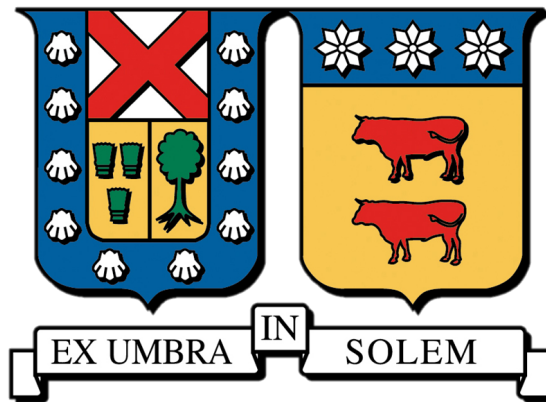


Universidad Técnica Federico Santa María
Departamento Ingeniería Mecánica
Valparaíso - Chile



“Análisis geométrico de la superficie de un soluto y su efecto en la aproximación de su energía de disolución”

Memoria de titulación para optar al título:
Ingeniero Civil Mecánico

Joaquín Echaíz Bielitz

Profesor guía: PhD. Christopher Cooper Villagrán

Profesor correferente: PhD. Harvey Zambrano Rodríguez

Diciembre - 2019

Agradecimientos

A Valentina, quien es mi compañera de trabajo y amiga, pasamos por esta etapa juntos desde el comienzo hasta el final. Sin ella esto hubiese sido para mi igual de importante pero menos entretenido.

También agradezco a mi pololo Maximiliano que me han acompañado y comprendido a lo largo de esta gran etapa.

A mi profesor Christopher por su guía y cercanía, sin duda es el mejor profesor que pude haber escogido.

Agradezco a mis padres y hermano que me dieron todo el cariño y todo lo que necesité para que pudiera estudiar.

Y finalmente agradezco a Alexandra Asanovna Elbakyan, fundadora de Sci-Hub, el sitio web que otorga acceso inmediato y gratuito a cualquier usuario a un enorme porcentaje de los artículos científicos.



Dedicatorias

Dedico este trabajo a mi hermana Bárbara y a mi sobrina Simona. Durante gran parte del tiempo pensé en ustedes, que debía dar este paso para poder viajar a donde quiera que estén, ser parte de sus vidas y que las echo mucho de menos.

Resumen

Se propone encontrar modelos implícitos capaces de aproximar la componente no-polar de la energía de solvatación ΔG_{np} , y la energía necesaria para la formación de la cavidad del soluto ΔG_{cav} . Debido a la naturaleza de estas componentes de la energía de solvatación, es que se modelan como funciones de la geometría de las superficies que produce la cavidad del soluto al interior del solvente.

Una de las propiedades más importantes de las interfaces soluto-solvente es la tensión superficial, cuyo comportamiento es afectado por la curvatura que tiene la interfaz y es caracterizado por la longitud de Tolman. Es por esto que el análisis de la curvatura de la superficie de la geometría que representa esta interfaz, podría llevar a una mejor aproximación de las componentes de la energía de solvatación.

Son diversas las formas en que se pueden caracterizar los aspectos de una geometría, considerando una malla triangular. Si bien el volumen y el área total se encuentran bien definidos, existen otros cuya complejidad depende de si se realiza una integración por vértice o por panel, como la curvatura, la normal y el área de integración locales. Para algunos aspectos se ha encontrado más de una manera de aproximación, las que se explorarán de forma excluyente en modelos separados.

Como resultado se han obtenido 1442 modelos, los cuales son una combinación de factores, agrupados en siete familias, para la aproximación de las energías ΔG_{np} y ΔG_{cav} . Un par de estos se han concebido con la intención de contrastar la importancia del análisis geométrico de la superficie. Los modelos obtenidos han sido jerarquizados según la correlación que tienen con los datos nominales con que se cuenta.

Los resultados apuntan a que el análisis geométrico de la superficie de los solutos logra una mejor aproximación de las componentes energéticas, obteniéndose una mejor correlación y menores errores que en los casos en que no se consideró un análisis de la geometría. Los resultados también indican que la mejor forma para aproximar la curvatura, para su utilización en modelos de este tipo, es hacerlo mediante un mejor ajuste de una esfera. Esto en contraste del tradicional empleo de la curvatura media.

Palabras clave: solvatación, energía libre de Gibbs, tensión superficial, longitud de Tolman, curvatura.

Abstract

It is proposed to find implicit models capable of approximating the non-polar component of solvent-free energy ΔG_{np} , and the energy required for the formation of the solute cavity ΔG_{cav} . Due to the nature of these solvent energy components, they are modeled as functions of the surface geometry produced by the solute cavity within the solvent.

One of the most important properties of solute-solvent interfaces is the surface tension, whose behavior is affected by the curvature of the interface and is characterized by the Tolman length. Hence the analysis of the curvature of the surface of the geometry that represents this interface could lead to a better approximation of the components of the energy of solvation.

There are several ways in which the aspects of a geometry can be characterized, considering a triangle mesh. Although the volume and the total area are well defined, there are others whose complexity depends on whether a vertex or panel integration is performed, such as the curvature, the normal and the local integration area. For some aspects more than one approach has been found, which will be explored exclusively in separate models.

The result has been 1442 models, which are a combination of factors, grouped into seven families, for the approximation of energies ΔG_{np} and ΔG_{cav} . A couple of these have been conceived with the intention of contrasting the importance of surface geometric analysis. The models obtained have been hierarchized according to the correlation they have with the nominal data available.

The results indicate that the geometric analysis of the surface of the solutes achieves a better approximation of the energetic components, obtaining a better correlation and fewer errors than in the cases in which an analysis of the geometry was not considered. The results also indicate that the best way to approximate the curvature, for use in models of this type, is to do it through the best fit of a sphere. This is in contrast to the traditional use of mean curvature.

Keywords: Solvation, Gibbs free energy, surface tension, Tolman length, curvature.

Glosario

anillo de puntos vecinos Perímetro formado con las aristas de los puntos que se conectan a un mismo punto.. , [15](#), [17–19](#), [22–24](#), [40](#)

MSMS Software que computa, para un grupo dado de esferas S y una probeta de radio rp , la superficie reducida y el modelo analítico de la .. , [16](#)

puntos vecinos Puntos con los que un punto forma aristas. , [17](#), [18](#)

Python lenguaje de programación multiparadigma, ya que soporta orientación a objetos, programación imperativa y, en menor medida, programación funcional. Es un lenguaje interpretado, dinámico y multiplataforma. , [11](#)

Acrónimos

AWPN Angle Weighted Pseudo Normal. , 15, 16, 32, 33, 40

EAM Error Absoluto Medio. , 35, 36, 39, 40

RECM Raíz del Error Cuadrático Medio. , 35, 36, 39, 40

SAS Solvent Accesible Surface. , 11, 12, 26, 42

SES Solvent Excluded Surface. , 11, 12, 26, 42

SPC Single Point Charge. , 8, 24, 25

SPC/E Single Point Charge/Extended. , 8, 24, 25, 41

Lista de símbolos

Símbolo	Descripción	Unidad
C_{esf}	centro de la esfera de mejor ajuste	-
H	curvatura media	Å^{-1}
P	presión	Pa
R_{esf}	radio de la esfera de mejor ajuste	Å
R	radio de curvatura medio	Å
ΔG_{cav}	energía correspondiente a la creación de una cavidad en el solvente	J
ΔG_{np}	componente no-polar de la energía de solvatación	J
ΔG_{vdw}	energía correspondiente a las interacciones de Van der Waals	J
ΔG	energía libre de Gibbs.	J
ΔH	entalpía	J
ΔS	entropía	J
ΔU	energía interna	J
δ	longitud de Tolman	Å
γ_{lv}	tensión superficial del agua en una interfaz plana	N/m
γ	tensión superficial.	N/m
ρ	coeficiente de pearson	-
H	curvatura media normal	Å^{-1}
\vec{R}_{cr}	radio centro-radio	Å
\vec{R}_{cv}	radio centro-vértice	Å

Índice general

1. Marco teórico	1
1.1. Longitud de Tolman	6
2. Desarrollo	9
2.1. Solutos esféricos no-polares	9
2.2. Geometrías	10
2.3. El modelo JUALE	11
2.4. Normal	13
2.4.1. Normal del panel	14
2.4.2. Normal en el vértice: AWPN	15
2.4.3. Normal en el vértice: MSMS	16
2.5. Curvatura	17
2.5.1. Curvatura media	17
2.5.2. Mejor ajuste de una esfera	18
2.5.3. Paneles curvos	20
2.6. Secciones cóncavas y convexas	20
2.7. Área	22
2.7.1. Por panel	22
2.7.2. Por vértice: área tercio	22
2.7.3. Por vértice: área mixta	22

2.8.	Tipo de aproximación de la longitud de Tolman	24
2.8.1.	SPC y SPC	24
2.8.2.	Longitud de Tolman constante, polinomio de grado 1 y tensión superficial contante	25
3.	Resultados	32
3.1.	Componente no-polar de la energía de solvatación	33
3.1.1.	Raíz del Error Cuadrático Medio y Error Absoluto Medio para componente no-polar de la energía de solvatación	35
3.2.	Energía requerida para la formación de una cavidad en el soluto	36
3.2.1.	Raíz del Error Cuadrático Medio y Error Absoluto Medio para energía correspondiente a la creación de una cavidad en el solvente	39
4.	Modelos con mayor correlación	40
5.	Conclusiones	42
	Referencias	45
A.	Mejor ajuste de una esfera	48
B.	Regresiones: Energía de la cavidad	50

Índice de figuras

1.1. Diagrama de energías involucradas en la solvatación	3
1.2. Energía de la formación de una cavidad ΔG_{cav} en función del tamaño de la cavidad.	4
2.1. Energía libre de solvatación normalizada por la superficie de la esfera en función del radio	10
2.2. Ejemplo de geometrías utilizadas	11
2.3. Representación de las distintas superficies	12
2.4. Ejemplo de cara de una malla triangular.	14
2.5. La AWPN es el promedio ponderado por el ángulo de las normales de las caras incidentes en un vértice.	16
2.6. Ángulos característicos para la aproximación de la curvatura media.	18
2.7. Un vértice V_i junto a su anillo de puntos vecinos son aproximados a una sección de esfera de radio R_i	19
2.8. Producto punto como herramienta de análisis de la concavidad de una superficie	21
2.9. División del área correspondiente a un vértice por tercios.	23
2.10. Área mixta y región de Voronoi	24
2.11. Regresiones de la longitud de Tolman para basadas en modelos SPC y SPC/E	25
2.12. Longitud de Tolman por radio promedio	28
2.13. Longitud de Tolman por radio ponderado por superficie	28
2.14. Longitud de Tolman por radio promedio de la malla	29

2.15. Longitud de Tolman por radio ponderado por superficie de la malla . . .	30
3.1. Cinco mejores modelos para componente no-polar de la energía de sol- vatación	34
3.2. Cinco mejores modelos para ΔG_{cav}	37
5.1. Correlación del área y el volumen frente a ΔG_{np}	43
5.2. Correlación del área y el volumen frente a ΔG_{cav}	44
B.1. Longitud de Tolman por radio promedio para la energía de formación de la cavidad	50
B.2. Longitud de Tolman por radio ponderado por la superficie para la ener- gía de formación de la cavidad	51
B.3. Longitud de Tolman por radio promedio de la malla para la energía de formación de la cavidad	51
B.4. Longitud de Tolman por radio ponderado por superficie de la malla para la energía de formación de la cavidad	52

Índice de tablas

1.1. Energías libres de solvatación para esferas Lennard-Jones neutrales en agua SPC y SPC/E	8
3.1. Los mejores modelos encontrados para la aproximación de ΔG_{np} jerarquizados según su ρ	34
3.2. Los mejores modelos encontrados para la aproximación de ΔG_{np} jerarquizados según su RECM.	35
3.3. Los mejores modelos encontrados para la aproximación de ΔG_{np} jerarquizados según su EAM.	36
3.4. Los mejores modelos encontrados para la aproximación de ΔG_{cav} jerarquizados según su ρ	38
3.5. Los mejores modelos encontrados para la aproximación de ΔG_{cav} jerarquizados según su RECM.	39
3.6. Los mejores modelos encontrados para la aproximación de ΔG_{cav} jerarquizados según su EAM.	39

Introducción

El estudio de los factores y energías que rigen el fenómeno de disolución ha ganado territorio en el interés científico, entre otras cosas, debido a la promesa de la edición de bio-moléculas al interior de bacterias y virus. Estas moléculas se encuentran inmersas en medios acuosos y van desde un par de átomos hasta arreglos de millones de ellos formando por ejemplo: una proteína o una membrana celular lipídica.

La forma más precisa de estimar la energía necesaria para traer una molécula desde el vacío al interior de un medio acuoso, es decir, la energía de solvatación, es a través de una simulación explícita de solventes, cuya limitación es la complejidad de la molécula.

Los modelos implícitos son una alternativa más eficiente, pero generalmente menos precisa, para resolver simulaciones explícitas de solventes. Estos modelos usualmente separan la energía de solvatación en dos componentes: una polar y una no-polar.

Este trabajo propone un modelo implícito para la aproximación de la componente no-polar de la energía de solvatación. La cual en un principio se modelaba como una función del área superficial S de la forma $\Delta G_{np} \approx \gamma S$. Con γ como una constante de energía por unidad de área.

Más tarde, la deficiencia de esos modelos llevó a la descomposición de la componente no-polar en dos: energía de formación de cavidad ΔG_{cav} y energía de las interacciones de Van der Waals ΔG_{vdw} .

En otros estudios no se describe γ como una constante, sino que como una cantidad específica que depende de la topología local de la interfaz soluto-solvente, y que correspondería a una desviación de la tensión superficial γ_{lv} , que es la cantidad de energía necesaria para aumentar la superficie plana de la interfaz en una unidad. Esta desviación o corrección se predice lineal con la curvatura de la superficie y ha sido demostrado que γ_{lv} es el valor asintótico de γ para esferas de radios muy grandes (tendencia al caso planar) de interfases líquido-vapor de agua.

Los modelos que aproximan este comportamiento de la tensión superficial, en la hidratación de esferas no-polares, que representan superficies con una curvatura constante, caracterizan este desplazamiento como una función que involucra la curvatura y una longitud característica: la longitud de Tolman δ .

El caso de bio-moléculas existe una gran cantidad de curvaturas locales, debido a la cantidad y tipos de átomos que la conforman. Sus estructuras pueden ser vistas como una aglomeración de esferas con distintos radios (curvaturas) constantes. Pero debido a la unión de estas, las interfaces que producen y sus superficies tienen incluso curvaturas locales cóncavas, complicando la utilización de modelos bien definidos para una esfera.

El modelo implícito que se plantea trabaja bajo esta idea: descomponer la superficie y caracterizar cada una de las secciones con una curvatura local. Para luego integrar las energías que producen estas secciones esféricas (curvadas) y como resultado dar con una aproximación de la componente no-polar de la energía de solvatación.

A lo largo de este trabajo se evaluarán distintas consideraciones para la aproximación de cada una de las partes de esta integración. Para hacerlo, se probarán las combinaciones de estas y se contrastarán con resultados conocidos para 502 especies (moléculas distintas).

Los objetivos específicos de este trabajo son:

- Analizar las variables que se encuentran involucradas en la creación de interfases y arreglos hidrofóbicos.
- Probar la cantidad de energía requerida para generar una cavidad esférica de distintos diámetros.
- Diseñar modelos que empleen un análisis superficial de las geometrías de los solutos para la aproximación de la componente no-polar de la energía de disolución.
- Evaluar los métodos con especies con resultados conocidos (David L. Mobley, Christopher I. Bayly, Matthew D. Cooper, Michael R. Shirts, and Ken A. Dill - 2009).
- Analizar los resultados del desempeño de los modelos diseñados.
- Elegir y presentar el modelo implícito que mejor aproxima la componente no-polar de la energía de solvatación.

1 | Marco teórico

La combinación de sustancias hidrofóbicas e hidrofílicas es un fenómeno ubicuo que ocurre incluso al interior de nuestras células. Cuando sustancias de naturaleza hidrofóbicas (apolares) se mezclan en un medio acuoso es de esperarse que no sean bien disueltas ya que son repelidas por la masa de agua, que es un solvente polar. Pero además de esto, las moléculas hidrofóbicas se agrupan formando cúmulos o estructuras hidrofóbicas.

Este mismo fenómeno hidrofóbico, que ocurre cuando un sustrato es sumergido en agua, requiere de una cierta cantidad de energía para llevarse a cabo. La cuantificación de esta energía ha sido difícil ya que la hidrofobicidad es un fenómeno que depende de muchas variables. Una de las variables más importantes es la escala del sustrato, más específicamente las relaciones que existen entre el volumen y su superficie.

La separación de los componentes hidrofílicos e hidrofóbicos del agua genera una interfaz. La cual se extiende más allá de la distancia a la cual las moléculas de un líquido homogéneo interactúan entre ellas.

El principal efecto que hace un soluto al ser introducido al agua es actuar como una cavidad en su interior, desplazando una cierta cantidad de líquido, al mismo tiempo interrumpiendo las fuerzas de Van der Waals existentes entre las moléculas de agua desplazadas y generando fuertes fuerzas atractivas mediante sus componentes hidrofílicos.

Esta reorganización, de las moléculas que se encuentran en esta interfaz, requiere de un costo termodinámico para lograrse y para que la interfaz permanezca también necesita de la compensación de fuerzas que favorezcan la separación del sistema en fases diferentes.

Es por lo anterior que mientras más grandes sea la superficie, también crecerá el costo energético de la interfaz y al mismo tiempo se requerirá de mayores fuerzas compensatorias a medida que el crecimiento de esta superficie implique un aumento de volumen. Es por esto que el núcleo puede ser estable o metaestable solo si excede un tamaño crítico.

Las unidades hidrofóbicas inducen cierto orden en las moléculas de agua que las ro-

dean. Este cambio de orden, cambio de entropía, significa que debe añadirse energía al sistema.

Termodinámicamente hablando, este proceso de disolución tiene una probabilidad de suceder, lo cual se cuantifica en términos de energía libre ΔG . En el contexto de la energía de solvatación de una molécula, ΔG es el trabajo reversible para que el solvente se reorganice y disuelva el soluto. La energía libre tiene dos componentes principales $\Delta G = \Delta U - T\Delta S$, donde ΔU y ΔS son los cambios de energía interna y entrópicos en los que se incurre durante la disolución. La componente de energía interna es medida como la energía potencial de las interacciones entre moléculas, mientras que la parte entrópica mide el orden o las correlaciones intermoleculares.

Un proceso que requiere cambios significativos en el número de interacciones moleculares, como el quiebre de puentes de hidrógeno para formar una interfaz líquido-vapor será dominado por los cambios de energía interna ΔU . Mientras que un proceso que requiere una reorganización espacial de puentes de hidrógeno tendrá una importante componente entrópica ΔS .

Los modelos implícitos son comúnmente usados en la química teórica para estudiar la solvatación de sistemas biomoleculares. Estos modelos proveen una alternativa más eficiente, pero generalmente menos precisa, para la resolver, mediante un modelo, las simulaciones explícitas de solventes. La energía libre en estos modelos es usualmente separada en no-polar ΔG_{np} y polar ΔG_p [1].

$$\Delta G = \Delta G_{np} + \Delta G_p \quad (1.1)$$

La componente no-polar de la energía de solvatación ΔG_{np} corresponde al costo energético de la formación de la cavidad, reorganización del solvente e interacciones soluto-solvente producidas cuando un soluto es traído desde el vacío al solvente. Esta componente es comúnmente aproximada mediante modelos que trabajan con el área, por ejemplo $\Delta G_{np} \approx \gamma S$, donde S es la Solvent Accesible Surface y γ es una constante de energía por unidad de área. Estos métodos son deficientes y han llevado a descomposición de esta componente en la suma de la energía para formar la cavidad ΔG_{cav} y de la energía de las interacciones de Van der Waals ΔG_{vdw} [1], como se puede ver en la figura 1.1.

Es por lo anterior que en esta investigación se buscaran modelos para aproximar ΔG_{np} y para ΔG_{cav} .

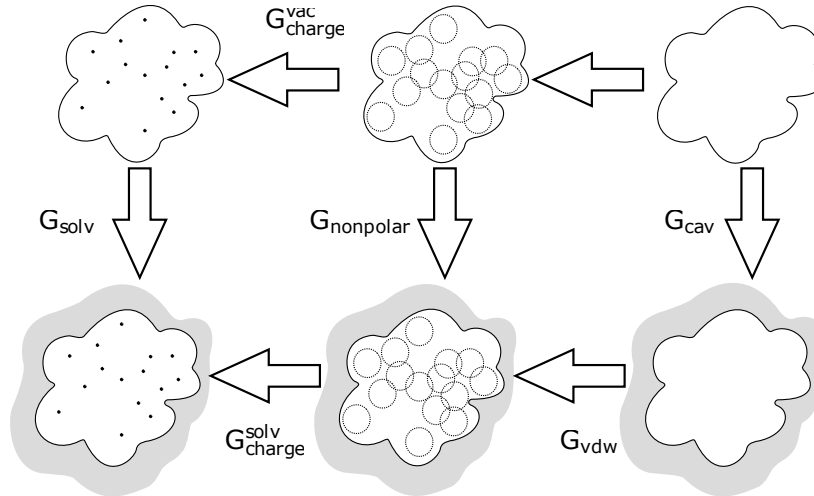


Figura 1.1: Energías involucradas en la solvatación [2].

En condiciones ambientales, es decir, a una presión de $P = 1[atm]$ y temperatura ambiente $T = 298[K]$, las fases líquida y gaseosas del agua coexisten y son sumamente cercanas. La diferencia de energía libre entre estas fases menor comparada con la energía necesaria para mover una molécula desde la fase líquida al vapor. Esta condición asegura que las cavidades en el agua tienen una interfaz entre las fases líquido-vapor, como ha sugerido [3], y se han confirmado mediante análisis teórico [4] [5] y simulaciones.

El costo de hidratar una cavidad esférica de radio R se encuentra bien modelado por:

$$\Delta G_{cav} \approx 4\pi R^2 \gamma + \frac{4\pi}{3} R^3 P \quad (1.2)$$

donde γ hace referencia a la tensión superficial entre las fases líquido vapor y P a la presión, ambas a las condiciones antes descritas [6].

Las simulaciones y trabajos previos en los que se basa esta investigación utilizaron una presión constante. Esto es considerado crucial, ya que el secado de la superficie es acompañado por significativas fluctuaciones del volumen. Una restricción de un volumen constante podría inhibir estas fluctuaciones y, por lo tanto, prevenir artificialmente

el secado de la superficie del soluto [5].

De esta ecuación se desprenden dos términos principales: el volumétrico, acompañado del volumen de la cavidad $\frac{4\pi}{3}R^3P$ y el superficie, caracterizado por la superficie de la cavidad $4\pi R^2\gamma$.

La influencia del término volumétrico es imperante para cavidades macroscópicas, pero pierde importancia, incluso se hace despreciable, a pequeñas escalas. La hidratación de pequeños solutos no interfiere con los puentes de hidrógeno, sino que los obliga a reordenarse. Esta interferencia u orden persiste solo hasta algunas moléculas de distancia, donde finalmente domina la organización aleatoria del líquido en estado puro.

Este cambio de comportamiento según las dimensiones del soluto no pueden esperarse que sea brusco, que a una cierta longitud característica comience a predominante una componente (superficial o volumétrica) de ΔG sobre la otra. La reorganización de moléculas de agua y separación de puentes de hidrógeno suceden de forma paulatina, tal y como se puede observar en la figura 1.2.

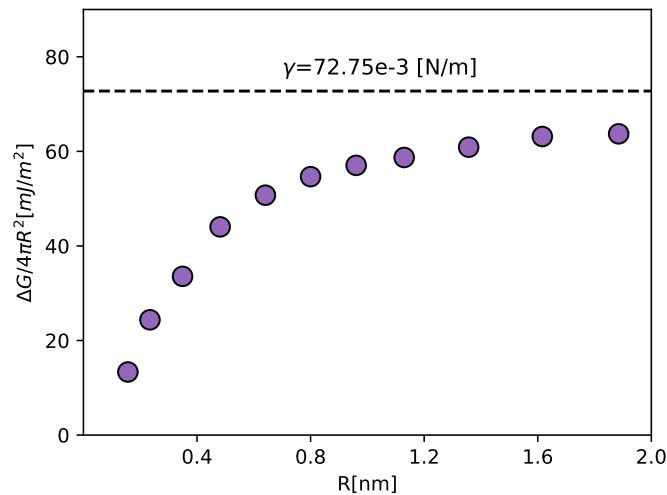


Figura 1.2: Energía de la formación de una cavidad ΔG_{cav} en función del tamaño de la cavidad. Adaptada de [6].

El parámetro γ se suele identificar con la tensión superficial de un solvente adyacente a un segundo medio. Esta tensión depende de las interacciones microscópicas entre los medios antes descritos. Estos efectos microscópicos son adecuadamente representados por una característica macroscópica como lo es γ si es que estas interacciones son

más pequeñas que las escalas que se estudiarán, como lo es el radio de curvatura o las distancias de la interacción entre las fases es entonces absorbido por el parámetro γ .

La dependencia de la curvatura ha sido un tema de investigación largamente debatido y que hasta el día de hoy sigue en discusión. Para el agua, que se encuentra cercana a su punto crítico a condiciones ambientales, γ no tiene una dependencia de la curvatura. Sin embargo la corrección de primer orden es aparentemente lineal con respecto a la curvatura según predice la teoría de escalamiento de partícula, la cual se emplea comúnmente para estudiar la solvatación de cavidades esféricas sólidas. Si bien este resultado es estrictamente para el caso de partículas esféricas, asumieron que se puede ser aplicado de manera local para cada uno de los puntos de la superficie [1].

Ha sido demostrado que la tensión superficial en una interfaz líquido-vapor plana γ_{lv} es un valor asintótico para la energía libre de solvatación por unidad de superficie para cavidades esféricas sólidas de grandes radios [4]. Este comportamiento asintótico también es apreciable en la figura 1.2.

Las superficies de (bio)moléculas presentan formas con curvaturas pronunciadas. Así que $\gamma(r)$ dependerá fuertemente de la geometría de la interfaz alrededor de r , una posición en la superficie. Una forma usual de modelar el desplazamiento del comportamiento de γ_{lv} debido a la curva que presenta localmente la superficie, es mediante la curvatura media H o del radio de curvatura medio R :

$$\gamma(r) = \gamma_{lv}(1 - 2\delta H(r)) = \gamma_{lv} \left(1 - \frac{2\delta}{R(r)} \right) \quad (1.3)$$

El mayor inconveniente de esta aproximación es que proporciona resultados no-físicos si el radio de la curvatura media R es más pequeño que el doble de la longitud de Tolman δ .

$$|R| < 2\delta \quad (1.4)$$

La longitud de Tolman δ es una longitud que indica hasta qué punto la tensión superficial de una interfaz curva se desvía del valor que tendría siendo plana γ_{lv} [7]. Se hablará más de esta longitud en la siguiente sección.

El método original [1] utilizaba las curvaturas locales principales de la superficie $k_1(r)$ y $k_2(r)$. Que junto a la siguiente expresión permiten la aproximación de la curvatura media local:

$$H(r) = \frac{k_1(r) + k_2(r)}{2} = R(r)^{-1} \quad (1.5)$$

En donde $R(r)$ es el radio medio de curvatura y ha sido obtenido en las cercanías de el vértice o sección de superficie caracterizado por r . A lo largo de este trabajo se utilizará la convención de que las curvaturas serán positivas para casos convexos (ej: cavidad esférica) y negativa para secciones de superficies cóncavas (ej: depresión esférica).

1.1. Longitud de Tolman

Para una superficie macroscópica, el coeficiente δ es la longitud de Tolman, la cual determina la energía libre en una interfaz curva.

En nuestros análisis de solutos microscópicos, δ describe el acercamiento hacia una asíntota en el escalamiento de ΔG con el área superficial y no necesariamente corresponde a la longitud de Tolman [5].

La longitud de Tolman puede entenderse como una diferencia entre dos radios que definen superficies equidensas (perpendiculares al gradiente de densidad) dentro de la zona de interfaz.

En una interfaz entre dos fases α y β que utilizan un determinado volumen V con una cierta cantidad de moles de ambas sustancias N_α y N_β por separado. Ocurre un fenómeno en que ambas fases se vuelven más densas respecto se aproximan a la zona de interfaz.

$$N_T = N_\alpha + N_\beta + N_S \quad (1.6)$$

La cantidad de moles N_S aparecen cuando una misma cantidad de volumen de ambas fases se encuentran en contacto y se vuelven mas densas en este.

El primer radio se define como el **radio equimolar** a_V y se caracteriza por ser el radio que define una superficie bajo la cual la contribución de N_S es nula, es decir, bajo esta superficie se cumple que: $N_T(r = a_V) = N_\alpha + N_\beta$.

La segunda superficie contempla que existe un radio al cual se cumple la relación de Kelvin, a esta superficie se le denomina **superficie de tensión** a_S [8].

$$\rho_\alpha - \rho_\beta = \frac{2\gamma_s}{a_s} \quad (1.7)$$

Donde γ_s es la tensión de la superficie curvada, es decir, es la tensión superficial que aparece corregida mediante la ecuación 1.3.

Una vez definidas estas superficies, la longitud de Tolman se define como:

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

Es esta diferencia entre las superficies definidas la que cambia según se curva la interfaz y genera un desplazamiento del comportamiento que tiene la tensión superficial en un plano [8].

En trabajos de otros autores [1] la longitud de Tolman se describe como un parámetro libre y se ha ajustado con el fin de dar con la energía de solvatación de los solutos. Algunos de estos ajustes nos resultan interesantes debido a la naturaleza de algunos de los solutos escogidos, gases nobles. Estos gases son inertes y se comportan como una cavidad esférica al interior del agua, siendo lo más próximo a una representación de la cavidad de un soluto no-polar, cuya interacción nos compete. Estas longitudes de Tolman del mejor ajuste son adjuntos en la tabla 1.1.

Tabla 1.1: Energías libres de solvatación ΔG para esferas Lennard-Jones neutrales en agua **Single Point Charge (SPC)** y **Single Point Charge/Extended (SPC/E)** [9], δ_{bf} es el valor de Tolman ajustado para obtener un resultado exacto, y R_{min} es el radio óptimo resultante que excluye solvente.

Soluto	ΔG_{sim} (kJ/mol)	δ_{bf} (Å)	R_{min} (Å)
Modelo SPC			
Ne	11.41	0.84	2.61
Ar	8.68	0.90	2.96
Kr	8.12	0.91	3.10
Xe	7.65	0.92	3.27
Modelo SPC/E			
Ne	11.65	0.88	2.60
Ar	8.83	0.96	2.94
Kr	8.20	0.98	3.09
Xe	7.58	1.00	3.25

2 | Desarrollo

2.1. Solutos esféricos no-polares

En primer lugar nos dedicaremos a analizar la ecuación 1.2 y así tener un mayor entendimiento de las componentes que dan forma al gráfico de la figura 1.2. El soluto que se modela mediante estas ecuaciones es perfectamente esférico y no-polar, disuelto en agua pura a condiciones ambientales (temperatura ambiente y presión 1[atm]).

Lo que se busca es encontrar la cantidad de energía por unidad de superficie que genera una esfera de radio R . De la ecuación 1.2 supusimos que γ se trataba de una constante y que por lo tanto esta relación quedaría de la forma:

$$\frac{\Delta G}{4\pi R^2} = \gamma_{lv} + \frac{R}{3} \quad (2.1)$$

En este caso la energía por unidad de área de una esfera crece linealmente con el radio de la esfera, divergiendo. Esta hipótesis fue descartada ya que en el caso extremo en que $R \rightarrow \infty$, que implica una superficie plana, la tensión superficial debería converger a la tensión superficial de un plano γ_{lv} .

Para el siguiente intento se utilizó la presión de una burbuja o estanque esférico para darle un modelo a la presión al interior de una interfaz curva. En este caso la presión sería la responsable de la curvatura de la interfaz.

$$P(R) = \frac{2\gamma_{lv}}{R} \rightarrow \frac{\Delta G}{4\pi R^2} = \frac{5\gamma_{lv}}{3} \quad (2.2)$$

Esta hipótesis también fue descartada debido a que en este caso la cantidad de energía por unidad de superficie es independiente del radio de la esfera y al mismo tiempo es siempre mayor a la tensión superficial del caso plano.

Finalmente se incluyó en el modelo la ecuación 1.3, cuya función es atenuar la participación del término superficial a medida que el crecimiento del radio vuelve plana la superficie para efectos de tensión superficial.

$$\Delta G = 4\pi R^2 \gamma_{lv} \left(1 - \frac{2\delta}{R}\right) + \frac{4\pi}{3} R^3 P \quad (2.3)$$

Los resultados de esta ecuación generan la figura 2.1, la que es similar a la figura 1.2. A esta escala el término volumétrico es muy inferior al superficial, el crecimiento de este último ocurre rápidamente, estabilizando su crecimiento en relación al área generada tras unos pocos nanómetros.

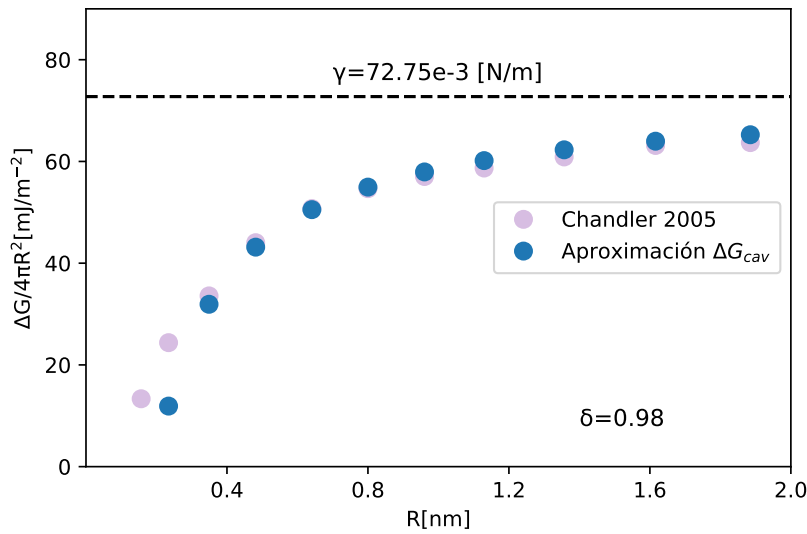


Figura 2.1: La tensión superficial $\gamma = 72,75e-3[N/m]$ se comporta como un asíntota de la cantidad de energía por unidad de área generada por un crecimiento del radio de la esfera sumergida. Para esta figura se ha escogido una longitud de Tolman constante $\delta = 0,98\text{\AA}$.

Estas son las variables involucradas en la disolución de esferas. La ecuación 2.3 será la que se empleará como punto de partida cuando se requiera calcular las componentes de ΔG para otras geometrías. Es por esto que se le dará especial tratamiento a la curvatura y superficie, debido a la importancia de la ecuación 1.3 en el modelo.

2.2. Geometrías

Los métodos presentados en este informe utilizan la geometría de la interfaz del soluto para calcular la componente no-polar de la energía libre de Gibbs. Esta geometría depende de la interacción entre el soluto y el solvente, las posiciones de los átomos que conforman al soluto levantan o contraen esta superficie, volviéndola más compleja.

Las mallas triangulares de las superficies **Solvent Excluded Surface (SES)** y **Solvent Accesible Surface (SAS)** fueron obtenidas usando software `msms` utilizando una probeta de $1,4\text{\AA}$, magnitud que corresponde al radio de la molécula de agua, como se puede ver en la figura 2.3. Las imperfecciones en las mallas generadas de este modo fueron reparadas mediante las herramientas proporcionadas por la librería de Python `trimesh`.

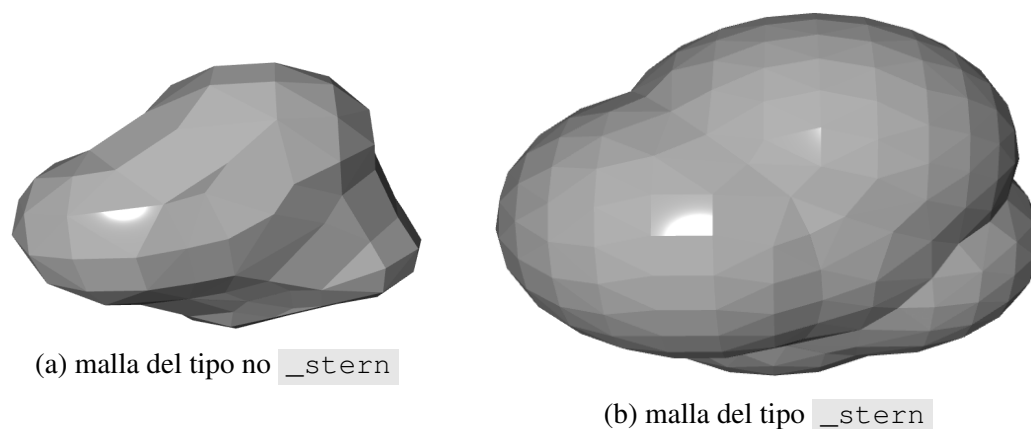


Figura 2.2:

Las geometrías utilizadas en este trabajo son del tipo `_stern` las cuales representan la **SAS**, mientras que geometrías sin `_stern` representan la **SES**.

Se puede observar que la malla `_stern` es más suave, tiene más vértices y caras, que la malla que no es de este tipo.

En ambos casos se trata de la especie *acetic acid*, su renderización se hizo gracias a librería `trimesh`.

Para este caso de estudio se cuenta con 502 especies, sus respectivas mallas y energías de solvatación no-polar ΔG_{np} , de Van der Waals ΔG_{vdw} y formación de la cavidad ΔG_{cav} [11].

2.3. El modelo JUALE

El propósito de los modelos que se explorarán a continuación es aproximar la componente no-polar de la energía libre de solvatación ΔG_{np} , análogamente se evaluará la aproximación mediante estos modelos de la energía de formación de cavidad ΔG_{cav} . Para ello utilizamos la ecuación 2.3 por cada i-especie.

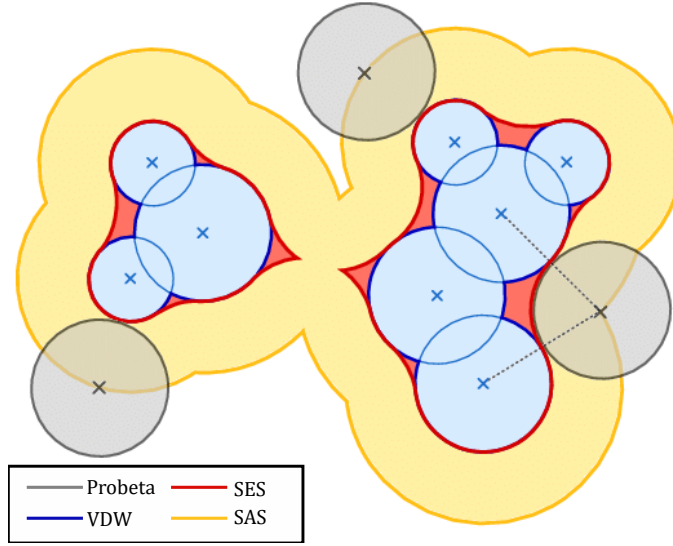


Figura 2.3: Representación de las distintas superficies. La superficie *VDW* es aproximada según las características de los átomos presentes en la molécula, luego una probeta la recorre trazando las *SAS* y *SES*. Imagen adaptada de [10].

$$\begin{aligned}
 \Delta G_{np_i} &= 4\pi R_i^2 \cdot \gamma_{lv} (1 - 2\delta H) + \frac{4\pi R_i^3}{3} P \\
 \Delta G_{np_i} &= A_i \cdot \gamma_{lv} (1 - 2\delta H) + V_i P \\
 \Delta G_{np_i} &= A_i \cdot \gamma_{lv} \left(1 - \frac{2\delta}{R} \right) + V_i P
 \end{aligned} \tag{2.4}$$

Comenzaremos por la componente volumétrica $V_i \cdot P$. Para aproximar el volumen utilizamos el método `mesh.volume` de la librería `trimesh` que nos permite conocer el volumen de la geometría analizada. Este, junto a la presión $P = 101300[Pa]$ conforman el término volumétrico.

El comportamiento de la ecuación 2.4 se encuentra bien definido para cavidades esféricas al interior de un solvente, esta configuración es un caso idealizado. Los solutos que podemos encontrar en la naturaleza tienen geometrías más complejas, al igual que las cavidades que estos forman. Es por esto que no es posible caracterizarlos por un solo radio R ni tampoco por una sola curvatura H .

Para trabajar las geometrías no esféricas de los solutos reales, se propone utilizar la forma integral[1] (de superficie) de la ecuación 2.4.

$$\Delta G_{np_i} = V_i \cdot P + \oint_V \gamma(r) ds = V_i \cdot P + \oint_V \gamma_{lv} \cdot \left(1 - \frac{2\delta(R(r))}{R(r)}\right) ds \quad (2.5)$$

Donde r es una porción de superficie. Para luego discretizar en pequeños m-cascos curvos y calcular la contribución de cada una de esta porciones.

$$\Delta G_{np_i} = V_i \cdot P + \left[\sum_{j=0}^m A_j \gamma_{lv} \cdot \left(1 - \frac{2\delta(R_j)}{R_j}\right) \right]_i \quad (2.6)$$

Para el término superficial $\left[\sum_{j=0}^m A_j \cdot \left(1 - \frac{2\delta(R_j)}{R_j}\right) \right]_i$ se dividirá la contribución por m-trozos de área que forman la superficie de la geometría. Para ello se evaluará la integración por vértice y por panel. En las siguientes secciones se detalla qué factores se ven involucrados en la selección de estos tipos de integración, como lo es el tipo de normal, área y curvatura que se emplearán para los modelos.

2.4. Normal

La efectividad de los algoritmos que manipulan y logran la visualización de mallas requieren información de las propiedades diferenciales de las mallas. Un ejemplo trivial es el de la normal de la superficie. La normal esta bien definida en las caras triangulares, pero cuando una malla se aproxima a una superficie suave se define generalmente en cada vértice y se interpola a las caras. Para hacerlo se requiere de alguna técnica para la estimación.

Todas las aproximaciones de curvaturas de una geometría son básicamente funciones derivadas de la normal, además la información contenida en esta magnitud es usada para muchos otros propósitos.

Frecuentemente deseamos computar la normal de la superficie en los vértices donde la normal es, en estricto sentido, no definida.

Para los modelos que se explorarán en este trabajo se utilizarán tres tipos de normal. Para la integración por panel se utilizará la bien definida normal del panel. Mientras que para la integración por vértice definiremos una forma de aproximación y también utilizaremos la información entregada junto con la geometría de las mallas.

2.4.1. Normal del panel

Para una malla triangular, la normal se encuentra claramente bien definida en las caras. Para un triangulo como el que se encuentra en la figura 2.4, la normal está dada por el producto cruz entre dos de las tres aristas que conforman sus vértices V_1 , V_2 y V_3 y la división de este vector por su propia magnitud como se muestra en la ecuación 2.7.

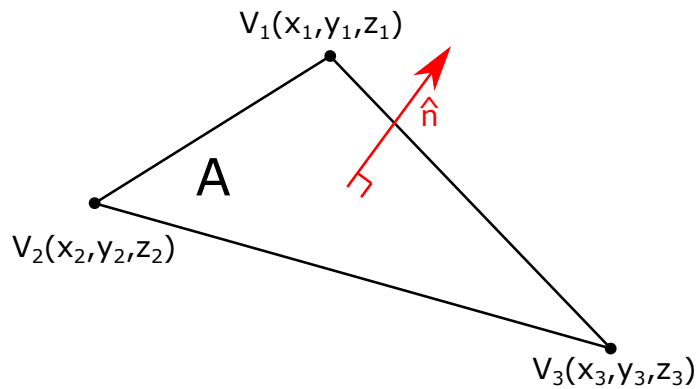


Figura 2.4: Ejemplo de cara de una malla triangular.

$$\hat{n} = \frac{\overrightarrow{V_2 - V_1} \times \overrightarrow{V_3 - V_2}}{\|\overrightarrow{V_2 - V_1} \times \overrightarrow{V_3 - V_2}\|} = \frac{\overrightarrow{V_3 - V_2} \times \overrightarrow{V_1 - V_3}}{\|\overrightarrow{V_3 - V_2} \times \overrightarrow{V_1 - V_3}\|} = \frac{\overrightarrow{V_1 - V_3} \times \overrightarrow{V_2 - V_1}}{\|\overrightarrow{V_1 - V_3} \times \overrightarrow{V_2 - V_1}\|} \quad (2.7)$$

Basta solo con una de las tres formas vistas en la ecuación 2.7 para definir la normal, ya que con dos vectores estaremos definiendo el plano en donde se encuentra la cara triangular.

El método `mesh.face_normals` retorna un arreglo con todas las normales de las caras de la malla. Una parte de las reparaciones realizadas a la malla en la sección 2.2 es el **winding**, la comprobación de que las normales apunten hacia el exterior del volumen. Para este efecto, el orden en el que se escogen los vectores para el producto cruz debe ser consistente para que todas apunten en el mismo sentido y horario si se quiere que sea hacia afuera.

2.4.2. Normal en el vértice: AWPN

Para una malla triangular, la normal se encuentra claramente definida en las caras, en dónde es un vector perpendicular a la cara, pero ¿Qué hacemos si la queremos definir en un vértice? ¿Qué propiedades requiere la normal en un vértice?.

Como primer requerimiento es que la normal debe ser similar a la normal de una superficie suave (si es que la malla se aproxima a esta superficie suave). En otras palabras, nos gustaría que la aproximación de esta normal converja a la normal de las superficie suave si es que refinamos la malla.

Si las caras son lo suficientemente refinadas, todas las normales alrededor del un vértice deberían tender a la misma normal a medida que refinamos la malla, y el promedio de las normales de las caras incidentes de un vértice deben satisfacer esta propiedad. Es más, una solución común para computar la normal en un vértice es simplemente tomar el promedio de las normales de las caras incidentes. Sin embargo esta no es la mejor solución. Para saber porqué, consideraremos 2 propiedades adicionales que necesitaremos para una buena normal del vértice.

- La normal debe ser independiente de la teselación y depender solo de la geometría de la malla. Por ejemplo: si quisiéramos añadir una arista a un [anillo de puntos vecinos](#) dividiendo una de las caras en dos triángulos tras añadir un vértice, esto no debería afectar la estimación de la normal en ese vértice.
- La normal debe permitirnos determinar si un punto se encuentra al interior o exterior de la malla.

La pseudo normal ponderada por el ángulo, en inglés [Angle Weighted Pseudo Normal \(AWPN\)](#), cumple con estas propiedades deseadas y se define para cada uno de los vértices cómo:

$$AWPN_i = \frac{\sum \alpha_j n_j}{\|\sum \alpha_j n_j\|} \quad (2.8)$$

donde i recorre los índices de todos los vértices de una malla y j recorre todas las caras incidentes dentro del [anillo de puntos vecinos](#). [12].

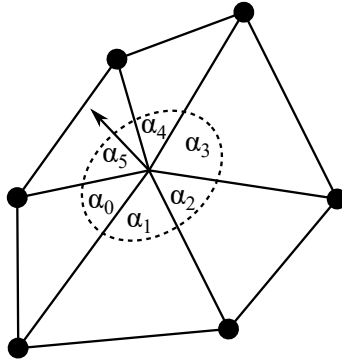


Figura 2.5: La AWPN es el promedio ponderado por el ángulo de las normales de las caras incidentes en un vértice.

2.4.3. Normal en el vértice: MSMS

El software [MSMS](#) con que se hicieron las mallas entrega un archivo para las caras, en donde se especifica qué vértices las conforman y un archivo con los vértices, en donde se indican las coordenadas de cada uno de ellos en las 3 primeras columnas. Este último archivo también contiene las normales, cuyo vector se especifica en las siguientes 3 columnas, calculadas mediante una técnica desconocida.

Estas normales fueron cargadas a la malla como datos de entrada. El método de la librería `trimesh mesh.vertex_normals` entrega la normal correspondiente a cada uno de los vértices. Si esta información no fuera cargada, las normales de los vértices sería calculada como el promedio de las normales de las caras incidentes de cada uno de los vértices, la cual como ya describimos en la sección [2.4.2](#) no es deseada.

El promedio de las normales de las caras incidentes, la [AWPN](#) y la normal entregada por el software [MSMS](#) son parecidas pero difieren en algunos puntos. Es por esto que probaremos sus desempeños por separado excepto la del promedio de las normales de las caras incidentes, cuyo comportamiento no es muy diferente de la [AWPN](#).

También probaremos la utilización de las normales de las caras, dividiendo la forma en que se realizará la integral en dos: por vértice y por panel.

2.5. Curvatura

La curvatura de la superficie que representa la interfaz entre el soluto y el solvente es de vital importancia para aproximar la desviación de la tensión superficial de su comportamiento en una interfaz plana, como se ha descrito en la ecuación 1.3.

Es comúnmente utilizada la curvatura media para caracterizar este desplazamiento del comportamiento, pero también creemos y exploraremos la utilización de cascos esféricos que logren un mejor ajuste en el [anillo de puntos vecinos](#). Ya que las esferas son la geometría que equilibra naturalmente las burbujas líquidas o gaseosas. Estos tipos de interfaces son dominadas por formas esféricas y simétricas [13].

2.5.1. Curvatura media

Podemos definir la curvatura normal en un punto como la curvatura de una sección normal. Una sección normal es la intersección de una superficie y un plano que contiene la normal. Rotando el plano alrededor de la normal, obtenemos diferentes curvaturas normales para cada ángulo. Para dos direcciones perpendiculares obtenemos respectivamente, la mas pequeña y la más grande de las curvaturas normales. La curvatura media es el promedio de estas curvaturas principales.

Para ponerlo de forma más explícita, imaginemos un punto p de una malla triangular y todos sus [puntos vecinos](#). Pensemos en el punto p tiene libre movimiento mientras que el resto del [anillo de puntos vecinos](#) se mantiene estático. Al moverse p , el parche al que pertenece se deforma sin alterar la superficie más allá del [anillo de puntos vecinos](#). Designemos A_{1-ring} como el área del parche. El área A_{1-ring} puede ser vista como una función de la posición de p y por tanto podemos definir el gradiente de A_{1-ring} .

La curvatura media está definida como el siguiente límite:

$$2H\vec{n} = 2\mathbf{H} = \lim_{A \rightarrow 0} \frac{\nabla A}{A} \quad (2.9)$$

donde el vector \mathbf{H} es la curvatura media. Para mallas triangulares este vector, por vértice, se define como el gradiente del área del [anillo de puntos vecinos](#) A_{1-ring} normalizado por el área del [anillo de puntos vecinos](#).

$$\mathbf{H}(p_i) = \frac{1}{4A_i^{1-ring}} \sum_{p_j \in \mathcal{N}_i} (\cot \alpha_{ij} + \cot \beta_{ik})(p_i - p_j) \quad (2.10)$$

En esta ecuación 2.10 sumamos todas las aristas haciendo un barrido del **anillo de puntos vecinos**, para luego ponderarlas con las cotangentes de los ángulos α_{ij} y β_{ik} opuestos a la arista que une a los puntos p_i y p_j , como se puede ver en la figura 2.6.

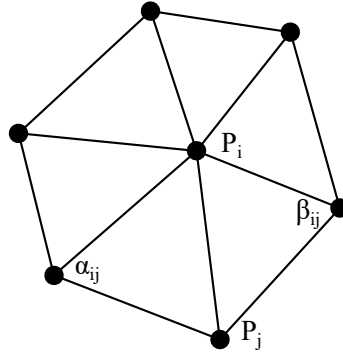


Figura 2.6: Punto principal p_i y su anillo de puntos vecinos, se ha escogido un punto p_j dentro de sus puntos vecinos para ejemplificar los ángulos α_{ij} y β_{ij} con los que se aproxima la curvatura media.

2.5.2. Mejor ajuste de una esfera

Gran parte de la investigación de la separación del comportamiento de la tensión superficial de una interfaz plana a una curva se ejemplifica mediante un soluto ideal que desplaza y crea un volumen perfectamente esférico.

Cuando nos enfrentamos a un soluto real, con geometría más compleja, notamos que su forma es básicamente un arreglo de esferas. Una de las primeras ideas en surgir fue descomponer esta geometría en cascos esféricos y tratarlos como partes de una esfera para la cual si estuviera bien estudiada su relación con la tensión superficial. Luego sumar la contribución de cada una de estos cascos para dar como resultado el comportamiento del soluto completo.

La curvatura en este caso, como se ve en la ecuación 1.5, es el inverso del radio de una esfera. Para un vértice y sus **puntos vecinos**, como el que se encuentra en la figura 2.7, la curvatura corresponde al inverso del radio de la esfera que logra el mejor ajuste de estos puntos a esta forma.

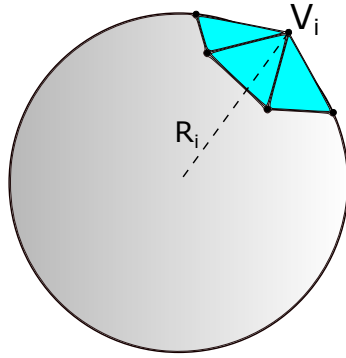


Figura 2.7: Un vértice V_i junto a su anillo de puntos vecinos son aproximados a una sección de esfera de radio R_i .

El mejor ajuste de una esfera para un grupo t ($t \geq 4$) de puntos se logra mediante mínimos cuadrados como se ve en el apéndice A. La función de la librería `trimesh` `trimesh.nsphere.fit_nsphere(points)` permite la obtención de esta esfera, centro y radio utilizando como datos de entrada un arreglo con las coordenadas de los t puntos.

Una vez obtenido el centro C_{esf} y radio R_{esf} de la esfera de mejor ajuste exploramos dos formas de caracterizar el vector de curvatura en un determinado vértice.

Centro y vértice

La primera forma es tomar las coordenadas del vértice V_i y restarle las del centro C_{esf} , de esta forma obtendremos un vector que apunta hacia el vértice y su magnitud es similar, en la mayoría de los casos, al radio R_{esf} obtenido por la función de `trimesh`.

$$\vec{R}_{cv} = \overrightarrow{V_i - C_{esf}} \quad (2.11)$$

Este vector \vec{R}_{cv} caracteriza la curvatura del vértice V_i con el centro de la esfera que hace el mejor ajuste a su **anillo de puntos vecinos**, pero mantiene la magnitud que corresponde a la distancia entre el centro y el vértice.

Centro y radio

La segunda forma utiliza la dirección del vector \vec{R}_{cv} pero se le asigna la magnitud R_{esf} con el propósito de rescatar la influencia del **anillo de puntos vecinos** sobre la curvatura

en el vértice V_i .

$$\vec{R}_{cr} = \frac{\vec{R}_{cv}}{\|\vec{R}_{cv}\|} \cdot R_{esf} \quad (2.12)$$

De esta forma se obtiene un vector que apunta hacia el vértice V_i cuya magnitud es la del radio de la esfera de mejor ajuste.

2.5.3. Paneles curvos

Con el propósito de usar la bien definida normal en un panel y su área, como se vio en la sección 2.4.1, se interpolará la curvatura de los vértices al panel mediante un promedio vectorial. Este promedio podrá ser tanto de la curvatura media (sección 2.5.1) o del radio de curvatura de la esfera de mejor ajuste (sección 2.5.2).

La curvaturas locales en las mallas puede ser considerada negativa, según se explica en la sección 2.6. Se analizará si los modelos se comportan mejor teniendo esta consideración en la integración de la ecuación 2.6 en la superficie.

2.6. Secciones cóncavas y convexas

Las geometrías analizadas distan de ser cavidades esféricas convexas, para las cuales es más conocida su contribución energética. Regularmente las superficies de las especies no presentan una curvatura constante (como el caso de la esfera), sino que cambian, algunas incluso tienen zonas convexas.

En las siguientes secciones se explorará el significado de la existencia curvaturas cóncavas en las geometrías de los solutos. Se probarán tres instancias, dentro del modelo, en la que estas zonas cóncavas podría restar o tener una menor contribución a ΔG_{np} o ΔG_{cav} , resaltadas en rojo en la ecuación 2.13.

$$\Delta G_{np_i} = V_i \cdot P + \left[\sum_{j=0}^m \pm A_j \cdot \left(1 - \frac{2\delta(\pm R_j)}{\pm R_j} \right) \right]_i \quad (2.13)$$

- **Área:** Es posible que una zona cóncava en la superficie de una especie reste

energía de solvatación. Es por esto que se probará la resta de la energía que corresponde a esa sección.

- **Radio de Tolman:** Se evaluará si la longitud de Tolman en una zona cóncava con un argumento (curvatura) negativo, mejora el modelo para la aproximación de la energías ΔG_{np} y ΔG_{cav} . Los modelos para la longitud de Tolman δ se describen más adelante en la sección 1.1.
- **Curvatura:** Se comparará el desempeño de modelos que consideren las curvaturas de las zonas cóncavas como negativas frente a los que solo consideren su magnitud (valor absoluto).

Para determinar si una sección de la geometría es cóncava o convexa se puede utilizar el producto punto de la curvatura de esa sección, vértice o panel, con su normal. Si el producto punto entre estos vectores es positivo quiere decir que entre estos vectores existe un ángulo inferior a $\frac{\pi}{2} [rad]$, por lo tanto tienen sentidos similares, indicando que se trata de una sección convexa de la malla. Por el contrario, si el producto punto es negativo quiere decir que entre estos vectores existe una diferencia angular superior a $\frac{\pi}{2} [rad]$ y por lo tanto los vectores tienen sentidos contrarios, indicando que se trata de una sección cóncava de la geometría.

Ejemplos los dos casos explicados en el párrafo anterior se pueden observar en la figura 2.8.

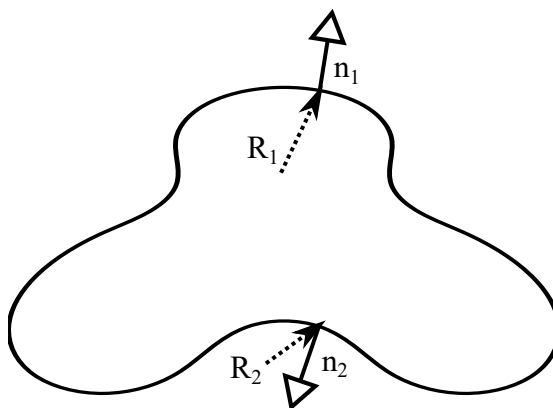


Figura 2.8: El producto punto entre el radio \vec{R}_1 y la normal \vec{n}_1 es positivo, mientras que el producto punto entre el radio \vec{R}_2 y la normal \vec{n}_2 es negativo. Ocurre de manera similar al trabajar con \vec{H} en lugar de \vec{R} .

Esta información respecto a la concavidades locales de la geometría se utilizará para determinar el signo, + o -, con el que se integrará la ecuación 2.13.

2.7. Área

Para el término superficial de ΔG_{np} se explorará la integración de la malla por vértice y por panel, con la finalidad encontrar el área que mejor caracteriza la zona en donde se debe integrar la curvatura.

2.7.1. Por panel

El área por panel se encuentra bien definida ya que se trata de triángulos y su suma corresponde a la superficie de la malla. El método `mesh.area_faces` entrega un arreglo con todas las áreas de las caras de la malla que se esté analizando.

2.7.2. Por vértice: área tercio

Para la integración por vértice, el área se vuelve más compleja ya que, en estricto sentido, el área solo se encuentra definida en los paneles. La forma más sencilla de dividir la superficie de la malla por vértice es fraccionar las caras en tercios y asignarle un tercio a cada uno de los vértices que forman el triángulo o, visto de otra forma, asignarle a cada vértice un tercio del área de su [anillo de puntos vecinos](#) como se ve en la figura 2.9.

2.7.3. Por vértice: área mixta

Asignar un tercio del área de los triángulos no es ideal, nos gustaría asignarle a cada vértice el área de la región más cercana a ese vértice. En otras palabras la región de Voronoi. Desafortunadamente, la región de Voronoi no esta contenida en el [anillo de puntos vecinos](#) en presencia de triángulos obtusángulos, como se puede ver en la figura 2.10. Para ello se ideó el siguiente algoritmo, que usa la simbología de la figura 2.10:

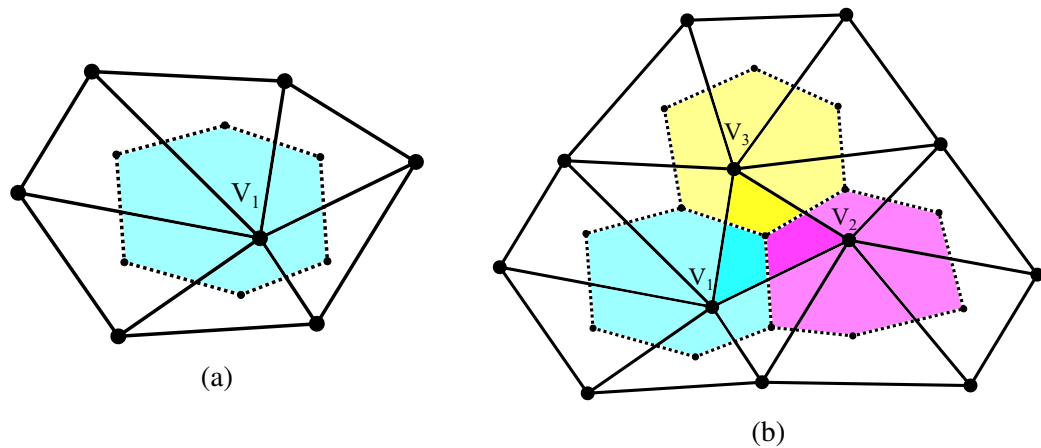


Figura 2.9:

a) Al vértice V_1 se le asigna un tercio de las áreas de sus caras incidentes, equivalente a un tercio del área de su **anillo de puntos vecinos**. b) Los vértices V_1 , V_2 y V_3 toman cada uno un tercio del triángulo formado por sus aristas. Esta forma de división de área cubre completamente la superficie y no se superpone.

Algoritmo para la aproximación del área mixta

1. Sea $A_{mix} = 0$
2. Para cada cara f incidente al vértice i :
3. Si f no es un triángulo obtusángulo:
4. $A_{mix} = A_{mix} + \frac{1}{8}(\cot(\alpha_{ij})\|P_i - P_k\|^2 + \cot(\beta_{ik})\|P_i - P_j\|^2)$
5. Sino, si f tiene su ángulo obtuso en el vértice i :
6. $A_{mix} = A_{mix} + \frac{1}{2}A(f)$
7. De lo contrario:
8. $A_{mix} = A_{mix} + \frac{1}{4}A(f)$

Este algoritmo usa la región de Voronoi tanto como es posible pero se restringe al área del **anillo de puntos vecinos**. Las áreas también cubren toda la superficie y no se superponen [12].

Así se puede caracterizar de forma más minuciosa el área al cual se le está asignando la curvatura del vértice.

Como se vio en la sección 2.6, la contribución del área podría ser negativa si se tratase de una sección cóncava. Esto significaría que la formación de esta región de interfaz,

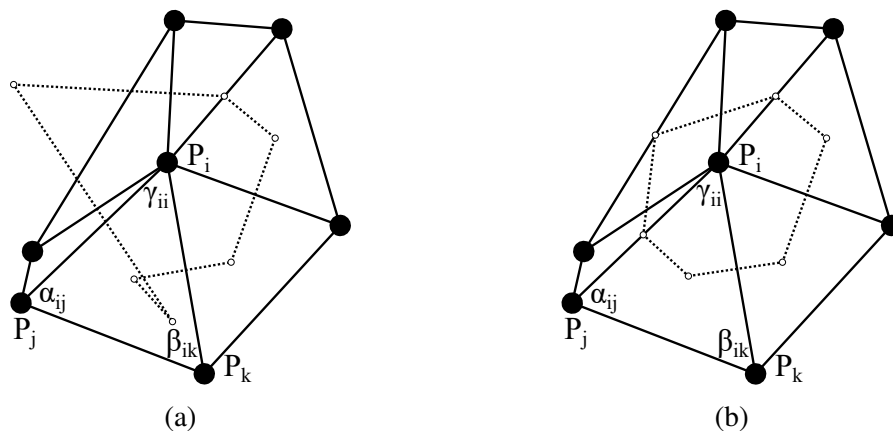


Figura 2.10:

a) La región de Voronoi no está incluida dentro del **anillo de puntos vecinos** en el caso de los triángulos obtusángulos. b) Región utilizada para el computo del área mixta, por cada triángulo obtusángulo se conectan los vértices de la región al punto medio de la arista opuesta al ángulo obtuso

visto desde la perspectiva plana, resta energía de solvatación.

2.8. Tipo de aproximación de la longitud de Tolman

La longitud de Tolman δ es la magnitud que caracteriza el desplazamiento del comportamiento de la tensión superficial γ_{lv} desde una interfaz plana a una curvada.

Lamentablemente no se cuenta con un modelo que nos permita aproximar δ . Lo que si se tiene es la restricción vista en la ecuación 1.4, pero esta información no es suficiente para conocer el comportamiento de esta característica, por lo que se buscarán formas simples de hacerlo.

Los modelos con los que se intentará abordar el comportamiento de esta longitud consideran que es una constante o que es una función cuya única dependencia es de la curvatura.

2.8.1. SPC y SPC/E

El primer intento por aproximar δ fue la utilización de datos que ya hubiesen sido calculados, para ello nos basamos en los datos de la tabla 1.1, los cuales fueron hechos para gases nobles. En esta tabla se encuentran los mejores ajustes de la longitud de Tolman

δ_{bf} y los radios generados por los solutos mono-atómicos R_{min} disueltos utilizando modelos de agua [SPC](#) y [SPC/E](#).

El objetivo es encontrar una función que relacione la longitud de Tolman con curvatura, es por esto que se hace una regresión sobre los puntos de esta tabla, como se ve en la figura 2.11.

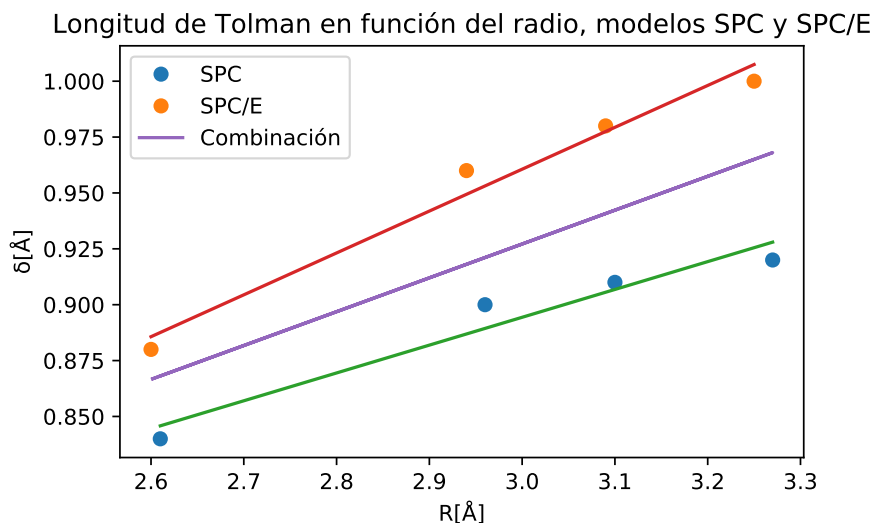


Figura 2.11: Las regresiones de los datos aquí mostrados para la aproximación de δ en modelos de agua [SPC](#) y [SPC/E](#) se encuentran en la ecuación 2.14.

$$\begin{aligned}\delta_{SPC}(R) &= 0,125R + 0,521 \\ \delta_{SPC/E}(R) &= 0,187R + 0,399\end{aligned}\tag{2.14}$$

Se ha decidido realizar líneas de tendencia de primer orden debido a que el grupo de puntos con que se cuenta es muy reducido al igual que su dominio. Buscar ajustes de grado dos o superiores genera grandes desviaciones para radios fuera del este dominio.

2.8.2. Longitud de Tolman constante, polinomio de grado 1 y tensión superficial contante

El segundo intento para la aproximación de la longitud de Tolman se desprende de los datos de las 502 especies con las que se esta trabajando, para las cuales se conoce su ΔG_{np} . Se hará una relación entre el valor de ΔG_{npi} (energía de solvatación no-polar

por i-especie) y un radio de curvatura característico R_{ci} de su geometría, para el cual se proponen y describen cuatro formas a lo largo de esta sección:

- \bar{R}_i : radio de curvatura promedio de los k-átomos que componen la molécula.
- \bar{R}_{si} : radio de curvatura ponderado por la superficie de los k-átomos que componen la molécula.
- \bar{R}_{in} : radio de curvatura promedio de los vértices de las mallas que representan la SAS de la molécula.
- \bar{R}_{ai} : radio de curvatura ponderado por el área asociada a los vértices de las mallas (como se ve en la sección 2.7.2) que representan la SAS de la molécula.

Dado un ΔG_{npi} por especie es posible despejar la tensión superficial por especie γ_i .

$$\Delta G_{npi} = V_i \cdot P + \gamma_i \cdot A_i \rightarrow \gamma_i = \frac{\Delta G_{npi} - V_i \cdot P}{A_i} \quad (2.15)$$

Donde V_i y A_i son el volumen y área por especie respectivamente. Conocer la tensión superficial por especie permite saber que tanto se desvía el comportamiento de esta interfaz respecto a una interfaz plana. El promedio de estas tensiones superficiales por especie resultó ser $\bar{\gamma}_{inp} = 4,18E-3[N/m]$, en lugar de $\gamma_{lv} = 72,75E-3[N/m]$ que es la tensión superficial del caso plano. Esta tensión superficial promedio se probará por si sola como modelo para la obtención de ΔG_{np} siguiendo la ecuación 2.15 bajo el nombre de `SoloGamma_gnp`.

Para el caso de ΔG_{cav} se hizo de forma análoga, cambiando ΔG_{npi} por ΔG_{cavi} (energía de cavitación por i-especie) en la ecuación 2.15. El promedio de las tensiones superficiales resultantes de este modo es de $\bar{\gamma}_{icav} = 25,06E-3[N/m]$. Esta tensión superficial promedio también se probará por si sola como modelo para la obtención de ΔG_{cav} con el nombre de `SoloGamma_cav`.

Para un radio característico R_{ci} se utilizarán los datos con los que contamos para la fabricación de las mallas, los que incluyen la localización de los centros de los átomos que forman las moléculas y sus radios. Estos radios corresponden a las esferas que dan origen a la SES y caracterizan la geometría.

Debido a que el radio de cada uno de los k -átomos que constituyen una molécula esta relacionado con la superficie que genera se ha decidido, además utilizar un promedio de los radios de las esferas de los átomos \bar{R}_i , se incluye la utilización de un promedio ponderado por la superficie de la esfera que generaría el átomo. Este radio por especie \bar{R}_{si} se calcula de la siguiente manera:

$$\bar{R}_{si} = \frac{\sum_{j=1}^k 4\pi R_j^2 \cdot R_j}{\sum_{j=1}^k 4\pi R_j^2} = \frac{\sum_{j=1}^k R_j^3}{\sum_{j=1}^k R_j^2} \quad (2.16)$$

La ponderación con la superficie se hace con el propósito de conservar la relación entre el radio, la superficie que genera y la energía involucrada en la creación de la interfaz que ocupa esa superficie. En otras palabras, el radio promedio de la superficie.

Conociendo la tensión superficial por especie γ_i y el radio característico R_{ci} de esta, se puede despejar el factor y por tanto la longitud de Tolman característica por especie δ_i involucrada en la desviación anteriormente presentada.

$$\gamma_i = \gamma_{lv} \left(1 - \frac{2\delta_i}{R_{ci}} \right) \rightarrow \delta_i = \frac{R_{ci}}{2} \left(1 - \frac{\gamma_i}{\gamma_{lv}} \right) \quad (2.17)$$

El promedio de las longitudes de Tolman calculadas de esta forma $\bar{\delta}_i$ para el radio promedio es de $\bar{\delta}_i(\bar{R}_i) = 1,4429[\text{Å}]$, mientras que para el radio ponderado por superficie es de $\bar{\delta}_i(\bar{R}_{si}) = 1,4672[\text{Å}]$.

Una vez obtenidos las longitudes de Tolman por especie δ_i y sus respectivos radios característicos R_{ci} , se utilizará una regresión por mínimos cuadrados como se ve en las figuras 2.18 y 2.13 para utilizar estas funciones como modelo. Las líneas de tendencia de estas figuras se encuentran en las ecuaciones 2.18 y 2.19 respectivamente.

$$\delta_{\bar{R}_i}(R) = 0,5028R - 0,0967 \quad (2.18)$$

$$\delta_{\bar{R}_{si}}(R) = 0,595R - 0,386 \quad (2.19)$$

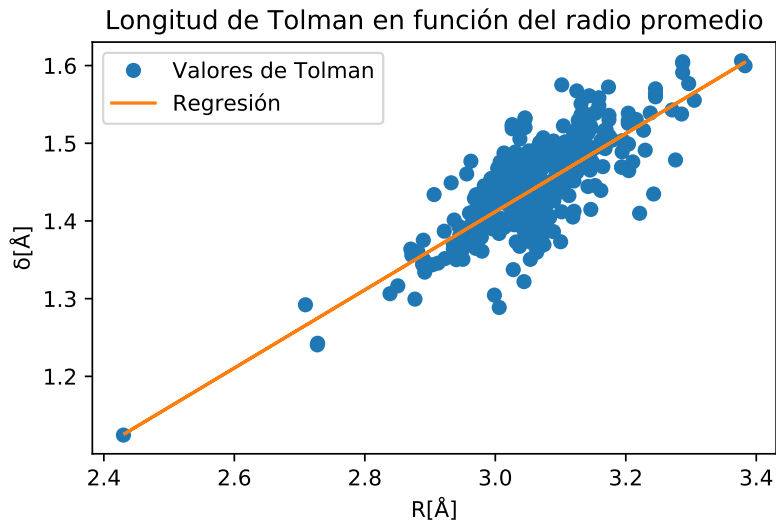


Figura 2.12: La regresión ($\rho = 0,7625$) de los datos aquí mostrados para la aproximación de δ utilizando un radio promedio se encuentra en la ecuación 2.18.

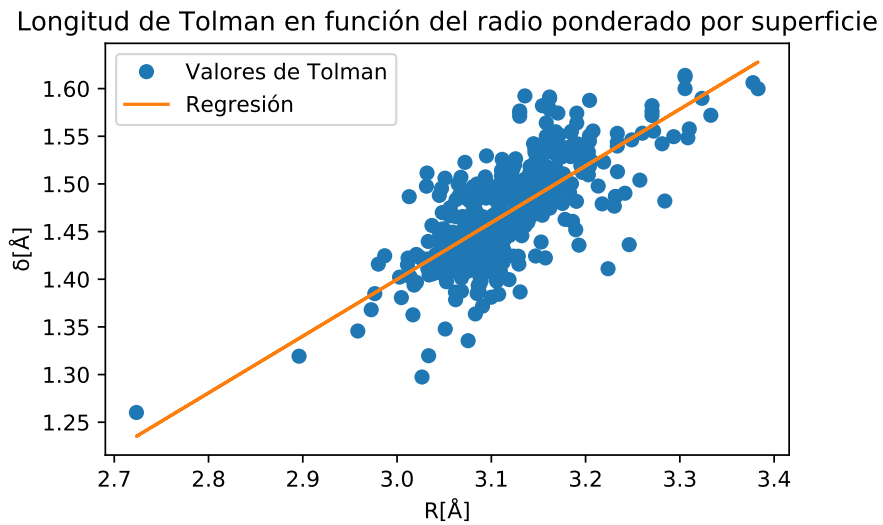


Figura 2.13: La regresión ($\rho = 0,7315$) de los datos aquí mostrados para la aproximación de δ utilizando un radio ponderado por superficie se encuentra en la ecuación 2.19.

Se ha decidido hacer regresiones de primer orden debido a la tendencia que presentan los datos en los gráficos.

Radios de las mallas

Siguiendo la misma lógica de las regresiones realizadas con el radio promedio y radio ponderado por superficie de los átomos se realizarán otras dos iteraciones. Esta vez en lugar de usar los radios de los k -átomos que dieron forma a la malla, se emplearán los radios del mejor ajuste de una esfera en los n -vértices de las mallas como se presentó en la sección 2.5.2.

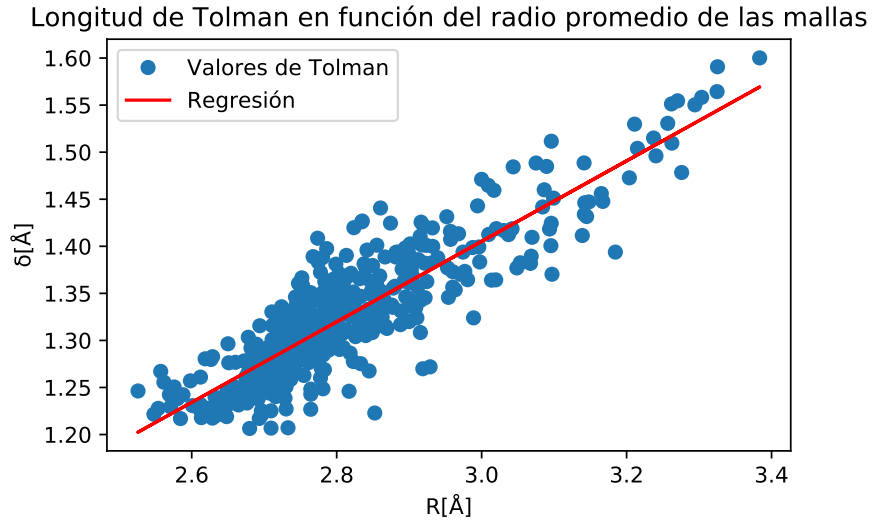


Figura 2.14: La regresión ($\rho = 0,8828$) de los datos aquí mostrados para la aproximación de δ utilizando un radio promedio de la malla se encuentra en la ecuación 2.20.

$$\delta_{\bar{R}_{in}}(R) = 0,4276R + 0,1224 \quad (2.20)$$

Para el radio ponderado por superficie basado en las mallas, se utilizará el área de cada n -vértice calculada de la forma vista en la sección 2.7.2. La ponderación de este radio se hace de forma análoga a la vista en la ecuación 2.16.

$$\bar{R}_{ai} = \frac{\sum_{j=1}^n A_j \cdot R_j}{\sum_{j=1}^n A_j} \quad (2.21)$$

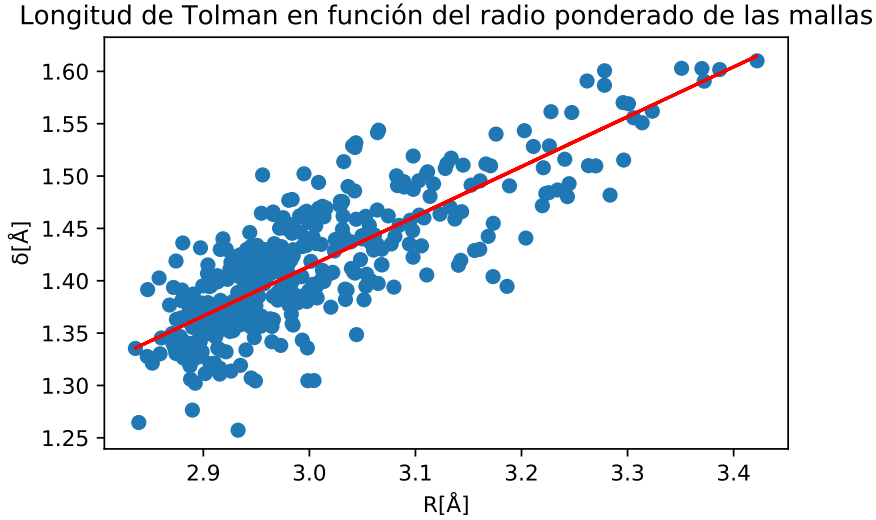


Figura 2.15: La regresión ($\rho = 0,8201$) de los datos aquí mostrados para la aproximación de δ utilizando un radio ponderado por la el área relacionado a los vértices de la malla se encuentra en la ecuación 2.22.

$$\delta_{R_{ai}}^-(R) = 0,4760R + -0,0141 \quad (2.22)$$

Análogamente a los modelos para la longitud de Tolman δ descritos en esta sección (2.8.2), que buscan una buena aproximación de ΔG_{np} , se hicieron modelos que buscaron aproximar ΔG_{cav} . Los cuales, por simplicidad de este informe, son agregados en el anexo B.

Las regresiones que relacionan la longitud de Tolman δ con los radios característicos de las geometrías exploradas en esta sección tienen un dominio positivo. Esto se debe a que para el caso de los radios característicos que fueron extraídos de los k-átomos, todos son considerados curvaturas convexas, lo mismo ocurre con las regresiones para los datos de la sección 2.8.1. En el caso de los radios característicos que utilizan la curvatura de la mallas, solo el 4,92% de las curvaturas de todas las mallas son cóncavas, es por esto que es esperable que tanto los radios promedios y como ponderados por especies también resulten ser positivos.

De la sección 2.6 se estima que el argumento de las funciones que se extraen de estas regresiones, la curvatura local de las mallas, podría ser negativa. Si bien ninguna de las regresiones vistas tiene dominio en radios de curvatura negativos, se considerarán mo-

delos que evalúen estas funciones para la longitud de Tolman δ con curvaturas negativas en las zonas de las mallas que sean cóncavas.

3 | Resultados

Tras haber expuesto las distintas formas de calcular las partes de la ecuación 2.6, se exponen los resultados de las combinaciones de los distintos factores de aproximación de la ΔG_{np} de las 502 especies. También se explorará la aproximación de la ΔG_{cav} utilizando estos modelos.

Los factores son todas las consideraciones y formas de aproximar los elementos de la ecuación 2.6. Sus combinaciones ascienden a un total de 1442 y son agrupados en 7 familias, las cuales se desglosan a continuación:

- **Tipo de curvatura:** Mean curvature `mc`, radio de mejor ajuste - centro/radio `bf_esf` y radio de mejor ajuste - centro/vértice `bf_cci`.
- **Tipo de normal:** Normal del panel `panel`, normal por vértice - [AWPN](#) `code-vert_AWPN`, normal por vértice - MSMS `vert_456`.
- **Tipo de área:** Área por panel `a_panel`, por vértice - área tercio `a/3` y por vértice - área mixta `a_mix`.
- **Tipo de aproximación de la longitud de Tolman (regresiones):**
 - Modelos de agua computacional SPC y SPC/E:** Factores basados en los datos de la tabla 1.1 `spc`, `spce`.
 - GNP:** constante `cte_gnp`, utilizando el radio de los átomos ponderado por superficie `p1_ponderado_gnp`, utilizando el radio promedio de los k-átomos `p1_promedio_gnp`, utilizando el radio de las mallas ponderado por superficie `p1_mesh_ponderado_gnp` y utilizando el radio promedio de las mallas `p1_mesh_promedio_gnp`.
 - CAV:** constante `cte_cav`, utilizando el radio de los k-átomos ponderado por superficie `p1_ponderado_cav`, utilizando el radio promedio de los k-átomos `p1_promedio_cav`, utilizando el radio de las mallas ponderado por superficie `p1_mesh_ponderado_cav` y utilizando el radio promedio de las mallas `p1_mesh_promedio_cav`.
- **Concavidad aplicada al área:** Áreas negativas para curvaturas cóncavas `area_c_s` y áreas siempre positivas `areas_s_s`.

- **Concavidad aplicada al radio de Tolman:** radios negativos para curvaturas cóncavas `radio_tolman_c_s` y radios siempre positivos `radio_tolman_s_s`.
- **Concavidad aplicada a la curvatura:** curvaturas en caso de ser cóncavas `mag_c_s` y curvaturas siempre positivas `mag_s_s`.

Cabe destacar que no todas las combinaciones entre estos factores es posible o tienen sentido, un ejemplo de esto sería utilizar el área de los paneles junto con la normal [AWPN](#) ya que se encuentran definidas para elementos distintos (paneles y vértices).

Dentro de los 1442 modelos nacidos de las combinaciones de factores, existen dos modelos que no contemplan ningún análisis de la superficie [`SoloGamma_gnp`] y [`SoloGamma_cav`]. La naturaleza de estos dos modelos se explica en la sección [2.8.2](#). Estos sirven para contrastar los resultados de los otros modelos que emplean el análisis de geométrico de la superficie de los solutos.

3.1. Componente no-polar de la energía de solvatación

Los cinco mejores modelos obtenidos de las combinaciones de factores anteriormente vistas y su comportamiento respecto a valores nominales se encuentran en la figura [3.1](#).

Estos modelos fueron jerarquizados según su [coeficiente de pearson](#), el cual es un índice que suele utilizarse para medir el grado de correlación lineal, por lo que estos modelos son los que mejor correlación tienen con los valores nominales.

También se ha incluido en la figura [3.1](#) los resultados del modelo `SoloGamma_gnp`, este modelo es la referencia a lo que sucede si se hace caso omiso al análisis de la superficie y su curvatura. Como se puede ver en la tabla [3.1](#), existen 411 modelos que tendrían una mejor correlación que `SoloGamma_gnp`.

Del gráfico [3.1](#) se observa que los datos no tienen muy buena correlación con los datos nominales, sin embargo, de los primeros 5 mejores modelos se puede decir que la consideración de la curvatura como el radio de una esfera de mejor ajuste es superior a la utilización de la curvatura media `mc` pero parece no tener una mayor diferencia entre sus sub categorías `bf_cci` y `bf_esf`. Lo mismo ocurre con el tipo de normal, las normales por vértice, `vert_AWPN` y `vert_456`, han demostrado un

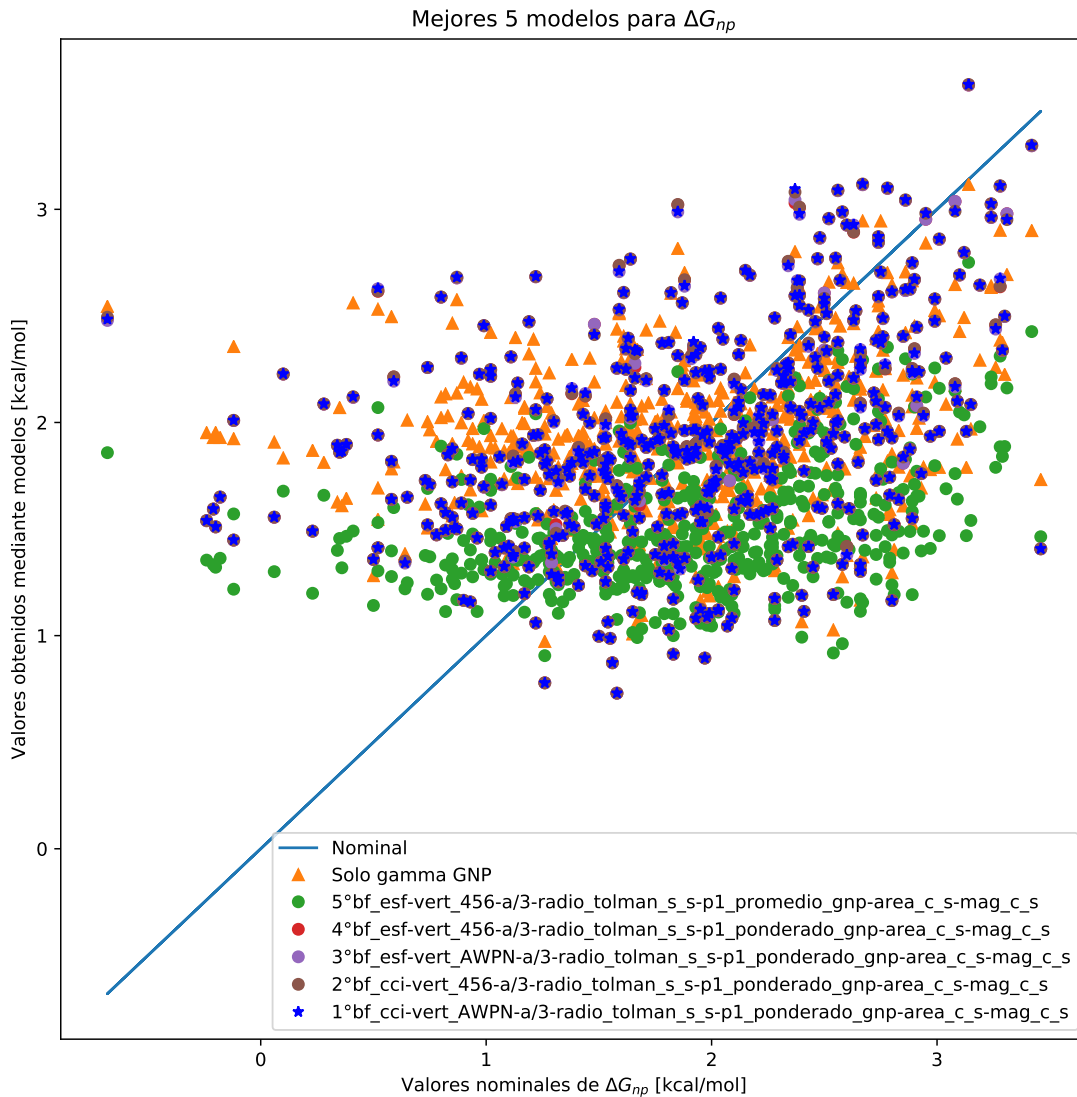


Figura 3.1: Cinco mejores modelos para ΔG_{np}

Tabla 3.1: Los mejores modelos encontrados para la aproximación de ΔG_{np} jerarquizados según su ρ .

Modelo	ρ	RECM	%EAM
1° bf_cci-vert_AWPN-a/3-radio_tolman_s_s-p1_ponderado_gnp-area_c_s-mag_c_s	0.3858	0.6933	61.08
2° bf_cci-vert_456-a/3-radio_tolman_s_s-p1_ponderado_gnp-area_c_s-mag_c_s	0.3856	0.6934	61.1
3° bf_esf-vert_AWPN-a/3-radio_tolman_s_s-p1_ponderado_gnp-area_c_s-mag_c_s	0.3855	0.6933	61.07
4° bf_esf-vert_456-a/3-radio_tolman_s_s-p1_ponderado_gnp-area_c_s-mag_c_s	0.3852	0.6933	61.09
5° bf_esf-vert_456-a/3-radio_tolman_s_s-p1_promedio_gnp-area_c_s-mag_c_s	0.3822	0.7667	54.52
412° SoloGamma_gnp	0.2159	0.7311	66.97

mejor comportamiento que las que se propusieron como integrales por panel `panel`. En cuanto al modelo de aproximación del comportamiento de la longitud de Tolman δ respecto a la curvatura, puede verse que cuatro de los cinco modelos presentados utilizan `p1_ponderado_gnp`, mientras que uno solo, el último de estos cinco, utiliza el promedio. Para el tipo de área el resultado es el mismo en esta muestra, todos estos modelos utilizan `a/3`.

Con respecto a la consideración de zonas cóncavas en la malla, como se vio en la sección 2.6, los modelos de Tolman en esta tabla muestran un mejor comportamiento cuando no es considerada negativamente las curvaturas cóncavas `tolman_s_s`. Pero en cuanto a las áreas y la curvatura de la ecuación 2.13, los modelos presentan un mejor comportamiento al considerar las curvaturas cóncavas como negativas `area_c_s` `mag_c_s`.

Como se puede ver en la tabla 3.1 los datos con mejor correlación no necesariamente son los que tienen una menor **Raíz del Error Cuadrático Medio (RECM)** ni **Error Absoluto Medio (EAM)** más favorable. Por ejemplo, el quinto modelo respecto a su grado de correlación, tiene un mejor **EAM** que los otros cuatro.

3.1.1. RECM y EAM para ΔG_{np}

La **RECM** es una medida de la precisión que sirve para comparar errores de predicción de diferentes modelos. Utiliza las diferencias entre los valores predichos por un modelo y los valores observados.

Los modelos jerarquizados por esta medida de precisión se encuentran en la tabla 3.2

Tabla 3.2: Los mejores modelos encontrados para la aproximación de ΔG_{np} jerarquizados según su RECM.

Modelo	ρ	RECM	%EAM
1° bf_cci-vert_AWPN-a_mix-radio_tolman_s_s-p1_ponderado_gnp-area_c_s-mag_c_s	0.3812	0.6908	60.4
2° bf_cci-vert_456-a_mix-radio_tolman_s_s-p1_ponderado_gnp-area_c_s-mag_c_s	0.381	0.6908	60.42
3° bf_esf-vert_AWPN-a_mix-radio_tolman_s_s-p1_ponderado_gnp-area_c_s-mag_c_s	0.3808	0.6909	60.39
4° bf_esf-vert_456-a_mix-radio_tolman_s_s-p1_ponderado_gnp-area_c_s-mag_c_s	0.3805	0.6909	60.41
5° bf_esf-vert_AWPN-a/3-radio_tolman_s_s-p1_ponderado_gnp-area_c_s-mag_c_s	0.3855	0.6933	61.07
89° SoloGamma_gnp	0.2159	0.7311	66.97

En donde el resultado es similar al visto en la tabla 3.1, solo que esta vez aparece el tipo de área `a_mix` en lugar de `a/3` como principal tipo de área utilizado por estos

modelos. El resultado `SoloGamma_gnp` ocupa el puesto 89°, por lo que existen otros 88 modelos que tienen una **RECM** más favorable gracias al análisis de la curvatura de su geometría.

El **EAM** es una medida de la diferencia entre dos variables continuas. Considerando dos series de datos (unos calculados y otros observados) relativos a un mismo fenómeno. Sirve para cuantificar la precisión de una técnica de predicción.

Los modelos jerarquizados por esta medida de precisión se encuentran en la tabla 3.3.

Tabla 3.3: Los mejores modelos encontrados para la aproximación de ΔG_{np} jerarquizados según su EAM.

Modelo	ρ	RECM	%EAM
1° bf_cci-vert_456-a/3-radio_tolman_s_s-p1_promedio_gnp-area_c_s-mag_c_s	0.3822	0.7661	54.52
2° bf_esf-vert_456-a/3-radio_tolman_s_s-p1_promedio_gnp-area_c_s-mag_c_s	0.3822	0.7667	54.52
3° bf_cci-vert_AWPN-a/3-radio_tolman_s_s-p1_promedio_gnp-area_c_s-mag_c_s	0.3821	0.7668	54.53
4° bf_esf-vert_AWPN-a/3-radio_tolman_s_s-p1_promedio_gnp-area_c_s-mag_c_s	0.3822	0.7673	54.54
5° bf_cci-vert_456-a_mix-radio_tolman_s_s-p1_promedio_gnp-area_c_s-mag_c_s	0.3742	0.7809	54.57
97° SoloGamma_gnp	0.2159	0.7311	66.97

En esta tabla 3.3 los resultados son similares a los vistos en el caso anterior (tabla 3.2). Con la diferencia de que el modelo empleado para la aproximación de la longitud de Tolman δ es en todos los casos `p1_promedio_gnp` en lugar de `p1_ponderado_gnp`. El resultado `SoloGamma_gnp` ocupa el puesto 97°, por lo que existen otros 96 modelos que tienen un **EAM** más favorable gracias al análisis de la curvatura de su geometría.

3.2. Energía requerida para la formación de una cavidad en el soluto

Tras haber conseguido un mejor modelo para la sección anterior, se nos sugirió comprobar si estos 1442 modos de aproximar ΔG_{np} servirían para aproximar ΔG_{cav} . La única diferencia es la comparación de las energías obtenidas, las cuales ahora hacen contraste de los valores nominales de ΔG_{cav} , los cuales se obtienen de las siguiente manera:

$$\Delta G_{np} = \Delta G_{cav} + \Delta G_{vdw} \quad (3.1)$$

Esta relación entre energías es complementaria a la figura 1.1. Las energías para ΔG_{vdw} nos las ha facilitado nuestro profesor Christopher Cooper, quien las obtuvo tras utilizar un modelo de integración de esta energía[14] en las 502 especies analizadas.

Los cinco mejores modelos obtenidos de las combinaciones de factores y su comportamiento respecto a valores nominales para ΔG_{cav} se encuentran en la figura 3.2.

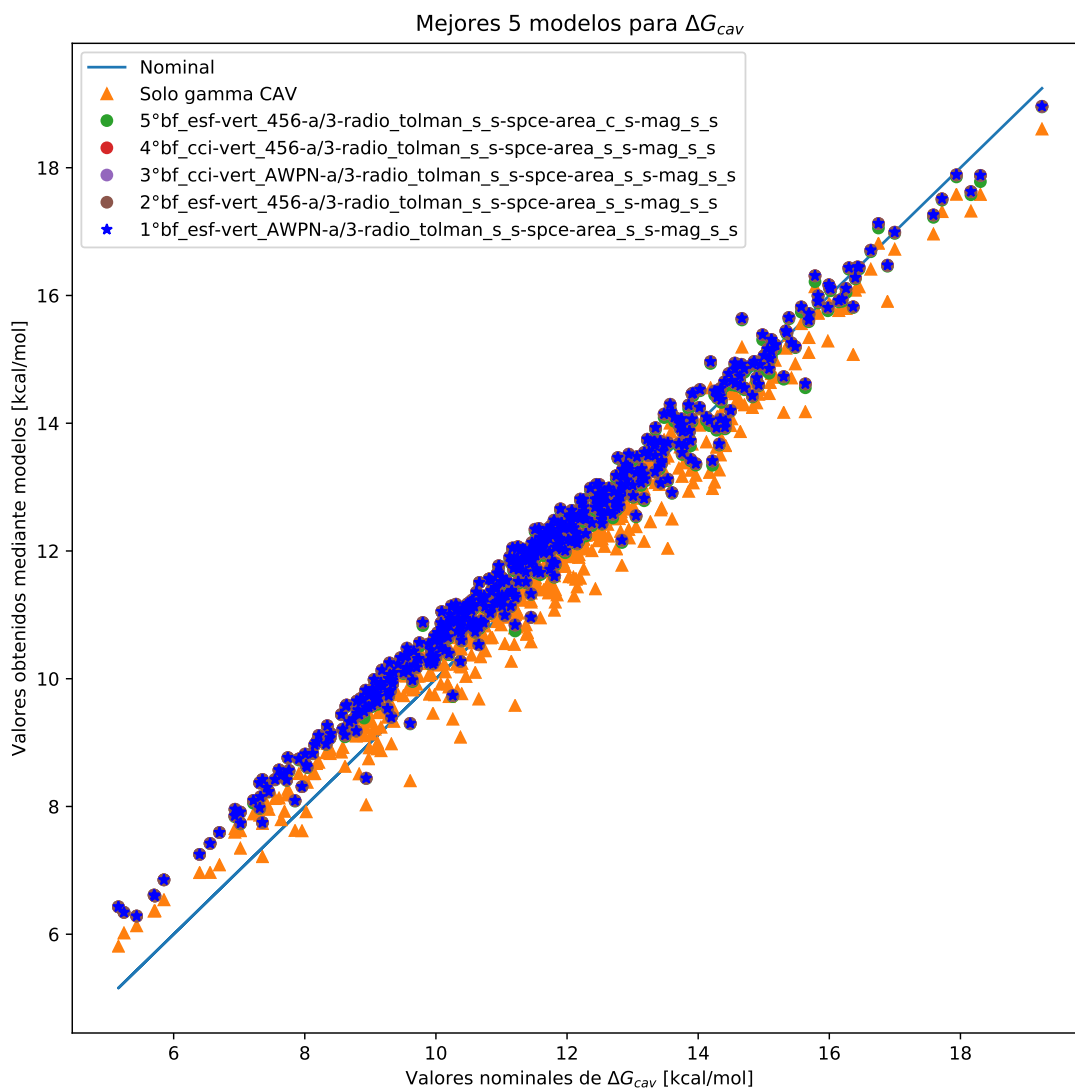


Figura 3.2: Cinco mejores modelos para ΔG_{cav} . A diferencia de los modelos para ΔG_{np} , estos se encuentran tan cercanos los unos a los otros que no es posible diferenciarlos en el gráfico.

En esta figura 3.2 se puede apreciar claramente una mejor correlación de los modelos al tipo de energía al que se está aproximando, en comparación a los datos vistos para

ΔG_{np} en la figura 3.1.

Estos modelos fueron jerarquizados según su [coeficiente de pearson](#). Se ha incluido en la figura 3.2 los resultados del modelo `SoloGamma_cav`, este modelo es la referencia a lo que sucede si se hace caso omiso al análisis de la superficie y su curvatura para ΔG_{cav} . Como se puede ver en la tabla 3.4, existen 543 modelos que tendrían una mejor correlación que `SoloGamma_cav`.

Tabla 3.4: Los mejores modelos encontrados para la aproximación de ΔG_{cav} jerarquizados según su ρ .

	Modelo	ρ	RECM	%EAM
1°	bf_esf-vert_AWPN-a/3-radio_tolman_s_s-spce-area_s_s-mag_s_s	0.9923	0.5201	4.37
2°	bf_esf-vert_456-a/3-radio_tolman_s_s-spce-area_s_s-mag_s_s	0.9923	0.5201	4.37
3°	bf_cci-vert_AWPN-a/3-radio_tolman_s_s-spce-area_s_s-mag_s_s	0.9923	0.5195	4.36
4°	bf_cci-vert_456-a/3-radio_tolman_s_s-spce-area_s_s-mag_s_s	0.9923	0.5195	4.36
5°	bf_esf-vert_456-a/3-radio_tolman_s_s-spce-area_c_s-mag_s_s	0.9922	0.5069	4.24
544°	SoloGamma_cav	0.9864	0.4333	3.12

Para los modelos de aproximación de ΔG_{cav} , los coeficientes de Pearson ρ son sumamente parecidos. Son cercanos a $\rho = 1$ lo cual indica una correlación positiva casi perfecta.

En la tabla 3.4, al igual que para el caso de ΔG_{np} los métodos de curvatura basados en el mejor ajuste de una esfera `bf_cci` y `bf_esf` dominan los primeros lugares frente a los modelos que utilizan la curvatura media `mc`. En cuanto a los tipos de normal, los métodos por vértice `vert_AWPN` y `vert_456` son superiores a los por panel `panel`. Para los tipos de área empleados ocurre lo mismo que al analizar la correlación del caso ΔG_{np} , los primeros lugares están dominados por el tipo de área `a/3`. Algo que si resulta completamente diferente para los modelos que mejor correlación tienen con ΔG_{cav} , es el modelo para la longitud de Tolman δ que domina los primeros lugares, el cual es `spce`.

En cuanto a la consideración de zonas cóncavas en las mallas, como se ve en la sección 2.6, los modelos de Tolman y curvatura en esta tabla muestran un mejor comportamiento cuando no es considerada negativamente su participación (`toman_s_s` y `mag_s_s`). Lo mismo ocurre para el área en los primeros cuatro lugares `area_s_s`, el quinto puesto lo tiene un modelo que si considera la concavidad para la aproximación.

3.2.1. RECM y EAM para ΔG_{cav}

Tabla 3.5: Los mejores modelos encontrados para la aproximación de ΔG_{cav} jerarquizados según su RECM.

Modelo	ρ	RECM	%EAM
1° bf_cci-vert_456-a_mix-radio_tolman_c_s-p1_mesh_ponderado_cav-area_s_s-mag_s_s	0.9899	0.398	3.08
2° bf_esf-vert_456-a_mix-radio_tolman_c_s-p1_mesh_ponderado_cav-area_s_s-mag_s_s	0.9899	0.3983	3.09
3° bf_cci-vert_AWPN-a_mix-radio_tolman_c_s-p1_mesh_ponderado_cav-area_s_s-mag_s_s	0.9899	0.3983	3.09
4° bf_esf-vert_AWPN-a_mix-radio_tolman_c_s-p1_mesh_ponderado_cav-area_s_s-mag_s_s	0.9899	0.3986	3.09
5° bf_esf-vert_456-a_mix-radio_tolman_c_s-p1_promedio_cav-area_s_s-mag_c_s	0.9905	0.3997	2.83
77° SoloGamma_cav	0.9864	0.4333	3.12

De la tabla 3.5 puede verse que los modelos con menor RECM siguen teniendo coeficientes de pearson ρ muy cercanos a la unidad, por lo que continúan estando relacionados. Nuevamente existe un dominio del análisis hecho mediante curvaturas extraídas del mejor ajuste de una esfera (bf_esf y bf_cci) y de la integración por vértice usando normales de los tipos vert_AWPN y vert_456. Para el área sin embargo, se presenta de manera análoga a lo sucedido para ΔG_{np} , existe dominio de los modelos que utilizan a_mix. Con respecto a la longitud de Tolman δ , los primeros cuatro de cinco modelos presentados utilizan p1_mesh_ponderado_cav y es el quinto el que utiliza p1_mesh_promedio_cav.

Tabla 3.6: Los mejores modelos encontrados para la aproximación de ΔG_{cav} jerarquizados según su EAM.

Modelo	ρ	RECM	%EAM
1° bf_esf-vert_456-a_mix-radio_tolman_c_s-p1_promedio_cav-area_s_s-mag_c_s	0.9905	0.3997	2.83
2° bf_esf-vert_AWPN-a_mix-radio_tolman_c_s-p1_promedio_cav-area_s_s-mag_c_s	0.9905	0.4001	2.84
3° bf_cci-vert_456-a_mix-radio_tolman_c_s-p1_promedio_cav-area_s_s-mag_c_s	0.9905	0.4003	2.84
4° bf_cci-vert_AWPN-a_mix-radio_tolman_c_s-p1_promedio_cav-area_s_s-mag_c_s	0.9905	0.4007	2.84
5° bf_esf-vert_456-a/3-radio_tolman_c_s-p1_promedio_cav-area_s_s-mag_c_s	0.9904	0.4083	2.94
59° SoloGamma_cav	0.9864	0.4333	3.12

Finalmente, de la tabla 3.6 se desprende de forma análoga que lo ocurrido para el caso de ΔG_{np} , que el EAM es favorecido de los modelos que emplean los comportamientos de la longitud de Tolman otorgados por el tipo p1_promedio_cav, el cual es similar a p1_promedio_gnp pero en base a ΔG_{cav} . El área del del tipo a/3 aparece en el quinto de estos cinco resultados, siendo dominante el tipo a_mix.

En cuanto a la concavidad de la malla, la jerarquización por EAM es la única para ΔG_{cav} que favorece la utilización de curvaturas negativas. Tanto para ρ como RECM en la consideración de zonas cóncavas de la malla no parecía favorable para los modelos.

4 | Análisis de los modelos con mayor correlación para ΔG_{np} y ΔG_{cav}

Como se vio en el capítulo anterior 3, la jerarquización de un mejor modelo, basándose en los resultados para ambas energías ΔG_{np} y ΔG_{cav} , dependerá de la medida de dispersión o correlación que se escoja para hacerlo.

Para este trabajo será de mayor peso la correlación de pearson ρ frente a la dispersión medida a través de sus *EAM* o *RECM*. Es por esto que de las tablas 3.1 y 3.4 los mejores resultados son:

ΔG_{np} : bf_cci-vert_AWPN-a/3-radio_tolman_s_s-pl_ponderado_gnp-area_c_s-mag_c_s

Este modelo tiene un coeficiente de Pearson $\rho = 0,3858$, una *RECM* = 0,6933 y un *EAM* = 61,08 %.

ΔG_{cav} : bf_esf-vert_AWPN-a/3-radio_tolman_s_s-spce-area_s_s-mag_s_s

Este modelo tiene un coeficiente de Pearson $\rho = 0,9923$, una *RECM* = 0,5201 y un *EAM* = 4,37 %.

Estos dos modelos son los que mejor correlación tienen con las energías nominales que aproximan respectivamente. En su estructuración se pueden ver factores comunes y factores que los diferencian, los cuales se exponen a continuación:

- **Tipo de curvatura:** Ambos modelos emplean curvaturas basadas en el mejor ajuste de una esfera al *anillo de puntos vecinos* (sección 2.5.2). Difieren en la categoría inferior del modelo de esta curvatura, ya que una es del tipo centro-vértice y la otra es del tipo centro-radio. De las tablas 3.1 y 3.4 se puede ver que los terceros lugares de estas corresponden a un modelo similar a los del primer lugar, con la única diferencia que está cambiada esta categoría inferior.
- **Tipo de normal:** El tipo de normal empleada por estos modelos es *AWPN*, cuyas propiedades son descritas en la sección 2.4.2.
- **Tipo de área:** El tipo de área *a/3* es empleado por ambos métodos.
- **Tipo de aproximación de la longitud de Tolman (regresiones):** Para el caso de ΔG_{np} el tipo de aproximación de la longitud de Tolman δ es el que emplea los

radios de los k-átomos que conforman la molécula y lo pondera por el área que estos formarían `p1_ponderado_gnp`. Mientras que en el caso de ΔG_{cav} , el tipo de aproximación de la longitud de Tolman δ es `spce`, la cual nace de el comportamiento de la solvatación de gases nobles en modelos de agua SPC/E vistos en la tabla 1.1.

- **Concavidad aplicada al área:** Para la concavidad aplicada al área los modelos son distintos, el caso de ΔG_{np} si considera que las zonas con curvaturas negativas deben ser tratadas como áreas negativas `area_c_s`, mientras que para el modelo representante de ΔG_{cav} las áreas son consideradas siempre positivas, sin importar la curvatura de la zona `area_s_s`.
- **Concavidad aplicada al radio de Tolman:** Ambos modelos consideran curvaturas para los modelos de aproximación de la longitud de Tolman δ siempre positivas `radio_tolman_s_s`, sin importar la concavidad de la zona que se esté integrando.
- **Concavidad aplicada a la curvatura:** Al igual que para el caso de la concavidad aplicada al área, en el caso del mejor modelo para ΔG_{np} si es considerada una curvatura negativa para zonas de la geometría cóncavas `mag_c_s`, mientras que para ΔG_{cav} sin importar la concavidad de la zona integrada se considerará una curvatura positiva `mag_s_s`.

5 | Conclusiones

El análisis geométrico de la SAS ha demostrado ser una herramienta útil para la aproximación de la componente no-polar de la energía de solvatación ΔG_{np} y de la energía de formación de cavidad de soluto ΔG_{cav} . En ambos casos se han logrado mejores aproximaciones a los valores nominales en comparación a solo considerar el área de estas interfases. Prueba de esto es la comparación hecha entre los cinco mejores modelos de cada una de las energías y sus respectivos modelos `SoloGamma_gnp` y `SoloGamma_cav`.

Se propusieron dos formas de integración de la geometría de las mallas de los solutos, por vértice y por panel. Esto se hizo con el propósito de aprovechar las características bien definidas en un panel, como lo es su área y su normal. Sin embargo resultaron mejores los modelos que empleaban una integración por vértice, en donde se encuentra mejor definida la curvatura, tanto por medio del mejor ajuste de una esfera como por medio de la curvatura media H .

La caracterización de la curvatura de la superficie de las mallas mediante un mejor ajuste de una esfera ha demostrado tener un mejor desempeño frente a hacerlo mediante la curvatura media H . Esto puede deberse principalmente a que las esferas son la geometría que equilibra naturalmente las burbujas líquidas o gaseosas y que estos tipos de interfases son dominadas por formas esféricas y simétricas [13].

Los modelos expuestos en este trabajo resultaron mejores para aproximar ΔG_{cav} que para ΔG_{np} . Esto puede deberse a la alta correlación ya existente entre la geometría de las mallas de la SES del soluto y la energía necesaria para la formación de la cavidad. Todos los modelos vistos en este trabajo dan distintas soluciones a la ecuación 2.3, la cual tiene dos términos, uno superficial y otro volumétrico. La correlación de estas características, área y volumen de las mallas, con los valores nominales de sus energías permite conocer la dificultad de aproximación que enfrentan estos modelos.

Como se puede ver en la figura 5.1 el área y el volumen de las mallas de los solutos no tienen buena correlación con la energía no-polar de solvatación ΔG_{np} . La correlación entre el área de las mallas y la ΔG_{np} es de $\rho = 0,2159$, mientras que para el caso del volumen es de $\rho = 0,2081$. Esto demuestra que la relación entre las variables area y volumen frente a la energía es remotamente directa.

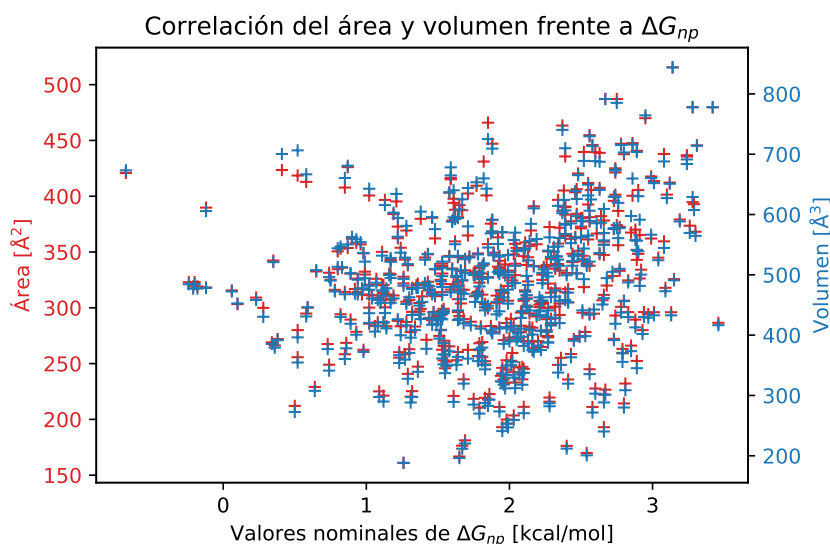


Figura 5.1: El área y el volumen de las mallas de los solutos frente a la energía ΔG_{np} tienen poca correlación.

En el caso de ΔG_{cav} , en la figura 5.2 se puede ver que tanto el volumen como el área de las mallas tiene una correlación casi perfecta ($\rho = 1$), para el área la correlación es de $\rho = 0,9864$ y para el volumen es de $\rho = 0,9860$ frente a la energía ΔG_{cav} .

Estas correlaciones también se pueden interpretar como una cota inferior para los modelos que trabajan estas variables, geometría y volumen. Estos son los puntos de partida para los modelos que buscan relacionar de mejor manera estas variables con las respectivas componentes de la energía de solvatación.

La principal diferencia entre un modelo encontrado para ΔG_{np} y ΔG_{cav} es la utilización de un modelo de aproximación de la longitud de Tolman δ . Para el caso de ΔG_{np} los mejores resultados emplearon modelos basados en regresiones hechas entre las longitudes de Tolman δ encontradas y los radios ponderados de sus superficies. En cambio en el caso de ΔG_{cav} los mejores modelos emplearon modelos de aproximación de la longitud de Tolman δ basados en la solvatación de gases nobles en modelos de agua computacionales. La solvatación de estos gases nobles en el solvente podrían ser una buena forma de generar modelos de aproximación más refinados para esta longitud que caracteriza la cavidad que generan. En un trabajo futuro de podría probar la eficiencia de un modelo de aproximación de la longitud de Tolman δ basándose en el comportamiento de los siete gases nobles en lugar de solo cuatro de ellos.

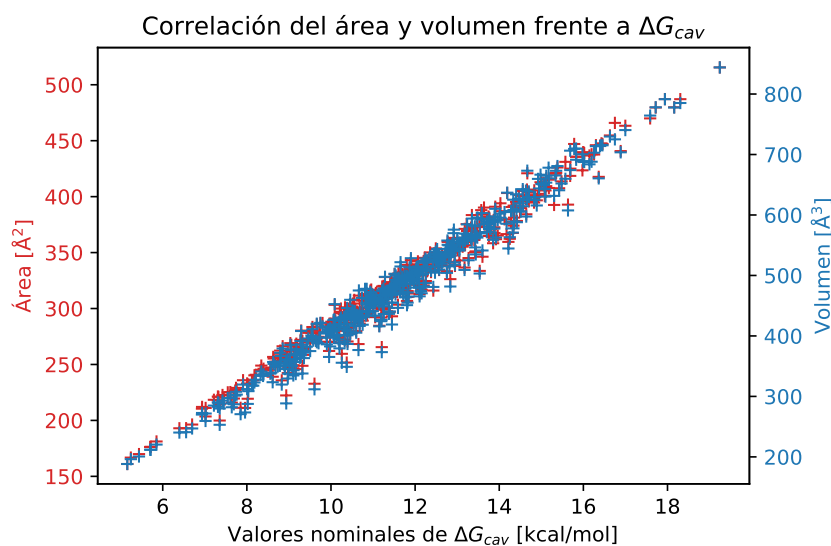


Figura 5.2: El área y el volumen de las mallas de los solutos frente a la energía ΔG_{cav} presentan una correlación casi perfecta.

Es la longitud de Tolman δ la que se encarga de gran parte del comportamiento de los modelos. Los valores nominales de las energías ΔG_{np} se encuentran en el rango $[-0,68; 3,46]$, mientras que para el caso de ΔG_{cav} el rango es $[5,16; 19,24]$, los modelos explorados utilizan las mismas variables de entrada y es este factor, la longitud de Tolman δ , el que se encarga de escalar apropiadamente la componente superficial de la ecuación 2.3 para generar esos cambios de magnitud en los resultados de un modelo.

Entre todas las mallas estudiadas solo el 4,92% de los vértices se encuentran en una zona cóncava de la geometría. Si bien es un pequeño porcentaje, los modelos fueron sensibles a esta consideración y se pudo establecer cuales son los factores afectados por esta característica.

Para los modelos de aproximación de la longitud de Tolman δ no es necesaria esta consideración en ninguno de los casos. Esto puede deberse al origen de los polinomios usados. Todos ellos fueron hechos a partir de regresiones respecto al radios, los cuales siempre estuvieron en un dominio positivo, como se puede ver en los gráficos de la sección 2.8.2.

Para el caso del área y la curvatura la consideración de la concavidad de la geometría depende del tipo de energía que se esté aproximando. Para el caso de ΔG_{np} si es

necesaria esta consideración, en cambio para ΔG_{cav} no.

Referencias

- [1] J. Dzubiella, J. M. Swanson, and J. A. McCammon, “Coupling nonpolar and polar solvation free energies in implicit solvent models,” *Journal of Chemical Physics*, vol. 124, no. 8, 2006.
- [2] C. Cooper and J. Bardhan, “A capacitor model for nonpolar solvation,” *Unpublished*, 2019.
- [3] F. H. Stillinger, “Structure in Aqueous Solutions of Nonpolar Solutes from the Standpoint of Scaled-Particle Theory,” *Journal of Solution Chemistry*, 1973.
- [4] K. Lum, D. Chandler, and J. D. Weeks, “Hydrophobicity at Small and Large Length Scales,” *Physical Chemistry B*, 1999.
- [5] 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.
- [6] D. Chandler, “Interfaces and the driving force of hydrophobic assembly,” 9 2005.
- [7] R. C. Tolman, “The Effect of Droplet Size on Surface Tension,” *The Journal of Chemical Physics*, 1949.
- [8] H. S. Green, S. Ono, S. Kondo, and F. P. Buff, *Structure of Liquids / Struktur der Flüssigkeiten*, vol. 3 / 10 of *Encyclopedia of Physics / Handbuch der Physik*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1960.
- [9] D. Paschek, “Temperature dependence of the hydrophobic hydration and interaction of simple solutes: An examination of five popular water models.,” *Journal of Chemical Physics*, 2004.
- [10] B. Kozlikova, 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*, p. 7, 11 2016.
- [11] D. L. Mobley, C. I. Bayly, M. D. Cooper, M. R. Shirts, and K. A. Dill, “Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations,” *Journal of Chemical Theory and Computation*, vol. 5, pp. 350–358, 2 2009.

- [12] J. A. Bærentzen, J. Gravesen, F. Anton, H. Aanæs, J. A. Bærentzen, J. Gravesen, F. Anton, and H. Aanæs, “Curvature in Triangle Meshes,” in *Guide to Computational Geometry Processing*, pp. 143–158, Springer London, 5 2012.
- [13] R. Bahadur and L. M. Russell, “Effect of Surface Tension from MD Simulations on Size-Dependent Deliquescence of NaCl Nanoparticles,” *Aerosol Science and Technology*, vol. 42, pp. 369–376, 3 2008.
- [14] M. Laudon, B. F. Romanowicz, Nano Science and Technology Institute., International Conference on Modeling and Simulation of Microsystems (8th : 2005), International Conference on Computational Nanoscience and Technology (5th : 2005), Workshop on Compact Modeling (2005), and NSTI Nanotech Ventures (2005), *2005 NSTI Nanotechnology Conference and Trade Show : NSTI Nanotech 2005 : Anaheim, May 8-12, 2005*. Nano Science and Technology Institute, 2005.

A | Mejor ajuste de una esfera

El mejor ajuste de una esfera para un grupo t ($t \geq 4$) de puntos, se logra mediante mínimos cuadrados.

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

En primer lugar se busca el centro de la esfera para luego dar con el radio de esta.

Para encontrar el centro de la esfera que más se aproxima a los puntos dados es necesario calcular el promedio de las coordenadas x , y , z para cada casco esférico.

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

Luego se procede a calcular las matrices A y B necesarias para la aproximación.

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

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

Operación para dar con las coordenadas del centro del casco:

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

Con las coordenadas del centro de la esfera, ahora es posible conocer el radio del casco, que estará dado por:

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

Una vez conocido el radio y el centro ya tenemos definida una esfera que hace el mejor ajuste a un grupo t ($t \geq 4$) de puntos.

B | Regresiones de la longitud de Tolman: ΔG_{cav}

De forma análoga a lo realizado para ΔG_{np} en la sección 2.8.2, se realizaron regresiones para ΔG_{cav} con la finalidad de ampliar la búsqueda de la relación de esta magnitud y la curvatura.

El valor de la tensión superficial promedio para las 502 especies en este caso resulta $\bar{\gamma}_i = 25,06E-3[N/m]$, el valor promedio de la longitud de Tolman es de $\bar{\delta}_i = 1,00[\text{\AA}]$.

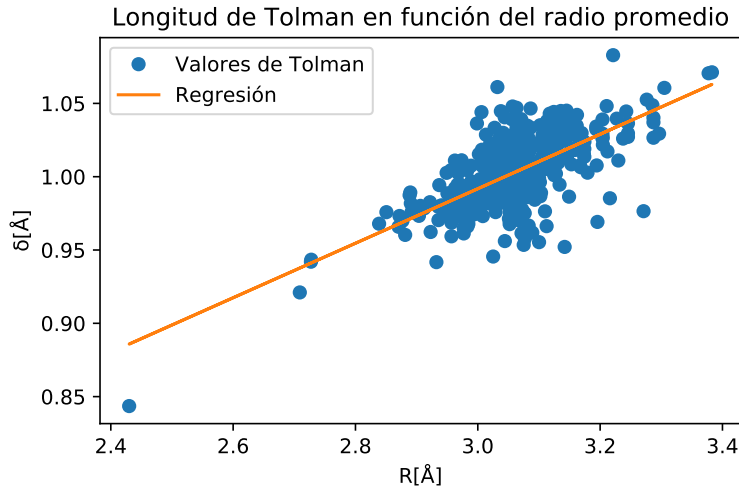


Figura B.1: La regresión ($\rho = 0,6476$) de los datos aquí mostrados para la aproximación de δ utilizando un radio promedio de la malla se encuentra en la ecuación B.1

$$\delta_{\bar{R}_i}(R) = 0,1859R + 0,4341 \quad (\text{B.1})$$

$$\delta_{\bar{R}_{si}}(R) = 0,1859R + 0,4341 \quad (\text{B.2})$$

$$\delta_{\bar{R}_{in}}(R) = 0,2949R + 0,0924 \quad (\text{B.3})$$

$$\delta_{\bar{R}_{ai}}(R) = 0,2444R + 0,2484 \quad (\text{B.4})$$

Longitud de Tolman en función del radio ponderado por superficie

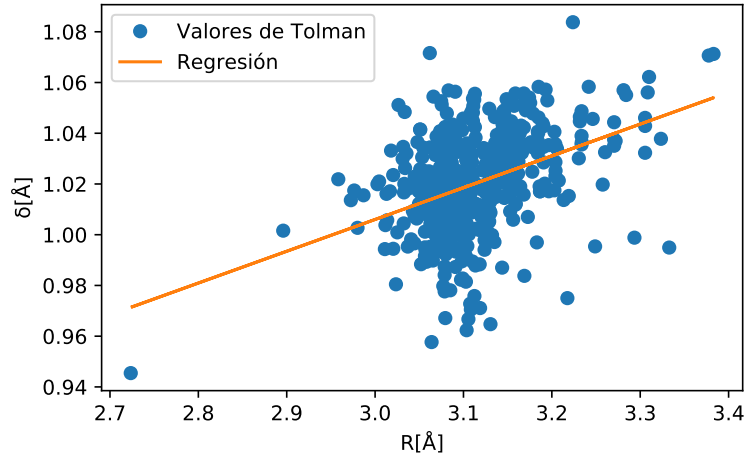


Figura B.2: La regresión ($\rho = 0,4060$) de los datos aquí mostrados para la aproximación de δ utilizando un radio ponderado por la superficie se encuentra en la ecuación B.2.

Longitud de Tolman en función del radio promedio de las mallas

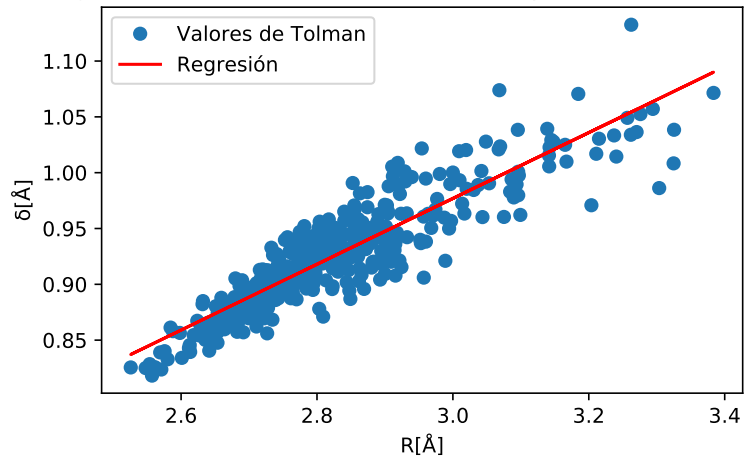


Figura B.3: La regresión ($\rho = 0,9013$) de los datos aquí mostrados para la aproximación de δ utilizando un radio promedio de la malla se encuentra en la ecuación B.3.

Longitud de Tolman en función del radio ponderado de las mallas

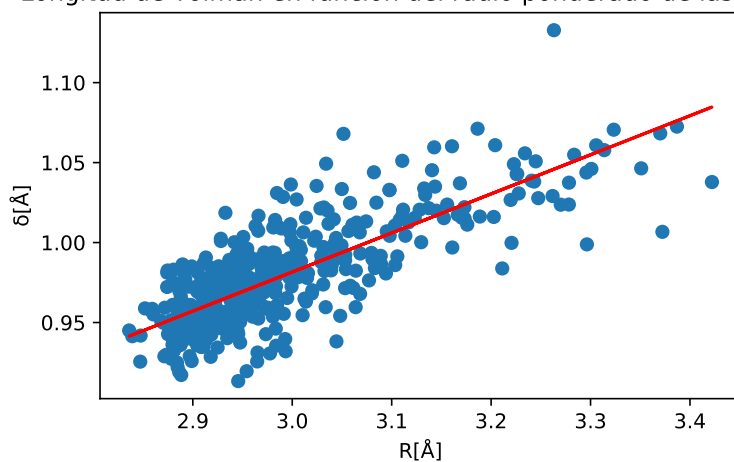


Figura B.4: La regresión ($\rho = 0,7920$) de los datos aquí mostrados para la aproximación de δ utilizando un radio ponderado por la el área relacionado a los vértices de la malla se encuentra en la ecuación B.4.