman, F. Jülicher, and C. A. Weber, "Local thermodynamics govern formation and dissolution of ¡i¿caenorhabditis¡/i¿ elegans p granule condensates," *Proceedings of the National Academy of Sciences*, vol. 118, no. 37, p. e2102772118, 2021.

[370] T. M. Franzmann, M. Jahnel, A. Pozniakovsky, J. Mahamid, A. S. Holehouse, E. Nüske, D. Richter, W. Baumeister, S. W. Grill, R. V. Pappu, A. A. Hyman, and S. Alberti, “Phase separation of a yeast prion protein promotes cellular fitness,” *Science*, vol. 359, no. 6371, p. eaao5654, 2018.

[371] O. Adame-Arana, C. A. Weber, V. Zaburdaev, J. Prost, and F. Jülicher, “Liquid Phase Separation Controlled by pH,” *Biophysical Journal*, vol. 119, pp. 1590–1605, Oct 2020.

[372] R. W. Thompson, R. F. Latypov, Y. Wang, A. Lomakin, J. A. Meyer, S. Vunnum, and G. B. Benedek, “Evaluation of effects of pH and ionic strength on colloidal stability of IgG solutions by PEG-induced liquid-liquid phase separation,” *The Journal of Chemical Physics*, vol. 145, no. 18, p. 185101, 2016.

[373] J. A. Riback, C. D. Katanski, J. L. Kear-Scott, E. V. Pilipenko, A. E. Rojek, T. R. Sosnick, and D. A. Drummond, “Stress-triggered phase separation is an adaptive, evolutionarily tuned response,” *Cell*, vol. 168, pp. 1028–1040.e19, Mar 2017.

[374] L. Zhao, G. Vecchi, M. Vendruscolo, R. Körner, M. Hayer-Hartl, and F. U. Hartl, “The Hsp70 Chaperone System Stabilizes a Thermo-sensitive Subproteome in E. coli,” *Cell Reports*, vol. 28, pp. 1335–1345.e6, Jul 2019.

[375] P. Gallardo, S. Salas-Pino, and R. R. Daga, “Reversible protein aggregation as cytoprotective mechanism against heat stress,” *Current Genetics*, vol. 67, pp. 849–855, Dec 2021.

[376] R. M. Vabulas, S. Raychaudhuri, M. Hayer-Hartl, and F. U. Hartl, “Protein folding in the cytoplasm and the heat shock response,” *Cold Spring Harbor Perspectives in Biology*, vol. 2, no. 12, 2010.

[377] R. J. Wood, A. R. Ormsby, M. Radwan, D. Cox, A. Sharma, T. Vöpel, S. Ebbinghaus, M. Oliveberg, G. E. Reid, A. Dickson, and D. M. Hatters, “A biosensor-based framework to measure latent proteostasis capacity,” *Nature Communications*, vol. 9, p. 287, Jan 2018.

[378] S. Ganesan, G. Rohde, K. Eckermann, K. Sroka, M. K. E. Schaefer, C. P. Dohm, P. Kermer, G. Haase, F. Wouters, M. Bähr, and J. H. Weishaupt, “Mutant SOD1 detoxification mechanisms in intact single cells,” *Cell Death & Differentiation*, vol. 15, pp. 312–321, Feb 2008.

[379] G. S. A. Wright, S. V. Antonyuk, and S. S. Hasnain, “The biophysics of superoxide dismutase-1 and amyotrophic lateral sclerosis,” *Quarterly Reviews of Biophysics*, vol. 52, p. e12, 2019.

[380] P. H. Nguyen, A. Ramamoorthy, B. R. Sahoo, J. Zheng, P. Faller, J. E. Straub, L. Dominguez, J.-E. Shea, N. V. Dokholyan, A. De Simone, B. Ma, R. Nussinov, S. Najafi, S. T. Ngo, A. Loquet, M. Chiricotto, P. Ganguly, J. McCarty, M. S. Li, C. Hall, Y. Wang, Y. Miller, S. Melchionna, B. Habenstein, S. Timr, J. Chen, B. Hnath, B. Strodel, R. Kayed, S. Lesné, G. Wei, F. Sterpone, A. J. Doig, and P. Derreumaux, "Amyloid Oligomers: A Joint Experimental/Computational Perspective on Alzheimer's Disease, Parkinson's Disease, Type II Diabetes, and Amyotrophic Lateral Sclerosis," *Chemical Reviews*, vol. 121, no. 4, pp. 2545–2647, 2021. PMID: 33543942.

[381] K. A. Burke, A. M. Janke, C. L. Rhine, and N. L. Fawzi, "Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II," *Molecular Cell*, vol. 60, pp. 231–241, Oct 2015.

[382] V. F. Thompson, R. A. Victor, A. A. Morera, M. Moinpour, M. N. Liu, C. C. Kisiel, K. Pickrel, C. E. Springhower, and J. C. Schwartz, "Transcription-Dependent Formation of Nuclear Granules Containing FUS and RNA Pol II," *Biochemistry*, vol. 57, no. 51, pp. 7021–7032, 2018.

[383] K. Kamagata, N. Iwaki, S. Kanbayashi, T. Banerjee, R. Chiba, V. Gaudon, B. Castaing, and S. Sakomoto, "Structure-dependent recruitment and diffusion of guest proteins in liquid droplets of FUS," *Scientific Reports*, vol. 12, p. 7101, May 2022.

[384] A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, and J. P. Taylor, "Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization," *Cell*, vol. 163, pp. 123–133, Sep 2015.

[385] P. S. Tsoi, M. D. Quan, K.-J. Choi, K. M. Dao, J. C. Ferreon, and A. C. M. Ferreon, "Electrostatic modulation of hnRNPA1 low-complexity domain liquid–liquid phase separation and aggregation," *Protein Science*, vol. 30, no. 7, pp. 1408–1417, 2021.

[386] F. G. Quiroz and A. Chilkoti, "Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers," *Nature Materials*, vol. 14, pp. 1164–1171, Nov 2015.

[387] G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal, "Sequence determinants of protein phase behavior from a coarse-grained model," *PLOS Computational Biology*, vol. 14, pp. 1–23, 01 2018.

[388] G. Rivas and A. P. Minton, "Macromolecular crowding in vitro, in vivo, and in between," *Trends in Biochemical Sciences*, vol. 41, pp. 970–981, Nov 2016.

[389] D. Gnutt, S. Timr, J. Ahlers, B. König, E. Manderfeld, M. Heyden, F. Sterpone, and S. Ebbinghaus, “Stability effect of quinary interactions reversed by single point mutations,” *Journal of the American Chemical Society*, vol. 141, no. 11, pp. 4660–4669, 2019. PMID: 30740972.

[390] N. M. Kanaan, C. Hamel, T. Grabinski, and B. Combs, “Liquid-liquid phase separation induces pathogenic tau conformations in vitro,” *Nature Communications*, vol. 11, p. 2809, Jun 2020.

[391] I. M. Kuznetsova, B. Y. Zaslavsky, L. Breydo, K. K. Turoverov, and V. N. Uversky, “Beyond the excluded volume effects: Mechanistic complexity of the crowded milieu,” *Molecules*, vol. 20, no. 1, pp. 1377–1409, 2015.

[392] D. S. Goodsell, *The Machinery of Life.* Copernicus New York, NY, 2nd ed., 2010.

Curriculum Vitae

Education

- **2017-ongoing** Ph D. in Physics, University of Barcelona

  Project: "Calculations of water free energy in bulk and large biological systems"

  Supervisor: Dr. Giancarlo Franzese

- **2015-ongoing** Bachelor's degree in Phiolosophy, University of Barcelona

- **2015-2016** Master's degree in Atomistic and Multiscale Computational Modelling in Physics, Chemistry and Biochemistry

  Master dissertation: "Liquid-Liquid Phase Transition Analysis for the 3D Many Body Water Model", University of Barcelona and Technical University of Catalonia

  Supervisor: Dr. Giancarlo Franzese

- **2011-2015** Bachelor's degree in Physics, University of Valladolid

List of publications

Directly related to the content of the Thesis

- **Luis E. Coronas**, Oriol Vilanova, Valentino Bianco, Francisco de los Santos, Giancarlo Franzese, "The Franzese-Stanley Coarse Grained Model for Hydration Water" Ed. F. Martelli in *Properties of Water from Numerical and Experimental Perspectives*, chapter 3. Science Publishers. (2022)

- **Luis E. Coronas**, Oriol Vialanova, Giancarlo Franzese, "Parallel distributed Monte Carlo algorithms for large-scale simulations of a coarse-grained water model" IN PREPARATION

- **Luis E. Coronas**, Oriol Vilanova, Giancarlo Franzese, "Parametrization around ambient conditions of a coarse-grained water model for large-scale simulations" IN PREPARATION. To be submitted to the Journal of Molecular Liquids.

- **Luis E. Coronas**, Giancarlo Franzese, “Second critical point on large-scale simulations of a cooperative coarse-grained water model” IN PREPARATION
- **Luis E. Coronas**, Emeline Laborie, Philippe Derremaux, Fabio Sterpone, Giancarlo Franzese, “Estimate of the hydration effect in large-scalebiomolecular simulations: the von Willebrand factor case” IN PREPARATION
- **Luis E. Coronas**, Emeline Laborie, Stepan Timr, Fabio Sterpone, Giancarlo Franzese, “Effect of crowding on SOD1 sequestration into FUS biocondensates” IN PREPARATION

Non-directly related to the content of the Thesis

- Paola Gallo, Johannes Bachler, Livia E. Bove, Roland Böhmer, Gaia Camisasca, **Luis E. Coronas**, Horacio R. Corti, Ingrid de Almeida Ribeiro, Maurice de Koning, Giancarlo Franzese, Violeta Fuentes-Landete, Catalin Gainaru, Thomas Loerting, Joan Manuel Montes de Oca, Peter H. Poole, Mauro Rovere, Francesco Sciortino, Christina M. Tonauer, and Gustavo A. Appignanesi “Advances in the study of supercooled water” The European Physical Journal E 44:143 (2021)

Conference contributions

Oral presentations

- “Parametrization of a bulk water model, with molecular hydrogen-bonding and coare-grained density, for large scale simulations in quantitative agreement with experiments under life-relevant conditions”.

  VI Workshop de la Red Española de Simulación Molecular. Baiona (Spain) 19-21 June 2022. Organized by Universidad de Huelva, Spain.

- “Coarse-grained model for water and water-protein interfaces”.

  IN2UB 2021 Annual Meeting. Online 14 October 2021. Organized by University of Barcelona, Spain.

- “Water contribution to the protein folding and its relevance in protein design and protein aggregation”.

  I Convegno SIFS 2021. Online 23-25 June 2021. Organized by Società Italiana di Fisica Statistica, Italy. (Poster + Oral)

- “Coarse-grained model for water and water-protein interfaces”.

  I Congreso FisEs Joven’ 21. Online 5-6 May 2021. Organized by Universidad de Zaragoza, Spain.

- “Coarse-grained model for water and water-protein interfaces”.

  V Workshop de la Red Española de Simulación Molecular. Online 29 January 2021. Organized by Universidad de Huelva, Spain.

- “Phase transition analysis between low density and high density nanoconfined water”.

  Joint EMLG/JMLG (European/Japanese Molecular Liquid Group) conference 2019. Kutná Hora (Czech Republic) 8-13 September 2019.

- “Phase transition analysis between low density and high density nanoconfined water”.

  Frontiers in Water Biophysics 2019. Erice (Italy) 21-26 July 2019.

- “Phase transition analysis between low density and high density nanoconfined water”.

  XXVI Sitges Conference on Statistical Mechanics: New trends on statistical physics. Sitges (Spain) 27-31 May 2019

Poster

- “A quantitative model for large- scale simulations at life-relevant conditions”.

  Joint EMLG/JMLG (European/Japanese Molecular Liquid Group) conference 2022. Barcelona (Spain) 12-16 September 2022. Organized by the University of Barcelona.

- “Reparametrization around ambient conditions of a water model for large scale simulations”.

  Mixed-Gen Season 2 – Session 7: Simulating non-equilibrium phenomena and rare events. Online 28 April 2022. Organized by CECAM-HQ-EPFL, Lausanne (Switzerland).

- “Coarse- grained model for water and water-protein interfaces”.

  NanoBio&Med International Conference. Barcelona (Spain) 19-21 November 2019.

- “Phase transition analysis between low density and high density nanoconfined water”.

  Water Europe: Multiscale simulations and coarse-grained models for water and aqueous systems. Laussane (Switzerland) 21-23 October 2019. Organized by CECAM-HQ-EPFL

- “Phase transition analysis between low density and high density nanoconfined water”.

  IN2UB 2019 Meeting. Barcelona, (Spain) 6 June 2019.

- “Towards a coherent picture of LD, HD and VHD forms of nanoconfined water”

  Primer encuentro de Física Estadística de No Equilibrio y sus aplicaciones multidisciplinares, Barcelona, España, 21-23 November 2018

- “Towards a coherent picture of LD, HD and VHD forms of nanoconfined water”.

  I Course of the ERICE School WATER AND THE WATER SYSTEM The hydrophobic effect, Erice (Italy) 04- 11 July 2018 (Poster + flash talk)

- “Toward a coherent picture of the low-density, high-density and very high-density forms of bulk water”.

  IN2UB 2018 Meeting, Barcelona (Spain) 5 June 2018.

Invited seminars

- “Water thermodynamics and its effects on biological interfaces”.

  LBT-IBPC Seminars. Institute de Biologie Physico-Chimique, CNRS, Paris (France) 16 December 2021.

- “Water thermodynamics and its effects on biological interfaces”.

  SIFS Young Seminars. Online 9 December 2021 (organized by Società Italiana di Fisica Statistica, Italy).