



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Efectos Inductivos en Aminoácidos y Péptidos:
Constantes de Ionización y Fluorescencia del Triptófano

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:
Jesús Lara Popoca

TUTOR PRINCIPAL
Dr. Enrique Rudiño Piñera
[Instituto de Biotecnología UNAM](#)

MIEMBROS DEL COMITÉ TUTOR
Dra. Adela Rodríguez Romero
[Instituto de Química UNAM](#)

Dr. Lorenzo Patrick Segovia Forcella
[Instituto de Biotecnología UNAM](#)

Ciudad de México. octubre, 2022



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**Efectos Inductivos en Aminoácidos y Péptidos:
Constantes de Ionización y Fluorescencia del
Triptófano**

Jesús Lara Popoca

19 de octubre de 2022

0

Índice general

1. Agradecimientos	5
2. Resumen	7
3. Introducción	9
3.1. Motivación	9
3.2. Hipótesis	10
3.3. Contribuciones	13
4. Cálculo de Proporciones de Tiol/Tiolato en Proteínas Modelo	15
4.1. Probabilidades individuales	15
4.2. Probabilidad de formación de puentes disulfuro	16
5. El Algoritmo CT	21
5.1. El algoritmo CT	21
5.2. Introducción	22
5.3. Invariantes estructurales	24
5.4. La aparente paradoja	25
5.5. Mejoras al Algoritmo CT	28
5.6. Correlaciones de <i>ii</i> con diferentes mediciones	35
6. Mediciones de Fluorescencia	39
6.1. Introducción	39
6.2. Diseño experimental	40
6.3. Resultados: Vida Media Fluorescente	41
6.4. Dipéptidos y tetrapéptidos	42
7. Perspectivas	45
A. Materiales y Métodos Experimentales	47
A.1. Aminoácidos y reactivos para titulaciones potenciométricas	47
A.2. Péptidos y reactivos para medidas de fluorescencia	47
A.3. Determinación de constantes de ionización	47
A.4. Mediciones de fluorescencia	48

B. Probabilidades de Formación	49
B.1. Probabilidad de formación de SS para 3 Cys	49
B.2. Probabilidad de formación de SS para 4 Cys	52

Capítulo 1

Agradecimientos

El autor agradece a las siguientes personas e instituciones por su valiosa contribución y apoyo en el proceso de completar la obtención de grado de Doctor en Ciencias Bioquímicas:

- Al Dr. Roberto Stock Silberman, por hacer el planteamiento original del proyecto de investigación que en parte se presenta en esta tesis, por haber sido mi tutor original y por haberme presentado una forma diferente de ver y pensar los problemas y modelos de la bioquímica.
- A todos en el laboratorio del Dr. Roberto Stock, por siempre apoyarme y ayudar a que la transición de la matemática a la bioquímica fuera amena y amable.
- Al Dr. Enrique Rudiño Piñera y su grupo de trabajo por todo el apoyo, la guía y el seguimiento, sin el cuál me habría sido imposible completar con éxito el doctorado.
- Al Dr. Héctor González por todo el apoyo y guía que me ha brindado desde que lo conozco, así como las largas pláticas y discusiones sobre temas de filosofía de la ciencia,
- Al Dr. Luís Bagatolli por haberme invitado y recibido en su laboratorio para realizar los experimentos necesarios para validar nuestros cálculos, así como brindarme la oportunidad de vivir una experiencia muy enriquecedora, tanto en términos académicos, como vivenciales.
- Al Dr. Henrik S. Thoke por su colaboración en los experimentos cuyos resultados se muestran en el presente trabajo, pero sobre todo, por haberme enseñado las técnicas y buenas prácticas necesarias en el laboratorio para poder realizar los experimentos sin necesidad de constante supervisión.
- Al CONACyT por apoyar parcialmente mi estancia en el programa de Doctorado en Ciencias Bioquímicas mediante la beca no. 354838, así como la beca mixta que recibí para hacer una estancia en el laboratorio del Dr. Bagatolli.

(Agradecimientos)

- A la Universidad Nacional Autónoma de México y al Instituto de Biotecnología por proveer el ambiente y espacio adecuado para estudiar y desarrollar investigación, rodeado del apoyo necesario para lograr esos objetivos.

Capítulo 2

Resumen

A pesar de que es sabido que los efectos inductivos en compuestos orgánicos influyen en propiedades químicas, tal como las constantes de ionización, la contribución específica a las propiedades y el comportamiento de aminoácidos y grupos funcionales en péptidos ha sido poco explorado. En el presente trabajo modificamos un algoritmo que permite hacer cálculos *ab initio* de la magnitud de los efectos inductivos para moléculas no aromáticas. El valor obtenido como resultado del algoritmo se denomina Índice Inductivo. Observamos una correlación alta ($R^2=0.9427$) entre nuestros cálculos y los valores de pK_a de los grupos α -amino de aminoácidos con cadenas laterales no aromáticas. Usamos una serie de aminoácidos modificados y observamos correlaciones altas ($R^2 > 0.9600$) entre los valores de Índices Inductivos y dos propiedades químicas diferentes, a saber, el pK_a de cadenas laterales ionizables y la fluorescencia del grupo indol del triptófano. Después de evaluar la aplicabilidad del algoritmo a nivel de aminoácidos, extendimos nuestro estudio a péptidos en los que el grupo indol del triptófano es el grupo reportero y pudimos observar que las contribuciones inductivas de las cadenas laterales vecinales son transmitidas a través de los enlaces peptídicos.

(Resumen)

Capítulo 3

Introducción

3.1. Motivación

El plegamiento de proteínas es uno de los problemas teóricos más importantes en la bioquímica moderna. El plegamiento oxidativo, que siguiendo la definición presentada en [3] se refiere al complejo proceso en el que la proteína llega a la conectividad nativa de puentes disulfuro, puede ser pensado como un proceso subordinado al plegamiento de la proteína. Dentro del espacio de proteínas posibles existe un pequeño grupo con poca estructura secundaria y terciaria, pero que sin embargo logra un plegamiento oxidativo cuantitativo. La ausencia de estructura secundaria/terciaria en este grupo de proteínas requiere de una explicación diferente al colapso hidrofóbico para justificar el que se logre plegamiento oxidativo cuantitativo. Nos referiremos a este grupo de proteínas como proteínas modelo.

Un primer acercamiento sugiere que, al menos para el caso de proteínas modelo, es posible estudiar el plegamiento oxidativo de manera independiente al plegamiento de la proteína. Es decir, podemos centrar nuestra atención en la formación de la conectividad nativa e ignorar la formación de las estructuras secundaria y terciaria de la proteína. Resulta necesario mencionar una suposición ampliamente aceptada en el estudio del plegamiento oxidativo, a saber, que en general las cisteínas en una proteína son químicamente indistinguibles entre sí. De manera más precisa, esto significa que si intentamos diferenciar de manera experimental entre dos o más cisteínas presentes en una proteína desnaturalizada, su reactividad será la misma, y por lo tanto, serán indistinguibles una de la otra. Esto es equivalente a decir que la única diferencia que podemos esperar (a nivel teórico), es la posición que ocupan en la proteína. Es decir, que para poder predecir la viabilidad de formación de un puente disulfuro la diferencia se limita a consideraciones estéricas. Una de las afirmaciones centrales del presente trabajo es que de hecho es posible diferenciar experimentalmente entre las cisteínas presentes en una proteína, por lo que resulta necesario contar con un modelo teórico adecuado para estudiar estas diferencias químicas, para posteriormente intentar predecir la formación de puentes disulfuro a partir de propiedades intrínsecas de la proteína,

3.2. Hipótesis (Introducción)

diferentes a meras consideraciones estéricas.

Gilbert Ling proporcionó un nuevo paradigma para estudiar los sistemas vivos [26], [27], [28] y por lo tanto abrió la puerta para poder desarrollar una serie de hipótesis novedosas sobre el estudio de estos sistemas. Las ideas de Ling sugirieron un posible camino teórico para estudiar el plegamiento oxidativo como consecuencia de un fenómeno inductivo (fenómeno químico). De manera más precisa, proponemos evaluar, mediante métodos computacionales, la contribución de los efectos inductivos de las cadenas laterales a través del esqueleto peptídico, en particular para el estudio de la formación de puentes disulfuro.

La suposición principal del presente proyecto es que el pK_a de grupos ionizables de cadenas laterales se modifica como resultado del efecto inductivo ejercido por los aminoácidos vecinales en una proteína. Esto sería en particular cierto para las cisteínas. Luego, es razonable suponer que tanto la conectividad nativa como el proceso de plegamiento oxidativo es consecuencia de un fenómeno de reactividad y selectividad química inducido por los efectos de los aminoácidos vecinales, por lo que las cisteínas en una proteína serían químicamente distinguibles como consecuencia de este efecto. Por esto, si logramos predecir de manera teórica las contribuciones inductivas de los aminoácidos vecinales, en principio sería posible predecir el cambio de propiedades de las cadenas laterales ionizables, en particular los valores de pK_a .

El trabajo de Chiang y Tai [6] proporciona una herramienta teórica que permite cuantificar (calcular) los efectos inductivos sobre un grupo reportero para una molécula, en principio de cualquier tamaño¹. Esto permite calcular y predecir los cambios de pK_a en grupos reporteros como función de los residuos vecinales en proteínas modelo.

Por lo anterior se vuelve indispensable que el presente trabajo cumpla con los siguientes requisitos:

1. mostrar que los efectos inductivos afectan el pK_a de los grupos ionizables de aminoácidos,
2. hacer los cálculos para los grupos ionizables de los aminoácidos usando el algoritmo de Chiang y Tai (CT)
3. obtener una buena correlación entre los valores calculados y las mediciones de pK_a para así,
4. intentar predecir el pK_a a partir de cálculos basados en la estructura primaria de péptidos y proteínas, usando el algoritmo CT y por último,

El presente trabajo es parte de un programa mayor que busca estudiar desde principios químicos el fenómeno del plegamiento de proteínas, y en particular, el plegamiento oxidativo.

3.2. Hipótesis

Planteamos como hipótesis operativa que las cisteínas de una proteína con

¹ Siempre que la molécula no contenga un grupo aromático

aminoácidos vecinales diferentes serán químicamente distintas², en la proteína desnaturalizada, como resultado del efecto inductivo ejercido por dichos aminoácidos vecinales. Esto será cierto hasta una distancia de separación entre la cisteína y los aminoácidos vecinales.

Esto implicaría que para dos cisteínas con aminoácidos vecinales idénticos (y en el mismo orden) esperamos observar un comportamiento similar en sus propiedades químicas. La expectativa de un comportamiento similar es porque hasta este punto no tenemos una medida precisa de la distancia, en número de residuos vecinales, en que el efecto inductivo sigue influyendo de manera significativa para hacer predicciones sobre propiedades químicas de los aminoácidos en cuestión, y en particular para las cisteínas. La razón para referirnos a la proteína desnaturalizada es para descartar los cambios en la reactividad aparente atribuible a efectos estéricos. La justificación de nuestra hipótesis es consecuencia de la siguiente predicción exitosa. La influencia ejercida por la cadena lateral de un aminoácido como consecuencia del efecto inductivo sobre los grupos α -amino y α -carboxilo es la misma para aminoácidos con cadena lateral no ionizable. Luego, si los valores de pK_a de los grupos alfa amino y carboxilo son mayoritariamente consecuencia del efecto inductivo ejercido por la cadena lateral, entonces esperaríamos observar una cierta proporcionalidad en los valores de pK_a .

Como consecuencia de las diferencias reportadas en los valores de pK_a para aminas y carboxilos no es razonable esperar que sean iguales pues, en general los grupos carboxilos son más ácidos que las aminas, además de que después de la desprotonación del grupo α -carboxilo, podemos esperar que la presencia de la carga formal negativa modifique (por efecto inductivo) el valor de pK_a del segundo grupo. Este comportamiento es común para todos los aminoácidos con cadena lateral no ionizable, por lo que la única diferencia para este grupo de aminoácidos es el efecto que cada cadena lateral ejerce sobre su respectivo esqueleto peptídico³. De manera sorprendente, en una primera exploración de los datos publicados de pK_a encontramos inconsistencias importantes en los valores reportados [14], [29], [30], [32] para el mismo grupo alfa del mismo aminoácido⁴. Es posible que sea por esta razón que la mencionada proporcionalidad no haya sido observada con anterioridad. Esto evidenció la necesidad de hacer las titulaciones correspondientes para poder validar o descartar nuestra predicción. En la figura 3.1 y la tabla 3.1 se evidencia la comprobación de nuestra predicción, a saber, que los valores de pK_a de los grupos alfa son proporcionales.

²Como se mencionó previamente, esto significa que sus propiedades químicas son distinguibles experimentalmente.

³Hasta donde sabemos, al momento de hacer esta predicción no existía una explicación satisfactoria de por qué los valores de pK_a deban ser proporcionales, o por qué existen diferencias entre los valores de pK_a de los grupos alfa de los diferentes aminoácidos.

⁴El principal problema con los valores de pK_a reportados de grupos alfa es que son recopilaciones de valores reportados por separado, algunas veces de publicaciones en la que se reporta el resultado de la titulación de uno o dos aminoácidos, por lo que es probable que la diferencia de valores se deba a la realización de las titulaciones usando diferentes instrumentos y bajo diferentes condiciones experimentales.

3.2. Hipótesis (Introducción)

Aminoácido	$\alpha - \text{NH}_3$	$\alpha - \text{COOH}$
Ala	9.80 ± 0.02	2.40 ± 0.04
Val	9.55 ± 0.01	2.33 ± 0.05
Gly	9.68 ± 0.01	2.40 ± 0.04
Ile	9.69 ± 0.01	2.38 ± 0.04
Gln	9.15 ± 0.01	2.22 ± 0.05
Met	9.15 ± 0.02	2.24 ± 0.05
Ser	9.11 ± 0.02	2.25 ± 0.05
Thr	9.01 ± 0.01	2.23 ± 0.05
Asn	8.80 ± 0.02	2.19 ± 0.05
Phe	9.15 ± 0.03	2.23 ± 0.05
Trp	9.42 ± 0.02	2.35 ± 0.04

Cuadro 3.1: Mediciones de valores de pK_a de grupos α -amino y α -carboxilo de los 12 aminoácidos con cadenas laterales no ionizables (\pm desviación estándar). Todos los valores de pK_a fueron determinados usando el mismo equipo, procedimientos y reactivos, tal como se menciona en el apéndice A.

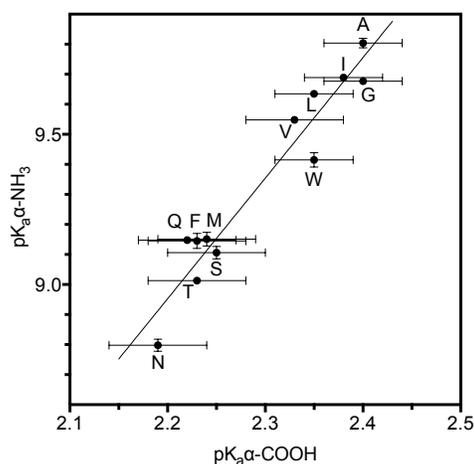


Figura 3.1: Gráfica de los valores de pK_a (\pm desviación estándar) de grupos α -amino (ordenada) contra α -carboxilo (abscisa) para todos los α -aminoácidos con cadenas laterales no ionizables junto con la recta de regresión lineal. Las barras de error más grandes para los valores de pK_a para los grupos α -carboxilo es consecuencia de la concentración mayor de iones de hidrógenos (y por lo tanto de una mayor fuerza iónica) en el punto de ionización (ver [1]).

Los efectos inductivos se definen como la polarización de un enlace causada por la diferencia de electronegatividades entre átomos de enlaces adyacentes. Estos efectos se propagan a través de la estructura de enlaces de la molécula, y son diferentes a los efectos electrostáticos o a los efectos de campo que por definición no se transmiten a través de los enlaces químicos. El objetivo del proyecto es realizar predicciones sobre las contribuciones inductivas de los aminoácidos vecinales sobre grupos reporteros adecuados. En particular proponemos correlacionar el cálculo *ab initio* de índices inductivos de Chiang y Tai [6] con los valores medidos de pK_a de grupos ionizables de los aminoácidos, para posteriormente estudiar de manera teórica las contribuciones

3.3. Contribuciones (Introducción)

inductivas sobre los grupos tioles de las cisteínas. El algoritmo CT original funciona razonablemente bien, sin embargo el ajuste (R^2) no es suficientemente bueno como para hacer predicciones *in silico*. Por lo que fue necesario hacer modificaciones al algoritmo original para intentar mejorar las correlaciones.

1. Para esto fue necesario hacer una revisión profunda del concepto de electronegatividad, central en el cálculo de índices inductivos,
2. para mejorar la correlación entre los valores de índices inductivos calculados y los valores de pK_a medidos en el laboratorio,
3. y el planteamiento formal de cálculo de la proporción de tiol/tiolato como función del pH del amortiguador (buffer) y del pK_a de las cisteínas presentes en la proteína.

A raíz del trabajo pionero de Gregorio Weber sobre la fluorescencia de aminoácidos aromáticos, principalmente el triptófano y la tirosina, es sabido que sus propiedades de fluorescencia se modifica siguiendo de cerca el estado de ionización de sus grupos alfa. Aumenta cuando los grupos se desprotonan y disminuye cuando se protonan. Esto sucede de tal manera que es posible determinar sus constantes de ionización a partir de los cambios de fluorescencia al realizar una titulación [44], [45]. Los mecanismos propuestos para explicar esta relación no han prestado suficiente atención a las contribuciones que los efectos inductivos puedan realizar, salvo algunos casos aislados [7], [8], [9]. El mencionado efecto también se ha observado y reportado para péptidos que contienen triptófano. En [15] se estudia el efecto de alejar el grupo reportero del grupo alfa ionizado al agregar residuos glicil. De estos experimentos se observa que el efecto disminuye con la distancia, medida en número de enlaces. A pesar de que en general se ha favorecido una explicación a partir de los efectos de campo, una explicación a partir de efectos inductivos es también consistente con los datos reportados.

Los resultados que se obtuvieron fueron los que se esperaban de acuerdo a las predicciones teóricas y a continuación se presenta el avance logrado.

3.3. Contribuciones

Las principales contribuciones del presente trabajo son:

1. importantes modificaciones al algoritmo propuesto por Chiang y Tai (CT),
2. la realización de experimentos para respaldar una explicación de variaciones en los valores del pK_a como consecuencia de los efectos inductivos,
3. el uso del algoritmo CT y el algoritmo CT mejorado (CTM) para predecir los cambios de pK_a en grupos alfa y grupos ionizables de cadenas laterales de alfa aminoácidos,
4. realizar experimentos para observar si los efectos inductivos se transmiten a través del esqueleto peptídico en péptidos,

3.3. Contribuciones (Introducción)

5. uso del algoritmo modificado (CTM) para hacer predicciones sobre la influencia de las contribuciones inductivas de residuos vecinales sobre grupos funcionales en péptidos,
6. uso del algoritmo modificado (CTM) para hacer predicciones sobre los cambios de fluorescencia y pK_a en grupos reporteros adecuados como consecuencia de las contribuciones inductivas presentes en la molécula.

Capítulo 4

Cálculo de Proporciones de Tiol/Tiolato en Proteínas Modelo

4.1. Probabilidades individuales

Como primer paso en el estudio del comportamiento de las cisteínas durante el proceso de plegamiento oxidativo estudiaremos la probabilidad de encontrar una cisteína en el estado protonado/desprotonado¹. Para esto definiremos dicha probabilidad como función del pH del medio y del pK_a de la cisteína. Por la ecuación de Henderson-Hasselbalch se tiene que para una cisteína Cys_1 ,

$$pH = pK_{a1} + \log \frac{[RS^-]}{[RSH]}, \quad (4.1)$$

en donde pK_{a1} se refiere al pK_a de la cisteína, $[RS^-]$ y $[RSH]$ denotan las concentraciones de la cisteína desprotonada y protonada respectivamente. Despejando de la ecuación 4.1 se obtiene que

$$\begin{aligned} pH &= pK_{a1} + \log \frac{[RS^-]}{[RSH]} \\ pH - pK_{a1} &= \log \frac{[RS^-]}{[RSH]} \\ 10^{(pH - pK_{a1})} &= \frac{[RS^-]}{[RSH]}. \end{aligned} \quad (4.2)$$

¹El desarrollo y las ecuaciones en el presente capítulo fueron desarrolladas completa e independientemente por el autor. La ecuación 4.6 aparece sin justificación en [46]. Considerando que la publicación precede por tres décadas al presente trabajo, el autor considera adecuado mencionar que el resultado ya era conocido.

4.2. Probabilidad de formación de puentes disulfuro (Cálculo de Proporciones de Tiol/Tiolato en Proteínas Modelo)

Para poder asignar una probabilidad a partir de esta ecuación sería necesario que las concentraciones cumplieran con lo siguiente

$$[RSH] + [RS^-] = 1. \quad (4.3)$$

Observemos que si reexpresamos la ecuación 4.3, no como concentraciones sino como la fracción molar χ_H de RSH y χ_- de RS^- , entonces en efecto se tiene que

$$\chi_H + \chi_- = 1. \quad (4.4)$$

Sustituyendo de manera apropiada χ_H por $[RSH]$ y χ_- por $[RS^-]$ en la ecuación 4.2 y observando que de la ecuación 4.4 se sigue que $\chi_- = 1 - \chi_H$, entonces

$$\begin{aligned} 10^{(pH-pK_{a1})} &= \frac{1 - \chi_H}{\chi_H} \\ 10^{(pH-pK_{a1})} \chi_H &= 1 - \chi_H \\ \chi_H + 10^{(pH-pK_{a1})} \chi_H &= 1 \\ \chi_H (1 + 10^{(pH-pK_{a1})}) &= 1 \\ \chi_H &= \frac{1}{1 + 10^{(pH-pK_{a1})}}. \end{aligned} \quad (4.5)$$

Por último, sustituyendo este resultado en la ecuación 4.4 se tiene que

$$\chi_- = 1 - \frac{1}{1 + 10^{(pH-pK_{a1})}}. \quad (4.6)$$

Las ecuaciones 4.5 y 4.6 nos dan la fracción molar de RSH y RS^- como función del pH y con parámetro pK_{a1} . Como estos números están entre 0 y 1, esto es igual a la probabilidad de encontrar a la Cys_1 en alguno de los dos estados. Luego, podemos definir

$$P(RSH) = \chi_H, \quad (4.7)$$

y

$$P(RS^-) = \chi_-. \quad (4.8)$$

Se sigue del desarrollo anterior que si podemos predecir de manera exitosa el pK_a de una cisteína en una proteína, entonces podemos calcular la probabilidad de que se encuentre protonada/desprotonada y, por lo tanto, su disponibilidad para formar un puente disulfuro con otra cisteína que se encuentre en el estado de protonación adecuado ².

4.2. Probabilidad de formación de puentes disulfuro

Notación: Para una cisteína Cys_i , denotaremos por RSH_i y RS_i^- sus estados protonado y desprotonado respectivamente. Al puente disulfuro formado entre Cys_i y Cys_j

²En [47] se presenta un estudio sobre la formación de puentes disulfuro en péptidos como función de la distancia en número de residuos entre dos cisteínas mientras que en [46] aparece la ecuación 4.5 como afirmación, sin desarrollo o justificación.

4.2. Probabilidad de formación de puentes disulfuro (Cálculo de Proporciones de Tiol/Tiolato en Proteínas Modelo)

lo denotaremos por SS_{ij} . A la probabilidad de RSH_i la denotaremos por P_{iH} y la probabilidad de RS_i^- por P_{i-} .

Nota: En adelante, y a menos de que explícitamente se mencione lo contrario, al referirnos a dos cisteínas Cys_i y Cys_j nos estaremos refiriendo a dos Cys en el mismo péptido o proteína.

Supongamos que tenemos dos cisteínas Cys_1, Cys_2 cuyas probabilidades de estar protonadas/desprotonadas están dadas por las ecuaciones 4.5 y 4.6. Entonces, asumiendo que no existen otras fuerzas que determinen la formación del puente SS_{12} , éste se puede formar mediante la interacción de RSH_1 con RS_2^- si ambas cisteínas se encuentran en el estado adecuado. Observemos que si suponemos que el estado de Cys_i es independiente del estado de Cys_j , entonces la probabilidad condicional de que ambas Cys estén en el estado adecuado para formar un puente SS se reduce al producto de sus probabilidades individuales³, luego la probabilidad de encontrar a las Cys en el estado adecuado para formar un puente SS como función de pH y con parámetros pK_{ai} estará dada por:

$$P(SS_{12}) = (P_{1H} \cdot P_{2-}) + (P_{1-} \cdot P_{2H}), \quad (4.9)$$

En donde los P_{iH} y P_{i-} estarán dados por las ecuaciones 4.7 y 4.8. Si además suponemos que siempre que dos Cys estén en el estado adecuado formarán un puente SS, entonces la ecuación 4.9 nos da precisamente esa probabilidad.

Esto equivale a suponer que:

- 1) el experimento es independiente del tiempo (o de manera alternativa que el sistema se encuentra en equilibrio),
- 2) el estado de protonación de las cisteínas es independiente para cada una, salvo en el caso de que sean vecinales (y como ya se mencionó, esto se considera en nuestro análisis),
- 3) que todos los puentes que se pueden formar lo van a hacer ,
- 4) que la formación de un puente (SS) sólo depende de la proporción de cisteínas que se encuentran en el estado adecuado (una protonada y la otra desprotonada),
- 5) estas probabilidades representan de manera teórica (y bajo las 3 suposiciones anteriores) la proporción de puentes que se formarían si no fuera posible el rearreglo⁴.

Empíricamente sabemos que las suposiciones anteriores no son compatibles con la realidad pues 1) los experimentos no son independientes del tiempo, 2) no todos los

³Para el caso en que las cisteínas sean vecinales, la expresión se vuelve más compleja pues es necesario contemplar la influencia inductiva mutua. Este desarrollo se contempla para un trabajo futuro.

⁴Rearreglo significa que una vez que se forma un puente disulfuro, puede ser modificado mediante la reacción con una tercera cisteína, modificando de esta forma la conectividad que se puede observar en un instante dado del experimento.

4.2. Probabilidad de formación de puentes disulfuro (Cálculo de Proporciones de Tiol/Tiolato en Proteínas Modelo)

puentes que son teóricamente posibles se forman (como en el caso de BPTI⁵), 3) la necesidad de tioles catalíticos durante el plegamiento oxidativo para la mayoría de las proteínas modelo y, 4) sabemos que existe rearreglo. Luego, los cálculos de la probabilidad de formación de puentes disulfuro muestran lo que podemos esperar en una situación en la que la formación sólo depende del estado de protonación de las cisteínas. La discrepancia entre los cálculos y los resultados experimentales sugieren la necesidad de incluir, o al menos considerar, hipótesis adicionales (ver capítulo 7).

A continuación se presentan las gráficas de las probabilidades para dos cisteínas. La idea es presentar un análisis de los diferentes estados que podemos esperar, suponiendo que nos es posible modificar los valores de pK_a de las cisteínas correspondientes. Cada gráfica muestra la probabilidad de encontrar dos cisteínas en el estado de protonación/desprotonación adecuado para formar un puente disulfuro como función de los valores de pK_a de cada cisteína.

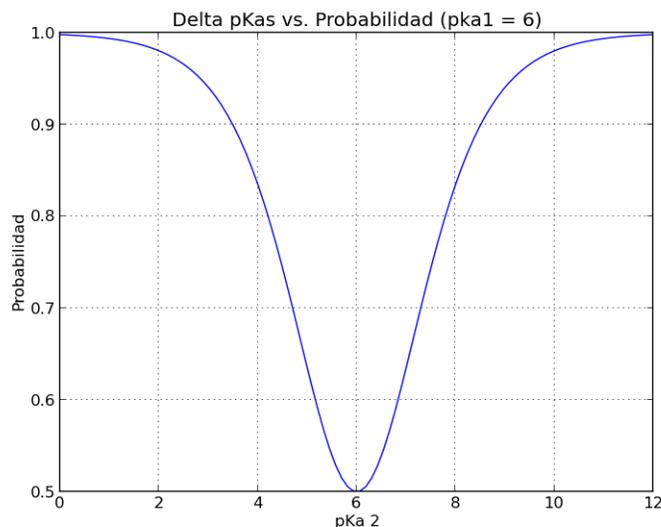


Figura 4.1: En esta figura suponemos la existencia de una cisteína Cys_1 con pK_a igual a 6, mientras que suponemos que el de la Cys_2 varía desde 0 hasta 12. Los valores de la gráfica muestran la probabilidad de encontrar ambas Cys en el estado de ionización adecuado para formar un puente disulfuro. La probabilidad se calcula usando la ecuación 4.9.

⁵En principio la formación del puente $SS = (14 - 38)$ debe tener probabilidad no nula, sin embargo no aparece de manera cuantitativa dentro de las rutas observadas en los experimentos de plegamiento oxidativo [3].

4.2. Probabilidad de formación de puentes disulfuro (Cálculo de Proporciones de Tiol/Tiolato en Proteínas Modelo)

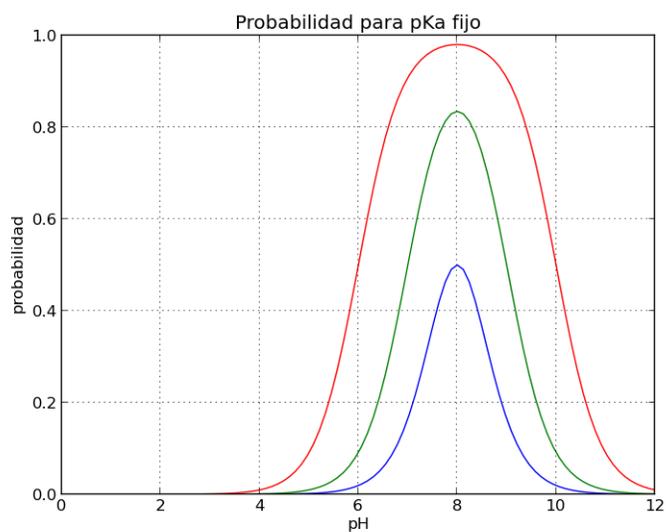


Figura 4.2: En esta gráfica se representan las probabilidades de formación de un puente disulfuro para dos Cys con valores simétricos con respecto a su pK_a y como función del pH. La gráfica azul es de dos Cys con pK_a igual a 8, la gráfica verde de dos Cys con pK_a s de 7 y 9 respectivamente, y la gráfica roja con pK_a s de 6 y 10. Observar que el máximo se alcanza en el promedio de ambos pK_a s.

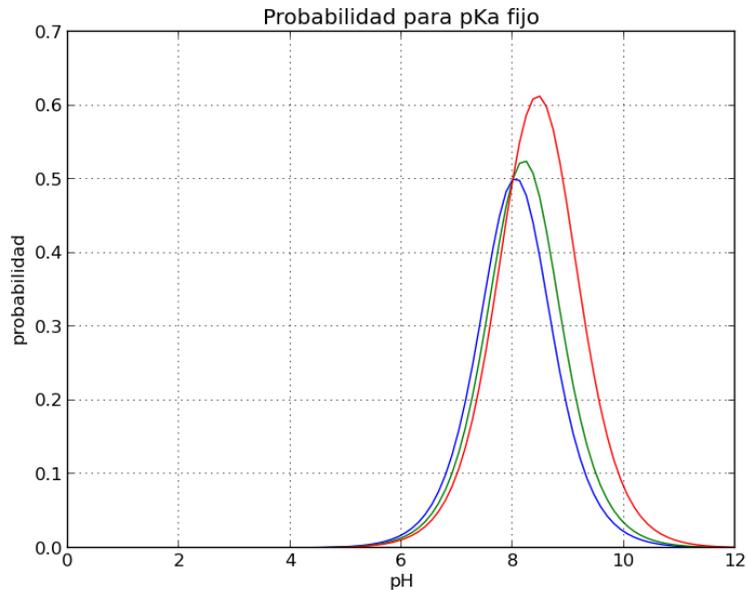


Figura 4.3: Gráficas para dos Cys (un sólo puente SS posible) en donde el pK_a de una de las Cys es de 8 y el de las demás varía de la siguiente manera: azul = 8.1, verde = 8.4 y rojo = 8.9

4.2. Probabilidad de formación de puentes disulfuro (Cálculo de Proporciones de Tiol/Tiolato en Proteínas Modelo)

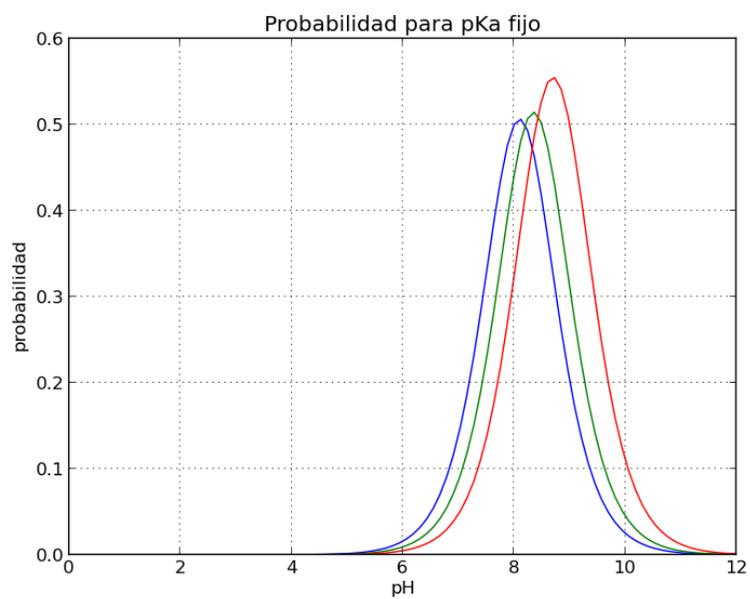


Figura 4.4: Gráficas para dos Cys (un sólo puente SS posible) en donde los pK_a s varían de la siguiente manera: azul = 8 y 8.2; verde = 8.2 y 8.5; rojo = 8.4 y 9.

Capítulo 5

El Algoritmo CT

5.1. El algoritmo CT

El algoritmo original de Chiang y Tai [6] permite calcular un valor numérico denominado *Índice Inductivo* (*ii*) sobre un átomo de una molécula dada, y que es una cuantificación de los efectos inductivos ejercidos por el resto de los átomos presentes en la molécula sobre el átomo en cuestión. Recordemos que los efectos inductivos son las modificaciones de la nube de electrones de cada átomo que conforma una molécula como resultado de las diferencias de electronegatividad de estos átomos. Como se explica en [35], esto impacta en las propiedades químicas de los restos¹, y como se muestra en [6], los cálculos de *ii* correlacionan con diversas medidas experimentales. Se sigue entonces que si tenemos una manera de cuantificar los efectos inductivos, en principio, podríamos predecir de manera cuantitativa los cambios en las propiedades químicas de los restos, y en general de los sitios activos de una molécula. En el artículo original de Chiang y Tai [6] se muestran ajustes lineales entre el valor calculado de *ii* con diferentes propiedades químicas, por lo que resulta razonable esperar que el algoritmo modificado CTM no sólo presente ajustes lineales con los datos obtenidos en el laboratorio, sino que esperamos que las modificaciones mejoren de manera importante el ajuste entre los cálculos y los datos experimentales. El objetivo de contar con una herramienta de cálculo mejor calibrada es el poder usar el resultado de los cálculos en sustitución de datos experimentales, que en general son más costosos y difíciles de obtener. Es por esta razón que enfocamos nuestros esfuerzos en mejorar la correlación entre valores medidos y calculados, al punto en que nos resultara posible usar los valores calculados con suficiente confianza de que presentarían una fuerte correlación con los valores experimentales².

El cálculo original de *ii* depende de tres parámetros, a saber, la electronegatividad de los átomos presentes en la molécula, la longitud de cada enlace presente en la molécula y un valor constante llamado *factor de transmisividad* que es igual a $\frac{1}{\alpha}$, en

¹En inglés, *moiety*.

²En el capítulo 7 se discute un posible siguiente paso en el uso de estos datos calculados para intentar predecir conectividades nativas en proteínas modelo.

5.2. Introducción (El Algoritmo CT)

donde $\alpha = 2.7$. El valor de ii se calcula como la suma de tres componentes (valores) diferentes. i_0 representa la contribución del átomo adyacente al átomo sobre el que se realiza el cálculo. i representa las contribuciones del resto de los átomos de la molécula y por último, i_{\pm} es la suma de las contribuciones de las cargas formales presentes en la molécula. Estos tres valores dependen de los parámetros antes mencionados, a saber, la electronegatividad, la longitud de los enlaces y el factor de transmisividad. A continuación se define formalmente cada uno de los valores:

$$i_0 = \frac{\delta_{AB}}{r_{AB}},$$

en donde r_{AB} representa la longitud del enlace entre los átomos A y B medida en Ångstroms (Å), y tal que δ_{AB} es

$$\delta_{AB} = \frac{\chi_B - \chi_A}{\chi_B + \chi_A},$$

y χ_A representa la electronegatividad del átomo A . Sea l_j el número de enlaces de orden³ j , entonces

$$i = \frac{1}{\alpha} \sum_1^{l_1} \left(\frac{\delta}{r}\right)_1 + \left(\frac{1}{\alpha}\right)^2 \sum_1^{l_2} \left(\frac{\delta}{r}\right)_2 + \left(\frac{1}{\alpha}\right)^3 \sum_1^{l_3} \left(\frac{\delta}{r}\right)_3 + \cdots + \left(\frac{1}{\alpha}\right)^n \sum_1^{l_n} \left(\frac{\delta}{r}\right)_n$$

en donde los δ correspondientes son las diferencias ponderadas entre las electronegatividades de los átomos involucrados en los enlaces en el orden correspondiente.

$$i_{\pm} = \left(\frac{\pm N_1}{r_{\pm}}\right)_0 + \frac{1}{\alpha} \sum_1^{l_1} \left(\frac{\pm N_2}{r_{\pm}}\right)_1 + \left(\frac{1}{\alpha}\right)^2 \sum_1^{l_2} \left(\frac{\pm N_3}{r_{\pm}}\right)_2 + \cdots + \left(\frac{1}{\alpha}\right)^n \sum_1^{l_n} \left(\frac{\pm N_{n+1}}{r_{\pm}}\right)_n,$$

en este caso r_{\pm} es la longitud del enlace en el grupo ionizado, y $\pm N$ es igual a 1 o -1 (i.e. los valores correspondientes de las cargas formales)⁴.

5.2. Introducción

El algoritmo de cálculo de índices inductivos *ab initio* de Chiang y Tai (CT) presenta un problema sistémico, a saber la incapacidad del algoritmo CT para distinguir los diferentes estados en los que puede aparecer un átomo en una molécula, i.e. su estado de oxidación y/o hibridación. La consecuencia de este problema es que la correlación entre valores de pK_a de los grupos alfa amino y el valor ii calculado mediante el algoritmo CT resulta poco satisfactoria⁵.

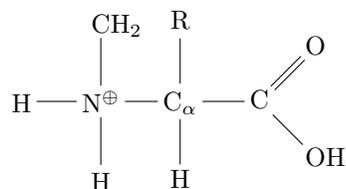
³El *orden* se refiere a la distancia, en número de enlaces, entre el átomo sobre el que se realiza el cálculo de índice inductivo, y el enlace del cual se está considerando su contribución. El exponente y el subíndice de este componente del cálculo indica la distancia en número de enlaces con respecto al átomo sobre el que se está realizando el cálculo. r es la longitud del enlace correspondiente.

⁴Observemos que se puede usar una notación más precisa para estos valores, a saber, representar cada valor como función de los tres parámetros mencionados, es decir, $i_0(\chi_j, \alpha, \delta_j) + i(\chi_j, \alpha, \delta_j) + i_{\pm}(\chi_j, \alpha, \delta_j)$, sin embargo consideramos que el contexto es suficiente para clarificar la dependencia de los valores i_0, i, i_{\pm} de los parámetros mencionados, por lo que evitamos el uso de esta notación más extensa.

⁵con un R^2 menor a 0.6 para ser más precisos.

5.2. Introducción (El Algoritmo CT)

Una consecuencia de la mencionada incapacidad es que el algoritmo no permite distinguir entre una amina primaria y una secundaria, lo que resulta especialmente relevante para el cálculo eficiente de índices inductivos para la prolina en proteínas modelo. Esto se traduce en valores calculados poco confiables y que presentan una correlación con las mediciones de pK_a que no es suficiente para hacer predicciones teóricas.



Estructura del esqueleto peptídico de la Prolina.

La prolina presenta un problema especial en el cálculo de índices inductivos pues, a diferencia del resto de los aminoácidos, la cadena lateral se enlaza al esqueleto peptídico en dos puntos y por lo tanto hace una contribución inductiva en dos puntos diferentes. Esto convierte al grupo alfa amino en una amina secundaria que tiene una carga formal positiva. Por lo anterior, el grupo alfa amino de la prolina resulta especialmente complicado de tratar mediante el algoritmo CT cuando es el grupo reportero⁶. La consecuencia que esto tiene es que los cálculos para la prolina no puedan ser comparados con plena confianza con los cálculos para el resto de los aminoácidos⁷. Como ya se mencionó, los valores de i_i dependen de tres parámetros, así que nuestra estrategia inicial para intentar mejorar las correlaciones entre valores calculados y valores medidos fue probar haciendo modificaciones en alguno de los tres parámetros, en este caso se exploró el efecto de usar diferentes escalas de electronegatividad. En principio no hay razón para comprometerse con la escala de Pauling para realizar los cálculos. Después de explorar diferentes escalas, la conclusión fue que no había razón para preferir la escala de Pauling sobre otras disponibles. Después de revisar y comparar diferentes escalas de electronegatividad, nos inclinamos por usar la escala de electronegatividad que se deriva del concepto de electronegatividad de orbital, tal como lo presentan Hinze y Jaffé [22]. La razón principal para usar esta escala es que en lugar de asignar un valor único a cada átomo, asigna diferentes valores, dependiendo del estado de los orbitales. Esto nos permitió refinar los cálculos, y eventualmente sugirió una de las modificaciones adicionales. Después de fijar los valores de electronegatividad a usar, organizamos los aminoácidos en función de su parecido estructural para ver de qué manera sería necesario modificar el algoritmo CT para mejorar el ajuste entre los datos experimentales y los valores calculados. Para esto fuimos observando cómo se ajustaban nuestros nuevos valores de electronegatividad para cadenas laterales progresivamente más complejas.

⁶Es decir que los cálculos se hacen sobre el grupo alfa amino, además de las mediciones de pK_a sobre ese grupo.

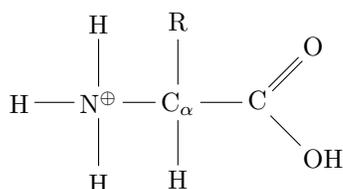
⁷Las modificaciones que más adelante proponemos al algoritmo CT no resuelven plenamente los problemas mencionados con respecto a la Prolina, sin embargo, este fue el primer problema que sugirió que era necesario hacer modificaciones importantes en el algoritmo CT.

5.3. Invariantes estructurales (El Algoritmo CT)

A continuación se presenta el orden adoptado para los grupos de aminoácidos. En el primer caso se trabajó con los aminoácidos alifáticos y con la glicina (V,A,L,I,G). Se realizaron los cálculos con el algoritmo CT original y eso se graficó contra el valor medido de pK_a sobre el grupo α -amino. Después de hacer alguna modificación al algoritmo, se volvían a realizar los cálculos para ver si la correlación había mejorado. Cada que logramos alguna mejora, se conservaba la modificación y se agregaba el siguiente grupo de aminoácidos, de tal forma que el grupo 2 contiene a los aminoácidos del grupo 1 más los propios del grupo 2. Es decir, cada caso contiene a todos los aminoácidos del caso anterior más los del nuevo grupo⁸. A lo largo del texto nos referiremos al caso correspondiente por el número en el que aparece:

1. alifáticas + Gly (5 aa)
2. + Ser y Thr (7 aa)
3. + Asn y Gln (9 aa)
4. + Met (10 aa)
5. + 5 ionizables (15 aa)
6. + Pro (16 aa)

Para simplificar la etapa de cálculo sólo se consideró la contribución de la cadena lateral sobre el carbono alfa, ignorando por el momento la contribución del esqueleto peptídico (BB) que es constante para los 15 aminoácidos que podemos calcular.



El objetivo es mejorar tanto como sea posible la correlación entre los valores de i_i calculados con las modificaciones propuestas al algoritmo CT con nuestras mediciones de pK_a sobre el grupo alfa amino. El mejorar la correlación nos indica que las modificaciones que realizamos nos dan una herramienta de cálculo mejorada, en comparación con el algoritmo original.

5.3. Invariantes estructurales

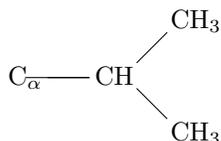
La suposición inicial presente en el algoritmo CT, y por lo tanto en nuestro trabajo, es que un grupo CH_3 tiene la misma contribución en donde sea que aparezca, y que lo único que se modifica es el orden (la distancia en número de enlaces), mediante la cual

⁸Cuando consideramos que era necesario hacer alguna modificación en los valores de electronegatividad, esa modificación se conservaba para los casos más complejos. Es decir que las modificaciones realizadas en el algoritmo CT se aplican de manera consistente para todos los aminoácidos.

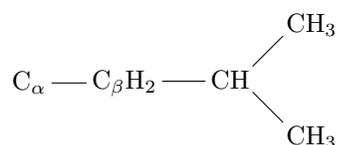
5.4. La aparente paradoja (El Algoritmo CT)

ejerce su contribución. Luego, la diferencia en la contribución de acuerdo al algoritmo original sólo puede variar por efecto de la distancia. En otras palabras, que la intensidad del efecto depende de la distancia pero no de la estructura.

A manera de ejemplo, el grupo



de la cadena lateral de la valina es igual que el mismo grupo de la leucina,

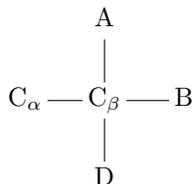


pero alejado en un orden del carbono alfa por el carbono beta de la leucina. La suposición implícita en el algoritmo CT es que la contribución es constante salvo en el caso de grupos ionizables, y que sólo aplica para el quinto caso.

Las cadenas laterales de los aminoácidos alifáticos muy pronto evidenciaron la necesidad de modificar este supuesto.

5.4. La aparente paradoja

Los casos que estaremos revisando tienen la siguiente forma:



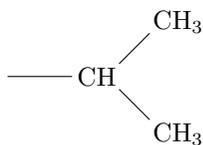
en donde A, B y D pueden representar un átomo (hidrógeno, oxígeno, etc.) o un grupo funcional (CH_3 , CH_2 , OH , etc.).

El contraste entre los valores de $\text{p}K_a$ de la alanina y la valina llamó nuestra atención por la incompatibilidad con los valores esperados. De acuerdo a la hipótesis inductiva, la modificación en el $\text{p}K_a$ del grupo alfa amino (o carboxilo) viene dada por las propiedades inductivas de la cadena lateral, a saber, la capacidad de donar o sustraer electrones. En este caso el $\text{p}K_a$ del grupo amino de la alanina es 9.80, el más alto de los alifáticos, Lo que significaría que el grupo metilo (CH_3) de la cadena lateral de la alanina



es donador de electrones. Luego, parece razonable esperar que para la valina, con un grupo metilo más que la alanina

5.4. La aparente paradoja (El Algoritmo CT)

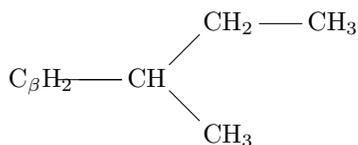


el pK_a sea parecido, o inclusive mayor. Sin embargo el pK_a del grupo amino de la valina es de 9.54, el más bajo de los aminoácidos alifáticos. Luego, los datos parecen sugerir que un grupo metilo es donador de electrones pero dos grupos son sustractores, lo que suena más a una hipótesis *ad hoc* que a una explicación de lo que realmente sucede. Como este no es el único caso en el que los datos experimentales parecen contradecir los supuestos básicos del presente proyecto, las mediciones experimentales hicieron evidente la necesidad de refinar o modificar nuestros supuestos.

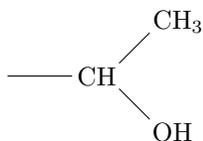
A continuación se exponen los demás casos pues la consistencia en la distribución de los valores experimentales de pK_a de grupos amino guiaron el desarrollo de las modificaciones que permitieron mejorar las correlaciones entre las predicciones y los datos experimentales.

El siguiente caso que consideramos fue el de la leucina, que podemos pensar como una valina pero con el grupo C_3H_7 alejado en un orden por el $C_\beta H_2$. Luego, en este caso parece que el grupo C_3H_7 efectivamente funciona como un donador de electrones, pero una vez más, el grupo donador se encuentra más alejado que para el caso de la valina, luego, alejar el grupo donador aumenta su capacidad donadora. Esto es una clara contradicción pues en [6], Chiang y Tai muestran que a mayor distancia (i.e. mayor número de enlaces que debe transitar el efecto) se aprecia una disminución importante de los efectos inductivos.

En resumen, el grupo CH_3 parece ser donador, pero el grupo C_3H_7 parece ser sustractor, a pesar de tener dos grupos CH_3 . Para concluir el argumento, observemos que el grupo amino de la isoleucina tiene un pK_a de 9.68 y su cadena lateral la podemos pensar como una valina, sólo que con un grupo CH_3 en lugar de un hidrógeno. De nuevo, al parecer alejar el grupo donador aumenta la capacidad donadora, lo cual es una clara contradicción.

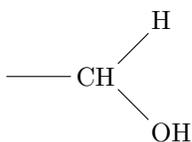


El comportamiento mencionado no se limita a los aminoácidos alifáticos. Observemos el caso de la serina:



y la treonina:

5.4. La aparente paradoja (El Algoritmo CT)



Aquí la diferencia es también un grupo metilo en lugar de un hidrógeno, por lo que esperaríamos que subiera el pK_a (por el carácter de donador de electrones del grupo metilo), sin embargo, una vez más, el pK_a de la treonina es más bajo por 0.1 unidades de pH .

La aparente paradoja sobre la contribución del grupo metilo y los valores de pK_a medidos se resuelve agregando el siguiente refinamiento a nuestros supuestos iniciales. Observemos que en realidad las cadenas laterales (alifáticas o no) están formadas por tres grupos diferentes de carbonos, a saber, los que aparecen en los grupos CH_3 , CH_2 y CH , y que el orden⁹ en el que aparecen determina su contribución.

Un grupo CH_3 efectivamente es un buen donador de electrones, siempre y cuando esa donación no tenga que pasar por un carbono terciario adyacente que amortigüe su efecto como donador. De hecho, parecería que cuando esto sucede, el efecto es el opuesto y no sólo se ve amortiguado el efecto donador, sino que el sistema $CH - CH_3$ se vuelve sustractor de electrones. Considerando que el valor ii arrojado por algoritmo CT es una función de la longitud de enlace (r_{ab}), la electronegatividad (χ_a) y la transmisividad (α), los ajustes necesarios se deben hacer en alguno de estos valores. Primero evaluamos qué tan pronunciado era el efecto de la longitud de enlace en el cálculo final. Para aminoácidos, todos los valores de r_{ab} fluctúan en un rango reducido, cercano a la unidad y, en general, no se percibió una mejora en la correlación final, con o sin el valor de la longitud de enlace como un parámetro más del algoritmo CT. Por esta razón se decidió adoptar una versión simplificada que sólo depende de la transmisividad (α) y de la electronegatividad (χ_a).

La observación de que un grupo $CH - CH_3$ se vuelve un sustractor de electrones sugirió el primer cambio de valores de electronegatividad pues eso significa que la nube electrónica es más densa hacia el CH_3 y por lo tanto, la electronegatividad del carbono CH debe ser menor. En la tabla 5.2 se presenta la lista completa de valores de electronegatividad usados en nuestros cálculos, y que son una versión modificada de los propuestos por Hinze y Jaffé en [22].

Es necesario mostrar que la estrategia usada para los otros 15 aminoácidos no puede ser aplicada a la prolina pues no es posible calcular de manera independiente la contribución de su cadena para después sumarle la contribución de la unidad BB. En este caso se piensa en la contribución de toda la molécula sobre los grupos (alfa) ionizables y se compara con los cálculos para los otros 15 aminoácidos más su unidad BB, i.e. la contribución de la molécula completa a sus grupos (alfa) ionizables. Con estas modificaciones en el algoritmo CT mediante la eliminación del valor de longitud de enlace y los ajustes en los valores de electronegatividad lograron una mejora sustancial¹⁰ en la correlación de los valores de pK_a e ii . Esta mejora permitió refinar los cálculos de ii

⁹En este caso es necesario aclarar que la palabra *orden* se refiere a la distancia en número de enlaces, de acuerdo a la forma en la que Chiang y Tai usan el término *orden* en [6], y no necesariamente se refiere a que los grupos, átomos o enlaces respeten un ordenamiento determinado.

¹⁰Ver la sección 5.5.

en péptidos para, de esta manera, tener datos calculados más confiables al momento de hacer predicciones.

5.5. Mejoras al Algoritmo CT

Se realizaron cuatro modificaciones importantes para mejorar el Algoritmo original de Chiang y Tai, y al que en adelante nos referimos como el algoritmo de Chiang y Tai Modificado (CTM)¹¹. Todas estas modificaciones se fueron haciendo de manera gradual (una a la vez) hasta lograr optimizar el ajuste. A continuación se describen los cambios:

1. usamos la escala de electronegatividad de Hinze-Jaffé en lugar de la escala de Pauling,
2. dejamos de considerar las longitudes de enlace en los denominadores de los términos que componen el algoritmo,
3. modificamos el valor de las cargas formales dando un valor de +1.2 a las positivas, y de -0.9 a las negativas,
4. y modificamos el valor del factor de transmisividad a 0.5263 ($\alpha = 1.9$).

La expresión para el cálculo del CTM queda de la siguiente forma: el Índice Inductivo (i) se compone de las mismas tres magnitudes que el algoritmo original:

$$i_0 + i + i_{\pm},$$

tal que $i_0 = \delta_{AB}$. Para la contribución de los átomos con orden mayor a 1 la expresión queda igual, salvo por la desaparición de las longitudes de enlace¹²:

$$i = \frac{1}{\alpha} \sum_1^{l_1} (\delta)_1 + \left(\frac{1}{\alpha}\right)^2 \sum_1^{l_2} (\delta)_2 + \left(\frac{1}{\alpha}\right)^3 \sum_1^{l_3} (\delta)_3 + \cdots + \left(\frac{1}{\alpha}\right)^n \sum_1^{l_n} (\delta)_n.$$

Para las contribuciones de las cargas formales, la expresión tuvo la misma modificación que para el caso de las contribuciones de orden superior (i):

$$i_{\pm} = (\pm N_1)_0 + \frac{1}{\alpha} \sum_1^{l_1} (\pm N_2)_1 + \left(\frac{1}{\alpha}\right)^2 \sum_1^{l_2} (\pm N_3)_2 + \cdots + \left(\frac{1}{\alpha}\right)^n \sum_1^{l_n} (\pm N_{n+1})_n.$$

Con esas modificaciones se logró mejorar de manera considerable la correlación (R^2) entre nuestros cálculos y los datos experimentales. A continuación se muestra la comparación de los resultados previos con el algoritmo CT y con el algoritmo CTM.

¹¹La "M" también puede significar "Mejorado".

¹²De manera alternativa se puede interpretar como que consideramos todas las longitudes iguales a 1.

5.5. Mejoras al Algoritmo CT (El Algoritmo CT)

AA	pK_a α -amino	I Std	I Std + BB	I Re Mull	I Re Mull + BB
Val	9.54	-0.0628	1.1544	-0.0600	1.1464
Leu	9.63	-0.0622	1.1544	-0.0263	1.1510
Gly	9.67	-0.0526	1.1558	-0.0531	1.1473
Ile	9.68	-0.0636	1.1543	-0.0395	1.1492
Ala	9.80	-0.0584	1.1550	-0.0943	1.1417
Ser	9.11	0.0154	1.1651	0.0781	1.1653
Thr	9.01	0.0277	1.1668	0.0907	1.1671
Asn	8.80	0.0487	1.1697	0.1567	1.1762
Gln	9.15	-0.0209	1.1602	0.0539	1.1621
Met	9.15	-0.0503	1.1561	0.0369	1.1597
Asp	9.76	-0.0694	1.1535	-0.0817	1.1435
Glu	9.74	-0.0540	1.1556	-0.0344	1.1499
Cys	10.33	-0.5760	1.0840	-0.2399	1.1219
Lys	9.15	-0.0418	1.1573	0.0156	1.1568
Arg	9.08	-0.0515	1.1560	0.0625	1.1632
Pro	10.64	1.0245	1.0245	1.1205	1.1205

Cuadro 5.1: Valores de pK_a y cálculos de Índices Inductivos (ii) para cada aminoácido. I Std se refiere al cálculo de la contribución de la cadena lateral sobre el carbono alfa con el algoritmo CT; I Std + BB se refiere al cálculo anterior más la contribución del esqueleto peptídico. En ambos casos se usa la escala de electronegatividad de Pauling. I Re Mull se refiere al cálculo de la contribución de la cadena lateral sobre el carbono alfa; I Re Mull + BB se refiere al mismo cálculo más la contribución de la unidad constante del esqueleto peptídico. En ambos casos se usa la escala de electronegatividad de Hinze-Jaffé y se omite el uso de la longitud de enlace. Para el caso de la prolina se calcula la contribución de toda la molécula sobre el grupo α -amino y se reporta este valor en ambas columnas. En las columnas marcadas como Re Mull se usó el algoritmo CTM.

5.5. Mejoras al Algoritmo CT (El Algoritmo CT)

Elemento	Pauling	Hinze-Jaffé
C_α	tetra	tetra
C tetra	2.48	15.95
C tri		11.19
CH_3		17
CH_2		14.5
CH		13
O tetra	5.93	30.5
O tri		33.47
O^-	4.43	14
N tetra	3.68	23.08
N sec		35
N^+	5.18	24
S tetra	3.21	20.27
S sec		45
S^-	2.93	25
H	2.21	14.34

Cuadro 5.2: Comparación de los valores de electronegatividad usados en el algoritmo CT tomados de la escala de Pauling contra los valores usados en el algoritmo CTM tomados de la escala de Hinze-Jaffé. En negritas aparecen los valores que se agregaron a los presentados por Hinze-Jaffé a partir de lo presentado en la sección 5.3.

Nota: Las siguientes gráficas muestran las correlaciones entre los valores calculados por el algoritmo CT y el CTM para grupos alfa de aminoácidos contra el valor medido de pK_a del mismo grupo. Los casos siguen la misma numeración que la lista que aparece en la página 24.

5.5. Mejoras al Algoritmo CT (El Algoritmo CT)

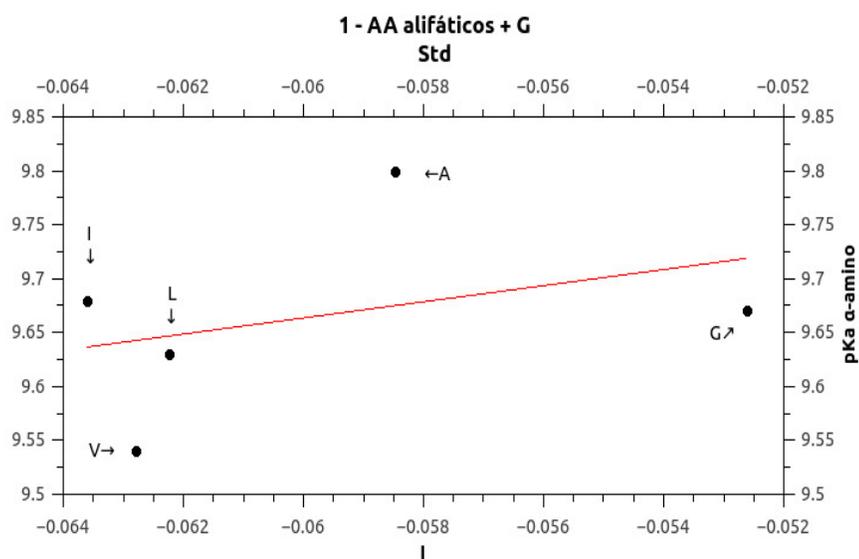


Figura 5.1: Gráfica del caso 1 con $R^2 = 0.13$. Algoritmo CT. Además de la baja correlación, la pendiente de la recta indica una correlación negativa. En este caso los aminoácidos más alcalinos tienen valores de I mayores, en contra de lo que debe ocurrir.

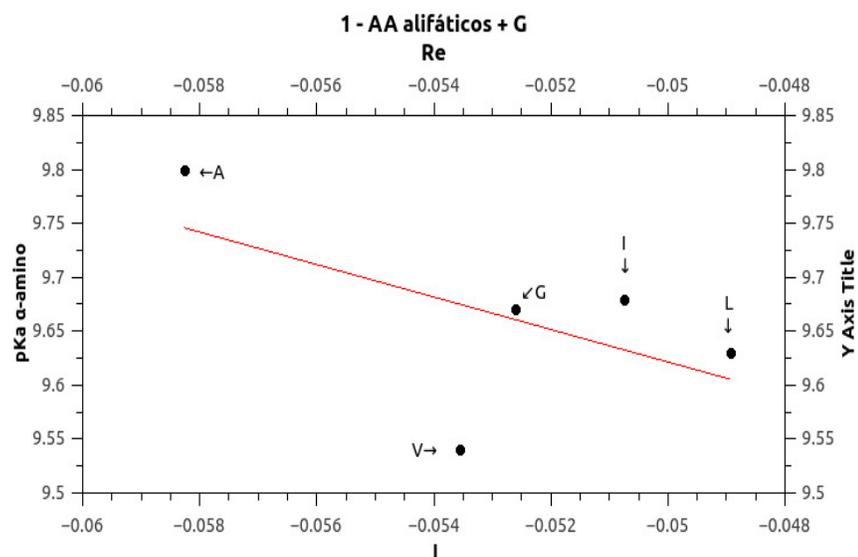


Figura 5.2: Gráfica del caso 1 con los valores modificados y con $R^2 = 0.32$. Algoritmo CTM. Una baja correlación que sin embargo es una mejora considerable con respecto a la que se tenía antes de hacer las modificaciones pues ahora la pendiente indica una correlación positiva.

5.5. Mejoras al Algoritmo CT (El Algoritmo CT)

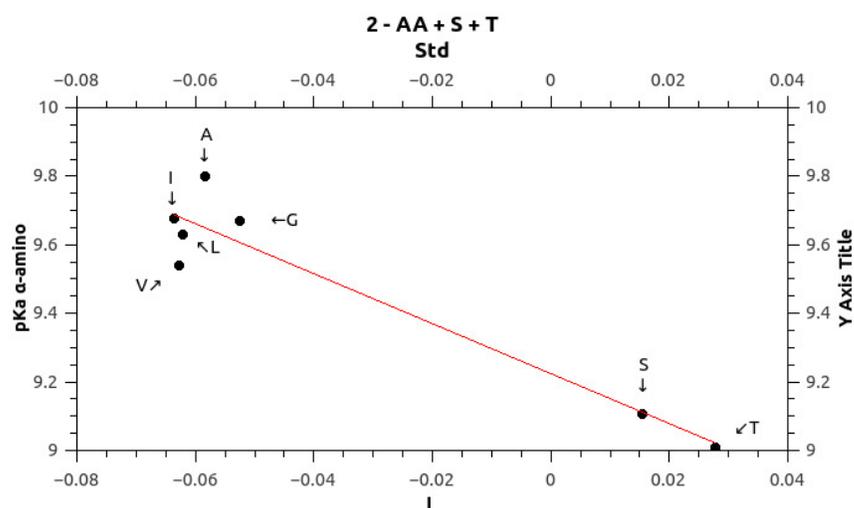


Figura 5.3: $R^2 = 0.91$. Algoritmo CT. Caso 1 más Ser y Thr.

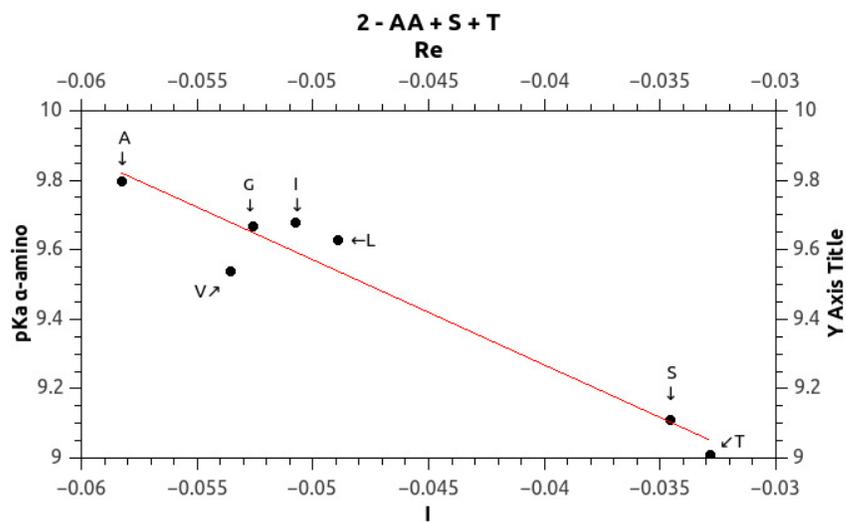


Figura 5.4: $R^2 = 0.93$. Algoritmo CTM. Caso 1 más Ser y Thr.

5.5. Mejoras al Algoritmo CT (El Algoritmo CT)

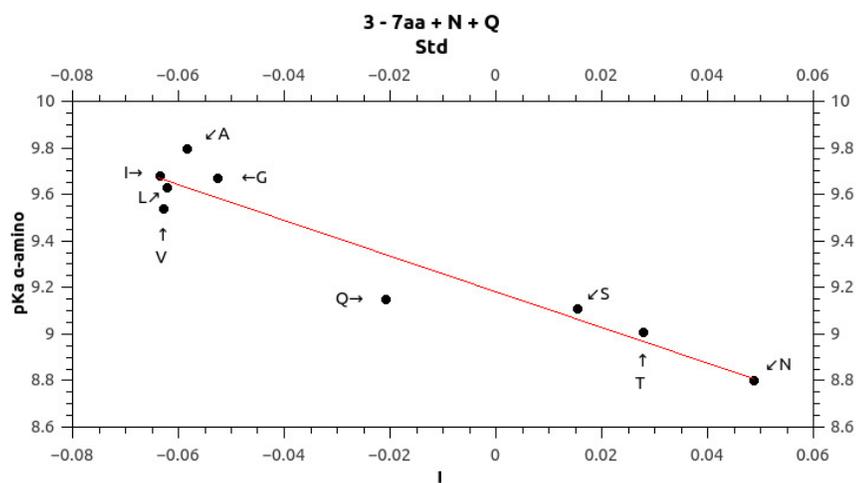


Figura 5.5: $R^2 = 0.91$. Caso 2 más Gln y Asn.

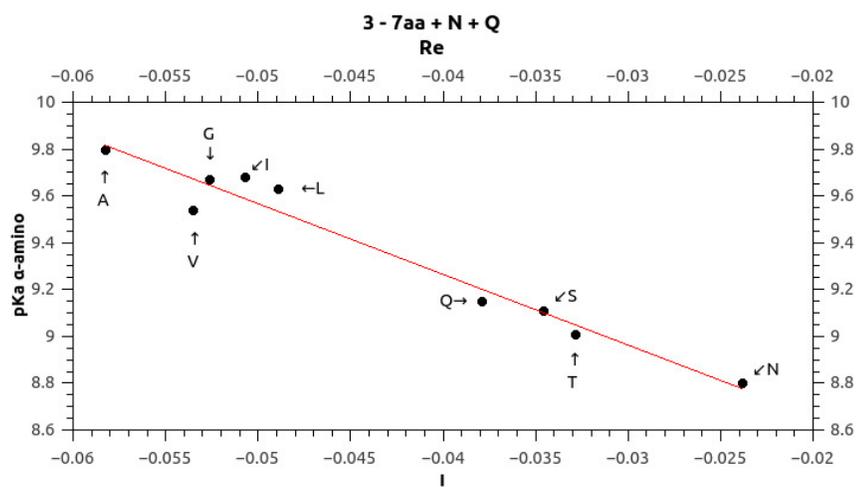


Figura 5.6: $R^2 = 0.96$. Caso 2 más Gln y Asn.

5.5. Mejoras al Algoritmo CT (El Algoritmo CT)

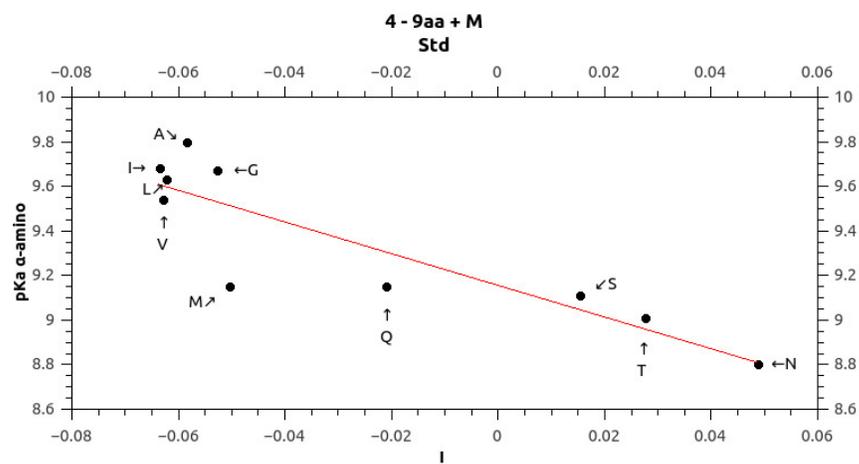


Figura 5.7: $R^2 = 0.71$. Caso 3 más Met.

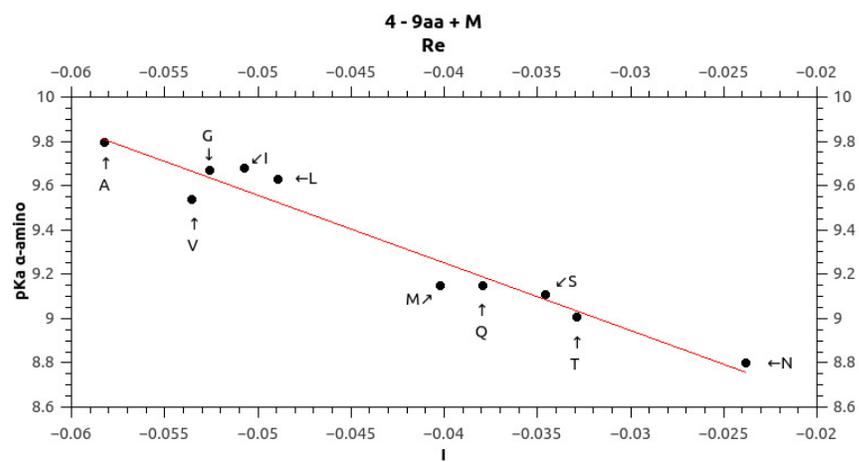


Figura 5.8: $R^2 = 0.95$. Caso 3 más Met.

5.6. Correlaciones de ii con diferentes mediciones (El Algoritmo CT)

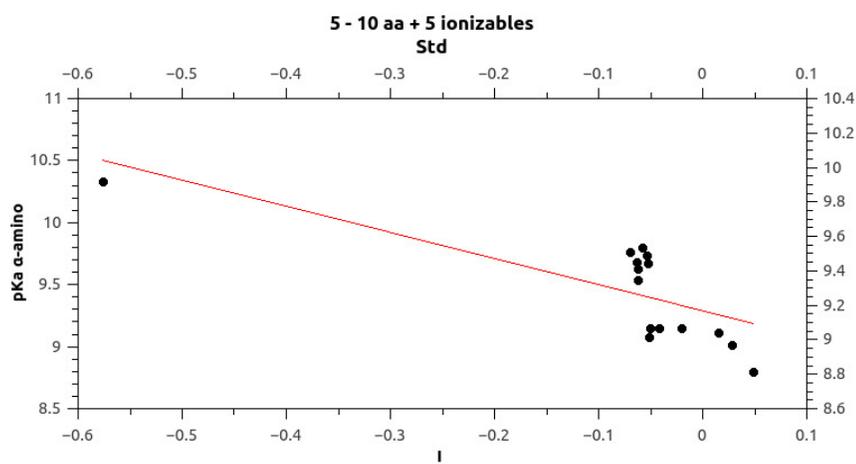


Figura 5.9: Caso 5 con los 15 aminoácidos, incluidos los cinco ionizables y con $R^2 = 0.55$. Esta gráfica hace evidente el problema con el algoritmo CT, y la falta de correlación una vez que se consideran los aminoácidos ionizables. Esto resulta particularmente importante pues las cisteínas son el actor principal en la dinámica del plegamiento oxidativo.

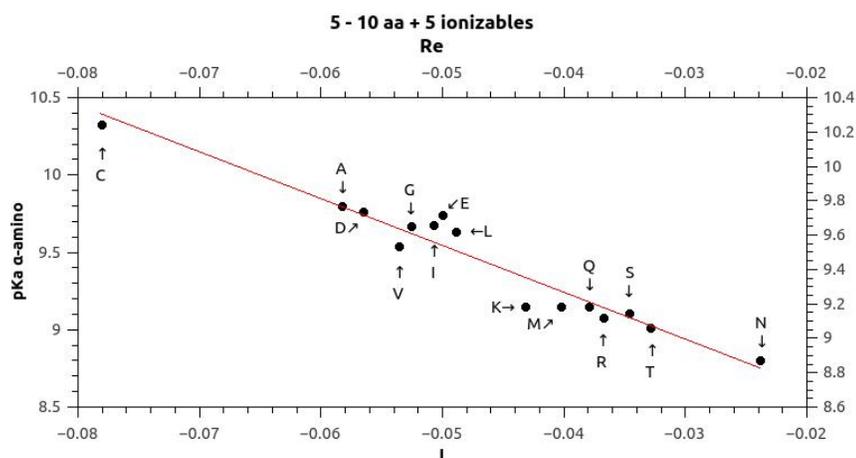


Figura 5.10: Caso 5 usando el algoritmo CTM para los 15 aminoácidos, incluidos los cinco ionizables y con $R^2 = 0.94$.

5.6. Correlaciones de ii con diferentes mediciones

La predicción inicial que motivó el presente trabajo, a saber, que si la influencia ejercida por la cadena lateral de un aminoácido como consecuencia del efecto inductivo sobre los grupos α -amino y α -carboxilo es la misma para aminoácidos con cadena lateral

5.6. Correlaciones de ii con diferentes mediciones (El Algoritmo CT)

Cadena Lateral	ii de cadena lateral sobre C_{α} ($\times 10^3$)	$\alpha - NH_3^+$ pK_a
Cys (tiol ionizado)	-339.6	10.33 ± 0.02
Asp (β -carboxilo ionizado)	-162.5	9.76 ± 0.02
Ala	-134.0	9.80 ± 0.02
Val	-127.1	9.55 ± 0.01
Ile	-96.8	9.69 ± 0.01
Glu (γ -carboxilo ionizado)	-91.3	9.74 ± 0.01
Leu	-72.8	9.64 ± 0.01
Gly	-53.2	9.68 ± 0.01
Met	19.7	9.15 ± 0.02
Arg (guanidino ionizado)	46.5	9.08 ± 0.01
Lys (ϵ -amino ionizado)	62.8	9.16 ± 0.01
Ser	81.5	9.11 ± 0.02
Gln	89.3	9.15 ± 0.01
Thr	92.3	9.01 ± 0.01
Asn	180.8	8.80 ± 0.02

Cuadro 5.3: Valores calculados de Índices Inductivos (ii) para cada cadena lateral sobre el carbono alfa de los 15 α -aminoácidos no aromáticos y valores de pK_a determinados experimentalmente para grupos α -amino. Los valores de pK_a son los mismos que en la tabla 3.1. Los estados de ionización de cadenas laterales se muestran entre paréntesis.

no ionizable, esperaríamos observar una cierta proporcionalidad en los valores de pK_a de los grupos α . Después de comprobar la proporcionalidad mediante mediciones experimentales, buscamos una herramienta que nos permitiera calcular de manera teórica la magnitud de los efectos inductivos. La herramienta que elegimos fue el algoritmo CT, que muestra ajustes lineales en todos los casos que se estudian en el trabajo original [6]. Para poder calcular el efecto inductivo de las cadenas laterales de aminoácidos sobre sus grupos α , para posteriormente extenderlos a péptidos y proteínas, sería necesario que el algoritmo CT permitiera realizar los cálculos para moléculas aromáticas. Como el algoritmo CT no es aplicable para moléculas aromáticas, resultó necesario interpolar los valores de los tres aminoácidos aromáticos. Esto fue posible por la proporcionalidad que muestran los valores de pK_a de los grupos α . Luego, al lograr obtener ajustes lineales entre nuestros cálculos y los datos experimentales, se sigue que es posible hacer la interpolación para los aminoácidos aromáticos con un error aproximado al error experimental que se observa en la figura 3.1. Las modificaciones que se propusieron para el algoritmo CTM son en los valores numéricos y no en la forma del algoritmo de cálculo, por lo que podemos esperar un mejor ajuste, pero no un cambio en la forma del ajuste¹³.

¹³Es decir, que los ajuste no pueden pasar de ser lineales a ser polinomiales o exponenciales.

5.6. Correlaciones de ii con diferentes mediciones (El Algoritmo CT)

Grupos α	pK_a de Cadena Lateral						τ (ns)de
	ii on C_α	Asp	Glu	His	Cys	Trp	
Zwitterion	653.0	3.81 ± 0.02	4.27 ± 0.01	6.02 ± 0.02	8.27 ± 0.01	2.55	
N - Ac ($\alpha - \text{COO}^-$)	-338.2	4.72 ± 0.01	4.83 ± 0.01	6.98 ± 0.01	9.66 ± 0.01	4.24	
C - NH_2 ($\alpha - \text{NH}_3^+$)	1305.2	2.94 ± 0.02	3.88 ± 0.01	5.35 ± 0.01	-	1.44	
C - OMe ($\alpha - \text{NH}_3^+$)	1385.7	-	-	5.26 ± 0.01	6.58 ± 0.01	-	
N - Ac/C - NH_2	314.0	4.01 ± 0.01	4.34 ± 0.02	-	-	2.71	
N - Ac/C - OMe	394.6	-	-	-	8.78 ± 0.01	-	
N - Ac/C - NHMe	304.9	-	-	6.42 ± 0.01	-	-	

Cuadro 5.4: Valores de Índice Inductivo (ii) calculados sobre el carbono α de aminoácidos con cadena lateral ionizable y triptófano, con y sin modificaciones covalentes en los grupos α amino y carboxilo; valores de pK_a medidos para la cadena lateral (Asp, Cys, Glu y His), y Vida Media Fluorescente (estos valores de τ son los usados en la figura 6.2)

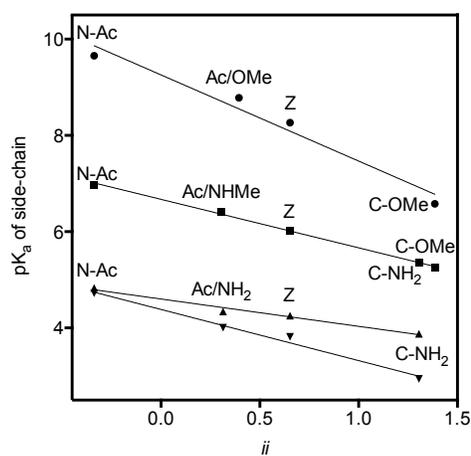


Figura 5.11: Valores medidos de pK_a (\pm desviación estándar) contra valores calculados de Índices Inductivos (ii sobre C_α de cada aminoácido modificado o Zwitterión: Cys (círculos), His (cuadrados), Glu (triángulos), y Asp (triángulos invertidos). Z es el aminoácido en su estado zwitteriónico (sin modificaciones) y N - Ac indica acetilación del grupo α -amino. C - NH_2 y C - OMe indican amidación y esterificación del grupo α -carboxilo, respectivamente. Los aminoácidos con doble modificación se indican con la combinación correspondiente (Ac/NH₂, Ac/OMe). En el caso de la His doblemente modificada, el grupo α -carboxilo fue N-metil amidado (Ac/NHMe).

5.6. Correlaciones de ii con diferentes mediciones (El Algoritmo CT)

Capítulo 6

Mediciones de Fluorescencia

6.1. Introducción

Después de los éxitos obtenidos en la predicción de pK_a al nivel de grupos alfa de aminoácidos y de cadenas laterales ionizables, se decidió explorar la transmisión de los efectos inductivos a través del esqueleto peptídico en péptidos que tuvieran un grupo reportero adecuado, para así poder validar la capacidad predictiva del algoritmo CTM con respecto a los cambios de pK_a como función de las contribuciones inductivas de los aminoácidos vecinales.

Inicialmente, se propuso realizar mediciones de pK_a por métodos potenciométricos sobre el grupo tiol de una cisteína que apareciera en un péptido dado. El objetivo general de estos experimentos era evaluar si el algoritmo CTM lograba hacer buenas predicciones con respecto a la transmisión del efecto inductivo a través del esqueleto peptídico, es decir, ver si existía una buena correlación entre valores de pK_a medidos para cada cisteína e índices inductivos. En particular, los cálculos realizados con el algoritmo de Chiang y Tai predicen una diferencia en los valores de pK_a medidos para cada cisteína como función de los aminoácidos vecinales en la estructura primaria del péptido. Luego, la medición de diferencias en los pK_a que estuvieran de acuerdo con las predicciones teóricas representarían un fuerte argumento en favor de la hipótesis central del presente trabajo. Al final se optó por realizar mediciones de fluorescencia tomando el triptófano como grupo reportero en lugar de la cisteína, para entonces intentar encontrar alguna correlación entre los valores de índice inductivo y algún parámetro de fluorescencia. La razón por la que nos inclinamos por realizar mediciones de fluorescencia son:

1. las concentraciones necesarias para realizar los experimentos son del orden de 10 micromolar, dos o tres órdenes de magnitud por debajo de lo necesario para mediciones por métodos potenciométricos, por lo que se minimiza la probabilidad de interacción intermolecular y de esta forma se reduce la posibilidad de tener artefactos experimentales;
2. la existencia de al menos un trabajo previo en el que se reporta una respuesta acorde con nuestras predicciones teóricas [15];

6.2. Diseño experimental (Mediciones de Fluorescencia)

3. el tener un conjunto de mediciones diferentes al pK_a , pero consistentes con la teoría para de esta manera, tener un segundo grupo de datos que sirvieran como comprobación independiente y que permitiera calibrar los cálculos mejorados de Chiang y Tai.
4. como ya se mencionó, es posible determinar las constantes de ionización a partir de los cambios de fluorescencia, por lo que es razonable esperar una correlación fuerte entre la fluorescencia y el pK_a .

La teoría explica las diferencias de pK_a de grupos ionizables de péptidos o proteínas, como el resultado de la modificación de la nube electrónica del grupo reportero como función o resultado de las diferencias en electronegatividad de los átomos que componen a la molécula (péptido o proteína) en cuestión. Luego, si se observa una modificación en la fluorescencia del triptófano en un péptido suficientemente pequeño tal que podamos garantizar la ausencia de estructura terciaria¹, y bajo condiciones similares (medio, temperatura, etc.) podemos concluir que las diferencias en fluorescencia son consecuencia de la contribución inductiva de los aminoácidos vecinales, y no del microambiente alrededor del grupo reportero como se interpretan comúnmente las diferencias de fluorescencia en proteínas.

6.2. Diseño experimental

Para esto se diseñaron 7 tetrapéptidos y 4 dipéptidos, todos con los grupos amino y carboxilo terminal bloqueados (acetilados y amidados respectivamente).

- los de la “familia D” (GGWD, GWGD y WGGD);
- los de la “familia N” (GGWN, GWGN y WGGN);
- un tetrapéptido que sólo contenía glicinas, además del triptófano (GGWG);
- los 4 dipéptidos (WN, WM, WV y WA).

El diseño de los péptidos responde a las siguientes directrices:

1. los grupos alfa ionizables aparecen bloqueados para eliminar el efecto de las cargas formales sobre el grupo reportero y poder observar sólo el efecto de las cadenas laterales;
2. todos los tetrapéptidos tienen un grupo reportero común (W), y un mismo aminoácido vecinal (G). Esto con la idea de homogeneizar al máximo las condiciones de cada sistema. Además se eligió como aminoácido vecinal a la glicina por tener la cadena lateral más pequeña, y por lo tanto, la que menos efecto estérico contribuye;
3. en general se busca ver el efecto de diferentes cadenas laterales sobre la fluorescencia del grupo reportero variando la distancia (en número de enlaces);

¹Y posiblemente secundaria también.

6.3. Resultados: Vida Media Fluorescente (Mediciones de Fluorescencia)

4. las “familias D y N” aportan datos sobre el efecto de la distancia en enlaces al alejar progresivamente el grupo reportero de la cadena lateral diferente (D y N);
5. para el caso de la “familia D”, la fluorescencia debe modificarse como función del pH como resultado de la ionización de la cadena lateral, y para el caso de la “familia N” la fluorescencia debería de mantenerse constante;
6. los aminoácidos (N, M, V, A) siguen una progresión ascendente de valores de pK_a ², luego, esperamos un comportamiento similar en la emisión de fluorescencia del triptófano enlazado a cada uno de estos aminoácidos.

6.3. Resultados: Vida Media Fluorescente

En la literatura ([36], [7], [8], [9]) se reporta que los cambios en eficiencia cuántica (*quantum efficiency*) correlacionan con la diferencia de electronegatividad de los sustituyentes de la molécula en cuestión, y por lo tanto con el efecto inductivo ejercido por estos. Como ya se mencionó, originalmente pensamos en medir eficiencia cuántica tomando como base estas publicaciones, sin embargo posteriormente nos inclinamos por mediciones de Vida Media Fluorescente (τ) para que nuestros resultados fueran independientes de concentración. A continuación se presentan los resultados de nuestros experimentos de medición de τ para péptidos con grupos alfa modificados de manera covalente contra los valores calculados de ii de los péptidos correspondientes.

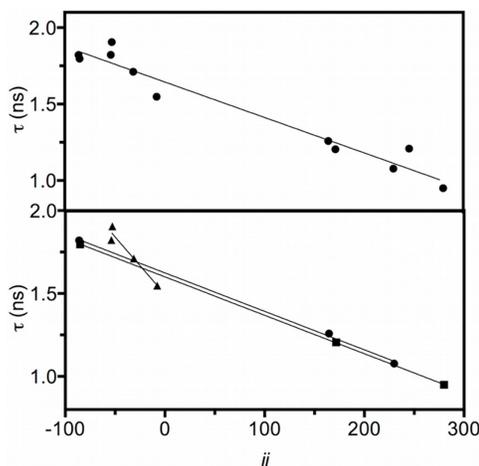


Figura 6.1: Correlación entre vida media fluorescente (τ) de todos los péptidos con ii ($R^2 = 0.9498$). La correlación aumenta cuando se analizan por separado los péptidos correspondientes a las diferentes familias. Para la familia D (círculos), se tiene un $R^2 = 0.9985$; para la familia N (cuadrados), se tiene un $R^2 = 1$; y para los dipéptidos (triángulos), se tiene un $R^2 = 0.9423$. Los valores se presentan en la tabla 6.1.

En este caso, el microambiente es esencialmente el mismo para todos los péptidos,

²Como ya se mostró, los valores de pK_a de grupos alfa amino y carboxilo son proporcionales para cada aminoácido no ionizable, salvo la prolina.

6.4. Dipéptidos y tetrapéptidos (Mediciones de Fluorescencia)

Péptido	\bar{i} en C_{α} de Trp ($\times 10^3$)	τ (ns)
WM	-7.8	1.55
WM	-31.3	1.71
WV	-52.7	1.91
WA	-53.7	1.82
GGWD ⁻	229.9	1.08
GWGD ⁻	164.5	1.26
WGGD ⁻	-86.1	1.80
GGWN	279.9	0.95
GWGN	171.8	1.21
WGGN	-85.1	1.80
GGWG	245.8	1.21

Cuadro 6.1: Valores calculados de Índice Inductivo sobre el carbono alfa del triptófano presente en los péptidos contra valores de Vida Media Fluorescente (τ). Todos los péptidos fueron modificados de manera covalente en los grupos alfa (C -amidación y N -acetilación). La carga negativa del ácido aspártico en los péptidos de la familia D indica que los experimentos fueron realizados a un pH adecuadamente superior al pK_a del grupo β -carboxilo.

por lo que cualquier variación debe ser consecuencia de la diferencia en contribución inductiva de las cadenas laterales presentes en el péptido.

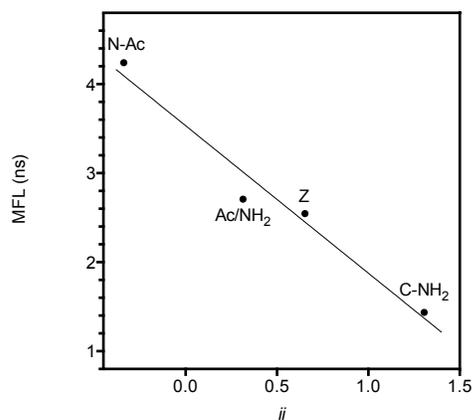


Figura 6.2: Correlación de Vida Media Fluorescente (τ) del grupo indol del triptófano contra valores de Índice Inductivo (\bar{i}) calculado sobre el carbono alfa del triptófano en diferentes estados, i.e., Zwitteriónico y con grupos alfa con diferentes modificaciones covalentes ($R^2 = 0.9678$). Las modificaciones y los valores de \bar{i} son los mismos que en la tabla 5.4. $N - Ac$ indica N -acetil-Trp, $C - NH_2$ indica Trp-amida y Ac/NH_2 indica N -acetil-Trp-amida (NATA).

6.4. Dipéptidos y tetrapéptidos

Para los 4 dipéptidos (WN, WM, WV, WA) se esperaba una importante correlación entre los valores de τ y los valores de pK_a para grupos alfa amino (y por la proporci-

6.4. Dipéptidos y tetrapéptidos (Mediciones de Fluorescencia)

nalidad también para los grupos alfa carboxilo) de los residuos que forman el dipéptido, y que no son el grupo reportero (N, M, V, A). Como en este caso todos los péptidos presentaban las mismas modificaciones, a saber, bloqueos covalentes en sus grupos alfa, la única diferencia presente es la cadena lateral del residuo vecinal al triptófano, por lo que cualquier modificación en la fluorescencia del triptófano se le puede atribuir a la influencia de la cadena lateral vecinal. Aunque la explicación que se favorece es la de la desactivación de fluorescencia (quenching), los resultados presentados no sólo son consistentes con una interpretación basada en los efectos inductivos, sino que el algoritmo CTM nos permite predecir la magnitud de esa influencia.

Para el caso de los tetrapéptidos, se intentó observar la influencia de las cadenas laterales vecinales conforme se modificaba la distancia (en número de enlaces) sobre la fluorescencia del grupo reportero. La asparagina fue seleccionada por ser una cadena lateral con un fuerte poder atractor de electrones, además de tener una cadena lateral que no es ionizable, por lo que su contribución debe de ser estable a diferentes valores de pH . Su parecido estructural con el ácido aspártico permite suponer que la diferencia en los valores de pK_a y/o fluorescencia entre ambos péptidos no depende principalmente de relaciones estéricas, por lo que la explicación de los cambios de fluorescencia se alejan de la interpretación estándar de desactivación de fluorescencia. Las decisiones experimentales respondieron a la necesidad de minimizar la necesidad de usar modelos o explicaciones tradicionales, y mostrar que una interpretación basada en la transferencia de efectos inductivos es plenamente consistente con los datos.

Con la “familia” del ácido aspártico esperábamos poder observar la influencia de la carga formal negativa presente en una cadena lateral sobre aminoácidos vecinales. Evidentemente la contribución inductiva será dependiente del estado de ionización de la cadena lateral. Observemos que en principio se debería poder deducir el pK_a de la cadena lateral del aspartato cuando forma parte de un péptido/proteína³ y la distancia a la que se sigue detectando el efecto de la carga formal negativa presente en la cadena lateral de D.

Los resultados confirmaron que las modificaciones que hicimos al algoritmo CT mejoran considerablemente las correlaciones y la capacidad del algoritmo CTM para hacer predicciones. Observamos una fuerte correlación entre nuestros valores calculados y las mediciones de pK_a y Vida Media Fluorescente.

Sin embargo el algoritmo CTM dista mucho de ser un producto terminado. El algoritmo sigue siendo incapaz de calcular valores de ii para moléculas o grupos aromáticos, además de que el factor de transmisividad es un valor constante que no depende del tipo de enlace o de los átomos o grupos involucrados y no es evidente que deba ser así. La idea es que el valor α sea una función de los átomos que forman el enlace tal que si los átomos en cuestión son A y B , el valor sería un valor variable dependiendo de las propiedades de A y B , i.e., $\alpha(A, B)$. Evidentemente que sea un valor constante simplifica de manera importante la facilidad para realizar los cálculos, por lo que este punto requiere de una revisión con mayor detalle. También es relevante mencionar que los ajustes que realizamos tenían el objetivo de mejorar las correlaciones para aminoácidos y péptidos, y no queda claro que las modificaciones realizadas sean

³En este caso por efecto inductivo esperamos que el pK_a de la cadena lateral sea diferente cuando D forma parte de un péptido/proteína al medido sobre el aminoácido individual.

6.4. Dipéptidos y tetrapéptidos (Mediciones de Fluorescencia)

una mejora universal que permita simplemente reemplazar al algoritmo CT para cualquier caso posible. Remediar esta situación también requiere de una revisión detallada y contrastada con las respectivas mediciones.

Con esto concluimos la parte de predicción y cuantificación de los efectos inductivos para aminoácidos y péptidos del programa de investigación mencionado en la Introducción, cuyo objetivo es explorar la influencia de los efectos inductivos en el plegamiento oxidativo y de manera general, en el plegamiento de proteínas.

Capítulo 7

Perspectivas

El primer objetivo del presente proyecto se logró con éxito, a saber, el mejorar las correlaciones entre los valores experimentales y calculados usando el algoritmo CTM.

A continuación se listan algunos puntos que quedan pendientes y que se planea retomar en el futuro.

1. Plantear la expresión adecuada para calcular la probabilidad de formación de un puente disulfuro a partir de las probabilidades individuales de las cisteínas en cuestión para cuando las cisteínas sean vecinales, o en su defecto, determinar a partir de cuantos aminoácidos vecinales podemos dejar de considerar el efecto ejercido por una cisteína vecinal.
2. Una de las modificaciones realizadas al algoritmo CT fue el usar un valor diferente para α . La modificación ayudó a mejorar las correlaciones pero seguimos usando un único valor de α . En principio no hay razón para esperar que la transmisividad de los efectos inductivos sea constante, por lo que suena razonable explorar los efectos de calcular *ii* tomando α no como una constante sino como una función que dependa de: 1) las electronegatividades de los átomos en cuestión, 2) los estados de hibridación de los orbitales en cuestión, 3) de si los átomos forman o no parte de un grupo funcional particular, 4) buscar algún otro parámetro que pueda influir en la transmisividad de los efectos inductivos¹.
3. Realizar los ajustes necesarios para poder usar con confianza el algoritmo CTM en moléculas cíclicas.
4. Chiang y Tai anunciaron que pronto publicarían una extensión del algoritmo CT que permitiera realizar cálculos para moléculas aromáticas. Ese trabajo nunca se publicó y sería una extensión bienvenida y necesaria para ampliar la aplicabilidad de los resultados presentados tanto en su trabajo original como en el presente trabajo.

¹Históricamente los primeros sistemas en los que se estudiaron los efectos inductivos fueron los sistemas aromáticos [19], [20], [21] pues se consideraba que la resonancia fomentaría la transmisividad. Posteriormente se mostró [41], [42], [43] que no era necesario tener sistemas resonantes para que se manifestaran de manera apreciable los efectos inductivos.

(Perspectivas)

5. Uno de los objetivos del presente proyecto es el poder partir de los valores calculados de ii en cisteínas de proteínas modelo para después buscar alguna diferencia numérica entre la conectividad nativa de puentes disulfuro y el resto de las posibilidades.
6. Explorar la pertinencia de extender los resultados del punto anterior (en caso de ser prometedores) para estudiar la formación de puentes de hidrógeno en la estructura secundaria de proteínas pequeñas (posiblemente buscar un equivalente de proteínas modelo para puentes de hidrógeno).

Apéndice A

Materiales y Métodos Experimentales

A.1. Aminoácidos y reactivos para titulaciones potenciométricas

Todos los L-aminoácidos fueron estándares analíticos de Sigma-Aldrich. Los aminoácidos modificados fueron de la mayor pureza disponible de Sigma-Aldrich y Research Organics. El agua fue doblemente destilada y de grado cristalográfico. Los tituladores (0.1 N NaOH o 0.1 N HCl) fueron soluciones estandarizadas de JT Baker. Los amortiguadores (buffers) de pH estándar para calibrar el electrodo fueron adquiridos de Hannah (HI 6000 series: pH 1.679, pH 4.010, pH 6.862, pH 7.010, pH 10.010).

A.2. Péptidos y reactivos para medidas de fluorescencia

Todos los péptidos se obtuvieron de JPT Peptide Technologies (Berlín, Alemania) y fueron de una pureza $> 95\%$. Todos los péptidos contenían Trp como fluoróforo, y los grupos alfa modificados de manera covalente, con el grupo amino acetilado y el grupo carboxilo amidado. Todos los tetrapéptidos tenían a la Gly como aminoácido estructural.

A.3. Determinación de constantes de ionización

Determinación de constantes de ionización de grupos alfa y cadenas laterales ionizables en aminoácidos modificados y no modificados por medios potenciométricos. Las titulaciones fueron realizadas de manera rigurosa bajo condiciones experimentales idénticas en una unidad automática de titulación Metrohm 853 equipada con un

A.4. Mediciones de fluorescencia (Materiales y Métodos Experimentales)

recipiente de titulación recubierto para control de temperatura ($25\text{ }^{\circ}\text{C} \pm 0.1\text{ }^{\circ}\text{C}$), con agitador magnético y un electrodo de *pH* Metrohm Biotrode. El volumen inicial fue de 10 ml para todos los casos; las titulaciones fueron realizadas bajo un flujo de nitrógeno húmedo y se usó una trampa de CO_2 . La variación del electrodo se estabilizó a < 0.5 mV/min y el electrodo fue calibrado cada día con cinco amortiguadores estándar. Para la titulación de los grupos alfa amino y cadenas laterales ionizables, se disolvió inicialmente los aminoácidos en 0.1 N HCl a 20mM y luego se titularon con 0.1 N NaOH; las constantes de Ionización (expresadas como $\text{p}K_a$) fueron determinadas por análisis de la curva de titulación mediante cálculos estándar de segunda derivada implementados con el software propio del sistema de titulación. Para el caso de las constantes de ionización de los grupos alfa carboxilos, los aminoácidos fueron inicialmente disueltos en agua a 20mM y se titularon agregando de manera secuencial 0.1 equivalentes de titulador, que en este caso fue 0.1 N de HCl. Los métodos experimentales y cálculos fueron realizados tal como se describe en [1]. Todas las titulaciones fueron replicadas al menos tres veces en días diferentes, y con dos lotes independientes para la mayoría de los aminoácidos.

A.4. Mediciones de fluorescencia

Las mediciones de Vida Media Fluorescente fueron realizadas en un fluorímetro Chronos ISS de cambio de fase de multifrecuencia y modulación (Champaign, IL, EUA) usando un LED de 280 nm (ISS modelo 90101) como excitador y *p*-terfenilo en EtOH ($\tau = 1.05$ ns) como referencia. Para eliminar la dispersión causada por la luz de excitación la emisión de fluorescencia fue colectada a través de un filtro de paso largo WG320. Se colectaron datos hasta que la desviación estándar de cada medición de retraso de fase (ϕ) y modulación (M) fueron a lo más 0.2° y 0.004 respectivamente. La frecuencia de modulación del diodo fue barrida de manera logarítmica de abajo hacia arriba del punto de cruce con ϕ y la modulación M en al menos 8 pasos (típicamente entre 20 MHz y 200 MHz), y para cada péptido fue analizado con el programa Vinci (ISS) asumiendo un único tiempo de vida con una distribución Lorentziana (en ambos casos el tiempo de vida y la longitud fueron variables). Los valores de χ^2 estuvieron por debajo de 1.40 y en general por debajo de 1. Un caso especial fue el del triptófano amida que fue el único compuesto que mostró dos tiempos de vida diferentes ($\tau = 1.44$ ns con una fracción de 0.71 y $\tau = 6.94$ ns con una fracción de 0.29). Para análisis subsecuente, se usó el tiempo de vida $\tau = 1.44$ ns, ya que ésta fue la contribución primaria. De manera similar se usó un sólo tiempo de vida para el triptófano pues el análisis usando dos no mejoró los resultados, y las fracciones de otros tiempos de vida resultaron despreciables. Todas las mediciones fueron realizadas usando agua de calidad Milli-Q a una temperatura de $25\text{ }^{\circ}\text{C} \pm 0.1\text{ }^{\circ}\text{C}$, y a un *pH* de 6.81.

Apéndice B

Probabilidades de Formación

B.1. Probabilidad de formación de SS para 3 Cys

Ahora extendemos el cálculo de probabilidades para 3 Cys. Observemos que para este caso el porcentaje de la Cys_1 desprotonado puede interactuar con el porcentaje protonado tanto de la Cys_2 como de la Cys_3 . Luego, para obtener la proporción real de las Cys_2 y Cys_3 que pueden interactuar con cada estado de la Cys_1 es necesario tomar $P_{2H} + P_{3H}$ como el nuevo total contra el que puede interactuar la población representada por P_{1-} . Además las fracciones molares se modifican de la siguiente manera:

$$\begin{aligned}\chi_{1H} + \chi_{1-} &= \frac{1}{3} \\ \chi_{2H} + \chi_{2-} &= \frac{1}{3} \\ \chi_{3H} + \chi_{3-} &= \frac{1}{3}.\end{aligned}\tag{B.1}$$

Luego, despejando χ_{i-} en cada caso y sustituyendo de la misma manera que para una Cys en la ecuación 4.2 se tiene que

$$P_{iH} = \chi_{iH} = \frac{1}{3} \left(\frac{1}{1 + 10^{(pH - pK_a)}} \right),\tag{B.2}$$

y

$$P_{i-} = \chi_{i-} = \frac{1}{3} \left(1 - \frac{1}{1 + 10^{(pH - pK_a)}} \right).\tag{B.3}$$

Bajo esta lógica se tiene que la probabilidad de encontrar a ambas cisteínas en el estado adecuado para formar el puente SS_{12} estará dada por

$$P_{12} = P_{1-} \cdot \left(\frac{P_{2H}}{P_{2H} + P_{3H}} \right) + P_{1H} \cdot \left(\frac{P_{2-}}{P_{2-} + P_{3-}} \right).\tag{B.4}$$

La ecuación B.4 nos da, *Mutatis Mutandis*, el planteamiento “general” para el cálculo de probabilidades de encontrar a dos de tres Cys en el estado adecuado para la forma-

B.1. Probabilidad de formación de SS para 3 Cys (Probabilidades de Formación)

ción de un puente disulfuro. Observemos que las suposiciones de la sección anterior también se aplican aquí.

En general se tiene que

$$P_{ij} = P(SS_{ij}) = P_{i-} \cdot \left(\frac{P_{jH}}{P_{jH} + P_{kH}} \right) + P_{iH} \cdot \left(\frac{P_{j-}}{P_{j-} + P_{k-}} \right), \quad (\text{B.5})$$

para i, j, k distintos entre sí y que cumplen con $i, j, k \in \{1, 2, 3\}$.

Ahora presentamos las gráficas correspondientes para la probabilidad de formación de puentes para 3 Cys.

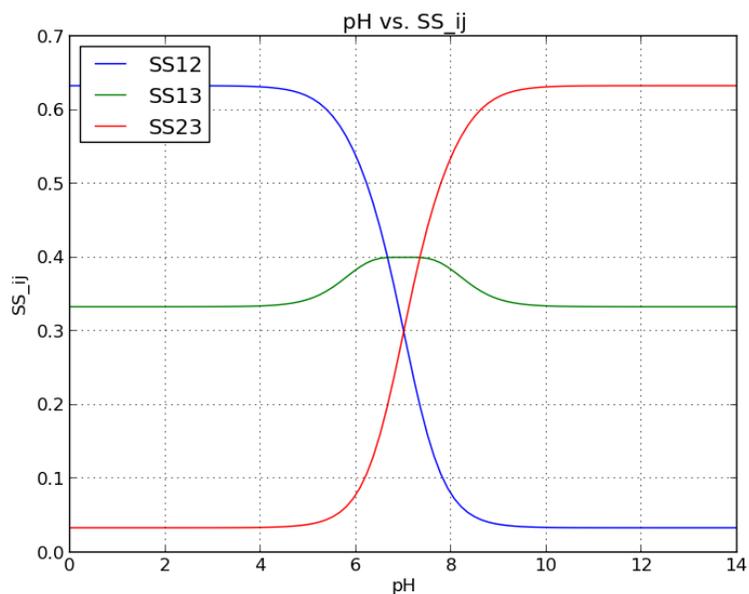


Figura B.1: Gráfica para 3 Cys con pK_a s respectivos de 6, 7 y 8.

B.1. Probabilidad de formación de SS para 3 Cys (Probabilidades de Formación)

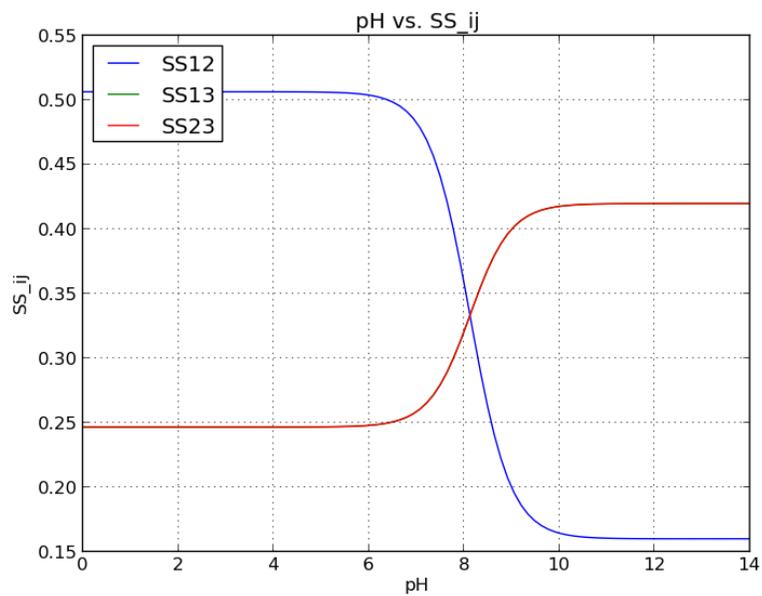


Figura B.2: Gráfica para dos Cys con pK_a igual a 8 y la tercera con pK_a igual a 8.5.

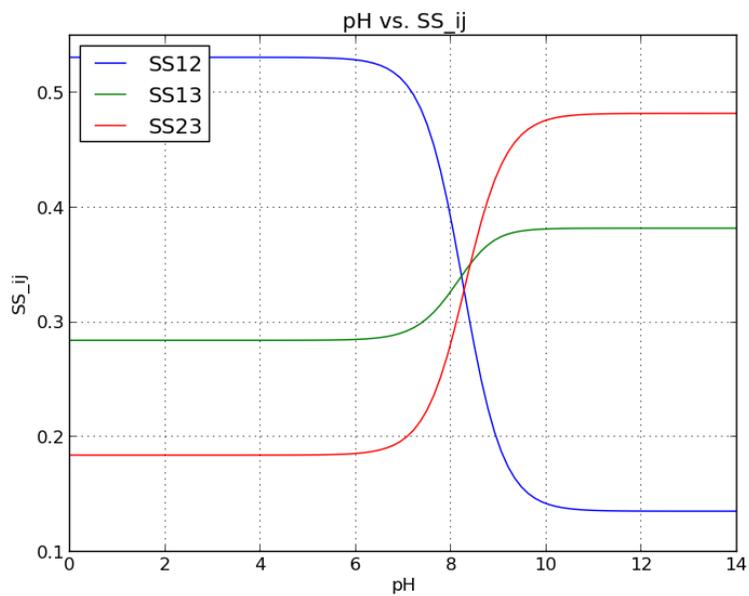


Figura B.3: Gráfica para 3 Cys con pK_a s igual a 8, 8.2 y 8.7.

B.2. Probabilidad de formación de SS para 4 Cys

En este caso se hace el cálculo de la misma manera que para la sección anterior. Como el número de puentes posibles es mayor, las gráficas se vuelven más complejas y un tanto difíciles de interpretar, por lo que se introduce una modificación en los algoritmos para hacer más limpias las gráficas. Esto realmente es una suposición necesaria para que las gráficas representen las hipótesis con las que se ha estado trabajando. La forma general de P_{ij} queda de la siguiente manera:

$$P_{ij} = P_{i-} \cdot \left(\frac{P_{jH}}{P_{jH} + P_{kH} + P_{lH}} \right), \quad (\text{B.6})$$

y para P_{ji} se tiene que

$$P_{ji} = P_{jH} \cdot \left(\frac{P_{i-}}{P_{i-} + P_{k-} + P_{l-}} \right). \quad (\text{B.7})$$

Luego, definimos entonces la probabilidad de interacción entre la Cys_i y la Cys_j como

$$SS_{ij} = P_{ij} + P_{ji}.$$

Los cálculos se realizan de la siguiente manera:

1. calculamos la probabilidad de encontrar a la Cys_i en sus estados protonada/desprotonada (ecuaciones 4.7 y 4.8)
2. calcular las probabilidades de interacción entre 2 de 4 Cys (sumando las ecuaciones B.6 y B.7)
3. **suposición adicional:** si se formó el puente SS_{12} , entonces el puente SS_{34} también lo hace (en las proporciones adecuadas). Por otro lado, si el puente que se formó fue el SS_{13} , entonces no queda más que la posibilidad de que se forme el puente SS_{24} y de hecho lo hace.

A continuación se presentan las gráficas del cálculo de probabilidades para la formación de puentes SS para 4 Cys.

B.2. Probabilidad de formación de SS para 4 Cys (Probabilidades de Formación)

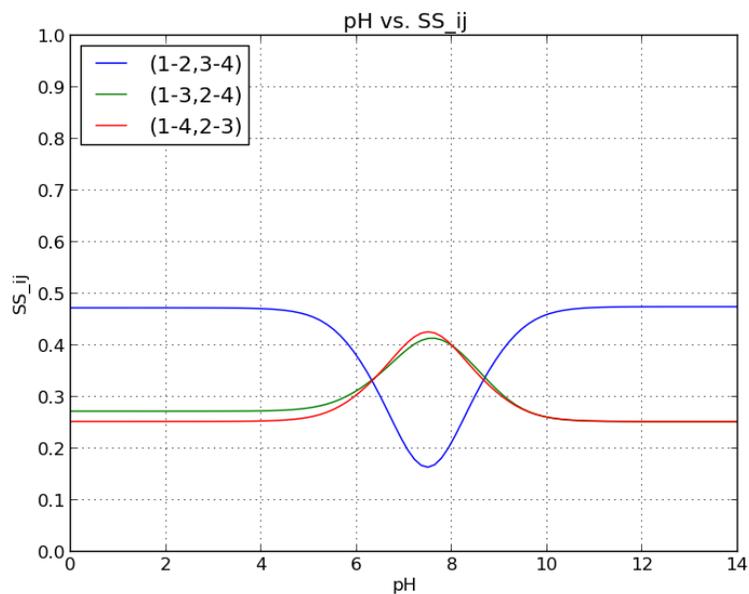


Figura B.4: Gráfica para 4 Cys con pK_a s de 6,7,8,9 respectivamente y donde se supone que si se forma el puente SS_{12} entonces se forma también el SS_{34} y de igual manera para las demás combinaciones.

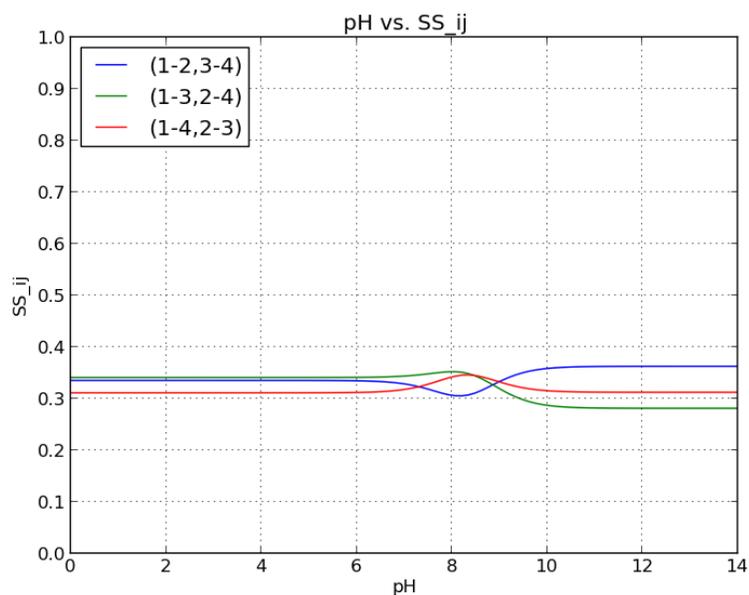


Figura B.5: Gráfica de probabilidad para 4 Cys con valores de pK_a de 8, 8.1, 8.3, 8.9 y bajo la suposición de que si se forma el puente SS_{12} entonces se forma también el SS_{34} y de igual manera para las demás combinaciones.

B.2. Probabilidad de formación de SS para 4 Cys (Probabilidades de Formación)

Bibliografía

- [1] Albert A, Serjeant EP. The Determination of Ionization Constants: A Laboratory Manual. Springer, Netherlands (1984).
- [2] Allred AL, Rochow EG. A Scale of Electronegativity Based On Electrostatic Force. *J. Inorg. Nucl. Chem.*, 5, 264-268.
- [3] Arolas JL, Aviles FX, Chang JY, Ventura S. 2006. Folding of small disulfide-rich proteins: clarifying the puzzle. *Trends Biochem Sci* 31: 292-301.
- [4] Batsanov SS. The Concept of Electronegativity. Conclusion and Prospects. *Russian Chemical Reviews*, 37 (5), 332-351.
- [5] Cantor CR, Schimmel PR. 1980. *Biophysical Chemistry, Part I: The conformation of biological macromolecules*. WH Free
- [6] Chiang MC, Tai TC. 1963. A quantitative relationship between molecular structure and chemical reactivity. *Scientia Sinica XII*: 785-867.
- [7] Cowgill RW. 1962. Fluorescence and the Structure of Proteins. I. Effects of Substituents on the Fluorescence of Indole and Phenol Compounds. *Archives of Biochemistry and Biophysics*, 100, 36-44.
- [8] Cowgill RW. 1963. Fluorescence and the Structure of Protein. II. Fluorescence of Peptides Containing Tryptophan and Tyrosine. *Biochim. Biophys. Acta*, 75, 272-273.
- [9] Cowgill RW. 1967. Fluorescence and Protein Structure. X. Reappraisal of Solvent and Structural Effects. *Biochim. Biophys. Acta*, 133, 6-18.
- [10] Creighton TE (Ed.) 1999. *Protein Folding*. WH Freeman and Company, New York.
- [11] Curtis C.-J. Lin, Jui-Yoa Chang. 2006. Pathway of Oxidative Folding of Secretory Leucocyte Protease Inhibitor: An 8-Disulfide Protein Exhibits a Unique Mechanism of Folding. *Biochemistry*, 45, 6231-6240.
- [12] Curtis C.-J. Lin, Bao-Yuan Lu, Jui-Yoa Chang. 2006. Conformational Stability of Secretory Leucocyte Protease Inhibitor: A Protein with no Hydrophobic Core and very Little Secondary Structure. *Biochimica et Biophysica Acta*. 1764, 1286-1291.

Bibliografía (Bibliografía)

- [13] Danehy JP, Noel CJ. 1960. The Relative Nucleophilic Character of Several Mercaptans toward Ethylene Oxide. *J. Am. Chem. Soc.*, Vol. 82, 2511-2515.
- [14] Dawson RMC. *Data for Biochemical Research*. Clarendon Press, Oxford (1959).
- [15] Edelhoch H, Brand L, Wilchek M. 1967. Fluorescence Studies with Tryptophyl Peptides. *Biochemistry*, 6 (2), 547-559.
- [16] Flory PJ. 1988. *Statistical Mechanics of Chain Molecules*. Hanser Publisher, Munich.
- [17] Gilbert, HF. 1990. Molecular and Cellular Aspects of Thiol-Disulfide Exchange. *Adv. Enzymol.*, 63, 69.
- [18] Gordy W. A New Method of Determining Electronegativity from Other Atomic Properties. *Phys. Rev.*, 69, 604-607.
- [19] Hammett LP. 1937. The effect of structure upon the reactions of organic compounds. Benzene derivatives. *J. Am. Chem. Soc.*, 59, p. 96.
- [20] Hammett LP. 1938. Linear free energy relationships in rate and equilibrium phenomena. *Trans. Faraday Soc.*, 34, p. 156.
- [21] Hammett LP. *Physical Organic Chemistry*. Mcgraw-Hill, New York (1940).
- [22] Hinze J, Jaffé HH. Electronegativity. I. Orbital Electronegativity of Neutral Atoms. *J. Am. Chem. Soc.*, 84, 540-546.
- [23] Hinze J, Whitehead MA, Jaffé HH. Electronegativity. I. Bond and Orbital Electronegativities. *J. Am. Chem. Soc.*, 85, 148-154.
- [24] Lara J, Et Al. 2020. Inductive effects in amino acids and peptides: Ionization constants and tryptophan fluorescence. *Biochemistry and Biophysics Reports* 24, 100802.
- [25] Lee AC, Crippen GM. 2009. Predicting pK_a . *J. Chem. Inf. Model.*, 49, 2013-2033.
- [26] Ling GN. 1962. *A Physical Theory of the Living State: the Association-Induction Hypothesis*. Blaisdell Publishing Company.
- [27] Ling GN. 1984. *In Search of the Physical Basis of Life*. Plenum Press, New York and London.
- [28] Ling GN. 1992. *A Revolution in the Physiology of the Living Cell*. Krieger Publishing Company, Malabar, Florida.
- [29] Lundblad RL, Macdonald FM. *Handbook of Biochemistry and Molecular Biology*. CRC Press Florida (2010), p. 3.
- [30] Martell AE, Smith RM. *Critical Stability Constants*. Plenum Press, New York (1974).

- [31] Mulliken RS. A New Electronaffinity Scale; Together with Data on Valence States and on Valence Ionization Potentials and Electron Affinities. *J. of Chem. Phys.*, 2, 782-793.
- [32] Nelson DL, Cox MM. *Lehninger Principles of Biochemistry*. W. H. Freeman, New York (2004), p. 78.
- [33] Pain RH (Ed). 1994. *Mechanisms of Protein Folding*. Oxford University Press. Oxford.
- [34] Pauling L. 1932. *The Nature of the Chemical Bond*. Cornell University Press. New York.
- [35] Pucci R, Angilella GGN. Density functional theory, chemical reactivity, and the Fukui functions. *Foundations of Chemistry* 24 (2022), 59-71.
- [36] Russell EC, Cowgill RW. 1968. Fluorescence and Protein Structure. XIII. Further Effects of Side-Chain Groups. *Biochim. Biophys. Acta*, 154, 231-233.
- [37] Sanderson RT. 1976. *Chemical Bonds and Bond Energy*. Academic Press. New York.
- [38] Sanderson RT. 1983. *Polar Covalence*. Academic Press. New York.
- [39] Saxena VP, Wetlaufer DB. 1970. Formation of three-dimensional structure in proteins. I. Rapid non-enzymic reactivation of reduced lysozyme. *Biochemistry*, 9, 5015-5023.
- [40] Stachurski J. 2009. *Economic Dynamics: Theory and Computation*. The MIT Press, Cambridge Massachusetts.
- [41] Taft RW. 1952. Linear free energy relationships from rates of esterification and hydrolysis of aliphatic and ortho-substituted benzoate esters. *J. Am. Chem. Soc.*, 74 (11), p. 2729.
- [42] Taft RW. 1952. Polar and steric substituent constants for aliphatic and o-benzoate groups from rates of esterification and hydrolysis of esters. *J. Am. Chem. Soc.*, 74, p. 3120.
- [43] Taft RW. 1952. The separation of relative free energies of activation to three basic contributing factors and the relationship of these to structure. *J. Am. Chem. Soc.*, 75 (18), p. 4538.
- [44] Teale FW, Weber G. Ultraviolet fluorescence of the aromatic amino acids. *Biochem. J.*, 65 (3) (1957), pp. 476-482.
- [45] White A. Effect of pH on fluorescence of tyrosine, tryptophan and related compounds. *Biochem. J.*, 71 (2) (1959), pp. 217-220.
- [46] Zhang R, Snyder GH. 1988. Kinetics of disulfide exchange reactions of monomer and dimer loops of cysteine-valine-cysteine peptides. *Biochemistry*, 1988, 27, 3785-3794.

Bibliografia (Bibliografia)

- [47] Zhang R, Snyder GH. 1989. Dependence of Formation of Small Disulfide Loops in Two-Cysteine Peptides on the Number and Types of Intervening Amino Acids. *J Biol Chem* 264: 18472-18479.