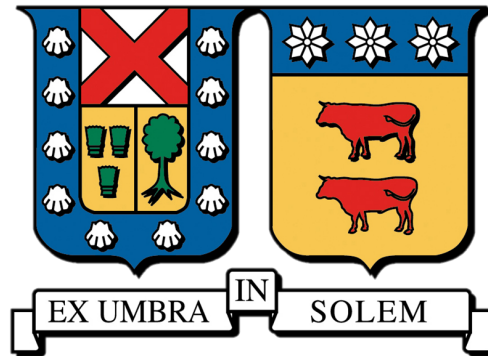


Universidad Técnica Federico Santa María
Departamento de Ingeniería Mecánica
Valparaíso - Chile



“Influencia de la longitud de Tolman en el cálculo de la energía necesaria para disolver un soluto, usando un modelo de medio continuo.”

Memoria de titulación para optar al título:
Ingeniero Civil Mecánico

Valentina Oyarzún Muñoz

Profesor guía: PhD. Christopher Cooper Villagrán
Profesor correferente: PhD. Harvey Zambrano Rodríguez

Enero - 2020

Agradecimientos

A mi compañero y amigo, la mejor iteración de Juaquin "Kunkin" Echaíz que existe en todo el multiverso.

A mi familia querida y amigos, quienes siempre me brindan amor y apoyo.

Finalmente, agradezco la valiente Alexandra Elbakyan, cuyo labor permitió la elaboración de toda esta investigación y miles más.

Resumen

Para calcular la energía de solvatación no-polar para una pequeña molécula orgánica neutral en un medio acuoso, se debe entender la naturaleza de la tensión superficial en una interfase entre agua líquida y vapor. La energía no-polar es la suma de la energía necesaria para generar una cavidad en el medio acuoso, y el costo energético debido a la interacción no electroestática entre la especie y el medio por [fuerzas débiles de Van der Waals](#). Es aceptado en la comunidad científica que la tensión superficial depende de el radio de curvatura por medio de la longitud de Tolman, sin embargo la forma de la variable de Tolman es desconocida y el valor numérico (hasta su signo) sigue siendo debatido.

Utilizamos aproximaciones de la energía no-polar para 502 moléculas orgánicas pequeñas y neutras obtenidas por Mobley et al. (2009) [1] por dinámica molecular de solvente explícito, y datos de Cooper et al. (2019) de la energía de Van der Waals [2]. Mediante un procedimiento general de mínimos cuadrados lineales, encontramos un valor para la función de corrección de la tensión superficial que se ajusta a los datos de Mobley. Evaluamos cuatro simplificaciones de la ecuación [Gibbs-Tolman-Koenig-Buff \(GTKB\)](#) para encontrar patrones en la variable de Tolman, graficando los resultados contra el radio de curvatura. Con estos resultados diseñamos distintos modelos para calcular la energía de formación de cavidad, generando el modelo con una pequeña muestra de las especies, y prediciendo la energía de formación de cavidad para el resto de las 502 especies totales. Además, empleamos modelos para predecir la energía de formación de cavidad considerando un valor de la longitud de Tolman constante para las simplificaciones de [GTKB](#).

Para un total de 16 modelos, la correlación con los datos de Mobley et al. (2009) es superior a 0,90 para 6, 13, y 16 modelos al usar 20, 50 y 100 especies para generarlos, respectivamente. La mayoría de los modelos empleados, 14 de 16, aumentan en exactitud al generar el modelo con más especies y predecir la energía de formación de cavidad para las especies restantes, lo que es esperado. El mejor modelo siempre presenta una correlación superior a 0,99. Los resultados obtenidos presentan poca correlación con los resultados de Mobley et al. (2009) para especies cuyo costo energético es muy bajo ($< 9 \text{ kcal/mol}$) o alto ($> 15 \text{ kcal/mol}$). Los modelos que entregan resultados con mayor correlación ($> 0,97$) consideran Tolman constante, y los gráficos que muestran el comportamiento de Tolman en función con el radio de curvatura no demuestran una clara dependencia entre estas.

Abstract

To estimate the non-polar solvation energy for an organic molecule functional group in water in bulk, we must first understand the nature of surface tension in a liquid-vapor interface. Non-polar solvation free energy is the sum of the energy necessary to create a cavity in bulk water, and the cost of the non electrostatic interaction between the functional group and medium through weak Van der Waal forces. The dependency of surface tension on the Tolman's length is widely accepted in the scientific community, however the form of the Tolman variable is unknown and the numeric value (even its sign) is still debated.

Our theoretical values are approximations of the non-polar energy for 502 functional groups published by Mobley et al. (2009) [1], obtained through the method of molecular dynamics, and approximations of the Van der Waals energy by Cooper (2019) [2]. Through least squares lineal regression, a fitted value for the correctional factor of the surface tension may be found. Additionally, four simplifications of the Gibbs-Tolman-Koenig-Buff (GTKB) equation are evaluated to find patterns in the Tolman variable, later graphing the results for the Tolman length against the radius of curvature. With these results, various models can be designed and used to predict the energy of cavity formation for other functional groups. Furthermore, constant Tolman length values are also tested, obtained through the four simplifications of the GTKB equation.

For a total of 16 models, the pearson correlation coeficient of prediction results and theoretical values is over 0,90 for 6, 13, and 16 models using 20, 50 and 100 functional groups to design them, respectively. 14 models improved in their linear relationship stregnth as the amount of functional groups used in their design increased as expected. The best performing model has a pearson correlation over 0,99 regardless of how many functional groups are used in its design. For species whose energetic cost of cavity formation is either too low ($< 9 \text{ kcal/mol}$) or too high ($> 15 \text{ kcal/mol}$), prediction results overestimate and subestimate energy costs, respectively. The best performing models ($> 0,97$) all consider a constant Tolman value, and graphing fitted Tolman values against the radius of curvature does not show a clear relationship between the two.

Acrónimos

AWPN Angle Weighted Pseudo Normal. , [14](#)

GTKB Gibbs-Tolman-Koenig-Buff. [ii](#), [iii](#), [x](#), [9](#), [12](#), [19](#), [23](#), [24](#), [26](#), [32](#)

IQR rango intercuartil. , [17](#)

SAS Solvent Accesible Surface. , [ix](#), [10](#), [11](#)

SES Solvent Excluded Surface. , [10](#)

VDW Van der Waals Surface. , [5](#), [10](#)

Glosario

fuerzas débiles de Van der Waals Las fuerzas de Van der Waals son fuerzas intermoleculares eléctricas relativamente débiles. Existen interacciones atractivas entre los átomos, que resulta de los dipolos inducidos, y interacciones repulsivas, que resulta de la superposición de las nubes de electrones de los dos átomos, cuando se encuentran a muy corta distancia.. [ii](#), [iii](#), [2](#)

one-ring neighborhood En mallas triangulares, el one-ring de un nodo n es el conjunto de todos los nodos en la malla que están directamente conectados por una arista a n .. , [13](#), [14](#)

partículas hidrofílicas Una sustancia es hidrofílica si tiene grupos polares fuertes que interactúan fácilmente con el agua, formando puentes de hidrógeno.. , [1](#)

partículas hidrofóbicas Una sustancia es hidrofóbica si no es miscible con el agua, no es capaz de interactuar con moléculas de agua por medio de interacciones ión-dipolo ni mediante puentes de hidrógeno.. , [1-3](#)

regla de Tukey Los métodos de Estadística Robusta producen series de datos de buen comportamiento. La Regla de Tukey es un método de estadística robusta que considera datos aberrantes aquellos que se alejan más de 1,5 veces el rango intercuartil de los cuartiles – bajo $Q1$ o sobre $Q3$.. , [17](#), [18](#), [26](#)

relación de Young-Laplace Describe la diferencia de presión entre el interior y exterior de una burbuja de gas dentro de un ambiente líquido del mismo componente.. , [7](#), [8](#)

superficie divisora de Gibbs Las superficies divisoras permiten definir a la interfaz líquido-vapor de radio constante (y por ende, de densidad constante) y sin espesor.. , [7](#), [8](#)

zonas húmedas Zona en donde la interfaz que define la cavidad se encuentra más cercana a la superficie molecular debido a la presencia de moléculas polares con las cuales se pueden formar enlaces de hidrógeno..

zonas secas Zona en donde la interfaz que define la cavidad se encuentra más alejada de la superficie molecular debido a la presencia de moléculas no-polares..

Lista de símbolos

Símbolo	Descripción	Unidad
P	Presión	Pa
V	Volumen	Å
ΔG_{cav}	Componente np-polar de la energía libre de solvatación debido a formación de cavidad esférica	J
ΔG_{np}	Componente no-polar de la energía libre de solvatación	J
ΔG_p	Componente polar de la energía libre de solvatación	J
ΔG_{vdw}	Componente no-polar de la energía libre debido a fuerzas Van der Waals	J
ΔG	Energía libre de solvatación	J
δ	Longitud de Tolman	Å
γ_{lv}	Tensión superficial del agua en una interfaz plana	N/m
γ	Tensión superficial	N/m
$f(R)$	Factor de corrección del valor planar de la tensión superficial	–

Índice general

1. Marco teórico	1
1.1. Energía de formación de cavidad	1
1.2. Energía libre de solvatación	4
1.3. Tensión superficial y la longitud de Tolman	6
1.4. Estudios actuales sobre la forma y valor de la longitud de Tolman	9
1.5. Generación de mallas y base de datos	10
2. Desarrollo	12
2.1. Reproduciendo resultados de Chandler(2005) y análisis de variables	12
2.2. Propiedades geométricas en mallas triangulares	12
2.3. Aproximación de la longitud de Tolman	14
2.3.1. Tolman Característico	14
2.3.2. Factor de corrección definido por Nodo	15
2.4. Capacidad de predicción del método	19
3. Resultados	26
4. Conclusiones	32
Referencias	34
A. Derivación de la energía de formación de cavidad	38

B. Mejor ajuste de una esfera	40
C. Solución del factor de corrección	42

Índice de figuras

1.1. Energía libre de solvatación para una cavidad esférica en función del tamaño de la cavidad.	3
1.2. Desglose de la energía de solvatación.	5
1.3. Sistema de dos fases de un solo componente con interfase plana.	6
1.4. Sistema de dos fases de un componente con interfase curva.	8
1.5. Esquema de superficies moleculares en 2D.	11
1.6. Mallas triangulares representativas de la superficie Solvent Accesible Surface (SAS), definidas por el software MSMS y visualizadas por la librería Trimesh de Python3.	11
2.1. Energía libre de solvatación calculada para una cavidad esférica, comparada con resultados de Chandler (2005) [3].	13
2.2. Definiciones de curvatura en mallas triangulares.	15
2.3. Histograma de radios de curvatura para 502 especies de Mobley et al. (2009) [1].	18
2.4. Ajuste de variable de Tolman usando la solución de $f(R)$ considerando 20 intervalos equidistantes.	19
2.5. Ajuste de variable de Tolman usando la solución de $f(R)$ considerando 20 intervalos equidistantes, eliminando datos aberrantes.	20
2.6. Ajuste de variable de Tolman usando la solución de $f(R)$ considerando 20 intervalos equidensos.	21
2.7. Ajuste de variable de Tolman usando la solución de $f(R)$ considerando 20 intervalos equidensos, eliminando datos aberrantes.	22

2.8. Construcción de la función constante por partes del factor de corrección de la tensión superficial	23
2.9. Construcción de la función linear por partes del factor de corrección de la tensión superficial	23
2.10. Tolman característico por especie según simplificaciones de GTKB.	24
2.11. Reproducción de resultados Chandler (2005) [3] para energía libre de cavitación esférica usando simplificaciones de GTKB.	25
3.1. Modelos generados con 20 especies de mejor correlación con datos teóricos de Mobley et al. (2009) [1].	27
3.2. Modelos generados con 50 especies de mejor correlación con datos teóricos de Mobley et al. (2009) [1].	28
3.3. Modelos generados con 100 especies de mejor correlación con datos teóricos de Mobley et al. (2009) [1].	29
3.4. Coeficiente de correlación de Pearson por modelo.	31
C.1. Corrección de la tensión superficial, resultados de 2.15 con 20 intervalos equidistantes sin eliminación de datos.	42
C.2. Corrección de la tensión superficial, resultados de 2.15 con 20 intervalos equidistantes con eliminación de datos aberrantes.	42
C.3. Corrección de la tensión superficial, resultados de 2.15 con 20 intervalos equidensos sin eliminación de datos.	43
C.4. Corrección de la tensión superficial, resultados de 2.15 con 20 intervalos equidensos con eliminación de datos aberrantes	43

Índice de tablas

1.1. Valores de la longitud de Tolman para una interfaz de agua líquido-vapor de distintas fuentes. Adaptado de Kalova et al. (2019) [4]	10
3.1. Coeficiente de correlación de Pearson por modelo.	30

Introducción

En esta investigación estudiamos la naturaleza de la tensión superficial para luego calcular la energía necesaria para introducir una pequeña molécula orgánica neutral a un medio acuoso desde el infinito y el vacío. Esta es la energía de solvatación, la cual se puede dividir en dos componentes, polar y no-polar. La energía no-polar es la suma de la energía necesaria para generar una cavidad en el medio acuoso, y el costo energético debido a la interacción no electrostática entre la especie y el medio, considerando exclusión repulsiva de corto rango y dispersión atractiva de un largo rango. Para poder calcular la energía libre de Gibbs de formación de cavidad en un medio acuoso, se debe entender la naturaleza de la tensión superficial en una interfase entre agua líquida y vapor.

La tensión superficial es una propiedad física muy interesante de los fluidos y su naturaleza sigue siendo altamente debatida dentro de la comunidad científica. Esta propiedad tiene numerosas aplicaciones en la industria, por ejemplo, en la fabricación de adhesivos, espumas y lubricantes. La tensión superficial está relacionada a muchos fenómenos físicos como la nucleación, humectación, el fenómeno de la capilaridad, y, el enfoque del presente estudio, la formación de cavidades nanoscópicas en un medio acuoso, de gran relevancia en el ámbito físico químico.

Para entender la naturaleza de la tensión superficial se debe conocer el estudio de Willard Gibbs (1976), "*On the equilibrium of heterogenous substances*" el cual forma la base de la termodinámica química y la química física hasta el día de hoy, en donde encontró que la tensión superficial se relaciona además con el radio de curvatura [5]. Luego, Tolman afirmó en el año 1949 que la tensión superficial disminuye con un aumento en la curvatura [6], sin embargo la forma de la variable de Tolman es desconocida y el valor numérico (hasta su signo) sigue siendo debatido [4].

En este trabajo estudiamos la relación entre la longitud de Tolman, variable que mide el grado en que la tensión superficial en una interfase curva se desvía de su valor planar, y la curvatura de la superficie de la interfase líquido-vapor. Esperamos desarrollar modelos que aproximen la longitud de Tolman para comprobar su relación con la curvatura, utilizando resultados de la energía de formación de cavidad ya disponibles. Comprobamos los métodos de aproximación evaluando su exactitud en calcular la energía libre de formación de cavidad en un medio acuoso para distintas moléculas orgánicas pequeñas y neutrales.

En este trabajo, se deben evaluar distintas formas de calcular la variable de Tolman y la energía de solvatación. Los objetivos específicos de este trabajo son:

- Estudiar el estado de arte e identificar las variables involucradas.
- Modelar resultados de Chandler (2005) para verificar la energía requerida para generar una cavidad esférica de distintos diámetros.
- Proponer posibles modelos de la longitud de Tolman en relación con el radio de curvatura con datos experimentales (Mobley, Bayly, Cooper, Shirts, y Dill 2009).
- Probar los modelos propuestos con un algoritmo para aproximar la componente no-polar de solvatación en mallas triangulares.
- Analizar exactitud, precisión y eficiencia de los modelos propuestos.
- Determinar el modelo que mejor aproxima la componente no-polar de la energía de solvatación.

1 | Marco teórico

Para comenzar a abordar la problemática, se requiere entender primero el fenómeno de la hidrofobicidad y la creación de cavidades.

Una partícula se denomina *polar* según la distribución de carga en ella. Las moléculas no-polares por lo general no se disuelven fácilmente en el agua, son **partículas hidrofóbicas**, debido a que no se forman enlaces de hidrógeno con el agua alrededor. Por otro lado, cuando existen regiones locales de carga negativa o positiva, se dice que la molécula es polar y se disuelve fácilmente en el agua, interactuando con ella por medio de enlaces de hidrógeno. Típicamente, las **partículas hidrofílicas** son polares, sin embargo existen excepciones. Cuando las interacciones soluto-soluto son más fuertes que las interacciones solvente-soluto, por ejemplo al introducir lípidos a un medio acuoso, estas partículas no-polares tienden a formar cúmulos de unidades hidrofóbicas [7]. Éste es el fenómeno de la hidrofobicidad, en donde se forman cúmulos debido a que existe menor gasto energético al disminuir el área superficial.

1.1. Energía de formación de cavidad

El costo termodinámico de una reacción es cuantificado por la energía libre de Gibbs ΔG e indica si el proceso es probable de ocurrir. La energía libre considera cambios entálpicos ΔH , o en la energía potencial media de interacción entre moléculas más el trabajo de flujo necesario para crear el volumen de control, y cambios entrópicos ΔS , una medida del orden intermolecular. En la ecuación 1.1, ΔG es la energía libre de Gibbs y T la temperatura del sistema.

$$\Delta G = \Delta H - T\Delta S \quad (1.1)$$

Consideremos entonces un cúmulo de partículas hidrofóbicas, no-polares, dentro de un medio acuoso. En condiciones ambientales, el agua líquida se encuentra cercana a la coexistencia con su fase gaseosa, siendo así, la diferencia entre la energía libre de Gibbs de la fase líquida y gaseosa del agua es menor a la energía térmica disponible en el sistema. Esta condición asegura la formación de cavidades en el agua, como las que se crean debido a cúmulos de partículas hidrofóbicas, que son acompañadas por un interfase similar a la de líquido-vapor [8].

Según análisis termodinámico, investigadores esperan una alta correlación entre la energía de formación de cavidad y el área y volumen de la molécula. La ecuación 1.2 entrega el costo asociado a la formación de una cavidad esférica de radio r , formada entorno a un cúmulo de partículas no-polares. Ésta es la energía libre de solvatación para una cavidad esférica, ver anexo A para su derivación.

$$\Delta G_{cav} = \frac{4\pi}{3}r^3P + \int \gamma dA \quad (1.2)$$

Para un soluto de radio constante:

$$\frac{\Delta G_{cav}}{4\pi r^2} = \frac{Pr}{3} + \gamma \cdot f(R) \quad (1.3)$$

El primer término a la derecha de la ecuación 1.3 representa el trabajo necesario para expandir una cavidad contra la presión externa, P . El segundo término describe el trabajo debido a la tensión superficial, incluyendo un factor de corrección $f(R)$. Esperamos que el aporte de el primer término sea inferior cuando el medio acuoso se encuentra en condiciones atmosféricas debido a que bajo estas condiciones el agua se encuentra cercana a su punto triple.

La figura 1.1 muestra la energía libre de solvatación normalizada al área superficial de la interfase con respecto a su radio.

Con respecto a radios menores a un nanómetro

Las partículas hidrofóbicas interactúan de varias maneras con el agua que las rodea, por ejemplo, ejerciendo fuerzas débiles de Van der Waals que reordenan los enlaces de hidrógeno entre las moléculas de agua adyacentes. En este proceso el agua responde a la presencia del cúmulo de forma elástica, adoptando una orientación que permite a la red de enlaces de hidrógeno que rodeen el soluto sin romperse. Este proceso requiere una organización espacial muy específica en donde el componente entrópico de la energía de Gibbs domina.

Esta reorganización persiste en el líquido adyacente a una distancia equivalente a una *longitud de correlación*. Debido a que la longitud de correlación en agua se extiende sobre la circunferencia de una molécula pequeña, sumando los puntos afectados alrededor de un pequeño soluto es mejor correlacionado con su volumen que con el área superficial [3]. Mientras

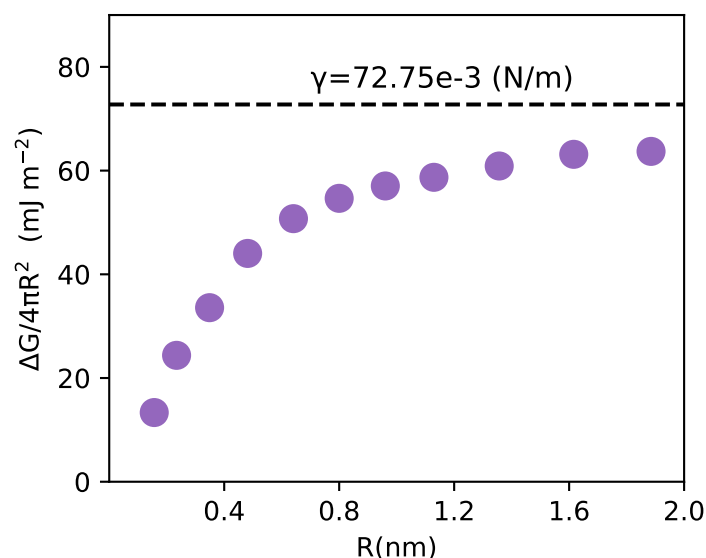


Figura 1.1: Energía libre de solvatación para una cavidad esférica en un medio acuoso en función del tamaño de la cavidad. La tensión superficial $\gamma_{lv} = 72,75 \times 10^{-3} \text{ N/m}$ se comporta como un asintota de la cantidad de energía por unidad de área de la cavidad. Adaptado de Chandler (2005) [3]

mayor sea el volumen, más enlaces requieren reorganización. Esta relación se visualiza en la figura 1.1, en donde para solutos de radios pequeños domina la componente volumétrica de la energía libre de formación de cavidad.

Con respecto a radios mayores a un nanómetro

Consideremos ahora un cúmulo de partículas hidrofóbicas con radio mayor a un nanómetro. Este cúmulo, cuya superficie se extiende con menor curvatura, imposibilita al agua adyacente a mantener su red de enlaces de hidrógeno intactas. En promedio, menos de un enlace de hidrógeno es sacrificado por molécula de agua. Como resultado, el agua tiende a separarse del cúmulo, creando una interfase.

Este proceso involucra cambios significativos en las interacciones moleculares, rompiendo enlaces de hidrógeno para formar la interfase. El gasto energético de este proceso es dominado por la componente de cambios entálpicos.

La energía libre de formación de cavidad es proporcional al área superficial de la interfase. Esto se puede visualizar en la figura 1.1, en donde para solutos de radio mayores a 1 nm , la

componente dependiente del volumen se hace más despreciable al aumentar el radio.

La transición ocurre de forma paulatina en la escala nanoscópica, y por esto se observa una zona de transición centrada aproximadamente en $r = 1 \text{ nm}$.

Fuerza impulsora

La ecuación 1.2 ayuda a explicar la fuerza que impulsa el fenómeno de la hidrofobicidad y la formación de cúmulos.

Una pelota suspendida en el aire caerá al piso en búsqueda de un estado de menor energía. Similarmente, la energía de formación de cúmulos de grandes radios es menor que la suma de las energías libres de solvatación de las n partículas no-polares que formarían el cúmulo. La formación de cúmulos es impulsado por la diferencia en la energía libre de Gibbs entre partículas de menor tamaño (dominan cambios entrópicos) y mayor tamaño (dominan cambios entálpicos), en búsqueda de un estado de menor energía.

Existen dos características físicas que permiten la formación de cúmulos y la creación de la interfase:

1. El solvente está en condiciones muy cercanas a la coexistencia con su fase gaseosa.
2. La interacción soluto-solvente es mucho menos atractiva que la interacción solvente-solvente.

1.2. Energía libre de solvatación

Consideremos una nueva agrupación de términos para el cálculo de la energía libre de solvatación: sus componentes polares y no-polares.

$$\Delta G = \Delta G_{np} + \Delta G_p \quad (1.4)$$

$$\Delta G_{np} = \Delta G_{cav} + \Delta G_{vdw} \quad (1.5)$$

La componente no-polar consiste en el costo energético de la formación de la cavidad, ΔG_{cav} , y la reorganización de las partículas del solvente adyacentes a la interfase. Esta reorganiza-

ción causa interacciones de dispersión en la relación soluto-solvente al generarse dipolos temporales, denotado la energía de Van der Waals Van der Waals Surface (VDW), ver figura 1.2. Las interacciones de Van der Waals son atractivas, por lo tanto VDW es negativo, favoreciendo la solvatación.

La componente polar describe la contribución electrostática, es la energía libre necesaria para cargar el soluto mono o multipolar en un medio dieléctrico.

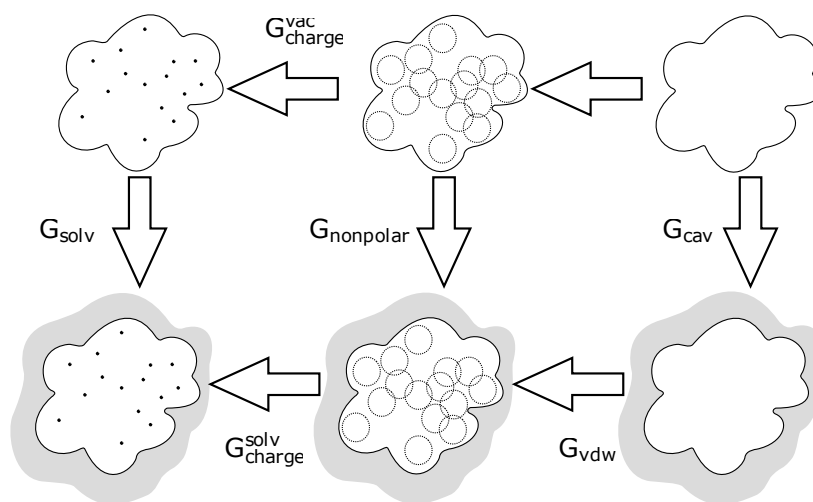


Figura 1.2: Desglose de la energía de solvatación. Fuente: Cooper (2019) [2]

Actualmente los modelos de dinámica molecular de solvente explícito han demostrado generar resultados de alta exactitud [1], sin embargo requieren bastantes recursos para generar resultados, simulando varios nanosegundos en un día de modelamiento computacional con software especializado. Los modelos de solvente implícitos proveen una forma menos exacta pero más eficiente de estudiar estos procesos.

Dentro de la formulación básica del método a utilizar, se debe considerar una forma de calcular la tensión superficial γ . Para una interfase nanoscópica y plana, γ es la tensión superficial del solvente adyacente, dependiente de las interacciones entre el medio y el solvente, y generalmente debilitado por fuerzas de dispersión atractiva y contribuciones electrostáticas de las partículas polares.

En un medio acuoso se ha demostrado que la tensión superficial entre las fases de líquido y vapor γ_{lv} es el valor asintótico de la energía libre de solvatación por unidad de área superficial

para cavidades esféricas de grandes radios [9]. Es decir, para grandes radios $\gamma \equiv \gamma_{lv}$. Las superficies reales, tal como las de biomoléculas, presentan formas irregulares y complejas con mucha variación en curvatura. Para aproximar el valor de la tensión superficial a las longitudes en sistemas nanoscópicos, se utiliza la longitud de Tolman δ , variable que representa la desviación de la tensión superficial en una interfase curva.

1.3. Tensión superficial y la longitud de Tolman

Supongamos que existe un sistema de dos fases, α y β , en equilibrio hidrostático con presión P dentro de un volumen V y separados por una interfase plana A , ver figura 1.3. Ahora imaginemos que por un proceso isotérmico e irreversible aumenta el volumen en dV , empujando la pared trasera más atrás. Entonces también aumenta el área del interfase, $A + dA$. Tal como explica Kondo en el texto "*Molecular Theory of Surface Tension in Liquids*" (1960) [10], el trabajo realizado durante el proceso se divide en dos:

$$dW = pdV - \gamma dA$$

Ahora imaginemos que empujamos la caja en su cara superior e inferior hasta llegar al volumen original, y así el único trabajo realizado es debido al cambio en área, $-\gamma dA$. Luego, la constante γ es la energía necesaria para aumentar el interfase por una unidad de área, llamada tensión superficial. Si γ fuese negativa, el trabajo para aumentar el área del interfase sería positivo y el proceso sería espontáneo, lo que implica estabilidad solo si γ es positivo.

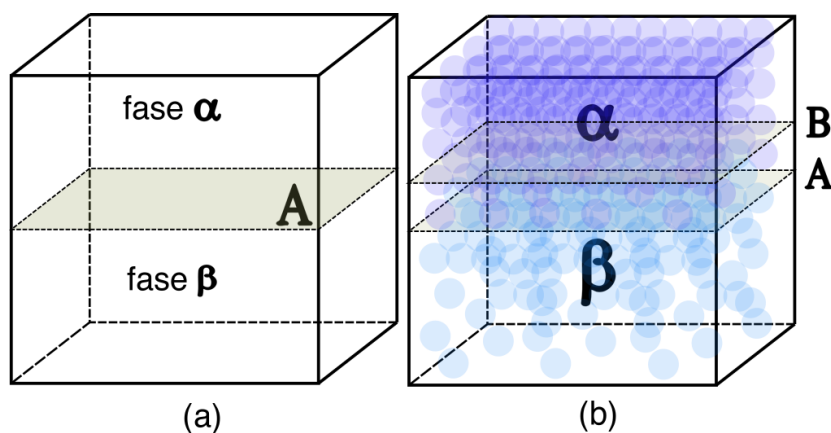


Figura 1.3: Sistema de dos fases, α y β , de un componente con interfase plana. Los planos A y B son superficies divisoras de Gibbs que se encuentran dentro de la zona de interfase.

Dentro del volumen de dos fases en la figura 1.3, ciertos componentes tienden a volverse más densos en la zona de transición. Se llaman **superficie divisora de Gibbs** de Gibbs aquellas que son superficies equidensas, por lo tanto en este caso serian planos paralelos al plano A .

La superficie divisora de Gibbs A divide el volumen en dos, V_α y V_β . Si los volúmenes V_α y V_β fuesen representativos de sus fases en abundancia (*in bulk*), lejos de alguna interfase, entonces las moléculas totales del sistema en conjunto sería $N_\alpha + N_\beta$. Sin embargo, debido al aumento de densidad cerca de la interfase, las moléculas totales son:

$$N_T = N_\alpha + N_\beta + N_S$$

Donde N_S es el aporte de moléculas al sistema debido a la presencia del interfase.

En el caso de una sola componente en dos fases, como agua líquida y vapor, se puede elegir la superficie divisora de Gibbs para cumplir la condición $N_S = 0$. Esta superficie en particular se le llama la *superficie divisora equimolar* definida por $r = a_v$, y es representada por el plano B en la figura 1.3 (b).

Consideremos ahora una interfase esférica, ver figura 1.4. Comparado con la interfase plana, este sistema no puede existir con una presión constante. En una condición de equilibrio se establece la **relación de Young-Laplace**:

$$p_\alpha - p_\beta = \frac{2\gamma}{a} \quad (1.6)$$

Donde p_α y p_β corresponden a la presión de cada fase y a es el radio de la interfase esférica. Es importante reconocer que la relación 1.6 es una aproximación en la mira nanoscópica debido a que a no es una interfase claramente definida, sino que existe una zona de transición.

La ecuación 1.7 es una generalización de la relación de Young-Laplace, y es válida a lo largo de la zona de transición llamada interfase [11].

$$\gamma + \frac{1}{2}a \left[\frac{\partial\gamma}{\partial a} \right] da = \frac{1}{2}a (p_\alpha - p_\beta) \quad (1.7)$$

El término $[\partial\gamma/\partial a] da$ es el cambio en la tensión superficial γ que se produce debido al

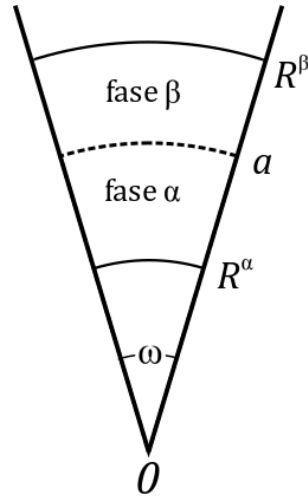


Figura 1.4: Sistema de dos fases de un componente con interfase curva.

desplazamiento matemático de la superficie divisora por una cantidad da , manteniendo constante las cantidad físicas dentro del sistema y las condiciones externas. Este término no se debe confundir con la expresión $\partial\gamma/\partial a$, "la derivada para la dependencia del radio actual" [10].

Sea a_s el radio correspondiente a la superficie divisora de Gibbs que anula el término $[\partial\gamma/\partial a]$ de la ecuación 1.7. Esta superficie divisora de Gibbs se llama *superficie de tensión*. Si y solo si $r = a_s$ se cumple la relación de Young-Laplace, ecuación 1.6:

$$p_\alpha - p_\beta = \frac{2\gamma_s}{a_s}$$

Esto implica que el efecto mecánico de la zona de transición en presencia de una interfase esférica y compleja se puede interpretar (o reemplazar matemáticamente) con una membrana flexible de grosor cero, y con tensión γ_s uniforme, ubicada en la superficie de tensión.

Finalmente, la longitud de Tolman es la distancia entre la superficie equimolar y la superficie de tensión:

$$\delta = a_v - a_s \quad (1.8)$$

Datos empíricos utilizando modelos de dinámica molecular apuntan a que δ es un valor positivo para una interfase de agua, líquido-vapor [12] [13], y esperamos que su magnitud sea de tamaño molecular (Tolman propuso un valor aproximado de 1 \AA [6]), sin embargo, aún es debatido su forma y valor.

1.4. Estudios actuales sobre la forma y valor de la longitud de Tolman

Tolman desarrolló la base matemática para la siguiente ecuación que describe el cambio en la tensión superficial debido a la curvatura del interfase, la ecuación **Gibbs-Tolman-Koenig-Buff (GTKB)** [11]:

$$\frac{1}{\gamma} \frac{d\gamma}{dr} = \frac{[2\delta/r^2] [1 + (\delta/r) + 1/3(\delta^2/r^2)]}{1 + [2\delta/r] [1 + (\delta/r) + 1/3(\delta^2/r^2)]} \quad (1.9)$$

en donde γ es la tensión superficial para una gota de fluido dentro con curvatura $1/r$. El parámetro δ es la longitud de Tolman. Las siguientes aproximaciones de la tensión superficial, simplificaciones de la ecuación GTKB, son utilizadas a menudo [4] [14]:

$$\gamma = \frac{\gamma_{lv}}{1 + \frac{2\delta_1}{r}} \quad (1.10)$$

$$\gamma = \gamma_{lv} \left(1 - \frac{2\delta_2}{r} \right) \quad (1.11)$$

$$\gamma = \gamma_{lv} \left(\frac{1}{1 + 1,42\delta_3/r} \right)^{1,45} \quad (1.12)$$

$$\gamma = \gamma_{lv} e^{-2\delta_4/r} \quad (1.13)$$

Estas ecuaciones son derivadas de la ecuación GTKB considerando distintos supuestos, el más común es que el valor de δ/r es tan pequeño que se considera insignificante.

Fuente	Año	Valor de Tolman, Å	Comentario
Holten [15]	2005	0.6	Teoría clásica de la nucleación (CNT)
Azouzi [16]	2013	-0.47	Cavitación, CNT, presión negativa
Homman [17]	2014	-0.04	Teoría funcional de la densidad (DFT)
Lu [18]	2005	0.36	Comparación de interfaces sólidas y líquidas

Fuente	Año	Valor de Tolman, Å	Comentario
Joswiak [19]	2013	-0.56	TIP4P/2005, CNT
Wilhelmsen [20]	2015	-0.5	CNT
Sedlmeier [21]	2012	-1.0	Simulaciones de dinámica molecular
Malek [22]	2019	0.55	Simulaciones de dinámica molecular
Kim [23]	2018	2.1	Condensación capilar

Tabla 1.1: Valores de la longitud de Tolman para una interfaz de agua líquido-vapor de distintas fuentes. Adaptado de Kalova et al. (2019) [4]

1.5. Generación de mallas y base de datos

Existen varias formas de definir una superficie molecular, incluyendo a la superficie SAS, Solvent Excluded Surface (SES), VDW, Gaussian Surface, etc. La superficie VDW se define como la unión de las superficies atómicas con el radio de Van der Waals de cada átomo en la molécula. Ambas SAS y SES son generadas al rodar una probeta esférica sobre VDW, como muestra la figura 1.5. La probeta representa una sola molécula disolvente, típicamente una molécula de agua con radio $r \approx 1,4$ Å.

El trabajo fue llevado a cabo con mallas triangulares creadas por el software MSMS para 502 especies, el cuál utiliza una probeta virtual de radio igual a 1,4 Å. para generar una superficie SAS. Debido al tamaño del estudio de Mobley et al. y el poco error demostrado, decidimos trabajar con la base de datos resultantes de esta investigación, la que incluye valores de ΔG_{np} obtenidos mediante simulaciones de dinámica molecular. Además, obtenemos valores de ΔG_{vdw} del estudio de Cooper (2019) [2].

Los archivos generados por el software MSMS contienen un listado de la posición en tres dimensiones de cada nodo en la malla, separados en filas, y además otro archivo con un listado de los nodos que forman cada triángulo. El estudio se llevará a cabo en el lenguaje de programación Python 3.7. Utilizamos principalmente las librerías NumPy para cálculos matriciales y Trimesh 3.4.0 para el análisis de mallas triangulares impermeables.

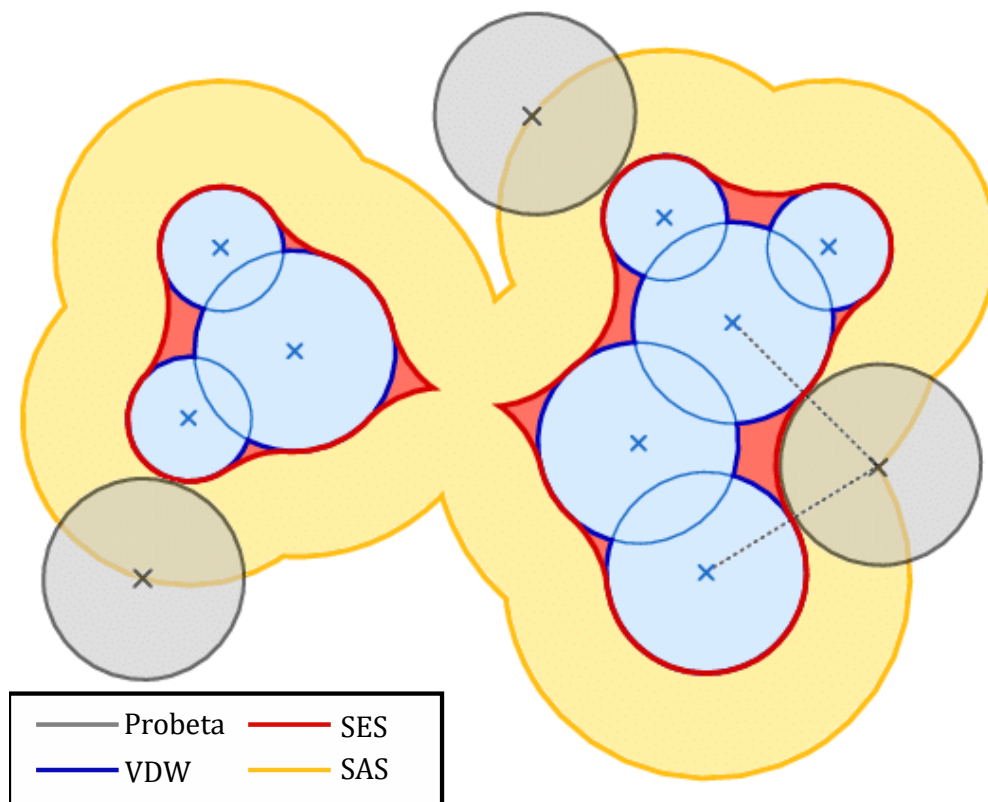


Figura 1.5: Esquema de definiciones de superficies moleculares y la probeta que las genera, en 2D. Adaptado de Kozlikova (2017) [24].

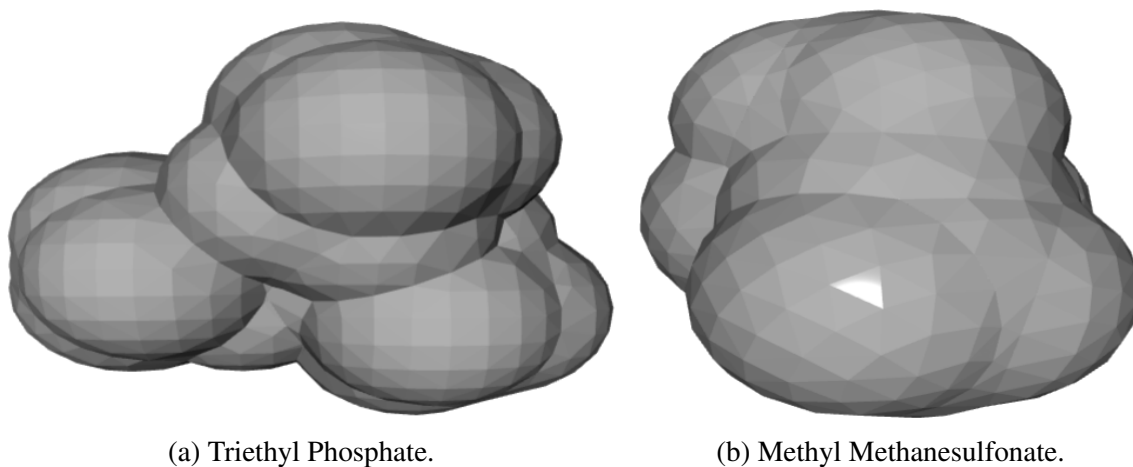


Figura 1.6: Mallas triangulares representativas de la superficie SAS, generadas por el software MSMS, visualizada por medio de la librería Trimesh de Python3.

2 | Desarrollo

El desarrollo de la problemática no debiese restringir el resultado del ajuste de la variable de Tolman, para así poder analizar libremente cuál simplificación de la ecuación [Gibbs-Tolman-Koenig-Buff \(GTKB\)](#), u otro modelo, aproxima mejor su valor y forma dado los resultados de Mobley et al. para 502 especies utilizando dinámica molecular.

2.1. Reproduciendo resultados de Chandler(2005) y análisis de variables

Para identificar y determinar las variables involucradas es importante poder reproducir los resultados de Chandler [\[3\]](#), ver figura [1.1](#).

La tensión superficial para una superficie plana de agua en abundancia es conocida, $\gamma_{lv} = 72,75 \times 10^{-3} [N/m]$. Utilizamos la misma simplificación de la ecuación [GTKB](#) utilizada por Chandler, ecuación [1.11](#).

Las simulaciones para aproximar el valor de Gibbs son llevadas a cabo a presión constante para permitir de manera artificial el cambio de volumen asociado al fenómeno de superficies secas y mojadas en donde el interfase se acerca a moléculas hidrofílicas y se aleja de moléculas hidrofóbicas [\[25\]](#) [\[13\]](#). Los resultados de Chandler consideran condiciones ambientales, $T = 298K$ y $P = 1 atm$. En esta condición el agua líquida está cercana al punto de coexistencia con su vapor, y el vapor puede ser considerado un gas ideal, lo que facilita algunas simulaciones de dinámica molecular.

La figura [2.1](#) muestra una reproducción del gráfico de Chandler considerando la longitud de Tolman constante e igual a $1,0 \text{ \AA}$, el valor propuesto por el mismo Tolman [\[6\]](#). Este valor es importante ya que ha sido propuesto por distintas fuentes que el valor de la longitud de Tolman está en la escala molecular [\[13\]](#) [\[10\]](#).

2.2. Propiedades geométricas en mallas triangulares

Para extender nuestro análisis a especies de múltiples átomos, debemos poder determinar cómo la noción de la curvatura se interpreta en nodos y caras de mallas triangulares. Los resultados de Mobley et al. incluyen 502 especies, cuyas geometrías están definidas por mallas triangulares compuestas por entre 328 y 1.247 nodos.

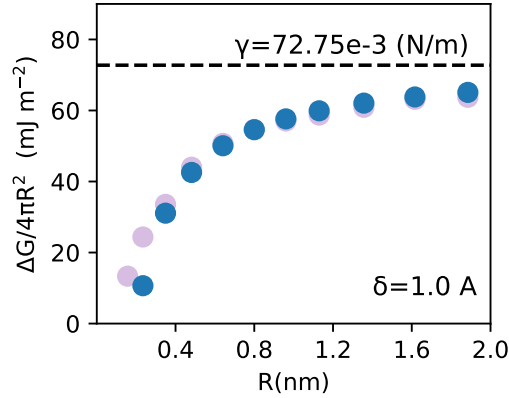


Figura 2.1: Serie azul: Aproximación de energía libre de solvatación para la formación de una cavidad esférica en agua, como función del tamaño de la cavidad. Serie lila: Resultados de Chandler (2005) [3].

Utilizando el factor de corrección a la tensión superficial para calcular la energía de formación de cavidad, consideramos la siguiente relación para cada especie, donde n es el número de nodos en su malla:

$$\Delta G_{cav} \approx \sum_{i=0}^n A_i \gamma_{lv} f(R)_i + V \cdot P \quad (2.1)$$

A cada nodo se le asigna un radio y un área correspondiente. Además, consideramos la convención en donde las curvaturas son determinadas positivas para superficies convexas y negativas para superficies cóncavas, por lo tanto buscamos la orientación de cada radio de la malla.

Las mallas triangulares están definidas por las coordenadas de sus nodos y la agrupación de nodos que crea cada triángulo en particular. Así, cada nodo tendrá por lo menos tres nodos vecinos, los que se definen como el **one-ring neighborhood** del nodo principal. Los triángulos que rodean a un nodo son los triángulos del one-ring neighborhood. Calculamos el área por nodo como un tercio del área total del one-ring neighborhood.

$$A_j = \frac{1}{3} \sum_{1-ring} A_i \quad (2.2)$$

El radio por nodo lo definimos por el método del mejor ajuste o *best fit* de una esfera. Para cuatro o más puntos en tres dimensiones se puede definir una esfera del mejor ajuste. Ya que cada nodo de la malla tiene por lo menos tres nodos vecinos en su one-ring neighborhood, podemos siempre determinar cuál es la esfera de mejor ajuste para este grupo de nodos. La metodología para calcular el centro y radio de una esfera de mejor ajuste es detallado en el anexo B.

La normal de una cara triangular puede ser definida claramente, sin embargo, debemos adoptar una forma de definir la pseudo-normal para cada nodo. Una solución común es igualar la pseudo-normal al promedio de las normales de las caras adyacentes al nodo. Para este análisis adoptamos el método [Angle Weighted Pseudo Normal \(AWPN\)](#), en donde el promedio se calcula ponderando la normal de cada triangulo adyacente por el ángulo de la cara en el vértice, como es ilustrado en la figura 2.2a.

$$n_\alpha = \frac{\sum \alpha_i n_i}{\|\sum \alpha_i n_i\|} \quad (2.3)$$

Para conocer la orientación del radio, y así determinar si corresponde a una curvatura cóncava o convexa, podemos comparar el vector del radio, ecuación 2.4, con la pseudo-normal. Si el ángulo formado entre estos dos vectores es mayor a 90 grados, entonces la superficie se determina cóncava, y en caso contrario, convexa.

$$v_R = p_i - p_R \quad (2.4)$$

2.3. Aproximación de la longitud de Tolman

2.3.1. Tolman Característico

Como primer intento, aproximamos un valor de Tolman característico por especie $\bar{\delta}$. Para lograr esto, no consideramos el aspecto acumulativo del término entrópico de la energía de formación de cavidad. Además debemos determinar un radio característico, \bar{R} . El radio característico se iguala al promedio del radio de cada nodo ponderado por el área superficial que le corresponde, ecuación 2.11.

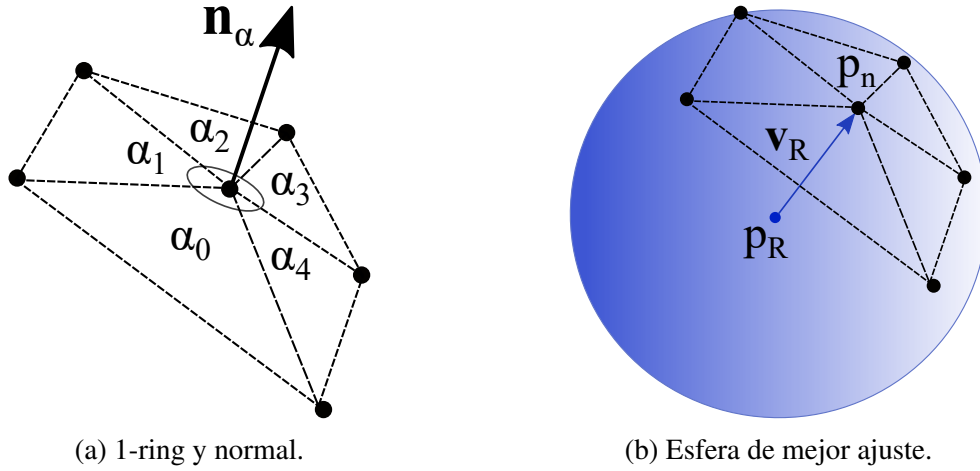


Figura 2.2: Para cada nodo, la normal n_α considera el ángulo α creado por los triángulos vecinos del punto principal p_n . La curvatura en el nodo depende de la orientación de n_α y v_R .

$$\bar{\delta}_1 = (1/\xi - 1) \cdot \frac{R}{2} \quad (2.5)$$

$$\bar{\delta}_2 = (1 - \xi) \cdot \frac{R}{2} \quad (2.6)$$

$$\bar{\delta}_3 = (1/\xi^{1/1,45} - 1) \cdot \frac{R}{1,42} \quad (2.7)$$

$$\bar{\delta}_4 = (\ln(\xi)) \cdot \frac{R}{2} \quad (2.8)$$

donde (2.9)

$$\xi = \frac{\Delta G_{cav} - VP}{S \cdot \gamma_{lv}} \quad (2.10)$$

$$\bar{R} = \frac{\sum A_n R_n}{\sum A_n} \quad (2.11)$$

2.3.2. Factor de corrección definido por Nodo

Contamos con una base de datos de 502 especies, para cada una tenemos el valor de la energía de solvatación para la formación de cavidad ΔG_k y volumen V_k . Existe una malla triangular para cada especie con n_k nodos, y para cada nodo definimos un radio R_j y área asociada A_j . La dependencia de Tolman con la curvatura ha sido demostrada teóricamente por medio de

teoría termodinámica y de estadística mecánica [12] [11], por lo tanto buscamos conocer el comportamiento del factor que corrige la tensión superficial $f(R)$.

$$\Delta G_k \approx \sum_{j=1}^{n_k} \gamma_{lv} \cdot f(R_j) \cdot A_j + P \cdot V_k \quad (2.12)$$

$$\frac{\Delta G_k - P \cdot V_k}{\gamma_{lv}} \approx \sum_{j=1}^{n_k} f(R_j) \cdot A_j \quad (2.13)$$

Siendo $f(R_j)$ uno de los factores de corrección de la tensión superficial:

$$f_1(R_j) = \left(1 - \frac{2\delta_j}{R_j}\right) \quad (2.14)$$

Este sistema de ecuaciones no es cuadrado; a cada especie le corresponde una ecuación de n_k incógnitas de la siguiente forma:

$$\xi_k \approx f(R_0) \cdot A_0 + f(R_1) \cdot A_1 + f(R_2) \cdot A_2 + \dots + f(R_{n_k}) \cdot A_{n_k}$$

donde $\xi_k = \frac{\Delta G_k - P \cdot V_k}{\gamma_{lv}}$

Para las 502 especies de Mobley et al., el número de nodos n_k varía entre 328 y 1247 dependiendo del tamaño de la especie y la complejidad de su geometría. Existen 502 ecuaciones y aproximadamente 388.000 incógnitas (tantas incógnitas como nodos individuales en todas las especies).

Abordaremos el problema ajustando las incógnitas para forzar un sistema con dimensiones que permitan el uso de matrices. Definimos intervalos de R_j en donde la variable $f(R_j)$ se mantiene constante. De esta forma, si hay 10 intervalos, la ecuación para cada especie sería la siguiente:

$$\xi_k \approx f(R_{k,b0}) \cdot A_{k,b0} + f(R_{b1}) \cdot A_{k,b1} + f(R_{b2}) \cdot A_{k,b2} + \dots + f(R_{b10}) \cdot A_{k,b10} \quad (2.15)$$

en donde $f(R_{b1})$ corresponde al factor de corrección de tensión superficial para los radios $R_{b1,min} \leq R_j \leq R_{b1,max}$, y A_{b1} es la suma de las áreas que corresponden a este intervalo de radios. El sistema matricial a resolver es el siguiente para l intervalos:

$$\begin{bmatrix} A_{0,b0} & A_{0,b1} & \dots & A_{0,bl} \\ A_{1,b0} & A_{1,b1} & \dots & A_{1,bl} \\ A_{2,b0} & A_{2,b1} & \dots & A_{2,bl} \\ A_{3,b0} & A_{3,b1} & \dots & A_{3,bl} \\ \vdots & \vdots & \ddots & \vdots \\ A_{501,b0} & A_{501,b1} & \dots & A_{501,bl} \end{bmatrix} \cdot \begin{bmatrix} f(R_{b0}) \\ f(R_{b1}) \\ \vdots \\ f(R_{bl}) \end{bmatrix} = \begin{bmatrix} \xi_0 \\ \xi_1 \\ \xi_2 \\ \xi_3 \\ \vdots \\ \xi_{501} \end{bmatrix}$$

$$[A] [X] = [b]$$

Este sistema se puede resolver mientras los intervalos, y por ende las incógnitas, sean igual o menor a 502. Esta solución, a través de un sistema linear de múltiples variables, es atractiva ya que no se asume ninguna relación entre las incógnitas, aunque en efecto sea una sola variable $f(R)$.

Consideramos dos formas de crear intervalos; Los *bins* se definen como segmentos equidistantes de radio. En un *bin*, la densidad, o cantidad de datos que corresponden a este intervalo, es variable. Pueden, también, existir intervalos con densidad igual a cero. Los datos también se pueden dividir en cuantiles, o segmentos de densidad constantes. La figura 2.3 muestra el 99.7 % de todos los datos, no se visualizan 1349 radios. Se puede apreciar que existe una concentración de datos alrededor de los 3 angstrom. De aquí nace la necesidad de considerar la estadística robusta, métodos estadísticos que no han sido indebidamente afectados por valores aberrantes. La [regla de Tukey](#) indica que los valores aberrantes están alejados en más de 1.5 veces el [rango intercuartil \(IQR\)](#), o diferencia entre el primer y tercer cuartil.

$$Q_1 - 1,5 \cdot IQR \leq R \leq Q_3 + 1,5 \cdot IQR \quad (2.16)$$

En la figura 2.3, los límites de la regla de Tukey están destacados con una línea roja.

Por lo tanto existen cuatro tipos de intervalos a evaluar:

- Intervalos equidistantes sin eliminación de datos.
- Intervalos equidistantes con eliminación de valores aberrantes por la regla de Tukey.
- Intervalos equidensos sin eliminación de datos.
- Intervalos equidensos con eliminación de valores aberrantes por la regla de Tukey.

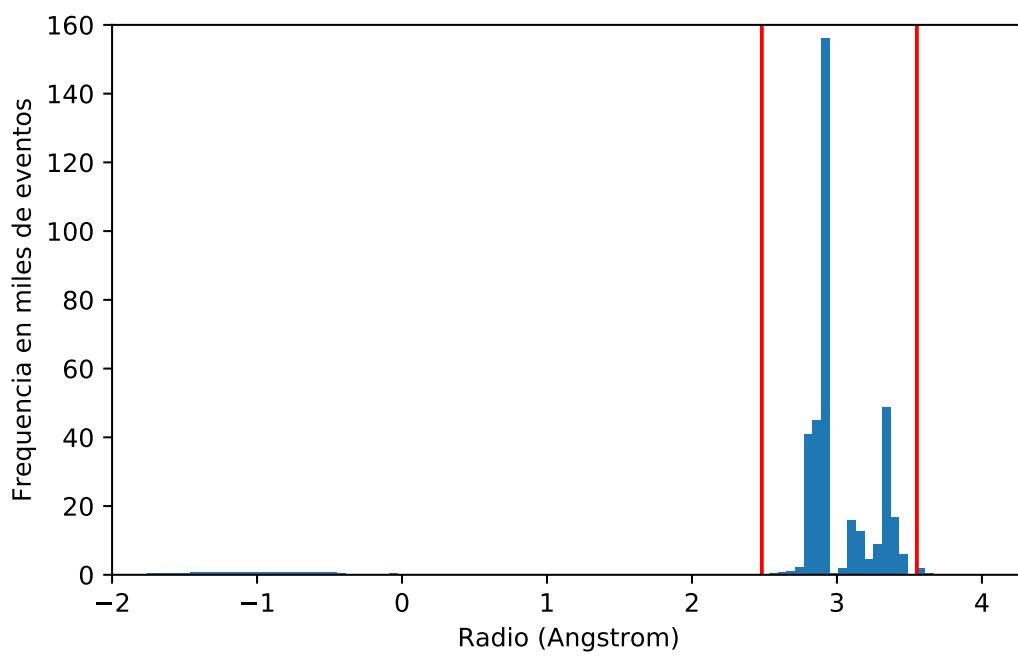
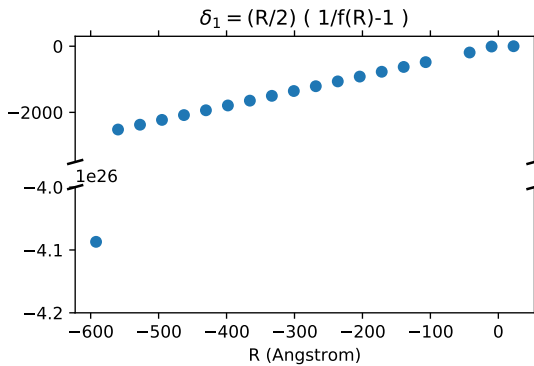
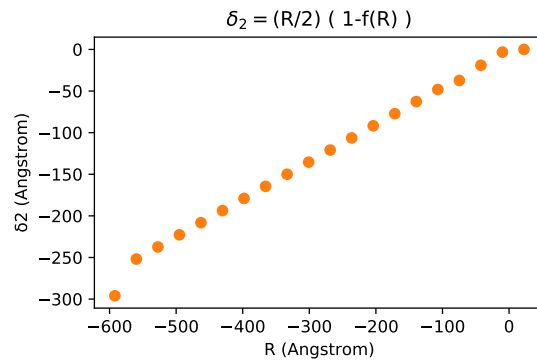


Figura 2.3: Histograma de radios de cada especie de Mobley et al., calculados en cada nodo por el método de esfera de mejor ajuste. Los radios negativos son asociados a superficies cóncavas y los positivos a superficies convexas. Las líneas rojas destacan los límites de estadística robusta por la regla de Tukey. Se visualizan el 99.7 % del espectro total de radios.

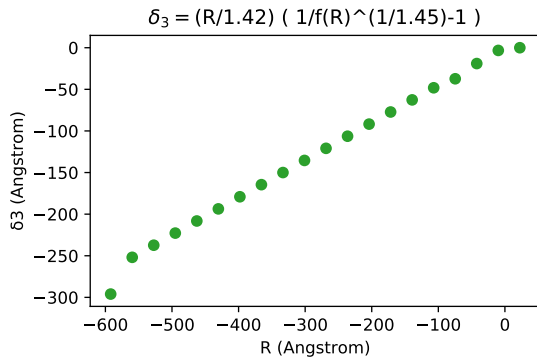
Obtenemos la solución de la ecuación 2.15 para 20 intervalos usando mínimos cuadrados con restricciones de 0 a 1, dando como resultado el valor de $f(R)$, ver anexo C. Luego, utilizando las simplificaciones para GTKB, utilizamos el resultado de $f(R)$ para resolver por la variable de Tolman δ . A continuación las figuras 2.4, 2.5, 2.6 y 2.7 muestran las soluciones del valor de Tolman obtenidas.



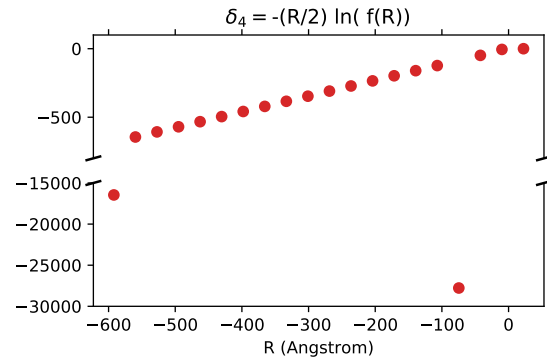
(a) Solución de la variable Tolman por ecuación 1.10.



(b) Solución de la variable Tolman por ecuación 1.11.



(c) Solución de la variable Tolman por ecuación 1.12.

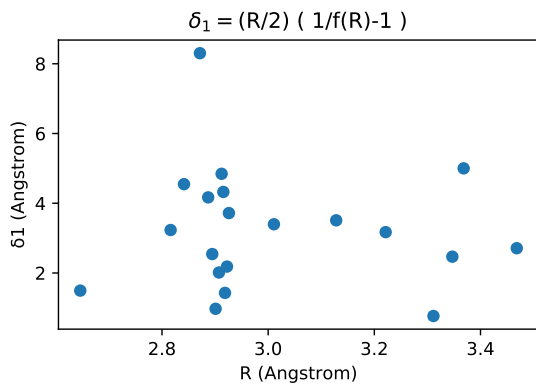


(d) Solución de la variable Tolman por ecuación 1.13.

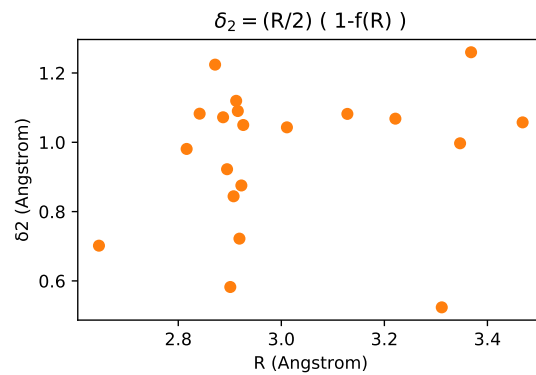
Figura 2.4: Ajuste de variable de Tolman usando la solución de $f(R)$ experimental calculado con el sistema de ecuaciones 2.13, usando 20 intervalos de radio equidistantes, considerando todo el espectro de radios disponible para las 502 especies.

2.4. Capacidad de predicción del método

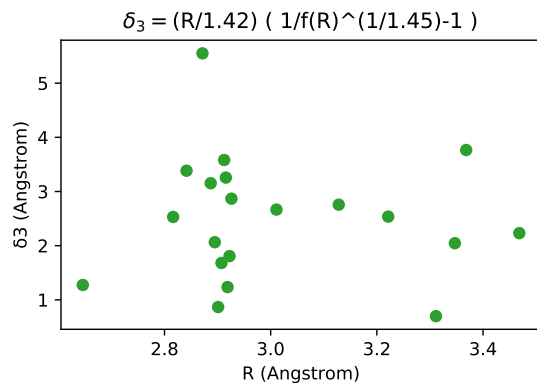
Para determinar la exactitud con la cual los métodos utilizados pueden predecir el valor de la energía de formación de cavidad por especie, se genera la solución por mínimos cuadrados de $f(R)$ para una muestra de k especies. En base a estos resultados construimos dos tipos de



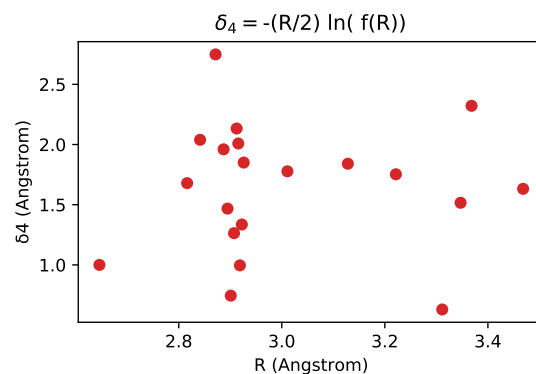
(a) Solución de la variable Tolman por ecuación 1.10.



(b) Solución de la variable Tolman por ecuación 1.11.

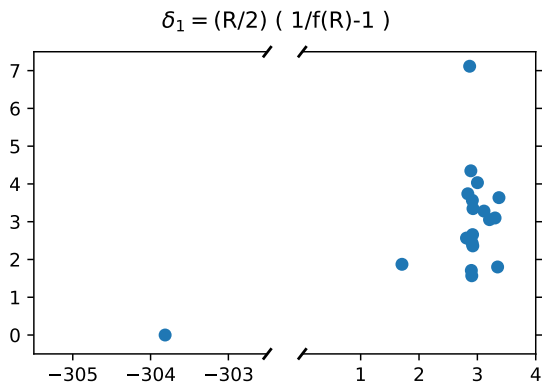


(c) Solución de la variable Tolman por ecuación 1.12.

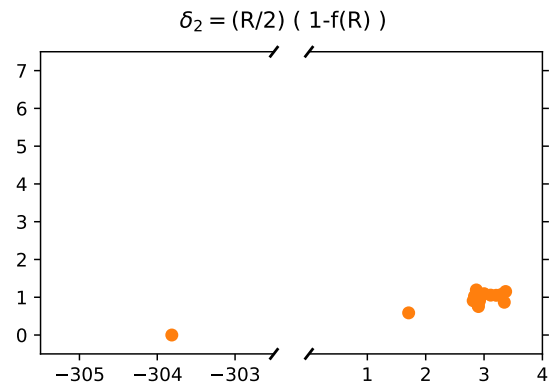


(d) Solución de la variable Tolman por ecuación 1.13.

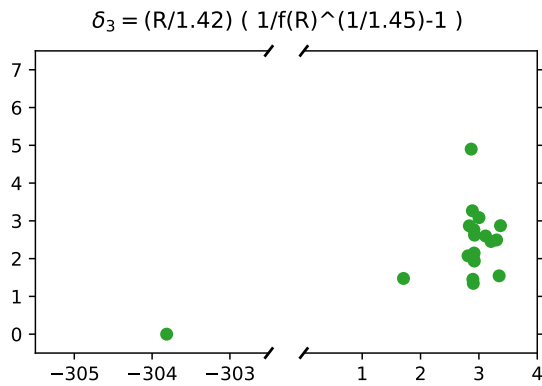
Figura 2.5: Ajuste de variable de Tolman usando la solución de $f(R)$ experimental calculado con el sistema de ecuaciones 2.13, usando 20 intervalos de radio equidistantes, eliminando radios del espectro que se consideran datos aberrantes.



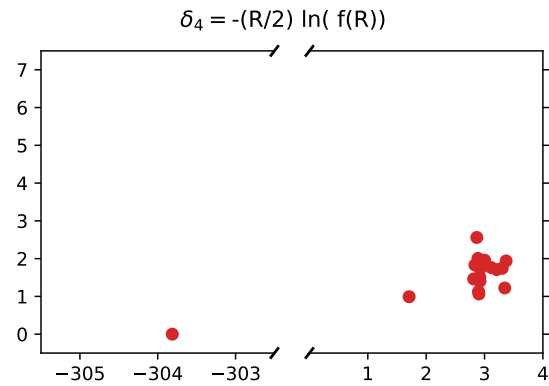
(a) Solución de la variable Tolman por ecuación 1.10.



(b) Solución de la variable Tolman por ecuación 1.11.

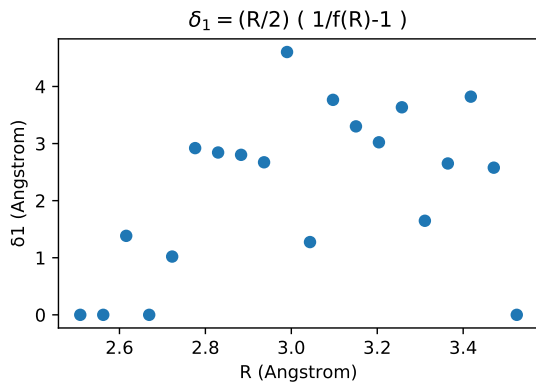


(c) Solución de la variable Tolman por ecuación 1.12.

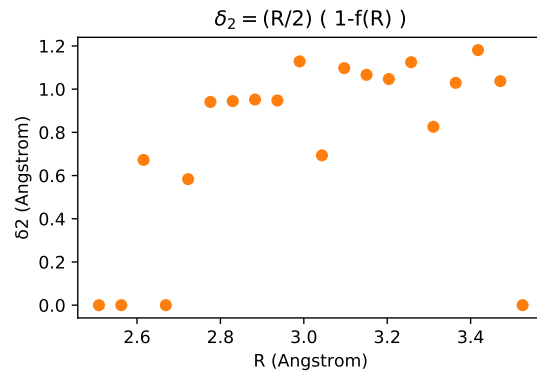


(d) Solución de la variable Tolman por ecuación 1.13.

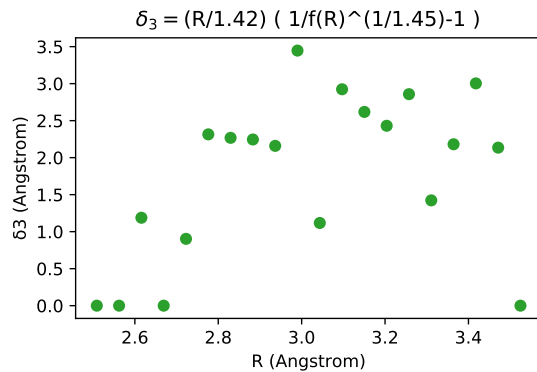
Figura 2.6: Ajuste de variable de Tolman usando la solución de $f(R)$ experimental calculado con el sistema de ecuaciones 2.13, usando 20 intervalos de radio equidensos, considerando todo el espectro de radios disponible para las 502 especies.



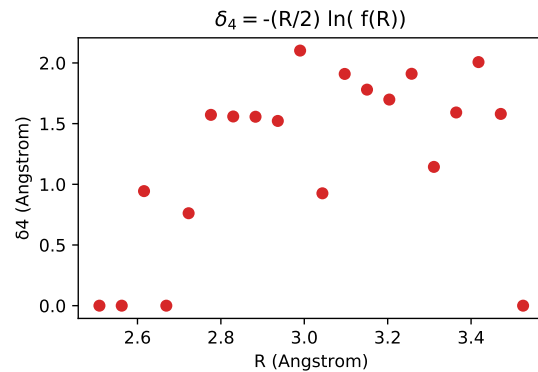
(a) Solución de la variable Tolman por ecuación 1.10.



(b) Solución de la variable Tolman por ecuación 1.11.



(c) Solución de la variable Tolman por ecuación 1.12.



(d) Solución de la variable Tolman por ecuación 1.13.

Figura 2.7: Ajuste de variable de Tolman usando la solución de $f(R)$ experimental calculado con el sistema de ecuaciones 2.13, usando 20 intervalos de radio equidensos, eliminando radios del espectro que se consideran datos aberrantes.

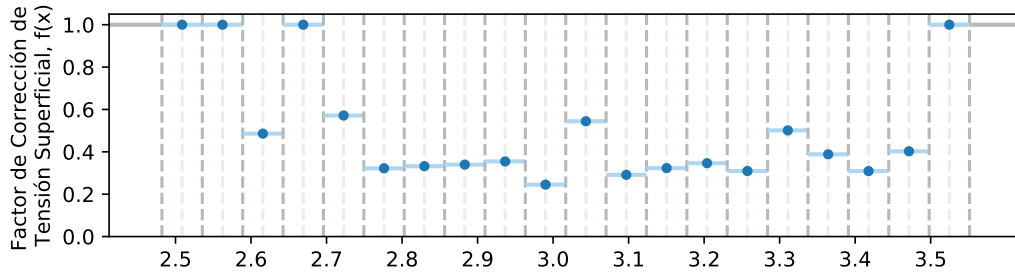


Figura 2.8: Construcción de la función de $f(R)$ constante por partes a partir de los resultados de la ecuación matricial 2.15 para intervalos equidistantes (líneas grises) con eliminación de datos aberrantes, ver figura C.2 en anexo C.

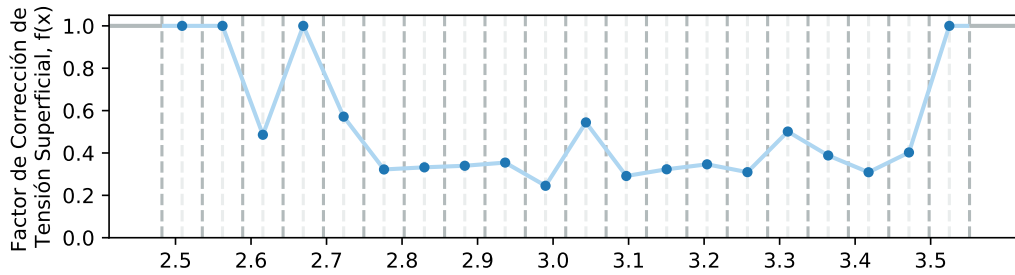


Figura 2.9: Construcción de la función de $f(R)$ lineal por partes a partir de los resultados de la ecuación matricial 2.15 para intervalos equidistantes (líneas grises) con eliminación de datos aberrantes, ver figura C.2 en anexo C.

funciones, constante por partes o lineal por partes, figuras 2.8 y 2.9 respectivamente. Luego utilizamos estas funciones con respecto al radio para evaluar la energía Gibbs de cavitación y para las restantes $(502 - k)$ especies.

Las soluciones de la ecuación 2.15 son ajustes del factor de corrección de la tensión superficial, que indican su desviación del valor planar, ecuación 2.17.

$$\gamma = \gamma_{lv} \cdot f(R) \quad (2.17)$$

Estas predicciones dan información sobre cuál de las 13 soluciones encontradas para calcular $f(R)$ es más exacto, sin embargo no nos entrega información sobre cuál de las simplificaciones de la ecuación GTKB es más exacta.

Consideramos además un modelo en donde el valor de Tolman sea constante. Para encontrar

el valor constante más adecuado, calculamos Tolman característico para las cuatro simplificaciones de GTKB, ajustando el resultado con los resultados conocidos de Mobley et al. (2009).

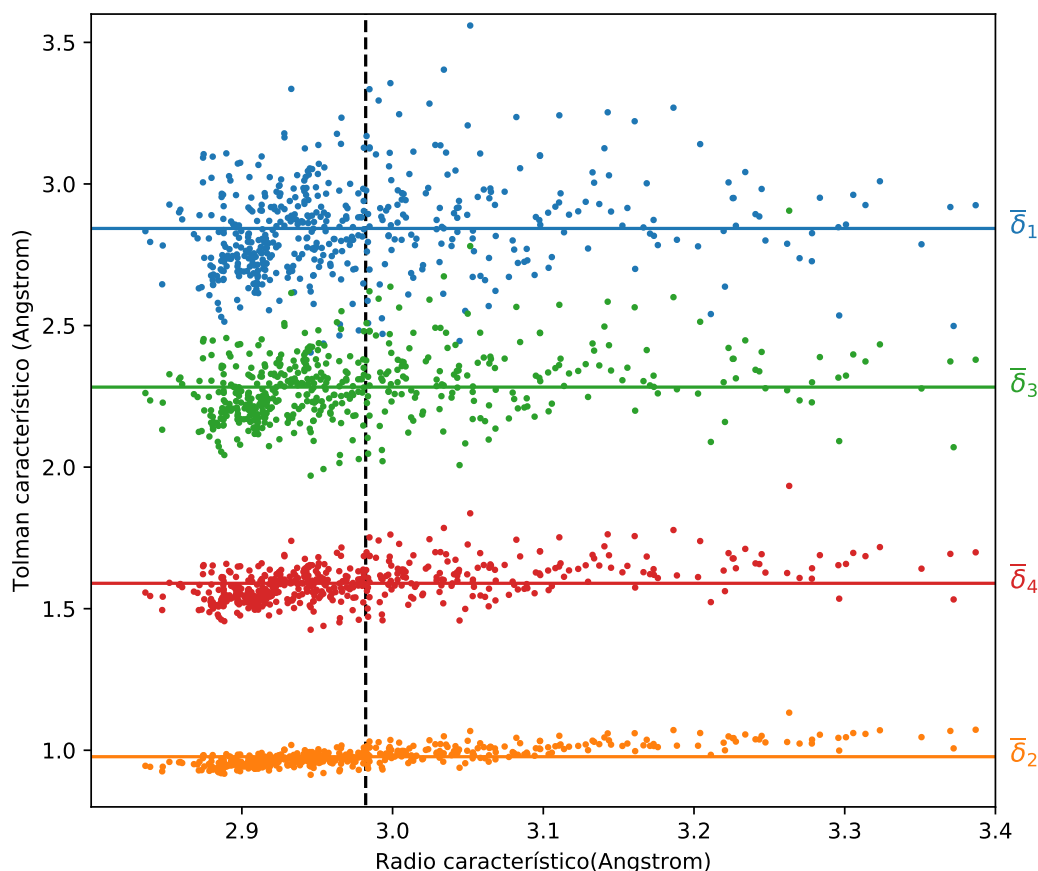


Figura 2.10: Valor de Tolman característico calculado para las 502 especies según ecuaciones 2.10 para la energía de solvatación de formación de cavidad. Las líneas rectas indican valores promedios.

Determinamos los valores de Tolman característicos, por especie, utilizando cuatro definiciones de tensión superficial, simplificaciones de la ecuación GTKB, ecuaciones 1.10–1.13. Sus valores promedio a continuación:

$$\bar{\delta}_1 = 2,84\text{Å}$$

$$\bar{\delta}_2 = 0,98\text{Å}$$

$$\bar{\delta}_3 = 2,28\text{Å}$$

$$\bar{\delta}_4 = 1,59\text{Å}$$

Para poder comparar estos resultados, se destaca que para Tolman constante igual a 0.98 angstrom, correspondiente al valor promedio de Tolman característico según ecuación 1.11, la reproducción del gráfico de Chandler es visualmente muy atractiva, figura 2.11.

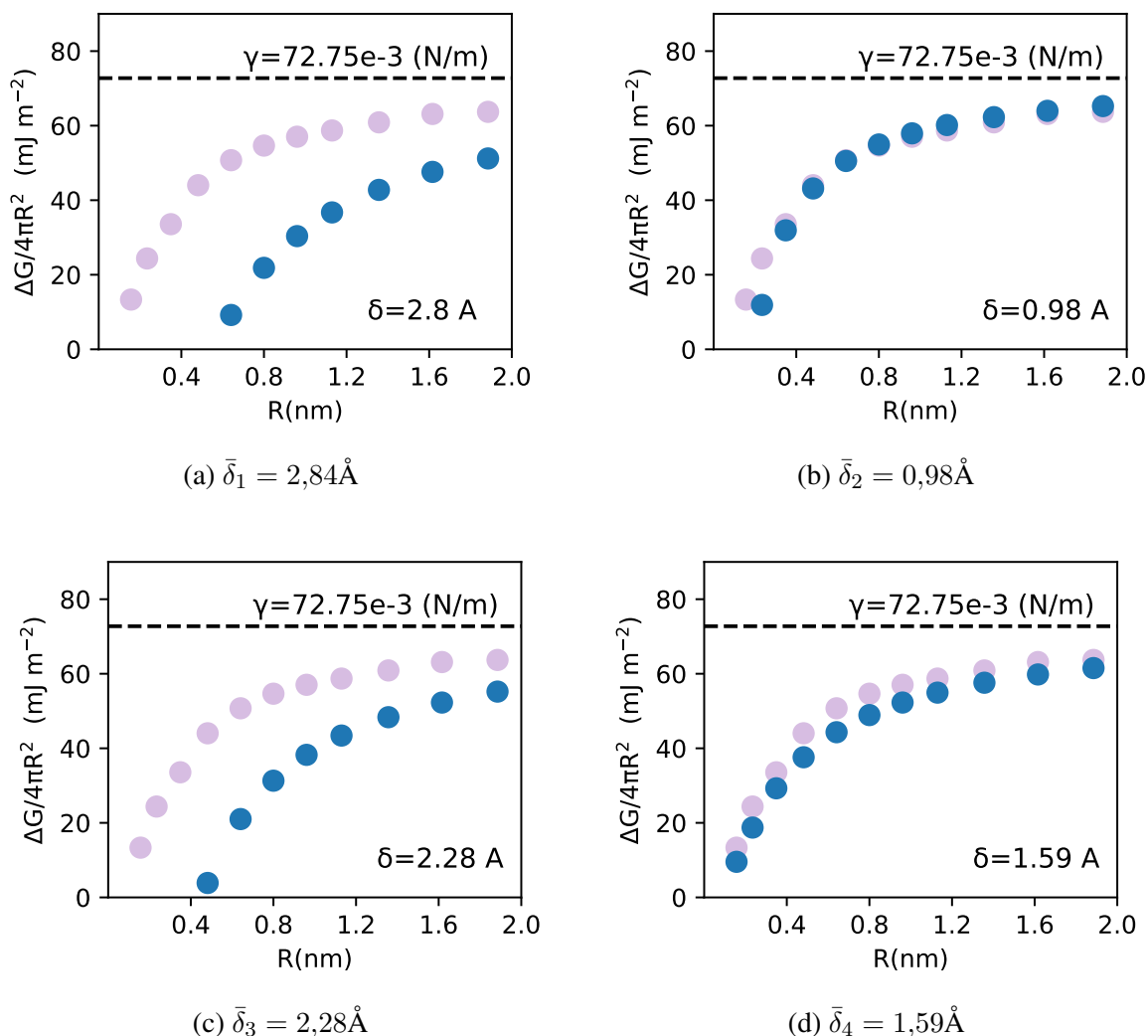


Figura 2.11: Serie azul: Aproximación de energía libre de solvatación para la formación de una cavidad esférica en agua, como función del tamaño de la cavidad. Se considera Tolman constante según resultados en sección 2.3.1. Serie lila: Resultados de Chandler (2005) [3].

3 | Resultados

Consideramos toda combinación de los siguientes métodos para aproximar el factor de corrección de la tensión superficial, $f(R)$:

- Para la solución del problema matricial por mínimos cuadrados, ecuación 2.15, considerar intervalos equidistantes (*bins*) o equidensos (cuantiles).
- Eliminación de valores aberrantes por medio de la regla de Tukey de estadística robusta (RS) o sin eliminación de datos (none).
- Construcción de una función para aproximar el factor de corrección de tensión superficial $f(R)$, constante por partes o lineal por partes.
- En los métodos en donde implementamos estadística robusta, considerar $f(R) = 0$ para radios fuera de los intervalos considerados (0), o que tomen el valor del intervalo del borde más cercano (1).
- Consideramos calcular la energía de formación de cavidad utilizando el promedio de los valores de Tolman característicos obtenidos por especie, utilizando las simplificaciones de $f(R)$ para la ecuación GTKB, ecuaciones 1.10–1.13.

Para cada grupo de datos obtenemos el coeficiente de correlación de Pearson, indicando cuál modelo posee una mejor correlación lineal con los valores experimentales esperados. Repetimos esto para una predicción utilizando 20, 50 y 100 especies, para cada uno de los 16 modelos.

Los valores experimentales utilizados para evaluar la eficiencia de los métodos probados vienen del cálculo de la energía no polar ΔG_{np} de Mobley et al. (2009) [1], obtenidos por simulaciones de dinámica molecular (RMSE de 1,24 kcal/mol y correlación de 0.89). Luego, a estos valores les restamos la energía libre asociada a fuerzas débiles de Van der Waals ΔG_{vdw} , calculado por Cooper y Bardhand (2019) [2].

$$\Delta G_{cav, experimental} = \Delta G_{np, Mobley} - \Delta G_{vdw, Cooper}$$

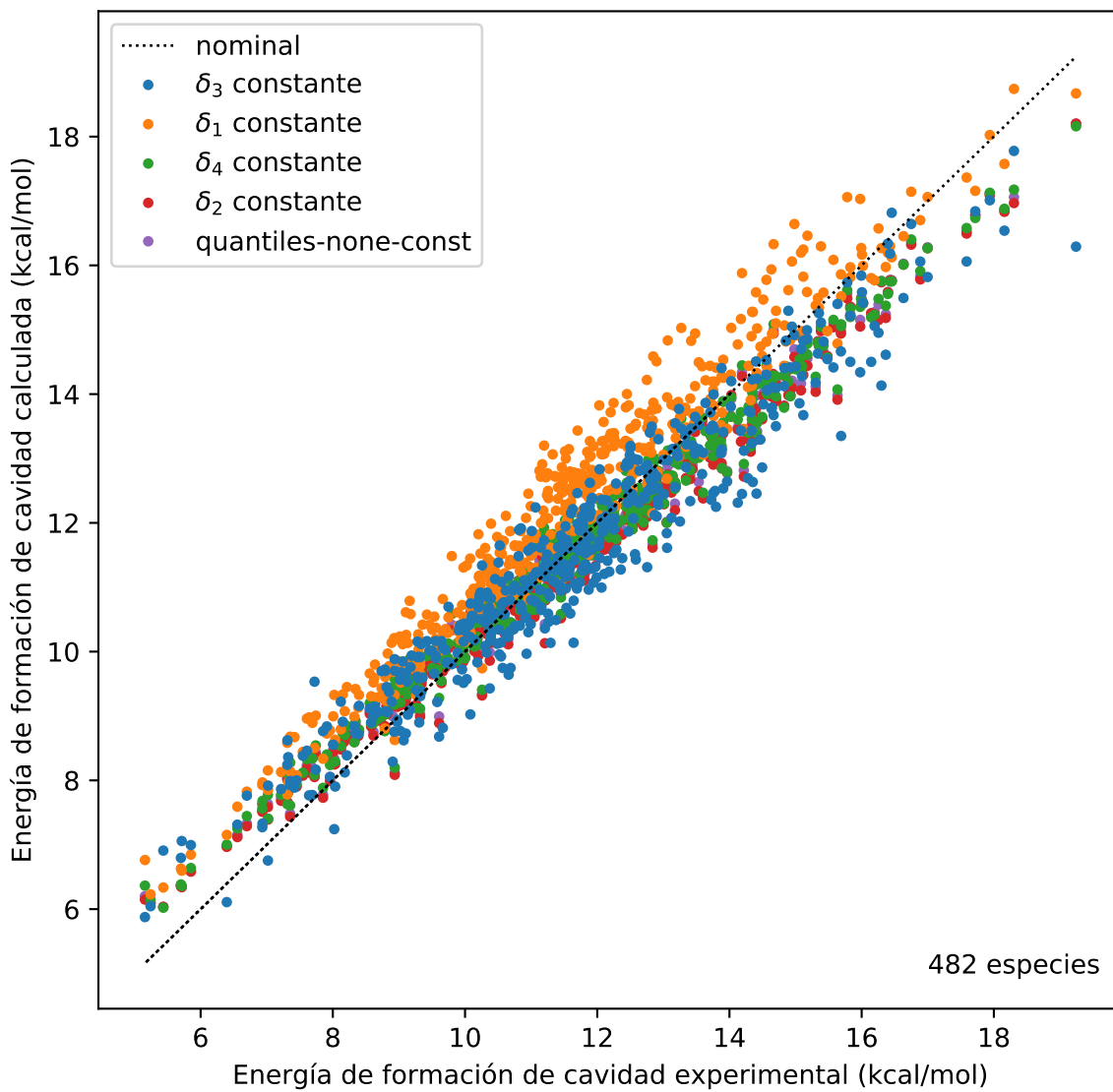


Figura 3.1: Energía de Gibbs calculada para 482 especies. Utilizamos las primeras 20 especies para crear funciones que aproximen el factor $f(R)$. Comparados con resultados teóricos de Mobley et al. y Cooper et al. se visualizan los 5 modelos que mejor correlación poseen según el coeficiente de Pearson.

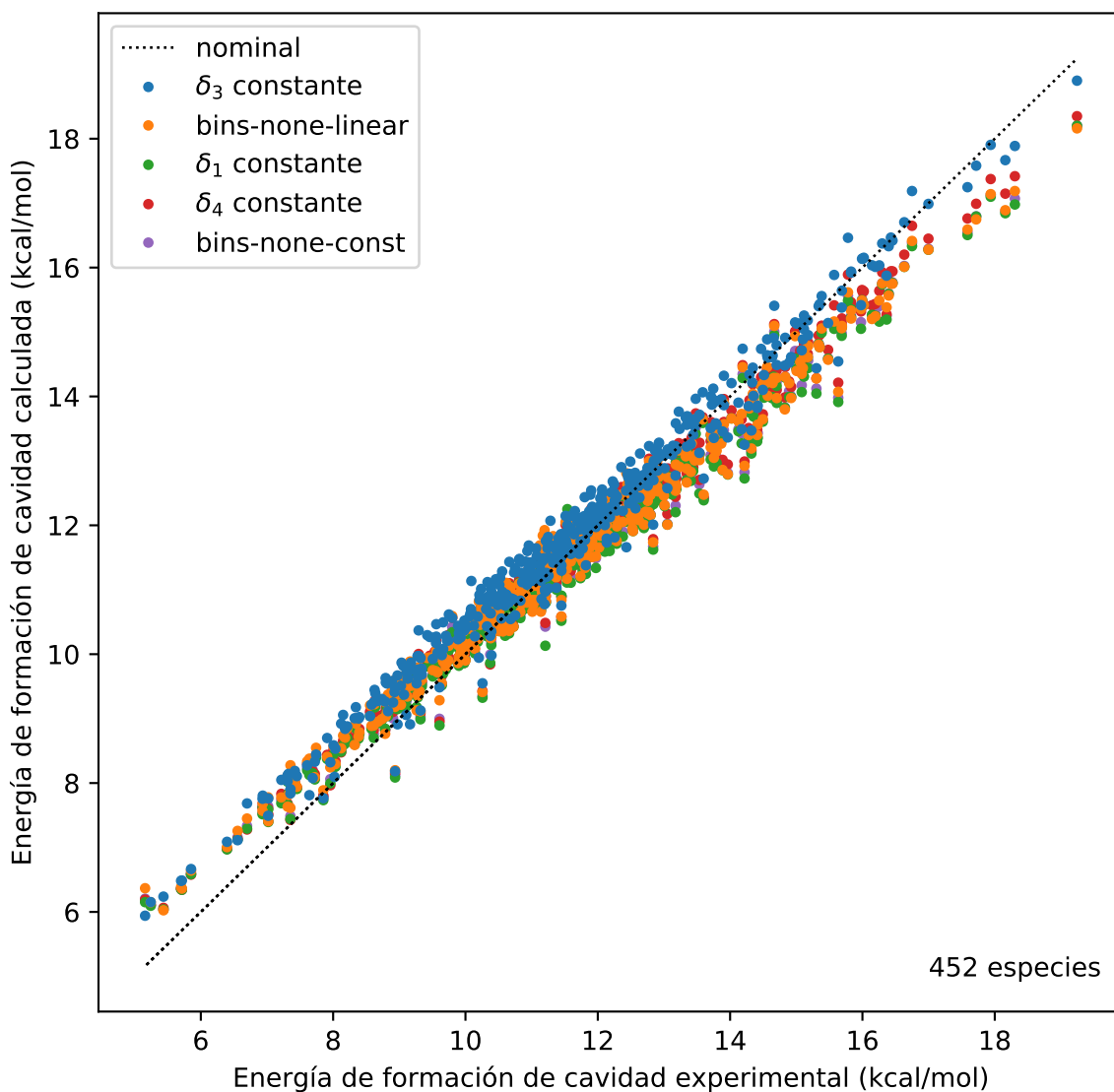


Figura 3.2: Energía de Gibbs calculada para 452 especies. Utilizamos las primeras 50 especies para crear funciones que aproximen el factor $f(R)$. Comparados con resultados de teóricos de Mobley et al. y Cooper et al. se visualizan los 5 modelos que mejor correlación poseen según el coeficiente de Pearson.

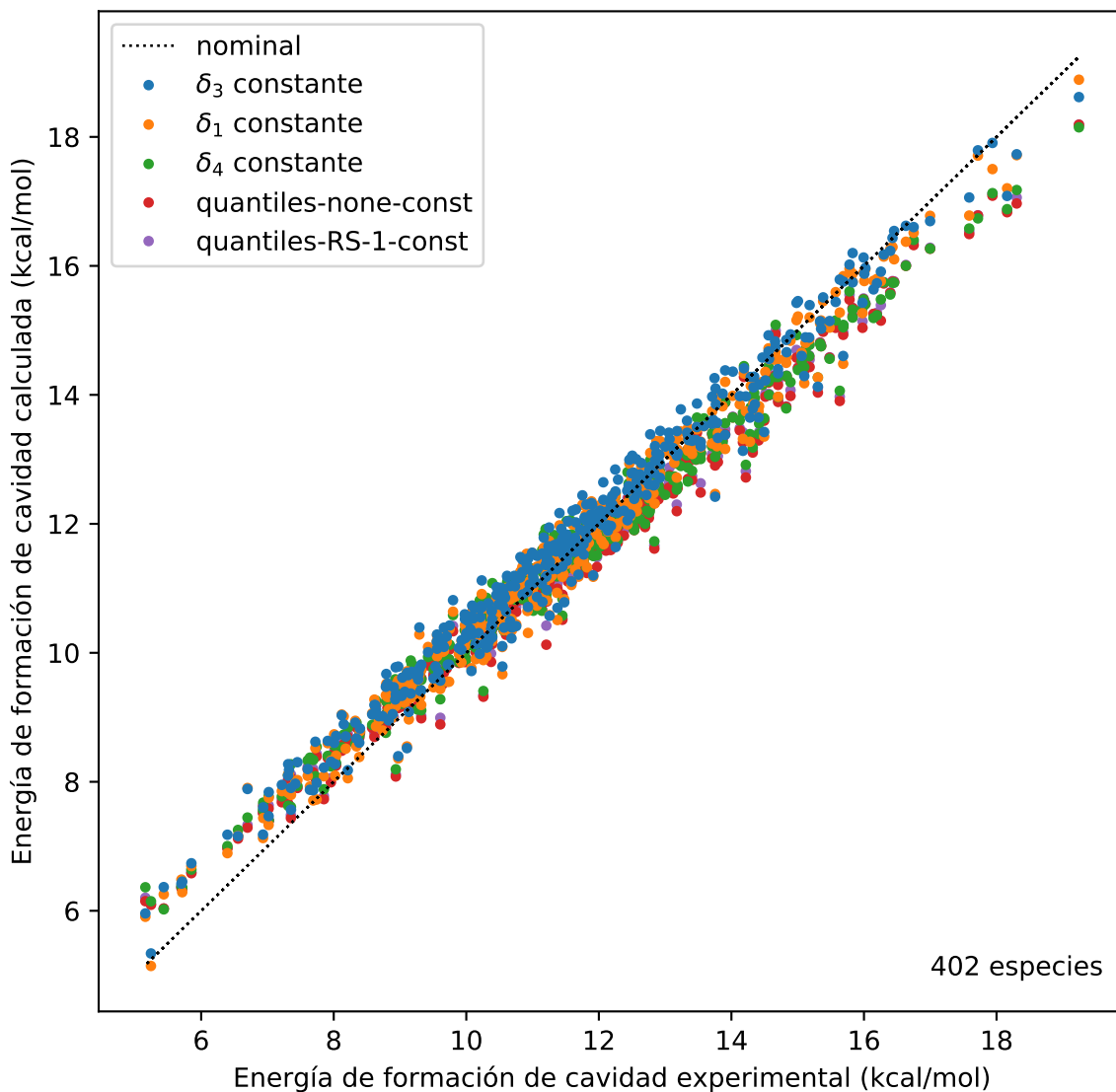
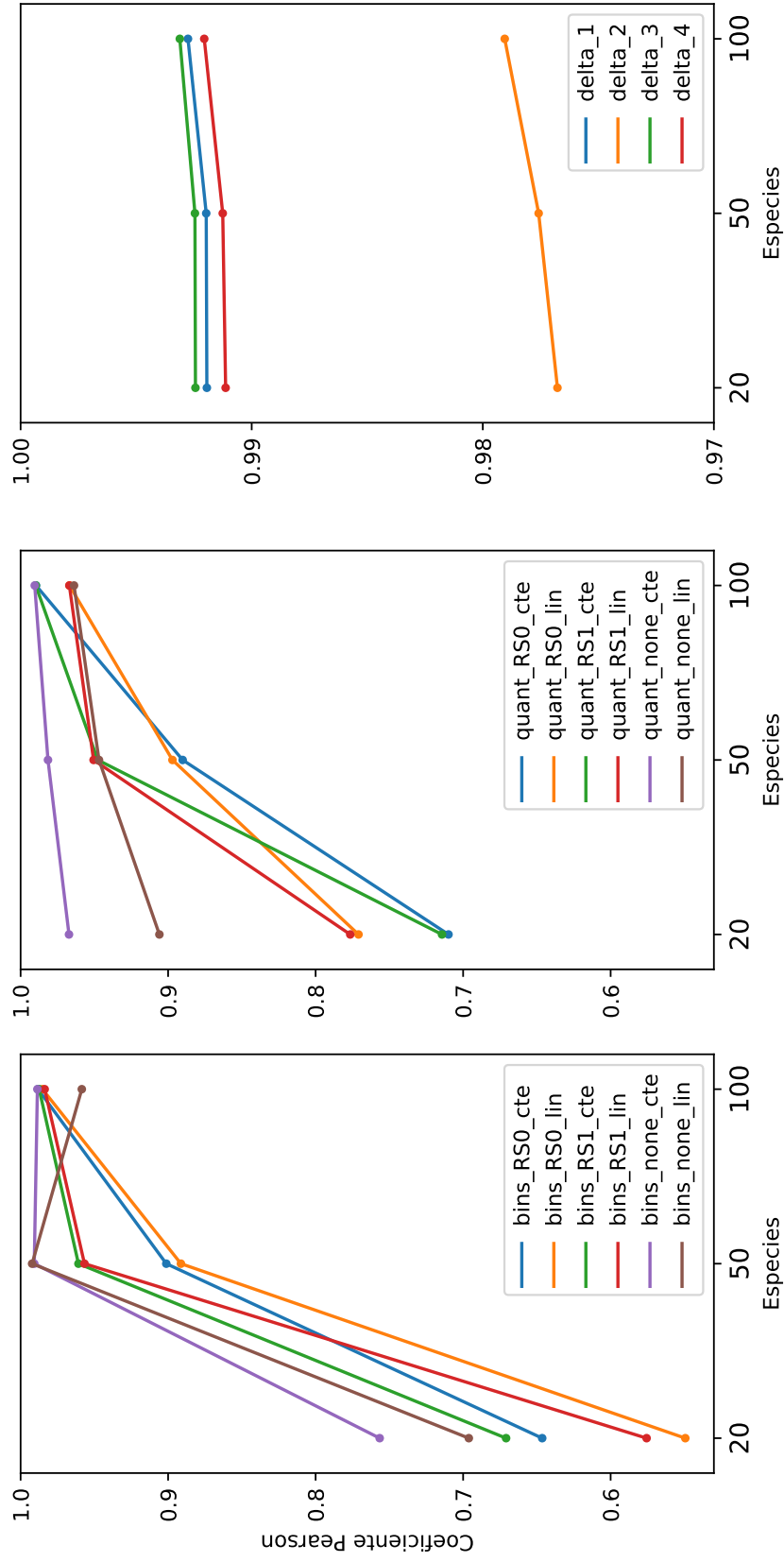


Figura 3.3: Energía de Gibbs calculada para 402 especies. Utilizamos las primeras 100 especies para crear funciones que aproximen el factor $f(R)$. Comparados con resultados de teóricos de Mobley et al. y Cooper et al. se visualizan los 5 modelos que mejor correlación poseen según el coeficiente de Pearson.

Lugar	Predicción con 20 especies		Predicción con 50 especies		Predicción con 100 especies	
	ρ	Modelo	ρ	Modelo	ρ	Modelo
1	0.99243	δ_3 constante	0.99246	δ_3 constante	0.99311	δ_3 constante
2	0.99194	δ_1 constante	0.9922	bins-none-linear	0.99276	δ_1 constante
3	0.99113	δ_4 constante	0.99197	δ_1 constante	0.99205	δ_4 constante
4	0.97677	δ_2 constante	0.99125	δ_4 constante	0.99043	quantiles-none-const
5	0.9672	quantiles-none-const	0.99074	bins-none-const	0.98973	quantiles-RS-1-const
6	0.90591	quantiles-none-linear	0.98144	quantiles-none-const	0.98956	quantiles-RS-0-const
7	0.77653	quantiles-RS-1-linear	0.97758	δ_2 constante	0.98862	bins-none-const
8	0.77093	quantiles-RS-0-linear	0.96079	bins-RS-1-const	0.98821	bins-RS-0-const
9	0.75657	bins-none-const	0.95671	bins-RS-1-linear	0.98743	bins-RS-1-const
10	0.71432	quantiles-RS-1-const	0.95053	quantiles-RS-1-linear	0.98523	bins-RS-0-linear
11	0.70992	quantiles-RS-0-const	0.9477	quantiles-RS-1-const	0.98369	bins-RS-1-linear
12	0.69614	bins-none-linear	0.94687	quantiles-none-linear	0.97905	δ_2 constante
13	0.67087	bins-RS-1-const	0.90124	bins-RS-0-const	0.96695	quantiles-RS-1-linear
14	0.6464	bins-RS-0-const	0.89707	quantiles-RS-0-linear	0.96684	quantiles-RS-0-linear
15	0.57561	bins-RS-1-linear	0.89131	bins-RS-0-linear	0.96379	quantiles-none-linear
16	0.54937	bins-RS-0-linear	0.89014	quantiles-RS-0-const	0.95852	bins-none-linear

Tabla 3.1: Coeficiente de correlación de Pearson por modelo, para generación de modelos utilizando 20, 50 y 100 especies. Visualizado en figura 3.4



(c) Modelos de Tolman constante.

(b) Modelos de intervalos equidensos.

(a) Modelos de intervalos equidistantes.

Figura 3.4: Coeficiente de correlación de Pearson por modelo, ver tabla 3.1, para generación de modelos utilizando 20, 50 y 100 especies.

4 | Conclusiones

El cálculo de la energía de formación de cavidades depende casi exclusivamente del costo energético debido al reordenamiento de partículas de solvente cercana a la interfase, $\int \gamma \cdot A$. El término volumétrico es ínfimo en comparación.

Las figuras 3.1, 3.2, y 3.3 muestran los cinco modelos con mejor correlación. Se puede apreciar que para energías de Mobley et al. muy menores o muy mayores, los modelos utilizados para aproximar la energía de formación de cavidad no son exactos. Antes de los 9 *kcal/mol* existe una sobre-estimación. Esto podría ser indicativo de una geometría más compleja, ya que una menor energía indica menor área superficial, y por ende especies nanoscópicas. Dentro de los modelos graficados se encuentran por mayoría los de Tolman constante, basados en las simplificaciones de la ecuación GTKB. Estas simplificaciones consideran un valor insignificante para la relación δ/r [4]. Esta suposición hace muy claro que para pequeños valores de r existirá mayor error. Existe además una sub-estimación para energías superiores a 15 *kcal/mol*, lo que es indicativo de una molécula de gran tamaño. Esperamos que para especies muy grandes, la aproximación de la energía de formación de cavidad presente una acumulación de los errores inherentes en los modelos utilizados.

Inicialmente, el factor de corrección de la tensión superficial $f(R)$ es ajustado para los resultados de Mobley et al [1]. Logramos despejar la variable Tolman de $f(R)$, por medio de cuatro simplificaciones de la ecuación GTKB, figuras 2.4–2.7. Es importante reconocer un caso en que no recibimos información debido a la divergencia extrema de los resultados, figura 2.4a. Sin embargo, de los resultados de Tolman para los aproximadamente 388.000 nodos individuales, correspondiendo a radios únicos en sus mallas por especie, no se percibe una correlación entre el valor de Tolman y el radio de curvatura. Para todas las soluciones obtenemos una gran mayoría de resultados positivos a escala molecular (cerca al valor del radio de una molécula de agua $r = 1,4 \text{ \AA}$). En comparación con otras investigaciones, tabla 1.1, comentamos que éste resultado origina de simulaciones de dinámica molecular ya que ajustamos el valor a resultados de Mobley et al. (2009) [1].

Luego de haber determinado una longitud de Tolman característica por especie, para cada una de las simplificaciones de GTKB, adoptamos modelos en donde Tolman es constante e igual al promedio de los valores característicos. Muchos investigadores han adoptado la ecuación 1.11 como simplificación, y muchos más utilizan un valor de tensión superficial corregido y constante [13] [25] [4], especialmente cuando se trata de calcular la energía libre de solvata-

ción en su totalidad. Es interesante observar que dentro de los modelos empleados, el utilizar un valor constante para predecir siempre entrega una correlación elevada. Dentro de los modelos con Tolman constante, δ_3 ecuación 1.12 siempre es el modelo de mayor exactitud, seguido por δ_1 ecuación 1.10, δ_4 ecuación 1.13 y finalmente δ_2 ecuación 1.11. Mientras más especies utilizamos para predecir el gasto energético en las demás, peor exactitud presenta el modelo de Tolman constante igual δ_2 . Dentro del estudio de Kalová y Mareš (2019) [4], se concluye que la simplificación δ_2 es la peor aproximación para un valor positivo de Tolman, afirmación que es apoyada por los resultados de este informe. En promedio las predicciones con modelos Tolman constante presentan un coeficiente de Pearson de 0,98.

Esperamos que en los modelos empleados aumente el coeficiente Pearson al aumentar la cantidad de especies con las cuales predecimos el comportamiento, ver figura 3.4. En caso contrario, podemos poner en duda la validez del modelo, siendo poco intuitivo que con más información empeore la capacidad de predicción. Existen dos modelos que empeoran en exactitud, según sus coeficientes de Pearson, ambos usan intervalos equidistantes para agrupar los radios de curvatura.

Esperamos que los modelos con intervalos equidistantes requieran bastante información para predecir con exactitud, debido a que para un rango de radios muy amplio, pueden existir muchos intervalos definidos sin datos. Este problema se ve acentuado por datos aberrantes. El histograma del rango de radios, figura 2.3, muestra muchos posibles intervalos en donde no existen datos. En promedio las predicciones con modelos de intervalos equidistantes presentan un coeficiente de Pearson de 0,86.

Los modelos que utilizan intervalos equidistantes requieren menos datos iniciales para entregar resultados con buena exactitud, ya que predicen información localizada donde existan grandes cantidades de datos; no existen intervalos vacíos. En promedio las predicciones con modelos de intervalos equidistantes presentan un coeficiente de Pearson de 0,91.

En promedio, modelos que consideran la eliminación de datos aberrantes presentan un coeficiente de Pearson disminuido en 0,06. Para los modelos que eliminan datos aberrantes, si al implementarse en la predicción de otras especies considera nulos los aportes por radios fueran de los límites decididos por estadística robusta, disminuye en 0,02 la correlación según coeficiente de Pearson. La diferencia entre considerar funciones continuas o lineales por parte para predecir el factor de corrección de la tensión superficial $f(R)$ es de 0,019.

La correlación con los datos obtenidos por Mobley es significativa, y ya que estos métodos no requieren grandes recursos de tiempo y tecnología, podemos afirmar que es favorable utilizar estos métodos sobre la simulación por dinámica molecular. Sin embargo, el aumento de la exactitud según el coeficiente de Pearson al utilizar un modelo de Tolman constante supera cualquier otro cambio que se pueda implementar.

Referencias

- [1] D. Mobley, C. Bayly, M. Cooper, M. Shirts, and K. Dill, “Small molecule hydration free energies in explicit solvent: An extensive test of atomistic simulations: Supporting Information,” *Journal of Chemical Theory and Computation*, 2009.
- [2] C. D. Cooper and J. P. Bardhan, “A capacitor model for nonpolar solvation,” *Unpublished*, 2019.
- [3] D. Chandler, “Interfaces and the driving force of hydrophobic assembly,” *Nature*, vol. 437, pp. 640–647, 9 2005.
- [4] J. Kalová and R. Mareš, “Size dependences of surface tension and measurement accuracy,” p. 020010, 2019.
- [5] b. J. Willard Gibbs, “Gibbs-Equilibrium of Heterogeneous Substances. 441 ART. LII.- On the Equilibrium of Heterogeneous Substances,” tech. rep.
- [6] R. C. Tolman, “The superficial density of matter at a liquid-vapor boundary,” *The Journal of Chemical Physics*, vol. 17, no. 2, pp. 118–127, 1949.
- [7] W. Kauzmann, “Some factors in the interpretation of protein denaturation,” *Advances in Protein Chemistry*, 1959.
- [8] F. H. Stillinger, “Structure in Aqueous Solutions of Nonpolar Solutes from the Standpoint of Scaled-Particle Theory,” *Journal of Solution Chemistry*, 1973.
- [9] K. Lum, D. Chandler, and J. Weeks, “Hydrophobicity at Small and Large Length Scales,” *Journal of Physical Chemistry B*, 1999.
- [10] S. Flugge, *Structure of Liquids*, vol. 10 of *Encyclopedia of Physics*. Springer, 1960.
- [11] S. Kondo, “Thermodynamical fundamental equation for spherical interface,” *The Journal of Chemical Physics*, vol. 25, no. 4, pp. 662–669, 1956.
- [12] Z. Ru-Zeng and W. Xiao-Song, “Thermodynamic theory of the Tolmans length,” *Chinese Physics B*, vol. 19, p. 76801, 7 2010.
- [13] J. Dzubiella, J. M. J. Swanson, and J. A. McCammon, “Coupling nonpolar and polar solvation free energies in implicit solvent models,” *The Journal of Chemical Physics*, 2006.

-
- [14] J. Kalová and R. Mareš, “Size Dependences of Surface Tension,” *International Journal of Thermophysics*, vol. 36, pp. 2862–2868, 11 2015.
- [15] V. Holten, D. G. Labetski, and M. E. Van Dongen, “Homogeneous nucleation of water between 200 and 240 K: New wave tube data and estimation of the Tolman length,” *Journal of Chemical Physics*, vol. 123, 9 2005.
- [16] M. E. M. Azouzi, C. Ramboz, J. F. Lenain, and F. Caupin, “A coherent picture of water at extreme negative pressure,” *Nature Physics*, vol. 9, no. 1, pp. 38–41, 2013.
- [17] A. A. Homman, E. Bourasseau, G. Stoltz, P. Malfreyt, L. Strafella, and A. Ghoufi, “Surface tension of spherical drops from surface of tension,” *Journal of Chemical Physics*, vol. 140, no. 3, 2014.
- [18] H. M. Lu and Q. Jiang, “Size-dependent surface tension and Tolman’s length of droplets,” *Langmuir*, vol. 21, pp. 779–781, 1 2005.
- [19] M. N. Joswiak, N. Duff, M. F. Doherty, and B. Peters, “Size-dependent surface free energy and tolman-corrected droplet nucleation of TIP4P/2005 water,” *Journal of Physical Chemistry Letters*, vol. 4, pp. 4267–4272, 12 2013.
- [20] O. Wilhelmson, D. Bedeaux, and D. Reguera, “Communication: Tolman length and rigidity constants of water and their role in nucleation,” *Journal of Chemical Physics*, vol. 142, 5 2015.
- [21] F. Sedlmeier and R. R. Netz, “The spontaneous curvature of the water-hydrophobe interface,” *Journal of Chemical Physics*, vol. 137, 10 2012.
- [22] S. M. Malek, P. H. Poole, and I. Saika-Voivod, “Surface tension of supercooled water nanodroplets from computer simulations,” *Journal of Chemical Physics*, vol. 150, 6 2019.
- [23] S. Kim, D. Kim, J. Kim, S. An, and W. Jhe, “Direct Evidence for Curvature-Dependent Surface Tension in Capillary Condensation: Kelvin Equation at Molecular Scale,” *Physical Review X*, vol. 8, 12 2018.
- [24] B. Kozlíková, M. Krone, M. Falk, N. Lindow, M. Baaden, D. Baum, I. Viola, J. Parulek, and H. C. Hege, “Visualization of Biomolecular Structures: State of the Art Revisited,” *Computer Graphics Forum*, vol. 36, pp. 178–204, 12 2017.

-
- [25] D. M. Huang, P. L. Geissler, and D. Chandler, “Scaling of hydrophobic solvation free energies,” *Journal of Physical Chemistry B*, vol. 105, pp. 6704–6709, 7 2001.

A | Derivación de la energía de formación de cavidad

Consideremos un sistema de dos fases en equilibrio hidrostático tal como es explicado en la sección 1.3, en donde aumenta el volumen, y por lo tanto el área del interfaz entre α y β .

Según la primera ley de la termodinámica, la energía interna en un sistema cerrado:

$$dU = dq - dw \quad (\text{A.1})$$

Según la segunda ley de la termodinámica, el calor absorbido por el sistema, dQ , para un proceso reversible se define,

$$dq = T \cdot ds \quad (\text{A.2})$$

El trabajo en el sistema es debido únicamente a la expansión de la cavidad, presión por volumen, y a la creación de la interfaz, área superficial.

$$dw = pdV - \gamma dA \quad (\text{A.3})$$

$$\therefore dU = T \cdot dS - pdV + \gamma dA \quad (\text{A.4})$$

La ecuación A.4 es válida para cualquier sistema en equilibrio mecánico de presión constante, p . Por otro lado, conocemos la definición de la energía libre de Gibbs:

$$G = H - TS \quad (\text{A.5})$$

Además la entalpía se define como $H = U + pV$,

$$G = (U + pV) - TS \quad (\text{A.6})$$

En su forma diferencial,

$$dG = dU + (pdV + Vdp) - (TdS + SdT) \quad (\text{A.7})$$

De la ecuación A.4,

$$\begin{aligned}dG &= [TdS - pdV + \gamma dA] + (pdV + Vdp) - (TdS + SdT) \\dG &= \gamma dA + Vdp - SdT\end{aligned}$$

Para temperatura y presión constante, $dT = 0$, $dp = 0$.

$$dG = \gamma dA \quad (\text{A.8})$$

Las unidades que pertenecen a la variable γ son iguales a las de tensión superficial. Se conoce que para una superficie plana, este valor corresponde a la tensión superficial del agua, γ_{lv} . Existen fuentes que han determinado la presencia de un factor de corrección de la tensión superficial según su curvatura.

$$\Delta G_{cav} = \int \gamma dA = \int \gamma_{lv} f(R) dA \quad (\text{A.9})$$

Para lograr un estado de menor energía, es favorable un interfaz curvo, el cual optimiza la relación área/volumen. Sin embargo, para el análisis presentado en este estudio, se adopta la forma de Gibbs utilizada en Chandler (2005) [3], Huang et al. (2001) [25], Dzubiella et al. (2006) [13].

$$\Delta G_{cav} = \int \gamma_{lv} f(R) dA + V \cdot P \quad (\text{A.10})$$

B | Mejor ajuste de una esfera

Para aproximar la curvatura de la malla en cada nodo, se calcula el radio de la esfera que mejor se ajuste al nodo y a sus nodos adyacentes. Los nodos adyacentes se definen como ellos que, junto al nodo principal, corresponden a vértices de algún triángulo en la malla. Para obtener el radio de la esfera de mejor ajuste, se debe considerar por lo menos cuatro nodos, condición que se cumple en toda malla triangular tridimensional. De esta forma se le otorgará un radio a cada nodo en la malla. Los t nodos por esfera tienen la forma:

$$\begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \\ x_4 & y_4 & z_4 \\ \vdots & \vdots & \vdots \\ x_t & y_t & z_t \end{bmatrix}$$

Promedio de las coordenadas x , y y z para cada esfera:

$$\bar{x} = \frac{1}{t} \sum_{i=1}^{i=t} x_i, \quad \bar{y} = \frac{1}{t} \sum_{i=1}^{i=t} y_i, \quad \bar{z} = \frac{1}{t} \sum_{i=1}^{i=t} z_i$$

Aproximación de coordenadas del centro de la esfera:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = (A^T \cdot A)^{-1} \cdot A^T \cdot B$$

Se definen entonces las matrices A y B :

$$A = 2 \begin{bmatrix} \sum_{i=1}^{i=t} \frac{x_i(x_i - \bar{x})}{t} & \sum_{i=1}^{i=t} \frac{x_i(y_i - \bar{y})}{t} & \sum_{i=1}^{i=t} \frac{x_i(z_i - \bar{z})}{t} \\ \sum_{i=1}^{i=t} \frac{y_i(x_i - \bar{x})}{t} & \sum_{i=1}^{i=t} \frac{y_i(y_i - \bar{y})}{t} & \sum_{i=1}^{i=t} \frac{y_i(z_i - \bar{z})}{t} \\ \sum_{i=1}^{i=t} \frac{z_i(x_i - \bar{x})}{t} & \sum_{i=1}^{i=t} \frac{z_i(y_i - \bar{y})}{t} & \sum_{i=1}^{i=t} \frac{z_i(z_i - \bar{z})}{t} \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{(x_i^2 + y_i^2 + z_i^2)(x_i - \bar{x})}{t} \\ \frac{(x_i^2 + y_i^2 + z_i^2)(y_i - \bar{y})}{t} \\ \frac{(x_i^2 + y_i^2 + z_i^2)(z_i - \bar{z})}{t} \end{bmatrix}$$

Finalmente, el radio de la esfera de mejor ajuste por nodo se aproxima como:

$$R_{bf} = \sqrt{\frac{\sum_{i=1}^{i=t} ((x_i - x_c)^2 + (y_i - y_c)^2 + (z_i - z_c)^2)}{t}}$$

Se define además el vector del ajuste esférico por nodo:

$$\vec{V}_{esf} = [(x_c - x_n), (y_c - y_n), (z_c - z_n)]$$

C | Solución del factor de corrección

El sistema de ecuaciones 2.15 se resuelve utilizando mínimos cuadrados, restringiendo la solución a valores entre 0 y 1. Se utilizan 20 intervalos, resultando en una matriz A de 502×20 .

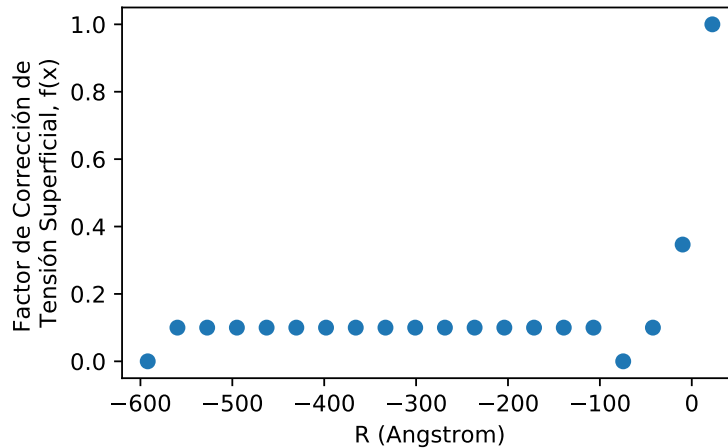


Figura C.1: Corrección de la tensión superficial, resultados de 2.15 con 20 intervalos equidistantes sin eliminación de datos.

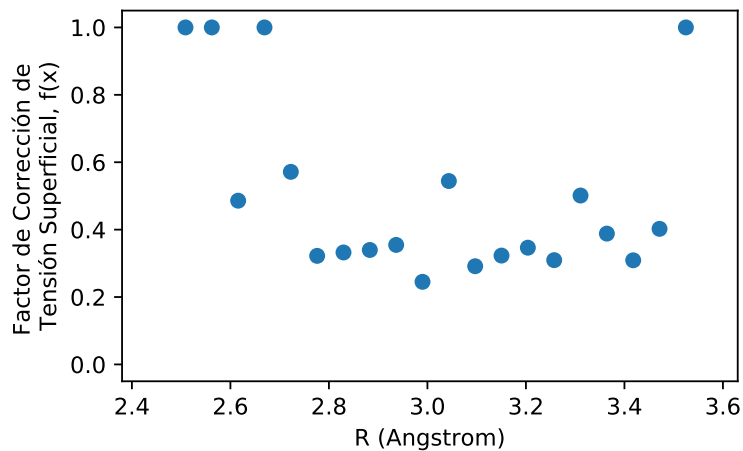


Figura C.2: Corrección de la tensión superficial, resultados de 2.15 con 20 intervalos equidistantes con eliminación de datos aberrantes.

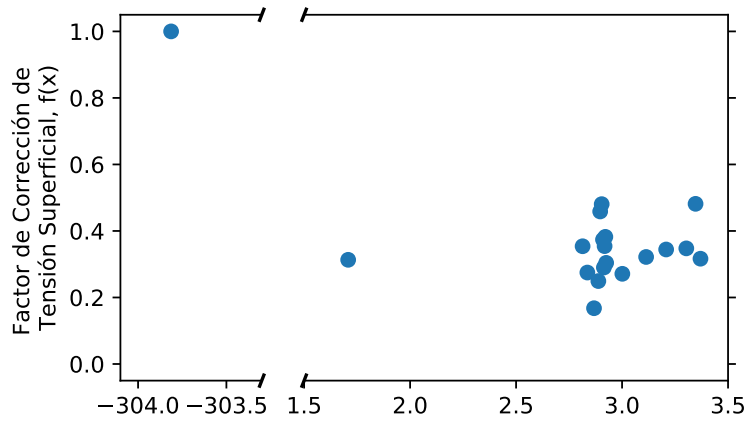


Figura C.3: Corrección de la tensión superficial, resultados de 2.15 con 20 intervalos equidensos sin eliminación de datos.

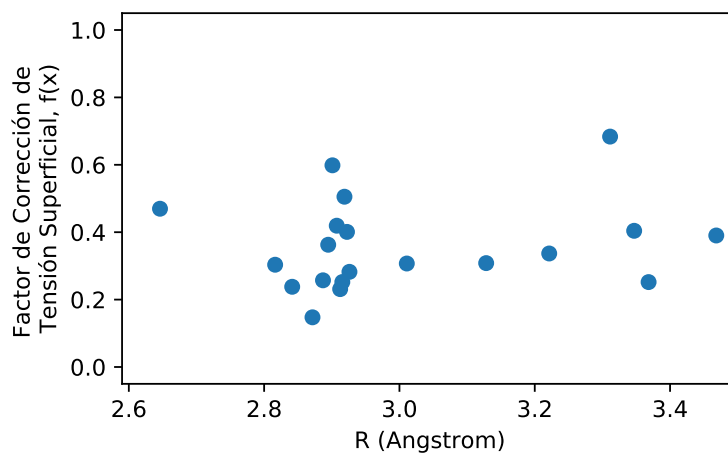


Figura C.4: Corrección de la tensión superficial, resultados de 2.15 con 20 intervalos equidensos con eliminación de datos aberrantes