

UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA QUÍMICA Y BIOTECNOLOGÍA

**PREDICCIÓN DEL COEFICIENTE DE PARTICIÓN DE PROTEÍNAS EN  
SISTEMAS DE DOS FASES ACUOSAS A TRAVÉS DE LA  
CARACTERIZACIÓN BIOINFORMÁTICA DE SU SUPERFICIE**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL EN  
BIOTECNOLOGÍA

MARÍA PAZ CORTÉS BURGOS

PROFESOR GUÍA:  
CRISTIAN SALGADO HERRERA

MIEMBROS DE LA COMISIÓN:  
BARBARA ANDREWS FARROW  
MARÍA ELENA LIENQUEO CONTRERAS

SANTIAGO DE CHILE  
ENERO 2008

## RESUMEN

El objetivo general del presente trabajo de título es determinar el efecto de la distribución de la hidrofobicidad en la superficie de las proteínas, en su comportamiento en sistemas de dos fases acuosas (ATPS), mediante el uso de modelos matemáticos para la descripción de esta distribución.

Para hacer esta descripción se utilizaron dos tipos de modelos matemáticos. El primer tipo de modelos corresponde a estadísticas obtenidas a partir de mediciones locales de la hidrofobicidad en la superficie de las proteínas. Estas mediciones se hicieron definiendo vecindarios alrededor de cada aminoácido expuesto en la superficie. El segundo tipo de modelos se define a partir del vector de desbalance hidrofóbico, correspondiente a la desviación del centro geométrico de la superficie de las proteínas por efecto de la hidrofobicidad.

Para estudiar la calidad de estos modelos, en términos de la predicción del comportamiento de proteínas en ATPS, se correlacionaron los valores obtenidos en un set de 8 proteínas con sus coeficientes de partición, en 12 sistemas distintos. Estos sistemas corresponden a 3 sistemas del tipo polímero/sal: PEG4/Fosfato, PEG/Sulfato y PEG/Citrato y un sistema PEG/Dextrano. Cada uno de ellos con concentraciones nula, baja y alta de NaCl. La predicción de  $K$  se midió utilizando el error cuadrático medio de Jack-Knife. Para estudiar la calidad de estas predicciones, los resultados generados se compararon con los obtenidos utilizando la hidrofobicidad superficial promedio y con los correspondientes a un modelo desarrollado previamente por Andrews *et al.*, en el que se correlacionó la hidrofobicidad de las proteínas, medida experimentalmente, con el coeficiente  $K$ .

Las correlaciones obtenidas utilizando los distintos modelos fueron de muy buena calidad. Los mejores resultados se obtuvieron con los modelos estadísticos donde en promedio la correlación fue de 0,88. Además las predicciones, en su mayoría, fueron superiores a las asociadas a los modelos de comparación.

Entre todos los modelos utilizados se seleccionó el correspondiente al descriptor ASPavg como el más apto para la predicción de  $K$ , en términos de la influencia que tiene la hidrofobicidad de las proteínas sobre este parámetro. ASPavg es un promedio de la hidrofobicidad superficial local de las proteínas. La selección se hizo debido a la preferencia del modelo por escalas de hidrofobicidad y al hecho que la resolución hidrofóbica mantiene características reportadas anteriormente.

Se concluye que, mediante la metodología seguida en este trabajo, es posible obtener una descripción de la distribución de la hidrofobicidad en la superficie de las proteínas, que sea de utilidad para modelar su comportamiento en ATPS.

# ÍNDICE

<b>1.</b>	<b>INTRODUCCIÓN .....</b>	<b>1</b>
<b>2.</b>	<b>ANTECEDENTES GENERALES .....</b>	<b>2</b>
2.1	SISTEMAS DE DOS FASES ACUOSAS .....	2
2.2	MODELOS PARA LA PARTICIÓN DE BIOMOLÉCULAS EN ATPS .....	5
<b>3.</b>	<b>OBJETIVOS.....</b>	<b>9</b>
<b>4.</b>	<b>METODOLOGÍA .....</b>	<b>10</b>
4.1	PROTEÍNAS Y SISTEMAS CONSIDERADOS EN EL ESTUDIO .....	10
4.2	CUANTIFICACIÓN DE LA DISTRIBUCIÓN DE LA HIDROFOBICIDAD EN LA SUPERFICIE DE LAS PROTEÍNAS.....	10
4.2.1	<i>Descriptores estadísticos.....</i>	<i>12</i>
4.2.2	<i>Descriptores basados en el vector de desbalance hidrofóbico.....</i>	<i>13</i>
4.3	PREDICCIÓN DEL COEFICIENTE DE PARTICIÓN DE PROTEÍNAS EN ATPS .....	14
<b>5</b>	<b>RESULTADOS Y DISCUSIÓN .....</b>	<b>15</b>
5.1	DESCRIPCIÓN DE LA DISTRIBUCIÓN DE LA HIDROFOBICIDAD EN LA SUPERFICIE DE LAS PROTEÍNAS 15	
5.2	DIFERENCIAS AL CONSIDERAR LA POSICIÓN DE LOS AMINOÁCIDOS EN LAS PROTEÍNAS COMO EL CENTRO GEOMÉTRICO DE SUS ÁTOMOS Y COMO LA UBICACIÓN DE SU CARBONO- $\beta$ .....	20
5.3	ANÁLISIS DE LOS DESCRIPTORES ESTADÍSTICOS .....	21
5.3.1	<i>Variaciones en los descriptores estadísticos en el set de proteínas.....</i>	<i>22</i>
5.3.2	<i>Efecto de las distintas escalas de propiedades utilizadas.....</i>	<i>22</i>
5.3.3	<i>Efecto de la variación de los radios de vecindarios asociados a los ASP locales.....</i>	<i>23</i>
5.3.4	<i>Correlación entre descriptores.....</i>	<i>26</i>
5.4	ANÁLISIS DE LOS DESCRIPTORES BASADOS EN EL DESBALANCE HIDROFÓBICO .....	27
5.4.1	<i>Variaciones en el conjunto de proteínas considerado.....</i>	<i>27</i>
5.4.2	<i>Efecto del uso de distintas escalas de propiedades .....</i>	<i>28</i>
5.4.3	<i>Correlación entre los descriptores .....</i>	<i>30</i>
5.5	MODELOS DE PREDICCIÓN DEL COEFICIENTE DE PARTICIÓN.....	30
5.5.1	<i>Calidad de los modelos obtenidos en la predicción de K.....</i>	<i>38</i>
5.5.2	<i>Resolución hidrofóbica e hidrofobicidad intrínseca de los sistemas: R y logPo.....</i>	<i>39</i>
5.5.3	<i>Radios seleccionados por los descriptores estadísticos .....</i>	<i>41</i>
5.5.4	<i>Predicción de K utilizando distintas escalas de propiedades.....</i>	<i>42</i>
5.5.5	<i>Comparación de los descriptores .....</i>	<i>50</i>
<b>6</b>	<b>CONCLUSIONES .....</b>	<b>53</b>
<b>7</b>	<b>BIBLIOGRAFÍA .....</b>	<b>55</b>

## ANEXOS

## NOMENCLATURA

$\gamma$	: Coeficiente de actividad
$\phi$	: Fracción volumétrica
$\tilde{\phi}$	: Fracción volumétrica modificada por efecto de la hidratación de las moléculas de PEG
$\chi_{ij}$	: Parámetro de interacción entre los compuestos i y j (modelo 3)
$\varphi$	: Propiedad del aminoácido i
$\delta E$	: Diferencia de energía electrostática de la proteína entre las fases del sistema
$\Delta S$	: Entropía
$\Delta\varphi$	: Diferencia del potencial eléctrico entre las fases
$\Lambda$	: Tasa entre los coeficientes de actividad de una especie neutra y cargada (modelo 7)
$\text{Å}$	: Angstrom
$a_x$	: Actividad de la especie x (modelo 7)
ASA	: Área superficial accesible
ASP	: Propiedad de superficie promedio
$ASP_{90^\circ, P}$	: ASP en el hemisferio hidrofóbico calculado con respecto a su superficie
$ASP_{90^\circ, T}$	: ASP en el hemisferio hidrofóbico calculado con respecto a la superficie total de la proteína
ASPavg	: Promedio de los ASP locales
ASPmax	: Máximo de los ASP locales
ASPmin	: Mínimo de los ASP locales
ASPstd	: Desviación estándar del promedio de los ASP locales
ATPS	: Sistemas de dos fases acuosas
$B_{Pi}$	: Coeficiente asociado a la interacción entre la proteína y el polímero i (modelo 1)
CB	: Carbono $\beta$ de los aminoácidos
CG	: Centro geométrico de los átomos en un aminoácido
$C_P$	: Concentración proteína
$d_i$	: N° de segmentos (monómeros) del polímero i (modelo 1)
Descriptor <sub>CG</sub>	: Valor de un descriptor calculado considerando la posición de los aminoácidos como el centro geométrico de sus átomos
HH	: Hemisferio hidrofóbico
HI	: Magnitud del vector de desbalance hidrofóbico
HPLC	: Cromatografía líquida de alto desempeño
k	: Especie iónica de referencia (modelo 2)
K	: Coeficiente de partición de proteínas en ATPS
$K_0$	: Coeficiente de partición en ausencia de $\Delta\varphi$
$K_{af}$	: Contribución del efecto de la afinidad al coeficiente K (ecuación 2.2)
$K_{el}$	: Contribución del efecto de fuerzas electrostáticas al coeficiente K (ecuación 2.2)
$K_{hidro}$	: Contribución del efecto de la hidrofobicidad al coeficiente K (ecuación 2.2)

$K_{sol}$	: Contribución del efecto de la solubilidad al coeficiente K (ecuación 2.2)
$K_{tamaño}$	: Contribución del efecto del tamaño de las partículas al coeficiente K (ecuación 2.2)
$K_{bl}$	: Constante de equilibrio para solutos negativos
$K_{cl}$	: Constante de equilibrio para solutos positivos
$K_i$	: Valor experimental del coeficiente de partición de la proteína i
$K_i'$	: Valor calculado del coeficiente de partición de la proteína i considerando un ajuste de los datos asociados a las proteínas restantes en el set
logPo	: Hidrofobicidad intrínseca de un ATPS
L	: Factor de esfericidad de la proteína
M	: Molalidad
$M_p$	: Peso molecular de la proteína
$MSE_{JK}$	: Error cuadrático medio de Jack-Knife
$n_i$	: Parámetro asociado al peso molecular de los polímeros (modelo 1)
Nr(k)	: Vecindario de radio alrededor del aminoácido k
P	: Hidrofobicidad
PDB	: Base de datos "Protein Data Bank"
Pearson	: Coeficiente de correlación
PEG	: Polietilenglicol
PEG4000	: Polietilenglicol de peso molecular 4000 Da
R	: Resolución hidrofóbica
r(i)	: Vector de posición del aminoácido i en una proteína
$r_{CG}$	: Centro geométrico de la superficie de proteínas
$r_H$	: Centro geométrico de la superficie de proteínas modificado por efecto de la hidrofobicidad
$r_{HI}$	: Vector de desbalance hidrofóbico
RP-HPLC	: Cromatografía líquida de alto desempeño de fase reversa
S	: Superficie de una proteína
V	: Volumen por mol
$W_1$	: Fracción másica del agua
$W_{2,3}$	: Fracciones másicas de los polímeros en el sistema
z	: Carga neta

# 1. INTRODUCCIÓN

Los procesos de separación y purificación de proteínas son de suma importancia para la obtención de productos libres de contaminantes. El uso de procesos eficientes y de fácil escalamiento es fundamental para la producción industrial de estas moléculas. Existen distintos métodos de purificación de proteínas siendo los procesos cromatográficos los más difundidos. Otra alternativa menos utilizada es el uso de sistemas de dos fases acuosas.

Los sistemas de dos fases acuosas o ATPS (*aqueous two phase systems*) están formados por dos fases inmiscibles originadas por la mezcla de dos polímeros o de un polímero y una sal en una solución con alto contenido de agua. Al incorporar proteínas en estos sistemas, estas moléculas mostrarán una preferencia por una u otra fase permitiendo así su separación.

La principal ventaja del uso de estos sistemas para la separación de proteínas es que en ellos se genera un ambiente inocuo para biomoléculas debido a su alto contenido de agua y baja tensión interfacial. Sin embargo, su uso en procesos de separación ha estado restringido principalmente a la escala de laboratorio, y no se ha difundido a nivel industrial debido principalmente a la falta de conocimiento sobre el mecanismo que gobierna la partición de las proteínas en el sistema [1]. Por esta razón, resulta de utilidad el desarrollo de modelos que describan el comportamiento de estas moléculas en ATPS.

El comportamiento de proteínas en estos sistemas depende de muchos factores asociados a las características de los ATPS como su pH, temperatura y el tipo y concentración de sales y polímeros utilizados. Además, propiedades de las mismas proteínas también tienen un rol en este fenómeno. Entre ellas se encuentran su peso molecular, carga e hidrofobicidad, siendo esta última una de las más relevantes.

Las proteínas se relacionan con su medio principalmente a través de su superficie. Así, una forma de cuantificar las propiedades que rigen esta interacción es mediante la consideración del aporte de los aminoácidos expuestos en la superficie.

En este trabajo se pretende estudiar la relevancia de la distribución de la hidrofobicidad en la superficie de las proteínas en su coeficiente de partición en ATPS. Para lograr este objetivo se utilizarán distintos modelos matemáticos para describir esta distribución de manera cuantitativa. Cada uno de estos modelos considera el área expuesta de los aminoácidos en la superficie de las proteínas y su hidrofobicidad. Una vez hecho los cálculos para distintas proteínas, se estudiará el nivel de correlación entre los valores obtenidos y los coeficientes de partición en ATPS así como también la capacidad predictiva de los modelos generados.

## 2. ANTECEDENTES GENERALES

### 2.1 *Sistemas de Dos Fases Acuosa*s

Los sistemas de dos fases acuosa se forman al mezclar dos polímeros inmiscibles o un polímero y una sal en alta presencia de agua. Al alcanzar el equilibrio el sistema se separa en dos fases, cada una enriquecida en uno de los componentes. Los polímeros que se utilizan normalmente son polietilenglicol (PEG) y dextrano y en el caso de las sales, se ocupan fosfato de potasio y sulfato de sodio, entre otras.

Los ATPS se pueden describir mediante un diagrama de fases como el que se muestra en la Figura 2.1. En este diagrama, la curva binodial marca el límite sobre el cual se forman las dos fases en los sistemas. De esta forma, sobre la curva existen dos fases y bajo ella sólo hay una.

La recta AC es la llamada Tie-line, la cual permite conocer la composición de las fases en equilibrio. Todos los sistemas que tengan una composición global dentro de esta Tie-line, tal como el punto B, en el equilibrio se separan en dos fases descritas por los puntos A y C. Estos puntos corresponden a las composiciones de la fase inferior y superior respectivamente.

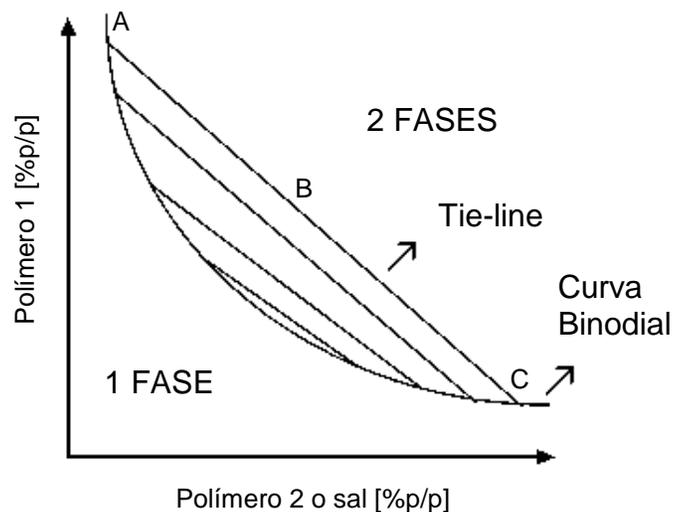


Figura 2.1: Diagrama de fases para un sistema de dos fases acuosa

Al incorporar una partícula en estos sistemas, como por ejemplo una proteína, ésta se distribuirá entre las dos fases. Esta partición se describe mediante el coeficiente de

partición  $K$ , definido como el cociente entre las concentraciones de la partícula en la fase superior e inferior del sistema:

$$K = \frac{C_{\text{sup}}}{C_{\text{inf}}} \quad (2.1)$$

La partición de biomoléculas en ATPS depende de distintos factores, los cuales a su vez, interactúan entre sí, lo que hace dificultoso el modelamiento de este fenómeno. Estas variables se dividen en factores asociados al sistema y los correspondientes a la naturaleza de la partícula en partición.

Entre las principales propiedades de los ATPS que afectan la partición de partículas están la temperatura y pH del sistema, los tipos de sales y polímeros utilizados y su concentración y el peso molecular de los polímeros involucrados.

El efecto de la temperatura del sistema en la partición de partículas no es claro. Se ha observado en sistemas formados por PEG y fosfato de potasio que un aumento de temperatura conduce a una alteración en la composición de las fases, llevando a un incremento de la concentración de PEG y sal en las fases superior e inferior respectivamente y a la obtención de un coeficiente de partición más elevado de la  $\beta$ -glucosidasa [2]. Este fenómeno se atribuye a una disminución de las moléculas de agua libres para la disolución de PEG y de moléculas menos polares que la sal en la fase inferior provocada por una intensificación del efecto de salting-out.

Por otro lado, en un estudio de la partición de mioglobina en sistemas formados por PEG y dextrano, una alteración de la temperatura de 22° C a 50° C no provocó un cambio significativo en la composición de las fases del sistema pero si llevó a que la distribución de esta proteína entre ellas fuese más uniforme a 50° C [3].

El pH del sistema también afecta la partición de biomoléculas. Se ha observado que una diferencia de pH positiva entre las fases superior e inferior de un sistema acuoso lleva a que moléculas con carga positiva prefieran la fase inferior y las con carga negativas la superior [4, 5]. De esta manera, producto de la interacción iónica entre el sistema y los solutos cargados, moléculas negativas tendrán un  $K$  superior que su equivalente neutro, mientras que las con carga positiva tendrán un coeficiente de partición menor.

Además, cambios en el pH del sistema provocan cambios en la carga neta de las proteínas. Un pH superior al punto isoeléctrico de una proteína provocará que su carga neta sea negativa mientras que uno inferior hará que sea positiva [2]. Así, es posible provocar cambios en la afinidad de las proteínas por las fases del sistema mediante el manejo del pH.

Con respecto a la concentración de polímeros utilizados para la formación del sistema, en general, se tiene que mientras mayor es esta concentración, la diferencia en la composición de las fases se hará más marcada, provocando que la preferencia por una u otra fase por parte de los solutos en partición sea mayor [6].

Otra característica de los polímeros que influye en la partición de biomoléculas en ATPS es su peso molecular. Cambios en este parámetro provoca alteraciones en la composición de las fases del sistema y en la interacción polímero-biomolécula. En sistemas de PEG/dextrano, se ha observado que el coeficiente de partición de proteínas que prefieren la fase inferior rica en dextrano se incrementa con el aumento del peso molecular del dextrano y disminuye con el del polietilenglicol [7]. Esto se explica debido a que a medida que aumenta el peso molecular de un polímero en solución la red que forman sus cadenas se hace más densa haciendo más difícil la entrada de otras moléculas en ella [8].

Otro factor que puede afectar la partición de biomoléculas en ATPS es el uso de polímeros con polidispersividad. Esta propiedad da cuenta de la distribución de un polímero en moléculas de distinto peso molecular. Esta variable puede significar cambios en la composición de las fases de los ATPS [9, 10], lo que a su vez puede influenciar la partición de biomoléculas en ellos. Este es el caso de sistemas de PEG/dextrano, donde se ha observado que la polidispersividad induce al fraccionamiento de las moléculas de PEG, tal que cadenas de alto peso molecular tienden a quedarse en la fase superior mientras que las de menor tamaño se mantienen en la fase inferior.

La incorporación de sales a los ATPS también afecta la partición de partículas en ellos, pues su presencia lleva a un cambio en la interacción de las moléculas en el sistema por un efecto de salting-out. Además, las especies iónicas asociadas a las sales no se distribuyen de manera uniforme entre las fases. Esto genera una diferencia en el potencial eléctrico entre ellas ( $\Delta\phi$ ) que también influencia el comportamiento de biomoléculas dentro del sistema acuoso.

Con respecto a las biomoléculas incorporadas al sistema, sus propiedades fisicoquímicas tales como hidrofobicidad, carga y tamaño también influyen su coeficiente de partición en ATPS. La magnitud del efecto de estas variables depende del sistema utilizado. Estas distintas propiedades se han relacionado con el coeficiente de partición de biomoléculas en ATPS a través de la siguiente expresión [11]:

$$K = K_{\text{hfob}} \cdot K_{\text{el}} \cdot K_{\text{tamaño}} \cdot K_{\text{sol}} \cdot K_{\text{af}} \quad (2.2)$$

Donde  $K$  se expresa como la multiplicación de factores asociados a la hidrofobicidad ( $K_{\text{hfob}}$ ), fuerzas electrostáticas ( $K_{\text{el}}$ ), tamaño ( $K_{\text{tamaño}}$ ), solubilidad ( $K_{\text{sol}}$ ), y afinidad de la partícula ( $K_{\text{af}}$ ).

En el caso de la hidrofobicidad, se ha observado que en sistemas PEG/sal las proteínas hidrofóbicas tienden a establecerse en la fase superior, rica en PEG. Además, se encontró que la incorporación de péptidos hidrofóbicos a una proteína modelo provoca un aumento de su coeficiente de partición en sistemas de PEG/fosfato [12, 13]. También se ha observado que la adición de NaCl a sistemas de dos fases acuosas lleva a una mayor correlación entre la hidrofobicidad de proteínas y su coeficiente de partición [14].

## 2.2 Modelos para la Partición de Biomoléculas en ATPS

Se han desarrollado distintos modelos para la predicción del coeficiente de partición de biomoléculas en sistemas de dos fases acuosas. Estos modelos se pueden separar principalmente en los basados en consideraciones termodinámicas asociadas al equilibrio en los sistemas acuosos y en los que asocian distintas características de las moléculas con su coeficiente de partición. Estos modelos se presentan en las Tablas 2.1 y 2.2 respectivamente.

**Tabla 2.1:** Modelos termodinámicos para la predicción del coeficiente de partición

N°	Autor	Modelo	Nomenclatura
1	Lin [7]	$\ln K = \ln \frac{W_1^T}{W_1^B} + 2M_P \left[ B_{P2}^* d_2^{n_2} \Delta \left( \frac{W_2}{W_1} \right) + B_{P3}^* d_3^{n_3} \Delta \left( \frac{W_3}{W_1} \right) \right]$	$B_{Pi}$ = Coeficiente asociado a la interacción entre la proteína y el polímero i $M_P$ = Peso molecular de la proteína $n_i$ = Parámetro asociado al peso molecular de los polímeros $d_i$ = N° de segmentos (monómeros) del polímero i $W_1$ = Fracción másica del agua $W_{2,3}$ = Fracciones másicas de los polímeros en el sistema
2	Madeira [15]	$\ln K = \ln \left( \frac{\gamma_p^b}{\gamma_p^t} \right) + \frac{z_p}{z_k} \left[ \ln \left( \frac{m_k^t}{m_k^b} \right) + \ln \left( \frac{\gamma_k^t}{\gamma_k^b} \right) \right]$	$\gamma$ = Coeficiente de actividad $k$ = Especie iónica de referencia $m$ = Molalidad $z$ = Carga neta
3	Pessôa – Mohamed [10]	$K = \frac{\gamma_{p,\infty}^b}{\gamma_{p,\infty}^t}$ $\ln \gamma_p = 1 + \lim_{C_p \rightarrow \infty} \ln \left( \frac{\phi_p}{C_p} \right) - v_b \sum_i \frac{\tilde{\phi}_i}{V_i} + \frac{V_b}{V_s} \left[ \sum_i \chi_{ip} \phi_i - \sum_i \sum_{k>j} \chi_{jk} \phi_j \phi_k \right]$	$\phi$ = Fracción volumétrica $\tilde{\phi}$ = Fracción volumétrica modificada por efecto de la hidratación de las moléculas de PEG $\chi_{ij}$ = Parámetro de interacción entre los compuestos i y j $C_P$ = Concentración proteína $V$ = Volumen por mol

**Tabla 2.2:** Modelos para la predicción del coeficiente de partición basados en propiedades de las biomoléculas

N°	Autor	Modelo	Nomenclatura
4	Albertsson [11]	$\ln K = \ln K_0 + \frac{Fz_p}{RT} \Delta\varphi$	$\Delta\varphi$ = Diferencia del potencial eléctrico entre las fases $K_0$ = Coeficiente de partición en ausencia de $\Delta\varphi$ $Z_p$ = Carga neta de la proteína
5	Hachem 1996, [16] Asenjo 1994 [17]	$\log K = R(\log P - \log P_0)$	$\log P$ = Hidrofobicidad proteína $\log P_0$ = Hidrofobicidad intrínseca $R$ = Resolución hidrofóbica
6	Eiteman [18]	$\ln K = D\Delta w_2 \log \frac{P}{P_0}$	$\Delta w_2$ = Diferencia de concentración entre las fases de uno de los componentes del sistema $\log P$ = Hidrofobicidad proteína $\log P_0$ = Hidrofobicidad intrínseca $D$ = Factor de discriminación
7	Eiteman [4]	$\frac{K}{K_0} = \frac{1 + \sum_{i=1}^m \Lambda_{i+} \frac{a_{H+}^i}{\prod_{l=1}^i K_{bl}} + \sum_{j=1}^n \Lambda_{j-} \frac{\prod_{l=1}^j K_{cl}}{a_{H+}^j}}{1 + \sum_{i=1}^m \Lambda_{i+} \frac{a_{H+}^i}{\prod_{l=1}^i K_{bl}} + \sum_{j=1}^n \Lambda_{j-} \frac{\prod_{l=1}^j K_{cl}}{a_{H+}^j}}$	$\Lambda$ = Tasa entre los coeficientes de actividad de una especie neutra y cargada $a_x$ = Actividad de la especie x $K_0$ = Coeficiente de partición proteínas neutras $K_{bl}$ = Constante de equilibrio para solutos negativos $K_{cl}$ = Constante de equilibrio para solutos positivos
8	Olivera-Nappa [19]	$\ln K = A(\delta E)^a (M_p)^b (L)^c + B(\delta E)^d (MW)^e (L)^f + C$	$\delta E$ = Diferencia de energía electrostática de la proteína entre las fases del sistema $M_p$ = Peso molecular de la proteína $L$ = Factor de esfericidad de la proteína
9	Olivera-Nappa [19]	$\ln K = A' \frac{(L)^{a'}}{(\text{pH})^{b'} (\delta E)^{c'} (M_p)^{d'}} - B' \frac{(\delta E)^{e'} (\text{pH})^{f'}}{(L)^{g'} (M_p)^{h'}} + C'$	$\delta E$ = Diferencia de energía electrostática de la proteína entre las fases del sistema $M_p$ = Peso molecular de la proteína $L$ = Factor de esfericidad de la proteína

El modelo 1, propuesto por Lin, Wu *et al.* [7], está basado en una modificación del modelo de Pitzer [20]. En él se relaciona K con la concentración y peso molecular de los polímeros que forman el sistema acuoso. El efecto de la concentración de los polímeros del sistema sobre K está representado por los términos  $\Delta(W_2/W_1)$  y  $\Delta(W_3/W_1)$ . Por otro lado, la influencia del peso molecular de estos polímeros se incluye en los términos  $n_i$ , de tal forma

que un valor negativo de  $n_i$  significa un aumento en  $K$  con el peso molecular del polímero  $i$  y uno positivo una disminución.

El modelo 1 se desarrolló para caracterizar la partición de proteínas en sistemas acuosos formados por dos polímeros, en la ausencia de sales y con pH cercano al punto isoeléctrico, por lo tanto, no considera el efecto de carga y ni el de la presencia de iones en el sistema.

El modelo 2, propuesto por Madeira, Xu *et al.* [15], se origina a partir de una modificación del modelo de Wilson [21] y fue desarrollado para sistemas de dos fases acuosas del tipo PEG/sal. En este caso el coeficiente de partición de proteínas se relaciona con la diferencia del potencial eléctrico ( $\Delta\phi$ ) entre las fases, generado producto de la distribución no uniforme de sustancias iónicas en el sistema. Este potencial se calcula a partir de la diferencia de potencial químico de una especie iónica de referencia  $k$  presente en sistema. Este cálculo lleva a que el coeficiente de partición de proteínas pase a depender de la carga y actividad de  $k$ .

El modelo 3 fue desarrollado por Pessôa y Mohamed [10], para la partición de biomoléculas en ATPS formados por PEG y dextrano. En este caso, el coeficiente de partición de biomoléculas se expresa a partir de sus coeficientes de actividad en las fases del sistema, considerando una dilución infinita. Los coeficientes de actividad se obtienen a partir de un modelo para el exceso de energía de Gibbs basado en la ecuación de Flory-Huggins con 2 modificaciones que consideran particularidades del sistema de dos fases acuosas estudiado.

La teoría de Flory-Huggins modela la termodinámica de polímeros en solución. La mezcla líquida se modela como una red cristalina en la cual los monómeros que forman los polímeros y las moléculas del solvente se distribuyen ocupando los sitios de la red. De esta manera, la entropía  $\Delta S$ , se calcula considerando todas las formas posibles de distribuir las moléculas de solvente y polímeros en la red.

La primera modificación a este modelo considera la interacción entre las moléculas de PEG y las de agua en el sistema. El modelo de Flory-Huggins asume que las moléculas de polímeros y del solvente se mezclan al azar, sin embargo, en el caso de PEG disuelto en agua, sus moléculas están fuertemente unidas mediante enlaces de hidrógeno. Así, las moléculas del polímero se pueden modelar como cadenas más largas debido a su unión con moléculas de agua, llevando a cambios en el cálculo de la entropía. La segunda modificación incorpora la polidispersividad de las moléculas de dextrano en el modelo, representando los distintos pesos moleculares del polímero mediante 6 pseudo-componentes.

El modelo 4, propuesto por Albertsson [11], expresa el coeficiente de partición de una biomolécula en función de su carga y de la diferencia del potencial eléctrico entre las fases del sistema.

El modelo 5 corresponde a una correlación entre el coeficiente  $K$  y la hidrofobicidad de las proteínas medida por precipitación. Los parámetros  $R$  y  $\log P_0$  corresponden a la resolución hidrofóbica e hidrofobicidad intrínseca del sistema. La resolución hidrofóbica representa cuan sensible es el sistema ante variaciones en la hidrofobicidad de las proteínas y la hidrofobicidad intrínseca corresponde a un umbral, de tal manera que si la hidrofobicidad de una proteína es igual a  $P_0$ , ésta se distribuirá de manera uniforme entre las fases.

Los modelos 6 y 7 fueron desarrollados por Eiteman [4, 18]. En el primero, el coeficiente de partición de biomoléculas en ATPS se expresa en función de su hidrofobicidad y la diferencia de concentración entre las fases de uno de los polímeros que la componen. El segundo toma en cuenta las cargas de las biomoléculas, incorporando el efecto de la diferencia de pH entre las fases del sistema sobre la partición de estas partículas. El modelo permite calcular el coeficiente de partición de moléculas cargadas con respecto su coeficiente en el caso neutro  $K_0$ , dado por el modelo 6.

Los modelos 8 y 9 fueron propuestos por Olivera-Nappa *et al.* [19]. Estos son modelos estadísticos que correlacionan el coeficiente de partición de proteínas con su peso molecular, su forma y la diferencia de su energía electrostática entre las fases del sistema. El modelo 9 además incorpora el pH del sistema como una variable más.

### **3. OBJETIVOS**

#### **Objetivo General**

Determinar la relevancia de la distribución de la hidrofobicidad en la superficie de las proteínas en su comportamiento en ATPS y establecer si es posible predecir su coeficiente de partición en base a esta distribución.

#### **Objetivos Específicos**

- Establecer una descripción cuantitativa de la distribución de la hidrofobicidad en la superficie de las proteínas.
- Estudiar la correlación entre esta distribución y el coeficiente de partición en sistemas de dos fases acuosas. En particular, estudiar la calidad de la descripción cuantitativa obtenida al utilizarla para la predicción del coeficiente de partición de proteínas en estos sistemas.

## 4. METODOLOGÍA

### 4.1 *Proteínas y Sistemas considerados en el estudio*

Para este estudio se consideraron 8 proteínas de estructura tridimensional conocida para las cuales los coeficientes de partición en 12 sistemas de dos fases acuosas fueron reportados en un estudio anterior [14]. Estos sistemas corresponden a 3 sistemas del tipo polímero/sal: PEG4000/Fosfato, PEG4000/Sulfato y PEG4000/Citrato y un sistema de dos polímeros PEG4000/Dextrano. Cada uno de ellos con concentraciones nula, baja y alta de NaCl de 0%, 0,6% y 8,8% p/p respectivamente.

Las proteínas seleccionadas junto a su código PDB<sup>1</sup>[22] se muestran en la Tabla 4.1.

**Tabla 4.1:** Proteínas seleccionadas para el estudio junto con su código PDB

Proteína	Código PDB
$\alpha$ -Lactalbúmina	1F6S
$\alpha$ -Quimotripsinógeno	2CGA
Conalbúmina	1OVT
Inhibidor de la Tripsina	1AVU
Lisozima	2LYM
Ovalbúmina	1OVA
Subtilisina	1SBC
Taumatina	1THV

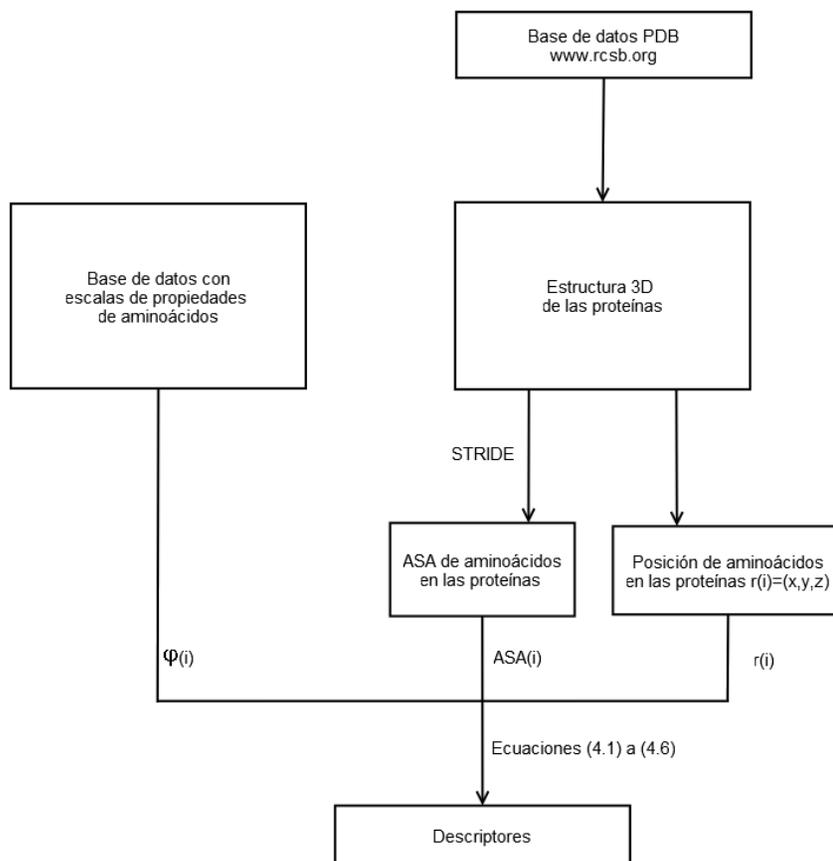
### 4.2 *Cuantificación de la distribución de la hidrofobicidad en la superficie de las proteínas*

Para cuantificar la distribución de la hidrofobicidad en la superficie de las proteínas, está se modeló utilizando 7 descriptores de dos tipos: 4 descriptores estadísticos y 3 descriptores basados en el vector de desbalance hidrofóbico. Estos descriptores fueron definidos previamente por Salgado *et al.*, para estudiar el comportamiento de proteínas en cromatografía de intercambio hidrofóbico [23, 24].

Todos estos descriptores se calculan a partir de información asociada a la estructura tridimensional de las proteínas. En particular, es necesario para su cálculo conocer, para cada aminoácido en una proteína, su posición, área superficial accesible (ASA) y grado de hidrofobicidad. La Figura 4.1 muestra un esquema de la metodología seguida para la obtención de los descriptores.

---

<sup>1</sup> [www.rcsb.org](http://www.rcsb.org)



**Figura 4.1: Esquema de la metodología seguida en el cálculo de los descriptores**

Para cada proteína considerada se obtuvo su estructura tridimensional de la base de datos PDB [22]. Esta información corresponde a la posición de los átomos que forman cada uno de los aminoácidos presentes en una proteína. A partir de estos datos se calculó el área superficial accesible de cada aminoácido en las proteínas utilizando el software STRIDE [25].

Para obtener la posición de los aminoácidos se establecieron dos alternativas. La primera corresponde a utilizar las coordenadas del carbono- $\beta$  de los aminoácidos como su posición (carbono- $\alpha$  para el caso de la glicina). La segunda define esta posición como el centro geométrico de los átomos del aminoácido.

La hidrofobicidad de los aminoácidos se obtuvo a partir de 37 escalas distintas correspondientes a esta propiedad [26-54]. Además se consideraron 37 escalas más, asociadas a distintas propiedades [36, 38, 40, 47, 54-65]. Entre ellas se encuentran peso molecular, refractividad, flexibilidad promedio y frecuencia en distintas estructuras secundarias<sup>2</sup>.

<sup>2</sup> Los detalles de estas escalas se presentan en Anexos.

## 4.2.1 Descriptores estadísticos

Los descriptores estadísticos se calcularon en base a la propiedad superficial promedio local (ASP local) [24]. Estos ASP son una forma de cuantificar una propiedad  $\varphi$ , en este caso, la hidrofobicidad, alrededor de cada uno de los aminoácidos presentes en la superficie proteica. Los ASP locales se calculan mediante la ecuación (4.1), definiendo un vecindario  $N_r$  de radio  $r$  alrededor de cada aminoácido superficial  $k$  y considerando los aminoácidos  $i$  dentro de este volumen (Figura 4.2).

$$ASP(k) = \frac{\sum_{i \in N_r(k)} ASA(i) \cdot \varphi(i)}{\sum_{i \in N_r(k)} ASA(i)} \quad (4.1)$$

De esta forma, se tendrá un conjunto de ASP locales asociados a la superficie de las proteínas a partir del cual se obtienen los descriptores estadísticos ASPavg, ASPmax, ASPmin y ASPstd correspondientes al promedio, máximo, mínimo y desviación estándar del promedio dentro de este conjunto respectivamente.

Estos descriptores se calcularán considerando vecindarios con radios de entre 5 y 25 Å, aumentando 1 Å cada vez. Así, se abarcará desde vecindarios pequeños, con pocos aminoácidos, a sectores de mayor tamaño que incluyan gran parte de la superficie de las proteínas.

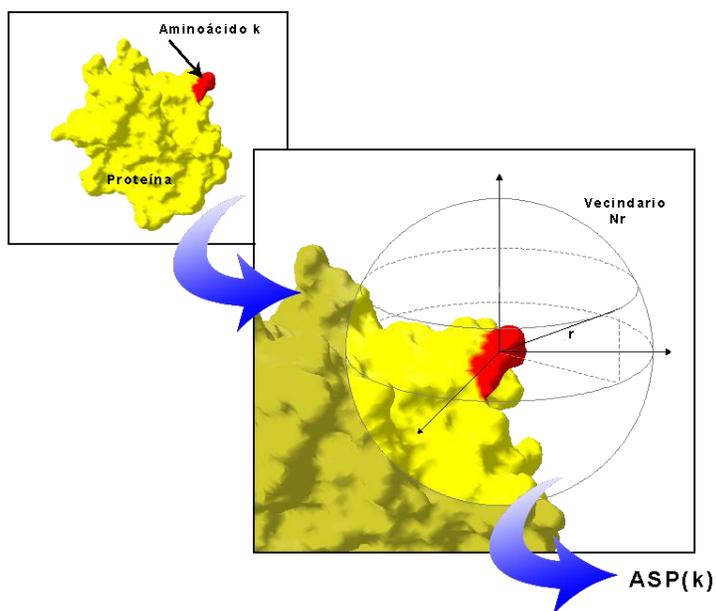


Figura 4.2: Esquema correspondiente a la obtención del ASP local para un aminoácido  $k$  en la superficie de una proteína

## 4.2.2 Descriptores basados en el vector de desbalance hidrofóbico

Una forma de cuantificar el efecto de la hidrofobicidad en la superficie de las proteínas es midiendo cuanto se desplaza su centro geométrico al tener en cuenta este factor. Este desplazamiento se representa mediante el vector de desbalance hidrofóbico  $r_{HI}$ , el cual se calcula como la diferencia entre el centro geométrico superficial  $r_{CG}$  y el centro geométrico desplazado por efecto de la hidrofobicidad de los aminoácidos en la superficie,  $r_H$  [23].

Para cuantificar la distribución de la hidrofobicidad de las proteínas se utilizó la magnitud del vector,  $HI$ , calculándola mediante las ecuaciones (4.2) a (4.4).

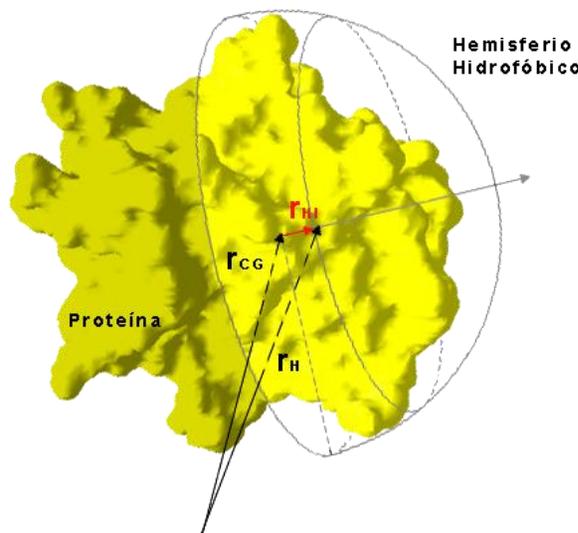
$$HI = |r_H - r_{CG}| \quad (4.2)$$

$$r_{CG} = \frac{\sum_{i \in S} ASA(i) \cdot r(i)}{\sum_{i \in S} ASA(i)} \quad (4.3)$$

$$r_H = \frac{\sum_{i \in S} ASA(i) \cdot r(i) \cdot \varphi(i)}{\sum_{i \in S} ASA(i) \cdot \varphi(i)} \quad (4.4)$$

Donde  $ASA(i)$ ,  $r(i)$  y  $\varphi(i)$  corresponden al área superficial accesible, posición e hidrofobicidad del aminoácido  $i$  respectivamente y  $S$  la superficie total de la proteína.

El vector  $r_{HI}$  apunta hacia el sector más hidrofóbico de la superficie de una proteína. Esto permite dividir la superficie proteica en dos hemisferios separados por un plano perpendicular al vector  $r_{HI}$ , donde el hemisferio hacia donde apunta este vector se define como el hemisferio hidrofóbico (Figura 4.3).



**Figura 4.3: Hemisferio Hidrofóbico de la superficie de una proteína ( $r_{CG}$  = centro geométrico superficial;  $r_H$  = centro geométrico desplazado;  $r_{HI}$  = vector de desbalance hidrofóbico).**

Como otra forma de describir la distribución de la hidrofobicidad en la superficie de las proteínas se utilizó el ASP local dentro del hemisferio hidrofóbico. Este se calculó mediante las ecuaciones (4.5) y (4.6), donde  $ASP_{90^\circ, T}$  y  $ASP_{90^\circ, P}$  corresponden al ASP calculado en relación a la superficie total de la proteína,  $S$ , y a la superficie del hemisferio hidrofóbico,  $HH$ , respectivamente.

$$ASP_{90^\circ, P} = \frac{\sum_{i \in HH} ASA(i) \cdot \varphi(i)}{\sum_{i \in HH} ASA(i)} \quad (4.5)$$

$$ASP_{90^\circ, T} = \frac{\sum_{i \in HH} ASA(i) \cdot \varphi(i)}{\sum_{i \in S} ASA(i)} \quad (4.6)$$

### 4.3 Predicción del coeficiente de partición de proteínas en ATPS

Para estudiar la capacidad de los descriptores definidos anteriormente para predecir el coeficiente de partición de proteínas en ATPS, se utilizó un modelo equivalente al modelo 5. En este caso, en vez de utilizar la medición de la hidrofobicidad por precipitación se utilizarán los valores obtenidos con los distintos descriptores  $D$ . Así el modelo utilizado será de la forma:

$$\log K = R(\log D - \log P_o) \quad (4.7)$$

Donde, como se mencionó anteriormente, los parámetros  $R$  y  $\log P_o$  corresponden a la resolución hidrofóbica e hidrofobicidad intrínseca del sistema respectivamente.

Para medir la calidad de la predicción de los modelos obtenidos se utiliza el error cuadrático medio de Jack-Knife,  $MSE_{JK}$ . Este error se obtiene mediante un proceso de ajuste reiterativo, dejando un elemento del set de datos afuera cada vez. Para calcularlo se realiza el siguiente procedimiento:

**Para cada proteína  $i$  en el grupo de proteínas considerado**

**Obtener los parámetros del modelo de la ecuación (4.7) mediante un ajuste de todos los datos excepto los asociados a la proteína  $i$**

**Utilizar este modelo para calcular el coeficiente de partición de la proteína  $i$ ,  $\log K_i'$**

**Calcular el error entre el valor experimental  $\log K_i$  y  $\log K_i'$**

**Calcular el promedio de los errores obtenidos mediante la siguiente ecuación:**

$$MSE_{JK} = \frac{1}{n} \sum_{i=1}^n (\log K_i - \log K_i')^2 \quad (4.8)$$

## 5 RESULTADOS Y DISCUSIÓN

### 5.1 Descripción de la distribución de la hidrofobicidad en la superficie de las proteínas

#### a) Proteínas seleccionadas

En la Tabla 5.1 se presentan las características de las 8 proteínas seleccionadas para este estudio.

**Tabla 5.1:** Código PDB [22], Peso molecular y punto isoelectrico de las 8 proteínas. Los pesos moleculares y puntos isoelectricos son medidas experimentales obtenidas de [14] . La superficie accesible se calculó utilizando el software STRIDE [25].

Proteína	Código PDB	Peso Molecular (KDa)	pI	Superficie Accesible (nm <sup>2</sup> )
$\alpha$ -Lactalbúmina	1F6S	14,4	4,6	70,6
$\alpha$ -Quimotripsinógeno	2CGA	23,6	8,2	107,5
Conalbúmina	1OVT	77,0	5,7	290,6
Inhibidor de la Tripsina	1AVU	24,5	4,2	84,9
Lisozima	2LYM	14,4	11	68,3
Ovalbúmina	1OVA	45,0	4,2	157,9
Subtilisina	1SBC	27,5	7,7	99,6
Taumatina	1THV	22,2	11	96,3

#### b) Descripción cualitativa de la distribución de hidrofobicidad en la superficie proteica

Para describir de manera cualitativa la distribución superficial de la hidrofobicidad en las proteínas, se utilizó una representación gráfica de su superficie destacando las zonas ocupadas por los aminoácidos hidrofóbicos: alanina, fenilalanina, isoleucina, leucina, metionina, prolina, y valina. El software Swiss-Pdb Viewer [66] se utilizó para este estudio cualitativo.

En las Figuras 5.1 a 5.8 se muestran las superficies de las 8 proteínas antes mencionadas.

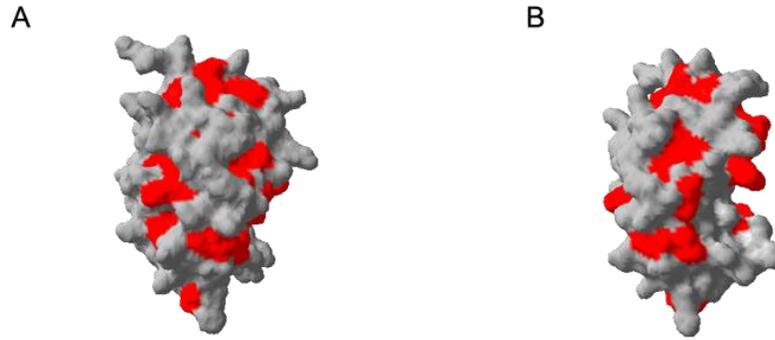


Figura 5.1: Superficie de la proteína  $\alpha$ -Lactalbúmina. Las zonas hidrofóbicas se destacan en color rojo.

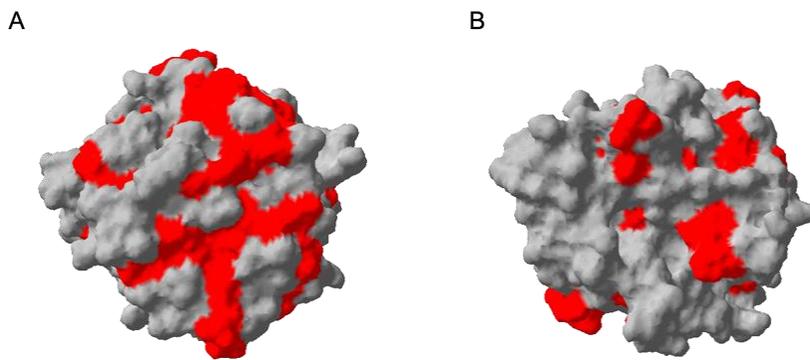


Figura 5.2: Superficie de la proteína  $\alpha$ -Quimotripsinógeno. Las zonas hidrofóbicas se destacan en color rojo.

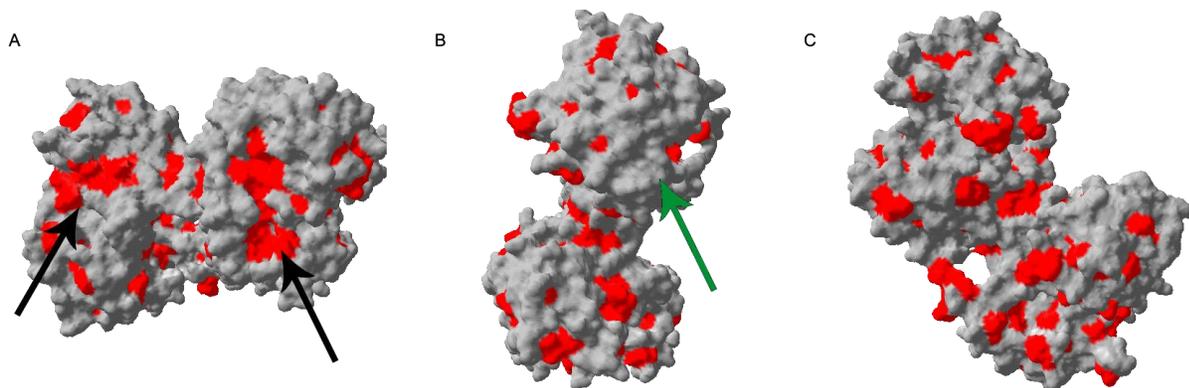
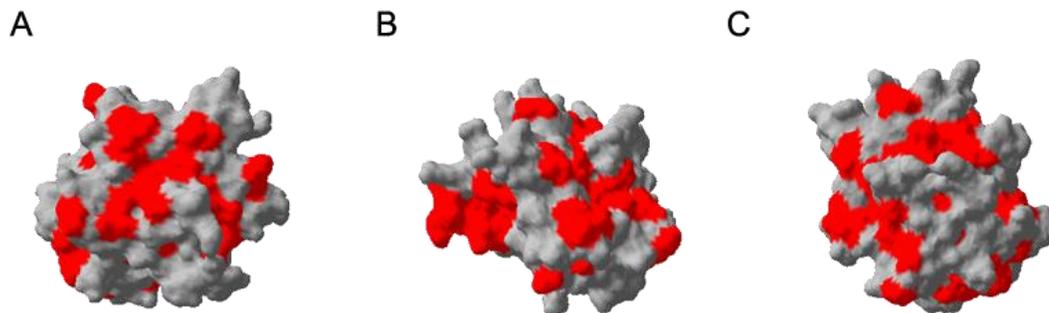
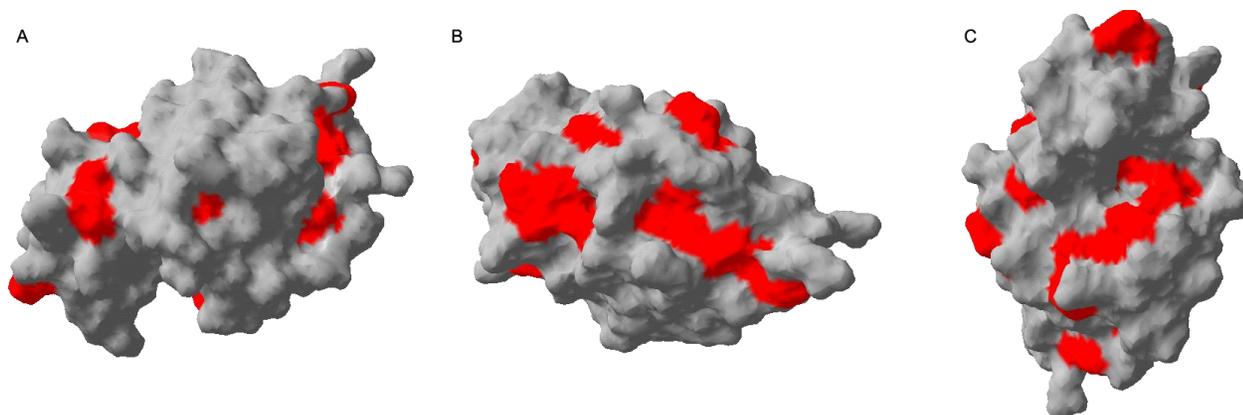


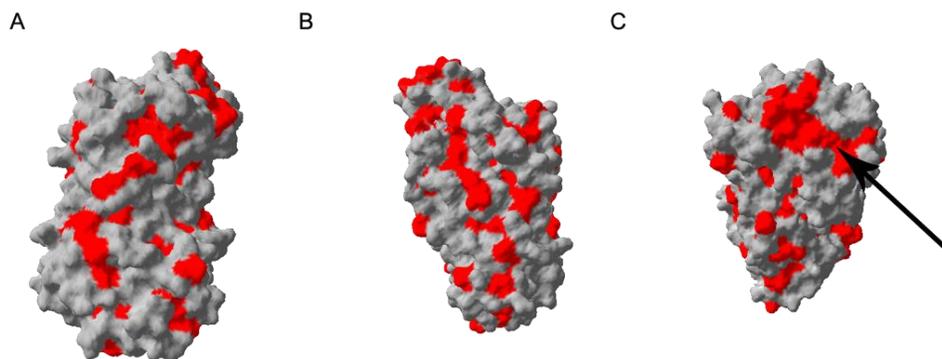
Figura 5.3: Superficie de la proteína Conalbúmina. Las zonas hidrofóbicas se destacan en color rojo. Las flechas negras y verdes indican las zonas de mayor y menor hidrofobicidad respectivamente.



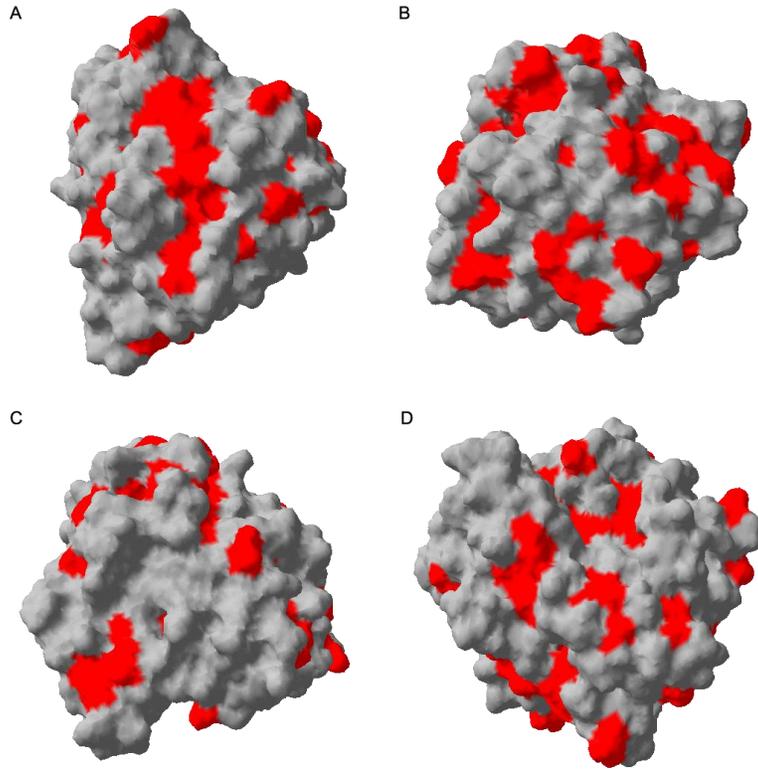
**Figura 5.4:** Superficie de la proteína Inhibidora de la Tripsina. Las zonas hidrofóbicas se destacan en color rojo.



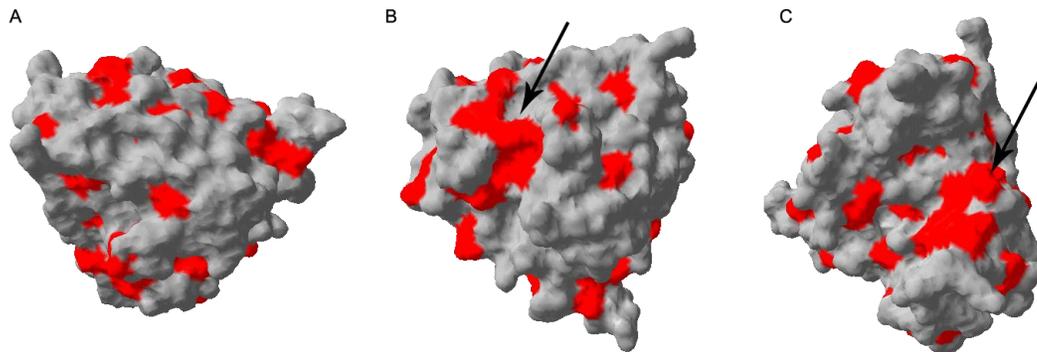
**Figura 5.5:** Superficie de la proteína Lisozima. Las zonas hidrofóbicas se destacan en color rojo.



**Figura 5.6:** Superficie de la proteína Ovalbúmina. Las zonas hidrofóbicas se destacan en color rojo. La flecha indica la zona de mayor hidrofobicidad.



**Figura 5.7:** Superficie de la proteína Subtilisina. Las zonas hidrofóbicas se destacan en color rojo.



**Figura 5.8:** Superficie de la proteína Taumatina. Las zonas hidrofóbicas se destacan en color rojo. Las flechas indican las zonas de mayor hidrofobicidad.

### **Proteína $\alpha$ -Lactalbúmina**

Esta proteína es una de las más pequeñas del grupo considerado. En ella, la hidrofobicidad se distribuye de manera homogénea, encontrándose principalmente parches hidrofóbicos de tamaño medio (Figura 5.1 A y B).

### **Proteína $\alpha$ -Quimotripsinógeno**

En el caso de esta proteína, la hidrofobicidad se distribuye a lo largo de toda su superficie. En relación a esta propiedad se pueden destacar dos zonas. Una en que se observa un manto de gran tamaño formado por aminoácidos hidrofóbicos (Figura 5.2 A) y otra cóncava, en la cual hay un alto predominio de residuos no hidrofóbicos (Figura 5.2 B).

### **Proteína Conalbúmina**

La distribución de la hidrofobicidad en la superficie de esta proteína es heterogénea. Se observan zonas con parches hidrofóbicos de gran tamaño (Figura 5.3 A), áreas de baja densidad hidrofóbica (Figura 5.3 B) y todo un costado que presenta parches hidrofóbicos de tamaño medio donde la distribución es más homogénea (Figura 5.3 C).

### **Proteína Inhibidor de la Tripsina**

Esta proteína tiene una forma relativamente esférica. En la Figura 5.4 se puede apreciar que la superficie de esta proteína presenta un alto contenido de aminoácidos hidrofóbicos. Estos se distribuyen de manera uniforme, principalmente en forma de parches de gran tamaño.

### **Proteína Lisozima**

En este caso, zonas no hidrofóbicas ocupan la mayor parte de la superficie de la proteína. Se puede observar que la distribución de la hidrofobicidad en su superficie no es homogénea. Presenta un sector convexo con una densidad hidrofóbica muy baja (Figura 5.5 A) y dos zonas en que ésta es mayor, donde se aprecia la existencia de parches hidrofóbicos de gran tamaño (Figura 5.5 B y C).

### **Proteína Ovalbúmina**

La superficie de esta proteína es muy amplia, siendo la segunda más grande del grupo con  $157,9 \text{ nm}^2$ . En la Figura 5.6 A y B se puede apreciar que la hidrofobicidad se reparte a lo largo de toda esta extensión, observándose una distribución aproximadamente homogénea en la forma de parches hidrofóbicos de tamaño pequeño y medio. Sin embargo, existe un sector en que esta propiedad se concentra en un gran parche, tal como se muestra en la Figura 5.6 C.

## Proteína Subtilisina

La superficie de esta proteína se asemeja a la de un cubo cortado por su diagonal. En la Figura 5.7 C se observa un sector de densidad hidrofóbica muy baja. En el resto de su superficie la hidrofobicidad se distribuye con relativa homogeneidad en parches de tamaño medio (Figura 5.7 B y D), con la excepción de una zona que presenta un área hidrofóbica de tamaño mayor.

## Proteína Taumatina

En esta proteína se observa que las zonas de menor densidad hidrofóbica en su superficie son mayores que los sectores donde esta propiedad está más presente (Figura 5.8 A, B y C). La hidrofobicidad se distribuye en parches de tamaño variable destacándose dos áreas hidrofóbicas un poco mayores (Figura 5.8 B y C).

En la Tabla 5.2 se muestra el resumen de las observaciones antes mencionadas, correspondiente al tipo de distribución de hidrofobicidad presente en la superficie de cada una de las proteínas.

**Tabla 5.2:** Tipo de distribución de la hidrofobicidad en la superficie de las proteínas

Proteína	Distribución superficial de la hidrofobicidad
$\alpha$ -Lactalbúmina	Homogénea
$\alpha$ -Quimotripsinógeno	Heterogénea
Conalbúmina	Heterogénea
Inhibidor de la Tripsina	Homogénea
Lisozima	Heterogénea
Ovalbúmina	Homogénea
Subtilisina	Heterogénea
Taumatina	Heterogénea

## 5.2 Diferencias al considerar la posición de los aminoácidos en las proteínas como el centro geométrico de sus átomos y como la ubicación de su carbono- $\beta$ .

Al calcular los distintos descriptores de la distribución de la hidrofobicidad se utilizaron dos alternativas: considerar la posición de los aminoácidos en las proteínas como el centro geométrico de sus átomos (CG) y tomar esta posición como las coordenadas de su carbono  $\beta$  (CB).

A continuación se muestran las diferencias entre estas alternativas a nivel de los valores de los descriptores considerados y en la predicción del coeficiente de partición hecha en base a estos descriptores.

### a) Diferencias en los valores de los descriptores calculados con CG y CB

En la Tabla 5.3 se muestran las diferencias en los valores de los descriptores al calcularlos mediante las 2 alternativas. Para la mayoría de los descriptores, se observa que las diferencias promedio son bajas, no superando el 5% del valor obtenido utilizando el centro geométrico de los aminoácidos. Una excepción es el caso de ASPmin en que la diferencia promedio es de un 12% y la diferencia máxima es muy alta. Esto se debe a casos correspondientes a vecindarios de radios pequeños donde el efecto del corrimiento del centro de los aminoácidos se hace más importante.

**Tabla 5.3:** Diferencias máximas y promedio en los valores de los descriptores al considerar la posición de los aminoácidos de las proteínas como el centro geométrico de sus átomos y como la ubicación de su carbono  $\beta$  en su cálculo. Las diferencias se muestran como porcentaje del valor obtenido utilizando el centro geométrico CG.

	Diferencia máxima (% descriptor <sub>CG</sub> )	Diferencia promedio (%descriptor <sub>CG</sub> )
ASP <sub>90°P</sub>	10,06	1,25
ASP <sub>90°T</sub>	17,68	2,18
HI	31,46	4,70
ASPprom	26,89	0,87
ASPdest	68,17	4,09
ASPmax	93,66	2,96
ASPmin	2728,62	11,98

### b) Diferencias en la predicción del coeficiente de partición al utilizar descriptores calculados con CB y CG

En general, al comparar los mejores resultados obtenidos con cada descriptor en la predicción de K, es decir aquellos modelos con MSE<sub>JK</sub> menores, se observa que se obtienen errores menores con los descriptores obtenidos con CG. Además, los casos en que aquellos calculados con CB son mejores, las diferencias promedio no superan el 20%.

Por esta razón y debido a que las diferencias en los valores de los descriptores obtenidos no son muy altas, en el análisis que sigue sólo se consideraron los descriptores calculados utilizando el centro geométrico de los átomos de los aminoácidos en las proteínas como su posición.

## 5.3 Análisis de los descriptores estadísticos

Los descriptores estadísticos ASPavg, ASPstd, ASPmax y ASPmin corresponden al promedio, desviación estándar, máximo y mínimo de los ASP locales asociados a vecindarios alrededor de los aminoácidos superficiales de las proteínas. A continuación se muestran las diferencias en estos descriptores dentro del conjunto de las 8 proteínas en

estudio y el efecto que tienen sobre ellos las escalas de propiedades utilizadas y los radios de los vecindarios.

### 5.3.1 Variaciones en los descriptores estadísticos en el set de proteínas

Se estudiaron las variaciones en los valores de los descriptores en el conjunto de las 8 proteínas consideradas para poder tener una idea de cuán disímiles son con respecto a los descriptores. Para hacer esto, para cada par de parámetros distinto -escala y radio- se calcularon los valores promedio de las estadísticas en el conjunto de proteínas y la desviación estándar de este promedio.

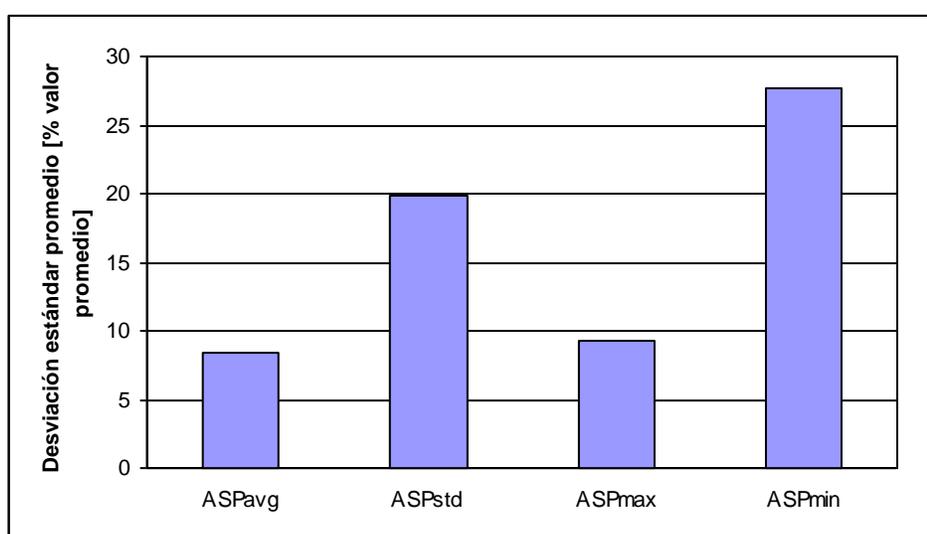


Figura 5.9: Desviación estándar promedio de los valores de los descriptores en el set de proteínas.

En la Figura 5.9 se observa que los valores de los 4 descriptores estadísticos utilizados no presentan variaciones demasiado elevadas en el set de proteínas. Las desviaciones estándares promedio mayores se obtuvieron para ASPmin y ASPstd, siendo éstas de 27,7% y 19,9% respectivamente. Los descriptores ASPavg y ASPmax son menos sensibles aún a las diferencias entre las proteínas, con desviaciones inferiores al 10%<sup>3</sup>.

### 5.3.2 Efecto de las distintas escalas de propiedades utilizadas

Para estudiar cuanto varían los descriptores al utilizar distintas escalas de propiedades se procedió de manera análoga al punto anterior. Para cada par (proteína, radio) distinto se calculó el valor promedio de los descriptores y la desviación estándar correspondiente. En la Figura 5.10 se observa que el descriptor más sensible a la escala de propiedades utilizadas es ASPmin, el cual exhibe desviaciones estándar de entre 28% y

<sup>3</sup> Los detalles de este cálculo se presentan en Anexos.

258% del promedio. Por otro lado, el descriptor que presenta menos variaciones en el grupo de escalas utilizado es ASPmax donde la desviación no supera el 29% del promedio. Para ASPavg y ASPstd, este valor está alrededor del 27%.

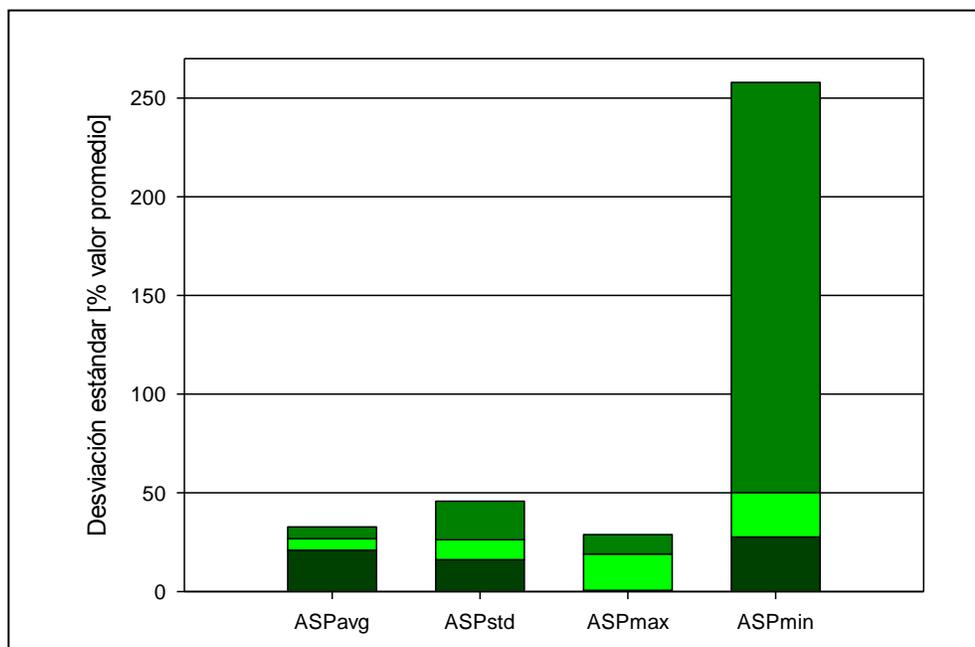


Figura 5.10: Desviación estándar de los descriptores en el set de escalas de propiedades. Cada columna incluye las desviaciones mínima, promedio y máxima.

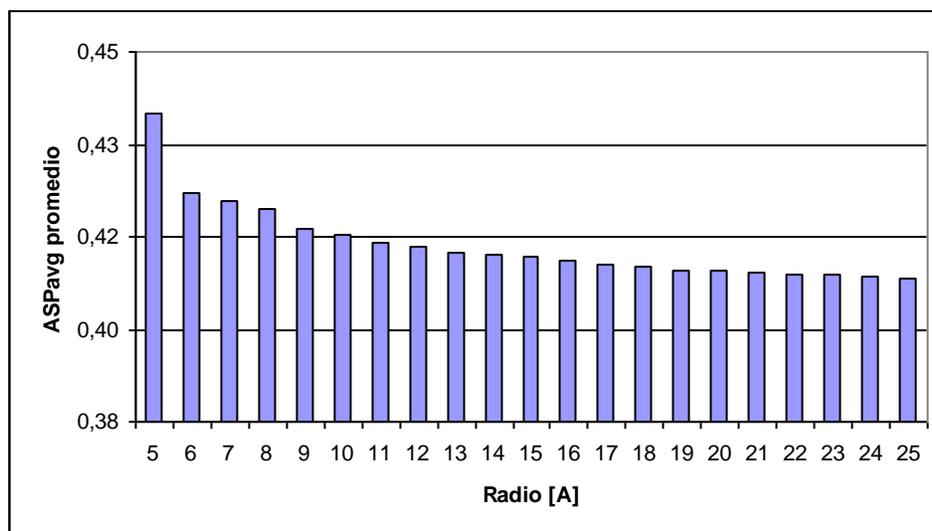
### 5.3.3 Efecto de la variación de los radios de vecindarios asociados a los ASP locales

En las Figuras 5.11 a 5.13 se observa que los valores promedio de ASPavg, ASPstd y ASPmax disminuyen al aumentar el radio de los vecindarios utilizados para el cálculo de los ASP locales. Además se puede ver que la tasa de esta disminución es decreciente, de tal manera que los valores de estos descriptores tienden a estabilizarse con radios superiores.

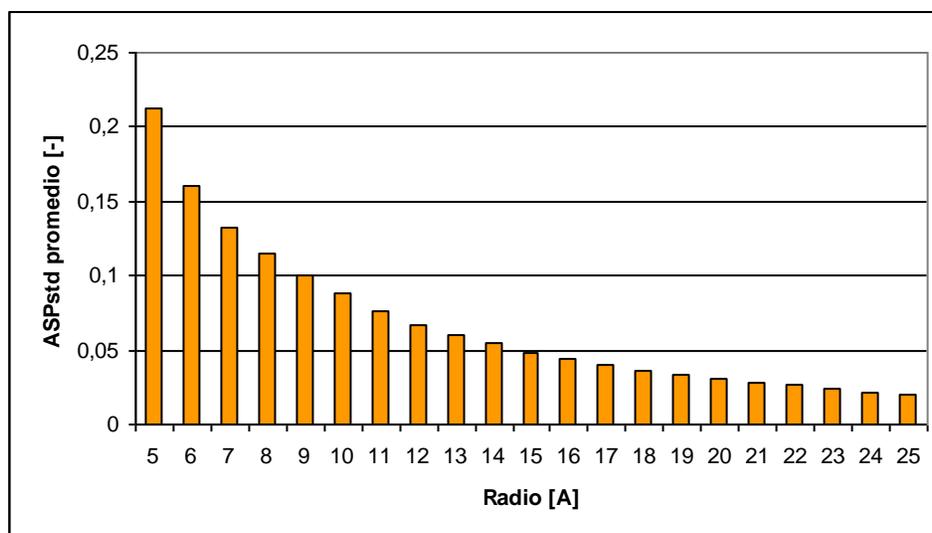
Para ASPavg, es interesante notar que al aumentar el radio sus valores tienden a convergen a los de la propiedad superficial promedio ASP, tal como se muestra en la Figura 5.15. El ASP es una medida global, equivalente a calcular un ASP local considerando un radio muy grande de tal forma que los vecindarios obtenidos contengan la superficie total de la proteína.

En el caso de ASPmin, la relación entre el valor de esta estadística y el radio del vecindario es opuesta a la de los casos anteriores, obteniéndose valores promedio mayores con radios más grandes y que la tasa de este aumento disminuye con el radio, como puede observarse en la Figura 5.14. Esto se explica debido a que vecindarios con radios mayores tienen un volumen superior y por lo tanto albergan un número más grande de aminoácidos en su interior.

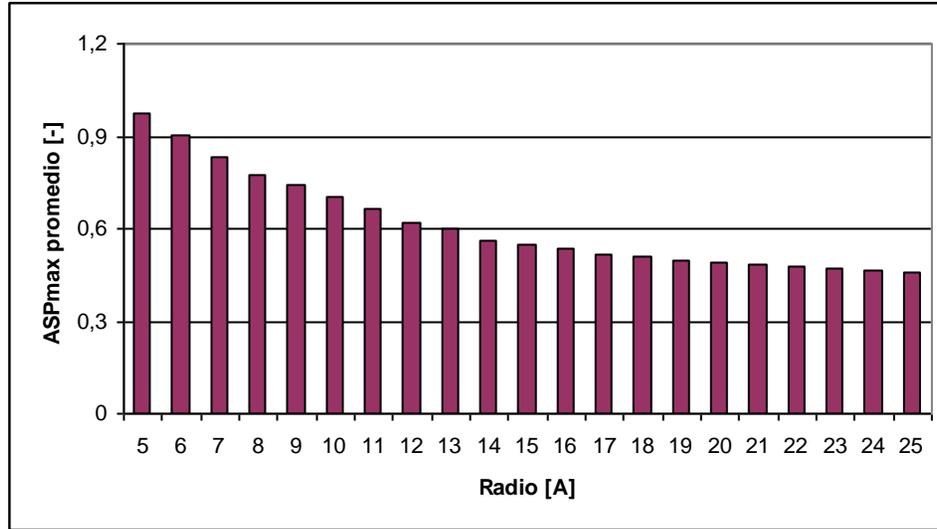
También es importante notar que en el rango de radios considerado, los valores promedio de ASPavg varían entre 0,4 y 0,44 al considerar todas las escalas de propiedades y proteínas. Un rango bastante pequeño, en comparación con los otros descriptores. Esto se puede explicar por el comportamiento opuesto que presentan ASPmax y ASPmin a medida que aumenta el radio de los vecindarios.



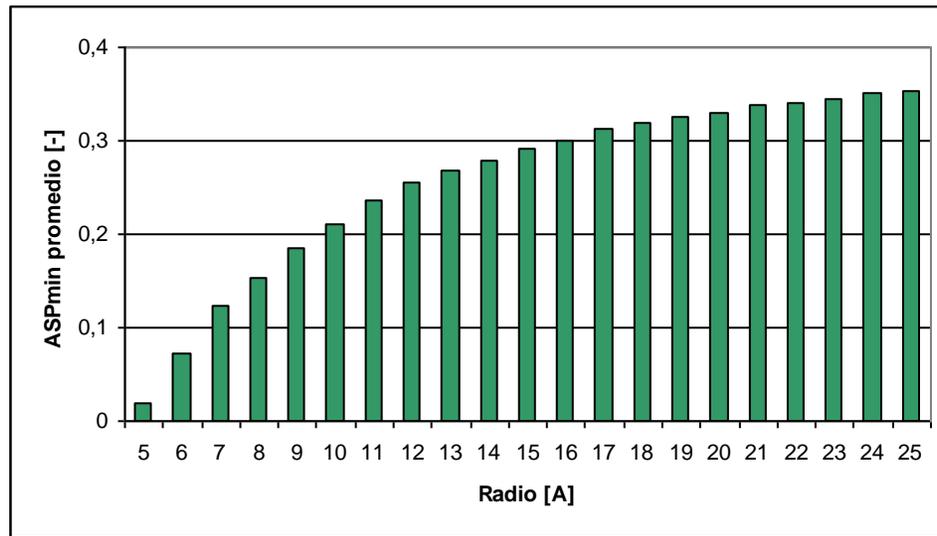
**Figura 5.11:** Efecto del radio del vecindario asociado a los ASP locales sobre ASPavg. Los valores de ASPavg corresponden al promedio de los resultados obtenidos utilizando las 74 escalas y las 8 proteínas para cada radio.



**Figura 5.12:** Efecto del radio del vecindario asociado a los ASP locales sobre ASPstd. Los valores de ASPstd corresponden al promedio de los resultados obtenidos utilizando las 74 escalas y las 8 proteínas para cada radio.



**Figura 5.13:** Efecto del radio del vecindario asociado a los ASP locales sobre ASPmax. Los valores de ASPmax corresponden al promedio de los resultados obtenidos utilizando las 74 escalas y las 8 proteínas para cada radio.



**Figura 5.14:** Efecto del radio del vecindario asociado a los ASP locales sobre ASPmin. Los valores de ASPmin corresponden al promedio de los resultados obtenidos utilizando las 74 escalas y las 8 proteínas para cada radio.

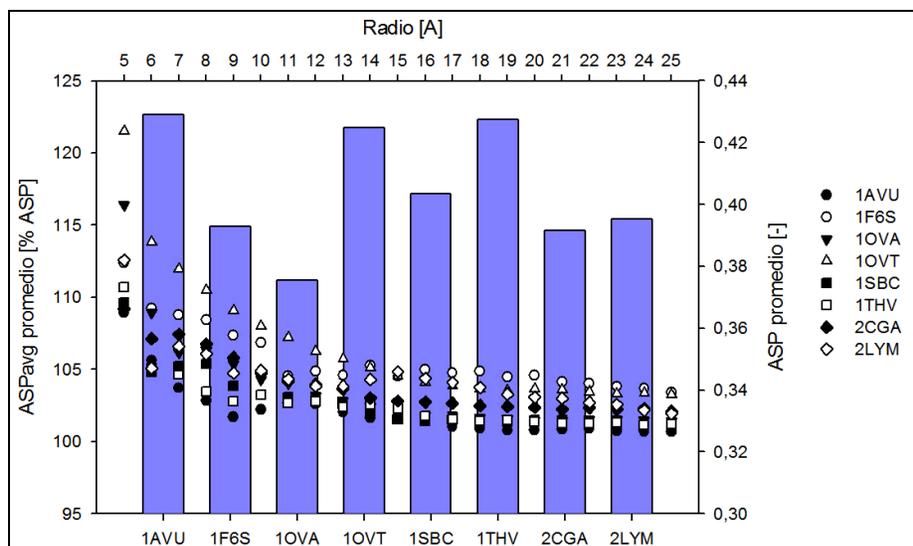


Figura 5.15: Convergencia de los valores de ASPavg promedio y valores promedio de ASP para las 8 proteínas consideradas en el estudio. Los valores promedio de ASP corresponden a las barras celestes y están asociados al eje derecho del gráfico. Los valores ASPavg promedio para las distintas proteínas se muestran para cada radio del eje superior y sus valores corresponden al eje izquierdo.

### 5.3.4 Correlación entre descriptores

En la Tabla 5.4 se muestra la correlación promedio entre los descriptores estadísticos para cada radio. En ella se puede observar que ASPstd presenta la menor correlación con el resto de los descriptores. Además se ve que a medida que aumenta el radio utilizado, la correlación entre los pares de descriptores ASPavg–ASPmax, ASPavg–ASPmin y ASPmax–ASPmin tiende a aumentar. Esto se explica considerando que los ASP locales muestran una desviación estándar del promedio que disminuye con el radio, llegando a ser muy bajos en radios de entre 24 y 25 Å (Figura 5.12). De esta forma, con vecindarios de mayor volumen, la información que entregan ASPmax, ASPmin y ASPavg, con respecto a la propiedad medida en la superficie de las proteínas, se hace más similar.

**Tabla 5.4:** Coeficientes de correlación promedio entre ASPavg, ASPstd, ASPmax y ASPmin en el set de proteínas considerando los valores obtenidos para las 74 escalas utilizadas.

Radio [Å]	ASPavg ASPstd	ASPavg ASPmax	ASPavg ASPmin	ASPstd ASPmax	ASPstd ASPmin	ASPmax ASPmin
5	-0,089	0,095	0,224	0,371	-0,373	-0,110
6	0,169	0,279	0,156	0,476	-0,374	-0,203
7	0,103	0,176	0,380	0,221	-0,371	-0,023
8	0,063	0,310	0,358	0,462	-0,373	-0,062
9	0,055	0,275	0,413	0,504	-0,428	-0,229
10	-0,011	0,146	0,494	0,515	-0,497	-0,222
11	0,044	0,336	0,589	0,585	-0,439	-0,110
12	0,028	0,513	0,701	0,534	-0,396	0,103
13	0,048	0,523	0,730	0,555	-0,389	0,155
14	0,080	0,617	0,743	0,547	-0,363	0,231
15	0,075	0,679	0,764	0,489	-0,339	0,325
16	0,081	0,731	0,759	0,448	-0,346	0,358
17	0,057	0,753	0,804	0,464	-0,382	0,411
18	0,040	0,733	0,819	0,479	-0,393	0,394
19	0,078	0,812	0,828	0,455	-0,353	0,514
20	0,084	0,825	0,833	0,464	-0,345	0,506
21	0,042	0,862	0,848	0,388	-0,368	0,574
22	0,042	0,880	0,836	0,383	-0,380	0,599
23	0,037	0,899	0,859	0,342	-0,361	0,641
24	0,032	0,907	0,877	0,351	-0,360	0,654
25	0,032	0,925	0,883	0,322	-0,341	0,688

## 5.4 *Análisis de los descriptores basados en el desbalance hidrofóbico*

Los descriptores basados en el desbalance hidrofóbico son la magnitud del vector HI, y  $ASP_{90^{\circ}, P}$  y  $ASP_{90^{\circ}, T}$ , los cuales están asociados al hemisferio hidrofóbico. Tal como para las estadísticas, se estudió la variación que muestran estos 3 descriptores en el conjunto de proteínas y el efecto que tiene en sus valores el uso de las distintas escalas de propiedades.

### 5.4.1 Variaciones en el conjunto de proteínas considerado

En general, se observó que los descriptores  $ASP_{90, P}$  y  $ASP_{90, T}$  no presentan diferencias muy elevadas en sus valores dentro del set de 8 proteínas considerado.

Para cada una de las 74 escalas de propiedades se calculó el promedio de estos descriptores para las 8 proteínas. En general, los valores no se alejan demasiado de este

promedio, encontrándose que la desviación estándar máxima es de 19,8% para  $ASP_{90,P}$  y de 24,8% para  $ASP_{90,T}$ .

En el caso del desbalance hidrofóbico HI, las diferencias en los valores obtenidos en el conjunto de proteínas son mayores que en los casos antes mencionados, observándose desviaciones estándar de entre 24 y 70% del promedio (Figura 5.16), lo que indica que este descriptor es más sensible a las diferencias presentes entre las superficies de las proteínas consideradas.

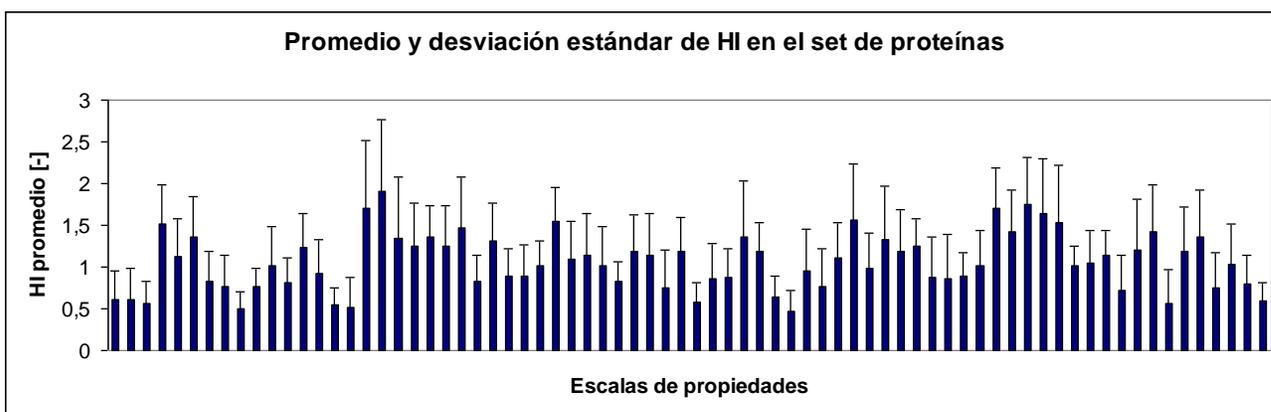
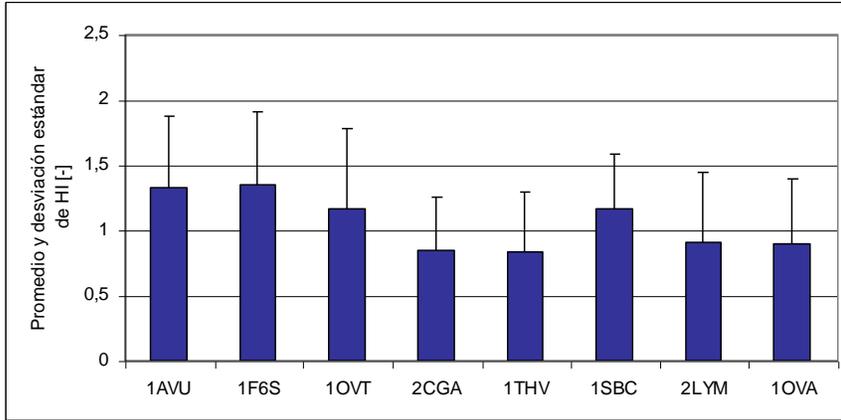


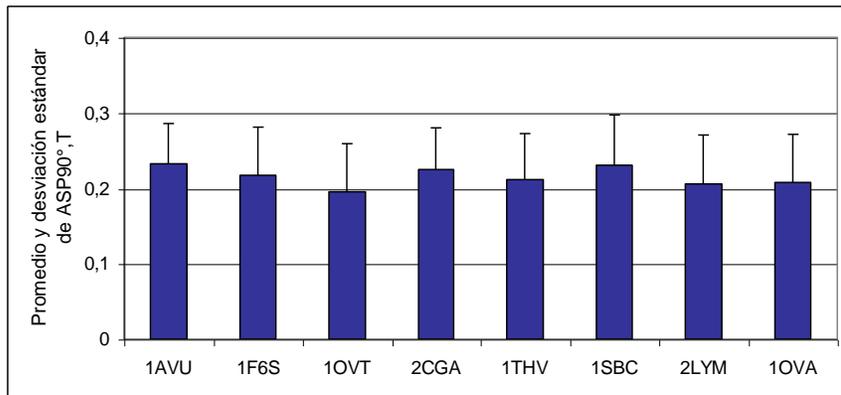
Figura 5.16: Promedio y desviación estándar de HI en el set de proteínas, para cada escala de propiedades utilizada.

#### 5.4.2 Efecto del uso de distintas escalas de propiedades

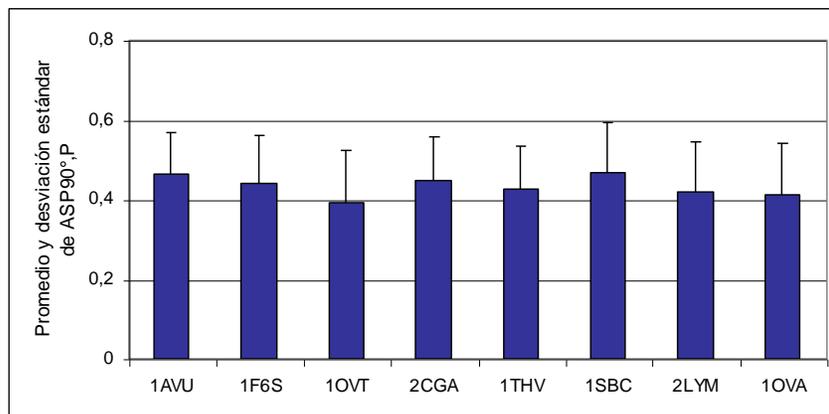
En las Figuras 5.17 a 5.19 se muestran los cambios en los descriptores al utilizar distintas escalas de propiedades. En el caso de  $ASP_{90^{\circ}, P}$  y  $ASP_{90^{\circ}, T}$  los valores para las distintas proteínas presentan desviaciones estándares de entre 22% y 33% del promedio, mientras que en el caso de HI éstas se encuentran en un rango superior de entre 36% y 59%. Esto muestra que HI también es más sensible a la escala utilizada que los otros dos descriptores.



**Figura 5.17:** Promedio y desviación estándar de HI para cada proteína, considerando el conjunto de las 74 escalas de propiedades.



**Figura 5.18:** Promedio y desviación estándar de  $ASP_{90^{\circ},T}$  para cada proteína, considerando el conjunto de las 74 escalas de propiedades.



**Figura 5.19:** Promedio y desviación estándar de  $ASP_{90^{\circ},P}$  para cada proteína, considerando el conjunto de las 74 escalas de propiedades.

### 5.4.3 Correlación entre los descriptores

En la Tabla 5.5 se muestran los coeficientes de correlación promedio entre los 3 descriptores basados en el desbalance hidrofóbico. Los descriptores  $ASP_{90^{\circ},P}$  y  $ASP_{90^{\circ},T}$  presentan una correlación relativamente alta, mientras que su correlación con HI es muy pobre.

**Tabla 5.5:** Coeficientes de correlación promedio entre  $ASP_{90^{\circ},P}$ ,  $ASP_{90^{\circ},T}$  y HI en el set de proteínas considerando los valores obtenidos para las 74 escalas utilizadas.

	$ASP_{90^{\circ},P}$	$ASP_{90^{\circ},T}$	HI
$ASP_{90^{\circ},P}$	1,00	0,87	0,16
$ASP_{90^{\circ},T}$		1,00	0,12
HI			1,00

### 5.5 Modelos de predicción del coeficiente de partición

Para predecir el coeficiente de partición se desarrollaron modelos lineales entre K y los descriptores calculados. En primera instancia, sólo se consideraron escalas asociadas a la hidrofobicidad de los aminoácidos. Para cada uno de los descriptores se seleccionaron los mejores modelos obtenidos para cada sistema, correspondientes a aquellos con los cuales se obtuvieron los  $MSE_{JK}$  más bajos, debido a que estos errores están asociados a la calidad en la predicción de K.

En las Tablas 5.8 a 5.11 se muestran los mejores resultados obtenidos para los descriptores estadísticos  $ASP_{avg}$ ,  $ASP_{std}$ ,  $ASP_{max}$  y  $ASP_{min}$ . Para el caso de los descriptores basados en el vector de desbalance hidrofóbico HI,  $ASP_{90^{\circ},P}$  y  $ASP_{90^{\circ},T}$ , estos resultados se encuentran en las Tablas 5.12 a 5.14.

Para evaluar la calidad de los modelos se hicieron dos comparaciones. En primer lugar, se compararon los resultados obtenidos con los asociados al *modelo 5*, reportados en [14]. Este modelo corresponde a la correlación entre K y la hidrofobicidad de las proteínas, medida experimentalmente, por precipitación. Los resultados obtenidos con este modelo se presentan en la Tabla 5.6. Para la segunda comparación, se utilizaron los mejores resultados obtenidos al correlacionar K con ASH. ASH corresponde al ASP global de las proteínas, calculado a partir del set de escalas de hidrofobicidad. Los resultados asociados a ASH se muestran en la Tabla 5.7.

**Tabla 5.6:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros correspondientes al modelo 5 para cada ATPS considerado.

Sistema	NaCl [% p/p]	Pearson [-]	$MSE_{JK}$ [-]	R [-]	logPo [-]
Fosfato	0	0,76	38,51	2,23	-0,27
	0,6	0,73	77,66	2,81	-0,27
	8,8	0,87	176,83	8,31	-0,44
Sulfato	0	0,38	271,51	2,20	-0,10
	0,6	0,55	214,62	3,06	-0,17
	8,8	0,92	161,51	9,76	-0,43
Citrato	0	0,88	23,18	2,58	-0,30
	0,6	0,85	77,56	4,36	-0,30
	8,8	0,89	297,53	11,19	-0,43
Dextrano	0	0,47	19,86	0,75	-0,44
	0,6	0,77	33,64	2,31	-0,39
	8,8	0,56	73,29	1,88	-0,54

**Tabla 5.7:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros de los modelos con menor  $MSE_{JK}$  asociados a ASH, para cada uno de los sistemas considerados

Sistema	NaCl [% p/p]	Escala	Propiedad	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	Cowan-Whittaker [32]	Índices de hidrofobicidad determinados en HPLC a pH 3.4	0,90	15,63	5,62	-0,30
	0,6	Welling <i>et al.</i> [50]	Valor de antigenicidad	0,82	36,73	11,23	-0,33
	8,8	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	0,96	58,67	22,37	-0,31
Sulfato	0	Welling <i>et al.</i> [50]	Valor de antigenicidad	0,70	196,42	14,38	-0,31
	0,6	Welling <i>et al.</i> [50]	Valor de antigenicidad	0,83	162,81	16,55	-0,32
	8,8	Grantham [36]	Polaridad	0,94	106,84	16,02	-0,42
Citrato	0	Rose <i>et al.</i> [47]	Área fraccional promedio perdida	0,90	12,16	5,34	-0,37
	0,6	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	0,85	58,64	10,75	-0,25
	8,8	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	0,97	85,91	29,87	-0,30
Dextrano	0	Grantham [36]	Polaridad	0,64	15,97	1,63	-0,42
	0,6	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	0,91	18,82	6,73	-0,29
	8,8	Grantham [36]	Polaridad (p)	0,72	47,99	3,87	-0,46

**Tabla 5.8:** Coeficiente de correlación, MSE<sub>JK</sub> y parámetros de los modelos con menor MSE<sub>JK</sub> obtenidos para cada sistema utilizando a ASPavg como descriptor de la hidrofobicidad.

Sistema	NaCl [% p/p]	Escala	Propiedad	Radio [Å]	Pearson [-]	MSE <sub>JK</sub> · 10 <sup>3</sup> [-]	R [-]	logPo [-]
Fosfato	0	Cowan-Whittaker [32]	Índices de hidrofobicidad determinados en HPLC a pH 3.4	11	0,89	16,45	6,33	-0,29
	0,6	Welling <i>et al.</i> [50]	Valor de antigenicidad	23	0,81	41,04	13,18	-0,33
	8,8	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	10	0,97	43,24	27,17	-0,29
Sulfato	0	Wertz-Scheraga [51]	Fracción de aminoácidos escondidos en 20 proteínas	5	0,52	173,53	8,28	-0,39
	0,6	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	5	0,74	168,33	16,37	-0,22
	8,8	Grantham [36]	Polaridad	15	0,95	88,87	18,87	-0,40
Citrato	0	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	5	0,91	12,02	10,62	-0,24
	0,6	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	5	0,92	33,06	18,73	-0,24
	8,8	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	10	0,98	60,81	36,29	-0,29
Dextrano	0	Fraga [35]	Factores de reconocimiento	9	0,55	13,15	1,25	-0,47
	0,6	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	11	0,90	19,37	7,88	-0,28
	8,8	Fraga [35]	Factores de reconocimiento	9	0,78	33,57	3,73	-0,52

**Tabla 5.9:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros de los modelos con menor  $MSE_{JK}$  obtenidos para cada sistema utilizando a ASPstd como descriptor de la hidrofobicidad.

Sistema	NaCl [% p/p]	Escala	Propiedad	Radio [Å]	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	Berggren [67]	Escala de hidrofobicidad en sistemas ATPS	25	0,78	32,72	2,36	1,38
	0,6	Berggren [67]	Escala de hidrofobicidad en sistemas ATPS	25	0,88	26,91	3,52	2,03
	8,8	Zimmerman <i>et al.</i> [54]	Polaridad	10	0,91	121,07	-17,24	-22,23
Sulfato	0	Chothia [31]	Proporción de residuos un 95% ocultos (considerando 12 proteínas)	11	0,50	181,48	2,41	2,68
	0,6	Welling <i>et al.</i> [50]	Valor de antigenicidad	6	0,90	95,83	-6,01	-5,48
	8,8	Zimmerman <i>et al.</i> [54]	Polaridad	12	0,85	280,40	-21,02	-23,77
Citrato	0	Zimmerman <i>et al.</i> [54]	Polaridad	10	0,85	22,58	-5,36	-6,45
	0,6	Zimmerman <i>et al.</i> [54]	Polaridad	10	0,96	20,08	-10,45	-12,65
	8,8	Zimmerman <i>et al.</i> [54]	Polaridad	10	0,91	226,10	-23,14	-29,60
Dextrano	0	Eisenberg <i>et al.</i> [33]	Escala de consenso normalizada de hidrofobicidad	13	0,91	4,79	-3,44	-2,88
	0,6	Roseman [48]	Hidrofobicidad	10	0,87	21,85	-4,22	-4,42
	8,8	Browne <i>et al.</i> [29]	Coeficiente de retención en HFBA	18	0,94	12,61	-6,00	-3,81

**Tabla 5.10:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros de los modelos con menor  $MSE_{JK}$  obtenidos para cada sistema utilizando a ASPmax como descriptor de la hidrofobicidad.

Sistema	NaCl [% p/p]	Escala	Propiedad	Radio [Å]	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	Welling <i>et al.</i> [50]	Valor de antigenicidad	19	0,93	9,02	2,42	-0,27
	0,6	Browne <i>et al.</i> [29]	Coeficiente de retención en TFA	10	0,88	28,79	-1,53	-0,23
	8,8	Berggren [67]	Escala de hidrofobicidad en sistemas ATPS	18	0,92	110,25	4,64	-0,65
Sulfato	0	Fraga [35]	Factores de reconocimiento	7	0,96	35,05	0,29	-0,03
	0,6	Zimmerman <i>et al.</i> [54]	Polaridad	12	0,83	166,42	-0,07	0,01
	8,8	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	22	0,84	263,04	9,82	-0,25
Citrato	0	Browne <i>et al.</i> [29]	Coeficiente de retención en HFBA	10	0,91	13,08	-1,09	-0,19
	0,6	Browne <i>et al.</i> [29]	Coeficiente de retención en HFBA	10	0,90	40,70	-1,86	-0,18
	8,8	Berggren [67]	Escala de hidrofobicidad en sistemas ATPS	18	0,90	284,10	5,84	-0,63
Dextrano	0	Browne <i>et al.</i> [29]	Coeficiente de retención en HFBA	12	0,76	10,03	-0,80	-0,20
	0,6	Welling <i>et al.</i> [50]	Valor de antigenicidad	6	0,85	27,26	-0,32	-0,10
	8,8	Browne <i>et al.</i> [29]	Coeficiente de retención en HFBA	12	0,86	24,38	-1,74	-0,18

**Tabla 5.11:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros de los modelos con menor  $MSE_{JK}$  obtenidos para cada sistema utilizando a ASPmin como descriptor de la hidrofobicidad.

Sistema	NaCl [% p/p]	Escala	Propiedad	Radio [Å]	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	Welling <i>et al.</i> [50]	Valor de antigenicidad	15	0,87	21,89	5,81	-0,39
	0,6	Rao-Argos [45]	Parámetro de hélice oculta en membrana	8	0,90	23,23	1,01	-0,82
	8,8	Zimmerman <i>et al.</i> [54]	Polaridad	8	0,98	35,99	3,41	-1,13
Sulfato	0	Wertz-Scheraga [51]	Fracción de aminoácidos escondidos en 20 proteínas	10	0,43	188,15	1,27	-0,50
	0,6	Rao-Argos [45]	Parámetro de hélice oculta en membrana	7	0,96	49,32	0,78	-0,55
	8,8	Zimmerman <i>et al.</i> [54]	Polaridad	10	0,98	28,96	2,92	-0,95
Citrato	0	Janin [40]	Energía libre de transferencia desde el interior al exterior de una proteína globular	16	0,94	7,30	3,51	-0,37
	0,6	Hopp <i>et al.</i> [39]	Hidrofilicidad	8	0,97	15,83	1,09	-1,01
	8,8	Chothia [31]	Proporción de residuos un 95% ocultos (considerando 12 proteínas)	10	0,98	48,52	17,88	-0,84
Dextrano	0	Aboderin [26]	Movilidad de aminoácidos en papel cromatográfico	16	0,89	3,69	1,64	-0,66
	0,6	Cowan-Whittaker [32]	Índices de hidrofobicidad determinados en HPLC a pH 7.5	16	0,98	2,82	3,55	-0,56
	8,8	Aboderin [26]	Movilidad de aminoácidos en papel cromatográfico	8	0,98	5,10	2,57	-1,04

**Tabla 5.12:** Coeficiente de correlación, MSE<sub>JK</sub> y parámetros de los modelos con menor MSE<sub>JK</sub> obtenidos para cada sistema utilizando a HI como descriptor de la hidrofobicidad

Sistema	NaCl [% p/p]	Escala	Propiedad	Pearson [-]	MSE <sub>JK</sub> ·10 <sup>3</sup> [-]	R [-]	logPo [-]
Fosfato	0	Meek [43]	Coeficiente de retención en HPLC a pH 2.1	0,73	39,68	-1,21	-0,12
	0,6	Meek [43]	Coeficiente de retención en HPLC a pH 2.1	0,65	62,67	-1,39	-0,14
	8,8	Welling <i>et al.</i> [50]	Valor de antigenicidad	0,76	288,23	3,00	-0,47
Sulfato	0	Roseman [48]	Escala de hidrofobicidad (pi-r)	0,64	176,95	-1,26	-0,47
	0,6	Browne <i>et al.</i> [29]	Coeficiente de retención en HFBA	0,36	206,37	-0,45	-1,81
	8,8	Hopp <i>et al.</i> [39]	Hidrofilicidad	0,68	485,58	-2,50	0,33
Citrato	0	Meek [43]	Coeficiente de retención en HPLC a pH 2.1	0,76	27,50	-1,25	-0,09
	0,6	Meek [43]	Coeficiente de retención en HPLC a pH 2.1	0,68	110,88	-1,96	-0,09
	8,8	Welling <i>et al.</i> [50]	Valor de antigenicidad	0,75	528,50	3,90	-0,43
Dextrano	0	Fauchere <i>et al.</i> [34]	Escala de hidrofobicidad (pi-r)	0,95	3,28	-0,81	0,16
	0,6	Browne <i>et al.</i> [29]	Coeficiente de retención en TFA	0,87	23,21	-0,79	-0,02
	8,8	Cowan-Whittaker [32]	Indices de hidrofobicidad determinados en HPLC a pH 7.5	0,88	23,51	-1,46	0,26

**Tabla 5.13:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros de los modelos con menor  $MSE_{JK}$  obtenidos para cada sistema utilizando a  $ASP_{90^{\circ},P}$  como descriptor de la hidrofobicidad

Sistema	NaCl [% p/p]	Escala	Propiedad	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	Miyazawa y Jerningan [44]	Escala de hidrofobicidad (Energía de contacto derivada de datos 3D)	0,83	21,93	3,78	-0,78
	0,6	Welling <i>et al.</i> [50]	Valor de antigenicidad	0,81	34,64	6,92	-0,60
	8,8	Grantham [36]	Polaridad (p)	0,80	254,52	10,97	-0,72
Sulfato	0	Janin [40]	Energía libre de transferencia desde el interior al exterior de una proteína globular	0,37	208,82	2,73	-0,36
	0,6	Welling <i>et al.</i> [50]	Valor de antigenicidad	0,92	76,98	11,41	-0,58
	8,8	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	0,79	375,13	17,03	-0,59
Citrato	0	Rose <i>et al.</i> [47]	Area fraccional promedio perdida (área promedio oculta / área en estado estándar)	0,85	17,14	4,02	-0,64
	0,6	Rose <i>et al.</i> [47]	Area fraccional promedio perdida (área promedio oculta / área en estado estándar)	0,80	69,53	6,60	-0,64
	8,8	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	0,81	404,52	20,91	-0,59
Dextrano	0	Grantham [36]	Polaridad (p)	0,49	18,45	1,13	-0,71
	0,6	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	0,82	30,43	5,04	-0,57
	8,8	Welling <i>et al.</i> [50]	Valor de antigenicidad	0,85	54,34	6,31	-0,69

**Tabla 5.14:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros de los modelos con menor  $MSE_{JK}$  obtenidos para cada sistema utilizando a  $ASP_{90^{\circ},T}$  como descriptor de la hidrofobicidad

Sistema	NaCl [% p/p]	Escala	Propiedad	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	Cowan-Whittaker [32]	Indices de hidrofobicidad determinados en HPLC a pH 3.4	0,75	37,55	4,40	-0,26
	0,6	Cowan-Whittaker [32]	Indices de hidrofobicidad determinados en HPLC a pH 7.5	0,62	67,63	4,10	-0,31
	8,8	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	0,81	235,41	19,96	-0,29
Sulfato	0	Chothia [31]	Proporción de residuos un 95% ocultos (considerando 12 proteínas)	0,26	221,91	1,98	-0,15
	0,6	Welling <i>et al.</i> [50]	Valor de antigenicidad	0,70	162,56	14,75	-0,29
	8,8	Meek [43]	Coeficiente de retención en HPLC a pH 2.1	0,85	292,86	16,69	-0,56
Citrato	0	Cowan-Whittaker [32]	Indices de hidrofobicidad determinados en HPLC a pH 7.5	0,79	24,66	4,05	-0,33
	0,6	Cowan-Whittaker [32]	Indices de hidrofobicidad determinados en HPLC a pH 7.5	0,75	84,92	6,70	-0,33
	8,8	Meek [43]	Coeficiente de retención en HPLC a pH 2.1	0,87	289,86	20,35	-0,55
Dextrano	0	Welling <i>et al.</i> [50]	Valor de antigenicidad	0,48	19,51	2,89	-0,35
	0,6	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	0,80	30,14	6,27	-0,27
	8,8	Welling <i>et al.</i> [50]	Valor de antigenicidad	0,64	61,51	8,10	-0,38

### 5.5.1 Calidad de los modelos obtenidos en la predicción de K

#### a) Descriptores estadísticos

Al comparar el modelo 5, el cual establece una relación lineal entre  $\log K$  y la hidrofobicidad de las proteínas medida por precipitación, con los modelos asociados a cada una de las estadísticas para los 12 sistemas, en casi todos los casos se observan mejores correlaciones y  $MSE_{JK}$  menores al utilizar los descriptores. La única excepción está en el sistema PEG/Sulfato con alta concentración de NaCl, donde los peores resultados se obtienen con el modelo asociado a  $ASP_{std}$ . En relación a los resultados calculados utilizando  $ASH$ , en general, también se obtiene mejores resultados con las estadísticas pero las diferencias son menores.

En términos de  $MSE_{JK}$ , las principales diferencias con el modelo 5, se observan con los resultados obtenidos con  $ASP_{min}$ , donde, en promedio los errores obtenidos con el

modelo 5 son 4,6 veces mayores. Para los casos correspondientes a ASPmax, los MSE<sub>JK</sub> asociados al modelo 5 son en promedio 1,4 veces superiores, mientras que para ASPavg y ASPstd son 1,3 veces mayores. Con respecto a ASH, con ASPavg y ASPmin se obtienen errores menores para la mayoría de los sistemas. Para ASPavg, las diferencias son bajas, mientras que para ASPmin son más marcadas, obteniéndose en promedio MSE<sub>JK</sub> 2,3 veces mayores con ASH. En relación a ASPmax y ASPstd, en una gran cantidad de sistemas se observan errores menores con ASH, sin embargo, en promedio estos siguen siendo superiores.

Con respecto al ajuste a los valores experimentales de logK, en todos los modelos correspondientes a las estadísticas se obtuvieron muy buenas correlaciones. Los coeficientes de correlación en la mayoría de los casos son altos, con valores entre 0,78 y 0,98. Además son en promedio entre 20,4% y 37,4% superiores a los obtenidos mediante el modelo 5. En comparación con los resultados obtenidos con ASH, para la mayoría de los sistemas, las correlaciones obtenidas son similares.

#### **b) Descriptores basados en el vector de desbalance hidrofóbico**

Para el caso de los descriptores calculados utilizando el vector de desbalance hidrofóbico, las correlaciones y los MSE<sub>JK</sub> obtenidos son de menor calidad que para las estadísticas. En general, se obtienen mejores resultados con el modelo 5, excepto para los sistemas de PEG/Dextrano donde las correlaciones obtenidas con los modelos de HI, ASP<sub>90°, P</sub> y ASP<sub>90°, T</sub>, fueron entre 21,7% y 56,6% superiores y los MSE<sub>JK</sub> son inferiores en los tres casos. Con respecto a ASH, en la gran mayoría de los casos se obtienen peores correlaciones y MSE<sub>JK</sub> más altos.

### **5.5.2 Resolución hidrofóbica e hidrofobicidad intrínseca de los sistemas: R y logPo**

#### **a) Descriptores estadísticos**

En las Tablas 5.8 a 5.11 se observa que el comportamiento del parámetro R cambia dependiendo de qué descriptor se utiliza para obtenerlo. Al comparar sus valores en estas tablas se pueden ver diferencias importantes en su magnitud y en su signo.

Al utilizar ASPstd y ASPmax se obtuvieron principalmente valores negativos, mientras que con ASPavg y ASPmin para todos los sistemas estos fueron positivos. En general, R muestra magnitudes menores en los modelos correspondientes a ASPmin y mayores para el caso de ASPavg. Para ASPavg, en promedio, la resolución hidrofóbica es 8,1 veces mayor que en el modelo de ASPmin.

Se ha reportado que en sistemas de dos fases acuosas, la resolución hidrofóbica en ATPS mejora al aumentar la concentración de NaCl. Este comportamiento es consistente con los resultados obtenidos con el modelo asociado a ASPavg, donde R crece con la concentración de NaCl en todos los sistemas excepto para el caso de PEG/Dextrano, como puede verse en la Figura 5.20.

Para el modelo de ASPstd, la magnitud de R también aumenta con la concentración de NaCl, sin embargo, el hecho que se observen signos negativos en este parámetro hace más difícil su interpretación. Aún así, es importante observar que en la mayoría de los casos en que se observan valores negativos de R, la escala seleccionada es el índice de polaridad de Zimmerman, de tal forma que este comportamiento podría deberse a que esta escala no es adecuada para medir la hidrofobicidad de los aminoácidos. Algo similar podría estar ocurriendo con del modelo asociado a ASPmax. En este caso, la mayoría de los valores negativos de R están asociados a 2 escalas de Browne. Ambas escalas corresponden a tiempos de retención en RP-HPLC y el coeficiente de correlación entre ellas es de 0,93. En el caso de ASPmin, todos los valores de R tienen signo positivo pero no presentan ningún comportamiento claro frente a la concentración de NaCl.

Con respecto a logPo, se obtuvieron valores negativos para todos los sistemas con los modelos asociados a las 4 estadísticas. Sin embargo, estos presentan diferencias importantes en su magnitud. En el modelo asociado a ASPstd los valores de logPo tienen en promedio una magnitud 5,3 veces mayor a sus análogos correspondientes a ASPmax, 3,4 veces superior que al utilizar ASPmin y 1,4 veces superiores que con ASPavg.

Además, al estudiar el comportamiento de la hidrofobicidad intrínseca con respecto a la concentración de NaCl en los distintos sistemas, no se observa ninguna tendencia evidente en ninguno de los modelos.

### **c) Descriptores basados en el vector de desbalance hidrofóbico**

Para los modelos obtenidos con  $ASP_{90^{\circ}, P}$  y  $ASP_{90^{\circ}, T}$ , los valores de R para todos los sistemas son de signo positivo y además aumentan con las concentraciones de NaCl tal como en el caso de ASPavg. Con respecto a logPo, los valores obtenidos con  $ASP_{90^{\circ}, P}$ , son en promedio 2 veces superiores en magnitud a los correspondientes a  $ASP_{90^{\circ}, T}$ . Además es importante notar que, en ambos casos tienen valores negativos, indicando la preferencia de las proteínas por la fase superior de los sistemas.

El comportamiento de los parámetros del modelo asociado a HI no es tan claro. Tanto R como logPo presentan variabilidad en su signo para los distintos sistemas. El hecho que R presente distintos signos para distintos sistemas hace que no sea posible determinar una relación clara entre HI y K. Este parámetro acompaña al valor del descriptor en el modelo utilizado. Así, la presencia de variaciones en su signo lleva a resultados

contradictorios, indicando que la relación entre K y HI es inversa en algunos sistemas y directa en otros. Además no se ve un claro aumento de R con la concentración de NaCl, sin embargo, en todos los sistemas R tiene su mayor magnitud en los casos de alta presencia de NaCl. Estas razones indican que el modelo asociado a HI no es adecuado para predecir la partición de proteínas en ATPS.

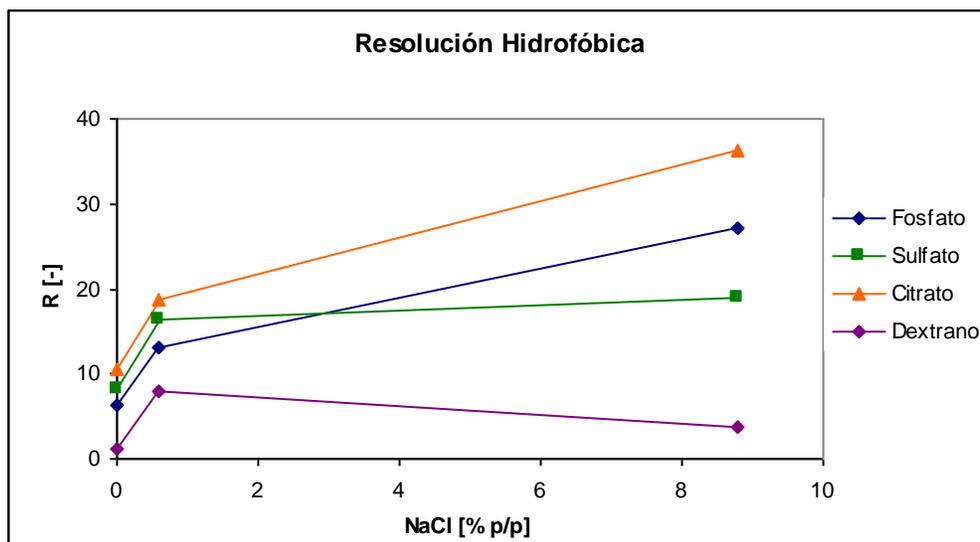


Figura 5.20: Resolución hidrofóbica obtenida con ASPavg para los 12 sistemas de dos fases acuosas considerados

### 5.5.3 Radios seleccionados por los descriptores estadísticos

Los radios que se muestran en las Tablas 5.8 a 5.11 son aquellos que definen el tamaño del vecindario alrededor de los aminoácidos de la superficie de las proteínas. Estos vecindarios son los que se consideran al calcular los ASP locales a partir de los cuales se obtienen los valores de los descriptores estadísticos.

Al comparar sus valores se puede ver que no existe un consenso en la selección de radios entre los modelos asociados a los descriptores estadísticos.

Para ASPstd en la mayoría de los sistemas los radios se encuentran entre 10 y 13 Å mientras que para el resto de los descriptores estadísticos el rango de radios seleccionados es más amplio.

Con respecto a los 12 sistemas estudiados, ASPprom, ASPdest, ASPmax y ASPmin seleccionan radios variados, siendo los sistemas de PEG/fosfato con baja concentración de NaCl y PEG/dextrano con alta concentración de NaCl los que presentan mayores variaciones.

El hecho de que exista tal variedad en los radios seleccionados por los distintos modelos indica que éstos no son capaces de seleccionar un radio único óptimo para todos

los sistemas y que es necesario recurrir a valores distintos para cada sistema para obtener los mejores modelos.

Sin embargo, es interesante notar que en varios casos existe una relación entre la escala y radio utilizado. Con ASPavg, para la mayoría de los sistemas se seleccionaron escalas de coeficientes de retención en HPLC desarrolladas por Meek y en todos estos casos el radio seleccionado fue de 5 o 10 Å. Lo mismo ocurre para ASPstd, donde en todos los casos en que la escala seleccionada es la de Zimmerman, el radio utilizado es de 10 Å. En el modelo correspondiente a ASPmax esto se observa en los casos en que se seleccionaron las escalas de Browne donde los radios siempre están entre 10 y 12 Å.

#### **5.5.4 Predicción de K utilizando distintas escalas de propiedades**

Los resultados que se muestran en las Tablas 5.8 a 5.14 están asociados a descriptores calculados teniendo en cuenta sólo escalas asociadas a la hidrofobicidad de los aminoácidos. Además, todos los descriptores fueron calculados utilizando 37 escalas más, correspondientes a otras propiedades. Estos resultados también se ajustaron a los valores experimentales de  $\log K$ . Así, es posible determinar si los modelos asociados a los distintos descriptores seleccionan escalas de hidrofobicidad, es decir si se obtienen  $MSE_{JK}$  menores y mejores correlaciones con estas escalas.

Las Tablas 5.15 a 5.21 muestran los mejores modelos obtenidos con los 7 descriptores utilizando todo el conjunto de escalas de propiedades.

**Tabla 5.15:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros correspondientes al modelo asociado a ASPavg, obtenidos utilizando 74 escalas de propiedades.

Sistema	NaCl [% p/p]	Escala	Propiedad	Radio [Å]	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	Cowan-Whittaker [32]	Índices de hidrofobicidad determinados en HPLC a pH 3.4	11	0,89	16,45	6,33	-0,29
	0,6	Welling <i>et al.</i> [50]	Valor de antigenicidad	23	0,81	41,04	13,18	-0,33
	8,8	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	10	0,97	43,24	27,17	-0,29
Sulfato	0	Hellberg <i>et al.</i> [38]	Análisis estadístico de propiedades aminoacídicas	5	0,89	89,23	-37,47	-0,39
	0,6	Varios textos	Número de codones	6	0,94	48,96	7,96	-0,25
	8,8	Levitt [62]	Frecuencia normalizada para hoja beta	15	0,97	71,46	20,86	-0,50
Citrato	0	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	5	0,91	12,02	10,62	-0,24
	0,6	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	5	0,92	33,06	18,73	-0,24
	8,8	Meek [43]	Coeficiente de retención en HPLC a pH 7.4	10	0,98	60,81	36,29	-0,29
Dextrano	0	Fraga [35]	Factores de reconocimiento	9	0,55	13,15	1,25	-0,47
	0,6	Levitt [62]	Frecuencia normalizada para hoja beta	25	0,89	18,48	5,13	-0,48
	8,8	Fraga [35]	Factores de reconocimiento	9	0,78	33,57	3,73	-0,52

**Tabla 5.16 :** Coeficiente de correlación,  $MSE_{JK}$  y parámetros correspondientes al modelo asociado a ASPstd, obtenidos utilizando 74 escalas de propiedades.

Sistema	NaCl [% p/p]	Escala	Propiedad	Radio [Å]	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	Dayhoff [58]	Mutabilidad relativa	11	0,90	15,11	-2,93	-1,17
	0,6	Bhaskaran-Ponnuswamy [56]	Indice de flexibilidad promedio	8	0,94	12,73	-6,69	-0,98
	8,8	Zimmerman <i>et al.</i> [54]	Polaridad	10	0,91	121,07	-22,23	-0,78
Sulfato	0	Dayhoff [58]	Mutabilidad relativa	22	0,94	32,88	-3,02	-1,75
	0,6	Bhaskaran-Ponnuswamy [56]	Indice de flexibilidad promedio	8	0,90	89,86	-9,33	-1,01
	8,8	Zimmerman <i>et al.</i> [54]	Polaridad	12	0,85	280,40	-23,77	-0,88
Citrato	0	Bhaskaran-Ponnuswamy [56]	Indice de flexibilidad promedio	8	0,88	16,37	-4,81	-0,98
	0,6	Zimmerman <i>et al.</i> [54]	Polaridad	10	0,96	20,08	-12,65	-0,83
	8,8	Zimmerman <i>et al.</i> [54]	Polaridad	10	0,91	226,10	-29,60	-0,78
Dextrano	0	Eisenberg <i>et al.</i> [33]	Escala de concenso normalizada de hidrofobicidad	13	0,91	4,79	-2,88	-1,19
	0,6	Jones [60]	Refractividad	8	0,91	15,60	3,50	-1,08
	8,8	Browne <i>et al.</i> [29]	Coeficiente de retención en HFBA	18	0,94	12,61	-3,81	-1,57

**Tabla 5.17:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros correspondientes al modelo asociado a ASPmax, obtenidos utilizando 74 escalas de propiedades.

Sistema	NaCl [% p/p]	Escala	Propiedad	Radio [Å]	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	Welling <i>et al.</i> [50]	Valor de antigenicidad	19	0,93	9,02	8,85	-0,27
	0,6	Sandberg <i>et al.</i> [65]	Análisis estadístico de propiedades aminoacídicas z4	8	0,94	13,03	-14,18	-0,18
	8,8	Deleage y Roux [59]	Parámetro conformacional de alfa hélice	23	0,96	68,31	-19,92	-0,20
Sulfato	0	Fraga [35]	Factores de reconocimiento	7	0,96	35,05	8,92	-0,03
	0,6	Varios textos	Número de codones	9	0,93	45,44	8,69	-0,01
	8,8	Jones [60]	Refractividad	10	0,96	71,77	-29,08	-0,22
Citrato	0	Deleage y Roux [59]	Parámetro conformacional de alfa hélice	8	0,92	11,36	-5,87	-0,13
	0,6	Deleage y Roux [59]	Parámetro conformacional de alfa hélice	8	0,94	23,11	-10,53	-0,12
	8,8	Deleage y Roux [59]	Parámetro conformacional de alfa hélice	12	0,94	146,80	-43,31	-0,14
Dextrano	0	Jonsson <i>et al.</i> [68]	Análisis estadístico de propiedades aminoacídicas z3	9	0,86	6,40	-4,00	-0,12
	0,6	Jones [60]	Refractividad	10	0,98	4,44	-8,45	-0,24
	8,8	Jones [60]	Refractividad	12	0,85	24,34	-10,19	-0,25

**Tabla 5.18:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros correspondientes al modelo asociado a ASPmin, obtenidos utilizando 74 escalas de propiedades.

Sistema	NaCl [% p/p]	Escala	Propiedad	Radio [Å]	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	Welling <i>et al.</i> [50]	Valor de antigenicidad	15	0,87	21,89	5,81	-0,39
	0,6	Rao-Argos [45]	Parámetro de hélice oculta en membrana	8	0,90	23,23	1,01	-0,82
	8,8	Janin [40]	Fracción molar de 2001 residuos ocultos	9	0,98	32,89	5,16	-1,11
Sulfato	0	Janin [40]	Fracción molar de 3220 residuos accesibles	7	0,52	170,31	-1,43	-1,18
	0,6	Rao-Argos [45]	Parámetro de hélice oculta en membrana	7	0,96	49,32	0,78	-0,55
	8,8	Zimmerman <i>et al.</i> [54]	Polaridad	10	0,98	28,96	2,92	-0,95
Citrato	0	Janin [40]	Energía libre de transferencia desde el interior al exterior de una proteína globular	16	0,94	7,30	3,51	-0,37
	0,6	Hopp <i>et al.</i> [39]	Hidrofilicidad	8	0,97	15,83	1,09	-1,01
	8,8	Chothia [31]	Proporción de residuos un 95% ocultos (considerando 12 proteínas)	10	0,98	48,52	17,88	-0,84
Dextrano	0	Aboderin [26]	Movilidad de aminoácidos en papel cromatográfico	16	0,89	3,69	1,64	-0,66
	0,6	Cowan-Whittaker [32]	Índices de hidrofobicidad determinados en HPLC a pH 7.5	16	0,98	2,82	3,55	-0,56
	8,8	Varios textos	Número de codones	22	0,98	2,58	3,26	-0,53

**Tabla 5.19:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros correspondientes al modelo asociado a HI, obtenidos utilizando 74 escalas de propiedades.

Sistema	NaCl [% p/p]	Escala	Propiedad	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	Meek [43]	Coeficiente de retención en HPLC a pH 2.1	0,73	39,68	-1,21	-0,12
	0,6	Meek [43]	Coeficiente de retención en HPLC a pH 2.1	0,65	62,67	-1,39	-0,14
	8,8	Welling <i>et al.</i> [50]	Valor de antigenicidad	0,76	288,23	3,00	-0,47
Sulfato	0	Dayhoff [58]	Mutabilidad relativa	0,79	137,43	-1,31	-0,59
	0,6	Rose <i>et al.</i> [47]	Área promedio oculta en la transferencia desde el estado estándar al de proteína plegada	0,40	189,77	0,86	0,57
	8,8	Hopp <i>et al.</i> [39]	Hidrofilicidad	0,68	485,58	-2,50	0,33
Citrato	0	Meek [43]	Coeficiente de retención en HPLC a pH 2.1	0,76	27,50	-1,25	-0,09
	0,6	Meek [43]	Coeficiente de retención en HPLC a pH 2.1	0,68	110,88	-1,96	-0,09
	8,8	Libros de texto	Número de codones	0,77	495,33	3,87	-0,15
Dextrano	0	Fauchere <i>et al.</i> [34]	Escala de hidrofobicidad	0,95	3,28	-0,81	0,16
	0,6	Browne <i>et al.</i> [29]	Coeficiente de retención en TFA	0,87	23,21	-0,79	-0,02
	8,8	Hellberg <i>et al.</i> [38]	Análisis estadístico de propiedades aminoacídicas	0,86	22,64	-1,41	0,35

**Tabla 5.20:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros correspondientes al modelo asociado a  $ASP_{90^\circ, P}$ , obtenidos utilizando 74 escalas de propiedades.

Sistema	NaCl [% p/p]	Escala	Propiedad	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	Miyazawa-Jerningan [44]	Escala de hidrofobicidad (Energía de contacto derivada de datos 3D)	0,83	21,93	3,78	-0,78
	0,6	Libros de texto	Número de codones	0,87	28,14	5,36	-0,55
	8,8	Lifson-Sander [63]	Preferencia conformacional para hoja beta antiparalela	0,88	148,48	14,73	-0,74
Sulfato	0	Deleage-Roux [59]	Parámetro conformacional de vuelta beta	0,85	108,35	-15,91	-0,61
	0,6	Welling <i>et al.</i> [50]	Valor de antigenicidad	0,92	76,98	11,41	-0,58
	8,8	Levitt [62]	Frecuencia normalizada para hoja beta	0,87	209,09	10,24	-0,79
Citrato	0	Lifson-Sander [63]	Preferencia conformacional para hoja beta total	0,90	13,49	5,20	-0,73
	0,6	Lifson-Sander [63]	Preferencia conformacional para hoja beta total	0,85	57,78	8,56	-0,73
	8,8	Lifson-Sander [63]	Preferencia conformacional para hoja beta antiparalela	0,89	260,24	19,97	-0,73
Dextrano	0	Janin [40]	Fracción molar de 3220 residuos accesibles	0,40	14,40	-1,43	-0,48
	0,6	Levitt [62]	Frecuencia normalizada para hoja beta	0,85	23,20	2,85	-0,75
	8,8	Libros de texto	Número de codones	0,85	27,80	4,63	-0,67

**Tabla 5.21:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros correspondientes al modelo asociado a  $ASP_{90^\circ, T}$ , obtenidos utilizando 74 escalas de propiedades.

Sistema	NaCl [% p/p]	Escala	Propiedad	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	Lifson-Sander [63]	Preferencia conformacional por hoja beta paralela	0,89	15,99	3,82	-0,58
	0,6	Levitt [62]	Frecuencia normalizada para alfa hélice	0,78	46,96	-4,91	-0,31
	8,8	Deleage-Roux [59]	Parámetro conformacional de alfa hélice	0,89	164,81	-17,30	-0,22
Sulfato	0	Grantham [36]	Razón de peso atómico de elementos hetero y carbonos en la cadena lateral	0,83	125,43	-18,45	-0,57
	0,6	Libros de texto	Número de codones	0,87	127,63	7,01	-0,21
	8,8	Levitt [62]	Frecuencia normalizada para hoja beta	0,92	135,53	11,10	-0,50
Citrato	0	Levitt [62]	Frecuencia normalizada para hoja beta	0,83	22,69	2,77	-0,38
	0,6	Deleage-Roux [59]	Parámetro conformacional de alfa hélice	0,82	64,23	-8,64	-0,29
	8,8	Chou-Pasman [57]	Parámetro conformacional para hoja beta calculados desde 29 proteínas	0,90	224,05	18,67	-0,40
Dextrano	0	Deleage-Roux [59]	Parámetro conformacional de vuelta beta	0,50	14,09	-3,92	-0,25
	0,6	Levitt [62]	Frecuencia normalizada para hoja beta	0,87	22,32	2,97	-0,46
	8,8	Libros de texto	Número de codones	0,92	29,90	4,49	-0,36

Los mejores resultados asociados a  $ASP_{avg}$ , en su mayoría se obtuvieron utilizando escalas de hidrofobicidad. La escala más seleccionada en este caso es la de Meek [43], la cual fue escogida para todos los sistemas de PEG/Citrato y para el sistema de PEG/Fosfato con alta concentración de NaCl.

En el caso de  $ASP_{std}$ , la naturaleza de las escalas seleccionadas para los distintos sistemas es más variada. Sólo en 6 de los 12 sistemas considerados se obtuvieron los mejores resultados con escalas asociadas a la hidrofobicidad, siendo la escala de Zimmerman [54] la más utilizada. Entre las otras escalas de propiedades seleccionadas, las con mayor frecuencia son el índice de mutabilidad relativa de Bhaskaran y Ponnuswamy [56] y la escala de mutabilidad relativa de Dayhoff [58] las cuales fueron utilizadas en los casos de concentración de NaCl baja y nula respectivamente, en los sistemas de PEG/Fosfato y PEG/Sulfato.

Para  $ASP_{max}$ , sólo se seleccionó una escala de hidrofobicidad en el caso del sistema de PEG/Fosfato sin NaCl, para el cual se escogió la escala de Welling [50]. Para el resto de los sistemas las escalas seleccionadas con más frecuencia son las de Deleage y Roux [59] y la escala de refractividad de Jones [60].

Además es importante notar que para ASPstd y ASPmax la calidad de predicción de K disminuye bastante al considerar sólo escalas de hidrofobicidad. En promedio los  $MSE_{JK}$  aumentan en un 26% para el caso de ASPstd y son 1,2 veces superiores para ASPmax.

El modelo correspondiente a ASPmin seleccionó principalmente escalas de hidrofobicidad, pero entre ellas se observa una gran variedad. La única escala seleccionada para más de un sistema es la de Rao y Argos [69], la cual fue escogida para los sistemas de PEG/Fosfato y PEG/Sulfato con baja concentración de NaCl.

Los modelos obtenidos con  $ASP_{90^\circ, T}$  y  $ASP_{90^\circ, P}$  en la gran mayoría de los casos no seleccionaron escalas de hidrofobicidad. Las escalas seleccionadas por estos descriptores principalmente corresponden a índices de frecuencia de aminoácidos en distintos tipos de estructuras secundarias.

Para el caso de HI, al considerar las 74 escalas de propiedades, en general se mantiene la preferencia por los índices de hidrofobicidad. Los únicos sistemas en los que se obtienen mejores resultados con otras escalas son los de PEG/Sulfato con concentraciones nula y baja de NaCl y PEG/Citrato y PEG/Dextrano con concentración de NaCl alta.

### 5.5.5 Comparación de los descriptores

Los descriptores basados en el desbalance hidrofóbico en general generan modelos con  $MSE_{JK}$  más altos que con los descriptores estadísticos y peores correlaciones, con coeficientes promedio de entre 0,74 y 0,83. Además para los casos correspondientes a  $ASP_{90^\circ, P}$  y  $ASP_{90^\circ, T}$  prácticamente no se seleccionaron escalas de hidrofobicidad. Al utilizar HI, se seleccionaron escalas asociadas a esta propiedad pero el modelo obtenido no permite establecer una relación clara entre este descriptor y K. Estos factores indican que como descriptores asociados a la hidrofobicidad en la superficie de las proteínas, estos no son suficientes para establecer la relación entre esta propiedad y el comportamiento de las proteínas en ATPS. En el caso de  $ASP_{90^\circ, P}$  y  $ASP_{90^\circ, T}$  esto puede explicarse debido al nivel de información asociado a estos descriptores. Ellos están asociados principalmente a un sector de la superficie de las proteínas correspondiente al hemisferio hidrofóbico. Sin embargo, en ATPS las proteínas se encuentran inmersas, de tal forma que toda su superficie interactúa con el sistema.

El descriptor con el que se obtuvo el mejor nivel de predicción, tanto al considerar todas las escalas de propiedades como al utilizar sólo el subconjunto correspondiente a la hidrofobicidad fue ASPmin. Además en la mayoría de los sistemas, al utilizar este descriptor los mejores resultados se obtuvieron utilizando escalas de hidrofobicidad. Sin

embargo, este descriptor entrega información con respecto a la zona de la superficie de las proteínas donde la hidrofobicidad es menor. Este hecho hace difícil el considerarlo como un descriptor representativo del comportamiento de esta propiedad sobre toda la superficie proteica, especialmente al utilizarlo para describir la relación que tiene esta propiedad con el coeficiente K.

En el caso de ASPavg, sus valores pueden entregar información que está más relacionada al comportamiento de la hidrofobicidad en toda la superficie de las proteínas debido a que corresponde a un valor promedio de los ASP locales. El modelo generado utilizando ASPavg se comporta de manera satisfactoria a pesar de que las correlaciones y los  $MSE_{JK}$  obtenidos no son los mejores. Además con este descriptor para la mayoría de los sistemas se seleccionaron escalas de hidrofobicidad y principalmente se utilizó una de escala de Meek de coeficientes de retención en HPLC.

Al considerar los mejores resultados obtenidos con esta escala se observa que existe una relación entre el radio utilizado y la concentración de NaCl en sistemas del tipo PEG/sal (Tabla 5.22). El modelo selecciona radios de 5 Å para todos los sistemas del tipo PEG/sal con concentraciones nulas y bajas de NaCl y radios de entre 10 y 12 Å para altas concentraciones de sal. Para el caso del sistema PEG/Dextrano los radios seleccionados son de 5 Å para los casos con concentración alta y nula de NaCl y 11 para el caso de concentración baja. Esto muestra que la selección de los radios por este modelo no se debe necesariamente al azar sino que puede tener relación con un ajuste al nivel de NaCl presente en los sistemas.

Además, al comparar las correlaciones y los  $MSE_{JK}$  obtenidos con el modelo asociado a ASPavg utilizando la escala de Meek con los correspondientes a ASH seleccionando la misma escala (Tabla 5.23) se observa que en casi todos los sistemas los resultados logrados con ASPavg son de mejor calidad.

Teniendo en consideración los puntos anteriores y al buen comportamiento de la resolución hidrofóbica en este modelo, es posible concluir que ASPavg es útil para una descripción de la hidrofobicidad en la superficie de las proteínas y como medio para estudiar la relación entre esta propiedad y el coeficiente K.

**Tabla 5.22:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros correspondientes al modelo asociado a ASPavg, obtenidos utilizando la escala de Meek de coeficientes de retención en HPLC a pH 7.4

Sistema	NaCl [% p/p]	Radio [A]	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	5	0,69	37,58	8,13	-0,23
	0,6	5	0,80	49,93	12,13	-0,23
	8,8	10	0,97	43,24	27,17	-0,29
Sulfato	0	5	0,46	279,12	10,39	-0,20
	0,6	5	0,74	168,33	16,37	-0,22
	8,8	12	0,94	131,65	28,21	-0,29
Citrato	0	5	0,91	12,02	10,62	-0,24
	0,6	5	0,92	33,06	18,73	-0,24
	8,8	10	0,98	60,81	36,29	-0,29
Dextrano	0	5	0,56	17,43	3,49	-0,27
	0,6	11	0,90	19,37	7,88	-0,28
	8,8	5	0,75	44,25	9,98	-0,29

**Tabla 5.23:** Coeficiente de correlación,  $MSE_{JK}$  y parámetros correspondientes al modelo asociado a ASH, obtenidos utilizando la escala de Meek de coeficientes de retención en HPLC a pH 7.4

Sistema	NaCl [% p/p]	Pearson [-]	$MSE_{JK} \cdot 10^3$ [-]	R [-]	logPo [-]
Fosfato	0	0,43	52,49	3,11	-0,21
	0,6	0,60	76,69	5,66	-0,23
	8,8	0,96	58,67	22,37	-0,31
Sulfato	0	0,21	282,64	2,95	-0,08
	0,6	0,55	196,60	7,47	-0,20
	8,8	0,93	149,65	24,18	-0,31
Citrato	0	0,76	29,63	5,44	-0,25
	0,6	0,85	58,64	10,75	-0,25
	8,8	0,97	85,91	29,87	-0,30
Dextrano	0	0,49	22,71	1,90	-0,31
	0,6	0,91	18,82	6,73	-0,29
	8,8	0,70	62,16	5,73	-0,33

## 6 CONCLUSIONES

- Al tener en cuenta los buenos resultados obtenidos al correlacionar el coeficiente de partición  $K$  con los descriptores de la distribución de la hidrofobicidad en la superficie de las proteínas utilizados, se concluye que esta distribución es un factor de relevancia en el comportamiento de proteínas en ATPS. Con respecto a la predicción del coeficiente de partición  $K$ , se estableció que los descriptores antes mencionados son de utilidad en esta tarea y que es posible obtener resultados de buena calidad mediante el uso de modelos lineales entre estos y el coeficiente  $K$ .
- Se encontraron muy buenas correlaciones entre los descriptores estadísticos y  $K$ . Los coeficientes de correlación son en promedio 0,84 para  $ASP_{avg}$ , 0,87 para  $ASP_{max}$  y 0,91 para  $ASP_{std}$  y  $ASP_{min}$ . En el caso de los descriptores basados en el desbalance hidrofóbico las correlaciones obtenidas no fueron tan altas. Para el coeficiente de partición promedio para  $ASP_{90^\circ, T}$  y  $HI$  son de 0,72 mientras que para  $ASP_{90^\circ, P}$ , este es de 0,76.
- La predicción del coeficiente de partición obtenida también fue de buena calidad. Los resultados se compararon con los correspondientes al modelo 5 desarrollado en [14], donde se correlaciona con  $K$  la hidrofobicidad de las proteínas, medida por precipitación. En todos los sistemas se obtuvieron mejores resultados con los descriptores. Los  $MSE_{JK}$  asociados al modelo 5 en promedio son 2,2 veces superiores a los de las estadísticas y 30% mayores que los correspondientes a los descriptores basados en el desbalance hidrofóbico. Además se hizo una comparación con los resultados obtenidos utilizando  $ASH$ , la hidrofobicidad promedio superficial. En este caso, en su mayoría, los descriptores estadísticos presentaron errores más bajos, pero las diferencias fueron menores que con el modelo 5, siendo los  $MSE_{JK}$  correspondientes a  $ASH$  un 85% superiores. Los menores errores en la predicción de  $K$  se obtuvieron con  $ASP_{min}$ .
- Al utilizar los descriptores basados en el desbalance hidrofóbico, en general, se obtuvieron modelos de menor calidad que con las estadísticas en términos de las correlaciones y  $MSE_{JK}$  obtenidos. En promedio, los coeficientes de correlación obtenidos con  $HI$ ,  $ASP_{90^\circ, P}$  y  $ASP_{90^\circ, T}$  corresponden al 84,5% de los valores obtenidos con las estadísticas y los  $MSE_{JK}$  son 2 veces mayores.
- Se estudió la relación entre  $K$  y los descriptores calculados con escalas de propiedades no asociadas a la hidrofobicidad. Se observó que los mejores resultados en términos de  $MSE_{JK}$  no siempre están asociados a escalas de hidrofobicidad.
- Se seleccionó un descriptor entre los 7 utilizados, teniendo en cuenta la calidad del modelo obtenido al correlacionarlo con  $K$ . Este descriptor fue  $ASP_{avg}$ , el promedio de los  $ASP$  locales. Su selección se hizo en base al buen comportamiento que se observa con este modelo en la resolución hidrofóbica en relación al nivel de la concentración de  $NaCl$  en los ATPS.

- Para este descriptor en la mayoría de los casos se seleccionó una de las escalas de Meek. Al utilizar esta escala se encontró una relación entre los radios seleccionados y el nivel de NaCl en los ATPS del tipo PEG/sal. Al comparar los mejores resultados con esa escala se observó que ASPavg selecciona radios de 5 Å para todos los sistemas del tipo PEG/sal con concentraciones nulas y bajas de NaCl y radios de entre 10 y 12 Å para altas concentraciones de sal. Para el caso del sistema PEG/Dextrano los radios seleccionados son de 5 Å para los casos con concentración alta y nula de NaCl y 11 para el caso de concentración baja.

## 7 BIBLIOGRAFÍA

1. Rito-Palomares, M., *Practical application of aqueous two-phase partition to process development for the recovery of biological products*. J Chromatogr B Analyt Technol Biomed Life Sci, 2004. **807**(1): p. 3-11.
2. Gautam, M. and Simon, L., *Partitioning of  $\beta$ -glucosidase from Trichoderma reesei in poly(ethylene glycol) and potassium phosphate aqueous two-phase systems: Influence of pH and temperature* Biochem. Eng. J., 2006. **30**(1): p. 104-108.
3. Carlsson, M., Linse, P. and Tjerneld, F., *Temperature-Dependent Protein Partitioning in Two-Phase Aqueous Polymer Systems*. Macromolecules, 1993. **26**: p. 1546-1554.
4. Eiteman, M.A., *Predicting partition coefficients of multi-charged solutes in aqueous two-phase systems*. J Chromatogr A, 1994. **668**: p. 21-30.
5. Eiteman, M.A. and Gainer, J.L., *Partition of isomeric dipeptides in poly(ethylene glycol)/magnesium sulfate aqueous two-phase systems*. Biochim Biophys Acta, 1991. **1073**(3): p. 451-5.
6. Johansson, G., Albertsson, P.-A. and Tjerneld, F., *Aqueous Two-Phase Separations*, in *Separation Processes in Biotechnology*, J.A. Asenjo, Editor. 1990, CRC Press. p. 287-328.
7. Lin, D.-Q., Wu, Y.-T., Mei, L.-H., Zhu, Z.-Q. and Yao, S.-J., *Modeling the protein partitioning in aqueous polymer two-phase systems: influence of polymer concentration and molecular weight*. Chem. Eng. Sci., 2003. **58**(13): p. 2963-2972.
8. Trindade, I.P., Diogo, M.M., Prazeres, D.M. and Marcos, J.C., *Purification of plasmid DNA vectors by aqueous two-phase extraction and hydrophobic interaction chromatography*. J Chromatogr A, 2005. **1082**(2): p. 176-84.
9. Pico, G., Romanini, D., Nerli, B. and Farruggia, B., *Polyethyleneglycol molecular mass and polydispersivity effect on protein partitioning in aqueous two-phase systems*. J Chromatogr B Analyt Technol Biomed Life Sci, 2006. **830**(2): p. 286-92.
10. Pessôa, P.A. and Mohamed, R.S., *Thermodynamic modeling of the partitioning of biomolecules in aqueous two-phase systems using a modified Flory-Huggins equation*. Proc Biochem, 2004. **39**(12): p. 2075-2083.
11. Albertsson, P.-A., *Partition of Cell Particles and Macromolecules*. 1986, New York: Wiley.
12. Eiteman, M.A., Hassinen, C. and Veide, A., *A mathematical model to predict the partitioning of peptides and peptide-modified proteins in aqueous two-phase systems*. Biotechnol Prog, 1994. **10**(5): p. 513-9.
13. Kepka, C., Collet, E., Roos, F., Tjerneld, F. and Veide, A., *Two-step recovery process for tryptophan tagged cutinase: interfacing aqueous two-phase extraction and hydrophobic interaction chromatography*. J Chromatogr A, 2005. **1075**(1-2): p. 33-41.
14. Andrews, B.A., Schmidt, A.S. and Asenjo, J.A., *Correlation for the partition behavior of proteins in aqueous two-phase systems: effect of surface hydrophobicity and charge*. Biotechnol Bioeng, 2005. **90**(3): p. 380-90.

15. Madeira, P., Xu, X., Teixeira, J. and Macedo, E., *Prediction of protein partition in polymer/salt aqueous two-phase systems using the modified Wilson model*. *Biochem Eng J*, 2004. **24**(2): p. 147-155.
16. Hachem, F., Andrews, B. and Asenjo, J., *Hydrophobic partitioning of proteins in aqueous two-phase systems*. *Enzyme Microb. Technol.*, 1996. **19**(7): p. 507-517
17. Asenjo, J.A., Schmidt, A.S., Hachem, F. and Andrews, B.A., *Model for predicting the partition behaviour of proteins in aqueous two-phase systems*. *J Chromatogr*, 1994. **668**: p. 47-54.
18. Eiteman, M.A. and Gainer, J.L., *Predicting partition coefficients in polyethylene glycol-potassium phosphate aqueous two-phase systems*. *J Chromatogr*, 1991. **586**: p. 341-346.
19. Olivera-Nappa, A., Lagomarsino, G., Andrews, B.A. and Asenjo, J.A., *Effect of electrostatic energy on partitioning of proteins in aqueous two-phase systems*. *J Chromatogr B Analyt Technol Biomed Life Sci*, 2004. **807**(1): p. 81-6.
20. Pitzer, K.S., *Thermodynamics of electrolytes. I. Theoretical basis and general equations*. *J. Phys. Chem.*, 1973. **77**(2): p. 268-277.
21. Wilson, G.M., *Vapor-Liquid Equilibrium. XI. A New Expression for the Excess Free Energy of Mixing*. *J. Am. Chem. Soc.*, 1964. **86**(2): p. 127-130.
22. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., *The Protein Data Bank*. *Nucleic Acids Res*, 2000. **28**(1): p. 235-42.
23. Salgado, J.C., Rapaport, I. and Asenjo, J.A., *Predicting the behaviour of proteins in hydrophobic interaction chromatography. 1: Using the hydrophobic imbalance (HI) to describe their surface amino acid distribution*. *J Chromatogr A*, 2006. **1107**(1-2): p. 110-9.
24. Salgado, J.C., Rapaport, I. and Asenjo, J.A., *Predicting the behaviour of proteins in hydrophobic interaction chromatography. 2. Using a statistical description of their surface amino acid distribution*. *J Chromatogr A*, 2006. **1107**(1-2): p. 120-9.
25. Frishman, D. and Argos, P., *Knowledge-Based Protein Secondary Structure Assignment* *Protein Struct Funct Genet*, 1995. **23**: p. 566-579.
26. Aboderin, A.A., *An empirical hydrophobicity scale for alpha-amino-acids and some of its applications*. *Int. J. Biochem*, 1971. **2**: p. 537-544.
27. Abraham, D.J. and Leo, A.J., *Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients*. *Proteins*, 1987. **2**(2): p. 130-52.
28. Black, S.D. and Mould, D.R., *Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications*. *Anal Biochem*, 1991. **193**(1): p. 72-82.
29. Browne, C.A., Bennett, H.P. and Solomon, S., *The isolation of peptides by high-performance liquid chromatography using predicted elution positions*. *Anal Biochem*, 1982. **124**(1): p. 201-8.
30. Bull, H.B. and Breese, K., *Surface tension of amino acid solutions: a hydrophobicity scale of the amino acid residues*. *Arch Biochem Biophys*, 1974. **161**(2): p. 665-70.
31. Chothia, C., *The nature of the accessible and buried surfaces in proteins*. *J Mol Biol*, 1976. **105**(1): p. 1-12.

32. Cowan, R. and Whittaker, R.G., *Hydrophobicity indices for amino acid residues as determined by high-performance liquid chromatography*. Pept Res, 1990. **3**(2): p. 75-80.
33. Eisenberg, D., Schwarz, E., Komaromy, M. and Wall, R., *Analysis of membrane and surface protein sequences with the hydrophobic moment plot*. J Mol Biol, 1984. **179**(1): p. 125-42.
34. Fauchere, J.-L. and Pliska, V.E., *Hydrophobic parameters of amino-acid side-chains from the partitioning of N-acetyl-amino-acid amides*. Eur. J. Med. Chem, 1983. **18**: p. 369-375.
35. Fraga, S., *Theoretical prediction of protein antigenic determinants from amino acid sequences*. Can. J. Chem, 1982. **60**: p. 2606-2610.
36. Grantham, R., *Amino acid difference formula to help explain protein evolution*. Science, 1974. **185**(4154): p. 862-4.
37. Guy, H.R., *Amino acid side-chain partition energies and distribution of residues in soluble proteins*. Biophys J, 1985. **47**(1): p. 61-70.
38. Hellberg, S., Sjostrom, M., Skagerberg, B. and Wold, S., *Peptide quantitative structure-activity relationships, a multivariate approach*. J Med Chem, 1987. **30**(7): p. 1126-35.
39. Hopp, T.P. and Woods, K.R., *Prediction of protein antigenic determinants from amino acid sequences*. Proc Natl Acad Sci U S A, 1981. **78**(6): p. 3824-8.
40. Janin, J., *Surface and inside volumes in globular proteins*. Nature, 1979. **277**(5696): p. 491-2.
41. Kyte, J. and Doolittle, R.F., *A simple method for displaying the hydropathic character of a protein*. J Mol Biol, 1982. **157**(1): p. 105-32.
42. Manavalan, P. and Ponnuswamy, P.K., *Hydrophobic character of amino acid residues in globular proteins*. Nature, 1978. **275**(5681): p. 673-4.
43. Meek, J.L., *Prediction of peptide retention times in high-pressure liquid chromatography on the basis of amino acid composition*. Proc Natl Acad Sci U S A, 1980. **77**(3): p. 1632-6.
44. Miyazawa, S. and Jernigen, R.L., *Estimation of Effective Interresidue Contact Energies from Protein Crystal Structures: Quasi-Chemical Approximation*. Macromolecules, 1985. **18**: p. 534-552.
45. Mohana Rao, J.K. and Argos, P., *A conformational preference parameter to predict helices in integral membrane proteins*. Biochim Biophys Acta, 1986. **869**(2): p. 197-214.
46. Parker, J.M., Guo, D. and Hodges, R.S., *New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites*. Biochemistry, 1986. **25**(19): p. 5425-32.
47. Rose, G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H. and Zehfus, M.H., *Hydrophobicity of amino acid residues in globular proteins*. Science, 1985. **229**(4716): p. 834-8.
48. Roseman, M.A., *Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds*. J Mol Biol, 1988. **200**(3): p. 513-22.
49. Sweet, R.M. and Eisenberg, D., *Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure*. J Mol Biol, 1983. **171**(4): p. 479-88.

50. Welling, G.W., Weijer, W.J., van der Zee, R. and Welling-Wester, S., *Prediction of sequential antigenic regions in proteins*. FEBS Lett, 1985. **188**(2): p. 215-8.
51. Wertz, D.H. and Scheraga, H.A., *Influence of water on protein structure. An analysis of the preferences of amino acid residues for the inside or outside and for specific conformations in a protein molecule*. Macromolecules, 1978. **11**(1): p. 9-15.
52. Wilson, K.J., Honegger, A., Stotzel, R.P. and Hughes, G.J., *The behaviour of peptides on reverse-phase supports during high-pressure liquid chromatography*. Biochem J, 1981. **199**(1): p. 31-41.
53. Wolfenden, R., Andersson, L., Cullis, P.M. and Southgate, C.C., *Affinities of amino acid side chains for solvent water*. Biochemistry, 1981. **20**(4): p. 849-55.
54. Zimmerman, J.M., Eliezer, N. and Simha, R., *The characterization of amino acid sequences in proteins by statistical methods*. J Theor Biol, 1968. **21**(2): p. 170-201.
55. Bairoch, A., *Release notes for UniProtKB/Swiss-Prot release 51.1 - November*. 2006.
56. Bhaskaran, R. and Ponnuswamy, P.K., *Positional flexibilities of amino acid residues in globular proteins*. Int J Pept Protein Res, 1988. **32**: p. 242-255.
57. Chou, P.Y. and Fasman, G.D., *Prediction of the secondary structure of proteins from their amino acid sequence*. Adv Enzymol Relat Areas Mol Biol, 1978. **47**: p. 45-148.
58. Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C., *Atlas of Protein Sequence and Structure*. Vol. 5. 1978.
59. Deleage, G. and Roux, B., *An algorithm for protein secondary structure prediction based on class prediction*. Protein Eng, 1987. **1**(4): p. 289-94.
60. Jones, D.D., *Amino acid properties and side-chain orientation in proteins: a cross correlation approach*. J Theor Biol, 1975. **50**(1): p. 167-83.
61. Jonsson, J., Eriksson, L., Hellberg, S., Sjostrom, M. and Wold, S., *Multivariate Parametrization of 55 Coded and Non-Coded Amino Acids*. Quant. Struct. Act. Relat., 1989. **8**: p. 204-209.
62. Levitt, M., *Conformational preferences of amino acids in globular proteins*. Biochemistry, 1978. **17**(20): p. 4277-85.
63. Lifson, S. and Sander, C., *Antiparallel and parallel beta-strands differ in amino acid residue preferences*. Nature, 1979. **282**(5734): p. 109-11.
64. McCaldon, P. and Argos, P., *Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences*. Proteins, 1988. **4**(2): p. 99-122.
65. Sandberg, M., Eriksson, L., Jonsson, J., Sjostrom, M. and Wold, S., *New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids*. J Med Chem, 1998. **41**(14): p. 2481-91.
66. Guex, N. and Peitsch, M.C., *SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling*. Electrophoresis, 1997. **18**(15): p. 2714-23.
67. Berggren, K., Wolf, A., Asenjo, J.A., Andrews, B.A. and Tjerneld, F., *The surface exposed amino acid residues of monomeric proteins determine the partitioning in aqueous two-phase systems*. Biochim Biophys Acta, 2002. **1596**(2): p. 253-68.
68. Hellberg, S., Sjostrom, M., Skagerberg, B. and Wold, S., *Quant Struct Act Relat*, 1989. **8**: p. 204-209.

69. Rao, M.J.K. and Argos, P., *A conformational preference parameter to predict helices in integral membrane proteins*. *Biochim. Biophys. Acta*, 1986. **869**: p. 197-214.
70. Eriksson, K.-O., *Hydrophobic Interaction Chromatography in Protein Purification: Principles, High-Resolution Methods and Applications*, J.-C. Janson and L. Ryden, Editors. 1998, Wiley-Liss: New York. p. 283.
71. Jesior, J.C., *Hydrophilic framework in proteins?* *J Protein Chem*, 2000. **19**(2): p. 93-103.

## **ANEXOS**

## ANEXO A

### Escalas de propiedades de aminoácidos utilizadas en este estudio

**Tabla A.1:** Escalas de propiedades de aminoácidos utilizadas en este estudio, normalizadas entre 0 y 1

Escala	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
1	0,909	0,571	0,403	0,519	0,052	0,351	0,636	0,766	0,117	0,506	1	0,571	0,143	0,338	0,494	0,727	0,584	0	0,247	0,688
2	0,78	0,484	0,37	0,488	0,05	0,327	0,64	0,683	0,129	0,561	1	0,57	0,141	0,344	0,442	0,692	0,525	0	0,233	0,654
3	0,606	0,383	0,617	0,723	0	0,457	0,511	0,617	0,17	0,202	0,415	1	0,011	0,16	0,415	0,904	0,649	0,053	0,447	0,383
4	0,904	0,436	0,106	0,468	0,138	0,574	1	0	0,457	0,543	0,681	0,628	0,936	0,596	0	0,213	0,277	0,543	0,128	0,521
5	0,811	0,463	0,4	0,547	0,621	0,789	0,968	0,042	0,737	0,474	0,821	0,747	1	0,579	0	0,316	0,316	0,495	0,211	0,411
6	0,985	0,723	0,277	0,427	0,468	0,664	1	0,018	0,505	0,505	0,735	0,672	0,861	0,695	0	0,244	0,29	0,591	0,292	0,492
7	0,361	0,451	0,15	0,038	0,617	0,571	0,15	0,105	0,526	0,842	0,632	0,241	0,504	0,609	0	0,338	0,662	1	0,947	0,835
8	0,147	0,614	0,25	0,216	0,429	0,348	0,315	0	0,575	0,588	0,626	0,325	0,68	0,811	0,186	0,14	0,27	1	0,71	0,483
9	0,25	0,958	0,667	0,875	0,208	0,792	0,833	1	0,083	0,667	0,292	0,708	0	0,042	0,875	0,875	0,583	0,042	0,5	0,375
10	0,346	0,421	0,391	0,128	0,617	0,549	0	0,286	0,376	0,925	0,699	0,278	0,511	0,759	0,135	0,286	0,617	0,752	0,827	1
11	0,306	0,412	0,141	0,094	0,118	0,188	0,129	0,329	0,518	0,953	0,447	0,153	0,388	0,8	0	0,365	0,671	0,588	0,718	1
12	0,22	0,351	0,155	0,116	0,52	0,302	0,132	0,188	0,316	0,897	0,563	0,228	0,531	0,645	0	0,356	0,538	0,591	0,566	1
13	0,174	0,44	1	0,908	0,661	0,468	0,248	1	0,44	0	0,11	0,495	0,119	0,119	0,963	0,881	0,45	0,45	0,615	0,028
14	0,24	0,313	0,58	0,667	0,267	0,38	0,387	0,82	0,18	0,067	0,113	0,367	0	0,12	1	0,607	0,42	0,233	0,427	0,04
15	0,338	0,415	0,822	0,591	0,448	0,467	0,561	1	0,451	0	0,265	0,656	0,121	0,237	0,725	0,664	0,308	0,189	0,343	0,091
16	0,443	0,596	0,516	0,453	0,551	0,605	0,557	0	0,563	0,985	0,985	0,674	0,703	0,898	0,768	0,332	0,677	1	0,801	0,995
17	0,863	0	0,194	0,194	0,29	0,089	0,105	0,911	0,121	0,653	0,903	0	0,113	0,371	0,177	0,605	0,355	0,137	0,169	1
18	0,081	0,169	0,521	0,56	0,246	0,239	0	0,629	0,394	0,16	0,063	0,175	0,063	0,046	1	0,474	0,497	0,231	0,447	0,014
19	0,858	0,041	0,162	0,014	0,797	0,162	0,108	0,676	0,399	0,912	0,932	0	0,899	1	0,304	0,595	0,669	0,615	0,5	0,865
20	0,616	0	0,236	0,028	0,68	0,251	0,043	0,501	0,165	0,943	0,943	0,283	0,738	1	0,711	0,359	0,45	0,878	0,88	0,825
21	0,137	0,107	0,031	0,137	0,233	0	0,176	0,061	0,107	0,924	1	0,195	0,622	0,95	0,435	0,21	0,26	0,828	0,916	0,656
22	0,627	0	0,186	0,237	0,831	0,102	0,288	0,593	0,271	1	0,746	0,034	0,661	0,831	0,288	0,356	0,373	0,441	0,237	0,898
23	0,7	0	0,111	0,111	0,778	0,111	0,111	0,456	0,144	1	0,922	0,067	0,711	0,811	0,322	0,411	0,422	0,4	0,356	0,967
24	0,806	0	0,448	0,417	0,721	0,43	0,458	0,77	0,545	1	0,918	0,263	0,811	0,951	0,678	0,601	0,634	0,854	0,714	0,923

Continuación Tabla A.1

Escala	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
25	0,405	0	0,126	0,074	0,782	0,242	0,113	0,31	0,35	0,862	0,831	0,006	0,687	0,859	0,531	0,298	0,39	1	0,604	0,684
26	0,449	0	0,355	0,28	0,826	0,238	0,268	0,392	0,598	0,754	0,767	0,127	0,868	1	0,293	0,345	0,457	0,6	0,526	0,789
27	0,778	0,148	0,481	0,444	1	0,407	0,407	0,778	0,63	0,926	0,852	0	0,815	0,852	0,556	0,63	0,593	0,778	0,519	0,889
28	0,577	0,006	0,226	0,427	0,605	0,347	0,421	0,489	0,487	0,98	0,98	0	0,709	0,996	0,747	0,321	0,407	1	0,814	0,834
29	0,436	0,179	0,117	0	0,778	0,187	0,214	0,325	0,27	0,992	0,833	0,105	0,728	0,648	0,107	0,078	0,173	0,634	0,529	1
30	0,391	0,202	0,125	0,105	0,819	0,151	0,115	0,252	0,354	0,967	0,908	0	0,987	1	0,151	0,188	0,253	0,775	0,484	0,77
31	0,51	0,2	0,06	0,07	0	0,14	0,18	0,41	0,16	0,93	1	0,13	0,87	0,96	0,49	0,31	0,35	0,92	0,8	0,85
32	0,395	0,29	0,15	0	0,43	0,2	0,11	0,215	0,395	0,9	0,96	0,215	0,71	0,96	0,395	0,175	0,24	1	0,595	0,685
33	0,66	0,176	0,306	0,433	0,763	0,323	0,467	0,557	0	1	0,998	0,061	0,846	0,983	0,768	0,401	0,494	0,914	0,682	0,885
34	0,628	0,163	0,291	0	0,731	0,307	0,05	0,54	0,377	1	0,992	0,153	0,817	0,965	0,751	0,382	0,472	0,879	0,638	0,872
35	0,698	0	0,328	0,023	0,675	0,426	0,167	0,635	0,532	0,928	0,928	0,19	0,789	1	0,794	0,436	0,474	0,977	0,871	0,844
36	0,564	0,308	0,282	0,256	1	0,256	0,256	0,513	0,667	0,923	0,846	0	0,846	0,923	0,308	0,359	0,462	0,846	0,615	0,872
37	0,282	0,223	0,121	0	0,458	0,124	0,028	0,198	0,207	0,793	0,783	0,198	0,721	1	0,254	0,235	0,319	0,56	0,923	0,687
38	0,283	0,364	0,558	0,354	0,62	0,463	0,549	0,712	0	0,867	0,34	0,152	1	0,65	0,524	0,485	0,512	0,611	0,429	0,466
39	0,287	0,217	0,296	0,191	0,861	0,296	0,313	0,417	0,2	0,687	0,887	0	0,53	0,965	0,504	0,261	0,122	1	0,93	0,826
40	0,98	0	0,459	0,402	0,837	0,472	0,436	1	0,433	0,989	0,995	0,466	0,827	0,859	0,893	0,666	0,674	0,629	0,619	0,982
41	0,547	0	0,438	0	0,625	0,438	0	0,469	0,547	0,75	0,75	0	0,672	0,859	0,469	0,422	0,531	1	0,828	0,703
42	0,289	0,117	0,23	0,184	0,207	0,195	0	0,273	0,324	0,754	0,684	0,168	0,57	0,836	0,605	0,148	0,352	1	0,613	0,422
43	0,547	0,557	0,557	0,274	0,318	0,381	0	0,531	0,421	0,969	0,808	0,535	0,682	0,947	0,723	0,569	0,616	1	0,723	0,616
44	0,567	0,545	0,358	0,358	0	0,502	0,212	0,374	0,508	0,788	0,913	0,368	0,573	0,903	0,62	0,336	0,48	1	0,564	0,511
45	0,565	0,192	0,12	0,216	0	0,305	0,072	0,274	0,243	0,541	1	0,188	0,507	0,973	0,49	0,175	0,342	0,873	0,517	0,435
46	0,109	0,767	0,442	0,45	0,357	0,55	0,558	0	0,62	0,434	0,434	0,55	0,574	0,698	0,31	0,233	0,341	1	0,822	0,326
47	0,6	1	0,2	0,2	0	0,2	0,2	0,6	0,2	0,4	1	0,2	0	0,2	0,6	1	0,6	0	0,2	0,6
48	0,306	0,17	0,111	0,094	0,268	0	0,132	0,217	0,043	0,987	0,485	0,132	0,515	0,434	0,03	0,179	0,132	0,26	0,34	1
49	0,605	0,309	0,173	0	0,926	0,309	0,086	0,494	0,321	0,963	1	0,21	0,901	0,963	0,617	0,469	0,543	0,938	0,84	0,877
50	1	0	0,935	0,044	0,972	0,932	0,04	1	0,008	0,998	0,998	0,048	0,973	0,993	0,97	0,968	0,968	0,96	0,969	0,998
51	0	0,236	0,484	0,502	1	0,324	0,335	0,269	0,211	0	0	0,12	0	0	0,142	0,516	0,258	0,047	0,073	0
52	0	0,586	0,552	0,103	0,379	0,31	0	0,207	0,207	0,345	0,241	0,31	0,069	0,103	0,448	1	0,517	0,897	0,207	0,379

Continuación Tabla A.1

Escala	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
53	0,102	0,627	0,282	0,312	0,841	0,406	0,413	0	0,513	0,442	0,448	0,501	0,509	0,691	0,257	0,149	0,259	1	0,741	0,327
54	0,707	0,405	1	0,759	0,017	0,647	0,724	0,267	0,414	0,672	0,19	0,328	0,655	0,198	0,328	0,879	0,681	0	0,198	0,483
55	0,369	0,376	0,142	0,057	0,539	0,39	0,149	0,149	0,376	1	0,638	0,213	0,56	0,603	0	0,298	0,511	0,809	0,801	1
56	0,375	0,321	0,196	0,107	0,929	0,071	0,125	0,179	0,696	0,857	0,821	0	0,804	1	0,071	0,321	0,125	0,982	0,589	0,732
57	0,231	0,225	0,024	0,175	0,403	0	0,178	0,027	0,231	0,836	0,576	0,435	0,443	0,761	0,735	0,019	1	0,019	0,708	0,496
58	0,184	0	0,251	0,099	0,184	0,251	0,099	0,106	0,184	0,24	0,24	0	0,184	0,792	0,184	0,184	0,184	1	0,869	0,24
59	0,639	0,138	0,071	0	0,672	0,253	0,004	0,288	0,195	0,967	1	0,054	0,843	0,918	0,295	0,072	0,22	0,476	0,378	0,868
60	0,592	0,096	0,077	0	0,652	0,148	0,016	0,275	0,2	0,932	0,998	0,082	0,815	0,862	0,294	0,167	0,256	0,48	0,378	0,867
61	0,734	0,12	0,103	0	0,73	0,117	0,024	0,384	0,203	0,982	0,998	0,04	0,892	0,883	0,346	0,183	0,368	0,641	0,543	0,937
62	0,513	0,021	0,306	0,536	0,74	0,324	0,409	0,411	0,49	0,977	0,977	0	0,74	0,995	0,708	0,414	0,515	1	0,786	0,809
63	0,64	0	0,227	0,325	0,826	0,248	0,296	0,546	0,393	0,969	0,969	0,201	0,78	1	0,791	0,532	0,604	0,947	0,408	0,851
64	0,338	0,071	0,113	0,184	0,594	0,337	0,346	0,176	0,035	0,931	0,954	0	0,853	0,949	0,481	0,165	0,413	1	0,943	0,775
65	0,214	0,783	0,681	0,587	0,336	0,535	0,493	0	0,799	0,296	0,353	0,643	0,518	0,721	0,559	0,361	0,23	1	0,754	0,177
66	0,718	0	0,763	0,749	0,903	0,34	0,348	0,455	0,574	0,348	0,436	0,157	0,569	0,672	1	0,715	0,287	0,549	0,514	0,313
67	0,417	0,089	0,049	0	0,342	0,171	0,065	0,165	0,144	0,944	0,915	0,093	0,716	1	0,568	0,196	0,318	0,98	0,588	0,739
68	0,403	0,875	0,756	0,72	0,487	0,654	0,638	0	0,788	0,408	0,481	0,751	0,565	0,739	0,693	0,414	0,363	1	0,852	0,314
69	0,466	0	0,565	0,766	1	0,304	0,445	0,494	0,601	0,318	0,325	0,04	0,4	0,514	0,749	0,53	0,269	0,567	0,456	0,284
70	0,448	0,055	0,112	0	0,376	0,267	0,104	0,231	0,181	0,944	0,99	0,203	0,819	0,983	0,676	0,191	0,387	1	0,782	0,788
71	0,218	0,82	0,71	0,624	0,299	0,57	0,54	0	0,751	0,291	0,345	0,619	0,48	0,75	0,541	0,374	0,235	1	0,813	0,178
72	0,569	0	0,63	0,753	1	0,286	0,47	0,535	0,521	0,248	0,279	0,14	0,551	0,632	0,741	0,645	0,33	0,567	0,545	0,272
73	0,418	0,725	0,272	0,084	0,464	0,245	0	0,32	1	0,317	0,334	0,653	0,718	0,516	0,539	0,238	0,228	0,934	0,444	0,316
74	0,849	0,533	0,916	0,731	0	0,712	0,516	0,488	0,589	0,626	0,751	0,637	0,359	0,437	1	0,714	0,484	0,228	0,254	0,566

**Tabla A.2:** Detalle de las escalas de propiedades de aminoácidos utilizadas en este estudio

<b>Escala</b>	<b>Características</b>	<b>Autor</b>
1	Composición aa	McCaldon y Argos, 1988 [64]
2	Composición aa en SwissProt	Bairoch, 2003 [55]
3	Fracción molar de 3220 residuos accesibles	Janin, 1979 [40]
4	Parámetro conformacional de alfa hélice calculados desde 29 proteínas	Chou y Fasman, 1978 [57]
5	Frecuencia normalizada para alfa hélice	Levitt, 1978 [62]
6	Parámetro conformacional de alfa hélice	Deleage y Roux, 1987 [59]
7	Preferencia conformacional para hoja beta antiparalela	Lifson y Sander, 1979 [63]
8	Area promedio oculta en la transferencia desde estado estándar al de proteína plegada	Rose <i>et al.</i> , 1985 [47]
9	Flexibilidad promedio	Bhaskaran y Ponnuswamy, 1988 [56]
10	Parámetro conformacional para hoja beta calculados desde 29 proteínas	Chou y Fasman, 1978 [57]
11	Frecuencia normalizada para hoja beta	Levitt, 1978 [62]
12	Parámetro conformacional de hoja beta	Deleage y Roux, 1987 [59]
13	Parámetro conformacional de vuelta beta calculados desde 29 proteínas	Chou y Fasman, 1978 [57]
14	Frecuencia normalizada para vuelta beta	Levitt, 1978 [62]
15	Parámetro conformacional de vuelta beta	Deleage y Roux, 1987 [59]
16	Voluminosidad	Zimmerman <i>et al.</i> , 1968 [54]
17	Fracción molar de 2001 residuos ocultos	Janin, 1979 [40]
18	Parámetro conformacional de coil	Deleage y Roux, 1987 [59]
19	Parámetro de hélice oculta en membrana	Rao y Argos, 1986 [45]
20	Hidrofobicidad de L-alfa aminoácidos	Black y Mould, 1991 [28]
21	Hidrofobicidad (Energía libre de transferencia a superficie)	Bull y Breese, 1974 [30]
22	Proporción de residuos un 95% ocultos (considerando 12 proteínas)	Chothia, 1976 [31]
23	Hidrofobicidad	Kyte y Doolittle, 1982 [41]
24	Escala de consenso normalizada de hidrofobicidad	Eisenberg <i>et al.</i> , 1984 [33]
25	Hidrofobicidad	Fauchere <i>et al.</i> , 1983 [34]
26	Hidrofobicidad basada en la energía libre de transferencia en sistemas etanol-agua	Guy, 1985 [37]
27	Energía libre de transferencia desde el interior al exterior de una proteína globular	Janin, 1979 [40]
28	Hidrofobicidad	Abraham y Leo, 1987 [27]
29	Hidrofobicidad promedio en un vecindario	Manavalan y Ponnuswamy, 1978 [42]
30	Hidrofobicidad (Energía de contacto derivada de datos 3D)	Miyazawa y Jerningan, 1985 [44]
31	Movilidad de aminoácidos en papel cromatográfico	Aboderin, 1971 [26]

Continuación Tabla A.2

Escala	Características	Autor
32	Hidrofilicidad derivada de tiempos de retención en HPLC	Parker <i>et al.</i> , 1986 [46]
33	Hidrofobicidad determinada en HPLC a pH 3.4	Cowan y Whittaker, 1990 [32]
34	Hidrofobicidad determinada en HPLC a pH 7.5	Cowan y Whittaker, 1990 [32]
35	Hidrofobicidad ( $\pi$ -r)	Roseman, 1988 [48]
36	Area fraccional promedio perdida	Rose <i>et al.</i> , 1985 [47]
37	Hidrofobicidad comparada optimizada	Sweet <i>et al.</i> , 1983 [49]
38	Valor de antigenicidad	Welling <i>et al.</i> , 1985 [50]
39	Constantes hidrofóbicas derivadas desde tiempos de retención en HPLC	Wilson <i>et al.</i> , 1981 [52]
40	Potencial de hidratación a 25°C	Wolfenden <i>et al.</i> , 1981 [53]
41	Hidrofilicidad	Hopp <i>et al.</i> , 1981 [39]
42	Coeficiente de retención en HPLC a pH 2.1	Meek, 1980 [43]
43	Coeficiente de retención en HPLC a pH 7.4	Meek, 1980 [43]
44	Coeficiente de retención en HFBA	Browne <i>et al.</i> , 1982 [29]
45	Coeficiente de retención en TFA	Browne <i>et al.</i> , 1982 [29]
46	Peso molecular	Varios textos
47	Número de codones	Varios textos
48	Preferencia conformacional por hoja beta paralela	Lifson y Sander, 1979 [63]
49	Polaridad	Grantham, 1974 [36]
50	Polaridad	Zimmerman <i>et al.</i> , 1968 [54]
51	Razón de peso atómico de elementos hetero y carbonos en la cadena lateral	Grantham, 1974 [36]
52	Factores de reconocimiento	Fraga, 1982 [35]
53	Refractividad	Jones, 1975 [60]
54	Mutabilidad relativa	Dayhoff, 1978 [58]
55	Preferencia conformacional para hoja beta total	Lifson y Sander, 1979 [63]
56	Fracción de aminoácidos escondidos en 20 proteínas	Wertz y Scheraga, 1978 [51]
57	Hidrofobicidad medida en base a la energía libre de transferencia en sistemas etanol/agua	Erikkson, 1998 [70]
58	Escala de hidrofobicidad en sistemas ATPS	Berggren, 2002 [67]
59	Composición en una vecindad espacial 4 A	Jesior, 2000 [71]
60	Composición en una vecindad espacial 6 A	Jesior, 2000 [71]
61	Composición en una vecindad espacial 8 A	Jesior, 2000 [71]

Continuación Tabla A.2

<b>Escala</b>	<b>Características</b>	<b>Autor</b>
62	Hidrofobicidad	Abraham y Leo, 1987 [27]
63	Hidrofilicidad	Roseman, 1988 [48]
64	Análisis estadístico de propiedades aminoacídicas (z1)	Jonsson <i>et al.</i> , 1989 [61]
65	Análisis estadístico de propiedades aminoacídicas (z2)	Jonsson <i>et al.</i> , 1989 [61]
66	Análisis estadístico de propiedades aminoacídicas (z3)	Jonsson <i>et al.</i> , 1989 [61]
67	Análisis estadístico de propiedades aminoacídicas (z1)	Hellberg <i>et al.</i> , 1987 [38]
68	Análisis estadístico de propiedades aminoacídicas (z2)	Hellberg <i>et al.</i> , 1987 [38]
69	Análisis estadístico de propiedades aminoacídicas (z3)	Hellberg <i>et al.</i> , 1987 [38]
70	Análisis estadístico de propiedades aminoacídicas (z1)	Sandberg <i>et al.</i> , 1998 [65]
71	Análisis estadístico de propiedades aminoacídicas (z2)	Sandberg <i>et al.</i> , 1998 [65]
72	Análisis estadístico de propiedades aminoacídicas (z3)	Sandberg <i>et al.</i> , 1998 [65]
73	Análisis estadístico de propiedades aminoacídicas (z4)	Sandberg <i>et al.</i> , 1998 [65]
74	Análisis estadístico de propiedades aminoacídicas (z5)	Sandberg <i>et al.</i> , 1998 [65]

## ANEXO B

### Efecto de las escalas de propiedades utilizadas en los descriptores estadísticos

**Tabla B.1:** Desviación estándar del promedio de los valores de los descriptores estadísticos en el set de 74 escalas de propiedades.

		Desviación estándar del promedio en el set de 74 escalas [% valor promedio]			
Radio [Å]	Código PDB	ASPavg	ASPstd	ASPmax	ASPmin
5	1AVU	20,93	21,19	6,38	155,41
5	1F6S	24,00	21,32	9,12	248,90
5	1OVA	23,31	18,90	2,27	192,57
5	1OVT	23,60	19,23	0,60	257,91
5	1SBC	24,79	18,57	9,65	192,31
5	1THV	24,39	17,46	6,53	213,52
5	2CGA	22,70	16,09	2,69	113,38
5	2LYM	25,77	17,79	3,12	208,02
6	1AVU	21,72	21,09	10,07	100,25
6	1F6S	24,97	20,29	9,28	89,94
6	1OVA	26,46	21,52	9,93	97,43
6	1OVT	26,65	21,29	6,66	144,35
6	1SBC	25,57	19,39	12,19	88,67
6	1THV	25,20	18,79	11,13	66,86
6	2CGA	22,21	18,04	6,11	65,19
6	2LYM	29,59	20,35	14,30	67,92
7	1AVU	22,21	19,67	11,24	65,07
7	1F6S	25,18	22,03	11,52	72,87
7	1OVA	27,58	22,14	10,63	66,84
7	1OVT	27,48	21,85	7,82	84,88
7	1SBC	25,00	19,30	14,42	53,79
7	1THV	25,47	19,50	10,97	66,29
7	2CGA	22,29	19,35	8,78	53,15
7	2LYM	27,63	18,87	12,40	52,84
8	1AVU	22,47	20,38	14,43	60,67
8	1F6S	25,44	21,18	13,55	67,29
8	1OVA	27,54	21,85	12,42	57,57
8	1OVT	28,20	23,13	9,94	66,81
8	1SBC	24,86	20,68	15,91	54,03
8	1THV	25,50	19,75	11,57	52,20
8	2CGA	22,08	20,55	10,25	44,29
8	2LYM	27,68	20,72	17,46	46,32
9	1AVU	22,99	21,28	14,57	56,81
9	1F6S	25,95	21,57	13,85	53,04
9	1OVA	28,02	21,48	13,56	62,18
9	1OVT	29,05	24,13	12,04	64,06

Continuación Tabla B.1

		Desviación estándar del promedio en el set de 74 escalas [% valor promedio]			
Radio [Å]	Código PDB	ASPavg	ASPstd	ASPmax	ASPmin
9	1SBC	25,54	21,26	12,82	46,09
9	1THV	25,64	19,77	13,12	41,98
9	2CGA	22,39	21,32	12,82	41,18
9	2LYM	28,31	22,21	20,92	44,88
10	1AVU	22,73	21,26	14,91	45,01
10	1F6S	26,32	23,42	14,80	53,66
10	1OVA	28,75	21,58	14,37	48,94
10	1OVT	29,69	24,76	12,81	60,10
10	1SBC	25,86	23,07	16,21	43,08
10	1THV	25,50	21,61	16,27	44,54
10	2CGA	22,75	21,95	15,03	40,43
10	2LYM	28,10	23,21	17,61	42,38
11	1AVU	22,58	21,61	15,79	44,02
11	1F6S	27,54	22,84	16,65	49,12
11	1OVA	28,83	21,26	14,86	45,89
11	1OVT	30,16	25,00	14,65	57,51
11	1SBC	25,95	23,14	15,51	37,41
11	1THV	25,76	22,42	17,13	39,82
11	2CGA	22,96	22,39	14,73	34,41
11	2LYM	28,55	24,02	20,48	42,48
12	1AVU	22,58	21,33	15,49	37,49
12	1F6S	27,40	23,18	15,96	44,14
12	1OVA	29,39	22,21	18,33	42,34
12	1OVT	30,82	24,92	15,24	56,78
12	1SBC	26,07	24,14	15,85	36,79
12	1THV	25,69	23,97	18,50	37,12
12	2CGA	23,04	21,81	15,69	34,56
12	2LYM	28,78	26,17	24,40	41,53
13	1AVU	22,90	21,15	15,75	35,13
13	1F6S	27,55	24,26	16,91	42,61
13	1OVA	29,67	22,27	19,08	41,39
13	1OVT	31,10	25,24	16,64	54,68
13	1SBC	26,27	25,52	19,03	36,02
13	1THV	25,88	25,14	20,00	37,87
13	2CGA	23,20	22,99	16,40	34,21
13	2LYM	28,74	27,22	22,01	39,17
14	1AVU	23,11	21,67	16,42	34,16
14	1F6S	27,15	24,60	20,33	41,67
14	1OVA	29,95	23,08	22,38	41,20
14	1OVT	31,49	25,22	18,23	54,02
14	1SBC	26,72	26,95	21,25	36,11
14	1THV	25,77	26,22	21,82	36,77

Continuación Tabla B.1

		Desviación estándar del promedio en el set de 74 escalas [% valor promedio]			
Radio [Å]	Código PDB	ASPavg	ASPstd	ASPmax	ASPmin
14	2CGA	23,50	23,21	17,07	34,05
14	2LYM	28,44	28,55	25,40	38,92
15	1AVU	23,07	20,97	16,49	33,65
15	1F6S	27,55	26,06	21,53	37,70
15	1OVA	30,15	24,07	23,31	39,05
15	1OVT	31,87	25,30	20,52	51,51
15	1SBC	26,91	28,07	21,42	34,63
15	1THV	25,91	26,77	21,04	36,49
15	2CGA	23,61	24,53	18,54	33,58
15	2LYM	28,11	33,67	27,00	38,24
16	1AVU	23,21	20,84	17,05	33,60
16	1F6S	27,27	26,92	20,82	36,84
16	1OVA	30,34	24,55	24,99	36,50
16	1OVT	32,21	25,50	22,71	49,22
16	1SBC	27,01	28,32	22,44	34,94
16	1THV	26,25	27,94	22,51	34,59
16	2CGA	23,63	24,72	19,26	33,16
16	2LYM	28,37	34,85	26,21	37,14
17	1AVU	23,47	21,07	17,76	31,33
17	1F6S	27,38	26,72	21,48	35,40
17	1OVA	30,42	25,42	25,57	36,58
17	1OVT	32,33	26,12	22,45	49,08
17	1SBC	26,95	29,20	22,84	34,29
17	1THV	26,42	27,81	21,87	33,15
17	2CGA	23,71	24,51	19,18	31,55
17	2LYM	28,53	38,28	25,96	33,36
18	1AVU	23,72	22,62	18,48	30,58
18	1F6S	27,29	28,72	22,00	35,36
18	1OVA	30,57	27,22	26,66	35,52
18	1OVT	32,45	26,62	21,87	47,94
18	1SBC	27,04	29,92	22,87	34,13
18	1THV	26,49	28,13	21,51	33,17
18	2CGA	23,83	25,41	20,26	30,81
18	2LYM	28,76	40,97	26,55	32,42
19	1AVU	23,69	22,90	18,31	30,34
19	1F6S	27,54	29,16	23,21	35,44
19	1OVA	30,75	28,45	26,81	36,34
19	1OVT	32,52	27,14	24,56	48,12
19	1SBC	27,05	30,13	23,30	34,55
19	1THV	26,48	29,61	22,75	31,89
19	2CGA	23,87	26,85	20,51	30,64
19	2LYM	29,06	40,46	27,03	32,77

Continuación Tabla B.1

		Desviación estándar del promedio en el set de 74 escalas [% valor promedio]			
Radio [Å]	Código PDB	ASPavg	ASPstd	ASPmax	ASPmin
20	1AVU	23,79	23,13	20,00	30,72
20	1F6S	27,46	30,08	23,56	36,19
20	1OVA	30,80	30,40	27,36	36,82
20	1OVT	32,47	27,48	24,92	46,69
20	1SBC	26,96	30,64	23,53	33,48
20	1THV	26,58	29,66	23,63	31,98
20	2CGA	23,91	27,66	21,19	30,59
20	2LYM	29,19	40,67	26,78	32,69
21	1AVU	23,93	24,74	20,82	30,32
21	1F6S	27,69	31,58	23,10	32,02
21	1OVA	30,81	32,03	27,50	36,17
21	1OVT	32,53	28,10	25,93	44,81
21	1SBC	26,97	30,27	23,97	33,30
21	1THV	26,64	30,04	22,18	31,66
21	2CGA	23,97	28,38	21,23	29,11
21	2LYM	29,22	41,36	27,16	32,70
22	1AVU	23,95	25,71	21,13	30,86
22	1F6S	27,79	32,53	23,59	32,45
22	1OVA	30,85	33,96	28,08	36,51
22	1OVT	32,62	28,47	26,41	43,77
22	1SBC	27,02	31,11	24,30	33,35
22	1THV	26,64	29,05	22,71	31,22
22	2CGA	23,92	28,07	21,10	28,31
22	2LYM	29,38	43,02	27,52	32,00
23	1AVU	23,78	25,69	20,62	29,85
23	1F6S	27,95	35,29	23,36	32,55
23	1OVA	30,83	35,85	28,44	36,39
23	1OVT	32,70	28,69	27,67	43,22
23	1SBC	27,15	31,54	24,17	32,12
23	1THV	26,58	30,63	24,14	30,61
23	2CGA	23,96	29,16	20,67	28,30
23	2LYM	29,47	44,03	26,74	31,63
24	1AVU	23,68	25,79	20,98	29,61
24	1F6S	28,06	37,91	25,16	31,92
24	1OVA	30,83	36,40	28,70	36,38
24	1OVT	32,65	29,06	28,12	42,66
24	1SBC	27,21	31,65	24,95	31,20
24	1THV	26,72	31,31	24,59	30,55
24	2CGA	23,94	29,76	21,28	27,61
24	2LYM	29,71	43,95	28,05	31,96
25	1AVU	23,68	25,50	21,22	28,98
25	1F6S	28,25	41,33	25,99	32,32

Continuación Tabla B.1

		<b>Desviación estándar del promedio en el set de 74 escalas [% valor promedio]</b>			
<b>Radio [Å]</b>	<b>Código PDB</b>	<b>ASPavg</b>	<b>ASPstd</b>	<b>ASPmax</b>	<b>ASPmin</b>
25	1OVA	30,86	38,48	28,82	36,72
25	1OVT	32,74	29,30	28,58	43,26
25	1SBC	27,32	31,65	25,05	31,16
25	1THV	26,68	31,47	24,12	30,18
25	2CGA	24,03	30,07	21,34	28,01
25	2LYM	29,86	45,74	28,12	31,67
<b>PROMEDIO</b>		<b>26,80</b>	<b>26,52</b>	<b>19,02</b>	<b>50,16</b>