UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE INGENIERÍA MECÁNICA.

Memoria de Titulación para optar al Titulo de Ingeniero Civil Mecánico

Influencia de la linealización de la ecuación de Poisson-Boltzmann para la estimación del potencial electroestático y la energía de solvatación.

Byron Alexander Vega Lagos

Profesor Guía: Ph.D. Christopher Cooper V.

Profesor Correferente: Ph.D. Harvey Zambrano R.

Septiembre, 2023

Agradecimientos

Quisiera agradecerle a todos los nombres, todos los rostros y todos los matices que estuvieron en mi recorrido universitario, principalmente todo mi cariño a mi familia que siempre los amaré sean cuales sean las tormentas que se avecinen, también los amigos que me han acompañado, y las instancias que he vivido, porque la vida cambia increíblemente deprisa, pero como decía un escritor que me gusta mucho, la soledad en si misma puede ser peligrosa. Finalmente un profundo agradecimiento a mi profesor guía, Christopher Cooper, a su paciencia y apoyo inconmensurable que me permitieron llegar al final de esta carretera.

Resumen

La energía de solvatación y el potencial electrostático, son medidas que nos permite entender las interacciones moleculares de una proteína y un medio. Dada la importancia que toma en investigación se requiere ser lo suficientemente minucioso con la forma y las suposiciones que se tomen al momento de entregar resultados.

A lo largo de esta memoria de investigación hemos realizado una serie de ensayos sobre distintas moléculas con variadas características, todo con el objetivo de responder cuál es el efecto sobre el potencial tras linealizar la ecuación de Poisson Boltzmann. En consideración de aquello se realizó un minucioso análisis de sensibilidad para cada una de las moléculas presentadas en el estudio, donde se definió que una distancia media de 20 [A] permite asegurar las condiciones de Dirichlet en el contorno.

Las moléculas analizadas las separamos en 3 baterías inicialmente, diferenciadas por su nivel de carga. La primera de ellas, denominada altamente cargada, contiene moléculas como la 9ANT, 280D, 437D y 1KX5 de donde concluimos cosas como que el comportamiento del error debido a la diferencia de potencial calculado de la linealización y la no linealización, se acentúa en el interfaz y en zonas relativamente cóncavas a comparación de la proteína, es decir, con cierto grado de cobertura por parte de la molécula. Otro factor de interés es que el error se acentúa mucho más en este tipo de batería de moléculas, pues al compararla con moléculas de menor carga o con ciertas propiedades hidrófobas, es decir que en su composición molecular existan cadenas de lípidos, estos presentan un orden de magnitud mucho mayor en el error, tomando importancia para la investigación el refinado de la malla.

Keywords – Poisson-Boltzmann, Diferencias Finitas, Potencial Electroestático

Abstract

The solvation energy and the electrostatic potential are measurements that allow us to understand the molecular interactions of a protein and a medium. Given their importance in research, it is necessary to be sufficiently thorough with the form and assumptions taken at the time of delivering results.

Throughout this research report we have performed a series of tests on different molecules with various characteristics, all with the aim of answering what is the effect on the potential after linearizing the Poisson Boltzmann equation. In consideration of this, a thorough sensitivity analysis was carried out for each of the molecules presented in the study, where it was defined that an average distance of 20[A] ensures the Dirichlet conditions in the boundary.

The analyzed molecules were initially separated into 3 batteries, differentiated by their charge level. The first one, called highly charged, contains molecules such as 9ANT, 280D, 437D and 1KX5 from which we conclude several separate notes. The error behavior due to the calculated potential difference of linearization and nonlinearization is accentuated at the interface and in relatively concave areas compared to the protein, i.e., with a certain degree of coverage by the molecule. Another factor of interest is that the error is much more accentuated in this type of molecule battery, because when compared with molecules of lower charge or with certain hydrophobic properties, i.e., with lipid chains in their molecular composition, these present a much higher order of magnitude in the error, making the refinement of the mesh much more important for the research.

Keywords – Poisson-Boltzmann, Finite Differences, Electrostatic Potential

Índice general

Agradecimientos I

Resumen

Abstract

1. Introducción 2
1.1. Objetivo General . . . 3
1.2. Objetivos Específicos . . . 3

2. Marco Teórico 5
2.1. Electroestática y Solvatación . . . 5
2.2. Modelo Solvente Implícito . . . 6
2.3. Ecuación de Poisson-Boltzmann . . . 6
2.4. Energía libre de Solvatación . . . 10

3. Metodología 14
3.1. Protein Data Bank . . . 14
3.1.1. Batería de alto nivel energético. . . . 15
3.1.1.1. 1KX5 . . . 15
3.1.1.2. 280D . . . 15
3.1.1.3. 9ANT . . . 16
3.1.1.4. 437D . . . 17
3.1.2. Batería de proteínas medianamente cargadas . . . 18
3.1.2.1. 1LYZ . . . 18
3.1.2.2. 1UBQ . . . 18
3.1.2.3. 1PGB . . . 19
3.1.3. Batería de compuestos orgánicos, baja carga . . . 19
3.1.3.1. Anilina . . . 19
3.1.3.2. Tolueno . . . 20
3.1.3.3. Etanol . . . 20
3.1.3.4. Methyl Octanoate . . . 21

3.2. Herramientas y softwares utilizados 21
3.2.1. PDB2PQR 21
3.2.2. APBS (Adaptive Poisson-Boltzmann Solver) 22
3.2.2.1. Keywords - APBS 23
3.2.3. H++ 26
3.2.4. MDAnalysis 26
3.2.5. Otras bibliotecas de Python Utilizadas 27
3.3. Análisis de sensibilidad 27
3.3.1. Tamaño de caja 27
3.3.2. Nivel de refinamiento 28
3.4. Potencial electroestático 29
3.5. Generación de Cortes 31
3.6. Nivel de cobertura 33
3.7. Recolección de data 35

4. Resultados y Discusión **36**
4.1. Batería de alto nivel energético 36
4.1.1. 1KX5 37
4.1.2. 280D 40
4.1.3. 9ANT 43
4.1.3.1. Sin proteína anexada 43
4.1.3.2. Con proteína anexada 46
4.1.4. 437D 47
4.1.5. Análisis de curvas características del comportamiento de 100.000 datos 50
4.1.6. Modelo del Error lineal no lineal en función del potencial total lineal 53
4.2. Batería medianamente cargadas 55
4.2.1. 1LYZ 55
4.2.2. 1UBQ 58
4.2.3. 1PGB 61
4.2.4. Análisis del grado de cobertura 64
4.3. Batería de baja carga 66
4.3.1. Anilina 66
4.3.2. Etanol 68
4.3.3. Tolueno 70
4.3.4. Metil Octanoato 72

5. Conclusión **74**

Referencias **76**

Apéndices **78**
A1. Registro de cortes - 1KX5 78
A1.1. Potencial total lineal - 1KX5 78

A1.2. Potencial de reacción lineal - 1KX5 79
A1.3. Error lineal no lineal - 1KX5 80
A2. Registro de cortes - 9ANT 81
A2.1. Potencial total lineal - 9ANT 81
A2.2. Potencial de reacción lineal - 9ANT 82
A2.3. Error lineal no lineal - 9ANT 83
A3. Registro de cortes - 280D 84
A3.1. Potencial total lineal - 280D 84
A3.2. Potencial de reacción lineal - 280D 85
A3.3. Error lineal no lineal - 280D 86
A4. Registro de cortes - 437D 87
A4.1. Potencial total lineal - 437D 87
A4.2. Potencial de reacción lineal - 437D 88
A4.3. Error lineal no lineal - 437D 89
A5. Registro de cortes - 1LYZ 90
A5.1. Potencial total lineal - 1LYZ 90
A5.2. Potencial de reacción lineal - 1LYZ 91
A5.3. Error lineal no lineal - 1LYZ 92
A6. Registro de cortes - 1UBQ 93
A6.1. Potencial total lineal - 1UBQ 93
A6.2. Potencial de reacción lineal - 1UBQ 94
A6.3. Error lineal no lineal - 1UBQ 95

Índice de figuras

2.1. Modelo de Poisson-Boltzmann (PB): Dominios; Ω1(molécula) y Ω2(solvente) con diferente dieléctricos constantes $\varepsilon 1$ y $\varepsilon 2$ respectivamente. Weihua Geng - A boundary integral Poisson-Boltzmann solvers package for solvated bimolecular simulations - Gruyter Open, pag 55. 7
2.2. Definiciones de superficie utilizadas para evaluar los coeficientes de PBE. (A) Representación de la superficie molecular para el coeficiente dieléctrico e(x). (B) Representación de la superficie de Van Der Waals. [1] 9
2.3. Discretización de la función del dieléctrico para diferencias finitas en un dominio 2D [2] 12

3.1. Representación de la molécula 1KX5 realizada con MDAnalysis 15
3.2. Representación de la molécula 280D realizada con MDAnalysis 16
3.3. Representación de la molécula 9ANT realizada con MDAnalysis 16
3.4. Representación de la molécula 437D realizada con MDAnalysis 17
3.5. Representación de la molécula 1LYZ realizada con MDAnalysis 18
3.6. Representación de la molécula 1UBQ realizada con MDAnalysis 19
3.7. Representación de la molécula 1PGB realizada con MDAnalysis 19
3.8. Estructura de la Anilina 20
3.9. Estructura del Tolueno 20
3.10. Estructura del Ethanol 21
3.11. Estructura del Methyl Octanoate 21
3.12. Línea de comandos en un archivo *.in* requerido por APBS para la molécula de ADN *9ANT* con una malla de 285x285x285 y una distancia entre puntos de 0.45 en toda dirección. 25
3.13. Logo del software H++ 26
3.14. Ejemplo de análisis de sensibilidad para el tamaño de caja 28
3.15. Ejemplo de análisis de sensibilidad para determinar el número de nodos por eje. 29
3.16. Cuadro de comandos en un archivo .in requerido por APBS, para el cálculo de potencial en el vacío. 30
3.17. Cuadro de comandos en un archivo .in requerido por APBS, para el cálculo de potencial total no linealizado. 30

3.18. Selección de máximos y mínimos locales por medio de particiones. 31
3.19. Representación del modelo de cobertura por medio de haces de luz. 33

4.1. Análisis de sensibilidad para la molécula 1KX5 37
4.2. Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula 1KX5 38
4.3. Error lineal no lineal, Comportamiento de 100.000 datos 39
4.4. Análisis de sensibilidad para la molécula 280D 40
4.5. Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula 280D 41
4.6. Error lineal no lineal, Comportamiento de 100.000 datos 42
4.7. Análisis de sensibilidad para la molécula 9ANT sin proteína anexada . . . 43
4.8. Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula 9ANT sin proteína anexada 44
4.9. Error lineal no lineal, Comportamiento de 100.000 datos, 9ANT sin proteína anexada 45
4.10. Error lineal no lineal, Comportamiento de 100.000 datos, 9ANT con proteína anexada 46
4.11. Análisis de sensibilidad para la molécula 437D 47
4.12. Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula 437D 48
4.13. Error lineal no lineal, Comportamiento de 100.000 datos, 437D 49
4.14. Valor absoluto del Error lineal - no lineal versus el potencial total lineal. . 50
4.15. Valor absoluto del Error lineal - no lineal versus el potencial de reacción lineal. 51
4.16. Valor absoluto del Error lineal - no lineal versus el potencial de Coulomb . 52
4.17. Figura superior, Corte del caso lineal no lineal calculado de la diferencia del Potencial total lineal menos el potencial total no lineal por APBS. Figura inferior, Modelo del error lineal no lineal dependiente del Potencial total lineal calculado por APBS. 1AJF 53
4.18. Figura superior, Corte del caso lineal no lineal calculado de la diferencia del Potencial total lineal menos el potencial total no lineal por APBS. Figura inferior, Modelo del error lineal no lineal dependiente del Potencial total lineal calculado por APBS. 1A4D 54
4.19. Análisis de sensibilidad para la molécula 1LYZ 55
4.20. Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula 1LYZ 56
4.21. Error lineal no lineal, Comportamiento de 100.000 datos, 1LYZ 57
4.22. Análisis de sensibilidad para la molécula 1UBQ 58
4.23. Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula 1UBQ 59
4.24. Error lineal no lineal, Comportamiento de 100.000 datos, 1UBQ 60
4.25. Análisis de sensibilidad para la molécula 1PGB 61

4.26. Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula 1PGB 62
4.27. Error lineal no lineal, Comportamiento de 100.000 datos, 1PGB 63
4.28. Análisis de cobertura para la molécula 1UBQ 64
4.29. Análisis de cobertura para la molécula 9ANT 65
4.30. Análisis de sensibilidad para la molécula de Anilina 66
4.31. Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula de Anilina 67
4.32. Análisis de sensibilidad para la molécula de Etanol 68
4.33. Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula de Etanol 69
4.34. Análisis de sensibilidad para la molécula de Tolueno 70
4.35. Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula de Tolueno 71
4.36. Análisis de sensibilidad para la molécula de Metil Octanoato 72
4.37. Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula de Metil Octanoato 73

A1.1.Potencial total lineal - 1KX5 78
A1.2.Potencial de reacción lineal - 1KX5 79
A1.3.Error lineal no lineal - 1KX5 80
A2.1.Descripción de ambas imágenes 81
A2.2.Potencial de reacción lineal - 9ANT 82
A2.3.Error lineal no lineal - 9ANT 83
A3.1.Potencial total lineal - 280D 84
A3.2.Potencial de reacción lineal - 280D 85
A3.3.Error lineal no lineal - 280D 86
A4.1.Registro de cortes - 437D 87
A4.2.Potencial de reacción lineal - 437D 88
A4.3.Error lineal no lineal - 437D 89
A5.1.Potencial total lineal - 1LYZ 90
A5.2.Potencial de reacción lineal - 1LYZ 91
A5.3.Error lineal no lineal - 1LYZ 92
A6.1.Potencial total lineal - 1UBQ 93
A6.2.Potencial de reacción lineal - 1UBQ 94
A6.3.Error lineal no lineal - 1UBQ 95

Capítulo 1

Introducción

La solvatación y la electrostática son dos factores críticos que influyen en la estructura y función de las biomoléculas. Estas interactúan en un nivel molecular para determinar la estabilidad de una estructura, así como para regular las interacciones moleculares en procesos biológicos, por ejemplo, el plegamiento de proteínas. Antiguamente, el entendimiento de estos fenómenos se basaba en experimentos que ofrecían una visión limitada debido a la complejidad y las escalas en las que se veían involucrados los análisis. Es por ello que la entrada a la era de la tecnología y la simulación computacional emergieron para complementar los métodos experimentales y las teorías. En este contexto, la modelación electrostática en biomoléculas se hace fundamental para entender varios procesos biológicos y biofísicos hoy en día.

Aunque existen diversas metodologías, como la dinámica molecular y los modelos continuos, cada uno tiene sus propias ventajas y limitaciones. En particular, los modelos continuos como la ecuación de Poisson-Boltzmann (PB) ofrecen una alternativa eficiente computacionalmente, a diferencia de modelos de dinámica molecular, que es costoso pero nos ofrece un mejor nivel de precisión. Y si bien, la ecuación de PB es útil, las soluciones analíticas generalmente sólo están disponibles para geometrías simples, lo que conlleva varias suposiciones que han simplificado el problema limitándole la precisión y aplicabilidad en formas complejas.

Es ideal buscar entender cómo estas simplificaciones y soluciones numéricas de la ecuación de PB afectan la precisión de las estimaciones, buscar una metodología que permita obviar

de la mejor manera posible el efecto de usar métodos numéricos, como elementos finitos o diferencias finitas, en el sentido de la geometría y las condiciones de contorno utilizadas. Es por ello que la utilización de Software como APBS o MDAnalysis requieren pasar por un análisis previo para entender qué tan sensibles son sus cálculos en relación a las condiciones que se utilicen, pues solo bajo parámetros adecuados, podemos obviar el efecto geométrico y concentrarnos en el efecto de una suposición, en particular la linealización de la ecuación de PB. Generar un conjunto de directrices o parámetros bajo los cuales las simplificaciones y el métodos numérico adoptado produce resultados confiables.

En términos de contenido, podemos dividir el siguiente documento en una sección introductoria, el marco teórico; que abarca en su medida la Ecuación de Poisson-Boltzmann y el cálculo de la energía libre de solvatación. Un capítulo de la metodología, en donde se presentan las baterías de moléculas, los softwares y bibliotecas de Python utilizadas, y cómo utilizarlas, cómo se desarrolló el análisis de sensibilidad, los códigos implementados, la teoría de cobertura y la recolección de información. Luego dando paso a una sección de resultados y análisis, donde por medio de la metodología se explican los resultados presentados, los análisis secundarios realizados y conclusiones individuales que fueron tomadas. Para así finalizar con las conclusiones, que vienen a resumir los hallazgos más importantes encontrados en los análisis y resultados.

1.1. Objetivo General

El objetivo general de este trabajo es encontrar condiciones o comportamientos que nos permitan estimar los cálculos energéticos y potenciales de manera robusta, donde se hace estudio de una batería de moléculas de variadas características.

1.2. Objetivos Específicos

Los objetivos específicos contemplados son

- Estudiar el estado del arte de la determinación del potencial electrostático y la energía de solvatación.
- Implementar cálculo de potencial electrostáticas en software disponibles como APBS.

- Implementar métodos para el análisis del potencial electrostático usando librerías como MDAnalysis.
- Estudiar el efecto de la linealización sobre la ecuación de Poisson-Boltzmann en moléculas de alta y baja carga.
- Implementar metodologías de análisis frente a la información recogida para cada grupo de moléculas estudiadas, buscando relaciones empíricas en el comportamiento de sus datos.

Capítulo 2

Marco Teórico

2.1. Electroestática y Solvatación

La solvatación como tal, es un proceso químico de interacción entre una sustancia y su entorno, que implica la formación de enlaces entre las moléculas del solvente y los iones o moléculas que se están solvatando. Esto nos permite en general entender cómo las moléculas interactúan y cómo se estabilizan en soluciones acuosas.

Desde un punto de vista básico cuando introducimos una sustancia (soluto) en agua (solvente), las moléculas de agua interactúan con el soluto a través de enlaces de hidrógeno y por medio de fuerzas electroestáticas. Por ejemplo, el caso de un soluto iónico, como un ion sodio (Na+) y un ion cloruro (Cl-) disueltos en agua, las moléculas de agua rodean los iones, con los extremos positivos orientados hacia los iones de cloruro (CL-) y los extremos negativos orientados hacia el sodio (NA+), lo que estabiliza la solución. Para otro tipo de molécula también se pueden generan enlaces de hidrógeno, fuerza dipolo-dipolo, lo que altera las propiedades tanto físicas como químicas del soluto.

Y es que para explicar el proceso de solvatación es necesario hacer el nexo con la electroestática, pues lo que interactúa, como se mencionaba anteriormente, son las cargas eléctricas de las moléculas o iones del solvente con las especies de la solución. El general la documentación nos habla de que el fenómeno nos permite comprender como la fuerza electroestática y la solvatación afectan a la estructura, las propiedades de transporte, la estabilidad y la interacción molecular [3].

2.2. Modelo Solvente Implícito

Los métodos computacionales para estudiar interacciones electrostáticas los podemos dividir en dos categorías: los métodos de solvente explícito y los de solvente implícito. El método de solvente explícito considera a cada molécula del solvente actuando de manera individual, es un método mucho más preciso pero conlleva grandes costos computacionales. Por otro lado, el método de solvente implícito trata al solvente como un medio continuo, lo que simplifica las consideraciones de un código, pues se supone que las moléculas del solvente (agua normalmente) se encuentran en un equilibrio de cargas permitiendo que las moléculas de interés no pueda distinguir las contribuciones individuales del solvente, sino más bien, un efecto generalizado del medio continuo.

2.3. Ecuación de Poisson-Boltzmann

La ecuación de Poisson describe el potencial eléctrico en un medio sin considerar la presencia de cargas móviles. Esta ecuación establece una relación directa entre el laplaciano del potencial y la densidad de carga en el sistema.

$$\nabla^2\phi(x) = \frac{\rho(x)}{\varepsilon} \quad \text{para } x \in \Omega \quad \text{donde } \phi(x) = g(x) \quad \text{para } x \in \partial\Omega \tag{2.1}$$

Sin embargo, en sistemas con cargas móviles, o sea soluciones acuosas con iones disueltos, es necesario considerar las interacciones térmicas y las distribuciones de carga asociadas. Es en este contexto que surge el enfoque clásico para tratar estas interacciones electroestáticas, que se basa en la ecuación de Poisson-Boltzmann (PBE). Derivada bajo la suposición de que el solvente que rodea la molécula es un continuo:

$$\nabla \cdot \varepsilon(x)\nabla\phi(x) = \rho(x) \quad \text{para } x \in \Omega \quad \text{donde } \phi(x) = g(x) \quad \text{para } x \in \partial\Omega \tag{2.2}$$

Donde $\phi(x)$ corresponde al potencial electroestático en unidades de kT/q (donde k es la constante de Boltzmann, T es la temperatura absoluta y q es la carga del protón), $\rho(x)$ es la distribución de la carga y $\varepsilon(x)$ es la constante dieléctrica. Donde en general se resuelve usando condiciones de borde de Dirichlet , considerando un dominio finito Ω y fijando un

potencial de $g(x)$ en el borde $\delta\Omega$.

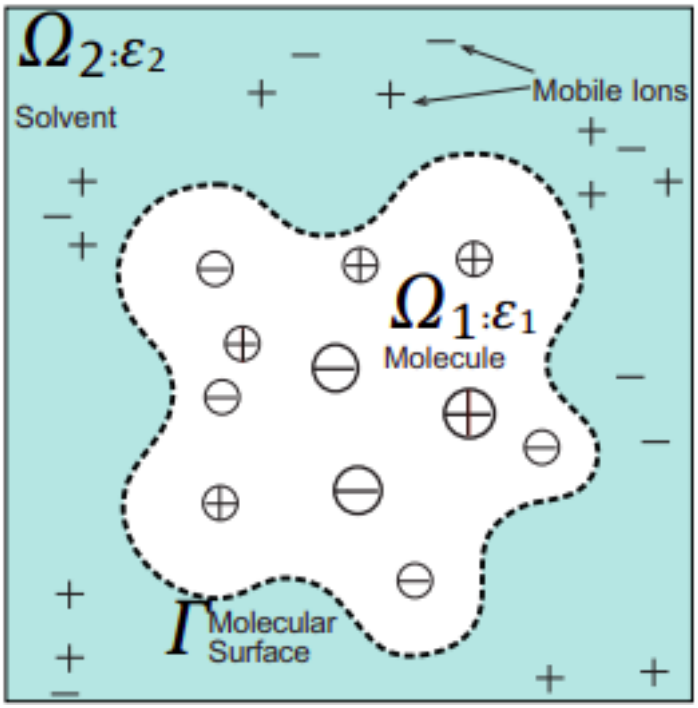

Figura 2.1: Modelo de Poisson-Boltzmann (PB): Dominios; Ω1(molécula) y Ω2(solvente) con diferente dieléctricos constantes $\varepsilon 1$ y $\varepsilon 2$ respectivamente. Weihua Geng - A boundary integral Poisson-Boltzmann solvers package for solvated bimolecular simulations - Gruyter Open, pag 55.

Sin embargo la función de distribución de carga se modela de forma distinta para cargas fijas y móviles; las fijas $\rho_f(x)$ modelan las M cargas parciales de la biomolécula como funciones de posición $\delta(x - x_i)$ ubicadas en el centro de cada átomo x_i con una magnitud Q_i. Con coeficientes adimensionales como e_c, la carga del electrón, y kT, la energía térmica del sistema.

$$\rho_{\mathrm{f}}(x) = \frac{4\pi e_{\mathrm{c}}^2}{kT} \sum_{i=1}^{M} Q_i \delta\left(x - x_i\right) \tag{2.3}$$

Las cargas móviles $\rho_m(x)$ por su parte, modelan la contribución de carga de las m especies de contra-iones de carga q_j, con concentraciones c_j y potencial estérico V_j (potencial que evita el solapamiento biomolécula-contraion), en el solvente mediante la distribución de Boltzmann.

$$\rho_{\mathrm{m}}(x) = \frac{4\pi e_{\mathrm{c}}^2}{kT} \sum_{j}^{m} c_j q_j \exp\left[-q_j \phi(x) - V_j(x)\right] \tag{2.4}$$

En el caso particular de soluciones salinas simétricas monovalentes como el NaCl, se reduce a la Eq. 2.5

$$\rho_{\mathrm{m}}(x) = \bar{\kappa}^2(x) \sinh \phi(x) \tag{2.5}$$

Donde $\bar{\kappa}^2(x)$ representa la función de accesibilidad de iones, la cual nos indica la existencia de alguna carga fija en una zona. Al reemplazar el término de la densidad de carga $\rho(x)$ de la Eq. 2.2 con las expresiones de distribución de carga fija Eq. 2.3 y cargas móviles Eq. 2.5 tenemos la siguiente expresión.

$$-\nabla \cdot \varepsilon(x) \nabla \phi(x) + \bar{\kappa}^2(x) \sinh((\phi(x)) = \frac{4\pi e_c^2}{kT} \sum q_i \delta\left(x - x_i\right) \tag{2.6}$$

Donde se observa que si $\bar{\kappa}(x) = 0$ la Eq. 2.6 se reduce a la ecuación de Poisson que a la vez se reduce a la ley de Coulomb cuando el dieléctrico es constante en todo el espacio. Sin embargo el motivo por el cual el agua se polariza mucho más fácilmente por un campo eléctrico que el resto de los solutos, es porque al menos se requieren dos $\varepsilon' s$ para capturar la física de las moléculas polares. De hecho el efecto de tener dos constantes dieléctricas expresadas en la Eq. 2.6 es porque la derivada de $\varepsilon(x)$ en el primer término no es cero en el interfaz, es decir cuando $\varepsilon(x)$ varía. Y para efecto del trabajo que se lleva a cabo, esta región corresponde a la superficie molecular, donde existe una discontinuidad dieléctrica entre el soluto de bajo dieléctrico y el disolvente de alto dieléctrico.

Se puede obtener una comprensión intuitiva del efecto de la forma de la interfase soluto-solvente a partir del concepto de carga superficial inducida, donde se induce una carga de polaridad opuesta a la de $\phi(x)$ en la superficie molecular. Un ejemplo es el caso simple de una carga positiva en el centro de una esfera de bajo dieléctrico, que induce una carga negativa distribuida sobre la superficie de la esfera. El potencial positivo en el solvente causado por la carga positiva fija será apantallado por la carga superficial negativa; cuanto mayor sea el valor del dieléctrico $\varepsilon(x)$ del disolvente, mayor será la carga superficial y, por lo tanto, mayor será el apantallamiento.

De manera similar, la interacción favorable de la carga positiva fija con su carga superficial

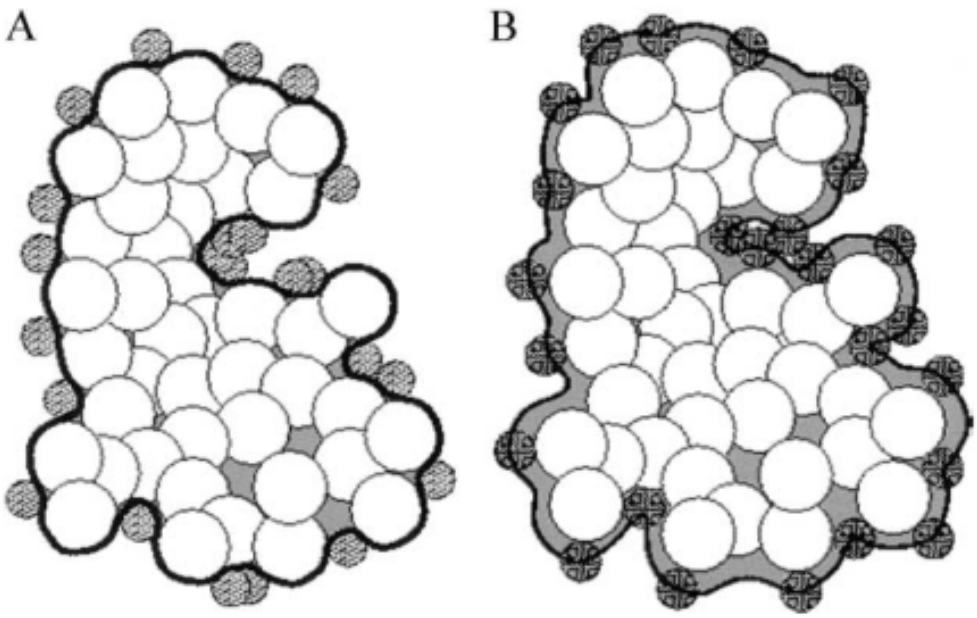

Figura 2.2: Definiciones de superficie utilizadas para evaluar los coeficientes de PBE. (A) Representación de la superficie molecular para el coeficiente dieléctrico e(x). (B) Representación de la superficie de Van Der Waals. [1]

negativa inducida (que produce un *campo de reacción*) es la base de los modelos de solución continua. Es tentador pensar que la carga superficial inducida corresponde a una alineación de dipolos de agua en la superficie molecular que produce un exceso de carga local y, de hecho, esto es parte del efecto. Sin embargo, debe darse cuenta de que la descripción de la carga superficial explica la polarización de todo el solvente.

Entre los modelos de coeficientes dieléctricos, los más populares han sido las definiciones basadas en las superficies moleculares y de Lee-Richards. En estos modelos, se considera que $\varepsilon(x)$ es discontinua a lo largo de la superficie biomolecular y asume valores dieléctricos de soluto en el interior y valores de solvente fuera de la superficie. Mientras que la función de accesibilidad de iones k(x) ha sido dimensionada en términos de la superficie de Van der Waals, Fig. 2.2B.

En condiciones en las que $\phi(x)$ es pequeño en relación con kT, el término *sinh* se puede linealizar estableciendo que $senh(\phi) = \phi$, lo que conocemos como PBE linealizado (que se reduce a la teoría de Debye-Huckel cuando no hay discontinuidad dieléctrica). Lamentablemente hay casos en los que se debe aplicar el PBE no lineal; por ejemplo, la

atmósfera de iones alrededor de macromoléculas altamente cargadas como el ADN o la Polilisina, pues se pierde parte de la simplicidad de la aproximación lineal, esto debido a que ignora las correlaciones y fluctuaciones de contraiones.

$$-\nabla \cdot \varepsilon(x)\nabla\phi(x) + \bar{\kappa}^2(x)\phi(x) = \frac{4\pi e_c^2}{kT}\sum q_i\delta\left(x - x_i\right) \tag{2.7}$$

En resumen, la teoría de Poisson-Boltzmann brinda resultados cuantitativos razonables para biomoléculas con baja densidad de carga lineal en soluciones salinas simétricas monovalentes; sin embargo, la teoría puede ser cualitativamente incorrecta para biomoléculas altamente cargadas o soluciones multivalentes más concentradas. Por lo tanto, la aplicación de la teoría y el software de Poisson-Boltzmann requiere cierta discreción y es una de las baterías de moléculas que se analizan en el trabajo a continuación.

2.4. Energía libre de Solvatación

Una vez el campo potencial, es decir la ecuación de PB, se resuelve por algún método numérico, se puede obtener distintos tipos de energía electroestática. El primer término, ΔG_{es}^{fixed}, está relacionado con el aporte de las cargas fijas del soluto. El segundo término, ΔG_{es}^{ions}, surge debido a los iones y se le denomina el término del potencial osmótico de la solución y para el caso de la investigación que se considera que la concentración no varía en el espacio, el término se hace 0. Y el tercero, ΔG_{es}^{sol} está relacionado con el solvente.

$$\Delta G_{es}^{t} = \Delta G_{es}^{fixed} + \Delta G_{es}^{ions} + \Delta G_{es}^{sol} \tag{2.8}$$

En general, la energía electroestática total se puede descomponer en la suma de tres términos, donde uno de ellos, denominado energía de solvatación toma nuestro interés para efecto de esta investigación, pues representa la energía necesaria para transferir el soluto desde el vacío hacia el solvente, al final refleja la cantidad de energía absorbida o liberada, si la energía es negativa la solvatación es favorable y el soluto tiene afinidad por el solvente.

En *Decherchi* [2], cuando linealizamos la ecuación de PB, se demuestra que todos los términos se reducen a ΔG_{es}^{fixed}. Para ello debemos partir sobre la idea de expandir el

término ya mencionado en otros tres; el primero de ellos, $\Delta G_{es}^{coulomb}$, que corresponde a la energía de coulomb de la interacción directa de las cargas fijas en un medio dieléctrico uniforme; el segundo término, ΔG_{es}^{rf}, se le conoce como la energía del campo de reacción, que es la energía debida a la interacción entre las cargas fijas y las cargas de polarización, que se encuentran en el interfaz donde hay un cambio en la constante dieléctrica; el tercer y último término, $\Delta G_{es}^{\text{excess}}$, que corresponde a la interacción de los iones móviles en la solución. Dadas estas definiciones, la energía electro-estática total se puede escribir como:

$$\Delta G_{es}^{t} = \Delta G_{es}^{\text{coulomb}} + \left(\Delta G_{es}^{rf} + \Delta G_{es}^{\text{excess}} + \Delta G_{es}^{\text{ions}} + \Delta G_{es}^{\text{sol}}\right) = \Delta G_{es}^{\text{coulomb}} + \Delta G_{es} \quad (2.9)$$

Agrupando los términos como se muestra en la ecuación 2.9 se diferencian dos contribuciones; la primera de ellas asociadas al término $\Delta G_{es}^{\text{coulomb}}$, que está relacionado directamente con la interacción de las cargas fijas como se mencionó con anterioridad y el segundo que agrupa todas las interacciones con el solvente y los iones en el medio.

$$\Delta G_{es}^{t} = \Delta G_{es}^{\text{conlomb}} + \Delta G_{es}^{rf} \quad (2.10)$$

De la linealización, podemos reducir la ecuación 2.9 de la energía total a la ecuación 2.10 en la que observamos que el único medio para que el soluto sienta el solvente es por medio de la energía del campo de reacción, que ocurre en la interfaz con el aporte de la interacción de los iones móviles.

En este punto para la resolución de la ecuación de PB es necesario tomar ciertas suposiciones, pues toma algunas características interesante al nivel del modelo geométrico. La primera de ellas tiene que ver con dieléctrico, pues existe una discontinuidad en el medio. Por otro lado se debe suponen la función de nube de iones, que normalmente se considera constante. Estos dos aspectos mencionados generan discontinuidad en el sistema.

Para solucionar el problema de la discontinuidad, anteriormente se había hecho mención de la superficie de Van der Waals, que es una de la formas de generar un volumen en torno a un cuerpo de interés, esto nos permite en métodos numéricos de elementos finitos o diferencias finitas, otorgar a cada punto medio de una rejilla un dieléctrico asociado, que

suaviza el factor de cambio del dieléctrico, tal como se muestra en la figura 2.3. El sistema lineal resultante puede ser solucionado por algún método iterativo.

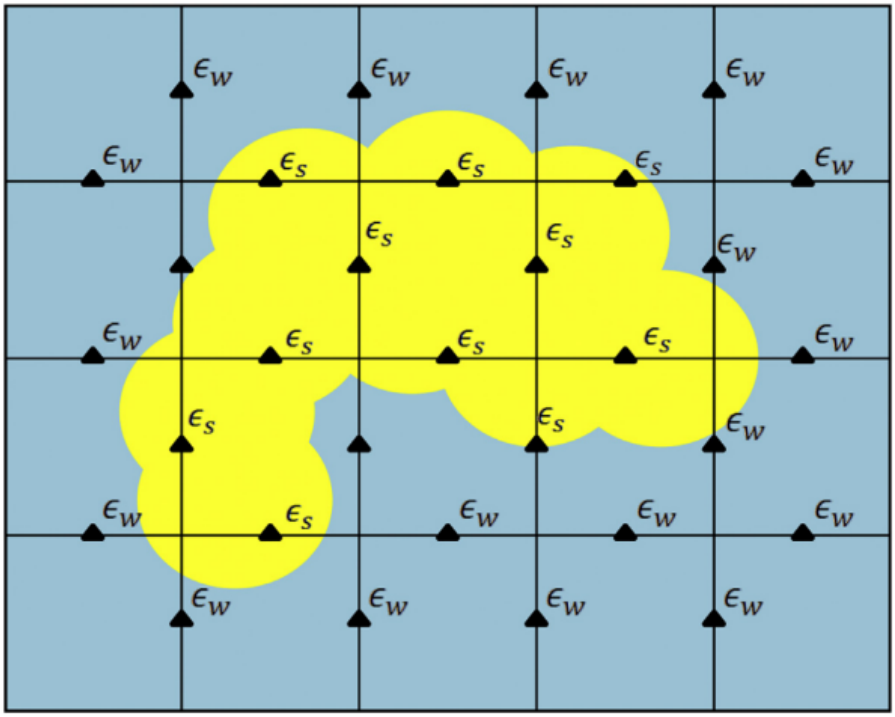

Figura 2.3: Discretización de la función del dieléctrico para diferencias finitas en un dominio 2D [2]

Uno de los métodos mencionados en *Decherchi* [2] corresponde a escribir la energía libre de solvatación en función de dos términos, un término polar (relacionado a la energía electroestática) y un término no polar (relacionado con el principio de exclusión de Pauli más la dispersión de Van der Waals), sin ahondar en definiciones, el término no polar suele aproximarse usando el método SASA que consiste en la utilización de radios de átomos pesados para definir un área de accesibilidad del solvente, la expresión queda explícita en la ecuación 2.11, donde γ es una constante con unidades de tensión superficial, que puede tomar valores de entre 20-30 [calmol^{-1}A]

$$\Delta G_{np} = \gamma A_{sasa} \tag{2.11}$$

Por otro lado el término polar, dado por la energía electroestática total, se muestra en la ecuación 2.12 que se separa en tres términos como se hacía alusión en la ecuación 2.9, la

energía debido a cargas fijas, la interacción entre cargas fijas y cargas de polarización, y el aporte de los iones móviles.

$$G_{es}[\phi] = \int_{\Omega} \left[\rho_f \phi - \frac{\varepsilon}{2}(\Delta\phi)^2 - \bar{\kappa}^2(\cosh(\phi) - 1) \right] dx \tag{2.12}$$

Donde se puede realizar la misma aproximación del cálculo de la PB para el $\cosh(\phi(x)) - 1 \approx \phi^2(x)/2$, quedando la ecuación 2.12

$$G_{es}[\phi] = \int_{\Omega} \left[\rho_f \phi - \frac{\varepsilon}{2}(\Delta\phi)^2 - \frac{\bar{\kappa}^2}{2}\phi^2 \right] dx = \frac{1}{2}\int_{\Omega} \rho_f \phi dx \tag{2.13}$$

Finalmente la ecuación de energía libre de solvatación, que es la suma de ambos aportes, queda como:

$$G_{es}[\phi] = \int_{\Omega} \left[\rho_f \phi - \frac{\varepsilon}{2}(\Delta\phi)^2 - \frac{\bar{\kappa}^2}{2}\phi^2 \right] dx + \gamma A_{sasa} \tag{2.14}$$

Capítulo 3

Metodología

3.1. Protein Data Bank

El RCSB PDB (Protein Data Bank) es un banco de datos que proporciona acceso a información estructural detallada de proteínas, ácidos nucleicos y otros macromoléculas biológicas, con posibilidad de descarga en un formato con extensión *.pdb*. El peso que tiene este banco de información es tal, que actualmente contiene más de 1 TB en estructuras y crece según estimaciones propias de ellos a un 10 % anual, con más de 2 millones de descargas en estructuras al día y contribuyendo a cerca de un millón de papers publicados, se posiciona como una herramienta casi de carácter obligatorio para investigadores, científicos, educadores, profesionales de la salud, sectores de la farmacéutica y biotecnología.

Para el estudio realizado se hace referencia a 3 baterías de moléculas; altamente cargada, que corresponden a ADN y ARN; medianamente cargado, proteínas; y moléculas pequeñas o esenciales.

3.1.1. Batería de alto nivel energético.

3.1.1.1. 1KX5

La proteína 1KX5, se clasifica como estructural y se ha estudiado en organismos como los Homo sapiens y Xenopus laevis (una especie de rana acuática comúnmente utilizada en investigación debido a su facilidad de manejo y desarrollo embrionario transparente). Está compuesto por 8 moléculas clasificadas como proteínas; histonas de distintos tipos (H3, H4, H2a.1, H2b.2) que permiten el empaquetamiento del ADN y 2 cadenas de nucleótidos. [4]

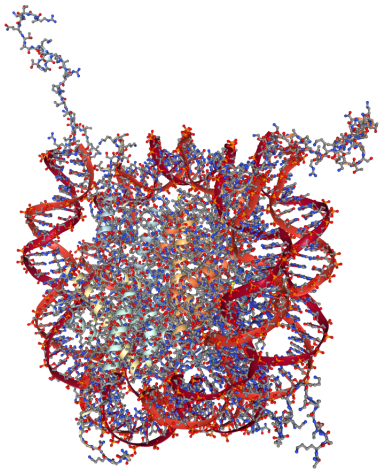

Figura 3.1: Representación de la molécula 1KX5 realizada con MDAnalysis

Descripción: Estructura de rayos X de la partícula central del nucleosoma, NCP147
Clasificación: PROTEÍNA ESTRUCTURAL/ADN
Organismo(s): Homo sapiens, Xenopus laevis
Sistema de expresión: Escherichia coli

3.1.1.2. 280D

280D corresponde a la estructura de un dodecámero de ARN, es decir, una molécula de ARN que consta de 12 nucleótidos (*RNA(5'r(*gp*gp*cp*gp*cp*up*up*gp*cp*gp*up*c)-3')*) con uniones específicas de las bases nitrogenadas donde se empareja un uracilo con un uracilo [4]. Como cadena de ARN está involucrada en la transmisión de información genética y la síntesis de proteínas.

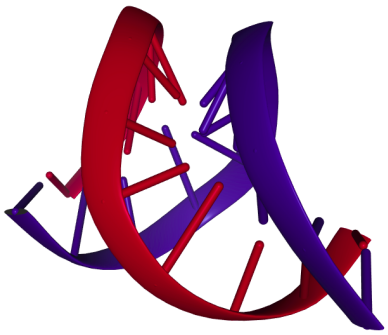

Figura 3.2: Representación de la molécula 280D realizada con MDAnalysis

Descripción: La estructura de un dodecámero de ARN muestra cómo los pares de bases u-u en tándem aumentan el rango de estructuras estables de ARN y la diversidad de sitios de reconocimiento.
Clasificación: ARN
Sistema de expresión: Escherichia coli BL21

3.1.1.3. 9ANT

La proteína 9ANT es un complejo entre el homeodominio de Antennapedia, que es un factor de transcripción implicado en la regulación genética durante el desarrollo embrionario de Drosophila melanogaster (Mosca de la fruta), y el ADN. Es crucial para controlar la expresión de genes específicos.

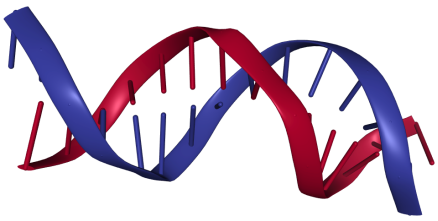

Figura 3.3: Representación de la molécula 9ANT realizada con MDAnalysis

Descripción: Antennapedia complejo homeodominio-ADN
Clasificación: Transcripción/ADN
Organismo(s): Drosophila melanogaster
Sistema de expresión: Escherichia coli

3.1.1.4. 437D

La proteína 437D es un pseudonudo de ARN del virus de la remolacha amarilla occidental, que participa en el cambio de marco ribosomal, que es un proceso en el que la maquinaria de traducción del ribosoma se desplaza a una posición diferente en el ARN mensajero, alterando la secuencia de aminoácidos de las proteínas sintetizadas.

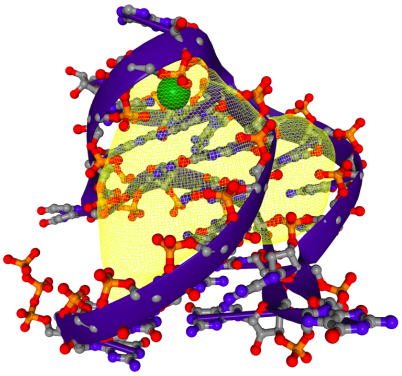

Figura 3.4: Representación de la molécula 437D realizada con MDAnalysis

Descripción: Estructura cristalina de un pseudoanillo de ARN del virus amarillo occidental de la remolacha implicado en el cambio de marco ribosómico.
Clasificación: ARN

3.1.2. Batería de proteínas medianamente cargadas

3.1.2.1. 1LYZ

La proteína 1LYZ corresponde a la estructura de la lisozima de la clara de huevo de gallina, que es una enzima clasificada como hidrolasa, y que cataliza la hidrólisis de los enlaces glucosídicos entre los residuos de N-acetilmurámico y N-acetilglucosamina en la pared celular de bacterias. [5]

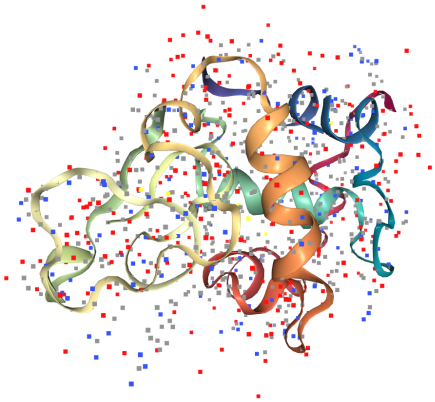

Figura 3.5: Representación de la molécula 1LYZ realizada con MDAnalysis

Descripción: Refinamiento en espacio real de la estructura de la lisozima de la clara de huevo de gallina
Clasificación: Hidrolasa
Organismo(s): Gallus gallus

3.1.2.2. 1UBQ

La ubiquitina, 1UBQ, es una pequeña proteína que desempeña un papel en la regulación de la degradación de proteínas y otros procesos celulares. Esta clasificada como una proteína cromosómica y se encuentra en organismos como el Homo sapiens.

Descripción: Estructura de la ubiquitina
Clasificación: Proteína cromosómica
Organismo(s): Homo sapiens

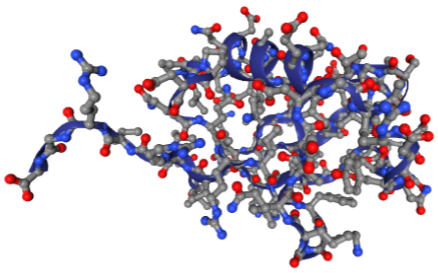

Figura 3.6: Representación de la molécula 1UBQ realizada con MDAnalysis

3.1.2.3. 1PGB

Se clasifica como una proteína de unión a inmunoglobulinas o anticuerpos presentes en el sistema inmunológico. El análisis de estas estructuras proporciona información sobre las interacciones entre la proteína G estreptocócica y las inmunoglobulinas, lo que es relevante para comprender la respuesta inmunológica.

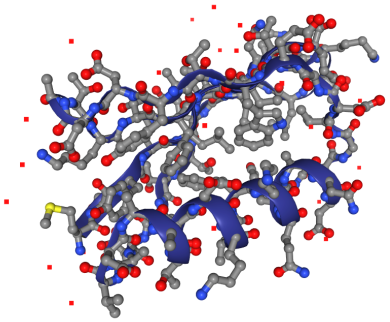

Figura 3.7: Representación de la molécula 1PGB realizada con MDAnalysis

Descripción: Dos estructuras cristalinas del dominio de unión a inmunoglobulina b1 de la proteína g estreptococal y comparación con la rnm
Clasificación: Proteína de unión a inmunoglobulinas
Organismo(s): Streptococcus sp. GX7805

3.1.3. Batería de compuestos orgánicos, baja carga

3.1.3.1. Anilina

La anilina es un compuesto químico orgánico que pertenece a la clase de las aminas aromáticas. Su fórmula química es C6H5NH2, y es líquida a temperatura ambiente con un

olor característico a amoníaco. Es soluble en la mayoría de los disolventes orgánicos pero es poco soluble en agua.

Figura 3.8: Estructura de la Anilina

3.1.3.2. Tolueno

El tolueno, también conocido como metilbenceno, es un compuesto químico orgánico que pertenece a la clase de los hidrocarburos aromáticos. Su fórmula química es C7H8. Es un líquido incoloro con un olor característico y es altamente volátil. Se encuentra naturalmente en el petróleo crudo y se utiliza ampliamente como disolvente en la industria química.

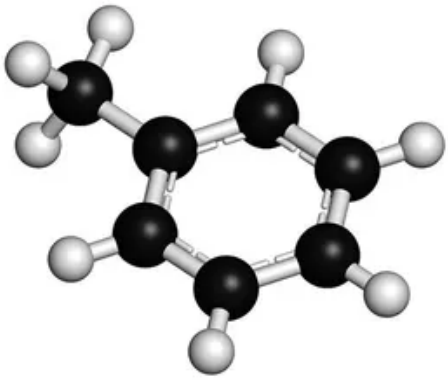

Figura 3.9: Estructura del Tolueno

3.1.3.3. Etanol

El etanol, también conocido como alcohol etílico, es un compuesto químico orgánico y una forma común de alcohol. Su fórmula química es C2H5OH. Es un líquido transparente e incoloro con un olor característico y se puede mezclar en agua.

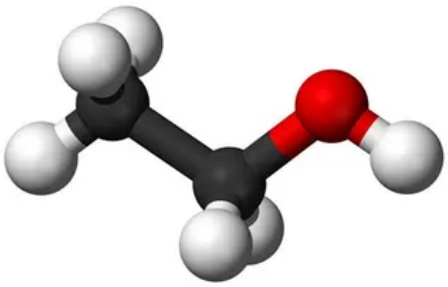

Figura 3.10: Estructura del Ethanol

3.1.3.4. Methyl Octanoate

El metil octanoato es un compuesto químico orgánico perteneciente a la clase de los ésteres. Su fórmula química es C9H18O2. Es un líquido incoloro con un aroma agradable a frutas y se utiliza en diversas aplicaciones.

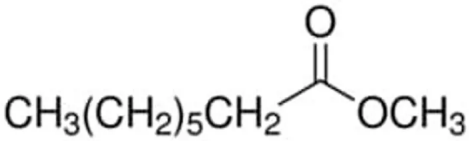

Figura 3.11: Estructura del Methyl Octanoate

3.2. Herramientas y softwares utilizados

3.2.1. PDB2PQR

Este software preparar las estructuras para cálculos de solvatación continua, así como otros tipos de modelado, análisis y simulación de estructuras biomoleculares. Estas tareas incluyen:

1. Agregar un número limitado de átomos pesados (no hidrógeno) faltantes en las estructuras biomoleculares.
2. Generar la protonación de moléculas biológicas, de manera que se generen enlaces de hidrógeno.
3. Asignar parámetros de carga y radio para una variedad de campos de fuerza.

4. Generar un archivo de tipo "PQR"que es compatible con varios paquetes de modelado y análisis computacional.

El software tiene como objetivo facilitar la configuración y ejecución de cálculos electrostáticos para cualquier usuario, ampliando así el acceso al análisis de solvatación y electrostática biomolecular en la comunidad [6].

La distribución de los datos de un archivo PQR tiene la siguiente forma: FieldName AtomNumber AtomName ResidueName ChainID ResidueNumber X Y Z Charge Radius.

- FieldName: Es la cadena de entrada y puede ser ATOM, que corresponden a los átomos pertenecientes a la biomolécula o HETATM que fueron aquellas ligadas por medio del software.
- AtomNumber: Es un índice que se le proporcionó a cada átomo, y corresponden a los naturales positivo.
- AtomName: Nombre del átomo.
- ResidueName: Tiene relación con el tipo de cadena de átomos al que está conectado el átomo, por ejemplo: GLU, LYZ, etc.
- X Y Z: Posición del átomo.
- Charge: Carga asociada al átomo en unidades de electrones.
- Radius: Radio atómico asociado al átomo en $[\mathring{A}]$.

3.2.2. APBS (Adaptive Poisson-Boltzmann Solver)

Por otro lado el software APBS resuelve las ecuaciones de electrostática continua. Fue diseñado asegurando su capacidad para interactuar con otros paquetes computacionales y evolucionar a medida que los métodos y aplicaciones cambien con el tiempo. [7]

Una vez transformado el archivo PDB en un archivo PQR se resuelve la ecuación de PB, para ello requiere como entrada un archivo en formato *.in* que contiene tres partes independientes para definir las condiciones del sistema:

1. READ: Como el nombre menciona hace referencia a la lectura de un archivo input, en el caso del estudio realizado, un archivos PQR.

2. ELEC: Después de este comando se especifican un listado de variables, condiciones y estados que se requieren. Donde adicionalmente pueden haber múltiples selecciones de ELEC en el mismo archivo. El desglose de los comandos se puede clasificar en tres partes; el *id* que es un nombre opcional al código para efectuar cálculos en la siguiente sección; El tipo de comando *type* que depende de la forma en que se quiere resolver o llevar a cabo los resultados, por ejemplo, tenemos *fe-manual* que es para hacer cálculos de elementos finitos con *FEtk* y *mg-manual* que es para el cálculo utilizando multigrid con configuraciones manuales. Otras instancias son: *mg-auto*, *mg-para*, *mg-dummy*; Y dependiendo del tipo de comando se extiende una lista de variables o *Keywords* que serán presentadas en breve con las suposiciones de uso.

3. PRINT: Como se indica, en esta sección se pueden manipular la información obtenida para realizar cálculos lineales en la salida del código.

3.2.2.1. Keywords - APBS

- dime : Número de puntos que se utilizan para discretizar la malla, donde para efecto de la investigación de usan valores de $n = 32x + 1$ con $x \in \mathbb{N} - \{1\}$. Tiene entrada en tres dimensiones {nx, ny, nz}.
- grid : Espaciado de la malla, tiene una entrada en tres dimensiones {hx, hy, hz}.
- gcent : Define el centro de la malla, puede ser de forma manual por un formato {gx, gy, gz} o utilizando el comando *mol* seguido del código verificador de molécula en caso de haber más de una en el modelo. Para efectos de la investigación utilizamos *mol 1*.
- mol : Especifica la molécula a la cual se le aplicará la ecuación de PB.
- lpbe/npbe: Define si la ecuación se resuelve por una aproximación lineal(*lpbe*) o de forma no lineal(*npbe*).
- bcfl: Especifica el tipo de condición de borde. Se escogió utilizar *mdh* (Multiple Debye Hückel), que considera múltiples esferas con puntos de carga donde el radio es igual al radio atómico de la molécula y la carga corresponde a la carga total de la proteína.
- ion: Permite definir la concentración de las cargas móviles, requiere el tipo de carga,

la concentración y el radio. Tiene como entrada el siguiente formato {charge {charge} conc {conc} radius {radius}}.

- pdie: Corresponde a la constante dieléctrica del soluto, que normalmente se encuentra en un rango de 2-20. Se considera un valor de 2 que supone tener en cuenta la polarización pura, pero existe data que para valores más altos, por ejemplo $\epsilon = 4$ tiene una mayor concordancia con los datos experimentales.
- sdie: Corresponde a la constante dieléctrica del solvente, que normalmente se encuentra en un rango de 78-80. Se considera un valor de 78.54 de medición empírica.
- srfm: Hace referencia al modelo utilizado para definir el coeficiente dieléctrico y el de accesibilidad iónica. Tras llamar al comando se tienen variadas opciones, entre ellas; mol, smol, spl2, spl4. En particular la que se usa es *smol*, que es una versión mejorada de *mol* en la que los coeficientes se suaviza mediante un promedio para suavizar los efectos de la malla.
- chgm: Corresponde al método con el cual se mapean las cargas puntuales para el posterior cálculo de la ecuación de PB, la cual depende del espaciado de la malla. Para aminorar los efectos generados por el tamaño de malla se hace uso del comando *spl2* que describe una discretización cúbica B-spline en la que se utilizan funciones de base cúbica para representar la distribución de carga de la malla.
- sdens: es la densidad de puntos por A^2, típicamente el valor recomendado por APBS es 10,0
- srad: Corresponde al radio de las moléculas del solvente, en el caso de estudio como se trabaja con agua se recomienda un radio de 1.4 [A].
- swin: Especifica la tasa de cambio para las definiciones de superficies basadas en *splines*, típicamente se recomienda un valor de 0.3.
- temp: Temperatura, en este caso se toma un valor de 298.15 [K]
- calcenergy: Controla la salida por pantalla de energía electroestática calculada. El comando *total* retorna la energía electroestática total de la molécula.
- calcforce: Controla la salida por pantalla de fuerza electroestática calculada. El comando *no* implica que no entrega el valor.

- write: un comando para generar archivos de texto con la información, para esta investigación se obtienen los datos en un formato de texto con terminación *.dx* .

En general los archivos de texto lucen como la figura 3.12

```
read
    mol pqr 9antmodif.pqr
end
elec name solv
    mg-manual
    dime 289 289 289
    grid 0.45 0.45 0.45
    gcent mol 1
    mol 1
    lpbe
    bcfl mdh
    ion charge 1 conc 0.125 radius 1.4
    ion charge -1 conc 0.125 radius 1.4
    pdie 2.0000
    sdie 78.5400
    srfm smol
    chgm spl2
    sdens 10.00
    srad 1.40
    swin 0.30
    temp 298.15
    calcenergy total
    calcforce no
    write pot dx test-solv
end
print elecEnergy solv end
quit
```

Figura 3.12: Línea de comandos en un archivo *.in* requerido por APBS para la molécula de ADN *9ANT* con una malla de 285x285x285 y una distancia entre puntos de 0.45 en toda dirección.

3.2.3. H++

Lamentablemente, PDB2PQR tiene ciertas complicaciones para transformar archivos que contengan nucleótidos, es por ello que el Instituto Politécnico de Virginia desarrolló un software de libre acceso denominado H++ perteneciente al grupo newbiophysics, que extiende el uso a este tipo de moléculas, incluso también desarrolla el labor de APBS, pero para motivos de la investigación solo se hace uso de su funcionalidad para obtener archivos PQR, que normalmente PDB2PQR genera de manera incompleta. [8]

Figura 3.13: Logo del software H++

3.2.4. MDAnalysis

MDAnalisys es una biblioteca orientada al cálculo de trayectoria en dinámica molecular, con un amplio espectro de alternativas y posibilidades, siendo muy flexible y relativamente rápido para algunas tareas que requieran análisis complejos. La documentación da una serie de casos y ejemplos en la que es posible aprovechar las virtudes de la biblioteca, por ejemplo: enlaces de hidrógenos en puentes de agua, membranas, ácidos nucleicos, polímeros, macromoléculas, líquidos, análisis volumétrico como; densidad, dieléctrico y dinámica del agua. Se mencionó lo flexible que es, pues tiene posibilidad de trabajar con archivos de tipo PDB, PQR, MOL2, GRO o GSD, por mencionar algunos.

En general, para uso de la investigación se hace uso de la flexibilidad que entrega para trabajar con data y archivos de formato *.dx*, lo que nos permite de una manera muy accesible realizar operaciones sobre mallas, adicionalmente las visualizaciones de las baterías de alta y mediana carga fueron obtenidas por medio de visualizadores de MDAnalysis.

3.2.5. Otras bibliotecas de Python Utilizadas

- numpy: Una biblioteca altamente usada en computación científica, dado su integración con el resto de bibliotecas, operaciones de álgebra lineal y funciones matemáticas.
- matplotlib: utilizada para visualización de datos.
- pandas: Para operaciones estadísticas y análisis de datos.
- csv: Para creación y manipulación de archivos .CSV

3.3. Análisis de sensibilidad

Para atenernos lo mejor posible a la diferencia de cálculo entre un método lineal y uno no lineal, se requiere un análisis de sensibilidad que nos permita establecer condiciones de caja que no alteren los resultados por imprecisión de los datos y construcción de un mallado poco refinado. Para ello se hace uso de la variable de energía de solvatación, que es un escalar obtenido del cálculo de la ecuación de PB. Vale mencionar que se toma como supuesto que si el comportamiento de la energía no cambia por efectos de la caja que se quiere utilizar, tampoco lo harán los potenciales.

3.3.1. Tamaño de caja

El primer análisis tiene relación con qué tan lejos deben estar las paredes del mallado que contiene a la molécula, para ello se establece la mayor distancia cardinal que contiene el plano cartesiano y tangente más lejano, a esta se le establece como la caja cero.

Se define una función *delta_ ensayo* que busca un punto inicial que se encuentre a una distancia de 1[A] o menos, esto como método comparativo para entender qué ocurre con cajas muy pegadas a las biomoléculas, aunque su objetivo principal tiene relación con encontrar una distancia en la que el cambio energético no sea considerable. Encontrado el *dx* se usa la función *espaciamiento* que nos permite determinar un vector de prueba en función de los *dimes* definidos por múltiples de 32i+1, para i>=2, los cuales para el *dx* definido y distintos valores del dime aplicados software APBS se calcula la energía de solvatación.

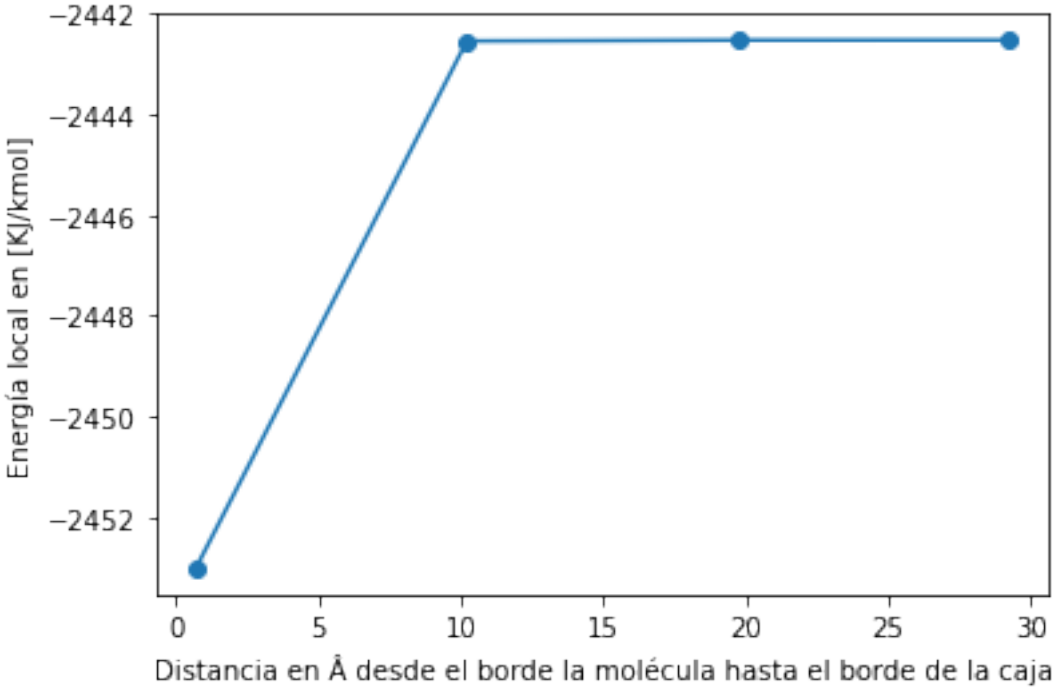

Figura 3.14: Ejemplo de análisis de sensibilidad para el tamaño de caja

En la figura 3.14 se observa un primer punto a una distancia pequeña inferior a 1[Å] y luego cálculos consecutivos al extender la caja. En algún punto es posible definir que el cambio porcentual es menor a 1 % y suponerlo como una distancia segura.

3.3.2. Nivel de refinamiento

Dada las dimensiones de las cajas, en el segundo análisis se busca determinar qué tan refinado tiene que ser el mallado, es decir, cuál es el número de nodos por eje óptimos que subdividan la malla de forma que los resultados del cálculo de la energía de solvatación no cambie significativamente. Por tanto a medida que aumentamos el número de puntos, disminuimos proporcionalmente el valor del dx para que se ajusten a la caja determinada en el primer ensayo.

La figura 3.15 muestra un ejemplo del cálculo realizado por APBS que muestra un cambio significativo de usar un mallado de 97 nodos versus uno de 353 nodos, en casi 100 [Kj/kmol].

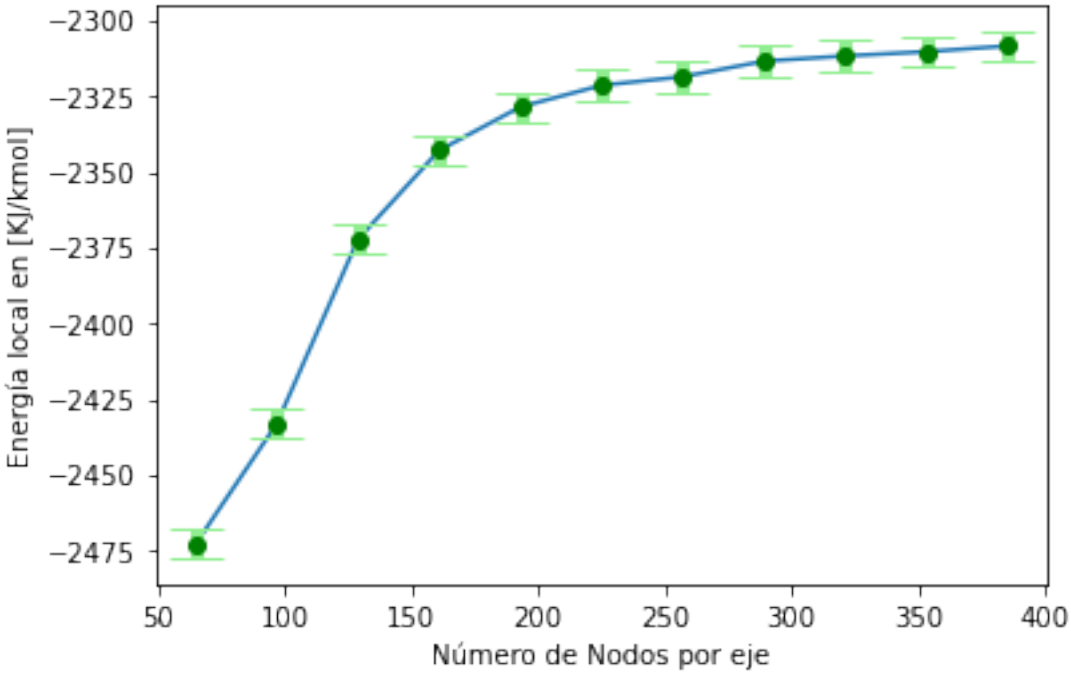

Figura 3.15: Ejemplo de análisis de sensibilidad para determinar el número de nodos por eje.

3.4. Potencial electroestático

Por medio de APBS se realizan 3 cálculo de la ecuación de PB; el primero de ellos para determinar el potencial total usando como parámetros los mostrados en la figura 3.12, con las consideraciones de caja establecidas en el análisis de sensibilidad, la nomenclatura *lpbe* para indicar linealidad, y estableciendo el dieléctrico del agua en 78.54; un segundo cálculo, que se ilustra en la figura 3.16, para determinar el potencial lineal en el vacío con la diferencia de que las lineas que indican las concentraciones ya no están y el dieléctrico del vacío se considera igual al de la molécula; y un tercero para determinar el potencial total no lineal que a diferencia del primero usa la nomenclatura *npbe*, ver figura 3.17

APBS permite realizar los 3 cálculos superponiendo los cuadro de códigos, e incluso realizar algunas operaciones matemáticas dentro del archivo *.in*, como se muestra al final de la figura 3.17, en particular se usó como parámetro para encontrar la diferencia entre el potencial total y el potencial en el vacío, lo que denominamos como energía libre de solvatación.

```
elec name vacc
    mg-manual
    dime 321 321 321
    grid 0.2 0.2 0.2
    gcent mol 1
    mol 1
    lpbe
    bcfl mdh
    pdie 2.0000
    sdie 2.0000
    srfm smol
    chgm spl2
    sdens 10.00
    srad 1.40
    swin 0.30
    temp 298.15
    calcenergy total
    calcforce no
    write pot dx 9ant-PotVacc-l
end
```

Figura 3.16: Cuadro de comandos en un archivo .in requerido por APBS, para el cálculo de potencial en el vacío.

```
elec name reac
    mg-manual
    dime 321 321 321
    grid 0.2 0.2 0.2
    gcent mol 1
    mol 1
    npbe
    bcfl mdh
    ion charge 1 conc 0.125 radius 1.4
    ion charge -1 conc 0.125 radius 1.4
    pdie 2.0000
    sdie 78.5400
    srfm smol
    chgm spl2
    sdens 10.00
    srad 1.40
    swin 0.30
    temp 298.15
    calcenergy total
    calcforce no
    write pot dx 9ant-PotTol-nl
end
print elecEnergy solv - vacc end
quit
```

Figura 3.17: Cuadro de comandos en un archivo .in requerido por APBS, para el cálculo de potencial total no linealizado.

3.5. Generación de Cortes

Dado que se quiere visualizar puntos de interés en la caja de estudio, se desarrolló un código de tres partes; la primera de ellas consiste en encontrar puntos de interés seccionando el cubo de trabajo en cubos más pequeños, y seleccionando los máximos y mínimos locales por zona, como muestra la figura 3.18

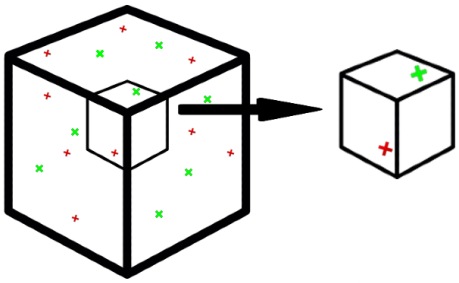

Figura 3.18: Selección de máximos y mínimos locales por medio de particiones.

El pseudocódigo [9] para esta iteración de muestra a continuación

```
Definir las listas vacías para cada una de las variables; x,y,z,maxvalues
Usar el dime asociado al mallado que se desea subdividir
Determinar número de particiones que se quieres, notar que si el valor es 2, se forman
8 cubos de análisis, crece por las dimensiones.
Generar el factor de cambio para realizar los saltos
for i en el rango de 0 al dime, haciendo saltos del valor factor do
    for j en el rango de 0 al dime, haciendo saltos del valor factor do
        for k en el rango de 0 al dime, haciendo saltos del valor factor do
            Encontrar máximo y mínimo local
            if Cumple condición then
                Agregar a la lista
            end if
        end for
    end for
end for
```

La segunda parte del código consiste en una función denominada *verticesenplano* que toma el archivo de vértices *.vert*, obtenido de usar *MSMS*, y determinar por medio de un plano

todos los vértices que vivan a una distancia segura, nosotros tomamos un valor de 0.5[A] y agregarlo en una data para su posterior utilización. La función tiene como entrada el valor del plano que se requiere, y una serie de *Flags* para cada eje. La idea es que una de esas banderas tome el valor de *True* para el eje de interés.

```
Se abre como lectura el archivo .vert
Se genera una lista vacía
Se obtienen los datos de la linea
if Flag1 == True then
    if el valor asociado a la coordenada es mayor o menor en 0.5[A] then
        Se agrega a la data
    end if
    Se repite para las restantes dos Flags
end if
```

El tercer código genera los gráficas que se visualizan luego en resultados como *Cortes* para visualizar los potenciales en la malla.

```
Se definen los mallados por medio de la función meshgrid para cada uno de los planos
for i en el total de los máximos o mínimos encontrados do
    Se generan listas vacías para cada plano
    for i en el valor del dime utilizado do
        Agregar en cada plano los valores de recorrer el mallado de potenical en en el
    array del máximo o mínimo respectivo
    end for
    Se plotea cada plano considerando incluir por medio de la función verticesenplano
    el ploteo de la superficie de la molécula sobre los planos de potenciales obtenidos
end for
```

Este proceso se puede iterar para determinar el potencial total lineal, el potencial de reacción y el error de potencial lineal menos no lineal.

Es importante aclarar por si no queda explícito anteriormente de que el valor del potencial de reacción es la diferencia del potencial total menos el potencial en el vacío. Y la diferencia lineal no lineal se realiza de la diferencia entre potencial total lineal y potencial total no lineal, que son data definida en el cálculo por medio de APBS.

3.6. Nivel de cobertura

Uno de los análisis que se requirió realizar, se toma como idea preliminar al momento de observar los distintos cortes de error o focos de potencial que se generaban en moléculas de alta y mediana carga. Para poner en términos cuantificables esta observación se definió como nivel de cobertura, a un porcentaje de 0 a 1 teórico. Para un punto ubicado en el espacio fuera de la proteína, al cual se le generaba una esfera alrededor y se iteraban con haces de luz por cada diferencial de área que tuviera la esfera, así se generaba una proporción si es que está rebotaba con una de las triangulaciones de la proteína.

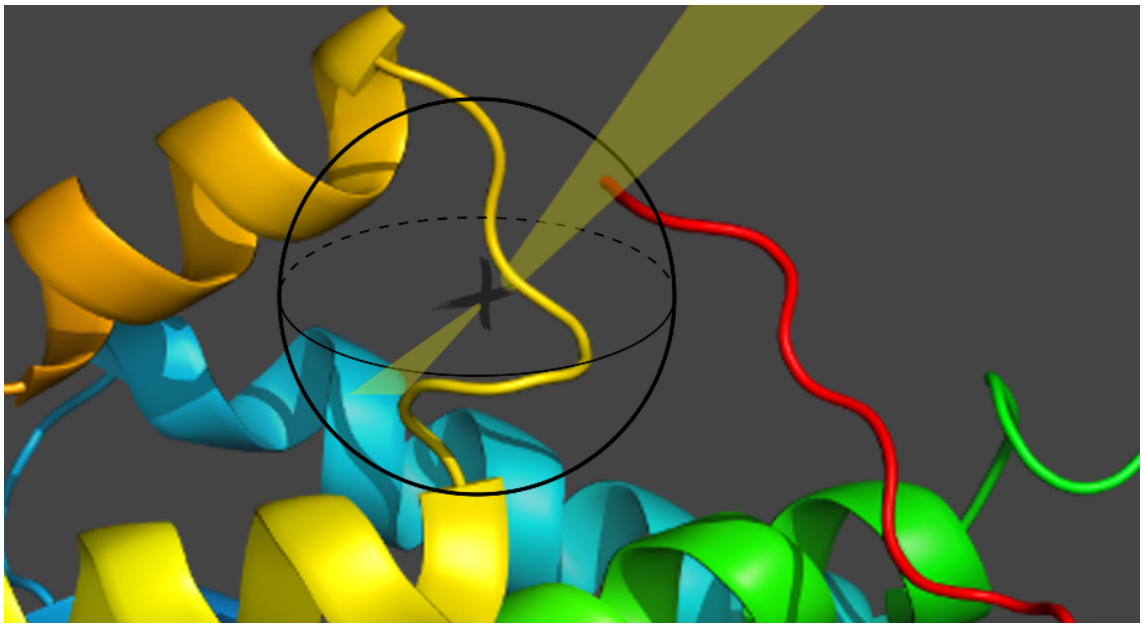

Figura 3.19: Representación del modelo de cobertura por medio de haces de luz.

La función de cobertura se puede ejemplificar por medio del siguiente pseudocódigo [9].

1: Generación de data relacionada a la esfera que rodea al punto de interés
2: Definir número de haces de luz igual a 0
3: **for** Para cada punto en la esfera **do**
4: Se realiza un vector posición con respecto al origen o punto de interés
5: **for** Para cada triangulación de la proteína **do**
6: Se mide si intersecta un haz y se calcula la distancia del origen a la triangulación
7: **if** La distancia es distinta de nula, es decir, intercepta. Y la distancia es menor que el radio de la esfera **then**

```
            Se aumenta en 1 el número de haces de luz que interceptan.
        end if
    end for
end for
```

A continuación se muestran algunos pseudocódigos [9] de las funciones utilizadas, primero aquellas para los puntos pertenecientes a la esfera. Y como relacionar las datas de vértices y de caras obtenidas por medio de MSMS.

```
Definir función Generar puntos en una esfera de radio R
Se define una lista vacía para contener los puntos
Se determinar el incremento en función de número de puntos
for Para cada i hasta el número de puntos requeridos do
    Se calcula la coordenada y y la coordenada r
    Se determina el ángulo Phi según el nivel de avance que se lleva de los puntos
    Se calcula la coordenada x y z por coordenadas polares
    Se agrega a la lista vacía
end for
```

También se muestra el pseudocódigo [9] de la función que realizar la intersección de un haz de luz con la triangulación de las caras de la proteína.

```
Calcular los bordes de un triangulo
Se define un valor h como el producto cruz entre la dirección del haz y uno de los
    lados de la triangulación
Se define un valor a como el producto escalar entre el otro lado de la triangulación y h
if El rayo está cerca de cero el rayo then
    Retorna None porque son paralelos
end if
Calcula las coordenadas baricéntricas
Comprueba si la distancia entre el haz y la triangulación está dentro de la esfera, sino
    retorna None
Calcula la distancia a la intersección
Devuelve t si hay una interacción, sino devuelve None
```

3.7. Recolección de data

Para proceder con la recolección de data es necesario generar funciones que nos permitan obtener datos de interés. Se escogieron varias opciones, entre ellas el potencial de Coulomb, la derivada del potencial de coulomb, distancia mínima al contorno de la proteína, nivel de cobertura, distancia mínima considerando radio atómico, entre otras. Aquellas con resultados coherentes fueron; Potencial de Coulomb, Potencial total lineal, Potencial de reacción y Distancias mínimas al contorno. Como el segundo y tercer análisis salen directos de los archivos *.dx* obtenidos con anterioridad, queda ejemplificar como calcular el potencial de Coulomb para cada punto, que se puede visualizar en el pseudocódigo [9] a continuación, en donde dado un punto de interés y la totalidad de las cargas de la proteína se procede a calcular el potencia de Coulomb y la menor distancia de una carga al punto de interés, tomando valores de dieléctrico igual a 1.

```
Se carga el archivo PQR
Se crean las variables de potencial y distancia mínima
for  cada átomo en el archivo PQR do
    Se calcula la distancia del punto de interés a la carga
    if Si la distancia es menor que la distancia mínima definida then
        distancia mínima = distancia
    end if
    Potencial += carga/distancia
end for
```

También mencionar que se establecieron funciones para otras variables como la derivada del potencial de Coulomb, que no resultaron fructíferas en los análisis.

Continuando con el desarrollo de la data se tomó un muestreo aleatorio de 100.000 datos tomando en consideración el potencial de Coulomb, Potencial de reacción lineal y potencial total lineal versus el error dado por la diferencia lineal-no lineal. En el GitHub [9] se muestra el código de obtención de la data y su posterior visualización.

Capítulo 4

Resultados y Discusión

4.1. Batería de alto nivel energético

Como se daba a entender implícitamente al momento de presentar el grupo o batería de moléculas que definen lo que denominamos *alto nivel energético* , estas consisten principalmente de una cadena de ADN o ARN, con ciertos casos como la 1KX5 que también contiene un empaquetamiento estructural. Las moléculas 280D y 9ANT fueron analizadas como cuerpos compuesto únicamente por aminoácidos, siendo esta última, la 9ANT, modificada para ese objetivo, y posteriormente analizada con la proteína que había sido retirada. En el caso de la 437D, que corresponde a una cadena de ARN mantiene la particularidad de que se enrolla sobre sí misma.

4.1.1. 1KX5

Algunos comentarios de la ejecución del análisis de la molécula 1KX5 están relacionados a lo grande que era, siendo la molécula de estudio que presentó la mayor magnitud de error lineal no lineal, alcanzando 91.413 [kT/e] como se muestra en la figura 4.2, donde el error está siempre muy cercano a la interfaz. En la figura 4.3 se muestra un comportamiento creciente para los potenciales con un difusión de los datos a medida la magnitud del valor analizado crece.

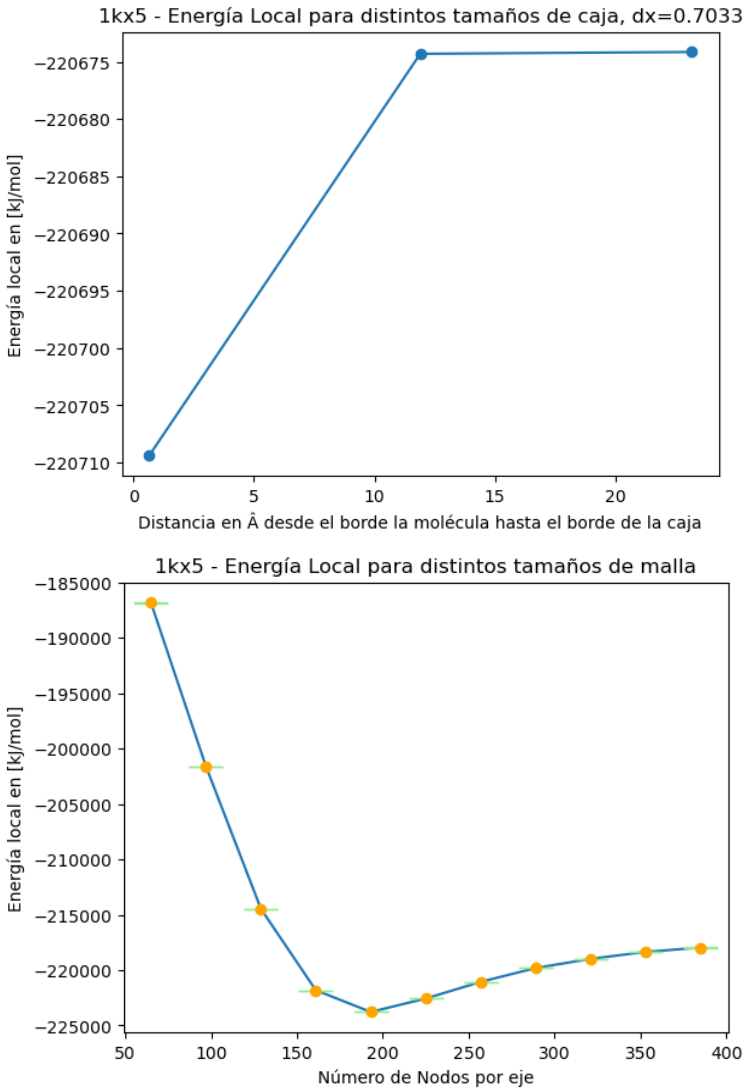

Figura 4.1: Análisis de sensibilidad para la molécula 1KX5

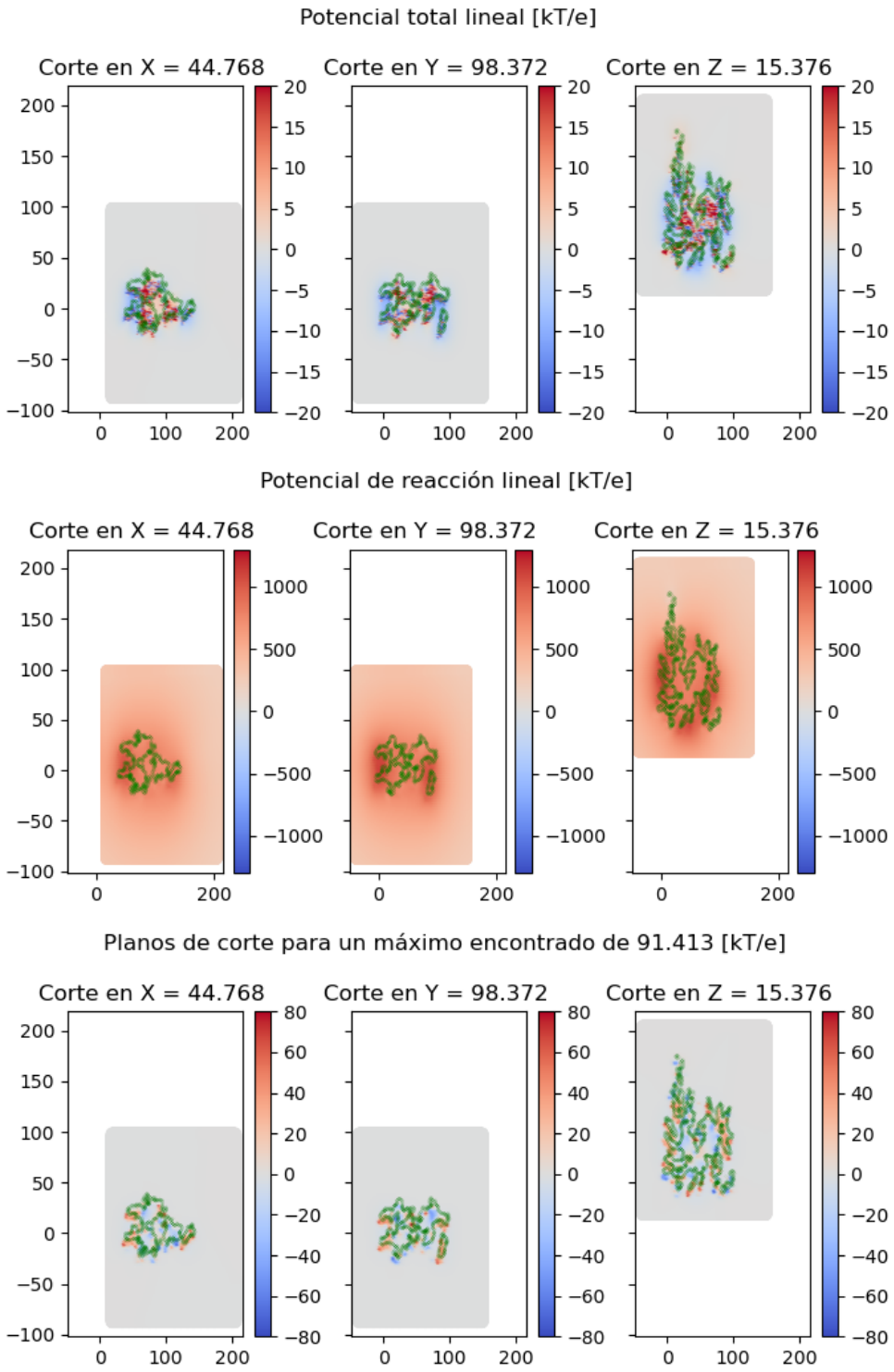

Figura 4.2: Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula 1KX5

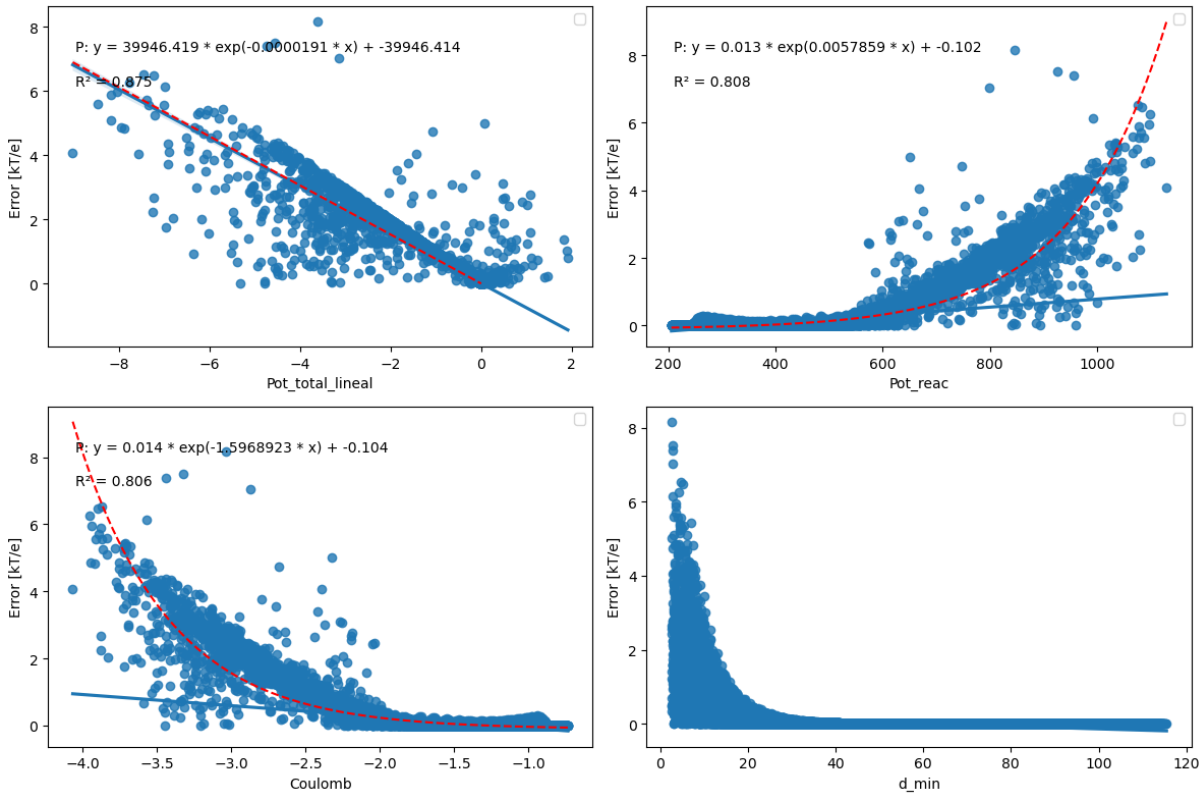

Figura 4.3: Error lineal no lineal, Comportamiento de 100.000 datos

4.1.2. 280D

Como se indicaba con anterioridad, la molécula 280D son dos cadenas de 12 nucleótidos que se conectan entre sí elemento a elemento. Una característica que se encuentra son los altos niveles del potencial de reacción en el interior de ésta, fig 4.5, siendo foco de los mayores errores y mostrando alta correlación de los datos al evaluar el efecto de esta sobre la magnitud del error, fig 4.6.

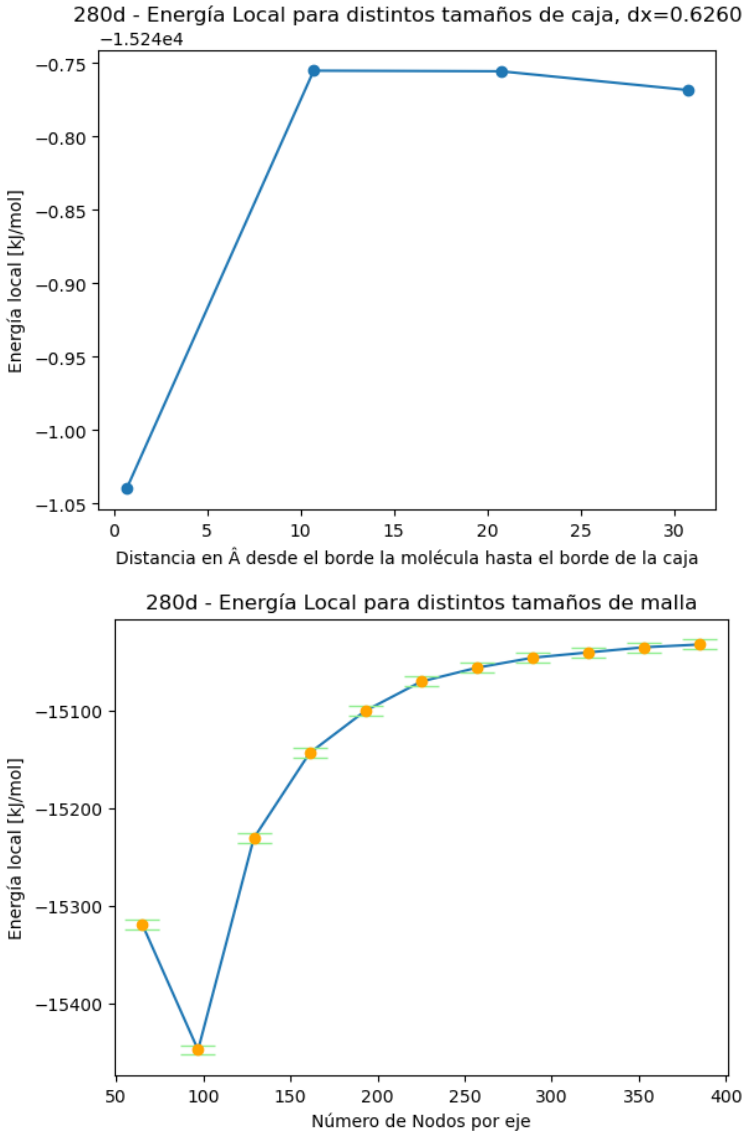

Figura 4.4: Análisis de sensibilidad para la molécula 280D

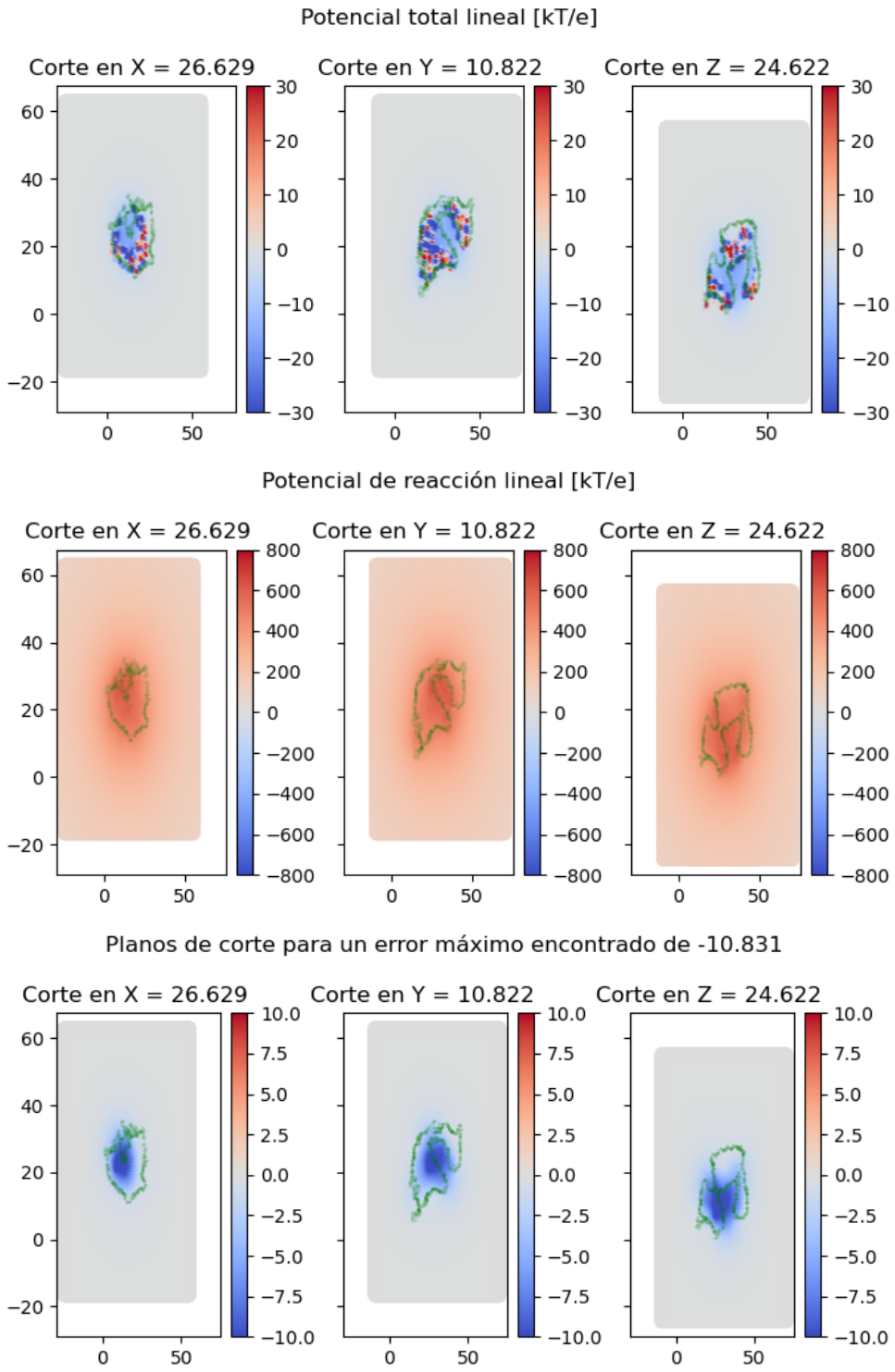

Figura 4.5: Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula 280D

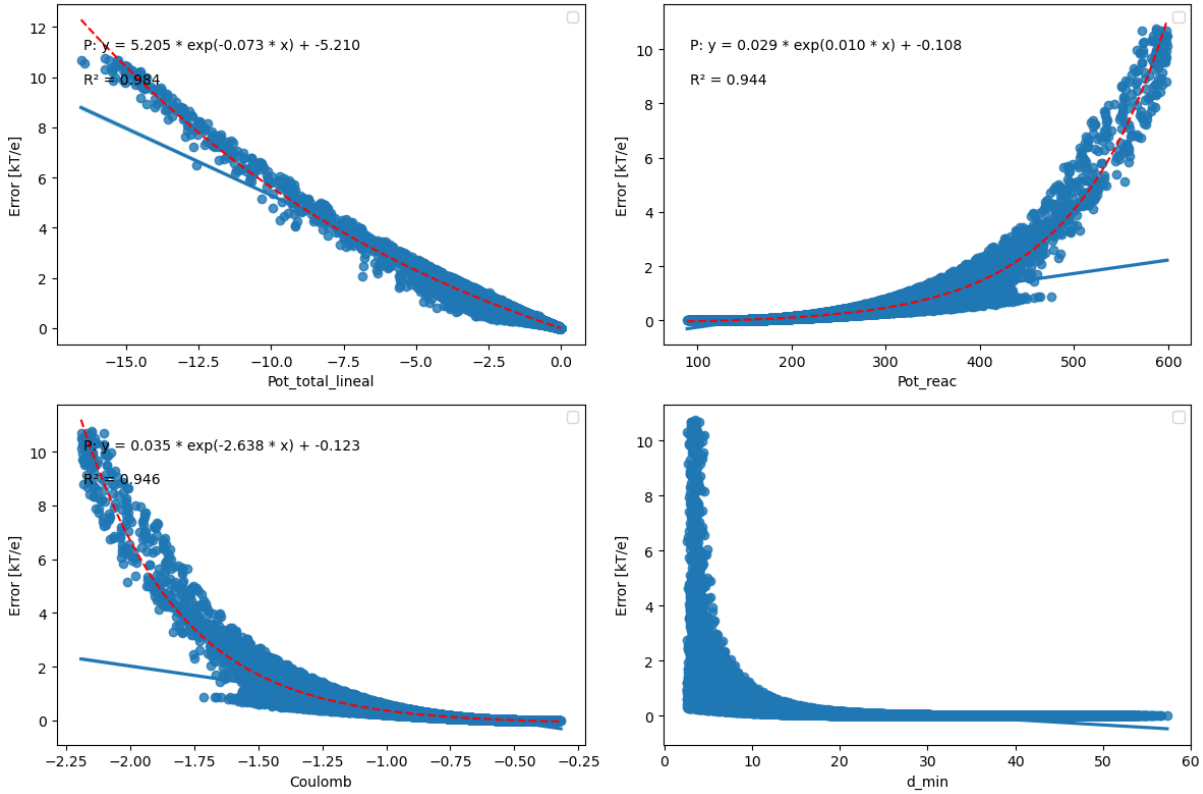

Figura 4.6: Error lineal no lineal, Comportamiento de 100.000 datos

4.1.3. 9ANT

El análisis de la molécula 9ANT nace como inspiración tras la lectura del paper de Yu Iwahara [10], que por medio de resonancia magnética nuclear logra evaluar el potencial electroestático entorno a una molécula de ADN en presencia de cargas móviles. Sin ahondar en ese tema, los resultados obtenidos de la molécula presentan un caso similar a la 280D, en la que los focos de error se concentran en las aberturas de las helicoides de la cadena y en la zona tubular interna.

4.1.3.1. Sin proteína anexada

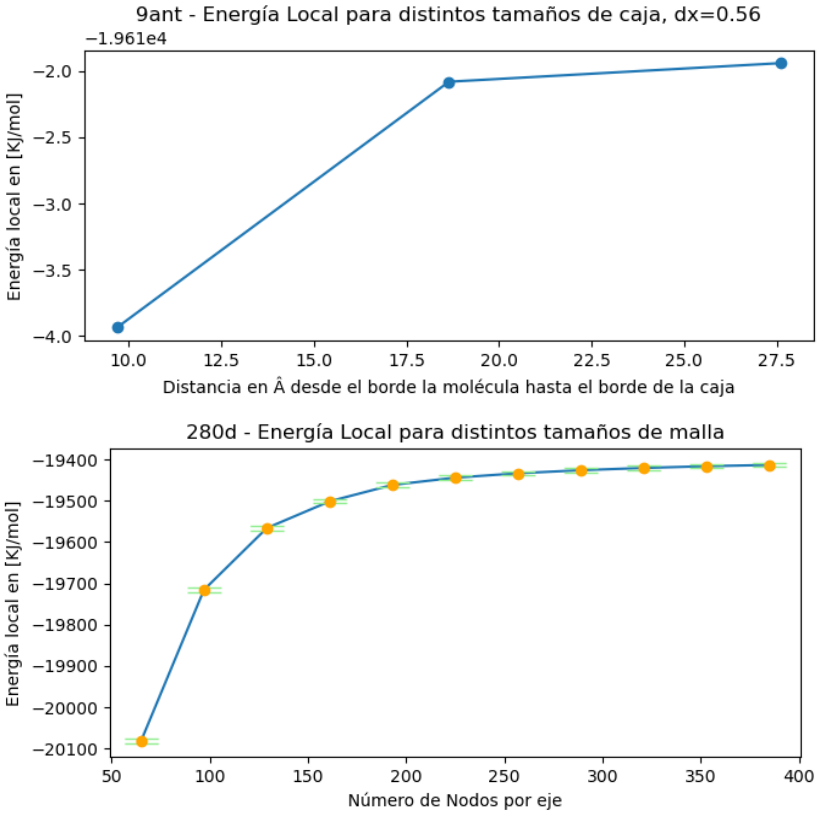

Figura 4.7: Análisis de sensibilidad para la molécula 9ANT sin proteína anexada

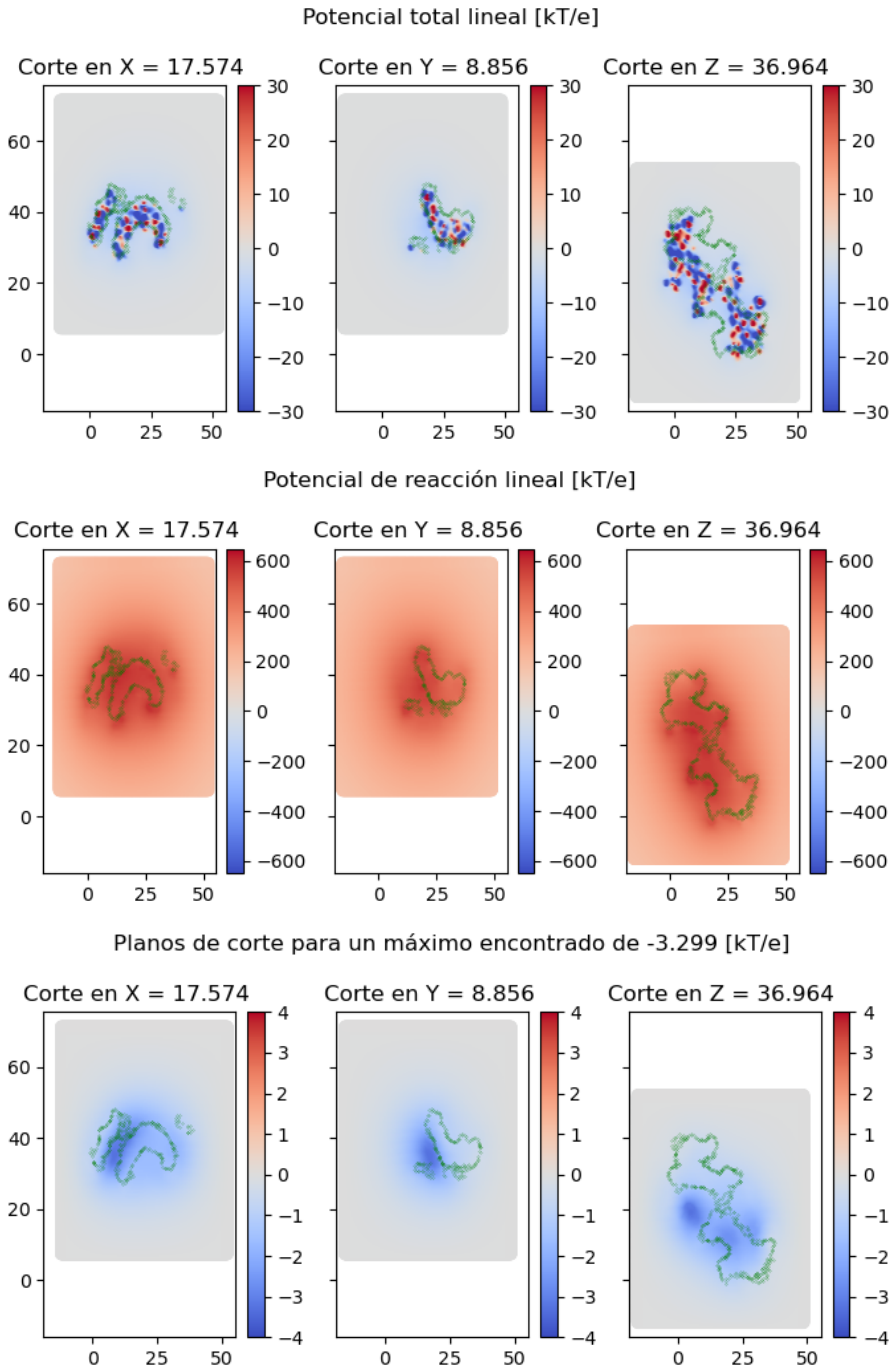

Figura 4.8: Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula 9ANT sin proteína anexada

Se realizan dos estudios con el objetivo de visualizar el efecto sobre los resultados de una molécula compuesta únicamente del ADN y otro con la proteína anexada que contiene normalmente. Para ello se eliminan manualmente la proteína del archivo PDB, obteniendo un comportamiento bastante acotado en torno a la curva exponencial modelo, versus el caso con cuerpo estructural que el potencial de Coulomb y el potencial de reacción tienen una dispersión mayor de los datos.

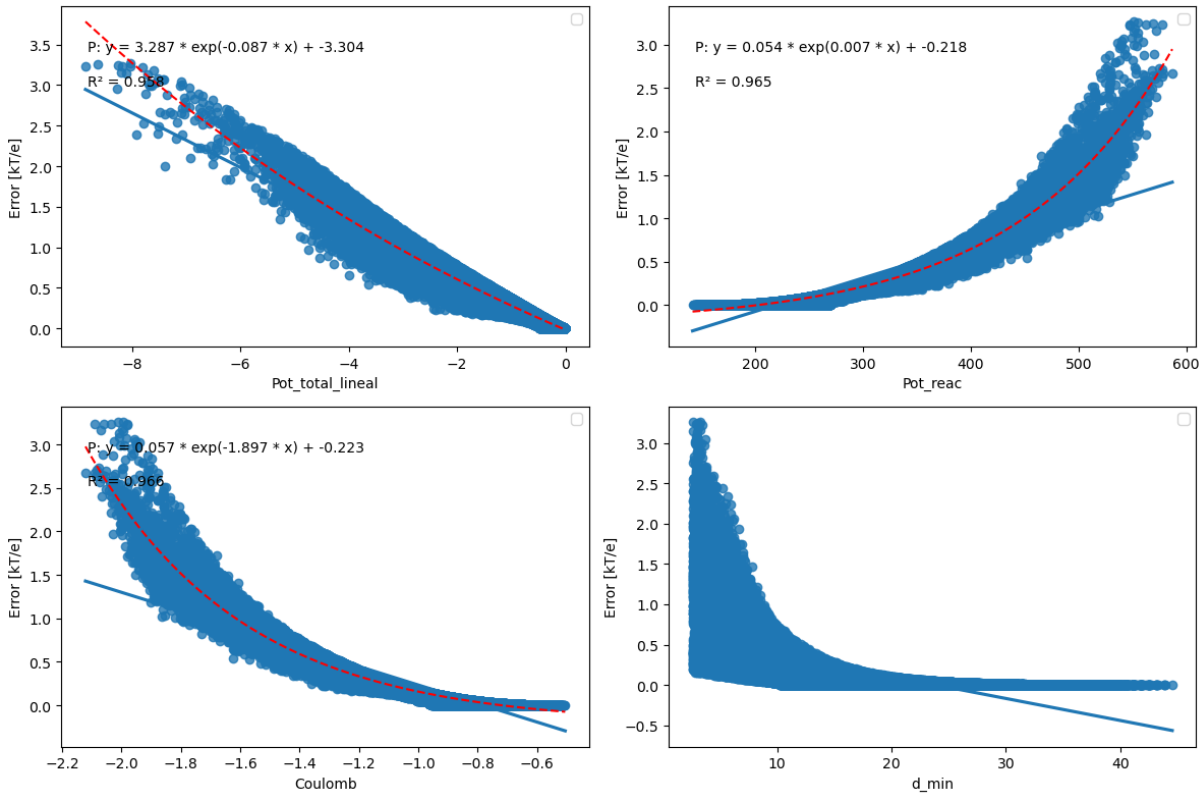

Figura 4.9: Error lineal no lineal, Comportamiento de 100.000 datos, 9ANT sin proteína anexada

4.1.3.2. Con proteína anexada

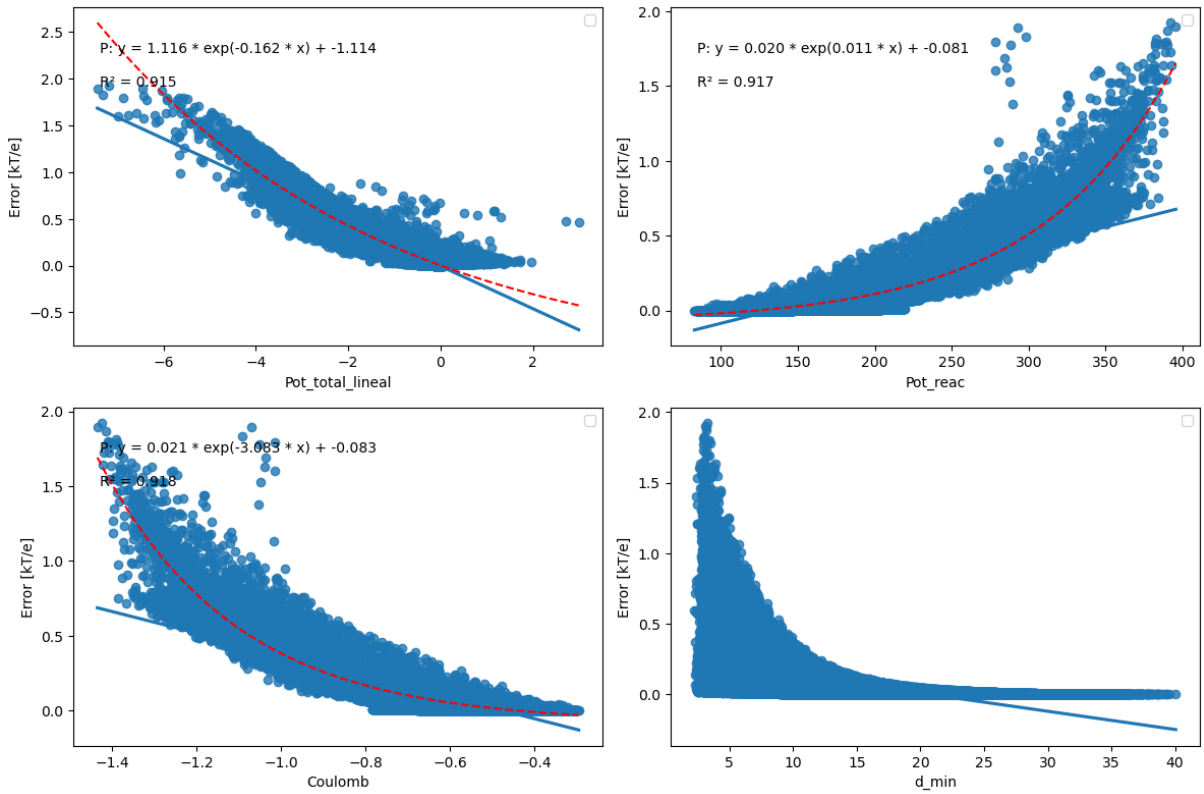

Figura 4.10: Error lineal no lineal, Comportamiento de 100.000 datos, 9ANT con proteína anexada

4.1.4. 437D

Tal como se mostraba en la figura 3.4, la cadena de ADN parece envolverse a sí misma, generando como se muestra en los cortes mostrados en la figura 4.12, un foco de error en el área contenida. Si bien se presencian los efectos sobre el interfaz externo de la molécula, los focos de error se dan en la zona interna, en donde existe un pico de la estructura. Los datos medidos muestran una buena correlación, figura 4.13 para las tres variables estudiadas.

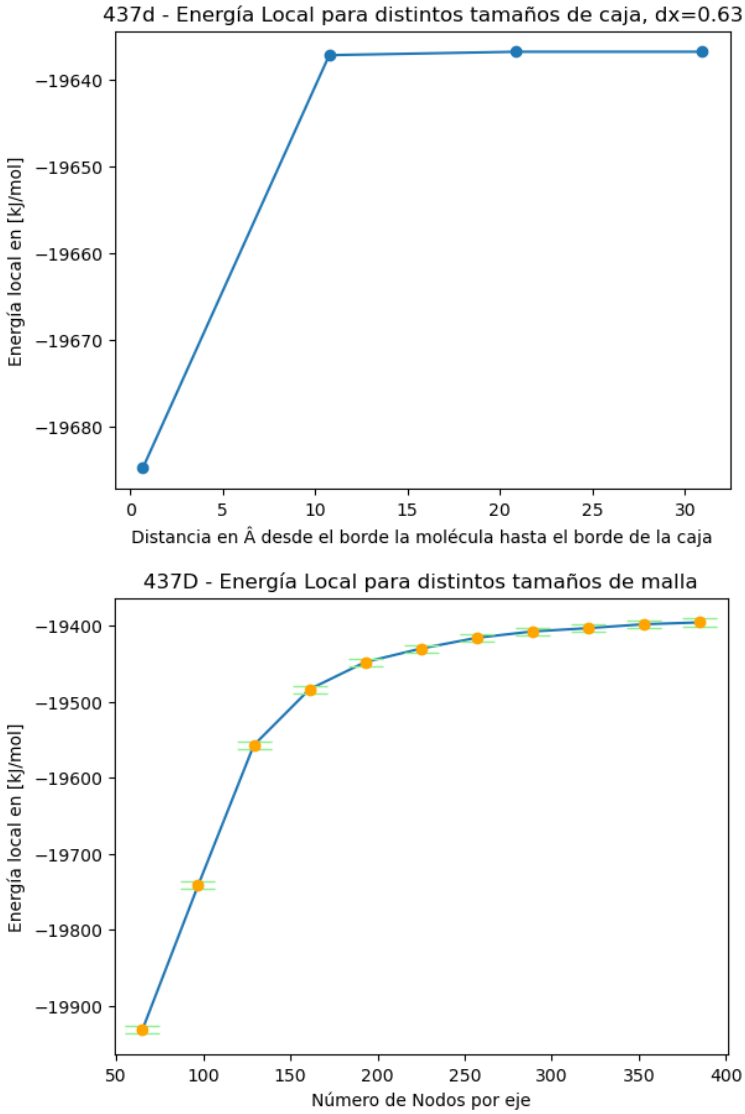

Figura 4.11: Análisis de sensibilidad para la molécula 437D

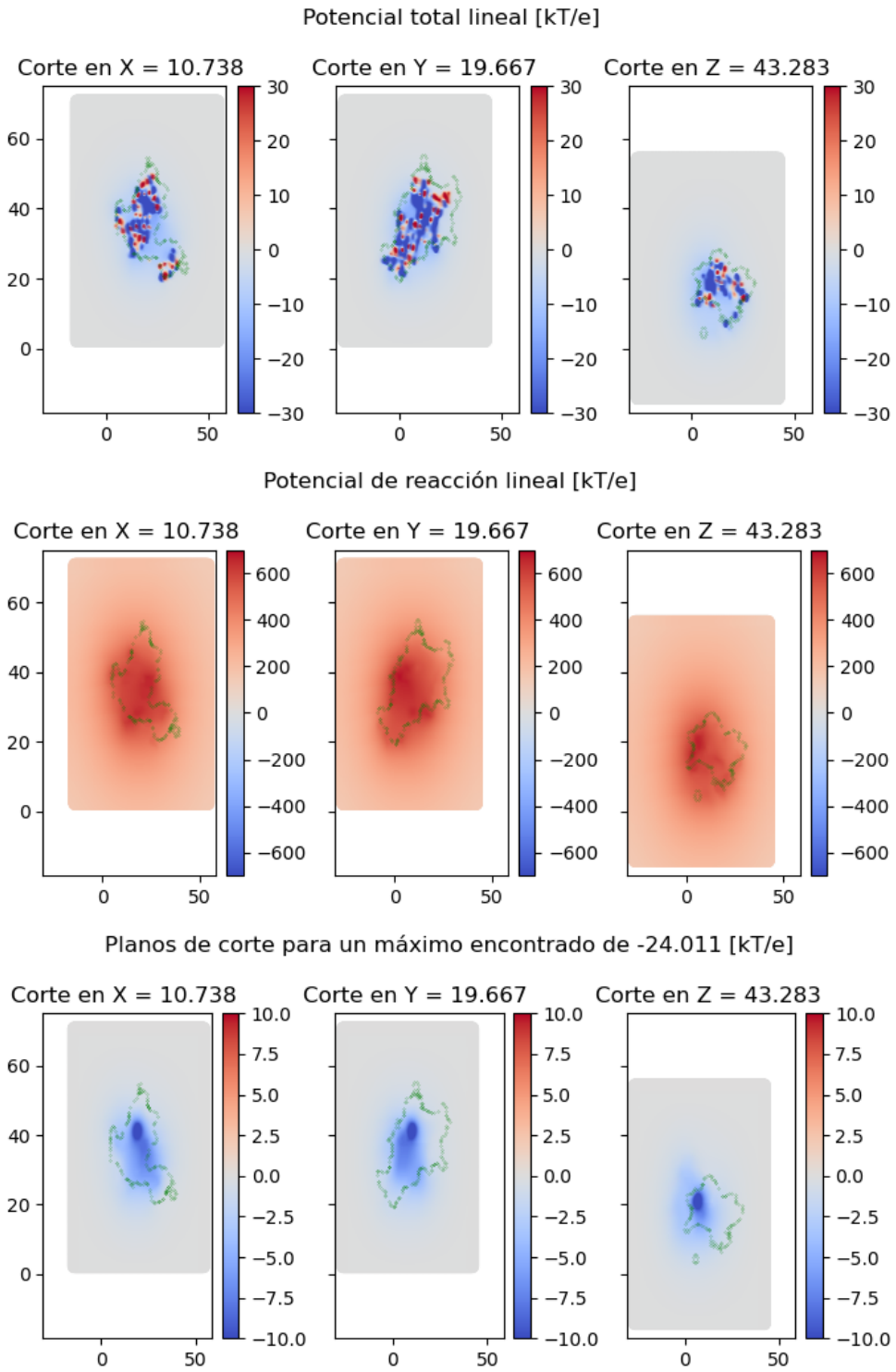

Figura 4.12: Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula 437D

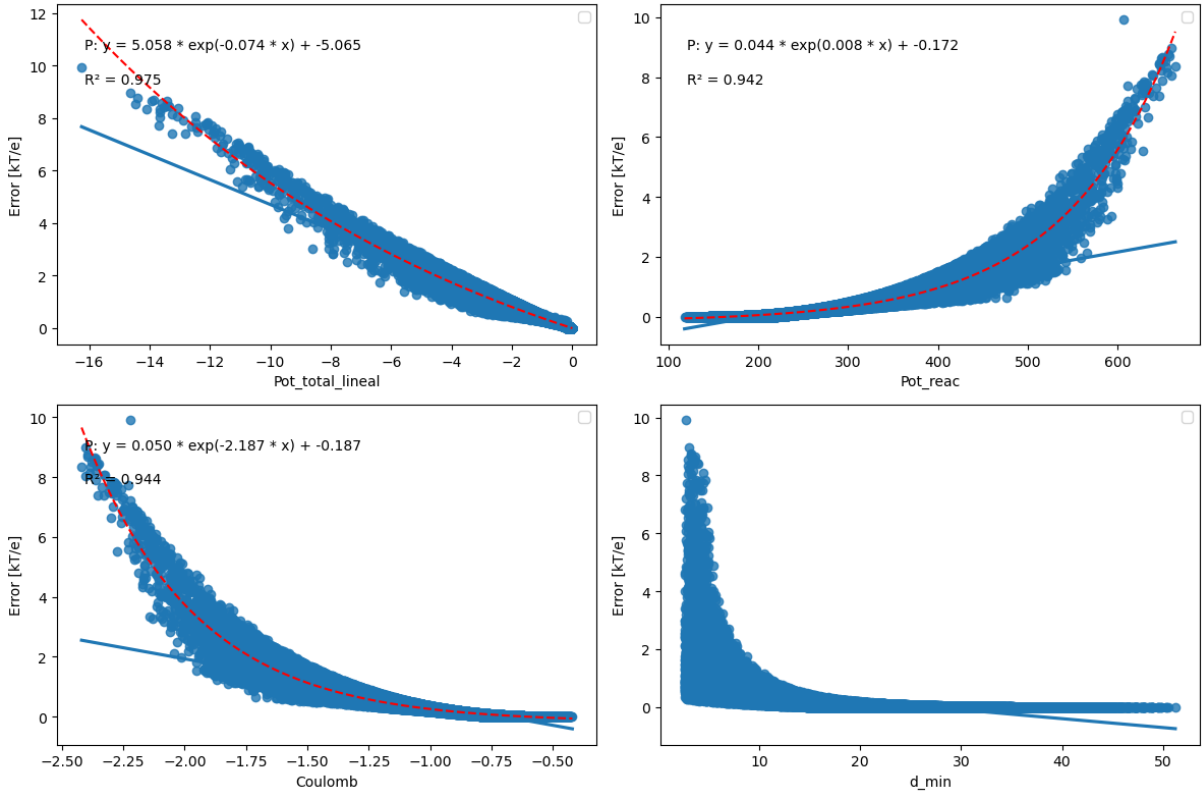

Figura 4.13: Error lineal no lineal, Comportamiento de 100.000 datos, 437D

4.1.5. Análisis de curvas características del comportamiento de 100.000 datos

La figura 4.14 nos muestra las curvas características obtenidas de los análisis previos de todas las moléculas de alta carga energética. En ellas podemos aislar un subgrupo de moléculas, que corresponden a aquellas que solo están compuestas por nucleótidos (280D, 9ANT-SoloADN, 437D) y otro grupo que tiene una proteína indexada (1KX5 y 9ANT). El primero de estos grupos nos hace suponer una dependencia cóncava de los datos, del cual es posible definir un modelo que nos permitiría estimar el error de la no linealidad para estructuras compuestas únicamente por nucleótidos.

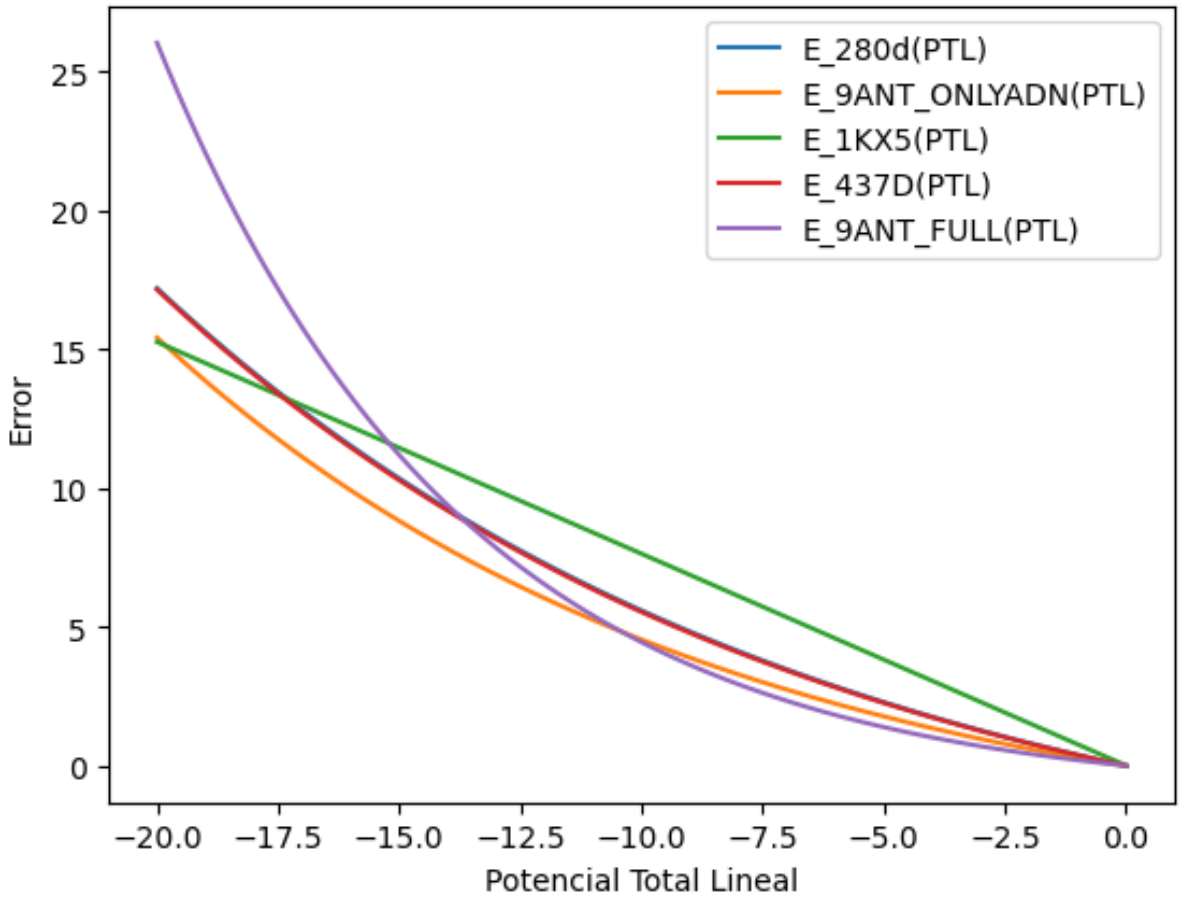

Figura 4.14: Valor absoluto del Error lineal - no lineal versus el potencial total lineal.

Del segundo análisis relacionado al potencial de reacción lineal, figura 4.15, no existe un modelo que se corresponda en esta batería de moléculas, más bien existe una tendencia clara de las curvas que nos permitirían normalizar en función de alguna variable que no se definido con éxito. Se dieron casos de normalización en función de error máximo registrado por molécula, e incluso por nivel energético, dejando abierto el análisis.

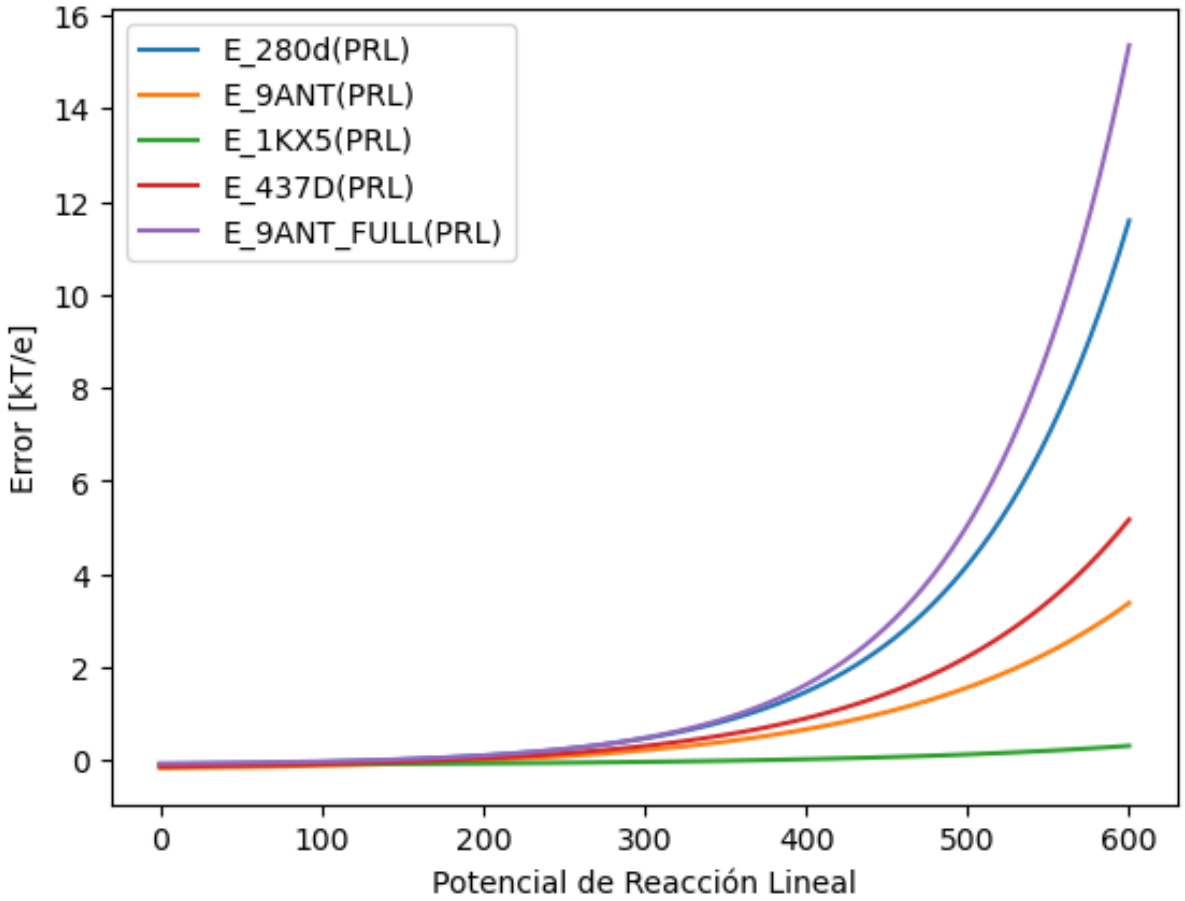

Figura 4.15: Valor absoluto del Error lineal - no lineal versus el potencial de reacción lineal.

Mismo caso es dado para el análisis del error lineal no lineal con el potencial de Coulomb, figura 4.16, en el que los intentos de normalización no fueron fructíferos.

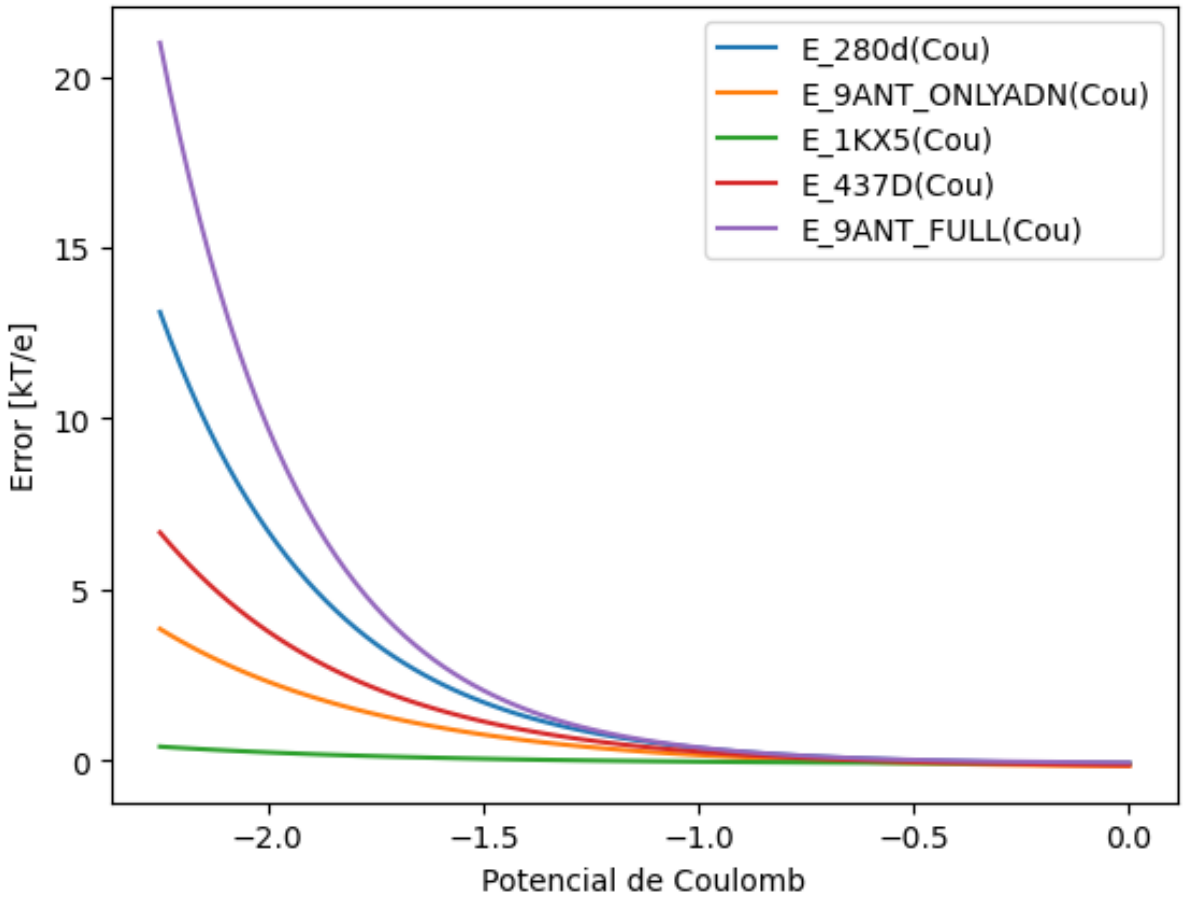

Figura 4.16: Valor absoluto del Error lineal - no lineal versus el potencial de Coulomb

De todos modos las curvas obtenidas si eran representativas de los datos para cada molécula en particular, quizás el enfoque que se debe tomar tiene que ser la selección adecuada de una familia de moléculas, por ejemplo; Todas aquellas moléculas compuestas por una cadena de ADN/ARN y una proteína indexada en uno de sus lados, o también, todas aquellas moléculas en las que la proteína recubre el ADN/ARN, etc.

4.1.6. Modelo del Error lineal no lineal en función del potencial total lineal

Como se mencionó con anterioridad, el subgrupo conformado por moléculas que en su estructura corresponden únicamente a nucleótidos, presentan un comportamiento similar, es por ello que se efectúa un modelo promediando las 3 curvas obtenidas. La hipótesis del modelo que estamos planteando dice que como las tres moléculas están únicamente compuestas por nucleótidos, el comportamiento debería replicarse para el resto.

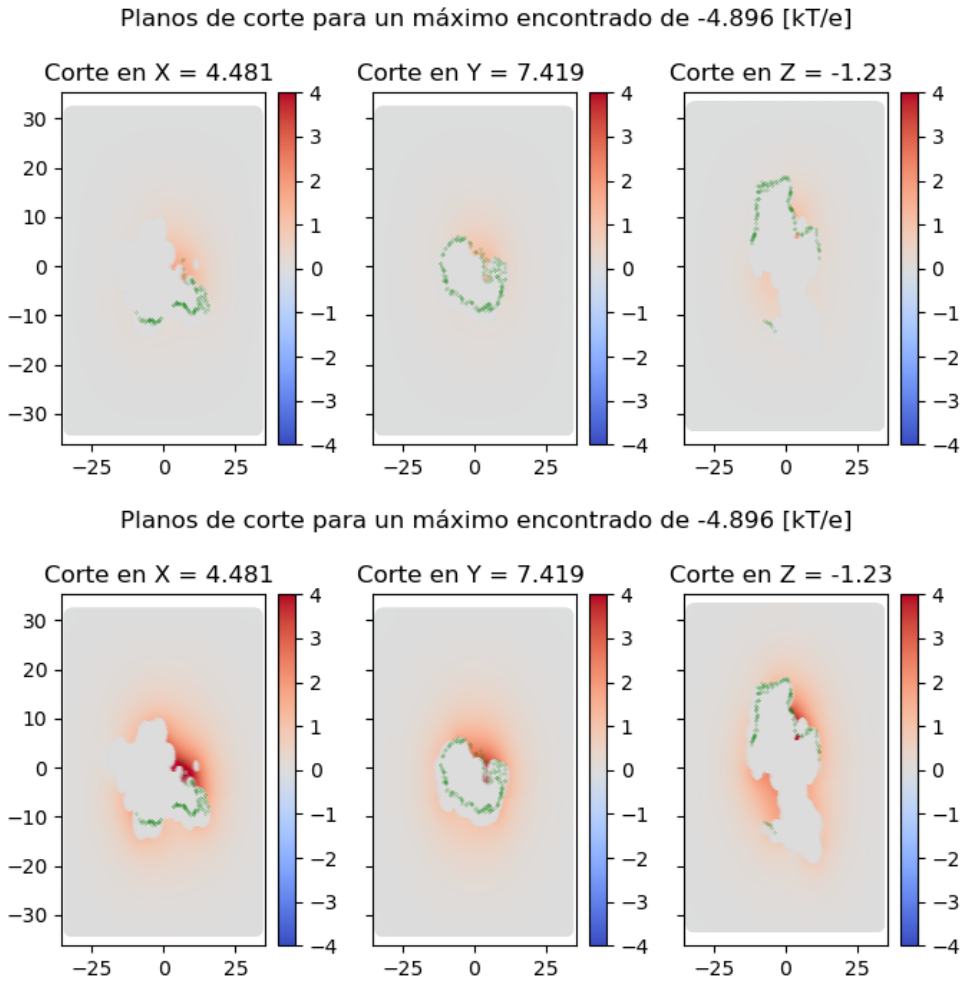

Figura 4.17: Figura superior, Corte del caso lineal no lineal calculado de la diferencia del Potencial total lineal menos el potencial total no lineal por APBS. Figura inferior, Modelo del error lineal no lineal dependiente del Potencial total lineal calculado por APBS. 1AJF

Para probar la hipótesis se tomaron dos moléculas aleatorias del RCSB PDB que tuvieran en su composición la regla estipulada. Las elegidas fueron la 1AJF, que tenía un nivel energético cercano a los 8.000 [kJ/mol], y la otra elegida fue la 1A4B con un nivel energético de 33.000 [kJ/mol]. Ellas nos entregan un buen contraste.

Los resultados se ven en las figuras 4.17 y 4.18, en donde se observa que el modelo descrito responde de buena manera al error lineal no lineal, aunque exagera estos errores. Desde el punto de vista de la investigación, esto no es malo, pues entrega una herramienta que permite identificar zonas de posible conflicto, en las que se hace necesario, o un mallado fino, o un cálculo lineal no lineal por zonas.

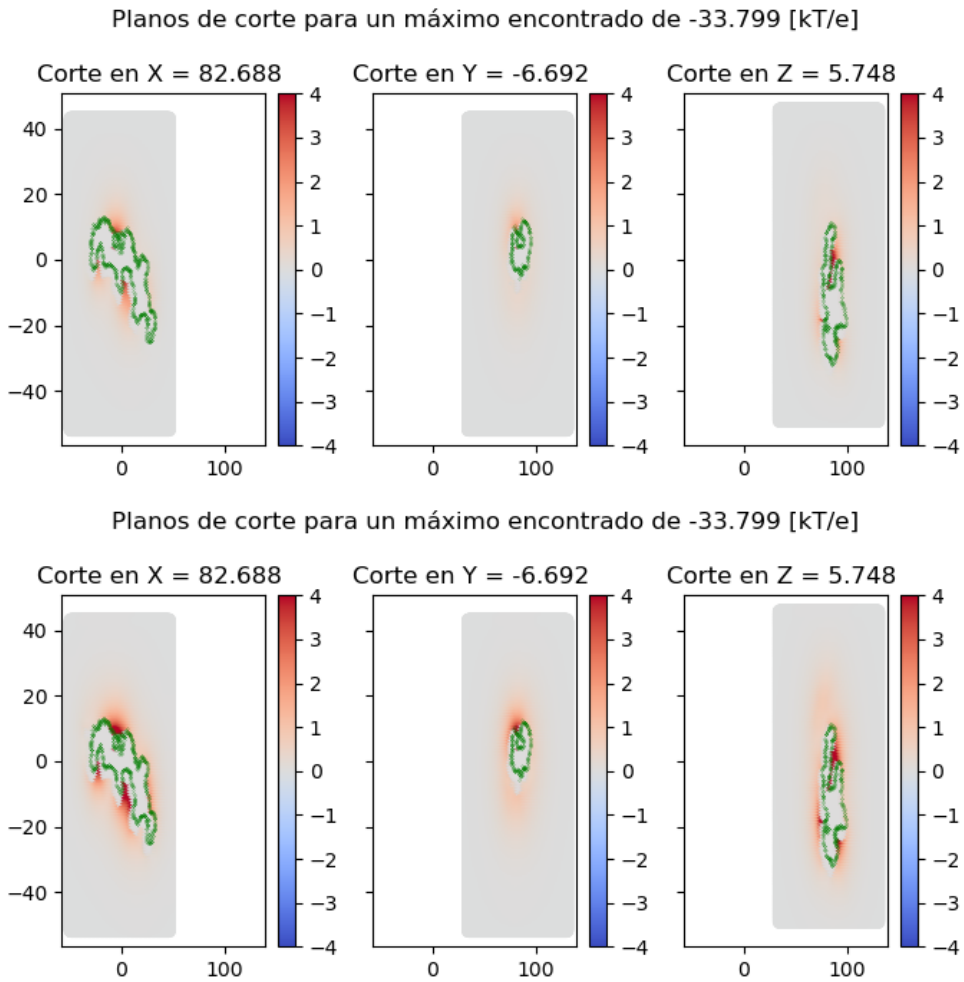

Figura 4.18: Figura superior, Corte del caso lineal no lineal calculado de la diferencia del Potencial total lineal menos el potencial total no lineal por APBS. Figura inferior, Modelo del error lineal no lineal dependiente del Potencial total lineal calculado por APBS. 1A4D

4.2. Batería medianamente cargadas

4.2.1. 1LYZ

Uno de los primeros elementos que se aprecia al poner en perspectiva la batería medianamente cargada con la altamente cargada, es el requerimiento del análisis de sensibilidad, que resultaba mucho menor, en términos del *dime* requerido y el *dx* solicitado. Al comparar los ordenes de magnitud de la figura 4.19 y las figuras en general de esta batería de moléculas versus aquellas de la batería altamente cargada, estamos hablando de cálculos de energía por local de un orden de magnitud menor.

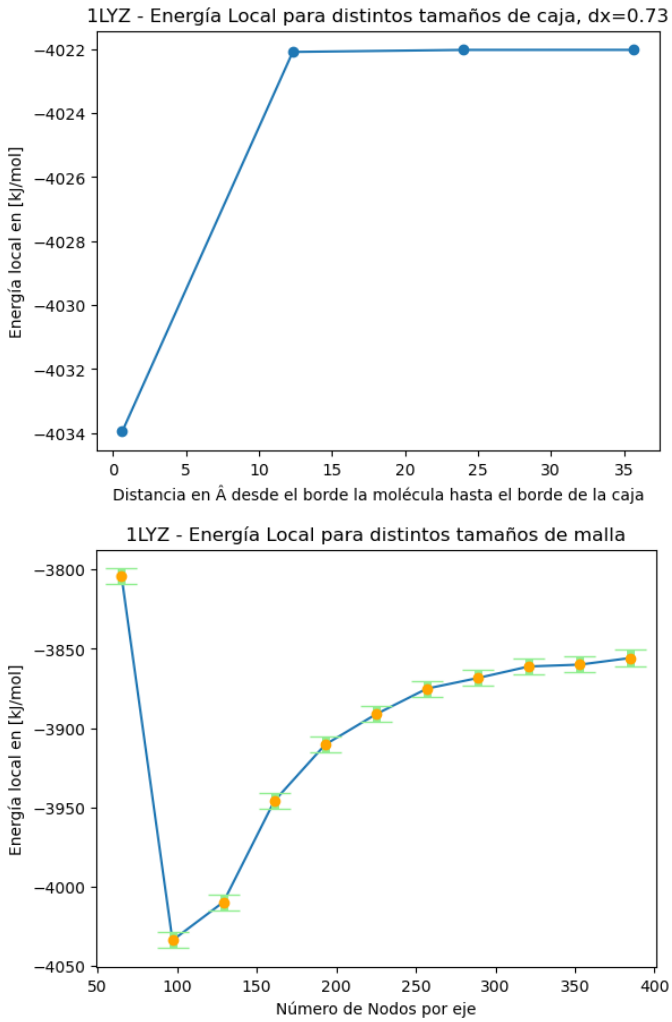

Figura 4.19: Análisis de sensibilidad para la molécula 1LYZ

Cuando observamos el error lineal no lineal en moléculas de esta batería, figura 4.20, se hace énfasis en un elemento que podría pasar desapercibido en moléculas más grandes debido a su complejidad, pero que son evidentes, y que son definidos como *focos de error*, en particular la molécula 1LYZ presenta sus dos focos de error en zonas que parecen estar parcialmente rodeadas o contenidas por la misma estructura.

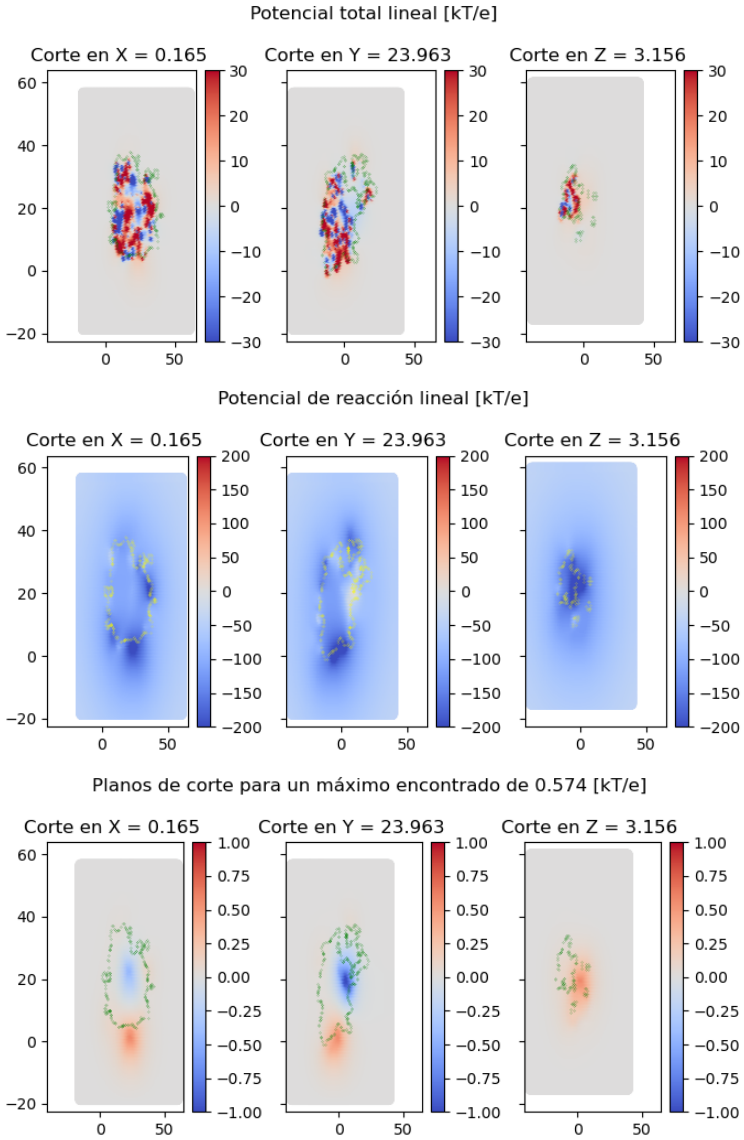

Figura 4.20: Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula 1LYZ

Por otro lado, existe una dispersión de los datos en el comportamiento de la toma de muestras, figura 4.21, que si bien muestra una tendencia en una dirección, en la otra se difumina.

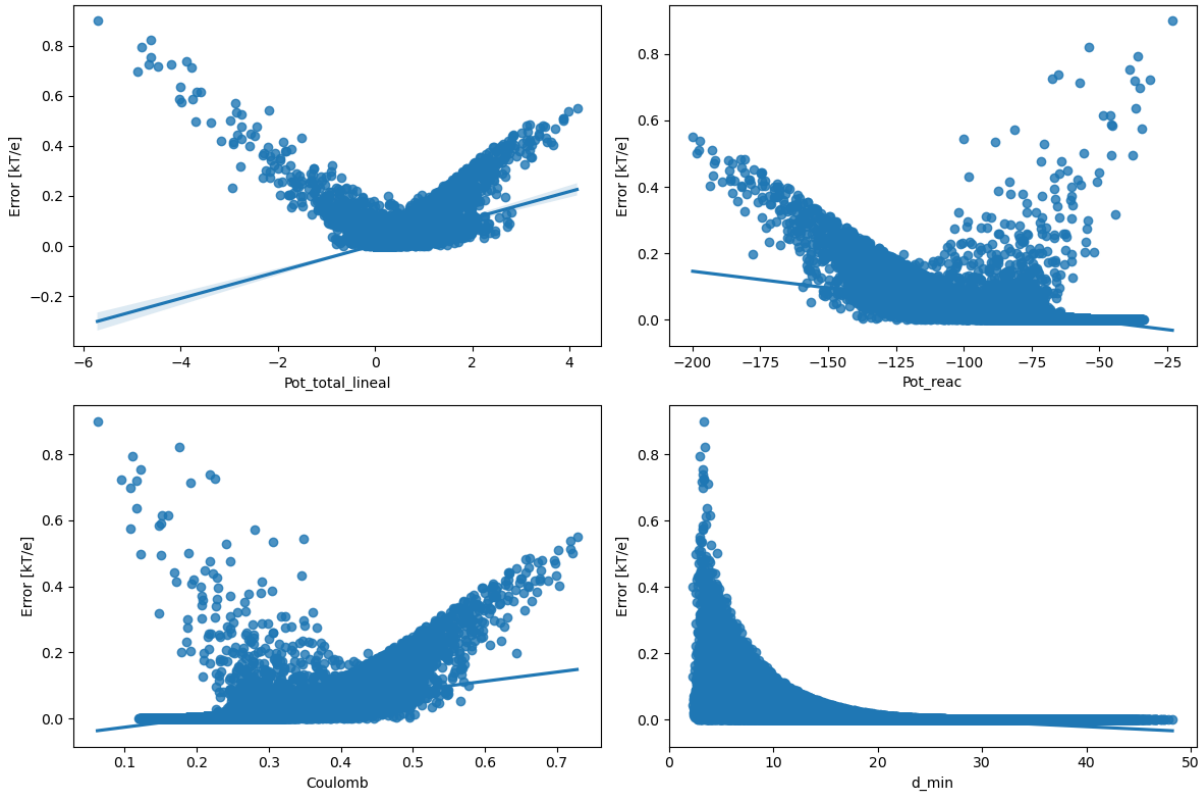

Figura 4.21: Error lineal no lineal, Comportamiento de 100.000 datos, 1LYZ

4.2.2. 1UBQ

En el corte realizado para el máximo error en la molécula 1UBQ, figura 4.23, se referencia por medio de un triángulo el lugar de interés, con el objetivo de remarcar la idea anterior sobre la cobertura de un punto. Si notamos el corte en el plano y, y el plano z, notamos que el triángulo está dentro de una zona cóncava, este análisis nos servirá luego en el estudio de *grado de concavidad*.

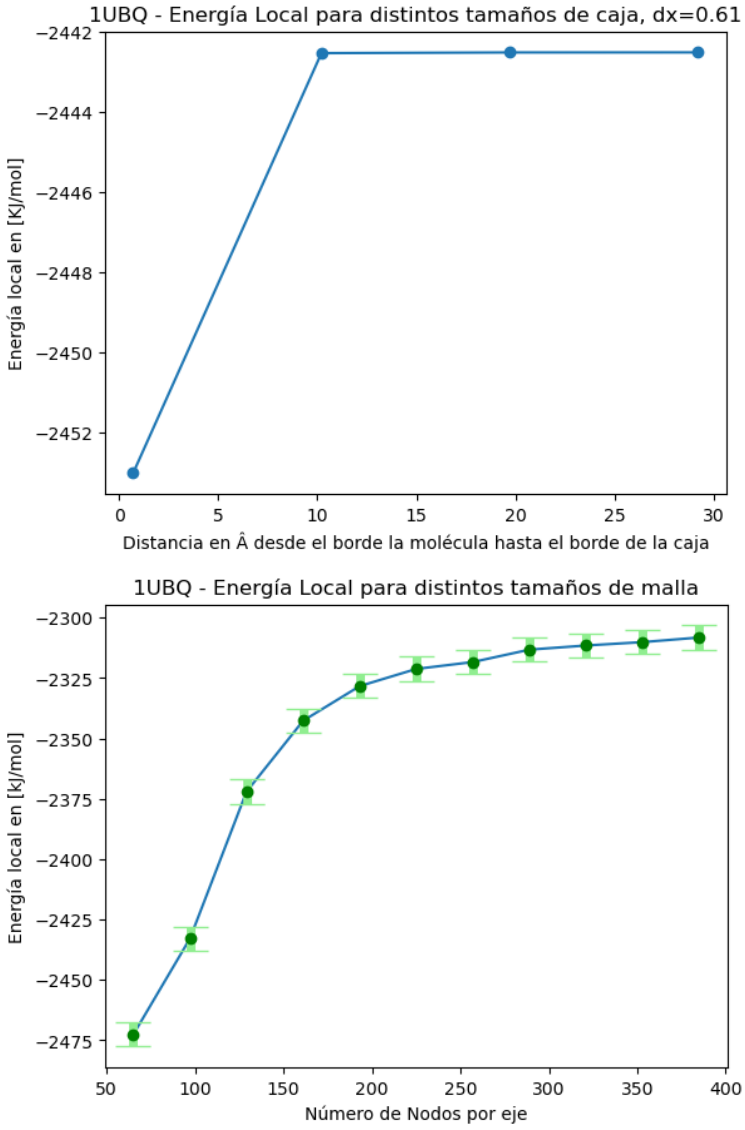

Figura 4.22: Análisis de sensibilidad para la molécula 1UBQ

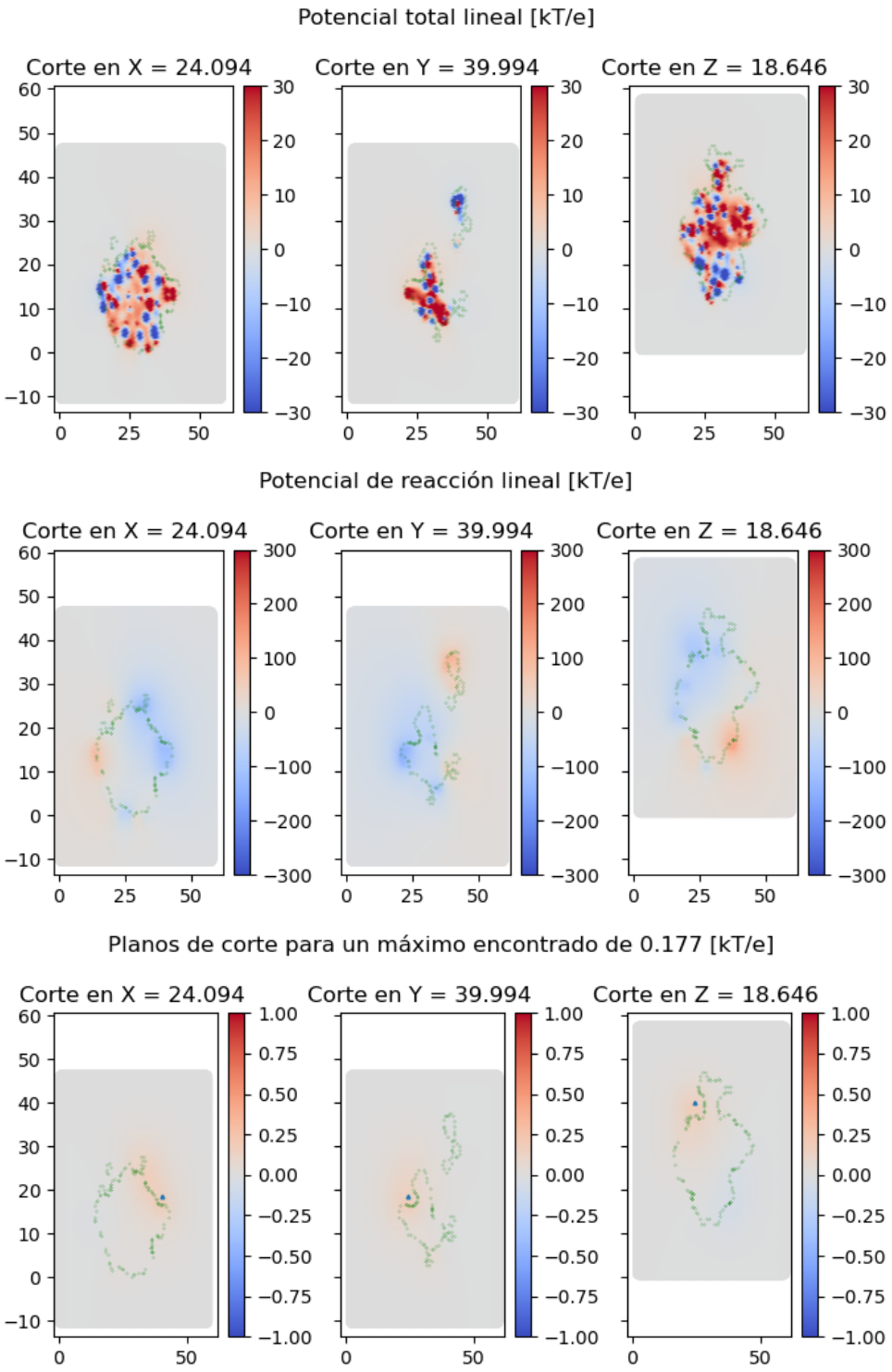

Figura 4.23: Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula 1UBQ

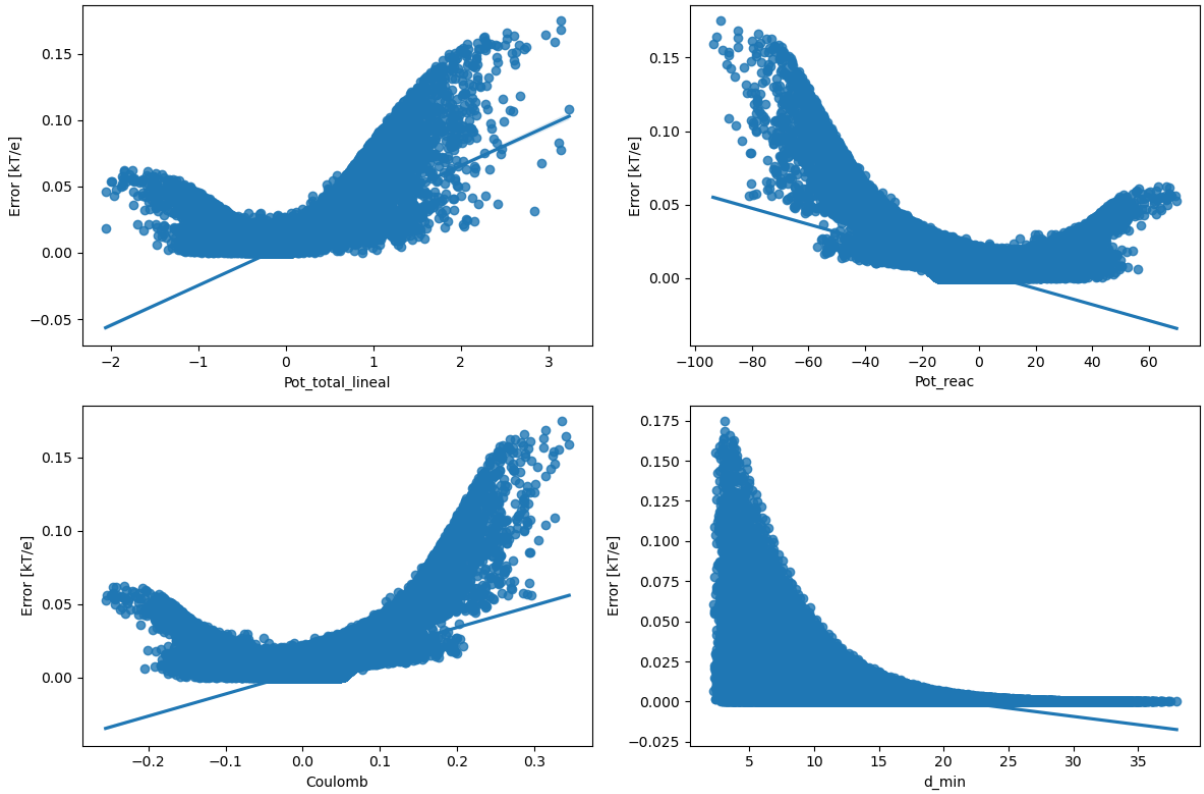

Figura 4.24: Error lineal no lineal, Comportamiento de 100.000 datos, 1UBQ

4.2.3. 1PGB

Finalmente para terminar la batería de moléculas medianamente cargadas, la 1PGB que corresponde a la inmunoglobulina. En donde observamos de la figura 4.26, el error se presenta en un trozo de interfaz en el que las cargas puntuales denotan cargas negativas rodeadas de un manto potencial en el interior de la molécula debido a las cargas puntuales positivas que contiene la proteína en el resto de la estructura.

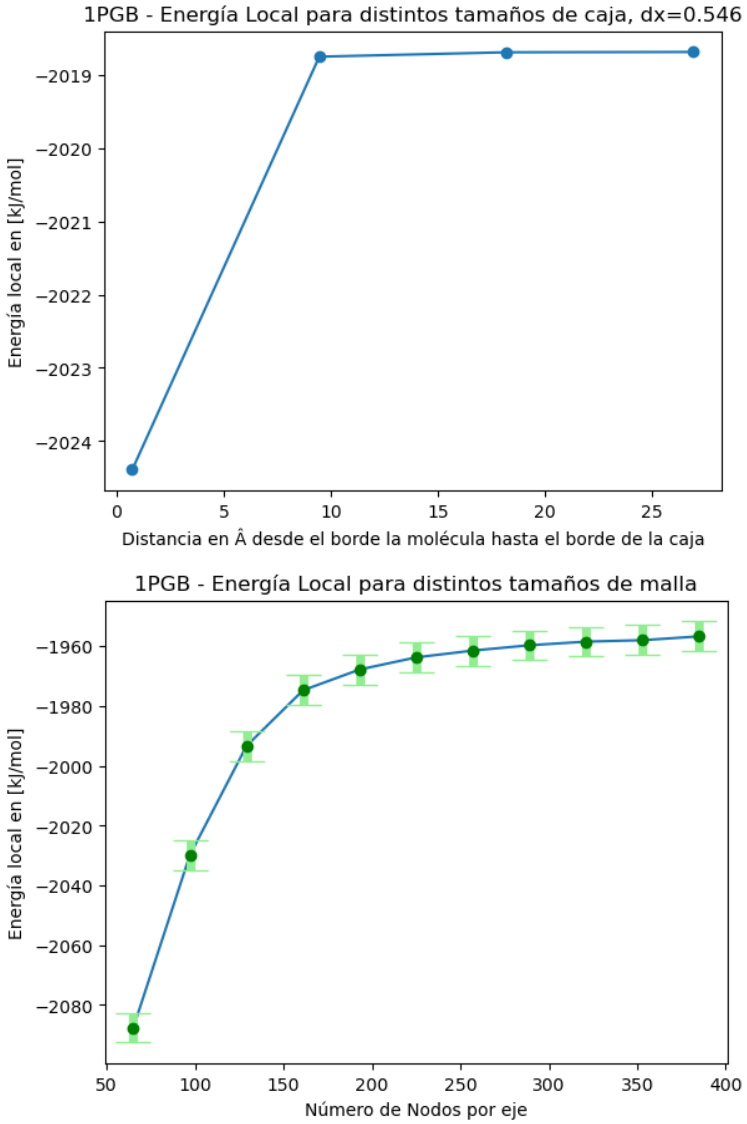

Figura 4.25: Análisis de sensibilidad para la molécula 1PGB

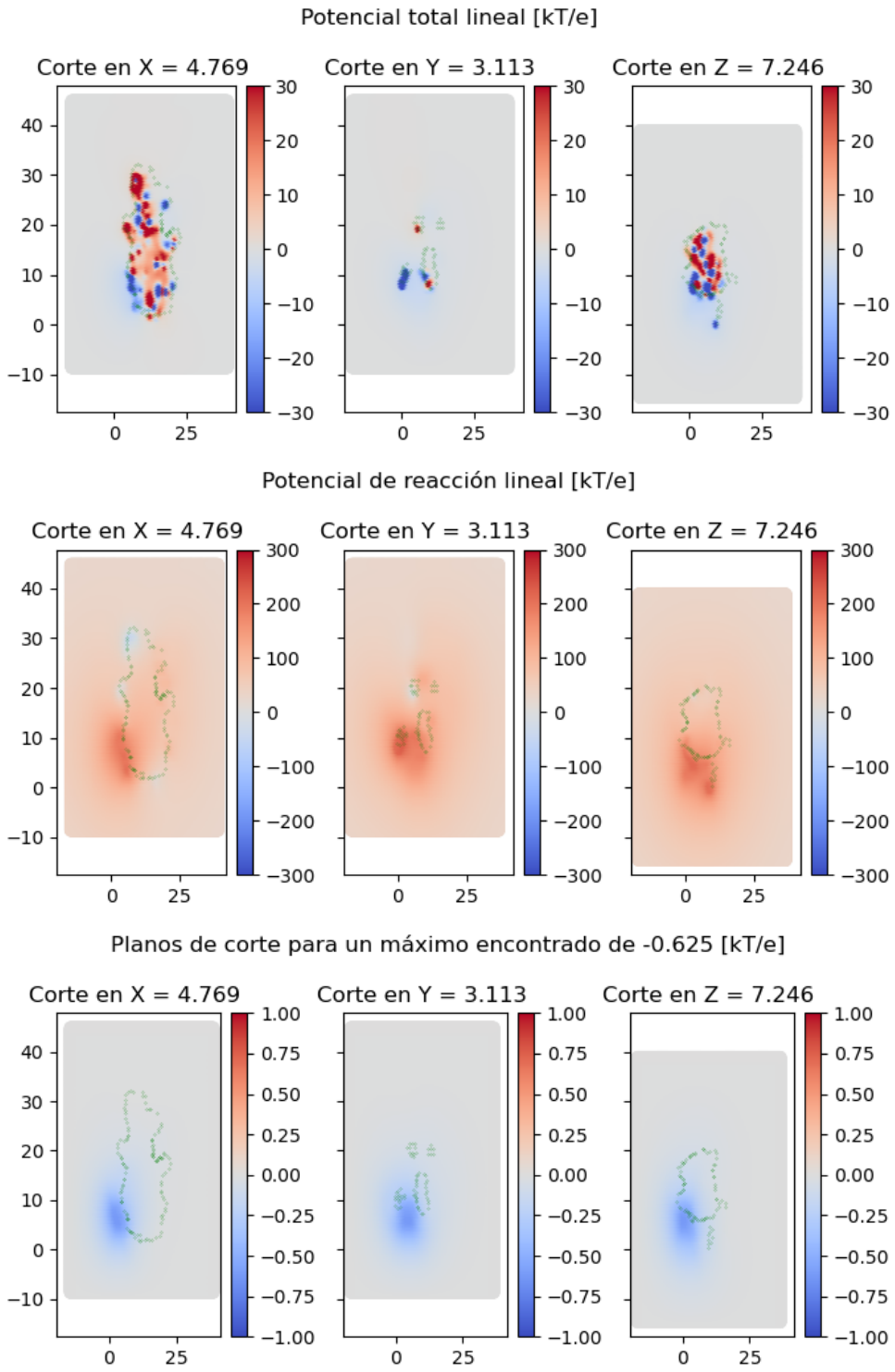

Figura 4.26: Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula 1PGB

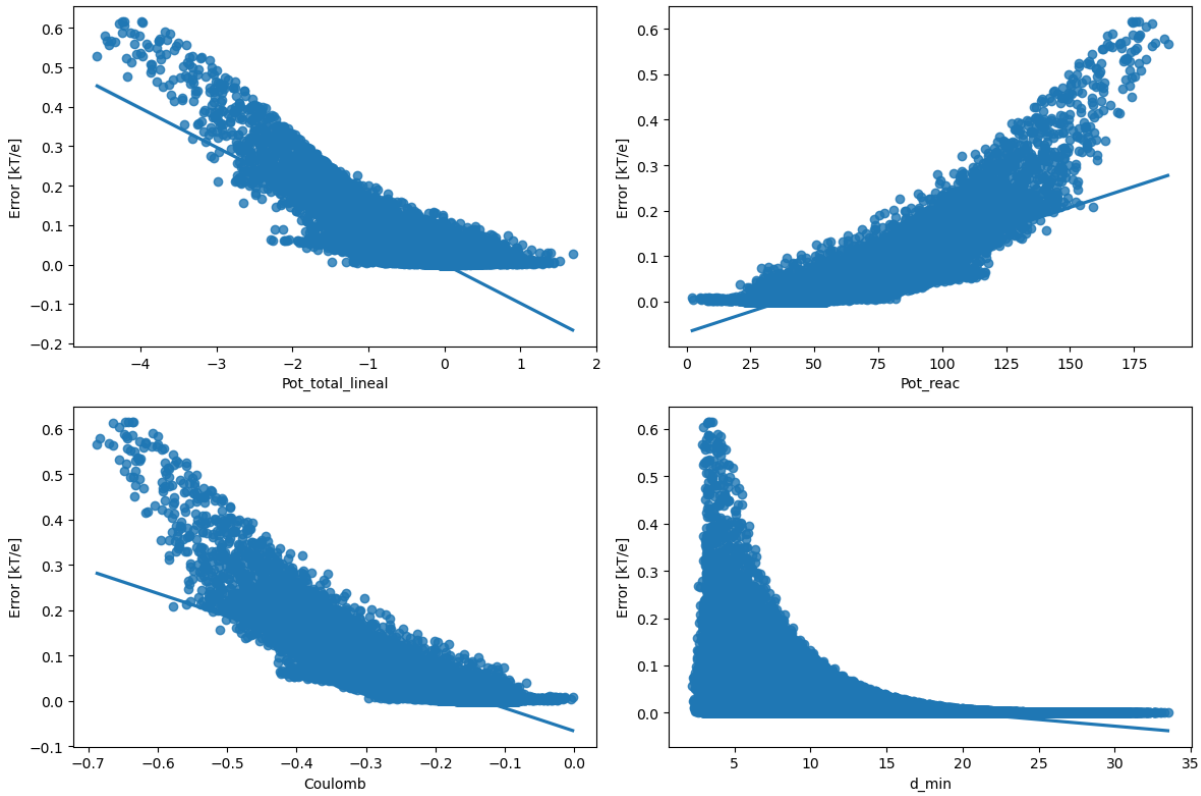

Figura 4.27: Error lineal no lineal, Comportamiento de 100.000 datos, 1PGB

4.2.4. Análisis del grado de cobertura

Otro de los análisis realizados se presenta durante el estudio de la molécula 1UBQ. Como se mencionó con anterioridad, se genera la necesidad de visualizar el punto crítico en relación a la cercanía con la molécula, dando como resultado el análisis progresivo de ambas baterías de moléculas mostradas hasta el momento, encontrando que lo errores se daban siempre en zonas donde existía curvatura de la molécula, es decir, zonas encerradas parcialmente.

Es por ello que se define una nueva variable de análisis denominada grado de cobertura, explicada previamente en metodología, la cual a grandes rasgos evalúa qué tantos diferenciales de área están cubiertos por alguna cara de la molécula a una distancia definida, tomando el valor de 1 para un punto rodeado completamente de caras de la molécula y 0 para un punto aislado, por ejemplo puntos en el contorno que en general están al menos a 15[A] de distancia.

El código es costoso computacionalmente, por lo que se tomó una muestra inicial de 500 datos. Los resultados se muestran en la figura 4.28 en donde parece tentativo relacionar que para grados de cobertura superiores a 0.3 una tendencia lineal.

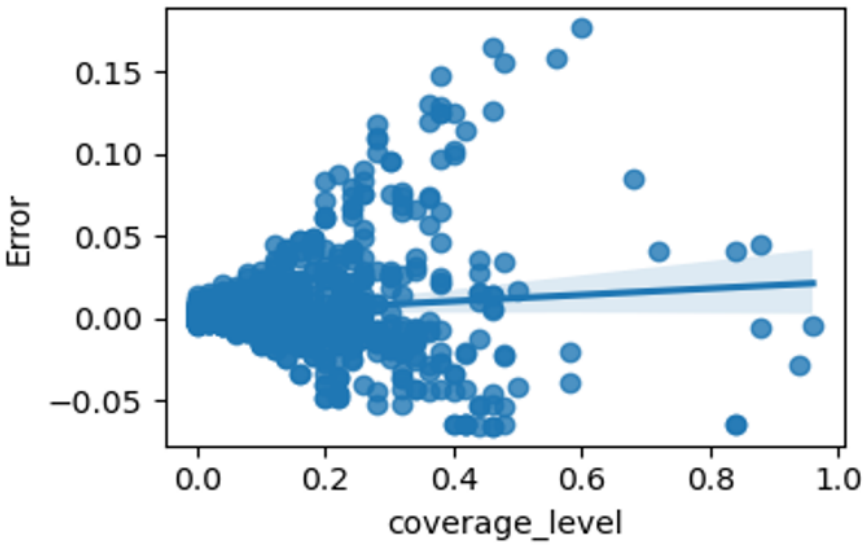

Figura 4.28: Análisis de cobertura para la molécula 1UBQ

También se hizo un ensayo para un grupo acotado de datos de la molécula 9ANT y los resultados fueron dispersos, como se observa en la figura 4.29, en donde los datos parecen estar distribuidos sobre dos líneas bases. Posteriormente el estudio se extendió a otras moléculas como la 280D y la 1LYZ, pero los resultados no fueron muy alentadores. Por lo que podemos decir en relación al grado de cobertura, que si bien, existe un tipo de dependencia, según se observan en los cortes estudiados , la forma propuesta para medirlo no es concluyente para establecer una norma o aproximación.

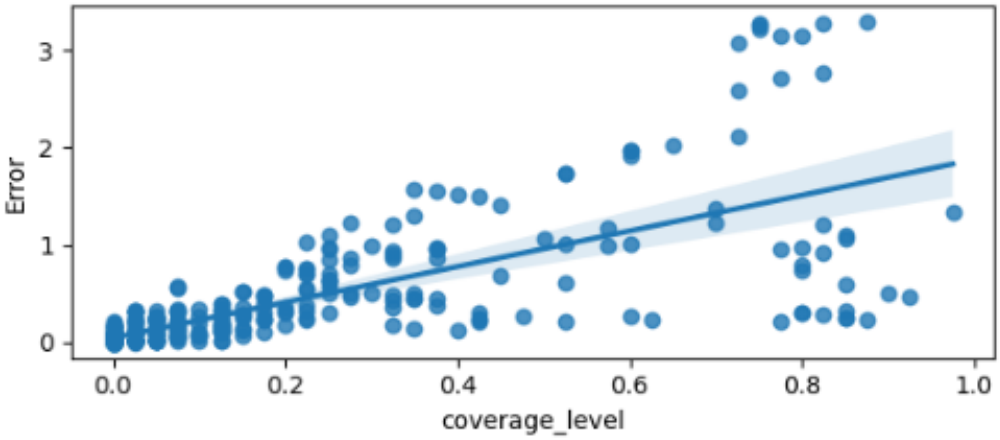

Figura 4.29: Análisis de cobertura para la molécula 9ANT

4.3. Batería de baja carga

En general las baterías de baja carga solo demostraron que con proteínas pequeñas el modelo lineal es imperceptible en comparación al no lineal, y es que el error máximo se lo llevó el etanol, figura 4.33, con $-0{,}001[kT/e]$ en un punto de una malla de más de 300.000 datos. A continuación se muestran los análisis de sensibilidad y cortes respectivos.

4.3.1. Anilina

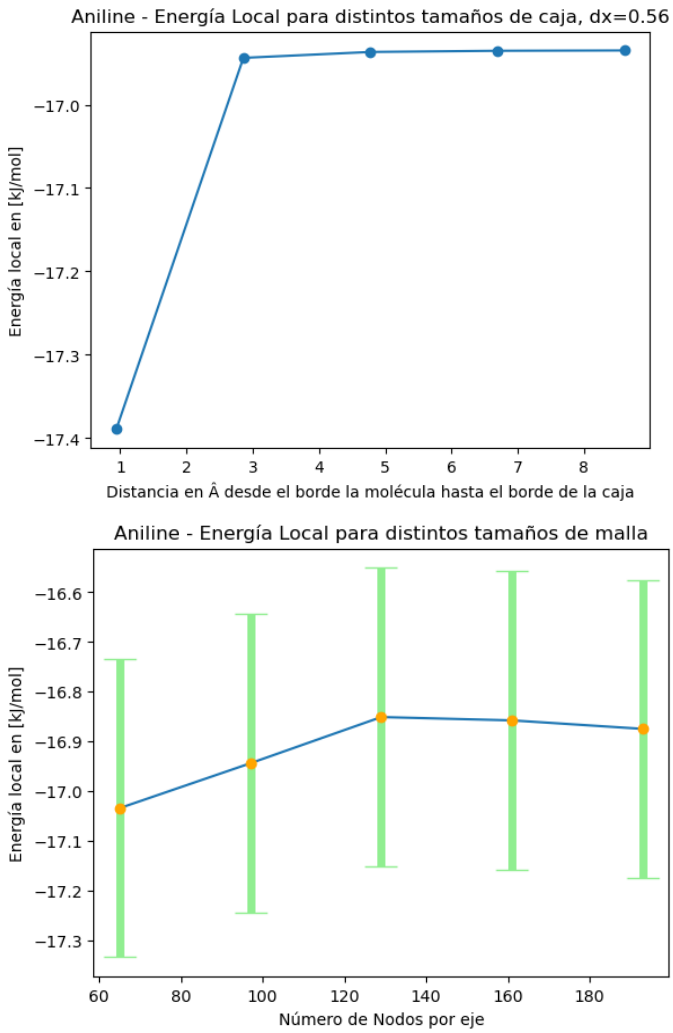

Figura 4.30: Análisis de sensibilidad para la molécula de Anilina

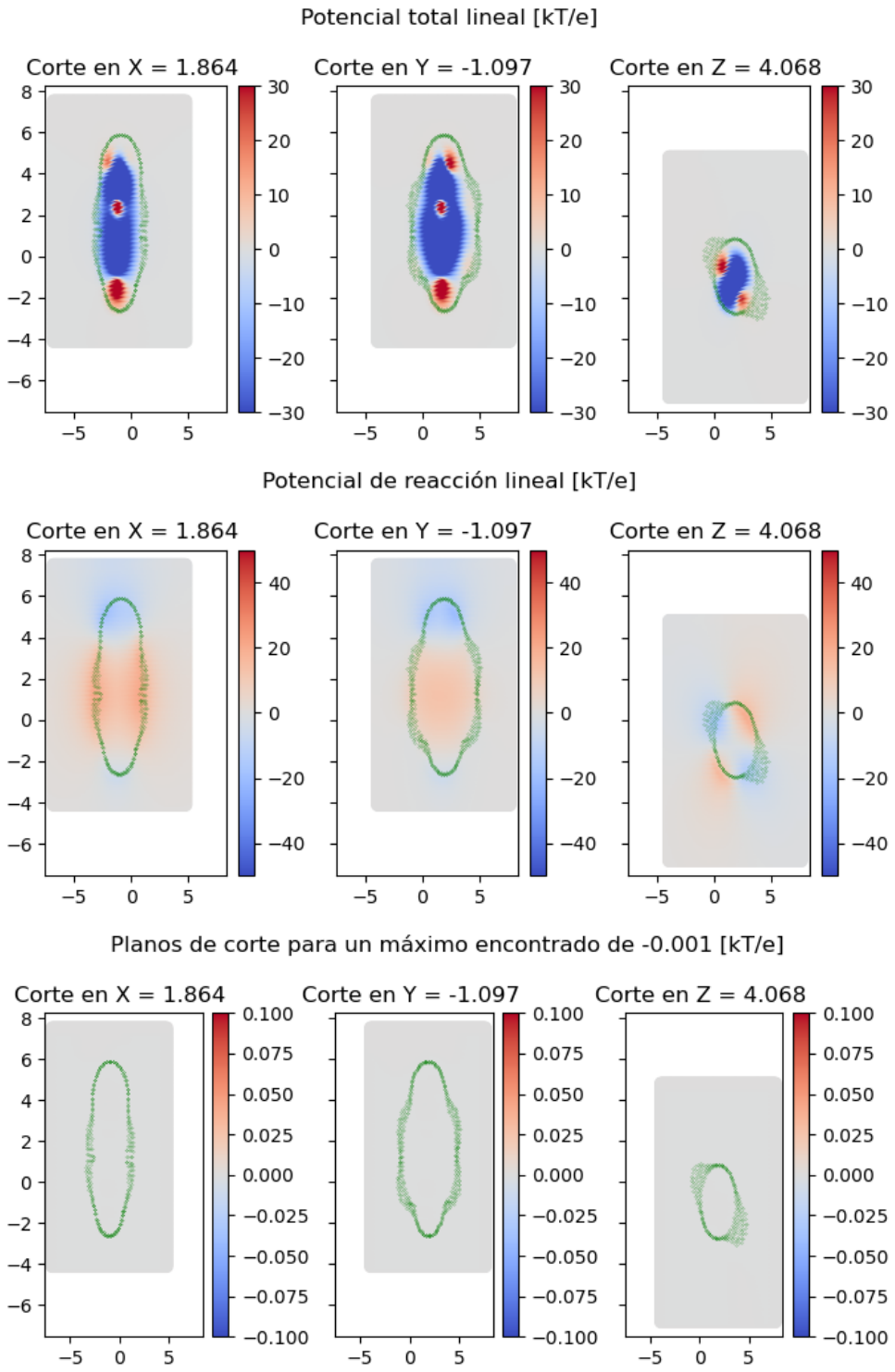

Figura 4.31: Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula de Anilina

4.3.2. Etanol

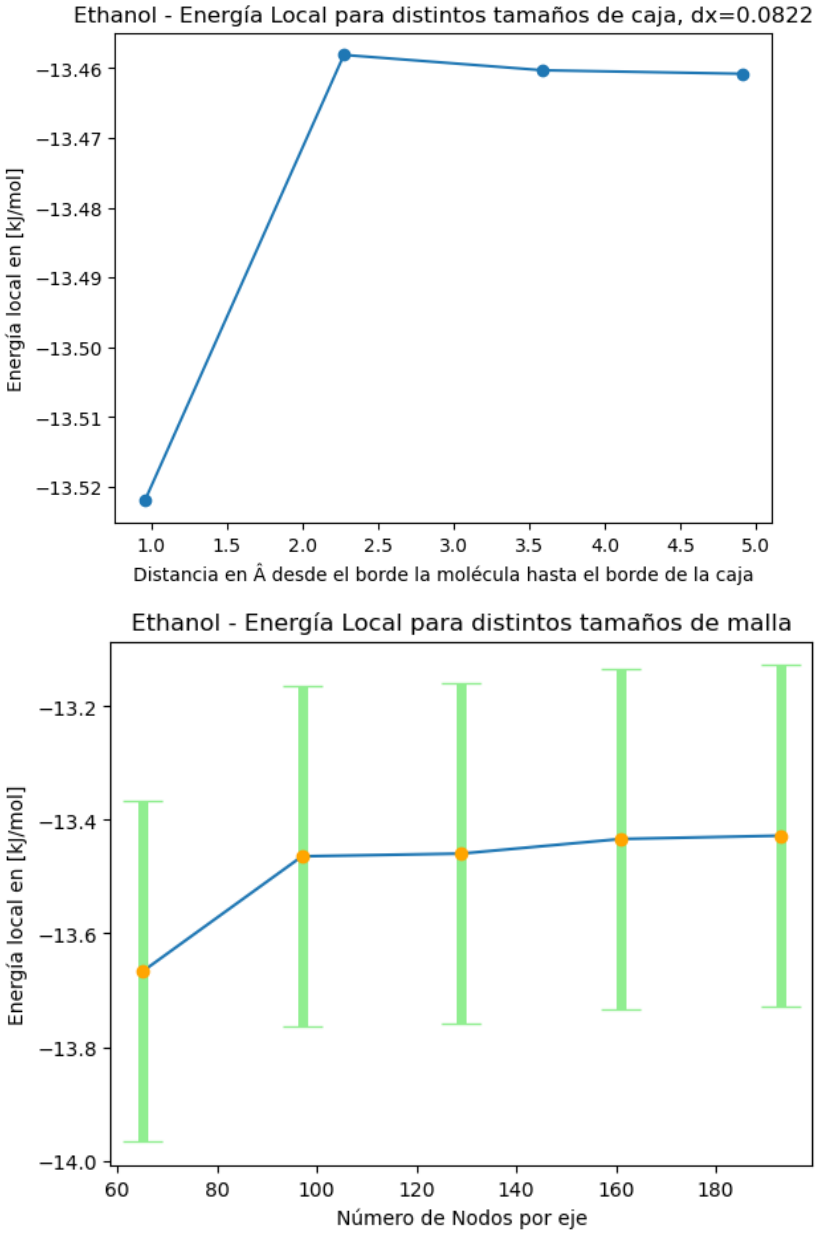

Figura 4.32: Análisis de sensibilidad para la molécula de Etanol

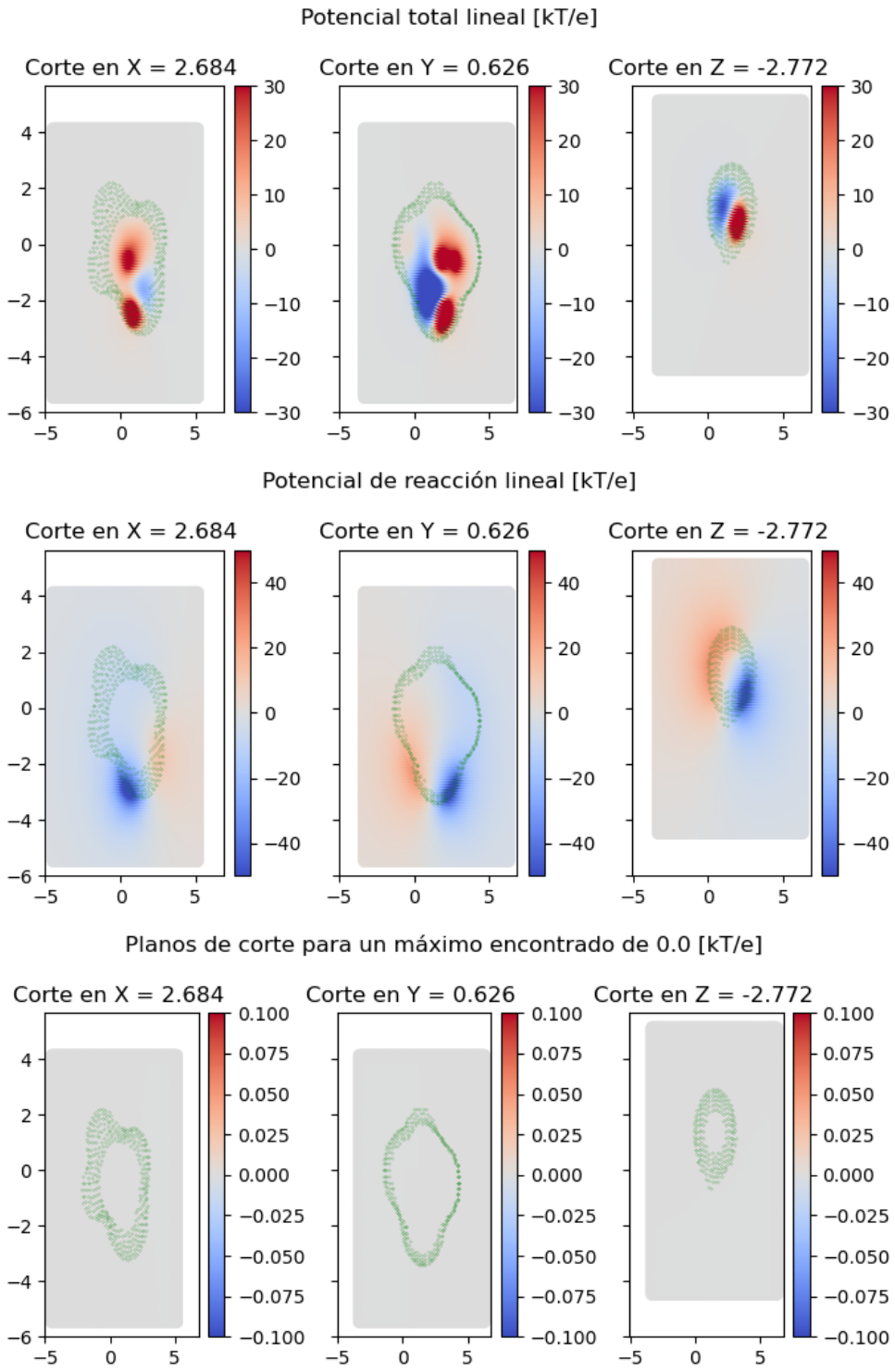

Figura 4.33: Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula de Etanol

4.3.3. Tolueno

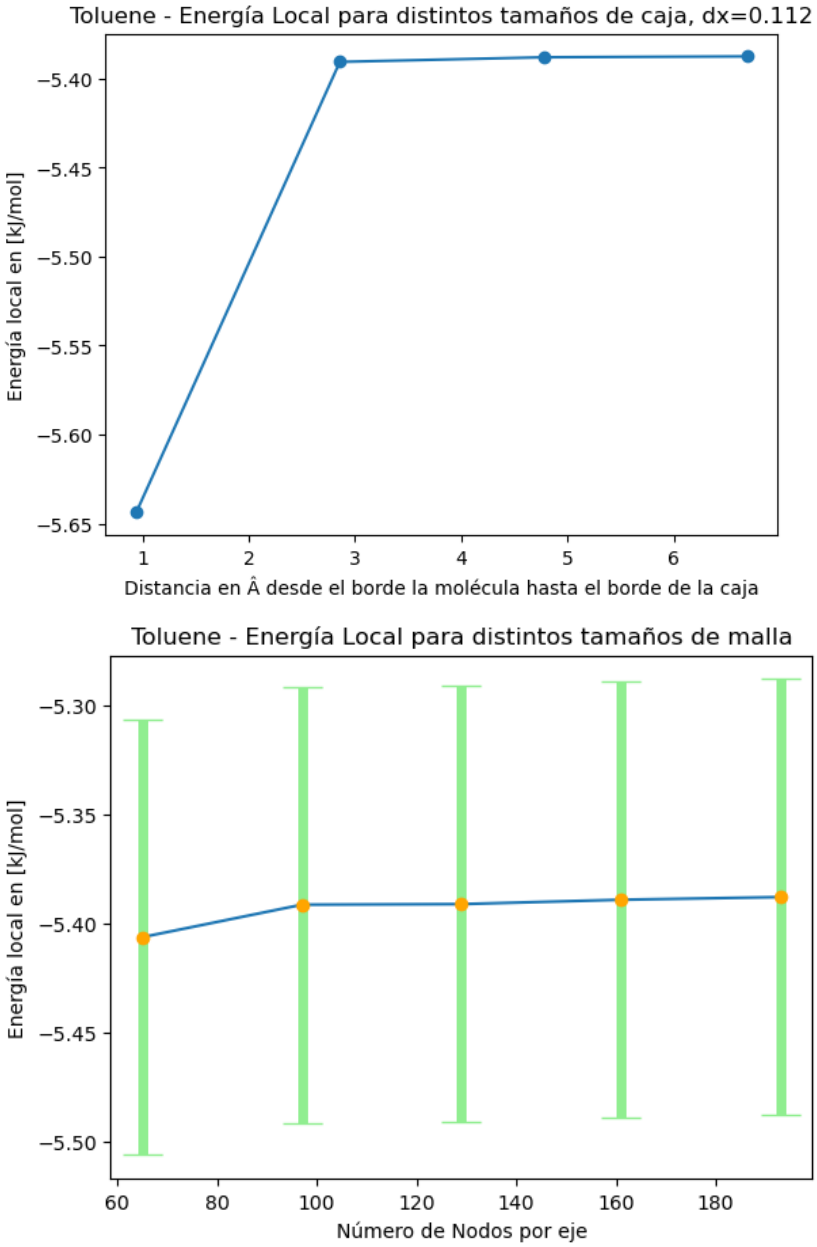

Figura 4.34: Análisis de sensibilidad para la molécula de Tolueno

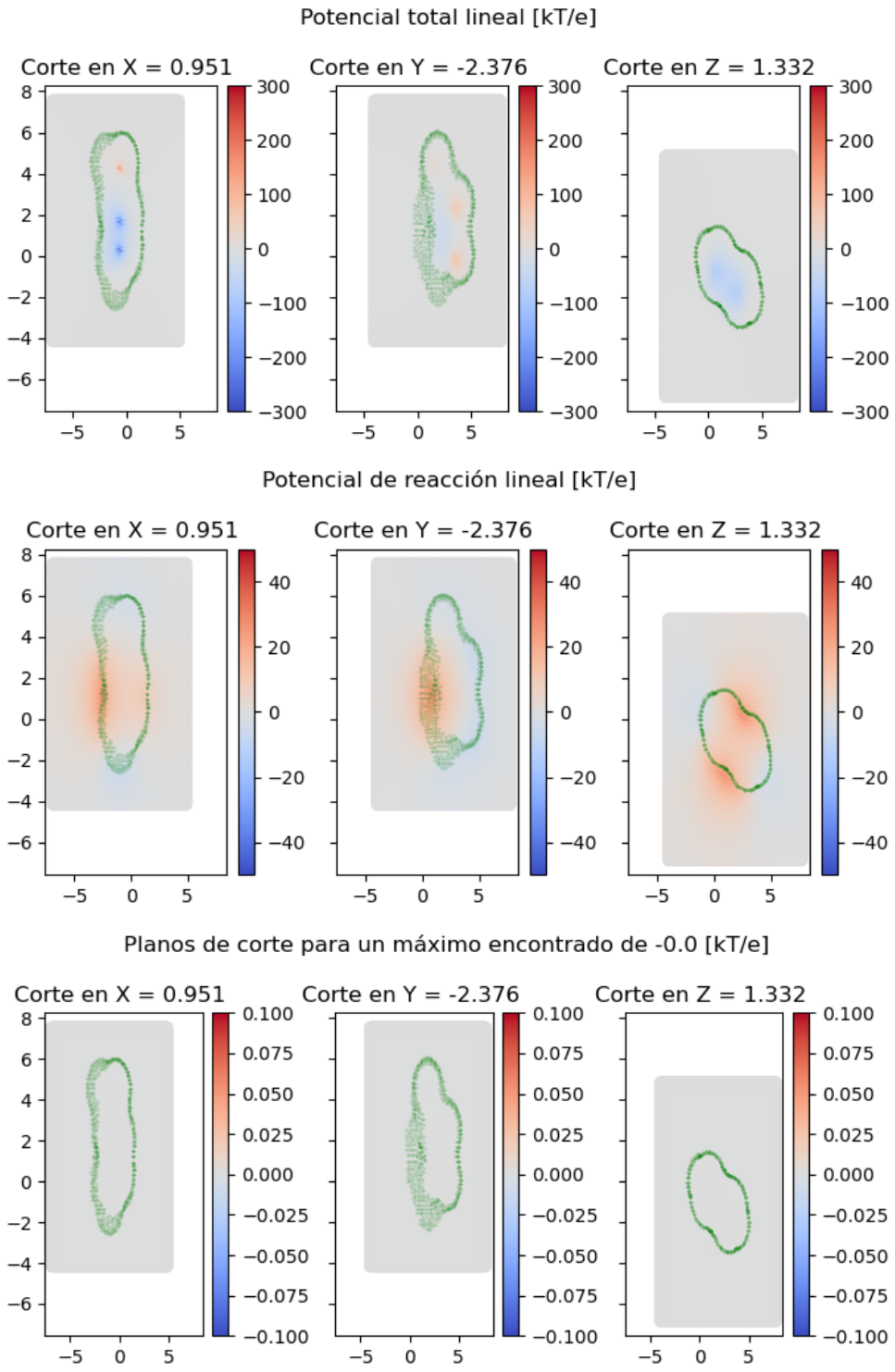

Figura 4.35: Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula de Tolueno

4.3.4. Metil Octanoato

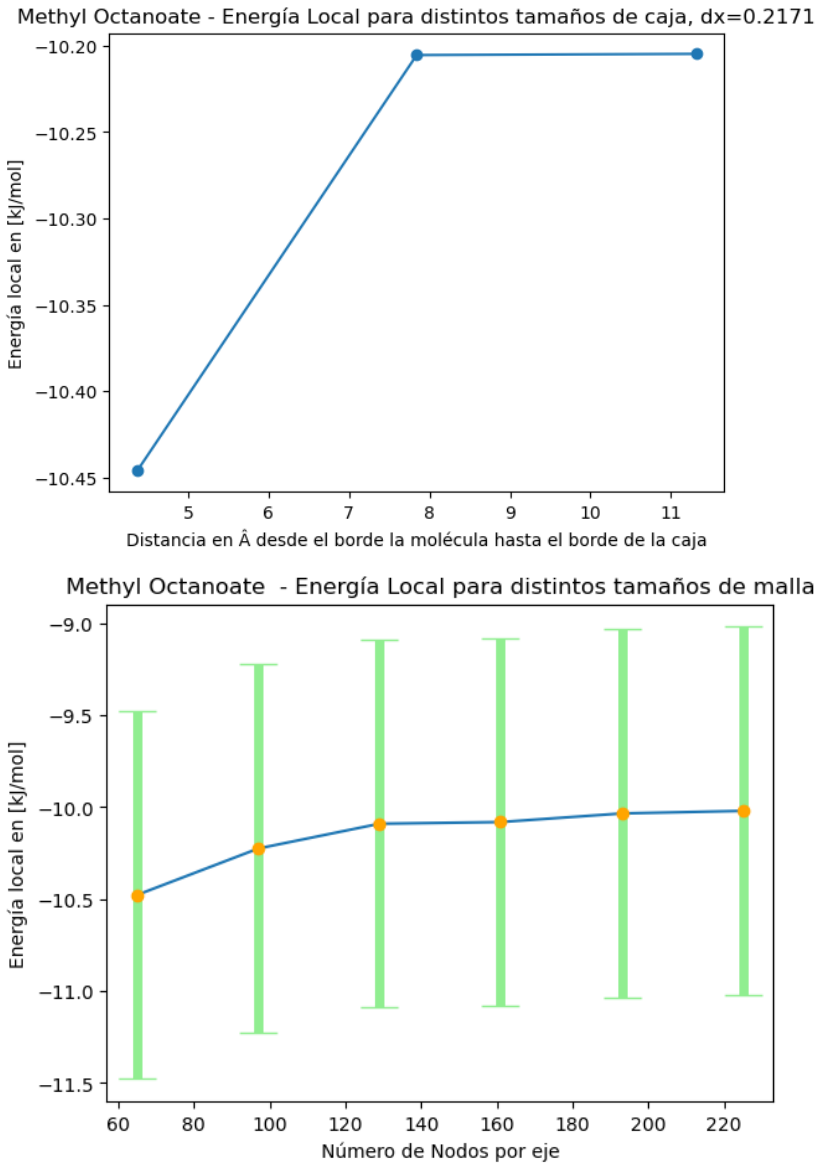

Figura 4.36: Análisis de sensibilidad para la molécula de Metil Octanoato

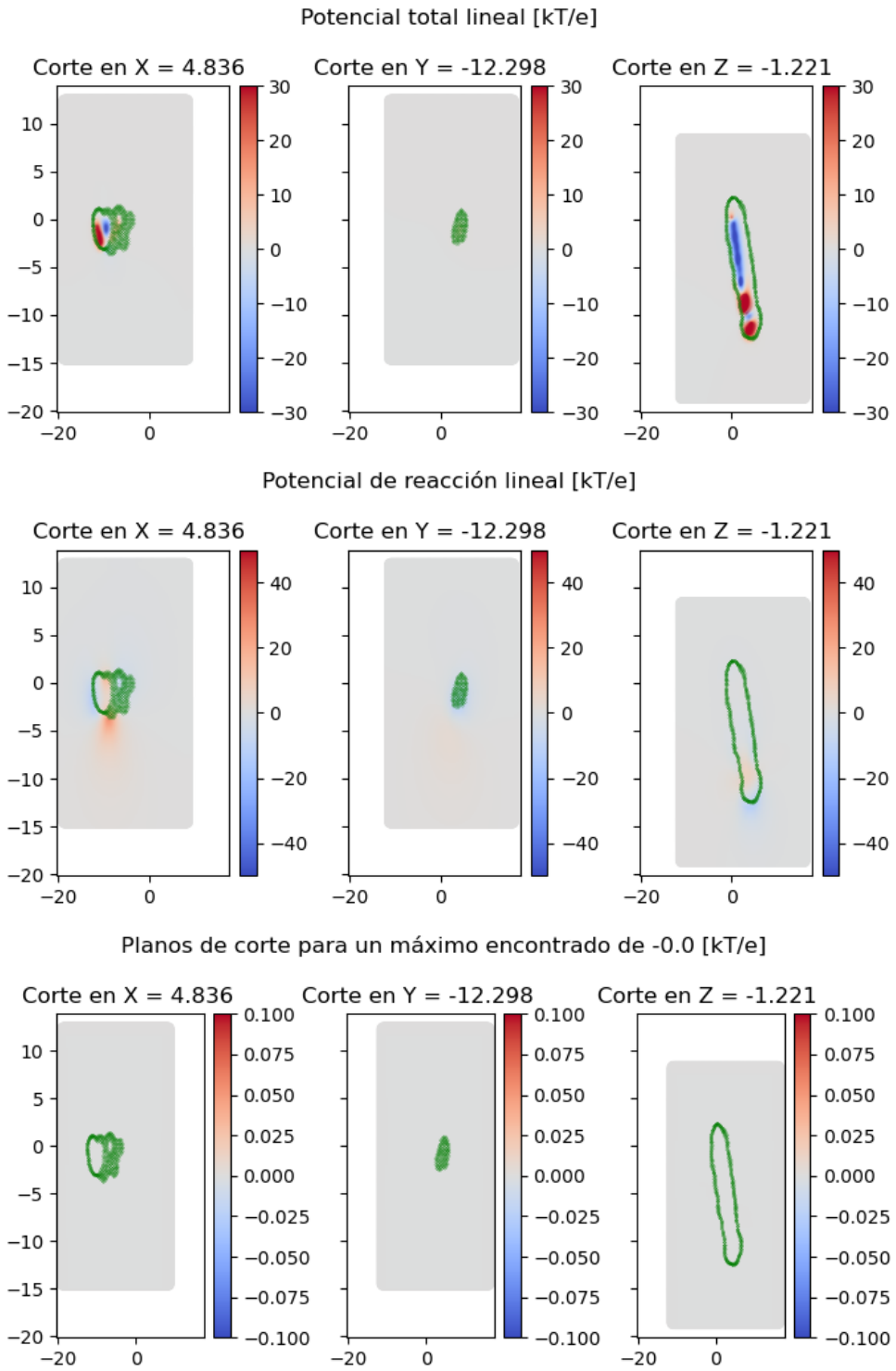

Figura 4.37: Potencial total lineal, potencial de reacción lineal y diferencia lineal-no lineal para la molécula de Metil Octanoato

Capítulo 5

Conclusión

En este trabajo hemos abordado la influencia de la linealización de la ecuación de Poisson-Boltzmann en diversas moléculas por medio del método de diferencias finitas, a sabiendas de las consideraciones y los límites que puede tener este tipo de método numérico, sobre todo en parámetros geométricos como el tamaño de la caja y el nivel de refinamiento de la malla que fueron estudiados para cada molécula con el fin de que los resultados obtenidos tuvieran un impacto mínimo debido a la geometría utilizada.

En relación a los objetivos específicos, revisamos el estado del arte e implementamos cálculos usando softwares como APBS y librerías como MDAnalysis. En rasgos generales se entiende la importancia de identificar características moleculares y efectos asociados a la linealización de la ecuación de PB.

Dentro de los resultados más significativos se concluye:

- En general, para los análisis de sensibilidad se encontró que una distancia media de 20[A] es suficiente para asegurar las condiciones de contorno de Dirichlet.
- El error generado de la diferencia entre el cálculo del potencial total linealizado y no linealizado tiende a concentrarse cerca de la interfaz de la molécula y en aberturas de helicoides, como se observa en la 9ANT.
- Se identificó un alto potencial en áreas cóncavas, y es una característica que se repite en todas las moléculas de las baterías medianamente y altamente cargadas. Si bien los resultados del grado de cobertura no fueron acertados, permiten poner en

discusión la idea de buscar un método que relacione el grado de cobertura de una zona, con el error producto de la diferencia de calcular el potencial linealizado y no linealizado.

- El orden de magnitud del error mencionado es mayor en la batería altamente cargada, tomando valores no despreciables en la molécula 1KX5.
- Se estableció un modelo para estimar el error en moléculas compuestas solo por nucleótidos en función de tres moléculas analizadas y luego probada en otras dos. Se deja como trabajo a futuro extender el uso, primero sobre moléculas que solo contengan ADN/ARN para que el modelo sea más representativo. Y luego, como un trabajo aparte, para familias de moléculas que puedan ser reconocidas como tal, por ejemplo cadena de nucleótidos que presenten una histoma anexada, etc. Por lo tanto se sugiere que una selección más precisa de una batería o familias de moléculas podría llevar a modelos más robustos.
- Las baterías de baja carga no mostraron diferencias significativas en los resultados entre modelos lineales y no lineales.

Si bien se lograron avances, también quedan varias incertidumbres; queda inconclusa una propuesta por medición de concavidad o grado de cobertura de las moléculas para determinar el error del método linealizado versus el no linealizado; también la falta de modelos para otro tipo de sub baterías generadas según su estructura o conformación. Estas brechas nos ofrecen oportunidades para investigaciones futuras en la estimación de zonas de error en el cálculo de potencial electroestático en biomoléculas.

Bibliografía

[1] Nathan A. Baker. Poisson–boltzmann methods for biomolecular electrostatics. In *Numerical Computer Methods, Part D*, volume 383 of *Methods in Enzymology*, pages 94–118. Academic Press, 2004.

[2] Sergio Decherchi, Matteo Masetti, Ivan Vyalov, and Walter Rocchia. Implicit solvent methods for free energy estimation. *European Journal of Medicinal Chemistry*, 91:27–42, 2015.

[3] M. P. Allen and D. J. & Tildesley. Computer simulation of liquids. *Oxford University Press*, 1987.

[4] Thomas Madej, Christopher J Lanczycki, Dachuan Zhang, Paul A Thiessen, Renata C Geer, Aron Marchler-Bauer, and Stephen H Bryant. Mmdb and vast+: tracking structural similarities between macromolecular complexes. *Nucleic acids research*, 42(D1):D297–D303, 2014.

[5] R Diamond. Real-space refinement of the structure of hen egg-white lysozyme. *Journal of molecular biology*, 82(3):371–391, 1974.

[6] Elizabeth Jurrus, Dave Engel, Keith Star, Kyle Monson, Juan Brandi, Lisa E Felberg, David H Brookes, Leighton Wilson, Jiahui Chen, Karina Liles, et al. Improvements to the apbs biomolecular solvation software suite. *Protein Science*, 27(1):112–128, 2018.

[7] Elizabeth Jurrus, Dave Engel, Keith Star, Kyle Monson, Juan Brandi, Lisa E. Felberg, David H. Brookes, Leighton Wilson, Jiahui Chen, Karina Liles, Minju Chun, Peter Li, David W. Gohara, Todd Dolinsky, Robert Konecny, David R. Koes, Jens Erik Nielsen, Teresa Head-Gordon, Weihua Geng, Robert Krasny, Guo-Wei Wei, Michael J. Holst, J. Andrew McCammon, and Nathan A. Baker. Improvements to the apbs biomolecular solvation software suite. *Protein Science*, 27(1):112–128, 2018.

[8] Ramu Anandakrishnan, Boris Aguilar, and Alexey V Onufriev. H++ 3.0: automating p k prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic acids research*, 40(W1):W537–W541, 2012.

[9] Byron Vega L. Potencial Electroestático, 9 2023. https://github.com/ByronVegaL/PotencialElectroestatico.git.

[10] Binhan Yu, Xi Wang, and Junji Iwahara. Measuring local electrostatic potentials around nucleic acids by paramagnetic nmr spectroscopy. *The Journal of Physical Chemistry Letters*, 13(42):10025–10029, 2022.

A1. Registro de cortes - 1KX5

A1.1. Potencial total lineal - 1KX5

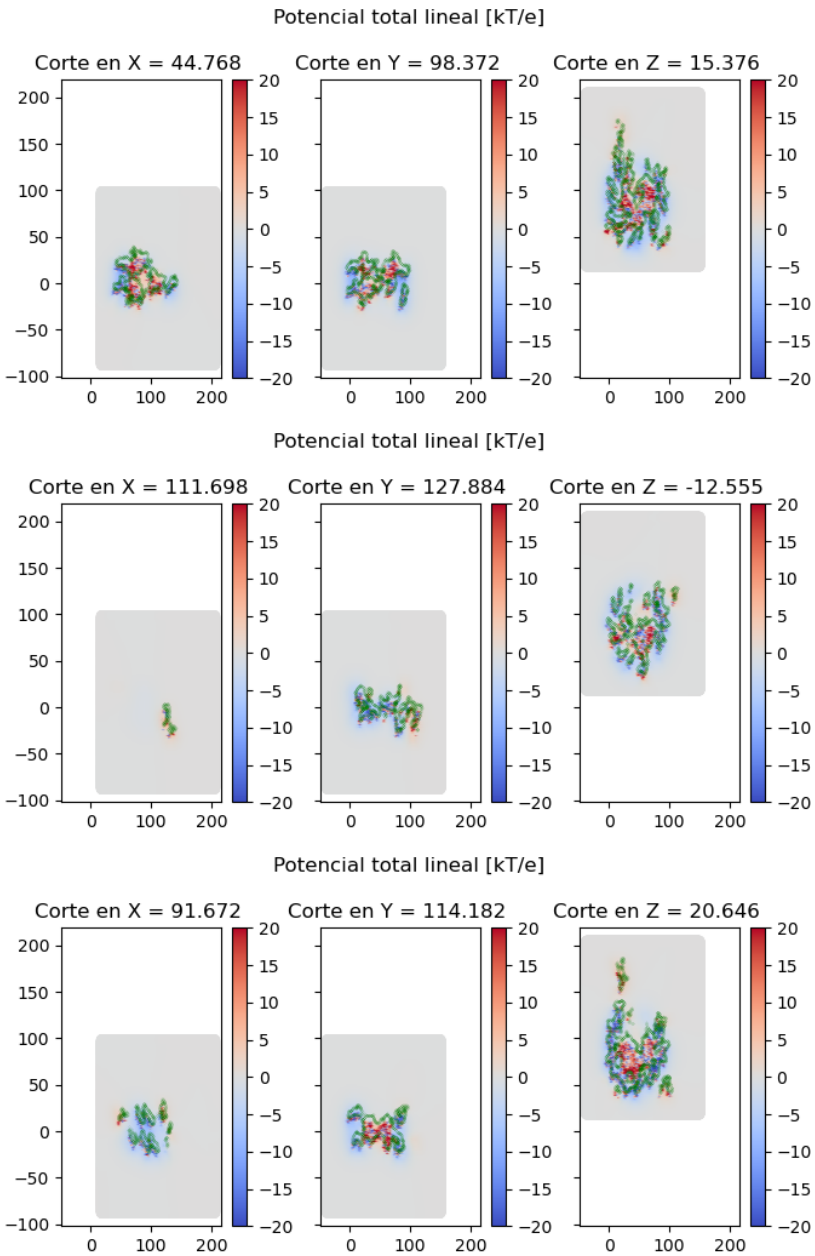

Figura A1.1: Potencial total lineal - 1KX5

A1.2. Potencial de reacción lineal - 1KX5

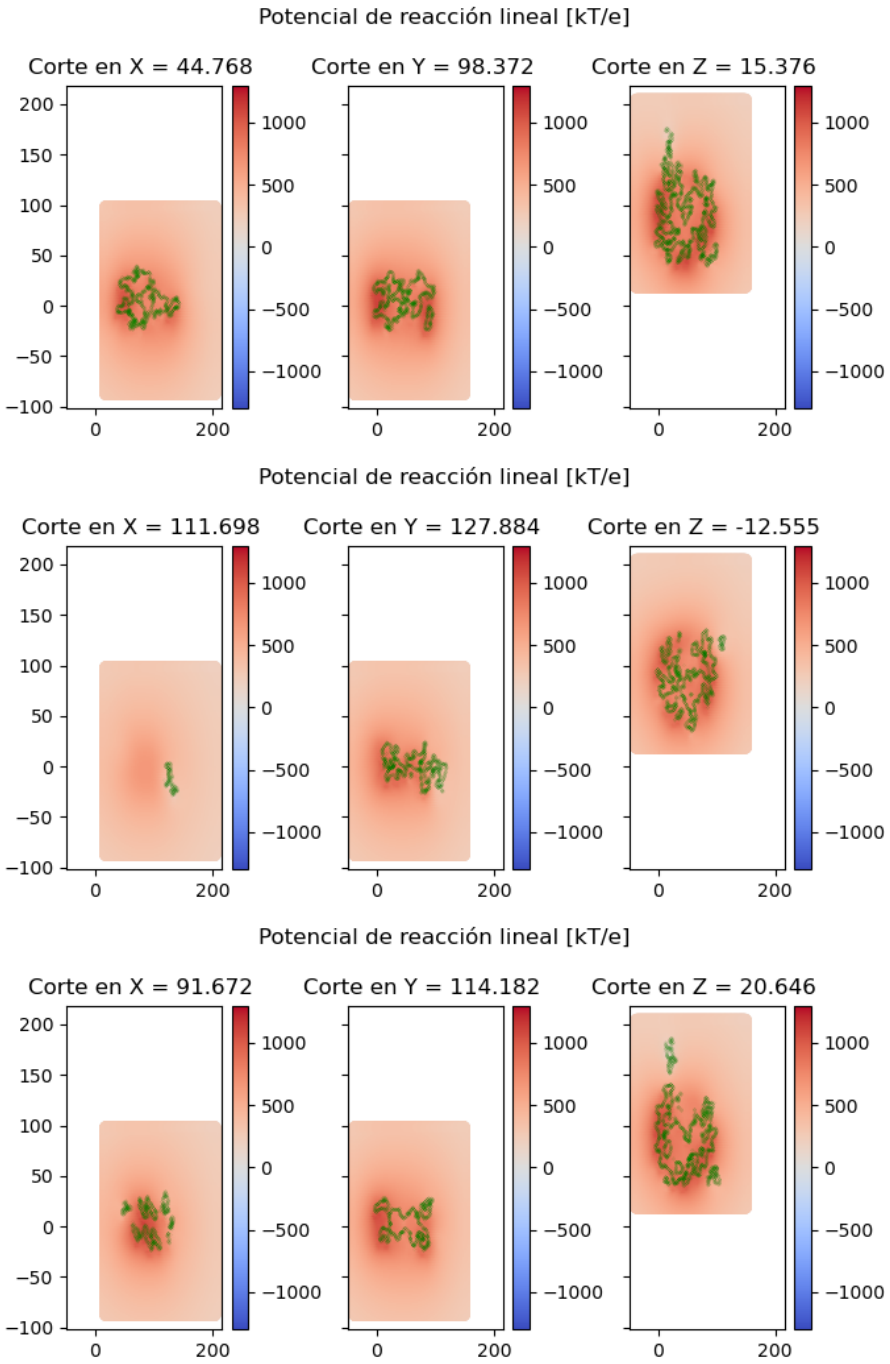

Figura A1.2: Potencial de reacción lineal - 1KX5

A1.3. Error lineal no lineal - 1KX5

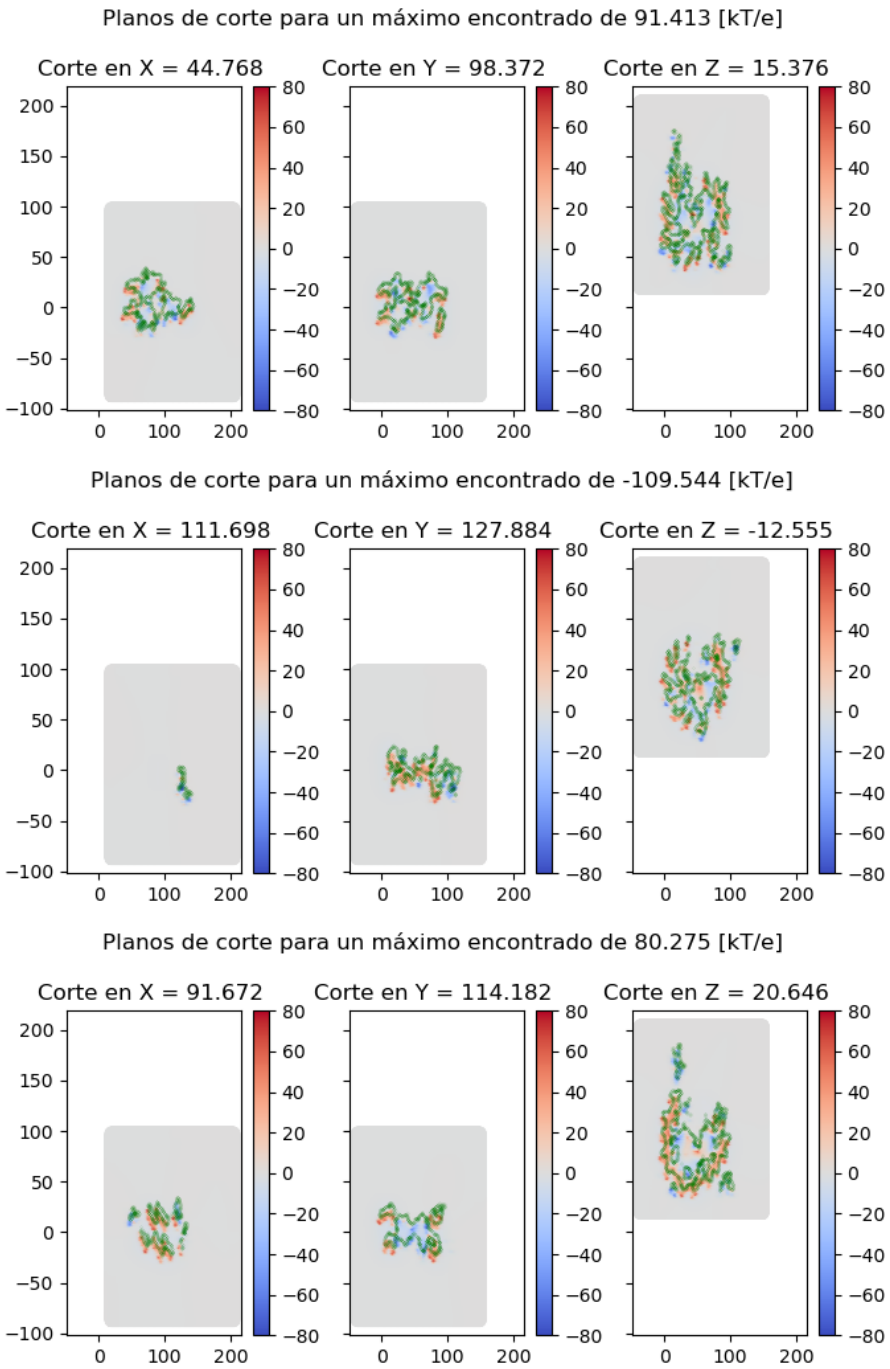

Figura A1.3: Error lineal no lineal - 1KX5

A2. Registro de cortes - 9ANT

A2.1. Potencial total lineal - 9ANT

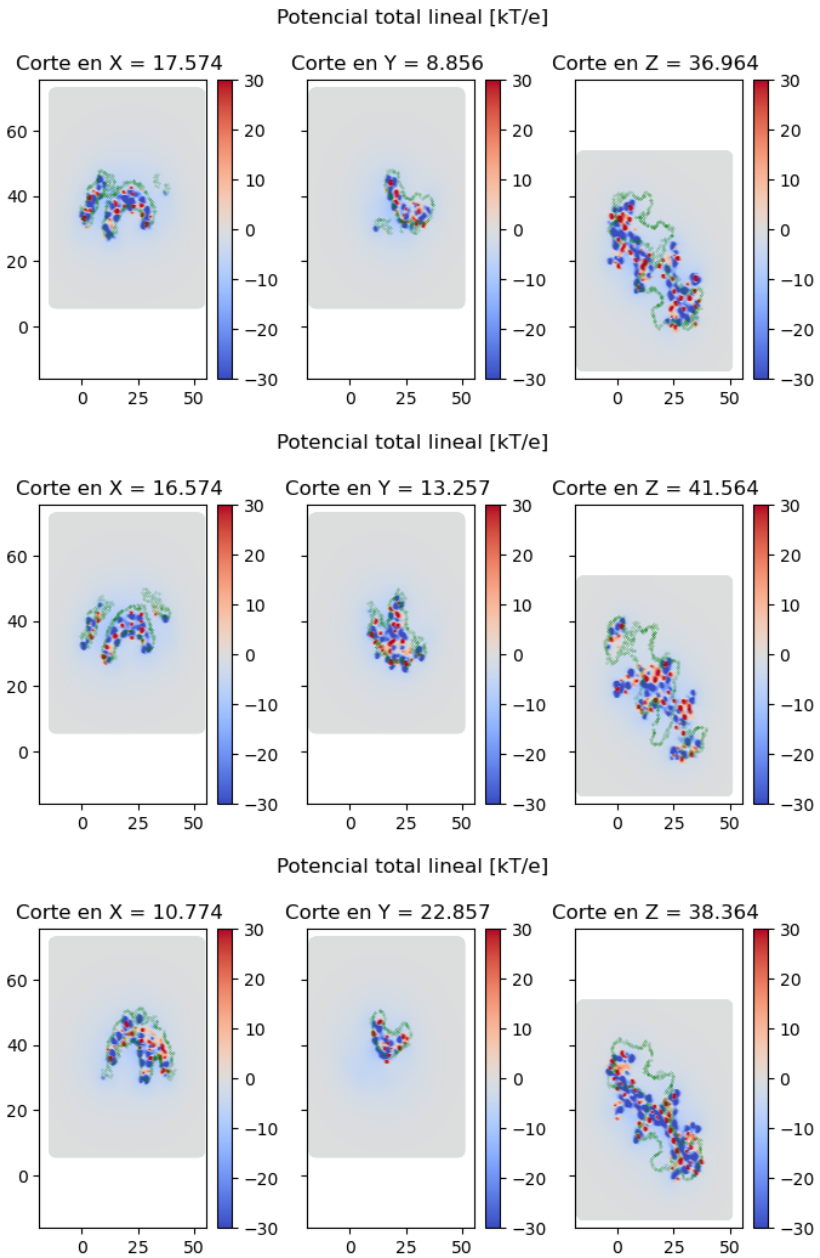

Figura A2.1: Descripción de ambas imágenes

A2.2. Potencial de reacción lineal - 9ANT

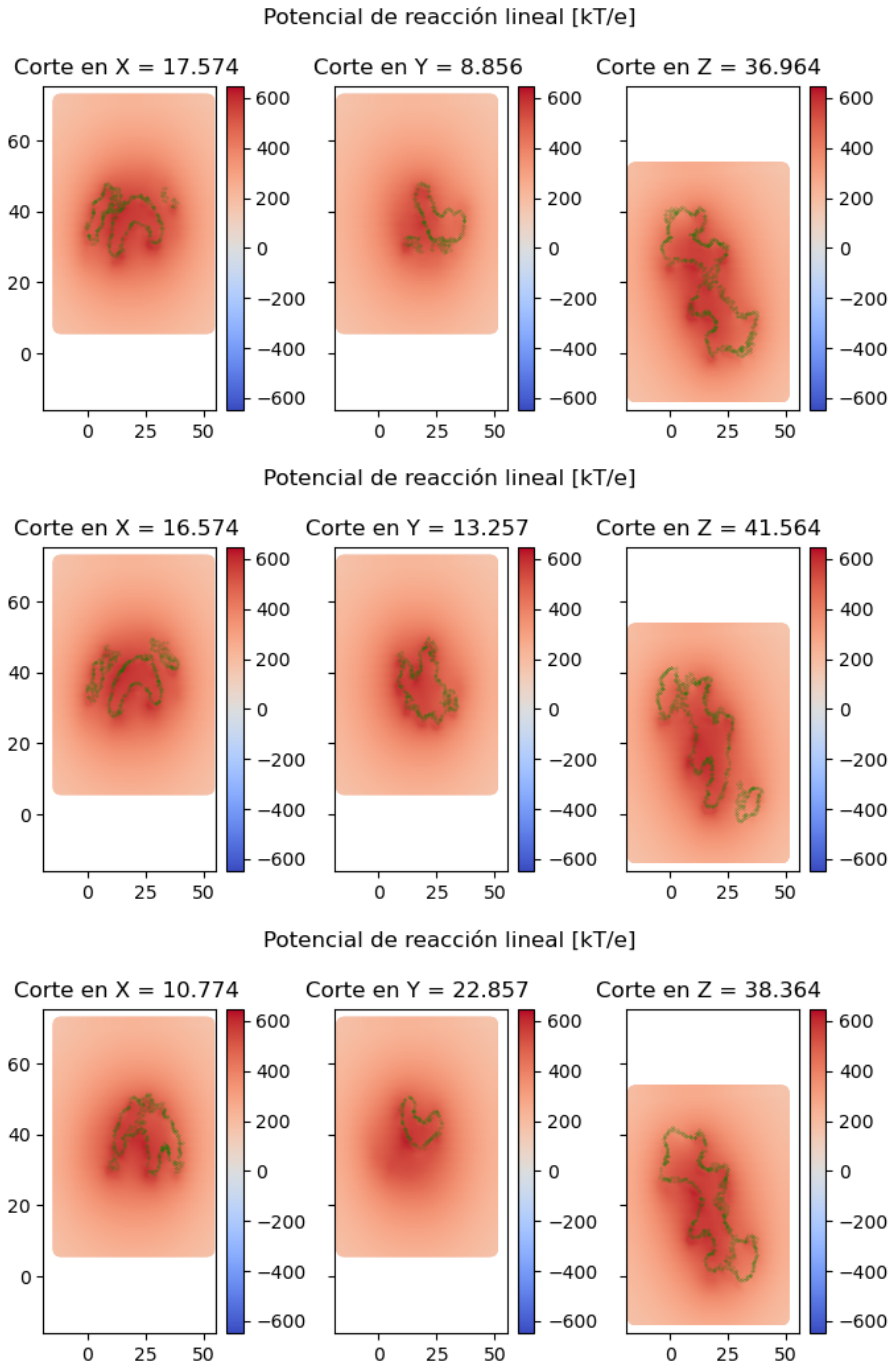

Figura A2.2: Potencial de reacción lineal - 9ANT

A2.3. Error lineal no lineal - 9ANT

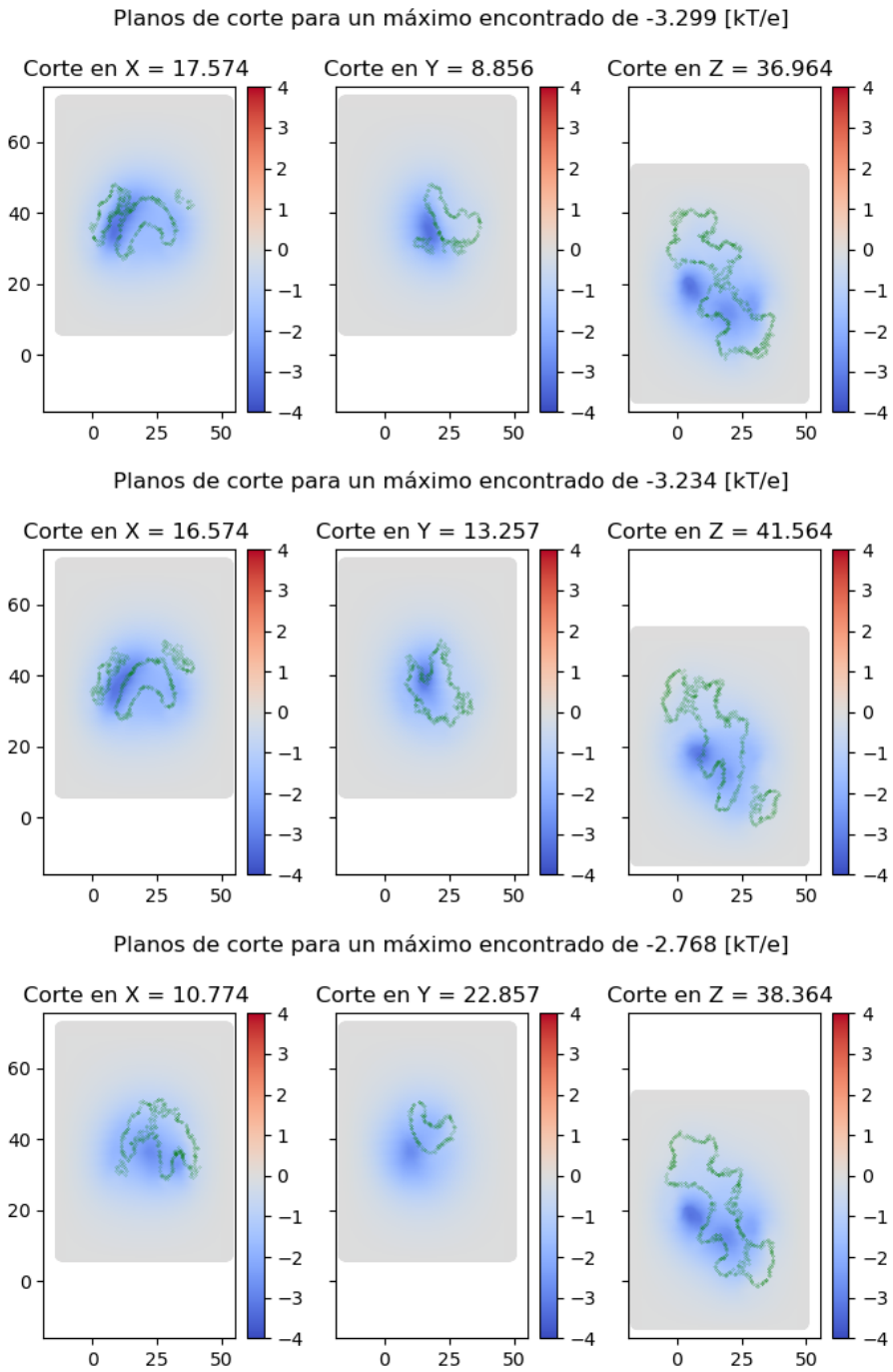

Figura A2.3: Error lineal no lineal - 9ANT

A3. Registro de cortes - 280D

A3.1. Potencial total lineal - 280D

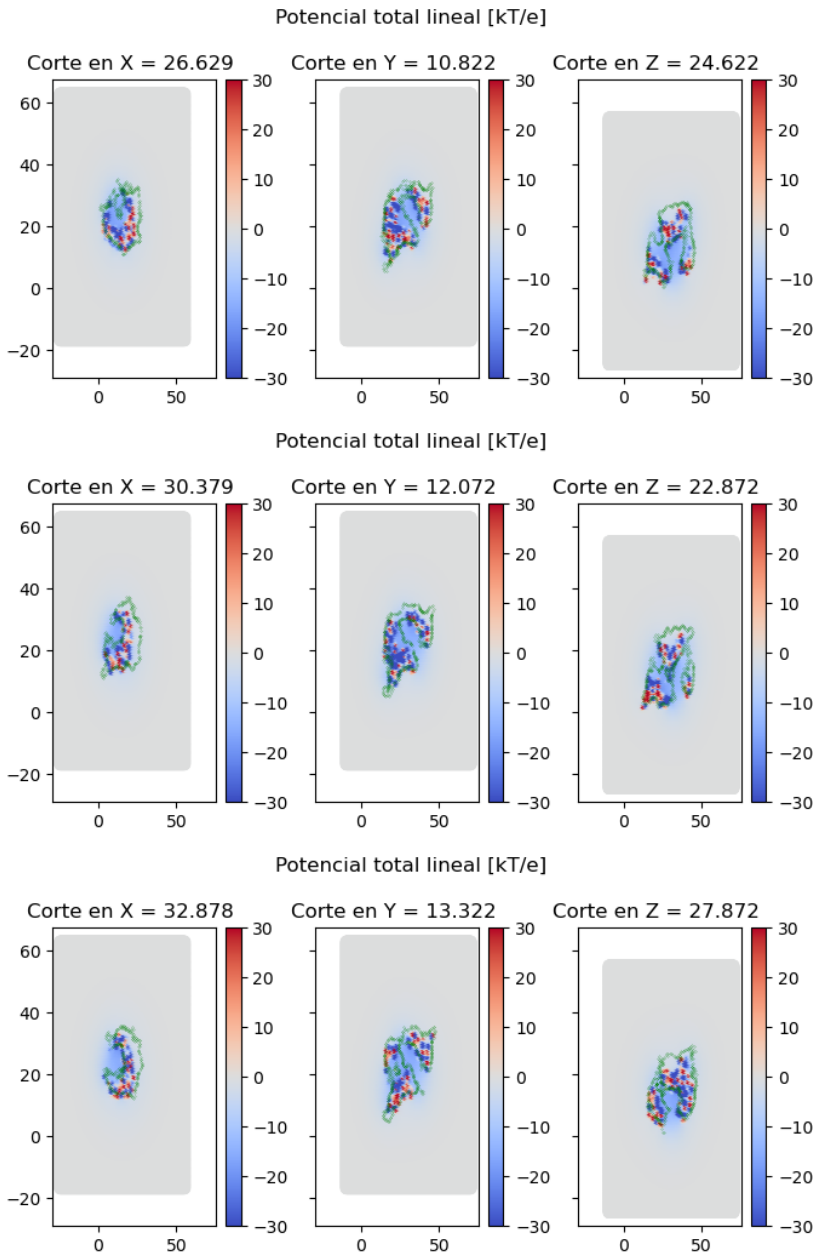

Figura A3.1: Potencial total lineal - 280D

A3.2. Potencial de reacción lineal - 280D

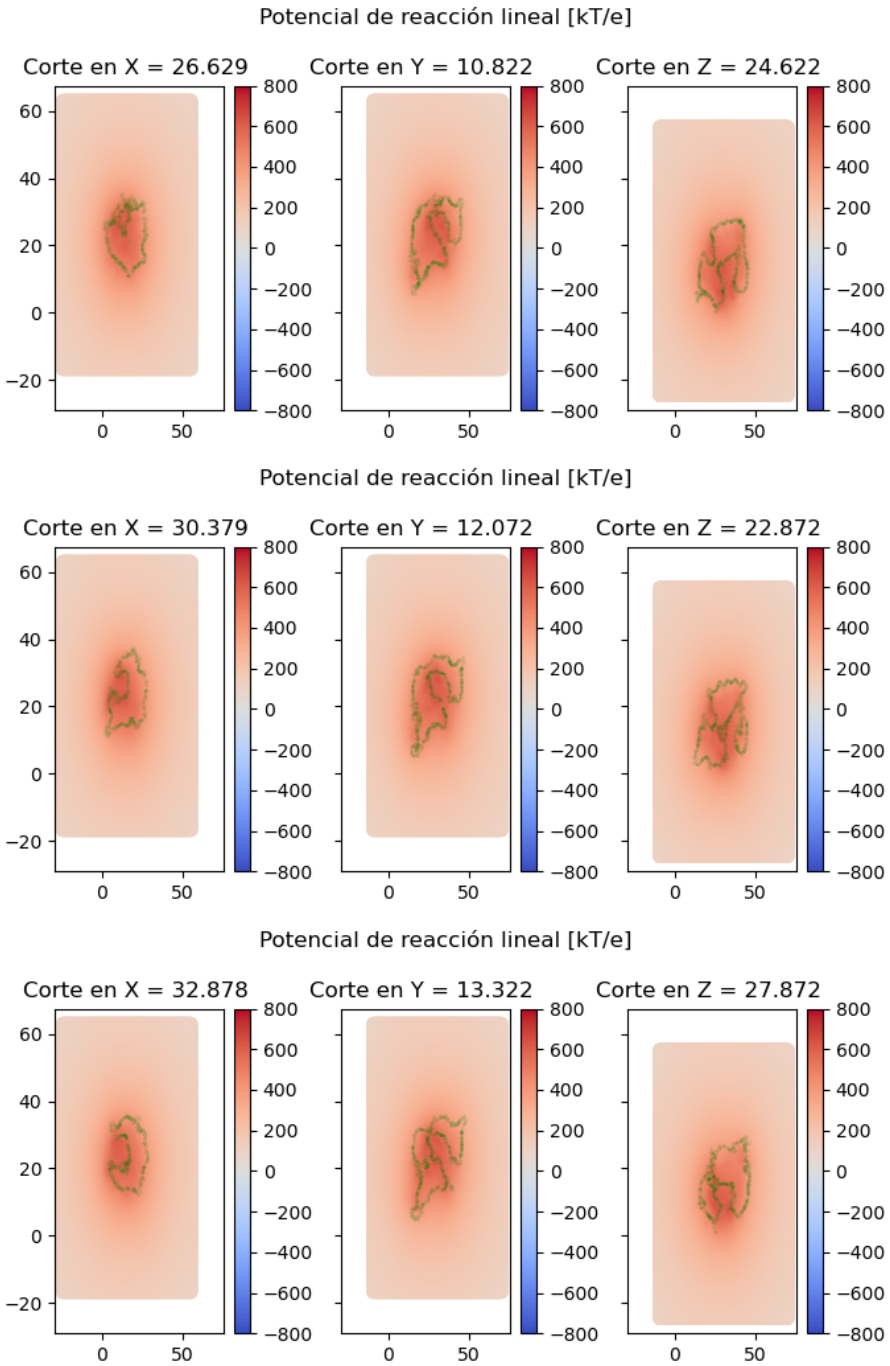

Figura A3.2: Potencial de reacción lineal - 280D

A3.3. Error lineal no lineal - 280D

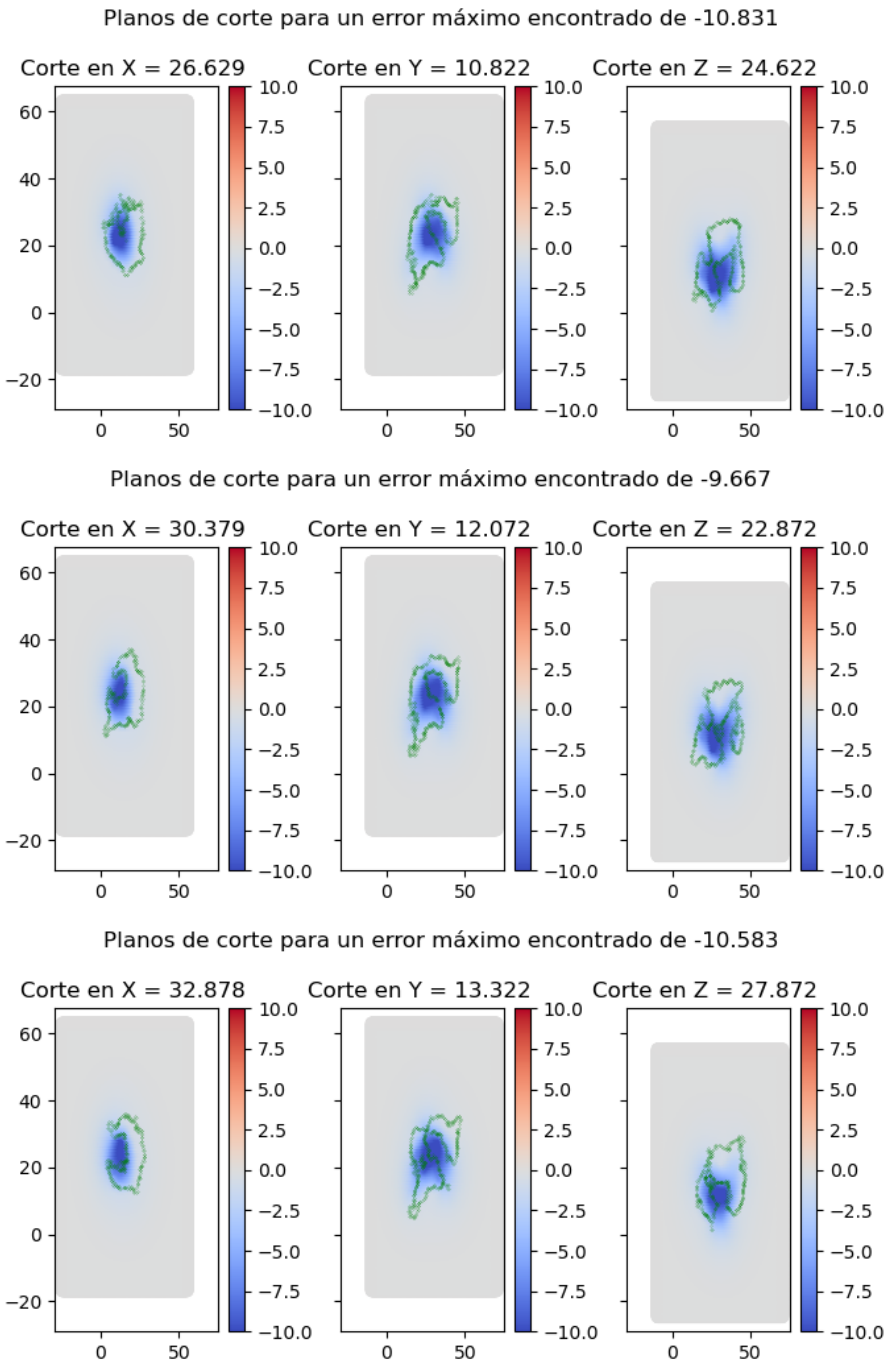

Figura A3.3: Error lineal no lineal - 280D

A4. Registro de cortes - 437D

A4.1. Potencial total lineal - 437D

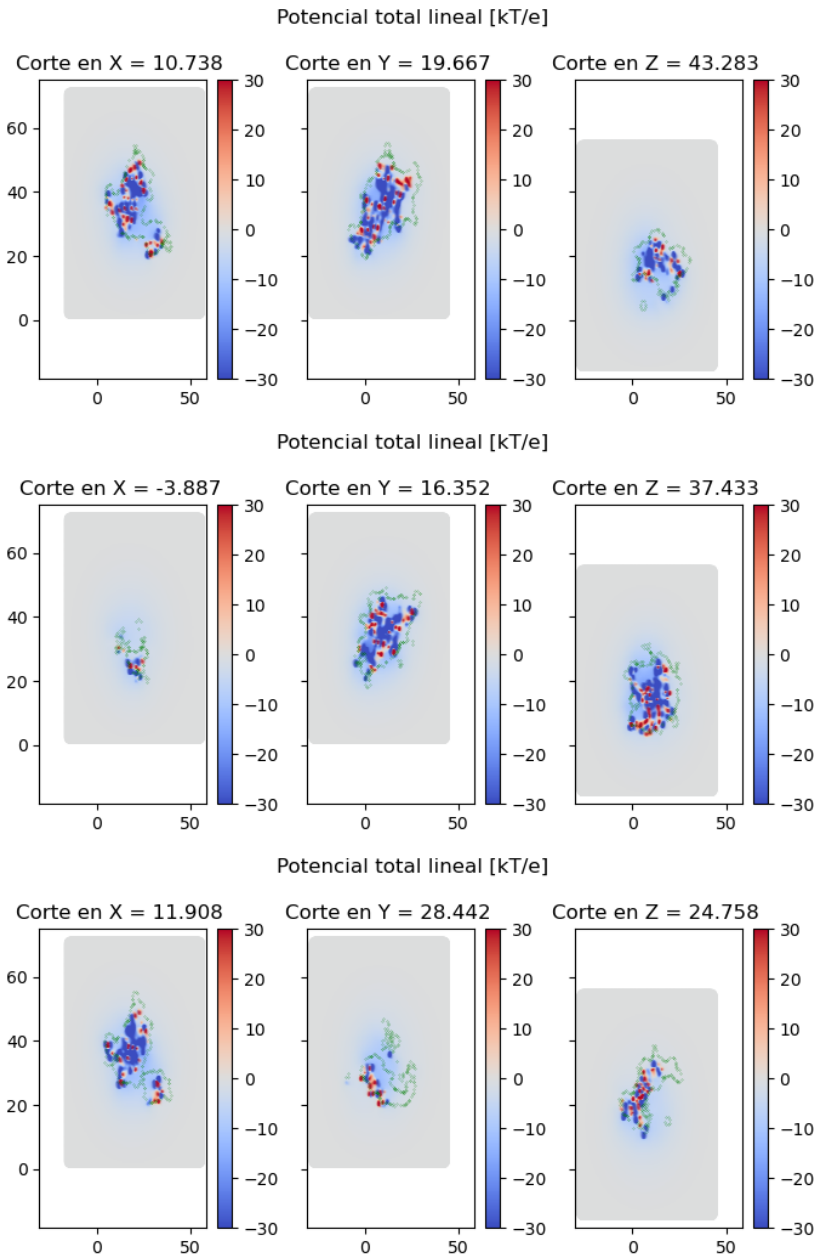

Figura A4.1: Registro de cortes - 437D

A4.2. Potencial de reacción lineal - 437D

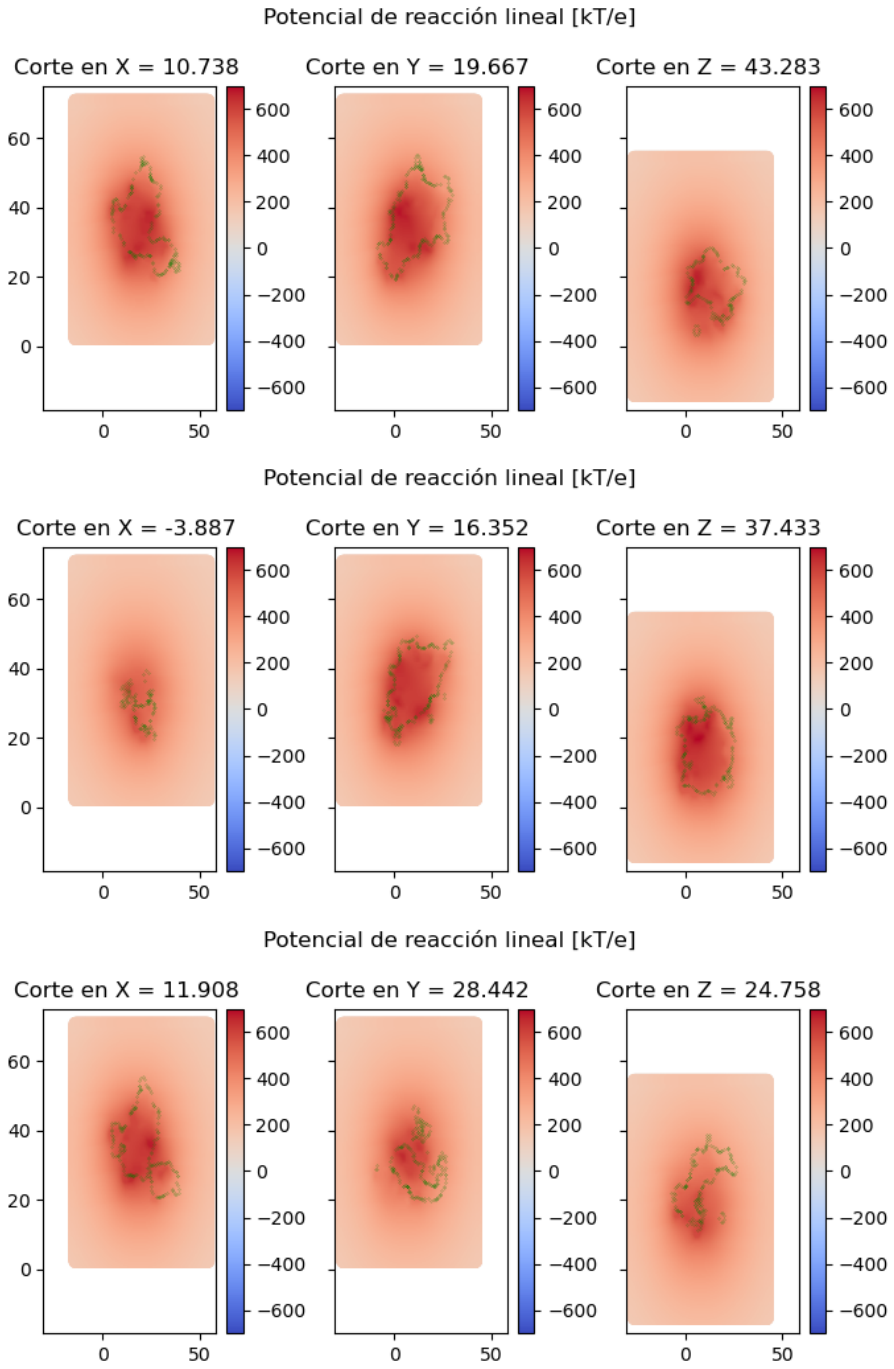

Figura A4.2: Potencial de reacción lineal - 437D

A4.3. Error lineal no lineal - 437D

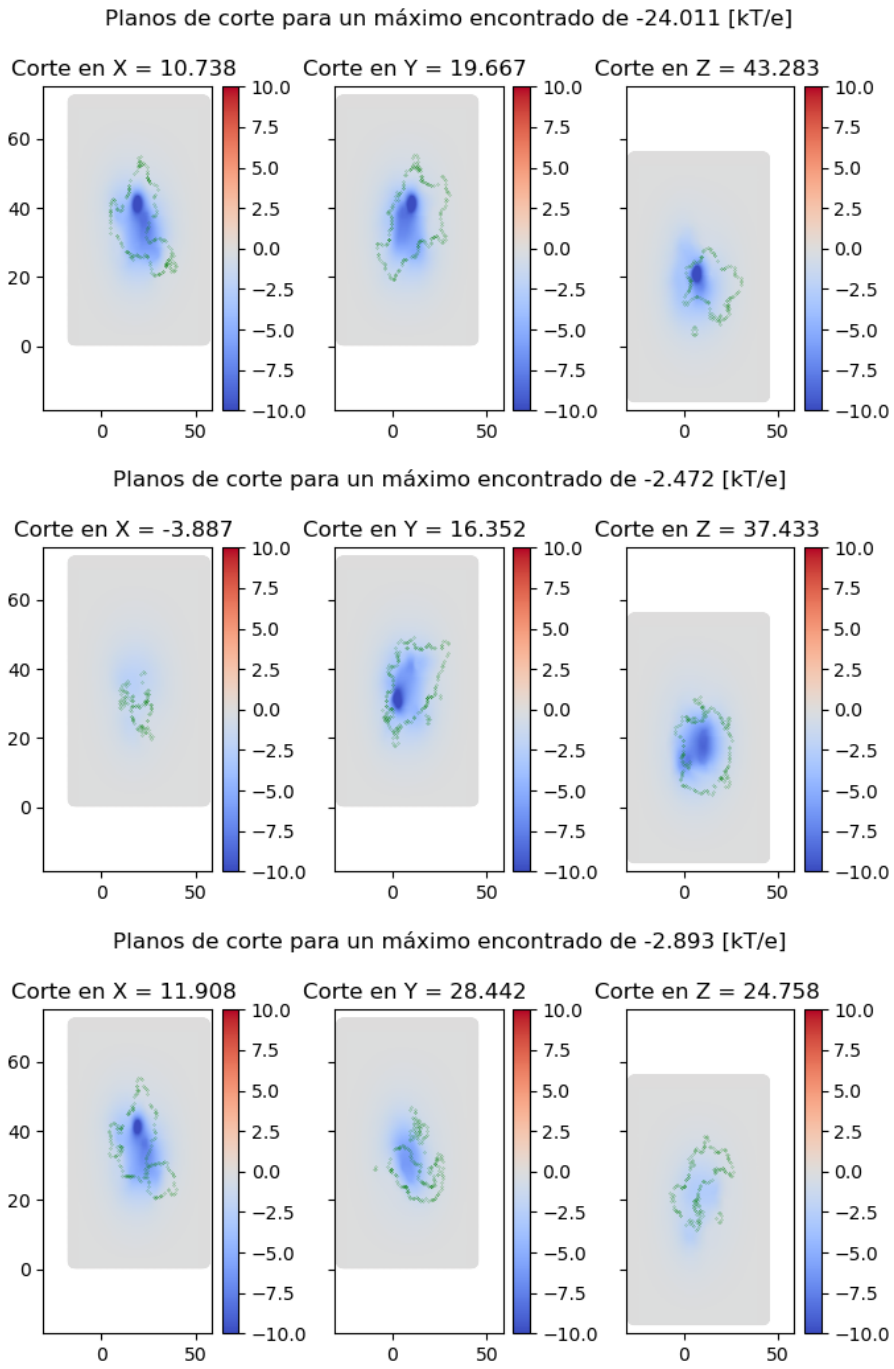

Figura A4.3: Error lineal no lineal - 437D

A5. Registro de cortes - 1LYZ

A5.1. Potencial total lineal - 1LYZ

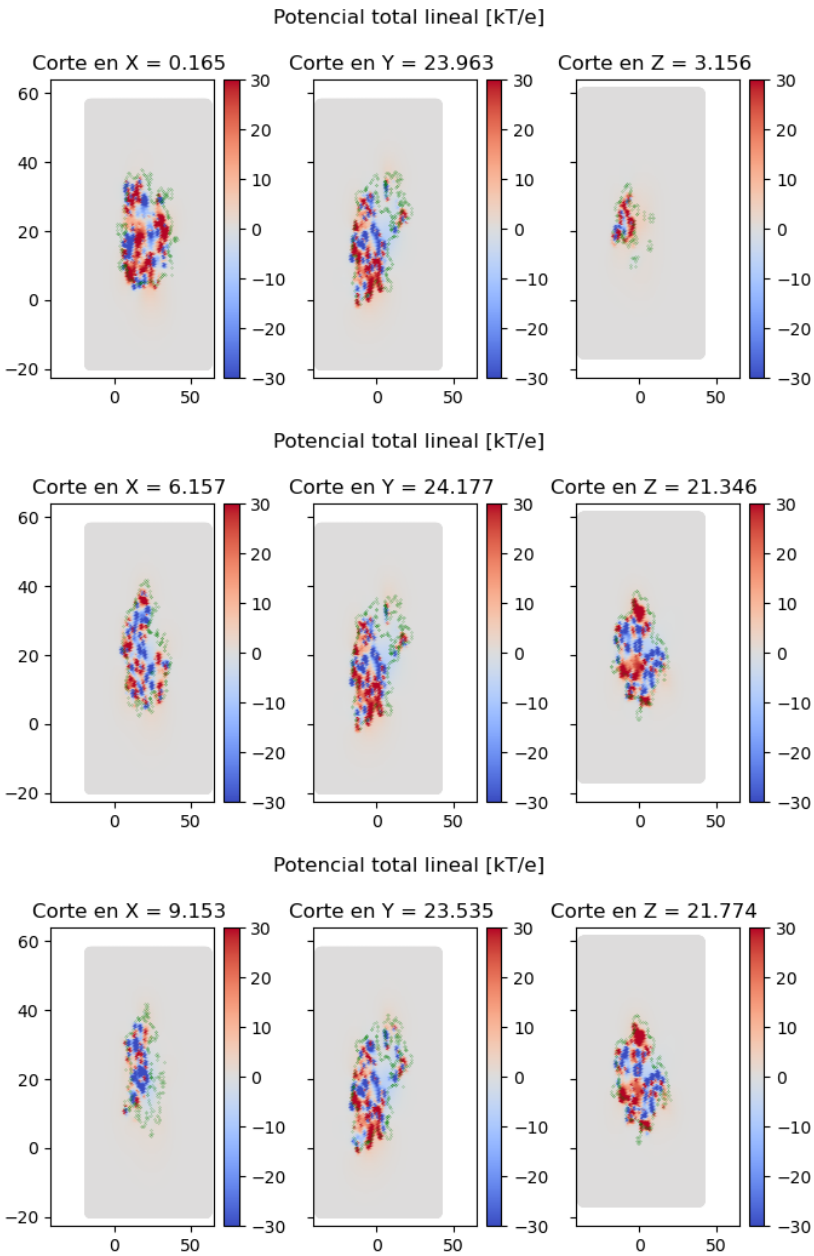

Figura A5.1: Potencial total lineal - 1LYZ

A5.2. Potencial de reacción lineal - 1LYZ

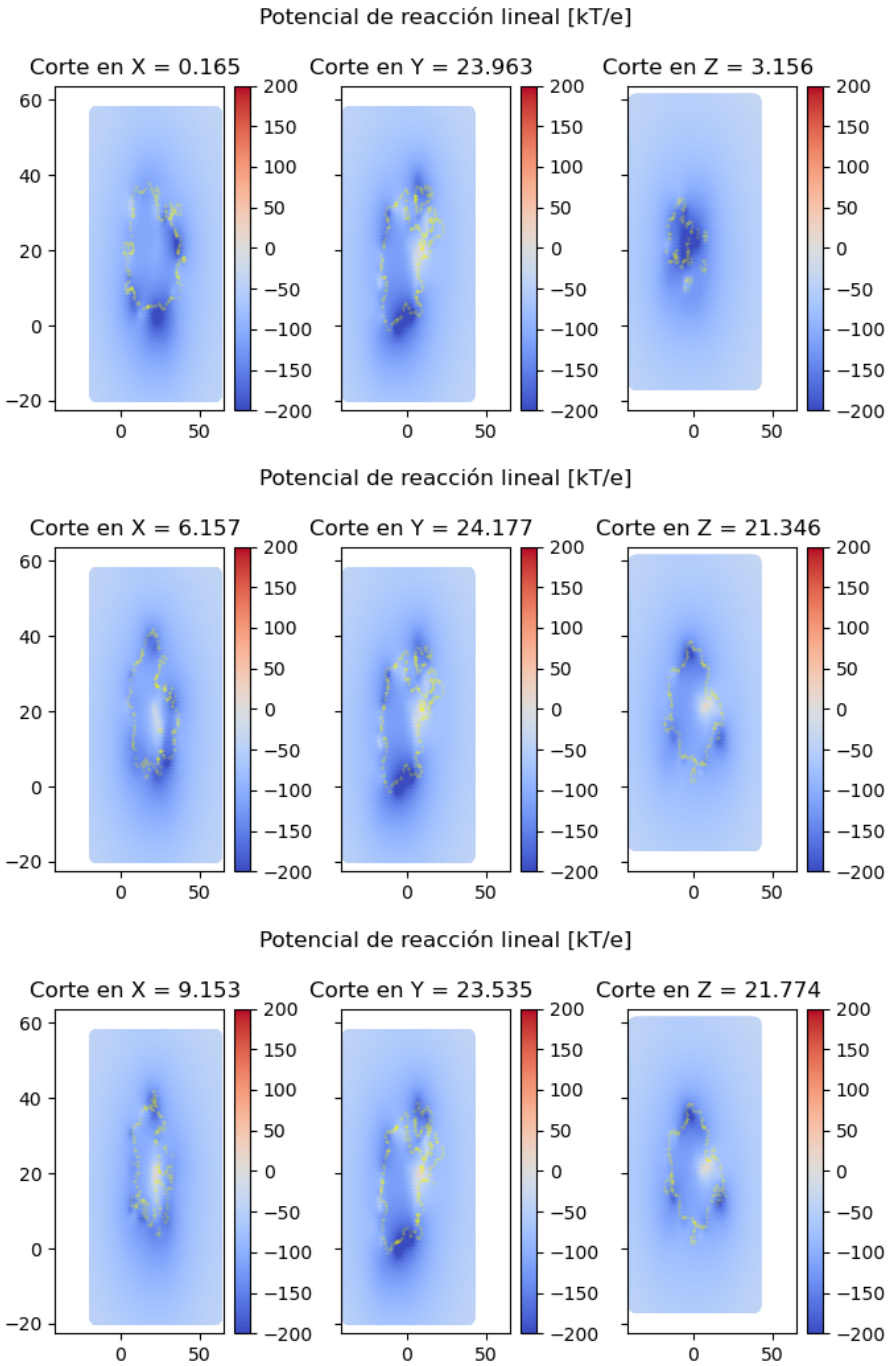

Figura A5.2: Potencial de reacción lineal - 1LYZ

A5.3. Error lineal no lineal - 1LYZ

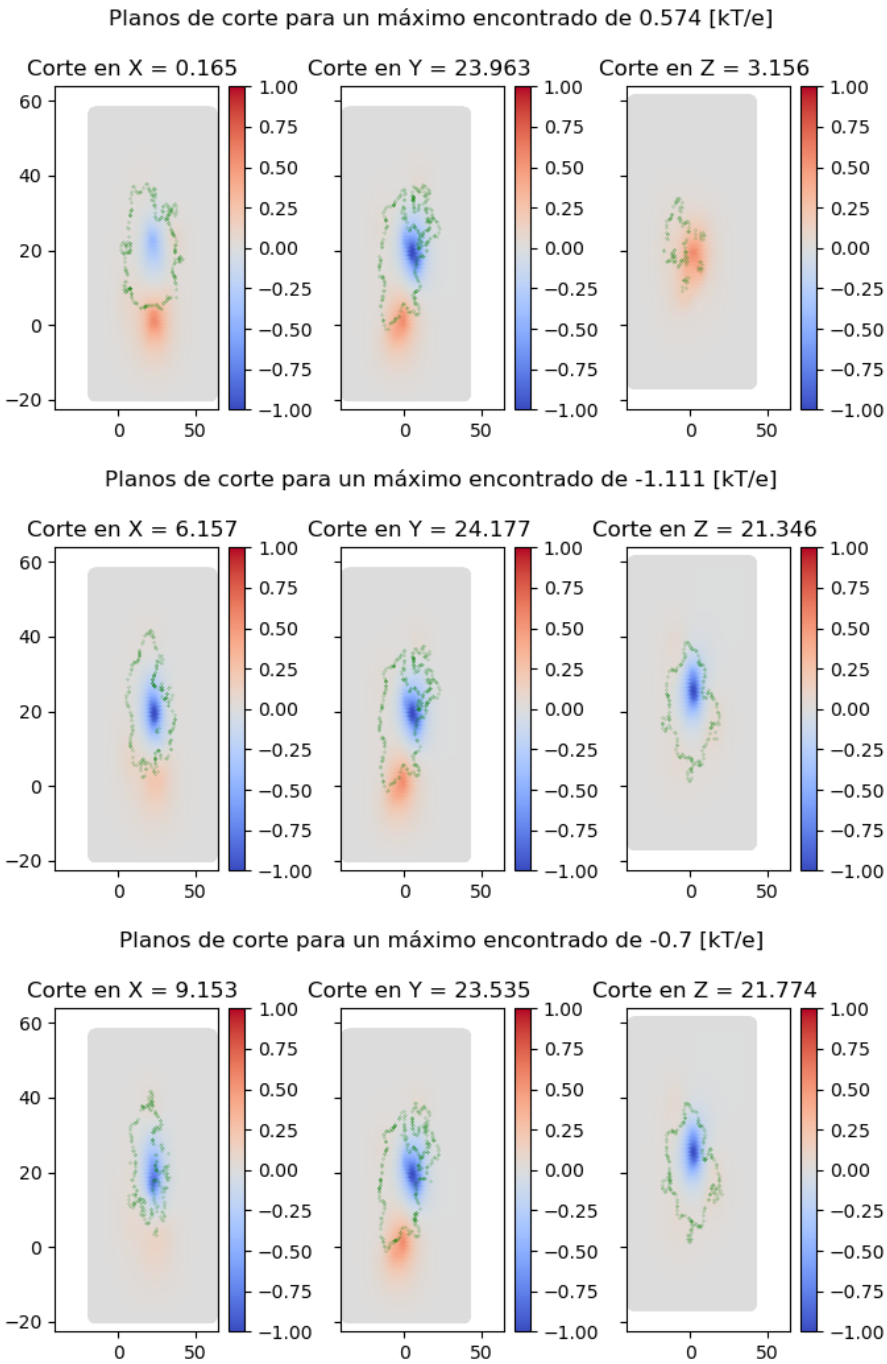

Figura A5.3: Error lineal no lineal - 1LYZ

A6. Registro de cortes - 1UBQ

A6.1. Potencial total lineal - 1UBQ

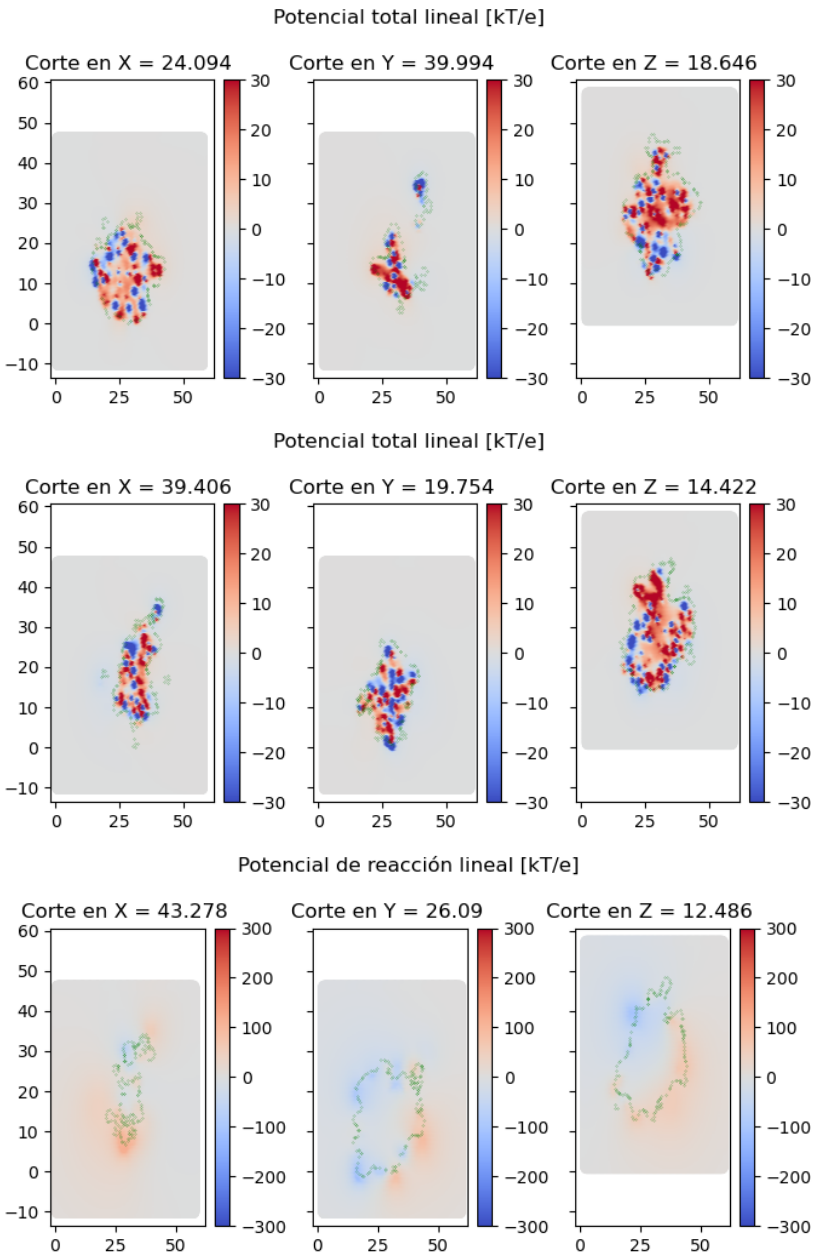

Figura A6.1: Potencial total lineal - 1UBQ

A6.2. Potencial de reacción lineal - 1UBQ

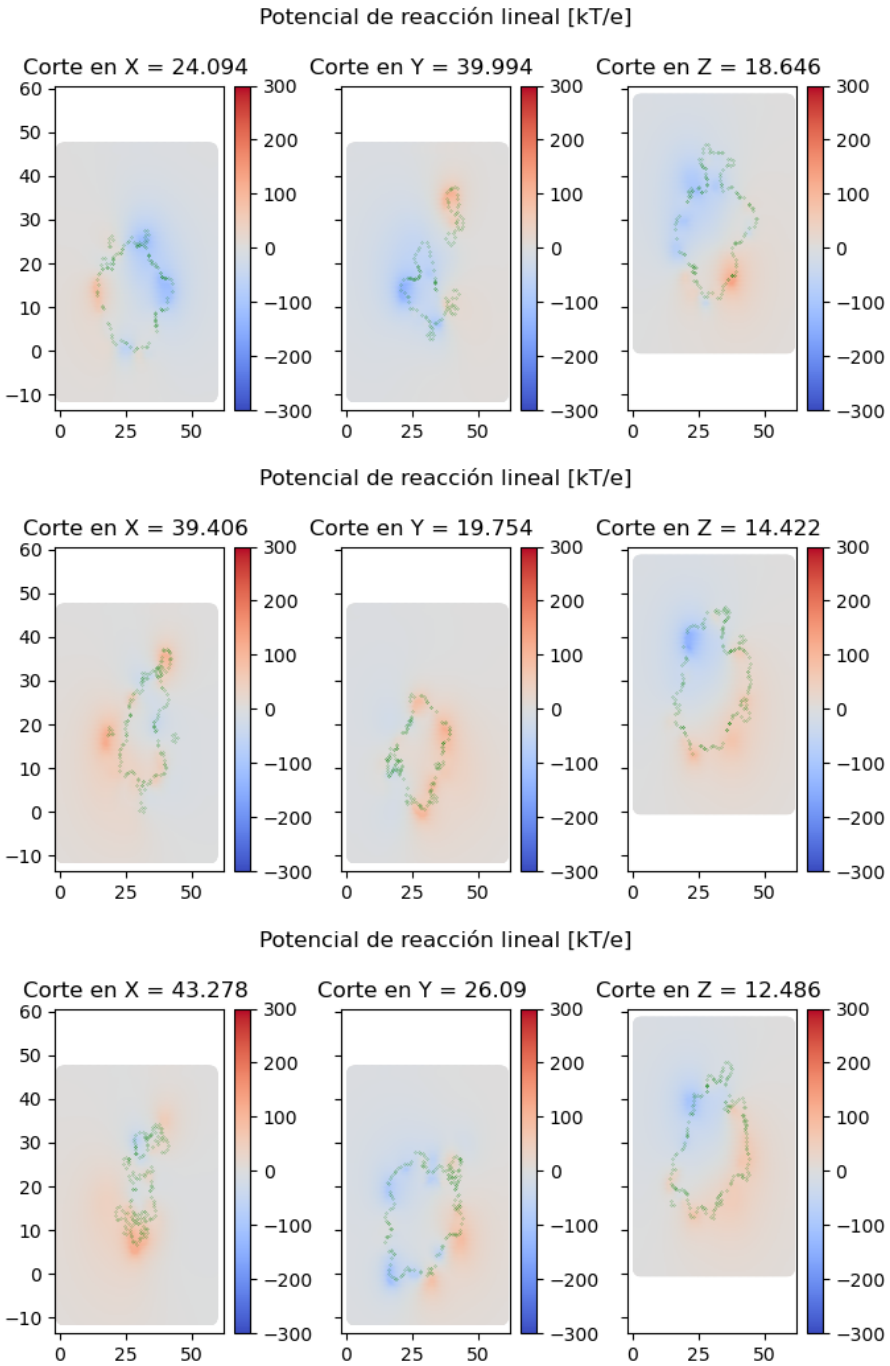

Figura A6.2: Potencial de reacción lineal - 1UBQ

A6.3. Error lineal no lineal - 1UBQ

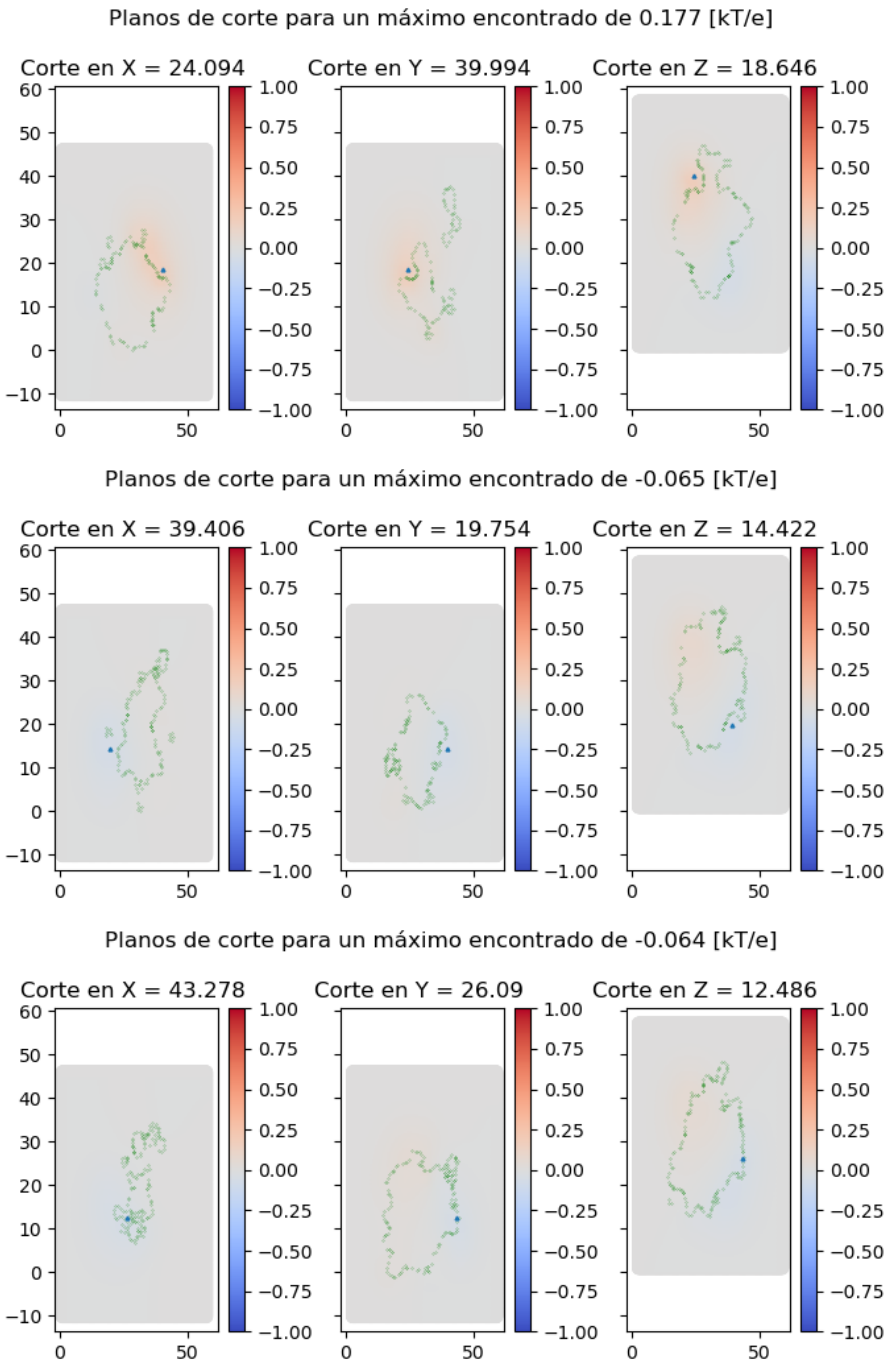

Figura A6.3: Error lineal no lineal - 1UBQ