

**UN MODELO EVOLUTIVO BIOFÍSICO DE LA ATP SINTETASA**

**FERNANDO CAYETANO OROZCO CALAMBAS**

**UNIVERSIDAD DEL CAUCA  
FACULTAD DE CIENCIAS NATURALES, EXACTAS Y DE LA EDUCACIÓN  
DEPARTAMENTO DE BIOLOGIA  
POPAYÁN  
2012**

**UN MODELO EVOLUTIVO BIOFÍSICO DE LA ATP SINTETASA**

**FERNANDO CAYETANO OROZCO CALAMBAS**

**Director**

**MSc. ERNESTO COMBARIZA**

**Profesor Titular Universidad del Valle**

**UNIVERSIDAD DEL CAUCA**

**FACULTAD DE CIENCIAS NATURALES, EXACTAS Y DE LA EDUCACIÓN**

**DEPARTAMENTO DE BIOLOGÍA**

**POPAYÁN**

**2012**

Nota de aceptación:

---

---

---

---

---

**MSc. Ernesto Combariza Cruz**  
Director

---

**MSc. Wayner Rivera**  
Jurado

---

**MSc. Nilza Velasco Palomino**  
Jurado

Fecha de sustentación: Popayán, diciembre 5 de 2012

## CONTENIDO

### RESUMEN

### INTRODUCCIÓN

<b>1. JUSTIFICACIÓN</b> .....	<b>5</b>
1.1. HIPÓTESIS.....	6
<b>2. OBJETIVOS</b> .....	<b>7</b>
2.1. OBJETIVO GENERAL.....	7
2.2. OBJETIVOS ESPECÍFICOS .....	7
<b>3. MARCO TEÓRICO</b> .....	<b>8</b>
3.1. FUNDAMENTOS BIOLÓGICOS.....	8
3.1.1. Teoría de la evolución Vs. Teoría del diseño .....	8
3.1.2. Las proteínas como estructuras constituidas por unidades jerárquicamente dispuestas .....	10
3.1.3. ATP-Sintetasa como unidad elemental del carácter orgánico vital .....	20
3.2. SOPORTE FÍSICO-MATEMÁTICO .....	25
3.2.1. Fundamentos del análisis wavelet.....	25
3.2.2. Análisis multiresolución .....	39
3.2.3. Conceptos Básicos de multifractalidad.....	41
<b>4. PROCEDIMIENTO METODOLÓGICO</b> .....	<b>47</b>
4.1. OBTENCIÓN DE SECUENCIAS.....	47
4.2. TRATAMIENTO DE DATOS.....	49

<b>5. RESULTADOS Y DISCUSIÓN .....</b>	<b>54</b>
<b>6. CONCLUSIONES .....</b>	<b>65</b>
<b>RECOMENDACIONES.....</b>	<b>68</b>
<b>REFERENCIAS .....</b>	<b>69</b>
<b>GLOSARIO.....</b>	<b>74</b>
<b>ANEXOS.....</b>	<b>75</b>

## ÍNDICE DE FIGURAS

Figura 3.1. Representación gráfica de los demonios de Maxwell.....	12
Figura 3.2. Cambios de energía en las proteínas.....	12
Figura 3.3. Los 20 aminoácidos proteicos codificados genéticamente.....	13
Figura 3.4. Estructura jerárquica de las proteínas.....	14
Figura 3.5. Estructura del enlace peptídico: estabilización por resonancia.....	14
Figura 3.6. Características estructurales del enlace peptídico.....	15
Figura 3.7. Conformaciones Trans y Cis del enlace peptídico.....	15
Figura 3.8. Estructura tridimensional de las proteínas.....	16
Figura 3.9. Superficie de energía de una proteína 1.....	17
Figura 3.10. Superficie de energía de una proteína 2.....	18
Figura 3.11. Superficie de energía de una proteína 3.....	19
Figura 3.12. Superficie de energía de una proteína 4.....	19
Figura 3.13. Superficie de energía de una proteína 5.....	20
Figura 3.14. Modelo de la estructura de la ATPasa.....	21
Figura 3.15. Unidad Fo.....	21
Figura 3.16. Unidad catalítica F1.....	22
Figura 3.17. Conexión de la unidades Fo y F1.....	23
Figura 3.18. Onda tipo Wavelet.....	28
Figura 3.19. Daubechies 4 db4: funciones wavelet y de escalamiento (detalle).....	33
Figura 3.20. Proceso de descomposición de una señal con análisis multiresolución....	41
Figura 3.21. Forma genérica de las funciones $\tau(q)$ y $\alpha(q)$ .....	46
Figura 4.1. Algoritmo piramidal .....	52
Figura 5.1. Espectro multifractal de análisis de la secuencia humano.....	57
Figura 5.2. Espectro multifractal de análisis de la secuencia rata.....	58
Figura 5.3. Espectro multifractal de análisis de la secuencia bovino.....	59
Figura 5.4. Espectro multifractal de detalle de la secuencia humano.....	60

Figura 5.5. Espectro multifractal de detalle de la secuencia rata..... 61

Figura 5.6. Espectro multifractal de detalle de la secuencia bovino..... 62

## ÍNDICE DE TABLAS

Tabla 4.1. Representación alfabética de los aminoácidos.....	48
Tabla 4.2. Valores de potencial de interacción ión-electrón (EIIP) para aminoácidos.....	50

## RESUMEN

Este trabajo propone un modelo desde el punto de vista biofísico de la enzima ATP-sintetasa (ATPasa) una de las más importantes en la actividad energética de los organismos vivos. La propuesta se enmarca dentro de desarrollos recientes de la biofísica, la bioquímica y la biomatemática; es decir, eminentemente transdisciplinaria y su lugar específico es la biología. Un supuesto fundamental es la capacidad de las ciencias básicas de dar cuenta de los fenómenos complejos a partir de las leyes de la naturaleza físico-matemática. Se explica esto diciendo que las leyes de la termodinámica y los fenómenos no lineales, han de dar cuenta del proceso biológico cuya característica es la finalidad sometida a leyes naturales.

El análisis evolutivo-comparativo de la secuencia de aminoácidos de la subunidad  $\beta$  del modulo  $F_1$  de la enzima ATPasa de humano, rata y bovino, es llevado a cabo inicialmente mediante la asignación del pseudopotencial (Potencial de Interacción Ion Electrón, EIIP por su siglas en ingles) de cada aminoácido, el cual arroja una serie numérica que se analiza por medio de la Transformada Wavelet Discreta Daubechies 12 (db12); con ello, definiendo la llamada función de partición que permite el análisis multifractal del que se puede obtener los factores bioinformáticos de estructura y función que poseen las tres secuencias de los modulos de ATPasa.

## INTRODUCCIÓN

Los conocimientos actuales de la síntesis de ATP son resultado de muchas propuestas que se desarrollaron en un período de casi 50 años. De gran importancia fue el descubrimiento del complejo de la enzima  $F_0-F_1$  ATP sintetasa, que es ampliamente distribuida en animales, plantas, hongos y bacterias. La combinación del trabajo investigativo dio como resultado la estructuración de la  $F_1$ -ATPasa, vía análisis de difracción de rayos X, de la mitocondria del corazón de vacuno, mitocondria de hígado de rata y de bacteria. Otros importantes temas concernientes a la fuerza conductora y el mecanismo de síntesis de ATP fueron aclarados por las definiciones teóricas de Peter Mitchell y por los estudios cinéticos de Paul Boyer [1].

A partir de estos trabajos se estructura una propuesta distinta de los modelos evolutivos Neodarwinianos, en el sentido de que parte y va hasta las ciencias básicas; de tal manera que no requiere un supuesto en el cual, azar y necesidad sean opuestos fundacionales, sino, dos funciones básicas de la termodinámica vistas a nivel de las leyes de la mecánica del micro mundo. Una ampliación de esto, consiste en el reto fundamental para la ciencia de cómo explicar lo irreversible junto con la temporalidad que es intrínseca a la biología y que debe ser explicada en un nivel más básico, sin recurrir a una propuesta del tipo reduccionista.

La teoría de complejidad, es una rama de gran actividad en la que han incursionado F. Crick, P. Anderson y M. Gelman, en la década de los noventa y su vigencia se prolongó con los desarrollos de nuevos elementos teóricos en el siglo XXI, de tal suerte que el aporte del trabajo consiste en acercar la formación curricular básica a los desarrollos más recientes, con miras a dar un marco explicativo más que descriptivo. Lo anterior para especificar que el propósito del

presente trabajo debe ser el de contribuir a una construcción teórica, que además de interpretar datos estadísticos, proponga unos principios investigativos en el campo de la bioenergética, que permitan el análisis correcto de mecanismos y técnicas tales como la Resonancia Magnética Nuclear, Los Rayos X, entre otros.

Para el trabajo se propone como método la Teoría Mejorada del Análisis de Fourier conocida como análisis Ondicular o *Wavelet*. La razón de esta ampliación es que éste tipo de análisis constituye el mecanismo más compacto de guardar información adecuada a una construcción jerárquica invariante de escala, que es la forma evolutiva de un arreglo molecular compacto. Las ondículas surgen de la transformada rápida de Fourier y en la necesidad de estudiar estructuras desordenadas, las cuales son portadoras de información.

Así como también se hace uso de la geometría fractal, puesto que esta geometría se puede considerar como la más cercana a la naturaleza, y además las formas fractales no sólo se presentan en las formas espaciales de los objetos sino que se observan en la propia dinámica evolutiva de los sistemas complejos. Dinámica que consta de ciclos (en los que partiendo de una realidad establecida simple acaban en la creación de una nueva realidad más compleja) que a su vez forman parte de ciclos más complejos los cuales forman parte del desarrollo de la dinámica de otro gran ciclo. Las evoluciones dinámicas de todos estos ciclos presentan las similitudes propias de los sistemas caóticos [31].

Por lo tanto con el objetivo de profundizar en el comportamiento y evolución de la ATPasa, se aplicó el análisis multifractal, a través del cual se pueden constatar patrones reproducibles a diferentes escalas.

En el presente trabajo una molécula compleja como la ATPasa que fue abordada en otros tiempos mediante el método de Vidrios de Spin y las propiedades del Grupo de Renormalización o metodología de sistemas complejos, hoy se aborda

desde el ferromagnetismo el modelo de espines desordenados portadores de información como única herramienta que explica la transición de fase. En nuestro caso mediante las wavelet se hace el modelo con la base adecuada, la cual supe las exigencias de no linealidad, que en el caso Fourier son limitadas debido a su linealidad.

El modelo de la ATPasa se representa por medio de una ondícula tridimensional cuyas propiedades de ortogonalidad se ajustan a ángulos diedros, recogiendo la disposición geométrica espacial de cada aminoácido. La ondícula es construida de forma que el algoritmo debe recoger la autosimilaridad (escalamiento) y la capacidad de portar información genética [5].

El modelo evolutivo bioenergético de la molécula ATPasa, consiste en una explicación de funcionamiento a partir de las ciencias naturales físico-química y matemática (de la unidad vital fundamental de los organismos vivos). Se parte de las leyes físicas representadas en un lenguaje matemático conocido como análisis ondicular o wavelet. Se aplica a los datos biológicos moleculares de la secuenciación de proteínas tomados de las bases de datos existentes y se analiza bioinformáticamente.

El modelo muestra que el funcionamiento de la ATPasa es intrínseco inmanente es decir descrito por una teoría físico-matemática en la que energía e información son una unidad fundamental de la descripción del universo.

## 1. JUSTIFICACIÓN

Desde la biofísica de Guohui Li inspirado en el trabajo de Florence Tama en el año 2002, se desarrolla para la ATP Sintetasa una discriminación dinámica entre modos lentos y rápidos que busca explicar la dinámica molecular de esta proteína gigante. Los desarrollos logrados por ellos sirven de base para el presente trabajo dado que otros modelos como Broglia y otros especializados en plegamiento molecular requieren de métodos muy complejos desde el punto de vista computacional. Este enfoque permite tratar la molécula como un sistema dinámico no lineal en el que los modos lentos determinan la actividad quinasa y fosfatasa de la región  $F_1$ , lo que en principio justifica la introducción de una base completa de soporte compacto como son las ondículas que se desarrollan en el presente trabajo.

Todos los sistemas vivos requieren mínimos de energía necesarios para desarrollar sus actividades. Esta energía se obtiene de los alimentos a través de procesos metabólicos celulares, de los cuales se pueden obtener ecuaciones energéticas referentes a la mecánica cuántica, las que incorporadas a los avances tecnológicos computacionales a nivel de software permitirán análisis de mayor precisión y rapidez.

La necesidad de sustentar teóricamente la reacción catalítica llevada a cabo por intermediación de la ATPasa, más allá de los aportes conocidos en la literatura, como por ejemplo la dinámica molecular, los desarrollos mecánico- cuánticos. La propuesta aquí apunta a fundamentar los cálculos de RMN, Rayos X y otros cuyo núcleo descriptivo es Fourier el cual se tomaba sólo como medio. De esta manera se lleva la información valiosa anterior hasta la categoría de explicación, basándose en los desarrollos recientes de los sistemas complejos.

La ATP Sintetasa funciona como un motor molecular en el que es preciso detallar cada uno de sus componentes, tomando el texto de los aminoácidos y llevándolos hasta la lectura funcional (temporal). Es decir lo que se ha detallado en los estudios previos lo ensamblamos como un colectivo oscilante (*wavelet*) que permite la expresión funcional de la proteína, buscando explicar así, la cinética molecular de las partículas estructurales que la componen; (la cual despliega su acción catalizadora entre la membrana interna mitocondrial).

Desde la biología tenemos una evolución de millones de años conteniendo una información proteómica sin procesar, lo que hace necesario una teoría biológica que interprete estos datos. Por ello es importante caracterizar las proteínas desde el punto de vista de la medida de la complejidad, puesto que no existe una medida que sea informáticamente clara.

### **1.1. HIPÓTESIS**

Un sistema portador de información y de actividad funcional como una proteína, ¿podrá considerarse un sistema que molecularmente se comporta ajustado a reglas de superposición como si fuera un sistema ondicular o altamente correlacionado, es decir, determinístico?.

Supuesto:

- **HIPÓTESIS ALTERNA:** La evolución es un sistema no lineal que vive en desorden macroscópico y en orden microscópico, dejando al azar subsumido en la determinación
- **HIPÓTESIS NULA:** La evolución replica o se repite con invariancia de escala.

## **2. OBJETIVOS**

### **2.1. OBJETIVO GENERAL**

Caracterizar bioinformáticamente la ATPasa a través del análisis wavelet y el cálculo de la medida de complejidad.

### **2.2. OBJETIVOS ESPECÍFICOS**

- Utilizar la transformada wavelet discreta como método eficiente alternativo a la continua utilizada tradicionalmente, en especial la wavelet Daubechies 12 como una estructura arborescente cuya invariancia de escala y compactación de información reflejen la evolución biológica desde el punto de vista del código más eficiente posible.
- Aplicar el análisis wavelet al cálculo de la multifractalidad.
- Calcular el índice de complejidad de la ATPasa por medio de un índice numérico.
- Relacionar el ancho de multifractalidad con el índice de complejidad.

### **3. MARCO TEÓRICO**

#### **3.1. FUNDAMENTOS BIOLÓGICOS EVOLUTIVOS DE LAS PROTEINAS**

##### **3.1.1. Teoría de la evolución Vs. Teoría del diseño**

La teoría de Darwin propone un modelo simple que explica el origen de las especies, consistente en la combinación de un mecanismo externo que regula un mecanismo interno. La expresión de este modelo la vemos en el caso de los pinzones en las Islas Galápagos.

Estos pinzones son de diversas especies en razón a las condiciones particulares de cada isla donde viven; sin embargo tienen un antecedente común que es la procedencia de una clase migratoria de pinzón que pobló las islas. Lo que corresponde a las condiciones externas donde viven. En tanto que la reproducción consiste en el mecanismo implícito interno.

Esta teoría tal como está planteada explica la diversidad de las especies pero no asume cual es el mecanismo asociado a la reproducción de cada individuo, pues su interés es fundamentalmente macroscópico.

Mendel por su parte estudiando el comportamiento de los guisantes (arvejas) descubre un mecanismo combinatorio que supone tácitamente la existencia de una unidad conservativa que en el siglo XX pasará a ser el gen. Este mecanismo implica un razonamiento microscópico, donde los factores de herencia generados por el descubrimiento del cromosoma nuclear en la célula materializan el mecanismo microscópico subyacente.

La teoría microscópica presenta por oposición a la teoría macroscópica un carácter discontinuo dada la naturaleza combinatoria de la teoría del gen, que se amplió con el concepto de mutación, marcando el contraste con la continuidad y parsimonia de la teoría darwiniana. Conciliar estos dos aspectos del saltacionismo con el continuismo fue el logro de la teoría sintética de Fisher, donde se reúne la genética con la mecánica estadística maxwelliana.

De los pinzones, pasando por la mosca del vinagre y los mohos llegamos hasta los virus y las bacterias E. coli donde gradualmente se alcanzan las dimensiones atómicas a la hora de tratar la teoría de la evolución.

La química orgánica de Liebig basaba el estudio de los organismos vivos en el concepto de fuerza vital. Con el surgimiento de la físico-química y el descubrimiento de la acción enzimática, el vitalismo recibió un duro golpe que se tradujo en el descubrimiento del ATP como proceso bioquímico asociado a las funciones de respiración y metabolismo en general. En el momento en que se descubrían las propiedades de los ciclos de las sustancias orgánicas, Max Delbruck heredero de Liebig intentaba utilizar el concepto de mecánica cuántica explicar la relación entre vida y mecanismo, intentando salvar la irreductibilidad de la fuerza vital por medio de la complementariedad bohriana.

Los aportes de la mecánica cuántica para la concepción del virus y la vida tienen un sabor ambiguo, pues en cierta forma presentan una alternativa para tratar los organismos pero introducen una dualidad e irreductibilidad que da cabida a la noción de intervención trascendental a los fenómenos vitales, que será de buen recibo para la teoría del diseño.

El descubrimiento del ADN en la mitad de siglo supuso llevarla hasta el nivel molecular trayendo como consecuencia el desarrollo de la noción de información y

la teoría de sistemas que son resumidas en la teoría del Operon de Monod y Jacob.

La noción de teleonomía pretendía suavizar el concepto direccional de teleología utilizando las leyes de la bioquímica. No obstante lograr organizar este proceso tan complejo significó recurrir a la regularidad de millones de años del ADN. Este supuesto como postulado fundamental de la necesidad creó un dualismo metodológico que en la escuela del diseño significó la introducción de un agente externo responsable del direccionamiento de los procesos.

El argumento del relojero ciego dio paso a la introducción de una inteligencia responsable de la cadena de reacciones necesarias para manejar la bioquímica de la célula. El ejemplo más representativo ha sido el flagelo de una bacteria operado mediante una estructura molecular como la ATPasa.

### **3.1.2. Las proteínas como estructuras constituidas por unidades jerárquicamente dispuestas**

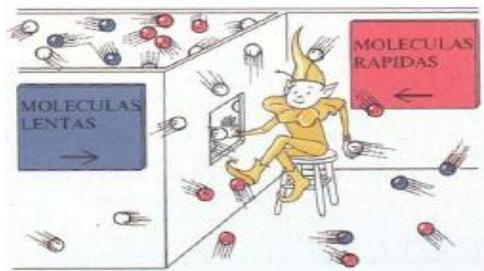
Las macromoléculas más versátiles y abundantes de los seres vivos son las proteínas, son los productos finales de los genes. Todos los caracteres genéticos de un individuo dependen de las proteínas que expresan. Aunque muchas de ellas están implicadas en roles estructurales y pueden ser metabolizadas, la mayor parte son enzimas, es decir catalizadores biológicos. Todas las proteínas son polímeros y los monómeros que se combinan para formarlas son los  $\alpha$ -aminoácidos [32].

Muchas proteínas están formadas por dominios, es decir por secuencias discretas de aminoácidos. Tales dominios presentan, en números casos funciones bien definidas que contribuyen a la actividad general de la proteína. Algunas de esas unidades modulares saltan frecuentemente de un punto a otro dentro de una

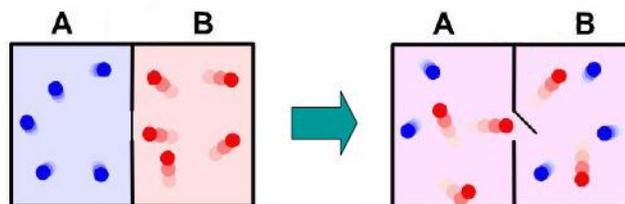
misma proteína y de una a otra proteína, en el curso de la evolución la movilidad evolutiva de los módulos no se circunscribe a los saltos operados en el material genético de una misma especie, en ciertos casos, los bloques en cuestión cruzan la barrera inter-específica para viajar, por ejemplo, de una célula animal a una bacteria [11].

Las proteínas son así cadenas largas de unas moléculas pequeñas denominadas aminoácidos. Veinte aminoácidos diferentes cada uno con su propia forma y estructura química, constituyen todas las proteínas que se encuentran en la naturaleza. Las propiedades de unas proteínas dependen de los aminoácidos utilizados en su construcción y en especial del orden en que están encadenados y lo que es más destacable, la secuencia de aminoácidos determina el modo de plegamiento de la proteína para dar una estructura tridimensional, funcional o activa [13].

Las proteínas (las enzimas) son el principal ejemplo que encontramos en la naturaleza de los demonios de Maxwell. (Ver Figura 5.1.)



(a)



(b)

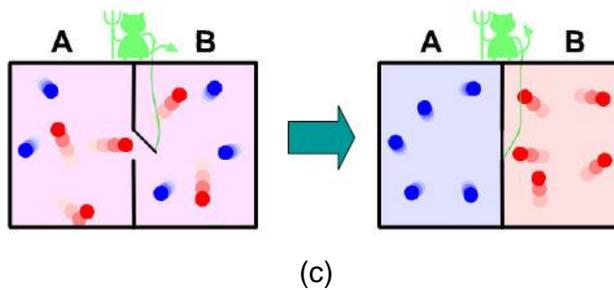


Figura 3.1. Representación gráfica de los Demonios de Maxwell: (a) Maxwell y sus Demonios. (b) Proceso espontaneo (c) Proceso controlado por un demonio de Maxwell. (Fuente: Mendieta, J., Estructura de Proteínas, Máster en Biofísica 2007-2008, CBM Severo Ochoa, Universidad Autónoma de Madrid)

Las enzimas aceleran las reacciones químicas sin violar el segundo principio de la termodinámica. Lo hacen disminuyendo la energía del estado de transición y por tanto la barrera de activación

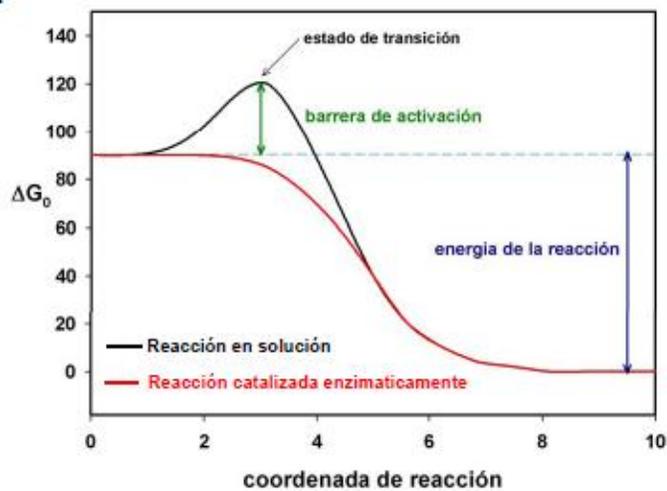


Figura 3.2. Cambios de energía en las proteínas. (Fuente: Mendieta, J., Estructura de Proteínas, Máster en Biofísica 2007-2008, CBM Severo Ochoa, Universidad Autónoma de Madrid)

La estabilización del estado de transición se realiza mediante la formación de interacciones no covalentes con los reactivos en una conformación idónea para dar lugar a la reacción química. Las proteínas son las únicas macromoléculas capaces de reconocer e interactuar con otras moléculas muy dispares [13].

## Definición estructural de proteína

Proteínas: son polipéptidos de longitud NO menor de 70 aminoácidos, que se PLIEGA en una estructura 3-D para realizar su función (ver Figura 3.3.).

Los sillares de las proteínas: químicamente las proteínas son heteropolímeros formados por la condensación de 20 unidades distintas: los aminoácidos proteicos.

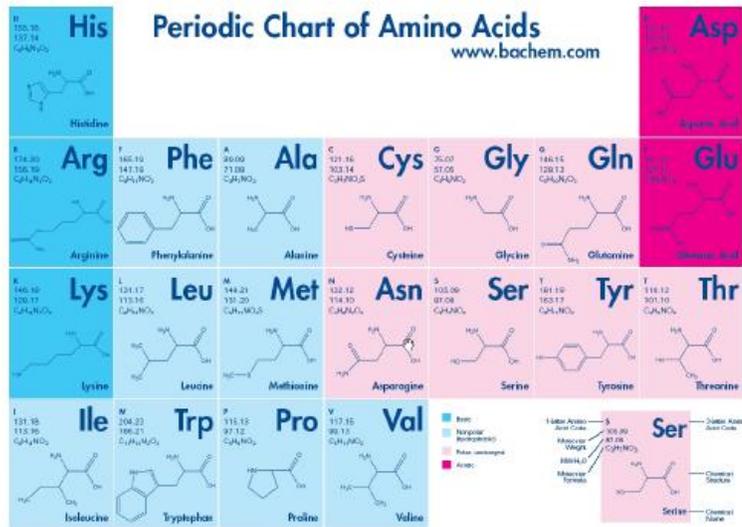


Figura 3.3. Los 20 aminoácidos proteicos codificados genéticamente. (Fuente: Mendieta, J., Estructura de Proteínas, Máster en Biofísica 2007-2008, CBM Severo Ochoa, Universidad Autónoma de Madrid)

## Organización jerárquica de la estructura de las proteínas

Las proteínas son sistemas complejos en el sentido científico del término, es decir, que sus propiedades no son exclusivamente la suma de las partes que las forman sino que se generan a partir de las relaciones entre sus componentes (propiedades emergentes). Una característica muy importante de las proteínas es su organización jerárquica: primaria, secundaria, terciaria y cuaternaria (ver Figura 3.4).

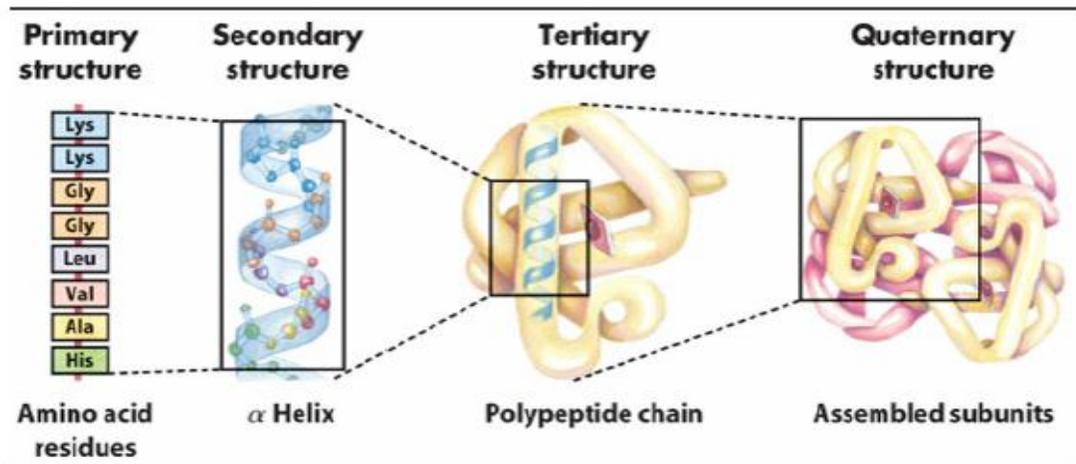


Figura 3.4. Estructura jerárquica de las proteínas. (Fuente: Mendieta, J., Estructura de Proteínas, Máster en Biofísica 2007-2008, CBM Severo Ochoa, Universidad Autónoma de Madrid)

Dentro de la estructura primaria se encuentra el enlace Peptídico (Figura 3.5.), que es un enlace covalente de tipo AMIDA.



Figura 3.5. Estructura del enlace peptídico: estabilización por resonancia. (Fuente: Mendieta, J., Estructura de Proteínas, Máster en Biofísica 2007-2008, CBM Severo Ochoa, Universidad Autónoma de Madrid)

Las características estructurales del enlace peptídico (Figura 3.6.) son: la estructura es planar, es decir, no hay rotación libre alrededor enlace C-N amida, pero sin embargo, la rotación es libre alrededor de otros enlaces de la traza peptídica (N-C $\alpha$ ; C-C $\alpha$ '). El oxígeno carbonílico tiene carga parcial negativa y el nitrógeno amida, carga parcial positiva (dipolo eléctrico).

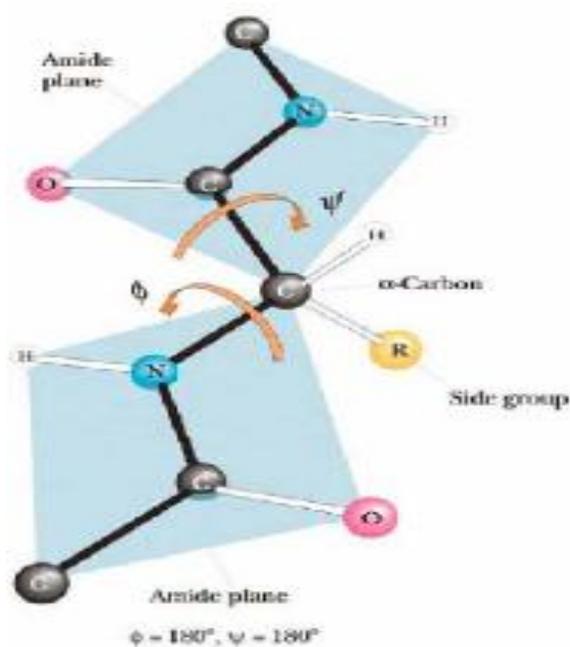


Figura 3.6. Características estructurales del enlace peptídico. (Fuente: Mendieta, J., Estructura de Proteínas, Máster en Biofísica 2007-2008, CBM Severo Ochoa, Universidad Autónoma de Madrid)

La conformación del esqueleto peptídico puede describirse en términos de dos ángulos diedros:  $\phi$  y  $\psi$ .  $\phi$  es el ángulo diedro para el enlace N-C $\alpha$  y  $\psi$  el ángulo diedro para el enlace C $\alpha$ -C.

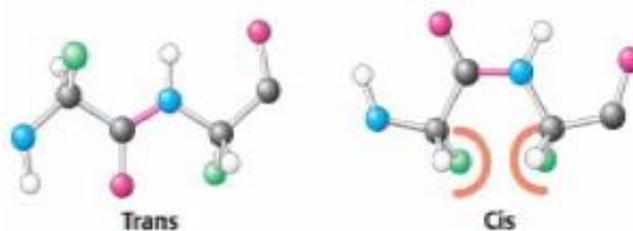


Figura 3.7. Conformaciones Trans y Cis del enlace peptídico. (Fuente: Mendieta, J., Estructura de Proteínas, Máster en Biofísica 2007-2008, CBM Severo Ochoa, Universidad Autónoma de Madrid)

Prácticamente todos los enlaces peptídicos en proteínas se disponen en configuración Trans (excepto Prolina).

Christian Anfinsen demostró que: La secuencia de aminoácidos especifica la estructura. (ver Figura 3.8.)

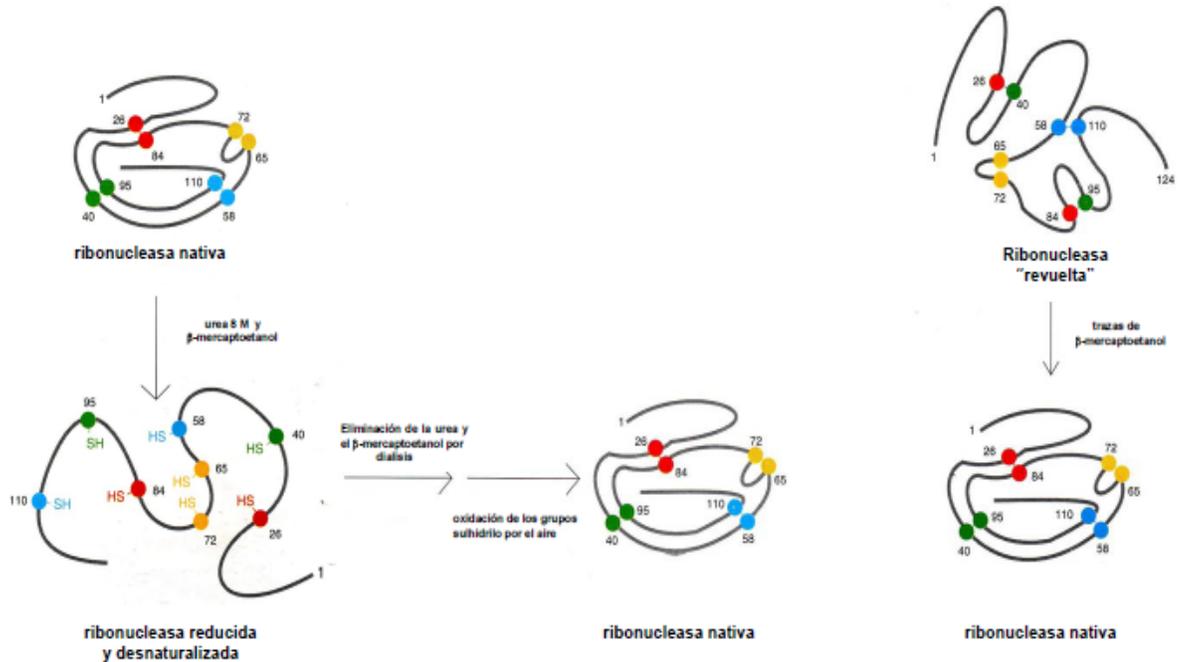


Figura 3.8. Estructura tridimensional de las proteínas. (Fuente: Mendieta, J., Estructura de Proteínas, Máster en Biofísica 2007-2008, CBM Severo Ochoa, Universidad Autónoma de Madrid)

### ¿Búsqueda aleatoria de la estructura de mínima energía?

La paradoja de Levinthal Cyrus Levinthal imagino una proteína simplificada de 100 residuos en la cual cada aminoácido fuese capaz de adoptar tres conformaciones distintas. El número total de conformaciones seria de  $3^{100}$  ( $5 \cdot 10^{47}$ ). Si consideramos que el tiempo que necesita una conformación para transformarse en otra es de  $10^{-13}$  s, el tiempo total de búsqueda de la estructura de menor energía seria de  $5 \cdot 10^{47} \times 10^{-13}$  o sea de  $5 \cdot 10^{34}$  s ( $1.6 \cdot 10^{27}$  años). (Ver Figura 3.9.)

Nota: La edad del universo se estima en 14000 millones de años ( $1.4 \cdot 10^{11}$  años)

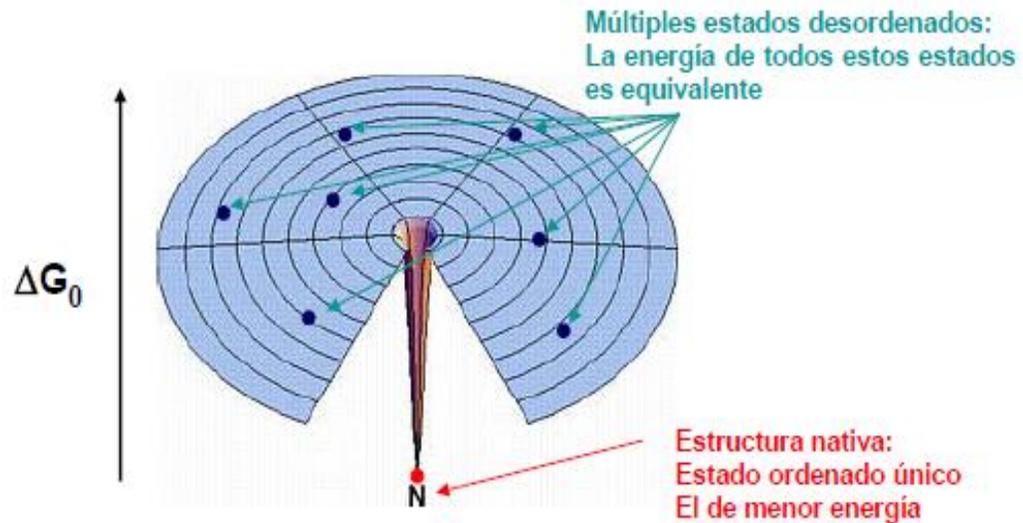


Figura 3.9. Superficie de energía de una proteína 1. (Fuente: Mendieta, J., Estructura de Proteínas, Máster en Biofísica 2007-2008, CBM Severo Ochoa, Universidad Autónoma de Madrid)

### ***¿Existe una ruta definida de plegamiento?***

Una forma de superar la paradoja de Levinthal sería la existencia de una ruta definida de plegamiento que pasara obligatoriamente por determinados estados (intermediarios) cada uno de los cuales tuviera menor energía que el anterior hasta llegar al estado nativo.

Experimentalmente se han detectado algunos intermediarios del proceso de plegamiento. En particular en varias proteínas se ha detectado una estructura compacta (globular) que retiene un elevado contenido de las conformaciones locales (estructura secundaria nativa), pero que no mantiene las interacciones de rango mayor (estructura terciaria). (Ver Figura 3.10.)

A dicha estructura se la conoce como estado de glóbulo fundido (molten globular state).

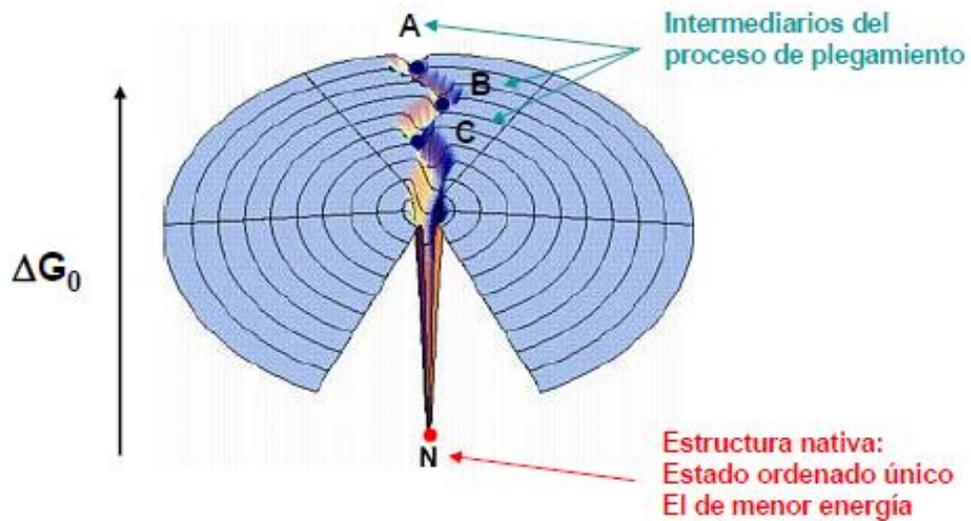


Figura 3.10. Superficie de energía de una proteína 2. (Fuente: Mendieta, J., Estructura de Proteínas, Máster en Biofísica 2007-2008, CBM Severo Ochoa, Universidad Autónoma de Madrid)

### ***El embudo de plegamiento***

En la actualidad, datos experimentales sugieren que la ruta de plegamiento en la mayoría de las proteínas no es única. Teniendo en cuenta estos datos, se ha propuesto una solución alternativa a la paradoja de Levinthal. Se supone que la superficie de energía de la proteína se asemeja a un embudo y que la proteína se pliega por rutas diferentes dependiendo de la conformación en la que se encuentre en el estado desplegado. Cada ruta seguiría la línea de menor energía desde esta conformación hasta la estructura nativa. Una imagen muy ilustrativa de este proceso nos la da la forma en que las gotas de agua alcanzan el fondo de un embudo siguiendo las líneas de máxima pendiente y sin necesidad de recorrer toda la superficie del embudo (ver Figuras 3.11 y 3.12).

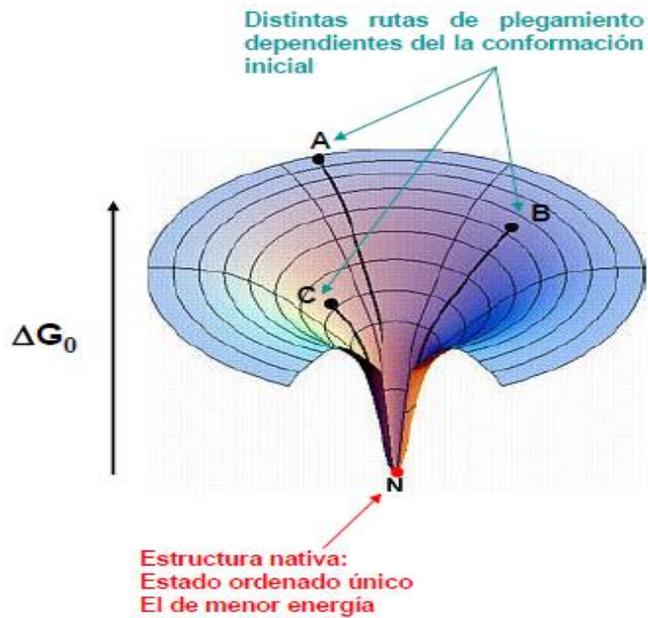


Figura 3.11. Superficie de energía de una proteína 3. (Fuente: Mendieta, J., Estructura de Proteínas, Máster en Biofísica 2007-2008, CBM Severo Ochoa, Universidad Autónoma de Madrid)

***¿Qué ocurre con los intermediarios?***

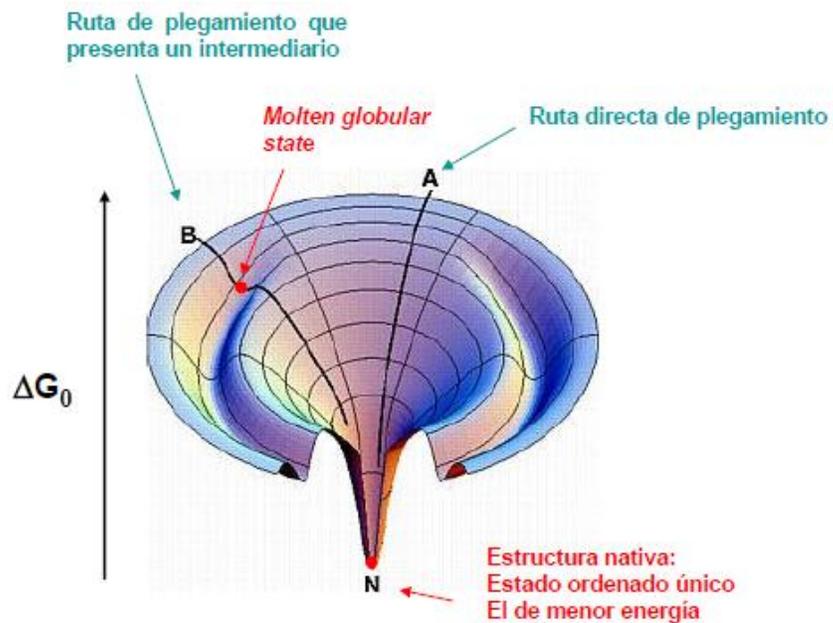


Figura 3.12. Superficie de energía de una proteína 4. (Fuente: Mendieta, J., Estructura de Proteínas, Máster en Biofísica 2007-2008, CBM Severo Ochoa, Universidad Autónoma de Madrid)

Una visión más realista de la superficie de energía de las proteínas. (Ver Figura 3.13.)

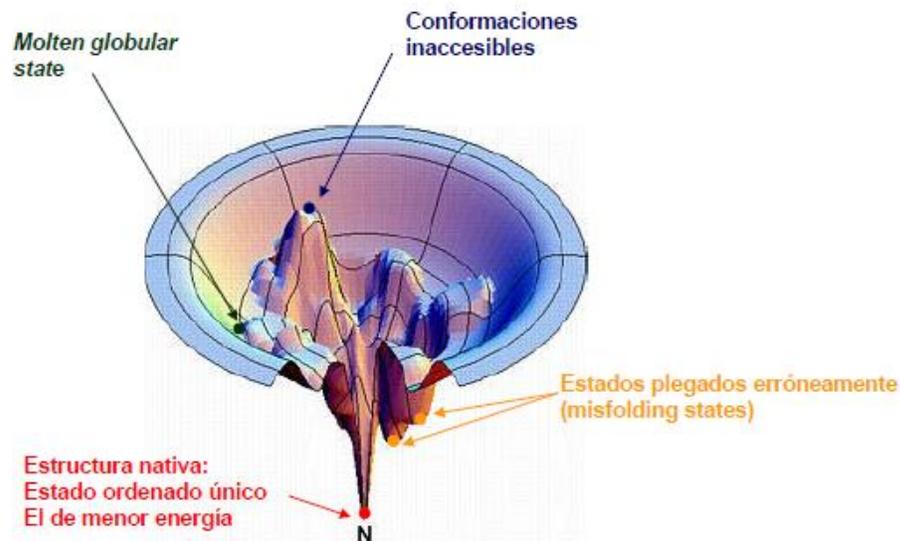


Figura 3.13. Superficie de energía de una proteína. (Fuente: Mendieta, J., Estructura de proteínas, Máster en Biofísica 2007-2008, CBM Severo Ochoa, Universidad Autónoma de Madrid)

### 3.1.3. ATPasa como unidad elemental del carácter orgánico vital

La ATPasa es una molécula gigante constituida por dos grandes unidades: Fo y F1, por lo que es también conocida como F<sub>0</sub>F<sub>1</sub> ATPasa (Figura 3.14), cataliza la formación de ATP (Adenosin Trifosfato) desde ADP (Adenosin Difosfato) y Pi (Fosfato inorgánico), en procesos conocidos como Fosforilación Oxidativa (conducida por oxidaciones en células animales y microorganismos) y fotofosforilación (conducida por luz en células de plantas). Una vez formado, el ATP es dividido en ADP y Pi proveyendo la energía para conducir un sinnúmero de procesos metabólicos incluyendo biosíntesis, contracción muscular, función cerebral y nerviosa.

## ATP sintasa

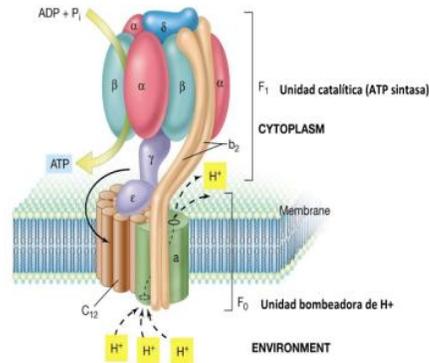


Figura 3.14. Modelo de la estructura de la ATPasa. (Fuente: ATP Sintetasa Estructura y función (Presentación). Roger Iván López Díaz)

La unidad  $F_0$  (Figura 3.15) tiene a su vez diez módulos transmembranales dispuestos simétricamente en forma cilíndrica. Cada módulo está formado por hélices alfa que a su vez están constituidos por aminoácidos hidrofóbicos, dispuestos regularmente. El eje del cilindro constituido por los 10 módulos, corresponde a una estructura de canal iónico. Los átomos de hidrógeno migran a través de estas unidades ubicándose en cada subunidad y migrando en el sentido del gradiente protónico.

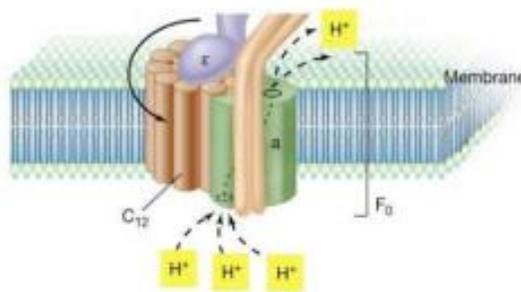


Figura 3.15. Unidad  $F_0$ . (Fuente: ATP Sintetasa Estructura y función (Presentación). Roger Iván López Díaz)

En este trabajo se asume la hipótesis consistente en un descenso de los iones siguiendo una trayectoria helicoidal desde la parte de alta concentración hasta una baja concentración que supone adaptarse a la simetría de la hélice alfa. Utilizando esta hipótesis, las wavelet, al hacer el cálculo de la multifractalidad, evidenciarán por medio de la invarianza de escala, la estructura modular correspondiente a la variable pseudopotencial asignada a cada aminoácido.

Por su parte la unidad F<sub>1</sub> (Figura 3.16) está constituida por 6 subunidades principales dispuestas axialmente con simetría hexagonal y subdividida a su vez en 3 grandes unidades alfa (3 $\alpha$ ), alternadas con tres grandes unidades beta (3 $\beta$ ), y por otras tres subunidades  $\gamma$ ,  $\delta$  y  $\epsilon$ . Estas unidades tienen una estructura terciaria constituida por hojas  $\beta$  con aminoácidos hidrofílicos dispuestos de cara al medio acuoso, donde se fijan las moléculas de ADP y los átomos de fósforo. Las subunidades  $\alpha$  y  $\beta$  hacen parte de la familia NTPasa y se encargan de la unión de nucleótidos, pero solamente  $\beta$  participa en la catálisis, debido a la interacción con  $\gamma$  cada subunidad  $\beta$  adopta una conformación diferente.

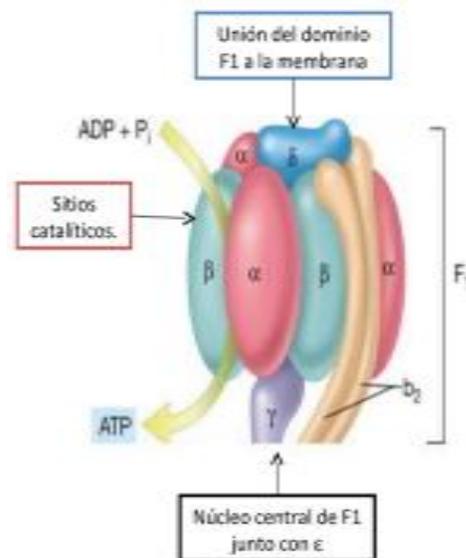


Figura 3.16. Unidad catalítica F<sub>1</sub>. (Fuente: ATP sintetasa Estructura y función (Presentación).

Roger Iván López Díaz)

Fo y F1 están conectadas por el tallo  $\gamma$  y  $\epsilon$  y por una columna externa formada por  $b_2$  y  $\delta$  (Figura 3.17). La enzima está formada por dos componentes funcionales: la unidad móvil, compuesta por el anillo c y tallo  $\gamma$  y  $\epsilon$ ; y, la unidad estática que es el resto de la molécula.

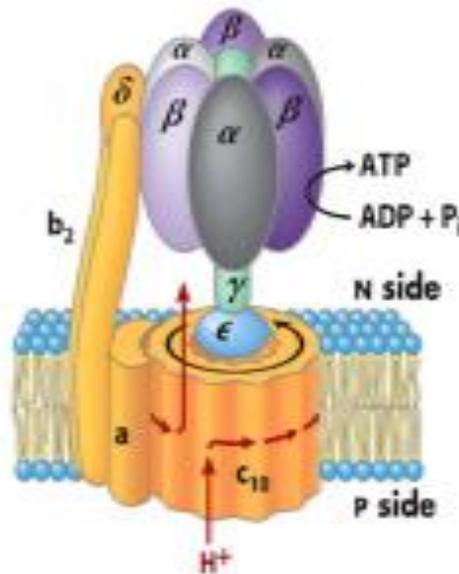


Figura 3.17. Conexión de las unidades Fo y F1. (Fuente: ATP Sintetasa Estructura y función (Presentación). Roger Iván López Díaz)

La unión fosforilizadora se consigue mediante la rotación de la unidad Fo que por medio de la subunidad  $\gamma$  transmite el movimiento axial rotacional a las unidades  $\alpha$  y  $\beta$  cuya simetría rotacional es alterada por esta rotación, generando un rompimiento de simetría que da origen a la catálisis de la formación de ATP. La operación de síntesis se denomina ATP sintetasa y la operación recíproca contraria al proceso descrito se conoce como ATP ASA.

La migración de protones a través de Fo y la composición de ATP es un proceso que ha sido abordado dada su regularidad, por medio de dinámica molecular

tratada adiabáticamente utilizando la separación movimientos rápidos separados de movimientos lentos. En este trabajo la relación de rápido-lento, corresponde al tratamiento de un sistema no lineal invariante de escala susceptible de ser tratado por medio de wavelet.

La ATPasa con su carácter de proteína dependiente del plegamiento tridimensional se hace más adecuada para la explicación del fenómeno vital que el concepto de información que es lo fundamental en el caso del virus, donde el carácter regular y cuasi cristalino pese a su complejidad mayor respecto a las proteínas no da para explicarlo como un fenómeno vivo. Por su parte el prion, que consiste en una proteína mal plegada, genera efectos devastadores, como es el caso de la encefalitis espongiforme (enfermedad de las vacas locas).

Esta característica de plegamiento tridimensional de las proteínas solo es detectable mediante las técnica RMN Nosy y Cosy, que permiten ubicar el plegamiento de las proteínas mediante el seguimiento de los puentes de hidrógeno.

Esta técnica es una aplicación de la mecánica cuántica consistente en la aplicación de campos magnéticos al comportamiento espacial de los protones caracterizados por la variable momento angular.

La química y la física reunidas mediante el seguimiento del protón responsable de los puentes de hidrógeno logrará una descripción fundamental a partir del seguimiento de la densidad protónica que químicamente corresponde al PH y físicamente corresponde a la densidad que en nuestro caso la base más adecuada para representarla es la wavelet.

## 3.2. SOPORTE FÍSICO-MATEMÁTICO

### 3.2.1. Fundamentos del análisis Wavelet

Las Wavelets<sup>1</sup> son un desarrollo relativamente reciente en matemática aplicada e introducida en el análisis de señales de principios de la década de 1980. Apareciendo de manera formal en el análisis de señales sísmicas con los trabajos de Morlet<sup>2</sup>, Grossman y Morlet<sup>3</sup>. Desde entonces significativos avances teórico-prácticos en el análisis Wavelet han sido desarrollados y usados en diversos campos.

Las características de esta técnica, en lo que respecta a la localización, tanto en el dominio del tiempo, como en el de las frecuencias de las señales que se pueden analizar mediante su implementación, le dan una ventaja considerable frente a las técnicas que le precedieron (primero la conocida transformada de Fourier, y luego su versión mejorada: la transformada de Fourier con ventana) [14]. En las últimas décadas se ha utilizado la transformada wavelet como un microscopio matemático, el cual se encuentra eficientemente adaptado para la revelación de la jerarquía que gobierna la distribución de las singularidades de las medidas multifractales [29].

El objetivo de este capítulo es hacer una descripción de los conceptos involucrados con la Transformada Wavelet y los necesarios para comprender como se realiza el Análisis Wavelet.

---

<sup>1</sup>Por razones a la nomenclatura encontrada en la literatura internacional y nacional, en este documento se ha decidido utilizar la designación wavelet en vez de su traducción: ondita u ondeleta.

<sup>2</sup>MORLET J.; ARENS G.; FORGEAU E.; y GIARD D. Wave Propagation and Sampling Theory. Geophysics, 47(2), 1982.

<sup>3</sup>MORLET, J.; y GROSSMANN, A. Decomposition of hardy functions into square integrable wavelets of constant shape, SIAM J. Math. Anal., 15(4), 1984.

### 3.2.1.1 De la transformada de Fourier a la transformada Wavelet

La transformada wavelet trabaja de manera similar a la transformada de Fourier, está última dice que una señal se puede descomponer y representar en una serie de funciones sinusoidales, logrando de esta forma realizar un análisis más sencillo.

La transformada de Fourier de una señal  $f(t)$  se define como:

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i2\pi\omega t} dt \quad (1)$$

Esta transformada presenta resultados satisfactorios si la señal  $f(t)$  es una señal periódica y suficientemente regular. Sin embargo para señales que presenten variaciones en el tiempo presenta un análisis más complejo con resultados deficientes. La transformada de Fourier detecta la presencia de una determinada frecuencia pero no brinda información adecuada acerca de la evolución en el tiempo de las características espectrales de la señal, del inicio y final de una señal finita, así como de la manifestación de una o varias singularidades.

Para contrarrestar las desventajas mencionadas anteriormente se origino la Transformada de Fourier con Ventana (TFV), que es una modificación de la anterior transformada. Para utilizar la TFV se toma una señal, la cual es dividida en una secuencia de segmentos de tiempo, en los que la señal puede ser considerada como cuasi-estacionaria y a los que se les aplica la TFV, definiéndose como:

$$Sf(\tau, \omega) \int_{-\infty}^{\infty} f(t)g(t - \tau)e^{-i\omega t} dt \quad (2)$$

Donde  $g(t - \tau)$  es una función ventana deslizante, la cual tiene un ancho fijo y se desplaza sobre la señal para determinar el espectro y  $\tau$  es el valor de ese desplazamiento.

Sin embargo esta transformada presenta desventaja en su análisis, teniendo que ver con el tamaño fijo de la ventana en tiempo y frecuencia, por lo que no podría adaptarse a las características de la señal en ciertos puntos de interés para su análisis. Por lo que para señales no estacionarias este método no sería muy adecuado.

Por lo anteriormente expuesto es que la Teoría Wavelet ha cobrado fuerza y vigencia, ya que permite un análisis más riguroso de señales no estacionarias, manteniendo los principios de análisis tiempo-frecuencia por medio de ventanas de análisis diferentes.

### 3.2.1.2. ¿Qué es una “Wavelet”?

Una “wavelet” es el nombre dado a una “pequeña onda u ondita” que tiene su energía concentrada en un periodo de tiempo determinado, lo cual proporciona una herramienta para el análisis de fenómenos transitorios no estacionarios, como se ilustra en la Figura 3.18, “Wavelet” es una función oscilatoria de longitud finita al estilo de la función ventana vista en el contexto de la Transformada de Fourier de Tiempo Corto.

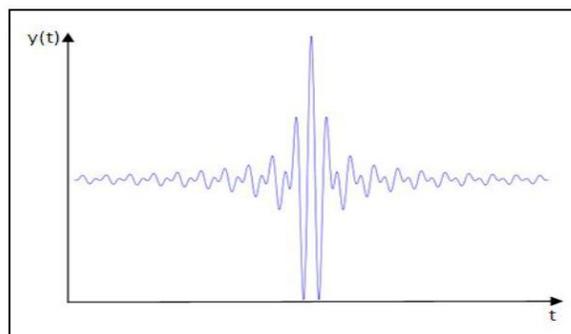


Figura No. 3.18. Onda tipo Wavelet

Esta Transformada es eficiente y de rápida transitoriedad para el análisis local de señales no estacionarias, al igual que la Transformada de Fourier con Ventana mapea la señal en una representación de tiempo-escala. El aspecto temporal de las señales es preservado. La diferencia está en que la Transformada Wavelet provee análisis de multiresolución con ventanas dilatadas. El análisis de las frecuencias de mayor rango se realiza usando ventanas angostas y el análisis de las frecuencias de menor rango se hace utilizando ventanas anchas [23].

Las Wavelets, funciones bases de la Transformada Wavelet, son generadas a partir de una función Wavelet básica, mediante traslaciones y dilataciones. Estas funciones permiten reconstruir la señal original a través de la Transformada Wavelet inversa. La Transformada Wavelet no es solamente local en tiempo, sino también en frecuencia.

### **3.2.1.3. Wavelet madre**

Como se menciona existe una wavelet básica, esta wavelet sin traslaciones ni dilataciones es llamada wavelet madre [14]:

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t - \tau}{s}\right) \quad (3)$$

Donde  $\tau$  representa el desplazamiento de la señal en el tiempo o espacio y  $s$  es el factor de escala de la wavelet, permitiendo contracciones o dilataciones de la señal;  $\tau$  y  $s$  son números reales que varían en el análisis dando origen a las wavelets hijas. Se utilizan siempre factores de escala  $s > 0$ . Las wavelets son dilatadas cuando la escala  $s > 1$ , y son contraídas cuando  $s < 1$ . Así, cambiando el valor de  $s$  se cubren rangos diferentes de frecuencias. Valores grandes del parámetro  $s$  corresponden a frecuencias de menor rango o una escala grande de

$\psi_{\tau,s}(t)$  . Valores pequeños de  $s$  corresponden a frecuencias de mayor rango o una escala muy pequeña de  $\psi_{\tau,s}(t)$  [10].

#### **3.2.1.4. Aplicaciones de las Wavelets**

La teoría de las “Wavelet” tiene muchas aplicaciones reales y aportes en diferentes campos, siendo esta herramienta muy joven en comparación con las Transformadas de Fourier.

La Transformada Wavelet es utilizada para múltiples fines como: detectar discontinuidades, puntos de rupturas, identificar frecuencias puras, reducción del ruido, comprimir información, aproximar funciones, solución de ecuaciones diferenciales en señales unidimensionales, dinámica molecular, la astrofísica, la geofísica de los sismos, la óptica, el estudio de las turbulencias y la mecánica cuántica; entre otros campos muy variados como el análisis de electrocardiogramas, el estudio del ADN y las proteínas, el análisis de sangre, el reconocimiento de voz, meteorología, en el campo de la biometría y análisis Multifractal.

Las imágenes de señales bidimensionales son la aplicación más famosa de esta transformada ya sea por la compresión o el procesamiento de estas, es decir reducción o ampliación. El eje central de las imágenes digitales JPEG-2000 y del método WSQ (del inglés Wavelet Scalar Quantization, cuantificación escalar de wavelets) que utilizó el FBI (Federal Bureau of Investigación) para comprimir su base de datos de huellas dactilares. Esta aplicación es posible gracias a los coeficientes “wavelet”, que se presentaran más adelante; información que sin ser la imagen misma posee los recursos suficientes. En este contexto, se puede pensar que las “wavelets” contienen los componentes básicos de una imagen.

Es importante aclarar que no son todas las aplicaciones que existen para dicha transformada, esto es solo una prueba muy pequeña de lo que se puede conseguir aplicándose por sí sola tal como lo marca la historia y el límite de nuestra imaginación.

### **3.2.1.5. Propiedades de las funciones Wavelet**

La elección de un tipo “wavelet” para aproximar una determinada señal requiere de un balance entre diferentes propiedades, tales como la suavidad, localización espacial y temporal, la localización de frecuencia, la habilidad para presentar funciones polinómicas locales, la ortogonalidad y la simetría, estas propiedades se discuten a continuación: [24]

Suavidad. En varias aplicaciones, las “wavelet” deben ser lo suficientemente suaves como para poder representar eficientemente las características de la señal que se desee aproximar. La suavidad en la “wavelets” se mide por el número de derivadas que existen, relacionándose también, con el número de momentos nulos.

Localización espacial y temporal. Una propiedad muy importante de las “wavelet” es su habilidad para localizar las características del fenómeno analizado en espacio y en tiempo.

Localización de frecuencias. Las wavelets no solo localizan características en tiempo y en espacio, sino también en frecuencia. En general, las funciones “wavelets” más suaves tienen mejores propiedades de localización de frecuencias.

La localización espacial y temporal, así como la frecuencial, es llamada o mejor conocida como soporte compacto, que le da la propiedad a la wavelet de ser de

duración finita, permitiendo una menor complejidad en los cálculos, mejor resolución en tiempo y frecuencias.

*Simetría.* Las “wavelets” ortogonales de soporte compacto no son simétricas, excepto la wavelet de Haar.

*Ortogonalidad.* Característica fundamental en el análisis con la transformada wavelet, ya que la información capturada por una función wavelet es completamente independiente de la información capturada por la función de escalamiento por lo tanto son mutuamente excluyentes. De esta manera no hay superposición de la representación de los datos en el análisis del dominio de la frecuencia.

Las ventanas wavelet  $\psi(t)$  deben cumplir las siguientes condiciones [17]:

1.  $\int_{-\infty}^{\infty} \psi(t)dt = 0$ , que indica que el promedio de la función wavelet debe ser cero.
2.  $C_{\psi} = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < \infty$ , llamada la condición de admisibilidad, donde  $C_{\psi}$  es un valor constante.

*Momentos de desvanecimiento.* Este es un concepto muy importante en la teoría wavelet que permite conocer la forma de la wavelet, siendo un parámetro que brinda información acerca de que "tan hábil" puede ser la wavelet para suprimir un polinomio determinado, posibilitando saber la suavidad de la wavelet, ya que esta dado por el número de momentos de desvanecimiento que se tenga. El i-ésimo momento de la wavelet se calcula con la siguiente integral: [4]

$$\int_{-\infty}^{\infty} \psi(t)t^i dt = 0 \quad (4)$$

De lo anterior se determina que una función tiene  $v$  momentos de desvanecimientos si la integral es igual a cero para  $i = 0, \dots, v - 1$ , de donde se toma que todas las señales que tengan la forma polinomial del tipo

$$y(t) = \sum_{m=0}^v c_m x^m \quad (5)$$

tienen cero coeficientes de wavelet, el concepto de coeficientes se explicará más adelante. El orden de la transformada wavelet es dado normalmente por el número de momentos de desvanecimientos que existan. Se puede percibir la palabra momentos como la referencia a un promedio, de tal manera que cuando el valor promedio de una wavelet sea cero, se tiene un momento de desvanecimiento, y los polinomios de grado menor a  $v$  serán eliminados.

### **3.2.1.6. Tipos de Wavelet**

Existen muchos y diferentes tipos de ventanas wavelet que son utilizadas y que tienen definiciones establecidas, por lo que su elección depende de la aplicación específica que se le vaya a dar. Entre las más conocidas están la Wavelet: Haar, Gaussiana, "Mexican Hat" o "Sombrero Mexicano", Symmlet, Coiflet, Morlet o las Daubechies, una de las cuales se utilizo en el presente trabajo.

### **Wavelets Daubechies**

La familia de Wavelets Daubechies, en honor a Ingrid Daubechies (una de las científicas más reconocidas y brillantes en la investigación sobre Wavelets) se le llama a una base de wavelets ortonormales, suaves y de soporte compacto. Se escribe  $dbN$  en donde  $N$  es el orden de la función y  $db$  el nombre que se le da a este conjunto,  $N$  depende del orden de desvanecimientos que se deseen, es un

entero positivo y denota el número de coeficientes de filtro que tiene esa wavelet; por ejemplo la Daubechies de orden 1 (*db1*), es la Wavelet Haar, que es la ventana wavelet más simple y antigua.

En 1988 Ingrid Daubechies [10], con el interés de construir bases ortogonales de funciones suaves y con el requerimiento de varios momentos nulos, pudo construir una familia de bases ortonormales, suaves y de soporte compacto. Las wavelets de Daubechies son las que tienen mayor cantidad de momentos nulos para su soporte.

En la figura 3.19 se puede observar la Daubechies de orden 12 (*db12*) que fue la que se utilizó en este trabajo, ya que en total son 40 tipos de wavelets Daubechies.

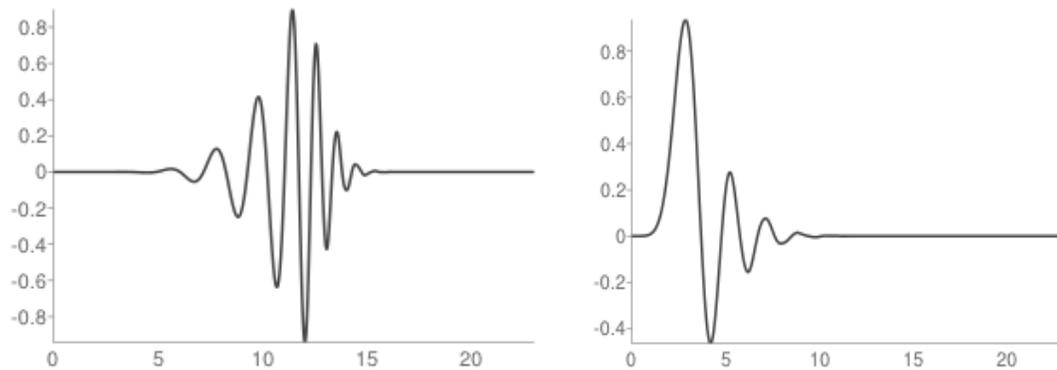


Figura No. 3.19. Daubechies 12 (*db12*) : funciones wavelet y de escalamiento

El conjunto  $\{\psi_{j,k}\}$  constituye una base ortonormal de wavelets, y la wavelet  $\psi(x)$  es suave, por lo tanto debe tener momentos nulos o de desvanecimiento, y cuanto mayor suavidad, más cantidad de momentos nulos, E4Resto fue demostrado por Walnut, [28], [27].

Esta familia de funciones se comportan muy bien para el análisis de señales discontinuas, no derivables y adicionalmente la proporción de coeficientes

resultantes iguales a cero es mayor, facilitando las operaciones y el almacenamiento de datos [6].

El análisis mediante wavelet puede hacerse para fenómenos continuos o discretos. A continuación se mencionara de manera breve la forma continua y de manera un poco más extendido la transformada discreta.

### 3.2.1.7. Transformada Wavelet continua

Para hacer el análisis de una señal se multiplicará cada punto de dicha señal por la wavelet que se haya elegido, cuyas características de escala y traslación serán permanentes para todo el proceso, después cada una de las muestras se suma y se obtiene la señal trasladada del dominio del tiempo al dominio de la frecuencia y el tiempo.

La transformada continua de Wavelets de una señal  $f(t)$  esta definida por la siguiente ecuación:

$$CWT(\tau, s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} f(t) \odot \psi\left(\frac{t - \tau}{s}\right) dt \quad (6)$$

Donde  $s$  es la escala y  $\tau$  es la traslación, son números reales y  $f(t)$  es la función a analizar.

Es importante aclarar que el proceso de transformación de una señal es reversible, a esto es lo que se le conoce como Transformada Inversa Continua de Wavelets, la cual permite volver a obtener la función original y está dada por la siguiente ecuación.

$$f(t) = \frac{1}{C_{\psi}} \int_0^{\infty} \int_{-\infty}^{\infty} CWT(\tau, s) \frac{1}{\sqrt{s}} \psi\left(\frac{t - \tau}{s}\right) d\tau \frac{ds}{s^2} \quad (7)$$

Donde  $C_\psi$  es una constante que se determina por la wavelet que se ha utilizado en la transformación, se conoce como constante de admisibilidad y se define como:

$$C_\psi = \sqrt{2\pi \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega} < \infty \quad (8)$$

Donde  $\hat{\psi}(\omega)$  es la Transformada de Fourier de  $\psi(\omega)$  que es la función madre de la Wavelet utilizada en la transformación inicial. Básicamente  $C_\psi$  debe tener un valor finito para que se pueda hacer una transformación inversa. A este proceso de reconstrucción de la señal se le conoce como síntesis de la señal y generalmente se realiza después del proceso de análisis o descomposición de esa misma señal [19].

En la transformación continua wavelet, los parámetros de dilatación  $s$  y traslación  $\tau$ , son números enteros reales, por lo tanto al analizar señales es necesario tomar algunos valores de  $s$  y  $\tau$  de las infinitas posibilidades que existen. Uno de los métodos consiste en hacer que las escalas varíen en potencia de dos, por la cual el muestreo de  $s$  se conoce como "dilatación binaria" y la traslación se realiza con base en múltiplos del tamaño de la ventana y recibe el nombre de "traslación diádica". Es decir:

$$s = 2^j \quad (9)$$

$$\tau = k * s = k * 2^j \quad (10)$$

Donde:

$j$ : Índice de discretización de  $s$ . Tomando valores enteros positivos.

$k$ : Índice de discretización de la  $\psi$ . Tomando valores enteros positivos.

Para un número de datos ( $n$ ) es necesario determinar el valor máximo de  $j(j_{m\acute{a}x})$  el cual garantiza que el valor mínimo de  $k$  sea entero. Este término nuevo  $j_{m\acute{a}x}$  se puede ver como la cantidad máxima de niveles para los que se puede analizar la señal.

$j_{m\acute{a}x} = \left\| \left\lfloor \frac{\log_2(n)}{2} \right\rfloor \right\|$ , las barras dobles simbolizan la función parte entera.

Reemplazando los parámetros de escala  $s$  y traslación  $\tau$  en la ecuación (6) por sus respectivos equivalentes discretizados dados en la ecuación (9) y (10) se llega a la siguiente expresión (11) que se denomina Transformada de Wavelet Continua en el Tiempo pero Discretizada para los parámetros  $a$  y  $b$ :

$$DCWT(j, k) = 2^{-j/2} * \int_{-\infty}^{\infty} f(t) \otimes \psi(2^{-j} * t - k) dt \quad (11)$$

El método de discretización de  $a$  y  $b$  se basa en potencias de dos, de tal forma que la serie de datos a analizar debe ser también en potencia de dos. En la implementación de la transformada, se muestra como se modificó el tamaño de las señales de análisis unidimensionales y bidimensionales para que fuesen potencia de dos.

### **3.2.1.8. Transformada Wavelet discreta**

El cálculo de la transformada wavelet para todas las posibles escalas supone una gran cantidad de información. Escoger solo aquellas escalas y posiciones que resulten interesantes para ciertos estudios es una tarea complicada. Si se escogen aquellas escalas y posiciones basadas en potencia de dos como se presentó en el marco continuo, los cálculos computacionales serían más ágiles. El análisis de la Transformada de Wavelet en Tiempo Discreto (*DWT*) se describe matemáticamente:

$$DWT(j, k) = 2^{-j/2} * \sum_{t=0}^{N-1} f(t) \otimes \psi(2^{-j} * t - k) \quad (12)$$

Es decir:

$$DWT(j, k) = 2^{-j/2} * \sum_{t=0}^{N-1} f(t) \otimes \psi_{jk}(t) \quad (13)$$

Al igual que para la transformación  $CWT$ , la señal se puede reconstruir a partir de los coeficientes de transformación, mediante una transformación inversa discreta:

$$f(t) = \sum_{\forall j} \sum_{\forall k} DWT(j, k) \otimes \psi_{jk}(t) \quad (14)$$

Para las señales la información más importante desde el punto de vista de su constitución podría decirse que se encuentra en las frecuencias bajas, mientras que en las altas frecuencias se encuentran los detalles o matices de la señal. El análisis wavelet en tiempo discreto permite descomponer la señal en aproximaciones o análisis  $A$  y detalles  $D$  para hacer un análisis por separado de la forma general y los detalles de la señal.

Para encontrar se debe partir de la expresión matemática que se tiene en la ecuación (11); si se fija una escala  $j$  y se nombrará todas las traslaciones  $k$  a lo largo de la escala de tiempo se obtiene una subseñal de detalle para esta escala que se bautizara  $D_j$ , como se muestra a continuación:

$$D_j(t) = \sum_{\forall k} DWT(j, k) \otimes \psi_{jk}(t) \quad (15)$$

Si se suman todas las subseñales de detalle a lo largo de las escalas  $j$  se recupera la señal original:

$$f(t) = \sum_{\forall j} D_j(t) \quad (16)$$

Como se observa en la anterior expresión, la señal original se cubre con la sumatoria de todos los detalles para infinitas escalas de análisis. Un procedimiento bastante largo y tedioso.

Si la traslación de las wavelets está limitada por la duración temporal de la señal es necesario un número finito ( $J$ ) de escalas para recuperar la señal. Al hacer esta transformación hasta un número finito de escalas " $j < J$ " se pretende reconstruir la señal original con cierta cantidad de escalas disponibles, lo cual es necesario completar la información correspondiente a  $DWT(j, k)$  para escalas superiores a  $J$ , en este caso se tiene:

$$f(t) = \sum_{j \leq J} D_j(t) + \text{complemento} = \sum_{j \leq J} D_j(t) + A_j(t) \quad (17)$$

Donde:  $A_j(t)$  es la subseñal de aproximación o de análisis

Los límites, ahora de la sumatoria han cambiado y las subseñales de aproximación se relacionan a diferentes escalas mediante:

$$A_{j-1} = A_j + D_j \quad (18)$$

Es importante tener en cuenta que la subseñal de aproximación y la subseñal de detalles son complementarias y ortogonales.

La cantidad de detalles que se tengan en cuenta en la sumatoria es la cantidad de *niveles*  $n$  que se analizan en una señal.

Para obtener la subseñal de aproximación  $A(t)$  en el análisis wavelets se recurre a unas funciones bases llamadas funciones de escalamiento, denotadas por:  $\phi_{j,k}(t)$ , lo cual para cada función de "wavelet" está definida una función de escalamiento,

que se presentarán más adelante. De tal forma la ecuación (14) se reescribe como:

$$f(t) = \sum_{\forall j} \sum_{\forall k} DWT^A(j, k) \otimes \phi_{jk}(t) + \sum_{j \leq J} \left( \sum_{\forall k} DWT(j, k) \otimes \psi_{jk}(t) \right) \quad (19)$$

Donde:

$\sum_{\forall k} DWT^A(j, k) \otimes \phi_{jk}(t)$  : Función de escala, principio fundamental del concepto de escala.

Para calcular dicho término se tiene:

$$DWT^A(j, k) = \int_{-\infty}^{\infty} f(t) \otimes \phi_{jk}(t) \quad (20)$$

En la ecuación (19) la escala "j" es finita, mientras que la traslación no está limitada, es decir "se mira a unas escalas la señal y el resto se aproxima". Esto conduce a la teoría matemática conocida como Análisis Multiresolución (MRA), la cual hace una descomposición jerárquica de la señal en subseñales de detalles y aproximación.

### 3.2.2. Análisis multiresolución

El método de multiresolución es una técnica destinada (al menos, originalmente) a rebajar el costo computacional asociado a los métodos de alta resolución. El análisis de multiresolución permite estudiar una señal en diferentes resoluciones y extraer características propias de cada resolución [16].

La idea central de este análisis consiste en estudiar ( $f(t)$ ) a partir de aproximaciones más y más burdas, donde a cada aproximación se cancelan

algunas de las altas frecuencias o de los “detalles” de la señal original. La información que se elimina de una aproximación a la siguiente, equivale a:

$$D_n(t) = A_{n-1}(t) - A_n(t) \quad (21)$$

En la Figura 3.20 se presenta como funcionan el análisis de una señal con la transformada wavelet. Se tiene una señal de la cual se desea hacer una descomposición de niveles:

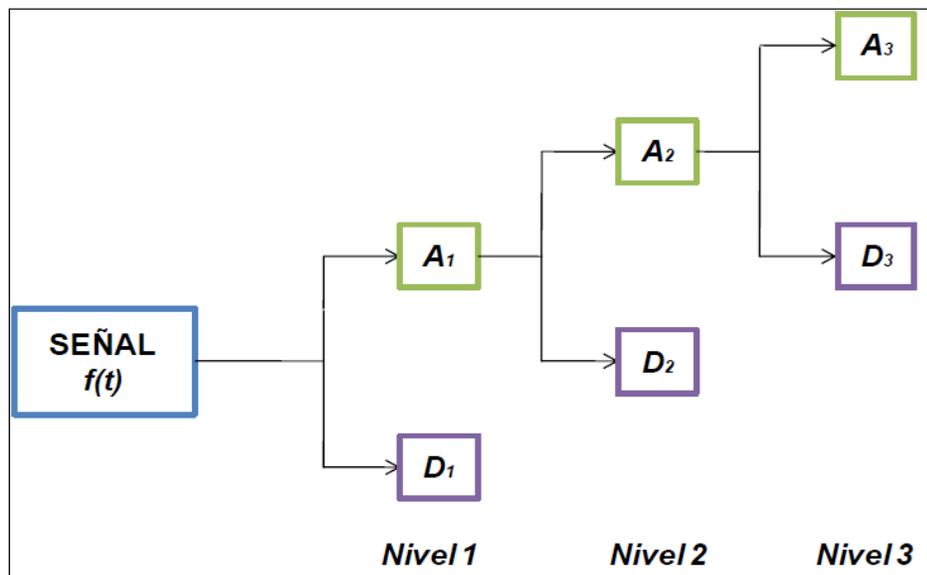


Figura 3.20. Proceso de descomposición de una señal con Análisis Multiresolución

Las cajas del diagrama que se observan en la Figura 5.20 nombradas con las letras  $C_A$  y  $C_D$  y corresponde a las subseñales de aproximación y detalles, respectivamente. El subíndice que acompaña las letras corresponde al nivel de descomposición. Por lo tanto para obtener la señal presentada en la Figura 5.20 a partir de la descomposición hecha y con respecto a la ecuación (19) se tiene:

$$SEÑAL = A_3 + D_3 + D_2 + D_1 \quad (22)$$

Es decir, la reconstrucción de la señal "*perfecta*" se da con las sumas de las subseñales de la descomposición de aproximación y detalles del último nivel y los detalles de los niveles inferiores.

A continuación se explica formalmente el Análisis Multiresolución que se define como una secuencia de subespacios cerrados  $V_j$ , que representa una aproximación en multiresolución del espacio  $L^2(\mathbb{R})$   $\{V_j \subset L^2(\mathbb{R}): j \in \mathbb{Z}\}$  las cuales cumplen las siguientes propiedades: [14]

- i.  $\dots V_{-2} \subset V_{-1} \subset V_{-0} \subset V_1 \subset V_2 \subset \dots \subset L^2(\mathbb{R})$ ; esto es  $V_j \subset V_{j+1}, \forall j \in \mathbb{Z}$
- ii.  $f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1}, \forall j \in \mathbb{Z}$
- iii.  $f(t) \in V_0 \Leftrightarrow f(t+1) \in V_0$
- iv.  $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$ , (Cuando la resolución se decrementa las aproximaciones contienen cada vez menos información y convergen a cero)  
 $\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R})$ , (Cuando la resolución se incrementa, la función aproximada debería converger a la función original)
- v.  $f(t) \in V_0 \Leftrightarrow f(t-k) \in V_0, \forall k \in \mathbb{Z}$
- vi. Existe una función de escala  $\phi(t) \in V_0$  tal que la familia  $\{\phi(t-k), k \in \mathbb{Z}\}$  constituye una base ortonormal de  $V$  [27].

### 3.2.3. Conceptos básicos de multifractalidad

La geometría fractal fue propuesta en los años sesenta y ganó aceptación inmediata en diversas áreas científicas debido a su flexibilidad para describir las formas irregulares encontradas en la naturaleza [20].

B. Mandelbrot postulo que "muchos de los patrones de la naturaleza son tan irregulares y fragmentarios que, comparándolos con la geometría euclidiana, muestran no solo un más alto grado, sino un muy distinto nivel de complejidad. El

número de las distintas escalas de extensión de patrones naturales es prácticamente infinito. La existencia de estos patrones nos reta a estudiar esas formas que la geometría euclidiana deja de lado por carecer de forma para investigar, por así decirlo, la morfología de lo amorfo. Respondiendo a este reto, he concebido y desarrollado una nueva geometría de la naturaleza, cuyo uso puede implementarse en diversos campos. Describe muchos de los patrones irregulares a nuestro alrededor, identificando una familia de formas a las que llamaré fractales" [3].

La palabra fractal (del latín *fractus*, roto, irregular, fragmentado, fracturado) fue introducida por Benoît Mandelbrot (1975, 1977) para hacer referencia a objetos demasiado irregulares como para ser descritos según la geometría tradicional, muy apropiada para objetos cuya dimensión es fraccionaria. Al estudio de los objetos fractales se le conoce, generalmente, como geometría fractal, que es una extensión de la geometría clásica y engloba la descripción, la clasificación y el análisis de subespacios geoméricamente 'complicados' [12].

Los fractales son entidades matemáticas, que abarcan otros espacios. Los primeros ejemplos de este tipo de objeto fueron figuras matemáticas como el polvo de Cantor, la curva de Koch (1904) y el triángulo de Sierpinski. Luego de éstos, que datan de finales de siglo XIX y principios del XX, vinieron los trabajos de Gaston Julia y Pierre Fatou sobre los fractales del conjunto de Julia (1918-19) y, varias décadas más tarde, los estudios de Benoît Mandelbrot y otros matemáticos sobre el conjunto de Mandelbrot, atractores extraños y bifurcaciones, entre otros. No obstante, los fractales están por todas partes. Hay muchos objetos "ordinarios" que, debido a su estructura o comportamiento, son considerados fractales naturales aunque no los reconozcamos como tales de primera instancia. Las nubes, las montañas, las costas, los árboles y los ríos son fractales naturales; se diferencian de sus contrapartes matemáticos por ser entidades finitas en vez de

infinitas. Ejemplos adicionales de fractales son el mercado de valores y el crecimiento poblacional [30].

Los fractales más simples son "auto-similares" y se generan repitiendo un patrón (generador) en un objeto inicial (iniciador). El generador puede adicionar o remover material del objeto original. La auto- semejanza de los fractales se puede entender como: el todo es igual a sus partes, salvo un factor de escala o como qué: el todo está formado por varias copias de sí mismo, solo que reducidas y colocadas en diferentes posiciones.

La principal teoría de los fractales es la invariancia de escala, término que indica que determinados rasgos de un sistema son independientes de la escala con la que se analizan. Un proceso fractal es aquel en el que el mismo proceso elemental tiene lugar a distintas escalas. Para el intervalo de escalas analizadas, los estadísticos del proceso siguen leyes potenciales caracterizadas por sus exponentes. La relación entre los estadísticos para las diferentes escalas consistirá simplemente en cocientes de escala [12].

### **3.2.3.1 Dimensión fractal**

*La dimensión fractal,  $D$* , es un número real que generaliza el concepto de dimensión ordinaria para objetos geométricos que no admiten espacio tangente [18]. La dimensión fractal es un exponente que da cuenta cuán completamente parece llenar un fractal el espacio, conforme se amplía el primero hacia escalas más y más finas. No existe una única dimensión fractal sino una serie de dimensiones que frecuentemente resultan equivalentes, pero que no siempre lo son. [7].

La noción de dimensión fractal (fraccional) provee una manera de medir qué tan rugosa es una curva. Normalmente, se considera que los puntos tienen dimensión

0, las líneas 1, las superficies 2 y los volúmenes 3. A este tipo de dimensión se lo llama dimensión topológica. Sin embargo una curva rugosa que recorre una superficie puede ser tan rugosa que casi llene la superficie en la que se encuentra. Se puede, entonces, pensar en la rugosidad como un incremento en la dimensión: una curva rugosa tiene una dimensión entre 1 y 2, y una superficie rugosa la tiene entre 2 y 3.

En términos generales, la dimensión fractal de un objeto geométrico es  $q$  si:

$$N(L)L^q = 1 \quad (23)$$

donde  $N(L)$  es el número de objetos elementales, o de unidades, de tamaño  $L$  que recubren, o que completan, el objeto, de donde se deduce, despejando  $q$ , que:

$$q = \log(N(L))/\log(1/L) \quad (24)$$

De esta manera, la dimensión fractal se ve expresada como la pendiente de una curva sobre un papel doblemente logarítmico [21].

El concepto de dimensión fractal describe el comportamiento de escala de las estructuras fractales [12] y tiene que ver con que al dividir una figura cuántas partes quedan de esta y a qué valor se debe elevar un exponente  $q$  para hacer la multiplicación entre el corte y las partes igual a 1.

El método del *Box-Counting* se utiliza para determinar la dimensión fractal de diversos objetos (usando la regresión no lineal planteada), en la cual al hacer la grafica se nota que la pendiente no es única, sino que cambia con la escala, deduciéndose que la medida del conjunto no es autosimilar. Para ello se utiliza el análisis multifractal en el cual no existe una dimensión única, sino un espectro de dimensiones.

### 3.2.3.2. Multifractalidad

Los fractales que necesitan ser descritos por un conjunto infinito de exponentes (dimensión fractal) se conocen como multifractales [15].

El término multifractal fue introducido por Frish y Parisi (1985), para hacer referencia a medidas caracterizadas por un conjunto infinito de índices de escala o dimensiones fractales [12]. En el que la función de partición  $Z(q, s)$  de orden  $q$  es calculada a partir de los valores  $DWT(\tau, s)$

$$Z(q, s) = \sum |DWT(\tau, s)|^q \quad (25)$$

Siendo  $q$  la dimensión,  $s$  la escala y  $DWTf(\tau, s)$  es la transformada wavelet de la función.

De lo anterior proviene una importante medida conocida como “Escala de Exponentes” que se representa con  $\tau(q)$  que es una de las funciones que caracterizan a los multifractales, y se define de la siguiente manera:

$$\tau(q) = \liminf_{s \rightarrow \infty} \left( \frac{\log Z(q, s)}{\log s} \right) \quad (26)$$

Y, la relación de  $Z(q, s)$  con la escala  $s$  es:

$$Z(q, s) \sim s^{\tau(q)} \quad (27)$$

El teorema de Jaffard [34] relacionan al  $\tau(q)$  como una transformada de Legendre de la función  $f(\alpha)$  para señales auto-similares, así:

$$\tau(q) = \min_{\alpha \in \Lambda} (q(\alpha + 1/2) - f(\alpha)) \quad (28)$$

donde  $f(\alpha)$  es el espectro de dimensiones fractales ( $\alpha$ ) o espectro de singularidades y  $\Lambda = [\alpha_{min}, \alpha_{max}]$  es el soporte de  $f(\alpha)$ .

La transformada de Legendre es invertible si y sólo si  $f(\alpha)$  es convexa, en cuyo caso:

$$f(\alpha) = \min_{q \in \Lambda} (q(\alpha + 1/2) - \tau(q)) \quad (29)$$

El espectro  $f(\alpha)$  de señales auto-similares es convexa. [35].

La forma genérica de las dos funciones se observa en la Figura 3.21.

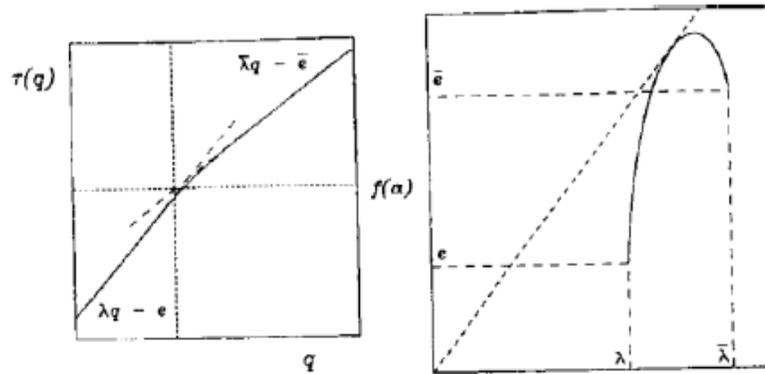


Figura 3.21. Forma genérica de las funciones  $\tau(q)$  y  $f(\alpha)$ . (Fuente: Solé, R., Manrubia, S. (1993). Orden y caos en sistemas complejos. Ediciones UPC., Cataluña).

## 4. PROCEDIMIENTO METODOLÓGICO

El análisis de la ATPasa propuesto en el presente trabajo se realizó mediante el desarrollo de los siguientes pasos: obtención de la secuencia de aminoácidos ATPasa, traducción de dicha secuencia en una serie numérica mediante la asignación del pseudopotencial, aplicación de la transformada wavelet, cálculo de la función de partición, obtención de la expresión para la energía libre y del nivel de complejidad o entropía, y finalmente la consecución de las graficas del espectro multifractal.

### 4.1. OBTENCIÓN DE SECUENCIAS

Se obtuvo y se trabajó con tres secuencias de aminoácidos de la proteína ATPasa de mitocondria de humano, bovino y rata, [2].

A continuación se presenta una tabla (Tabla 4.1.) en la que se encuentra el tipo de aminoácido y su correspondiente identificación alfabética, con la cual se forman las respectivas secuencias de aminoácidos.

Aminoácido	Identificación
Leu	L
Ile	I
Asn	N
Gly	G
Val	V
Glu	E
Pro	P
His	H
Lys	K
Ala	A
Tyr	Y
Trp	W
Gln	Q

Met	M
Ser	S
Cys	C
Thr	T
Phe	F
Arg	R
Asp	D

Tabla 4.1. Representación alfabética de los aminoácidos

Las secuencias de aminoácidos utilizadas de humano, bovino y rata se detallan a continuación.

#### Secuencia ATPasa de humano

```
MLGFVGRVAAAPASGALRRLTPSASLPPAQLLLRAAPTAVHPVRDYAAQT
SPSPKAGAATGRIVAVIGAVVDVQFDEGLPPILNALEVQGRETRLVLEVAQ
HLGESTVRTIAMDGTEGLVRGQKVLDSGAPIKIPVGPETLGRIMNVIGEPID
ERGPICKQFAPIHAEAPEFMEMSVEQEILVTGIKVVDLLAPYAKGGKIGLF
GGAGVGKTVLIMELINNVAKAHGGYSVFAGVGERTREGNDLYHEMIESGVI
NLKDATSKVALVYGQMNEPPGARARVALTGLTVAEYFRDQEGQDVLLFID
NIFRFTQAGSEVSALLGRIPSAVGYQPTLATDMGMTMQUERITTTKKSITSVQ
AIYVPADDLTDPAATTFHLDATTVLSRAIAELGIYPAVDPLDSTSRIMDPNI
VGSEHYDVARGVQKILQDYKSLQDIIAILGMDELSEEDKLTVSRARKIQRFL
SQPFQVAEVFTGHMGKLVPLKETIKGFQQILAGEYDHLPEQAFYMGPIEE
AVAKADKLAEHSS
```

#### Secuencia ATPasa de bovino

```
MLGLVGRVVAASASGALRGLSPSAPLPQAQLLLRAAPAALQPARDYAAQA
SPSPKAGATTGRIVAVIGAVVDVQFDEGLPPILNALEVQGRETRLVLEVAQH
LGESTVRTIAMDGTEGLVRGQKVLDSGAPIRIPVGPETLGRIMNVIGEPIDE
RGPIKTKQFAAIHAEAPEFVEMSVEQEILVTGIKVVDLLAPYAKGGKIGLFGG
AGVGKTVLIMELINNVAKAHGGYSVFAGVGERTREGNDLYHEMIESGVINL
KDATSKVALVYGQMNEPPGARARVALTGLTVAEYFRDQEGQDVLLFIDNIF
RFTQAGSEVSALLGRIPSAVGYQPTLATDMGMTMQUERITTTKKSITSVQAIY
VPADDLTDPAATTFHLDATTVLSRAIAELGIYPAVDPLDSTSRIMDPNIVG
SEHYDVARGVQKILQDYKSLQDIIAILGMDELSEEDKLTVSRARKIQRFLSQ
PFQVAEVFTGHLGKLVPLKETIKGFQQILAGEYDHLPEQAFYMGPIEEAVA
KADKLAEHSS
```

## Secuencia ATPasa de rata

```
MLSLVGRVASASASGALRGLNPLAALPQAHLLLRTAPAGVHPARDYAAQS  
SAAPKAGTATGQIVAVIGAVVDVQFDEGLPPILNALEVQGRESRLVLEVAQ  
HLGESTVRTIAMDGTEGLVRGQKVLDSGAPIKIPVGPETLGRIMNVIGEPID  
ERGPICKQFAPIHAEAPEFIEMSVEQEILVTGIKVVDLLAPYAKGGKIGLFG  
GAGVGKTVLIMELINNVAKAHGGYSVFAGVGERTREGNDLYHEMIESGVIN  
LKDATSKVALVYGQMNEPPGARARVALTGLTVAEYFRDQEGQDVLLFIDNI  
FRFTQAGSEVSALLGRIPSAVGYQPTLATDMGTMQERITTTKKSITSVQAI  
YVPADDLTDPAPATTF AHLDATTVLSRAIAELGIYPAVDPLDSTSRIMDPNIV  
GSEHYDVARGVQKILQDYKSLQDIIAILGMDELSEEDKLTVSRARKIQRFLS  
QPFQVAEVFTGHMGKLVPLKETIKGFQQILAGDYDHLPEQAFYMVGPPIEEA  
VAKADKLAEEHGS
```

### 4.2. TRATAMIENTO DE DATOS

El procesamiento de datos que se describe a continuación es realizado en el software MAThematical LABoratory (MATLAB) R2009a (Ver Anexo 1).

La secuencia de aminoácidos que codifica la estructura y función de la proteína es transformada en una serie numérica. Cada aminoácido es representado por el valor de interacción ión-electrón (EIIP), el cual describe los estados promedio de energía de todos los electrones de valencia en aminoácidos particulares (Tabla 4.2.) (Estos valores fueron tomados de Protein Sequence comparison based on the wavelet transform approach, Chafia Hejase de Trad, Qiang Fang and Irena Cosic, Protein Engineering Vol. No. 15 no 3 pp 193-203, 2002.)

Aminoácido	EIIP
Leu	0.0000
Ile	0.0000
Asn	0.0036
Gly	0.0050
Val	0.0057
Glu	0.0058
Pro	0.0198

His	0.0242
Lys	0,0371
Ala	0.373
Tyr	0.516
Trp	0.0548
Gln	0.0761
Met	0.0823
Ser	0.0829
Cys	0.0829
Thr	0.0941
Phe	0.0946
Arg	0.0959
Asp	0.1263

Tabla 4.2. Valores de potencial de interacción ión-electrón (EIIP) para aminoácidos (Cosic, *Macromolecular Bioactivity: Is it Resonant Interaction Between Macromolecules? - Theory and Applications*, 1994)

Después de obtener la secuencia numérica se procede a convertirla en valores a escala, para ello se utiliza una transformada matemática.

A diferencia de la Transformada de Fourier, la Transformada Wavelet se puede implementar sobre numerosas bases. Las diferentes categorías de wavelets (continuas, discretas, ortogonales, etc.) y los varios tipos de funciones wavelets dentro de cada categoría proveen una gran cantidad de opciones para analizar una señal de interés. Esto permite elegir la base de funciones cuya forma se aproxime mejor a las características de la señal que se desea representar o analizar.

En particular, para este trabajo se emplearon las *bases wavelet de Daubechies* para el procesamiento y análisis de las secuencias, las cuales tienen la propiedad de formar una base ortonormal y poseen soporte compacto. Debido a la condición de ortonormalidad, se asegura la independencia de la representación de la señal

en los diferentes niveles de descomposición, es decir, que no se genera información redundante de la señal, y así, se evita la aparición de información falsa. Además, las bases de Daubechies permiten calcular la Transformada Wavelet mediante un algoritmo menos complejo, con un bajo costo computacional y numéricamente estable (los cálculos realizados son confiables dentro de la precisión numérica del procesador), lo cual las hace eficientes frente a las bases no ortonormales. Por esta razón, son adecuadas para el análisis de señales con soporte finito y en particular para el análisis y procesamiento de secuencias proteínicas.

Las bases wavelets de Daubechies, incluye un rango de funciones que se extiende desde funciones altamente localizadas a funciones altamente suavizadas, entre las que se encuentran DAUB4, DAUB6, DAUB8, etc., hasta DAUB20, donde el número de Daubechies indica la cantidad de coeficientes distintos de cero. La decisión con respecto a cuál base de Daubechies es la más conveniente debe basarse en la forma de la señal a analizar, eligiendo la wavelet de Daubechies cuya forma más se le aproxime. Esto es, se debe utilizar la wavelet que provea la mejor descripción de la señal con el menor número de coeficientes no nulos.

Para la selección de la wavelet madre se tuvieron criterios como: Forma de onda, que debe tener una similitud con la señal, la cantidad de singularidades que se presenten en la señal. En el caso que se tengan pocas singularidades, la mejor opción es una wavelet madre con altos momentos de desvanecimientos para asegurar un análisis más exacto. En cambio si la cantidad de singularidades se incrementa por unidad de tiempo se debe hacer uso de una wavelet madre con menor número de desvanecimientos, para evitar un mayor coste computacional [22]. Por lo anteriormente expuesto se utilizó la base Daubechies DB12, la cual tiene un número promedio de desvanecimientos.

Estas Wavelet se desarrollaron para descomponer señales de tipo discreto. La idea consiste en tener una representación tiempo-escala de una señal discreta. La señal se pasa a través de filtros paso alto para analizar las componentes de alta frecuencia, y se pasa a través de filtros paso bajo para analizar las componentes de baja frecuencia, ver Figura 4.1. Estas operaciones cambian la resolución de la señal, y la escala se cambia mediante operaciones de interpolación y submuestreo. Los términos obtenidos de la transformación son conocidos como coeficientes de análisis y coeficientes de detalle [26].

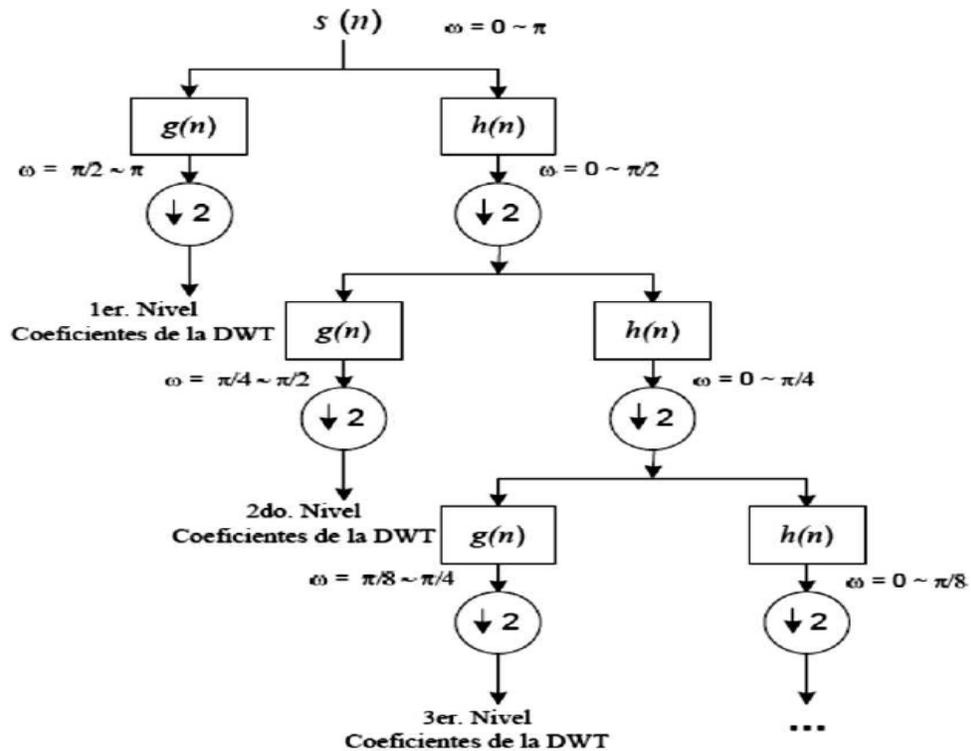


Figura 4.1. Algoritmo piramidal. La parte superior es la de análisis con (h) siendo el filtro paso bajo y (g) el filtro paso alto, mientras que la inferior es el detalle. (Fuente: Tovar S., Orozco N., Muñoz G., Murillo W., & Granada, Transformada Wavelet y Máquinas de Soporte Vectorial para la Identificación de Arritmias, Revista Universidad del Quindío (19): 104114, Armenia, 2009)

Se realizó el análisis wavelet para 8 escalas de análisis y 4 escalas para detalles.

Posteriormente se calcula la función de partición con los datos obtenidos de la transformada (coeficientes de análisis y detalle). Estos datos se elevan a la potencia  $q$ , donde  $q$  es el inverso de la temperatura. Elevar a la potencia  $q$  permite determinar si los datos son o no al azar, si no es aleatorio, se está calculando las correlaciones, se advierte si se presentan correlaciones de orden  $q$ .

Con lo anterior se tiene una función de la escala en análisis y detalle, en función de la temperatura.

Una vez se obtiene la función de partición, se extrae el logaritmo de esta función, que es equivalente a la energía libre-mecánico-estadístico interpretada cuánticamente en función de la densidad protónica.

Subsiguientemente se realiza la transformada de Legendre de la energía libre, con el que se determina la entropía o nivel de complejidad, conocido como multifractalidad.

Y por último, para las diversas secuencias de estudio (humano, bovino y rata) tanto para el análisis como el detalle, se obtienen las gráficas respectivas del espectro multifractal, que es dimensión fractal (de cada exponente) vs. exponente de fractalidad.

## 5. RESULTADOS Y DISCUSIÓN

Las proteínas son las secuencias de aminoácidos vistos como palabras representadas por ceros (0) y unos (1) no pueden ser tan largas porque se satura el canal o tan cortas porque se introducen errores, solo con una longitud apropiada se garantizara en la hebra proteica la redundancia, que nos permitirá ver un orden o sentido evidenciado en el código, con la incorporación de lo nuevo que ahora se ha hecho repetitivo, de esta manera se integra azar y diseño.

La noción de autodualidad ha sido expresada en lenguaje matemático como sinónimo de solución exacta de problemas en casos no lineales. La dualidad es la simetría más importante en la física moderna. Para (el código) nuestro caso la noción de transformación de la palabra código por diversas longitudes de secuencias es la forma auto-dual en la que un sistema no periódico, con modificación evolutiva, es decir mutaciones, se puede describir matemáticamente. La teoría de sistemas dinámicos no lineales (caos), tiene en la invariancia de escala o fractalidad o auto-similaridad la forma de descubrir un orden dentro del caos. Lo que se ha descrito es una convergencia entre teoría de códigos eficiente y la teoría de sistemas dinámicos no lineales o sea proteínas.

El ejemplo de ilustrar unas proteínas sencillas conocidas; la ATPasa, insulina, hemoglobina. A continuación al módulo se le aplica una proteína transmembranal como es un canal sensible de voltaje que puede ser de  $\text{Na}^+\text{K}^+$  (ATPasa). De este estudio obtenemos los patrones de auto-similaridad de manera directa sin ser necesario los levantamientos realizados.

Las proteínas desde el punto de vista bioquímico las podemos considerar como un gran aminoácido que expresa funcionalmente el mensaje contenido en el ADN. La

hipótesis de un gen, una proteína es una hipótesis de grano grueso que presenta fluctuaciones en la medida que la relación genoma-protonema no es estrictamente lineal, es en este sentido en el que abordamos a lo largo del trabajo la discusión o punto central del modelamiento de proteínas. Ampliando la idea de un gran aminoácido, lo que se quiere decir es que el rango de pH decide el extremo ácido al extremo básico pasando por una serie de valores intermedios que al observarlos en la secuencia de una proteína presentan un carácter aparentemente aleatorio. Nuestra hipótesis fundamental plantea que la proteína es una secuencia o serie de tiempo que expresa una sección de una gran distribución de bifurcaciones desde moléculas ancestrales que en una evolución a lo largo de millones de años dejan ver que el paisaje de energía (máximos y mínimos de energía) por el que discurren las reacciones químicas presenta una estructura compleja distinta a la concepción darwiniana del estilo Fisher donde convergían la idea de genes con el modelo seleccionista. Ese proceso de ascenso unidireccional es una aproximación unilateral y simplificada. Con la anterior anotación se quiere indicar que el estudio de la evolución de proteínas refleja los adelantos de la biología molecular reciente en el seguimiento de las mutaciones y múltiples bifurcaciones que en medio de la complejidad debe de existir una sistematicidad que en nuestro caso corresponde a un sistema dinámico no lineal.

Si las proteínas revelan patrones significativos que la evolución ha registrado en dominios y subdominios, independientemente del individuo o de la especificidad de cada individuo, es decir lo común a todas las especies.

El procedimiento consiste en integrar una secuencia asignada a los aminoácidos y un valor característico fisicoquímico que puede ser la hidrofobicidad o el potencial de interacción. De esta forma la secuencia se convierte en una serie. Esta serie se puede disponer en un arreglo matricial al cual le aplicamos el análisis Wavelet.

La transformada ondicular es una operación que transforma una secuencia en una nueva secuencia que va eliminando detalles superfluos, es decir promedia localmente de acuerdo a una regla que identifica variaciones importantes. Por ejemplo, si una secuencia repite un aminoácido de una estructura tipo  $\alpha$ , la transformada le asigna una variable única. Esta propiedad permite reconocer las variaciones o cambios de la secuencia, analíticamente significa que para diversas escalas la transformada expresa las variaciones de las secuencias dependiendo del zoom. Esta cantidad desde la termo-estadística se conoce como función de partición; es decir el significado de la distribución de energía microscópica que tiene una colección de aminoácidos para cada configuración posible que mediría la estabilidad del sistema. La transformada wavelet consigue una medida a partir de hacer una métrica definida por exponentes enteros distintos de dos que como sabemos dan para este caso la métrica pitagórica.

Los exponentes son una cantidad análoga a la temperatura, que en este caso ponderan la importancia de cada escala en la secuencia de aminoácidos. La proximidad de cada secuencia ponderada se comprime mediante la identificación de una función de distribución de exponentes que llamamos Energía libre informática, la cual mediante una transformación geométrica de Legendre nos conduce a la entropía.

La transformada ondicular de la secuencia de la proteína portadora de la información sobre la estructura tridimensional que representa la diversidad y la complejidad de la función. Los enfoques informáticos muestran que la longitud de la frase, la irreductibilidad del programa, son un indicador cuantitativo de esta. Este número corresponde al cálculo de la multifractalidad es decir la entropía del párrafo anterior. Las curvas auto-contenidas son una representación de la estructura fractal, en el caso de la ATPasa la estructura no es simple, son muchos fractales superpuestos, es lo múltiple de la secuencia. Un fractal puro se genera con una secuencia completamente aleatoria, un multifractal implica una

correlación asociada al carácter energético de los aminoácidos que deben expresar una función precisa. La energía funcional mínima corresponde a una función con invariancia de escala, que permite la estabilidad robustez del mensaje.

A continuación se presentan como resultados las gráficas del espectro multifractal  $f(\alpha)$  Vs.  $\alpha$  para las secuencias de Humano, Rata y Bovino, que se obtuvieron tanto para los coeficientes de análisis (aproximación o escalamiento), como para los de detalle (wavelet).

Las curvas de multifractalidad de la ATPasa dan cuenta de la diversidad genética de las proteínas, lo que se conoce por la bioinformática actual, no obstante el enfoque informático y energético justamente en estas proteínas, muestran la posibilidad de reunir el enfoque termodinámico y el informático.

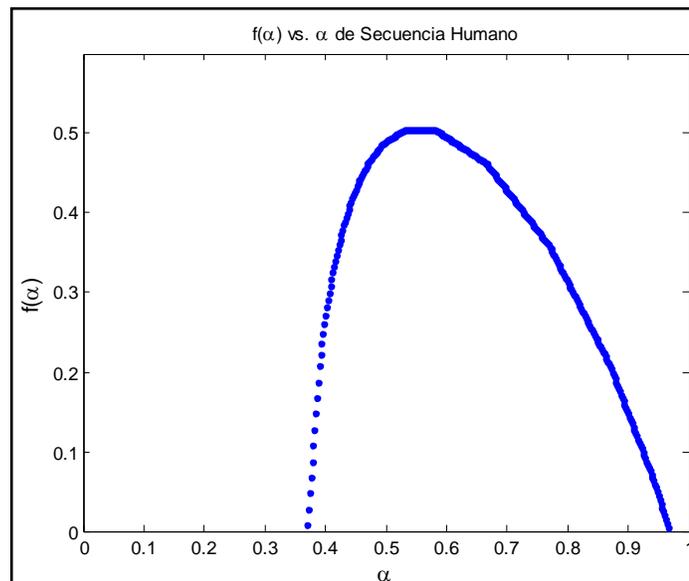


Figura 5.1. Espectro multifractal de análisis de la secuencia humano

Análisis humano: tiene fuerza multifractal (anchura) alrededor de 1,5 significando mayor complejidad, si amplificamos los detalles debe significar variabilidad con gran incidencia de mutaciones. El valor superior a 1 indica que la dimensión fractal

par a esos datos son altamente correlacionados en el caso de los análisis; indicando además que los aminoácidos en la gráfica secuencial se distribuyen de manera uniforme.

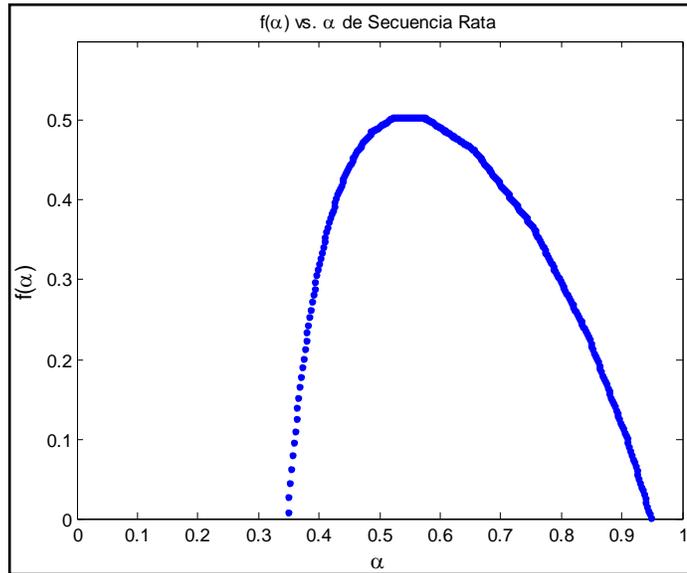


Figura 5.2. Espectro multifractal de análisis de la secuencia rata

Análisis rata: la rata esta en un punto crítico entre la novedad y la redundancia lo cual debe reflejarse al mostrar los detalles amplificados donde aparece el hecho de que la curva de análisis llegue a 1,5.

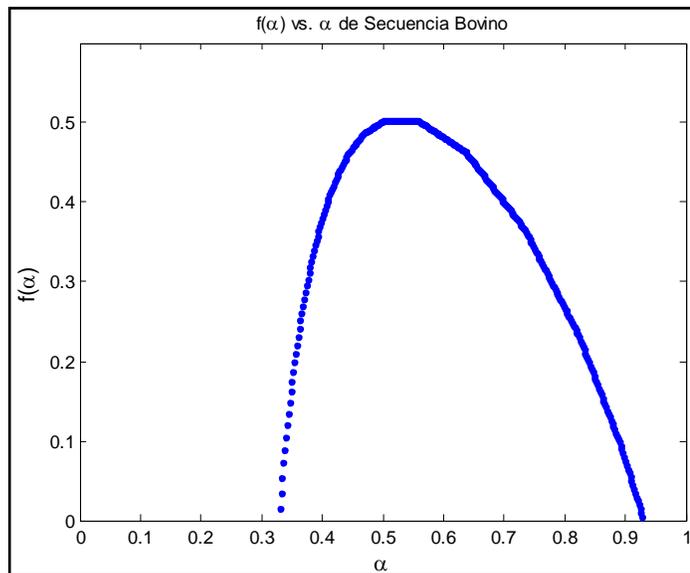


Figura 5.3. Espectro multifractal de análisis de la secuencia bovino

Análisis bovino: El ancho es el más pequeño, lo que quiere decir que es de menor complejidad. En los detalles que muestran una amplificación de la variedad genética se observa que hay un tramo mayor de 0,5 lo que implica que hay una alta correlación en términos informáticos, bastante redundancia, al existir menos complejidad, existirá menos información.

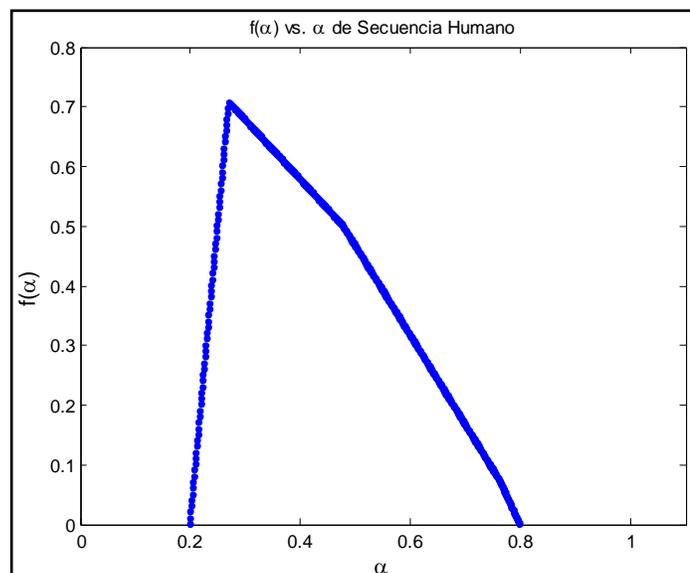


Figura 5.4. Espectro multifractal de detalle de la secuencia humano

Detalle humano: los detalles no tienen correlación fuerte estando por debajo de la distribución gaussiana de 0,5 resultando consistente con el ancho de las aproximaciones de la gráfica de análisis, significando que las mutaciones y variabilidad no tienen prácticamente correlación alguna o redundancia, que es un costo que se paga por el aumento de la complejidad ganada con el aumento de la información (debilidad).

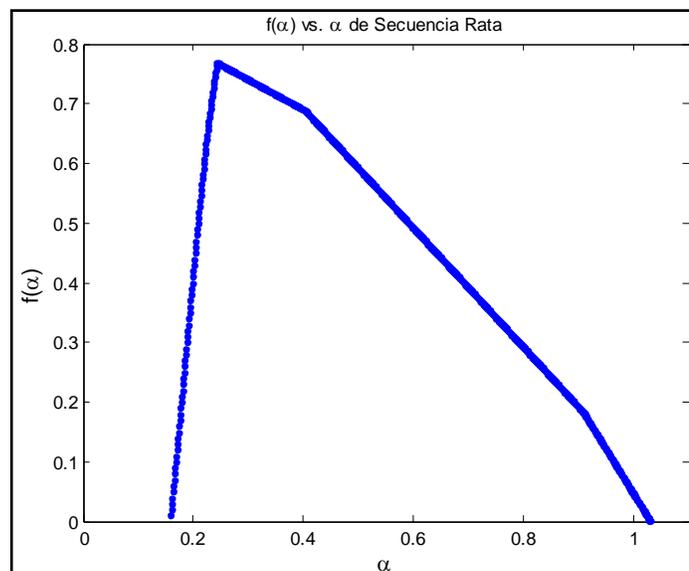


Figura 5.5. Espectro multifractal de detalle de la secuencia rata

Detalle rata: en los detalles la aleatoriedad tiene características de divergencia y convergencia de mutaciones variando en los rangos en los cuales las potencias que miden el grado de aleatoriedad solo convergen para exponentes negativos como se muestra en la figura del detalle rata y divergen para exponentes positivos, queriendo decir que variabilidad y mutaciones no guardan correlación de fase y se disparan para exponentes positivos dando como consecuencia una ambigüedad que significa el carácter crítico de las mutaciones; lo que puede hacer a las ratas

muy exitosas desde el punto de vista evolutivo, mirándolo con cuidado tratándose de un punto crítico que en nuestro caso corresponde a 1.

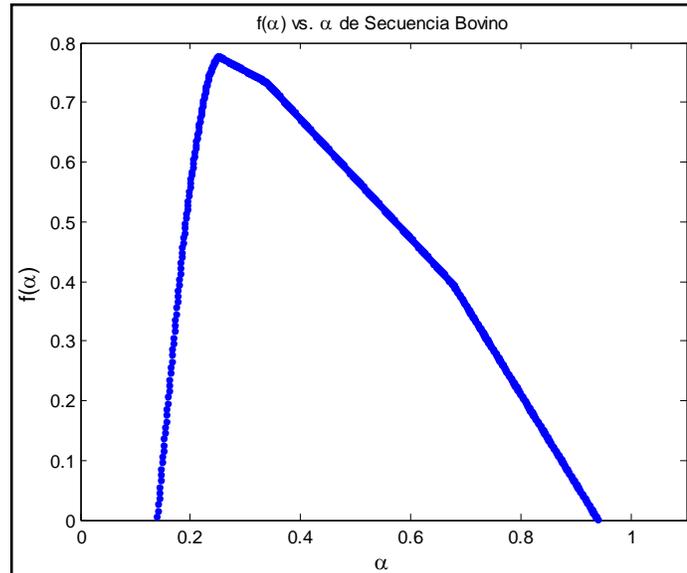


Figura 5.6. Espectro multifractal de detalle de la secuencia bovino

Detalle bovino: cuando se miran los detalles nos indican la novedad y variabilidad (las mutaciones). Observando ésta gráfica se deduce que tienen en buena parte significando que predomina la redundancia, la correlación contra la información, la cual es coherente con la gráfica de aproximación cuyo ancho no es muy grande.

Cuando en una gráfica multifractal, la fuerza multifractal (ancho) está alrededor de 0,5 esto significa que la distribución es gaussiana o aleatoria.

Existe un rango de complejidad que caracteriza la ATPasa de bovino, rata y humano. Con el rango de multifractalidad se determina una jerarquía biológica de las diferentes especies que tiene una explicación evolutiva dada la similaridad que las wavelet tienen con los árboles cladísticos.

Relación de los árboles y la invariancia de escala: La convexidad puede explicar la evolución o la fractalidad de la proteína ya que la estructura de la proteína analizada con wavelet es no divergente, siendo esta la propiedad de un sistema dinámico, es decir representado por ondas superpuestas, que se cierra sobre sí mismo; cuando eso sucede el sistema es determinístico, no aleatorio, llamándose al conjunto convergente un atractor y lo que hace el análisis wavelet es convertir este en convexidad.

Para una proteína significa que si hay un atractor entonces el sistema geoméricamente es autosimilar y autoreplicante, hecho que sucede con las polimerizaciones que se replican pero no exactamente, entendiendo que cada polimerización es un proceso bioquímico. Cuando hay iteraciones o ciclo de replicación una proteína repite porque químicamente solo sabe hacer esto, pero si un sistema interactúa tratando de mantener identidad, su forma es casi periódica, lo que resulta es bifurcaciones tipo cantor, los cantor son los conjuntos que representan una forma simple logística, la escalera del diablo es la generalización de esto, la secuencia simula algo como esta escalera y el resultado es convexo, por su estructura autoreplicante.

Lo que hace el algoritmo multifractal es dar la atenuación, mostrar que para cada escala la amplitud de las wavelet decae y lo hace con un exponente característico que es la dimensión fractal, que representada gráficamente nos dará convexidad y para la entropía que es la curva mas convexa los limites son las cotas para valores negativos y positivos del parámetro térmico y configuracional.

De la estructura autoreplicante se puede decir que genera convexidad, entendiendo que la distribución, que es entropía, se mide con la temperatura y no puede irse a la deriva, eso significa que la curva debe tener una serie de máximos que expresan convexidad, llegando a lo que se puede llamar equilibrio térmico informático. En el máximo esta la dimensión fractal representativa y las ramas

descendientes identifican entropías de sacudidas térmicas y de diversidad química que es acomodamiento de aminoácidos en una secuencia. Las wavelet por ser de convergencia óptima vuelven los datos mucho más convexos.

Así, de la convexidad se puede explicar que la rama izquierda es complejidad energética y será mayor en tanto más ancha sea esta rama en relación con el centro; la derecha es informática evolutiva y será mayor por la misma razón de anchura, existiendo un equilibrio entre lo energético y lo informático dado por el reacomodamiento de los aminoácidos; evolutivamente significa que el sistema siempre de manera convexa hace novedad temporal y repetición informática.

La dimensión fractal que se evidencia en el ancho de la gráfica, físicamente en la proteína representa o mide el grado de importancia de lo energético y lo configuracional, la complejidad es resultado de la articulación de los dos efectos y el ancho caracteriza la acotación de las dos contribuciones, la una es redundancia y la otra es novedad, su amplitud es la potencia multifractal, si fuera cero sería un fractal puro lo que no daría información de novedad, sería como un cristal puro redundante.

Podría decirse que los aminoácidos que se incluyen o reacomodan en el proceso evolutivo de la proteína generan entropía y por lo tanto caos pero en el tiempo se estabiliza el sistema por pérdida de energía generando permanencia evolutiva.

## 6. CONCLUSIONES

1. La Transformada Wavelet permitió descomponer las secuencias de estudio en un conjunto de señales restringidas para ciertas bandas de frecuencia, las cuales forman una base del espacio de funciones, con ciertas propiedades como ortogonalidad, tamaño, suavidad, etc. La descomposición se realiza a partir de funciones más complejas, en las cuales no se varía la frecuencia, sino su posición y su escala temporal.
2. Se comprobó que el análisis multiresolución permite analizar una señal en distintas resoluciones, al expandir una señal en una base de funciones y extraer características propias de cada resolución.
3. De los datos no nulos obtenidos de las transformadas wavelet se puede calcular el valor de la función de partición que es la pieza fundamental del análisis multifractal, que es finalmente lo que nos indica si existen índices bioinformáticos, es decir, si hay multifractalidad hay información.
4. La convexidad que caracteriza al espectro multifractal para las secuencias de Humano, Rata y Bovino indica una naturaleza multifractal de las proteínas de estudio.
5. Teniendo en cuenta la diferencia entre el mínimo y el máximo de  $\alpha$  (ancho), del espectro multifractal en las tres gráficas de "análisis", la secuencia de Humano muestra ser el de mayor información evolutiva y por lo tanto el de mayor variabilidad e información genética; la gráfica de la secuencia Bovino estaría registrando la menor información evolutiva o sea la menor variabilidad genética; y, la gráfica de la secuencia Rata se ubicaría en un punto intermedio

6. de la información evolutiva que correspondería a una variabilidad genética media. De esta manera se puede observar que la relación energía y configuración espacial expresan el grado de complejidad evolutiva desde la parte al todo, es decir, proteína- especie.
7. De acuerdo a los espectros multifractales de “análisis” obtenidos, se puede observar que la secuencia de Rata registra ligeramente una mayor diferencia entre el mínimo y máximo de alfa en la rama izquierda sugiriendo mayor flujo de energía metabólica, flujo que a su vez se puede traducir en entropía generadora de mejores posibilidades de cambios evolutivos, es decir, aumento de las posibilidades de variabilidad genética y por ende de mejores adaptaciones a diferentes medios. Para los casos de las secuencias Humano-Bovino se observan anchos similares, de lo que se puede decir que registran una estabilidad energética similar que puede manifestarse en una constante evolutiva y de lo cual se puede desprender la idea de un menor rango de adaptabilidad a cambios bruscos ambientales, en contraste con las posibilidades de la Rata.
8. Al observar los espectro multifractales obtenidos se puede decir que son muy similares en su forma convexa y a manera de hipótesis se dirá que el “análisis” como expresión de las frecuencias bajas registra bajos niveles de entropía, hecho que posibilita una gran estabilidad evolutiva en el tiempo-espacio, explicando el por qué la proteína en términos de complejidad energética e informática se comporta similarmente en las tres especies, mostrando la generalidad térmica en el conjunto proteico.
9. Los espectro multifractales que se obtuvieron de la función de escalamiento denominadas “detalle”, que son el resultado del análisis de frecuencias altas, y, a manera de hipótesis se puede inferir que registra altos niveles de entropía y por ende mayores posibilidades de cambios bruscos e inestabilidad

evolutiva, justificando así las diferencias marcadas entre sí. Las singularidades que nos muestra el “detalle” bien pueden ser expresión de las frecuencias altas.

10. El espectro multifractal (ancho) derecho respecto del punto intercepto del máximo en alfa rata es mayor, seguido de bovino; registrando el menor rango el humano, sugiriendo que el grado de estabilidad alcanzado es mayor; en contraste la rata y bovino registran mayores niveles de información fluctuante, en consecuencia, mayores posibilidades de mutación genética.
11. Respecto del ancho izquierdo, el promedio mayor lo registra el bovino, seguido de la rata, porcentajes que nos sugieren mayor consumo energético, mayores niveles de entropía. El humano registra el menor promedio sugiriendo con ello que hay mayor estabilidad energética es decir mayor eficiencia metabólica y quizá un mejor control de los cambios bruscos.

## RECOMENDACIONES

El presente trabajo se enmarca dentro de una línea de investigación interdisciplinar como es la bioinformática, debido a que existen diferentes sugerencias y perspectivas, como se enuncian a continuación:

1. Utilizar la secuencia completa de la ATPasa, es decir realizar el análisis con las unidades  $F_0$  y  $F_1$  y sus respectivas subunidades; puesto que se cuenta con una cantidad mayor de información que permitiría realizar una caracterización más amplia de la proteína.
2. Aplicar la transformada wavelet a la secuencia de ADN que generan la ATPasa, realizándose de forma sencilla ya que se aplica directamente, debido a que las bases pueden ser convertidas al sistema binario, permitiendo hacer un análisis evolutivo más de primeros principios.
3. Caracterización del canal de  $Na^+$ ,  $K^+$ ,  $Ca^{++}$  de la ATPasa en sus compuertas utilizando transformada wavelet discreta.

## REFERENCIAS

- [1] Abrahams, J. P., Leslie, A. G., Lutter, R., & Wal, J. E. (1994). *Structure at 2.8Å Resolution of F1 –ATPase from Bovine Heart Mitochondria*. *Nature* , 370, 621-628.
- [2] Alconada Rodríguez, A. (1993). *Homología Estructural y Funcional entre las Familias Proteicas de las Chaperoninas y de las Subunidades Alfa del Complejo F1-ATPasa*. Madrid: Universidad Complutense de Madrid.
- [3] Arenas D., G., & Sabogal P., S. M. (2006). *Presentación del proyecto: Una introducción a la geometría fractal a nivel universitario*. XXII Coloquio Distrital de Matemáticas y Estadística. Bogotá: UIS - UNAL.
- [4] Arguedas Oses, D. (2010). *Diseño e implementación de una arquitectura para la transformada discreta wavelet empleando Verilog*. Ciudad Universitaria Rodrigo Facio: Universidad de Costa Rica.
- [5] Broglia, R. A. (Junio de 2002). *De los núcleos atómicos a las proteínas*. *Investigación y Ciencia* , 54-60.
- [6] Chui, C. K. (1992). *An Introduction to Wavelets*. Academic press.
- [7] Colmenares R., R. A. (2011). *Análisis Fractal Basado en la Transformada de Ondícula para el Reconocimiento de Facies*, Campo Lama, Lago de Maracaibo. Sartenejas: Universidad Simón Bolívar, Ingeniería GeoFísica.
- [8] Cosic, I. (1994). *Macromolecular Bioactivity: Is it Resonant Interaction Between Macromolecules? - Theory and Applications*. *IEEE Trans. on Biomedical Engineering* , 41, 1101-1114.

- [9] Dana Mackenzie, I. D. (2001). *Wavelets: ver el bosque y los arboles*.
- [10] Daubechies, I. (1992). *Ten Lectures on Wavelets*. The Society for Industrial and Applied.
- [11] Doolittle, R. F., & Bork, P. (Diciembre de 1993). *Módulos Móviles en la Evolución de las Proteínas*. *Investigación y Ciencia* , 22 - 29.
- [12] García Marín, A. P. (2007). *Análisis multifractal de series de datos pluviométricos en Andalucía*. Tesis Doctoral, Universidad de Cordoba . Córdoba: Universidad de Cordoba- Departamento de Ingeniería Rural.
- [13] Harper. (2007). *Bioquímica* (Vol. 17 Edición). España: El Manual Moderno S.A de C.V.
- [14] Hinestroza, D. (2009). *Cursillo Introducción a las Wavelets*. Cali: Universidad del Valle.
- [15] Jiménez-Hornero, F., Giráldez, J., & Laguna, A. (2007). *Análisis Multifractal de la Velocidad de Flujo Simulada con un Modelo de Malla en Medios Porosos Bidimensionales Idealizados*. *Estudios de la Zona No Saturada del Suelo* , III.
- [16] Lugo, J. O., Sampallo, G. M., & Gonzáles Thomas, A. *Detección de Zonas de Diferentes Texturas Usando Análisis de Multiresolución*. FACENA UNNE.
- [17] Mallat, S. (2009). *A Wavelet Tour of Signal Processing*. Burlington: Academic Press, 3rd edition.
- [18] Mandelbrot, B. (1982). *La Geometría Fractal de la Naturaleza*. Barcelona: Tusquets Editores.

- [19] Polikar, R. (1996). *The Wavelet Tutorial*. U.S.A: Universidad Estatal de Iowa. Centro de Computación.
- [20] Posadas, A., Giménez, D., & Quiroz, R. (2002). *Análisis Multifractal de la Variabilidad Espacial de la Conductividad Hidráulica en un Suelo Estratificado*. *Revista de Investigación de Física* , 36-43.
- [21] Quezada Len, A. (2006). *Fractales y Opinión Pública: una aplicación del exponente de Hurst al estudio de la dinámica de la identificación ideológica*. Barcelona: Tesis de Doctorado - Universidad de Barcelona, Facultad de Psicología, Influencia Social: relaciones, procesos y efectos 1999-2000.
- [22] Sepulveda, A., & Castellanos, G. (2004). *Estimación de la Frecuencia Fundamental de Señales de Voz Usando Transformada Wavelet*. *Sienta et Technica* (24), 7-11.
- [23] Sheng, Y. (1996). *The Transforms and Applications Handbook*. CRC Press.
- [24] Strang, G., & Truong, N. (1996). *Wavelet and Filter Banks*. Wellesley – Cambridge Press.
- [25] *The Wavelet Tutorial*. Dept. of Electrical and Computer Engineering. Rowan University. 1995.
- [26] Tovar S., D. A., Orozco N., A. J., Muñoz G., P. A., Murillo W., H., & Granada, J. A. (2009). *Transformada Wavelet y Máquinas de Soporte Vectorial para la Identificación de Arritmias*. *Revista Universidad del Quindío* (19), 104-114.

- [27] Vampa, V. (2011). *Desarrollo de Herramientas basadas en la Transformada Wavelet para su Aplicación en la Resolución Numérica de Ecuaciones Diferenciales*. (T. Doctoral, Ed.) La Plata, Argentina: Universidad Nacional de la Plata.
- [28] Walnut, D. (2001). *An Introduction to Wavelet Analysis*. Birkhauser.
- [29] Walnut, D. (2002). *An Introduction to Wavelet Analysis, Applied and Numerical Harmonic Analysis Series*. Boston: Birkhäuser Eds.
- [30] Web, V. (s.f.). [http://www.fractovia.org/art/es/what\\_es3.shtml](http://www.fractovia.org/art/es/what_es3.shtml).
- [31] Wikipedia. (s.f.). Recuperado el Junio de 2012, de <http://es.wikipedia.org/wiki/Fractal>.
- [32] Mathews, C.K. and Ahern, K.G. and Van Holde, K.E. (2002). *Bioquímica*. Edición. Madrid: Addison Wesley Pearson.
- [33] Solé, R., Manrubia, Susanna. (1993). *Orden y caos en sistemas complejos*. Ediciones UPC., Cataluña.
- [34] Jaffard, S. (1997). *Multifractal formalism for functions parts I and II*. SIAM J. of Mathematical Analysis, 28 (2): 944-998.
- [35] Mallat, S. (1999). *A Wavelet Tour of Signal Processing*. Academic Press. USA.

## GLOSARIO

**ATPasa:** enzima encargada de la fosforilación del Adenosin Difosfato (ADP), creando la molécula energética de ATP.

**Enzimas:** proteínas que cumplen una función catalizadora de procesos bioquímicos.

**Medida de la complejidad:** cantidad de información necesaria para describir un sistema. Cuanto más complejo es un ser vivo más información hay contenida en el.

**Multifractalidad:** se considera una familia uniparamétrica de dimensiones fractales diferentes. La multifractalidad hace referencia a tener múltiples fractales en un sistema o señal.

**Transformada wavelet:** es un tipo especial de transformada de Fourier la cual representa una señal de versiones trasladadas y dilatadas de una onda finita, denominada wavelet madre.

**Proteómica:** estudio a gran escala de las proteínas, especialmente de la estructura y función. La palabra proteoma es la fusión de proteína y genoma. La descripción del proteoma permite tener una imagen dinámica de todas las proteínas expresadas.

**Pseudopotencial:** falso potencial asignado porque es la información obtenida a partir de los aminoácidos apantallado en entorno acuoso o cristalino.

## ANEXOS

### ANEXO 1

Se anexa el programa desarrollado en Matlab que fue utilizado en el presente trabajo.

```
%Analiza secuencias con wavelets asignando valores a cada aminoácido de
%acuerdo al potencial de interacción
```

```
clear X1
clear Y1
```

```
[nombre_aminoacidos,camino_aminoacidos] = uigetfile('aminoacidos.txt','Seleccione la base de
aminoacidos-aminoacidos');
[nombre_secuencia2,camino_secuencia2] = uigetfile('ATPAmino.txt','Seleccione el secuencia con
la palabra a leer: Humano, Bovino o Rata');
camino_y_nombre_aminoacidos = strcat(camino_aminoacidos,nombre_aminoacidos);
camino_y_nombre_secuencia2 = strcat(camino_secuencia2,nombre_secuencia2);
[X1_vector] = textread(camino_y_nombre_aminoacidos,'%s','whitespace','\n'); %y11_vector es un
cell array of strings
[Y1_vector] = textread(camino_y_nombre_secuencia2,'%s','whitespace','\n'); %y11_vector es un
cell array of strings
X1= char(X1_vector);% X1 es un character array
Y1= char(Y1_vector);% Y1 es un character array
posicion_en_X1=[];
posicion_en_Y1=[];
```

```
X1; %20 aminoacidos
Y1; %Secuencia seleccionada
[e1,d1]=size(X1);%calcula el numero d1 de datos en secuencia1.txt
d1; %tamaño del vector de aminoaciids
[e2,d2]=size(Y1);%calcula el numero d2 de datos en secuencia2.txt
d2; %tamaño del vector de la secuencia seleccionada
```

```
for inx=1:d1
    X(inx)=0;
end
```

```
X(1)=0.0373;%A
X(2)=0.0829;%C
X(3)=0.1263;%D
X(4)=0.0058;%E
```

```

X(5)=0.0946;%F
X(6)=0.0050;%G
X(7)=0.0242;%H
X(8)=0;%I
X(9)=0.0371;%K
X(10)=0;%L
X(11)=0.0823;%M
X(12)=0.036;%N
X(13)=0.0198;%P
X(14)=0.0761;%Q
X(15)=0.0959;%R
X(16)=0.0829;%S
X(17)=0.0941;%T
X(18)=0.057;%V
X(19)=0.0548;%W
X(20)=0.0516;%Y

```

```

for ini=1:d2
    Y(ini)=0;
    XX(ini)=ini;
end

```

```

for ni=1:d1 % indices para los aminoácidos
    for mi=1:d2 % indices para la secuencia
        c=eq(Y1(mi),X1(ni));
        if c
            Y(mi)=X(ni);
        end
    end
end
end

```

```

Y;
ild2=int16(log2(d2));
ld2=log2(d2);
c1=gt(ild2,ld2);
i=0;

```

```

if c1
    for ni=0:d1 % indices para los aminoácidos
        ihelp1=d2+1;
        ihelp2=2^(ild2);
        for mi=ihelp1:ihelp2 % indices para la secuencia
            i=1+i;
            Y(mi)=Y(i);
        end
    end
    d2=ihelp2;
end
c2=lt(ild2,ld2);

```

```

if c2
    for ni=0:d1 % indices para los aminoácidos
        ihelp1=d2+1;

```

```

ihelp3=2^ild2+1;
for mi=ihelp1:ihelp3 % indices para la secuencia
    i=1+i;
    Y(mi)=Y(i);
end
end
d2=ihelp3;
end

clear ii
clear iiq
clear Z
clear ZZ
clear Zp
clear ZZp
iq=0;
clear iiq

for q=-10:0.5:10
    ca1=Y;
    iq=1+iq;
    iiq(iq)=q;
    imax=5;
    imin=2;
    di=imax-imin+1;
    i=0;
    for ih=imin:imax
        i=1+i;
        ii(i)=ih;
        [ca1,cd1] = dwt(ca1,'db4');
        [filascd1,columnascd1]=size(cd1);
        helpZ1=0;
        for zi=1:columnascd1
            helpZ1=helpZ1+(abs(cd1(1,zi)/columnascd1))^q;
        end
        ZZ(iq,i)=(helpZ1);
        ZZp(iq,i)=log2(ZZ(iq,i));
    end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% CALCULO DE LA PENDIENTE DE %
% tao = Log2[Z(q)]/Log2[1/a] %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clear tao
for q1=1:iq
    n=0;
    clear Z11
    for k1=1:di
        n=n+1;
        Z11(n)=ZZp(q1,k1);
    end
    Z11;
end

```

```

p = polyfit(ii,Z11,1);
tao(q1)=p(1);      %p(1)PENDIENTES y p(2) CORTES
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%D(q)=q*(alfa+0.5)-tao(q): Do(q)%
%PARA EL CALCULO DE D(alfa)  %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clear Do
clear D
clear alfa1
alfa22=0;

for alfa=0:1:1200
    alfa1=0.001*alfa;
    for q=1:iq
        Do(q)=iiq(q)*(alfa1+0.5)-tao(q);
    end
    if c
        alfa22=alfa22+1;
        alfa11(alfa22)=alfa1;
        D(alfa22)=min(Do);
    end
end

figure(4),plot(alfa11,D,'.','color','black')
xlabel('\alpha','FontSize',12)
ylabel('D(\alpha)','FontSize',12)
title('D(\alpha) vs. \alpha de Secuencias Bovino-Humano-Rata')

```